
Newsroom Textual Analysis and Visualization Tools Built With R Shiny

Aleszu Bajak

Northeastern University
Boston, MA, USA
a.bajak@neu.edu

John Wihbey

Northeastern University
Boston, MA, USA
j.wihbey@neu.edu

Gibson Free

Northeastern University
Boston, MA, USA
free.g@husky.neu.edu

Paxtyn Merten

Northeastern University
Boston, MA, USA
merten.p@husky.neu.edu

Abstract

While word cloud visualizations and similar types of simple tools are widely available on the web, the more sophisticated textual analysis software and code unfortunately remain the domain of experts and users of languages like R and Python. To address this, we created two prototypes that would allow newsrooms to harness the power of R's textual and sentiment analysis packages in a simple drag-and-drop format through public-facing R Shiny apps. These two apps show how powerful data wrangling, analysis and visualization functions may be increasingly democratized for time-pressed journalists. We demonstrate use cases through: 1) exploratory analysis of public speeches; 2) exploration of sentiment analysis around large bodies of political advertising, such as the new archive on the Facebook platform; and 3) comparative analysis of one outlet's coverage on a specific topic as compared to that of the competition.

R is an open-source statistical software environment with a robust community of developers that collaborate and share knowledge, tools and code.¹ Newsrooms are increasingly adopting R. Currently, R has statistical and textual analysis packages that allow for rich and generative forms of interpretation. In addition, R Shiny, which deploys code and datasets to a server, is easily customizable and user-friendly; it allows for a text-and-numerical dataset to be dropped in and a set of questions to be asked of

it, facilitated by a pre-coded operational workflow.

These apps would allow reporters to drop in spreadsheets or text documents of interest and ask a series of cross-cutting questions. In this project, we have customized R's "tidytext" and "plotly" packages and UVM's "LabMTsimple" sentiment dictionary for textual and sentiment analysis and added several intervention points for users to subset the data, ask it guided questions and produce exploratory data visualizations.^{2 3 4}

We consider the possibilities enabled by our applications by looking at two use cases. Our first use case with a prototype Shiny app involves looking at a corpus of texts or headlines such as the public speeches of a politician or coverage of a news topic to understand rhetorical patterns, popular formulations and sentiment (See Figure 1). Techniques such as descriptive statistics, sentiment analysis and n-gram analysis are employed. Our second use case with a prototype Shiny app involves discerning patterns in political advertisements on Facebook, allowing for unique insights into the strategies employed in political messaging (See Figure 2). Techniques such as textual and sentiment analysis are employed.

These prototypes point to the possibility of a broader ecosystem of similar deadline-friendly

apps for newsrooms that could provide them with greater analytical power and higher-level insights.

Optional

Some datasets you might be curious to explore.

- Three months of r/politics headlines
- One year of Washington Post Politics headlines
- One year of New York Times Politics headlines
- Three months of Sen. Heidi Heitkamp articles
- @realDonaldTrump tweets through Nov. 2016

Have a TXT file for sentiment analysis?
[Try out my other drag-and-drop app.](#)

This analysis uses the R package 'tidytext' and the 'labMT' sentiment dictionary from Andy Reagan. Created by Aleszu Bajak.

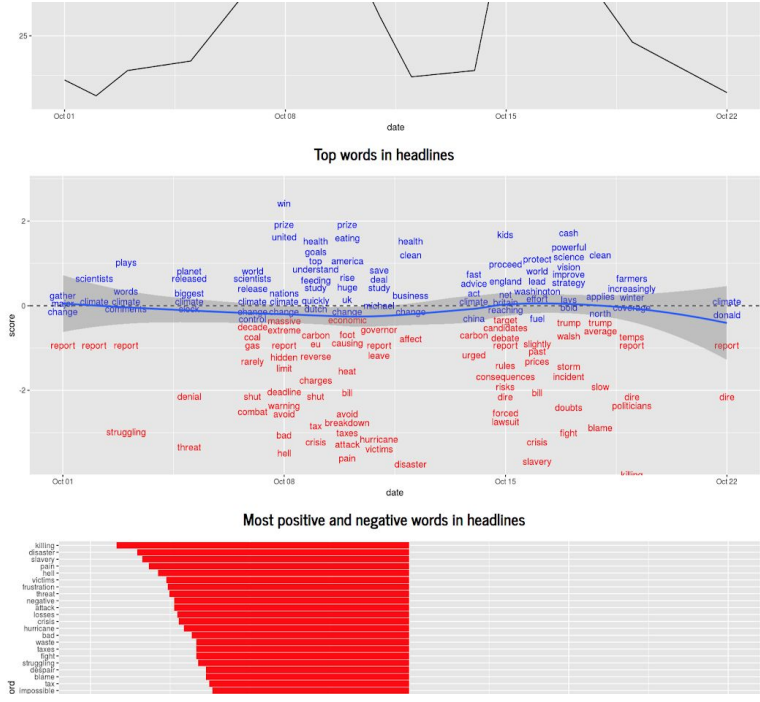


Figure 1: One R Shiny prototype allows users to discern patterns in media coverage or political speeches. Techniques such as textual and sentiment analysis are employed. The app can be found at storybench.shinyapps.io/csvanalysis and storybench.shinyapps.io/textanalysis

Facebook ads analysis

This Shiny app uses Facebook ads collected by ProPublica

[Download the data.](#)

Search for specific politician or keyword

Try Beto, Cruz, immigration, Russia...

Change color #1

Change color #2

Change color #3

Have a TXT file for textual and sentiment analysis?
[Try out our TXT Shiny app.](#)

Have a CSV file for textual and sentiment analysis?
[Try out our CSV Shiny app.](#)

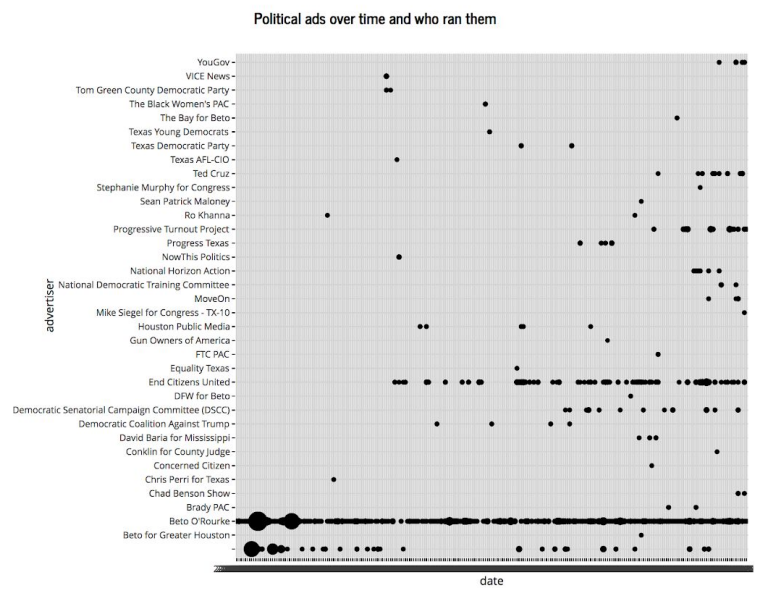


Figure 2: One R Shiny prototype allows users to discern patterns in political advertisements on Facebook, allowing for unique insights into the strategies employed in political messaging. Techniques such as n-gram and sentiment analysis are employed. The app can be found at storybench.shinyapps.io/facebook

References

- [1] "The R Project for Statistical Computing" Retrieved from <https://cran.r-project.org/>
- [2] Robinson, D. and Silge, J. 2018. "tidytext: Text Mining using 'dplyr', 'ggplot2', and Other Tidy Tools" Retrieved from <https://cran.r-project.org/web/packages/tidytext/index.html>
- [3] Ropensci. 2018. "Plotly: an interactive graphing library for R" Retrieved from <https://github.com/ropensci/plotly>
- [4] Reagan, A. 2018. "labMTsimple Documentation" Retrieved from <https://media.readthedocs.org/pdf/labmt-simple/latest/labmt-simple.pdf>