Homework 2

Research in Health Economics, Spring 2025

Megan Zheng

In this assignment, I wil be analyzing the HCRIS data to pull informtion from their reports and estimate ATEs. The analysis and the raw data can be found in another script/document.

The GitHub repository for this work is available here. Enjoy!

Summarize the Data

Question 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.

Number of Hospitals Filing More than One Report per Year
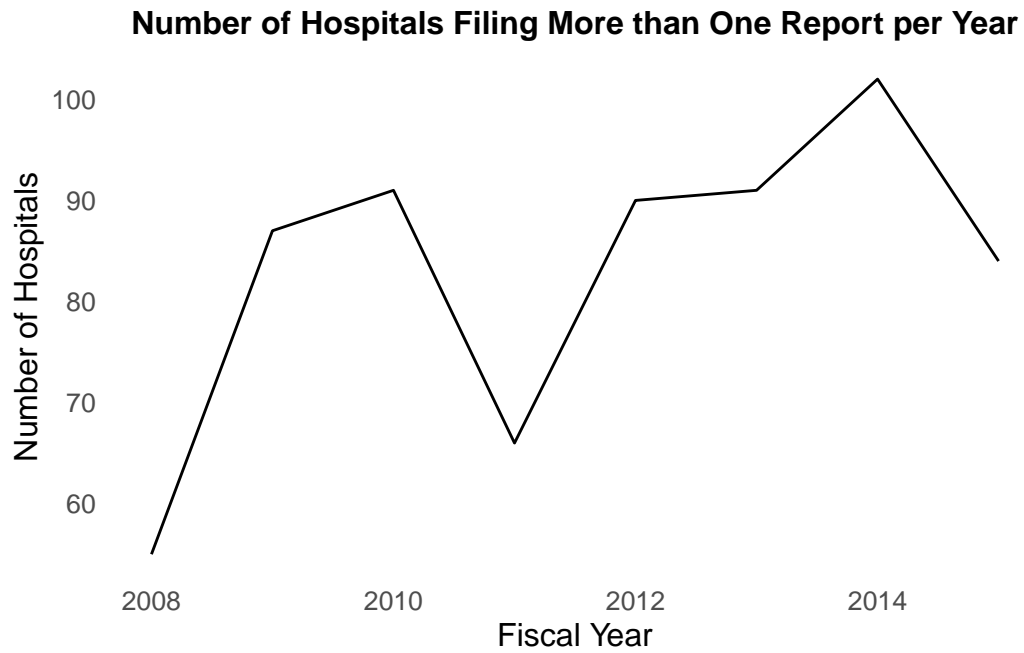
Question 2. After removing/combining multiple reports, how many unique hospital IDs exist in the data?

Table 1: Unique Hospital IDs per Year

Year	Unique Providers
2007	3
2008	1654
2009	2723
2010	2724
2011	2662
2012	2711
2013	2667
2014	2643
2015	2633
2016	1114

Question 3. What is the distribution of total charges in each year? Show your results with a "violin" plot.

Distribution of Total Charges by Year
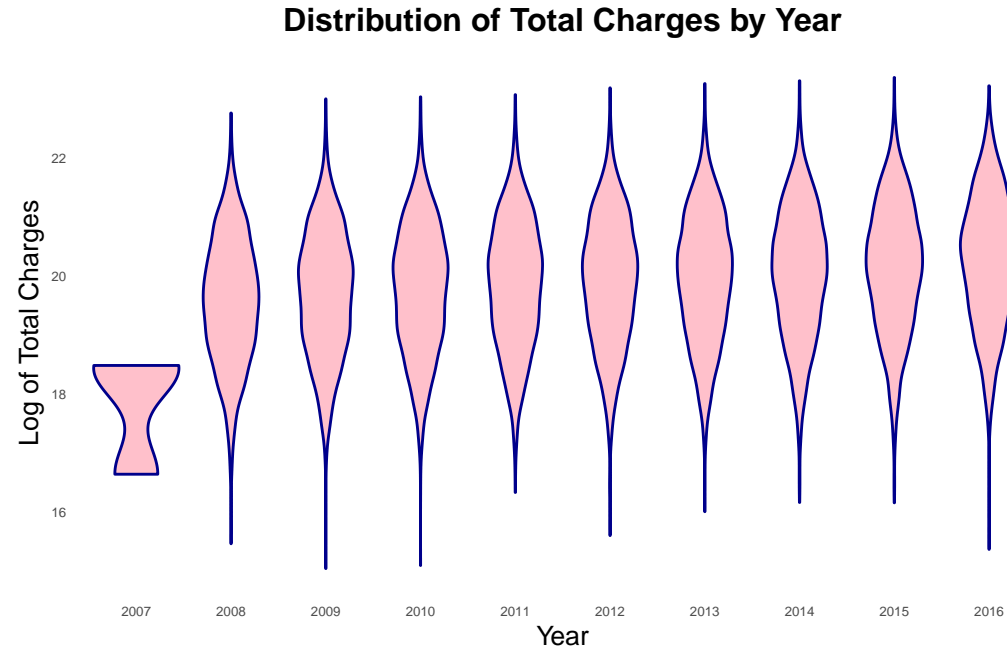
Question 4. What is the distribution of estimated prices in each year? Present your results
with a violin plot.

```
[1] 21534
```


Distribution of Estimated Prices by Year
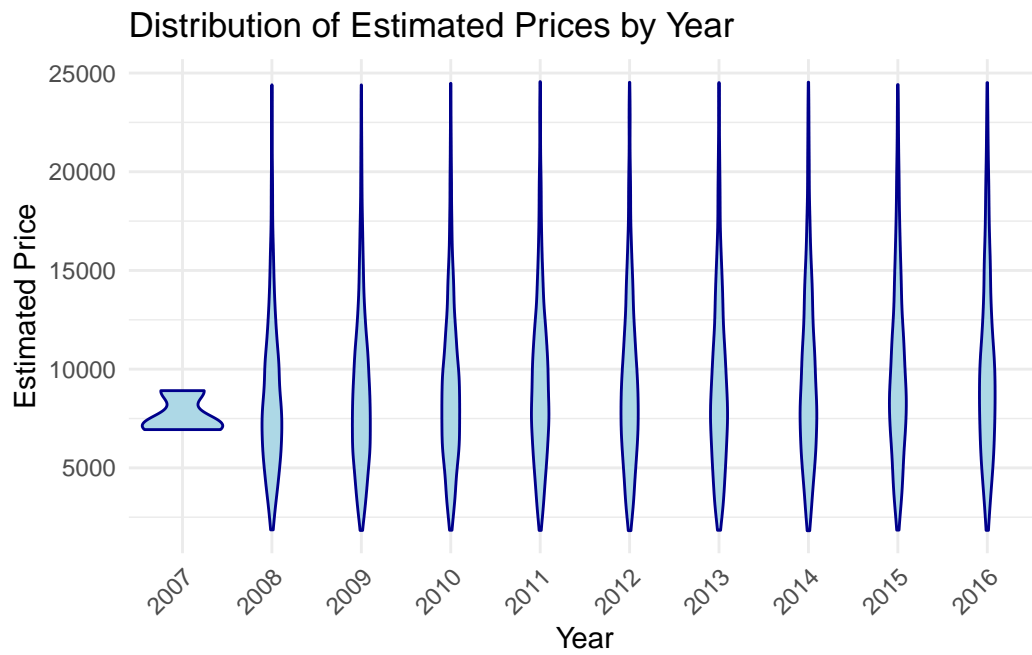
Estimating ATEs

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

From my analysis, the average price among penalized hospitals is $9685 and $9323 among non-penalized hospitals.

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

Table 2: Average Prices by Bed Quartile and Penalty

Bed Size	No Penalty	Penalty
Q1	7562.219	7530.446
Q2	8336.514	9592.664
Q3	9563.395	11026.711
Q4	11933.160	12882.861

Question 7. Find the average treatment effect using each of the following estimators, and present your results in a single table.

Estimator	ATE
Nearest Neighbor (Inverse Variance)	272.36395
Nearest Neighbor (Mahalanobis)	427.92620
Inverse Propensity Weighting	0.00000
Linear Regression	-31.77374

Question 8

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

The different methods give similar but not identical results. Each method gives a different estimate of how much the penalty affects hospital prices. One method finds an increase of $272, another $428, and one finds no effect at all ($0). These differences happen because each method balances the data in a unique way, but most suggest penalized hospitals charge more.

Question 9

Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences)

The penalty might not be the true reason for the price differences. Penalized hospitals already have higher prices on average ($9,685 vs. $9,323 for non-penalized hospitals), so something else could be influencing costs. Since different methods give different answers, we can't be sure the penalty alone is causing price changes.

Question 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.

This data was really interesting to work with, it gave me a chance to work with data as I would in the real world. But this also means I faced all of those challenges. It was frustrating when my data wasn't loading properly, and with so much data to work with, I was constantly fighting a battle with my computer just shutting down in the middle of running a file.