Project 3: "State of the Union" Text Analysis

Stat 133, Spring 2022

Introduction

In this project, you will build a shiny app to visualize the results from a text analysis performed on the State of the Union messages given by various presidents of the U.S. from 2001 to 2022.

- Part I on Data
- Part II on Text Analysis
- Part III on Shiny App

```
# suggested packages
library(tidyverse)  # tidy-data ecosystem of packages
library(tidytext)   # for text mining
```

We are assuming that you have reviewed the learning materials of weeks 11, 12, and 13 (see bCourses).

Part I: Data

This section introduces the data for this project.

Annual Messages on the State of the Union

The data for this project involves the text of the "Annual Messages to Congress on the State of the Union" given by various presidents of the U.S. from 2001 to 2022.

The data was webscraped by Prof. Sanchez from the *The American Presidency Project*:

https://www.presidency.ucsb.edu/

BWT: Prof. Sanchez has no background in Political Science; he is interested in the State of the Union messages purely from a text-data analysis standpoint.

Figure 1: Screenshot of the "American Presidency Project" website

BARACK OBAMA

44th President of the United States: 2009 - 2017

Address Before a Joint Session of the Congress

February 24, 2009

Madam Speaker, Mr. Vice President, Members of Congress, the First Lady of the United States--she's around here somewhere: I have come here tonight not only to address the distinguished men and women in this great Chamber, but to speak frankly and directly to the men and women who sent us here.

I know that for many Americans watching right now, the state of our economy is a concern that rises above all others, and rightly so. If you haven't been personally affected by this recession, you probably know someone who has: a friend, a neighbor, a member of your family. You don't need to hear another list of statistics to know that our economy is in crisis, because you live it every day. It's the worry you wake up with and the source of sleepless nights. It's the job you thought you'd retire from but now have lost, the business you built your dreams upon that's now hanging by a thread, the college acceptance letter your child had to put back in the envelope. The impact of this recession is real, and it is everywhere.

Figure 2: Screenshot of 2009 message by President Barack Obama

1.1) Data file `state-union-2001-2022.csv`

We are providing a CSV file `state-union-2001-2022.csv` located in the folder containing this pdf of instructions (see bCourses folder `Files/hws/project3`)

This data set is fairly simple—in terms of its structure—although the text content is far from being tidy. The dataset has 22 rows and four columns:

1) `president`: name of president
2) `year`: year of state of the union message
3) `party`: president's party
4) `message`: text with content of the message

Part II: Text Analysis

> This section provides some of the suggested text analysis that you can perform for this project.

Listed below are four major text analysis ideas for you to get inspiration from. We are also including recommended readings (some available in bCourses, some available in the book "Text Mining with R", by Silge & Robinson).

Out of the four listed types of text analysis (2A-2D) you will have to choose two of them in order to create the shiny app.

2.A) Word Frequency Analysis

Taking into account **all** the "words" (i.e. tokens)

- what are the top-5, or top-10, or top-20 (or any other number of) most frequent words used in the State of the Union messages (among all presidents)?
- what are the top-5, or top-10, or top-20 (or any other number of) most frequent words **for a given president**?
- what are the top-5, or top-10, or top-20 (or any other number of) most frequent words **per message** (hint: facet by year)?

After removing **stopwords**:

- what are the top-5, or top-10, or top-20 (or any other number of) most frequent words used in the State of the Union messages (among all presidents)?
- what are the top-5, or top-10, or top-20 (or any other number of) most frequent words **for a given president**?

- what are the top-5, or top-10, or top-20 (or any other number of) most frequent words **per message** (hint: facet by album)?

Suggested reading

- `text-mining-1-pride-and-prejudice.html` (see `Files/readings` in bCourses)

2.B) Sentiment Analysis

You can also perform a sentiment analysis. For example, but not limited to:

- Compute a sentiment score for each State of the Union message, or for each president. And then rank them from more positive to more negative.
- Which messages have "relatively large" positive scores? And/or what words contribute the most for the score?
- Which messages have "relatively large" negative scores? And/or what words contribute the most for the score?

Suggested reading

- `text-mining-3-sentiment-analysis.html` (see `Files/readings` in bCourses)
- See also section 2.4 "Most common positive and negative words" (in "Text Mining with R"; link below)

https://www.tidytextmining.com/sentiment.html#most-positive-negative

2.C) n-gram Analysis

Another type of analysis involves studying so-called **n-grams** (e.g. bigrams, trigrams, etc) for answering questions like:

- what kind of words tend to be associated with other words?

Suggested reading

- `text-mining-2-pride-and-prejudice.html` (see `Files/readings` in bCourses)
- See chapter 4 "Relationships between words: n-grams and correlations" (in "Text Mining with R"; link below)

https://www.tidytextmining.com/ngrams.html

2.D) Word Trend Analysis

Lastly, you can also do some word-trend analysis. For example, how do words (or other words) such as "freedom", "democracy", "leader", "challenge", "economy", "war", have been used in the State of the Union messages over the years?

Suggested reading

- See figure 5.4 in "Text Mining with R" (link below) to get a rough idea about this type of trends over time.

https://www.tidytextmining.com/dtm.html#tidying-dfm-objects

Part III: Shiny App

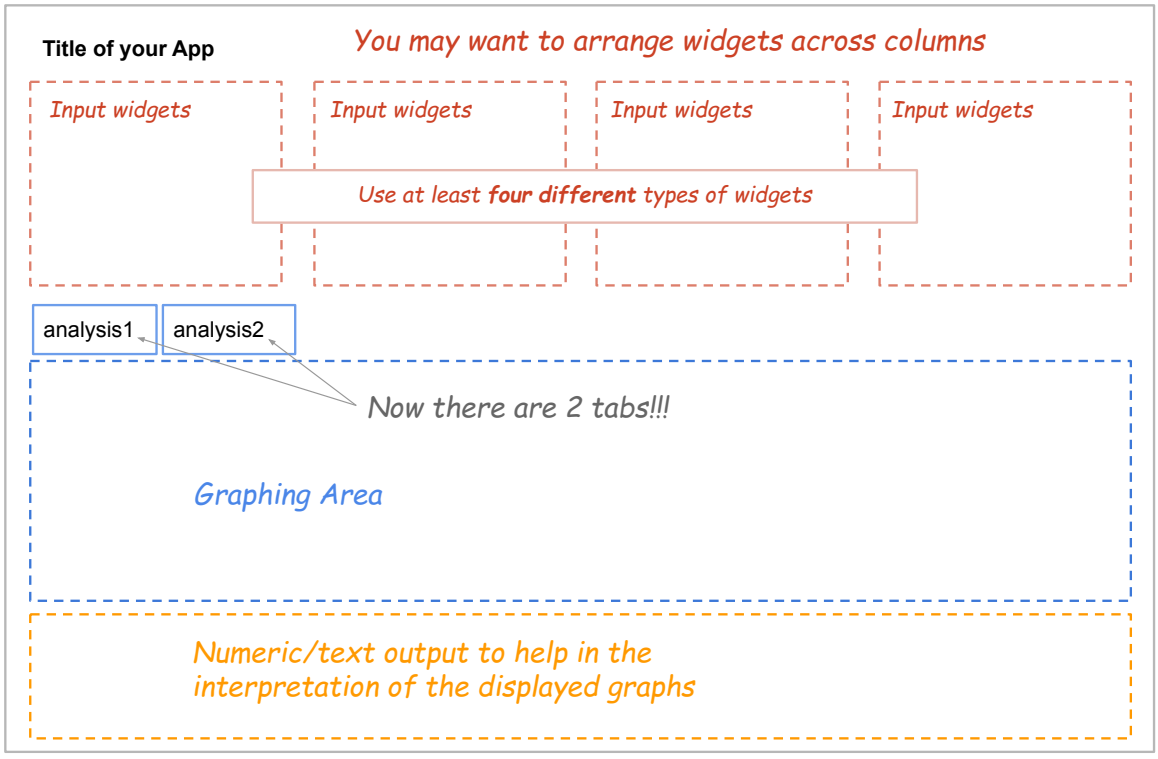
This section describes generic specifications of the project's shiny app.

3) Shiny App

The main data product to be delivered for this assignment is a shiny app that allows the user to explore the results of two types of text analysis.

3.1) Layout

You can find a template R script file `app-template.R` in the folder containing this pdf of instructions (see bCourses folder `Files/hws/project3`).



From the diagram above, note that there are four distinctive sections in the layout—see template file `app-template.R`:

- **title**: main title for your app (give it a meaningful name).
- **input widgets**: the template already contains five input widgets arranged in four columns; but you can change this configuration, as well as the types and number of widgets.
- **plot**: an output area to display graph(s).
- **stats**: an output area (e.g. for a table, text, etc) to display numeric/text output.

As you can tell, the layout of the app is very similar to the shiny app of project 2. The main difference in the app for this third project is in the fact that it uses **two tabs**:

1) **Analysis1**: this tab is for displaying the results for one type of text analysis (for example: word frequency analysis)
2) **Analysis2**: this tab is for displaying the results for another type of text analysis (for example: sentiment analysis)

For example, you can choose 1) a word frequency analysis, and a 2) sentiment analysis. Keep in mind that even if two (or more) students choose to work on the same type of analyses, there is still enough room to approach them in slightly different ways, therefore producing different shiny apps, with different scopes, and of course different data visualizations and outputs.

4) Submission

1) **R file**: You will have to submit the source `app.R` file (do NOT confuse with an `Rmd` file) containing the code of your app.
2) **Link of published app**: You will also have to submit the link of your published app in shinyapps.io (the free version). Share the link with us in the comments section of the submission in bCourses.
3) **Video**: In addition to the `app.R` file and the link of your published app, you will also have to record a video—maximum length of 4 mins—in which you show us your published shiny app, how to use it, and a description of its outputs. As usual, make sure that both your screen and your face are captured, also check that the resolution of the video is okay, without too much background noise, avoiding very low volume or inaudible audio.
4) **Important:** You do NOT have to submit any Rmd or html files this time. Also, **we will not accept any content sent by email**. We will only grade the `app.R` file submitted to bCourses, the public link of the video, and the link of your app in `shinyapps.io`.

5) Some of the things we will pay attention to

We will pay attention to the visual appearance of the graphics (e.g. type of graph, use of colors, supporting elements such as grid lines, text, labels, legends, annotations, etc.). This does not mean that your graphic must have all possible visual elements. Instead, this means that we will assess the effectiveness of your graph in terms of the displayed information, taking into account good practices of data visualization.

We will also evaluate the effectiveness of the numeric and/or text output displayed in your shiny app, in terms of providing understanding and insight for each of the analysis.

Likewise, we will also assess your video. Make sure that the image and sound quality of your video are acceptable (avoid background noise, inaudible voice, highly pixelated images, trembling camera movements, and things like that). You may need to rehearse what you will say in your video a couple of times before its definitive recording.

Above all, put yourself in the place of a generic user who will use your app without you being there to explain them how to use it, or to tell them how to make sense of the displayed information. We will examine your published app without necessarily watching your video at the same time. If something needs an explanation, make sure to include it in your app (not just in your video).

Resources

You may want to take a look at the Shiny gallery: https://shiny.rstudio.com/gallery/

Shiny widgets gallery:

https://shiny.rstudio.com/gallery/widget-gallery.html

Share you app with `shinyapps.io`:

https://vimeo.com/rstudioinc/review/131218530/212d8a5a7a/#t=30m35s

Of course, you can take a look at other apps displayed in the Shiny gallery to get some inspiration.