



Hospital Consumer Assessment of Healthcare Providers and Systems CIS – 5270 R-Project

Submitted By :-
Vrunda B Shah
Anusha Manjappa

ABSTRACT

The main goal of this project is to understand and measure variables affecting patient satisfaction and healthcare quality. The HCAHPS survey is considered in measuring these factors as it is based on nine key topics based on communication with doctors, communication with nurses, responsiveness of hospital staff, pain management, communication about medicines, discharge information, cleanliness of the hospital environment, quietness of the hospital environment, and transition of care.

DATASET

The dataset url:

- Hospital General Information

<https://data.medicare.gov/Hospital-Compare/Hospital-General-Information/xubh-q36u>

- HCAHPS – Hospital

<https://data.medicare.gov/Hospital-Compare/Patient-survey-HCAHPS-Hospital/dgck-sy fz>

- HCAHPS – State

<https://data.medicare.gov/Hospital-Compare/Patient-survey-HCAHPS-State/84jm-wiui>

We have used the above mentioned three public data sets provided by HCAHPS to evaluate the HCAHPS questions. We will be able to find out foremost hospitals in various states and will also be able to compare hospitals overall rating with other states hospitals and their ratings based on patient satisfaction. The survey provides star rating to all the above nine measures for all the

hospitals located in different states of USA which further helps consumer to choose hospitals during exigency.

The hospital general information dataset has nearly 5k rows and 22 columns. It contains details about various hospitals in different states like the services provided by the hospitals, ownership of the hospital, emergency services provided by them, overall general rating of the hospital and the comparison of various facilities provided by these hospitals. The hospital dataset contains detailed information about each hospital and the state dataset contains detailed information about all the star rating questions pertaining to each state for all the HCAHPS questions, their ratings, percentage of the ratings, etc. The hospital general information sheet gives us a general overview whereas the hospital and state sheets gives us detailed information about the questions pertaining to each state and each hospital. Therefore, the state and hospital sheets are more useful in getting a more detailed and specific visualization.

DATA CLEANING

The dataset required very little cleaning to be done. The corrections made on the existing dataset are as follows:

1) Data Cleaning Approach1

Using the **split** feature to separate data for better clarity.

Screenshot where we have merged data

Hospital Consumer Assessment of Healthcare Providers and Systems

State	County	Hospital.Type	Hospital.Ownership	Emergency.Services	Hospital.overall.rating	Hospital.overall.rating.foot
AL	MARSHALL	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	3	Not Available
AL	RANDOLPH	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	4	Not Available
AL	MADISON	Acute Care Hospitals	Proprietary	TRUE	3	Not Available
AK	ANCHORAGE	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	3	Not Available
AK	BETHEL	Acute Care Hospitals	Tribal	TRUE	3	This hospital's star rating o
AK	DILLINGHAM	Critical Access Hospitals	Voluntary non-profit - Private	TRUE	NA	There are too few measures
AK	NORTHWEST ARCTIC	Critical Access Hospitals	Tribal	TRUE	NA	Data are shown only for hos
AZ	APACHE	Acute Care Hospitals	Government - Federal	TRUE	NA	There are too few measures
AZ	COCHISE	Acute Care Hospitals	Voluntary non-profit - Other	FALSE	3	This hospital's star rating o
AZ	PIMA	Acute Care Hospitals	Government - Federal	FALSE	NA	There are too few measures
AZ	GILA	Acute Care Hospitals	Government - Federal	FALSE	NA	There are too few measures
AZ	APACHE	Acute Care Hospitals	Government - Federal	TRUE	2	This hospital's star rating o
AR	PLAQUEMINE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	2	Not Available
AZ	MUHAMMAD	Critical Access Hospitals	Tribal	FALSE	NA	There are too few measures
AZ	APACHE	Critical Access Hospitals	Voluntary non-profit - Private	FALSE	NA	There are too few measures
AR	CONWAY	Critical Access Hospitals	Voluntary non-profit - Church	TRUE	NA	There are too few measures
AR	CHICOT	Critical Access Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
CA	RIVERSIDE	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	3	Not Available
CA	SAN FRANCISCO	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	2	Not Available
CA	ORANGE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
CA	PLACER	Acute Care Hospitals	Voluntary non-profit - Other	TRUE	4	Not Available
CA	SANTA CLARA	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
CA	SAN JUAN	Acute Care Hospitals	Voluntary non-profit - Other	TRUE	3	Not Available
CA	ORANGE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
CA	SAN JOSE	Acute Care Hospitals	Government - Local	FALSE	3	Not Available
CA	ORANGE	Children's	Voluntary non-profit - Private	TRUE	NA	Data are shown only for hos
CO	BROOMFIELD	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	4	Not Available
CT	HARTFORD	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
DE	NEW CASTLE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
FL	OKALOOSA	Acute Care Hospitals	Proprietary	TRUE	4	Not Available
FL	OKALOOSA	Acute Care Hospitals	Proprietary	TRUE	2	Not Available
FL	MIAMIDADE	Acute Care Hospitals	Proprietary	TRUE	3	Not Available
AK	Not Available	Critical Access Hospitals	Government - Local	TRUE	NA	There are too few measures
AK	Not Available	Critical Access Hospitals	Government - Local	TRUE	NA	There are too few measures

Code for separating merged data via **separate** function

```
> data1 <- read.csv("GIDData Cleaning.csv", header=TRUE)
> View(data1)
> keepcols <- c(1,4,5,6,7,8,9)
> data2 <- data1[,keepcols]
> View(data2)
>
> library("tidyr")
> s_seperate<-separate(data2,State.County,c("State","ounty"),sep="-")
```

Warning message:

Too many values at 21 locations: 32, 301, 481, 515, 606, 657, 711, 743, 797, 814, 828, 829, 835, 903, 1052, 10

Hospital Consumer Assessment of Healthcare Providers and Systems

> View(s_seperate)

Screenshot after implementation of separate function where we have separated state and county

The screenshot shows the RStudio interface. The main window displays a data table with the following columns: State, County, Hospital.Type, Hospital.Ownership, Emergency.Services, Hospital.Overall.rating, and Hospital.Overall.rating. The table contains 28 rows of data, with the first row being: 1 AL MARSHALL Acute Care Hospitals Government - Hospital District or Authority TRUE 3 Not Available. The 's_seperate' function is highlighted in the Environment pane on the right, showing its arguments: all_wars_matrix, data1, data2, data3, data4, and s_seperate. The Environment pane also shows the values for these arguments, such as 'all_wars_matrix' being a numeric vector of length 461,290,309,474,311.

State	County	Hospital.Type	Hospital.Ownership	Emergency.Services	Hospital.Overall.rating	Hospital.Overall.rating
1 AL	MARSHALL	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	3	Not Available
2 AL	RANDOLPH	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	4	Not Available
3 AL	MADISON	Acute Care Hospitals	Proprietary	TRUE	3	Not Available
4 AK	ANCHORAGE	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	3	Not Available
5 AK	BETHEL	Acute Care Hospitals	Tribal	TRUE	3	This hospital's star rat
6 AK	DILLINGHAM	Critical Access Hospitals	Voluntary non-profit - Private	TRUE	NA	There are too few meas
7 AK	NORTHWEST ARCTIC	Critical Access Hospitals	Tribal	TRUE	NA	Data are shown only for
8 AZ	APACHE	Acute Care Hospitals	Government - Federal	TRUE	NA	There are too few meas
9 AZ	COCONINO	Acute Care Hospitals	Voluntary non-profit - Other	FALSE	3	This hospital's star rat
10 AZ	PIMA	Acute Care Hospitals	Government - Federal	FALSE	NA	There are too few meas
11 AZ	GILA	Acute Care Hospitals	Government - Federal	FALSE	NA	There are too few meas
12 AZ	APACHE	Acute Care Hospitals	Government - Federal	TRUE	2	This hospital's star rat
13 AR	PULASKI	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	2	Not Available
14 AZ	NAVAJO	Critical Access Hospitals	Tribal	FALSE	NA	There are too few meas
15 AZ	APACHE	Critical Access Hospitals	Voluntary non-profit - Private	FALSE	NA	There are too few meas
16 AR	CONWAY	Critical Access Hospitals	Voluntary non-profit - Church	TRUE	NA	There are too few meas
17 AR	CHICOT	Critical Access Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
18 CA	RIVERSIDE	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	3	Not Available
19 CA	SAN FRANCISCO	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	2	Not Available
20 CA	ORANGE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
21 CA	PLACER	Acute Care Hospitals	Voluntary non-profit - Other	TRUE	4	Not Available
22 CA	SANTA CLARA	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
23 CA	ALAMEDA	Acute Care Hospitals	Voluntary non-profit - Other	TRUE	3	Not Available
24 CA	ORANGE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
25 CA	LOS ANGELES	Acute Care Hospitals	Government - Local	FALSE	3	Not Available
26 CA	ORANGE	Childrens	Voluntary non-profit - Private	TRUE	NA	Data are shown only for
27 CO	BROOMFIELD	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	4	Not Available
28 CT	HARTFORD	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available

Hospital Consumer Assessment of Healthcare Providers and Systems

The screenshot shows the RStudio interface with a dataset loaded from a CSV file. The dataset has columns: State, county, Hospital.Type, Hospital.Ownership, Emergency.Services, Hospital.overall.rating, and Hospital.overall.rating.2. The data is displayed in a table with 18 rows. The 'Hospital.overall.rating' column contains several NA values. The console shows the following code and output:

```
data1 <- read.csv("GIData Cleaning.csv", header=TRUE)
> View(data1)
> keepcols <- c(1,4,5,6,7,8,9)
> data2 <- data1[,keepcols]
> View(data2)
> library("tidyr")
> s_separate<-separate(data2,State,county,c("State","county"),sep="-")
Warning message:
Too many values at 21 locations: 32, 301, 481, 515, 606, 657, 711, 743, 797, 814, 828, 829, 835, 903, 1052, 1053, 1093, 1154, 1162, 1231, ...
> View(s_separate)
```

The Environment pane shows the following objects:

- Global Environment
- all_wars_matrix: num [1:6, 1:2] 461 290 309 474 311 ...
- data1: 4819 obs. of 23 variables
- data2: 4819 obs. of 7 variables
- data3: 104 obs. of 3 variables
- data4: 52 obs. of 3 variables
- s_separate: 4819 obs. of 8 variables
- star_wars_matrix: num [1:9, 1:2] 461 290 309 314 248 ...
- star_wars_matrix2: num [1:9, 1:2] 474 311 380 552 339 ...
- visitor_us_n_nonus: num [1:6, 1:2] 92.2 58.1 61.9 94.9 62.1 ...
- A_new_hope: num [1:2] 461 314
- Attack_of_the_clones: num [1:2] 311 339
- keepcols: num [1:7] 1 4 5 6 7 8 9

2) Data Cleaning Approach2

Removal of NA values present in the dataset

In the dataset shown below we have NA values

Hospital Consumer Assessment of Healthcare Providers and Systems

The screenshot shows the RStudio interface with a data frame loaded. The data frame has columns: State, County, Hospital.Type, Hospital.Ownership, Emergency.Services, Hospital.Overall.Rating, and Hospital.Overall.Rating.Footnote. A red rectangle highlights a section of the data frame where many values are NA, specifically in the 'Hospital.Overall.Rating' and 'Hospital.Overall.Rating.Footnote' columns for rows 11 through 22. The Environment pane on the right shows the loaded objects: all_hars_matrix, data1, data2, data3, s_separate, star_hars_matrix, star_hars_matrix2, visitor_us_h_horus, A_new_hope, Attack_of_the_clones, and keepcols.

	State	County	Hospital.Type	Hospital.Ownership	Emergency.Services	Hospital.Overall.Rating	Hospital.Overall.Rating.Footnote
1	AL	MARSHALL	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	3	Not Available
2	AL	RANDOLPH	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	4	Not Available
3	AL	MADISON	Acute Care Hospitals	Proprietary	TRUE	3	Not Available
4	AK	ANCHORAGE	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	3	Not Available
5	AK	BETHEL	Acute Care Hospitals	Tribal	TRUE	3	The hospital's star rating o
6	AK	DILLINGHAM	Critical Access Hospitals	Voluntary non-profit - Private	TRUE	NA	There are too few measures
7	AK	NORTHWEST ARCTIC	Critical Access Hospitals	Tribal	TRUE	NA	Data are shown only for hos
8	AZ	APACHE	Acute Care Hospitals	Government - Federal	TRUE	NA	There are too few measures
9	AZ	COCHISE	Acute Care Hospitals	Voluntary non-profit - Other	FALSE	3	The hospital's star rating o
10	AZ	PIMA	Acute Care Hospitals	Government - Federal	FALSE	NA	There are too few measures
11	AZ	GILA	Acute Care Hospitals	Government - Federal	FALSE	NA	There are too few measures
12	AZ	APACHE	Acute Care Hospitals	Government - Federal	TRUE	2	The hospital's star rating o
13	AR	PLAQUEMINE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	2	Not Available
14	AZ	MUDD	Critical Access Hospitals	Tribal	FALSE	NA	There are too few measures
15	AZ	APACHE	Critical Access Hospitals	Voluntary non-profit - Private	FALSE	NA	There are too few measures
16	AR	CONWAY	Critical Access Hospitals	Voluntary non-profit - Church	TRUE	NA	There are too few measures
17	AR	CHICOT	Critical Access Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
18	CA	RIVERSIDE	Acute Care Hospitals	Government - Hospital District or Authority	TRUE	3	Not Available
19	CA	SAN FRANCISCO	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	2	Not Available
20	CA	ORANGE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
21	CA	PLACER	Acute Care Hospitals	Voluntary non-profit - Other	TRUE	4	Not Available
22	CA	SANTA CLARA	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
23	CA	SAN JUAN	Acute Care Hospitals	Voluntary non-profit - Other	TRUE	3	Not Available
24	CA	ORANGE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
25	CA	SAN JOSE	Acute Care Hospitals	Government - Local	FALSE	3	Not Available
26	CA	ORANGE	Childrens	Voluntary non-profit - Private	TRUE	NA	Data are shown only for hos
27	CO	BROOMFIELD	Acute Care Hospitals	Voluntary non-profit - Church	TRUE	4	Not Available
28	CT	HARTFORD	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	3	Not Available
29	DE	NEW CASTLE	Acute Care Hospitals	Voluntary non-profit - Private	TRUE	4	Not Available
30	FL	KALAMAZOO	Acute Care Hospitals	Proprietary	TRUE	4	Not Available
31	FL	POLK	Acute Care Hospitals	Proprietary	TRUE	2	Not Available
32	FL	MIAMI-DADE	Acute Care Hospitals	Proprietary	TRUE	3	Not Available
33	AK	Not Available	Critical Access Hospitals	Government - Local	TRUE	NA	There are too few measures
34	AK	Not Available	Critical Access Hospitals	Government - Local	TRUE	NA	There are too few measures

Code for removal of NA values, replacing it with the mean

```
> data1 <- read.csv("GIData Cleaning.csv", header=TRUE)

> View(data1)

> keepcols <- c(1,4,5,6,7,8,9)

> data2 <- data1[,keepcols]

> View(data2)

> for(i in 1:ncol(data2)){

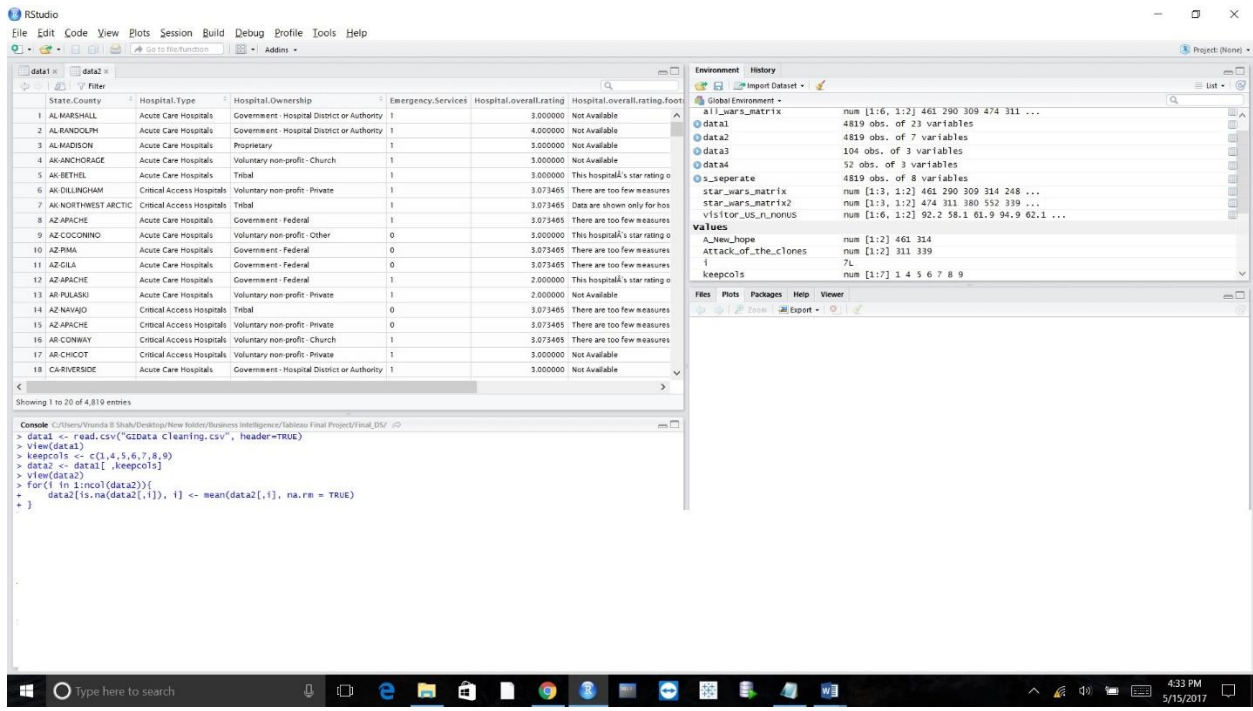
+   data2[is.na(data2[,i]), i] <- mean(data2[,i], na.rm = TRUE)

+ }
```

Screenshot after removal of NA values and replacing it with mean



Hospital Consumer Assessment of Healthcare Providers and Systems



RStudio interface showing data cleaning for Hospital Consumer Assessment of Healthcare Providers and Systems. The console displays the following R code and output:

```
> data1 <- read.csv("C:/Users/Phanda/Desktop/Healthcare/Business Intelligence/tables/Final Project/Final_DS1_01")
> View(data1)
> keepcols <- c(1,4,5,6,7,8,9)
> data2 <- data1[,keepcols]
> View(data2)
> for(i in 1:nrow(data2)){
+   data2[is.na(data2[,i]), i] <- mean(data2[,i], na.rm = TRUE)
+ }
```

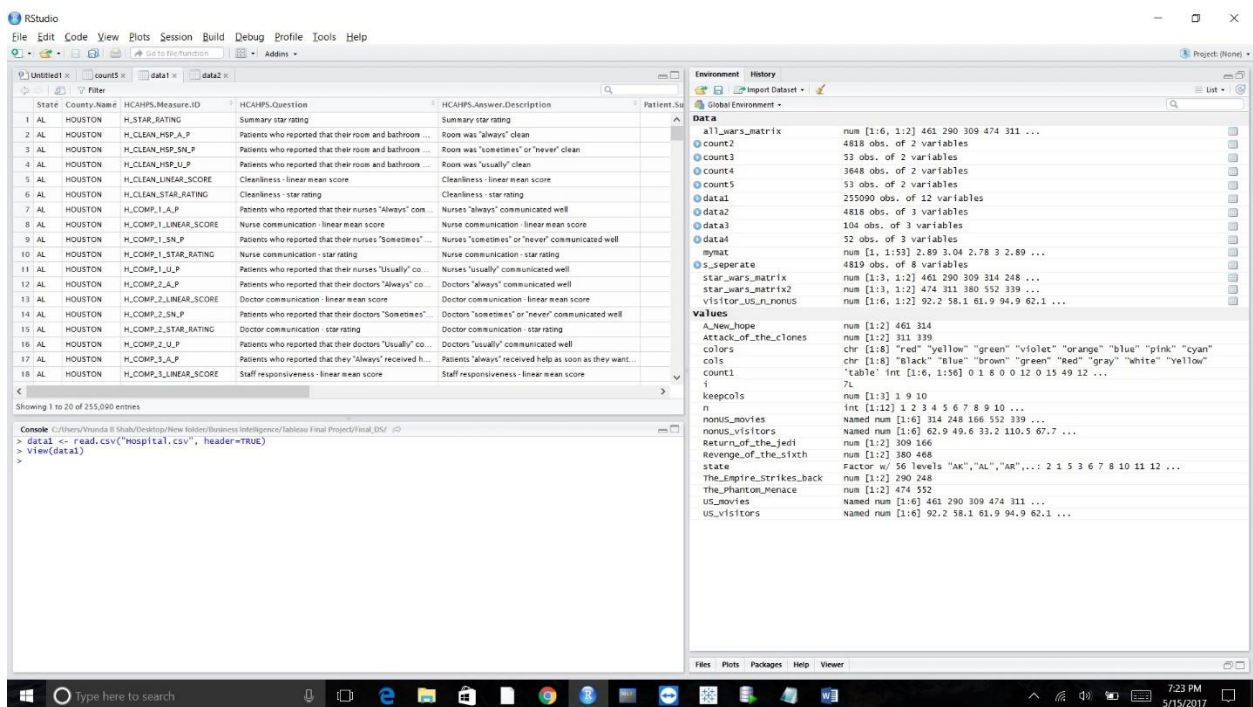
The Environment pane shows the following objects:

- all_hosp_matrix: num [1:6, 1:2] 461 290 309 474 311 ...
- data1: 4819 obs. of 23 variables
- data2: 4819 obs. of 7 variables
- data3: 104 obs. of 3 variables
- data4: 52 obs. of 3 variables
- s_separate: 4819 obs. of 8 variables
- star_wars_matrix: num [1:3, 1:2] 461 290 309 314 248 ...
- star_wars_matrix2: num [1:3, 1:2] 474 311 380 552 339 ...
- visitor_us_norhus: num [1:6, 1:2] 92.2 58.1 61.9 94.9 62.1 ...
- A_new_hope: num [1:2] 461 314
- Attack_of_the_clones: num [1:2] 311 339
- i: 7L
- keepcols: num [1:7] 1 4 5 6 7 8 9

3) Data Cleaning Approach3

We have used this approach to remove the rows that we do not require for our analysis.

Removal of unwanted columns.



RStudio interface showing data cleaning for Hospital Consumer Assessment of Healthcare Providers and Systems. The console displays the following R code and output:

```
> data1 <- read.csv("C:/Users/Phanda/Desktop/Healthcare/Business Intelligence/tables/Final Project/Final_DS1_01")
> View(data1)
> keepcols <- c(1,4,5,6,7,8,9)
> data2 <- data1[,keepcols]
> View(data2)
> for(i in 1:nrow(data2)){
+   data2[is.na(data2[,i]), i] <- mean(data2[,i], na.rm = TRUE)
+ }
```

The Environment pane shows the following objects:

- all_hosp_matrix: num [1:6, 1:2] 461 290 309 474 311 ...
- count2: 4819 obs. of 2 variables
- count3: 53 obs. of 2 variables
- count4: 3648 obs. of 2 variables
- count5: 53 obs. of 2 variables
- data1: 255090 obs. of 22 variables
- data2: 4819 obs. of 3 variables
- data3: 104 obs. of 3 variables
- data4: 52 obs. of 3 variables
- mydat: num [1, 1:53] 2.89 3.04 2.78 3 2.89 ...
- s_separate: 4819 obs. of 8 variables
- star_wars_matrix: num [1:3, 1:2] 461 290 309 314 248 ...
- star_wars_matrix2: num [1:3, 1:2] 474 311 380 552 339 ...
- visitor_us_norhus: num [1:6, 1:2] 92.2 58.1 61.9 94.9 62.1 ...
- A_new_hope: num [1:2] 461 314
- Attack_of_the_clones: num [1:2] 311 339
- colors: chr [1:8] "red" "yellow" "green" "violet" "orange" "blue" "pink" "cyan"
- cols: chr [1:8] "black" "blue" "brown" "green" "red" "gray" "white" "yellow"
- count1: "table" int [1:6, 1:56] 0 1 6 0 0 12 0 15 49 12 ...
- i: 7L
- keepcols: num [1:3] 1 9 10
- n: int [1:12] 1 2 3 4 5 6 7 8 9 10 ...
- nonus_movies: named num [1:6] 314 248 186 552 339 ...
- nonus_visitors: named num [1:6] 62.9 49.6 33.2 110.5 67.7 ...
- Return_of_the_jedi: num [1:2] 309 166
- Revenge_of_the_Sith: num [1:2] 380 468
- state: Factor w/ 56 levels "AK","AL","AR",...: 2 1 5 3 6 7 8 10 11 12 ...
- The_Empire_Strikes_Back: num [1:2] 290 248
- The_Phantom_Menace: num [1:2] 474 552
- US_movies: named num [1:6] 461 290 309 474 311 ...
- US_visitors: named num [1:6] 92.2 58.1 61.9 94.9 62.1 ...

Hospital Consumer Assessment of Healthcare Providers and Systems

Code for removal of unwanted columns

```
> data1 <- read.csv("Hospital.csv", header=TRUE)

> View(data1)

> keepcols <- c(1,9,10)

> data2<- data1[,keepcols]

> View(data2)
```

Data after removal of unwanted columns

The screenshot displays the RStudio interface with the following components:

- Environment Panel:** Lists objects in the global environment, including `all_wars_matrix`, `count2`, `count3`, `count4`, `count5`, `data1`, `data2`, `data3`, `data4`, `mydata`, `s_separate`, `star_wars_matrix`, `star_wars_matrix2`, `visitor_us_n_nones`, `A_new_hope`, `Attack_of_the_clones`, `colors`, `col1`, `col2`, `count1`, `i`, `keepcols`, `n`, `nonus_movies`, `nonus_visitors`, `Return_of_the_jedi`, `Revenge_of_the_sith`, `state`, `The_empire_strikes_back`, `The_phantom_menace`, `US_movies`, and `US_visitors`.
- Console:** Shows the execution of the following R code:

```
> data1 <- read.csv("Hospital.csv", header=TRUE)
> View(data1)
> keepcols <- c(1,9,10)
> data2 <- data1[,keepcols]
> View(data2)
>
```
- Data Viewer:** Displays a table with columns: `State`, `Number.of.Completed.Surveys`, `Survey.Response.Rate.Percent`, and `State`. The data is filtered to show 1 to 20 of 255,090 entries.

DATA VISUALISATION

Question1

Compare hospital overall rating of each state across USA

Reference Sheet: GI1.csv

Elements used: Bar Chart

Insights:

We have used **bar chart** to analyze the data, we can view that following state comes under Top5. The HCAHPS Summary Star Rating combines all of the HCAHPS Star Ratings. It is the average of all of the Star Ratings of the HCAHPS measures. It is constructed based on 7 HCAHPS Composite measures, single Star Rating for the HCAHPS Individual Items, single Star Rating for the HCAHPS Global Items.

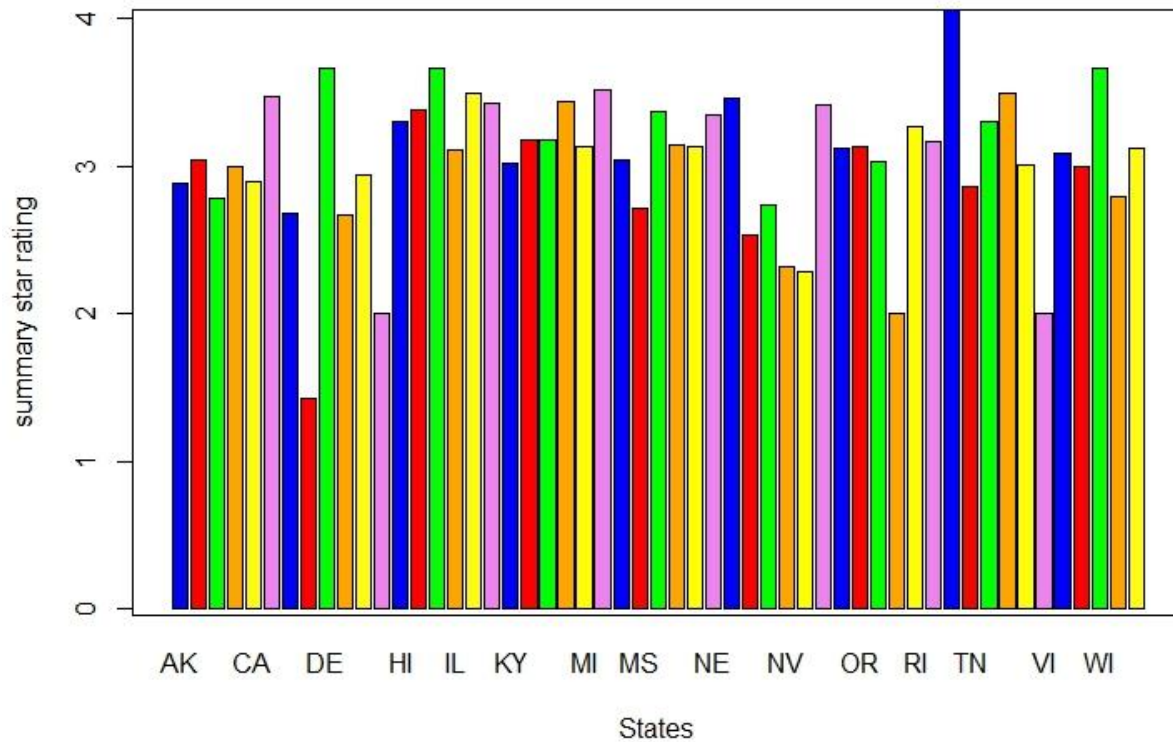
From the below visualization, top 5 state with highest average rating is mentioned below in tabular format.

<u>State</u>	<u>Average Rating</u>
SD	4.05
DE	3.66
ID	3.66

Hospital Consumer Assessment of Healthcare Providers and Systems

WE	3.66
MN	3.51

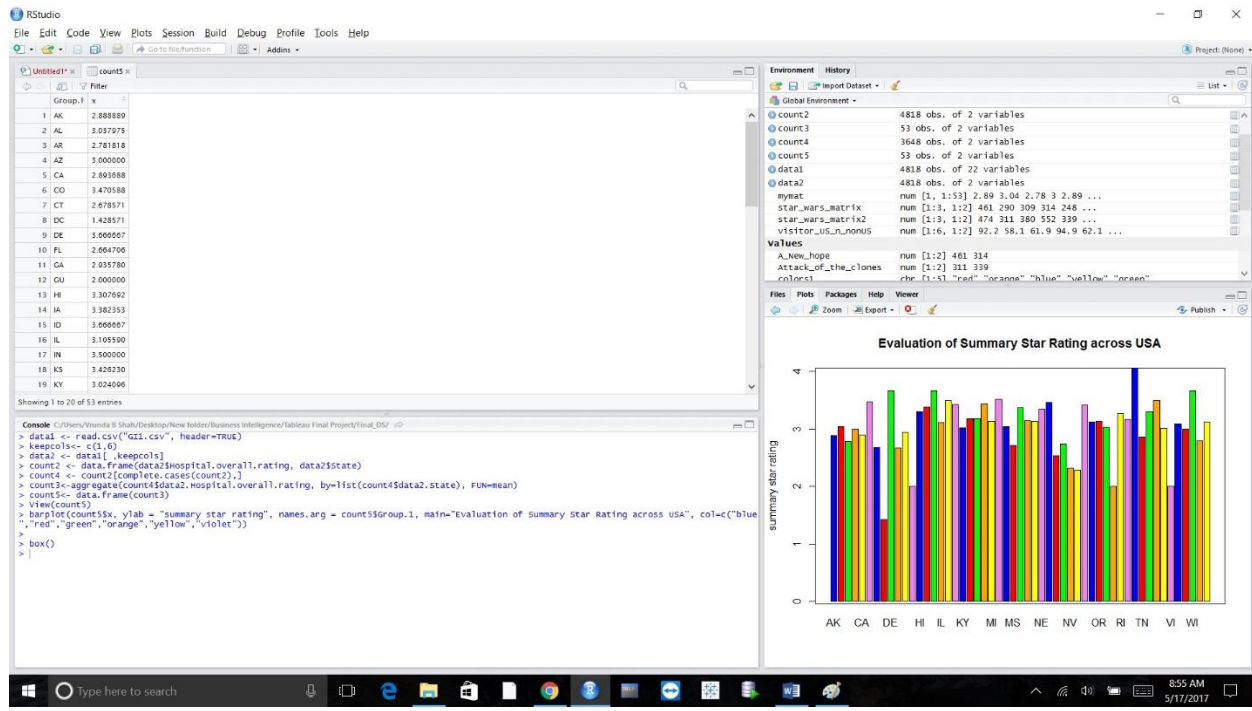
Evaluation of Summary Star Rating across USA



Axis Specifications

<u>X-Axis</u>	<u>Y-Axis</u>
States	Summary Star Rating

Hospital Consumer Assessment of Healthcare Providers and Systems



Question2- Find out how many surveys we completed for each state and what was its survey response rate

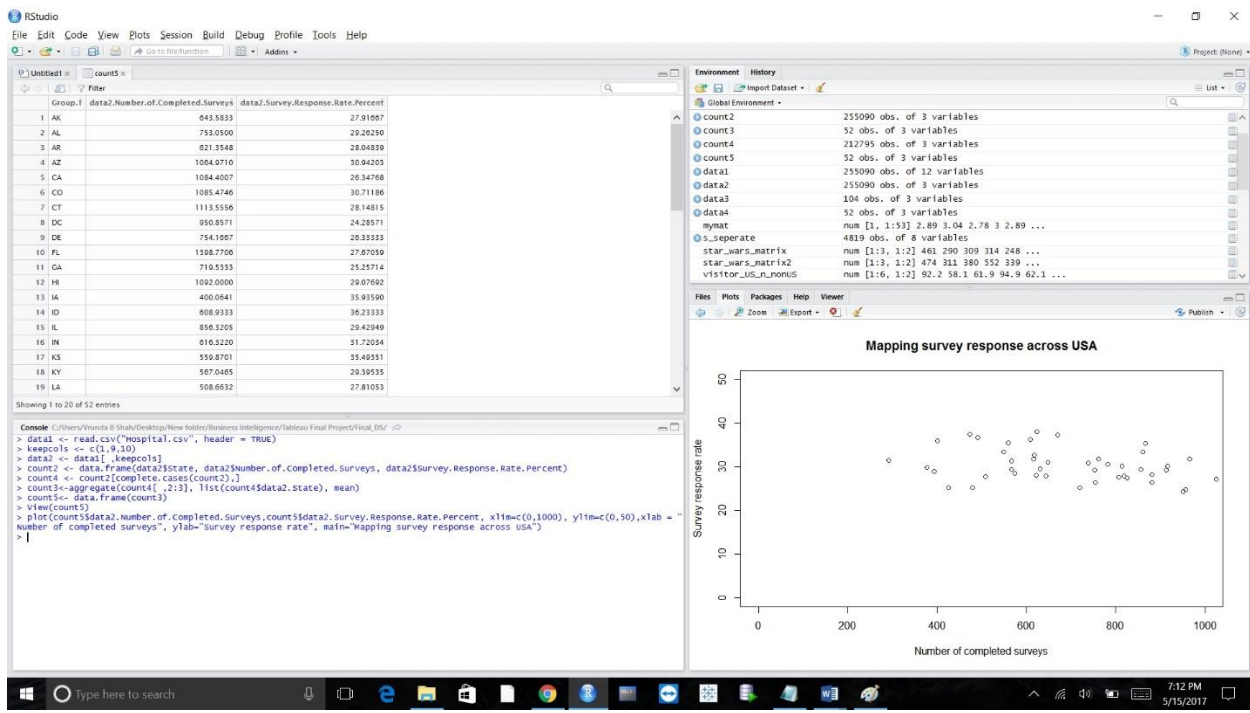
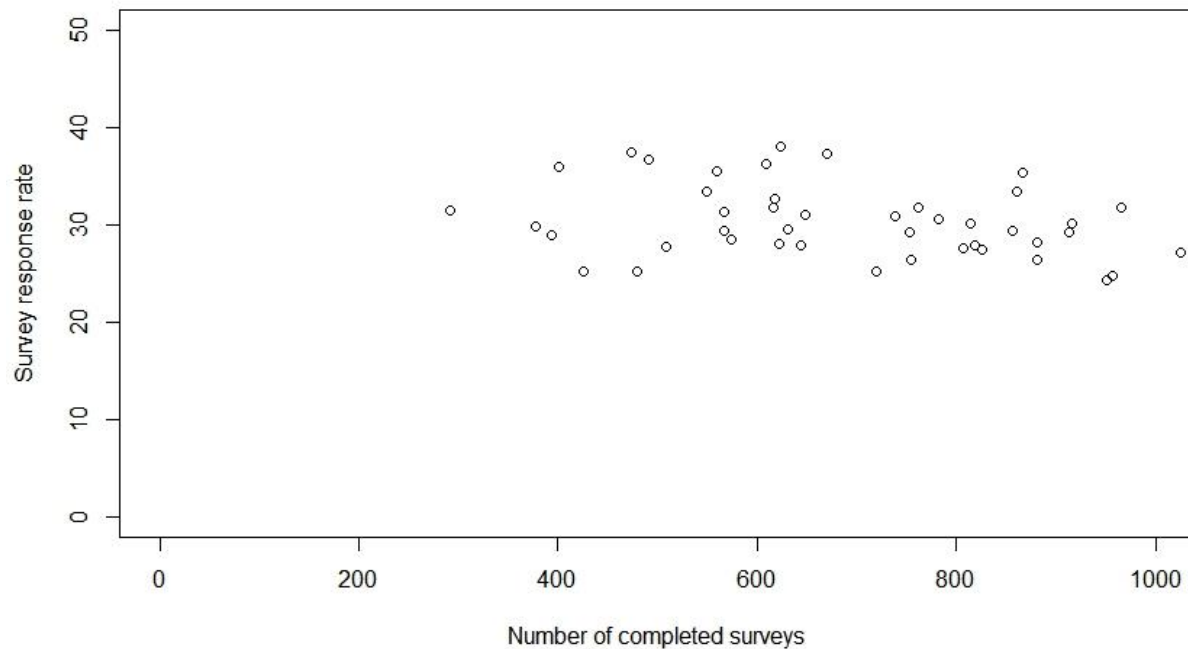
Reference Sheet: Hospital.csv

Elements used: Scatter Plot

Insights: We have used **scatter plot** to analyze the data, we found that state who completed 400 to 600 surveys has highest response state across USA than states which has completed 1000-1400 surveys

In the below diagram you can see WI state which has completed on an average 623 surveys has highest response rate 37.97%. Similarly FL state completed 1399 surveys on an average and has lowest average response rate 27.60%

Mapping survey response across USA



Axis Specifications

<u>X-Axis</u>	<u>Y-Axis</u>
Number of Completed Survey	Survey Response Rate

Question 3

What is the average Patient Survey Star Rating for all the star rating HCAHPS Question?

Reference Sheet: HCAHPS – Hospital Sheet

Elements used: Bar graph, Histogram, Function, Script

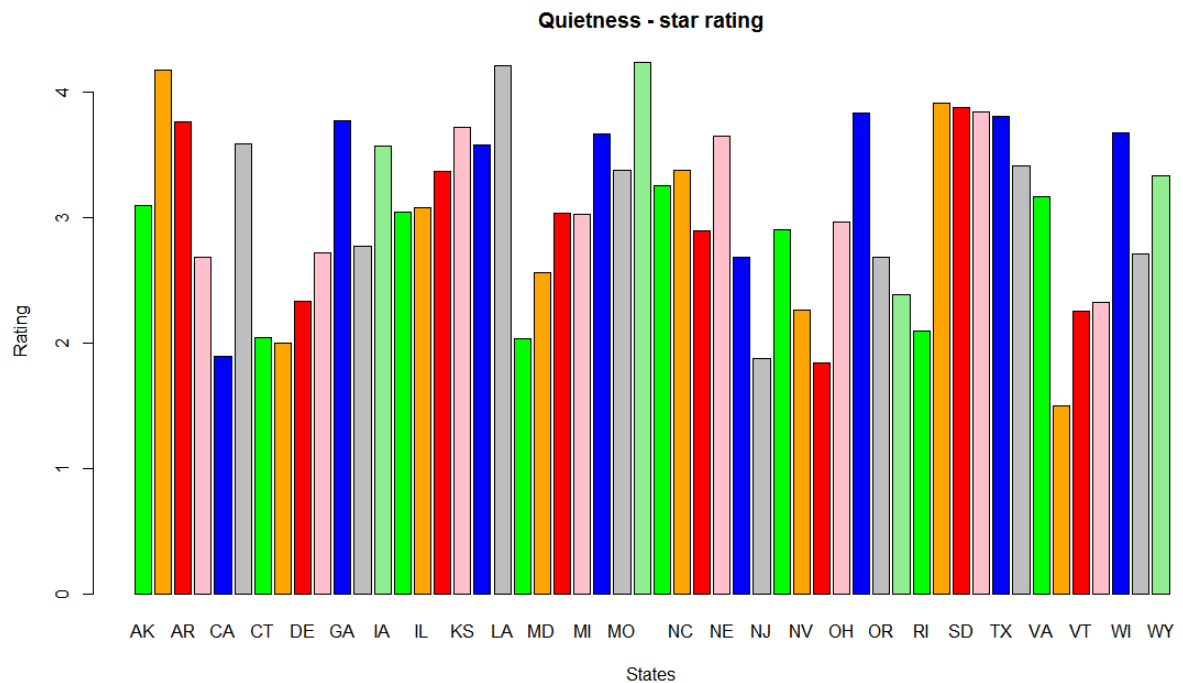
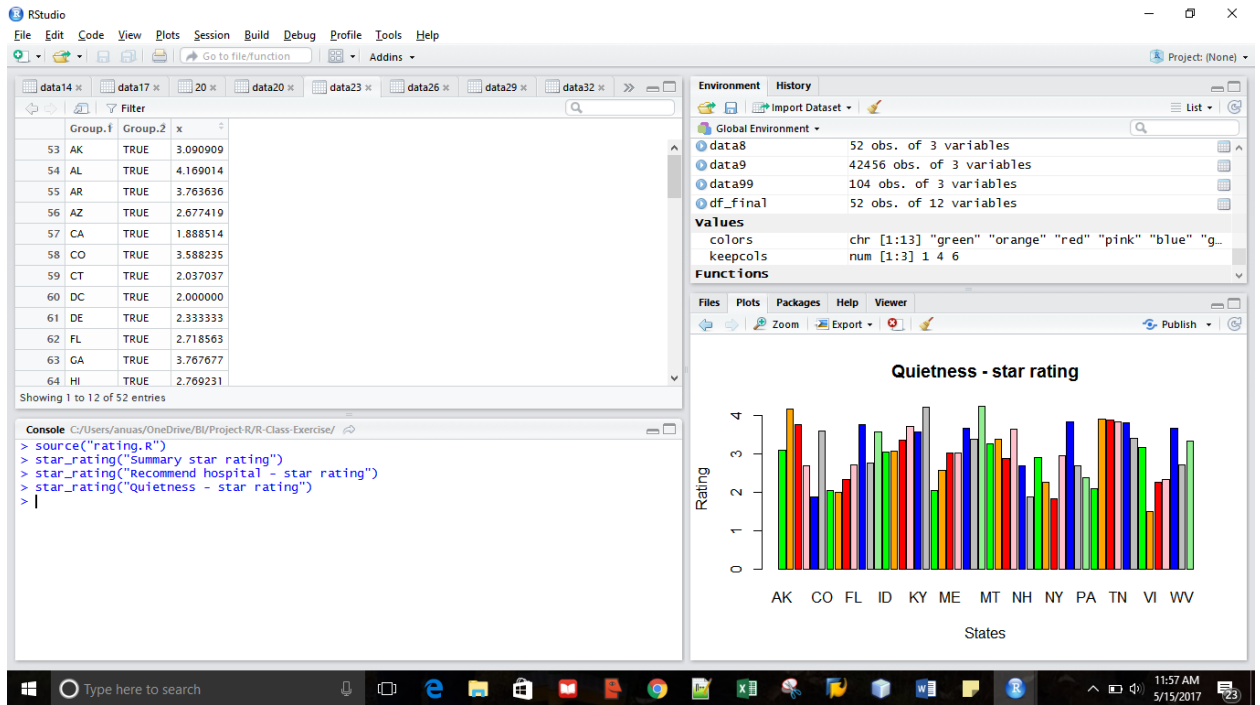
a) Visualization 1: -

Insights: We have used a **Bar** plot and a **Histogram** to visualize the average Patient Survey Star Rating for all the star rating HCAHPS Question. We have 11 different star rating HCAHPS Questions. We have found out the state-wise average Patient Survey Star Rating for each question. This is implemented using a function Rating.R . The user is asked to enter the star rating question for which the average rating is to be calculated and the respective averages for each state of that question is displayed.

- Output1

User wants to know the average rating in each state for the question- “Quietness star rating “. Therefore, on running the function **star_rating** with an input string parameter “**Quietness star rating** “, the respective dataframe is displayed on the top on the screen and the respective graph is shown on the right of the screen.

Hospital Consumer Assessment of Healthcare Providers and Systems

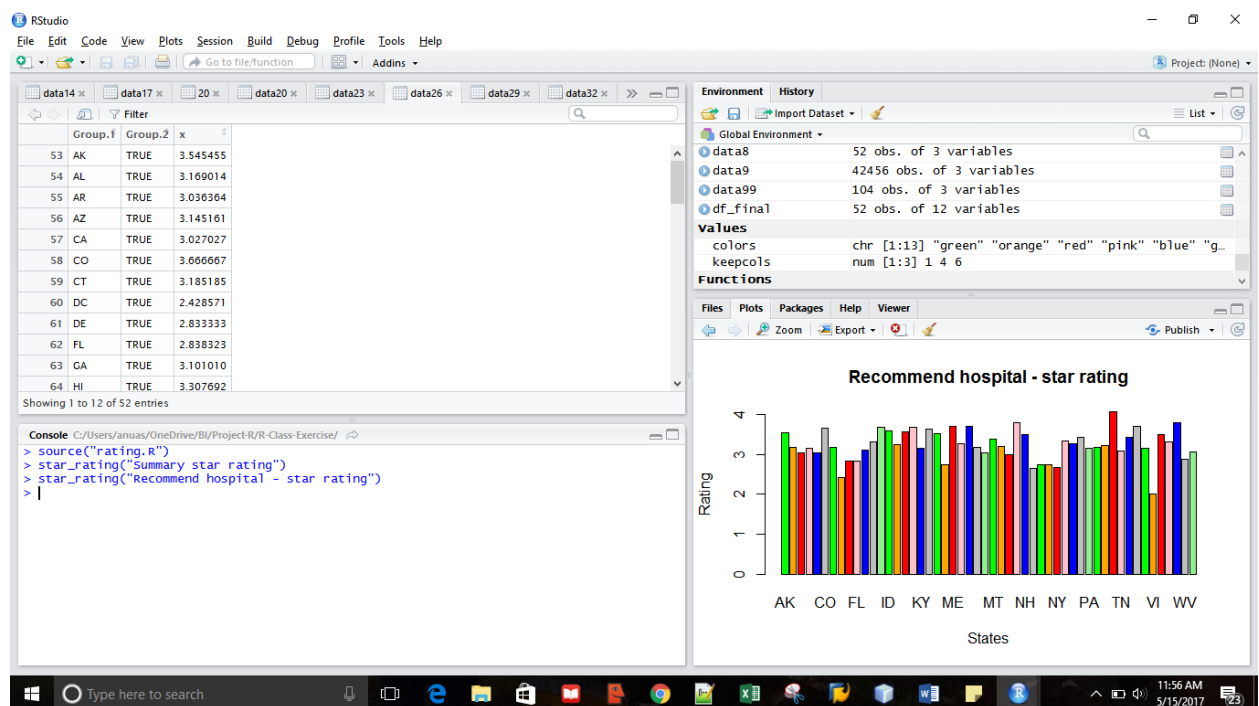


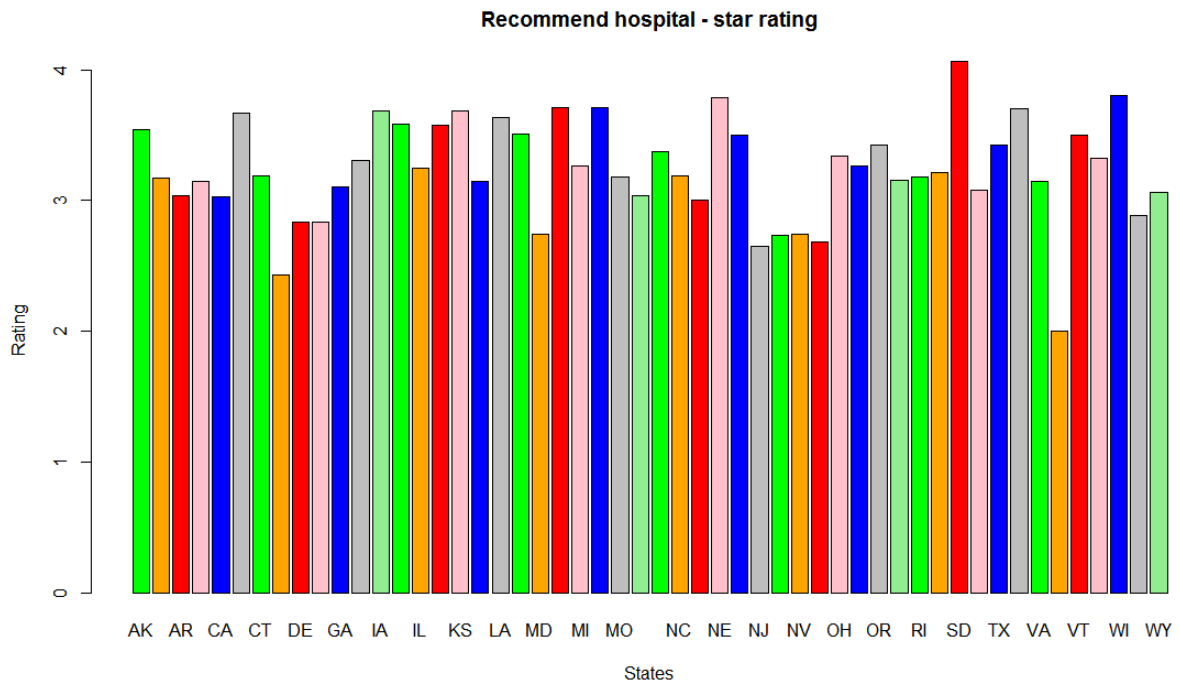
Axis Specifications

<u>X-Axis</u>	<u>Y-Axis</u>
States	HCAHPS Questions star Rating

- Output2

User wants to know the average rating in each state for the question- “Recommend hospital- star rating “. Therefore, on running the function **star_rating** with an input string parameter “**Recommend hospital- star rating** “, the respective dataframe is displayed on the top on the screen and the respective graph is shown on the right of the screen.

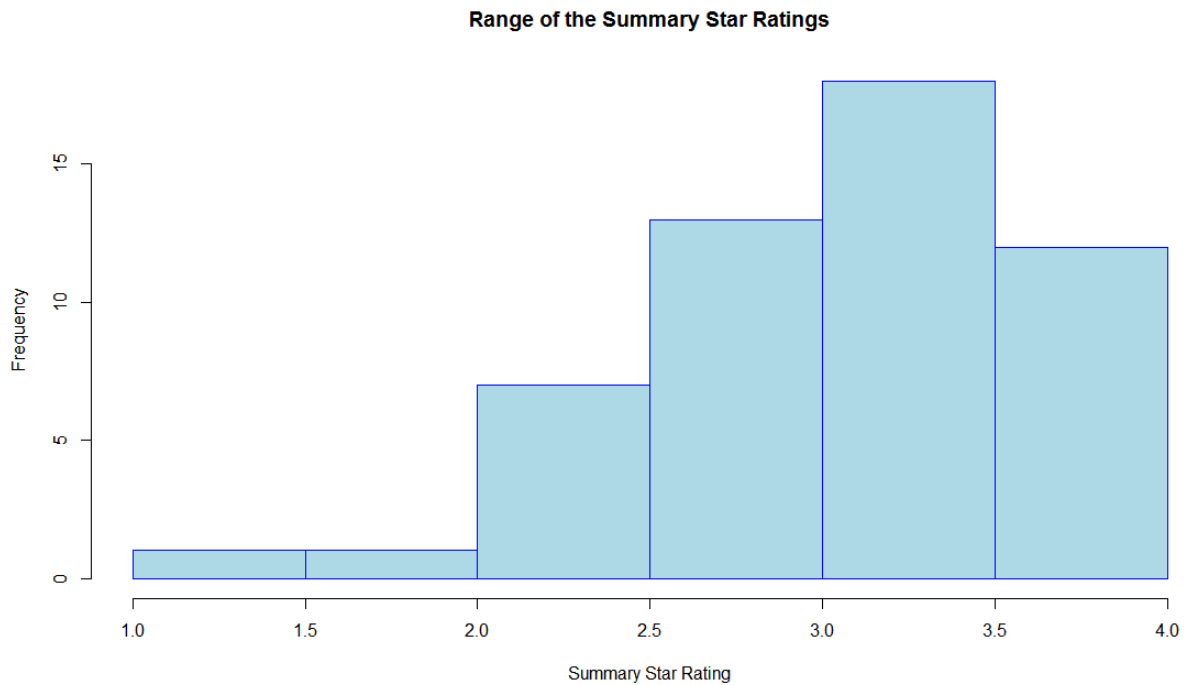
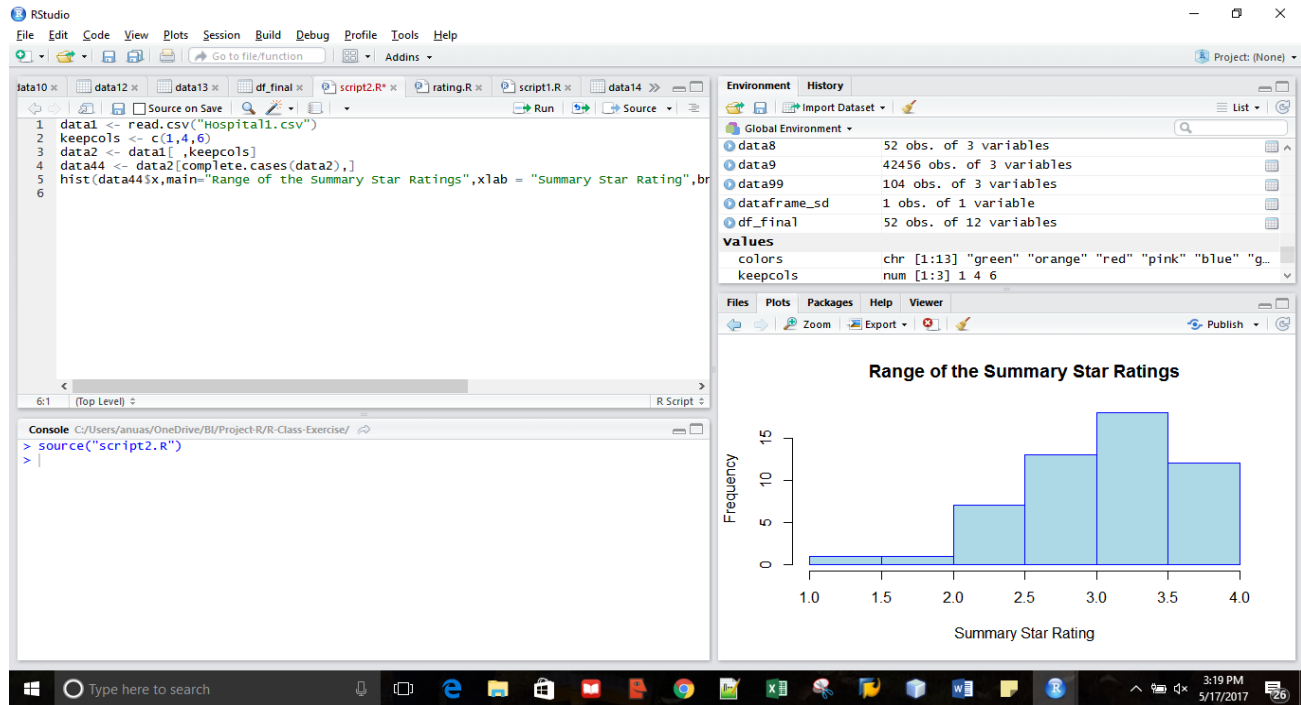




b) Visualization 2: -

Insights: A script was written to see in which range most of the ratings of the “Summary star rating HCAHPS Question” lies considering all the states in which this question was rated. A histogram was plotted to visualize this. It was seen that more than 15 states had a rating of 3-3.5 for this question which was the maximum number of states when compared to all the other ranges and around 10-13 states to be precise, had a good rating of 3.5-4. It can also be seen that only a few states had a low rating of 1-2.

Hospital Consumer Assessment of Healthcare Providers and Systems



Axis Specifications

<u>X-Axis</u>	<u>Y-Axis</u>
Summary Star Rating	Frequency

Codes Related to the Project: -

Question1: -

- Code to create dataframes, calculate mean of overall hospital rating and create a bar chart to visualize the same.

```
> data1 <- read.csv("GI1.csv", header=TRUE)
> keepcols<- c(1,6)
> data2 <- data1[,keepcols]
> count2 <- data.frame(data2$Hospital.overall.rating, data2$State)
> count4 <- count2[complete.cases(count2),]
> count3<-aggregate(count4$data2.Hospital.overall.rating, by=list(count4$data
2.State), FUN=mean)
> count5<- data.frame(count3)
>View(count5)
> barplot(count5$x, xlab= "States", ylab = "summary star rating", names.arg =
count5$Group.1, main="Evaluation of Summary Star Rating across USA", col=c("b
lue","red","green","orange","yellow","violet"))
> box()
```

Question2:-

- Code to create dataframes, calculate mean of the number of completed surveys and the survey response rate and create a scatter plot to represent the relationship between these two parameters calculated.


```
> data1 <- read.csv("Hospital.csv", header = TRUE)
> keepcols <- c(1,9,10)
> data2 <- data1[,keepcols]
> count2 <- data.frame(data2$State, data2$Number.of.Completed.Surveys, data2$
Survey.Response.Rate.Percent)
> count4 <- count2[complete.cases(count2),]
> count3<-aggregate(count4[,2:3], list(count4$data2.State), mean)
> count5<- data.frame(count3)
> View(count5)
>plot(count5$data2.Number.of.Completed.Surveys,count5$data2.Survey.Response.R
ate.Percent, xlim=c(0,1000), ylim=c(0,50),xlab = "Number of completed surveys
", ylab="Survey response rate", main="Mapping survey response across USA")
```

Question3 (Implemented Using Function and script):-

- We have selected a few columns and performed mean on the ratings of the star rating questions state wise and represented the same on a bar graph using a function in visualization1. The code to plot a histogram for the range of ratings has been implemented in the script.

```
> setwd("C:/Users/anuas/OneDrive/BI/Project-R/R-Class-Exercise")
> data1 <- read.csv("Hospital1.csv")
> View(data1)
> keepcols <- c(1,4,6)
> data2 <- data1[,keepcols]
> colors <- c("green","orange","brown",)
```

Function Used: -

- This function takes in the HCAHPS question for which the state wise mean rating is to be displayed in string format and gives a dataframe with all the mean values and the respective visualization in a bar-graph as the output.

Code to call the function: -

```
> source("script2.R")
> star_rating("Quietness - star rating")
> star_rating("Recommend hospital - star rating")
```

Code of the function

```
star_rating<-function(type)

{

  data1 <- read.csv("Hospital1.csv")

  keepcols <- c(1,4,6)

  data2 <- data1[ ,keepcols]

  if(type=="Summary star rating")

  {

    data30 <- data2[complete.cases(data2),]

    data31<-aggregate(data30$Patient.Survey.Star.Rating,

by=list(data30$State,data30$HCAHPS.Question=="Summary star rating"), FUN=mean)

    data32<-data31[which(data31$Group.2==TRUE), ]

    View(data32)

    barplot(data32$x,main = "Summary star rating",names.arg = data44$Group.1,xlab =

"States",ylab = "Rating",col = colors)

  }

  else if(type=="Care transition - star rating")

  {

    data44 <- data2[complete.cases(data2),]

    data33<-aggregate(data44$Patient.Survey.Star.Rating,

by=list(data44$State,data44$HCAHPS.Question=="Care transition - star rating"),

FUN=mean)
```

Hospital Consumer Assessment of Healthcare Providers and Systems

```
data44<-data33[which(data33$Group.2==TRUE), ]

View(data44)

barplot(data44$x,main = "Summary star rating",names.arg = data44$Group.1,xlab =
"States",ylab = "Rating",col = colors)

}

else if(type=="Cleanliness - star rating")

{

data5 <- data2[complete.cases(data2),]

data55<-aggregate(data5$Patient.Survey.Star.Rating,
by=list(data5$State,data5$HCAHPS.Question=="Cleanliness - star rating"), FUN=mean)

data6<-data55[which(data55$Group.2==TRUE), ]

View(data6)

barplot(data6$x,main = "Cleanliness - star rating",names.arg = data44$Group.1,xlab =
"States",ylab = "Rating",col = colors)

}

else if(type=="Communication about medicines - star rating")

{

data7 <- data2[complete.cases(data2),]

data77<-aggregate(data7$Patient.Survey.Star.Rating,
by=list(data7$State,data7$HCAHPS.Question=="Communication about medicines - star
rating"), FUN=mean)

data8<-data77[which(data77$Group.2==TRUE), ]

View(data8)
```

```
    barplot(data8$x,main = "Communication about medicines - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

  }

  else if(type=="Discharge information - star rating")

  {

    data9 <- data2[complete.cases(data2),]

    data99<-aggregate(data9$Patient.Survey.Star.Rating,
by=list(data9$State,data9$HCAHPS.Question=="Discharge information - star rating"),
FUN=mean)

    data10<-data99[which(data99$Group.2==TRUE), ]

    View(data10)

    barplot(data10$x,main = "Discharge information - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

  }

  else if(type=="Doctor communication - star rating")

  {

    data12 <- data2[complete.cases(data2),]

    data13<-aggregate(data12$Patient.Survey.Star.Rating,
by=list(data12$State,data12$HCAHPS.Question=="Doctor communication - star
rating"), FUN=mean)

    data14<-data13[which(data13$Group.2==TRUE), ]

    View(data14)
```

Hospital Consumer Assessment of Healthcare Providers and Systems

```
    barplot(data14$x,main = "Doctor communication - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

  }

  else if(type=="Nurse communication - star rating")

  {

    data15 <- data2[complete.cases(data2),]

    data16<-aggregate(data15$Patient.Survey.Star.Rating,
by=list(data15$State,data15$HCAHPS.Question=="Nurse communication - star rating"),
FUN=mean)

    data17<-data16[which(data16$Group.2==TRUE), ]

    View(data17)

    barplot(data17$x,main = "Nurse communication - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

  }

  else if(type=="Overall hospital rating - star rating")

  {

    data18 <- data2[complete.cases(data2),]

    data19<-aggregate(data18$Patient.Survey.Star.Rating,
by=list(data18$State,data18$HCAHPS.Question=="Overall hospital rating - star rating"),
FUN=mean)

    data20<-data19[which(data19$Group.2==TRUE), ]

    View(data20)
```

Hospital Consumer Assessment of Healthcare Providers and Systems

```
    barplot(data20$x,main = "Overall hospital rating - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

  }

  else if(type=="Quietness - star rating")

  {

    data21 <- data2[complete.cases(data2),]

    data22<-aggregate(data21$Patient.Survey.Star.Rating,
by=list(data21$State,data21$HCAHPS.Question=="Quietness - star rating"),
FUN=mean)

    data23<-data22[which(data22$Group.2==TRUE), ]

    View(data23)

    barplot(data23$x,main = "Quietness - star rating",names.arg = data44$Group.1,xlab =
"States",ylab = "Rating",col = colors)

  }

  else if(type=="Recommend hospital - star rating")

  {

    data24 <- data2[complete.cases(data2),]

    data25<-aggregate(data24$Patient.Survey.Star.Rating,
by=list(data24$State,data24$HCAHPS.Question=="Recommend hospital - star rating"),
FUN=mean)

    data26<-data25[which(data25$Group.2==TRUE), ]

    View(data26)
```


Hospital Consumer Assessment of Healthcare Providers and Systems

```
barplot(data26$x,main = "Recommend hospital - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

}

else if(type=="Staff responsiveness - star rating")

{

  data27 <- data2[complete.cases(data2),]

  data28<-aggregate(data27$Patient.Survey.Star.Rating,
by=list(data27$State,data27$HCAHPS.Question=="Staff responsiveness - star rating"),
FUN=mean)

  data29<-data28[which(data28$Group.2==TRUE), ]

  View(data29)

  barplot(data29$x,main = "Staff responsiveness - star rating",names.arg =
data44$Group.1,xlab = "States",ylab = "Rating",col = colors)

}

else

{

  cat("Enter the proper rating discription")

}

}
```

- Code to display a dataframe with all the mean values of all the question state-wise: -

```
> df_final<-data.frame(data44$Group.1,data44$x,data6$x,data8$x,data10$x
,data14$x,data17$x,data20$x,data23$x,data26$x,data29$x,data32$x)
> View(df_final)
```

Script2.R

- This script code is to display a histogram to show the range of rating values that fall in each rating category.

Code to Run the Script

```
> source("script2.R")
```

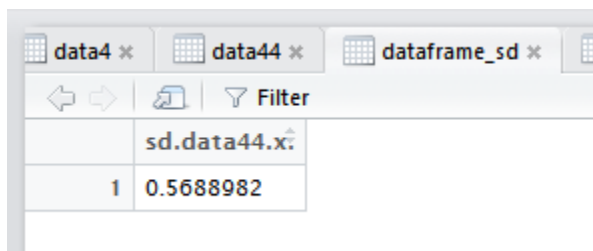
Code written inside the Script

```
data1 <- read.csv("Hospital1.csv")
keepcols <- c(1,4,6)
data2 <- data1[,keepcols]
data44 <- data2[complete.cases(data2),]
hist(data44$x,main="Range of the Summary Star Ratings",xlab = "Summary Star Rating",breaks
=1,col = "light blue",border = "blue")
```

Statistical Functions: -

- Standard Deviation

```
> setwd("C:/Users/anuas/OneDrive/BI/Project-R/R-Class-Exercise")
> data1 <- read.csv("Hospital1.csv")
> View(data1)
> keepcols <- c(1,4,6)
> data2 <- data1[,keepcols]
data44 <- data2[complete.cases(data2),]
> data33<-aggregate(data44$Patient.Survey.Star.Rating, by=list(data44$State,data44$HCAHPS.
Question=="Care transition - star rating"), FUN=mean)
> data44<-data33[which(data33$Group.2==TRUE), ]
> View(data44)
> dataframe_sd<-data.frame(sd(data44$x))
> View(dataframe_sd)
```



	sd.data44.x
1	0.5688982

- Mean

Hospital Consumer Assessment of Healthcare Providers and Systems

```
> setwd("C:/Users/anuas/OneDrive/BI/Project-R/R-Class-Exercise")
> data1 <- read.csv("Hospital1.csv")
> View(data1)
> keepcols <- c(1,4,6)
> data2 <- data1[,keepcols]
data44 <- data2[complete.cases(data2),]
> data33<-aggregate(data44$Patient.Survey.Star.Rating, by=list(data44$State,data44$HCAHPS.
Question=="Care transition - star rating"), FUN=mean)
> data44<-data33[which(data33$Group.2==TRUE), ]
> View(data44)
```

	Group.1	Group.2	x
53	AK	TRUE	3.363636
54	AL	TRUE	2.887324
55	AR	TRUE	2.672727
56	AZ	TRUE	2.709677
57	CA	TRUE	2.472973
58	CO	TRUE	3.607843
59	CT	TRUE	2.851852
60	DC	TRUE	1.714286
61	DE	TRUE	3.166667
62	FL	TRUE	2.425150
63	GA	TRUE	2.838384
64	HI	TRUE	3.230769

