Term Project

Sharad Bajpai

11/23/2019

Libraries required

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.0     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(rvest)
```

```
## Loading required package: xml2
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:purrr':
##
##     pluck
```

```
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
library(knitr)
library(readxl)
library(png)
```

Importing datasets for Private Insurance:

```
library(readr)
California <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/Priva
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
```

```
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_double(),
##   Treatment = col_character()
## )
```

```
head(California)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Kaiser Fou~ CA    HMO   Yes   2.5
## 2              4.5 Kaiser Fou~ CA    HMO   Yes   3
## 3              4.5 Sharp Heal~ CA    HMO   Yes   4
## 4              3.5 Aetna Heal~ CA    HMO/~ Yes   2
## 5              3.5 Anthem Blu~ CA    PPO/~ Yes   2.5
## 6              3.5 Blue Cross~ CA    HMO/~ Yes   2.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
library(readr)
NewYork <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/Private/
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_character(),
##   Treatment = col_double()
## )
```

```
head(NewYork)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              5   Capital Di~ NY    HMO   Yes   5
## 2              5   Capital Di~ NY    HMO/~ Yes   5
## 3              4.5 Capital Di~ NY    PPO   Yes   4.5
## 4              4.5 CDPHP Univ~ NY    PPO   Yes   4.5
## 5              4.5 HealthNow ~ NY    HMO/~ Yes   4
## 6              4.5 Independen~ NY    HMO/~ Yes   4
## # ... with 2 more variables: Prevention <chr>, Treatment <dbl>
```

```
library(readr)
Pennsylvania <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/Pri
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   `Consumer Satisfaction` = col_double(),
##   Prevention = col_double(),
##   Treatment = col_double(),
##   NCQA = col_character()
## )
```

```
head(Pennsylvania)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  `Consumer Satis~ Prevention
##              <dbl> <chr>       <chr> <chr>            <dbl>      <dbl>
## 1              4.5 Martin's P~ MA, ~ HMO                5          3.5
## 2              4.5 UPMC Benef~ PA    HMO                3.5        4.5
## 3              4.5 UPMC Healt~ PA    HMO                3.5        4.5
## 4              4.5 UPMC Healt~ PA    HMO                3.5        4.5
## 5              4   Aetna Heal~ PA    HMO/~              3          3.5
## 6              4   Aetna Life~ PA    PPO/~              3.5        3.5
## # ... with 2 more variables: Treatment <dbl>, NCQA <chr>
```

```
library(readr)
Virginia <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/Private
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(Virginia)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              5   Kaiser Fou~ DC, ~ HMO   Yes   3.5
## 2              4   Optima Hea~ VA    HMO/~ Yes   3.5
## 3              3.5 Aetna Heal~ DC, ~ HMO/~ Yes   3.5
## 4              3.5 Aetna Life~ VA    PPO/~ Yes   3
## 5              3.5 Anthem Hea~ VA    PPO/~ Yes   3
## 6              3.5 CareFirst ~ DC, ~ HMO/~ Yes   3
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
library(readr)
Wisconsin <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/Priva
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(Wisconsin)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Dean Healt~ WI    HMO   Yes   4
## 2              4.5 Group Heal~ WI    HMO   Yes   3.5
## 3              4.5 HealthPart~ IA, ~ HMO/~ Yes   3.5
## 4              4.5 Network He~ WI    HMO/~ Yes   3.5
## 5              4   Aspirus Ar~ WI    HMO/~ Yes   2.5
## 6              4   Group Heal~ WI    HMO   No    3.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

Cleaning the State column name:

```
California$State<-gsub("[[:upper:]]","",California$State)
California$State<-'CA'
head(California)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Kaiser Fou~ CA    HMO   Yes   2.5
## 2              4.5 Kaiser Fou~ CA    HMO   Yes   3
## 3              4.5 Sharp Heal~ CA    HMO   Yes   4
## 4              3.5 Aetna Heal~ CA    HMO/~ Yes   2
## 5              3.5 Anthem Blu~ CA    PPO/~ Yes   2.5
## 6              3.5 Blue Cross~ CA    HMO/~ Yes   2.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
NewYork$State<-gsub("[[:upper:]]","",NewYork$State)
NewYork$State<-'NY'
head(NewYork)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
```

```
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              5   Capital Di~ NY    HMO   Yes   5
## 2              5   Capital Di~ NY    HMO/~ Yes   5
## 3              4.5 Capital Di~ NY    PPO   Yes   4.5
## 4              4.5 CDPHP Univ~ NY    PPO   Yes   4.5
## 5              4.5 HealthNow ~ NY    HMO/~ Yes   4
## 6              4.5 Independen~ NY    HMO/~ Yes   4
## # ... with 2 more variables: Prevention <chr>, Treatment <dbl>
```

```
Pennsylvania$State<-gsub("[[:upper:]]","",Pennsylvania$State)
Pennsylvania$State<-'PA'
head(Pennsylvania)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  `Consumer Satis~ Prevention
##              <dbl> <chr>       <chr> <chr>            <dbl>      <dbl>
## 1              4.5 Martin's P~ PA    HMO                5          3.5
## 2              4.5 UPMC Benef~ PA    HMO                3.5        4.5
## 3              4.5 UPMC Healt~ PA    HMO                3.5        4.5
## 4              4.5 UPMC Healt~ PA    HMO                3.5        4.5
## 5              4   Aetna Heal~ PA    HMO/~              3          3.5
## 6              4   Aetna Life~ PA    PPO/~              3.5        3.5
## # ... with 2 more variables: Treatment <dbl>, NCQA <chr>
```

```
Virginia$State<-gsub("[[:upper:]]","",Virginia$State)
Virginia$State<-'VA'
head(Virginia)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              5   Kaiser Fou~ VA    HMO   Yes   3.5
## 2              4   Optima Hea~ VA    HMO/~ Yes   3.5
## 3              3.5 Aetna Heal~ VA    HMO/~ Yes   3.5
## 4              3.5 Aetna Life~ VA    PPO/~ Yes   3
## 5              3.5 Anthem Hea~ VA    PPO/~ Yes   3
## 6              3.5 CareFirst ~ VA    HMO/~ Yes   3
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
Wisconsin$State<-gsub("[[:upper:]]","",Wisconsin$State)
Wisconsin$State<-'WI'
head(Wisconsin)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Dean Healt~ WI    HMO   Yes   4
## 2              4.5 Group Heal~ WI    HMO   Yes   3.5
## 3              4.5 HealthPart~ WI    HMO/~ Yes   3.5
## 4              4.5 Network He~ WI    HMO/~ Yes   3.5
## 5              4   Aspirus Ar~ WI    HMO/~ Yes   2.5
## 6              4   Group Heal~ WI    HMO   No    3.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

Combining these datasets into one.

```
Insurance<-rbind(California,NewYork,Pennsylvania,Virginia, Wisconsin)
head(Insurance)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Kaiser Fou~ CA    HMO   Yes   2.5
## 2              4.5 Kaiser Fou~ CA    HMO   Yes   3
## 3              4.5 Sharp Heal~ CA    HMO   Yes   4
## 4              3.5 Aetna Heal~ CA    HMO/~ Yes   2
## 5              3.5 Anthem Blu~ CA    PPO/~ Yes   2.5
## 6              3.5 Blue Cross~ CA    HMO/~ Yes   2.5
## # ... with 2 more variables: Prevention <chr>, Treatment <chr>
```

```
Ins_Id <-1
Insurance <- cbind(Ins_Id,Insurance)
HealthPlanId<- c(101:200)
HealthPlanId<-as.data.frame(HealthPlanId)
Insurance <- cbind(HealthPlanId,Insurance)
head(Insurance, 8)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1          101      1            4.5
## 2          102      1            4.5
## 3          103      1            4.5
## 4          104      1            3.5
## 5          105      1            3.5
## 6          106      1            3.5
## 7          107      1            3.5
## 8          108      1            3.5
##                                                 Plan Name State    Type
## 1  Kaiser Foundation Health Plan Inc. - Southern California    CA     HMO
## 2 Kaiser Foundation Health Plan, Inc. - Northern California    CA     HMO
## 3                                       Sharp Health Plan    CA     HMO
## 4                             Aetna Health of California Inc.    CA HMO/POS
## 5       Anthem Blue Cross Life and Health Insurance Company    CA PPO/EPO
## 6            Blue Cross of California dba Anthem Blue Cross    CA HMO/POS
## 7            Blue Cross of California dba Anthem Blue Cross    CA PPO/EPO
## 8                                  Blue Shield of California    CA HMO/POS
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                  2.5          5       4.5
## 2  Yes                    3          5       4.5
## 3  Yes                    4          4         4
## 4  Yes                    2        2.5         3
## 5  Yes                  2.5          3         3
## 6  Yes                  2.5        3.5         3
## 7  Yes                  2.5          3         3
## 8  Yes                  2.5        3.5         3
```

Removing Special Characters from the columns:

```
Insurance$`Consumer Satisfaction`<-gsub("I","0",Insurance$`Consumer Satisfaction`)
Insurance$Prevention<-gsub("I","0", Insurance$Prevention)
Insurance$Treatment<-gsub("I","0", Insurance$Treatment)
Insurance$`Overall Rating`<-as.numeric(Insurance$`Overall Rating`)
Insurance$`Consumer Satisfaction`<-as.numeric(Insurance$`Consumer Satisfaction`)
Insurance$Prevention<-as.numeric(Insurance$Prevention)
Insurance$Treatment<-as.numeric(Insurance$Treatment)
head(Insurance)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1          101      1            4.5
## 2          102      1            4.5
## 3          103      1            4.5
## 4          104      1            3.5
## 5          105      1            3.5
## 6          106      1            3.5
##                                                 Plan Name State    Type
## 1  Kaiser Foundation Health Plan Inc. - Southern California    CA     HMO
## 2 Kaiser Foundation Health Plan, Inc. - Northern California    CA     HMO
## 3                                       Sharp Health Plan    CA     HMO
## 4                             Aetna Health of California Inc.    CA HMO/POS
## 5       Anthem Blue Cross Life and Health Insurance Company    CA PPO/EPO
## 6            Blue Cross of California dba Anthem Blue Cross    CA HMO/POS
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                  2.5        5.0       4.5
## 2  Yes                  3.0        5.0       4.5
## 3  Yes                  4.0        4.0       4.0
## 4  Yes                  2.0        2.5       3.0
## 5  Yes                  2.5        3.0       3.0
## 6  Yes                  2.5        3.5       3.0
```

Importing dataset for Medicare:

```
library(readr)
CaliforniaMedicare <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Da
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(CaliforniaMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
```

```
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              5   Kaiser Fou~ CA    HMO   Yes   3.5
## 2              5   Kaiser Fou~ CA    HMO   Yes   3.5
## 3              4   Aetna Life~ CA    PPO   Yes   3.5
## 4              4   Sierra Hea~ AZ, ~ PPO   Yes   4
## 5              3.5 Aetna Heal~ CA    HMO   Yes   2.5
## 6              3.5 Blue Shiel~ CA    HMO   Yes   2.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
library(readr)
NewYorkMedicare <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_double(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(NewYorkMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr>            <dbl>
## 1              4.5 Capital Di~ NY    HMO   Yes                4
## 2              4.5 Excellus H~ NY    HMO   Yes                4
## 3              4.5 Excellus H~ NY    HMO   Yes                4
## 4              4   Aetna Life~ NY    PPO   Yes                3.5
## 5              4   CDPHP Univ~ NY    PPO   No                 5
## 6              4   Excellus H~ NY    PPO   Yes                3.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
library(readr)
PennsylvaniaMedicare <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_double(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(PennsylvaniaMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr>            <dbl>
## 1              4.5 Geisinger ~ PA    HMO   Yes                4.5
## 2              4   Aetna Heal~ PA    HMO   Yes                4.5
## 3              4   Aetna Life~ PA    PPO   Yes                3.5
## 4              4   Keystone H~ PA    HMO   Yes                3.5
## 5              4   QCC Insura~ PA    PPO   Yes                4.5
## 6              4   Sierra Hea~ CO, ~ HMO/~ No                 3.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
library(readr)
VirginiaMedicare <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_character(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_double(),
##   Treatment = col_character()
## )
```

```
head(VirginiaMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Kaiser Fou~ DC, ~ HMO   Yes   3.5
## 2 4                Aetna Life~ VA    PPO   Yes   3.5
## 3 4                Sierra Hea~ CO, ~ HMO/~ No    3.5
## 4 4                Sierra Hea~ DE, ~ PPO   Yes   4
## 5 4                UnitedHeal~ KY, ~ HMO/~ Yes   3.5
## 6 3.5              Aetna Heal~ DC, ~ HMO   Yes   3
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
library(readr)
WisconsinMedicare <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Da
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_character(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
```

```
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(WisconsinMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Dean Healt~ WI    HMO   No    5
## 2 4.5              Medical As~ WI    HMO   Yes   4.5
## 3 4.5              Network He~ WI    PPO   Yes   4
## 4 4.5              Quartz Hea~ IA, ~ HMO   Yes   5
## 5 4                Aetna Life~ WI    PPO   Yes   3.5
## 6 4                Dean Healt~ WI    HMO/~ No    4.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

Cleaning the State column name:

```
CaliforniaMedicare$State<-gsub("[[:upper:]]","",CaliforniaMedicare$State)
CaliforniaMedicare$State<-'CA'
head(California)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Kaiser Fou~ CA    HMO   Yes   2.5
## 2              4.5 Kaiser Fou~ CA    HMO   Yes   3
## 3              4.5 Sharp Heal~ CA    HMO   Yes   4
## 4              3.5 Aetna Heal~ CA    HMO/~ Yes   2
## 5              3.5 Anthem Blu~ CA    PPO/~ Yes   2.5
## 6              3.5 Blue Cross~ CA    HMO/~ Yes   2.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
NewYorkMedicare$State<-gsub("[[:upper:]]","",NewYorkMedicare$State)
NewYorkMedicare$State<-'NY'
head(NewYorkMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr>            <dbl>
## 1              4.5 Capital Di~ NY    HMO   Yes                4
## 2              4.5 Excellus H~ NY    HMO   Yes                4
## 3              4.5 Excellus H~ NY    HMO   Yes                4
## 4              4   Aetna Life~ NY    PPO   Yes                3.5
## 5              4   CDPHP Univ~ NY    PPO   No                 5
## 6              4   Excellus H~ NY    PPO   Yes                3.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
PennsylvaniaMedicare$State<-gsub("[[:upper:]]","",PennsylvaniaMedicare$State)
PennsylvaniaMedicare$State<-'PA'
head(PennsylvaniaMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr>            <dbl>
## 1              4.5 Geisinger ~ PA    HMO   Yes                4.5
## 2              4   Aetna Heal~ PA    HMO   Yes                4.5
## 3              4   Aetna Life~ PA    PPO   Yes                3.5
## 4              4   Keystone H~ PA    HMO   Yes                3.5
## 5              4   QCC Insura~ PA    PPO   Yes                4.5
## 6              4   Sierra Hea~ PA    HMO/~ No                 3.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
VirginiaMedicare$State<-gsub("[[:upper:]]","",VirginiaMedicare$State)
VirginiaMedicare$State<-'VA'
head(VirginiaMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Kaiser Fou~ VA    HMO   Yes   3.5
## 2 4                Aetna Life~ VA    PPO   Yes   3.5
## 3 4                Sierra Hea~ VA    HMO/~ No    3.5
## 4 4                Sierra Hea~ VA    PPO   Yes   4
## 5 4                UnitedHeal~ VA    HMO/~ Yes   3.5
## 6 3.5              Aetna Heal~ VA    HMO   Yes   3
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
WisconsinMedicare$State<-gsub("[[:upper:]]","",WisconsinMedicare$State)
WisconsinMedicare$State<-'WI'
head(WisconsinMedicare)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Dean Healt~ WI    HMO   No    5
## 2 4.5              Medical As~ WI    HMO   Yes   4.5
## 3 4.5              Network He~ WI    PPO   Yes   4
## 4 4.5              Quartz Hea~ WI    HMO   Yes   5
## 5 4                Aetna Life~ WI    PPO   Yes   3.5
## 6 4                Dean Healt~ WI    HMO/~ No    4.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

Combining these Medicare datasets into one.

```
Medicare<-rbind(CaliforniaMedicare,NewYorkMedicare,PennsylvaniaMedicare,VirginiaMedicare, WisconsinMedi
head(Medicare)
```

```
## # A tibble: 6 x 8
```

```
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 5                Kaiser Fou~ CA    HMO   Yes   3.5
## 2 5                Kaiser Fou~ CA    HMO   Yes   3.5
## 3 4                Aetna Life~ CA    PPO   Yes   3.5
## 4 4                Sierra Hea~ CA    PPO   Yes   4
## 5 3.5              Aetna Heal~ CA    HMO   Yes   2.5
## 6 3.5              Blue Shiel~ CA    HMO   Yes   2.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
Ins_Id<-2
Medicare<-cbind(Ins_Id,Medicare)
HealthPlanId<- c(1001:1099)
HealthPlanId<-as.data.frame(HealthPlanId)
Medicare <- cbind(HealthPlanId,Medicare)
head(Medicare)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1         1001      2              5
## 2         1002      2              5
## 3         1003      2              4
## 4         1004      2              4
## 5         1005      2            3.5
## 6         1006      2            3.5
##                                                 Plan Name State Type
## 1  Kaiser Foundation Health Plan Inc. - Southern California    CA  HMO
## 2 Kaiser Foundation Health Plan, Inc. - Northern California    CA  HMO
## 3                 Aetna Life Insurance Company (California)    CA  PPO
## 4 Sierra Health and Life Insurance Company, Inc. (AZ,CA,NV)    CA  PPO
## 5                           Aetna Health of California Inc.    CA  HMO
## 6                                Blue Shield of California    CA  HMO
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                  3.5        5.0       4.5
## 2  Yes                  3.5        5.0       4.5
## 3  Yes                  3.5        3.5       3.5
## 4  Yes                    4        4.5       3.5
## 5  Yes                  2.5        3.0       3.5
## 6  Yes                  2.5        3.5       3.5
```

Removing Special Characters from the columns:

```
Medicare$`Consumer Satisfaction`<-gsub("I","0",Medicare$`Consumer Satisfaction`)
Medicare$Prevention<-gsub("I|O","0", Medicare$Prevention)
Medicare$Treatment<-gsub("I","0", Medicare$Treatment)
Medicare$`Overall Rating` <- gsub("Partial Data Reported|No Data Reported","0", Medicare$`Overall Ratin
Medicare$`Overall Rating`<-as.numeric(Medicare$`Overall Rating`)
Medicare$`Consumer Satisfaction`<-as.numeric(Medicare$`Consumer Satisfaction`)
```

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

```
Medicare$Prevention<-as.numeric(Medicare$Prevention)
Medicare$Treatment<-as.numeric(Medicare$Treatment)
head(Medicare)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1         1001      2            5.0
## 2         1002      2            5.0
## 3         1003      2            4.0
## 4         1004      2            4.0
## 5         1005      2            3.5
## 6         1006      2            3.5
##                                                  Plan Name State Type
## 1  Kaiser Foundation Health Plan Inc. - Southern California    CA  HMO
## 2 Kaiser Foundation Health Plan, Inc. - Northern California    CA  HMO
## 3                 Aetna Life Insurance Company (California)    CA  PPO
## 4 Sierra Health and Life Insurance Company, Inc. (AZ,CA,NV)    CA  PPO
## 5                           Aetna Health of California Inc.    CA  HMO
## 6                                 Blue Shield of California    CA  HMO
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                   3.5        5.0       4.5
## 2  Yes                   3.5        5.0       4.5
## 3  Yes                   3.5        3.5       3.5
## 4  Yes                   4.0        4.5       3.5
## 5  Yes                   2.5        3.0       3.5
## 6  Yes                   2.5        3.5       3.5
```

Fixing multiple values in Consumer Satsifaction column:

```
Medicare$`Consumer Satisfaction`[17]<-1.5
Medicare$`Consumer Satisfaction`[20]<-2.0
head(Medicare)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1         1001      2            5.0
## 2         1002      2            5.0
## 3         1003      2            4.0
## 4         1004      2            4.0
## 5         1005      2            3.5
## 6         1006      2            3.5
##                                                  Plan Name State Type
## 1  Kaiser Foundation Health Plan Inc. - Southern California    CA  HMO
## 2 Kaiser Foundation Health Plan, Inc. - Northern California    CA  HMO
## 3                 Aetna Life Insurance Company (California)    CA  PPO
## 4 Sierra Health and Life Insurance Company, Inc. (AZ,CA,NV)    CA  PPO
## 5                           Aetna Health of California Inc.    CA  HMO
## 6                                 Blue Shield of California    CA  HMO
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                   3.5        5.0       4.5
## 2  Yes                   3.5        5.0       4.5
## 3  Yes                   3.5        3.5       3.5
## 4  Yes                   4.0        4.5       3.5
## 5  Yes                   2.5        3.0       3.5
## 6  Yes                   2.5        3.5       3.5
```

Importing dataset for Medicaid:

```
library(readr)
CaliforniaMedicaid <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Da
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_character(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_character(),
##   Treatment = col_character()
## )
```

```
head(CaliforniaMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4                Alameda Al~ CA    HMO   Yes   3
## 2 4                Community ~ CA    HMO   Yes   3
## 3 4                Local Init~ CA    HMO   Yes   2
## 4 4                Orange Cou~ CA    HMO   Yes   2.5
## 5 4                San Franci~ CA    HMO   Yes   3
## 6 3.5              Contra Cos~ CA    HMO   Yes   2
## # ... with 2 more variables: Prevention <chr>, Treatment <chr>
```

```
library(readr)
NewYorkMedicaid <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data/
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_character(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_character(),
##   Treatment = col_character()
## )
```

```
head(NewYorkMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Capital Di~ NY    HMO   Yes   3.5
## 2 4.5              Excellus H~ NY    HMO   Yes   4
## 3 4.5              Excellus H~ NY    HMO   Yes   4
```

```
## 4 4               HealthPlus~ NY    HMO   Yes   2.5
## 5 3.5             HealthNow ~ NY    HMO   No    3.5
## 6 3.5             UnitedHeal~ NY    HMO   Yes   2.5
## # ... with 2 more variables: Prevention <chr>, Treatment <chr>
```

```
library(readr)
PennsylvaniaMedicaid <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_character(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_character(),
##   Treatment = col_double()
## )
```

```
head(PennsylvaniaMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Health Par~ PA    HMO   Yes   3
## 2 4.5              Vista Heal~ PA    HMO   Yes   4.5
## 3 4                Gateway He~ PA    HMO   Yes   3.5
## 4 4                Geisinger ~ PA    HMO   Yes   3.5
## 5 4                UPMC For Y~ PA    HMO   Yes   3
## 6 4                Vista Heal~ PA    HMO   Yes   3.5
## # ... with 2 more variables: Prevention <chr>, Treatment <dbl>
```

```
library(readr)
VirginiaMedicaid <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Data
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_double(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_double(),
##   Prevention = col_double(),
##   Treatment = col_double()
## )
```

```
head(VirginiaMedicaid)
```

```
## # A tibble: 6 x 8
```

```
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr>            <dbl>
## 1              3.5 HealthKeep~ VA    HMO   Yes                3.5
## 2              3.5 Optima Hea~ VA    HMO   Yes                4.5
## 3              3.5 Virginia P~ VA    HMO   Yes                4
## 4              3   Coventry H~ VA    HMO   Yes                3
## 5              3   UnitedHeal~ VA    HMO   Yes ~              3
## 6              2.5 Magellan C~ VA    HMO   Yes ~              3
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
library(readr)
WisconsinMedicaid <- read_csv("C:/Users/bajpa/Downloads/Collecting, Storing, Retrieving Data/Project/Da
```

```
## Parsed with column specification:
## cols(
##   `Overall Rating` = col_character(),
##   `Plan Name` = col_character(),
##   State = col_character(),
##   Type = col_character(),
##   NCQA = col_character(),
##   `Consumer Satisfaction` = col_character(),
##   Prevention = col_character(),
##   Treatment = col_double()
## )
```

```
head(WisconsinMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4                Children's~ WI    HMO   Yes   3.5
## 2 4                Security H~ WI    HMO   Yes   3
## 3 4                UnitedHeal~ WI    HMO   Yes   3.5
## 4 3.5              Compcare H~ WI    HMO   Yes   3
## 5 3.5              Managed He~ WI    HMO   Yes   3
## 6 3.5              Molina Hea~ WI    HMO   Yes   3.5
## # ... with 2 more variables: Prevention <chr>, Treatment <dbl>
```

Cleaning the State column name:

```
CaliforniaMedicaid$State<-gsub("[[:upper:]]","",CaliforniaMedicaid$State)
CaliforniaMedicaid$State<-'CA'
head(California)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr> <chr>
## 1              4.5 Kaiser Fou~ CA    HMO   Yes   2.5
## 2              4.5 Kaiser Fou~ CA    HMO   Yes   3
## 3              4.5 Sharp Heal~ CA    HMO   Yes   4
## 4              3.5 Aetna Heal~ CA    HMO/~ Yes   2
## 5              3.5 Anthem Blu~ CA    PPO/~ Yes   2.5
## 6              3.5 Blue Cross~ CA    HMO/~ Yes   2.5
## # ... with 2 more variables: Prevention <dbl>, Treatment <chr>
```

```
NewYorkMedicaid$State<-gsub("[[:upper:]]","",NewYorkMedicaid$State)
NewYorkMedicaid$State<-'NY'
head(NewYorkMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Capital Di~ NY    HMO   Yes   3.5
## 2 4.5              Excellus H~ NY    HMO   Yes   4
## 3 4.5              Excellus H~ NY    HMO   Yes   4
## 4 4                HealthPlus~ NY    HMO   Yes   2.5
## 5 3.5              HealthNow ~ NY    HMO   No    3.5
## 6 3.5              UnitedHeal~ NY    HMO   Yes   2.5
## # ... with 2 more variables: Prevention <chr>, Treatment <chr>
```

```
PennsylvaniaMedicaid$State<-gsub("[[:upper:]]","",PennsylvaniaMedicaid$State)
PennsylvaniaMedicaid$State<-'PA'
head(PennsylvaniaMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4.5              Health Par~ PA    HMO   Yes   3
## 2 4.5              Vista Heal~ PA    HMO   Yes   4.5
## 3 4                Gateway He~ PA    HMO   Yes   3.5
## 4 4                Geisinger ~ PA    HMO   Yes   3.5
## 5 4                UPMC For Y~ PA    HMO   Yes   3
## 6 4                Vista Heal~ PA    HMO   Yes   3.5
## # ... with 2 more variables: Prevention <chr>, Treatment <dbl>
```

```
VirginiaMedicaid$State<-gsub("[[:upper:]]","",VirginiaMedicaid$State)
VirginiaMedicaid$State<-'VA'
head(VirginiaMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##              <dbl> <chr>       <chr> <chr> <chr>            <dbl>
## 1              3.5 HealthKeep~ VA    HMO   Yes                3.5
## 2              3.5 Optima Hea~ VA    HMO   Yes                4.5
## 3              3.5 Virginia P~ VA    HMO   Yes                4
## 4              3   Coventry H~ VA    HMO   Yes                3
## 5              3   UnitedHeal~ VA    HMO   Yes ~              3
## 6              2.5 Magellan C~ VA    HMO   Yes ~              3
## # ... with 2 more variables: Prevention <dbl>, Treatment <dbl>
```

```
WisconsinMedicaid$State<-gsub("[[:upper:]]","",WisconsinMedicaid$State)
WisconsinMedicaid$State<-'WI'
head(WisconsinMedicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
```

```
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4                Children's~ WI    HMO   Yes   3.5
## 2 4                Security H~ WI    HMO   Yes   3
## 3 4                UnitedHeal~ WI    HMO   Yes   3.5
## 4 3.5              Compcare H~ WI    HMO   Yes   3
## 5 3.5              Managed He~ WI    HMO   Yes   3
## 6 3.5              Molina Hea~ WI    HMO   Yes   3.5
## # ... with 2 more variables: Prevention <chr>, Treatment <dbl>
```

Combining these datasets into one.

```
Medicaid<-rbind(CaliforniaMedicaid,NewYorkMedicaid,PennsylvaniaMedicaid,VirginiaMedicaid, WisconsinMedi
head(Medicaid)
```

```
## # A tibble: 6 x 8
##   `Overall Rating` `Plan Name` State Type  NCQA  `Consumer Satis~
##   <chr>            <chr>       <chr> <chr> <chr> <chr>
## 1 4                Alameda Al~ CA    HMO   Yes   3
## 2 4                Community ~ CA    HMO   Yes   3
## 3 4                Local Init~ CA    HMO   Yes   2
## 4 4                Orange Cou~ CA    HMO   Yes   2.5
## 5 4                San Franci~ CA    HMO   Yes   3
## 6 3.5              Contra Cos~ CA    HMO   Yes   2
## # ... with 2 more variables: Prevention <chr>, Treatment <chr>
```

```
Ins_Id<-3
Medicaid<-cbind(Ins_Id,Medicaid)
HealthPlanId<- c(5001:5073)
HealthPlanId<-as.data.frame(HealthPlanId)
Medicaid <- cbind(HealthPlanId,Medicaid)
head(Medicaid)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1         5001      3              4
## 2         5002      3              4
## 3         5003      3              4
## 4         5004      3              4
## 5         5005      3              4
## 6         5006      3            3.5
##                                                 Plan Name State Type
## 1                                 Alameda Alliance for Health    CA  HMO
## 2                                     Community Health Group    CA  HMO
## 3 Local Initiative Health Authority, dba L.A. Care Health Plan    CA  HMO
## 4               Orange County Health Authority - dba CalOptima    CA  HMO
## 5                    San Francisco Community Health Authority    CA  HMO
## 6                                   Contra Costa Health Plan    CA  HMO
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                     3        4.5         3
## 2  Yes                     3          4       3.5
## 3  Yes                     2        3.5       3.5
## 4  Yes                   2.5          4       3.5
## 5  Yes                     3          4       3.5
## 6  Yes                     2        3.5         3
```

Removing Special Characters from the columns:

```
Medicaid$`Consumer Satisfaction`<-gsub("I","0",Medicaid$`Consumer Satisfaction`)
Medicaid$Prevention<-gsub("I|0","0", Medicaid$Prevention)
Medicaid$Treatment<-gsub("I","0", Medicaid$Treatment)
Medicaid$`Overall Rating` <- gsub("Partial Data Reported|No Data Reported","0", Medicaid$`Overall Rating
Medicaid$NCQA<-gsub("(In Process)|(Scheduled)|(Interim)|(In process)|[[:punct:]]"," ", Medicaid$NCQA)
Medicaid$`Overall Rating`<-as.numeric(Medicaid$`Overall Rating`)
Medicaid$`Consumer Satisfaction`<-as.numeric(Medicaid$`Consumer Satisfaction`)
Medicaid$Prevention<-as.numeric(Medicaid$Prevention)
Medicaid$Treatment<-as.numeric(Medicaid$Treatment)
head(Medicaid)
```

```
##   HealthPlanId Ins_Id Overall Rating
## 1         5001      3            4.0
## 2         5002      3            4.0
## 3         5003      3            4.0
## 4         5004      3            4.0
## 5         5005      3            4.0
## 6         5006      3            3.5
##                                                  Plan Name State Type
## 1                                Alameda Alliance for Health    CA  HMO
## 2                                    Community Health Group    CA  HMO
## 3 Local Initiative Health Authority, dba L.A. Care Health Plan    CA  HMO
## 4               Orange County Health Authority - dba CalOptima    CA  HMO
## 5                     San Francisco Community Health Authority    CA  HMO
## 6                                  Contra Costa Health Plan    CA  HMO
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                  3.0        4.5       3.0
## 2  Yes                  3.0        4.0       3.5
## 3  Yes                  2.0        3.5       3.5
## 4  Yes                  2.5        4.0       3.5
## 5  Yes                  3.0        4.0       3.5
## 6  Yes                  2.0        3.5       3.0
```

Fixing multiple values in Prevention column:

```
Medicaid$Prevention[67]<-0.5
view(Medicaid)
```

Creating the Central Entity that connects all the other tables.

```
Ins_Id<-c(1,2,3)
Ins_type<-c("Private_Insurance", "Medicare", "Medicaid")
Ins <- cbind(Ins_Id,Ins_type)
Ins<-as.data.frame(Ins)
head(Ins)
```

```
##   Ins_Id          Ins_type
## 1      1 Private_Insurance
## 2      2          Medicare
## 3      3          Medicaid
```

The four dataframes are Insurnace (Private Insurnace), Medicare and Medicaid and Ins

```
view(Insurance)
view(Medicare)
view(Medicaid)
view(Ins)
```

Connecting to database:

```
library(RSQLite)
database <- dbConnect(SQLite(), dbname="US_Insurance")
dbWriteTable(conn = database, name = 'Insurance', value = Insurance, row.names=F, header=T, overwrite=T
dbWriteTable(conn = database, name = 'Medicare', value = Medicare, row.names=F, header=T, overwrite=T)
dbWriteTable(conn = database, name = "Medicaid", value = Medicaid, row.names=F, header=T, overwrite=T)
dbWriteTable(conn = database, name = "Ins", value = Ins, row.names=F, header=T, overwrite=T)
dbListTables(database)
```

```
## [1] "Ins"       "Insurance" "Medicaid"  "Medicare"
```

Checking the database by running custom queries:

```
dbGetQuery(database, "select Max([Overall Rating]), [Plan Name], State
           from Insurance
           where (State='CA')")
```

```
##   Max([Overall Rating])
## 1                   4.5
##                                                 Plan Name State
## 1 Kaiser Foundation Health Plan Inc. - Southern California    CA
```

```
dbGetQuery(database, "select [Overall Rating], [Plan Name], State, Type, NCQA, [Consumer Satisfaction],
           from Insurance
           where ([Consumer Satisfaction]>3.5 AND [Consumer Satisfaction]!='NA')")
```

```
##    Overall Rating
## 1             4.5
## 2             5.0
## 3             5.0
## 4             4.5
## 5             4.5
## 6             4.5
## 7             4.5
## 8             4.5
## 9             3.5
## 10            3.5
## 11            4.5
## 12            4.0
## 13            4.0
## 14            4.0
## 15            4.0
## 16            4.5
## 17            4.0
```

```
## 18            4.0
## 19            4.0
## 20            3.5
##                                                                                  Plan Name
## 1                                                                          Sharp Health Plan
## 2                                      Capital District Physicians' Health Plan, Inc. (CDPHP)
## 3                                Capital District Physicians' Healthcare Network, Inc. (CDPHN)
## 4                                Capital District Physicians' Healthcare Network, Inc. (CDPHN)
## 5                                                    CDPHP Universal Benefits, Inc.
## 6                                                          HealthNow New York, Inc.
## 7                                                 Independent Health Association, Inc.
## 8                                       Martin's Point US Family Health Plan (NH, NY, PA, VT)
## 9                           Empire HealthChoice HMO, Inc. d/b/a Empire BlueCross BlueShield HMO
## 10                                                                  YourCare Health Plan
## 11                                      Martin's Point US Family Health Plan (NH, NY, PA, VT)
## 12                                                            Capital BlueCross - FEP
## 13                                                            Highmark Choice Company
## 14                                                 Independence Hospital Indemnity Plan
## 15                                               QCC Insurance Company (Personal Choice)
## 16                                                           Dean Health Plan, Inc.
## 17    Medical Associates Clinic Health Plan of Wisconsin dba Medical Associates Health Plans
## 18                                             United HealthCare Services, Inc. (Wisconsin)
## 19                                           UnitedHealthcare Insurance Company (Wisconsin)
## 20 Blue Cross and Blue Shield of Wisconsin dba Anthem Blue Cross and Blue Shield in Wisconsin
##    State        Type NCQA Consumer Satisfaction Prevention Treatment
## 1     CA         HMO  Yes                   4.0        4.0       4.0
## 2     NY         HMO  Yes                   5.0        4.5       4.0
## 3     NY     HMO/POS  Yes                   5.0        4.5       4.0
## 4     NY         PPO  Yes                   4.5        4.5       4.0
## 5     NY         PPO  Yes                   4.5        4.5       4.0
## 6     NY HMO/POS/PPO  Yes                   4.0        4.0       4.0
## 7     NY     HMO/POS  Yes                   4.0        4.5       4.0
## 8     NY         HMO  Yes                   5.0        3.5       3.5
## 9     NY     HMO/POS  Yes                   4.0        3.0       3.0
## 10    NY         HMO   No                   4.0        0.0       3.0
## 11    PA         HMO  Yes                   5.0        3.5       3.5
## 12    PA         PPO  Yes                   4.5        3.5       3.5
## 13    PA         HMO  Yes                   4.0        4.0       3.5
## 14    PA         PPO  Yes                   4.0        4.0       3.0
## 15    PA         PPO  Yes                   4.0        4.0       3.0
## 16    WI         HMO  Yes                   4.0        4.5       4.0
## 17    WI     HMO/POS  Yes                   4.0        3.5       3.5
## 18    WI         PPO  Yes                   4.0        3.5       3.5
## 19    WI         PPO  Yes                   4.0        3.5       3.5
## 20    WI         PPO  Yes                   4.5        2.5       2.5
```

```
dbGetQuery(database, "select [Overall Rating], [Plan Name], State, Type, NCQA, [Consumer Satisfaction],
          from Medicare
          where ([Overall Rating]>4 AND [Consumer Satisfaction]>4 AND[Consumer Satisfaction]!='NA') AN
```

```
##   Overall Rating
## 1            4.5
## 2            4.5
## 3            4.5
```

```
## 4           4.5
##                                                                                  Plan Name
## 1                                                                    Geisinger Health Plan
## 2                                                                   Dean Health Plan, Inc.
## 3 Medical Associates Clinic Health Plan of Wisconsin dba Medical Associates Health Plans
## 4                                                          Quartz Health Plan Corporation
##   State Type NCQA Consumer Satisfaction Prevention Treatment
## 1    PA  HMO  Yes                  4.5        3.5       3.5
## 2    WI  HMO   No                  5.0        4.0       5.0
## 3    WI  HMO  Yes                  4.5        3.0       4.0
## 4    WI  HMO  Yes                  5.0        5.0       3.5
```

```
dbGetQuery(database, "select [Overall Rating], [Plan Name], State, Type, NCQA, [Consumer Satisfaction],
           from Medicare
           where ([Treatment]>4 AND [Prevention]>4 AND [Treatment]!='NA')")
```

```
##   Overall Rating
## 1            5.0
## 2            5.0
## 3            4.5
##                                                  Plan Name State
## 1       Kaiser Foundation Health Plan Inc. - Southern California    CA
## 2      Kaiser Foundation Health Plan, Inc. - Northern California    CA
## 3 Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.    VA
##   Type NCQA Consumer Satisfaction Prevention Treatment
## 1  HMO  Yes                  3.5          5       4.5
## 2  HMO  Yes                  3.5          5       4.5
## 3  HMO  Yes                  3.5          5       4.5
```

```
dbGetQuery(database, "select [Overall Rating], [Plan Name], State, Type, NCQA, [Consumer Satisfaction],
           from Medicaid
           where ([Overall Rating]>4 AND [Prevention]<4 AND [Prevention]!='NA' AND [Overall Rating]!='NA
```

```
##   Overall Rating                                          Plan Name
## 1            4.5 Vista Health Plan DBA AmeriHealth Caritas Pennsylvania
##   State Type NCQA Consumer Satisfaction Prevention Treatment
## 1    PA  HMO  Yes                  4.5        3.5         4
```

```
dbGetQuery(database, "select [Overall Rating], [Plan Name], State, Type, NCQA, [Consumer Satisfaction],
           from Medicaid
           where ([Overall Rating]=4.5 AND [Overall Rating]!='NA') AND State='NY'")
```

```
##   Overall Rating
## 1            4.5
## 2            4.5
## 3            4.5
##                                                Plan Name State Type
## 1       Capital District Physicians' Health Plan, Inc. (CDPHP)    NY  HMO
## 2 Excellus Health Plan, Inc. dba Excellus BlueCross BlueShield    NY  HMO
## 3            Excellus Health Plan, Inc. dba Univera Healthcare    NY  HMO
##   NCQA Consumer Satisfaction Prevention Treatment
## 1  Yes                  3.5          4         4
```

```
## 2  Yes                   4.0          4         4
## 3  Yes                   4.0          4         4
```

```
dbGetQuery(database, "select *
                      from Ins
           ")
```

```
##   Ins_Id          Ins_type
## 1      1 Private_Insurance
## 2      2          Medicare
## 3      3          Medicaid
```

As you can see, I have tested all the four tables and the database is working efficiently. Now I wish to carry out analysis to answer my research questions, 1. Average State ratings (of all the three insurance types) based on consumer satisfaction, treatment, and prevention. Private Insurance Analysis

```
B<- Insurance
B<-na.omit(B)
view(B)

b <- B %>%
  na.omit() %>%
  group_by(State) %>%
  summarise_at(vars(-NCQA, -'Plan Name',-Type,-HealthPlanId, -Ins_Id),funs(mean(., na.rm=TRUE)))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
view(b)

Ans1_a<- b[order(-b$`Overall Rating`),]
head(Ans1_a)
```

```
## # A tibble: 5 x 5
##   State `Overall Rating` `Consumer Satisfaction` Prevention Treatment
##   <chr>            <dbl>                   <dbl>      <dbl>     <dbl>
## 1 PA                4.1                     3.58       3.78      3.45
## 2 WI                4                       2.85       3.75      3.55
## 3 NY                3.82                    3.3        3.22      3.32
## 4 CA                3.52                    2.17       3.35      2.95
## 5 VA                3.45                    2.8        3.15      3.02
```

In the Private Insurance analysis, Pennsylvania has the highest overall, consumer satisfaction and prevention rating whereas, Wisconsin has the highest treatment rating.

Medicare Analysis:

```
C<- Medicare
C<-na.omit(C)
view(C)

c <- C %>%
  na.omit() %>%
  group_by(State) %>%
  summarise_at(vars(-NCQA, -'Plan Name',-Type, -HealthPlanId, -Ins_Id),funs(mean(., na.rm=TRUE)))
view(c)

Ans1_b<- c[order(-c$`Overall Rating`),]
head(Ans1_b)
```

```
## # A tibble: 5 x 5
##   State `Overall Rating` `Consumer Satisfaction` Prevention Treatment
##   <chr>            <dbl>                   <dbl>      <dbl>     <dbl>
## 1 NY                3.85                    3.68       3.95      3.45
## 2 PA                3.72                    3.7        3.62      3.25
## 3 WI                3.66                    3.59       3.66      3.59
## 4 CA                3.5                     2.52       3.4       3.3
## 5 VA                2.29                    2.12       2.74      2.85
```

In the Medicare analysis, New York has the highest overall and prevention rating whereas, Wisconsin has the highest treatment rating and Pennsylvania has highest consumer satisfaction rating.

Medicaid Analysis:

```
D<- Medicaid
D<-na.omit(D)
view(D)

d <- D %>%
  na.omit() %>%
  group_by(State) %>%
  summarise_at(vars(-NCQA, -'Plan Name',-Type,-HealthPlanId, -Ins_Id),funs(mean(., na.rm=TRUE)))
view(d)

Ans1_c<- d[order(-d$`Overall Rating`),]
head(Ans1_c)
```

```
## # A tibble: 5 x 5
##   State `Overall Rating` `Consumer Satisfaction` Prevention Treatment
##   <chr>            <dbl>                   <dbl>      <dbl>     <dbl>
## 1 PA                4                       3.11       3.33      3.56
## 2 VA                3.3                     3.6        2.2       2.7
## 3 NY                2.42                    1.81       3.42      3.31
## 4 CA                2.33                    1.52       2.7       2.3
## 5 WI                1.61                    1.39       2.86      3.21
```

In the Medicaid Analysis, Pennsylvania has the highest Overall Rating and Treatment rating, Virginia has the highest Consumer satisfaction ratings whereas New York has the highest prevention rating.

2. Which insurance plan provides the best customer satisfaction, treatment, and prevention? Health Plans for the Private Insurance:

```
Ans2_a_1 <- B[order(-B$`Consumer Satisfaction`),]
Ans2_a_1 <- Ans2_a_1 %>%
  select(`Plan Name`, State, `Consumer Satisfaction`)
head(Ans2_a_1,4)
```

```
##                                                    Plan Name State
## 21        Capital District Physicians' Health Plan, Inc. (CDPHP)    NY
## 22 Capital District Physicians' Healthcare Network, Inc. (CDPHN)    NY
## 27         Martin's Point US Family Health Plan (NH, NY, PA, VT)    NY
## 41         Martin's Point US Family Health Plan (NH, NY, PA, VT)    PA
##    Consumer Satisfaction
## 21                     5
## 22                     5
## 27                     5
## 41                     5
```

```
Ans2_a_2 <- B[order(-B$Prevention),]
Ans2_a_2 <- Ans2_a_2 %>%
  select(`Plan Name`, State, Prevention)
head(Ans2_a_2,4)
```

```
##                                                     Plan Name State
## 1        Kaiser Foundation Health Plan Inc. - Southern California    CA
## 2       Kaiser Foundation Health Plan, Inc. - Northern California    CA
## 61 Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.    VA
## 21         Capital District Physicians' Health Plan, Inc. (CDPHP)    NY
##    Prevention
## 1         5.0
## 2         5.0
## 61        5.0
## 21        4.5
```

```
Ans2_a_3 <- B[order(-B$Treatment),]
Ans2_a_3 <- Ans2_a_3 %>%
  select(`Plan Name`, State, Treatment)
head(Ans2_a_3,4)
```

```
##                                                     Plan Name State
## 1        Kaiser Foundation Health Plan Inc. - Southern California    CA
## 2       Kaiser Foundation Health Plan, Inc. - Northern California    CA
## 61 Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.    VA
## 3                                                 Sharp Health Plan    CA
##    Treatment
## 1        4.5
## 2        4.5
## 61       4.5
## 3        4.0
```

Health Plan for Medicare:

```
Ans2_b_1 <- C[order(-C$`Consumer Satisfaction`),]
Ans2_b_1 <- Ans2_b_1 %>%
  select(`Plan Name`, State, `Consumer Satisfaction`)
head(Ans2_b_1,4)
```

```
##                                                 Plan Name State
## 25                          CDPHP Universal Benefits, Inc.    NY
## 81                                  Dean Health Plan, Inc.    WI
## 84                          Quartz Health Plan Corporation    WI
## 32 UnitedHealthcare Insurance Company of New York (Medicare)    NY
##    Consumer Satisfaction
## 25                   5.0
## 81                   5.0
## 84                   5.0
## 32                   4.5
```

```
Ans2_b_2 <- C[order(-C$Prevention),]
Ans2_b_2 <- Ans2_b_2 %>%
  select(`Plan Name`, State, Prevention)
head(Ans2_b_2,4)
```

```
##                                                 Plan Name State
## 1   Kaiser Foundation Health Plan Inc. - Southern California    CA
## 2  Kaiser Foundation Health Plan, Inc. - Northern California    CA
## 25                          CDPHP Universal Benefits, Inc.    NY
## 32 UnitedHealthcare Insurance Company of New York (Medicare)    NY
##    Prevention
## 1           5
## 2           5
## 25          5
## 32          5
```

```
Ans2_b_3 <- C[order(-C$Treatment),]
Ans2_b_3 <- Ans2_b_3 %>%
  select(`Plan Name`, State, Treatment)
head(Ans2_b_3,4)
```

```
##                                                      Plan Name State
## 81                                       Dean Health Plan, Inc.    WI
## 1        Kaiser Foundation Health Plan Inc. - Southern California    CA
## 2       Kaiser Foundation Health Plan, Inc. - Northern California    CA
## 61 Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.    VA
##    Treatment
## 81       5.0
## 1        4.5
## 2        4.5
## 61       4.5
```

Health Plan for Medicaid:

```r
Ans2_c_1 <- D[order(-D$`Consumer Satisfaction`),]
Ans2_c_1 <- Ans2_c_1 %>%
  select(`Plan Name`, State, `Consumer Satisfaction`)
head(Ans2_c_1,4)
```

```
##                                                  Plan Name State
## 38       Vista Health Plan DBA AmeriHealth Caritas Pennsylvania    PA
## 50                                          Optima Health Plan    VA
## 22 Excellus Health Plan, Inc. dba Excellus BlueCross BlueShield    NY
## 23            Excellus Health Plan, Inc. dba Univera Healthcare    NY
##    Consumer Satisfaction
## 38                   4.5
## 50                   4.5
## 22                   4.0
## 23                   4.0
```

```r
Ans2_c_2 <- D[order(-D$Prevention),]
Ans2_c_2 <- Ans2_c_2 %>%
  select(`Plan Name`, State, Prevention)
head(Ans2_c_2,4)
```

```
##                                  Plan Name State Prevention
## 1                    Alameda Alliance for Health    CA        4.5
## 37                         Health Partners Plans    PA        4.5
## 2                         Community Health Group    CA        4.0
## 4  Orange County Health Authority - dba CalOptima    CA        4.0
```

```r
Ans2_c_3 <- D[order(-D$Treatment),]
Ans2_c_3 <- Ans2_c_3 %>%
  select(`Plan Name`, State, Treatment)
head(Ans2_c_3,4)
```

```
##                                                  Plan Name State
## 21       Capital District Physicians' Health Plan, Inc. (CDPHP)    NY
## 22 Excellus Health Plan, Inc. dba Excellus BlueCross BlueShield    NY
## 23            Excellus Health Plan, Inc. dba Univera Healthcare    NY
## 38       Vista Health Plan DBA AmeriHealth Caritas Pennsylvania    PA
##    Treatment
## 21         4
## 22         4
## 23         4
## 38         4
```