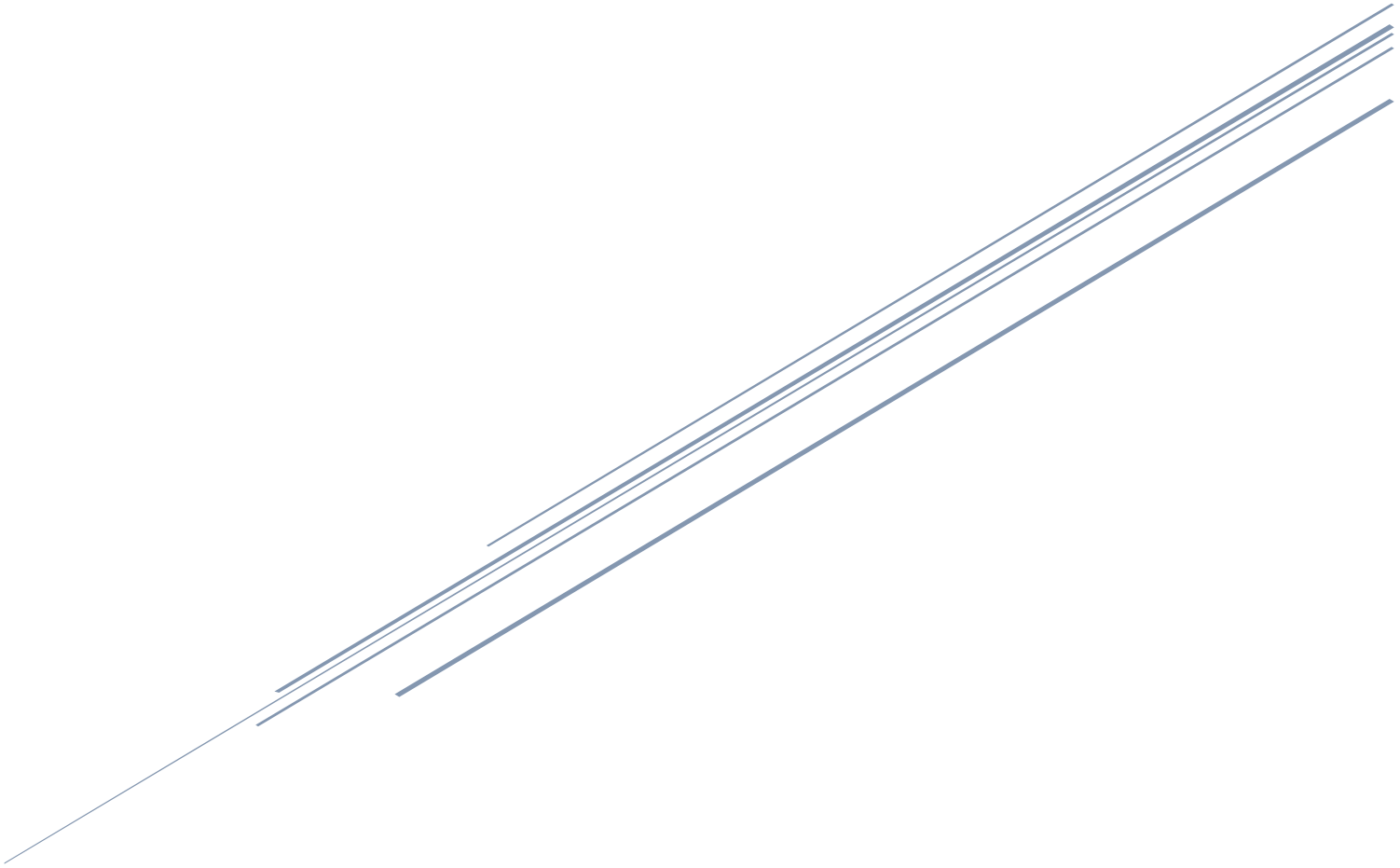
PISA DATA 2012 ANALYSIS

Shirley Ping Hsuan Chen



Links to story

First version

Final version

Summary

In this analysis, there are several findings along the story told:

- Most of the top-performing countries are developed countries (mostly OECD countries).
- Asian students tend to perform better on math than their pals in rest of the world, while top-performing countries tend to be located outside of Asia.
- Generally speaking, female students outperform male students on reading; males do better on math rather than other subjects.
- Those perform better on PISA test tend to have positive attitude toward. On the other have, perseverance does not have that much of impact on students' test performance.
- Higher-educated parents tend to have positive impact on their child's PISA score
- The wealthier the student is, the more PISA score he/she is likely to get. But there's exceptions observed in Vietnam and Qatar.

Design

When designing the structure of this analysis, my first idea is to look at a general picture of scores around the world, and then zoom in to some factors that I find might be related to PISA performance, such as gender, attitude, and wealth.

Comparison among Countries

My initial design is to use filled map, bar chart and mark of score to show the PISA performance of different countries as an overview. As many countries are involved in the test, using a filled map with average PISA score as color scale provides a great general idea of how each country performs – readers can easily find out countries that perform better by identifying nations with darker colors.
Also I chose bar charts to display the top-10 countries of each subjects. As here I have narrowed down to 10 countries, differences in height of bar charts can help readers see how each top-performer do better than its up runner. Initially I included score figures in bar charts, but as my reviewer pointed out, these figures do not mean much to readers so I removed them. I've also follow my reviewer's advice to encode continents with colors in my bar charts, so readers can have a better idea of what Asian, European, and American students are good at respectively.
I put the filled map at the top of canvas so readers can have a general idea at the first sight; and put bar charts of top-10s at the bottom so when readers look down they can see more in-depth information of PISA score.

Comparison between Genders

For gender analysis, I used dot chart with subjects encoded with colors to show the difference between two genders. I added color legend so readers can know which color stands for which subject. Leveraging Tableau's interactivity, I show mathematics, reading, and science score respectively of top-10/bottom-10 males/females – readers can judge how male and female perform differently in each subjects by looking at the gap of dots, while see more details of each perform-gender pair by hovering over to dots to see tooltips; average score are shown in area chart to provide information but avoid confusion with other dots.
For overall score comparison between genders, I use dot chart again to illustrate the percentage difference (male as baseline.)

Comparison among Parent Education Classes and Wealth

I used boxplot to show the distribution of average PISA score for students whose parents are in different education levels. Boxplot is good at illustrating data with different levels – in this case the plot help me to illustrate student's average score while also showing both of his/her parents' education level.
And I used scatter plot to show how wealth affect PISA performance. I chose scatter plot because it's often best choice when presenting relationship of two variables, which is exactly the case here (wealth and average score.) After collecting feedback, I changed color coding from average PISA score to whether a country is in OECD league.

I plan to show how attitudes and perseverance affect PISA score, but the result doesn't seem to be satisfactory and cannot explain much, so I remove this part from my story.

Feedback

I turned to one of my friends who's more familiar with data analysis to provide feedback on my first sketch. Here's his feedback:

What do you notice in the visualization?

At the first glance, the color of China is quite eye-catching. Except for this, I can't see much from the visualization.

What questions do you have about the data?

Story point 1: Since I don't have sense to PISA score, showing score figure doesn't make difference to me. You can show which part of the world (from continent perspective) consists more of the top 10 list, to align with your description.

Story point 2: You wrote 'tend to be bigger - males of top 10s' – what's your explanation to this statement? The display on this story point is interesting, but you need to come up with more explanation for your visualization.

Story point 3: It's difficult to see the connectivity in the matrixes on the top. Moreover, though the histograms at the bottom are clear to understand, it is hard to see any trend or pattern.

Story point 4: I don't quite understand what the color stands for in the plot on the right.

What relationships do you notice?

The relationship between wealth and students' scores, as this is the most straightforward. Other than that maybe the connectivity of score and gender.

What do you think is the main takeaway from this visualization?

The relationship between wealth and students' score. Tableau is a really nice tool.

Is there something you don't understand in the graphic?

I listed them down in question 2 already. My recommendation is that you need to make visualizations easy to understand, and make them easy to be connected to readers' thought. You might want to think deeper in order to tell an insightful story.

My reviewer says my final sketch looks way better that the first one.

Resources

PISA OECD
Udacity Data Analysis Forum
Tableau Community
Stack Overflow
Levels of Education
Pisa2012_subset.csv (I used pandas package to rename column and remove some fields that I don't intend to use, this csv file is the process file)