

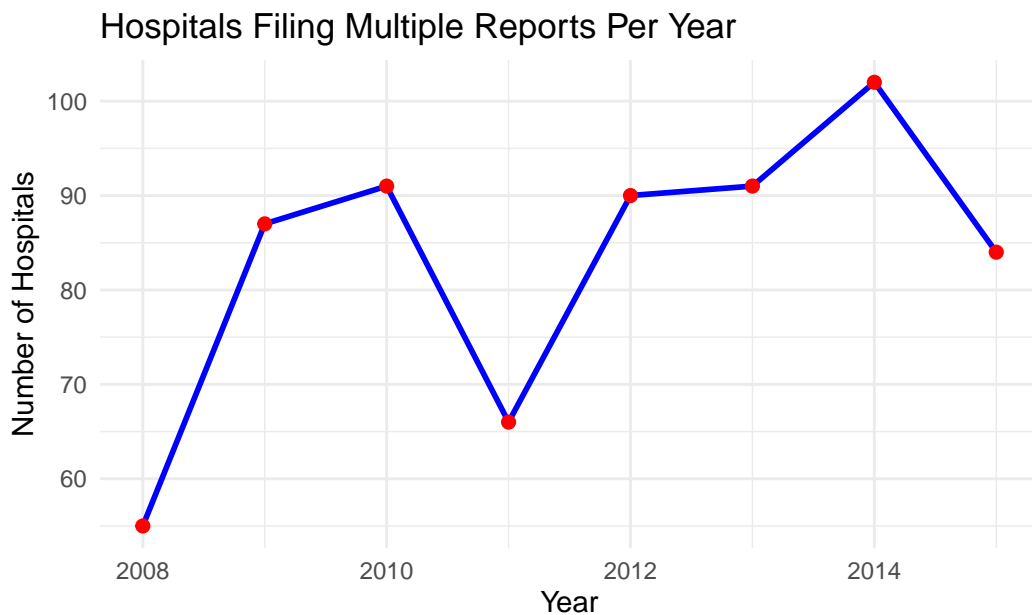
# Homework 1

Research Methods, Spring 2025

Genevieve DeBell

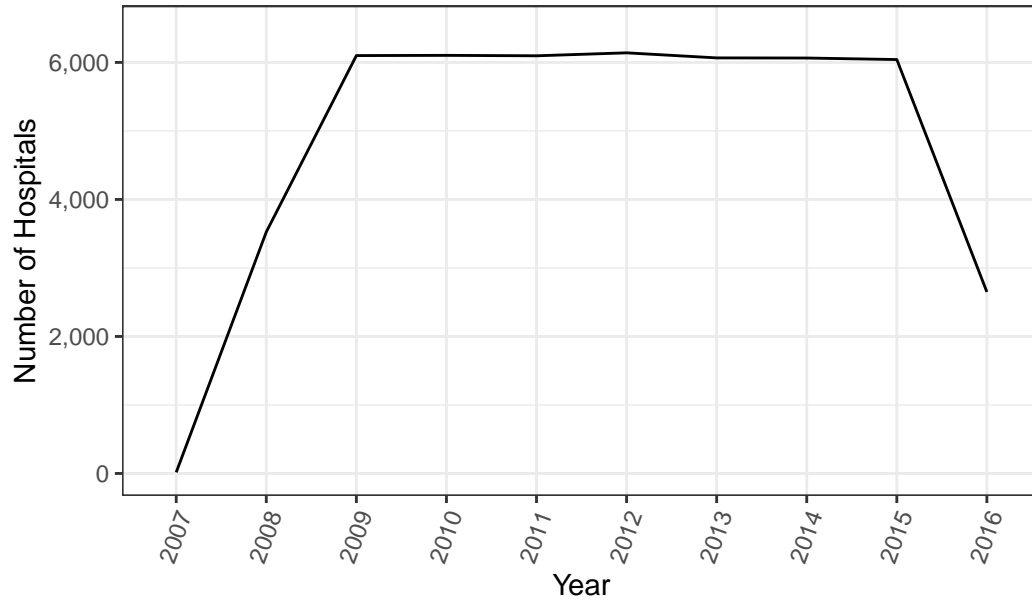
My answers to the homework questions are described below. The GitHub repository for this work is available [here](#). Enjoy!

**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.**

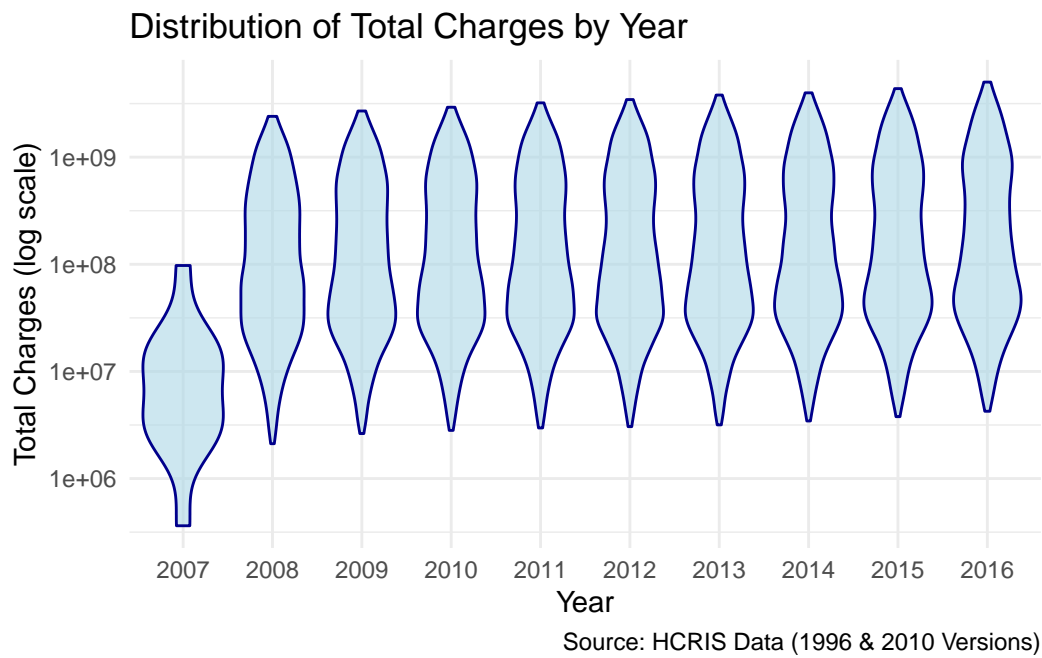


Source: HCRIS Data (1996 & 2010 Versions)

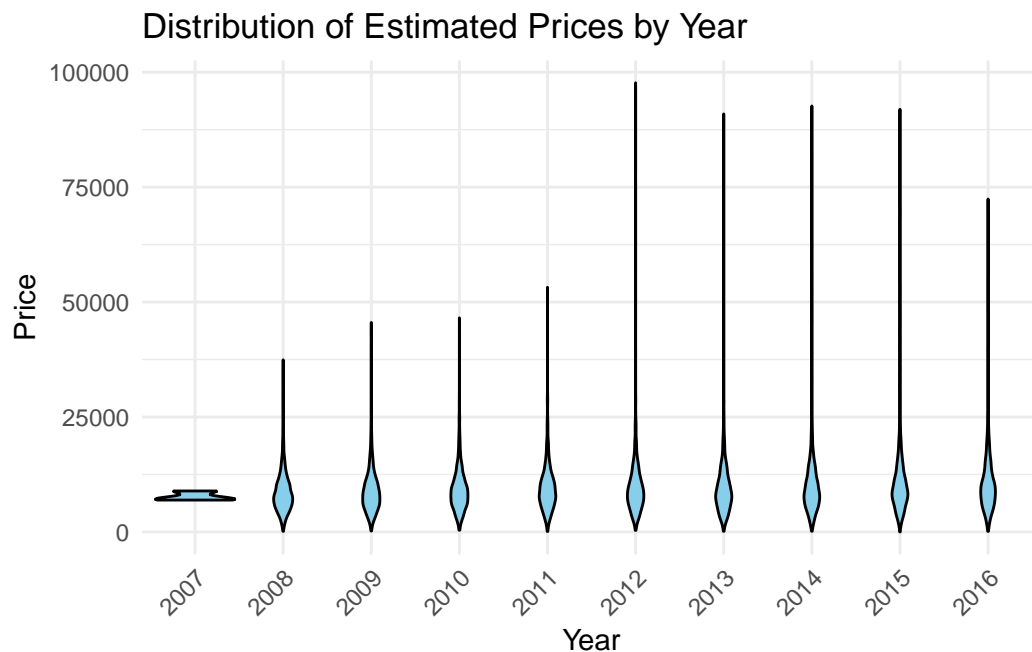
**2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?**



**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?**



**5. Calculate the average price among penalized versus non-penalized hospitals, only including observations from 2012.**

Average price among non-penalized hospitals: 9560.41

Average price among penalized hospitals: 9896.31

**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.**

Table 1: Average Prices by Quartile and Penalty Status

Bed Quartile	No Penalty	Penalty
Q1	7834.98	7802.32
Q2	8327.12	9083.82
Q3	9356.47	10144.62
Q4	10633.66	10971.42

**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.**

Table 2: ATE Estimates from Different Methods

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?**

With these treatment estimators, I got identical results.

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

I don't think I've estimated a causal effect of the penalty. While the matching and weighting techniques help create a more balanced comparison between penalized and non-penalized hospitals, we only estimated the effect based on quartiles of bed size, so there could still be unobserved confounders influencing both the likelihood of receiving a penalty and the hospital's pricing behavior. For example, we didn't take into account the health of the hospitals' patient populations which could significantly affect prices. Only matching on one variable makes it hard to claim that we estimated a causal effect.

**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.**

Working with this data was definitely more frustrating than the first homework. I found that it demanded that I be much more organized with my code. One thing that really aggravated me was the multiple rounds of data cleaning that we did to final.hcris.data because I found it hard to keep track of all the changes that I was making and what the different versions of the data were called. However, I think this assignment helped me learn how applying those changes to the data can affect the estimates you get when analyzing the data. I was getting pretty weird values at one point and had to go back to earlier versions of the data to figure out where things went wrong. This experience taught me the importance of maintaining clear and consistent documentation of data cleaning steps.