



# Endangered Data Preservation

## Organizing and Visualizing Thoreau's Botanical Observations

Jodi Coalter, MLIS  
University of Maryland, College Park

### ABSTRACT

The aim of this work was to create a usable data set from botanical data gathered in the 1800's by Henry David Thoreau for visualization and scientific purposes. A special focus was to preserve the work completed by Ray Angelo in his botanical index for future work. A Python script was written to scrape each HTML page on Angelo's website to gather the botanical information and index information provided. The data has been useful in drawing out patterns in Thoreau's exploration of botany. Specific plants are watched more closely, though his interest wandered throughout the botanical world. Future data work will be needed or helpful, including attribution of dates to specific pages and volumes, as well as updated botanical terms..

### REFERENCES

Angelo, R. (2017, May 4, 2017). Botanical index to the journal of Henry David Thoreau. Retrieved from <http://www.ray-a.com/ThoreauBotIdx/index.html>

Primack, R. B., & Miller-Rushing, A. J. (2012). Uncovering, collecting, and analyzing records to investigate the ecological impacts of climate change: A template from Thoreau's Concord. *BioScience*, 62(2), 170-181. doi:10.1525/bio.2012.62.2.10

### CONTACT

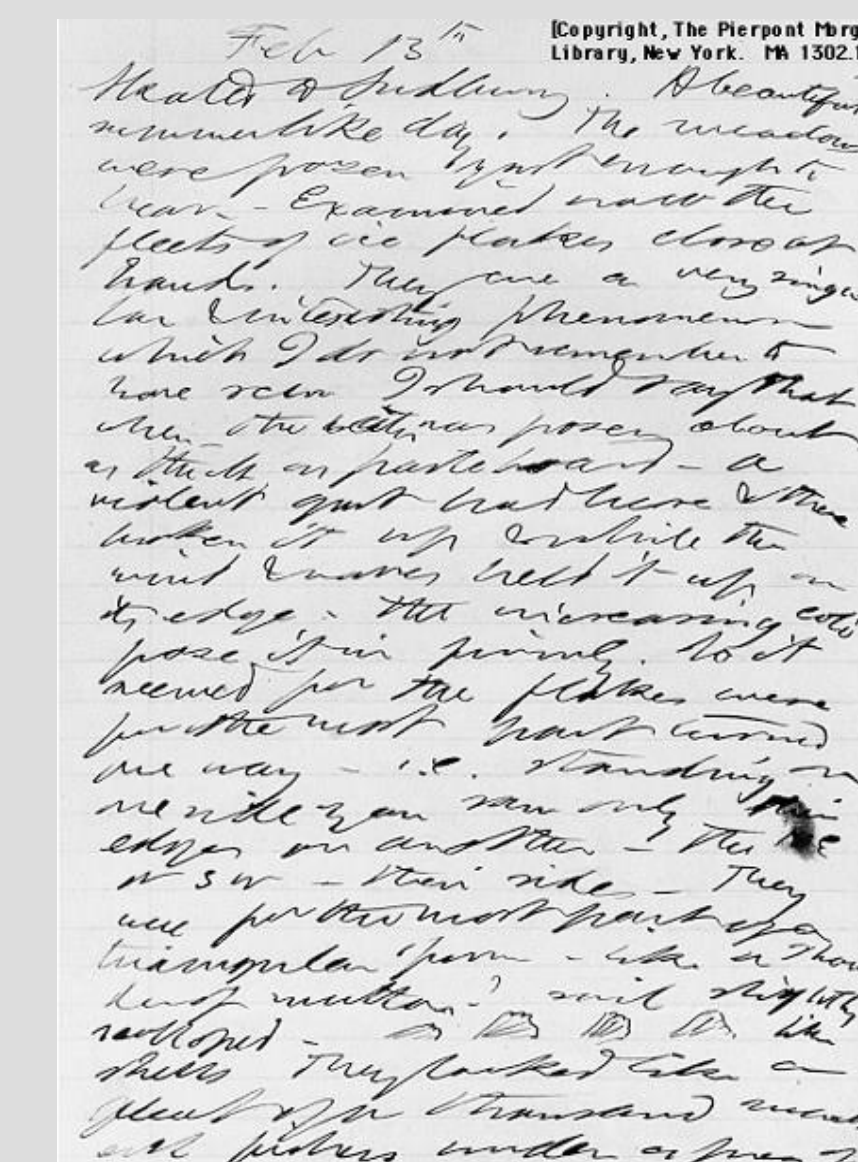
Jodi Coalter  
University of Maryland, College Park  
Email: [jcoalter@umd.edu](mailto:jcoalter@umd.edu)  
Phone: 301-405-9147

### The Data Path



Photograph by Herbert Gleason  
<https://www.walden.org/collection/journals/>

### METHODS AND MATERIALS



Example journal page from Thoreau's diary.

#### SAMPLE TRANSCRIPTION

Feb 13th  
Skated to Sudbury. A beautiful summerlike day. The meadows were frozen just enough to bear-- Examined now the fleets of ice flakes close at hand. They are a very singular & interesting phenomenon which I do not remember to have seen I should say that when the water was frozen about as thick as pasteboard--a violent gust had here & there broken it up & while the wind & waves held it up on its edge--the increasing cold froze it in seemed for the flakes firmly. So it were for the most part turned one way--i.e. standing on one side you saw only their edges on another--the N E or S W--their--sides-- They were for the most part of a triangular form--like a shoulder of mutton? sail slightly scolloped--(drawing) like shells They looked like a fleet of a thousand mackeral fishers under a press of

[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]	[K]	[L]	[M]	[N]	[O]
A. Ribes <sup>3</sup>	=	Ribes	spp.	(GOOSEBERRY; CURRANT)	—	IX	60							
Aaron's-rod	=	Sedum telephium	(LIVE-FOREVER)	Sedum purpureum	(LIVE-FOREVER)	—	V	342, 353, XIV	187					
Abele	=	Populus alba	(WHITE POPLAR)	—	VI	221, XI	316, 443, XII	133						
Abele (Silvery)	=	Populus alba	(WHITE POPLAR)	—	VII	334, XI	201, 262, 264, 276, 297, 301, 309							
Abies alba	(SILVER FIR)	—	see Picea pectinata											
Abies alba	(old usage)	=	Picea glauca	(WHITE SPRUCE)	—	II	190							
Abies balsamea	(BALSAM FIR)	—	see Abies balsamea											
Abies balsamea	=	Abies balsamea	(BALSAM FIR)	—	II	279								
Abies canadensis	=	Tsuga canadensis	(EASTERN HEMLOCK)	—	II	199								
Abies douglasii	=	Pseudotsuga menziesii	(DOUGLAS FIR)	—	II	363, X	185							
Abies excelsa	=	Picea abies	(NORWAY SPRUCE)	—	II	363, X	185							
Abies fraseri	(FRASER FIR)	—	see Picea canadensis											
Abies grandis	(GRAND FIR)	—	see Picea canadensis											
Abies nigra	=	Picea mariana	(BLACK SPRUCE)	Picea rubens	(RED SPRUCE)	—	X	464						
Acacia (Three-thorned)	=	Gleditsia triacanthos	(HONEY LOCUST)	—	II	361, XI	199							
Acalypha	=	Acalypha	spp.	(THREE-SEEDED MERCURY)	—	VIII	441, XI	184						

Ray Angelo's botanical index of Thoreau's journals.  
<http://www.ray-a.com/ThoreauBotIdx/>

### RESULTS

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	1	1	1	1

### CONCLUSIONS

Further research is needed to complete the data sets. For example, dates (specially month and year) would be helpful in determining what Thoreau cited in winter months vs summer months. In addition, Ray's index uses the Grey's Manual of Botany 8<sup>th</sup> Edition. Since its use, other editions have been printed, and +plant taxonomy has continued to improve. Updating botanical terms will be required.

### Acknowledgments

This work was completed with significant help from Dr. Peter Hook, Cornell University, and Dr. Timothy Bowman, Wayne State University.

### DISCUSSION

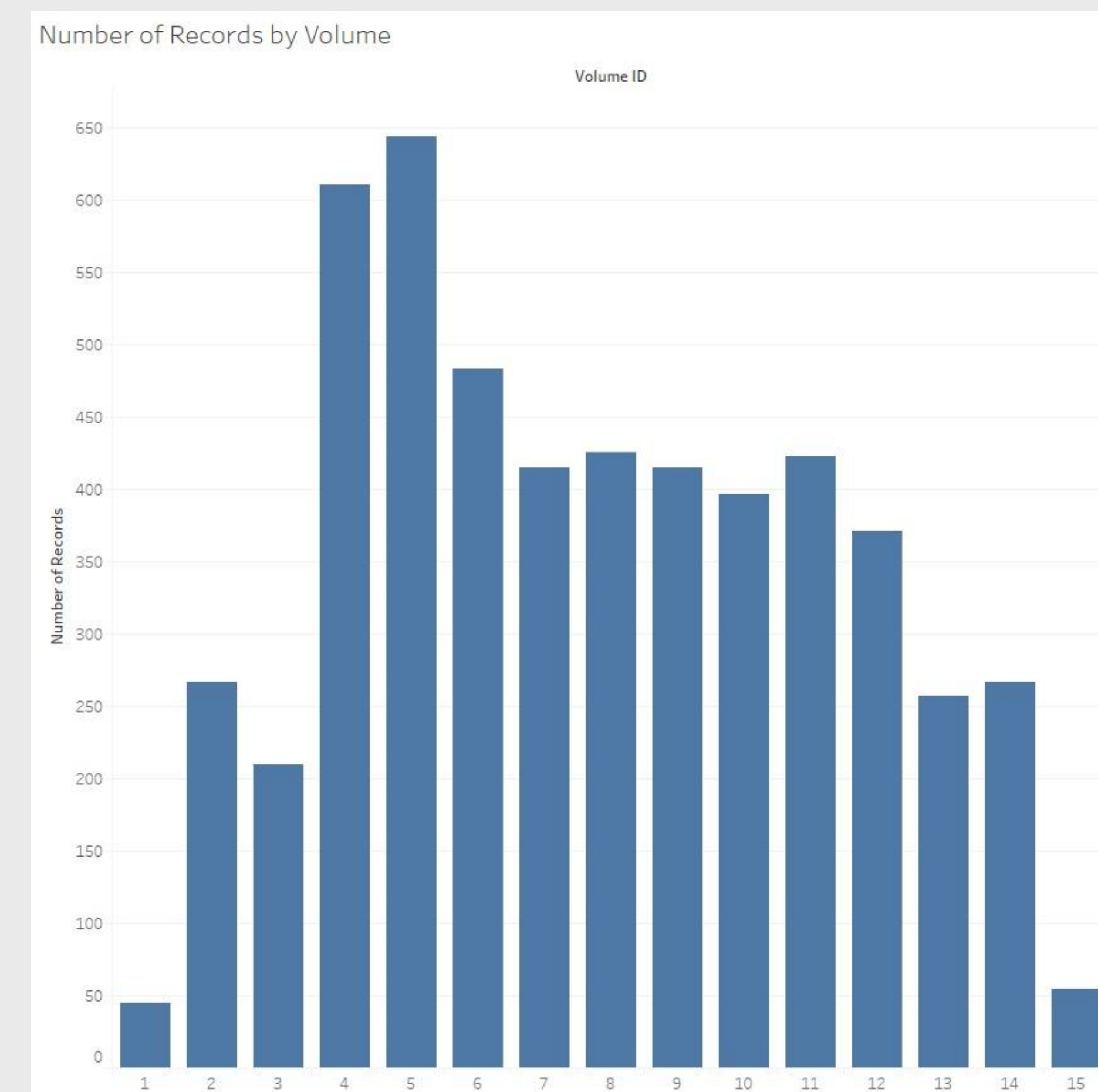


Figure 1: Distribution of citations by volume

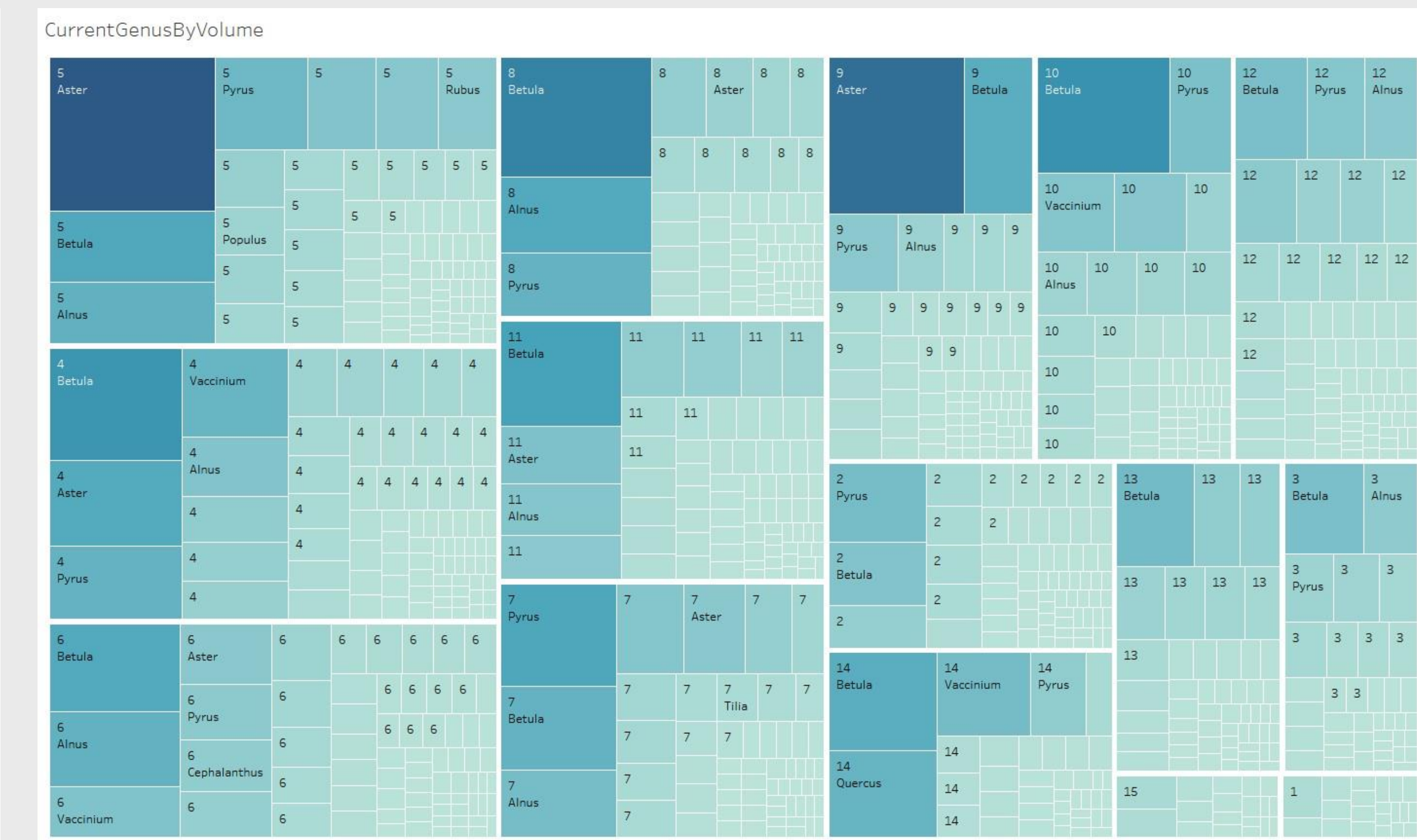


Figure 2: Current Genus by volume



Figure 3 & 4: Example of Thoreau's use names normalized by current Genus.

