Problem Set 5

Collin DeVore

March 2023

Question 3

As simple as this may seem, this is my first time data scraping without any supervision. That being said, I wanted to start with a very simple example, but one that could potentially be useful to me in the future. This table is under the keywords "Airline Hub" on Wikipedia. Generally, in the airline literature, it is common to sort of guess where the hubs are using the data and econometric modeling specifications. It is also common to use whatever definition the author decides to come up with for what a hub is. Using this table, however, I can easily see which airline has which hub and where it is, and it should be pretty easy to merge (left_join) into a larger dataset using the tidyverse package. In this way, I can pull this data and merge it to create a dataset that tracks where the top twenty-four airline hubs are located. The only notes that I used for this are the notes that we used in class. Specifically, I mimicked the same code and used Grant McDermott's lecture notes for this, though some of it (namely which copy to use after using inspect) came from my memory.

Question 4

Originally, I was going to use data from the FAA. After trying to find data with an API, however, I ended up finding this dataset from the FRED database. As stated before, this may seem like a simple dataset to grab, but it took me a while to understand how it works. For instance, I was not aware that R would store my API Key so that I do not have to worry about getting a new one. That was something I learned doing this. As for the data, it is interesting because I was working on a similar dataset last semester. I was working on wages between different groups of workers at different airlines after the end of the CARES Act. Finding this dataset was interesting after that project. Looking at the graph, airlines in aggregate faced a mass exodus from May 2020 to June 2020, and it is the largest drop since 1990 (which probably means it is the largest of all time since the airlines were deregulated in 1989). This was not supposed to happen. The CARES Act did not allow recipients to fire or furlough employees without their consent, which means that these workers had to have left voluntarily. This

is why this data is so interesting to me, and similar data may help me as I further develop my CARES Act idea. As far as packages that I used, I only had to use the fredr package that we learned in class. I have wanted to experiment with the FRED datasets since I was an undergraduate, so this was a fun experience.