Final Project–R

Kevin Ducat

2018-11-30

Load necessary libraries here.

```
#install.packages("data.table")
library(data.table)
#install.packages("readr")
library(readr)
#install.packages("syuzhet")
library(syuzhet)
```

Read the CSV

```
twitter_data <- as.data.frame(read_csv('Macys23.csv',locale = locale(encoding
= 'UTF-8'), col_names = c("tweet_time", "tweet_descr")))
## Parsed with column specification:
## cols(
##   tweet_time = col_datetime(format = ""),
##   tweet_descr = col_character()
## )
```

Summarize Data

```
#head(twitter_data)
```

Create a subset of the data that contains information about the “kiss” and the “prom” the musical.

```
twitter_data2 <-subset(twitter_data, tweet_descr %like% "kiss" |
tolower(tweet_descr) %like% "prom")
```

Clean the tweets

```
twitter_data2$tweet_descr <- gsub("http.*","",twitter_data2$tweet_descr)

twitter_data2$tweet_descr <- gsub("https.*","",twitter_data2$tweet_descr)
```

Get Sentiment

```
word.df <- as.vector(twitter_data2)

emotion.df <- get_nrc_sentiment(word.df$tweet_descr)

emotion.df2 <- cbind(twitter_data2,emotion.df)

####
sent.value <- get_sentiment(word.df$tweet_descr)

### Split the data into 3 categories, Positive, Negative, and Neutral.
category_sent <- ifelse(sent.value < 0, "Negative", ifelse(sent.value > 0,
"Positive", "Neutral"))

#Load the results into a Table.
table(category_sent)
## category_sent
## Negative  Neutral Positive
##      282      351     3358
#category_sent
#Negative  Neutral Positive
#     282      351      3358 #File23

neg_sent = 282
#print(neg_sent)

neutral_sent = 351
#print(neutral_sent)

pos_sent = 3358
#print(pos_sent)

total_sent =  pos_sent + neg_sent + neutral_sent
print(total_sent)
## [1] 3991
pos_pct  =    pos_sent     / total_sent
neg_pct  =    neg_sent     / total_sent
neutral_pct = neutral_sent / total_sent

#print(pos_pct)
#0.8413
#print(neg_pct)
#0.0706
#print(neutral_pct)
#0.0879
```

Pie Chart with Percentages

```
slices <- c(neg_sent, pos_sent, neutral_sent)

lbls <- c("Negative", "Positive", "Neutral")

pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col=rainbow(length(lbls)),
```

```
    main="Sentiment of Tweets about The Musical Prom during the Macys
Parade")
```

Perform T test.

```
t.test(sent.value > 0, sent.value < 0, mu = 0.67, alternative = "two.sided")
##
##  Welch Two Sample t-test
##
## data:  sent.value > 0 and sent.value < 0
## t = 14.26, df = 7151.2, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0.67
## 95 percent confidence interval:
##  0.7568861 0.7845822
## sample estimates:
##  mean of x  mean of y
## 0.84139313 0.07065898
#
#   Welch Two Sample t-test
#
#data:  sent.value > 0 and sent.value < 0
#t = 14.26, df = 7151.2, p-value < 2.2e-16
#alternative hypothesis: true difference in means is not equal to 0.67
#95 percent confidence interval:
# 0.7568861 0.7845822
#sample estimates:
# mean of x  mean of y
#0.84139313 0.07065898
#
```