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-syfz

- 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.

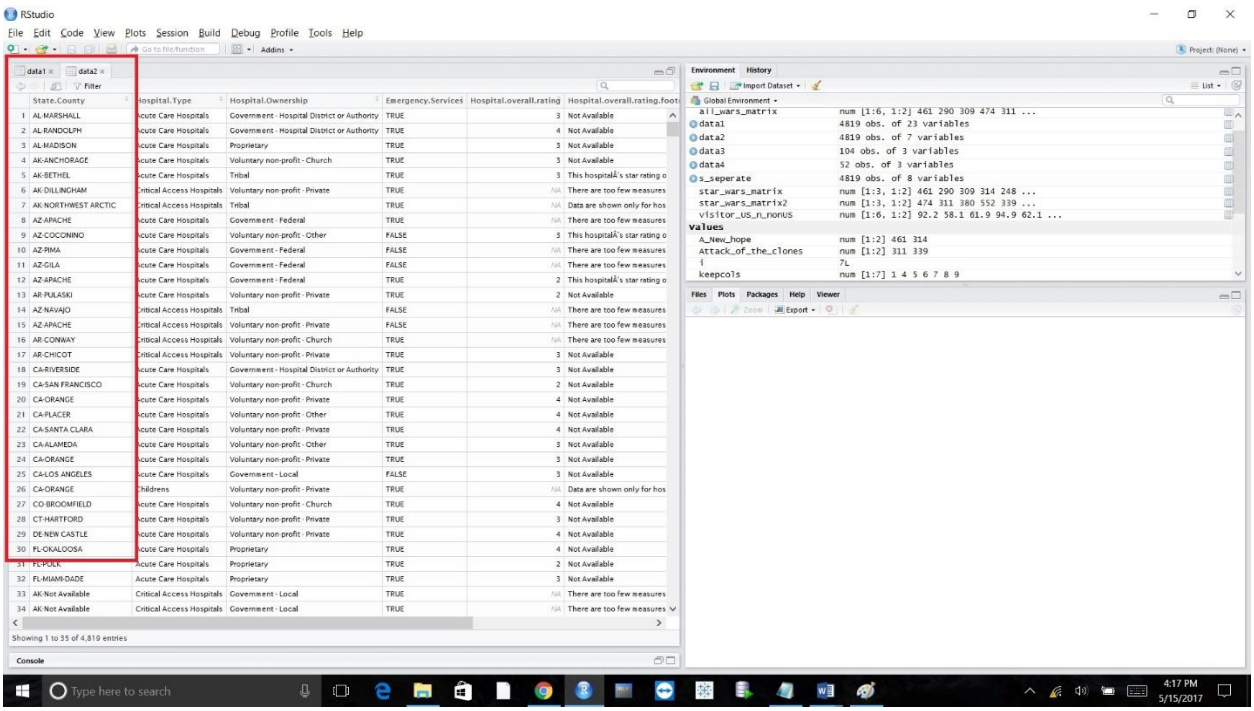
DATA CLEANING

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

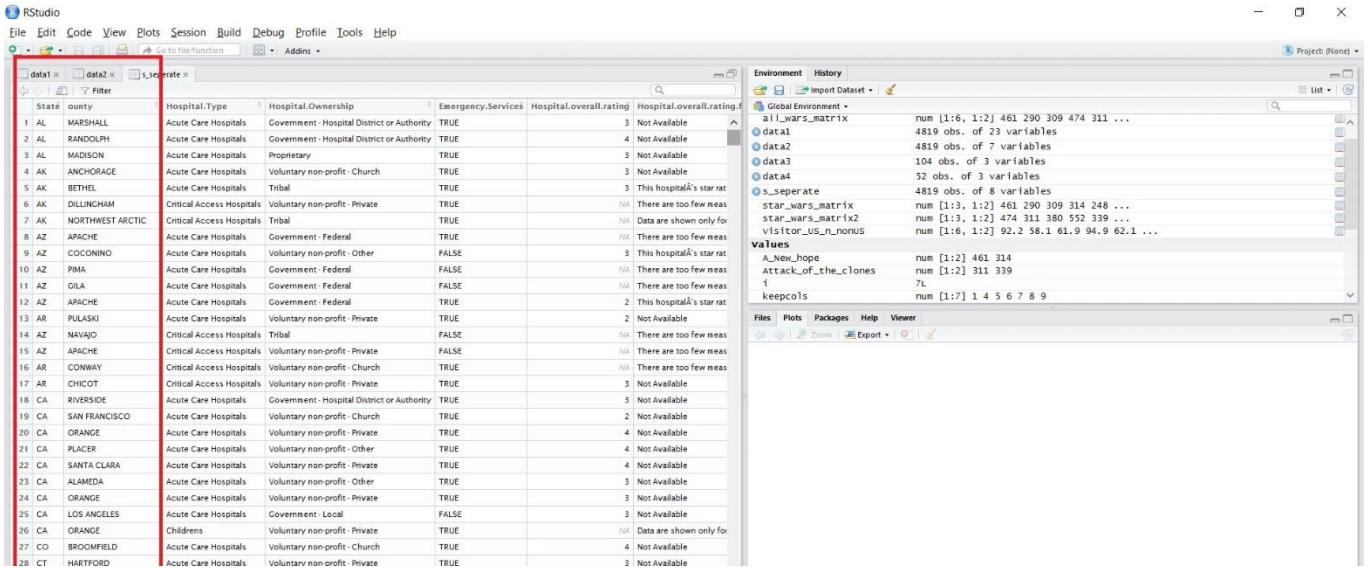


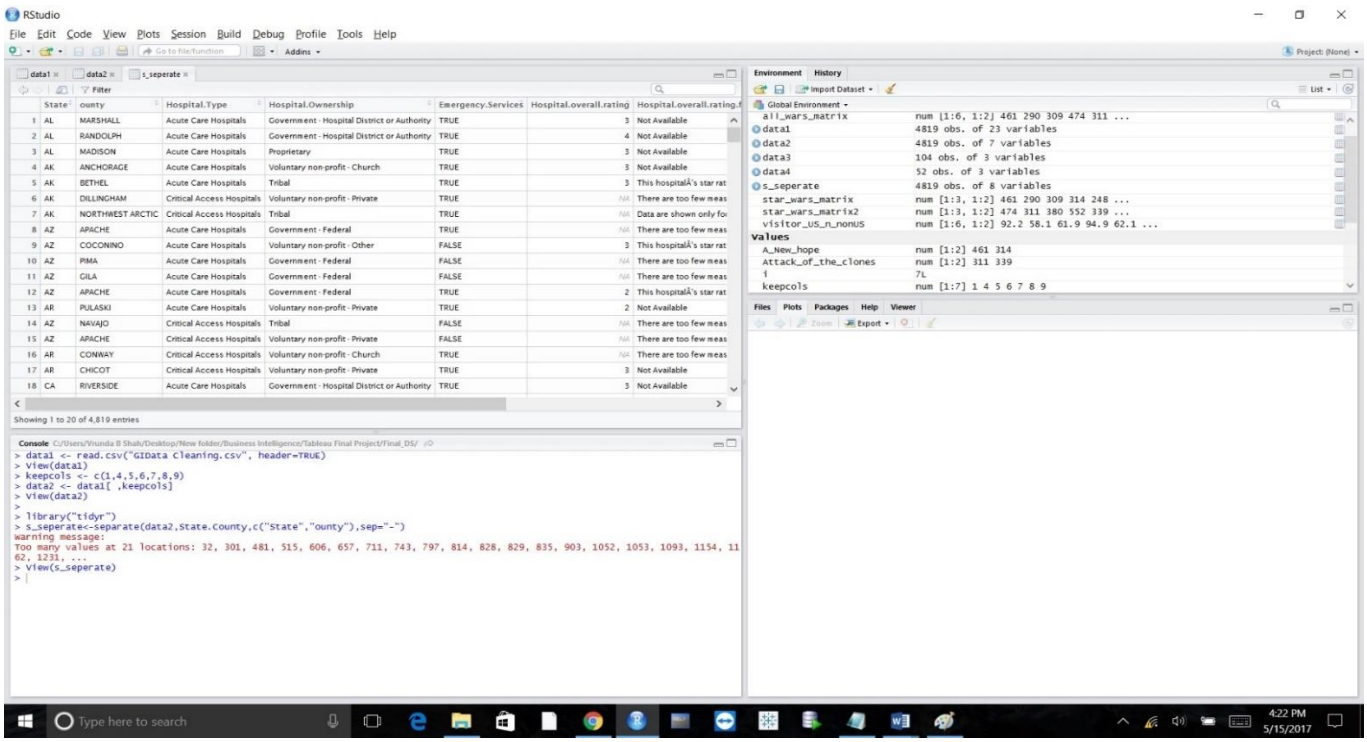
Code for separating merged data via **separate** function

```
> 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_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, 105
```

> View(s_seperate)

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

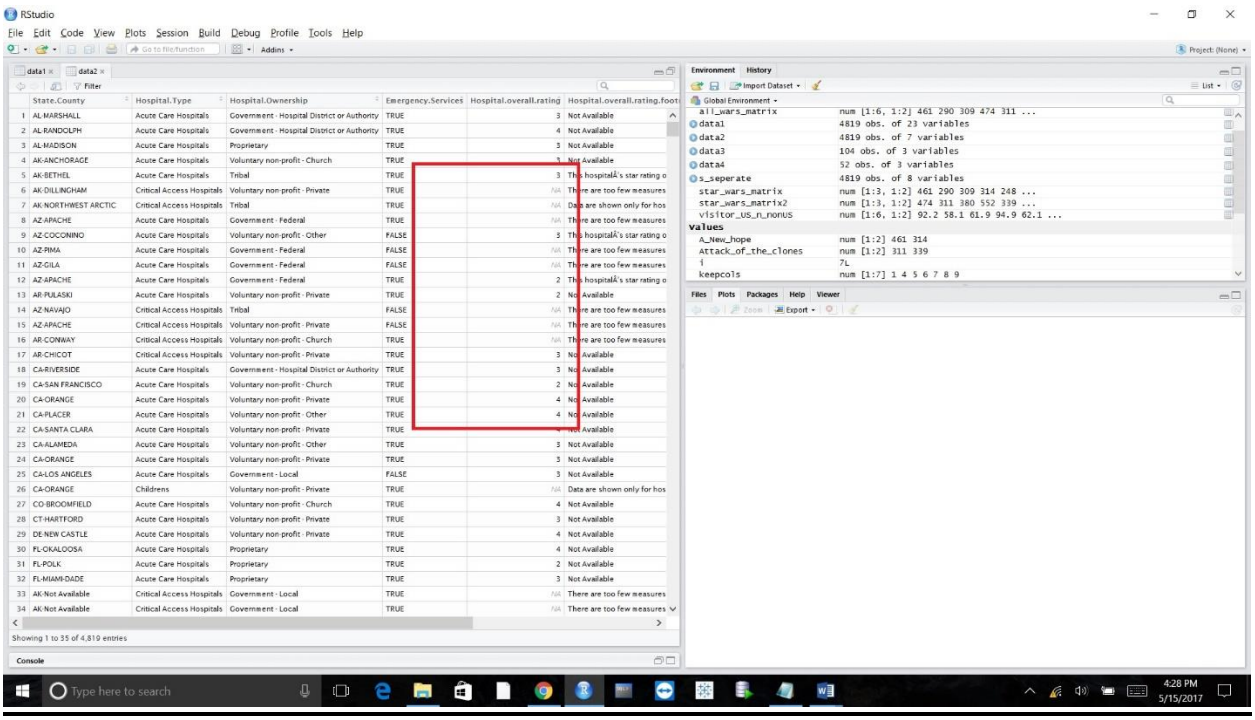




2) Data Cleaning Approach2

Removal of NA values present in the dataset

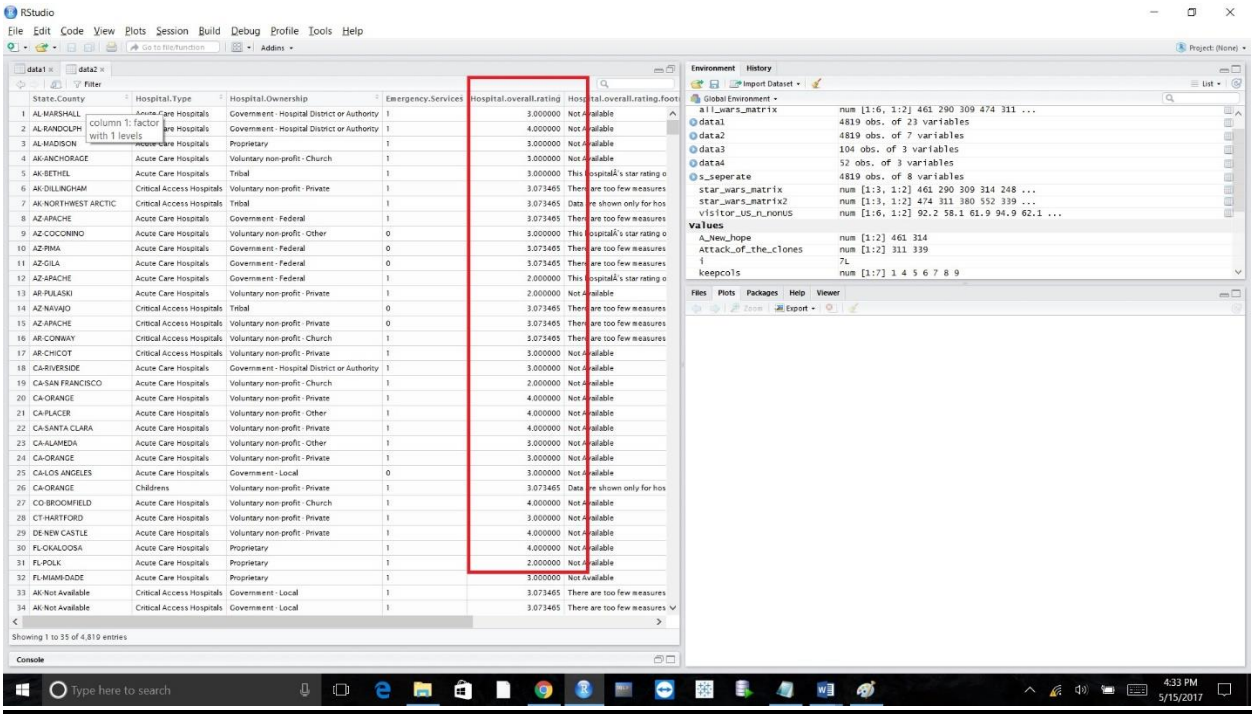
In the dataset shown below we have **NA** values

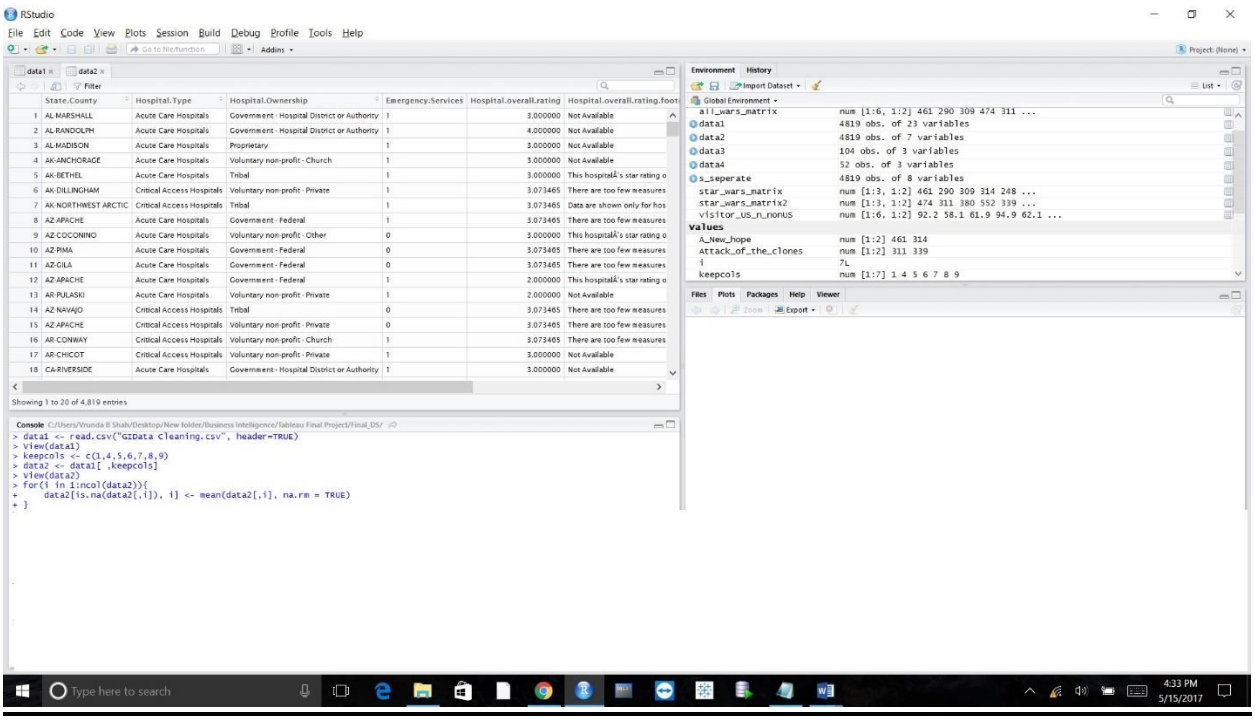


Code for removal of NA values, replacing it with the mean

```
> 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

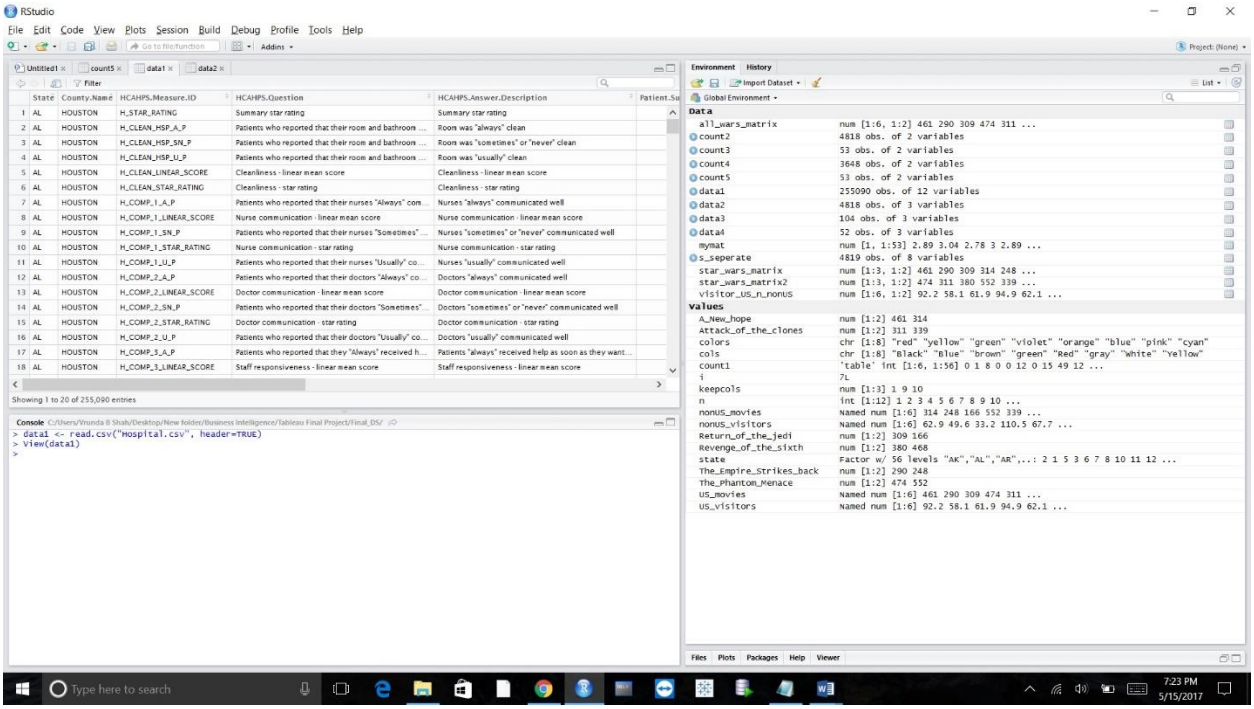




3) Data Cleaning Approach3

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

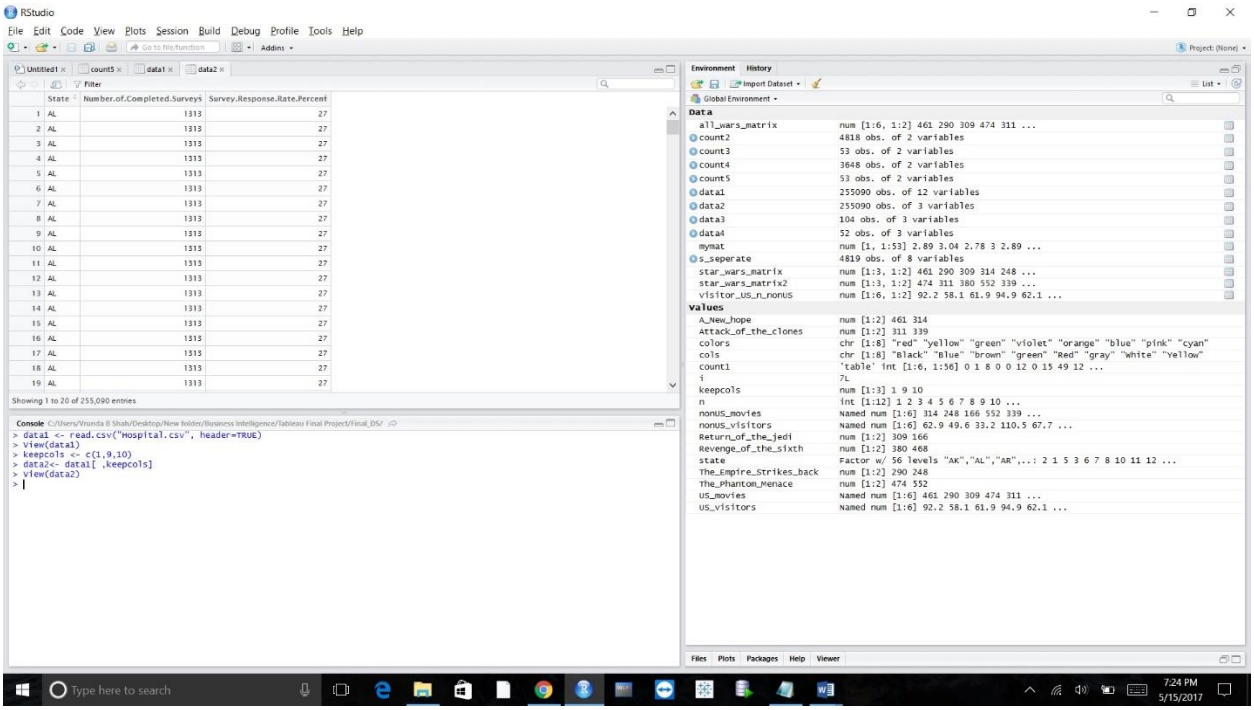
Removal of unwanted columns.



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



DATA VISUALISATION

Question1

Compare hospital overall rating of each state across USA

Reference Sheet: GI1.csv

Elements used: Bar Chart

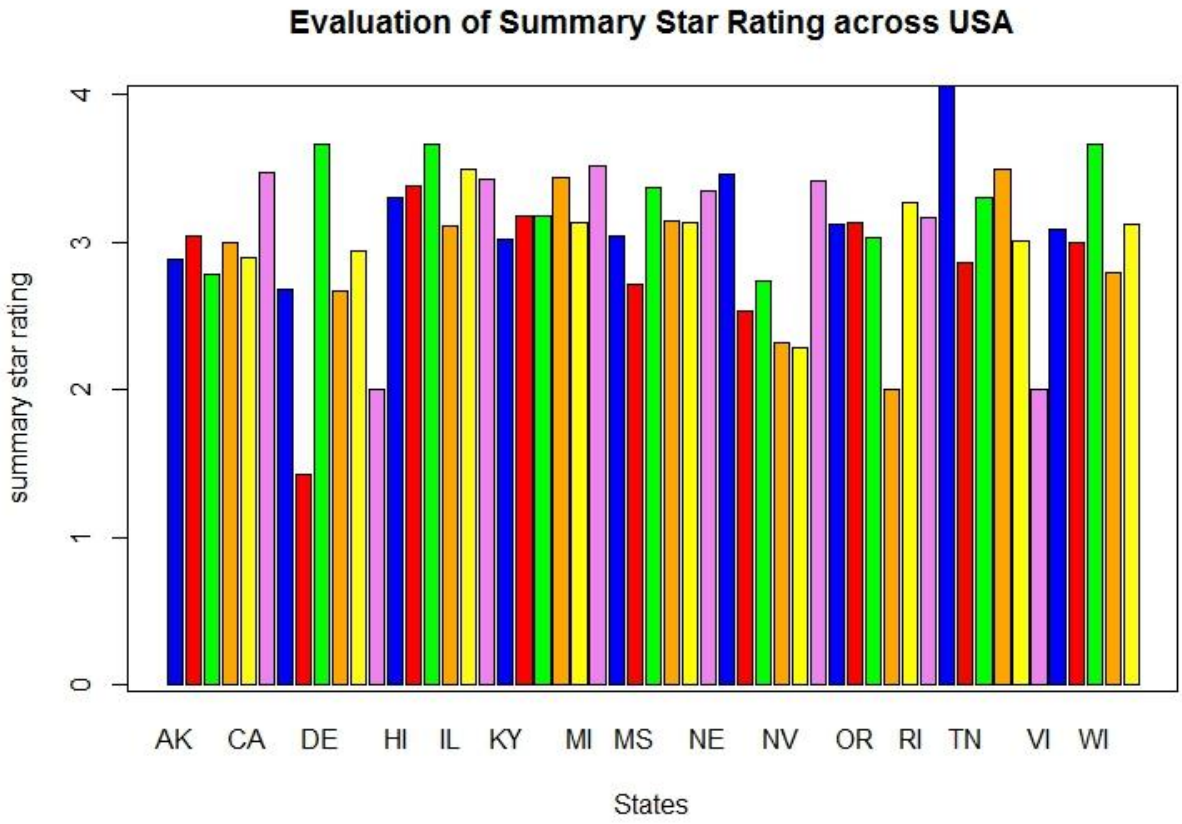
Insights:

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 kin tabular format.

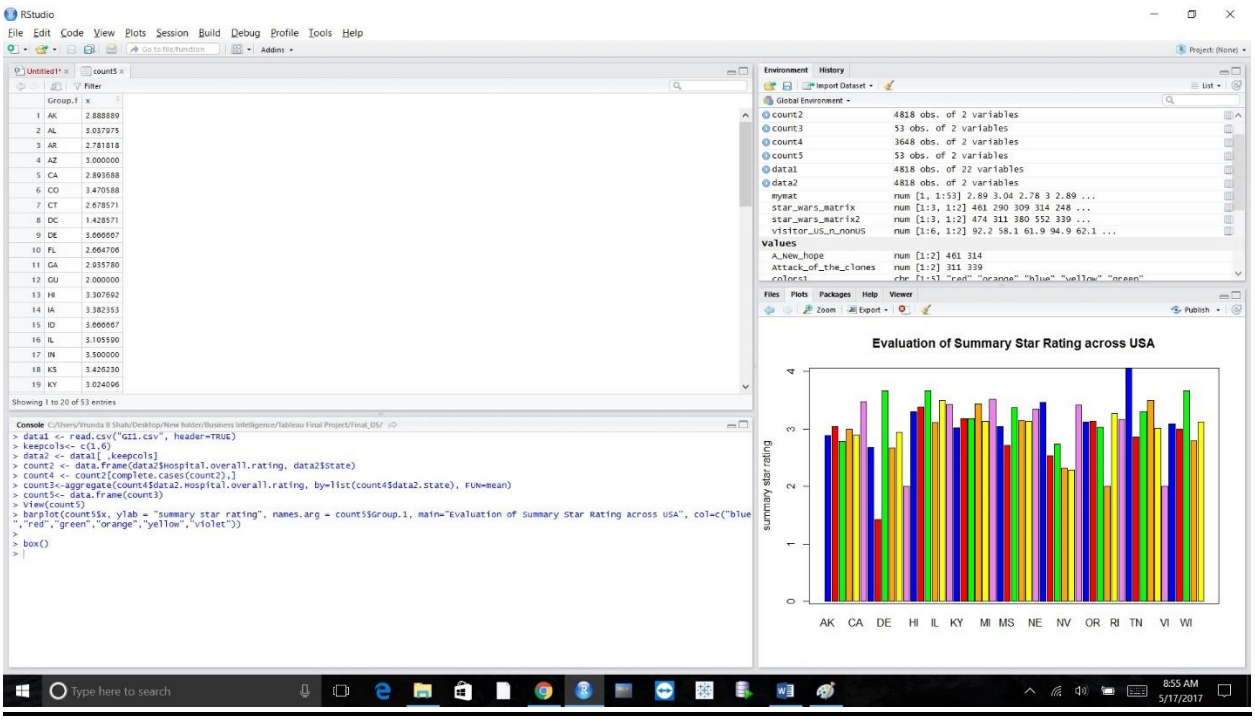
State	Average Rating
SD	4.05
DE	3.66
ID	3.66

WE	3.66
MN	3.51



Axis Specifications

X-Axis	**Y-Axis**
States	Summary Star Rating



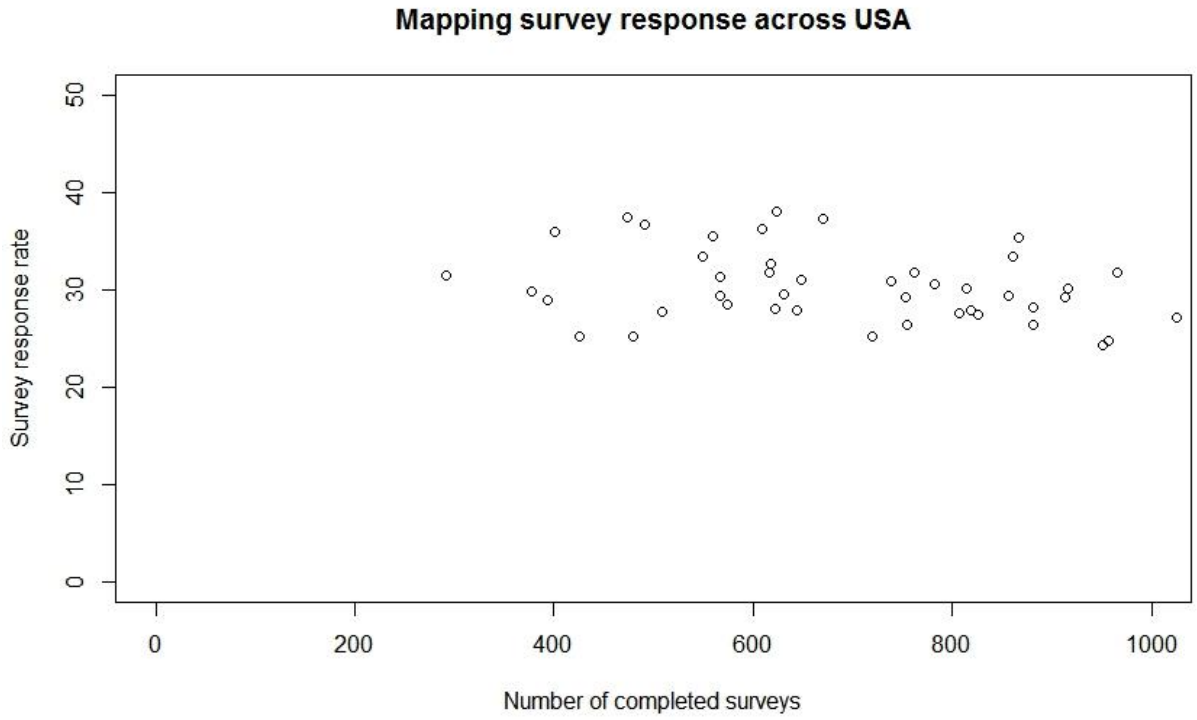
Question2- Find out how many surveys we completed for each state and what was its survey response rate

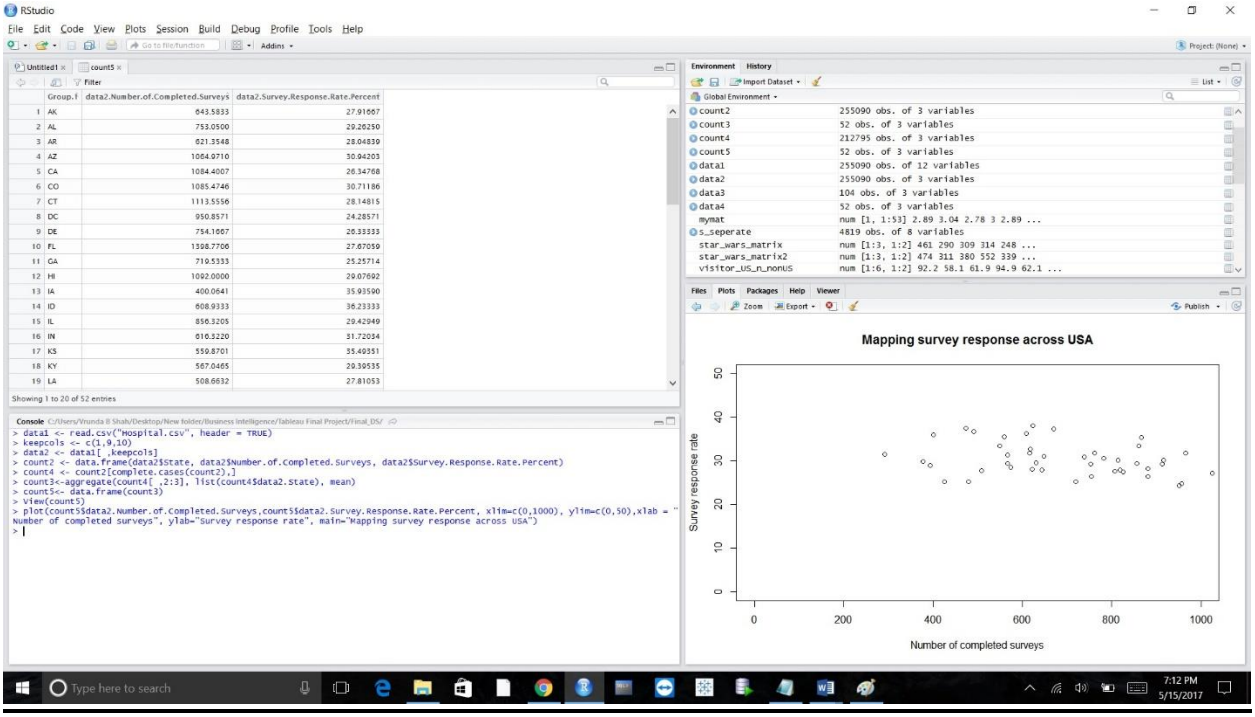
Reference Sheet: Hospital.csv

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
Survey response rate
0
10
20
30
40
50
0
200
400
600
800
1000
Number of completed surveys


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Mapping survey response across USA
Survey response rate
Number of completed surveys
Type here to search
7:12 PM
5/15/2017

Axis Specifications

X-Axis	**Y-Axis**
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

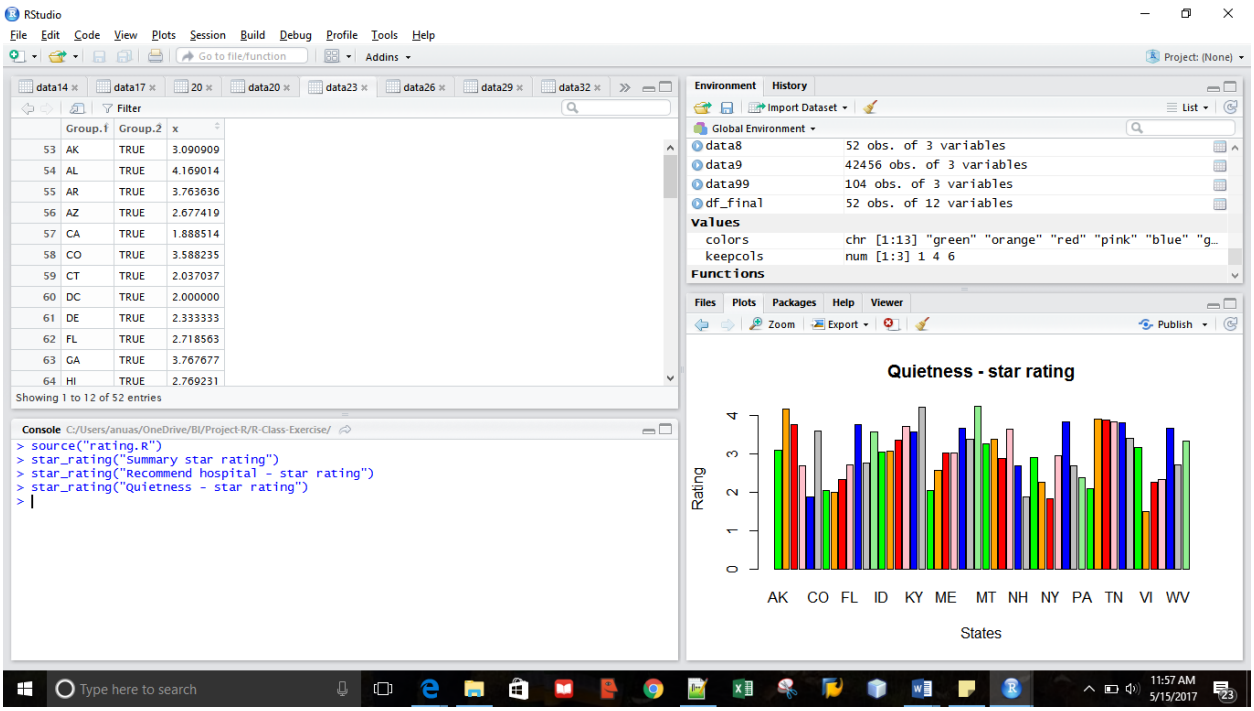
Elements used: Bar graph, Histogram, Function, Script

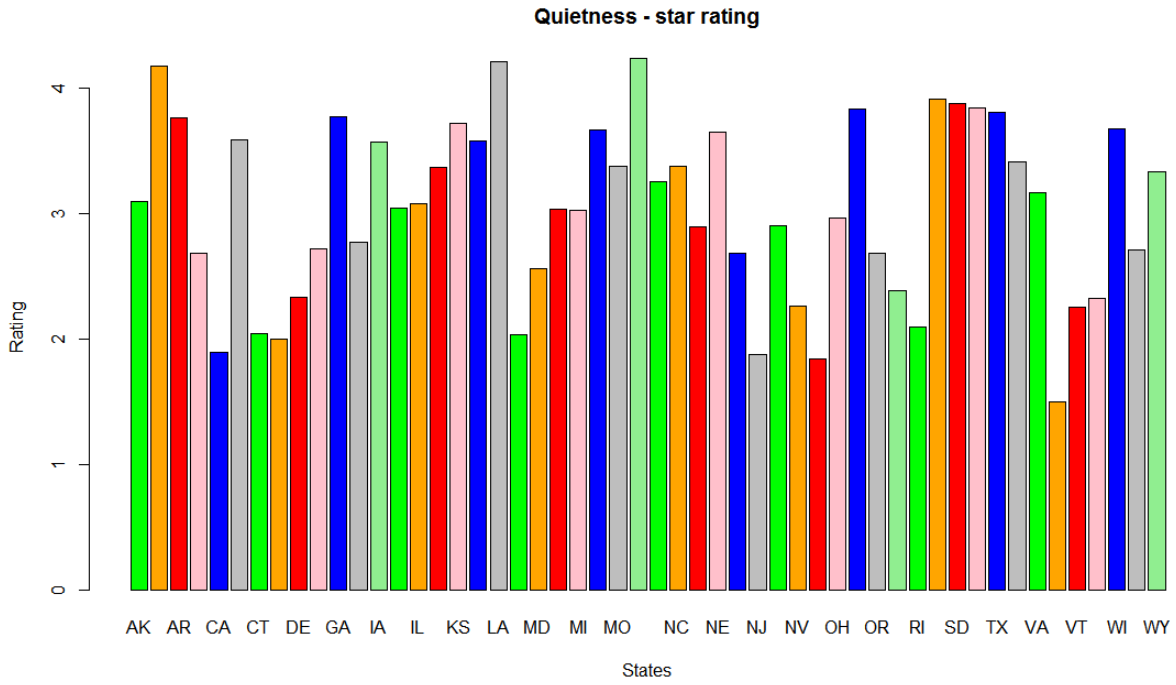
a) **Visualization 1: -**

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.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History
Import Dataset
Global Environment
data8 52 obs. of 3 variables
data9 42456 obs. of 3 variables
data99 104 obs. of 3 variables
df_final 52 obs. of 12 variables
Values
colors chr [1:13] "green" "orange" "red" "pink" "blue" "g...
keepcols num [1:3] 1 4 6
Functions
Showing 1 to 12 of 52 entries
Console C:/Users/anuas/OneDrive/BI/Project-R/R-Class-Exercise/
> source("rating.R")
> star_rating("Summary star rating")
> star_rating("Recommend hospital - star rating")
> star_rating("Quietness - star rating")
Files Plots Packages Help Viewer
Zoom Export Publish
Quietness - star rating
Rating
States
AK CO FL ID KY ME MT NH NY PA TN VI WV
Type here to search
11:57 AM
5/15/2017

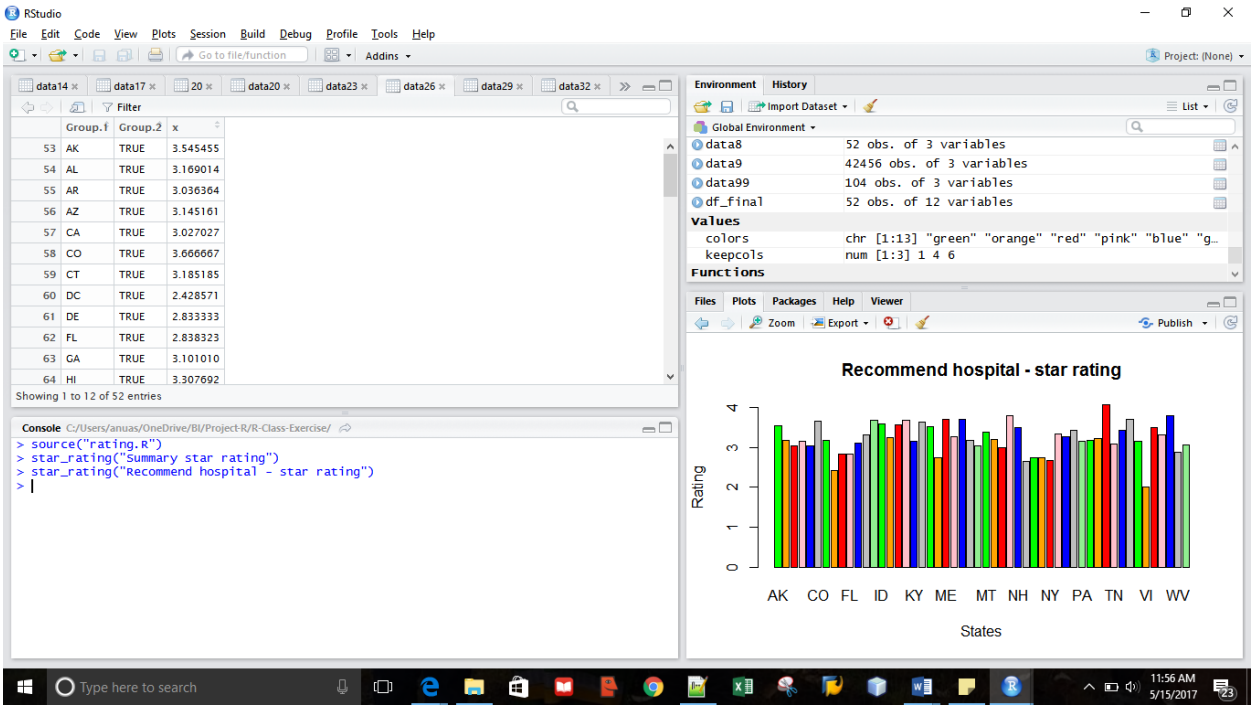

Quietness - star rating
Rating
States
AK AR CA CT DE GA IA IL KS LA MD MI MO NC NE NJ NV OH OR RI SD TX VA VT WI WY

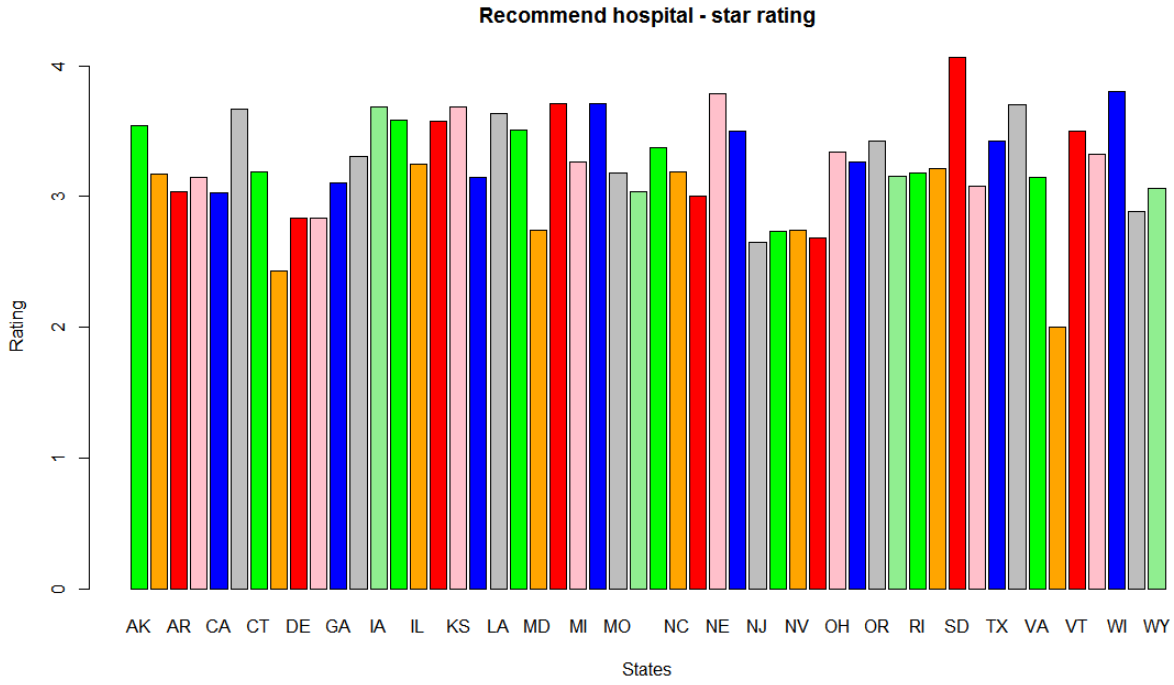
Axis Specifications

X-Axis	**Y-Axis**
States	HCAHPS Questions star Rating

- Output2

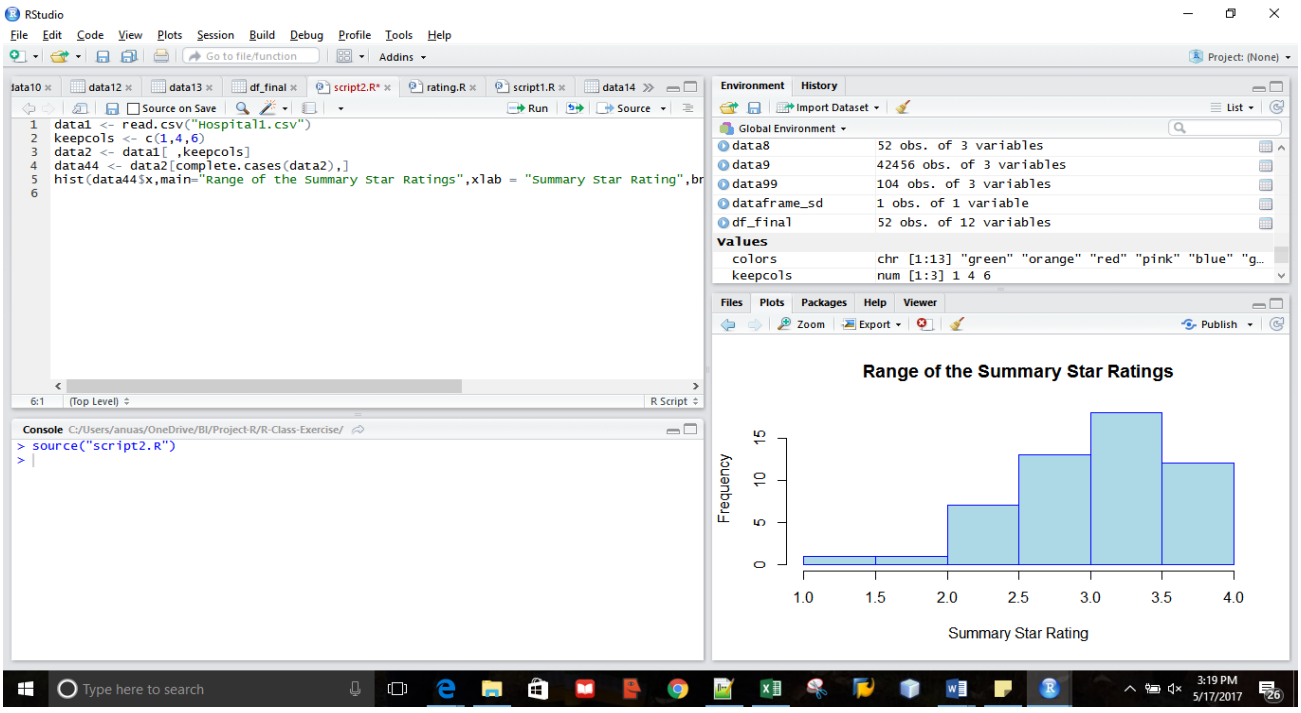
User wants to know the average rating in each state for the question- "Recommend hospital- star rating ". Therefore, on running the function **star_rating** with an input string parameter "**Recommend hospital- star rating** ", the respective dataframe is displayed on the top on the screen and the respective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

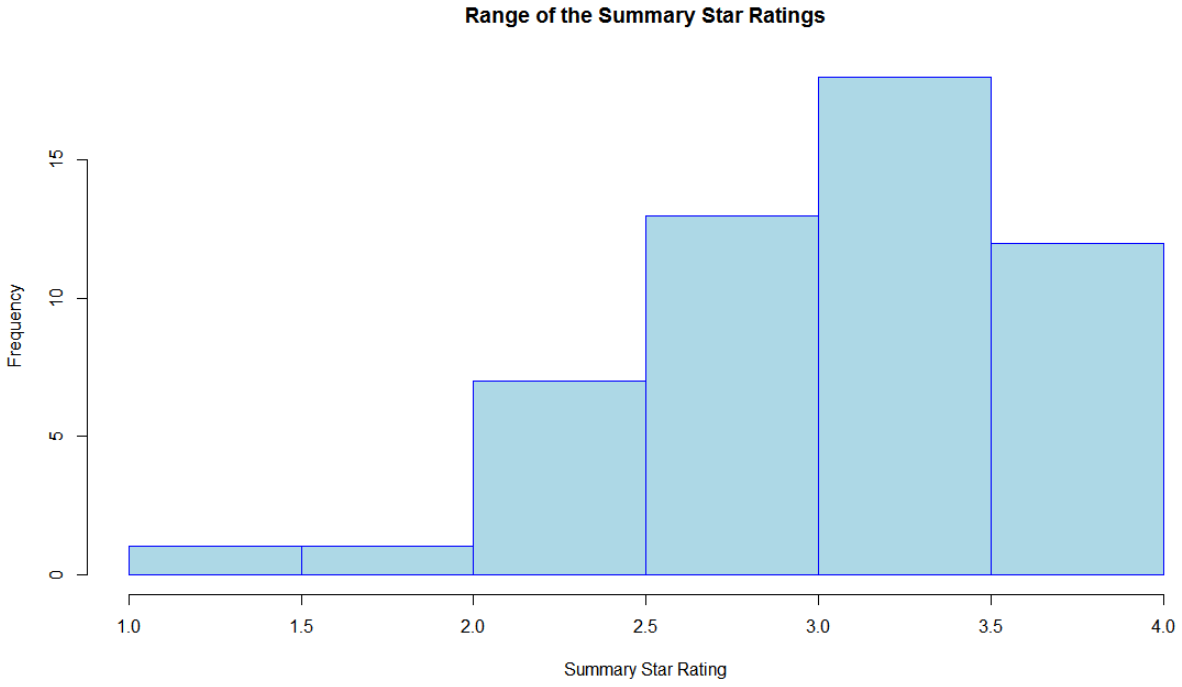




b) <u>Visualization 2: -</u>

<u>Insights</u>: 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.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
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",br
> source("script2.R")
Range of the Summary Star Ratings
Frequency
Summary Star Rating
0
5
10
15
1.0
1.5
2.0
2.5
3.0
3.5
4.0


Range of the Summary Star Ratings
Frequency
Summary Star Rating
0
5
10
15
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Axis Specifications

X-Axis	**Y-Axis**
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)
```

```
    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)
```

```
   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)
```

```
   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)
```

```
    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")
```

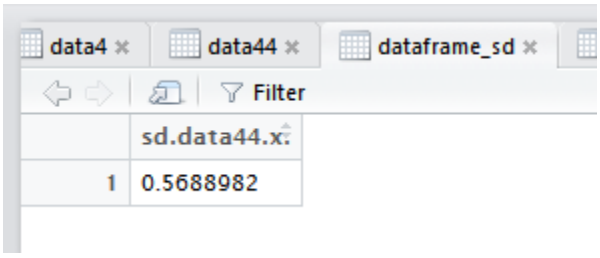
Code written inside the Script

```
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: -

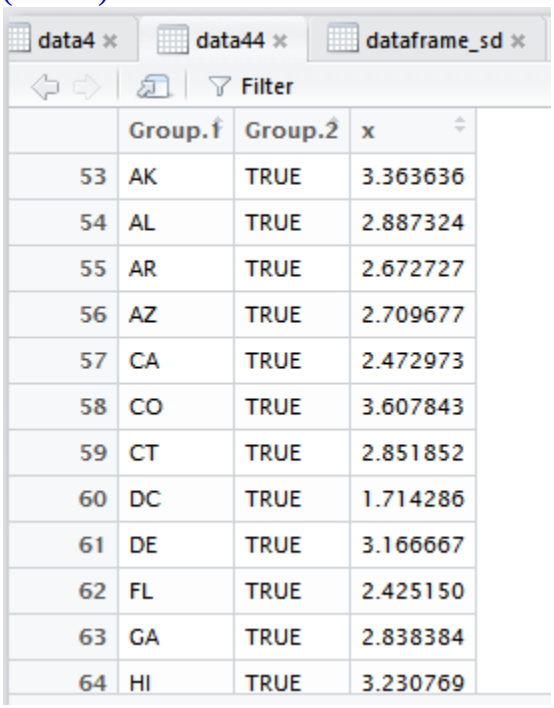
- **Standard Deviation**

```
> 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)
```



- **Mean**

```
> 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)
```



data4 × | data44 × | dataframe_sd ×

Filter

	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