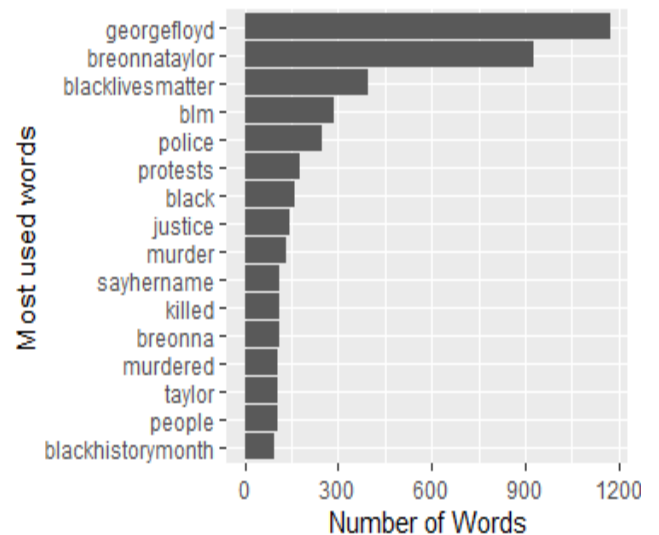


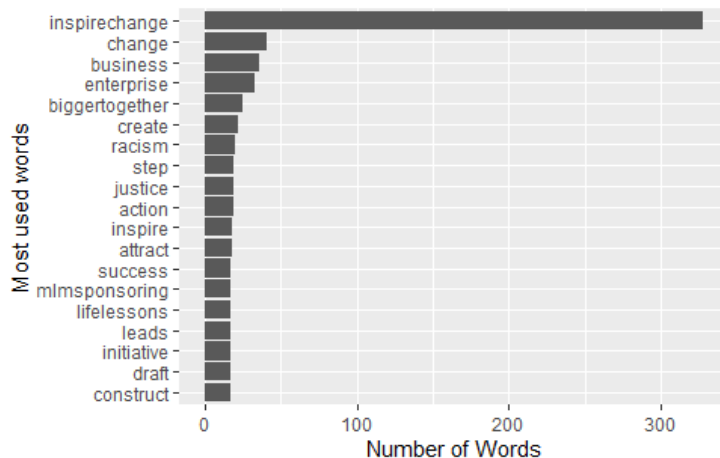
Racism covid and super bowl

Society had suffer from a cronique disease for centuries. This disease is racism, during 2020 there were two event that exalted this social-problem. This report is based on tweeters obtain form "Rtweet" library. The # used for this analysis were #Georgefloyd #breonnataylor, #inspirechange, #nfl, #racism, #blacklivematter, #racism, #covid, #Cov-19. The aim of this report is to demonstrate how racism has been a cronique dease for society nowadays, anlysisng form two perspective, if there is any relation between pandemic COVID-19 and Racism. The second perspecitve see if superbowl commercil inspire change had a possitive impact in this social problema.

Floyd & Taylor



Inspire NFL

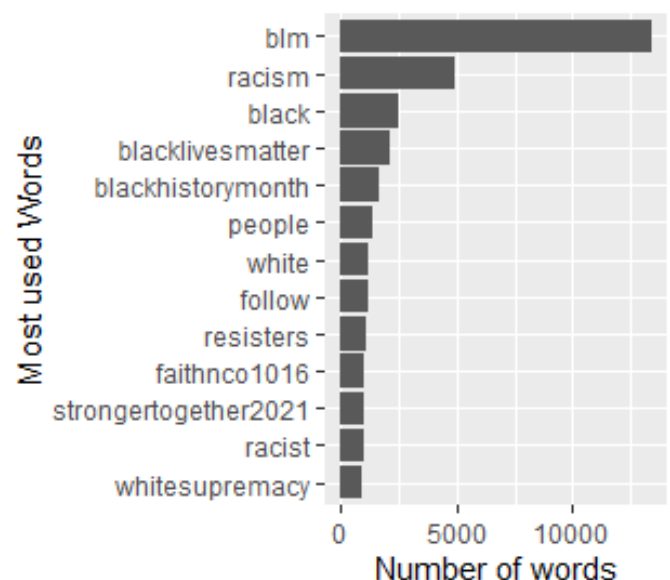


The histograms of this page show the most common word used in this queries. It is important to analyze the words: Racism, police, attack and justice. Theses word can be observe in the three histograms. Thus, it shows that this feeling of racism comes from injustice. In NFL commercial "inspire change" the message is that united we are better. However, it seems that the commercial brings to the table feeling of anger or sadness. On the Racism histogram there is a feeling of

white supremacy which show the injustice that people are appealing. Racism can be observe all around our lifes. For example, Google photos, tag black people as "gorillas"(Christian, 2020). Google had this problem with its algorithm, they already worked on it to fixed it. However, while they fixe one algorithm this problems continue happening throughout lifes.

During this days, society is going through one of the biggest challenges we've ever had. Pandemic of covid-19. As lockdowns increase, suiciedes and mental illlness increase (Rickards, 2021). Likewise, racism has increase due to different fake news around cultures and races.

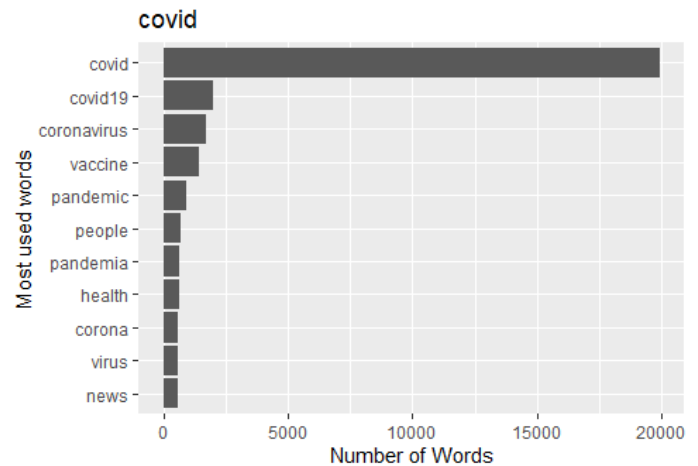
Racism



Juan Camilo Salazar

Text Analytics

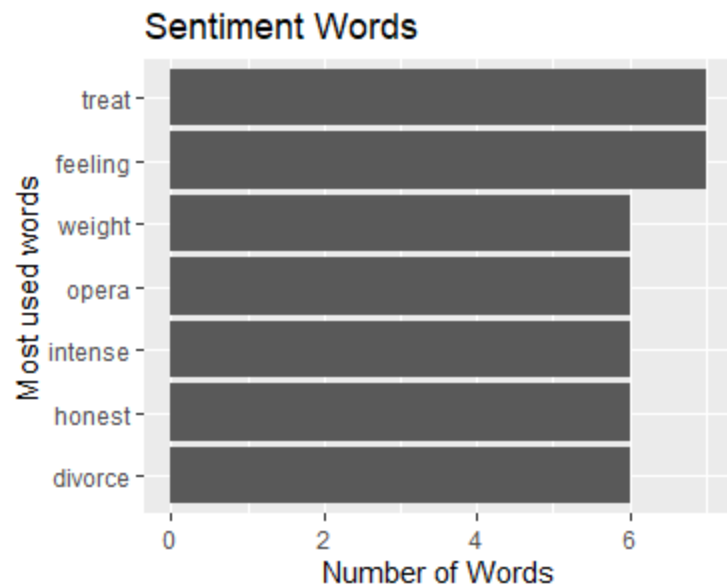
The covid histogram shows that there is a feeling of unity around PEOPLE and also news have had a important roll in this pandemic. With this words is not possible to observe any feeling of racism or anger against other cultures or race. Most of these words are focus on seeking of information about what is happening with covid on these days.



Sentiments

To observe which were the dominant feelings and sentiments it was done a analysis with the nrc dataframe. The nrc dataframe has seven levels of sentiments: Anger, Disgust, Fear, Sadness, Joy, Trust and Surprise. There are three positive and four negative.

On the Sentiment word histogram, which is composed from the combination of the four datasets. It is shown that most of the words used are negative in this histogram. Treat, intense, divorce are word consider as negative feelings.

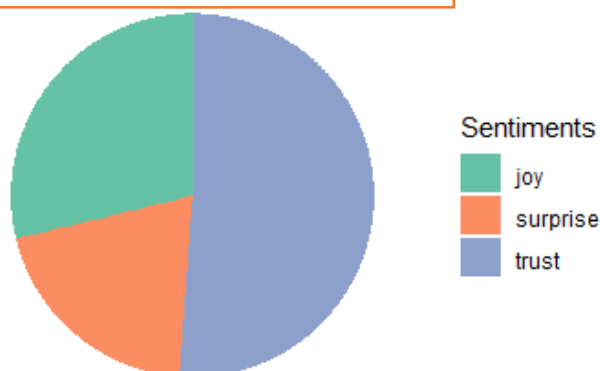


On the other hand, pie chart of the negative sentiments it can be observe that the most common sentiments are anger and fear. The feeling of anger is express when there is injustice and it cause your body to realese adrenaline(Mayo, 2021) what can justify some of the actions of the 26 of may in the United States.

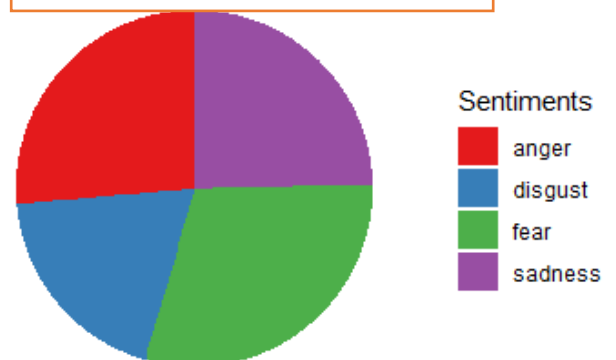
As to positive sentiment the most relevant is trust.

People still have some faith in the society that one day people change and understand that we are the same humans and we live in the same planet.

Pie Chart Positive Sentiments



Pie Chart Negative Sentiments

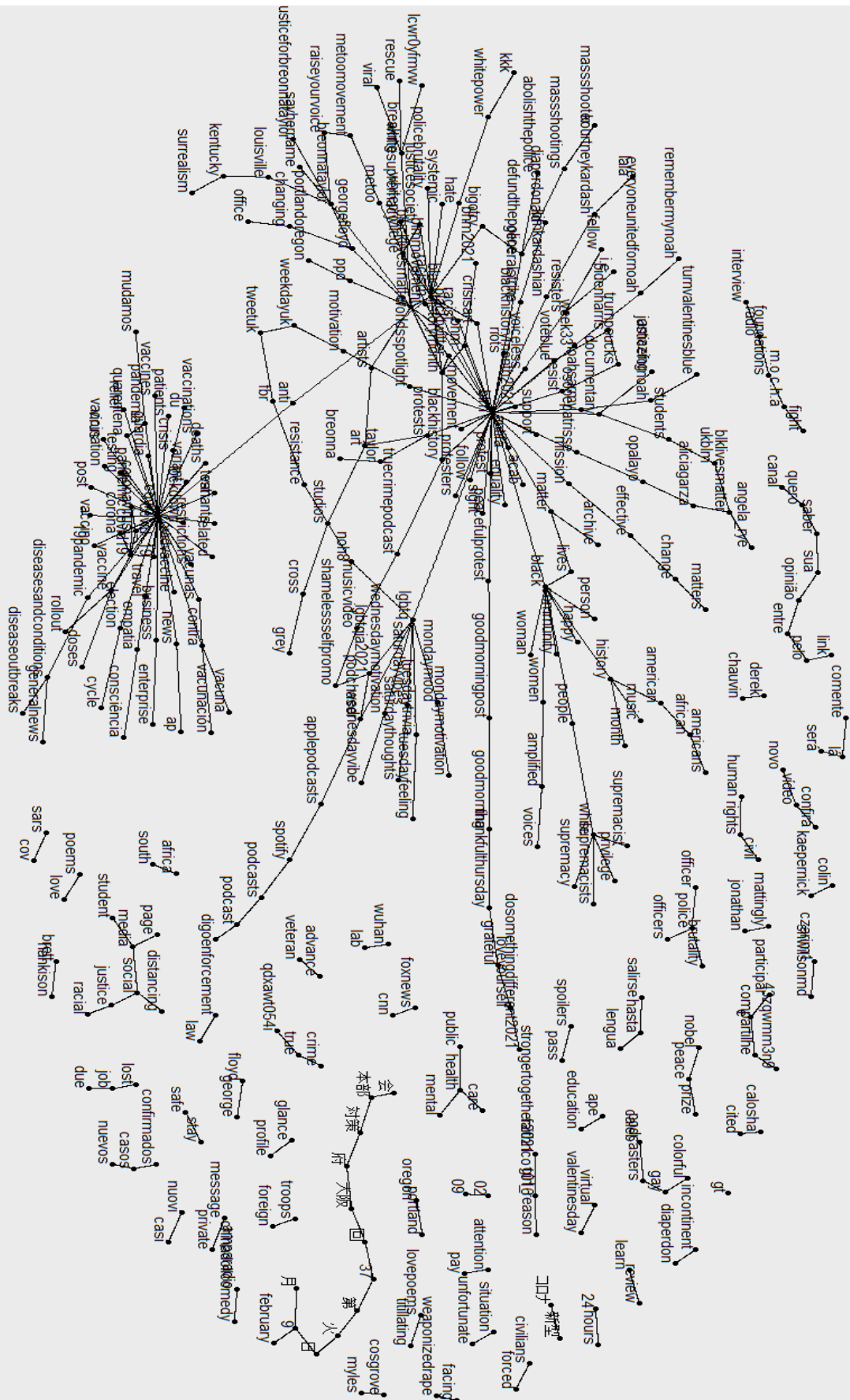


Juan Camilo Salazar

Text Analytics

Finally, the word cloud shows two nodes. The outstanding is the racism word from which more words start to come out. There are hundreds of words that are correlated to the word racism no matter the area. Politics with the election of 2020 which evoke a feeling of division and supremacy. Likewise, as it can be observed there is a connection between the COVID node and Racism, which confirms that it is possible that COVID situation has increase racism due to how media and we people use our language in our daily life.

In conclusion, campaigns such as the NFL commercial during the Superbowl evoke a feeling of change and trust. However, society must continue working in this process of change and learning. Understand that we live in the same planet and we all are equal.



Juan Camilo Salazar

Text Analytics

References

Christian, B. (2020). *The Alignment Problem*. New York: W.W Northon Company .

Rickards, J. (2021). *The New Great Depression*. Penguin.

Appendix

Code

```
library(rtweet)
```

```
library(dplyr)
```

```
library (dplyr)
```

```
library (stringr)
```

```
library(scales)
```

```
library(tidyr)
```

```
library(ggplot2)
```

```
##### Racism twitters #####
```

```
Racism_blm<- search_tweets(
```

```
  " #Racism OR #blm OR #whitesupremacy", n = 18000, include_rts = FALSE,
```

```
  fromDate = "202102070000", toDate = "202102072359"
```

```
)
```

```
c<- c("https", "t.co", "sur", "a", "b", "c", "d", "e", "f", "g", "h", "i", "j", "k", "l", "m", "n", "o", "p", "19",  
      "di", "il",
```

```
"q", "r", "s", "t", "u", "v", "w", "x", "y", "z", "es", "por", "won", "su", "2", "al", "las", "fue", "half", "pepsi", "half",  
me", "3", "tiempo", "fans", "7", "tampabaybuccaneers", "5", "más", "day", "lv",
```

```
"2021", "bay", "2021", "superbowl2021", "superbowlweeknd", "para", "una", "avec", "les", "top",  
"qui", "est", "s21",
```

```
"superbowl", "le", "sblv", "", "chipotle", "commercial", "halftime", "chipotleismylife", "chipotleismylife",  
"chipotleminis", "watch", "1", "team", "mahomes", "con", "ad", "ads", "lo",  
"weeknd", "se", "bucs", "gobucs", "win", "theweeknd", "superbowl", "super",
```

Juan Camilo Salazar

Text Analytics

```
"bowl", "de", "del", "chiefs", "game", "tampa", "amp", "time", "buccaneers",  
"los", "el", "la", "en", "tom", "nfl", "tombrady", "brady", "superbowlcommercials",  
"superbowlsunday")
```

```
c <- as.data.frame(c)
```

```
names(c)[names(c) == "c"] <- "word"
```

```
### unnest token and atin join stop words an other words #####
```

```
tidy_mydf_racism <- Racism_blm %>%
```

```
  unnest_tokens(word, text) %>%
```

```
  anti_join(stop_words) %>%
```

```
  anti_join(c) %>%
```

```
  count(word, sort = TRUE)
```

```
print(tidy_mydf_racism)
```

```
##### create histogram for the racism dataset #####
```

```
freq_hist_Racism <- tidy_mydf_racism %>%
```

```
  filter(n > 950) %>% # we need this to eliminate all the low count words
```

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

```
  ggplot(aes(word, n))+
```

```
  geom_col()+
```

```
  xlab(NULL)+
```

```
  coord_flip()
```

```
freq_hist_Racism <- freq_hist_Racism + ggtitle("Racism") + xlab("Most used Words") +  
ylab("Number of words")
```

```
print(freq_hist_Racism)
```

new search about inspire change commercial superbowl

```
inspire <- search_tweets("#inspirechange", n = 18000 , include_rts = FALSE,  
                        fromDate = "202102070000", toDate = "202102072359")
```

```
tidy_mydf_isnpire <- inspire %>%  
  unnest_tokens(word, text) %>%  
  anti_join(stop_words) %>%  
  anti_join(c) %>%  
  count(word, sort = TRUE)  
print(tidy_mydf_isnpire)
```

```
freq_hist_inspire <- tidy_mydf_isnpire %>%  
  filter(n > 16) %>%  
  mutate (word = reorder(word, n)) %>%  
  ggplot(aes(word, n)) +  
  geom_col() +  
  xlab(NULL) +  
  coord_flip()  
freq_hist_inspire <- freq_hist_inspire + ggtitle("Inspire NFL") + xlab("Most used words") +  
  ylab("Number of Words")  
print(freq_hist_inspire)
```

new search for george floyd

Juan Camilo Salazar

Text Analytics

```
George_floyd <- search_tweets("#Georgefloyd OR #BreonnaTaylor", n = 18000 , include_rts = FALSE,
```

```
    fromDate = "202102070000", toDate = "202102072359")
```

```
tidy_mydf_george <- George_floyd %>%
```

```
  unnest_tokens(word, text) %>%
```

```
  anti_join(stop_words) %>%
```

```
  anti_join(c) %>%
```

```
  count(word, sort = TRUE)
```

```
print(tidy_mydf_george)
```

```
freq_hist_george <- tidy_mydf_george %>%
```

```
  filter(n > 90) %>%
```

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

```
  ggplot(aes(word, n)) +
```

```
  geom_col() +
```

```
  xlab(NULL) +
```

```
  coord_flip()
```

```
freq_hist_george <- freq_hist_george + ggtitle("Floyd & Taylor") + xlab("Most used words") +  
ylab("Number of Words")
```

```
print(freq_hist_george)
```

```
##### COVID#####
```

```
covid <- search_tweets("#COVID OR #COV-19", n = 18000 , include_rts = FALSE,
```

```
    fromDate = "202102070000", toDate = "202102072359")
```

```
tidy_mydf_covid <- covid %>%
```

```
  unnest_tokens(word, text) %>%
```

```
  anti_join(stop_words) %>%
```

```
  anti_join(c) %>%
```


Juan Camilo Salazar

Text Analytics

```
count(word, sort = TRUE)

print(tidy_mydf_covid)

freq_hist_covid <- tidy_mydf_covid%>%
  filter(n > 550) %>%
  mutate (word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()

freq_hist_covid <- freq_hist_covid + ggtitle("covid") + xlab("Most used words") + ylab("Number of Words")

print(freq_hist_covid)
```

correlogram

```
frequency <- bind_rows(mutate(tidy_mydf_racism, author="Racism"),
                        mutate(tidy_mydf_isnpire, author= "Inspire Change"),
                        mutate(tidy_mydf_george, author= "George"),
                        mutate(tidy_mydf_covid, author="Jane Austen")
) %>% #closing bind_rows

mutate(word=str_extract(word, "[a-z']+")) %>% # remove numbers

count(author, word) %>%

group_by(author) %>%

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

Juan Camilo Salazar

Text Analytics

```
select(-n) %>% #remove n we dont need n

spread(author, proportion) %>%

gather(author, proportion, `Inspire Change`, `Racism`, `George` )
```

#let's plot the correlograms:

```
ggplot(frequency, aes(x=proportion, y=`Jane Austen`,
                      color = abs(`Jane Austen` - proportion)))+
  geom_abline(color="grey40", lty=3)+
  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=3)+
  theme(legend.position = "none")+
  labs(y= "COVID", x=NULL)
```

Sentiment Analysis

```
sentiments_d <- bind_rows(mutate(tidy_mydf_racism, author="Racism"),
                           mutate(tidy_mydf_isnpire, author= "Inspire Change"),
                           mutate(tidy_mydf_george, author= "George"),
                           mutate(tidy_mydf_covid, author="Jane Austen")
)
```

Juan Camilo Salazar

Text Analytics

```
afinn <- get_sentiments("afinn")
```

```
nrc <- get_sentiments("nrc") #flavor of sentiments
```

```
bing <- get_sentiments("bing") # binary
```

```
sentiments <- bind_rows(mutate(afinn, lexicon="afinn"),
```

```
  mutate(nrc, lexicon= "nrc"),
```

```
  mutate(bing, lexicon="bing")
```

```
)
```

```
sentiments %>%
```

```
  filter(lexicon == "nrc")
```

```
nrc_1 <- nrc
```

```
nrc_1$sentiment <- as.factor(nrc_1$sentiment)
```

```
levels(nrc_1$sentiment)
```

```
nrcsurprise <- get_sentiments("nrc") %>%
```

```
  filter(sentiment == "fear" )
```

```
#inner joining the emma book and the surprise sentiments
```

```
fear <- sentiments_d %>%
```

```
  inner_join(nrcsurprise) %>% #intersection between 2
```

```
  count(word, sort=T) %>%
```

```
  mutate(sentiment = "fear")
```

```
nrcsurprise <- get_sentiments("nrc") %>%
```

Juan Camilo Salazar

Text Analytics

```
filter(sentiment == "anger" )
```

#inner joining the emma book and the surprise sentiments

```
anger <- sentiments_d%>%
```

```
inner_join(nrcsurprise) %>% #intersection between 2
```

```
count(word, sort=T) %>%
```

```
mutate(sentiment = "anger")
```

```
nrcsurprise <- get_sentiments("nrc") %>%
```

```
filter(sentiment == "sadness" )
```

#inner joining the emma book and the surprise sentiments

```
sadness <- sentiments_d%>%
```

```
inner_join(nrcsurprise) %>% #intersection between 2
```

```
count(word, sort=T) %>%
```

```
mutate(sentiment = "sadness")
```

```
nrcsurprise <- get_sentiments("nrc") %>%
```

```
filter(sentiment == "disgust" )
```

#inner joining the emma book and the surprise sentiments

```
disgust <- sentiments_d%>%
```

```
inner_join(nrcsurprise) %>% #intersection between 2
```

```
count(word, sort=T) %>%
```

```
mutate(sentiment = "disgust")
```

```
negative_s <- rbind(disgust,sadness,anger,fear)
```

```
negative_top10 <- subset(negative_a[1:10, ])
```

```
negative_a <- negative_top10 %>%
```

```
  ggplot(aes(word, n)) +
```

```
  geom_col() +
```

```
  xlab(NULL) +
```

```
  coord_flip()
```

```
freq_hist_negative <- negative_a + ggtitle("Negative Words") + xlab("Most used words") +  
ylab("Number of Words")
```

```
print(freq_hist_negative)
```

```
bp_N<- ggplot(negative_s,labels = piepercent, aes(x="", y=n, fill=sentiment))+
```

```
  geom_bar(width = 1, stat = "identity")
```

```
bp_N
```

```
pie_N <- bp_N + coord_polar("y", start=0)
```

```
pie_N
```

```
blank_theme <- theme_minimal()+
```

```
  theme(
```

```
    axis.title.x = element_blank(),
```

```
    axis.title.y = element_blank(),
```

```
    panel.border = element_blank(),
```

Juan Camilo Salazar

Text Analytics

```
panel.grid=element_blank(),  
axis.ticks = element_blank(),  
plot.title=element_text(size=14, face="bold")  
)
```

```
pie_negative <- pie_N + scale_fill_brewer("Sentiments", palette = "Set1") + blank_theme +  
  theme(axis.text.x=element_blank())  
print(pie_negative)
```

```
##### positive #####
```

```
nrcsurprise <- get_sentiments("nrc") %>%  
  filter(sentiment == "joy" )
```

#inner joining the emma book and the surprise sentiments

```
joy <- sentiments_d %>%  
  inner_join(nrcsurprise) %>% #intersection between 2  
  count(word, sort=T) %>%  
  mutate(sentiment = "joy")
```

```
nrcsurprise <- get_sentiments("nrc") %>%  
  filter(sentiment == "trust" )
```

#inner joining the emma book and the surprise sentiments

```
trust <- sentiments_d %>%  
  inner_join(nrcsurprise) %>% #intersection between 2  
  count(word, sort=T) %>%  
  mutate(sentiment = "trust")
```

```
nrcsurprise <- get_sentiments("nrc") %>%
```

Juan Camilo Salazar

Text Analytics

```
filter(sentiment == "surprise" )
```

#inner joining the emma book and the surprise sentiments

```
surprise <- sentiments_d%>%
```

```
inner_join(nrdsurprise) %>% #intersection between 2
```

```
count(word, sort=T) %>%
```

```
mutate(sentiment = "surprise")
```

```
positive_s <- rbind(joy,trust,surprise)
```

```
positive_b <- positive_s %>%
```

```
count(word, sort = TRUE)
```

```
positive_top10 <- subset(positive_b[1:10, ])
```

```
positive_a <- positive_top10 %>%
```

```
ggplot(aes(word, n)) +
```

```
geom_col() +
```

```
xlab(NULL) +
```

```
coord_flip()
```

```
freq_hist_positive <- positive_a +ggtitle("Positive Words") + xlab("Most used words") +  
ylab("Number of Words")
```

```
print(freq_hist_positive)
```

```
bp_N<- ggplot(positive_s, aes(x="", y=n, fill=sentiment))+
```

Juan Camilo Salazar

Text Analytics

```
geom_bar(width = 1, stat = "identity")
```

```
bp_N
```

```
pie_N <- bp_N + coord_polar("y", start=0)
```

```
pie_N
```

```
blank_theme <- theme_minimal()+
```

```
theme(
```

```
  axis.title.x = element_blank(),
```

```
  axis.title.y = element_blank(),
```

```
  panel.border = element_blank(),
```

```
  panel.grid=element_blank(),
```

```
  axis.ticks = element_blank(),
```

```
  plot.title=element_text(size=14, face="bold")
```

```
)
```

```
pie_positive <- pie_N + scale_fill_brewer("Sentiments", palette="Set2") + blank_theme +
```

```
  theme(axis.text.x=element_blank())
```

```
print(pie_positive)
```

```
#####
```

```
total <- rbind(positive_s, negative_s)
```

```
total_1 <- total%>%
```

```
  count(word, sort = TRUE) %>%
```

```
  top_n(5, n) %>%
```

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


Juan Camilo Salazar

Text Analytics

```
ggplot(aes(word, n)) +  
geom_col() +  
xlab(NULL) +  
coord_flip()  
freq_hist_covid_1 <- total_1 + ggtitle("Sentiment Words") + xlab("Most used words") +  
ylab("Number of Words")  
print(freq_hist_covid_1)
```

```
bp_N<- ggplot(total, aes(x="", y=n, fill=sentiment))+  
  geom_bar(width = 1, stat = "identity")  
bp_N
```

```
pie_T <- bp_N + coord_polar("y", start=0)  
pie_T
```

```
blank_theme <- theme_minimal()+  
  theme(  
    axis.title.x = element_blank(),  
    axis.title.y = element_blank(),  
    panel.border = element_blank(),  
    panel.grid=element_blank(),  
    axis.ticks = element_blank(),  
    plot.title=element_text(size=14, face="bold")  
  )
```

```
pie_Total <- pie_T + scale_fill_brewer("Sentiments", palette="Set2") + blank_theme +
```

Juan Camilo Salazar

Text Analytics

```
theme(axis.text.x=element_blank())  
print(pie_Total)
```

```
#####
```

```
Racism_blm
```

```
full_tweet <- rbind(Racism_blm,covid,inspire,George_floyd)
```

```
austen_bigrams <- full_tweet %>%
```

```
unnest_tokens(bigram, text, token = "ngrams", n=2) %>%
```

```
filter(!is.na(bigram))
```

```
austen_bigrams #We want to see the bigrams (words that appear together, "pairs")
```

```
austen_bigrams %>%
```

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

```
#to remove stop words from the bigram data, we need to use the separate function:
```

```
library(tidyr)
```

```
bigrams_separated <- austen_bigrams %>%
```

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

```
bigrams_filtered <- bigrams_separated %>%
```

```
filter(!word1 %in% stop_words$word) %>%
```

```
filter(!word2 %in% stop_words$word) %>%
```

Juan Camilo Salazar

Text Analytics

```
filter(!word1 %in% c$word) %>%
```

```
filter(!word2 %in% c$word)
```

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

```
bigram_counts <- bigrams_filtered %>%
```

```
count(word1, word2, sort = TRUE)
```

```
#want to see the new bigrams
```

```
bigram_counts
```

```
bigram_graph <- bigram_counts %>%
```

```
filter(n>30) %>%
```

```
graph_from_data_frame()
```

```
bigram_graph
```

```
#install.packages("xxxxxxx")
```

```
#library(xxxxxxxx)
```

```
ggraph(bigram_graph, layout = "fr") +
```

```
geom_edge_link()+
```

```
geom_node_point()+
```

```
geom_node_text(aes(label=name), vjust =1, hjust=1)
```

Juan Camilo Salazar

Text Analytics

Outputs

