Homework 2 - Submission 3

ECON 470

Baran Pasa

Homework 2 Analysis

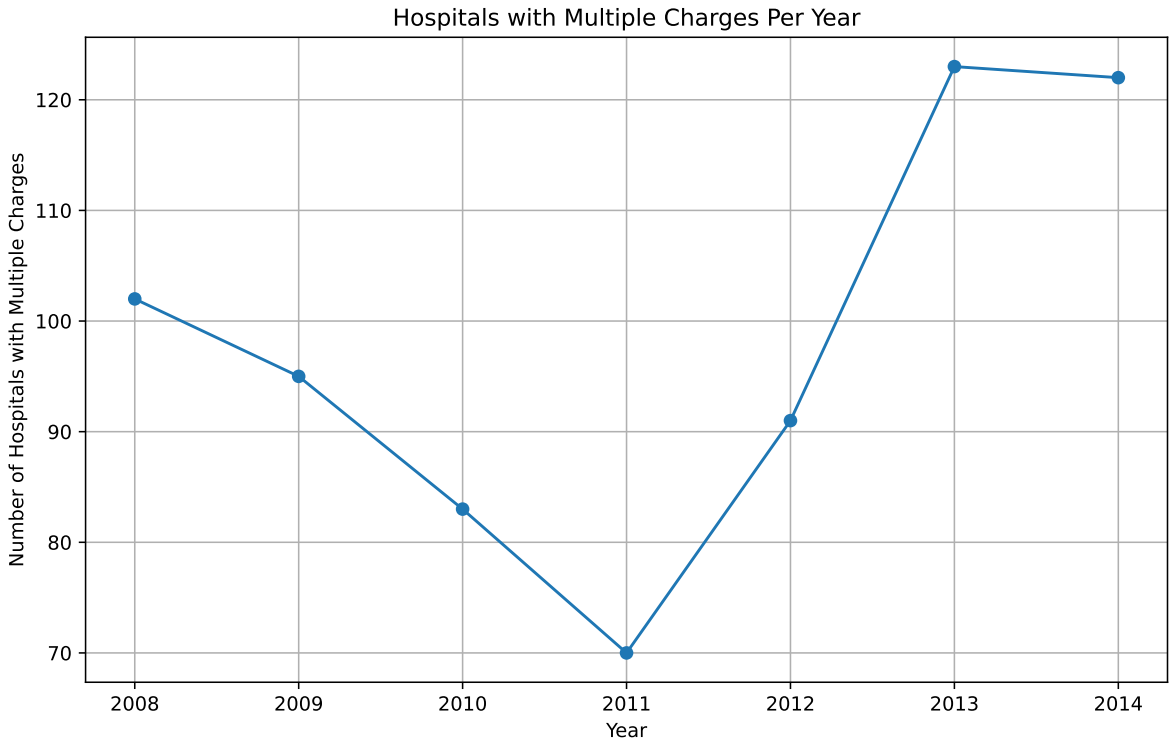
Link to Github

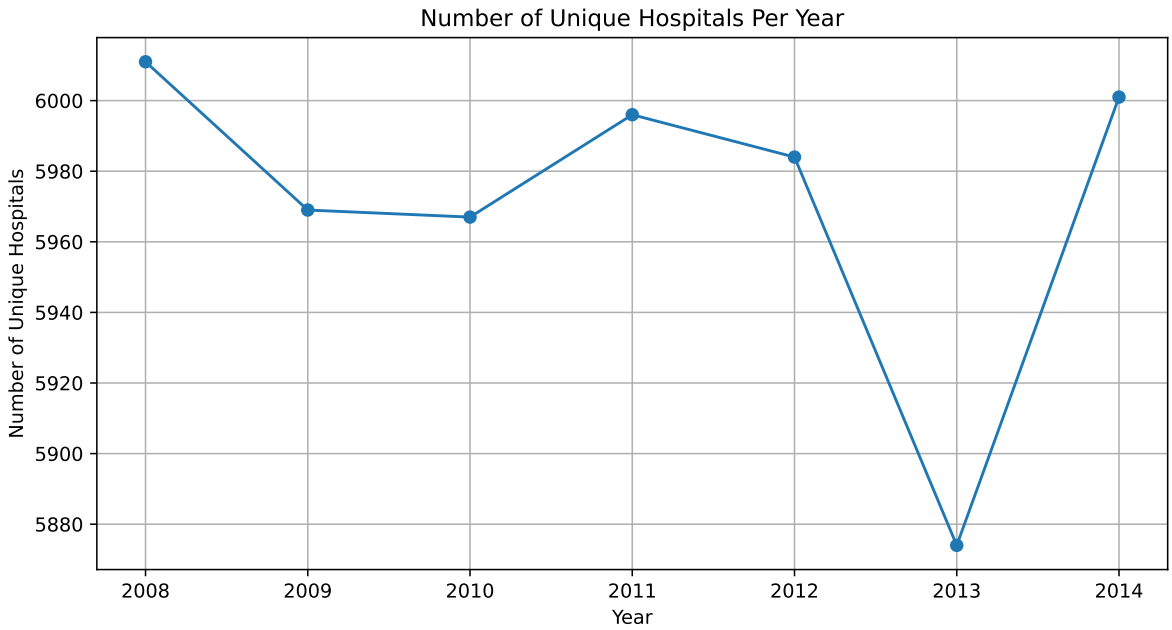
Part 0: Importing and cleaning the data

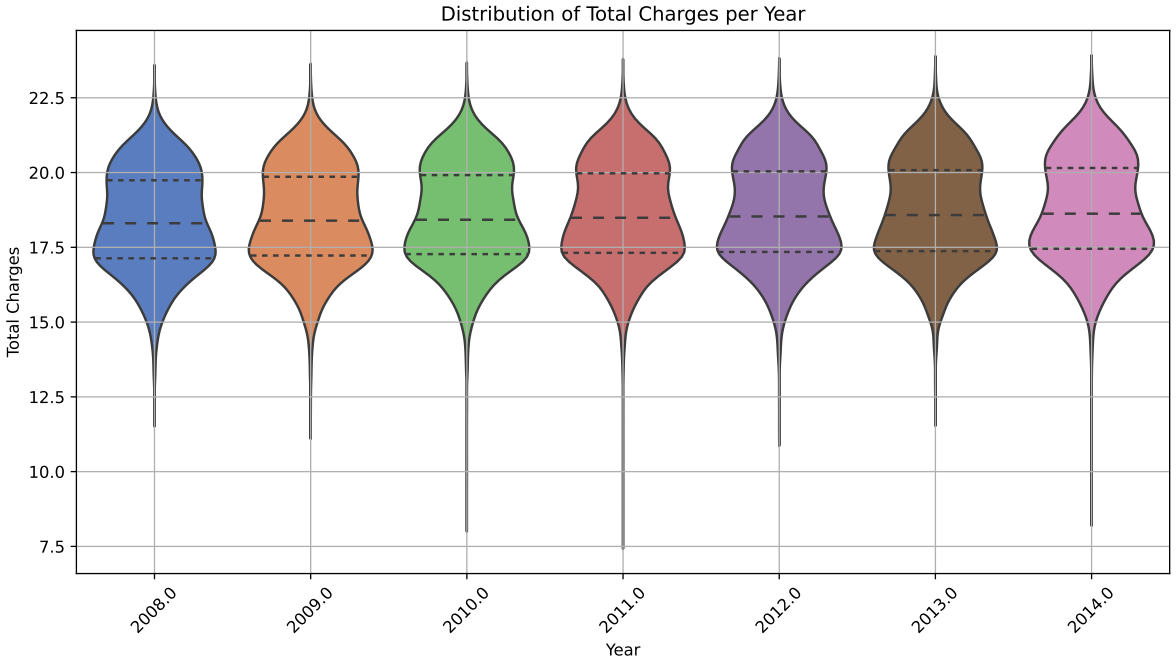
The data was cleaned in order to make summarizing and analysis easier. My cleaning process relied heavily on the code kindly provided to us by our TA Pablo. Additinally, data was only used from 2008 to 2015 due to my computer's lack of power.

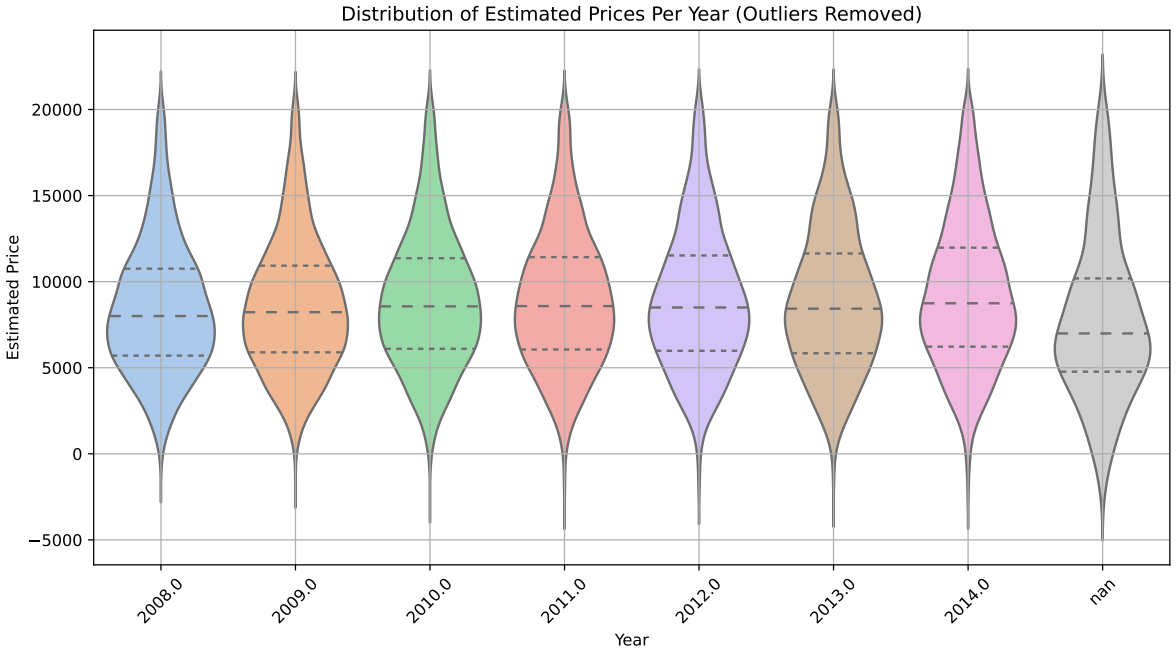
Part 1 Summarizing the Data

The graphs on the next page provide a small summary of the data. Violin plots were used in order to display the general distribution of the data for each given year, providing better insights into the data. As the violin plots show distributions, I took the log of the data so that while the shape was maintained, outliers did not distort the graphs and hinder interpretation.


Hospitals with Multiple Charges Per Year
Number of Hospitals with Multiple Charges
120
110
100
90
80
70
2008
2009
2010
2011
2012
2013
2014
Year


Number of Unique Hospitals Per Year
Number of Unique Hospitals
6000
5980
5960
5940
5920
5900
5880
2008
2009
2010
2011
2012
2013
2014
Year





As you can see, only the years 2008 through 2014 are visible. You can also see that there is distribution of NaN on the final graph. This is an issue that needs to be furhter addressed, and is most likely occuring to due to faulty code.

Part 2: Estimating ATE

For the rest of the assignment, we are only working within the year 2012.

Question 5

First we begin by penalizing the hospitals that have a negative HRRP and HBVP sum and calculate their respective mean price:

```
Average price for penalized hospitals in 2012: 10721.38
Average price for non-penalized hospitals in 2012: 9423.59
```

As you can see, non-penalized hospitals have a lower mean price than penalized hospitals.

Question 6

We then sort all hospitals into 4 quartiles based on the number of beds provided.

Quartile	Penalized_Mean_Price	Non_penalized_Mean_Price
1	8008.29	7456.69
2	10186.6	8522.48
3	11260.1	9470.05
4	13518.2	12254.5

As shown above, penalized hospitals have a higher mean price across all quartiles.

Quesiton 7:

Now we run a regression 4 times, using 4 different regression models, and compare the results.

	INV	MAH	IPW	OLS
ATE	1313.77	1313.77	1313.77	1313.77
SE	280.21	280.21	280.11	279.96

Question 8

Across all four regressions, the results are identical, but the standard errors are different. The standard error variance makes sense, as each regression method handles clutering and correlation in a data differently, and also assigns different weights while regressing.

Question 9

I do not believe that I accuratly estimated causal effect of the penalty. There are many more variables that impact prices that must be taken into account, such as the amount paid to physicians, the CoL for the area in which the hospital resides, etc. Such a simple regression is not able to establish a causal effect of the penalty on hospital pricing.

Question 10

Working with this data was quite the struggle. It took me a long time to clean the data properly, and for hours I was not able to get any values for neither payment types. Almost everything aggrivated me at some point, but few things are as rewarding as when everything clicks and my code runs and outputs a clean results. Moving forward, I need to pay much closer attention to how my code cleans the data, so that I do not struggle like I did here.