ABSTRACT

Title of dissertation: DATA-DRIVEN STORYTELLING
FOR CASUAL USERS

Zhenpeng Zhao
Doctor of Philosophy, 2019

Dissertation directed by: Professor Niklas Elmqvist
College of Information Studies

Today's overwhelming volume of data has made effective analysis virtually inaccessible for the general public. The emerging practice of data-driven storytelling is addressing this by framing data using familiar mechanisms such as slideshows, videos, and comics to make even highly complex phenomena understandable. However, current data stories still do not utilize the full potential of the storytelling domain. One reason for this is that current data-driven storytelling practice does not leverage the full repertoire of media that can be used for storytelling, such as speech, e-learning, and video games.

In this dissertation, we propose a taxonomy focused specifically on media types for the purpose of widening the purview of data-driven storytelling by putting more tools in the hands of designers. We expand the idea of data-driven storytelling into the group of casual users, who are the consumers of information and non-professionals with limited time, skills, and motivation , to bridge the data gap between the advanced data analytics tools and everyday internet users. To prove the

effectiveness and the wide acceptance of our taxonomy and data-driven storytelling among the casual users, we have collected examples for data-driven storytelling by finding, reviewing, and classifying ninety-one examples.

Using our taxonomy as a generative tool, we also explored two novel storytelling mechanisms, including live-streaming analytics videos—DataTV—and sequential art (comics) that dynamically incorporates visual representations—Data Comics [1]. Meanwhile, we widened the genres we explored to fill the gaps in the literature. We also evaluated Data Comics and DataTV with user studies and expert reviews. The results show that Data Comics facilitates data-driven storytelling in terms of inviting reading, aiding memory, and viewing as a story. The results also show that an integrated system as DataTV encourages authors to create and present data stories.

DATA-DRIVEN STORYTELLING FOR CASUAL USERS

by

Zhenpeng Zhao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Niklas Elmqvist, Chair/Advisor
Professor Leilani Battle
Professor Ben Bederson
Professor Héctor Corrada Bravo
Professor Matthias Zwicker

Dedication

I dedicate this dissertation to people I loved and those I still love.

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Professor Niklas Elmqvist, for giving me an invaluable opportunity to work on challenging and extremely interesting projects over the past six years. He has always been patient to me throughout the years. He has always made himself available for help and advice and there has never been an occasion when I've knocked on his door and he hasn't given me time. It has been a pleasure to work with and learn from such an extraordinary individual.

I would also like thank all my committee members: Prof. Leilani Battle, Prof. Ben Bederson, Prof. Hector Corrado Bravo, and Prof. Matthias Zwicker. All the professors tried their best to advice me from my preliminary exam to my defense. although Prof. Alan Sussman was unfortunately unable to attend my defense. They encouraged and tried their best to accommodate my schedule.

My colleagues from HCIL also gave me extensive help. They helped me during my research and gave me lots of feedbacks for this dissertation.

It is impossible to remember all, and I apologize to those I've inadvertently left out.

Lastly, thank you all and thank God!

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables xi

List of Figures xii

1 Introduction 1
- 1.1 Motivation: Harnessing the Data Deluge 2
  - 1.1.1 Opportunity: The Scale of Big Data 3
  - 1.1.2 Problem: The Digital Data Divide 4
  - 1.1.3 Setting: Data for Casual Users 6
  - 1.1.4 Solution: Data-Driven Storytelling 7
- 1.2 Background: Data Visualization 8
  - 1.2.1 Definition 8
  - 1.2.2 Examples of Data Visualization 8
  - 1.2.3 Data Visualization for the Masses 9
- 1.3 Background: Data-Driven Storytelling 10
  - 1.3.1 Definition of Storytelling, Story, and Data-driven Storytelling 10
  - 1.3.2 Existing Frameworks 12
  - 1.3.3 Brief History of Storytelling 14
- 1.4 Purpose of This Work 14

2 Related Work 16
- 2.1 Visualization and Visual Analytics 16
  - 2.1.1 Foundational Work on Data Visualization 16
  - 2.1.2 Examples of Classic and Well-known Visualizations 17
- 2.2 Production, Presentation, and Dissemination 18
- 2.3 Visual Communication and Visualization 19
- 2.4 Visual Storytelling 20
- 2.5 Data-Driven Storytelling 21
- 2.6 Comics as a Storytelling Medium 21

2.7 Animated Graphics as a Storytelling Medium 23

3 Taxonomy of Media for Data-Driven Storytelling 25
3.1 Method 25
3.2 Evidence 26
3.2.1 Exhibit 1: Storytelling in Movies and Documentaries 27
3.2.2 Exhibit 2: Dance 29
3.2.3 Exhibit 3: Data-Driven Storytelling in Video Games 30
3.3 Taxonomy Dimensions 31
3.3.1 Audience Cardinality (A) 33
3.3.2 Space and Time (S/T) 34
3.3.3 Visual Components (VC) 36
3.3.4 Data Components (DC) 37
3.3.5 Media Viewing Sequence (SQ) 38
3.3.6 Storage and Persistence (S/P) 39
3.4 Examples Using the Taxonomy (Applications) 39
3.4.1 Data Comics 43
3.4.2 Data TV 44
3.5 Implication for the Taxonomy 45

4 Designing Data-Driven Tools for Casual Users 51
4.1 Casual Users 51
4.2 Casual Information Visualization 51
4.3 Why Data-driven Storytelling for Casual Users 52
4.4 Which Media to Use 53

5 Data Comics 54
5.1 Definition 55
5.1.1 Basic Model 55
5.1.2 Panel Content 56
5.1.3 Characters, Annotation, and Effects 56
5.1.4 Layout Management 57
5.1.5 Viewing 57
5.2 Implementing Data Comics 58
5.2.1 Clipper 58
5.2.2 Decorator 59
5.2.3 Composer 59
5.2.4 Presenter 60
5.3 Telling Stories Using Data Comics 61
5.3.1 Comics Narration 61
5.3.2 Panel Content and Size 62
5.3.3 Textual Narration 63
5.4 DataComics Examples 64
5.4.1 Euro Debt Crisis 65
5.4.2 U.S. Census Population Pyramid 66

5.4.3 Scientific Journal Comparisons . . . 66
5.5 Evaluation . . . 67
5.5.1 Study 1: Authoring Data Comics . . . 67
5.5.1.1 Method . . . 68
5.5.2 Study 2: Presenting Data Comics . . . 70
5.5.2.1 Participants . . . 70
5.5.2.2 Apparatus . . . 70
5.5.2.3 Task and Datasets . . . 71
5.5.2.4 Metrics . . . 72
5.5.2.5 Factors . . . 73
5.5.2.6 Procedure . . . 73
5.5.2.7 Quantitative Results . . . 74
5.5.2.8 Qualitative Feedback . . . 75
5.5.3 Study 3: Partitioning and Sequence in Storytelling . . . 76
5.5.3.1 Participants . . . 77
5.5.3.2 Apparatus . . . 77
5.5.3.3 Task and Datasets . . . 77
5.5.3.4 Metrics . . . 79
5.5.3.5 Factors . . . 80
5.5.3.6 Procedure . . . 80
5.5.3.7 Quantitative Results . . . 82
5.5.3.8 Qualitative Feedback . . . 82

6 DataTV 84
6.1 Design: Supporting Streaming Data Video Production . . . 84
6.2 DataTV: A Streaming Data Video Editor . . . 86
6.2.1 Workflow . . . 87
6.2.2 Source Management . . . 89
6.2.3 Scene Management . . . 89
6.2.4 Video Annotation . . . 90
6.2.5 Recording and Live Streaming . . . 91
6.2.6 Implementation . . . 91
6.3 DataTV Examples . . . 92
6.3.1 Nobel Prize Data Analysis with Keshif . . . 93
6.3.2 Stock Market Data Analysis with TimeFork . . . 96
6.3.3 NY Times Comment Data Analysis with CommentIQ . . . 97
6.4 Qualitative Evaluation . . . 99
6.4.1 Participants . . . 101
6.4.2 Apparatus . . . 102
6.4.3 Tasks . . . 102
6.4.4 Procedure . . . 103
6.4.5 Results . . . 104

7 Discussion 108
7.1 Explaining the Findings . . . 108
7.1.1 DataComics . . . 108
7.1.2 DataTV . . . 109
7.2 Generalizing the Findings . . . 110
7.2.1 DataComics . . . 110
7.2.2 DataTV . . . 111
7.3 Limitations . . . 111
7.3.1 DataComics . . . 111
7.3.2 DataTV . . . 112
7.4 Guidelines . . . 113
7.4.1 Be Open to Unique Media for Storytelling . . . 113
7.4.2 Avoid Relying on Artistic Skill . . . 113
7.4.3 Start from Existing Examples, Don't Be Too Unique . . . 114

8 Conclusion 115

9 Future Work 117

A Survey of Data-Driven Storytelling Media 119
A.1 Storytelling in Movies and Documentaries . . . 119
A.1.1 Stop Marine Plastic Pollution . . . 119
A.1.2 A Beautiful Planet . . . 121
A.1.3 ABC News for Irma Hurricane . . . 123
A.1.4 Is Height All in Our Gene . . . 124
A.1.5 Ancient Greece in 18 minutes . . . 125
A.1.6 The history of Asia: every year . . . 126
A.1.7 Wealth Inequality in America . . . 128
A.1.8 The Joy of Stats . . . 129
A.1.9 Religions and babies . . . 130
A.1.10 Gene Pool Decline . . . 132
A.1.11 How To End Poverty . . . 133
A.1.12 China's Geography Problem . . . 134
A.1.13 Imaginary Numbers Are Real . . . 135
A.1.14 Big Data Revolution . . . 137
A.1.15 The Truth About Population . . . 138
A.1.16 Inside the mind of a master procrastinator . . . 139
A.1.17 How data will transform business . . . 140
A.1.18 Will Saving Poor Children Lead to Overpopulation . . . 142
A.2 Data Comics . . . 143
A.2.1 PhD Comic . . . 143
A.2.2 NFL Player Data . . . 145
A.2.3 Graphic Comic . . . 147
A.2.4 Comic style Dashboard . . . 148
A.2.5 Infographic Comic . . . 150

A.2.6 NYC Restaurant Data Vis Comic . . . 151
A.2.7 Marvel vs DC Comics . . . 153
A.2.8 Body Cartoon . . . 155
A.2.9 Spider Man Comic Visualization . . . 157
A.2.10 Cell Phone Comic . . . 158
A.2.11 Linear Regression Comic . . . 160
A.2.12 Vocation Stress Comic . . . 161
A.2.13 Desk Entropy Comic . . . 162
A.2.14 PhD Grooming Comic . . . 164
A.2.15 PhD Procrastination . . . 165
A.2.16 Day of Life an American . . . 167
A.2.17 Curve Fitting Comic . . . 169
A.2.18 Seashell Probability Comic . . . 171
A.3 Web Article with Data Visualization . . . 172
A.3.1 The Two Americas . . . 173
A.3.2 Strikeouts on the Rise . . . 177
A.3.3 1.5 Million Missing Black Men . . . 179
A.4 Visualization Tools . . . 181
A.4.1 VisJocky . . . 181
A.4.2 ChartAccent . . . 183
A.5 Sketch Tools . . . 184
A.5.1 SketchStory . . . 184
A.5.2 Sketcholution . . . 185
A.5.3 DataSketches– Royal Constellations . . . 186
A.5.4 DataSketches–Carcaptor Sakura . . . 187
A.5.5 The Big Short Movie Explained Animated . . . 188
A.6 Infographics . . . 189
A.6.1 New Orleans Housing Population . . . 190
A.6.2 Top writers for best sellers . . . 191
A.6.3 Public Library Report . . . 193
A.6.4 US Poverty . . . 195
A.6.5 London March . . . 196
A.6.6 Yemen War . . . 197
A.6.7 Space Industry . . . 198
A.6.8 Household Air Pollution . . . 199
A.6.9 Air Pollution Linked Death . . . 200
A.6.10 North and South Korean Comparison . . . 201
A.6.11 In the Shadow of Foreclosure . . . 202
A.6.12 Word of Democrats and Republicans . . . 204
A.6.13 UK and US Firearms . . . 205
A.6.14 White House Correspondent Dinner . . . 206
A.6.15 Day vs. Night: What NYCs Population Looks Like . . . 207
A.6.16 Who owns everything: Big Data Today . . . 208
A.6.17 Big Welsh Coast Walk . . . 210
A.6.18 Hungry USA . . . 211

A.6.19 NYC Celebrity Map . . . 212
A.7 Games . . . 214
A.7.1 Halo: Reach . . . 214
A.7.2 Call of Duty: : Black Ops . . . 215
A.7.3 StarCraft II . . . 216
A.8 Social Media . . . 217
A.8.1 TwitterSheep . . . 217
A.8.2 Twitter Interactive Games of Thrones . . . 219
A.8.3 Twitter Interactive: How tweets spread . . . 220
A.9 Augmented Reality . . . 221
A.9.1 Uber Mobile Visualization . . . 221
A.9.2 AR Data Visualization Design . . . 223
A.9.3 AR 3D Design . . . 225
A.9.4 AR Flight Data . . . 226
A.9.5 AR Street Visualization . . . 227
A.9.6 AR pipeline . . . 229
A.9.7 AR Infrastructure Visualization . . . 230
A.9.8 AR Bio-Chemical Visualization . . . 232
A.10 Virtual Reality . . . 233
A.10.1 Adobe VR Data Visualization . . . 233
A.10.2 VR Baseball training . . . 235
A.10.3 VR Big Data Analysis . . . 236
A.10.4 VR Lens Big Data . . . 237
A.10.5 VR Immersive visualization for Big Data . . . 239
A.10.6 VR Bio-informatics Visualization . . . 240
A.10.7 VR Geo Map Visualization . . . 241
A.10.8 VR Data Analysis . . . 243

B Data Comics Evaluation Protocol 245
B.1 Evaluation: DataComics vs PowerPoint: Test Cases and Scripts . . . 245
B.1.1 Twitter Heatmap for Stocks . . . 245
B.1.2 the U.S. Census Population pyramid . . . 246
B.1.3 World Happiness . . . 247
B.1.4 Star Wars . . . 248
B.2 Evaluation: Single Frame vs Frame Panels: Test Cases and Scripts . . . 250
B.2.1 The Origin of Major Beer Types . . . 250
B.2.2 The Arabic-Israel War . . . 251
B.2.3 Cell Phone Phishing . . . 252
B.2.4 Global Wealthy Population . . . 253
B.3 Evaluation: Expert Review . . . 254

C Data TV Evaluation Protocol 255
C.1 Questions and Protocol . . . 255
C.2 Data Sets . . . 256
C.2.1 Data Set One . . . 256

C.2.2 Data Set Two 256

Bibliography 258

List of Tables

3.1 Classification of our representative sample of different media used for data-driven storytelling–one. 40
3.2 Classification of our representative sample of different media used for data-driven storytelling–two. 41
3.3 Classification of our representative sample of different media used for data-driven storytelling–three. 42

List of Figures

1.1 Genres of narrative visualization by Segel and Heer [2]. 11

2.1 Visualization of locations of cholera cases and wells [3]. 17
2.2 Visualization of Napoleon'ss failed campaign into Russia [4]. 18

3.1 The distribution of all the sources of examples. 26
3.2 Image from the video *A Day in the Life of the (Polluted) Ocean*. . . . 27
3.3 Physical dance demonstration of a bubble sort algorithm. 28
3.4 Data-driven story of a session in the PC game *Civilization 3*. 30
3.5 The figure shows the parallel coordinate graph of the design space of the examples categorized by the taxonomy 48
3.6 Data Comic on the European debt crisis. 49
3.7 Webcam and Keshif visualization being recorded for the Nobel Prize Winner data video. A data video can be made with user as the presenter. 50

5.1 (a) Basic DataComicsJS interface with the Composer in the middle of the workspace and Decorator on the left side. The Decorator is composed of four expendable menus including data importing, operations, comic elements, and collection of clips. (b) Data import is used to load and update visualization clips from a server where the user can also save finished work and reload them for presentation. SVG clips is the content captured by the Clipper. 58
5.2 Viewing individual panels using the Presenter in slideshow mode. Buttons allow for navigating slides in sequence. 60
5.3 (a) The use of text to drive narrative. (b) Encapsulation of moments in panels juxtaposed yield closure as the viewer connects them and sees the whole. (c) Panels organized into tiers with horizontal gutters; multiple tiers form pages with vertical gutters. 60
5.4 Examples of Data Comics. (a) The role of Greece in the European debt crisis. (b) Trends in scientific journals in neuroscience. 64
5.5 Data Comics of the U.S. baby boom of the 1960s. 65

5.6 A comparison of the story frames of Birth data from the U.S. Census Bureau with DataComics and PowerPoint. Stories are composed with different methods but with one-to-one correspondence in details to make the user study fair. 71
5.7 Comparison between DataComics (DC) and PowerPoint (PPT) of subjective ratings (Likert 1-5 scale) for engagement, speed, space-efficiency, ease of use, and enjoyability. 73
5.8 (a) The original infographics without any changes. (b) Adding red dotted boxes to highlight the locations for the panels. 78
5.9 (a) The infographic is partitioned into panels. (b) The infographic is partitioned into panels and captions are added to help address the storyline. 79
5.10 Comparison between single visualization (V), visualization with caption (VC), data comic panels without captions (DC) and data comic panels with captions (DCC) of subjective ratings (Likert 1-5 scale) for engagement, speed, space-efficiency, ease of use, and enjoyability. . 81

6.1 Main user interface of the DataTV prototype tool. A live mode toolbar allows for panning and zooming sources as well as scribbling directly on top of the output. The list panes at the bottom of the interface allow for controlling the scenes and sources being displayed. 87
6.2 DataTV's live mode interface where users can zoom and pan in a data source as well as annotate using a pen, eraser, and highlighter. . 91
6.3 Keshif browser for the Nobel Prize Winners dataset. 93
6.4 A live DataTV session composing a data video using the Keshif system for the Nobel Prize Winners dataset. The user imports an external visualization tool to display the economy of South Africa. 94
6.5 Webcam and Keshif visualization being recorded for the Nobel Prize Winner data video. The author is recording a live video "talking head" view using a webcam and microphone source as input, which is composed into the final video output. 95
6.6 96
6.7 98
6.8 Interface of CommentIQ system supporting multidimensional analysis for online article comments. 99
6.9 DataTV being used to record a streaming data video on race, murder rate, and SWAT activity. 100
6.10 DataTV recording the CommentIQ system being used to filter comments over time. 101
6.11 Expert 1 reviewed the U.S. budget trend in 2013 during the time of President Obama. The expert presented their insights using DataTV. 104
6.12 Expert 2 reviewed the population change of Florida. Insights on housing and population gain were presented using DataTV. 105

A.1 Marine Plastic Pollution [5]. 119

A.2 Scene that shows relations between CO2 and temperature change. . . 120
A.3 Ice are melting with temperature increases. 121
A.4 Scene that seasons are shifting and causing problems for animals . . . 121
A.5 News for Irma Hurricane [5]. Part of Miami will be flooded. 123
A.6 Height change [6] along with time. The average height of human increases with time. 124
A.7 Ancient Greek [7] ruler. 125
A.8 The change of Asia [8] viewed from a map. 126
A.9 The documentary [9] shows how one percent of the population occupies a large percent of wealth. 128
A.10 The documentary [10] shows how data visualization works. 129
A.11 The documentary [11] shows how numbers of babies varies with religion.130
A.12 The documentary [12] shows how gene pool of human declines. 132
A.13 The documentary [13] shows how to end poverty. 133
A.14 The documentary [14] shows what China's problem with its neighbours.134
A.15 The documentary [15] shows where real number comes from. 135
A.16 The documentary [16] shows how the revolution of big data take place.137
A.17 The documentary [17] shows the relation between wealth and the size of population. 138
A.18 The documentary [18] shows how exactly some procrastinator thinks. 139
A.19 The documentary [19] shows how big data transform business 140
A.20 The video [20] shows how the poor has more children than others. . . 142
A.21 Phd comic [21]. The ambition decreases along time. 143
A.22 NFL player report [22] 145
A.23 Graph comic for European relations. 147
A.24 Comic style dashboard [23] 148
A.25 Infographical Comics [24] 150
A.26 Comics for data of restaurants in NYC [25] 151
A.27 Comics of the comparison of characters from Marvel and DC [26] . . 153
A.28 Comics of someone having tatoo [27] on his arm as data visualization. 155
A.29 Spider Man Visualization in Comics [28] 157
A.30 Cell Phone Visualization in Comics [29] 158
A.31 Linear Regression Visualization in Comics [30] 160
A.32 Visualization to show the stress level change with vocation in Comics [31]161
A.33 The Increase of Entropy of Visualization in Comics [32] 162
A.34 The Need of Grooming of Visualization in Comics [33] 164
A.35 The Change of Procrastination and Stress Level in Comics [34] 165
A.36 The Change of Procrastination and Stress Level in Comics [35] 167
A.37 Curve Fitting Visualization in Comics [36] 169
A.38 Illustration of Conditional Probability Visualized in Comics [37] . . . 171
A.39 The Two Americas 173
A.40 The Two America: Trump. 174
A.41 The Two America: Cliton 174
A.42 The Two America: description of each side. 175
A.43 The Two America: description of Cliton side 175

A.44 The Two America: comparison of two sides. 176
A.45 Strikeouts on the rise 177
A.46 Strikeouts on the rise [38] for each player. 178
A.47 Missing men for different races [39] 179
A.48 Which areas have men missing 180
A.49 Missing men for blacks and whites. 180
A.50 VisJocky interface [40] 181
A.51 ChartAccentc interface [41] 183
A.52 Process of design story with SketchStory 184
A.53 Sketcholution comic strip and summary. [42] 185
A.54 The sketch visualization of the royal members [43]. 186
A.55 The sketch connects the royal members 186
A.56 Visual Explanation of the Relationship of a Cartoon Series [44] 187
A.57 How mortgage bond combined into sub-prime mortgage [45]. 188
A.58 New Orleans housing population decreases [46] 190
A.59 Top writers for best sellers [47] 191
A.60 Public library report [48] 193
A.61 How poverty is distributed in the U.S. [49] 195
A.62 Distribution of London Marches in Visualization [50] 196
A.63 Geo-distribution of All the Forces in Yemen Civil War in Visualization [51] 197
A.64 The Stats of Space Industry and Technology in Visualization [52] . . 198
A.65 The Source of Household Air Pollution in Visualization [53] 199
A.66 The Source of Death Caused by Air Pollution in Visualization [54] . . 200
A.67 The Comparison of North and South Korea in Visualization [55] . . . 201
A.68 The situation of foreclosure across the U.S. [56] 202
A.69 The comparison for the Democrats and Republicans during election [57]204
A.70 The comparison of the firearm possession of the U.K. and the U.S. [58]205
A.71 The guest distribution of the White House Correspondent Dinner [59] 206
A.72 The change of population during days and nights [60] 207
A.73 The wealth distribution [61] 208
A.74 The participants and locations of the big Welsh coast walk [62] 210
A.75 Public library report [63] 211
A.76 The locations of celebrities' home [64] 212
A.77 Visualization in Halo: Reach [65] 214
A.78 Visualization in Call of Duty: : Black Ops [66] 215
A.79 Visualization in StarCraft II [67] 216
A.80 TwitterSheep interface [68] 217
A.81 Twitter interactive Games of Thrones [69] 219
A.82 Twitter Interactive [70]: how world cup news spread. 220
A.83 The architecture for the augmented reality [71] to have a data visualization 221
A.84 The architecture for the augmented reality [72] to have a data visualization 223
A.85 How AR [72] is used to visualize the world population 224

A.86 The process for the engineers with augmented reality [73] to design a building with data visualization 225
A.87 The process for passengers to view flight data in augmented reality [74] 226
A.88 Street viewers obtain information from augmented reality [75] on the street 227
A.89 Engineers have the pipes shown with augmented reality [76] 229
A.90 Engineers have the underground infrastructure shown with augmented reality [77] 230
A.91 A Bio-chemistry researcher has a structure of molecular shown with augmented reality [78] 232
A.92 How virtual reality tool [79] is used to visualize the baseball training data. 233
A.93 How virtual reality tool [80] is used to visualize the baseball training data. 235
A.94 How virtual reality tool [81] is used to visualize big data in 3D. . . . 236
A.95 How virtual reality tool [82] is used to visualize big data with virtual objects. 237
A.96 How virtual reality tool [83] is used to do data analytics. 239
A.97 How virtual reality tool [84] is used to visualize gene information with graph in 3D. 240
A.98 How virtual reality tool [85] is used to visualize geographical information. 241
A.99 How virtual reality tool [86] is used to do data analysis. 243

B.1 Stock heat map in comic style 245
B.2 US census data in comic style 246
B.3 World happiness data in comic style 247
B.4 Star data in comic style 249
B.5 The geo origin of beer around world in sequenced panels 250
B.6 The Arabic-Israel War in sequenced panels 251
B.7 The data about cellphone phished in sequenced panels 252
B.8 The data about global wealthy population in sequenced panels 253

Chapter 1: Introduction

If raw *data* is the crude oil of the information age, then digital tools are the refineries needed to turn data into *information*—or crude into gas—ready for consumption. Unfortunately, current data science tools typically require significant expertise far beyond that of the normal citizen. *Visualization*, which uses interactive graphical representations of data to aid cognition [87], has the potential to lower the threshold of understanding data by virtue of using visual representations that are both accessible to casual users [88], as well as scale gracefully to large data [89]. However, even interactive visualization tools such as Tableau [90], Spotfire [91], and QlikView [92], which provide point-and-click interfaces on standard computers and devices, require specialized knowledge in mathematics, statistics, and data management as well as access to clean, complete, and properly formatted data sources. These barriers are essentially insurmountable to the average person, or what we here call a *casual user* [88]: a person with average knowledge in science, technology, and mathematics, with normal computer savvy, and with moderate interest in harnessing data to enrich or enhance their lives.

This dissertation proposes methods for data-driven communication [93] for casual users [88] using so-called *data-driven storytelling* [2]: interactive visualization in conjunction with storytelling methods. The focus is primarily on casual users as consumers of information, not as authors.

In this chapter, we set the stage for this dissertation: the data deluge in today's society, the use of storytelling to make all this data accessible to casual users, and the gaps in current data-driven storytelling practice that this work will address.

1.1 Motivation: Harnessing the Data Deluge

Our society is under a veritable *deluge* of data [94]; we are inundated with massive, detailed, and complex datasets, and we need digital tools to help us stay afloat. This deluge is a double-edged sword: It's an *opportunity*, because never before in human history has so much of our world and our society, of science and of arts, of medicine and of life itself been captured by sensors and stored in machine readable form. If we were to harness this data effectively, there is tremendous potential in improving the collective lives of millions—even billions—of people around the world. For example, reserving informative patterns and knowledge for big data provide the public sector an opportunity to improve productivity and higher levels of efficiency and effectiveness [95, 96]. The opportunities lie in fields such as E-health, Internet of Things, public utilities, transportation, logistics, public services, government monitoring, and so on [97].

It's a *challenge*, because the data is complex, large, heterogeneous, uncertain, and fleeting [98]. Extracting actionable information from such data requires a careful and deliberate approach. When handling big data problems, there are difficulties in data capturing, storage, searching, sharing, analysis, and visualization [96,99]. Furthermore, in a world of ever-increasing surveillance, there are several ethics concerns that must also be considered [100,101].

1.1.1 Opportunity: The Scale of Big Data

Today's world is facing the problem that the amount of available data is exploding [102]. Until 2003, humans had created approximately 5 exabytes (10^{18} bytes) of data [103]. By 2012, this amount of information was created in two days, and the digital world of data had expanded to 2.72 zettabytes (10^{21} bytes) [103]. In 2012, IBM indicated that every day 2.5 exabytes of data was created and 90% of the data produced by then was produced from 2010 to 2012 [98]. By the year 2020, 50 billion devices will be connected to networks and the Internet [104]. According to estimates, the volume of business data worldwide, for most companies, doubles every 1.2 years [95, 96]. Several resources exist surveying the scope of big data [94,105]. To take YouTube as an example, in 2014, there were about 300 hours of video uploaded to YouTube every minute [106], and almost 1 billion minutes of videos watched on YouTube every single day [107].

In an era when a common smartphone holds more computing power than all the computing power of NASA in 1969 combined [108], it is no surprise that humans are producing a huge amount of data everyday and everywhere. Accordingly, the ability required to handle big data is increasingly demanding for virtually all fields, including physics simulation, finance, business, and personal care [94,105]. Furthermore, beyond sheer scale, the data comes in many shapes and forms, can be faulty, and is produced at significant speed.

Obviously, the opportunity in harnessing this data for governmental, societal, humanistic, or business interests is significant. Many of these approaches are automatic and based on machine learning and artificial intelligence. For example, today's recommendation engines know the preferences of their customers, and are able to anticipate which products the consumer will want to buy [109]. YouTube, as the largest and most popular online video community, can analyze personal preferences of its large user group [110]. Modern web services, such as Amazon Web Services [111], Microsoft Azure [112], and Google Cloud [113] have successfully served businesses across the world with very large data volume and storage requirements.

1.1.2 Problem: The Digital Data Divide

For the average American who is not specialized in data analysis, dealing with all this data is increasingly challenging. In today's information society, it is difficult to be an effective citizen without coming into contact with the internet and the digital sphere. Most people have to deal with different kinds of data in their daily lives, from health data collected by sport apps on their smartphones, to stock market data governing their finances [102]. They have to use the internet to balance their checkbooks, pay their bills, and manage their mortgages. In this case, the amount and type of data makes it increasingly difficult for the average American to handle.

Given the ubiquity of data, there is a risk that people not capable of analyzing and understanding data will fall behind in society [114]. This effect is only exacerbated by the increasing scope and magnitude of the aforementioned data deluge and the lack of effective, accessible, and approachable tools for casual data analysis. We refer to the gap between the difficulty of processing large amount of data and the lack of data analysis skill and processing power of casual users as the *digital data divide*.

Overcoming this barrier requires tackling both the lack of ICTs (information and communication technologies, akin to the original definition of the "digital divide" [115,116]), as well as the lack of specialized data analytics knowledge. For the former, technology progress is fortunately helping; there is more computing power available for casual users than ever before. A Samsung flagship smartphone today possesses more floating point computing power than the supercomputer Deep Blue [117, 118], which beat the world champion in chess [118], and mobile devices are coming down in price while the interfaces are improving. Even elderly users can make full use of most apps on a smartphone with basic training. These factors all contribute to shrinking the divide in terms of device access.

However, this still leaves a significant gap in data analytics capability. Most current data science tools are simply not designed for novice users. One possible way to overcome this is to leverage *data visualization*: the use of interactive graphical representations of data to amplify cognition [87]. Unfortunately, even commercial point-and-click data visualization tools—such as Tableau [119], Spotfire [120], and Qlik [92]—rely on significant training and intimate knowledge of many advanced mathematical and statistical concepts.

1.1.3 Setting: Data for Casual Users

While visualization is more approachable than typical data science tools, visualizations must still be deliberately designed to be appropriate for *casual users*: non-professionals with limited time, skills, and motivation. In 2007, Pousman et al. [88] proposed the topic of *casual information visualization* as visualization systems that are not designed for professional work tasks. They went on to paint a vision for using computer mediated-tools to depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations.

Many examples of casual visualizations exist; we give a sampling here. Some help normal citizens understand complex scientific phenomena [121]. Embedding visualization into presentation software such as PowerPoint [122], Tableau [123], and Excel [124] can give people who are not familiar with visualization skills great leverage over interpreting and analyzing large amount of data. Other tools dealing with data for everyday users are also rich in visualization applications. For example, line charts can be used to show the moving average of changes in the stock market [125], health apps use histograms and pie charts to show the workout daily distribution [126], and personal accounting software uses visualizations such as line charts, treemaps, and histograms to show the annual change of a person's financial situation [127].

However, while casual visualization is a powerful method for conveying data to the average user, this entire topic represents a large design space. Applying casual visualization to aid data-driven communication requires adopting a specific strategy. In this dissertation, we use storytelling for this purpose.

1.1.4 Solution: Data-Driven Storytelling

Data-driven storytelling [128] is the use of traditional storytelling techniques—such as from oral narratives, fiction, and film—to convey narratives about data. This approach to visual communication [93] is particularly appealing because of its high familiarity and accessibility to casual users. Thus, our focus here is on **data-driven storytelling for casual users**. As defined above, a casual user is a non-professional consumer of data that has limited time, skills, and motivation. In addition, our emphasis here is primarily on consumers as casual users; the authors of a data-driven story may still need to be an expert.

Data-driven storytelling has seen increasing popularity in the visualization and human-computer interaction communities. Using methods such as narration, animation, cinematography, art, and design, these methods aim to shape findings from complex and large-scale data into a story form that is more amenable to human understanding than typical data displays. In particular, a recent book on data-driven storytelling by Riche et al. [128] reviews this burgeoning field. However, the repertoire of available media types for such data-driven stories is still limited.

1.2 Background: Data Visualization

Here we take a step back and review the research area of data visualization and its related topics and technologies.

1.2.1 Definition

Data visualization represents data using visual displays to improve cognition [129]. It can also be considered as a medium to capture and share thoughts on data with others [130]. More specifically, it is a process of exploration, analysis, and presentation for data with graphics information and tools [131, 132].

1.2.2 Examples of Data Visualization

Data visualizations are designed for both visualization professional and casual users to understand and analyze data. While visualization has been around for a long time [133], the academic field was formally founded in 1987 with a special volume on the use of computer graphics for scientific and engineering applications [134]. Here we review some representative examples of data visualization.

Multiple visualization tools exist, both commercial and academic. Tableau [90, 135] is a popular commercial visual analytics tool used for creating visualizations. This tool is mostly used by people with background of business analytics or information visualization. Additional commercial tools include Spotfire [91], Qlik [92], and Keshif [136].

In the last decade, visualization has increasingly moved from personal computers to a web-based ecosystem. D3.js [137] is the primary JavaScript package used by front-end developers to create attractive data visualizations. Google Charts [138] is another widely used package for front-end development to create rich interactive data visualizations. The most recent development is the introduction of visualization grammars, such as Vega [139] and Vega-Lite [140], which enable creating charts using declarative specifications.

1.2.3 Data Visualization for the Masses

More and more people are facing the need to analyze, present and interpret data [141]. Data visualization is not only used by professionals, but also by casual users who adopt visualization as a tool to improve presentation, exploration, and analysis. Wattenberg et al. [142] gave an example of how a visualization of baby names can bring together communities of people on the internet. Many Eyes [143] was one of the first visualization systems "for the masses" that enabled people, often casual users, to collaborate on data using just their web browsers. This has shown that visualization has been both a challenge and an opportunity for the masses when dealing with large range of visualization tools to handle and analyze data. Viegas and Wattenberg call this idea *communication-minded visualization* [93].

Today, data visualization has affected lives in many different ways in education, medicine, finance, meteorology, astronomy, etc. We have our smartphones loaded with apps with data visualization such as Uber, Google Map, and stock market apps. We see people tweet about their workout condition with visualization of their physical data. We use Google Map everyday for navigation. We predict stock price, house price, and other price trends with line charts or other dashboards. We also use data visualization to present data in our PowerPoint slides.

Adopting visualization for a mass audience requires adopting casual visualization practices. Freyne et al. [114] note that next generation visualization tools are set to play an important role to overcome the challenge of large amount of raw data, but current generation of visualization tools are sometimes too complex for typical users. Infographics is a widely used media that primarily consists of data visualization, and tools such as Infogram [144] provides means to create such artifacts easily. More effort is needed to make this kind of mechanisms accessible to casual users.

1.3 Background: Data-Driven Storytelling

In data-driven storytelling, we harness the age-old practice of storytelling to create narratives about data. This section describes the definition, background, and existing work for data-driven storytelling.

1.3.1 Definition of Storytelling, Story, and Data-driven Storytelling

Storytelling is the conveyance of a sequence of events (often involving characters and places)—*stories*—using speech, sound, visuals [145], and other multisensory stimuli, and has a history spanning thousands of years [146, 147]. It is one of the oldest form of human communication, record-keeping, and entertainment, well predating the written word [146–148]. Stories—sequences of events involving characters

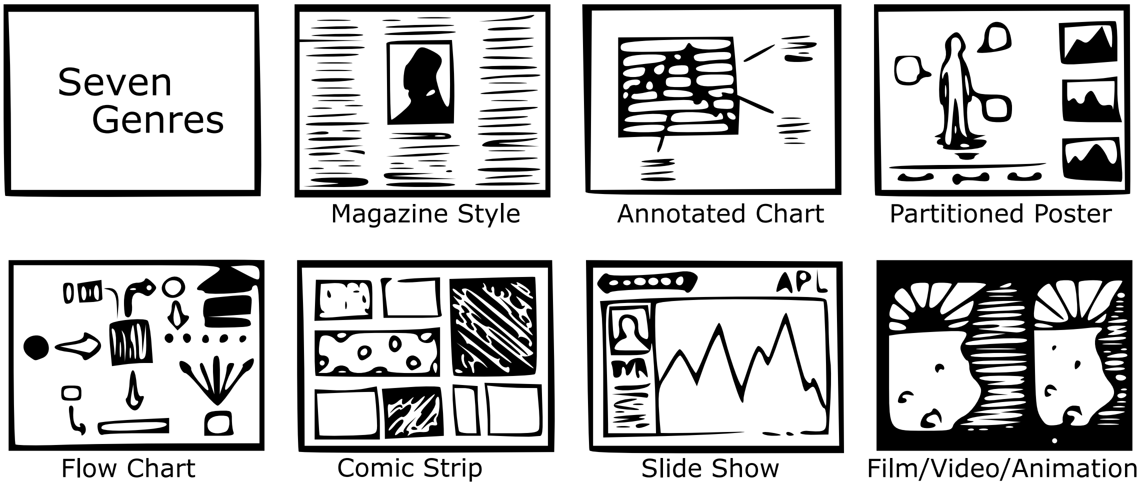


Figure 1.1: Genres of narrative visualization by Segel and Heer [2].

and places—are particularly well suited for this purpose because their chronological structure enables memorization and recall, entices listeners, and facilitates understanding [145, 148]. For this reason, narration and storytelling retain important roles even in today's information society, where these properties are particularly important in helping people understand an increasingly complex world. This has recently given rise to *data-driven storytelling* where narrative techniques are utilized for telling stories about data [2], often using visual media [149, 150].

A *visual narrative* is a story told primarily using visual media, such as illustrations, photographs, animations, video, and—now—visualization [149, 150]. In particular, visualization has a specific proclivity for communication by virtue of its graphical form, yielding the notion of *communication-minded visualization* [93] to support collaborative analysis. Combining the idea of communication-driven visualization with storytelling yields the notion of *data-driven storytelling*: narrative techniques for telling stories about data [2, 128].

1.3.2 Existing Frameworks

Our focus in this thesis is on exploring communication mechanisms—or *media*—used for conveying data-driven stories that are specifically suitable for casual users. Here, we define “media” as the channel or the tools used to store and deliver information. While data-driven storytelling is a nascent research topic in visual communication and visualization, there has so far been no specific focus directed to the specific media used. Instead, existing efforts tend to revolve around the seven genres of narrative visualization proposed by Segel and Heer [2] (Figure 1.1):

- **Magazine Style:** A data-driven image integrated in a page of text, where the text refers to and explains the image.

- **Annotated Chart:** Chart adorned with descriptive text and labels for the purpose of explaining its contents.

- **Partitioned Poster:** A poster or dashboard consisting of multiple images, each with separate data.

- **Flow Chart:** Visually ordered sequence of images and annotations designed to tell a story.

- **Comic Strip:** Sequence of frames containing images and text organized in a comic-style strip layout.

- **Slide Show:** Deck of slides combining images and text to sequentially tell a story.

- **Film/Video/Animation:** Motion graphics that incorporate data-driven imagery and visualizations, often animated.

However, as readily admitted by Segel and Heer, their findings are limited to a sample of 58 examples. They also do not claim that their genres are exhaustive, noting for example that their work did not include video games or e-learning tools. Furthermore, the above seven genres conflate the media used for storytelling with the format, method, and components employed. The seven genres have proven to be extremely powerful for categorizing research in this field. They have even played a prescriptive role, with Graph Comics [151] and Data Comics [1, 152] arising as examples of the Comic Strip genre, Data Videos [153, 154] drawing inspiration from the Film/Video/Animation genre, and infographics [155] from the combination of text and visualization However, there certainly is room for expanding the framework further.

This suggests that the community may be limiting itself by needlessly adhering to a framework that was intended to be generative rather than delimiting. What about using the spoken word for data-driven storytelling, i.e., supporting speakers talking to an audience? What about the written word, i.e., data-driven prose, such as for inclusion in a textual report? Much innovation remains in data-driven storytelling, but this requires going beyond existing frameworks.

1.3.3 Brief History of Storytelling

From the hunter returning from his latest foray to tell tall tales of stalking his prey, to the shaman spinning a yarn about the origins of the gods, the stars, and the moon, *storytelling* is one of the oldest ways for people to record, communicate, and inherit information. There are drawings portraying fighters against wild animals before stone ages. The famous Homeric Hymns has provided us countless inspirations for ancient Greek history. We learned the ways how ancient people were thinking and behaving by all kinds of historical literature in the form of stories. Oral, written, and drawn artifacts were the three major ways of storytelling in ancient times.

Modern technologies have helped storytelling evolve into multimedia forms including audio, video, games, etc. Recently, virtual reality and argument reality have joined the arsenal of methods and media for storytelling.

1.4 Purpose of This Work

The overall purpose of this Ph.D. thesis is ***to expand the horizon of media for data-driven storytelling to aid casual users viewing, analyzing, and understanding data.*** To achieve this, our work collects extensive examples and explores the use of different media types for this purpose. From the taxonomy and guideline derived, we investigate media types particularly useful for casual users with little professional training or background in data visualization and analysis.

In particular, based on recommendations from Segel and Heer [2], this work was the first to propose a practical system for leveraging comics (sequential art [156]) for data-driven storytelling [1] (first submitted in September 2014). Furthermore, our work is the first to propose live-streaming data videos for this purpose (in submission). Finally, our taxonomy reviewing media types for data-driven storytelling is, to the best of our knowledge, the first of its kind (in submission).

Chapter 2: Related Work

In this chapter, we will examine the literature that helped setup the background, definition and other relevant materials.

2.1 Visualization and Visual Analytics

Visualization is defined as translation from data to image [157], or the use of interactive graphical representations of data to aid cognition [87]. Visual analytics is an analytics discourse to make the processing of information and data more transparent [158]. In "Illuminating the Path" [159], visual analytics is defined as analytical reasoning science with interactive visual interfaces.

2.1.1 Foundational Work on Data Visualization

There are many early attempts on building taxonomies of visualization system for its data [160] and algorithms [161]. The design space of information visualization has laid foundations for research on the components of visualization [162]. Both information visualization and visual analytics are possible ways to handle the problem of information overload [158].

2.1.2 Examples of Classic and Well-known Visualizations

While only recently recognized as a research field in its own right, visualization has a long history [133]. For instance, the famous Napoleon March Map by Charles Minard [4] in Figure 2.2 is a well-known example of using visualization to tell the story of Napoleon's failed campaign into Russia. John Snow's cholera map [3] in Figure 2.1 is an early example of using a dot graph to map the relation between cholera death and location of water wells. The diagram of the deaths during the Crimean war by Florence Nightingale shows the relation between causes of death and time [163].

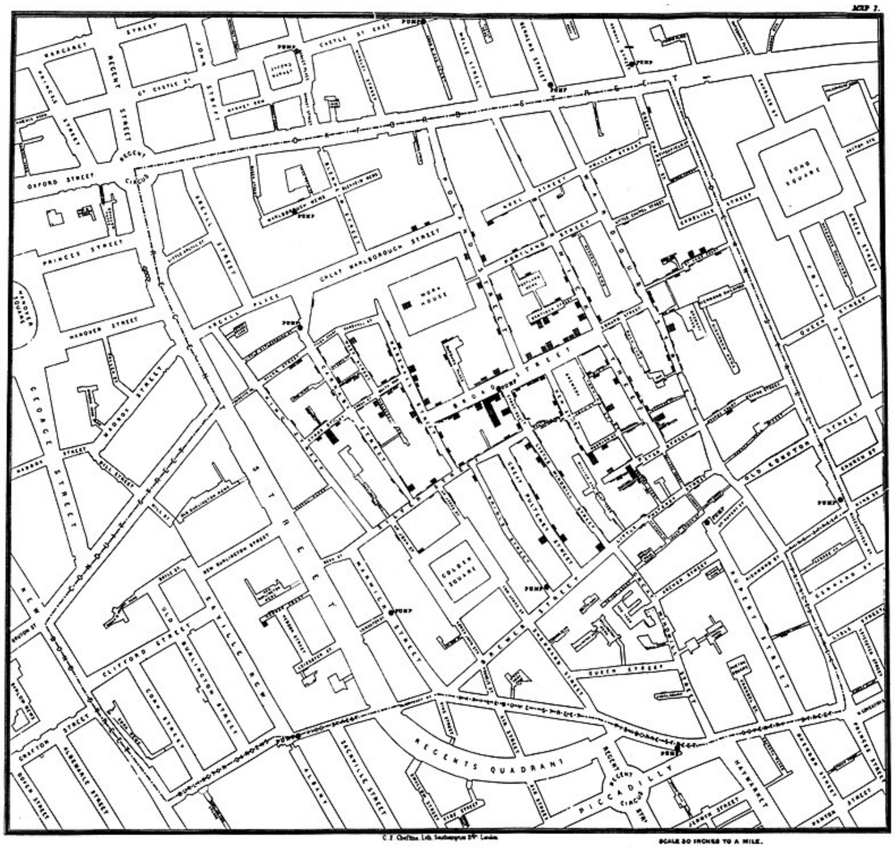
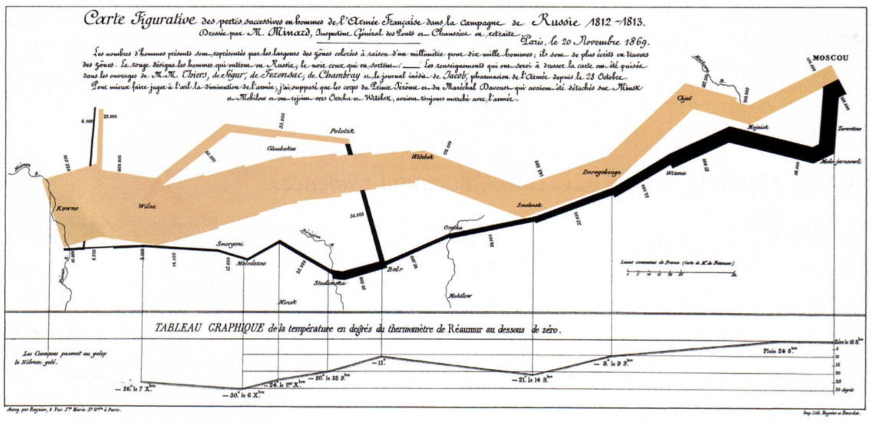


Figure 2.1: Visualization of locations of cholera cases and wells [3].


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.
Paris, le 20 Novembre 1869.
MOSCOU
Kowno
Wilna
Polotzk
Gloubokoe
Witebsk
Smolensk
Dorogobouge
Chjat
Mojaisk
Wizma
Tarentino
Malo-jarosewli
Orscha
Bobr
Studienska
Minsk
Mohilow
Smorgoni
Molodezno
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
– 26° le 7 X.bre
– 30° le 6 X.bre
– 24° le 1er X.bre
– 20° le 28 9.bre
– 11°
– 21° le 14 9.bre
– 8° le 9 9.bre
Pluie 24 8.bre
Zéro le 18 8.bre
30 degrés

This map drawn by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen, the thick band shows the size of the army (422,000 men) as it invaded Russia. The width of the band indicates the size of the army at each position. In September, the army reached Moscow with 100,000 men. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. The remains of the Grande Armée struggled out of Russia with 10,000 men. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn. Napoleon's March poster $14 postpaid; English/French version $18 postpaid.

Figure 2.2: Visualization of Napoleon'ss failed campaign into Russia [4].

2.2 Production, Presentation, and Dissemination

The research agenda for visual analytics called for a focus on the production, presentation, and dissemination of analytics results to stakeholders, policymakers, and the public [159].

- **The case for communication:** Viégas and Wattenberg remarked upon the proclivity of visualization for communication by virtue of its graphical form. They encouraged focusing on so-called *communication-minded visualization* [93] where communication enables collaborative analysis.

- **Embedded dissemination:** To reach its full potential, communication capabilities should be integrated into the visualization tools themselves [159]; for example, Tableau now incorporates the story points feature [119], and most commercial tools support exporting interactive dashboards and workspaces to the web.

- **Literacy and casual viewers:** Presenting insights from data to the masses requires taking the visualization literacy [164] of everyday users into account. Thus, the notion of "casual visualization" [88] is important.

2.3 Visual Communication and Visualization

Visual aids such as images, signs, typography, icons, and drawings have long been used as a particularly effective medium for communication [150]. Beyond human perceptual factors, part of the reason for this effectiveness is the mutual knowledge, mutual beliefs, and mutual assumptions that visual communication enjoys. These mutually agreed-upon conventions allow a particular medium—such as the visual—to encode significant amounts of information with minimal resources given.

Visualization is a particular form of visual language traditionally used for solitary sense-making. The notion of communication-minded visualization (CMV) [93] builds on ideas from visual communication by noting that visualization can often be used for more than just individual insights. There are many examples of CMV systems that exist as precursors to storytelling, including Themail [165], the Baby Name Explorer [142], and Isis [166].

2.4 Visual Storytelling

Storytelling is when a sequence of events are conveyed using plots, locations, and characters. It is a particularly effective communication medium because of the typically high degree of common ground shared between narrator and listeners. Visual storytelling draws upon imagery—both static and dynamic—for this purpose, and includes media such as film, television, animation, design, and even art.

Already in 2001, Gershon and Page [167] suggested using storytelling in visualization to improve its use for visual communication. In fact, the newly popular infographics practiced on the web is built on many such storytelling principles. Despite this, it is only recently that data-driven storytelling was fully embraced by the visualization community, such as the survey by Segel and Heer in 2010 [2], and successful workshops at the annual conference in 2010 and 2011 [168, 169]. Hullman and Diakopoulos followed this up by studying how framing, context, and design impact the rhetoric of a narrative [170]. Since then, several practical methods and techniques have been proposed, including using free-form sketching for narration [171], story points in Tableau [119], and automatic spatialization for visual exploration [172]. Most recently, Hullman et al. [173] studied sequence in narrative visualization, proposing a graph-driven approach for transitioning between views to minimize load on the viewer.

2.5 Data-Driven Storytelling

Combining the idea of communication-driven visualization with storytelling yields the notion of *data-driven storytelling*: narrative techniques for telling stories about data [2]. Gershon and Page first proposed using storytelling for visualization [167], and their work has since been followed up by workshops [168, 169], surveys [2, 170], and even commercial tools [119].

The purview of data-driven storytelling has quickly grown, from dashboards and slideshow presentations [119] to more esoteric formats such as sketching [171], journeys in time and space, and even comics [151, 174]. In their survey of narrative visualization, Segel and Heer [2] identified seven genres—magazine style, annotated charts, partitioned posters, flow charts, comic strips, slideshows, and videos—and also suggested that future storytelling approaches may combine genres.

2.6 Comics as a Storytelling Medium

Comics are often defined as sequences of images—"sequential art" [175] or "sequential images" [176]—that combine to tell a story using graphical means [149]. It is therefore a visual communication medium. Because of its familiarity for many, it enjoys a high level of common ground for conveying information. Furthermore, the visual language of comics is often clear, concise, and intuitive [176, 177].

How comics affect the reader has long been a topic of study. Dorfman et al. [178] discussed the ideology in Disney comics from the perspective of culture and economics. McCloud [175] built on these ideas by suggesting that the engagement in comics mainly arises from the simplified and non-photorealistic appearance of faces and characters, increasing recognizability and facilitating imagination. In a way, this “vague and unspecific” nature and lack of fixation in comics helps bridge the gap between books and film [179].

Some efforts have tried to harness comics for visualization. Jin and Szekely [174, 180] proposed a visual query environment that uses a comic-strip metaphor for querying and presenting temporal patterns. However, their system solely uses comics strip for layout, and does not leverage the full potential of comics as a visual communication medium.

The most recent and most relevant work in this vein is Graph Comics [151], which are data-driven comics used for telling stories about dynamic networks. Simultaneously invented as our work in this dissertation, Graph Comics draw on the same comic medium principles as Data Comics, but are restricted to node-link diagrams. Furthermore, whereas the Graph Comics work merely proposes the idea and explores its utility, it does not provide any authoring support for creating them. Thus the Graph Comics effort is complementary to the more general Data Comics approach proposed here.

There are other relevant works including applications and surveys. Kim et al. [181] proposed the DataToon for interactive data comics and dynamic networks. Moore et at. [182] introduced the comic-strip narratives in time geography. Bryan et al. [183] proposed an approach for interactive annotation to narrative visualization with comic-strip style snapshots. Wang et al. [184] compared the effectiveness and engagement of data comics and infographics. Wang et al. [185] also explored how to teach data visualization and data-driven storytelling with data comics. Bach et al. [186] explored the design patterns of data comics for data-driven storytelling.

2.7 Animated Graphics as a Storytelling Medium

Video-based storytelling is making significant inroads on the internet as a whole. Never mind streaming services such as Netflix, Hulu, and Amazon Prime, which according to the network service company Sandvine accounts for up to 70% of peak internet traffic [187] (December 2015), the new generation of so-called "YouTubers", or YouTube celebrities, are challenging the boundaries of the medium through novel formats such as "vlogging" (short for video blogging), "reaction videos" (recording of a person reacting to an event), or "unboxings" (video of someone unpacking a new product, commonly a high-tech one). One format in particular is fascinating: so-called "Let's Play" video capture the gameplay as well as live audio and video of a person playing a computer or console game. Sometimes likened to watching a friend play a game while sitting on a couch in their home, Let's Play videos are characterized by focus on the often humorous, irreverent, and sometimes profanity-filled commentary that the recorded person provides. Popularize by Swedish YouTuber (and now general celebrity) Felix Kjellberg, better known by his YouTube alias PewDiePie and holding the distinction of having more than 42 million follower and more than 10 billion views, this phenomenon has since given rise to live-streaming Let's Play videos, such as through the online streaming service

Twitch.tv.[1] Combined with their focus on broadcasting eSports, where professional gamers play electronic games for prize money and salaries, Twitch has quickly risen to be the fourth largest source of internet traffic in the United States [188].

Amini et al. [153, 154] recently identified data videos as motion graphics that combine both sound and visuals to tell a data story. Pointing to prominent examples from the New York Times and the Guardian, their work is descriptive and formative in nature, engaging professional storytellers to use visuals to craft their narratives. While obviously deeply influential to our work, their treatment focuses on the careful and deliberate production of painstakingly designed data videos through ideation, sketching, storyboarding, capturing, and editing; live streaming is not covered or even considered, and the software, resources, and skills needed of their approach is substantial. Furthermore, unlike our DataTV platform, their work provides no technical interventions to support data video authoring.

[1] `http://www.twitch.tv/`

Chapter 3: Taxonomy of Media for Data-Driven Storytelling

We present a new taxonomy focused on the media for data-driven storytelling with the purpose of opening the field to a wider set of future possibilities. Our work started with collecting a significant corpus of evidence of data-driven storytelling using novel and diverse media, from the spoken word to interpretative dance and choreography. We then use these wealth of data to derive a taxonomy and classify all of these examples into a coherent framework. Finally, by generalizing across storytelling practice for different media, we derive design guidelines for data-driven storytelling and discuss how future narrative techniques about data may look.

3.1 Method

To collect examples with enough quality and quantity, we start from data features and storytelling features, by collecting examples with proper data components and clear storytline. The sources of examples are online articles, blogs, visualization tools, video websites, books, research papers, and software packages. In order to generate a robust taxonomy, we chose the examples from a wide range of genres including infographics, documentaries/data videos, data comics, visual analytics/visualization tools, virtual reality visualization tools, augmented reality visualization tools, Computer game tools, and etc. Figure. 3.1 shows the distribution of all the sources of examples. We started from about one hundred and thirty examples. Ninety-one are left after removing examples without clear storyline, explicit data for a story, or clear group audiences.

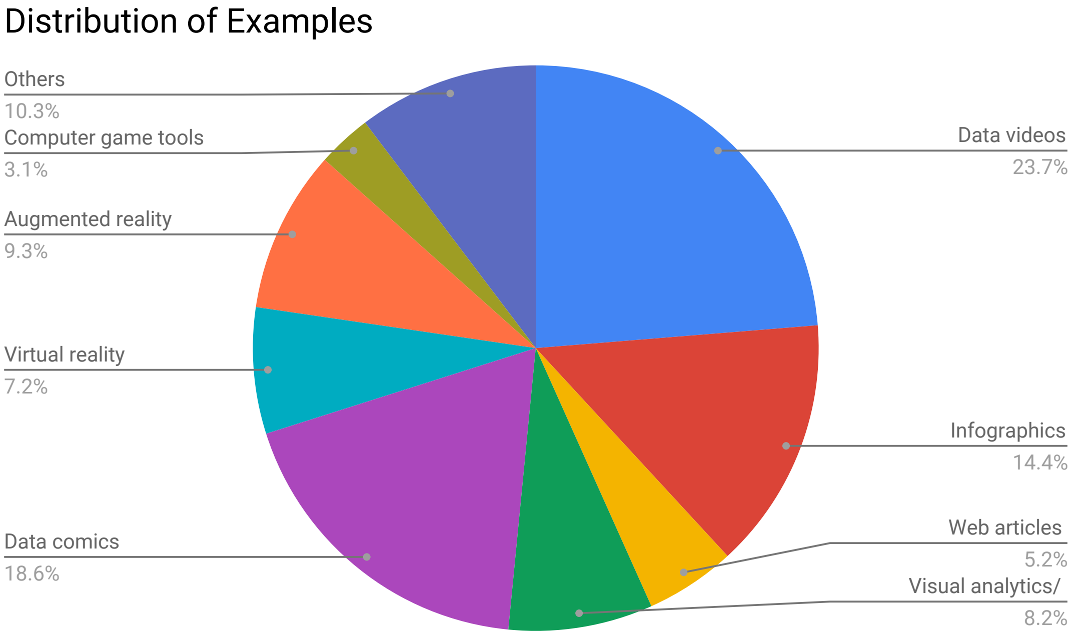


Figure 3.1: The distribution of all the sources of examples.

3.2 Evidence

To illustrate the prevalence and variety of different data-driven narratives in the world, we here enumerate and discuss a set of representative and innovative such examples. We only list three representative ones; the rest are explained in Appendix A with classification in Table 3.1. The purpose is to provide a basis for a taxonomy that can be used to classify the storytelling media used for the data-driven narrative. For each example, we use an informal classification scheme to describe the media in more detail. This scheme will then feed into our taxonomy in the following section.

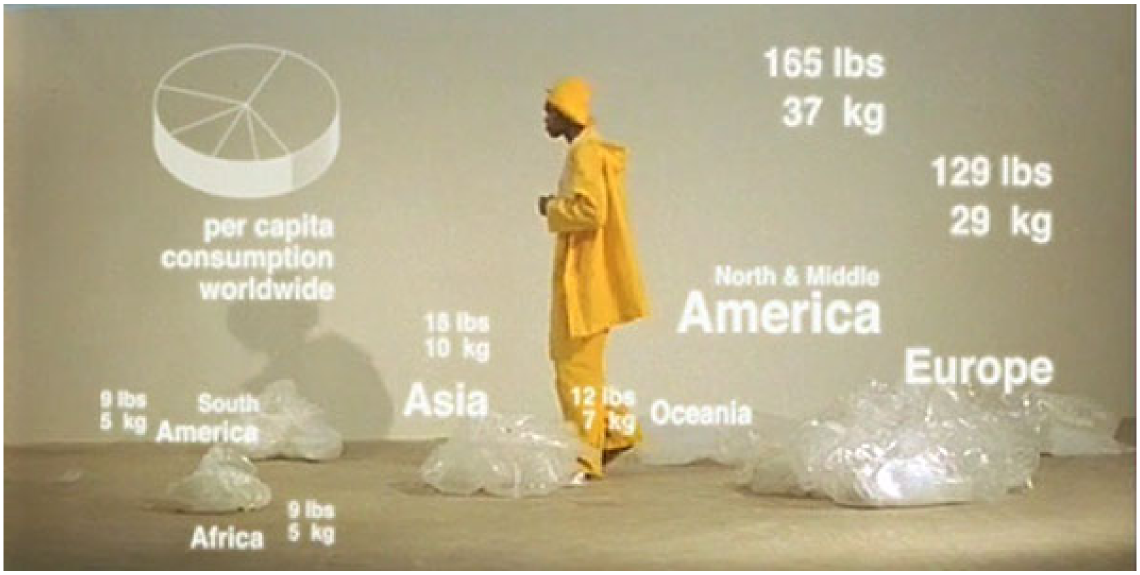


Figure 3.2: Image from the video *A Day in the Life of the (Polluted) Ocean*.

3.2.1 Exhibit 1: Storytelling in Movies and Documentaries

The movie ”A Day in the Life of the (Polluted) Ocean” [189] talks about the pollution emission of the main character (the person in yellow) for one day. The storyline moves with him as a regular person getting up in the morning and proceeds with his normal day. The idea is to show how much pollution a normal person can produce with simple activities.

Figure 3.2 depicts one scene that shows that the amount of plastic waste produced for every capita in each continent. In this scene, the underlying data is the weight of plastic waste, and the story is that there is too much plastic waste produced for generating a certain amount of value. There is also a simple chart on the left upper corner illustrating the portion of plastic waste per capita for each continent. This scene has both simple numbers and data visualizations to facilitate the storytelling.

Informal classification: As a movie that is *shared online*, the audience is potentially *many people* on the internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* The *visual components* used include full-motion video, animated graphics, text, and non-interactive static data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The typical way to view a video is in *sequence.* However, it is also *replicable*, as it is stored and can be played back at any time.

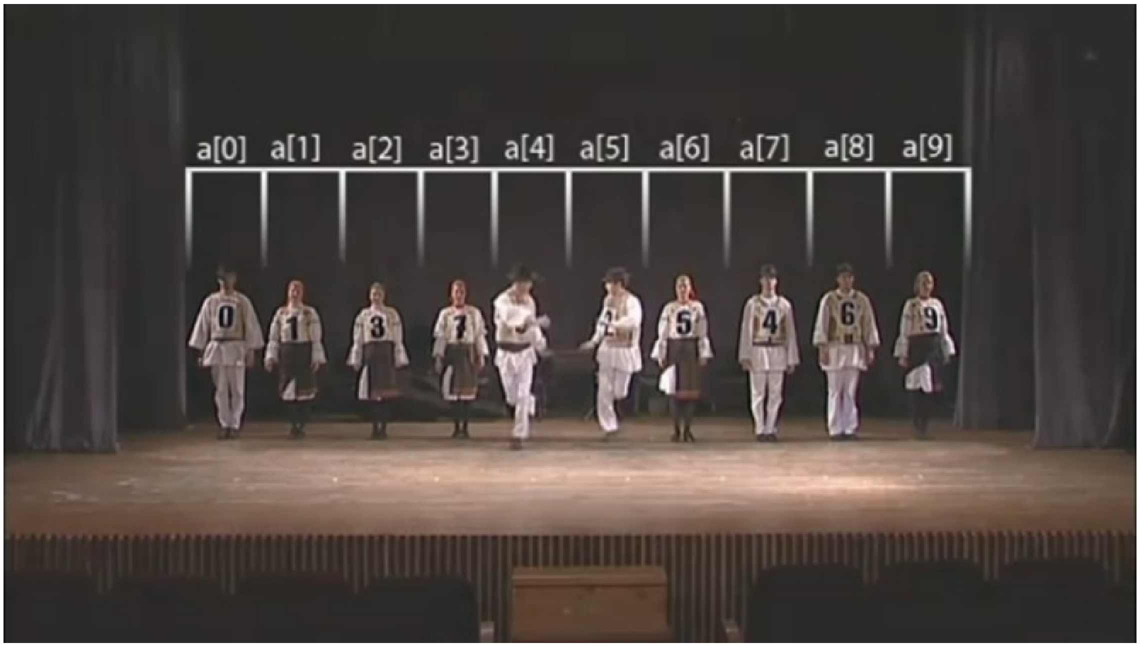


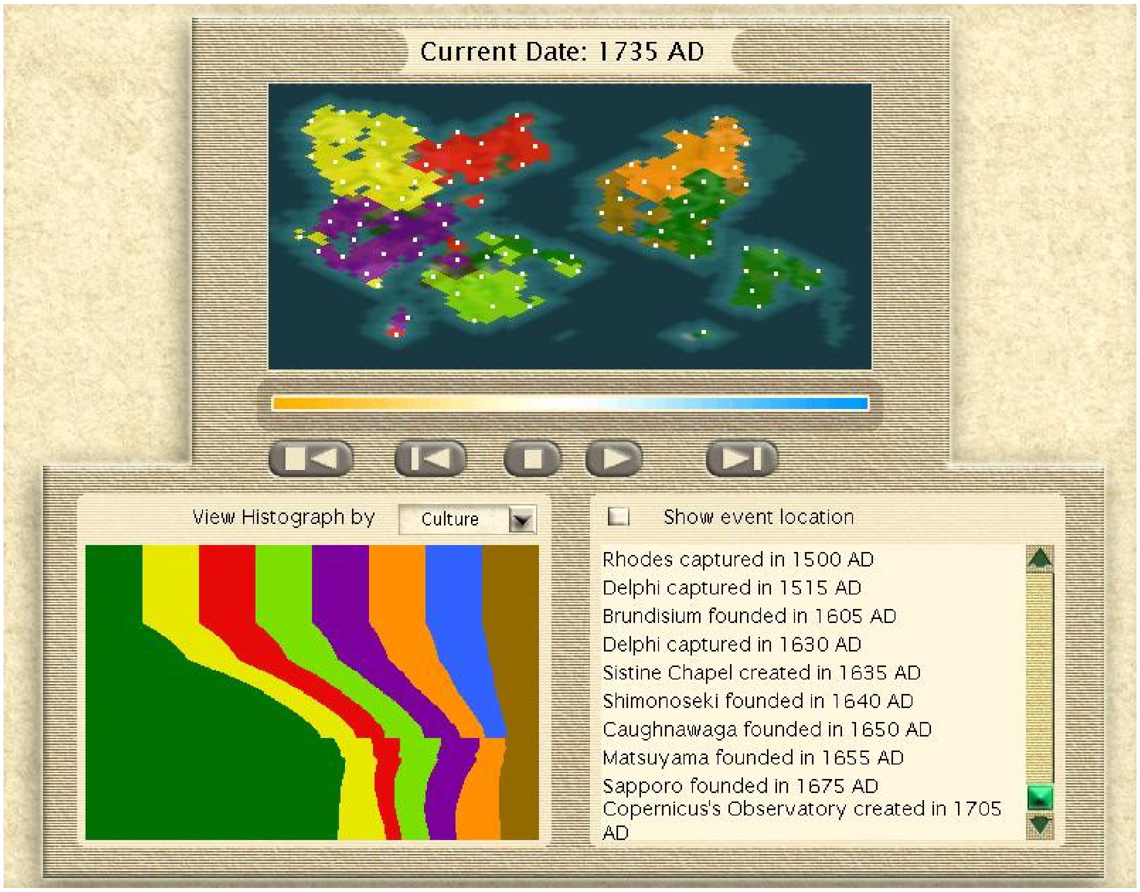
Figure 3.3: Physical dance demonstration of a bubble sort algorithm.

3.2.2 Exhibit 2: Dance

Figure 3.3 shows an image from a performance where dancers demonstrate how the bubble sort algorithm works [190]. This algorithm pushes the largest element to the right and forms a ordered sequence of numbers. For clarity, each dance is labeled in the this picture. In reality, the dancers wear uniforms with numbers on them, where each dancer stands for a different element in the array. The storyline is the movement of all the array elements. The data is the ordered sequence.

A similar data-driven dance performance was created by the NSF-funded Dance.Draw project [191], where the movements of dancers in a physical space was conveyed using visual representations. This mechanism could also be used as a vehicle for conveying data-driven stories.

Informal classification: Dance performances typically take place in an auditorium or studio, which supports a *large audience*. Disregarding video recordings of the performance (which would be another form of media), this does require the audience to be physically *co-located* with the performance, and to consume it in *synchronously*, in real-time. The sorting process demonstrated by sorting is view in *sequence*. This means that the performance is not stored; it is *ephemeral*. The components of the performance are human bodies in motion over the duration of the performance, and can also include text, sound, light, and visuals (typically projected in the workspace).



Rhodes captured in 1500 AD
Delphi captured in 1515 AD
Brundisium founded in 1605 AD
Delphi captured in 1630 AD
Sistine Chapel created in 1635 AD
Shimonoseki founded in 1640 AD
Caughnawaga founded in 1650 AD
Matsuyama founded in 1655 AD
Sapporo founded in 1675 AD
Copernicus's Observatory created in 1705 AD

Figure 3.4: Data-driven story of a session in the PC game *Civilization 3*.

3.2.3 Exhibit 3: Data-Driven Storytelling in Video Games

While Segel and Heer [2] explicitly note that they chose not to include video games in their survey, games have long been instances where visualization is often integrated. As it turns out, they have also been used for data-driven storytelling. Figure 3.4 depicts an image from the replay session that is shown at the end of a completed game session, i.e., after one of the players (human or computer) has achieved one of the victory conditions: conquering all opponents, winning the space race, taking over most of the land, or scoring a diplomatic or cultural victory. The replay shows a history of how each civilization was founded, expanded across the world, and was eventually defeated. While the interaction is limited, playback controls allows the user to go back and forward in the history. Similar session playback functionality—often called *theater mode*—can be found in *Halo 3* and *Call of Duty: Black Ops.*

Informal classification: The audience for most theater modes is *individual* players who wants to study their own and other players' performance. However, many theater modes typically also allow the player to cut and paste clips together, eventually producing a resulting *video to share* with others. The resulting video will have the same features as a data video (see above). A playback session typically uses a map view, so the bandwidth requirement is lower than full-motion video, thus reducing the cognitive load. One feature of most theater modes is that they make it easy to navigate in 2D or 3D in the scene, thus changing the viewpoint. The viewing sequence of a playback system is in *sequence*. While the action itself cannot be changed (since it represents events that already happened), this interaction is powerful in that it can, for example, allow a player to put themselves in the shoes of another player to see what an encounter looked like from their viewpoint.

3.3 Taxonomy Dimensions

We propose to identify, study, and classify *media* that have traditionally been used for storytelling. The purpose of this activity is to expand existing genres of narrative visualization to encompass the entire scope of storytelling in society. This would generate a wealth of new research ideas for how to best use such media for data-driven storytelling. We found the following dimensions useful in classifying such media, based on our survey of the existing evidence of a wide variety of media for data-driven storytelling:

- **Audience Cardinality (A):** Who is the intended recipient for the story?
- **Space and Time (S/T):** What is the temporal and spatial delivery mechanism for the story?
- **Visual Components (VC):** What are the visual and sound building blocks employed? Are the components interactive?
- **Data Components (DC):** How is the data conveyed to the viewer?
- **Media Viewing Sequence (SQ)** How is the media viewed by the viewer?
- **Storage and Persistence (SP):** How is the media preserved. Is it temporary or replicable?

The methods we used to derive these dimensions included reviewing a sample of about one hundred and thirty examples of data-driven storytelling, and then narrowing our selection down a set of 91 representative examples. We looked at the fundamentals of the process regarding creating data stories, broadcasting them, and interacting with them. From the dimensions we selected, we are able to determine a profile of a data-driven storytelling media type, so that the media types are descriptive enough in our taxonomy. For example, the dimension Data Bandwidth is highly correlated with data components. But this dimension is very important for storyteller when choosing a media type to convey large amount of data in real time. These examples have been classified into the taxonomy and can be viewed in Table 3.1.

3.3.1 Audience Cardinality (A)

In our notion of the audience of the data-driven story, we also include the storytellers: whether it is one or several people who are creating or viewing the narrative, respectively. A group performance, such as the bubble sort dance in the example above, would be an example of many storytellers (the dancers) conveying data to many recipients (the audience in the dance studio). When choosing the Audience Cardinality, the storytellers can adjust the way of the story is conveyed. For example, by changing the way computer game is recorded, we have changed the media type from local video to online broadcasting, and thus we have changed the Audience Cardinality from one-to-one to one-to-many.

Audience has one of the below values:

- *One-to-one* (1:1): one storyteller and one recipient, such as in a private conversation. The traditional storytelling process from one storyteller to an audience is a typical one-to-one process.

- *One-to-many* (1:N): one storyteller and many recipients, such as a speaker giving a talk to a group. Most media with broadcast ability is one-to-many, since one storyteller often creates the data story and communicates it to more than one receipt. For example, the creating and broadcasting process of infographics, data comics, and data videos.

- *Many-to-many* (N:N): many storytellers and many recipients, such as a dance troupe giving a performance to an audience. There are media types that allow multiple storytellers to broadcast at the same time. For example, in the advanced version of Data TV, the system allows more than one user to tell data stories at the same time.

Although the values of this dimension seem straightforward, there are situations from which choosing values can be hard. For example, during the live streaming with data stories, the video is broadcast to multiple viewers. This seems typical one-to-many cardinality. However, for extreme cases, such that there is only one recipient, the broadcast of the media becomes one-to-one. In general, we only cover the basic cases here.

3.3.2 Space and Time (S/T)

We borrow the notion of space and time from CSCW [192], where the space-time matrix has long been used to characterize forms of groupware based on the spatial and temporal relations of the human users. Some media types naturally require that the storytellers and recipients are located in the same place, such as dancers and their audiences. Other media types, such as infographics, movies, and comics are easily recorded and broadcasted. The storytellers and the recipients can stay in different places and time periods. It is a useful property because both space and time have a significant impact on the delivery and storage mechanism for the data-driven story.

Space and time has two values, one for each dimension:

- *Space*: relative physical locations of storyteller and recipient.
  - *Co-located* (coloc): the storyteller and the recipient are in the same physical space. Traditionally, the storyteller is in the same location with the audience when communicating a story, such as giving a speech, showing a dancing performance, or giving a presentation in person.
  - *Distributed* (distr): the storyteller and the recipient are **not** in the same physical space. Some examples of this space attribute are media types using modern technology such as live video streaming, phone meeting, and other collaborative work tools.
- *Time*: temporal locations of storyteller and recipient.
  - *Synchronous* (sync): the storyteller is delivering the story to the recipient in real time. This time attribute applies when it is required that the storytellers and recipients share the same temporal locations. Such instances can be included for media types that have a broadcasting ability and have no method of saving a record, such as live streaming and dance performances.
  - *Asynchronous* (async): the storyteller is delivering the story in a form that will be consumed by the recipient at a later time. Examples of this include media types that have default saving capabilities, such as

infographics, in which the recipient can consume the story anytime after the story has been delivered.

The dimension values *Co-located* and *Distributed* are very important in terms of determine the media type to use when there are limited resources. For example, if the storyteller and audience are located separately, the storyteller has to choose a media type that is able to convey the story in a *Distributed* way.

This dimension can be heavily affected by the habits and preferences of the storytellers and recipients. For example, dancing performance are often recorded by video. As a result, the habitual way to tell this story is by distributing in both time and space, allowing recipients to view it at a different location in the future.

3.3.3 Visual Components (VC)

The nature of the media captures the composition of the media being used for data-driven storytelling. Since we are typically talking about composite media types, the Media Components is a set variable that can include one or several of the below:

- *Audio* (aud): audio, such as speech recordings, ambient noise, or sampled sound effects.
- *Photographs* (pho): pixmap images.
- *Live video* (vid): animated pixmap images with basic control.
- *Static graphics* (gra): non-dynamic vector graphics.

- *Animated graphics* (ani): dynamic vector graphics with control.

- *Text* (txt): textual representations.

We covered the media components that used for data-driven storytelling, and did not include other components that do not contribute as much. This list of media components may not be complete, as we only covered the major media components shown by the examples.

3.3.4 Data Components (DC)

The core purpose of a data-driven storytelling artifact is to convey data from the storyteller to the viewer. The form that this takes is captured in the Data Components (DC) dimension. For this dimension, we enumerate the most common components used by the examples. When designing the Data Components for data-driven storytelling media, storytellers are choosing what is the way that data will be presented. Our enumeration of the components is sufficient to cover most of the common data expressions. For example, we know that table is used for structured data, map is used for geological information, and time visualization is used for continuous time series. It is a set variable that will take or several of the below values:

- *Table* (tab)

- *Map* (map)

- *Statistical graphics* (stat)

- *Discrete event visualization* (event)

- *Continuous time visualization* (time)

- *Graph visualization* (graph)

- *Text visualization* (txtvis)

- *Formula* (fmla)

This dimension covers the major artifact that carries data from the examples. The list of values can also suffer from incompleteness caused the limited scope of the examples.

3.3.5 Media Viewing Sequence (SQ)

From the nature of the media types, there are viewing sequences better accepted than others. For example, a comic strip can be viewed in sequence or with branches in most of the cases. A video is mostly viewed in sequence. Infographics can be viewed in parallel, which makes them capable of conveying information faster in certain situations, and thus can be more difficult to understand [1].

- *Sequence* (seq): viewed linearly, the stroyline is in sequence.

- *Branch* (brch): viewed in sequence, but with branches.

- *Parallel* (prll): be able to viewed in parallel.

3.3.6 Storage and Persistence (S/P)

While related to the delivery mechanism of the story artifact, the storage and persistence aspects of the artifact governs how and when it can be consumed. It takes one of the following values:

- *Ephemeral* (eph): generated once and not recorded, not easily replicated. Some media types have stories received by the audience immediately after the stories are generated and told, such as gestures, speeches and facial expressions. These types of storytelling process relies on other tools to record them and are hard to replicate. These media types can be recorded by other media types, but we only consider the media types by themselves.
- *Replicable* (rep): generated once, but can be replicated by the storyteller on demand. Some media types are easily replicated and stored, such as video, audio and infographics.

3.4 Examples Using the Taxonomy (Applications)

With our taxonomy, we explored a few applications with new designs. By changing the values of certain dimensions, we were able to test or propose new media for data-driven storytelling.

Media Evidence	A	S/T	VC	DC	SQ	SP
Video game replay	1:1	coloc/sync	vid	map | events	seq	rep
Life of the Ocean [189]	1:N	distr/async	vid	tab | stat | txtvis	seq	rep
A Beautiful Planet [193]	1:N	distr/async	vid	tab | stat | txtvis	seq	rep
News Irma Hurricane [5]	1:N	distr/async	vid	map | graph	seq	rep
Height Gene [6]	1:N	distr/async	vid	tab | stat | txtvis	seq	rep
NFL Players Data [22]	1:N	distr/async	gra | txt	tab | stat | txtvis | event | time	seq | brch	rep
Graph Comics [151]	1:N	distr/async	gra | txt	graph	seq | brch	rep
The Two Americas [194]	1:N	distr/async	gra | txt	tab | stat | txtvis | event	seq | brch	rep
Strikeouts on the Rise [38]	1:N	distr/async	gra | txt	tab | time | event	prll	rep
Missing Black Men [39]	1:N	distr/async	gra | txt	tab | time | event	prll	rep
VisJocky for Stocks [40]	1:N	distr/async	gra | txt	time | event	seq	rep
ChartAccent for Countries [41]	1:N	distr/async	gra | txt	time | event	seq	rep
SketchStory [171]	1:N	distr/async	gra | txt	tab | time | event | graph	seq	rep
Skecholution [42]	1:N	distr/async	gra	event | graph	seq	rep
New Orleans Housing [46]	1:N	distr/async	gra | txt	event | graph	prll	rep
Top writers for best sellers [47]	1:N	distr/async	gra | txt	tab | event	prll	rep
Public Library Report [48]	1:N	distr/async	gra | txt	tab | stat | txtvis | event	prll	rep
HALO Reach [65]	N:N	distr/async	gra	map	seq	rep
Call of Duty: Black Ops [66]	1:1	coloc/async	gra	event	seq	rep
StarCraft II [67]	1:1	coloc/async	gra	event	seq	rep
Twittersheep [68]	1:N	distr/async	txt	txtvis	seq	eph
Twitter Interactive GOT [69]	1:N	distr/async	gra | txt	graph | txtvis	prll	eph
Interactive: Tweets Spread [70]	1:N	distr/async	ani | txt	map | stat | graph | time	seq	rep
Dance of Sorting [190]	N:N	coloc/sync	music | txt	txtvis	seq	eph
Microsoft PowerPoint	1:N	coloc/sync	aud | gra | ani | txt	tab | stat | event | time | txtvis	seq	rep
IBM Many Eyes [143]	1:N	distr/async	gra | txt	tab | stat | event | time | txtvis	seq	rep
Storyfy [195]	1:N	distr/async	gra | ani | txt	tab | stat | event | time | txtvis	seq	rep

Table 3.1: Classification of our representative sample of different media used for data-driven storytelling–one.

Media Evidence	A	S/T	VC	DC	SQ	SP
Height Gene [6]	1:N	distr/async	vid	tab | stat |time | txtvis	seq	rep
Ancient Greece [7]	1:N	distr/async	vid	tab | map |stat |graph |event	seq	rep
The history of Asia [8]	1:N	distr/async	vid	map | graph	seq	rep
U.S. Wealth Inequality [9]	1:N	distr/async	vid	stat | map | graph |txtvis	seq	rep
Saving Poor Children [20]	1:N	distr/async	vid	stat	seq	rep
Carcaptor Sakura [44]	1:N	distr/async	gra | txt	tab | stat | graph	seq | brch	eph
Movie Explained Animated [45]	1:N	distr/async	vid	txtvis | event	seq	rep
Comic style Dashboard [23]	1:N	distr/async	gra | txt	tab | stat | graph	seq | brch	rep
Marvel&DC Comic [26]	1:N	distr/async	gra | txt	tab | stat	seq | brch	rep
Body Cartoon Comic [27]	1:N	distr/async	gra | txt	stat | time	seq | brch	eph
Spiderman Suit [28]	1:N	distr/async	gra | txt	txtvis	seq | brch	rep
Phone Comic [29]	1:N	distr/async	gra | txt	stat | time	seq | brch	rep
Linear Regression [30]	1:N	distr/async	gra | txt	tab | graph	seq | brch	rep
Curve Fitting [36]	1:N	distr/async	gra | txt	tab | graph	seq | brch	rep
Seashell Comic [37]	1:N	distr/async	gra | txt	fmla	seq | brch	rep
Infographic Comic [24]	1:N	distr/async	gra | txt	event | stat	seq | brch	rep
DataSketches [43]	1:N	distr/async	gra | txt	event | graph	seq | brch	eph
Shadow of Foreclosure [56]	1:N	distr/async	gra | txt	map | event	prll	rep
US Poverty [49]	1:N	distr/async	gra | txt	tab | stat | map	prll	rep
Democrats and Republicans [57]	1:N	distr/async	gra | txt	tab | stat | txtvis	prll	rep
UK and US Firearms [58]	1:N	distr/async	gra | txt	stat | time	prll	rep
Correspondent Dinner [59]	1:N	distr/async	gra | txt	tab | event | txtvis	prll	rep
Day vs. Night: NYC [60]	1:N	distr/async	gra | txt	map | event	prll	rep
Who owns everything [61]	1:N	distr/async	gra | txt	graph | txtvis | event	prll	rep
Big Welsh Coast Walk [62]	1:N	distr/async	gra | txt	tab | stat | txtvis	prll	rep
Hangry USA [63]	1:N	distr/async	gra | txt	tab | stat | map	prll	rep
NYC Celebrity Map [64]	1:N	distr/async	gra | txt	map	prll	rep
VR Data Analysis [86]	1:N	distr/async	txt | ani | gra	tab| map | stat | event | time | txtvis	seq	rep
VR Baseball training [80]	1:N	distr/async	gra | txt | ani	tab | stat	seq	rep
Adobe VR Data Vis [79]	1:N	distr/async	gra | ani | txt	tab | stat | graph | time	seq	rep

Table 3.2: Classification of our representative sample of different media used for data-driven storytelling–two.

Media Evidence	A	S/T	VC	DC	SQ	SP
Gene Pool Decline [12]	1:N	distr/async	vid	stat | graph	seq	rep
The Joy of Stats [10]	1:N	distr/async	vid	stat	seq	rep
End Poverty [13]	1:N	distr/async	vid	tab | stat	seq	rep
China Geo Problem [14]	1:N	distr/async	vid	map	seq	rep
Data Trans Biz [19]	1:N	distr/async	vid	tab | stat | event | time	seq	rep
Bigdata Revolution [16]	1:N	distr/async	vid	tab | stat | map | time	seq	rep
PhD Comic [21]	1:N	distr/async	gra | txt	stat | txtvis | time	seq | brch	rep
Vocation Comic [31]	1:N	distr/async	gra | txt	time	seq | brch	rep
Desk Entropy Comic [32]	1:N	distr/async	gra | txt	time	seq | brch	rep
PhD Grooming Comic [33]	1:N	distr/async	gra | txt	time	seq | brch	rep
Procrastination Comic [34]	1:N	distr/async	gra | txt	time	seq | brch	rep
Procrastinator Mind [18]	1:N	distr/async	vid	tab | stat | event	seq	rep
Imaginary Numbers [15]	1:N	distr/async	vid	stat | txtvis	seq	rep
Religions and babies [11]	1:N	distr/async	vid	tab | stat | map	seq	rep
Population truth [17]	1:N	distr/async	vid	stat | txtvis	seq	rep
Household Pollution [53]	1:N	distr/async	gra | txt	tab | txtvis	prll	rep
Air Pollution [54]	1:N	distr/async	gra | txt	stat | txtvis	prll	rep
Yemen War [51]	1:N	distr/async	gra | txt	map	prll	rep
London March [50]	1:N	distr/async	gra | txt	stat|txtvis	prll	rep
Space Industry [52]	1:N	distr/async	gra | txt	stat|txtvis	prll	rep
NYC Restaurant Vis [25]	1:N	distr/async	gra | txt	tab | graph | stat | txtvis	prll	rep
VR Geo Map [85]	1:N	distr/async	gra | txt | tab | ani	map | stat | txtvis	seq	rep
VR BioInfo Vis [84]	1:N	distr/async	gra | txt | ani	tab | stat |graph | txtvis	seq	rep
VR Immersive [83]	1:N	distr/async	gra | txt | ani	stat | tab	seq	rep
VR Lens [82]	1:N	distr/async	gra | txt | ani	stat | tab | map	seq	rep
VR Big Data [81]	1:N	distr/async	gra | txt | ani	stat | tab | map	seq	rep
AR Flight [74]	1:N	distr/async	gra | ani | txt	map | stat | graph | time	seq	eph
AR Street View [75]	1:N	distr/async	gra | ani | txt	map | stat | event	seq	rep
AR Pipeline [76]	1:N	coloc/sync	gra | ani | txt	stat | graph | event	seq	eph
AR BioChemical [78]	1:N	distr/async	gra | ani | txt	graph	seq	rep
AR Infrastructure [76]	1:N	coloc/sync	gra | ani | txt	graph | time	seq	eph
Uber Mobile Vis [71]	1:N	coloc/sync	gra | ani | txt	tab | stat | event | txtvis	seq	eph
AR Data Vis Design [72]	1:N	distr/async	gra | txt	tab | stat | event | time | txtvis	seq	rep
AR 3D Design [73]	1:N	distr/async	gra | ani | txt	tab | stat | graph	seq	rep

Table 3.3: Classification of our representative sample of different media used for data-driven storytelling–three.

3.4.1 Data Comics

In an effort to show how sequential art—also known as comics—can be used as a novel method for storytelling, we presented this example with a technique of creating sequential art for data-driven storytelling. We developed the Data Comics system with the design principle from our taxonomy—combining storytelling with the comic media [152].

The data comics technique, specifically introduced in appendix [152], allows the users to build narratives using comic layouts of panels containing both snapshots and live visualizations. Comic features are seamlessly emerged into the comic layout panel. The storytelling process leverages the continuous frames as the organization of storyline and the comic features, such as speech bubbles, comic figures, and directional arrows, to bring the data-based storytelling into another dimension.

The above figure is an example of the data comic frame produced with our design philosophy by the data comic system. The design of the data-driven storytelling in data comic system is not much different from traditional oral way, users are still dominant in the system as the system is not capable of generate the storyline by itself. More specifically, users start by collecting the data-visualization by cropping a subgraph and decide the size and location of cropped part. Such cropping finish allows the user to choose the level focus and decontextualization, which encourages the users to summarize the most interested part of the whole visualization. Then they are free to map the data visualization with frames of any size and sequence.

3.4.2 Data TV

DataTV is a prototype system for authoring live-streaming data videos using single, integrated interface. The DataTV prototype is based on OBS Studio[1], a popular Open Source project for game streaming that supports multiple different operating systems. The prototype supports three separate modes for (1) production, (2) recording, and (3) editing, in a highly streamlined and optimized workflow that allows a single content creator to control the entire process, even during live streaming. The tool incorporates multimedia sources such as live webcams, live audio recordings, web browsers, image viewers, and full-motion video. In particular, it supports live recording of any selected window presumably containing an interactive visualization, such as a web browser or dedicated application window.

Furthermore, the tool incorporates advanced video functionality, such as chroma keying (making parts of a stream transparent, such as for blue or green screens), picture-in-picture, hand-drawn annotations (for highlighting important parts of a stream), viewport control (zooming and panning), and advanced source composition operations (transitions, stretching, and fitting). To validate the DataTV platform, we engaged three visualization tools which are the results of published work in the information visualization community and created data videos analyzing a certain topic with them. Further more, we asked two visualization experts to create a data video less than one minute long with DataTV on a random topic to test if our system is adaptive on a wide range of topics.

[1] Open Broadcaster Software: http://obsproject.com/

Our DataTV interface supplies as much space as needed for the web-based visualization to support its interactive features (Figure 3.7).

3.5 Implication for the Taxonomy

We have successfully labeled about 91 existing examples with our taxonomy shown as Table 3.1, 3.2, 3.3, and Figure 3.5. The examples cover movies, documentaries, web articles with data visualization, infographics, comics, social media, visualization tools, games, dance, and sketching tools, which are a large part of the major categories for storytelling. This comprehensive classification provides good evidence that our taxonomy is sufficient and complete.

From Figure 3.5 we can see that there are thick paths that indicate there are more examples categorized with the values in these dimensions. The parts with light paths or without paths are the **gaps** between examples. To **fill the gaps** of the taxonomy and the examples, we can find examples based on light path or the dimension values that have not been explored. For example, to create media type data comics from infographics, we can follow the design space and dimensions of infographics, then change the viewing sequence from parallel to sequence or branch. Using partitioning and sequencing, we can reorganize the content of infographics and add static comic graph to create data comics. The detailed process and evaluation is mentioned in Chapter 5 later.

From collecting the examples, we have found media types primarily in the form of data videos (23%), infographics (14%), web articles (5%), visual analytics/visualization tools (8%), data comics (18%), virtual reality visualization tools (7%), augmented reality visualization tools (9%), and computer game tools (3%). We also observe that the values of some dimensions are more diverse than others. For example, the dimension Data Component has the most diversity, meanwhile some of examples share part of the values for this dimension.

Our taxonomy is an extension that builds on the foundation that Segel and Heer [2] laid in 2010. While this foundation has proven instrumental in the guiding the development of data-driven storytelling, their model is limited in scope and conflates the delivery mechanism with the media used for the message. We believe our taxonomy provides a more comprehensive view of media for data-driven storytelling while still building on their foundational work. Using our taxonomy, designers will be able to widen the horizon of data-driven storytelling. By providing a taxonomy with detailed dimensions, we explored the possible values for each dimension. By expanding the list of dimensions and dimensions values, we can also keep tracking of the emerging media types. In particular, the terms and dimensions in our taxonomy provides a standardized vocabulary to use when discussing data-driven storytelling. This enables researchers and practitioners alike to classify their own work so that existing and new media can be systematically organized with a common ground.

However, the true value of a taxonomy such as ours is in generating new ideas by identifying gaps in the literature. By grouping and labeling the dimensions of existing media, our taxonomy can help researchers identify new areas to explore in the future. For example, this new design space can be generated by exploring previously untested combinations of dimensions. While we have done so in the previous section, specifically in terms of the Data Comics and Data TV applications, the space is wide open for even more radical ideas. For example, consider employing data-driven storytelling in e-learning, social media, or even video games. What about interpretative dance, theater, improv, music, or even song for data-driven storytelling? The possibilities are endless.

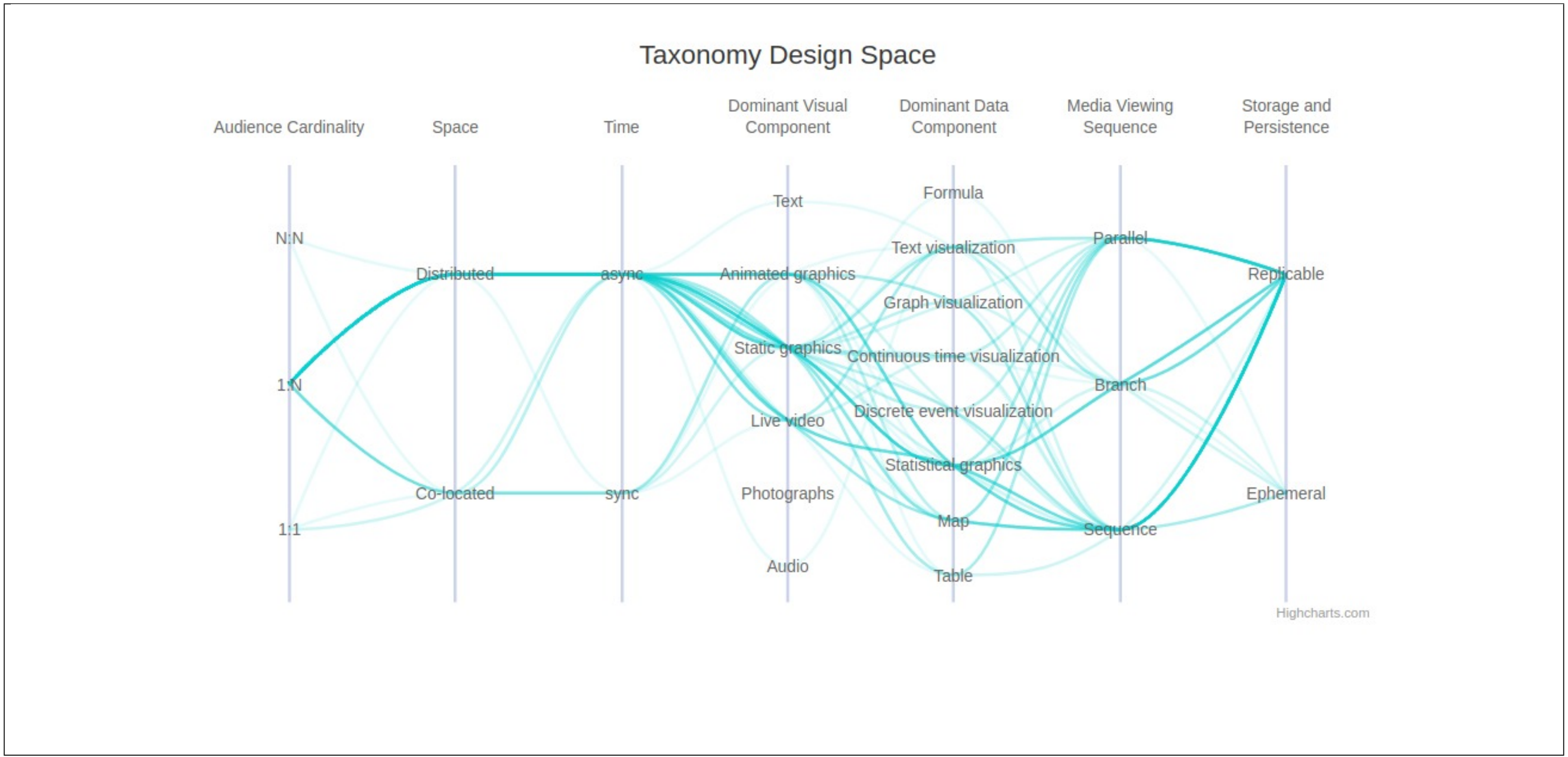


Figure 3.5: The figure shows the parallel coordinate graph of the design space of the examples categorized by the taxonomy

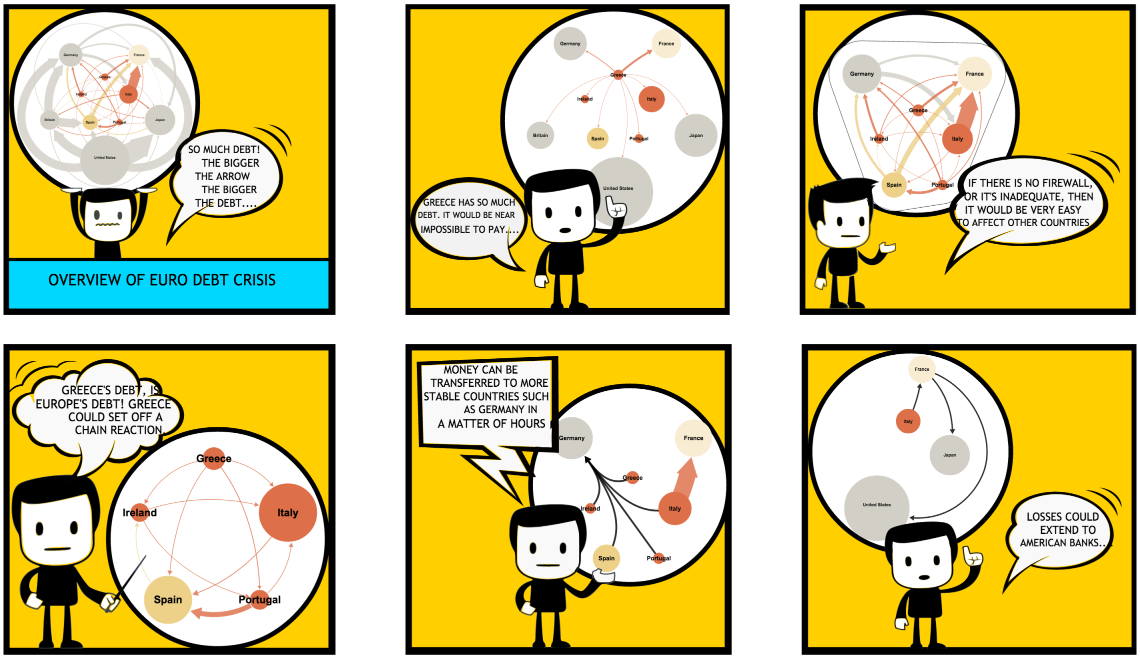
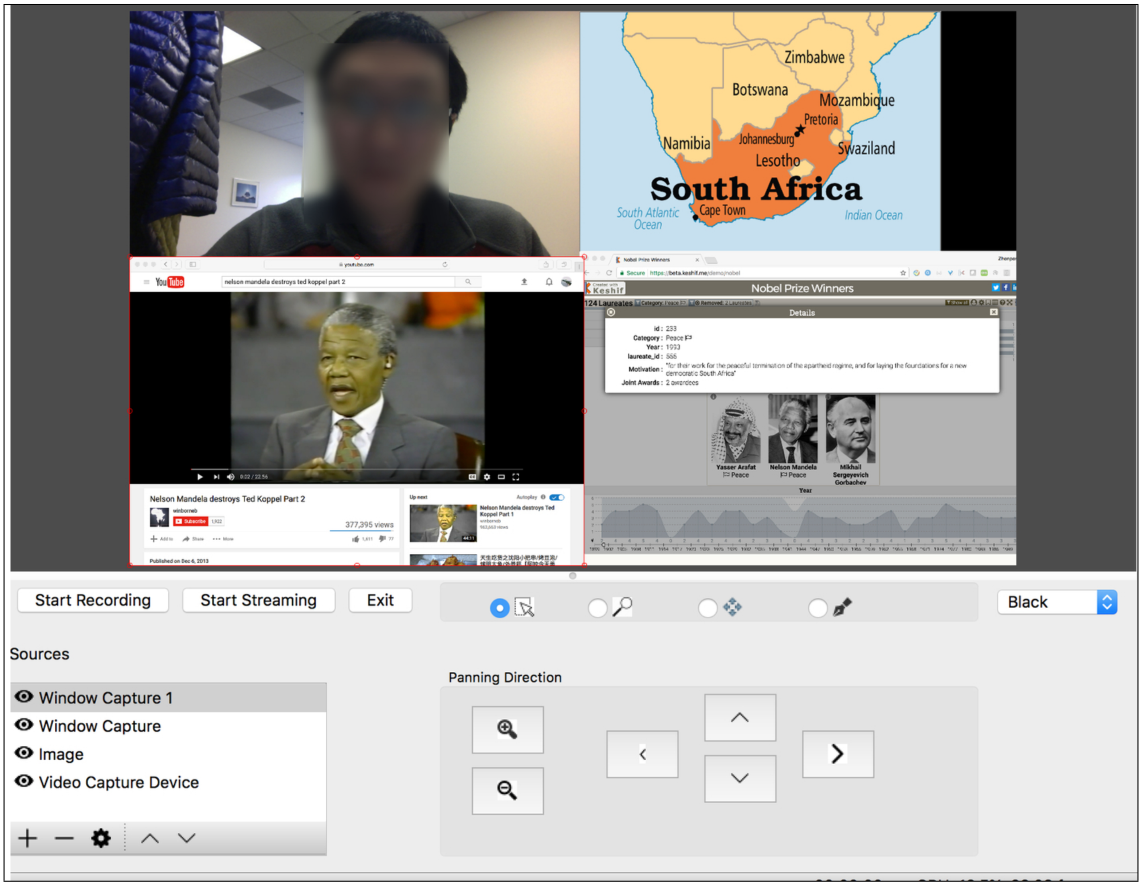

Germany
France
Greece
Ireland
Italy
Britain
Spain
Portugal
Japan
United States
SO MUCH DEBT! THE BIGGER THE ARROW THE BIGGER THE DEBT....
OVERVIEW OF EURO DEBT CRISIS
Germany
France
Greece
Ireland
Italy
Britain
Spain
Portugal
Japan
United States
GREECE HAS SO MUCH DEBT. IT WOULD BE NEAR IMPOSSIBLE TO PAY....
Germany
France
Greece
Ireland
Italy
Spain
Portugal
IF THERE IS NO FIREWALL, OR IT'S INADEQUATE, THEN IT WOULD BE VERY EASY TO AFFECT OTHER COUNTRIES
GREECE'S DEBT, IS EUROPE'S DEBT! GREECE COULD SET OFF A CHAIN REACTION
Greece
Ireland
Italy
Spain
Portugal
MONEY CAN BE TRANSFERRED TO MORE STABLE COUNTRIES SUCH AS GERMANY IN A MATTER OF HOURS
Germany
France
Greece
Ireland
Italy
Spain
Portugal
France
Italy
Japan
United States
LOSSES COULD EXTEND TO AMERICAN BANKS...

Figure 3.6: Data Comic on the European debt crisis.



id : 233
Category : Peace
Year : 1993
laureate_id : 555
Motivation : "for their work for the peaceful termination of the apartheid regime, and for laying the foundations for a new democratic South Africa"
Joint Awards : 2 awardees

Figure 3.7: Webcam and Keshif visualization being recorded for the Nobel Prize Winner data video. A data video can be made with user as the presenter.

Chapter 4: Designing Data-Driven Tools for Casual Users

The overall purpose of this Ph.D. thesis is ***to expand the horizon of media for data-driven storytelling to aid casual users viewing, analyzing, and understanding data.*** Traditionally, infovis systems are designed for experts and professionals with strong domain knowledge.

4.1 Casual Users

To expand to a larger population including everyday internet users, the elderly, and young people, we consider casual users as the group of consumers of information such as data stories, who leverage storytelling and infovis systems to gain meaningful information for both everyday work and non-work situations.

4.2 Casual Information Visualization

Pousman et al. [88] proposed casual information visualization by expanding traditional visualization research to edge cases such as ambient infovis, social infovis, artistic infovis. The three sub-domains seem far from the core infovis research.

- **Ambient Infovis** are systems that can be loosely defined as infovis, with very sparse and abstract expressions of data.

- **Social Infovis** are collaborative visualization systems that work on social information.
- **Artistic Infovis** are information visualization systems that work on data-driven art.

All these different kinds of infovis are not strictly bounded by the traditional visualization tools which are widely used by the professionals such as Tableau [135], Microsoft Power BI [196], and Many Eyes [143].

4.3 Why Data-driven Storytelling for Casual Users

To make a greater impact on the largest user group, we choose to focus on the casual users. As smartphones are widely accepted by people of different ages, from young kids to elderly, large amount of computing power is spread as well. We want to leverage the computing power to level the digital data gap between experts and casual users with data-driven storytelling media.

In today's internet world, the majority of the users are young and open to new media types. One of the most popular Youtubers, "PewDiePie" [197], who makes original videos and live streaming on Youtube for a living, has a subscriber count of 70 million. This size is larger than some that of the population of large countries.

For the casual users without professional training, using visualization tools can be both an opportunity and a challenge. The visualization tools have already provided a powerful way to analyze, understand, present, and interpret raw data, but current visualization tools are often too difficult for inexperienced users.

4.4 Which Media to Use

Most of the media types on the internet, such as online articles, online videos, live streaming, online infographics and etc, are well accepted by casual users. For instance, we developed DataTV using live streaming and data visualization to help data videos broadcast to casual users. . We also developed Data Comics with comic as media, as well as an online application to develop data comic for online users.

Chapter 5: Data Comics

In this chapter, we present Data Comic, a technique that can be created by juxtaposing multiple visualizations into comic strip layouts for casual users as consumers of information. Data Comic consists of a sequence of panels, each annotated with both visual and textual elements, and arranged into a sequence that progressively develops the overarching story told in the comic.

To facilitate the creation of Data Comics, we present DataComicsJS (Figure 5.1(b)), a Google Chrome extension that consists of four components: (1) the Clipper, for collecting both snapshots of visualizations and images as well as raw data from any webpage viewed in the browser; (2) the Decorator, for editing the visual design of an individual panel, including clips, images, captions, and comic-style visual elements; (3) the Composer, for managing the layout, size, and position of panels making up the comic; and (4) the Presenter, for ultimately allowing a viewer to navigate in a finished Data Comic, including viewing the entire comic as a whole, as well as view single panels in sequence.

5.1 Definition

Data Comics is a visual storytelling method based on sequential images consisting of data-driven visual representations. Its purpose is to support for expert users to build engaging narratives about data. Our inspiration for this method came from several sources, including the recent focus on storytelling for visualization [198], the increasing use of comics for "serious" applications (e.g [175, 199, 200]), and the EuroVis 2011 keynote by Scott McCloud on comics [156].

Our motivation is to take advantage of both the plethora of existing visualizations on the web and the familiar visual language of comics, including layout, characters, and comic elements such as motion lines, speech bubbles, and arrows. Below we review an operational model for data comics and then discuss each of its aspects, including creating panels, managing their layout, and letting a viewer navigate the comic.

5.1.1 Basic Model

For the purposes of this chapter, a comic consists of a sequence of *panels* organized into one-dimensional *tiers* (or strips) and separated by *gutters*, or spacing, between the panels [149, 175]. The panels in a tier are organized to be read from left to right to form a narrative (at least in Western cultures). Tiers can in turn be organized into *pages*, where each tier becomes a row separated by a vertical gutter, and several pages can be linked together into a *book* (or comic book).

5.1.2 Panel Content

Unlike a normal comic, most panels in a Data Comic consist of visualizations that convey information using graphical means.[1] These could be simple and familiar statistical graphics such as barcharts, time-series charts, and piecharts, or more advanced visualizations such as treemaps [201], node-link diagrams, or even parallel coordinate plots [202], all depending on the visualization literacy of the intended audience and the instructional annotations in the panel.

Because of this focus on data-driven graphics, the designer is largely relieved from creating artistic content, which requires drawing skills that only few people have. Instead, the visual content can be constructed by either creating entirely new visualizations from raw data, or by clipping a snapshot from an existing visualization.

5.1.3 Characters, Annotation, and Effects

A Data Comic would not truly be a comic if it did not also leverage the visual language of comics. Designers creating Data Comics can be given access to this in several ways:

Comic-style rendering: To emphasize the comic medium, content can be drawn using non-photorealistic rendering (e.g. [203]).

Characters: Characters often drive the narrative and complement the data-driven visualizations. Because this requires artistic talent, designers should be given

[1] For engagement and effect, a few panels may consist solely of artistic content, but this puts corresponding artistic burden on the designer.

a library of characters.

Comic elements: Designers should also be given access to common visual elements used in comics, such as motion lines, highlights, or even onomatopoeia (words that mimic sounds).

Captions, speech, thoughts: Visuals are often scaffolded by text in captions as well as speech and thought balloons [149, 175].

5.1.4 Layout Management

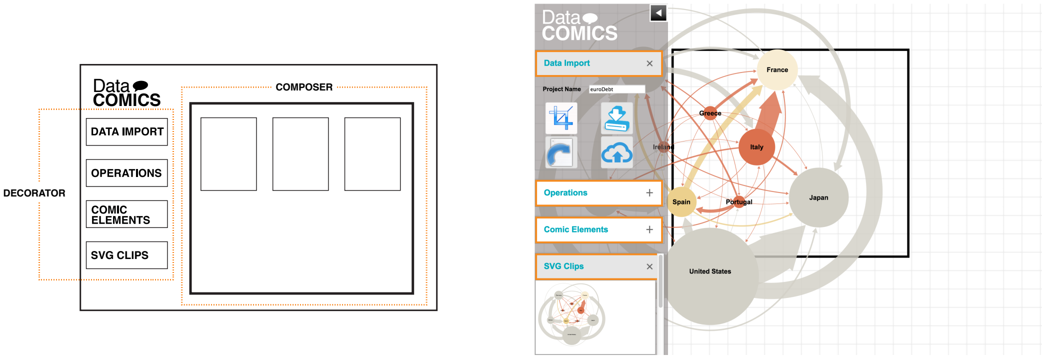
The layout of a Data Comic—the organization of panels into tiers and pages—is an important consideration in creating a narrative. To facilitate easily constructing a narrative, the Data Comics model should allow a designer to easily change the order of panels.

5.1.5 Viewing

Finally, after a Data Comic has been created, its purpose is to be viewed by its intended audience to convey its designer's story (and message). Just like a traditional comic, the default view for a Data Comic is to view an entire page, with all of the panels visible. Since screens are different from the written page, however, it also makes sense to support a single-panel navigation mode, where the viewer can sequentially navigate backwards and forwards in the comic. This is not unlike traditional slideshows in PowerPoint or Keynote.

5.2 Implementing Data Comics

We have implemented Data Comics as a web application called DataComicsJS. It is a hybrid application consisting of both client-side and server-side components. Client-side components are built using JQuery for DOM manipulation and D3 for visualization. The content is stored on the server-side backend, implemented as a simple Python server communicating using JSON-RPC.



(a) Composer in DataComicsJS.

(b) Data importing.

Figure 5.1: (a) Basic DataComicsJS interface with the Composer in the middle of the workspace and Decorator on the left side. The Decorator is composed of four expendable menus including data importing, operations, comic elements, and collection of clips. (b) Data import is used to load and update visualization clips from a server where the user can also save finished work and reload them for presentation. SVG clips is the content captured by the Clipper.

5.2.1 Clipper

The Clipper component of DataComicsJS is implemented as a Google Chrome extension that users download and install in their local browser. This allows the system to integrate with and extend the browser so that users can clip content from any website they visit. This is achieved by traversing the DOM from the element that the user indicates. Three different types of content can be clipped: (1) Raw data: structured data, such as in an HTML table element or tab-delimited text, can be parsed and saved; (2) Snapshot (SVG): Parts or all of an SVG element can be clipped, including any CSS that affect its appearance; and (3) Snapshot (raster): A specified bounding box of the webpage can also be clipped as a raster screenshot.

5.2.2 Decorator

The panel is the basic building block of a Data Comic, and panels are created using the Decorator, which consists of a toolbox (graphical tools, such as translation, scaling, rotation, annotation, cropping, etc), the workspace (a blank area with panel borders that can be dragged to change the aspect ratio of the panel), the clip collection (the visual and raw data content that the user has collected using the Clipper; the user can simply drag and drop elements from the collection, or drop raw datasets to bring up a dialog to select the visualization), and a symbols library (comic-style elements and direct access to the Noun Project free icon library).

5.2.3 Composer

The Composer gives the user control over the layout and organization of panels for a specific Data Comic. The layout flows to fit the page from left to right, top to bottom based on the number of panels and the page width. The interface provides buttons for adding and removing panels as well as dragging and dropping them to change their order. Double-clicking on a panel will open it in the Decorator.

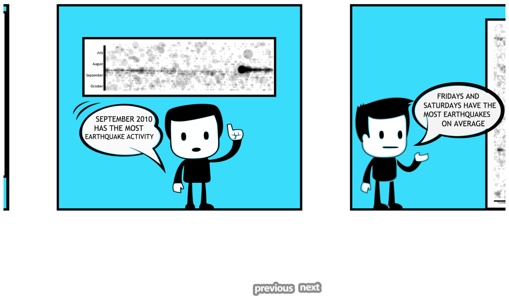
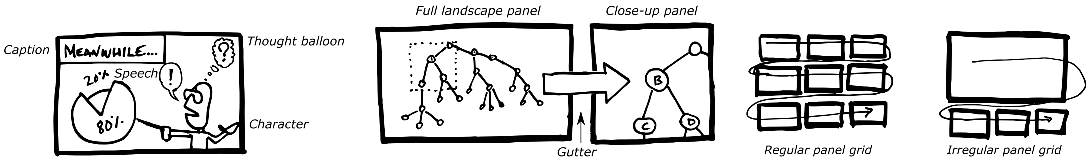

SEPTEMBER 2010 HAS THE MOST EARTHQUAKE ACTIVITY
FRIDAYS AND SATURDAYS HAVE THE MOST EARTHQUAKES ON AVERAGE
previous
next

Figure 5.2: Viewing individual panels using the Presenter in slideshow mode. Buttons allow for navigating slides in sequence.



(a) Text driving narrative.

(b) Seeing the whole from parts.

(c) Tiers and pages.

Figure 5.3: (a) The use of text to drive narrative. (b) Encapsulation of moments in panels juxtaposed yield closure as the viewer connects them and sees the whole. (c) Panels organized into tiers with horizontal gutters; multiple tiers form pages with vertical gutters.

5.2.4 Presenter

Once the designer has created and exported a Data Comic, he or she can share it widely using a unique URL that can be sent via e-mail, shared on social media, or posted on a forum. Everyone opening the URL will be given the Presenter view of the comic, which is a read-only viewer. The Presenter supports two viewing (and printing) modes: (1) viewing the entire comic, as is common for normal comics, and (2) viewing individual panels by navigating using on-screen buttons or the arrow keys (Figure 5.2)). The latter presentation mode is implemented as smooth animated transitions from one panel to the next to reinforce the comic metaphor.

5.3 Telling Stories Using Data Comics

Creating a Data Comics requires some knowledge of the narrative style of comics; the layout, content, and size of panels; as well as the use of text in driving the story. Here we draw on the general literature of comics to operationalize these concepts for data stories [175].

5.3.1 Comics Narration

While comics narration is generally linear, the medium by necessity cannot present the continuous narrative made possible by media such as digital video or film. Instead, comics narration depends on *encapsulation*, the focus on particular moments in the possible narrative arc to be represented in individual panels. Writers and pencillers exercise syntagmatic choice in establishing a series of juxtaposed images in the panels. The edges of the panels represent the limits of representation within this smaller bit of the page layout, but each panel also contributes to the larger unit, leading to an interpretive experience that is both linear and holistic since each panel is interpreted individually but also as part of the larger page. The gutters, while delineating the limits of each panel, also indicate the necessity for the reader to engage in *closure* in order to generate a more coherent narrative (Figure 5.3(b)).

The progression of syntagmatically related panels can follow a number of patterns. The progression may be temporal, following the same event as it unfolds in time. It may be spatial or spatial/temporal, moving between a number of locations at a given moment or during a specified time, or else presenting a series of localized micro-events happening within a larger framework (akin to a series of reaction shots to the same event used in film). Panels could also relate conceptually, presenting a series of related abstractions and calling upon the reader to form associations between them. The panels in Figure 5.5, operating in a relatively static timeframe and without a specific location, might best be thought of as such a conceptual arrangement, isolating and explaining particular phenomena within the broader structure of the U.S. Census data.

5.3.2 Panel Content and Size

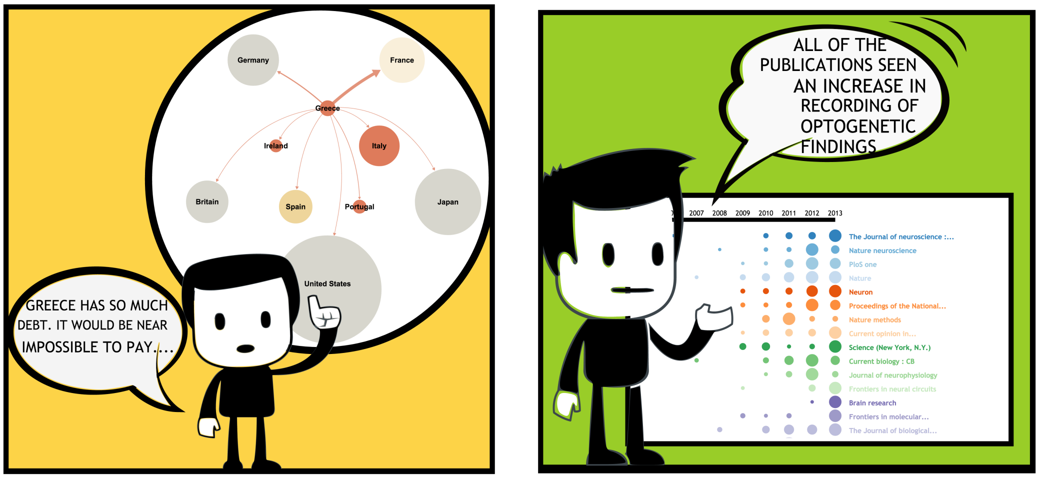
The comic panel is among the most fungible methods of representation, capable of presenting visual data ranging from a full landscape or urban horizon to a mid-range "street scene" shot to a close-up portrait shot or even the so-called extreme close-up of a particular detail. While layout possibilities are nearly limitless, most American (or European)-produced comics follow some variation of a grid pattern in which rows are read left to right while proceeding from the top to the bottom of the page. Panels might all be of equal size, as in Figure 5.5, but it is typical to encounter panels with particularly important details occupying more space on the page (Figure 5.3(c)). The extreme of this narration is the "splash page," in which one panel occupies an entire page, or even two pages, a rare panel known as a "double-page spread."

The combination of larger and smaller panels can often be used to convey a large quantity of visual information in a large panel while highlighting particular elements of the larger panel's content in smaller panels, such as presenting a landscape or a cutaway view of a house in a larger panel and then presenting close-ups of various visual elements in smaller panels collected near the larger panel. (The visual content need not be a natural or built environment, however: a similar layout for Figure 5.5 with a single large panel depicting the overview of the U.S. Census data could readily be supplemented by rows of smaller panels presenting the individual elements of the data set.)

5.3.3 Textual Narration

Comics panels may feature dialogue or textual narration, or these elements may be omitted, a stylistic choice embedded in the form's eighteenth-century origins. The satirical engravings of the British artist William Hogarth generally omitted dialogue, whereas politics cartoons dating from the eighteenth and nineteenth centuries often featured multiple characters speaking (to all appearances simultaneously) in a one-panel image. Narration and dialogue are typically omitted for action sequences, focusing the reader's attention on the implied movement of the characters or objects in the panels. Dialogue is generally placed in speech balloons that are superimposed on the image. Narrative captions can be placed in a box, usually located at the top of the panel, and delivered in the first, second, or third person (Figure 5.3(a)).

An entirely different method of narration is available in the form of direct address using a character in the comic, usually anthropomorphized to one degree or another. Breaking the so-called "fourth wall" between the page and the reader, such characters can function as avatars drawing the reader into the narrative or as interlocutors presenting essential information to the reader. (The narrator in Figure 5.5 represents the latter approach.) Scott McCloud combines both functions in his *Understanding Comics* [175] guide to the visual interpretation of the medium, deploying a cartooned avatar of himself to walk the reader through the essential physical elements of comics layouts and illustration styles while also dispensing information necessary to the reader's interpretation of these elements.


Germany
France
Greece
Ireland
Italy
Britain
Spain
Portugal
Japan
United States
GREECE HAS SO MUCH DEBT. IT WOULD BE NEAR IMPOSSIBLE TO PAY....
ALL OF THE PUBLICATIONS SEEN AN INCREASE IN RECORDING OF OPTOGENETIC FINDINGS
2007
2008
2009
2010
2011
2012
2013
The Journal of neuroscience :...
Nature neuroscience
PloS one
Nature
Neuron
Proceedings of the National...
Nature methods
Current opinion in...
Science (New York, N.Y.)
Current biology : CB
Journal of neurophysiology
Frontiers in neural circuits
Brain research
Frontiers in molecular...
The Journal of biological...

(a) The Greek debt crisis.

(b) Journals in neuroscience.

Figure 5.4: Examples of Data Comics. (a) The role of Greece in the European debt crisis. (b) Trends in scientific journals in neuroscience.

5.4 DataComics Examples

To validate the Data Comics method and give a concrete idea of what it is capable of, we present a few examples based on real-life data. The inspiration comes from online data visualizations and infographics we experience on a daily basis. Our work here reformulates some of these insights using the comics medium.

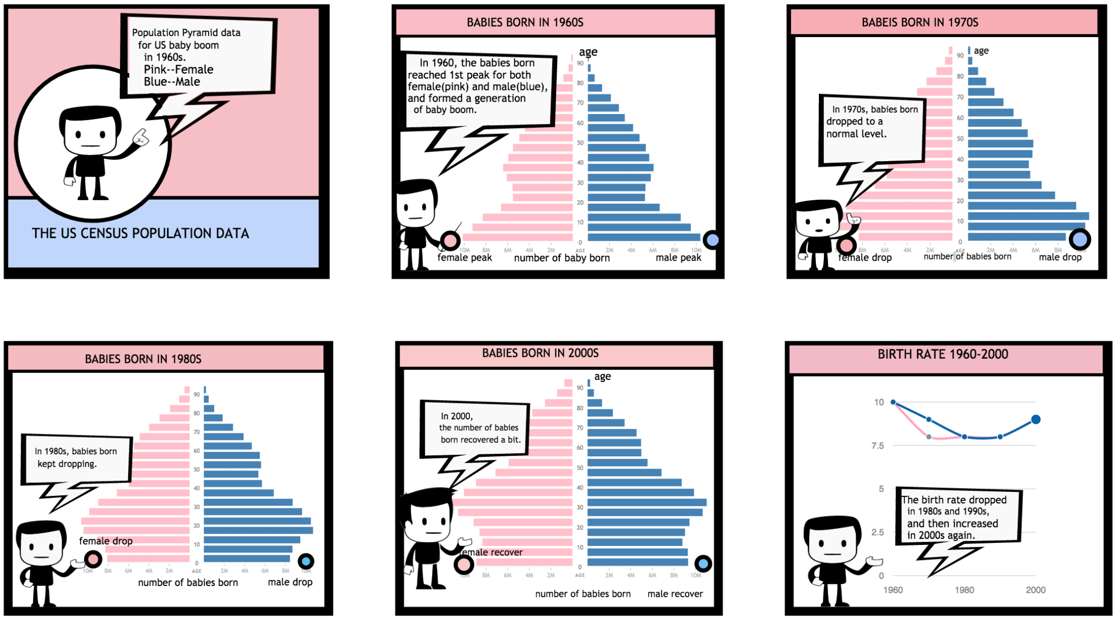

Population Pyramid data for US baby boom in 1960s. Pink--Female Blue--Male
THE US CENSUS POPULATION DATA
BABIES BORN IN 1960S
In 1960, the babies born reached 1st peak for both female(pink) and male(blue), and formed a generation of baby boom.
age
female peak
number of baby born
male peak
BABEIS BORN IN 1970S
In 1970s, babies born dropped to a normal level.
age
female drop
number of babies born
male drop
BABIES BORN IN 1980S
In 1980s, babies born kept dropping.
female drop
number of babies born
male drop
BABIES BORN IN 2000S
In 2000, the number of babies born recovered a bit.
age
female recover
number of babies born
male recover
BIRTH RATE 1960-2000
10
7.5
5
2.5
0
1960
1980
2000
The birth rate dropped in 1980s and 1990s, and then increased in 2000s again.

Figure 5.5: Data Comics of the U.S. baby boom of the 1960s.

5.4.1 Euro Debt Crisis

Dataset: The comic is based on a European Debt visualization available from The New York Times.[2] The visualization is a flow chart describing the debt relations between major European countries, the United States, and Japan.

Insight: The original visualization shows the direction and amount of the debt. It is obvious that the debt system is tangled, and the debt is huge. The Data Comic can tell the story in smaller, more manageable chunks.

Construction and Visualization: We snapshot the visualization and add comic figure and comic style text diagram to it. Figure 5.4(a) shows that the Greek debt crisis has a great negative impact on the economy of the whole Europe. The exaggerated lightning sign and the character figure is intended to show the seriousness

[2] `http://www.nytimes.com/interactive/2011/10/23/sunday-review/an-overview-of-the-euro-crisis.html`

of this crisis while adding a human angle. This shows how the Data Comic method can change the style of an existing visualization.

5.4.2 U.S. Census Population Pyramid

Dataset: Figure 5.5 is the Data Comics version of a U.S. Census dataset from 1960s to 2000s. [3]

Insight: The horizontal bar chart illustrates a population distribution based on an age scale and time scale in decades, partitioned by gender. We clearly see the babies born in each decade and the trend of the population growth. We want to make a story about the population growth during and after baby boom to reveal how the babies born are the key factor to shape the population structure.

Construction and Visualization: To track the changes of babies born during the period of 1960 to 2000, we captured the bar charts of population pyramid during this time period, and then generated another line chart based on these data. Within each panel, we labeled dimensions such as age, number of babies, and changes in births from previous years. The panels clearly show that the number of babies born decreased during the middle two of the four decades. Later, the number of babies born increased in the last bar chart.

5.4.3 Scientific Journal Comparisons

Dataset: The data is from a visualization comparing publication counts over time in scientific journals covering different topics in neuroscience and brain stimu-

[3] `http://vis.stanford.edu/jheer/d3/pyramid/shift.html`

lation. Data gathered from the U.S. National Library of Medicine.[4]

Insight: The comparison among different publications over time is vastly distributed on the visualization space. Through comparison, we found that the Journal Brain Stimulations has steady growth in the field of transcranial magnetic stimulation.

Construction and Visualization: We captured the visualization, and then added a hand-shaped indicator and a female figure to help guide the narrative. As part of the whole story of six panels we built for this topic, in Figure 5.4(b), the Data Comic performs the role as a personalized guide highlighting the difference for comparison.

5.5 Evaluation

We focus on two separate aspects of Data Comics in this chapter: first, how an analyst can go about *authoring* a data comic, and second, how an audience will respond to the *presentation* using a data comic. No single user study can explore both of these aspects, so below we report on two separate studies: authoring as well as presentation.

5.5.1 Study 1: Authoring Data Comics

Our primary contribution in this work is the concept of Data Comics rather than the prototype tool we have built to facilitate authoring them. After all, given sufficient time and effort, the comics shown in this chapter can be replicated using

[4] `http://neuralengr.com/asifr/journals/`

traditional drawing tools. However, we still wanted to study how the presence of a specialized authoring tool affected their creation. Our hypothesis is that when explicitly framing data-driven storytelling through the lens of comics, new modalities for visual communication arise.

5.5.1.1 Method

We performed an expert review with two data visualization professionals using the basic method proposed by Tory and Möller [204]. Two independent experts from a neighboring research group at our university used DataComicsJS to create a three-panel data comic based on their current work.These two experts have extensive experiences and knowledge over data visualization theory and tools. They were asked to follow an informal think-aloud protocol and allotted a total of one hour session.

P1: Visual Analytics Researcher: P1 is a visual analytics and human-computer interaction researcher at our university working on online visualization system design with more than three years of experience.

During the expert review, P1 raised several points: (1) *Better insight.* The process of having to determine a narrative structure for the data comic helped the expert gain a better understanding of the data. This may be common to storytelling in general, but the expert felt that the highly visual and visceral comic medium made this particularly clear. (2) *Better evidence.* The process of assembling sufficient material to compose the story was also beneficial in general in collecting evidence for the comic. (3) *Thinking in comics.* The expert noted that the fact that our DataComicsJS tool gives ready access to the visual language of comics—including layout, narrative structure, and visual elements—reduced the workload and helped P1 to think purely of comic storytelling rather than the mechanics of the interface. (4) *Web integration.* Being able to effortlessly clip imagery and data from the web was cited by the expert as one of the main benefits of the tool.

P2: Information Visualization Researcher P2 is a researcher specialized in web-based information visualization system design, particularly for big data. P2 has more than three years of experience in designing infographics and interactive dashboards.

During the expert review, P2 raised several points: (1) *Accessibility:* Comics, by its nature, tell stories in a very friendly and accessible way, which can be beneficial for data that is complex or even intimidating. (2) *Sequence:* Infographics typically relies on layout to deliver the story and there is no explicit sequence. If readers do not follow the layout in the way designed by the creator, they can lose the causality of the story. Comics use a pre-determined temporal sequence where the causal relation is obvious (see Section 5.3). (3) *Motivational.* The expert (P2) noted that creating data stories in a comic style was intrinsically fun and motivated the design process. It also invited thinking about how to best present the data using the visual languge of comics.

5.5.2 Study 2: Presenting Data Comics

Our working hypothesis in this chapter is that Data Comics provide a compelling way of telling stories about data. To empirically explore the virtues of this premise, we conducted a qualitative user study comparing Data Comics to traditional PowerPoint slideshows. The reason we chose PowerPoint is not to prove that our prototype implementation—DataComicsJS—is in any way superior to PowerPoint or any presentation software, but to pick an application and style of presentation widely used in the real world. Here we describe the methods and results from this evaluation.

5.5.2.1 Participants

Participants: We recruited 12 paid participants (6 male, 6 female) to participate in the user study. The participants were self-selected from the student population at our university, were aged between 20 and 31 years of age, had normal or corrected-to-normal vision, and were proficient computer users (all demographics were self-reported).

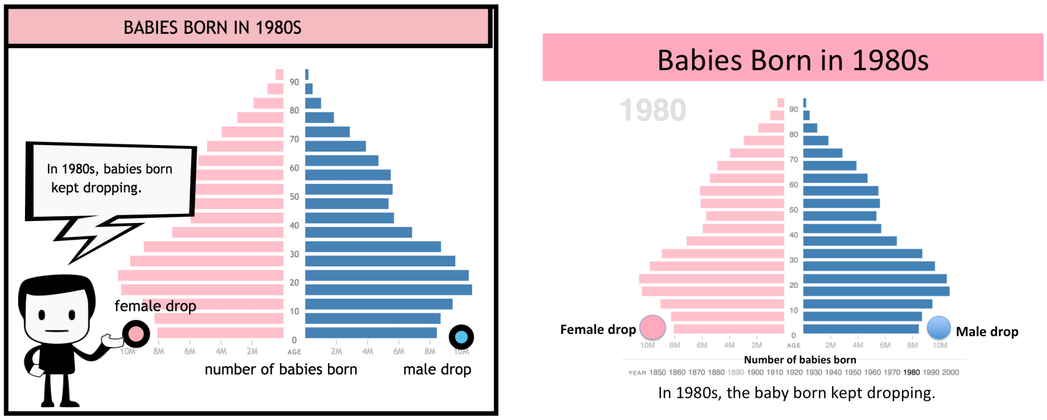
5.5.2.2 Apparatus

Apparatus: We conducted the experiment on a laptop computer equipped with a 15-inch 1280 × 800 LCD screen, a standard keyboard, and a three-button mouse. Both the Data Comics prototype and Microsoft PowerPoint was maximized to fill the screen during the experiment.

5.5.2.3 Task and Datasets

Task and Datasets: Each trial consisted of the participant using either MS PowerPoint or Data Comics to answer questions about a data story. Each story consisted of several panels of narration. The number of panels were limited to five or six to keep the simplicity of layout while providing sufficient information for the story. Each story presentation focused on a single topic of visualized data and came with an associated list of 7 to 9 questions designed to make the participant focus on details of the visualization. Participants had access to all questions during the time they were interacting with the story.

We created four stories for the evaluation: Twitter heatmap for stocks (S_1), the U.S. Census Population pyramid (S_2), world happiness (S_3), and Star Wars character fans' personality rating (S_4).



(a) Panel in a DataComicJS story.

(b) Slide in a PowerPoint story.

Figure 5.6: A comparison of the story frames of Birth data from the U.S. Census Bureau with DataComics and PowerPoint. Stories are composed with different methods but with one-to-one correspondence in details to make the user study fair.

The stories were first created as a Data Comic by clipping from online data sources and visualizations, creating an appropriate number of panels to tell the story, and finally decorating the panels with comic annotations, characters, and captions. We then created a corresponding PowerPoint slideshow with the exact same number of slides as the number of panels in the Data Comic. There was a one-to-one correspondence between the visualizations and captions between panels and slides as Figure 5.6; the only difference between the two versions was that the PowerPoint slideshow only used visualizations and text, whereas the Data Comic included characters and comic-style symbology. Furthermore, while the text explanations were identical across versions, the Data Comic version integrated them in comic-style captions or speech and thought balloons.

5.5.2.4 Metrics

Our focus with the evaluation was not to primarily study quantitative metrics, such as time and accuracy, but to collect subjective and qualitative feedback on the difference between Data Comics and traditional slideshow presentations. For this reason, we developed a questionnaire polling participants on their subjective experience of a story. This was administered to participants directly after each story, and consisted of the following 1–5 Likert-scale questions: engagement, speed, space efficiency, ease of use, and enjoyability. We also asked participants for general feedback on the tool.

5.5.2.5 Factors

We included two factors in the experiment, described below.

Presentation (P): This factor modeled the presentation technique P given for solving questions in a trial: *1) Data Comics:* The narrative visualization is presented as a Data Comic in the Presenter in our prototype implementation. Participants were able to view the entire comic and navigate panel by panel in the comic. *2) PowerPoint:* The narrative is presented as a PowerPoint slideshow. Participants can navigate backwards and forwards in the slideshow. They can also view all of the slides at once in the "slide sorter" view.

Story (S): We also hypothesized that the specific story and topic of the data visualization may impact our outcome. Thus we added a factor S to model the different stories.

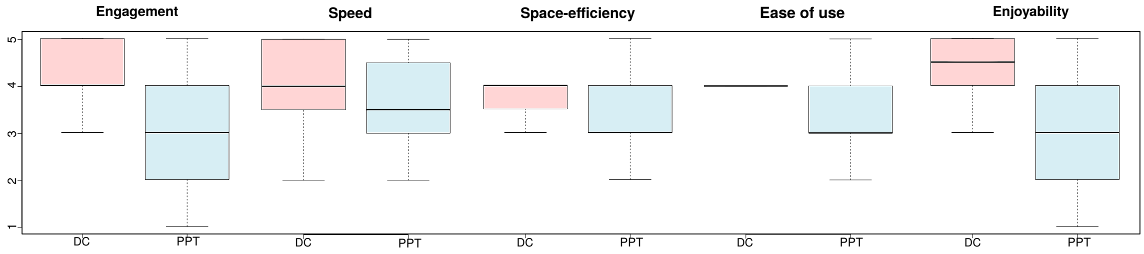


Figure 5.7: Comparison between DataComics (DC) and PowerPoint (PPT) of subjective ratings (Likert 1-5 scale) for engagement, speed, space-efficiency, ease of use, and enjoyability.

5.5.2.6 Procedure

An experimental session started with the participant arriving, reading and signing the consent form, and being assigned an identifier and story order. The administrator then explained the general goals and task. Each trial started with the administrator demonstrating how to use a Data Comic. The participant was then given two examples, one Data Comic and one PowerPoint, and was allowed to ask questions about the examples and task during this time.

When the participant finished the training, they were given a story opened in the appropriate tool and a paper sheet with questions. They were given up to 10 minutes to answer the questions, and were encouraged to use all of the time. After answering all questions, the participant was given the subjective questionnaire polling their experience in the trial. This was repeated for all four stories—two using Data Comics, two using PowerPoint. A full session lasted approximately 50 minutes, including training and questionnaires.

5.5.2.7 Quantitative Results

Figure 5.7 depicts boxplots of the subjective ratings for both Data Comics and PowerPoint on engagement, speed, space-efficiency, ease of use, and enjoyability (Q1 through Q5). We analyzed the 5-point Likert scale of subjective ratings for effects of presentation technique P (Data Comic vs. PowerPoint), and found that the engagement (Q1), efficiency (Q3), and enjoyability (Q5) were significantly different between the two techniques (Friedman tests, $p < .05$), but the speed (Q2) and ease of use (Q3) had no significant difference (Friedman tests, $p = .51$ and $p = .08$ separately). We also found no significant effect of story S on any of the metrics.

5.5.2.8 Qualitative Feedback

Inviting Reading: Nine out of twelve participants mentioned that the comic-style rendering helped them view the materials as a whole story from the very beginning without any explicit direction. They noted that the speech balloon helps focus by creating a feeling that there is a virtual conversation going on, and the comic figure kept them more involved in the scenario of the story. All these comments from the participants suggest that Data Comics invite reading, even when incorporating only simple and trivial comic elements.

Viewing as a Story: The Presenter is organized to show not just the current panel, but also the two surrounding ones (Figure 5.2). While this nominally is a waste of visual space—a slideshow shows each slide in full-screen mode, yielding more pixels to complex visualizations—participants seemed to enjoy this view, presumably because it suggests continuity and story flow ("there is more to see beyond this panel") and it evokes the "comic state of mind" we seek. Our observations and interviews confirmed this fact; the extra context panels seem to encourage participants to keep reading. Also, our Likert scale results show that participants actually felt that comics were more space-efficient (Q3) than the slideshows.

In fact, we observed that all but two of our participants would start each Data Comic task by first reading through the entire comic from beginning to end. Thus, the comics format seems to invite reading. This is in contrast to the PowerPoint slideshows, which no participants were observed to read fully before answering questions. Several participants remarked that the slideshows did not "feel" like stories, but rather information sheets that they were just flipping through to find information. While there is no intrinsic value to this story aspect of Data Comics, we do think it increases user engagement, as evidenced by our quantitative results.

Facilitating Memory: Eight out of twelve participants mentioned that the comic version of each story helped them remember the contents, even down to the individual panel for specific information. Figures are naturally memorable, even when they are not relevant to the current topic; this mirrors findings by Bateman et al. [205] on the beneficial effect of "chart junk" on recall in visualizations and infographics. We also noted that a Data Comic does not need to be designed in a very artistic way, but can incorporate basic clipart-like imagery and graphics. Participants also mentioned that even a little variation of comic figures can help distinguish frames.

5.5.3 Study 3: Partitioning and Sequence in Storytelling

Our working hypothesis in this chapter is that data comics provide a more effective way of telling stories than a single visualization. To empirically explore the virtues of this premise, we conducted a qualitative user study comparing multi-panel data comics with a single infographic-style visualization for the same data. Here we describe the methods and results from this evaluation.

5.5.3.1 Participants

We recruited 12 paid participants (9 male, 3 female) to participate in the study. The participants were self-selected from the student population at our university, were aged between 20 and 31 years of age, had normal or corrected-to-normal vision, and were proficient computer users (all demographics were self-reported).

5.5.3.2 Apparatus

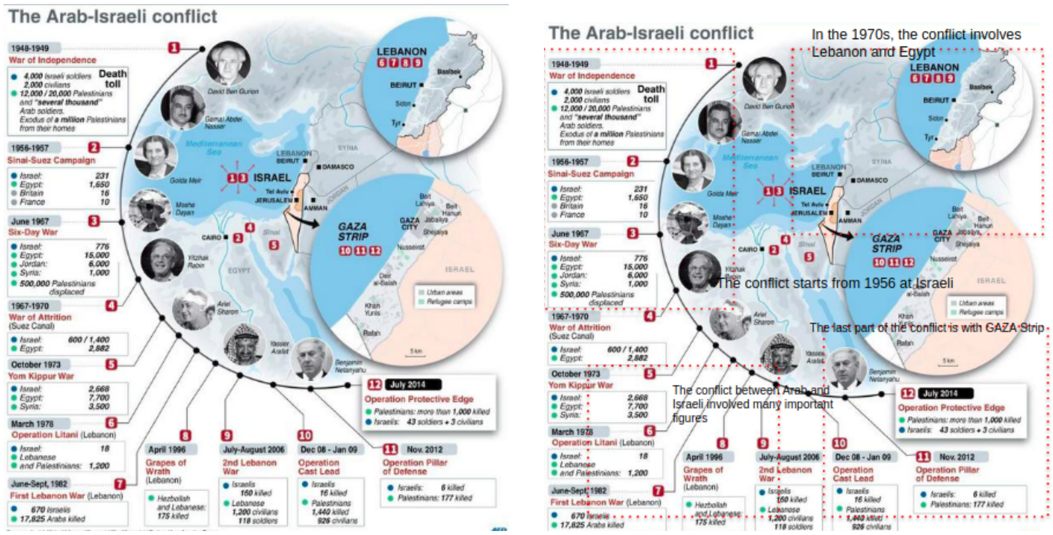
We conducted the experiment on a laptop computer equipped with a 15-inch 1280 × 800 LCD screen, a standard keyboard, and a three-button mouse. Both the data comics and the infographic were maximized to fill the screen during the experiment.

5.5.3.3 Task and Datasets

Each trial consisted of the participant using four types of data stories to answer questions about a data story: a single infographic versus a data comic, with and without captions for the different parts (see Section 5.5.3.5 below). Each story consisted of several panels of narration. The number of panels was limited to five or six to keep the simplicity of layout while providing sufficient information for the story. Each story presentation focused on a single topic of visualized data and came with an associated list of 7 to 9 questions designed to make the participant focus on details of the visualization. Participants had access to all questions during the time they were interacting with the story.

We created four stories for the evaluation: Beer Origin Map (S_1), the Arab-Israeli Conflict (S_2), Smart Phishing Attacks (S_3), and World Wealthy People Distribution (S_4).

The sequences of stories assigned to all the participants are the same. We mixed the methods for stories to counter-balance the learning effect for the methods. The four methods and the four stories are mix-matched into sixteen combinations, where we pick twelve to assign to twelve participants randomly.

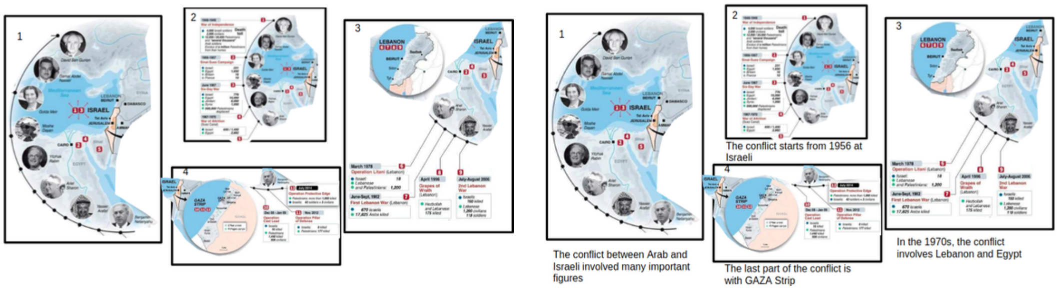

The Arab-Israeli conflict
1948-1949 War of Independence
1956-1957 Sinai-Suez Campaign
June 1967 Six-Day War
1967-1970 War of Attrition (Suez Canal)
October 1973 Yom Kippur War
March 1978 Operation Litani (Lebanon)
June-Sept, 1982 First Lebanon War (Lebanon)
April 1996 Grapes of Wrath (Lebanon)
July-August 2006 2nd Lebanon War
Dec 08 - Jan 09 Operation Cast Lead
Nov. 2012 Operation Pillar of Defense
July 2014 Operation Protective Edge
LEBANON
ISRAEL
GAZA STRIP
In the 1970s, the conflict involves Lebanon and Egypt
The conflict starts from 1956 at Israeli
The last part of the conflict is with GAZA Strip
The conflict between Arab and Israeli involved many important figures

(a) Composer in DataComicsJS.

(b) Data importing.

Figure 5.8: (a) The original infographics without any changes. (b) Adding red dotted boxes to highlight the locations for the panels.

The stories were created based on existing infographics that we found online. The data stories are based on the topic, structure, and organization of online visualization. The organization of an infographics can be classified into several genres [2], including overview-detail, cause-effect, chronological, etc. We first follow these patterns to find stories from our selected infographics. Then we partition storyline into panels and write captions to form the final data comic.


1
2
3
4
1
2
3
4
The conflict starts from 1956 at Israeli
The conflict between Arab and Israeli involved many important figures
The last part of the conflict is with GAZA Strip
In the 1970s, the conflict involves Lebanon and Egypt

(a) Composer in DataComicsJS.

(b) Data importing.

Figure 5.9: (a) The infographic is partitioned into panels. (b) The infographic is partitioned into panels and captions are added to help address the storyline.

5.5.3.4 Metrics

Our focus with the evaluation was to both study quantitative metrics, such as time and accuracy, as well as to collect subjective and qualitative feedback on the difference between data stories of data comics and original visualization. For this reason, we developed a questionnaire polling participants on their subjective experience of a story. This was administered to participants directly after each story, and consisted of the following 1–5 Likert-scale questions: engagement, speed, space efficiency, ease of use, and enjoyability. We also asked participants for general feedback on the tool.

Moreover, we forced the participants to answer correctly. We also found that the amount of time spent by each participant with different techniques is comparable. In this case, we only study the subject score of 15 Likert-scale questions on engagement, speed, space efficiency, ease of use, and enjoyability.

5.5.3.5 Factors

We included three factors in the experiment, described below.

- **Presentation (P)**: This factor modeled the presentation technique P given for solving questions: infographic (IG) or data comic (DC). In other words, this was the primary factor intended to differentiate between different data story mechanisms.

- **Captions (C):** Whether or not the participant had access to the partitions and captions. For the infographic, having access to the captions would show the bounding boxes of the partitions as well as the associated caption we had written for each partition (Figure 5.8, Figure 5.9).

- **Story (S)**: We also hypothesized that the specific story and topic of the data visualization may impact our outcome. Thus we added a factor S to model the different stories.

5.5.3.6 Procedure

An experimental session started with the participant arriving, reading and signing the consent form, and being assigned an identifier and story order. The administrator then explained the general goals and task. Each trial started with the administrator demonstrating how to read a data comic. The participant was then given four examples, two with original visualizations (w/wt highlight of panel loca-

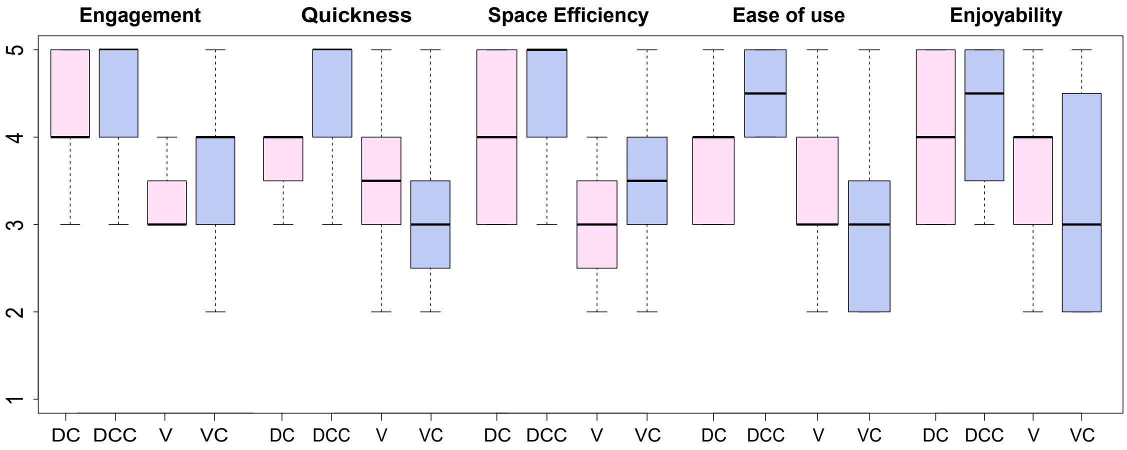


Figure 5.10: Comparison between single visualization (V), visualization with caption (VC), data comic panels without captions (DC) and data comic panels with captions (DCC) of subjective ratings (Likert 1-5 scale) for engagement, speed, space-efficiency, ease of use, and enjoyability.

tions) and two with data comic (w/wt captions), and was allowed to ask questions about the examples and task during this time.

When the participant finished the training, they were given a story opened in the appropriate tool and a paper sheet with questions. They were given up to 10 minutes to answer the questions, and were encouraged to use all of the time. After answering all questions, the participant was given the subjective questionnaire polling their experience in the trial. This was repeated for all four stories—one with the infographic, one with the infographic with captions, one with panels without captions, and one with panels with captions. A full session lasted approximately 50 minutes, including training and questionnaires.

5.5.3.7 Quantitative Results

Figure 5.10 depicts boxplots of the subjective ratings for the four types of tasks: a) infographic, b) infographic with highlights of panel focus locations and captions, c) data comic without captions, and d) data comic with captions. The ratings are for following effects on a 5-point Likert scale: engagement, speed, space-efficiency, ease of use, and enjoyability (Q1 through Q5). We analyzed the 5-point Likert scale of subjective ratings for effects of the technique P (infographic vs. data comic) and captions C (no captions vs. with captions), and found that the engagement (Q1), speed (Q2) efficiency (Q3), and enjoyability (Q4) were significantly different between the four techniques (Friedman tests, $p < .05$), but enjoyability (Q5) had no significant difference (Friedman tests, $p = .12$). We also found no significant effect of story S on any of the metrics.

5.5.3.8 Qualitative Feedback

Easy to Start and Follow the Story: Reading a infographic isn't always easy. Some infographics are designed for professionals as they are packed with information of all kinds. Partitioning the visualization into panels, especially when adding captions, can help the reader follow the sequence of panels and captions to generate a thread. One participant mentioned that the Arab-Israeli conflict example is overwhelming from the first look, but that he first data comic panel is a great summary for the whole visualization with all the excessive supplementary information partitioned to other panels. Another participant mentioned that she can skip a few panels when reading through the data comic panels while still following the big picture of the story.

Facilitating Focusing: The data comic panels are organized in a sequence following the storyline suggested by each infographic. The audiences' attention is directed by the panel, so that the important information is contained by certain panels. One participant mentioned that during the user study, he was able to easily locate a couple of panels whenever he needed a certain kind of information. Another participant mentioned that having captions and panels is like having labels for the whole visualization.

Facilitating Memory: Individuals have different habits when reading an infographic. People still start from different position to read a big infographics. In our study, several participants mentioned that the panels suggested a structured and progressive way of reading to build up the information through the sequence of the panels. Five of the twelve participants mentioned that reading the data comic panels helped them remember the information when answering questions. Even they did not catch the information in details, they were able to go back to the correct panels most of the time. This result matches the findings from Borkin et al. [206], which shows that the visualizations are more memorable when including pictograms or cartoons of a recognizable image.

Chapter 6: DataTV

Past work has shown that animated narratives can be particularly effective for data-driven storytelling [153, 154]. In this chapter, we describe DataTV, an approach to live data video production using online streaming technologies. DataTV is implemented as an integrated system, which provides a tool for the casual users as consumers of information to view data video stories created from narration, interactive visualization and annotation, and so on.

6.1 Design: Supporting Streaming Data Video Production

We claim that there is a need for a multimedia platform for creating *live data videos* at a pace and scale where they can be streamed and uploaded to an online video sharing service, such as YouTube or Twitch. Here we describe the major design decisions of the DataTV platform we design to meet this need.

D1 **Standalone application:** We design our tool to be a standalone desktop application rather than a web-based one.

- Video production and real-time streaming requires high performance processing and significant storage.

- No existing streaming software is entirely web-based; in fact, some even employ specialized hardware.
- *Alternative:* A web-based tool is platform-agnostic, but the performance demands for real-time video capture are too high.

D2 **Web integration:** We embed a web browser as a capture source to support web-based visualizations.

- Toolkits such as D3 [137] have made the web a unified platform for delivering visualization to the masses.
- Modern web browsers are full-featured multimedia platforms supporting a wide range of content, including video, sound, speech, vectors, 3D, etc.
- *Alternative:* A web-based tool would trivially support web technologies, but is not practical due to performance constraints (see above).

D3 **Optimized workflow:** The tool supports streaming with a single user acting as talent, engineer, and producer.

- Sustainable workflows for creating streams on a weekly or even daily basis must be time-efficient.
- Most streamers on Twitch—even established ones with thousands of subscribers—operate alone.
- *Alternative:* Abandoning real-time control would prevent streaming.

D4 **Native media support:** We provide source handlers for many media types, such as video, music, webcams, etc.

  - Effective data videos incorporate multiple media types beyond “just” the visualizations themselves [153].
  - Capturing directly from on-screen windows trivially enables native support for all applications.
  - *Alternative:* Using special software for specific media breaks the workflow, requires expertise, and reduces efficiency, but requires integrating all of the media handlers in the same tool.

D5 **Simplified video production elements:** We design the tool to provide simplified video production operations using easily accessible actions.

  - Typical analysts do not have a background in video editing, much less storytelling using motion graphics.
  - Providing the building blocks of video production will help creators think in terms of storytelling rather than mundane tool operations.
  - *Alternative:* Dedicated video editing software have a richer set of video production elements, but their use would break the workflow.

6.2 DataTV: A Streaming Data Video Editor

We present DataTV, a prototype data video streaming utility. DataTV is a standalone desktop tool for multiple platforms that allows recording multiple video sources from any number of windows on the user’s desktop.

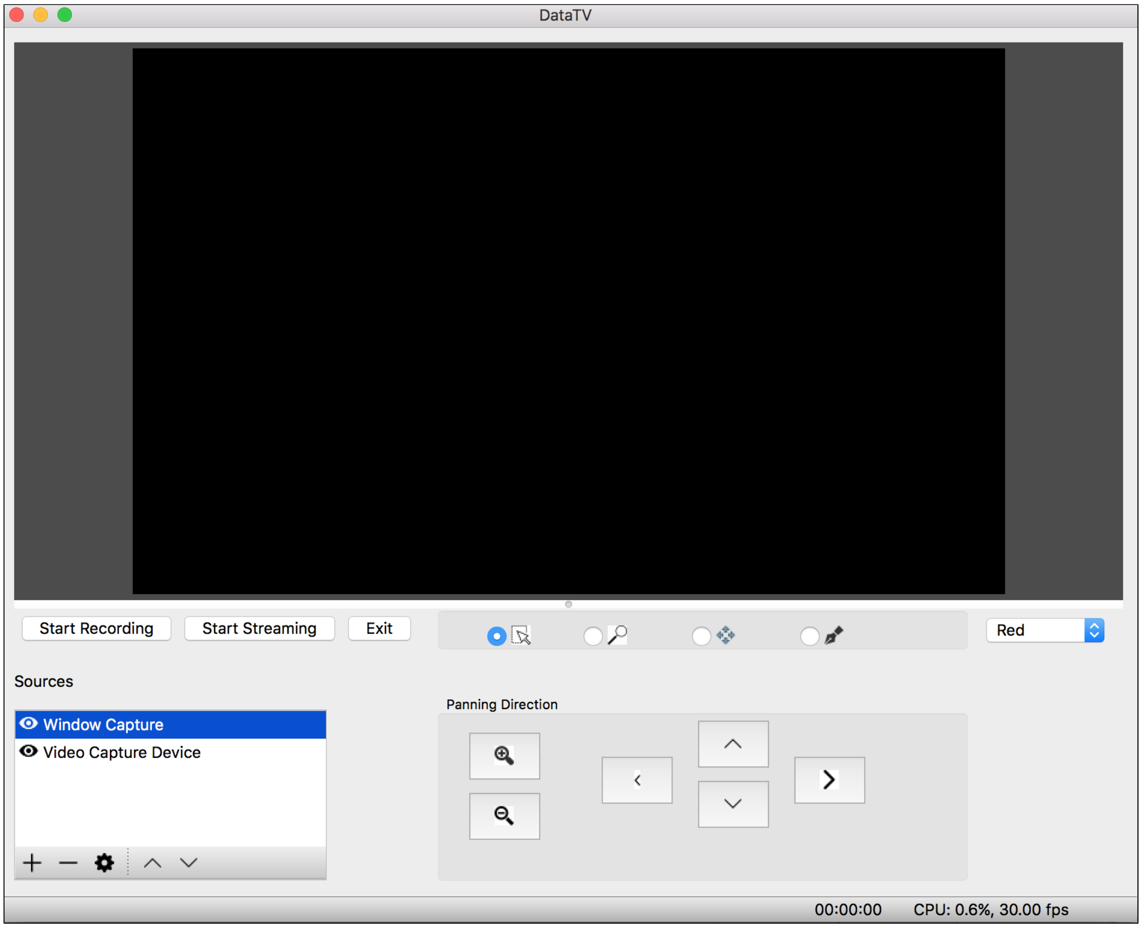


Figure 6.1: Main user interface of the DataTV prototype tool. A live mode toolbar allows for panning and zooming sources as well as scribbling directly on top of the output. The list panes at the bottom of the interface allow for controlling the scenes and sources being displayed.

6.2.1 Workflow

The DataTV tool is modeled along the standard workflow used in game streaming software such as OBS, XSplit, and GameShow, where the streamed output of the tool is managed using the concept of *sources* and *scenes*:

- **Source:** Streaming input such as from a window, specific application, audio source, or multimedia content.

- **Scene:** Composition of sources on an empty display that can be recorded and streamed as output from the tool.

DataTV operates in one of two distinct modes: (a) *offline* mode, where the user configures sources and scenes, or (b) *live* mode, where scenes are recorded (and possibly streamed). Most operations, such as creating, modifying, and deleting scenes and sources, can be performed in both modes to allow for users responding to live events (e.g., adding a new web-based visualization to the stream on a spur-of-the-moment idea), but the most common workflow is as follows:

1. *Preparation:* The user prepares all scenes and sources for recording in offline mode. This involves creating the sources, creating the scenes, and composing the sources for each scene on the output display canvas.

2. *Recording:* The user switches to live mode, sets up the stream settings (if enabled), and begins the recording. If the video is being streamed live, the user cannot easily switch back to preparation; instead, any changes to scenes or sources must be made while recording. If the video is merely being recorded and not streamed, the user can stop recording and go back to preparation, allowing the clips to be edited together at a later stage.

3. *Storytelling:* The user creates the data video in live mode. This entails switching between different scenes (by selecting the scene to display in the scene manager), managing specific sources (transforming, toggling on and off, annotating, etc), interacting with visualizations and other windows, and potentially narrating and/or capturing webcam video of the user.

6.2.2 Source Management

The tool maintains a list of currently available sources in the source listing window (Figure 6.1). Using this widget, sources can be created, toggled on and off, and deleted both during offline and live modes. Sources are created by selecting the source type and then associating the source with the appropriate object, such as a specific Google Chrome browser window on the desktop. Supported sources include the following:

- **Window capture:** Video output from a specific window on the desktop based on the window title, class, or executable.
- **Video capture:** Video output from an external device, such as a webcam or video camera.
- **Audio capture:** Audio output from a microphone.
- **Media objects:** Sources based on static images, video files, and rich text.

6.2.3 Scene Management

A *scene* in DataTV is an empty canvas containing sources that form the current output of the tool. Only one scene can be active at a time in DataTV; the active scene is displayed in the main composition window (Figure 6.1) and governs what is recorded and streamed when switching to live mode. The user can easily switch scenes using the scene list.

Managing a scene essentially entails managing the sources involved in the scene. In offline mode, most scene management operations are “heavyweight” in that they require significant setup that is not amenable to live recording. Examples of such operations include adding sources to the scene, managing their depth order (governing the drawing order of the sources), and transforming them (translating, scaling, rotating, and cropping). These operations are achieved by interacting with the sources in the composition window.

In live mode, users should mostly use “lightweight” operations that are designed for easy interaction while recording. This includes toggling the visibility of sources, zooming and panning in a source using mouse dragging and the mouse wheel (such as to zoom in on a particular part of a visualization or window), and scribbling using the annotation feature.

6.2.4 Video Annotation

Sportscasters regularly use annotation to scribble symbols and highlights directly on the video feed, such as to explain specific events in an instant reply of a touchdown or goal. Presentation software such as Microsoft PowerPoint supports similar “ink annotations” where the presenter can draw pen and highlight strokes directly on a slide to illustrate a specific point. DataTV supports a similar video annotation feature through its live mode interface (Figure 6.2), which provides a pen, highlighter, and eraser tool. The interface also allows the user to select the drawing color as well as to clear all of the annotation from the output when moving to a new scene.

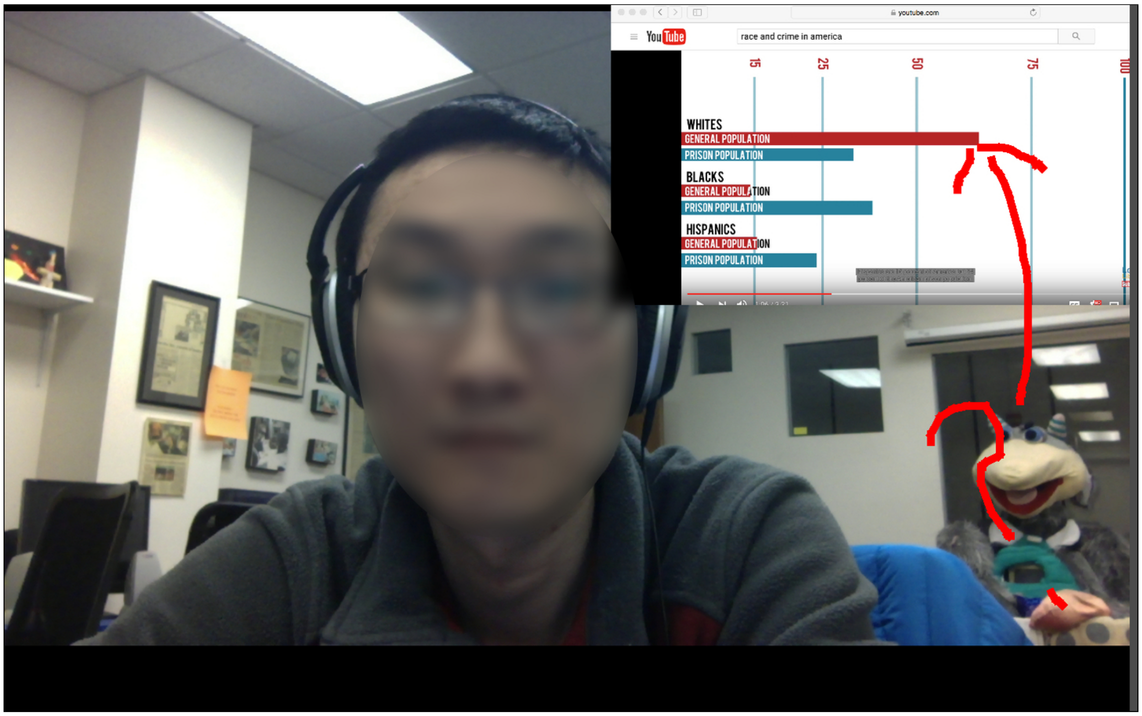


Figure 6.2: DataTV's live mode interface where users can zoom and pan in a data source as well as annotate using a pen, eraser, and highlighter.

6.2.5 Recording and Live Streaming

Our DataTV prototype supports recording to standard native video file formats (FLV, MP4, MOV, etc) as well as live streaming output to services such as Youtube and Twitch.tv. The tool also provides full control over stream settings for both video and audio output.

6.2.6 Implementation

We implemented our DataTV prototype based on OBS (Open Broadcaster Software) Studio, an Open Source live streaming package for multiple platforms.

Our extensions were made in C++ and significantly modifies the workflow of the tool to include a streamlined live mode interface, including scene management, zooming and panning, and live video annotation.

6.3 DataTV Examples

To validate the DataTV prototype and to give a concrete idea of what it is capable of, we present a few examples based on real-life data and existing visualization systems. The inspiration for these DataTV examples comes from creating and understanding web-based data visualizations and infographics we experience on a daily basis. Our work here reformulates and reimagines some of the insights from the data video creation process by Amini et al. [153, 154].

The data videos below consist of the following media sources:

- Live webcam video;
- Live microphone recordings;
- Web-based visualizations in a browser; and
- Video and images.

The media source selection for each story varies depending on the needs of each particular story.

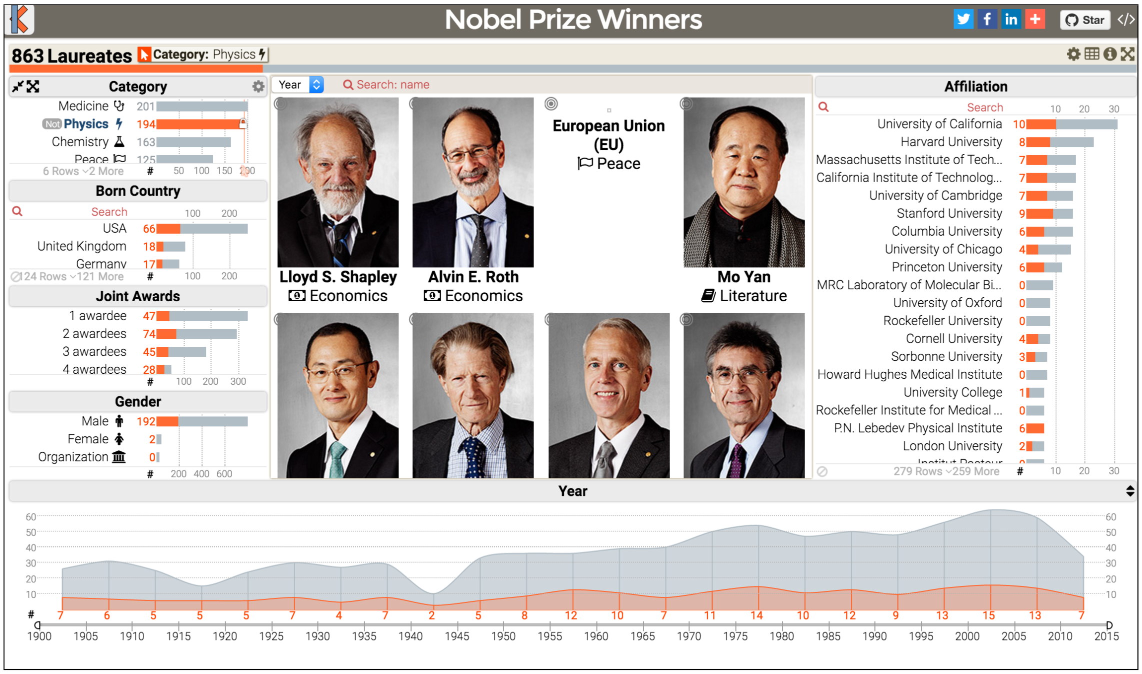


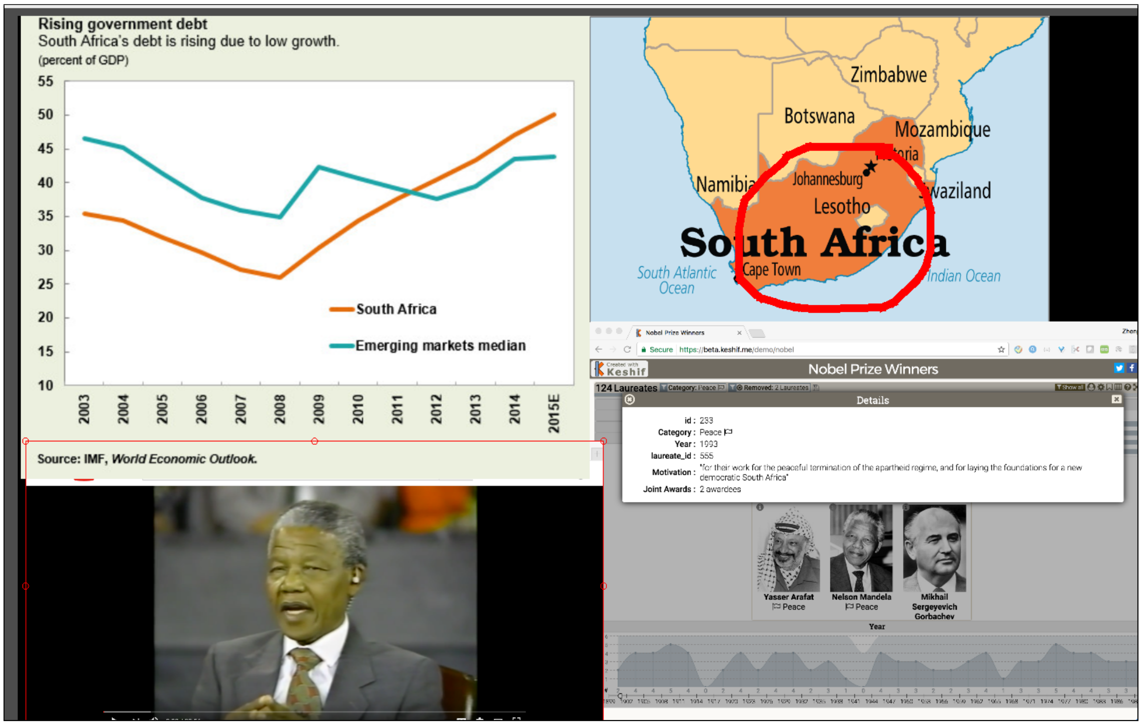
Figure 6.3: Keshif browser for the Nobel Prize Winners dataset.

6.3.1 Nobel Prize Data Analysis with Keshif

The first example uses the Keshif [136] system[1] to explore the "Nobel Prize Winners" dataset. Keshif is a visualization system designed for interacting with multi-dimensional data using a sophisticated faceted browser. The key interaction in the Keshif system is *linked selection*, which is a generalization of brushing and linking supporting highlighting, filtering, and comparison.

The Nobel Prize Winners dataset is composed of the basic profiles of all winners, including their pictures, year of winning the prize, nationality, etc. The overall story of this example is to briefly introduce South Africa, particularly its former leader and Nobel Peace Prize winner, the late Nelson Mandela. The analyst uses a YouTube video of Mandela (Figure 6.4) giving a speech as the introduction. Live

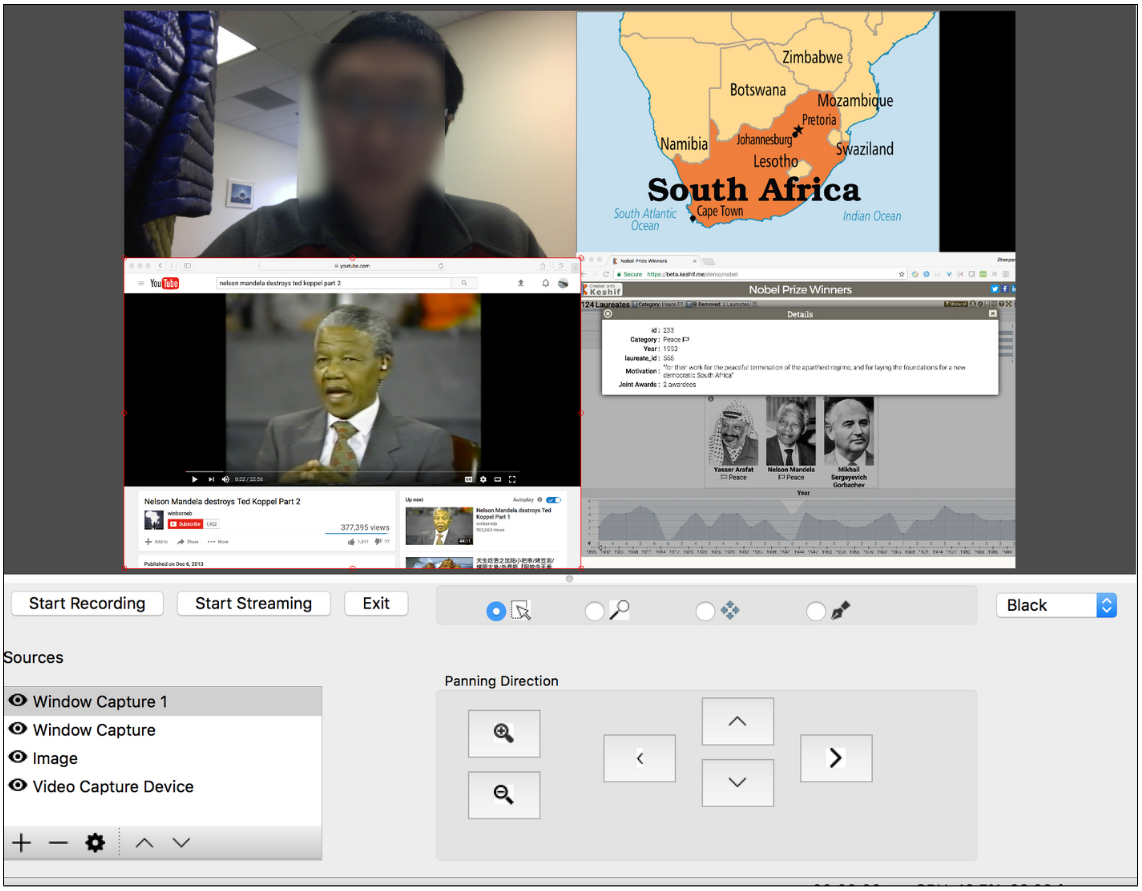
[1] http://keshif.me/



id : 233
Category : Peace
Year : 1993
laureate_id : 555
Motivation : "for their work for the peaceful termination of the apartheid regime, and for laying the foundations for a new democratic South Africa"
Joint Awards : 2 awardees

Figure 6.4: A live DataTV session composing a data video using the Keshif system for the Nobel Prize Winners dataset. The user imports an external visualization tool to display the economy of South Africa.

video used in this way will make the presentation engaging and draw in the viewer. It is easily achieved in DataTV using a YouTube source and controlled live by the user; no specific off-line editing is needed. Then we turn off the video and use a live window with Keshif to lead the story from introducing the categories of Nobel Prize to their age distribution. Again, the storyteller can do this in real-time simply by interacting with the Keshif visualization in a normal web browser window, potentially narrating his findings using the webcam and microphone. The DataTV platform will capture all of these sources, compose them in real-time, and stream them to a remote server. Compared to static media types, such as infographics, the ability to use an interactive interface enables the analyst to change the topic and approach during the storytelling. For example, potentially in response to an audience question through the Twitch chat service, the analyst may decide to give a little history of the Nobel Peace Prize, as well as Nelson Mandela's term as president, culminating in him winning the Nobel Prize. We then conclude the session with a short YouTube video of him giving another speech.



Details

id : 233
Category : Peace
Year : 1993
laureate_id : 555
Motivation : "for their work for the peaceful termination of the apartheid regime, and for laying the foundations for a new democratic South Africa"
Joint Awards : 2 awardees

Figure 6.5: Webcam and Keshif visualization being recorded for the Nobel Prize Winner data video. The author is recording a live video "talking head" view using a webcam and microphone source as input, which is composed into the final video output.

Figure 6.3 shows the original interface of the Keshif system, which clearly requires a significant amount of screen space to use properly. Since our DataTV prototype supports real-time controls for zooming and scaling, the presenter can easily adapt the size of the browser in the composite video output, even zooming in to show specific features (Figure 6.5).

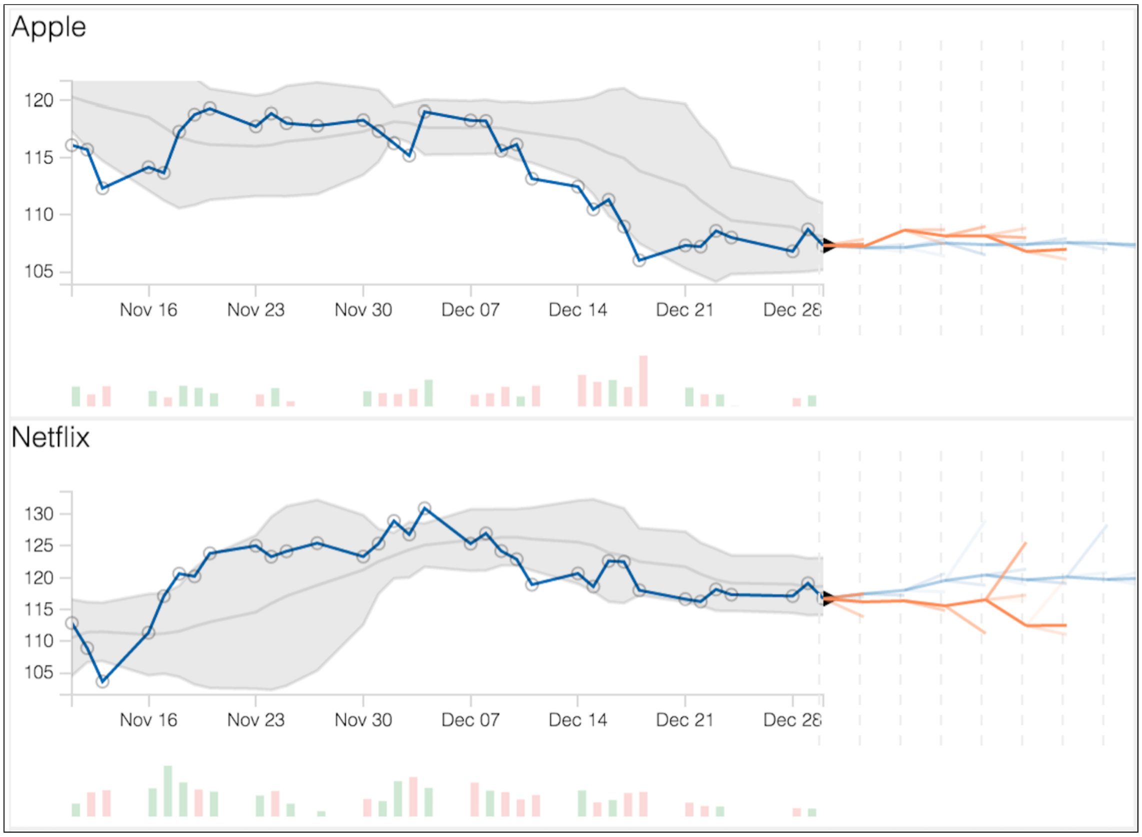


Figure 6.6: TimeFork [207] prediction space for stock market data.

6.3.2 Stock Market Data Analysis with TimeFork

For this example, we use TimeFork [207], an interactive visual prediction technique to support users exploring the future of time-series data. The TimeFork implementation allows an analyst to explore a multitude of potential futures for specific stocks by initiating a dialogue between the analyst and the user. Here we use TimeFork to create a narrative for tech market stocks (in this case Apple and Netflix).

Our user starts off the session with an introductory scene involving a YouTube video featuring Warren Buffett talking about the current stock market. Then the user switches to a scene incorporating the TimeFork tool, using it to predict the current trend of hot tech stocks Apple and Netflix by simply interacting with the tool in a web browser. The user can even switch to a window of a desktop visualization tool such as Tableau or Spotfire and include that into the video if desired. Meanwhile, the user is narrating the interaction and the findings using a live recording of himself using the computer's webcam and its built-in microphone. All of these media sources are composed, recorded, and streamed in the DataTV tool in real-time using the active scene specification. The main narrative of the data video would describe a scenario for stock market trends, similar to what you may hear on financial news. The user closes the video with a PowerPoint information slide that summarizes the main trends.

6.3.3 NY Times Comment Data Analysis with CommentIQ

The CommentIQ [208] visual analytics system is designed to help online community moderators manage large amount of comments associated with online articles by automatically ranking them based on criteria such as relevance, readability, personal experience, and length.

In this example, the user wants to author a streaming data video about the community response on an article from The New York Times[2] titled "City Reacts: State of Emergency" during the 2015 racial unrest in Ferguson, Missouri following the death of Michael Brown at the hands of a white policeman. The user collects

[2] http://www.nytimes.com/

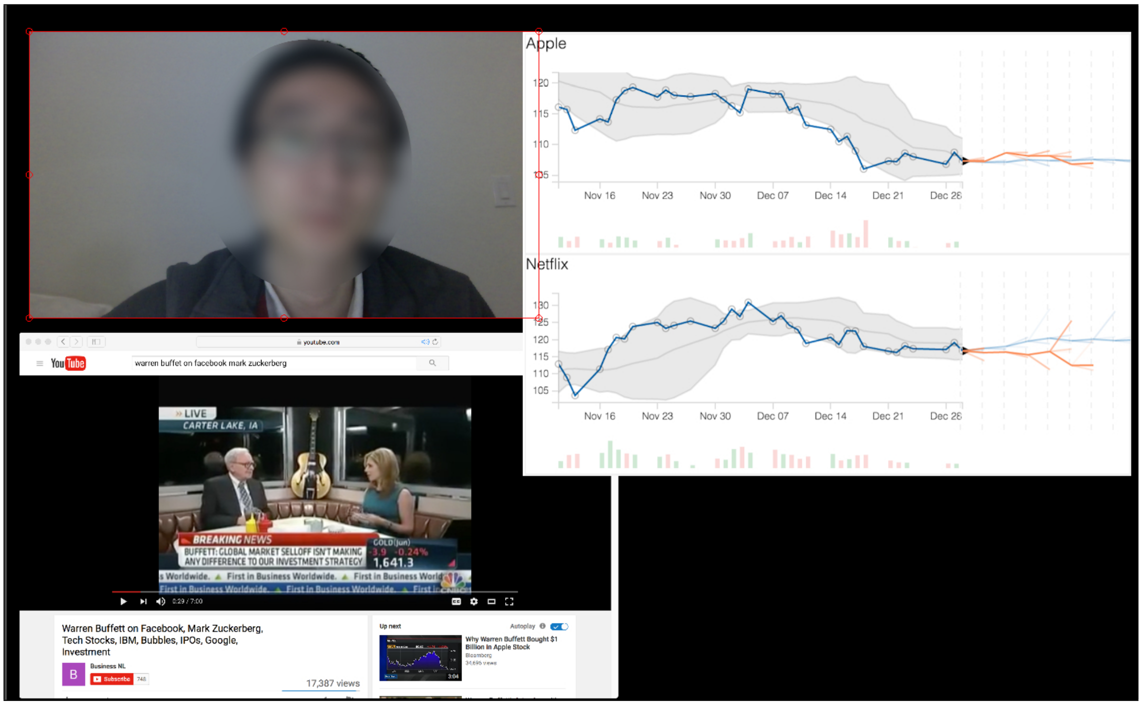
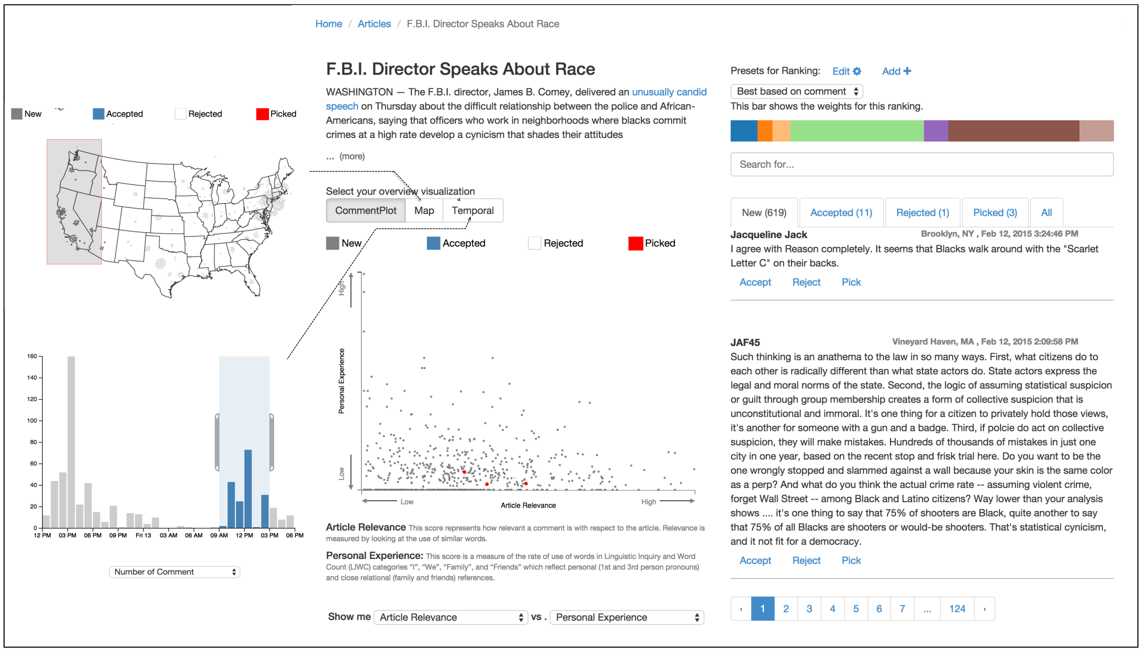

Apple
120
115
110
105
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Dec 28
Netflix
130
125
120
115
110
105
Nov 16
Nov 23
Nov 30
Dec 07
Dec 14
Dec 21
Dec 28
youtube.com
YouTube
warren buffet on facebook mark zuckerberg
LIVE
CARTER LAKE, IA
BREAKING NEWS
BUFFETT: GLOBAL MARKET SELLOFF ISN'T MAKING ANY DIFFERENCE TO OUR INVESTMENT STRATEGY
GOLD (jun)
1,641.3
First in Business Worldwide
Warren Buffett on Facebook, Mark Zuckerberg, Tech Stocks, IBM, Bubbles, IPOs, Google, Investment
Business NL
Subscribe
17,387 views
Up next
Autoplay
Why Warren Buffett Bought $1 Billion in Apple Stock

Figure 6.7: DataTV recording session for a stock market prediction story involving a YouTube clip of Warren Buffet talking about the current state of the stock market as well as the TimeFork web-based visualization tool [207] for predicting the stock price of Apple and Netflix.

two infographics with topics on murder rate across races, and the SWAT deployment rate of different races. The online interactive visualization CommentIQ looks deeper into the comments of the article from The New York Times.

This data video leverages the advantages of each of the three types of media source: video, infographics, and visualization. First, the narrator uses a YouTube video of the Ferguson incident as the introduction. This video shows the confrontation between the protesters and the police, providing a suitable framing to the video that emphasizes the direneness of the situation. The two infographics (Figure 6.9) give background information by showing an overview of the guns and crimes in the area. Finally, the CommentIQ visualization allows the narrator to discover trends in how



Home / Articles / F.B.I. Director Speaks About Race

F.B.I. Director Speaks About Race

WASHINGTON — The F.B.I. director, James B. Comey, delivered an unusually candid speech on Thursday about the difficult relationship between the police and African-Americans, saying that officers who work in neighborhoods where blacks commit crimes at a high rate develop a cynicism that shades their attitudes

... (more)

Select your overview visualization

CommentPlot | Map | Temporal

New Accepted Rejected Picked

Article Relevance This score represents how relevant a comment is with respect to the article. Relevance is measured by looking at the use of similar words.

Personal Experience: This score is a measure of the rate of use of words in Linguistic Inquiry and Word Count (LIWC) categories “I”, “We”, “Family”, and “Friends” which reflect personal (1st and 3rd person pronouns) and close relational (family and friends) references.

Show me Article Relevance vs . Personal Experience

Presets for Ranking: Edit ⚙ Add +

Best based on comment

This bar shows the weights for this ranking.

Search for...

New (619) | Accepted (11) | Rejected (1) | Picked (3) | All

Jacqueline Jack Brooklyn, NY , Feb 12, 2015 3:24:46 PM

I agree with Reason completely. It seems that Blacks walk around with the "Scarlet Letter C" on their backs.

Accept Reject Pick

JAF45 Vineyard Haven, MA , Feb 12, 2015 2:09:58 PM

Such thinking is an anathema to the law in so many ways. First, what citizens do to each other is radically different than what state actors do. State actors express the legal and moral norms of the state. Second, the logic of assuming statistical suspicion or guilt through group membership creates a form of collective suspicion that is unconstitutional and immoral. It's one thing for a citizen to privately hold those views, it's another for someone with a gun and a badge. Third, if polcie do act on collective suspicion, they will make mistakes. Hundreds of thousands of mistakes in just one city in one year, based on the recent stop and frisk trial here. Do you want to be the one wrongly stopped and slammed against a wall because your skin is the same color as a perp? And what do you think the actual crime rate -- assuming violent crime, forget Wall Street -- among Black and Latino citizens? Way lower than your analysis shows .... it's one thing to say that 75% of shooters are Black, quite another to say that 75% of all Blacks are shooters or would-be shooters. That's statistical cynicism, and it not fit for a democracy.

Accept Reject Pick

‹ 1 2 3 4 5 6 7 ... 124 ›

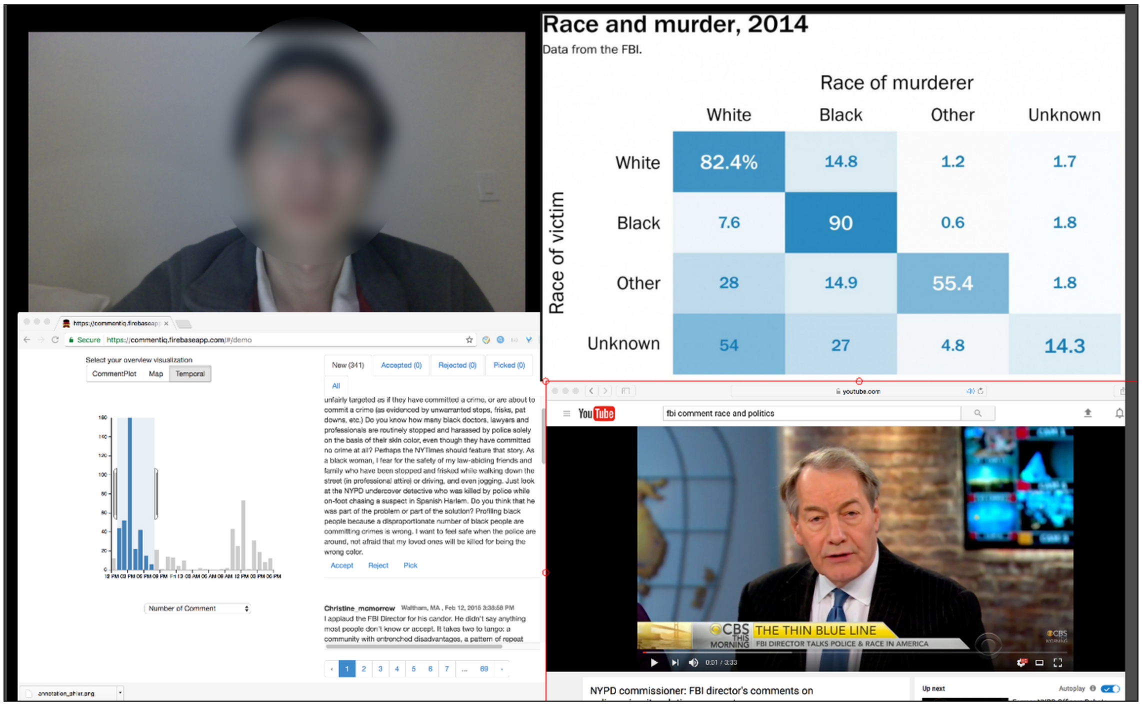
Figure 6.8: Interface of CommentIQ system supporting multidimensional analysis for online article comments.

the NYT commenter community responded to the article. In all of these cases, the narrator is able to scribble directly on the composited video output to highlight interesting or important aspects of the video, such as outliers or trends.

The live streaming functionality of the DataTV platform opens up an entirely new potential for the New York Times to provide a live complement to go with their online comment system. Using the DataTV live stream, community moderators could aggregate and discuss comments in real time, for example when polling voter panels for political debates.

6.4 Qualitative Evaluation

In order to understand how experts and practitioners from the field of information science and HCI would use DataTV, and to identify potential advantages



https://commentiq.firebaseapp
Secure https://commentiq.firebaseapp.com/#/demo
Select your overview visualization
CommentPlot
Map
Temporal
160
140
120
100
80
60
40
20
0
12 PM 03 PM 06 PM 09 PM Fri 13 03 AM 06 AM 09 AM 12 PM 03 PM 06 PM
Number of Comment
annotation_phixr.png
New (341)
Accepted (0)
Rejected (0)
Picked (0)
All
Accept
Reject
Pick
1 2 3 4 5 6 7 … 69
youtube.com
YouTube
fbi comment race and politics
CBS THIS MORNING
THE THIN BLUE LINE
FBI DIRECTOR TALKS POLICE & RACE IN AMERICA
0:01 / 3:33
Up next
Autoplay

unfairly targeted as if they have committed a crime, or are about to commit a crime (as evidenced by unwarranted stops, frisks, pat downs, etc.) Do you know how many black doctors, lawyers and professionals are routinely stopped and harassed by police solely on the basis of their skin color, even though they have committed no crime at all? Perhaps the NYTimes should feature that story. As a black woman, I fear for the safety of my law-abiding friends and family who have been stopped and frisked while walking down the street (in professional attire) or driving, and even jogging. Just look at the NYPD undercover detective who was killed by police while on-foot chasing a suspect in Spanish Harlem. Do you think that he was part of the problem or part of the solution? Profiling black people because a disproportionate number of black people are committing crimes is wrong. I want to feel safe when the police are around, not afraid that my loved ones will be killed for being the wrong color.

Christine_mcmorrow Waltham, MA , Feb 12, 2015 3:38:58 PM
I applaud the FBI Director for his candor. He didn't say anything most people don't know or accept. It takes two to tango: a community with entrenched disadvantages, a pattern of repeat

Race and murder, 2014

Data from the FBI.

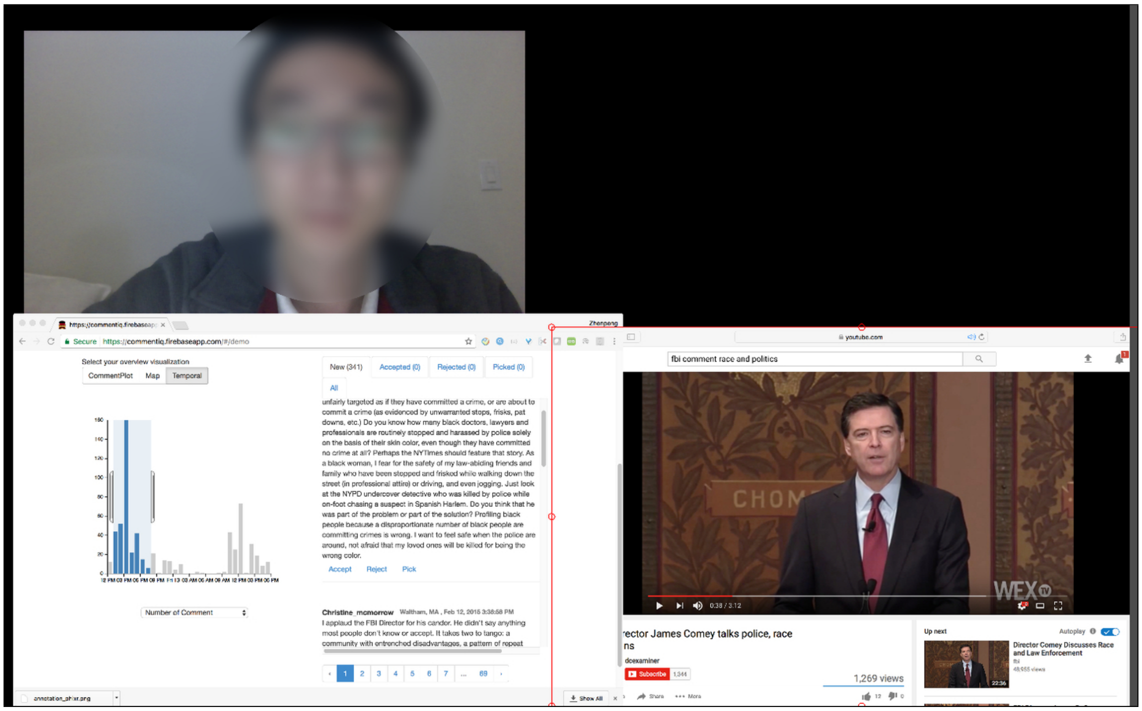
Race of victim	Race of murderer: White	Black	Other	Unknown
White	82.4%	14.8	1.2	1.7
Black	7.6	90	0.6	1.8
Other	28	14.9	55.4	1.8
Unknown	54	27	4.8	14.3

NYPD commissioner: FBI director's comments on

Figure 6.9: DataTV being used to record a streaming data video on race, murder rate, and SWAT activity.

and challenges in using DataTV, we conducted a usability study. Because there is no comparable tool for creating live-streamed data videos in real time, we opted to not perform a comparative study, but instead to focus on the affordances and capabilities of DataTV in a qualitative evaluation.

Our intent with the evaluation was to understand how DataTV can be used in different scenarios. We thus picked two separate data representations, and designed a set of tasks for each data representation such that the tasks were of similar complexity across the two scenarios. Each participant was assigned one data representation, with the associated set of tasks. During the process we use the basic expert review method proposed by Tory and Möller [204], which includes experts evaluating a tool using pre-defined heuristics. The purpose of the study was to test



unfairly targeted as if they have committed a crime, or are about to commit a crime (as evidenced by unwarranted stops, frisks, pat downs, etc.) Do you know how many black doctors, lawyers and professionals are routinely stopped and harassed by police solely on the basis of their skin color, even though they have committed no crime at all? Perhaps the NYTimes should feature that story. As a black woman, I fear for the safety of my law-abiding friends and family who have been stopped and frisked while walking down the street (in professional attire) or driving, and even jogging. Just look at the NYPD undercover detective who was killed by police while on-foot chasing a suspect in Spanish Harlem. Do you think that he was part of the problem or part of the solution? Profiling black people because a disproportionate number of black people are committing crimes is wrong. I want to feel safe when the police are around, not afraid that my loved ones will be killed for being the wrong color.

Christine_mcmorrow Waltham, MA , Feb 12, 2015 3:38:58 PM
I applaud the FBI Director for his candor. He didn't say anything most people don't know or accept. It takes two to tango: a community with entrenched disadvantages, a pattern of repeat

Figure 6.10: DataTV recording the CommentIQ system being used to filter comments over time.

the usability under the context of the works of the experts.

6.4.1 Participants

We engaged two volunteer participants—Expert 1 and 2—to evaluate our system. The two participants have extensive experiences for using information visualization to tell data stories. They are both very knowledgeable for the existing information visualization and visual analytics system. They have created and reviewed tools in both fields. The participants were Ph.D. students in the field of HCI or information visualization with at least three years of experience. Both were male.

6.4.2 Apparatus

We conducted the experiment on a standard laptop computer equipped with a 15-inch LCD screen (resolution 1280×800), a standard keyboard, and a three-button mouse. The built-in camera was used for recording the user's speech with voice.

6.4.3 Tasks

Each participant's task was to create data videos using DataTV and the data visualization randomly assigned to them. Participants were required to use at least one interactive visualization in their video. They were allowed to pick any other appropriate media sources on the web to support their stories. The tasks were devised so that the experts would need to explore the visualizations, select media sources, and sift through information to create a narrative. The data stories we asked them to create were inspired by our examples from Section 6.3:

- ***Obama Budget:*** A visualization illustrating the components of government budget in the year 2013.
- ***U.S. Census:*** Demographics for the state of Florida.

A set of tasks was prepared for each scenario, requiring the participants to explore the data visualizations in detail and to look for supporting information online, before finally creating a data video to answer the question. A sample set of tasks for one of the scenarios (Obama budget) is given below:

- Explain roughly how the total ($3.7 trillion) were allocated (using a 30-second video);
- Explain the main types of spending (using a 30-second video);
- Explain how the spending has changed since the last budget (using a 40-second video); and
- Describe how much of the budget was allocated to social security (using a 60-second video).

6.4.4 Procedure

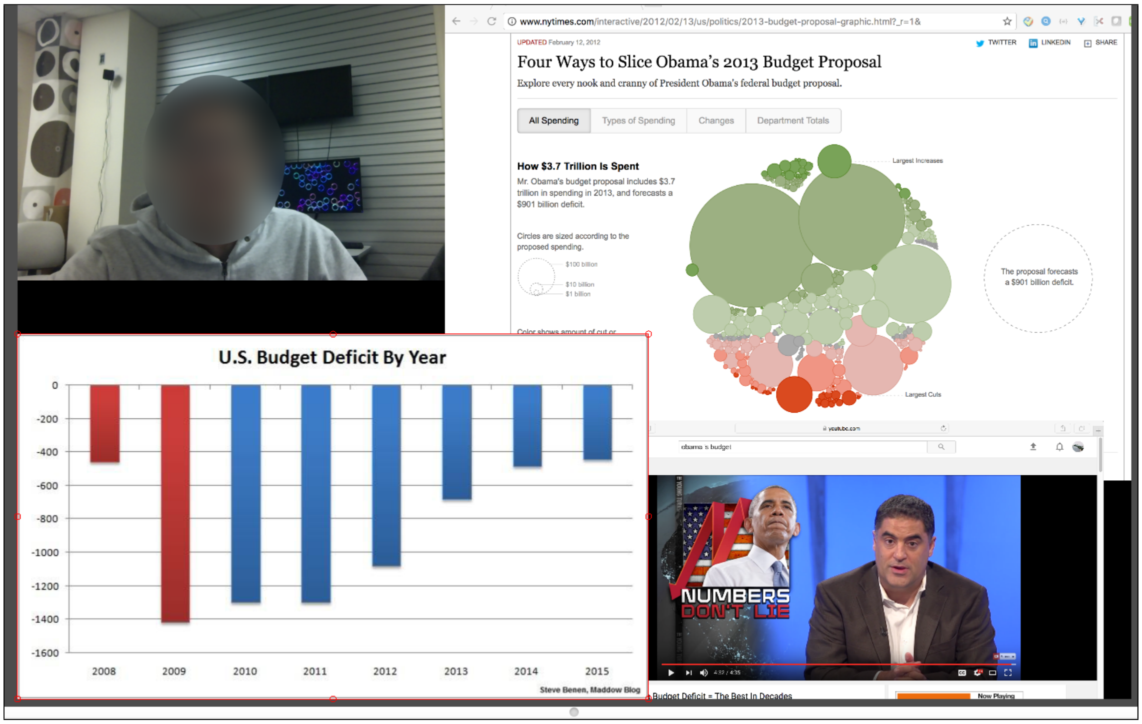
Participants were shown a demonstration of DataTV and given as much time as they needed to explore and familiarize themselves with the system. We then gave them a sheet of tasks and asked them to create a data video in response to each task. The resulting data video would ideally make use of multiple media types, interactive visualization, and innovative storytelling techniques. During the creation process, which was capped at 60 minutes, the experts were allowed to ask questions about the interface. We followed a think-aloud protocol with the participants, and recorded their behavior via video and written observations.

For the purpose of this study, the participant-generated stories were recorded and saved (instead of streamed dynamically). The final participant-generated videos were limited to a duration of one minute. During the process, the participants used the DataTV tool to design, sketch, record, and produce their videos. After completing their tasks, we followed up with an interview where participants explained their process and provided feedback on the system.

6.4.5 Results

We collected and analyzed both the products as well as the observations and interview feedback from each expert review session.

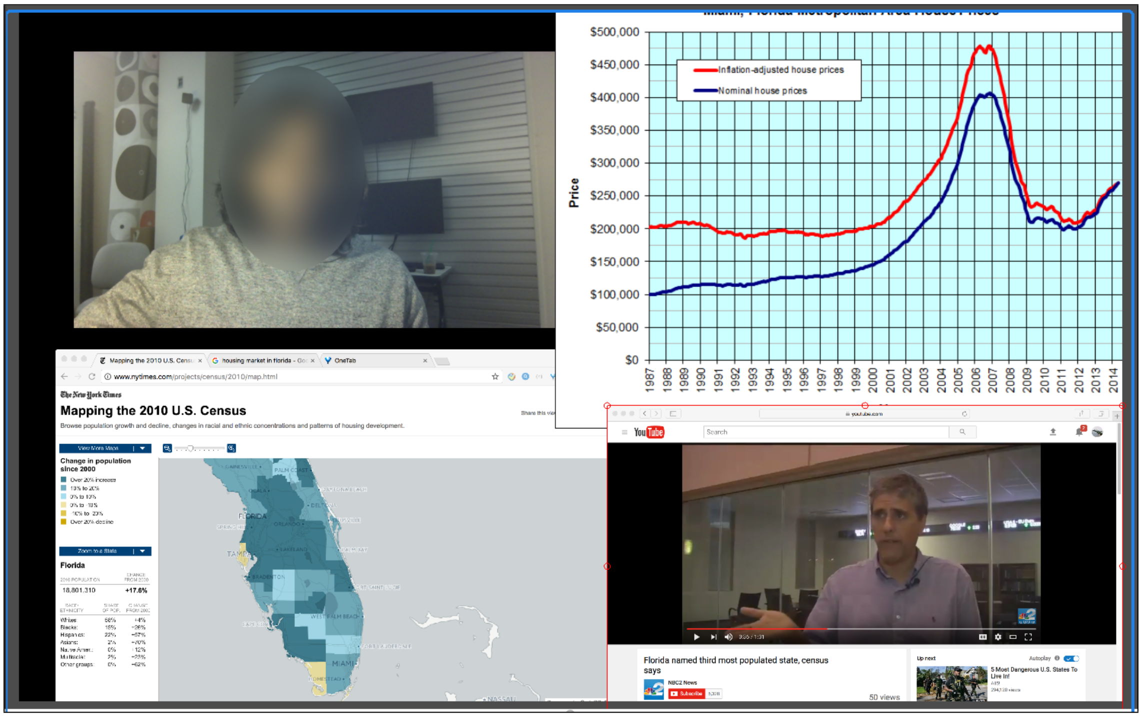
Products: Each expert made several data videos, all less than one minute long. Representative data videos are attached as supplemental materials. We also captured screenshots of the ending exact workspace for each expert (Figure 6.11, Figure 6.12). All in all, the resulting data videos were all of good quality and suggest that the DataTV platform was instrumental in the process.



www.nytimes.com/interactive/2012/02/13/us/politics/2013-budget-proposal-graphic.html?_r=1&
UPDATED February 12, 2012
TWITTER
LINKEDIN
SHARE
Four Ways to Slice Obama's 2013 Budget Proposal
Explore every nook and cranny of President Obama's federal budget proposal.
All Spending
Types of Spending
Changes
Department Totals
How $3.7 Trillion Is Spent
Mr. Obama's budget proposal includes $3.7 trillion in spending in 2013, and forecasts a $901 billion deficit.
Circles are sized according to the proposed spending.
$100 billion
$10 billion
$1 billion
Color shows amount of cut or
Largest Increases
The proposal forecasts a $901 billion deficit.
Largest Cuts
U.S. Budget Deficit By Year
0
-200
-400
-600
-800
-1000
-1200
-1400
-1600
2008
2009
2010
2011
2012
2013
2014
2015
Steve Benen, Maddow Blog
youtube.com
obama s budget
NUMBERS DON'T LIE
Budget Deficit = The Best In Decades
Now Playing

Figure 6.11: Expert 1 reviewed the U.S. budget trend in 2013 during the time of President Obama. The expert presented their insights using DataTV.

Observations: Overall, participants used the tool with little training. Both



The New York Times

Mapping the 2010 U.S. Census

Browse population growth and decline, changes in racial and ethnic concentrations and patterns of housing development.

Change in population since 2000

- Over 20% increase
- 10% to 20%
- 0% to 10%
- 0% to -10%
- -10% to -20%
- Over 20% decline

Florida

2010 POPULATION	CHANGE FROM 2000
18,801,310	+17.6%

RACE/ ETHNICITY	SHARE OF POP.	CHANGE FROM 2000
Whites:	58%	+4%
Blacks:	15%	+26%
Hispanics:	22%	+57%
Asians:	2%	+70%
Native Amer.:	0%	+12%
Multiracial:	2%	+25%
Other groups:	0%	+62%

Florida named third most populated state, census says

NBC2 News

50 views

5 Most Dangerous U.S. States To Live In!

294,120 views

Figure 6.12: Expert 2 reviewed the population change of Florida. Insights on housing and population gain were presented using DataTV.

experts familiarized themselves with the basic operation by first making a few trial videos that were recorded and saved locally. Once satisfied with the basic mental model, they spent a considerable time (20-25 minutes) finding source materials, selecting visualizations, and writing notes. They then spent an additional 10 minutes to create their scenes, including arranging the layout and size of the different media sources. The actual recording of each video was surprisingly quick; given the preparation, the experts were both able to record their videos in a single take and with few mistakes.

From our observations, it appeared as if the experts did not need much prompting to get familiar with the DataTV interface. The experts seemed to think the interface behaved in a logical and predictable fashion, and they quickly became proficient with very little instruction and within 10 minutes of training. Most importantly, observations and think-aloud remarks seem to indicate that the experts rapidly internalized the DataTV controls and were able to focus on the craft of data-driven storytelling. This was indicated by their utterances increasingly dealing with how to best organize and present insights rather than minutae of the interface.

Interview Feedback: During the structured interview after completing the tasks, the experts were asked to give feedback on the system, including advantages and disadvantages of DataTV over existing approaches. We summarize their feedback below:

- **Positive:** Expert 1 thought that the DataTV interface was simple and straightforward, with few opportunities to make mistakes. Expert 2 remarked that compared to using multiple software platforms, our system makes the video creation process seamless as it requires less operations and vital tools such as annotation, compositing, and transformations are very accessible.

  Expert 2 also thought that the use of our system was surprisingly easy, particularly the streamlined workflow where very little preparation is necessary. The expert noted that the composite video output was helpful so as to always know what is being streamed and recorded, striking a good balance between real-time control and accuracy.

- **Negative:** Both experts remarked on the lack of video editing capabilities. We responded with the fact that such editing functionality would have precluded real-time streaming of the tool, and they both remarked that the compromise was acceptable. The first expert complained that the DataTV interface should provide better control over media sources in the workspace.

Chapter 7: Discussion

7.1 Explaining the Findings

7.1.1 DataComics

Our qualitative evaluation indicated that data comics, especially with captions were significantly more engaging, space-efficient, faster and easier to use than original visualization/infograhic. The feedbacks from the participants also indicate that panels of partitions help focus and memory. The captions are particularly helpful when following the story and remembering the details. The participants mostly felt that the data comic was more effective that it invites reading and helps build up the story. The sequence of the panels in data comics are important especially when helping the participants recall detailed information on one of the panels with the help of captions. The overall sequence is more important than sequence of a small range, i.e. two panels about the topics parallel to each other can be changed without harming the whole storyline.

A future evaluation with more participants regarding the change of eye focus during the experiment from the participants will definitely help. However, we don t have the equipment or time to conduct the experiment in such short time.

7.1.2 DataTV

Our informal evaluation indicated that DataTV facilitated live data-driven storytelling. It should be noted that the main contribution of the chapter 6 is the method of live-streaming data videos, whereas our implementation is merely a prototype to show the validity of the concept. We have feedback from both experts that an easy to use, well-built, and integrated system is facilitating the storytellers to create data video stories.

It is important to note that all of the functionality of the DataTV tool can be replicated in a combination of desktop recording tools—such as VLC—and video editing tools—such as Adobe Premiere Pro—with sufficient time and effort. While DataTV makes constructing data videos easy with its integration of media source picking functionality, media label editing, and video recording, each of the DataTV videos showcased in this chapter 6 can be built using other tools. However, the argument for the DataTV platform and related software is two-fold: (1) a single unified platform is needed to allow for live streaming and rapid production, and (2) the integration of all of these data-driven storytelling features in a single tool enables the analyst to think about data videos more in terms of storytelling rather than low-level software, mechanics, and features. Results from our qualitative evaluation support these two arguments.

We believe our work surfaces several new issues that were not considered in the past. For example, while Amini et al. already suggested the data videos concept in 2015 [153], their work still results in a static and prerecorded video, not a live-streamed one. One of the benefits may be that it is easier to quickly produce a data video using DataTV than painstakingly using a suite of tools such as screen recorders, video editors, and audio production tools. However, DataClips [154], presented in 2017, does provide functionality for quickly assembling several clips using predefined visualizations. On the other hand, a live data video can be responsive to an audience, for example in responding to questions or requests for more information. In this way, DataTV is much more of an interactive presentation tool than typical data video production tools (such as DataClips). This is reinforced by the emphasis on live video in DataTV, whereas the narrator is typically disembodied in most existing data videos. We think this suggests that live data videos as those supported in our work is a unique data-driven storytelling medium in its own right. It is also the reason that we found no easy baseline for a comparative evaluation. We leave comparisons to live presentation software, such as Microsoft PowerPoint, to future work.

7.2 Generalizing the Findings

How general are these findings? We discuss this below.

7.2.1 DataComics

We explicitly chose not to measure time or correctness. There is likely little difference between data comics versus single visualization/infographic, and this perception was also confirmed by participants in our experiment. Rather, the strength of data comics comes from its approachable, compelling, and intuitive format. This is further validated by Lee et al. [171], who only collected subjective ratings from participants in their SketchStory evaluation.

7.2.2 DataTV

The utility of live-streaming data videos as a concept can be questioned. It is certainly true that we do not foresee “Let’s Analyze” videos to dethrone the “Let’s Play” category on Twitch or YouTube anytime soon. However, the power of the internet as both a medium as well as an audience should not be underestimated. There is already a small but growing group of Twitch communities devoted to non-gaming, such as painting, gardening, and programming. The step is not too far from such topics to data analysis. Besides, even if live data videos never become popular, many of the real-time authoring techniques pioneered in DataTV will be invaluable for creating normal, non-streaming data videos, going beyond what even tools such as DataClips [154] can do.

7.3 Limitations

7.3.1 DataComics

First of all, much of our argumentation of using sequential art for data is based on two assumptions: that the audience has (a) prior experience, and (b) a favorable opinion about comics. With no prior experience, much of the benefit of an established common ground in the visual language of comics is lost. Furthermore, given the sometimes questionable respectability of comics [149, 175], its use as a communication medium may be problematic. For example, it can be argued that a data comic may not be the best vehicle for presentations in very formal settings, such as a boardroom meeting. Similarly, the intrinsically light-hearted nature of comics may be inappropriate for sensitive or difficult topics, such as natural disasters, emergency situations, and other types of crises or stories on the loss of lives or livelihoods.

7.3.2 DataTV

First of all, much of our argumentation of using data video for storytelling is based on two assumptions: that the audience has (a) enough knowledge for understanding the data video, and (b) a favorable opinion about video storytelling. Without enough knowledge, much of the benefit of an established common ground in the visual language of data video is lost. Furthermore, given the sometimes the higher requirement of environment—playing video might be inappropriate in some communication situations—data video might not be the perfect choice for information distribution under situation that noise level is sensitive or displaying device is not well equipped. For example, it can be argued that a DataTV may not be the best vehicle for presentations in very noisy settings, such as a couple people discussing a topic in a train station, where static material might be more suitable. It should also be noted that the content of DataTV, often including personal webcam video, can be inappropriate for public broadcasting.

7.4 Guidelines

After exploration of the taxonomy and examples of data-driven storytelling media. We suggest the following guidelines for others to conduct examination and exploration for data-driven storytelling media.

7.4.1 Be Open to Unique Media for Storytelling

When facing new type of media, one should not be so restrict that only certain types of media are suitable for data-driven storytelling. Dancing is not often considered as a viable way of data-driven storytelling. However, with proper labelling and moves, dancing can tell a story about how to conduct a bubble sort very intuitively [190]. Augmented reality (AR) and virtual reality (VR) have gained much attention recently. We found examples [72, 78, 80, 86]that use AR and VR extensively for data analysis and exploration. The applications were not invented to conduct data-driven storytelling, but these applications show that VR and AR, as newly introduced media, make great example of how new technologies can be integrated in the scope of data-driven storytelling.

7.4.2 Avoid Relying on Artistic Skill

Not everyone is an artist. Many data-driven storytelling media require artistic skill. For instance, documentaries [12, 13, 15] with data stories require editing and video shooting skills. DataComics is another example that requires a certain level of artistic skills. Conducting data-driven storytelling with comics [21,151,156] requires that the creator choose comic figure and other comic features accordingly, so that the generated datacomic is enjoyable and easy to follow.

To make the creating process less dependent on the creator's artistic stills, the applications for new media should be loaded with more automatic features, such as template recommendation [143, 196], computer vision for caption recommendation, and natural language processing for caption generation. The applications should have the mostly used functions well integrated in the interface to create an environment convenient for the storytellers so that the storytellers can focus on the content instead of the user interface.

7.4.3 Start from Existing Examples, Don't Be Too Unique

When creating new data-driven storytelling media types, it can be started from existing examples. It is unnecessary to think about new media types totally unique. One can use the taxonomy to determine each value of the dimensions, and then change values of certain dimensions.

Some examples for data-driven storytelling with augmented reality and virtual reality show that the latest technology are all practical on existing applications such as design [72], construction [76], and data analysis [81, 86].

Chapter 8: Conclusion

In this dissertation, we studied data-driven storytelling media for casual users as the consumers of information by expanding its horizon, and exploring how it aids casual users in viewing, analyzing, and understanding data. In order to achieve this, we present a new taxonomy focused on media types for data-driven storytelling with the purpose of opening the field to a wider set of future possibilities. Our work started with collecting a large amount of evidence of data-driven storytelling using novel and diverse media, from the spoken word to interpretative dance and choreography. From the taxonomy and guideline derived, we investigate media types particularly useful for casual users with little professional training or background in data visualization and analysis.

With our taxonomy and the guidelines derived, we proposed two examples: DataComics, leveraging comics (sequential art [156]) for data-driven storytelling [1], and DataTV, live-streaming data videos for this purpose.

Through collecting and studying the examples as well as the two systems we proposed, we found several common phenomena:

- Many of the new media types that we studied are not well investigated under the scope of data-driven storytelling. For example, there are still very few dedicated tools for story authoring with Augmented Reality or Virtual Reality for data visualization.

- Most of the media types for data-driven storytelling we examined can already be leveraged using existing software systems. For example, a data comic could be created using only Microsoft PowerPoint, or a data video using Adobe Premiere.

- However, it is very important to have an integrated system for authoring and presenting data-driven stories. Most of existing authoring tools had alternatives before they were invented. For example, Adobe Premiere, with its advanced artificial intelligence video editing functions, can be replaced by multiple simpler software systems, such as basic video editing tools, and more human effort. The availability of dedicated or automatic tools allow the user to focus on the subject matter rather than the technical or logistical aspects of the process.

- Evaluation of data-driven storytelling system should not only study the presentation of data-driven stories, but also the authoring experience. Although it is easier to assess the effect on presentation, having a tool well integrated and easy to use for the authoring process is equally important and is the key to encourage creating new data-driven storytelling.

Chapter 9: Future Work

In the future, our taxonomy can still be further refined with exploring more types of data-driven storytelling media. More specifically, the current list of dimensions and values for each media type can be further justified to make the taxonomy more robust and thorough. Our exploration of data-driven storytelling media has been mostly focused on and following our own taxonomy, but connecting with similar taxonomies–taxonomies of general storytelling– and taxonomies of other fields–taxonomies of visualization techniques–are also potential beneficial. Such connections can be used for expanding and completing the current taxonomy, as well as inspecting our taxonomy and guideline from other directions.

When new types of data-driven storytelling media is invented over time, they are still yet to be categorized using our taxonomy. This thesis does not cover all the possible media types of data-driven storytelling. There are new media types coming out all the time and their combination with data-driven storytelling is pending further investigation. Also, many existing media types are available to use for data-driven storytelling.

For example, imagine creating a data-driven storytelling tool designed to support speech. Such a tool may be supported by the natural language processing technology and automatically generates statistical diagrams to illustrate the ideas extracted from the speech. Clearly, there is ample opportunity for leveraging data-driven storytelling in many other forms than has currently been studied in the visualization community.

The process of creating new data-driven storytelling media for casual users is based on the guidelines we proposed. We have shown the effectiveness of the guidelines through creating DataTV and Data Comics [152]. However, the guideline may not be sufficient for all future situations. It is likely that new guidelines will be added to the current collection.

Appendix A: Survey of Data-Driven Storytelling Media

A.1 Storytelling in Movies and Documentaries

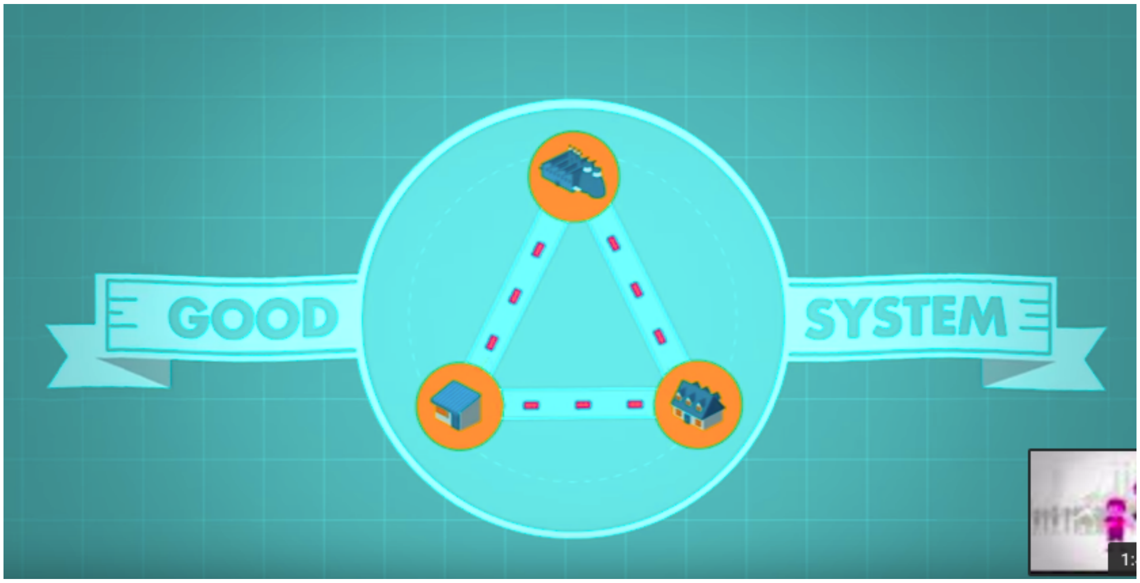


Figure A.1: Marine Plastic Pollution [5].

A.1.1 Stop Marine Plastic Pollution

The marine plastic pollution condition in Figure. A.1 shows the sources and conditions of marine pollution. The map shows the distribution of pollution condition.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many peo-*

ple on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload.* The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence.* However, it is also *replicable*, as it is stored and can be played back at any time.

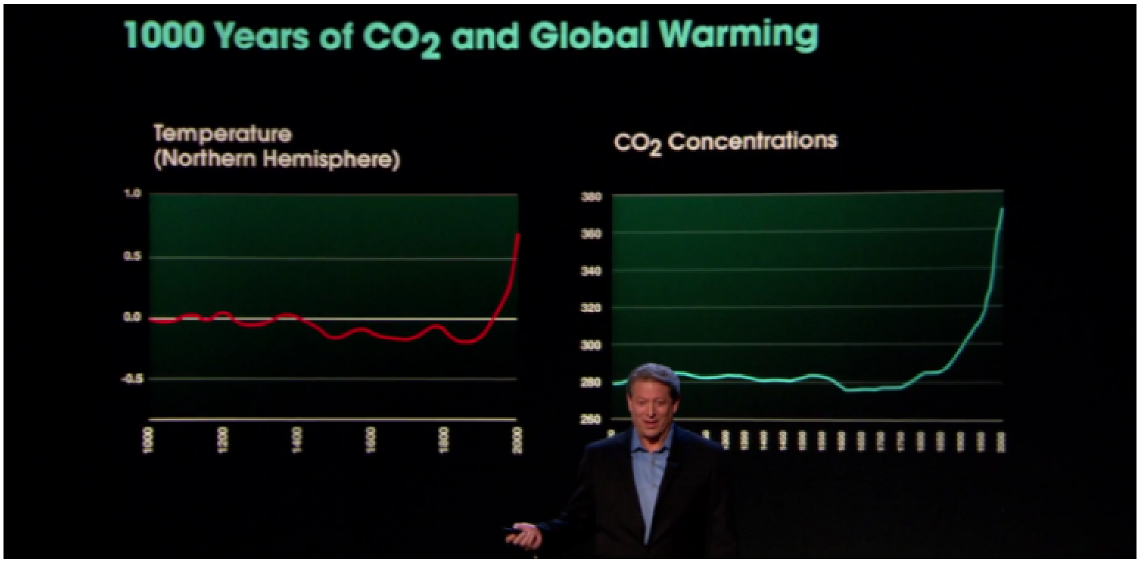


Figure A.2: Scene that shows relations between CO2 and temperature change.

Figure A.3: Ice are melting with temperature increases.

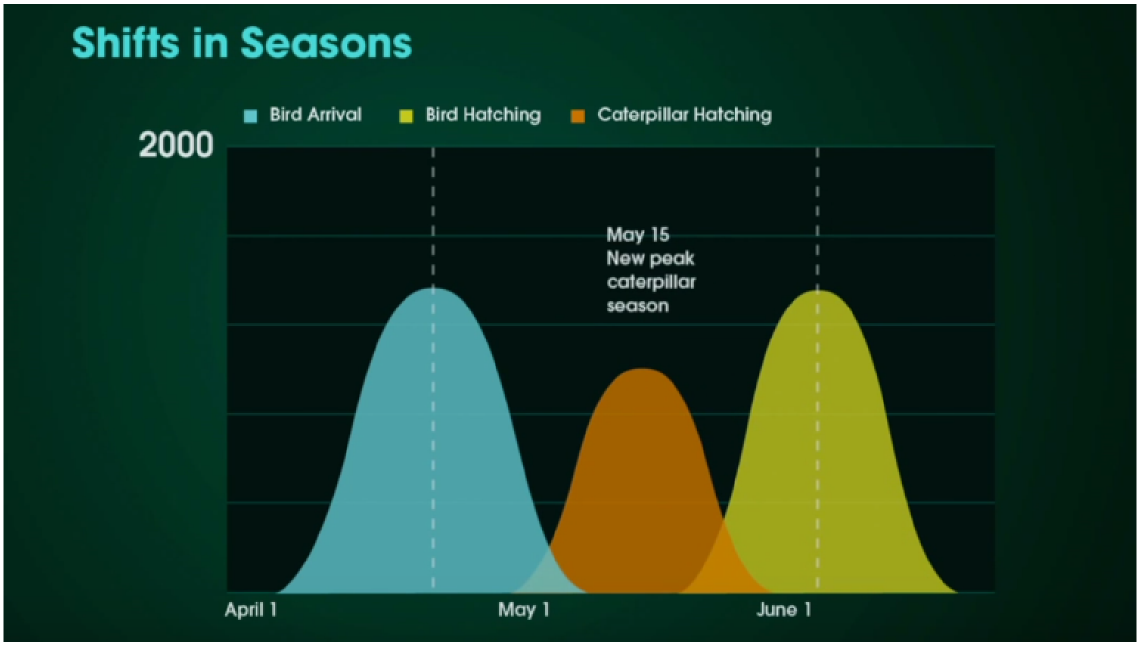


Figure A.4: Scene that seasons are shifting and causing problems for animals

A.1.2 A Beautiful Planet

This is a documentary [193] that telling a story about global warming. The movie consists of scenes of warming signs such as melting glacier, illustration of correlation between temperature and CO2, and animal birth timing changes. The movie is not just a video of one guy giving presentation with data visualization, whereas a large portion of it is footage of natural scenes and interviews. The storytelling is to use data to prove that global warming is an inevitable truth that needs peoples attention immediately. The images below can show that both data (1. the changes of both temperature, CO2 level, and the correlation between them. 2. The birth timing change.) and natural footage are important to compose such a story.

Figure(A.2, A.3, A.4), can show that both data (1. the changes of both temperature, CO2 level, and the correlation between them. 2. The birth timing change.) and natural footages are important to compose such a a story.

Informal classification: As a movie that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

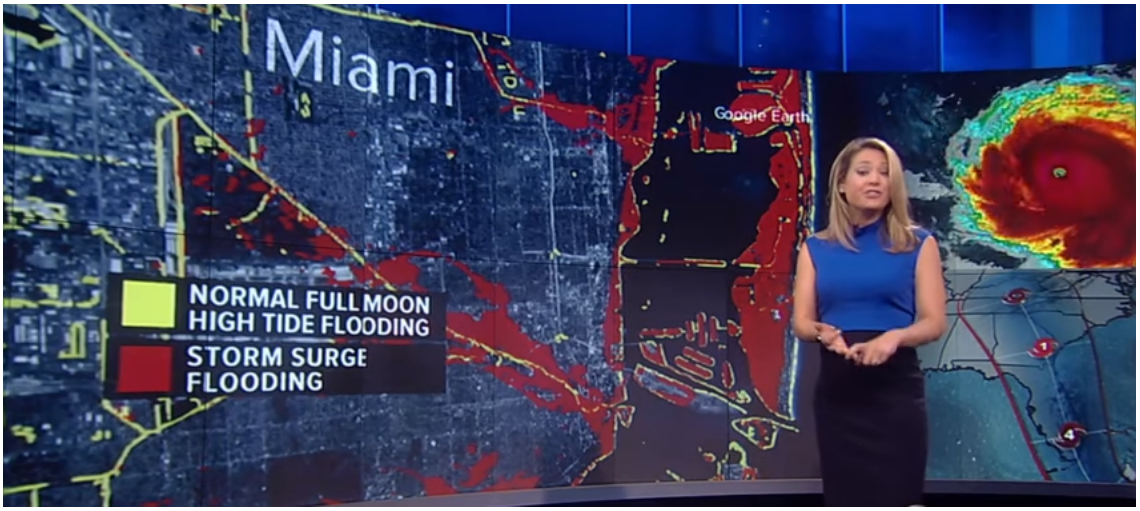


Figure A.5: News for Irma Hurricane [5]. Part of Miami will be flooded.

A.1.3 ABC News for Irma Hurricane

The ABC News in Figure. A.5 shows the forecast of Hurrican Irma and its potential damage. The heatmap illustrate the flooded area and the intensity of the wind.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.4 Is Height All in Our Gene

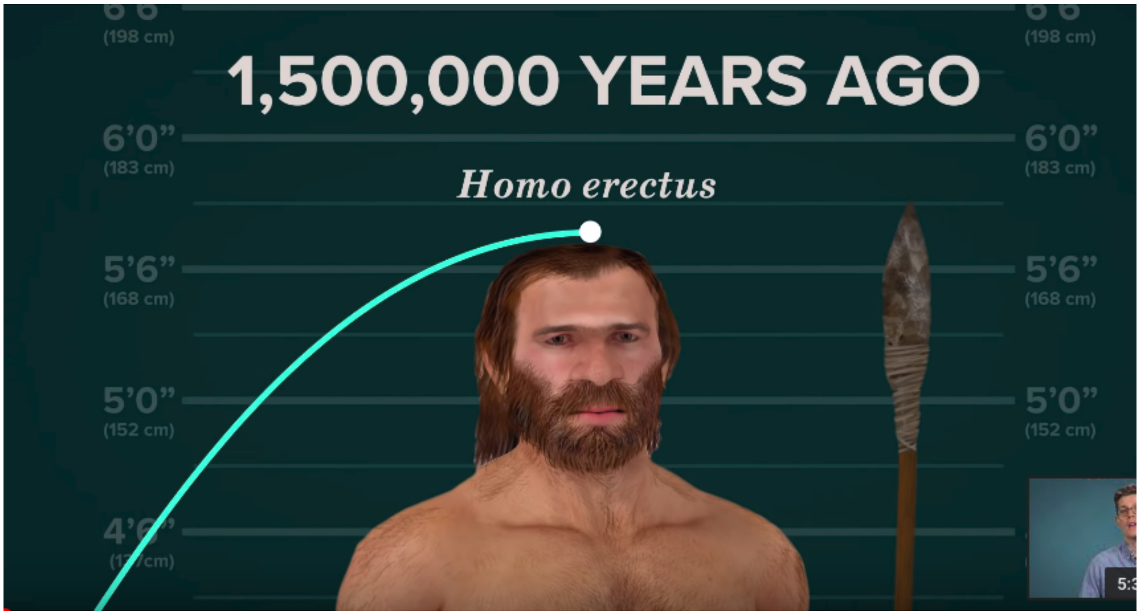


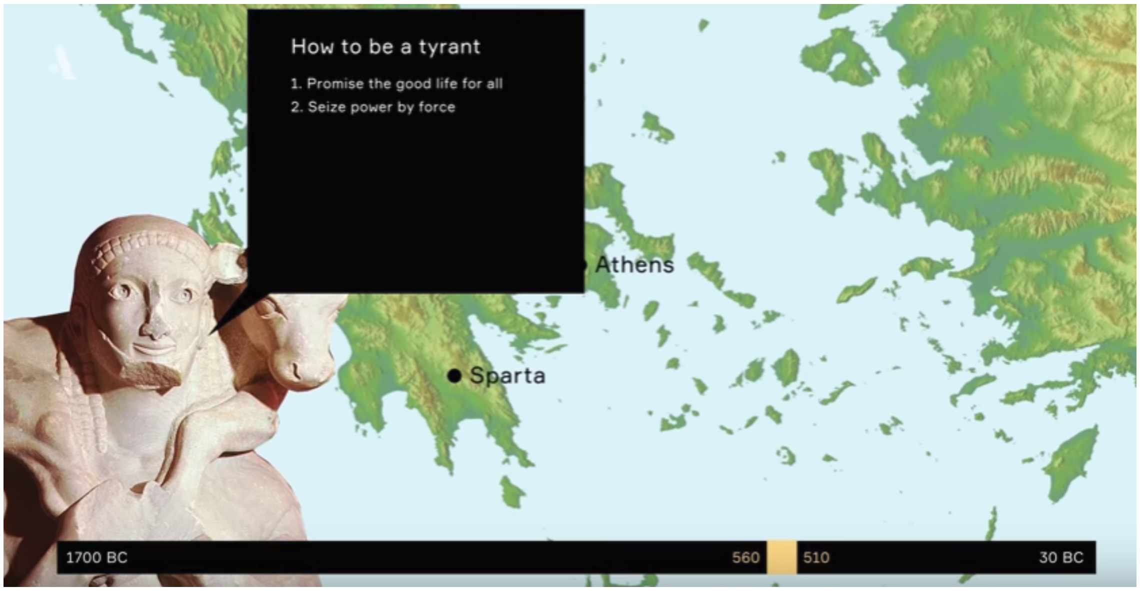
Figure A.6: Height change [6] along with time. The average height of human increases with time.

The data-driven video shows how human height is affected by gene. Line chart is used to demonstrate the change of height with time in Figure. A.6. The story is the change of human height along time and among people within the same period. The data is the height of time and people of a specific group.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, table, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.5 Ancient Greece in 18 minutes


How to be a tyrant

1. Promise the good life for all
2. Seize power by force

Figure A.7: Ancient Greek [7] ruler.

The data-driven video shows how Greece changed its territory during the history. Colored map graph is used to demonstrate different countries in Figure. A.7. The story is the change of Greece narrated by text and time is visualized with bar chart as data.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.6 The history of Asia: every year

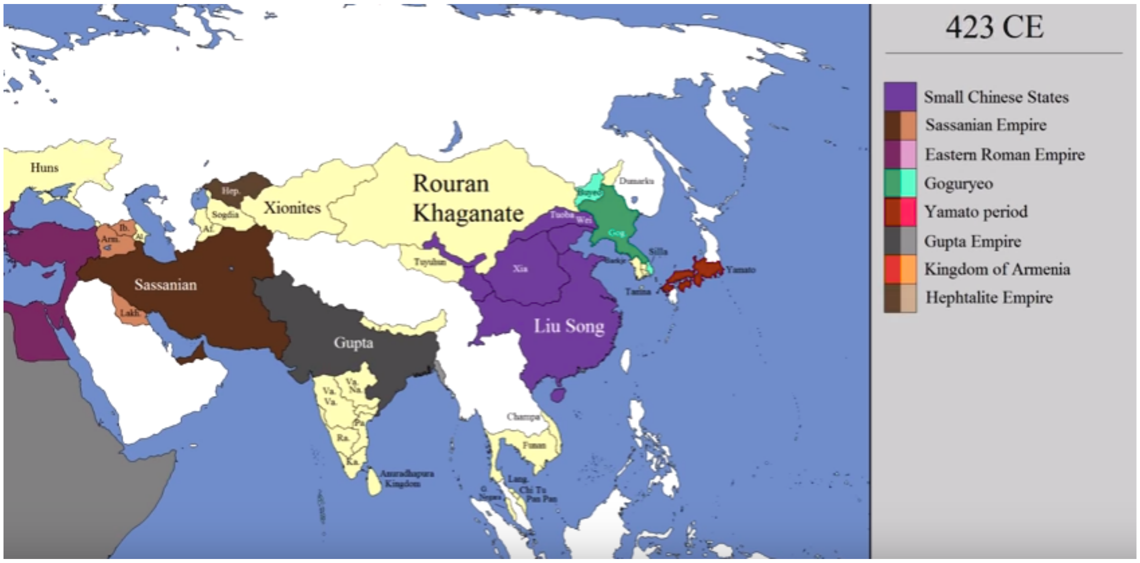


Figure A.8: The change of Asia [8] viewed from a map.

The data-driven video shows how Asian countries changed territories during the history. Colored map graph is used to demonstrate different countries in Figure. A.8.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.7 Wealth Inequality in America

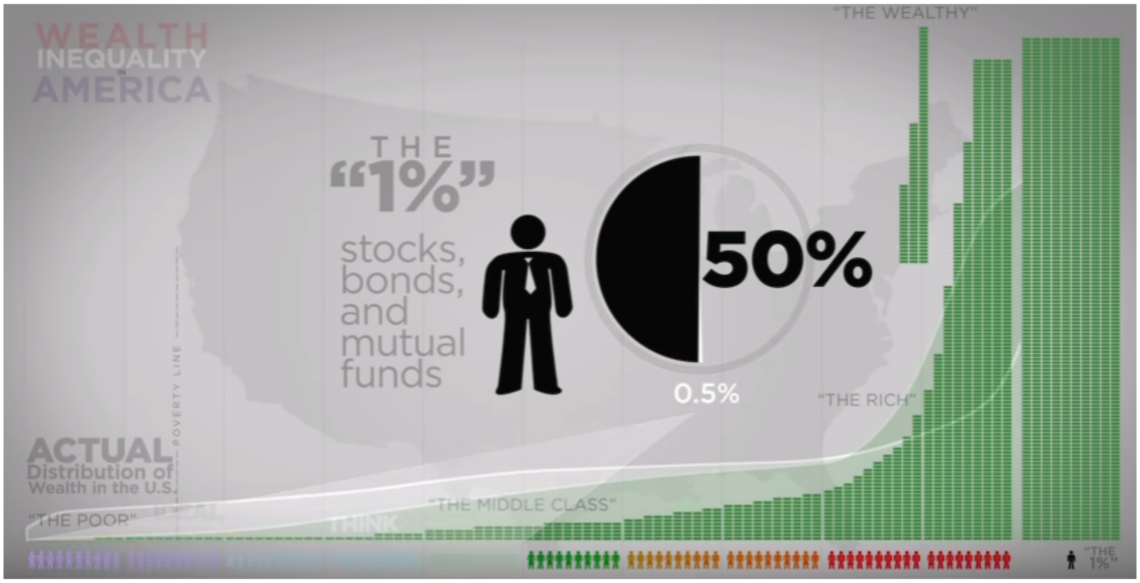


Figure A.9: The documentary [9] shows how one percent of the population occupies a large percent of wealth.

The data-driven video shows how unbalanced the wealth are owned by different group of people in Figure. A.9 The bar chart and annotations show that the half of the stocks, bonds, and mutual funds are owned by one percent of the people. The super rich one percent has more wealth that the bar chart is not able to hold in the current view.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.8 The Joy of Stats

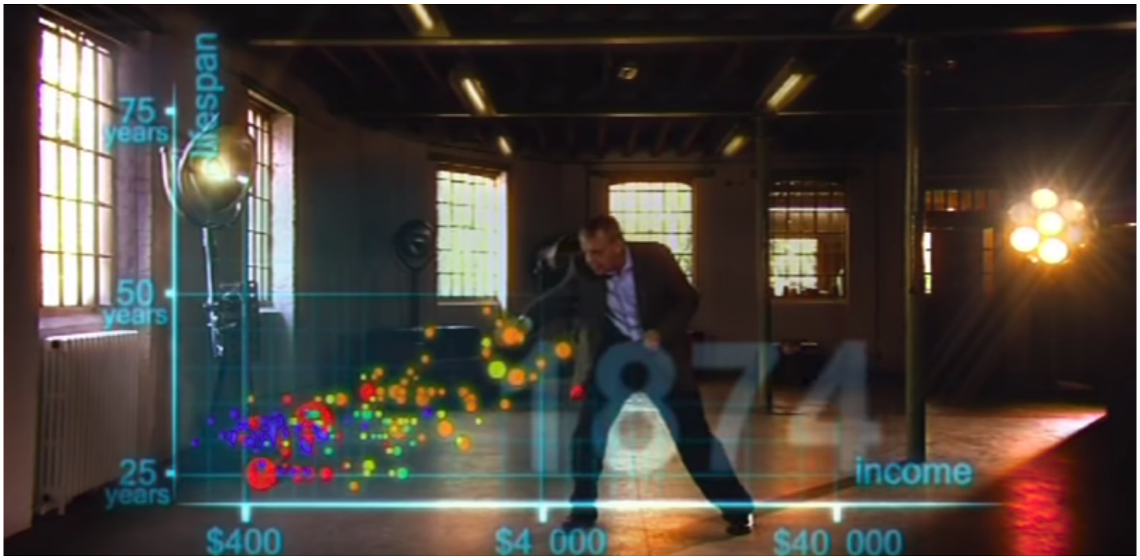


Figure A.10: The documentary [10] shows how data visualization works.

The data-driven video shows statistical visualization can summarize the history over two hundred years in a few minutes in Figure. A.10 Scatter plots show how income and population change during the time.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.9 Religions and babies

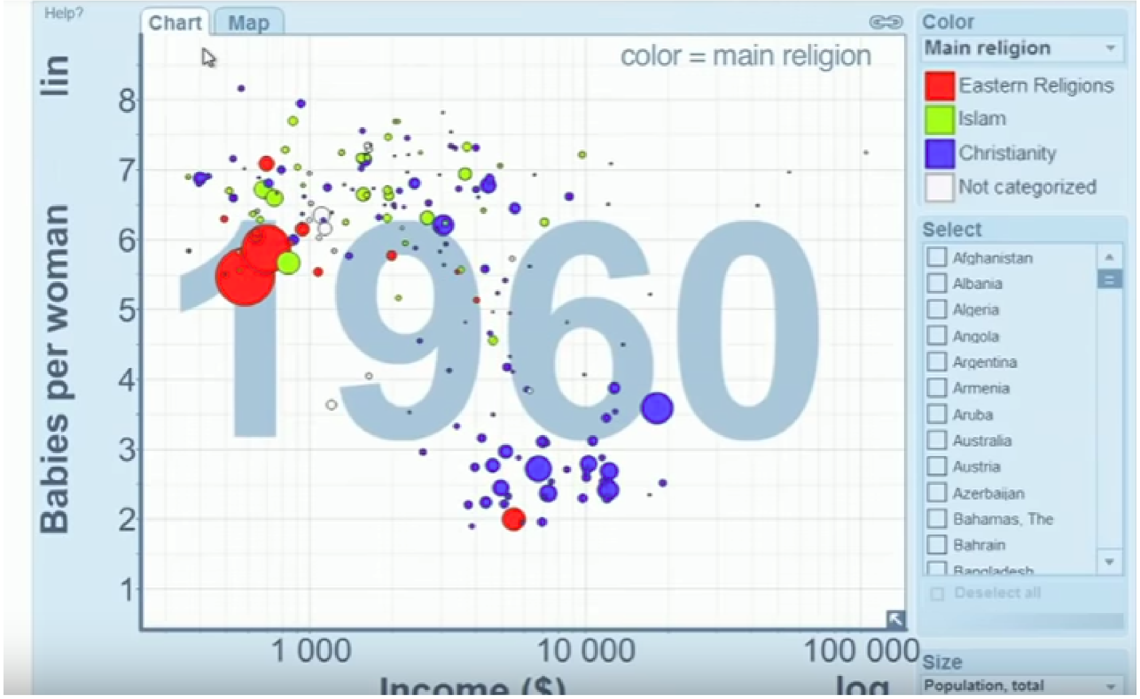


Figure A.11: The documentary [11] shows how numbers of babies varies with religion.

The data-driven video shows relation of baby population and religion with different years in Figure. A.11 The scatter plot shows the distribution of baby population size which is easy to map on the visualization.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.10 Gene Pool Decline

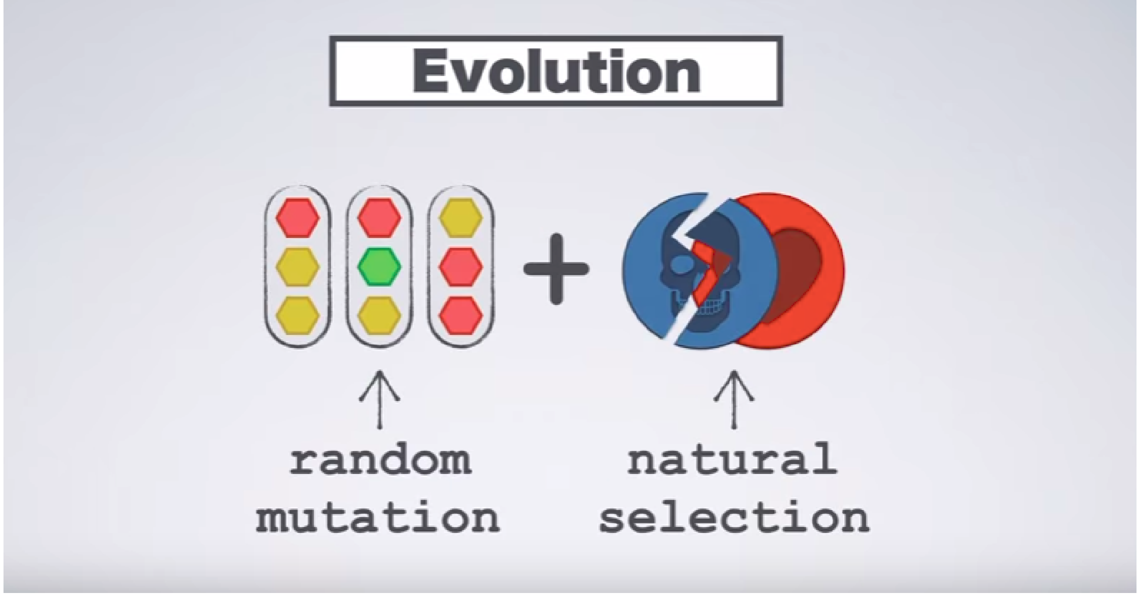


Figure A.12: The documentary [12] shows how gene pool of human declines.

The data-driven video shows how gene pool declines with time as we are more relied on medical treatment. A.12 The graph and statistical visualization shows how human gene pool evolves and regresses with time.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.11 How To End Poverty

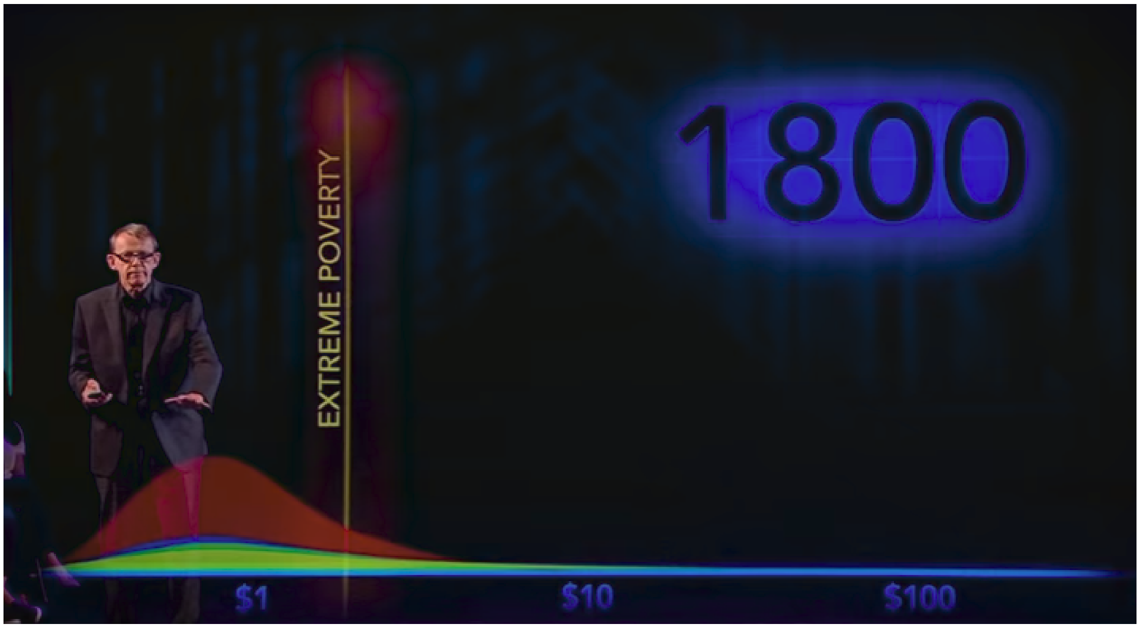


Figure A.13: The documentary [13] shows how to end poverty.

The data-driven video shows how poverty is distributed with locations and time. A.13 The statistical graph and map shows how the poverty decreases along time from ancient time to modern days.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.12 China's Geography Problem

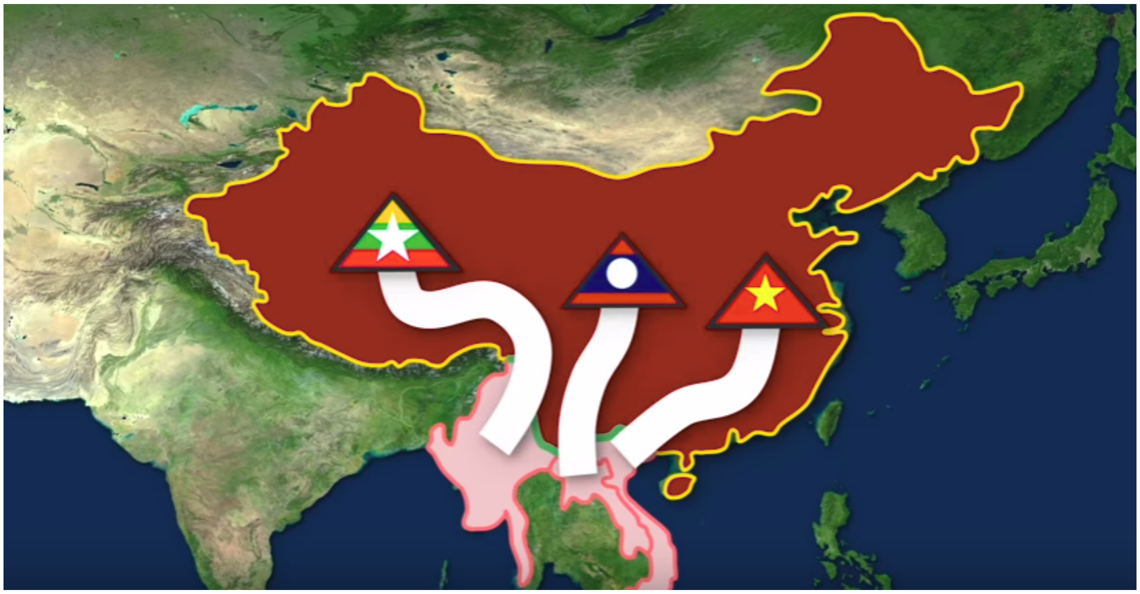


Figure A.14: The documentary [14] shows what China's problem with its neighbours.

The data-driven video shows how China has a problem with its geography situations. A.14 The video uses map and event visualization to show the conflicts between China and surrounding countries.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.13 Imaginary Numbers Are Real

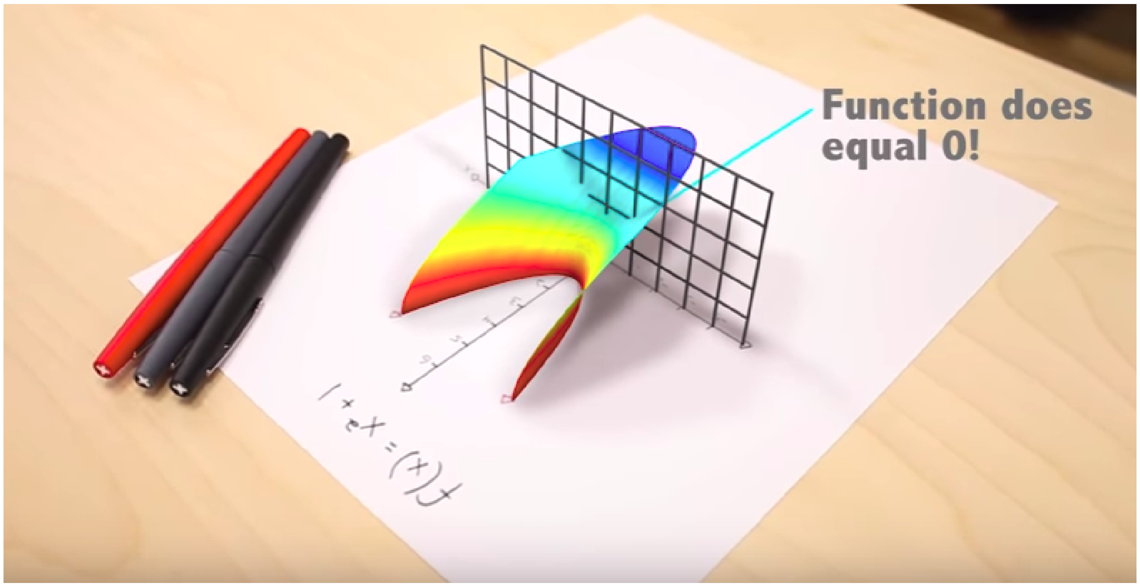


Figure A.15: The documentary [15] shows where real number comes from.

The data-driven video shows how imaginary numbers are generated and why they are useful. A.15 The diagrams and numbers shows how the imaginary numbers are different compared to rational numbers.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.14 Big Data Revolution

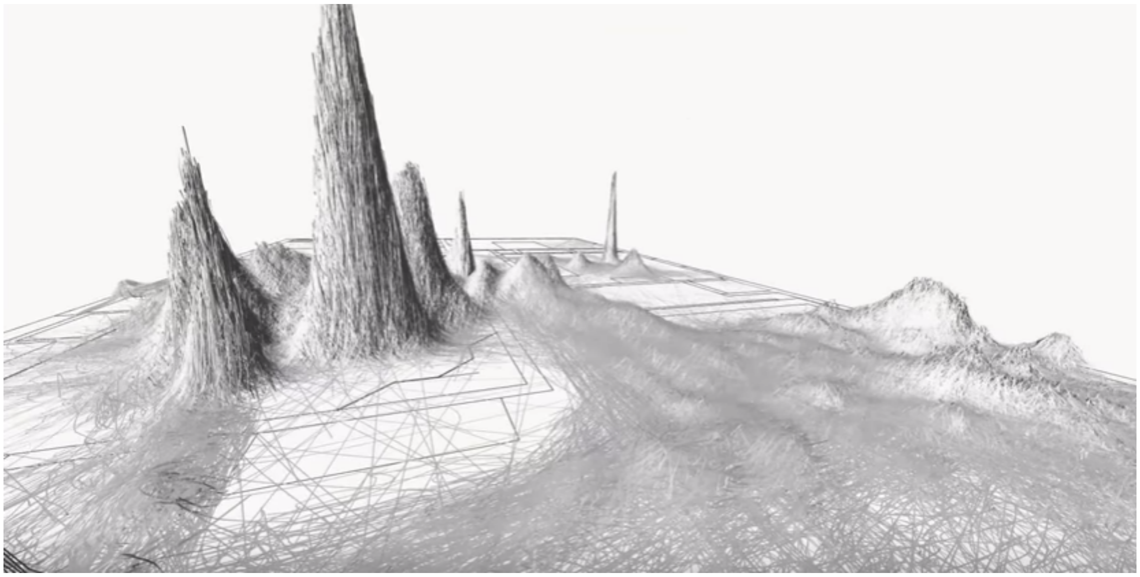


Figure A.16: The documentary [16] shows how the revolution of big data take place.

The data-driven video shows how animation and statistical visualization with VR and AR can change the use of big data in daily life. A.16

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.15 The Truth About Population

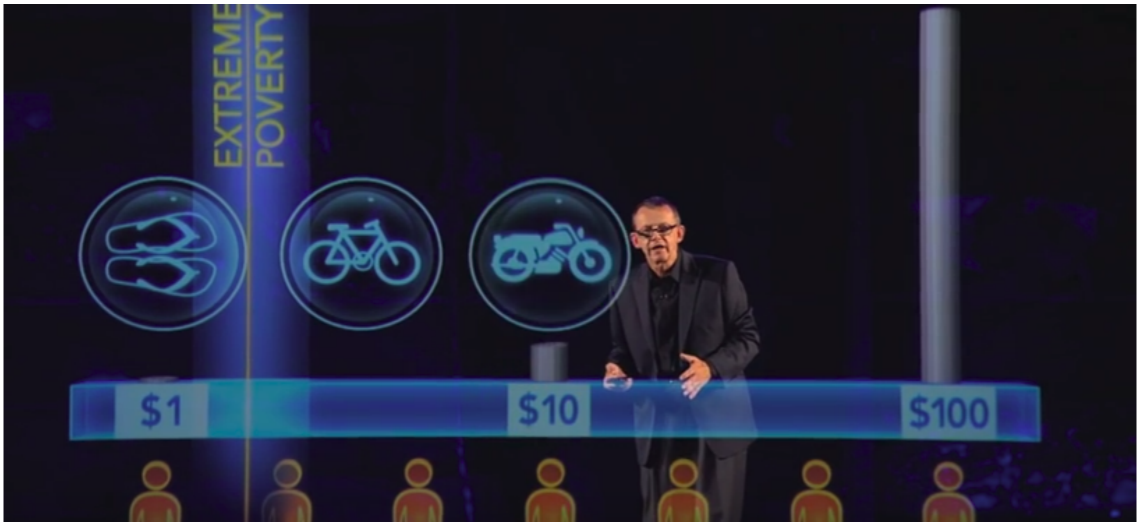


Figure A.17: The documentary [17] shows the relation between wealth and the size of population.

The data-driven video shows how population group with different levels of income will be treated in the society Figure. A.17 The barchart shows that the group with 100 dollar income will ignore the different between groups of 10 dollar income and 1 dollar income.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.16 Inside the mind of a master procrastinator

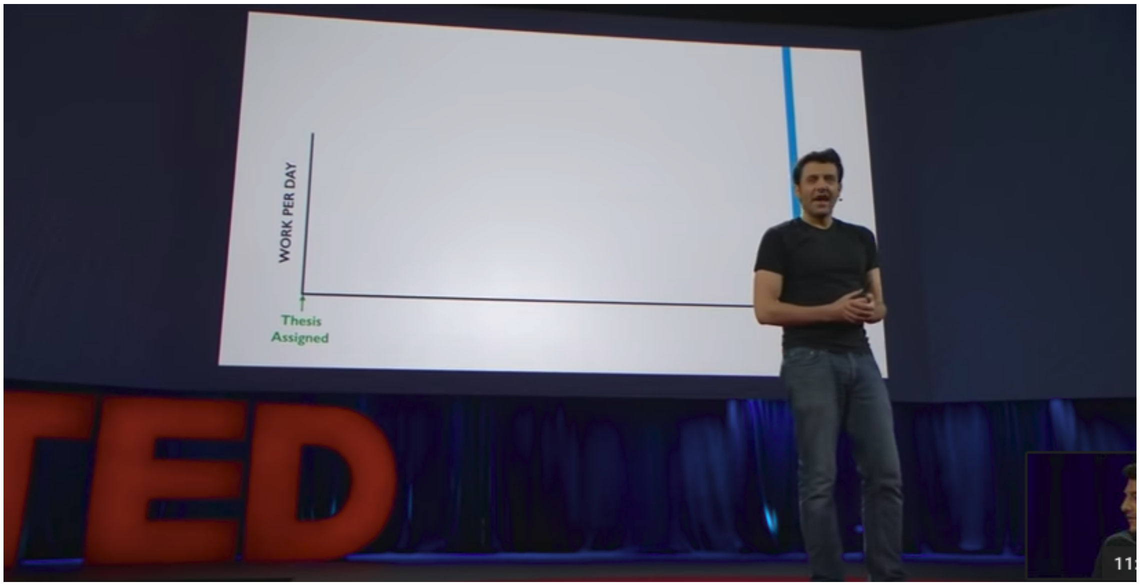


Figure A.18: The documentary [18] shows how exactly some procrastinator thinks.

The data-driven video shows the process that someone used to procrastinating gradually turned into a master procrastinator Figure. A.18 The animation of event and statistical visualization shows how his time is distributed.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many peo-*

ple on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload.* The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.17 How data will transform business

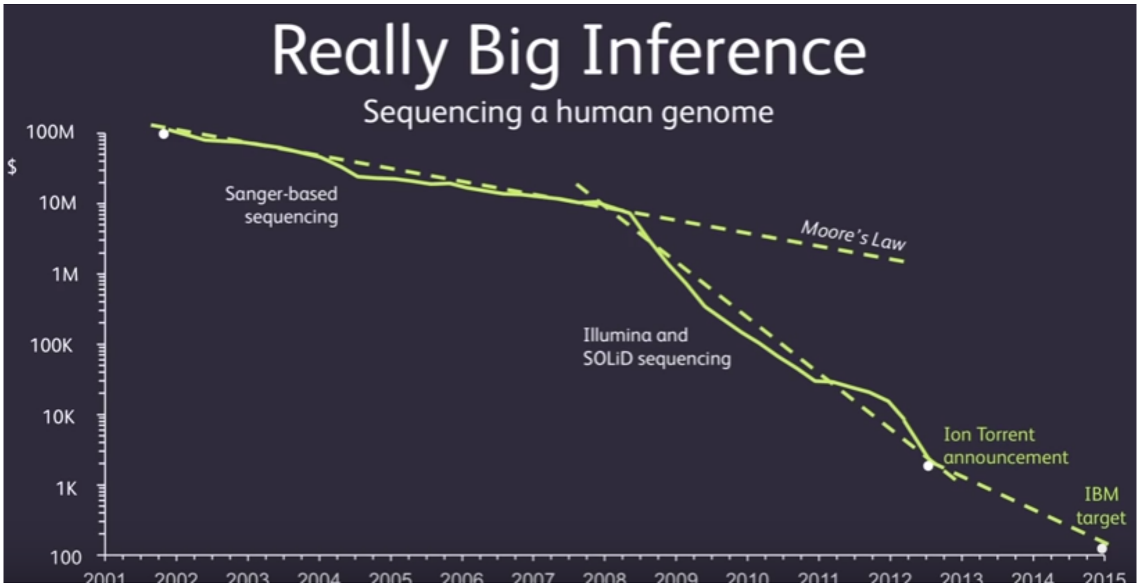


Figure A.19: The documentary [19] shows how big data transform business

The data-driven video shows how digital revolution takes place and how the stock of data is booming with statistical, event and continuous visualization in

Figure A.19.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.1.18 Will Saving Poor Children Lead to Overpopulation

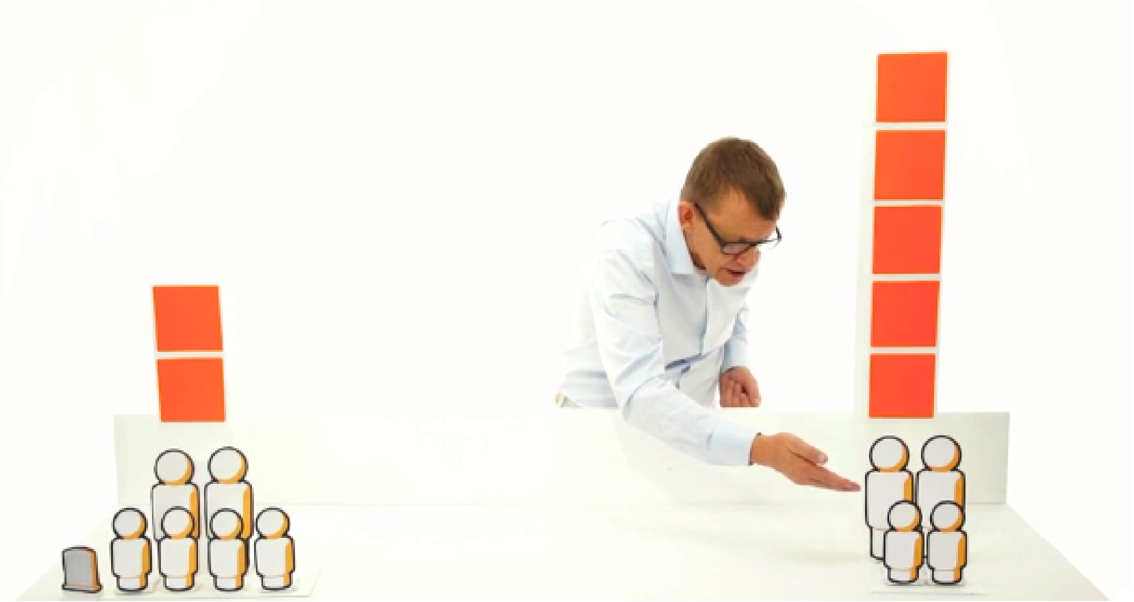


Figure A.20: The video [20] shows how the poor has more children than others.

The data-driven video shows how saving poor children will stop overpopulation in Figure. A.20 The bar chart shows that the majority of the population have family patterns that two parents only have two children, while the poor family have more children than two on average with some children dying at young age.

Informal classification:

As a documentary that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Because full-motion video is a *high-bandwidth medium*, and the user has *no control* over the pacing (except to pause the playback) there is significant potential for *cognitive overload*. The *visual components* used include full-motion video, map, animated graphics, text, and non-interactive data visualizations. Video is *static* in that it cannot be manipulated or interacted with (except for controlling the playback) by the audience. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2 Data Comics

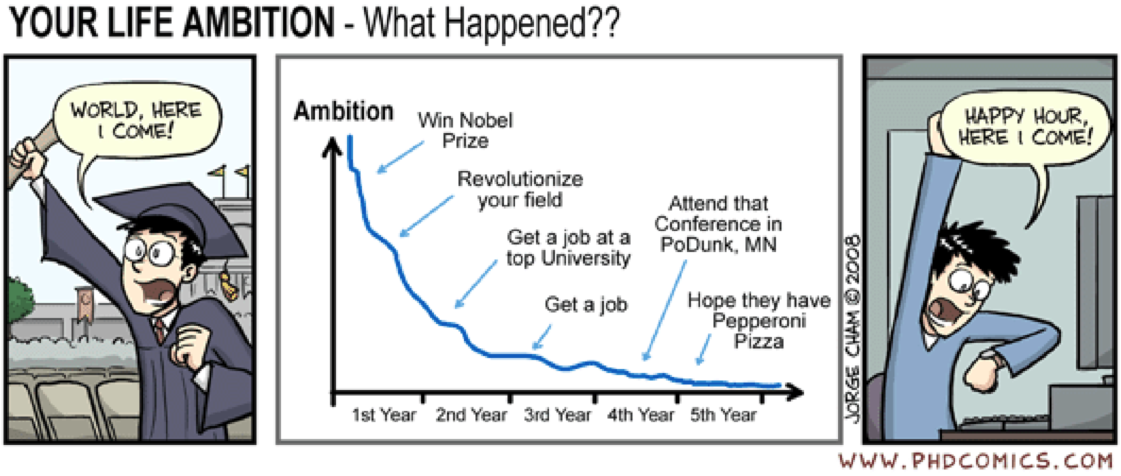


Figure A.21: Phd comic [21]. The ambition decreases along time.

A.2.1 PhD Comic

This comics in Figure. A.21 has a story that the PhDs are stressed and their ambitions are decreasing. The content and results of the two stories are shown by the first and last frame separately. The data, which is shown in a qualitative way in the middle frame showing a decreasing trend.

Informal classification: As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no control.* The *visual components* used include full- text, photographics, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch.* However, it is also *replicable*, as it is stored and can be played back at any time.

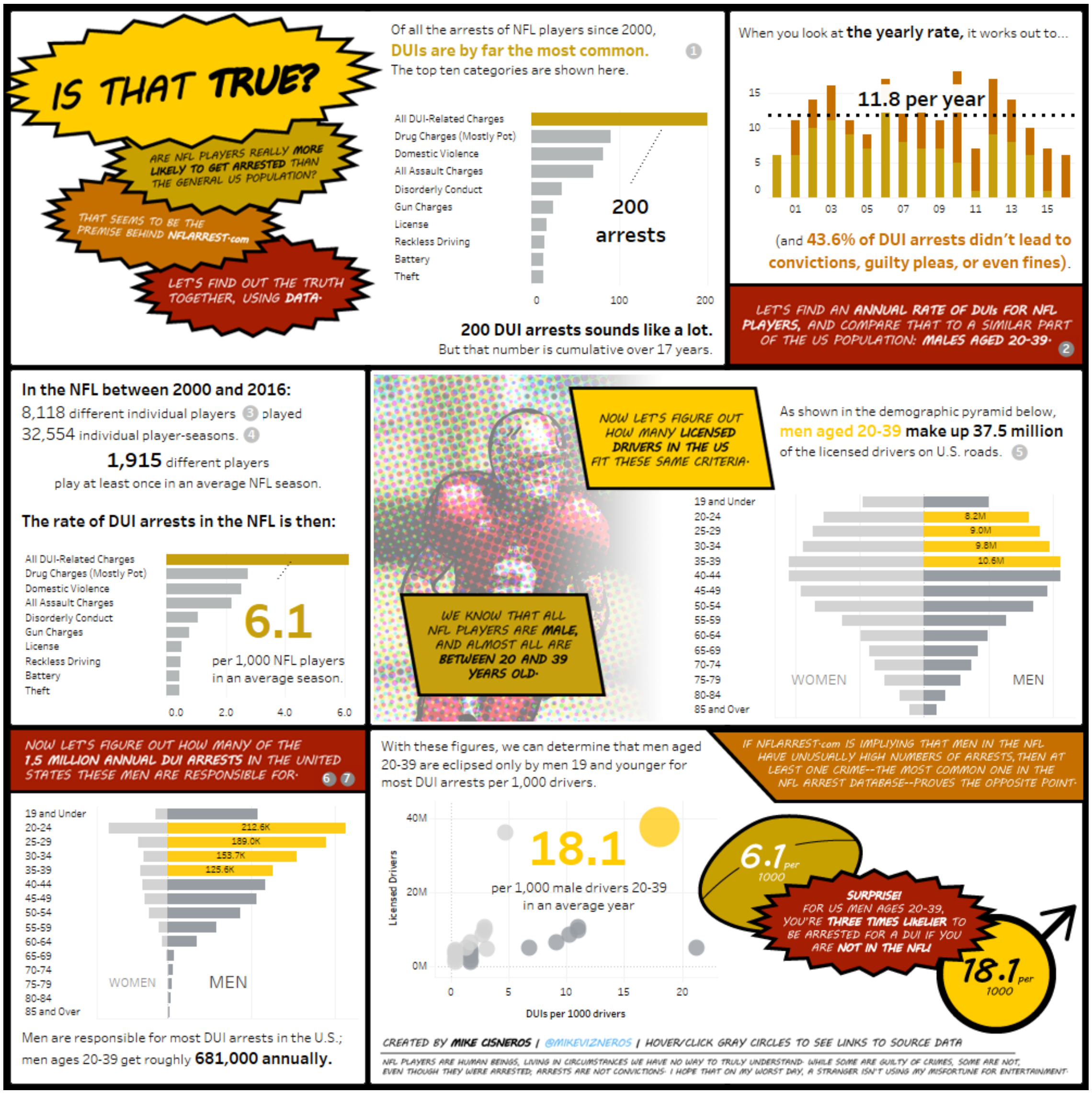

IS THAT TRUE?
ARE NFL PLAYERS REALLY MORE LIKELY TO GET ARRESTED THAN THE GENERAL US POPULATION?
THAT SEEMS TO BE THE PREMISE BEHIND NFLARREST·com
LET'S FIND OUT THE TRUTH TOGETHER, USING DATA·
Of all the arrests of NFL players since 2000,
DUIs are by far the most common. 1
The top ten categories are shown here.
All DUI-Related Charges
Drug Charges (Mostly Pot)
Domestic Violence
All Assault Charges
Disorderly Conduct
Gun Charges
License
Reckless Driving
Battery
Theft
0
100
200
200 arrests
200 DUI arrests sounds like a lot.
But that number is cumulative over 17 years.
When you look at the yearly rate, it works out to...
15
10
5
0
11.8 per year
01
03
05
07
09
11
13
15
(and 43.6% of DUI arrests didn't lead to convictions, guilty pleas, or even fines).
LET'S FIND AN ANNUAL RATE OF DUIs FOR NFL PLAYERS, AND COMPARE THAT TO A SIMILAR PART OF THE US POPULATION: MALES AGED 20-39· 2
In the NFL between 2000 and 2016:
8,118 different individual players 3 played
32,554 individual player-seasons. 4
1,915 different players
play at least once in an average NFL season.
The rate of DUI arrests in the NFL is then:
All DUI-Related Charges
Drug Charges (Mostly Pot)
Domestic Violence
All Assault Charges
Disorderly Conduct
Gun Charges
License
Reckless Driving
Battery
Theft
6.1
per 1,000 NFL players in an average season.
0.0
2.0
4.0
6.0
NOW LET'S FIGURE OUT HOW MANY LICENSED DRIVERS IN THE US FIT THESE SAME CRITERIA·
WE KNOW THAT ALL NFL PLAYERS ARE MALE, AND ALMOST ALL ARE BETWEEN 20 AND 39 YEARS OLD·
As shown in the demographic pyramid below, men aged 20-39 make up 37.5 million of the licensed drivers on U.S. roads. 5
19 and Under
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85 and Over
8.2M
9.0M
9.8M
10.6M
WOMEN
MEN
NOW LET'S FIGURE OUT HOW MANY OF THE 1.5 MILLION ANNUAL DUI ARRESTS IN THE UNITED STATES THESE MEN ARE RESPONSIBLE FOR· 6 7
19 and Under
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85 and Over
212.6K
189.0K
153.7K
125.6K
WOMEN
MEN
Men are responsible for most DUI arrests in the U.S.; men ages 20-39 get roughly 681,000 annually.
With these figures, we can determine that men aged 20-39 are eclipsed only by men 19 and younger for most DUI arrests per 1,000 drivers.
IF NFLARREST·com IS IMPLYING THAT MEN IN THE NFL HAVE UNUSUALLY HIGH NUMBERS OF ARRESTS, THEN AT LEAST ONE CRIME--THE MOST COMMON ONE IN THE NFL ARREST DATABASE--PROVES THE OPPOSITE POINT·
40M
20M
0M
Licensed Drivers
18.1
per 1,000 male drivers 20-39 in an average year
0
5
10
15
20
DUIs per 1000 drivers
6.1 per 1000
SURPRISE! FOR US MEN AGES 20-39, YOU'RE THREE TIMES LIKELIER TO BE ARRESTED FOR A DUI IF YOU ARE NOT IN THE NFL!
18.1 per 1000
CREATED BY MIKE CISNEROS | @MIKEVIZNEROS | HOVER/CLICK GRAY CIRCLES TO SEE LINKS TO SOURCE DATA
NFL PLAYERS ARE HUMAN BEINGS, LIVING IN CIRCUMSTANCES WE HAVE NO WAY TO TRULY UNDERSTAND· WHILE SOME ARE GUILTY OF CRIMES, SOME ARE NOT, EVEN THOUGH THEY WERE ARRESTED; ARRESTS ARE NOT CONVICTIONS· I HOPE THAT ON MY WORST DAY, A STRANGER ISN'T USING MY MISFORTUNE FOR ENTERTAINMENT·

Figure A.22: NFL player report [22]

A.2.2 NFL Player Data

The comic in Figure. A.22 is another example of the application of the design of data comics. The story is about the arrests data of NFL players. Firstly, the layout is divided into irregular frames like typical comic strips. The comic features such as speech bubble and comic figures are also leading the storyline. The texts in the speech bubble are transition sentences that guide the readers about what about happen and what to expect in this or next few frames. The data is represented as data visualization for bar chart and scatter plot, illustrating the detailed number such as number of DUI and number of arrests. Within the data visualizations comic features such as speech bubble and directional arrows are used to highlight numbers. The original page of this datacomics is interactive that the visualizations can be clicked to show more information.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no control.* The *visual components* used include full- text, photographics, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch.* However, it is also *replicable*, as it is stored and can be played back at any time.

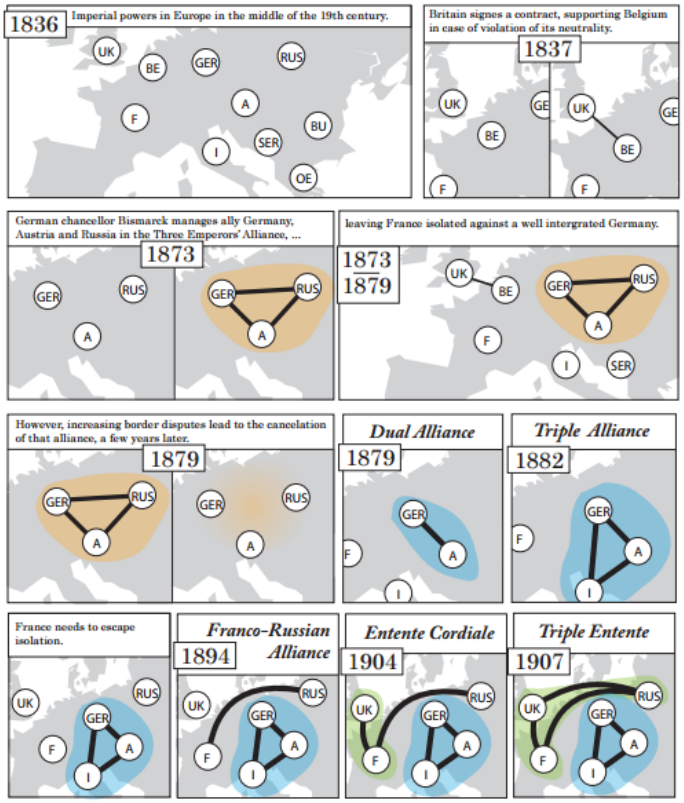

1836
Imperial powers in Europe in the middle of the 19th century.
UK BE GER RUS F A BU SER I OE
Britain signes a contract, supporting Belgium in case of violation of its neutrality.
1837
UK BE GE F
German chancellor Bismarck manages ally Germany, Austria and Russia in the Three Emperors' Alliance, ...
1873
GER RUS A
leaving France isolated against a well intergrated Germany.
1873
1879
UK BE GER RUS A F I SER
However, increasing border disputes lead to the cancelation of that alliance, a few years later.
1879
Dual Alliance
1879
GER A F I
Triple Alliance
1882
GER A I F
France needs to escape isolation.
UK RUS GER F A I
Franco-Russian Alliance
1894
Entente Cordiale
1904
Triple Entente
1907

Figure A.23: Graph comic for European relations.

A.2.3 Graphic Comic

Bach et al. [151] expended the design of data comics into network graphs A.23. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into strips of frames like comic strips. The story is about how European countries formed alliance during the early 1900s. The data is about the process that how different countries changed their relationships along time. The storytelling process of this example is simple with text description and comic style network graphs, but it enables the general audience to fast understand complicated temporal changes.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, photographics, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.4 Comic style Dashboard

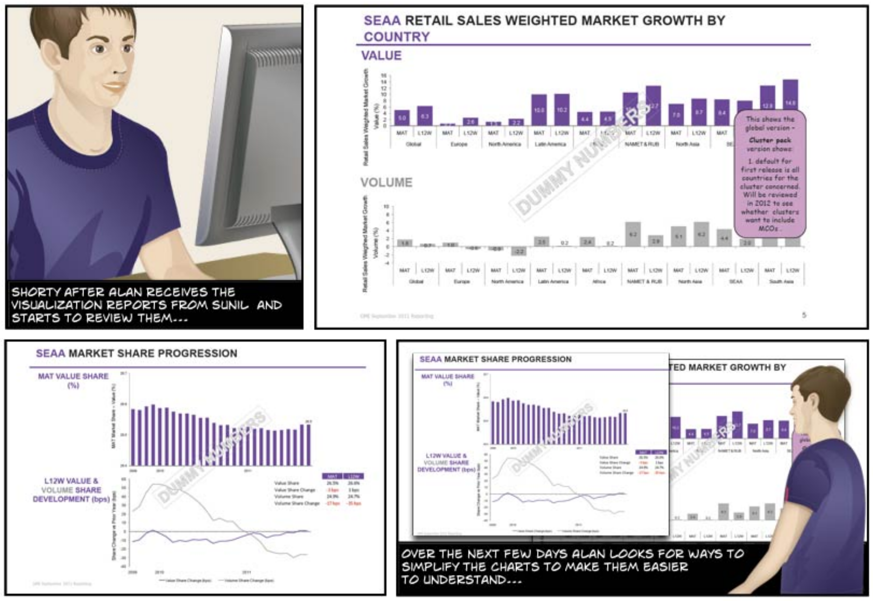

SHORTY AFTER ALAN RECEIVES THE VISUALIZATION REPORTS FROM SUNIL AND STARTS TO REVIEW THEM...
SEAA RETAIL SALES WEIGHTED MARKET GROWTH BY COUNTRY
VALUE
VOLUME
DUMMY NUMBERS
SEAA MARKET SHARE PROGRESSION
MAT VALUE SHARE (%)
L12W VALUE & VOLUME SHARE DEVELOPMENT (bps)
OVER THE NEXT FEW DAYS ALAN LOOKS FOR WAYS TO SIMPLIFY THE CHARTS TO MAKE THEM EASIER TO UNDERSTAND...

Figure A.24: Comic style dashboard [23]

The design of comic style dashboard with data visualization A.24 can be expended into panels with comics and data visualization. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into sequence of panels. The story is about how a commentator can help the illustration of an idea with data visualization. The data is the marketing and sales trend.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.5 Infographic Comic

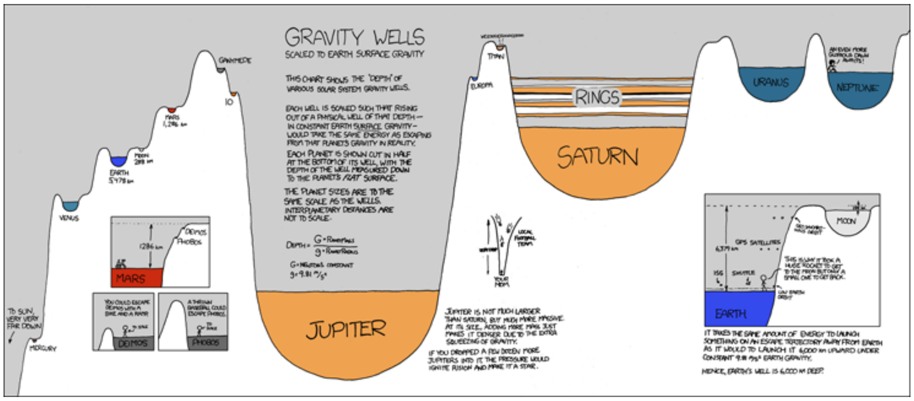

GRAVITY WELLS
SCALED TO EARTH SURFACE GRAVITY
THIS CHART SHOWS THE 'DEPTH' OF VARIOUS SOLAR SYSTEM GRAVITY WELLS.
EACH WELL IS SCALED SUCH THAT RISING OUT OF A PHYSICAL WELL OF THAT DEPTH— IN CONSTANT EARTH SURFACE GRAVITY— WOULD TAKE THE SAME ENERGY AS ESCAPING FROM THAT PLANET'S GRAVITY IN REALITY.
EACH PLANET IS SHOWN CUT IN HALF AT THE BOTTOM OF ITS WELL, WITH THE DEPTH OF THE WELL MEASURED DOWN TO THE PLANET'S /FLAT/ SURFACE.
THE PLANET SIZES ARE TO THE SAME SCALE AS THE WELLS. INTERPLANETARY DISTANCES ARE NOT TO SCALE.
DEPTH = G·M_PLANET / g·R_PLANET
G = NEWTON'S CONSTANT
g = 9.81 m/s²
TO SUN, VERY VERY FAR DOWN
MERCURY
VENUS
EARTH 5,478 km
MOON 288 km
MARS 1,286 km
GANYMEDE
IO
1286 km
DEIMOS
PHOBOS
MARS
YOU COULD ESCAPE DEIMOS WITH A BIKE AND A RAMP
TO SPACE
DEIMOS
A THROWN BASEBALL COULD ESCAPE PHOBOS.
TO SPACE
PHOBOS
JUPITER
JUPITER IS NOT MUCH LARGER THAN SATURN, BUT MUCH MORE MASSIVE. AT ITS SIZE, ADDING MORE MASS JUST MAKES IT DENSER DUE TO THE EXTRA SQUEEZING OF GRAVITY.
IF YOU DROPPED A FEW DOZEN MORE JUPITERS INTO IT, THE PRESSURE WOULD IGNITE FUSION AND MAKE IT A STAR.
EUROPA
TITAN
RINGS
SATURN
LOCAL FOOTBALL TEAM
YOUR MOM
URANUS
AN EVEN MORE GLORIOUS DAWN AWAITS!
NEPTUNE
MOON
GEOSYNCHRONOUS ORBIT
GPS SATELLITES
6,379 km
ISS
SHUTTLE
THIS IS WHY IT TOOK A HUGE ROCKET TO GET TO THE MOON BUT ONLY A SMALL ONE TO GET BACK.
LOW EARTH ORBIT
EARTH
IT TAKES THE SAME AMOUNT OF ENERGY TO LAUNCH SOMETHING ON AN ESCAPE TRAJECTORY AWAY FROM EARTH AS IT WOULD TO LAUNCH IT 6,000 km UPWARD UNDER CONSTANT 9.81 m/s² EARTH GRAVITY.
HENCE, EARTH'S WELL IS 6,000 km DEEP.

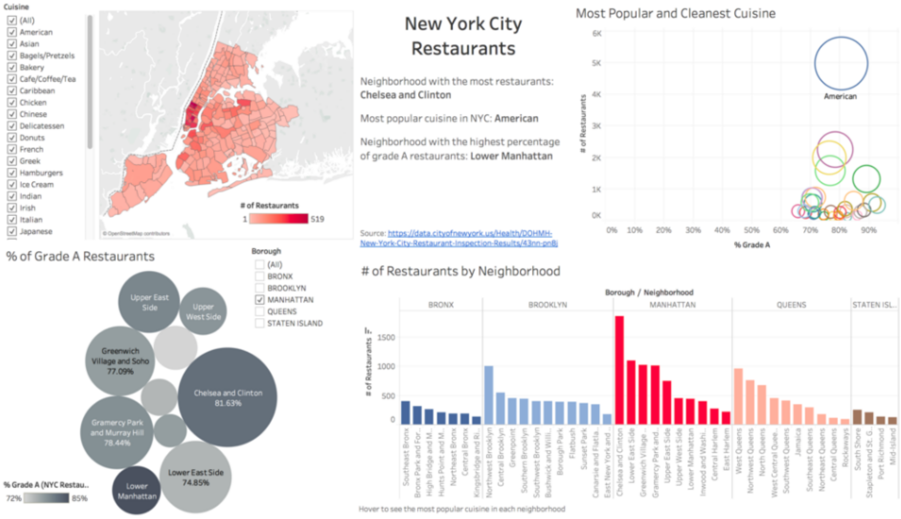
Figure A.25: Infographical Comics [24]

The design of data comics into network graphs A.25 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a big infographic. The story is about how gravity changes from the edge of the solar system to the inner circle such as Venus and Mars. The data is the gravity levels demonstrated as the level of water and the altitude of mountain.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.6 NYC Restaurant Data Vis Comic



New York City Restaurants

Neighborhood with the most restaurants: **Chelsea and Clinton**

Most popular cuisine in NYC: **American**

Neighborhood with the highest percentage of grade A restaurants: **Lower Manhattan**

Source: https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j

Hover to see the most popular cuisine in each neighborhood

Figure A.26: Comics for data of restaurants in NYC [25]

The design of data comics with statistic graphs A.26 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The story is about how the style and location of restaurants in New York City are distributed. Different visualization types are used to demonstrate the different data fields of the restaurants.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.7 Marvel vs DC Comics

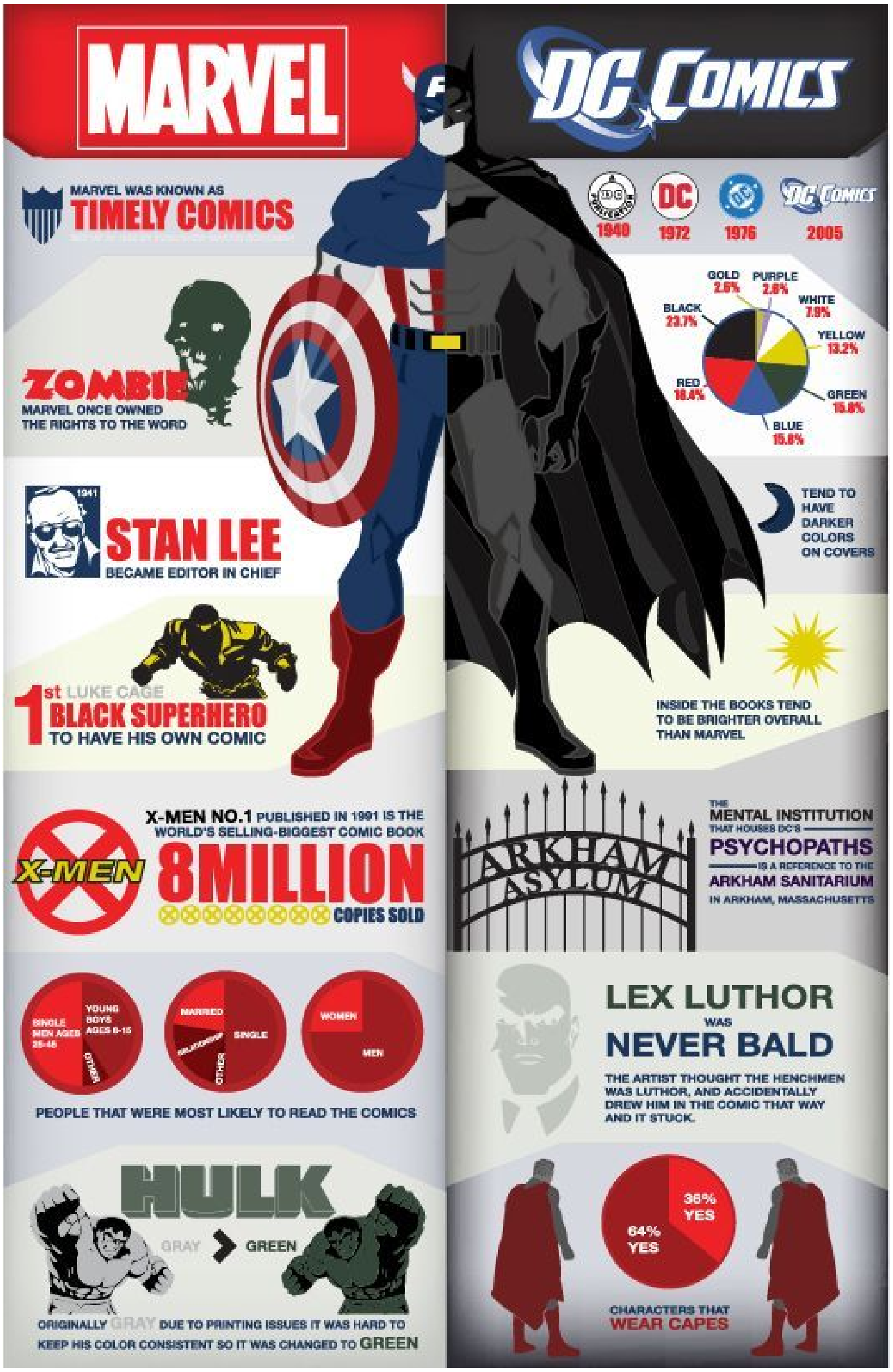


Figure A.27: Comics of the comparison of characters from Marvel and DC [26]

The design of data comics with statistic graphs A.27 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a big infographic. The story is about the comparison of the comic figures from Marvel and DC series. The data is the visualization of figure distribution and comparison of different fields such as dressing and mental status.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.8 Body Cartoon

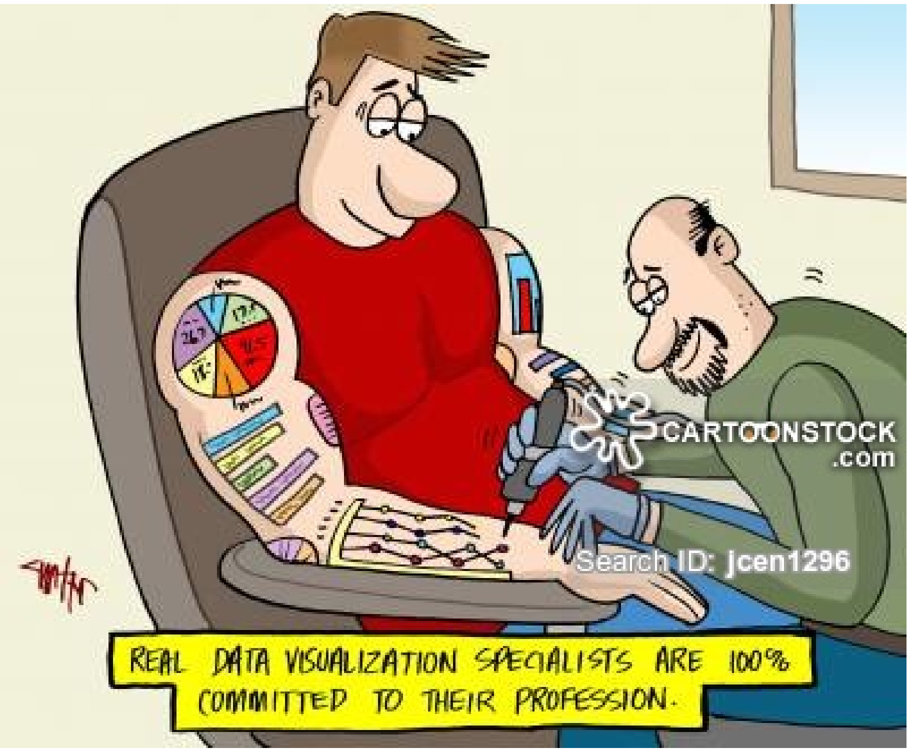

CARTOONSTOCK
.com
Search ID: jcen1296
REAL DATA VISUALIZATION SPECIALISTS ARE 100% COMMITTED TO THEIR PROFESSION.

Figure A.28: Comics of someone having tatoo [27] on his arm as data visualization.

The design of data comics with statistic graphs A.28 can be expended into data visualization with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a few visualizations on one human arm. The story is about how the visualization is rendered as tattoo on a human body. The data is the location of the visualization and the data visualizations themselves.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.9 Spider Man Comic Visualization

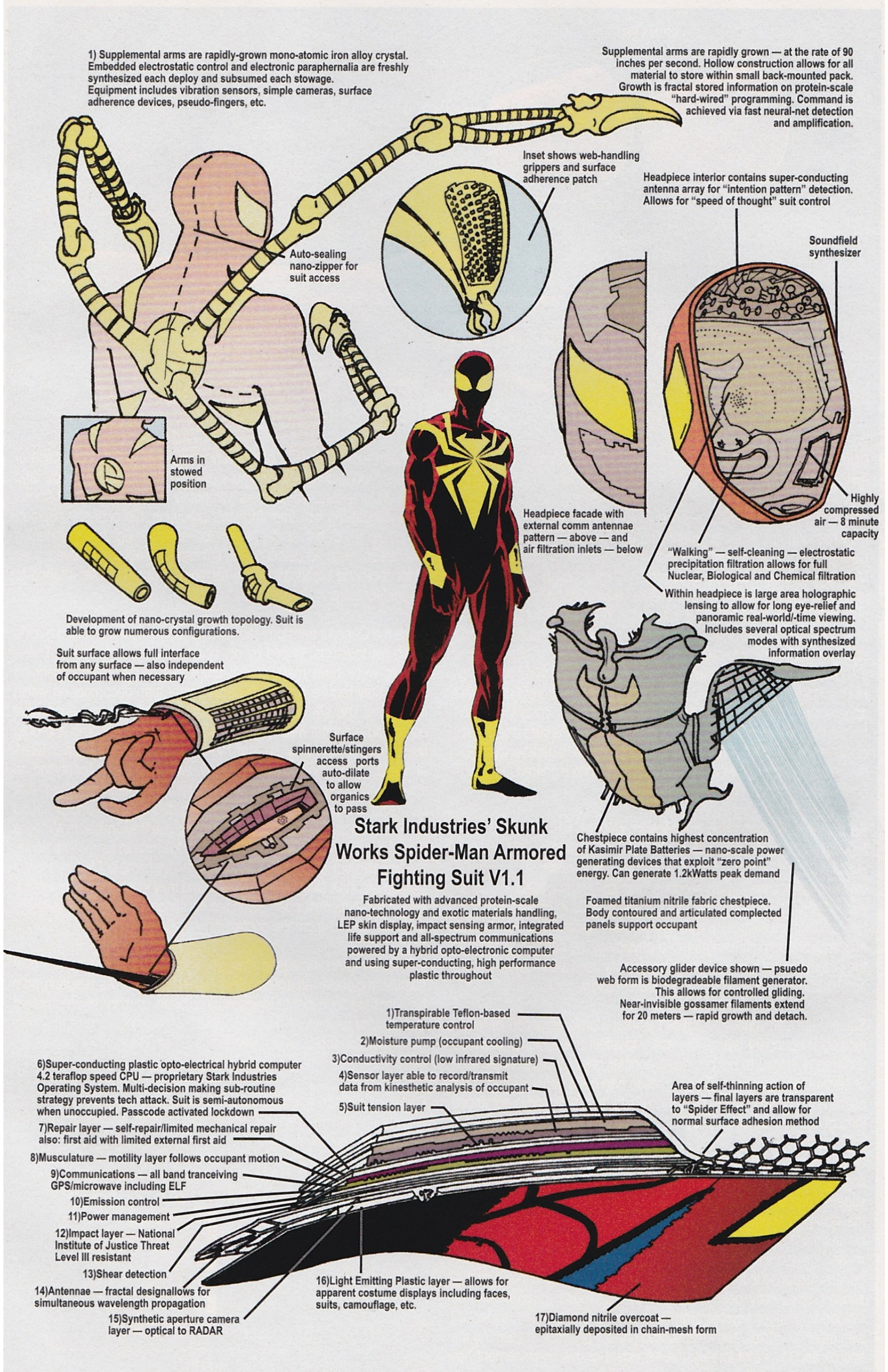

1) Supplemental arms are rapidly-grown mono-atomic iron alloy crystal. Embedded electrostatic control and electronic paraphernalia are freshly synthesized each deploy and subsumed each stowage. Equipment includes vibration sensors, simple cameras, surface adherence devices, pseudo-fingers, etc.
Supplemental arms are rapidly grown — at the rate of 90 inches per second. Hollow construction allows for all material to store within small back-mounted pack. Growth is fractal stored information on protein-scale "hard-wired" programming. Command is achieved via fast neural-net detection and amplification.
Inset shows web-handling grippers and surface adherence patch
Headpiece interior contains super-conducting antenna array for "intention pattern" detection. Allows for "speed of thought" suit control
Soundfield synthesizer
Auto-sealing nano-zipper for suit access
Arms in stowed position
Highly compressed air — 8 minute capacity
Headpiece facade with external comm antennae pattern — above — and air filtration inlets — below
"Walking" — self-cleaning — electrostatic precipitation filtration allows for full Nuclear, Biological and Chemical filtration
Within headpiece is large area holographic lensing to allow for long eye-relief and panoramic real-world/-time viewing. Includes several optical spectrum modes with synthesized information overlay
Development of nano-crystal growth topology. Suit is able to grow numerous configurations.
Suit surface allows full interface from any surface — also independent of occupant when necessary
Surface spinnerette/stingers access ports auto-dilate to allow organics to pass
Stark Industries' Skunk Works Spider-Man Armored Fighting Suit V1.1
Fabricated with advanced protein-scale nano-technology and exotic materials handling, LEP skin display, impact sensing armor, integrated life support and all-spectrum communications powered by a hybrid opto-electronic computer and using super-conducting, high performance plastic throughout
Chestpiece contains highest concentration of Kasimir Plate Batteries — nano-scale power generating devices that exploit "zero point" energy. Can generate 1.2kWatts peak demand
Foamed titanium nitrile fabric chestpiece. Body contoured and articulated complected panels support occupant
Accessory glider device shown — psuedo web form is biodegradeable filament generator. This allows for controlled gliding. Near-invisible gossamer filaments extend for 20 meters — rapid growth and detach.
1)Transpirable Teflon-based temperature control
2)Moisture pump (occupant cooling)
3)Conductivity control (low infrared signature)
4)Sensor layer able to record/transmit data from kinesthetic analysis of occupant
5)Suit tension layer
6)Super-conducting plastic opto-electrical hybrid computer 4.2 teraflop speed CPU — proprietary Stark Industries Operating System. Multi-decision making sub-routine strategy prevents tech attack. Suit is semi-autonomous when unoccupied. Passcode activated lockdown
7)Repair layer — self-repair/limited mechanical repair also: first aid with limited external first aid
8)Musculature — motility layer follows occupant motion
9)Communications — all band tranceiving GPS/microwave including ELF
10)Emission control
11)Power management
12)Impact layer — National Institute of Justice Threat Level III resistant
13)Shear detection
14)Antennae — fractal designallows for simultaneous wavelength propagation
15)Synthetic aperture camera layer — optical to RADAR
16)Light Emitting Plastic layer — allows for apparent costume displays including faces, suits, camouflage, etc.
17)Diamond nitrile overcoat — epitaxially deposited in chain-mesh form
Area of self-thinning action of layers — final layers are transparent to "Spider Effect" and allow for normal surface adhesion method

Figure A.29: Spider Man Visualization in Comics [28]

The design of data comics into network graphs A.29 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a big infographic. The story is about the technology that the Spiderman's suit uses. The data is the layers and functions of the suit.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.10 Cell Phone Comic

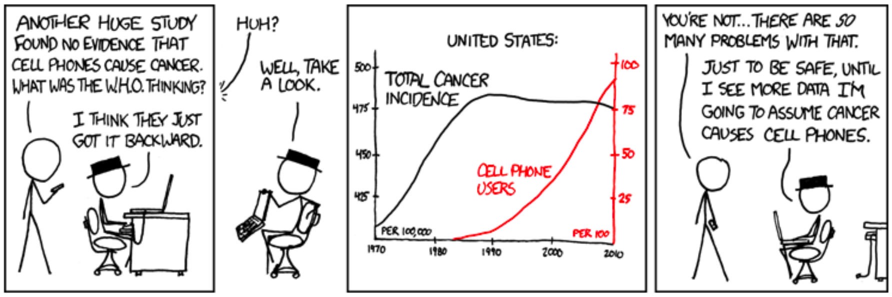

ANOTHER HUGE STUDY FOUND NO EVIDENCE THAT CELL PHONES CAUSE CANCER. WHAT WAS THE W.H.O. THINKING?
I THINK THEY JUST GOT IT BACKWARD.
HUH?
WELL, TAKE A LOOK.
UNITED STATES:
TOTAL CANCER INCIDENCE
CELL PHONE USERS
500
475
450
425
PER 100,000
100
75
50
25
PER 100
1970
1980
1990
2000
2010
YOU'RE NOT... THERE ARE SO MANY PROBLEMS WITH THAT.
JUST TO BE SAFE, UNTIL I SEE MORE DATA I'M GOING TO ASSUME CANCER CAUSES CELL PHONES.

Figure A.30: Cell Phone Visualization in Comics [29]

The design of data comics into network graphs A.30 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about how the trends of cancer and cellphone changing.The data is the visualization that illustrate the number change of cancer incidents and cellphones.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.11 Linear Regression Comic

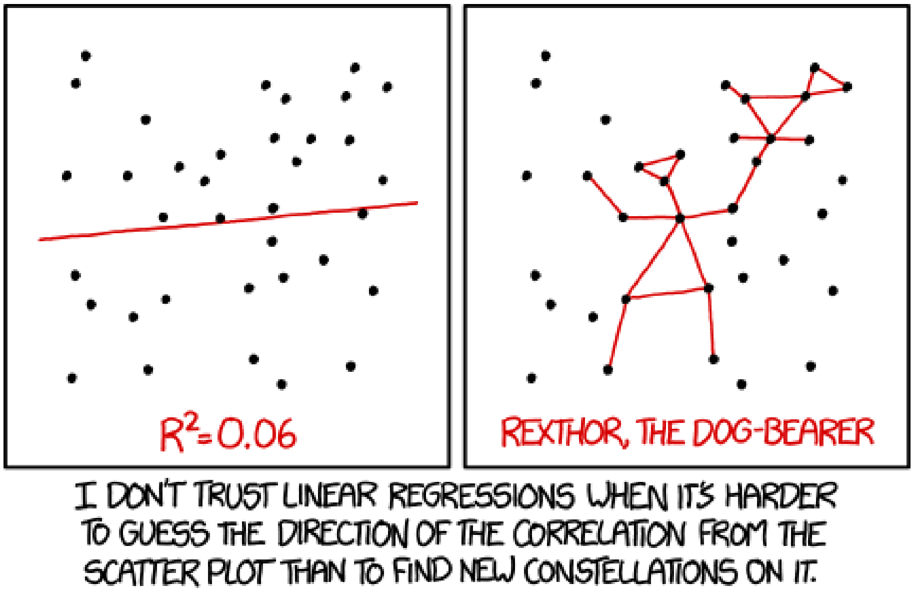

R^2=0.06
REXTHOR, THE DOG-BEARER
I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Figure A.31: Linear Regression Visualization in Comics [30]

The design of data comics into network graphs A.31 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the visualization of a linear regression.The data is the visualization comparison of linear regression and a randomly drawn diagram based on the data points.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.12 Vocation Stress Comic

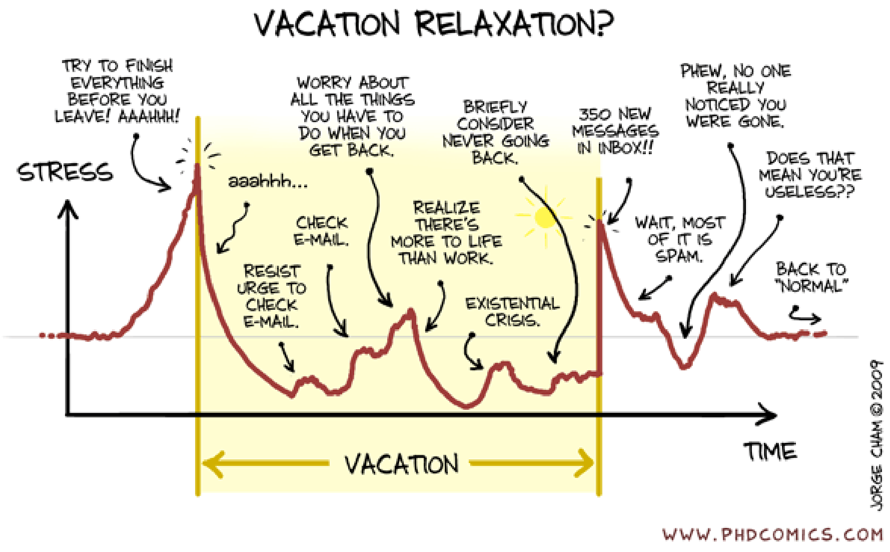

VACATION RELAXATION?
TRY TO FINISH EVERYTHING BEFORE YOU LEAVE! AAAHHH!
STRESS
aaahhh...
WORRY ABOUT ALL THE THINGS YOU HAVE TO DO WHEN YOU GET BACK.
CHECK E-MAIL.
RESIST URGE TO CHECK E-MAIL.
BRIEFLY CONSIDER NEVER GOING BACK.
REALIZE THERE'S MORE TO LIFE THAN WORK.
EXISTENTIAL CRISIS.
350 NEW MESSAGES IN INBOX!!
WAIT, MOST OF IT IS SPAM.
PHEW, NO ONE REALLY NOTICED YOU WERE GONE.
DOES THAT MEAN YOU'RE USELESS??
BACK TO "NORMAL"
VACATION
TIME
JORGE CHAM © 2009
WWW.PHDCOMICS.COM

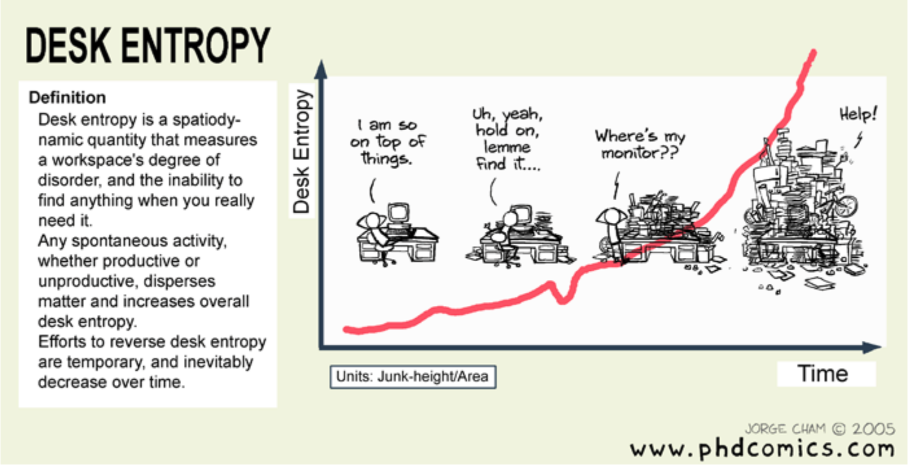
Figure A.32: Visualization to show the stress level change with vocation in Comics [31]

The design of data comics into network graphs A.32 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the visualization of the stress level as the vocation starts and ends.The data is the visualization of stress level changed not as expected for being interrupted by worrying about work during vocation.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.13 Desk Entropy Comic



DESK ENTROPY

Definition

Desk entropy is a spatiodynamic quantity that measures a workspace's degree of disorder, and the inability to find anything when you really need it.

Any spontaneous activity, whether productive or unproductive, disperses matter and increases overall desk entropy.

Efforts to reverse desk entropy are temporary, and inevitably decrease over time.

Figure A.33: The Increase of Entropy of Visualization in Comics [32]

The design of data comics into network graphs A.33 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the visualization about how messy the desk can be for a PhD student.The data is the visualization that a PhD student's desk is getting more and more messy during time.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.14 PhD Grooming Comic

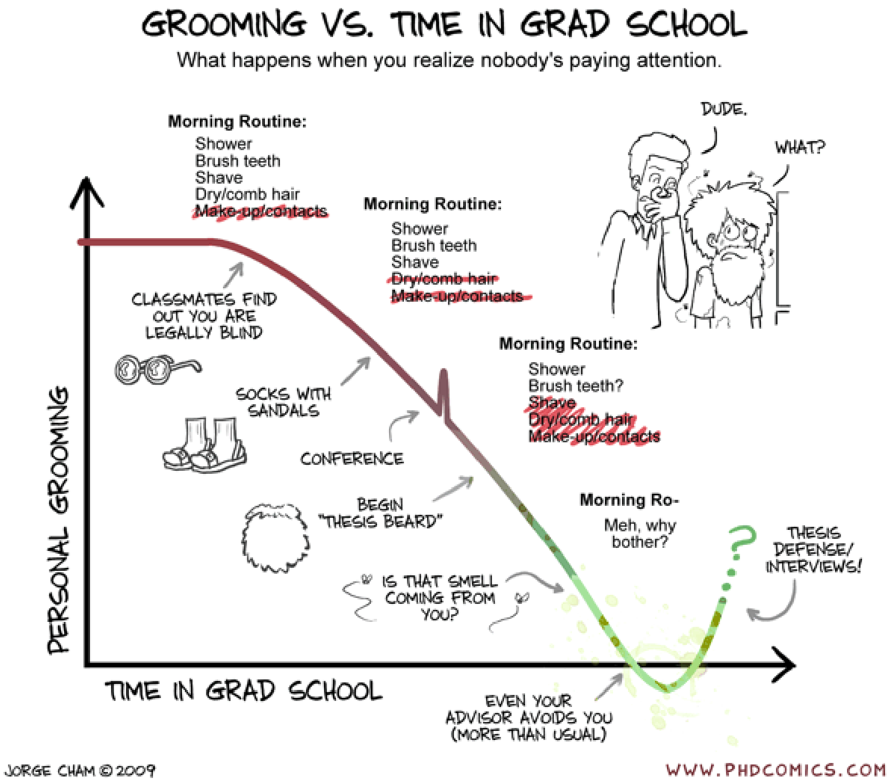


Figure A.34: The Need of Grooming of Visualization in Comics [33]

The design of data comics into network graphs A.33 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the visualization of the grooming condition of a PhD student.The data is the visualization of the grooming condition getting worse through time.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many* *people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.15 PhD Procrastination

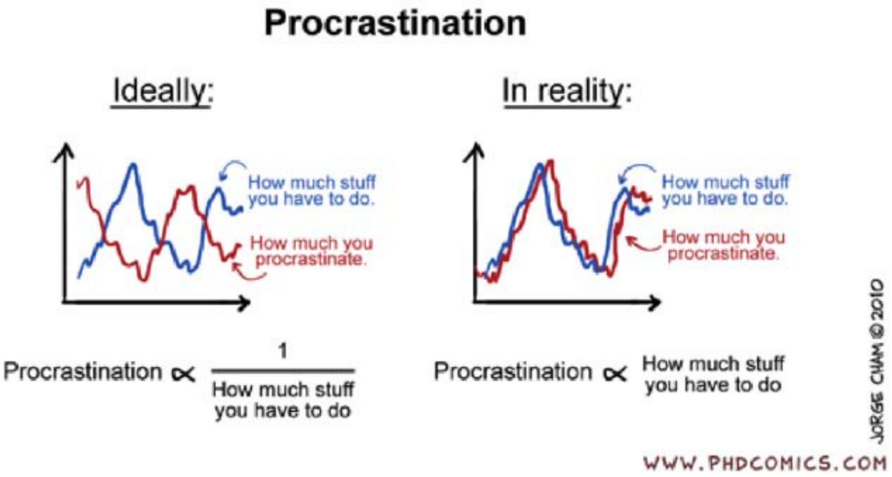

Procrastination
Ideally:
In reality:
How much stuff you have to do.
How much you procrastinate.
How much stuff you have to do.
How much you procrastinate.
Procrastination ∝ 1 / How much stuff you have to do
Procrastination ∝ How much stuff you have to do
JORGE CHAM © 2010
WWW.PHDCOMICS.COM

Figure A.35: The Change of Procrastination and Stress Level in Comics [34]

The design of data comics into network graphs A.35 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the visualization of procrastination condition. The data is the visualization of how level of procrastination increases when word load is high in reality.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.16 Day of Life an American

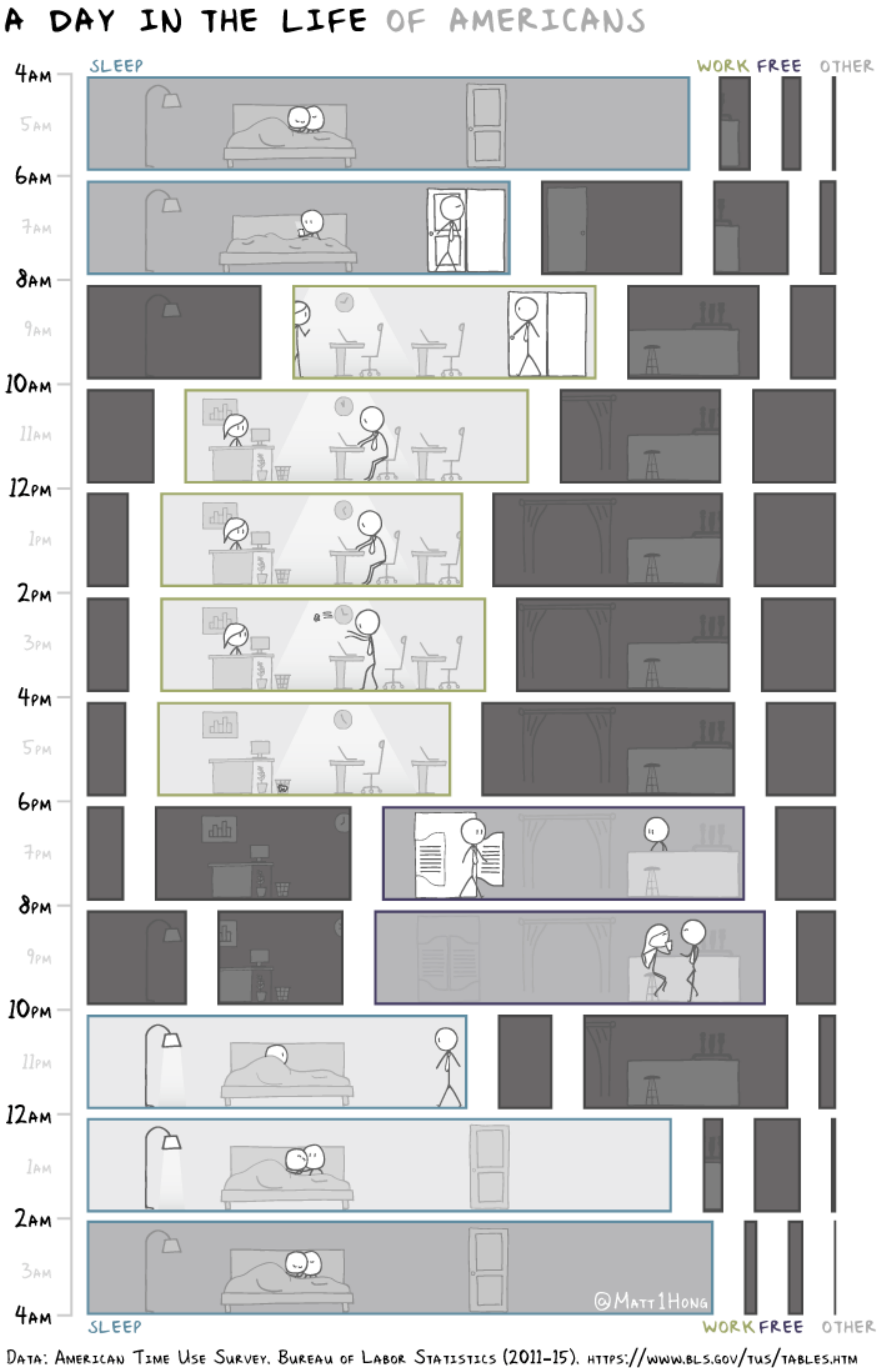

A DAY IN THE LIFE OF AMERICANS
SLEEP
WORK FREE OTHER
4AM
5AM
6AM
7AM
8AM
9AM
10AM
11AM
12PM
1PM
2PM
3PM
4PM
5PM
6PM
7PM
8PM
9PM
10PM
11PM
12AM
1AM
2AM
3AM
4AM
@Matt1Hong
SLEEP
WORK FREE OTHER
Data: American Time Use Survey, Bureau of Labor Statistics (2011–15). https://www.bls.gov/tus/tables.htm

Figure A.36: The Change of Procrastination and Stress Level in Comics [35]

The design of data comics into network graphs A.36 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the visualization of an average American's life. The data is the comic strips that shows what this person does throughout the day.

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.17 Curve Fitting Comic

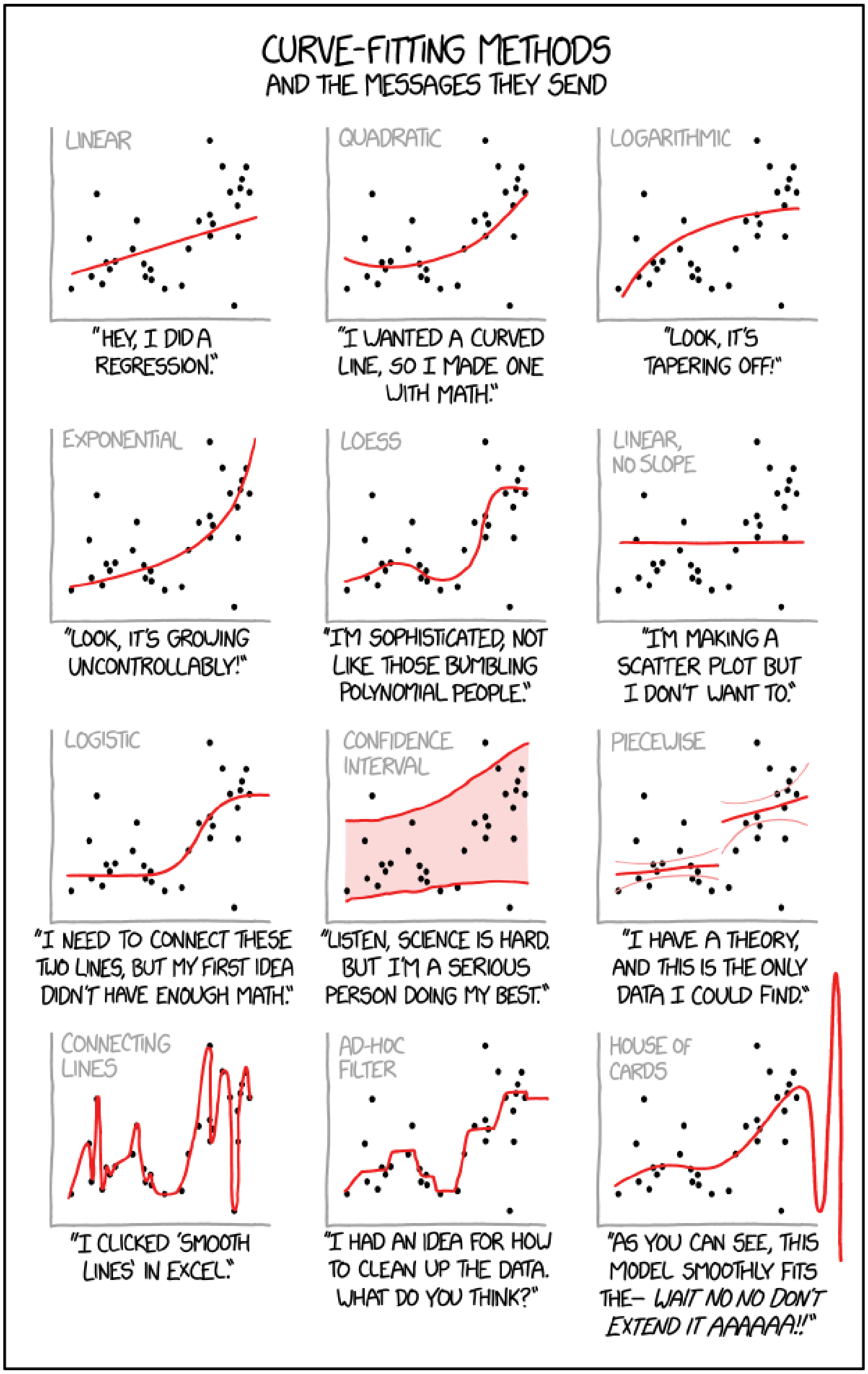

CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND
LINEAR
"HEY, I DID A REGRESSION."
QUADRATIC
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPERING OFF!"
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY!"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
LINEAR, NO SLOPE
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."
LOGISTIC
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."
CONFIDENCE INTERVAL
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."
CONNECTING LINES
"I CLICKED 'SMOOTH LINES' IN EXCEL."
AD-HOC FILTER
"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"
HOUSE OF CARDS
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE- WAIT NO NO DON'T EXTEND IT AAAAAA!!"

Figure A.37: Curve Fitting Visualization in Comics [36]

The design of data comics into network graphs A.37 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the comparison of all the curve fitting methods.The data is the visualization of the matching of different .

Informal classification:

As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no control.* The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.2.18 Seashell Probability Comic

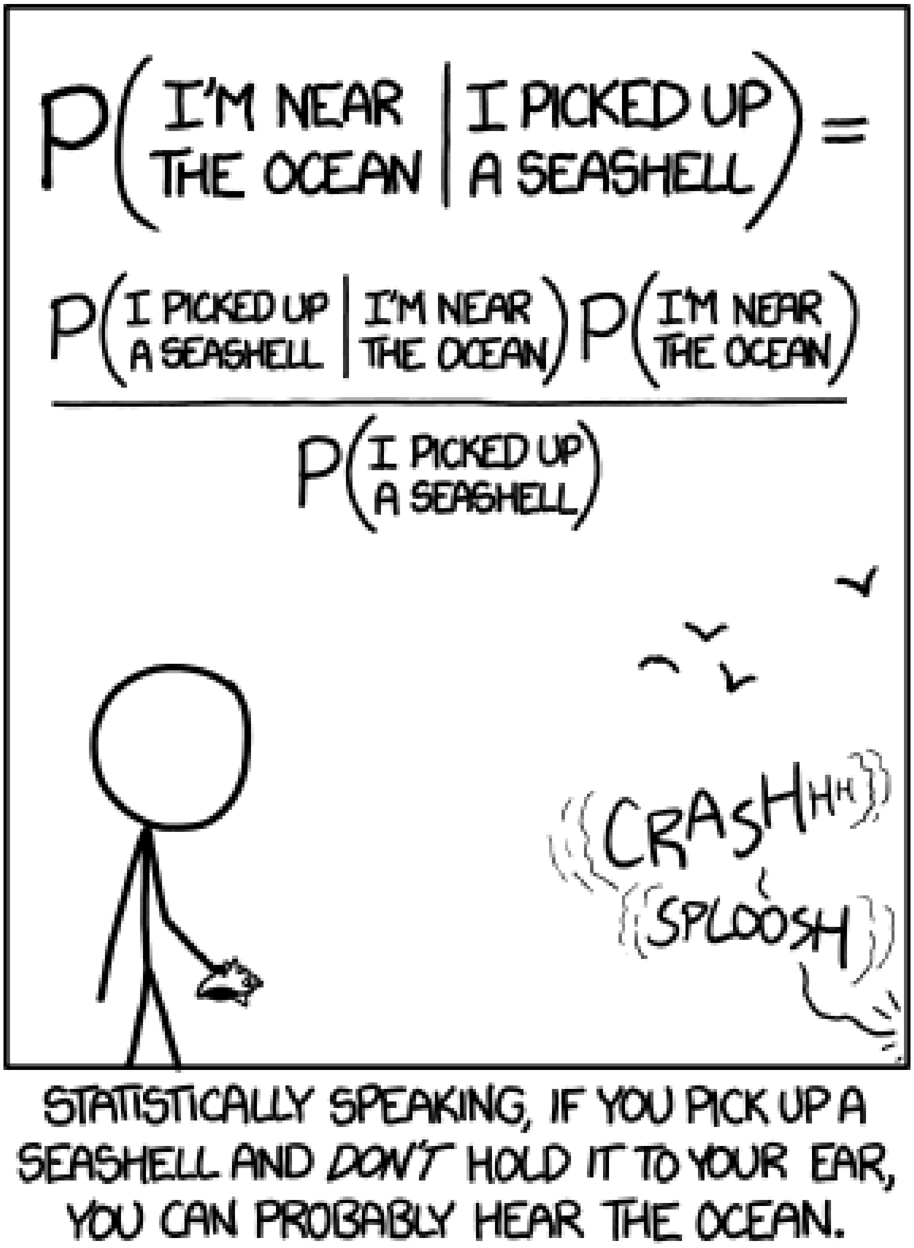

P(I'M NEAR THE OCEAN | I PICKED UP A SEASHELL) =
P(I PICKED UP A SEASHELL | I'M NEAR THE OCEAN) P(I'M NEAR THE OCEAN)
P(I PICKED UP A SEASHELL)
CRASHHH
SPLOOSH
STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND DON'T HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

Figure A.38: Illustration of Conditional Probability Visualized in Comics [37]

The design of data comics into network graphs A.38 can be expended into infographics with comics. This still falls into the category of telling data-driven stories with comics. The comic styles graphs are organized into a strip of comic with information visualizations. The story is about the comparison of all the curve fitting methods.The data is the visualization of the matching of different .

Informal classification:

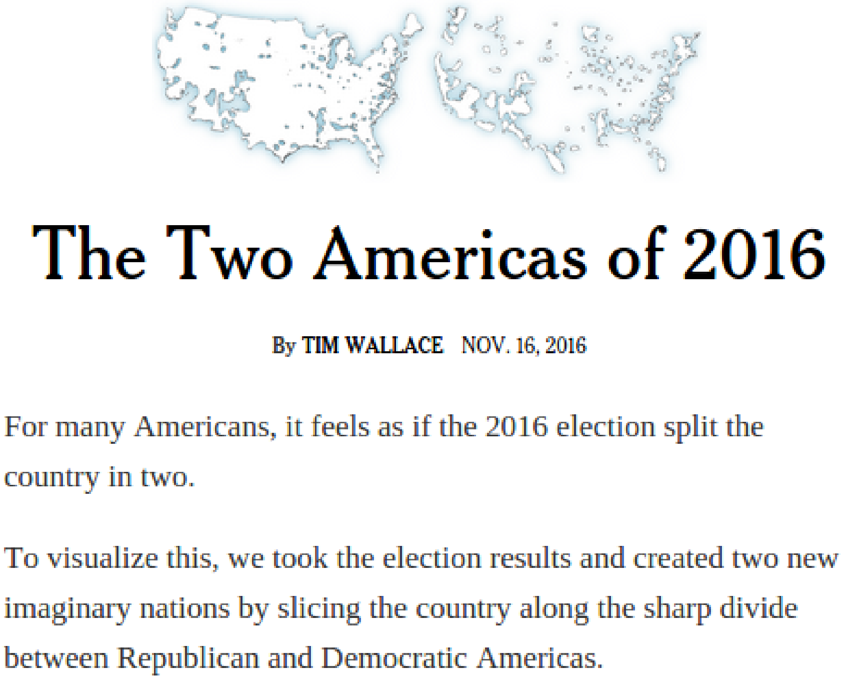
As a published comic that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no control*. The *visual components* used include full- text, comic figures, and data visualizations. The data visualization in this comic is *static* in that it cannot be manipulated or interacted with. The viewing sequence is in *sequence* or *branch*. However, it is also *replicable*, as it is stored and can be played back at any time.x

A.3 Web Article with Data Visualization

Some web pages consist of very rich media including text, image, and visualization. Being a bit different from infographics, rich web pages are more interactive, as all the media types inside a web page can be dynamic. The data can be updated in real time, and the visualization can be interactive. The forms of data in a web page is more versatile than infographics.

A.3.1 The Two Americas

Figure. (A.39, A.40, A.41, A.42, A.43, A.44) [194] are showing an example of a web page telling a data story about the comparison of election between Donald Trump and Hillary Clinton. The story is composed with visualization map, text description and table.



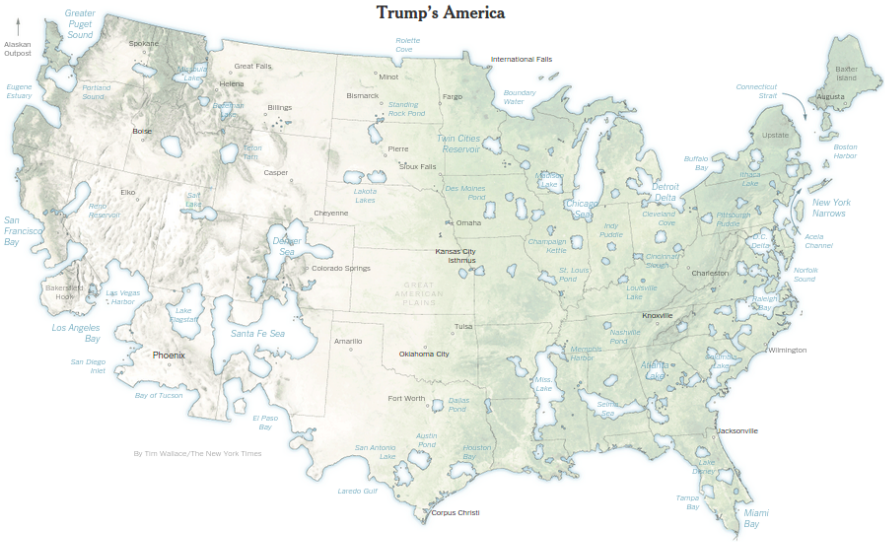
The Two Americas of 2016

By TIM WALLACE NOV. 16, 2016

For many Americans, it feels as if the 2016 election split the country in two.

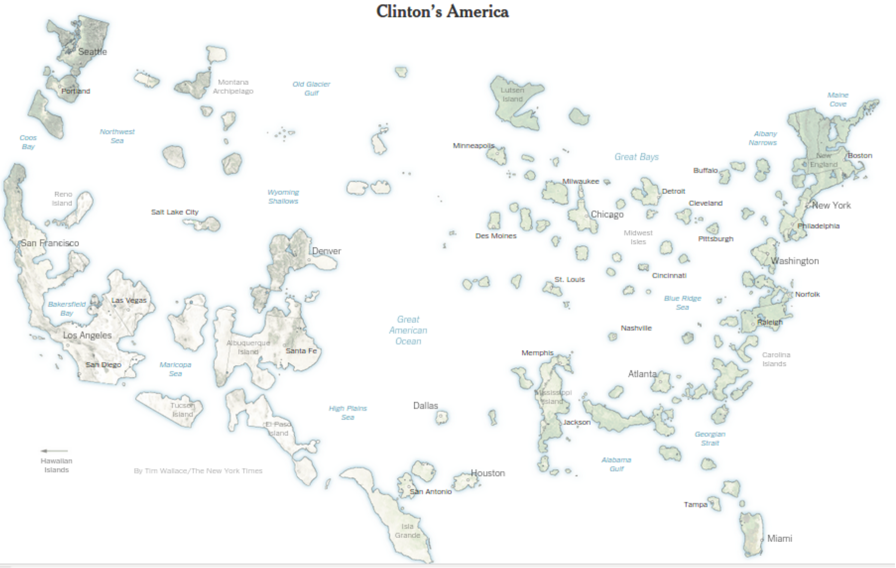
To visualize this, we took the election results and created two new imaginary nations by slicing the country along the sharp divide between Republican and Democratic Americas.

Figure A.39: The Two Americas



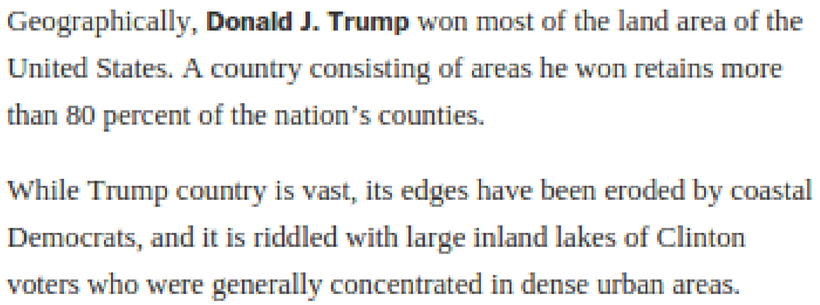
By Tim Wallace/The New York Times

Figure A.40: The Two America: Trump.



By Tim Wallace/The New York Times

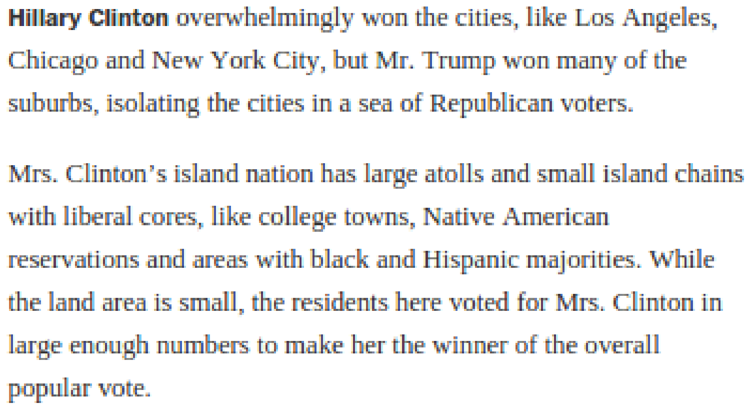
Figure A.41: The Two America: Cliton



Geographically, **Donald J. Trump** won most of the land area of the United States. A country consisting of areas he won retains more than 80 percent of the nation’s counties.

While Trump country is vast, its edges have been eroded by coastal Democrats, and it is riddled with large inland lakes of Clinton voters who were generally concentrated in dense urban areas.

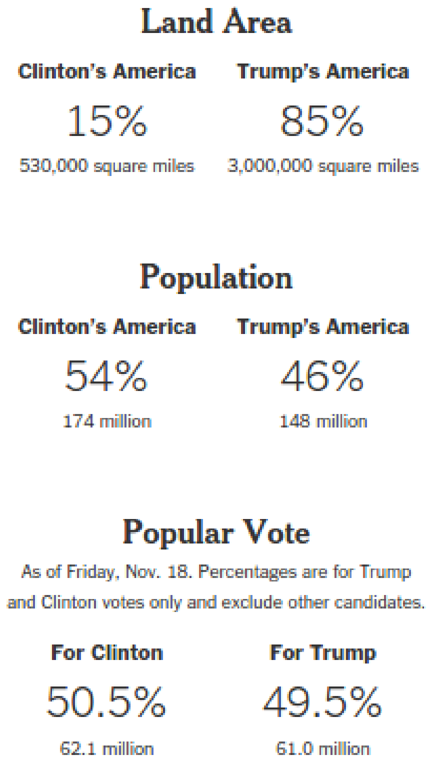
Figure A.42: The Two America: description of each side.



Hillary Clinton overwhelmingly won the cities, like Los Angeles, Chicago and New York City, but Mr. Trump won many of the suburbs, isolating the cities in a sea of Republican voters.

Mrs. Clinton’s island nation has large atolls and small island chains with liberal cores, like college towns, Native American reservations and areas with black and Hispanic majorities. While the land area is small, the residents here voted for Mrs. Clinton in large enough numbers to make her the winner of the overall popular vote.

Figure A.43: The Two America: description of Cliton side



Land Area

Clinton's America	Trump's America
15%	85%
530,000 square miles	3,000,000 square miles

Population

Clinton's America	Trump's America
54%	46%
174 million	148 million

Popular Vote

As of Friday, Nov. 18. Percentages are for Trump and Clinton votes only and exclude other candidates.

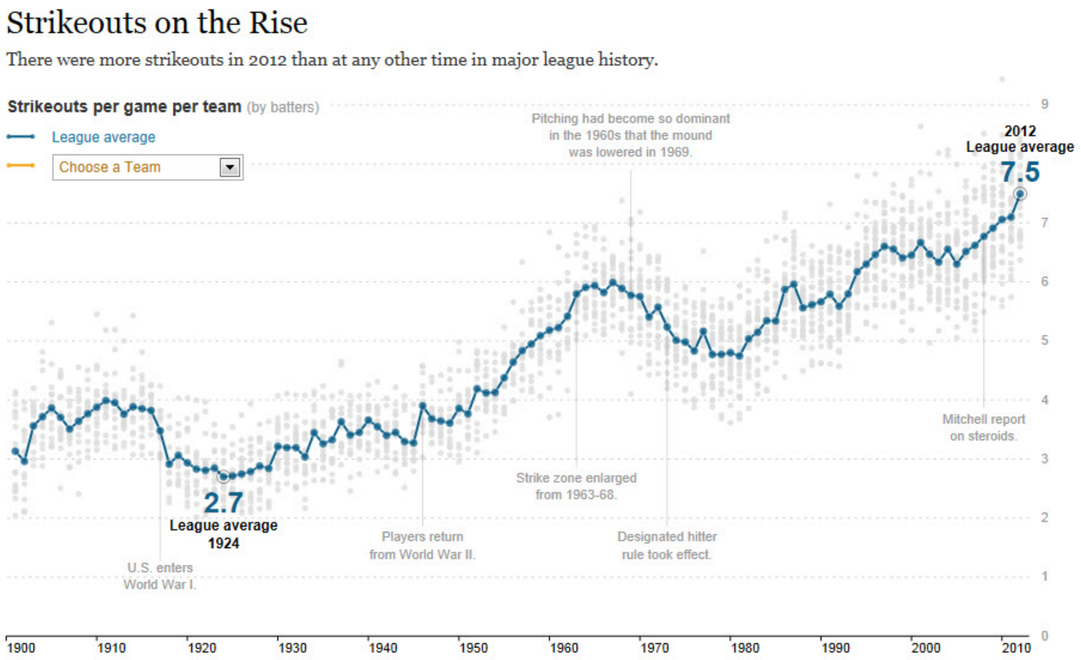
For Clinton	For Trump
50.5%	49.5%
62.1 million	61.0 million

Figure A.44: The Two America: comparison of two sides.

Informal classification:

As a published article that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium.* The *visual components* used include full- text, photographics, and data visualizations. The data visualization in this article is *basically interactive* in that it supports hovering, clicking, and other ways from the data visualization. The viewing sequence is in parallel. However, it is also *replicable*, as it is stored and can be played back at any time.

A.3.2 Strikeouts on the Rise


Strikeouts on the Rise

There were more strikeouts in 2012 than at any other time in major league history.

Pitching had become so dominant in the 1960s that the mound was lowered in 1969.

U.S. enters World War I.

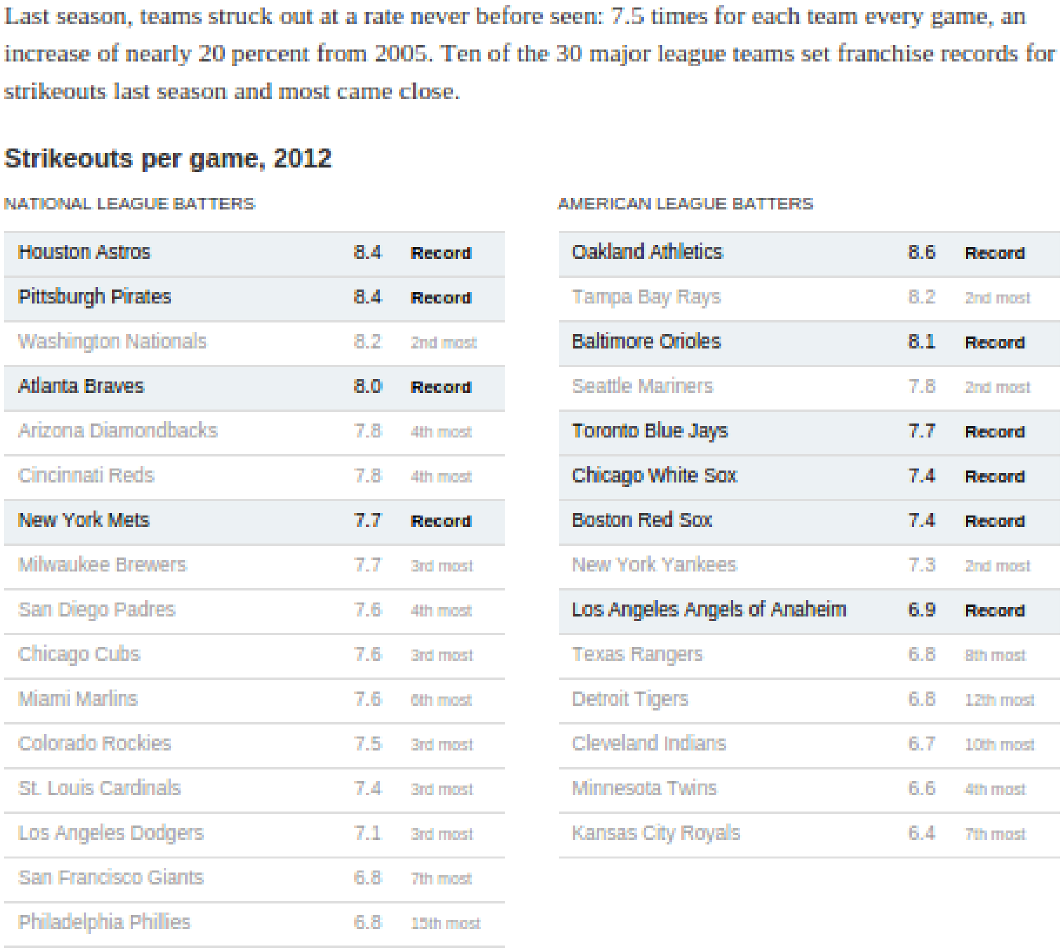
Players return from World War II.

Strike zone enlarged from 1963-68.

Designated hitter rule took effect.

Mitchell report on steroids.

Figure A.45: Strikeouts on the rise



Last season, teams struck out at a rate never before seen: 7.5 times for each team every game, an increase of nearly 20 percent from 2005. Ten of the 30 major league teams set franchise records for strikeouts last season and most came close.

Strikeouts per game, 2012

NATIONAL LEAGUE BATTERS

Houston Astros	8.4	Record
Pittsburgh Pirates	8.4	Record
Washington Nationals	8.2	2nd most
Atlanta Braves	8.0	Record
Arizona Diamondbacks	7.8	4th most
Cincinnati Reds	7.8	4th most
New York Mets	7.7	Record
Milwaukee Brewers	7.7	3rd most
San Diego Padres	7.6	4th most
Chicago Cubs	7.6	3rd most
Miami Marlins	7.6	6th most
Colorado Rockies	7.5	3rd most
St. Louis Cardinals	7.4	3rd most
Los Angeles Dodgers	7.1	3rd most
San Francisco Giants	6.8	7th most
Philadelphia Phillies	6.8	15th most

AMERICAN LEAGUE BATTERS

Oakland Athletics	8.6	Record
Tampa Bay Rays	8.2	2nd most
Baltimore Orioles	8.1	Record
Seattle Mariners	7.8	2nd most
Toronto Blue Jays	7.7	Record
Chicago White Sox	7.4	Record
Boston Red Sox	7.4	Record
New York Yankees	7.3	2nd most
Los Angeles Angels of Anaheim	6.9	Record
Texas Rangers	6.8	8th most
Detroit Tigers	6.8	12th most
Cleveland Indians	6.7	10th most
Minnesota Twins	6.6	4th most
Kansas City Royals	6.4	7th most

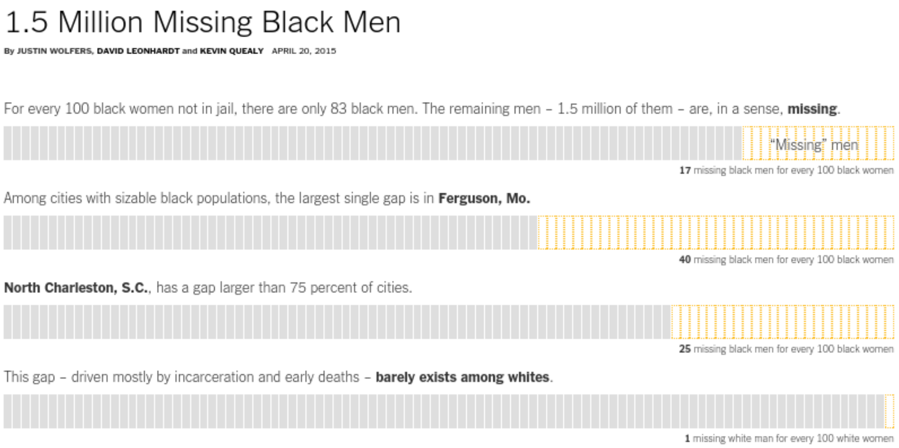
Figure A.46: Strikeouts on the rise [38] for each player.

Figure. (A.45, A.46) show that the visualization can be interactive and more informative. This story shows the trend of strikeouts is on the rise for the league in 2012. The line chart combined with scatter plot enables the audience to hover each point and see the accurate data for that time. For more detailed information, a table with hyper-link is provided for each team. Here I only show part of the web page but the audience can easily draw a story with detailed information from the visualization, text and table.

Informal classification:

As a published article that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*. The *visual components* used include full- text, photographics, and data visualizations. The data visualization in this article is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.3.3 1.5 Million Missing Black Men



1.5 Million Missing Black Men

By **JUSTIN WOLFERS**, **DAVID LEONHARDT** and **KEVIN QUEALY** APRIL 20, 2015

For every 100 black women not in jail, there are only 83 black men. The remaining men – 1.5 million of them – are, in a sense, **missing**.

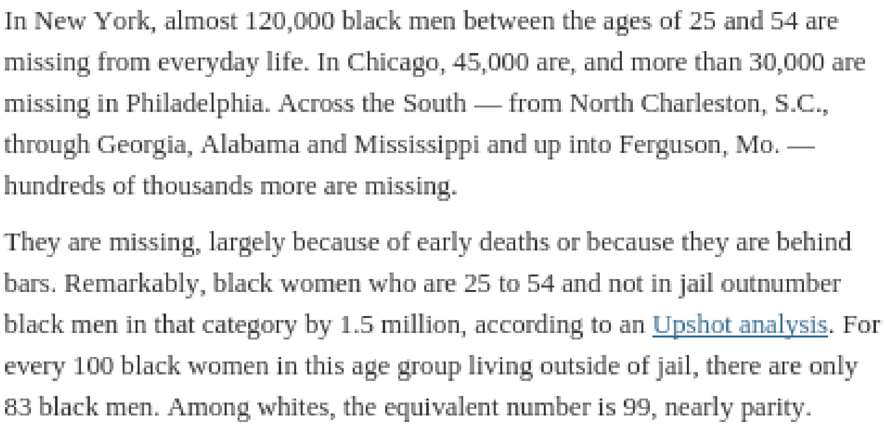
Among cities with sizable black populations, the largest single gap is in **Ferguson, Mo.**

North Charleston, S.C., has a gap larger than 75 percent of cities.

This gap – driven mostly by incarceration and early deaths – **barely exists among whites**.

"Missing" men
17 missing black men for every 100 black women
40 missing black men for every 100 black women
25 missing black men for every 100 black women
1 missing white man for every 100 white women

Figure A.47: Missing men for different races [39]



In New York, almost 120,000 black men between the ages of 25 and 54 are missing from everyday life. In Chicago, 45,000 are, and more than 30,000 are missing in Philadelphia. Across the South — from North Charleston, S.C., through Georgia, Alabama and Mississippi and up into Ferguson, Mo. — hundreds of thousands more are missing.

They are missing, largely because of early deaths or because they are behind bars. Remarkably, black women who are 25 to 54 and not in jail outnumber black men in that category by 1.5 million, according to an Upshot analysis. For every 100 black women in this age group living outside of jail, there are only 83 black men. Among whites, the equivalent number is 99, nearly parity.

Figure A.48: Which areas have men missing

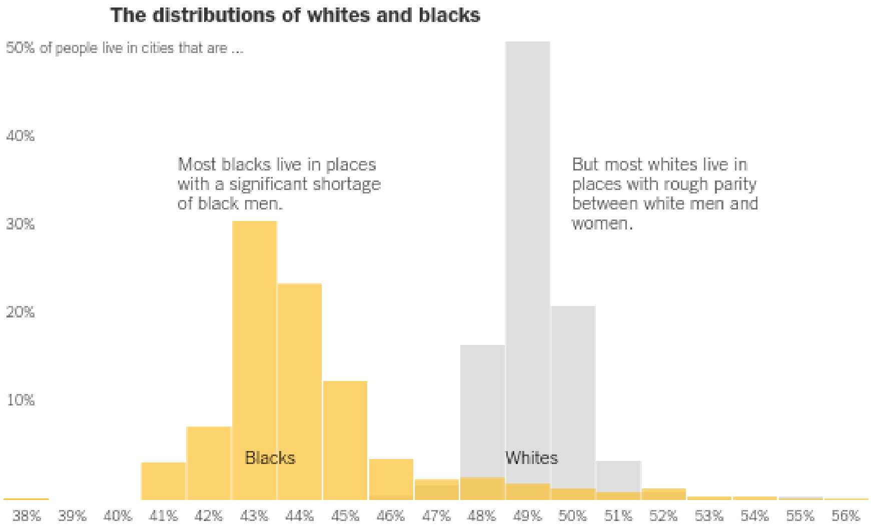


Figure A.49: Missing men for blacks and whites.

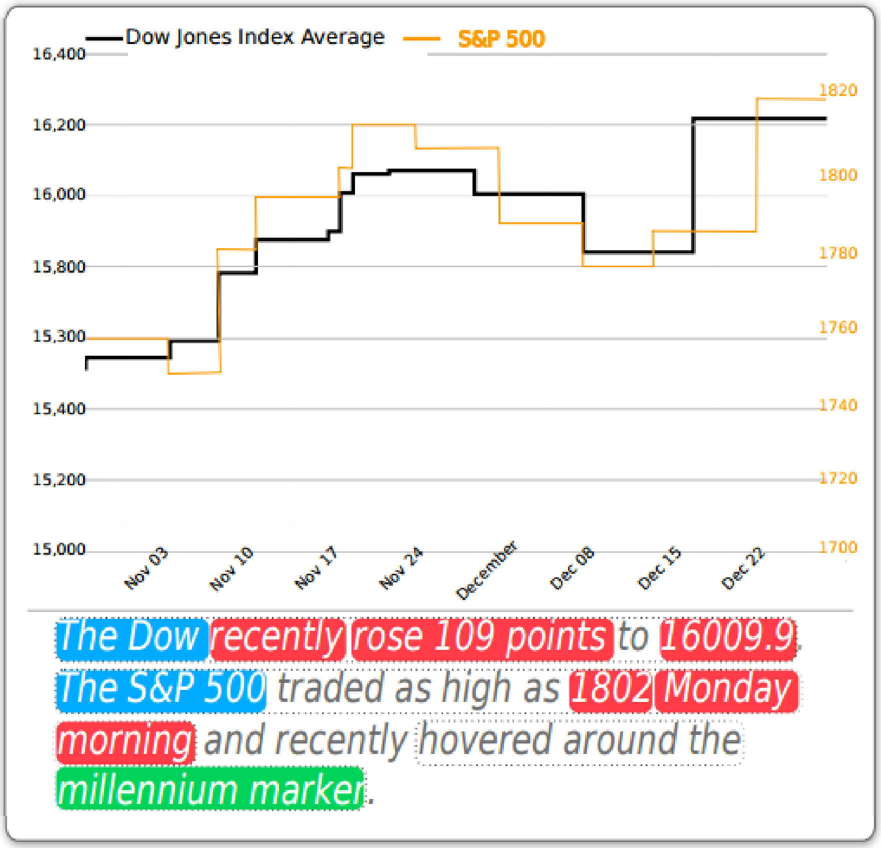
The example in Figure.(A.47, A.48, A.48) shows the black men missing for every one hundred population. The story is that more black men are missing then other racess.

Informal classification:

As a published article that is *shared on-line*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*. The *visual components* used include full- text, photographics, and data visualizations. The data visualization in this article is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.4 Visualization Tools

A.4.1 VisJocky



The Dow recently rose 109 points to 16009.9. The S&P 500 traded as high as 1802 Monday morning and recently hovered around the millennium marker.

Figure A.50: VisJocky interface [40]

Figure. A.50 is to simply visualize and add annotation to data to form a story. The tool creates a basic line chart with highlighted text annotation as description. The original data is a table of Dow Jones Index average and S&P 500 index for a few months. The story is that the Dow index loses 109 points and S&P index surged.

Informal classification:

As a visualization that is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has basic *controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization in this article is *interactive* in that it supports hovering, clicking, and other ways from the data visualization. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.4.2 ChartAccent

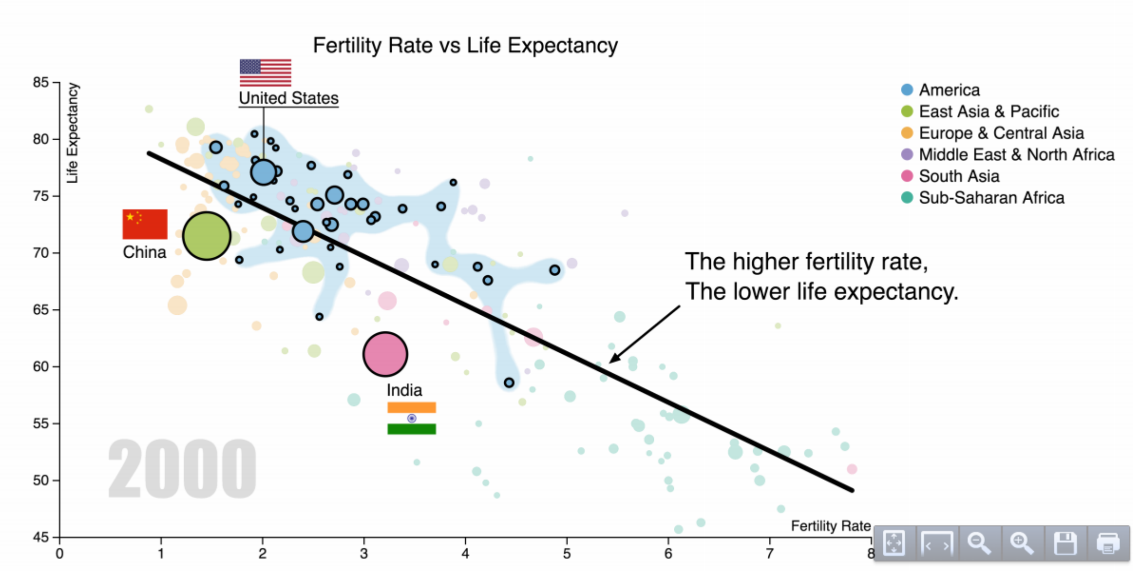


Figure A.51: ChartAccentc interface [41]

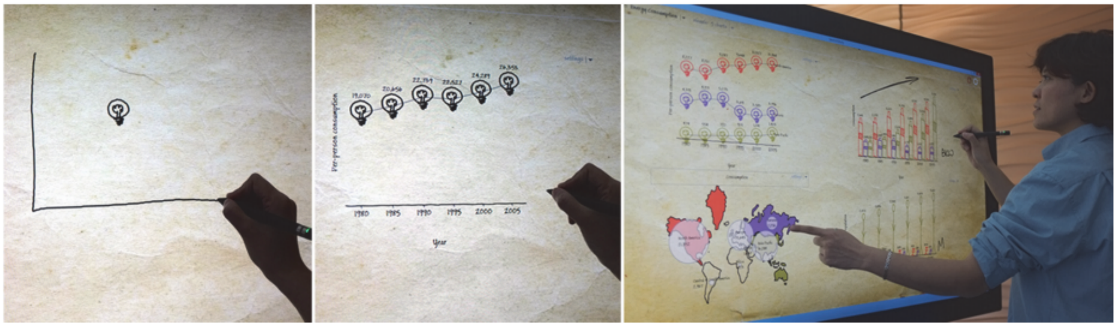
This tool shown in Figure. A.51 is used to add additional annotation and highlighting to existing data visualization. The blue area connecting United States and other small dots shows a subset of countries that are located in North and South America. Data: table of countries with their life expectancy and fertility rate. Story: The story of this added annotation is to show a group of North and South America countries about the distribution of their life expectancy and fertility rate. The connected highlighted area is a clear indication that the distribution of America countries are spreaded very sparse.

Informal classification:

As a visualization that is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has basic *controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization in this article is *interactive* in that it supports hovering, clicking, and other ways from the data visualization. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.5 Sketch Tools

A.5.1 SketchStory



(a) (b) (c)

Figure A.52: Process of design story with SketchStory

How to tell a story using SketchStory [171]: The presenter can start from (a) sketching icon and chart axis as example, (b) let SketchStory system finish the rest of the chart by combining sketches and underlying data, to (c) the produced media, whose chart can be interacted with presenter.

Informal classification:

As a sketching tool that is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has basic *controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *interactive* in that it supports hovering, clicking, and other ways from the data visualization. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

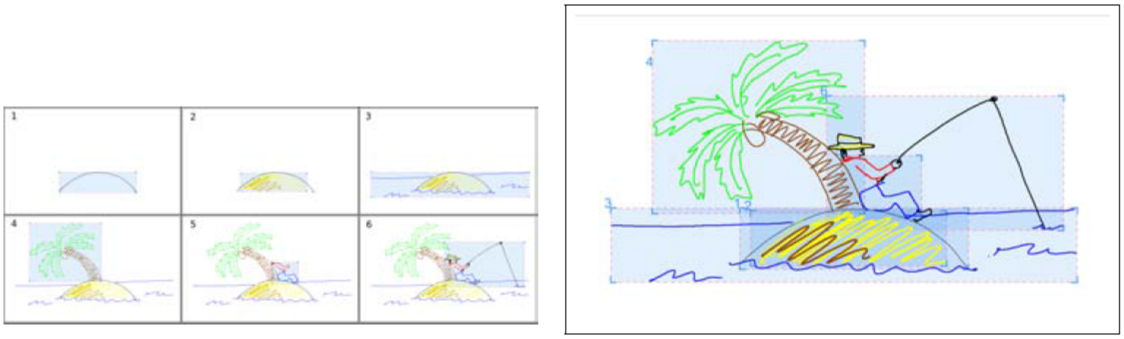


Figure A.53: Sketcholution comic strip and summary. [42]

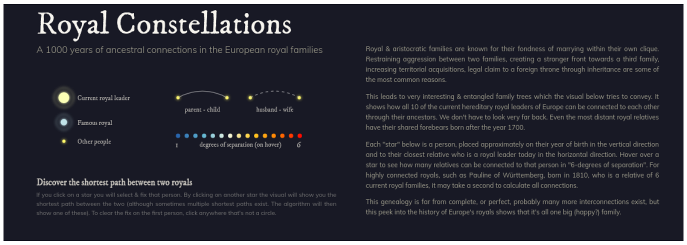
A.5.2 Sketcholution

Sketcholution is a method to create visual histories of hand sketches automatically. The resulting aggregation dendrogram in Figure. A.53 is able be adjust at any level based on display space. It can also be used to create a visual history in two styles including comic-strip for highlighting differences and a single summary frame annotating each object in the sketch scene.

Informal classification:

As a sketching tool that is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has basic *controls*. The *visual components* used include full-text, photographics, and data visualizations. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.5.3 DataSketches– Royal Constellations



Royal Constellations

A 1000 years of ancestral connections in the European royal families

Discover the shortest path between two royals

If you click on a star you will select & fix that person. By clicking on another star the visual will show you the shortest path between the two (although sometimes multiple shortest paths exist. The algorithm will then show one of these). To clear the fix on the first person, click anywhere that's not a circle.

Royal & aristocratic families are known for their fondness of marrying within their own clique. Restraining aggression between two families, creating a stronger front towards a third family, increasing territorial acquisitions, legal claim to a foreign throne through inheritance are some of the most common reasons.

This leads to very interesting & entangled family trees which the visual below tries to convey. It shows how all 10 of the current hereditary royal leaders of Europe can be connected to each other through their ancestors. We don't have to look very far back. Even the most distant royal relatives have their shared forebears born after the year 1700.

Each "star" below is a person, placed approximately on their year of birth in the vertical direction and to their closest relative who is a royal leader today in the horizontal direction. Hover over a star to see how many relatives can be connected to that person in "6-degrees of separation". For highly connected royals, such as Pauline of Württemberg, born in 1810, who is a relative of 6 current royal families, it may take a second to calculate all connections.

This genealogy is far from complete, or perfect, probably many more interconnections exist, but this peek into the history of Europe's royals shows that it's all one big (happy?) family.

Figure A.54: The sketch visualization of the royal members [43].

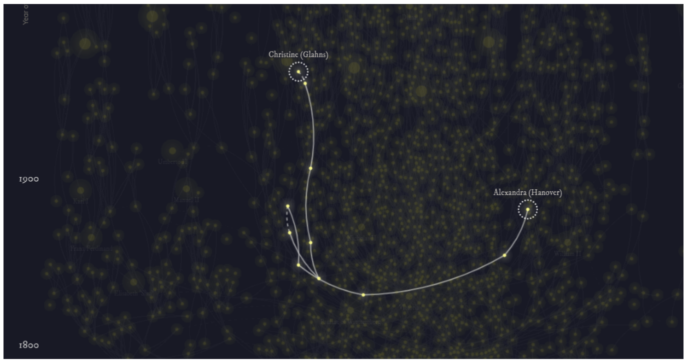


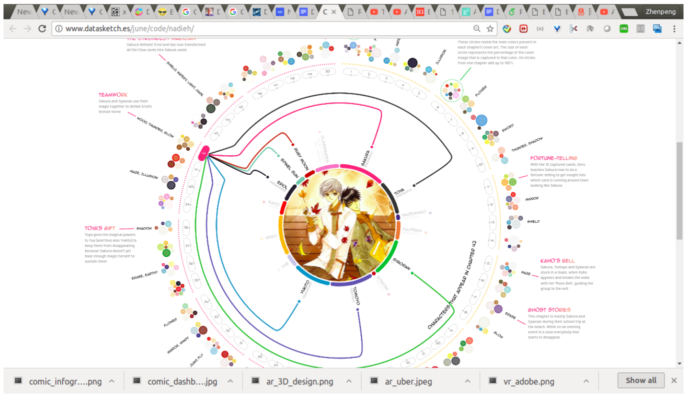
Figure A.55: The sketch connects the royal members

The example shows that a sketch based visualization tool is a method to create visual demonstration of the relationship of a 1000 years of ancestral connections in the European royal families. The resulting animated sketches in Figure. A.54 is to show the relationships between two royal figures by sketches. The story also includes the annotation from the storyteller. The data is the information of the relationship between royal family members.

Informal classification:

As a sketching tool that is used *individually*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *synchronous*. Data visualization is a *low-bandwidth medium*, but the user has basic *controls*. The *visual components* used include full- text, sketches, and data visualizations. The data visualization produced by this tool is *interactive*. The viewing sequence is in *sequence* or *branch*. However, it is also *ephemeral*, as it is stored and can be played back at any time.

A.5.4 DataSketches–Carcaptor Sakura



www.datasketch.es/june/code/nadieh/

Sakura 'defeats' Eriol and has now transformed all the Clow cards into Sakura cards

These circles reveal the main colors present in each chapter's cover art. The size of each circle represents the percentage of the cover image that is captured in that color. All circles from one chapter add up to 100%

TEAMWORK
Sakura and Syaoran use their magic together to defeat Eriol's bronze horse

FORTUNE-TELLING
With the 10 captured cards, Kero teaches Sakura how to do a fortune-telling to get insight into which card is running around town looking like Sakura

TOYA'S GIFT
Toya gives his magical powers to Yue (and thus also Yukito) to keep them from disappearing because Sakura doesn't yet have enough magic herself to sustain them

KAHO'S BELL
Sakura, Tomoyo and Syaoran are stuck in a maze, when Kaho appears and breaks the walls with her 'Moon Bell', guiding the group to the exit

GHOST STORIES
This chapter is mostly Sakura and Syaoran during their school trip at the beach. While on an evening event in a cave everybody else starts to disappear

Figure A.56: Visual Explanation of the Relationship of a Cartoon Series [44]

The example shows that a sketch based animation is a method to create visual explanation of the relationship of a cartoon series. The resulting animated sketches in Figure. A.56 is to show the relationships between selected cartoon figures by sketches. The story also includes the annotation from the storyteller. The data is the information of the relationship networks in the cartoon series.

Informal classification:

As a sketching tool that is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *synchronous*. Data visualization is a *low-bandwidth medium*, but the user has basic *controls*. The *visual components* used include full-text, sketches, and data visualizations. The data visualization produced by this tool is *interactive*. The viewing sequence is in *sequence* or *branch*. However, it is also *ephemeral*, as it is stored and can be played back at any time.

A.5.5 The Big Short Movie Explained Animated

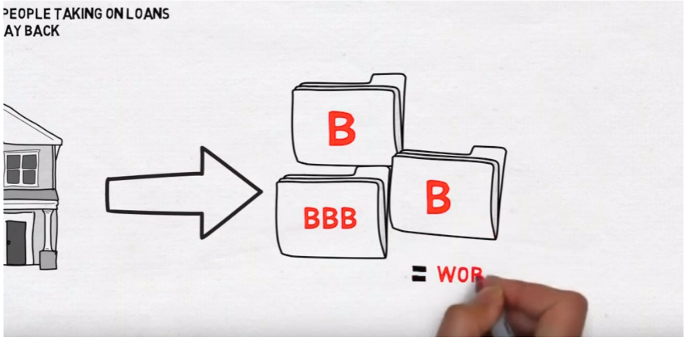


Figure A.57: How mortgage bond combined into sub-prime mortgage [45].

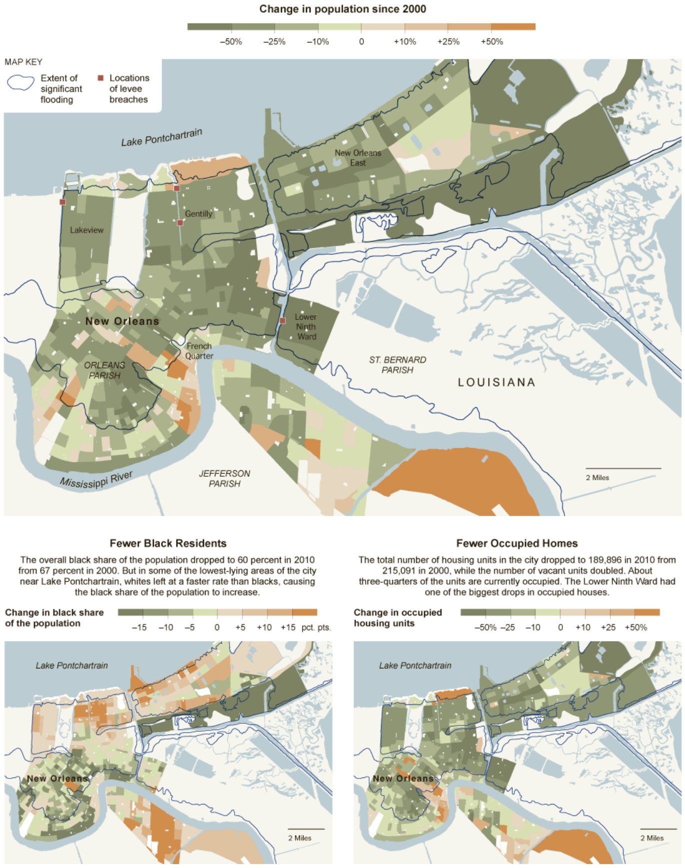
The example shows that a sketch based animation is a method to create visual explanation of a movie story line. The resulting animated sketches in Figure. A.57 is to show the story of the movie into steps. The story also includes the annotation and the personal understanding of the movie from the storyteller. The data is the information of the financial product in the movie.

Informal classification:

As a sketching tool that is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has no *controls*. The *visual components* used include full- text, sketches, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6 Infographics

Infographics is another example of storytelling media that combines a few basic forms of media to interpret a data story. Most infographics involve certain forms of data visualization and text explanation. Infographics are very widely used from entertainment to education purpose. The data for the story is mostly represented by data visualization and tables.



Fewer Black Residents

The overall black share of the population dropped to 60 percent in 2010 from 67 percent in 2000. But in some of the lowest-lying areas of the city near Lake Pontchartrain, whites left at a faster rate than blacks, causing the black share of the population to increase.

Fewer Occupied Homes

The total number of housing units in the city dropped to 189,896 in 2010 from 215,091 in 2000, while the number of vacant units doubled. About three-quarters of the units are currently occupied. The Lower Ninth Ward had one of the biggest drops in occupied houses.

Figure A.58: New Orleans housing population decreases [46]

A.6.1 New Orleans Housing Population

For the example in Figure. A.58, the story is about he population change of New Orleans. It clearly shows a trend that the coast area and downtown of New Orleans has an increasing black share of the population, as well as a decreasing occupation of the house units. The story is both revealed by the heat map and text description. Most of the time, the infographics is not meant to be interactive, the story is well organized by the positioning and annotation of the data visualization.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.2 Top writers for best sellers

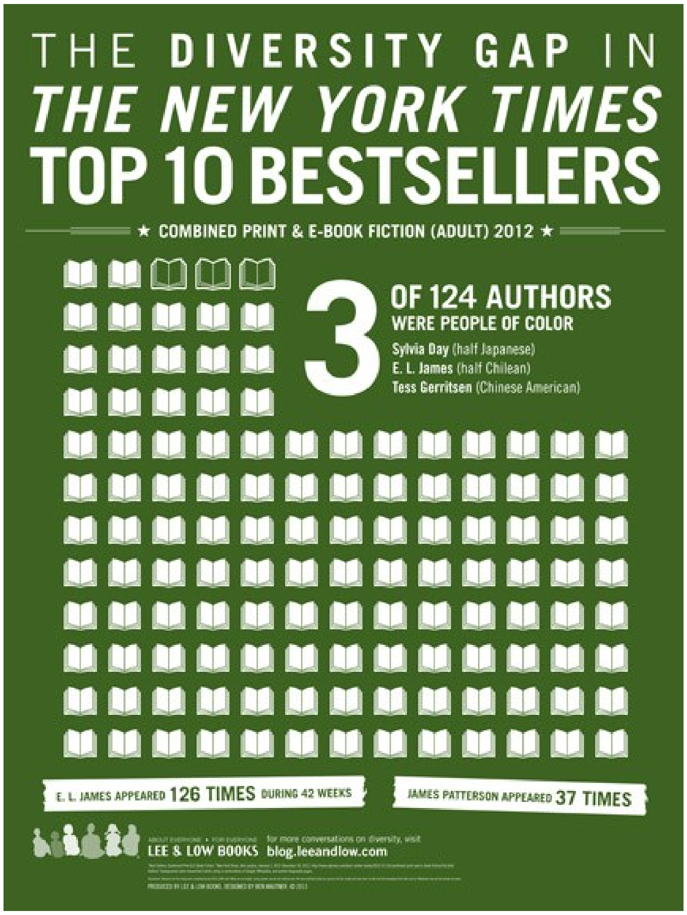


Figure A.59: Top writers for best sellers [47]

This infographic in Figure. A.59 is to show that very few colored writers were ranked top bestsellers. Data is the number of colored writers and two top sellers ranked the most times. The story is the writers of the bestsellers do not have very good diversity.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.3 Public Library Report

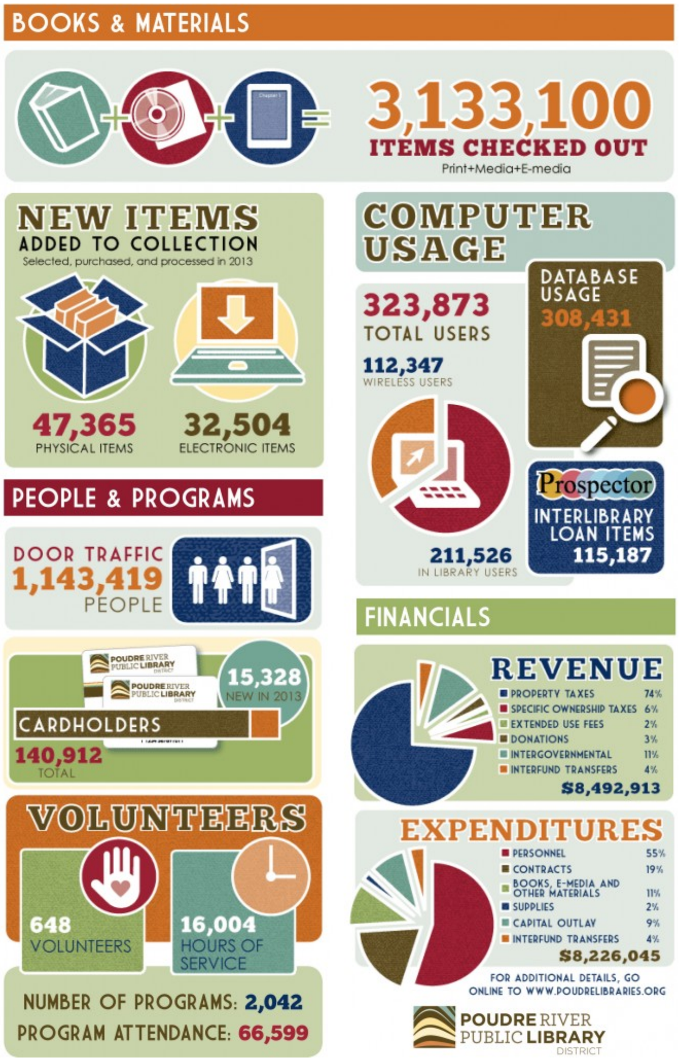


Figure A.60: Public library report [48]

The infographics in Figure. A.60 is show the managing status for a public library. It includes data of people, inventory and financials. Data: The original data is a data sheet of the running conditions of public library Story: The story is that the business is running well. All major data including people, usage, products showed the library is in a good condition.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.4 US Poverty

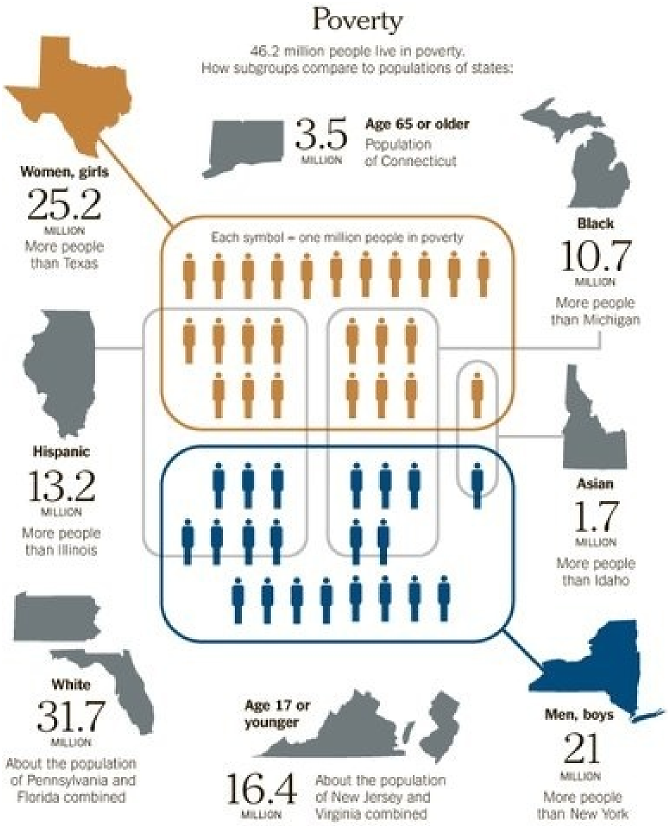


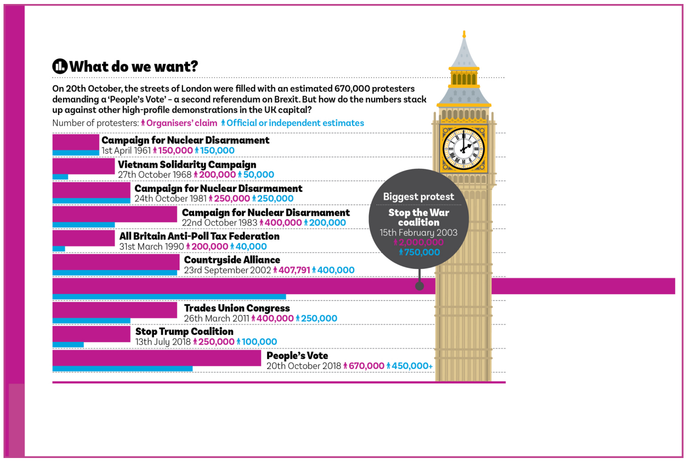
Figure A.61: How poverty is distributed in the U.S. [49]

The infographics in Figure. A.61 shows the percentage of poor people and the ethic distribution around the US. Data: The distribution of poor people. Story: The way the poor people are distributed among subgroups.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth* *medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.5 London March



What do we want?

On 20th October, the streets of London were filled with an estimated 670,000 protesters demanding a 'People's Vote' – a second referendum on Brexit. But how do the numbers stack up against other high-profile demonstrations in the UK capital?

Figure A.62: Distribution of London Marches in Visualization [50]

The infographics in Figure. A.62 shows the distribution of features of marches happened in London. Data: The distribution of Marches. Story: The way the marches in London are distributed for their time and size.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-

text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.6 Yemen War

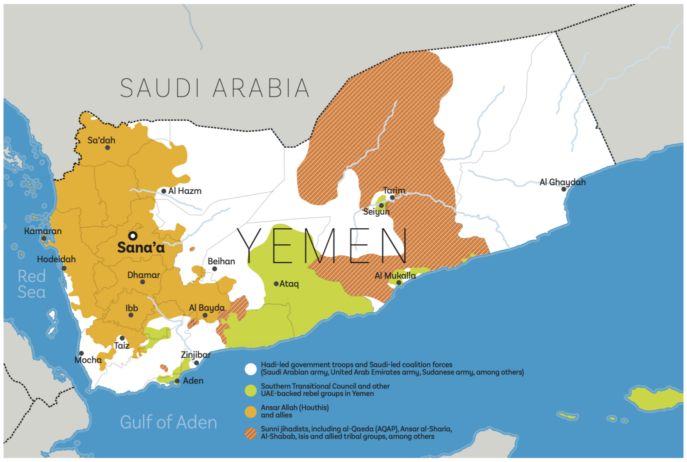

SAUDI ARABIA
Sa'dah
Al Hazm
Kamaran
Sana'a
Hodeidah
Red Sea
Dhamar
Ibb
Al Bayda
Beihan
YEMEN
Ataq
Tarim
Seiyun
Al Ghaydah
Al Mukalla
Taiz
Mocha
Zinjibar
Aden
Gulf of Aden
○ Hadi-led government troops and Saudi-led coalition forces (Saudi Arabian army, United Arab Emirates army, Sudanese army, among others)
● Southern Transitional Council and other UAE-backed rebel groups in Yemen
● Ansar Allah (Houthis) and allies
▨ Sunni jihadists, including al-Qaeda (AQAP), Ansar al-Sharia, Al-Shabab, Isis and allied tribal groups, among others

Figure A.63: Geo-distribution of All the Forces in Yemen Civil War in Visualization [51]

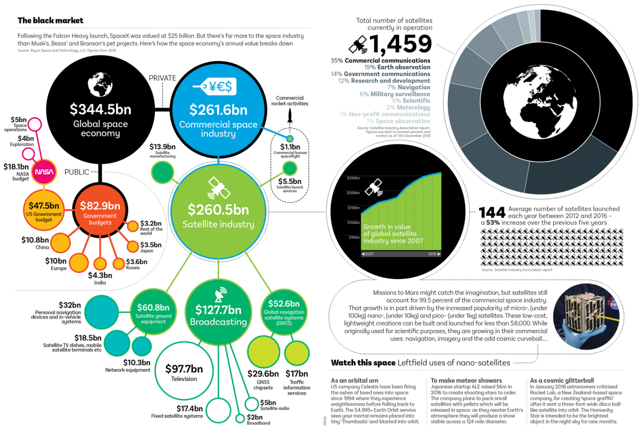
The infographics in Figure. A.63 shows the distribution of all forces' position of Yemen Civil War. Data: The distribution of all forces. Story: The way all different forces are distributed and fight against each other.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-

text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.7 Space Industry



The black market

Following the Falcon Heavy launch, SpaceX was valued at $25 billion. But there's far more to the space industry than Musk's, Bezos' and Branson's pet projects. Here's how the space economy's annual value breaks down

Source: Bryce Space and Technology, LLC, figures from 2016

Source: Satellite Industry Association report, figures rounded to nearest percent and correct as of 31st December 2016

144 Average number of satellites launched each year between 2012 and 2016 – a **53%** increase over the previous five years

Source: Satellite Industry Accociation report

Missions to Mars might catch the imagination, but satellites still account for 99.5 percent of the commercial space industry. That growth is in part driven by the increased popularity of micro-, (under 100kg) nano-, (under 10kg) and pico- (under 1kg) satellites. These low-cost, lightweight creations can be built and launched for less than $8,000. While originally used for scientific purposes, they are growing in their commercial uses: navigation, imagery and the odd cosmic curveball....

Watch this space Leftfield uses of nano-satellites

As an orbital urn
US company Celestis have been firing the ashes of loved ones into space since 1994 where they experience weightlessness before falling back to Earth. The $4,995+ Earth Orbit service sees your mortal remains placed into tiny 'Thumbsats' and blasted into orbit.

To make meteor showers
Japanese startup ALE raised $6m in 2016 to create shooting stars to order. The company plans to pack small satellites with pellets which will be released in space: as they reenter Earth's atmosphere they will produce a show visible across a 124 mile diameter.

As a cosmic glitterball
In January 2018 astronomers criticised Rocket Lab, a New Zealand-based space company, for creating 'space graffiti' after it sent a three-foot-wide disco ball-like satellite into orbit. The Humanity Star is intended to be the brightest object in the night sky for nine months.

Figure A.64: The Stats of Space Industry and Technology in Visualization [52]

The infographics in Figure. A.64 shows the distribution of all kinds of satellite launches in the space industry. Data: The distribution of cost of all kinds of commercial launches. Story: The way all different kinds of satellite launches including governmental and commercial purpose.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-

text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.8 Household Air Pollution

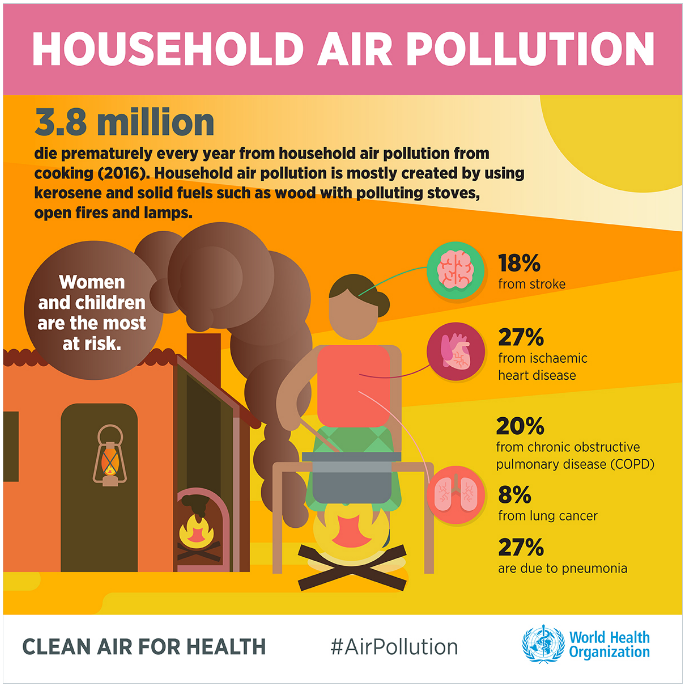

HOUSEHOLD AIR POLLUTION
3.8 million
die prematurely every year from household air pollution from cooking (2016). Household air pollution is mostly created by using kerosene and solid fuels such as wood with polluting stoves, open fires and lamps.
Women and children are the most at risk.
18%
from stroke
27%
from ischaemic heart disease
20%
from chronic obstructive pulmonary disease (COPD)
8%
from lung cancer
27%
are due to pneumonia
CLEAN AIR FOR HEALTH
#AirPollution
World Health Organization

Figure A.65: The Source of Household Air Pollution in Visualization [53]

The infographics in Figure. A.65 shows the death caused by household air pollution. Data: The distribution of all kinds of death caused by household air pollution. Story: The household air pollution cause all kinds of death.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.9 Air Pollution Linked Death

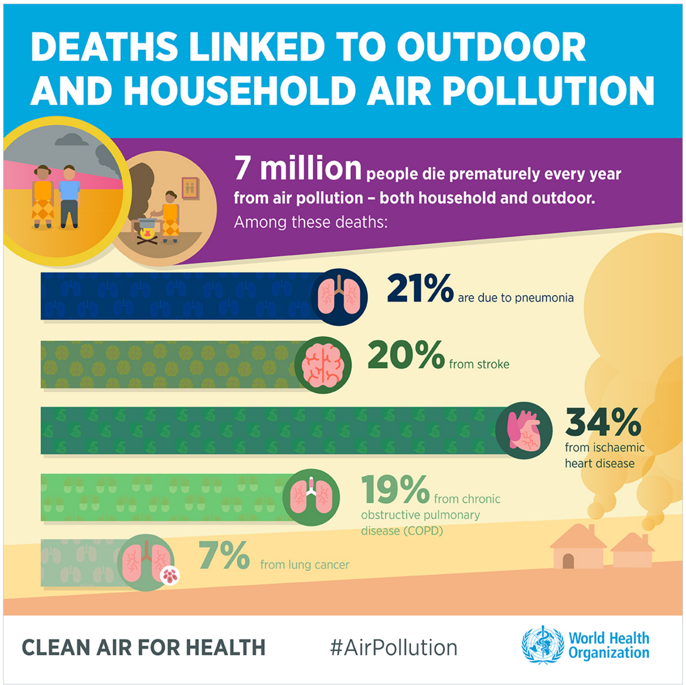


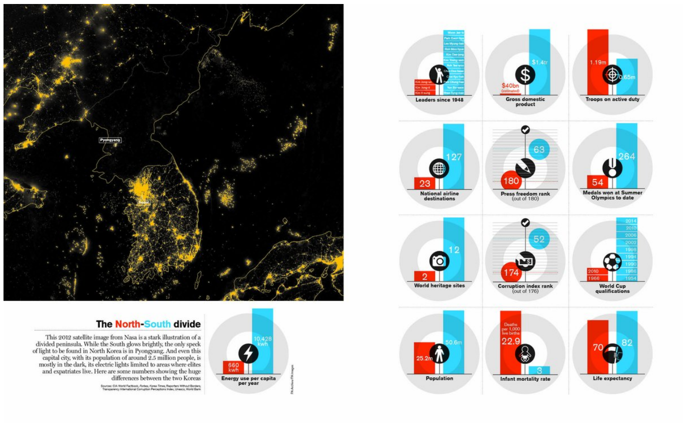
Figure A.66: The Source of Death Caused by Air Pollution in Visualization [54]

The infographics in Figure. A.66 shows the death caused by air pollution. Data: The distribution of all kinds of death caused by air pollution. Story: The death can be caused by household and outdoor air pollution.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.10 North and South Korean Comparison



This 2012 satellite image from Nasa is a stark illustration of a divided peninsula. While the South glows brightly, the only speck of light to be found in North Korea is in Pyongyang. And even this capital city, with its population of around 2.5 million people, is mostly in the dark, its electric lights limited to areas where elites and expatriates live. Here are some numbers showing the huge differences between the two Koreas

Figure A.67: The Comparison of North and South Korea in Visualization [55]

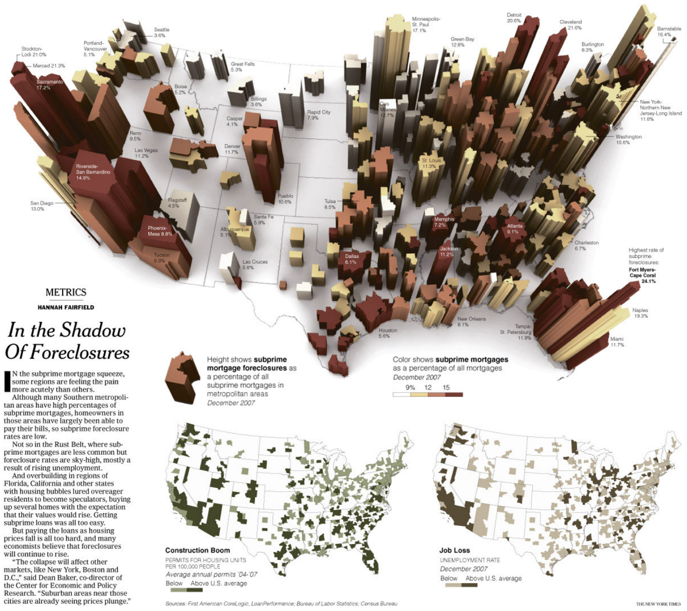
The infographics in Figure. A.67 shows the comparison of all aspects of north and south Korea. Data: The distribution of all kinds of death caused by air pollution. Story: The death can be caused by household and outdoor air pollution.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *per-*

sonal devices (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.11 In the Shadow of Foreclosure



METRICS

HANNAH FAIRFIELD

In the Shadow Of Foreclosures

IN the subprime mortgage squeeze, some regions are feeling the pain more acutely than others.

Although many Southern metropolitan areas have high percentages of subprime mortgages, homeowners in those areas have largely been able to pay their bills, so subprime foreclosure rates are low.

Not so in the Rust Belt, where subprime mortgages are less common but foreclosure rates are sky-high, mostly a result of rising unemployment.

And overbuilding in regions of Florida, California and other states with housing bubbles lured overeager residents to become speculators, buying up several homes with the expectation that their values would rise. Getting subprime loans was all too easy.

But paying the loans as housing prices fall is all too hard, and many economists believe that foreclosures will continue to rise.

“The collapse will affect other markets, like New York, Boston and D.C.,” said Dean Baker, co-director of the Center for Economic and Policy Research. “Suburban areas near those cities are already seeing prices plunge.”

Sources: First American CoreLogic, LoanPerformance; Bureau of Labor Statistics; Census Bureau

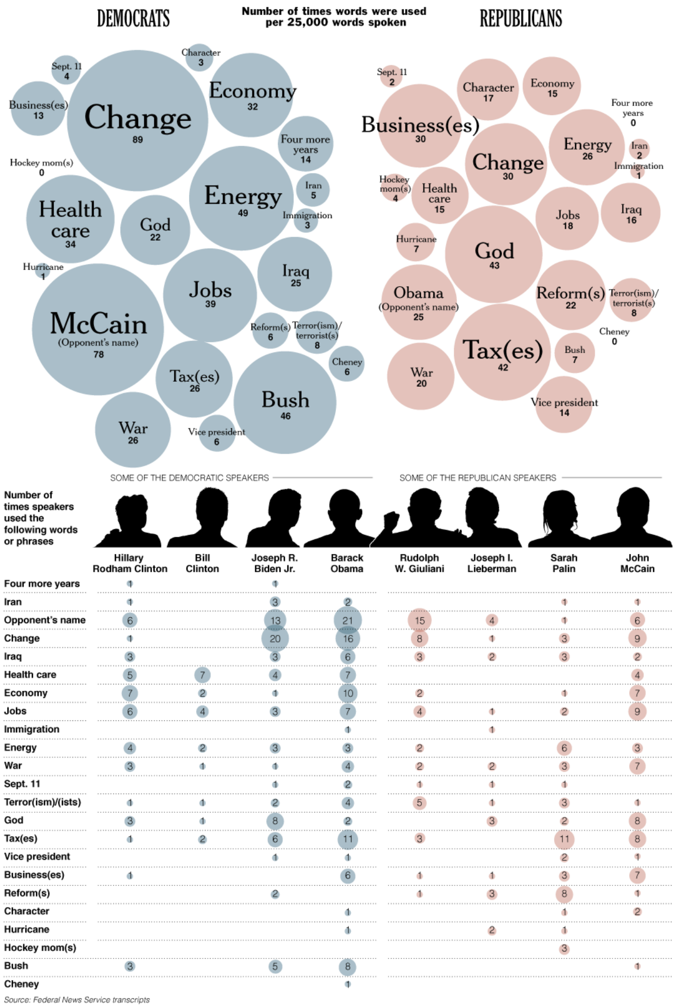
Figure A.68: The situation of foreclosure across the U.S. [56]

The infographics in Figure. A.68 shows the distribution of foreclosure percent. The west coast and big Florida area are the heavily influenced locations. Data: The percent of houses foreclosed in each state. Story: Giving a impression the level of pain across US due to the sub-prime mortgage crisis.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.12 Word of Democrats and Republicans



Number of times speakers used the following words or phrases	Hillary Rodham Clinton	Bill Clinton	Joseph R. Biden Jr.	Barack Obama	Rudolph W. Giuliani	Joseph I. Lieberman	Sarah Palin	John McCain
Four more years	1		1					
Iran	1		3	2			1	1
Opponent's name	6		13	21	15	4	1	6
Change	1		20	16	8	1	3	9
Iraq	3		3	6	3	2	3	2
Health care	5	7	4	7				4
Economy	7	2	1	10	2		1	7
Jobs	6	4	3	7	4	1	2	9
Immigration				1		1		
Energy	4	2	3	3	2		6	3
War	3	1	1	4	2	2	3	7
Sept. 11			1	2	1	1	1	
Terror(ism)/(ists)	1	1	2	4	5	1	3	1
God	3	1	8	2		3	2	8
Tax(es)	1	2	6	11	3		11	8
Vice president			1	1			2	1
Business(es)	1			6	1	1	3	7
Reform(s)			2		1	3	8	1
Character				1			1	2
Hurricane				1		2	1	
Hockey mom(s)							3	
Bush	3		5	8				1
Cheney				1				

Source: Federal News Service transcripts

Figure A.69: The comparison for the Democrats and Republicans during election [57]

The infographics in Figure. A.69 shows the most frequent word the Democrats and Republicans usually say during the election Data: The heat map of different words and mention of countries. Story: The story is that the focus of Democrats and Republicans have very different focus of topics during election.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.13 UK and US Firearms

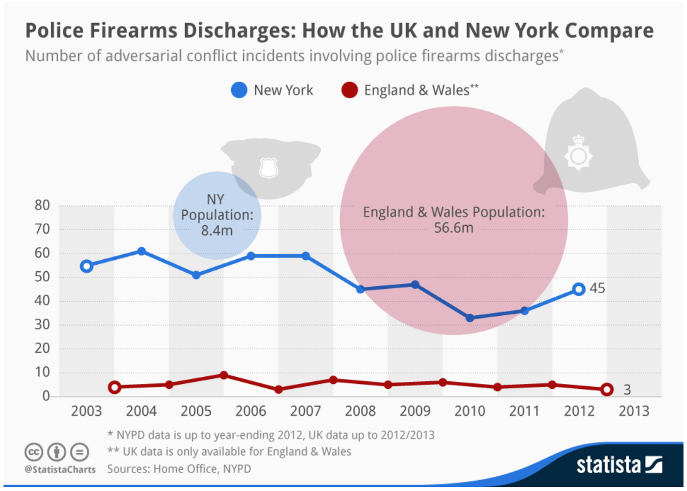


Figure A.70: The comparison of the firearm possession of the U.K. and the U.S. [58]

The infographics in Figure. A.70 the managing status for a public library. It includes data of people, inventory and financials. Data: The original data is a data sheet of the running conditions of public library Story: The story is that the business is running well. All major data including people, usage, products showed the library is in a good condition.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.14 White House Correspondent Dinner

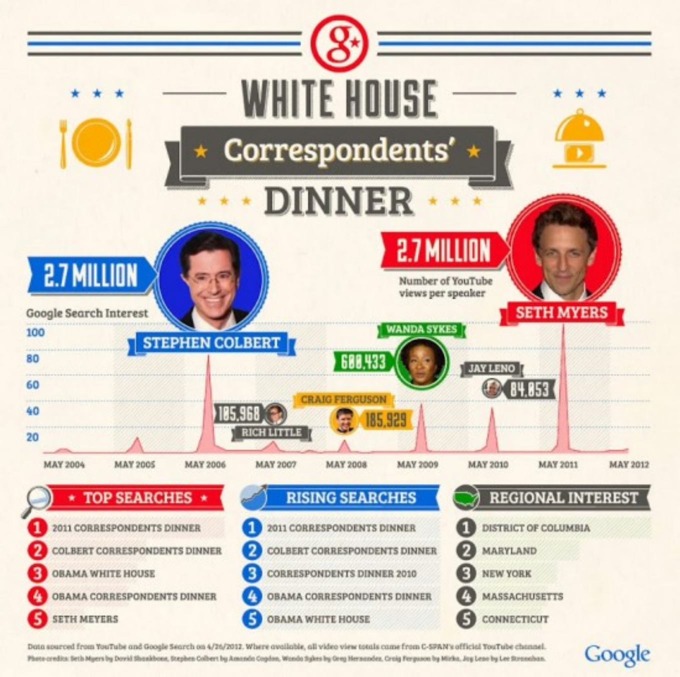


Figure A.71: The guest distribution of the White House Correspondent Dinner [59]

The infographics in Figure. A.71 shows the comparison among the invited to White House correspondent dinner. Data: the specs of different correspondents along the time. Story: The story is the trend which correspondent is more popular.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.15 Day vs. Night: What NYCs Population Looks Like

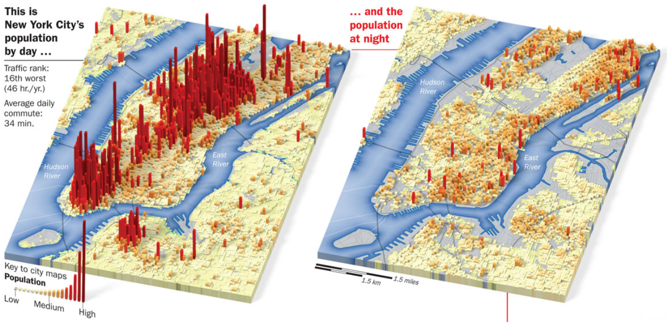

This is New York City's population by day ...
Traffic rank: 16th worst (46 hr./yr.)
Average daily commute: 34 min.
Hudson River
East River
Key to city maps
Population
Low
Medium
High
... and the population at night
Hudson River
East River
1.5 km
1.5 miles

Figure A.72: The change of population during days and nights [60]

The infographics in Figure. A.72 shows the different between day and night for population distribution in New York City. There are more people in midtown and downtown in the day time than in the night. Data: The actual distribution of population density in the day and night of New York City. Story: There are way more people in Manhattan in the daytime than in the night. The average commute time is 34 minutes.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.16 Who owns everything: Big Data Today

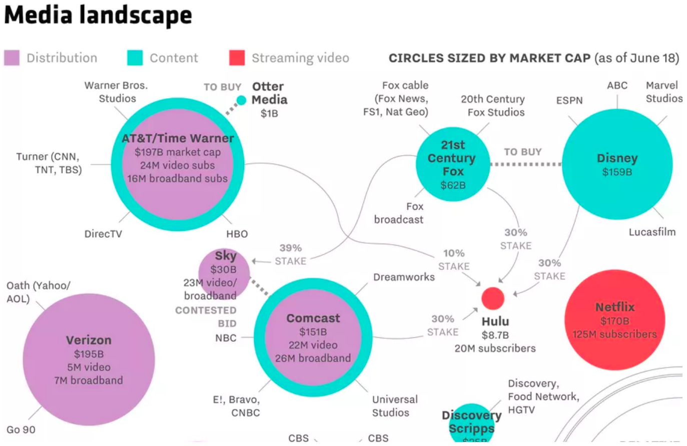


Figure A.73: The wealth distribution [61]

The infographics in Figure. A.73 shows the ownership between large cooperation in US. It includes data of the capital value of each company and the share of each company to other companies. Data: The ownership data between companies. Story: The story is how the giant cooperations have share from each other.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.17 Big Welsh Coast Walk

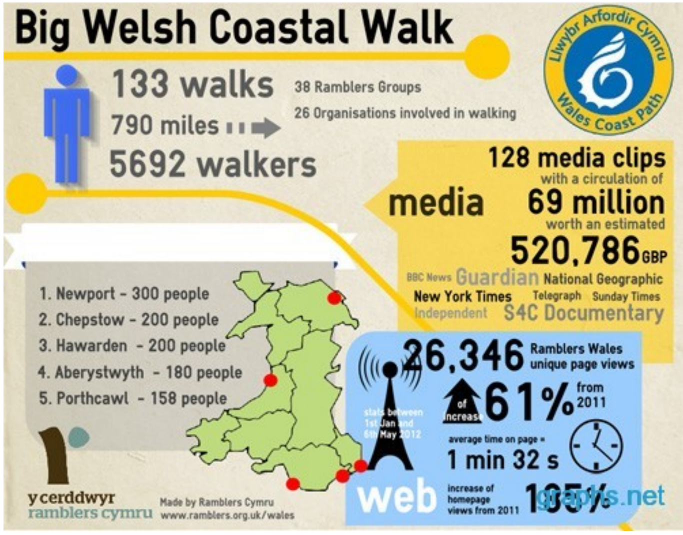


Figure A.74: The participants and locations of the big Welsh coast walk [62]

The infographics in Figure. A.74 shows the coast walk activity in Welsh. It includes people distribution among the cities and the money raised in total. Data: The original data is a data sheet of the running conditions of public library Story: The story is that the business is running well. All major data including people, usage, products showed the library is in a good condition.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.18 Hungry USA

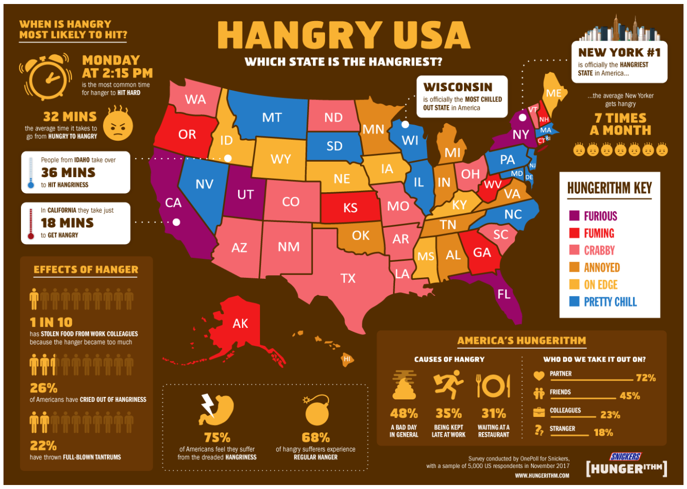

WHEN IS HANGRY MOST LIKELY TO HIT?
MONDAY AT 2:15 PM
is the most common time for hanger to HIT HARD
32 MINS
the average time it takes to go from HUNGRY TO HANGRY
People from IDAHO take over
36 MINS
to HIT HANGRINESS
In CALIFORNIA they take just
18 MINS
to GET HANGRY
EFFECTS OF HANGER
1 IN 10
has STOLEN FOOD FROM WORK COLLEAGUES because the hanger became too much
26%
of Americans have CRIED OUT OF HANGRINESS
22%
have thrown FULL-BLOWN TANTRUMS
HANGRY USA
WHICH STATE IS THE HANGRIEST?
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MI
IN
OH
KY
TN
MS
AL
GA
FL
SC
NC
VA
WV
PA
MD
DE
NJ
NY
VT
NH
ME
MA
CT
RI
AK
HI
WISCONSIN
is officially the MOST CHILLED OUT STATE in America
NEW YORK #1
is officially the HANGRIEST STATE in America...
...the average New Yorker gets hangry
7 TIMES A MONTH
HUNGERITHM KEY
FURIOUS
FUMING
CRABBY
ANNOYED
ON EDGE
PRETTY CHILL
75%
of Americans feel they suffer from the dreaded HANGRINESS
68%
of hangry sufferers experience REGULAR HANGER
AMERICA'S HUNGERITHM
CAUSES OF HANGRY
48%
A BAD DAY IN GENERAL
35%
BEING KEPT LATE AT WORK
31%
WAITING AT A RESTAURANT
WHO DO WE TAKE IT OUT ON?
PARTNER 72%
FRIENDS 45%
COLLEAGUES 23%
STRANGER 18%
Survey conducted by OnePoll for Snickers, with a sample of 5,000 US respondents in November 2017
WWW.HUNGERITHM.COM
SNICKERS
[HUNGERITHM]

Figure A.75: Public library report [63]

The infographics in Figure. A.75 shows hangry level of US states as a colored map. It includes the cause and effect of hangryness on average level Data: The level of hangryness across the US and the average data for different states. Story: The states of New York and California have the highest level of hangryness across the country. People in South Dakota and Illinois are pretty chill.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *low-bandwidth* *medium*, but the user has *no controls*. The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive*. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.6.19 NYC Celebrity Map

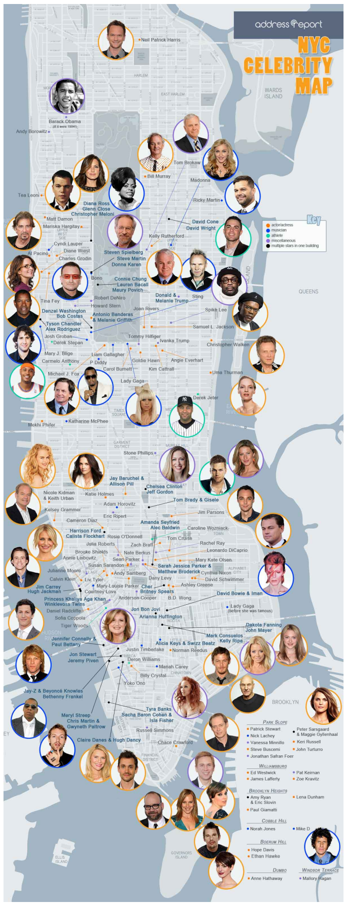


Figure A.76: The locations of celebrities' home [64]

The infographics in Figure. A.76 is show the managing status for a public library. It includes data of people, inventory and financials. Data: The original data is a data sheet of the running conditions of public library Story: The story is that the business is running well. All major data including people, usage, products showed the library is in a good condition.

Informal classification:

As an infographic is used *individually* but shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *no controls.* The *visual components* used include full-text, photographics, and data visualizations. The data visualization produced by this tool is *not interactive.* The viewing sequence is in *parallel.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.7 Games

A.7.1 Halo: Reach

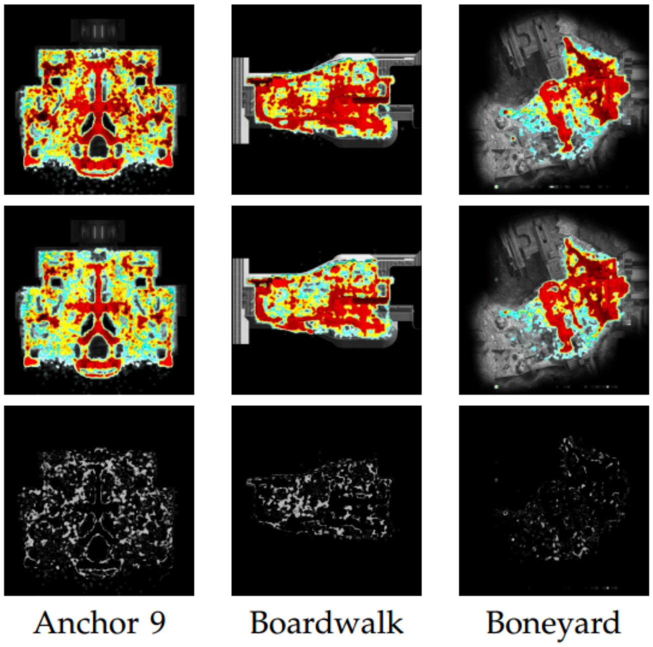


Figure A.77: Visualization in Halo: Reach [65]

The heatmaps in Figure. A.77 is to compare three players for their deaths (top), kills (middle) and their differences (bottom row) separately in the game of Halo: Reach [65]. The difference map is to show the areas in the map, at which one player have advantages over others on the possibility of survival.

Informal classification:

As a visualization in the game is used *individually* but can be shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *basic controls*. The *visual components* used include full- text and data visualizations. The viewing sequence is in *parallel*. However, it is also *replicable*, as it is stored and can be played back at any time.

A.7.2 Call of Duty: : Black Ops

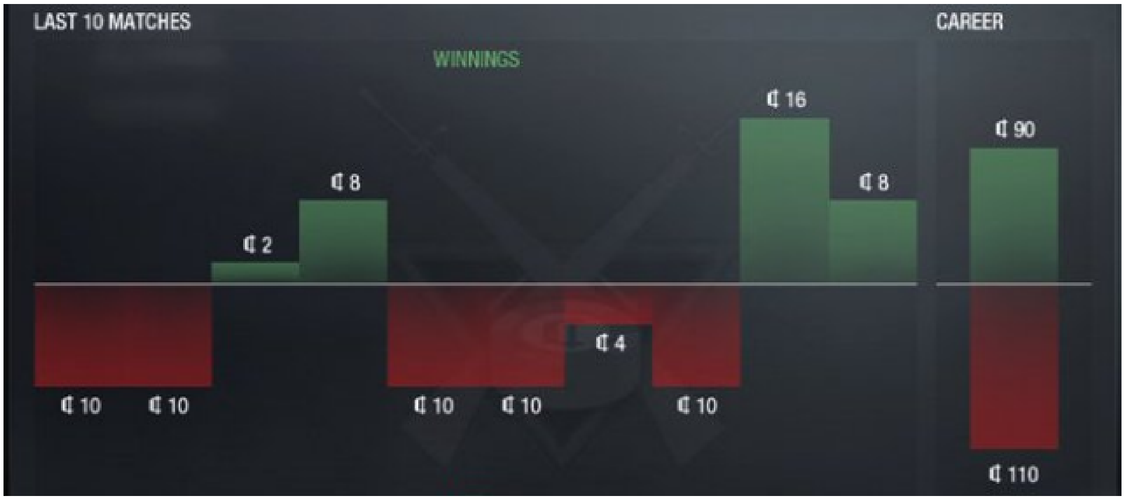


Figure A.78: Visualization in Call of Duty: : Black Ops [66]

The bar chart from Figure. A.78 is to shown the wager earnings for different players. The players are competitive from the game Call of Duty: Black Ops [66]

Informal classification:

As a visualization in the game is used *individually* but can be shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *basic controls*. The *visual components* used include full- text and data visualizations.The viewing sequence is in *sequence*.

However, it is also *replicable*, as it is stored and can be played back at any time.

A.7.3 StarCraft II

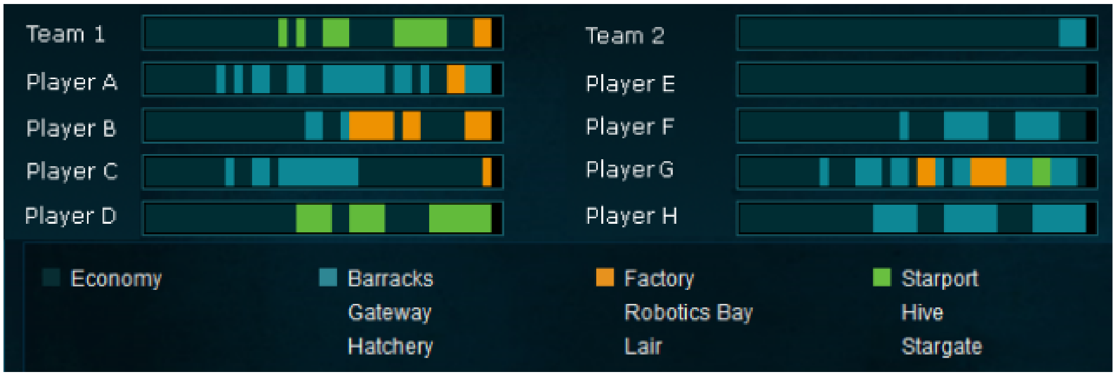


Figure A.79: Visualization in StarCraft II [67]

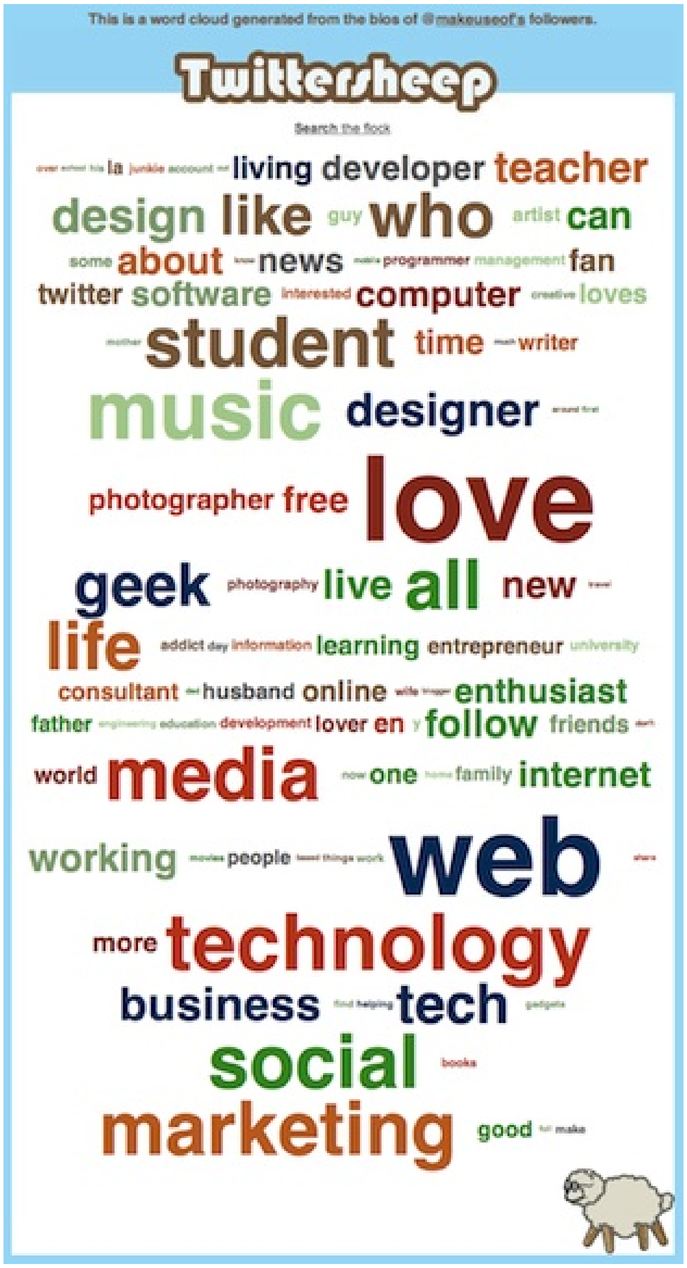
The Figure. A.79 [67] show that a comparison of the two opposing teams for their building order of installations. Players are listed separately on the left and right columns.

Informal classification:

As a visualization in the game is used *individually* but can be shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *high-bandwidth medium*, but the user has *basic controls.* The *visual components* used include full- text and data visualizations.The viewing sequence is in *sequence.* However, it is also *replicable*, as it is stored and can be played back at any time.

A.8 Social Media

A.8.1 TwitterSheep



This is a word cloud generated from the bios of @makeuseof's followers.

Twittersheep
Search the flock
over school his la junkie account living developer teacher
design like guy who artist can
some about news programmer management fan
twitter software interested computer creative loves
mother student time writer
music designer around first
photographer free love
geek photography live all new travel
life addict day information learning entrepreneur university
consultant husband online wife blogger enthusiast
father engineering education development lover en follow friends
world media now one home family internet
working movies people things work web
more technology
business find helping tech gadgets
social books
marketing good make

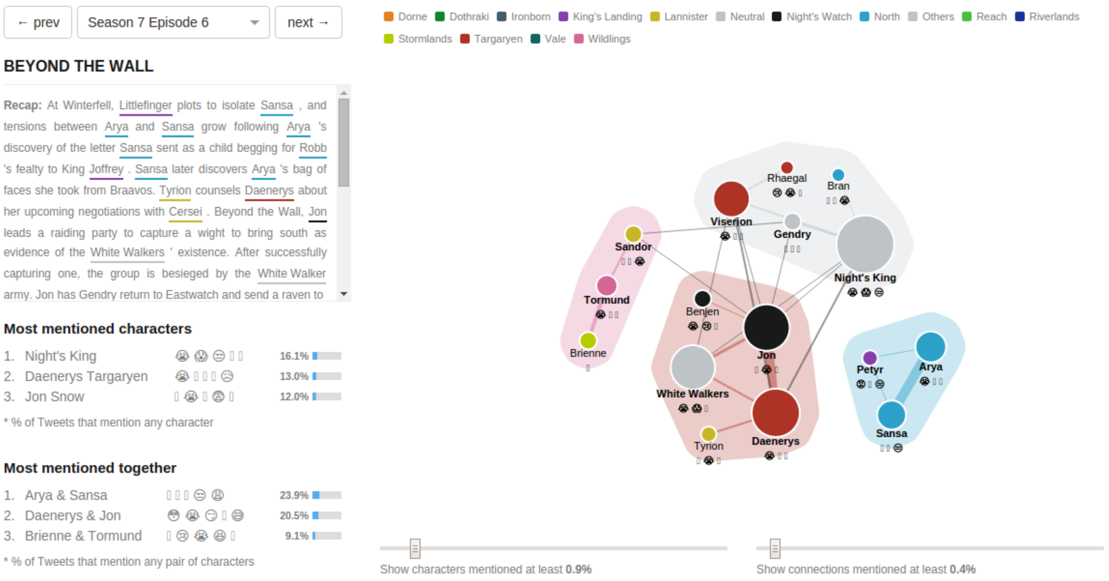
Figure A.80: TwitterSheep interface [68]

The application shown in Figure. A.80 gathers the topics for ones tweets. This example is the result for a software developer working in IT industry. The data are all the texts of tweets for one account The story is that the summary of major topics of a software engineers tweets. He is interested web and technology.

Informal classification:

As a visualization in the social media is used *individually* but can be shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *low-bandwidth medium*, but the user has *basic controls.* The *visual components* used include full- text and data visualizations.The viewing sequence is in *sequence.* However, it is also *replicable*, as it is stored a images.

A.8.2 Twitter Interactive Games of Thrones



← prev | Season 7 Episode 6 | next →

BEYOND THE WALL

Recap: At Winterfell, Littlefinger plots to isolate Sansa , and tensions between Arya and Sansa grow following Arya 's discovery of the letter Sansa sent as a child begging for Robb 's fealty to King Joffrey . Sansa later discovers Arya 's bag of faces she took from Braavos. Tyrion counsels Daenerys about her upcoming negotiations with Cersei . Beyond the Wall, Jon leads a raiding party to capture a wight to bring south as evidence of the White Walkers ' existence. After successfully capturing one, the group is besieged by the White Walker army. Jon has Gendry return to Eastwatch and send a raven to

Most mentioned characters

1. Night's King 16.1%
2. Daenerys Targaryen 13.0%
3. Jon Snow 12.0%

* % of Tweets that mention any character

Most mentioned together

1. Arya & Sansa 23.9%
2. Daenerys & Jon 20.5%
3. Brienne & Tormund 9.1%

* % of Tweets that mention any pair of characters

🟧 Dorne 🟩 Dothraki ■ Ironborn 🟪 King's Landing 🟨 Lannister ■ Neutral ■ Night's Watch 🟦 North ■ Others 🟩 Reach 🟦 Riverlands 🟩 Stormlands 🟥 Targaryen ■ Vale 🟪 Wildlings

Show characters mentioned at least **0.9%**

Show connections mentioned at least **0.4%**

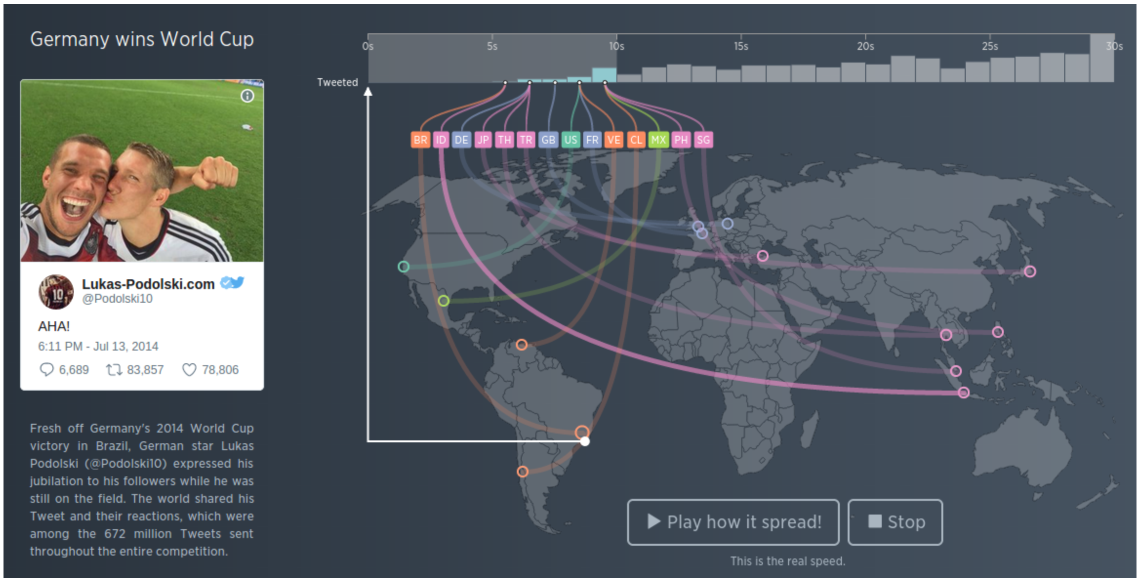
Figure A.81: Twitter interactive Games of Thrones [69]

The text tool in Figure. A.81 is to show the story summary of each episode. The Visualization on the right side is a connection graph showing relationship between different districts. The colored points are coded based on families and sides. Data: the texts and highlighted names of main character, their background and relationship. Story: The interaction frequencies and targets of main characters.

Informal classification:

As a visualization in the social media is used *individually* but can be shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual* *components* used include full- text, animation, sound, and data visualizations.The viewing sequence is in *sequence*. The data visualization produced is *interactive*. However, it is also *ephemeral*, as it is stored.

A.8.3 Twitter Interactive: How tweets spread



Germany wins World Cup

Lukas-Podolski.com
@Podolski10

AHA!

6:11 PM - Jul 13, 2014

6,689 83,857 78,806

Fresh off Germany's 2014 World Cup victory in Brazil, German star Lukas Podolski (@Podolski10) expressed his jubilation to his followers while he was still on the field. The world shared his Tweet and their reactions, which were among the 672 million Tweets sent throughout the entire competition.

Figure A.82: Twitter Interactive [70]: how world cup news spread.

This example in Figure. A.82 in NYT-smis a twitter application designed to show the spreading trend for news of the 2010 German soccer world cup. The application is running with animation in the speed as the real speed of the news spreading. Data: The spreading location based on the tweets location of 2010 German soccer world cup Story: The sequence and speed of news spreading internationally. The bar chart at the top is to show the frequency of tweets.

Informal classification:

As a visualization in the social media is used *individually* but can be shared *publicly*, the audience is potentially *many people* on the Internet, typically viewed individually on their *personal devices* (although public mass viewings are certainly possible). The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *basic controls*. The *visual components* used include full- text , animaiton, and data visualizations. The data visualization produced is *interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.9 Augmented Reality

A.9.1 Uber Mobile Visualization

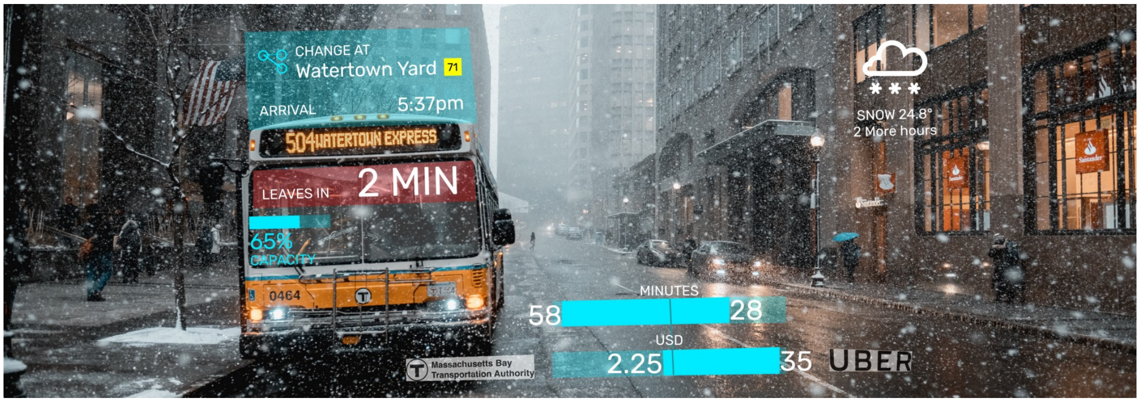


Figure A.83: The architecture for the augmented reality [71] to have a data visualization

This example in Figure. A.83 is an augmented reality application designed to show the road condition for drivers. The application is running with animation and data visualization in the speed according to the road condition. Data: The road condition and traffic information Story: The information on top of the bus is showing the bus schedule and warning the it's leaving in two minutes. The bar charts and numbers on the right shows the navigation and weather information.

Informal classification: This example is primarily intended to be *general.* As a visualization in the social media is used *individually*, the audience is potentially *one driver*, typically viewed individually on their *personal devices.* The delivery method is thus *distributed* but *synchronous.* Data visualization is a *high-bandwidth medium*, but the user has *basic controls.* The *visual components* used include full-text , animaiton, and data visualizations. The data visualization produced is *not interactive.* The viewing sequence is in *sequence.* However, it is also *ephemeral.*

A.9.2 AR Data Visualization Design

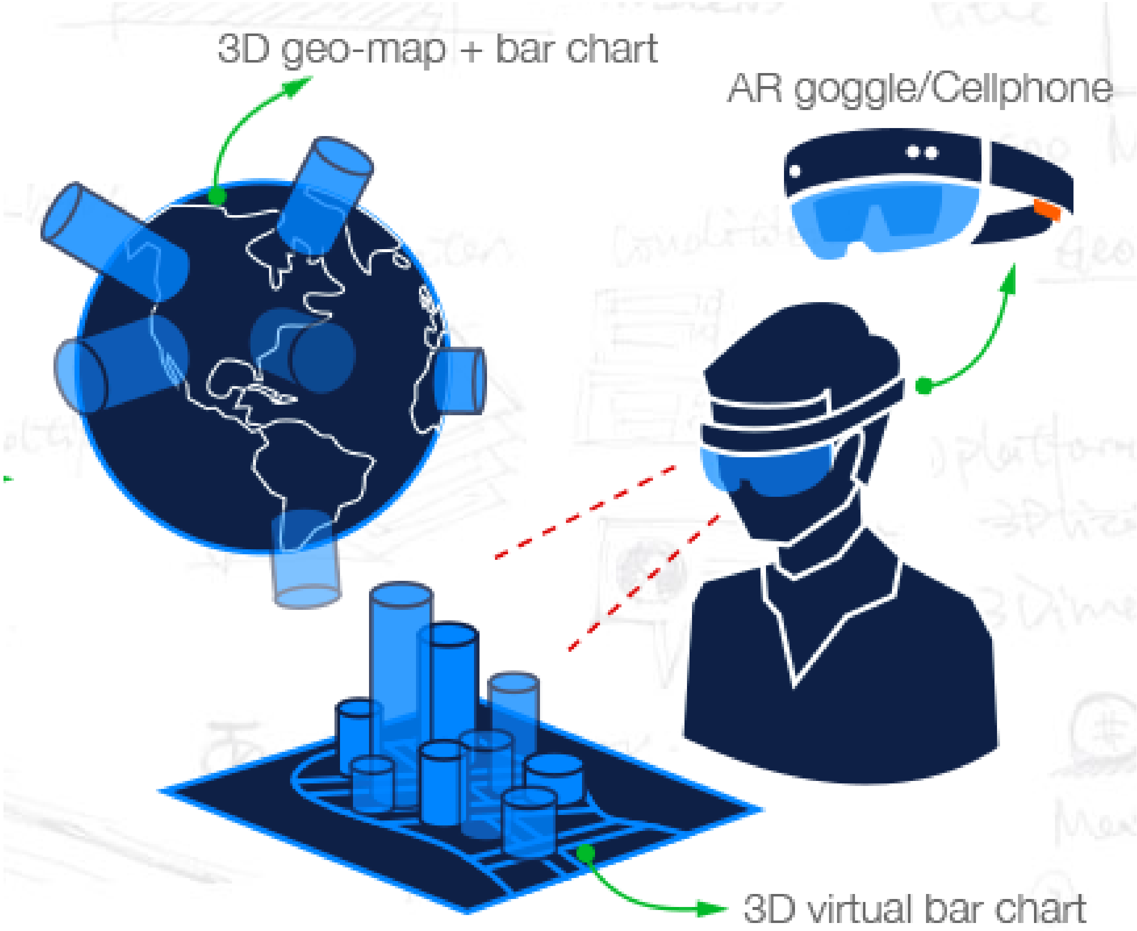


Figure A.84: The architecture for the augmented reality [72] to have a data visualization

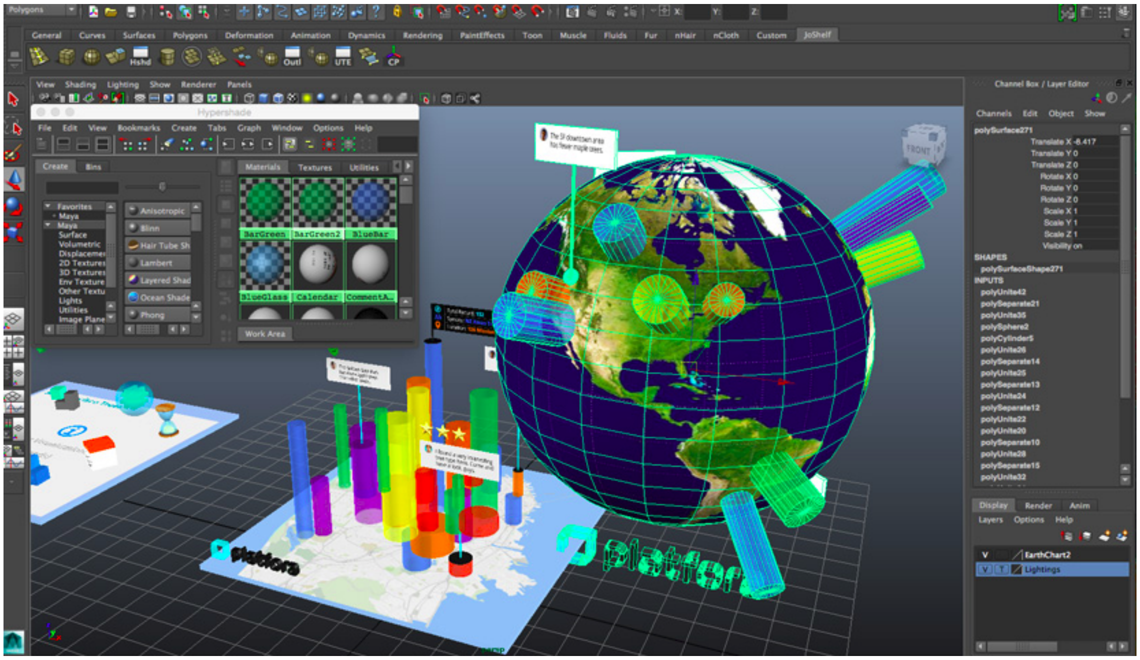


Figure A.85: How AR [72] is used to visualize the world population

This example in Figure. A.84 and Figure. A.85 is an augmented reality application designed to show the data visualization. The application is running with animation and data visualization with user's interaction Data: The data visualization and annotation of the world population Story: The information are shown as the bar chart to demonstrate the distribution of the world population.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual components* used include full-text , animaiton, and data visualizations. The viewing sequence is in *sequence*.

However, it is also *replicable*.

A.9.3 AR 3D Design

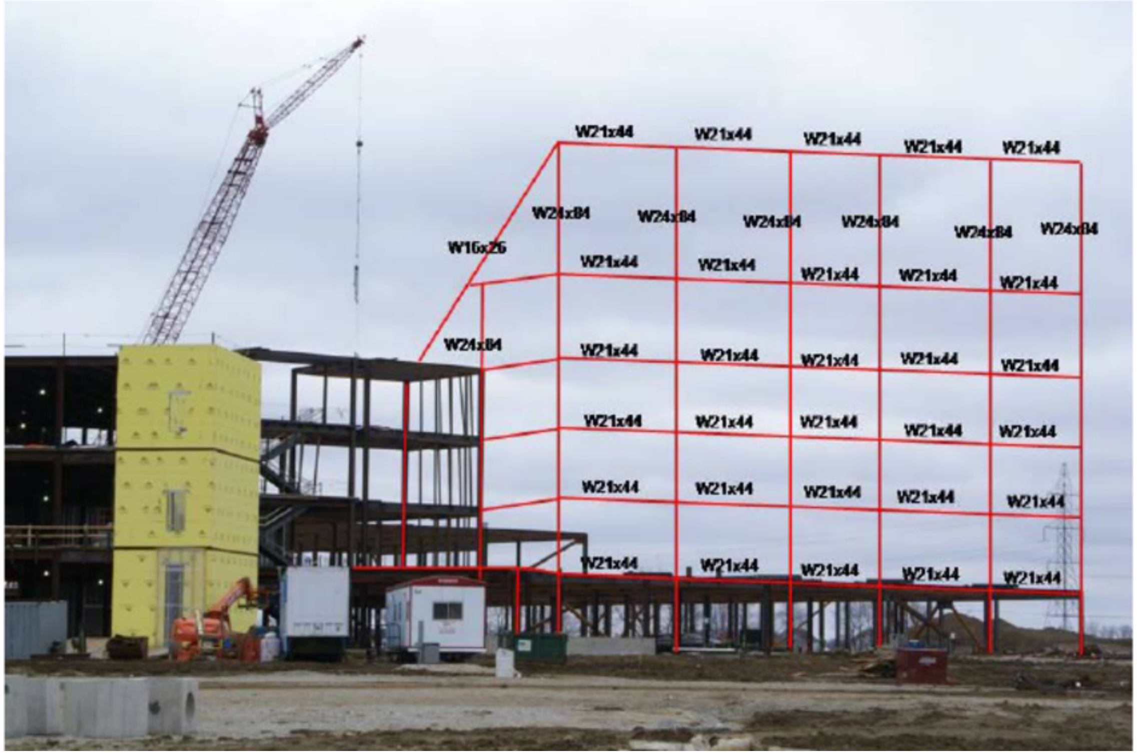


Figure A.86: The process for the engineers with augmented reality [73] to design a building with data visualization

This example in Figure. A.86 is an augmented reality application designed to show the data visualization combined with real working site. The application is running with animation and data visualization according to the user's interaction. Data: The data visualization, images, and annotation of the world population Story: The information are shown as frames, line chart, and text to demonstrate the meta data and expected layout of the finished building.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.9.4 AR Flight Data

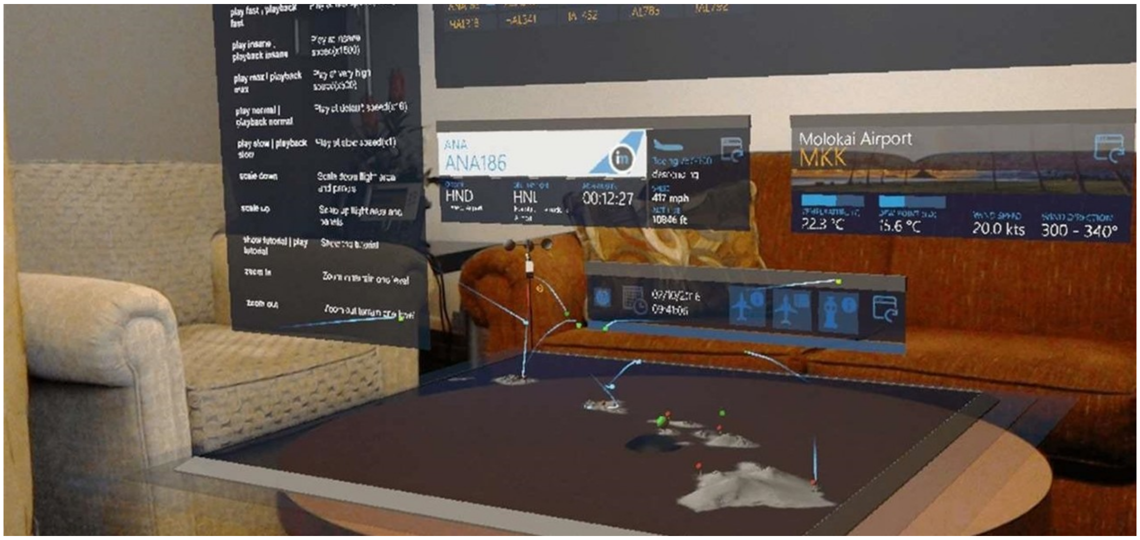


Figure A.87: The process for passengers to view flight data in augmented reality [74]

This example in Figure. A.87 is an augmented reality application designed to show the data visualization of flight tracking information. The application is running with animation and data visualization to give users a virtual route of flight positions. Data: The data visualization, animation, and annotation for flight information Story: The information is shown as animated tracks to show the plane heights and locations in real time

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *basic controls*. The *visual components* used include full- text , images, animation, and data visualizations. The viewing sequence is in *sequence*. However, it is also *ephemeral*.

A.9.5 AR Street Visualization

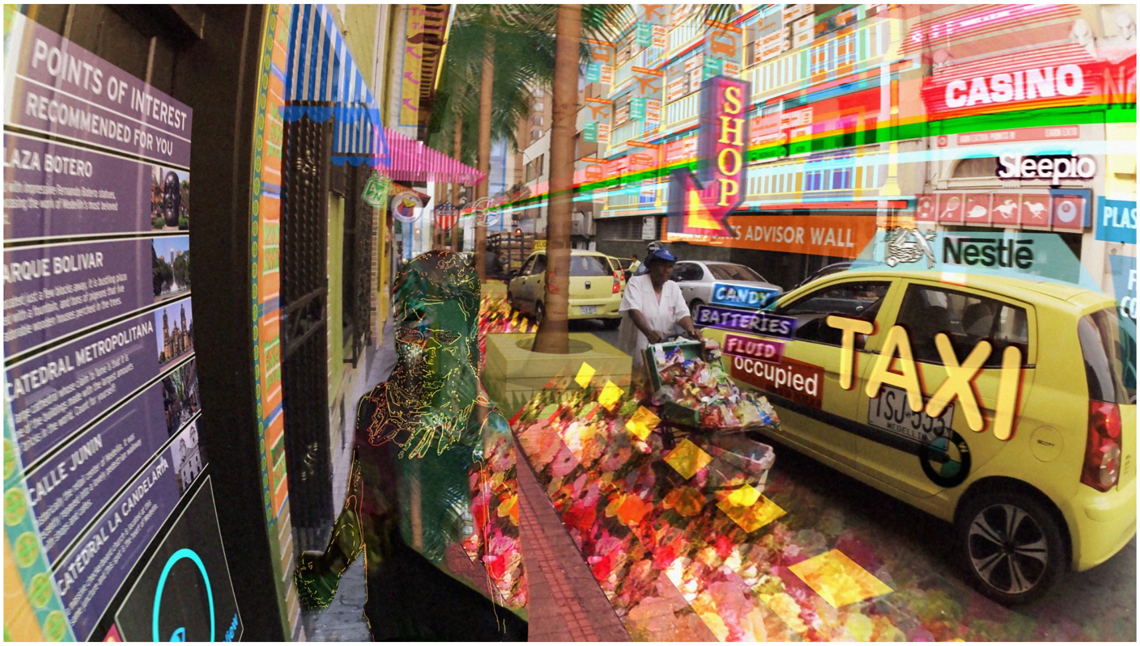


Figure A.88: Street viewers obtain information from augmented reality [75] on the street

This example in Figure. A.88 is an augmented reality application designed to show the data visualization combined with real working site. The application is running with animation and data visualization in the speed according to the user's interaction. Data: The data visualization, images, texts, and annotation of the street view. Story: The information is shown to annotate the street view in real time.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *basic controls*. The *visual components* used include full- text , images, animation, and data visualizations. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.9.6 AR pipeline

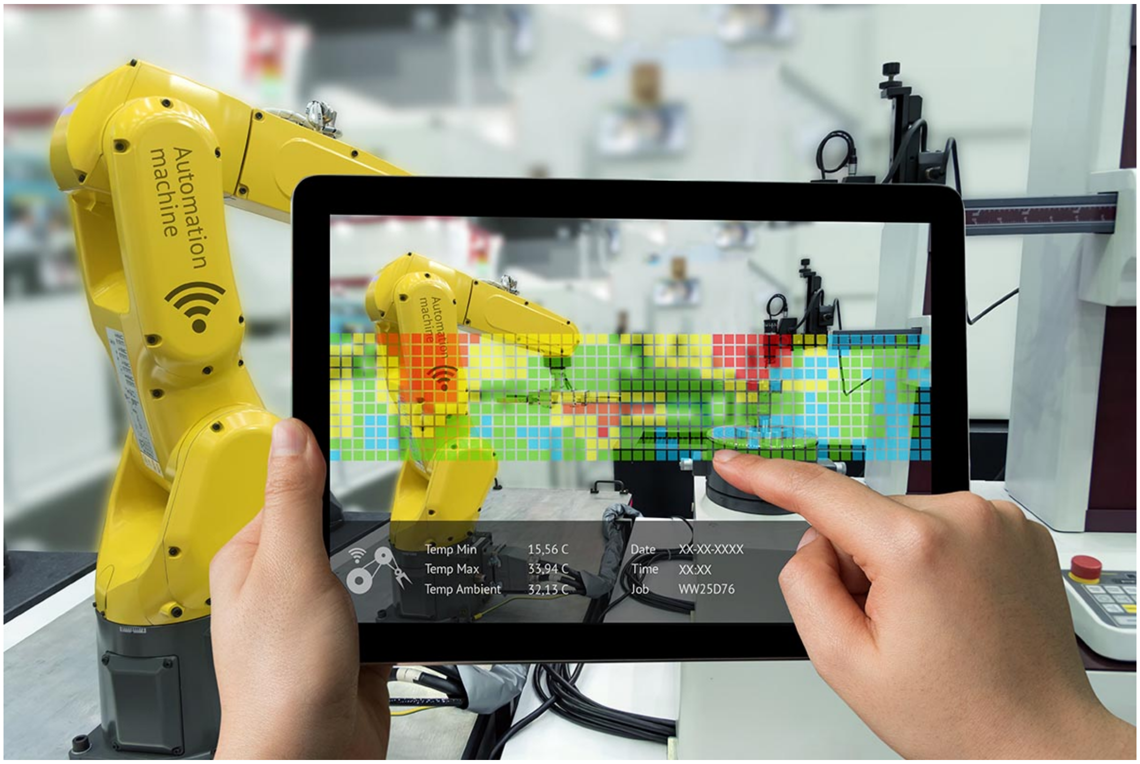


Figure A.89: Engineers have the pipes shown with augmented reality [76]

This example in Figure. A.89 is an augmented reality application designed to show the data visualization combined with production pipeline. The application is running with animation and data visualization according to the user's interaction. Data: The data visualization, images, and annotation of the pipeline and devices. Story: The information is shown to the supporting information about a pipeline for its production rate, efficiency, and running condition.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *no controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *not interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.9.7 AR Infrastructure Visualization

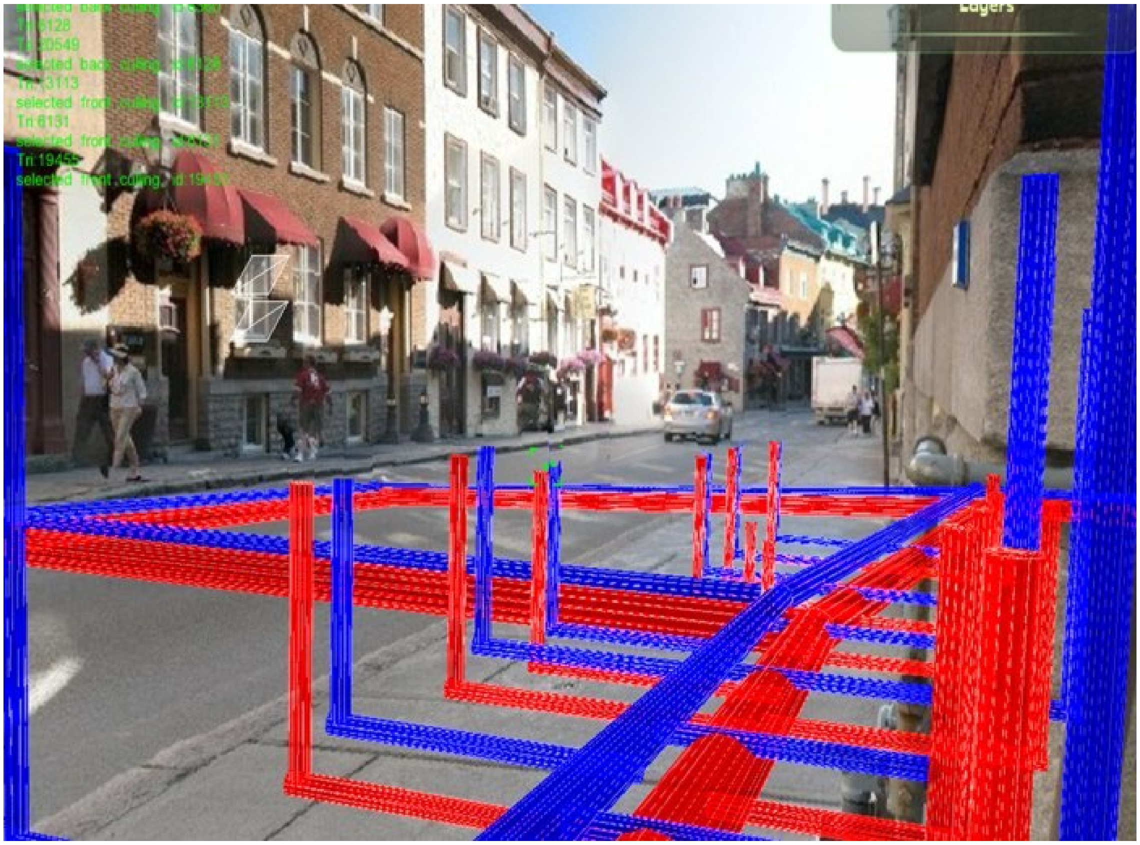


Figure A.90: Engineers have the underground infrastructure shown with augmented reality [77]

This example in Figure. A.90 is an augmented reality application designed to show the data visualization combined with real working site. The application is running with animation and data visualization in the speed according to the view. Data: The data visualization, images, and annotation of the street view and highlighted pipes. Story: The information are shown as frames, line chart, and text to demonstrate the meta data and expected layout of underground pipes.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *high-bandwidth medium*, but the user has *no controls.* The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *not interactive.* The viewing sequence is in *sequence.* However, it is also *ephemeral.*

A.9.8 AR Bio-Chemical Visualization

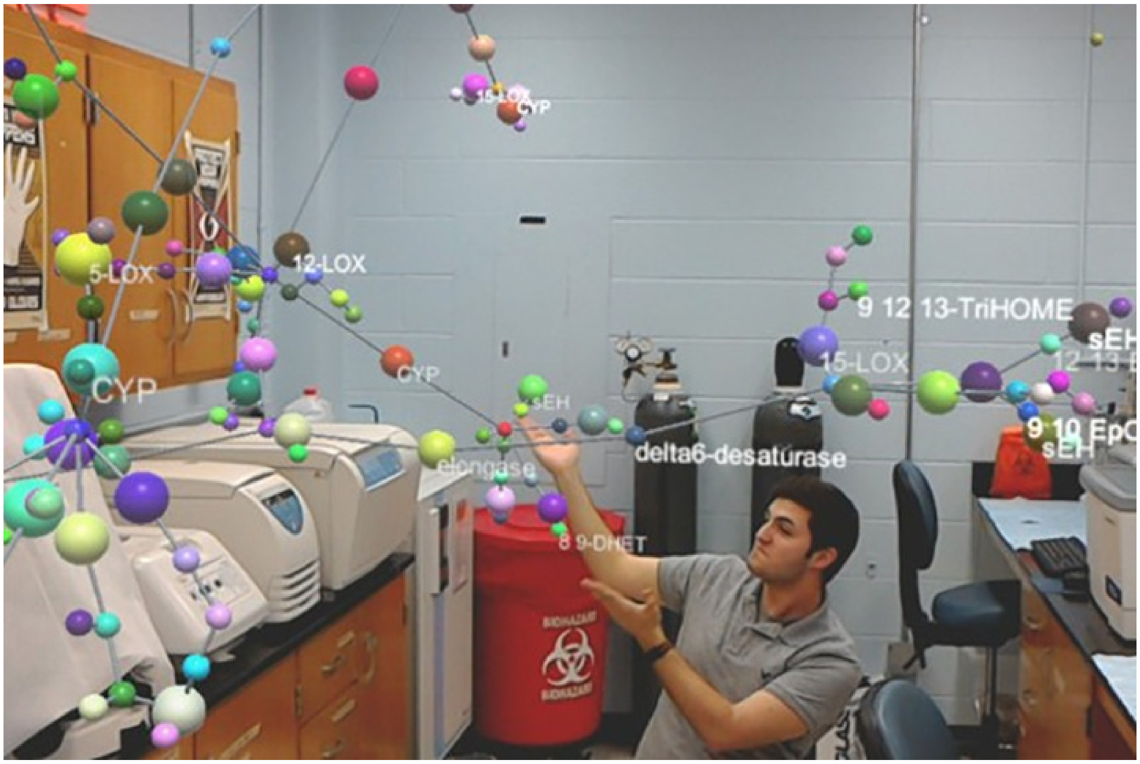


Figure A.91: A Bio-chemistry researcher has a structure of molecular shown with augmented reality [78]

This example in Figure. A.91 is an augmented reality application designed to show the data visualization of molecular structure of Bio-chemical materials. The application is running with animation and data visualization according to the user's interaction. Data: The data visualization, graph visualization, and annotation of the molecular structure Story: The information shown is to help the bio-chemical researchers understand the structures of different molecular.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *basic controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *basicly interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.10 Virtual Reality

A.10.1 Adobe VR Data Visualization

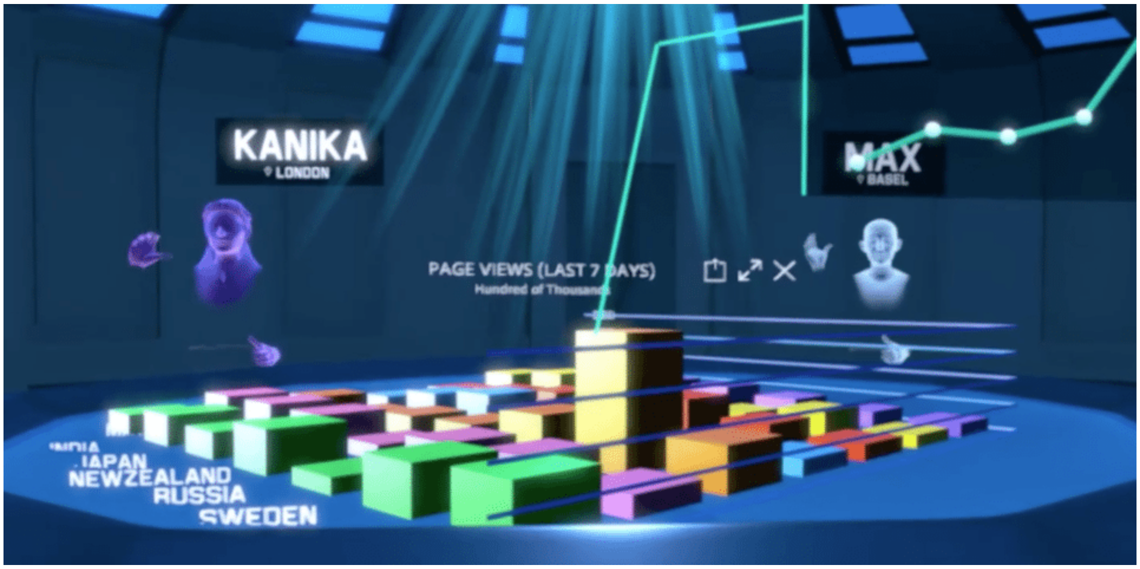


Figure A.92: How virtual reality tool [79] is used to visualize the baseball training data.

This example in Figure. A.92 is an virtual reality application designed to show the data visualization in virtual space. The application is running with animation, texts, and data visualization according to the user's interaction. Data: The data visualization, images, and annotation of page views in different parts of the world. Story: The information is to show the comparison of different parts of the world in terms of internet web page view. The trend of each part can be highlighted.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and the actual working site, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *high-bandwidth medium*, but the user has *full controls.* The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive.* The viewing sequence is in *sequence.* However, it is also *replicable.*

A.10.2 VR Baseball training

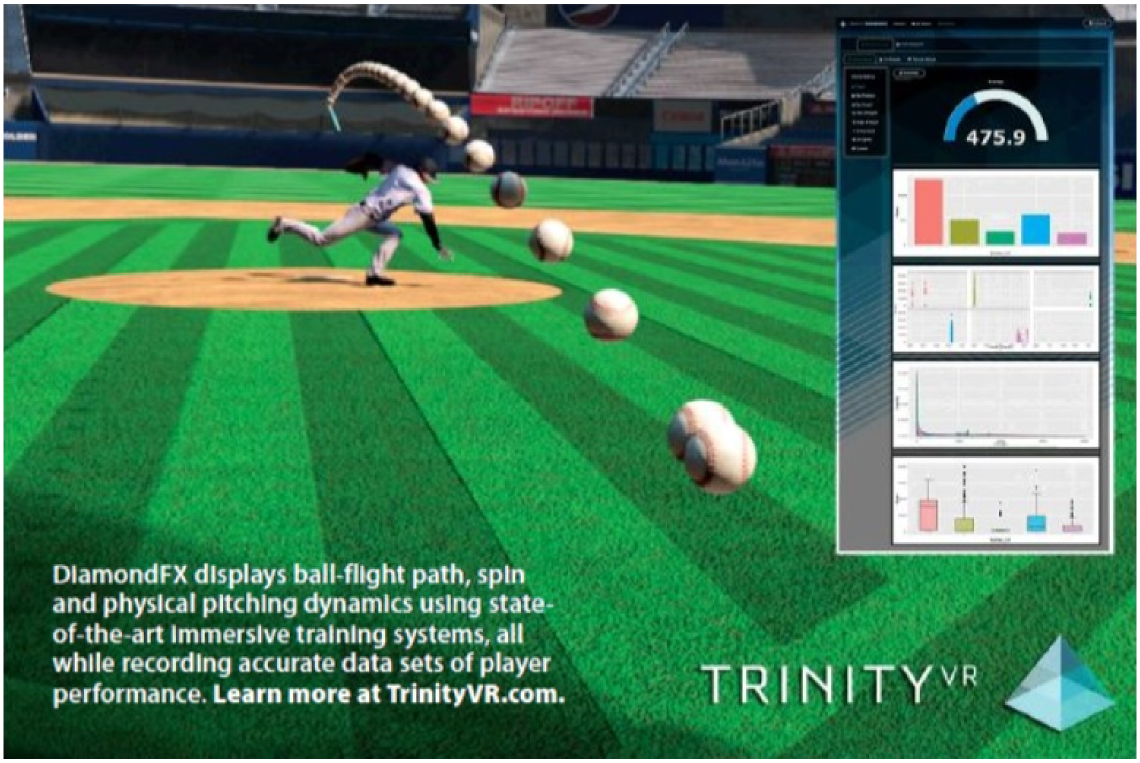

475.9
DiamondFX displays ball-flight path, spin and physical pitching dynamics using state-of-the-art immersive training systems, all while recording accurate data sets of player performance. Learn more at TrinityVR.com.
TRINITY VR

Figure A.93: How virtual reality tool [80] is used to visualize the baseball training data.

This example in Figure. A.93 is an virtual reality application designed to facilitate baseball training by highlighting the baseball trace. The application is running with animation, texts, and data visualization in the speed according to the user's interaction. Data: The data visualization, images, and highlighting of the baseball trajectory. Story: The information are shown as frames, line chart, and text to help the user correct the baseball training by showing the correct trajectory.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and try to improve baseball techniques, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *fully controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *full interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.10.3 VR Big Data Analysis

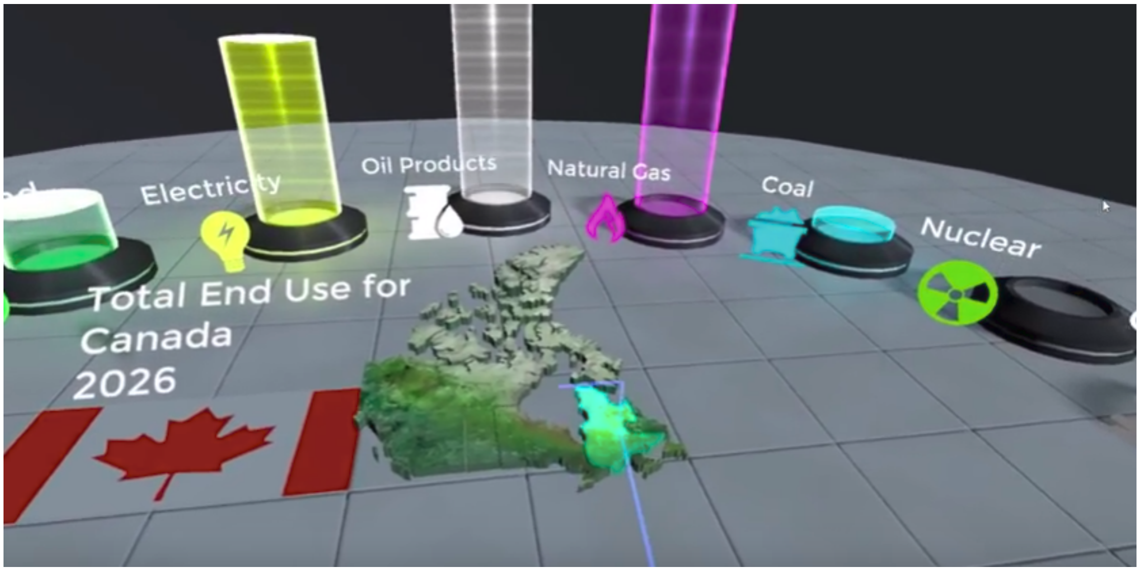


Figure A.94: How virtual reality tool [81] is used to visualize big data in 3D.

This example in Figure. A.94 is an virtual reality application designed to facilitate understanding the natural level globally. The application is running with animation, texts, and data visualization in the speed according to the user's interaction. Data: The data visualization, images, and map highlighting of the natural resources. Story: The information is to show the comparison of natural resources of different countries with map visualization.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and try to improve baseball techniques, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous.* Data visualization is a *high-bandwidth medium*, but the user has *basic controls.* The *visual components* used include full- text , images, animation, and data visualizations. However, it is also *replicable.*

A.10.4 VR Lens Big Data

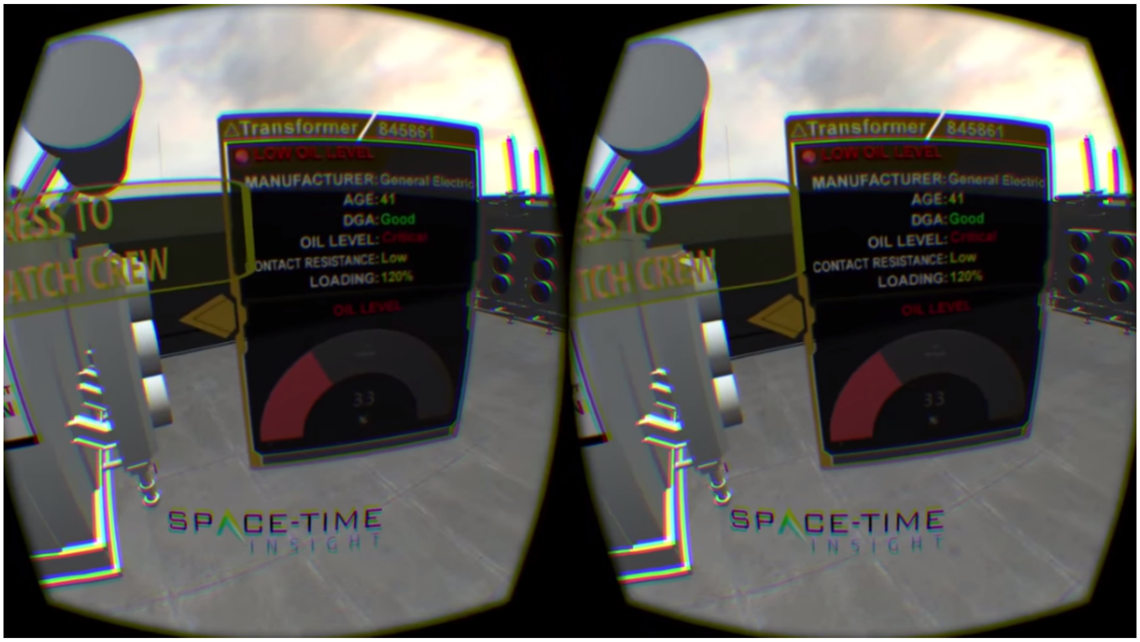


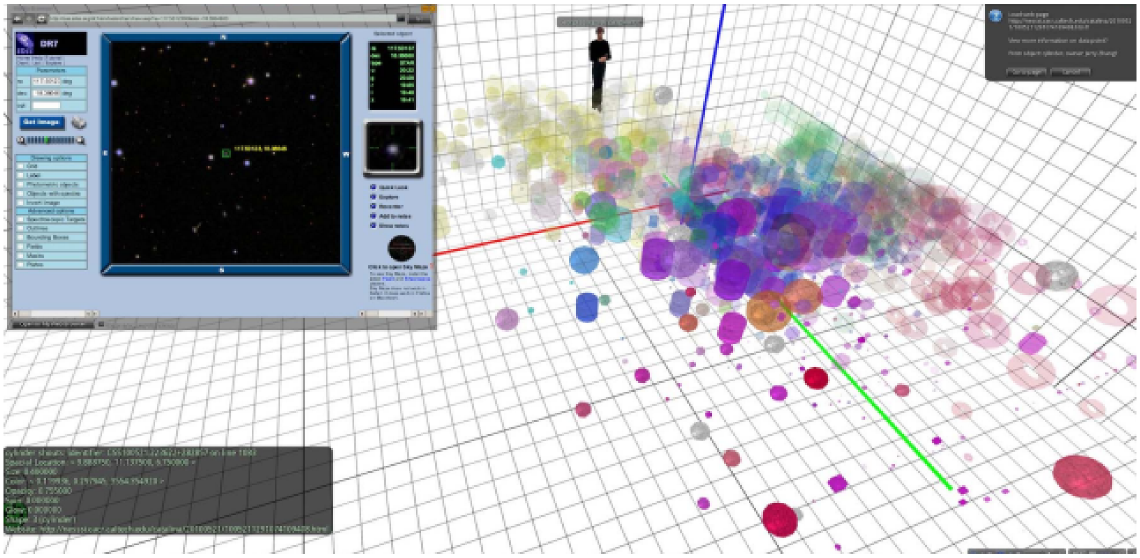
Figure A.95: How virtual reality tool [82] is used to visualize big data with virtual objects.

This example in Figure. A.95 is an virtual reality application designed to facilitate training in the environment of a new industrial facility. The application is running with animation, texts, and data visualization according to the user's interaction. Data: The data visualization, images, and highlighting of the information of each devices. Story: The information is shown to show the trainee what is the condition of each device in a virtual industrial facility.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and try to improve baseball techniques, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.10.5 VR Immersive visualization for Big Data



Cylinder shows: Identifier: [illegible]
Spacial Location: < [illegible] >
Size: [illegible]
Color: < [illegible] >
Opacity: [illegible]
Spin: [illegible]
Glow: [illegible]
Shape: 3 (cylinder)
Website: [illegible]

Figure A.96: How virtual reality tool [83] is used to do data analytics.

This example in Figure. A.96 is an virtual reality application designed to facilitate the understanding of high dimensional data. The application is running with animation, texts, and data visualization according to the user's interaction. Data: The data visualization, images, and virtual viewer figure. Story: The information is to show how to visualize high dimensional data generated by the users from a user interface in 2D.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and try to improve baseball techniques, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual* *components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.10.6 VR Bio-informatics Visualization

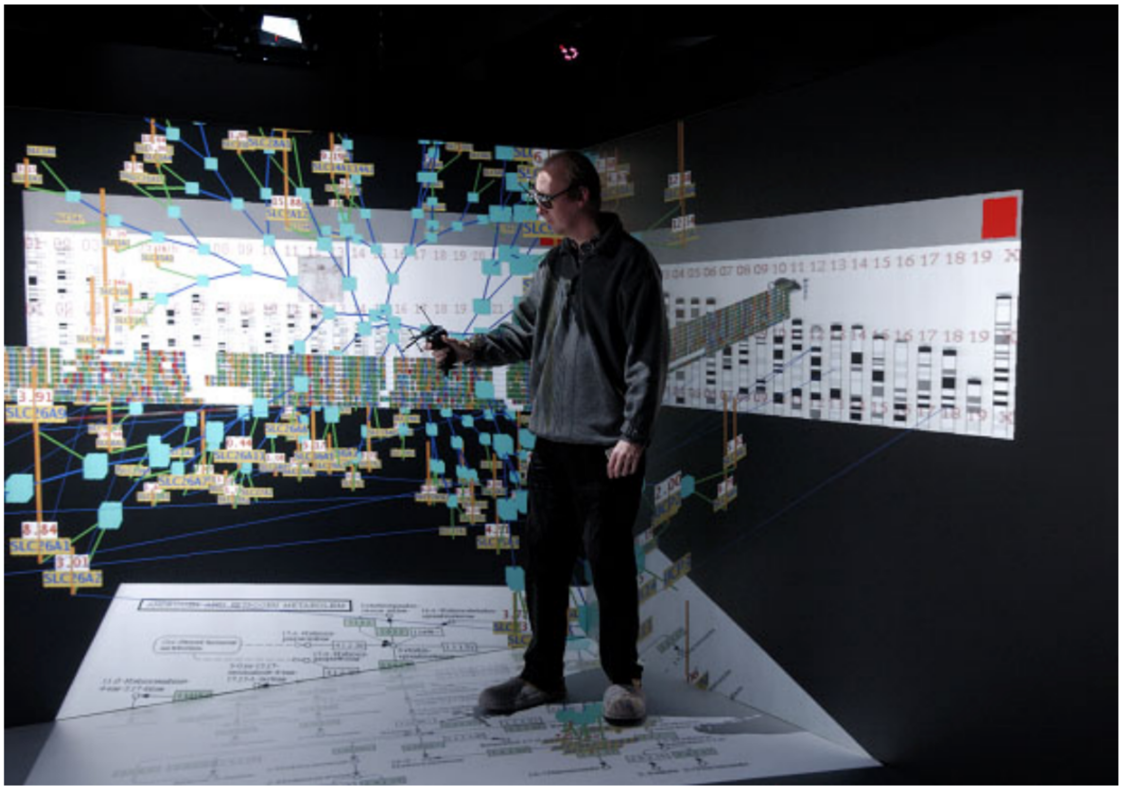


Figure A.97: How virtual reality tool [84] is used to visualize gene information with graph in 3D.

This example in Figure. A.97 is an virtual reality application designed to facilitate understanding of the gene sequence. The application is running with animation, texts, and data visualization according to the user's interaction. Data: The data visualization, images, and the connection of each gene sequence. Story: The information is to show how a researcher is in the virtual space with label gene sequence surrounding him.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and try to improve baseball techniques, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.10.7 VR Geo Map Visualization

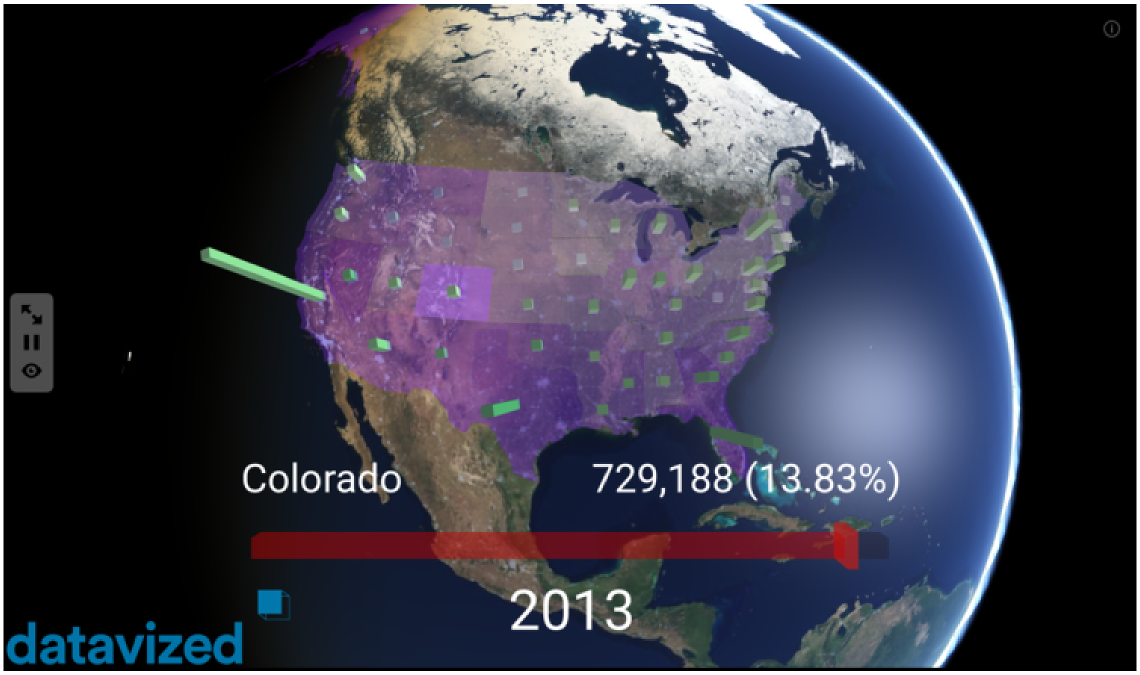


Figure A.98: How virtual reality tool [85] is used to visualize geographical information.

This example in Figure. A.98 is an virtual reality application designed to facilitate comparison of population of certain areas on earth. The application is running with animation, texts, and data visualization according to the user's interaction. Data: The data visualization, images, and map. Story: The information is to show the population as barchart in 3D on a virtual earth.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *one personal* who is working on the dataset and try to improve baseball techniques, although public mass viewings are certainly possible. The delivery method is thus *distributed* and *asynchronous*. Data visualization is a *high-bandwidth medium*, but the user has *full controls*. The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive*. The viewing sequence is in *sequence*. However, it is also *replicable*.

A.10.8 VR Data Analysis

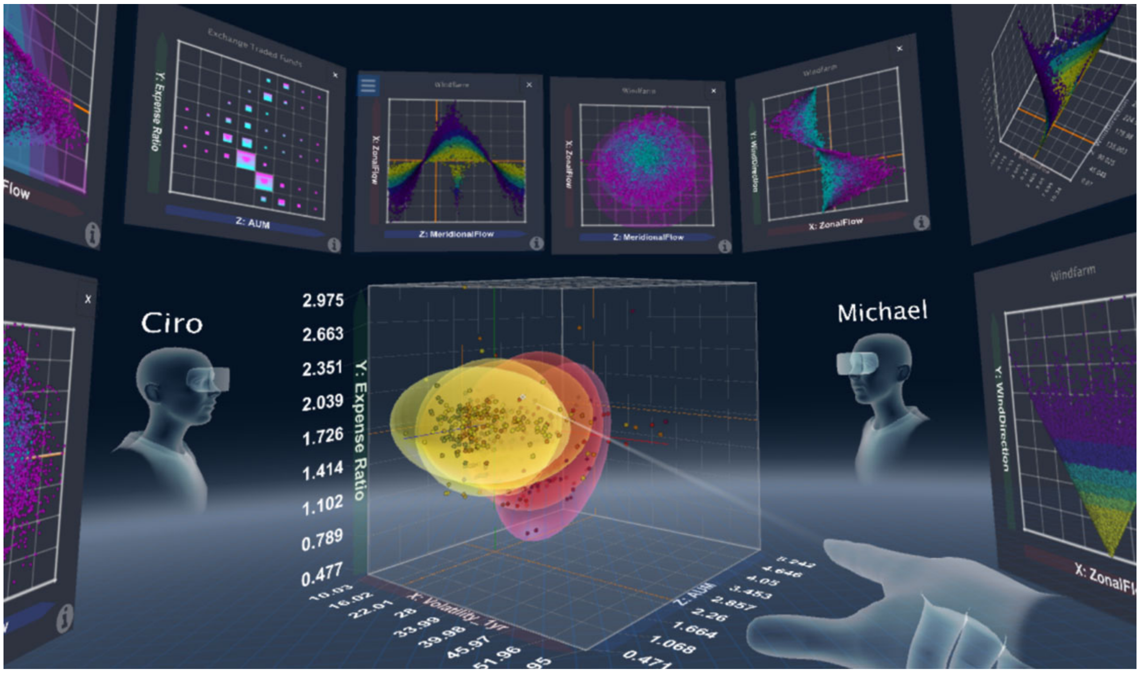


Figure A.99: How virtual reality tool [86] is used to do data analysis.

This example in Figure. A.99 is an virtual reality application designed to show the data visualization in virtual space. The application is running with animation, texts, virtual figure, and data visualization in the speed according to the user's interaction. Data: The data visualization, images, and annotation of the a multidimensional data set. Story: The information is shown as frames, charts, and text to demonstrate how data can be analysis from multiple dimensions with virtual reality.

Informal classification:

As a visualization rendered with personal tools, it is used *individually* but can be shared *publicly*, the audience is potentially *multiple users* who is working on the dataset, although one personal user is certainly possible. The delivery method is thus *distributed* and *synchronous.* Data visualization is a *high-bandwidth medium*, but the user has *basic controls.* The *visual components* used include full- text , images, animation, and data visualizations. The data visualization produced is *fully interactive.* The viewing sequence is in *sequence.* However, it is also *replicable.*

Appendix B: Data Comics Evaluation Protocol

B.1 Evaluation: DataComics vs PowerPoint: Test Cases and Scripts

B.1.1 Twitter Heatmap for Stocks

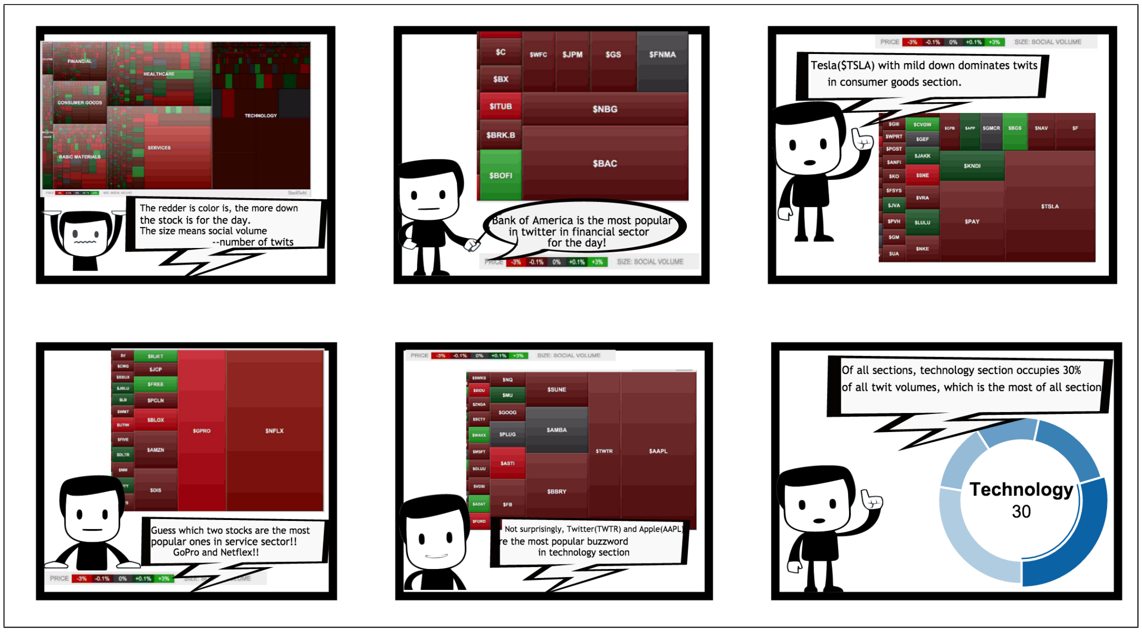

FINANCIAL
HEALTHCARE
CONSUMER GOODS
TECHNOLOGY
BASIC MATERIALS
SERVICES
The redder is color is, the more down the stock is for the day. The size means social volume --number of twits
$C
$WFC
$JPM
$GS
$FNMA
$BX
$ITUB
$NBG
$BRK.B
$BAC
$BOFI
Bank of America is the most popular in twitter in financial sector for the day!
PRICE -3% -0.1% 0% +0.1% +3% SIZE: SOCIAL VOLUME
Tesla($TSLA) with mild down dominates twits in consumer goods section.
$KNDI
$PAY
$TSLA
Guess which two stocks are the most popular ones in service sector!! GoPro and Netflex!!
$GPRO
$NFLX
$AMZN
$DIS
Not surprisingly, Twitter(TWTR) and Apple(AAPL) re the most popular buzzword in technology section
$SUNE
$AMBA
$BBRY
$TWTR
$AAPL
Of all sections, technology section occupies 30% of all twit volumes, which is the most of all section
Technology
30

Figure B.1: Stock heat map in comic style

1. What are the six sections of stocks this story talks about _____,_____, _____, _____, _____,_____

2. Which stock has largest volume for financial sector? _____

3. Which stock has largest volume for consumer goods section?_____

4. Which stock has largest volume for service section? _____

5. Which stock has largest volume for technology section? _____

6. Which section has 30 percent of tweets volume of all sections? _____

7. Which stock has second largest volume for consumer goods section? _____

8. Which stock has second largest volume for service section? _____

9. Which stock has second largest volume for financial section? _____

B.1.2 the U.S. Census Population pyramid

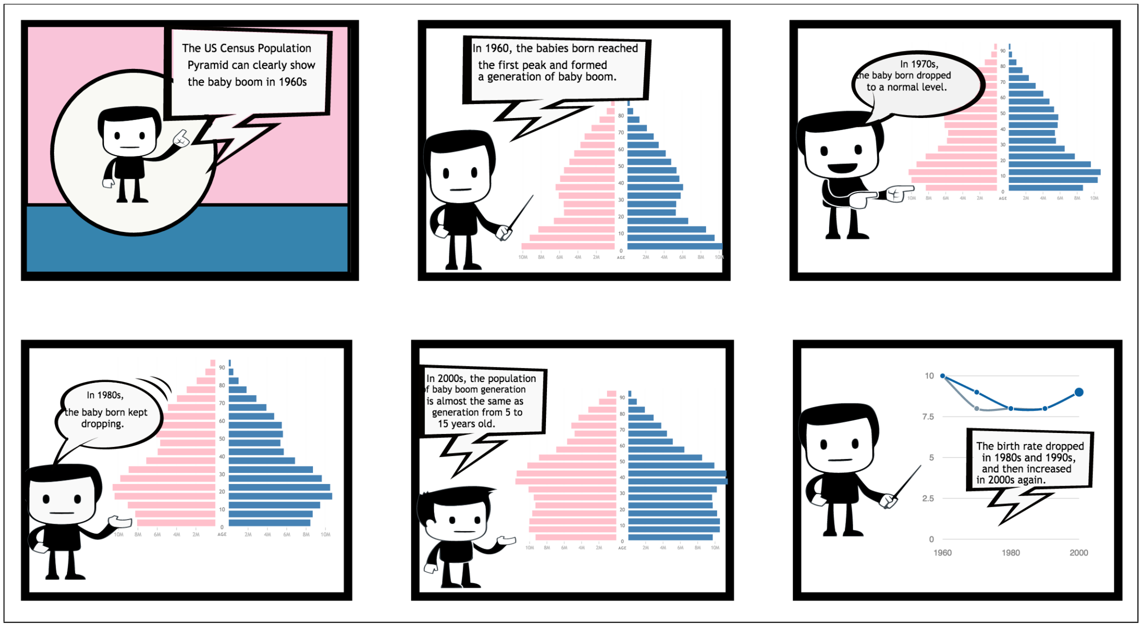

The US Census Population Pyramid can clearly show the baby boom in 1960s
In 1960, the babies born reached the first peak and formed a generation of baby boom.
In 1970s, the baby born dropped to a normal level.
In 1980s, the baby born kept dropping.
In 2000s, the population of baby boom generation is almost the same as generation from 5 to 15 years old.
The birth rate dropped in 1980s and 1990s, and then increased in 2000s again.
10
7.5
5
2.5
0
1960
1980
2000

Figure B.2: US census data in comic style

1. At what year did the baby boom start? _____

2. What is the level of baby born in 1970s? Boom or Normal (circle one)

3. What is the level of baby born in 1980s? Boom or Normal or Keep dropping (circle one)

4. In 2000s, the population of the baby boom generation is almost the same as the population aged from _____ to _____.

5. At which two decades the baby dropped? _____, _____

6. How many babies were born in 1970_____

7. How many babies were born in 1980_____

8. How many babies were born in 2000_____

B.1.3 World Happiness

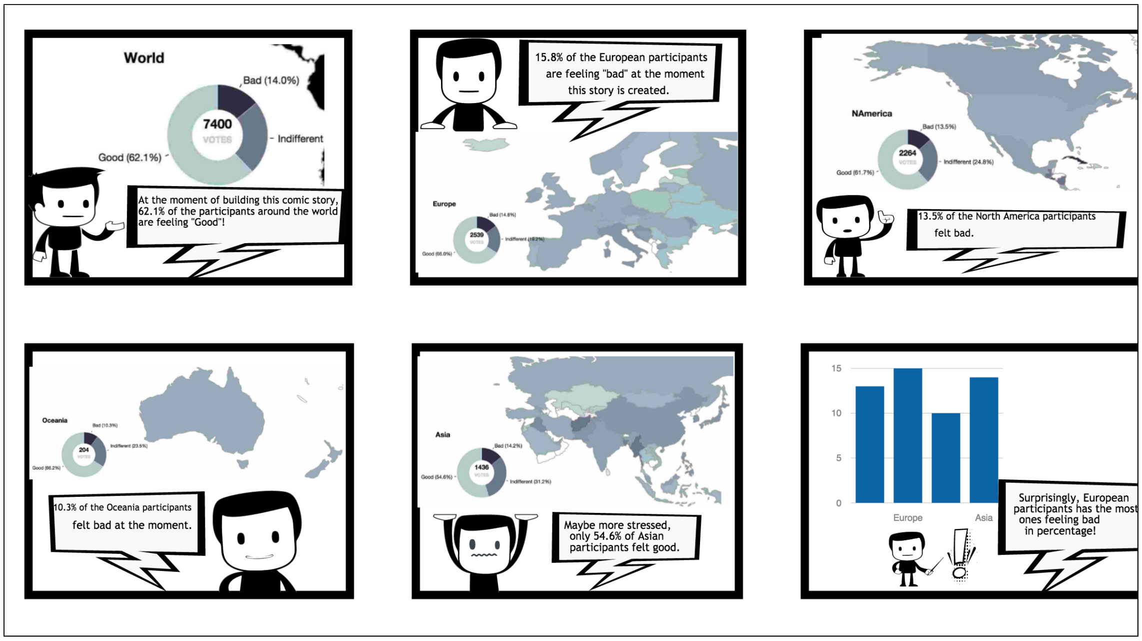

World
Bad (14.0%)
7400
VOTES
Indifferent
Good (62.1%)
At the moment of building this comic story, 62.1% of the participants around the world are feeling "Good"!
15.8% of the European participants are feeling "bad" at the moment this story is created.
Europe
Bad (14.8%)
2539
VOTES
Indifferent (19.2%)
Good (66.0%)
NAmerica
Bad (13.5%)
2264
VOTES
Indifferent (24.8%)
Good (61.7%)
13.5% of the North America participants felt bad.
Oceania
Bad (10.3%)
204
VOTES
Indifferent (23.5%)
Good (66.2%)
10.3% of the Oceania participants felt bad at the moment.
Asia
Bad (14.2%)
1436
VOTES
Indifferent (31.2%)
Good (54.6%)
Maybe more stressed, only 54.6% of Asian participants felt good.
15
10
5
0
Europe
Asia
Surprisingly, European participants has the most ones feeling bad in percentage!

Figure B.3: World happiness data in comic style

1. How many votes worldwide are there for this happiness data? _____

2. What is the percentage of European participants that are feeling bad at the moment? _____

3. What is the percentage of North America participants that are feeling bad at the moment? _____

4. What is the percentage of Oceania participants that are feeling bad at the moment? _____

5. What is the percentage of Asian participants that are feeling Good at the moment? _____

6. Frame six is the comparison of bad feeling participants among five different continents. Of these continents, which one do you think has the most percentage of participants feeling bad?

7. What is the percentage of Oceania participants that are feeling good at the moment?_____

8. How many votes in Asia are there for this happiness data? _____

9. How many votes Oceania are there for this happiness data? _____

B.1.4 Star Wars

1. We like people _____ to us.

2. Which movie series is this datacomics story talking about ? _____

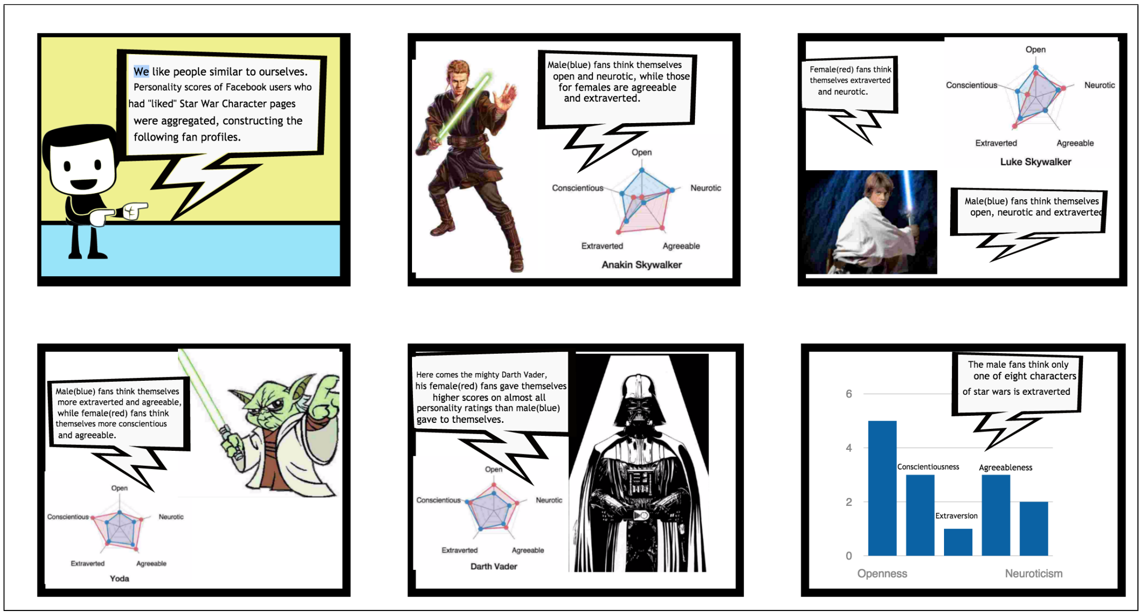

We like people similar to ourselves. Personality scores of Facebook users who had "liked" Star War Character pages were aggregated, constructing the following fan profiles.
Male(blue) fans think themselves open and neurotic, while those for females are agreeable and extraverted.
Open
Conscientious
Neurotic
Extraverted
Agreeable
Anakin Skywalker
Female(red) fans think themselves extraverted and neurotic.
Open
Conscientious
Neurotic
Extraverted
Agreeable
Luke Skywalker
Male(blue) fans think themselves open, neurotic and extraverted
Male(blue) fans think themselves more extraverted and agreeable, while female(red) fans think themselves more conscientious and agreeable.
Open
Conscientious
Neurotic
Extraverted
Agreeable
Yoda
Here comes the mighty Darth Vader, his female(red) fans gave themselves higher scores on almost all personality ratings than male(blue) gave to themselves.
Open
Conscientious
Neurotic
Extraverted
Agreeable
Darth Vader
The male fans think only one of eight characters of star wars is extraverted
6
4
2
0
Conscientiousness
Agreeableness
Extraversion
Openness
Neuroticism

Figure B.4: Star data in comic style

3. For the male fans of Anakin Skywalker, what are the two personalities they rate themselves with highest scores? (_____ and _____)

4. For the female fans of Anakin Skywalker, what are the three personalities they rate themselves with highest scores? (_____, _____ and _____)

5. For the female fans of Luke Skywalker, which personality is best rated? _____.

6. For the male fans of Master Yoda, which two personalities are best rated? _____, _____

7. For the female fans of Master Yoda, which two personalities are the best rated. _____,_____

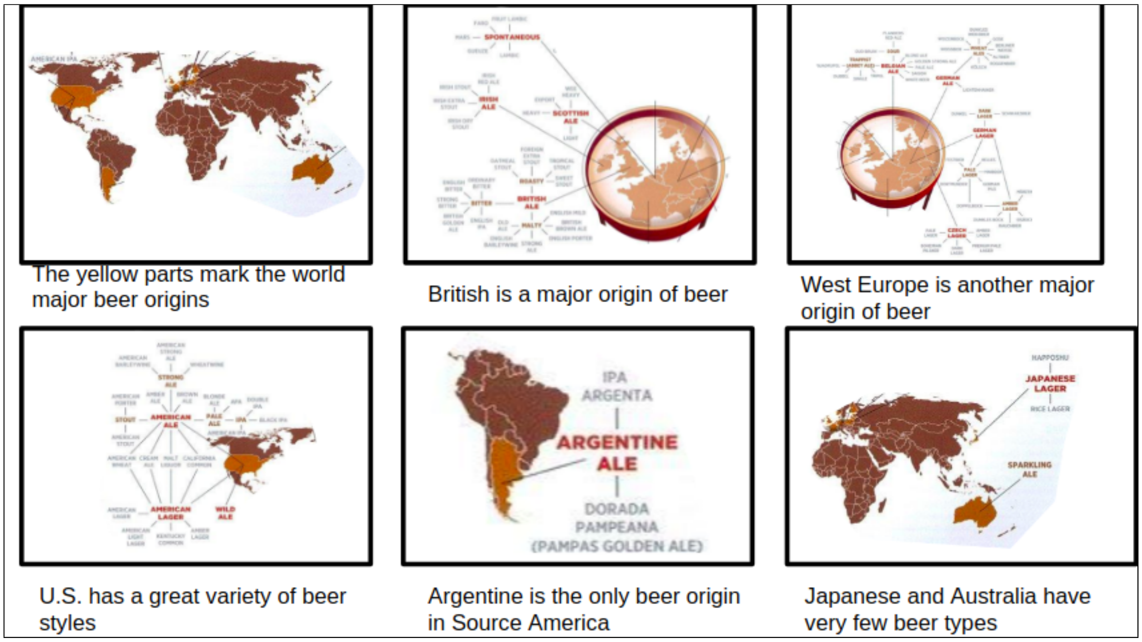
8. For the male fans of Darth Vader, which personality is best rated?_____

9. There are eight Star war characters counted in during the like collecting process

from the fans. The last frame shows how many times each one personality is rated with highest score by the male fans. From the bar chart/table, please tell which personality is rated the highest with the least times. ____

B.2 Evaluation: Single Frame vs Frame Panels: Test Cases and Scripts

B.2.1 The Origin of Major Beer Types



The yellow parts mark the world major beer origins

British is a major origin of beer

West Europe is another major origin of beer

U.S. has a great variety of beer styles

Argentine is the only beer origin in Source America

Japanese and Australia have very few beer types

Figure B.5: The geo origin of beer around world in sequenced panels

1. What are the countries that most of beer styles are originated from

2. List the styles of Scottish Ale

3. List the styles of Japanese Larger

4. Do America and Belgium share any beer styles?

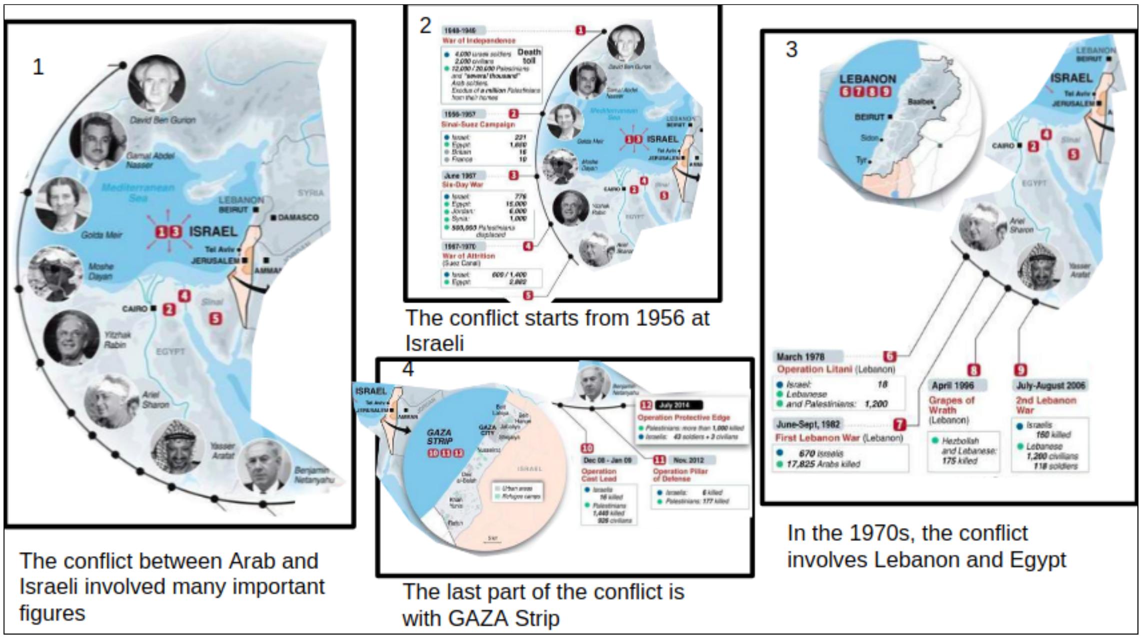
5. Do America and German share any beer styles?

6. Do America and British share any beer styles?

7. What is the country in Asia that mentioned for the styles of beer

8. Does Australia have any beer style? If yes, what are they?

B.2.2 The Arabic-Israel War



The conflict between Arab and Israeli involved many important figures

The conflict starts from 1956 at Israeli

The last part of the conflict is with GAZA Strip

In the 1970s, the conflict involves Lebanon and Egypt

Figure B.6: The Arabic-Israel War in sequenced panels

1. How many important persons are involved in the Arab-Israeli conflict?

2. Which conflict happened in the 1980s?

3. In which country/countries did the conflict 6,7,8,9 happened?

4. In which country/countries did the conflict 1,2,3,4,5 happened?

5. How many conflict happened in the 1970s

6. How many conflict happened in the 1960s

7. Who is the important figure involved in the conflict happened in 6,7,8,9

8. Who is the important figure involved in the conflict happened in 10, 11, 12

B.2.3 Cell Phone Phishing

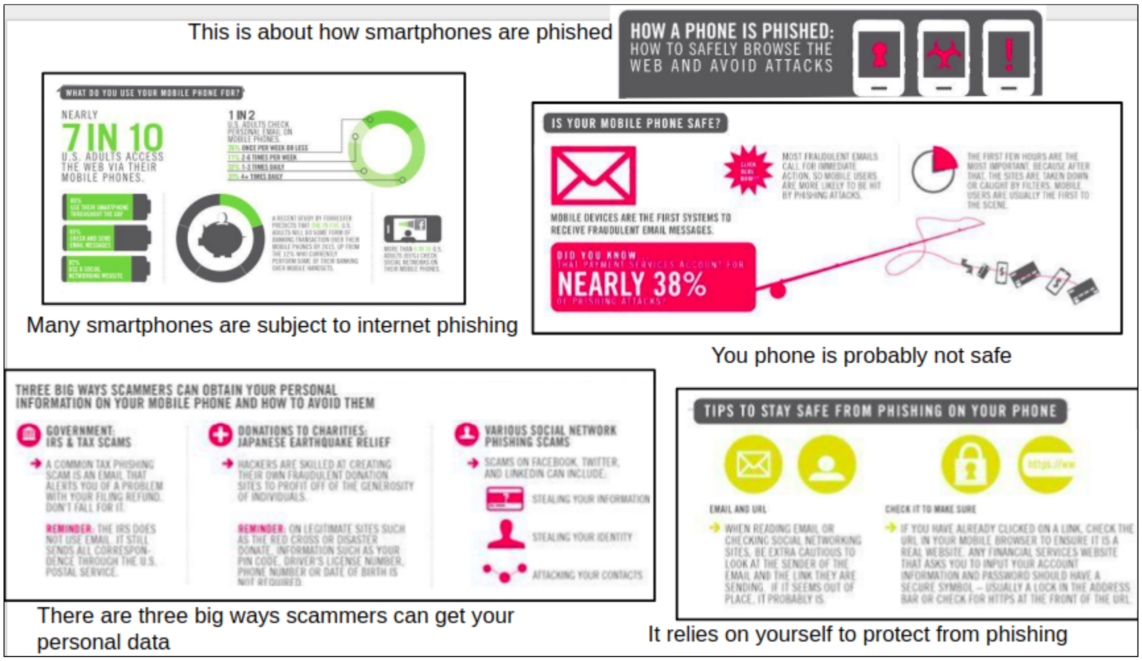

This is about how smartphones are phished
HOW A PHONE IS PHISHED:
HOW TO SAFELY BROWSE THE WEB AND AVOID ATTACKS
WHAT DO YOU USE YOUR MOBILE PHONE FOR?
NEARLY 7 IN 10 U.S. ADULTS ACCESS THE WEB VIA THEIR MOBILE PHONES.
1 IN 2 U.S. ADULTS CHECK PERSONAL EMAIL ON MOBILE PHONES.
26% ONCE PER WEEK OR LESS
11% 2-6 TIMES PER WEEK
32% 1-3 TIMES DAILY
31% 4+ TIMES DAILY
Many smartphones are subject to internet phishing
IS YOUR MOBILE PHONE SAFE?
MOBILE DEVICES ARE THE FIRST SYSTEMS TO RECEIVE FRAUDULENT EMAIL MESSAGES.
MOST FRAUDULENT EMAILS CALL FOR IMMEDIATE ACTION, SO MOBILE USERS ARE MORE LIKELY TO BE HIT BY PHISHING ATTACKS.
THE FIRST FEW HOURS ARE THE MOST IMPORTANT, BECAUSE AFTER THAT, THE SITES ARE TAKEN DOWN OR CAUGHT BY FILTERS. MOBILE USERS ARE USUALLY THE FIRST TO THE SCENE.
DID YOU KNOW... THAT PAYMENT SERVICES ACCOUNT FOR NEARLY 38% OF PHISHING ATTACKS?
You phone is probably not safe
THREE BIG WAYS SCAMMERS CAN OBTAIN YOUR PERSONAL INFORMATION ON YOUR MOBILE PHONE AND HOW TO AVOID THEM
GOVERNMENT: IRS & TAX SCAMS
A COMMON TAX PHISHING SCAM IS AN EMAIL THAT ALERTS YOU OF A PROBLEM WITH YOUR FILING REFUND. DON'T FALL FOR IT.
REMINDER: THE IRS DOES NOT USE EMAIL. IT STILL SENDS ALL CORRESPONDENCE THROUGH THE U.S. POSTAL SERVICE.
DONATIONS TO CHARITIES: JAPANESE EARTHQUAKE RELIEF
HACKERS ARE SKILLED AT CREATING THEIR OWN FRAUDULENT DONATION SITES TO PROFIT OFF OF THE GENEROSITY OF INDIVIDUALS.
REMINDER: ON LEGITIMATE SITES SUCH AS THE RED CROSS OR DISASTER DONATE, INFORMATION SUCH AS YOUR PIN CODE, DRIVER'S LICENSE NUMBER, PHONE NUMBER OR DATE OF BIRTH IS NOT REQUIRED.
VARIOUS SOCIAL NETWORK PHISHING SCAMS
SCAMS ON FACEBOOK, TWITTER, AND LINKEDIN CAN INCLUDE:
STEALING YOUR INFORMATION
STEALING YOUR IDENTITY
ATTACKING YOUR CONTACTS
There are three big ways scammers can get your personal data
TIPS TO STAY SAFE FROM PHISHING ON YOUR PHONE
https://ww
EMAIL AND URL
WHEN READING EMAIL OR CHECKING SOCIAL NETWORKING SITES, BE EXTRA CAUTIOUS TO LOOK AT THE SENDER OF THE EMAIL AND THE LINK THEY ARE SENDING. IF IT SEEMS OUT OF PLACE, IT PROBABLY IS.
CHECK IT TO MAKE SURE
IF YOU HAVE ALREADY CLICKED ON A LINK, CHECK THE URL IN YOUR MOBILE BROWSER TO ENSURE IT IS A REAL WEBSITE. ANY FINANCIAL SERVICES WEBSITE THAT ASKS YOU TO INPUT YOUR ACCOUNT INFORMATION AND PASSWORD SHOULD HAVE A SECURE SYMBOL – USUALLY A LOCK IN THE ADDRESS BAR OR CHECK FOR HTTPS AT THE FRONT OF THE URL.
It relies on yourself to protect from phishing

Figure B.7: The data about cellphone phished in sequenced panels

1. What is the percentage of adults in U.S. access web via cellphone

2. What is the percentage of adults in U.S check their email 1-3 times with cell phone

3. Is mobile users easier to submit their login information than PC users?

4. How many Facebook users have been attacked by Phishing sites

5. Is there any scam involving IRS and tax refund mentioned?

6. Is there any scam involving support kids in Africa mentioned?

7. Should we be cautious of the sender of an email regarding preventing from phishing attacks.

8. Should we install anti-phishing apps on cell phones to prevent phishing attacks

B.2.4 Global Wealthy Population

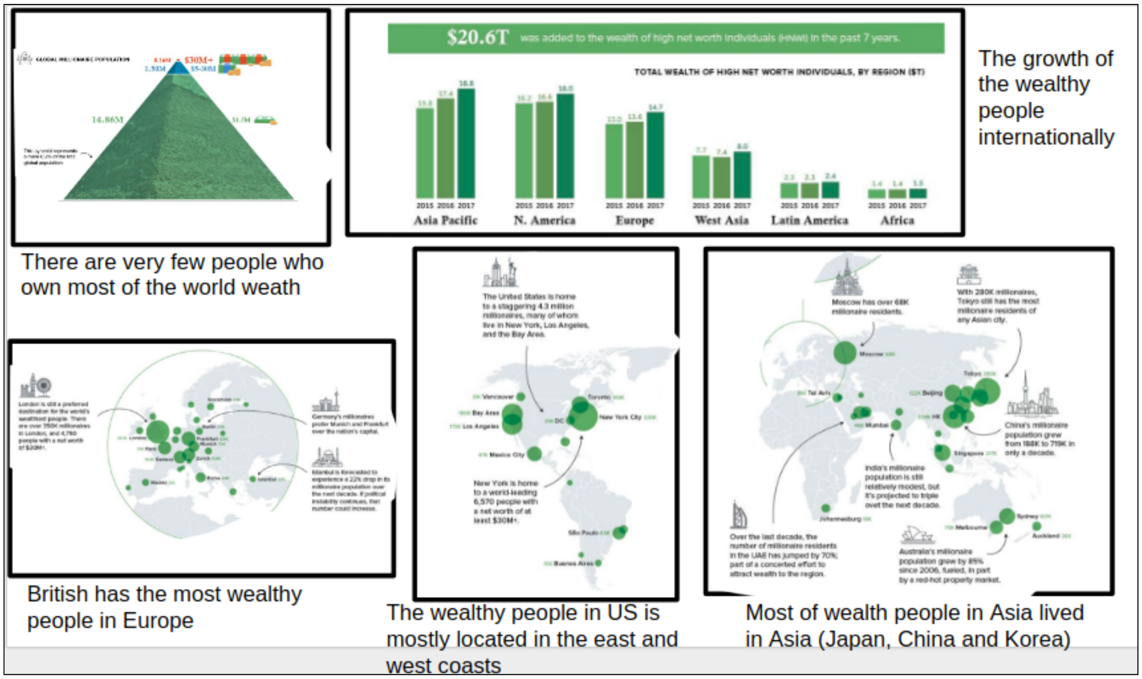


Figure B.8: The data about global wealthy population in sequenced panels

1. Which area has the most number of wealthy people among all the continents?

2. What is the number of millionaires worth more than 30M dollar?
3. What is the percentage of millionaires of the global population?
4. Which city in Europe has the most millionaires?
5. Among North America and Asia, which city has the most millionaires?
6. Which city in Asia has the most millionaires?
7. Which area has the fastest rate of growth of number of millionaires?
8. What kind of pattern we can obtain in term of the relation between number and worth of millionaires?

B.3 Evaluation: Expert Review

The experts are interviewed with the following questions:

1. Does the system help storytelling
2. If yes, how does it help storytelling
3. Do you think the partitioning helps storytelling and understanding of the story
4. Please explain how does the system help storytelling and user to understand the story

Appendix C: Data TV Evaluation Protocol

During the process we use the basic expert review method proposed by Tory and Möller [204], which includes experts evaluating a tool using pre-defined heuristics. We had two datasets prepared for the experts.

C.1 Questions and Protocol

The experts are given but not restricted to a few directions for recommendation as part of the training process:

1. What do you think about overall of the interface?

2. What do you think about ease of use of the interface?

3. What do you think about the functions of the system?

The experts are interviewed with the following questions:

1. Do you think this system will help storytelling?

2. If yes, how does it help?

3. Which functions do you think help you the most

4. Which part of the system do you have the most impression

5. Which part of the system do you not like the most

6. Any suggestion for the improvement of the system

C.2 Data Sets

Each participant's task was to create data videos using DataTV and the data visualization randomly assigned to them. Participants were required to use at least one interactive visualization in their video.

The resulting data video would ideally make use of multiple media types, interactive visualization, and innovative storytelling techniques.

C.2.1 Data Set One

For this data set, we use TimeFork [207], an interactive visual prediction technique to support users exploring the future of time-series data. Here the expert use TimeFork to create a narrative for tech market stocks (here Apple and Netflix).

C.2.2 Data Set Two

The CommentIQ [208] system is designed to help community moderators manage large amount of comments associated with online articles by automatically ranking them based on criteria such as relevance, readability, personal experience, and length.Here the expert wants to author a streaming data video about the community response on an article from The New York Times[1] titled "City Reacts: State of

[1] `http://www.nytimes.com/`

Emergency" during the 2015 racial unrest in Ferguson, Missouri following the death of Michael Brown at the hands of a white policeman.

Bibliography

[1] Zhenpeng Zhao, Rachael Marr, Jason Shaffer, and Niklas Elmqvist. Understanding partitioning and sequence in data-driven storytelling. In *Proceedings of the iConference*, 2019. To appear.

[2] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[3] John Snow. *1854 Broad Street cholera outbreak*. 1855.

[4] Charles Joseph Minard. *Graphic Storytelling and Visual Narrative*. 1812.

[5] Hurricane irma news, 2017. https://www.youtube.com/watch?v=KDmFbAhlh3w.

[6] Is height all in our gene, 2018. https://www.youtube.com/watch?v=0cuO5OSDMbw&t=328s.

[7] Ancient greece in 18 minutes, 2017. https://www.youtube.com/watch?v=gFRxmi4uCGo.

[8] The history of asia: every year, 2017. https://www.youtube.com/watch?v=c8TNvvjoqvw.

[9] Wealth inequality in america, 2012. https://www.youtube.com/watch?v=QPKKQnijnsM.

[10] The joy of stats, 2010. https://www.youtube.com/watch?v=jbkSRLYSojo.

[11] Religions and babies, 2012. https://www.youtube.com/watch?v=ezVk1ahRF78.

[12] Gene pool decline, 2018. https://www.youtube.com/watch?v=k2N4ZO57fjE.

[13] How to end poverty, 2015. https://www.youtube.com/watch?v=5JiYcV_mg6A.

[14] China's geography problem, 2017. https://www.youtube.com/watch?v=GiBF6v5UAAE.

[15] Imaginary numbers are real, 2015. https://www.youtube.com/watch?v=T647CGsuOVU.

[16] Big data revolution, 2013. https://www.youtube.com/watch?v=bIY3LUZ7i8Y.

[17] The truth about population, 2013. https://www.youtube.com/watch?v=QpdyCJi3Ib4.

[18] Inside the mind of a master procrastinator, 2016. https://www.youtube.com/watch?v=arj7oStGLkU&t=588s.

[19] How data will transform business, 2014. https://www.youtube.com/watch?v=EHTmxmuhZ10.

[20] Will saving poor children lead to overpopulation, 2012. https://www.gapminder.org/answers/will-saving-poor-children-lead-to-overpopulation/.

[21] Phd comic your life ambition, 2017. http://phdcomics.com/comics/archive.php?comicid=1012.

[22] Nfl player data report, 2017. https://public.tableau.com/profile/mikevizneros#!/vizhome/IsThatRight/IsThatTrue.

[23] Comic style dashboard, 2012. https://atrowpoole.wordpress.com/portfolio/data-visualization/.

[24] Infographic comic, 2010. https://xkcd.com/681/.

[25] The new york restaurant vis, 2017. https://www.menglugao.com/blog/2017/12/7/new-york-city-restaurants-data-visualization.

[26] Marvel vs dc comics, 2018. http://jobloving.com/infographics/data-visualization/data-visualization-infographics-marvel-vs-dc-comics/.

[27] Body cartoon, 2018. https://www.cartoonstock.com/cartoonview.asp?catref=jcen1296/.

[28] Spider man comic visualization, 2018. https://vignette.wikia.nocookie.net/marveldatabase/images/b/bc/Iron_Spider_Armor_V1.1_from_Official_Handbook_of_the_Marvel_Universe_Vol_5_Spider-Man_-_Back_in_Black.jpg/revision/latest?cb=20141204030850/.

[29] Cell phone comic, 2018. https://vignette.wikia.nocookie.net/marveldatabase/images/b/bc/Iron_Spider_Armor_V1.1_from_Official_Handbook_of_the_Marvel_Universe_Vol_5_Spider-Man_-_Back_in_Black.jpg/revision/latest?cb=20141204030850/.

[30] Linear regression comic, 2018. https://xkcd.com/1725/.

[31] Comic for vocation stress, 2016. https://twitter.com/pvermeul_peter/status/756524771702149121.

[32] Comic for desk entropy, 2005. http://phdcomics.com/comics/archive.php?comicid=575.

[33] Comic for phd grooming, 2010. https://strangelyincoherentloveletters.files.wordpress.com/2010/12/phd061209s.gif.

[34] Comic for phd procrastination, 2016. https://substance.etsmtl.ca/en/power-procrastination-according-phd-comics.

[35] A day of an american life, 2019. https://flowingdata.com/2019/04/02/data-comic-shows-an-average-american-day/.

[36] Curve fitting comic, 2018. https://xkcd.com/2048/.

[37] Seashell comic, 2018. https://xkcd.com/1236/.

[38] Strikeouts on the rise, 2017. http://www.nytimes.com/interactive/2013/03/29/sports/baseball/Strikeouts-Are-Still-Soaring.html.

[39] 1.5 million missing black men, 2017. https://www.nytimes.com/interactive/2015/04/20/upshot/missing-black-men.html.

[40] Bum Chul Kwon, Florian Stoffel, Dominik Jckle, Bongshin Lee, and Daniel Keim. Visjockey: Enriching data stories through orchestrated interactive visualization. In *Proceedings of the Symposium on Computation+Journalism*, January 2014.

[41] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Hollerer, and Eun Kyoung Choe. ChartAccent: Annotation for data-driven storytelling. In *Proceedings of the IEEE Pacific Symposium on Visualization*. IEEE, April 2017.

[42] Zhenpeng Zhao, William Benjamin, Niklas Elmqvist, and Karthik Ramani. Sketcholution. *International Journal of Human-Computer Studies*, 82(C):11–20, October 2015.

[43] Datasketches–royal constellations, 2010. http://www.datasketch.es/october/code/nadieh/.

[44] Datasketches–carcaptor sakura, 2010. http://www.datasketch.es/june/code/nadieh/.

[45] The big short movie explained animated, 2016. https://www.youtube.com/watch?v=UFlHwkiAmyU.

[46] Population decline in new orleans, 2017. http://www.nytimes.com/interactive/2011/02/03/us/0203-nat-census-orleans.html.

[47] The ny times top 10 bestsellers, 2017. http://blog.leeandlow.com/2013/12/10/wheres-the-diversity-the-ny-times-top-10-bestsellers-list/.

[48] Non profit revenue report, 2017. goo.gl/HsvQDB.

[49] Attack of the little people, 2011. https://graphicviolence.wordpress.com/2011/09/18/attack-of-the-little-people/.

[50] London marches, 2018. https://www.slow-journalism.com/infographics/infographic-londons-largest-protest-marches.

[51] Yemen civil war, 2018. hhttps://www.slow-journalism.com/infographics/map-the-conflict-in-yemen-june-2018.

[52] Global space industry, 2018. https://www.slow-journalism.com/infographics/infographic-starship-enterprises-charting-the-new-space-race.

[53] Air pollution for household, 2016. https://www.who.int/airpollution/infographics/Air-pollution-INFOGRAPHICS-English-5-1200px.jpg?ua=1.

[54] Air pollution linked death, 2016. https://www.who.int/airpollution/infographics/Air-pollution-INFOGRAPHICS-English-2-1200px.jpg?ua=1.

[55] North and south korean comparison, 2012. https://www.slow-journalism.com/infographics/the-great-divide-north-and-south-korea-compared.

[56] In the shadow of foreclosure, 2017. http://infographicsnews.blogspot.com/2009/02/.

[57] Words democrats and republican used, 2011. https://archive.nytimes.com/www.nytimes.com/interactive/2008/09/04/us/politics/20080905_WORDS_GRAPHIC.html?_r=0.

[58] Uk and us firearms, 2014. https://www.statista.com/chart/2628/police-firearms-discharges/.

[59] White correspondent dinner, 2015. https://www.6sqft.com/what-nycs-population-looks-like-day-vs-night/.

[60] Day and night: Nyc population, 2015. https://www.6sqft.com/what-nycs-population-looks-like-day-vs-night/.

[61] Who owns everything: Big data today, 2015. https://www.6sqft.com/what-nycs-population-looks-like-day-vs-night/.

[62] Big welsh coast walk, 2015. https://graphs.net/easel-ly-infographics.html.

[63] Hangry usa, 2015. https://kfbk.iheart.com/content/2018-01-10-california-is-hangry-are-you/.

[64] Nyc celebrity map, 2015. https://www.addressreport.com/blog/nyc-celebrity-map-star-map/.

[65] Bungie, inc.,halo: Reach, 2010. Microsoft Game Studios.

[66] Treyarch, call of duty: Black ops, 2010. Activision.

[67] Starcraft ii: Wings of liberty, 2010. Blizzard.

[68] Twittersheep, 2017. http://www.twittersheep.com/.

[69] Game of thrones discussion of twitter, 2017. http://www.twittersheep.com/.

[70] How tweets spread, 2017. https://interactive.twitter.com/tenyears/#?lang=EN.

[71] Uber mobile visualization, 2016. https://hackernoon.com/can-augmented-reality-solve-mobile-visualization-f06c008f8f84.

[72] Ar data visualization design, 2016. http://www.jolamux.com/v3.5/works/ar/dataViz.html.

[73] Ar 3d design, 2010. https://www.researchgate.net/figure/A-conceptual-image-of-AR-overlay-of-3D-design-and-contextual-data-Dunston-Sh
fig4_221906912.

[74] Ar flight data, 2016. https://hololens.reality.news/news/holoflight-turns-flight-data-into-cool-mixed-reality-visualizations-0173138/

[75] Ar street visualization, 2017. https://hackernoon.com/silent-augmented-reality-f0f7614cab32.

[76] Ar pipeline, 2017. http://thearea.org/augmented-reality-and-the-internet-of-things-boost-human-performance/.

[77] Ar underground infrastructure, 2018. https://communities.bentley.com/other/old_site_member_blogs/bentley_employees/b/stephanecotes_blog/posts/augmentation-of-subsurface-utilities-the-problem-of-spatial-perception.

[78] Ar bio-chemical vis, 2018. https://ideastations.org/science-matters/science-news/augmented-reality-revolutionizes-surgery-and-data-visualization

[79] Adobe vr 3d design, 2018. https://edgylabs.com/project-new-view-leverages-vr-ai-tools-for-3d-immersive-data-visualization.

[80] Vr baseball training, 2018. https://www.baseballamerica.com/stories/better-data-equals-better-training-with-trinityvr/.

[81] Vr big data visualization, 2016. https://www.youtube.com/watch?v=wacNaAVGXdU.

[82] Vr lens for big data, 2016. https://www.datanami.com/2015/03/09/a-virtual-reality-lens-for-big-data-visualization/.

[83] C. Donalek, S. G. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, S. Davidoff, J. S. Norris, and G. Longo. Immersive and collaborative data visualization using virtual reality platforms. In *Proceedings of the IEEE International Conference on Big Data*, pages 609–614, Oct 2014.

[84] Vr bio-informatics, 2017. https://vrsconference.com/2017/10/not-playing-games-enterprise-vr/bio-informatics-data-visualization-at-greenlight-insights-vrs-2017/.

[85] Vr geo map visualization, 2015. https://ocean.sagepub.com/blog/2018/6/20/experimenting-with-data-visualization-in-vr/.

[86] Vr data analysis, 2018. https://www.springwise.com/vr-enables-immersive-3d-data-analysis/.

[87] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, CA, 1999.

[88] Zachary Pousman, John T. Stasko, and Michael Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, 2007.

[89] Jean-Daniel Fekete and Catherine Plaisant. Interactive information visualization of a million items. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 117–124, 2002.

[90] Chris Stolte, Diane Tang, and Pat Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):52–65, 2002.

[91] Christopher Ahlberg. Spotfire: An information exploration environment. *SIGMOD Record*, 25(4):25–29, 1996.

[92] Qlik. `https://www.qlik.com/`.

[93] Fernanda Viégas and Martin Wattenberg. Communication-minded visualization: A call to action. *IBM Systems Journal*, 45(4):801–812, 2006.

[94] Okyay Kaynak and Shen Yin. Big data for modern industry: Challenges and trends [point of view]. *Proceedings of the IEEE*, 103:143–146, 02 2015.

[95] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. Big data: The next frontier for innovation, competition, and productivity. 05 2011.

[96] C.L. Philip Chen and Chun-Yang Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275:314 – 347, 2014.

[97] Ahmed Oussous, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, and Samir Belfkih. Big data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4):431 – 448, 2018.

[98] Sachchidanand Singh and Nirmala Singh. Big data analytics. *2012 International Conference on Communication, Information & Computing Technology (ICCICT)*, pages 1–4, 2012.

[99] J. Ahrens, B. Hendrickson, G. Long, S. Miller, R. Ross, and D. Williams. Data-intensive science in the us doe: Case studies and future challenges. *Computing in Science Engineering*, 13(6):14–24, Nov 2011.

[100] Katie Shilton. Values and ethics in human-computer interaction. *Foundations and Trends HumanComputer Interaction*, 12(2):107–171, 2018.

[101] Andrej Zwitter. Big data ethics. *Big Data & Society*, 1(2):2053951714559253, 2014.

[102] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, and William Money. Big data: Issues and challenges moving forward. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 995–1004, Washington, DC, USA, 2013.

[103] S. Sagiroglu and D. Sinanc. Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 42–47, May 2013.

[104] Griffin K. Gerhardt, B. and R. Klemann. Unlocking value in the fragmented world of big data analytics. *Cisco Internet Business Solutions Group*, 2012.

[105] Vasant Dhar. Data science and prediction. *Communactions of the ACM*, 56(12):64–73, 2013.

[106] Hours uploaded, 2014. https://tubularinsights.com/300-hours-video-youtube-advertisers/.

[107] Hours watched, 2014. https://techcrunch.com/2017/02/28/people-now-watch-1-billion-hours-of-youtube-per-day/.

[108] Cellphone more powerful than old NASA computers, 2014. https://www.zmescience.com/research/technology/smartphone-power-compared-to-apollo-432/.

[109] Michael J. Pazzani and Daniel Billsus. *Content-Based Recommendation Systems*, pages 325–341. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[110] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. The YouTube video recommendation system. In *Proceedings of the ACM Conference on Recommender Systems*, pages 293–296, New York, NY, USA, 2010. ACM.

[111] Andreas Wittig and Michael Wittig. *Amazon Web Services in Action.* Manning Publications Co., Greenwich, CT, USA, 1st edition, 2015.

[112] Sherif Talaat. *Pro PowerShell for Microsoft Azure*. Apress, Berkely, CA, USA, 1st edition, 2015.

[113] Asit K. Mishra, Joseph L. Hellerstein, Walfredo Cirne, and Chita R. Das. Towards characterizing cloud backend workloads: Insights from google compute clusters. *SIGMETRICS Performance Evaluation Review*, 37(4):34–41, March 2010.

[114] Jill Freyne and Barry Smyth. Visualization for the masses: Learning from the experts. In *Case-Based Reasoning. Research and Development*, pages 111–125, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[115] Digit divide. https://www.internetworldstats.com/links10.htm.

[116] Digit divide definition. https://cs.stanford.edu/people/eroberts/cs181/projects/digital-divide/start.html.

[117] Phone vs computer, 2014. https://www.phonearena.com/news/A-modern-smartphone-or-a-vintage-supercomputer-which-is-more-powerful_id57149.

[118] Murray Campbell, A. Joseph Hoane, and Feng hsiung Hsu. Deep blue. *Artificial Intelligence*, 134(1):57 – 83, 2002.

[119] Robert Kosara. Story points in Tableau Software. Keynote at Tableau Customer Conference, September 2013.

[120] Spotfire. `https://www.tibco.com/products/tibco-spotfire`.

[121] Peter Fox and James Hendler. Changing the equation on scientific data visualization. *Science*, 331(6018):705–708, 2011.

[122] N Erdemir. The effect of PowerPoint and traditional lectures on students' achievement in physics. *Journal of Turkish Science Education*, 8:176–189, 09 2011.

[123] Dan Murray. *Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software*. Wiley Publishing, 1st edition, 2013.

[124] Stephen Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, USA, 1st edition, 2009.

[125] Pu Shen. The p/e ratio and stock market performance. *Economic Review*, pages 23–36, 01 2000.

[126] M Silver, T Sakata, H C Su, C Herman, Steven Dolins, and M J O'Shea. Case study: How to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management*, 15:155–64, 02 2001.

[127] Miriam Lux. Visualization of financial information. In *Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation*, pages 58–61, New York, NY, USA, 1997. ACM.

[128] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale. *Data-Driven Storytelling*. A. K. Peters, Ltd., Natick, MA, USA, 1st edition, 2018.

[129] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[130] Steven F. Roth. Capstone address: Visualization as a medium for capturing and sharing thoughts. In *Proceedings of the IEEE Symposium on Information Visualization*, 2004.

[131] Peter R. Keller and Mary M. Keller. *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press, Los Alamitos, CA, USA, 1994.

[132] Chun-houh Chen, Wolfgang Hrdle, Antony Unwin, Chun-houh Chen, Wolfgang Hrdle, and Antony Unwin. *Handbook of Data Visualization (Springer Handbooks of Computational Statistics)*. Springer-Verlag TELOS, Santa Clara, CA, USA, 1 edition, 2008.

[133] Michael Friendly. A brief history of data visualization. *Handbook of Computational Statistics: Data Visualization*, III, 2007.

[134] Thomas A. Defanti and Maxine D. Brown. Visualization in scientific computing. volume 33 of *Advances in Computers*, pages 247–307. Elsevier, 1991.

[135] Dan Murray. *Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software*. Wiley Publishing, 1st edition, 2013.

[136] M. Adil Yalcin, Niklas Elmqvist, and Benjamin B. Bederson. Keshif: Rapid and expressive tabular data exploration for novices. *IEEE Transactions on Visualization and Computer Graphics*, 2017.

[137] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[138] Google charts api, 2007. `https://developers.google.com/chart/`.

[139] Arvind Satyanarayan, Kanit Wongsuphasawat, and Jeffrey Heer. Declarative interaction design for data visualization. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 669–678, 2014.

[140] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2017.

[141] Andy Kirk. *Data Visualization: A Successful Design Process*. Community experience distilled. Packt Pub., 2012.

[142] Martin Wattenberg. Baby names, visualization, and social data analysis. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 1–7, 2005.

[143] Many eyes, 2007. `http://www.boostlabs.com/ibms-many-eyes-online-data-visualization-tool/`.

[144] Infogram, 2017. `https://infogram.com/`.

[145] Jonathan Gottschall. *The Storytelling Animal: How Stories Make Us Human*. Mariner Books, 2012.

[146] Roger C. Schank and Robert P. Abelson. Knowledge and memory: The real story. In Jr. Robert S. Wyer, editor, *Knowledge and Memory: The Real Story*, pages 1–85, Hillsdale, NJ, 1995. Lawrence Erlbaum Associates.

[147] Jan Vansina. *Oral Tradition as History*. University of Wisconsin Press, Madison, WI, 1985.

[148] Thomas M. Leitch. *What Stories Are: Narrative Theory and Interpretation.* Pennsylvania State University Press, University Park, PA, 1986.

[149] Will Eisner. *Graphic Storytelling and Visual Narrative.* W. W. Norton & Company, 2008.

[150] David Sless. *Learning and Visual Communication.* Wiley, 1981.

[151] Benjamin Bach, Natalie Kerracher, Kyle Wm. Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. Telling stories about dynamic networks with Graph Comics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 3670–3682, 2016.

[152] Zhenpeng Zhao, Rachael Marr, and Niklas Elmqvist. Data comics: Sequential art for data-driven storytelling. Technical Report HCIL-15-15, University of Maryland, College Park, October 2015.

[153] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1459–1468, 2015.

[154] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernández, and Pourang Irani. Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):501–510, 2017.

[155] Robert L. Harris. *Information Graphics: A Comprehensive Illustrated Reference.* Oxford University Press, Oxford, United Kingdom, 1999.

[156] Scott McCloud. Comics: A medium in transition. *Computer Graphics Forum*, 30(3):xiii, 2011.

[157] Caroline Ziemkiewicz and Robert Kosara. *Embedding Information Visualization within Visual Representation.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[158] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, and etc. *Visual Analytics: Definition, Process, and Challenges*, pages 154–175. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[159] James J. Thomas and Kristin A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* IEEE Computer Society, 2005.

[160] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, Sept 1996.

[161] Melanie Tory and Torsten Moller. Rethinking visualization: A high-level taxonomy. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 151–158, 2004.

[162] Stuart K. Card and Jock Mackinlay. The structure of the information visualization design space. In *Proceedings of the IEEE Conference on Visualization*, pages 92–99, Oct 1997.

[163] Florence Nightingale. *The causes of mortality in the army in the East*.

[164] Jeremy Boy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, 2014.

[165] Fernanda B. Viégas, Scott Golder, and Judith Donath. Visualizing email content: portraying relationships from conversational histories. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 979–988, 2006.

[166] Doantam Phan, Andreas Paepcke, and Terry Winograd. Progressive multiples for communication-minded visualization. In *Proceedings of Graphics Interface*, pages 225–232, 2007.

[167] Nahum D. Gershon and Ward Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001.

[168] Nick Diakopoulos, Joan DiMicco, Jessica Hullman, Karrie Karahalios, and Adam Perer. Telling stories with data: The next chapter—a visweek 2011 workshop, 2011.

[169] Joan DiMicco, Matt McKeon, and Karrie Karahalios. Telling stories with data—a visweek 2010 workshop, 2010.

[170] Jessica Hullman and Nicholas Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, 2011.

[171] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. SketchStory: Telling more engaging stories with data through freeform sketching. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2416–2425, 2013.

[172] Waqas Javed and Niklas Elmqvist. ExPlates: Spatializing interactive analysis to scaffold visual exploration. *Computer Graphics Forum*, 32(3pt4):441–450, 2013.

[173] Jessica Hullman, Steven M. Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2406–2415, 2013.

[174] Jing Jin and Pedro A. Szekely. Interactive querying of temporal data using a comic strip metaphor. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 163–170, 2010.

[175] Scott McCloud. *Understanding Comics: The Invisible Art.* William Morrow Paperbacks, 1994.

[176] Neil Cohn. *The Visual Language of Comics: Introduction to the structure and cognition of sequential images.* Bloomsbury, London, 2014.

[177] Neil Cohn. Beyond speech balloons and thought bubbles: The integration of text and image. *Semiotica*, 197:35–63, 2013.

[178] Ariel Dorfman and Armand Mattelart. *How to Read Donald Duck: Imperialist Ideology in the Disney Comic.* Intl General, 1984.

[179] Hans-Christian Christiansen. Comics and film: a narrative perspective. In Anne Magnussen and Hans-Christian Christiansen, editors, *Comics & culture: Analytical and theoretical approaches to comics*, pages 107–122. Museum Tusculanum Press, University of Copenhagen, 2000.

[180] Jing Jin and Pedro A. Szekely. QueryMarvel: A visual query language for temporal patterns using comic strips. In *Proceedings of the IEEE Conference on Visual Languages and Human-Centered Computing*, pages 207–214, 2009.

[181] Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J. McGuffin, and Hanspeter Pfister. Datatoon: Drawing dynamic network comics with pen + touch interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 105:1–105:12, New York, NY, USA, 2019. ACM.

[182] Antoni B. Moore, Mariusz Nowostawski, Christopher Frantz, and Christina Hulbe. Comic strip narratives in time geography. *ISPRS International Journal of Geo-Information*, 7(7), 2018.

[183] C. Bryan, K. Ma, and J. Woodring. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):511–520, Jan 2017.

[184] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 253:1–253:12, New York, NY, USA, 2019. ACM.

[185] Zezhong Wang, Harvey Dingwall, and Benjamin Bach. Teaching data visualization and storytelling with data comic workshops. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pages CS26:1–CS26:9, New York, NY, USA, 2019. ACM.

[186] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. Design patterns for data comics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 38:1–38:12, New York, NY, USA, 2018. ACM.

[187] Sandvine. Global internet phenomena report. Technical report, Sandvine Incorporated, December 2015.

[188] David M. Ewalt. The ESPN of video games. *Forbes*, (December 2), 2013.

[189] Stop marine plastic pollution, 2014. `https://www.youtube.com/watch?v=02WjKxk1veQ`.

[190] Sorting algorithms shown by dance, 2017. `http://www.pdviz.com/different-sorting-algorithm-demonstrated-with`.

[191] Celine Latulipe, David Wilson, Sybil Huskey, Berto Gonzalez, and Melissa Word. Temporal integration of interactive technology in dance: Creative process impacts. In *Proceedings of the ACM Conference on Creativity & Cognition*, pages 107–116, 2011.

[192] Ronald M. Baecker. *Readings in Groupware and Computer-Supported Cooperative Work*. Morgan Kaufmann Publishers, San Francisco, 1993.

[193] A beautiful planet, 2016. `http://www.imdb.com/title/tt2800050/`.

[194] The two americas, 2017. `https://www.nytimes.com/interactive/2016/11/16/us/politics/the-two-americas-of-2016.html?smid=pl-share&_r=0`.

[195] Storyfy, 2017. `https://storify.com/`.

[196] Power bi, 2014. `https://powerbi.microsoft.com/en-us/blog/tag/pdf/`.

[197] Youtuber pewdiepie. `https://www.youtube.com/channel/UC-lHJZR3Gqxm24_Vd_AJ5Yw`.

[198] Robert Kosara and Jock D. Mackinlay. Storytelling: The next step for visualization. *IEEE Computer*, 46(5):44–50, 2013.

[199] Larry Gonick and Art Huffman. *The Cartoon Guide to Physics*. HarperPerennial, New York, 1990.

[200] Larry Gonick and Woollcott Smith. *The Cartoon Guide to Statistics*. HarperCollins, New York, 1993.

[201] Ben Shneiderman. Tree visualization with tree-maps: A 2-D space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99, January 1992.

[202] Alfred Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

[203] John P. Collomosse, D. Rowntree, and P. M. Hall. Rendering cartoon-style motion cues in post-production video. *Graphical Models*, 67(6):549–564, 2005.

[204] Melanie Tory and Torsten Möller. Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, 25(5):8–11, 2005.

[205] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher A. Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 2573–2582. ACM, 2010.

[206] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, Dec 2013.

[207] Sriram Karthik Badam, J. Zhao, N. Elmqvist, and D. S. Ebert. TimeFork: Interactive prediction of time series. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 5409–5420, 2016.

[208] Deok Gun Park, Simranjit Singh, Nicholas Diakopoulos, and Niklas Elmqvist. Supporting comment moderators in identifying high quality online news comments. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1114–1125, 2016.