

Conclusion and Resources

In the context of the data-driven knowledge economy and ubiquitous screen culture, the role of visualization as a communication tool has increased considerably. Visuals clearly and effectively convey factual information extracted from data. But the power of the visuals does not stop there. Visual representations of data also reveal patterns and trends hidden in data and thus enable further exploration and addressing new research questions.

One way of looking at library digital collections is to view them as repositories of digital data that can be explored and analyzed. Visualization as a method for exploring digital content provides unique insights into collections for both library patrons and collection curators. Applying visualization allows viewing a digital collection through analysis of its data. As more and more technologies have become available to facilitate the creation of visualizations, this method for exploring data and communicating their findings is now at our fingertips.

The R programming language has been specifically designed to support analysis and visualization of data. R is free, open source, and actively supported by an established community of developers, coders, and enthusiasts. R also is relatively easy to learn and experiment with.

As demonstrated in this report, a few lines of R code may instantly create charts that address a range of questions regarding a digital collection, starting from content overviews and ending at identifying specific metadata issues. Eighteen visualizations of the Clyde W. Tombaugh Papers digital collection described in this report directly reflect and help interpret the content of this collection and also indirectly shed some light on the life and career of a scientist. Based on these visualizations, new findings emerged.

As discussed throughout this report, visualizations provide holistic views of collection content and invite users to engage with historic documents actively and creatively. At the very least, visualizations fill the gap between an item-level description found on a digital collection's display and an extensive collection-level description issued in an archival finding aid. But

visualization may also inform the creation of metadata. Now when we know what information is needed to produce useful charts, we may reevaluate metadata elements that describe our collections.

This report is meant to encourage the creation of visualizations for digital collections and inspire further exploration of collection data. R scripts, provided along with explanations of code, are to serve only as examples of possible solutions.

Resources

Installing R and RStudio

- The R Project for Statistical Computing website, <https://www.r-project.org/>
- CRAN Mirrors, <https://cran.r-project.org/mirrors.html>
- RStudio website, <https://rstudio.com/>
- How to Articles, RStudio Support, <https://support.rstudio.com/hc/en-us/categories/200035113-Documentation>

Learning R

- What is R? Introduction to R on [r-project.org](https://www.r-project.org/about.html), <https://www.r-project.org/about.html>
- R Manuals on [r-project.org](https://cran.r-project.org/manuals.html), <https://cran.r-project.org/manuals.html>
- Documentation on [r-project.org](https://www.r-project.org/other-docs.html), <https://www.r-project.org/other-docs.html>
- *The R Journal* home page, <https://journal.r-project.org/>
- *The R Blog*, <https://developer.r-project.org/Blog/public/>
- The R Foundation @ Twitter, https://twitter.com/_R_Foundation
- Books Related to R on [r-project.org](https://www.r-project.org/doc/bib/R-books.html), <https://www.r-project.org/doc/bib/R-books.html>

Other Books

- Hadley Wickham and Garrett Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize and Model Data* (Sebastopol, CA: O'Reilly Media, 2017).
- J. D. Long, and Paul Teetor, *R Cookbook: Proven Recipes for Data Analysis, Statistics and Graphics* (Sebastopol, CA: O'Reilly Media, 2019).

Platforms That Provide Courses on R

- Codecademy, <https://www.codecademy.com/>
- Coursera, <https://www.coursera.org/>
- DataCamp, <https://www.datacamp.com/>
- Quick-R by DataCamp, <https://www.statmethods.net/>
- edX, <https://www.edx.org/>
- Udacity, <https://www.udacity.com/>
- Udemy, <https://www.udemy.com/>
- Library Carpentry, <https://librarycarpentry.org/>

Learning RStudio

- How to Articles, RStudio Support, <https://support.rstudio.com/hc/en-us/categories/200035113-Documentation>
- RStudio Cheatsheets, <https://rstudio.com/resources/cheatsheets/>
- RStudio Essentials, <https://rstudio.com/collections/rstudio-essentials/>

- RStudio Webinars, <https://rstudio.com/resources/webinars/>
- *RStudio Blog*, <https://blog.rstudio.com/>
- RStudio Community, <https://community.rstudio.com/>

Cleaning Data

- OpenRefine, <https://openrefine.org/>
- Library Carpentry, <https://librarycarpentry.org/>

Selecting Charts

- The Extreme Presentation Method, by Andrew Abela, <https://extremepresentation.com/>
- The Data Visualisation Catalogue, by Severino Ribecca, <https://datavizcatalogue.com/index.html>
- The Visualization Universe, by Anna Vital, Mark Vital, Alexander Vushkan, Simon Roger, and Alberto Cairo, <http://visualizationuniverse.com/>
- Chartopedia, by AnyChart, <https://www.anychart.com/chartopedia>
- From Data to Viz, by Yan Holtz and Conor Healy, <https://www.data-to-viz.com/>
- Scott Berinato, *Good Charts: The HRB Guide to Making Smarter, More Persuasive Data Visualizations*, (Boston: Harvard Business Review Press, 2016).