Homework 2

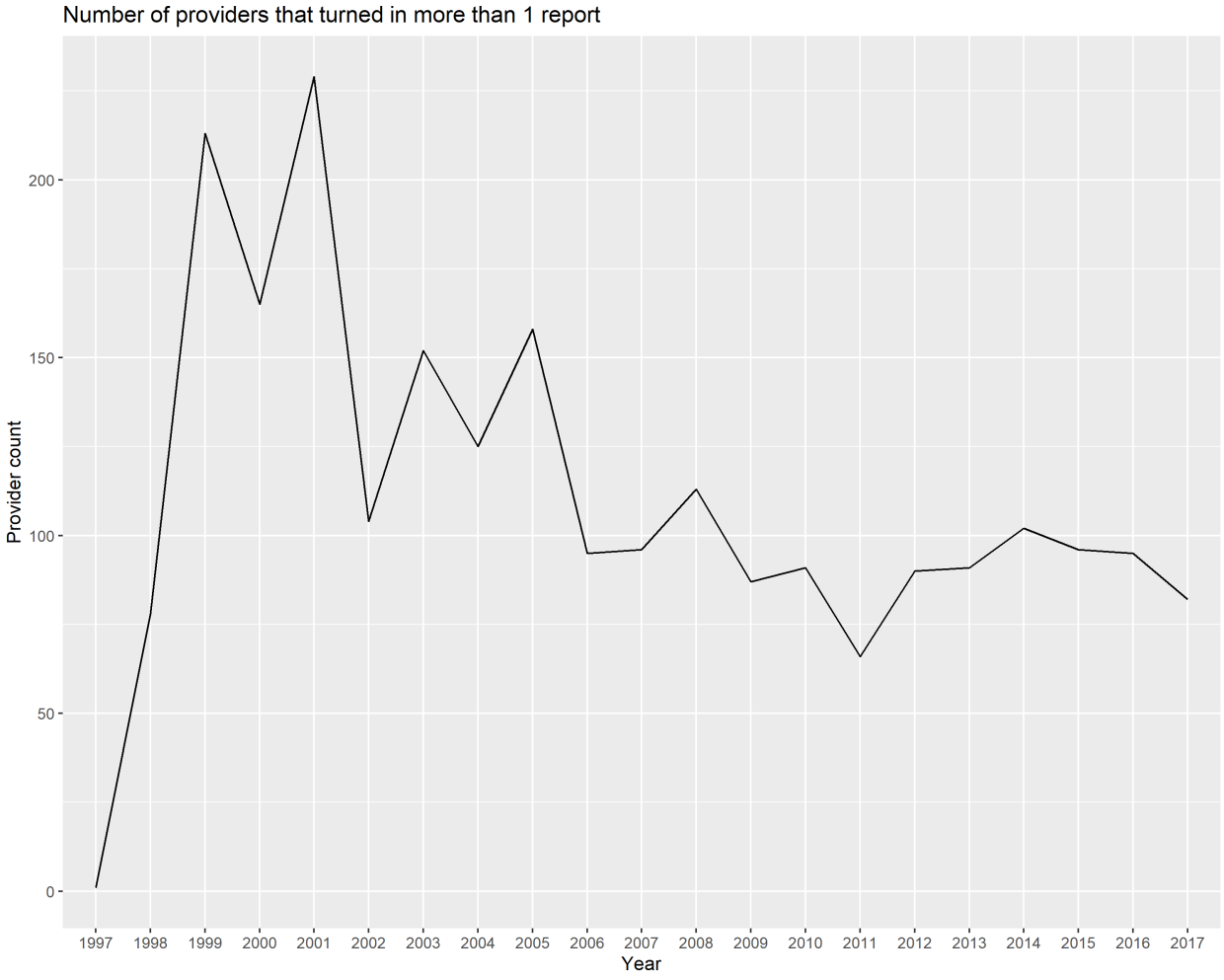
Sam Cohen

2023-02-11

Answers

1.

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

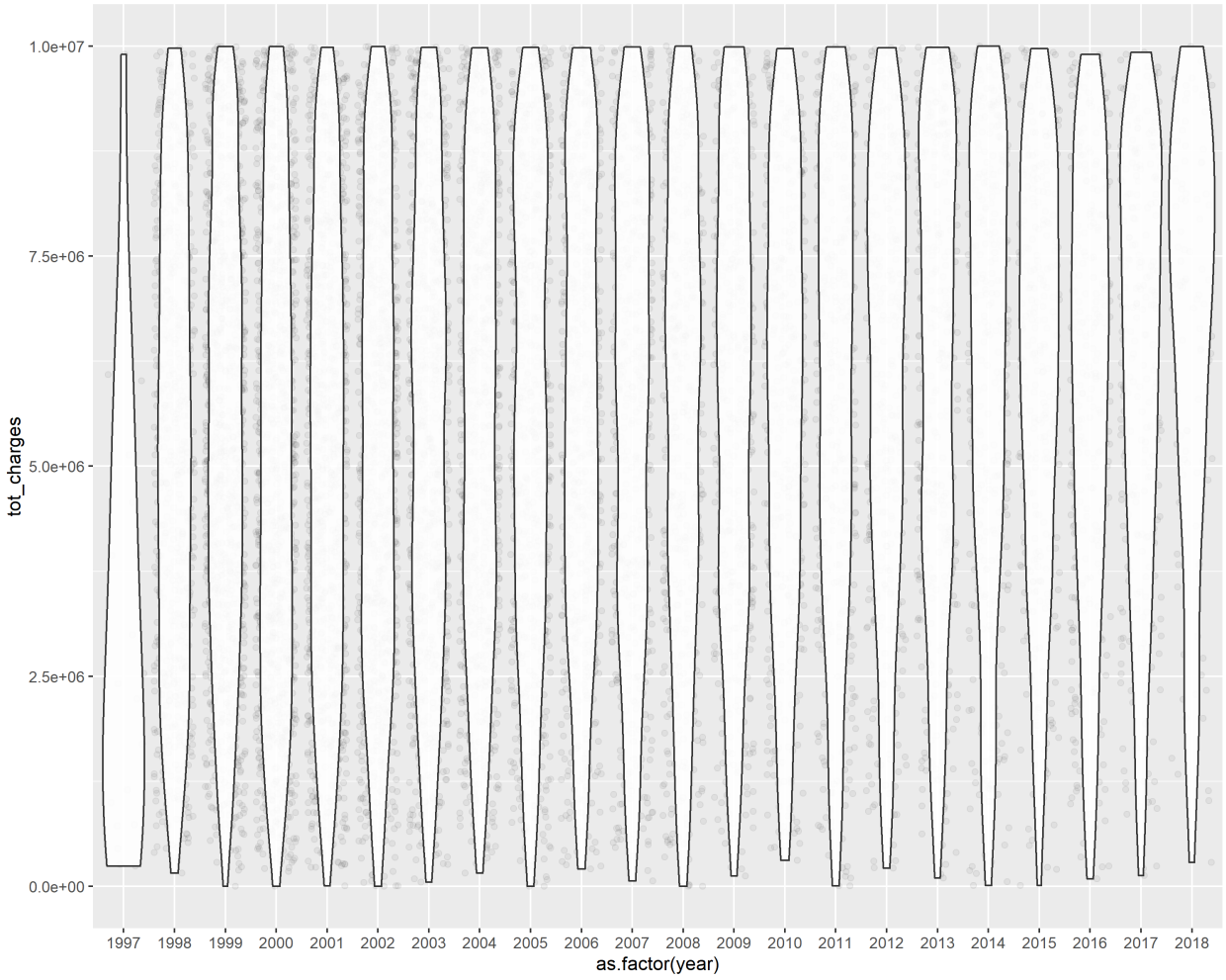


2.

There are 9323 unique provider numbers over the years.

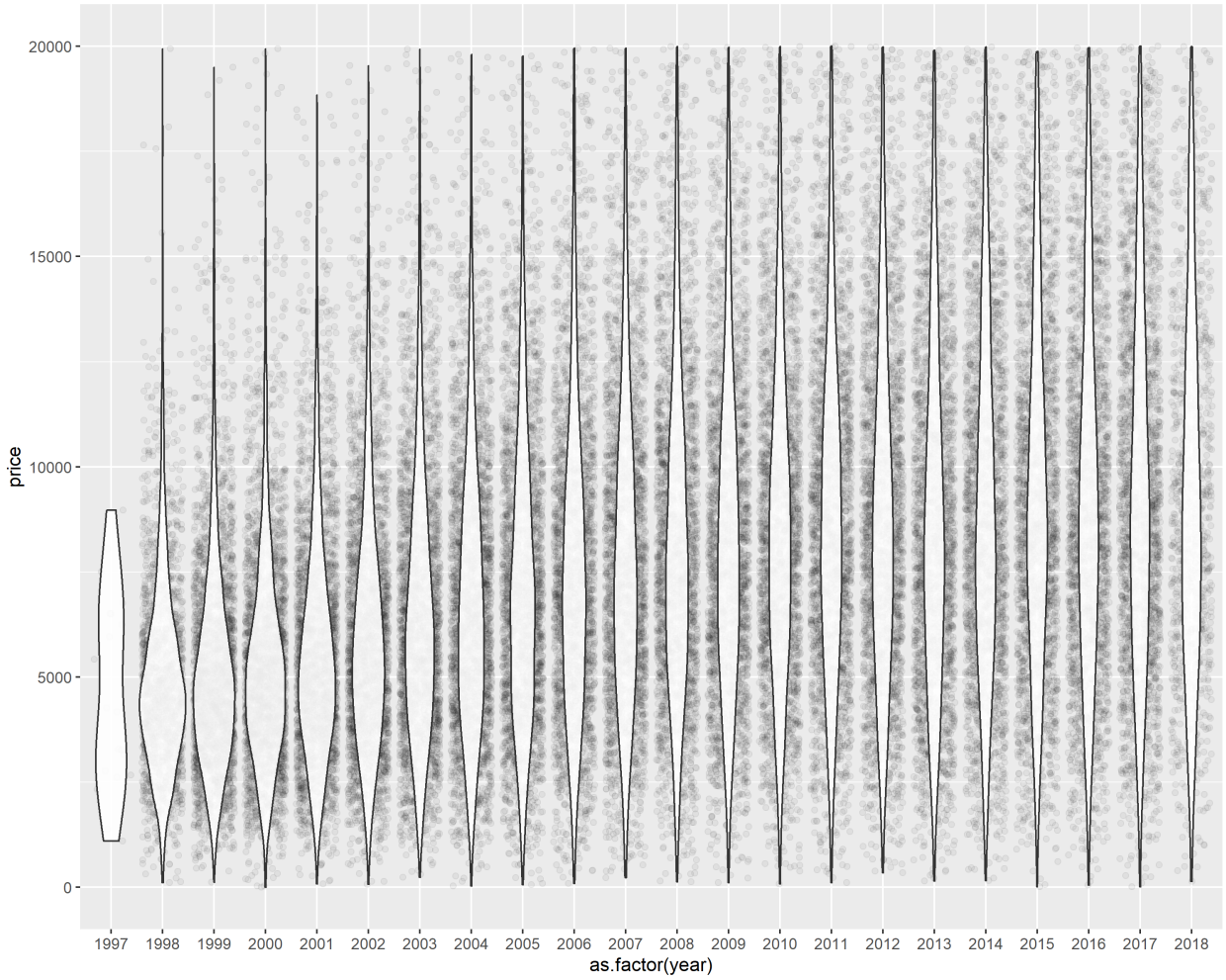
3.

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.



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.



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.

6

```
## `summarise()` has grouped output by 'group', 'quartile1', 'quartile2',
## 'quartile3'. You can override using the `.groups` argument.
```

```
## # A tibble: 8 x 6
## # Groups:   group, quartile1, quartile2, quartile3 [8]
##   group     quartile1 quartile2 quartile3 quartile4 price_mean
##   <chr>         <dbl>     <dbl>     <dbl>     <dbl>      <dbl>
## 1 Control           0         0         0         1     11056.
## 2 Control           0         0         1         0      8467.
## 3 Control           0         1         0         0      7326.
## 4 Control           1         0         0         0      7612.
## 5 Treatment         0         0         0         1     11184.
## 6 Treatment         0         0         1         0      8519.
```

```
## 7 Treatment         0         1         0         0      7997.
## 8 Treatment         1         0         0         0      6981.
```

7

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

	term	estimate	std.error	statistic	p.value
Linear Regression	penalized	171.831396492062	208.532964376467	7.82400112138556	8.41001066526468
Inverse Propensity Weighting	penalized	190.559537148193	188.657422118603	3.01008237581235	0.312544948325758
Nearest Neighbor (Inverse Variance)	penalized	190.55953714855	208.217565693355	NA	NA
Nearest Neighbor (Mahalanobis)	penalized	190.55953714855	208.217565693355	NA	NA

8

Each of these estimators are different, but there certainly some that are closer than others. For example, the linear regression, as well as inverse propensity weighing estimates are very close, 1037 and 1022. These differ greater though from the two nearest neighbor matchings.

9

It is unlikely we have sufficiently estimated the effect of the penalty as we are not controlling for enough factors. Number of beds is most likely not the only metric we should use to control or determine similarity between providers, in order to determine an accurate estimate.

10

Honestly I struggled a lot with 6 and 7. With 6, the first part was fairly easy, but for whatever reason, creating a table as you specified was very difficult for me. I assume the best way would be to use the pivot_wider function, but I was not able to get it to work properly. In addition to this, 7 was similarly difficult as I found it hard to know exactly how to incorperate the quartiles, as well as how to make a good table.