

ISLR Chapter2

Labs

ACO

2025-01-03

Contents

Q8	2
(a)	2
(b)	2
(c)	2
i	3
ii	4
iii	5
iv	5
v	7
vi	8
Q9	9
(a)	9
(b)	9
(c)	10
(d)	10
(e)	11
(f)	11
Q10	12
(a)	12
(b)	13
(c)	13
(d)	13
(e)	14
(f)	14
(g)	15
(h)	15

```
#  
library(ISLR)  
library(RColorBrewer)
```

Q8

(a)

```
# (a)  
College <- read.csv("College.csv")  
head(College)

##                                     X Private Apps Accept Enroll Top10perc Top25perc  
## 1 Abilene Christian University     Yes 1660   1232    721      23      52  
## 2 Adelphi University             Yes 2186   1924    512      16      29  
## 3 Adrian College                Yes 1428   1097    336      22      50  
## 4 Agnes Scott College            Yes  417    349     137      60      89  
## 5 Alaska Pacific University     Yes  193    146      55      16      44  
## 6 Albertson College              Yes  587    479     158      38      62  
##   F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD Terminal  
## 1      2885          537    7440     3300    450    2200    70      78  
## 2      2683          1227   12280     6450    750    1500    29      30  
## 3      1036           99   11250     3750    400    1165    53      66  
## 4       510            63   12960     5450    450     875    92      97  
## 5       249            869   7560     4120    800    1500    76      72  
## 6       678            41  13500     3335    500     675    67      73  
##   S.F.Ratio perc.alumni Expend Grad.Rate  
## 1      18.1           12   7041      60  
## 2      12.2           16  10527      56  
## 3      12.9           30  8735      54  
## 4       7.7           37 19016      59  
## 5      11.9            2 10922      15  
## 6      9.4            11  9727      55
```

(b)

```
# (b)  
rownames(College) <- College[,1]  
College <- College[,-1]
```

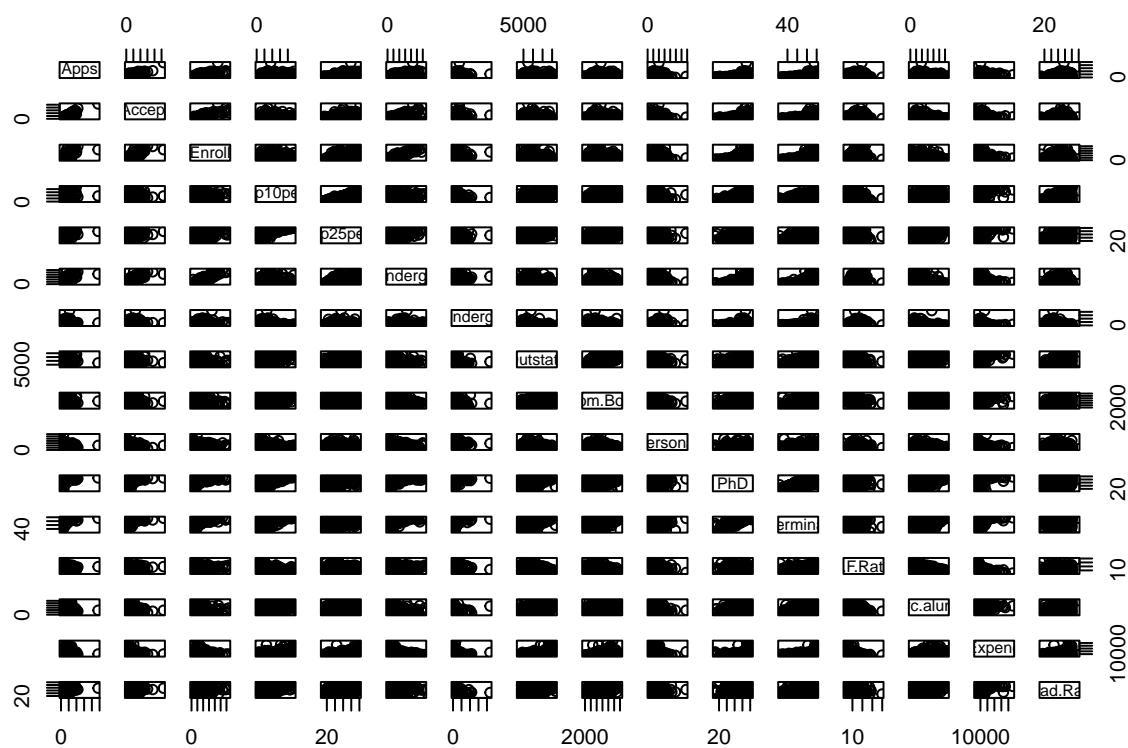
(c)

```
# (c) i  
num_cols = unlist(lapply(College, is.numeric), use.names = F)  
summary(College[,num_cols])
```

i

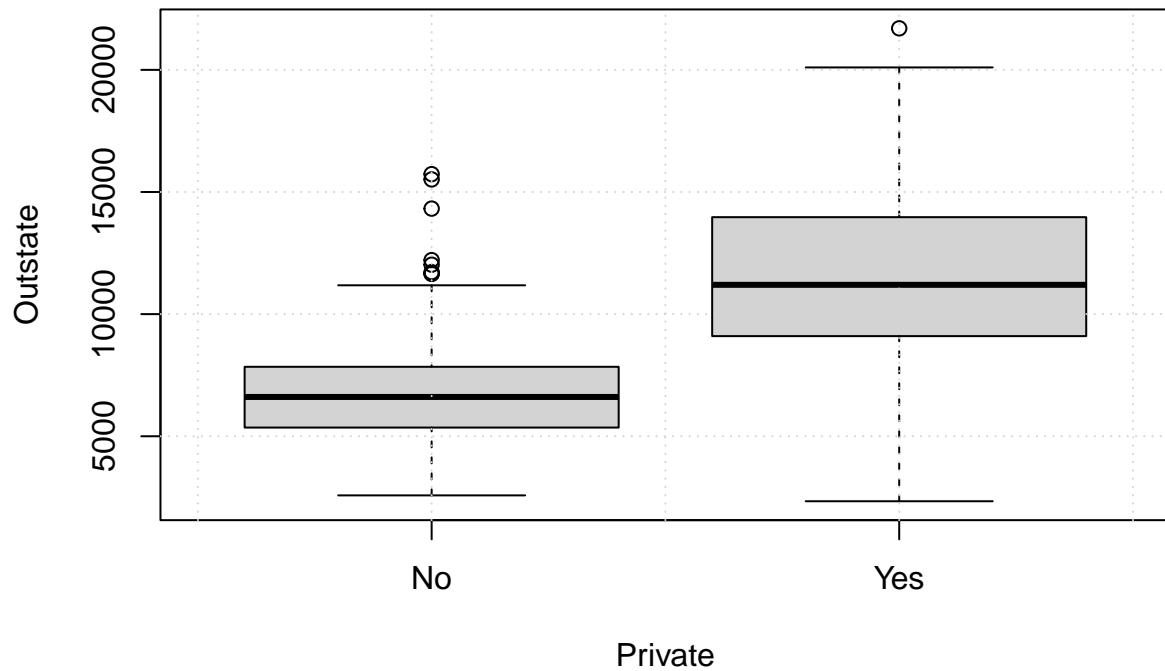
```
##      Apps        Accept       Enroll     Top10perc     Top25perc
##  Min.   :  81   Min.   :  72   Min.   : 35   Min.   :1.00   Min.   : 9.0
##  1st Qu.: 776   1st Qu.: 604   1st Qu.:242   1st Qu.:15.00  1st Qu.:41.0
##  Median :1558   Median :1110   Median :434    Median :23.00  Median :54.0
##  Mean   :3002   Mean   :2019   Mean   :780    Mean   :27.56  Mean   :55.8
##  3rd Qu.:3624   3rd Qu.:2424   3rd Qu.:902   3rd Qu.:35.00  3rd Qu.:69.0
##  Max.   :48094  Max.   :26330  Max.   :6392   Max.   :96.00  Max.   :100.0
##      F.Undergrad    P.Undergrad      Outstate     Room.Board
##  Min.   : 139   Min.   : 1.0   Min.   :2340   Min.   :1780
##  1st Qu.: 992   1st Qu.: 95.0  1st Qu.:7320   1st Qu.:3597
##  Median :1707   Median :353.0  Median :9990   Median :4200
##  Mean   :3700   Mean   :855.3  Mean   :10441  Mean   :4358
##  3rd Qu.:4005   3rd Qu.:967.0  3rd Qu.:12925 3rd Qu.:5050
##  Max.   :31643  Max.   :21836.0 Max.   :21700  Max.   :8124
##      Books        Personal      PhD        Terminal
##  Min.   : 96.0   Min.   :250   Min.   : 8.00   Min.   :24.0
##  1st Qu.:470.0   1st Qu.:850   1st Qu.:62.00   1st Qu.:71.0
##  Median :500.0   Median :1200   Median :75.00   Median :82.0
##  Mean   :549.4   Mean   :1341   Mean   :72.66   Mean   :79.7
##  3rd Qu.:600.0   3rd Qu.:1700   3rd Qu.:85.00   3rd Qu.:92.0
##  Max.   :2340.0  Max.   :6800   Max.   :103.00  Max.   :100.0
##      S.F.Ratio    perc.alumni      Expend     Grad.Rate
##  Min.   : 2.50   Min.   : 0.00   Min.   :3186   Min.   : 10.00
##  1st Qu.:11.50  1st Qu.:13.00  1st Qu.:6751   1st Qu.: 53.00
##  Median :13.60  Median :21.00  Median :8377   Median : 65.00
##  Mean   :14.09  Mean   :22.74  Mean   :9660   Mean   : 65.46
##  3rd Qu.:16.50  3rd Qu.:31.00  3rd Qu.:10830  3rd Qu.: 78.00
##  Max.   :39.80  Max.   :64.00  Max.   :56233  Max.   :118.00
```

```
# (c) ii
pairs(College[,num_cols[1:10]])
```



ii

```
# (c) iii
boxplot(Outstate ~ Private, data = College)
grid()
```

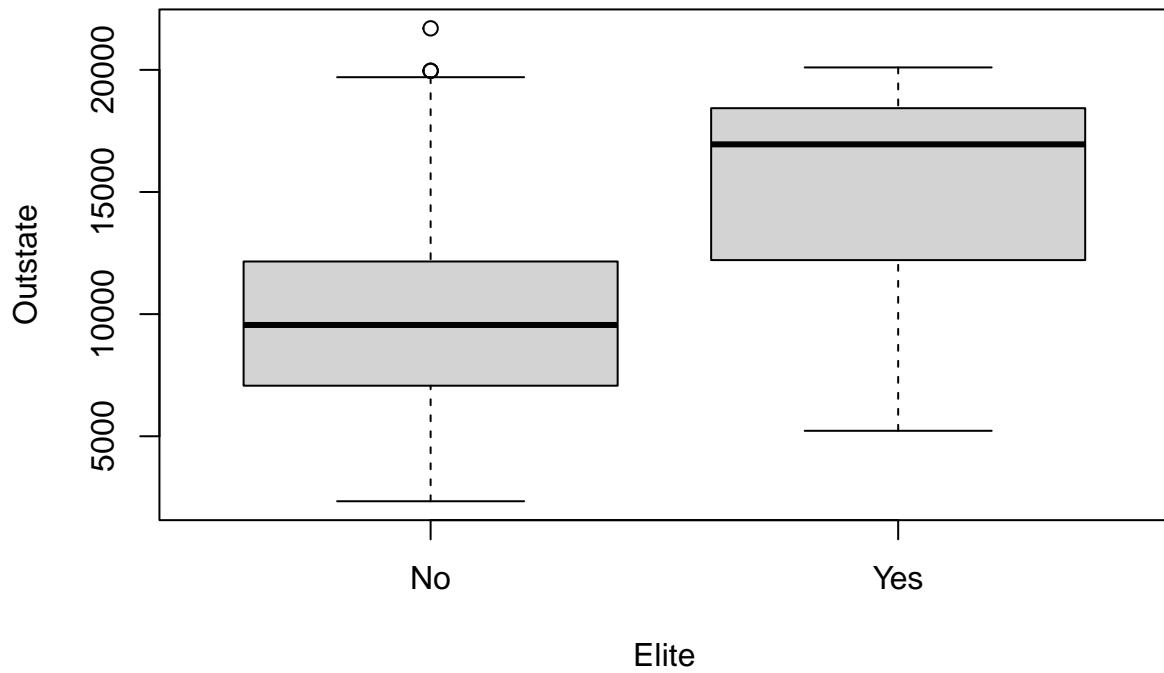


iii

```
# (c) iv
College$Elite <- "No"
College$Elite[College$Top10perc > 50] <- "Yes"
#
summary(College$Elite)
```

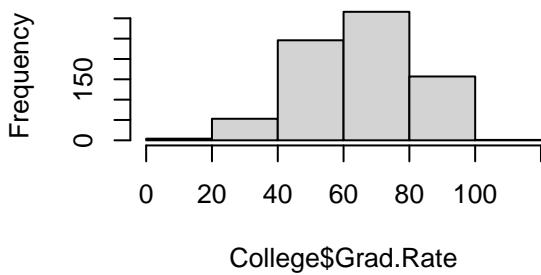
iv

```
##      Length     Class      Mode
##      777 character character
#
# boxplot(Outstate ~ Elite, data = College)
```

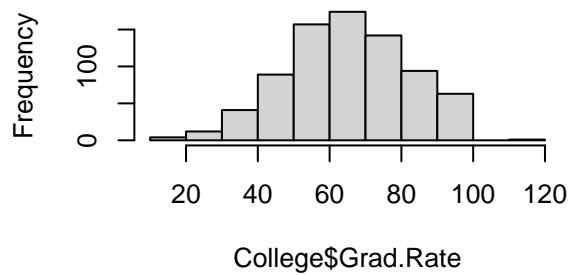


```
# (c) v
par(mfrow = c(2,2))
hist(College$Grad.Rate, breaks = 5)
hist(College$Grad.Rate, breaks = 10)
hist(College$Grad.Rate, breaks = 15)
hist(College$Grad.Rate, breaks = 20)
```

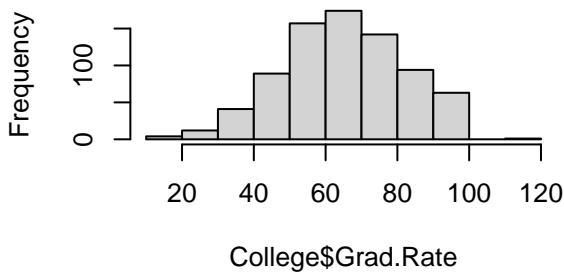
Histogram of College\$Grad.Rate



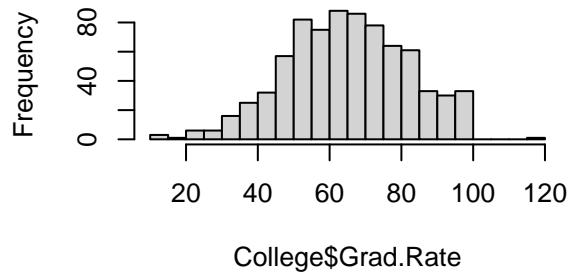
Histogram of College\$Grad.Rate



Histogram of College\$Grad.Rate



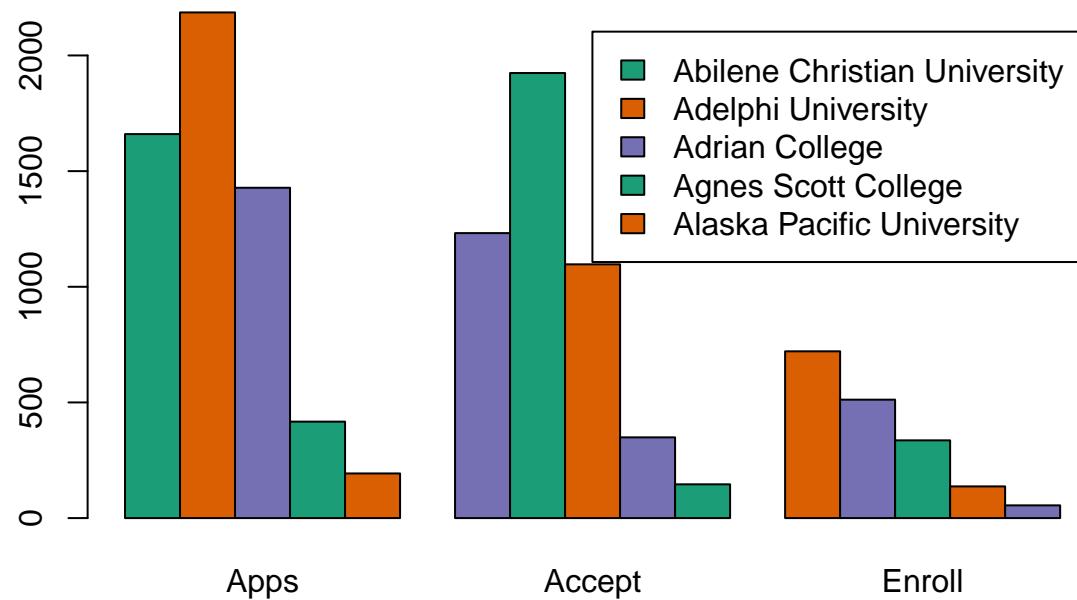
Histogram of College\$Grad.Rate



v

```
par(mfrow = c(1,1))
```

```
# (c) vi
bar_plot = barplot(
  as.matrix(College[1:5,2:4]), beside = T,
  legend.text = rownames(College[1:5,]), col = brewer.pal(3, "Dark2")
)
```



vi

Q9

(a)

```
data("Auto")
#
str(Auto)

## 'data.frame':   392 obs. of  9 variables:
## $ mpg          : num  18 15 18 16 17 15 14 14 14 15 ...
## $ cylinders    : num  8 8 8 8 8 8 8 8 8 ...
## $ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
## $ horsepower   : num  130 165 150 150 140 198 220 215 225 190 ...
## $ weight       : num  3504 3693 3436 3433 3449 ...
## $ acceleration: num  12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
## $ year         : num  70 70 70 70 70 70 70 70 70 70 ...
## $ origin       : num  1 1 1 1 1 1 1 1 1 1 ...
## $ name         : Factor w/ 304 levels "amc ambassador brougham",...: 49 36 231 14 161 141 54 223 241 241 ...

# Predictors
# Quantitative:
names(Auto)[unlist(lapply(Auto, is.numeric)), use.names = T]

## [1] "mpg"           "cylinders"      "displacement"   "horsepower"     "weight"
## [6] "acceleration" "year"          "origin"

# Qualitative:
names(Auto)[!unlist(lapply(Auto, is.numeric)), use.names = T]

## [1] "name"
```

(b)

```
# (b)
numeric_cols <- unlist(lapply(Auto, is.numeric), use.names = T)
numeric_cols

##      mpg    cylinders displacement horsepower      weight acceleration
##      TRUE      TRUE        TRUE      TRUE      TRUE        TRUE
##      year    origin      name
##      TRUE      TRUE       FALSE

apply(Auto[,numeric_cols], MARGIN = 2, FUN = function(x) range(x))

##      mpg cylinders displacement horsepower weight acceleration year origin
## [1,]  9.0        3            68          46  1613        8.0       70      1
## [2,] 46.6        8            455         230  5140       24.8       82      3
```

(c)

```
# (c)
apply(Auto[,numeric_cols], MARGIN = 2, FUN = function(x) mean(x))

##          mpg cylinders displacement horsepower      weight acceleration
##    23.445918      5.471939     194.411990     104.469388   2977.584184    15.541327
##          year      origin
##    75.979592      1.576531

apply(Auto[,numeric_cols], MARGIN = 2, FUN = function(x) sd(x))

##          mpg cylinders displacement horsepower      weight acceleration
##    7.8050075     1.7057832    104.6440039     38.4911599   849.4025600    2.7588641
##          year      origin
##    3.6837365     0.8055182

apply(Auto[,numeric_cols], MARGIN = 2, FUN = function(x) summary(x))

##          mpg cylinders displacement horsepower      weight acceleration
## Min.    9.000000 3.0000000      68.00000 46.00000 1613.00000      8.000000
## 1st Qu. 17.000000 4.0000000     105.00000 75.00000 2225.25000     13.775000
## Median  22.750000 4.0000000     151.00000 93.50000 2803.50000     15.500000
## Mean    23.445920 5.4719390     194.41200 104.46940 2977.58400     15.541330
## 3rd Qu. 29.000000 8.0000000     275.75000 126.00000 3614.75000     17.025000
## Max.   46.600000 8.0000000     455.00000 230.00000 5140.00000     24.800000
##          year      origin
## Min.    70.000000 1.0000000
## 1st Qu. 73.000000 1.0000000
## Median  76.000000 1.0000000
## Mean    75.979591 1.576531
## 3rd Qu. 79.000000 2.0000000
## Max.   82.000000 3.0000000
```

(d)

```
# (d)
apply(Auto[c(-10,-85),numeric_cols], MARGIN = 2, FUN = function(x) summary(x))

##          mpg cylinders displacement horsepower      weight acceleration
## Min.    9.000000 3.0000000      68.00000 46.00000 1613.00000      8.00000
## 1st Qu. 17.500000 4.0000000     105.00000 75.00000 2223.75000     13.82500
## Median  23.000000 4.0000000     148.50000 92.50000 2797.50000     15.50000
## Mean    23.494360 5.4589740     193.51150 104.06920 2972.46900     15.56590
## 3rd Qu. 29.000000 8.0000000     262.00000 125.00000 3608.00000     17.07500
## Max.   46.600000 8.0000000     455.00000 230.00000 5140.00000     24.80000
##          year      origin
## Min.    70.000000 1.0000000
```

```

## 1st Qu. 73.00000 1.000000
## Median  76.00000 1.000000
## Mean    76.00256 1.579487
## 3rd Qu. 79.00000 2.000000
## Max.    82.00000 3.000000

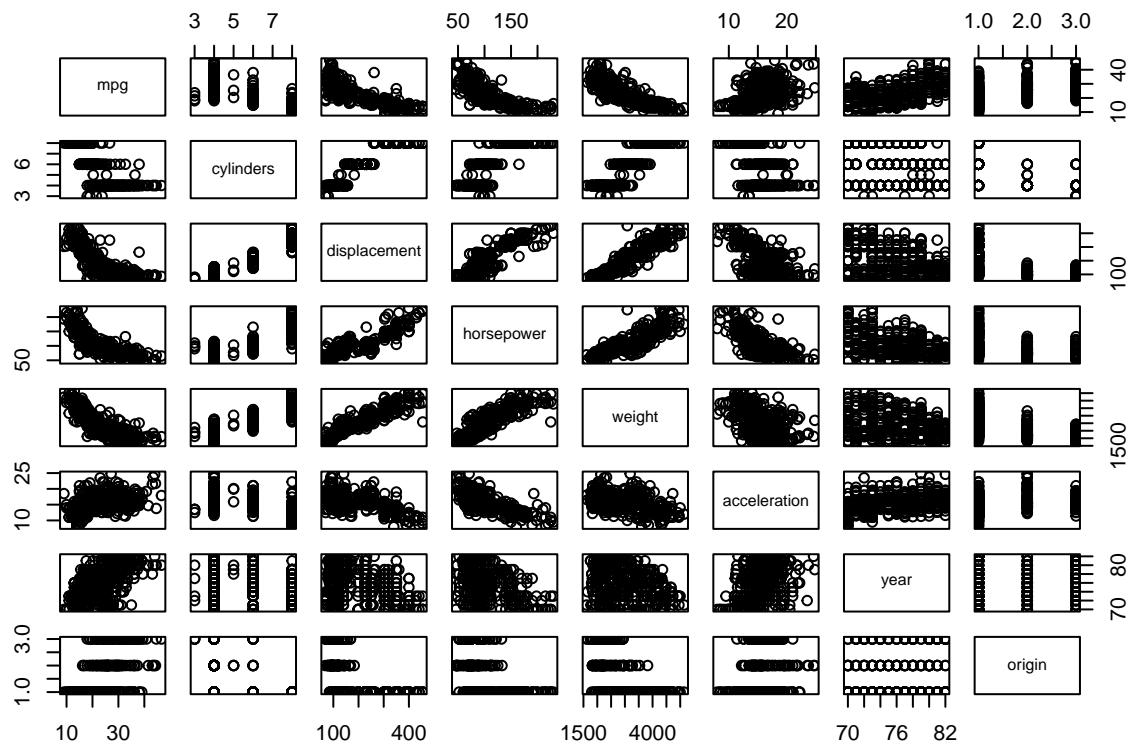
```

(e)

```

# (e)
pairs(Auto[, numeric_cols])

```

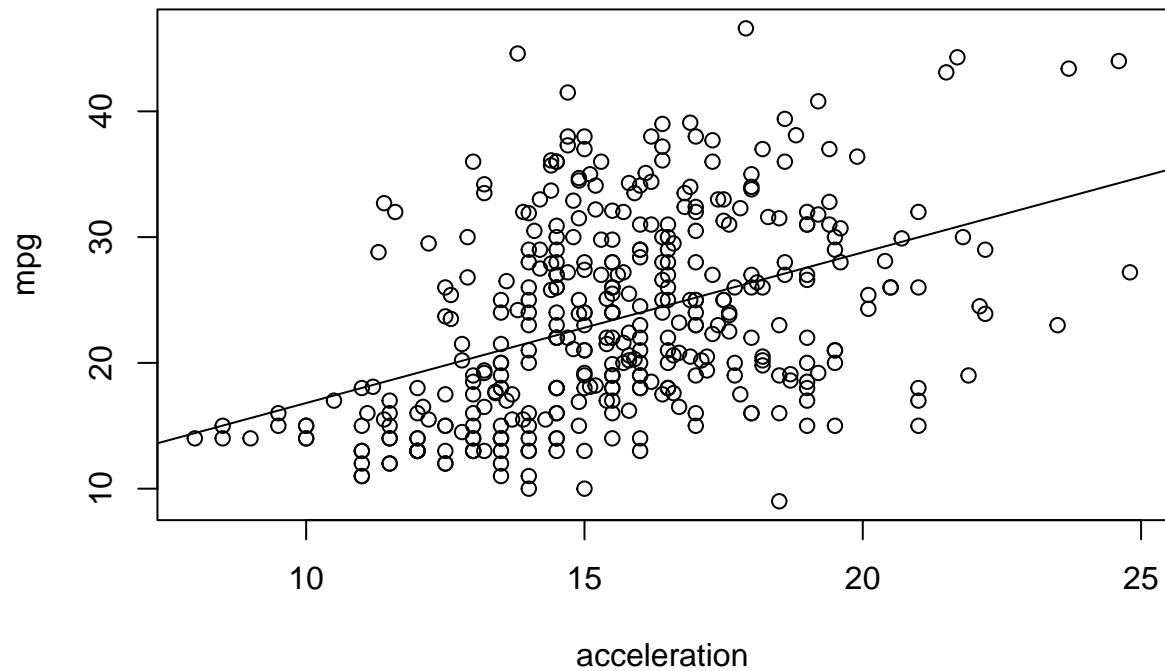


(f)

```

# (f)
# Acceleration
plot(mpg~acceleration, data = Auto)
abline(lm(mpg~acceleration, data = Auto))

```



Q10

```
# Q10 #####
```

(a)

```
# (a)
library(ISLR2)
```

```
##
## Attaching package: 'ISLR2'

## The following objects are masked _by_ '.GlobalEnv':
##   Auto, College

## The following objects are masked from 'package:ISLR':
##   Auto, Credit
```

```
data("Boston")
```

```
#
```

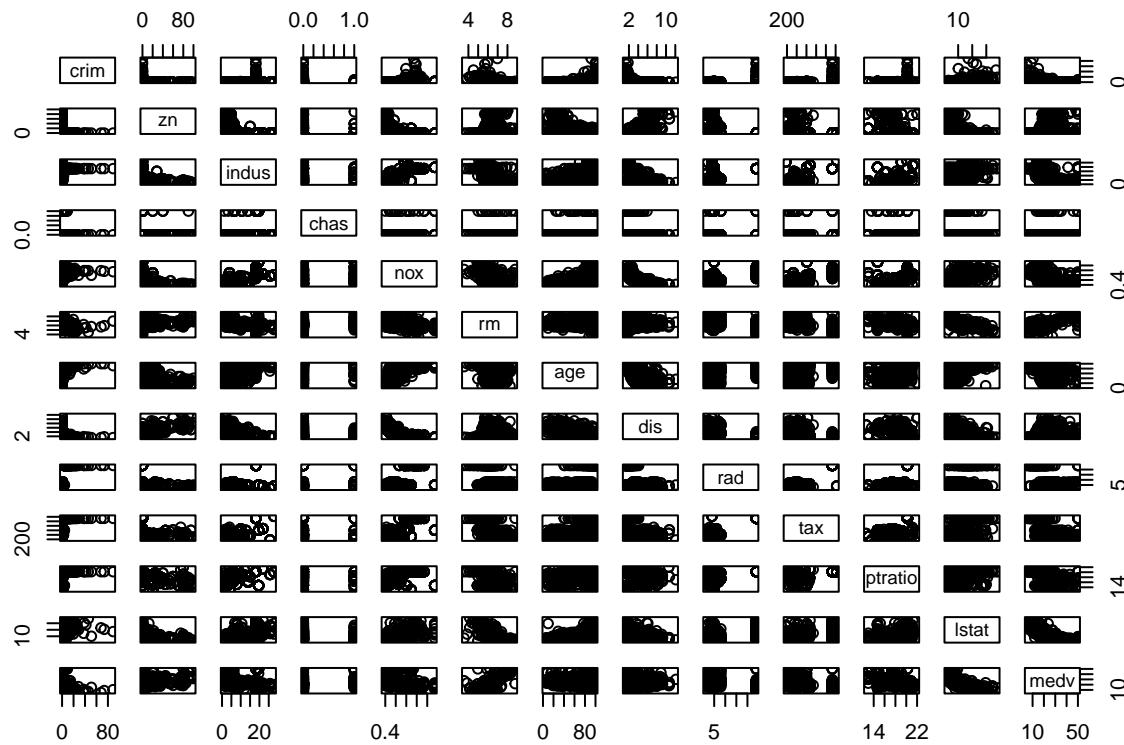
```
dim(Boston)
```

```
## [1] 506 13
```

(b)

```
# (b)
```

```
pairs(Boston)
```



(c)

```
# (c)
```

```
# relationship of crime to other variables not clearly defined via viz
```

(d)

```
# (d)
summary(Boston)
```

```
##      crim             zn            indus            chas
##  Min. : 0.00632   Min. : 0.00   Min. : 0.46   Min. :0.00000
##  1st Qu.: 0.08205  1st Qu.: 0.00   1st Qu.: 5.19   1st Qu.:0.00000
##  Median : 0.25651  Median : 0.00   Median : 9.69   Median :0.00000
##  Mean   : 3.61352  Mean   : 11.36  Mean   :11.14  Mean   :0.06917
##  3rd Qu.: 3.67708  3rd Qu.: 12.50  3rd Qu.:18.10  3rd Qu.:0.00000
##  Max.   :88.97620  Max.   :100.00  Max.   :27.74  Max.   :1.00000
##      nox              rm            age            dis
##  Min. :0.3850    Min. :3.561   Min. : 2.90   Min. : 1.130
##  1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02  1st Qu.: 2.100
##  Median :0.5380   Median :6.208   Median : 77.50  Median : 3.207
##  Mean   :0.5547   Mean   :6.285   Mean   : 68.57  Mean   : 3.795
##  3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08  3rd Qu.: 5.188
##  Max.   :0.8710   Max.   :8.780   Max.   :100.00  Max.   :12.127
##      rad              tax            ptratio          lstat
##  Min. : 1.000   Min. :187.0   Min. :12.60   Min. : 1.73
##  1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40  1st Qu.: 6.95
##  Median : 5.000   Median :330.0   Median :19.05   Median :11.36
##  Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :12.65
##  3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20  3rd Qu.:16.95
##  Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :37.97
##      medv
##  Min. : 5.00
##  1st Qu.:17.02
##  Median :21.20
##  Mean   :22.53
##  3rd Qu.:25.00
##  Max.   :50.00
```

(e)

```
# (e)
colnames(Boston)
```

```
## [1] "crim"      "zn"        "indus"      "chas"       "nox"        "rm"        "age"
## [8] "dis"        "rad"       "tax"        "ptratio"    "lstat"     "medv"
```

(f)

```
# (f)
# subset(Boston, chas == 1)
table(Boston$chas)
```

```
##
##    0    1
##  471  35
```

```
#  
median(Boston$ptratio)
```

```
## [1] 19.05
```

(g)

```
# (g)  
subset(Boston, medv == min(Boston$medv))
```

```
##      crim zn indus chas   nox     rm age     dis rad tax ptratio lstat medv  
## 399 38.3518 0 18.1    0 0.693 5.453 100 1.4896 24 666    20.2 30.59    5  
## 406 67.9208 0 18.1    0 0.693 5.683 100 1.4254 24 666    20.2 22.98    5
```

- high crim, per capita crime rate by town.
- low zn proportion of residential land zoned for lots over 25,000 sq.ft.
- high indus: proportion of non-retail business acres per town.
- no bounds to chas: Charles River
- high nox: nitrogen oxides concentration (parts per 10 million).
- close to 1st qu rm:average number of rooms per dwelling.
- maximum age: proportion of owner-occupied units built prior to 1940.
- above average dis: weighted mean of distances to five Boston employment centres
- max rad: index of accessibility to radial highways.
- 3rd Qu tax: full-value property-tax rate per \$10,000.
- 3rd Qu ptratio: pupil-teacher ratio by town.
- Above 3rd Qu lstat: lower status of the population (percent).
- minimum medv: median value of owner-occupied homes in \$1000s.

(h)

```
# (h)  
Boston[Boston$rm>7,] |> nrow()
```

```
## [1] 64
```

```
Boston[Boston$rm>8,] |> nrow()
```

```
## [1] 13
```

```
#
```