**Text Analytics Assignment : Business Insight Report**

**What discussion strategy will Buttigieg be able to beat Trump?**

**Dong Hoon Lee**

**Cohort 2**

**Part 1. Business Insights**

**Conclusion**

The report is based on an analysis of Buttigieg's Iowa Caucus speech and Trump's speech using Text analytics technique. Based on the Iowa Caucus rise, if Buttigieg is chosen as the Democratic presidential candidate, there are three things to focus on in a series of primaries and debates with Trump: First, we should focus more on the brevity and substance of the words we use in our speeches. Second, it is important to exclude overlaps in Trump and the use of the term, and to select and reiterate important key words. Third, if there is an expression related to the performance emphasized by Trump in the word bundle, it should be prepared to strike back on objective grounds.

**Background**

Interest in former Indiana Mayor Pete Buttigieg , 38, is rising in the Iowa caucuses, where he will pick the Democratic presidential candidate. Buttigieg stood out in the Democratic race because of his calm, logical rhetoric and young, fresh image. Also, he didn’t end up in the "buddy" storm but rose to the top of the race because of his moves reminiscent of former President Barack Obama. Compared to relatively radical candidates such as Warren and Sanders, the Democratic Party supporters are rather considered moderate and moderate. Unlike Buttigieg g, he is a 38-year-old young candidate. Therefore, Buttigieg is the best choice for centrist supporters who want a generational shift.

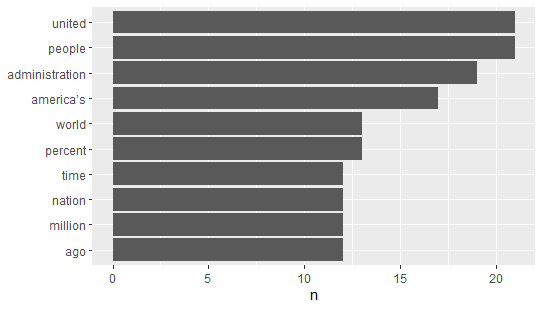
Speech, such as TV debates and campaigning, is the most important item among the fast-rising Buttigieg ’s winning ways to compete with President Trump. Therefore, we compared the speeches of the two and analyzed the strategies that Buttigieg should take in the discussions and speeches.

**Text Sources**I used a script for President Trump's recent speech by Buttigieg . For Trump, the State of the Union Speech and the Iowa Speech transcript were used for 2020 State of the Union Speech and Buttigieg . Based on the data in this analysis 1) Token frequency histograms 2) Corelograms 3) Bigram were analyzed to find the words, meanings and points of the Trump presidency (the "Trump") that Buttigieg should strategically target in the TV debate.

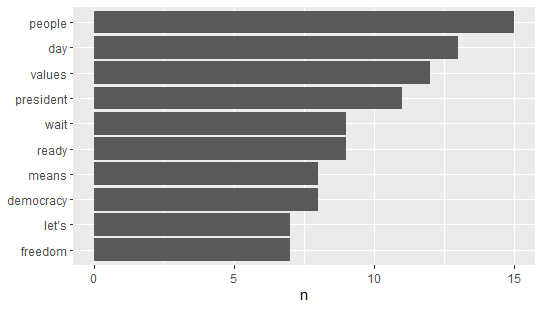
**Words frequency analysis**

The term Trump uses is ‘United’, ‘People’, ‘administration’, ‘America’ and ‘world’. On the other hand, in the case of Buttigieg , words such as ‘People’, ‘Day’, ‘Present’, ‘Value’, ‘Wait’ and ‘Ready’ were frequently used. Buttigieg needs to reduce the use of abstract words such as ‘President’ and ‘Value’ in future speech strategies.

**[ Figure 1. Token Frequencies for Trump ]**



**[ Figure 2. Token Frequencies for Buttigieg]**



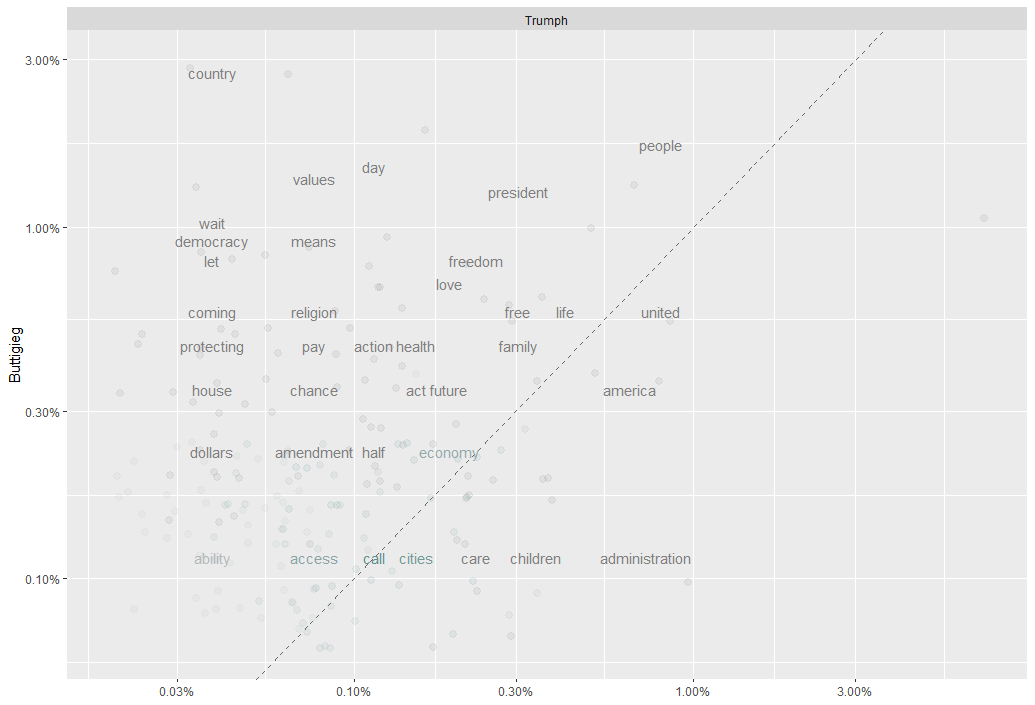
**Correlograms analysis**

The results of comparing the two people by calculating the application of the words used in the speech by Buttigieg and Trump are as follows.

Although there was more data on the Trumps used in the analysis, the Top 10 word use portion of the Trumps was largely constant, while Buttigieg had various terms calculated in the Correlogram. In other words, Buttigieg can be interpreted as using a richer vocabulary than Trump.

It was also analyzed that although Trump is not used very well, Buttigieg uses a lot of words such as ‘democracy’, ‘value’ and ‘protecting’. In doing so, it is important to recognize that there is a risk where the key message is diluted through a mixture of key words, although this may differentiate it from Trump.

**[ Figure 3. Correlograms ]**



**Bigram analysis**

The results of comparing the two people by calculating the application of the words used in the speech by Buttigieg and Trump are as follows. In the case of Trump, word bundles that are close to his or her own performance, such as ‘unemployment rate’, ‘Illegal aliens’, ‘American families’ and ‘criminal alien’ were used a lot. Buttigieg , on the other hand, used a lot of nouns related to ‘white house’, ‘American people’, ‘gun security’, ‘national security’ and ‘politic’.

From Buttigieg ’s perspective, it is important to respond if there are expressions that the other party is using a lot in preparation for future Trumps and speeches and debate contests. Unlike Trump, Buttigieg ’s word-collection can reduce clarity because the field is diverse and mentions a lot of political and social issues.

**[ Table 1. Bigram between Trump and Buttigieg ]**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Trump** | | | | **Buttigieg** | | | |
| Rank | word1 | word2 | n | Rank | word1 | word2 | n |
| 1 | unemployment | rate | 6 | 1 | white | house | 3 |
| 2 | days | ago | 5 | 2 | american | people | 2 |
| 3 | illegal | aliens | 4 | 3 | awful | lot | 2 |
| 4 | al | baghdadi | 3 | 4 | gun | violence | 2 |
| 5 | american | families | 3 | 5 | keeping | government | 2 |
| 6 | criminal | aliens | 3 | 6 | national | security | 2 |
| 7 | million | americans | 3 | 7 | political | opponent | 2 |
| 8 | opportunity | zones | 3 | 8 | political | party | 2 |
| 9 | previous | administration | 3 | 9 | precinct | leaders | 2 |
| 10 | sanctuary | cities | 3 | 10 | public | sector | 2 |

**References**

* Rev.com (2020, February 4) Transcript: Speeches at the Iowa Caucuses – Bernie Sanders, Joe Biden, Elizabeth Warren, Andrew Yang, Pete Buttigieg , Amy Klobuchar Speak. Rev.com. Retrieved from

[https://www.rev.com/blog/transcripts/transcript-speeches-at-the-iowa-caucuses-bernie-sanders-joe-biden-elizabeth-warren-andrew-yang-pete-Buttigieg -amy-klobuchar-speak](https://www.rev.com/blog/transcripts/transcript-speeches-at-the-iowa-caucuses-bernie-sanders-joe-biden-elizabeth-warren-andrew-yang-pete-buttigieg-amy-klobuchar-speak)

* Politico (2020, February 6) Trump's 2020 State of the Union address. Politico. Retrieved from

<https://www.politico.com/interactives/2020/trump-state-of-the-union-2020-live-fact-check-transcript-2-4-20/>

**Part 2. R Code and Output**

library(twitteR)

library(tm)

library(dplyr)

library(tidyverse)

library(tidytext)

library(stringr)

library(tidyr)

library(scales)

##################### Importing files from temp folder ######################

Trumph <- readLines('C:/temp/2020 Trumph State of the Union Speech.txt')

Buttigieg <- readLines('C:/temp/2020 Buttigieg Iowa Speech.txt')

my\_trump <- data\_frame(line=1:102, text = Trumph)

my\_butti <- data\_frame(line=1:48, text = Buttigieg )

# Tokenization

my\_token\_trump <- my\_trump %>%

unnest\_tokens(word, text)%>%

anti\_join(stop\_words)%>% #delete stop word

count(word, sort=TRUE)#sort and count

my\_token\_butti <- my\_butti %>%

unnest\_tokens(word, text)%>%

anti\_join(stop\_words)%>% #delete stop word

count(word, sort=TRUE)#sort and count

##################### Frequency Histogram ############################

# User defined for Lexicon

my\_junk <- data\_frame(

word = c("american","america", "americans","country","tonight",

"Congress","congress"),

lexicon = rep("junk", each = 7)

)

library(ggplot2)

freq\_hist1 <- my\_trump %>%

unnest\_tokens(word, text) %>%

anti\_join(stop\_words) %>%

anti\_join(my\_junk) %>%

count(word, sort=TRUE) %>%

mutate(word=reorder(word,n)) %>%

top\_n(10) %>%

ggplot(aes(word, n))+

geom\_col()+

xlab(NULL)+

coord\_flip()

print(freq\_hist1)

library(ggplot2)

freq\_hist2 <- my\_butti %>%

unnest\_tokens(word, text) %>%

anti\_join(stop\_words) %>%

anti\_join(my\_junk) %>%

count(word, sort=TRUE) %>%

mutate(word=reorder(word,n)) %>%

top\_n(10) %>%

ggplot(aes(word, n))+

geom\_col()+

xlab(NULL)+

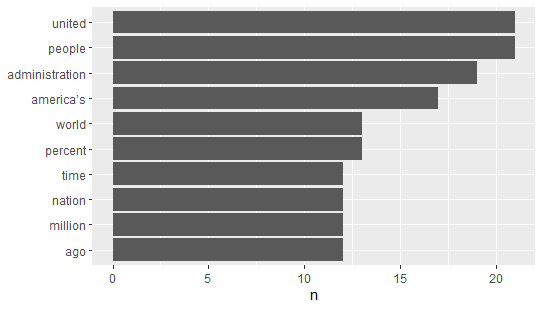
coord\_flip()

print(freq\_hist2)

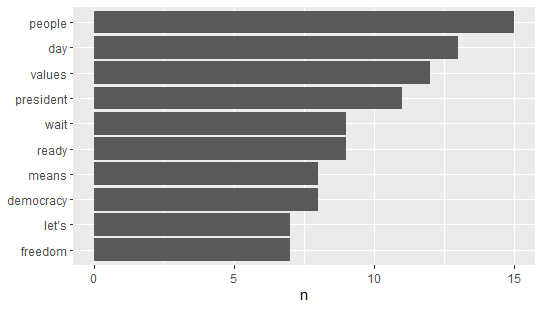
---------------------------------------------------------------------------------------------------------------------------------------

**[ R-Output]**

**Trump - Frequency**



**Buttigieg - Frequency**



##################### Frequency correlograms ############################

my\_tidy\_trump <- my\_trump %>%

unnest\_tokens(word, text)%>%

anti\_join(stop\_words)%>% #delete stop word

anti\_join(my\_junk)

my\_tidy\_butti <- my\_butti %>%

unnest\_tokens(word, text)%>%

anti\_join(stop\_words) %>%

anti\_join(my\_junk)

#Frequency grouping

frequency <- bind\_rows(mutate(my\_tidy\_trump, author= "Trumph"),

mutate(my\_tidy\_butti, author= "Buttigieg ")

)%>%#closing bind\_rows

mutate(word=str\_extract(word, "[a-z']+")) %>%

count(author, word) %>%

group\_by(author) %>%

mutate(proportion = n/sum(n))%>%

select(-n) %>%

spread(author, proportion)%>%

gather(author, proportion, `Trumph`)

head(frequency)

#Plotting Frequency

ggplot(frequency, aes(x=proportion, y=`Buttigieg `,

color = abs(`Buttigieg `- proportion)))+

geom\_abline(color="grey40", lty=2)+

geom\_jitter(alpha=.1, size=2.5, width=0.3, height=0.3)+

geom\_text(aes(label=word), check\_overlap = TRUE, vjust=1.5) +

scale\_x\_log10(labels = percent\_format())+

scale\_y\_log10(labels= percent\_format())+

scale\_color\_gradient(limits = c(0,0.001), low = "darkslategray4", high = "gray75")+

facet\_wrap(~author, ncol=2)+

theme(legend.position = "none")+

labs(y= "Buttigieg ", x=NULL)

**[ R-Output]**

> head(frequency)

# A tibble: 6 x 4

word Buttigieg author proportion

*<chr>* *<dbl>* *<chr>* *<dbl>*

1 abandon NA Trumph 0.000386

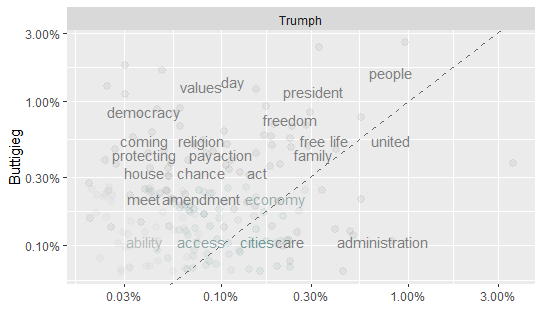
2 abiding NA Trumph 0.000386

3 ability 0.00130 Trumph 0.000386

4 abolish NA Trumph 0.000386

5 abortion NA Trumph 0.000386

6 abraham NA Trumph 0.000386



################## Bigram - Pair Words Analysis #########################

my\_trump\_bigrams <- my\_trump %>%

unnest\_tokens(bigram, text, token = "ngrams", n=2)

my\_butti\_bigrams <- my\_butti %>%

unnest\_tokens(bigram, text, token = "ngrams", n=2)

my\_trump\_bigrams %>%

count(bigram, sort = TRUE) #this has many stop words, need to remove them

my\_butti\_bigrams %>%

count(bigram, sort = TRUE) #this has many stop words, need to remove them

# separate function:

trump\_big\_sep <- my\_trump\_bigrams %>%

separate(bigram, c("word1", "word2"), sep = " ")

butti\_big\_sep <- my\_butti\_bigrams %>%

separate(bigram, c("word1", "word2"), sep = " ")

turmp\_big\_fil <- trump\_big\_sep %>%

filter(!word1 %in% stop\_words$word) %>%

filter(!word2 %in% stop\_words$word)

butti\_big\_fil <- butti\_big\_sep %>%

filter(!word1 %in% stop\_words$word) %>%

filter(!word2 %in% stop\_words$word)

#creating the new bigram, "no-stop-words":

trump\_big\_counts <- turmp\_big\_fil %>%

count(word1, word2, sort = TRUE)

#want to see the new bigrams

print(trump\_big\_counts)

butti\_big\_counts <- butti\_big\_fil %>%

count(word1, word2, sort = TRUE)

#want to see the new bigrams

print(butti\_big\_counts)

**[ R-Output]**

> print(trump\_big\_counts)

# A tibble: 1,008 x 3

word1 word2 n

*<chr>* *<chr>* *<int>*

1 unemployment rate 6

2 days ago 5

3 illegal aliens 4

4 al baghdadi 3

5 american families 3

6 criminal aliens 3

7 million americans 3

8 opportunity zones 3

9 previous administration 3

10 sanctuary cities 3

# ... with 998 more rows

> print(butti\_big\_counts)

# A tibble: 154 x 3

word1 word2 n

*<chr>* *<chr>* *<int>*

1 white house 3

2 american people 2

3 awful lot 2

4 gun violence 2

5 keeping government 2

6 national security 2

7 political opponent 2

8 political party 2

9 precinct leaders 2

10 public sector 2

# ... with 144 more rows