

## R Code for my Word Cloud

```
Ads.Lang

library(ggplot2)
library(grid)

names(Ads.Lang) <- c("AdID", "Title")
Ads.Lang <- Ads.Lang[Ads.Lang$Country == "USA", ]

library(tm)
library(NLP)

Ads.Title <- paste(Ads.Lang$Title, collapse=" ")
Ads.Title <- VectorSource(Ads.Title)
myCorpus <- Corpus(Ads.Title)

myCorpus <- tm_map(myCorpus, tolower)
myCorpus <- tm_map(myCorpus, removePunctuation)
myCorpus <- tm_map(myCorpus, removeNumbers)
myCorpus <- tm_map(myCorpus, stripWhitespace)
myCorpus <- tm_map(myCorpus, gsub, pattern = "american", replacement = "america")
myCorpus <- tm_map(myCorpus, gsub, pattern = "americas", replacement = "america")
myCorpus <- tm_map(myCorpus, gsub, pattern = "youll", replacement = "you")
#myCorpus <- tm_map(myCorpus, gsub, pattern = "your", replacement = "you")

## [...]

myStopwords <- c(stopwords("english"), "often", "even", "got", "wouldnt", "theres",
"vermont", "chicagos", "heres", "dont", "cant", "says", "santa", "every", "whatever",
"something", "one", "two", "will", "set")
## leave out to get results without "you"
myStopwords <- setdiff(myStopwords, c("i", "our", "us", "we", "you", "your", "they"))
myCorpus <- tm_map(myCorpus, removeWords, myStopwords)
myCorpus <- tm_map(myCorpus, PlainTextDocument)

docTermMatrix <- DocumentTermMatrix(myCorpus)
dtm <- DocumentTermMatrix(myCorpus)
dtm2 <- as.matrix(dtm)
frequency <- colSums(dtm2)
frequency <- sort(frequency, decreasing=TRUE)
library(RColorBrewer)
pal2 <- brewer.pal(8, "Dark2")

## for phrase use [1:50], for title/slogan use [1:100]
library(wordcloud)
words <- names(frequency)
wordcloud(words[1:50], frequency[1:50], colors = pal2)
```