IST 687
Homework 10
Due Date: 12/14

Code requires the following packages to run

```
library(tm) #install.packages('tm')
library(wordcloud) #install.pacakges('wordcloud')
library(stringr) #install.packages('stringr')
library(dplyr) #install.packages('dplyr')
library(ggplot2) #install.packages('ggplot2')
```

read in the AFINN word list

```
AFINNtext <- "AFINN-96.txt"
AFINNlist <- scan(AFINNtext, character(0), sep="\n")
AFINNlist <- str_split_fixed(AFINNlist, "\t", 2)
AFINNlist <- data.frame(AFINNlist)
```

compute the overall score for the MLK speech

```
#prepare the data
MLKtext <- "MLKspeech.txt"
MLKspeech <- scan(MLKtext, character(0), sep="\n")
words.vec <- VectorSource(MLKspeech)
words.corpus <- Corpus(words.vec)
words.corpus <- tm_map(words.corpus, content_transformer(tolower))
words.corpus <- tm_map(words.corpus, removePunctuation)
words.corpus <- tm_map(words.corpus, removeNumbers)
words.corpus <- tm_map(words.corpus, removeWords, stopwords("english"))
tdm <- TermDocumentMatrix(words.corpus)
m <- as.matrix(tdm)
wordCounts <- rowSums(m)
wordCounts <- sort(wordCounts, decreasing=TRUE)
totalWords <- sum(wordCounts)
words <- names(wordCounts)









#compute the overall score
MLKscore <- data.frame(wordCounts)
MLKscore$word <- rownames(MLKscore)
```

```
rownames(MLKscore) <- NULL
colnames(MLKscore) <- c("Count", "Word")
colnames(AFINNlist) <- c("Word", "Score")
MLKscore <- left_join(MLKscore, AFINNlist, on = "Word")
## Joining, by = "Word"

MLKscore$Score <- as.numeric(MLKscore$Score)
sum(MLKscore$Score, na.rm=TRUE)

## [1] 12
```

Compute the sentiment score for each quarter of the speech

```
#break the speech up into quarters
MLKspeechQ1 <- words.vec[1:7]
MLKspeechQ2 <- words.vec[8:15]
MLKspeechQ3 <- words.vec[16:23]
MLKspeechQ4 <- words.vec[24:29]

#create a function to calculate the score
SentimentScore <- function(text){
words.vec <- VectorSource(text)
words.corpus <- Corpus(words.vec)
words.corpus <- tm_map(words.corpus, content_transformer(tolower))
words.corpus <- tm_map(words.corpus, removePunctuation)
words.corpus <- tm_map(words.corpus, removeNumbers)
words.corpus <- tm_map(words.corpus, removeWords, stopwords("english"))
tdm <- TermDocumentMatrix(words.corpus)
m <- as.matrix(tdm)
wordCounts <- rowSums(m)
wordCounts <- sort(wordCounts, decreasing=TRUE)
totalWords <- sum(wordCounts)
words <- names(wordCounts)
MLKscore <- data.frame(wordCounts)
MLKscore$word <- rownames(MLKscore)
rownames(MLKscore) <- NULL
colnames(MLKscore) <- c("Count", "Word")
colnames(AFINNlist) <- c("Word", "Score")
MLKscore <- left_join(MLKscore, AFINNlist, on = "Word")
MLKscore$Score <- as.numeric(MLKscore$Score)
return(sum(MLKscore$Score, na.rm=TRUE))}


#Use function on each quarter of the speech
SentimentScore(MLKspeechQ1)

## [1] 10

SentimentScore(MLKspeechQ2)
```

```
## [1] -1

SentimentScore(MLKspeechQ3)

## [1] 12

SentimentScore(MLKspeechQ4)

## [1] 6
```

Plot the results via bar chart

```
Q1score <- SentimentScore(MLKspeechQ1)
Q2score <- SentimentScore(MLKspeechQ2)
Q3score <- SentimentScore(MLKspeechQ3)
Q4score <- SentimentScore(MLKspeechQ4)
ScoreChart <- data.frame(c(Q1score = Q1score, Q2score = Q2score, Q3score = Q3score, Q4score = Q4sc
ore))
ScoreChart$Quarter <- rownames(ScoreChart)
rownames(ScoreChart) <- NULL
colnames(ScoreChart) <- c("Score", "Quarter")
ggplot(ScoreChart, aes(x=Quarter, y=Score)) + geom_col()
```