Homework 2

Kendall Pollard

Here is the link to my GitHub Repository: https://github.com/kpollard8/Homework-2

Here are my answers for Homework 2. I do the coding in a separate `R` script, but here is the cleaned-up version. I run the analysis separately, save the workspace with only the summary stats, figures, and tables that I need, and then load the workspace in the final qmd. My analysis file with answers and code to all the questions is available in the analysis folder.

Duplicate Report

1. How many hospitals filed more than one report in the same year?

The answer to quesiton 1 is

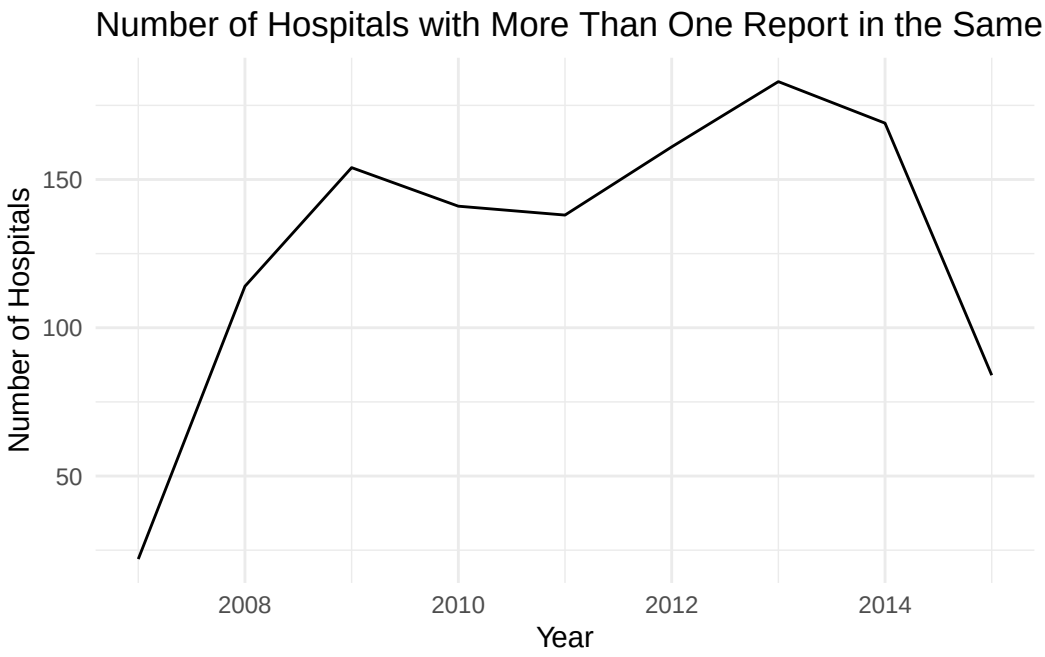

Figure 1: Hosptials with Mult Reports

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

The answer for question 2 is

```
num_unique_hospital_ids
```

```
[1] 48803
```

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.

```
Warning: Removed 2578 rows containing non-finite values (`stat_ydensity()`).
```

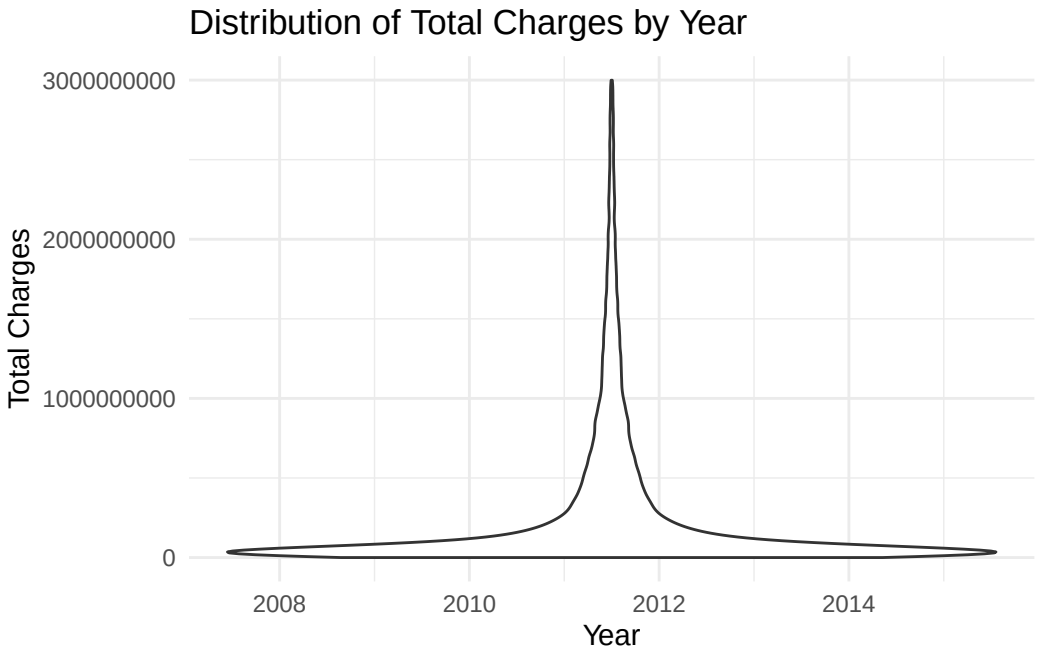

Figure 2: totalcharges

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

I used the equation from class and filtered outliers/negative prices by removing prices lower than 0 and setting a custom upper limit.

```
Warning: Removed 26937 rows containing non-finite values (`stat_ydensity()`).
```

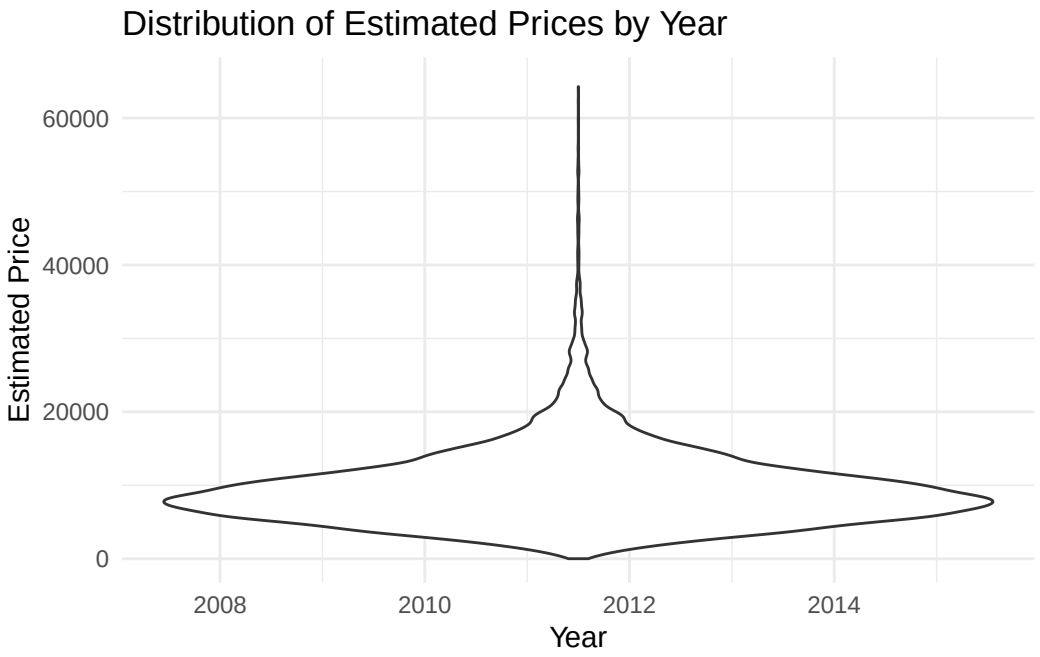

Figure 3: prices

5. Calculate the average price among penalized versus non-penalized hospitals. Average price among penalized: 9,896.31 Average price among non-penalized: 9,560.41

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

```
Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
resized.
```

Table 1: Bed size quartlies

Quartile	Penalized	Non-Penalized
1	8,318.709	7,684.240
2	8,690.891	8,510.959
3	10,127.130	9,856.928
4	12,068.479	12,355.606

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

In this first attempt, the ATE for the first three estimators is not working. I am running into errors with the matching, where the error said User estimand="ATE" Error in Matching::MatchY = lp.vars$price, Tr = lp.vars$penalty, X = lp.covs2, : length(Tr) != nrow(X) When I tried to fix it, I coded it so that it would tell me if it was "null" before even trying to match it. That happened for all 3 estimators. Also, I used "beds" as the x, but then I thought I should do it by qaurtile_1, 2, 3, 4, and neither seemed to work. I was able to do the simple linear regression for some reason, so the table only shows that. >

```
Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
resized.
```

Table 2: ATE 4 different ways

Blank	Estimator	ATE
penaltyTRUE	Simple Linear Regression	257.568

8. With these different treatment effect estimators, are the results similar, identical, very different? If the numbers align once I'm able to correctly do the ATEs for all estimators, then the results should be similar. Because they are all using different types of estimators, they are not going to be identical, but my hypothesis is that they would be similar.

9. Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences) I don't think we have established a causal effect because we are using various estimators to find an average treatment effect, not necessarily a causal effect. With ATEs that are statistically significant, we can say there is very high correlation/association, but still it's difficult to say it was directly 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.

This time around, I was definitely more comfortable working with VSCode (Quarto, the setup, etc.), but there were a few hiccups. I learned how to create violin plots which were not as hard as I thought they'd be. One thing that aggravated me was figuring out how to call all the variables/graphs/tables into the Quarto document. It took a lot of time to figure out how to correctly call them.