Final_project.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:

Dependent 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.

Indivendent 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)

Indivendent 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)

Cleaned and Merged

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

Summary statistics of the variables

Statistic N Mean St. Dev. Min Max

math 63 469.743 52.225 368.103 573.468
reading 63 471.675 45.780 384.151 544.600
ps 42 485.544 44.181 399.166 562.421
GDPperc 63 31,350.570 28,867.050 1,754.548 149,160.800 pop 63 36,014,700.000 60,090,250.000 36,791 314,102,623 popd 63 635.244 2,625.005 2.959 18,654.280 rteacher 51 15.061 4.335 7.444 28.016
eyear 62 10.226 1.750 6 14
expend 39 4.897 1.270 2.922 7.656
internet 63 66.580 18.362 14.520 96.210
mobile 63 128.770 33.182 79.568 289.782

Table:Regression Estimates of Problem Solving Score

	Dependent variable:		
	Problem Solving Score		
	(1)	(2)	(3)
log(GDPperc)	30.639***	29.944***	16.218*
	(7.175)	(7.404)	(8.606)
log(pop)	5.746	10.372**	8.667*

(3.764) (4.617) (4.288)
popd 0.004* 0.003 0.003
(0.002) (0.002) (0.002)
rteacher -3.424** -1.065
(1.530) (1.667)
eyear -16.169*** -11.386**
(4.192) (4.211)
expend 4.506 -2.768
(5.686) (5.851)
internet 1.561**
(0.625)
mobile 0.125
(0.223)
(intercept) 77.963 214.289* 201.180*
(88.520) (101.400) (104.076)

Observations 42 23 23
R2 0.453 0.760 0.834
Adjusted R2 0.410 0.670 0.740
Residual Std. Error 33.948 (df = 38) 24.438 (df = 16) 21.703 (df = 14)
F Statistic 10.481*** (df = 3; 38) 8.445*** (df = 6; 16) 8.817*** (df = 8; 14) ==========
Note: *p<0.1;* ***p<0.05;*** p<0.01

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.

Table 3: Table:Regression Estimates of Math Score

	Dependent variable:		
	Math Score		
	(1)	(2)	(3)
log(GDPperc)	29.666***	22.703**	−7.883
	(5.794)	(9.524)	(10.680)
log(pop)	1.642	7.587	11.018**
	(3.249)	(6.643)	(5.264)
popd	0.004*	0.005	0.003
	(0.002)	(0.003)	(0.004)
rteacher		−4.035	−0.956
		(2.616)	(2.173)
eyear		−8.544	−7.042
		(5.699)	(4.473)
expend		11.785	−2.215
		(7.097)	(6.412)
internet			2.522***
			(0.602)
mobile			0.452
			(0.319)
(intercept)	145.762	216.904	240.336
	(87.914)	(156.559)	(143.650)
Observations	63	34	34
R^2	0.388	0.451	0.690
Adjusted R^2	0.357	0.328	0.591
Residual Std. Error	41.891 (df = 59)	45.818 (df = 27)	35.750 (df = 25)
F Statistic	12.455*** (df = 3; 59)	3.689*** (df = 6; 27)	6.963*** (df = 8; 25)
Note:			*p<0.1; **p<0.05; ***p<0.01

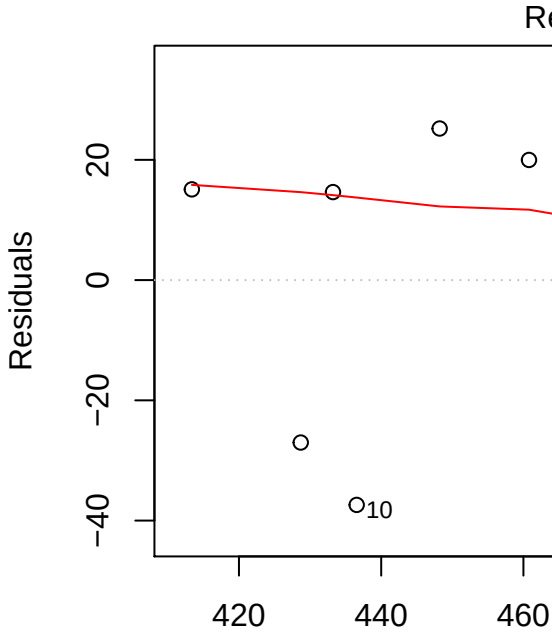
Table 4: Table:Regression Estimates of Reading Score

	Dependent variable:		
	Reading Score		
	(1)	(2)	(3)
log(GDPperc)	27.480***	20.996**	−2.677
	(5.174)	(7.897)	(9.208)
log(pop)	3.107	7.920	10.655**
	(2.901)	(5.508)	(4.538)
popd	0.002	0.003	0.002
	(0.002)	(0.002)	(0.003)
rteacher		−2.788	−0.350
		(2.169)	(1.873)
eyear		−7.206	−5.995
		(4.725)	(3.857)
expend		13.659**	2.736
		(5.884)	(5.528)
internet			1.960***
			(0.519)
mobile			0.383
			(0.275)
(intercept)	146.816*	187.641	198.843
	(78.509)	(129.802)	(123.845)
Observations	63	34	34
R^2	0.365	0.463	0.673
Adjusted R^2	0.332	0.343	0.568
Residual Std. Error	37.409 (df = 59)	37.987 (df = 27)	30.822 (df = 25)
F Statistic	11.284*** (df = 3; 59)	3.877*** (df = 6; 27)	6.419*** (df = 8; 25)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Table:Regression Estimates of Three Modeles

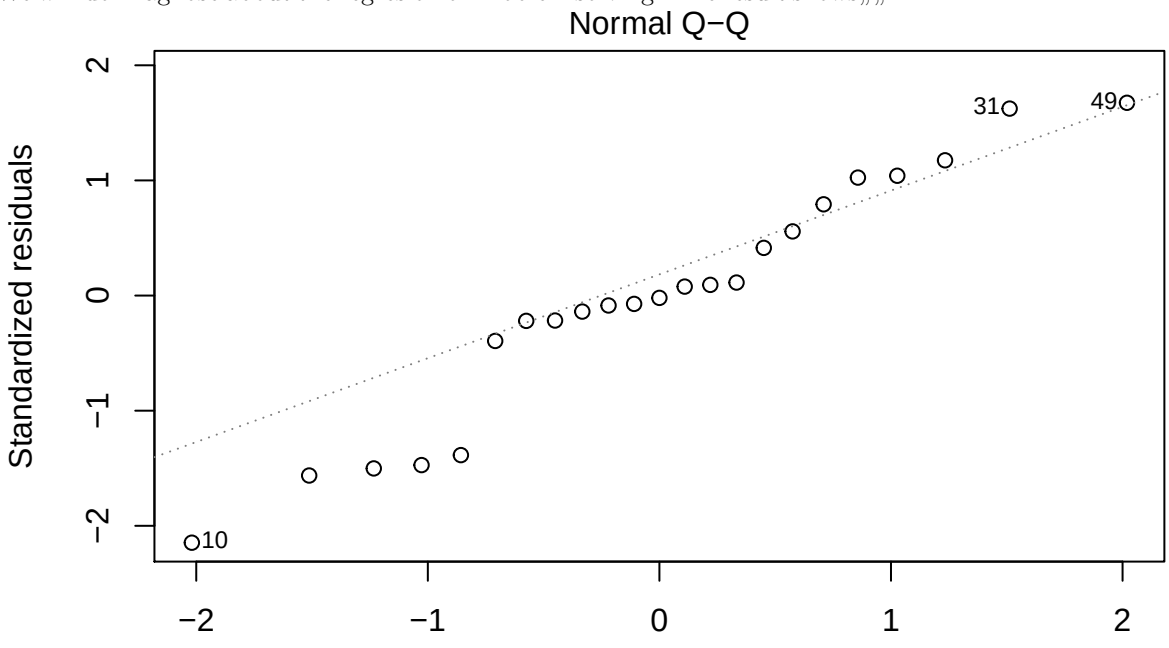
	Dependent variable:		
	math	reading	ps
	(1)	(2)	(3)
log(GDPperc)	−7.883	−2.677	16.218*
	(10.680)	(9.208)	(8.606)
log(pop)	11.018**	10.655**	8.667*
	(5.264)	(4.538)	(4.288)
popd	0.003	0.002	0.003
	(0.004)	(0.003)	(0.002)
rteacher	−0.956	−0.350	−1.065
	(2.173)	(1.873)	(1.667)
eyear	−7.042	−5.995	−11.386**
	(4.473)	(3.857)	(4.211)
expend	−2.215	2.736	−2.768
	(6.412)	(5.528)	(5.851)
internet	2.522***	1.960***	1.561**
	(0.602)	(0.519)	(0.625)
mobile	0.452	0.383	0.125
	(0.319)	(0.275)	(0.223)
(intercept)	240.336	198.843	201.180*
	(143.650)	(123.845)	(104.076)
Observations	34	34	23
R^2	0.690	0.673	0.834
Adjusted R^2	0.591	0.568	0.740
Residual Std. Error	35.750 (df = 25)	30.822 (df = 25)	21.703 (df = 14)
F Statistic	6.963*** (df = 8; 25)	6.419*** (df = 8; 25)	8.817*** (df = 8; 14)
Note:			*p<0.1; **p<0.05; ***p<0.01

Correlation between Problem Solving and Internet

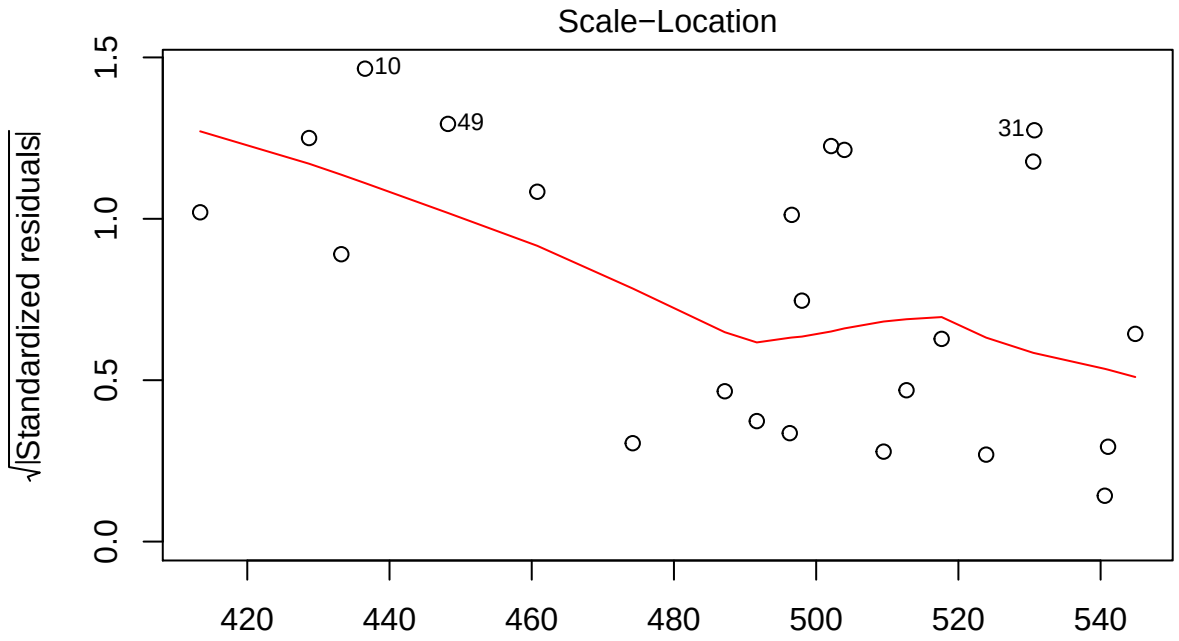


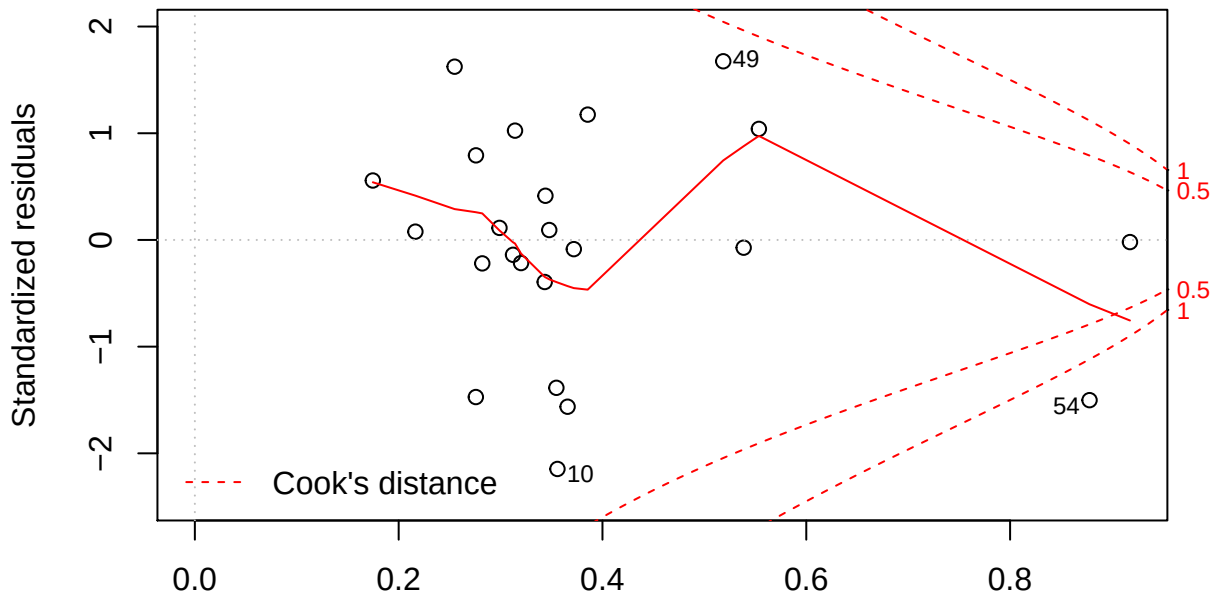
lm(ps ~ log(GDPperc) + log(pop)

We will do Diagnose about the regresion of Problem solving. The result shows„ „



lm(ps ~ log(GDPperc) + log(pop) + popd + rteacher + eyear + expend + intern ...





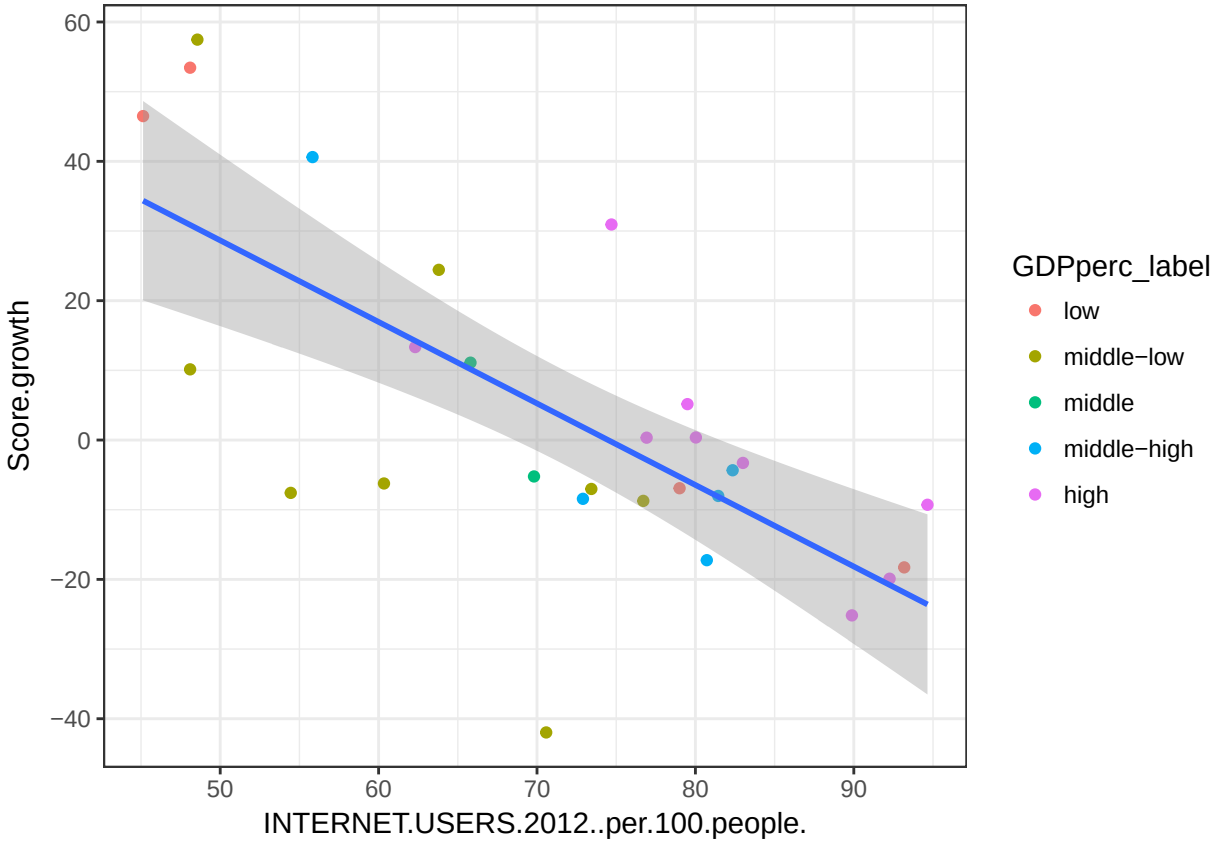
2.5 % 97.5 % (Intercept) -22.041157797 424.400379451 log(GDPperc) -2.241003905 34.676962325 log(pop) -0.530610003 17.863716671 popd -0.002706016 0.007889451 rteacher -4.641069200 2.510159015 eyear -20.416747602 -2.354732539 expend -15.316598142 9.781025773 internet 0.221215890 2.901338214 mobile -0.352832156 0.603257434 Min. 1st Qu. Median Mean 3rd Qu. Max. 1755 10690 20580 31350 45720 149200

Analysis about score growth

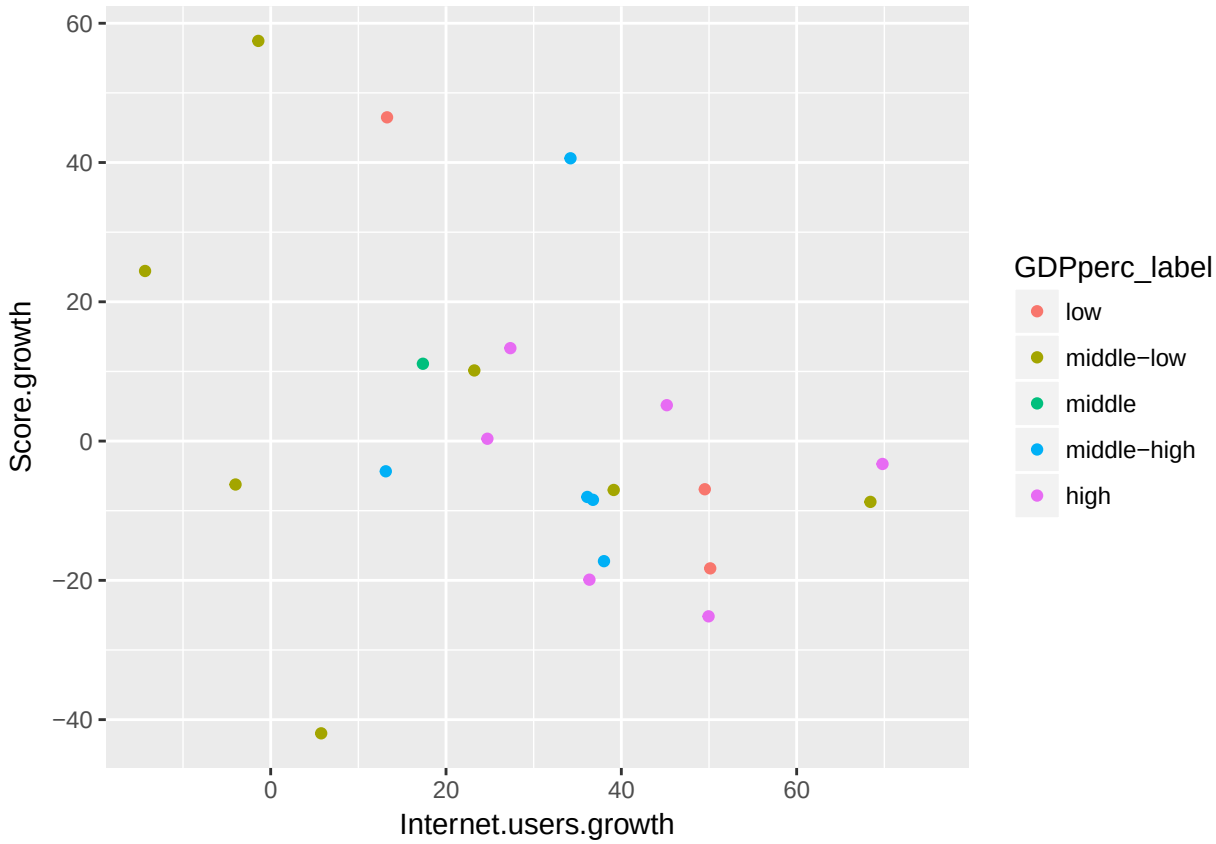
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3989   13140   24450   29960   44730  101600
```

```
## Warning: Removed 13 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 13 rows containing missing values (geom_point).
```




```
## Warning: Removed 19 rows containing missing values (geom_point).
```


60
40
20
0
−20
−40
Score.growth
0
20
40
60
Internet.users.growth
GDPperc_label
low
middle−low
middle
middle−high
high