CMS RATINGS FOR HOSPITALS

CAPTSONE PROJECT

CMS Ratings for Hospitals –Business Understanding

CMS rates hospitals in the US on a scale of 1-5 with the objective to make it easier for patients and consumers to compare the quality of hospitals. Data to be analyzed is collected from all the hospitals , various surveys conducted during different time periods. The ratings directly influence the choice of the hospital made by consumers and may have a significant impact on the revenue earned by hospitals. The purpose of this project is to develop an approach to calculate Hospital Ratings as done by CMS.

CMS Ratings for Hospitals – Data Requirement

CMS collected information for various measures through different quality reporting programs and surveys conducted in hospitals. All the measures were grouped based on the quality measure they define. The measures in the data are broadly classified into below Measure Groups.

- Mortality
- Safety of Care
- Readmission
- Patient Experience
- Effectiveness of Care
- Timeliness of Care
- Efficient Use of Medical Imaging
- Models is build for each of the Measure Groups and a weighted approach is used for star rating calculations from these measure group scores.

CMS Ratings for Hospitals – Initial Observation

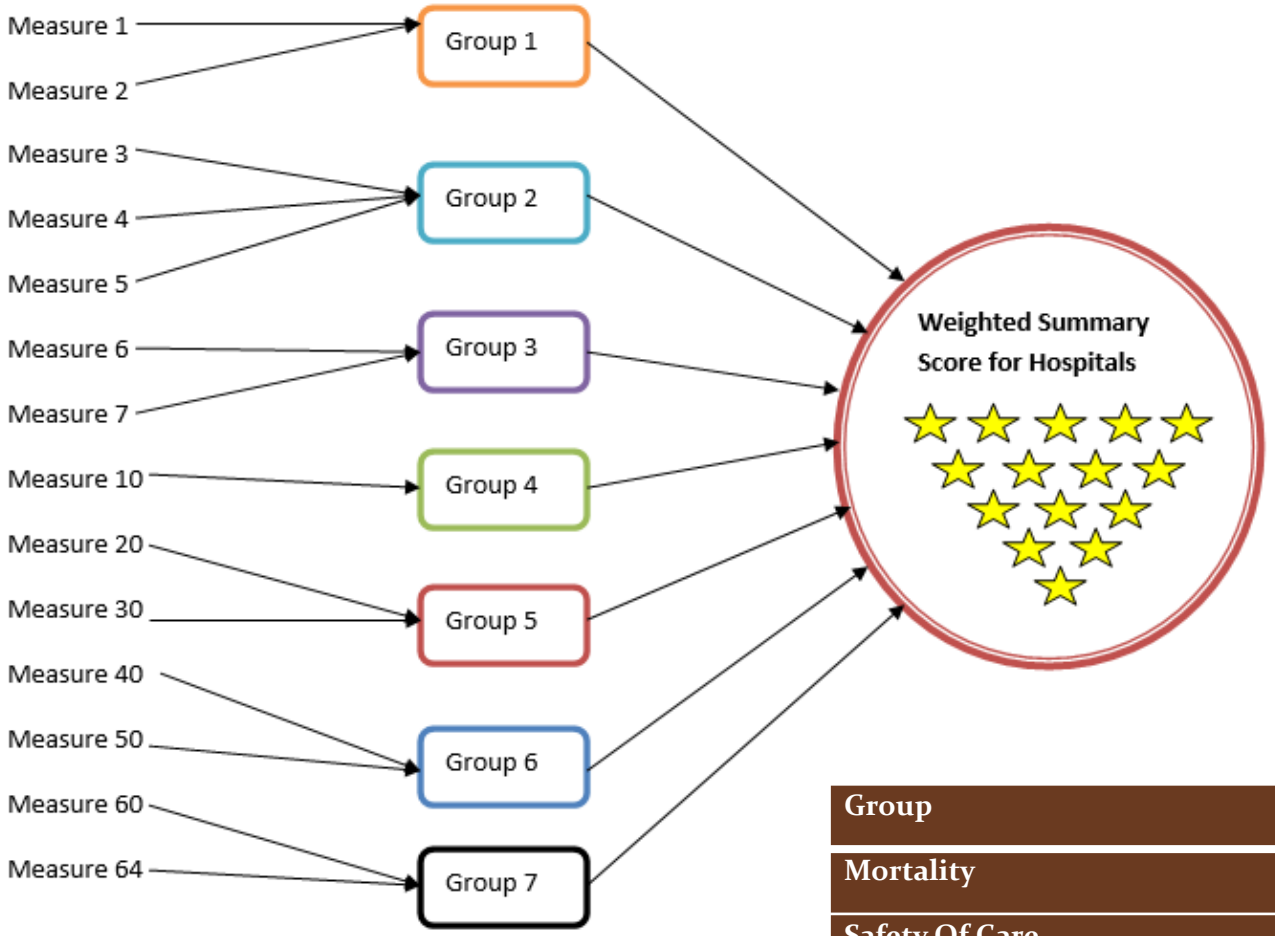
Total of about 58 files are available in the dataset that capture data for about 64 measures. All the datasheets with HOSPITAL in the name capture the measure values for each provider. There are datasheets pertaining to NATIONAL and STATE average for the measures in each measure group , that will be used for comparison and provide the recommendations. There are other sheets like MeasureDates and Footnotes that are the master tables for Measure collection periods and Footnote descriptions.

As part of EDA , following steps were done on each of the data sheets

- Check for NA and handle it.
- Classify the variables as Numerical and Categorical
- Standardize the values .
- Handle missing values if any
- Scale all numerical values to a common scale
- Transform the MeasureID – MeasureValue rows to dataset with MeasureId as columnNames and Measure Value as value.
- Merge all data pertaining to single Measure group into a common dataset
- This will give about 7 data sets, each pertaining to a Measure Group.
- EDA will be conducted on each group of measures.
- Correlation between measures will be checked and those measures will be captured which have the highest influence to calculate the Group score.
- Weights for each category is already been specified by CMS as below table. Weighted average method will be used to calculating the start rating.

Once EDA has been performed. For second analysis, we will be also performing Exploratory Factor Analysis (EFA). As part of EFA , following steps were done on each of the data sheets

- Check for NA and handle it.
- Classify the variables as Numerical and Categorical
- Standardize the values .
- Handle missing values if any
- Scale all numerical values to a common scale
- Transform the MeasureID – MeasureValue rows to dataset with MeasureId as columnNames and Measure Value as value.
- Merge all data pertaining to single Measure group into a common dataset
- This will give about 7 data sets, each pertaining to a Measure Group.
- EFA will be conducted on each group of measures.
- Correlation between measures will be checked and those measures will be captured which have the highest influence to calculate the Group score.
- Weights for each category is already been specified by CMS as below table. Weighted average method will be used to calculating the start rating.



Group	Group Star Rating Weight
Mortality	22%
Safety Of Care	22%
Readmission	22%
Patient Experience	22%
Effectiveness of Care	4%
Efficient Use of Medical Imaging	4%

CMS Ratings for Hospitals – Exploratory Data Analysis

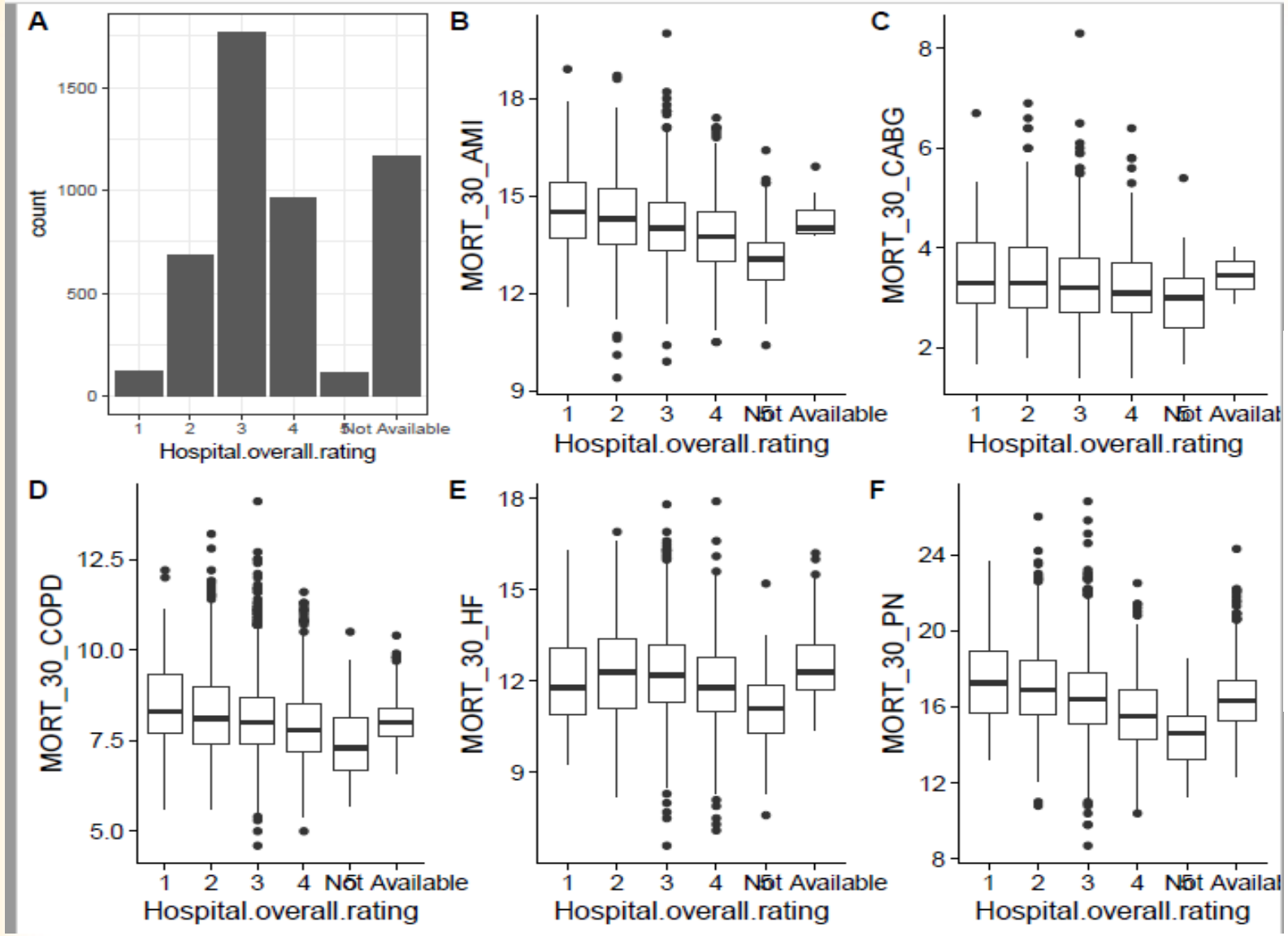
As part of the Supervised Learning, we took to the EDA Approach as mentioned previously. The data sets provided were took to analytical tests and we have obtained the results. In the forthcoming slides, we will be reflecting on the results obtained for each Group.

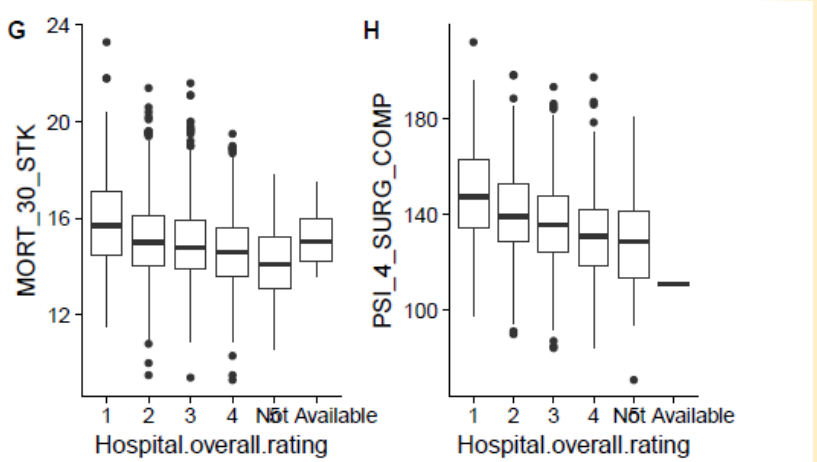
Group 1 – Mortality Measures– Relation with Target variable Over All Rating

For mortality measures we found the following inferences.

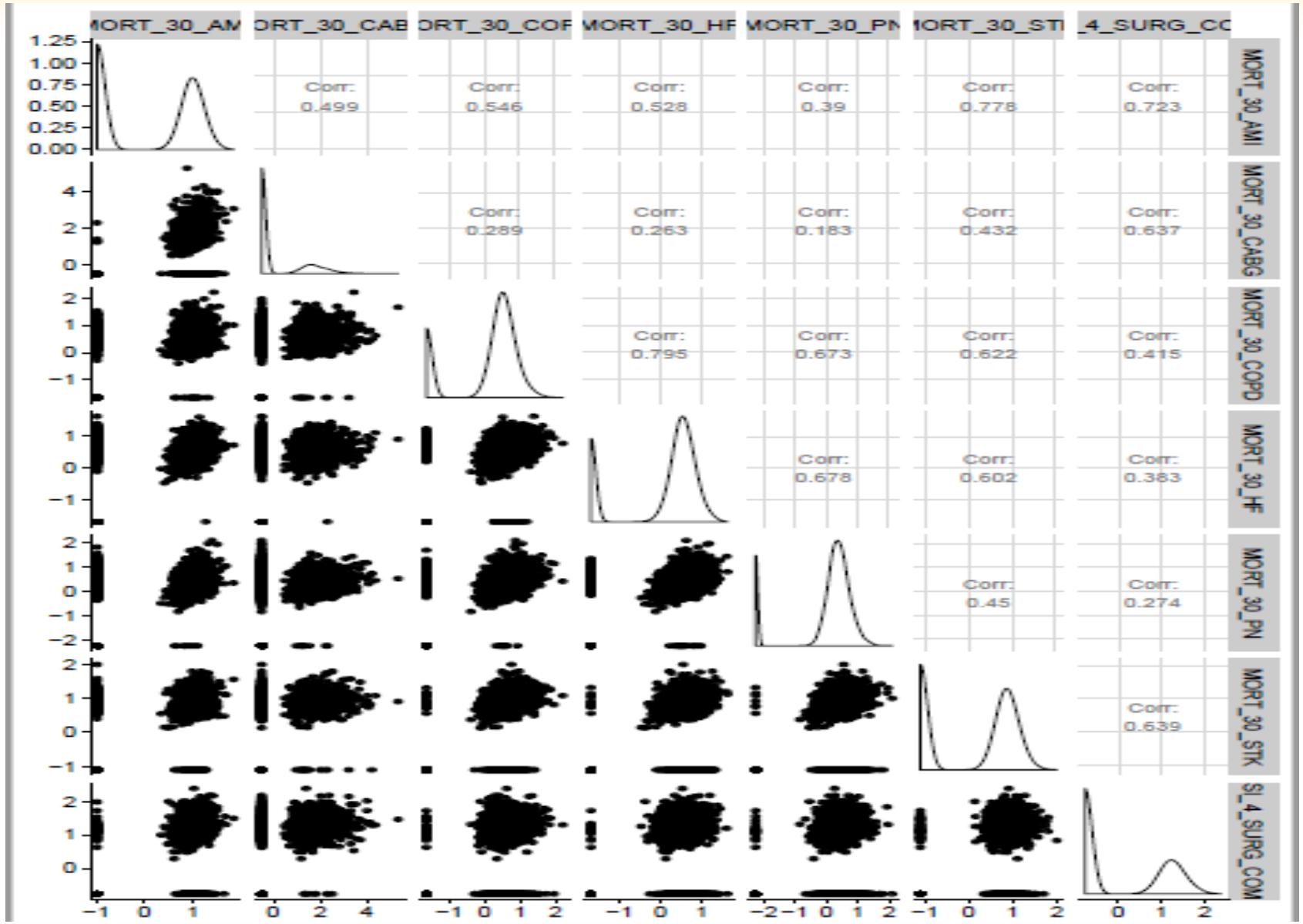
Inferences -

- Most of the hospitals have got 3 star rating
- Low rating Hospitals has have Death rate of heart attack patients
- Hospital rated 3 have a lot of outliers.
- High rated hospitals have low Death rate of pneumonia patients.
- Low rated hospitals have high Death rate of stroke patients.


A
count
1500
1000
500
0
1
2
3
4
5
Not Available
Hospital.overall.rating
B
MORT_30_AMI
18
15
12
9
Hospital.overall.rating
C
MORT_30_CABG
8
6
4
2
Hospital.overall.rating
D
MORT_30_COPD
12.5
10.0
7.5
5.0
Hospital.overall.rating
E
MORT_30_HF
18
15
12
9
Hospital.overall.rating
F
MORT_30_PN
24
20
16
12
8
Hospital.overall.rating


G
MORT_30_STK
24
20
16
12
Hospital.overall.rating
H
PSI_4_SURG_COMP
180
140
100
Hospital.overall.rating

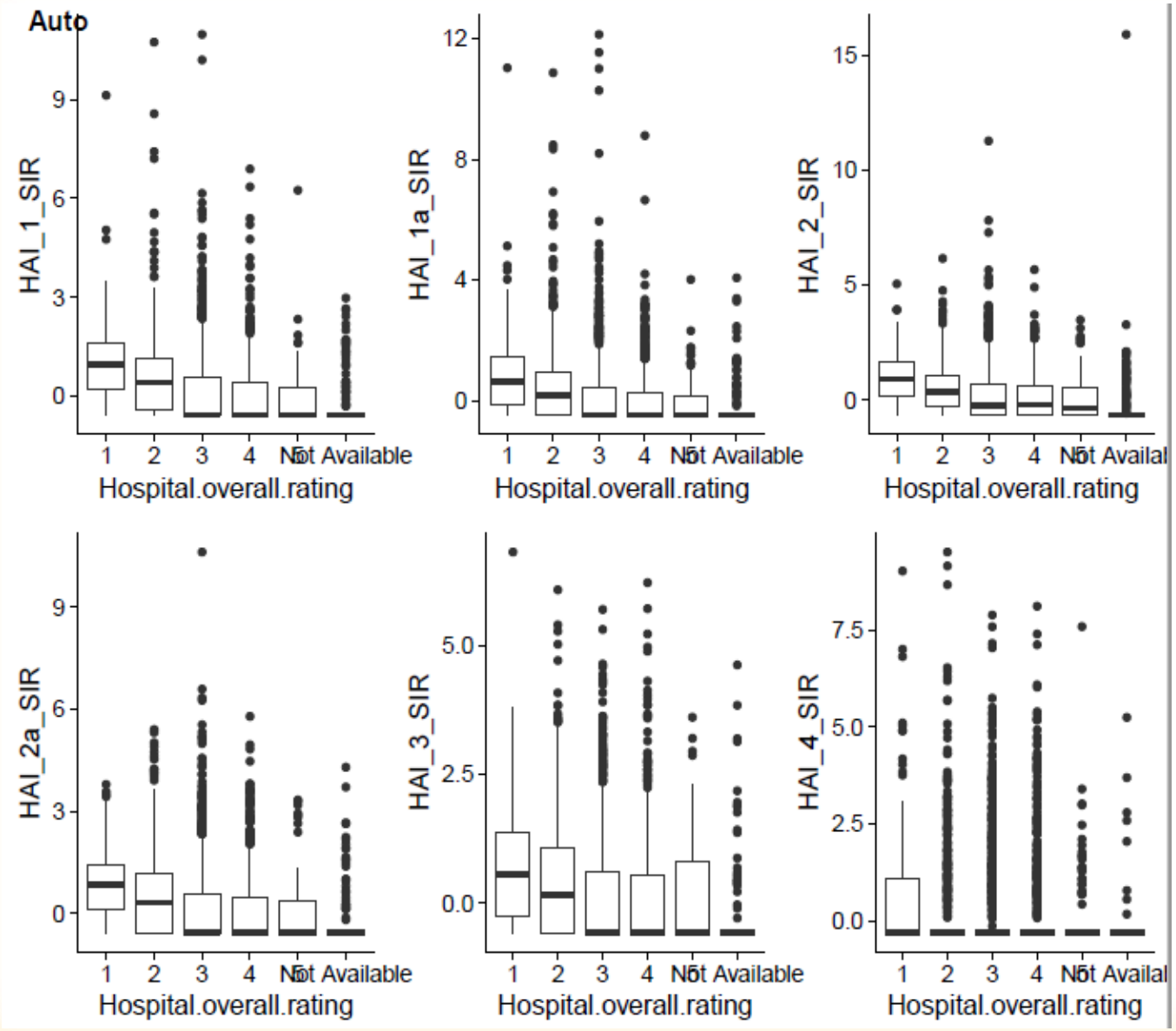
Correlation – Mortality Measures

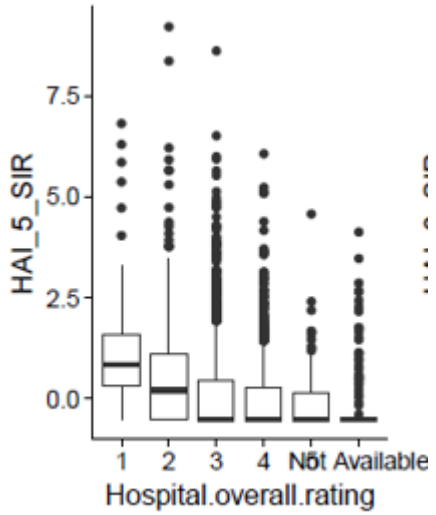


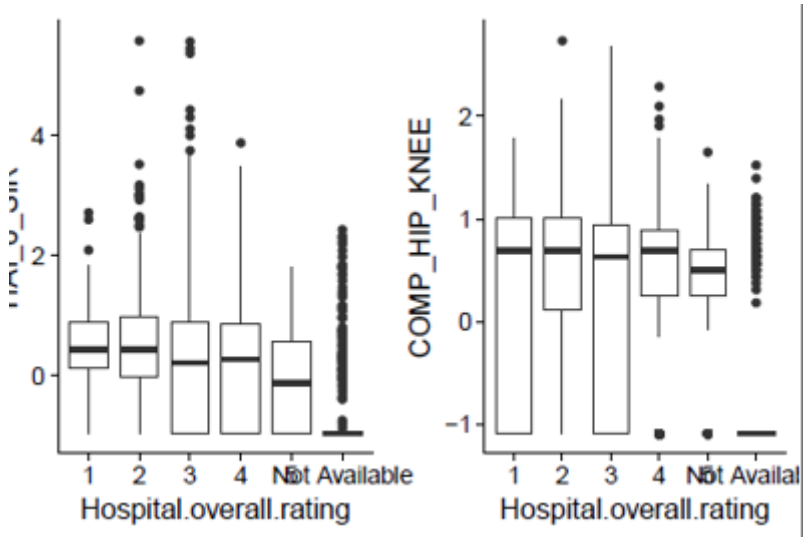
Group 2 – Safety of Care – Variance with Target Variable – Overall Hospital Rating

Inferences –

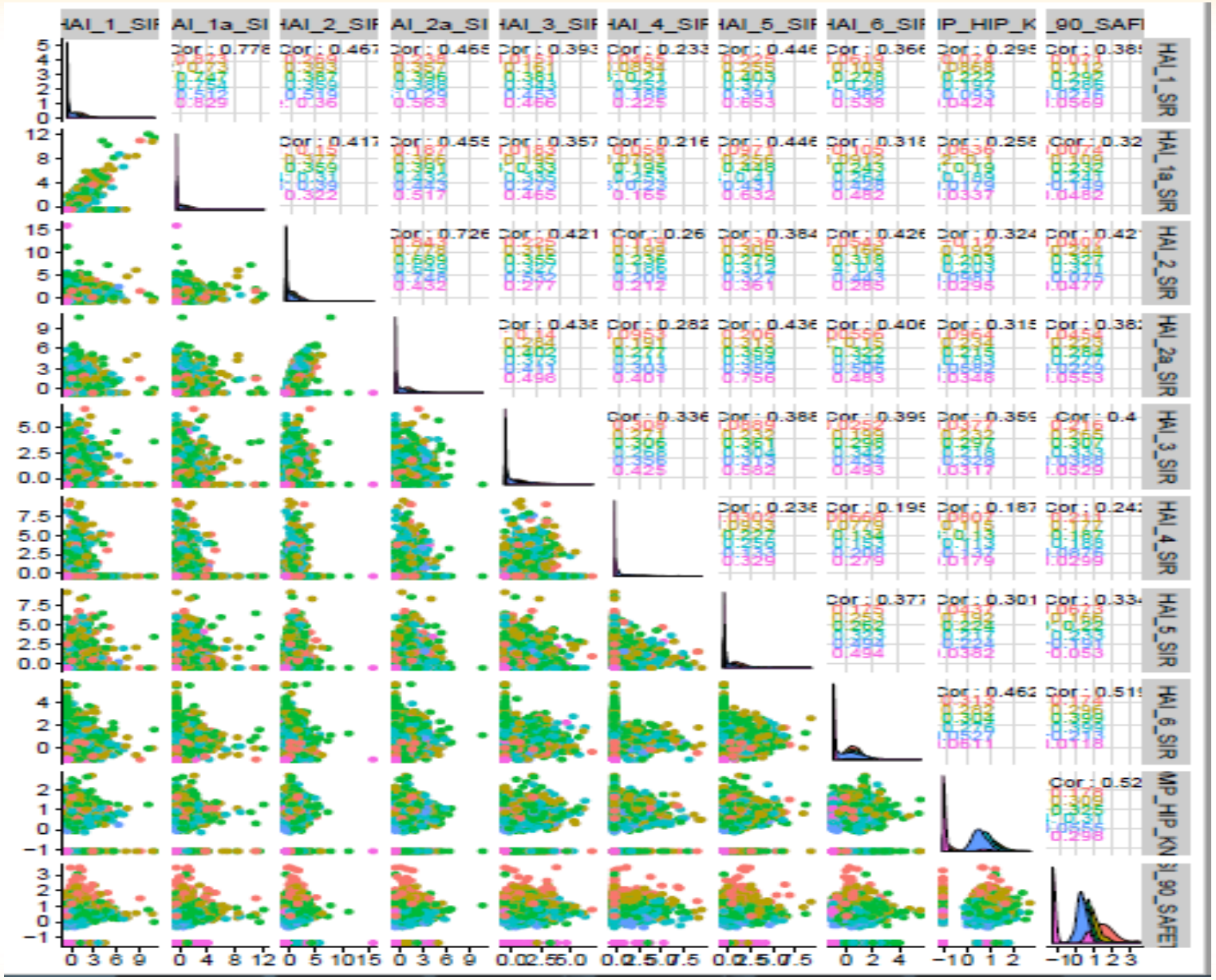
- There is a high Correlation for Last two Measures with target variable. This means that for the target customers that is the patients the measures in terms of safety has a very high correlation. This affliction of higher correlation is because the safer the patients feel under the care, the better they get and feel in terms of recovery from the ailment.
- Hospitals with Lower Rating have High Values for PSI_90_SAFETY and COMP_HIP_KNEE


Auto
HAI_1_SIR
HAI_1a_SIR
HAI_2_SIR
HAI_2a_SIR
HAI_3_SIR
HAI_4_SIR
1 2 3 4 5 Not Available
Hospital.overall.rating


HAI_5_SIR
1 2 3 4 5 Not Available
Hospital.overall.rating


COMP_HIP_KNEE
1 2 3 4 5 Not Available
Hospital.overall.rating

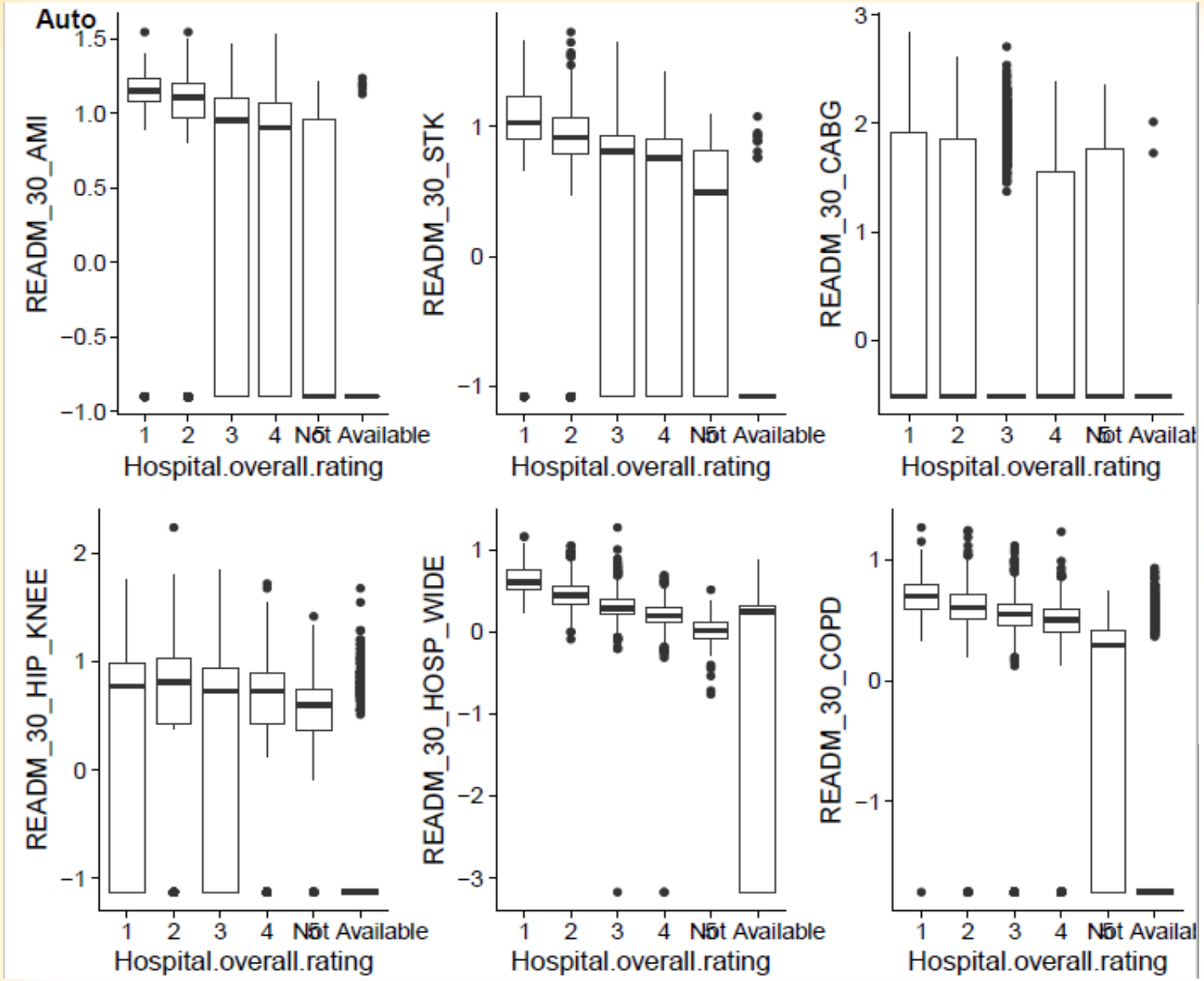
Correlation – Safety Care Measures

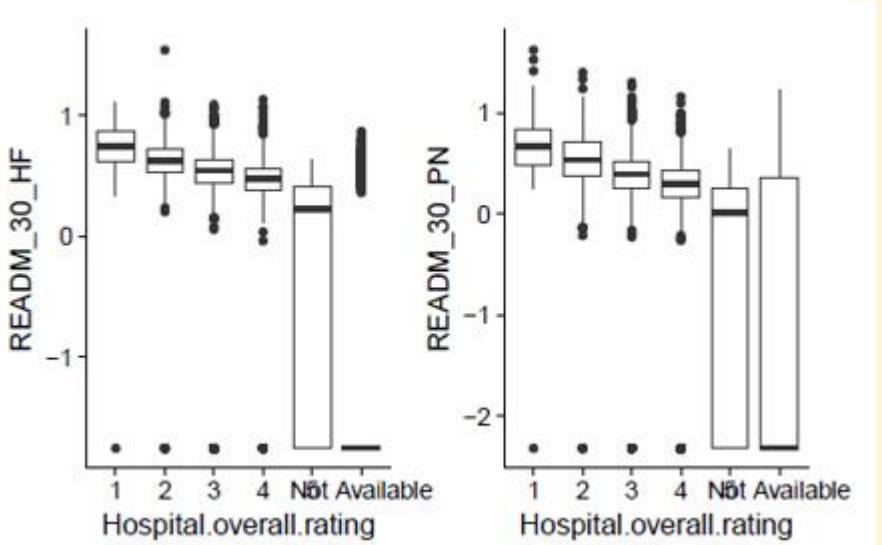


Measure Group 3 – Readmissions –Variance with Target Variable

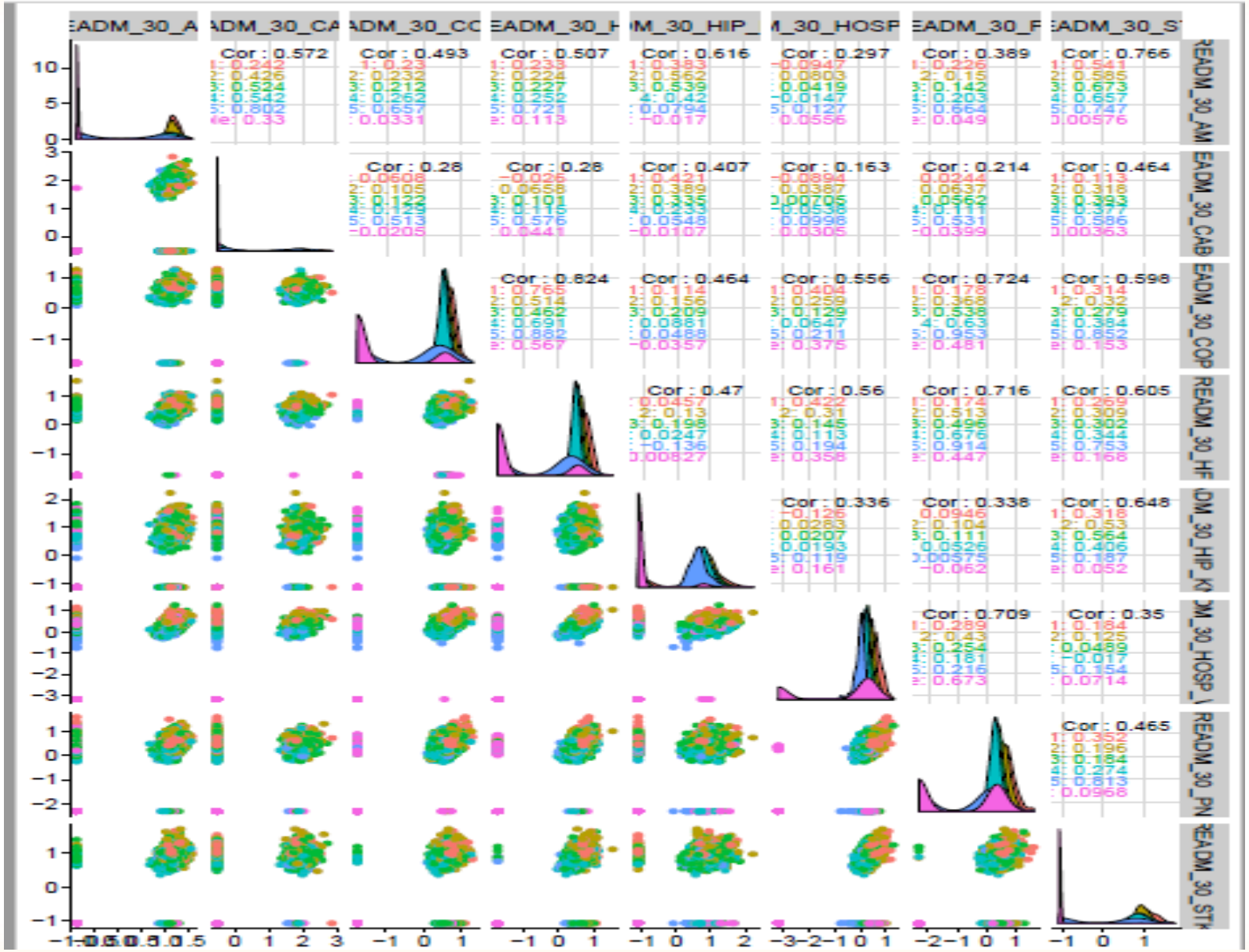
Inferences –

- Hospitals with lower rating have high readmission rate for heart attack & stroke patients
- Highest rated hospital have lower readmissions of COPD, heart failure & pneumonia patients


Auto
READM_30_AMI
READM_30_STK
READM_30_CABG
READM_30_HIP_KNEE
READM_30_HOSP_WIDE
READM_30_COPD
1 2 3 4 Not Available
Hospital.overall.rating


READM_30_HF
READM_30_PN
1 2 3 4 Not Available
Hospital.overall.rating

Correlation – Readmissions

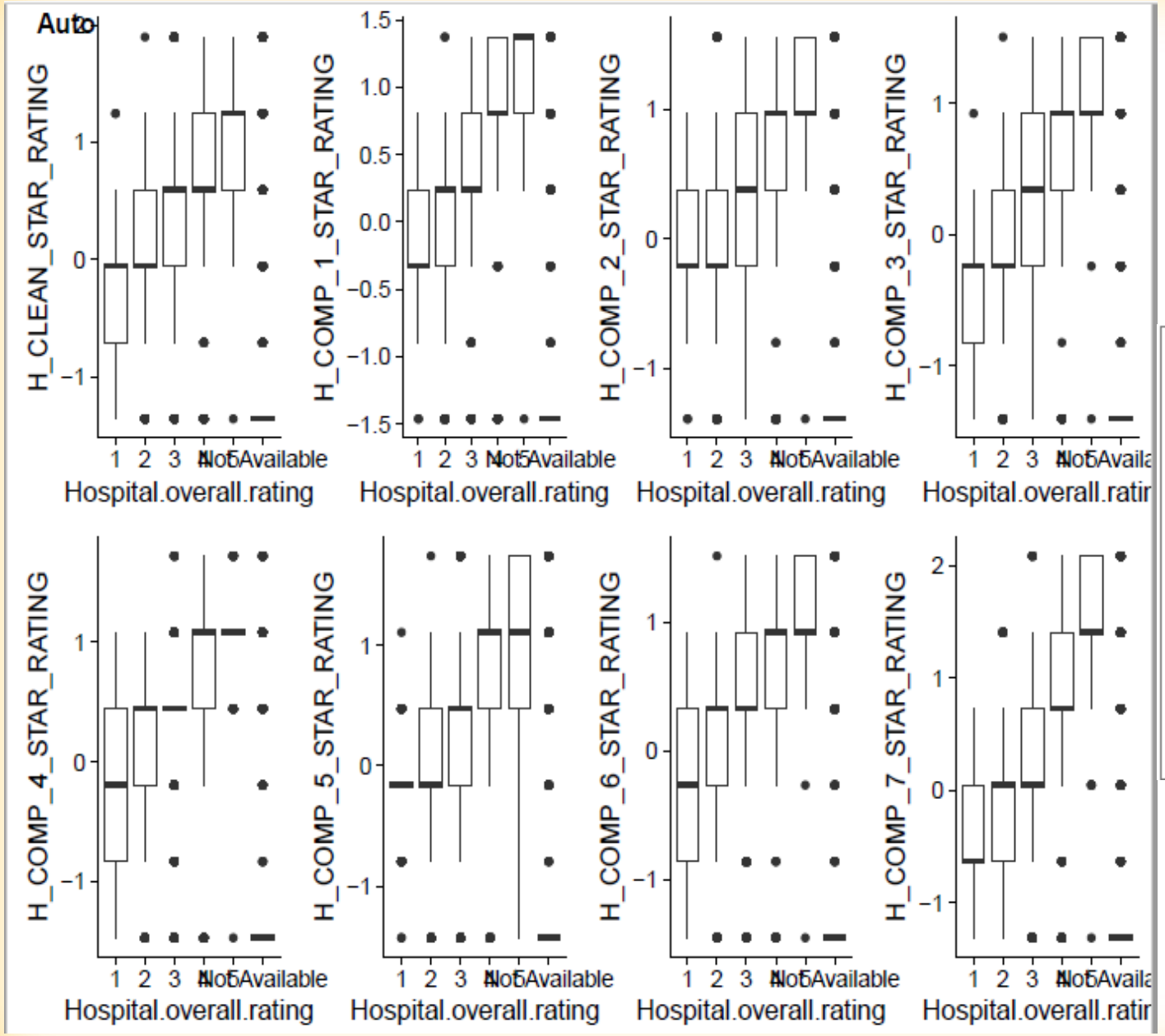


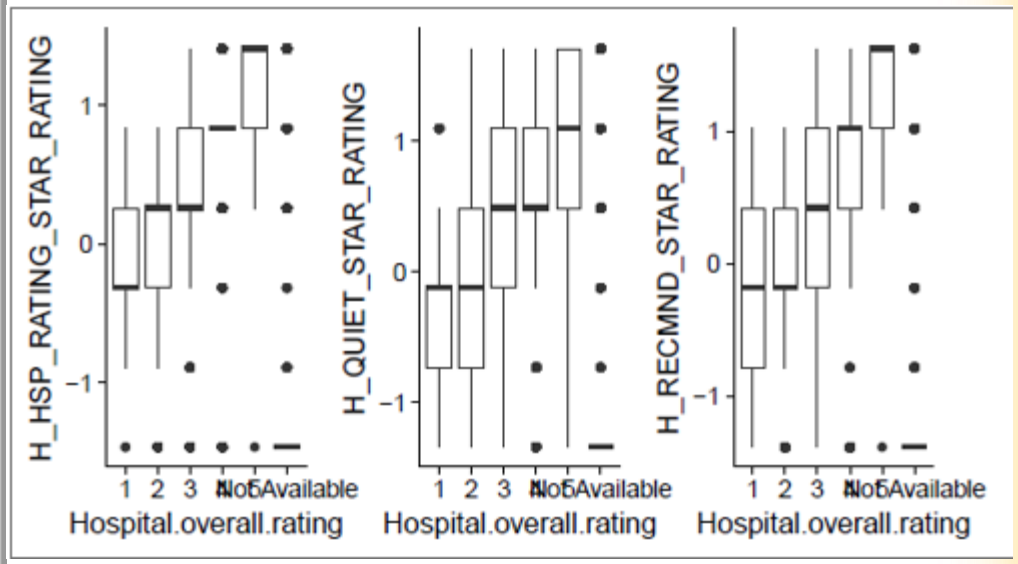
Measure Group 4 – Patient Experience – Variance with Target variable

Inferences –

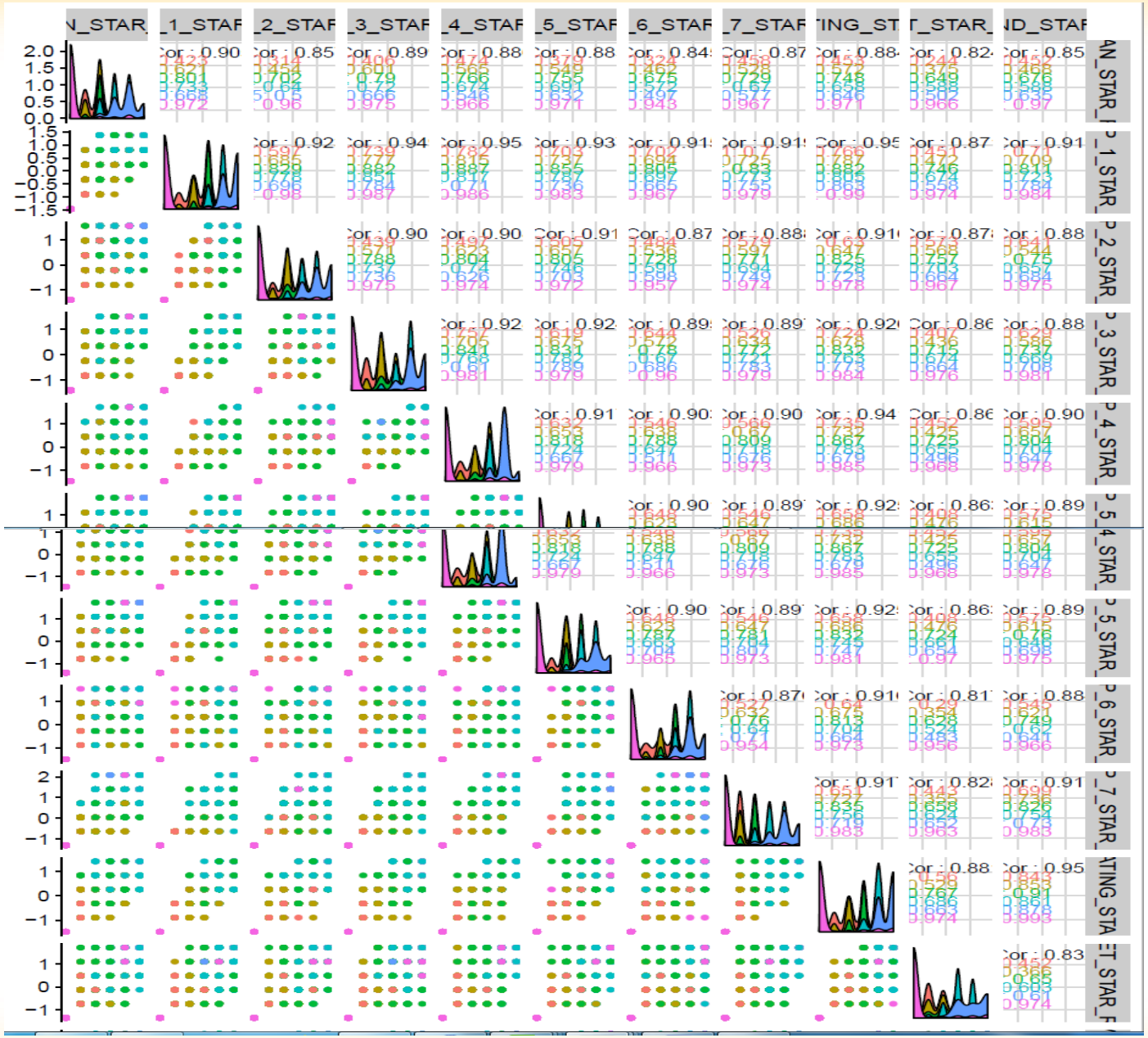
High Rated hospitals provided better Patient experience –

- They had better communication with their patients
- Good Cleanliness for bathrooms
- Patient Understood the care better
- Patients in emergency were treated immediately to ease pain


H_CLEAN_STAR_RATING
H_COMP_1_STAR_RATING
H_COMP_2_STAR_RATING
H_COMP_3_STAR_RATING
H_COMP_4_STAR_RATING
H_COMP_5_STAR_RATING
H_COMP_6_STAR_RATING
H_COMP_7_STAR_RATING
1 2 3 Not Available
Hospital.overall.rating


H_HSP_RATING_STAR_RATING
H_QUIET_STAR_RATING
H_RECMND_STAR_RATING
1 2 3 Not Available
Hospital.overall.rating

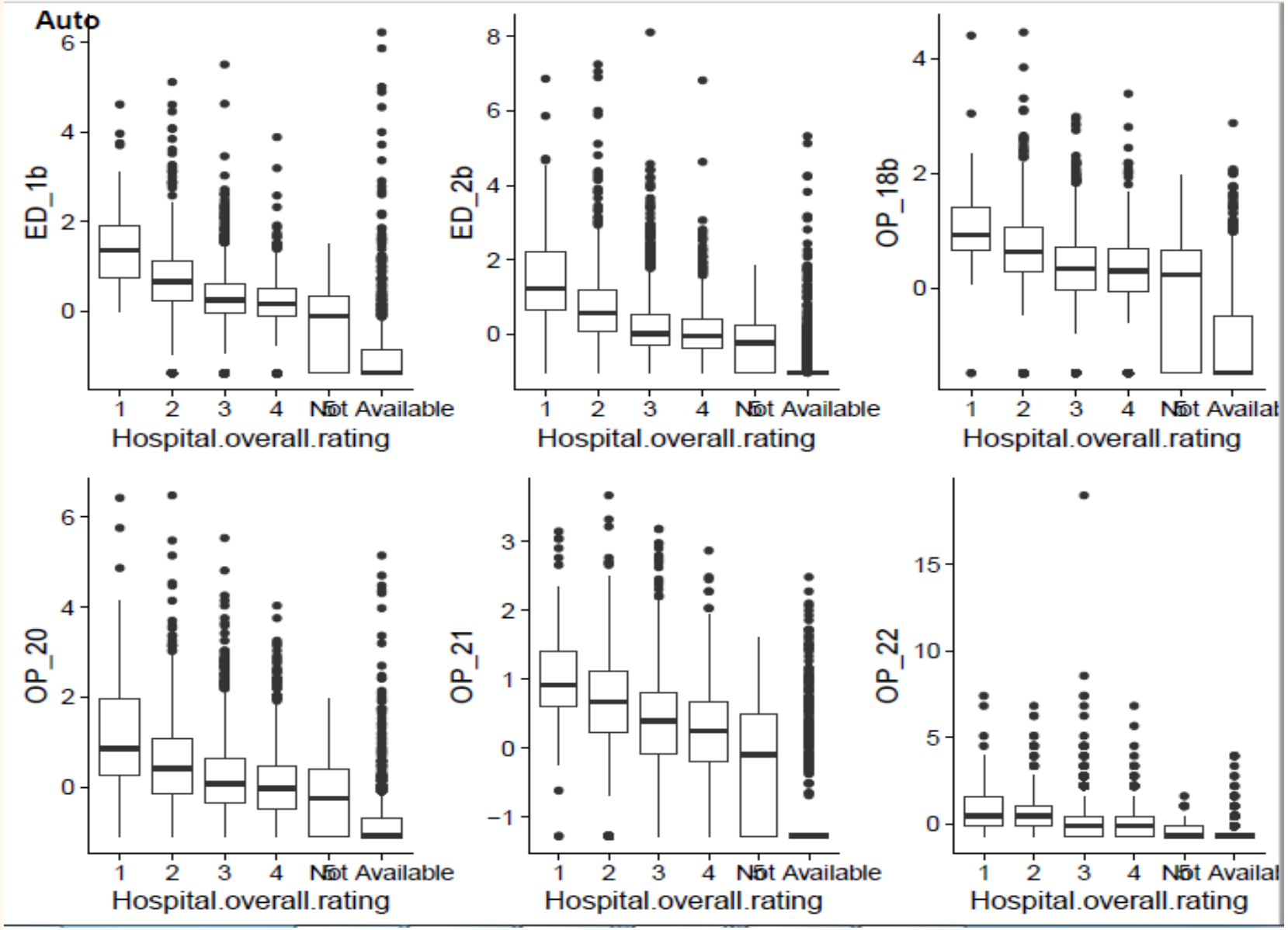
Correlation – Patient Experience

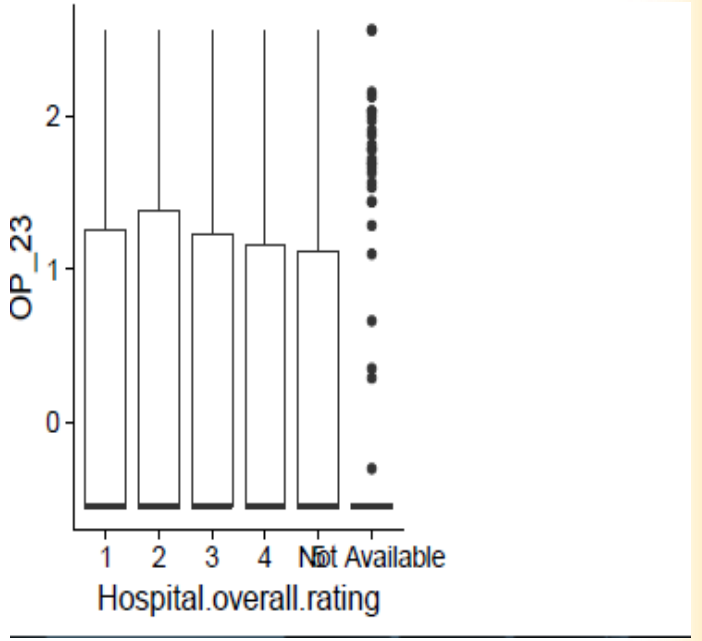


Measure Group 5 –Timeliness of Care – Variance

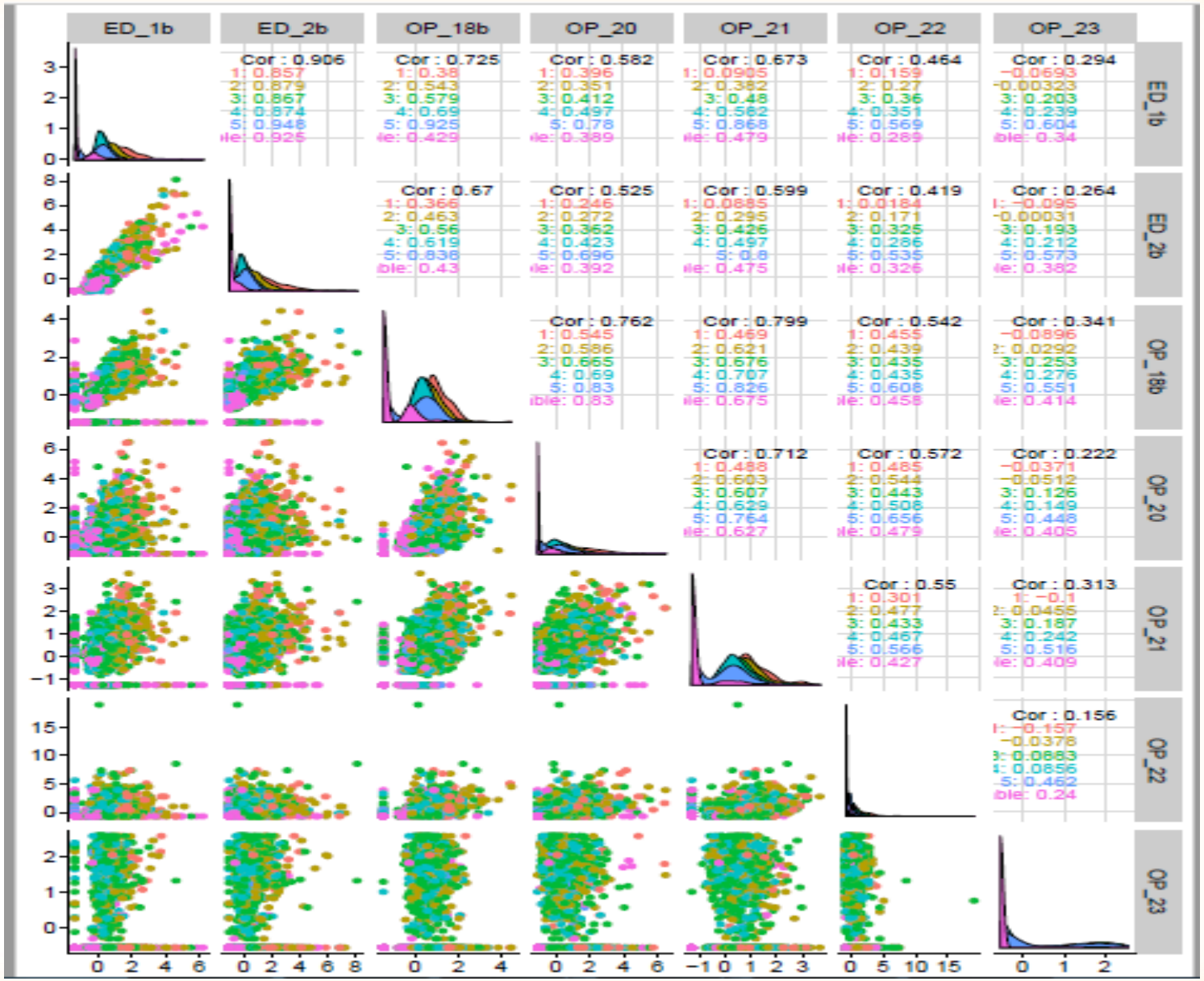
Inferences –

- 1. Low rated hospitals had higher wait time compared to High rated factors
- 2. Wait time was calculated for Emergency care , availability of doctor
- 3. Wait time for Heart patients before CPR
- 4. Wait time for people with bone injuries


Auto
ED_1b
ED_2b
OP_18b
OP_20
OP_21
OP_22
1 2 3 4 Not Available
Hospital.overall.rating


OP_23
1 2 3 4 Not Available
Hospital.overall.rating

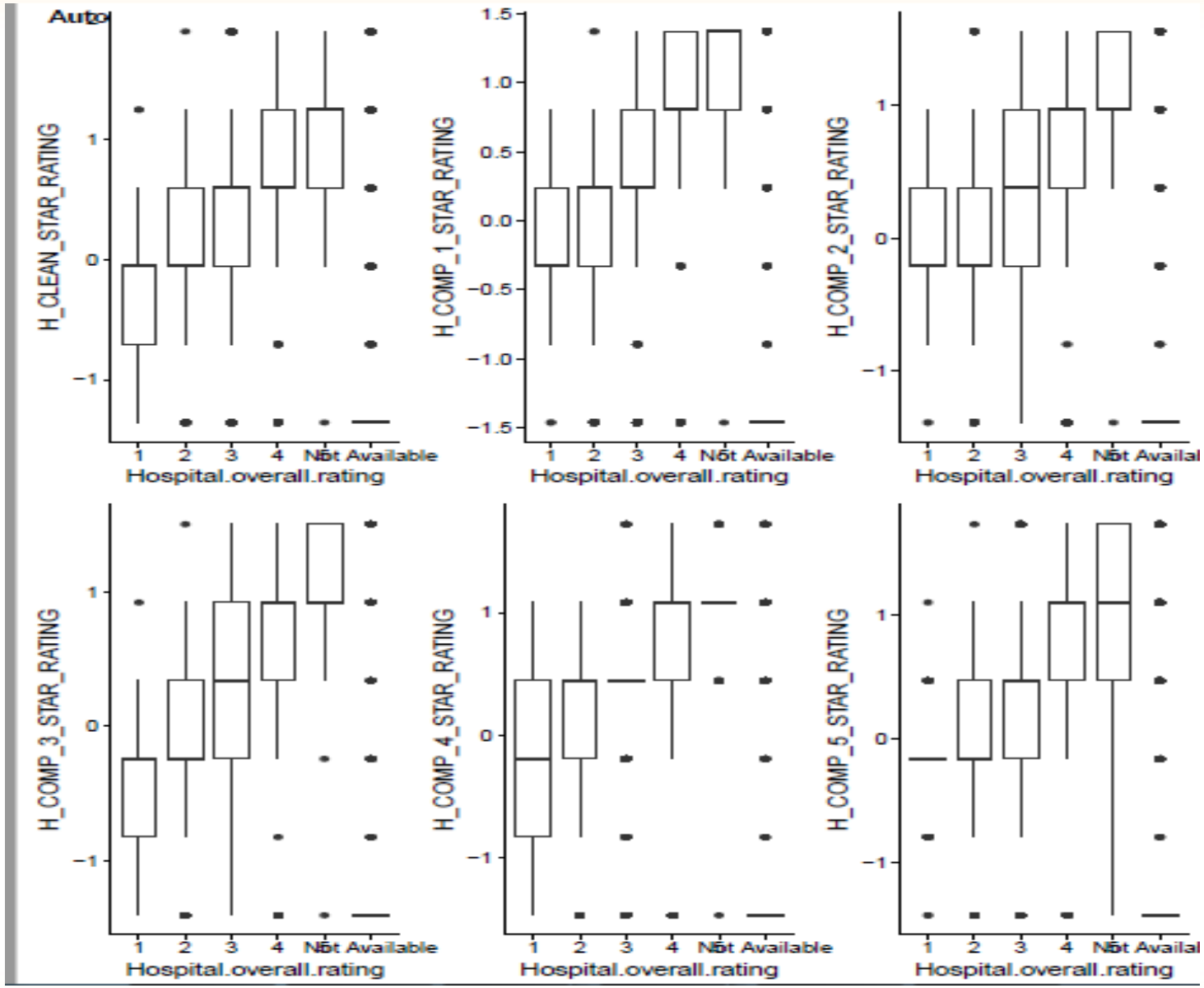
Correlation – Timeliness of Care



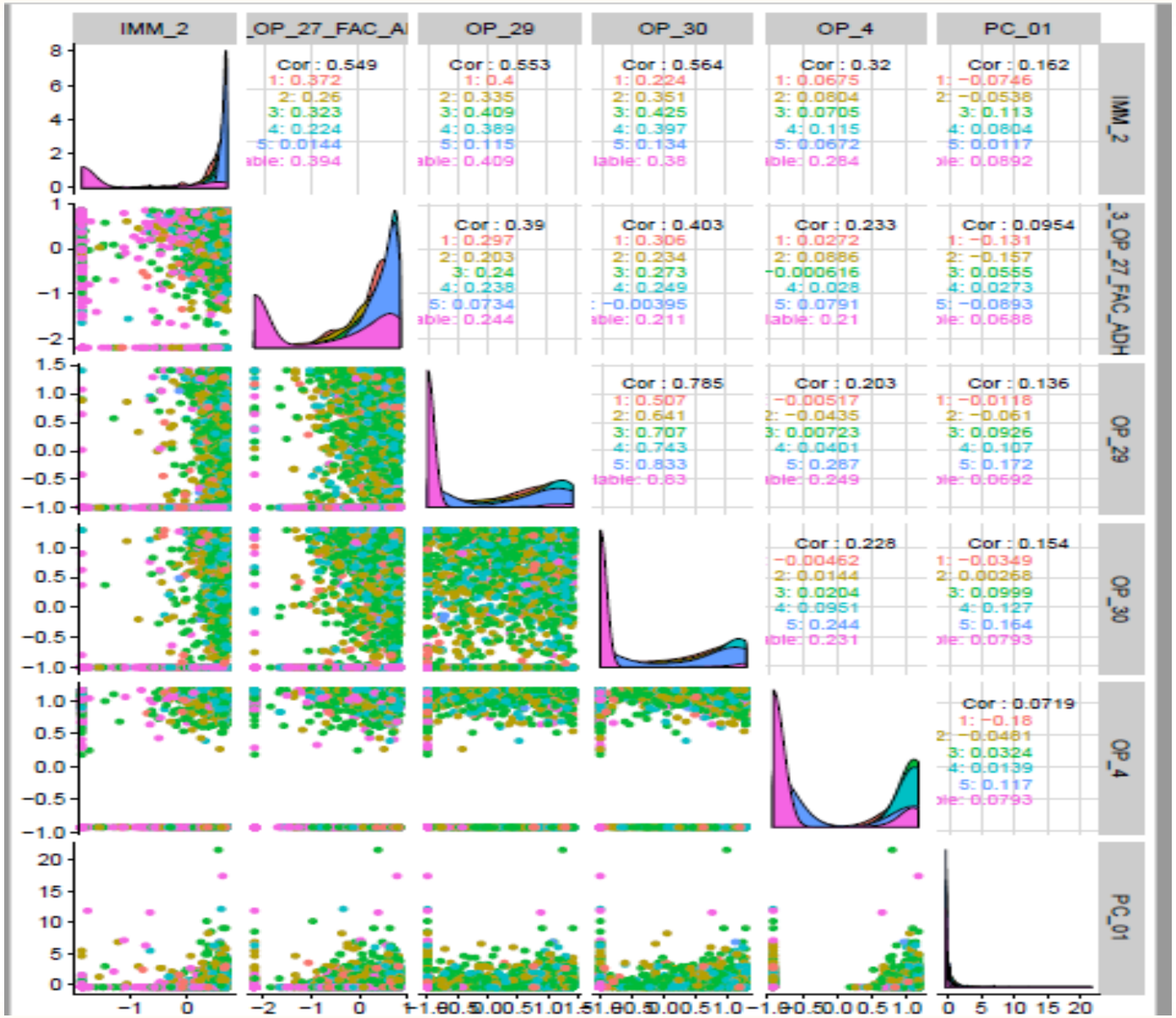
Measure Group 6 – Effectiveness of Care – Variance

Inferences –

- High rated hospitals have better rating for effectiveness of care for measures like Asprin given to patients within 24 hours
- Emergency care
- Getting reports early for radiology and colonoscopy
- Low rated hospitals have higher values for measures like Patients who developed blood clot


H_CLEAN_STAR_RATING
H_COMP_1_STAR_RATING
H_COMP_2_STAR_RATING
H_COMP_3_STAR_RATING
H_COMP_4_STAR_RATING
H_COMP_5_STAR_RATING
1
2
3
4
Not Available
Hospital.overall.rating

Correlation – Effectiveness of Care

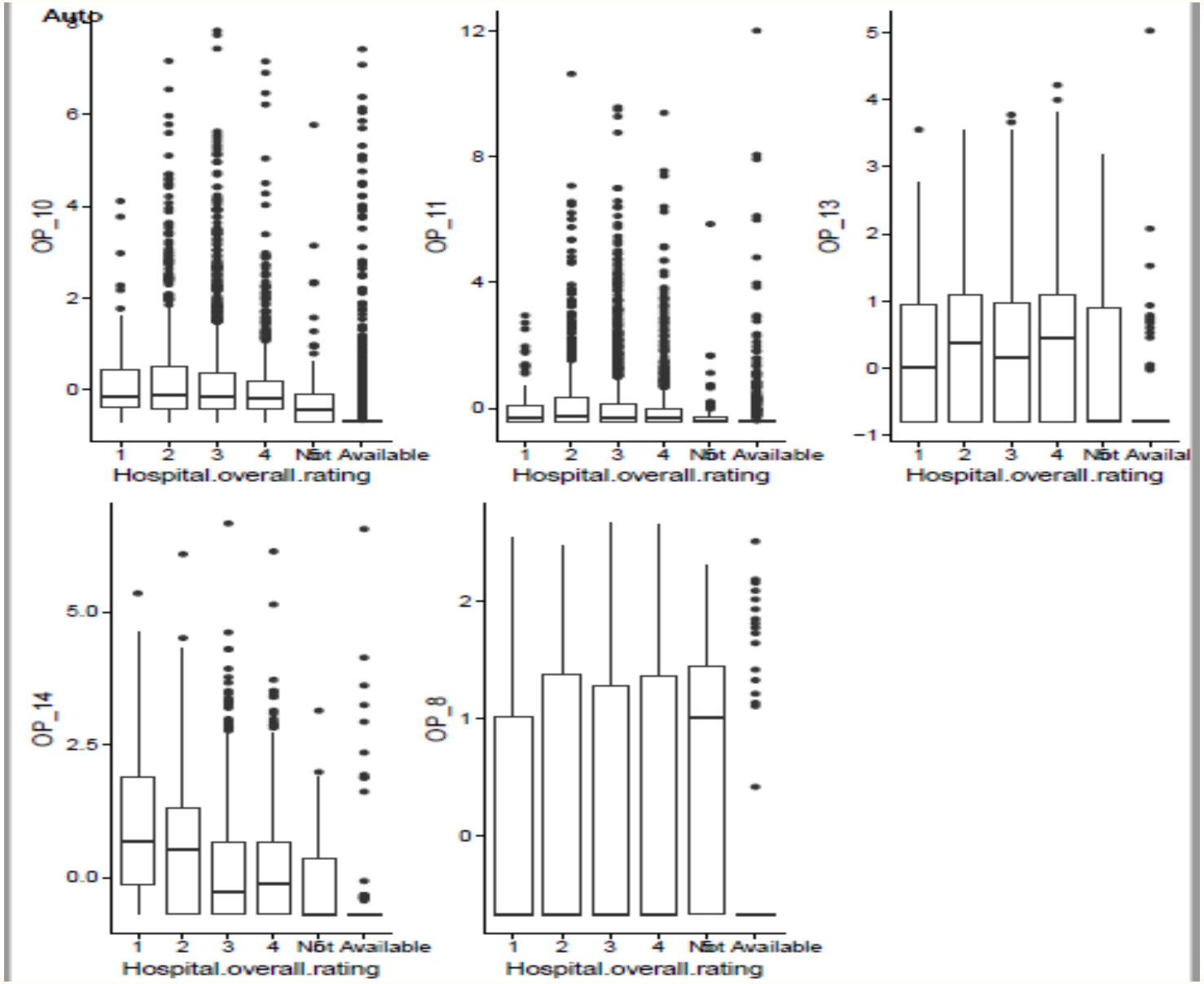


Measure Group 7 – Medical Imaging – Variance

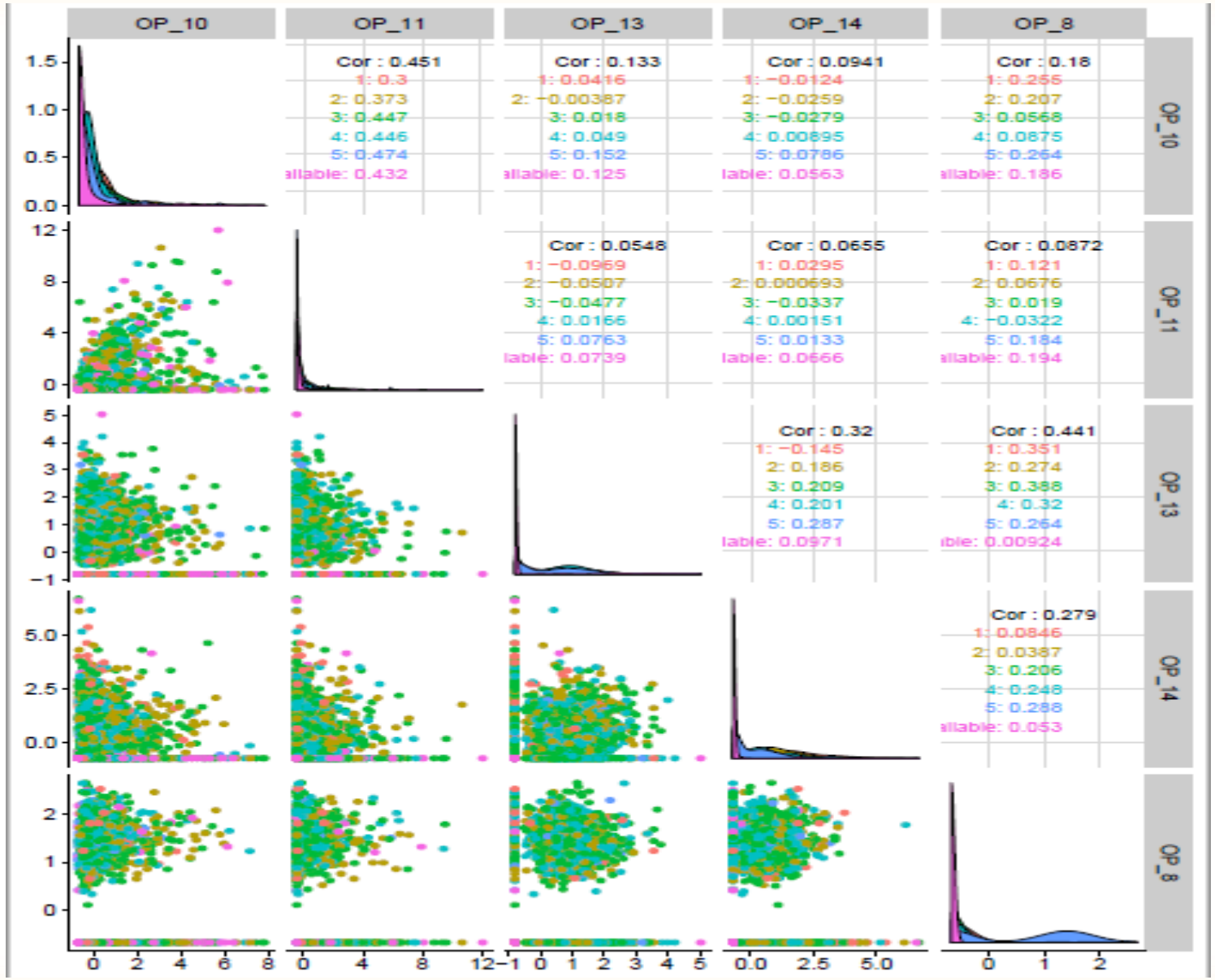
Medical Imaging is the last Measure Group used to perform EDA on.

Inferences –

- Low rated hospitals have higher values of Double scans , Scans recommended for low risk patients
- High rated hospitals have higher values for all the facilities for Medical imaging readily available


Auto
OP_10
OP_11
OP_13
OP_14
OP_8
Hospital.overall.rating
1
2
3
4
Not Available

Correlation – Medical Imaging



CMS Ratings for Hospitals – Random Factor Analysis

As part of the Supervised Learning, we took to the EDA Approach as mentioned previously. Once the EDA is performed, we further applied Random Factor Analysis approach to obtain the Star Rating for the every hospital.

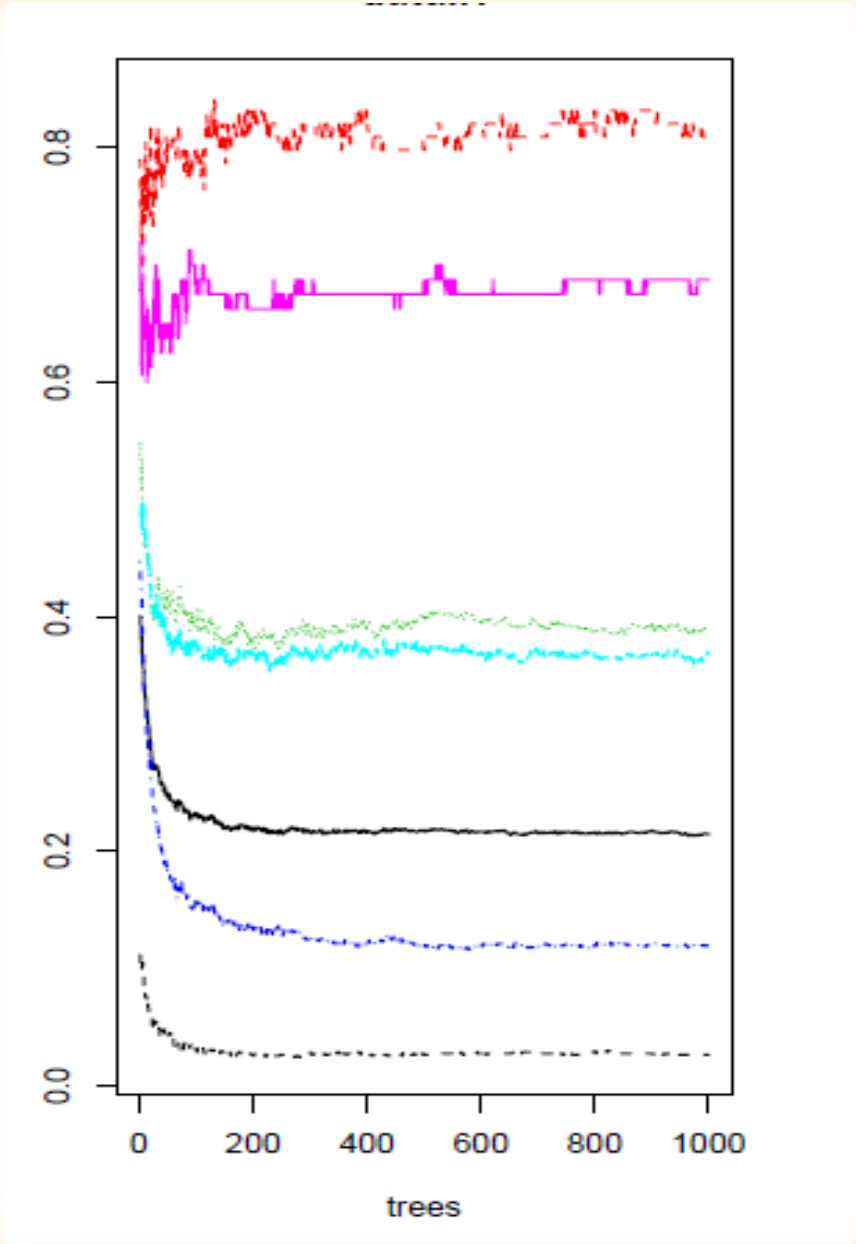
This has been obtained by combining all the 7 models of every hospital to provide one rating.

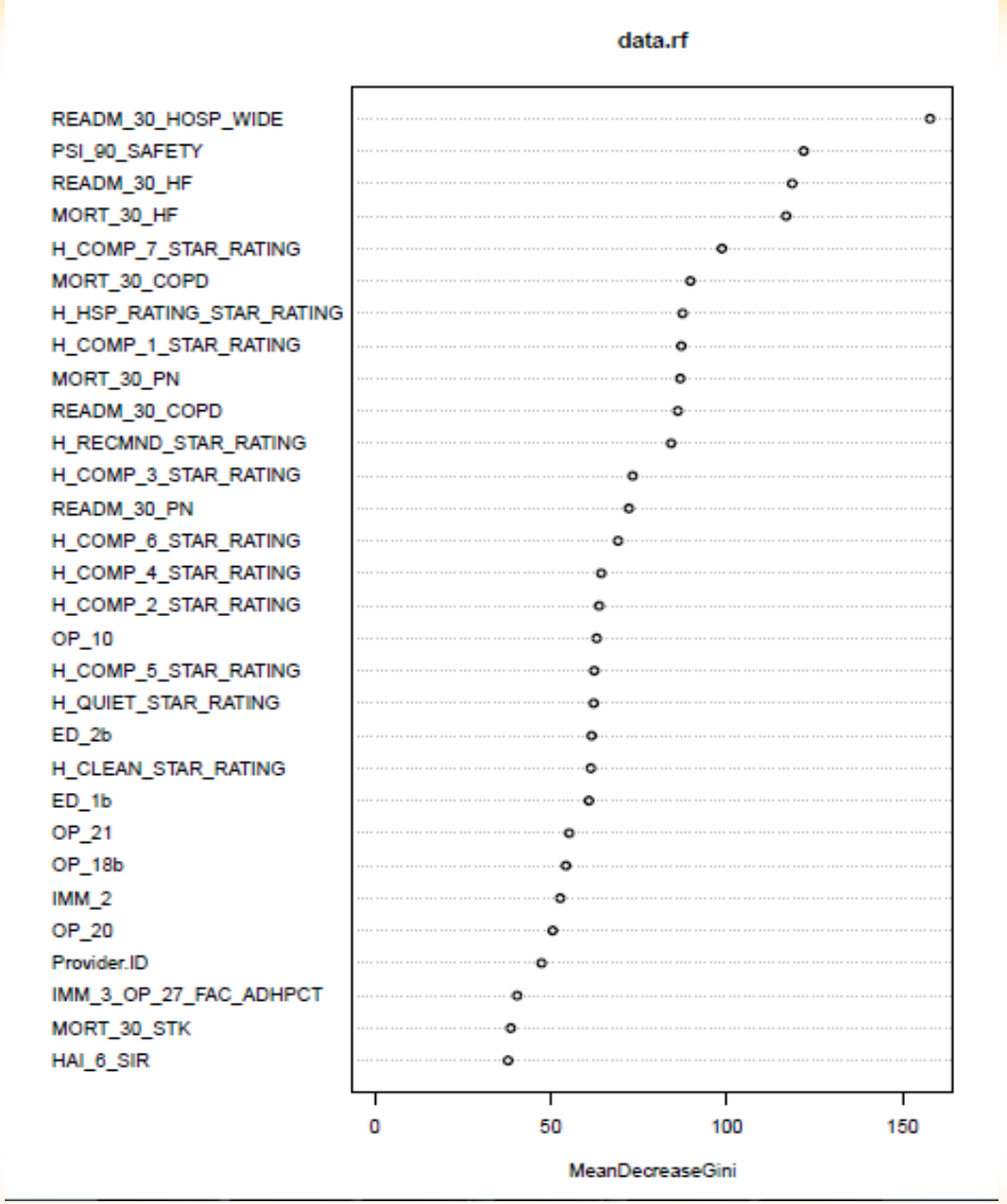
We chose to perform Random Forest of type – classification with 5 variables. This has resulted in provision of 1000 tree structure from top to bottom.

For this particular Random Forest generator, the error rate is the lowest we have observed that is at 21.46%.

In the consequent slides, we have provided a Plot as a result of the Random Forest application and also Variable Importance Factor index used to obtain the results.

As a result of the plot, you can observe that index value at a very healthy rate for many of the hospitals.


0.0
0.2
0.4
0.6
0.8
0
200
400
600
800
1000
trees


data.rf
READM_30_HOSP_WIDE
PSI_90_SAFETY
READM_30_HF
MORT_30_HF
H_COMP_7_STAR_RATING
MORT_30_COPD
H_HSP_RATING_STAR_RATING
H_COMP_1_STAR_RATING
MORT_30_PN
READM_30_COPD
H_RECMND_STAR_RATING
H_COMP_3_STAR_RATING
READM_30_PN
H_COMP_6_STAR_RATING
H_COMP_4_STAR_RATING
H_COMP_2_STAR_RATING
OP_10
H_COMP_5_STAR_RATING
H_QUIET_STAR_RATING
ED_2b
H_CLEAN_STAR_RATING
ED_1b
OP_21
OP_18b
IMM_2
OP_20
Provider.ID
IMM_3_OP_27_FAC_ADHPCT
MORT_30_STK
HAI_6_SIR
0
50
100
150
MeanDecreaseGini

CMS Ratings for Hospitals – Exploratory Factor Analysis

As part of the Unsupervised Learning, we took to the Exploratory Factor Analysis Approach as mentioned previously. Once the EDA is performed, we will further touch upon Random Factor Analysis approach used to obtain the Star Rating for the every hospital.

As previously obtained during the EDA approach, the 7 models are used again as the base for performing the analysis.

Initially, we deduce the number of factors required for the analysis using the Principal Component analysis. This will provide the number of factors required for each model.

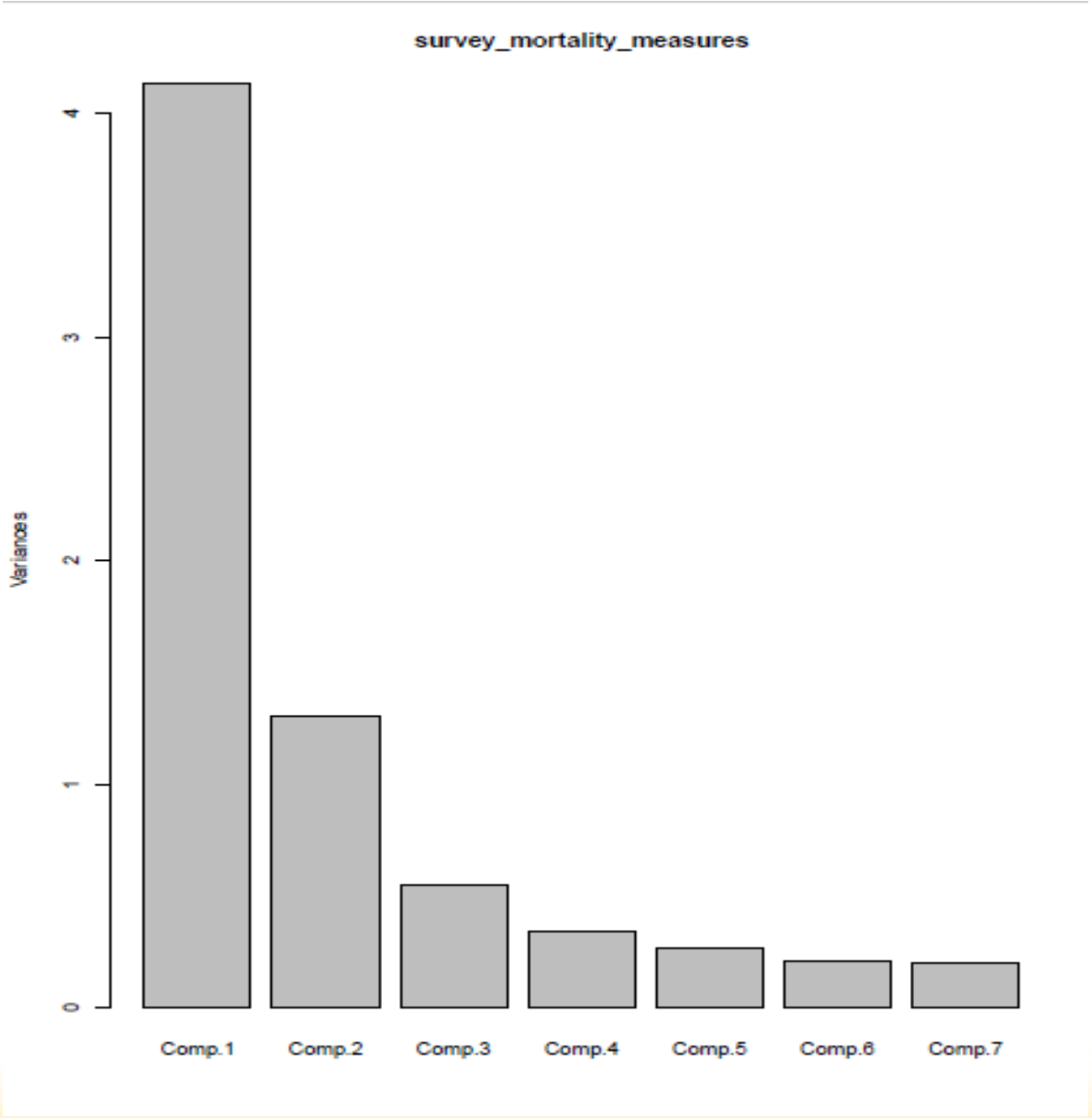
On this we have performed the Factor Analysis for which the plots have been provided for every model in subsequent slides.

However, it has to be noted that for every factor analysis request, we have performed scoring as well. The objectivity was not well received with this model. This is because we obtained few outliers when analysed. The in-depth analysis on these outliers doesn't justify the other results that we do have achieved with the factor analysis.

Measure Group 1 – Mortality

Inference –

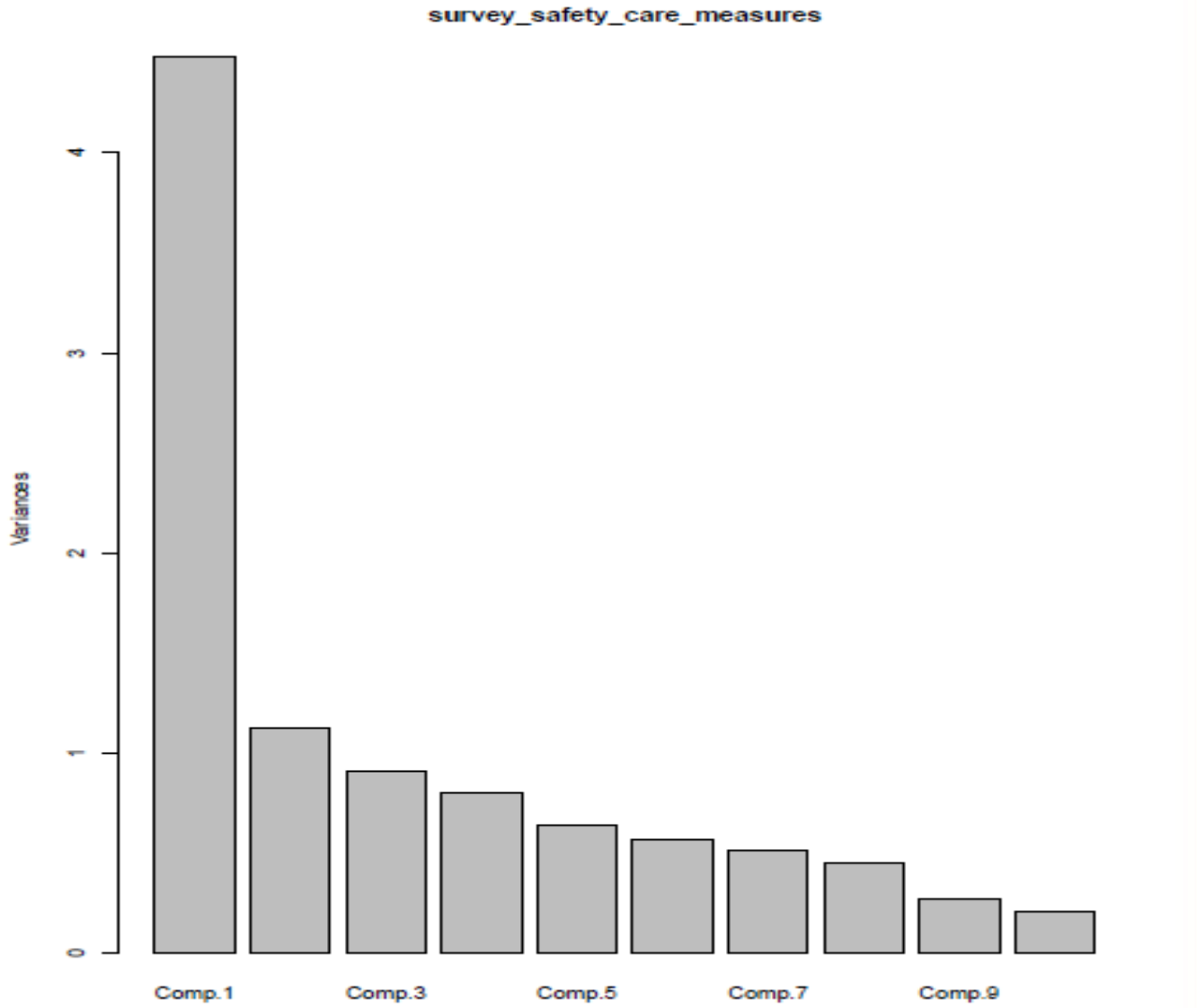
- The component 1 high variance reflects that Mortality Measure is a very high rated measure which is addressed to very immediately.
- However, we do have certain hospitals which are rated on the other end. This is due to multiple reasons which we don't entire analysis / grasp on.



Measure Group 2 – Safety Care

Inference –

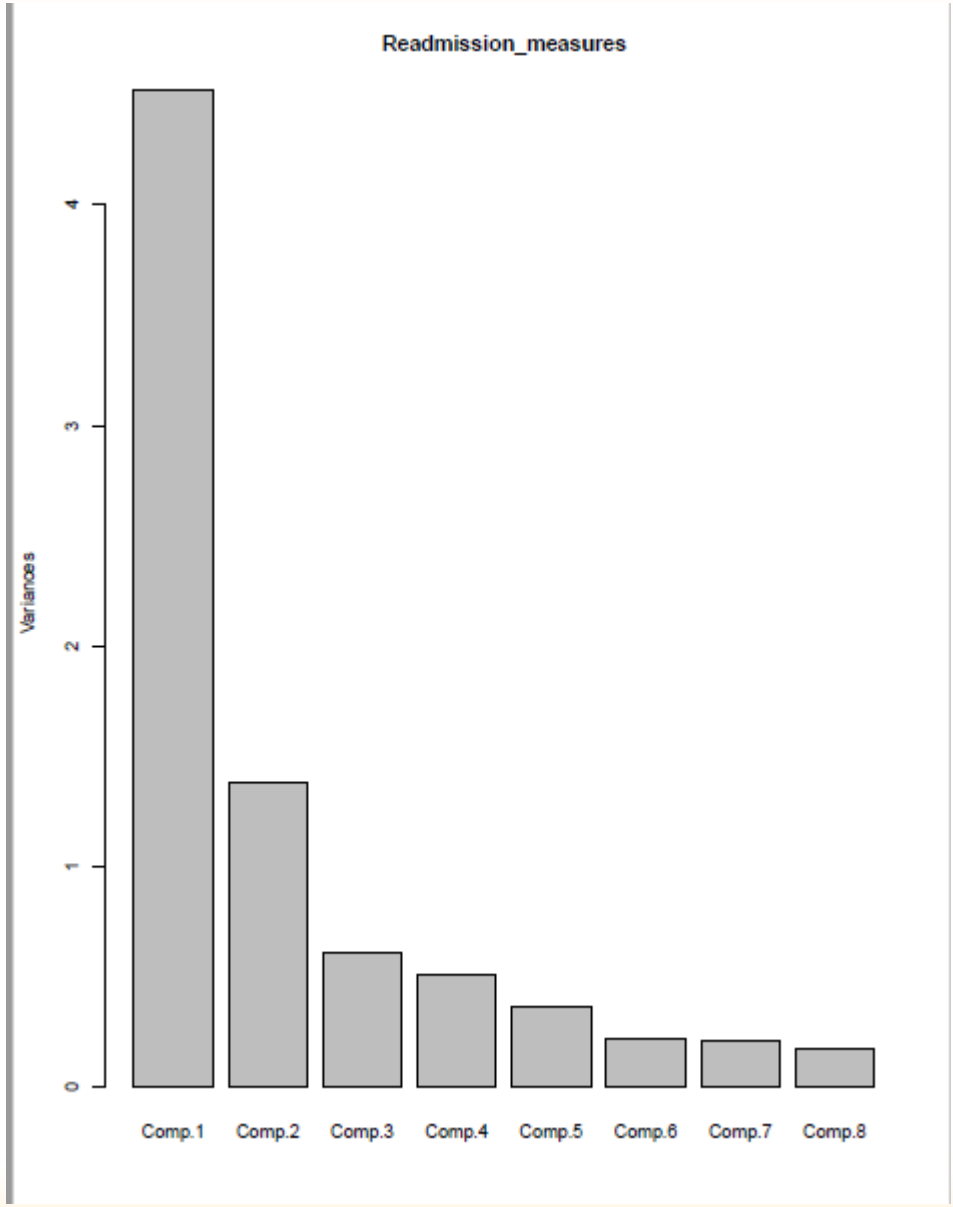
- The component 1 high variance reflects that Safety Care Measure is a very high rated measure which is addressed to very immediately similar to Mortality.
- However, we do have certain hospitals which are rated on the other end. This is due to multiple reasons which we don't entire analysis / grasp on similar to Mortality.



Measure Group 3 – Readmission Measures

Inference –

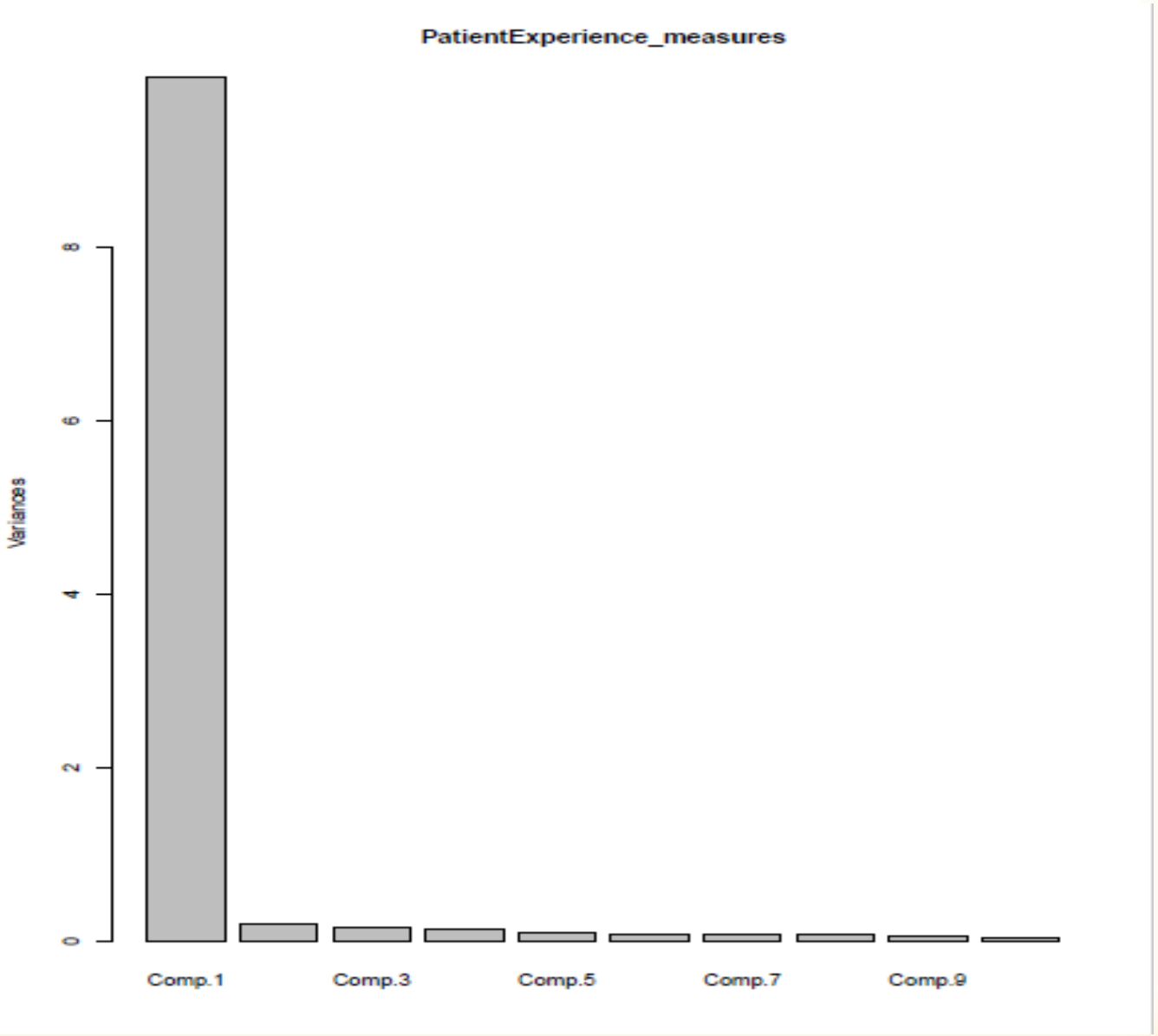
- The component 1 high variance reflects that Mortality Measure is a very high rated measure . This is because Re-admission is very critical factor to Hospital and its administration.
- However, we do have certain hospitals which are rated on the other end. This might due to certain factors such as Cold and Cough which needs a consultation by the patient which here for analysis becomes an outlier.



Measure Group 4 – Patient Experience

Inference –

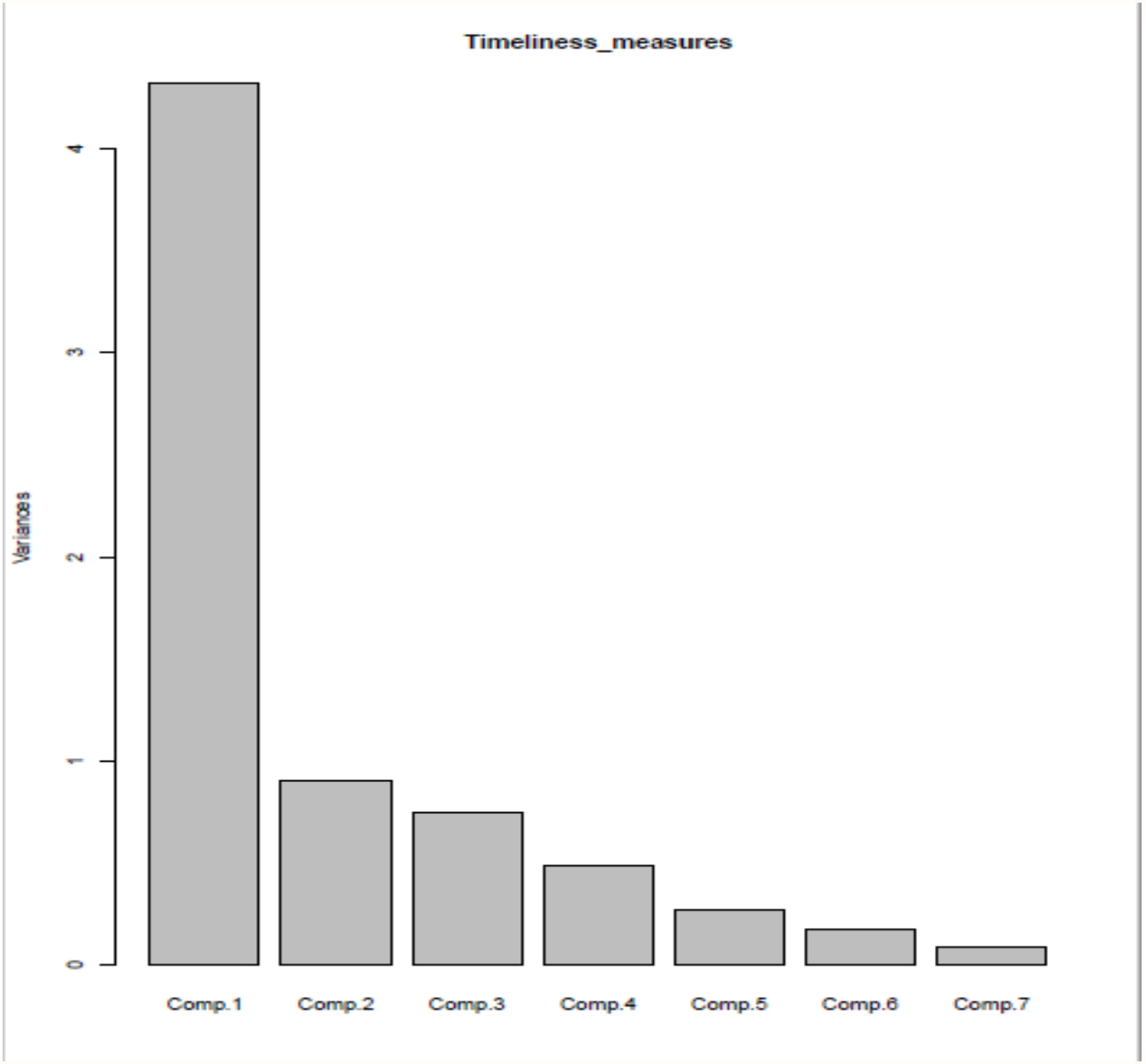
- The component 1 high variance reflects that Patient Experince is a very high rated measure which is addressed to very immediately.
- There are not many hospitals which rate on the other end. This might be due to multiple factors such as lawsuits and ethics which has a stringent following in the Hospitals.



Measure Group 5 – Timeliness Measures

Inference –

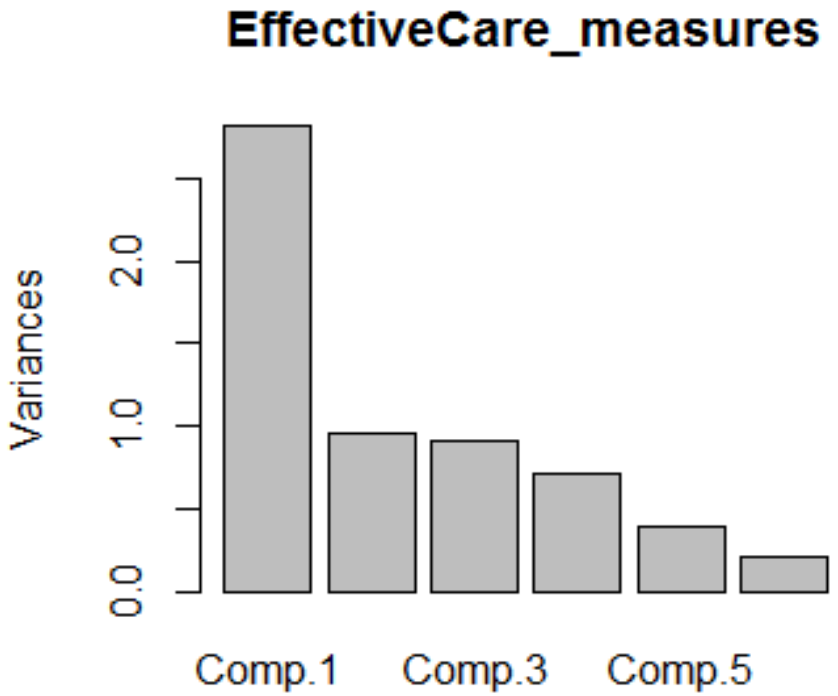
- The component 1 high variance reflects that Timeliness Measure is a high rated measure which does need more work on.
- Here though Timeliness does have many hospitals lying the in the safe region. This is because Patient Care comes first as we have seen earlier. That measure has an effect here with causing delay.



Measure Group 6 – Effective Care

Inference –

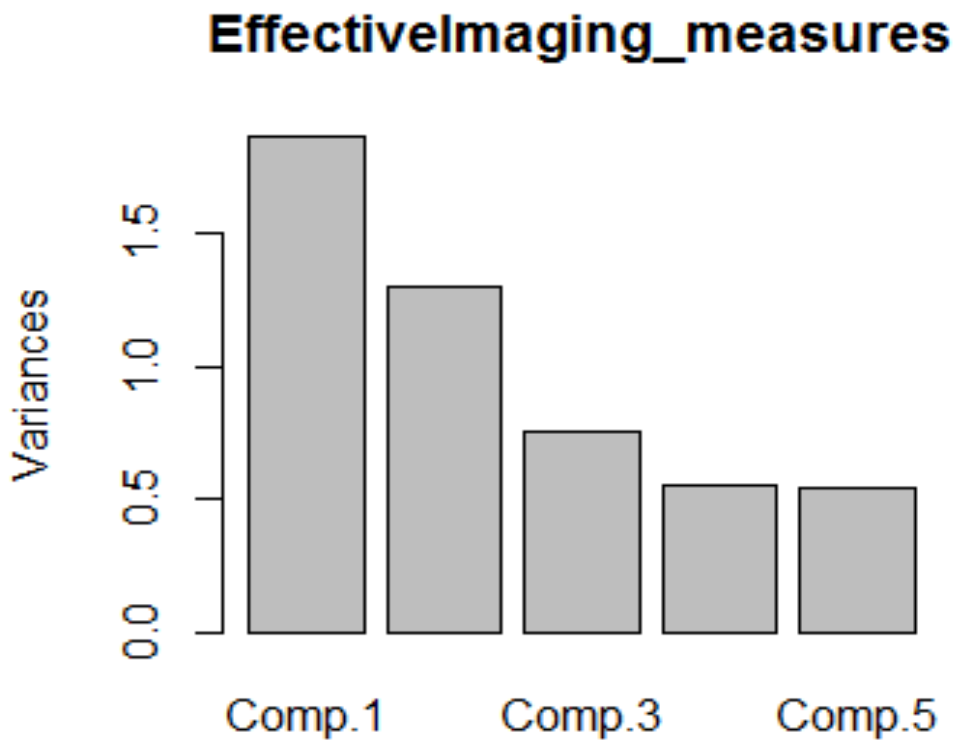
- The effective care is more of a result of the previous measures. If the right care and timeliness for example is provided then the effective care will also get tended to.
- Since effective care does rely on multiple measure this is a measure which cannot entirely justify rating of a hospital on its own.



Measure Group 7 – Effective Imaging

Inference –

- The result that we have observed is that similar to Effective Imaging, this is a measure we cannot entirely depend on to make a Star Rating for the Hospital.



CMS Ratings for Hospitals – Exploratory Factor Analysis Clustering

As part of the Unsupervised Learning, we took to the Exploratory Factor Analysis Approach as mentioned previously. Once the EDA is performed, we will further touch upon Random Factor Analysis approach used to obtain the Star Rating for the every hospital.

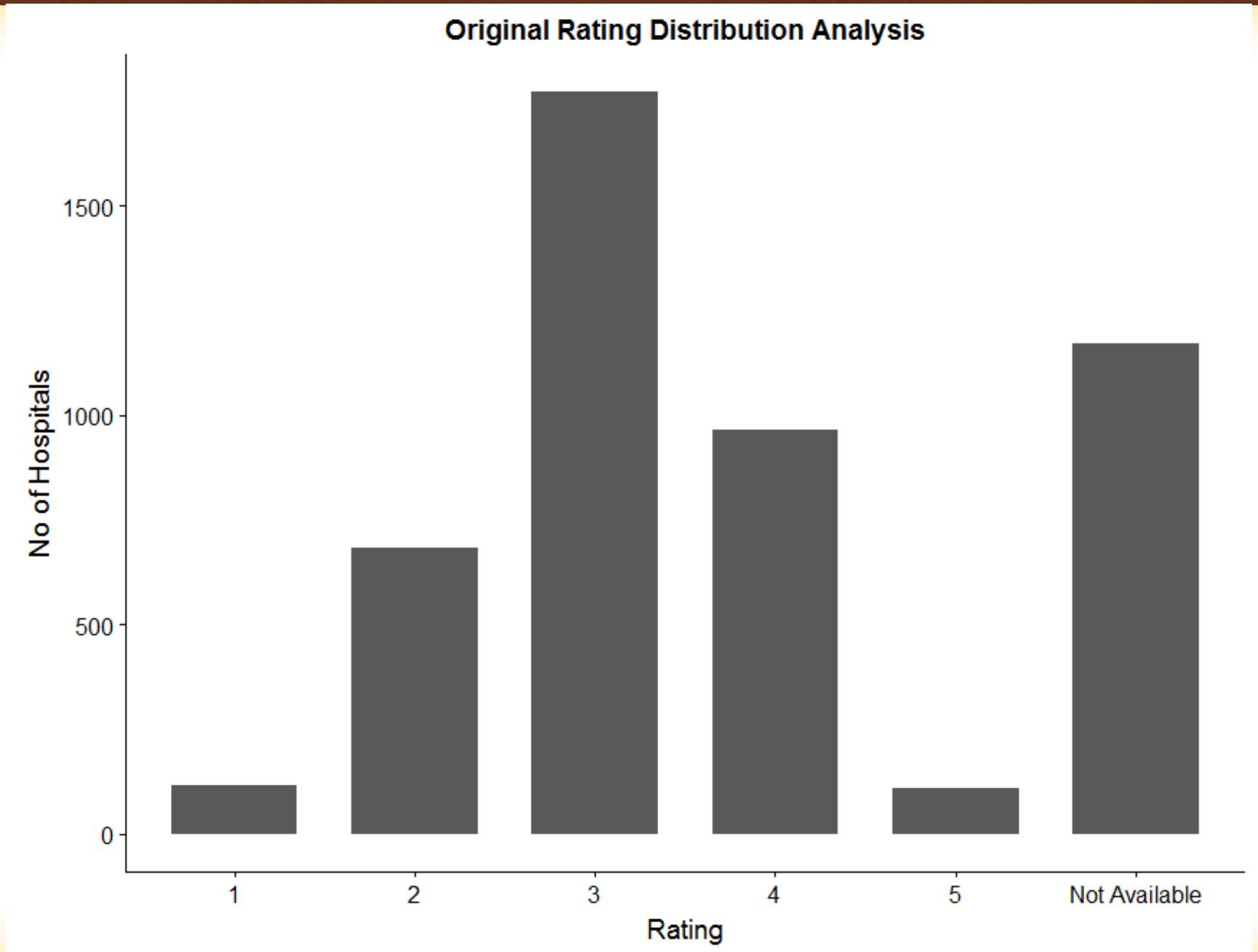
Once the Exploratory Factor Analysis has been performed for the entire set of 7 models, we performed clustering for the 7 models to obtain one set of results for the entire Hospital to obtain to a Star Clustering.

The following slides provide an insight into the Clustering results that has been plotted.

This is more of an extra point that has been taken care as a precaution to gather in-depth knowledge on the analysis route that has been taken during the Exploring Factor Analysis structure.

Inference –

- From the histogram plot, we can observe that the common trend of Hospitals is lying towards the Star Rating of 3.
- But we can also observe a huge chunk lying under the Rating of Unavailable. This is due to multiple outliers and factors which could not be consumed under the current Factor Analysis structure.


Original Rating Distribution Analysis
No of Hospitals
1500
1000
500
0
1
2
3
4
5
Not Available
Rating

END OF PRESENTATION