Sentiment Analysis

```
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")
```

```
text <- readLines("glassdoortest.csv")

TextDoc <- Corpus(VectorSource(text))
```

```
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
TextDoc <- tm_map(TextDoc, toSpace, "/")
TextDoc <- tm_map(TextDoc, toSpace, "@")
TextDoc <- tm_map(TextDoc, toSpace, "\\|")
```

```
TextDoc <- tm_map(TextDoc, content_transformer(tolower))
```

```
TextDoc <- tm_map(TextDoc, removeNumbers)
```

```
TextDoc <- tm_map(TextDoc, removeWords, stopwords("english"))

TextDoc <- tm_map(TextDoc, removeWords, c("s", "company", "team"))

TextDoc <- tm_map(TextDoc, removePunctuation)
TextDoc <- tm_map(TextDoc, stripWhitespace)
TextDoc <- tm_map(TextDoc, stemDocument)
```

```
TextDoc_dtm <- TermDocumentMatrix(TextDoc)
dtm_m <- as.matrix(TextDoc_dtm)
```

```
dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)
dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)
```

```
head(dtm_d, 5)
```

```
##            word freq
## work       work 1597
## manag     manag  772
## employe employe  702
## peopl     peopl  671
## good       good  592
```

```
wordcloud(words = dtm_d$word, freq = dtm_d$freq, min.freq = 1, max.words=50,random.order=FALSE, rot.per=
          colors=brewer.pal(8, "Dark2"))
```


```
#findAssocs(TextDoc_dtm, terms = c("exam","cbse","cancel"),
            #corlimit = 0.25)
#findAssocs(TextDoc_dtm, terms = findFreqTerms(TextDoc_dtm,
             #                                lowfreq = 6),
            #corlimit = 0.25)
```

```
syuzhet_vector <- get_sentiment(text, method="syuzhet")
head(syuzhet_vector)
```

```
## [1] 0.00 4.30 2.65 1.50 3.90 3.70
```

```
summary(syuzhet_vector)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -4.100   1.050   2.050   2.295   3.200  22.550
```

```
bing_vector <- get_sentiment(text, method="bing")
head(bing_vector)
```

```
## [1] 0 5 2 2 1 2
```

```
afinn_vector <- get_sentiment(text, method="afinn")
head(afinn_vector)
```

```
## [1] 0 7 5 5 2 4
```

```
summary(afinn_vector)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -22.000   1.000   4.000   4.108   7.000  32.000
```

Including Plots

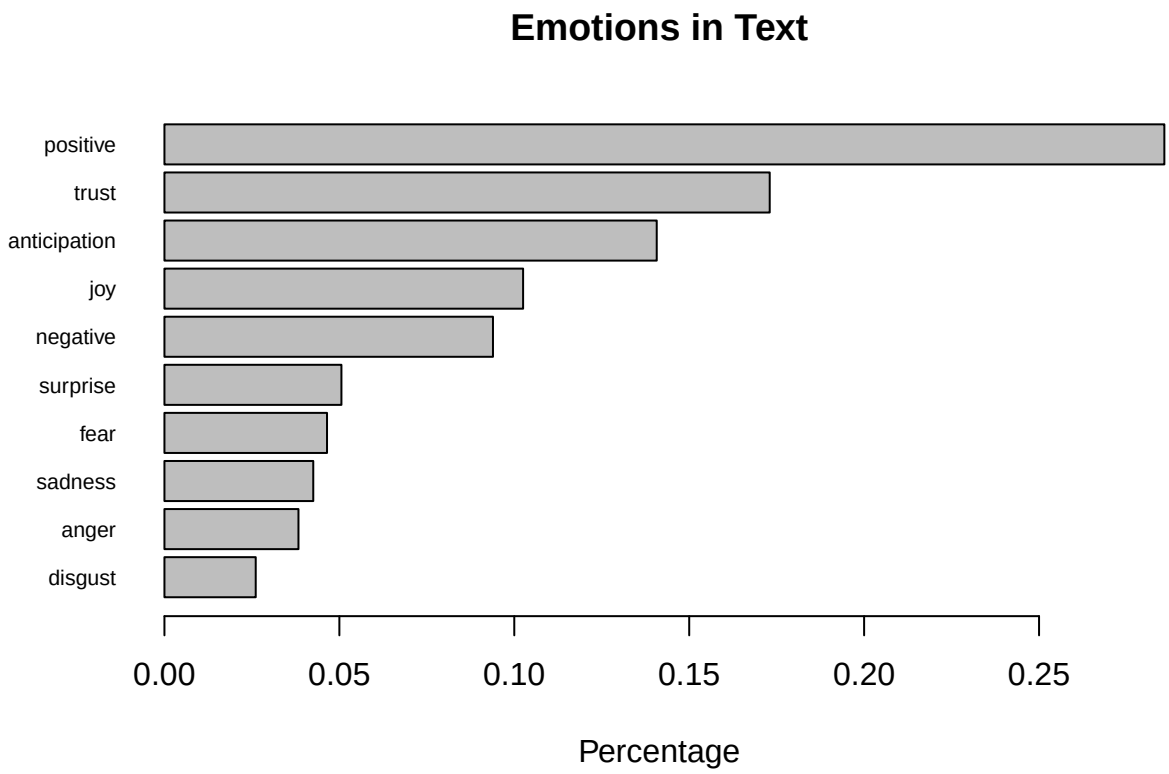
You can also embed plots, for example:

```
d<-get_nrc_sentiment(text)
head (d,10)
```

```
##    anger anticipation disgust fear joy sadness surprise trust negative positive
## 1      0            0       0    0   0       0        0     1        0        1
## 2      0            3       0    0   3       0        1     2        0        4
## 3      0            1       0    0   1       1        0     2        1        3
## 4      0            0       0    0   0       0        0     0        0        2
## 5      0            0       0    0   0       0        0     2        0        3
## 6      0            3       0    0   2       0        1     4        0        7
## 7      0            0       0    0   0       0        0     1        0        1
## 8      0            1       0    0   1       0        1     1        0        1
## 9      0            0       0    0   0       1        1     1        1        3
## 10     0            1       0    0   2       0        0     2        0        2
```

```
td<-data.frame(t(d))
```

```
barplot(
  sort(colSums(prop.table(d))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions in Text", xlab="Percentage"
)
```


Emotions in Text
positive
trust
anticipation
joy
negative
surprise
fear
sadness
anger
disgust
0.00
0.05
0.10
0.15
0.20
0.25
Percentage