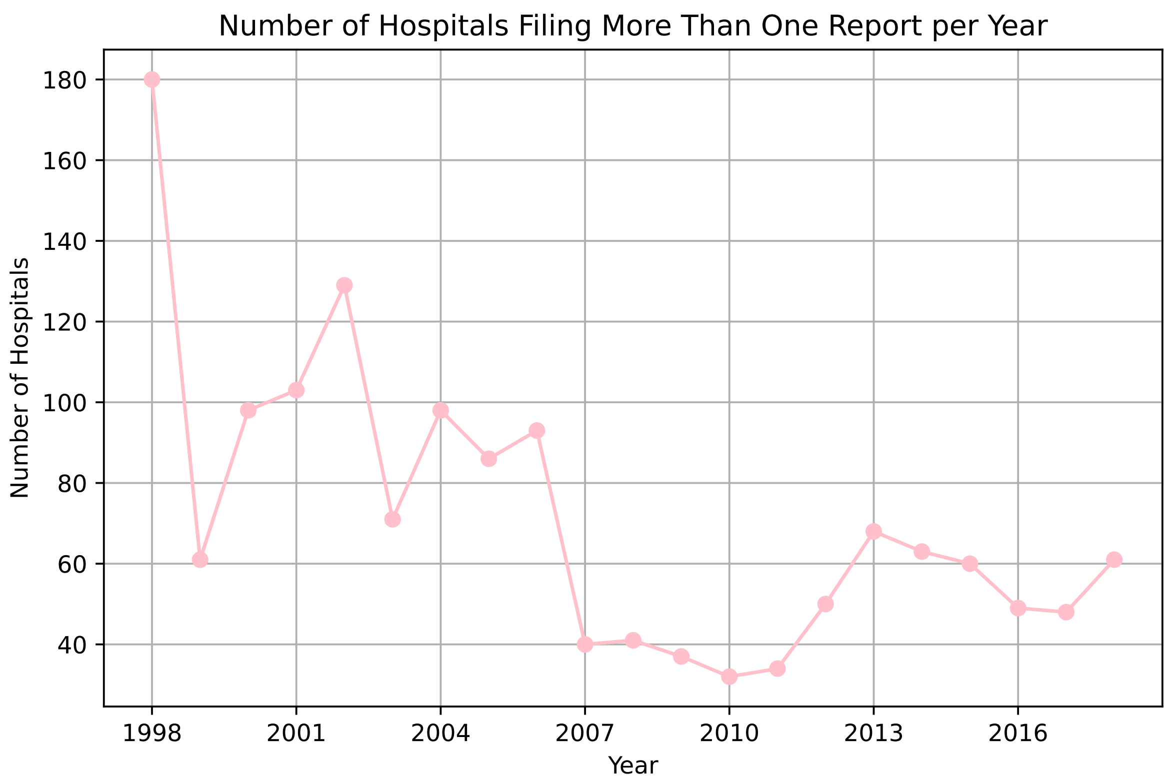
ECON 470 Homework 1

Ellen Wu

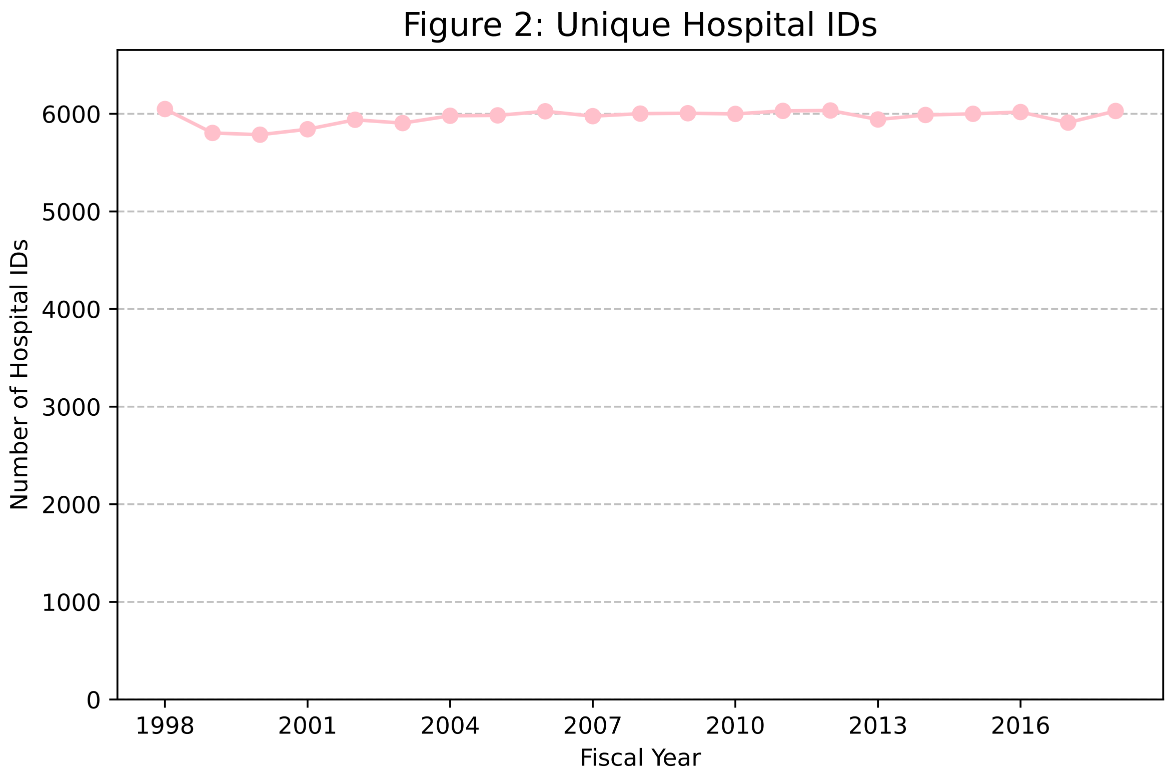
The link to my repository: https://github.com/ellenwu-git/homework2.git

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.

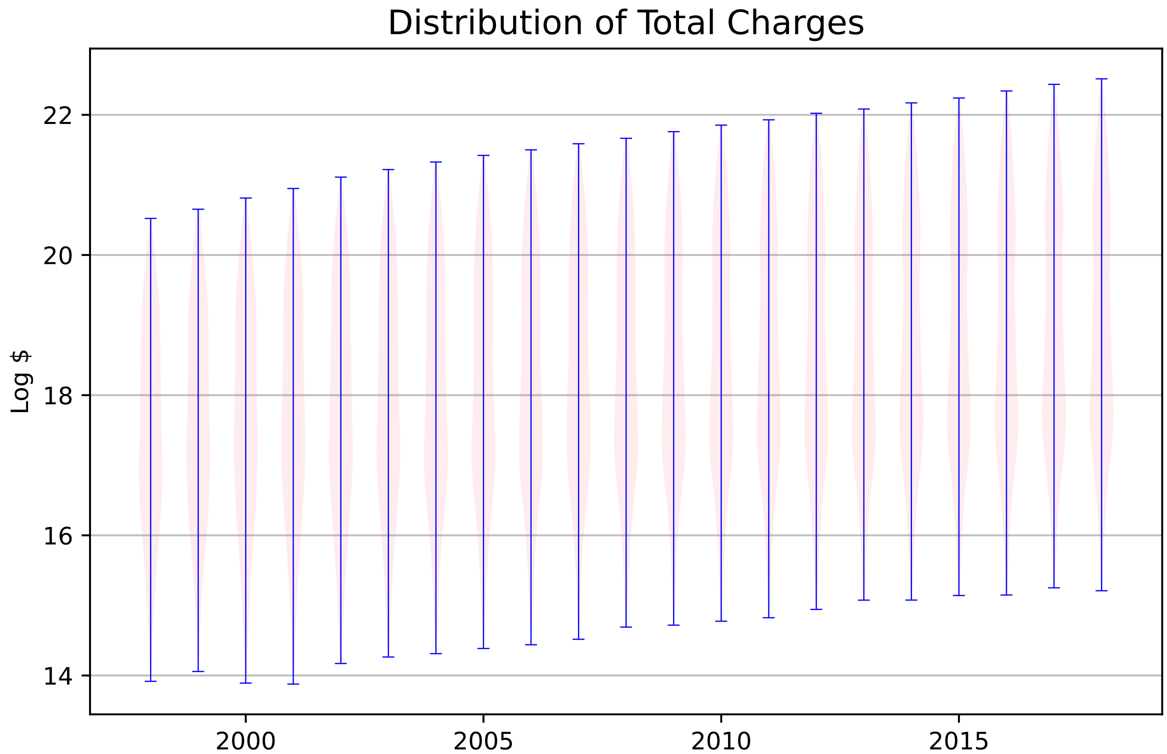


2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

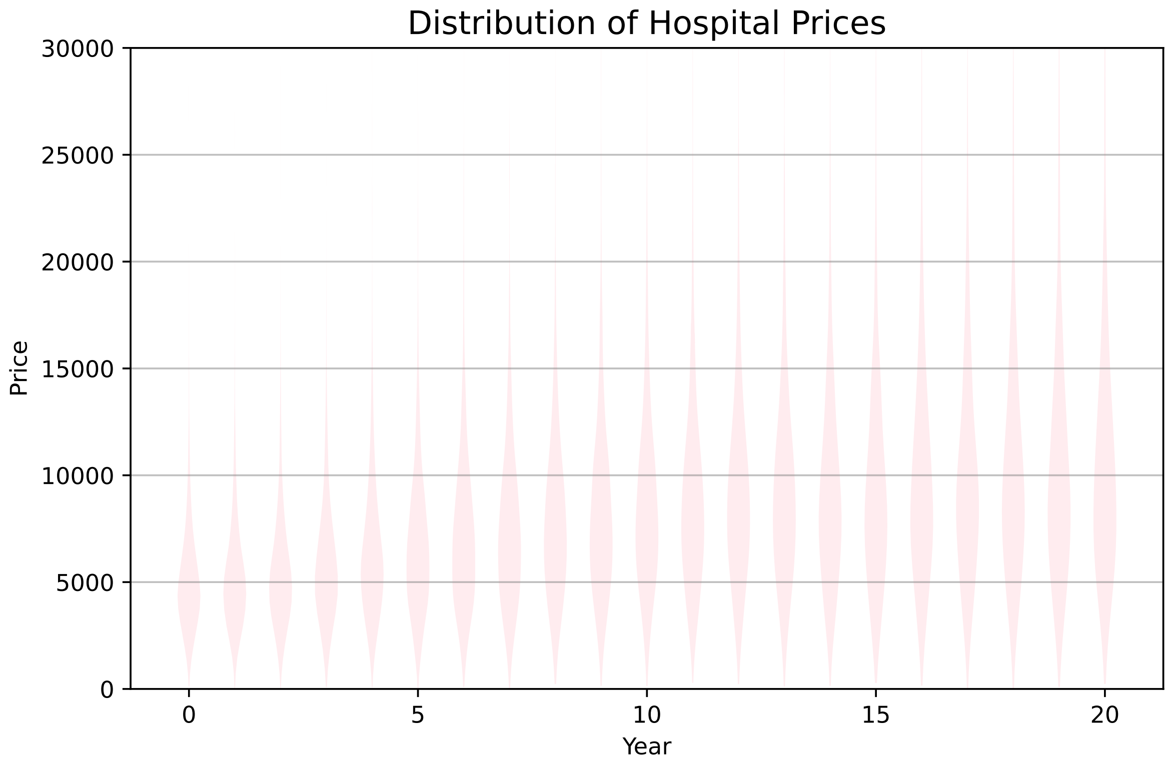
```
The number of unique hospital IDs is: 9382
```



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.



4. What is the distribution of estimated prices in each year? Again present your results with a violin plot, and recall our formula for estimating prices from class. Be sure to do something about outliers and/or negative prices in the data.



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

Table 1: Avg. Price for Penalty Status

Penalty Status	Average Price
Penalized	10171.5
Non-Penalized	9651.82

6. Provide a table of the average price among treated/control groups for each quartile.

Table 2: Avg. Price Among Treated & Control Groups

Quartile	Penalized Avg. Price	Non Penalized Avg. Price
1	$7,277.14	$7,759.76
2	$8,877.66	$8,638.68
3	$10,036.28	$9,521.91
4	$12,690.44	$12,193.30

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

```
<causalinference.causal.CausalModel object at 0x174609750>
```

Average Treatment Effect Estimates

```
|     |   NN-INV |   NN-MAH |     IPW |     OLS |
|-----|----------|----------|---------|---------|
| ATE |  187.251 |  187.251 | 187.251 | 187.251 |
| SE  |  203.755 |  203.755 | 206.161 | 206.016 |
```

	NN-INV	NN-MAH	IPW	OLS
ATE	187.251	187.251	187.251	187.251
SE	203.755	203.755	206.161	206.016

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

The results are very similar across all estimators, with the average treatment effect (ATE) being the same across all four estimators. The standard error (SE) slightly vary, but still very close.

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

Through matching (nearest neighbor) and inverse propensity weighting, causal identification strengthens by balancing observable covariates between penalized and non-penalized hospitals. However, these methods only control for observed confounders, meaning unmeasured factors could still bias the estimates. An example of this could be hospital management quality. Therefore, while the estimates are close to a causal effect, they are not fully causal unless we assume all relevant confounders have been accounted for.

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 a steep learning curve. I was feeling a lot of frustration because even though I was running the data cleaning code we were provided in class, it would still produce me a dataset with missing years. In many attempts I would be missing values for 2012-2015. It definitely took some time to debug this error by running each code line by line, which eventually paid off.