PISA2012_wenchit

Abstract

In this research, I want to explore the education resource distribution in Australia. For insatnce, comparing the education resources between remote and city areas and the discrepancies between private or public schools. Based on the PISA2012 mathematics literacy survey data. Also, I will break down the resource allocation within different states across Australia.

Introduction

- PISA2012: the term PISA stands for Programme for International Student Assessment is a survey that randomly select 15 year-old students as samples for assessment. In 2012, a total of 65 OECD countries and economics and about half a million 15 year-old students participated in the PISA assessment. Generally, the assessed reaults lies in 5 categories: Level 5 are high performers, whereas students who lies below the international standard baseline level 2 are cosidered low performers. In 2012, 775 Australian schools and 14,481 studnets participated in this assessment. To ensure the authenticity, an amount of indigenous students were also sampled.
- Mathematic Lieteracy: in the mathematic literacy domain, the assessment focused on students' ability to solve mathematic problems described in a real-life situation. In PISA2012 framework(OECD2012, p25), it defined mathematic literacy as follows: Mathematic literacy is an individual's capacity to formulate, employ and interpret mathematics in a variety of contexts. It includes reasoning mathematically nd using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that Mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.

6-levels of student performances

According to the technical manual released by PISA, on a performance level of 1~6, students are supposed to be able to:

Level6: conceptualise, generalise and use information.They are capable of advanced mathematical thinking and reasoning; have a mastery of symbolic and formal mathematical operations and relationships; and can formulate and precisely communicate their ndings, interpretations and arguments.

Level5: develop and work with models for complex situations; select, compare and evaluate appropriate problem-solving strategies for dealing with complex problems; work strategically using broad, well-developed thinking and reasoning skills; and re ect on their work and formulate and communicate their interpretations and reasoning.

Level4: work effectively with explicit models for complex, concrete situations; select and integrate different representations, including symbolic representations; utilise their skills and reason with insight; and construct and communicate explanations and arguments.

Level3: execute clearly described procedures, including those that require sequential decisions; select and apply simple problem-solving strategies; interpret and use representations; typically show some ability to handle percentages, fractions and decimal numbers, and to work with proportional relationships; and provide solutions that re ect that they have engaged in basic interpretation and reasoning.

Level2: interpret and recognise situations in contexts that require no more than direct inference; extract relevant information from a single source and make use of a single representational mode; employ basic algorithms, formulas, procedures or conventions to solve problems involving whole numbers; and make literal interpretations of the results.

Level1: answer questions involving familiar contexts where all relevant information is present and the questions are clearly de ned; identify information and carry out routine procedures; and perform actions that are almost always obvious and follow immediately from the given stimuli.

Below 1: not demonstrate even the most basic types of mathematical literacy that PISA measures. These students are likely to be seriously disadvantaged in their lives beyond school.

Overview of Australian Education System:

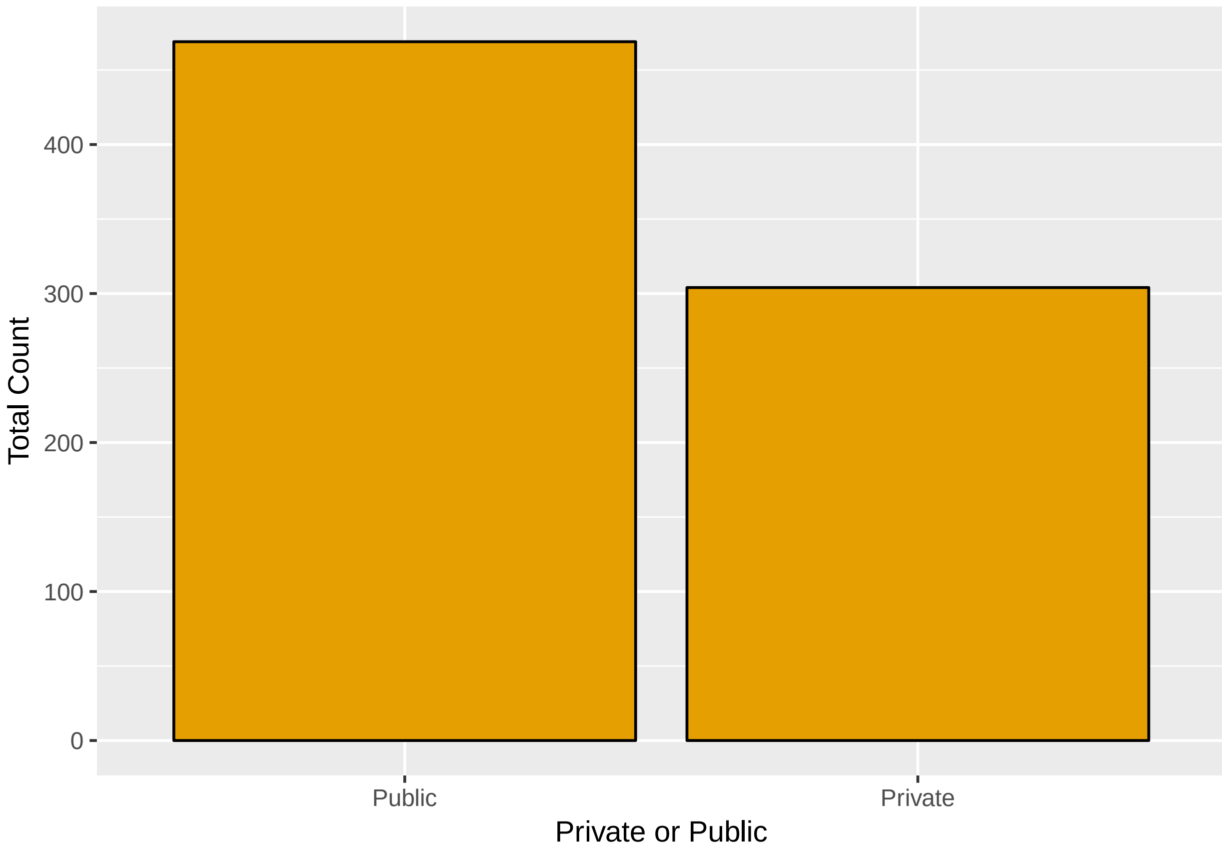
Methods

Linear Mixed Model: including linear regression model and ANOVA test:

Results

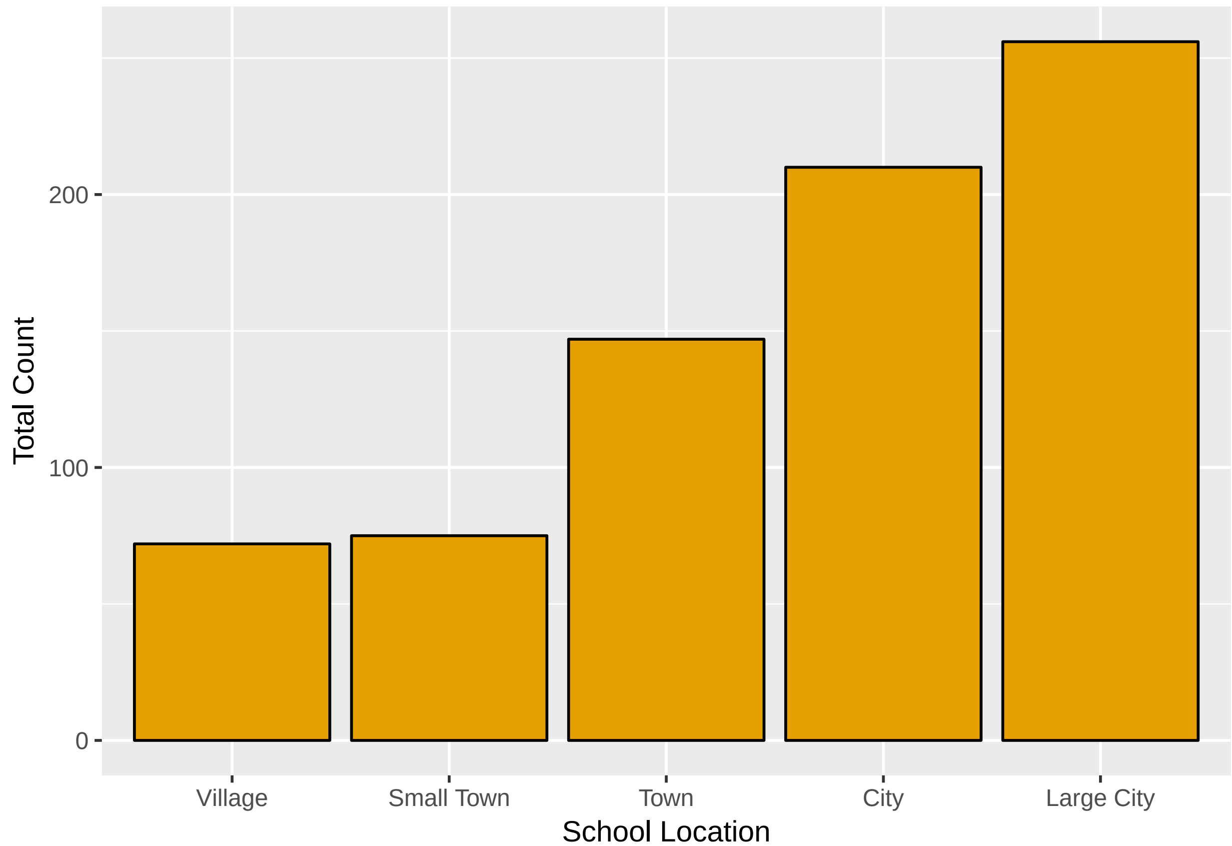
The following graph shows the amount of private and public schools in Australia that participated:

```
library(data.table)
library(ggplot2)
library(PISA2012lite)
data("school2012")
setDT(school2012)

ggplot(school2012[NC == "Australia" & !is.na(SC01Q01)], aes(SC01Q01))+geom_bar(fill="#E69F00", colour="
```



The following is a graph about the distribution of schools in Australia that participated:

```
library(data.table)
library(ggplot2)
library(PISA2012lite)
data("school2012")
setDT(school2012)

ggplot(school2012[NC == "Australia" & !is.na(SC03Q01)], aes(SC03Q01))+geom_bar(fill="#E69F00", colour="
```



The following is a calculation of math literacy scores by schools participated all across Australia and summarise by its house weight and to obtain coefficient in the null model:

```
library(data.table)
library(ggplot2)
library(PISA2012lite)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##     between, first, last

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
data("computerStudent2012")
setDT(computerStudent2012)
```

```r
n <- nrow(computerStudent2012)
computerStudent2012$W_HOUSEWHT <- n * computerStudent2012$W_FSTUWT / sum(computerStudent2012$W_FSTUWT)
#calculate the weighted average of math literacy scores
computerStudent2012[NC == "Australia"] %>%
  group_by(SCHOOLID) %>%
  summarise(avg1 = weighted.mean(PV1MATH, w = W_HOUSEWHT))
```

```
## # A tibble: 775 x 2
##    SCHOOLID     avg1
##      <fctr>    <dbl>
##  1  0000001 562.9773
##  2  0000002 480.2559
##  3  0000003 566.4070
##  4  0000004 429.4231
##  5  0000005 464.8072
##  6  0000006 500.0640
##  7  0000007 484.0204
##  8  0000008 554.4547
##  9  0000009 431.5852
## 10  0000010 600.4370
## # ... with 765 more rows
```

```r
HLM0 <- lmer(PV1MATH ~ (1 | SCHOOLID), data = computerStudent2012[NC == "Australia"],
             weights = W_HOUSEWHT)
summary(HLM0)
```

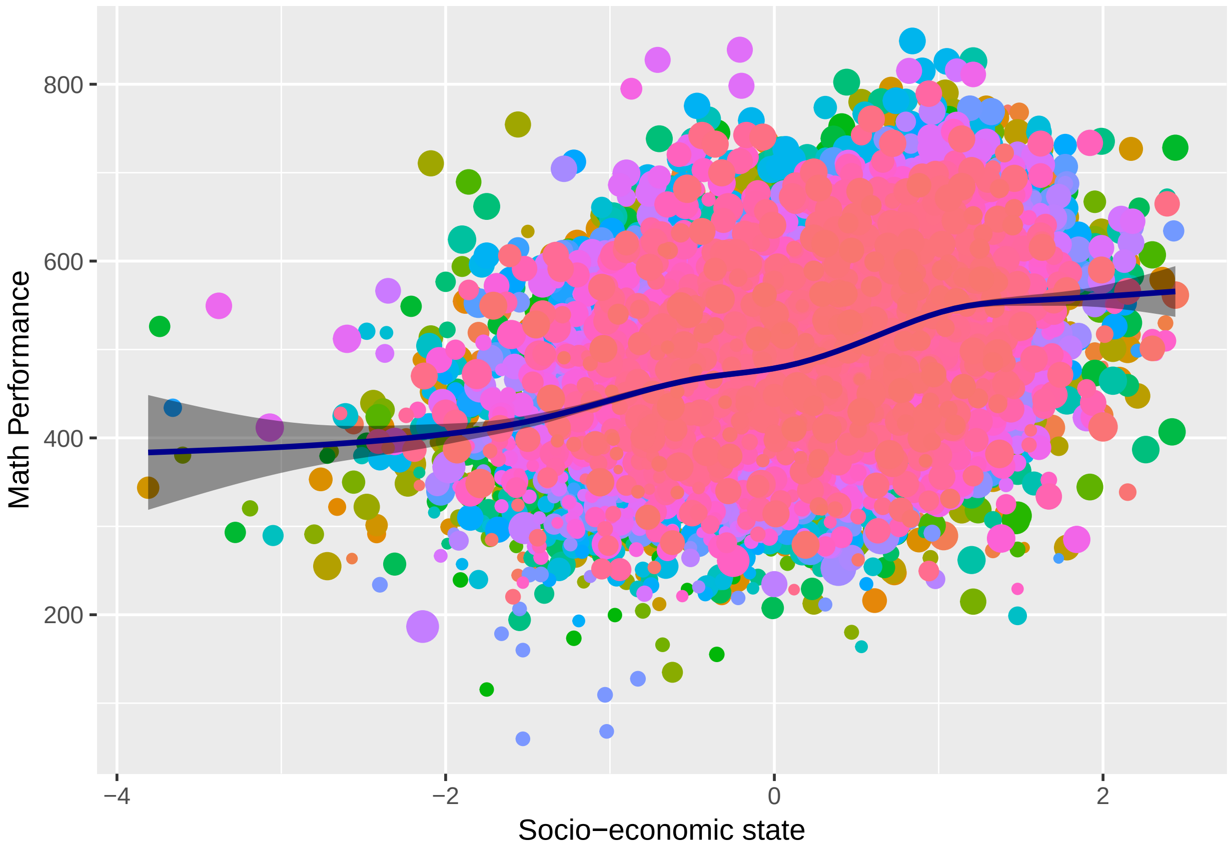
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: PV1MATH ~ (1 | SCHOOLID)
##    Data: computerStudent2012[NC == "Australia"]
## Weights: W_HOUSEWHT
##
## REML criterion at convergence: 174059.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5983 -0.6028 -0.0587  0.5508  4.5737
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SCHOOLID (Intercept) 2598     50.97
##  Residual             1850     43.01
## Number of obs: 14481, groups:  SCHOOLID, 775
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  498.342      2.019   246.8
```

The following cosidering the socio-economic state of the student

```r
library(data.table)
library(ggplot2)
library(PISA2012lite)
library(dplyr)
library(lme4)
data("computerStudent2012")
```

```r
setDT(computerStudent2012)
n <- nrow(computerStudent2012)
computerStudent2012$W_HOUSEWHT <- n * computerStudent2012$W_FSTUWT / sum(computerStudent2012$W_FSTUWT)

ggplot(data = computerStudent2012[NC == "Australia"], aes(x = ESCS, y = PV1MATH, size = W_HOUSEWHT)) +
  geom_point(aes(colour = SCHOOLID)) + geom_smooth(fill="black", colour="darkblue", size=1)+ theme(leger
```

```
## `geom_smooth()` using method = 'gam'

## Warning: Removed 371 rows containing non-finite values (stat_smooth).

## Warning: Removed 371 rows containing missing values (geom_point).
```



Now we want to look into the relationship between schools and math performances:

```r
library(data.table)
library(PISA2012lite)
library("dplyr")
data("computerStudent2012")
setDT(computerStudent2012)
data("school2012")
setDT(school2012)

#filter data
school_view <- school2012[NC =="Australia" & !is.na(SC03Q01)]
school_plot_data <- select(school_view, SC03Q01, SCHOOLID)
#filter data
n <- nrow(computerStudent2012)
```

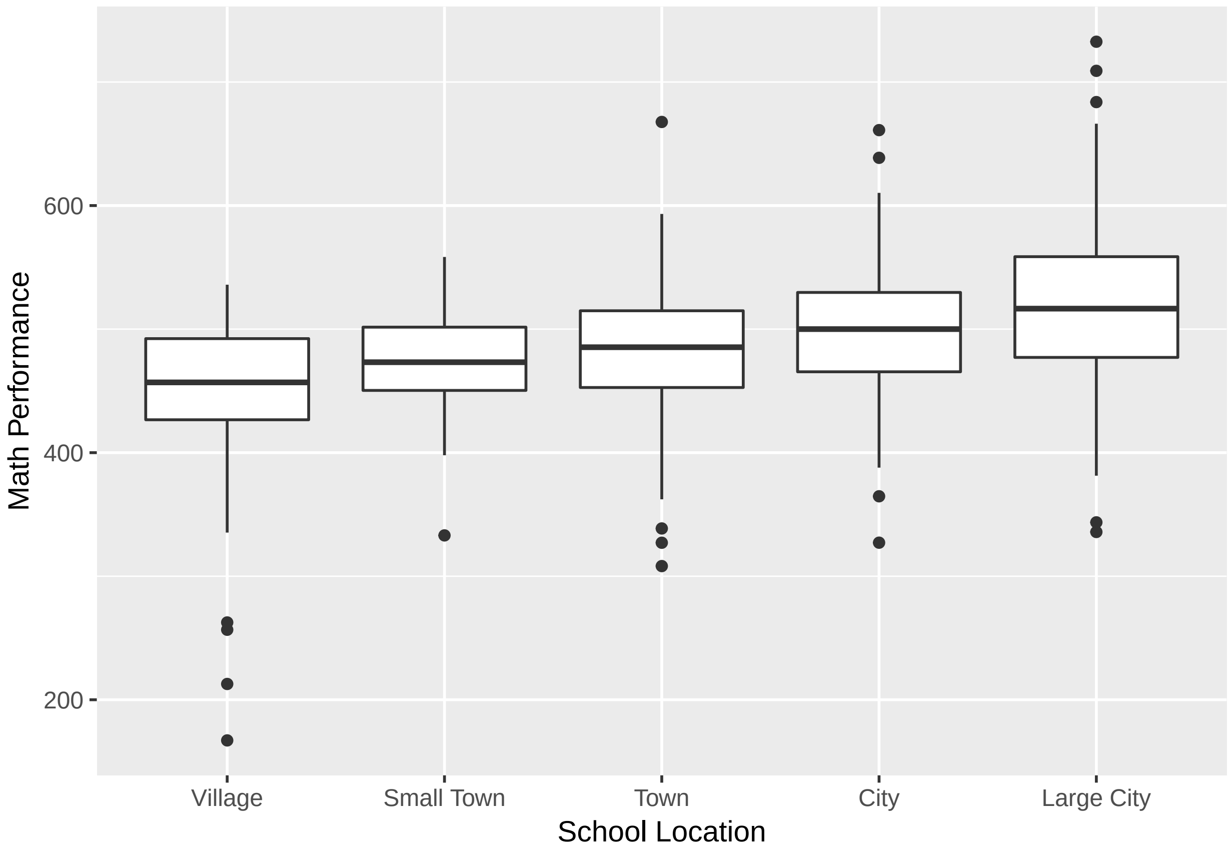
```r
computerStudent2012$W_HOUSEWHT <- n * computerStudent2012$W_FSTUWT / sum(computerStudent2012$W_FSTUWT)
student_data <- computerStudent2012[NC == "Australia" & !is.na(PV1MATH), .(avg= weighted.mean(PV1MATH, w

#merge data
plot_data <- merge(student_data, school_plot_data, by="SCHOOLID")
fit <- lm(avg~SC03Q01, data=plot_data)
summary(fit)
```

```
##
## Call:
## lm(formula = avg ~ SC03Q01, data = plot_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -279.715  -35.140    1.428   35.492  214.570
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         446.794      6.525  68.472  < 2e-16 ***
## SC03Q01Small Town    25.048      9.135   2.742  0.00625 **
## SC03Q01Town          37.020      7.965   4.648 3.95e-06 ***
## SC03Q01City          53.345      7.562   7.055 3.92e-12 ***
## SC03Q01Large City    71.274      7.386   9.650  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.37 on 755 degrees of freedom
## Multiple R-squared:  0.1371, Adjusted R-squared:  0.1325
## F-statistic: 29.98 on 4 and 755 DF,  p-value: < 2.2e-16
```

```r
plot_anova <- aov(avg~SC03Q01, data = plot_data)
summary(plot_anova)
```

```
##              Df  Sum Sq Mean Sq F value Pr(>F)
## SC03Q01       4  367681   91920   29.98 <2e-16 ***
## Residuals   755 2314567    3066
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
ggplot(plot_data, aes(x=factor(SC03Q01), y=avg))+geom_boxplot()+labs(x="School Location", y="Math Perfo
```




```
library(data.table)
library(PISA2012lite)
library("dplyr")
data("computerStudent2012")
setDT(computerStudent2012)
data("school2012")
setDT(school2012)

#filter data
school_view <- school2012[NC =="Australia" & !is.na(SC01Q01)]
school_plot_data <- select(school_view, SC01Q01, SCHOOLID)
#filter data
n <- nrow(computerStudent2012)
computerStudent2012$W_HOUSEWHT <- n * computerStudent2012$W_FSTUWT / sum(computerStudent2012$W_FSTUWT)
student_data <- computerStudent2012[NC == "Australia" & !is.na(PV1MATH), .(avg= weighted.mean(PV1MATH, w

#merge data
plot_data <- merge(student_data, school_plot_data, by="SCHOOLID")
fit <- lm(avg~SC01Q01, data=plot_data)
summary(fit)
```

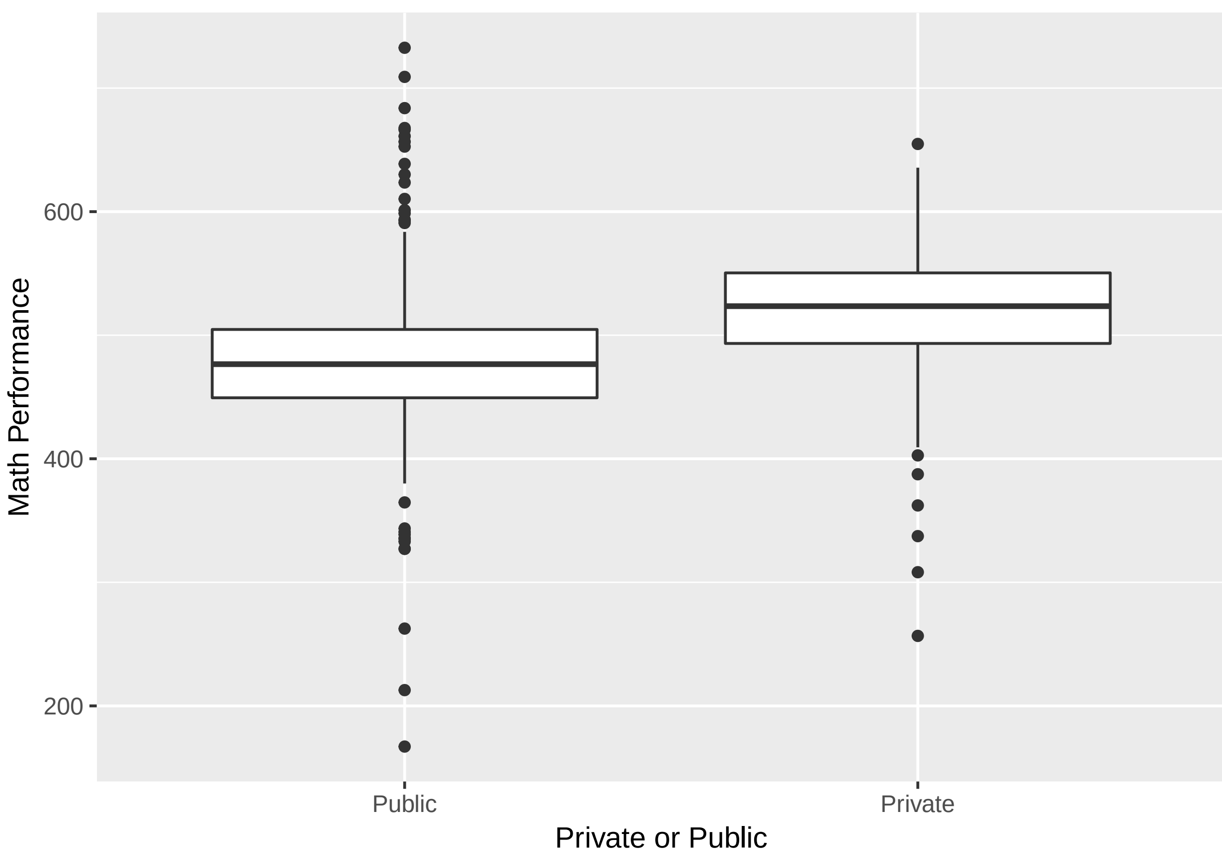
```
##
## Call:
## lm(formula = avg ~ SC01Q01, data = plot_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -311.569  -29.259    0.429   27.836  253.990
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      478.648      2.575  185.87   <2e-16 ***
## SC01Q01Private    41.856      4.106   10.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.77 on 771 degrees of freedom
## Multiple R-squared:  0.1188, Adjusted R-squared:  0.1176
## F-statistic: 103.9 on 1 and 771 DF,  p-value: < 2.2e-16
```

```
plot_anova <- aov(avg~SC01Q01, data = plot_data)
summary(plot_anova)
```

```
##              Df  Sum Sq Mean Sq F value Pr(>F)
## SC01Q01       1  323134  323134   103.9 <2e-16 ***
## Residuals   771 2397990    3110
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(plot_data, aes(x=factor(SC01Q01), y=avg))+geom_boxplot()+labs(x="Private or Public", y="Math Per
```



A few points can be concluded accroding to the analysis result: 1. Students from private/ independent schools significantly outperform public/government students. 2. Students from large city outperform students from other scetors of students. 2. 1/4 of students from public/government schools falls in low performance in mathematic leteracy whereas only 9% of students from private schools falls in this category. 3. Students from private schools have an average of ## scores whereas students from public schools have an average of ## scores in mathematic literacy. 4. The standard deviation of private shcool student is ## while the standard deviation of public schools is ##. 5. According to the regression model,

Discussion

As we can see from the analysis above, there is a positive relationship between the student's socio-economic background and their math performances. Big city students own more education resources than of villages students.

Reference

1. Pmc.gov.au. (2017). Remote School Attendance Strategy | Department of the Prime Minister and Cabinet. [online] Available at: https://www.pmc.gov.au/indigenous-affairs/education/remote-school-attendance-strategy [Accessed 29 Nov. 2017].
2. REGIONAL AND REMOTE EARLY CHILDHOOD EDUCATION. (2017). [pdf] NSW Department of Education. Available at: https://education.nsw.gov.au/media/ecec/pdf-documents/2017/Regional-and-Remote_Literature-Review.pdf [Accessed 30 Nov. 2017].
3. Thomson, S., Bortoli, L. and Buckley, S. (2013). PISA 2012: How Australia measures up. [pdf] Australian Council for Educational Research. Available at: https://research.acer.edu.au/cgi/viewcontent.cgi?article=1015&context=ozpisa [Accessed 2 Dec. 2017].
4. En.wikipedia.org. (2017). Structural equation modeling. [online] Available at: https://en.wikipedia.org/wiki/Structural_equation_modeling [Accessed 4 Dec. 2017].
5. Hox, J. and Bechger, T. (n.d.). An Introduction to Structural Equation Modeling. [ebook] Available at: http://joophox.net/publist/semfamre.pdf [Accessed 4 Dec. 2017].
6. Bulut, O., Delen, E. and Kaya, F. (2012). An SEM Model based on PISA 2009 in Turkey: How Does the use of Technology and Self-regulation Activities Predict Reading Scores?. Procedia - Social and Behavioral Sciences, 64, pp.564-573.