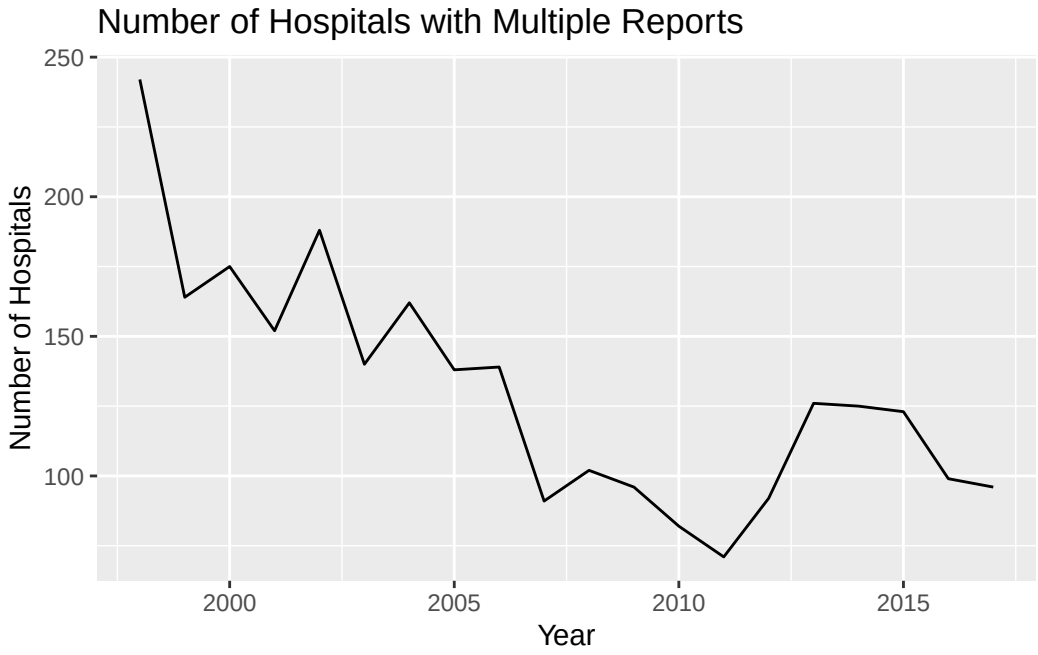
Homework 2

Varun Saxena

Repository: https://github.com/varunsaxena2/saxena-v-hwk2-2/tree/main

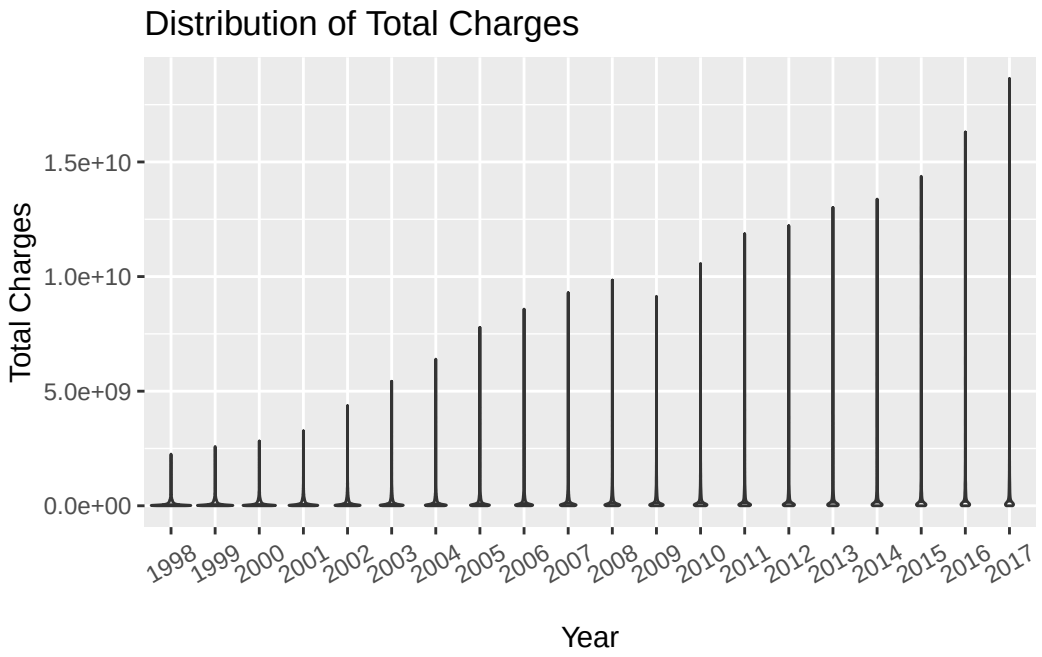
1. Here is the line graph with the number of hospitals filing multiple reports per year

```
Loading required package: pacman
```

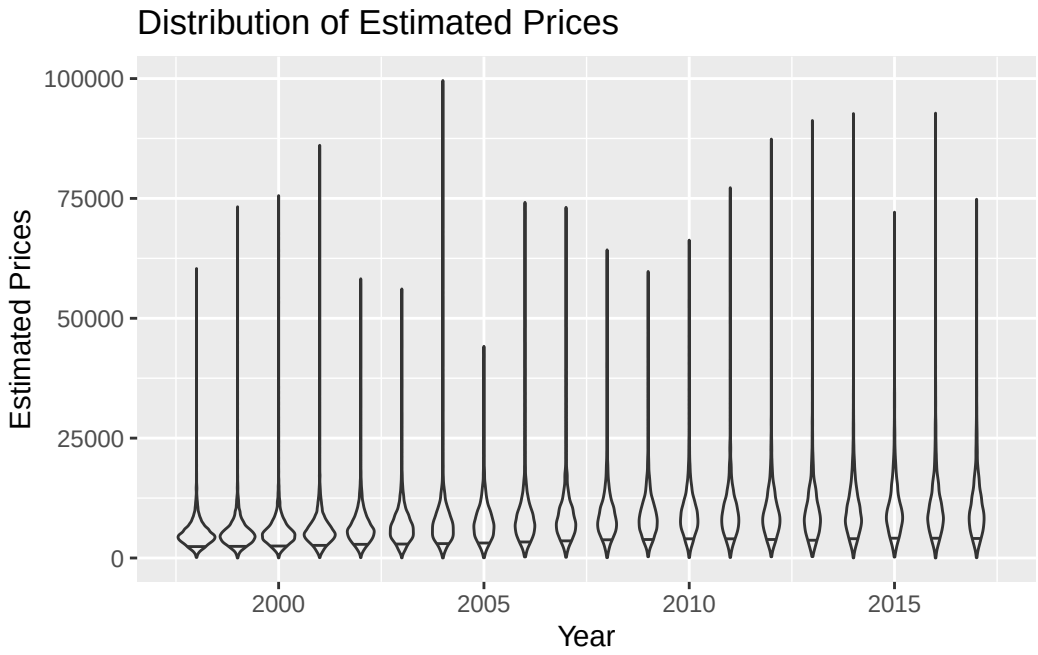


2. The number of unique hospitals based on the hospital ID is 9325.

3.


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

4.


Distribution of Estimated Prices
Estimated Prices
100000
75000
50000
25000
0
2000
2005
2010
2015
Year

5. The average price among penalized hospitals is 9799. The average price among non-penalized hospitals is 9335.

6.

```
# A tibble: 4 x 2
  quartile      b
     <int>  <dbl>
1        1  7628.
2        2  8745.
3        3  9681.
4        4 12324.
```

7.

	Estimator	ATE
1	Nearest Neighbor (Inverse Variance Distance)	88.49866
2	Nearest Neighbor (Mahalanobis Distance)	88.49866
3	Inverse Propensity Weighting	88.49866
4	Simple Linear Regression	88.49866

8. The results are practically identical between the four estimators. This result is quite surprising, as I expected some variance between the estimators.

9. Not exactly. I think that the differences associated with the penalty may or may not be a direct result on the penalty itself. In other words, the causality of this analysis cannot be determined. However, there is a notable discrepancy between the penalized and non-penalized hospitals, so causality cannot be ruled out.

10. So far, I have been able to figure out everything fairly well. The violin plots would ideally be formatted better, but I am still figuring out the best way to do that. Also, the tibble for 6 and datafame for 7 would ideally look neater if I knew how to make them more presentable.