

HAP 725 Case Study

Koustav Roy

Loading the Library

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Declaring Variables

```
V_Provider_ID_Hospital_to_Analyze <- 240010  
V_Provider_ID_Comp1 <- 360180  
V_Provider_ID_Comp2 <- 50510  
V_Provider_ID_Comp3 <- 0  
  
V_Attribute1 <- "H_COMP_1_A_P"  
V_Attribute2 <- "H_COMP_7_LINEAR_SCORE"  
V_Attribute3 <- "H_RECMND_PY"  
V_Attribute4 <- "H_COMP_2_U_P"  
V_Attribute5 <- ""  
  
V_Analysis_starting_Year <- 2015
```

Importing The Dataset

#Reading 2015 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2015/hos
_revised_flatfiles_archive_12_2015.zip", "hos_revised_flatfiles_archive_12_2015.zip")
unzip("hos_revised_flatfiles_archive_12_2015.zip")
df_2015 <- read.csv("HCAHPS - Hospital.csv", na.strings = c("", "NA"))
head(df_2015)
```

##	Provider.ID	Hospital.Name	Address	City	
## 1	10001 SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN		
## 2	10001 SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN		
## 3	10001 SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN		
## 4	10001 SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN		
## 5	10001 SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN		
## 6	10001 SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN		
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	3347938701	H_STAR_RATING
## 2	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_A_P
## 3	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_SN_P
## 4	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_U_P
## 5	AL	36301	HOUSTON	3347938701	H_CLEAN_LINEAR_SCORE
## 6	AL	36301	HOUSTON	3347938701	H_CLEAN_STAR_RATING
##	HCAHPS.Question				
## 1	Summary star rating				
## 2	Patients who reported that their room and bathroom were "Always" clean				
## 3	Patients who reported that their room and bathroom were "Sometimes" or "Never" clean				
## 4	Patients who reported that their room and bathroom were "Usually" clean				
## 5	Cleanliness - linear mean score				
## 6	Cleanliness - star rating				
##	HCAHPS.Answer.Description	Patient.Survey.Star.Rating			
## 1	Summary star rating	3			
## 2	Room was "always" clean	Not Applicable			
## 3	Room was "sometimes" or "never" clean	Not Applicable			
## 4	Room was "usually" clean	Not Applicable			
## 5	Cleanliness - linear mean score	Not Applicable			
## 6	Cleanliness - star rating	2			
##	Patient.Survey.Star.Rating.Footnote	HCAHPS.Answer.Percent			
## 1	<NA>	Not Applicable			
## 2	<NA>	66			
## 3	<NA>	12			
## 4	<NA>	22			
## 5	<NA>	Not Applicable			
## 6	<NA>	Not Applicable			
##	HCAHPS.Answer.Percent.Footnote	HCAHPS.Linear.Mean.Value			
## 1	<NA>	Not Applicable			
## 2	<NA>	Not Applicable			
## 3	<NA>	Not Applicable			
## 4	<NA>	Not Applicable			
## 5	<NA>	84			
## 6	<NA>	Not Applicable			
##	Number.of.Completed.Surveys	Number.of.Completed.Surveys.Footnote			
## 1	1313	<NA>			
## 2	1313	<NA>			
## 3	1313	<NA>			
## 4	1313	<NA>			
## 5	1313	<NA>			
## 6	1313	<NA>			
##	Survey.Response.Rate.Percent	Survey.Response.Rate.Percent.Footnote			
## 1	27	<NA>			
## 2	27	<NA>			

## 3	27	<NA>
## 4	27	<NA>
## 5	27	<NA>
## 6	27	<NA>
##	Measure.Start.Date	Measure.End.Date
## 1	04/01/2014	03/31/2015
## 2	04/01/2014	03/31/2015
## 3	04/01/2014	03/31/2015
## 4	04/01/2014	03/31/2015
## 5	04/01/2014	03/31/2015
## 6	04/01/2014	03/31/2015

#Reading 2016 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2016/hos_revised_flatfiles_archive_12_2016.zip", "hos_revised_flatfiles_archive_12_2016.zip")
unzip("hos_revised_flatfiles_archive_12_2016.zip")
df_2016 <- read.csv("HCAHPS - Hospital.csv", na.strings = c("", "NA"))
head(df_2016)
```

##	Provider.ID	Hospital.Name	Address	City	
## 1	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	3347938701	H_STAR_RATING
## 2	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_A_P
## 3	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_SN_P
## 4	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_U_P
## 5	AL	36301	HOUSTON	3347938701	H_CLEAN_LINEAR_SCORE
## 6	AL	36301	HOUSTON	3347938701	H_CLEAN_STAR_RATING
##	HCAHPS.Question				
## 1	Summary star rating				
## 2	Patients who reported that their room and bathroom were "Always" clean				
## 3	Patients who reported that their room and bathroom were "Sometimes" or "Never" clean				
## 4	Patients who reported that their room and bathroom were "Usually" clean				
## 5	Cleanliness - linear mean score				
## 6	Cleanliness - star rating				
##	HCAHPS.Answer.Description		Patient.Survey.Star.Rating		
## 1	Summary star rating		3		
## 2	Room was "always" clean		Not Applicable		
## 3	Room was "sometimes" or "never" clean		Not Applicable		
## 4	Room was "usually" clean		Not Applicable		
## 5	Cleanliness - linear mean score		Not Applicable		
## 6	Cleanliness - star rating		2		
##	Patient.Survey.Star.Rating.Footnote		HCAHPS.Answer.Percent		
## 1	<NA>		Not Applicable		
## 2	<NA>		65		
## 3	<NA>		12		
## 4	<NA>		23		
## 5	<NA>		Not Applicable		
## 6	<NA>		Not Applicable		
##	HCAHPS.Answer.Percent.Footnote		HCAHPS.Linear.Mean.Value		
## 1	<NA>		Not Applicable		
## 2	<NA>		Not Applicable		
## 3	<NA>		Not Applicable		
## 4	<NA>		Not Applicable		
## 5	<NA>		84		
## 6	<NA>		Not Applicable		
##	Number.of.Completed.Surveys		Number.of.Completed.Surveys.Footnote		
## 1	1213		<NA>		
## 2	1213		<NA>		
## 3	1213		<NA>		
## 4	1213		<NA>		
## 5	1213		<NA>		
## 6	1213		<NA>		
##	Survey.Response.Rate.Percent		Survey.Response.Rate.Percent.Footnote		
## 1	27		<NA>		
## 2	27		<NA>		

## 3	27	<NA>
## 4	27	<NA>
## 5	27	<NA>
## 6	27	<NA>
##	Measure.Start.Date	Measure.End.Date
## 1	04/01/2015	03/31/2016
## 2	04/01/2015	03/31/2016
## 3	04/01/2015	03/31/2016
## 4	04/01/2015	03/31/2016
## 5	04/01/2015	03/31/2016
## 6	04/01/2015	03/31/2016

#Reading 2017 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2017/hos_revised_flatfiles_archive_10_2017.zip", "hos_revised_flatfiles_archive_10_2017.zip")
unzip("hos_revised_flatfiles_archive_10_2017.zip")
df_2017 <- read.csv("HCAHPS - Hospital.csv", na.strings = c("", "NA"))
head(df_2017)
```

##	Provider.ID	Hospital.Name	Address	City	
## 1	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	10001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	3347938701	H_STAR_RATING
## 2	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_A_P
## 3	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_SN_P
## 4	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_U_P
## 5	AL	36301	HOUSTON	3347938701	H_CLEAN_LINEAR_SCORE
## 6	AL	36301	HOUSTON	3347938701	H_CLEAN_STAR_RATING
##	HCAHPS.Question				
## 1	Summary star rating				
## 2	Patients who reported that their room and bathroom were "Always" clean				
## 3	Patients who reported that their room and bathroom were "Sometimes" or "Never" clean				
## 4	Patients who reported that their room and bathroom were "Usually" clean				
## 5	Cleanliness - linear mean score				
## 6	Cleanliness - star rating				
##	HCAHPS.Answer.Description	Patient.Survey.Star.Rating			
## 1	Summary star rating	3			
## 2	Room was "always" clean	Not Applicable			
## 3	Room was "sometimes" or "never" clean	Not Applicable			
## 4	Room was "usually" clean	Not Applicable			
## 5	Cleanliness - linear mean score	Not Applicable			
## 6	Cleanliness - star rating	2			
##	Patient.Survey.Star.Rating.Footnote	HCAHPS.Answer.Percent			
## 1	<NA>	Not Applicable			
## 2	<NA>	67			
## 3	<NA>	12			
## 4	<NA>	21			
## 5	<NA>	Not Applicable			
## 6	<NA>	Not Applicable			
##	HCAHPS.Answer.Percent.Footnote	HCAHPS.Linear.Mean.Value			
## 1	<NA>	Not Applicable			
## 2	<NA>	Not Applicable			
## 3	<NA>	Not Applicable			
## 4	<NA>	Not Applicable			
## 5	<NA>	84			
## 6	<NA>	Not Applicable			
##	Number.of.Completed.Surveys	Number.of.Completed.Surveys.Footnote			
## 1	1332	<NA>			
## 2	1332	<NA>			
## 3	1332	<NA>			
## 4	1332	<NA>			
## 5	1332	<NA>			
## 6	1332	<NA>			
##	Survey.Response.Rate.Percent	Survey.Response.Rate.Percent.Footnote			
## 1	25	<NA>			
## 2	25	<NA>			

## 3	25	<NA>
## 4	25	<NA>
## 5	25	<NA>
## 6	25	<NA>
##	Measure.Start.Date	Measure.End.Date
## 1	01/01/2016	12/31/2016
## 2	01/01/2016	12/31/2016
## 3	01/01/2016	12/31/2016
## 4	01/01/2016	12/31/2016
## 5	01/01/2016	12/31/2016
## 6	01/01/2016	12/31/2016

#Reading 2018 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2018/hos_revised_flatfiles_archive_10_2018.zip", "hos_revised_flatfiles_archive_10_2018.zip")
unzip("hos_revised_flatfiles_archive_10_2018.zip")
df_2018 <- read.csv("HCAHPS - Hospital.csv", na.strings = c("", "NA"))
head(df_2018)
```

##	Provider.ID	Hospital.Name	Address	City	
## 1	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	3347938701	H_STAR_RATING
## 2	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_A_P
## 3	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_SN_P
## 4	AL	36301	HOUSTON	3347938701	H_CLEAN_HSP_U_P
## 5	AL	36301	HOUSTON	3347938701	H_CLEAN_LINEAR_SCORE
## 6	AL	36301	HOUSTON	3347938701	H_CLEAN_STAR_RATING
##	HCAHPS.Question				
## 1	Summary star rating				
## 2	Patients who reported that their room and bathroom were "Always" clean				
## 3	Patients who reported that their room and bathroom were "Sometimes" or "Never" clean				
## 4	Patients who reported that their room and bathroom were "Usually" clean				
## 5	Cleanliness - linear mean score				
## 6	Cleanliness - star rating				
##	HCAHPS.Answer.Description		Patient.Survey.Star.Rating		
## 1	Summary star rating		2		
## 2	Room was "always" clean		Not Applicable		
## 3	Room was "sometimes" or "never" clean		Not Applicable		
## 4	Room was "usually" clean		Not Applicable		
## 5	Cleanliness - linear mean score		Not Applicable		
## 6	Cleanliness - star rating		2		
##	Patient.Survey.Star.Rating.Footnote		HCAHPS.Answer.Percent		
## 1	<NA>		Not Applicable		
## 2	<NA>		65		
## 3	<NA>		13		
## 4	<NA>		22		
## 5	<NA>		Not Applicable		
## 6	<NA>		Not Applicable		
##	HCAHPS.Answer.Percent.Footnote		HCAHPS.Linear.Mean.Value		
## 1	<NA>		Not Applicable		
## 2	<NA>		Not Applicable		
## 3	<NA>		Not Applicable		
## 4	<NA>		Not Applicable		
## 5	<NA>		83		
## 6	<NA>		Not Applicable		
##	Number.of.Completed.Surveys		Number.of.Completed.Surveys.Footnote		
## 1	463		<NA>		
## 2	463		<NA>		
## 3	463		<NA>		
## 4	463		<NA>		
## 5	463		<NA>		
## 6	463		<NA>		
##	Survey.Response.Rate.Percent		Survey.Response.Rate.Percent.Footnote		
## 1	20		<NA>		
## 2	20		<NA>		

## 3	20	<NA>
## 4	20	<NA>
## 5	20	<NA>
## 6	20	<NA>
##	Measure.Start.Date	Measure.End.Date
## 1	01/01/2017	12/31/2017
## 2	01/01/2017	12/31/2017
## 3	01/01/2017	12/31/2017
## 4	01/01/2017	12/31/2017
## 5	01/01/2017	12/31/2017
## 6	01/01/2017	12/31/2017

#Reading 2019 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2019/hos_revised_flatfiles_archive_10_2019.zip", "hos_revised_flatfiles_archive_10_2019.zip")
unzip("hos_revised_flatfiles_archive_10_2019.zip")
df_2019 <- read.csv("HCAHPS - Hospital.csv", na.strings = c("", "NA"))
head(df_2019)
```

##	Facility.ID	Facility.Name	Address	City	
## 1	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_A_P
## 2	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_SN_P
## 3	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_U_P
## 4	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_LINEAR_SCORE
## 5	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_STAR_RATING
## 6	AL	36301	HOUSTON	(334) 793-8701	H_NURSE_RESPECT_A_P
##	HCAHPS.Question				
## 1	Patients who reported that their nurses "Always" communicated well				
## 2	Patients who reported that their nurses "Sometimes" or "Never" communicated well				
## 3	Patients who reported that their nurses "Usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Patients who reported that their nurses "Always" treated them with courtesy and respect				
##	HCAHPS.Answer.Description				
## 1	Nurses "always" communicated well				
## 2	Nurses "sometimes" or "never" communicated well				
## 3	Nurses "usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Nurses "always" treated them with courtesy and respect				
##	Patient.Survey.Star.Rating	Patient.Survey.Star.Rating	Footnote		
## 1	Not Applicable		<NA>		
## 2	Not Applicable		<NA>		
## 3	Not Applicable		<NA>		
## 4	Not Applicable		<NA>		
## 5	2		<NA>		
## 6	Not Applicable		<NA>		
##	HCAHPS.Answer.Percent	HCAHPS.Answer.Percent	Footnote	HCAHPS.Linear.Mean.Value	
## 1	74		<NA>	Not Applicable	
## 2	8		<NA>	Not Applicable	
## 3	18		<NA>	Not Applicable	
## 4	Not Applicable		<NA>	88	
## 5	Not Applicable		<NA>	Not Applicable	
## 6	79		<NA>	Not Applicable	
##	Number.of.Completed.Surveys	Number.of.Completed.Surveys	Footnote		
## 1	558		<NA>		
## 2	558		<NA>		
## 3	558		<NA>		
## 4	558		<NA>		
## 5	558		<NA>		
## 6	558		<NA>		
##	Survey.Response.Rate.Percent	Survey.Response.Rate.Percent	Footnote	Start.Date	
## 1	23		<NA>	01/01/2018	
## 2	23		<NA>	01/01/2018	

```
## 3      23      <NA> 01/01/2018
## 4      23      <NA> 01/01/2018
## 5      23      <NA> 01/01/2018
## 6      23      <NA> 01/01/2018
##      End.Date
## 1 12/31/2018
## 2 12/31/2018
## 3 12/31/2018
## 4 12/31/2018
## 5 12/31/2018
## 6 12/31/2018
```

#Reading 2020 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2020/hospitals_archive_10_2020.zip", "hospitals_archive_10_2020.zip")
unzip("hospitals_archive_10_2020.zip")
df_2020 <- read.csv("dgck-syfz.csv", na.strings = c("", "NA"))
head(df_2020)
```

##	Facility.ID	Facility.Name	Address	City	
## 1	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	010001	SOUTHEAST ALABAMA MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_A_P
## 2	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_SN_P
## 3	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_U_P
## 4	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_LINEAR_SCORE
## 5	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_STAR_RATING
## 6	AL	36301	HOUSTON	(334) 793-8701	H_NURSE_RESPECT_A_P
##	HCAHPS.Question				
## 1	Patients who reported that their nurses "Always" communicated well				
## 2	Patients who reported that their nurses "Sometimes" or "Never" communicated well				
## 3	Patients who reported that their nurses "Usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Patients who reported that their nurses "Always" treated them with courtesy and respect				
##	HCAHPS.Answer.Description				
## 1	Nurses "always" communicated well				
## 2	Nurses "sometimes" or "never" communicated well				
## 3	Nurses "usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Nurses "always" treated them with courtesy and respect				
##	Patient.Survey.Star.Rating	Patient.Survey.Star.Rating	Footnote		
## 1	Not Applicable		<NA>		
## 2	Not Applicable		<NA>		
## 3	Not Applicable		<NA>		
## 4	Not Applicable		<NA>		
## 5	3		<NA>		
## 6	Not Applicable		<NA>		
##	HCAHPS.Answer.Percent	HCAHPS.Answer.Percent	Footnote	HCAHPS.Linear.Mean.Value	
## 1	77		<NA>	Not Applicable	
## 2	7		<NA>	Not Applicable	
## 3	16		<NA>	Not Applicable	
## 4	Not Applicable		<NA>	90	
## 5	Not Applicable		<NA>	Not Applicable	
## 6	84		<NA>	Not Applicable	
##	Number.of.Completed.Surveys	Number.of.Completed.Surveys	Footnote		
## 1	507		<NA>		
## 2	507		<NA>		
## 3	507		<NA>		
## 4	507		<NA>		
## 5	507		<NA>		
## 6	507		<NA>		
##	Survey.Response.Rate.Percent	Survey.Response.Rate.Percent	Footnote	Start.Date	
## 1	21		<NA>	01/01/2019	
## 2	21		<NA>	01/01/2019	

```
## 3      21      <NA> 01/01/2019
## 4      21      <NA> 01/01/2019
## 5      21      <NA> 01/01/2019
## 6      21      <NA> 01/01/2019
##      End.Date
## 1 12/31/2019
## 2 12/31/2019
## 3 12/31/2019
## 4 12/31/2019
## 5 12/31/2019
## 6 12/31/2019
```

#Reading 2021 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2021/hospitals_10_2021.zip", "hospitals_10_2021.zip")
unzip("hospitals_10_2021.zip")
df_2021 <- read.csv("HCAHPS-Hospital.csv", na.strings = c("", "NA"))
head(df_2021)
```

##	Facility.ID	Facility.Name	Address	City	
## 1	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	Houston	(334) 793-8701	H_COMP_1_A_P
## 2	AL	36301	Houston	(334) 793-8701	H_COMP_1_SN_P
## 3	AL	36301	Houston	(334) 793-8701	H_COMP_1_U_P
## 4	AL	36301	Houston	(334) 793-8701	H_COMP_1_LINEAR_SCORE
## 5	AL	36301	Houston	(334) 793-8701	H_COMP_1_STAR_RATING
## 6	AL	36301	Houston	(334) 793-8701	H_NURSE_RESPECT_A_P
##	HCAHPS.Question				
## 1	Patients who reported that their nurses "Always" communicated well				
## 2	Patients who reported that their nurses "Sometimes" or "Never" communicated well				
## 3	Patients who reported that their nurses "Usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Patients who reported that their nurses "Always" treated them with courtesy and respect				
##	HCAHPS.Answer.Description				
## 1	Nurses "always" communicated well				
## 2	Nurses "sometimes" or "never" communicated well				
## 3	Nurses "usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Nurses "always" treated them with courtesy and respect				
##	Patient.Survey.Star.Rating	Patient.Survey.Star.Rating	Footnote		
## 1	Not Applicable		<NA>		
## 2	Not Applicable		<NA>		
## 3	Not Applicable		<NA>		
## 4	Not Applicable		<NA>		
## 5	3		<NA>		
## 6	Not Applicable		<NA>		
##	HCAHPS.Answer.Percent	HCAHPS.Answer.Percent	Footnote	HCAHPS.Linear.Mean.Value	
## 1	77		<NA>	Not Applicable	
## 2	7		<NA>	Not Applicable	
## 3	16		<NA>	Not Applicable	
## 4	Not Applicable		<NA>	90	
## 5	Not Applicable		<NA>	Not Applicable	
## 6	81		<NA>	Not Applicable	
##	Number.of.Completed.Surveys	Number.of.Completed.Surveys	Footnote		
## 1	306		<NA>		
## 2	306		<NA>		
## 3	306		<NA>		
## 4	306		<NA>		
## 5	306		<NA>		
## 6	306		<NA>		
##	Survey.Response.Rate.Percent	Survey.Response.Rate.Percent	Footnote	Start.Date	
## 1	25		<NA>	07/01/2020	
## 2	25		<NA>	07/01/2020	

```
## 3                25                <NA> 07/01/2020
## 4                25                <NA> 07/01/2020
## 5                25                <NA> 07/01/2020
## 6                25                <NA> 07/01/2020
##      End.Date
## 1 12/31/2020
## 2 12/31/2020
## 3 12/31/2020
## 4 12/31/2020
## 5 12/31/2020
## 6 12/31/2020
```

#Reading 2022 Data

```
download.file("https://data.cms.gov/provider-data/sites/default/files/archive/Hospitals/2022/hospitals_10_2022.zip", "hospitals_10_22_zip")
unzip("hospitals_10_22_zip")
df_2022 <- read.csv("HCAHPS-Hospital.csv", na.strings = c("", "NA"))
head(df_2022)
```

##	Facility.ID	Facility.Name	Address	City	
## 1	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 2	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 3	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 4	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 5	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
## 6	010001	SOUTHEAST HEALTH MEDICAL CENTER	1108 ROSS CLARK CIRCLE	DOTHAN	
##	State	ZIP.Code	County.Name	Phone.Number	HCAHPS.Measure.ID
## 1	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_A_P
## 2	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_SN_P
## 3	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_U_P
## 4	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_LINEAR_SCORE
## 5	AL	36301	HOUSTON	(334) 793-8701	H_COMP_1_STAR_RATING
## 6	AL	36301	HOUSTON	(334) 793-8701	H_NURSE_RESPECT_A_P
##	HCAHPS.Question				
## 1	Patients who reported that their nurses "Always" communicated well				
## 2	Patients who reported that their nurses "Sometimes" or "Never" communicated well				
## 3	Patients who reported that their nurses "Usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Patients who reported that their nurses "Always" treated them with courtesy and respect				
##	HCAHPS.Answer.Description				
## 1	Nurses "always" communicated well				
## 2	Nurses "sometimes" or "never" communicated well				
## 3	Nurses "usually" communicated well				
## 4	Nurse communication - linear mean score				
## 5	Nurse communication - star rating				
## 6	Nurses "always" treated them with courtesy and respect				
##	Patient.Survey.Star.Rating	Patient.Survey.Star.Rating	Footnote		
## 1	Not Applicable		<NA>		
## 2	Not Applicable		<NA>		
## 3	Not Applicable		<NA>		
## 4	Not Applicable		<NA>		
## 5	3		<NA>		
## 6	Not Applicable		<NA>		
##	HCAHPS.Answer.Percent	HCAHPS.Answer.Percent	Footnote	HCAHPS.Linear.Mean.Value	
## 1	77		<NA>	Not Applicable	
## 2	7		<NA>	Not Applicable	
## 3	16		<NA>	Not Applicable	
## 4	Not Applicable		<NA>	90	
## 5	Not Applicable		<NA>	Not Applicable	
## 6	82		<NA>	Not Applicable	
##	Number.of.Completed.Surveys	Number.of.Completed.Surveys	Footnote		
## 1	421		<NA>		
## 2	421		<NA>		
## 3	421		<NA>		
## 4	421		<NA>		
## 5	421		<NA>		
## 6	421		<NA>		
##	Survey.Response.Rate.Percent	Survey.Response.Rate.Percent	Footnote	Start.Date	
## 1	16		<NA>	01/01/2021	
## 2	16		<NA>	01/01/2021	

```
## 3 16 <NA> 01/01/2021
## 4 16 <NA> 01/01/2021
## 5 16 <NA> 01/01/2021
## 6 16 <NA> 01/01/2021
## End.Date
## 1 12/31/2021
## 2 12/31/2021
## 3 12/31/2021
## 4 12/31/2021
## 5 12/31/2021
## 6 12/31/2021
```

#From 2019 onwards file, the first column name is Facility ID

```
names(df_2019)[1] <- 'Provider.ID'
names(df_2020)[1] <- 'Provider.ID'
names(df_2021)[1] <- 'Provider.ID'
names(df_2022)[1] <- 'Provider.ID'
```

#From 2019 onwards file, the second column name is Facility Name

```
names(df_2019)[2] <- 'Hospital.Name'
names(df_2020)[2] <- 'Hospital.Name'
names(df_2021)[2] <- 'Hospital.Name'
names(df_2022)[2] <- 'Hospital.Name'
```

From 2018 onwards, the Provider ID is imported as Character, there might be some junk value
df_2018\$Provider.ID = as.integer(df_2018\$Provider.ID)

```
## Warning: NAs introduced by coercion
```

```
df_2019$Provider.ID = as.integer(df_2019$Provider.ID)
```

```
## Warning: NAs introduced by coercion
```

```
df_2020$Provider.ID = as.integer(df_2020$Provider.ID)
```

```
## Warning: NAs introduced by coercion
```

```
df_2021$Provider.ID = as.integer(df_2021$Provider.ID)
```

```
## Warning: NAs introduced by coercion
```

```
df_2022$Provider.ID = as.integer(df_2022$Provider.ID)
```

```
## Warning: NAs introduced by coercion
```

Creating Functions

Creating Function for Data Filtering

```
data_filtering <- function(df) {

  df %>%
    filter( Provider.ID == V_Provider_ID_Hospital_to_Analyze | Provider.ID == V_Provider_ID_Comp
1 | Provider.ID == V_Provider_ID_Comp2 | Provider.ID == V_Provider_ID_Comp3, HCAHPS.Measure.ID =
= V_Attribute1 | HCAHPS.Measure.ID == V_Attribute2 | HCAHPS.Measure.ID == V_Attribute3 | HCAHPS.
Measure.ID == V_Attribute4 | HCAHPS.Measure.ID == V_Attribute5) %>%
    select(Provider.ID,Hospital.Name,HCAHPS.Measure.ID,HCAHPS.Question,HCAHPS.Answer.Percent,Pat
ient.Survey.Star.Rating,HCAHPS.Linear.Mean.Value)

}
```

Creating Function for Data Manipulation

```
data_manipulation <- function(df,field) {

  df %>%
    filter( HCAHPS.Measure.ID == field) %>%
    mutate(UCL = mean(HCAHPS.Answer.Percent,na.rm=T)+ 1.96* sd(HCAHPS.Answer.Percent,na.rm=T),
           LCL = mean(HCAHPS.Answer.Percent,na.rm=T)- 1.96* sd(HCAHPS.Answer.Percent,na.rm=T)
           ) %>%
    filter(Provider.ID == V_Provider_ID_Hospital_to_Analyze)

}
```

Creating Function for Data Plotting

```
data_plotting <- function(df){

  df %>% ggplot(aes(x=Year)) +
    geom_line(aes(y=HCAHPS.Answer.Percent,group = Provider.ID),color="blue",size=1.)+
    geom_point(aes(y=HCAHPS.Answer.Percent),color="blue",size=2)+
    geom_line(aes(y=UCL,group = Provider.ID),color="red",size=1.2)+
    geom_line(aes(y=LCL,group = Provider.ID),color="red",size=1.2) +
    labs(y="Value") + ggtitle(paste(df$HCAHPS.Measure.ID, "\n",df$HCAHPS.Question))

}
```

Data Filtering

```
#Callinf the data_filter function and creating one new column Year
df_2015_filter <- data_filtering(df_2015) %>% mutate(Year = 2015)
df_2016_filter <- data_filtering(df_2016) %>% mutate(Year = 2016)
df_2017_filter <- data_filtering(df_2017) %>% mutate(Year = 2017)
df_2018_filter <- data_filtering(df_2018) %>% mutate(Year = 2018)
df_2019_filter <- data_filtering(df_2019) %>% mutate(Year = 2019)
df_2020_filter <- data_filtering(df_2020) %>% mutate(Year = 2020)
df_2021_filter <- data_filtering(df_2021) %>% mutate(Year = 2021)
df_2022_filter <- data_filtering(df_2022) %>% mutate(Year = 2022)

df_2015_filter
```

##	Provider.ID	Hospital.Name	HCAHPS.Measure.ID
## 1	50510	KAISER FOUNDATION HOSPITAL	H_COMP_1_A_P
## 2	50510	KAISER FOUNDATION HOSPITAL	H_COMP_2_U_P
## 3	50510	KAISER FOUNDATION HOSPITAL	H_COMP_7_LINEAR_SCORE
## 4	50510	KAISER FOUNDATION HOSPITAL	H_RECMND_PY
## 5	240010	MAYO CLINIC HOSPITAL ROCHESTER	H_COMP_1_A_P
## 6	240010	MAYO CLINIC HOSPITAL ROCHESTER	H_COMP_2_U_P
## 7	240010	MAYO CLINIC HOSPITAL ROCHESTER	H_COMP_7_LINEAR_SCORE
## 8	240010	MAYO CLINIC HOSPITAL ROCHESTER	H_RECMND_PY
## 9	360180	CLEVELAND CLINIC	H_COMP_1_A_P
## 10	360180	CLEVELAND CLINIC	H_COMP_2_U_P
## 11	360180	CLEVELAND CLINIC	H_COMP_7_LINEAR_SCORE
## 12	360180	CLEVELAND CLINIC	H_RECMND_PY
##			HCAHPS.Question
## 1			Patients who reported that their nurses "Always" communicated well
## 2			Patients who reported that their doctors "Usually" communicated well
## 3			Care transition - linear mean score
## 4			Patients who reported YES, they would probably recommend the hospital
## 5			Patients who reported that their nurses "Always" communicated well
## 6			Patients who reported that their doctors "Usually" communicated well
## 7			Care transition - linear mean score
## 8			Patients who reported YES, they would probably recommend the hospital
## 9			Patients who reported that their nurses "Always" communicated well
## 10			Patients who reported that their doctors "Usually" communicated well
## 11			Care transition - linear mean score
## 12			Patients who reported YES, they would probably recommend the hospital
##	HCAHPS.Answer.Percent	Patient.Survey.Star.Rating	HCAHPS.Linear.Mean.Value
## 1	78	Not Applicable	Not Applicable
## 2	12	Not Applicable	Not Applicable
## 3	Not Applicable	Not Applicable	83
## 4	18	Not Applicable	Not Applicable
## 5	83	Not Applicable	Not Applicable
## 6	14	Not Applicable	Not Applicable
## 7	Not Applicable	Not Applicable	87
## 8	10	Not Applicable	Not Applicable
## 9	83	Not Applicable	Not Applicable
## 10	15	Not Applicable	Not Applicable
## 11	Not Applicable	Not Applicable	85
## 12	12	Not Applicable	Not Applicable
##	Year		
## 1	2015		
## 2	2015		
## 3	2015		
## 4	2015		
## 5	2015		
## 6	2015		
## 7	2015		
## 8	2015		
## 9	2015		
## 10	2015		
## 11	2015		
## 12	2015		

Data Manipulation

```
# Combining all the filtered dataframe
df_concat <- rbind(df_2015_filter,df_2016_filter,df_2017_filter,df_2018_filter,df_2019_filter,df_2020_filter,df_2021_filter,df_2022_filter) %>%
  filter(Year >= V_Analysis_starting_Year )

#Converting Year to Character
df_concat$Year <- as.character(df_concat$Year)

# Changing Datatype of 3 fields to numeric, as we have to calculate mean and SD
df_concat$HCAHPS.Answer.Percent <- as.numeric(df_concat$HCAHPS.Answer.Percent)
```

```
## Warning: NAs introduced by coercion
```

```
df_concat$Patient.Survey.Star.Rating <- as.numeric(df_concat$Patient.Survey.Star.Rating)
```

```
## Warning: NAs introduced by coercion
```

```
df_concat$HCAHPS.Linear.Mean.Value <- as.numeric(df_concat$HCAHPS.Linear.Mean.Value)
```

```
## Warning: NAs introduced by coercion
```

```
# Selecting the not null value amongst the 3 fields
df_concat <- df_concat %>%
  mutate(HCAHPS.Answer.Percent = coalesce(HCAHPS.Answer.Percent, Patient.Survey.Star.Rating, HCAHPS.Linear.Mean.Value))
```

```
#Calling the data_manipulation function
final_data_v1<- data_manipulation(df_concat,V_Attribute1)
final_data_v2<- data_manipulation(df_concat,V_Attribute2)
final_data_v3<- data_manipulation(df_concat,V_Attribute3)
final_data_v4<- data_manipulation(df_concat,V_Attribute4)
final_data_v5<- data_manipulation(df_concat,V_Attribute5)

final_data_v1
```

```
## Provider.ID Hospital.Name HCAHPS.Measure.ID
## 1 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 2 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 3 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 4 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 5 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 6 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 7 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## 8 240010 MAYO CLINIC HOSPITAL ROCHESTER H_COMP_1_A_P
## HCAHPS.Question
## 1 Patients who reported that their nurses "Always" communicated well
## 2 Patients who reported that their nurses "Always" communicated well
## 3 Patients who reported that their nurses "Always" communicated well
## 4 Patients who reported that their nurses "Always" communicated well
## 5 Patients who reported that their nurses "Always" communicated well
## 6 Patients who reported that their nurses "Always" communicated well
## 7 Patients who reported that their nurses "Always" communicated well
## 8 Patients who reported that their nurses "Always" communicated well
## HCAHPS.Answer.Percent Patient.Survey.Star.Rating HCAHPS.Linear.Mean.Value
## 1 83 NA NA
## 2 83 NA NA
## 3 83 NA NA
## 4 83 NA NA
## 5 84 NA NA
## 6 88 NA NA
## 7 88 NA NA
## 8 86 NA NA
## Year UCL LCL
## 1 2015 87.9548 75.5452
## 2 2016 87.9548 75.5452
## 3 2017 87.9548 75.5452
## 4 2018 87.9548 75.5452
## 5 2019 87.9548 75.5452
## 6 2020 87.9548 75.5452
## 7 2021 87.9548 75.5452
## 8 2022 87.9548 75.5452
```

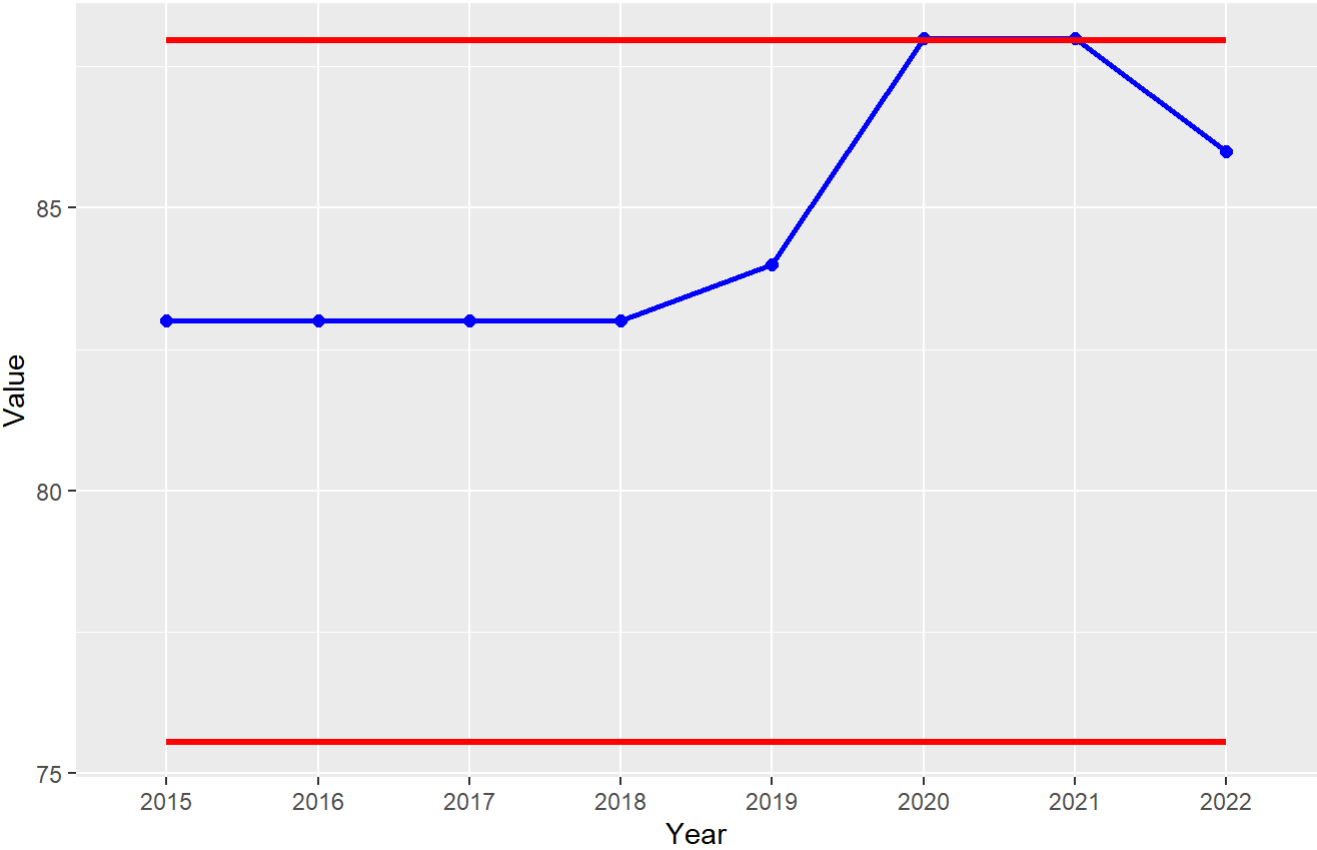
Data Plotting

```
#Calling the data_plotting function
data_plotting(final_data_v1)
```

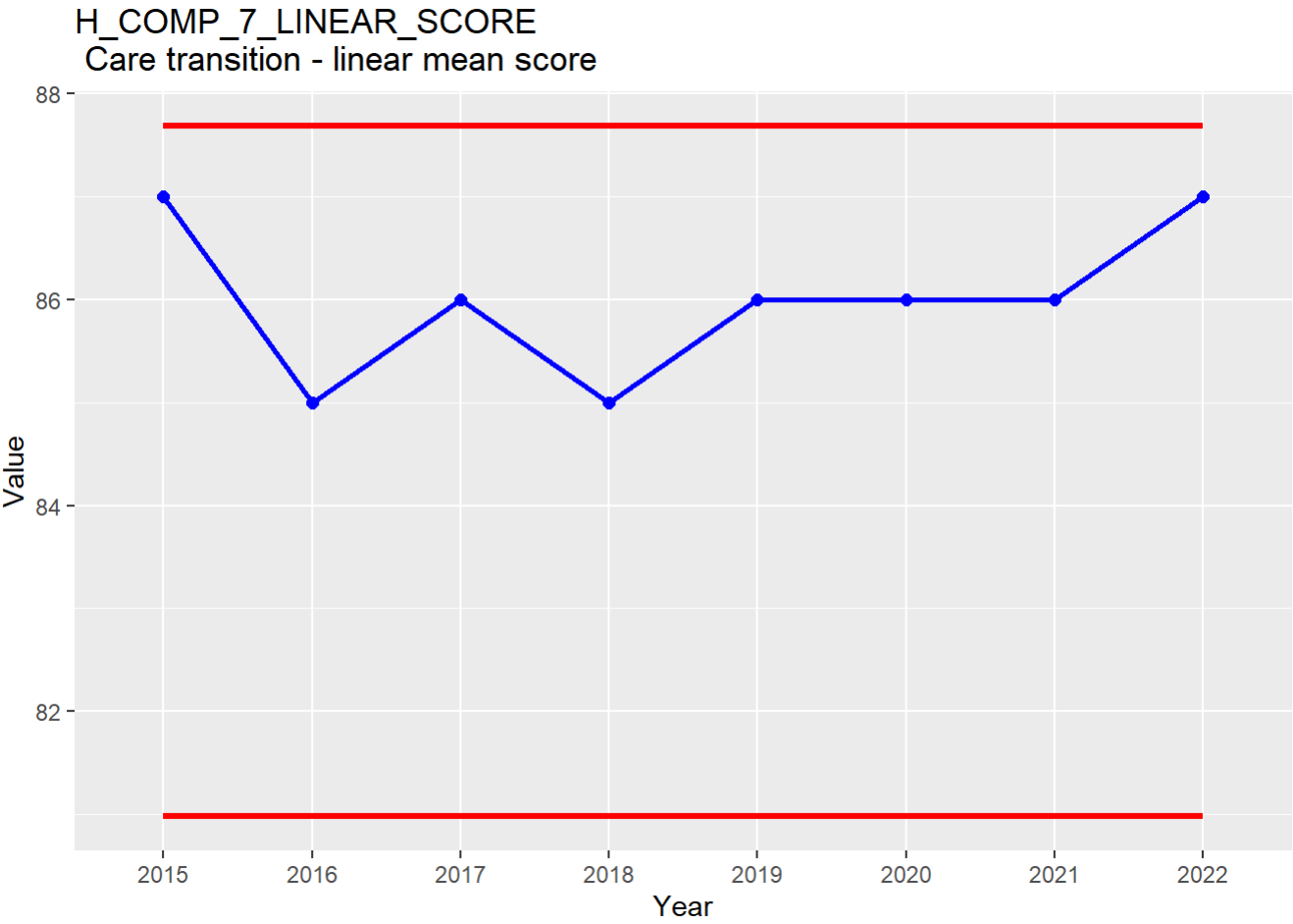
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```

H_COMP_1_A_P

Patients who reported that their nurses "Always" communicated well



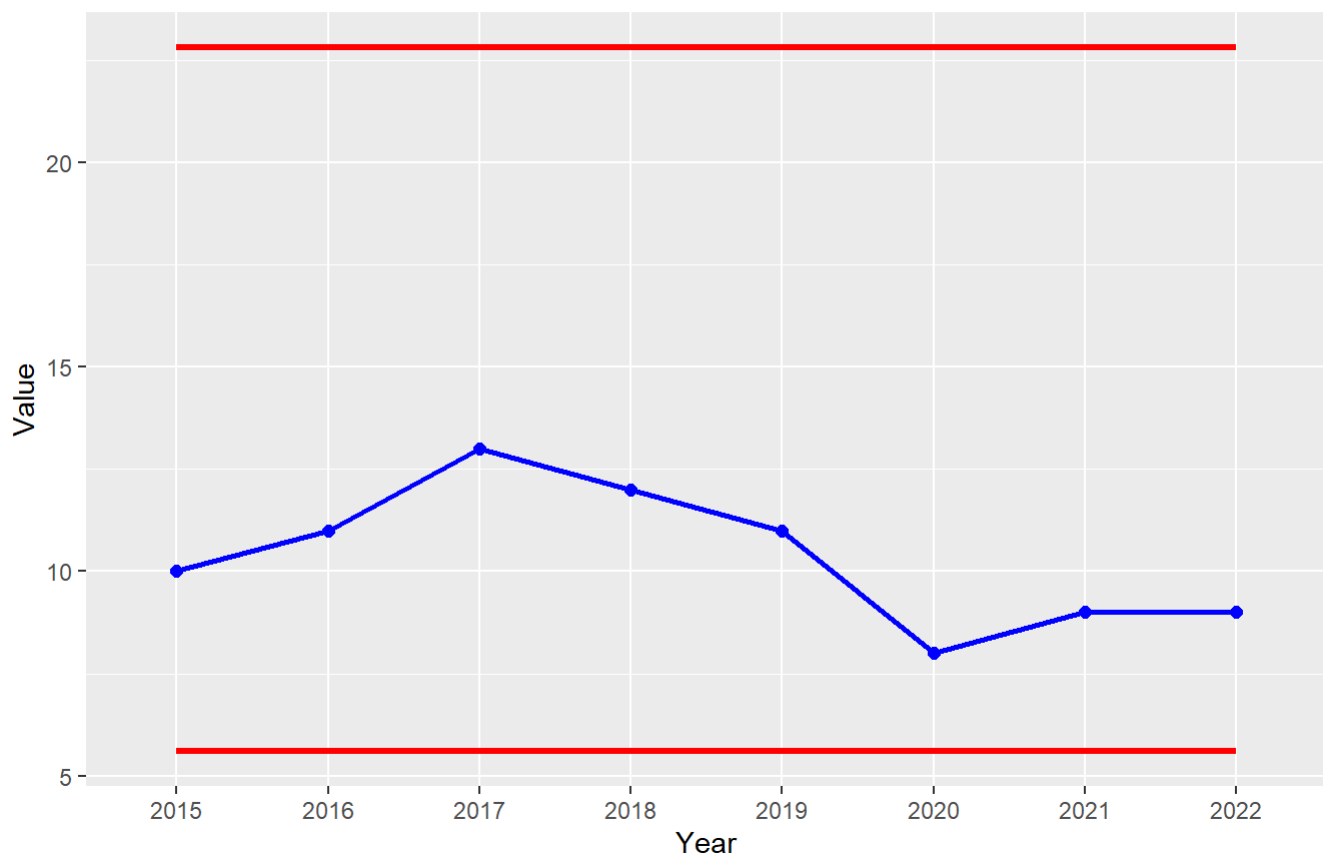
```
data_plotting(final_data_v2)
```



```
data_plotting(final_data_v3)
```

H_RECMND_PY

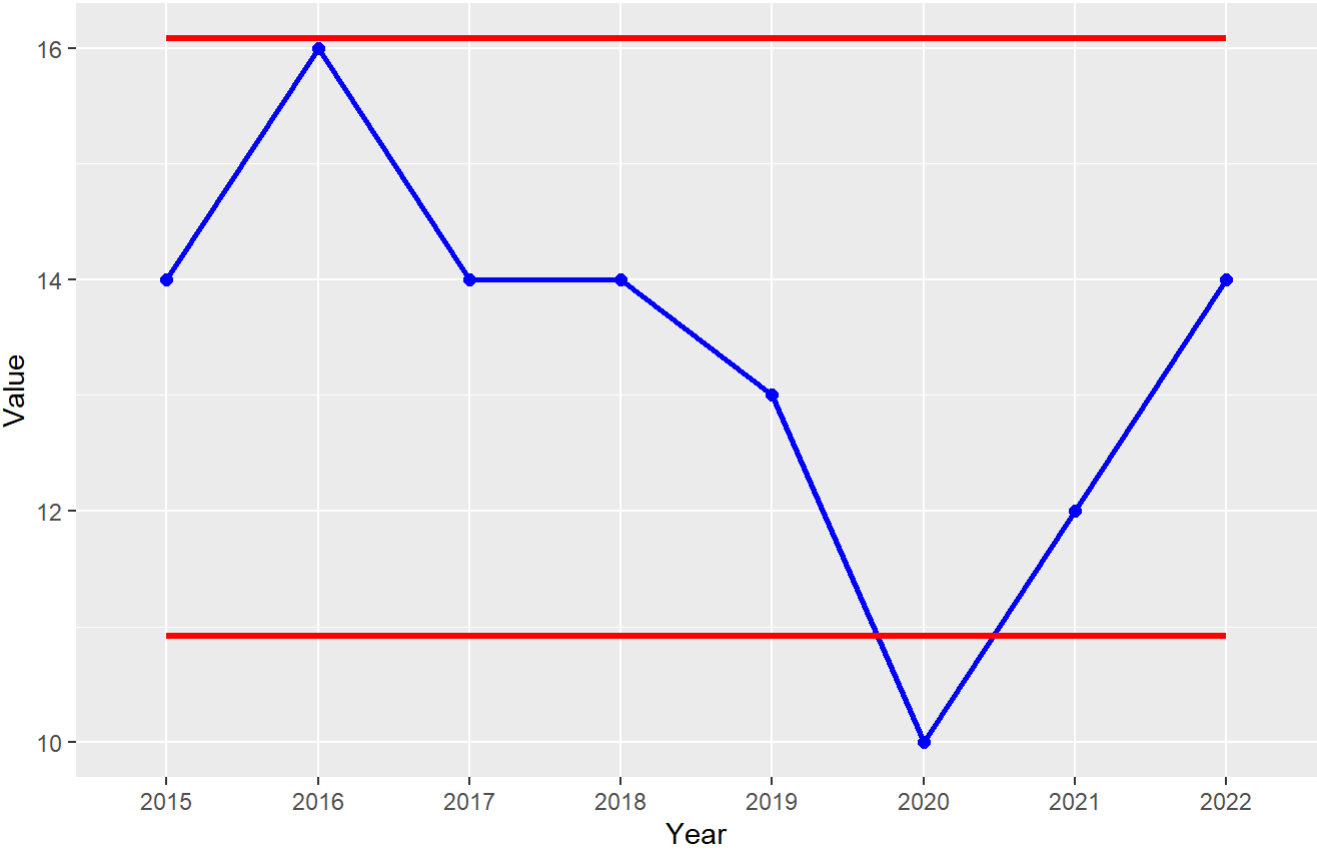
Patients who reported YES, they would probably recommend the hospital



```
data_plotting(final_data_v4)
```

H_COMP_2_U_P

Patients who reported that their doctors "Usually" communicated well



```
data_plotting(final_data_v5)
```

