## Challenge 2

Problems created by Pablo Barbera, Dan Cervone « Text Analysis Module

Write commands that help you answer the following questions about the bullying dataset.

1. What are the most popular words in the dataset, other than bullying words?

```
## Loading required package: NLP

df.tweets <- read.csv("bullying.csv", header = TRUE, stringsAsFactors = FALSE)

corpus <- VCorpus(VectorSource(df.tweets$text))
corpus <- tm_map(corpus, content_transformer(tolower))
corpus <- tm_map(corpus, removeWords, stopwords("english"))
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, stripWhitespace)
corpus <- tm_map(corpus, stemDocument)
dtm <- DocumentTermMatrix(corpus)
findFreqTerms(dtm, 50)</pre>
```

2. Create a wordcloud comparison between bullying tweets and non-bullying tweets.

```
# Identify posts with and without bullying traces and create large documents
no_bullying <- paste(df.tweets$text[df.tweets$bullying_traces=="n"], collapse=" ")</pre>
yes_bullying <- paste(df.tweets$text[df.tweets$bullying_traces=="y"], collapse=" ")</pre>
# Create DTM and preprocess
groups <- VCorpus(VectorSource(c("No bullying" = no_bullying, "Yes bullying" = yes_bullying)))</pre>
groups <- tm_map(groups, content_transformer(tolower))</pre>
groups <- tm_map(groups, removePunctuation)</pre>
groups <- tm map(groups, stripWhitespace)</pre>
dtm <- DocumentTermMatrix(groups)</pre>
## Label the two groups
dtm$dimnames$Docs = c("No bullying", "Yes bullying")
## Transpose matrix so that we can use it with comparison.cloud
tdm <- t(dtm)
## Compute TF-IDF transformation
tdm <- as.matrix(weightTfIdf(tdm))</pre>
library(wordcloud)
## Loading required package: RColorBrewer
comparison.cloud(tdm, max.words=100, colors=c("red", "blue"))
```

## No bullying

```
pessoa escola fingirfazer
sofrer estava agora sofreu
naosolteiros mas plos
delbeatdown
brincando meu uma pra oquem pode
quando cara
sua hoje COMISSO para vou
porque frio por não contra fazendo
assim o vai quem pode pessoas
sua hoje COMISSO para vou
porque frio por não contra fazendo
assim o vai quem pode pessoas
sua hoje COMISSO para vou
porque frio por não contra fazendo
assim o vai quem pode pessoas
sor sobre jogar
comigo
darryn sofre vai comigo
gente se shelly jeff helped huh
po esta shelly jeff helped perry private
spending hoes cobsbigorother door
refuse punk kept bb13 keeps mexican
talked wanted asked deleted should be sho
```

## Yes bullying