Web Scraping II

Recap of Web Scraping I

- Final 3 Data Frames From Last Lecture Should All Be Saved to CSV's on PC
  - FINAL_VIOLENT.CSV
  - FINAL_ZIP.CSV
  - FINAL_STATE_ABBREV.CSV
- Think About What Other City Information Could Potentially Be a Factor in Violent Crimes
- Think About What Other City Information Could Potentially Be Influenced by the Prevalence of Violent Crimes

Supplement Introduction

- Step 1: Open Supplement
- Step 2: Unzip and Open .Rmd
- Step 2: Ensure You Have the Following R Packages Installed
  - tidyverse
  - rvest (Requires Internet)
- Step 3: Switch Knitter
- Step 4: Read the Introduction

Part 1: Connection to Population Change and Density

- Step 1: Select the Link and Observe the Following Table

Rank	Name	State	2020 Pop ▾	2010 Census	Change	Density (km²)	Area (km²)
1	New York	New York	8,323,340	8,190,360	1.62%	10,699	778
2	Los Angeles	California	4,015,940	3,795,370	5.81%	3,306	1,215
3	Chicago	Illinois	2,694,240	2,697,530	-0.12%	4,575	589
4	Houston	Texas	2,340,890	2,098,450	11.55%	1,412	1,658
5	Phoenix	Arizona	1,703,080	1,449,300	17.51%	1,270	1,341
6	Philadelphia	Pennsylvania	1,591,800	1,528,290	4.16%	4,577	348
7	San Antonio	Texas	1,578,030	1,332,880	18.39%	1,256	1,256
8	San Diego	California	1,447,100	1,305,970	10.81%	1,715	844
9	Dallas	Texas	1,382,270	1,200,370	15.15%	1,571	880
10	San Jose	California	1,033,670	954,492	8.30%	2,245	461
11	Austin	Texas	988,218	806,423	22.54%	1,193	829

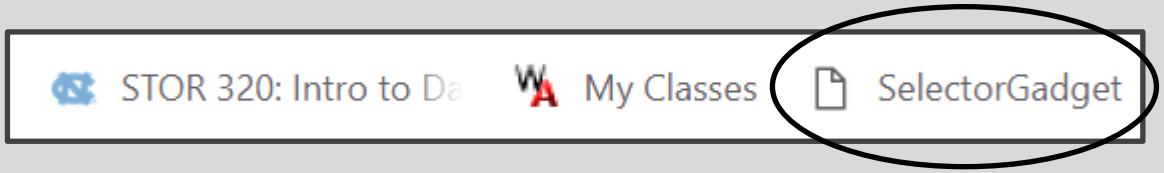
- Step 2: Questions?
  - What is the Connection to Violent Crimes?
  - How is this Useful When Related to Violent Crimes?

Part 1: Connection to Population Change and Density

- Step 3: Run Chunk 1
  - What is required to convert the Percentage Change to a numeric variable?
  - What is required to convert the 2019 Density to a numeric variable?
- Step 4: Run Chunk 2
- Step 5: No-Knitter

Part 2: Inclusion of Expert Opinion

- Step 1: Selector Gadget Website
  - Open Source
  - Chrome Extension Exists
  - Easy: Drag Link to Bookmark Bar as Webpage Explains



- Step 2: Observe the Article on 2018's Safest and Most Dangerous States
  - What info could be of use?
  - Do you agree identification?

Part 2: Inclusion of Expert Opinion

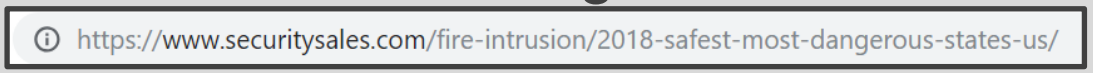
- Step 3: Information of Interest
  - Safe vs Dangerous

1. Vermont	1. Mississippi
2. Maine	2. Louisiana
3. Minnesota	3. Oklahoma
4. Utah	4. Texas
5. New Hampshire	5. Florida
6. Connecticut	6. Arkansas
7. Rhode Island	7. Alabama
8. Hawaii	8. Missouri
9. Massachusetts	9. Alaska
10. Washington	10. South Carolina

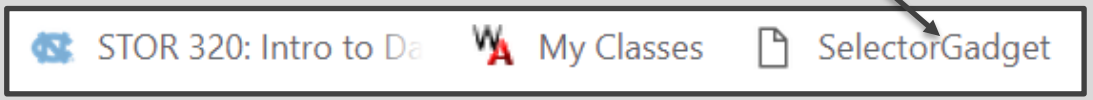
  - Goal: Scrape this Information into Vectors in R to Create a Table

Part 2: Inclusion of Expert Opinion

- Step 4: Identifying CSS Selector
  - Go to Web Page



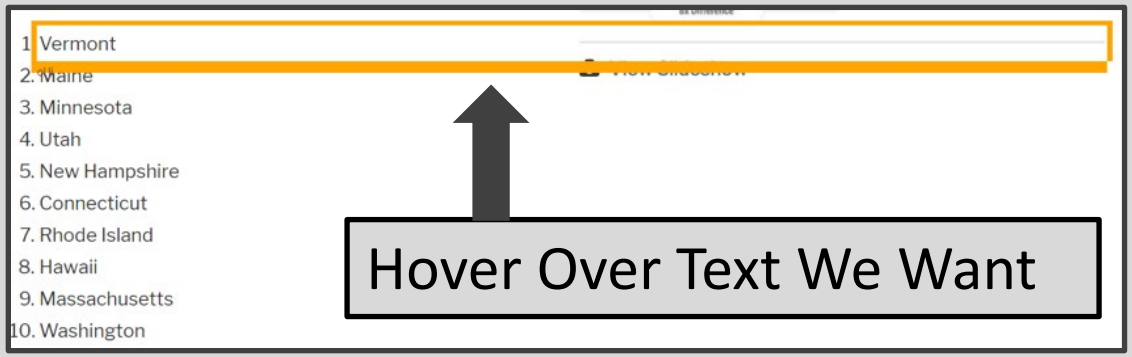
  - Choose SelectorGadget in Bookmark Tab



  - Locate This Box

Part 2: Inclusion of Expert Opinion

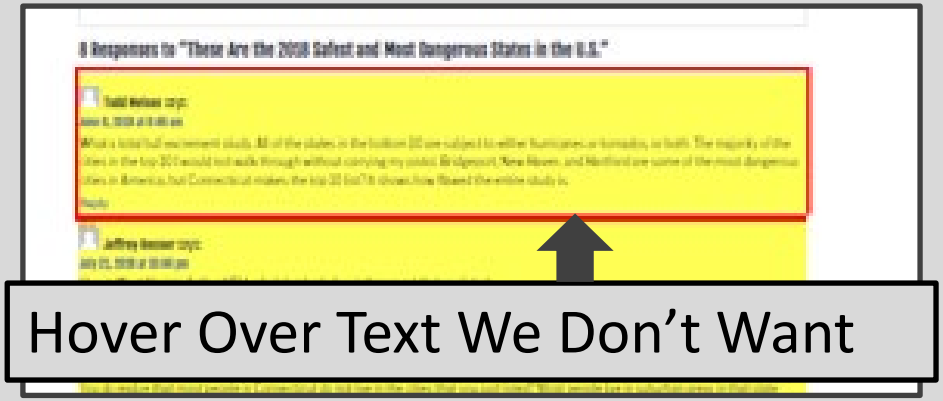
- Step 4: Continued
  - Find Content You Want



- Point and Click to Select Info
- Info We Want is Highlighted
- Info We Don't Want, As Well

Part 2: Inclusion of Expert Opinion

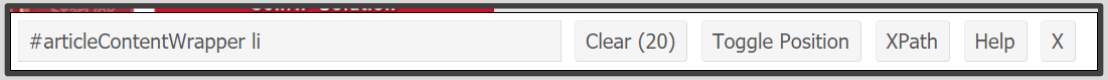
- Step 4: Continued
  - Find Content You Don't Want



  - Point and Click to Deselect
  - Locate This Box

Part 2: Inclusion of Expert Opinion

- Step 4: Continued
  - Locate This Box



  - Copy CSS Selector "#articleContentWrapper li"
- Step 5: Run Chunk 1

```
SAFE_VS_DANGEROUS = URL.SAFE_VS_DANGEROUS %>%
                    read_html() %>%
                    html_nodes(css="#articleContentWrapper li") %>%
                    html_text()
```

- Step 6: Run Chunk 2
  - What About the Other States?
- Step 7: Walk-off Knit

Closing

Disperse and Make Reasonable Decisions