

Name: Chandana Burramukku

ISL Assignment 2

1.a

```

R packages available: Chandana_Burramukku_ISL_Assignment_2 College
Source on Save Go to R function Addins
1 w1.a ISLR 2.4 Applied Problem 8
2 library(ISLR)
3 setwd("C:/R")
4 reading the file from the directory
5 college <- read.csv("College.csv")
6
7 R.122 - C/R
8 11296 8124 600 850 83 93 10.3 33 12580 91
9 5360 600 1800 76 78 12.6 11 9084 72
10 11296 12580 91 12.6 11 9084 72
11 12850 5400 400 800 78 89 12.2 30 8954 73
12 3664 650 900 63 111 13 7614 49
13 8840 2950 750 1290 74 82 11.1 31 6668 84
14 9680 4800 1200 2480 78 85 11.2 16 7550 52
15 3616 355 715 87 95 11.1 26 12957 69
16 16304 2750 660 1800 57 62 19.6 16 3752 46
17 9550 3850 1000 210 64 14.1 18 3722 58
18 21700 4100 600 500 35 59 10.1 33 16364 55
19 13800 5510 630 850 87 87 17.5 20 10941 82
20 3850 3500 2300 890 16.3 17 1111 63
21 8740 3363 550 1700 62 68 11.6 29 7718 48
22 540 * 3580 500 1400 61 80 8.8 32 8324 56
[ reached 'max' in getoptions('max.print') -- omitted 725 rows ]
> college<-read.csv('College.csv')
> View(college)
> view.college
> view(college)
> View(College)
> college <- read.csv("College.csv")
>

```

b.

```

File Edit Code View Plots Session Build Debug Profile Tools Help
Go to Rfunction Addins
R packages available: Chandana_Burramukku_ISL_Assignment_2 college College
Private Apps Accept Enroll Top10perc Top25perc F.Undergrad P.Undergrad Outstate Room.Board Be
Abilene Christian University Yes 1660 1232 721 23 52 2885 557 7440 3300
Adolphus University Yes 2194 1804 512 16 29 2663 1227 12280 6450
Adrian College Yes 1428 1097 536 22 50 1036 99 11230 2750
Agnes Scott College Yes 417 349 137 60 89 510 63 12960 5450
Alaska Pacific University Yes 119 95 45 16 44 249 889 7560 4250
Albertson College Yes 937 479 158 38 62 676 41 13030 3339
Albertus Magnus College Yes 333 340 108 17 45 416 230 13290 6720
Alderson College Yes 1899 1720 489 37 68 1594 32 13668 4624
Albright College Yes 1039 839 227 30 63 973 306 15595 4400
Alderson Broaddus College Yes 382 498 172 21 44 799 78 10468 3380
Alma College Yes 1773 1420 437 37 79 1835 111 14548 5640
Allegheny College Yes 2652 1900 454 44 77 1707 44 17000 4440
Albion Coll. of St. Francis de Sales Yes 1178 780 290 58 64 1130 658 9880 4783
Alma College Yes 1267 1000 385 44 73 1306 28 12572 4552
Alymer College Yes 494 313 157 23 46 1317 1239 8352 3640
American International College Yes 1420 1093 220 9 22 1018 287 8700 4780
Amherst College Yes 4302 992 416 63 96 1993 5 19760 3300
Amherst College Yes 4302 992 416 63 96 1993 5 19760 3300
Showing 1 to 18 of 7777 entries, 18 total columns
Console Terminal Jobs
R 4.1.2 - C/R
> library(ISLR)
> head(college)
Private Apps Accept Enroll Top10perc Top25perc F.Undergrad P.Undergrad Outstate Room.Board Be
Abilene Christian University Yes 1660 1232 721 23 52 2885 557 7440 3300
Adolphus University Yes 2186 1924 512 16 29 2683 1227 12280 6450
Adrian College Yes 1428 1097 536 22 50 1036 99 11230 2750
Agnes Scott College Yes 417 349 137 60 89 510 63 12960 5450
Alaska Pacific University Yes 119 95 45 16 44 249 889 7560 4250
Albertson College Yes 587 479 158 38 62 678 41 13500 3335
Books.Fees.Rate Total.Fees.Terminal.S.F.Rate perc.Alumni.Expend.Grad.Rate
Abilene Christian University 450 2200 70 78 18.1 12 7041 60
Adolphus University 400 1600 100 100 10.0 12 7041 60
Adrian College 400 1165 53 66 12.9 30 8735 54
Agnes Scott College 400 1000 100 97 7.7 30 8735 54
Alaska Pacific University 800 1500 76 72 11.0 12 10922 15
Albertson College 600 675 67 73 9.4 11 9727 55
-> read.csv("College.csv")
-> view.college
-> view.college <- college[, 1]
-> view.college
-> view.college <- college[, 1]
-> colnames(college) <- c("College", "1")
-> view(College)

```

1.c.i

RStudio Environment View

```

#1.a ISLR 2.4 Applied Problem 8
2 setting up the directory
3 setwd("C:/R")
4 #reading the File from the directory
5 college <- read.csv("college.csv")
6 head(college)
7
8 colnames (college) <- college[, 1]
9 view (college)
10 college[1:10, -1]
11 view (college)
12 #I.c.i Use the summary() function to produce a numerical summary
13 #recall that you can use the data set
14 summary (college)
15

```

RStudio Files View

```

19:1 (Top Level) z R Script
Console Terminal Jobs
# R4.12 - C:\R\z
> #not the variables in the data set
> sum(is.na(college))
Private          Apps      Accept    Enroll Top10perc Top25perc F.Undergrad
Length: 1800   Min. :  0.0   Min. : 96.0   Min. : 1250   Min. :  8.00   Min. : 24.0
Class: character 1st Qu.: 776   1st Qu.: 604   1st Qu.: 242   1st Qu.: 15.00   1st Qu.: 41.0   1st Qu.: 992
Mode: character Median : 1558  Median : 1110  Median : 434  Median : 23.00  Median : 54.0   Median : 1707
          Mean   : 1555.0  Mean   : 1109.0  Mean   : 434.0  Mean   : 23.00  Mean   : 54.0   Mean   : 1707
          3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.: 35.00   3rd Qu.: 69.0   3rd Qu.: 4005
          Max.  : 48094   Max.  : 26330  Max.  :16392  Max.  :196.00  Max.  :1000.00  Max.  :11643
P.Undergrad <- outstate ~ room.Board ~ Books ~ Persons ~ PHP ~ Terminal
Min. : 1.0   Min. : 1240   Min. : 11780  Min. : 1.250   Min. :  8.00   Min. : 24.0
1st Qu.: 11.50  1st Qu.:113.00  1st Qu.: 6751   1st Qu.: 12.000   1st Qu.: 65.00   1st Qu.: 10.0
Median : 353.0  Median : 9990  Median :14200   Median :12000   Median : 75.00   Median : 82.0
          Mean   : 355.0  Mean   :10441   Mean   : 9398   Mean   :1341   Mean   : 75.66   Mean   : 79.7
          3rd Qu.: 116.00  3rd Qu.:113.00  3rd Qu.: 6751   3rd Qu.: 15.00   3rd Qu.: 69.00   3rd Qu.: 4005
          Max.  :399.80   Max.  :16392   Max.  :196.00  Max.  :1000.00  Max.  :103.00   Max.  :1100.00
>

```

1.c.ii

RStudio Environment View

```

#1.a ISLR 2.4 Applied Problem 8
2 setting up the directory
3 setwd("C:/R")
4 #reading the File from the directory
5 college <- read.csv("college.csv")
6 head(college)
7
8 colnames (college) <- college[, 1]
9 view (college)
10 college[1:10, -1]
11 view (college)
12 #I.c.i Use the summary() function to produce a numerical summary
13 #recall that you can use the data set
14 summary (college)
15 #I.c.ii Use the pairs() function to produce a scatterplot matrix of the first ten columns or variables of the data.
16 #recall that you can use the first ten columns of a matrix A using A[,1:10].
17 #pairs (college[, 1:10])
18 pairs (college[, 1:10])
19

```

RStudio Files View

```

19:1 (Top Level) z R Script
Console Terminal Jobs
# R4.12 - C:\R\z
> #not the variables in the data set
> sum(is.na(college))
Private          Apps      Accept    Enroll Top10perc Top25perc F.Undergrad
Length: 1800   Min. :  0.0   Min. : 96.0   Min. : 1250   Min. :  8.00   Min. : 24.0
Class: character 1st Qu.: 776   1st Qu.: 604   1st Qu.: 242   1st Qu.: 15.00   1st Qu.: 41.0   1st Qu.: 992
Mode: character Median : 1558  Median : 1110  Median : 434  Median : 23.00  Median : 54.0   Median : 1707
          Mean   : 1555.0  Mean   : 1109.0  Mean   : 434.0  Mean   : 23.00  Mean   : 54.0   Mean   : 1707
          3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.: 35.00   3rd Qu.: 69.0   3rd Qu.: 4005
          Max.  : 48094   Max.  : 26330  Max.  :16392  Max.  :196.00  Max.  :1000.00  Max.  :11643
P.Undergrad <- outstate ~ room.Board ~ Books ~ Persons ~ PHP ~ Terminal
Min. : 1.0   Min. : 1240   Min. : 11780  Min. : 1.250   Min. :  8.00   Min. : 24.0
1st Qu.: 11.50  1st Qu.:113.00  1st Qu.: 6751   1st Qu.: 12.000   1st Qu.: 65.00   1st Qu.: 10.0
Median : 353.0  Median : 9990  Median :14200   Median :12000   Median : 75.00   Median : 82.0
          Mean   : 355.0  Mean   :10441   Mean   : 9398   Mean   :1341   Mean   : 75.66   Mean   : 79.7
          3rd Qu.: 116.00  3rd Qu.:113.00  3rd Qu.: 6751   3rd Qu.: 15.00   3rd Qu.: 69.00   3rd Qu.: 4005
          Max.  :399.80   Max.  :16392   Max.  :196.00  Max.  :1000.00  Max.  :103.00   Max.  :1100.00
>

```

RStudio Plots View

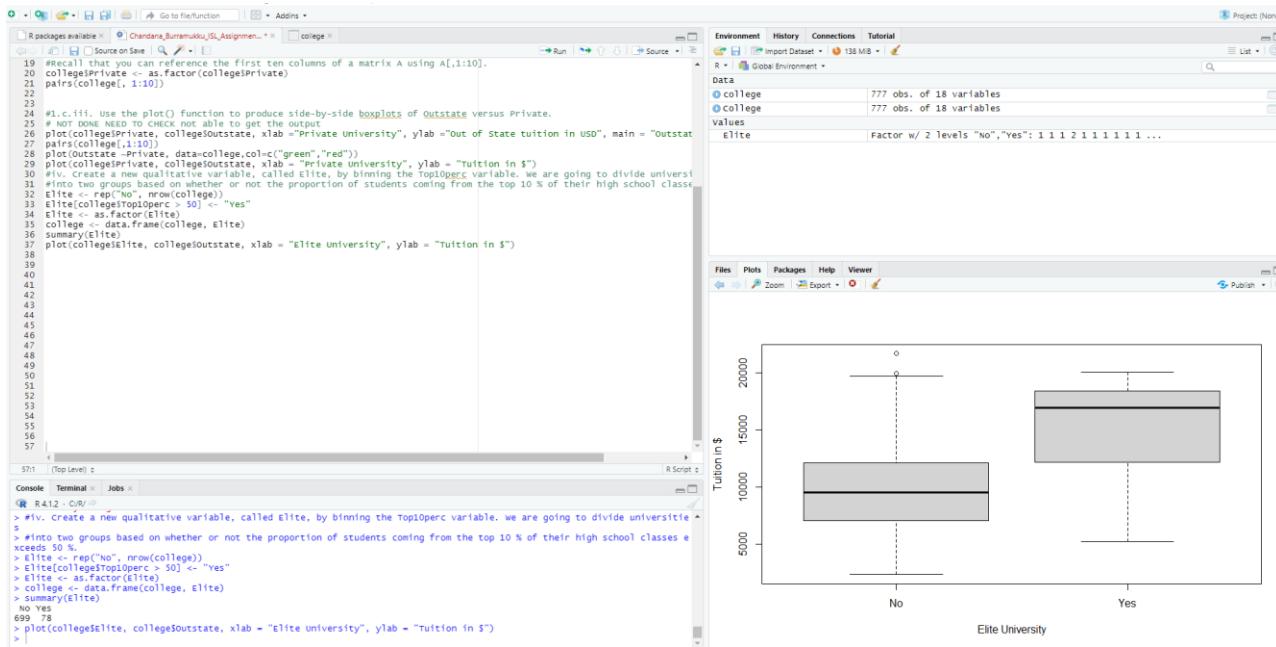
1.c.iii

```

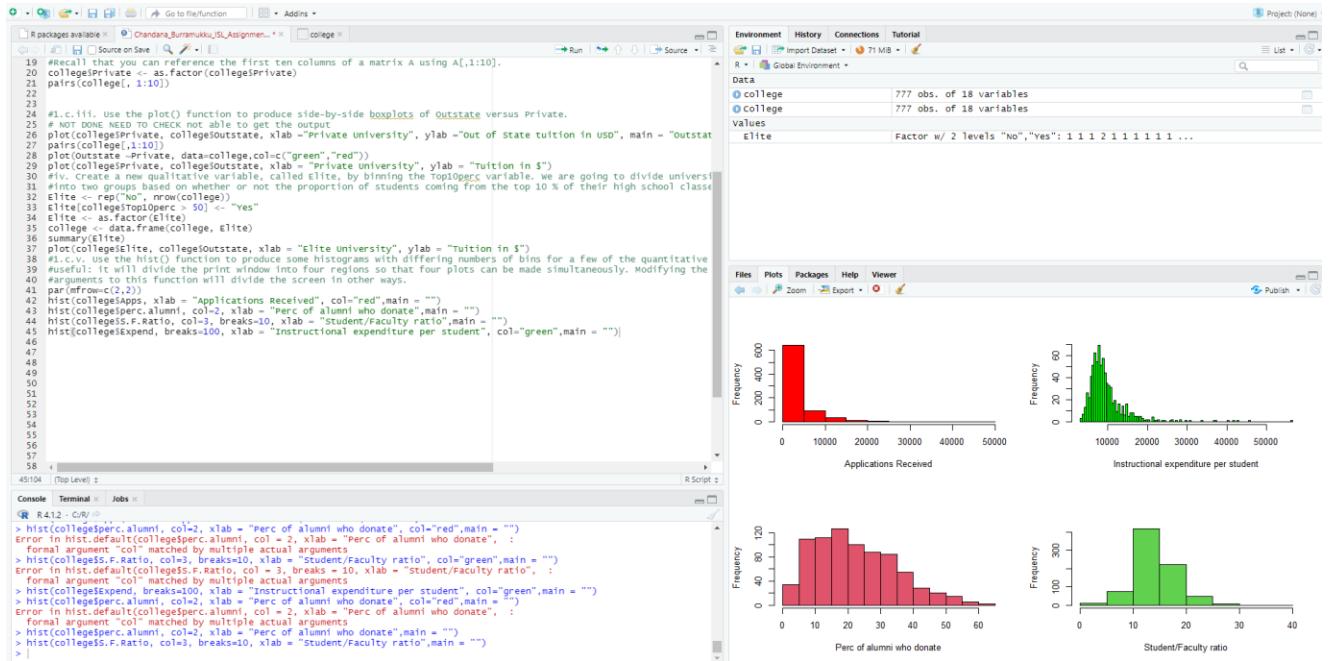
# package available - C:\Users\Arunmaku\IIS\Assignment -> Boston > college
1 view (college)
2 # view (college) function to produce a numerical summary
3 # of the variables in the data set
4
5 #<--(1:10) tell the pairs() function to produce a scatterplot matrix of the first ten columns or variables of the data.
6 #<--(1:10) tell the pairs() function to produce a scatterplot matrix of the first ten columns of a matrix A using A[,1:10].
7
8 collegePrivate <- as.factor(college$Private)
9 head(college)
10
11
12 #<--(1:10) use the plot() function to produce side-by-side boxplots of outstate versus private.
13 plot(college$private, college$outstate, xlab = "Private", ylab = "Out-of-state tuition in dollars")
14
15
16 #iv. Create a new qualitative variable, called elite, by binning the top10perc variable. we are going to divide universities
17 #into two groups based on whether or not the proportion of students coming from the top 10 % of their high school classes exceeds
18 #30%.
19 elite <- rep("No", nrow(college))
20 elite[college$top10perc >= 30] <- "Yes"
21 elite <- as.factor(elite)
22 college <- data.frame(college, elite)
23 summary(elite)
24 plot(elite, college$outstate, xlab = "Elite university", ylab = "Tuition in $")
25
26 #<--(1:10) use the hist() function to produce some histograms with different numbers of bins for a few of the quantitative variable
27 #in the data set. Note that the first histogram has 10 bins, and that your plots can be modified simultaneously. Modifying the
28 #arguments in this function will divide the screen in other ways.
29 par(mfrow=c(2,2))
30 hist(college$tuexpnd, breaks=10, xlab = "Instructional expenditure per student", col="green", main = "")
31 hist(college$alumni, col=2, xlab = "perc of alumni who donate", main = "")
32 hist(college$tuexpnd, breaks=50, xlab = "Instructional expenditure per student", col="red", main = "")
33 hist(college$tuexpnd, breaks=100, xlab = "Instructional expenditure per student", col="green", main = "")
34
35 summary(college$tuexpnd)
36
37 #<--(1:10) use the pairs() function to produce side-by-side boxplots of outstate versus private.
38 plot(college$private, college$outstate, xlab = "Private", ylab = "Out-of-state tuition in dollars")
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

```

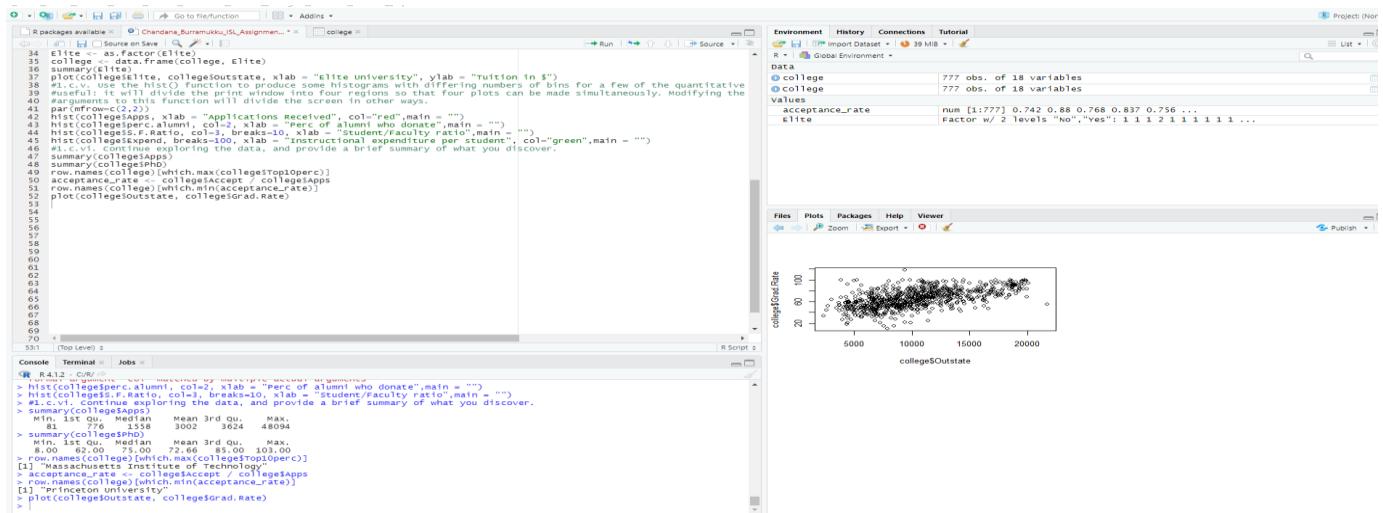
1.c.iv



1.c.v



1.c.vi



2.a

The screenshot shows the RStudio interface with the following details:

- Code Editor:** The left pane displays R code for generating a scatter plot. The code includes imports for ggplot2, dplyr, and magrittr, and uses the mtcars dataset to create a plot of college spending versus tuition.
- Environment View:** The top right pane shows the environment with variables like `college`, `collegeOutstate`, `collegeSpendRate`, `elite`, `mtcars`, and `summary`.
- Plots View:** The bottom right pane contains a scatter plot titled "collegeSpendRate" with "collegeOutstate" on the x-axis and "collegeSpendRate" on the y-axis. The plot shows a strong positive correlation between the two variables.

2.b

The screenshot shows the RStudio interface with two main panes. The left pane is the 'Script' editor containing R code for data analysis, and the right pane is a 'Plots' window showing a scatter plot.

Script Editor Content:

```
# Load libraries
library(tidyverse)
library(ggplot2)
library(dplyr)

# Load data
college_gpa = read_csv("College GPA.csv")
college_gdpstate = read_csv("College GDP State.csv")

# Merge datasets
college = dplyr::left_join(college_gpa, college_gdpstate, by = "name")

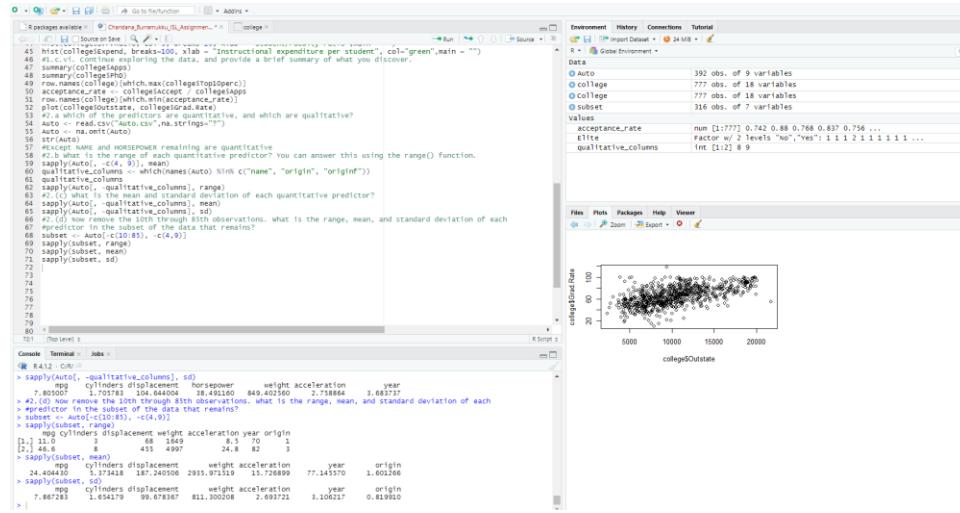
# Plotting
ggplot(college, aes(x = collegeGDPstate, y = collegeGPA)) +
  geom_point() +
  ggtitle("Elite university vs tuition in $")
```

Plots Window Content:

A scatter plot titled "Elite university vs tuition in \$" showing a strong positive correlation between collegeGDPstate (X-axis) and collegeGPA (Y-axis). The X-axis ranges from approximately 5000 to 20000, and the Y-axis ranges from 20 to 80.

2.c

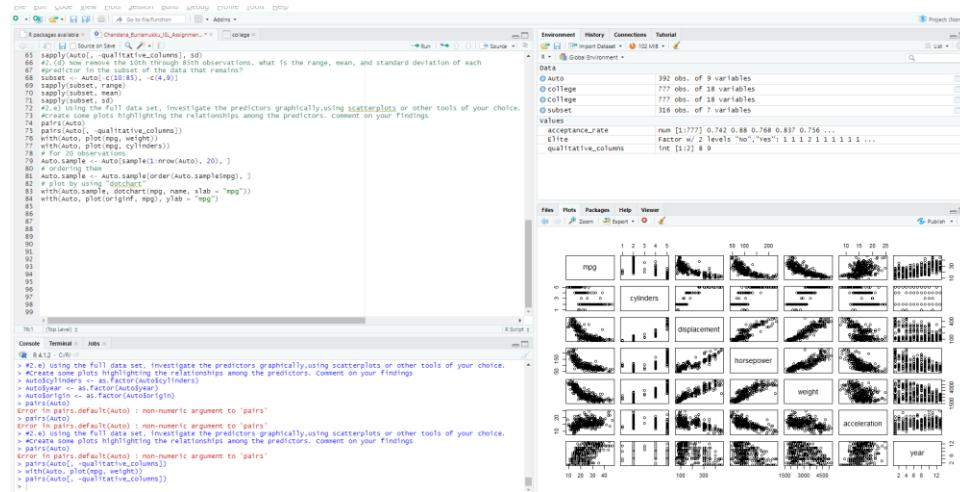
2.d



The screenshot shows the RStudio interface with the following details:

- Code Editor:** Displays R code for analyzing the college dataset. The code includes:
 - Exploratory data analysis (e.g., histograms, summary statistics).
 - Filtering observations (e.g., `college\$outstate == "NC"`).
 - Calculating descriptive statistics (e.g., range, mean, standard deviation) for quantitative variables.
 - Subsetting the data to remove specific observations (e.g., `subset ~ -c(10:85)`).
 - Plotting `college\$displacement` vs `college\$outstate`.
- Environment View:** Shows the following objects:
 - Auto: 392 obs. of 9 variables
 - college: 777 obs. of 18 variables
 - College: 777 obs. of 18 variables
 - subset: 316 obs. of 7 variables
 - values: num [1:777] 0.742 0.88 0.768 0.837 0.756 ...
 - acceptance_rate: Factor w/ 2 levels: "no","yes": 1 1 1 2 1 1 1 1 1 1 ...
 - qualitative_columns: int [1:2] 8 9
- Plots:** A scatter plot showing the relationship between `college\$displacement` (X-axis, 5000 to 20000) and `college\$outstate` (Y-axis, 20 to 100).

2.e



The screenshot shows the RStudio interface with the following details:

- Code Editor:** Displays R code for analyzing the auto dataset. The code includes:
 - Exploratory data analysis (e.g., histograms, summary statistics).
 - Filtering observations (e.g., `subset ~ -c(10:85)`).
 - Calculating descriptive statistics (e.g., range, mean, standard deviation) for quantitative variables.
 - Subsetting the data to remove specific observations (e.g., `subset ~ -c(10:85)`).
 - Plotting various pairs of variables (e.g., mpg vs cylinders, displacement, horsepower, weight, acceleration, year, origin).
- Environment View:** Shows the following objects:
 - Auto: 392 obs. of 9 variables
 - college: 777 obs. of 18 variables
 - College: 777 obs. of 18 variables
 - subset: 316 obs. of 7 variables
 - values: num [1:777] 0.742 0.88 0.768 0.837 0.756 ...
 - acceptance_rate: Factor w/ 2 levels: "no","yes": 1 1 1 2 1 1 1 1 1 1 ...
 - qualitative_columns: int [1:2] 8 9
- Plots:** A grid of 12x12 scatter plots showing pairwise relationships between all variables in the auto dataset.

File Edit Code View Plots Session Build Debug Profile Tools Help

R packages available: [] Chandra_Arunmukul_Asgnmen... | conda |

```

65 #apply(Auto, -qualitative_columns, sd)
66 #2.(d) now remove the 10th through 80th observations. what is the range, mean, and standard deviation of each
67 #subset <- Auto[-c(10:80), -c(4,9)]
68 #subset <- Auto[-c(10:80), -c(4,9)]
69 #subset <- Auto[-c(10:80), -c(4,9)]
70 #apply(subset, sd)
71 #apply(subset, sd)
72 #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
73 #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
74 #pairs(Auto[, qualitative_columns])
75 #pairs(Auto[, qualitative_columns])
76 #with(Auto, plot(mpg, weight))
77 #with(Auto, plot(mpg, cylinders))
78 #Auto.sample <- Auto[sample(1:nrow(Auto), 20), ]
79 #ordering them
80 #plot by using "dotchart"
81 #with(Auto.sample, dotchart(mpg, name, xlab = "mpg"))
82 #with(Auto.sample, dotchart(cylinders, mpg, ylab = "mpg"))
83 #with(Auto, plot(cylinders, mpg, ylab = "mpg"))
84 #with(Auto, plot(cylinders, weight))
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

771 [Top Level] R Script

Console Terminal Jobs

R 4.1.2 - CRAN

> #create some plots highlighting the relationships among the predictors. Comment on your findings
> Auto\$cylinders <- as.factor(Auto\$cylinders)
> Auto\$year <- as.factor(Auto\$year)
> Auto\$origin <- as.factor(Auto\$origin)
> pairs(Auto)
Error in pairs.default(Auto) : non-numeric argument to 'pairs'
> pairs(Auto[, default(Auto) : non-numeric argument to 'pairs'
> #2.(d) using the full data set, investigate the predictors graphically,using scatterplots or other tools of your choice.
> #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
> pairs(Auto[, qualitative_columns])
Error in pairs.default(Auto) : non-numeric argument to 'pairs'
> pairs(Auto[, qualitative_columns])
> with(Auto, plot(mpg, weight))
> with(Auto, plot(mpg, cylinders))
> Auto.sample <- Auto[sample(1:nrow(Auto), 20),]
> Auto.sample <- Auto.sample[order(Auto.sample\$mpg),]
> with(Auto.sample, dotchart(mpg, name, xlab = "mpg"))
> with(Auto.sample, dotchart(cylinders, mpg, ylab = "mpg"))
> with(Auto, plot(cylinders, mpg, ylab = "mpg"))
> with(Auto, plot(cylinders, weight))
>

File Edit Code View Plots Session Build Debug Profile Tools Help

R packages available: [] Chandra_Arunmukul_Asgnmen... | conda |

```

65 #apply(Auto, -qualitative_columns, sd)
66 #2.(d) now remove the 10th through 80th observations. what is the range, mean, and standard deviation of each
67 #subset <- Auto[-c(10:80), -c(4,9)]
68 #subset <- Auto[-c(10:80), -c(4,9)]
69 #subset <- Auto[-c(10:80), -c(4,9)]
70 #apply(subset, range)
71 #apply(subset, mean)
72 #2.(d) using the full data set, investigate the predictors graphically,using scatterplots or other tools of your choice.
73 #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
74 #pairs(Auto)
75 #pairs(Auto[, qualitative_columns])
76 #with(Auto, plot(mpg, weight))
77 #with(Auto, plot(mpg, cylinders))
78 #Auto.sample <- Auto[sample(1:nrow(Auto), 20), ]
79 #Auto.sample <- Auto.sample[order(Auto.sample$mpg), ]
80 #with(Auto.sample, dotchart(mpg, name, xlab = "mpg"))
81 #with(Auto.sample, dotchart(cylinders, mpg, ylab = "mpg"))
82 #with(Auto, plot(cylinders, mpg, ylab = "mpg"))
83 #with(Auto, plot(cylinders, weight))
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

771 [Top Level] R Script

Console Terminal Jobs

R 4.1.2 - CRAN

> #create some plots highlighting the relationships among the predictors. Comment on your findings
> Auto\$cylinders <- as.factor(Auto\$cylinders)
> Auto\$year <- as.factor(Auto\$year)
> Auto\$origin <- as.factor(Auto\$origin)
> pairs(Auto)
Error in pairs.default(Auto) : non-numeric argument to 'pairs'
> pairs(Auto[, non-numeric argument to 'pairs'
> #2.(d) using the full data set, investigate the predictors graphically,using scatterplots or other tools of your choice.
> #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
> pairs(Auto[, qualitative_columns])
> with(Auto, plot(mpg, weight))
> with(Auto, plot(mpg, cylinders))
> Auto.sample <- Auto[sample(1:nrow(Auto), 20),]
> Auto.sample <- Auto.sample[order(Auto.sample\$mpg),]
> with(Auto.sample, dotchart(mpg, name, xlab = "mpg"))
> with(Auto.sample, dotchart(cylinders, mpg, ylab = "mpg"))
> with(Auto, plot(cylinders, mpg, ylab = "mpg"))
> with(Auto, plot(cylinders, weight))
>

File Edit Code View Plots Session Build Debug Profile Tools Help

R packages available: [] Chandra_Arunmukul_Asgnmen... | conda |

```

65 #apply(Auto, -qualitative_columns, sd)
66 #2.(d) now remove the 10th through 80th observations. what is the range, mean, and standard deviation of each
67 #subset <- Auto[-c(10:80), -c(4,9)]
68 #subset <- Auto[-c(10:80), -c(4,9)]
69 #subset <- Auto[-c(10:80), -c(4,9)]
70 #apply(subset, sd)
71 #apply(subset, sd)
72 #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
73 #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
74 #pairs(Auto[, qualitative_columns])
75 #pairs(Auto[, qualitative_columns])
76 #with(Auto, plot(mpg, weight))
77 #with(Auto, plot(mpg, cylinders))
78 #for 20 observations
79 #Auto.sample <- Auto[sample(1:nrow(Auto), 20), ]
80 #ordering them
81 #plot by using "dotchart"
82 #with(Auto.sample, dotchart(mpg, name, xlab = "mpg"))
83 #with(Auto.sample, dotchart(cylinders, mpg, ylab = "mpg"))
84 #with(Auto, plot(cylinders, mpg, ylab = "mpg"))
85 #with(Auto, plot(cylinders, weight))
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

821 [Top Level] R Script

Console Terminal Jobs

R 4.1.2 - CRAN

> #create some plots highlighting the relationships among the predictors. Comment on your findings
> Auto\$cylinders <- as.factor(Auto\$cylinders)
> Auto\$year <- as.factor(Auto\$year)
> Auto\$origin <- as.factor(Auto\$origin)
> pairs(Auto)
Error in pairs.default(Auto) : non-numeric argument to 'pairs'
> pairs(Auto[, non-numeric argument to 'pairs'
> #2.(d) using the full data set, investigate the predictors graphically,using scatterplots or other tools of your choice.
> #Kreate some plots highlighting the relationships among the predictors. Comment on your findings
> pairs(Auto[, qualitative_columns])
Error in pairs.default(Auto) : non-numeric argument to 'pairs'
> pairs(Auto[, qualitative_columns])
> with(Auto, plot(mpg, weight))
> with(Auto, plot(mpg, cylinders))
> with(Auto, plot(mpg, weight))
> with(Auto, plot(mpg, cylinders))
> for 20 observations
> Auto.sample <- Auto[sample(1:nrow(Auto), 20),]
> ordering them
> Auto.sample <- Auto.sample[order(Auto.sample\$mpg),]
>

The screenshot shows the RStudio interface with the following components:

- Script Editor:** Displays the R code for the assignment.
- Environment:** Shows the global environment with variables like `Auto`, `Auto.sample`, `college`, and `qualitative_columns`.
- Plots:**
 - A scatter plot of `weight` vs `mpg` showing a strong negative correlation.
 - A box plot of `origin` vs `mpg` showing median mpg values for different car origins.

2.f

3.a

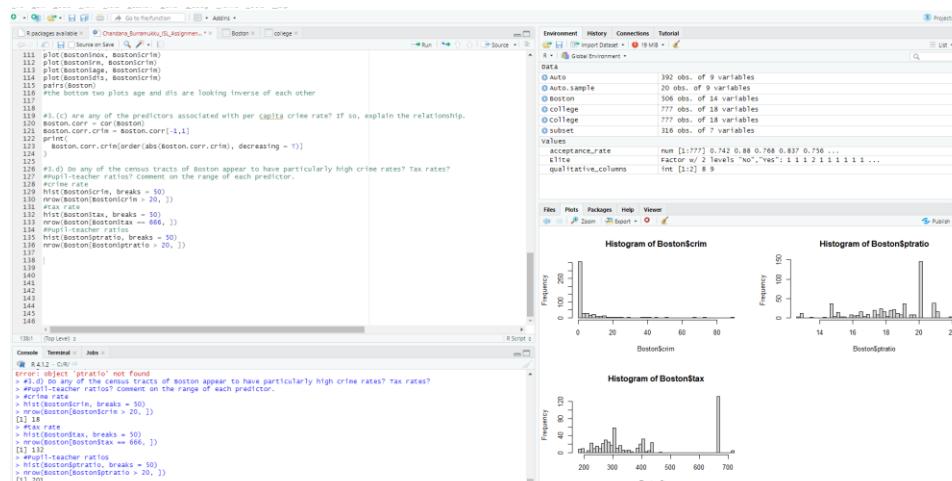
The screenshot shows the RStudio interface with the following details:

- Environment:** Shows objects like `auto`, `auto.sample`, `Boston`, `cylinders`, `college`, `subset`, `values`, `acceptance_rate`, `Elitism`, and `qualitative_columns`.
- Data View:** Displays the first few rows of the `auto` dataset.
- Plots:** Two scatter plots are displayed. The top plot shows `mpg` on the x-axis (ranging from 10 to 45) and `weight` on the y-axis (ranging from 1000 to 4500). The bottom plot shows `cylinders` on the y-axis (ranging from 1 to 8) and `mpg` on the x-axis (ranging from 10 to 45).
- Code Editor:** The code editor contains R scripts for data manipulation and visualization, including loading datasets, creating sample datasets, and plotting relationships between variables like `mpg`, `horsepower`, `displacement`, and `cylinders`.

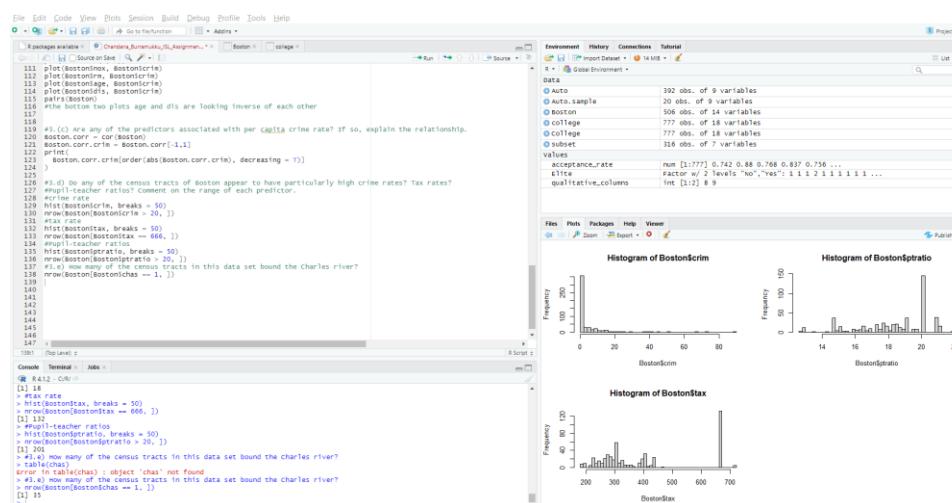
3.b

3.c

3.d



3.e



3.f

```

111 plot(boston$nox, boston$crime)
112 plot(boston$nitro, boston$crime)
113 plot(boston$dis, boston$crime)
114 plot(boston$rad, boston$crime)
115 plot(boston$tax, boston$crime)
116 #the better two plots age and dis are looking inverse of each other
117
118 #(.1.) are any of the predictors associated with per capita crime rate? if so, explain the relationship.
119 boston.corr = cor(boston)
120 boston.corr.crim = -0.1
121
122 Boston.corr.crim[order(abs(Boston.corr.crim), decreasing = TRUE)]
123
124
125 #(.d.) do any of the census tracts of Boston appear to have particularly high crime rates? Tax rates?
126 ##pupil-teacher ratios? Comment on the range of each predictor.
127
128 hist(boston$tax, breaks = 50)
129 hist(boston$dis, breaks = 20, )
130 hist(boston$rad, breaks = 50)
131 hist(boston$nox, breaks = 50)
132 hist(boston$age, breaks = 50)
133 nrow(Boston$Bostoncrim >= 666, )
134 ##pupil-teacher ratios
135 hist(boston$tax, breaks = 50)
136 nrow(Boston$Bostonpratio >= 20, )
137 #(.e.) how many of the census tracts in this data set bound the Charles river?
138 nrow(Boston$Bostonchas == 1, )
139 #(.f.) what is the median pupil-teacher ratio among the towns in this data set?
140 median(Boston$pratio)
141
142
143
144
145
146
147

```

(Top Level) Terminal > Jobs >

```

R 4.1.2 - C:\R\...
> #row(Boston$Bostontax == 666, )
148 [1] 1
149 ##pupil-teacher ratios
150 hist(boston$tax, breaks = 50)
151 nrow(Boston$Bostonpratio >= 20, )
152 [1] 20
153 #(.g.) are any of the census tracts in this data set bound the Charles river?
154 #<table>(chaz <-- table(chaz))
155 #<table>(chaz)
156 #(.h.) what is the median pupil-teacher ratio among the towns in this data set?
157 median(Boston$pratio)
158
159
160

```

3.g

```

141 [1] 1
142 #(.g.) which census tract of Boston has lowest median value on unenclosed houses? what are the values of the other predictors
143 #for that census tract, and how do those values compare to the overall ranges for those predictors? Comment on your findings.
144 tisubset(boston,medv==min(boston$medv)))
145 range(boston$tax)
146 Boston$tax[min(boston$medv),]
147
148
149
150
151
152
153
154
155
156
157
158
159
160

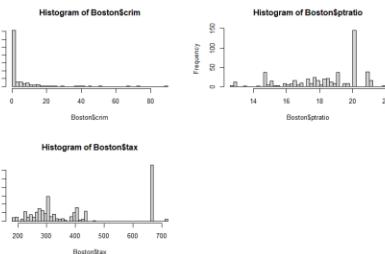
```

(Top Level) Terminal > Jobs >

```

R 4.1.2 - C:\R\...
> #(.g.) which census tract of Boston has lowest median value on unenclosed houses? what are the values of the other predictors
161 #for that census tract, and how do those values compare to the overall ranges for those predictors? Comment on your findings.
162 tisubset(boston,medv==min(boston$medv)))
163 range(boston$tax)
164 Boston$tax[min(boston$medv),]
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180

```



3.h

```

177
178 #(.g.) are any of the predictors associated with per capita crime rate? if so, explain the relationship.
179 boston.corr = cor(boston)
180 boston.corr.crim = -0.1
181
182 Boston.corr.crim[order(abs(Boston.corr.crim), decreasing = TRUE)]
183
184
185 #(.d.) do any of the census tracts of Boston appear to have particularly high crime rates? Tax rates?
186 ##pupil-teacher ratios? Comment on the range of each predictor.
187
188 hist(boston$tax, breaks = 50)
189 hist(boston$dis, breaks = 20, )
190 hist(boston$rad, breaks = 50)
191 hist(boston$nox, breaks = 50)
192 hist(boston$age, breaks = 50)
193 nrow(Boston$Bostoncrim >= 666, )
194 ##pupil-teacher ratios
195 hist(boston$tax, breaks = 50)
196 nrow(Boston$Bostonpratio >= 20, )
197 #(.e.) how many of the census tracts in this data set bound the Charles river?
198 nrow(Boston$Bostonchas == 1, )
199 #(.f.) what is the median pupil-teacher ratio among the towns in this data set?
200 median(Boston$pratio)
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220

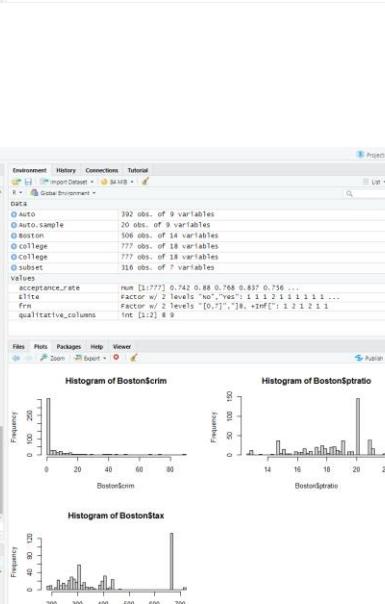
```

(Top Level) Terminal > Jobs >

```

R 4.1.2 - C:\R\...
> Boston$tax <- medv<-30, '
Error in r <- : ...
complain: 'r' is possible only for atomic and list types
> range(boston$tax)
[1] 64
> nrow(boston$Bostonpratio >= 8, )
[1] 13
>

```



4.a.i

The screenshot shows the RStudio interface with the following details:

- Console:**

```
R> #4.a.i: Is there a relationship between the predictor and the response?
R> lm(Boston$displacement ~ horsepower, data = Auto)
<--> #4.a.i: Is there a relationship between the predictor and the response?
```
- Environment:** Shows objects like `Auto`, `Auto.sample`, `Boston`, `college`, `Im.Fit`, `subset`, `valence`, `acceptance_rate`, `slite`, `fit`, and `qualitative_columns`.
- Plots:**
 - Histogram of `Boston$displacement`
 - Histogram of `Boston$tax`
 - Histogram of `Boston$nox`

4.a.ii

R file

4.a.iii

R file

4.a.iv

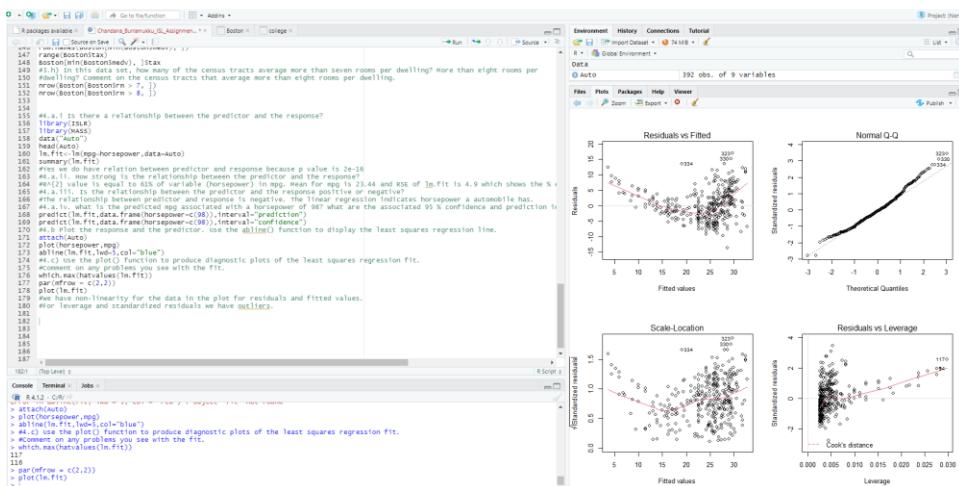
The screenshot shows the RStudio interface with the following details:

- Console:**

```
R> #4.a.iv: Is there a relationship between the predictor and the response?
R> lm(Boston$displacement ~ horsepower, data = Auto)
<--> #4.a.iv: Is there a relationship between the predictor and the response?
```
- Environment:** Shows objects like `Auto`, `Auto.sample`, `Boston`, `college`, `Im.Fit`, `subset`, `valence`, `acceptance_rate`, `slite`, `fit`, and `qualitative_columns`.
- Plots:**
 - Histogram of `Boston$displacement`
 - Histogram of `Boston$tax`
 - Histogram of `Boston$nox`

4.b

4.c



5.a

5.b

The figure shows two side-by-side panels from RStudio. The left panel displays a scatterplot matrix for the 'mtcars' dataset, with variables including mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb, and origin. The right panel shows diagnostic plots for a linear regression model, including histograms of residuals and standardized residuals, and plots of residuals versus fitted values.

5.c.i

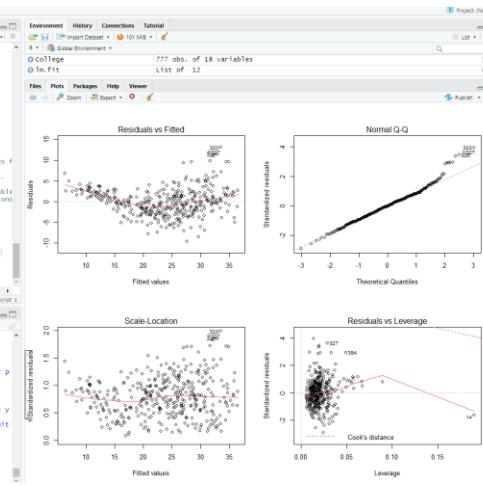
5.c.ii

R file

5.c.iii

R file

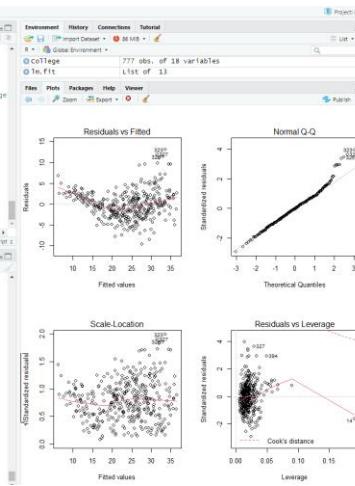
5.d



5.e

```
# package exists -> library(lmtest)
# <-- source('BivariateLM.R')
# <-- lmtest::lemon(fit)

193 #, c.11, which predictors appear to have a statistically significant relationship to the response.
194 #> #> We can see this by looking at the p-values. We can also do this by seeing the p-value
195 #> #> c.11, does that coefficient for the year variable suggest?
196 #> #> The p-value of 0.0001 suggests that the coefficients are nearly 1 year from fuel economy.
197 #> #> The coefficient of year variable shows the effect of raise: For one year it is equivalent to 6.75 in mpg.
198 #> #> The residual function to produce diagnostic plots of the linear regression fit. Comment on any problem
199 #> #> residuals(fit) to identify any unusual or large residuals from the regression fit.
200 #> #> which(max(abs(residuals(fit))))
201 #> #> plot(fit)
202 #> #> plot(fit, 1)
203 #> #> We have a non-linear relationship between the response and predictors for the first graph.
204 #> #> lm(fit) residuals are distributed normally and are skewed towards right for second one.
205 #> #> plot(fit, 2)
206 #> #> lm(fit) residuals are distributed normally and are skewed towards left for third one.
207 #> #> We do not have any leverage points for fourth graph, but this one stands out as potential leverage point.
208 #> #> lm(fit) leverage
209 #> #> lm(fit) influence
210 #> #> summary(fit)
211 #> #> We have a statistically significant interaction between displacement and weight for p-values.
212 #> #> lm(fit)
213 #> #> lm(fit)
214
215 #> #> lm(fit)
216
```



5.f

The figure consists of three vertically stacked screenshots of the RStudio interface.

- Top Screenshot:** Shows the R console with the following code and output. The code includes a plot of log(mf1) against displacement, a summary of lm.f1, and a scatter plot of log(placement) vs log(displacement). The output shows coefficients for a model with interaction terms between displacement and weight.

```

#> 202 plot(log(mf1))
#> 203 #> Non-linear relationship between the response and predictors. For the first graph.
#> 204 #> The residuals are distributed normally and are skewed towards right for second one.
#> 205 #> For this model the error assumption is not true for constant variance for third graph.
#> 206 #> We can use log transformation to overcome this problem. This is called residual leverage point.
#> 207 #> r.sq use the r.sq symbols to fit linear regression models with interaction effects.
#> 208 #> log(placement) ~ log(displacement)*weight, data = Auto
#> 209 lm.f1 <- lm(log(placement) ~ log(displacement)*weight, data = Auto)
#> 210 summary(lm.f1)
#> 211 #> we have statistically significant interaction between displacement and weight for p-values.
#> 212 #> we do not have any interaction between displacement and cylinders.
#> 213 #> we can use log transformation of the variables, such as log(X), yX, x2, comment on your findings.
#> 214 par(mfrow=c(1,2), mar=c(5,4,4,2))
#> 215 plot(log(placement), Autolab)
#> 216 plot(log(placement), Autolab)
#> 217 plot(log(placement), Autolab)
#> 218 lm.f1 <- lm(log(placement)~log(displacement)+log(displacement)*weight, data = Auto)
#> 219 summary(lm.f1)
#> 220 #> we get the most linear plot for log transformation. we are predicting all the things based on horsepower.
#> 221
#> 222
```

- Middle Screenshot:** Shows the Environment tab with a global environment containing a variable 'college' with 77 observations and 18 variables. It also shows a scatter plot of 'Ausding' vs 'log(AutoHorsepower)'.

6.a

6.b

R file

6.c

R file

6.d

R file

```
Project [Boston] Environment History Connections Tutorial
File [Fits] Import Dataset [0 MB] Help
File [fits] Close Environment [List of 11]
File [lm.flt] List of 13
File [Cansats] Cansats [ISLR]
File [Plots] Packages Help Viewer
R. Sales of Child Car Seats - Find in Topics
Cansats [ISLR]
R Documentation



## Sales of Child Car Seats



Description



A simulated data set containing sales of child car seats at 400 different stores.



Usage



Cansats



Format



A data frame with 400 observations on the following 11 variables.



| Call:                           | formula = Sales ~ Price + US, data = Cansats                    |
|---------------------------------|-----------------------------------------------------------------|
| <b>Residuals:</b>               | [1] 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 |
| <b>Min:</b>                     | [1] -6.9269 -1.6286 -0.0574 1.7986 7.0155                       |
| <b>Median:</b>                  | [1] 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 |
| <b>Max:</b>                     | [1] 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 |
| <b>coefficients:</b>            | [1] (Intercept) std. error t value Pr(> t )                     |
|                                 | [1] 11.0309 0.63098 20.612 <2e-16 ***                           |
|                                 | [2] US -0.034689 -0.0044588 -0.08627 <2e-16 ***                 |
|                                 | [3] Price 0.119894 0.218847 4.645 4.24e-06 ***                  |
| <b>Signif. codes:</b>           | [1] **** ** * . 1                                               |
| <b>Residual standard error:</b> | [1] 2.469 on 397 degrees of freedom                             |
| <b>Multiple R-squared:</b>      | [1] 0.2399,                                                     |
| <b>Adjusted R-squared:</b>      | [1] 0.2354                                                      |
| <b>F-statistic:</b>             | [1] 40.43 on 2 and 397 DF, p-value: < 2.2e-16                   |



[1]


```

6.f

R file

6.g

6.h

