Assessing campaigns

Gov 1347: Election Analytics

Kiara Hernandez

December 1, 2022

Harvard University

Today's agenda

1. Definitions: **campaign narrative** and **testable implication**

- District-level application for the final assignment

2. Applying Vavreck's (2009) campaign assessment methodology to our districts

- Text-as-data: what is text preprocessing and analysis?
- Basic text analysis example

3. The End :) and stickers!

Reminder

Final assignment on Campaign Narratives is due Wednesday, December 7th by 9pm

What is a narrative?

Campaign narrative:

- Social scientific findings: rigorous quantitative or qualitative causal relationships about general phenomenon using repeated observations

What is a narrative?

Campaign narrative:

- Social scientific findings: rigorous quantitative or qualitative causal relationships about general phenomenon using repeated observations
- Campaign narratives: less rigorous causal relationships about a specific phenomenon using repeated or anecdotal observations

What is a narrative?

Campaign narrative:

- Social scientific findings: rigorous quantitative or qualitative causal relationships about general phenomenon using repeated observations
- Campaign narratives: less rigorous causal relationships about a specific phenomenon using repeated or anecdotal observations

Testable implication:

- If narrative `X` is true $\rightsquigarrow$ we should observe relationship `Y`
- If we don't observe relationship `Y` $\rightsquigarrow$ narrative `X` is unlikely

Thinking through the final assignment

- To start/continue thinking about the final assignment, let's turn to a partner and

(1) discuss two campaign narratives that apply to your district

(2) two testable implications drawn from those narratives

Reviewing Vavreck (2009)

TABLE 3.2
A Campaign Typology: Clarifying and Insurgent Campaigns

Economy Helps Candidate	**Economy Does Not Help Candidate**
Incumbent party in good economy or challenging party in bad economy.	Incumbent party in bad economy or challenging party in good economy.
Clarifying Campaign	**Insurgent Campaign**
Talk about economy more than anything else in the campaign to reduce voters' uncertainty about your relationship to current economic situation.	Choose an issue on which insurgent candidate benefits from public opinion more than clarifying candidate (the more lopsided the distribution of opinion, the better).
Talk about the economy more than the insurgent candidate to prevent him or her from increasing voters' uncertainty about your relationship to current economic situation.	Must be an issue on which clarifying candidate is committed to or constrained by previously taken unpopular position.

- Dependent on whether the party is the incumbent or challenger party $\rightarrow$

(a) Clarifying campaign: Economy = #1
(b) Insurgent campaign: Public opinion = #1

- In Discussion, we started to assess the campaigns in our districts using Vavreck's typology. How can we do this more formally?

Text-as-data analytical approaches

- What do we mean by "text as data"?
  - Text-as-data methods are a broad set of techniques and approaches relying on the automated or semi-automated analysis of text. This methodology comes out of advances in Machine Learning, but as we'll see, it is a pretty straightforward and powerful tool for textual analysis.
- The typical workflow for text analysis:

  - We start with a corpus: a corpus is a collection of texts, usually stored electronically, on which we perform our analysis. A corpus might be a collection of news articles from Reuters, the published works of Shakespeare, and in our case, the "About" page on our candidate's campaign website.
  - Within each corpus we will have separate articles, stories, volumes, each treated as a separate entity or record. Each unit is called a "document." Each row in the relevant file is a document, and columns are text and metadata (information about each document). For our purposes today, we only have one "document."

Text-as-data workflow

Many text analysis applications follow a similar 'recipe' for preprocessing the text, including:

1. Tokenizing the text to unigrams (or bigrams, or trigrams)
2. Converting all characters to lowercase
3. Removing punctuation
4. Removing numbers
5. Removing Stop Words, including custom stop words
6. "Stemming" words, or lemmitization. There are several stemming alogrithms. Porter is the most popular.
7. Creating a Document-Term Matrix

a. A document term matrix is simply a matrix with documents as the rows and terms as the columns and a count of the frequency of words as the cells of the matrix.

8. Weighting features
9. Removing Sparse Terms

Example: analyzing Pappas' "About You" page

Today, you'll see that I'm only interested in quickly assessing the frequency of bigrams. There are more sophisticated techniques to then classify those terms, but for the purposes of the final assignment, it would suffice to simply get the frequency and do this classification "by hand," i.e. simply assess for yourself whether the candidate's are campaigning on the economy or other insurgent issues.

```
library(quanteda)
library(tidyverse)
#install.packages('quanteda.textstats')
library(quanteda.textstats)
#install.packages('quanteda.textplots')
library(quanteda.textplots)
library(ggplot2)

# read in txt file
df <- read_delim("about_pappas.txt", delim="|")

cat("\nmaking a corpus")
```

```
##
## making a corpus
```

```
my_corpus <- corpus(df,
                    text_field = "text")
```

Example: analyzing Pappas' "About You" page pt.2

```
# preprocessing
cat("\nmaking tokens")
```

```
##
## making tokens
```

```
my_tokens <- tokens(my_corpus,
                    remove_punct = TRUE,
                    remove_symbols = TRUE,
                    remove_numbers = TRUE,
                    remove_url = TRUE) %>%
  tokens_tolower() %>%
  tokens_remove(pattern=stopwords("en")) %>%
  tokens_wordstem() %>%
  tokens_select(min_nchar=3) %>%
  tokens_ngrams(n=2) # creating bigrams

cat("\nmaking a DFM")
```

```
##
## making a DFM
```

```
my_dfm <- dfm(my_tokens,
              tolower = TRUE)
```

Example: analyzing Pappas' "About You" page pt.3

```
# quick summaries
tstat_freq <- textstat_frequency(my_dfm, n = 10)

head(tstat_freq, 100)
```

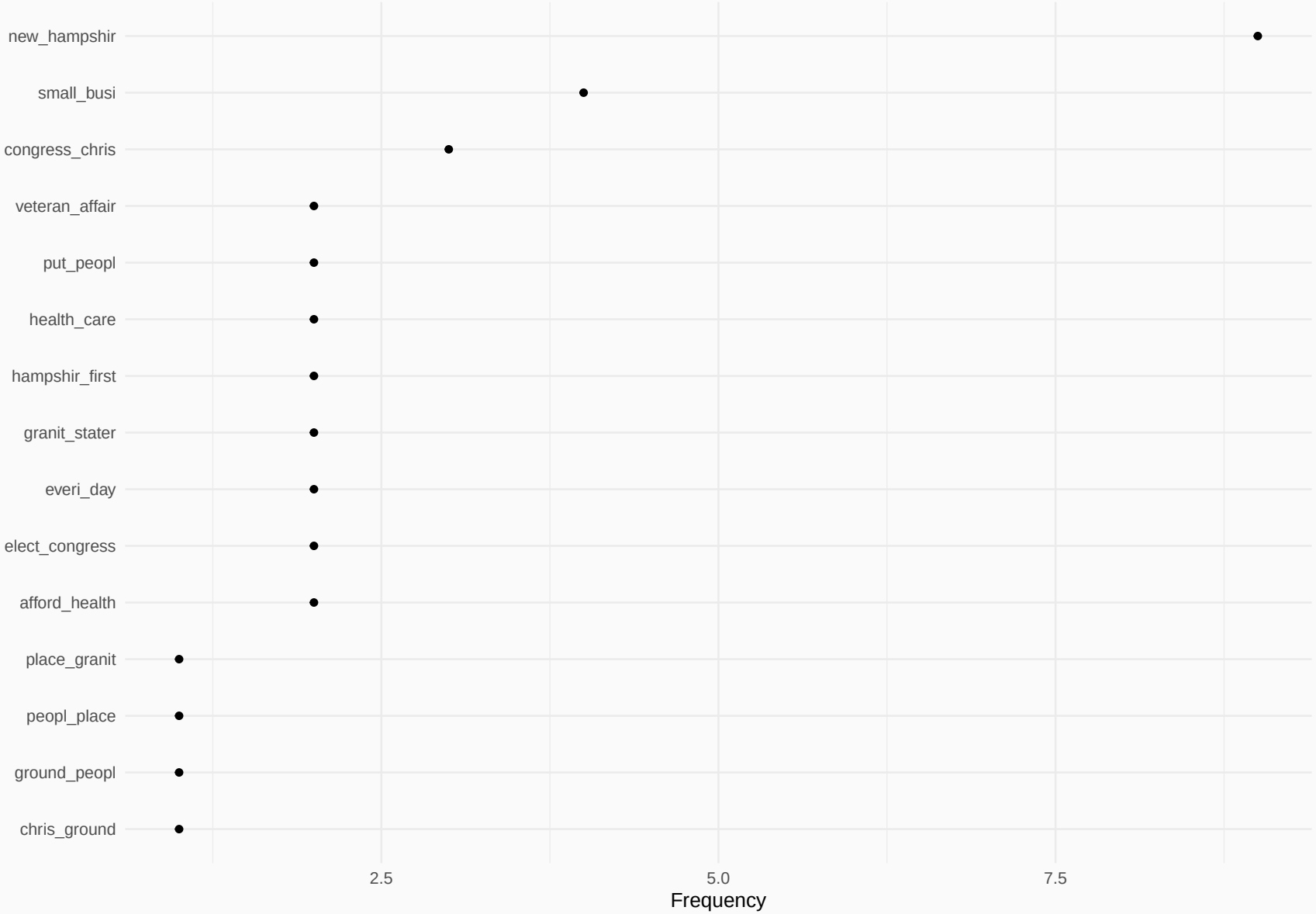
```
##           feature frequency rank docfreq group
## 1    new_hampshir         9    1       1   all
## 2      small_busi         4    2       1   all
## 3  congress_chris         3    3       1   all
## 4       put_peopl         2    4       1   all
## 5  hampshir_first         2    4       1   all
## 6       everi_day         2    4       1   all
## 7   afford_health         2    4       1   all
## 8     health_care         2    4       1   all
## 9   granit_stater         2    4       1   all
## 10 veteran_affair         2    4       1   all
```

```
set.seed(132)
textplot_wordcloud(my_dfm, max_words = 10)
```

Example: analyzing Pappas' "About You" page pt.4

```
# quick summaries
my_dfm %>%
  textstat_frequency(n = 15) %>%
  ggplot(aes(x = reorder(feature, frequency), y = frequency)) +
  geom_point() +
  coord_flip() +
  labs(x = NULL, y = "Frequency") +
  theme_minimal()
```

Examining bigram frequencies



So, how did Pappas do?

Thanks for a great semester!

This semester, we learned how elections are won and lost. We covered a ton of territory methodologically, but also substantively:

- Voter psychology (retrospective voting, generic ballot responses (polling), voter persuasion)
- Campaign strategy (political advertisements and on-the-ground campaign tactics (mobilization vs. persuasion))
- Institutional and structural constraints (redistricting, election administration, polarization)
- The "error" term (unexpected events, group-level demographic surges)

The goal of our class: kick off your career as professional election analyst but also give you skills broadly useful for any analytics profession:

- data scientist
- statistical consultant
- campaign strategist
- media/advertising analyst
- quantitative finance
- quantitative social scientist (us!)

Best of luck in your journey!