Homework 1-STAT 6365E

Shuangyan Wu (811349623)

6/24/2022

Contents

Q1

(a)

Plot the violent crime rate versus time:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```r
library(readxl)
data1<-read_excel("table-1.xls", skip=3,col_names = TRUE,na="")
```

```
## New names:
## * `` -> `...23`
## * `` -> `...24`
```

```r
crime<-data1[1:(31-3-7-1),1:22]
crime$Year[crime$Year==20015]<-2001
crime$Year[crime$Year==20166]<-2016
colnames(crime) <- c("Year","Population","VC","VCR","MNM","MNMR" , "RP_RD" ,"RR_RD",
                    "RP_LD","RR_LD",
  "RB","RBR","AA", "AAR","PC","PCR", "BG","BGR","LT","LTR","MVT","MVTR")
ggplot(data=crime) +
  geom_line(mapping = aes(x=Year, y=VCR, group=1), color="black")+
  geom_point(mapping = aes(x=Year, y=VCR, group=1))+
  ggtitle("Violent Crime Rate vs Year")+
  ylab("Violent Crime Rate")
```


(b)
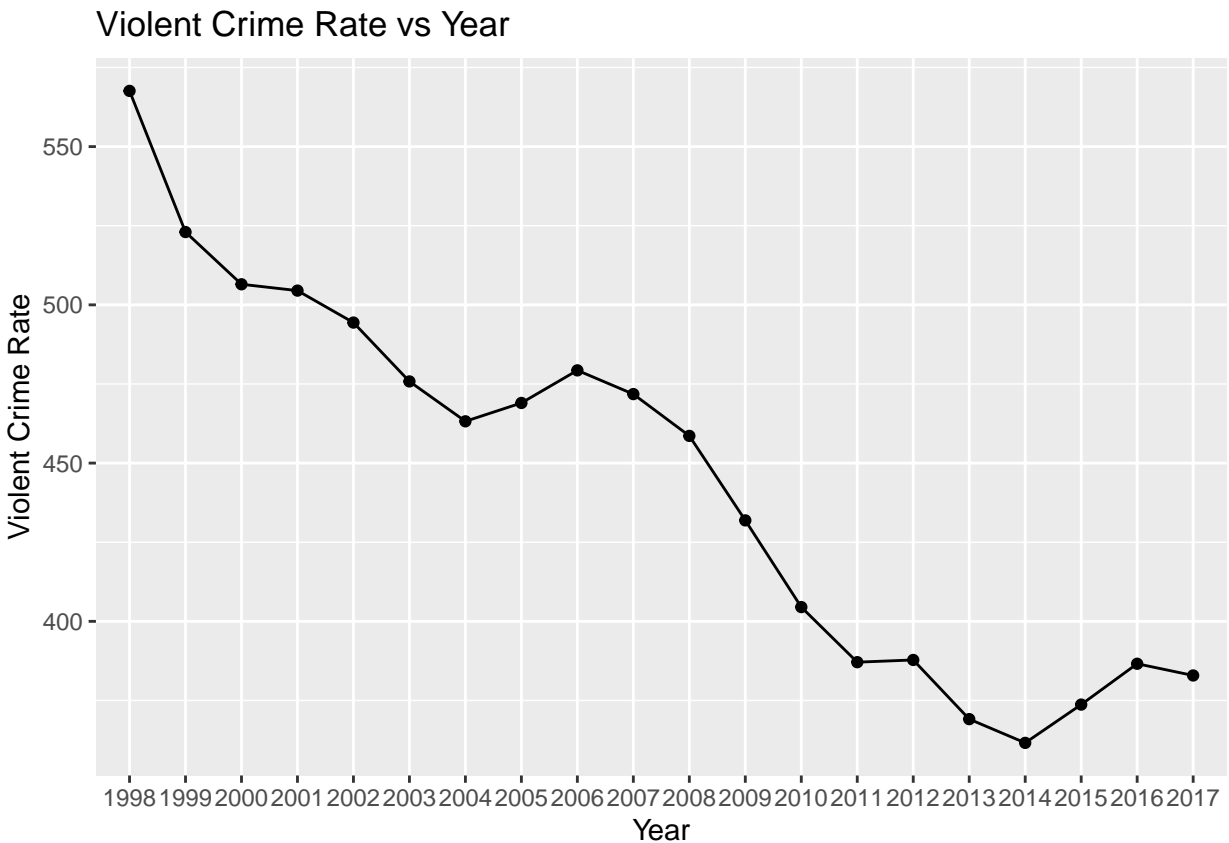
```r
options(warn = - 1)
ggplot(data=crime,aes(x = Year)) +
  geom_line(aes(y=MNMR,group=1,colour="Murder and NM"),size = 1)+
  geom_line(aes(y=RR_RD,group=1,colour="Rape_RD"),size = 1)+
```

```
    geom_line(aes(y=RR_LD,group=1,colour="Rape_LD"),size = 1)+
    geom_line(aes(y=RBR,group=1,colour="Robbery"),size = 1)+
    geom_line(aes(y=AAR,group=1,colour="Aggravated assault"),size = 1)+
    geom_line(aes(y=VCR,group=1,colour="Violent crime"),size = 1)+
    scale_colour_manual("", breaks = c("Murder and NM", "Rape_RD", "Rape_LD","Robbery",
    "Aggravated assault", "Violent crime"),
    values = c("black", "blue",
"yellow","brown","green","red"))+
    theme(legend.position="bottom")+
    ggtitle("Violent Crime Rates vs Year")+ ylab("Rate")
```
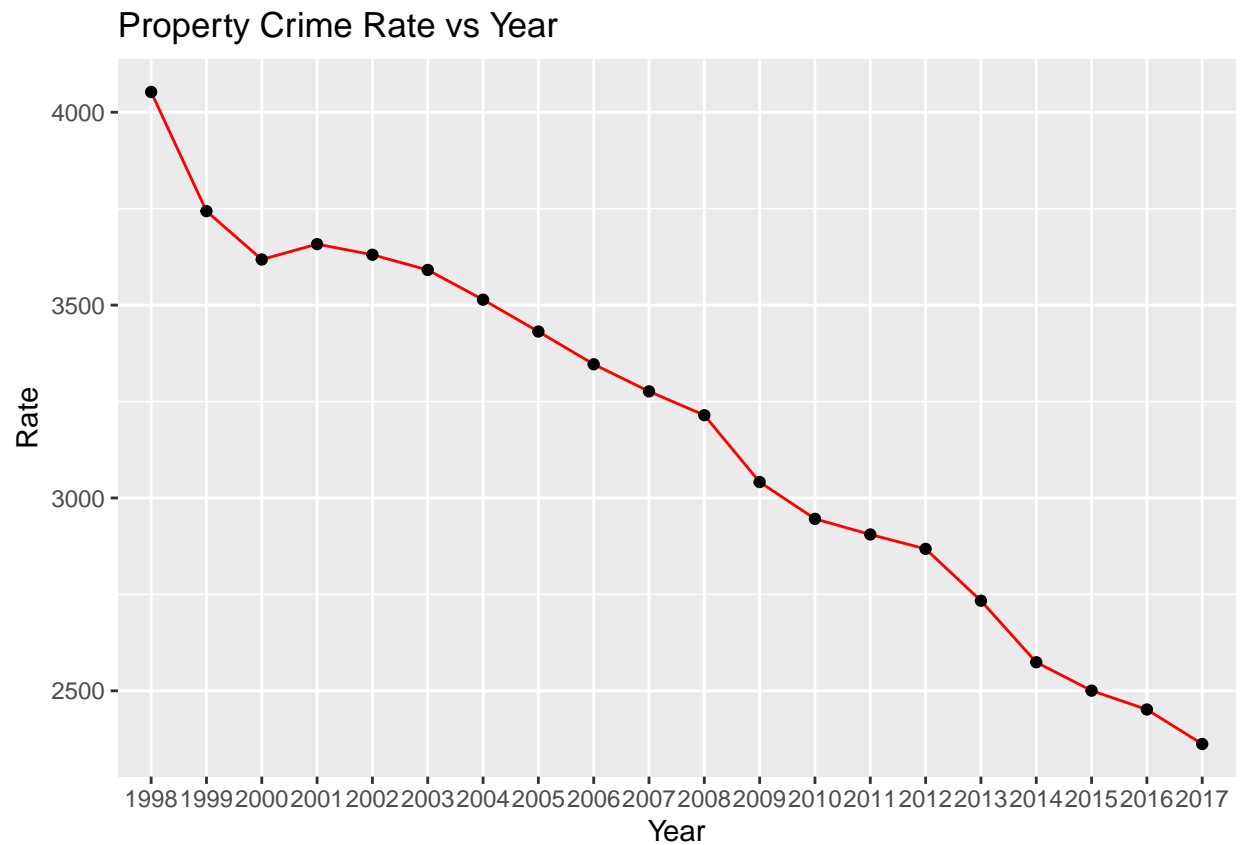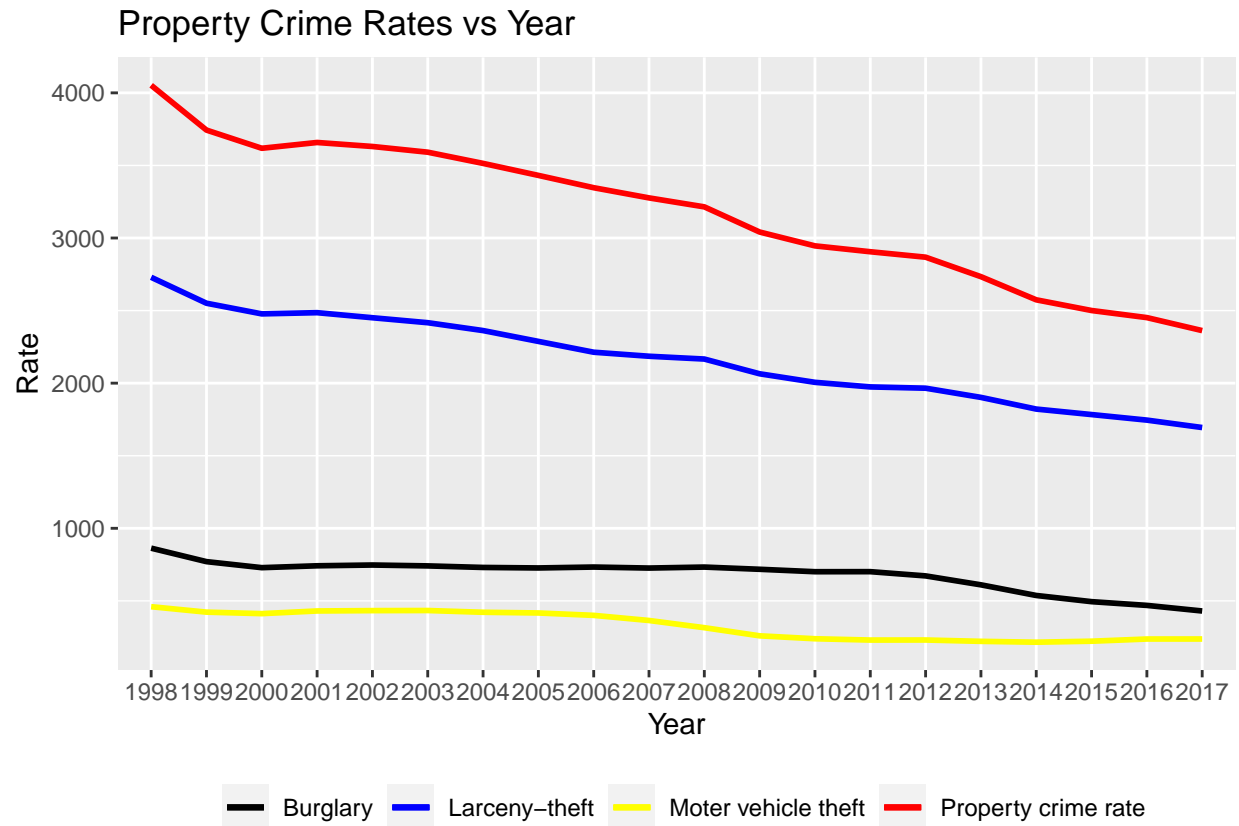

Violent Crime Rates vs Year
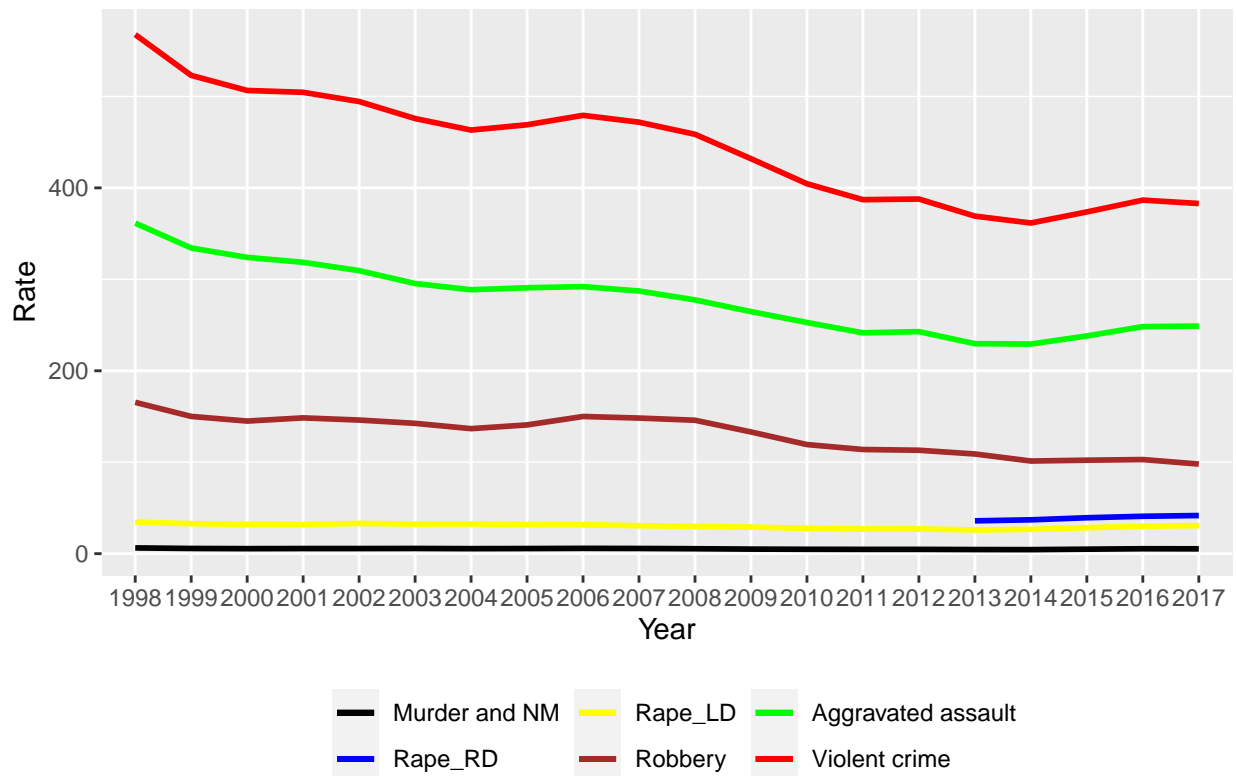
(c)

```
ggplot(data=crime) +
  geom_line(mapping = aes(x=Year, y=PCR, group=1), color="red")+
  geom_point(mapping = aes(x=Year, y=PCR, group=1))+
  ggtitle("Property Crime Rate vs Year")+
  ylab("Rate")
```

Property Crime Rate vs Year

(d)

```
ggplot(data=crime,aes(x = Year)) +
  geom_line(aes(y=BGR,group=1,colour="Burglary"),size = 1)+
  geom_line(aes(y=LTR,group=1,colour="Larceny-theft"),size = 1)+
  geom_line(aes(y=MVTR,group=1,colour="Moter vehicle theft"),size = 1)+
  geom_line(aes(y=PCR,group=1,colour="Property crime rate"),size = 1)+
  scale_colour_manual("", breaks = c("Burglary", "Larceny-theft",
"Moter vehicle theft","Property crime rate"),
values = c("black", "blue", "yellow","red"))+
  theme(legend.position="bottom")+
  ggtitle("Property Crime Rates vs Year")+ylab("Rate")
```

Property Crime Rates vs Year

(e)

From the violent crime rate vs year plot in 1(a), it is observed that the violent crime rate decreases overall from year 1998 to 2017. While some fluctuations can be seen around 2005-2006 and 2015-2017. From the graph for different types of violent crime rates along years in 1(b), it is seen that the overall crime rate order is: aggravated assault rate > robbery rate > rape (legacy definition) rate> murder and nonnegligent manslaughter rate. The rape (revised definition) rate has limited amount of data. It shows close number to the rape (legacy definition) rate. Similar to the total violent crime rate, assault rate and robbery rate both show decreasing trends along years.

For the Property Crime Rate vs Year graph in 1(c), an overall decreasing trend along years from 1989 to 2017 is also observed. By comparing different types of property crime rates along years in 1(d), the order of those property crime rates is found to be: Larceny-theft rate > Burglary rate > Motor Vehicle Theft rate. A decreasing rate can be seen for all three types of property crime along years to different extents. The decreasing rates for violent crime rates and property crime rates could be suggesting an improved public security, education, and economy.

Q2

(a)

```
score<-read_csv("MERGED2015_16_PP.csv",na=c(" ","NA","NULL","Privacy_Suppressed"),
              show_col_types = FALSE)
show_col_types = FALSE
```

```r
score_t<-tibble(score)
library(dplyr)
library(ggpubr)
score_1<-score_t %>% rowwise() %>% mutate(percentage_p1 = sum(NUM41_PUB,NUM41_PRIV,
na.rm = TRUE)/sum(NUM41_PUB,NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
p1<-ggplot(score_1, aes(x=percentage_p1)) + geom_histogram(binwidth=0.03)+
  xlab("Percentage")+ggtitle("Student %-family income $0-$30,000");

score_2<-score_1 %>% rowwise() %>% mutate(percentage_p2 = sum(NUM42_PUB,NUM42_PRIV,
na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
p2<-ggplot(score_2, aes(x=percentage_p2)) + geom_histogram(binwidth=0.03)+
  xlab("Percentage")+ggtitle("Student %-family income $30,001-$48,000");

score_3<-score_2 %>% rowwise() %>% mutate(percentage_p3 = sum(NUM43_PUB,NUM43_PRIV,
na.rm = TRUE)/sum(NUM41_PUB,
                  NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
p3<-ggplot(score_3, aes(x=percentage_p3)) + geom_histogram(binwidth=0.03)+
  xlab("Percentage")+ggtitle("Student %-family income $48,001-$75,000");

score_4<-score_3 %>% rowwise() %>% mutate(percentage_p4 = sum(NUM44_PUB,NUM44_PRIV,
 na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
p4<-ggplot(score_4, aes(x=percentage_p4)) + geom_histogram(binwidth=0.03)+
  xlab("Percentage")+ggtitle("Student %-family income $75,001-$110,000");

score_5<-score_4 %>% rowwise() %>% mutate(percentage_p5 = sum(NUM45_PUB,NUM45_PRIV,
 na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
p5<-ggplot(score_5, aes(x=percentage_p5)) + geom_histogram(binwidth=0.03)+
  xlab("Percentage")+ggtitle("Student %-family income $110,000+");

library(gridExtra)
```
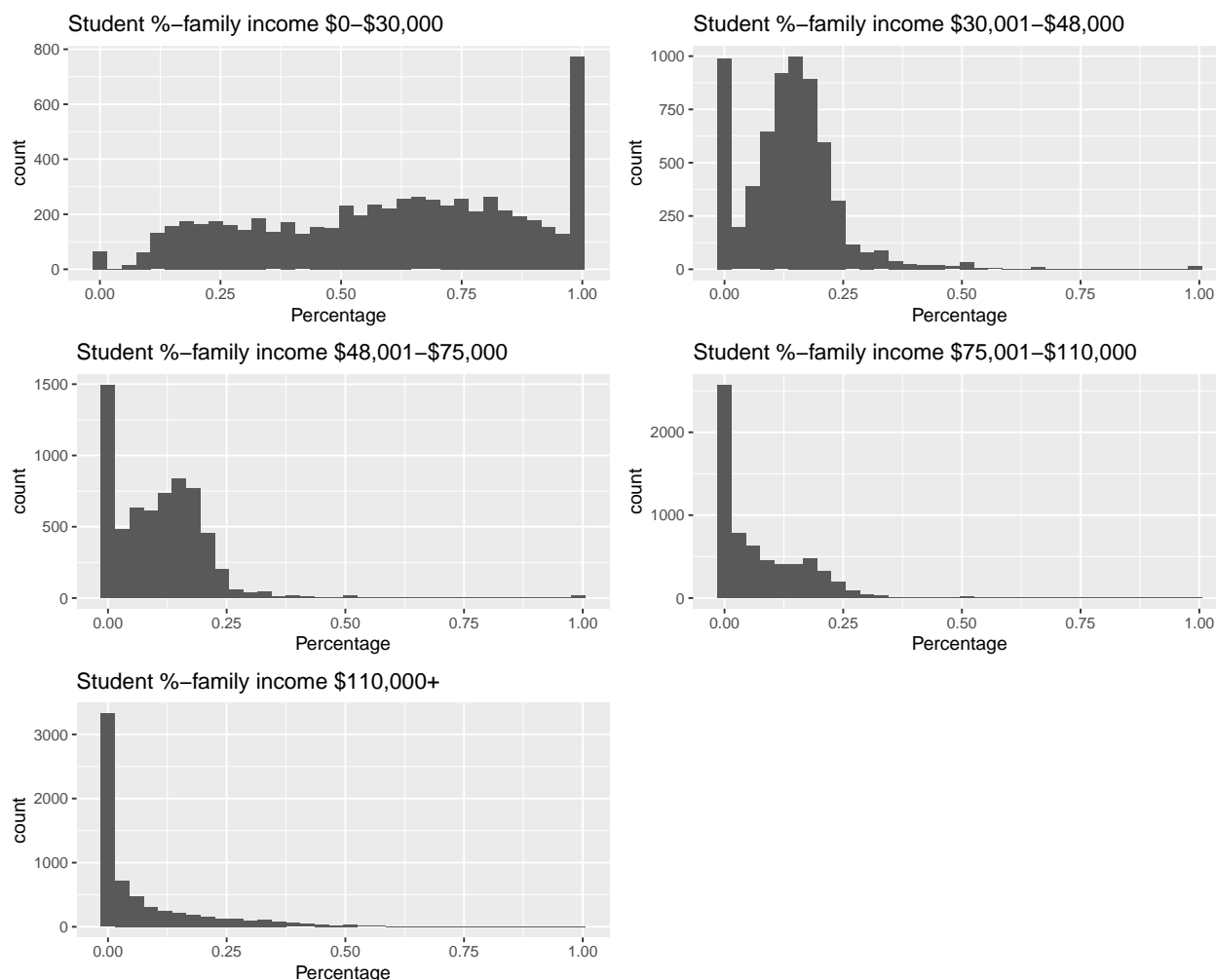
```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
grid.arrange(p1, p2, p3, p4,p5, ncol=2, nrow =3)
```

Student %–family income $0–$30,000

Student %–family income $30,001–$48,000

Student %–family income $48,001–$75,000

Student %–family income $75,001–$110,000

Student %–family income $110,000+

(b)

```
score_1<-score_t %>% rowwise() %>% mutate(percentage_1 =
sum(NUM41_PUB,NUM41_PRIV,na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
pt1<-ggplot(score_1, aes(x=percentage_1,fill=factor(ICLEVEL))) +
  geom_histogram(binwidth=0.03)+xlab("Percentage")+
  ggtitle("Student %-family income $0-$30,000");

score_2<-score_1 %>% rowwise() %>% mutate(percentage_2 =
                                 sum(NUM42_PUB,NUM42_PRIV,na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
pt2<-ggplot(score_2, aes(x=percentage_2,fill=factor(ICLEVEL))) +
  geom_histogram(binwidth=0.03)+xlab("Percentage")+
  ggtitle("Student %-family income $30,001-$48,000");

score_3<-score_2 %>% rowwise() %>% mutate(percentage_3 =
                  sum(NUM43_PUB,NUM43_PRIV,na.rm = TRUE)/sum(NUM41_PUB,
```

```r
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
pt3<-ggplot(score_3, aes(x=percentage_3,fill=factor(ICLEVEL))) +
  geom_histogram(binwidth=0.03)+xlab("Percentage")+
  ggtitle("Student %-family income $48,001-$75,000");

score_4<-score_3 %>% rowwise() %>% mutate(percentage_4 =
                          sum(NUM44_PUB,NUM44_PRIV,na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,NUM41_PRIV,
NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
pt4<-ggplot(score_4, aes(x=percentage_4,fill=factor(ICLEVEL))) +
  geom_histogram(binwidth=0.03)+xlab("Percentage")+
  ggtitle("Student %-family income $75,001-$110,000");

score_5<-score_4 %>% rowwise() %>% mutate(percentage_5 =
                          sum(NUM45_PUB,NUM45_PRIV,na.rm = TRUE)/sum(NUM41_PUB,
NUM42_PUB,NUM43_PUB,NUM44_PUB,NUM45_PUB,
NUM41_PRIV,NUM42_PRIV,NUM43_PRIV,NUM44_PRIV,NUM45_PRIV,na.rm = TRUE))
pt5<-ggplot(score_5, aes(x=percentage_5,fill=factor(ICLEVEL))) +
  geom_histogram(binwidth=0.03)+xlab("Percentage")+
  ggtitle("Student %-family income $110,000+");
grid.arrange(pt1, pt2, pt3, pt4, pt5, ncol=2, nrow =3)
```

Student %–family income $0–$30,000

Student %–family income $30,001–$48,000

Student %–family income $48,001–$75,000

Student %–family income $75,001–$110,000

Student %–family income $110,000+
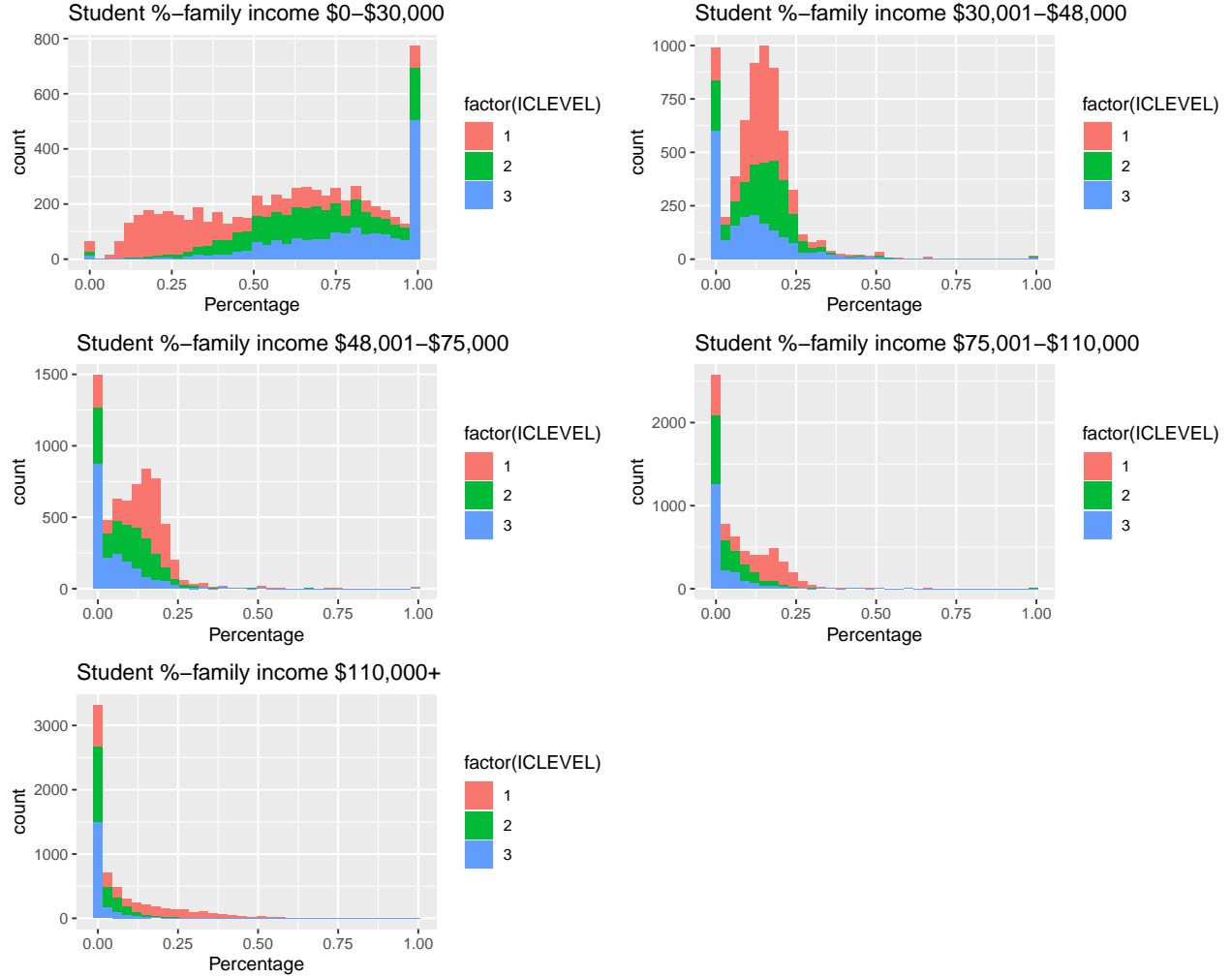
(c)

For the percentage of students with the family income $0-30000, all institutions show relatively even distribution below 100%, while almost 800 of the institutions have about 100% students from families with income $0-30000. Two small peaks can be seen at around 25% and 75%. This means most colleges have more or fewer students from the income level 1. And about 800 colleges have only students from this range of family income. This may be because many or most of the institutions accept federal and state financial aid which provides a green path for students from lower range income families.

For the percentage of students from income level 2 ($30001-48000), the institution counts show approximately normal distribution at percentage range 0-50% with the mean value around 15%. Almost 1000 colleges have no students from the family income level 2 mainly because of the colleges that have only students from income level 1. Similarly, increasing amounts of colleges with no students from income level 3, 4, and 5 can be observed in the figure. And the percentages in the institutions are mainly in the ranges of 0-30%, 0-30%, and 0-25% for income level 3,4, and 5, respectively. As noticed, with the income range increases, the percentage of students from that income range decreases in colleges. For example, most of the colleges had 0% students from families with income $110000+. This is possibly due to the higher need of college education for students from lower income families to prepare for better career. It is also because of the reduced number of students from families with higher incomes.

The figures in 2(b) show not only the distribution of institutions regarding their percentages of students in

different levels of family incomes, but also the divided distributions based on the levels of the institution (level 1: 4-years, level 2: 2-years, level 3: less than 2 years). Overall, there are more level-1 colleges than level-2 colleges, and then level-3 colleges. Different levels of colleges show similar distribution for student percentages from different levels of family income. Higher-level colleges tend to have more students from income level-1 and fewer students from increased family incomes. It may be explained by the education quality and tuition order of the institution: level 1>level 2> level 3. Students from low-income families may be able to afford reduced year of education (level 3: <2 years). And students from high-income families may be better prepared for higher-quality colleges (level 1) due to their better financial condition and education resources.

(d)

```
pt6<-ggplot(data=score_5) +
  geom_point(aes(x=AVGFACSAL, y=percentage_1, group=1,color=factor(ICLEVEL)))+
  geom_smooth(method = "gam",aes(x=AVGFACSAL,y=percentage_1,group=1),
              formula = y ~ s(x, bs = "cs"))+
  ggtitle("Student%($0-$30,000) vs faculty salary")+
  ylab("Student Percentage")+xlab("Faculty Salary")

  pt7<-ggplot(data=score_5) +
    geom_point(aes(x=AVGFACSAL, y=percentage_2, group=1,color=factor(ICLEVEL)))+
    geom_smooth(method = "gam",aes(x=AVGFACSAL,y=percentage_2,group=1),
                formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student%($30,001-$48,000) vs faculty salary")+
    ylab("Student Percentage")+
    xlab("Faculty Salary")

 pt8<- ggplot(data=score_5) +
    geom_point(aes(x=AVGFACSAL, y=percentage_3, group=1,color=factor(ICLEVEL)))+
    geom_smooth(method = "gam",aes(x=AVGFACSAL,y=percentage_3,group=1),
                formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student%($48,001-$75,000) vs faculty salary")+
    ylab("Student Percentage")+
   xlab("Faculty Salary")

 pt9<- ggplot(data=score_5) +
    geom_point(aes(x=AVGFACSAL, y=percentage_4, group=1,color=factor(ICLEVEL)))+
    geom_smooth(method = "gam",aes(x=AVGFACSAL,y=percentage_4,group=1),
                formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student%($75,001-$110,000) vs faculty salary")+
    ylab("Student Percentage")+xlab("Faculty Salary")

pt10<-ggplot(data=score_5) +
    geom_point(aes(x=AVGFACSAL, y=percentage_5, group=1,
                   color=factor(ICLEVEL)))+
    geom_smooth(method = "gam",aes(x=AVGFACSAL,y=percentage_5,group=1),
                formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student%($110,000+) vs faculty salary")+
    ylab("Student Percentage")+
  xlab("Faculty Salary")

    grid.arrange(pt6, pt7, pt8, pt9, pt10, ncol=2, nrow =3)
```

Student%($0–$30,000) vs faculty salary

Student%($30,001–$48,000) vs faculty salary

Student%($48,001–$75,000) vs faculty salary

Student%($75,001–$110,000) vs faculty salary

Student%($110,000+) vs faculty salary

(e)

The figure in 2(d) shows that the institutions with increasing faculty salary have a decreasing percentage of students from income level-1. Since faculty salary could be positively related to the college quality, therefore higher tuition. This leads to the lower percentage of students from low-income family. For the percentage of students from income level 2, no obvious trend is seen except a slight increase with faculty salary higher than $15000. A more significant increasing trend is observed for the student percentage at income level 3, 4, and 5. This could be due to the improved financial condition and education background of these students. Which provide better conditions for pursuing higher-quality institutions (higher faculty salary). However, a slight decrease can be observed at faculty salary higher than $15000 for the percentage of students from income level 4 and 5. It may be related with the special financial funds proved by high-quality colleges and their selection of various types of students to meet certain ratios, which lower the percentage of students from high-income families. In addition, all five distributions are sub grouped by the college levels (level 1: 4-years, level 2: 2-years, level 3: less than 2 years). Most plotted institutions are level 1 or level 2. Salary of faculty in level-2 colleges is in a lower range (about $0-10000) than it in level-1 colleges ($0-20000). And compared to level-1 schools, level-2 schools have overall higher percentages of students from lower income families and lower percentages of students from higher income families. This is due to shorter programs in level-2 schools (2 years) and cheaper tuition leading to more students with lower family income. In contrast, students from high income family can normally afford 4-year education in level-1 schools and mostly pursue a better education in level-1 schools.

Q4

```
earning<-read_csv("MERGED2007_08_PP.csv",na=c(" ","NA","NULL","PrivacySuppressed"),
                  show_col_types = FALSE)
library(dplyr)
earning_t<-tibble(earning)
sfg<-ggplot(earning_t,aes(x=FIRST_GEN, y=MN_EARN_WNE_P10)) +
     geom_point()+geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))+
     ggtitle("Student mean earning vs Share of first generation students")+
     ylab("Mean Earnings")+xlab("Share of first generation students");sfg
```

Student mean earning vs Share of first generation students
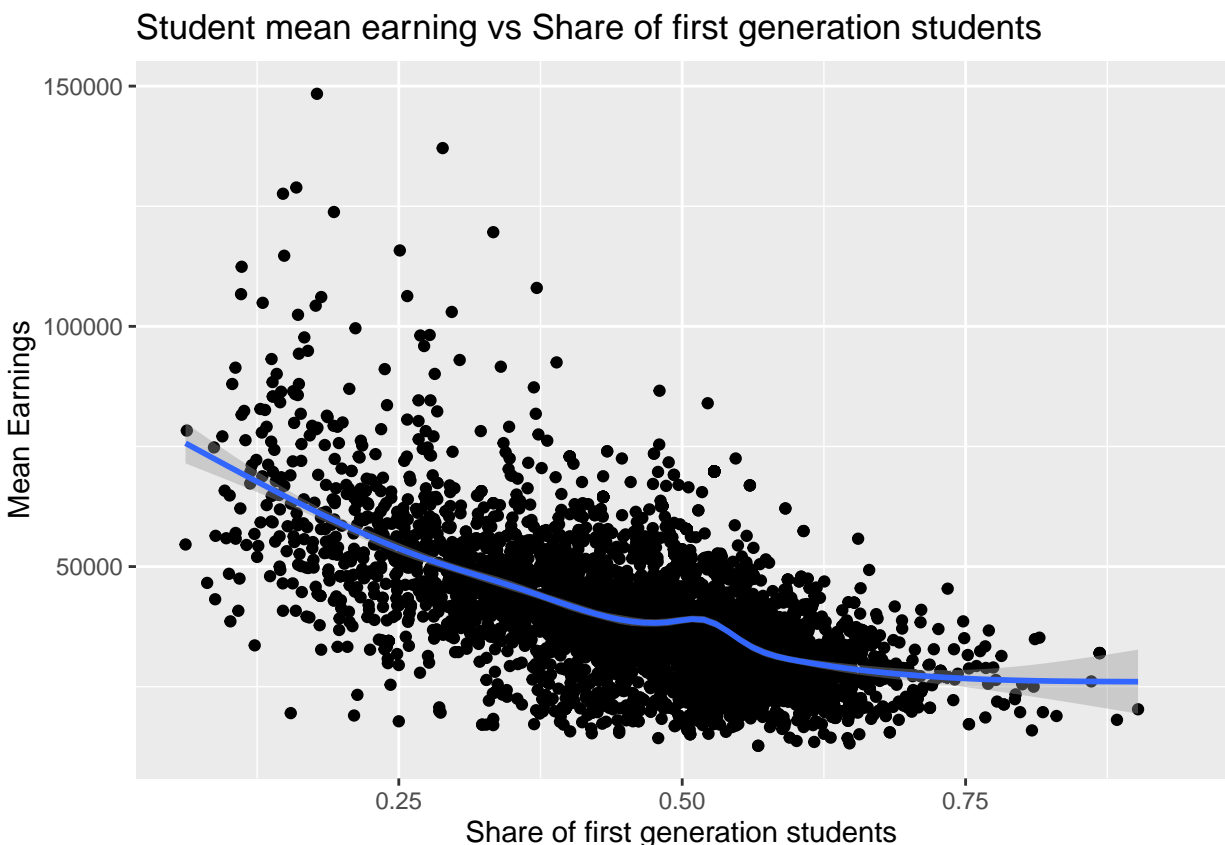
The plot shows the relationship between student mean earnings and the share of first-generation students. As observed, student mean earnings in the institutions decrease when the share of first-generation students in the institutions increases. This suggests the high importance of family education background. When an institution has a lower percentage of first-generation students and a higher percentage of non-first-generation students, it tends to show an increasing mean student earning. This suggests students with educated families may have a better career with higher earnings after graduation. This could be attributed to important support, additional education resource, and guidance these students received from their families. In addition, the student mean earnings show much lower variation at higher share of first-generation students than at lower share. It indicates institutions with high share of non-first-generation students may not necessarily have high mean student earnings, however, institutions with high share of first-generation students are highly likely to have low mean student earnings. It reflects the importance of family education background in this case.

Q3 Report

1. Introduction

Different information about academic institutions in 2007-2008 has been collected in the College Scorecard data. Some of such data includes variables of (1) mean earnings of students working and not enrolled 10 years after entry, (2) institution's admission rate, (3) average SAT equivalent score of students admitted, (4) region of the country, (5) average faculty salary per month, (6) mean earnings of female students working and not enrolled 10 years after entry, and (7) mean earnings of male students working and not enrolled 10 years after entry [1]. Valuable relationship information can be extracted by plotting and analyzing the data set. Herein, the relationship between the mean earning of students and different variables ((2)-(5)) has been analyzed. In addition, possible relationship between mean earning of female and male students ((6) vs (7)) has been discussed.

2. Analysis methods

To plotting and editing the report, statistical software including R studio and R markdown has been adopted. Packages including "tidyverse" and "dplyr" have been applied.

3. Results and discussion

3.1. Student earnings vs admission rate

To explore the relationship between student earnings and admission rate, the data points of the mean student earnings are plotted with institutions' admission rate in Fig. 1. A smooth line is added. As seen in the figure, mean student earnings show a decreasing non-linear trend with an increasing admission rate from 0 to 1. This suggests that students admitted through higher admission standard (low admission rate) may have higher required abilities, which contribute to their increasing earnings per month. The figure also shows fewer datapoints at higher admission rate than at lower admission rate, which is due to the low amount of student in institutions with high standards. Majority institutions have the admission rate in the range of approximately 0.3-1. It is also observed the mean earnings at admission rate of 0.50-0.75 are relatively stable with no obvious increase or increase, which indicates different institutions with admission rate 0.5 to 0.75 may not significantly affect their students' salary in this case.
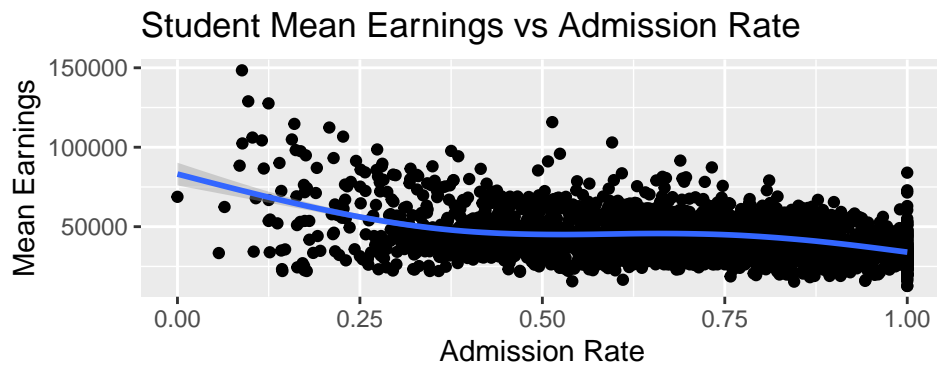

Fig. 1. The relationship between mean student earnings (10 years after entry) and admission rate of the institutions.

3.2. Student earnings vs in-state tuition

Fig. 2 shows the plot of mean student earnings with the in-state tuition in the institutions. The added smooth line exhibits an overall increasing trend with higher in-state tuition. This suggests institutions with high in-state tuition may provide higher-quality education, better resources and faculties, better social network, etc. These lead to better development of students and future jobs with higher salary after their graduation. It can also be observed that tuition of most collected institutions is cheaper than about $30000, only a small number of institutions request tuition than 30000.
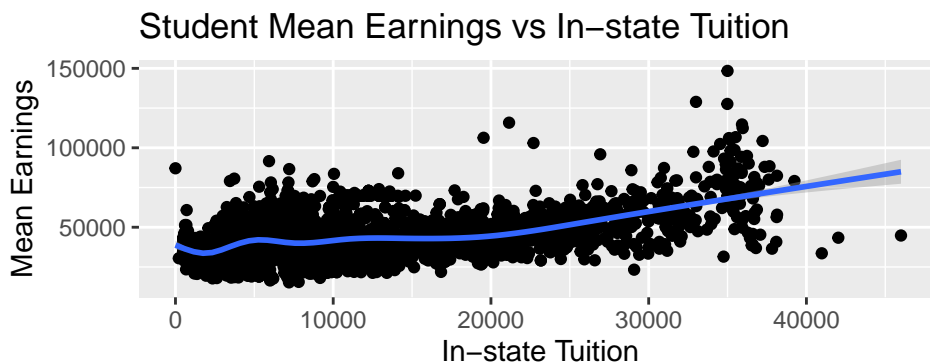

Fig. 2. The relationship between mean student earnings (10 years after entry) and in-state tuition of the institutions.

3.3. Student earnings vs SAT scores

Fig. 3 presents the relationship between mean student earnings and the average SAT scores of students from the institutions. As observed, when SAT scores increase, the mean student earnings also increase in a non-linear manner. Since a higher average SAT score of a college indicates its higher admission standard and higher capabilities and better background of admitted students. This results in better ability of graduated students and their increased salary on average. It is noticed that the average SAT scores in majority colleges are around 900-1100. With further increase of average SAT score in colleges, students' earnings seem to show an exponential increase.
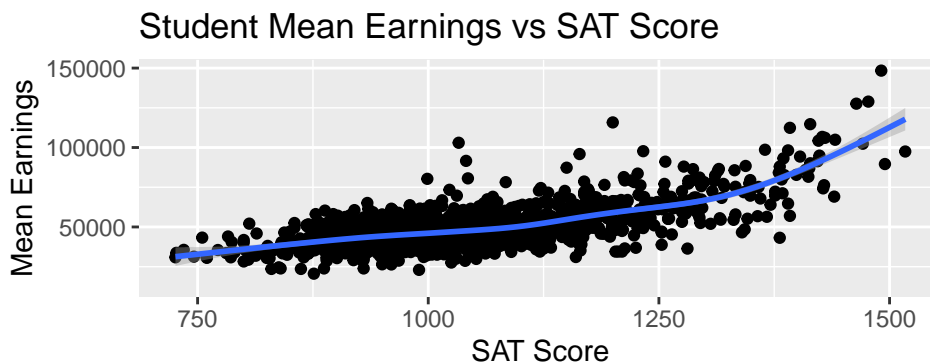

Fig. 3. The relationship between mean student earnings (10 years after entry) and SAT scores of the institutions.

3.4. Student earnings vs region

Fig. 4 represents the boxplot of the mean student earnings for the institutions in different regions. 10 levels of regions (in order: service schools, New England, Mid East, Great Lacks, Plains, Southeast, Southwest, Rocky

Mountains, Far West, outlying areas) are included. As seen, there are data outliers with higher earnings for regions 1 to 9. The average student earnings from colleges in regions 1-8 show no significant difference. Although the student earnings in service schools (#0) shows a much higher average than those in region 1-8, a very limited number of data points were collected. And the mean earnings in region 9 are mostly lower than student earnings from other regions. This may be because of the highly limited education quality and isolated condition of the colleges in outlying areas, those lead to limited skills/abilities and earnings of the students.
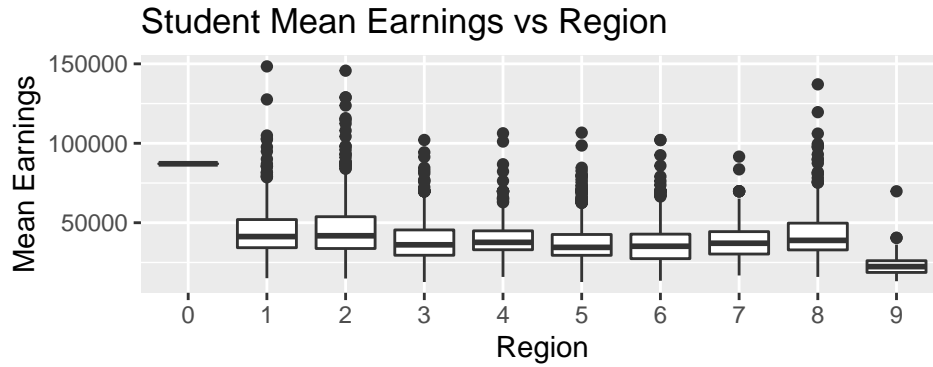

Fig. 4. The relationship between mean student earnings (10 years after entry) and country region of the institutions.

3.5. Student earnings vs faculty salary

Fig. 5 shows a main increasing trend of the main student earnings with higher faculty salary of the institutions. At the faculty salary lower than ~100, only few colleges can be found, the prediction through the added smooth line may not be accurate. Overall, higher faculty salary may represent a higher level of the institution, better ability of the faculty, and better quality of the education. Therefore, the related institutions can better prepare their students for future careers and higher earnings.
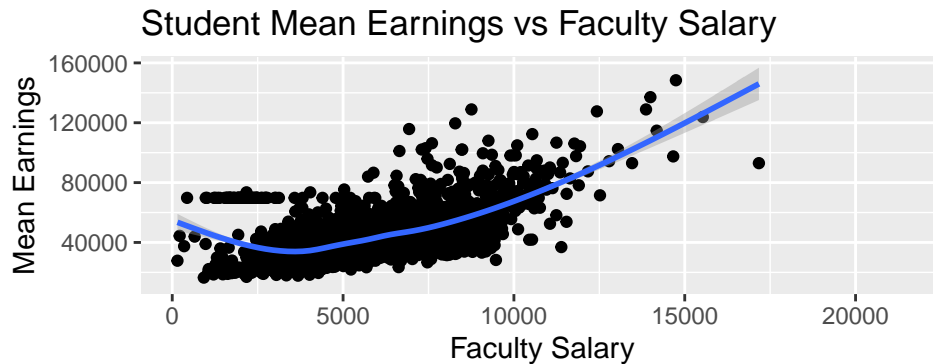

Fig. 5. The relationship between mean student earnings (10 years after entry) and faculty salary of the institutions.

3.6. Earnings - males and female students

Fig. 6 shows average male and female student earnings follow an approximately positive linear relationship. Average male earnings increase when average female earnings increase. However, female students seem to have slightly lower earnings than male students in average according to the slope of the smooth line (<1). Most of average female earnings are in the range of \$1500-6000, while most of the average male earnings are in the range of \$2000-8000.
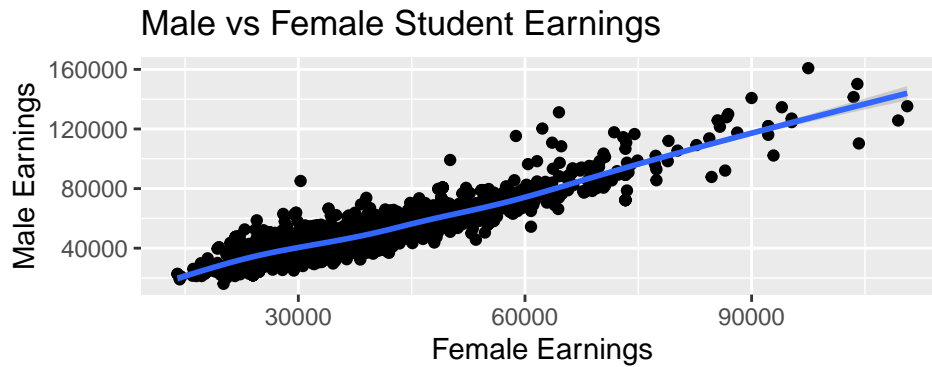
Fig. 6. The relationship between mean male student earnings and female student earnings (10 years after entry) of the institutions.

4. Conclusion

In summary, the average earning of student not enrolled 10 years after entry in U.S. institutions is highly related with different factors of their institutions in 2007-2008. As analyzed, these factors include institution's admission rate, in-state tuition, average SAT score, region of the country, and average faculty salary. Students graduated from institutions with higher admission rate, in-state tuition, SAT score, and average faculty salary tend to have increased earnings. While students from colleges in outlying areas show lower average earnings mostly. Earnings of female and male students show positive linear relationship, while female students have slightly lower earnings than male students in average.

Appendix: R code for Q3

```
library(tidyverse)
options(warn = - 1)

earning<-read_csv("MERGED2007_08_PP.csv",na=c(" ","NA","NULL","PrivacySuppressed"),
                show_col_types = FALSE)
library(dplyr)
earning_t<-tibble(earning)
adm<-ggplot(earning_t,aes(x=ADM_RATE, y=MN_EARN_WNE_P10)) +
    geom_point()+geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student Mean Earnings vs Admission Rate")+
    ylab("Mean Earnings")+xlab("Admission Rate"); adm

tui<-ggplot(earning_t,aes(x=TUITIONFEE_IN, y=MN_EARN_WNE_P10)) +
    geom_point()+geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student Mean Earnings vs In-state Tuition")+
    ylab("Mean Earnings")+xlab("In-state Tuition"); tui

sat<-ggplot(earning_t,aes(x=SAT_AVG, y=MN_EARN_WNE_P10)) +
    geom_point()+geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student Mean Earnings vs SAT Score")+
    ylab("Mean Earnings")+xlab("SAT Score"); sat

rig<-ggplot(earning_t,aes(x=factor(REGION), y=MN_EARN_WNE_P10)) +
    geom_boxplot()+
    ggtitle("Student Mean Earnings vs Region")+
    ylab("Mean Earnings")+xlab("Region");rig
```

```r
fsa<-ggplot(earning_t,aes(x=AVGFACSAL, y=MN_EARN_WNE_P10)) +
    geom_point()+geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))+
    ggtitle("Student Mean Earnings vs Faculty Salary")+
    ylab("Mean Earnings")+xlab("Faculty Salary");fsa

gender<-ggplot(earning_t,aes(x=MN_EARN_WNE_MALE0_P10, y=MN_EARN_WNE_MALE1_P10)) +
    geom_point()+geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))+
    ggtitle("Male vs Female Student Earnings")+
    ylab("Male Earnings")+xlab("Female Earnings");gender
```