

Zhao FENG

Education Equality

An analysis of PISA 2018



Contents of the Report

Part 1: Introduction

Part 2: My audience

Part 3: The course goals

Part 4: My reflection

Part 5: My contact



Introduction

PISA is the OECD's Programme for International Student Assessment. It is a worldwide study by the Organisation for Economic Co-operation and Development (OECD) in member and non-member nations intended to evaluate educational systems by measuring 15-year-old school pupils' scholastic performance on mathematics, science, and reading. While PISA aims to assist countries' decision making and education policies improvement, the results also reveal the inequality across the world.

My project is about to investigate global education inequality by examining PISA 2018 results. The work is driven by my long-lasting interest in education and my passion for open data. I also took the opportunity to refining my skills in data visualisation and storytelling.

You can find the details of my work from the following links.

My final deliverable:

[The Leaning PISA](#)

My code:

[PISA 2018 Analysis](#)



My Audience

My ideal audience can be:

- Students majoring education, having interest in international education or psychometrics

The other audience groups can be:

- People having interest in education equality, quality education, and social justice
- Data analytics practitioner



Course Goals

Skill 7 – Communicate problems, recommendations and insights adapted to the intended target audience

Skill 10 – Plan, execute and identify resources for carrying out experiments to draw data-informed conclusions

Competence 1 – Process and analyse data in such a way that it leads to further learning and professional development

Competence 3 – Independently be able to work methodically and flexibly in various projects and processes



S7

Communicate problems, recommendations and insights adapted to the intended target audience

In my project, I took the following approaches to make the communication clear and effective.

- Setting up a scope. By providing **background**, my audience and I clearly know what the work is about and what kind of content we are expecting.
- Creating the trust. In the **motivation** part, I frankly shared my intention and my academic backgrounds, which narrows down the distance between my audience and me, also conveys a message that I am the person to do the work.
- Taking the right **tone**. The article tailors for my audience, a group of knowledgeable people. I would like it to be neutral, general, and readable. As an analytical article, a few of terms might pop up but they are still in my audience' horizon.
- Keeping it **structured**. It's challenging to read and track everything what the article is long. I divided the article into small sections, and the questions give readers contexts to think about.



S10

Plan, execute and identify resources for carrying out experiments to draw data-informed conclusions

Since the beginning, I clearly knew I want to execute my work using open and public data, with a social impact.

Data collection

- The PISA-dataset was downloaded from OECD Data. [[PISA 2018 Database](#)]
- The PISA-dataset doesn't include nations' information, like income-group and region. To supplement the dataset, I downloaded additional data from the World Bank Open Data. [<https://databank.worldbank.org/source/world-development-indicators>]
- The datasets from OECD and the World Bank are lack of geo data, so that I take world map data generated by GeoPandas, a Python library.

Obstacle and solution

- PISA data files are in SAS™ and SPSS™ formats, and they requires specific softwares.
- By using sas7bdat_converter, I convert the file from SAS format to .csv file.



C1

Process and analyse data in such a way that it leads to further learning and professional development

I conducted the processing and analysing work on Jupyter Notebook, using Python, Pandas and other libraries.

Data file types

- sas7bdat
- csv
- excel

Libraries

Data processing

- sas7bdat_converter
- pandas
- numpy

Data visualizing

- matplotlib
- seaborn
- geopandas



C1

Process and analyse data in such a way that it leads to further learning and professional development

How did I conduct the project besides using the technologies?

- Focusing my working on the the school-questionnaire data file to deliver the final deliverable on time.
- Following the instructions carefully to make sure I processed the data properly.
- Revisiting the statistical courses on Coursera to understand "strata" better.

What is my further learning:

- Diving into the student data files to generate more findings and insights.
- Bringing in the weighted value and to practicing statistical analysis.
- Redoing the work using SPSS, a software I used two decades ago.
- Presenting my work in an interactive map with enhanced data.





Independently be able to work methodically and flexibly in various projects and processes

Using *sas7bdat_converter* to convert the original data files (left) to a workable csv file (right).

SAS (TM) Data Files (compressed)

Databases including Global Competence variables were uploaded in October 2020.

- Student questionnaire data files (597 MB)
- School questionnaire data file (3.7 MB)
- Teacher questionnaire data file (17.3 MB)
- Cognitive item data file (584 MB)
- Moscow City data file (23.6 MB)
- Questionnaire timing data files (184 MB)
- Additional data file for Viet Nam (PVs and cognitive) (2.1 MB)
- Financial literacy data file (249 MB)
- Cognitive items total time/visits data file (193 MB) See PISA 2018 Technical Report Annex K for details

SPSS (TM) Data Files (compressed)

Databases including Global Competence variables were uploaded in October 2020.

- Student questionnaire data file (489 MB)
- School questionnaire data file (3.1 MB)
- Teacher questionnaire data file (12.8 MB)
- Cognitive item data file (466 MB)
- Moscow City data file (20 MB)
- Questionnaire timing data files (188 MB)
- Additional data file for Viet Nam (PVs and Cognitive) (1.1 MB)
- Financial literacy data file (206 MB)
- Cognitive items total time/visits data file (180 MB) See PISA 2018 Technical Report Annex K for details

```
pip install sas7bdat-converter
```

```
import sas7bdat_converter
```

```
# to_csv(sas7bdat_file, export_file), the pathes and the names, including the file extensions
```

```
sas7bdat_converter.to_csv('SCH/cy07_msu_sch_qqq.sas7bdat', 'SCH/results.csv')
```

```
result = pd.read_csv('SCH/results.csv')
result.tail()
```

	CNTRYID	CNT	CNTSCHID	CYC	NatCen	Region	STRATUM	SUBNATIO	OECD	ADMINMODE	...	STAFFSHORT
21898	983.0	QRT	98300802.0	07MS	64300	64387.0	QRT8787	9830000	0.0	2.0	...	0.1964
21899	983.0	QRT	98300803.0	07MS	64300	64387.0	QRT8787	9830000	0.0	2.0	...	0.6736
21900	983.0	QRT	98300804.0	07MS	64300	64387.0	QRT8787	9830000	0.0	2.0	...	-1.4551
21901	983.0	QRT	98300805.0	07MS	64300	64387.0	QRT8787	9830000	0.0	2.0	...	-0.0019
21902	983.0	QRT	98300806.0	07MS	64300	64387.0	QRT8787	9830000	0.0	2.0	...	-0.1953

The original data files are in SAS™ and SAS™ formats. Considering I haven't touched SPSS for 20 years and haven't had a copy since then, I needed to find a solution to utilise my existing skillset and technology settings. To resolve the read-in issue, I first installed *sas7bdat_converter*, then converted the SAS™ file into csv file to proceed it.



My Reflection

On the **hard skills** side, I am proud of what I have learnt and mastered:

- I have learnt how to work with unfamiliar data types and figure out by using work-around.
- One important lesson I have absorbed is to utilise the arguments. Many errors and low-efficiency are avoidable if I can include more accurate and comprehensive arguments in my lines.
- I have maximised my coding skills of *matplotlib* and *seaborn*, especially I applied 'strata' concept into the data visualisation. Furthermore, I have developed my technique to envision the geospatial data by using GeoPandas.

In term of **project management and working process**, I would like to take away with:

- Knowing your capability and your ambition. Working hard to make progress approaching your vision.
- Setting up a routine. The structure will help you to go through the two-month-individual learning period.
- Supporting your classmates and receiving supports from them. Sharing your knowledge, skills, experiences, and affections with the people surrounding you.
- Trying alternatives. Take a breath, and be creative.
- There is no silver bullet.



My Contact

Zhao FENG

Data Analyst Student at Hyper Island

- E-mail: zhao.feng.hall@gmail.com
- LinkedIn: [zhao-feng](#)
- Project repository: [PISA_2018_Analysis](#)

