Final Project - NFL Tweets

Matt Rock

12/9/2019

The tweets collected will be used to see if there is a similar proportion of tweets supporting the Packers, Giants, or people just acknowledging the existince of this particular football game as there is retweets of each kind. We'll categorize each tweet, then use a chi-squared omega test to compare proportions of the tweets and retweets of each kind.

NFL Tweets During Gametime

```
packersgiantstweets = read.csv("C:/Users/Matt Rock/Documents/GitHub/ds710fall2019finalproject/packersgia
packersgiantsretweets = read.csv("C:/Users/Matt Rock/Documents/GitHub/ds710fall2019finalproject/packersg

#This is me regretting not grabbing the text for retweets. I'm purposely dropping any retweets of tweet
#that happened outside of my timeframe.

pgretweets = sqldf("select text, original_id, retweet_time, retweeter, retweeter_name
                   from packersgiantstweets
                   join packersgiantsretweets on packersgiantstweets.tweet_id = packersgiantsretweets.o
```

Text Analysis

Adding true/false columns for each search term. Using the above booleans to classify each tweet as interested in the Packers (by containing @packers or #GoPackGo), the Giants (@Giants or #GiantsPride) or neutral (either only contains #GBvsNYG or has both a Packer and Giant-supporting mention) Then, same analysis for retweets.

```
packersgiantstweets <- packersgiantstweets %>%
  mutate(GoPackGo = grepl("#GoPackGo", packersgiantstweets$text, ignore.case = TRUE))

packersgiantstweets <- packersgiantstweets %>%
  mutate(Packers = grepl("@Packers", packersgiantstweets$text, ignore.case = TRUE))

packersgiantstweets <- packersgiantstweets %>%
  mutate(GBvsNYG = grepl("#GBvsNYG", packersgiantstweets$text, ignore.case = TRUE))

packersgiantstweets <- packersgiantstweets %>%
  mutate(GiantsPride = grepl("#GiantsPride", packersgiantstweets$text, ignore.case = TRUE))

packersgiantstweets <- packersgiantstweets %>%
  mutate(Giants = grepl("@Giants", packersgiantstweets$text, ignore.case = TRUE))

packersgiantstweets <- packersgiantstweets %>%
  mutate(Association = case_when (
    (Giants | GiantsPride) & !(Packers | GoPackGo) ~ 'GiantsTweets',
    !(Giants | GiantsPride) & (Packers | GoPackGo) ~ 'PackersTweets',
    GBvsNYG | ((Giants | GiantsPride) & (Packers | GoPackGo)) ~ 'NeutralTweets',
```

```
    TRUE ~ 'ThisShouldNotExist') #this was extremely helpful in catching my own mistakes
  )

packersgiantstweets <- packersgiantstweets %>%
  mutate(Association = as.factor(Association))

pgretweets <- pgretweets %>%
  mutate(GoPackGo = grepl("#GoPackGo", pgretweets$text, ignore.case = TRUE))

pgretweets <- pgretweets %>%
  mutate(Packers = grepl("@Packers", pgretweets$text, ignore.case = TRUE))

pgretweets <- pgretweets %>%
  mutate(GBvsNYG = grepl("#GBvsNYG", pgretweets$text, ignore.case = TRUE))

pgretweets <- pgretweets %>%
  mutate(GiantsPride = grepl("#GiantsPride", pgretweets$text, ignore.case = TRUE))

pgretweets <- pgretweets %>%
  mutate(Giants = grepl("@Giants", pgretweets$text, ignore.case = TRUE))

pgretweets <- pgretweets %>%
  mutate(Association = case_when (
    (Giants | GiantsPride) & !(Packers | GoPackGo) ~ 'GiantsRetweets',
    !(Giants | GiantsPride) & (Packers | GoPackGo) ~ 'PackersRetweets',
    GBvsNYG | ((Giants | GiantsPride) & (Packers | GoPackGo)) ~ 'NeutralRetweets',
    TRUE ~ 'ThisShouldNotExist') #this was extremely helpful in catching my own mistakes
  )

pgretweets <- pgretweets %>%
  mutate(Association = as.factor(Association))
```

Combining tweets and retweets

```
tweets <- select(pgretweets, text, Association)
retweets <- select(packersgiantstweets, text, Association)

alltweets = rbind(tweets, retweets)
```

#Getting the number of each kind of tweet and retweet

```
table(alltweets$Association)
```

```
##
##  GiantsRetweets NeutralRetweets PackersRetweets    GiantsTweets
##            1662            1608           10787             379
##   NeutralTweets   PackersTweets
##            2366            4923
```

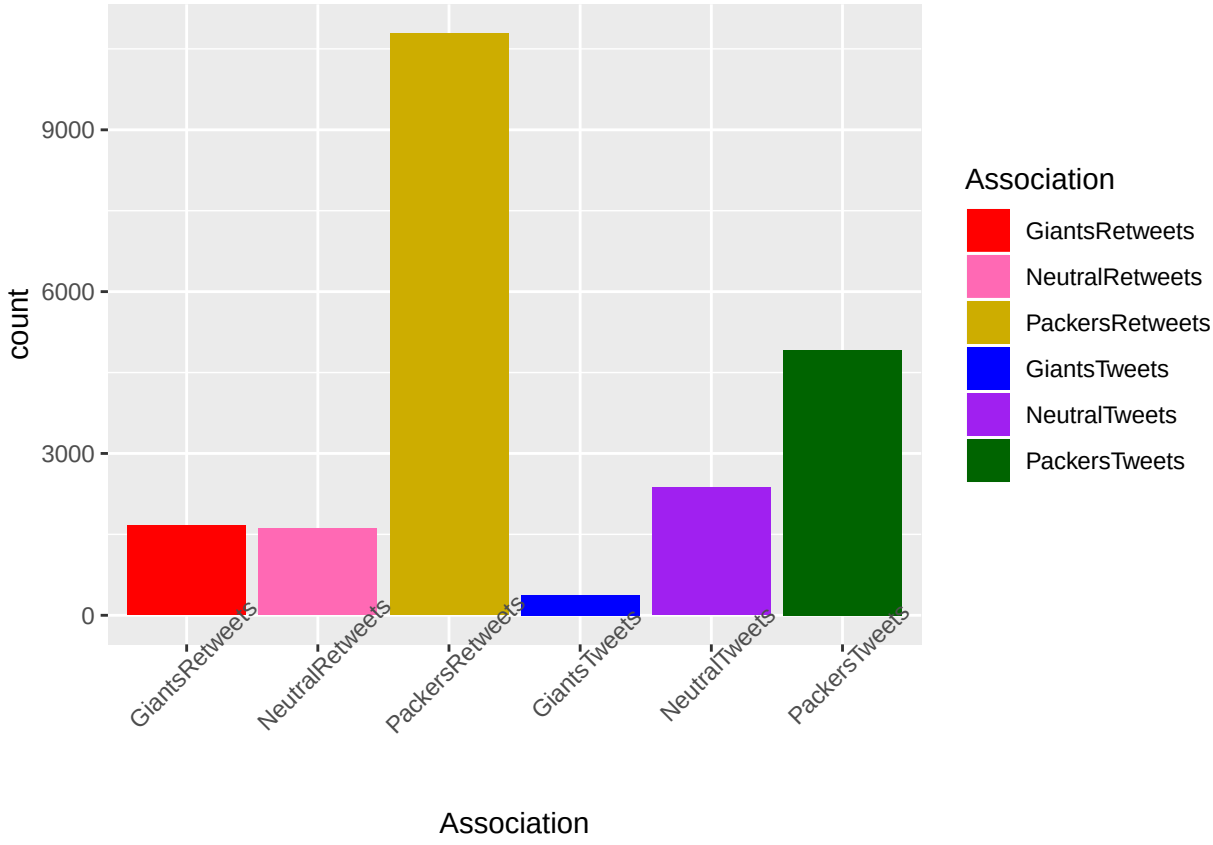
#Chi-Squared Omega Test Proportions of the tweets: Giants = .04943 Neutral = .30856 Packers = .64219

```r
tweetproportions = c(.04943, .30856, .64201)
retweetcount = c(1662, 1608, 10787)

chisq.test(retweetcount, p=tweetproportions)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  retweetcount
## X-squared = 3407.9, df = 2, p-value < 0.00000000000000022
```

#Creating a chart of number of each kind of tweets during the Packers - Giants game I am very happy I got 'dark green' because regular is the hideous neon.

```r
colorset = c('GiantsTweets' = 'blue', 'PackersTweets' = 'dark green', 'NeutralTweets' = 'purple',
             'GiantsRetweets' = 'red', 'PackersRetweets' = 'gold3', 'NeutralRetweets' = 'hotpink')

gf_bar( ~ Association, data = alltweets, fill =~Association) + scale_fill_manual(values=colorset) + ther
```



#Exporting samples of the dataframes

```r
write.csv(top_n(alltweets, 100), "RData - Combined Tweets-Retweets And Associations.csv")
```

```
## Selecting by Association
```

```
write.csv(top_n(pgretweets, 100), "RData - Tweets And Associations.csv")
```

```
## Selecting by Association
```

```
write.csv(top_n(packersgiantstweets, 100), "RData - RetweetsR And Associations.csv")
```

```
## Selecting by Association
```