Homework 4

Research Methods, Spring 2024

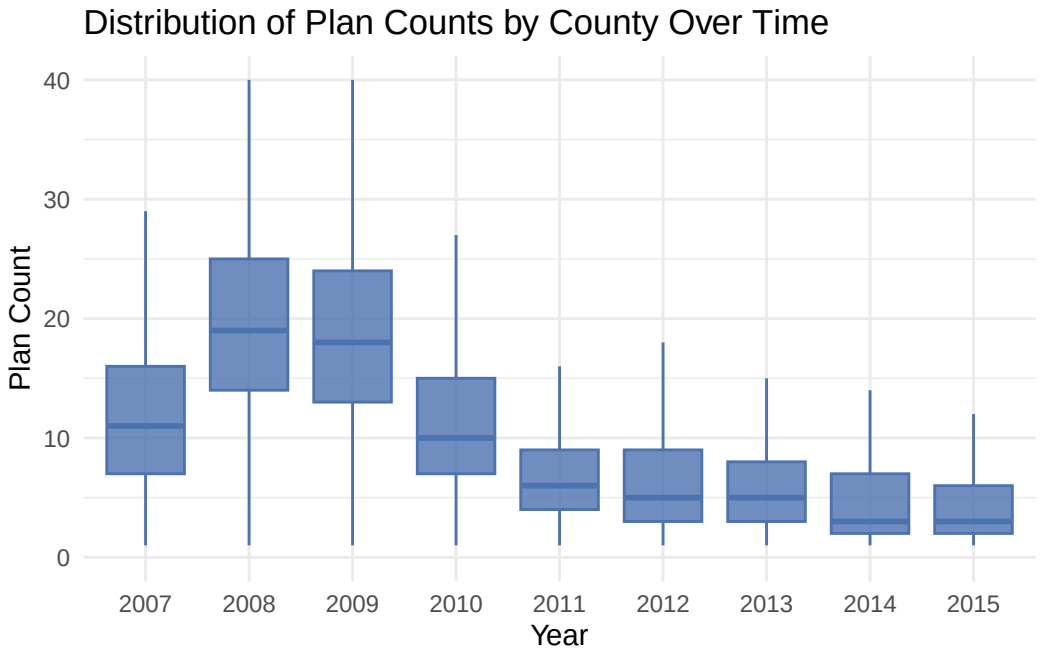
Taeyoung Yim

Summarize the Data

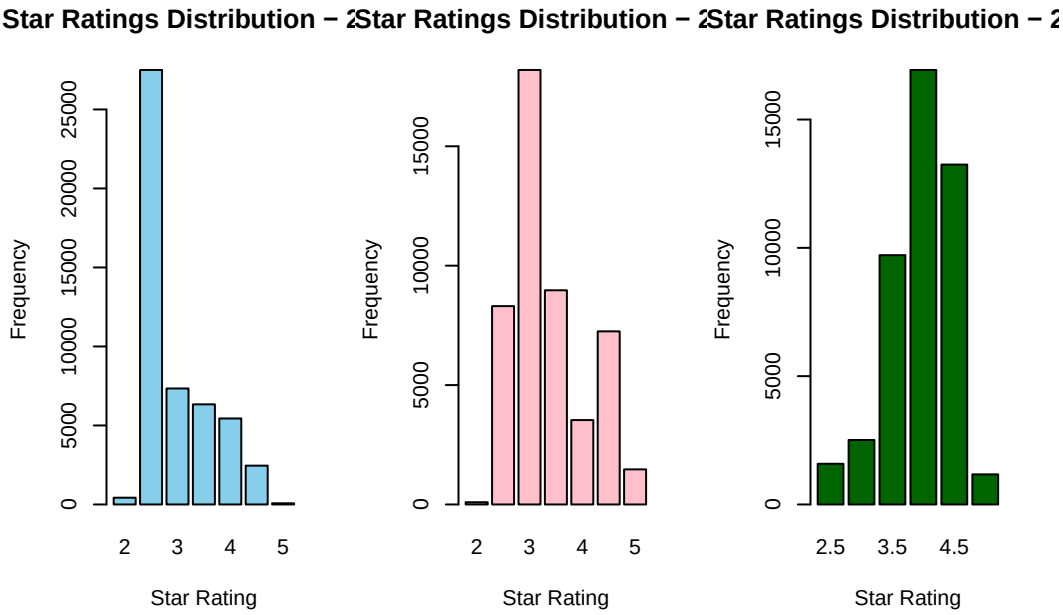
1. Remove all SNPs, 800-series plans, and prescription drug only plans (i.e., plans that do not offer Part C benefits). Provide a box and whisker plot showing the distribution of plan counts by county over time. Do you think that the number of plans is sufficient, too few, or too many?

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

```
Warning: Removed 907 rows containing non-finite values (`stat_boxplot()`).
```

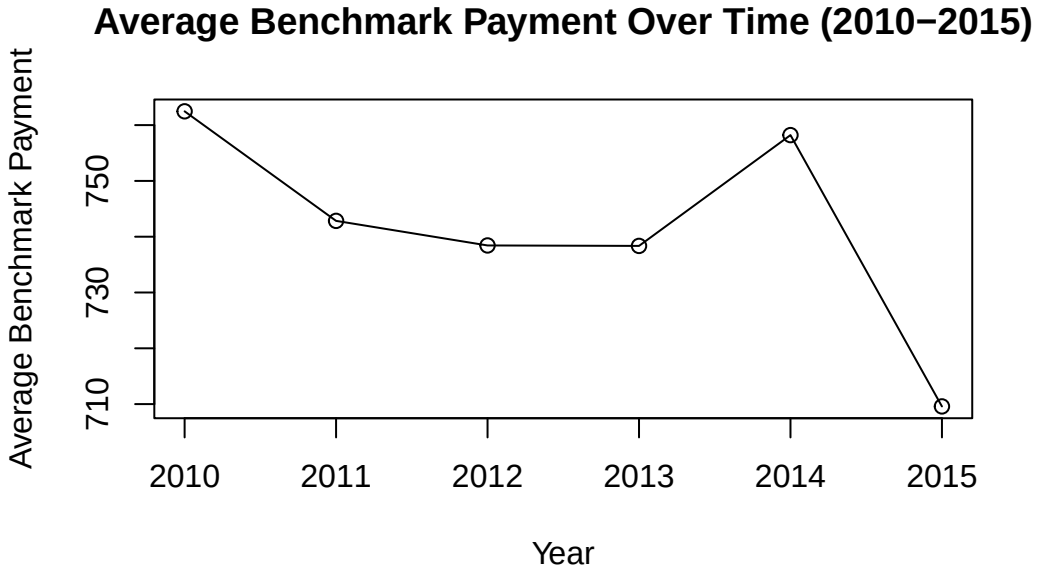


2. Provide bar graphs showing the distribution of star ratings in 2010, 2012, and 2015. How has this distribution changed over time?



The mean start ratings seems to increase over time.

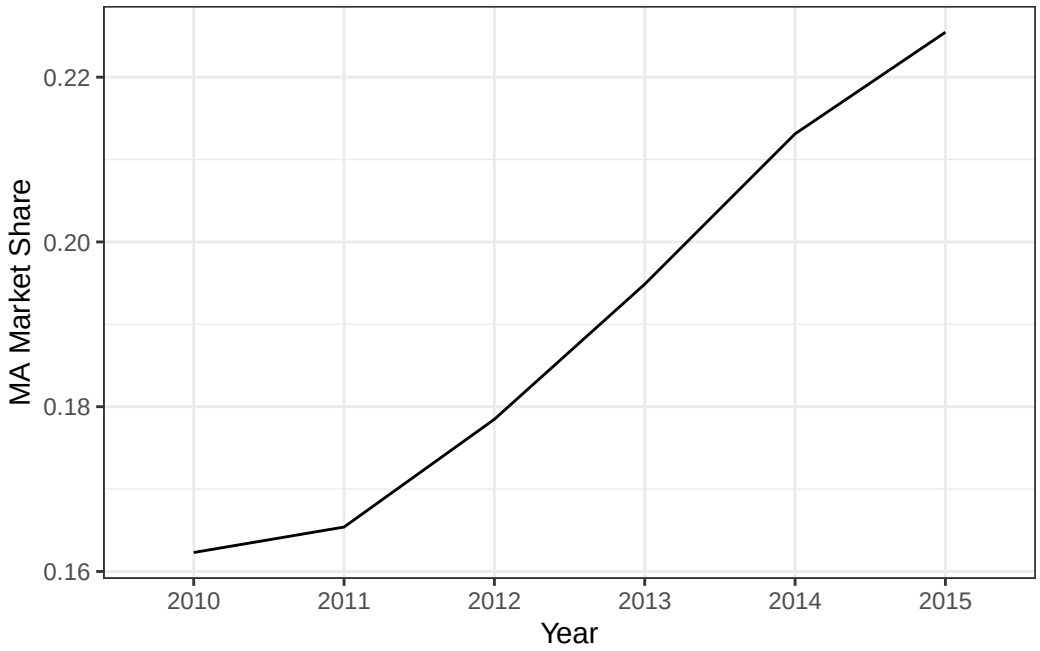
3.Plot the average benchmark payment over time from 2010 through 2015. How much has the average benchmark payment risen over the years?.




```
The average benchmark payment has risen by -52.89 over the years (2010-2015).
```

4.Plot the average share of Medicare Advantage (relative to all Medicare eligibles) over time from 2010 through 2015. Has Medicare Advantage increased or decreased in popularity? How does this share correlate with benchmark payments?

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



Medical advantage increased overtime. Comparing this trend with the benchmark payment trend, the over all trend seems to be positively correlated, although the trend is slight different between 2011 and 2013.

Estimate ATEs

For the rest of the assignment, we'll use a regression discontinuity design to estimate the average treatment effect from receiving a marginally higher rating. We'll focus only on 2010. 5.Calculate the running variable underlying the star rating. Provide a table showing the

number of plans that are rounded up into a 3-star, 3.5-star, 4-star, 4.5-star, and 5-star rating.

```
# A tibble: 8 x 2
  rounded_rating count
           <dbl> <int>
1            2     428
2            2.5 27498
3            3    7341
4            3.5  6337
5            4    5442
6            4.5  2457
7            5      74
8           NA   39808
```

6.Using the RD estimator with a bandwidth of 0.125, provide an estimate of the effect of receiving a 3-star versus a 2.5 star rating on enrollments. Repeat the exercise to estimate the effects at 3.5 stars, and summarize your results in a table.

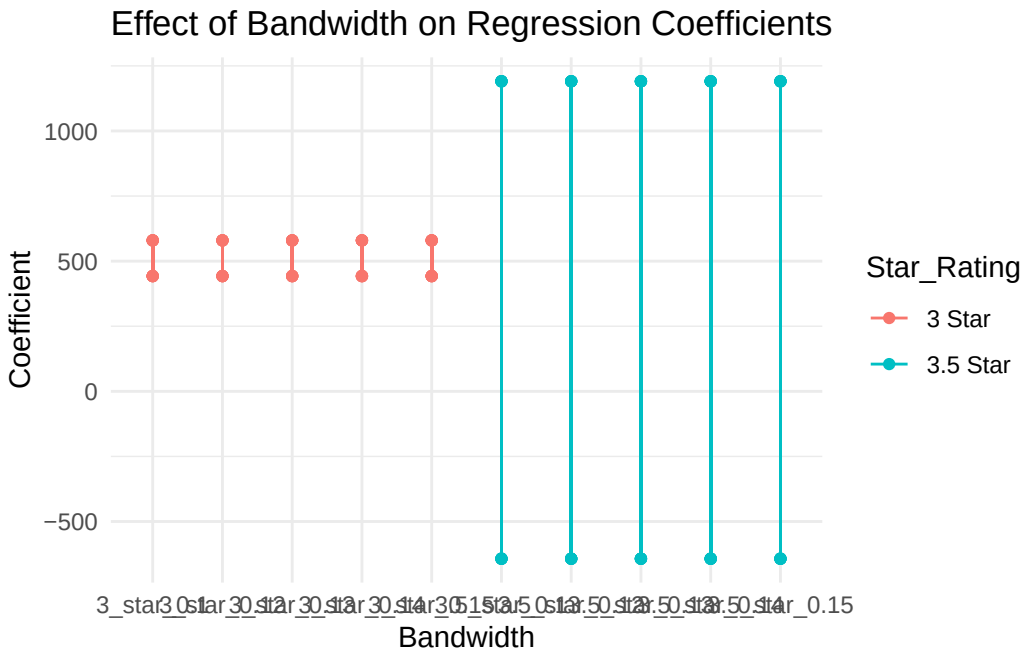
```
Call:
lm(formula = avg_enrollment ~ score + treat + treat_score, data = star30_data_2010)

Residuals:
   Min     1Q Median     3Q    Max
  -727   -132   -115    -53  41766

Coefficients: (2 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   442.69      16.73   26.45   <2e-16 ***
score        1190.82      66.94   17.79   <2e-16 ***
treat             NA         NA      NA       NA
treat_score       NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1538 on 9994 degrees of freedom
  (24843 observations deleted due to missingness)
Multiple R-squared:  0.0307,    Adjusted R-squared:  0.0306
F-statistic: 316.5 on 1 and 9994 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = avg_enrollment ~ score + treat + treat_score, data = star35_data_2010)

Residuals:
   Min     1Q Median     3Q    Max
  -727   -650   -389   -227  41228

Coefficients: (2 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   579.67      26.72  21.696  < 2e-16 ***
score        -642.91     106.87  -6.016  1.9e-09 ***
treat             NA         NA      NA       NA
treat_score       NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2054 on 5911 degrees of freedom
  (7765 observations deleted due to missingness)
Multiple R-squared:  0.006085,  Adjusted R-squared:  0.005917
F-statistic: 36.19 on 1 and 5911 DF,  p-value: 1.898e-09
```

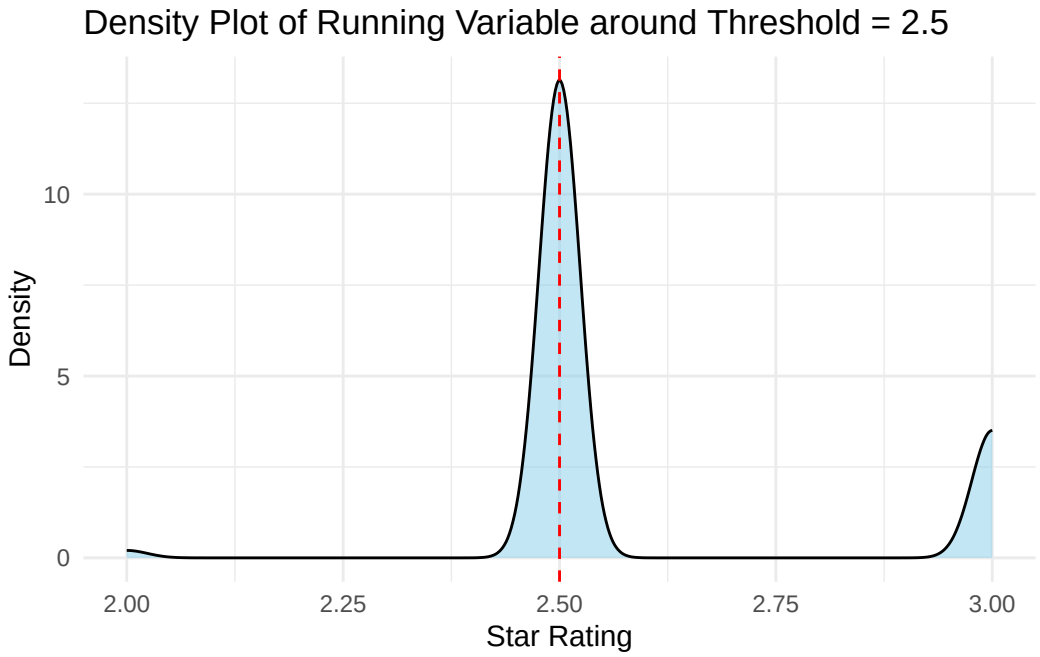
7.Repeat your results for bandwidhts of 0.1, 0.12, 0.13, 0.14, and 0.15 (again for 3 and 3.5 stars). Show all of the results in a graph. How sensitive are your findings to the choice of bandwidth?

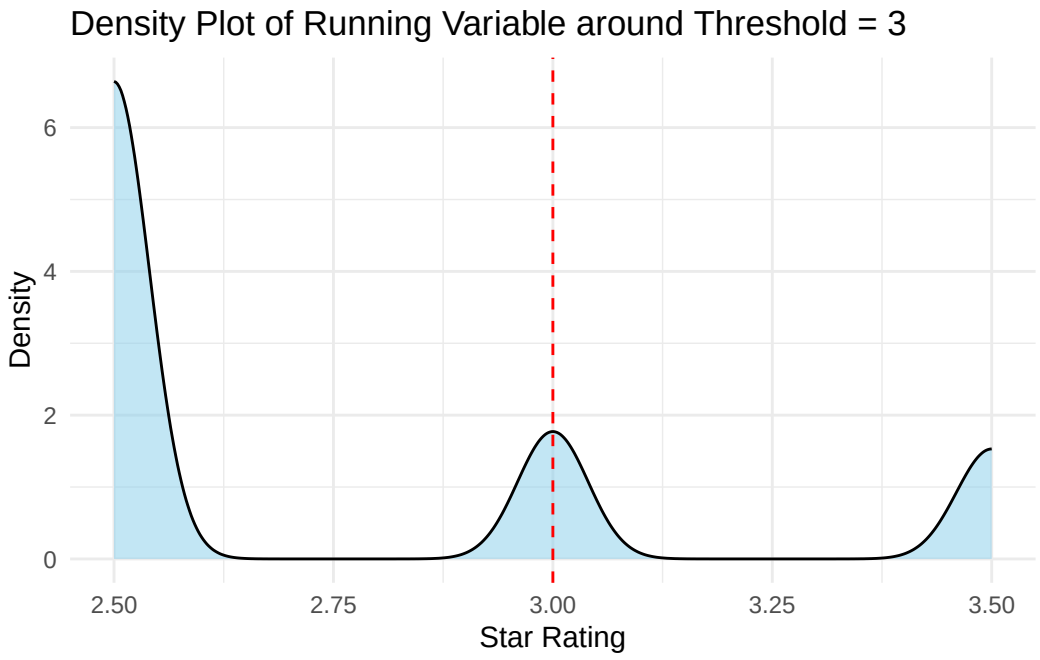
```
Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded

Warning in data.frame(Bandwidth = names(results), Coefficients =
x$coefficients[, : row names were found from a short variable and have been
discarded
```

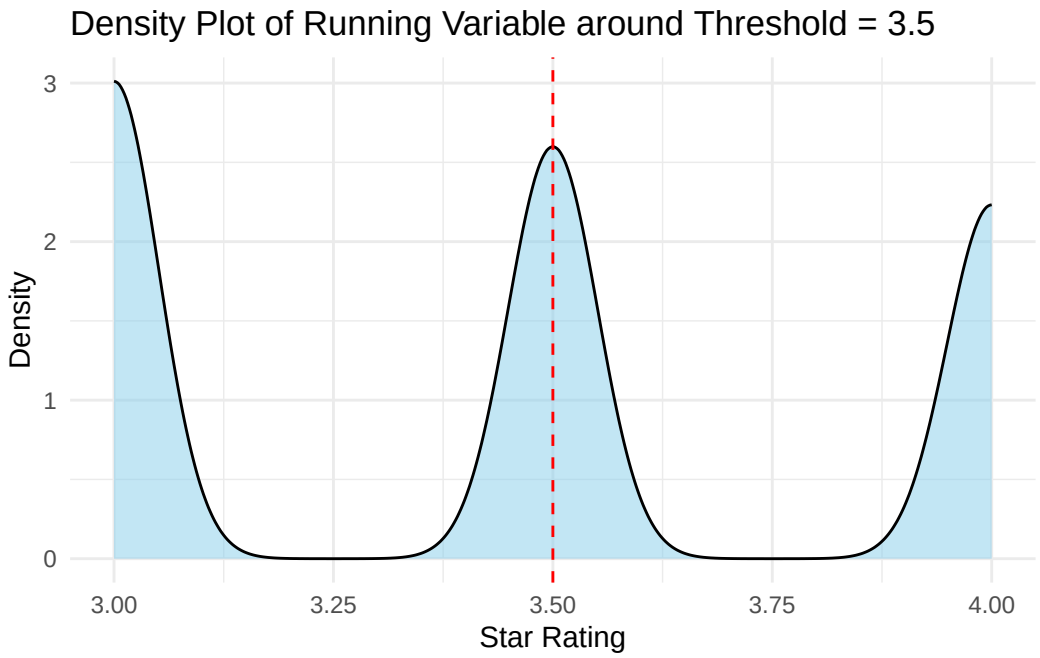


I am getting "null" for all, so it is not possible to tell.

8.Examine (graphically) whether contracts appear to manipulate the running variable. In other words, look at the distribution of the running variable before and after the relevent threshold values. What do you find?.

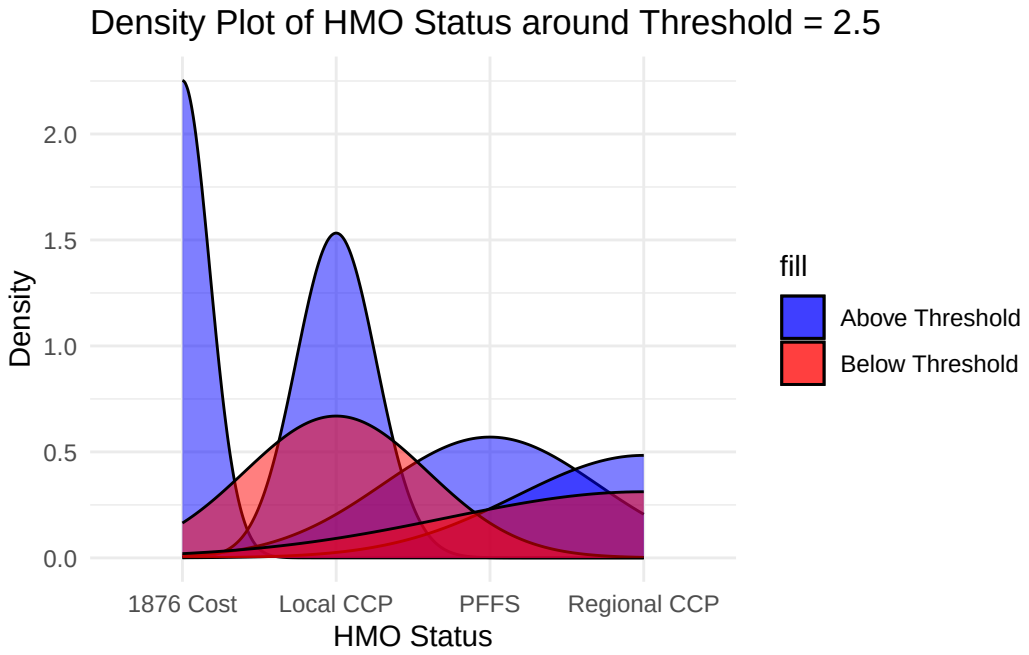


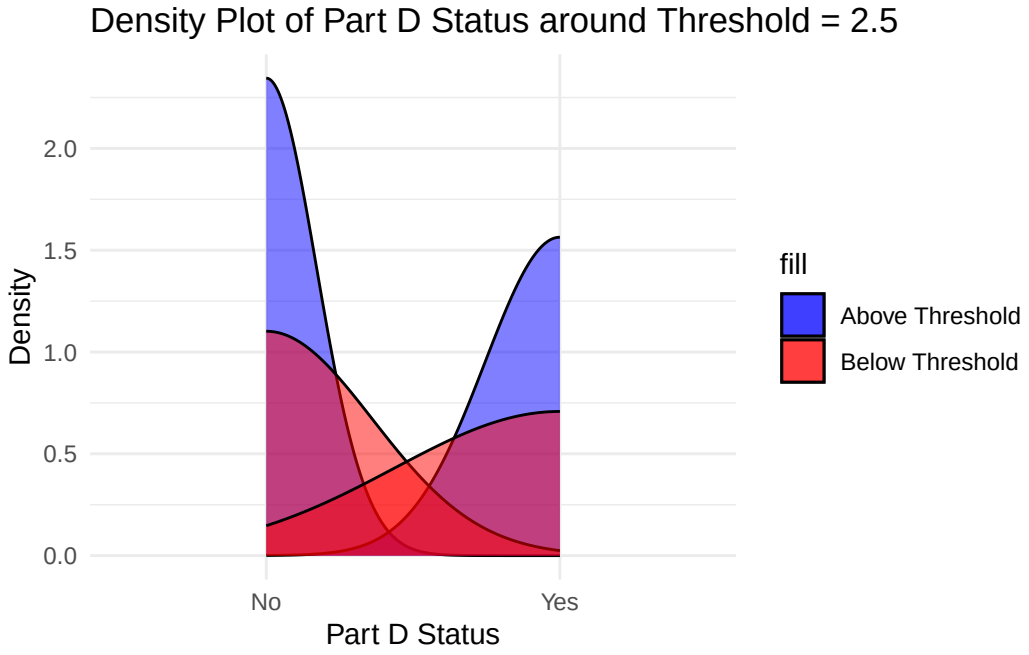


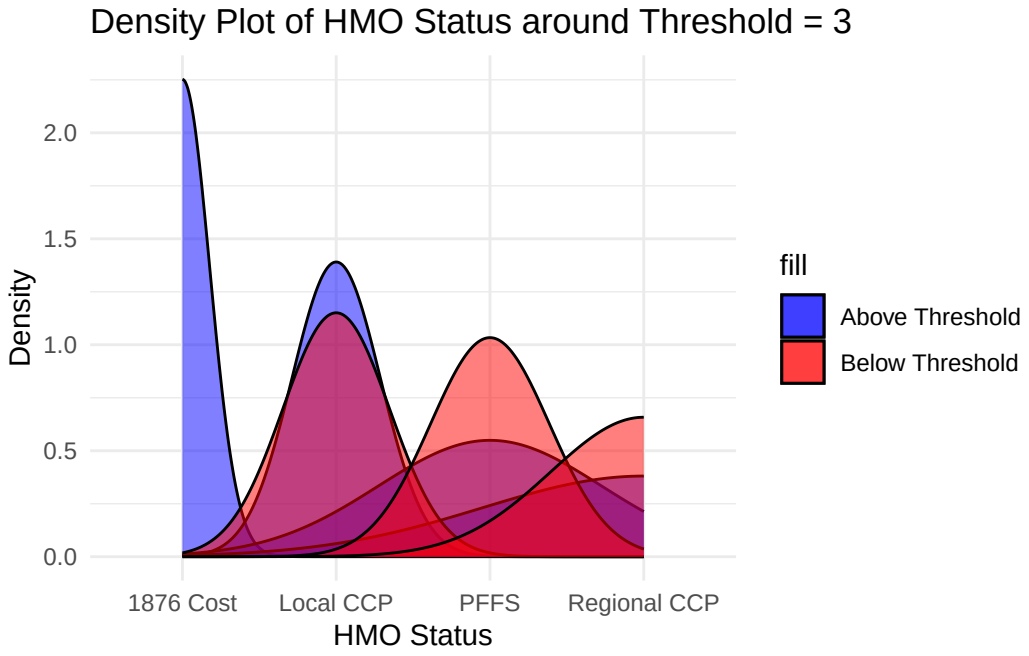


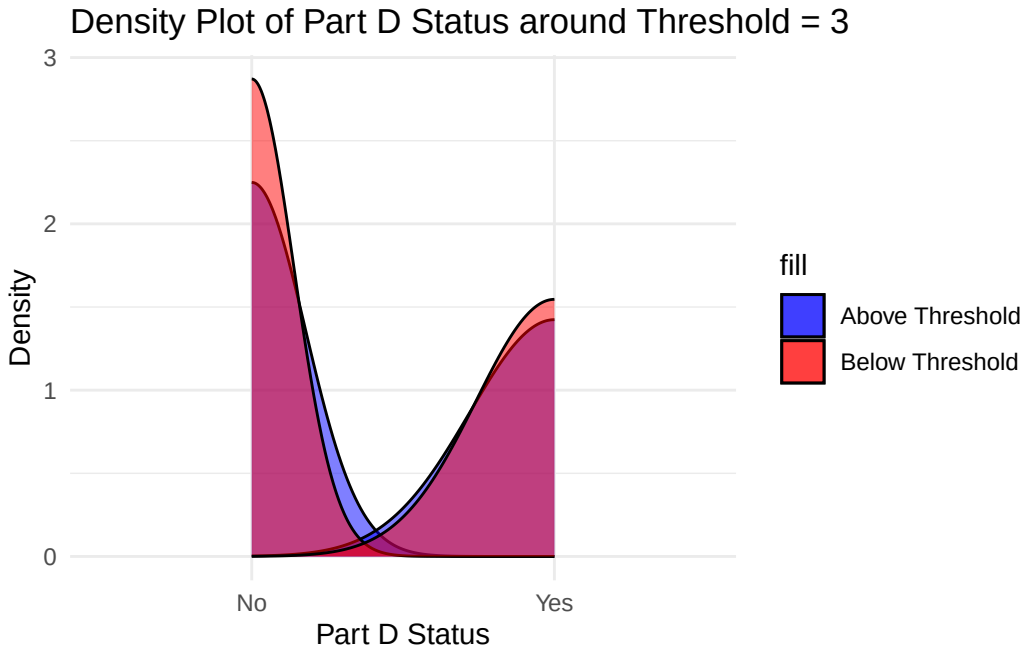
It peaks at 3, 3.5, and 4.

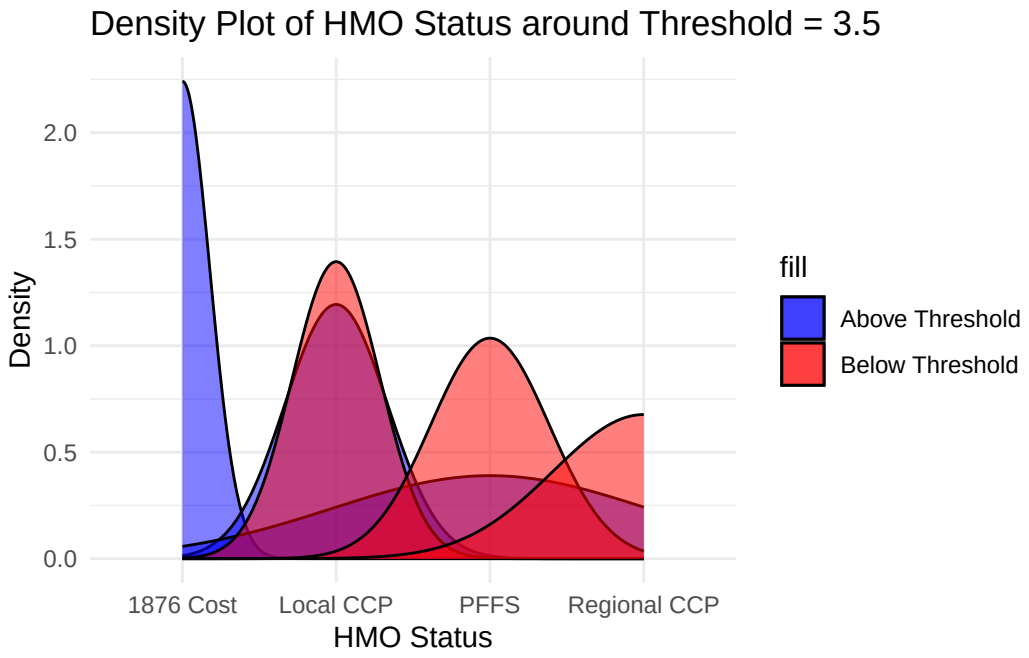
9.Similar to question 8, examine whether plans just above the threshold values have different characteristics than contracts just below the threshold values. Use HMO and Part D status as your plan characteristics.

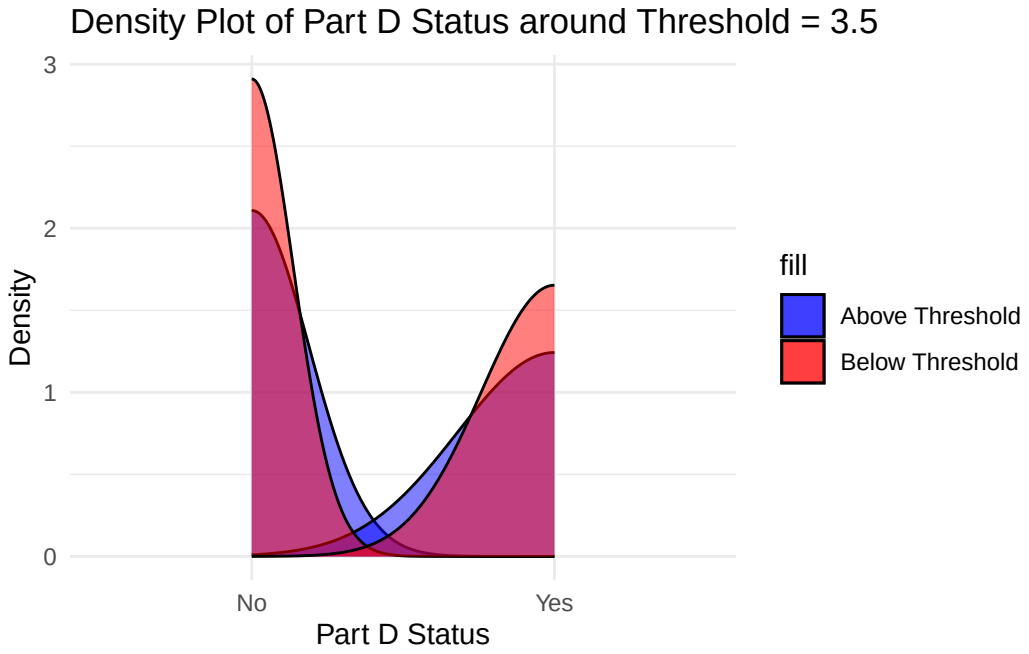





Density Plot of HMO Status around Threshold = 3
Density
2.0
1.5
1.0
0.5
0.0
1876 Cost
Local CCP
PFFS
Regional CCP
HMO Status
fill
Above Threshold
Below Threshold


Density Plot of Part D Status around Threshold = 3
Density
3
2
1
0
No
Yes
Part D Status
fill
Above Threshold
Below Threshold


Density Plot of HMO Status around Threshold = 3.5
Density
2.0
1.5
1.0
0.5
0.0
1876 Cost
Local CCP
PFFS
Regional CCP
HMO Status
fill
Above Threshold
Below Threshold


Density Plot of Part D Status around Threshold = 3.5
Density
3
2
1
0
No
Yes
Part D Status
fill
Above Threshold
Below Threshold

10.Summarize your findings from 5-9. What is the effect of increasing a star rating on enrollments? Briefly explain your results.

Number 5, 6 and 7 clearly seems to be wrong at this point, so it is hard to clearly tell what is happening. My guess is that it should have positive impact on enrollment rate.