Homework 2
Research in Health Economics, Spring 2024

Ria Dharnidharka

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.
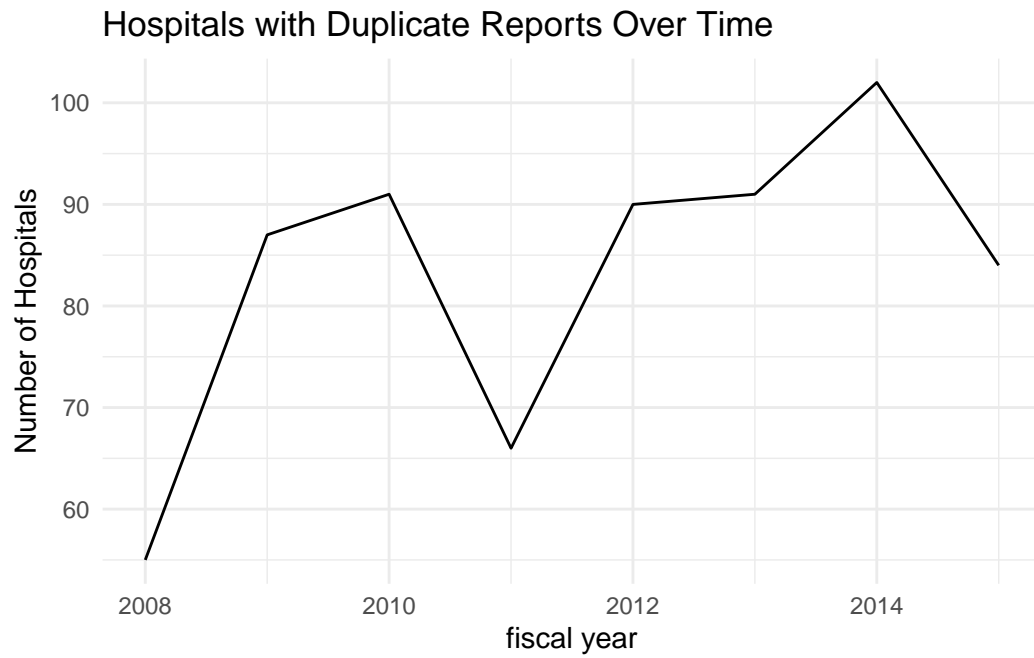

Figure 1: Hospitals with Duplicates

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

There are 48803 unique hospital ids in the dataset

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. For a nice tutorial on violin plots, look at Violin Plots with ggplot2.

```
Warning: Removed 2578 rows containing non-finite values (`stat_ydensity()`).
```
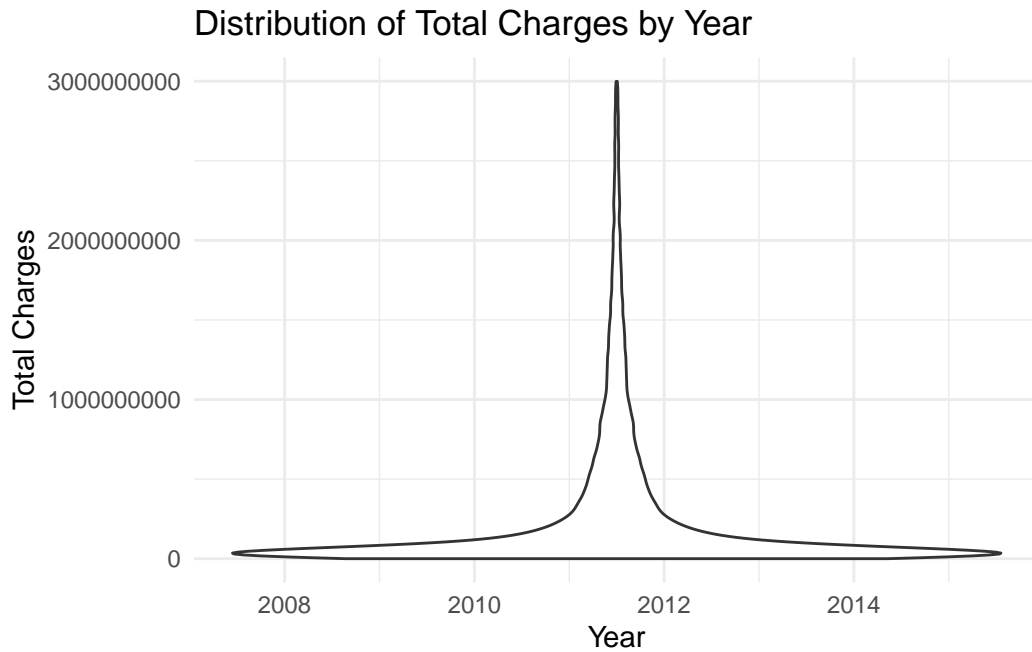

Figure 2: Total Charges

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.

```
Warning: Removed 20 rows containing non-finite values (`stat_ydensity()`).
```
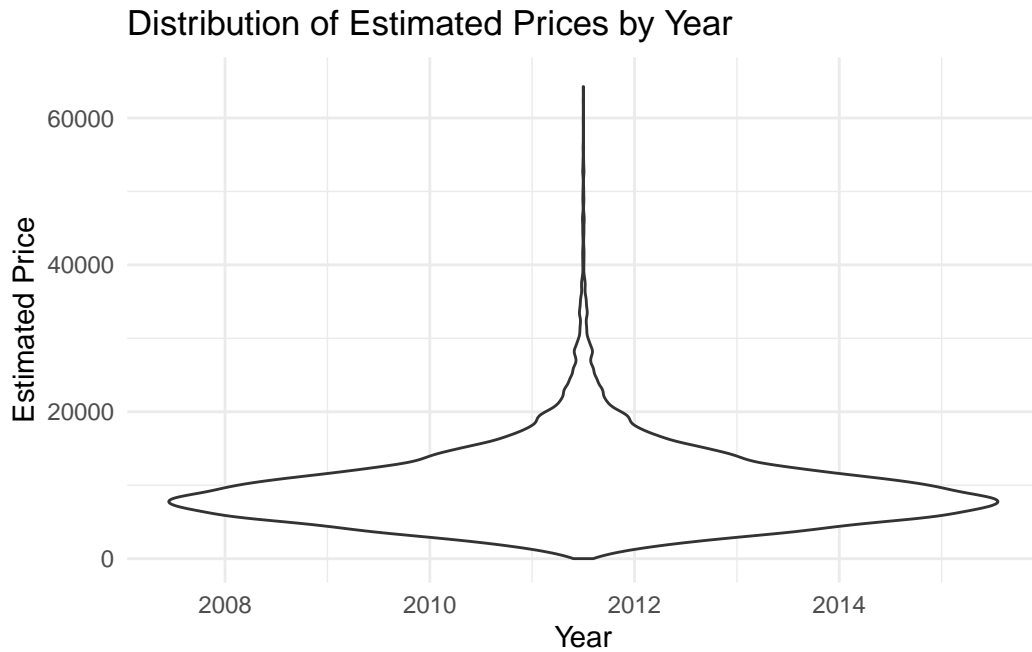
Distribution of Estimated Prices by Year

Figure 3: Estimated Prices