# Bing-Jw_Wu_HW10

**Bing-Je Wu**

**6/9/2019**

# 1. First read in the AFINN word list.

Note that each line is both a word and a score(between -5 and 5). You will need to split the line and create two vectors (one for words and one for scores).

```
#- read the files -#
afinn_list <- read.delim(file='AFINN-en-165.txt',header=FALSE, stringsAsFactors=FALSE)
words <- afinn_list[,1]
head(words)
```

```
## [1] "abandon"    "abandoned"  "abandons"   "abducted"   "abduction"
## [6] "abductions"
```

```
scores <- afinn_list[,2]
head(scores)
```

```
## [1] -2 -2 -2 -2 -2 -2
```

# 2. Compute the overall score for the MLK speech using the AFINN word list

(as opposed to the positive and negative word lists).

```
#- Read a file by using readLine() function -#
MLKFile <- "MLK.txt"
MLK <- readLines(MLKFile)
head(MLK,3)
```

```
## [1] "    Delivered on the steps at the Lincoln Memorial in Washington D.C. on August
28, 1963"
## [2] "    Five score years ago, a great American, in whose symbolic shadow we stand si
gned the Emancipation Proclamation. This momentous decree came as a great beacon light o
f hope to millions of Negro slaves who had been seared in the flames of withering injust
ice. It came as a joyous daybreak to end the long night of captivity."
## [3] "    But one hundred years later, we must face the tragic fact that the Negro is
still not free. One hundred years later, the life of the Negro is still sadly crippled b
y the manacles of segregation and the chains of discrimination. One hundred years later,
the Negro lives on a lonely island of poverty in the midst of a vast ocean of material p
rosperity. One hundred years later, the Negro is still languishing in the corners of Ame
rican society and finds himself an exile in his own land. So we have come here today to
dramatize an appalling condition."
```

```
str(MLK)  #-- 31 elements in the MLK vector
```

```
##  chr [1:31] "    Delivered on the steps at the Lincoln Memorial in Washington D.C. on
August 28, 1963" ...
```

```
library(tm)
#- Interprets each element of the MLK vector as a document -#
MLK.vec <- VectorSource(MLK)
str(MLK.vec)  #-- 31 documents in the MLK.vec vector
```

```
## List of 31
##  $ encoding: chr "    Delivered on the steps at the Lincoln Memorial in Washington D.
C. on August 28, 1963"
##  $ length  : chr "    Five score years ago, a great American, in whose symbolic shado
w we stand signed the Emancipation Proclamat"| __truncated__
##  $ position: chr "    But one hundred years later, we must face the tragic fact that
the Negro is still not free. One hundred yea"| __truncated__
##  $ reader  : chr "    In a sense we have come to our nation's capital to cash a chec
k. When the architects of our republic wrote "| __truncated__
##  $ content : chr "    It is obvious today that America has defaulted on this promisso
ry note insofar as her citizens of color are"| __truncated__
##  $ NA: chr "    It would be fatal for the nation to overlook the urgency of the momen
t and to underestimate the determinati"| __truncated__
##  $ NA: chr "    But there is something that I must say to my people who stand on the
warm threshold which leads into the pa"| __truncated__
##  $ NA: chr "    We must forever conduct our struggle on the high plane of dignity and
discipline. We must not allow our cre"| __truncated__
##  $ NA: chr "    And as we walk, we must make the pledge that we shall march ahead. We
cannot turn back. There are those who"| __truncated__
##  $ NA: chr "    I am not unmindful that some of you have come here out of great trial
s and tribulations. Some of you have c"| __truncated__
##  $ NA: chr "    Go back to Mississippi, go back to Alabama, go back to Georgia, go ba
ck to Louisiana, go back to the slums "| __truncated__
##  $ NA: chr "    I say to you today, my friends, that in spite of the difficulties and
frustrations of the moment, I still h"| __truncated__
##  $ NA: chr "    I have a dream that one day this nation will rise up and live out the
true meaning of its creed: \"We hold "| __truncated__
##  $ NA: chr "    I have a dream that one day on the red hills of Georgia the sons of f
ormer slaves and the sons of former sl"| __truncated__
##  $ NA: chr "    I have a dream that one day even the state of Mississippi, a desert s
tate, sweltering with the heat of inju"| __truncated__
##  $ NA: chr "    I have a dream that my four children will one day live in a nation wh
ere they will not be judged by the col"| __truncated__
##  $ NA: chr "    I have a dream today."
##  $ NA: chr "    I have a dream that one day the state of Alabama, whose governor's li
ps are presently dripping with the wor"| __truncated__
##  $ NA: chr "    I have a dream today."
##  $ NA: chr "    I have a dream that one day every valley shall be exalted, every hill
and mountain shall be made low, the r"| __truncated__
##  $ NA: chr "    This is our hope. This is the faith with which I return to the South.
With this faith we will be able to he"| __truncated__
##  $ NA: chr "    This will be the day when all of God's children will be able to sing
with a new meaning, \"My country, 'tis"| __truncated__
##  $ NA: chr "    And if America is to be a great nation this must become true. So let
freedom ring from the prodigious hillt"| __truncated__
##  $ NA: chr "    Let freedom ring from the snowcapped Rockies of Colorado!"
##  $ NA: chr "    Let freedom ring from the curvaceous peaks of California!"
##  $ NA: chr "    But not only that; let freedom ring from Stone Mountain of Georgia!"
##  $ NA: chr "    Let freedom ring from Lookout Mountain of Tennessee!"
##  $ NA: chr "    Let freedom ring from every hill and every molehill of Mississippi. F
rom every mountainside, let freedom ring."
##  $ NA: chr "    When we let freedom ring, when we let it ring from every village and
every hamlet, from every state and eve"| __truncated__
```

```
##  $ NA: chr ""
##  $ NA: chr ""
##  - attr(*, "class")= chr [1:3] "VectorSource" "SimpleSource" "Source"
```

```
#- Build a corpus class of MLK.vec -#
MLK.corpus <- Corpus(MLK.vec)
#- Interface to apply transformation functions to corpus -#
MLK.corpus <- tm_map(MLK.corpus, content_transformer(tolower)) # lower caps
MLK.corpus <- tm_map(MLK.corpus, removePunctuation) # remove Punctuation
MLK.corpus <- tm_map(MLK.corpus, removeNumbers) # remove numbers
MLK.corpus <- tm_map(MLK.corpus, removeWords,stopwords("english")) # remove stop words
#- Create Term-document Matrix -#
MLK.tdm <- TermDocumentMatrix(MLK.corpus)
MLK.tdm
```

```
## <<TermDocumentMatrix (terms: 451, documents: 31)>>
## Non-/sparse entries: 659/13322
## Sparsity           : 95%
## Maximal term length: 14
## Weighting          : term frequency (tf)
```

```
str(MLK.tdm)
```

```
## List of 6
##  $ i       : int [1:659] 1 2 3 4 5 6 7 8 9 10 ...
##  $ j       : int [1:659] 1 1 1 1 1 1 2 2 2 2 ...
##  $ v       : num [1:659] 1 1 1 1 1 1 1 1 1 2 ...
##  $ nrow    : int 451
##  $ ncol    : int 31
##  $ dimnames:List of 2
##   ..$ Terms: chr [1:451] "august" "delivered" "lincoln" "memorial" ...
##   ..$ Docs : chr [1:31] "1" "2" "3" "4" ...
##  - attr(*, "class")= chr [1:2] "TermDocumentMatrix" "simple_triplet_matrix"
##  - attr(*, "weighting")= chr [1:2] "term frequency" "tf"
```

```
inspect(MLK.tdm)
```

```
## <<TermDocumentMatrix (terms: 451, documents: 31)>>
## Non-/sparse entries: 659/13322
## Sparsity           : 95%
## Maximal term length: 14
## Weighting          : term frequency (tf)
## Sample             :
##          Docs
## Terms     18 2 21 29 3 4 5 6 8 9
##   come     0 0  0  0 1 1 3 0 1 0
##   day      1 0  1  1 0 0 0 1 0 0
##   dream    1 0  0  0 0 0 0 0 0 0
##   every    0 0  0  4 0 1 0 0 0 0
##   freedom  0 0  1  1 0 0 1 1 2 0
##   let      0 0  0  2 0 0 0 0 0 0
##   negro    0 1  0  1 4 0 1 3 1 2
##   one      1 0  1  0 4 0 0 0 0 1
##   ring     0 0  0  2 0 0 0 0 0 0
##   will     2 0  4  2 0 0 1 5 0 2
```

```
#- Convert Term-document Matrix into Matrix for computation-#
MLK.m<- as.matrix(MLK.tdm)
str(MLK.m)
```

```
##  num [1:451, 1:31] 1 1 1 1 1 1 0 0 0 0 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ Terms: chr [1:451] "august" "delivered" "lincoln" "memorial" ...
##   ..$ Docs : chr [1:31] "1" "2" "3" "4" ...
```

```
head(MLK.m)
```

```
##             Docs
## Terms        1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
##   august     1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   delivered  1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   lincoln    1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   memorial   1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   steps      1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   washington 1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##             Docs
## Terms        24 25 26 27 28 29 30 31
##   august      0  0  0  0  0  0  0  0
##   delivered   0  0  0  0  0  0  0  0
##   lincoln     0  0  0  0  0  0  0  0
##   memorial    0  0  0  0  0  0  0  0
##   steps       0  0  0  0  0  0  0  0
##   washington  0  0  0  0  0  0  0  0
```

```
MLKwordCounts <- rowSums(MLK.m)
MLKwordCounts <- sort(MLKwordCounts,decreasing = T)
head(MLKwordCounts)
```

```
##    will freedom     let   negro     one    ring
##      26      19      14      13      12      12
```

```
#- create dataframe for afinn_df and MLKwordCounts_df
afinn_df<- data.frame(words, scores)
#- put matched-word counts into MLKwordCounts_df
#-- get a list of named indexing Vector
indexname <- names(MLKwordCounts)
#-- combine indexname into MLKwordCounts
MLKwordCounts_df <- data.frame(mwords=indexname, counts=MLKwordCounts)
rownames(MLKwordCounts_df) <- NULL
head(MLKwordCounts_df)
```

```
##    mwords counts
## 1    will     26
## 2 freedom     19
## 3     let     14
## 4   negro     13
## 5     one     12
## 6    ring     12
```

```
#- merge adinn_df and mWordslist
matchWords_df <- merge(afinn_df, MLKwordCounts_df ,by.x = "words", by.y="mwords")
#- compute the score and counts
matchWords_df$subtotal <- matchWords_df$scores*matchWords_df$counts
#- sum of the total score
sum(matchWords_df$subtotal)
```

```
## [1] 96
```

```
### The overall score for the MLK speech is 96
```

# 3. Compute the sentiment score for each quarter

Then, just as in class, compute the sentiment score for each quarter (25%) of the speech to see how this sentiment analysis is the same or different than what was computing with just the positive and negative word files. Note that since you will be doing almost the exact same thing 4 times (once for each quarter of the speech), you should create a function to do most of the work, and call it 4 times.

```
library(BurStMisc)
Temp <- ntile(MLK,4)
MLK1<- Temp[1]
MLK2<- Temp[2]
MLK3<- Temp[3]
MLK4<- Temp[4]

# create sentiment analysis function
sentiment <- function(text_vector){

  library(tm)
  #- Interprets each element of the text_vector as a document -#
  text_vector.vec <- VectorSource(text_vector)
  #- Build a corpus class of text_vector.vec -#
  text_vector.corpus <- Corpus(text_vector.vec)
  #- Interface to apply transformation functions to corpus -#
  text_vector.corpus <- tm_map(text_vector.corpus, content_transformer(tolower))
  text_vector.corpus <- tm_map(text_vector.corpus, removePunctuation)
  text_vector.corpus <- tm_map(text_vector.corpus, removeNumbers)
  text_vector.corpus <- tm_map(text_vector.corpus, removeWords,stopwords("english"))
  #- Create Term-document Matrix -#
  text_vector.tdm <- TermDocumentMatrix(text_vector.corpus)
  #- Convert Term-document Matrix into Matrix for computation -#
  text_vector.m<- as.matrix(text_vector.tdm)
  text_vector_wordCounts <- rowSums(text_vector.m)

  #- create dataframe for afinn_df and MLKwordCounts_df -#
  #- read the afinn word list files -#

  afinn_list <- read.delim(file='AFINN-en-165.txt',
                           header=FALSE, stringsAsFactors=FALSE)
  words <- afinn_list[,1]
  scores <- afinn_list[,2]
  afinn_df<- data.frame(words, scores)

  #- put matched-word counts into MLKwordCounts_df -#
  #-- get a list of named indexing Vector  --#
  indexname <- names(text_vector_wordCounts)
  #-- combine indexname into MLKwordCounts --#
  text_vector_wordCounts_df <-
    data.frame(mwords=indexname, counts=text_vector_wordCounts)
  rownames(text_vector_wordCounts_df) <- NULL

  #- merge adinn_df and MLKwordCounts_df -#
  matchWords_df <- merge(afinn_df, text_vector_wordCounts_df,
                         by.x = "words", by.y="mwords")
  #- compute the score and counts -#
  matchWords_df$subtotal <- matchWords_df$scores*matchWords_df$counts
  #- sum of the total score -#
  result <- sum(matchWords_df$subtotal)

  sprintf("The sentiment score: %s", result)
}
```

```
# First quarter sentiment score
sentiment(MLK1)
```

```
## [1] "The sentiment score: 29"
```

```
# Second quarter sentiment score
sentiment(MLK2)
```

```
## [1] "The sentiment score: 14"
```

```
# Third quarter sentiment score
sentiment(MLK3)
```

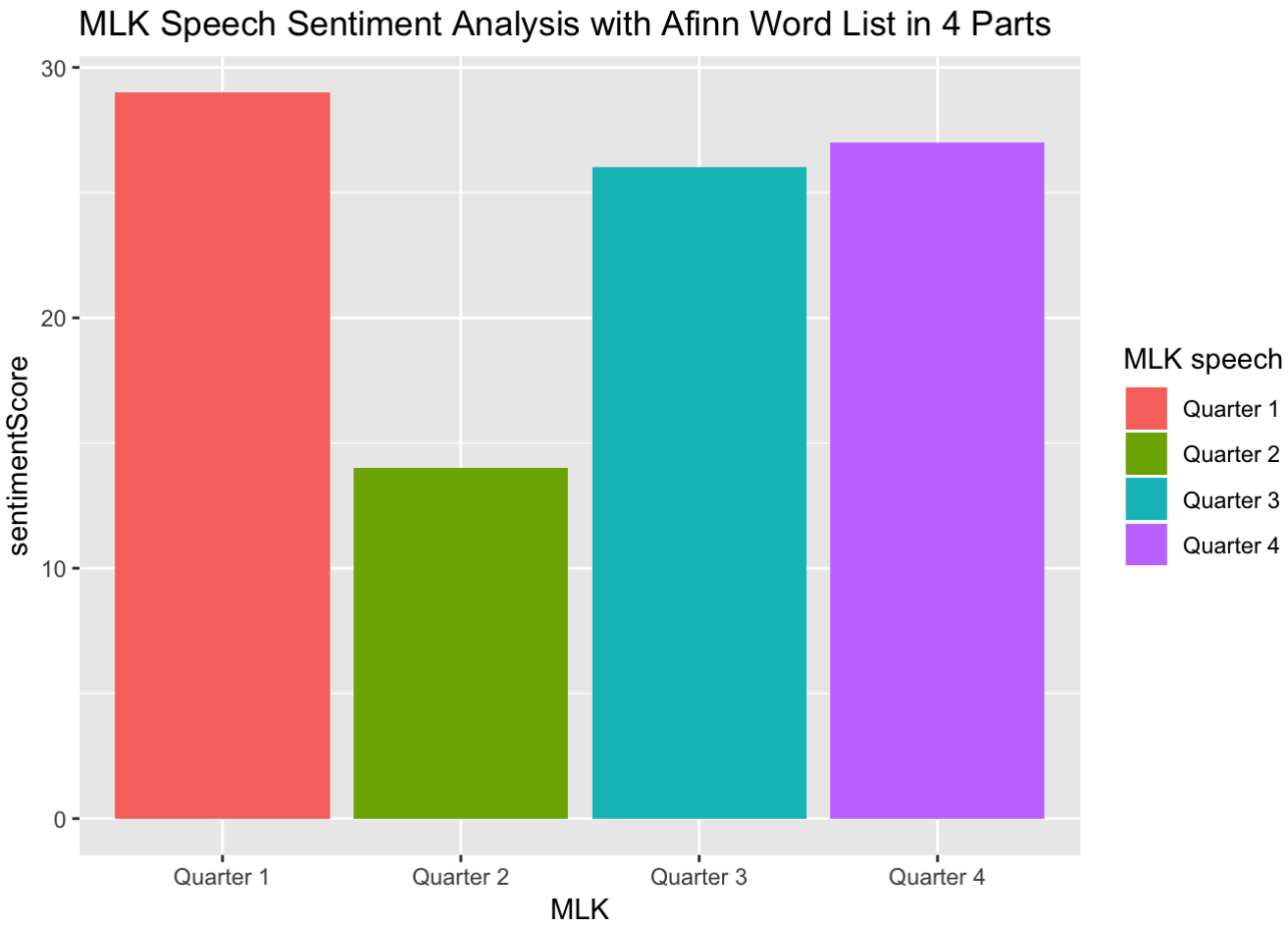
```
## [1] "The sentiment score: 26"
```

```
# Fourth quarter sentiment score
sentiment(MLK4)
```

```
## [1] "The sentiment score: 27"
```

# 4. Finally, plot the results (i.e, 4 numbers) via a bar chart.

```
sentimentplot <- data.frame(MLK=c("Quarter 1", "Quarter 2", "Quarter 3", "Quarter 4"),
                            sentimentScore = c(29,14,26,27))
library(ggplot2)
ggplot(sentimentplot, aes(x=MLK, y=sentimentScore)) +
  geom_bar(stat = "identity", aes(fill=as.factor(MLK))) +
  labs(fill = "MLK speech") +
  ggtitle("MLK Speech Sentiment Analysis with Afinn Word List in 4 Parts")
```


MLK Speech Sentiment Analysis with Afinn Word List in 4 Parts
30
20
10
0
sentimentScore
Quarter 1
Quarter 2
Quarter 3
Quarter 4
MLK
MLK speech
Quarter 1
Quarter 2
Quarter 3
Quarter 4