

## U.S. Presidents:

### What did they say? How did they say? Is Trump an alien?

Author: Jihan wei



The inauguration speech is treated as one of the most important speeches for U.S. Presidents. In this project, we applied texting mining techniques to explore what they said during their inauguration speech, what strategies they adopted, and what kind of emotion they intended to convey. Intriguingly, we identified a significant trend on their speaking strategy and interesting clusters of their topics.

Once we related their topics with the America history, we can obtain more inspiration. And based on the analysis of all the presidents, we also treated Trump as our interest to check whether he is alien in the American history.

---

## Part 1: Sentence Analysis:

Starting from the easiest method, we analyze the number and length of sentences for each president to identify interesting trend.

For the clear structure of our following analysis, we represent each president by the index of term rather than their name. And we refer to the following table to get the president name for necessary interpretation.

```
tables<-as.matrix(order)
colnames(tables)<-c("President Name")
rownames(tables)<-c(1:58)
tables
```

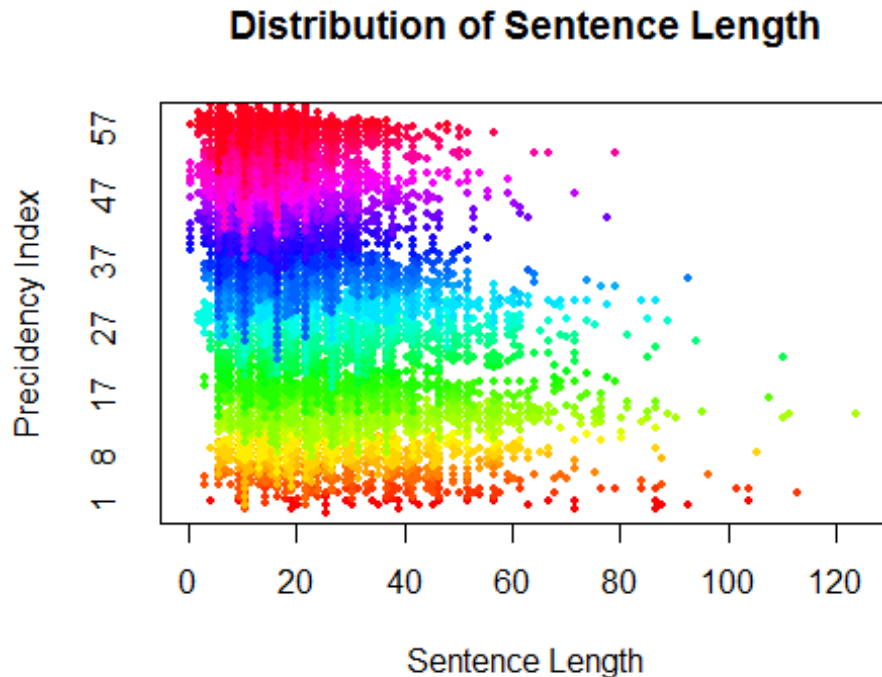
```
##      President Name
## 1  "GeorgeWashington-1"
```

## 2 "GeorgeWashington-2"  
## 3 "JohnAdams-1"  
## 4 "ThomasJefferson-1"  
## 5 "ThomasJefferson-2"  
## 6 "JamesMadison-1"  
## 7 "JamesMadison-2"  
## 8 "JamesMonroe-1"  
## 9 "JamesMonroe-2"  
## 10 "JohnQuincyAdams-1"  
## 11 "AndrewJackson-1"  
## 12 "AndrewJackson-2"  
## 13 "MartinvanBuren-1"  
## 14 "WilliamHenryHarrison-1"  
## 15 "JamesKPolk-1"  
## 16 "ZacharyTaylor-1"  
## 17 "FranklinPierce-1"  
## 18 "JamesBuchanan-1"  
## 19 "AbrahamLincoln-1"  
## 20 "AbrahamLincoln-2"  
## 21 "UlyssesSGrant-1"  
## 22 "UlyssesSGrant-2"  
## 23 "RutherfordBHayes-1"  
## 24 "JamesGarfield-1"  
## 25 "GroverCleveland-I-1"  
## 26 "BenjaminHarrison-1"  
## 27 "GroverCleveland-II-2"  
## 28 "WilliamMcKinley-1"  
## 29 "WilliamMcKinley-2"  
## 30 "TheodoreRoosevelt-1"  
## 31 "WilliamHowardTaft-1"  
## 32 "WoodrowWilson-1"  
## 33 "WoodrowWilson-2"  
## 34 "WarrenGHarding-1"  
## 35 "CalvinCoolidge-1"  
## 36 "HerbertHoover-1"  
## 37 "FranklinDRoosevelt-1"  
## 38 "FranklinDRoosevelt-2"  
## 39 "FranklinDRoosevelt-3"  
## 40 "FranklinDRoosevelt-4"  
## 41 "HarrySTruman-1"  
## 42 "DwightDEisenhower-1"  
## 43 "DwightDEisenhower-2"  
## 44 "JohnFKennedy-1"  
## 45 "LyndonBJohnson-1"  
## 46 "RichardNixon-1"  
## 47 "RichardNixon-2"  
## 48 "JimmyCarter-1"  
## 49 "RonaldReagan-1"  
## 50 "RonaldReagan-2"  
## 51 "GeorgeBush-1"  
## 52 "WilliamJClinton-1"  
## 53 "WilliamJClinton-2"  
## 54 "GeorgeWBush-1"  
## 55 "GeorgeWBush-2"  
## 56 "BarackObama-1"

```
## 57 "BarackObama-2"  
## 58 "DonaldJTrump-1"
```

Then, we use plot to show the trend:

```
beeswarm(sen_length,col=rainbow(58),pch=19,  
method="hex", cex=0.5, horizontal=TRUE,  
xlab="Sentence Length", ylab="Precidency Index",  
main="Distribution of Sentence Length")
```



The beeswarm plot indicates that from George Washington, presidents tend to use shorter sentences in their inauguration speech.

What's the interpretation? As we know:

**Shorter sentences are easier to follow**

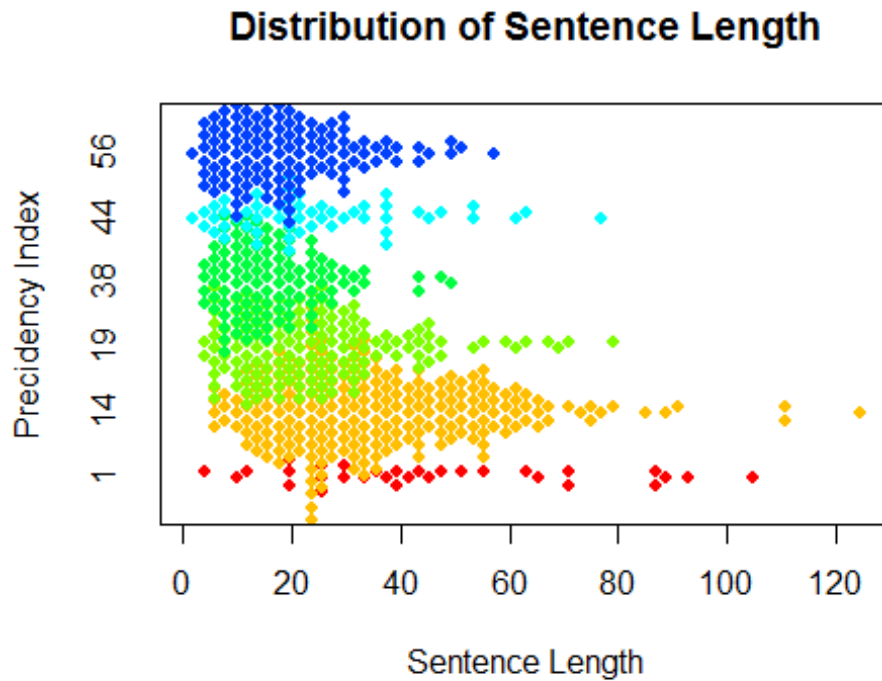
**Shorter sentences make more sense**

**Shorter sentences indicate approachability**

That is, they focus more on the communication with the public. They try to make sure that the public understand what they indent to convey and build a down-to-earth image. Then we select several famous presidents in different historical period: George Washington(1), John Tyler(14) Abraham Lincoln(19) Franklin D. Roosevelt (38), John F. Kennedy (44) and Barack Obama (56) to make comparision. The trend is evident.

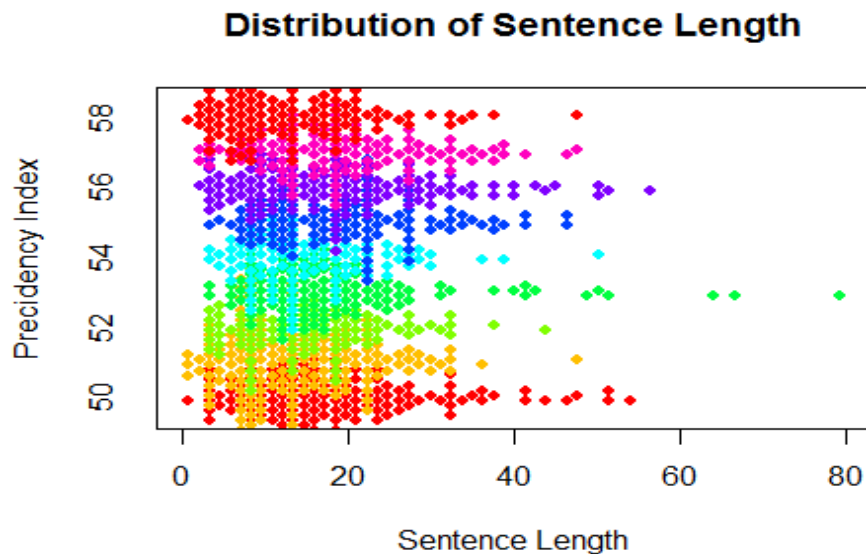
```
#  
ex1<-sen_length[c(1,14,19,38,44,56)]  
beeswarm(ex1,col=rainbow(8),pch=19,  
method="hex", cex=0.8, horizontal=TRUE,
```

```
xlab="Sentence Length", ylab="Precidency Index",
main="Distribution of Sentence Length")
```



Based on the historical trend, we explore the contemporary era and focus on Donald Trump(58).

```
###当代
ex2<-sen_length[50:58]
beeswarm(ex2,col=rainbow(8),pch=19,
method="hex", cex=0.8, horizontal=TRUE,
xlab="Sentence Length", ylab="Precidency Index",
main="Distribution of Sentence Length")
```



From the above picture, we can't detect significant difference between Trump and other presidents.

## Part 1 Conclusion:

Presidents tend to use more shorter sentences to keep effective communication between them and the public. In terms of this, Trump is not heterogeneous and adopt the same speaking strategy.



---

## Part 2: Topic Modelling

### 2.1: Topic Allocation

Besides how they give a speech, we may be also curious about what they say. Moreover, we are interested in how indicative their topics is to reveal the current situation of America. In part 2, we adopt topic modeling method to explore their speech topics and the weights on each topic for each president.

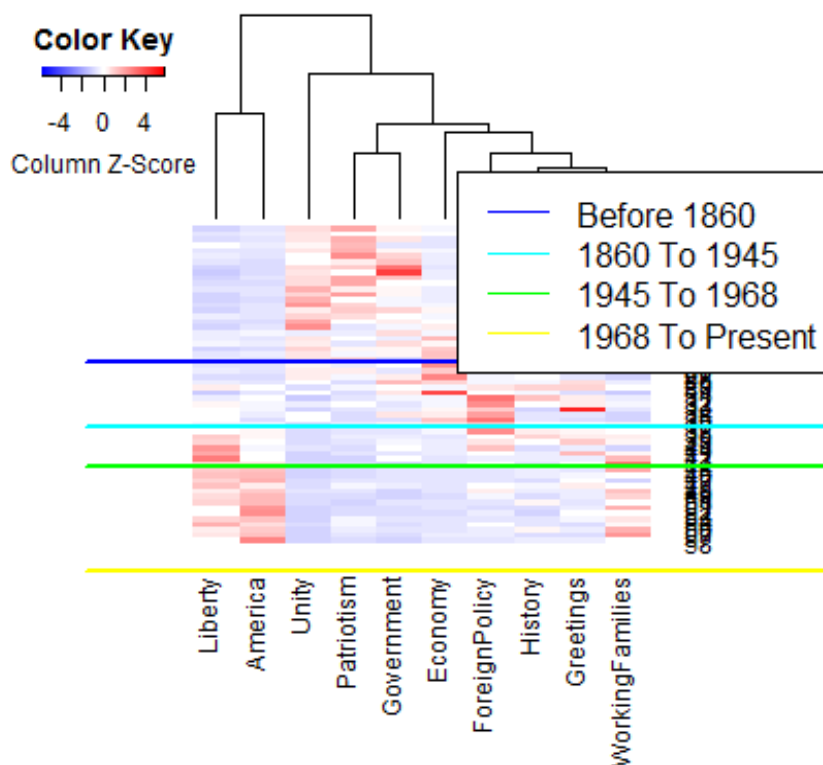
I set the topic numbers to be 10 and I manually tag them as "Patriotism", "Liberty", "America", "Economy", "History", "WorkingFamilies", "Greetings", "ForeignPolicy", "Government" and "Unity" after analyzing the keywords for each topic.



For example, I manually set Topic 1 as “Patriotism” for the keywords: Nation, Duty, Patriot. And I tag topic 2 as “Liberty”, since it contains keywords: People, Freedom, Right, Liberty; I tag Topic 3 as “Government”, since it contains keywords like: Govern, Law, Congress, Secure; I tag Topic 9 as “ForeignPolicy” for the keywords: War, Nation, Peace, Import. Moreover, I tag Topic 14 as “Unity” for the keywords: One, Union, People, Can, State.

```
topics.hash<-c("Patriotism","Liberty","America","Economy","WorkingFamilies", "Greetings",
"History", "ForeignPolicy", "Government","Unity")
ldatopic<-as.vector(ldaOut.topics)
ldahash<-topics.hash[ldaOut.topics]
colnames(topicProbabilities)<-topics.hash
```

Then, I use heatmap plot to see the weight allocation of topics for each president. And an indicative trend appears.



Note that red color indicates more weights on that topic. From the heatmap, we identify interesting trend: presidents that in same era tends to have similar weight allocation among the topics. And when relating them with the America history, the results are not spurring and even indicative.

### Interpretation:

We add reference lines to indicate different era. We can see, before 1860, which is the early years of the new American nation, presidents talked more about “Unity”, “Government” and “Patriotism”. It reveals that in that period, presidents wanted to show the power of government and unify people to contribute to the new country.

Between 1860 to 1945, which is the period of American Civil War and the world wars, presidents tended to focus more on “Economy”. This is reasonable since after several wars (especially after the Civil War), developing economy is the most urgent task.



And from 1945 to 1968, which is treated as the Post-war period, presidents needed to develop the economy and establish good diplomatic relationships, thus they talked more about “Economy” and “Foreign Policy”.

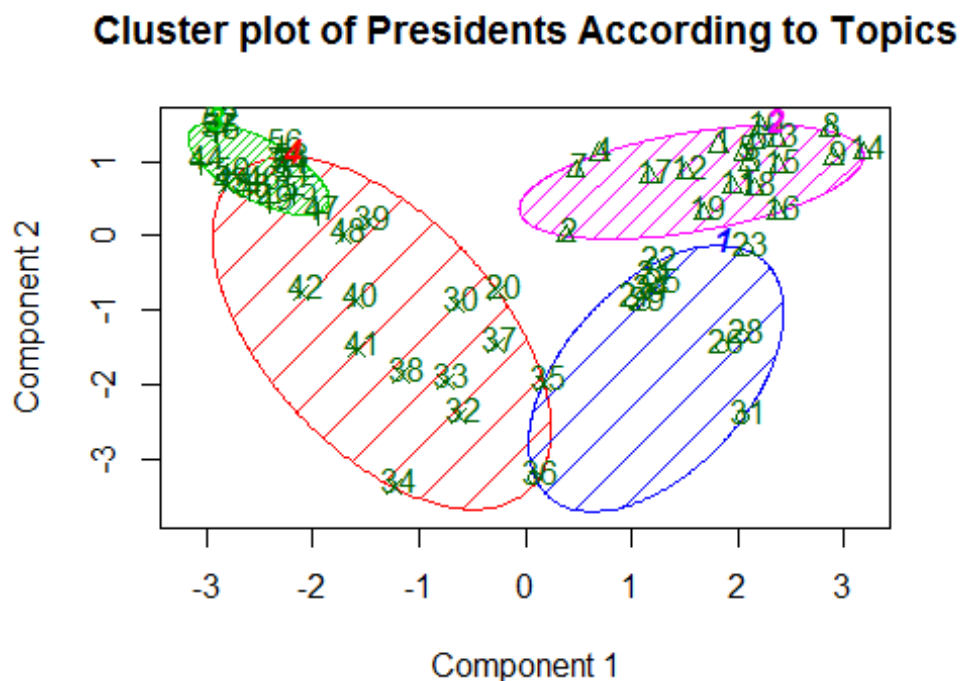
As for during the contemporary era, after severe problems have been released, presidents put their emphasis on individuals and even focus more the spiritual level so they usually talked about “Liberty”, “America” and “Working Families”. Moreover, they tend to focus on working families to gain more support.

In all, we believe what they say can indicate what challenge America is faced.

## 2.2: Clustering of presidents according to topics

Suddenly coming to the above interpretation may be not convincing. In the part, we perform presidents clustering based on the topics to see whether our belief is correct.

```
####Clustering
mydata<-as.matrix(topicProbabilities)
fit <- kmeans(mydata,4,nstart=20)
#fit$cluster[58]
clusplot(mydata, fit$cluster, color=TRUE, shade=TRUE, labels=2,lines = 0,main="Cluster plot of Presidents According to Topics")
```



From the plot, we can see, the K-mean clustering provide support for our interpretation:

**Contemporaneous presidents are clustered in same group**

This means that during different historical periods, America is faced with different challenges and presidents Inauguration speech will talk more about those challenges. We can gain information about current situations about America from president’s speech.

Now, let's have a further analysis.

```
fit$cluster

## [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 1 1 1 1 1 1 1 1 1 4 1 4 4 4 4
## [36] 1 4 4 4 4 4 4 4 3 3 3 3 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

fit$cluster[58]

## [1] 3
```

We see that cluster 3 mainly contains the contemporary presidents, and Trump is assigned to that Group. We may conclude that Trump doesn't deviate a lot from the main topics of current America. Within the basic structure, Trump may have his own trait. Now, let's have a look at what Trump said:

```
docs2<-Corpus(VectorSource(unlist(text.list)))
docs2<-tm_map(docs2, stripWhitespace)
docs2<-tm_map(docs2, content_transformer(tolower))
docs2<-tm_map(docs2, removeWords, stopwords("english"))
docs2<-tm_map(docs2, removeWords, character(0))
docs2<-tm_map(docs2, removePunctuation)
dtm2<- DocumentTermMatrix(docs2, control = list(weighting = function(x) weightTfIdf(x, normalize = FALSE), stopwords = TRUE))
ff.dtm2<-tidy(dtm2)
wordcloud(ff.dtm2$term[ff.dtm2$document==58], ff.dtm2$count[ff.dtm2$document==58], scale=c(2.5,.2), min.freq=2,
          max.words=100, random.order=FALSE, rot.per=.15, colors=brewer.pal(8, "Dark2"))
```



The most significant words are “America”, “Borders”, “Protection”, “Dreams” and “Everyone”.

## Part 2 Conclusion:

From the topic modelling, we conclude that inauguration speech can reveal the problems that American was faced and can indicate the emphasis of the president's policy.

For Trump, as one of the controversial president, he doesn't make himself so different in term of the topics. While the key words show that he may focus more on focus more on:

protectionism, nationalism and “tried” to “protect” “America” and each American individual.

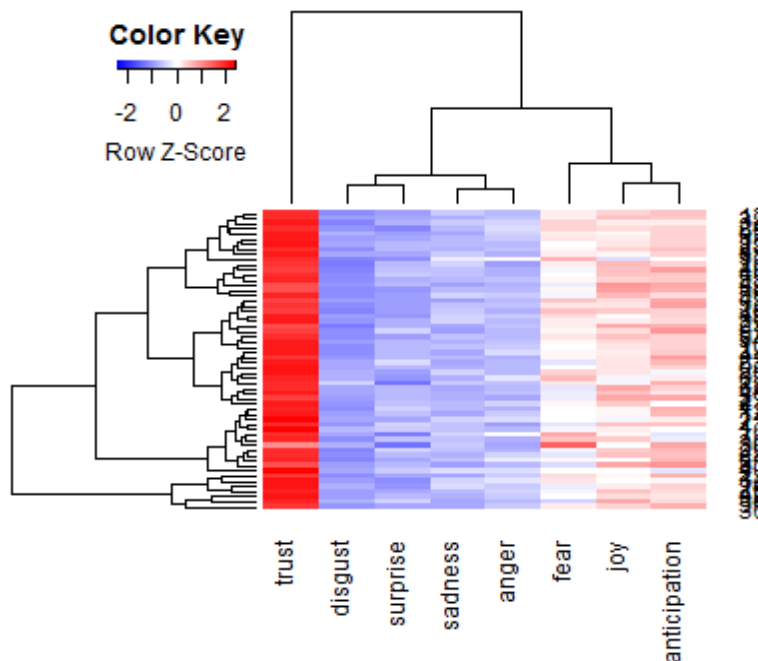




### Part 3: Sentiment Analysis

Apart from the topics and sentence lengths, we also want to know what kind of emotions they wanted to convey during their speech. Therefore, we conducted the sentiment analysis to identify the emotion contained in each sentence of their speeches.

We incorporate “trust”, “disgust”, “surprise”, “sadness”, “fear”, “joy”, and “anticipation” as emotional categories. Based on the result for each sentence, we summarize them and plot the heat map to see the weights for different emotion.



The above graph indicates that, different from the weight allocation of topics, nearly all the presidents convey more positive emotion like Trust, Joy and Anticipation.

**Yes! The presidents want to convey a positive emotion to the public.**

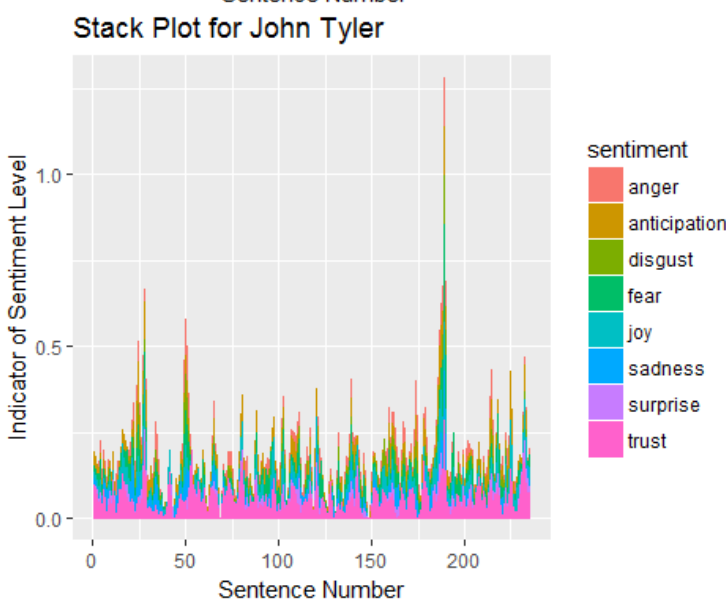
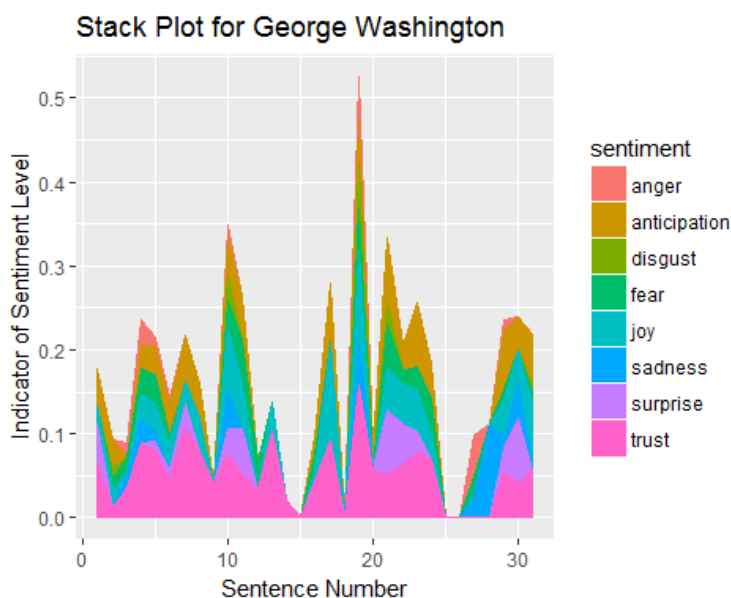
Now, it could be interesting to see who is the most “negative” president.

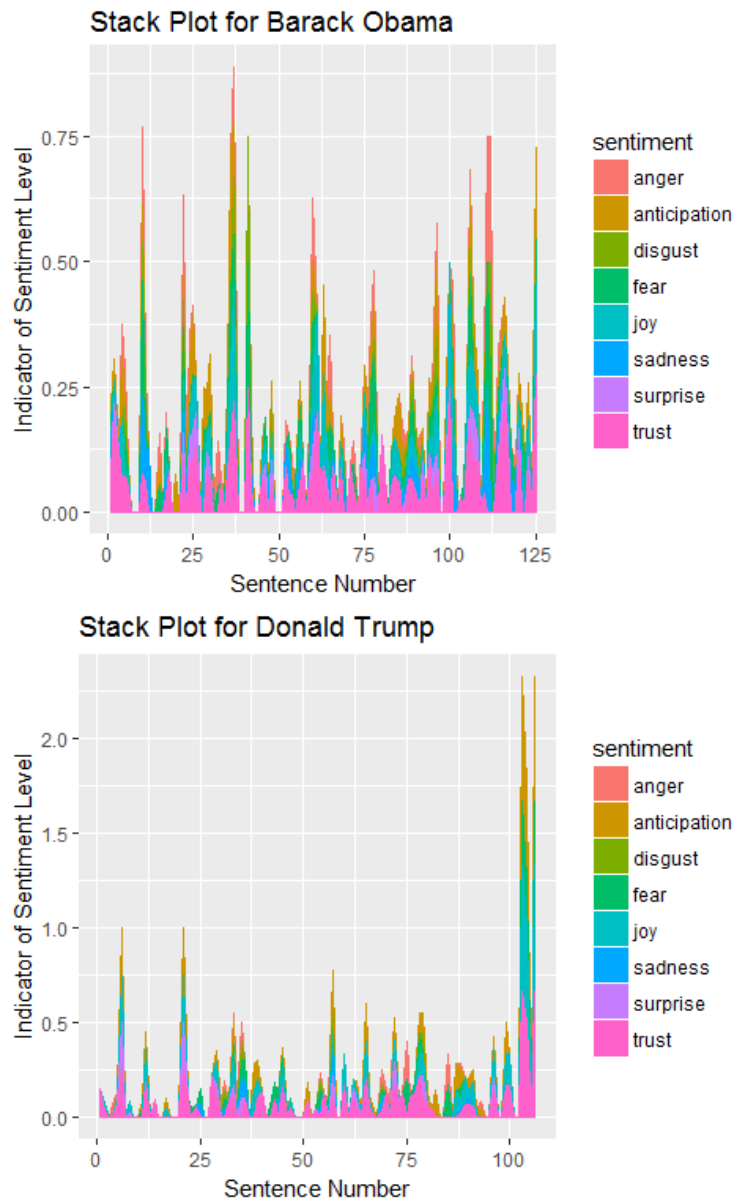
```
n<-which.max(data.frame(emotion.total)$negative)
order[n]
```

```
## [1] "WilliamHenryHarrison-1"
```

It's William Henry Harrison! You could find it understandable once you know more about his stormy life experience and the fact that he is the “shortest-life” president in American history.

Generally, all presidents adopted a positive attitude during their speech. But, further analyses indicate that the flow of the emotion are different: some are moderate, some are more passionate. We can make the stack plot to further explore it.





From the stack plot of several selected presidents, we can see, for John Tyler, there are few fluctuations of emotion from sentence to sentence. However, for Obama, who is often treated as an excellent orator, he is more spirited and full of emotional fluctuations. As for Trump, he has his own way of speaking: he kept relatively placid at the beginning and tried to kindle people in the end.

## What do we get?

- In all, presidents tend to use more shorter sentences in their speech to interest people and to build a down-to earth image; Their speech topics cluster according to different historical era, which provides valuable information about national conditions of America. Moreover, nearly all of them adopt a positive attitude in their speech and focus on “Trust” heavily, and this reveals the nature of the inauguration speech.
- As for Trump, although interesting differences exist, we can’t treat him as an alien president based on the speech analysis. After all, that they do is much more important than what he says.
- Is Trump an alien? Maybe time will tell us.

---

## Reference:

1. US.Presidents:  
From[https://www.google.com/search?q=%E7%BE%8E%E5%9B%BD%E5%8E%86%E5%8F%B2%E6%80%BB%E7%BB%9F&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiWxY6puOzRAhXNdSYKHb96D44Q\\_AUICSgC&biw=1563&bih=892](https://www.google.com/search?q=%E7%BE%8E%E5%9B%BD%E5%8E%86%E5%8F%B2%E6%80%BB%E7%BB%9F&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiWxY6puOzRAhXNdSYKHb96D44Q_AUICSgC&biw=1563&bih=892)
2. Trump1:  
From[https://www.google.com/search?q=TRUMP+CARTOON&biw=1563&bih=892&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjUo43k0-zRAhWFJCYKHcIFBlgQ\\_AUIBigB](https://www.google.com/search?q=TRUMP+CARTOON&biw=1563&bih=892&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjUo43k0-zRAhWFJCYKHcIFBlgQ_AUIBigB)
3. Trump2:  
From[https://www.google.com/search?q=%E7%BE%8E%E5%9B%BD%E5%8E%86%E5%8F%B2%E6%80%BB%E7%BB%9F&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiWxY6puOzRAhXNdSYKHb96D44Q\\_AUICSgC&biw=1563&bih=892](https://www.google.com/search?q=%E7%BE%8E%E5%9B%BD%E5%8E%86%E5%8F%B2%E6%80%BB%E7%BB%9F&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiWxY6puOzRAhXNdSYKHb96D44Q_AUICSgC&biw=1563&bih=892)