

bae-j-hwk-2-3

Jennifer Bae

2023-02-20

## Summarize the Data

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.

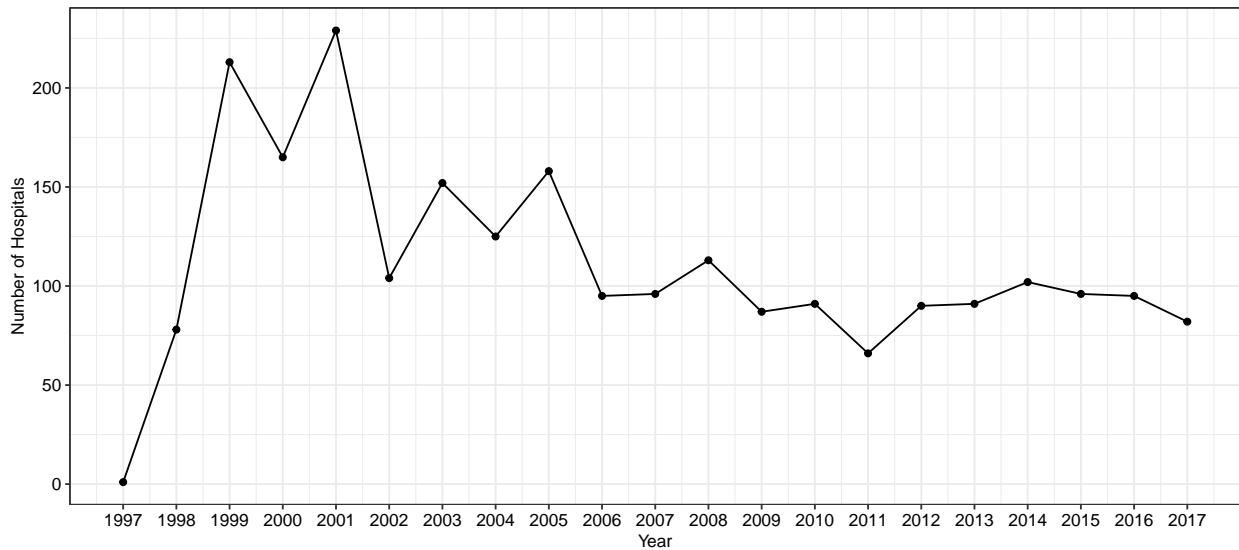


Figure 1: Number of Hospitals with more than 1 Report in Each Year from 1997 to 2018

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

After removing and combining multiple reports, there are 9,323 unique hospital IDs in the data.

3. What is the distribution of total charges (tot\_charges in the data) in each year? Show your results with a “violin” plot, with charges on the y-axis and years on the x-axis.

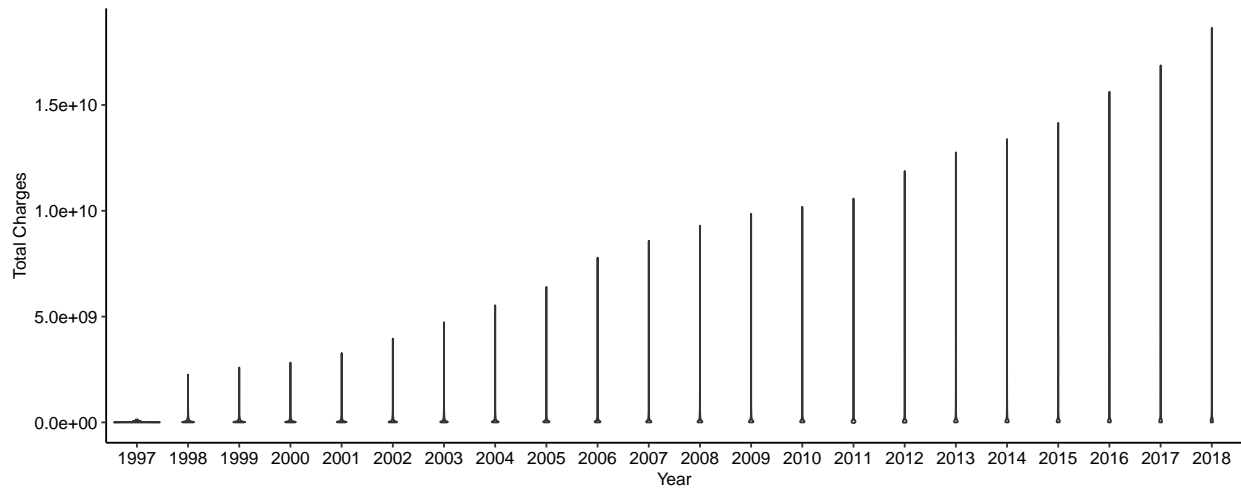


Figure 2: Distribution of Total Charges in Each Year from 1997 to 2018

4. What is the distribution of estimated prices in each year? Present your results with a violin plot, and recall our formula for estimating prices from class.

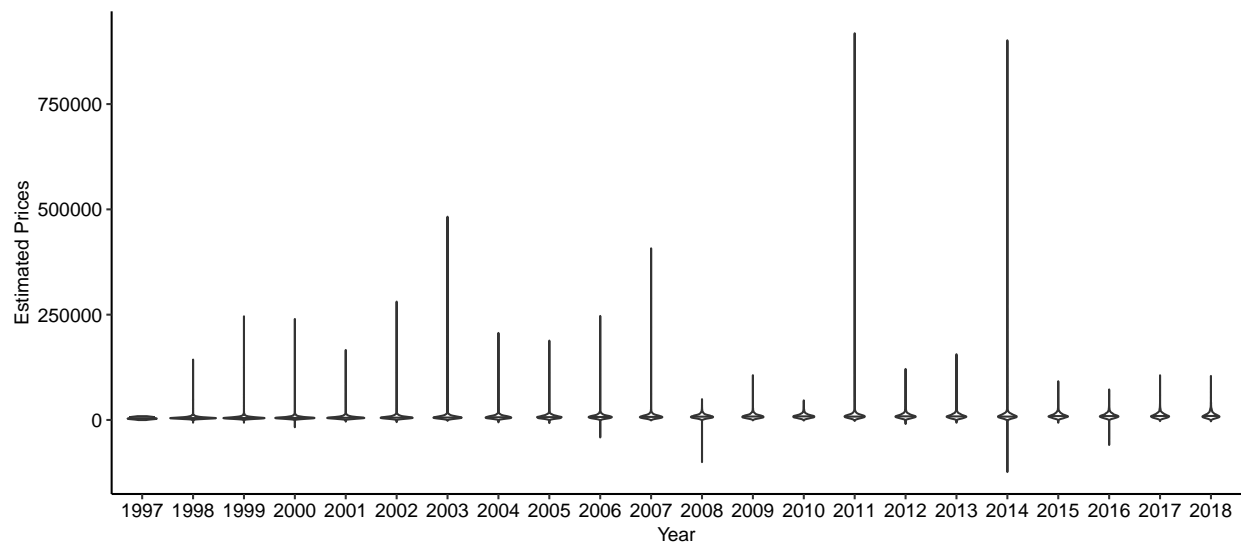


Figure 3: Distribution of Estimated Prices in Each Year from 1997 to 2018

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

The average price for penalized hospitals is 9,896.31 and the average price for non-penalized hospitals is 9,560.41

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

quartile	Non-Penalized	Penalized
1st Quartile	7684.240	8318.709
2nd Quartile	8510.959	8690.891
3rd Quartile	9856.928	10127.130
4th Quartile	12355.606	12068.479

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

Estimators	ATE
NN Matching, inverse variance	199.5281
NN Matching, mahalanobis	199.5281
Inverse pscore weighting	199.5281
Linear Regression	199.5281

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

As shown, the results with different treatment effect estimators were identical. This is due to the fact that we used dummy variables, such as the quartiles of bed size to draw these estimations. In this case, discrete variables were used, suggesting that the results would have been different if they had been a continuous variable instead.

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

I don't think I have estimated a causal effect of the penalty since we only took bed size into consideration when drawing our estimation and regression analysis. We examined the effect of penalty on price when only considering bed size while we omitted other variables for the purpose of this assignment, making our results prone to omitted variables bias and selection bias – as we discussed in class.

**10. Briefly describe your experience working with these data. Tell me one thing you learned and one thing that really aggravated you.**

Once again, these data were just a lot to handle at first because of the sheer volume of it. Also, I had a difficult time figuring out how to sort, filter data because I have zero experience, and there was not much guidance provided on where to start. However, my teammates who have more experience with coding in R and handling data sets have been very helpful.