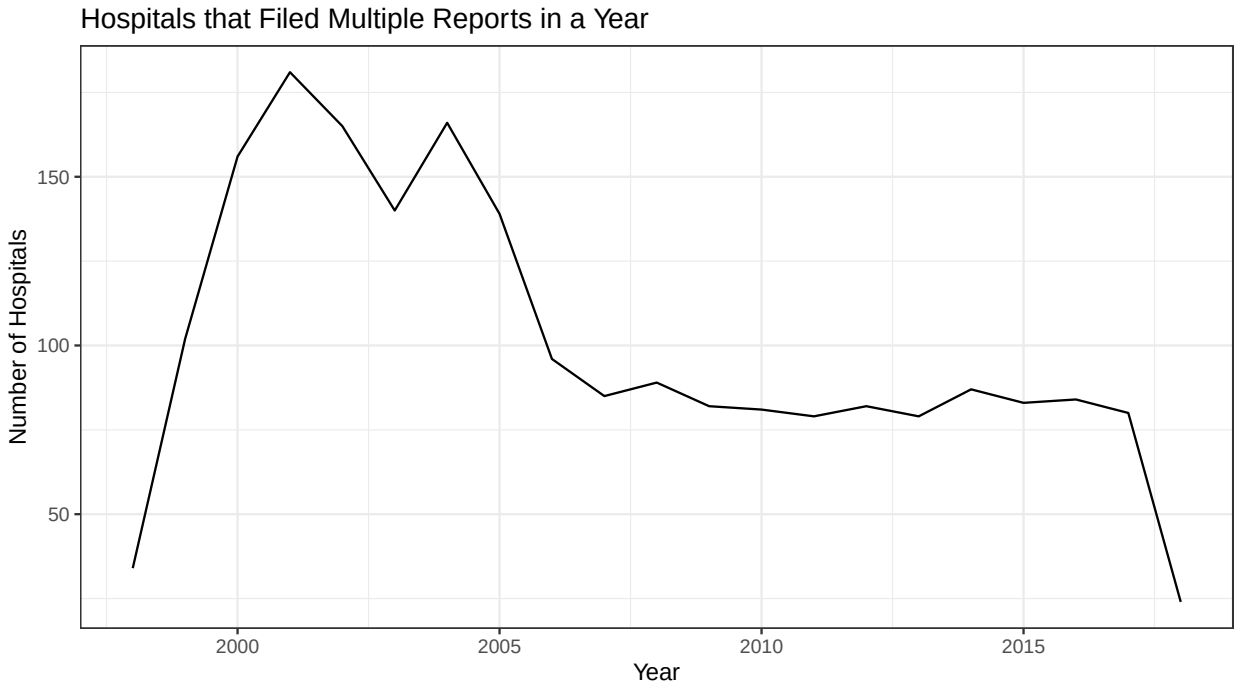
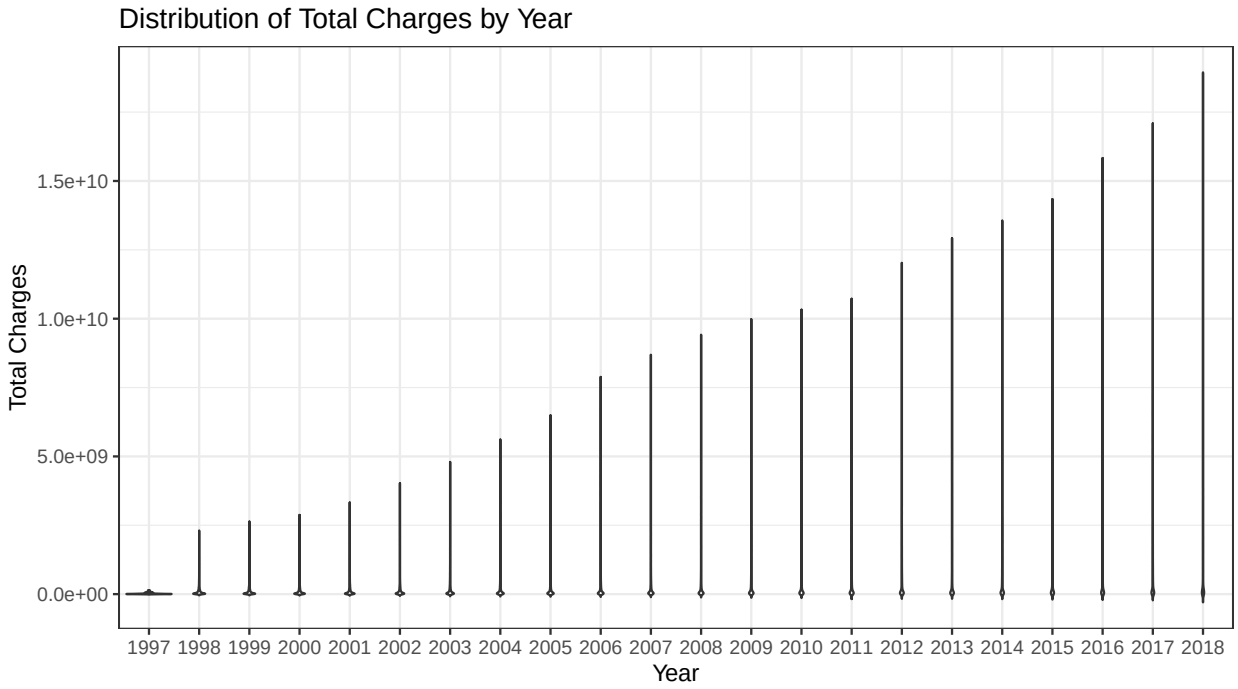
Bhasin-S-hwk2-2

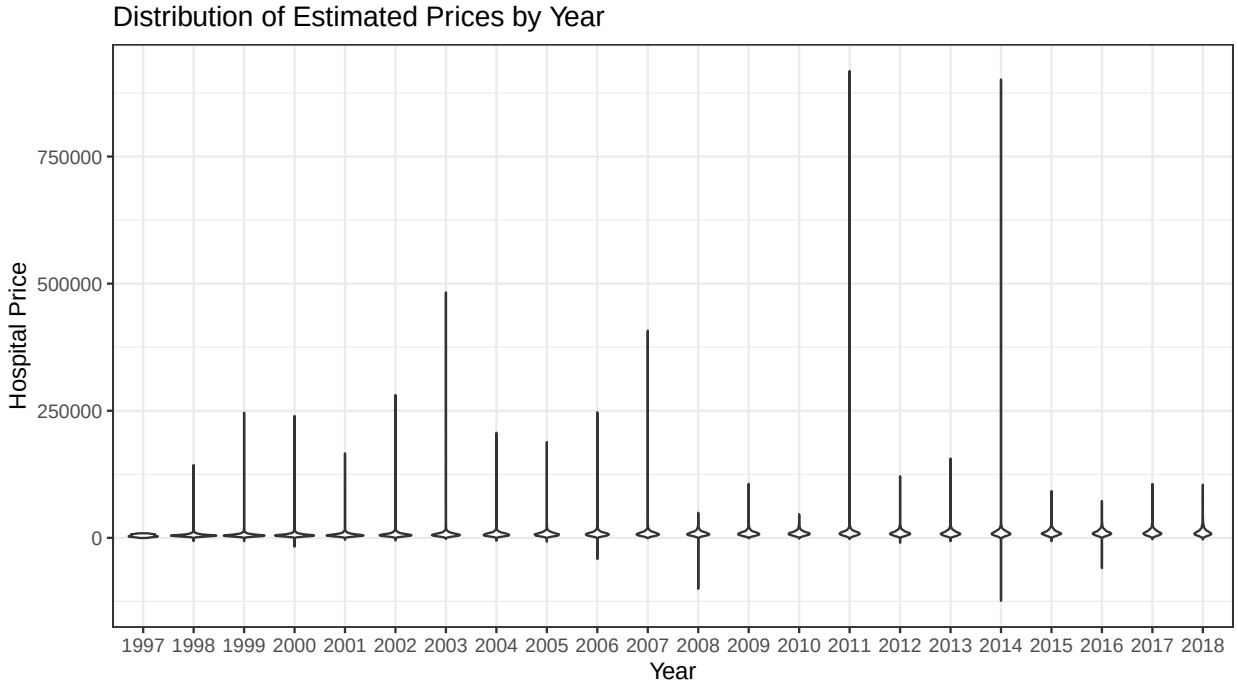
Sachi Bhasin

2023-02-14



There are 9323 unique hospital IDs in the data.






```
## # A tibble: 2 x 2
##   penalty  price
##     <dbl>  <dbl>
## 1       0  9752.
## 2       1 10235.
```

```
## # A tibble: 8 x 3
## # Groups:   quartile [4]
```

```
##   quartile penalty avg_price
##      <int>   <dbl>     <dbl>
## 1        1       0     8409.
## 2        1       1     7653.
## 3        2       0     8315.
## 4        2       1    10833.
## 5        3       0    10501.
## 6        3       1     9339.
## 7        4       0    11726.
## 8        4       1    12435.
```

```
##                   Length Class  Mode
## est                    1 -none- numeric
## se                     1 -none- numeric
## est.noadj              1 -none- numeric
## se.standard            1 -none- numeric
## se.cond                1 -none- numeric
## mdata                  4 -none- list
## index.treated      24812 -none- numeric
## index.control      24812 -none- numeric
## index.dropped          0 -none- NULL
## weights            24812 -none- numeric
## orig.nobs              1 -none- numeric
## orig.wnobs             1 -none- numeric
## orig.treated.nobs      1 -none- numeric
## nobs                   1 -none- numeric
## wnobs                  1 -none- numeric
## caliper                0 -none- NULL
## ecaliper               0 -none- NULL
## exact                  0 -none- NULL
## ndrops                 1 -none- numeric
## ndrops.matches         1 -none- numeric
## MatchLoopC        124060 -none- numeric
## version                1 -none- character
## estimand               1 -none- character
```

```
##                   Length Class  Mode
## est                    1 -none- numeric
## se                     1 -none- numeric
## est.noadj              1 -none- numeric
## se.standard            1 -none- numeric
## se.cond                1 -none- numeric
## mdata                  4 -none- list
## index.treated      24812 -none- numeric
## index.control      24812 -none- numeric
## index.dropped          0 -none- NULL
## weights            24812 -none- numeric
## orig.nobs              1 -none- numeric
## orig.wnobs             1 -none- numeric
## orig.treated.nobs      1 -none- numeric
## nobs                   1 -none- numeric
## wnobs                  1 -none- numeric
## caliper                0 -none- NULL
## ecaliper               0 -none- NULL
```

```
## exact                  0 -none- NULL
## ndrops                 1 -none- numeric
## ndrops.matches         1 -none- numeric
## MatchLoopC        124060 -none- numeric
## version                1 -none- character
## estimand               1 -none- character
```

propensity score

I was unable to get the code to run for the propensity scores. When I ran the following code, I kept getting an error that 'ps' must be size 1, not 673 and where this error happened. I think R i trying to tell me I am inputting too many values in the log regression model. I am unsure how to fix this problem right now. I think I may have to make a data frame to adjust the number of input values.

logit.model <- glm(penalty ~ quartile_1 + quartile_2 + quartile_3 + quartile_4, data = year_2012, family = binomial)

ps <- fitted(logit.model)

year_2012 <- year_2012 %>% mutate(ps = predict(logit.model, type = 'response')) %>% filter(ps>0 & ps<1)

IPW Weights

year_2012 <- year_2012 %>% mutate(ipw = case_when(penalty == 1 ~ 1/ps, penalty == 0 ~ 1/(1=ps), TRUE~NA_real_))

view(year_2012)

mean.t1 <- year_2012 %>% # filter(penalty==1) %>% # dplyr::select(price, ipw) %>% #summarize(mean_y=weighted.mean(price, w=ipw)) mean.t0 <- year_2012 %>% filter(penalty==0) %>% dplyr::select(price, ipw) %>% summarize(mean_y=weighted.mean(price, w=ipw))

mean.t1$mean_y$ − $mean.t$0mean_y reg.ipw <- lm(price ~ penalty, data=year_2012, weights=ipw)

reg.ipw

```
## [1] 482.5144
```

8 Although I was unable to run the code for all of the different treatments, based on the results I got thus far, the results are similar. The two neighbor matching estimators are equal but the linear regression one differs.

9 I think I have estimated a causal effect of the penalty. I think getting similar estimators by matching, weighting, and running regressions on the data allowed us to control for potential confounding variables, which would suggest a causal effect.

10 I found working with this data challenging but easier than homework 1 as I am getting more comfortable trouble shooting and working with R studio. I learned how to create a dummy variable and quartiles for a data set. It has been very aggravating to troubleshoot the error I keep getting with the propensity score.

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:

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.