

# COVID-19 Sentiment Analysis

Richard Yim

5/13/2020

## Reading in Cleaned Data and Basic Reformatting

```
headlines <- read.csv("data/news_clean.csv")

# Names of variables
names(headlines)

## [1] "id"          "timestamp"   "source"      "title"       "description"
## [6] "date"

# Load all of the data into the working environment
attach(headlines)
# We ignore the id, timestamp and source, and focus on the title, description and date
headlines <- data.frame(title, description, date)
# Reformat and separate year month and day
headlines$month = substr(date, 6, 7)
headlines$day <- substr(date, 9, 10)
attach(headlines)

## The following objects are masked from headlines (pos = 3):
##
##   date, description, title

We want to check how many days there are in the dataset:

begin <- headlines$date[1]
end <- headlines$date[length(headlines$date)]
as.Date(end) - as.Date(begin)

## Time difference of 52 days

cat(begin, end)

## 1 53
```

## Frequency of Strings

```
library("ggplot2") #beautiful plots
library("reshape2") #melt-function

plot_word_in_day <- function(phrase){
  freq <- NULL
  days <- unique(headlines$date)
```

```

    for(day in days)
    {
        matching_day <- headlines$title[headlines$date == day]
        freq[length(freq)+1] <- sum(grepl(phrase, matching_day, ignore.case = FALSE))
    }
    return(freq)
}

list <- c("recession", "downturn", "covid-19", "trump")
time <- data.frame(seq(as.Date("2020-03-21"),by="day",length.out=53))

# turn off warnings temporarily
options(warn=-1)
for(i in list)
{
    val <- plot_word_in_day(i)
    time <- cbind(time, val)
}
options(warn = 0)

colnames(time) <- c("date", list)
charts <- melt(time, id="date")
names(charts) <- c('x', 'func', 'value')

g=ggplot() +
  geom_line(data = charts, aes(x = x, y = value, color = func),size=1) +
  xlab("year") +
  ylab("frequency")
g+scale_color_manual(values=c("#FF0000", "#FF00D0", "#0007FF", "#09FF00"))

```

