Aimed audience :

Our general aimed audience are people who is interested or who has professions that is related to education, because our datasets and analysis provide profound insights into seeing the student performance difference between two genders across different countries. That results can be utilised by people who is working in education industry, such as gaining marketing insights from our results and then come up with their own marketing campaign, etc.

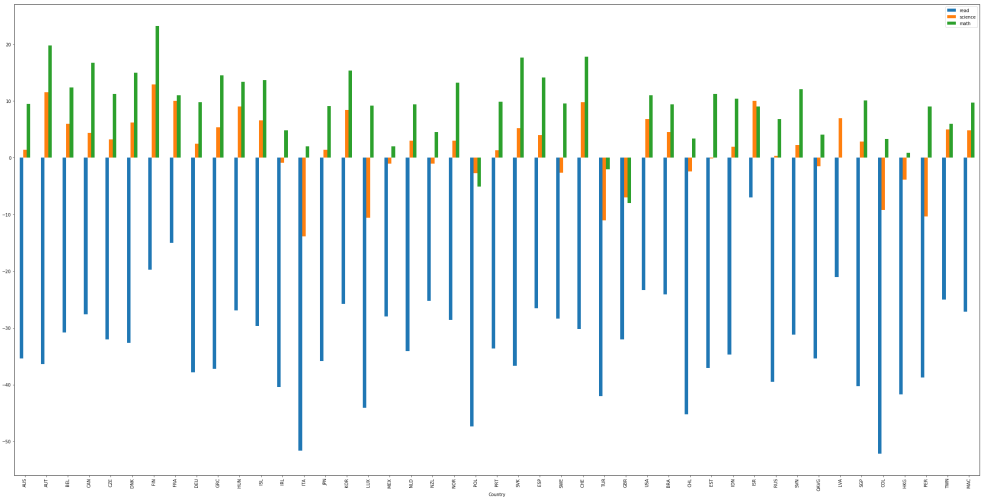
- describe the domain situation, :

  Those datasets we are using are obtained from measurement of the scientific literacy of a 15 year-old in the use of scientific knowledge to identify questions, acquire new knowledge, explain scientific phenomena, and draw evidence-based conclusions about science-related issues. The mean score is the measure. The datasets includes data entries over all the continent and the sample size for each country range from 401 to 559(Indonesia and Singapore). The attributes include location, indicator, subject, value, time, and frequency.

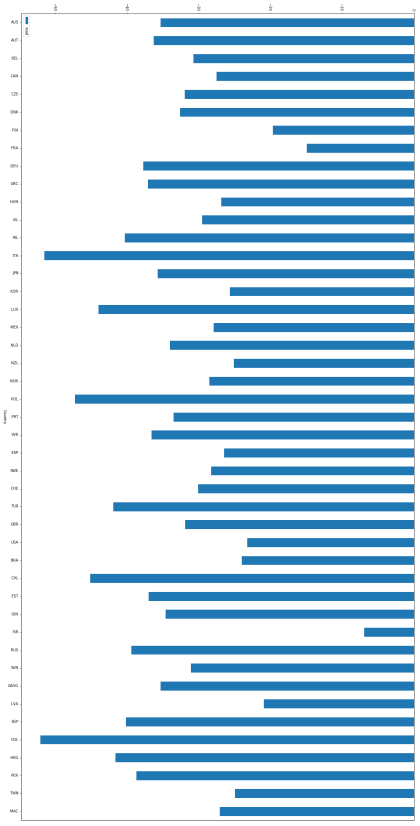
- the origin of the data you used, and then :

  obtained from canvas provided cleaned dataset, and three other datasets are obtained from https://data.oecd.org which are Reading performance(PISA), Science performance(PISA), and Math performance(PISA) respectively.

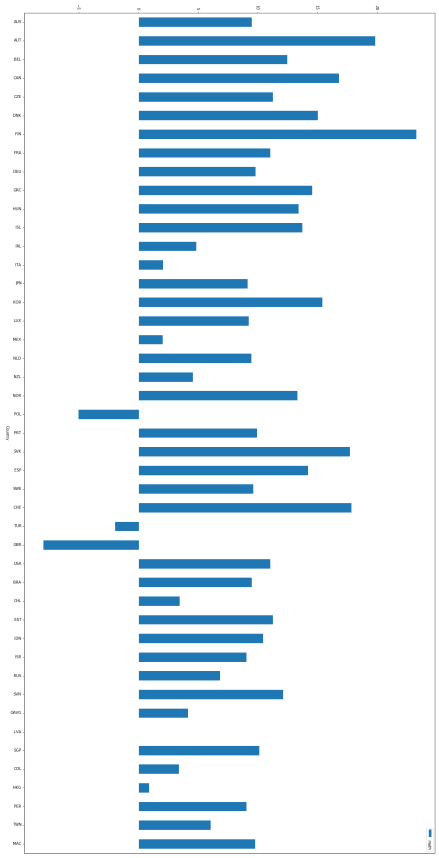
- present what your analysis has revealed about the domain.



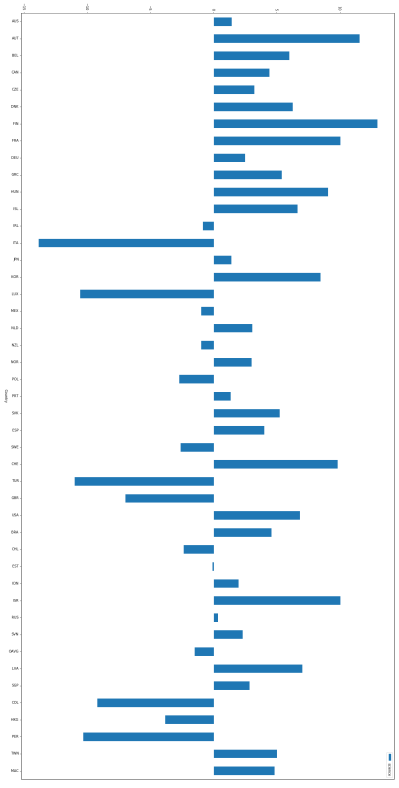
This is the plot that includes overall attributes of the datasets and it demonstrates a visual representation of the difference between the score of boy's and girl's of each country and three different colours indicate three different domains which are science, maths, and reading.



This plot indicates the difference of score in reading between girl's and boy's (score of girl - score of boy), and we can see clearly that the difference are all positive numbers, therefore it demonstrates visually that girls on average score much higher than boys across all countries that are in the data sample. There is only one colour to indicate the bars because we are extracting one attribute which is reading out of the three subject attributes from our dataset. However, given the conclusion that girls do much better than boys at reading among most countries, girls in countries such as Israel, Finland, and France perform still better at reading than boys but not to the marginal amount of difference like the other countries. Other factors might have caused this phenomenon, although we are not able to pin point the exact reasons due to lack of variables.



This plot extract another attribute out of the subject attributes which is math, from the plot we can see clearly that the comparison of difference between the score of boy's and girl's. It is visually clear that the average score of math of boy is much higher than the average score of girl across the majority of countries within the dataset. Apart from countries such as Britain, Poland, and Turkey that has girl's score higher. By looking at the whole picture of the dataset, we can determine these three countries are the anomalies of our dataset. There are confounding factor that might cause this phenomenon to happen, could be cultural difference or environmental difference. However, without more factors from the dataset, we are unable to answer this question.



This last plot is the visual demonstration of the difference of score between girls and boys in the domain of science. It seems relatively clear that even though some countries girls out perform in science to a marginal amount, some countries boys outperform girls to a marginal amount as well. And the rest of the countries don't really have such a strong feature in difference. Therefore we can conclude that within the domain of science, the performance between girls and boys are relatively balanced.