Third Assignment

Introduction

The main goal of this third assignment is to gather web based data from at least two sources, merge the data sets, conduct basic descriptive and some inferential statistics on the data to address a relevant research question and briefly describe the results including with dynamically generated tables and figures.

Hypothesis and Methodology

Our hypothesis is the follow **there is a direct correlation between Internet accessibility and student's academic ability and economic growth**.

To prove this hypothesis we plot these variables of all countries that we have data, which have different internet accessibility rate, student's academic ability and economic growth. Then, we have analyzed the correlation between these variables.

Data gathering and merging process

We have gathered different datasets from two different data sources. The first dataset is from **the World Bank**, and the second from **the Quality of Government Institute (QOG)**, which is part of the University of Gothenburg. Both datasets are open and can be found in their respective webpages.

In order to produce a tangible statistical analysis we have cleaned and merged the relevant data of both datasets. Then, we have generate tables and figures to relate the data with the aim of addressing, as well as possible, our researching question to begin to determine whether our hypothesis is correct or not.

Tables and Figures

In this section, we have created three different plots in order to expose in a visual way the data that we have selected to answer our research question. Two of these plots are related to education and Internet accessibility.

In order to display, as best as possible, our data we have chosen different figures to make understandable the correlation between two different data.

Each correlation has been displayed with an axis with two vectors, each one representing a different variable (either Internet Users, GDP per capita, Average of Schooling Years or Quality of Education). In this axis, the tendency in the correlation between the two different variables it is represented by points, each one representing a different country. Altogether, it gives a clear picture of whether there is a tendency or not that confirms the hypothesis of our researching question.
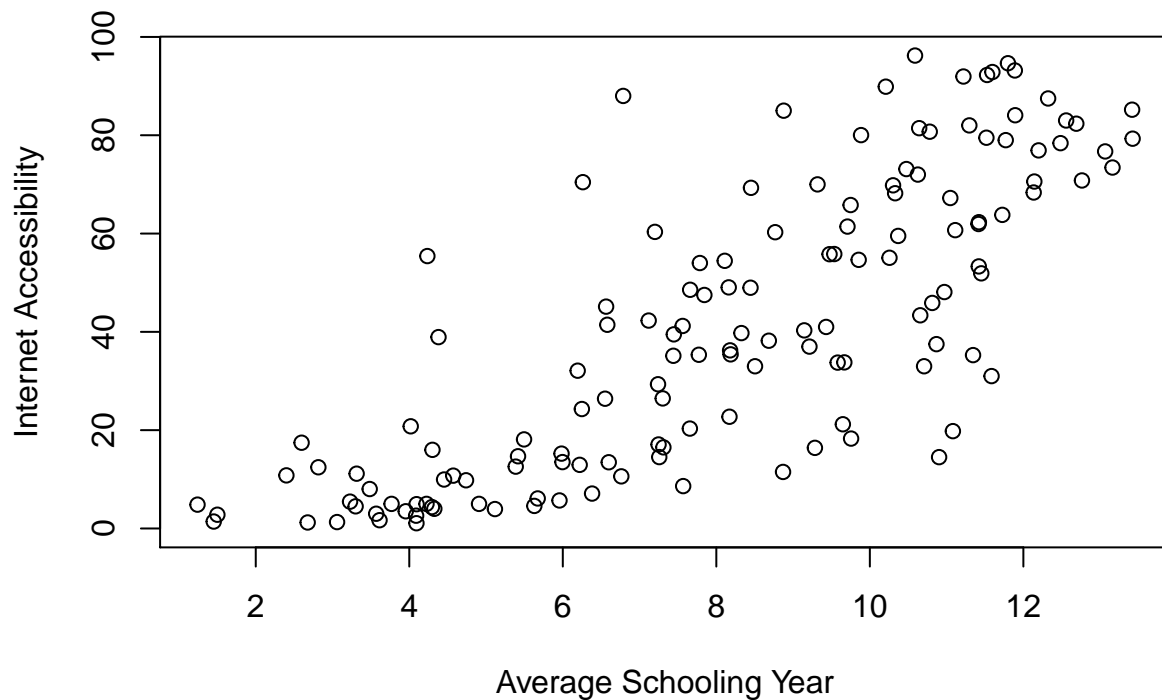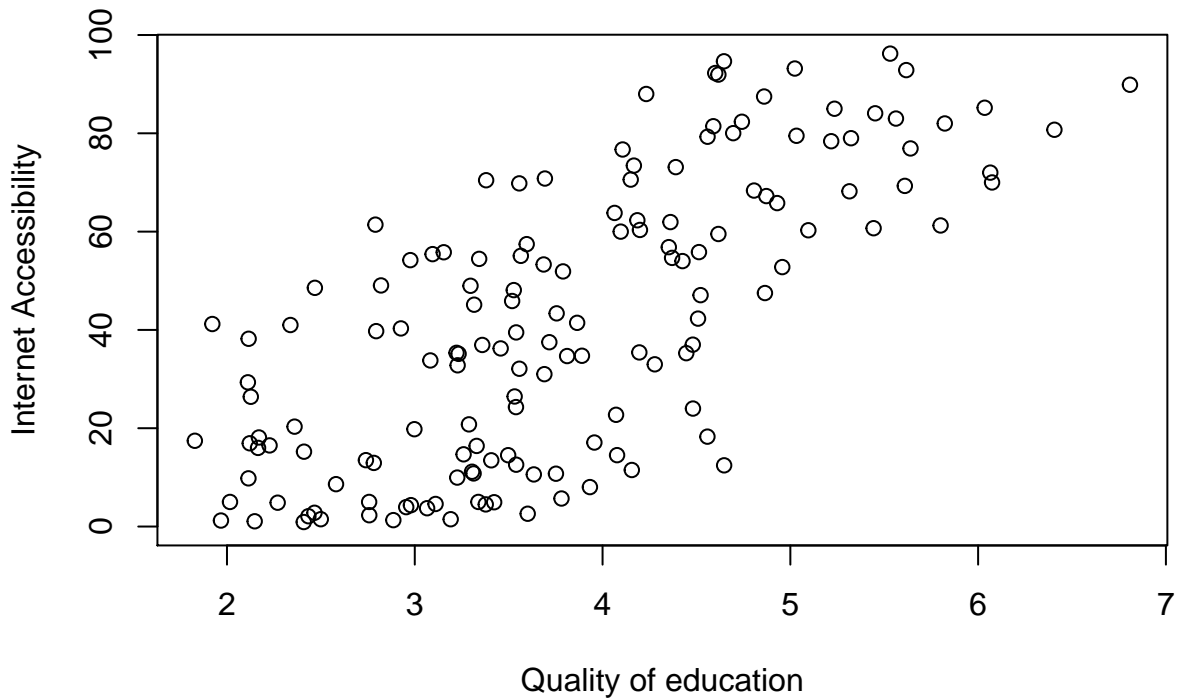
Education and Internet Accessibility

Data Source

- QOGData: http://qog.pol.gu.se/data/datadownloads/qogstandarddata

We downloaded The QoG Basic Cross-Section Data (version January 2016), and picked the following three datas up to create the new data set: - internetuse Internet users (per 100 people) - Average Schooling Years, Female and Male (25+) - Quality of the educational system

Plots

In the first one there is a comparison between the quality of the education and Internet access and the second one between average schooling year and Internet users of all countries in 2016.

These plots show the strong correlation between internet accessibility and average schoolong year and quality of education system.

Conducting hypothesis tests

To analyzing whether the correlation is statistically significant or not, conductig hypothesis test.

```
##
##  Pearson's product-moment correlation
##
## data:  Q and I
## t = 12.396, df = 146, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6271975 0.7865483
## sample estimates:
##       cor
## 0.7160782
```

```
##
##  Pearson's product-moment correlation
##
## data:  Y and I
## t = 15.355, df = 139, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7225957 0.8473676
## sample estimates:
##       cor
## 0.7931617
```

The both result shows P-value = 0, so we can reject the null hypothesis that the true correlatio is equal to zero, and conclude there is a strong correlation between internet accesibility and average schoolong year and quality of education system.

- internet accessibility and average schoolong year: 0.72
- internet accessibility and quality of education system: 0.80
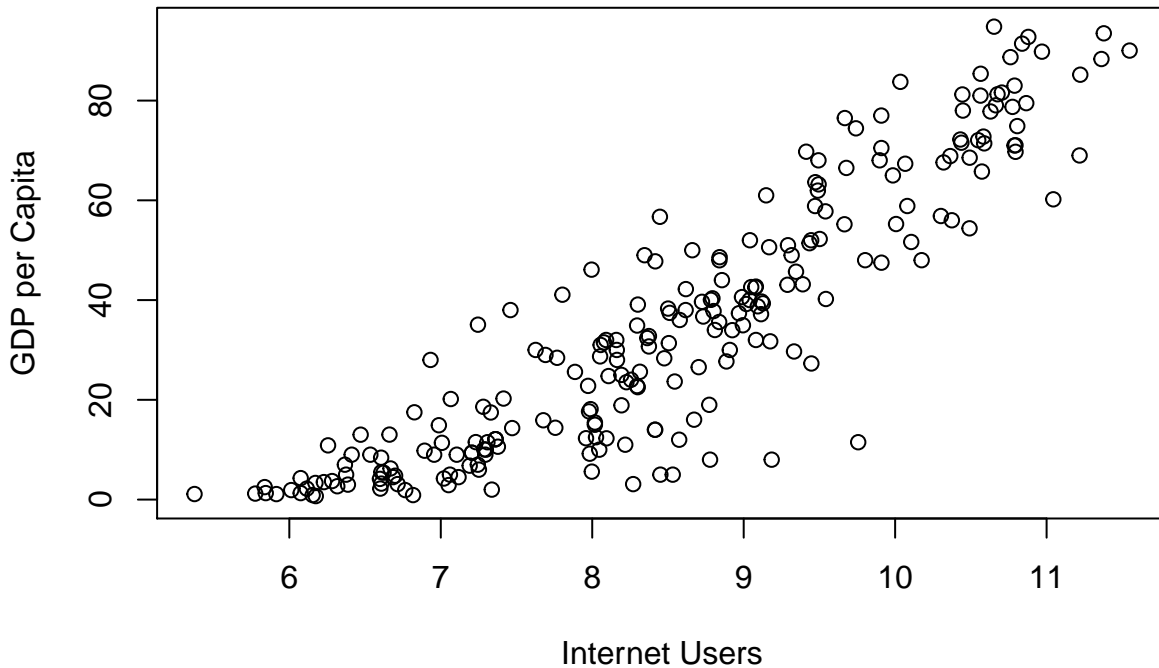
GDP and Internet Accessibility

Data Sorces

- The World Bank

We picked the R package(WDI) and use the following two datas. - Internet users (per 100 people, 2011) : http://data.worldbank.org/indicator/IT.NET.USER.P2 - GDP per capita (constant 2000 US$, 2011)

Conducting basic descriptive statistics

This plot focuses its attention on the GDP and makes and compares GDP per capita and Internet users of all countries in 2011.

The result shows a strong correlation between internet accessibility and GDP per capita.

Conducting hypothesis tests

To analyzing whether the correlation is statistically significant or not, conductig hypothesis test.

```
##
##  Pearson's product-moment correlation
##
## data:  G and I2
## t = 20.74, df = 232, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7558780 0.8467173
## sample estimates:
##       cor
## 0.8059925
```

The result shows P-value = 0, so we can reject the null hypothesis that the true correlatio is equal to zero, and conclude there is a strong correlation between internet accesibility and GDP per capita (0.81).

Conclusion

The results obtained in the three different figures have been similar. **We can observe that all the data introduced in every different figure, that is all the points, draw a clear upward curve tendency, and correlatons are statistically significant**. Therefore, we can conclude our hypothesis is true.

However, the drowned upward curve is not exactly the same in all three cases. The upward curve between Internet Accessibility and Quality of Education is far more dispersed than GDP per capita and Internet users. The latter is much more uniform with just a few points clearly away from the main tendency of the curve. Likewise, in the second figure it can be appreciated that the uniformity of the upward curve it is between the first and the third figure.