Assignment 2

Promothesh Chatterjee*

Webscraping in R

1) Go to website https://fourminutebooks.com/nudge-summary/. This a brief summary of the famous behavioral book by Thaler (Nobel Laureate) and Cass Sunstein.

Use the rvest package to scrape different elements (paragraphs, headings etc.) of the webpage. Create a wordcloud from all the scraped text.

2) Use Developer Tools in Chrome (using xpath) to scrape the table of top one hundred surnames in the U.S. from the Wikipedia website below:

https://en.wikipedia.org/wiki/List_of_most_common_surnames_in_North_America

3) Go to the following website https://www.carfax.com/Used-Cars-in-Salt-Lake-City-UT_c12314

Use the Selector Gadget to scrape the following things from the website.

1) Car Name
2) Car Prices
3) Car Mileage

Create a tibble of the above variables.

Do not worry about data wrangling of the dataset at this point. That's an exercise for later, right now our interest is to get the data from the webpage.

Using APIs in R

Use HTTR

Go to https://superheroapi.com/

1) Get access token

2) Read the documentation to figure out how to get data regarding Spiderman's work, appearance, name

3) Optional Question: Create an iterative process to get the same info for multiple characters of interest, and combine them in a dataframe.