IMT 573: Problem Set 4 - Data Analysis

LEE CHEN HSIN

Due: Tuesday, November 2, 2021

Collaborators:

Instructions: Before beginning this assignment, please ensure you have access to R and RStudio.

1. Download the `problemset4.Rmd` file from Canvas. Open `problemset4.Rmd` in RStudio and supply your solutions to the assignment by editing `problemset4.Rmd`.
2. Replace the "Insert Your Name Here" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.
3. All materials and resources that you use (with the exception of lecture slides) must be appropriately referenced within your assignment. In particular, note that Stack Overflow is licenses as Creative Commons (CC-BY-SA). This means you have to attribute any code you refer from SO.
4. Partial credit will be awarded for each question for which a serious attempt at finding an answer has been shown. But please **DO NOT** submit pages and pages of hard-to-read code and attempts that is impossible to grade. That is, avoid redundancy. Remember that one of the key goals of a data scientist is to produce coherent reports that others can easily follow. Students are *strongly* encouraged to attempt each question and to document their reasoning process even if they cannot find the correct answer. If you would like to include R code to show this process, but it does not run without errors you can do so with the `eval=FALSE` option as follows:

```
a + b # these object dont' exist
# if you run this on its own it with give an error
```

6. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit PDF`, rename the knitted PDF file to `ps4_ourLastName_YourFirstName.pdf`, and submit the PDF file on Canvas.
7. Collaboration is often fun and useful, but each student must turn in an individual write-up in their own words as well as code/work that is their own. Regardless of whether you work with others, what you turn in must be your own work; this includes code and interpretation of results. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.

Setup In this problem set you will need, at minimum, the following R packages.

```
# Load standard libraries
library(tidyverse)
library(grid)
```

Problem 1: 50 States in the USA In this problem we will use the `state` dataset, available as part of the R statistical computing platform. This data is related to the 50 states of the United States of America. Load the data and use it to answer the following questions.

```
data('state')
str(state.abb)
```

(a) Describe the data and each variable it contains. Tidy the data, preparing it for a data analysis.

```
##  chr [1:50] "AL" "AK" "AZ" "AR" "CA" "CO" "CT" "DE" "FL" "GA" "HI" "ID" ...
```

```
str(state.area)
```

```
##  num [1:50] 51609 589757 113909 53104 158693 ...
```

```
str(state.center)
```

```
## List of 2
##  $ x: num [1:50] -86.8 -127.2 -111.6 -92.3 -119.8 ...
##  $ y: num [1:50] 32.6 49.2 34.2 34.7 36.5 ...
```

```
str(state.division)
```

```
##  Factor w/ 9 levels "New England",..: 4 9 8 5 9 8 1 3 3 3 ...
```

```
str(state.name)
```

```
##  chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" "California" "Colorado" ...
```

```
str(state.region)
```

```
##  Factor w/ 4 levels "Northeast","South",..: 2 4 4 2 4 4 1 2 2 2 ...
```

```
str(state.x77)
```

```
##  num [1:50, 1:8] 3615 365 2212 2110 21198 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
##   ..$ : chr [1:8] "Population" "Income" "Illiteracy" "Life Exp" ...
```

```
data_state.x77 <- data.frame(state.x77)
tidy_data <- cbind(state.abb,state.area,state.center,state.division,state.name, data_state.x77, state.r

colnames(tidy_data)[1] <- "State"                # Rename first column
colnames(tidy_data)[2] <- "Region"              # Rename the 2th column
colnames(tidy_data)[5] <- "Division"
head(tidy_data)
```

```
##            State Region         x       y           Division state.name
## Alabama       AL  51609  -86.7509 32.5901 East South Central    Alabama
## Alaska        AK 589757 -127.2500 49.2500            Pacific     Alaska
## Arizona       AZ 113909 -111.6250 34.2192           Mountain    Arizona
## Arkansas      AR  53104  -92.2992 34.7336 West South Central   Arkansas
## California    CA 158693 -119.7730 36.5341            Pacific California
## Colorado      CO 104247 -105.5130 38.6777           Mountain   Colorado
##            Population Income Illiteracy Life.Exp Murder HS.Grad Frost   Area
## Alabama          3615   3624        2.1    69.05   15.1    41.3    20  50708
## Alaska            365   6315        1.5    69.31   11.3    66.7   152 566432
## Arizona          2212   4530        1.8    70.55    7.8    58.1    15 113417
## Arkansas         2110   3378        1.9    70.66   10.1    39.9    65  51945
## California      21198   5114        1.1    71.71   10.3    62.6    20 156361
## Colorado         2541   4884        0.7    72.06    6.8    63.9   166 103766
```

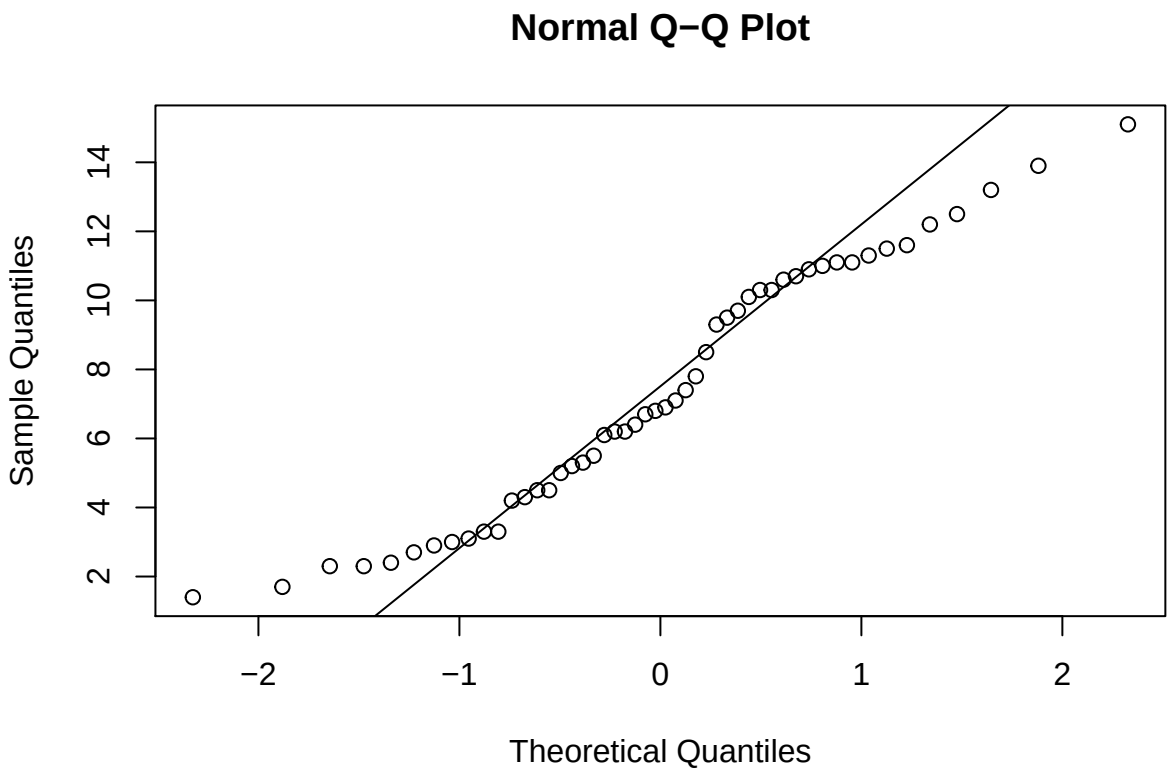
```
##            state.region
## Alabama           South
## Alaska             West
## Arizona            West
## Arkansas          South
## California         West
## Colorado           West
```

```
summary(tidy_data)
```

```
##      State        Region             x                y
##  AK     : 1   Min.   :  1214   Min.   :-127.25   Min.   :27.87
##  AL     : 1   1st Qu.: 37317   1st Qu.:-104.16   1st Qu.:35.55
##  AR     : 1   Median : 56222   Median : -89.90   Median :39.62
##  AZ     : 1   Mean   : 72368   Mean   : -92.46   Mean   :39.41
##  CA     : 1   3rd Qu.: 83234   3rd Qu.: -78.98   3rd Qu.:43.14
##  CO     : 1   Max.   :589757   Max.   : -68.98   Max.   :49.25
##  (Other):44
##                Division       state.name   Population        Income
##  South Atlantic    : 8   Alabama    : 1   Min.   :  365   Min.   :3098
##  Mountain          : 8   Alaska     : 1   1st Qu.: 1080   1st Qu.:3993
##  West North Central: 7   Arizona    : 1   Median : 2838   Median :4519
##  New England       : 6   Arkansas   : 1   Mean   : 4246   Mean   :4436
##  East North Central: 5   California: 1   3rd Qu.: 4968   3rd Qu.:4814
##  Pacific           : 5   Colorado   : 1   Max.   :21198   Max.   :6315
##  (Other)           :11   (Other)    :44
##    Illiteracy       Life.Exp         Murder          HS.Grad
##  Min.   :0.500   Min.   :67.96   Min.   : 1.400   Min.   :37.80
##  1st Qu.:0.625   1st Qu.:70.12   1st Qu.: 4.350   1st Qu.:48.05
##  Median :0.950   Median :70.67   Median : 6.850   Median :53.25
##  Mean   :1.170   Mean   :70.88   Mean   : 7.378   Mean   :53.11
##  3rd Qu.:1.575   3rd Qu.:71.89   3rd Qu.:10.675   3rd Qu.:59.15
##  Max.   :2.800   Max.   :73.60   Max.   :15.100   Max.   :67.30
##
##      Frost             Area               state.region
##  Min.   :  0.00   Min.   :  1049   Northeast    : 9
##  1st Qu.: 66.25   1st Qu.: 36985   South        :16
##  Median :114.50   Median : 54277   North Central:12
##  Mean   :104.46   Mean   : 70736   West         :13
##  3rd Qu.:139.75   3rd Qu.: 81162
##  Max.   :188.00   Max.   :566432
##
```

```
# Q-Q plot: Do the sample quantiles almost fall into a straight line? If yes, then the variable more li
qqnorm(tidy_data$Murder)
qqline(tidy_data$Murder)
```

(b) Suppose you want to explore the relationship between a state's `Murder` rate and other characteristics of the state, for example population, illiteracy rate, and more. Begin by examining the bivariate relationships present in the data. What does your analysis suggest might be important varibles to consider in building a model to explain variation in murder rates?




```
# To visualize the linear relationship among variables in a plot, a scatter matrix is the best choice.

library(ellipse)
```

```
##
## Attaching package: 'ellipse'
```

```
## The following object is masked from 'package:graphics':
##
##     pairs
```

```
library(corrplot)
```

```
## corrplot 0.90 loaded
```

```
# Negative correlation
cor(tidy_data$Murder, tidy_data$Income)
```

```
## [1] -0.2300776
```

```
cor(tidy_data$Murder, tidy_data$Life.Exp)
```

```
## [1] -0.7808458
```

```
# Positive correlation
cor(tidy_data$Murder, tidy_data$Illiteracy)
```

```
## [1] 0.7029752
```

```
cor(tidy_data$Murder, tidy_data$Population)
```

```
## [1] 0.3436428
```

```
#According to the analysis, murder rate has a negative correlation with Income
#and life expectation;murder rate has a positive correlation with Illiteracy.
```

```
#How many percent of state has the high murder rate(>=10%)?
Murder=as.integer(tidy_data$Murder)
murderratebigthan10=filter(tidy_data,Murder>=10)
murderratebigthan10
```

(c) Develop a new research question of your own that you can address using the state dataset. Clearly state the question you are going to address. Provide at least one visualization to support your exploration of this question. Discuss what you find.

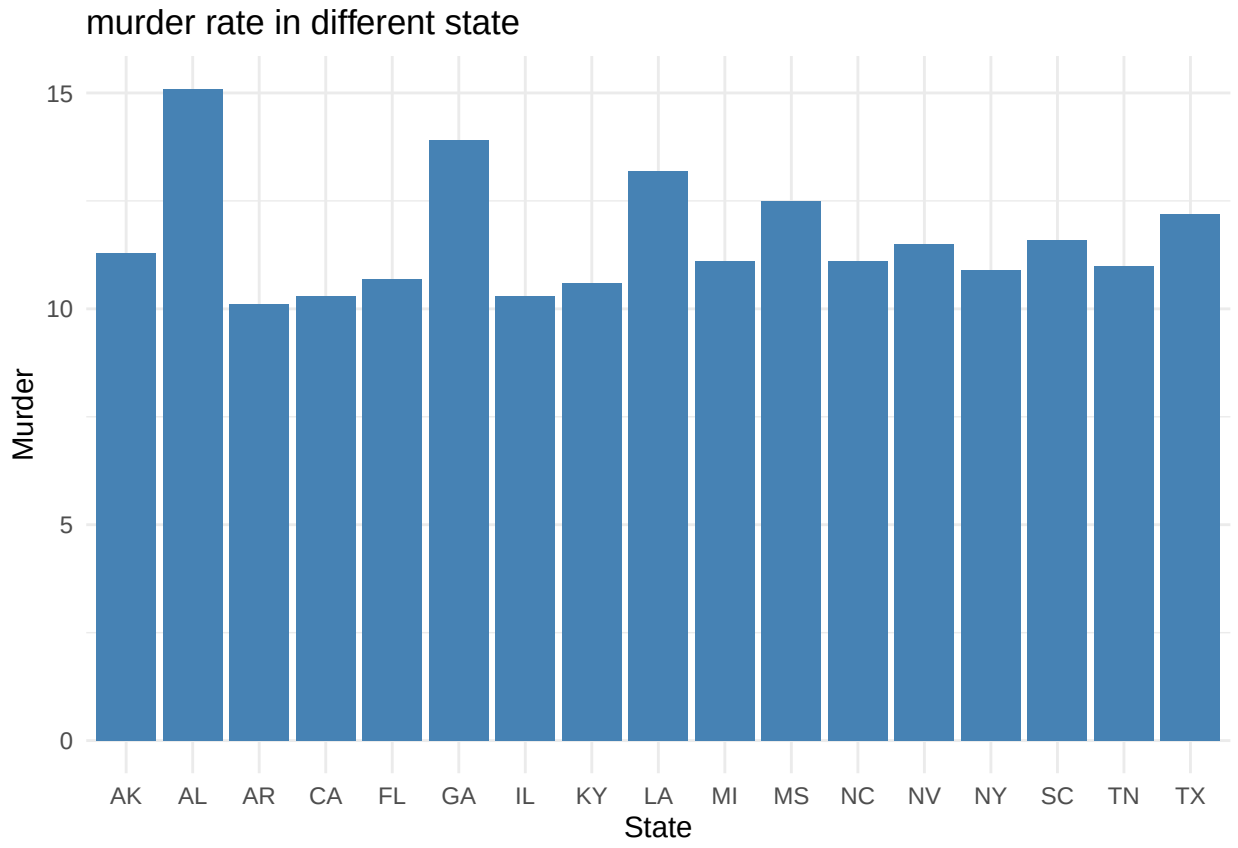
```
##                State Region         x       y           Division     state.name
## Alabama           AL  51609  -86.7509 32.5901 East South Central        Alabama
## Alaska            AK 589757 -127.2500 49.2500            Pacific         Alaska
## Arkansas          AR  53104  -92.2992 34.7336 West South Central       Arkansas
## California        CA 158693 -119.7730 36.5341            Pacific     California
## Florida           FL  58560  -81.6850 27.8744     South Atlantic        Florida
## Georgia           GA  58876  -83.3736 32.3329     South Atlantic        Georgia
## Illinois          IL  56400  -89.3776 40.0495 East North Central       Illinois
## Kentucky          KY  40395  -84.7674 37.3915 East South Central       Kentucky
## Louisiana         LA  48523  -92.2724 30.6181 West South Central      Louisiana
## Michigan          MI  58216  -84.6870 43.1361 East North Central       Michigan
## Mississippi       MS  47716  -89.8065 32.6758 East South Central    Mississippi
## Nevada            NV 110540 -116.8510 39.1063           Mountain         Nevada
## New York          NY  49576  -75.1449 43.1361    Middle Atlantic       New York
## North Carolina    NC  52586  -78.4686 35.4195     South Atlantic North Carolina
## South Carolina    SC  31055  -80.5056 33.6190     South Atlantic South Carolina
## Tennessee         TN  42244  -86.4560 35.6767 East South Central      Tennessee
## Texas             TX 267339  -98.7857 31.3897 West South Central          Texas
##                Population Income Illiteracy Life.Exp Murder HS.Grad Frost
## Alabama              3615   3624        2.1    69.05   15.1    41.3    20
## Alaska                365   6315        1.5    69.31   11.3    66.7   152
## Arkansas             2110   3378        1.9    70.66   10.1    39.9    65
## California          21198   5114        1.1    71.71   10.3    62.6    20
## Florida              8277   4815        1.3    70.66   10.7    52.6    11
## Georgia              4931   4091        2.0    68.54   13.9    40.6    60
## Illinois            11197   5107        0.9    70.14   10.3    52.6   127
## Kentucky             3387   3712        1.6    70.10   10.6    38.5    95
## Louisiana            3806   3545        2.8    68.76   13.2    42.2    12
## Michigan             9111   4751        0.9    70.63   11.1    52.8   125
## Mississippi          2341   3098        2.4    68.09   12.5    41.0    50
## Nevada                590   5149        0.5    69.03   11.5    65.2   188
## New York            18076   4903        1.4    70.55   10.9    52.7    82
## North Carolina       5441   3875        1.8    69.21   11.1    38.5    80
## South Carolina       2816   3635        2.3    67.96   11.6    37.8    65
## Tennessee            4173   3821        1.7    70.11   11.0    41.8    70
## Texas               12237   4188        2.2    70.90   12.2    47.4    35
##                  Area  state.region
## Alabama         50708         South
```

```
## Alaska          566432          West
## Arkansas         51945         South
## California      156361          West
## Florida          54090         South
## Georgia          58073         South
## Illinois         55748 North Central
## Kentucky         39650         South
## Louisiana        44930         South
## Michigan         56817 North Central
## Mississippi      47296         South
## Nevada          109889          West
## New York         47831     Northeast
## North Carolina   48798         South
## South Carolina   30225         South
## Tennessee        41328         South
## Texas           262134         South
```

```
murderratebigthan10rate=nrow(murderratebigthan10)/nrow(tidy_data)
murderratebigthan10rate
```

```
## [1] 0.34
```

```
ggplot(data=murderratebigthan10, aes(x=State, y=Murder)) +
  geom_bar(stat="identity", fill="steelblue")+
  theme_minimal()+ ggtitle("murder rate in different state")+xlab("State") + ylab("Murder")
```



Problem 2: Asking Data Science Questions: Crime and Educational Attainment In Problem Set 3, you joined data about crimes and educational attainment. Here you will use this new combined dataset to examine questions around crimes in Seattle and the educational attainment of people living in the areas in which the crime occurred.

(a) Develop a Data Science Question Develop your own question to address in this analysis. Your question should be specific and measurable, and it should be able to be addressed through a basic analysis of the crime dataset you compiled in Problem Set 3.

```
load("/Users/leechenhsin/Desktop/study@USA/07_UW_School/IMT573/beat_census.RData")

#Which kinds of educational level has a higher crime total number?
#Find the correlation between educational level and total crime number

subset_beat_census=as.data.frame(beat_census[,c(4,24:28)])
str(subset_beat_census)
```

```
## 'data.frame':    33 obs. of  6 variables:
##  $ total                     : int  2045 4155 3095 3524 4487 3524 2122 2806 4145 5077 ...
##  $ associates_degree         : int  114 310 295 261 106 111 127 106 244 362 ...
##  $ bachelors_degree          : int  661 1301 1360 1391 1786 1310 845 1175 1794 2122 ...
##  $ masters_degree            : int  461 760 560 748 985 864 351 659 998 841 ...
##  $ professional_school_degree: int  76 64 102 205 323 212 175 144 190 257 ...
##  $ doctorate_degree          : int  61 137 85 254 288 97 164 134 158 217 ...
```

```
i <- c(1:6)
# Specify own function within apply
subset_beat_census[,i] <- apply(subset_beat_census[,i],2,
                    function(x) as.numeric(x))
corMatrix <- cor(subset_beat_census)
```

(b) Describe and Summarize Briefly summarize the dataset, describing what data exists and its basic properties. Comment on any issues that need to be resolved before you can proceed with your analysis.

```
summary(beat_census)
```

```
##                   GEO.id      GEO.id2
##  1400000US53033009300: 2   Min.   :5.303e+10
##  1400000US53033010800: 2   1st Qu.:5.303e+10
##  1400000US53033001100: 1   Median :5.303e+10
##  1400000US53033001400: 1   Mean   :5.303e+10
##  1400000US53033001701: 1   3rd Qu.:5.303e+10
##  1400000US53033002600: 1   Max.   :5.303e+10
##  (Other)             :25
##                                    GEO.display.label     total
##  Census Tract 108, King County, Washington   : 2      Min.   : 939
##  Census Tract 93, King County, Washington    : 2      1st Qu.:2806
##  Census Tract 100.01, King County, Washington: 1      Median :3416
##  Census Tract 102, King County, Washington   : 1      Mean   :3390
##  Census Tract 109, King County, Washington   : 1      3rd Qu.:4145
##  Census Tract 11, King County, Washington    : 1      Max.   :5424
##  (Other)                                     :25
##   no_schooling    nursery_school    kindergarten      X1st_grade
##  Min.   :  0.00   Min.   :0.0000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:  0.00   1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.: 0.000
```

```
##  Median : 15.00   Median :0.0000   Median : 0.000   Median : 0.000
##  Mean   : 49.76   Mean   :0.1818   Mean   : 2.576   Mean   : 3.091
##  3rd Qu.: 68.00   3rd Qu.:0.0000   3rd Qu.: 0.000   3rd Qu.: 0.000
##  Max.   :333.00   Max.   :6.0000   Max.   :44.000   Max.   :61.000
##
##    X2nd_grade       X3rd_grade      X4th_grade       X5th_grade
##  Min.   : 0.000   Min.   : 0.00   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 0.000   1st Qu.: 0.000
##  Median : 0.000   Median : 0.00   Median : 0.000   Median : 0.000
##  Mean   : 1.909   Mean   :11.18   Mean   : 8.121   Mean   : 7.212
##  3rd Qu.: 0.000   3rd Qu.:15.00   3rd Qu.: 4.000   3rd Qu.: 4.000
##  Max.   :49.000   Max.   :79.00   Max.   :87.000   Max.   :80.000
##
##    X6th_grade       X7th_grade       X8th_grade       X9th_grade
##  Min.   :  0.00   Min.   : 0.000   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:  0.00   1st Qu.: 0.000   1st Qu.:  0.00   1st Qu.:  1.00
##  Median :  3.00   Median : 0.000   Median :  0.00   Median : 17.00
##  Mean   : 22.85   Mean   : 6.606   Mean   : 21.91   Mean   : 23.18
##  3rd Qu.: 24.00   3rd Qu.: 7.000   3rd Qu.: 33.00   3rd Qu.: 31.00
##  Max.   :243.00   Max.   :62.000   Max.   :136.00   Max.   :135.00
##
##   X10th_grade     X11th_grade     X12th_grade_no_diploma high_school_diploma
##  Min.   :  0.0   Min.   :  0.00   Min.   :  0.00         Min.   : 23.0
##  1st Qu.:  0.0   1st Qu.:  0.00   1st Qu.:  8.00         1st Qu.:160.0
##  Median :  5.0   Median : 19.00   Median : 34.00         Median :207.0
##  Mean   : 16.3   Mean   : 23.64   Mean   : 38.18         Mean   :275.4
##  3rd Qu.: 17.0   3rd Qu.: 38.00   3rd Qu.: 63.00         3rd Qu.:383.0
##  Max.   :110.0   Max.   :141.00   Max.   :119.00         Max.   :653.0
##
##  ged_or_alternative_credential some_college_less_than_1_year
##  Min.   :  0.00                Min.   : 30.0
##  1st Qu.: 15.00                1st Qu.: 84.0
##  Median : 53.00                Median :134.0
##  Mean   : 58.27                Mean   :128.4
##  3rd Qu.: 88.00                3rd Qu.:155.0
##  Max.   :165.00                Max.   :338.0
##
##  some_college_1_or_more_years_no_degree associates_degree bachelors_degree
##  Min.   : 111.0                         Min.   :106.0     Min.   : 348
##  1st Qu.: 304.0                         1st Qu.:145.0     1st Qu.: 803
##  Median : 401.0                         Median :212.0     Median :1175
##  Mean   : 428.2                         Mean   :233.2     Mean   :1178
##  3rd Qu.: 512.0                         3rd Qu.:281.0     3rd Qu.:1554
##  Max.   :1289.0                         Max.   :551.0     Max.   :2122
##
##  masters_degree professional_school_degree doctorate_degree  digitcode
##  Min.   : 100   Min.   :  7                Min.   :  0.0    Length:33
##  1st Qu.: 351   1st Qu.: 81                1st Qu.: 56.0    Class :character
##  Median : 563   Median :144                Median :110.0    Mode  :character
##  Mean   : 562   Mean   :173                Mean   :116.9
##  3rd Qu.: 760   3rd Qu.:212                3rd Qu.:169.0
##  Max.   :1030   Max.   :543                Max.   :288.0
##
##       Name                                   Location.1    Latitude
```

```
##  B1     : 1   (47.5254502461741, -122.365817548329): 1   Min.   :47.53
##  B3     : 1   (47.5261052985115, -122.336388313318): 1   1st Qu.:47.57
##  C1     : 1   (47.5345836385751, -122.303020266287): 1   Median :47.61
##  C3     : 1   (47.5439339496481, -122.286476209963): 1   Mean   :47.61
##  D2     : 1   (47.5478566154038, -122.361787408364): 1   3rd Qu.:47.66
##  D3     : 1   (47.5484146593035, -122.354809670155): 1   Max.   :47.71
##  (Other):27   (Other)                                :27
##    Longitude      geoloaction         Statecode         Countrycode
##  Min.   :-122.4   Length:33          Length:33          Length:33
##  1st Qu.:-122.4   Class :character   Class :character   Class :character
##  Median :-122.3   Mode  :character   Mode  :character   Mode  :character
##  Mean   :-122.3
##  3rd Qu.:-122.3
##  Max.   :-122.3
##
```

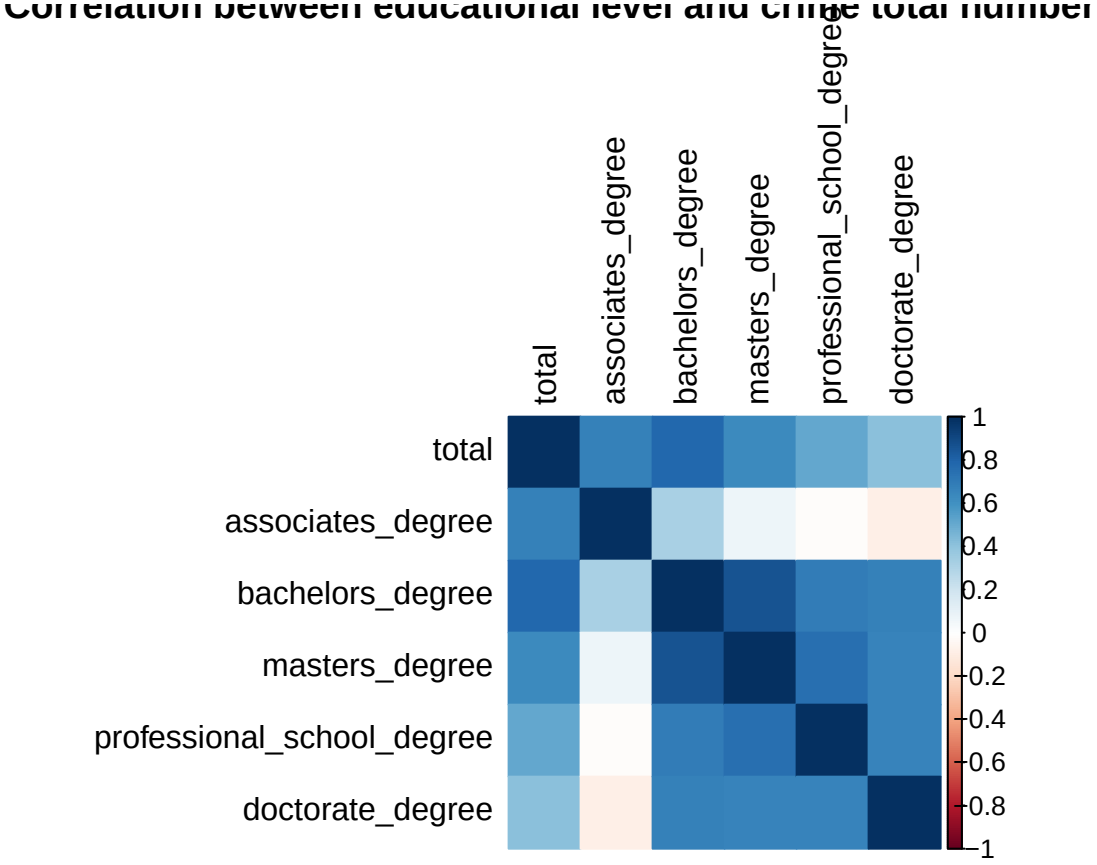
```
summary(subset_beat_census)
```

```
##      total      associates_degree bachelors_degree masters_degree
##  Min.   : 939   Min.   :106.0     Min.   : 348     Min.   : 100
##  1st Qu.:2806   1st Qu.:145.0     1st Qu.: 803     1st Qu.: 351
##  Median :3416   Median :212.0     Median :1175     Median : 563
##  Mean   :3390   Mean   :233.2     Mean   :1178     Mean   : 562
##  3rd Qu.:4145   3rd Qu.:281.0     3rd Qu.:1554     3rd Qu.: 760
##  Max.   :5424   Max.   :551.0     Max.   :2122     Max.   :1030
##  professional_school_degree doctorate_degree
##  Min.   :  7                Min.   :  0.0
##  1st Qu.: 81                1st Qu.: 56.0
##  Median :144                Median :110.0
##  Mean   :173                Mean   :116.9
##  3rd Qu.:212                3rd Qu.:169.0
##  Max.   :543                Max.   :288.0
```

```
#Some columns in the database are not numeric so before conducting the data
#analysis, it is necessary to convert them into numeric data type.
```

(c) Data Analysis Use the dataset to provide empirical evidence that addressed your question from part (a). Discuss your results. Provide at least one visualization to support your narrative.

```
# The variable total has the most positive correlation with bachelors_degree
#0.78.However, the variables of associates_degree and masters_degree are also important.

library(corrplot)
corrplot(corMatrix, method="color",tl.col="black",title ="Correlation between educational level and cri
```



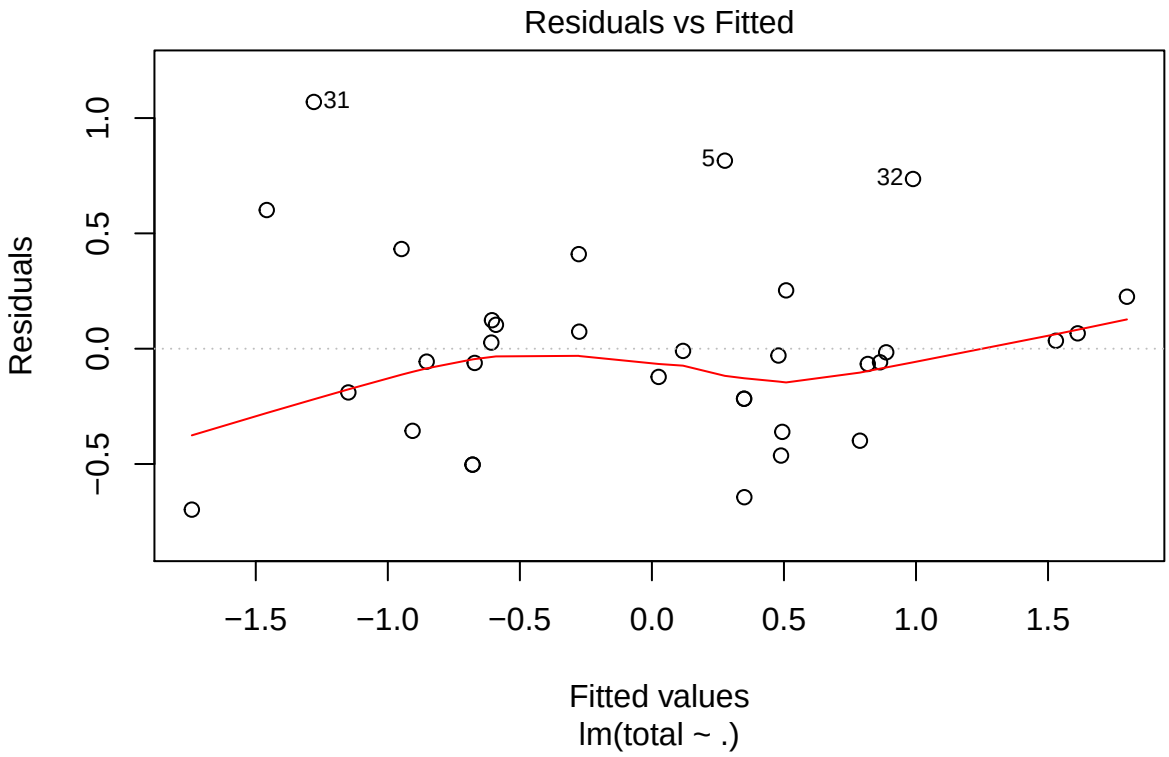

```
# scale function is used to normalization
set.seed(100)
z_sub_beat_census <- as.data.frame(scale(subset_beat_census))
fit.lm <- lm(total ~ ., data=z_sub_beat_census)

coef(fit.lm)
```

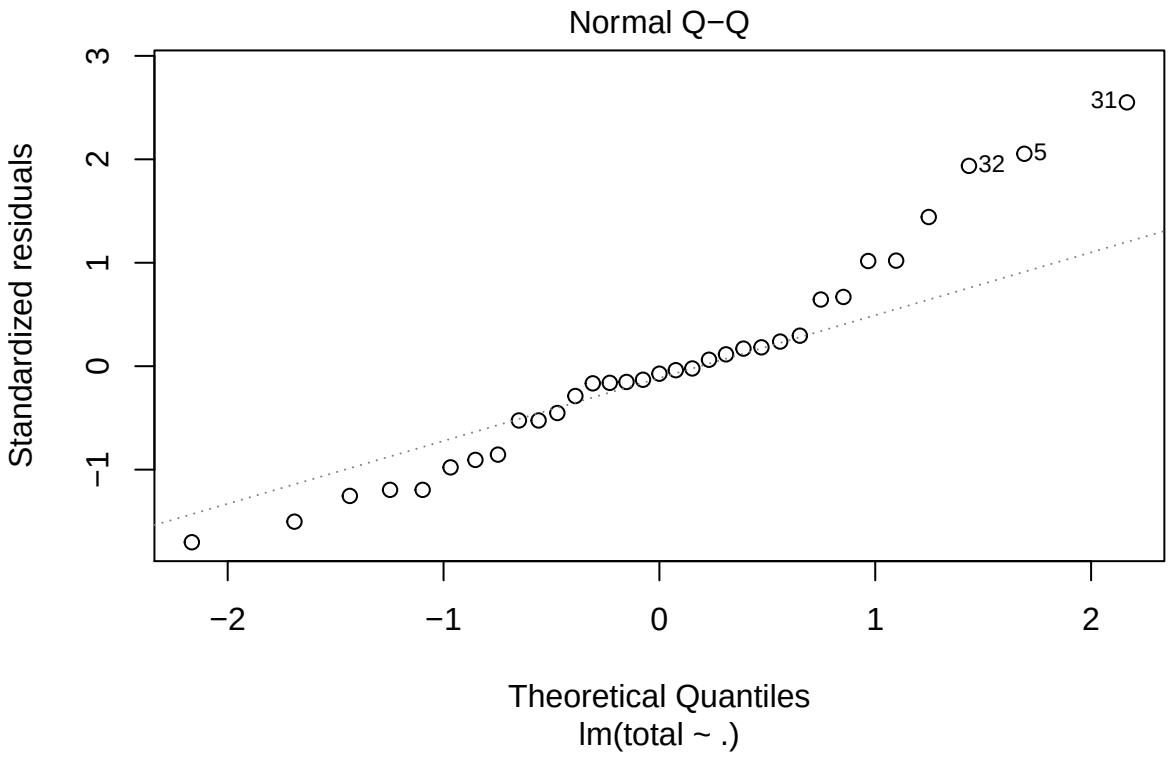
```
##                (Intercept)          associates_degree
##              -2.532432e-16              5.610916e-01
##           bachelors_degree             masters_degree
##               3.083639e-01              2.033184e-01
## professional_school_degree          doctorate_degree
##               1.423894e-01              2.225825e-02
```

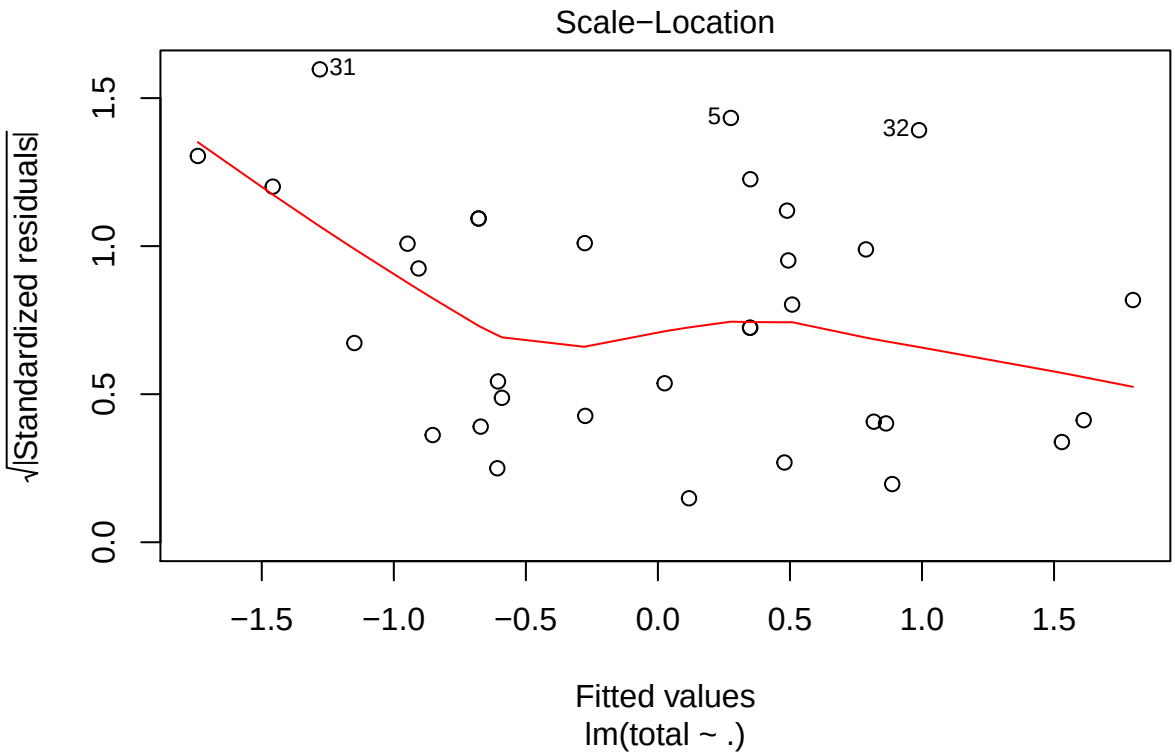
```
round(coef(fit.lm),digits=2)
```

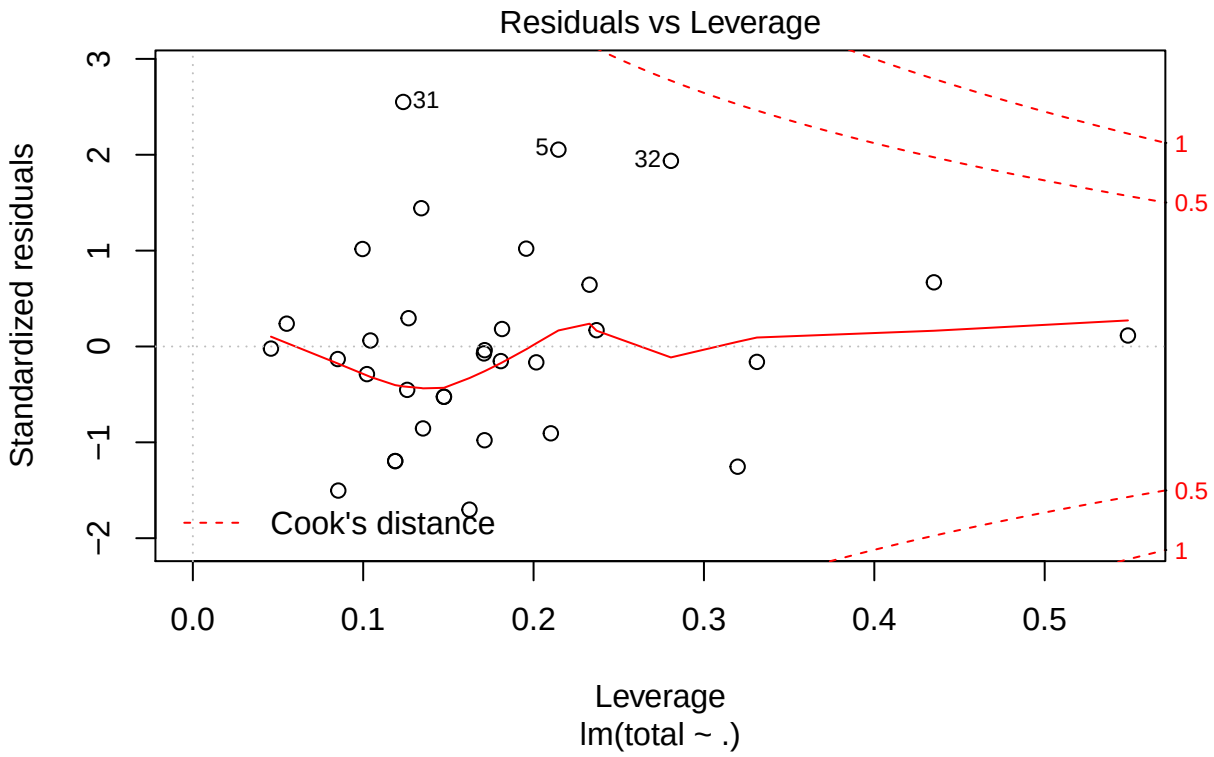
```
##                (Intercept)          associates_degree
##                       0.00                       0.56
##           bachelors_degree             masters_degree
##                       0.31                       0.20
## professional_school_degree          doctorate_degree
##                       0.14                       0.02
```

```
plot(fit.lm)
```


Residuals vs Fitted
Residuals
Fitted values
lm(total ~ .)
31
5
32
1.0
0.5
0.0
−0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5


Normal Q−Q
Standardized residuals
3
2
1
0
−1
31
32
5
−2
−1
0
1
2
Theoretical Quantiles
lm(total ~ .)


Scale−Location
31
5
32
√|Standardized residuals|
0.0
0.5
1.0
1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Fitted values
lm(total ~ .)




```
summary(fit.lm)
```

```
##
## Call:
## lm(formula = total ~ ., data = z_sub_beat_census)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6978 -0.2170 -0.0296  0.1235  1.0697
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -2.532e-16  7.798e-02   0.000    1.000
## associates_degree           5.611e-01  1.016e-01   5.521 7.54e-06 ***
## bachelors_degree            3.084e-01  2.039e-01   1.513    0.142
## masters_degree              2.033e-01  1.835e-01   1.108    0.278
## professional_school_degree  1.424e-01  1.291e-01   1.103    0.280
## doctorate_degree            2.226e-02  1.237e-01   0.180    0.859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4479 on 27 degrees of freedom
## Multiple R-squared:  0.8307, Adjusted R-squared:  0.7993
## F-statistic:  26.5 on 5 and 27 DF,  p-value: 1.272e-09
```

```
# Multiple R-squared:  0.8061,  Adjusted R-squared:  0.7932
# The variable of associates_degree is significant
```

(d) Reflect and Question Comment the questions (and answers) in this analysis. Were you able to answer all of these questions? Are all questions well defined? Is the data good enough to answer all these?

```
# From the functions of avPlots and crPlots, there are some outliers in dataset.
# Thereafter, We can process these outliers or use nonlinear regression in the
# near future.
```