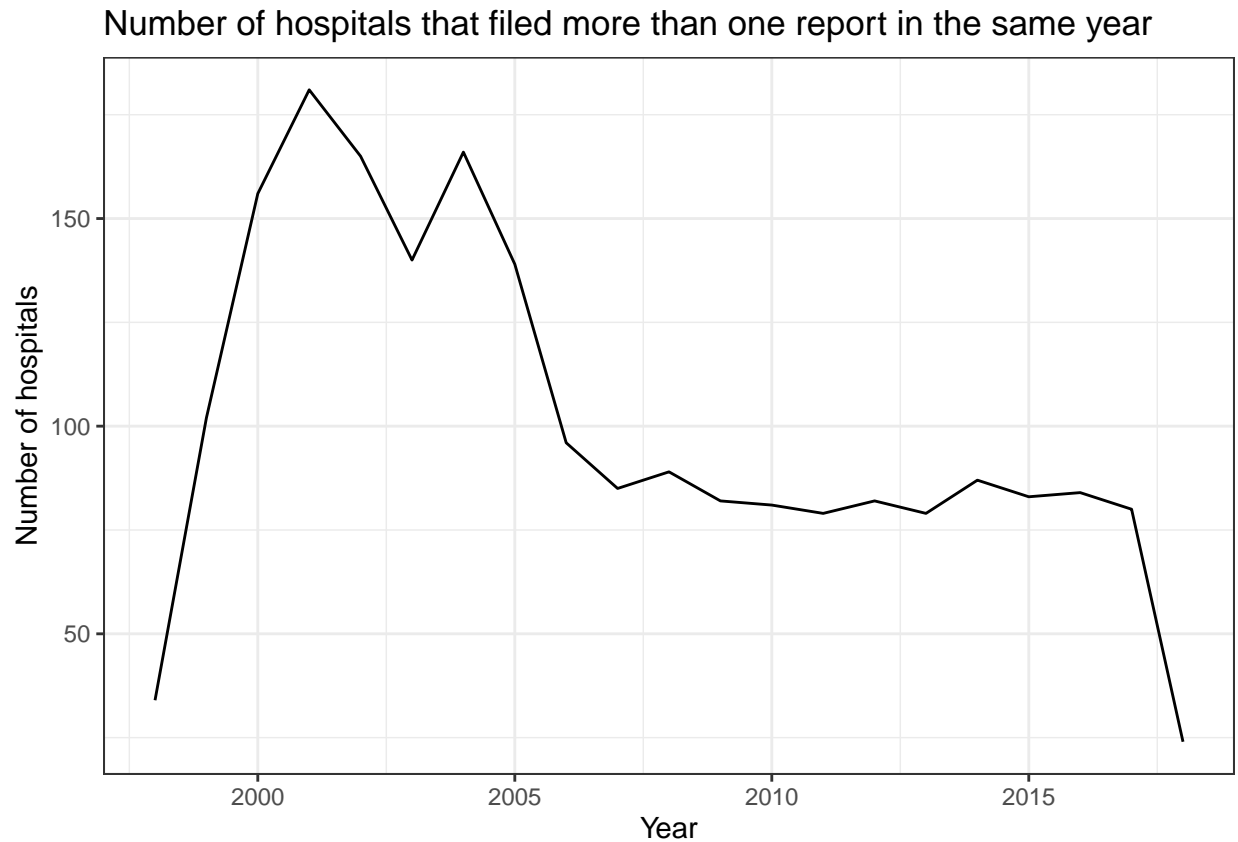# bae-j-hwk-2-1
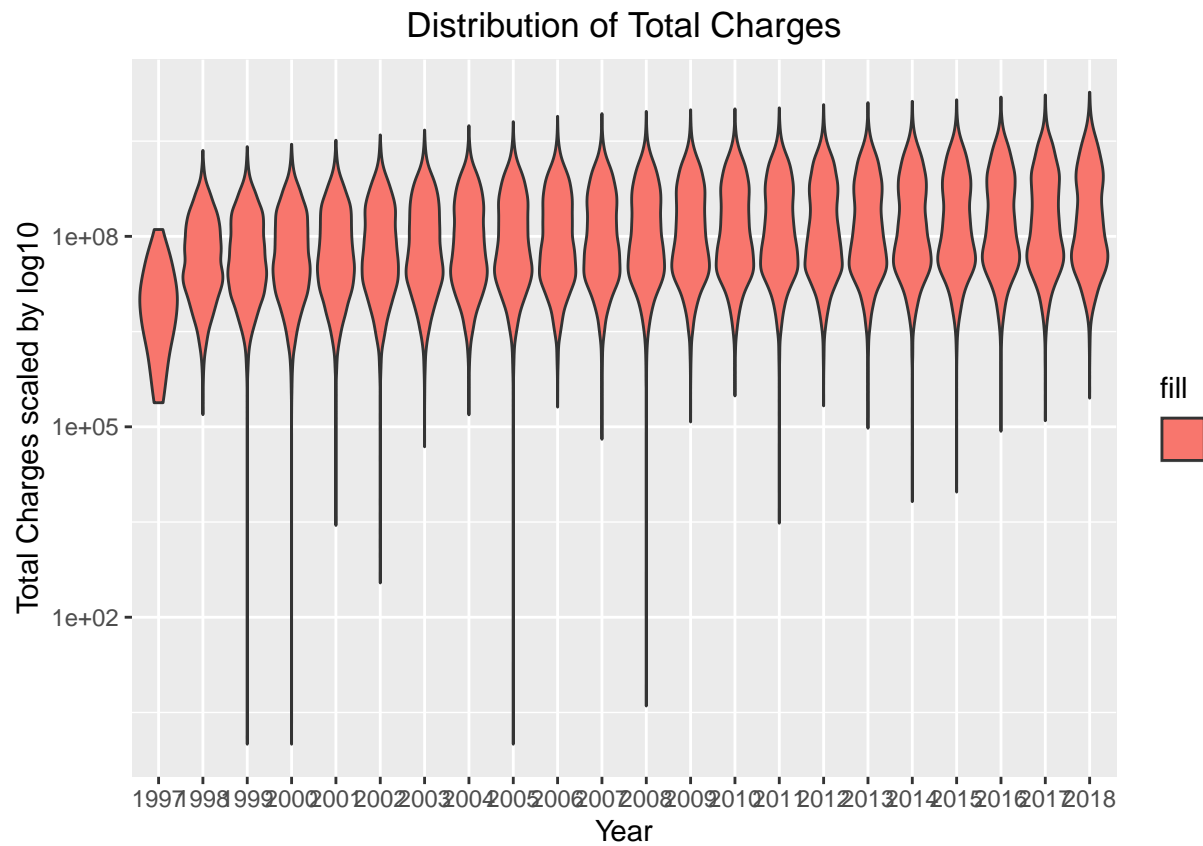
Jennifer Bae

2023-02-10

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.

Number of hospitals that filed more than one report in the same year
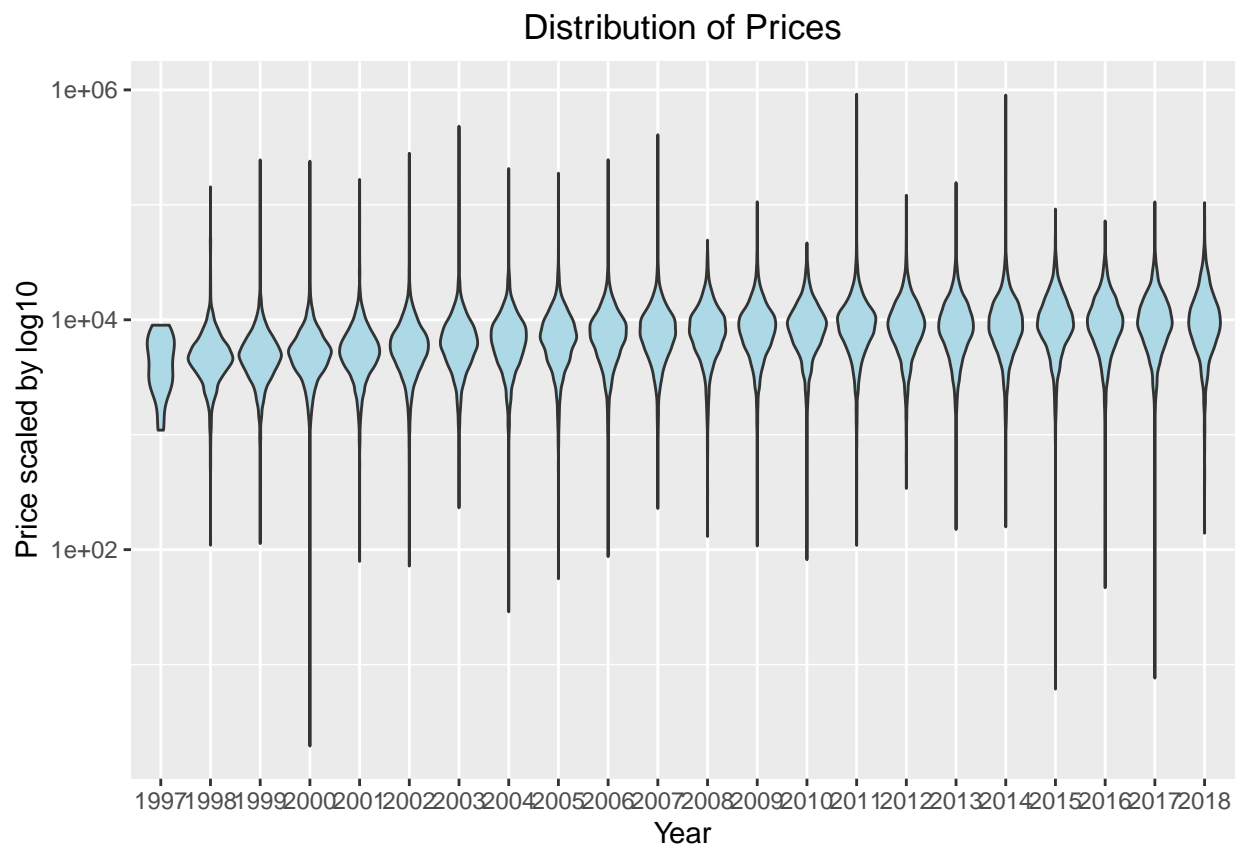

2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

After removing/combining multiple reports, there are 9323 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.

Distribution of Total Charges

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.

Distribution of Prices

Estimate the ATEs

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

The average price for penalized hospitals = 10,234.98 The average price for non-penalized hospitals = 9752.46

6. Split hospitals into quartiles based on bed size

```
## 'summarise()' has grouped output by 'quartile'. You can override using the
## '.groups' argument.
```

quartile	penalty	avg_price
1	0	NaN
1	1	NaN
2	0	10031.890
2	1	6680.007
3	0	8004.319
3	1	10079.437
4	0	11075.752
4	1	11263.644

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

As seen on the R file, I attempted to code this problem, but couldn't troubleshoot the error successfully by the due date. Will be working on it.

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

I'm not too sure without having the result from #7, but I would want to estimate that the results for different treatment effect estimators would be very different.

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

If the results for different treatment effect estimators were very different, I would want to say that I have estimated a causal effect of the penalty. The three criteria for establishing causality is covariation, termporal precedence, and control for confounding variables; however, I do not have enough information to ascertain the causality.

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 you.

Once again, these data were just a lot to handle at first because of the sheer volume of it. Also, I had a very 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.