Homework 2

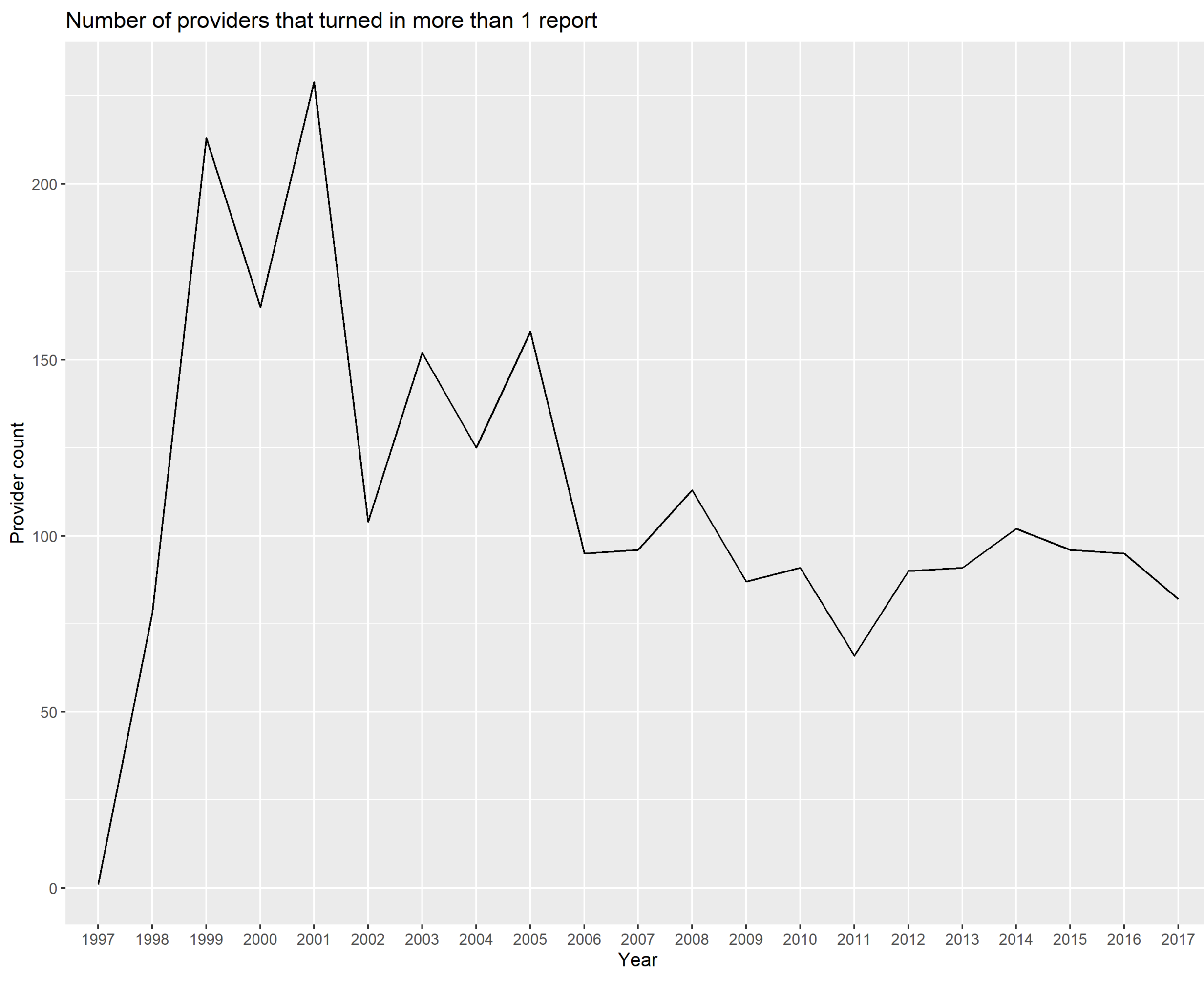
Sam Cohen

2023-02-19

0.1 Answers

0.1.1 1.

There were 2329 providers across the years that submitted more than 1 report in a single year.

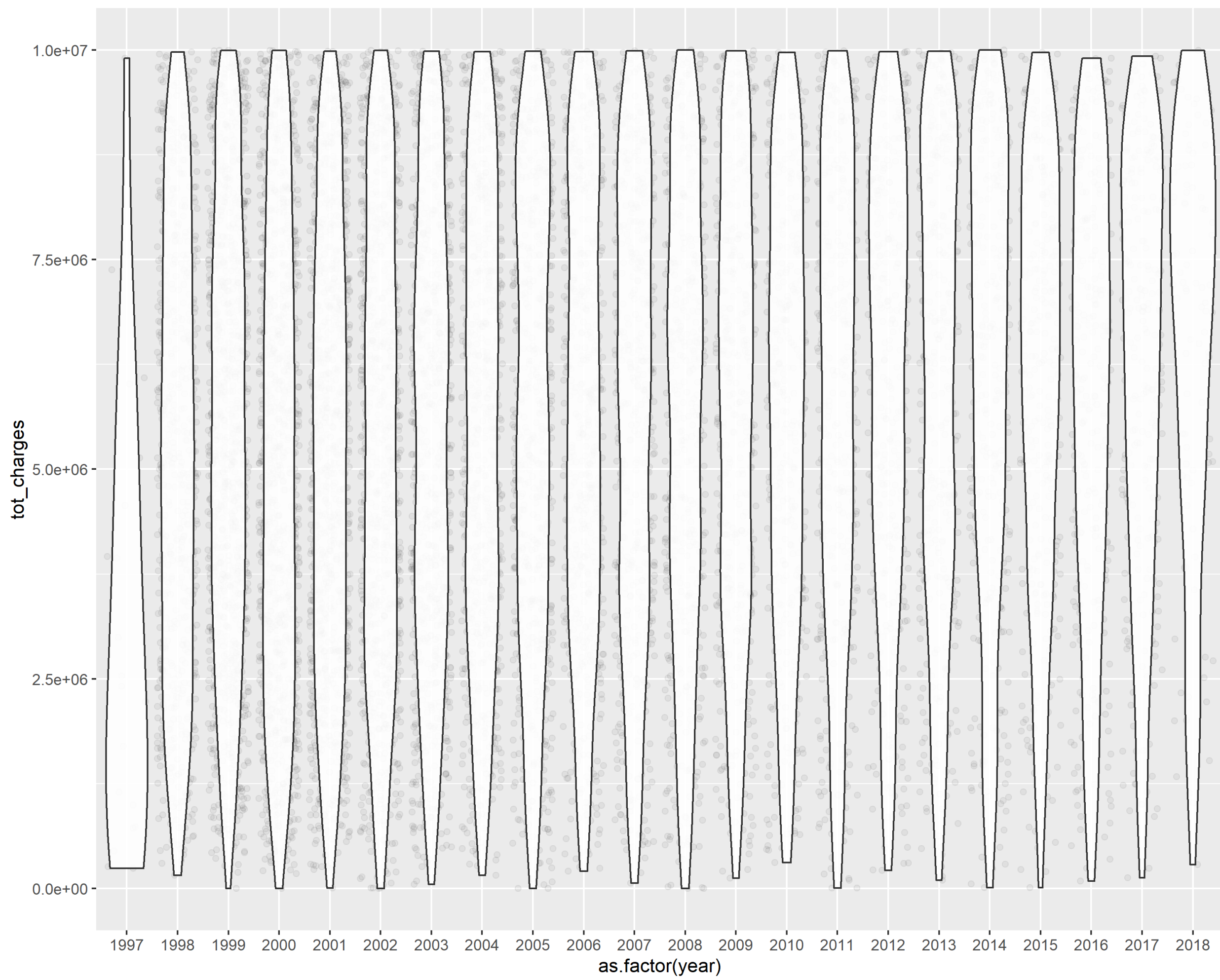


0.1.2 2.

There are 9323 unique provider numbers over the years.

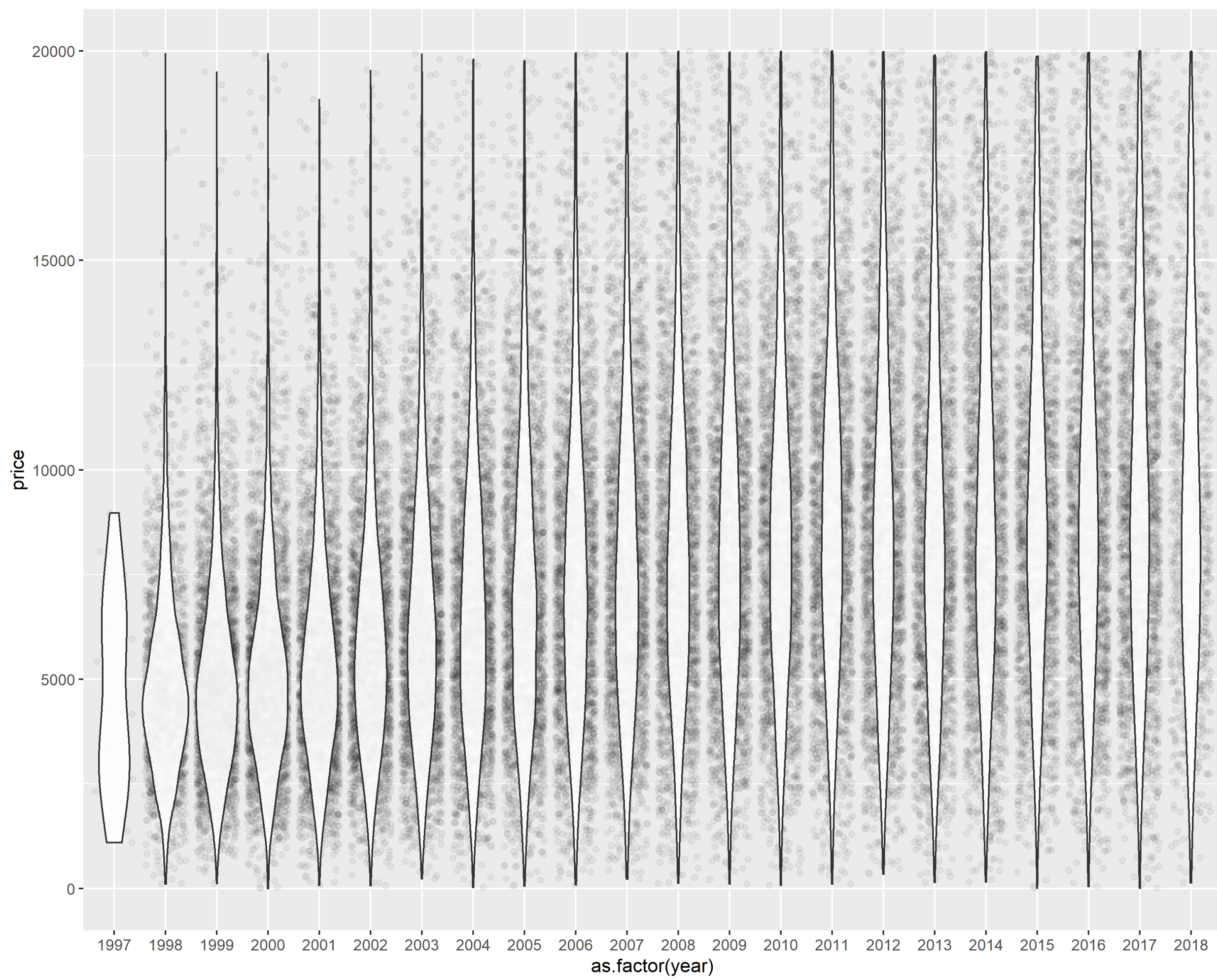
0.1.3 3.

This graph demonstrates the distribution of charges for each year. For the sake of the graph, observations with total charges of over 10,000,000 have been dropped.



0.1.4 4.

This graph represents the distribution of prices over the years, as prices calculated by all charges for a hospital multiplied by the discount factor and subtracting medicare payments, all divided by the total number of discharges outside of medicare patients. For the sake of the graph, observations with the price greater than 20,000 have been excluded.



0.1.5 5

The average price for penalized hospitals is 9,789.33 whereas the average price for non-penalized providers is 9,398.832. In this calculation I dropped all prices over 50,000 as well any prices below 0.

0.1.6 6

```
## `summarise()` has grouped output by 'Quartiles'. You can override using the
## `.groups` argument.
```

Quartiles	mean_price_Control	mean_price_Treatment
1	7598.543	8062.777
2	8507.619	8662.349
3	9728.542	10102.451
4	12247.019	12068.479

0.1.7 7

```
## Warning: package 'knitr' was built under R version 4.2.2
```

	term	estimate	std.error
Linear Regression	penalized	196.794408937835	215.369756455923
Inverse Propensity Weighting	penalized	204.806886033375	195.530680369108
Nearest Neighbor (Inverse Variance)	penalized	204.806886034367	214.139281669362
Nearest Neighbor (Mahalanobis)	penalized	204.806886034367	214.139281669362

0.1.8 8

We see now that each of the estimators is the exact same, except for the linear regression, which is still very close. This makes sense as the others are weighted by the same metric, although in different ways, whereas the linear regression was not weighted.

0.1.9 9

I now feel as though this is a semi-accurate representation of the effect of the penalty on prices. The numbers are similar to just comparing means, but smaller, which can make sense as we are trying to accurately take into account any other factor that can cause a penalty. I always struggle to know if it is truly a causal link here, but if the code works as I think it did, then there is a causal effect of penalties on prices.

0.1.10 10

By the third version, a lot of it makes more sense and I am glad most things came together in a sense. I still struggled the most with 6 I would say in trying to make a nice table as well as trying to see if my code was right for 7 as I did not really understand the code the first time I implemented it. Now I feel I better understand each of the arguments of the functions for the various estimators.