Part One - Data Collection

Samad Patel

9/29/2018

Summary

This document involves the process of collecting data regarding Trump's followers.

Libraries

```r
# Load relevant libraries
library(rvest)
library(dplyr)
library(zoo)
library(stringr)
library(lubridate)
library(ggplot2)
library(ggthemes)
library(readr)
library(kableExtra)
```

Scraping

```r
### Load the data into a dataframe ###
url <- 'https://www.trackalytics.com/twitter/profile/realdonaldtrump/'
page <- read_html(url)
data <- page %>% html_nodes('#summary') %>% html_table(fill = T)
df <- data[[1]]
```

Basic Cleaning

```r
# Column names are the first row; remove first column
colnames(df) <- df[1,]
df <- df[-1,]

# Extract number of followers for each date
followers <- str_extract_all(string = df$`Followers  (change)`, pattern = '.*\\s', simplify = T) %>%
   trimws()
# Remove commas
followers <- str_replace_all(followers, ',', '') %>% as.numeric()

# Extract change for each date
follower_change <- str_extract_all(string = df$`Followers  (change)`, pattern = '(?<=\\().*',
                       simplify = T) %>%
   str_remove_all(pattern = '\\)') %>% # Replace end parenthesis
   str_remove_all(pattern = '\\+') %>% # Replace + sign
```

```r
    str_remove_all(pattern = ',') %>%  # Replace commas
    as.numeric()

# Extract number of tweets per day
tweets_that_day <- str_extract_all(string = df$`Tweets  (change)`, pattern = '(?<=\\().*',
                  simplify = T) %>%
    str_remove_all(pattern = '\\)') %>% # Replace end parenthesis
    str_remove_all(pattern = '\\+') %>% as.numeric() # Replace + sign

# Make Date a datetime object
date <- mdy(df$Date)

# Place into df
follows <- tibble(date, followers, follower_change, tweets_that_day)
colnames(follows) <- c('Date', 'Followers', 'Follower_Change', 'Num_Tweets')

# Print
kable(head(follows)) %>% kable_styling(position = 'center', 'striped', row_label_position = 'c')
```

Date	Followers	Follower_Change	Num_Tweets
2018-09-29	54722152	23589	1
2018-09-28	54698563	26089	1
2018-09-27	54672474	21128	12
2018-09-26	54651346	20751	13
2018-09-25	54630595	17609	12
2018-09-24	54612986	17292	3