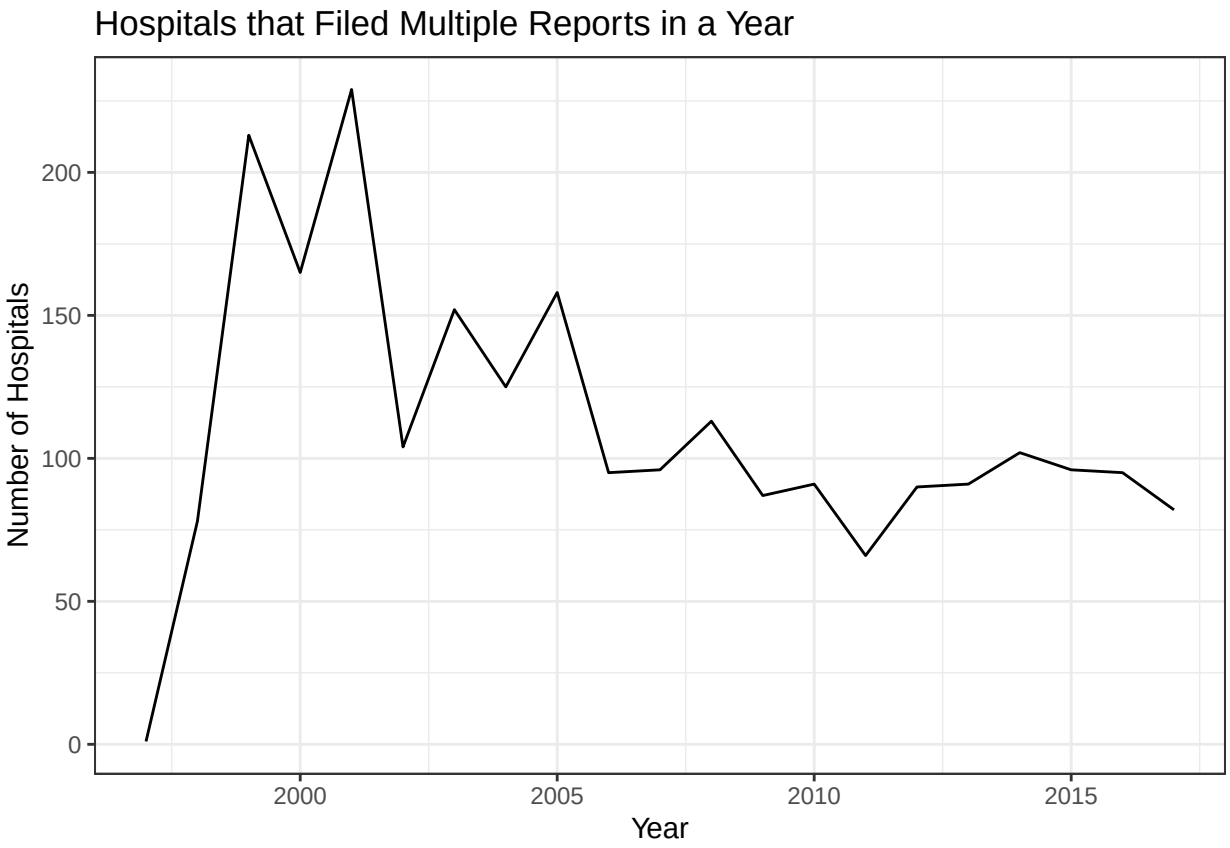
Bhasin-S-hwk2-3

Sachi Bhasin

Question 1

2,329 hospitals filed more than one report in the same year.

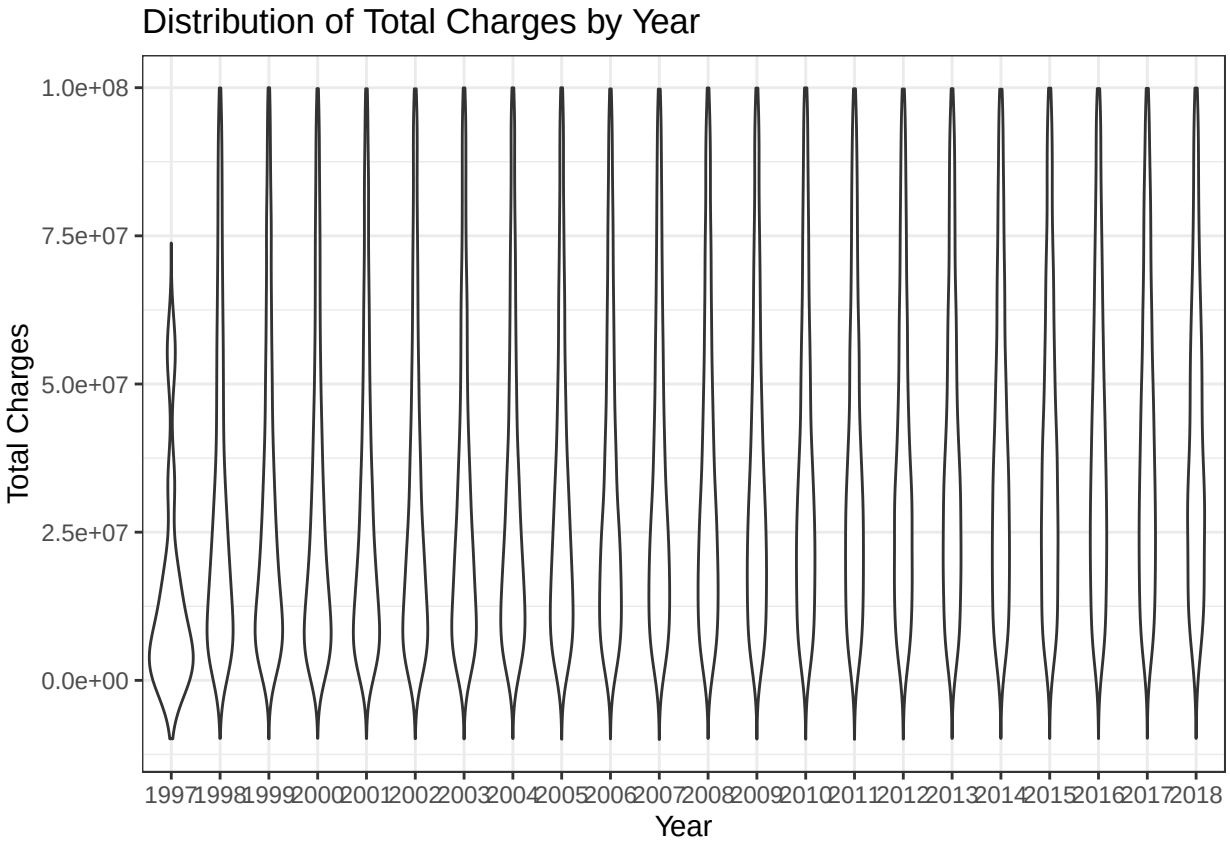


Question 2

After removing/combining multiple reports, there are 9,323 unique hospital IDs in the data.

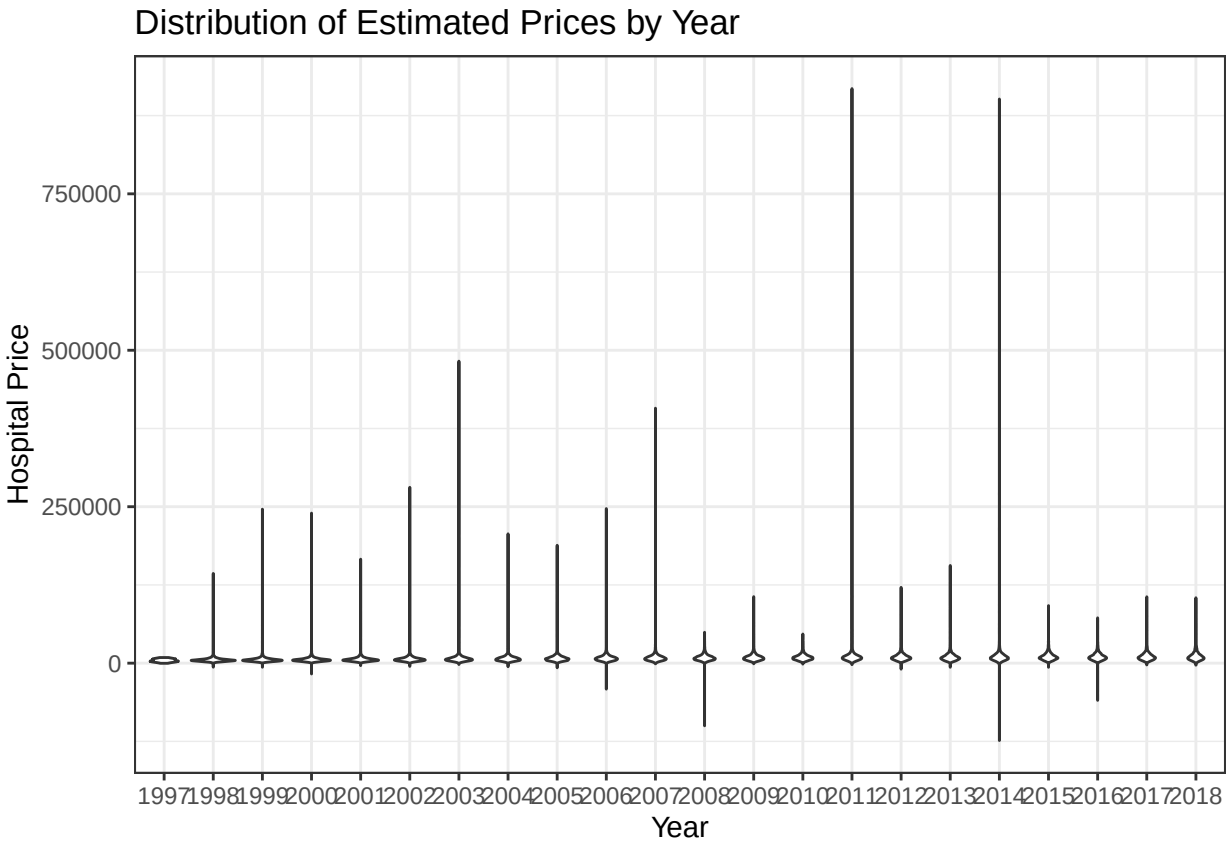
```
## [1] 9323
```

Question 3


Distribution of Total Charges by Year
Total Charges
1.0e+08
7.5e+07
5.0e+07
2.5e+07
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

Question 4

This graph made me realize we had to filter our data since some of the prices were negative and some of tthe values were extreme.



Question 5

Before calculating the average price among penalized versus non-penalized hospitals, prices were filtered to be positive and below 100,000 to get rid of outliers.

```
## # 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

```
##         Matching.Inverse.Variance Mahalanobis.Matching
## penalty                  286.4789             286.4789
##         Inverse.Propensity.Weighting Linear.Regression
## penalty                     286.4789          286.4789
```

Question 8

With these different treatment effect estimators, the results are identical.

Question 9

I do not think I have estimated a causal effect of the penalty. There are several variables that affect hospital price and we only controlled for one of them.

Question 10

I found working with this data challenging but easier than homework 1 as I am getting more comfortable trouble shooting and working with this application. I learned how to create a dummy variable and quartiles for a data set. It was very aggravating to troubleshoot the error I kept getting with the propensity score and number 7 in general.

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

0.1 Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.