Analyzing and Predicting Life Expectancy

Shiqi She

December 18, 2015

Introduction

Life expectancy is a statistical measure of how long a person or organism may live and is probably the most important measure of health. It is readily comparable across countries and asks the most fundamental question concerning health: how long can the typical person expect to live? Worldwide, the average life expectancy at birth was 71.0 years (68.5 years for males and 73.5 years for females) over the period 2010–2013 according to United Nations World Population Prospects 2012 Revision. On the country level, there are a lot of factors that would affect the life expentancy and could be considered as predictors. With environmental issues and healthcare catching an increasing amount of attention, this report is to analyze those underlying factors and try to predict the life expectancy range given those parameters.

Description of Data Used

The data used in this report comes from World Bank Group (http://www.worldbank.org/)'s World Bank Open Data (http://data.worldbank.org/). The ordinary data set includes 214 countries, covering 50 variables from 2000 to 2013. We will first treat Year as a variable, therefore we will have 214x14=2996 entries and each entry will have 51 variables. The description of the 50 variables is provided by World Bank Open Data and is attached "Definition and Source" sheet in Data_Life Expectancy2_raw.xlsx. However, among the 2996x50=149800 cells, nearly half (70256) are "..", which indicates n.a. value. Due to the high degree of incompleteness of some variables, we can not simply delete data entries that contain n.a. value, which will need to delete all! Therefore, we have to delete some variables with low completeness. The entire process is described in Appendix A, which gives us a data set with no null values (Data_Life Expectency2_no null.csv: 1779 entries and 20 variables including Country.Name and Year).

Part I Data Exploring

We will start by exploring the data set produced by Appendix A.

```r
data=read.csv("Data_Life Expectancy2_no null.csv", header=TRUE)
dim(data)
```

```
## [1] 1779   20
```

```r
sum(is.na(data))
```

```
## [1] 0
```

```r
# change names of variables
# we will denote Life Expenctancy by LE for ther rest of this report
colnames(data)[2] <- "LE"
colnames(data)[6] <- "Urban.population.percentage"
colnames(data)[7] <- "Household.consumption.expenditure"
colnames(data)[9] <- "Forest.area.percentage"
colnames(data)[10] <- "Access.to.improved.water.source.percentage"
colnames(data)[12] <- "Arable.land.per.capita"
colnames(data)[13] <- "Health.expenditure.percentage.of.GDP"
colnames(data)[14] <- "Immunization.DPT.percentage.of.children"
colnames(data)[15] <- "Access.to.improved.sanitation.facilities.percentage"
colnames(data)[16] <- "Immunization.measles.percentage.of.children"
colnames(data)[17] <- "Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure"
# summary(data) is attached in Appendix B
```

Here is the basic exploration of each variable and the plots are attached as Appendix C:

most countries have GDP.per.capita below $10000

most countries have Population.density below 500 people per sq km of land area

Urban.population.percentage looks like a normal distribution

most countries have total Household.consumption.expenditure below $500000

Unemployment looks like negatively related to LE

most countries have Forest.area.percentage less than 40%

most countries have Access.to.improved.water.source.percentage more than 95%

most countries have Food.production.index around 100

most countries have Arable.land.per.capita less than 0.5

Health.expenditure.percentage.of.GDP looks like positively related to LE

most countries have Immunization.DPT.percentage.of.children more than 90%

most countries have Access.to.improved.sanitation.facilities.percentage more than 90%

most countries have Immunization.measles.percentage.of.children more than 90%

Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure is rather diperse, but looks like there is a little negative effects on LE

rather flat except high frequency for values lower than 5 for Fixed.telephone.subscriptions.per.100.people, Mobile.cellular.subscriptions.per.100.people and Internet.users.per.100.people

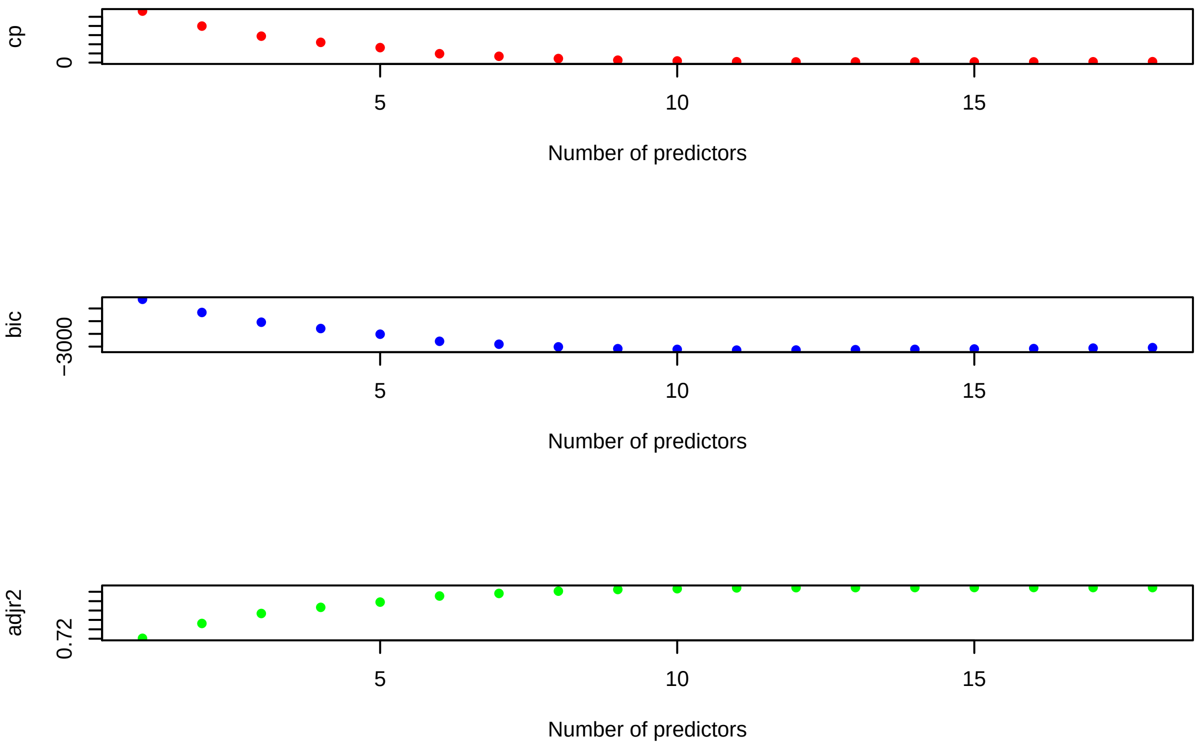
LE is growing slowly each year

Part II Basic Multiple Regression

As mentioned above, our first analysis is to find a basic multple regression model, treating Year as a normal variable. The goal is to identify significatn factors and their relationship with LE.

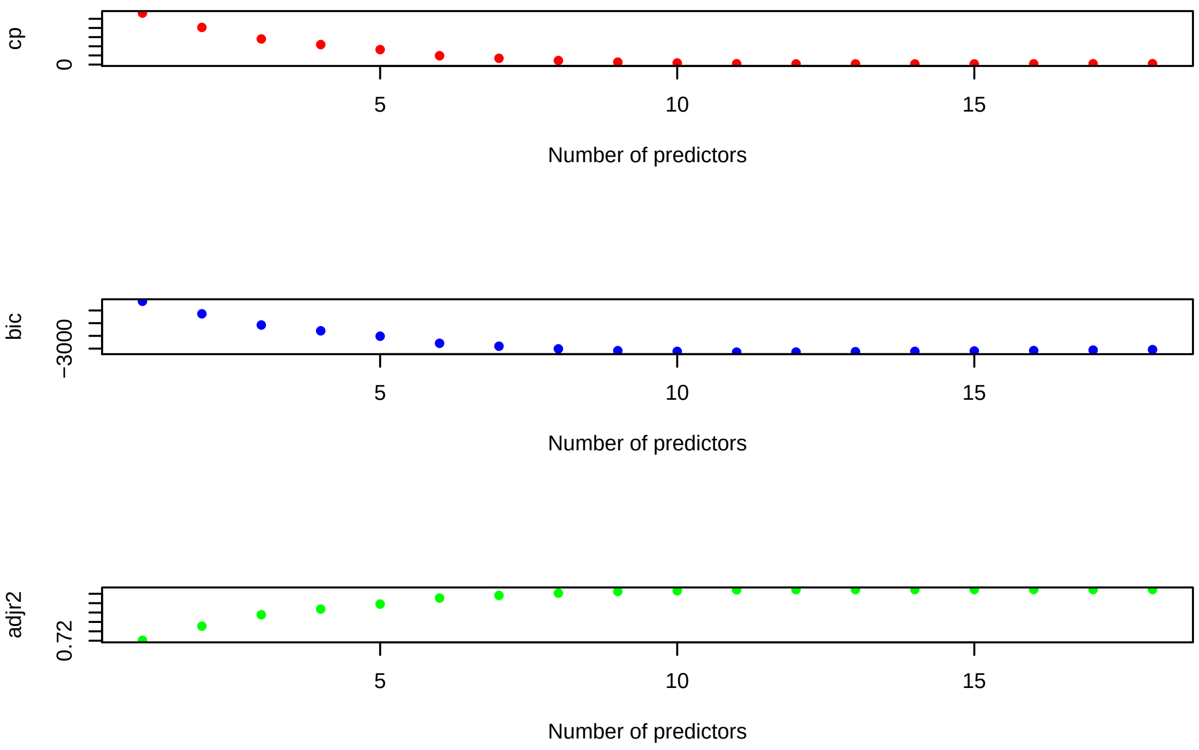
```
library(leaps)

# choose the appropriate model using forward method
fit.forward=regsubsets(LE~.-Country.Name, data, nvmax=20, method="forward")
f.f=summary(fit.forward)
par(mfrow=c(3,1))
plot(f.f$cp, xlab="Number of predictors", ylab="cp", col="red", type="p", pch=16)
plot(f.f$bic, xlab="Number of predictors", ylab="bic", col="blue", type="p", pch=16)
plot(f.f$adjr2, xlab="Number of predictors", ylab="adjr2", col="green", type="p", pch=16)
```



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

# choose the appropriate model using backward method
fit.backward=regsubsets(LE~.-Country.Name, data, nvmax=20, method="backward")
f.b=summary(fit.backward)
par(mfrow=c(3,1))
plot(f.b$cp, xlab="Number of predictors", ylab="cp", col="red", type="p", pch=16)
plot(f.b$bic, xlab="Number of predictors", ylab="bic", col="blue", type="p", pch=16)
plot(f.b$adjr2, xlab="Number of predictors", ylab="adjr2", col="green", type="p", pch=16)
```



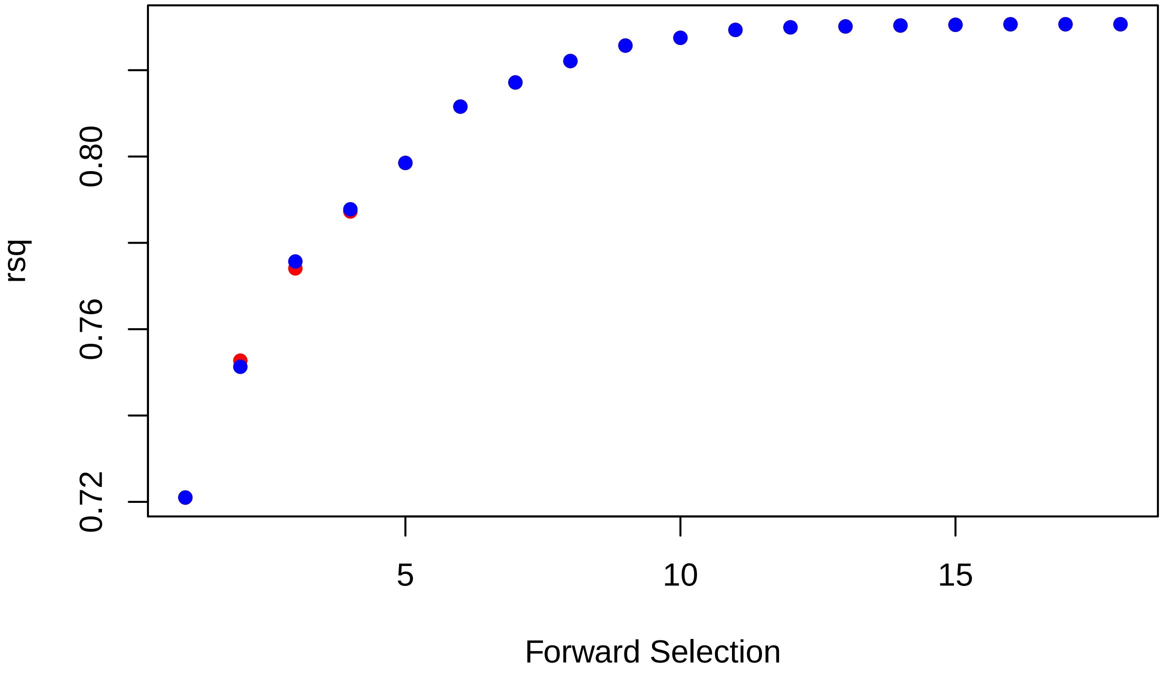
```r
par(mfrow=c(1,1))
coef(fit.forward, 7)
```

```
##                                                             (Intercept) 
##                                                             35.09661288 
##                                            Urban.population.percentage 
##                                                             0.09266439 
##                                                           Unemployment 
##                                                            -0.22491359 
##                                   Health.expenditure.percentage.of.GDP 
##                                                             0.61587790 
##                                Immunization.DPT.percentage.of.children 
##                                                             0.12299744 
##                    Access.to.improved.sanitation.facilities.percentage 
##                                                             0.17457739 
## Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure 
##                                                             0.10956803 
##                                        Internet.users..per.100.people. 
##                                                             0.05198543 
```

```r
coef(fit.backward, 7)
```

```
##                                                             (Intercept) 
##                                                             35.09661288 
##                                            Urban.population.percentage 
##                                                             0.09266439 
##                                                           Unemployment 
##                                                            -0.22491359 
##                                   Health.expenditure.percentage.of.GDP 
##                                                             0.61587790
##                                   Immunization.DPT.percentage.of.children
##                                                             0.12299744
##                         Access.to.improved.sanitation.facilities.percentage
##                                                             0.17457739
## Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure
##                                                             0.10956803
##                                            Internet.users..per.100.people.
##                                                             0.05198543
```

```
# compare the two methods
plot(f.f$rsq, ylab="rsq", col="red", type="p", pch=16,
     xlab="Forward Selection")
lines(f.b$rsq, ylab="rsq", col="blue", type="p", pch=16,
      xlab="All Subset Selection")
```



Based on cp, bic and adjusted r2, we focus on the model with 7 variables, which we believe shows a good balance of sophisticatin and explanation power. Here is the detailed anaysis of the model:

Urban.population.percentage: 0.09266439 The percentage of the total population living in urban areas is a good measure of the degree of urbanization of a population. Urbanization is relevant to a range of disciplines, including geography, sociology, economics, urban planning, and public health. Generally speaking, people living in urban areas usually have more access to healthcare services and better infrastructure, which leads to higher LE. Although this is in line with our intuition, we might expect some change in the future as most rural areas have similiar level of healthcare services and infrastructure.

Unemployment: -0.22491359 As we have seen before, unemployment rate show negative coorelation with LE. This makes sense since higher level of unemployment indicates more people with no constant income, which will lead to worse living condiction both physically and pychologically.

Health.expenditure.percentage.of.GDP: 0.61587790 Health expenditure percentage of GDP is another indicator directly related to healthcare services. We believe that it shows both its importance to a nation and a nation's financial capability to provide such services. Although higher expenditure not always reflects higher LE, a positive relationship is still expected here.

Immunization.DPT.percentage.of.children: 0.12299744 DPT refers to a class of combination vaccines against three infectious diseases in humans: diphtheria, pertussis (whooping cough), and tetanus. A child is considered adequately immunized against these diseases after receiving three doses of vaccine. Therefore, higher perentage of immunization leads to lower percentage of being infected, which leads to higher LE. Thus, the positive relationship makes sense.

Access.to.improved.sanitation.facilities.percentage: 0.17457739 According to the description, improved sanitation facilities are likely to ensure hygienic separation of human excreta from human contact. They include flush/pour flush (to piped sewer system, septic tank, pit latrine), ventilated improved pit (VIP) latrine, pit latrine with slab, and composting toilet. Therefore, it is highly likely to improve the LE.

Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure: 0.10956803 Total health expenditure is the sum of public and private health expenditure. Out-of-pocket is a major part of private health expenditure, showing the amount of money coming from individuals' pockets. Intuitively, the lower the level of out-of-pocket health expenditure, the higher the country's social wellfare and the lower the financial burden of healthcare on an individual, which will result in longer LE as expected. However, as this indicator is a just percentage, it is possible that a country with higher out-of-pocket health expenditure percentage still has a higher absolute value of health expenditure per capita.

Internet.users..per.100.people: 0.05198543 Percentage of Internet users is considered as a indicator of modernization. We believe it is an integrated indicator of a country's infrastruture, techonology and education. Therefore, a high level of Internet users per 100 people usually has a positive coorelation with LE.

Part III Predicting the LE Compared with World Average

As we have seen in Part I, there is an increasing trend in LE versus Year. Thus, we could provide a more meaningful result by comparing the prediction with the world average.

To be more speficic, we construct a new data set using the following steps:

1.Find the variables that have a high correlation with Year (plots attached in Appendix D)

2.Find the world average each year for each of these variables (e.g. world average of LE in 2000 is 81.32978723). These values are compiled in "Data_Life Expectancy3_calculated average.xlsx"

3.Modify the values to be the percentage compared with that year's world average (e.g. in 2012, United States is 1.12 times 2012 world average of LE)

Therefore, we have the new data set called "Data_Life Expectancy2_no null_percentage of average.csv"

Then, we could read the date, change some colomns' names and separate the data into train data and test data (randomly select 1000 entries as train data).

```r
data1=read.csv("Data_Life Expectancy2_no null_percentage of average.csv", header=TRUE)
data2=data1[,-c(2, 5, 12, 17, 19, 21, 24, 26, 28)]
colnames(data2)[2] <- "LE"
colnames(data2)[4] <- "GDP.per.capita"
colnames(data2)[6] <- "Urban.population.percentage"
colnames(data2)[7] <- "Household.consumption.expenditure"
colnames(data2)[9] <- "Forest.area.percentage"
colnames(data2)[10] <- "Access.to.improved.water.source.percentage"
colnames(data2)[12] <- "Arable.land.per.capita"
colnames(data2)[13] <- "Health.expenditure.percentage.of.GDP"
colnames(data2)[14] <- "Immunization.DPT.percentage.of.children"
colnames(data2)[15] <- "Access.to.improved.sanitation.facilities.percentage"
colnames(data2)[16] <- "Immunization.measles.percentage.of.children"
colnames(data2)[17] <- "Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure"
colnames(data2)[18] <- "Fixed.telephone.subscriptions.per.100.people"
colnames(data2)[19] <- "Mobile.cellular.subscriptions.per.100.people"
colnames(data2)[20] <- "Internet.users.per.100.people"

N=length(data2$LE)
set.seed(10)  # set a random seed so that we will be able to reproduce the random sample
index.train=sample(N, 1000) # Take a random sample of n=1000 from 1 to N=1779
data2.train=data2[index.train,] # Split the 1000 randomly chosen subjects as a training data
data2.test=data2[-index.train,] # The remaining subjects will be reserved for testing
```

We also divide the LE into four categories: “lower than 90% of the world average”, “between 90% to 100% of the world average”, “between 100% and 110% of the world average”, “more than 110% of the world average”

```
data2.train$LE=cut(data2.train$LE, br = c(0, 0.9, 1, 1.1, max(data2.train$LE)))
summary(data2.train$LE)
```

```
##    (0,0.9]    (0.9,1]    (1,1.1] (1.1,1.21]
##        242        153        350        255
```

```
data2.train$LE=as.factor(data2.train$LE)
levels(data2.train$LE) <- c("lower than 90% of the world average", "between 90% to 100% of the world ave

data2.test$LE=cut(data2.test$LE, br = c(0, 0.9, 1, 1.1, max(data2.test$LE)))
summary(data2.test$LE)
```

```
##    (0,0.9]    (0.9,1]    (1,1.1] (1.1,1.21]
##        188        108        288        195
```

```
data2.test$LE=as.factor(data2.test$LE)
levels(data2.test$LE) <- c("lower than 90% of the world average", "between 90% to 100% of the world aver
```

We then build different classifiers and try to find the best one. Selected models are presented here:

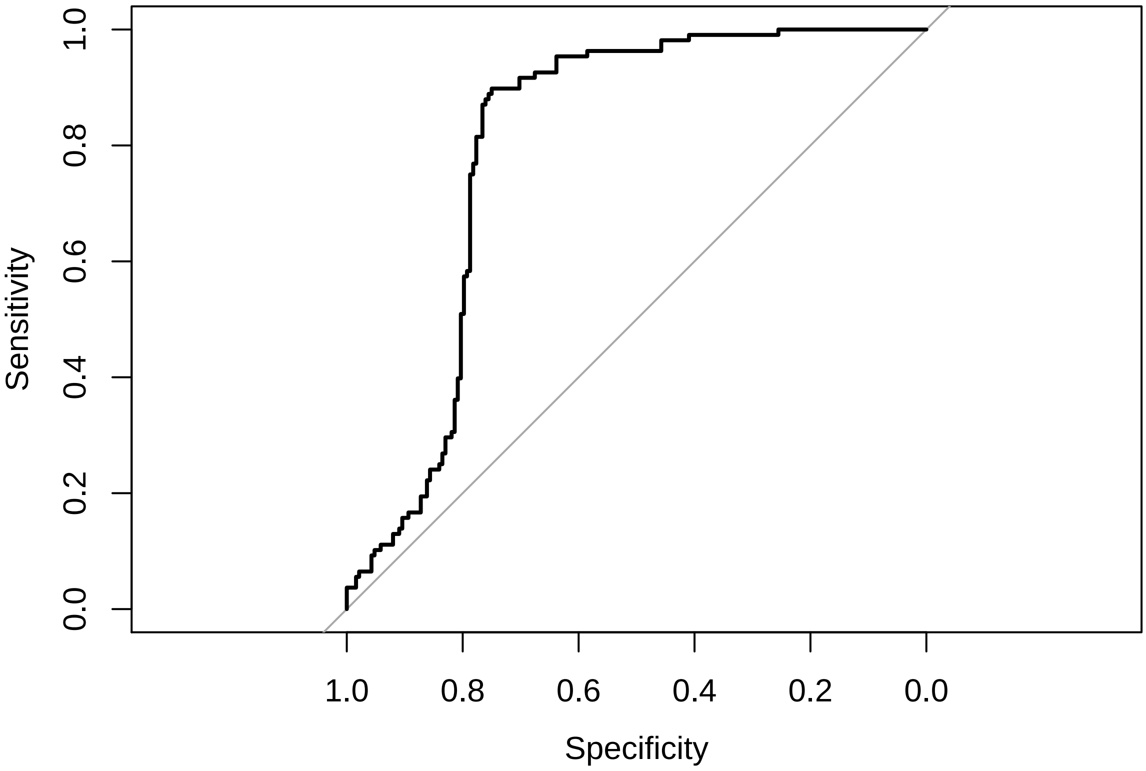
```
library(MASS)
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
##
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
# The LDA model with variables selected in Part II
lda.fit1=lda(LE~Urban.population.percentage+Unemployment+Health.expenditure.percentage.of.GDP+Immunizat
lda.fit.predict1=predict(lda.fit1, data2.test)
lda.erro1=sum(data2.test$LE != lda.fit.predict1$class)/length(data2.test$LE)
lda.erro1
```

```
## [1] 0.2567394
```

```
lda.fit.roc1=roc(data2.test$LE, lda.fit.predict1$posterior[, 2], plot=T)
```



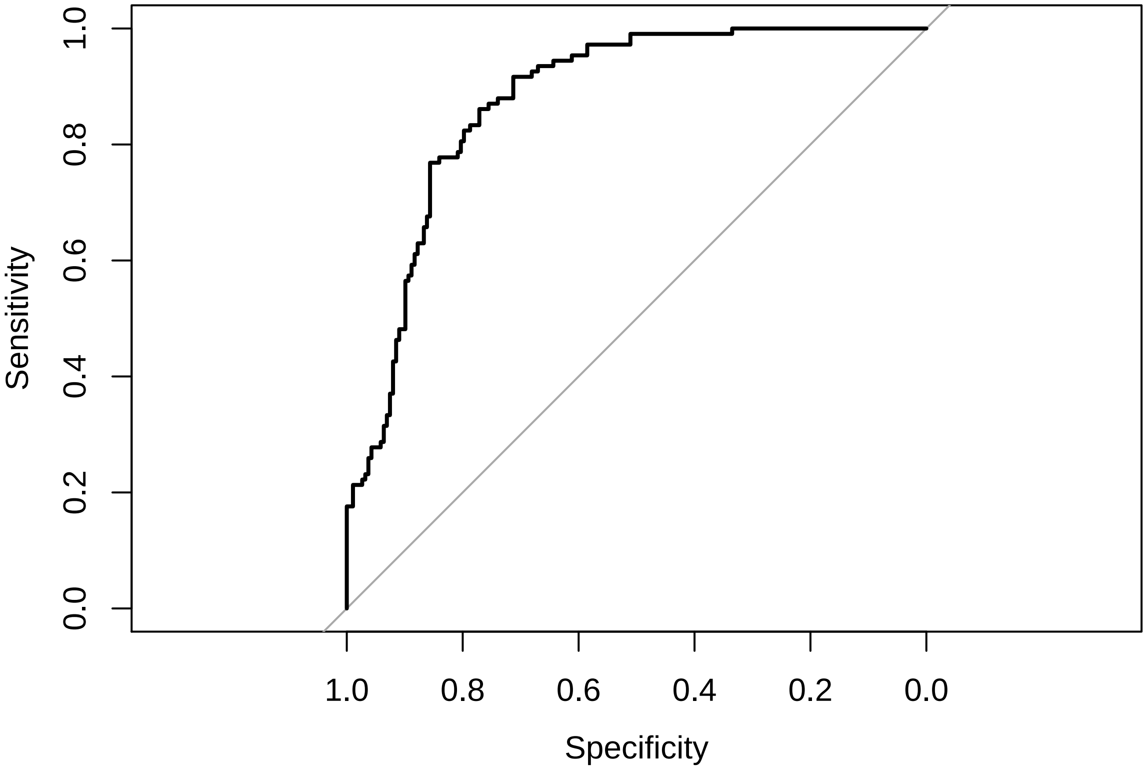
```
auc(lda.fit.roc1)
```

```
## Area under the curve: 0.8025
```

```
# The QDA model with variables selected in Part II
qda.fit1=qda(LE~Urban.population.percentage+Unemployment+Health.expenditure.percentage.of.GDP+Immunizat
qda.fit.predict1=predict(qda.fit1, data2.test)
qda.erro1=sum(data2.test$LE != qda.fit.predict1$class)/length(data2.test$LE)
qda.erro1
```

```
## [1] 0.2297818
```

```
qda.fit.roc1=roc(data2.test$LE, qda.fit.predict1$posterior[, 2], plot=T)
```



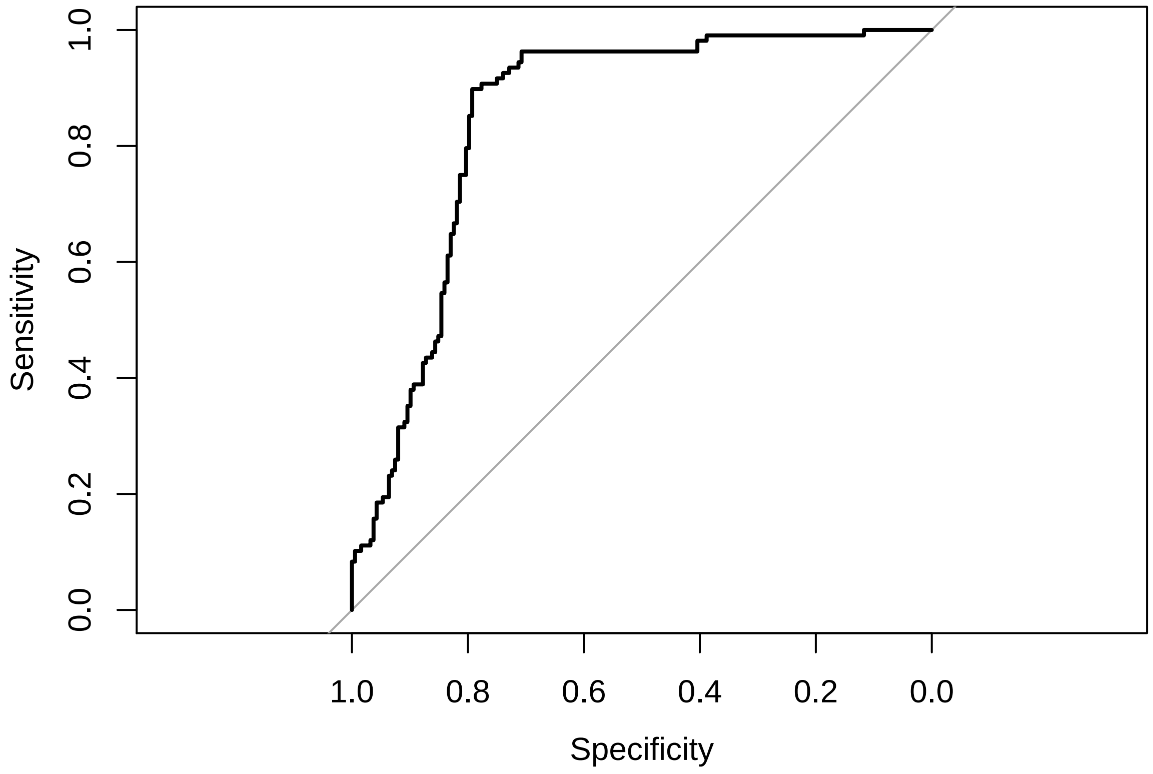
```
auc(qda.fit.roc1)
```

```
## Area under the curve: 0.8761
```

```
# The full LDA model
lda.fit=lda(LE~.-Country.Name-Year, data2.train)
lda.fit.predict=predict(lda.fit, data2.test)
lda.erro=sum(data2.test$LE != lda.fit.predict$class)/length(data2.test$LE)
lda.erro
```

```
## [1] 0.1784339
```

```
lda.fit.roc=roc(data2.test$LE, lda.fit.predict$posterior[, 2], plot=T)
```



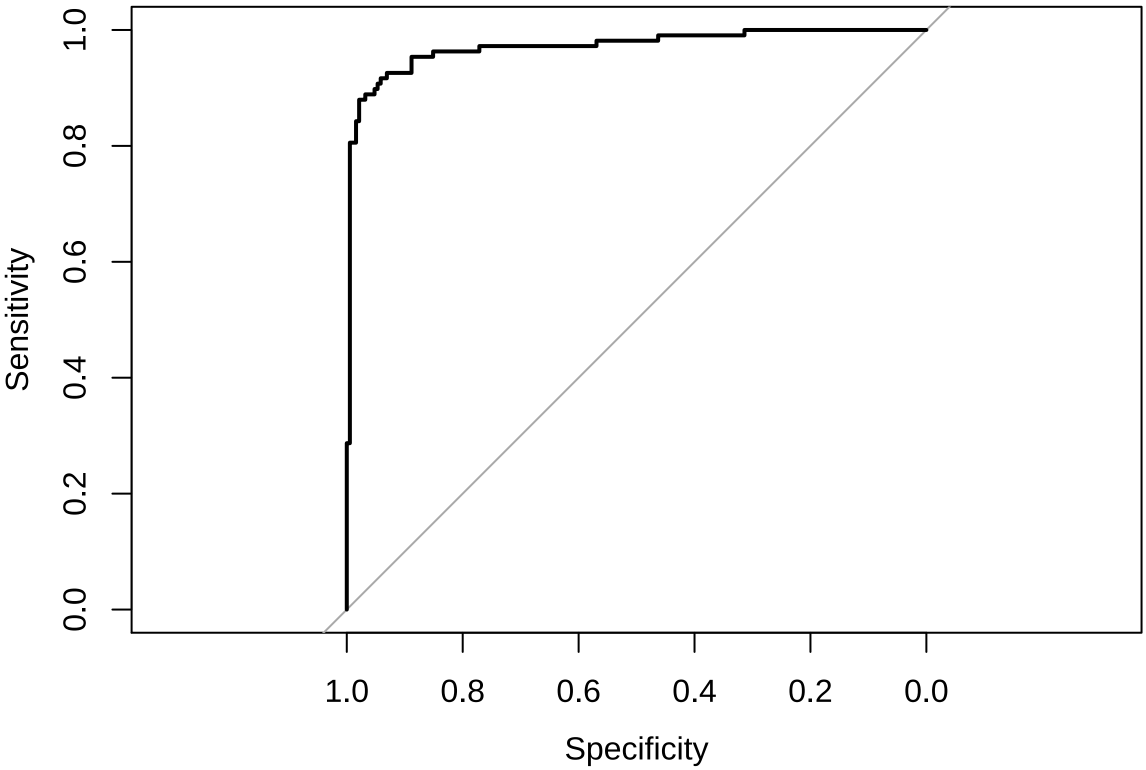
```
auc(lda.fit.roc)
```

```
## Area under the curve: 0.8502
```

```
# The full QDA model
qda.fit=qda(LE~.-Country.Name-Year, data2.train)
qda.fit.predict=predict(qda.fit, data2.test)
qda.erro=sum(data2.test$LE != qda.fit.predict$class)/length(data2.test$LE)
qda.erro
```

```
## [1] 0.1206675
```

```
qda.fit.roc=roc(data2.test$LE, qda.fit.predict$posterior[, 2], plot=T)
```



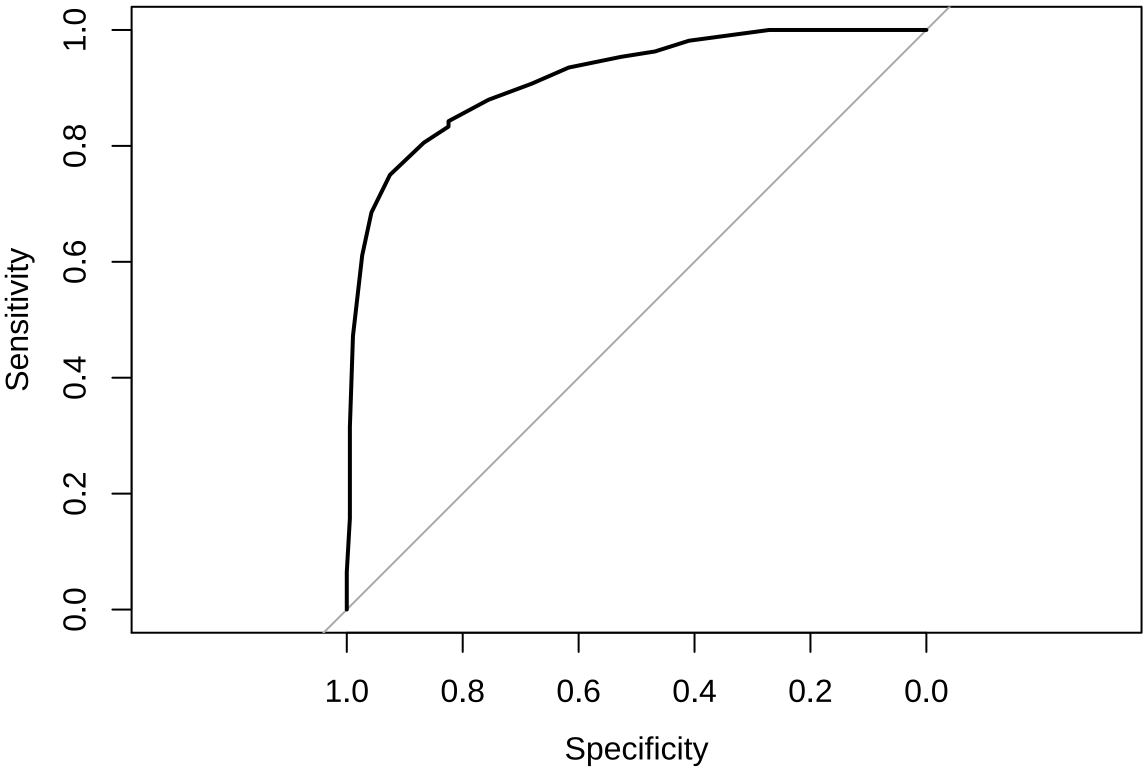
```
auc(qda.fit.roc)
```

```
## Area under the curve: 0.9715
```

```
# The KNN model
library(class)
knn.pred=knn(scale(data2.train[, -c(1,2,3)]), scale(data2.test[, -c(1,2,3)]), data2.train[, 2], k=30, p
knn.testing.erro=sum(data2.test$LE != knn.pred)/length(data2.test$LE)
knn.testing.erro
```

```
## [1] 0.1745828
```

```
posterior=attributes(knn.pred)$prob
knn.fit.roc=roc(data2.test$LE, posterior, plot=T)
```



```
auc(knn.fit.roc)
```

```
## Area under the curve: 0.9165
```

Therefore, we choose the qda model with all variables, which gives us auc of 0.9715 and error rate of 0.1206675 on the test data. The QDA models generally provide better prediction than LDA models due to their flexibility. QDA also performs better than KNN, which has a even higher flexibility, since it assumes a quadratic desision boundry given the limited number of obervations. And the models with variables selected in Part II have auc of 0.8761 and 0.8025 for qda and lda respectively, which are pretty good results as well. Therefore, we believe the variables described above do have high predicting power.

Part IV Conclusion

In this report, we selected 50 variables from World Bank Open Data and examined 20 of them closely. In the multple regression model, we found that Urban Population Percentage, Unemployment Rate, Health Expenditure Percentage of GDP, Immunization DPT Percentage of Children, Access to Improved Sanitation Facilities Percentage, Out-of-pocket Health Expenditure Percentage of Total Health Expenditure and Internet Users per 100 People are significatnt factors that affect Life Expectancy. When we are predicting a country's life expectancy compared with the world average, we found that a qda model with all variables gives us the highest auc. And 7 variables selected before are also very useful in predicting, giving us an auc of 0.8761.

Generally speaking, life expectancy is higher in a country where level of modernization is higher, public health service (including infrastructure and immunication, etc.) is better, citizens' financial capibility is higher and overall education is better.

Appendix A - Procedure to Delete N.A balues

Start with 14*214 entries 50 variables 70256 empty values delete entries with 40 or more empty values 2837 entries left delete variables with more than 2500 na values delete entries without life expectancy delete entries with 30 or more empty values 2641 entries left delete variables with more than 2000 na values delete entries with 18 or more empty values 2532 entries left delete variables with more than 1700 na values delete entries with 14 or more empty values 2484 entries left delete variables with more than 1000 na values delete entries with 5 or more empty values 1924 entries left delete variables with more than 100 na values delete entries with empty values 1779 entries left

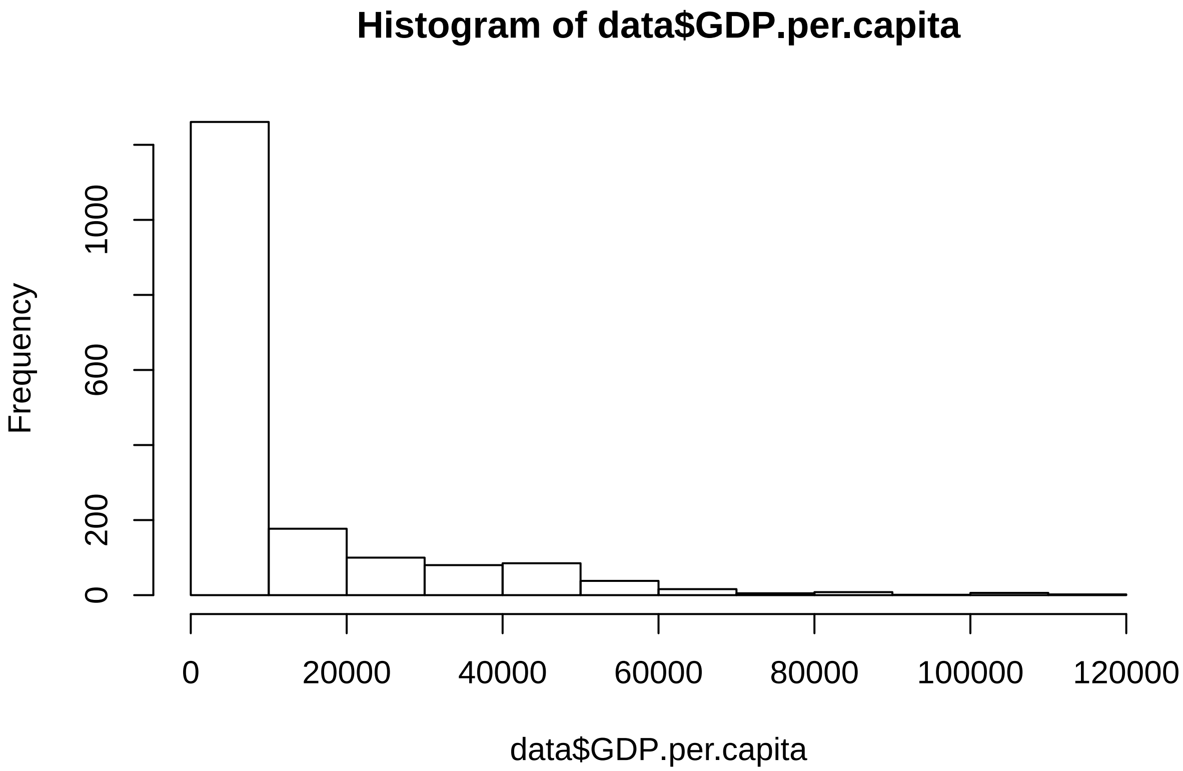
Appendix B Summary of Data

```
summary(data)
```

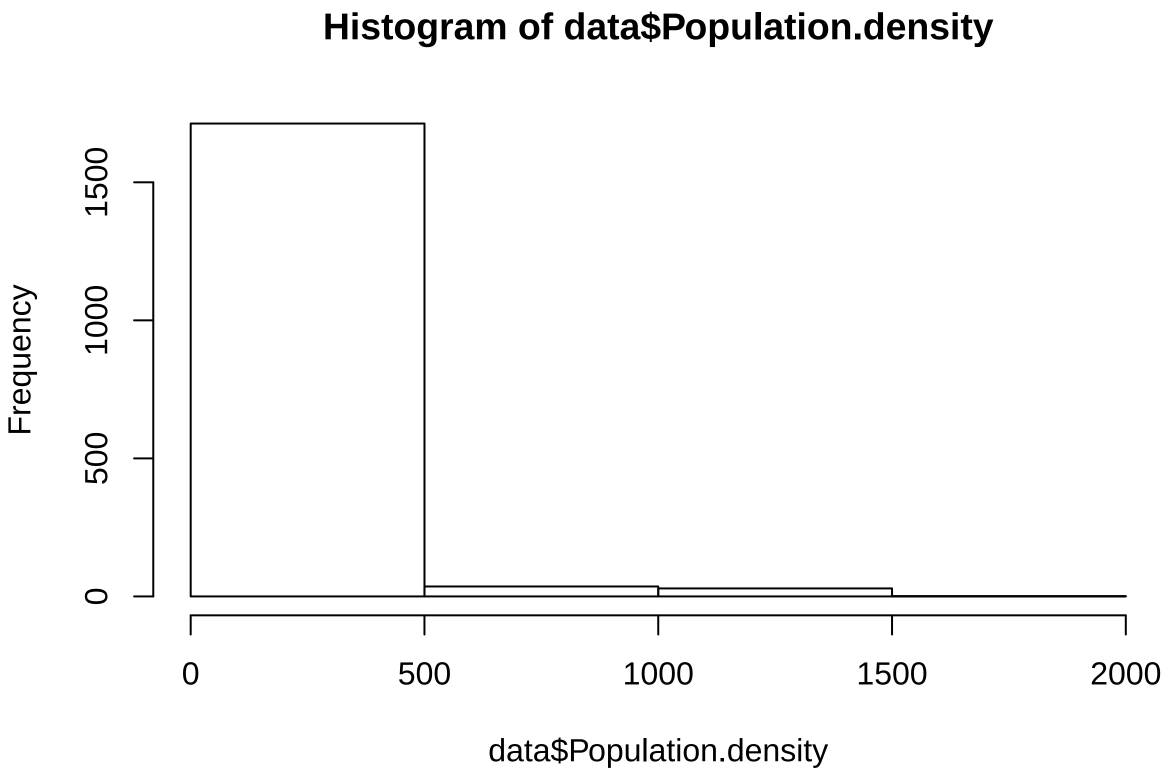
```
##        Country.Name        LE             Year      GDP.per.capita
##  Austria     :  14   Min.   :38.11   Min.   :2000   Min.   :   106.0
##  Belgium     :  14   1st Qu.:63.46   1st Qu.:2003   1st Qu.:   961.3
##  Benin       :  14   Median :72.02   Median :2006   Median :  3645.9
##  Bulgaria    :  14   Mean   :68.94   Mean   :2006   Mean   : 11226.9
##  Burkina Faso:  14   3rd Qu.:76.28   3rd Qu.:2010   3rd Qu.: 13295.7
##  Burundi     :  14   Max.   :83.10   Max.   :2013   Max.   :113239.6
##  (Other)     :1695
##  Population.density Urban.population.percentage
##  Min.   :   1.54    Min.   : 8.25
##  1st Qu.:  31.43    1st Qu.:37.40
##  Median :  72.32    Median :57.13
##  Mean   : 130.56    Mean   :55.98
##  3rd Qu.: 130.85    3rd Qu.:73.53
##  Max.   :1702.76    Max.   :98.81
##
##  Household.consumption.expenditure  Unemployment    Forest.area.percentage
##  Min.   :     205                  Min.   : 0.100   Min.   : 0.00
##  1st Qu.:    5338                  1st Qu.: 4.500   1st Qu.:11.05
##  Median :   17937                  Median : 7.200   Median :28.93
##  Mean   :  215865                  Mean   : 8.699   Mean   :29.38
##  3rd Qu.:  104728                  3rd Qu.:10.600   3rd Qu.:45.30
##  Max.   :11497182                  Max.   :38.600   Max.   :98.63
##
##  Access.to.improved.water.source.percentage Food.production.index
##  Min.   : 35.50                             Min.   : 43.63
##  1st Qu.: 80.10                             1st Qu.: 94.34
##  Median : 92.40                             Median :100.93
##  Mean   : 86.33                             Mean   :103.11
##  3rd Qu.: 99.10                             3rd Qu.:109.48
##  Max.   :100.00                             Max.   :190.76
##
##  Arable.land.per.capita Health.expenditure.percentage.of.GDP
##  Min.   :0.0000         Min.   : 0.160
##  1st Qu.:0.1000         1st Qu.: 2.110
##  Median :0.1900         Median : 3.280
##  Mean   :0.2571         Mean   : 3.739
##  3rd Qu.:0.3100         3rd Qu.: 5.065
##  Max.   :2.5700         Max.   :11.250
##
##  Immunization.DPT.percentage.of.children
##  Min.   :19.00
##  1st Qu.:83.00
##  Median :93.00
##  Mean   :87.44
##  3rd Qu.:97.00
##  Max.   :99.00
##
##  Access.to.improved.sanitation.facilities.percentage
##  Min.   :  6.60
##  1st Qu.: 47.70
##  Median : 84.80
##  Mean   : 71.41
##  3rd Qu.: 97.10
##  Max.   :100.00
##
##  Immunization.measles.percentage.of.children
##  Min.   :16.00
##  1st Qu.:80.00
##  Median :92.00
##  Mean   :86.44
##  3rd Qu.:96.00
##  Max.   :99.00
##
##  Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure
##  Min.   : 2.98
##  1st Qu.:18.84
##  Median :33.40
##  Mean   :34.81
##  3rd Qu.:49.31
##  Max.   :87.86
##
##  Fixed.telephone.subscriptions..per.100.people.
##  Min.   : 0.00
##  1st Qu.: 3.14
##  Median :13.66
##  Mean   :19.65
##  3rd Qu.:31.23
##  Max.   :74.76
##
##  Mobile.cellular.subscriptions..per.100.people.
##  Min.   :  0.00
##  1st Qu.: 16.88
##  Median : 58.91
##  Mean   : 60.73
##  3rd Qu.: 98.57
##  Max.   :208.94
##
##  Internet.users..per.100.people.
##  Min.   : 0.02
##  1st Qu.: 2.90
##  Median :12.80
##   Mean   :24.54
##   3rd Qu.:40.12
##   Max.   :96.21
##
```

Appendix C Plots of Data

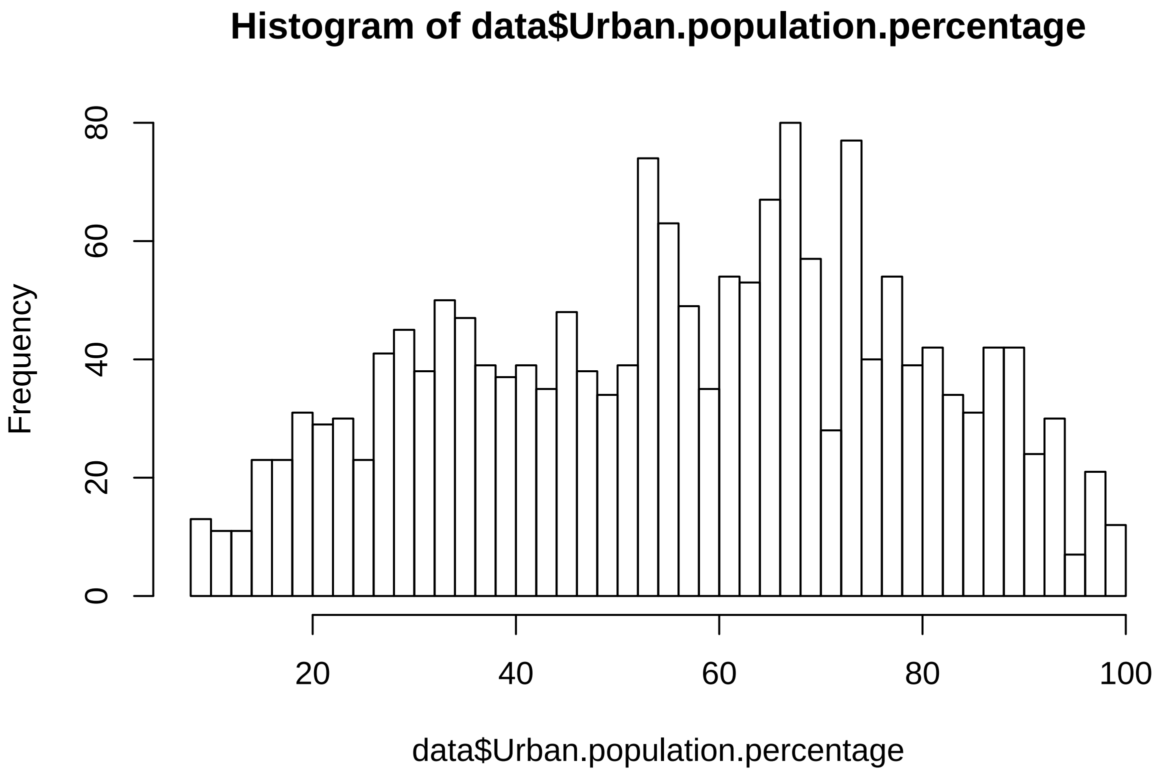
```
# most countries have GDP.per.capita below $10000
hist(data$GDP.per.capita, breaks=10)
```



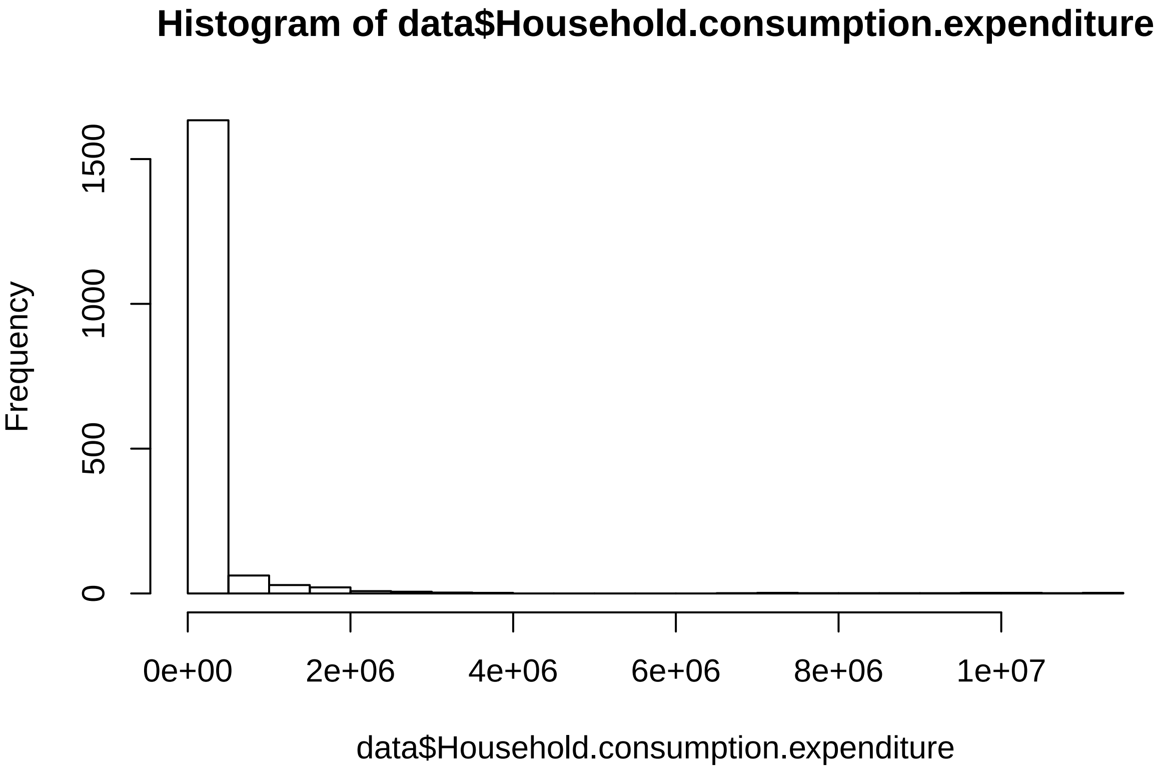
```
# most countries have Population.density below 500 people per sq km of land area
hist(data$Population.density, breaks=4)
```



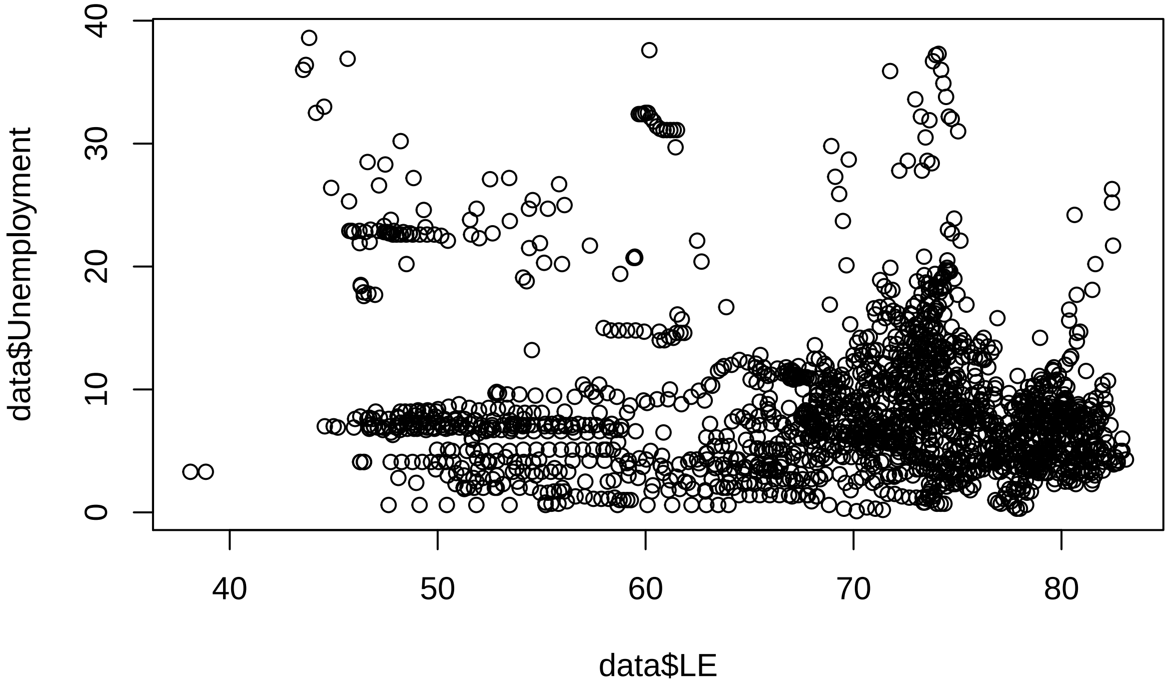
```
# Urban.population.percentage looks like a normal distribution
hist(data$Urban.population.percentage, breaks=50)
```



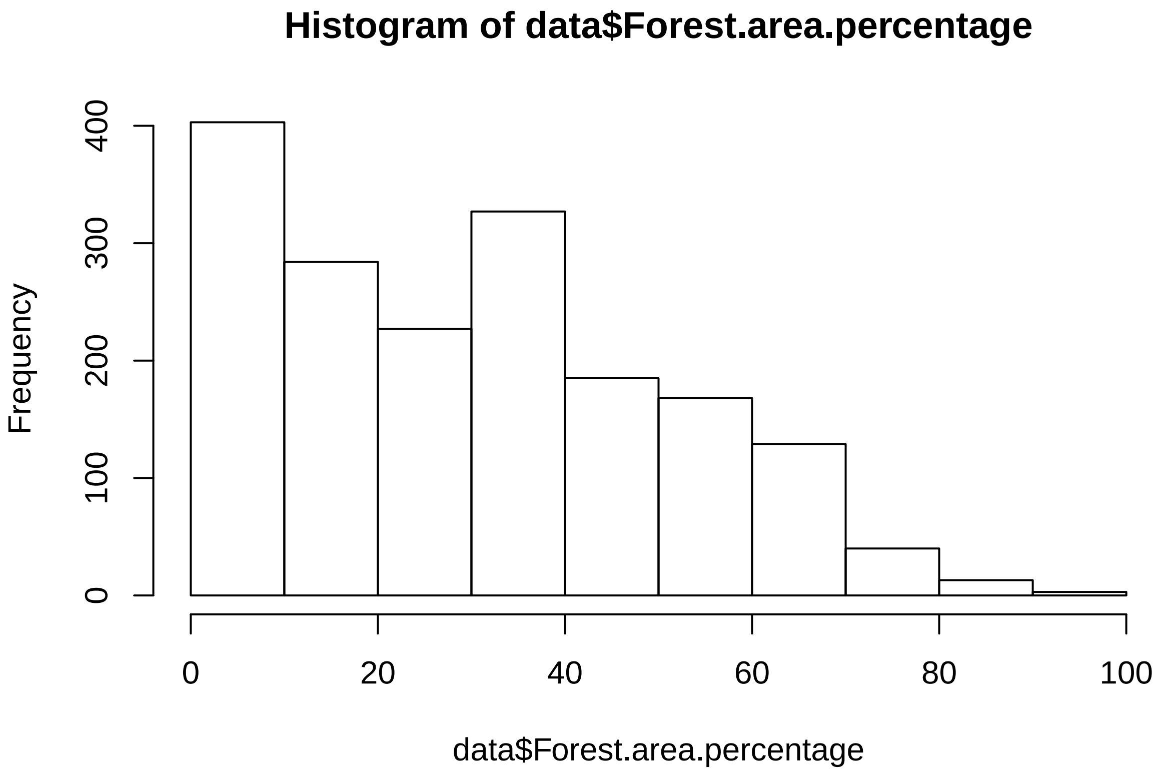
```r
# most countries have total Household.consumption.expenditure below $500000
hist(data$Household.consumption.expenditure, breaks=20)
```



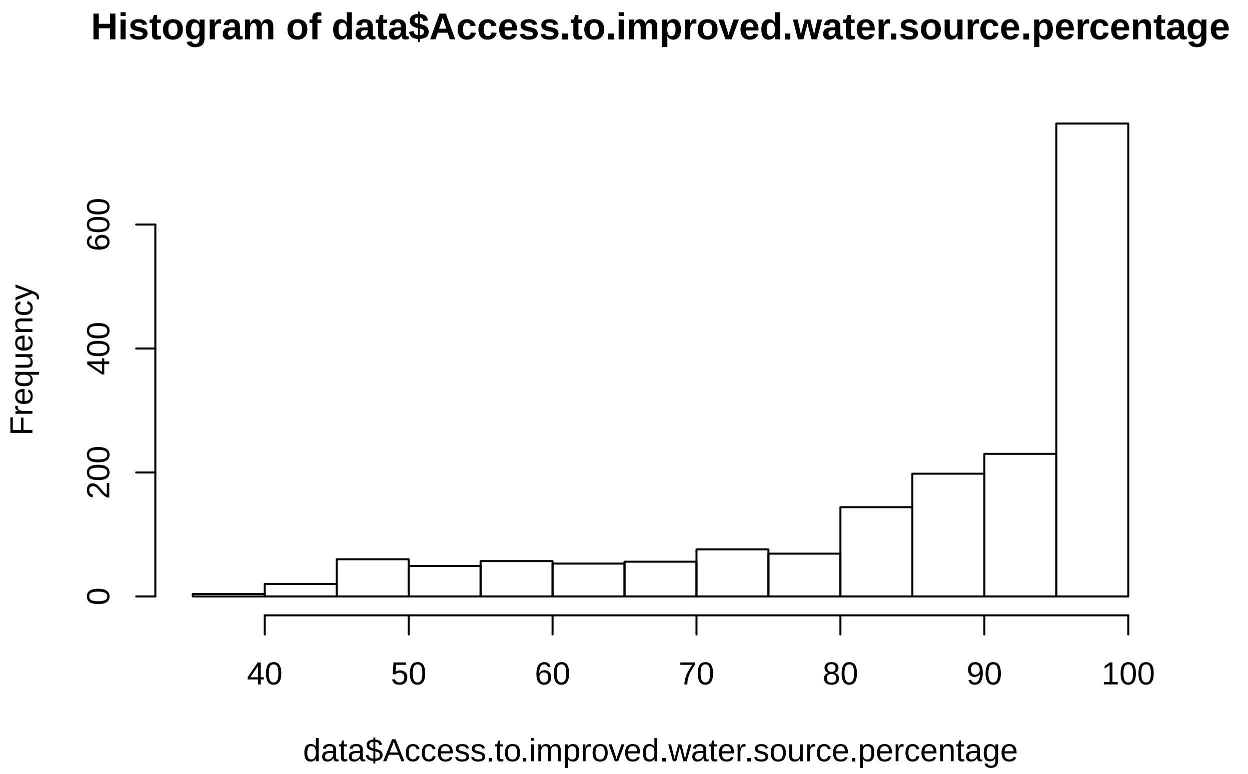
```r
# Unemployment looks like negatively related to LE
plot(data$LE, data$Unemployment)
```



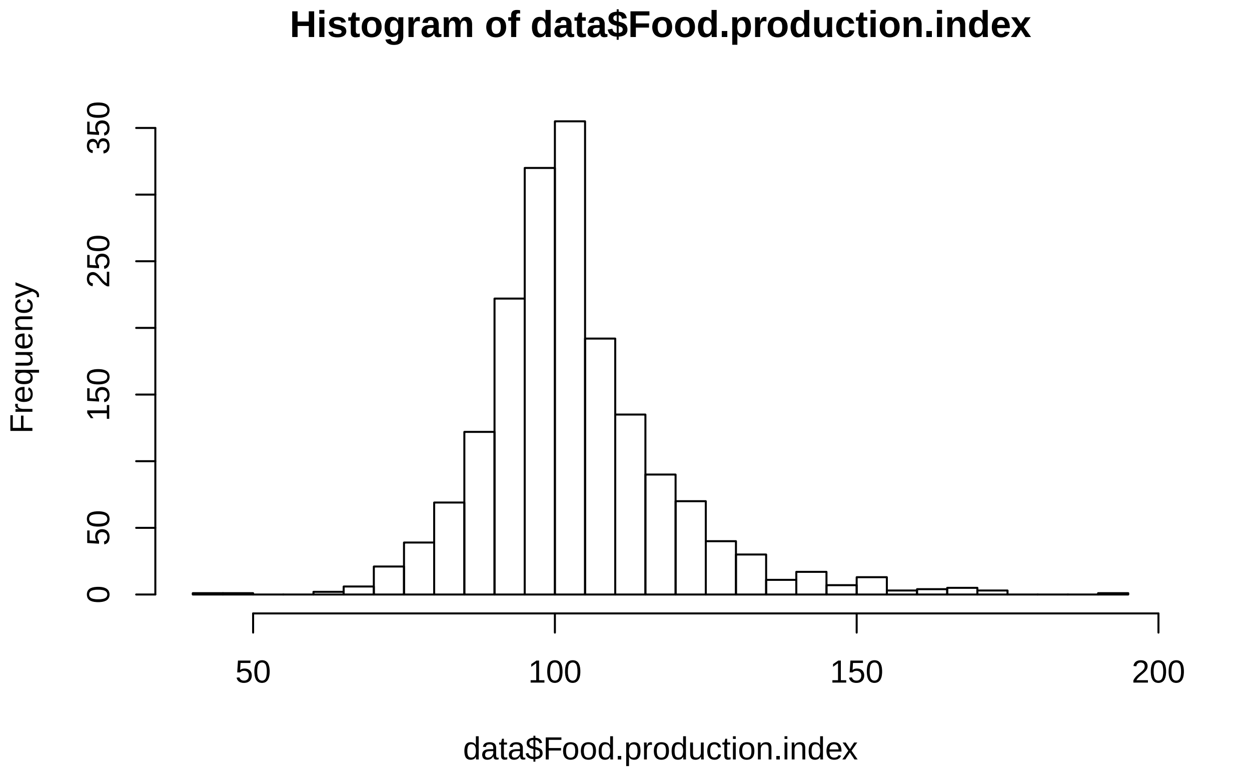
```r
# most countries have Forest.area.percentage less than 40%
hist(data$Forest.area.percentage, breaks=10)
```



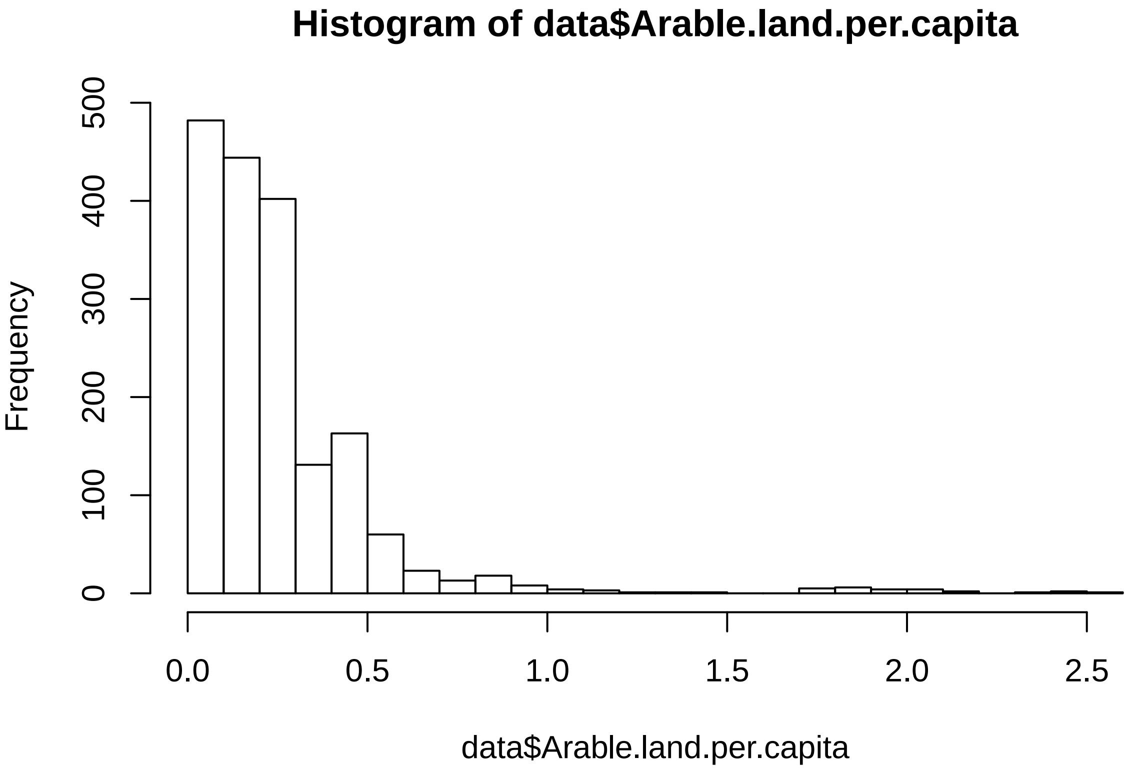
```r
# most countries have Access.to.improved.water.source.percentage more than 95%
hist(data$Access.to.improved.water.source.percentage, breaks=10)
```



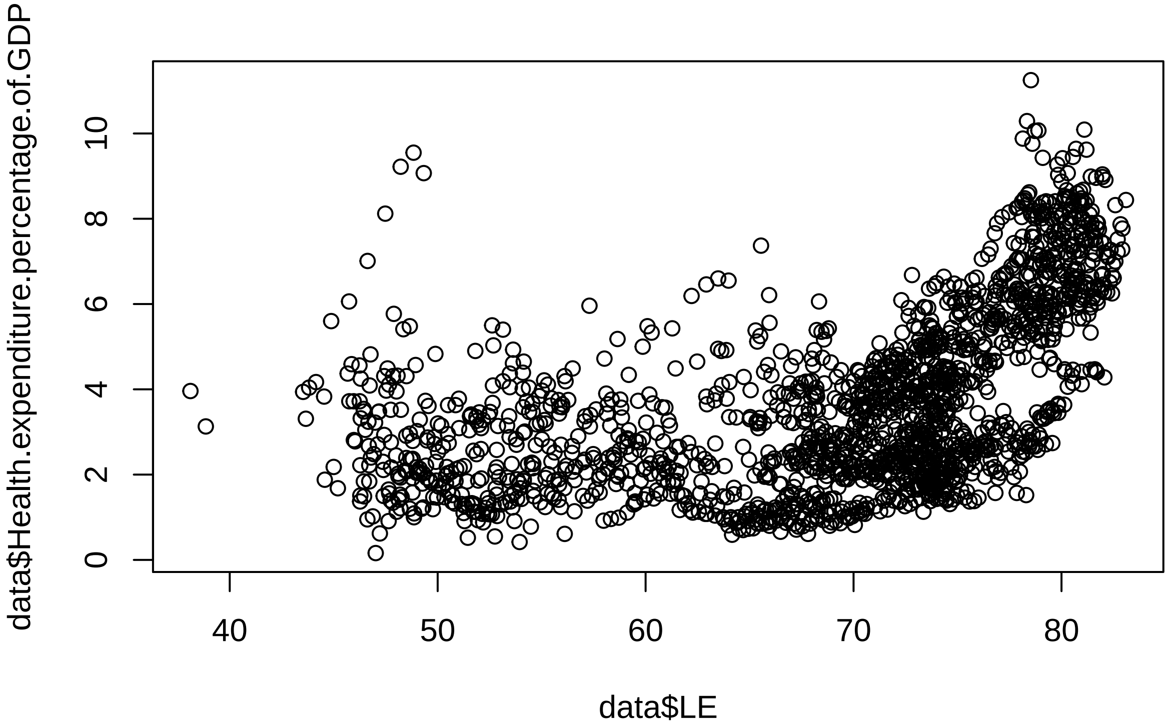
```
# most countries have Food.production.index around 100
hist(data$Food.production.index, breaks=30)
```



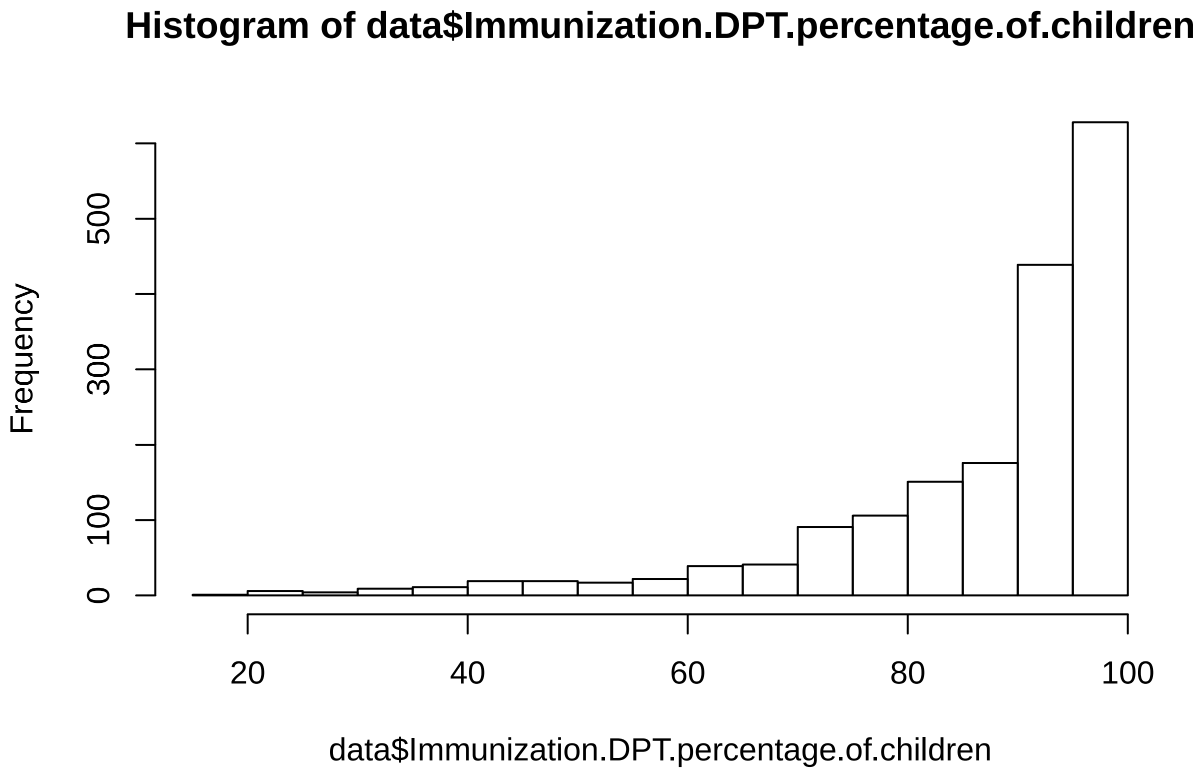
```r
# most countries have Arable.land.per.capita less than 0.5
hist(data$Arable.land.per.capita, breaks=20)
```



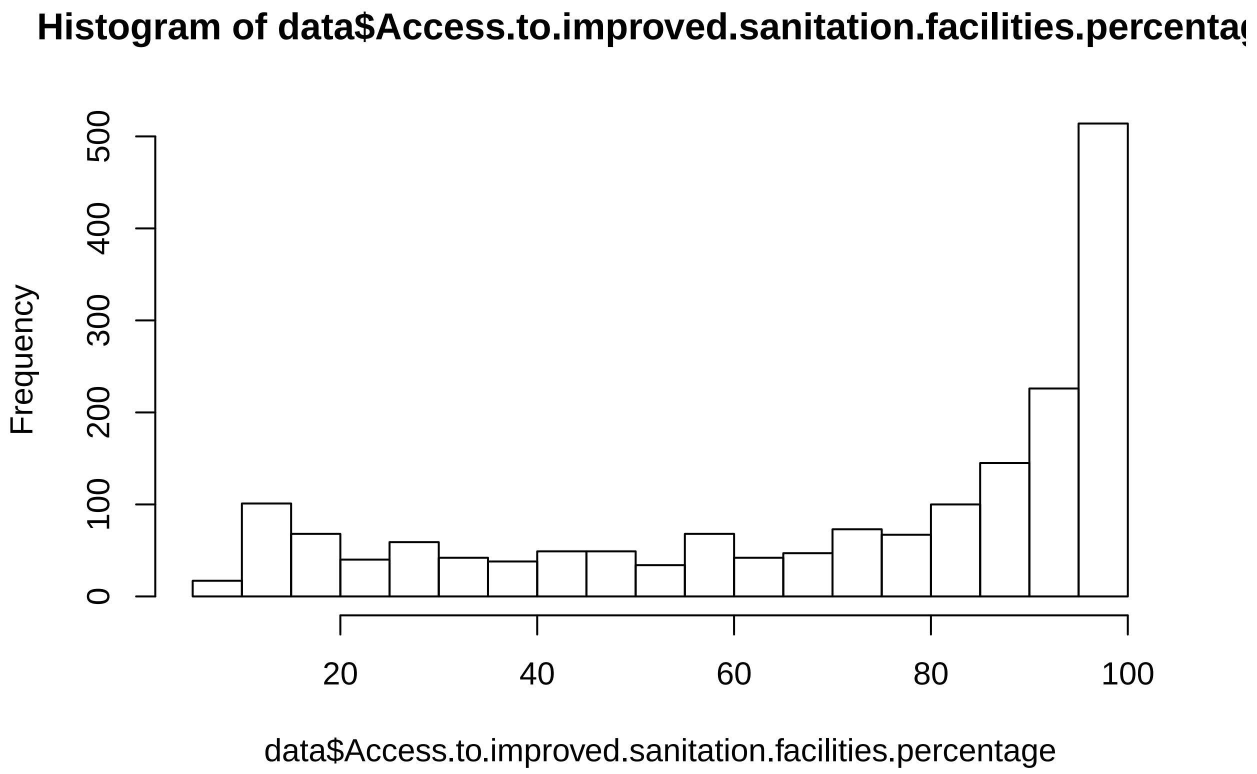
```r
# Health.expenditure.percentage.of.GDP looks like positively related to LE
plot(data$LE, data$Health.expenditure.percentage.of.GDP)
```



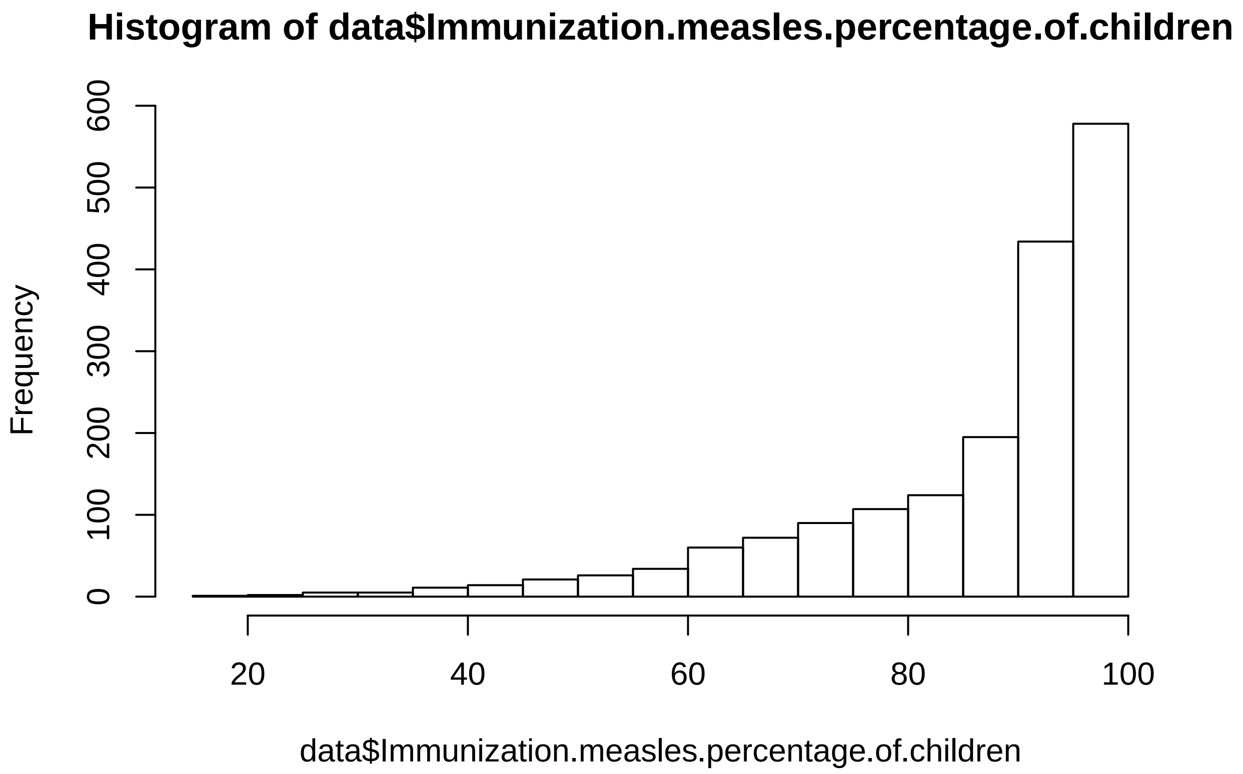
```r
# most countries have Immunization.DPT.percentage.of.children more than 90%
hist(data$Immunization.DPT.percentage.of.children, breaks=20)
```



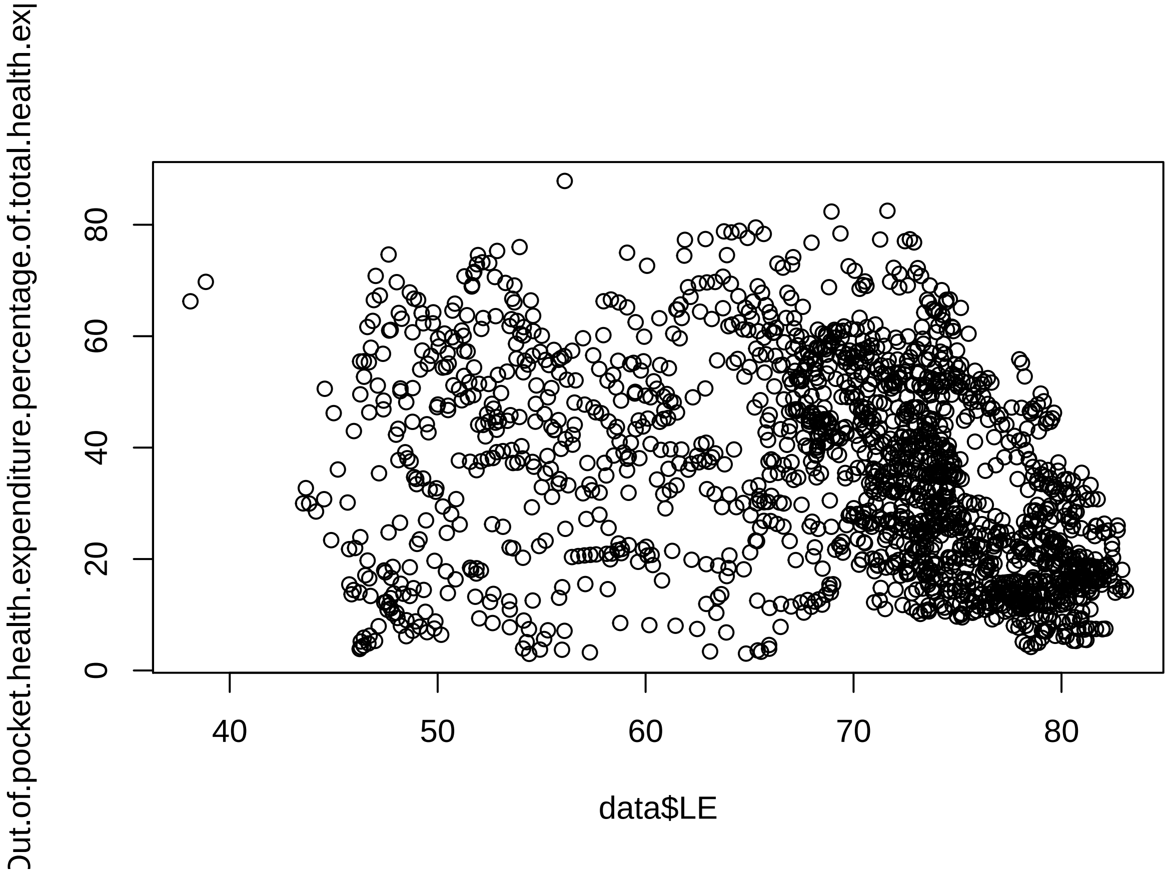
```r
# most countries have Access.to.improved.sanitation.facilities.percentage more than 90%
hist(data$Access.to.improved.sanitation.facilities.percentage, breaks=20)
```



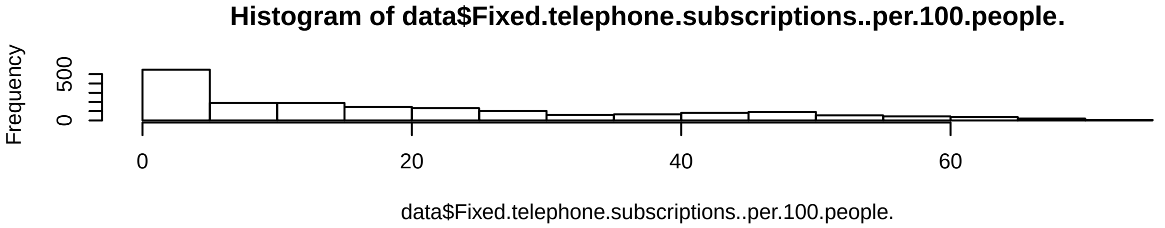
```r
# most countries have Immunization.measles.percentage.of.children more than 90%
hist(data$Immunization.measles.percentage.of.children, breaks=20)
```

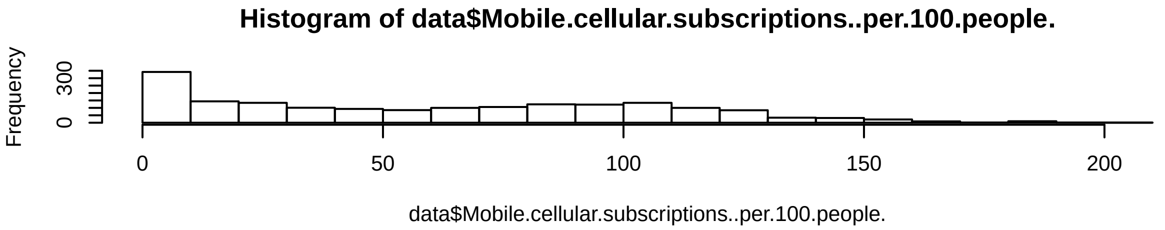


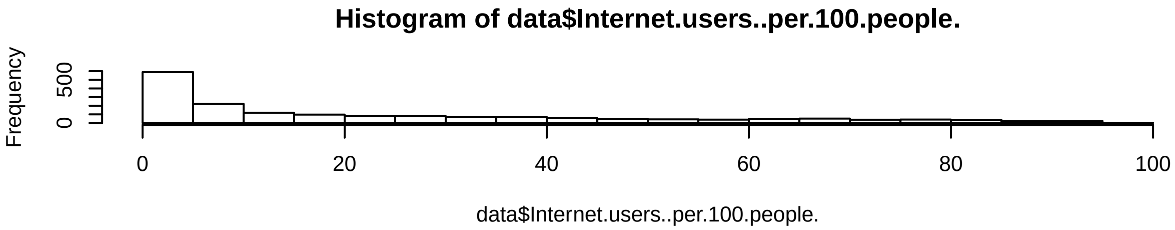
```
# rather diperse, but looks like there is a little negative effects on LE
plot(data$LE, data$Out.of.pocket.health.expenditure.percentage.of.total.health.expenditure)
```



```
# rather flat except high frequency for values lwoer than 5
par(mfrow=c(3,1))
hist(data$Fixed.telephone.subscriptions..per.100.people., breaks=20)
hist(data$Mobile.cellular.subscriptions..per.100.people., breaks=20)
hist(data$Internet.users..per.100.people., breaks=20)
```

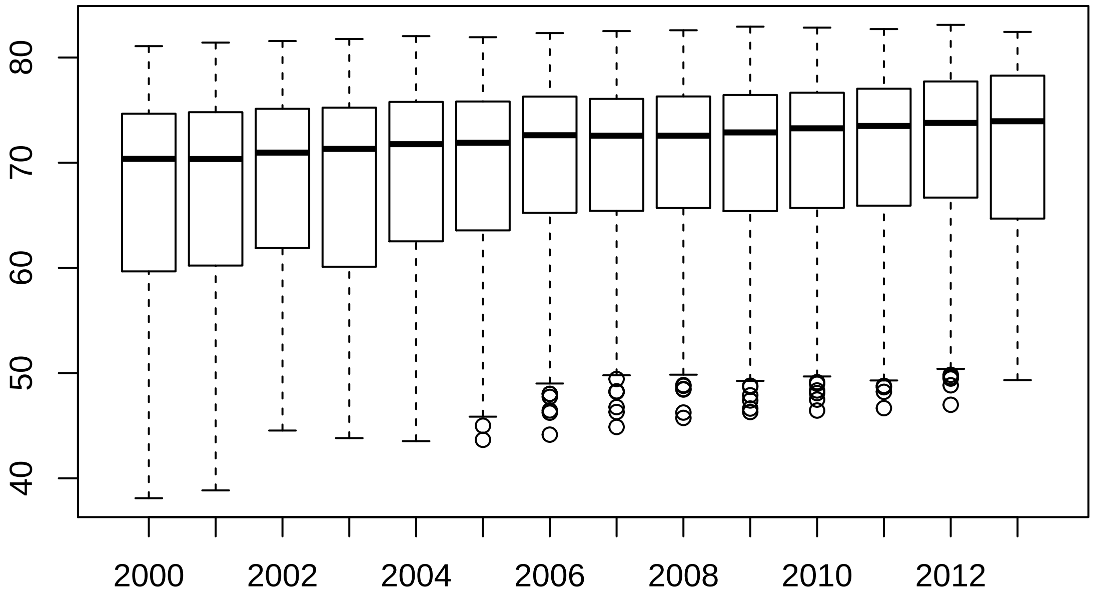






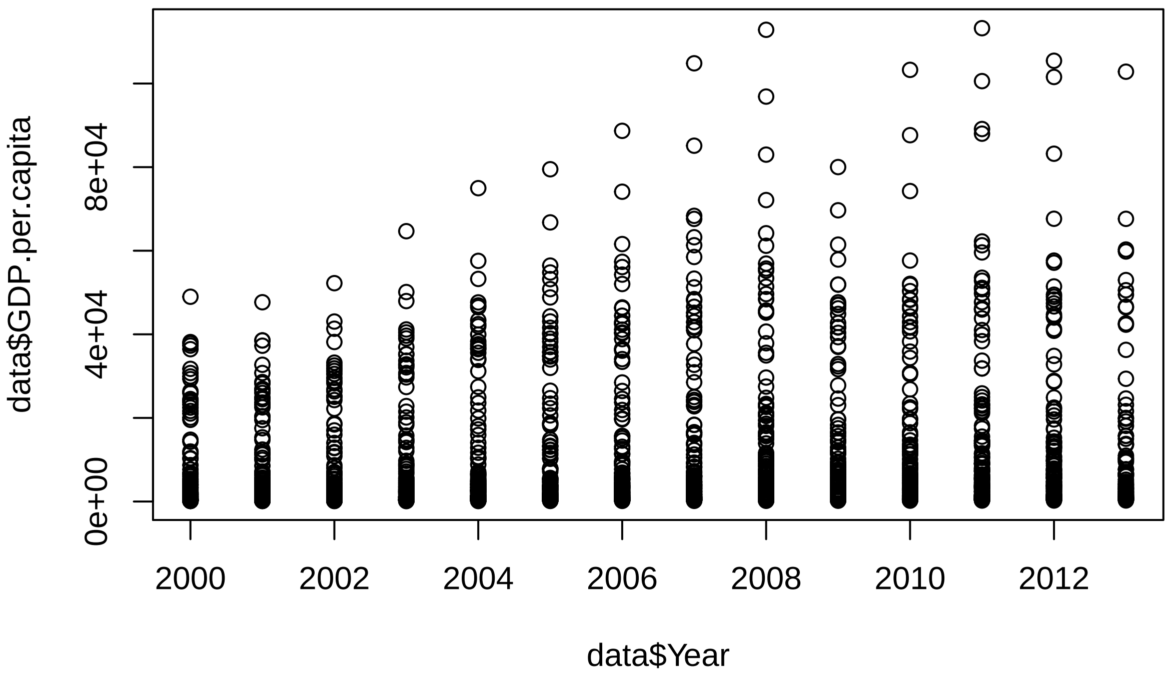
```r
par(mfrow=c(1,1))

# LE is growing slowly each year
boxplot(data$LE~data$Year)
```

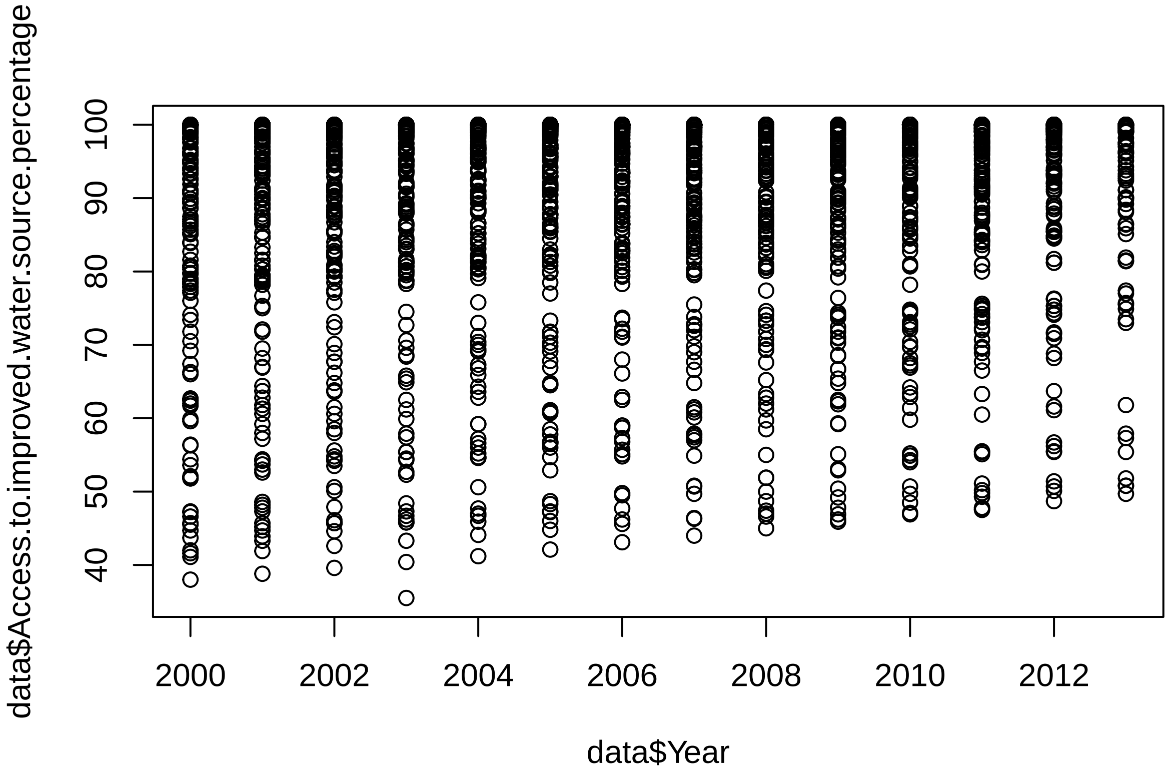


Appendix D Plots of Variables with High Coorelation versus Year

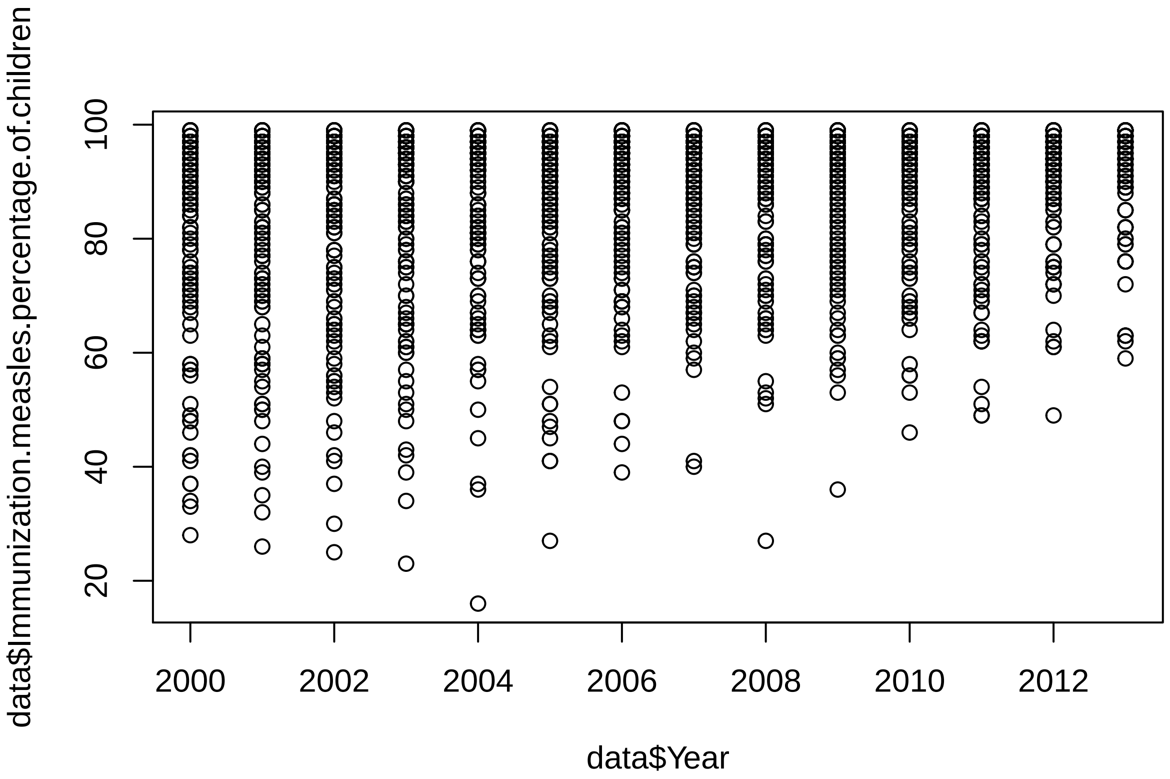
```
plot(data$Year, data$GDP.per.capita)
```



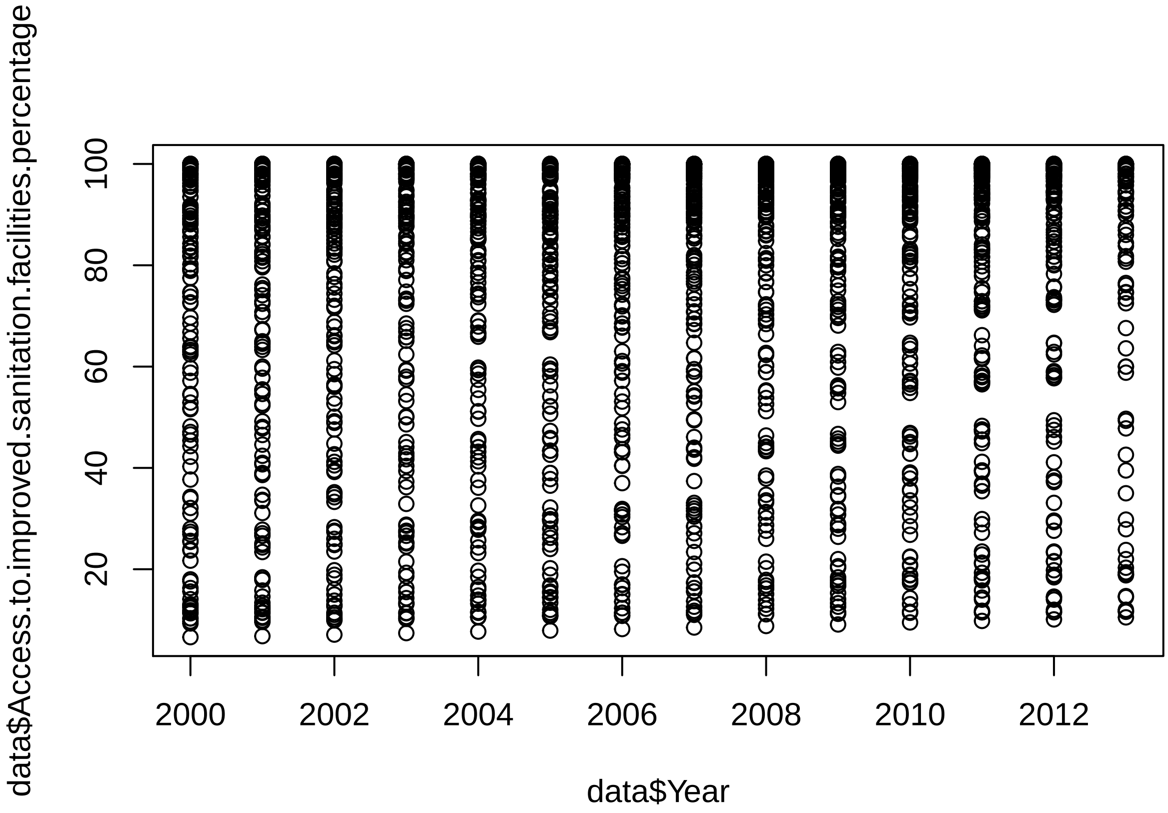
```
plot(data$Year, data$Access.to.improved.water.source.percentage)
```



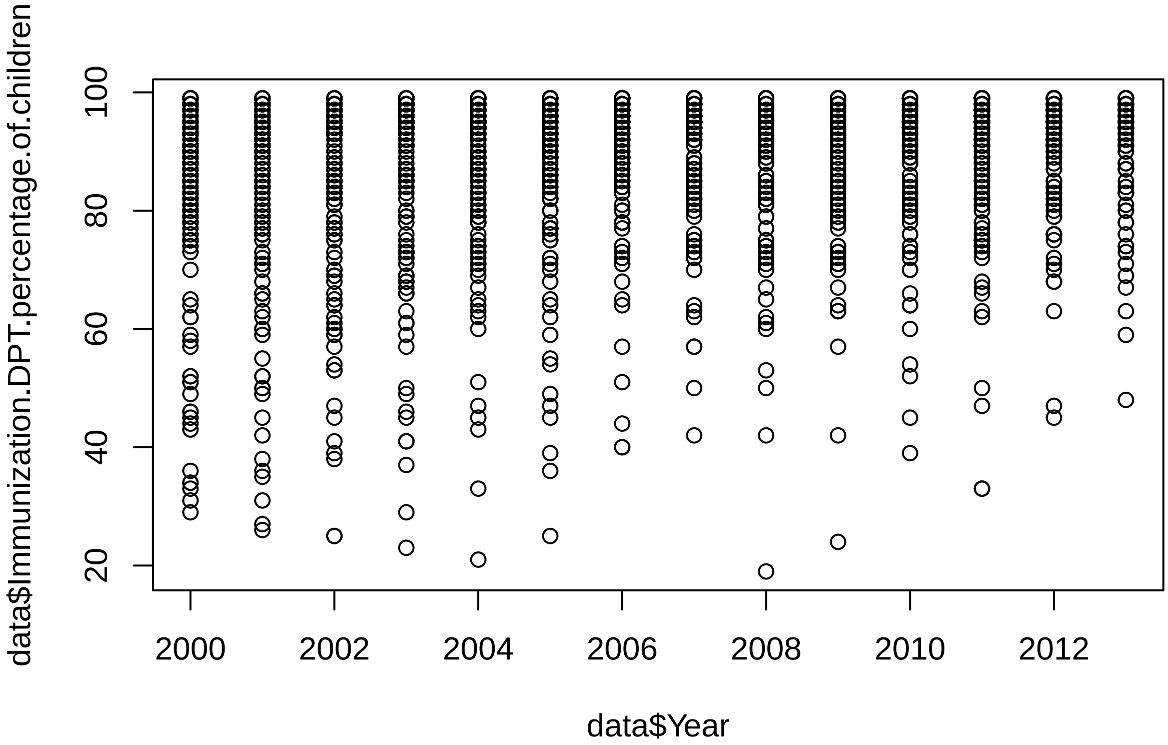
```
plot(data$Year, data$Immunization.measles.percentage.of.children)
```



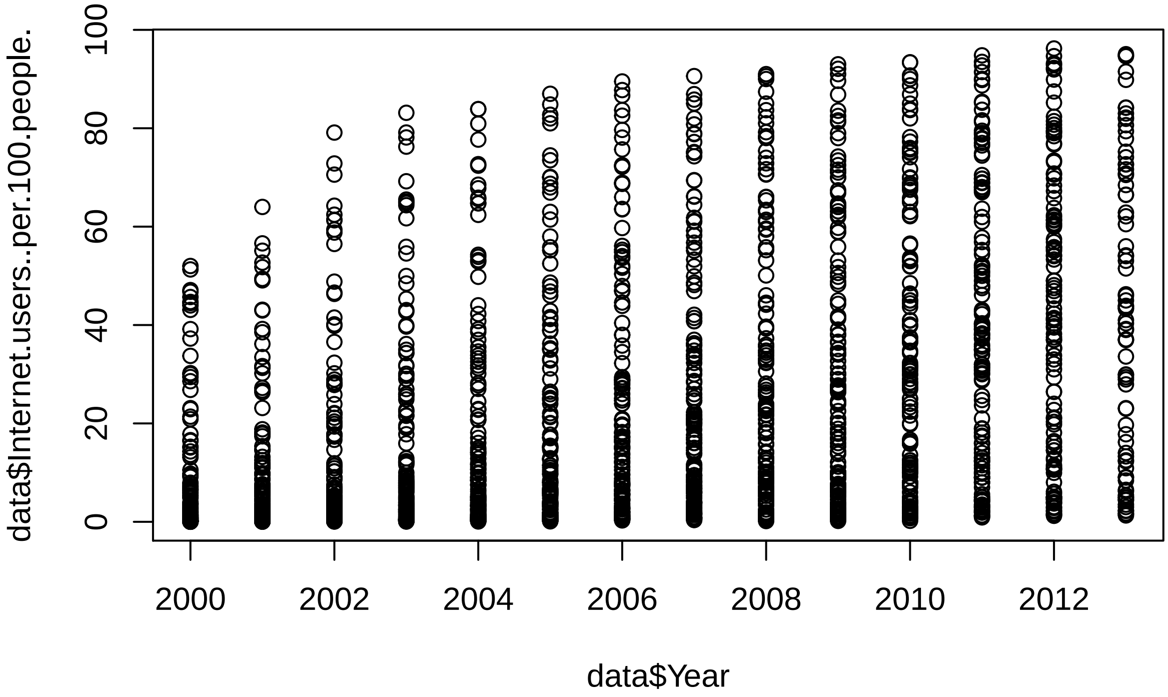
```
plot(data$Year, data$Access.to.improved.sanitation.facilities.percentage)
```



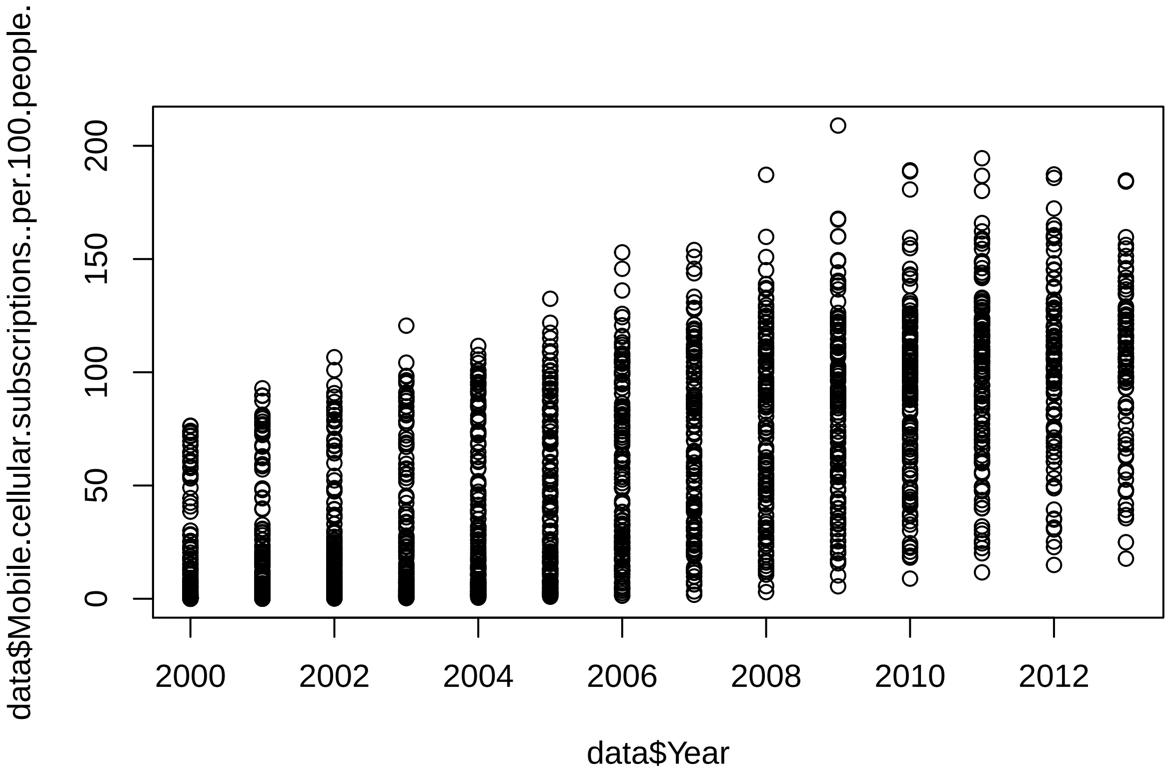
```r
plot(data$Year, data$Immunization.DPT.percentage.of.children)
```



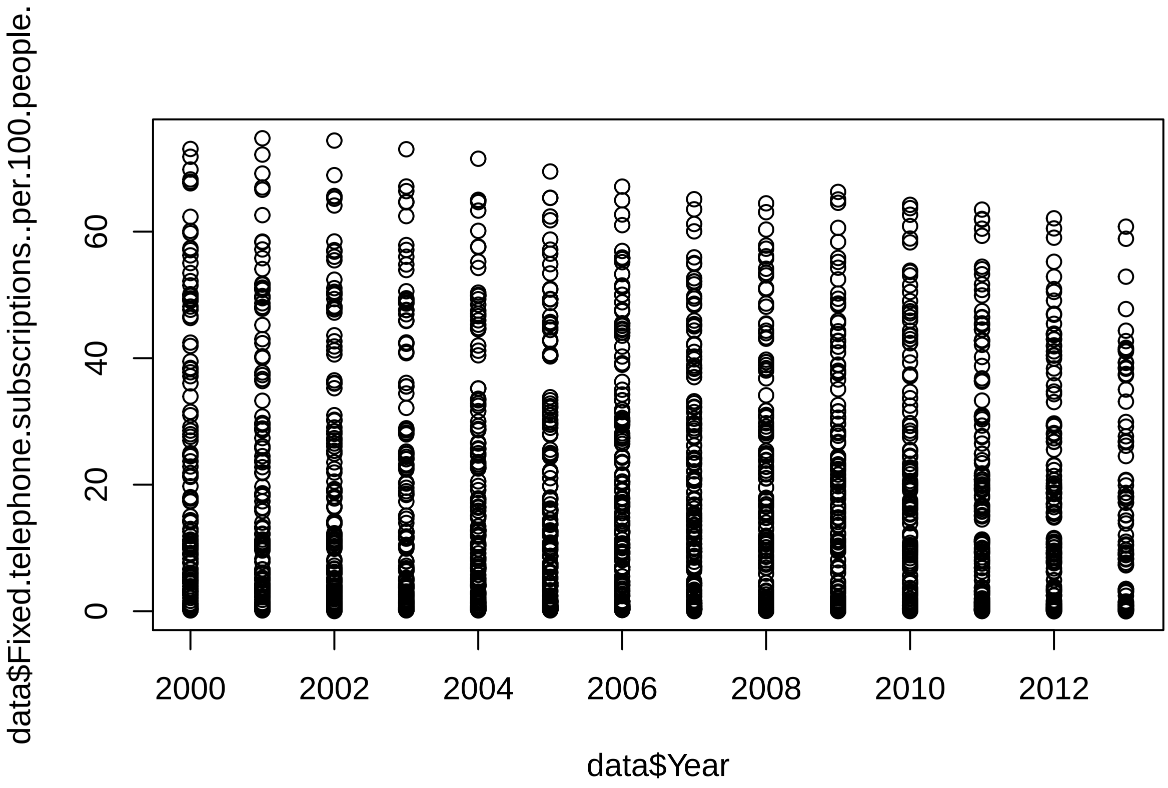
```
plot(data$Year, data$Internet.users..per.100.people.)
```



```
plot(data$Year, data$Mobile.cellular.subscriptions..per.100.people.)
```



```
plot(data$Year, data$Fixed.telephone.subscriptions..per.100.people.)
```


data$Fixed.telephone.subscriptions..per.100.people.
data$Year