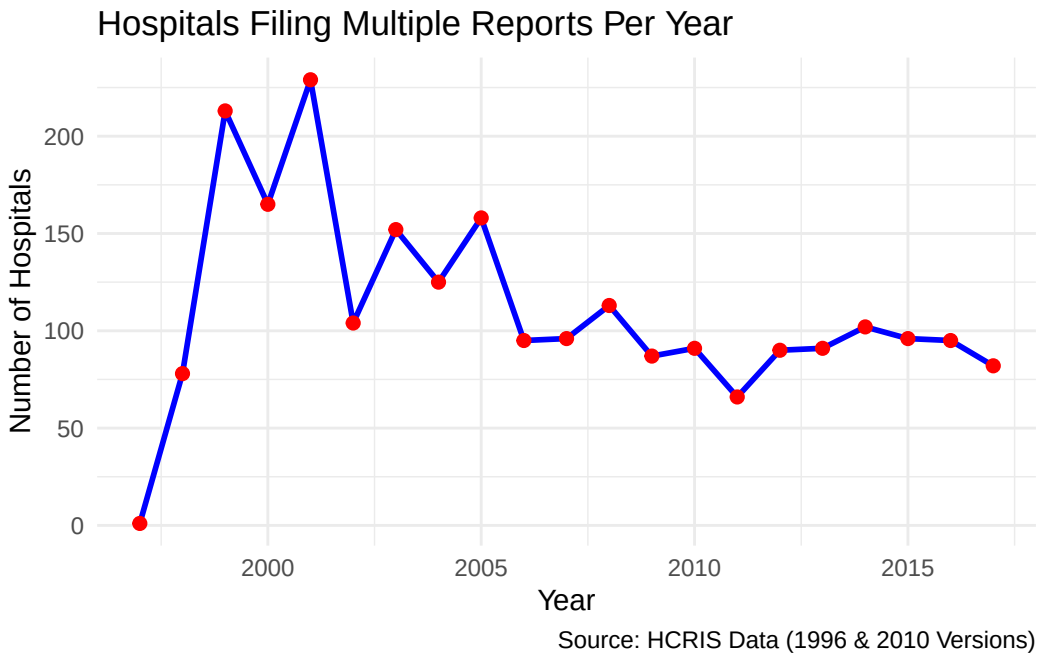
Homework 2

ECON 470, Spring 2025

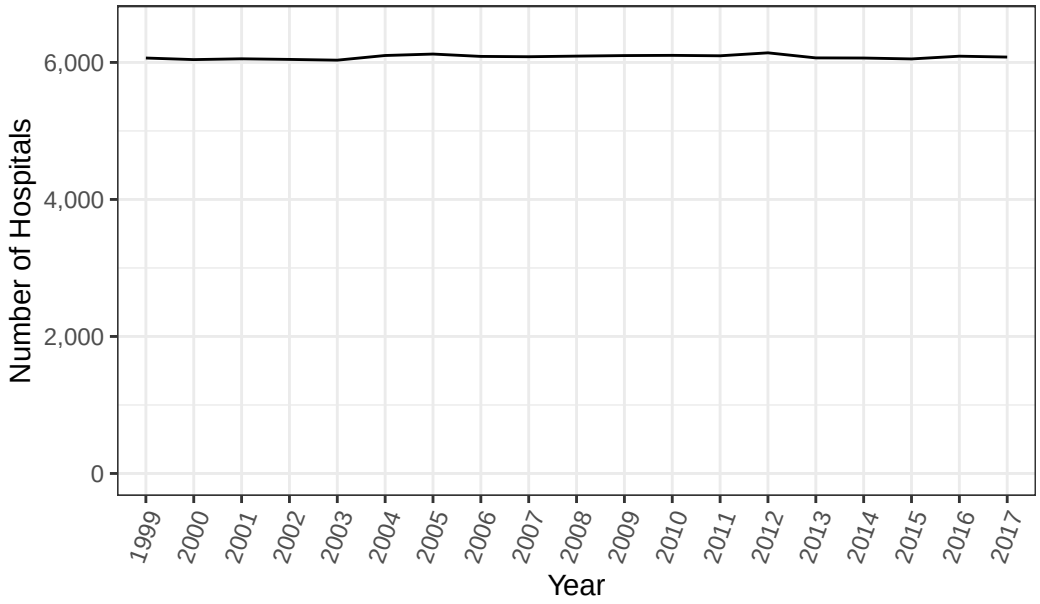
Molly Catlin

Here is a link to my repository: {https://github.com/mollyjc02/Homework_2.git}

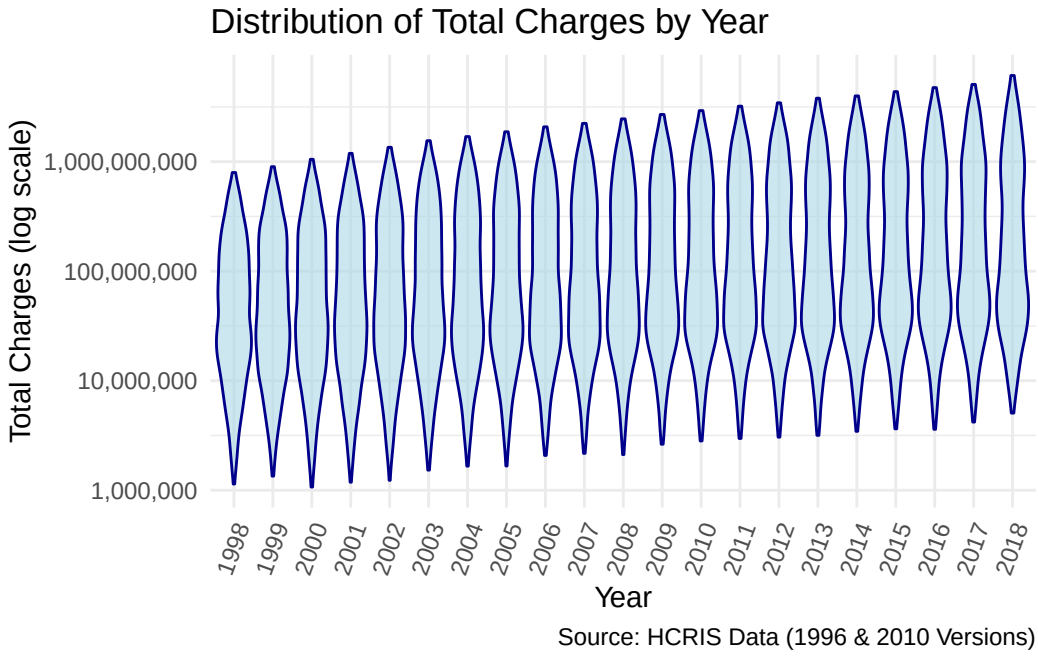
1. How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.



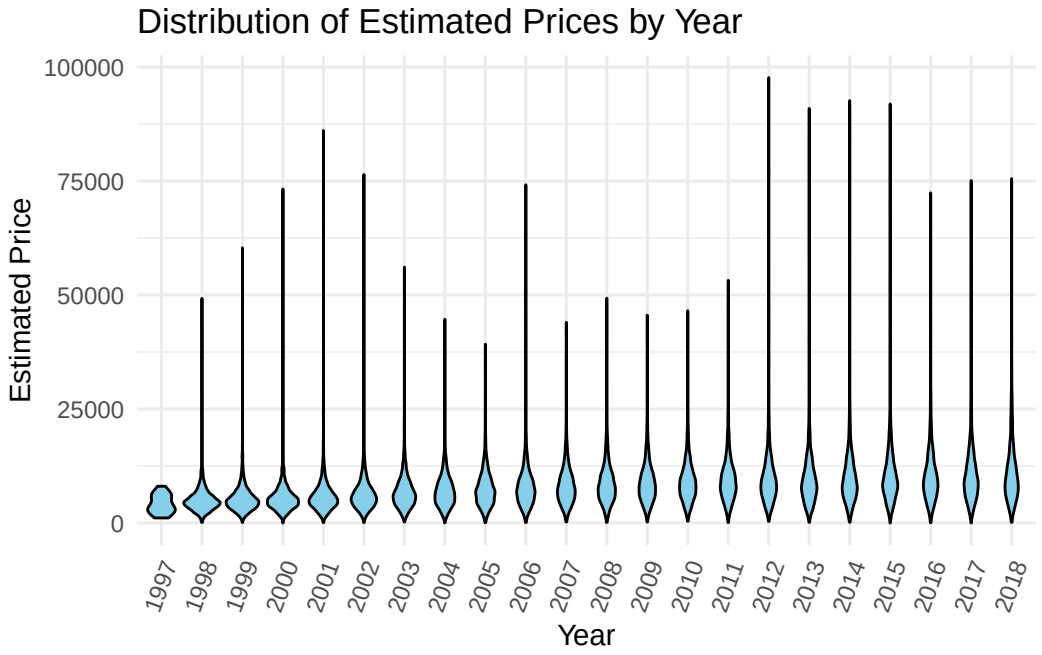
2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?



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



4. What is the distribution of estimated prices in each year?



For the rest of the assignment, I have included only observations in 2012. So I am now dealing with cross-sectional data in which some hospitals are penalized and some are not.

5. Calculate the average price among penalized versus non-penalized hospitals.

The average price among penalized hospitals is 9896.31, while the average price among non-penalized hospitals is 9560.41.

6. Split hospitals into quartiles based on bed size and provide a table of the average price among treated/control groups for each quartile.

Table 1: Average Price by Treatment Status and Bed Size

Bed Quartile	No Penalty	Penalty
Q1	7834.979	7802.316
Q2	8327.120	9083.821
Q3	9356.467	10144.617
Q4	10633.655	10971.422

7. Find the average treatment effect based on quartiles of bed size using each of the following estimators: nearest neighbor matching with inverse variance distance, nearest neighbor matching with Mahalanobis distance, inverse propensity weighting, and simple linear regression.

Table 2: ATE Estimates

Method	ATE Estimate
Nearest Matching (Inverse Variance)	505.7106
Nearest Matching (Mahalanobis Distance)	505.7106
Inverse Propensity Weighting (IPW)	505.7106
Linear Regression	505.7106

8. With these different treatment effect estimators, are the results similar, identical, very different?

Using nearest neighbor matching with inverse variance distance, nearest neighbor matching with Mahalanobis distance, inverse propensity weighting, and simple linear regression as estimators based on quartiles of bed size, I got identical results for the average treatment effect of instilling penalties.

9. Do you think you've estimated a causal effect of the penalty? Why or why not?

No, I do not think I've estimated a causal effect of the penalty. Although the estimating techniques used (such as matching and weighting) do improve the validity of relationships being measured, these results are still based on observational data and thus cannot determine causality. Additionally, in this data, I only looked into the estimated effect based on quartiles of bed size. There are a number of other unincluded confounders that could be influencing both the likelihood of receiving a penalty and the hospital's pricing behavior. So, while matching techniques do help control for confounders in the data, matching on only one variable is not sufficient to claim that the estimated effect is causal.

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.

In working with HCRIS data, I found the process of data cleaning to be extremely frustrating. Particularly, I struggled to know when it was appropriate to impose restrictions (e.g., cap the data at the 99th percentile) and how aggressive to be with these impositions (e.g., capping results at the 99th versus 95th percentile). That said, I did learn a lot about how to best structure my data cleaning and realized that it is really important to keep track of alterations made to variables in the data, which I think will help prevent me from running into similar issues in future assignments. Additionally, although not specific to this data necessarily, I found that forming my Quarto document was very frustrating. I spent nearly as much time figuring out why my Quarto wasn't rendering and making the document look presentable as I did on writing my code.However, I do feel I have now learned how to better and more efficiently work within Quarto, as well as how to interpret the errors I receive, which will definitely be a useful skillset to have.