

## What Does the Data Collected During PISA Testing of Teenagers Tell Us About Education Around the World?

Luke Fostveldt, Alex Shum, Ian Lyttle, Di Cook

Reducing inequities in educational outcomes is of great interest to politicians and the public, alike. Dissecting the causes of inequality is a much tougher challenge with no panacea. Nonetheless, tackling inequality in educational outcomes is essential from both moral and economic viewpoints. The Programme for International Student Assessment (PISA) is a triennial survey conducted by the Organization for Economic Cooperation and Development (OECD) with a rotating emphasis on one of mathematics, reading, or science. In 2012, the emphasis was on mathematics. All 34 member countries of the OECD and 31 partnering countries and economies participated in the survey. This represents over 80% of the global economy. The OECD estimates that if all students could reach a level-2 proficiency in mathematics -- a level-2 proficiency means that a student can only handle the "simplest and most obvious tasks" -- it would add US\$200 trillion to the world's collective GDP.

The OECD PISA Results in Focus report describes the survey as "the world's global metric for quality, equity and efficiency in school education". The goal of the PISA survey is to assess the workforce readiness of 15-year old students. Nearly 500,000 students were tested across 65 countries and 18,000 schools. Students were examined on how well they can apply the knowledge they learned in school to applications outside of school. The reported scores range between 0-1000. Information about the students, parents, and schools is also collected. The students completed a questionnaire providing information about themselves, their homes, their schools, and a variety of psychological views regarding factors they believe affect their performance in school. School principals responded to a questionnaire covering their school system and learning experiences for their students. In some countries, parents completed a questionnaire requesting information about their perceptions regarding the school system, expectations for their child, and their involvement in their child's schooling.

The data was made available by the OECD as part of a data challenge for the userR! 2014 conference. Entries to the competition can be found at <http://beta.icm.edu.pl/PISAcontest/>. We learned a lot from this data, and won one of the prizes. This article is a description of the seven main things we learned about 15-year olds, their math, science, and reading skills, and some of the factors that affect these skills.

## 1. The gender gap in math is not universal but the reading gap is, in favor of girls.

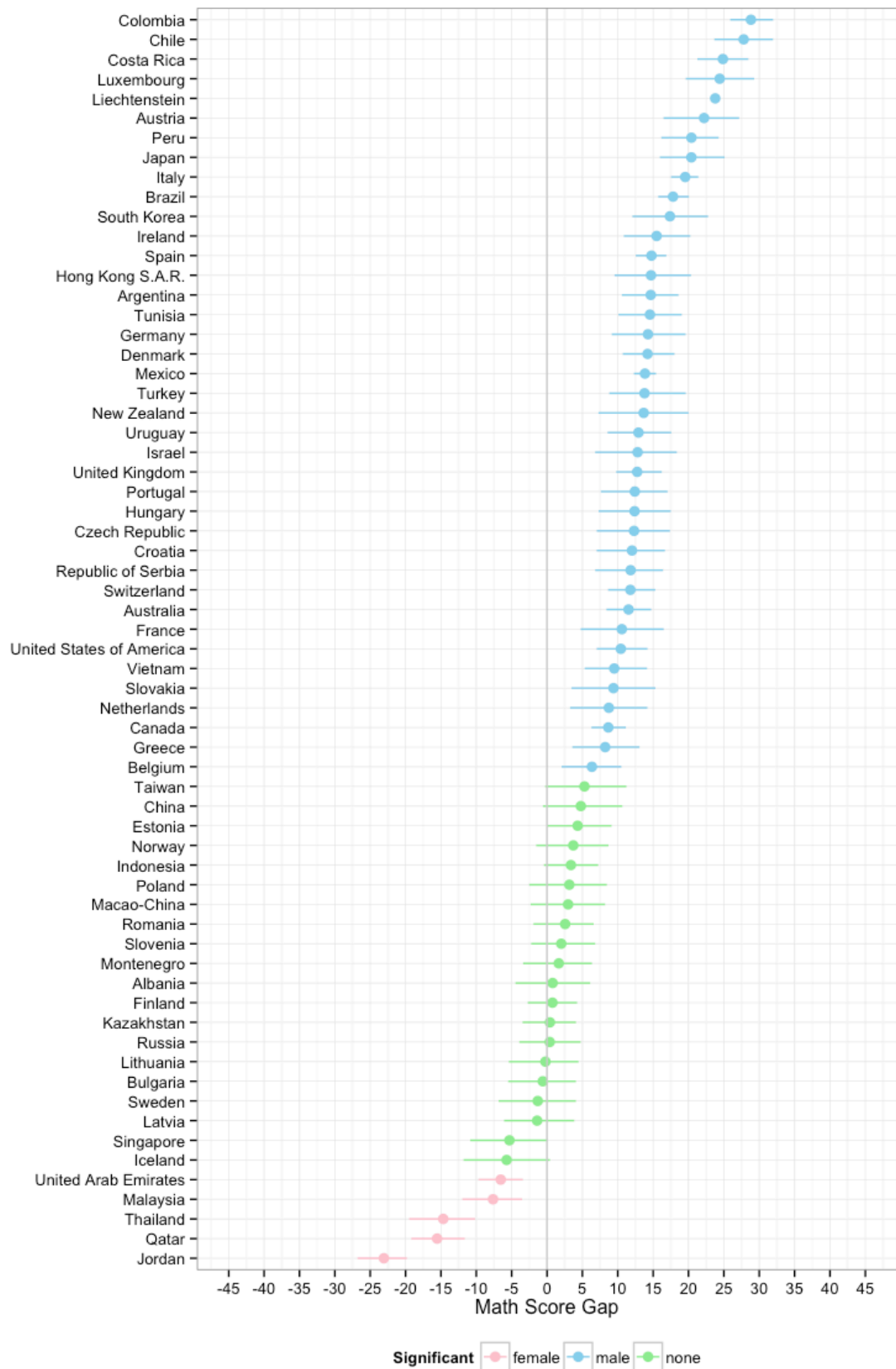
For each country, average scores on math and reading are calculated separately for boys and girls. These averages are differenced to compute the "gap" between boys and girls. To gauge the statistical significance of the difference, t-statistics and bootstrapped 95% confidence intervals were calculated. The results are plotted in the graphs below. Each dot corresponds to a difference between girls' and boys' scores, for each country, with lines giving the confidence interval. Wider intervals indicate less data support the statistics, for example, Liechtenstein tests few students so there is more uncertainty about the mean. Color indicates statistical significance (blue in favor of boys, pink in favor of girls, and green means there was no difference). The maps color the countries by the presence (or absence) of a gender gap.

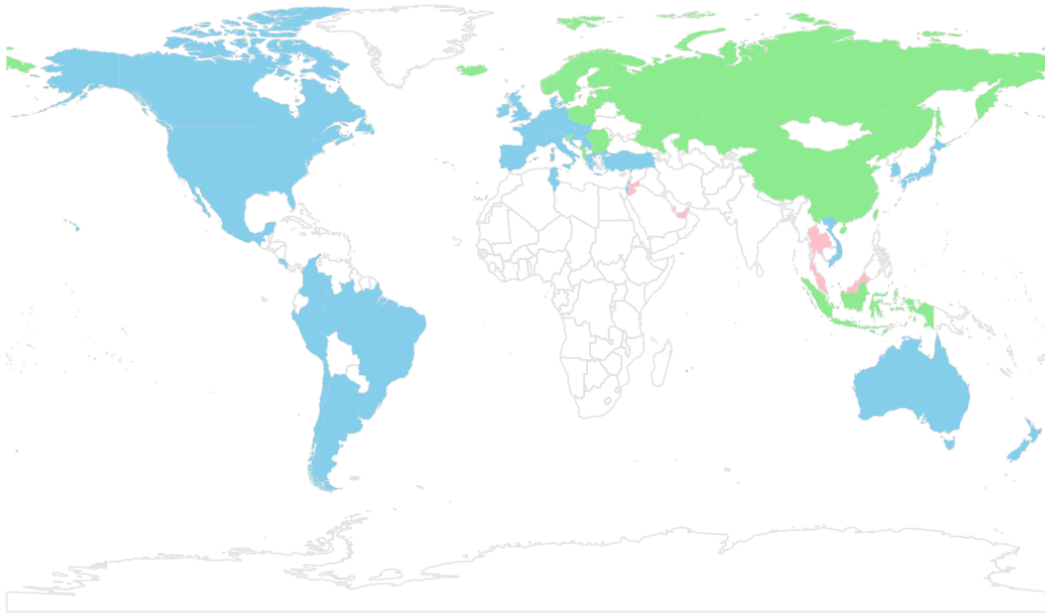
The countries are sorted from biggest gap to smallest gap. Most countries have a gender gap in math scores in favor of boys. Colombia has the biggest gender gap, about 30 points, on average boys do better than girls. It should be noted that this is small, though, because this 30 points is out of a possible 1000 points.

Five countries have gender gaps in favor of girls; these are middle-eastern countries, Malaysia and Thailand. Given popular perception of gender issues in the Middle East, this may be a surprising result. Perhaps this is an example of how data can be used to show that reality is more nuanced than popular perception.

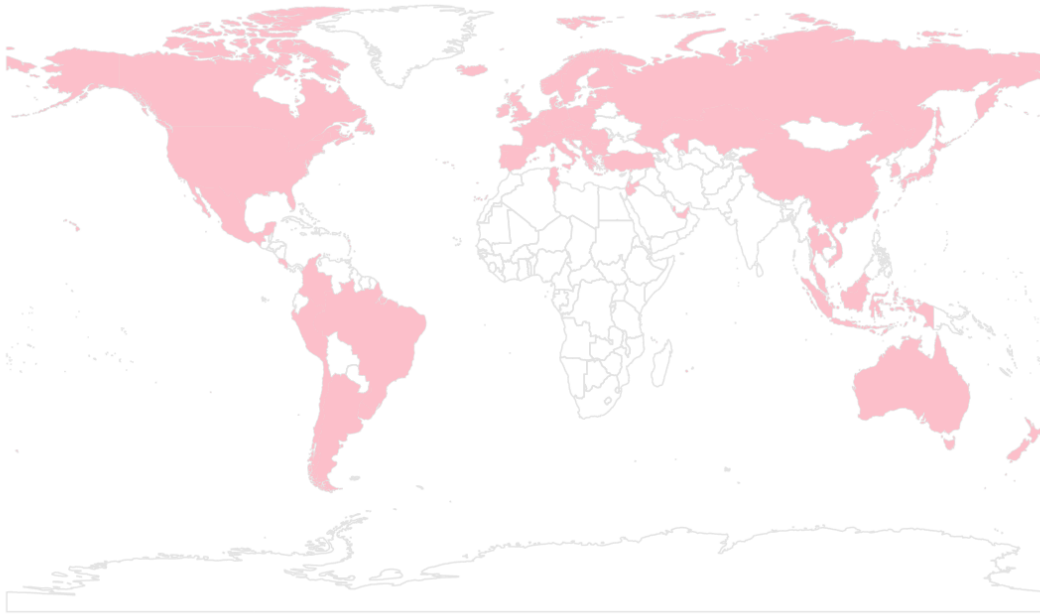
Conversely, though, looking at reading scores, the story is entirely different. Universally, girls score significantly better than boys.

Although not shown here, it also should be noted that this is the difference between the average scores for boys and girls. There is a slightly different story for each country if we look at the top math score for each gender only. We would notice that the top score in the USA is achieved by a girl, and is substantially higher than the highest boys' score. There is far more individual variability than the overall average differences - who you are is more important than your demographic.





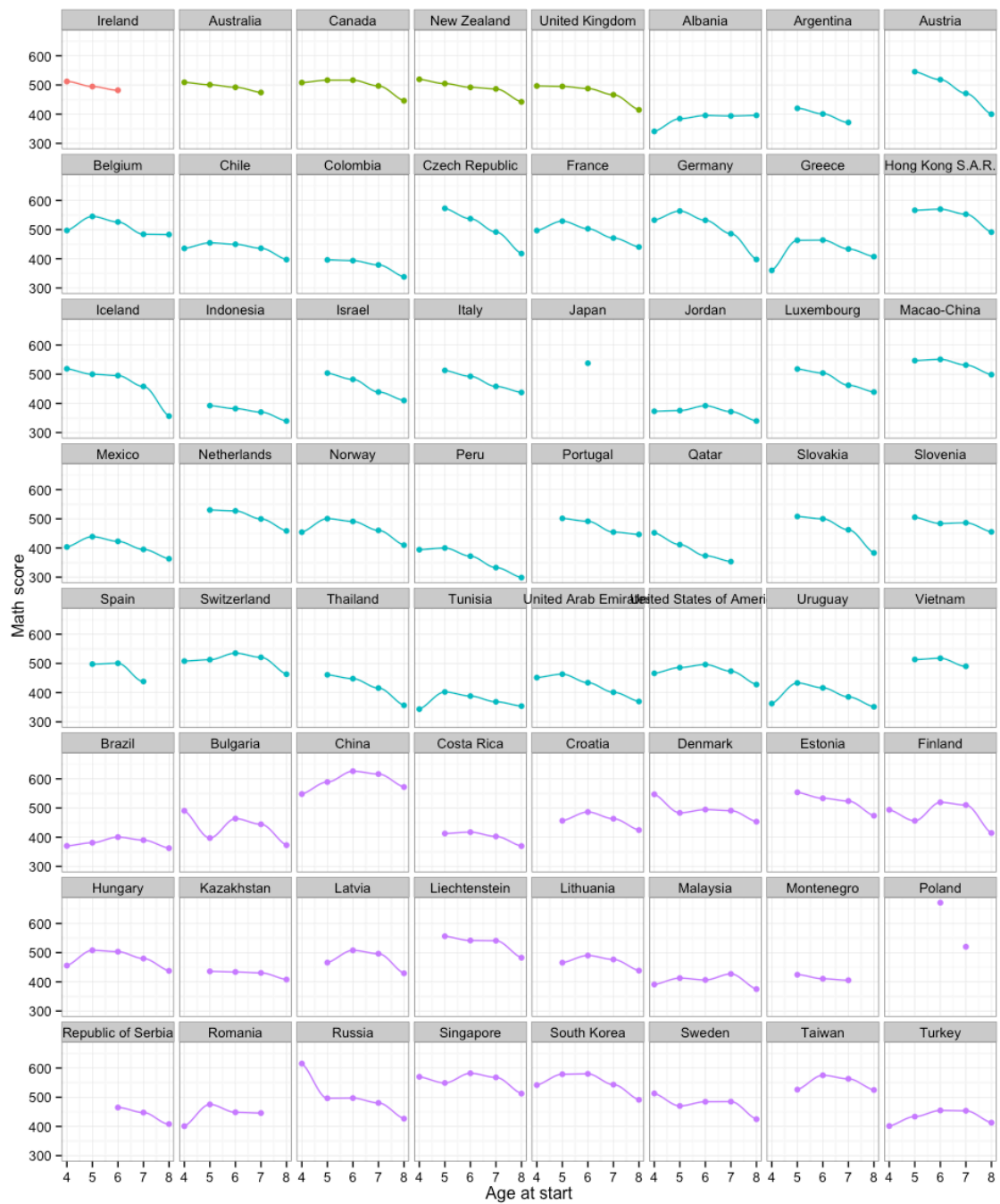


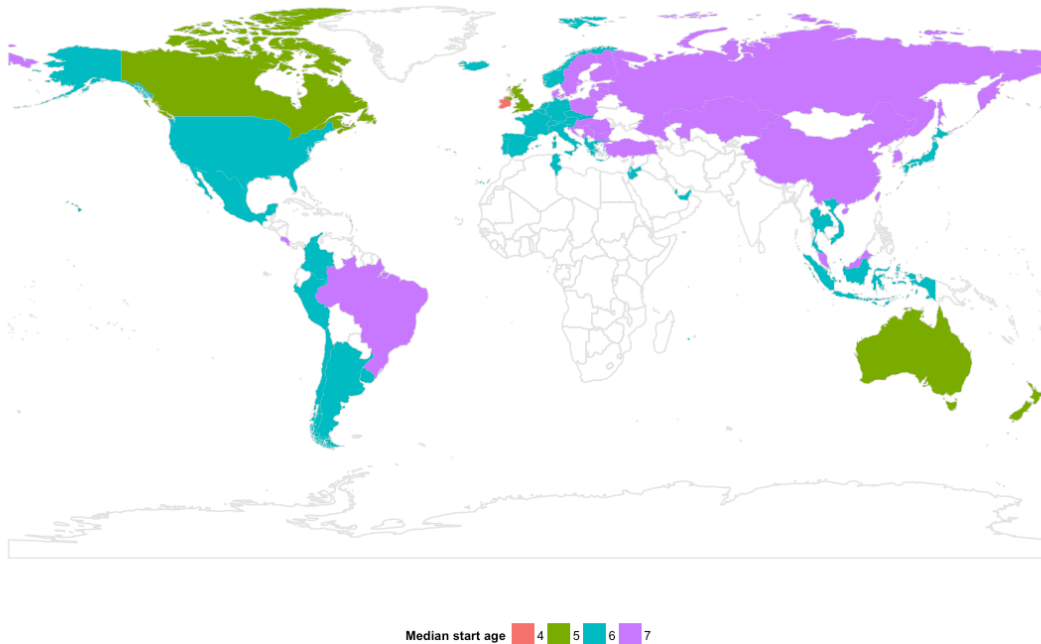


## **2. Starting school at an older age is associated with lower average math scores.**

For each country, we have computed the average math score by age that the students started school. In addition, for each country we computed the median age that children in the study started school. There are differences in typical school start age between countries. Ireland has the youngest median start age - children head to school at age 4. Britain, and its past colonies, Australia, Canada and New Zealand, have a median start age of 5. For many other countries the median start age is 6, and for northern Europe, much of Asia, and Brazil median start age is as late as 7. Some countries appear to have rigid starting ages, e.g. in Japan all children start at age 6.

Within a country, looking at the average math score for each start age, the score typically declines, especially after age 6. This is almost uniformly the pattern across all of the countries in the study. For several countries - Russia, Denmark, Bulgaria - if a child starts school at age 4 their math scores were much better on average than any other age group. School start age appears to matter: with starting school in the 4-6 age range being associated with higher average math scores. There could be many other factors confounding this observation, most notably being that it is the more mathematically gifted students that start earlier, leading to a difference in the population characteristics of the different age (at school start) groups.



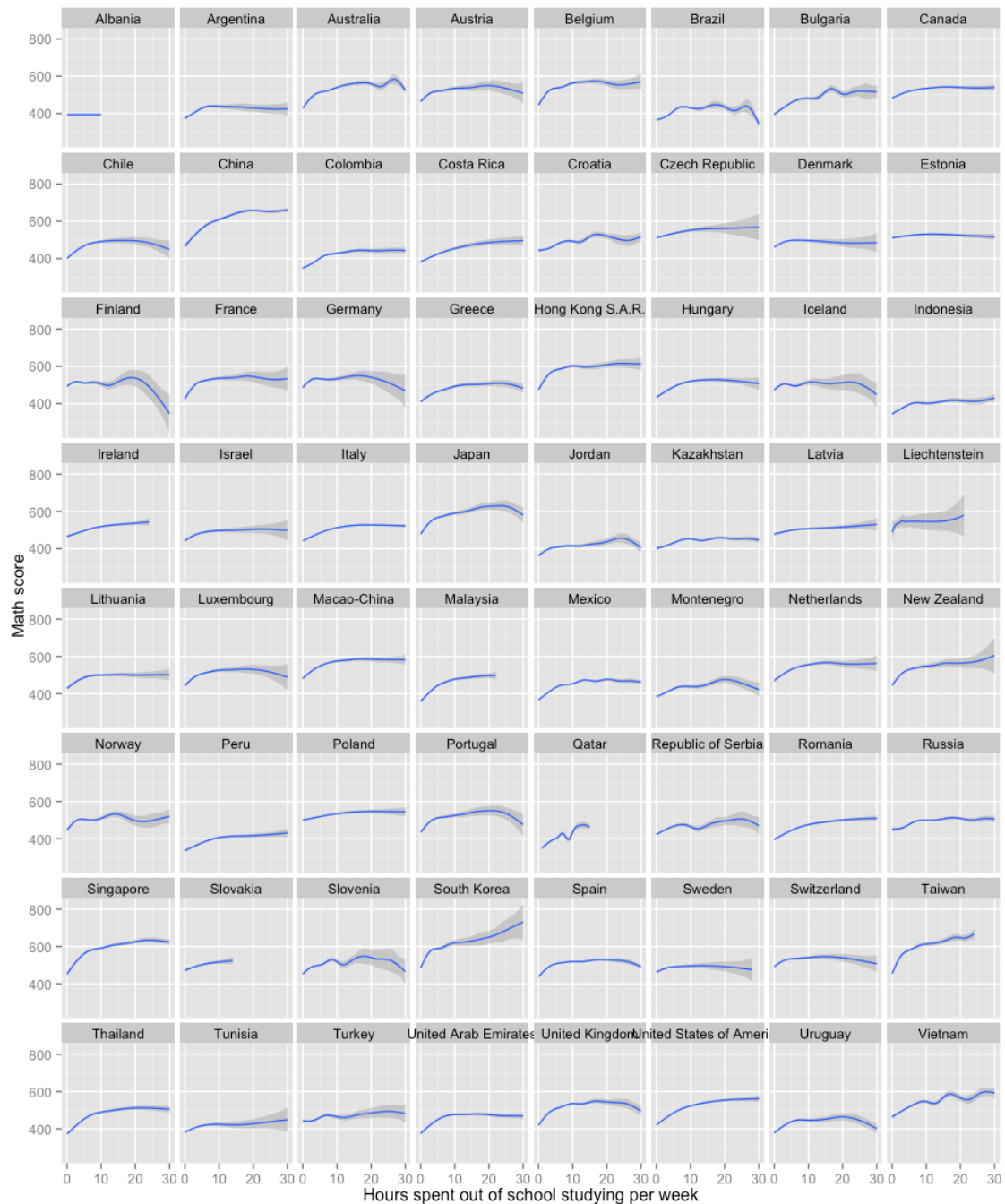


### 3. Time spent out of school studying is important, but only up to a point.

For each country we have provided loess-smoothed curves of the relationship between the math score and the time spent studying outside of school. Math scores increase on average the more time spent studying, up to about 10 hours. Beyond that scores stay flat, with the exception being several Asian countries: Vietnam, China, South Korea, Taiwan. For China, the gain tails off around 20 hours of study.

The vast majority of students spend less than 10 hours studying per week. The pattern suggests decreasing benefits to increased out of school study time after 10 hours. The thickness of the standard-error bars in the plot are an indication of the sample size, with thicker line indicating less certainty about the math value.



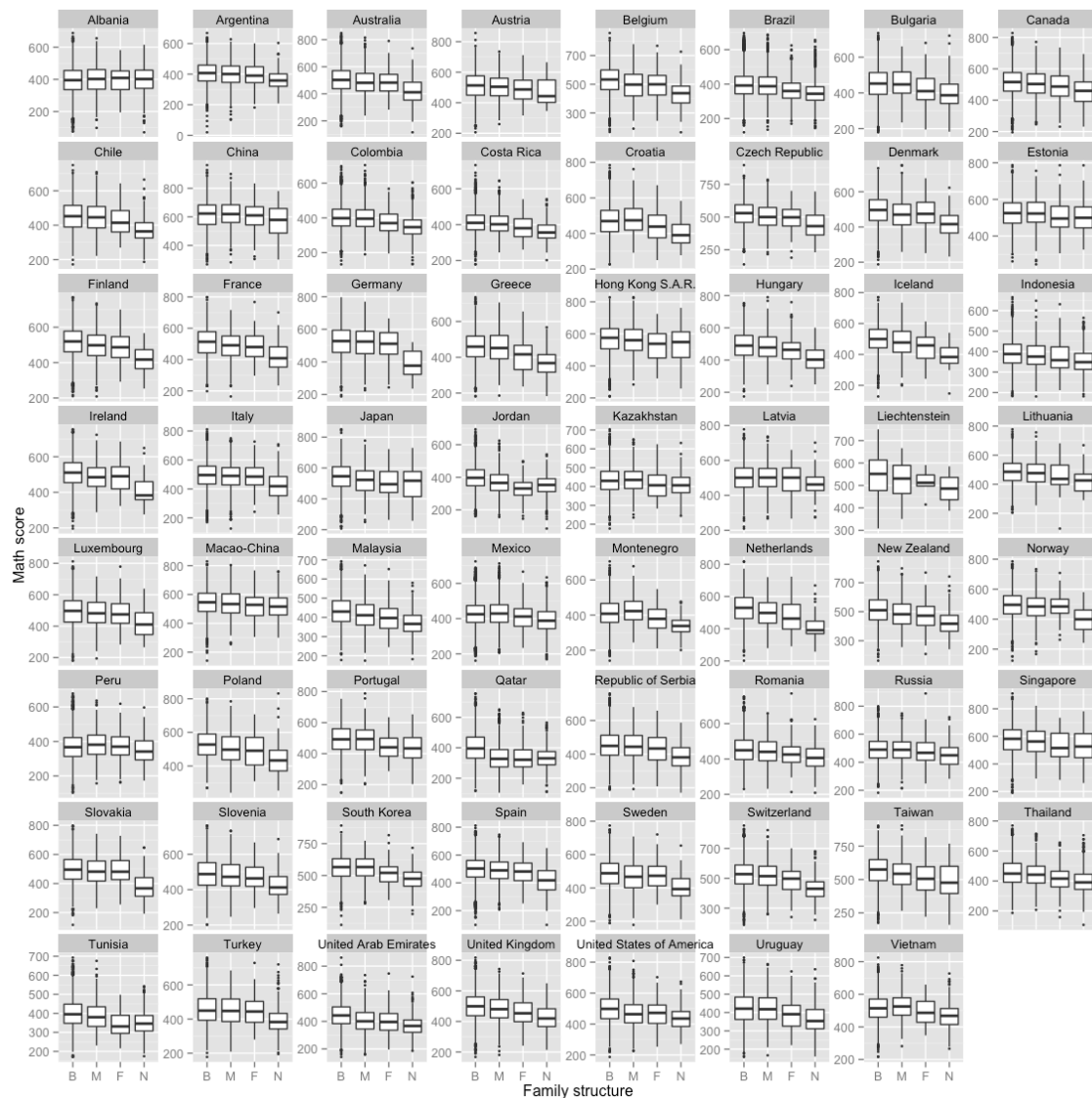


## 4. Parents matter!

### a. Number of parents in the home

The role of parents in educational achievement is the subject of much research. Looking at the number of parents in a household shows that students in two-parent households are, on average, performing better in math than students not in a two-

parent households. The majority of students (97%) live in two-parent households. Of the single-parent households, a single mother is five times more common than a single father. Scores of students with just a mother at home tend to be higher than those with just a father at home. In many countries, mostly Asian, students from single-mother households achieve comparably to those from two-parent households. Students in households with no parents tend to perform the worst. This pattern is consistent across all countries.

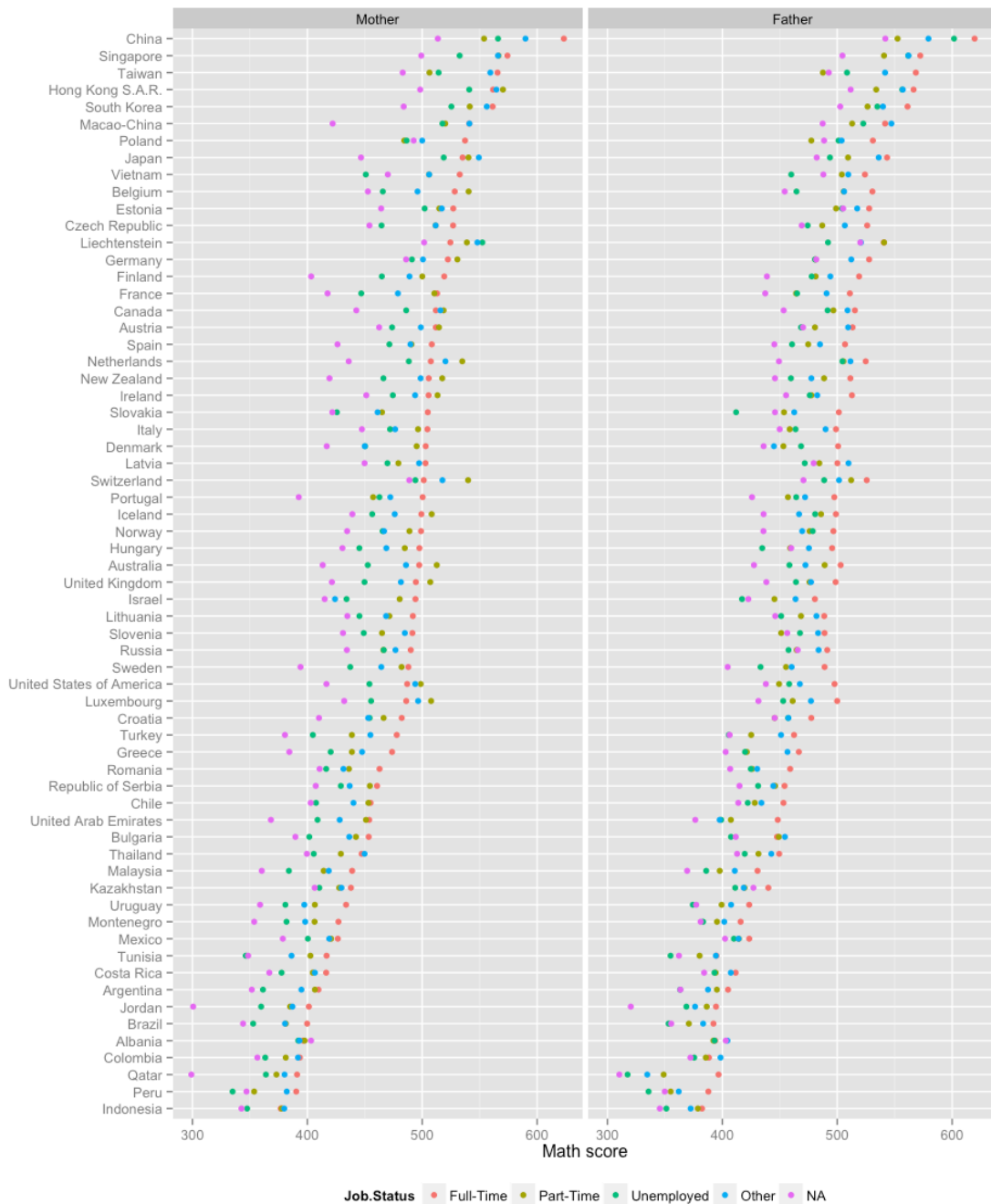


## b. Parent's Occupation Status

One of the variables the survey addressed was the occupation status of each of the parents. There were five possible categories: "Full-Time", "Part-Time", "Not working, but looking for a job", "Other (e.g. stay-at-home, retired)", and no response. Fathers primarily have full-time jobs. However, there are large numbers of mothers who work full-time, part-time, or are categorized as "Other". With women making

up a larger portion of the work-force, it is interesting to look at how the parents' occupations relate to students' performance in math.

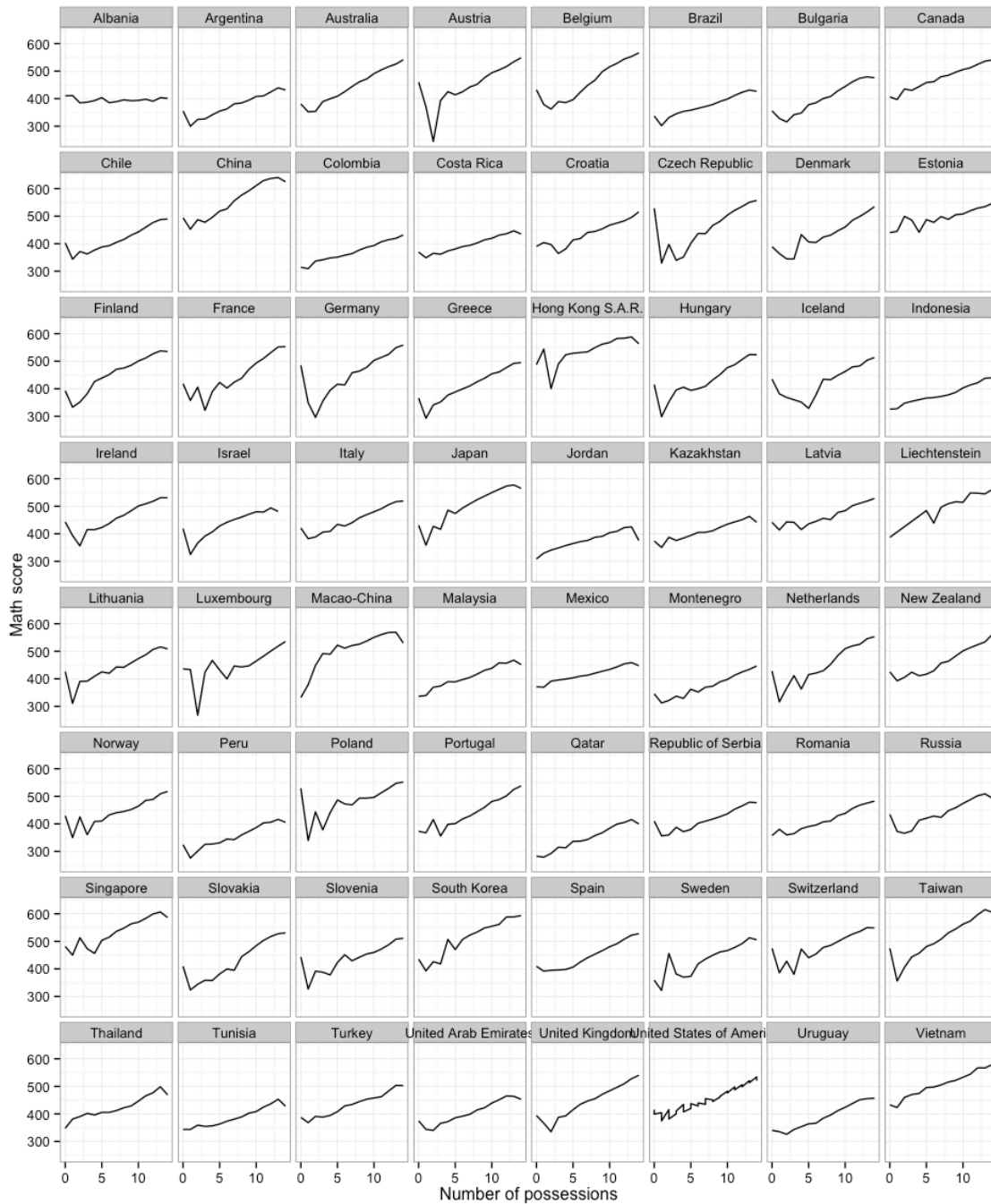
The plot below shows the average student math score by the parents' occupational status. The countries are sorted according to the average score for students with a mother who is "full-time" in the work force. Occupational status is distinguished by color. Students with mothers who work part-time or full-time tend outperform those students with mothers labelled as "Other". This pattern holds with only a few exceptions, Switzerland, Netherlands, USA, Japan. Meanwhile, when the father is employed part-time, students tend to have much lower scores than if the father is labelled as "Other". Those students whose fathers have full-time jobs have the highest average scores among the possible work statuses for the father. The patterns described also hold across reading and science.



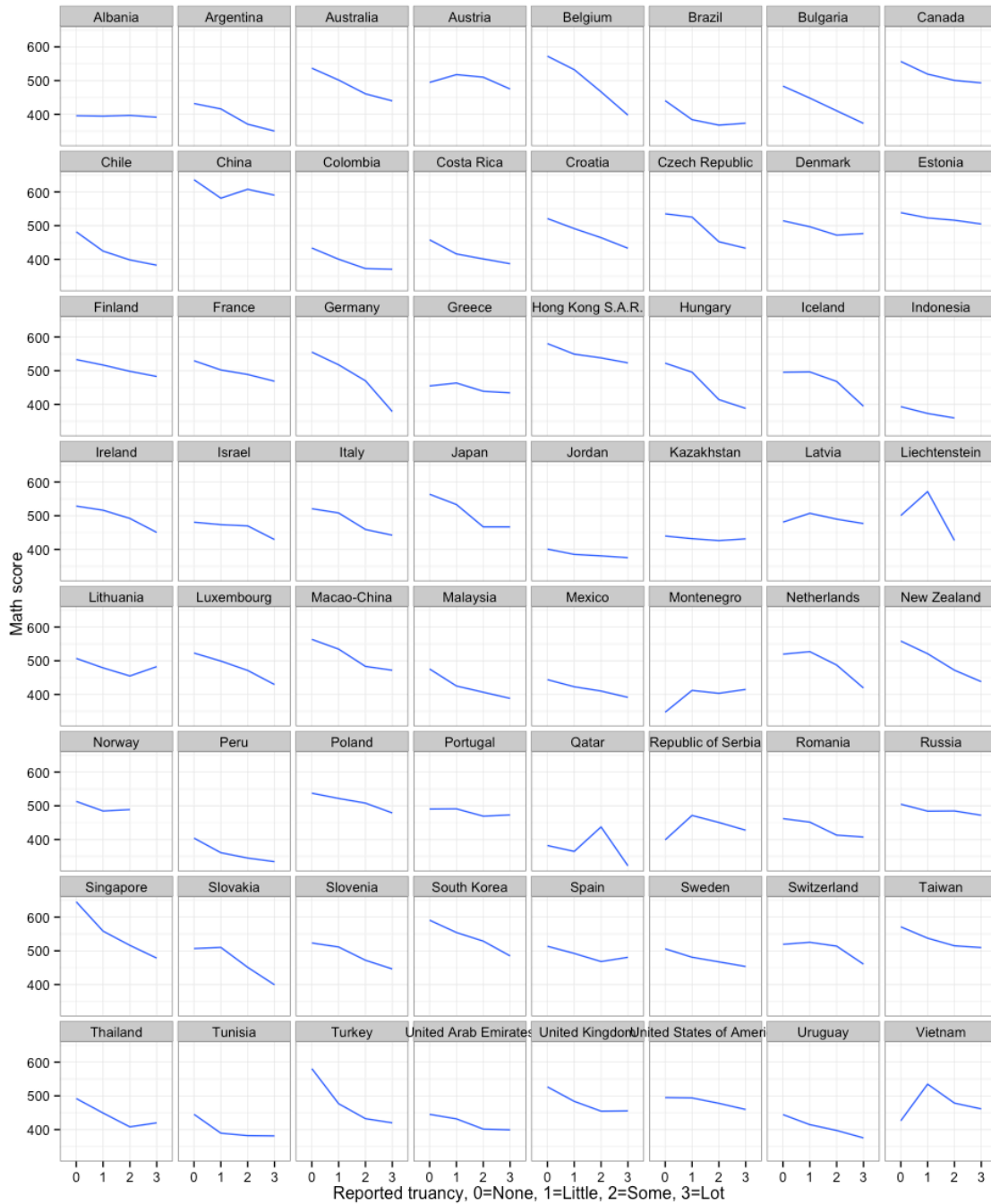
## 5. Albania is not like anyone else!

Albania appears to be an outlier in many educational trends. Many socioeconomic and school factors that are associated with test scores in other nations are unrelated to Albania's test scores. For example, there is no difference in math scores based on household structure in Albania, or in terms of hours spent studying.

The lack of association where we would expect association is also evident in number of possessions. Generally, students with more possessions score higher on average. The number of possessions is a surrogate for a student's economic standing. The survey records whether students have desks, computers, internet, books, a place to study and dozens of other possessions. After aggregating these number of possessions together, to get a total, we can see that scores increase on average as possessions increase in all nations, except Albania.



A similar aberration is seen with school reported truancy rates. Each school reports the level of student truancy: not at all, very little, to some extent, or a lot. We group schools together by reported truancy level and aggregate the test scores for students at these schools. Generally students at schools with high truancy rates average lower test scores than students at schools with lower truancy rates, except for schools in Albania.



These patterns, especially all together, would most likely indicate that there are some problems with the way the data was collected for Albania.

## 6. On average, more TVs yield higher math scores in the developing world, but lower in the developed countries.

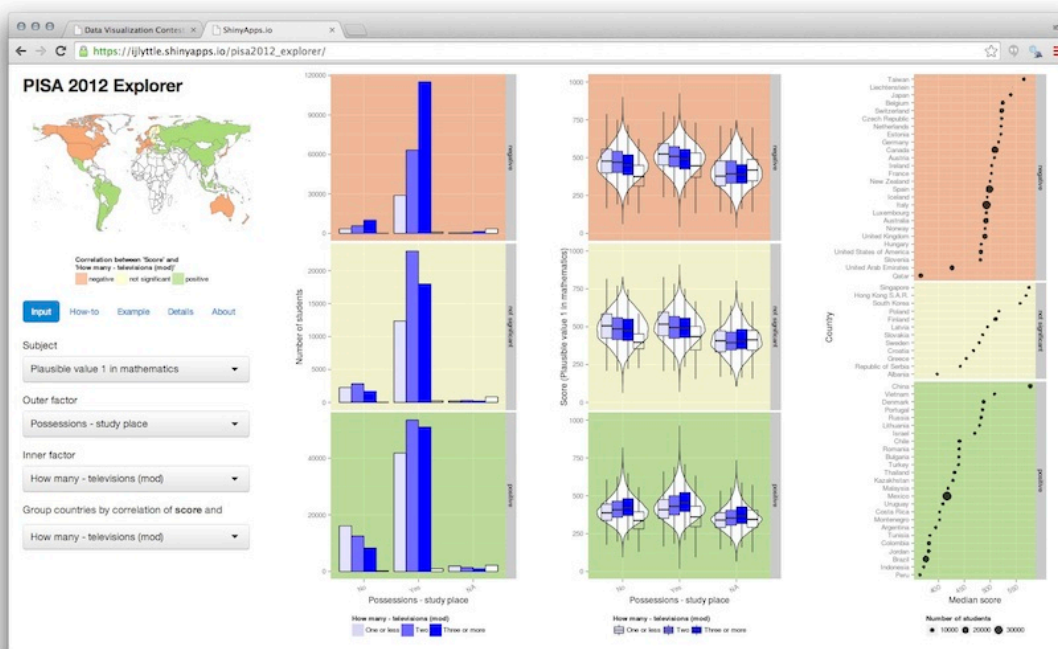
One of the factors in the data is the number of televisions in the student's household. Possible values were: "zero", "one", "two", or "three or more". Because there were so (relatively) few observations of "zero", for this investigation, the "zero" and "one" categories were combined to form a "one or less" category.

A linear model was fitted for the students of each country, using the math score as the dependent variable and the number of televisions as the independent variable (as an ordered factor). The direction and significance of the correlation was determined for each country. Significance is indicated by the p-value for a given regression being less than 0.05.

For many countries, a significant relationship was be found between math scores and number of household televisions. The interesting part was to notice that those countries with a significant *positive* relationship appear to include much of the developing world (with lower median-scores), while those countries with a significant *negative* relationship appear to include much of the developed world (with higher median-scores).

The figure below illustrates the relationship, along with another factor, the student has a place to study. It is a complex diagram, but of interest is the middle column of plots - the boxplots overlaid on other information show the math scores of students relative to the number of TVs in the household, separately plotted for countries falling into one of three categories, negative, zero or positive correlation. The white regions are violin plots of math scores by presence or absence of a place to study. A place to study universally leads to better math scores on average. The map is colored by correlation between number of TVs and math scores. The bar charts indicates sample size for the categories. The dot plots show average math scores by country, separately for each correlation group.

It is interesting that the more TVs in the household in developed countries leads to lower math scores on average. Although results shown here are for math only, a similar set of relationships is found using the reading and science scores. It seems that television is a great equalizer.



## Play with the data yourself.

The PISA dataset contains hundreds of factors, and this article can describe just a few findings. The reader is welcome to explore the data themselves using an [interactive web tool](https://ijlyttle.shinyapps.io/pisa2012_explorer/), made with RStudio's [Shiny](https://www.rstudio.com/shiny/). Shiny is an open-source tool that allows an investigator to create web-based applications powered by R.

## Discussion

The PISA data is a very large data set, with 500,000 students measured on more than 600 variables, and additional information collected about school and home life. The selection of students into the study follows a carefully designed survey instrument, which means that each student also has a sample weight which reflects how their demographic is represented in the sample. These allow the inference of findings made from this sample of students to the entire student population in the country. Strictly speaking, for the PISA data the survey weights primarily apply to the test score components, covering different types of schools and ensuring good representation of both genders. There is no guarantee that all socioeconomic groups are represented equally, or that households with all sorts of possible possessions are represented. The sampling design focused on school types and gender, only.

We have conducted what would be considered to be an exploratory analysis of the PISA data. There is a big difference between exploratory analyses and confirmatory analyses. When many people think of statistics they have in mind confirmatory methods, which involve deciding on a hypothesis, collecting the data, and doing a



statistical test of significance. Exploratory data analysis is different, it focuses on the data at hand, and the goal is to let the data inform, to use various methods, primarily visual to find relationships and patterns and anomalies that may not have been anticipated ahead of the data collection. Exploratory analyses often lead to designed experiments where conditions are controlled, or specific data collection with all the controls to allow for hypothesis testing. Exploratory analysis is particularly important for today's big data society. Data is available everywhere, and in many situations it is collected because it is easy to. Mining data, like we have done for the PISA survey data, can lead to constructing hypotheses that can be rigorously tested with new data.

To organize an exploratory analysis, start by formulating several questions to answer, based on the description of the data. These will form the starting point for the analysis, but they might also lead to more questions being raised as the exploration ensues. For our analysis, these are the questions that we started with:

- Is there a gender gap, like we hear about in the media?
- Do more possessions, essentially higher socioeconomic status, lead to higher scores?
- Is time spent studying outside school instrumental in higher scores?
- Is there a housewife effect - that students with stay-at-home mothers score better?
- What is the relationship between books and televisions and scores?
- Does age when starting school influence scores?
- Do parents occupations affect scores?

Most of these were answered in the analysis, with the exception of occupations because it turned out that there was too much missing data.

Exploratory analysis, like confirmatory analysis, relies on probability, to some extent. If the sample data being examined is not representative of the population of interest then no inference is valid. If you search for many, many different associations it is likely that you will find something that looks important. This is the interplay of probability that helps determine whether what you see in the sample infers structure in the population.

The PISA data is a carefully designed survey, but the student and household questionnaires are more loosely administered, yielding many missing values, and perhaps some questionable data. When results are surprising, and contradict current thinking, it can be helpful to research the problem in more details. For example, in follow up exploration of the PISA data during an introductory statistics class, we found that South Koreans do not use the internet as much as students in other countries. This is surprising, given the high tech nature of the society, and leads us to be skeptical about that part of the data. There are several plausible reasons: (1) the mobile phone network dominates internet access more than internet providers, and so internet usage is masked by phone usage, (2) individual

identification on the questionnaires might lead students not to answer honestly. Both problems could be fixed by re-designing the survey instrument.

Our findings about test scores and other factors, documented in this article, some of the results are consistent with current thinking, e.g. parents matter, socioeconomic status matters, time spent studying matters, but also contained some results that surprised us. These findings are what this data suggests, but they are only as good as the data. If the collection was not well done, and the collection is the responsibility of the local committee in each country, then the findings cannot be trusted. An example of this is the data from Albania, which our analysis suggests that the PISA committee needs to re-assess what was done by that local committee, because the results from the questionnaire are at odds with all other countries.

Overall, our analysis yielded several interesting insights into factors that affect the average test scores. The data is rich, and should reliably reflect the larger population of 15 year olds. Average scores are not the individual experience. We found that the major source of variation in the data is from individual to individual. For example, even though, on average boys perform better on the math tests in the USA, and girls on the reading tests, the top math score was achieved by a girl, and the top reading score was achieved by a boy. You as a person are more than your demographic - you count, your individual efforts matter.

## Acknowledgements

The following R packages were used to conduct the data exploration: shiny, plyr, dplyr, ggplot2, stringr, reshape2, lubridate, maps, ggmap, htmltools, rworldmap, grid, scales, doBy, boot. Principles of reproducible research were used for the analysis and the markdown script is available from [https://github.com/ijlyttle/isu\\_pisa/blob/master/paper/paper.Rmd](https://github.com/ijlyttle/isu_pisa/blob/master/paper/paper.Rmd). By downloading the data from the OECD site, all of our results can be reproduced in the free, and open source, software R, particularly using the RStudio interface.

## Reference

OECD (2014) "Pisa 2012 results in focus: What 15-year-olds know and what they can do with what they know." Technical Report, Organization for Economic Development and Cooperation.