index.Rmd

Introduction

The main goal of our reserch is finding variables which affect the student's ability generally through comparison of test scores from differend countries. Especially, we will focus on how inerenet accessibility bring effects on student's ability. Recently, some reserchers found that internet accessibility might bring reverse afffect on student's ability because they spend more time on net surfing and gathering information through internet without thinking.

Methodology

To analyze our research theme, we used the Data from Programme for International Student Assessment(PISA) that conducts assesment about the ..... We picked up the result of **Mathmatics, Reading and Problem Solving** that is conductet in 2012. Also, we picked the result of 2003 Problem solving as well, which is a direct assesment of life competencies that apply across different areas of the school curriculum. This data is benefitial to analyze student's ability that is not measured by academic ability.

Picked up coutries are depended on PISA data avairability,.....

Then, we have analyzed the correlation between these variables.

Data gathering and merging process

The first dataset is from **PISA**, and the second from **the World Bank**. Both datasets are open and can be found in their respective webpages.

Data Source

1. PISA: We downloaded .... and picked the following three datas up to use as dependent variables:

Variable	Variable Name	Description
DV	math	Mathmatics mean score(2012)
DV	reading	Readind mean score(2012)
DV	ps	Prolem Solving mean score(2012)

2. The World Bank: Taking aveirability into account, we picked the following variables up as explanetory variables from the World Bank Data.

Variable	Variable Name	Description
IV	GDPperc	GDP per Capita (current US$)
IV	expend	Government expenditure on education, total (% of GDP)
IV	pop	Population, total
IV	popd	Population density (people per sq. km of land area)

Variable	Variable Name	Description
IV	rteacher	Primary school pupil-teacher ratio is the average number of pupils per teacher in primary school
IV	eyear	Number of years that children are legally obliged to attend school
IV	internet	internet users (per 100 people). Internet users are individuals who have used the Internet (from any location) in the last 12 months. Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.
IV	mobile	Mobile cellular subscriptions (per 100 people)

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.

Multiple Regression

```
##
## Call:
## lm(formula = math ~ log(GDPperc) + log(pop) + popd, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -133.704  -26.788    2.763   23.346  112.728
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.458e+02  8.791e+01   1.658   0.1026
## log(GDPperc) 2.967e+01  5.794e+00   5.120 3.51e-06 ***
## log(pop)     1.642e+00  3.249e+00   0.505   0.6152
## popd         4.191e-03  2.112e-03   1.984   0.0519 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.89 on 59 degrees of freedom
## Multiple R-squared:  0.3877, Adjusted R-squared:  0.3566
## F-statistic: 12.46 on 3 and 59 DF,  p-value: 2.053e-06
```

```
##
## Call:
## lm(formula = math ~ log(GDPperc) + log(pop) + popd, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -133.704  -26.788    2.763   23.346  112.728
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.458e+02  8.791e+01   1.658   0.1026
```

```
## log(GDPperc) 2.967e+01  5.794e+00   5.120 3.51e-06 ***
## log(pop)     1.642e+00  3.249e+00   0.505   0.6152
## popd         4.191e-03  2.112e-03   1.984   0.0519 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.89 on 59 degrees of freedom
## Multiple R-squared:  0.3877, Adjusted R-squared:  0.3566
## F-statistic: 12.46 on 3 and 59 DF,  p-value: 2.053e-06
```

```
##
## Call:
## lm(formula = math ~ log(GDPperc) + log(pop) + popd + rteacher +
##     eyear + expend, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -136.710  -18.988    3.514   28.777   65.504
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  216.904042 156.559363   1.385   0.1773
## log(GDPperc)  22.703295   9.524314   2.384   0.0244 *
## log(pop)       7.587348   6.643184   1.142   0.2634
## popd           0.004735   0.002787   1.699   0.1008
## rteacher      -4.034771   2.615527  -1.543   0.1346
## eyear         -8.544046   5.698842  -1.499   0.1454
## expend        11.785467   7.097490   1.661   0.1084
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.82 on 27 degrees of freedom
##   (29 observations deleted due to missingness)
## Multiple R-squared:  0.4505, Adjusted R-squared:  0.3284
## F-statistic: 3.689 on 6 and 27 DF,  p-value: 0.008318
```

```
##
## Call:
## lm(formula = reading ~ log(GDPperc) + log(pop) + popd + rteacher +
##     eyear + expend, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -111.950  -16.008    3.598   22.750   50.636
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  187.641002 129.801722   1.446   0.1598
## log(GDPperc)  20.996377   7.896508   2.659   0.0130 *
## log(pop)       7.919981   5.507794   1.438   0.1619
## popd           0.003439   0.002311   1.488   0.1483
## rteacher      -2.787892   2.168506  -1.286   0.2095
## eyear         -7.205603   4.724850  -1.525   0.1389
## expend        13.659393   5.884454   2.321   0.0281 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.99 on 27 degrees of freedom
##   (29 observations deleted due to missingness)
## Multiple R-squared:  0.4628, Adjusted R-squared:  0.3435
## F-statistic: 3.877 on 6 and 27 DF,  p-value: 0.006411
```

```
##
## Call:
## lm(formula = reading ~ internet + mobile + log(GDPperc) + log(pop) +
##     popd + rteacher + eyear + expend, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -72.220  -8.297   1.841  18.542  46.734
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  198.843065 123.844972   1.606 0.120925
## internet       1.959607   0.519093   3.775 0.000881 ***
## mobile         0.383294   0.274898   1.394 0.175492
## log(GDPperc)  -2.676702   9.207544  -0.291 0.773673
## log(pop)      10.655010   4.537883   2.348 0.027090 *
## popd           0.001603   0.003215   0.499 0.622419
## rteacher      -0.349618   1.873351  -0.187 0.853459
## eyear         -5.995144   3.856552  -1.555 0.132626
## expend         2.735772   5.527635   0.495 0.624974
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.82 on 25 degrees of freedom
##   (29 observations deleted due to missingness)
## Multiple R-squared:  0.6726, Adjusted R-squared:  0.5678
## F-statistic: 6.419 on 8 and 25 DF,  p-value: 0.0001491
```

```
##
## Call:
## lm(formula = ps ~ log(GDPperc) + log(pop) + popd + rteacher +
##     eyear + expend, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.338  -5.947   4.057   8.202  44.884
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  214.289111 101.400223   2.113  0.05064 .
## log(GDPperc)  29.944236   7.403851   4.044  0.00094 ***
## log(pop)      10.371807   4.617193   2.246  0.03915 *
## popd           0.002725   0.001792   1.520  0.14800
## rteacher      -3.424028   1.530232  -2.238  0.03983 *
## eyear        -16.168661   4.191807  -3.857  0.00139 **
## expend         4.506369   5.685975   0.793  0.43964
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 24.44 on 16 degrees of freedom
##   (40 observations deleted due to missingness)
## Multiple R-squared:   0.76,  Adjusted R-squared:   0.67
## F-statistic: 8.445 on 6 and 16 DF,  p-value: 0.0003067
```

```
##
## Call:
## lm(formula = ps ~ internet + mobile + log(GDPperc) + log(pop) +
##     popd + rteacher + eyear + expend, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.389  -5.489  -0.126  12.811  30.420
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  201.179611 104.075976   1.933   0.0737 .
## internet       1.561277   0.624799   2.499   0.0255 *
## mobile         0.125213   0.222887   0.562   0.5832
## log(GDPperc)  16.217979   8.606442   1.884   0.0804 .
## log(pop)       8.666553   4.288148   2.021   0.0628 .
## popd           0.002592   0.002470   1.049   0.3118
## rteacher      -1.065455   1.667119  -0.639   0.5331
## eyear        -11.385740   4.210679  -2.704   0.0171 *
## expend        -2.767786   5.850844  -0.473   0.6435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.7 on 14 degrees of freedom
##   (40 observations deleted due to missingness)
## Multiple R-squared:  0.8344, Adjusted R-squared:  0.7397
## F-statistic: 8.817 on 8 and 14 DF,  p-value: 0.0002568
```

Analysis

According to the result, in the secound regressio, these variables are statistically significant. - Math:GDP per capita - Reading: GDP per capita, expenditure - Problem solving: GDP per capita, population, pupills-teacher rate and duraion educational year in elementary

In the third regression, GDP per capita is not statistically significant any more. Instead of GDP per capita, the number of internet users and population become statistically significant.

Especially, the number of internet users is significant at 0.1% level in math and reading. We can see it bring plus effect on the both score.However, regarding the problem solving, internet is statistically significant only at 5 % level and the effect is weaker than other 2 scores.

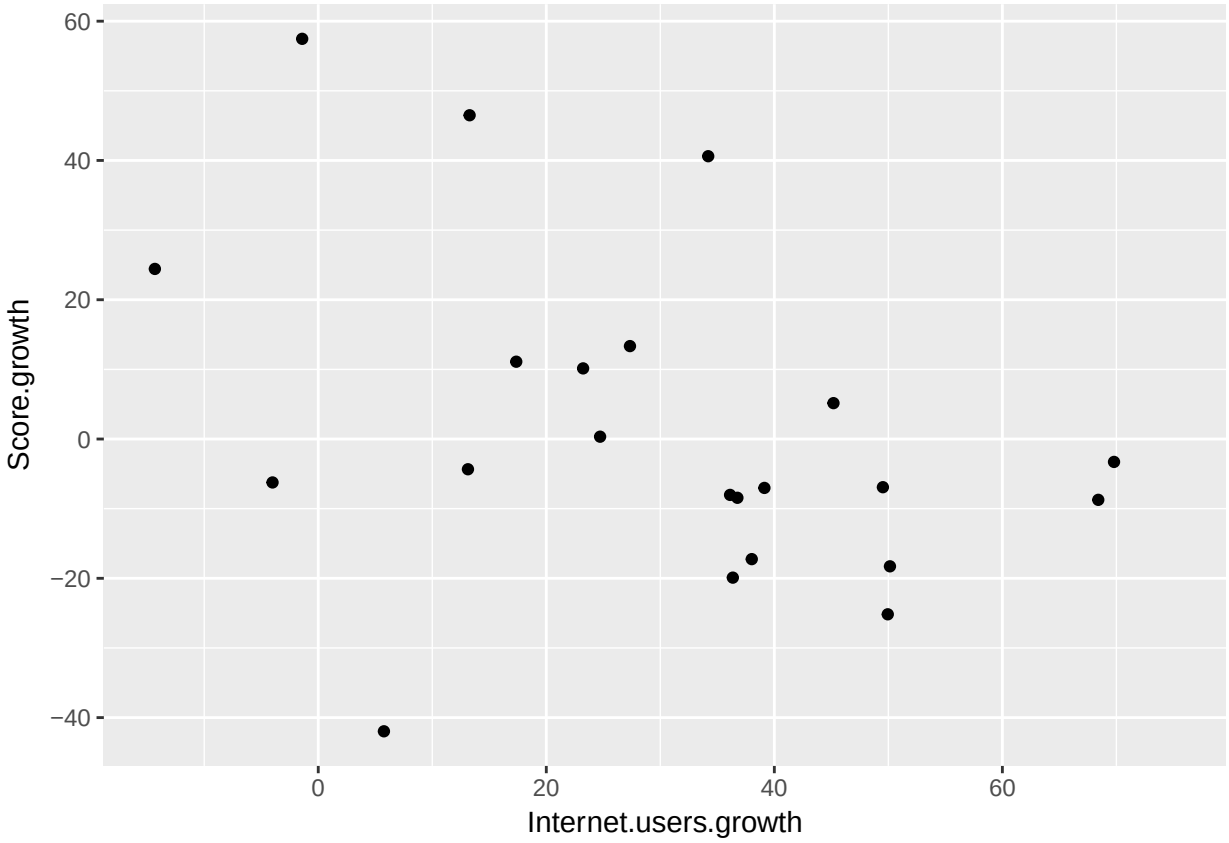
In addition, regarding GDP per capita of math ane reading, it bring the minues effect on the both scores after the variable internet and mobile are added, though it still bring the plus effect on problem solving. Even though it is not statistically significant for math and reading, but it might possible to infere if the internet accessibility would be the same level, economically strong country's math and reading scores might be lower than economically weak countries.

Also, we noticed that about the academic ability (math and reading,) internet accessibility would bring the some impact on the scores. However, it would bring less impact on the problem solving ability.

Actually, there is some discussion that internet accessibility might bring reverse effect on student's thinkig ability. Therefore, next we focus on the problem solving score.

Correlation between scpre growth

```
## Warning: Removed 33 rows containing missing values (geom_point).
```




```
##
## Call:
## lm(formula = Internet.users.growth ~ Score.growth, data = pisa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.780  -8.265   1.494  12.432  38.128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   30.4856     4.4649   6.828 1.23e-06 ***
## Score.growth  -0.3587     0.1917  -1.872   0.0759 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.9 on 20 degrees of freedom
##   (33 observations deleted due to missingness)
## Multiple R-squared:  0.1491, Adjusted R-squared:  0.1065
## F-statistic: 3.503 on 1 and 20 DF,  p-value: 0.07594
```

The result shows that we cannot find any correlation between internet accessibility growth and score growth. . . ..