Homework 2

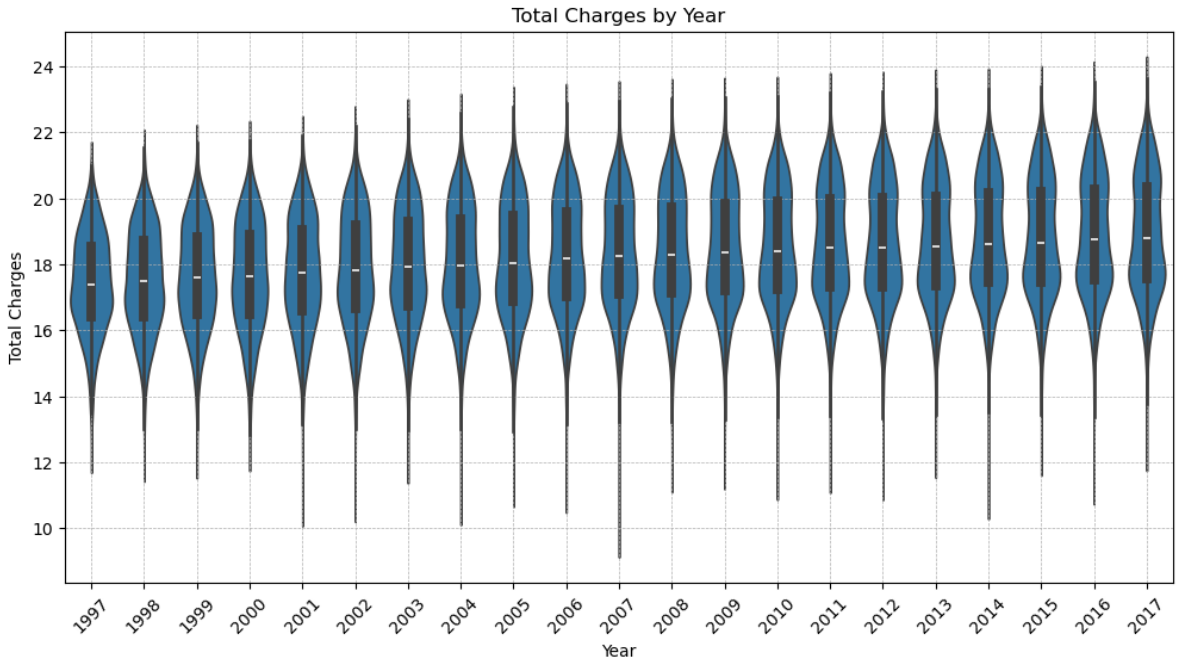
By Avanth Pakanati

Question 1: Graph of Hospitals That Filed More Than One Report

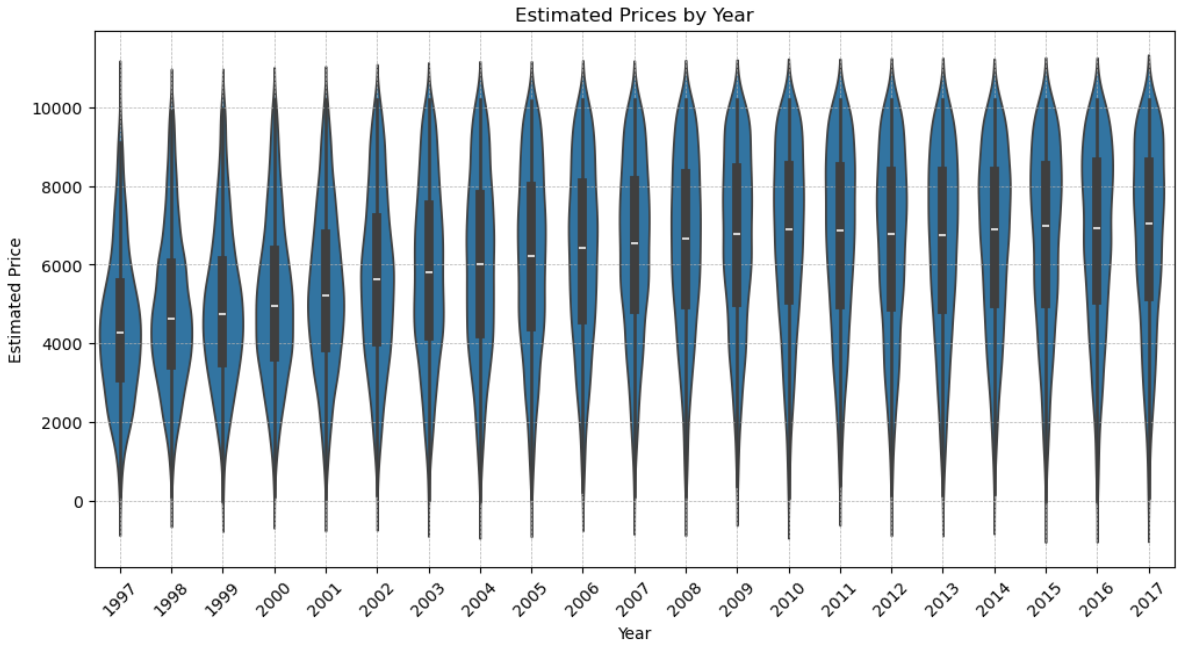
Question 2

```
Number of Unique Hospital IDs: 9323
```

Question 3 - For the graph, I dropped all negative values and outliers (Top 5% values)



Question 4 - For the graph, I dropped all negative values and outliers (Top 25% values)



Question 5

```
Average price for penalized hospitals in 2012: 10087.73
Average price for non-penalized hospitals in 2012: 9388.11
```

Question 6 - Average Price for Each Quartile Group

	Quartile	Penalized Mean Price	Non Penalized Mean Price
0	1	7558.71	7408.88
1	2	9444.01	8387.51
2	3	10956.85	9429.43
3	4	12678.08	12308.15

Question 7 - Average Treatment Effect Across All Four Methods

	INV	MAH	IPW	OLS
ATE	774.168691	774.168691	774.168691	774.168691
SE	249.737628	249.737628	248.914803	248.696831

Question 8 - The different treatment effect estimators are exactly identical. This is the result that I expected.

Question 9 - No I do not think we've estimated a causal effect. In our model, we have purely adjusted for bed size and there many other factors that we are not accounting for. These other factors would need to be controlled for in order to estimate a causal effect of the penalty. For example, patient acuity and quality of the hospital are both factors that could definetely effect penalty.

Question 10 - Overall, I had a difficult experience working with this data, but learned a lot of valuable lessons along the way. When I have worked with data in other classes in the past, I have always been given a nice, clean dataset and am able to begin analyzing the dataset easily. When working with this data, a lot of the work is cleaning and merging the data, which I definitely found to be challenging. However, I definitely recognize how important these skills are. When working with data in the real world, it is rarely ever clean, so I am really happy to be getting experience with this. Once the dataset was cleaned, I had some issues creating the Violin Plots but was able to work through it by filtering out negative values and outliers.