Understanding the evolution of the American's concerns leveraging the recent Presidential Speeches

Vassily Carantino - vc2434

Wednesday, September 20, 2017

Introduction: Presidential Speeches are telling us a lot about American people's concerns.

Figure 1: Concerns of the American people are evolving

Inaugural speech is the first official speech of any president of United States. The President elected is supposed to address American people concerns and offer a plan to answer these concerns during his mandate. As a result, the Presidential inaugural speeches tell a lot about those concerns and their evolution. In this work, we will leverage the presidential inaugural speeches to analyse the evolution of the concerns of American people.

Even if the President elected is supposed to represent the entire nation, it is clear that the points the president will address are mainly corresponding to the one of the very people that voted for him. We will then have to see what are the respective concerns of Republican and Democrat people, and how these concerns are evolving. In order to have a more interesting analysis, we will focus on recent elections only: starting from Bill Clinton in 1993, which gives approximately 25 years of recent American History.

Step 0 - Data Preparation:

Step 0 (i)- change directory:

```
setwd('C:/Users/vc2434/Documents/fall2017-project1-Flaar-master')
```

Step 0 (ii) - Check and install needed packages. Load the libraries and functions.

This notebook was prepared with the following environmental settings.

```
##                _
## platform       x86_64-w64-mingw32
## arch           x86_64
## os             mingw32
## system         x86_64, mingw32
## status
## major          3
## minor          3.2
## year           2016
## month          10
## day            31
## svn rev        71607
## language       R
## version.string R version 3.3.2 (2016-10-31)
## nickname       Sincere Pumpkin Patch
```

Step 1 - load the speeches from Bill Clinton to Donald Trump

```
folder.path=paste0(getwd(),"/data/InauguralSpeeches/")
speeches=list.files(path = folder.path, pattern = "*.txt")
speeches=speeches[c(5,6,9,20,21,52,53)] #we select the speeches from Bill Clinton first mandate to Dona
prex.out=substr(speeches, 6, nchar(speeches)-4)

ff.all<-Corpus(DirSource(folder.path))
```

Step 2 - Text formating and computation of TF-IDF weighted document-term matrices for individual speeches

```
ff.all<-tm_map(ff.all, stripWhitespace)
ff.all<-tm_map(ff.all, content_transformer(tolower))
ff.all<-tm_map(ff.all, removeWords, stopwords("english"))
ff.all<-tm_map(ff.all, removeWords, character(0))
ff.all<-tm_map(ff.all, removePunctuation)

# compute TF-IDF weighted document-term matrices
tdm.all <- DocumentTermMatrix(ff.all,
                              control = list(weighting = function(x)
```

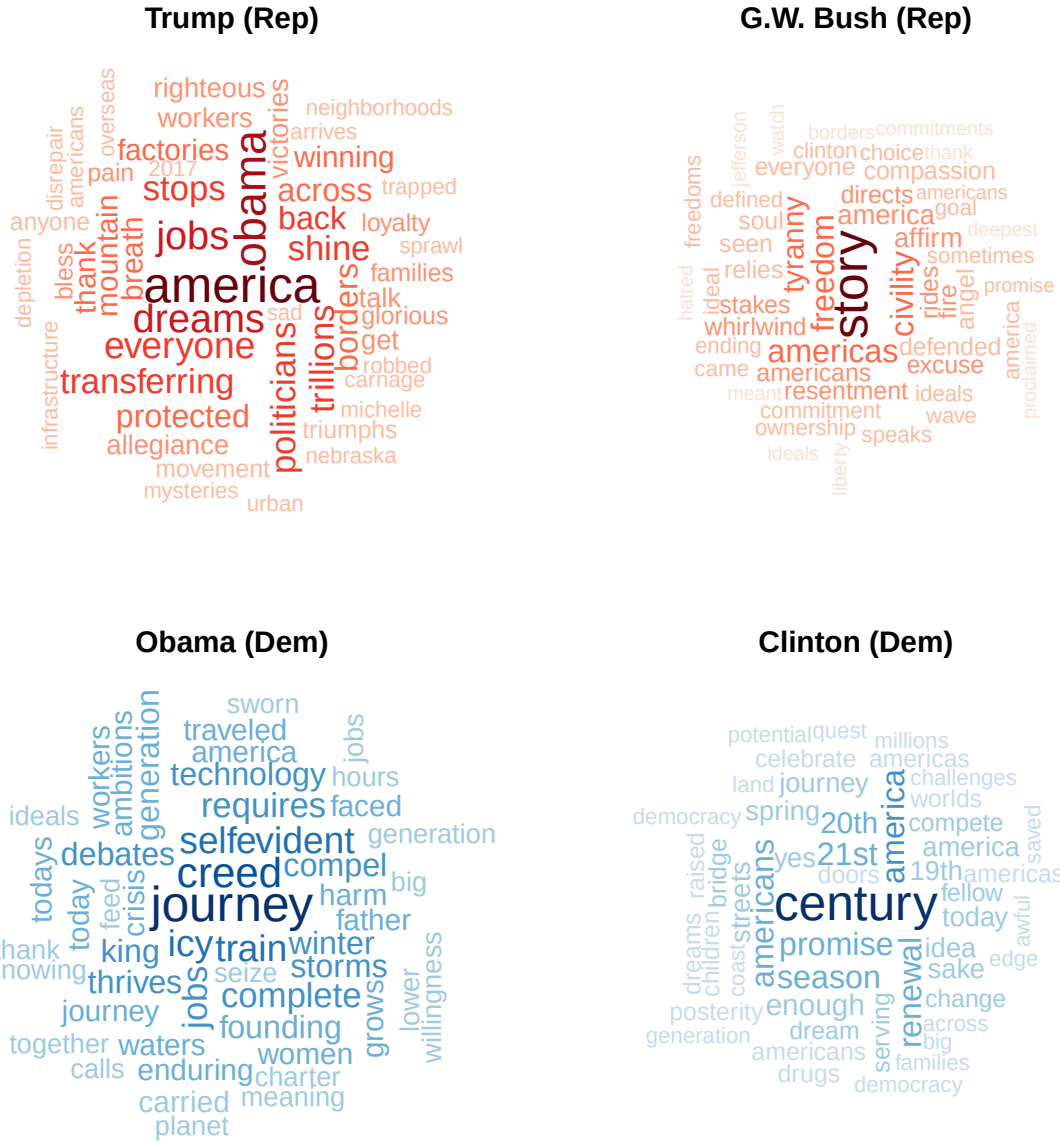
```
                                                weightTfIdf(x,
                                                            normalize =FALSE),

                                            stopwords = TRUE))
tdm.tidy.all=tidy(tdm.all)


# select only speeched of interest and create a Document term Matrix for each one of the president

tdm.tidy= subset(tdm.tidy.all,document %in% speeches)

tdm.trump = subset(tdm.tidy,document == 'inaugDonaldJTrump-1.txt')
tdm.obama = subset(tdm.tidy,document == "inaugBarackObama-1.txt" | document =="inaugBarackObama-2.txt")
tdm.obama<-tdm.obama[order(tdm.obama$count,decreasing = TRUE),]
tdm.obama = tdm.obama[-1,]
tdm.bush = subset(tdm.tidy,document == "inaugGeorgeWBush-1.txt" | document =="inaugGeorgeWBush-2.txt")
tdm.clinton = subset(tdm.tidy,document == 'inaugWilliamJClinton-1.txt'| document =="inaugWilliamJClinto
```

Step 3 - Analzing each president style:

We first analyse each president verbal style, in order to extract potential verbal tics and repetitive elements of langague.

```
op = par(mfrow=c(2,2), mar = c(0, 0, 3, 0))

wordcloud(tdm.trump$term, tdm.trump$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Reds"))
title(main = 'Trump (Rep)')

wordcloud(tdm.bush$term, tdm.bush$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Reds"))
title(main = 'G.W. Bush (Rep)')

wordcloud(tdm.obama$term, tdm.obama$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
```

```
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
title(main = 'Obama (Dem)')

wordcloud(tdm.clinton$term, tdm.clinton$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
title(main = 'Clinton (Dem)')
```




```
par(op)
```

On these words clouds we can see that every and each president has his own style and some words might happen as very frequent just because of the president style. The more blatent example of this is with President Georges W. Bush that keeps using the word 'story' that doesn't tell us much about the American people worries. we want to get rid of these effects and manually suppress the most obvious elements of language that doesn't tell much about the speech contempt.

step 4 - Cleaning data, mannerisms of speech and verbal tics:

```
"%ni%" <- Negate("%in%")

tdm.obama<-tdm.obama[order(tdm.obama$count,decreasing = TRUE),]
tdm.obama = subset(tdm.obama,term %ni% c("requires","complete","compel","today"))


tdm.bush<-tdm.bush[order(tdm.bush$count,decreasing = TRUE),]
tdm.bush = subset(tdm.bush,term %ni% c("story","sometimes","seen"))


tdm.clinton<-tdm.clinton[order(tdm.clinton$count,decreasing = TRUE),]
tdm.clinton = subset(tdm.clinton,term %ni% c("enough","yes","today"))

tdm.trump<-tdm.trump[order(tdm.trump$count,decreasing = TRUE),]
tdm.trump = subset(tdm.trump,term %ni% c("get","talk"))



# Create a Document term Matrix for each presidential inauguration speech, in order to analyse their ev

tdm.obama.1=subset(tdm.obama,document == "inaugBarackObama-1.txt")
tdm.obama.2=subset(tdm.obama,document == "inaugBarackObama-2.txt")
tdm.bush.1 = subset(tdm.bush,document == "inaugGeorgeWBush-1.txt")
tdm.bush.2 = subset(tdm.bush,document == "inaugGeorgeWBush-2.txt")
tdm.clinton.1 = subset(tdm.clinton,document == 'inaugWilliamJClinton-1.txt')
tdm.clinton.2 = subset(tdm.clinton,document == 'inaugWilliamJClinton-2.txt')
```

step 5 - Plotting the wordClouds of all speeches in chronological order:

```
op = par(mfrow=c(2,4), mar = c(0, 0, 3, 0))


wordcloud(tdm.clinton.1$term, tdm.clinton.1$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
title(main = 'Clinton Term 1 (Dem)')


wordcloud(tdm.clinton.2$term, tdm.clinton.2$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
```

```
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
title(main = 'Clinton Term 2 (Dem)')

wordcloud(tdm.bush.1$term, tdm.bush.1$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Reds"))
title(main = 'G.W. Bush term 1 (Rep)')

wordcloud(tdm.bush.2$term, tdm.bush.2$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Reds"))
title(main = 'G.W. Bush Term 2(Rep)')

wordcloud(tdm.obama.1$term, tdm.obama.1$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
title(main = 'Obama Term 1 (Dem)')

wordcloud(tdm.obama.2$term, tdm.obama.2$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
title(main = 'Obama Term 2 (Dem)')


wordcloud(tdm.trump$term, tdm.trump$count,
          scale=c(2,0.5),
          max.words=50,
          min.freq=1,
```

```
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Reds"))
title(main = 'Trump (Rep)')

par(op)
```

Clinton Term 1 (Dem)

Clinton Term 2 (Dem)

G.W. Bush term 1 (Rep)

G.W. Bush Term 2(Rep)

Obama Term 1 (Dem)

Obama Term 2 (Dem)

Trump (Rep)

Here is the wordclouds of all the different presidential inaugural speeches since 1993. We have corrected for verbal tics and we can say that these clouds represent the concerns of the American people.

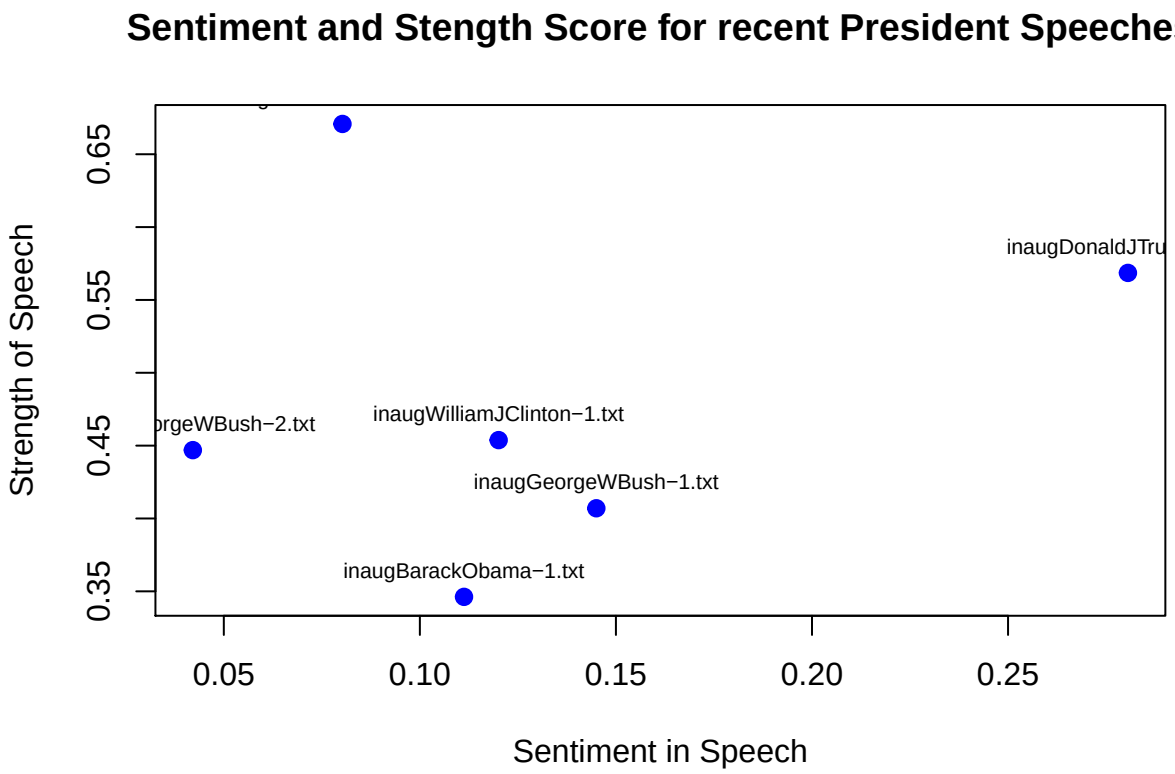
step 6 - Analysis

- Clinton first speech is very positive, it's the begining of a new 'season', where 'america' will have a 'renewal'. We see that at the time, the election has mainly be driven by a desire of 'change' as the previous Bush administration was in a 'deadlock'.
- Clinton second mandate speech is also a speech of hope just a few years before the change of century. Clinton is focusing on this transition from the 19th to the 20st century, and seems to be presenting the 'dream' and the future 'journey' of america in this new century. We can also see that new points of concerns arose: Clinton also stress the importance of 'education', the 'values' of 'family', 'labor'. He is also putting some stress on 'African american people'.
- Bush first mandate is very different with less diversity in terms of words. The main points conveyed by his speach are the values of 'civility', 'courage' and 'commitment'. For the first time, we can see that religion has an important place in the debate with words like: 'mosquee', 'pastor', 'synagogue'.
- The second term of Bush is very different. We are after the 9/11 events and the USA are in Irak. we can see a lot of terms that are linked to these events: Americans has to 'defend' themselves against 'tyranny', 'hatred', 'resentment', to preserve their 'freedom'. For the first time again, we can see the apparition of the 'borders' and 'immigration' issue.
- Obama first speech is very different. It focuses on a lot of new concerns that arised in the society or that were ignored by the precedent administration. We are in 2009 and the word is facing an important financial crisis. Obama is of course adressing this with words like 'jobs', 'winter', 'hours', 'grows'. But Obama is also focusing on climate change for the first time, with the words: 'icy', 'storms', 'planet', 'waters'. Once again, the religion is again very present in his speech with words such as: muslims, nonbelievers, ideals. For the first time, 'women' appears in the word cloud.
- Obama second speech is again addressing new challenges that are araising: the one related to 'technology', but also the one of the US health system with words such as 'medicaid'. The challenge of guns legislation is also present here with words like 'Newtown' and 'harm'. -Finally the more recent speach is the one of Donald Trump. His speech is focusing on 'America' and 'Americans' that Donald Trump want to 'protect' bringing 'back' 'jobs' and 'factories', closing the 'borders'. We can see that his campaign is mainly directed towards the people that hated the 'Obama' administration and 'politicians' in general.

step 7 - Sentiment Analysis of the presidential Speeches:

```
data(hash_sentiment_nrc)
data(hash_strength)
tdm.tidy.all=tidy(tdm.all)

# Adding the sentiment scores to each word:
tdm.tidy.all = merge(tdm.tidy.all, hash_sentiment_nrc, by.x = 'term', by.y = 'x')
names(tdm.tidy.all)[dim(tdm.tidy.all)[2]] = 'sentiment'
tdm.tidy.all = merge(tdm.tidy.all, hash_strength, by.x = 'term', by.y = 'x')
names(tdm.tidy.all)[dim(tdm.tidy.all)[2]] = 'strength'


# Weighting the results
tdm.tidy.all$sentiment = tdm.tidy.all$count * tdm.tidy.all$sentiment
tdm.tidy.all$strength = tdm.tidy.all$count * tdm.tidy.all$strength


# Computing average sentiment and strength of the speech: we obviously need to take into account the sp
tdm.tidy.all = summarise(group_by(tdm.tidy.all, document), sum(count),sum(sentiment),sum(strength))
tdm.tidy.all$average_sentiment = tdm.tidy.all$`sum(sentiment)`/tdm.tidy.all$`sum(count)`
tdm.tidy.all$average_strength = tdm.tidy.all$`sum(strength)`/tdm.tidy.all$`sum(count)`
```

```
# Table: presidential speeches of interest:
Sentiment_Strength_score_subset <-subset(tdm.tidy.all,document %in%  c('inaugWilliamJClinton-1.txt','ina
Speech_Score_subset<-Sentiment_Strength_score_subset[,c('document', 'average_sentiment', 'average_streng

plot(Speech_Score_subset$average_sentiment, Speech_Score_subset$average_strength, main= 'Sentiment and S
text(Speech_Score_subset$average_sentiment, Speech_Score_subset$average_strength, labels= Speech_Score_s
```



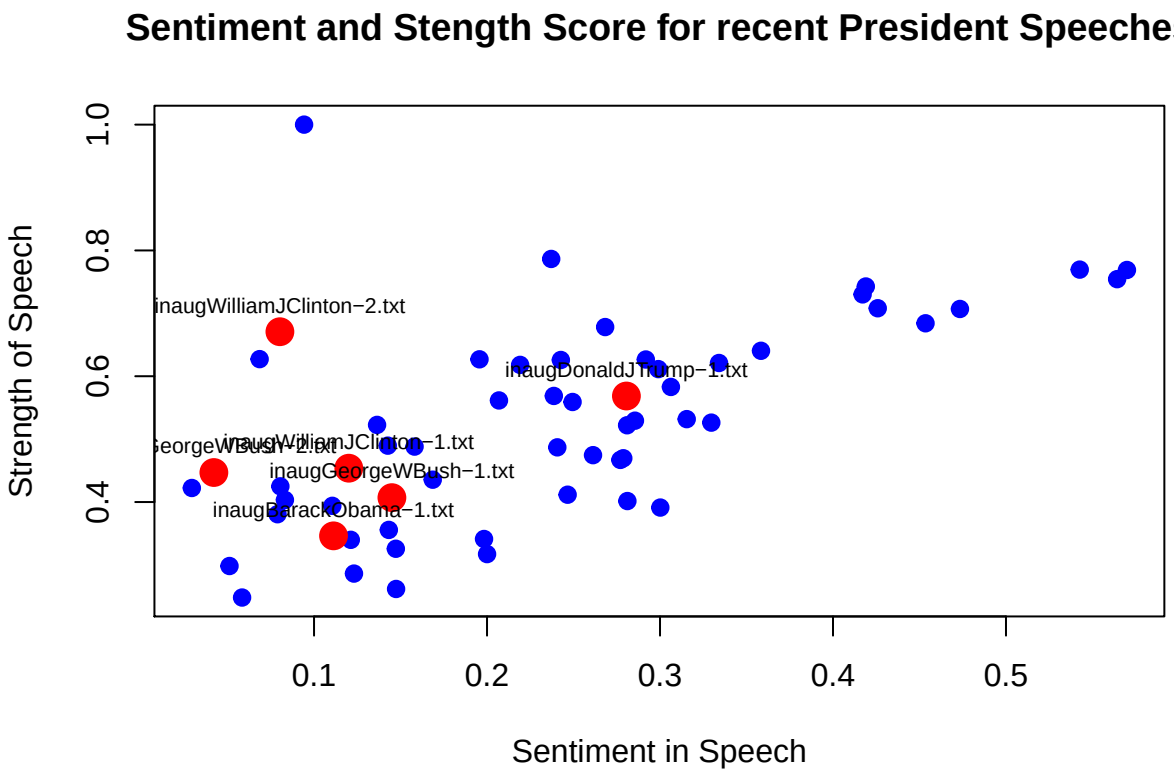
In this part, I use the hash_sentiment_nrc and the hash_strength datasets to assign to each token a sentiment and a strength score that can be in the set (0,1,-1). We then multiply this by the occurence of the world in the speech and average it using the number of words prononced in the speech. This method gives us a score for each speech in terms of its average strength and the average sentiments the speech is using.

We can see that the presidents have very different style and their speech convey very different levels of strength and sentiment. But we also notice that depending on the speech, the message conveyed can be very different for the same president.

We can note here that - according to this simple analysis- Trump style seems to be the most effective as it has a both high strength and high sentiments in his speech. On the contraty, Obama migth use a much more elaborated and metaphorical vocabulary that makes him have a somewhat low score both in terms of strength and sentiment. We note that there is no clear patterns of appartenance to one political movement in this data. Let's use the whole data set and compare the score of the recent presidents against all the previous US presidents.

step 8 - Comparing recent President speeches to past ones:

```
tdm.tidy.all<-tdm.tidy.all[,c('document', 'average_sentiment', 'average_strength')]

plot(tdm.tidy.all$average_sentiment, tdm.tidy.all$average_strength, main= 'Sentiment and Stength Score f


points(subset(tdm.tidy.all,document %in%  c('inaugWilliamJClinton-1.txt','inaugWilliamJClinton-2.txt',
text(subset(tdm.tidy.all,document %in%  c('inaugWilliamJClinton-1.txt','inaugWilliamJClinton-2.txt', 'in
```



We can see that Trump is finally not an outlier, past Presidents have used much more strength and sentiments in their inaugural speeches. This data is very rich and we could conduct a lot of different analysis on it: clustering analysis, correlation with political party and so forth. I unfortunately didn't get the time for all these.

Conclusion

This analysis allowed us to roughly understand the evolution of the worries and concerns of American people. We observe that a new presidency is a moment of transition where one of the main concern being addressed is the desire of change. This desire of change can follow a presidency where a large part of the US population disagreed with the administration choices and decision, in that case we observe a lot of negative words. But these inaugural speeches also lay out a new future and bring a lot of hope.

But this analysis - and this could be detailed- also allowed us to learn about the evolution of the US society concerns. for instance, the concern of terrorism appears after the terrible events of 9/11, adressing climate

change and equal wages for women also appeared quite late in the presidential speeches, showing the evolving importance of these issues.

Finally, a sentiment analysis can help us understand the strentgh and the sentiments that conveys a speech. Donald Trump was elected as a movement to radically change politics and the way the US society is working. As a result, his speech conveys a lot of hope (positive sentiments) and also has a lot of strength.

Sources: http://www.statmethods.net/advstats/cluster.html https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf