

# Sentiment analysis - Tweets

*Kristian Wichmann*

*6 jan 2017*

## Load tweets

```
tweets <- read.csv("potus_tweets.csv", fileEncoding="UTF-8")
```

## Load AFINN-111 dataset

```
afinn <- read.csv("AFINN-111.txt", sep = "\t", header = FALSE)
colnames(afinn) <- c("word", "score")
```

## Sentiment of word

```
sentiment <- function(word) {
  lookup <- grep(paste0("^", word, "$"), afinn$word)
  if (length(lookup) == 0) {
    return(0)
  }
  afinn[lookup,]$score[1]
}
```

## Sentiment analysis

```
sent_analyze <- function(tweet) {
  text <- gsub('[\".,!?;()]', '', tweet)
  text <- tolower(text)
  words <- strsplit(text, ' ')[[1]]
  sentiments <- sapply(words, sentiment)
  sum(sentiments)
}
```

```
tweets$sentiment <- sapply(tweets$text, sent_analyze)
```

## Histogram

```
hist(tweets$sentiment,
     main = "Sentiment distribution for POTUS tweets",
     xlab = "Sentiment",
     col = "blue")
```

**Sentiment distribution for POTUS tweets**

