Maryland Department of Natural Resources
Partner Network Requirements and Data Assessment

Authors: Avram Baicu, Amanu Huq, Jennifer Tatham, Farhana Uddin, Ady Weng

Under the supervision of TJ Rainsford

Course: INST490 Integrated Capstone for Information Science

College of Information Studies
University of Maryland - College Park

December 2023

PALS
An initiative of the National Center for Smart Growth
Kathryn Howell, NCSG Executive Director
Kim Fisher, PALS Director

Contents

Abstract

The Maryland Department of Natural Resources' (MDNR) Chesapeake & Coastal Service (CCS) unit ensures that the state is equipped to deal with the needs and demands of Chesapeake economic and coastal resource use. CCS supports MDNR goals by partnering with various stakeholders, including federal and state government agencies; regional, national, and international institutions; formal and informal educational programs; nongovernmental organizations; and the private sector.

The iConsultancy Team was tasked with assisting CCS in their partnerships with stakeholders. Namely, CCS requested visualizations of data collected about partner organizations to help enable CCS staff to better understand the composition of their partner network (e.g., geography, service).

To accomplish this, the team cleaned and assessed a large data set of interactions between CCS and its stakeholder communities so it could be better visualized, conducted user research with key CCS staff to better understand and define their goals for stakeholder interactions, and provided recommendations for information gaps and additional data collection.

The team worked with the MDNR's PALS team—Marlo Atkinson, Kate Vogel, and Dylan Tallie—throughout the project.

Deliverables

The project comprised three primary deliverables: a normalized clean data set, interactive visualizations, and a collection of static, immutable visualizations. Other independent deliverables are mentioned in the scope of work.

The first phase was to clean the *Community and Partners* dataset, which was compiled from responses to a CCS Survey123 of their stakeholders. After discussion with CCS, the iConsultancy Team identified seven categories in the raw dataset to clean and use for analysis:

- CCS Contact Name
- County
- Organizations
- Engagement Category
- Engagement
- and two columns for x and y coordinates.

After analyzing and exploring the data, the team proceeded to create data subsets for cleaning and analysis.

The interactive visualizations were completed in Tableau. They comprised a table that filters partner organizations by county and engagement category, a map depicting the geographical distribution of partner organizations that facilitates easy information access, and an interactive table organized by organizations and contact information. It is important to note that feedback and user interviews came from the CCS clients and the directors. Individual stakeholder clients were not interviewed, and all participants were made aware of this in the project's initial stages.

Method

The CCS contact name category required normalization for analysis. For instance, pairs of names are in different order. Some rows used slash separators, and some didn't. All name pairs were put in the same order and a uniform separator was used to prepare for data analysis.

The engagement category, which had multiple values, was managed by splitting it into distinct columns. In most database management software columns with multiple values are difficult to search and filter. For instance, an organization listed with a single engagement category like "Climate, DEIJ" as a unified entity wouldn't be easily searchable for activities related to both Climate and DEIJ. Splitting engagement categories into separate columns allows for a more detailed representation of each organization's involvement and ensures accurate attribution to distinct categories. This strategic restructuring improved searchability and enabled the team to showcase and effectively use specific organizations and their corresponding engagement categories.

After the categories were successfully split each engagement category was quantified. These categories were translated into numerical form and their total counts were aggregated using a pivot table. Despite repeating counties due to the multiplication of engagement columns, a new sheet further aggregated the data, yielding comprehensive totals for each county. By leveraging various functions, the team generated additional tables and calculations to create ranks, charts, and graphs.

Finally, the x and y coordinate column categories were missing many coordinates. The missing coordinates were filled in by referencing the county they were assigned to. A new column indicated which ones were added. It's important to note that there were missing counties, and they could not be assigned because multiple organizations were assigned to a single county. Since geographical coordinates are important for visualizations, filler coordinates were used with Maryland coordinates, which is noted in the add-in column.

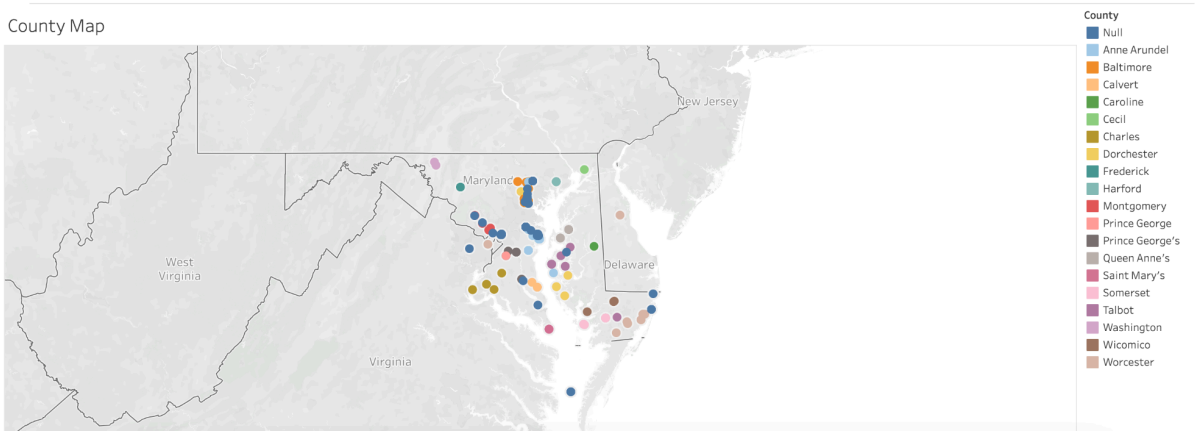
Interactive Visualizations

CCS Interactive Map

The map in Figure 1 represents CCS contact names, with organization and county. The team thought this would be the best visually comprehensive tool. The geographical x-y categorizes each colored dot with coordinates from the raw data. When a point is toggled, contact name, organization and county appear.

The interactive legend (on the right) depicts counties by color on the map. The map also allows users to filter by county they want to see on the map. For example, if a user wants to see all existing queries for Anne Arundel County, they click on the small pen at the top right of the County legend and then on Anne Arundel.

Figure 1



Method

Since geographical location was an important aspect of this data, it was the map's primary marker. The goal was to provide a visual representation of CCS's different partnerships bonding information.

To create the map, x-coordinates were set in columns, and y-coordinates were set in rows. To prepare the x and y coordinate variables, the geographical role was changed to longitude and latitude, respectively, on the Tableaus Table tab. Then, we inserted the "County" column in the Color Marks Category to color-code all the counties.

To ensure the county legend is present, right-click "County" in the Marks table and ensure there is a check next to "Include in Tooltip." The team put the CCS contact name and organization columns in the Detail Marks category so CCS information would be present when toggled. To ensure all the points are on the map, click on the Analysis dropdown on the header, and make sure that "Aggregate Measures" is *not checked.* The missing counties are present because the

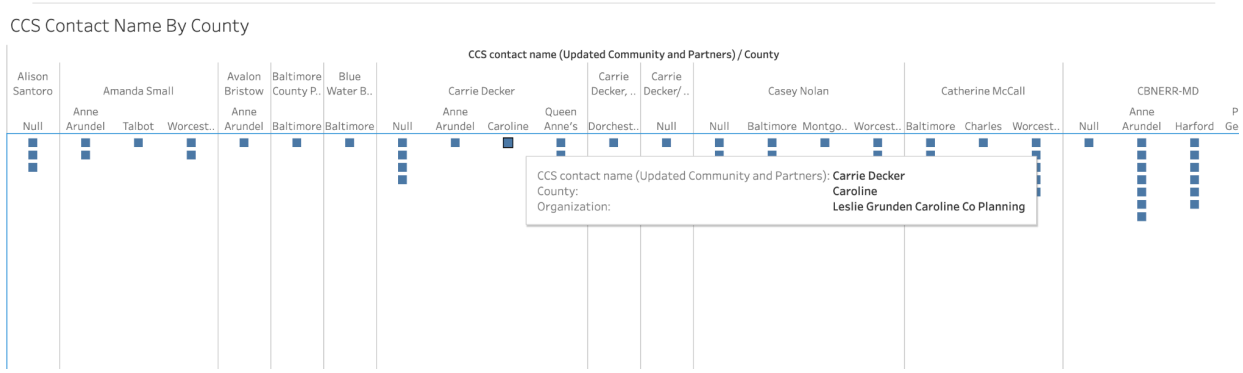
data cleaning process could not be filled in due to multiple organizations being present under one contact. Hence, filler coordinates of Maryland's geographical coordinates were used so they could have a placeholder in the map.

Contact Name Dashboard

The dashboard for Contact Name by County is an interactive way to identify the counties and organizations CCS has worked with. The dashboard's categories are CCS contact names and the blue boxes depict the counties they've worked with.

When toggled, county and organization, information appears. The dashboard is scrollable from left to right. The null values represent the absence of certain counties in the data, which could not be filled during the data processing stage due to ambiguities identified during data cleaning. There are no quantitative findings in this interactive dashboard because this dashboard is intended as a tool for users.

Figure 2



Method

To create this dashboard, CCS Contact Names and County were put in the column category. This created the breakdown of contact names and the counties they worked with. Then, the organization column was put in the Detail category in the Markers Table to present additional information when the blue boxes are toggled. As mentioned, the Null Values Category represents the missing counties in the data cleaning process that could not be filled in due to multiple organizations present under one contact. Hence, filler coordinates of Maryland's geographic coordinates were used.

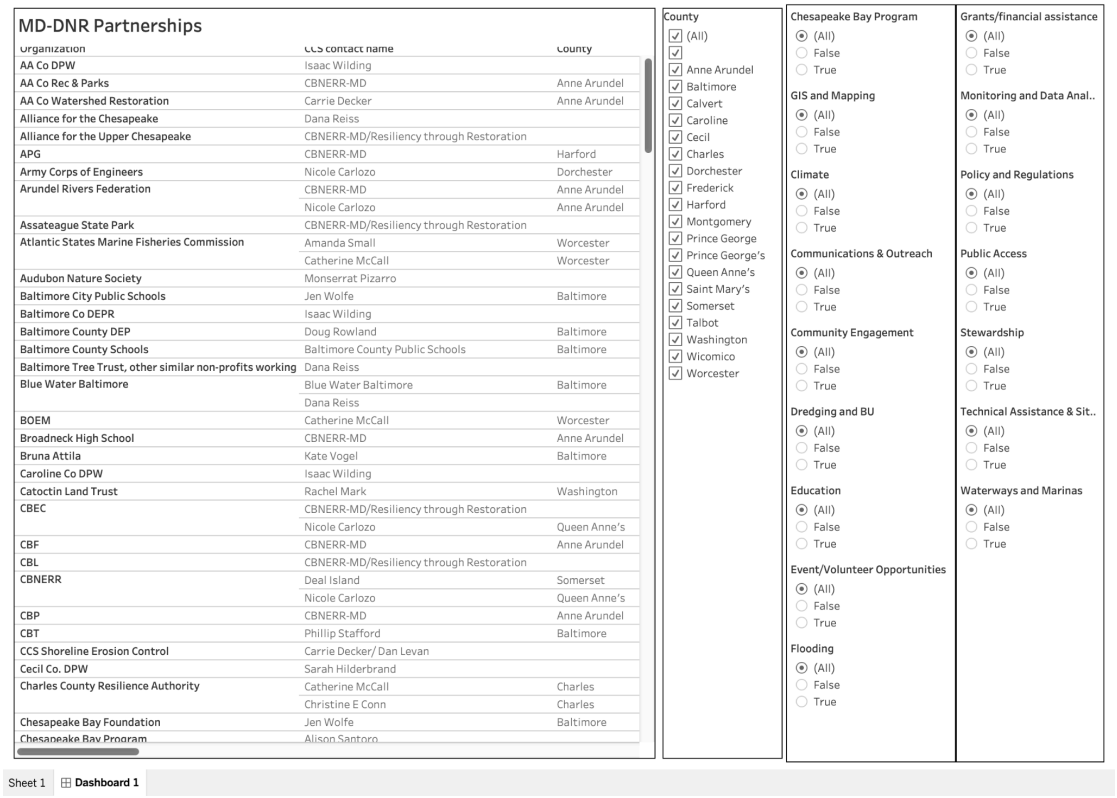
Filterable Dashboard

The interactive dashboard in Figure 3 lists organizations and their associated CCS contact names and counties. This dashboard enables the clients to select specific "views" of organizations based on their county and categories they engage while excluding non-relevant values. For instance, if

a CCS staff member wants to find all partner organizations who work in Baltimore in the climate engagement category, they would select those respective checkboxes and radio buttons.

This visualization doesn't provide any quantitative analysis (as the static visualizations do) but does give CCS staff an efficient way to search for the organizations based on specified criteria. Consequently, it advances the project's goal to form data visualizations that empower CCS staff to better understand their partner organizations.

Figure 3



Method

This filterable dashboard was created with suggestions from CCS clients about the most frequently used search criteria used to build new partnerships. CCS staff most frequently search for partner organizations based on their location and engagement categories. Consequently, the team used Tableau to create a searchable table.

First, using Google Sheets, the engagement category data was cleaned by splitting the values of one original engagement category column into 14 distinct values and columns, enabling Tableau to isolate individual engagement categories (*e.g.,* "Climate, DEIJ" vs. "Climate", "DEIJ"). Then, using Tableau's drag-and-drop interface, the Organization, CCS contact name, County, and Engagement Category columns were included in a Tableau sheet (Sheet 1), which was uploaded

into a Tableau dashboard (Dashboard 1). The team selected those columns as filters with distinct values by right-clicking on the engagement categories.

Static Visualizations

Pie Charts

The iConsultancy Team also used Google Sheets to create a series of static visualizations to summarize the partner organizations and facilitate quantitative analysis. For example, the team chose to represent the distribution of partner organizations throughout engagement categories and counties through pie charts.

These charts allow the team to make recommendations about gaps in the CCS coverage. For instance, Figure 4 illustrates that MDNR and CCS should consider partnering with organizations that participate in education, waterways and marinas, policy and regulations, and dredging and BU work.

Figure 4

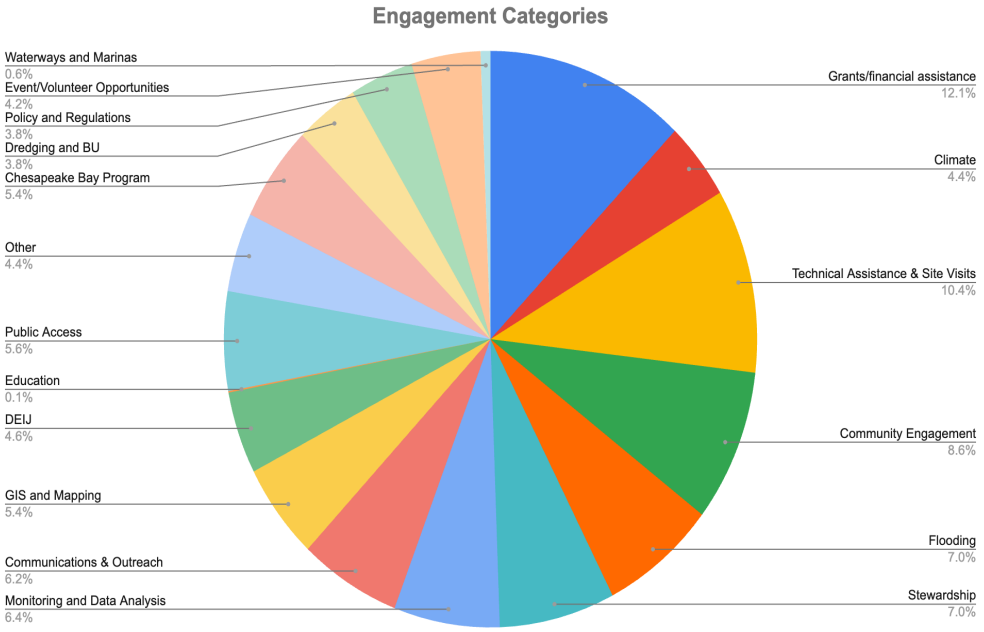
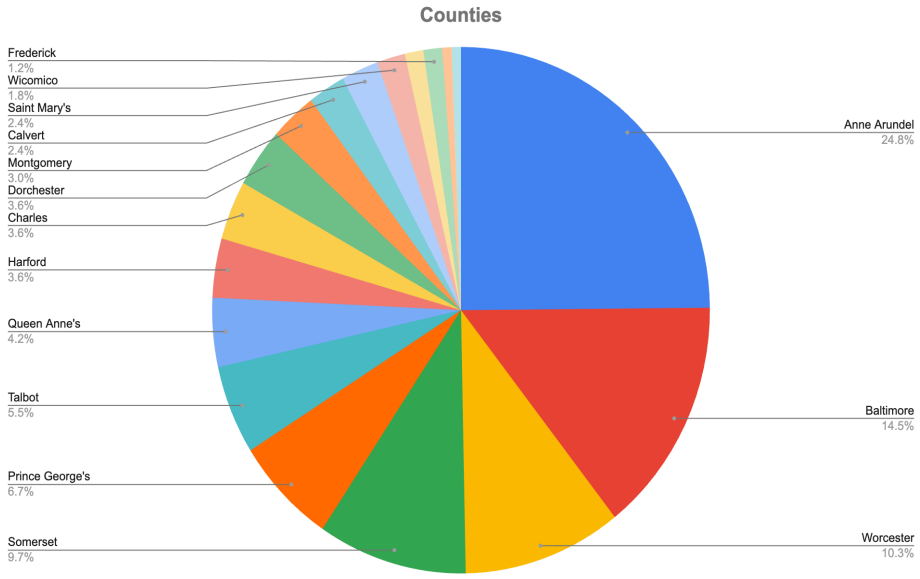


Figure 5 indicates that MDNR and CCS should consider conducting additional outreach and partnerships in Frederick, and in Wicomico, Saint Mary's, and Calvert Counties. Although these counties do not directly abut with the Chesapeake Bay (and might be considered outside the purview of CCS-specific services), CCS might consider education or policy and regulation related programs in those locales.

Figure 5



Method

Data cleaning was necessary to create the data visualizations; the team used Google Sheets to create a sheet of formulas to denote each organization as a “Yes” or “No” for each engagement category. This made it easy to count the number of “Yes” values in an engagement category and facilitated the sorting of partnerships by engagement categories.

The sorted pie chart of partnerships by distribution was created using the same data in Google Sheets. Specifically, the team used the “COUNTIF” function in formulas to count and note the occurrences of specific engagement categories and counties e.g., =COUNTIF('Engagement Categories (Y/N)'!W:W, "Climate"), =COUNTIF('Community and Partners'!D:D, "Anne Arundel").

With these counts in a separate backend spreadsheet, the team then inserted two new pie charts comprising the data containing the counts. For instance, the team inserted a new pie chart representing data from the Counts spreadsheet between cells B2 and B19, which contain all counts of engagement categories. Similarly, the team created a pie chart of data between cells B22 and B39 to represent the distribution of counties in partner organizations.

Dynamic Pivot Tables

The team developed two tables displaying counts of CCS contacts within counties (Figure 6) and by engagement categories (Figure 7). This data empowers CCS staff to quickly identify the most frequently used contacts in each county, enabling them to find the best contacts when approaching new organizations and partnerships.

For instance, in the case depicted in Figure 6, if a CCS staff member wanted to engage in partnerships in Anne Arundel County, they should consider reaching out to Sarah Lane or CBNERR-MD. CCS staff can see that CBNERR-MD and Sarah Lane are also the best contacts for partnering with organizations working on the Chesapeake Bay Program.

```
[7]: df_contact_engagement = melted_df.groupby("Value")["CCS contact name"].value_counts()
     pds.set_option('display.max_rows', len(df_contact_engagement))
     df_contact_engagement
```

Figure 6

	A	B	C
96	Total		105
97	Anne Arundel	Amanda Small	2
98		Avalon Bristow	1
99		Carrie Decker	1
100		CBNERR-MD	7
101		Dylan Taillie	1
102		Elliott Campbell	4
103		Maggie Cavey	1
104		Mayor Gavin Buckly	1
105		MD iMap Technical Committee	1
106		Nicole Carlozo	5
107		NOAA - Maryland	1
108		Pat Gitlin	1
109		Rachel Marks	1
110		Sadie Drescher	1
111		Sarah Lane	9
112		Sasha Land	1
113		Shannon Sprague	2
114		UMCES	1
115	Anne Arundel Total		41
116	Baltimore	Baltimore County Public Schools	1
117		Blue Water Baltimore	1
118		Casey Nolan	4
119		Catherine McCall	3

Figure 7

E	F	G
Engagement Category	***CCS contact name***	Count
Chesapeake Bay Program		
Chesapeake Bay Program	CBNERR-MD	20
	CBNERR-MD/Resiliency through Restoration	12
	Sarah Lane	7
	Phillip Stafford	4
	Jen Wolfe	3
	Alison Santoro	2
	Shannon Sprague	2
	Sasha Land/Nicole Carlozo	2
	Stephanie Tuckfield	1
	UMCES	1
	Olivia Wisner	1
	NOAA	1
	Laura Collard, Mary Westlund	1
	Elliott Campbell	1
	Claudia Donegan	1
	Catherine McCall	1
	Amanda Small	1
Climate		
	CBNERR-MD	20
	Nicole Carlozo	15
	Jackie Specht, The Nature	

Method

For Figure 6, the team used the Pivot Table functionality in Google Sheets to determine the contact names for organizations in each county. The pivot table includes the county and contact name values from the broader Community and Partners spreadsheet, counting each occurrence of the latter within each reference to the former.

For Figure 7 the team used Python (i.e., the Pandas library) to analyze the data because Google Sheets offers limited functionality for developing a table of contact names by engagement category in its current data structure. Within Python, the team "melted down" the original data spreadsheet to ensure each engagement category column value was represented on an individual row with their respective organization. Then, using split-apply-combine methodology, the team created a new data frame that counted the number of occurrences of each CCS contact name by engagement category. Finally, the team manually copied and pasted this table from Python into Google Sheets so the clients could easily access all visualizations simultaneously.

County Ranking Data Table

The team also developed a Google Sheets table to gather the top five engagement categories per county (Figure 8). This gives insight into the most engaged categories, giving the MD-DNR and CCS organizations essential information on strategic decision making and targeted initiatives. They might use this information to enhance engagement in specific counties or to comprehend

the prevalent focuses in those counties. Additionally, they might work to elevate neighboring counties to the same standard as others and pursue various initiatives. Certain counties may not have five engagement categories, but this table gives insight into what they could work toward. Furthermore, it shows the most prominent engagement category per county, further informing our understanding of the specific notable areas requiring more focused attention and tailored strategies. The table indicates climate is a dominant engagement factor in most counties.

Figure 8

County	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Worcester	Climate	GIS and Mapping	Waterways and Marinas	Technical Assistance & Site Visits	Community Engagement
Wicomico	Climate				
Washington	Climate	Technical Assistance & Site Visits			
Talbot	Climate	Public Access			
Somerset	Education	Climate			
Saint Mary's	Climate				
Queen Anne's	Climate	Community Engagement			
Prince George's	Education	DEIJ	Climate		
Prince George	Education				
Montgomery	Climate	Grants/financial assistance	Monitoring and Data Analysis		
Harford	Education				
Frederick	Climate				
Dorchester	Climate	Education	Education	Flooding	Policy and Regulations
Charles	Education	Climate	Waterways and Marinas		
Cecil	Other				
Caroline	DEIJ				
Calvert	Climate				
Baltimore	Education	Climate	Chesapeake Bay Program	GIS and Mapping	Waterways and Marinas
Anne Arundel	Climate	Education	Chesapeake Bay Program	Community Engagement	Community Engagement
Other	DEIJ	Education	Climate	Waterways and Marinas	Technical Assistance & Site Visits

Method

This table originated from the updated Community and Partners dataset, which the team also used to assign numerical values to each engagement category. Using counted engagement categories in the backend (counts for Ranking 2), the team aggregated counties and the total count of engagement categories, to provide an overview of the counties and their engagement categories.

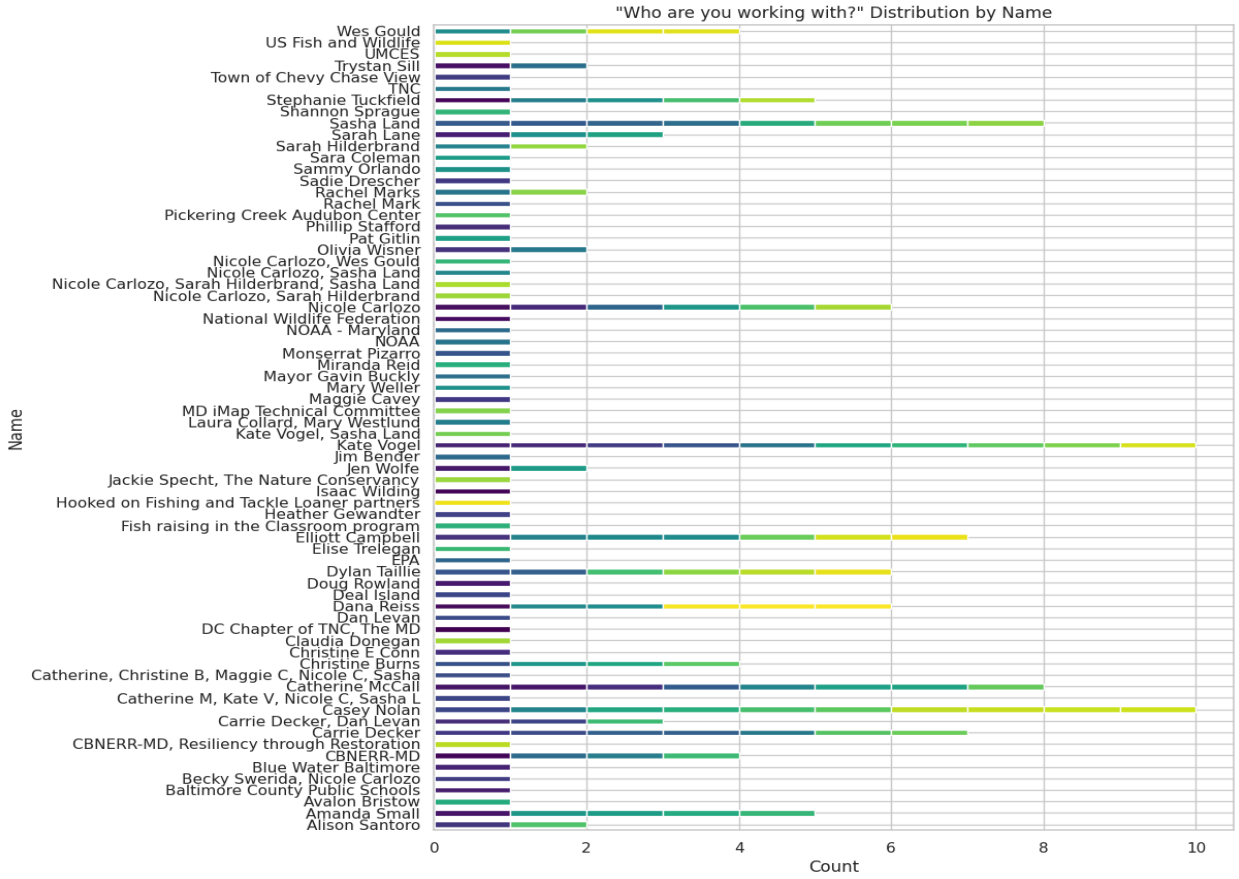
In another table, the team employed functions "ARRAYFORMULA" and "SUMIF" to collect the number of engagement categories per county. Additionally, they utilized "IFERROR" in conjunction with "IF," "LARGE," ">", "INDEX," and "MATCH" to identify the most prominent engagement category. This approach helps identify categories with the highest counts, providing valuable insights and facilitating the gathering of feedback.

Distribution Table

One Community and Partners dataset category is “Who are you working with?” (or CCS). It included more details about the geographical location. For example, if the “County” category is Montgomery, this variable will show further location information about a specific project. Row seven of the dataset lists the “City of Rockville” as part of Montgomery County. This

visualization shows the CCS contact names with the most affiliations for the “Who are you working with?” category. There are 70 unique CCS contact names in total.

Figure 9



According to the data in Figure 9, Casey Nolan and Kate Vogel have the most affiliations. They are associated with ten different associations. When researching Casey Nolan further, she was associated with three other counties (available in the dataset)—Worcester, Montgomery, and Baltimore Counties. Kate Vogel is associated with six different counties—Talbot, Prince George’s, Charles, Cecil, Calvert, and Baltimore Counties. This is a valuable technique to find the most active contact names for a contact.

Method

The CCS Contact Name required some normalization for analysis. Pairs of names were in different orders and only some rows used slash separators. All the name pairs were put in the same order. For example, one name combination, Carrie Decker and Dan Levan, appeared as “Carrie Decker/Dan Levan” in one row and “Dan Levan/Carrie Decker” in another. A uniform separator was used, and the combination appeared uniformly as “Carrie Decker, Dan Levan.”

These were necessary steps to allow data analysis. After cleaning the data, the CCS contact names and affiliations column was grouped in a separate dataset. In Python, that subset was turned into a horizontal bar graph.

There weren't particular challenges in creating this visualization. However, the data gaps didn't help. For example, it was challenging to cross-reference the "Who are you working with?" column with the county because a lot of values in the county section were missing. This was overcome by using judgment to figure out the missing information. Although the missing information didn't cause any harm because it was in a different column, it would have been best to include all the information.

Recommendations

Data Tableau

Tableau is a valuable data visualization and business intelligence tools that provides various methods to process large datasets. The platform allows users to upload large datasets for visualization, connect with collaborators, transfer data through the cloud, and share data simultaneously. However, it does have a user subscription fee. Three subscriptions options vary by the features offered.

A Tableau Creator ($75/month, billed annually) receives a Tableau desktop and prep builder that allows complete back-end work and data cleaning. A Tableau Explorer ($42/month, billed annually) receives a Tableau cloud license with back-end work (with limited functionalities compared to Tableau Creator). Tableau Viewer ($15/month, billed annually) provides users with a Tableau cloud viewer license that allows viewership and interaction of Tableau dashboards.

A Creator or Explorer account is required to update or make edits. The platform offers multiple techniques to refresh uploaded data. The user can make changes on Google Sheets that are then updated in Tableau. If dealing with large datasets, then saving subsets of the data can help. **The iConsultancy team strongly recommends using this technique.**

Users can also publish the data source. However, this technique is not recommended for this project. It is challenging to manage published data sources and update the server.

One of Tableau's most significant advantages is a user-friendly interface. Easy navigation makes it less complicated to create data visualizations despite the size of the dataset. Interactive dashboards are easily made with Tableau's drag-and-drop features that allow users to quickly turn data into filterable databases.

Different visualizations based on user data can also be created—for example, the pie charts, maps, and tables in this report. The program can also create interactive and statistical visualizations based on categories of interest. Tableau also allows users to customize dashboards based on device. For example, the dashboards will be best fit to appear on smartphones and laptops. Overall, Tableau is a vital tool when it comes to data visualization.

Alternatives

There are cost-friendly options for data visualizations. For example, Google Sheets comes with no cost and is available through a Google account. It uses static and non-interactive visualizations. It can quickly create dashboards, pie charts, bar charts, line graphs, and other statistical analyses. However, it can be more challenging due to the updatability of cells using formulas. Unlike Tableau, Google Sheets doesn't have a drag-and-drop interface and no option for geographical visualizations. Data can be quickly updated through Google Surveys, survey platform integration, or services like Zapier.

Python is another alternative. Although it is primarily used for data analysis and algorithm design, it is also helpful for data visualizations. Access to Python comes without fees through Google Colaboratory. Datasets in comma-separated values (CSV) formats can be easily connected to the platform for the intended use. However, Python requires much more back-end work than Tableau. It lacks a user-friendly interface and requires coding skills to produce visualizations. The software can be used for both static and interactive visualizations. The interactive visualizations require additional steps when compared to Tableau. The updatability on Google Colaboratory is quickly done by updating the CSV file.

Data Maintenance

In the team's effort to refine the data to make it easier to work with, several considerations emerged that warrant attention for maintaining the data. The iConsultancy team recommends the following.

1. It is crucial to create consistency and acknowledgment of every organization involved. Separating each organization is essential to ensure each contributor receives due credit. To streamline this process, we recommend developing a program that can autonomously separate organizations and update engagement categories. Alternatively, incorporating functionality into the questionnaire that automatically accomplishes this separation could be a solution. This approach guarantees that each organization receives proper credit for its contributions, fostering transparency and accuracy in recognizing its engagement efforts.
2. The questionnaire should be restructured to align with the data, specifically focusing on a method for capturing various engagement categories. The team has developed 14 distinct engagement category columns, and implementing a section in the questionnaire for users to choose from these categories or propose new ones would significantly enhance the

organization of data from the outset. This approach ensures a comprehensive and standardized representation of engagement types, facilitating more effective data analysis and interpretation.

3. Consistent formatting is important to keep data usable for ranking. Ensuring that data is written in the same manner throughout is integral. The ranking table uses the updated Community and Partners table to generate ranks based on the provided data. The table containing all engagement categories has a blank entry for the county labeled as “Other” because of not listing all counties. As a result, adjustments must be made for the “Other” category in subsequent tables mentioning all counties. This modification will enhance the accuracy and completeness of data representation in subsequent tables.
4. The team suggests that future users refer to the “User Guide Documentation” in the final deliverable folder.

Next Steps

While the iConsultancy Team has successfully delivered a set of visualizations, there are additional steps to enhance the accuracy and utility of the data:

1. Incorporate missing data will ensure a more comprehensive representation. This may involve updating the primary table with engagement categories to include all counties, addressing the “Other” category, filling in data gaps, and achieving more accurate measures.
2. Restructure the questionnaire to align seamlessly with the data. The team has proposed developing a method for capturing various engagement categories. Implementing this recommendation would contribute to a more organized and streamlined data collection process.
3. Ensure consistent names of partner organization to optimize data use. This could involve establishing standardized naming conventions or conducting a thorough review to identify and rectify discrepancies. More information on the organization side of the data could bring findings that lead to value.
4. Lead effective change by using visualizations to focus efforts. The visualizations can support more informed decision making for the organizations and partners.
5. Address the many missing county data in the cleaning process by resolving the multiple organizations assigned to one contact.

By addressing these next steps, the organization can further leverage the visualizations and data insights to make informed decisions, drive positive change, and enhance stakeholder engagement.

Contact Information
Amanu Huq: ahuq00@terpmail.umd.edu
Ady Weng: aweng@terpmail.umd.edu
Avram Baicu: abaicu@terpmail.umd.edu
Farahan Uddin: fuddin1@terpmail.umd.edu
Jennifer Tatham: jtatham@umd.edu

Deliverable Links
Cleaned Data Sheet
Interactive Map Visualization
Interactive Contact Name Dashboard
Interactive Filterable Dashboard
Static Visualizations Google Sheet