Rumainum-HW1

Lince Rumainum

August 27, 2019

Problem 1

```
# library
library(moments)
library(survival)

# Problem 1a
x <- c(3, 12, 6, -5, 0, 8, 15, 1, -10, 7)
x

##  [1]   3  12   6  -5   0   8  15   1 -10   7

# Problem 1b
y <- seq(min(x), max(x), length = 10)
y

##  [1] -10.000000  -7.222222  -4.444444  -1.666667   1.111111   3.888889
##  [7]   6.666667   9.444444  12.222222  15.000000

# Problem 1c
sum(x)

## [1] 37

mean(x)

## [1] 3.7

sd(x)

## [1] 7.572611

var(x)

## [1] 57.34444

mad(x)

## [1] 5.9304

quantile(x, prob=round(seq(0,1,length=4),digits=2))

##     0%    33%    67%   100%
## -10.00   0.97   7.03  15.00
```

```
quantile(x, prob=seq(0,1,length=5))

##     0%    25%    50%    75%   100%
## -10.00   0.25   4.50   7.75  15.00

sum(y)

## [1] 25

mean(y)

## [1] 2.5

sd(y)

## [1] 8.41014

var(y)

## [1] 70.73045

mad(y)

## [1] 10.29583

quantile(y, prob=round(seq(0,1,length=4),digits=2))

##     0%    33%    67%   100%
## -10.00  -1.75   6.75  15.00

quantile(y, prob=seq(0,1,length=5))

##     0%    25%    50%    75%   100%
## -10.00  -3.75   2.50   8.75  15.00

# Problem 1d
z = sample(x, size = 7, replace = TRUE)
z

## [1]  6 12 -5  6  7 15  1

# Problem 1e
data(kidney, package="survival")

kidney

##    id time status age sex disease frail
## 1   1    8      1  28   1   Other   2.3
## 2   1   16      1  28   1   Other   2.3
## 3   2   23      1  48   2      GN   1.9
## 4   2   13      0  48   2      GN   1.9
## 5   3   22      1  32   1   Other   1.2
## 6   3   28      1  32   1   Other   1.2
## 7   4  447      1  31   2   Other   0.5
## 8   4  318      1  32   2   Other   0.5
## 9   5   30      1  10   1   Other   1.5
## 10  5   12      1  10   1   Other   1.5
```

```
## 11  6   24       1  16   2   Other   1.1
## 12  6  245       1  17   2   Other   1.1
## 13  7    7       1  51   1      GN   3.0
## 14  7    9       1  51   1      GN   3.0
## 15  8  511       1  55   2      GN   0.5
## 16  8   30       1  56   2      GN   0.5
## 17  9   53       1  69   2      AN   0.7
## 18  9  196       1  69   2      AN   0.7
## 19 10   15       1  51   1      GN   0.4
## 20 10  154       1  52   1      GN   0.4
## 21 11    7       1  44   2      AN   0.6
## 22 11  333       1  44   2      AN   0.6
## 23 12  141       1  34   2   Other   1.2
## 24 12    8       0  34   2   Other   1.2
## 25 13   96       1  35   2      AN   1.4
## 26 13   38       1  35   2      AN   1.4
## 27 14  149       0  42   2      AN   0.4
## 28 14   70       0  42   2      AN   0.4
## 29 15  536       1  17   2   Other   0.4
## 30 15   25       0  17   2   Other   0.4
## 31 16   17       1  60   1      AN   1.1
## 32 16    4       0  60   1      AN   1.1
## 33 17  185       1  60   2   Other   0.8
## 34 17  177       1  60   2   Other   0.8
## 35 18  292       1  43   2   Other   0.8
## 36 18  114       1  44   2   Other   0.8
## 37 19   22       0  53   2      GN   0.5
## 38 19  159       0  53   2      GN   0.5
## 39 20   15       1  44   2   Other   1.3
## 40 20  108       0  44   2   Other   1.3
## 41 21  152       1  46   1     PKD   0.2
## 42 21  562       1  47   1     PKD   0.2
## 43 22  402       1  30   2   Other   0.6
## 44 22   24       0  30   2   Other   0.6
## 45 23   13       1  62   2      AN   1.7
## 46 23   66       1  63   2      AN   1.7
## 47 24   39       1  42   2      AN   1.0
## 48 24   46       0  43   2      AN   1.0
## 49 25   12       1  43   1      AN   0.7
## 50 25   40       1  43   1      AN   0.7
## 51 26  113       0  57   2      AN   0.5
## 52 26  201       1  58   2      AN   0.5
## 53 27  132       1  10   2      GN   1.1
## 54 27  156       1  10   2      GN   1.1
## 55 28   34       1  52   2      AN   1.8
## 56 28   30       1  52   2      AN   1.8
## 57 29    2       1  53   1      GN   1.5
## 58 29   25       1  53   1      GN   1.5
## 59 30  130       1  54   2      GN   1.5
## 60 30   26       1  54   2      GN   1.5
## 61 31   27       1  56   2      AN   1.7
## 62 31   58       1  56   2      AN   1.7
## 63 32    5       0  50   2      AN   1.3
```

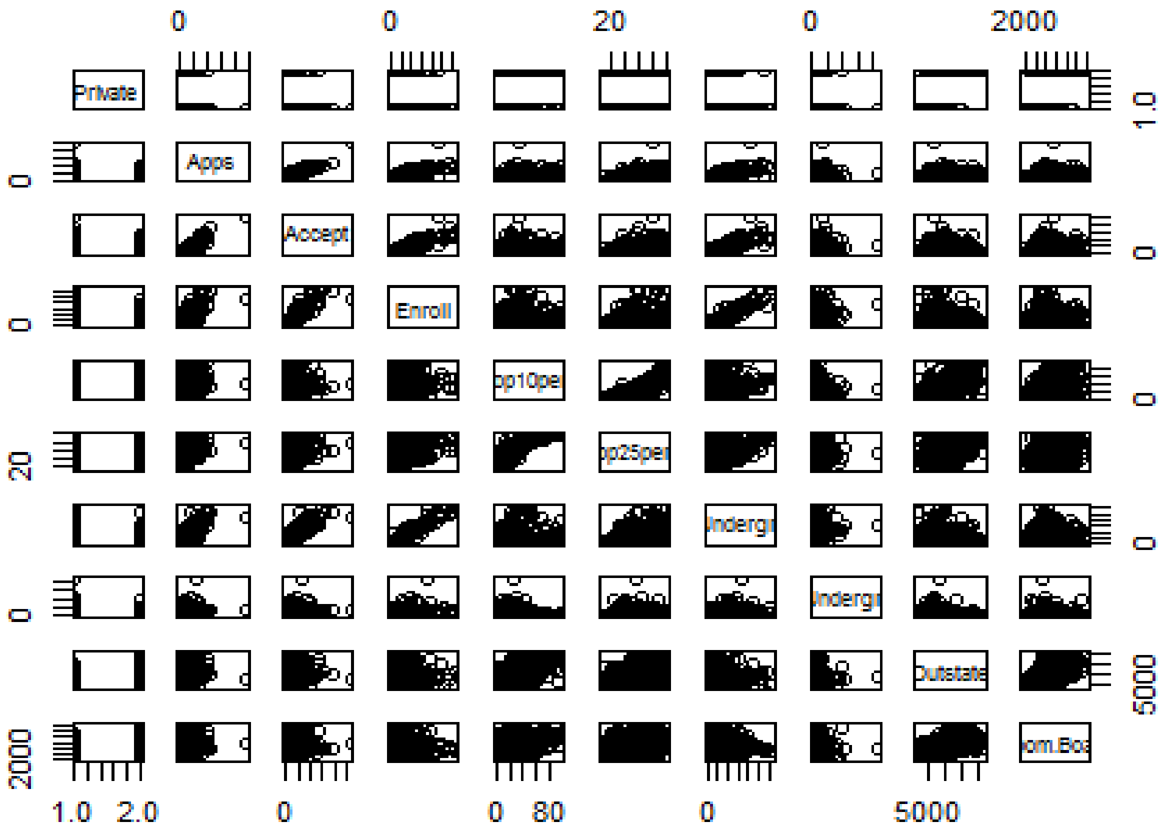
```
## 64 32   43       1  51   2      AN   1.3
## 65 33  152       1  57   2     PKD   2.9
## 66 33   30       1  57   2     PKD   2.9
## 67 34  190       1  44   2      GN   0.7
## 68 34    5       0  45   2      GN   0.7
## 69 35  119       1  22   2   Other   2.2
## 70 35    8       1  22   2   Other   2.2
## 71 36   54       0  42   2   Other   0.7
## 72 36   16       0  42   2   Other   0.7
## 73 37    6       0  52   2     PKD   2.1
## 74 37   78       1  52   2     PKD   2.1
## 75 38   63       1  60   1     PKD   1.2
## 76 38    8       0  60   1     PKD   1.2

skewness(kidney$time)

## [1] 1.968044

kurtosis(kidney$time)

## [1] 6.449392

# Problem 1f
# Are the differences in means significant?
# No, it does not, the mean difference is only 1.2 from mean of x = 3.7 and the
mean of y = 2.5 (stated below in t.test()).
t.test(x,y)

##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 0.33531, df = 17.805, p-value = 0.7413
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.324578  8.724578
## sample estimates:
## mean of x mean of y
##       3.7       2.5

# Problem 1g
x.sort <- sort(x)
x.sort

##  [1] -10  -5   0   1   3   6   7   8  12  15

t.test(x.sort, y, paired = TRUE)

##
##  Paired t-test
##
## data:  x.sort and y
## t = 2.164, df = 9, p-value = 0.05868
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.05440584  2.45440584
## sample estimates:
## mean of the differences
##                      1.2

# Problem 1h
x.logical <- NULL
x.logical = x < 0
x.logical

##  [1] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE

# Problem 1i
x[x.logical == TRUE] <- NA
x <- x[!is.na(x)]
x

## [1]  3 12  6  0  8 15  1  7

####################
# END OF PROBLEM 1 #
####################
```

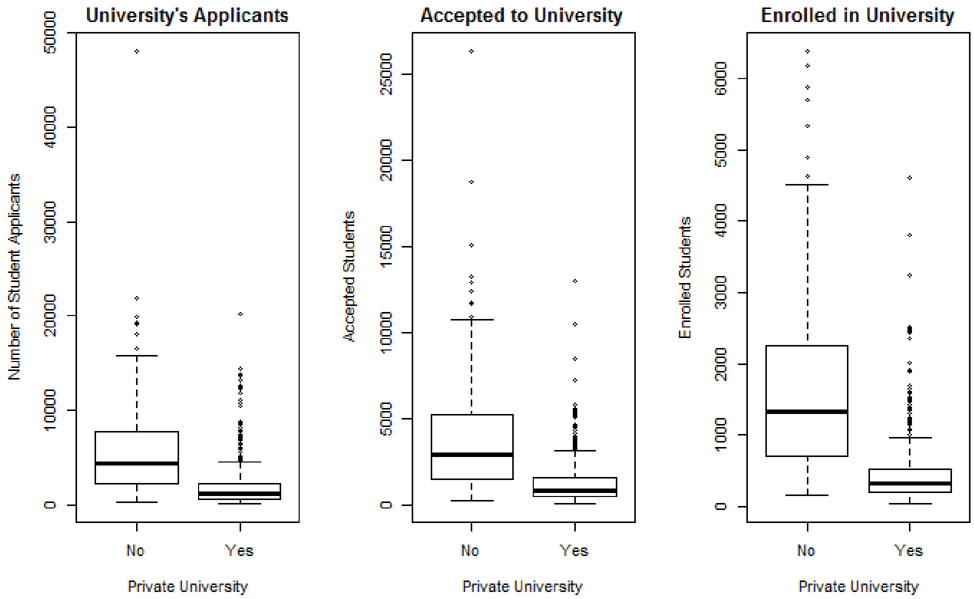
Problem 2

```r
# Problem 2a
college <- read.csv(file="college.csv", header=TRUE, sep=",")

# Problem 2b
rownames (college) <- college [,1]
View (college)

college <- college [,-1]

# Problem 2c-i
summary(college)

##  Private        Apps           Accept          Enroll       Top10perc    
##  No :212   Min.   :   81   Min.   :   72   Min.   :  35   Min.   : 1.00  
##  Yes:565   1st Qu.:  776   1st Qu.:  604   1st Qu.: 242   1st Qu.:15.00  
##             Median : 1558   Median : 1110   Median : 434   Median :23.00  
##             Mean   : 3002   Mean   : 2019   Mean   : 780   Mean   :27.56  
##             3rd Qu.: 3624   3rd Qu.: 2424   3rd Qu.: 902   3rd Qu.:35.00  
##             Max.   :48094   Max.   :26330   Max.   :6392   Max.   :96.00  
##    Top25perc      F.Undergrad     P.Undergrad         Outstate    
##  Min.   :  9.0   Min.   :  139   Min.   :    1.0   Min.   : 2340  
##  1st Qu.: 41.0   1st Qu.:  992   1st Qu.:   95.0   1st Qu.: 7320  
##  Median : 54.0   Median : 1707   Median :  353.0   Median : 9990  
##  Mean   : 55.8   Mean   : 3700   Mean   :  855.3   Mean   :10441  
##  3rd Qu.: 69.0   3rd Qu.: 4005   3rd Qu.:  967.0   3rd Qu.:12925  
##  Max.   :100.0   Max.   :31643   Max.   :21836.0   Max.   :21700  
##    Room.Board       Books           Personal         PhD        
##  Min.   :1780   Min.   :  96.0   Min.   : 250   Min.   :  8.00  
##  1st Qu.:3597   1st Qu.: 470.0   1st Qu.: 850   1st Qu.: 62.00  
##  Median :4200   Median : 500.0   Median :1200   Median : 75.00  
##  Mean   :4358   Mean   : 549.4   Mean   :1341   Mean   : 72.66  
##  3rd Qu.:5050   3rd Qu.: 600.0   3rd Qu.:1700   3rd Qu.: 85.00  
##  Max.   :8124   Max.   :2340.0   Max.   :6800   Max.   :103.00  
##     Terminal       S.F.Ratio      perc.alumni        Expend     
##  Min.   : 24.0   Min.   : 2.50   Min.   : 0.00   Min.   : 3186  
##  1st Qu.: 71.0   1st Qu.:11.50   1st Qu.:13.00   1st Qu.: 6751  
##  Median : 82.0   Median :13.60   Median :21.00   Median : 8377  
##  Mean   : 79.7   Mean   :14.09   Mean   :22.74   Mean   : 9660  
##  3rd Qu.: 92.0   3rd Qu.:16.50   3rd Qu.:31.00   3rd Qu.:10830  
##  Max.   :100.0   Max.   :39.80   Max.   :64.00   Max.   :56233  
##    Grad.Rate     
##  Min.   : 10.00  
##  1st Qu.: 53.00  
##  Median : 65.00  
##  Mean   : 65.46  
##  3rd Qu.: 78.00  
##  Max.   :118.00  
```

```r
# Problem 2c-ii
pairs(college[,1:10])
```

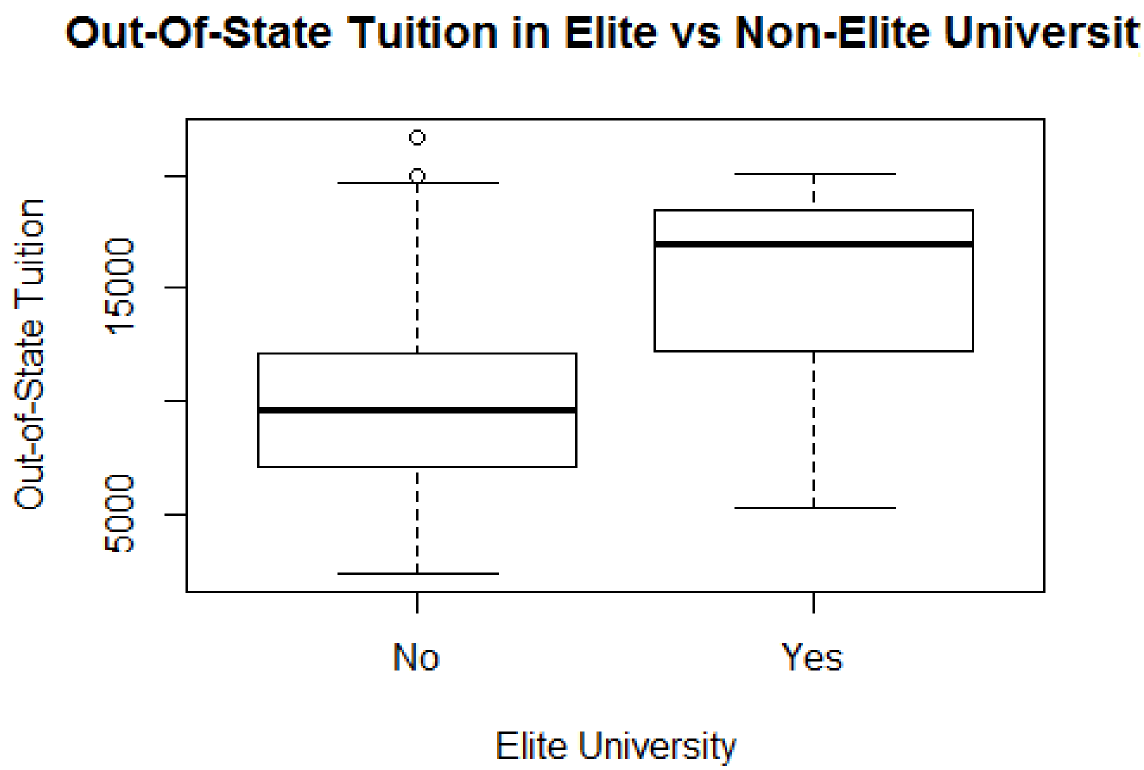


```r
# Problem 2c-iii
par(mfrow=c(1,3))
plot(college$Private, college$Apps, xlab = "Private University", ylab ="Number of Student Applicants", main = "Applicants to University")
plot(college$Private, college$Accept, xlab = "Private University", ylab ="Accepted Students", main = "Accepted to University")
plot(college$Private, college$Enroll, xlab = "Private University", ylab ="Enrolled Students", main = "Enrolled in University")
```

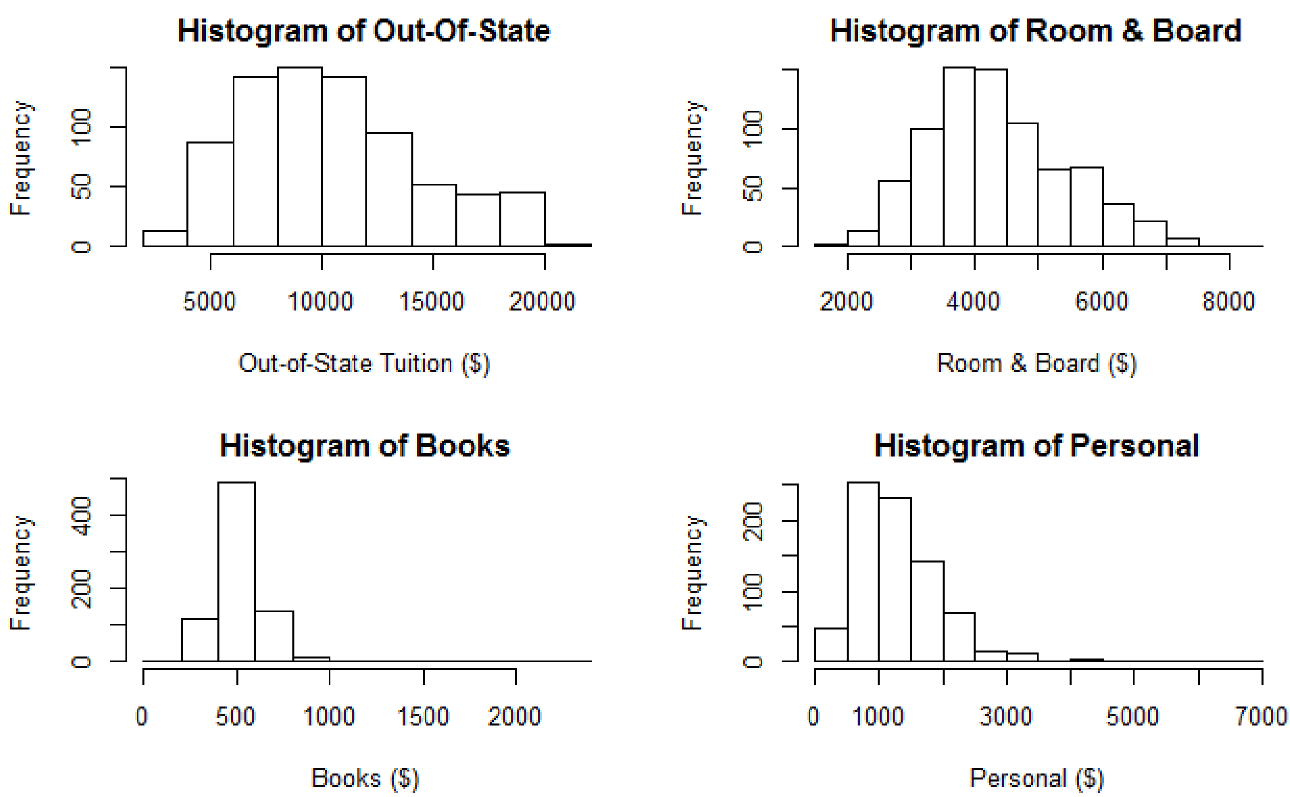



```r
# Problem 2c-iv

Elite <- rep ("No", nrow(college)) # replicates "No" for the total number of rows in college data frame
Elite [college$Top10perc > 50] <- "Yes" # if the Top10perc value is > 50 change No to Yes
Elite <- as.factor (Elite) # create Elite as factor
college <- data.frame(college, Elite) # create new column for Elite in the college data frame

# Problem 2c-v
summary(Elite) # Number of Yes is the total number of Elite Universities

##  No Yes
## 699  78
```

```r
# Problem 2c-vi
par(mfrow = c(1, 1))
plot(college$Elite, college$Outstate, xlab = "Elite University", ylab ="Out-of-State Tuition", main = "Out-Of-State Tuition in Elite vs Non-Elite University")
```



```
# Problem 2c-vii
par(mfrow = c(2, 2))
hist(college$Outstate, xlab = "Out-of-State Tuition ($)", main = "Histogram of Out-Of-State Tuition")
hist(college$Room.Board, xlab = "Room & Board ($)", main = "Histogram of Room & Board")
hist(college$Books, xlab = "Books ($)", main = "Histogram of Books Expenses")
hist(college$Personal, xlab = "Personal ($)", main = "Histogram of Personal Expenses")
```



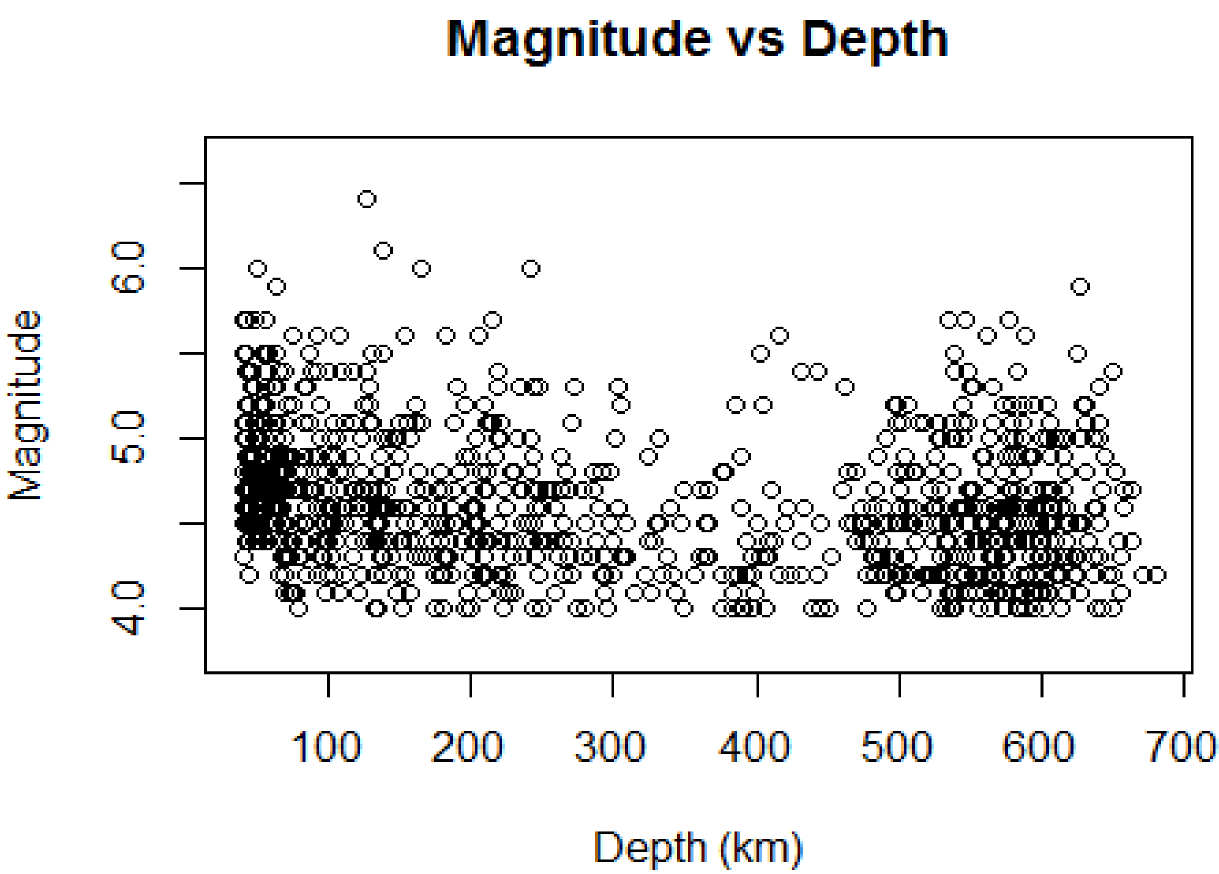

```
####################
# END OF PROBLEM 2 #
####################
```

Problem 3

```r
# library
library(plyr)

# Problem 3a
data ("baseball",package = "plyr")
?baseball

## starting httpd help server ... done

# Problem 3b
baseball$sf [baseball$year < 1954] <- 0
baseball$hbp [is.na(baseball$hbp)] <- 0
baseball <- baseball[!(baseball$ab < 50),]

# Problem 3c
obp <- rep (0, nrow(baseball))
obp <- (baseball$h + baseball$bb + baseball$hbp) / (baseball$ab + baseball$bb +
baseball$hbp + baseball$sf)

# Problem 3d
obp <- as.factor(obp)
baseball <- data.frame(baseball, obp)
baseball <- baseball[order(baseball$obp, decreasing = TRUE),]
baseball[1:5,c(2,1,ncol(baseball))]

##       year        id               obp
## 84983 2004 bondsba01 0.609400324149109
## 82594 2002 bondsba01 0.581699346405229
## 29489 1941 willite01 0.552805280528053
## 7772  1899 mcgrajo01 0.547486033519553
## 19883 1923  ruthba01 0.544540229885057

####################
# END OF PROBLEM 3 #
####################
```

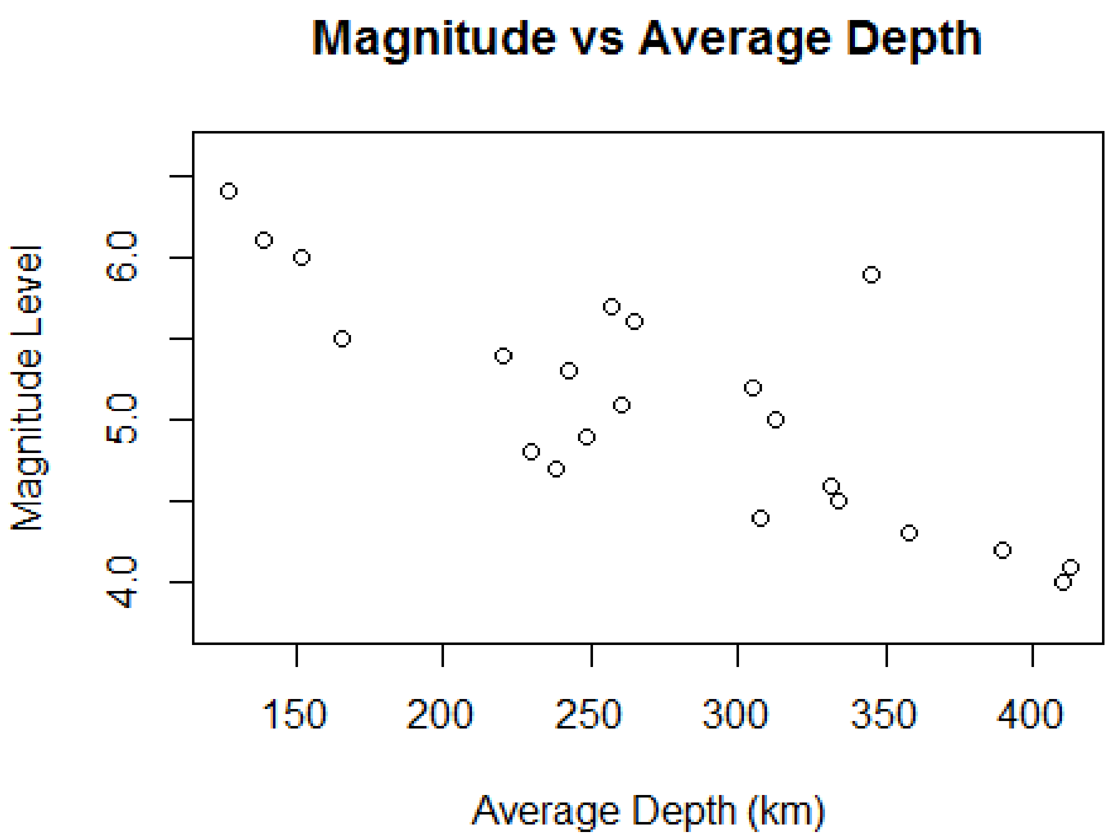
Problem 4

```r
# Problem 4a
# library
library(datasets)

# Problem 4a
data("quakes", package = "datasets")

# Problem 4b
plot(quakes$depth, quakes$mag, xlab = "Depth (km)", ylab = "Magnitude", main =
"Magnitude vs Depth", ylim = range(min(quakes$mag)-0.25, max(quakes$mag)+0.25))
```




```r
# Problem 4c
quakeAvgDepth <- aggregate(quakes,by = list(quakes$mag), FUN = mean)

# Problem 4d
colnames(quakeAvgDepth) <- c("mag.level", "ave.lat", "ave.long", "ave.depth",
"ave.mag", "ave.stations")
```

```r
# Problem 4e
plot(quakeAvgDepth$ave.depth, quakeAvgDepth$mag.level, xlab = "Average Depth (km)",
ylab = "Magnitude Level", main = "Magnitude vs Average Depth", ylim =
range(min(quakeAvgDepth$mag.level)-0.25, max(quakeAvgDepth$mag.level)+0.25))
```




```r
# Problem 4f

# Yes, the plot shows that the earthquakes at shallow depth create a higher
magnitude level earthquakes.

####################
# END OF PROBLEM 4 #
####################
```