What is the trend of U.S. presidents' inaugrual speeches?

Xiaoxiao Guo; xg2282

Jan 28, 2018

Data Description

The data comprises of 58 inaugrual speeches, from that of George Washington to that of Donald Trump which was delivered earlier this year.

Research Question

First of all, I must admit that I am not good at politics, so I choose to tell my story just from the data itself and do not make any comment on politics.

The main question of my research is to analyse the trend of presidents' inaugrual speeches according to the 58 inaugrual speeches we have from 1789-2017. Specifically, I want to tell a story from the following three aspects:

1. Change in length of sentences in presidents' inaugrual speeches
2. Change in speakers' emotions with presidents' inaugrual speeches
3. Change in topics of presidents' inaugrual speeches

In order to analyse these changes of presidents' inaugrual speeches in these years, I split the data into three parts according to the date of these speeches:
Period 1: 1789-1865, George Washington-Abraham Lincoln (20 speeches)
Period 2: 1866-1945, Ulysses S. Grant-Franklin D. Roosevelt (20 speeches)
Period 3: 1946-2017, Harry S. Truman-Donald J. Trump (18 speeches)

In this report, I will focus my study on changes among speeches of these three periods and tell my own story.

Part I: Change in Length of Sentences

First, I look at the change in length of sentences of presidents' inaugrual speeches.

Data Collection and Processing

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

Step 1: Read in the speeches

Step 2: Generate list of sentences
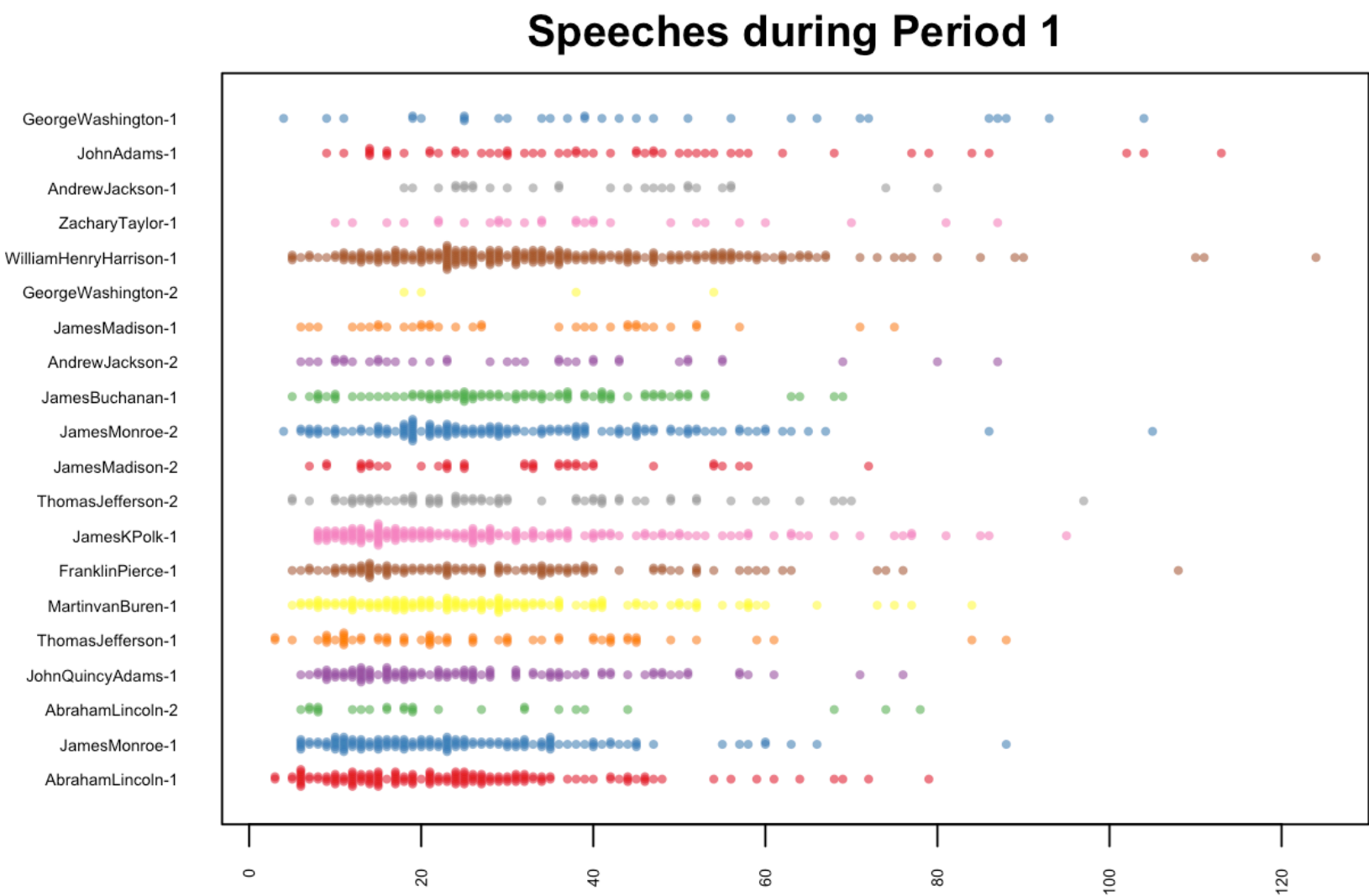
Data Analysis — Length of Sentences

Overview of sentence length distribution by different time periods of speeches
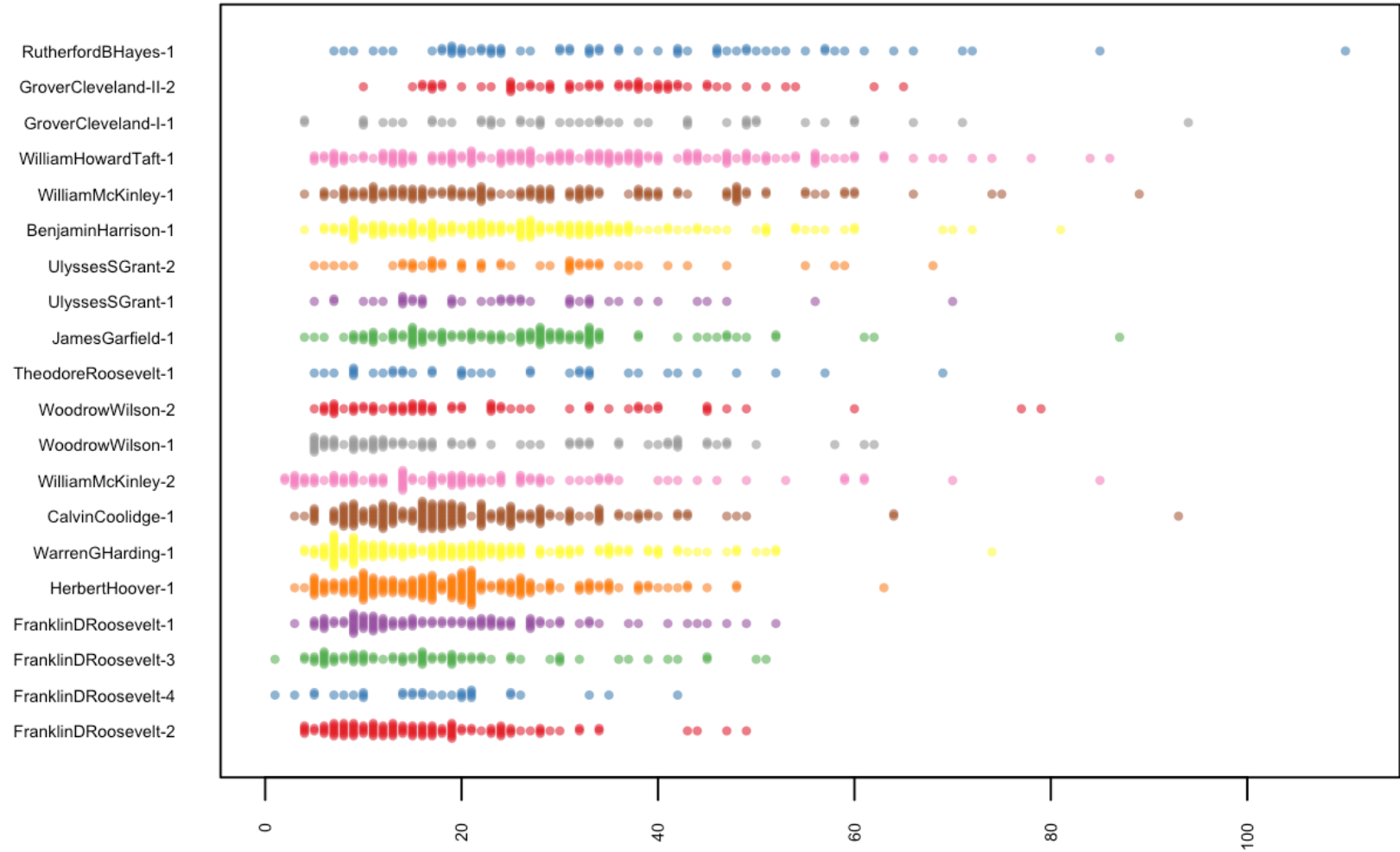
Period 1 (1789-1865)

Speeches during Period 1

What are these short sentences?

sentences
<fctr>
I trust not.
freedom of religion;
I think not.
Unanimity is impossible.
4 rows

Period 2 (1866-1945)

Speeches during Period 2
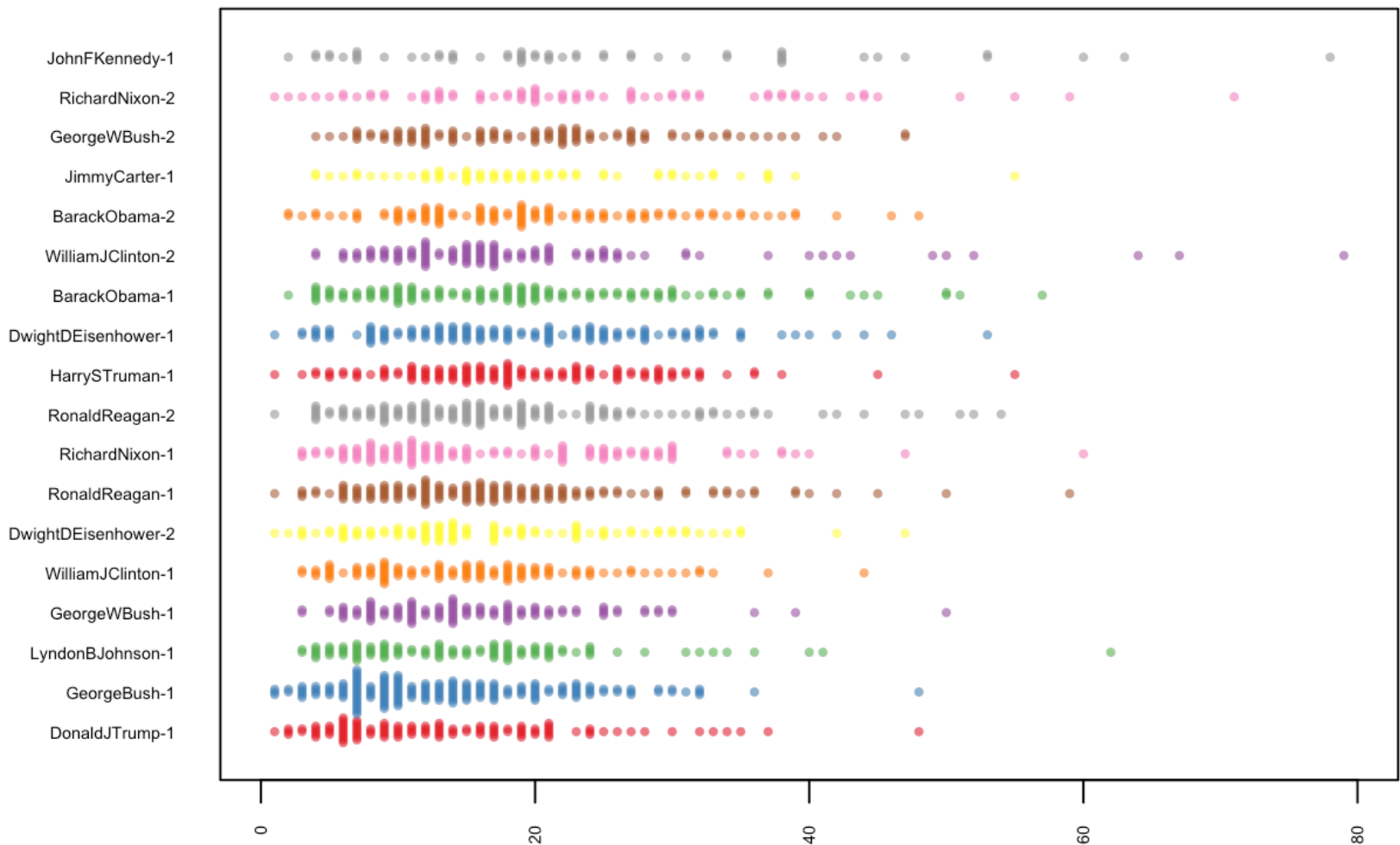

What are these short sentences?

sentences
<fctr>

None exists now.

It became inevitable;

It came.

We are reunited.

Sectionalism has disappeared.

They are inconvenient.

They triumphed.

That verdict stands.

Crime is increasing.

taxes have risen;

1-10 of 13 rows   Previous **1** 2 Next

Period 3 (1946-2017)

Speeches during Period 3

What are these short sentences?

<div style="text-align: right">Code</div>

sentences
<fctr>

Mr.

Vice President, Mr.

Amen.

Nations amass wealth.

These principles are: 1.

Mr.

Chairman, Mr.

Vice President, Mr.

Chief Justice, Mr.

Our population grows.

1-10 of 48 rows   Previous **1** 2 3 4 5 Next

The three plots above show that the length of sentences in presendents' inaugrual speeches has become shorter and shorter.

In Period 1, the largest number of words in a sentence is above 120, and some sentences has more than 80 words. Most sentences has words less than 60.

In Period 2, the largest number of words in a sentence is above 100, and some sentences has more than 60 words. Most sentences has words less than 40.

In Period 3, the largest number of words in a sentence is below 100, and some sentences has more than 40 words. Most sentences has words less than 30.

The pattern that the length of sentences in presendents' inaugrual speeches has become shorter and shorter is consistent with the change of human's expressing behavior.

Nowadays, people prefer to express themselves in an easier way to make their ideas clear rather than use long sentences with difficult structures.

Part II: Change in Emotions of Speakers

I continue my research with sentiment changes of speakers. In this part, I also use one sentence as a unit.

## Data Analysis — Length of Sentences with Sentiment Analysis

First, I want to make a further study on sentence length variation over the course of the speech, with emotions. This is a combination of sentiment change and sentence length change.

It makes sense that I use darker colors to represent more intense emotions. In addition, blues stand for good feelings (joy, trust, anticipation) and reds stand for bad feelings (anger, disgust, fear, sadness), blueviolet stands for surprise which is neutral. I also use dark to show negative attitude and lightgreen to show positive attitude.
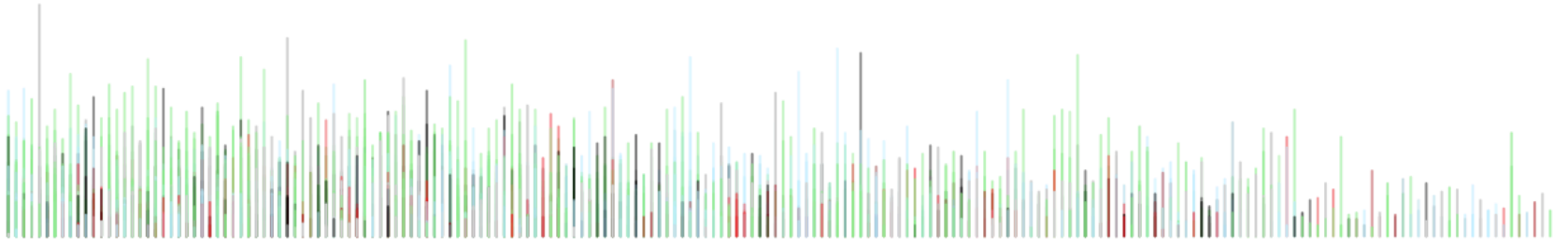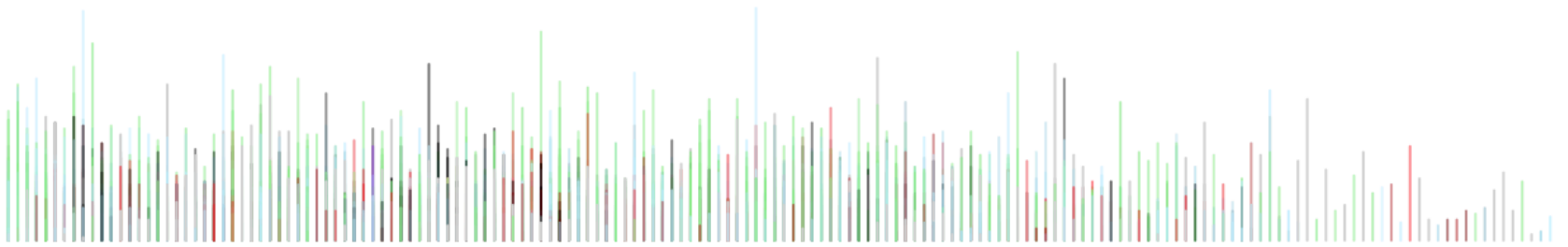
Code

Code

Period 1
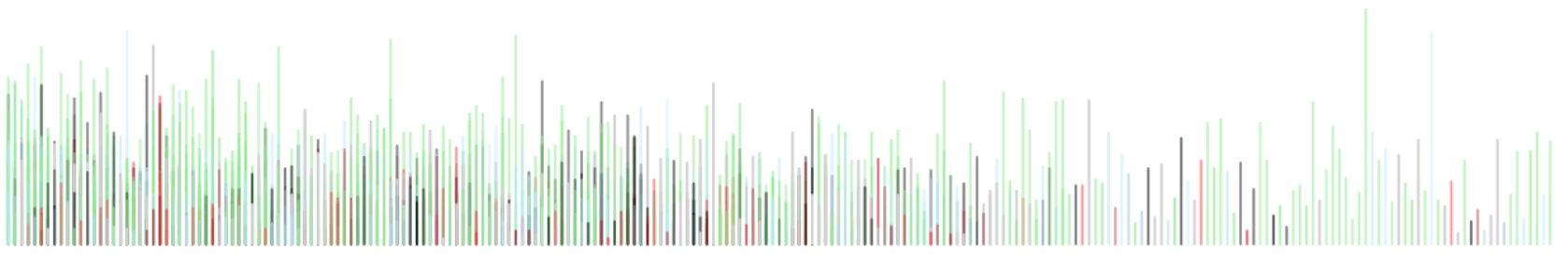
Period 2

Period 3

Now we can see that there are few reds in these plots, which is consistent with our intuition that presidents always express good emotions when giving speeches. Moreover, the colors are becoming lighter and lighter from Period 1 to Period 3, which means that speakers' emotions are becoming calmer and calmer.

This change is rational because people now can control their emotions better than before. In addition, the following presidents may gain some experience about how to express their emotions in speeches from the previous president. Therefore, they can express themselves in a less radical way.
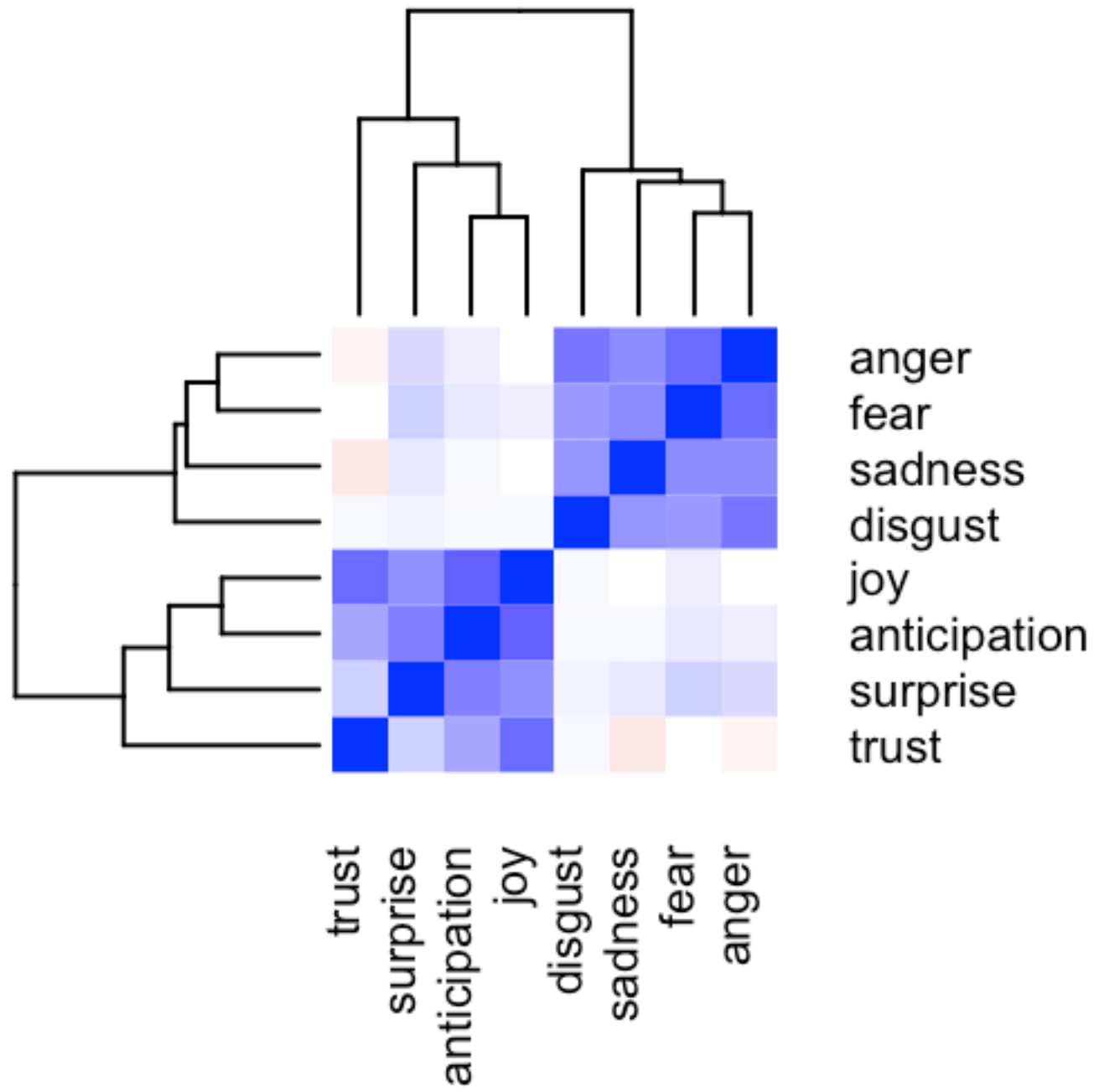
Moreover, these plots also show the pattern that the length of sentences in speeches is becoming shorter and shorter.

Data Analysis — Clustering of Emotions

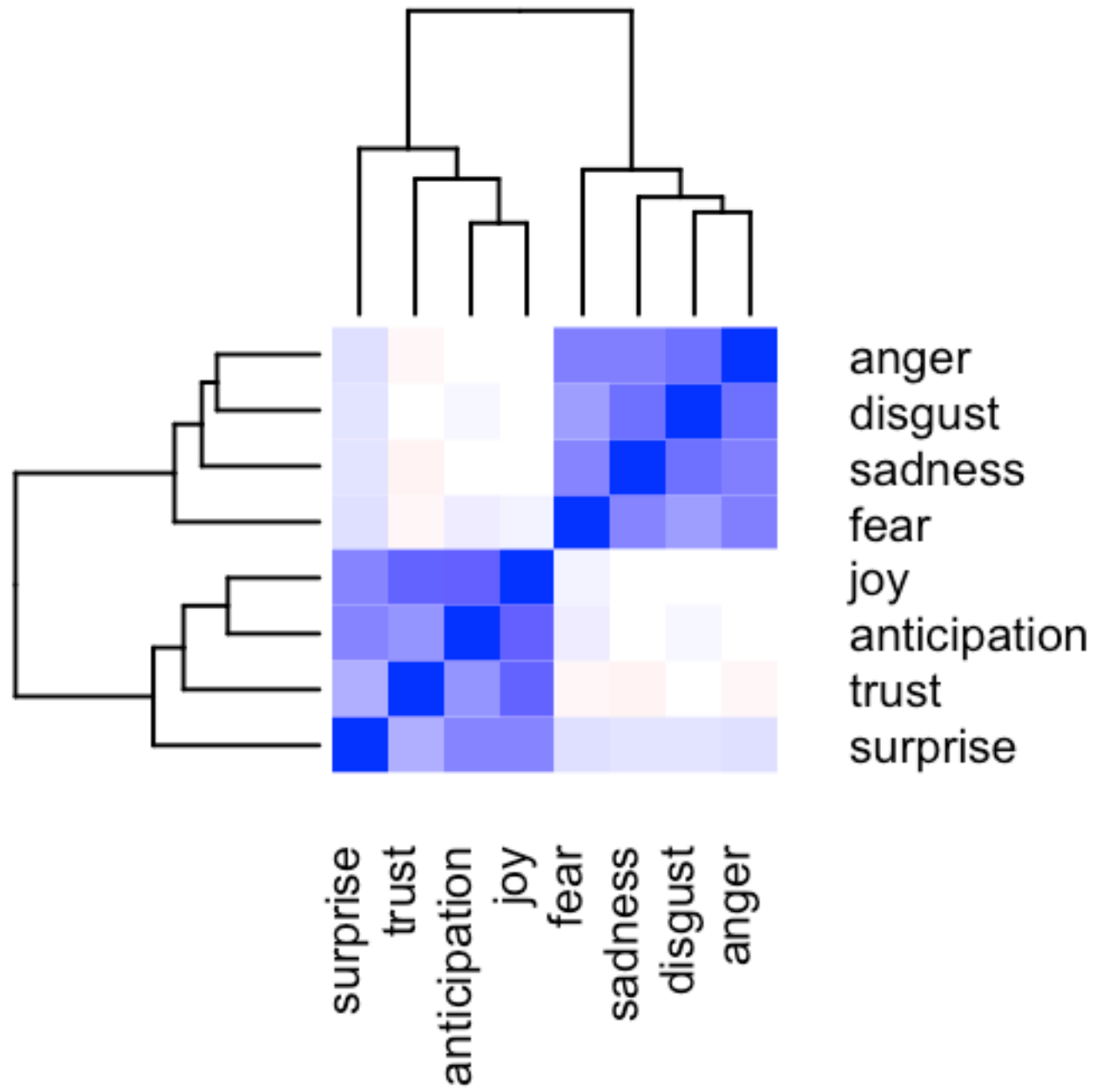## Next, I will show a more detailed study on emotion change of speakers among three different periods.

First, I give the heatmaps of speaker's emotions in these three periods.

Code

anger
fear
sadness
disgust
joy
anticipation
surprise
trust

trust
surprise
anticipation
joy
disgust
sadness
fear
anger

Code

anger
disgust
sadness
fear
joy
anticipation
trust
surprise
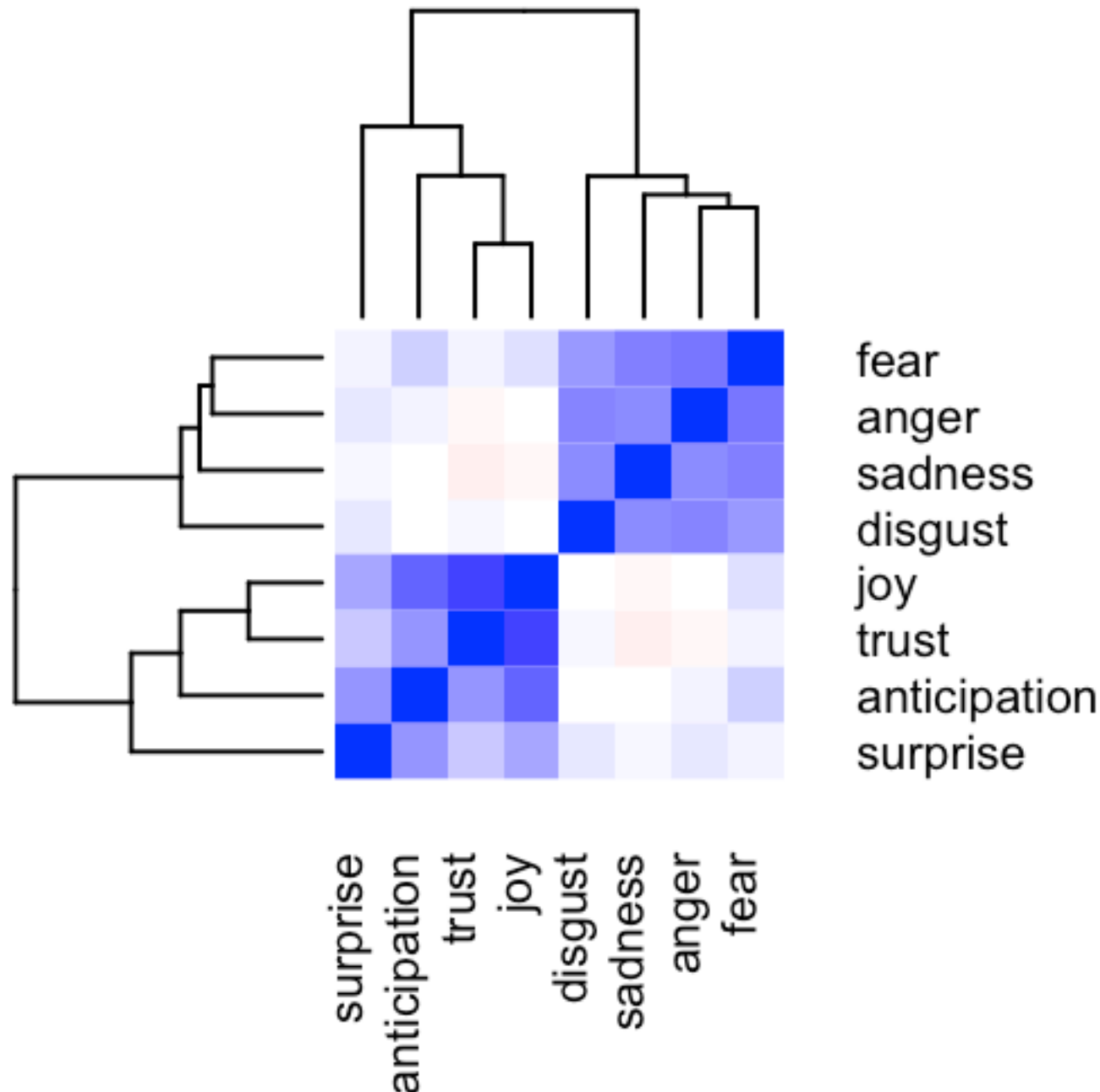
surprise
trust
anticipation
joy
fear
sadness
disgust
anger

Code

In my opinion, the extend of darkness represent the degree of correlation between two emotions. Therefore, I think the correlation between positive emotions bocome higher and the correlation between negative emotions bocome lower than before, which means **presidents now express more positive emotions than before.**

The following plots give us a straight view of the distributions of emotions in these three periods.

Code

Code
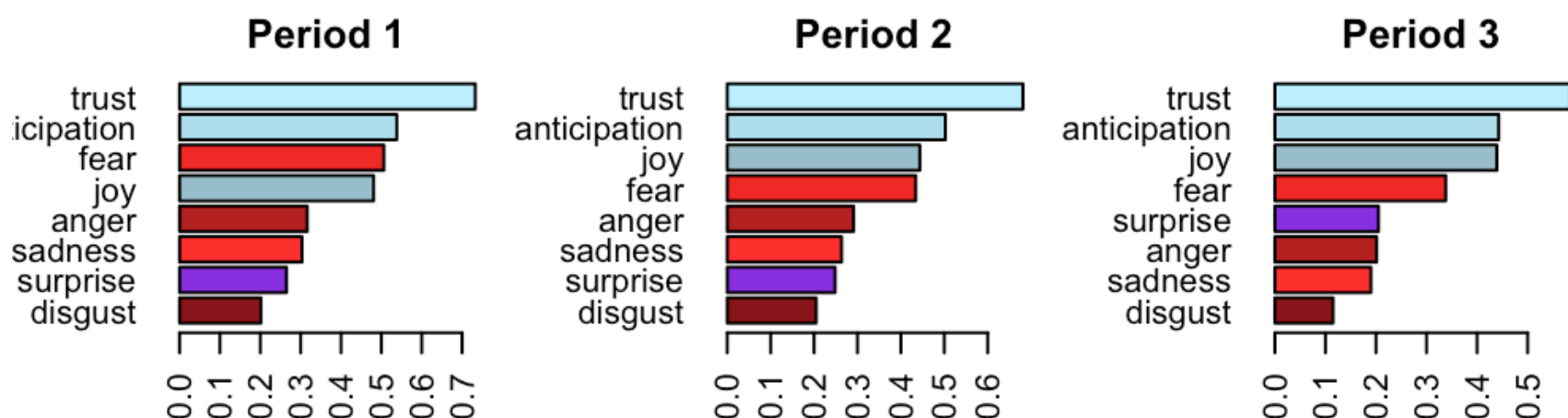
These plots give us a better view. It is clear that good feelings (blues) rank higher than bad feelings (reds) in general. And in Period 1, fear is in the third place of the emotion expression. However, in Period 2, joy become the third place of the emotion expression. In addition, the negative emotions like disgust, sadness and anger have a less weight than before.

Thus, I will say that presidents today express more positive emotions than before and this change is also reasonable.

The following plot illustrates my view clearly.

Code
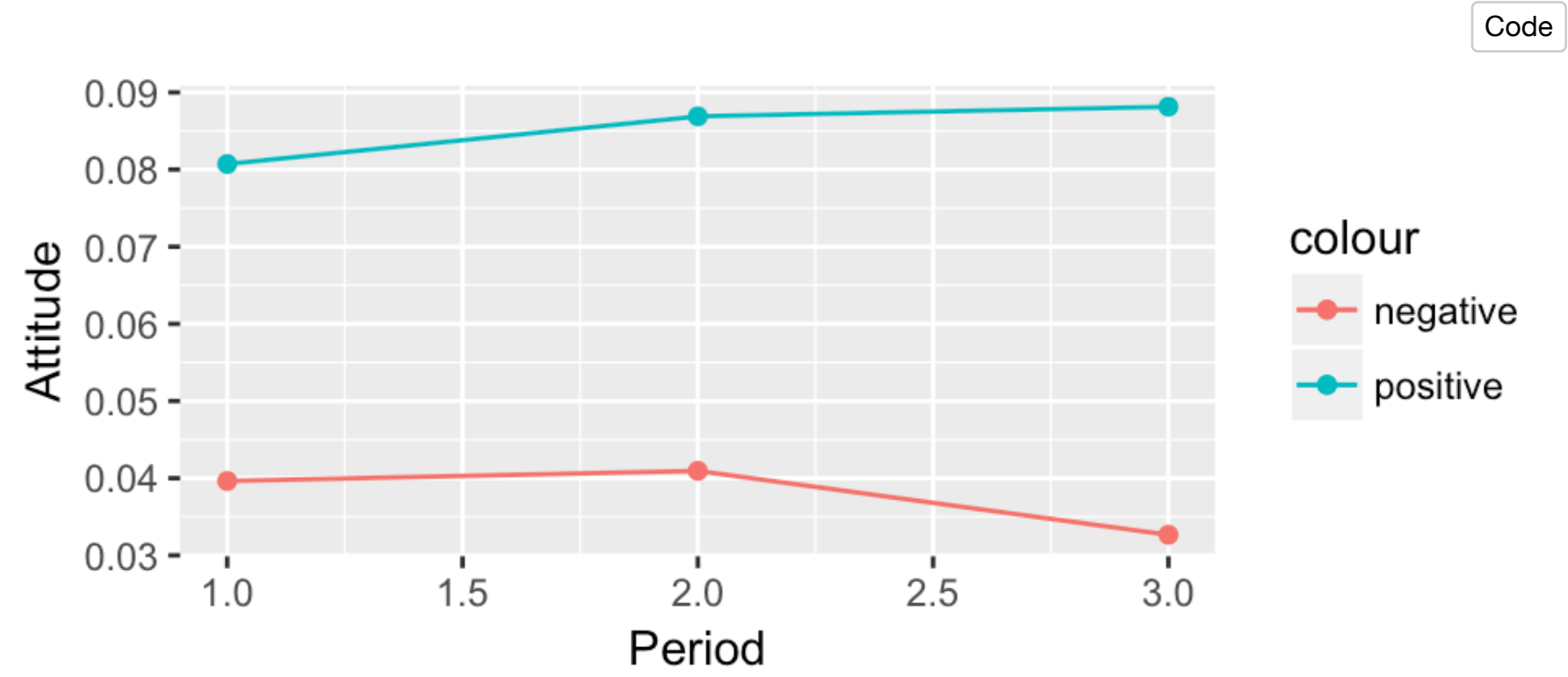
From this plot we can see that presidents always express more positive attitude than negative. Meanwhile, there is a trend that they express more positive and less negative attitude than before.

Part III: Change in Topics of Presidents' Inaugrual Speeches

Data Collection and Processing

For topic modeling, I still use one sentence as a unit. I prepare a corpus of sentence snippets as follows.

Code

Data Analysis — Topic Modeling

Step 1: Gengerate document-term matrices.

Code

Step 2: Run LDA with 10 topics

The following table shows the number of sentences in each topic.

Code

```
   1    2    3    4    5    6    7    8    9   10
 508  580  504  593  727  464  835  503  350  488
```

[Code]

I also give the top 20 most popular words in each topic.

[Code]

```
         Topic 1          Topic 2      Topic 3      Topic 4       Topic 5      Topic 6       Top
ic 7
 [1,] "countri"        "nation"     "govern"     "public"      "time"       "shall"      "wo
rld"
 [2,] "great"          "war"        "state"      "use"         "year"       "peopl"      "ne
w"
 [3,] "interest"       "peac"       "power"      "industri"    "now"        "may"        "am
erica"
 [4,] "everi"          "great"      "right"      "mean"        "man"        "duti"       "am
erican"
 [5,] "mani"           "unit"       "constitut"  "interest"    "come"       "upon"       "wo
rk"
 [6,] "much"           "polici"     "union"      "secur"       "day"        "faith"      "le
t"
 [7,] "present"        "forc"       "preserv"    "revenu"      "first"      "purpos"     "li
ve"
 [8,] "danger"         "foreign"    "within"     "product"     "histori"    "offic"      "na
tion"
 [9,] "spirit"         "made"       "general"    "condit"      "futur"      "best"       "pe
ac"
[10,] "caus"           "secur"      "unit"       "busi"        "god"        "support"    "se
ek"
[11,] "patriot"        "maintain"   "protect"    "may"         "presid"     "confid"     "ho
pe"
[12,] "experi"         "less"       "principl"   "demand"      "stand"      "public"     "wa
y"
[13,] "success"        "establish"  "author"     "need"        "today"      "servic"     "he
lp"
[14,] "continu"        "relat"      "form"       "well"        "generat"    "will"       "to
geth"
[15,] "happi"          "countri"    "equal"      "econom"      "heart"      "respons"    "ol
d"
[16,] "mind"           "progress"   "feder"      "encourag"    "still"      "trust"      "la
nd"
[17,] "charact"        "intern"     "object"     "economi"     "last"       "honor"      "se
e"
[18,] "fellowcitizen"  "extend"     "whole"      "necessari"   "american"   "take"       "ma
ke"
[19,] "result"         "state"      "citizen"    "labor"       "face"       "execut"     "pr
omis"
[20,] "exist"          "defens"     "limit"      "trade"       "past"       "oblig"      "st
rength"
         Topic 8      Topic 9      Topic 10
 [1,] "law"         "will"        "peopl"
```

```
 [2,] "upon"      "can"         "freedom"
 [3,] "congress"  "must"        "free"
 [4,] "parti"     "one"         "life"
 [5,] "question"  "nation"      "liberti"
 [6,] "polit"     "everi"       "human"
 [7,] "effect"    "good"        "men"
 [8,] "subject"   "make"        "justic"
 [9,] "opinion"   "citizen"     "believ"
[10,] "execut"    "never"       "thing"
[11,] "legisl"    "like"        "civil"
[12,] "administr" "other"       "long"
[13,] "elect"     "need"        "republ"
[14,] "control"   "anoth"       "know"
[15,] "reason"    "communiti"   "order"
[16,] "given"     "race"        "democraci"
[17,] "enforc"    "done"        "common"
[18,] "certain"   "noth"        "spirit"
[19,] "action"    "whether"     "moral"
[20,] "made"      "among"       "fear"
```

Based on the most popular terms, I assign a hashtag to each topic. This part is subjective because I set up these topics manually.

Code

Frequencies of topics in different periods.

The following pie plots show the distributions of the topics of speeches in these three periods.

Code

Code

Code
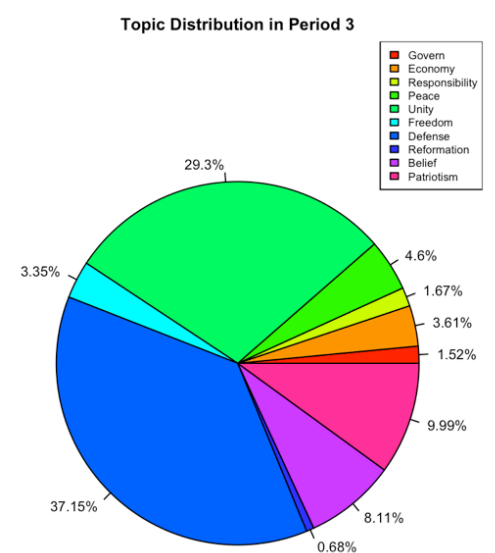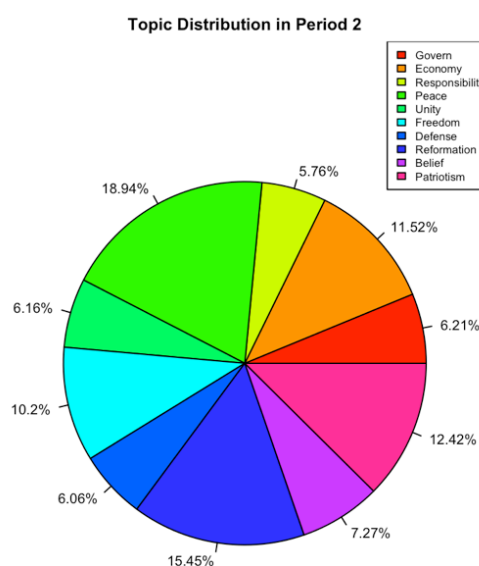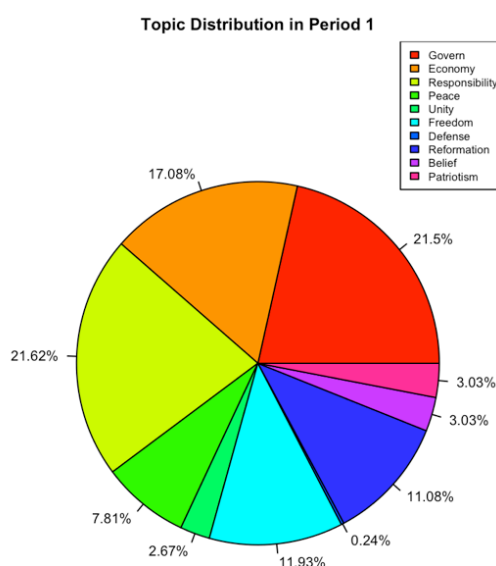
From the plots, we know that Responsibility and Govern play important roles in Period 1, Peace and Denfense play important roles in both Period 2 and Period 3. However, Peace and Denfense cover over 2/3 topic in Period 3.

I think this change of topics is also consistent with the development of our society, since people care more about living a peaceful life now.

Conclusion:

From my study, I will draw a conclusion that the trend of presidents' speeches from 1789 to 2017 corresponds with the development of our human behaviours.

1. Sentences become shorter than before, which implies that people are pursuing an simpler way to express themselves.

2. Less radical in sentiment means people now can control their emotion well and can always keep calm. In addition, people now become more possitive than before.

3. Topic change of speeches reflects our thoughts. It shows what kind of life we are living in or we want to live in.

Appendix: Word Cloud

In the end, I want to show the word clouds of speeches in three different periods.
(This is not part of my story, I just do this for fun.)

Data Collection and Processing

Step 0: Install and load libraries

Code

Step 1: Read in the speeches

Code

Step 2: Text processing

Code

Data Analysis — Wordclouds

Make wordclouds of speeches in three different periods using TF-IDF weighted document-term matrices.

As I would like to identify interesting words for speeches in each period, I use TF-IDF (https://en.wikipedia.org/wiki/Tf%E2%80%93idf) to weigh each term within speeches in each period. It highlights terms that are more widely-used in a particular period.

Code

Period 1

Period 2

Period 3