Is Trump really an alien compared to other presidents after the cold war?

Lin Han(lh2810)

As we know, America have become the only superpower after the cold war and Donald J Trump became the 45th America presidents on Jaunary 20th, 2017. Compared to other presidents in the past, he seems so different. He seems highly emotional, highly bold, highly mischievous and highly sociable which conflicts with our traditional knowledge of politicians. So people may ask "Is he really like an alien compared to other presidents?" In this project, in order to coincide with the background of the times, we will investigate about it through inaugural speeches of presdents after cold war.

Step 0: check and install needed packages. Load the libraries and functions.

```r
#packages.used=c("tm", "wordcloud", "RColorBrewer",
#                "dplyr", "tydytext","rvest", "tibble", "qdap",
#                "sentimentr", "gplots", "dpl.yr",
#                "tm", "syuzhet", "factoextra",
#                "beeswarm", "scales", "RColorBrewer",
#                "RANN", "tm", "topicmodels")

# check packages that need to be installed.
#packages.needed=setdiff(packages.used,
#                        intersect(installed.packages()[,1],
#                                  packages.used))
# install additional packages
#if(length(packages.needed)>0){
#  install.packages(packages.needed, dependencies = TRUE)
#}
#install.packages("tidytext",dependencies=TRUE,repos='http://cran.rstudio.com/')
# load packages
# You mlibrary("rvest")
# sudo ln -f -s $(/usr/libexec/java_home)/jre/lib/server/libjvm.dylib /usr/local/lib
# in order to load qdap

library("rvest")
library("tibble")
library(tm)
library(wordcloud)
library(RColorBrewer)
library(dplyr)
library(tidytext)
library("qdap")
library("sentimentr")
library("gplots")
library("dplyr")
library("tm")
library("syuzhet")
library("factoextra")
library("beeswarm")
library("scales")
library("RColorBrewer")
library("RANN")
library("tm")
library("topicmodels")

source("/Users/linhan/Desktop/semester 3/ADS/project1/Fall2017-Project1-RNotebook/lib
/plotstacked.R")
source("/Users/linhan/Desktop/semester 3/ADS/project1/Fall2017-Project1-RNotebook/lib
/speechFuncs.R")
```

Step1: WordCloud Comparison

In the first step, we will investigate the main words of Trump's and other presidents' speeches to find some inspirations.

```
folder.path="/Users/linhan/Desktop/semester 3/ADS/project1/Fall2017-Project1-RNoteboo
k/data/InauguralSpeeches/"
speeches=list.files(path = folder.path, pattern = "*.txt")
prex.out=substr(speeches, 6, nchar(speeches)-4)
ff.all<-Corpus(DirSource(folder.path))
```
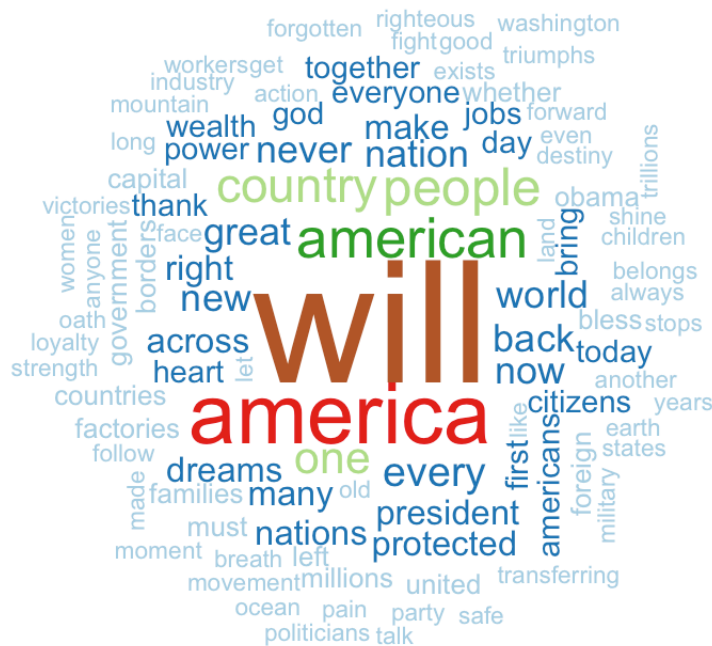
```
ff.all<-tm_map(ff.all, stripWhitespace)
ff.all<-tm_map(ff.all, content_transformer(tolower))
ff.all<-tm_map(ff.all, removeWords, stopwords("english"))
ff.all<-tm_map(ff.all, removeWords, character(0))
ff.all<-tm_map(ff.all, removePunctuation)
```

```
#all other presidents' WordCloud

tdm.all<-TermDocumentMatrix(ff.all)
tdm.tidy=tidy(tdm.all)
tdm.all = subset(tdm.tidy,document == 'inaugGeorgeBush-1.txt'|document =='inaugWillia
mJClinton-1.txt' |document =='inaugWilliamJClinton-2.txt' |document == 'inaugGeorgeWB
ush-1.txt'|document =='inaugGeorgeWBush-2.txt'|document =='inaugBarackObama-1.txt'|do
cument =='inaugBarackObama-2.txt')
#tdm.overall=summarise(group_by(tdm.tidy, term), sum(count))
wordcloud(tdm.all$term, tdm.all$count,
          scale=c(3,0.5),
          max.words=100,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.3,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(9,"Blues"))
```

```
#Trump's WordCloud
tdm.trump = subset(tdm.tidy,document == 'inaugDonaldJTrump-1.txt')
wordcloud(tdm.trump$term, tdm.trump$count,
          scale=c(5,0.5),
          max.words=100,
          min.freq=1,
          random.order=FALSE,
          rot.per=0.15,
          use.r.layout=T,
          random.color=FALSE,
          colors=brewer.pal(12,"Paired"))
```

We can see from two pictures that Trump and other presidents all focus on people's rights and America future's develpment. However, there are four obvious differences between Trump and other presidents.

Firstly, Trump is more concerned about the national economy, such as jobs, wealth while the other presidents are more concerned about spiritual aspects of things like freedom and liberty. This is easily to understand for the reason that the US economy has not been completely recovered from the 2008 economic crisis and people are really looking forward to the new president to take more measures to reduce the unemployment rate and improve the national wealth value.

Secondly, compared to other presidents, Trump reduced the importance of word–'world' and we can find that some words like 'protected', 'erveryone' and 'every' appear in the picture. This corrsponds to Trump' altitude towards immigration and his focus on protecting US citizens' career opportunities.

Thirdly, Trump stressed the word–'america' and 'great' compared to other presidents. Besides, he also used the word–'back' and 'dreams'. This means Trump tried to give people more confidence and this also corrsponded to his nostalgic slogan–'Make America Great Aagin.'

Lastly, on the one hand, those other presidents mostly used some affirmative words like 'will' and 'must'. On the other hand, Trump often used the word–'never'. This shows his personality character to some extent.

step2:data Processing — generate list of sentences

In the second step, we want to look into the speeches by seperating the speeches into sentences.

```
main.page <- read_html(x = "http://www.presidency.ucsb.edu/inaugurals.php")
inaug=f.speechlinks(main.page)
#as.Date(inaug[,1], format="%B %e, %Y")
inaug=inaug[-nrow(inaug), ]
inaug.list=read.csv("/Users/linhan/Desktop/semester 3/ADS/project1/Fall2017-Project1-
RNotebook/data/inauglist.csv", stringsAsFactors = FALSE)
speech.list= inaug.list
speech.url=inaug
speech.list=cbind(speech.list, speech.url)

speech.list$fulltext=NA
for(i in seq(nrow(speech.list))) {
  text <- read_html(speech.list$urls[i]) %>% # load the page
    html_nodes(".displaytext") %>% # isloate the text
    html_text() # get the text
  speech.list$fulltext[i]=text
  # Create the file name
  filename <- paste0("/Users/linhan/Desktop/semester 3/ADS/project1/Fall2017-Project1
-RNotebook/data/fulltext/",
                     speech.list$File[i], "-",
                     speech.list$Term[i], ".txt")
  sink(file = filename) %>% # open file to write
  cat(text)  # write the file
  sink() # close the file
}

sentence.list=NULL
for(i in 1:nrow(speech.list)){
  sentences=sent_detect(speech.list$fulltext[i],
                        endmarks = c("?", ".", "!", "|",";"))
  if(length(sentences)>0){
    emotions=get_nrc_sentiment(sentences)
    word.count=word_count(sentences)
    # colnames(emotions)=paste0("emo.", colnames(emotions))
    # in case the word counts are zeros?
    emotions=diag(1/(word.count+0.01))%*%as.matrix(emotions)
    sentence.list=rbind(sentence.list,
                        cbind(speech.list[i,-ncol(speech.list)],
                              sentences=as.character(sentences),
                              word.count,
                              emotions,
                              sent.id=1:length(sentences)
                              )
```

```
      )
    }
  }
sentence.list=
    sentence.list%>%
    filter(!is.na(word.count))
sel.comparison=c("DonaldJTrump","JohnMcCain", "GeorgeBush", "MittRomney", "GeorgeWBus
h",
                "RonaldReagan","AlbertGore,Jr", "HillaryClinton","JohnFKerry",
                "WilliamJClinton","HarrySTruman", "BarackObama", "LyndonBJohnson",
                "GeraldRFord", "JimmyCarter", "DwightDEisenhower", "FranklinDRooseve
lt",
                "HerbertHoover","JohnFKennedy","RichardNixon","WoodrowWilson",
                "AbrahamLincoln", "TheodoreRoosevelt", "JamesGarfield",
                "JohnQuincyAdams", "UlyssesSGrant", "ThomasJefferson",
                "GeorgeWashington", "WilliamHowardTaft", "AndrewJackson",
                "WilliamHenryHarrison", "JohnAdams"
      )
```

Next, we want to compare the sentence length of all presidents to find some pattern.

```
sentence.list.sel=filter(sentence.list,
                        File%in%sel.comparison)
sentence.list.sel$File=factor(sentence.list.sel$File)

sentence.list.sel$FileOrdered=reorder(sentence.list.sel$File,
                                      sentence.list.sel$word.count,
                                      mean,
                                      order=T)

beeswarm(word.count~FileOrdered,
        method="square",
        data=sentence.list.sel,
        horizontal = TRUE,
        pch=11, col=alpha(brewer.pal(9, "Set1"), 0.6),
        cex=0.5, cex.axis=0.4, cex.lab=0.5,
        spacing=5/nlevels(sentence.list.sel$FileOrdered),
        las=2, xlab="Number of words in a sentence.", ylab="",
        main="inaugspeech sentence length")
```
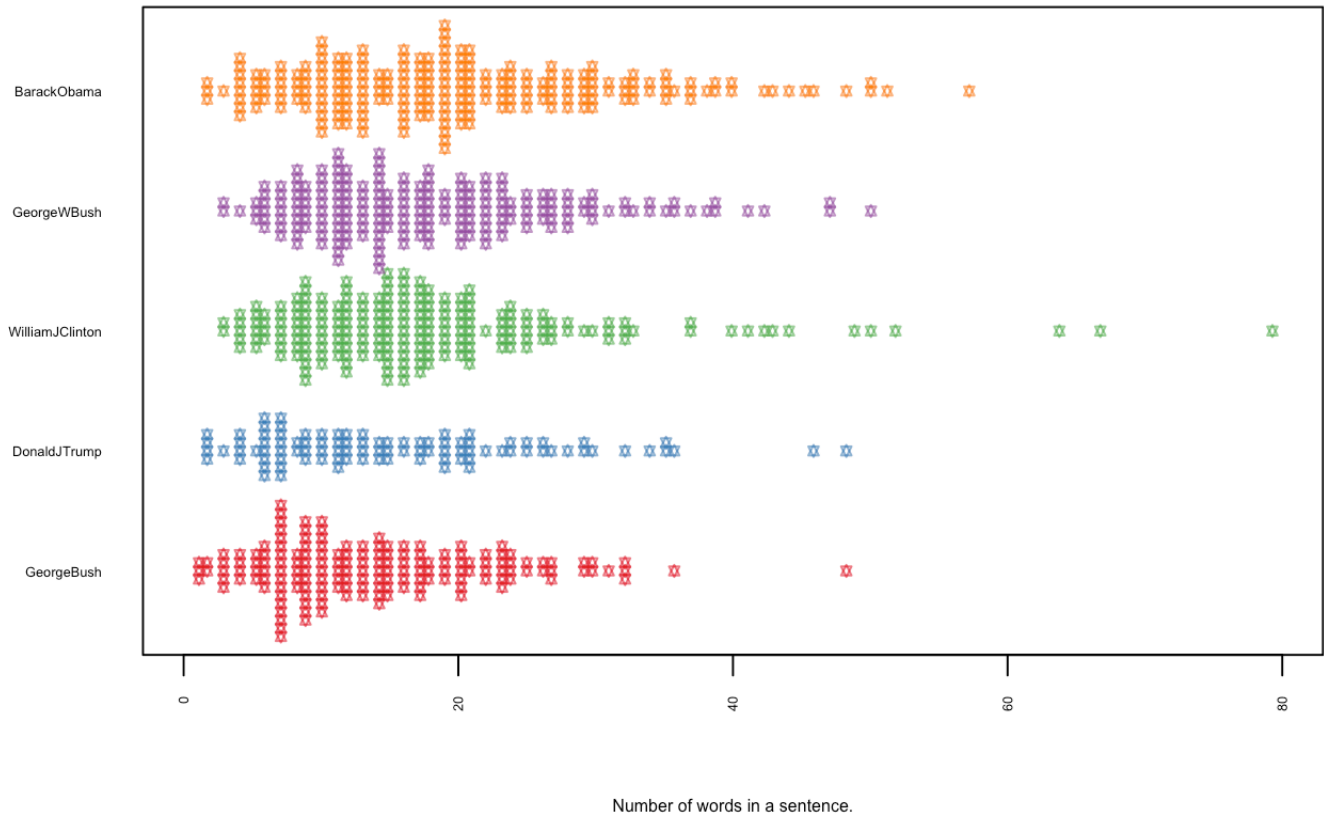
inaugspeech sentence length

Number of words in a sentence.

We can see from the senetence length distribution graph, almost all presidents after the cold war had shorter sentences compared to other presidents from other time period. This means presidents tend to use short sentences to express themselves over time. The reason is really clear. With short sentences, the listener gets the whole thing in one go. They can easily to remember, easily to understand. Compared to longer sentences, short sentences are more powerful communication.

Now, we can compare those presidents' speeches after cold war to find some similarities and differences between Trump and other presidents.

```
sel.comparison1=c("DonaldJTrump","GeorgeBush","GeorgeWBush",
                  "WilliamJClinton" ,"BarackObama")
sentence.list.sel=filter(sentence.list,
                         File%in%sel.comparison1)
sentence.list.sel$File=factor(sentence.list.sel$File)

sentence.list.sel$FileOrdered=reorder(sentence.list.sel$File,
                                      sentence.list.sel$word.count,
                                      mean,
                                      order=T)

beeswarm(word.count~FileOrdered,
         method="square",
         data=sentence.list.sel,
         horizontal = TRUE,
         pch=11, col=alpha(brewer.pal(9, "Set1"), 0.6),
         cex=0.5, cex.axis=0.4, cex.lab=0.5,
         spacing=5/nlevels(sentence.list.sel$FileOrdered),
         las=2, xlab="Number of words in a sentence.", ylab="",
         main="inaugspeech sentence length of presidents after cold war")
```

inaugspeech sentence length of presidents after cold war

Number of words in a sentence.

When we only focus on presidents after the cold war, we can see that among all 5 presidents, Trump and George Bush all used shorter sentences.It means Trump has very simple and direct speaking styles which may somehow explain his popularity. The reason is that , whether used well or poorly, the language of a typical modern politician has a distinctive sound to it. It sounds complex and careful and may sometimes disingenuous.

Step 3: Sentiment analysis

Emotion Comparison

Since we have studied the main words and sentences length, let's focus on emotions of those speeches.

```
sel.comparison2=c("GeorgeBush","GeorgeWBush",
                  "WilliamJClinton" ,"BarackObama")
sentence.list1=filter(sentence.list,
                      File%in%sel.comparison2)
heatmap.2(cor(sentence.list1%>%select(anger:trust)),
          scale = "none",
          col = bluered(100), , margin=c(6, 6), key=F,
          trace = "none", density.info = "none",main="other presidents' heatmap")
```
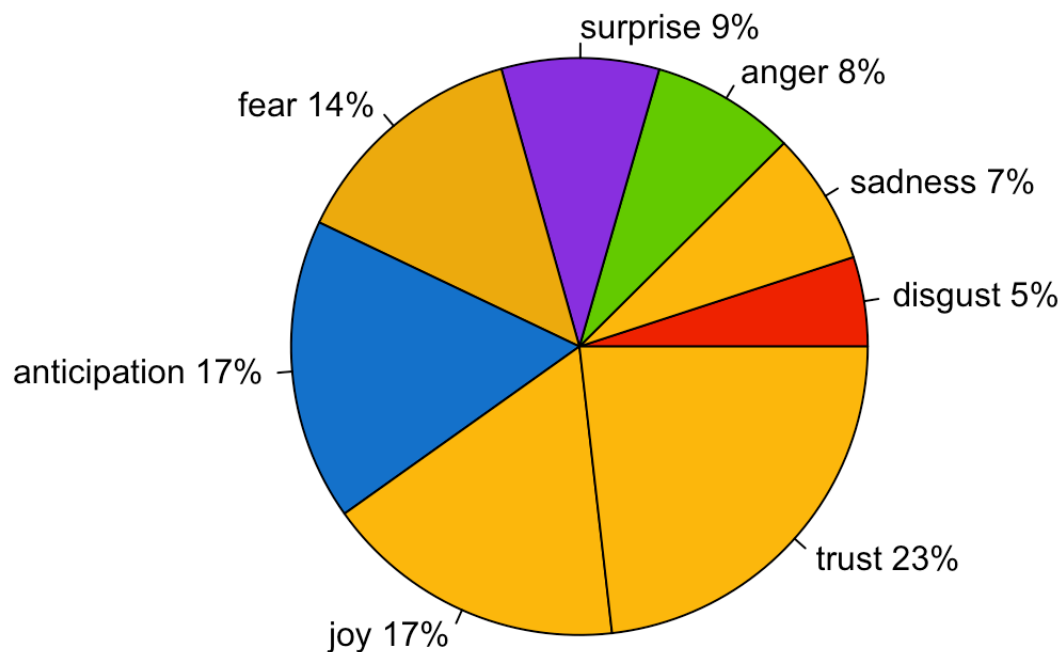
other presidents' heatmap
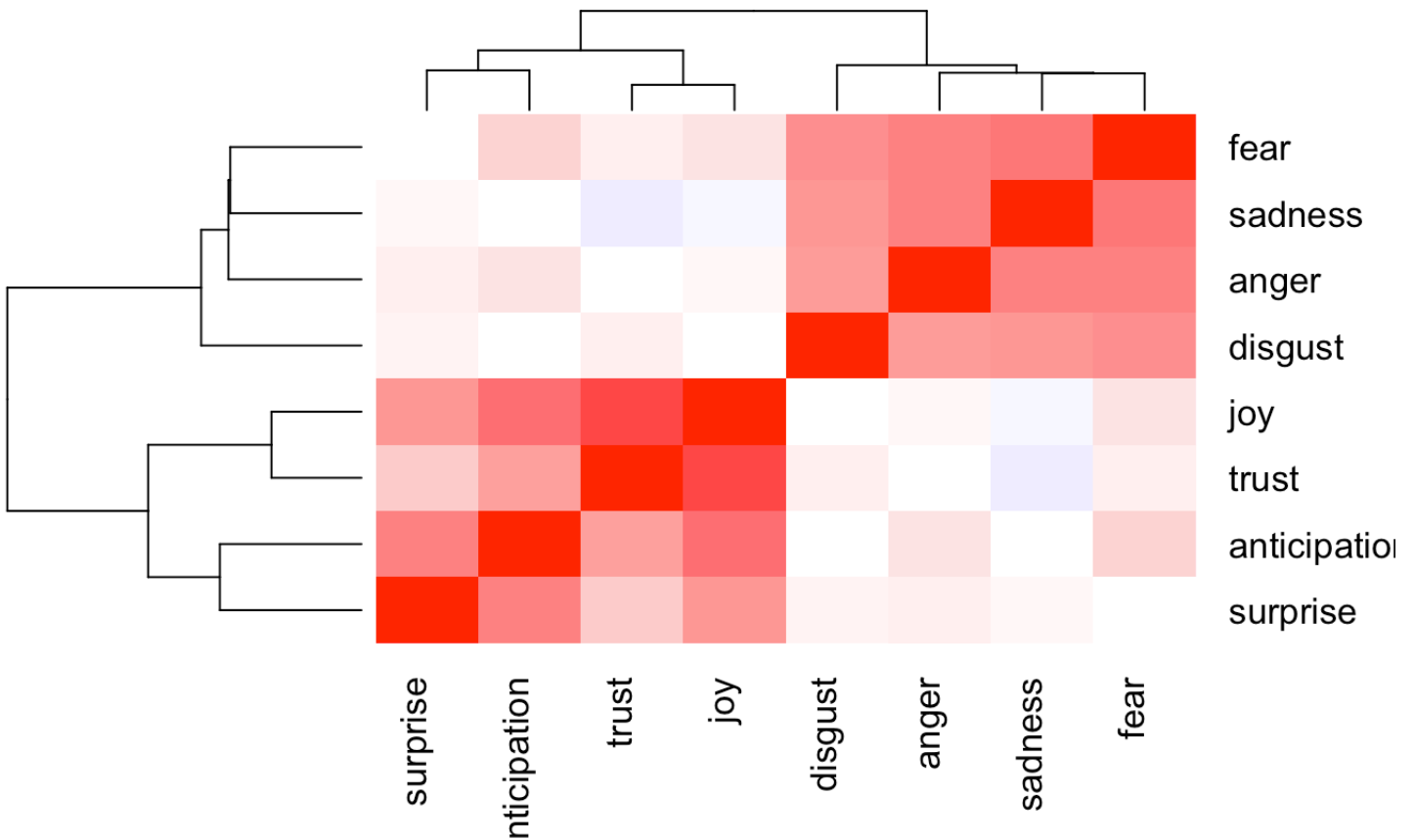
```
par(mar=c(4, 6, 2, 1))
emo.means=colMeans(select(sentence.list1, anger:trust)>0.01)
ord = order(emo.means)
emo.means= emo.means[ord]
col.use=c("red2", "darkgoldenrod1",
            "chartreuse3", "blueviolet",
            "darkgoldenrod2", "dodgerblue3",
            "darkgoldenrod1", "darkgoldenrod1")
pct <- round(emo.means/sum(emo.means)*100)
lbls<- c(names(emo.means))
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(emo.means, labels= lbls, col=col.use[order(emo.means)],  main= "other presidents'
emotion")
```

other presidents' emotion
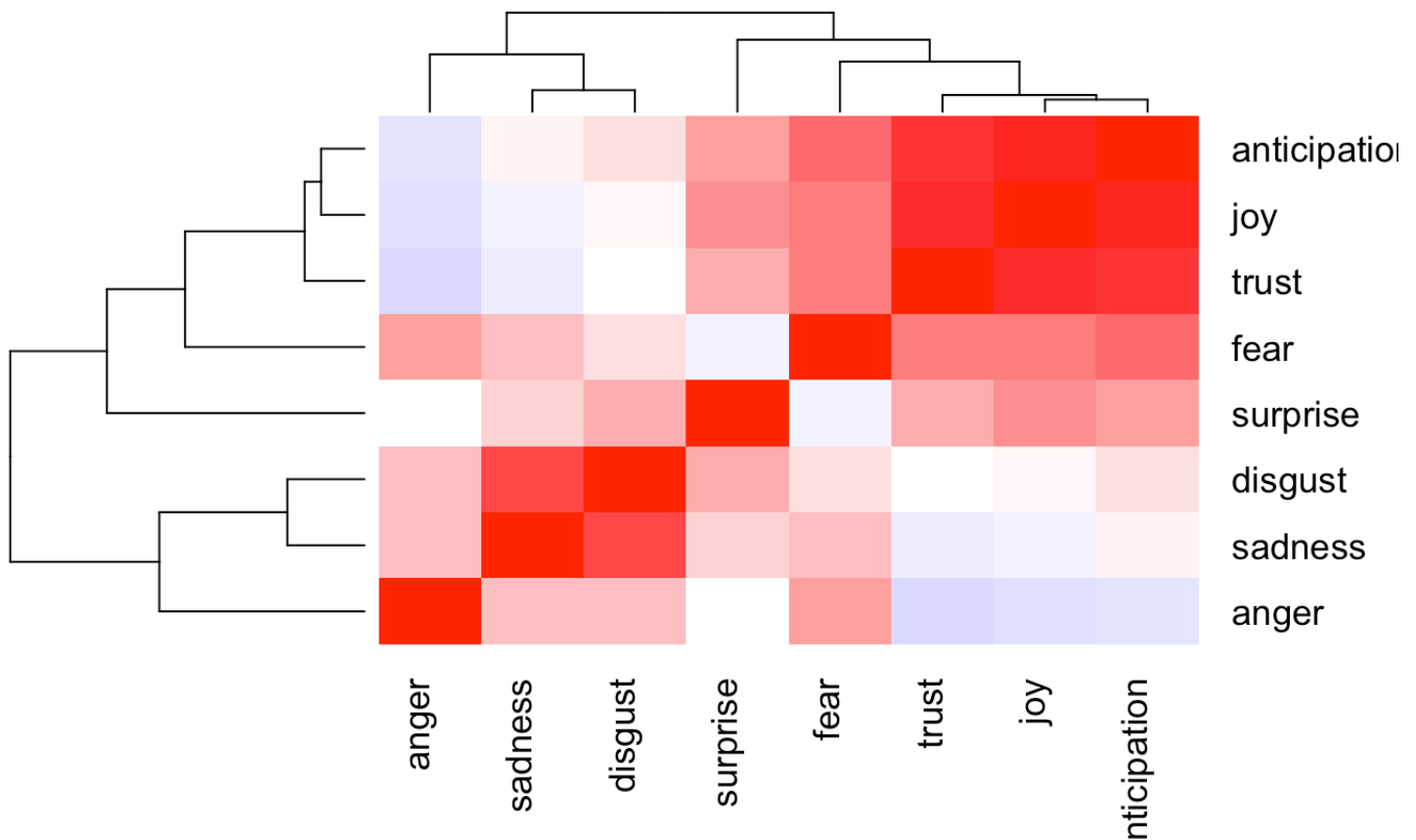
```
sel.comparison3=c("DonaldJTrump")
sentence.list2=filter(sentence.list,
                      File%in%sel.comparison3)
heatmap.2(cor(sentence.list2%>%select(anger:trust)),
          scale = "none",
          col = bluered(100), , margin=c(6, 6), key=F,
          trace = "none", density.info = "none", main= "Trump heatmap")
```
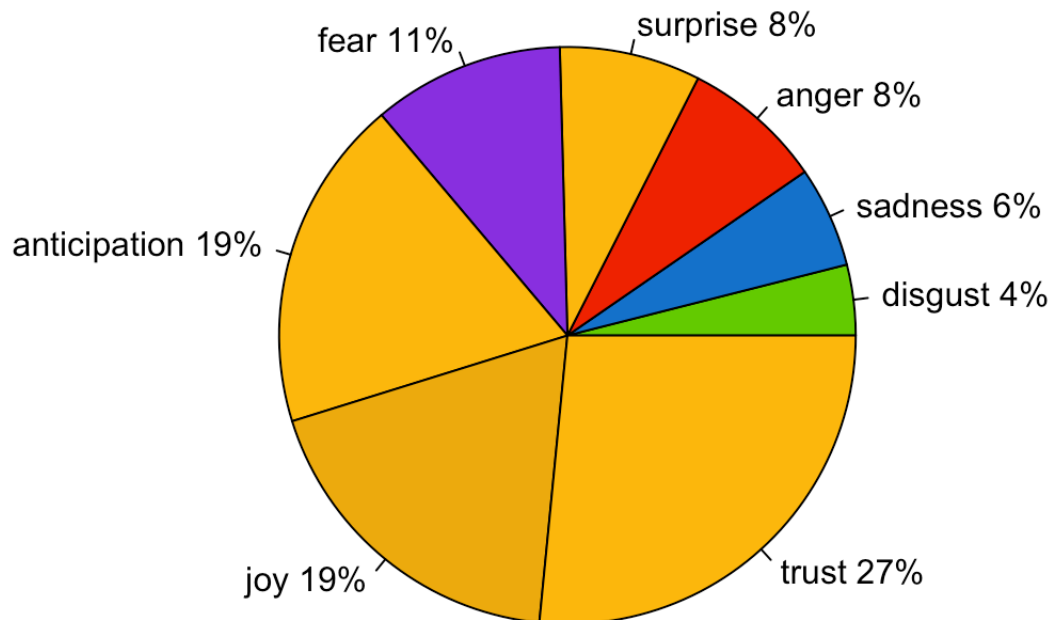
Trump heatmap

```
par(mar=c(4, 6, 2, 1))
emo.means1=colMeans(select(sentence.list2, anger:trust)>0.01)
col.use=c("red2", "darkgoldenrod1",
          "chartreuse3", "blueviolet",
          "darkgoldenrod2", "dodgerblue3",
          "darkgoldenrod1", "darkgoldenrod1")
ord1 = order(emo.means1)
emo.means1= emo.means1[ord1]
pct1 <- round(emo.means1/sum(emo.means1)*100)
lbls1<- c(names(emo.means1))
lbls1 <- paste(lbls1, pct1) # add percents to labels
lbls1 <- paste(lbls1,"%",sep="") # ad % to labels
pie(emo.means1, las=2, labels= lbls1, col=col.use[ord1], main="Trump's piechart")
```

Trump's piechart

We can see from the emotion heatmap of Trump and other presidents, anticipation, joy and trust had strong connections for all of their speeches. Anger, sadness, disgust were always connected, too. However, we can see some differences between Trump and other presidents. Trump had strong emotion of anticipation,joy and trust than other presidents. These emotions were also connected to fear. This is quite different from other presidents. For other presidents, fear was consisted with sadness, anger and disgust which are known as

"negetive" emotions. As for "positive" emotions, other presidents combined surprise, anticipation, trust and joy together, but for Trump, 'surprise' doesn't seem so important for his 'positive' emotions. In conclusion, from heatmap, Trump had strong 'positive' emotions than other presidents, but these emotions are also connected with fear, this means Trump had concerns while he had anticipation.

Next, from emotion pie chart, we can get that the sum percentage of those "negetive" emotions like disgust, sadness,anger, fear are almost the same between trump and other presidents, but Trump have relatively smaller percentage of "negetive"" emotions. As for the percentage of those "positive" emotions, we can see that for all emotions– trust, joy and anticipation, Trump have a higher percentage. So we can know that Trump have stronger 'positive' emotions and smaller percentage of 'negetive' emotions than other presidents. Then we can get the conclusion that Trump have much confidence in his personality and he is more likely to be emotional than other presidents.

Step 4: Topic Modeling

Finally, we will investigate about the topic differences between Trump and other presidents. #Text processing

```
sel.comparison1=c("DonaldJTrump","GeorgeBush","GeorgeWBush",
                  "WilliamJClinton" ,"BarackObama")
sentence.list=filter(sentence.list,
                       File%in%sel.comparison1)
corpus.list=sentence.list[2:(nrow(sentence.list)-1), ]
sentence.pre=sentence.list$sentences[1:(nrow(sentence.list)-2)]
sentence.post=sentence.list$sentences[3:(nrow(sentence.list)-1)]
corpus.list$snipets=paste(sentence.pre, corpus.list$sentences, sentence.post, sep=" "
)
rm.rows=(1:nrow(corpus.list))[corpus.list$sent.id==1]
rm.rows=c(rm.rows, rm.rows-1)
corpus.list=corpus.list[-rm.rows, ]
```

```
docs <- Corpus(VectorSource(corpus.list$snipets))
docs <-tm_map(docs,content_transformer(tolower))
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, stripWhitespace)
docs <- tm_map(docs,stemDocument)
```

```
dtm <- DocumentTermMatrix(docs)
#convert rownames to filenames#convert rownames to filenames
rownames(dtm) <- paste(corpus.list$File,
                       corpus.list$Term, corpus.list$sent.id, sep="_")

rowTotals <- apply(dtm , 1, sum) #Find the sum of words in each Document

dtm  <- dtm[rowTotals> 0, ]
corpus.list=corpus.list[rowTotals>0, ]
```

```
#Set parameters for Gibbs sampling
burnin <- 4000
iter <- 2000
thin <- 500
seed <-list(2003,5,63,100001,765)
nstart <- 5
best <- TRUE

#Number of topics
k <- 15

#Run LDA using Gibbs sampling
ldaOut <-LDA(dtm, k, method="Gibbs", control=list(nstart=nstart,
                                          seed = seed, best=best,
                                          burnin = burnin, iter = iter,
                                          thin=thin))
#write out results
#docs to topics
ldaOut.topics <- as.matrix(topics(ldaOut))
table(c(1:k, ldaOut.topics))
```

```
##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
## 64 68 61 69 70 53 79 68 73 49 51 53 45 59 44
```

```
write.csv(ldaOut.topics,file=paste("/Users/linhan/Desktop/semester 3/ADS/project1/Fal
l2017-Project1-RNotebook/output/LDAGibbs",k,"DocsToTopics.csv"))

#top 6 terms in each topic
ldaOut.terms <- as.matrix(terms(ldaOut,20))
write.csv(ldaOut.terms,file=paste("/Users/linhan/Desktop/semester 3/ADS/project1/Fall
2017-Project1-RNotebook/output/LDAGibbs",k,"TopicsToTerms.csv"))

#probabilities associated with each topic assignment
topicProbabilities <- as.data.frame(ldaOut@gamma)
write.csv(topicProbabilities,file=paste("/Users/linhan/Desktop/semester 3/ADS/project
1/Fall2017-Project1-RNotebook/output/LDAGibbs",k,"TopicProbabilities.csv"))
```

```r
terms.beta=ldaOut@beta
terms.beta=scale(terms.beta)
topics.terms=NULL
for(i in 1:k){
  topics.terms=rbind(topics.terms, ldaOut@terms[order(terms.beta[i,], decreasing = TR
UE)[1:7]])
}
#topics.terms
#ldaOut.terms
```

```r
topics.hash=c("Economy", "America", "Defense", "Belief", "Election", "Patriotism", "U
nity", "Government", "Reform", "Temporal", "WorkingFamilies", "Freedom", "Equality",
"Misc", "Legislation")
corpus.list$ldatopic=as.vector(ldaOut.topics)
corpus.list$ldahash=topics.hash[ldaOut.topics]

colnames(topicProbabilities)=topics.hash
corpus.list.df=cbind(corpus.list, topicProbabilities)
```

```r
par(mar=c(1,1,1,1))
topic.summary=tbl_df(corpus.list.df)%>%
              select(File, Economy:Legislation)%>%
              group_by(File)%>%
              summarise_each(funs(mean))
topic.summary=as.data.frame(topic.summary)
rownames(topic.summary)=topic.summary[,1]
topic.plot=c(1, 2, 3,6,7,8,11,12,13,15)
print(topics.hash[topic.plot])
```
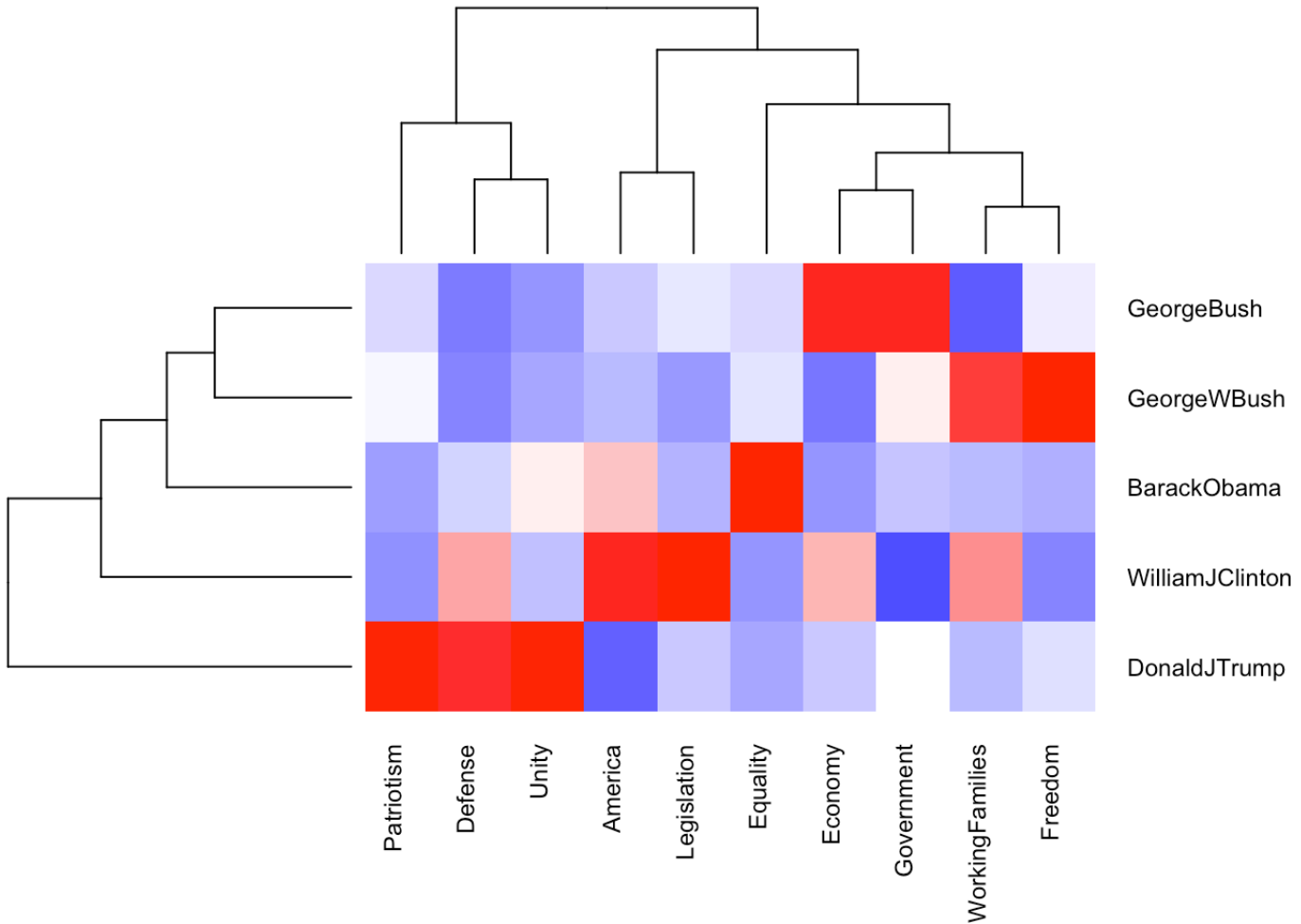
```
##  [1] "Economy"         "America"         "Defense"
##  [4] "Patriotism"      "Unity"           "Government"
##  [7] "WorkingFamilies" "Freedom"         "Equality"
## [10] "Legislation"
```

```r
heatmap.2(as.matrix(topic.summary[,topic.plot+1]),
          scale = "column", key=F,
          col = bluered(100),
          cexRow = 0.9, cexCol = 0.9, margins = c(8, 8),
          trace = "none", density.info = "none")
```

From topic heatmap, we can easily find some big differences between Trump and other presidents.

Compared to other presidents, Trump focused more on topics about Patriotism, defense and unity. This shows Trump's personal tendency. Actually, Trump always delivers speeches heavy with patriotic themes. This kind of personal tendency corresponds to his immigration policies and the old slogan he mentioned–"America first". We can also see that 'America' is also a really important topic of his speech. However, topics like Legislation, government, freedom become less important in his speech compared to other presidents. These topic tendency can somewhat explain why Trump is so controversial compared to other presidents, especially many extreme patriotic activities appeared after Trump became the president and many people considered Trump as one of incentives. what's more, Trump is sometimes criticized for underestimating the importance of law and we can see the reason from his speech topic focus.

Step 5: Final Conclusion

By main word analysis(Wordcloud), sentence length analysis, emotion analysis and topic anlysis, we can get some final conclusions.

First of all, Trump is not so different from other presidents. They share some similarities.Their speeches are all fcocusd on people's rights and America's future development. Their speeches mostly expressed "positive'"emotions like trust, joy, anticipation.

Secondly, there are obviuos differences between Trump and other presidents after cold war. Trump was more concerned about jobs, wealth and other presidents were more likely to focus on freedom and liberty. However, these tendencies are highly influenced by the background of America, so we cannot say "Trump is an alien."

based on this.

Thirdly, Trump is apparently more emotional than other presidents. He had strong emotion expression and really liked using shorter sentence in his speech. This shows his personality– straightforward and emotional which can soemhow explain his personal charm.

Finally, from the topic focus, we can see big differences between Trump and other presidents. Trump is obviously more focused on Patriotism, defense and unity which correspond with his immigration policies and the mentioned slogan 'America first'.

In conclusion, Trump and other presidents have some similarities but in many ways, he is very different from other presidents, not only in political point of view but also in personal charaters. We can see why he is so controversial based on our analysis. We can also undertand this from his Twitter!

Refernce: http://www.marketwatch.com/story/how-trump-stacks-up-to-reagan-jfk-and-other-presidential-greats-2017-05-30 (http://www.marketwatch.com/story/how-trump-stacks-up-to-reagan-jfk-and-other-presidential-greats-2017-05-30) http://us-presidents.insidegov.com/compare/7-13-134/Richard-Nixon-vs-Ronald-Reagan-vs-Donald-Trump (http://us-presidents.insidegov.com/compare/7-13-134/Richard-Nixon-vs-Ronald-Reagan-vs-Donald-Trump)