

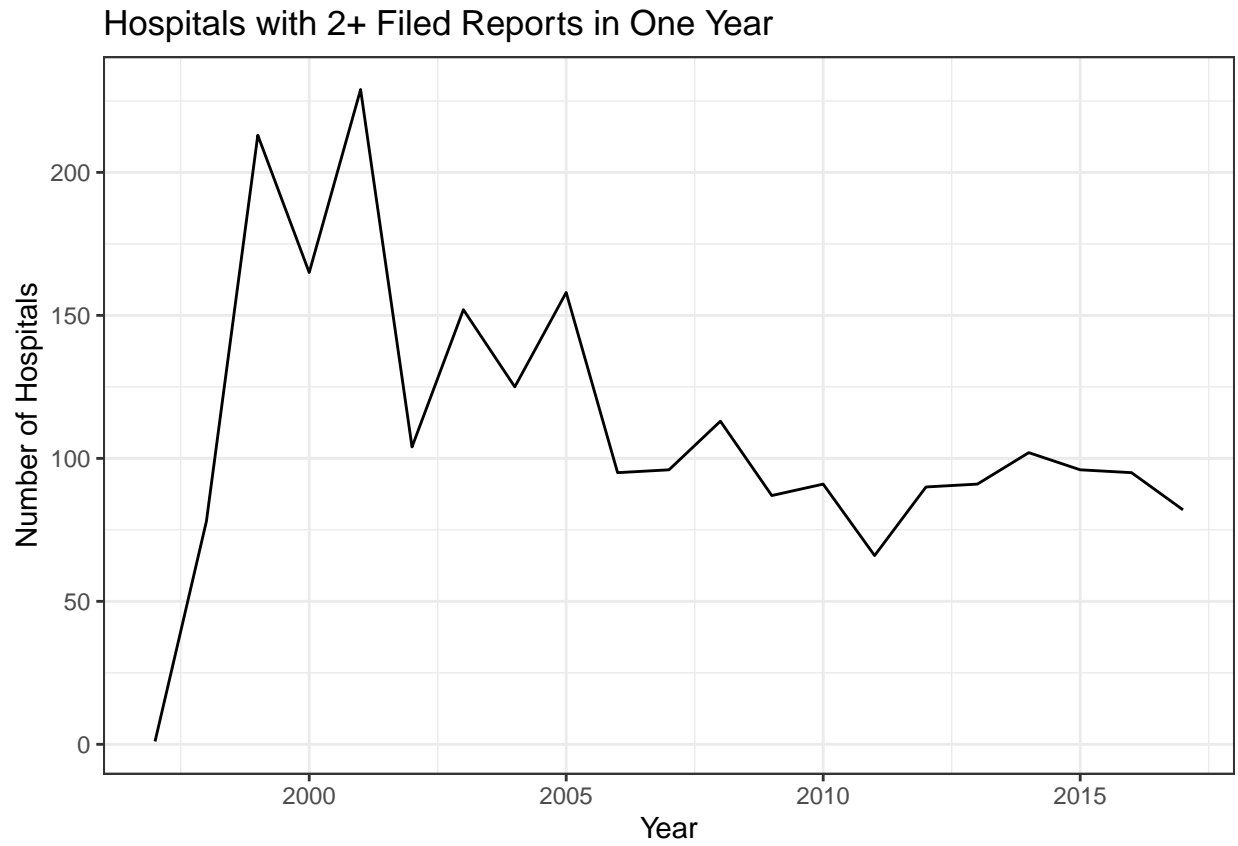
# Homework 2

Virginia Sanson

17 February 2023

### Question 1

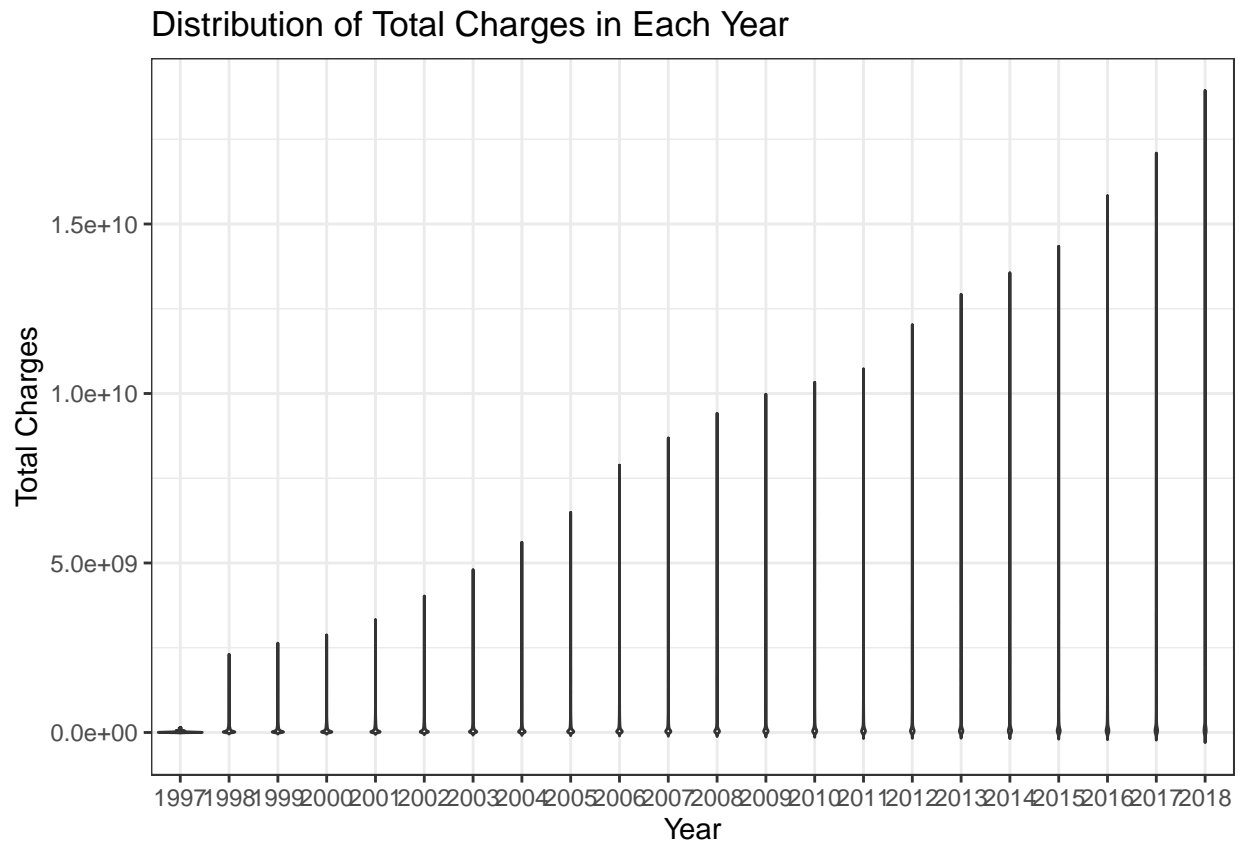
2,329 hospitals filed more than one report in the same year, from 1997 to 2018.



## Question 2

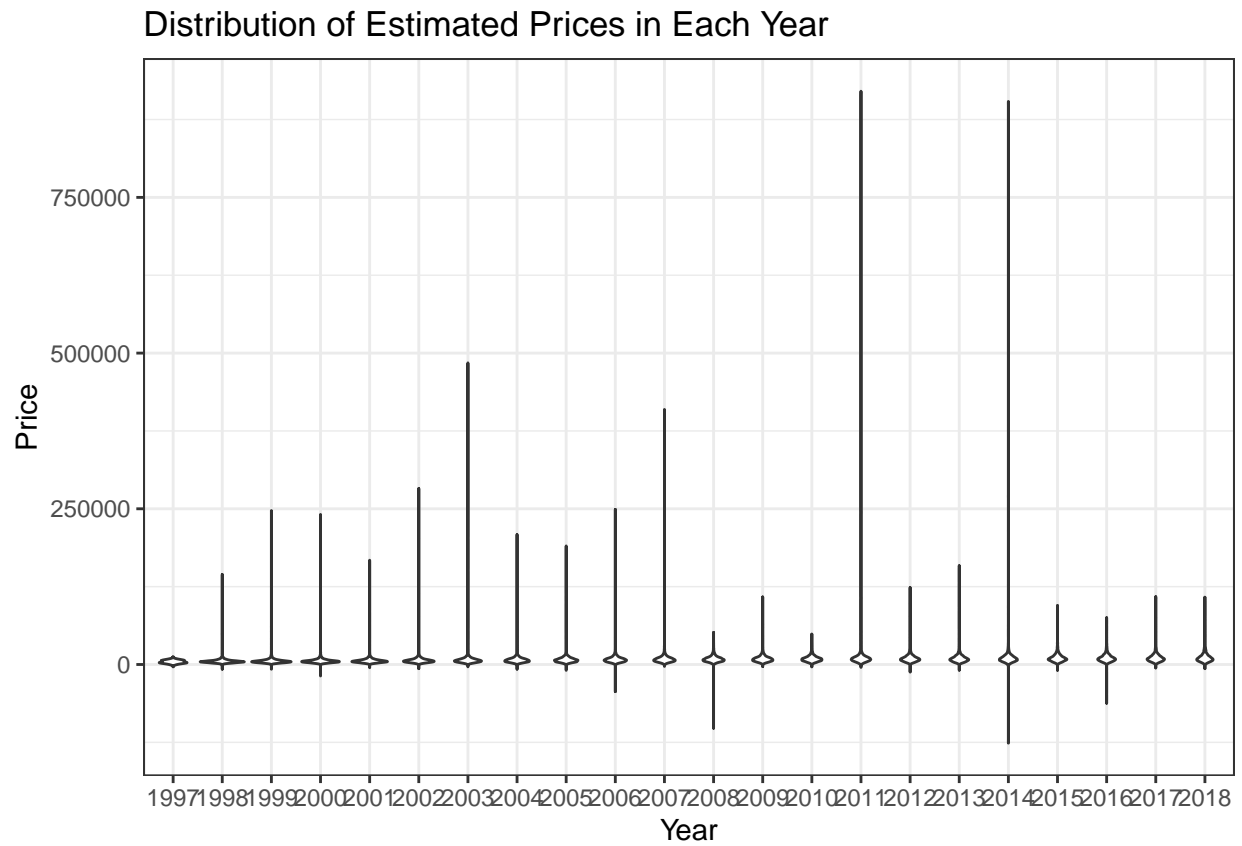
The total number of unique hospital IDs in the dataset is 9323.

Question 3



#### Question 4

The negative values for price on this y-axis made me realize that the data needed to be filtered for positive values. This was done thereafter.



### Question 5

Prices were filtered to be positive values and penalties were limited to below 100,000 to eliminate outliers in the data.

```
## # A tibble: 2 x 2
##   penalty price
##   <dbl> <dbl>
## 1      0 9791.
## 2      1 10235.
```

# Question 6

```
## # A tibble: 8 x 3
## # Groups:   penalty [2]
##   penalty quartile avg_price
##   <dbl>    <int>    <dbl>
## 1      0        1      8482.
## 2      0        2      8361.
## 3      0        3     10521.
## 4      0        4     11749.
## 5      1        1      7653.
## 6      1        2     10833.
## 7      1        3      9339.
## 8      1        4     12435.
```

### Question 7

```
##          Inverse.Variance Mahalanobis Inverse.P propensity.Weight
## penalty          286.4789      286.4789          286.4789
##          Simple.Regression
## penalty          286.4789
```



#### Question 8

With these different treatment effect estimators, the results are identical for the inverse variance distance based matching, Mahalanobis distance based matching, and the inverse propensity score weighting. These three estimators yield an average treatment effect (ATE) of 286.5. The result from the simple linear regression, however, is similar but not identical as it yields an ATE of 287.

#### Question 9

Propensity score weighting and matching based on inverse variance distance and Mahalanobis distance are tools for causal inference, and as such, I believe we are estimating a causal effect of the penalty on hospitals. These tools reduce the effects of confounding variables, and the simple linear regression model provided draws a causal relationship between two variables.

#### Question 10

My experience working with this data set was a lot better than last homework's, however some parts are definitely still time-consuming. I learned how to create a violin plot in R; my previous work with violin plots was very basic ones in Python instead. However, I still wish I could make it a more aesthetically pleasing graph with graphics on density. Question 7 continued to aggravate me as I was initially getting could not understand why the linear regression result was very different from the three previous estimators.