DAT-3113 PROJECT

Hospital West

AUTHOR
Luke Stucky

PUBLISHED
December 3, 2023

INTRODUCTION

In this project, we'll be working as public policy analysts for Senator James Lankford, focusing on healthcare quality and patient safety, particularly regarding hospital-acquired infections. Our goal is to understand how hospital safety impacts patients' perception of care quality and whether this perception varies based on the size of the hospital.

PART A: DATASET

Importing Dataset

Summarizing Data Before Wrangling, Munging, and Preparation

Summarizing Data

```
── Data Summary ────────────────────────
                           Values
Name                       Piped data
Number of rows             940
Number of columns          10
_______________________
Column type frequency:
  numeric                  10
________________________
Group variables            None

── Variable type: numeric ──────────────────────────────────────────────────────
   skim_variable          n_missing        mean          sd         p50
 1 CLABSI_Device_Days           183     4962.         7415.        2131
 2 CLABSI_Cases                 183        4.17          8.04         1
 3 CLABSI_National_Compare      505        0.0506        0.432        0
 4 cdiff_Patient_Days           142    35735.        43193.       18288
 5 cdiff_Cases                  142       10.4          17.3          4
 6 cdiff_National_Compare       323        0.509         0.519        1
 7 total_payments_received      330 42422622.     54184999.    26118708.
 8 total_discharges             329     2023.         1953.        1384
 9 cms_risk_score               330        2.01          0.503        1.93
10 nbr_stars_overall            283        3.16          0.884        3
```

After importing the data set, I found that there are many missing values at first glance. Then when looking closer, I found that there are some hospitals with just a couple missing values, others with many, and even some that had all missing values. On average, excluding the missing variables, western hospitals had better infection rates than the national average in both cdiff and CLABSI. The average star rating was 3.19 while a rating of 3 was the median.

PART B: DATA WRANGLING, MUNGING, AND PREPARATION

This is the section for all of the data transformation and preparation so that proper analyses can be conducted.

Removing useless data

In this section, I found all the hospitals that did not turn in any data and removed them. This totaled out to be 118 hospitals out of the 940 starting variables. I then removed any variables that would not be useful for the rest of the project.

Imputing for missing variables

In this section, I used different imputation methods to replace the rest of the NAs in the data set. I incorporated mean imputation, mode imputation, and removed cases with dependent variables.

Creating the variable for Hospital Size

Using clustering to identify three clusters of discharges, the variable hospital size is created.

Creating a new dataset for states

In this section, I created a new dataset separating the data by individual states. This allows for better visualization of different variables.

PART C: SUMMARY MEASURES AND DATA VISUALIZATION

Summary Measures

For ratio and interval data, include the complete set of summary measures, including skewness and kurtosis. Use tables and other methods to show categorical variables.
###Installing package

Using the psych package allows for the describe function to work.

Function to find mode

CLABSI Summary

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	5658.62	7560.699	3297	4125.691	3974.851	3	61179	61176	3.188887	14.22901	317.5195

```
$value
[1] 4725.279

$frequency
[1] 16

IQR : 6315

Variability : 57164164
```

CLABSI Device Days shows a mean of 5659, a median of 3297, and a mode of 4725. The IQR is 6315 which shows where 50% of the data lies. There is an extremely high standard deviation which means that on average, a value is far from the mean. It has a high positive skew of 3.2 which means that there is a significant amount of values above the mean and it is positively skewed. The kurtosis is 17.29 which shows that it is heavily peaked.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	4.784282	8.163333	2	2.999315	2.9652	0	63	63	3.713256	17.68734	0.3428278

```
$value
[1] 0

$frequency
[1] 190

IQR : 6

Variability : 66.64001
```

CLABSI Cases shows a mean of 4.78, a median of 2, and a mode of 0. With a standard deviation of 8.16, a selected value is usually off by +/- 8.16 from the mean of 4.78. It has a skewness of 3.71 and a kurtosis value of 17.69 which show that it is positively skewed and it is highly peaked.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	0.0246914	0.345732	0	0	0	-1	1	2	0.3832448	5.253918	0.0145194

```
$value
[1] 0

$frequency
[1] 499

IQR : 0
```

```
Variability : 0.1195306
```

CLABSI National Comparison shows a mean of .025, a median of 0, and a mode of 0. The frequency of 0 is 499 times which explains why the mean is so close to 0.

cdiff Summary

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	42634.13	43033.11	30992	35410.21	35115.38	417	283659	283242	1.753586	3.948487	1807.221

```
$value
[1] 34263.14

$frequency
[1] 11

IQR : 51644.5

Variability : 1851848442
```

Cdiff Patient Days shows a mean of 42634.13 and a median of 30992. With a mean higher than the median, it is positively skewed. After viewing the skewness of 1.75, this is proven to be true.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	12.39647	17.56079	7	8.775381	8.8956	0	132	132	3.220156	13.94892	0.7374838

```
$value
[1] 0

$frequency
[1] 81

IQR : 13

Variability : 308.3813
```

Cdiff Cases shows a mean of 12.4 and a median value of 7. This means that the hospitals with more than 7 cases have a large number of cases. This also explains why the skewness is 3.22 or positively skewed.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	0.5679012	0.516744	1	0.5978022	0	-1	1	2	-0.502827	-1.220816	0.0217012

```
$value
[1] 1

$frequency
[1] 328
```

```
IQR : 1

Variability : 0.2670244
```

The cdiff National Comparison shows a mean of 0.57, a median of 1, and a mode of 1. The frequency of 1 is 328 which shows that a greater percentage of hospitals in the west have below average cdiff cases.

Total Payments Summary

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
X1	1	567	43885093	52116210	32262133	35113118	27634389	943251	560769926	559826675	4.788228	36.10763

```
$value
[1] 41878958

$frequency
[1] 74

IQR : 38220855

Variability : 2.716099e+15
```

Total payments shows a mean of 43885093 and a median of 32262133. The IQR is 38220855 meaning that 50 % of the data lay within that range. The extremely high kurtosis value of 36.11 indicates an extreme distribution.

Total Discharges Summary

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	2059.764	1805.453	1704	1781.047	1500.391	37	17019	16982	2.179599	9.461294	75.82189

```
$value
[1] 1966.977

$frequency
[1] 74

IQR : 2052.5

Variability : 3259659
```

Total Discharges shows a mean of 2059.76 and a median of 1704 which identify the central tendency of the discharges.The high kurtosis value of 9.46 suggests a more extreme distribution.

cms Risk Score Summary

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	2.042779	0.4649276	2.032636	2.011083	0.3682243	0.6556	3.915	3.2594	0.673245	1.019182	0.0195251

```
$value
[1] 2.032636

$frequency
[1] 74
```

```
IQR : 0.473
```

```
Variability : 0.2161577
```

Cms Risk Score shows a mean of 2.04 and a median of 2.03 meaning that the mean is extremely close to the center. The variance of 0.22 indicates that the values hover pretty closely to the mean. It has a skewness of 0.67 which means that it is skewed to the right slightly, and it has a kurtosis value of 1.02 which indicates slight peakedness.

nbr Stars Overall Summary

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	567	3.190476	0.8811218	3	3.217582	1.4826	1	5	4	-0.2398211	-0.2189365	0.0370036

```
$value
[1] 3

$frequency
[1] 236
```

```
IQR : 1
```

```
Variability : 0.7763756
```

Nbr Stars Overall shows a mean of 3.19, a median of 3. The value 3 occurs 236 times and the majority is 3 and above which explains the above average rating of 3.19. Assuming that the average rating for a hospital is 3, hospitals in the west are rated higher on average.

Data Visualization

Data visualization is used to help understand the data and see trends between different variables.

Contingency Tables

CLABSI National Comparison to nbr stars overall

```
    -1   0   1
  1   2  14   1
  2  10  85   4
  3  10 200  26
  4   5 174  10
  5   0  26   0
```

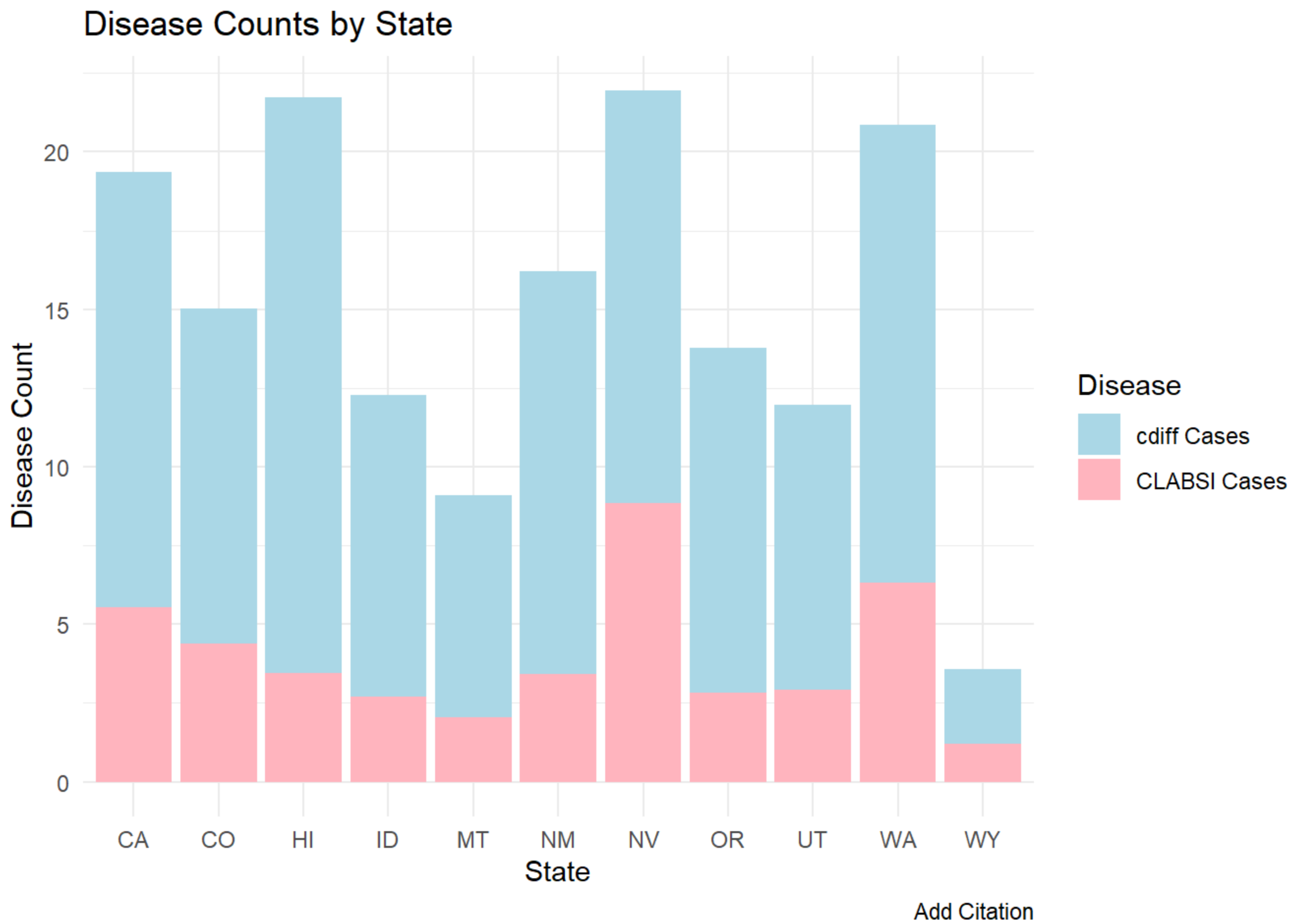
Three starts are most common and the majority of them come from hospital with average CLABSI cases. There are barely any one star and five star ratings and the majority of them also come from hospitals with average CLABSI cases.

cdiff National Comparison to nbr stars overall

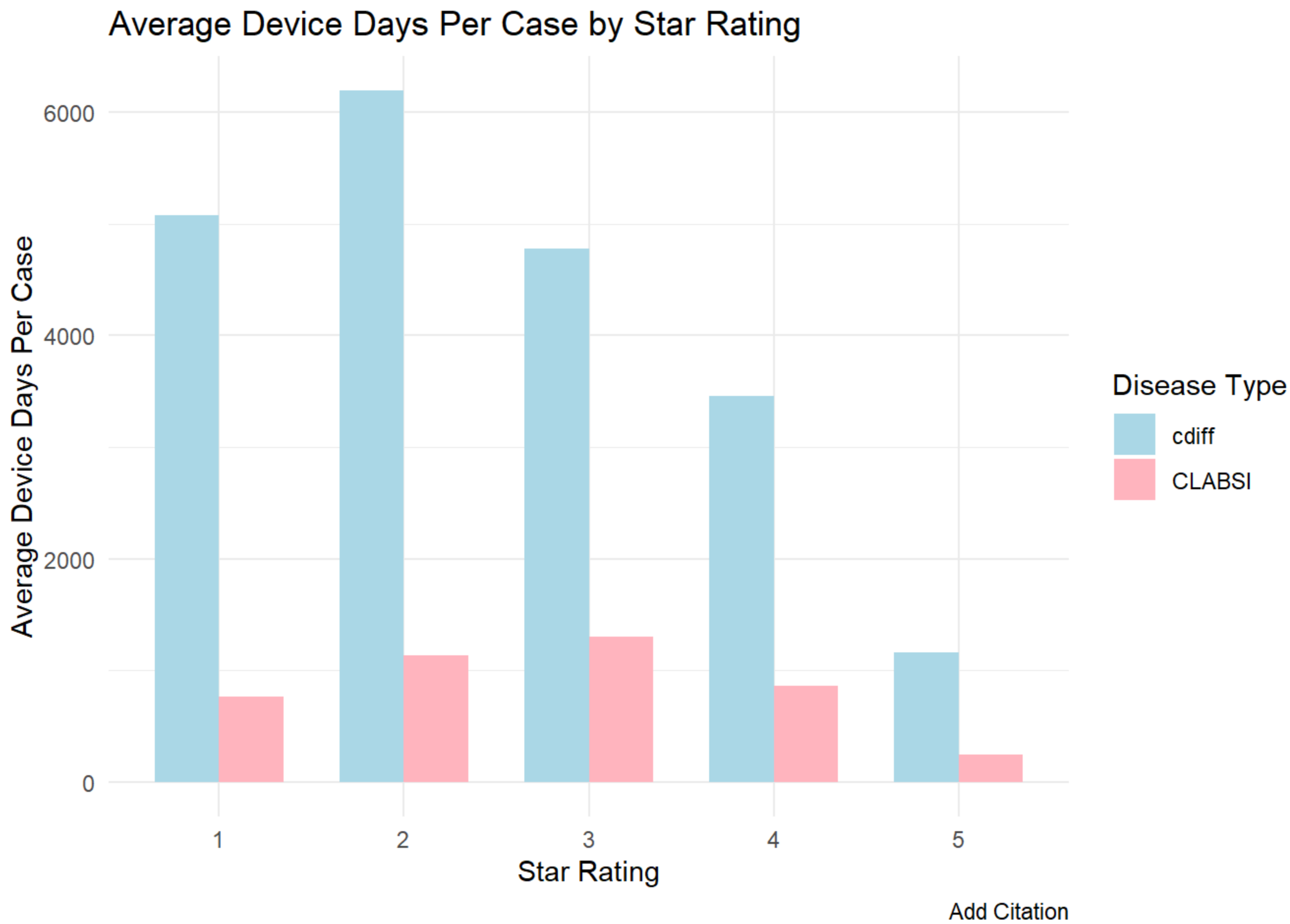
```
    -1   0   1
  1   0  10   7
  2   0  31  68
  3   3  97 136
  4   2  86 101
  5   1   9  16
```

The contingency table for cdiff national comparison and nbr stars overall shows the majority of the ratings being three star. The majority of five star hospitals were above average in national case comparison. One hospital that got a five star rating had a below average national case comparison. There were few total one and two star hospitals.

Stacked Bar Chart



This stacked bar chart analyzes the number of both cdiff and CLABSI cases in each state in the west. From the chart, a few conclusions can be drawn. The more populated states typically have more cases and the less have fewer. Hawaii is a remote state and it has the most cdiff cases.

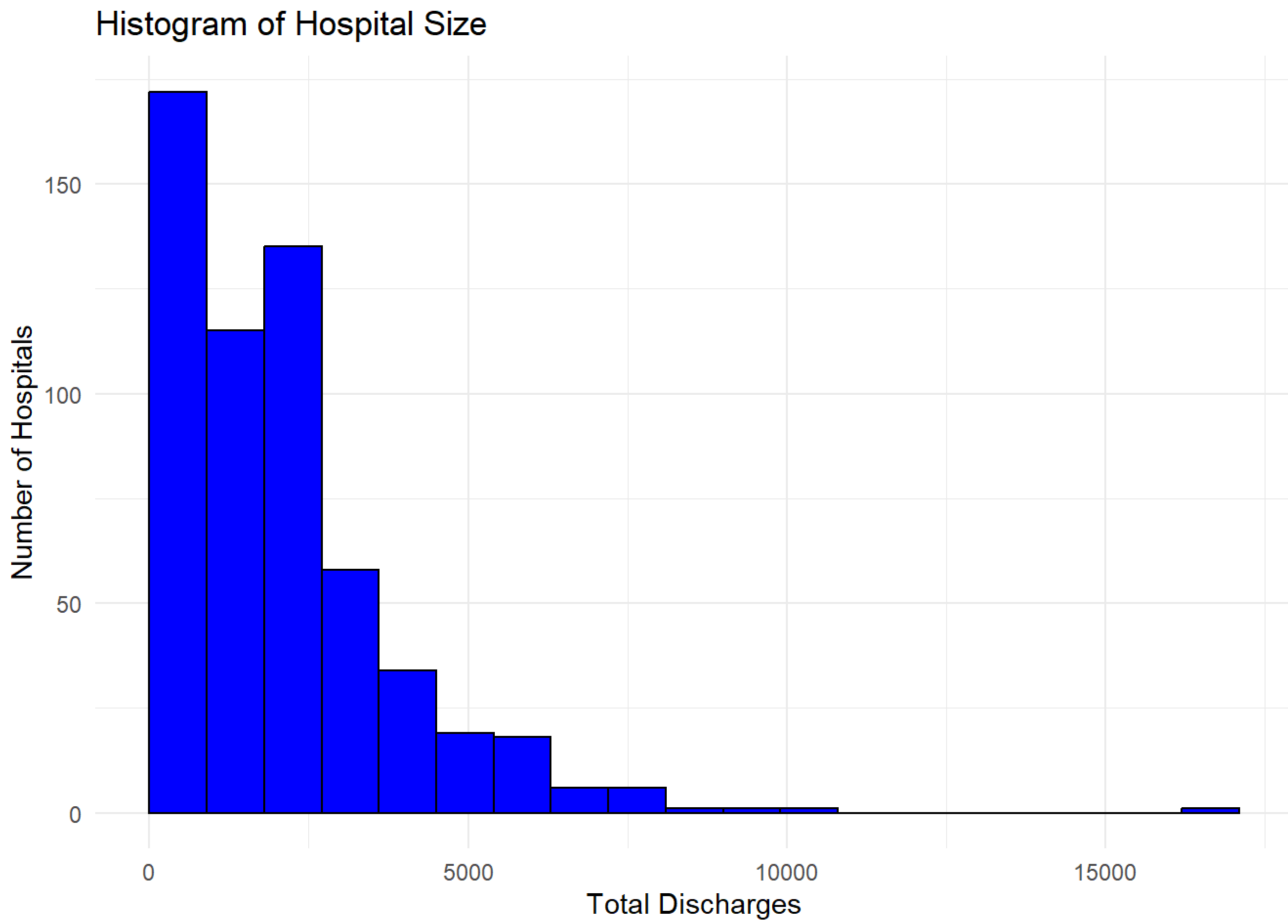


Creating a dodged bar chart looking at stars overall and the average days for each disease, There is a negative slope which is expected. The greater number of days in the hospital, the lower the star rating.

Density Plot

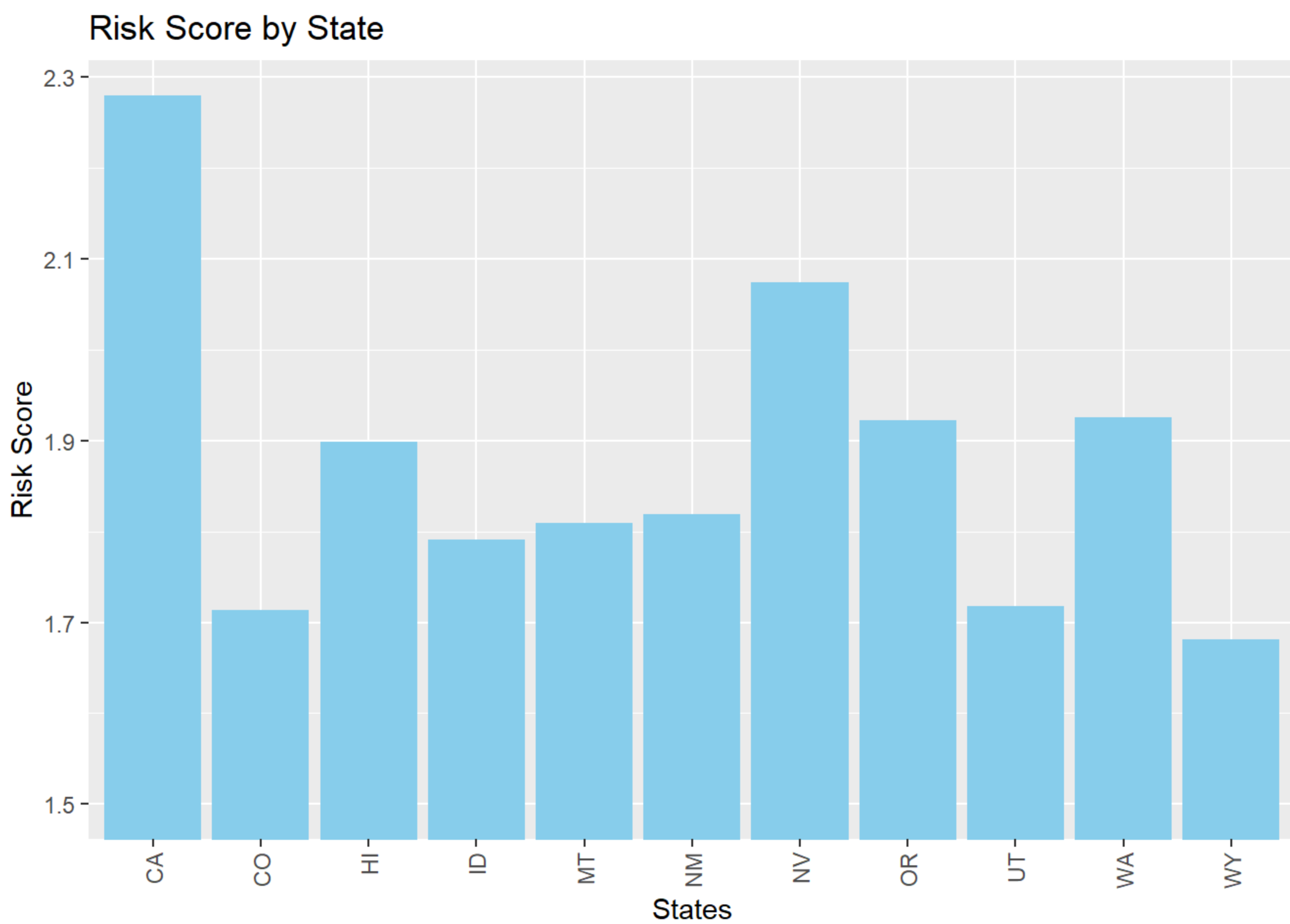
Creating a density plot enabled viewing if the quality perception of hospitals varied by size of hospital hospital. In interpreting the plot, there appears to be a trend. The higher star ratings are more common for medium sized hospitals, and the lower star ratings are more common for the small hospitals.

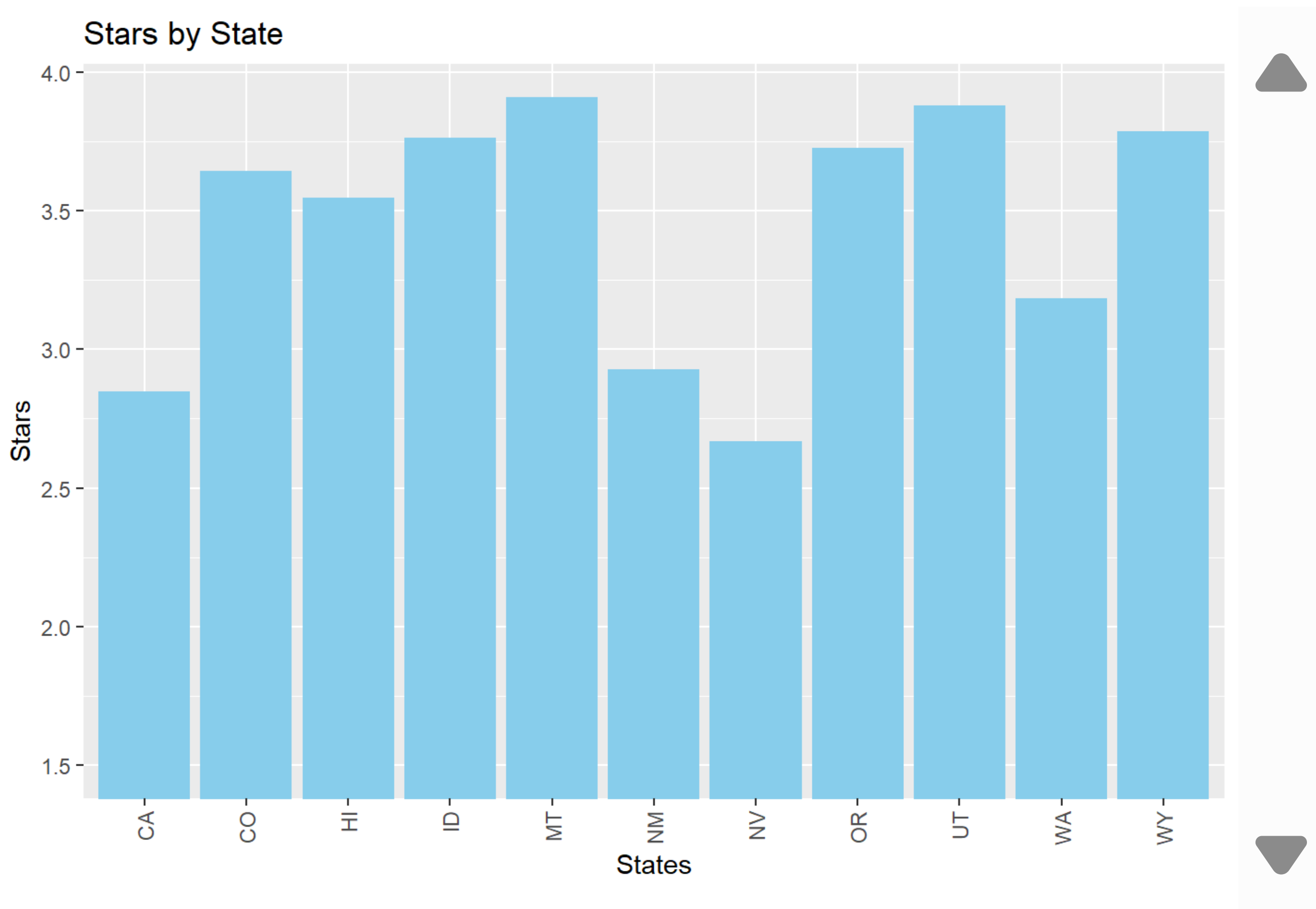
Histogram



The graph shows a negative slope. The number of discharges is typically small for the majority of hospitals. There is an outlier that is different by 5000 discharges.

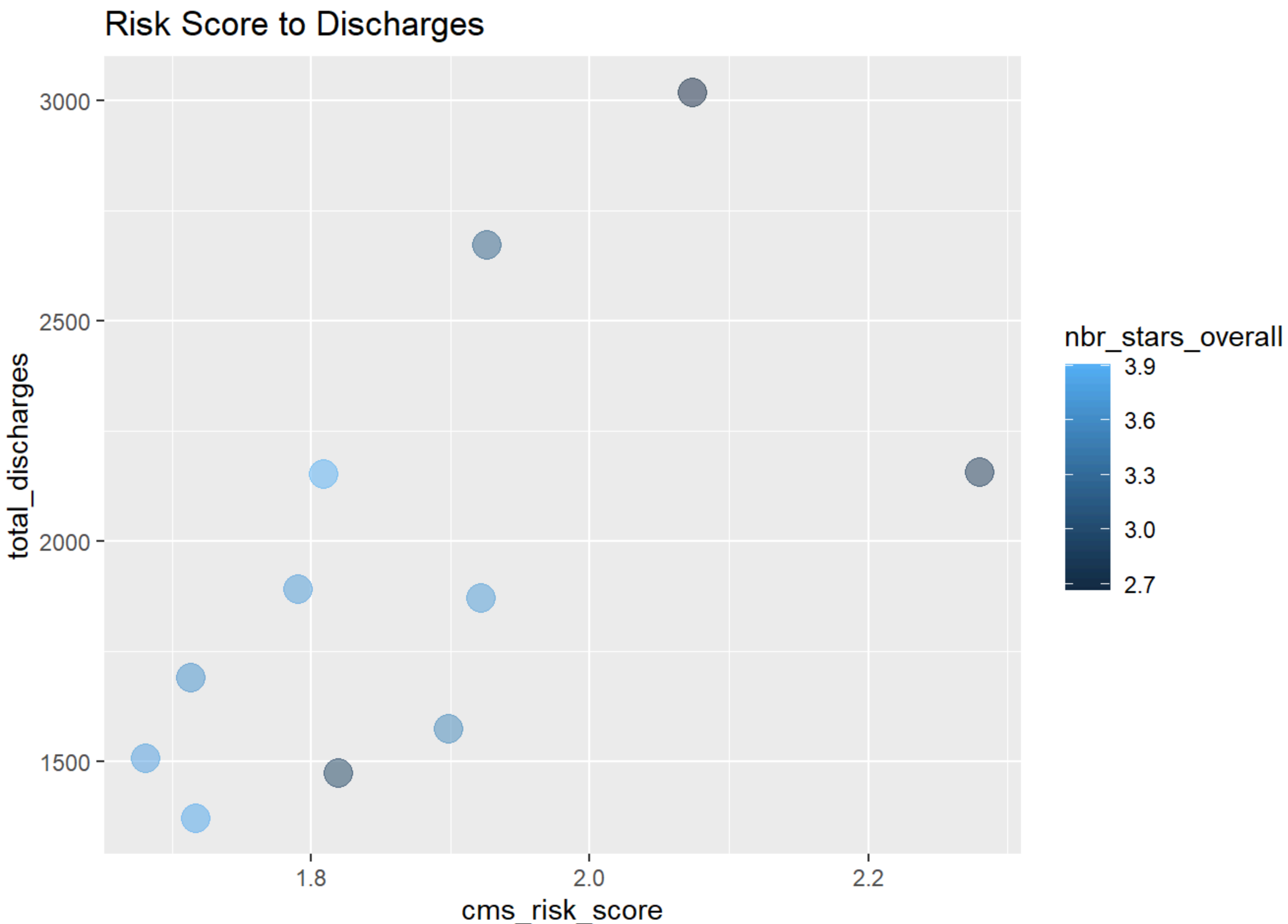
Bar Charts


Risk Score by State
Risk Score
2.3
2.1
1.9
1.7
1.5
CA
CO
HI
ID
MT
NM
NV
OR
UT
WA
WY
States


Stars by State
Stars
4.0
3.5
3.0
2.5
2.0
1.5
CA
CO
HI
ID
MT
NM
NV
OR
UT
WA
WY
States

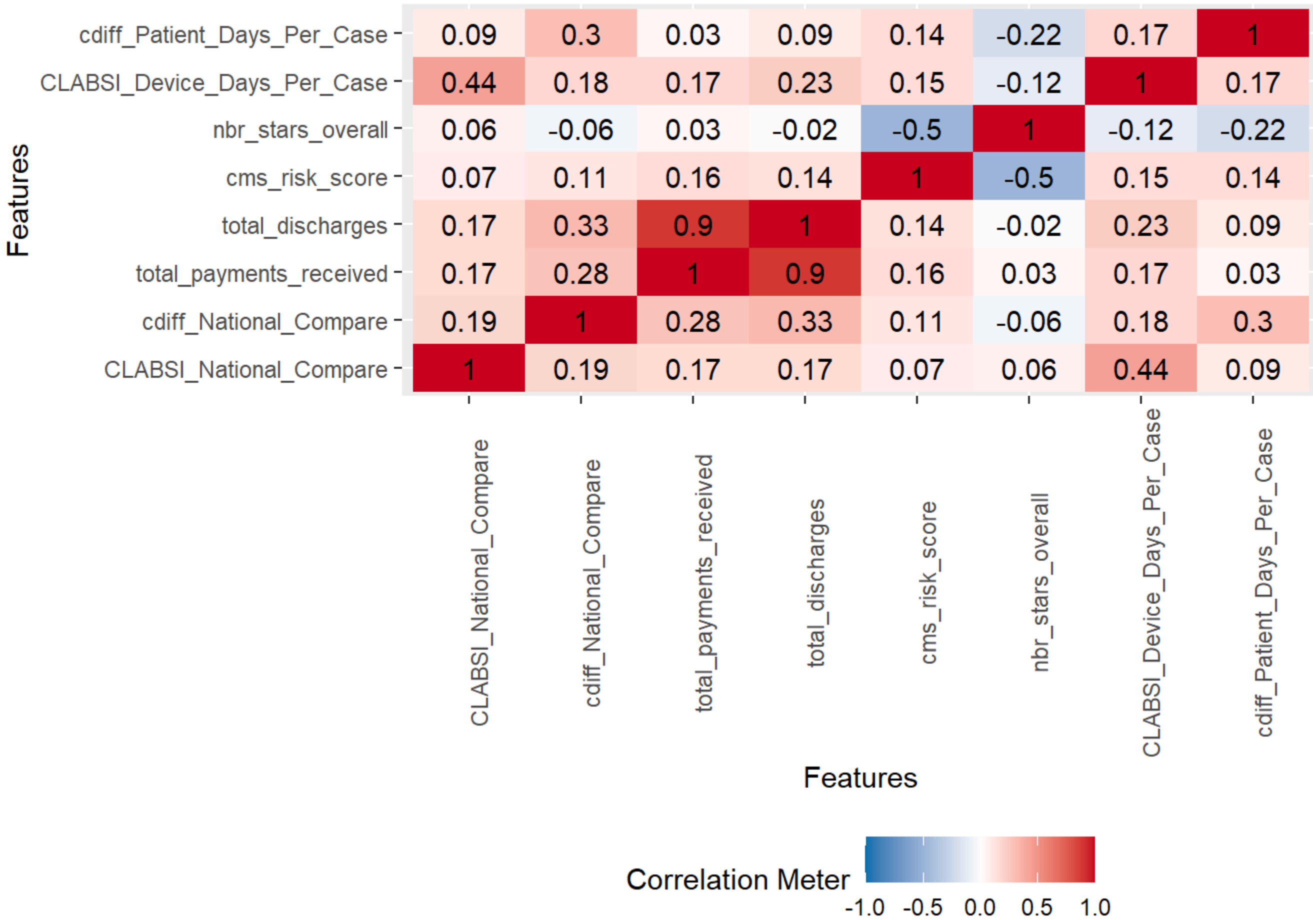
When breaking the risk score and the stars down by state, bar charts can be created for insights. At first glance, it appears that there is negative correlation between the two, which is expected. Seeing this visualization, it will be important to check for correlation later.

Scatterplot



Using the aggregated data to compare risk score and discharges, a scatterplot can be created to view the relationship. A positive slope is expected as the number of discharges should increase with a higher risk score. This is exactly what appears after the creation of the scatterplot. By adding in the color changing based on star rating, it can be observed that typically, the lower the risk score, the higher the star rating.

Heat Map



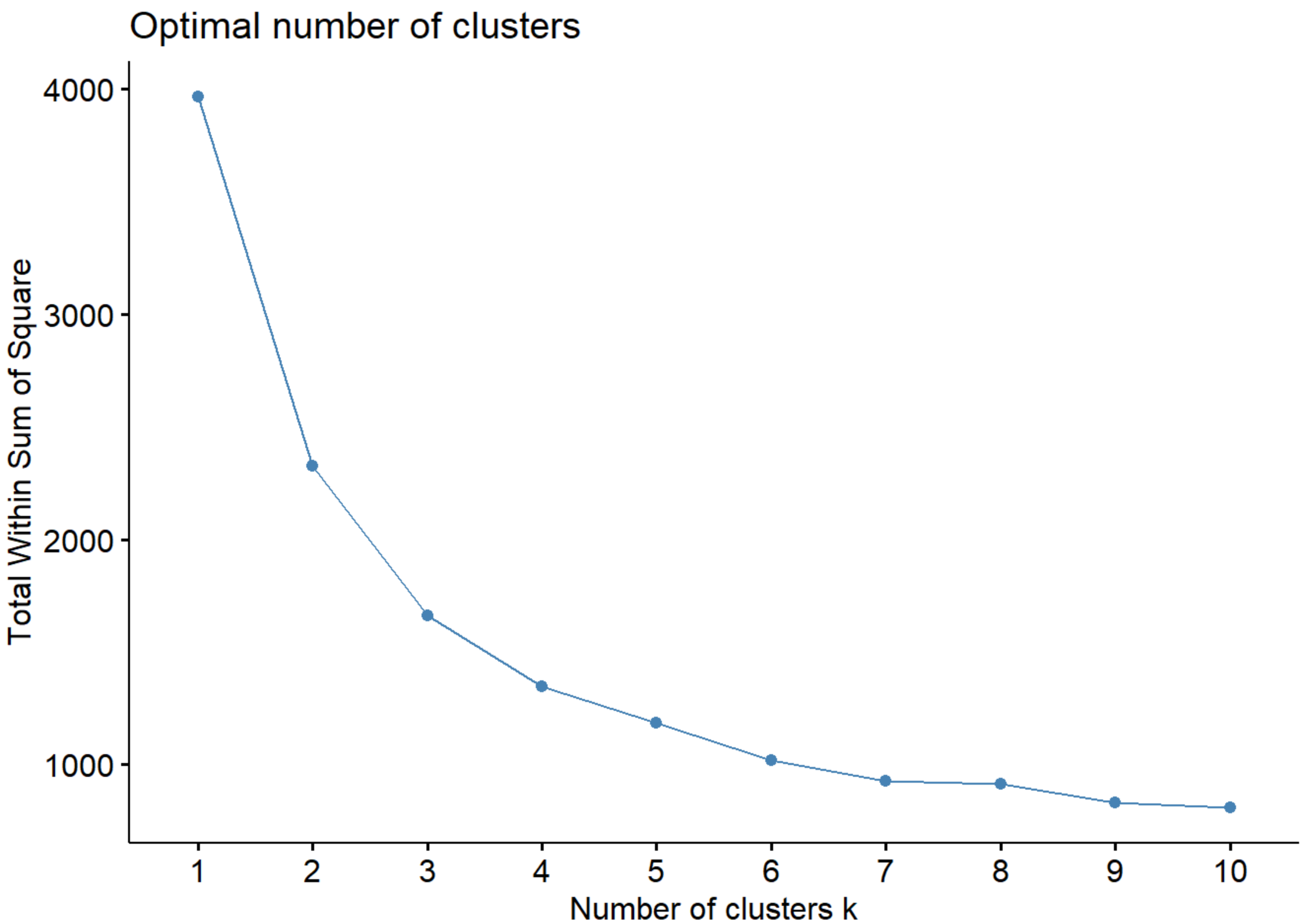

```
[1] -0.4967444
```

Running a heat map for correlation allows for contributing factors to star ratings to be identified. The first observation that is heavily correlated is payments and discharges, which makes sense, but is not useful. The correlation between star rating and risk score is negatively correlated by -0.5. This shows that as the risk score goes up, the star rating typically goes down.

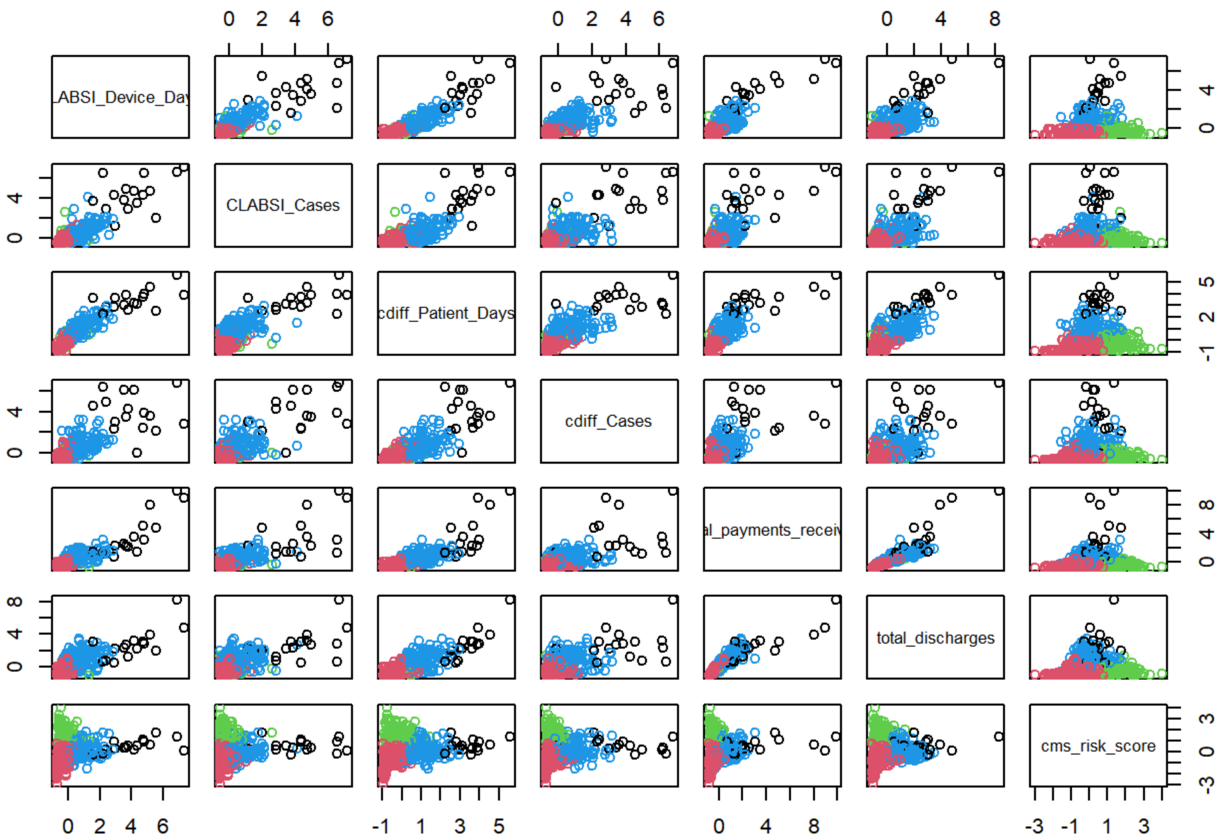
PART D: ANALYTIC MODELS

Clustering

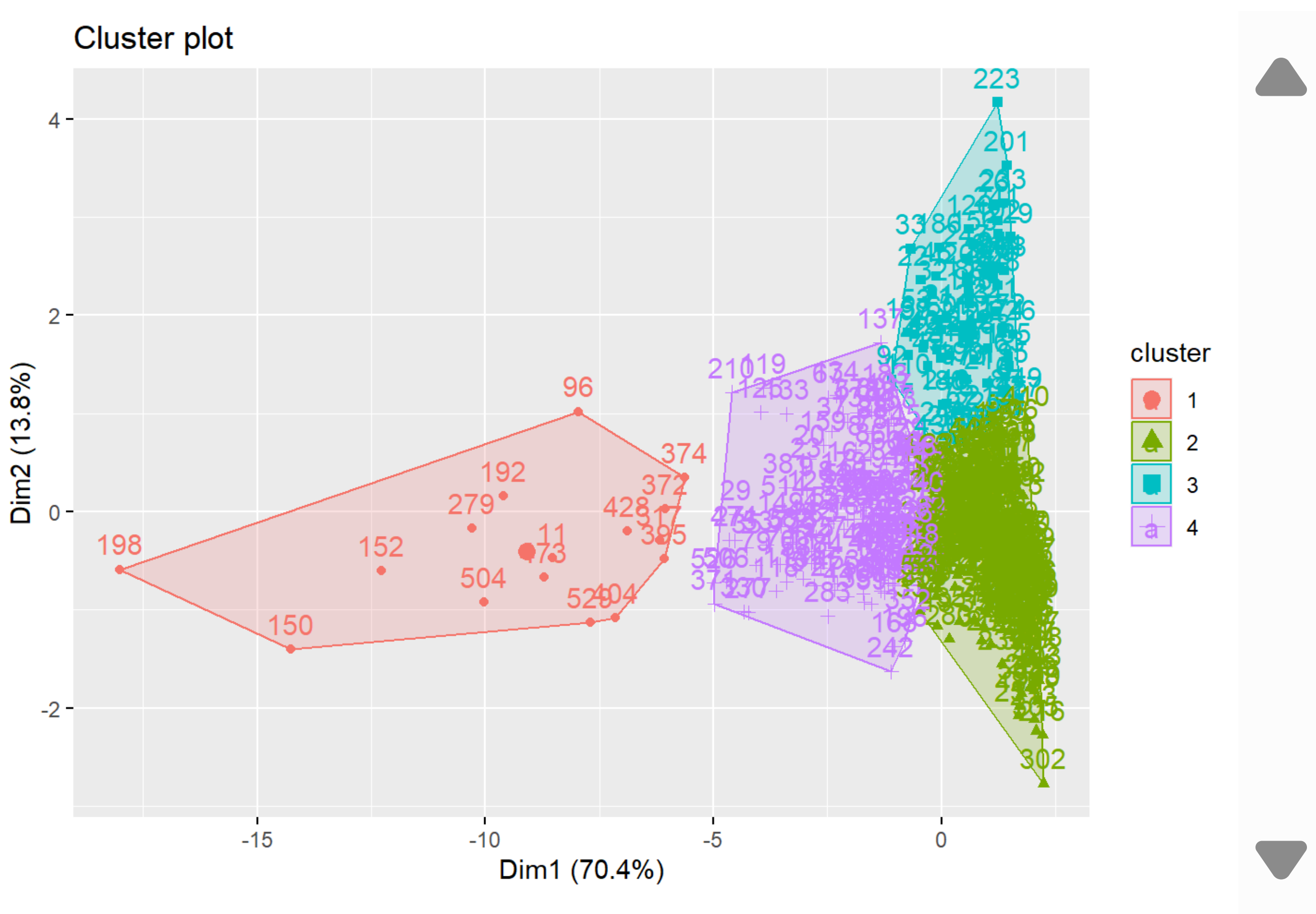
Clustering analysis is important for analytic modeling as it allows for patterns to be discovered and predictions to be made.



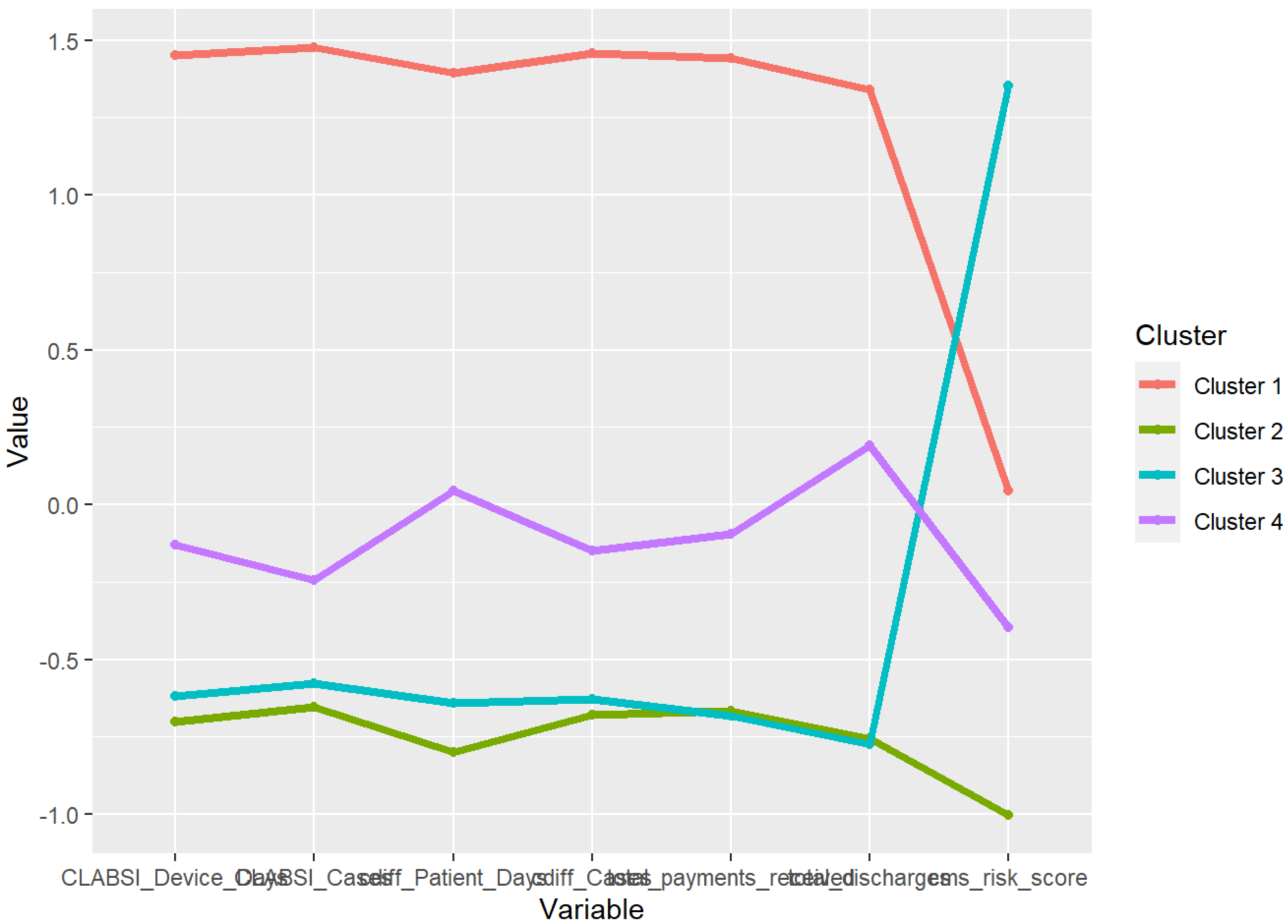

```
[1]  16 337  76 138
```



CLABSI_Device_Days
CLABSI_Cases
cdiff_Patient_Days
cdiff_Cases
total_payments_received
total_discharges
cms_risk_score



Cluster plot
Dim1 (70.4%)
Dim2 (13.8%)
cluster
1
2
3
4




```
# A tibble: 4 × 9
  cluster     N xCLABSI_Device_Days xCLABSI_Cases xcdiff_Patient_Days
    <int> <int>               <dbl>         <dbl>               <dbl>
1       1    16              36785.         40.2              191220.
2       2   337               2293.          1.72              20116.
3       3    76               3587.          3.05              32505.
4       4   138              11410.          9.12              85976.
# ℹ 4 more variables: xcdiff_Cases <dbl>, xtotal_payments_received <dbl>,
#   xtotal_discharges <dbl>, xcms_risk_score <dbl>
```

After scaling the data, it is observed that 4 clusters would be most beneficial. The cluster plot shows four fairly distinct sections that will be useful for predictive analysis.By profiling the data, the centers can be observed. This shows the distinction between each of the clusters. There is a little overlap that can be observed. After appending the data, it appears that the large hospitals have a high values of everything and the small hospitals have lower values. The two clusters in the middle are fairly similar besides the risk score which varies.

Regression

The regression analysis will be useful for checking the relationship between the variables. It will help determine which relationships matter and which do not.

```
Start:  AIC=-345.35
nbr_stars_overall ~ CLABSI_Device_Days + CLABSI_Cases + cdiff_Patient_Days +
    cdiff_Cases + total_discharges + cms_risk_score + CLABSI_Device_Days_Per_Case +
```

```
    cdiff_Patient_Days_Per_Case + BelowAverageCLABSI + AboveAverageCLABSI +
    BelowAveragecdiff + AboveAveragecdiff

                              Df Sum of Sq    RSS     AIC
- BelowAveragecdiff            1     0.182 294.72 -347.00
- AboveAverageCLABSI           1     0.722 295.26 -345.96
<none>                                     294.54 -345.35
- BelowAverageCLABSI           1     1.815 296.36 -343.87
- CLABSI_Cases                 1     2.218 296.76 -343.10
- CLABSI_Device_Days_Per_Case  1     2.868 297.41 -341.85
- CLABSI_Device_Days           1     2.993 297.54 -341.62
- AboveAveragecdiff            1     3.865 298.41 -339.96
- cdiff_Patient_Days_Per_Case  1     4.502 299.04 -338.75
- cdiff_Cases                  1     6.227 300.77 -335.49
- total_discharges             1     7.245 301.79 -333.57
- cdiff_Patient_Days           1    12.069 306.61 -324.58
- cms_risk_score               1    85.384 379.93 -203.02

Step:  AIC=-347
nbr_stars_overall ~ CLABSI_Device_Days + CLABSI_Cases + cdiff_Patient_Days +
    cdiff_Cases + total_discharges + cms_risk_score + CLABSI_Device_Days_Per_Case +
    cdiff_Patient_Days_Per_Case + BelowAverageCLABSI + AboveAverageCLABSI +
    AboveAveragecdiff

                              Df Sum of Sq    RSS     AIC
- AboveAverageCLABSI           1     0.747 295.47 -347.56
<none>                                     294.72 -347.00
- BelowAverageCLABSI           1     1.926 296.65 -345.31
- CLABSI_Cases                 1     2.091 296.82 -344.99
- CLABSI_Device_Days_Per_Case  1     2.846 297.57 -343.55
- CLABSI_Device_Days           1     2.871 297.59 -343.50
- AboveAveragecdiff            1     3.799 298.52 -341.74
- cdiff_Patient_Days_Per_Case  1     4.445 299.17 -340.51
- cdiff_Cases                  1     6.735 301.46 -336.19
- total_discharges             1     7.243 301.97 -335.23
- cdiff_Patient_Days           1    12.400 307.12 -325.63
- cms_risk_score               1    86.076 380.80 -203.71

Step:  AIC=-347.56
nbr_stars_overall ~ CLABSI_Device_Days + CLABSI_Cases + cdiff_Patient_Days +
    cdiff_Cases + total_discharges + cms_risk_score + CLABSI_Device_Days_Per_Case +
    cdiff_Patient_Days_Per_Case + BelowAverageCLABSI + AboveAveragecdiff

                              Df Sum of Sq    RSS     AIC
<none>                                     295.47 -347.56
- BelowAverageCLABSI           1     1.547 297.02 -346.60
- CLABSI_Device_Days_Per_Case  1     3.017 298.49 -343.80
- AboveAveragecdiff            1     4.104 299.58 -341.74
- cdiff_Patient_Days_Per_Case  1     4.527 300.00 -340.94
- CLABSI_Cases                 1     4.910 300.38 -340.22
- CLABSI_Device_Days           1     5.046 300.52 -339.96
```

```
- cdiff_Cases                 1     7.074 302.55 -336.15
- total_discharges            1     7.269 302.74 -335.78
- cdiff_Patient_Days          1    12.742 308.21 -325.63
- cms_risk_score              1    87.253 382.72 -202.86

Call:
lm(formula = nbr_stars_overall ~ CLABSI_Device_Days + CLABSI_Cases + 
    cdiff_Patient_Days + cdiff_Cases + total_discharges + cms_risk_score + 
    CLABSI_Device_Days_Per_Case + cdiff_Patient_Days_Per_Case + 
    BelowAverageCLABSI + AboveAveragecdiff, data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.58540 -0.46397  0.01301  0.55468  1.74669 

Coefficients:
                                Estimate   Std. Error t value
(Intercept)                  5.017228166  0.144196752  34.794
CLABSI_Device_Days           0.000039291  0.000012751   3.081
CLABSI_Cases                -0.028513653  0.009380295  -3.040
cdiff_Patient_Days          -0.000011378  0.000002324  -4.897
cdiff_Cases                  0.013534744  0.003709632   3.649
total_discharges             0.000104724  0.000028316   3.698
cms_risk_score              -0.875411414  0.068319288 -12.814
CLABSI_Device_Days_Per_Case -0.000068394  0.000028705  -2.383
cdiff_Patient_Days_Per_Case -0.000020440  0.000007003  -2.919
BelowAverageCLABSI           0.257319162  0.150834420   1.706
AboveAveragecdiff            0.210160067  0.075625631   2.779
                                        Pr(>|t|)    
(Intercept)                 < 0.0000000000000002 ***
CLABSI_Device_Days                      0.002162 ** 
CLABSI_Cases                            0.002479 ** 
cdiff_Patient_Days                    0.00000128 ***
cdiff_Cases                             0.000289 ***
total_discharges                        0.000238 ***
cms_risk_score              < 0.0000000000000002 ***
CLABSI_Device_Days_Per_Case             0.017524 *  
cdiff_Patient_Days_Per_Case             0.003658 ** 
BelowAverageCLABSI                      0.088572 .  
AboveAveragecdiff                       0.005638 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.729 on 556 degrees of freedom
Multiple R-squared:  0.3276,    Adjusted R-squared:  0.3155 
F-statistic: 27.09 on 10 and 556 DF,  p-value: < 0.00000000000000022
```

Through running the regression model, 32.76% of the variability in the star rating can be explained. The null hypothesis of the model is that all the coefficients in the model are equal to zero, and therefore do not affect the star rating. This can be rejected with the low p-value indicating that there is a variable contributing to the change in stars. On average, the model is off by +/- .144 stars. With every one unit increase in cms risk score, the star rating goes down by -0.875.

PART E: ETHICAL IMPLICATIONS

When dealing with hospital data, it is important to look at the ethical considerations to ensure responsible and transparent data science practices. It is crucial to address the ethical implications that may arise during the collection, analysis, and interpretation of hospital data. Data scientists must adhere to ethical guidelines and principles to protect both themselves and the clients involved. Rule 8 part a of Data Science Ethics puts an emphasis on transparency. Data scientists are obliged to inform the client about all data science results and material facts, regardless of whether they are good or bad. This ensures that clients are equipped with the necessary information to make informed decisions based on the findings of the data analysis.

The quality of data is a crucial ethical consideration in data science. Rule 8 part b emphasizes that data scientists should assess the data, rate it, and disclose that information to the customer. This practice is essential for maintaining the integrity of the analysis and ensuring that the client understands the reliability and usefulness of the data used. Rule 8 part j discourages data scientists from engaging in cherry-picking, which is the selective use of data to support a particular conclusion while ignoring contradictory data. This barrier reinforces the commitment to objectivity and unbiased analysis, promoting ethical practices in data analysis. By following ethical standards, data scientists ensure transparency and reliability in their work, fostering trust and credibility.

CONCLUSIONS

In this project, we worked with hospital data to understand how hospital safety impacts patients' perception of care quality and whether this perception varies based on the size of the hospital. After running summary measures, visualizations, and analytic models, it was concluded that there was a relationship between perception of care quality and hospital safety. Specifically, the cms risk score was moderately correlated to the hospital stars. It was then found that the disease level decreased with the higher care quality. To give Senator Lankford an answer on what to do, he should focus on lowering the risk scores of hospitals to elevate the hospital star ratings.

The second part of Senator Lankford's question dealt with whether the perception of care quality varied by hospital size. After running a density plot, it was discovered that there was a pattern in the hospital size. The medium size hospitals had a higher quality perception and the small hospitals had a lower quality perception. This observation may be a trend for the west, but it is not the same throughout the nation. After comparing with other members of the group, it was discovered that their results showed the opposite, where medium hospitals had the low quality perception. So to give Senator Lankford an answer for his second question, there are trends in different sections of the nation, but not one nationwide trend.

REFERENCES

Here is useful code to get R, RStudio, and R packages citations.
Remember to also cite your data source using the html and any other citations required by that source.

```
Version R version 4.3.1 (2023-06-16 ucrt)

[1] "Xie, Y. (2022). knitr: A General-Purpose Package for Dynamic Report Generation in R. R
package version 1.43. URL: https://cran.r-project.org/package=knitr"

[1] "Mcnamara, A., Arino De La Rubia, E., & Quinn, M. (2022). skimr: Compact and flexible
summaries of data. R package version 2.1.5. URL: https://cran.r-project.org/package=skimr"

[1] "Revelle, W. (2022). psych: Procedures for Psychological, Psychometric, and Personality
Research. R package version 2.3.6. URL: https://cran.r-project.org/package=psych"

[1] "Wickham, H., & Bryan, J. (2022). readxl: Read Excel Files. R package version 1.4.3. URL:
https://cran.r-project.org/package=readxl"

[1] "Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of
Statistical Software, 40(3), 1-25. R package version 1.9.2. URL: https://cran.r-
project.org/package=lubridate"

[1] "Wickham, H. (2022). forcats: Tools for Working with Categorical Variables (Factors). R
package version 1.0.0. URL: https://cran.r-project.org/package=forcats"

[1] "Wickham, H. (2022). stringr: Simple, Consistent Wrappers for Common String Operations. R
package version 1.5.0. URL: https://cran.r-project.org/package=stringr"

[1] "Henry, L., & Wickham, H. (2022). purrr: Functional Programming Tools. R package version
1.0.2. URL: https://cran.r-project.org/package=purrr"

[1] "Wickham, H., Hester, J., & Francois, R. (2022). readr: Read Rectangular Text Data. R package
version 2.1.4. URL: https://cran.r-project.org/package=readr"

[1] "Müller, K., & Wickham, H. (2022). tibble: Simple Data Frames. R package version 3.2.1. URL:
https://cran.r-project.org/package=tibble"

[1] "Wickham, H., Averick, M., Bryan, J., et al. (2022). tidyverse: Easily Install and Load the
'Tidyverse'. R package version 2.0.0. URL: https://cran.r-project.org/package=tidyverse"

[1] "Schloerke, B., Crowley, J., Cook, D., et al. (2022). GGally: Extension to 'ggplot2'. R
package version 2.1.2. URL: https://cran.r-project.org/package=GGally"

[1] "Maechler, M., Rousseeuw, P., Struyf, A., et al. (2022). cluster: Cluster Analysis Basics and
Extensions. R package version 2.1.4. URL: https://cran.r-project.org/package=cluster"

[1] "Kassambara, A., & Mundt, F. (2022). factoextra: Extract and Visualize the Results of
Multivariate Data Analyses. R package version 1.0.7. URL: https://cran.r-
project.org/package=factoextra"

[1] "Persson, A. (2022). DataExplorer: Data Exploration and Visualization. R package version
0.8.2. URL: https://cran.r-project.org/package=DataExplorer"
```

```
[1] "Wickham, H., & Henry, L. (2022). tidyr: Tidy Messy Data. R package version 1.3.0. URL:
https://cran.r-project.org/package=tidyr"
```

```
[1] "Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. R
package version 3.4.3. URL: https://cran.r-project.org/package=ggplot2"
```

```
[1] "Wickham, H., Francois, R., Henry, L., & Müller, K. (2022). dplyr: A Grammar of Data
Manipulation. R package version 1.1.3. URL: https://cran.r-project.org/package=dplyr"
```

```
[1] "Komsta, L., & Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurtosis and
related tests. R package version 0.14.1. URL: https://cran.r-project.org/package=moments"
```

```
[1] "Centers for Medicare and Medicaid Services. (2023). Dataset Title. Retrieved on September 3,
2023, from https://data.cms.gov/provider-data/dataset/77hc-ibv8"
```

```
[1] "Centers for Medicare and Medicaid Services. (2023). Medicare Inpatient Hospitals Provider
Summary by Type of Service. Retrieved on September 3, 2023, from https://data.cms.gov/provider-
summary-by-type-of-service/medicare-inpatient-hospitals/medicare-inpatient-hospitals-by-
provider/data"
```

```
[1] "Centers for Medicare and Medicaid Services. (2023). HCAHPS Dataset. Retrieved on September
3, 2023, from https://data.cms.gov/provider-data/dataset/dgck-syfz"
```

```
[1] "Data Science Association. (2023). Code of Conduct for Data Science. Retrieved from
https://datascienceassn.org/code-of-conduct.html"
```


END