

# Problem Set 5

Karley Nadolski

March 8, 2021

## 1 Webscraping

For this section of the assignment, I first tried to scrape information from a different Wikipedia page to practice in a situation similar to what we walked through in class. I was asking my roommates for inspiration, and someone mentioned that I should look into the Wikipedia page titled "List of Unusual Deaths." Format-wise this suggestion worked out perfectly because all of the information was already organized in tables. I chose to scrape the information about unusual deaths during the 1990s. From this data, I created the dataframe "deaths1990s."

I also wanted practice working with data that wasn't already formatted in a table. I looked online for different walkthrough tutorials and found a tutorial for scraping information from Craigslist. I used a step-by-step guide from Dr. Nikhil Kaza's website. I specifically scraped Oklahoma City's Craigslist for information about Apartments for rent. Dr. Kaza's tutorial was really helpful in figuring out how to organize visually diffuse data into a single data frame. This data is interesting now just to have a glimpse into the rental real estate market in the OKC area as told by Craigslist, but it could also be really helpful in helping me directly compare rental opportunities should I be in a situation where I have to search for housing.

## 2 Webscraping with an API

Last semester, in Dr. Wong's class, we practiced webscraping via Twitter's API for text analysis. I referenced the code I worked on during that class to write code that scraped the 500 most recent tweets from the Norman area that had mention of the word 'vaccine.' Today (Monday, March 8) turned out to be a huge day for vaccine news because the state government announced that they would begin Phase 3 in their vaccine roll-out plan. To do this, I used the `twitterR` package in R with the "searchTwitter" function.

Data like this has always been interesting to me. One summer, I interned with the Communications department of the Governor's Office back home in Wisconsin. One of my main jobs was compiling regular "Media Market" reports that scanned social media accounts and news sources in different regions of the

state to monitor regional, political talking points. Code that can scrape data from Twitter and clean/organize it in a meaningful way could have been really helpful in automating an otherwise time-intensive process. I also just find social media data really interesting more generally. I'm fascinated by what gets people talking.

### **3 Sources**

Nikhil Kaza. "Scraping Craigslist Posts," February 4, 2020. <http://www.nikhilkaza.com>