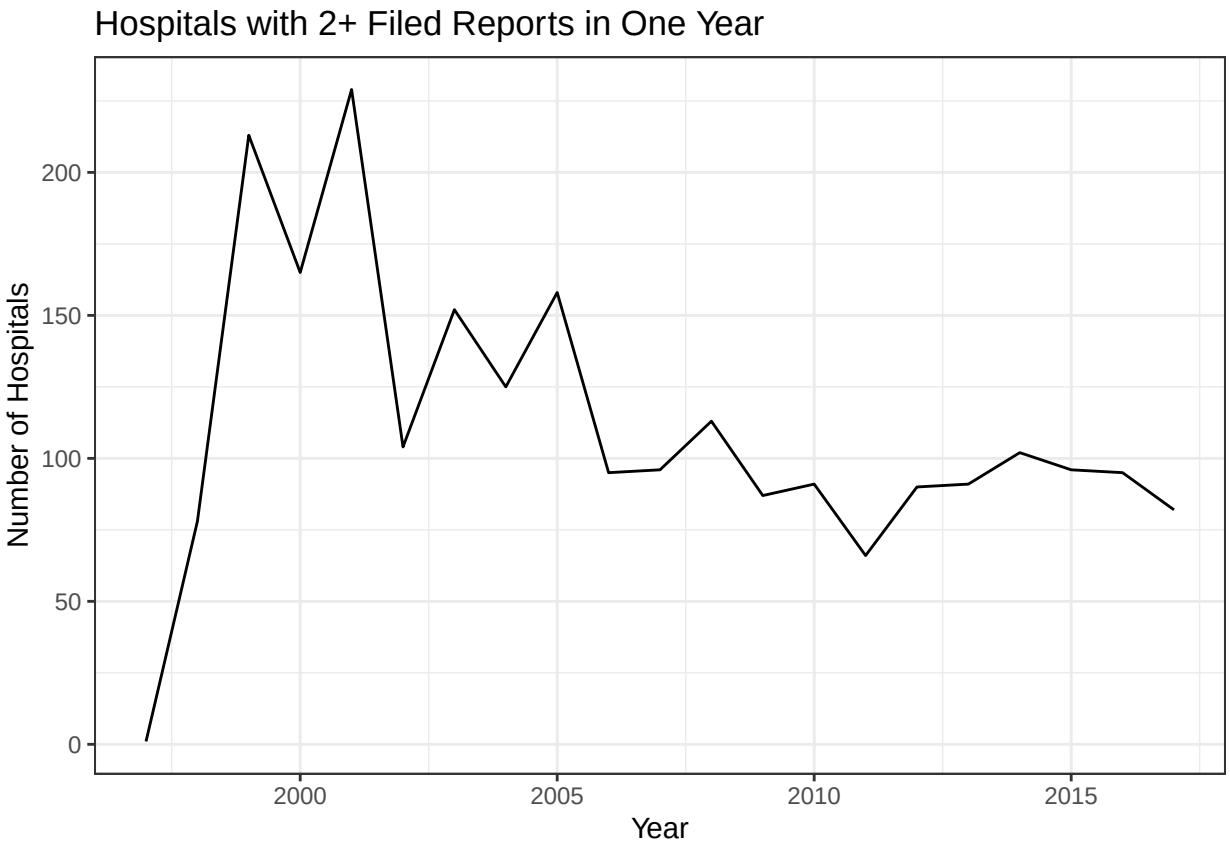
Homework 2

Virginia Sanson

17 February 2023

Question 1

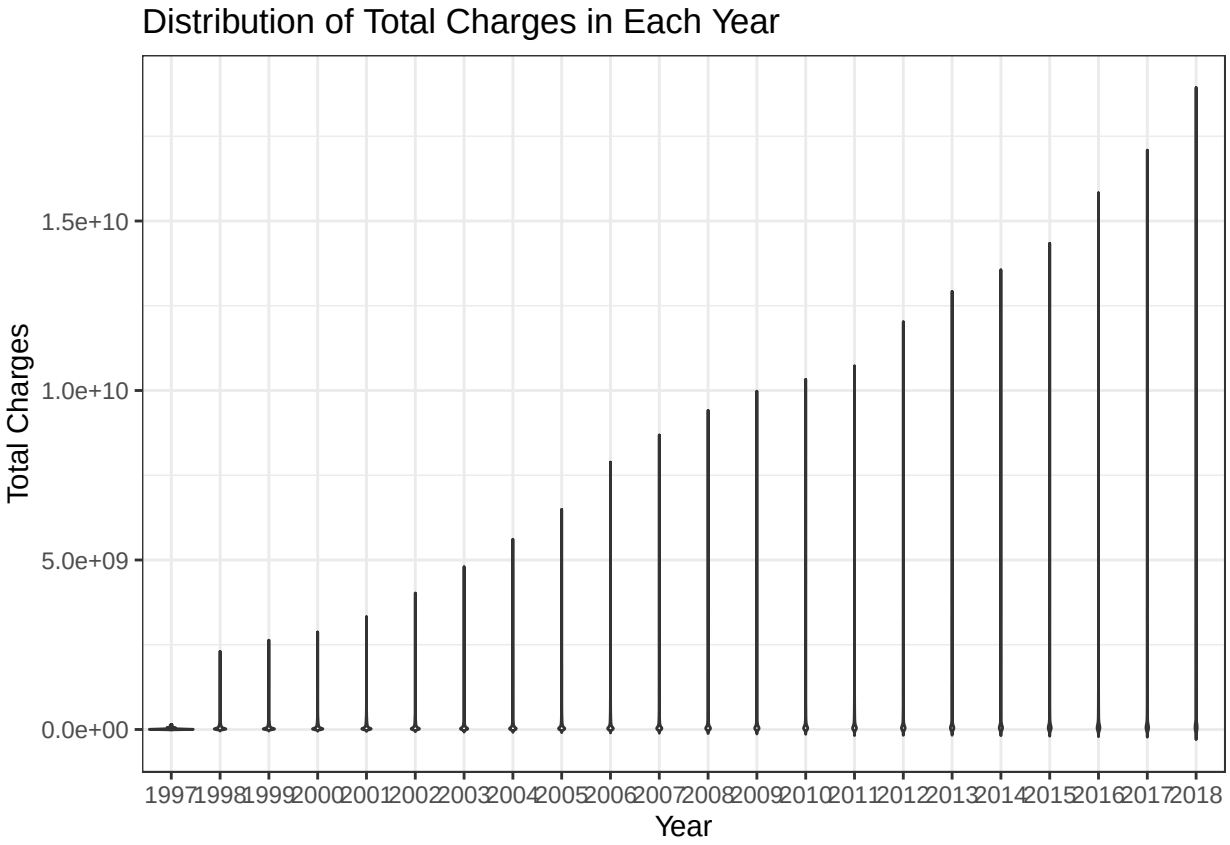
2,329 hospitals filed more than one report in the same year, from 1997 to 2018.

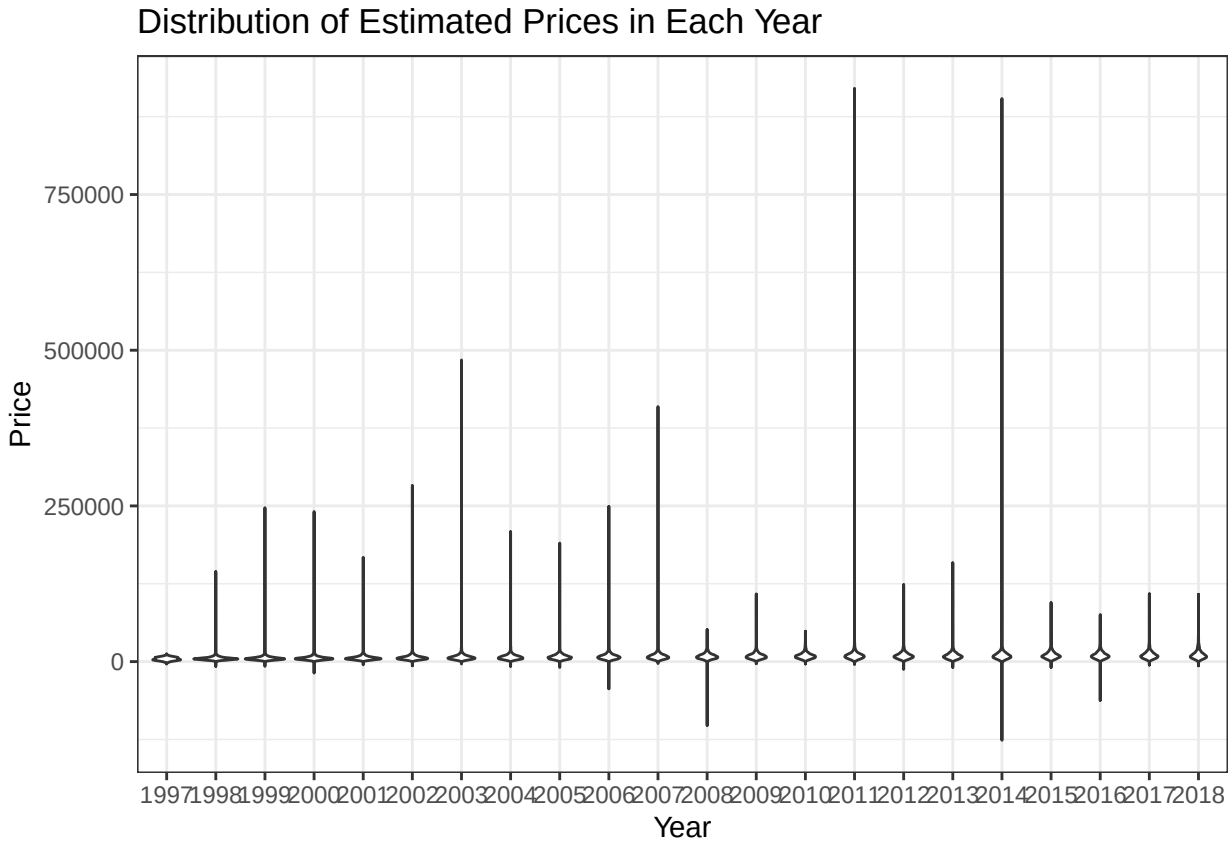


Question 2

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

Question 3


Distribution of Total Charges in Each Year
Total Charges
1.5e+10
1.0e+10
5.0e+09
0.0e+00
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Year


Distribution of Estimated Prices in Each Year
Price
750000
500000
250000
0
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Year

Question 4

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 Part A

Estimate... 286.48 AI SE..... 681.72 T-stat..... 0.42023 p.val..... 0.67432

Original number of observations..... 671 Original number of treated obs..... 85 Matched number of observations..... 671 Matched number of observations (unweighted). 24724

```
##                   Length Class  Mode
## est                    1 -none- numeric
## se                     1 -none- numeric
## est.noadj              1 -none- numeric
## se.standard            1 -none- numeric
## se.cond                1 -none- numeric
## mdata                  4 -none- list
## index.treated      24724 -none- numeric
## index.control      24724 -none- numeric
## index.dropped          0 -none- NULL
## weights            24724 -none- numeric
## orig.nobs              1 -none- numeric
## orig.wnobs             1 -none- numeric
## orig.treated.nobs      1 -none- numeric
## nobs                   1 -none- numeric
## wnobs                  1 -none- numeric
## caliper                0 -none- NULL
## ecaliper               0 -none- NULL
## exact                  0 -none- NULL
## ndrops                 1 -none- numeric
## ndrops.matches         1 -none- numeric
## MatchLoopC        123620 -none- numeric
## version                1 -none- character
## estimand               1 -none- character
```

Question 7 Part B

Estimate... 286.48 AI SE..... 681.72 T-stat..... 0.42023 p.val..... 0.67432

Original number of observations..... 671 Original number of treated obs..... 85 Matched number of observations..... 671 Matched number of observations (unweighted). 24724

```
##                   Length Class  Mode
## est                    1 -none- numeric
## se                     1 -none- numeric
## est.noadj              1 -none- numeric
## se.standard            1 -none- numeric
## se.cond                1 -none- numeric
## mdata                  4 -none- list
## index.treated      24724 -none- numeric
## index.control      24724 -none- numeric
## index.dropped          0 -none- NULL
## weights            24724 -none- numeric
## orig.nobs              1 -none- numeric
## orig.wnobs             1 -none- numeric
## orig.treated.nobs      1 -none- numeric
## nobs                   1 -none- numeric
## wnobs                  1 -none- numeric
## caliper                0 -none- NULL
## ecaliper               0 -none- NULL
## exact                  0 -none- NULL
## ndrops                 1 -none- numeric
## ndrops.matches         1 -none- numeric
## MatchLoopC        123620 -none- numeric
## version                1 -none- character
## estimand               1 -none- character
```

Question 7 Part C

```
##
## Call:
## lm(formula = price ~ penalty, data = obs_2012, weights = ipw)
##
## Coefficients:
## (Intercept)      penalty
##      9775.2        286.5
```

Question 7 Part D

```
## [1] 443.5711
```

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 not identical but similar as it yields an ATE of ____.

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 intially getting could not understand why I continued to get errors for parts C and D, so there is definitely a knowledge gap of what I wanted to do and what I could.