

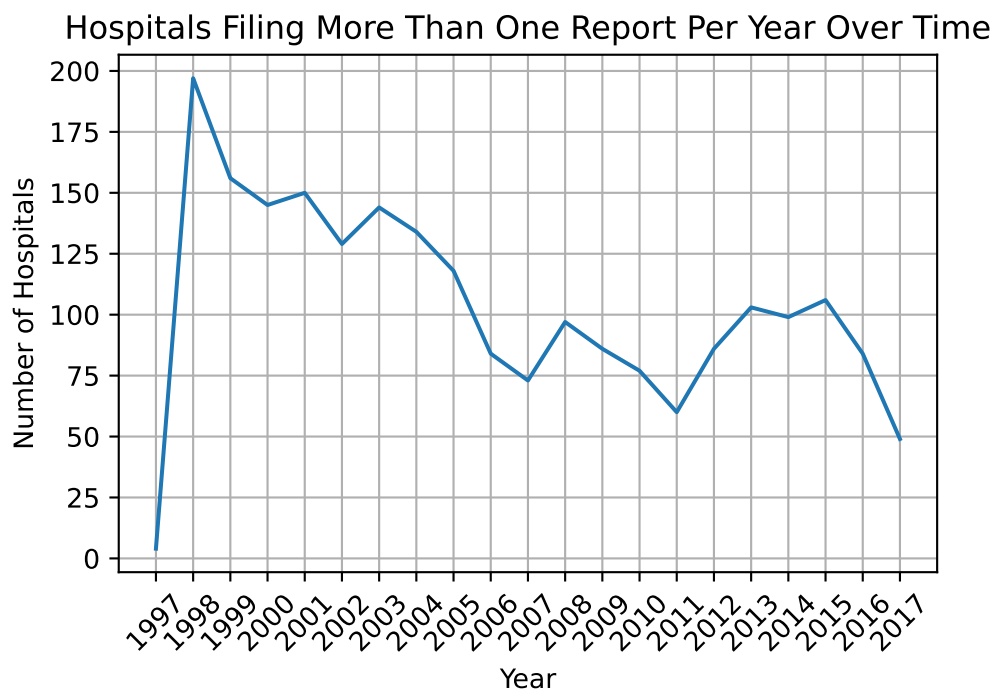
Homework 2

Research Methods, Spring 2025

Ryan Scholte

You can access the [Repository](#)

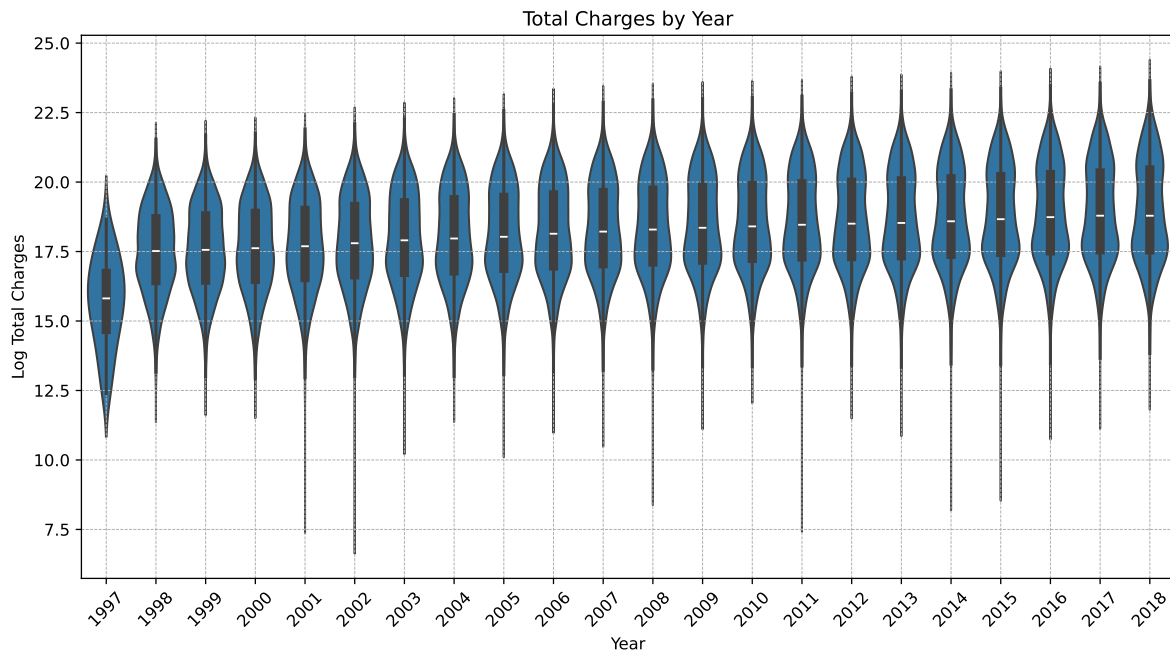
Question 1



Question 2

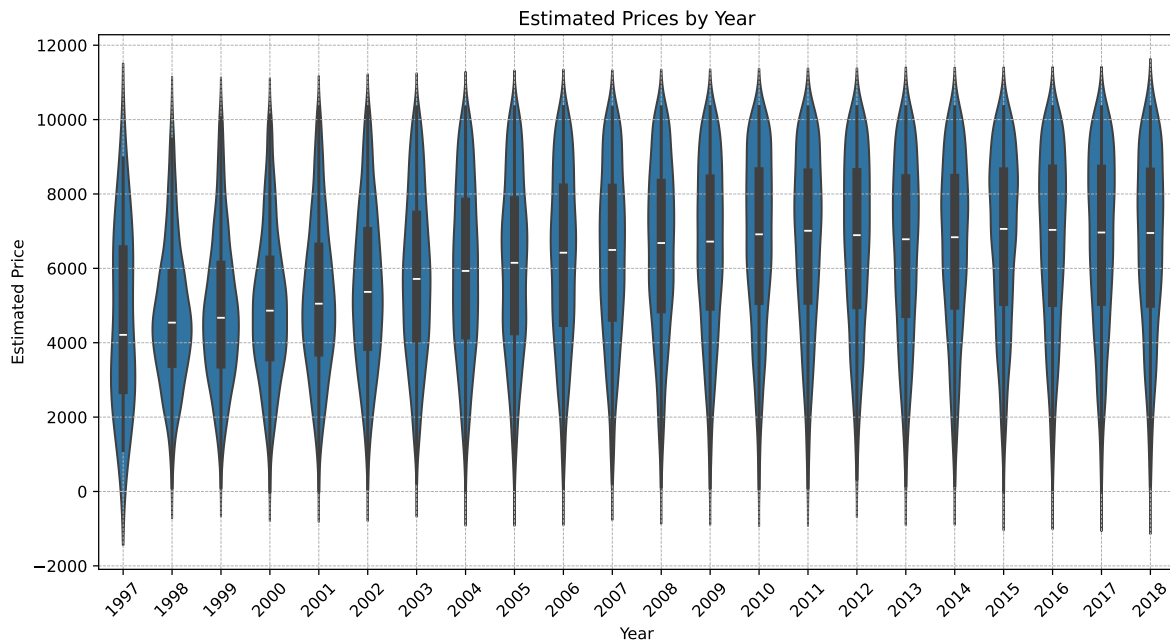
Number of unique hospital IDs: 9323

Question 3



I filtered the data by: dropped all na, took absolute value of all charges, limited to charges above 1000, limited to charges in the bottom 90% to cut out outliers, used a log scale to represent

Question 4



I filtered the data by: dropped all na, calculated estimated price following slides, filtered to only positive value of estimate prices, limited to charges in the bottom 75% to cut out outliers

Question 5

Mean price for penalized hospitals: 10171.54

Mean price for non-penalized hospitals: 9651.82

Question 6

Quartile	Penalized_Mean_Price	Non_penalized_Mean_Price
1	7581.19	7871.93
2	9835.01	8584.81
3	9982.35	9930.10
4	12435.12	12263.70

Question 7

	INV	MAH	IPW	OLS
ATE	299.275537	299.275537	299.275537	299.275537
SE	624.023041	624.023041	573.963946	573.857613

8

all ATE are the same. The ATE assumes that a penalized hospital will have a \$299 increase in price when accounting for number of hospital beds.

9

I think there are many other issues we talked about that we cant assume to determine the causality between a hospital being penalized increasing the estimated price. Those can be unobserved factors like patient acuity or hospital location that affect the chance of a hospital being penalized. This means penalized hospitals arent assigned randomly and have other factors that are not fully controlled for.

10

It was very difficult especially if the cleaning is not done correctly. Also using a bunch new packages is tough. I learned that file paths are super frustrating and can change all the time but there a bunch of differeent solutions and I haven't found a consistent or one I have learned to use consistently yet. I also learned after talking with classmates and Pablo there is almost always a simpler way than what you thought or tried even if you get the same answer. I often did manual calculations rather than using the packages properly. Getting familiar with different packages and capabilities can save a bunch of time if you invest in that at the start.