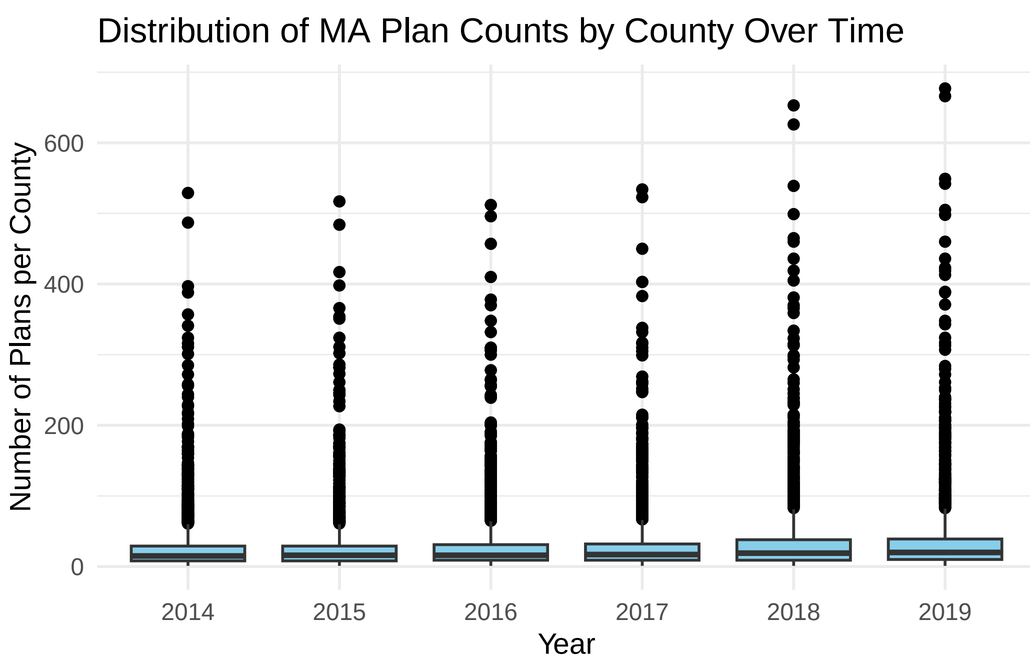
ECON 470 Homework 2 Submission 3

Tammy Dang

Part I: Summarize the Data

1. Remove all SNPs, 800-series plans, and prescription drug only plans (i.e., plans that do not offer Part C benefits). Provide a box and whisker plot showing the distribution of plan counts by county over time. Do you think that the number of plans is sufficient, too few, or too many? To answer this question, you need only work with the "plan data" and the "service area" data just as we did in homework 1, but for more years in this case.

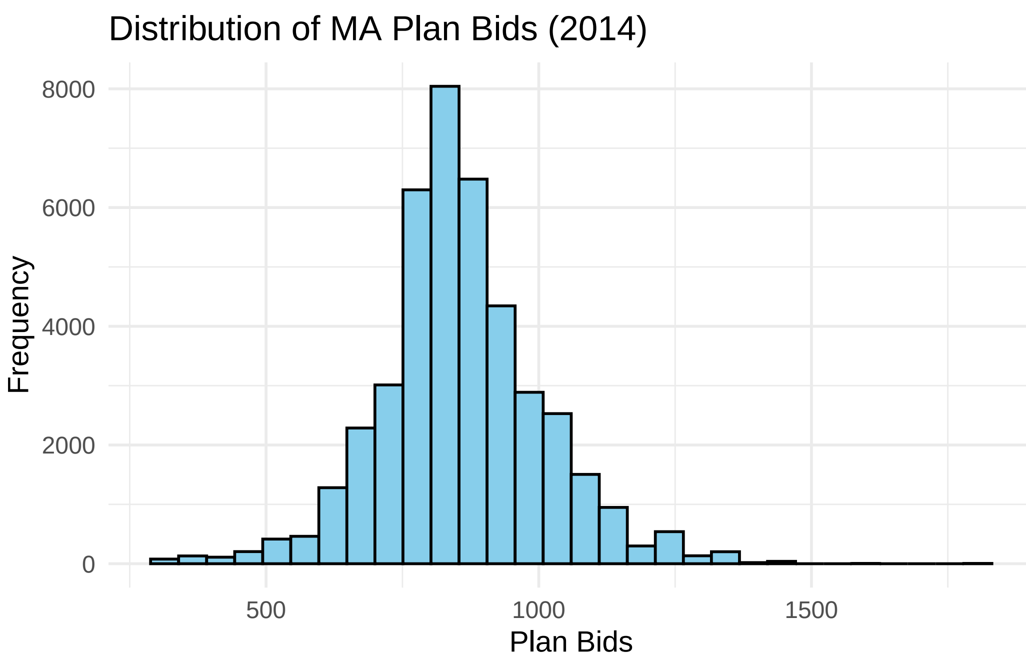
Based on the box-and-whisker plot, the number of plans per county for each year are extremely varied per county. There tends to be many options within the counties with the large number of outliers, which indicates many options or plans for each county. This could possibly overhwlem the consumer.



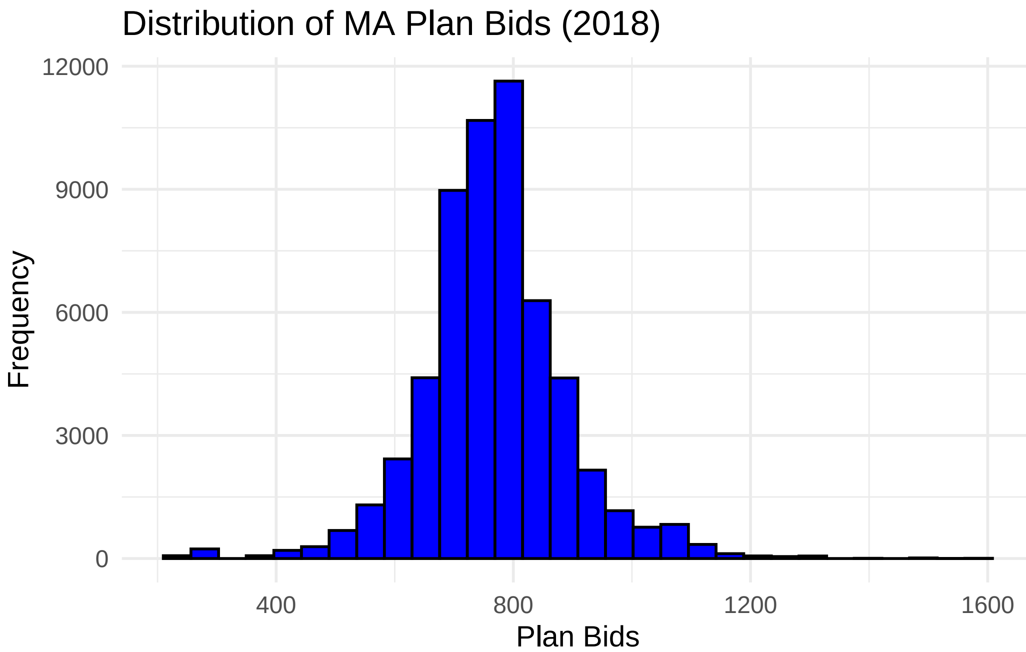
2. Provide frequency histograms showing the distribution of plan bids in 2014 and 2018. How has this distribution changed over time? To properly measure plan bids, you need to incorporate the landscape files (in the plan characteristics code files) and the risk/rebate data. The build data code file in the Medicare Advantage GitHub repository shows how to back out plan bids from the observed premium and rebate data.

The highest distribution of plan bids in 2014 is slightly more than 8000. The highest distribution of plan bids in 2018 is around 10000. This reveals that plan bids have increased, over time.

```
Warning: Removed 7635 rows containing non-finite outside the scale range
(`stat_bin()`).
```

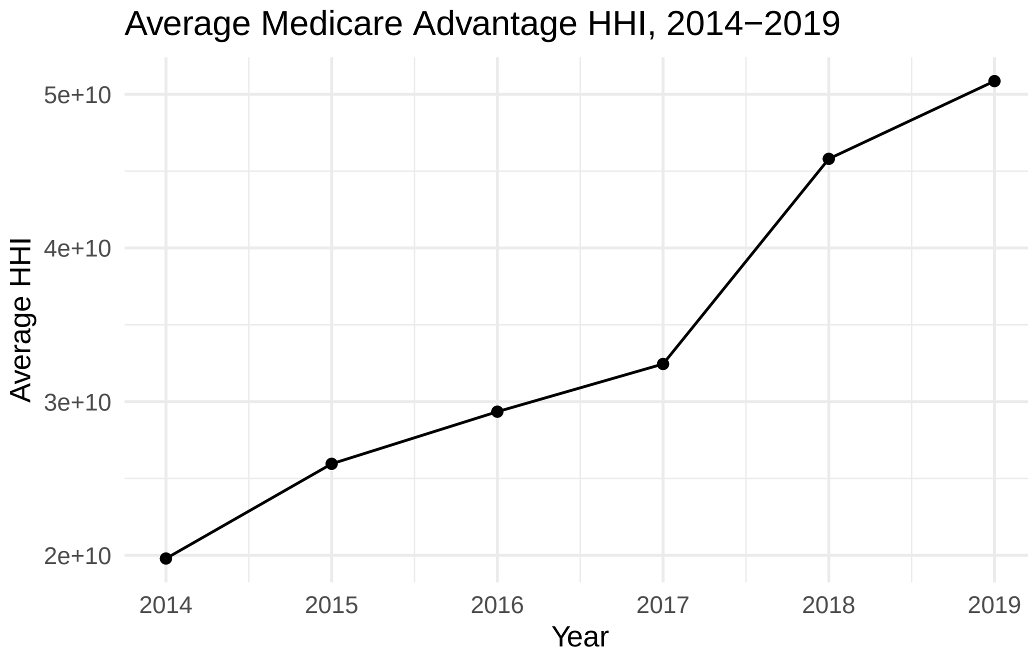



```
Warning: Removed 7192 rows containing non-finite outside the scale range
(`stat_bin()`).
```

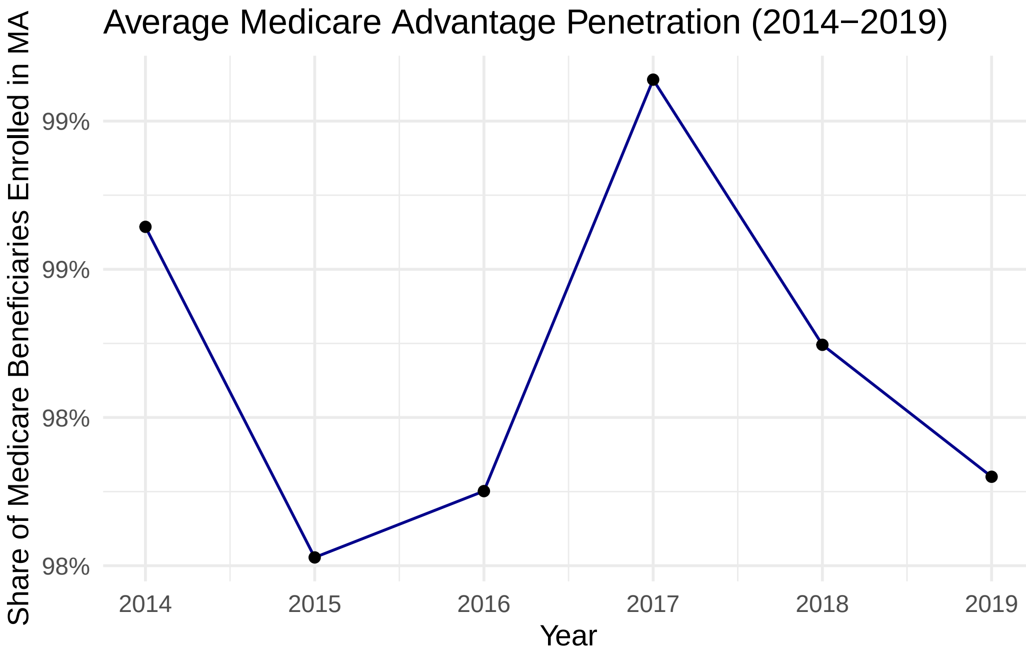


3. Plot the average HHI over time from 2014 through 2019. How has the HHI changed over time? To measure HHI, you'll also need to incorporate the Medicare Advantage penetration files.

The average HHI for Medicare Advantage markets between 2014-2019 shows how MA markets are concentrated and become even more highly concentrated over time. With high increases in average HHI across years suggests growing concentration and consolidation in the counties over time. Despite growth in the number of plans, enrollments continues to get more and mroe concentrated among a smaller number of insurers, decreasing competition in the market over time.


Average Medicare Advantage HHI, 2014−2019
Average HHI
5e+10
4e+10
3e+10
2e+10
2014
2015
2016
2017
2018
2019
Year

4. Plot the average share of Medicare Advantage (relative to all Medicare eligibles) over time from 2014 through 2019. Has Medicare Advantage increased or decreased in popularity?



Part II: Estimate ATEs

Construct this ma_hhi2 for part 2:

For the rest of the assignment, you should include only observations in 2018. As we did in class, please define "competitive" markets as those with HHIs in the lower 33rd percentile of the national distribution of HHI, and define "concentrated" or "uncompetitive" markets as with with HHIs in the upper 66th percentile. This is somewhat arbitrary but it allows us to define a binary treatment variable in a way that we can more easily implement the methods in this module.

5. Calculate the average bid among competitive versus uncompetitive markets.

```
# A tibble: 2 x 2
  market_type            avg_bid
  <chr>                    <dbl>
1 Competitive (control)     797.
```

```
2 Concentrated (treated)    776.
```

6. Split markets into quartiles based on Medicare fee-for-service (FFS) costs. To do this, create 4 new indicator variables, where each variable is set to 1 if the FFS costs falls into the relevant quartile. Provide a table of the average bid among treated/control groups for each quartile. For this, you'll need to incorporate the FFS costs data as well.

```
# A tibble: 4 x 3
  ffs_quartile `competitive (control)` `concentrated (treated)`
  <chr>                          <dbl>                    <dbl>
1 Q1                              821.                     786.
2 Q2                              803.                     785.
3 Q3                              780.                     766.
4 Q4                              789.                     759.
```

7. Find the average treatment effect using each of the following estimators, and present your results in a single table:

- Nearest neighbor matching (1-to-1) with inverse variance distance based on quartiles of FFS costs
- Nearest neighbor matching (1-to-1) with Mahalanobis distance based on quartiles of FFS costs
- Inverse propensity weighting, where the propensity scores are based on quartiles of FFS costs
- Simple linear regression, adjusting for quartiles of FFS costs using dummy variables and appropriate interactions as discussed in class

Nearest neighbor matching (1-to-1) with inverse variance

```
Increasing memory because of ties: allocating a matrix of size 3 times 231800 doubles.
I would be faster with the ties=FALSE option.
```

```
Warning in MatchLoopCfast(N = s1$N, xvars = Kx, All = s1$All, M = s1$M, :
Increasing memory because of ties.  I would be faster with the ties=FALSE
option.
```

```
Estimate...  -24.052
AI SE......  3.5529
T-stat.....  -6.7697
```

```
p.val......  1.2908e-11

Original number of observations..............  1159
Original number of treated obs...............  604
Matched number of observations...............  1159
Matched number of observations  (unweighted).  162808
```

nearest neighbor matching with mahalnobis

```
Increasing memory because of ties: allocating a matrix of size 3 times 231800 doubles.
I would be faster with the ties=FALSE option.

Warning in MatchLoopCfast(N = s1$N, xvars = Kx, All = s1$All, M = s1$M, :
Increasing memory because of ties.  I would be faster with the ties=FALSE
option.


Estimate...  -24.052
AI SE......  3.5529
T-stat.....  -6.7697
p.val......  1.2908e-11

Original number of observations..............  1159
Original number of treated obs...............  604
Matched number of observations...............  1159
Matched number of observations  (unweighted).  162808
```

inverse propoensity weighting

```
     mean_y
1 -24.05216
```

SLRegression w/ the 4 quartiles

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -34.11  -34.11  -18.26  -24.05  -18.26  -13.58
```

8. With these different treatment effect estimators, are the results similar, identical, very different?

The results of each ATE, ipw, or estimates are identical to each other.

9. Pick your favorite flavor of estimators in this section (matching, weighting, regression, etc) and re-estimate treatment effects using the continuous FFS costs variable as well as total Medicare beneficiaries as your covariates. How does this result compare to the analogous estimate when matching/weighting only on FFS quartile?

I will use nearest neighbor matching with malhanobis. The result of this estimate is less robust in terms of unweighted matched number of observations in comparison to when we only match on FFS quartile. In addition, the results are not statistically significant, with a wider standard error overall.

```
Estimate...  -9.3117
AI SE......  10.077
T-stat.....  -0.92407
p.val......  0.35545

Original number of observations..............  1159
Original number of treated obs...............  604
Matched number of observations...............  1159
Matched number of observations  (unweighted).  1170
```

10. Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated or surprised you.

Working with this data, I was really aggravated by the amount of time and errors it took to combine the dataset together. One thing that surprised me was how I kept having to go back to the data cleaning part to retrieve something I accidentally left out. It goes to show how important the data cleaning process is, and how intentional you have to be when cleaning it in order to have a smooth process when answering the questions. I'm honestly still not sure how I could fix the dataset because I'm getting different results and distributions.