

Introduction

The US citizen has just recently celebrated the annual National Football League as called Super Bowl in Raymond James Stadium in Tampa, Florida on Sunday. The Super Bowl matchup between Patrick Mahomes and the Chiefs and Tom Brady and the Buccaneers will host the smallest number of attendances in Super Bowl history which limited to vaccinated health care workers only. With 23 first-time advertisers, the pandemic's impact on commercial production, Super Bowl 2021 is a new experience (Liffreing, 2021) and Twitter offers a platform of choice for viewers to share their opinions. Based on that, we decided to use Twitter API to pull the data to analyze our concerns, which are:

- What is the most tweeted topic during Super Bowl? By analyzing the topic or word posted on Twitter, we can further predict attendance's interest and concern for the Super Bowl game.
- What is the most mention activity during the 'Half Time Show' performance? In order to analyze this, we will pull the tweets that related to Half Time Show only.
- What is the most favorite commercial on the Super Bowl this year? Based on CNN, Super Bowl commercials matters just as much as the game itself.
- How do they feel about the Super Bowl this year? As we know that pandemic has limit people for gathering and watching the game together as they usually do.

We will use some hashtags below to give us insights to analyze the topic.

#SBLV

#SuperBowl + #halftimeshow

#superbowlcommercials

Insights

When we pull the Twitter API for Super Bowl 2021 tweets, we put it into the Wordcloud to show a general wider insight of the tweeted topic during Super Bowl 2021 (**appendix 1**) and we can see a lot of enthusiasm from the fans that supporting either Buccaneers or Chiefs. Also, we can tell that Tom Brady and Mahomes get the most support as the captain of the team. From the Wordcloud, we can tell that Half Time Show is the most waited and high-lighted performance of the Super Bowl especially for The Weeknd performance. Further, we see an excitement for the return of retirement in 2020, Gronk or Rob Gronkowski, as the savor of Super Bowl with his old friend, Tom Brady. We can spot the word "GOAT" stands for "Greatest of All Time" from the Wordcloud that refers to Tom Brady as he just nominated as the best player in football.

Based on the commercial hashtag, we can see the top 5 most-tweeted commercials on Super Bowl 2021 are Jeep with Bruce Springsteen, Cheetos with Mila Kunis, Shaggy and Ashton Kutcher, Amazon Alexa with Michael Jordan, Reddit and Cadillac with Scissorhands.

When we dig deeper to the tweets with bi-grams and visualize it, we find a really interesting pairs such as farmer protest (**appendix 2**). Based on the The Indian Express, the farmer protest ad has been played during Super Bowl to raise help for the farmer in India as they demand the

government to repeal the farm laws and legal guarantee of minimum support price for their crops.

Further, we want to explore Half Time Show performance to the most trending topic as and we can see that Pepsi is the biggest sponsor for Half Time Show (**appendix 3**), however, again, we see The Weeknd achieved high engagement with Super Bowl fans as the solo headliner of the Half Time Show compare to Pepsihalftime tweets. Based on AdAge, The Weeknd has contributed 65% of mentions during the game.

Lastly, we would like to know what people feel about Super Bowl this year and based on our NRC sentiment analysis (**appendix 4**), we can see that regarding the game itself, people have anticipated that Tom Brady and the Buccaneers will win the game as the score gap differences are huge, however, we see a lot of happiness, hope and love for the game. On the negative sides, we can see that mostly has pandemic aspect in the criteria. There are few assumptions happening for the raise of pandemic tweets in Twitter. First, people think the stadium is full of people with no concern of Covid. Instead, the stadium is filled with vaccinated health-care workers as appreciation and it looks full because of the cardboard cutouts and second assumption is cultural activity. People used to gather, sit on the stadium or bar with family and friends watching the game but now they have to stay at home and only gather with people in their house for social distancing purpose.

Conclusion

Based on our insight, The Weeknd performance for the Half Time Show has the most significant engagement of the people based on their mentions via Twitter with their most popular song called Blinding Lights. Also, Tom Brady's achievement to himself and the Buccaneers is enormous because we can see his name keeps showing in various Super Bowl hashtag on Twitter other than winning the game.

Super Bowl commercial is also the most waited moment of the game. It is becoming a competition to win the best commercial in the Super Bowl. As mention by Business Insider, studies have shown that commercials that aired during the Super Bowl are found to boost the weekend sales by more than twice the cost of the ad time. Not only serve to drive sales, but Super Bowl commercials can also be used to deliver strong message to reach millions of people. When we analyze deeply with bi-gram, we find that this year, Super Bowl shows more than just interactive commercials but also farmer's protest video known as the largest protest in history. The purpose is to raise awareness about India's farmers' protest that are still going on.

Enthusiasm for the Half Time Show has a really huge impact for the business not only for the Super Bowl but also for the publicity of the artist. It is a win-win situation where the game wants revenue from their commercial and show while all the commercials and artist will get attention they need for publicity to increase their sales (Heyen, 2021).

Lastly, in spite of all the sadness, fear, anger and negative from pandemic and watch the game at home, there is still hope, trust and joy from the Super Bowl game, commercial and performance.

References:

Asmelash, L. (2021, February 8). *Here are the best Super Bowl commercials.*

CNN. <https://www.cnn.com/2021/02/08/us/best-super-bowl-commercials-trnd/index.html>

Davis, N. (2021, February 8). *Miserable to masterful: Rob Gronkowski basks in glory of Super Bowl 55 victory following brief retirement.* USA

TODAY. <https://www.usatoday.com/story/sports/nfl/columnist/nate-davis/2021/02/08/super-bowl-rob-gronkowski-buccaneers/4427608001/>

Liffreing, I. (2021, February 8). ***WHICH BRANDS WON (AND LOST) THE SUPER BOWL ON***

TWITTER. AdAge. <https://adage.com/article/special-report-super-bowl/which-brands-won-and-lost-super-bowl-twitter/2312916>

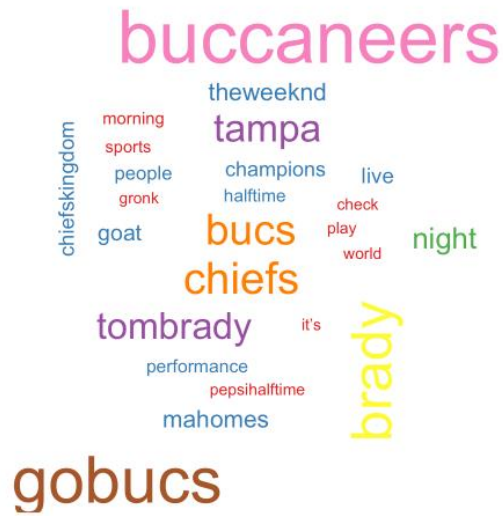
Frank, J. (2020, January 31). ***How the Super Bowl became the championship of advertising and how to win it today.*** Business. <https://www.businessinsider.com/how-the-super-bowl-became-the-championship-of-advertising-2020-1>

Heyen, B. (2021, February 7). ***Why is The Weeknd performing at Super Bowl 55? Explaining the NFL's 2021 halftime show decision.*** Sportingnews.

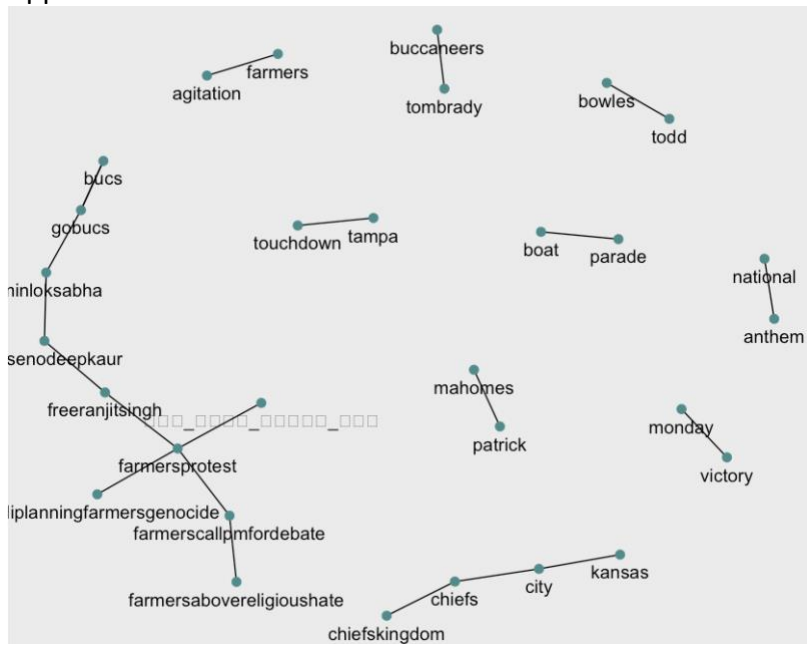
<https://www.dashboard.visme.co/v2/projects/own>.

Appendix:

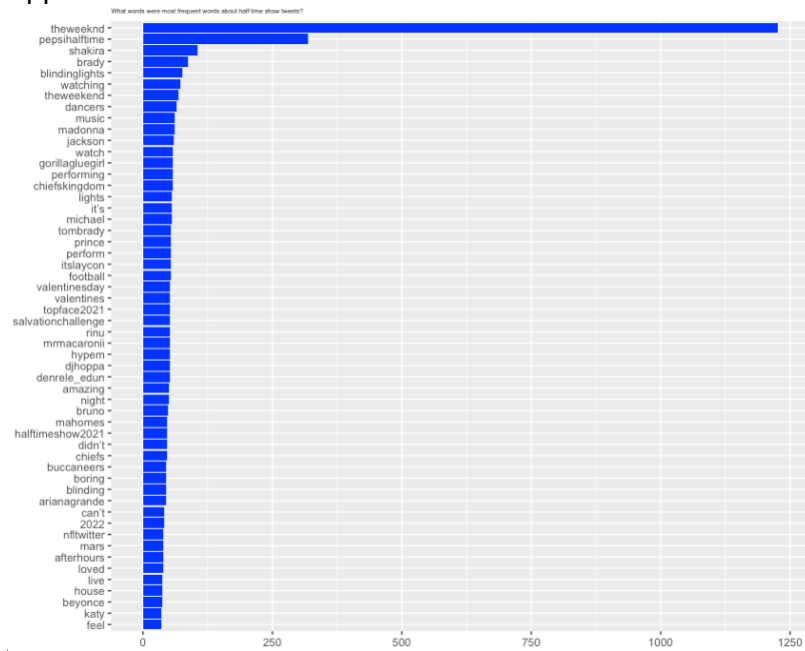
Appendix 1



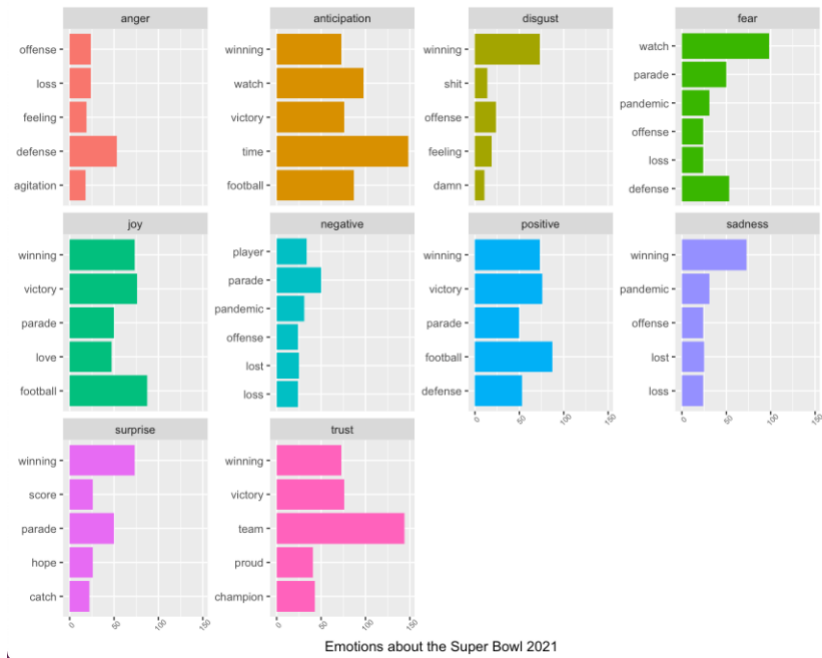
Appendix 2



Appendix 3



Appendix 4



R code :

```
#####  
# importing all necessary packages  
#####  
#install.packages("rtweet")  
#install.packages("stopwords")  
#install.packages("RColorBrewer")  
#install.packages("widyr")  
#install.packages("igraph")  
library(rtweet)  
library(dplyr)  
library(tidytext)  
library(tidyverse)  
library(ggplot2)  
library(stopwords)  
library(tm)  
library(wordcloud)  
library(RColorBrewer) #adding some color to our wordcloud  
library(widyr) #for n-grams  
library(tidyr)  
library(igraph)  
library(ggraph)  
library(dplyr)  
library(tidytext)
```

```
> #install.packages("rtweet")  
> #install.packages("stopwords")  
> #install.packages("RColorBrewer")  
> #install.packages("widyr")  
> #install.packages("igraph")  
> library(rtweet)  
> library(dplyr)  
> library(tidytext)  
> library(tidyverse)  
> library(ggplot2)  
> library(stopwords)  
> library(tm)  
> library(wordcloud)  
> library(RColorBrewer) #adding some color to our wordcloud
```

```
#####  
# create library from Twitter  
#####  
##### first library: sblv #####  
sblv <- search_tweets(  
  "#SBLV" , n = 3000, lang = "en", include_rts = FALSE  
)
```

```
##### second library: sbhts #####
```

```
sbhts <- search_tweets(
  "#SuperBowl + #halftimeshow" , n = 3000, lang= "en", include_rts = FALSE
)
```

```
##### third library: sbcomm #####
sbcomm <- search_tweets(
  "#superbowlcommercials" , n = 3000, lang= "en", include_rts = FALSE
)
```

```
Downloading [=====] 100%
>
> ##### second library: sbhts #####
> sbhts <- search_tweets(
+   "#SuperBowl + #halftimeshow" , n = 3000, lang= "en", include_rts = FALSE
+ )
Downloading [=====] 100%
>
> ##### third library: sbcomm #####
> sbcomm <- search_tweets(
+   "#superbowlcommercials" , n = 3000, lang= "en", include_rts = FALSE
+ )
Downloading [=====] 100%
```

```
#####
# data cleaning part 1
#####
```

```
##### sblv #####
#Removing the URLs
sblv$stripped_text <- gsub("http.*", "", sblv$text)
```

```
#tokenizing: tidying the text
sblv_tweets_clean <- sblv %>%
  dplyr::select(stripped_text) %>%
  unnest_tokens(word, stripped_text)
```

```
##### halftimeshow #####
#Removing the URLs
sbhts$stripped_text <- gsub("http.*", "", sbhts$text)
```

```
#tokenizing: tidying the text
sbhts_tweets_clean <- sbhts %>%
  dplyr::select(stripped_text) %>%
  unnest_tokens(word, stripped_text)
```

```
##### super bowl commercial #####
#Removing the URLs
sbcomm$stripped_text <- gsub("http.*", "", sbcomm$text)
```

#tokenizing: tidying the text

```
sbcomm_tweets_clean <- sbcomm %>%
```

```
  dplyr::select(stripped_text) %>%
```

```
  unnest_tokens(word, stripped_text)
```

```
> ##### sblv #####
> #Removing the URLs
> sblv$stripped_text <- gsub("http.*", "", sblv$text)
>
> #tokenizing: tidying the text
> sblv_tweets_clean <- sblv %>%
+   dplyr::select(stripped_text) %>%
+   unnest_tokens(word, stripped_text)
>
> ##### halftimeshow #####
> #Removing the URLs
> sbhts$stripped_text <- gsub("http.*", "", sbhts$text)
>
> #tokenizing: tidying the text
> sbhts_tweets_clean <- sbhts %>%
+   dplyr::select(stripped_text) %>%
+   unnest_tokens(word, stripped_text)
>
> ##### super bowl commercial #####
> #Removing the URLs
> sbcomm$stripped_text <- gsub("http.*", "", sbcomm$text)
>
> #tokenizing: tidying the text
> sbcomm_tweets_clean <- sbcomm %>%
+   dplyr::select(stripped_text) %>%
```

#####

data cleaning part 2 and Workcloud

#####

first library

#my list of words to not include

```
unnecessary_words1 <- c("congratulations", "winning", "game", "team", "congrats", "fans",
"football", "season", "time", "watch", "victory", "2021", "superbowl", "bowl", "super",
"superbowlv", "sblv", "superbowlsunday", "superbowl2021", "city")
```

#creating a Wordcloud

```
sblv_tweets_clean %>%
```

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

```
  anti_join(stop_words) %>% #exclude stopwords
```

```
  filter(!word %in% unnecessary_words1) %>% #excluding the unnecessary words
```

```
  filter(nchar(word)>3) %>% #only including words that have at least 3 chr
```

```
  with(wordcloud(word, #add the word
```

```
    n, #word-count
```

```
    main= "Wordcloud for #SBLV",
```

```
    scale=c(4,0.5),
```

```
    use.r.layout=FALSE,
```

```
    max.words = 25, #limit the numer of words
```



```
colors=brewer.pal(8, "Set1")))
```

```
> #creating a Wordcloud
> sblv_tweets_clean %>%
+   count(word, sort=TRUE) %>%
+   anti_join(stop_words) %>% #exclude stopwords
+   filter(!word %in% unnecessary_words1) %>% #excluding the unnecessary words
+   filter(nchar(word)>3) %>% #only including words that have at least 3 chr
+   with(wordcloud(word, #add the word
+                   n, #word-count
+                   main= "Wordcloud for #SBLV",
+                   scale=c(4,0.5),
+                   use.r.layout=FALSE,
+                   max.words = 25, #limit the numer of words
+                   colors=brewer.pal(8, "Set1"))))
Joining, by = "word"
```



second library

#my list of words to not include

```
unnecessary_words2 <- c("superbowlads", "game", "commercials", "superbowlv", "sblv",
"superbowlsunday", "spot", "superbowl2021", "2021", "it's", "superbowl", "bowl", "super",
"superbowlhalftimeshow", "performing game", "love", "half", "performance", "people", "time",
"halftimeshow", "halftime", "weeknd", "superbowlweeknd", "commercial", "night", "watch")
```

#creating a Wordcloud

```
sbcomm_tweets_clean %>%
count(word, sort=TRUE) %>%
anti_join(stop_words) %>% #exclude stopwords
filter(!word %in% unnecessary_words2) %>% #excluding the unnecessary words
filter(nchar(word)>3) %>% #only including words that have at least 3 chr
with(wordcloud(word, #add the word
                n, #word-count
                scale=c(5,0.5),
                use.r.layout=FALSE,
                max.words = 25, #limit the numer of words
                colors=brewer.pal(8, "Set1")))
```

```

> #creating a Wordcloud
> sbcomm_tweets_clean %>%
+   count(word, sort=TRUE) %>%
+   anti_join(stop_words) %>% #exclude stopwords
+   filter(!word %in% unnecessary_words2) %>% #excluding the unnecessary words
+   filter(nchar(word)>3) %>% #only including words that have at least 3 chr
+   with(wordcloud(word, #add the word
+                   n, #word-count
+                   scale=c(5,0.5),
+                   use.r.layout=FALSE,
+                   max.words = 25, #limit the number of words
+                   colors=brewer.pal(8, "Set1"))))
Joining, by = "word"

```



```

#####
# create bi-grams
#####
data("stop_words")

```

#as we create bi-grams, we will notice unnecessary words again, therefore we decide to filter it again

```

unnecessary_words3 <- c("superbowlads", "game", "commercials", "superbowlv", "sblv",
"superbowlsunday", "spot", "superbowl2021", "2021", "it's", "superbowl", "bowl", "super",
"superbowlhalftimeshow", "performing game", "love", "half", "performance", "people", "time",
"halftimeshow", "halftime", "weeknd", "superbowlweeknd", "commercial", "night", "watch")

```

#creating the bigrams

```

sblv_paired_words <- sblv %>%
  dplyr::select(stripped_text) %>%
  unnest_tokens(paired_words, stripped_text, token="ngrams", n=2) #bigrams

```

#word-pair counts

```

sblv_paired_words %>%
  count(paired_words, sort=TRUE)

```

```

> #creating the bigrams
> sblv_paired_words <- sblv %>%
+   dplyr::select(stripped_text) %>%
+   unnest_tokens(paired_words, stripped_text, token="ngrams", n=2) #bigrams
>
> #word-pair counts
> sblv_paired_words %>%
+   count(paired_words, sort=TRUE)
# A tibble: 38,871 x 2
  paired_words      n
  <chr>          <int>
1 super bowl      450
2 of the          237
3 tom brady       187
4 in the          186
5 for the         156
6 gobucs sblv     150
7 on the          150
8 to the          142
9 last night      136
10 sblv superbowl 136
# ... with 38,861 more rows

```

#separating to filter

```

sblv_separated_words <- sblv_paired_words %>%
  separate(paired_words, c("word1", "word2"), sep = " ")

```

#exclude the stop words from n-grams

```

sblv_tweets_filter <- sblv_separated_words %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(nchar(word1)>3) %>%
  filter(!word1 %in% unnecessary_words3) %>%
  filter(!word2 %in% stop_words$word) %>%
  filter(!word2 %in% unnecessary_words4) %>%
  filter(nchar(word2)>3)

```

#view the filtered word pairs

```

sblv_word_counts <- sblv_tweets_filter %>%
  count(word1, word2, sort=TRUE)

```

#visualize bigrams

```

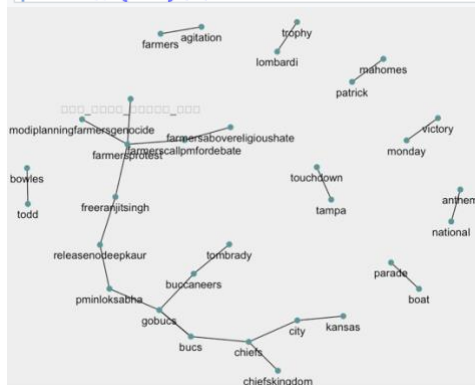
sblv_word_counts %>%
  filter(n>17) %>%
  graph_from_data_frame() %>%
  ggraph(layout = "fr") +
  geom_edge_link(aes()) +
  geom_node_point(color = "darkslategray4", size = 3) +
  geom_node_text(aes(label= name), vjust= 1.8, size= 5)

```

```

> #separating to filter
> sblv_separated_words <- sblv_paired_words %>%
+   separate(paired_words, c("word1", "word2"), sep = " ")
>
> #exclude the stop words from n-grams
> sblv_tweets_filter <- sblv_separated_words %>%
+   filter(!word1 %in% stop_words$word) %>%
+   filter(nchar(word1)>3) %>%
+   filter(!word1 %in% unnecessary_words3) %>%
+   filter(!word2 %in% stop_words$word) %>%
+   filter(!word2 %in% unnecessary_words4) %>%
+   filter(nchar(word2)>3)
>
> #view the filtered word pairs
> sblv_word_counts <- sblv_tweets_filter %>%
+   count(word1, word2, sort=TRUE)
>
> #visualize bigrams
> sblv_word_counts %>%
+   filter(n>17) %>%

```



```

#####
# explore Half Time Show most tweets words
#####

```

```

unnecessary_words4 <- c("superbowlv", "sblv", "superbowlsunday", "superbowl2021", "2021",
, "superbowl", "bowl", "super", "superbowlhalftimeshow", "performing game", "love", "half",
"performance", "people", "time", "halftimeshow", "halftime", "weeknd", "superbowlweeknd",
"pepsi", "abel", "weekend", "game")

```

```

#plotting the frequency plot for the words mentioned under the hashtag
sbhts_tweets_clean %>%
  count(word, sort=TRUE) %>%
  top_n(200) %>%
  anti_join(stop_words) %>%
  filter(!word %in% unnecessary_words4) %>% #excluding the unnecessary words
  filter(nchar(word)>3) %>% #only including words that have at least 3 chr
  mutate(word= reorder(word, n)) %>%
  ggplot(aes(x=word, y = n)) +
  geom_bar(stat= "identity",
    fill="blue") + #coloring the bars blue
  xlab(NULL) + #removing the x-axis label

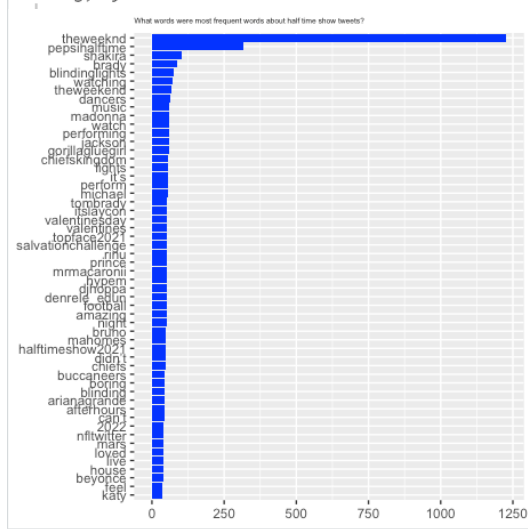
```

```
ylab(NULL) + #removing the y-axis label
ggtitle("What words were most frequent words about half time show tweets?")+
theme(plot.title = element_text(size=5))+ #changing the text size
coord_flip()
```

```
> #plotting the frequency plot for the words mentioned under the hashtag
> sbhts_tweets_clean %>%
+ count(word, sort=TRUE) %>%
+ top_n(200) %>%
+ anti_join(stop_words) %>%
+ filter(!word %in% unnecessary_words4) %>% #excluding the unnecessary words
+ filter(nchar(word)>3) %>% #only including words that have at least 3 chr
+ mutate(word= reorder(word, n)) %>%
+ ggplot(aes(x=word, y = n)) +
+ geom_bar(stat= "identity",
+ fill="blue") + #coloring the bars blue
+ xlab(NULL) + #removing the x-axis label
+ ylab(NULL) + #removing the y-axis label
+ ggtitle("What words were most frequent words about half time show tweets?")+
+ theme(plot.title = element_text(size=5))+ #changing the text size
+ coord_flip()
```

Selecting by n

Joining, by = "word"



```
#####
# explore audiences' feeling for Super Bowl this year
#####
```

```
unnecessary_words5 <- c("2021", "superbowl", "bowl", "super")
```

```
#sentiment: NRC
```

```
sblv_tweets_clean %>%
```

```
inner_join(get_sentiments("nrc")) %>% #using the nrc sentiment
```

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

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

```
distinct() %>%
```

```
filter(!word %in% unnecessary_words5) %>% #excluding the unnecessary words
```

```
filter(nchar(word)>3) %>%
```

```

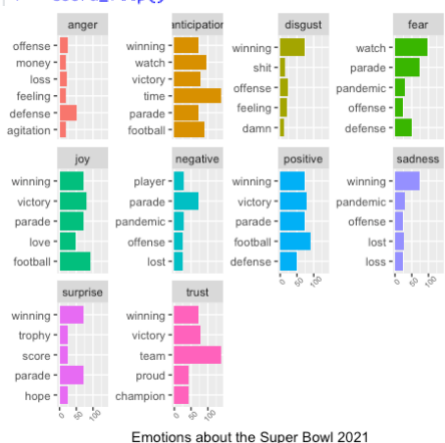
group_by(sentiment) %>%
top_n(5) %>% #showing the top 3 words for every emotion
ungroup() %>%
ggplot(aes(word, n, fill=sentiment))+
geom_col(show.legend = FALSE) + #hiding the legend
facet_wrap(~sentiment, scales="free_y") + #adding multiple graphs
xlab(NULL) +
ylab("Emotions about the Super Bowl 2021")+
theme(axis.text.x=element_text(size=6, angle=45)) +
coord_flip()

```

```

> #sentiment: NRC
> sblv_tweets_clean %>%
+   inner_join(get_sentiments("nrc")) %>% #using the nrc sentiment
+   count(word, sentiment, sort=TRUE) %>%
+   anti_join(stop_words) %>%
+   distinct() %>%
+   filter(!word %in% unnecessary_words5) %>% #excluding the unnecessary words
+   filter(nchar(word)>3) %>%
+   group_by(sentiment) %>%
+   top_n(5) %>% #showing the top 3 words for every emotion
+   ungroup() %>%
+   ggplot(aes(word, n, fill=sentiment))+
+   geom_col(show.legend = FALSE) + #hiding the legend
+   facet_wrap(~sentiment, scales="free_y") + #adding multiple graphs
+   xlab(NULL) +
+   ylab("Emotions about the Super Bowl 2021")+
+   theme(axis.text.x=element_text(size=6, angle=45)) +
+   coord_flip()

```



Emotions about the Super Bowl 2021