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

Cities, Towns & Places

The population of all cities, towns and unincorporated places in the United States of America inhabitants according to census results and latest offcial estimates.

Name	Status	Adm.	Population Census (C) 1990-04-01	Population Census (C) 2000-04-01	Population Census (C) 2010-04-01	Population Census (C) 2020-04-01	Population Estimate (E) 2022-07-01
Abilene	City	TX	106,785	116,028	117,509	125,178	127,385
Akron	City	OH	223,142	216,899	199,077	190,416	188,509
Albany	City	NY	100,022	94,952	97,836	99,233	100,826
Albuquerque	City	NM	387,035	449,807	546,121	564,581	561,008
Alexandria	City	VA	111,183	128,278	139,998	159,461	155,525
Allen	City	TX	19,433	43,576	84,273	104,645	111,551
Allentown	City	PA	105,719	106,661	118,018	125,858	125,094
Amarillo	City	TX	157,680	173,562	190,675	200,377	201,291
Anaheim	City	CA	266,229	327,920	336,332	346,825	344,461
Anchorage	Mun	AK	226,338	260,299	292,252	291,244	287,145
Ann Arbor	City	MI	110,324	114,638	114,008	123,872	119,875

- 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 create a Percentage Change variable?
  - What is required to create a Population Density 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

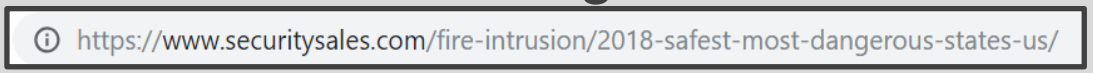
- Step 3: Information of Interest
  - Safe vs Dangerous

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

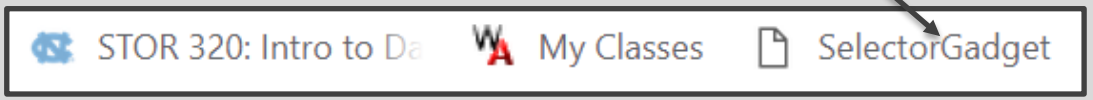
  - Goal: Scrape this Information into Vectors in R to Create a Table

Part 2: Inclusion of Expert Opinion

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



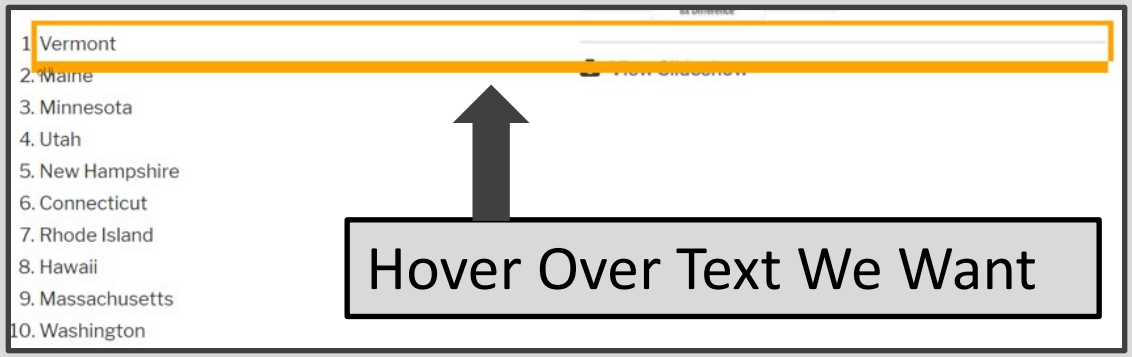
  - Choose SelectorGadget in Bookmark Tab



  - Locate This Box

Part 2: Inclusion of Expert Opinion

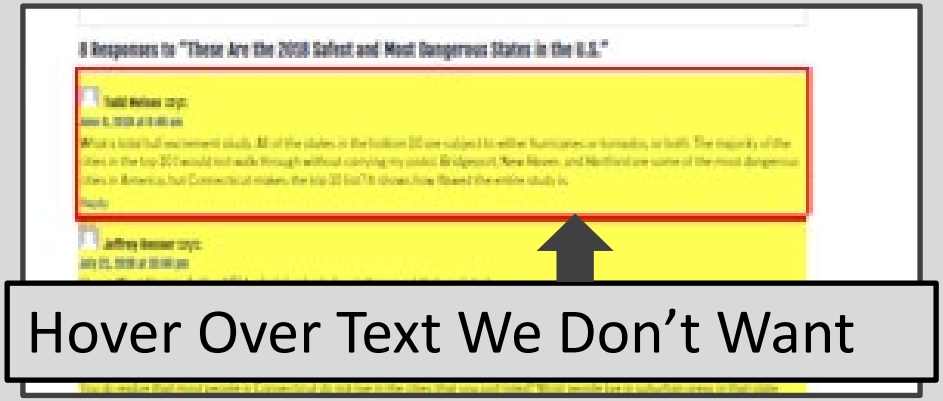
- Step 4: Continued
  - Find Content You Want



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

Part 2: Inclusion of Expert Opinion

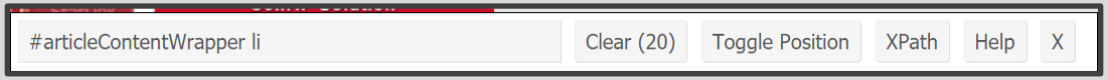
- Step 4: Continued
  - Find Content You Don't Want



  - 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