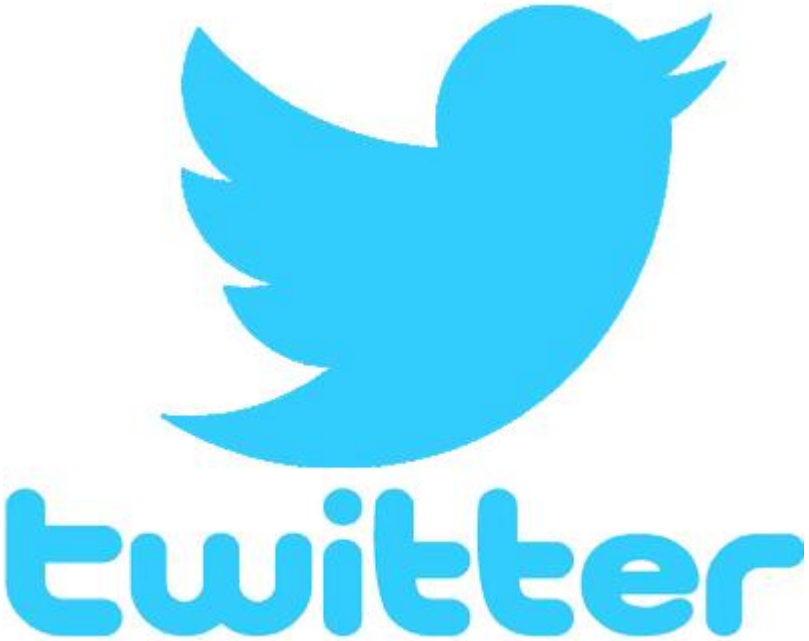
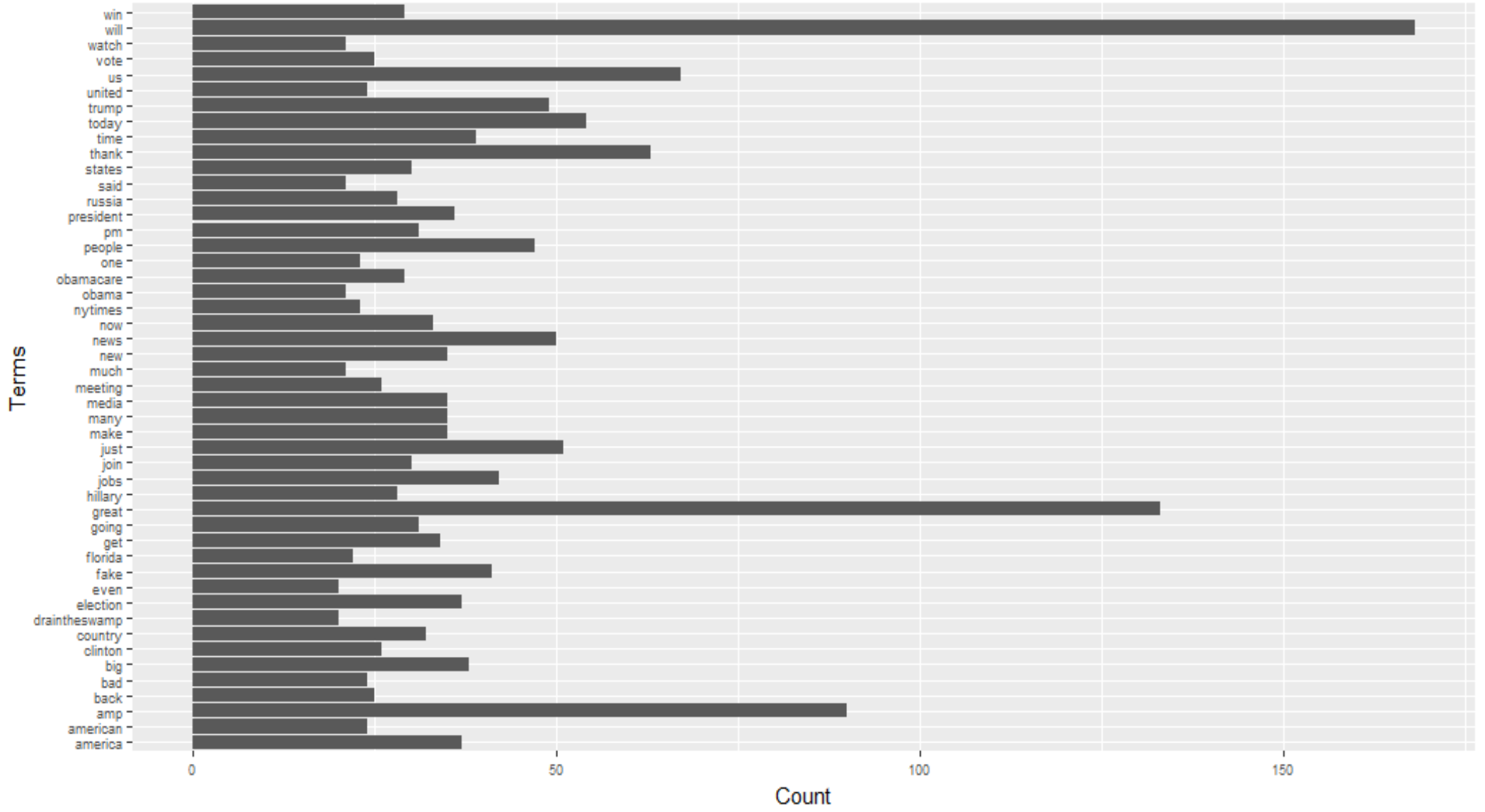

TEXT MINING
DATASETS
ANALYTICS
PROVIDER
FRAMEWORK
TEXTUAL
ENTITY
MANAGEMENT
CUSTOMER
RETRIEVAL
TEXT
LARGE
PROCESS
SOFTWARE
APPLICATION
INFORMATION
STATISTICAL
LITERATURE
SCIENCES
COMPUTATIONAL
ABSTRACTED
SET
ATTITUDINAL
PREDICTIVE
AFFINITIES
NATURAL
AFFECTIVITY
CLUSTER
CATEGORIZATION
SUMMARIZATION
SYNTACTIC
NEWS
SEMANTIC
TOOL
MINING
SPECIFIC
DATA
EXTRACTION
SENTIMENT
VISUALIZATION
MEDIA
LANGUAGE
BUSINESS
SUITE
PARSING
LINGUISTIC
SOCIAL
ANNOTATION
LEARNING
DOCUMENT
TECHNIQUES
INTELLIGENCE
CONTENT


twitter

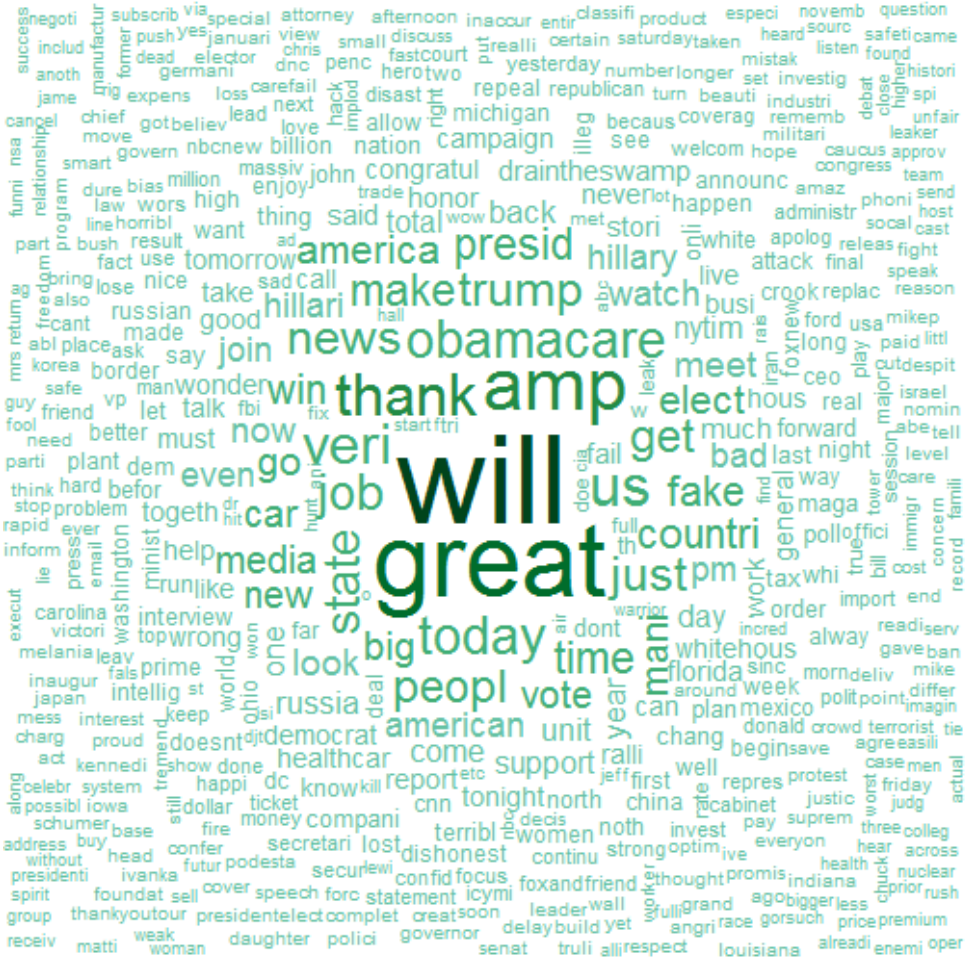
Text Mining in R *(Twitter Analysis)*

- Text mining on Donald Trump's Twitter account
- Freq words: vote, thank, maga, Obamacare, Hillary, nytimes, jobs, great, florida, drain the swamp, fake, election, Russia
- Wordcount :
  - Hillary Clinton : 60
  - Obamacare : 33
  - Jobs : 34
  - Russia : 35
  - Drain the swamp : 20

Frequency of the words



Word Cloud in R



Background in R

```r
library(twitteR)
library(ROAuth)

api_key <- "[redacted]"
api_secret <- "[redacted]"
access_token <- "[redacted]"
access_token_secret <- "[redacted]"

setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)

tweets <- userTimeline("realDonaldTrump", n = 100)
n.tweet <- length(tweets)

# convert tweets to a data frame
tweets.df <- twListToDF(tweets)

# tweet #190
tweets.df[10, c("id", "created", "screenName", "replyToSN","favoriteCount", "retweetCount", "longitude", "latitu

# print tweet #190 and make text fit for slide width
writeLines(strwrap(tweets.df$text[10], 60))

library(tm)
# build a corpus, and specify the source to be character vectors
myCorpus <- Corpus(VectorSource(tweets.df$text))


# remove URLs
removeURL <- function(x) gsub("http[^[:space:]]*", "", x)
myCorpus <- tm_map(myCorpus, content_transformer(removeURL))

# remove anything other than English letters or space
removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]*", "", x)
myCorpus <- tm_map(myCorpus, content_transformer(removeNumPunct))

# convert to lower case
myCorpus <- tm_map(myCorpus, content_transformer(tolower))

# remove stopwords
myStopwords <- c(setdiff(stopwords('english'), c("r", "big")),"use", "see", "used", "via", "amp")
myCorpus <- tm_map(myCorpus, removeWords, myStopwords)
# remove extra whitespace
myCorpus <- tm_map(myCorpus, stripWhitespace)

# keep a copy for stem completion later
myCorpusCopy <- myCorpus


myCorpus <- tm_map(myCorpus, stemDocument) # stem words
writeLines(strwrap(myCorpus[[190]]$content, 60))

stemCompletion2 <- function(x, dictionary) {
  x <- unlist(strsplit(as.character(x), " "))
  x <- x[x != ""]
  x <- stemCompletion(x, dictionary=dictionary)
  x <- paste(x, sep="", collapse=" ")
```

92:37 (Top Level)