Kevin Pan

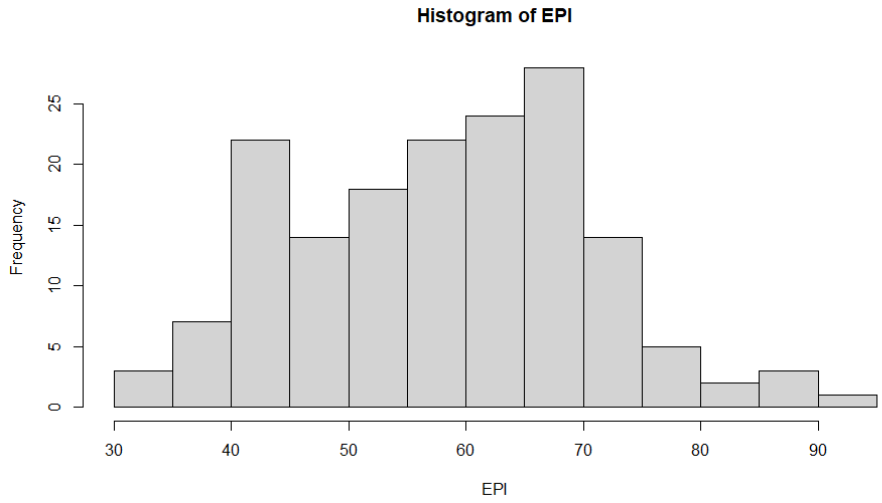
Data Analytics (Level 6000)

Lab2

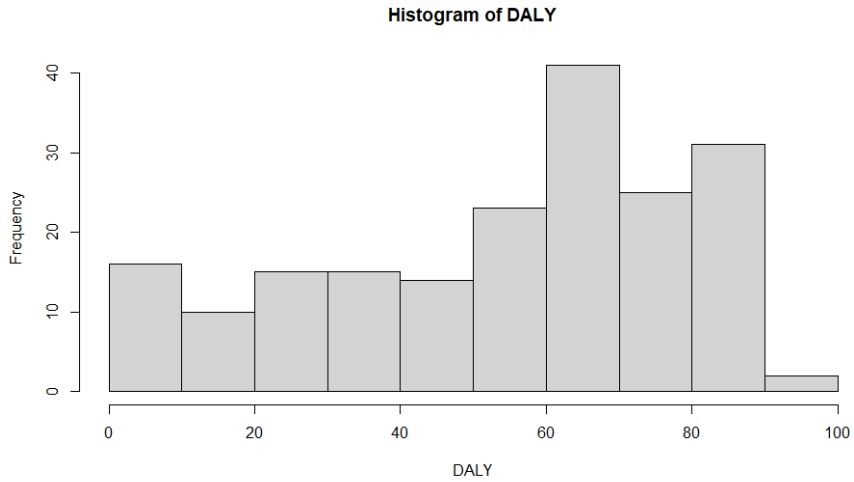
Due: March. 04, 2021

Part 1

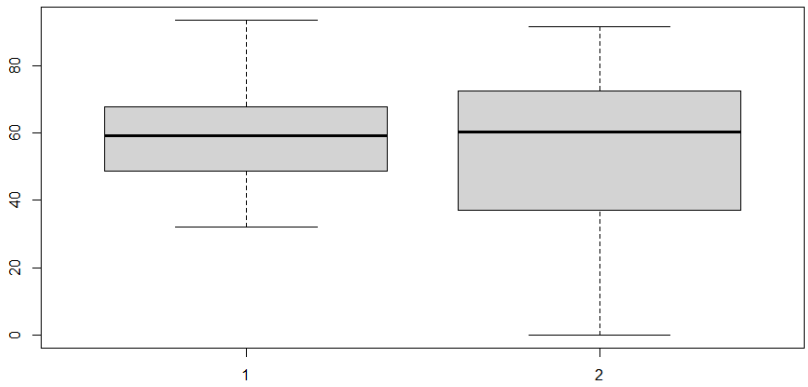
```
> ###################### Lab2 Part1 a) ######################
> EPI_data <- read.csv('EPI_data.csv')
> View(EPI_data)
> attach(EPI_data)
> fix(EPI_data)
> EPI <- EPI_data$EPI[!is.na(EPI)]
> DALY <- EPI_data$DALY[!is.na(DALY)]
> summary(EPI)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  32.10   48.60   59.20   58.37   67.60   93.50
> summary(DALY)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   37.19   60.35   53.94   71.97   91.50
>
> # central tendency
> getmode <- function(v) {
+   uniqv <- unique(v)
+   uniqv[which.max(tabulate(match(v, uniqv)))]
+ }
> print(getmode(EPI))
[1] 44.6
> print(getmode(DALY))
[1] 86.86
>
> # summary statistic
> mean(EPI, na.rm = "TRUE")
[1] 58.37055
> median(EPI, na.rm = "TRUE")
[1] 59.2
> mean(DALY, na.rm = "TRUE")
[1] 53.94313
> median(DALY, na.rm = "TRUE")
[1] 60.35
>
> # histogram
> hist(EPI)
```



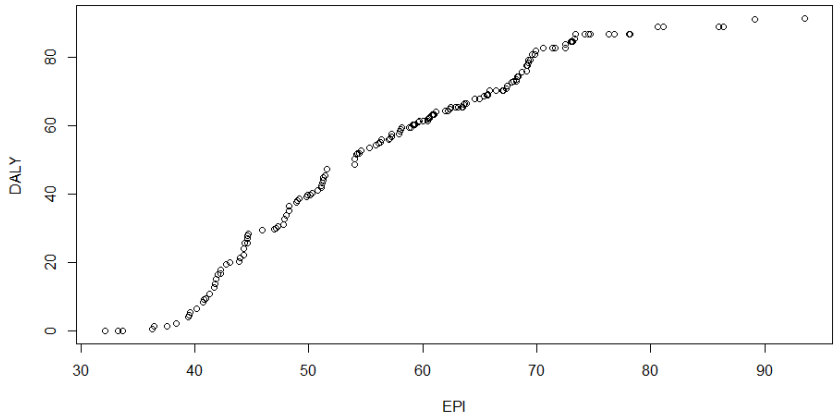

```
> hist(DALY)
```



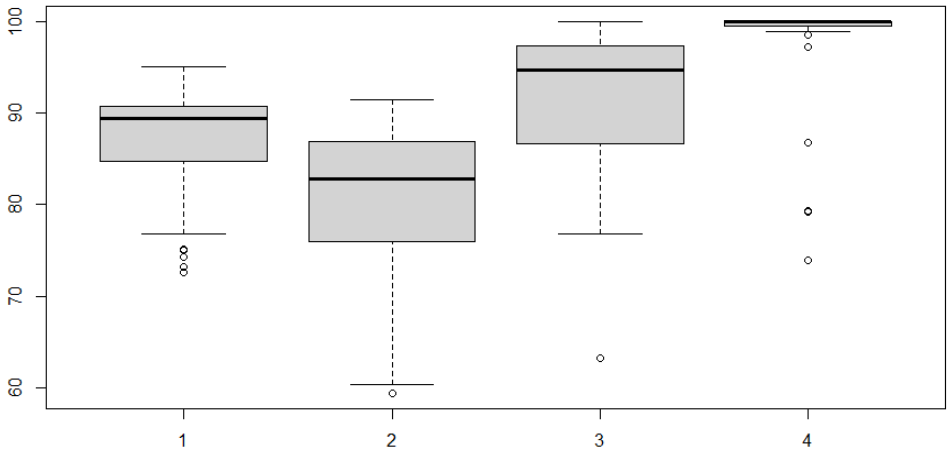

```
>
> # box plot and qqplot
> boxplot(EPI, DALY)
```




```
> qqplot(EPI, DALY)
```



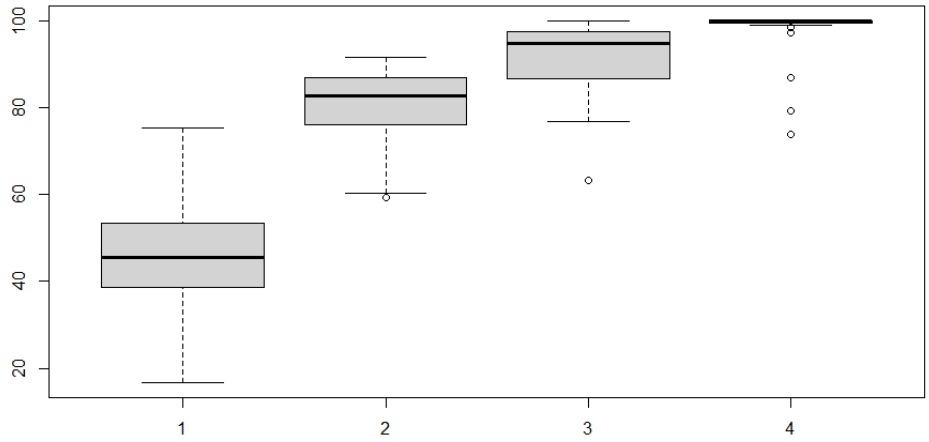

```
> detach(EPI_data)

> ######################## Lab2 Part1 b) ######################
> EPI_data <- subset(EPI_data, EPI_regions=="Europe", na.rm=TRUE)
> DALY <- EPI_data$DALY
> ENVHEALTH <- EPI_data$ENVHEALTH
> AIR_H <- EPI_data$AIR_H
> WATER_H <- EPI_data$WATER_H
> boxplot(ENVHEALTH,DALY,AIR_H,WATER_H)
```




```
> lmENVH <- lm(ENVHEALTH~DALY+AIR_H+WATER_H)
> lmENVH

Call:
lm(formula = ENVHEALTH ~ DALY + AIR_H + WATER_H)

Coefficients:
(Intercept)        DALY       AIR_H     WATER_H
  -0.009899     0.499971     0.250009     0.250121

> summary(lmENVH)

Call:
lm(formula = ENVHEALTH ~ DALY + AIR_H + WATER_H)

Residuals:
      Min        1Q    Median        3Q       Max
-0.0057883 -0.0029477 -0.0000453  0.0018432  0.0044693

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) -9.899e-03  9.735e-03   -1.017    0.318
DALY         5.000e-01  7.827e-05 6387.891   <2e-16 ***
AIR_H        2.500e-01  7.143e-05 3500.063   <2e-16 ***
WATER_H      2.501e-01  1.059e-04 2361.839   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003125 on 28 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:      1,   Adjusted R-squared:      1
F-statistic: 4.438e+07 on 3 and 28 DF,  p-value: < 2.2e-16

> cENVH<-coef(lmENVH)
> cENVH
```

```
 (Intercept)         DALY       AIR_H      WATER_H
-0.009899081  0.499971352  0.250009155  0.250120508
> DALYNEW<-c(seq(1,39,1))
> AIR_HNEW<-c(seq(1,39,1))
> WATER_HNEW<-c(seq(1,39,1))
> NEW <-data.frame(DALYNEW,AIR_HNEW,WATER_HNEW)
> pENV<-na.omit(predict(lmENVH,NEW,interval="prediction"))
> cENV<-na.omit(predict(lmENVH,NEW,interval="confidence"))
> pENV
       fit      lwr      upr
1  90.20556 90.19902 90.21210
2  89.46793 89.46126 89.47461
3  89.05570 89.04917 89.06222
4  73.21068 73.20267 73.21870
5  92.29043 92.28380 92.29706
6  87.64037 87.63343 87.64731
7  86.87579 86.86919 86.88239
8  90.74817 90.74160 90.75474
9  89.82322 89.81665 89.82979
10 88.71300 88.70639 88.71961
11 76.84595 76.83872 76.85317
12 90.74817 90.74160 90.75474
13 90.74817 90.74160 90.75474
15 89.82322 89.81665 89.82979
17 86.60039 86.59372 86.60707
18 82.66363 82.65663 82.67063
20 91.72811 91.72153 91.73470
21 95.09292 95.08617 95.09967
22 90.86799 90.86140 90.87457
24 74.33882 74.33154 74.34610
25 90.74817 90.74160 90.75474
26 75.04972 75.04274 75.05670
27 92.11553 92.10895 92.12212
28 92.11553 92.10895 92.12212
29 89.33293 89.32624 89.33962
30 90.74817 90.74160 90.75474
31 75.19088 75.18353 75.19822
32 87.57307 87.56655 87.57958
33 72.67047 72.66285 72.67809
36 84.49853 84.49171 84.50535
37 84.97571 84.96908 84.98233
38 92.77305 92.76644 92.77967
attr(,"na.action")
14 16 19 23 34 35 39
14 16 19 23 34 35 39
attr(,"class")
[1] "omit"
> cENV
       fit      lwr      upr
1  90.20556 90.20423 90.20689
2  89.46793 89.46605 89.46982
3  89.05570 89.05442 89.05697
4  73.21068 73.20587 73.21550
5  92.29043 92.28870 92.29216
```

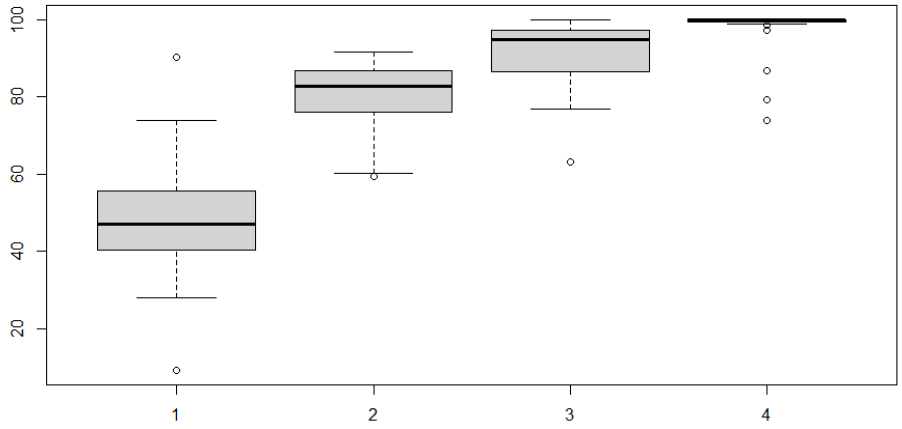
```
6  87.64037 87.63769 87.64305
7  86.87579 86.87419 86.87738
8  90.74817 90.74669 90.74965
9  89.82322 89.82175 89.82470
10 88.71300 88.71136 88.71465
11 76.84595 76.84260 76.84930
12 90.74817 90.74669 90.74965
13 90.74817 90.74669 90.74965
15 89.82322 89.82175 89.82470
17 86.60039 86.59851 86.60228
18 82.66363 82.66080 82.66647
20 91.72811 91.72657 91.72966
21 95.09292 95.09078 95.09507
22 90.86799 90.86646 90.86951
24 74.33882 74.33536 74.34228
25 90.74817 90.74669 90.74965
26 75.04972 75.04694 75.05251
27 92.11553 92.11399 92.11707
28 92.11553 92.11399 92.11707
29 89.33293 89.33099 89.33487
30 90.74817 90.74669 90.74965
31 75.19088 75.18727 75.19448
32 87.57307 87.57187 87.57426
33 72.67047 72.66634 72.67460
36 84.49853 84.49618 84.50088
37 84.97571 84.97401 84.97740
38 92.77305 92.77137 92.77474
attr(,"na.action")
14 16 19 23 34 35 39
14 16 19 23 34 35 39
attr(,"class")
[1] "omit"


> # repeat for AIR_E
> AIR_E <- EPI_data$AIR_E
> boxplot(AIR_E,DALY,AIR_H,WATER_H)
```




```
> lmENVH <- lm(AIR_E~DALY+AIR_H+WATER_H)
> lmENVH

Call:
lm(formula = ENVHEALTH ~ DALY + AIR_H + WATER_H)

Coefficients:
(Intercept)        DALY       AIR_H     WATER_H
  -0.009899     0.499971     0.250009     0.250121

> summary(lmENVH)

Call:
lm(formula = ENVHEALTH ~ DALY + AIR_H + WATER_H)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0057883 -0.0029477 -0.0000453  0.0018432  0.0044693

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) -9.899e-03  9.735e-03   -1.017    0.318
DALY         5.000e-01  7.827e-05 6387.891   <2e-16 ***
AIR_H        2.500e-01  7.143e-05 3500.063   <2e-16 ***
WATER_H      2.501e-01  1.059e-04 2361.839   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003125 on 28 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:      1,   Adjusted R-squared:      1
F-statistic: 4.438e+07 on 3 and 28 DF,  p-value: < 2.2e-16

> cENVH<-coef(lmENVH)
> cENVH
```

```
(Intercept)        DALY       AIR_H     WATER_H
48.68530192 -0.60512629  0.43338822  0.05382547
> DALYNEW<-c(seq(1,39,1))
> AIR_HNEW<-c(seq(1,39,1))
> WATER_HNEW<-c(seq(1,39,1))
> NEW <-data.frame(DALYNEW,AIR_HNEW,WATER_HNEW)
> pENV<-na.omit(predict(lmENVH,NEW,interval="prediction"))
> cENV<-na.omit(predict(lmENVH,NEW,interval="confidence"))
> pENV
       fit      lwr      upr
1  42.33097 20.62056 64.04138
2  37.97620 15.81567 60.13673
3  45.94533 24.26973 67.62093
4  41.77178 15.17149 68.37208
5  39.57209 17.55203 61.59215
6  34.80813 11.76511 57.85115
7  49.73826 27.83396 71.64256
8  46.15635 24.34134 67.97136
9  47.27584 25.46424 69.08744
10 39.74364 17.79734 61.68994
11 57.56473 33.57837 81.55108
12 46.15635 24.34134 67.97136
13 46.15635 24.34134 67.97136
15 47.27584 25.46424 69.08744
17 39.39155 17.23049 61.55261
18 55.94124 32.69907 79.18342
20 44.97030 23.10325 66.83736
21 40.89780 18.48031 63.31530
22 40.40317 18.54936 62.25698
24 58.63276 34.47462 82.79090
25 46.15635 24.34134 67.97136
26 58.40045 35.22307 81.57784
27 42.56578 20.70190 64.42966
28 42.56578 20.70190 64.42966
29 37.74217 15.53467 59.94967
30 46.15635 24.34134 67.97136
31 45.46565 21.07364 69.85766
32 46.50487 24.88283 68.12691
33 51.81152 26.51378 77.10926
36 53.72043 31.07961 76.36125
37 43.90635 21.91935 65.89335
38 43.70559 21.72854 65.68264
attr(,"na.action")
14 16 19 23 34 35 39
14 16 19 23 34 35 39
attr(,"class")
[1] "omit"
> cENV
       fit      lwr      upr
1  42.33097 37.90968 46.75227
2  37.97620 31.70763 44.24476
3  45.94533 41.69827 50.19240
4  41.77178 25.77861 57.76496
5  39.57209 33.81978 45.32440
```

```
6  34.80813 25.90943 43.70683
7  49.73826 44.44635 55.03016
8  46.15635 41.24711 51.06560
9  47.27584 42.38177 52.16990
10 39.74364 34.28050 45.20678
11 57.56473 46.44937 68.68009
12 46.15635 41.24711 51.06560
13 46.15635 41.24711 51.06560
15 47.27584 42.38177 52.16990
17 39.39155 33.12111 45.66199
18 55.94124 46.53887 65.34362
20 44.97030 39.83473 50.10588
21 40.89780 33.77389 48.02172
22 40.40317 35.32431 45.48204
24 58.63276 47.15140 70.11412
25 46.15635 41.24711 51.06560
26 58.40045 49.15941 67.64150
27 42.56578 37.44374 47.68782
28 42.56578 37.44374 47.68782
29 37.74217 31.30951 44.17483
30 46.15635 41.24711 51.06560
31 45.46565 33.50003 57.43127
32 46.50487 42.54021 50.46953
33 51.81152 38.09359 65.52945
36 53.72043 45.92219 61.51867
37 43.90635 38.28193 49.53076
38 43.70559 38.12020 49.29098
attr(,"na.action")
14 16 19 23 34 35 39
14 16 19 23 34 35 39
attr(,"class")
[1] "omit


> # repeat for CLIMATE
> CLIMATE <- EPI_data$CLIMATE
> boxplot(CLIMATE,DALY,AIR_H,WATER_H)
```




```
> lmENVH <- lm(CLIMATE~DALY+AIR_H+WATER_H)
> lmENVH

Call:
lm(formula = ENVHEALTH ~ DALY + AIR_H + WATER_H)

Coefficients:
(Intercept)        DALY       AIR_H      WATER_H
  -0.009899     0.499971     0.250009     0.250121

> summary(lmENVH)

Call:
lm(formula = ENVHEALTH ~ DALY + AIR_H + WATER_H)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0057883 -0.0029477 -0.0000453  0.0018432  0.0044693

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) -9.899e-03  9.735e-03   -1.017    0.318
DALY         5.000e-01  7.827e-05 6387.891   <2e-16 ***
AIR_H        2.500e-01  7.143e-05 3500.063   <2e-16 ***
WATER_H      2.501e-01  1.059e-04 2361.839   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003125 on 28 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:      1,   Adjusted R-squared:      1
F-statistic: 4.438e+07 on 3 and 28 DF,  p-value: < 2.2e-16

> cENVH<-coef(lmENVH)
> cENVH
(Intercept)        DALY       AIR_H      WATER_H
```

```
  4.4595074   0.4639057   0.7930328  -0.6677696
> DALYNEW<-c(seq(1,39,1))
> AIR_HNEW<-c(seq(1,39,1))
> WATER_HNEW<-c(seq(1,39,1))
> NEW <-data.frame(DALYNEW,AIR_HNEW,WATER_HNEW)
> pENV<-na.omit(predict(lmENVH,NEW,interval="prediction"))
> cENV<-na.omit(predict(lmENVH,NEW,interval="confidence"))
> pENV
       fit       lwr      upr
1  49.39587  20.917238 77.87450
2  44.71111  15.639420 73.78279
3  50.02334  21.611154 78.43554
4  19.18386 -15.671018 54.03873
5  51.15081  22.233133 80.06848
6  38.91404   8.697346 69.13073
7  49.53695  20.847717 78.22618
8  53.31618  24.705375 81.92699
9  52.45796  23.863894 81.05202
10 44.66146  15.882596 73.44033
11 39.57902   8.165744 70.99230
12 53.31618  24.705375 81.92699
13 53.31618  24.705375 81.92699
15 52.45796  23.863894 81.05202
17 41.79587  12.750040 70.84169
18 45.81483  15.391364 76.23829
20 54.22544  25.531985 82.91889
21 57.34752  27.873506 86.82154
22 49.15209  20.472668 77.83151
24 56.72637  25.102671 88.35006
25 53.31618  24.705375 81.92699
26 49.07739  18.733324 79.42146
27 53.10932  24.407802 81.81084
28 53.10932  24.407802 81.81084
29 44.28287  15.150523 73.41521
30 53.31618  24.705375 81.92699
31 48.31233  16.380319 80.24434
32 49.37000  21.037108 77.70289
33 54.70738  21.593362 87.82141
36 47.51736  17.878857 77.15586
37 40.91884  12.122364 69.71531
38 55.19500  26.340503 84.04950
attr(,"na.action")
14 16 19 23 34 35 39
14 16 19 23 34 35 39
attr(,"class")
[1] "omit"
> cENV
       fit       lwr      upr
1  49.39587 43.317855 55.47388
2  44.71111 36.280634 53.14158
3  50.02334 44.264657 55.78203
4  19.18386 -1.810719 40.17843
5  51.15081 43.267799 59.03382
6  38.91404 27.125980 50.70209
```

```
7  49.53695 42.538243 56.53565
8  53.31618 46.646235 59.98613
9  52.45796 45.860209 59.05570
10 44.66146 37.303950 52.01897
11 39.57902 24.994648 54.16340
12 53.31618 46.646235 59.98613
13 53.31618 46.646235 59.98613
15 52.45796 45.860209 59.05570
17 41.79587 33.455008 50.13672
18 45.81483 33.506418 58.12323
20 54.22544 47.209448 61.24142
21 57.34752 47.619853 67.07519
22 49.15209 42.193705 56.11047
24 56.72637 41.694131 71.75860
25 53.31618 46.646235 59.98613
26 49.07739 36.966554 61.18823
27 53.10932 46.060406 60.15824
28 53.10932 46.060406 60.15824
29 44.28287 35.645541 52.92020
30 53.31618 46.646235 59.98613
31 48.31233 32.641873 63.98279
32 49.37000 44.016239 54.72376
33 54.70738 36.750537 72.66423
36 47.51736 37.302139 57.73258
37 40.91884 33.492753 48.34492
38 55.19500 47.547000 62.84300
attr(,"na.action")
14 16 19 23 34 35 39
14 16 19 23 34 35 39
attr(,"class")
[1] "omit"
```

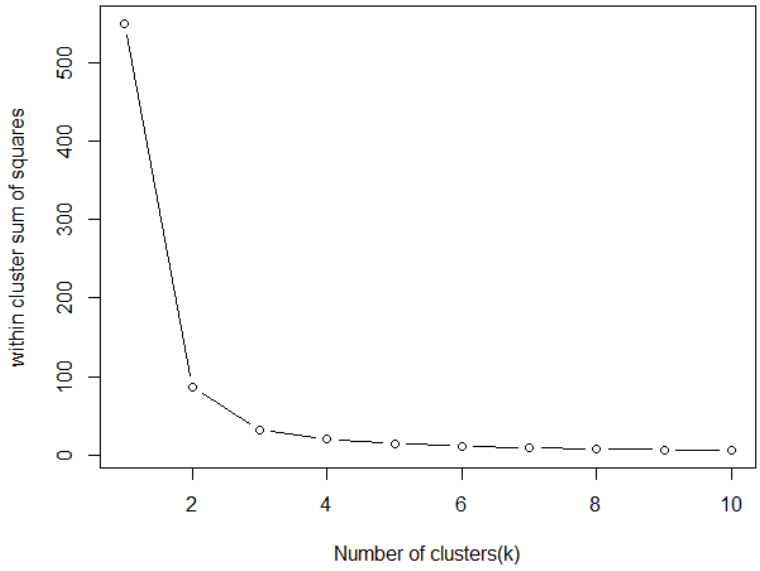
Part 2

```
> ###################### exercise1 ####################
> multiReg <- read.csv("dataset_multipleRegression.csv")
> attach(multiReg)
> View(multiReg)
> # using the linear model
> linearModel <- lm(ROLL ~ UNEM + HGRAD)
> UNEM = c(7)
> HGRAD = c(90000)
> INC = c(25000)
> UNEM_HGRAD <- data.frame(UNEM,HGRAD)
> predictedRoll1 <- predict(linearModel, UNEM_HGRAD, interval='prediction')
> predictedRoll1
       fit      lwr      upr
1 81437.04 68082.31 94791.78
> UNEM_HGRAD_INC <- data.frame(UNEM, HGRAD, INC)
> predictedRoll2 <- predict(linearModel, UNEM_HGRAD_INC, interval='prediction')
```

```
> predictedRoll2
      fit      lwr      upr
1 81437.04 68082.31 94791.78
> detach(multiReg)


> ##################### exercise2 ####################
> abalone <- read.csv("abalone.csv")
> attach(abalone)
> colnames(abalone) <-
c("sex","length",'diameter','height','whole_weight','shucked_wieght','viscera_wieght','shell_weight', 'rings' )
> summary(abalone)
    sex              length         diameter         height        whole_weight    shucked_wieght
 Length:4177        Min.   :0.075   Min.   :0.0550   Min.   :0.0000   Min.   :0.0020   Min.   :0.0010
 Class :character   1st Qu.:0.450   1st Qu.:0.3500   1st Qu.:0.1150   1st Qu.:0.4415   1st Qu.:0.1860
 Mode  :character   Median :0.545   Median :0.4250   Median :0.1400   Median :0.7995   Median :0.3360
                    Mean   :0.524   Mean   :0.4079   Mean   :0.1395   Mean   :0.8287   Mean   :0.3594
                    3rd Qu.:0.615   3rd Qu.:0.4800   3rd Qu.:0.1650   3rd Qu.:1.1530   3rd Qu.:0.5020
                    Max.   :0.815   Max.   :0.6500   Max.   :1.1300   Max.   :2.8255   Max.   :1.4880
 viscera_wieght    shell_weight       rings
 Min.   :0.0005   Min.   :0.0015   Min.   : 1.000
 1st Qu.:0.0935   1st Qu.:0.1300   1st Qu.: 8.000
 Median :0.1710   Median :0.2340   Median : 9.000
 Mean   :0.1806   Mean   :0.2388   Mean   : 9.934
 3rd Qu.:0.2530   3rd Qu.:0.3290   3rd Qu.:11.000
 Max.   :0.7600   Max.   :1.0050   Max.   :29.000
> str(abalone)
'data.frame':        4177 obs. of  9 variables:
 $ sex          : chr  "M" "M" "F" "M" ...
 $ length       : num  0.455 0.35 0.53 0.44 0.33 0.425 0.53 0.545 0.475 0.55 ...
 $ diameter     : num  0.365 0.265 0.42 0.365 0.255 0.3 0.415 0.425 0.37 0.44 ...
 $ height       : num  0.095 0.09 0.135 0.125 0.08 0.095 0.15 0.125 0.125 0.15 ...
 $ whole_weight : num  0.514 0.226 0.677 0.516 0.205 ...
 $ shucked_wieght: num  0.2245 0.0995 0.2565 0.2155 0.0895 ...
 $ viscera_wieght: num  0.101 0.0485 0.1415 0.114 0.0395 ...
 $ shell_weight : num  0.15 0.07 0.21 0.155 0.055 0.12 0.33 0.26 0.165 0.32 ...
 $ rings        : int  15 7 9 10 7 8 20 16 9 19 ...
> summary(abalone$rings)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   8.000   9.000   9.934  11.000  29.000
> abalone$rings <- as.numeric(abalone$rings)
> abalone$rings <- cut(abalone$rings, br=c(-1,8,11,35), labels = c("young", 'adult', 'old'))
> abalone$rings <- as.factor(abalone$rings)
> summary(abalone$rings)
young adult   old
 1407  1810   960
> aba <- abalone
> aba$sex <- NULL
> normalize <- function(x) {
+   return ((x - min(x)) / (max(x) - min(x)))
+ }
> aba[1:7] <- as.data.frame(lapply(aba[1:7], normalize))
> summary(aba$shucked_wieght)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
 0.0000  0.1244  0.2253  0.2410  0.3369  1.0000
> ind <- sample(2, nrow(aba), replace=TRUE, prob=c(0.7, 0.3))
> KNNtrain <- aba[ind==1,]
> KNNtest <- aba[ind==2,]
> sqrt(2918)
[1] 54.01852
> library(class)
> KNNpred <- knn(train = KNNtrain[1:7], test = KNNtest[1:7], cl = KNNtrain$rings, k = 55)
> KNNpred
  [1] adult old   young adult young young young adult adult adult adult young adult young young adult young
 [18] adult old   young adult adult adult old   adult old   adult young young young young young young young
 [35] young young adult adult adult young young old   adult adult adult old   old   adult young young young
 [52] adult adult adult adult young adult adult old   adult young old   young young old   young adult adult
 [69] adult old   young young young young young old   young old   adult old   old   adult old   young young
 [86] young young old   old   young young old   young old   young old   adult adult adult adult adult adult
[103] old   old   old   old   old   adult adult young young young adult young young adult old   adult adult
[120] old   adult young old   young adult adult adult young young young old   old   old   adult adult
[137] adult adult adult adult adult adult adult old   adult old   adult old   adult old   young young adult
[154] adult young young young young young adult adult young adult old   adult old   adult old   old   young
[171] young young old   adult young young young adult adult adult adult young young adult adult adult old
[188] old   old   old   adult young adult young old   adult young adult adult young young young young young
[205] young young old   young young old   adult adult young old   old   adult adult old   old   adult old
[222] adult old   old   adult adult adult adult adult old   young adult young young young young young young
[239] young adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult
[256] adult adult old   young young young young young young young young young young young young young adult young
[273] young young young young young adult adult young adult adult adult adult adult adult adult adult adult
[290] adult adult adult adult adult adult adult adult adult adult adult adult old   adult adult adult adult old
[307] adult young young young young young young young young young young young young adult young young young adult
[324] adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult
[341] adult adult adult adult adult old   adult adult old   adult old   young young young young young young
[358] young young young young young adult adult young adult adult adult adult adult adult adult adult adult
[375] adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult
[392] adult adult adult adult adult adult adult adult old   young young young young young young young young
[409] young adult young adult adult adult adult adult adult adult adult adult adult adult adult adult adult
[426] adult adult young young young young young young young young young young young young young adult young adult
[443] old   adult adult adult adult adult adult adult adult adult adult adult adult adult adult old   adult
[460] adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult
[477] adult adult adult adult adult adult adult adult old   adult adult adult adult adult adult adult young
[494] young young adult old   adult adult young adult adult adult adult adult adult adult adult adult old
[511] adult adult adult young young young young young young young young adult young adult adult young adult
[528] adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult
[545] adult adult adult adult adult adult old   adult adult old   adult adult adult adult adult young young
[562] young young young young young young young adult young young adult young adult adult adult adult adult
[579] adult adult young young young young young adult young adult young adult adult adult adult old   young
[596] adult old   young young young young adult young young old   adult young adult young old   young young
[613] young young adult old   adult old   adult young adult young young old   adult young adult adult adult
[630] young young old   old   young old   old   adult adult adult old   old   old   young adult young old
[647] adult adult old   old   old   old   old   adult adult old   young young adult adult young adult young
[664] adult adult young adult adult young young adult adult adult adult adult young young old   adult adult
[681] old   adult old   adult adult old   young young young young young young adult adult young old   adult
[698] young old   adult adult young young young adult old   old   adult old   old   old   adult young adult
[715] young young adult old   young young young young young young young young young adult adult adult adult
[732] adult adult adult adult young young young young young young young adult young adult adult adult adult
[749] adult adult adult adult adult adult adult adult adult adult young young young young young young young
```

```
 [766] young adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult adult
 [783] adult adult young young young young young young young young young young young young young young
young
 [800] young adult young adult adult young adult adult adult adult adult adult adult adult adult adult
 [817] adult adult adult adult adult adult young young young young young young young adult adult adult adult
 [834] adult adult adult young adult young adult adult adult adult adult adult adult adult adult adult
 [851] adult adult adult adult adult adult adult adult old   adult adult adult adult adult adult young
 [868] adult adult adult adult adult young young young young adult young adult adult adult adult adult old
 [885] adult adult adult old   adult young adult adult young young young young adult adult adult adult
 [902] adult adult adult adult adult adult adult old   young old   adult old   old   young old   old   young
 [919] young adult adult young young adult old   young young young adult old   young young young old   young
 [936] adult adult adult adult adult adult young young young adult adult adult adult adult old   old
 [953] adult adult adult adult adult young young adult adult young young adult young adult adult young young
 [970] adult adult young young adult young old   young old   adult young young old   adult adult adult adult
 [987] young young young young young young young adult adult adult adult adult young young
 [ reached getOption("max.print") -- omitted 225 entries ]
Levels: young adult old
> table(KNNpred)
KNNpred
young adult   old
  415   674   136
> detach(abalone)



> ################### exercise3 ###################
> library(ggplot2)
> View(iris)
> sapply(iris[,-5],var)
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
   0.6856935    0.1899794    3.1162779    0.5810063
> k.max<-10
> wss <- sapply(1:k.max,function(k){kmeans(iris[,3:4],k,nstart=20,iter.max=1000)$tot.withinss})
> wss
 [1] 550.895333  86.390220  31.371359  19.465989  13.916909  11.025145   9.236596   7.924764   6.687056
[10]   5.528149
> plot(1:k.max,wss,type="b",xlab = "Number of clusters(k)",ylab="within cluster sum of squares")
```




```
> icluster<-kmeans(iris[,3:4],3,nstart=20)
> table(icluster$cluster,iris$Species)
```

	setosa	versicolor	virginica
1	50	0	0
2	0	2	46
3	0	48	4