Untitled.R

riteshdabas

2021-11-06

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
data <- read_csv("Life Expectancy Data.csv")
```

```
## Rows: 2938 Columns: 22
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): Country, Status
## dbl (20): Year, Life expectancy, Adult Mortality, infant deaths, Alcohol, pe...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(data)
```

```
## # A tibble: 6 x 22
##   Country  Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol
##   <chr>   <dbl> <chr>             <dbl>            <dbl>           <dbl>   <dbl>
## 1 Afghan~  2015 Devel~             65                263              62    0.01
## 2 Afghan~  2014 Devel~             59.9              271              64    0.01
## 3 Afghan~  2013 Devel~             59.9              268              66    0.01
## 4 Afghan~  2012 Devel~             59.5              272              69    0.01
## 5 Afghan~  2011 Devel~             59.2              275              71    0.01
## 6 Afghan~  2010 Devel~             58.8              279              74    0.01
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,
## #   Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,
## #   Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,
## #   Population <dbl>, thinness  1-19 years <dbl>, thinness 5-9 years <dbl>,
## #   Income composition of resources <dbl>, Schooling <dbl>
```

```
summary(data)
```

```
##    Country               Year         Status          Life expectancy
##  Length:2938        Min.   :2000   Length:2938        Min.   :36.30
```

```
##  Class :character   1st Qu.:2004   Class :character   1st Qu.:63.10
##  Mode  :character   Median :2008   Mode  :character   Median :72.10
##                     Mean   :2008                      Mean   :69.22
##                     3rd Qu.:2012                      3rd Qu.:75.70
##                     Max.   :2015                      Max.   :89.00
##                                                       NA's   :10
##  Adult Mortality infant deaths       Alcohol        percentage expenditure
##  Min.   :  1.0   Min.   :   0.0   Min.   : 0.0100   Min.   :    0.000
##  1st Qu.: 74.0   1st Qu.:   0.0   1st Qu.: 0.8775   1st Qu.:    4.685
##  Median :144.0   Median :   3.0   Median : 3.7550   Median :   64.913
##  Mean   :164.8   Mean   :  30.3   Mean   : 4.6029   Mean   :  738.251
##  3rd Qu.:228.0   3rd Qu.:  22.0   3rd Qu.: 7.7025   3rd Qu.:  441.534
##  Max.   :723.0   Max.   :1800.0   Max.   :17.8700   Max.   :19479.912
##  NA's   :10                       NA's   :194
##   Hepatitis B       Measles               BMI        under-five deaths
##  Min.   : 1.00   Min.   :     0.0   Min.   : 1.00   Min.   :   0.00
##  1st Qu.:77.00   1st Qu.:     0.0   1st Qu.:19.30   1st Qu.:   0.00
##  Median :92.00   Median :    17.0   Median :43.50   Median :   4.00
##  Mean   :80.94   Mean   :  2419.6   Mean   :38.32   Mean   :  42.04
##  3rd Qu.:97.00   3rd Qu.:   360.2   3rd Qu.:56.20   3rd Qu.:  28.00
##  Max.   :99.00   Max.   :212183.0   Max.   :87.30   Max.   :2500.00
##  NA's   :553                        NA's   :34
##      Polio       Total expenditure   Diphtheria       HIV/AIDS
##  Min.   : 3.00   Min.   : 0.370    Min.   : 2.00   Min.   : 0.100
##  1st Qu.:78.00   1st Qu.: 4.260    1st Qu.:78.00   1st Qu.: 0.100
##  Median :93.00   Median : 5.755    Median :93.00   Median : 0.100
##  Mean   :82.55   Mean   : 5.938    Mean   :82.32   Mean   : 1.742
##  3rd Qu.:97.00   3rd Qu.: 7.492    3rd Qu.:97.00   3rd Qu.: 0.800
##  Max.   :99.00   Max.   :17.600    Max.   :99.00   Max.   :50.600
##  NA's   :19      NA's   :226       NA's   :19
##       GDP              Population        thinness  1-19 years
##  Min.   :     1.68   Min.   :3.400e+01   Min.   : 0.10
##  1st Qu.:   463.94   1st Qu.:1.958e+05   1st Qu.: 1.60
##  Median :  1766.95   Median :1.387e+06   Median : 3.30
##  Mean   :  7483.16   Mean   :1.275e+07   Mean   : 4.84
##  3rd Qu.:  5910.81   3rd Qu.:7.420e+06   3rd Qu.: 7.20
##  Max.   :119172.74   Max.   :1.294e+09   Max.   :27.70
##  NA's   :448         NA's   :652         NA's   :34
##  thinness 5-9 years Income composition of resources   Schooling
##  Min.   : 0.10      Min.   :0.0000                  Min.   : 0.00
##  1st Qu.: 1.50      1st Qu.:0.4930                  1st Qu.:10.10
##  Median : 3.30      Median :0.6770                  Median :12.30
##  Mean   : 4.87      Mean   :0.6276                  Mean   :11.99
##  3rd Qu.: 7.20      3rd Qu.:0.7790                  3rd Qu.:14.30
##  Max.   :28.60      Max.   :0.9480                  Max.   :20.70
##  NA's   :34         NA's   :167                     NA's   :163
```

```r
# data format : factor
data$Country <- as.factor(data$Country)
data$Status <-  as.factor(data$Status)
View(data)

# Counting missing values in each column
missing <-sapply(data, function(x) sum(length(which(is.na(x)))))
```

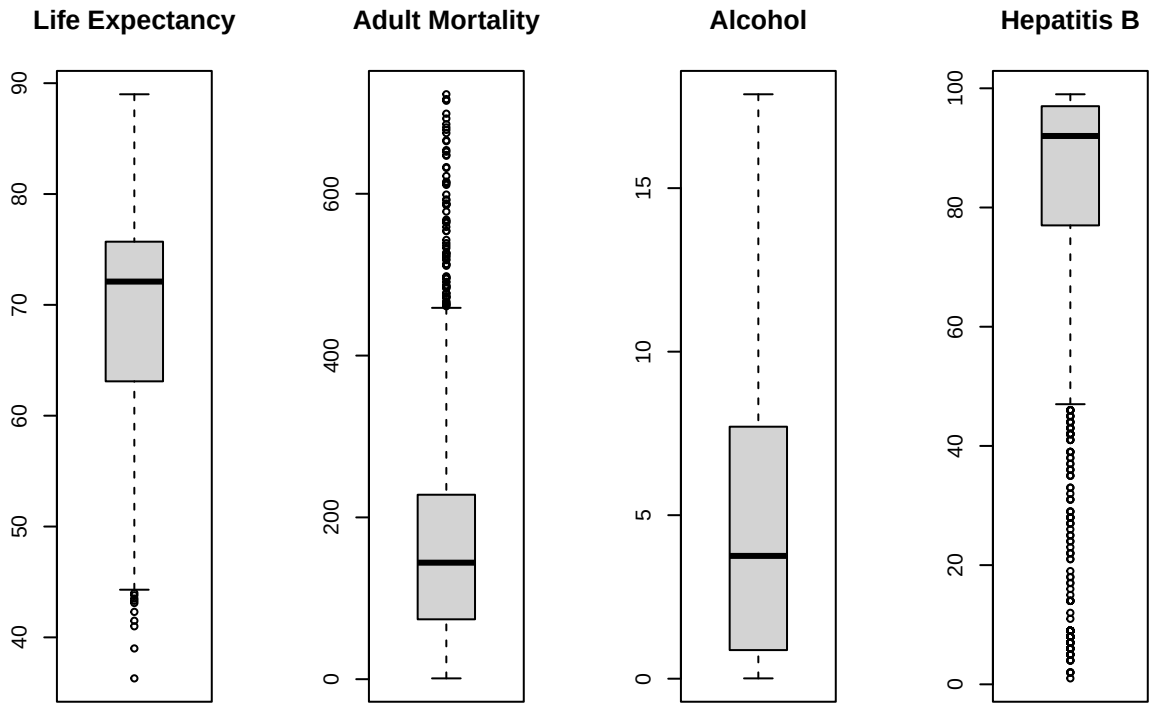
	missing
Country	0
Year	0
Status	0
Life expectancy	10
Adult Mortality	10
infant deaths	0
Alcohol	194
percentage expenditure	0
Hepatitis B	553
Measles	0
BMI	34
under-five deaths	0
Polio	19
Total expenditure	226
Diphtheria	19
HIV/AIDS	0
GDP	448
Population	652
thinness 1-19 years	34
thinness 5-9 years	34
Income composition of resources	167
Schooling	163

```
library('kableExtra')
```

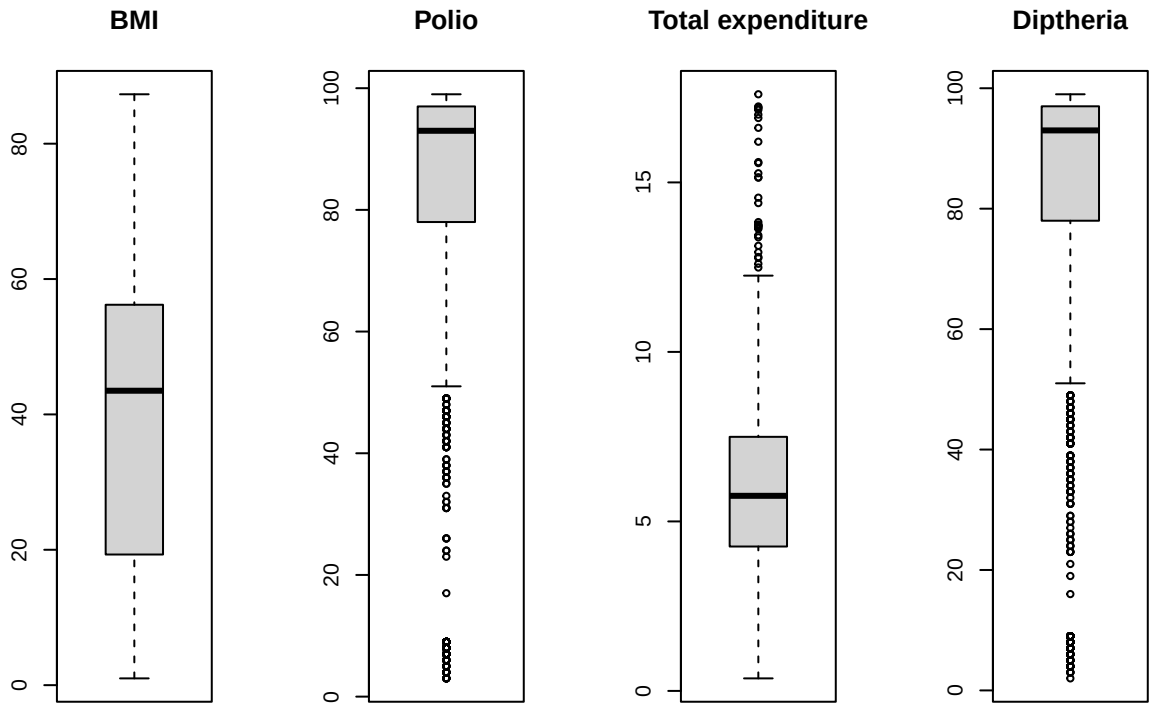
```
## Warning: package 'kableExtra' was built under R version 4.1.1
```

```
# Result table
kable(as.data.frame(missing)) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive")) %>%
  scroll_box(width = "100%", height = "250px")
```

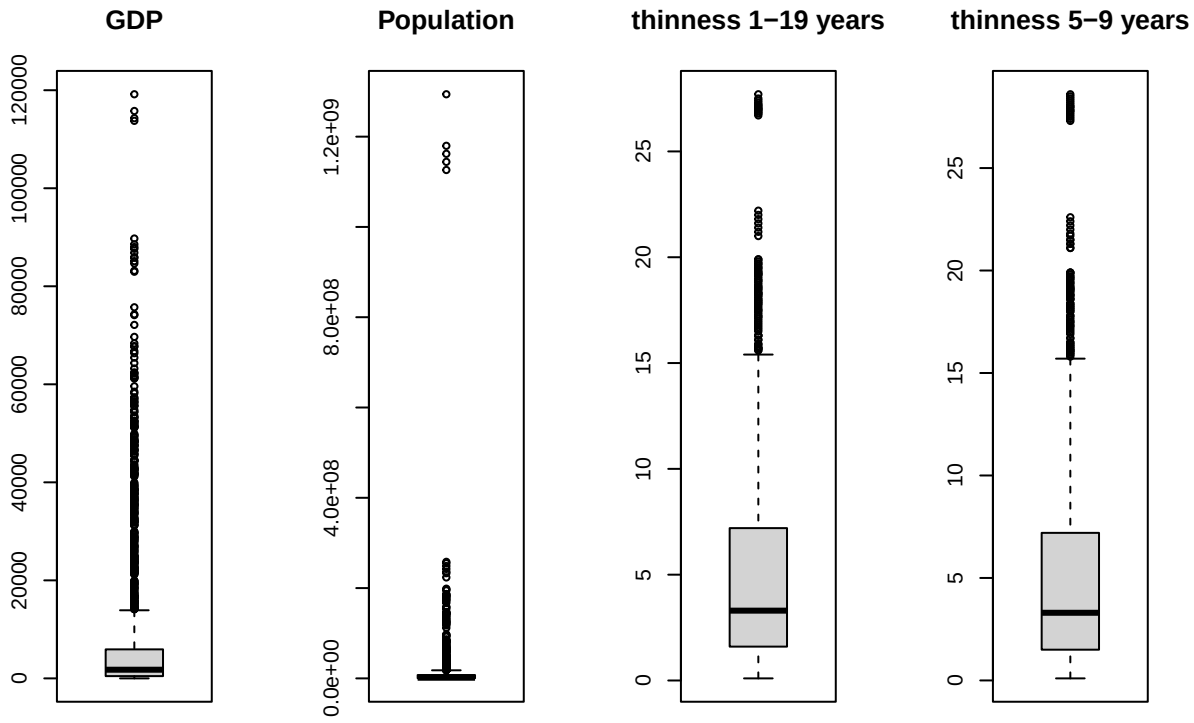
```
#boxplot for outliers
par(mfrow=c(1,4))

boxplot(data$`Life expectancy`,main='Life Expectancy')
boxplot(data$`Adult Mortality`,main='Adult Mortality')
boxplot(data$`Alcohol`,main='Alcohol')
boxplot(data$`Hepatitis B`,main='Hepatitis B')
```




```
boxplot(data$`BMI`,main='BMI')
boxplot(data$`Polio`,main='Polio')
boxplot(data$`Total expenditure`,main='Total expenditure')
boxplot(data$Diphtheria,main='Diptheria')
```

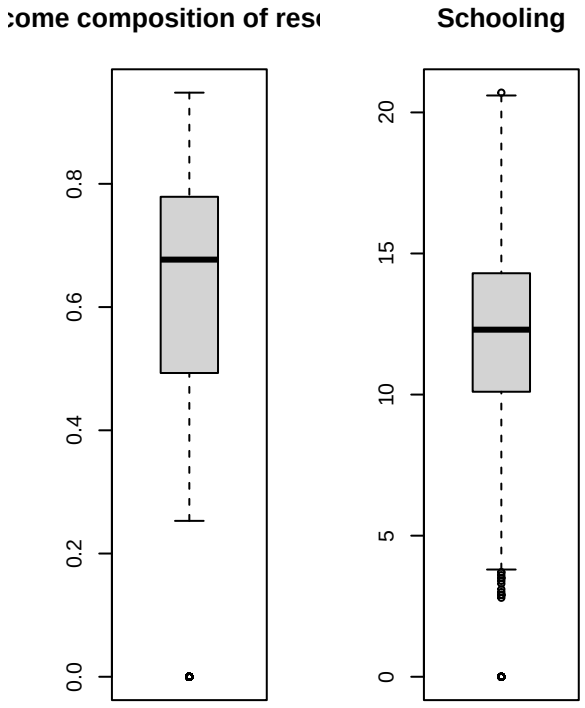



```
boxplot(data$GDP,main='GDP')
boxplot(data$Population,main='Population')
boxplot(data$`thinness  1-19 years`,main='thinness 1-19 years')
boxplot(data$`thinness 5-9 years`,main='thinness 5-9 years')
```




```
boxplot(data$`Income composition of resources`,main='income composition of resources')
boxplot(data$Schooling,main='Schooling')

#finding median values for columns with high outliers
life_md <- median(data$`Life expectancy`,  na.rm = TRUE)
mortality_md<- median(data$`Adult Mortality`,  na.rm = TRUE)
hepatitis_md <- median(data$`Hepatitis B`,  na.rm = TRUE)
polio_md <- median(data$Polio,  na.rm = TRUE)
diph_md <- median(data$Diphtheria,  na.rm = TRUE)
exp_md <- median(data$`Total expenditure`,  na.rm = TRUE)
gdp_md <- median(data$GDP,  na.rm = TRUE)
pop_md <- median(data$Population,  na.rm = TRUE)
thin19_md <- median(data$`thinness  1-19 years`,  na.rm = TRUE)
thin9_md <- median(data$`thinness 5-9 years`,  na.rm = TRUE)
school_md <- median(data$Schooling,  na.rm = TRUE)

#finding mean values for columns with low outliers
alcohol_mean <- mean(data$Alcohol,  na.rm = TRUE)
bmi_mean <- mean(data$BMI,  na.rm = TRUE)
income_mean <- mean(data$`Income composition of resources`,  na.rm = TRUE)

#putting median and mean values in place of NA
data$`Life expectancy`[is.na(data$`Life expectancy`)] <- life_md
data$`Adult Mortality`[is.na(data$`Adult Mortality`)] <- mortality_md
data$`Hepatitis B`[is.na(data$`Hepatitis B`)] <- hepatitis_md
data$Polio[is.na(data$Polio)] <- polio_md
```

```
data$Diphtheria[is.na(data$Diphtheria)] <- diph_md
data$`Total expenditure`[is.na(data$`Total expenditure`)] <- exp_md
data$GDP[is.na(data$GDP)] <- gdp_md
data$Population[is.na(data$Population)] <- pop_md
data$`thinness  1-19 years`[is.na(data$`thinness  1-19 years`)] <- thin19_md
data$`thinness 5-9 years`[is.na(data$`thinness 5-9 years`)] <- thin9_md
data$Schooling[is.na(data$Schooling)] <- school_md

data$Alcohol[is.na(data$Alcohol)] <- alcohol_mean
data$BMI[is.na(data$BMI)] <- bmi_mean
data$`Income composition of resources`[is.na(data$`Income composition of resources`)] <- income_mean

View(data)

par(mfrow=c(1,2))
```



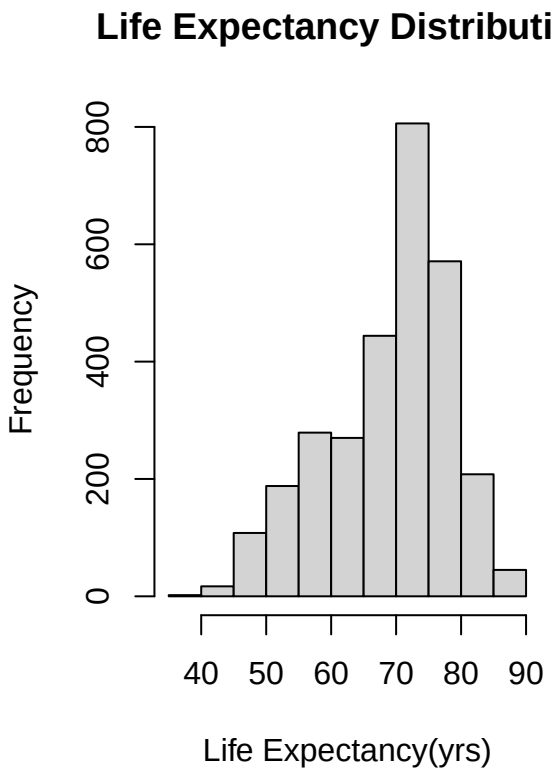

```
# target variable
# histogram
hist(data$`Life expectancy`,
     main = "Life Expectancy Distribution",
     xlab = "Life Expectancy(yrs)")
# kernel density plot with a vertical indication of location of the mean
plot(density(data$`Life expectancy`),
     main = "Distribution of Life Expectancy",
     xlab = "Life Expectancy (yrs)")
```

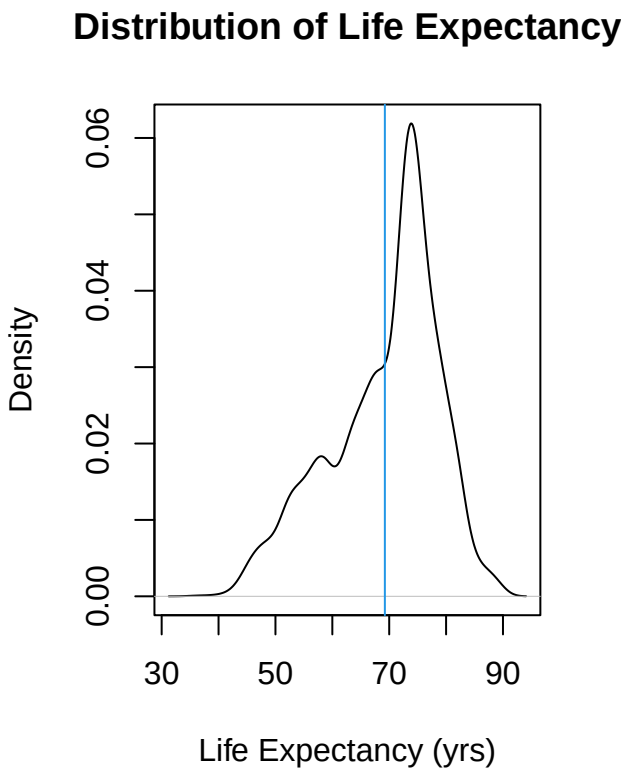
```
x = mean(data$`Life expectancy`)
x
```

```
## [1] 69.23472
```

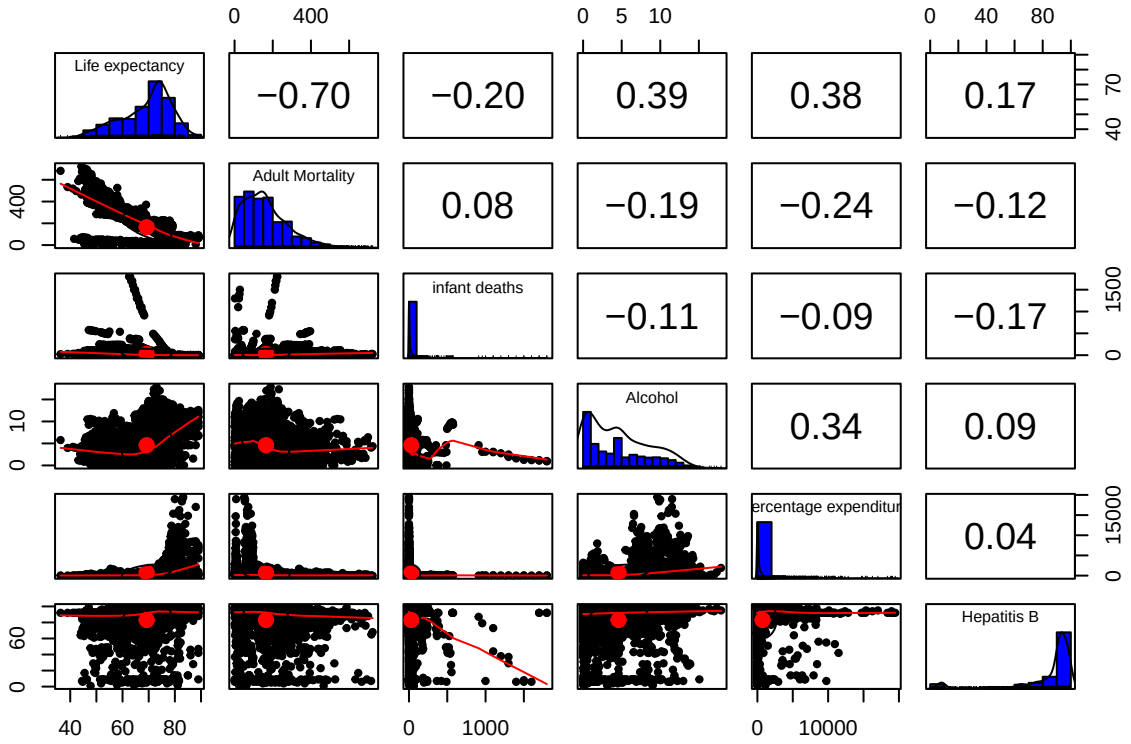
```
abline(v=mean(data$`Life expectancy`), col = 4)

#Correlations
# check correlations of the target variable with the first 5 predictors using Pearson correlation
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.1.1
```

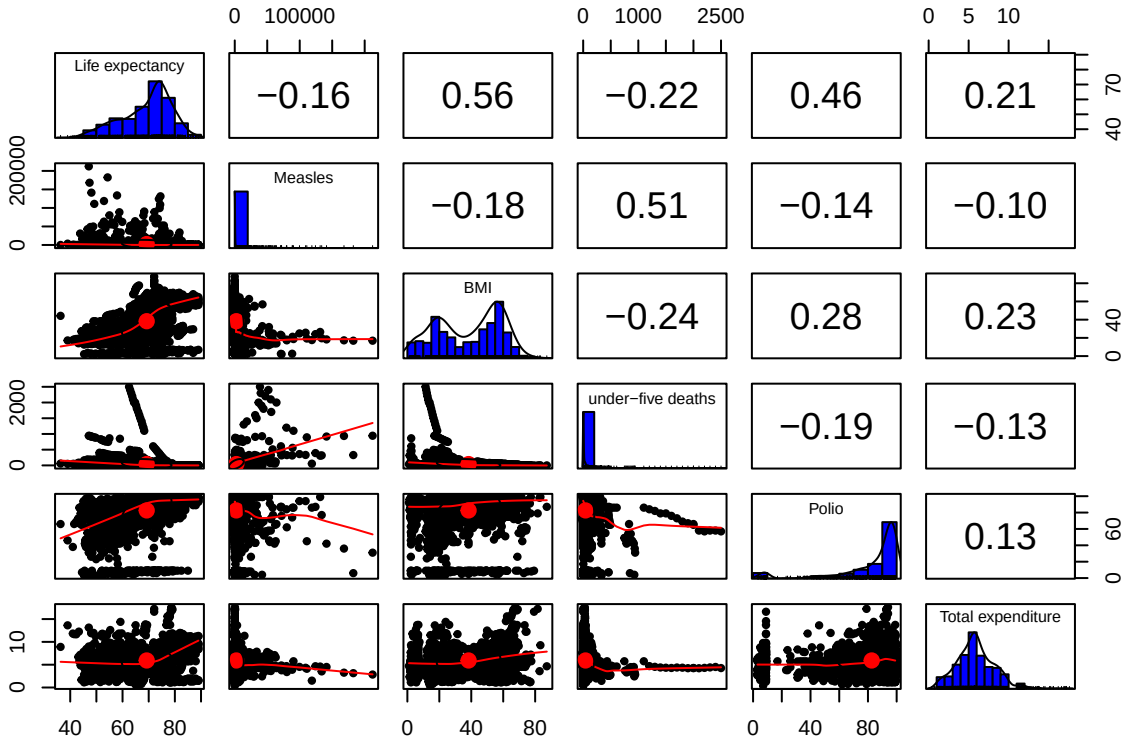




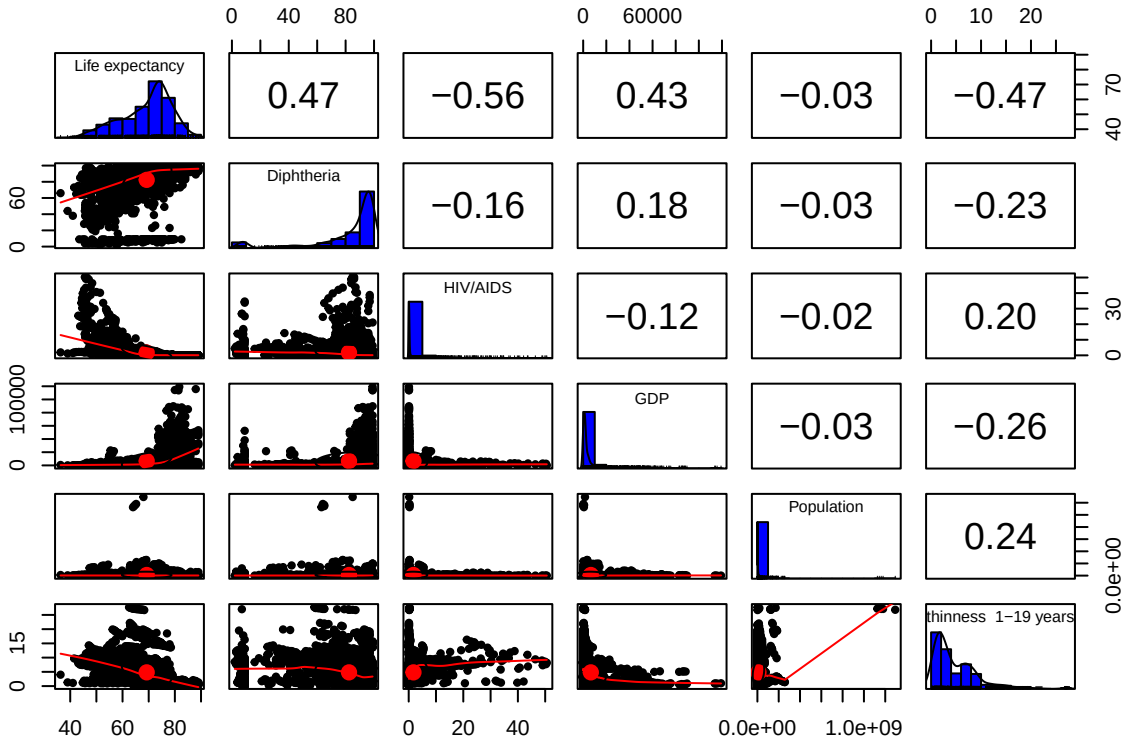

```
pairs.panels(data[,4:9],
             method = "pearson",
             hist.col = "blue",
)
```



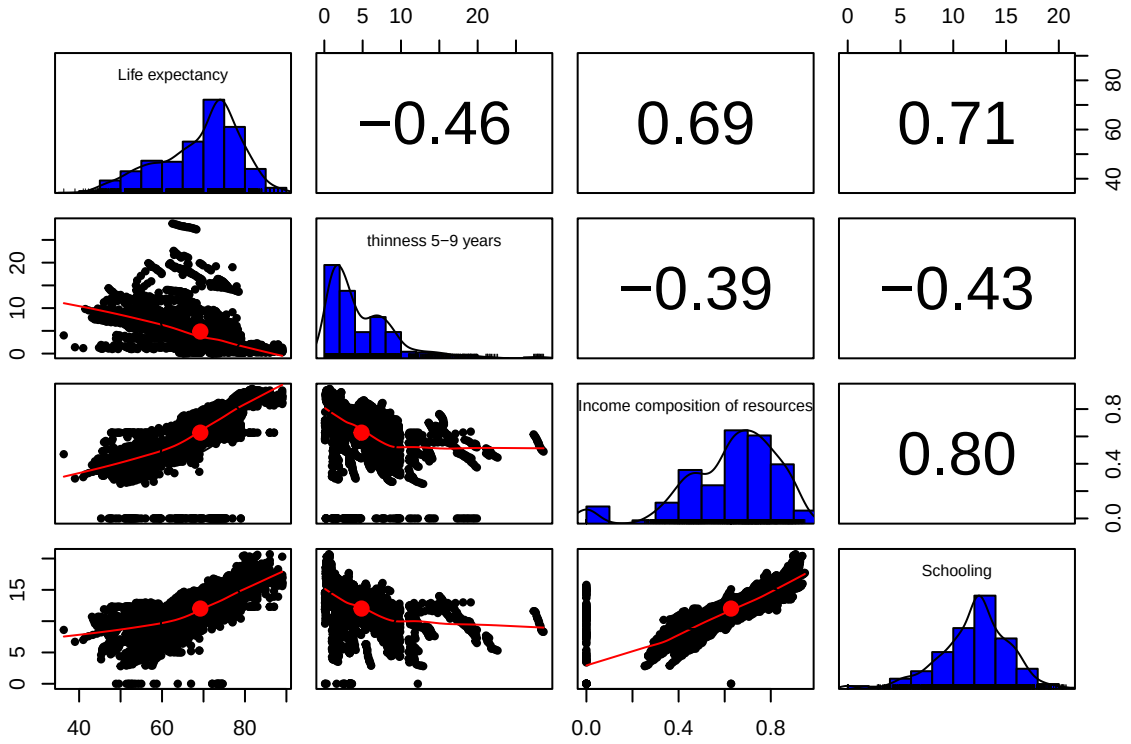

```
pairs.panels(data[,c(4,10:14)],
             method = "pearson",
             hist.col = "blue",
)
```




```
pairs.panels(data[,c(4,15:19)],
             method = "pearson",
             hist.col = "blue",
)
```




```
pairs.panels(data[,c(4,20:22)],
             method = "pearson",
             hist.col = "blue",
)
```




```
library(caTools)

dt = sort(sample(nrow(data), nrow(data)*.7))
train<-data[dt,]
test<-data[-dt,]

View(test)
View(train)

head(test)
```

```
## # A tibble: 6 x 22
##   Country  Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol
##   <fct>   <dbl> <fct>             <dbl>            <dbl>           <dbl>   <dbl>
## 1 Afghan~  2010 Devel~             58.8              279              74    0.01
## 2 Afghan~  2008 Devel~             58.1              287              80    0.03
## 3 Afghan~  2007 Devel~             57.5              295              82    0.02
## 4 Afghan~  2006 Devel~             57.3              295              84    0.03
## 5 Afghan~  2005 Devel~             57.3              291              85    0.02
## 6 Afghan~  2004 Devel~             57                293              87    0.02
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,
## #   Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,
## #   Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,
## #   Population <dbl>, thinness  1-19 years <dbl>, thinness 5-9 years <dbl>,
## #   Income composition of resources <dbl>, Schooling <dbl>
```

```
head(train)
```

```
## # A tibble: 6 x 22
##   Country  Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol
##   <fct>   <dbl> <fct>             <dbl>            <dbl>           <dbl>   <dbl>
## 1 Afghan~  2015 Devel~             65                263              62    0.01
## 2 Afghan~  2014 Devel~             59.9              271              64    0.01
## 3 Afghan~  2013 Devel~             59.9              268              66    0.01
## 4 Afghan~  2012 Devel~             59.5              272              69    0.01
## 5 Afghan~  2011 Devel~             59.2              275              71    0.01
## 6 Afghan~  2009 Devel~             58.6              281              77    0.01
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,
## #   Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,
## #   Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,
## #   Population <dbl>, thinness  1-19 years <dbl>, thinness 5-9 years <dbl>,
## #   Income composition of resources <dbl>, Schooling <dbl>
```

```
# model with predictors that are strongly correlated to target variable
Model <- lm(`Life expectancy` ~ Schooling +  `Adult Mortality` + `Income composition of resources`,
               data = train)

# check the model
summary(Model)
```

```
##
## Call:
## lm(formula = `Life expectancy` ~ Schooling + `Adult Mortality` +
##     `Income composition of resources`, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.9676  -1.9580   0.3652   2.6402  22.7544
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      57.092473   0.572304   99.76   <2e-16 ***
## Schooling                         1.021798   0.056286   18.15   <2e-16 ***
## `Adult Mortality`                -0.035693   0.001004  -35.54   <2e-16 ***
## `Income composition of resources`  9.357604   0.895053   10.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.029 on 2052 degrees of freedom
## Multiple R-squared:  0.7179, Adjusted R-squared:  0.7174
## F-statistic:  1740 on 3 and 2052 DF,  p-value: < 2.2e-16
```

```
#applying on test data
library(Metrics)

pred <- predict(Model, test)
pred
```

```
##        1        2        3        4        5        6        7        8
## 60.72681 59.79001 59.02948 58.62937 58.48356 57.14783 78.43760 75.70620
##        9       10       11       12       13       14       15       16
## 75.56931 75.20249 74.00245 73.99582 73.69026 78.08089 71.36840 74.84547
##       17       18       19       20       21       22       23       24
## 69.36943 69.08085 61.75263 61.25119 54.95829 53.73780 53.07130 64.39197
##       25       26       27       28       29       30       31       32
## 73.85656 73.90117 74.09300 74.07701 51.52433 78.22339 81.77271 77.44460
##       33       34       35       36       37       38       39       40
## 76.32803 75.32157 72.79149 72.32227 71.68656 70.95183 69.38066 84.27757
##       41       42       43       44       45       46       47       48
## 83.54685 81.15697 78.11872 80.38580 81.14807 72.34783 71.70277 75.48005
##       49       50       51       52       53       54       55       56
## 68.91084 69.07069 69.08638 68.80412 68.33708 76.77919 71.38951 71.03302
##       57       58       59       60       61       62       63       64
## 70.13940 69.99978 79.28514 76.69205 76.78042 76.66572 76.36190 75.52455
##       65       66       67       68       69       70       71       72
## 74.65151 68.29101 67.93278 67.53049 71.29756 66.12436 64.32409 79.72957
##       73       74       75       76       77       78       79       80
## 79.54175 76.01618 74.64258 74.01623 73.95420 73.50402 73.52862 69.26200
##       81       82       83       84       85       86       87       88
## 68.06598 67.98884 79.15237 83.82269 76.23766 70.05208 69.56850 69.85762
##       89       90       91       92       93       94       95       96
## 74.73523 68.30252 61.95376 61.54286 59.90644 58.25896 57.50280 67.13777
##       97       98       99      100      101      102      103      104
## 60.60292 59.94925 59.46914 56.69698 63.82304 69.78038 69.73587 69.74577
##      105      106      107      108      109      110      111      112
## 76.75906 78.15861 74.15405 73.92019 69.02791 64.91200 76.18654 74.47411
##      113      114      115      116      117      118      119      120
## 72.39658 71.63061 71.97829 72.20418 71.72506 76.59659 76.68942 76.46416
##      121      122      123      124      125      126      127      128
## 79.52888 76.21703 75.66251 73.71433 73.25392 72.95456 71.56291 67.46042
##      129      130      131      132      133      134      135      136
## 57.65070 56.98692 56.35121 55.55891 50.31423 49.12516 48.70664 48.14535
##      137      138      139      140      141      142      143      144
## 61.10312 70.36780 60.15825 56.95978 51.52456 50.66435 50.45651 73.85538
##      145      146      147      148      149      150      151      152
## 59.93503 58.65007 58.65007 72.03746 69.19985 68.85152 68.68067 68.16586
##      153      154      155      156      157      158      159      160
## 67.01574 65.92703 59.78650 59.41500 59.18975 70.17492 56.81013 55.67516
##      161      162      163      164      165      166      167      168
## 79.55314 79.17228 78.89752 53.42412 51.97768 51.21104 47.86692 63.95119
##      169      170      171      172      173      174      175      176
## 63.48762 55.28012 54.37281 54.31493 65.41881 51.71702 63.73293 63.60419
##      177      178      179      180      181      182      183      184
## 64.76379 77.66149 77.43178 75.55023 74.72130 73.60028 72.62083 71.48108
##      185      186      187      188      189      190      191      192
## 72.94262 72.45240 72.04881 68.15032 72.16948 72.09190 76.13962 75.69348
##      193      194      195      196      197      198      199      200
## 70.73844 69.51350 72.26422 63.49504 61.57072 57.19808 56.41724 56.60561
##      201      202      203      204      205      206      207      208
## 75.46329 74.61975 74.32355 74.04161 73.66467 73.06607 75.85821 76.11122
##      209      210      211      212      213      214      215      216
## 76.73262 79.13747 73.07766 71.94064 75.24512 77.96730 82.10332 77.65195
```

```
##      217      218      219      220      221      222      223      224
## 76.57053 71.97519 77.83953 79.23027 76.77594 76.29607 76.16582 75.64654
##      225      226      227      228      229      230      231      232
## 72.46335 74.96187 74.85479 75.10465 71.03562 70.46453 70.07190 69.60789
##      233      234      235      236      237      238      239      240
## 69.64358 65.43178 65.00346 74.24801 64.32529 63.68281 81.80858 79.73844
##      241      242      243      244      245      246      247      248
## 79.59567 79.29631 81.74252 81.57136 58.73427 55.01814 51.92359 71.67632
##      249      250      251      252      253      254      255      256
## 71.74607 70.87084 69.94883 75.21860 69.09059 71.94610 76.28686 69.31009
##      257      258      259      260      261      262      263      264
## 68.54827 68.09647 69.57645 75.60067 68.49295 59.91988 59.58452 58.93815
##      265      266      267      268      269      270      271      272
## 56.77830 56.58112 61.39661 39.91141 77.08752 76.76597 74.55486 74.19142
##      273      274      275      276      277      278      279      280
## 73.70740 61.77131 61.21198 61.02242 58.07997 66.34266 51.71705 47.12144
##      281      282      283      284      285      286      287      288
## 78.54581 77.67535 77.09851 76.32581 79.80832 79.50722 79.43093 79.41004
##      289      290      291      292      293      294      295      296
## 82.25756 83.49482 80.19277 80.06251 79.05465 78.60904 78.53591 77.59465
##      297      298      299      300      301      302      303      304
## 68.28756 67.93651 65.76306 65.21799 65.41170 64.15624 64.36866 74.26330
##      305      306      307      308      309      310      311      312
## 61.04577 60.89364 59.32612 66.08410 73.38301 77.32940 72.87679 72.29277
##      313      314      315      316      317      318      319      320
## 72.78713 71.06534 70.41635 80.39821 82.21869 79.08816 65.23398 65.40711
##      321      322      323      324      325      326      327      328
## 65.11091 64.63179 64.02687 68.28073 68.09563 79.97146 79.57482 79.17546
##      329      330      331      332      333      334      335      336
## 81.56517 75.07783 79.63547 66.87230 65.35079 66.63134 66.45876 72.55263
##      337      338      339      340      341      342      343      344
## 65.96823 71.12333 59.44007 67.45847 66.97935 56.36380 53.32387 53.24673
##      345      346      347      348      349      350      351      352
## 60.61693 60.11103 59.80993 68.62901 63.94621 63.09798 65.75923 66.02759
##      353      354      355      356      357      358      359      360
## 66.05750 62.20315 61.17940 69.44481 69.11950 69.20568 69.30394 73.17997
##      361      362      363      364      365      366      367      368
## 67.50488 67.17298 66.31692 75.31826 74.80300 73.08607 71.52781 81.39501
##      369      370      371      372      373      374      375      376
## 81.64051 79.98670 79.54327 67.77372 73.01830 70.14829 66.56874 66.86694
##      377      378      379      380      381      382      383      384
## 66.90188 71.33769 70.25508 68.71304 71.99917 64.85900 66.59374 66.38275
##      385      386      387      388      389      390      391      392
## 82.42255 81.14034 80.67646 80.50878 79.87307 79.47295 78.86739 78.95184
##      393      394      395      396      397      398      399      400
## 78.98264 78.08544 81.76756 79.67418 71.96632 71.89048 71.26368 68.48970
##      401      402      403      404      405      406      407      408
## 69.01932 79.10396 78.53398 78.29838 78.05931 77.99555 77.23993 73.63135
##      409      410      411      412      413      414      415      416
## 73.49130 73.19064 72.75657 79.01734 71.71734 70.62480 69.89494 68.16482
##      417      418      419      420      421      422      423      424
## 67.42706 67.09387 64.28804 64.06279 61.52077 55.08086 54.36757 73.75299
##      425      426      427      428      429      430      431      432
## 73.64244 73.54832 62.12442 61.30208 59.90236 74.81344 75.41672 75.48320
```

```
##      433      434      435      436      437      438      439      440
## 75.55415 70.64551 75.70329 68.65991 67.44626 66.95332 72.16656 64.12693
##      441      442      443      444      445      446      447      448
## 60.18751 75.72828 74.96500 71.37904 74.28861 74.36414 74.17012 73.90472
##      449      450      451      452      453      454      455      456
## 73.61614 70.86492 53.76167 51.07818 49.44707 48.77162 48.02206 48.80949
##      457      458      459      460      461      462      463      464
## 50.84619 59.45154 61.81808 61.10435 60.58398 69.69609 58.90654 69.60826
##      465      466      467      468      469      470      471      472
## 59.54755 74.38293 74.88447 75.25902 74.92125 73.63010 80.59170 78.29767
##      473      474      475      476      477      478      479      480
## 77.43121 77.21531 77.09125 76.48361 76.08751 75.34941 64.34904 71.54971
##      481      482      483      484      485      486      487      488
## 63.86185 63.62898 62.04741 61.59463 60.22375 55.16568 59.52601 56.33760
##      489      490      491      492      493      494      495      496
## 53.56631 49.47763 68.11983 68.95848 50.69692 72.99451 71.98526 70.77597
##      497      498      499      500      501      502      503      504
## 69.07677 60.27981 59.78643 58.09859 56.46492 56.32096 63.35838 76.59718
##      505      506      507      508      509      510      511      512
## 57.82502 61.95051 61.74887 61.14885 61.28183 74.70217 74.02457 72.50428
##      513      514      515      516      517      518      519      520
## 72.13669 70.96714 70.18518 69.86711 69.53838 72.51258 72.24718 72.11073
##      521      522      523      524      525      526      527      528
## 71.78918 70.99831 69.08323 67.19655 50.48923 71.00580 70.18250 66.46953
##      529      530      531      532      533      534      535      536
## 63.63052 79.42900 75.89951 74.52591 77.14376 72.40506 52.23820 52.23820
##      537      538      539      540      541      542      543      544
## 72.10103 72.01855 71.92671 73.76376 69.25807 68.17849 66.54003 70.60200
##      545      546      547      548      549      550      551      552
## 65.20484 69.92485 52.97706 70.65869 70.48611 70.34922 61.18215 60.54079
##      553      554      555      556      557      558      559      560
## 59.98463 66.15639 65.00422 62.83567 59.04503 68.16499 70.08668 69.98419
##      561      562      563      564      565      566      567      568
## 79.58960 79.48524 79.72312 79.44563 79.12310 82.58263 82.33594 82.20568
##      569      570      571      572      573      574      575      576
## 81.91122 81.34591 80.18059 69.61417 69.13461 69.39796 68.08234 67.78124
##      577      578      579      580      581      582      583      584
## 73.31196 66.94747 66.16422 65.10945 53.86576 53.34747 52.82819 59.26482
##      585      586      587      588      589      590      591      592
## 58.81649 57.00544 63.21135 81.27472 83.43602 80.87199 80.37143 80.24563
##      593      594      595      596      597      598      599      600
## 78.04941 77.91154 71.84010 69.83993 64.75040 63.11509 63.03435 62.31072
##      601      602      603      604      605      606      607      608
## 67.81121 65.99397 73.75176 72.80041 59.66429 55.97079 55.38187 70.92484
##      609      610      611      612      613      614      615      616
## 70.12787 69.88761 69.54636 70.15182 73.29086 73.20075 72.07778 72.05123
##      617      618      619      620      621      622      623      624
## 71.69702 67.50965 67.15458 67.41607 67.22695 67.09496 67.07135 67.11565
##      625      626      627      628      629      630      631      632
## 66.67320 80.24152 76.01688 79.74956 75.32187 74.79378 74.53327 74.63643
##      633      634      635      636      637      638      639      640
## 73.82052 78.66676 78.56730 77.89861 77.21666 74.77140 75.61790 73.07013
##      641      642      643      644      645      646      647      648
## 72.14211 69.92913 67.75184 66.89520 66.68105 67.50199 74.83307 73.73661
```

```
##      649      650      651      652      653      654      655      656
## 71.97368 71.24341 70.41052 69.28361 71.80045 71.31370 69.18707 68.56875
##      657      658      659      660      661      662      663      664
## 66.41891 70.63078 61.41802 61.75232 68.39440 52.84639 72.18525 71.69949
##      665      666      667      668      669      670      671      672
## 71.29274 70.02686 71.81603 71.67914 76.64464 70.56712 70.19018 72.38105
##      673      674      675      676      677      678      679      680
## 72.26462 71.95100 71.50049 74.80262 64.80302 63.68917 63.16642 78.30959
##      681      682      683      684      685      686      687      688
## 79.24183 74.73845 74.27804 76.28790 75.69061 63.42175 60.38682 67.19674
##      689      690      691      692      693      694      695      696
## 58.42930 73.48519 73.12794 72.59267 72.28364 72.29398 71.91214 77.14135
##      697      698      699      700      701      702      703      704
## 69.51560 69.52496 54.25993 68.78230 50.26103 65.63241 79.40265 75.02986
##      705      706      707      708      709      710      711      712
## 74.66054 76.02124 75.82232 74.19910 73.70126 73.52912 72.64672 72.64672
##      713      714      715      716      717      718      719      720
## 80.67299 79.57415 79.66207 82.20881 76.67002 75.39761 64.18009 67.93984
##      721      722      723      724      725      726      727      728
## 63.54004 74.31939 63.61143 63.11172 59.96566 58.26861 58.69867 60.94939
##      729      730      731      732      733      734      735      736
## 53.79606 53.68724 53.44260 43.92168 43.42197 43.42197 43.45766 43.49336
##      737      738      739      740      741      742      743      744
## 55.73613 81.21253 79.42085 78.63128 78.00764 77.73832 73.62119 73.20856
##      745      746      747      748      749      750      751      752
## 71.13055 69.73703 60.98496 59.88546 56.69864 56.26238 70.55285 70.47211
##      753      754      755      756      757      758      759      760
## 75.53274 68.97325 60.27048 60.44405 59.75728 71.34260 69.38726 68.83848
##      761      762      763      764      765      766      767      768
## 51.05639 52.39433 79.78744 79.65229 79.31549 79.36327 78.94008 80.46971
##      769      770      771      772      773      774      775      776
## 79.92856 81.08339 81.06467 78.68054 70.78746 68.95586 68.43931 68.27336
##      777      778      779      780      781      782      783      784
## 68.10568 67.95845 68.19878 67.97353 67.31420 67.25316 66.91191 66.35160
##      785      786      787      788      789      790      791      792
## 66.09122 64.78465 72.61224 72.38699 72.41604 71.88741 76.41122 73.24495
##      793      794      795      796      797      798      799      800
## 75.76463 68.82355 64.75452 64.65234 70.02713 69.97273 69.54998 47.24114
##      801      802      803      804      805      806      807      808
## 63.63918 63.62634 59.79048 58.74071 58.47218 73.67575 73.58565 73.53297
##      809      810      811      812      813      814      815      816
## 77.96927 73.22894 73.10587 70.92791 70.19992 69.97782 69.35834 69.14560
##      817      818      819      820      821      822      823      824
## 68.65049 78.74058 78.38637 77.83555 76.52148 78.40418 71.01997 70.95893
##      825      826      827      828      829      830      831      832
## 66.89188 66.40091 59.12201 61.49024 60.10374 59.73527 58.97778 56.32531
##      833      834      835      836      837      838      839      840
## 70.86289 50.90666 70.94061 69.09361 67.84749 75.82835 75.71954 75.65577
##      841      842      843      844      845      846      847      848
## 74.33787 75.94477 72.96305 72.57043 63.11172 62.39786 74.89049 75.14034
##      849      850      851      852      853      854      855      856
## 71.42824 71.46394 76.26733 75.41180 75.52205 69.03564 68.48679 68.92213
##      857      858      859      860      861      862      863      864
## 61.99356 61.23281 73.29635 73.17719 72.93465 72.81375 69.24413 68.19667
```

```
##      865      866      867      868      869      870      871      872
## 71.61260 69.56751 69.10189 68.38785 62.14102 61.77442 69.48688 61.53317
##      873      874      875      876      877      878      879      880
## 74.07705 63.44462 61.18065 71.85101 55.65578 69.37587 68.89946 60.36838
##      881      882
## 52.58004 48.38535
```

```
rmse(test$`Life expectancy`,pred)
```

```
## [1] 5.32198
```

```
# actual vs predicted in test data
plot(y = pred,
     x = test$`Life expectancy`,
     main = "Actual vs Predicted in test data ",
     xlab = "Actual Life Expectancy",
     ylab = "Predicted Life Expectancy",
     pch = 19)
abline(0,1, col = "green", lwd = 2)  # this is a perfect prediction - 45 degree line

# add the regression line
abline(lm(pred ~ test$`Life expectancy`),
       col = "red", lwd = 2)
```