Homework Assignment

Martinna Roldan

February 28, 2025

1 Summarize the Data

1.1 Question 1

How many hospitals filed more than one report in the same year? The answer is shown in the graph below.

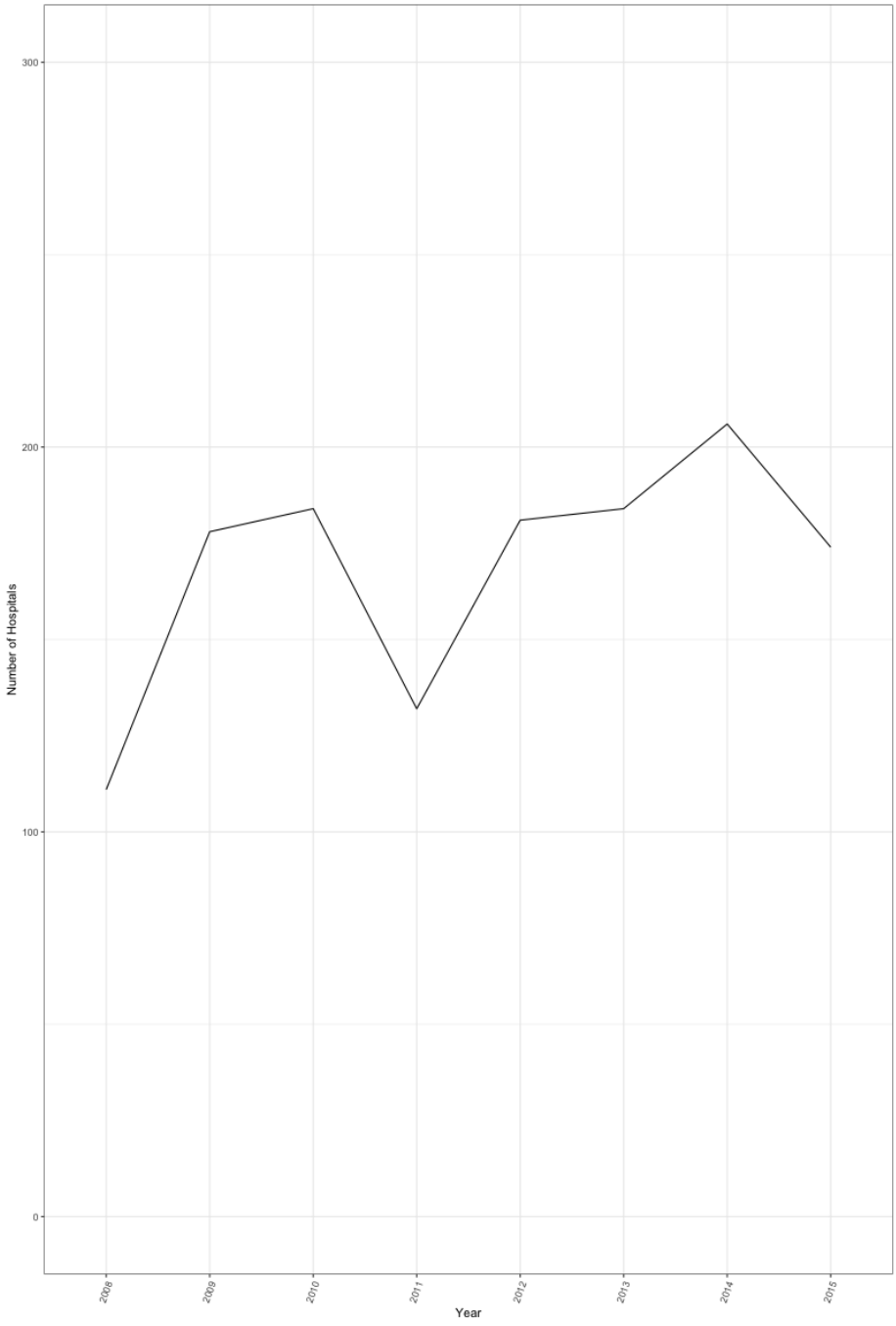

Figure 1: Number of hospitals filing more than one report over time.

1.2 Question 2

The number of unique hospital IDs (Medicare provider numbers) in the data after removing/combining multiple reports is shown in the graph below.

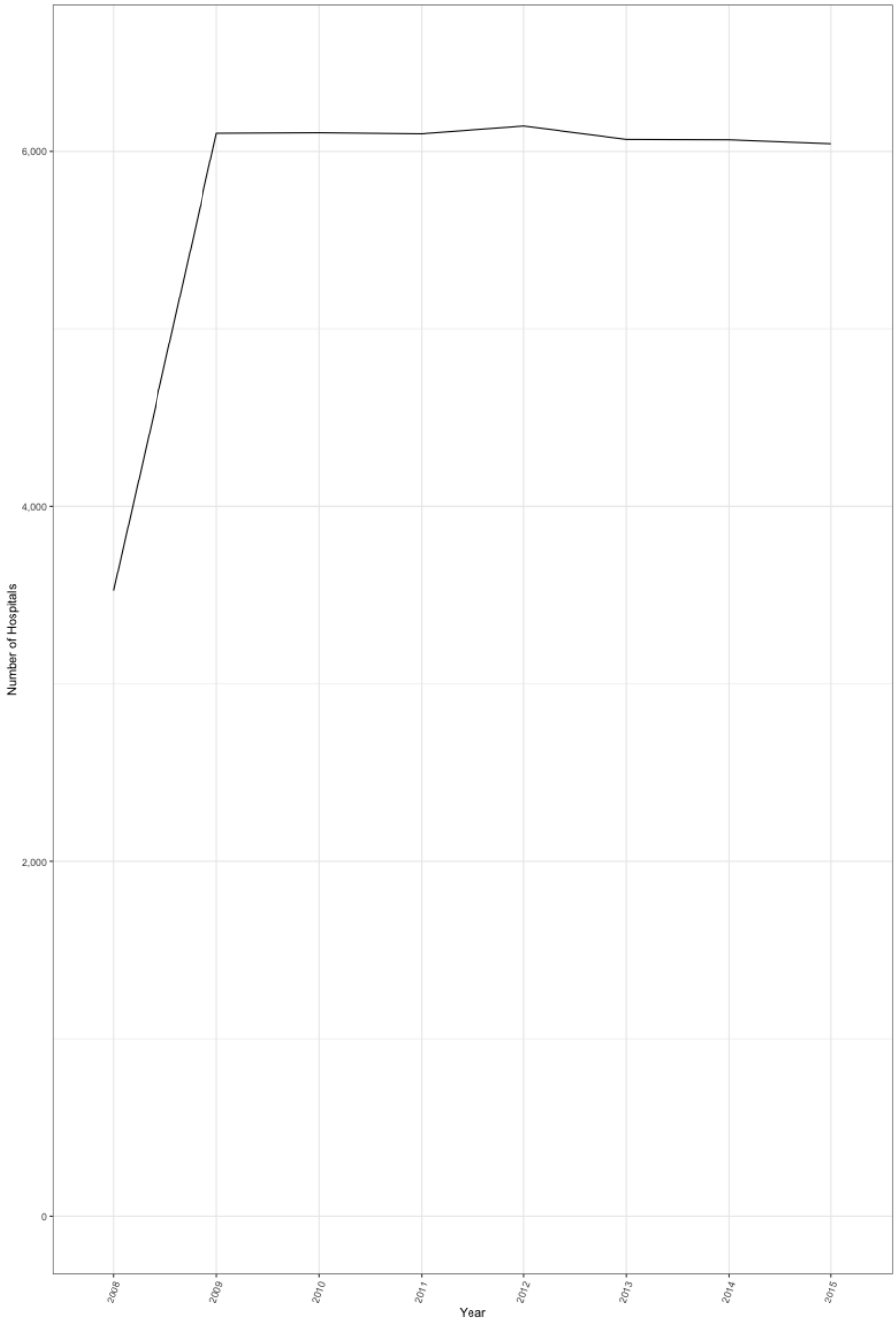

Figure 2: Unique hospital IDs over time.

1.3 Question 3

The distribution of total charges (tot_charges) in each year is shown in the violin plot below.

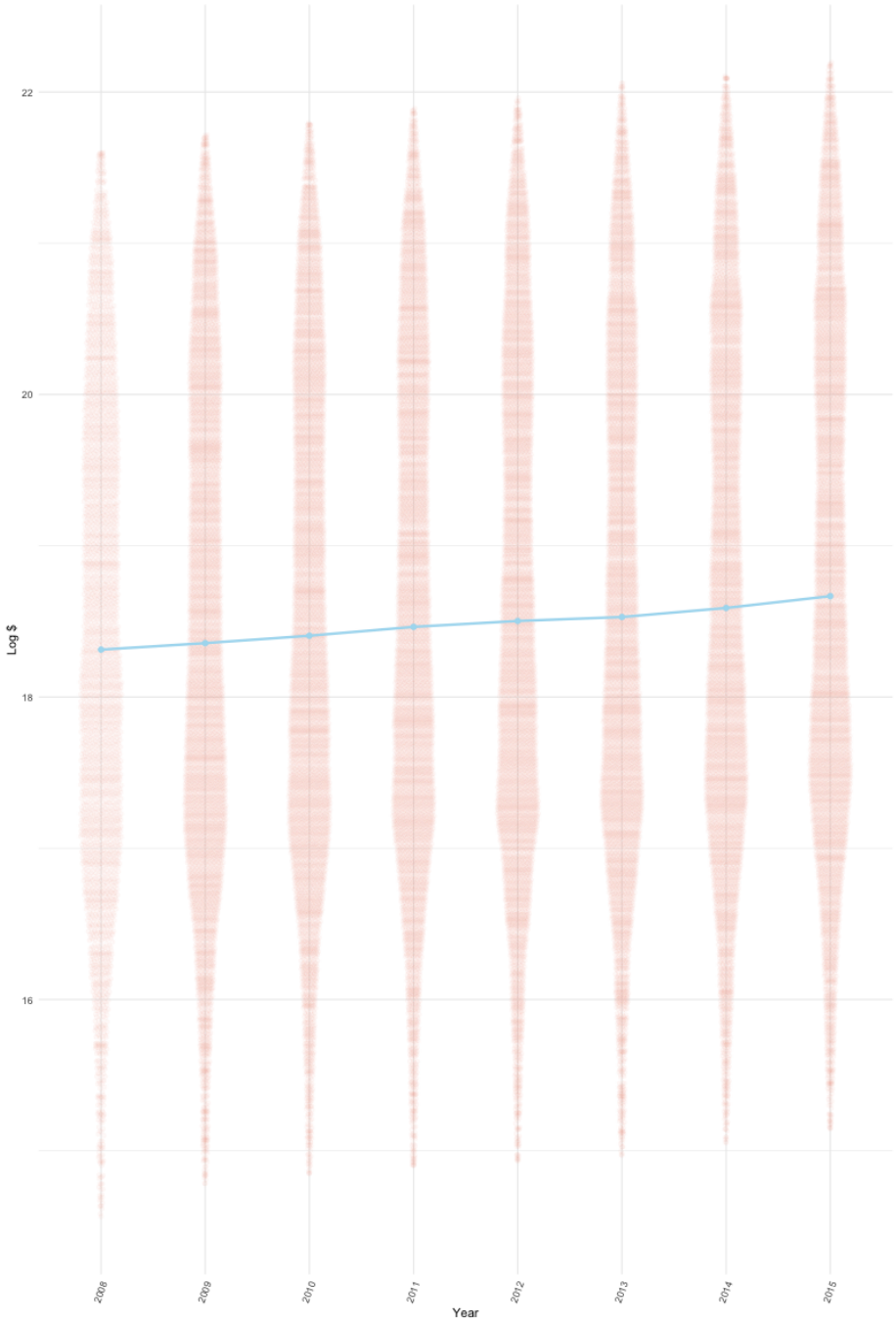

Figure 3: Violin plot of total charges per year.

1.4 Question 4

The distribution of estimated prices in each year is shown in the violin plot below.

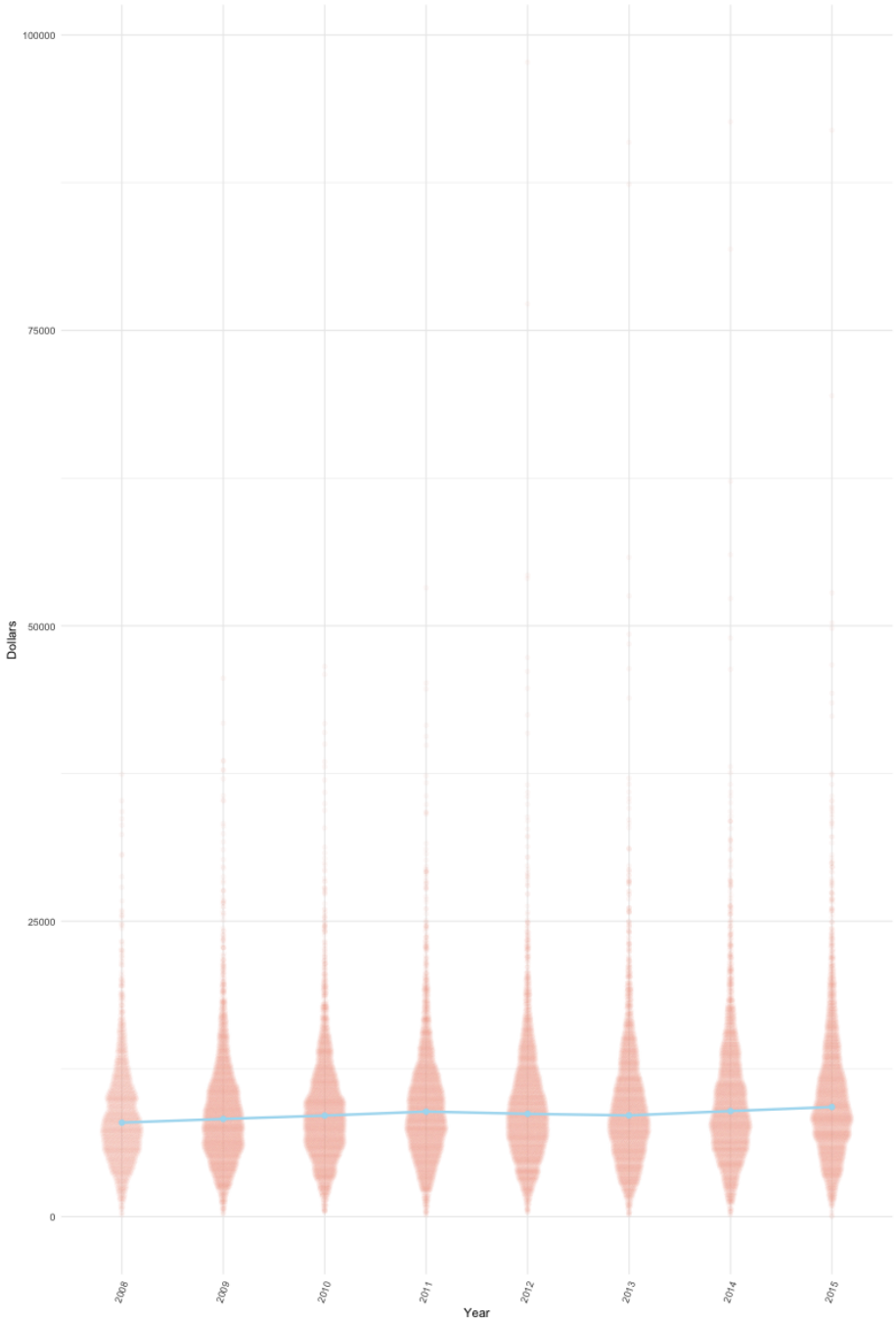

Figure 4: Violin plot of estimated prices per year.

2 Estimate ATEs

2.1 Question 5

The average price among penalized versus non-penalized hospitals is presented in the table below.

Group	Mean Price
Non-Penalized	9550.936
Penalized	9881.975

Table 1: Average price comparison between penalized and non-penalized hospitals.

2.2 Question 6

Below is a table showing the average price among treated/control groups based on bed size quartiles.

Bed Quartile	Control Avg. Price (price_0)	Treated Avg. Price (price_1)
1	7474.241	8730.965
2	8453.519	8751.843
3	9839.936	10141.490
4	12416.085	11890.251

Table 2: Average prices by bed size quartiles.

2.3 Question 7

The table below presents the average treatment effect using different estimators.

	INV	MAH	IPW	OLS
Penalty	331.039	331.039	331.039	331.039
	(222.605)	(222.605)	(211.467)	(226.979)

Table 3: ATE estimates using different estimators.

2.4 Question 8

The results of the different estimators are identical across all methods. Each estimator—Inverse Variance Matching (INV), Mahalanobis Distance Matching (MAH), Inverse Propensity Weighting (IPW), and Ordinary Least Squares (OLS)—produced the same ATE estimate of 331.039 with slight variations in standard errors. This consistency suggests that the treatment effect estimation is stable across different methodologies.

2.5 Question 9

The estimates suggest a consistent association between the penalty and hospital pricing, but they do not necessarily indicate a causal effect. One key potential confounding factor is the severity and quality of patients treated at different hospitals. Some hospitals may serve a disproportionately sicker patient population, leading to higher costs and potentially higher penalties. If this factor is not fully accounted for in the matching or regression models, the observed association between penalties and pricing could be partially driven by differences in patient health rather than the penalty itself. Therefore, while the estimates are robust across methods, causality remains uncertain due to possible unobserved confounders.

2.6 Question 10

I learned that working with HCRIS data requires handling a variety of inconsistencies, including duplicate reports due to hospitals changing their fiscal years and missing variables in certain years. The presence of two form versions (1996 and 2010) also introduces challenges in ensuring consistency across different reporting formats. Despite CMS's efforts to provide accurate and comprehensive data, the process of validation and verification means that data accuracy is only guaranteed at the point of reporting, and derived conclusions must account for potential discrepancies. One of the biggest challenges was the inconsistency in available columns across the two forms and the impact of removing certain information to create a standardized dataset, particularly during matching. The 1996 and 2010 versions contained different sets of variables, requiring trade-offs between preserving key variables for matching. This raised concerns about potential bias in the estimates, as excluding certain variables could alter the comparability of hospitals and affect the accuracy of treatment effect estimations.