Homework 2

Sam Cohen

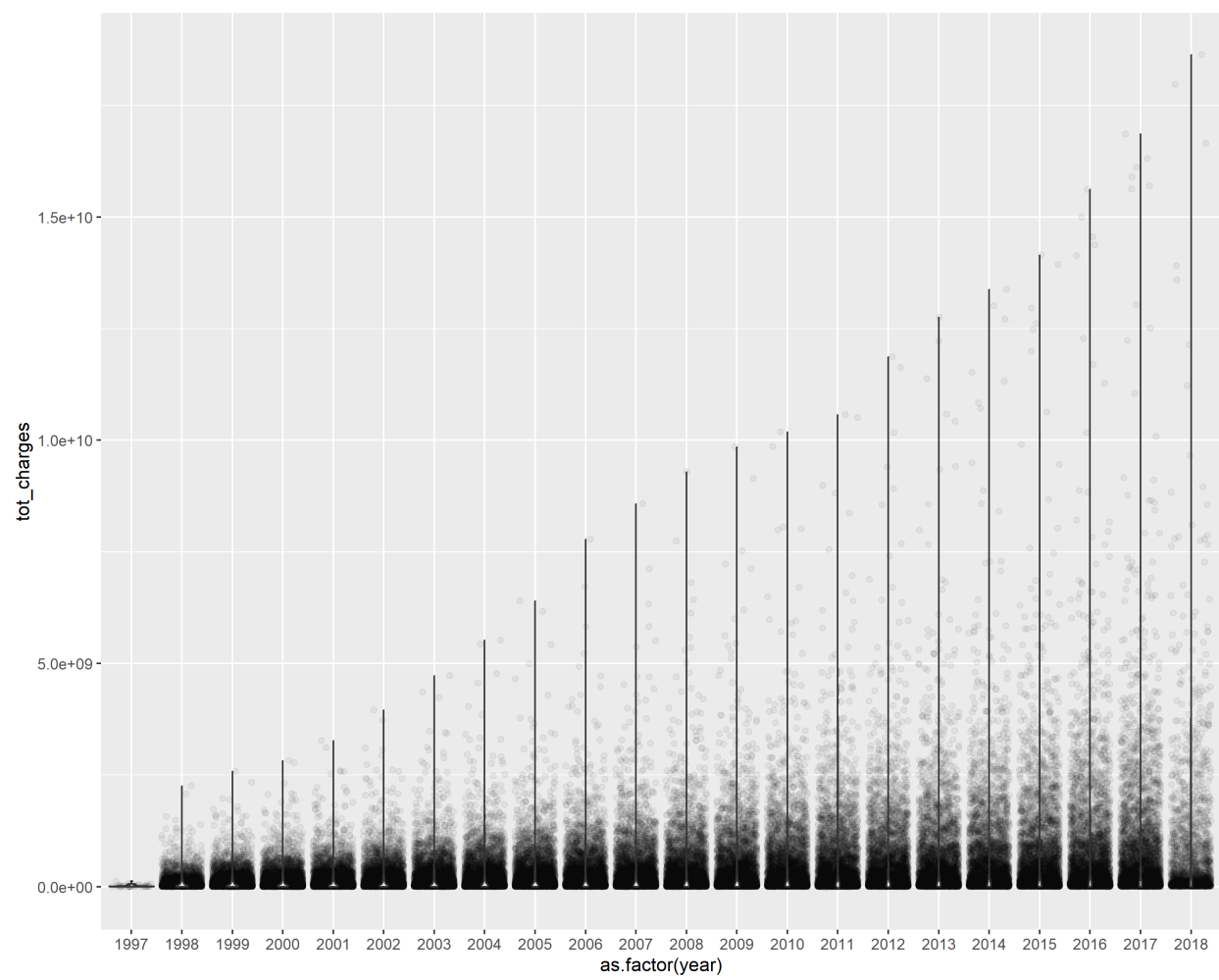### 2023-02-11

Answers

1.

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

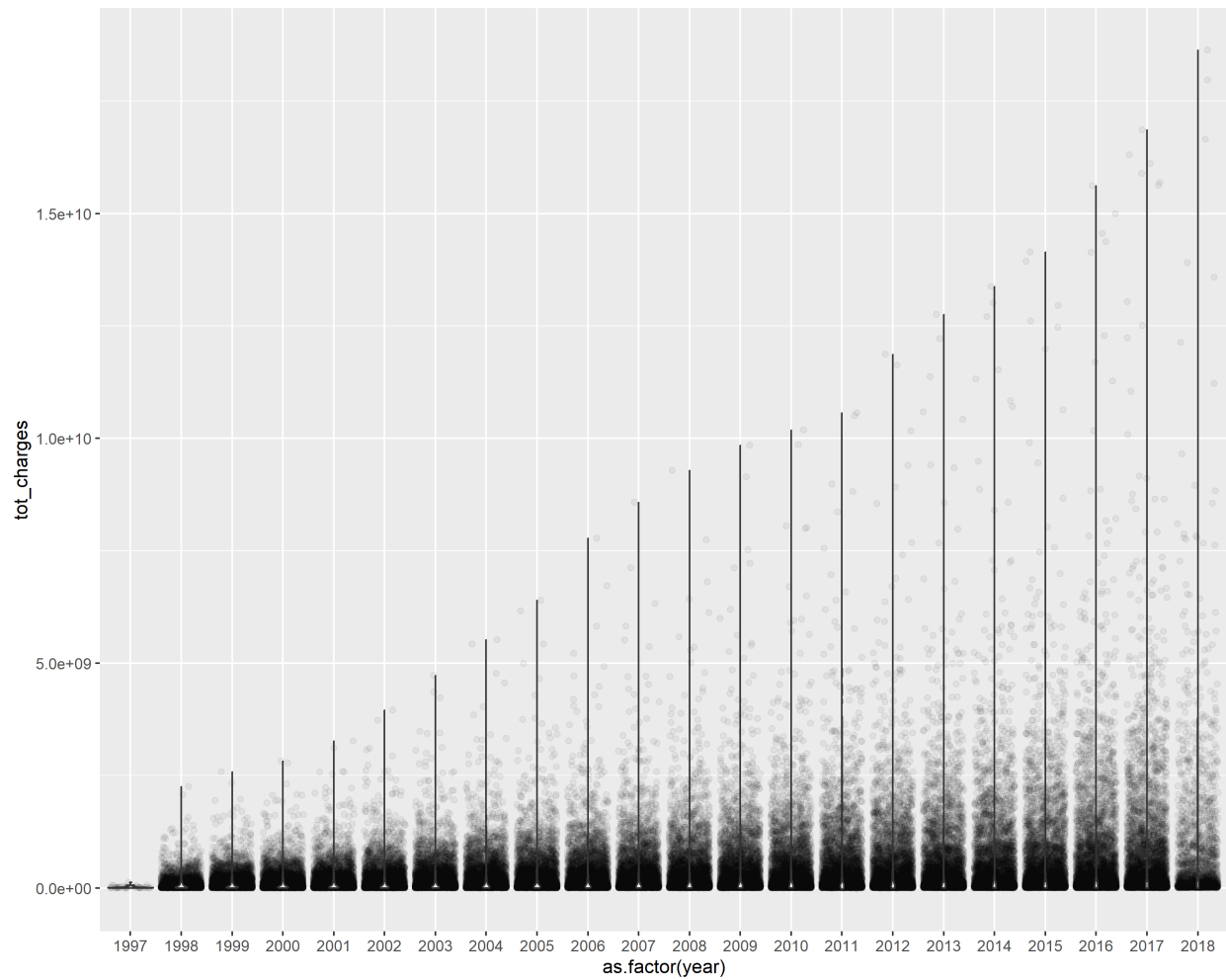**Number of providers that turned in more than 1 report**

2.

There are 9323 unique provider numbers over the years.

3.

4.

5

The average price for penalized hospitals is 10,234.98 whereas the average price for non-penalized providers
is 9,653.264.

6

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

## # A tibble: 9 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         0        NaN
## 2 Control          0         0         0         1     11099.
## 3 Control          0         0         1         0      8570.
## 4 Control          0         1         0         0      7529.
```

```
## 5 Control          1        0        0        0      7522.
## 6 Treatment        0        0        0        1     12036.
## 7 Treatment        0        0        1        0     10000.
## 8 Treatment        0        1        0        0      9088.
## 9 Treatment        1        0        0        0       NaN
```

7

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

term	estimate	std.error	statistic	p.value
(Intercept)	11235.6217399732	165.89526653265	67.7271990624393	0
penalized	1037.67957523629	455.500269684014	2.27810968357086	0.022814475146533
quartile1	-6662.80762943178	3926.07224808145	-1.69706699429377	0.08982385259093
quartile2	-3926.33412690053	367.657543576845	-10.6793242665505	5.35907085702224e-26
quartile3	-2526.28675457336	256.204274869984	-9.8604395100565	1.77158333666337e-22
(Intercept)	9833.39833182917	160.890764829397	61.1184758942265	0
penalized	1022.95696978204	227.584958385215	4.49483558597296	7.31904965569591e-06
penalized	1431.58752804429	501.60050134474	NA	NA
penalized	1124.61938518242	499.081205743406	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.