

```
library(tm)
library(topicmodels)

text <- readLines( "/Users/Shariq/Desktop/tweets.csv")

doc.vec <- VectorSource(text)
doc.corpus <- Corpus(doc.vec)
doc.corpus <- tm_map(doc.corpus, function(x) iconv(enc2utf8(x$content), sub = "byte"))
doc.corpus <- tm_map(doc.corpus, PlainTextDocument)
doc.corpus <- tm_map(doc.corpus, content_transformer(tolower))
doc.corpus <- tm_map(doc.corpus, removeWords, stopwords('english'))
doc.corpus <- tm_map(doc.corpus, removePunctuation)
doc.corpus <- tm_map(doc.corpus, removeNumbers)
doc.corpus <- tm_map(doc.corpus, stripWhitespace)

dtm <- DocumentTermMatrix(doc.corpus)
dtm <- removeSparseTerms(dtm, 0.98)
x <- as.matrix(dtm)
x <- x[which(rowSums(x) > 0),]
rownames(x) <- 1:nrow(x)
lda <- LDA(x, 5)
```