# Geographical Inequality in Chinese Education

Team Edvocator: Liujing Ren, Zijing Lu, Ziyu Han, Chengyu Chen

## Introduction

News is constantly saying that teenagers in East Asia, and China, in particular, are doing well in exams. However, for all the high-achieving, hard-working students that China churns out, there are hundreds more that are left behind. Educational inequality is a complex and persistent issue in China but received poor attention from the public and media. Society has not reached a coordinated effort to tackle the issue. Data visualization is powerful in boosting public awareness, which might encourage discussions towards the issue and contribute to the trigger of social change. The user will first be familiar with data associated with geographical educational inequality such as pupil-teacher ratio, teacher's qualifications, or admission quotas universities allot to the provinces through visualizations. Then, the websites will introduce the Mathew effect in education to the users.

## Related Work

Chinese education has made remarkable achievements but the regional educational gaps also increased. The research found that the regional inequalities are deeply entrenched in China, resulting in a historical and continuing link between provinces of birth and different educational chances since the 1960s (Wang & Meiyan, 2006). On the other hand, educational inequality indicators such as the fair allocation of public education resources (Wang, 2008) and the fair access to education, especially higher education (Li, 2008) are also studied.

In the education inequality area, visualization is usually used for illustrating the correlation in the research papers. As an example, Betthäuser, Kaiser, & Trinh (2021) highlight the importance of care for cross-national inequality of education access within counties or regions in Europe and encode the data with colors. The correlation between Gini and educational outcomes in different counties of the US is also explored and visualized (Hyrne, 2015). Some NGOs also provide applications that allow users to compare education outcomes at different levels and factors (World Bank, 2021).

However, few data visualizations about the education inequality in China are found. There is only one narrative visualization on this topic which uses the line chart and bar charts to represent the educational Gini coefficients of different provinces. Our design aimed at visualizing Chinese educational inequality from different perspectives and motivated audiences to research and improve the educational inequality in China.

## Methods

Research by Paul Slovic (2015) has demonstrated that statistics do not work when humans are affected by "statistical numbing": the incapacity to wrap their heads around numbers and the lack of empathy elicited by them. Therefore, our design incorporates the storyline of the main character. Following the experience in the current education system, the readers will unfold the facts of inequality.

The Latent Class Analysis is used to prepare for the data used in the animated Sankey chart. To get students' compositions, it is necessary to divide students into different categories. Latent Class Analysis (LCA) is a statistical method proposed by sociologists Lazarsfeld and Henry in the 1950s, which is based on the assumption that latent class variables can explain the probability distributions of exogenous variables. The composition of unobservable latent variables or potential categories is estimated from a set of observable episodic variables(Finch & Marchant, 2013). Using LCA, it is possible to cluster identical individuals according to their response probabilities on the observed metrics to obtain different classifications. Moreover, compared with the K-means clustering which has high requirements for data, the LCA has more relaxed requirements for data and does not require indicators to obey normal distribution.

We analyzed the science literacy performance of 26,609 students from 665 schools in the four regions of Taipei, Hong Kong, Macau, and mainland China that participated in PISA 2015 using the LCA. Based on each student's performance on the eight sub-dimensions of science literacy, students were classified into three categories: A, B, and C. Schools were also classified into three categories: T-shaped, spindle, and pyramidal, based on the proportion of schools containing students from different categories.

**Results**

In storytelling, we used several visualizations to depict a Chinese student **Mei**'s school life, and the possible education inequality she encountered. There are four different charts we used to communicate readers with the Chinese geographical educational inequality.

Following is our informal observation of 2 testing participants via think aloud. The first and second charts (Figure 1 & 2) are based on the map, participants claimed it is easy to make comparisons between different provinces. The participants reported having an overview of the inequality.

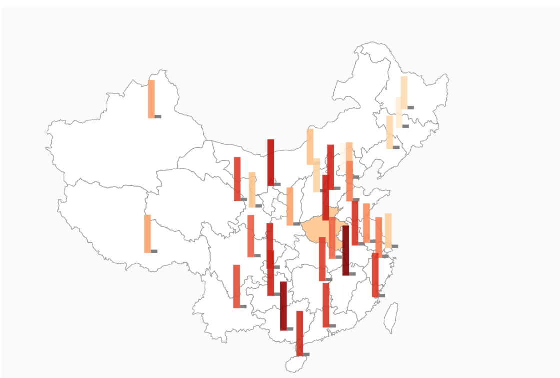


**Fig 1** The Pupil-Teacher Ratio of junior schools in different provinces in China

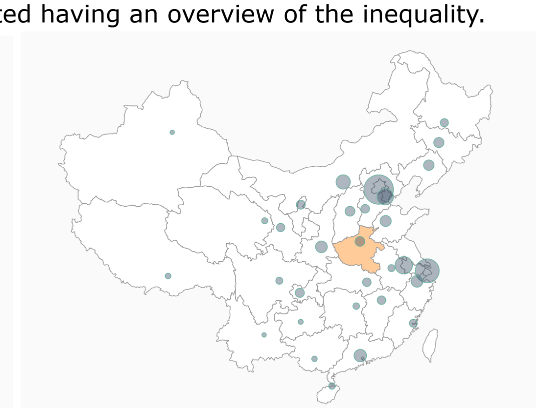


**Fig 2** The Ratio of the teachers with graduate degrees in different provinces in China

Following chart (Figure 3 & 4) demonstrates how the students are distributed between different types of schools. As an important factor of the factor, the student educational experience is showed through animations.

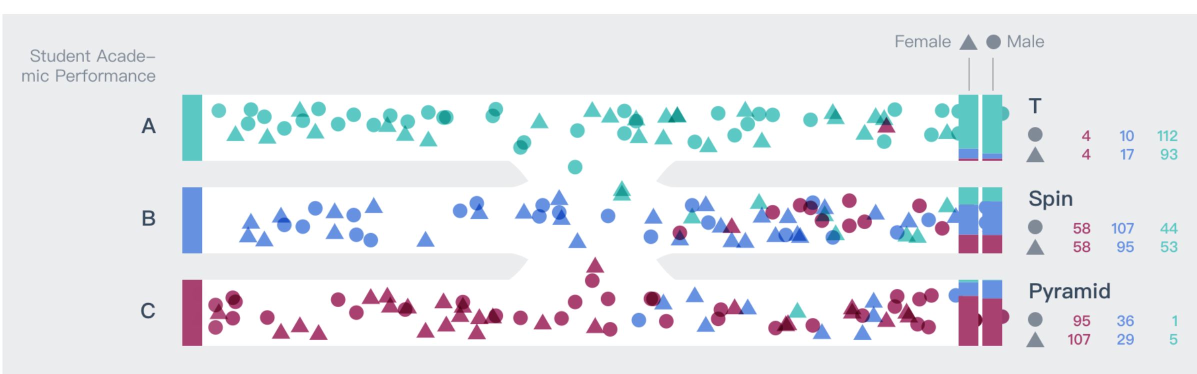



**Fig 3** Simulated data based on OECD's Programme for International Student Assessment

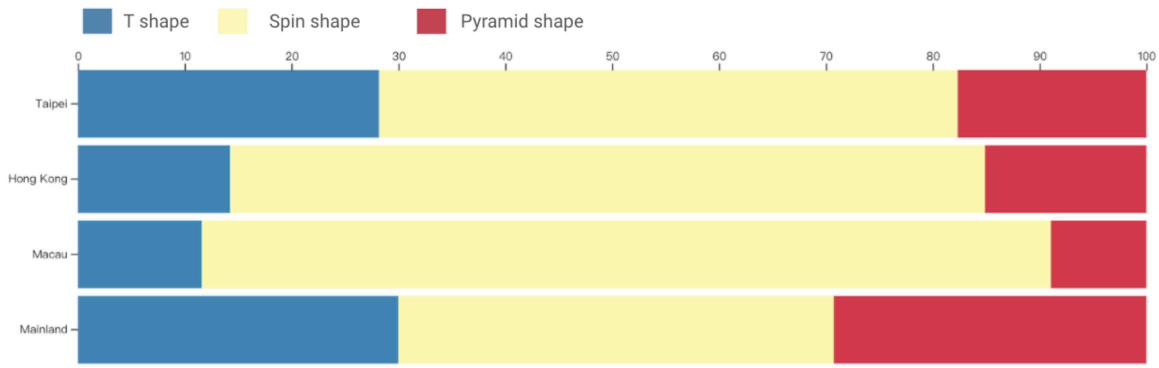
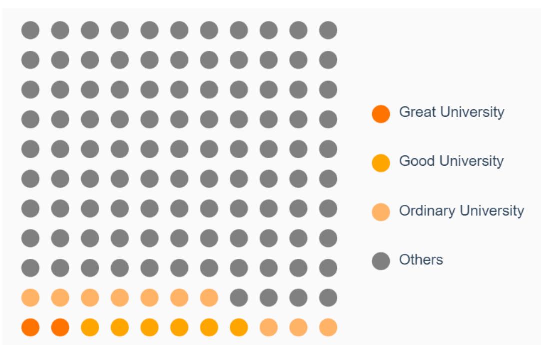



**Fig 4** Simulated data based on OECD's Programme for International Student Assessment

From the dot plot (Figure 3), the participants realized the unequal college quota to students in different provinces. They appreciated the strong comparisons between the small percentage that get into college and the large quantity that don't. Through the donut chart (Figure 4), the participants learned the different ratios of students entering different levels of universities from the circle sizes.



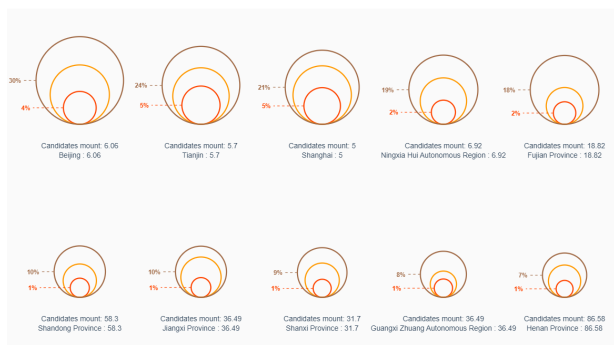




**Fig 5** The difficulty of getting into a good university in Henan

**Fig 6** Differences in acceptance rates between provinces (top five and bottom five)

They also noticed Henan province in particular due to Mei's story. The participants were able to understand concepts of T, spin, and pyramid shape school and Matthew effect. It is reported that they still want to know more about the causes, which suggests the design successfully motivates the audience to explore the education inequality issues.

**Discussion**

Through the three indicators of geographical education inequality, the audiences have learned the imbalance of resources in different provinces from the visualization. Through the narration, they've further explored the related issues and potential consequences of inequality. There is a huge gap in educational opportunities between students from more urbanized southeastern coastal areas and less urbanized inland areas. The Matthew effect is reflected in the fact that urban kids occupy more educational resources and have easier access to better higher-level schools, which results in a system that discriminates against less wealthy and well-connected citizens, thwarting social mobility at every step.

Based on the informal observation of use, the major limitation falls on a less immersive experience for audiences unfamiliar with China. Though we have considered the negative outcome of using statistics and try to foster an emotional connection by the story of Mei, the effect is not significant. We found our work is in an emerging domain called **Anthropographics** when attending IEEE VIS 2021. Anthropographics are the visualizations that represent data about people in a way that is intended to promote prosocial feelings (e.g., compassion or empathy) or prosocial behavior (Morais, Jansen, Andrade, & Dragicevic, 2021). More domain knowledge is needed, to fully empathize with the dilemma of Mei.

Background knowledge of China is needed for all readers. Inland provinces such as Henan are usually economically underdeveloped, less urbanized, and suffer from the outflow of talents. Some users unfamiliar with China fail to associate education with the social and geographical characteristics of the provinces. Even if the user is an individual with the knowledge in mind, there is a need to activate the prior knowledge for a better understanding.

Another issue is the inadequacy of analysis. The design aimed at boosting social awareness so the content mainly shows the facts and effects of educational inequality. However, in the observation of use, we found users' attention diverge from educational inequality and its consequences to its causes.

**Future Work**

For future work, we would like to develop this visualization in two aspects:

1) Conducting user research and improving the user experience in terms of empathy.
As said before, one common problem in data storytelling is failing to evoke the user's thoughts by simply representing the data. In our future improvement, we would develop more immersive visualizations by adjusting the visual encoding, supplying other material like video and images, and adding more interactions for users to explore and dive deeper into.

2) Revealing the social causes behind using visualization.
Emphasizing the analysis size of our visualizations would be another aspect that we want to develop. For example, creating visualizations correlating current examination difficulty to the economic, population data, finding historical data to see the tendency and changes of educational inequality would be effective ways to explore the social causes of education problems. Also, how to add these analyses to our storyline and the current chart would be another great challenge.

**Reference**


Betthäuser, B.A., Kaiser, C., & Trinh, N.A. (2021). Regional Variation in Inequality of Educational Opportunity across Europe. Sage Journals. https://doi.org/10.1177/23780231211019890

Finch, W. H. , & Marchant, G. J. (2013). Application of multilevel latent class analysis to identify achievement and socio-economic typologies in the 20 wealthiest countries. Journal of Educational & Developmental Psychology, 3(1).

Hyrne., K. (2015). Data visualization experiment: education and inequality.

Morais., L., Jansen., Y., Andrade, N., & Dragicevic, P. (2021). Showing Data about People: A Design Space of Anthropographics. IEEE Transactions on Visualization and Computer Graphics, Institute of Electrical and Electronics Engineers, In press. Ffhal-02931257v2f.

Shanmai Wang. (2008). Analytical framework and evaluation indicators of educational equity. Journal of Beijing Normal University: Social Science Edition, 3, 93-97.

Wang Li. (2008). Education inequality in China: problems of policies on access to higher education, Journal of Asian Public Policy, 1:1, 115-123, DOI: 10.1080/17516230801900444

Wang, Meiyan. (2006). Geography and educational inequality in China. China Economic Review. 17. 253-265. 10.1016/j.chieco.2006.04.003.

World Bank. (2021). World Inequality Database on Education. Retrieved Nov 27 from https://www.education-inequalities.org/