607 Project 3

Katherine Evers

3/18/2019

```r
library("tidyverse")
library("rvest")
library("stringi")
library("xml2")
library("kableExtra")
library(RCurl)
library(plyr)
library(RColorBrewer)
library(dplyr)
library(ggplot2)
library(tm)
library(wordcloud)
library(tidytext)
library(xtable)
library(readr)
library(tidytext)
library(knitr)
library(phrasemachine)
library(quanteda)
library(tidyr)
library(scales)
library(forcats)
```

New final web scraping code

Extract location url

```r
#Import url (indeed search results for full time data sceintist positions)
url <- "https://www.indeed.com/jobs?q=data+scientist&jt=fulltime"
page <- read_html(url)

#Extract urls from left side of page
location <- page %>%
  html_nodes("li") %>%
  html_nodes(xpath = '//*[@rel="nofollow"]') %>%
  html_attr("href")

#Extract top 5 location urls based on indexes
# location2 <- location[c(8:12)]
location2 <- location[c(8:21)]    ### All locations
```