State of the Union Addresses

Gina Reynolds, Susie Q., Johny J.

2/2/2019

Contents

1 Abstract

State of the Union addresses are a classic corpus for students of text analysis. To the author's knowledge, less has been done to characterize the timing of addresses.[1] In this paper you'll see time-in-year by year plots and a model of the timing.

2 Introduction

The 2019 government shutdown had many people wondering, when will US State of the Union address be delivered by the President. Will it be pushed later than what is considered "normal."[2] I might have another tangential comment.[3]

3 Literature Review

The texts of the State of the Union addresses have been studied by scholars including Benoit, Munger and Spirling (2018); they offer an interesting analysis of State of the Union addresses by US presidents, noting that SOTU that are not delivered orally tend to use more complex language. There are several R text analysis packages (Silge and Robinson 2016; Benoit and Nulty 2016). These entries are included in the literature/bibliography.bib document that is referenced in the YMAL.

In this section you might write quota a large selection which you will want to indent.

---

[1] Motivation for addressing this question actually comes from Kenneth Benoit who conducted a quick analysis on this question.

[2] Here is my tangential comment (footnote). It will appear at the end of the page or bottom of the document.

[3] Here is a second tangential point.

Table 2: This table contains presidents and the number of SOTU that they have given

President	Number of Addresses
Donald J. Trump	3
Barack Obama	8
George W. Bush	8
William J. Clinton	8
George Bush	4
Ronald Reagan	8
Jimmy Carter	7
Gerald R. Ford	3
Richard Nixon	12
Lyndon B. Johnson	6
John F. Kennedy	3
Dwight D. Eisenhower	10
Harry S. Truman	8
Franklin D. Roosevelt	13

4 Theory and Hypotheses

This analysis was exploratory. It is well known that modern SOTU addresses are delivered early in the year. After plotting the data, what seemed to matter was year of delivery, and if the SOTU followed an election year. We tried to follow the advice in the e-book The Fundamentals of Data Visualization.

The principles in TFODV are listed in Table 1

Table 1: This is the caption for the table about data visualization categories from *"The Fundamentals of Data Visualization."*

Category	Description
Bad	Misrepresents data or confuse
Ugly	Not pleasing
Good	Not bad and not ugly

5 Data

The corpus data was made available by on github.
President in general give one SOTU address per year they serve as president but there are some exceptions as seen in table 2

6 Analysis

6.1 Visualization of relationships

Here I show some of the relationships in the data visually:

In figure 1 we use LOESS smoothing to summarize the relationship between year and time in year for the categories "following election year" and "not following". In the linear modelling exercise below, we remove the 1973 cases, given that they are so unusual — in light of the Nixon impeachment, there were multiple SOTU's delivered that year.

How late are State of the Union addresses usually delivered?
Data: Speeches from 1934 to 2019 from www.presidency.ucsb.edu
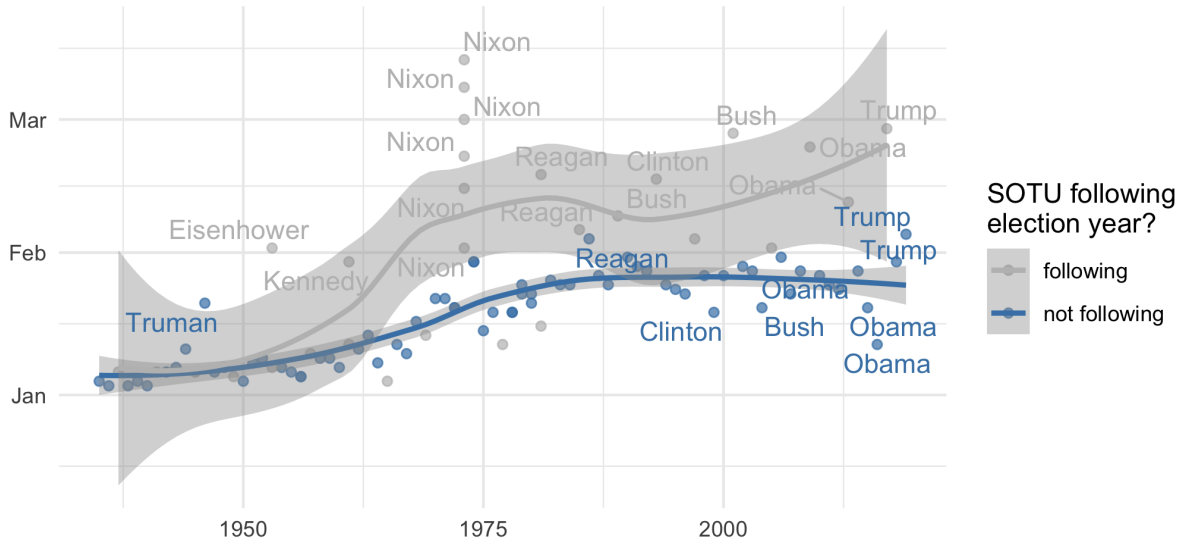
Vis: @EvaMaeRey | Data Collection: github/BrianWeinstein/state-of-the-union

Figure 1: Timing of state of the union by year. SOTUs that follow election years are colored with blue.

6.2 Modeling

We load the models that we've estimated and save out,then prepare the table using stargazer.

If compiling to pdf: in the ymal you will change: bookdown::pdf_document2. Note that setting the label is done differently than the knit to html version.

As you can see in the regression table 3, the Rˆ2 for the full model with an interaction term is 0.704.

The full model formula with the interaction is:

$$DaysSinceJan1 = \beta_0 + \beta_1 Year + B_2 FollowingElection + B_3 Year * FollowingElection + \epsilon$$

Need to put figures side-by-side? Use fig.show = hold in the code chunk options, and specify the width

Note that the visualization of the linear models do not represent the full model with the interaction but rather independent models for each category "following", "not following", the default for ggplot2. Not going to worry about this.

7 Conclusion

The analysis shows the timing of State of the Union addresses is increasingly late in the year and is dependent on whether it follows an election year. While the 2019 State of the Union address is in February, which is unusual for a large departure from the trend. In future work, we might look at plotting prediction intervals as well as confidence intervals Now is a good time to spell check (Edit -> Spell Check. . .)!

Table 3: Models of time elapsed in year before state of the union address

	Dependent variable:		
	Days elapsed since January 1st		
	(1)	(2)	(3)
year	0.401***		0.320***
	(0.038)		(0.036)
post election		7.700**	−653.628***
		(2.991)	(139.861)
year*post election			0.335***
			(0.071)
Constant	−773.950***	17.300***	−615.478***
	(75.412)	(1.511)	(71.731)
Observations	94	94	94
R^2	0.546	0.067	0.704
Adjusted R^2	0.541	0.057	0.695
Residual Std. Error	8.821 (df = 92)	12.644 (df = 92)	7.196 (df = 90)
F Statistic	110.654*** (df = 1; 92)	6.628** (df = 1; 92)	71.493*** (df = 3; 90)

Note: *p<0.1; **p<0.05; ***p<0.01
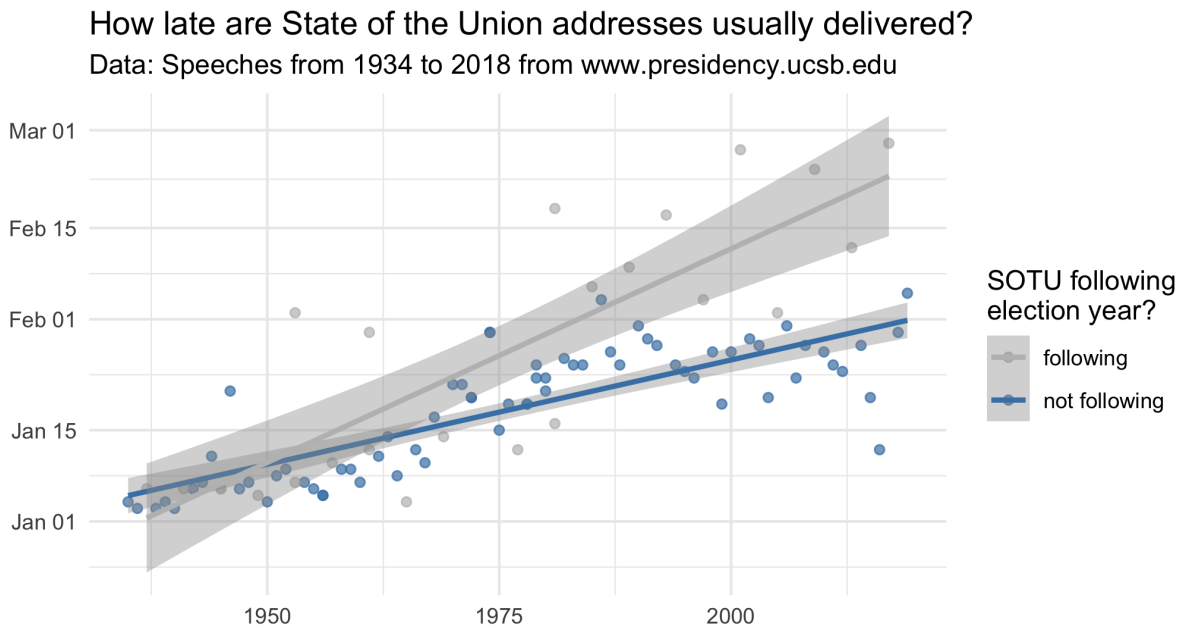
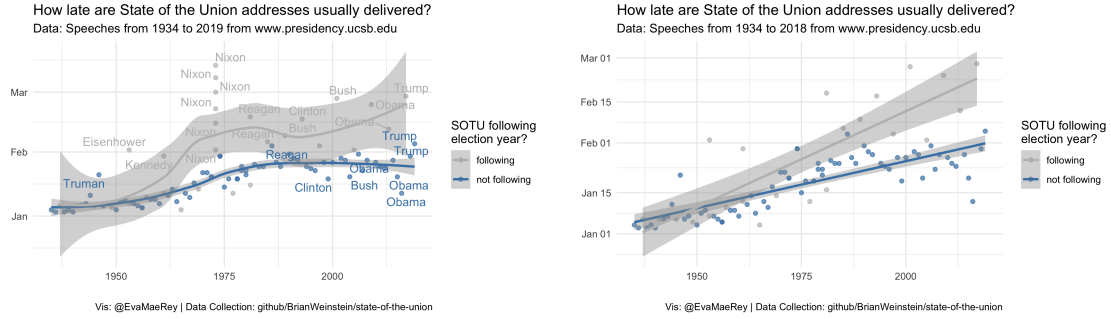Figure 2: Timing of state of the union by year. SOTUs that follow election years are colored with blue.

Figure 3: Timing of state of the union by year. SOTUs that follow election years are colored with blue.

8 References

Benoit, Kenneth, Kevin Munger, and Arthur Spirling. 2018. "Measuring and Explaining Political Sophistication Through Textual Complexity." *Available at SSRN 3062061*.

Benoit, Kenneth, and Paul Nulty. 2016. "Quanteda: Quantitative Analysis of Textual Data." *R Package Version 0.9* 8.

Silge, Julia, and David Robinson. 2016. "Tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *The Journal of Open Source Software* 1 (3): 37.