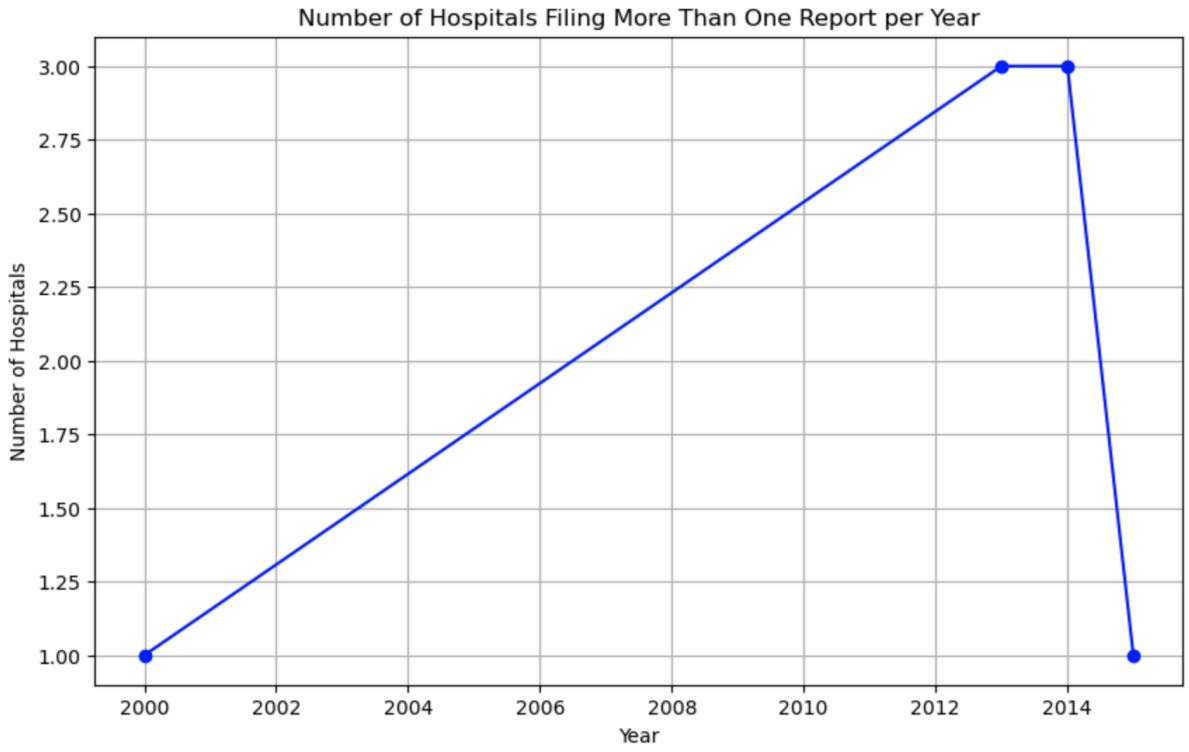
tan-s-homework2-1

Sarina Tan

ECON 470

Here is the link to my repository: https://github.com/sarina-tan/HLTH470hw2

1. How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.



2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

8
9323

3. What is the distribution of total charges (tot_charges in the data) in each year?
Need to fix data


```
: (8, 9323)
```

5. Calculate the average price among penalized versus non-penalized hospitals.
Need to fix data

6. Split hospitals into quartiles based on bed size. To do this, create 4 new indicator variables, where each variable is set to 1 if the hospital's bed size falls into the relevant quartile. Provide a table of the average price among treated/control groups for each quartile.
Need to fix data

7. Find the average treatment effect using each of the following estimators, and present your results in a single table:
Nearest neighbor matching (1-to-1) with inverse variance distance based on quartiles of bed size

Nearest neighbor matching (1-to-1) with Mahalanobis distance based on quartiles of bed size

Inverse propensity weighting, where the propensity scores are based on quartiles of bed size

Simple linear regression, adjusting for quartiles of bed size using dummy variables and appropriate interactions as discussed in class

I do not know

8. With these different treatment effect estimators, are the results similar, identical, very different?
I do not know

9. Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences)
idk my data is NaN

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated or surprised you.

My experience working with this data was a bit frustrating. The data took a long time to load onto my laptop as well as processing to make the new cleaned csv files. One thing that I learned is that with a lot of data, there are also a lot of blanks that need to be filled in and/or removed while merging files together. While I was able to make the final HCRIS data pretty smoothly, it was aggravating to then see that there were still blanks and spots that said NaN that made me unable to analyze it.