

COMPSCI 1710 Process Book

FALL SEMESTER 2025

https://docs.google.com/document/d/1K_8hEUijRf8wQL9pCJqkn5q0xLVhGP1lcWnQWBLXGtM/edit?usp=sharing

Team Members

Juhong Eom (juhong_eom@gse.harvard.edu)

Solo Project Agreement

Although I am embarking on this project alone, I recognize the importance of setting clear commitments, self-discipline, and accountability. The following agreement is my personal promise to uphold high standards throughout the process.

Roles and Responsibilities

As the sole member of this project, I am responsible for all aspects, including:

- Project planning and management
- Research and ideation
- Design, development, and documentation
- Testing, iteration, and final deliverables I will dedicate myself to maintaining organization, thorough documentation, and professional-level output in all phases.

Communication Protocols

- I will set regular, scheduled check-ins with myself, such as weekly reviews, to track progress and reassess goals.
- I will maintain detailed notes and progress logs for reference and reflection.
- If guidance is needed, I will proactively reach out to instructors, peers, or online communities in a timely manner.

Working Arrangements

- I will establish and stick to routine weekly work blocks, treating these as non-negotiable work sessions.
- I will set clear milestones and internal deadlines to keep the project on track.
- I will regularly update my process book to reflect advancements and challenges.

Work Allocation

- I will break the project into distinct tasks, using a to-do list or project management tool for transparency.

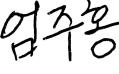
-
- I will ensure balanced attention to planning, execution, and review phases, recognizing that some tasks will require more effort and time than others.
 - I will monitor time spent on each major component to ensure no area is neglected.

Accountability

- If I miss a deadline or feel progress has stalled, I will honestly assess the reasons, adjust future plans, and document any takeaways or changes.
- I will periodically share progress with a mentor, classmate, or instructor for external feedback and accountability.
- I will recognize both my strengths and areas for improvement, and celebrate progress as a way to stay motivated.

By signing below, I commit to holding myself accountable to the standards and goals outlined in this agreement, ensuring a productive and rewarding solo project experience.

Name: Juhong Eom

Signature: 

Date: 09/11/2025

Proposal

Basic Information

Project Title

Visualizing the Relationship Between Student Empathy and Creativity

Team Members

Juhong Eom | juhong_eom@gse.harvard.edu

Team Name

Idea Weavers

Abstract

This project investigates how students' individual creativity connects with their underlying socio-emotional skills—specifically, empathy and creativity in social problem solving. Leveraging data from the PISA background questionnaires and pilot peer surveys, the study aims to visualize the complex web of relationships among what students care about, the kinds of impacts they hope to achieve, and key dispositions that shape these ambitions.

This project focuses on the interplay between empathy and creativity in social problem solving. Through correlation matrices and complementary visualizations, I examine whether higher empathy can foster creative approaches to addressing social issues, and whether this, in turn, contributes to students' overall creative capacity. Additionally, this project explores differences on the magnitude of relationships between countries. Collectively, these visual analytics offer insights into how learners' creativity formulates—illuminating the social-emotional competencies that may empower transformative, collective inquiry.

Background and Motivation

This project is motivated by my ongoing interest in learning analytics, affective learning, and student agency—specifically, how students' motivations and social-emotional competencies shape meaningful group inquiry. I am particularly intrigued by how qualities such as empathy and creative problem solving connect with students' interests, and how they differ across countries.

A key insight underlying my approach is that framing creative learning around socio-emotional skills empowers students to pursue projects that authentically matter to them. This philosophy stems from my own teaching journey. Experiencing their own potential to help and influence others becomes pure motivation for students to engage in learning, even for those low-motivated. Also, those learning experiences led them not only to participate in overall learning activities, but also to pursue their own projects. This experience reinforced the importance of connecting social responsibility with creativity. In this visualization project, I seek to support the relationship and emphasize social-emotional strengths more transparently.

I am planning to encode responses at reliable surveys conducted nationally, and make the data visualization interpretable to students and teachers. These visuals help both students and educators reflect on learning practices and explore ways to improve students' creativity within formal and informal learning environments. Prior research and tools have also informed my direction. Visualization libraries such as Plotly provide numerous way to depict information, and those such as D3 enable effective data visualization management and interactive UI/UX platform. My project builds on and extends these ideas: by applying them to student empathy, and creativity across countries, I aim to develop a visual platform that makes the convergence of diverse strengths and perspectives within a group both visible and actionable—for students, educators, and researchers alike.

This project would be meaningful. I would like to keep working on this project for the second phase to develop interactive network and flow visualizations that facilitate finding group project topics based on their interests and desired social impacts. I would like to scaffold the process by asking consecutive questions created with GenAI that help them specify their topic.

Data

PISA (Programme for International Student Assessment)

These surveys include items on students' interests in learning, career aspirations, and attitudes toward global issues such as environment, equity, and innovation. These variables provide a large-scale dataset that can be mapped into interests, academic topics, and social impacts. In the visualization, reported importance of interests will be represented by node size, academic topics will appear as tags, and desired social impacts will be expressed through color saturation. Public data are available through the OECD's PISA database: <https://www.oecd.org/pisa/data/>.

Survey on Social and Emotional Skills (SSES)

To capture various aspects of social and emotional skills, I am going to utilize the dataset of SSES 2023 data. The OECD survey is an international survey that identifies and assesses the conditions and practices that foster or hinder the development of social and emotional skills for 10- and 15-year-old students. The survey took place in 2023 in sixteen sites - six countries and ten sub-national entities around the world.

<https://www.oecd.org/en/about/programmes/oecd-survey-on-social-and-emotional-skills.html#data>

User Input

To make the platform more interactive, I am going to build pages that users can take the test and see the result, comparing with others. This feature will make the data visualization richer, and provide users the opportunity to interpret the data in their personalized context.

Preprocessing Plan

I will review and categorize PISA variables in comparison with SSES data. Then, I will encode countries as colors, population units as sizes, and locate edges based on students' level of empathy and creativity.

Map

Audience

Audience Profile

The primary audience for this visualization is primary and middle school teachers, who want to facilitate creative learning experiences and foster socio-emotional skills. Specifically, the project targets teachers from those countries whose students lack creativity.

Secondary audiences include:

- 11- to 15-year-old Students interested in exploring creativity and socio-emotional skills of peers within their age group, either within their school or more broadly.
- Teachers who wish to gain insights into students' creativity and socio-emotional skills for more effective project guidance.

Overall, the visualization aims to empower both students and educators by making visible the connections between individual motivation, social-emotional skills, collective interest patterns, and prospects for social impact, thereby supporting more informed, inclusive, and self-directed group topic development.

Prior Knowledge and Visualization Literacy

Teachers, equipped with data literacy and facilitation skills, are the primary audience for this data visualization. The degree of 'how' creative or socially emotional might be different depending on individuals' teaching experience or backgrounds. Therefore, it would be crucial to provide clear criteria provided by the dataset codebook, and provide enough comparative context.

As the secondary audiences encompass students, it is important to make the data visualization accessible and interpretable at their cognitive developmental level.

Project-based and problem-based learning are central learning practices to foster creativity and socio-emotional skills. They are usually implemented at the end of upper elementary (grades 5–6) and through middle school (grades 7–9). Therefore, the secondary audience of

this project entails those age groups. The visual literacy of the audience can be inferred from the Korean National Curriculum for primary and lower secondary education (National Curriculum Information Center, 2022). Students are expected to develop the following knowledge and skills:

[Elementary School (Grades 5-6)]

Mathematics (Graphical and Data Literacy):

- Understanding the concept of averages, collecting data, calculating and interpreting averages.
- Collecting data and representing it with band (stacked bar) graphs and pie charts, and interpreting results.
- Setting inquiry problems, gathering and organizing appropriate data, representing it with suitable graphs, and interpreting the outcomes.
- Expressing and comparing the likelihood of events in words and numbers.
- Using data for basic probability reasoning and making decisions based on likelihoods.

Practical Course and Computer Education (Computational and Data Literacy):

- Exploring case studies of using computers to solve everyday problems, and expressing algorithms for everyday solutions in various ways.
- Experiencing how to command computers and writing simple programs to address given problems.
- Collaboratively developing programs to solve real-life problems and sharing the results with others.
- Understanding the characteristics of digital and analog data, and exploring types and forms of data that can be used in artificial intelligence (AI).
- Experiencing the process of creating AI and exploring the impacts of AI on society.

[Middle School (Grades 7–9)]

Mathematics (Statistical and Data Literacy):

-
- Understanding and interpreting various types of graphs (scatter plots, histograms, box plots).
 - Calculating and interpreting measures of central tendency (mean, median, mode) and variability (range, deviation).
 - Constructing and analyzing frequency tables and distributions from real-life data.
 - Understanding and reasoning with probabilities, conducting simple experiments, and interpreting random events.
 - Analyzing survey results, representing them visually, and making data-driven decisions.
 - Interpreting and critically evaluating statistical information in social, scientific, and media contexts.

Information/Computer Class (Digital, Computational, and Data Literacy):

- Collecting, processing, and analyzing large-scale and diverse real-world data.
- Programming basic algorithms to handle data, visualize outcomes, and model real-world situations.
- Understanding basic principles of artificial intelligence (AI), machine learning, and their impact on society.
- Using spreadsheet and data analysis tools, constructing comparative visualizations, and interpreting digital data results.
- Collaborating in solving data-driven problems and sharing findings using digital platforms.

Given this curricular background, middle school students are generally able to read, interpret, and construct a wide variety of visual data representations (bar, pie, line, scatter, histogram, box plot, etc.), understand statistical summaries, and engage in introductory data-driven exploration using digital tools. With these curricular competencies, both upper elementary and middle school students are well-positioned to engage in accessible, age-appropriate exploratory data visualizations.

Level of Detail and Customization

Information will be presented at multiple levels, which the audience can choose to explore:

-
- Individual Level: The audience can see visualizations focused solely on those newly participated in the survey.
 - National Level: The audience can view aggregated data from their own country.
 - International Level: As capacity allows, the audience may compare results across different countries.
 - Age Filtering: Future versions may allow users to filter data by age or grade for more tailored comparisons.

This flexible detail granularity empowers the audience to contextualize the dataset and start with what is most comfortable and gradually expand their perspective, supporting both metacognitive learning and collaborative awareness.

Expected Questions

Self-Reflection

- How do my empathy and creativity influence what interests me or motivates me to take action?
- Is my empathy connected to my creativity?

Peer level

- How does my way of thinking (empathetic, creative, etc.) align with or differ from my peers?
- **What is the creativity level of peers who have the same level of empathy as I do?**

National level

- To what extent is students' creativity associated with their empathy?
- What is the distribution of student creativity and empathy in my country?
- How do my own, or my students' levels of creativity and empathy compare to those of others in my country?
- Would integrating socio-emotional skills enhance students' creative learning experiences?

Global level

- **How do levels of creativity and empathy among 15-year-olds vary around the world?**
- **How do students in my country perform in creativity and socio-emotional skills compared to their peers in other countries?**
- What common characteristics do countries with high levels of creativity and socio-emotional skills share?

Data Plan

For this project, I plan to use and synthesize several sources of data to represent the relationship between students' empathy .

Data Sources

- Primary Research - **PISA 2022 Student Questionnaire Data:** To explore how students engage in self-directed school projects and to identify their expressed interests and concerns about social impacts.
- Secondary Research: Additional insights will be drawn from sources such as the OECD dataset of Survey on Social and Emotional Skills (SSES).
- Custom Survey/Platform Input: The platform will allow direct student input to capture measured creativity and self-reported empathy in real time.

Envisioned Data and Types

Index	Datatype	Example/Description
Student Name / ID	Categorical	Each student's name or anonymized identifier
Region / Country	Categorical	South Korea, USA, other country, or specific city/region
Age	Categorical	11-15 years, others(can be grouped or retained at year-level)
Creativity	Ordinal/Quantitative	Likert scale (e.g., 1-10) indicating measured level of creativity
Empathy	Ordinal/Quantitative	Likert scale (e.g., 1-10) indicating self-reported level of empathy

Input Source	Categorical	Platform input as survey response
--------------	-------------	-----------------------------------

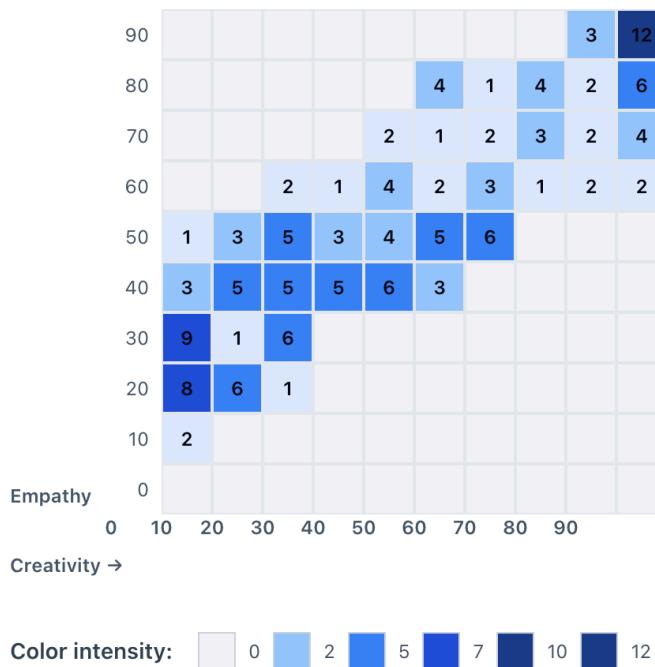
**Bolded data indexes are the primary focus for visualization and analysis.*

Where possible, I will harmonize scales (such as using a 10-point Likert scale) so that students can articulate detailed gradations of creativity or empathy. The platform will also enable filtering and aggregation by region, age, and interest, supporting both personal and comparative exploration.

Initial Visualizations

1. Heatmap of Student Distribution by Score Ranges (Claude)

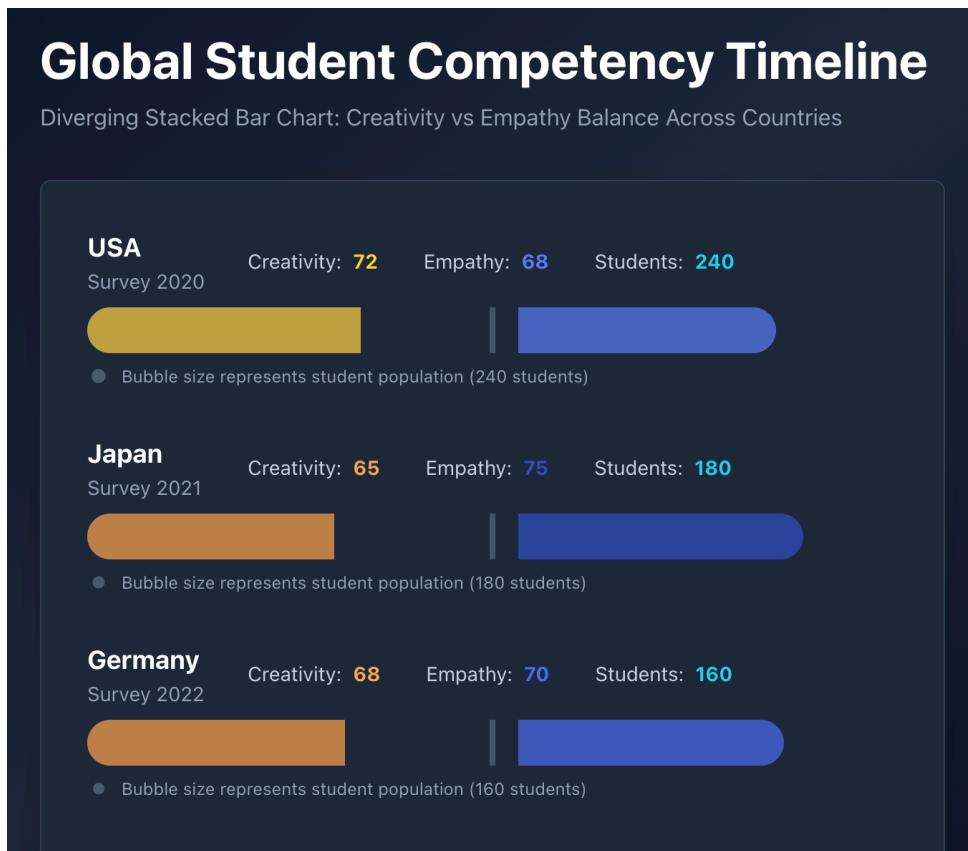
Heatmap: Student Distribution by Score Ranges



This visualization displays the student distribution by their levels of creativity and empathy. It could answer questions such as "To what extent is students' creativity associated with their empathy?" or "What is the distribution of student creativity and empathy in my country?". Also, it is possible to locate new input on the plane. On the other hand, it makes

me worried that students as audiences might be overwhelmed in the case that their level of creativity or socio-emotional skills are low.

2. Global Student Competency Timeline (Claude)



The question addressed in this example is one at a wider community level - "How do levels of creativity and empathy among 15-year-olds vary around the world?". The question is directly related to the research question, though it cannot provide answers to other questions. It is hard to know the distribution of students as only the national average is provided. Additionally, this timeline is to get information not about each country's specific level of creativity or empathy, but about subjective differences between countries. However, it is not easy to compare them intuitively due to spaces between bars. Additionally, it makes me worried that the obviousness of 'the longer the better' would make education more competitive to just comparing the level of empathy or creativity, rather than focusing on what we can do to foster those skills.

3. Global Competency Galaxy (Vibe Coding, Claude)



This constellation is compelling to answer the question of wider community level - "How do levels of creativity and empathy among 15-year-olds vary around the world?". It is positive that it does not induce competition, using visual channels that are less linear. The creativity is depicted as the distance between planets and the *center sun*-Global. The size of each planet depicts each country's population size. Also, the saturation intensity shows stronger competencies of empathy. On the other hand, it could be less intuitive to understand the graphic. It would be hard to compare others between countries as it is hard to gauge differences in saturation and distance. Additional circles such as dart targets would be helpful to compare. Additionally, it is hard to answer the original research question to show the association between student creativity and empathy.

I decided to focus on one of the research questions: *How do students' level of creativity and empathy vary across countries?*. Compared to more complex or abstract inquiries, this approach keeps the visualization intuitive and relevant for the intended audience. By transforming abstract ideas about "creativity" and "empathy" into accessible, interactive visuals, I believe the sketch makes it easier for teachers to design creative and meaningful project-based learning experiences.

Data

Data Source

This project's aim is to understand the distribution of students' creativity and empathy as one of socio-emotional skills. It would aid teachers to determine whether empathy indirectly promotes overall creativity through its effect on social problem solving creativity.

PISA 2022

- Source: I plan to use the data from PISA 2022. PISA is a triennial survey of 15-year-old students around the world that assesses the extent to which they have acquired key knowledge and skills essential for full participation in social and economic life (OECD, 2024a).
- Dataset: I used the student questionnaire data file and the cognitive item data file. These files include data from questionnaire items and creative thinking items.
 - The PISA target population consists of students who are between 15 years 3 months and 16 years 2 months old at the time of assessment, are enrolled in school, and have completed at least six years of formal schooling. This definition applies regardless of the type of institution (academic or vocational, public, private, or foreign schools within the country), and includes both full-time and part-time students (OECD, 2024a).
 - Creative Assessment
 - This test assessed 15-year-old students' capacity for creative thinking, defined as the competence to generate, evaluate, and improve original and diverse ideas. The creative thinking units were organized into five mutually exclusive 30-minute blocks or clusters, which were rotated according to an integrated design (OECD, 2024b). Approximately 28% of the PISA student sample was administered the creative thinking assessment—these students spent one hour on

creative thinking test items, with the remaining hour devoted to one of the other core domains (mathematics, reading, or scientific literacy) (OECD, 2024a).

- There were four domain contexts in the PISA 2022 Creative Thinking assessment: written expression, visual expression, social problem solving, and scientific problem solving. Additionally, there were three ideation processes assessed: generating diverse ideas, generating creative ideas, and evaluating and improving ideas.
- File format: SAS and SAS7BDAT file (CY08MSP_STU_QQQ.SAV, CY08MSP_CRT_COG.SAS7BDAT)
- Usage Restrictions:
 - Public Use Files containing individual unit record data in SAS and SPSS formats are available for downloading for all countries participating in the PISA 2022 Survey.
 - I completed a form to download and access the data. The terms of use are as follows:

Terms of Use

1. Data Access:

- The PISA Data must not be distributed, disclosed, or made available to any third party.
- Access to the Dataset also requires compliance with the terms and conditions available on www.oecd.org or any other platform used to make the PUF available.
- The OECD reserves the right to remove access to the PUF on its website at any time.

2. Data Linking:

- No attempts will be made to link the Dataset to other information that enables the identification of participants or schools.
- Any attempt to re-identify participants or schools from the Dataset is strictly prohibited.
- The User(s) agree(s) to notify the OECD if there is any reason for the User(s) to believe that the Dataset is subject to re-identification risks.

3. Data Reporting:

-
- The PISA S Data will not be reported or published in a report in a manner that could allow for the identification of individual participants or schools' responses.
 - Aggregated data may be included, provided that individual participants or schools' responses cannot be identified.
 - The OECD will be acknowledged as the source of the PISA Data in all uses of the Dataset (e.g. presentations, publications, research proposals), as follows: "Programme for International Student Assessment (PISA) Organisation for Economic Co-operation and Development (OECD), Paris".

4. Data Security Measures:

- All User(s) will implement and maintain appropriate technical and organisational measures to protect the Dataset from unauthorised access, use, disclosure, alteration, or destruction.
- Users must immediately notify the OECD of any actual or suspected security breaches or unauthorised access to the Dataset, by email to edu.pisa@oecd.org, with a copy to infosec@oecd.org and dpo@oecd.org.

5. Disclaimers:

- Use of the Dataset is subject to the disclaimers applicable to OECD data, which can be consulted at: www.oecd.org/en/about/terms-conditions/oecd-disclaimers

6. Liability:

- The User(s) shall be responsible for any claims, losses, damages, costs or liabilities, including those of third parties, arising out of or in connection with any breach of these Terms of Use herein. The Institution assumes full responsibility for any breach of these Terms of Use by the requestor and any other persons who are permitted to access and use the Dataset by the Institution or the requestor. The OECD reserves the right to seek any recourse or remedies available to it for any breach of the Terms of Use against any User(s), as the OECD deems appropriate.

7. Applicable law and dispute resolution:

- Given the status of the OECD as an international organisation, the rights and obligations of the OECD and the rights and obligations of the User(s) shall be governed exclusively by these Terms of Use.
- Any dispute, controversy or claim arising out of or relating to the interpretation, application or performance of the Terms of Use, including their existence, validity or termination, shall be settled by final and binding arbitration in accordance with the Permanent Court of Arbitration (PCA) Arbitration Rules 2012. The number of arbitrators shall be one. The language to be used in the arbitral proceedings shall be English. The place of arbitration shall be Paris (France). The parties expressly renounce their right to seek the annulment or setting-aside of any award rendered by the arbitral tribunal, or if this renunciation is not legally possible, the parties expressly agree that if an award rendered by the arbitral tribunal is annulled, the jurisdiction ruling on the annulment proceedings cannot rule on the merits of the case. The dispute will therefore be settled by new

arbitral proceedings in accordance with this clause. The parties expressly renounce their right to seek the annulment or setting-aside of any award rendered by the arbitral tribunal.

8. Privileges and Immunities:

- Nothing herein shall constitute or be considered to be a limitation upon or a waiver of the privileges and immunities of the OECD, which are specifically reserved.

9. Modification of Terms of Use:

- The OECD may amend the Terms of Use herein at any time at its sole discretion, by providing notice to the requestor. A User should immediately cease to use or access the Dataset if they cannot comply with the amended Terms of Use.

Data Protection Notice

The OECD will use the personal data collected in this form to maintain a record of those who have access to the Dataset and contact them if necessary. These records will be retained for a maximum period of six years and will not be shared outside the OECD, except where it is necessary to share the personal data with external advisors for the purpose of auditing, security, investigation or legal claims, including in the event of unlawful activity and/or violation of these Terms of Use.

Under the OECD Data Protection [Rules](#), you have rights to access and rectify your personal data, as well as to object to its processing, and request erasure in certain circumstances. To exercise these rights in connection with your use of this website please contact edu.pisa@oecd.org. If you have further queries or complaints related to the processing of your personal data, please contact the [Data Protection Officer](#) [dpo@oecd.org]. If you need further assistance in resolving claims related to personal data protection you can contact the [Data Protection Commissioner](#) [dpc@oecd.org].

Data Inventory Creation

Index	Data type	Value ranges	Example/Description
CNT	Categorical (CHAR)	(3 characters)	Country code 3-character KOR (Korea), ARG (Argentina), etc.
CNTSTUID	Ordinal (NUM)	800001.0-86007492.0 (8 numbers)	Intl. Student ID
STRATUM	Categorical (CHAR)	KOR01 - KOR05	Stratum ID 5-character(cnt+original stratum ID) KOR01 (Lower Secondary/General), KOR02 (Upper Secondary/General),

			KOR03 (Upper Secondary/Vocational), etc.
ST001D01 T	Categorical (INT)	7-13	Student International Grade 7 (Grade 7), 8 (Grade 8), etc.
ST004D01 T	Categorical (INT)	1, 2	Student Gender 1 (Female), 2(Male)

I selected several questions relevant to my research question. First, I selected demographic data in order to filter for responses from the Republic of Korea within each file. Since only 28% of students participated in the creativity test, I needed to cross-check two datasets. To address this issue, I used each student's ID to match their responses across the datasets.

Index	Datatype	Value	Example/Description
ST311Q01JA	Categorical, Quantitative (NUM)	1: "Strongly disagree", 2: "Disagree", 3: "Neither agree or disagree", 4: "Agree", 5: "Strongly agree"	Agree/disagree: I do not care what happens to other people.
ST311Q02JA			Agree/disagree: I can sense how others feel.
ST311Q03JA			Agree/disagree: It is important to me that my friends are okay.
ST311Q04JA			Agree/disagree: I can see situations from my friends' perspectives.
ST311Q05JA			Agree/disagree: I ignore the feelings of others.
ST311Q06JA			Agree/disagree: I am more compassionate than most people I know.
ST311Q07JA			Agree/disagree: It is difficult for me to sense what others think.
ST311Q08JA			Agree/disagree: I predict the needs of others.
ST311Q09JA			Agree/disagree: I get upset if bad things happen to other people.
ST311Q10JA			Agree/disagree: I understand what others want.

Second, I selected questions regarding students' self-reported empathy, which is an aspect of social-emotional development. These data are included in the CY08MSP_STU_QQQ.SAV file. Students' ratings of their agreement with statements about various behaviors indicative of empathy were combined into an 'Empathy' index. This scale utilized a within-construct matrix sampling design, and each of the 10 items had five response options (OECD, 2024b). Some data were missing due to extreme straightlining or, for comparison purposes, because there were not enough responses (i.e., fewer than three

responses for the scale). In both cases, scale scores were coded as "99" in the SPSS file and ".M" in the SAS file. Upon reviewing the data, I found that in Korea, 0.7% of responses were identified as extreme straightlining and 1.4% as not having enough responses, out of a total of 6,454 respondents.

Index	Ideation Process	Data type	Value ranges	Example/Description
DT400Q01C	Generate diverse ideas			SaveTheBees Q01: suggest three different ideas to raise awareness about the importance of bees.
DT400Q02C2	Generate creative ideas			SaveTheBees Q02: Suggest an original idea that could be used to increase awareness about the importance of bees.
DT400Q03C2	Evaluate and improve ideas	Numerical, Quantitative	0: "No credit", 1: "Partial credit", 2: "Full credit"	SaveTheBees Q03: Describe an original idea to make the event of the "Save the Bees" club more engaging for participants.
DT500Q01C	Generate diverse ideas			Library Accessibility Q01: Describe 3 different ideas for how to improve the wheelchair accessibility of the library
DT500Q02C2	Evaluate and improve ideas			Library Accessibility Q02: Describe one original idea for a modification or feature that could be added to the ramp that will improve the ability of wheelchair users to access books in the library.
DT630Q01C2	Evaluate and improve ideas			Carpooling Q01: Think of an original way that the initiative to promote carpooling can be expanded and improved upon.

Third, I selected questions related to students' measured creativity from the CY08MSP_CRT_COG.SAS7BDAT file. Among these, I focused on the social problem solving domain. Social problem solving involves understanding different perspectives, addressing the needs of others, and finding innovative and functional solutions for the parties involved. This domain comprises 3 units and 5 items.

The creative thinking scale for PISA 2022 is constructed as a bounded scale ranging from 0 to 60 points. The maximum sum score of 60 represents the total number of points possible

in a hypothetical test containing all 32 items within the creative thinking item pool. Student scores on the creative thinking scale can, therefore, be interpreted as their estimated score (i.e., the sum of their partial and full credit responses) if they were to take a test covering all 32 creative thinking items.

Data Exploration

First of all, I extracted the columns with R.

```
pisa = read_sav('/Users/juhong/Desktop/CS1710/Data/PISA2022/CY08MSP_STU_QQQ.SAV')
pisa_sub <- pisa%>%select(CNTSTUID, CNT, ST311Q02JA, ST311Q03JA, ST311Q08JA, ST311Q01JA, ST311Q04JA,
ST311Q05JA, ST311Q06JA, ST311Q07JA, ST311Q09JA, ST311Q10JA)

write_sav(pisa_sub, '/Users/juhong/Desktop/CS1710/Data/PISA2022/CY08MSP_STU_QQQ_SUB2.SAV')

pisa_cog <- read_sav('/Users/juhong/Desktop/CS1710/Data/PISA2022/CY08MSP_CRT_COG.SAV')
pisa_cog_kor <- pisa_cog%>%filter(CNT=='KOR', na.rm=TRUE)
crt_sub <- pisa_cog_kor%>%select(CNT, DT400Q01C, DT400Q02C2, DT400Q03C2, DT500Q01C, DT500Q02C2,
DT630Q01C2, CNTSTUID)

write.csv(crt_sub, file='/Users/juhong/Desktop/CS1710/Data/PISA2022/crt_sub.csv')
```

Then, I merged them using the CNTSTUID as a key.

```
�
pisa_sub = read_sav('/Users/juhong/Desktop/CS1710/Data/PISA2022/CY08MSP_STU_QQQ_SUB2.SAV')
pisa_kor <- pisa_sub%>%filter(CNT=='KOR', na.rm=TRUE)
pisa_kor%>%head()

pisa_kor_selected <- pisa_kor%>%select(CNTSTUID, CNT, ST311Q02JA, ST311Q03JA, ST311Q08JA)%>%drop_na()

pisa_kor_selected%>%head()

crt_kor <- read.csv('/Users/juhong/Desktop/CS1710/Data/PISA2022/crt_sub.csv')
crt_kor %>% head()

merged_data <- merge(pisa_kor, crt_kor, by="CNTSTUID")
merged_data_clean <- merged_data%>%drop_na()
```

	CNTSTUID	CNT.x	ST311Q02JA	ST311Q03JA	ST311Q08JA
		Country code 3-character	Agree/disagree: I can sense how others feel.	Agree/disagree: It is important to me that my friends are o...	Agree/disagree: I predict the needs of others.
1	41000001	KOR	NA	NA	4
2	41000002	KOR	NA	4	NA
3	41000003	KOR	NA	NA	4
4	41000004	KOR	4	NA	NA
5	41000005	KOR	4	3	4
6	41000006	KOR	NA	NA	4
7	41000007	KOR	NA	NA	3
8	41000008	KOR	NA	NA	3
9	41000009	KOR	4	NA	NA
10	41000010	KOR	5	5	5
11	41000011	KOR	NA	NA	2
12	41000012	KOR	4	NA	5
13	41000013	KOR	NA	4	5
14	41000014	KOR	4	NA	4
15	41000015	KOR	2	NA	NA
16	41000016	KOR	NA	4	4
17	41000017	KOR	4	4	NA
18	41000018	KOR	NA	5	5
19	41000020	KOR	2	NA	NA
20	41000021	KOR	3	4	NA
21	41000022	KOR	NA	NA	4
22	41000023	KOR	NA	NA	NA
23	41000024	KOR	NA	4	NA
24	41000025	KOR	NA	NA	5
25	41000026	KOR	NA	4	NA
26	41000027	KOR	NA	NA	5
27	41000028	KOR	5	5	5

Showing 1 to 27 of 6,454 entries, 20 total columns

Because students were divided into multiple clusters, some students participated in specific combinations of questions about empathy and creativity. I also analyzed the OECD reports for PISA 2022. According to the OECD report on creativity assessment, Korea ranks second among the highest-performing systems in creative thinking, with a mean creative thinking score of 38.

Data Cleaning

```

26 #diverse questions
27 beeQ1 <- merged_data %>% drop_na(DT400Q01C, ST311Q02JA)
28 beeQ1_unique <- beeQ1[!duplicated(beeQ1), ]
29 na_cols <- colSums(is.na(beeQ1_unique)) > 0
30 beeQ1_cleaned <- beeQ1_unique[, !na_cols]
31

```

I dropped all the rows that contain NA for certain columns that I selected. Then, I removed duplicates with the R library, and dropped any columns that contain NA values.

Data	
beeQ1_1	338 obs. of 21 variables
beeQ1_1_cleaned	338 obs. of 7 variables
beeQ1_1_unique	338 obs. of 21 variables
beeQ1_2	341 obs. of 21 variables
beeQ1_2_cleaned	341 obs. of 7 variables
beeQ1_2_unique	341 obs. of 21 variables
beeQ1_3	351 obs. of 21 variables
beeQ1_3_cleaned	351 obs. of 7 variables
beeQ1_3_unique	351 obs. of 21 variables
merged_data	6454 obs. of 21 variables

I cross-examined the answer for the first question from unit 1 Save The Bee, and those for three selected empathy questions.

▲	X.1	CNTSTUID	CNT.x	ST311Q08JA	X	CNT.y	DT400Q01C
1	5	41000005	KOR		4	5	KOR
2	6	41000006	KOR		4	6	KOR
3	34	41000035	KOR		3	34	KOR
4	52	41000054	KOR		4	52	KOR
5	66	41000069	KOR		4	66	KOR
6	74	41000077	KOR		4	74	KOR

▲	X.1	CNTSTUID	CNT.x	ST311Q02JA	X	CNT.y	DT400Q01C
1	5	41000005	KOR		4	5	KOR
2	15	41000015	KOR		2	15	KOR
3	32	41000033	KOR		4	32	KOR
4	35	41000036	KOR		2	35	KOR
5	52	41000054	KOR		5	52	KOR
6	80	41000083	KOR		4	80	KOR

There were some students who answered more than two questions among the selected empathy questions. For instance, the students 41000005 and 41000083, who answered to the first question of the bee unit, responded to both ST411Q08JA and ST311Q02JA. As they were randomly selected, it was inevitable to find those cases. Therefore, I decided to narrow down the question type for creativity and cross-examine those with empathy questions one by one.

Final selection for questions are as follows:

X axis:

Index	Datatype	Value	Example/Description
ST311Q02JA	Categorical, Quantitative (NUM)	1: "Strongly disagree", 2: "Disagree", 3: "Neither agree or disagree", 4: "Agree", 5: "Strongly agree"	Agree/disagree: I can sense how others feel.
ST311Q03JA			Agree/disagree: It is important to me that my friends are okay.
ST311Q07JA			Agree/disagree: It is difficult for me to sense what others think.
ST311Q08JA			Agree/disagree: I predict the needs of others.

Y axis:

Index	Ideation Process	Data type	Value ranges	Example/Description
DT400Q01C	Generate diverse ideas			SaveTheBees Q01: suggest three different ideas to raise awareness about the importance of bees.
DT500Q01C				Library Accessibility Q01: Describe 3 different ideas for how to improve the wheelchair accessibility of the library
DT400Q02C2	Generate creative ideas		0: "No credit", 1: "Partial credit", 2: "Full credit"	SaveTheBees Q02: Suggest an original idea that could be used to increase awareness about the importance of bees.
DT400Q03C2				SaveTheBees Q03: Describe an original idea to make the event of the "Save the Bees" club more engaging for participants.
DT500Q02C2	Evaluate and improve ideas			Library Accessibility Q02: Describe one original idea for a modification or feature that could be added to the ramp that will improve the ability of wheelchair users to access books in the library.
DT630Q01C2				Carpooling Q01: Think of an original way that the initiative to promote carpooling can be expanded and improved upon.

However, I open the possibility to compare between different domains of creativity.

Reflection

While delving into the data, there were two challenges. First, it was difficult to find data on students' interests and their desired social impacts. Although I found some questions related to desired social impacts in the PISA dataset, it was much harder to locate data on students' own interests. There are some surveys conducted by private organizations, but I should count into the probability of biased data. I even discussed this with my Statistics professor, however, we were unable to find the specific data that I wanted. I wondered why there is so little official data focused on students' interests at their age. Since the PISA dataset is one of the most reliable data sources available, I ultimately decided to base my analysis on it, even though this required me to revise my research question. Additionally, as the PISA is conducted with 15-year-olds, its findings may not be meaningful for the 4th to 6th graders, who were my initial target audience. I might need to reconsider the group I focus on.

The second challenge was to find design guidelines for data visualization intended for children. Given their cognitive developmental stage, I expected to find such guidelines, perhaps based on students' achievements on math for interpreting graphics and data literacy for understanding data. However, to my surprise, I discovered that very little such guidance exists, though numerous dashboards have already been developed for children.

Reference

The National Curriculum Information Center. (2022). *The Korean national framework of math for the elementary and secondary curriculum.* <https://ncic.re.kr>

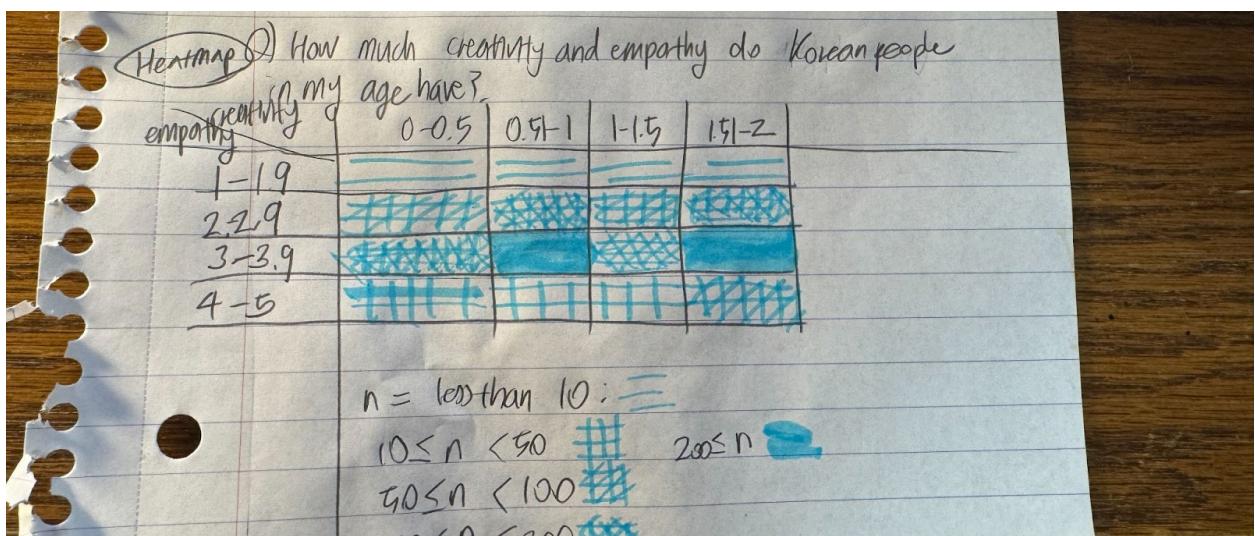
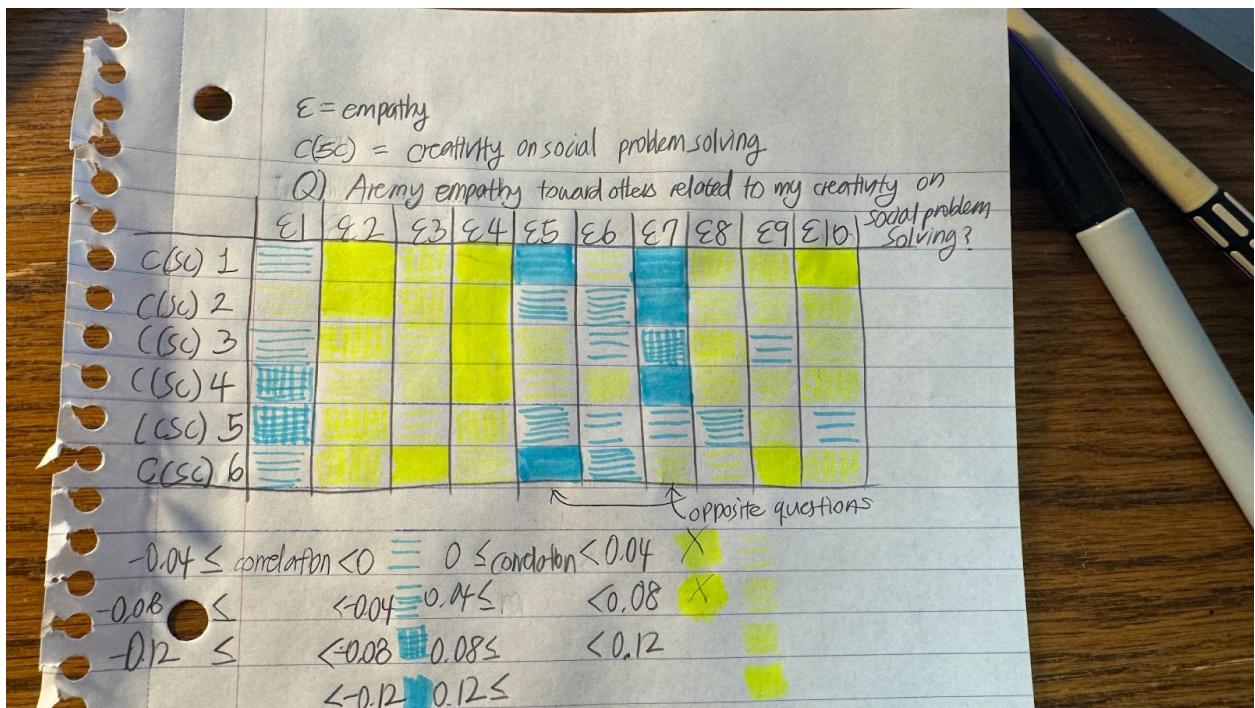
OECD. (2024a). *PISA 2022 RESULTS (VOLUME III).*

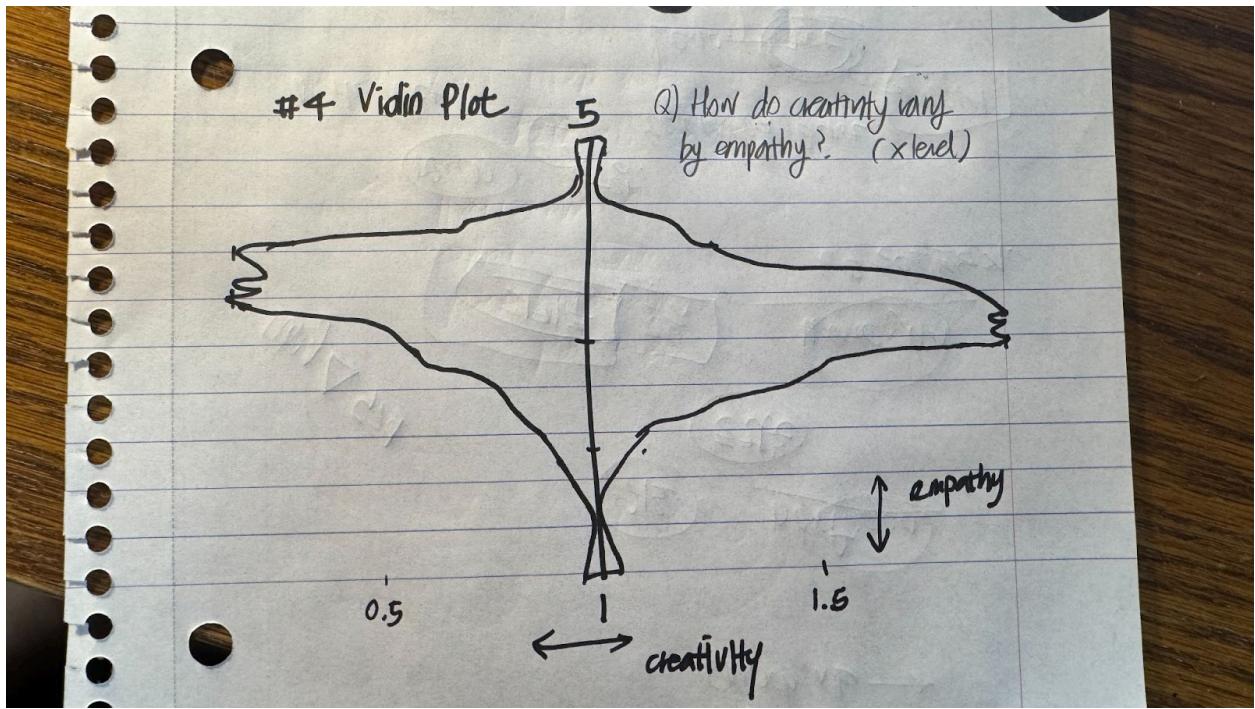
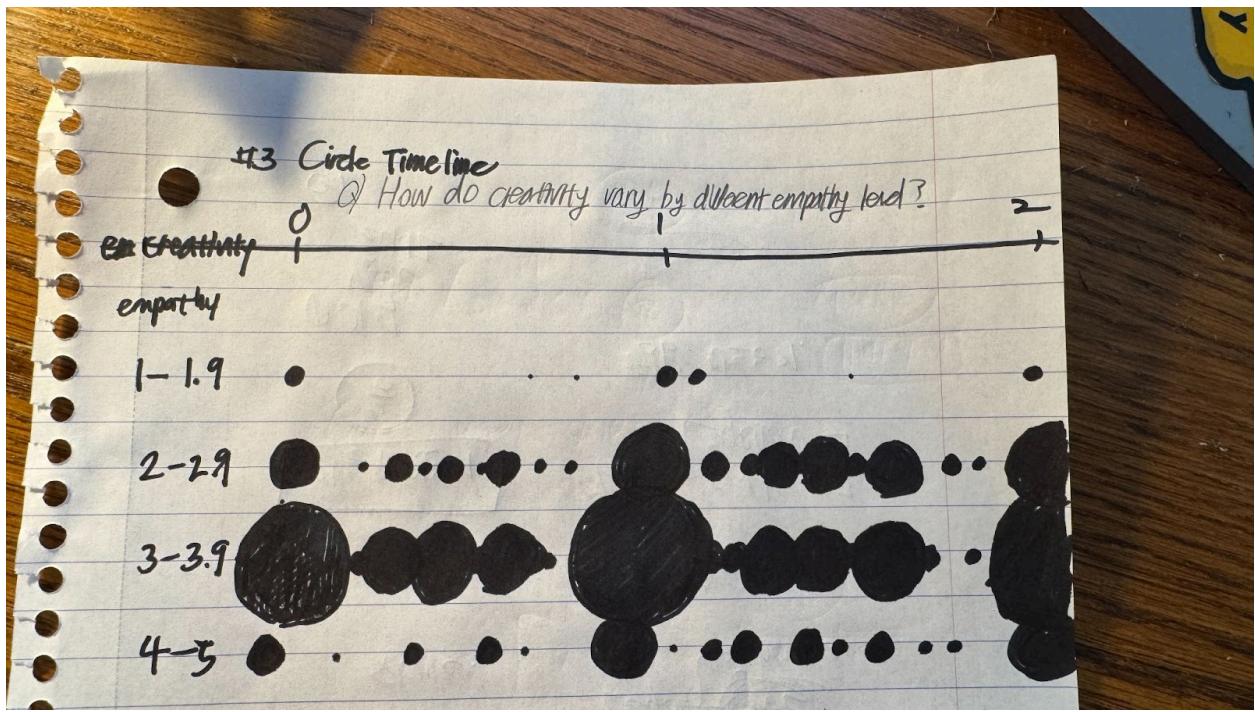
https://www.oecd.org/en/publications/pisa-2022-results-volume-iii_765ee8c2-en.html

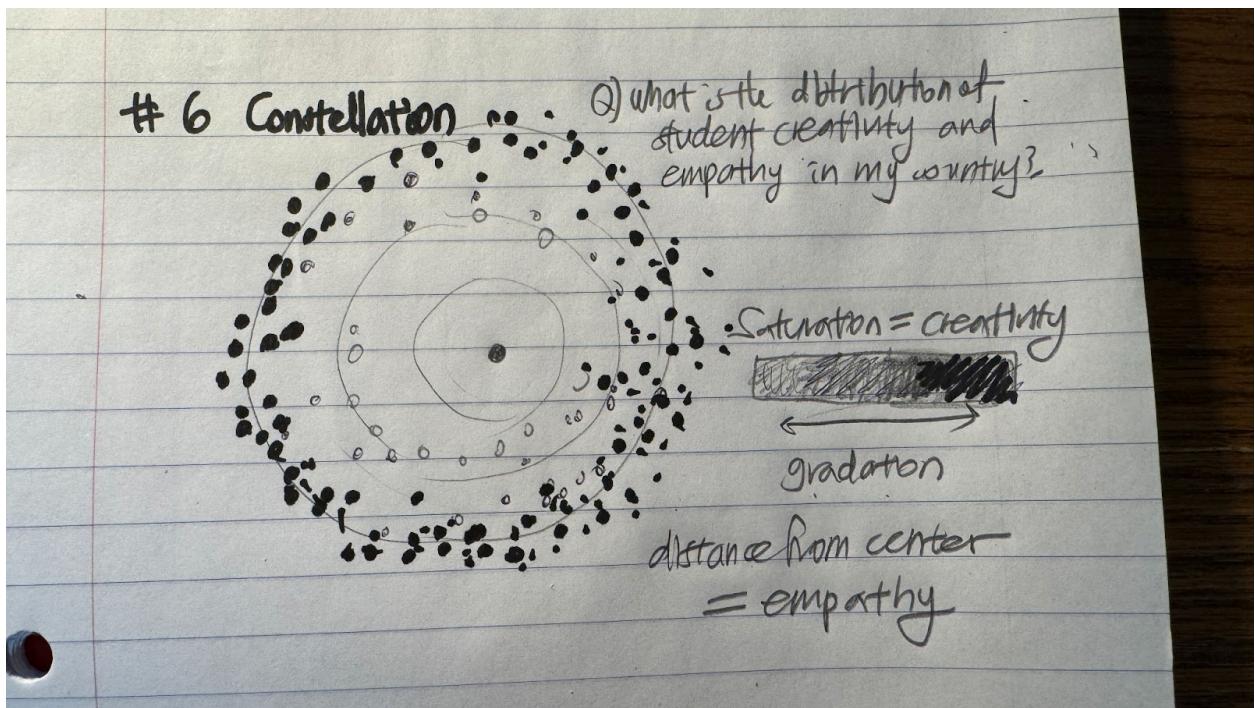
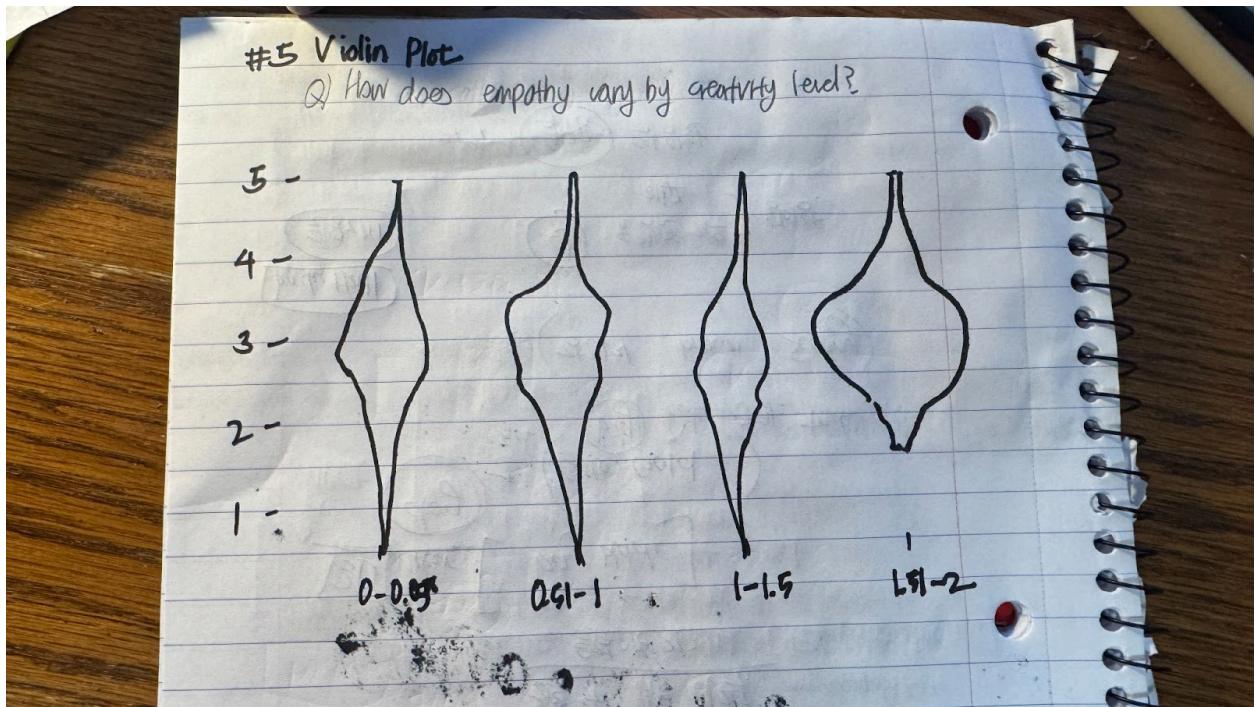
OECD. (2024b). *PISA 2022 Technical Report.*

https://www.oecd.org/en/publications/pisa-2022-technical-report_01820d6d-en.html

Sketch







Decide

Affinity Diagramming

- Question ID
 - 1: What is the creativity level of peers who have the same level of empathy as I do? (Comparison)
 - 2: What is the distribution of student creativity and empathy? (country/world)
 - 3: To what extent is students' creativity associated with their empathy?
 - 4: How does creativity vary across different levels of empathy?

Sketch ID	Question ID	Author
1	3	Juhong
2	1, 2, 4	Juhong
3	1, 2, 4	Juhong
4	1, 2, 4	Juhong
5	1, 2, 4	Juhong
6	1, 2, 4	Juhong

Design decision

I chose to focus on comparing the distributions of students' creativity and empathy between individuals and countries. Although Sketch No. 1 is meaningful to calculate and compare all the associations between the test questionnaires, it is more meaningful to researchers, not for teachers or students. Additionally, No. 2 is hard to read for target-aged students to interpret the data. The visualization should be more intuitive, and it would be better if other channels inspired by nature objects are added. In that context, the last one is my favorite.

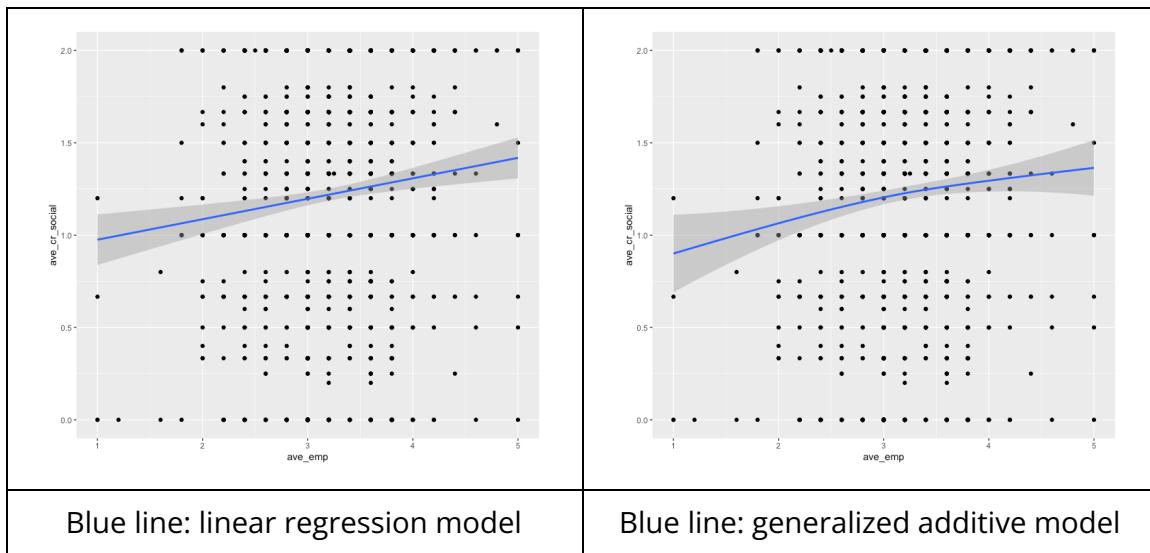
Storyboard

Main Message

Insights

1. **Korean students with higher self-reported empathy tend to show higher achievements in creativity regarding social problem solving.**

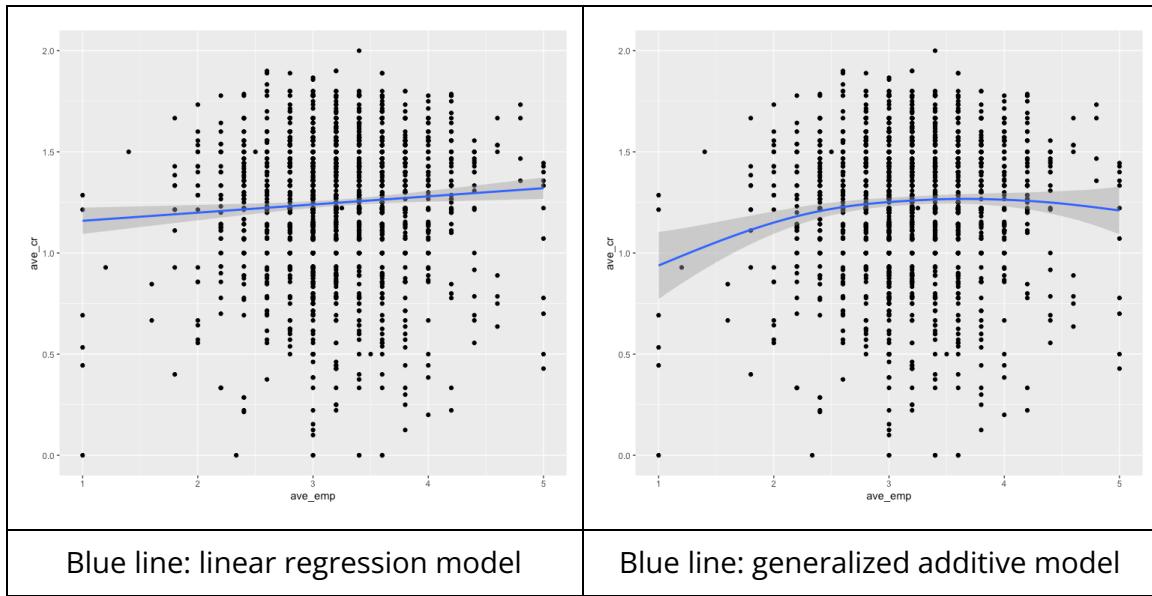
Using R, I examined the correlation between students' empathy scores and creativity in social problem solving. Using a Pearson correlation analysis, I found a positive association ($r = .09$) between students' self-reported empathy levels and their creativity in solving social problems. The result was statistically significant ($t(df = 1561) = 3.65, p < .001$), which means this association exists in the population.



2. **Korean students with higher self-reported empathy tend to show higher achievements in overall creativity.**

Using R, I examined the correlation between students' empathy scores and overall creativity domains. Using a Pearson correlation analysis, I found a positive association ($r = .06$) between students' self-reported empathy levels and their creativity. The result was statistically significant ($t(df = 1774) = 2.80, p = .005$),

which means this association exists in the population.



3. The association between Korean students' empathy and creativity holds regardless of gender.

Using R, I separated the data by gender and conducted correlation tests between students' self-reported empathy levels and their creativity. Both of them turned out to be positive, while female students show a stronger correlation ($N = 876, r = .08, p = .01$) than male students ($N = 900, r = .04, p = .18$).

4. The positive association between Korean students 'empathy and creativity is more distinct in upper secondary general schools compared to those in vocational schools.

Using R, I grouped data based on school type and conducted Pearson correlation analyses between students' self-reported empathy levels and their creativity.

School type	correlation	t-statistic
Lower Secondary School, General	0.15	$t(df = 89) = 1.40, p = .16$
Upper Secondary School, General	0.08	$t(df = 1407) = 3.19, p = .001$
Upper Secondary School, Vocational	-0.08	$t(df = 250) = -1.21, p = .22$
Mixed Secondary School, General	0.46	$t(df = 12) = 1.79, p = .09$

Mixed Secondary School, Vocational	-0.05	$t(df = 8) = -0.17, p = .90$
------------------------------------	-------	------------------------------

There was a positive association between empathy and creativity in general schools. However, it reached statistical significance only in upper secondary general schools. In contrast, there was a negative association in vocational schools, while this result was not statistically significant. This suggests that the positive association between empathy and creativity is more pronounced among students in upper secondary general schools compared to those in vocational schools.

5. The positive association between Korean students' empathy and creativity is present across different grades.

Although the result for ninth grade was not statistically significant, there was a positive association between students' empathy and creativity in both ninth ($N = 95, r = .14, p = .17$) and tenth graders ($N = 1680, r = .06, p = .01$). There were four students in eleventh grade, while it was impossible to conduct a Pearson correlation analysis due to lack of samples.

Insight Review

Insight Number	Compelling (1-5)	Audience(1-5)	Data(1-5)
1	5 The relationship is quite intuitive and apparent rather than proposing insights. It is obvious that students who care more about others would know or discuss more about social issues. They would come up with new, multiple, and various ideas to help them.	5 Social conflicts among people are persistent and serious challenges in South Korea. Despite their significance, there are few creative solutions currently being proposed or implemented. Additionally, students nowadays don't even consider the problems as theirs. Teachers can address this by introducing real-life social issues into the classroom. By doing so, they can encourage students to develop agency and creative social problem solving skills to address these challenges.	5 The results of correlation tests are statistically significant.
2	5 The relationship between empathy as socio-emotional skills and creativity is a new idea that isn't mentioned within articles before.	5 Teachers may verify the association with the data by themselves, and include socio-emotional skills in creative activities. Additionally, this insight provides teachers with the flexibility to adjust and diversify their approach to fostering creativity in the classroom.	5 The results of correlation tests are statistically significant, although the correlation is not strong.
3	4 This insight is meaningful as	5 Teachers would be able to consider the	3 Both female and male students

	gender composition significantly influences students' participation and preference within activities.	gender composition of their classrooms when deciding how much emphasis to place on socio-emotional elements within creative activities.	showed positive correlations between students' empathy and creativity. Nonetheless, with this Korean dataset, statistical significance was observed in only the female group. It means that the associations across genders may differ in strength and reliability.
4	2 There is a much variety in the results across schools. It would be more compelling to deliver a single main message.	5 Curricula and learning environments differ to a great degree between different types of schools. Therefore, it would be helpful for teachers to tailor their instructional approaches and creative activities to fit the specific context and needs of their school type.	2 The association is statistically significant only in upper secondary general schools.
5	4 This insight is related to the main research question.	5 This insight would be helpful for teachers to decide which grade they should merge socio-emotional elements into creative activities.	3 The majority of the sample data is centered on the tenth grade, which makes it hard to compare between different grades.

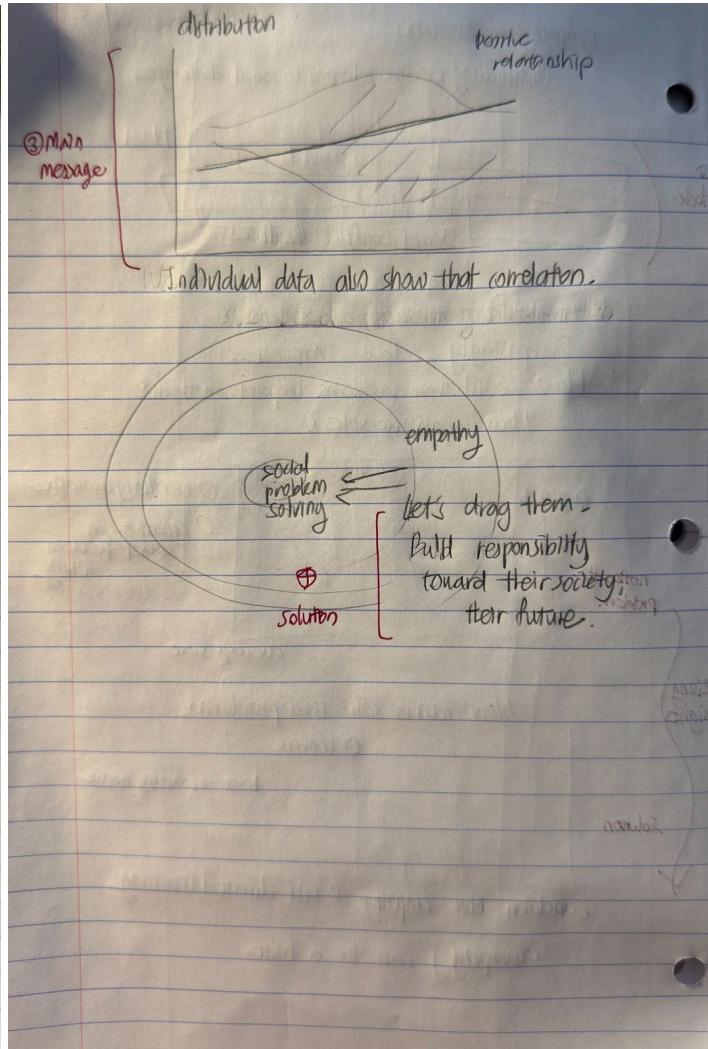
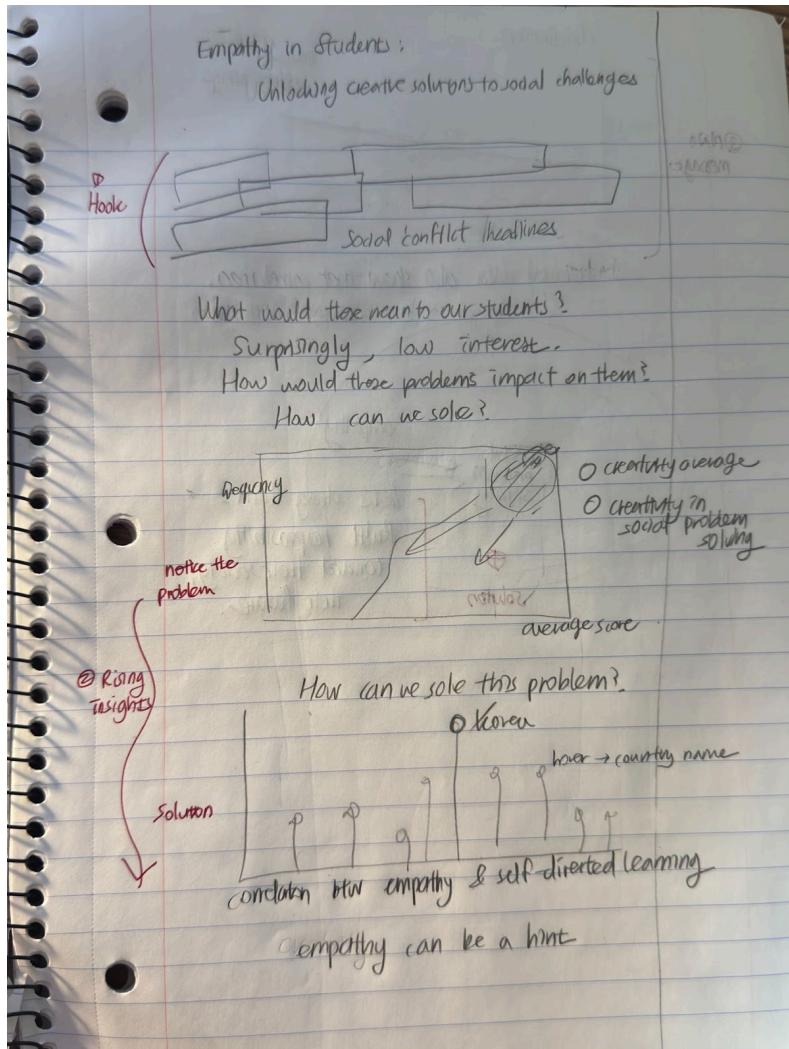
Decision

Main insight: Empathy in students: Unlocking creative solutions to social challenges

To foster students' creativity in addressing social challenges, include elements of socio-emotional skills, such as empathy, in educational activities.

- Our team selected this main message because it is strongly supported by our data and offers a fresh perspective to the discussion. We found a statistically significant positive relationship between students' empathy and their ability to generate creative solutions to social problems.
- This insight is actionable for teachers. It suggests that teachers can promote both creativity and agency among students by integrating socio-emotional skill development into creative classroom activities focused on social issues. Moreover, teachers have the flexibility to adapt and diversify creative activities based on the unique problems that students face. Data visualizations drawn from this research can offer teachers meaningful guidance for tailoring instruction, helping students become more effective, creative problem-solvers in the face of real-world social conflicts.

Data Story



Prototype v.1

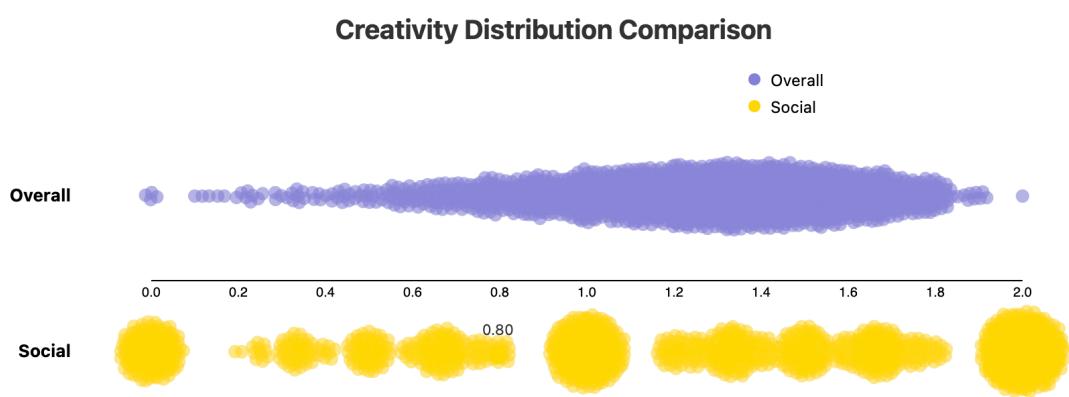
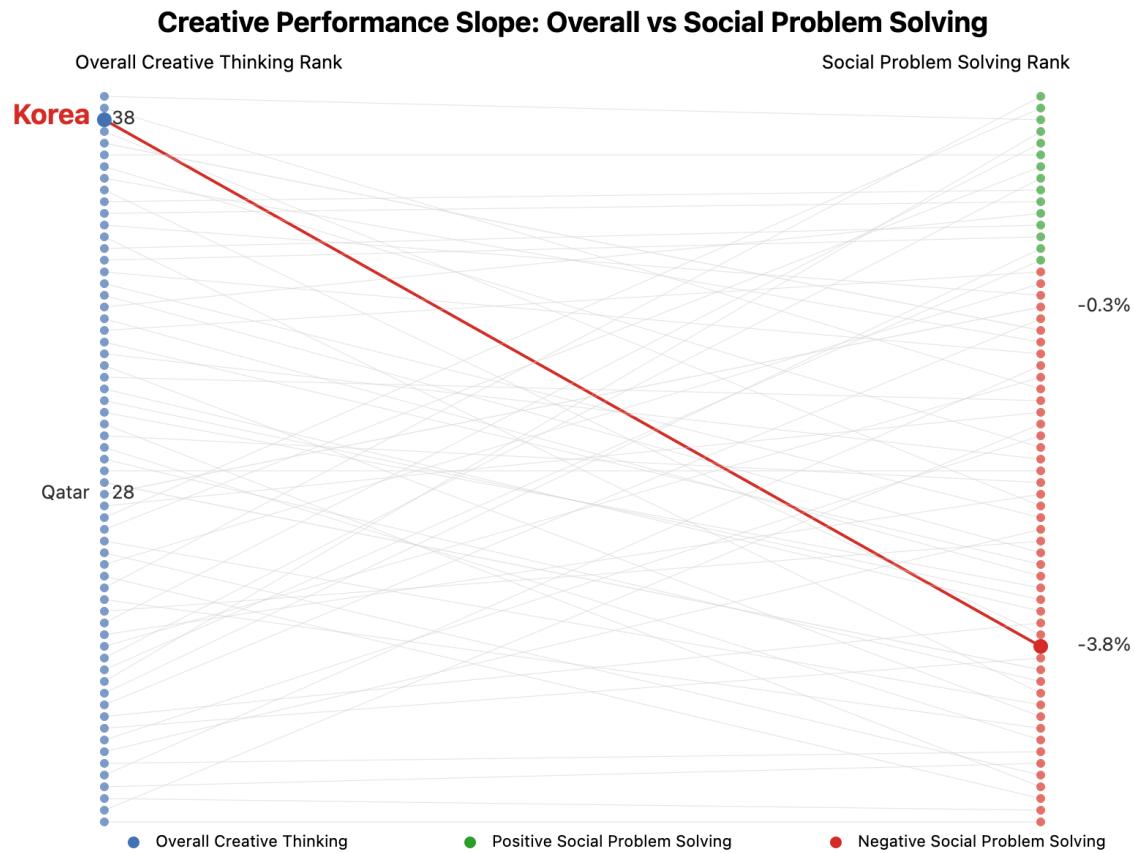
Team member documentation: Juhong Eom

Complete data pipeline

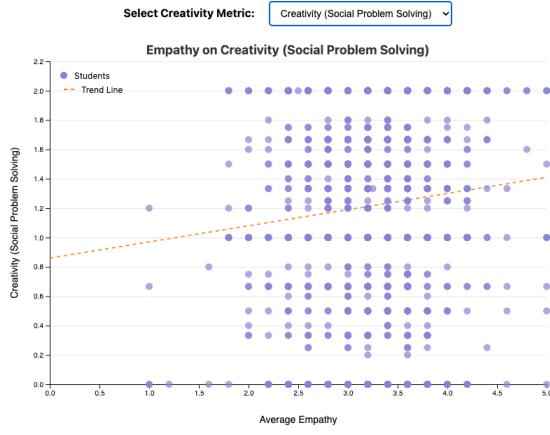
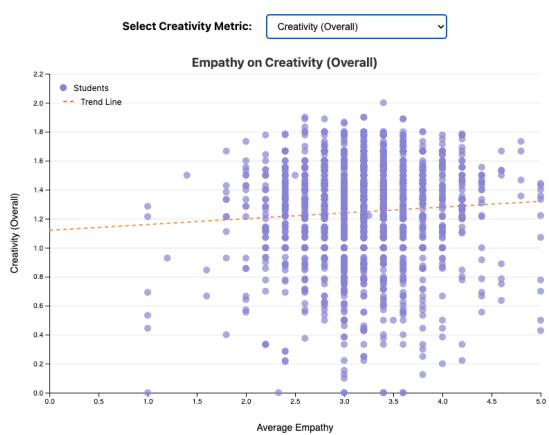
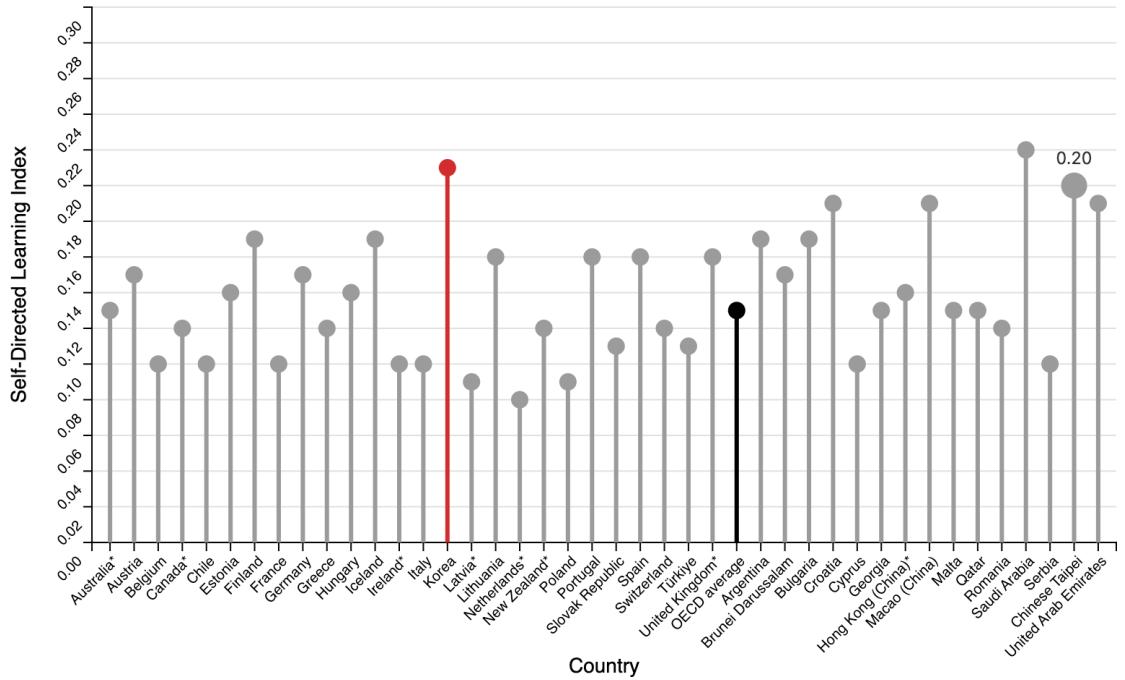
Dataset	Visualization
Overall and social problem solving creativity performance by country	[slope chart] Creative Performance Slope: Overall vs Social Problem Solving
Overall and social problem solving creativity performance of Korean students	[Beeswarm chart] Creativity Distribution Comparison
	[Scatterplot] Relationship between Korean students' self-reported empathy score and performance on overall/social problem solving creativity
	[Circular Scatterplot] Distribution of Korean students' self-reported empathy
Association between self-directed learning and empathy as a social and emotional skill	[Lollipop chart] Power of Empathy on Self-Directed Learning

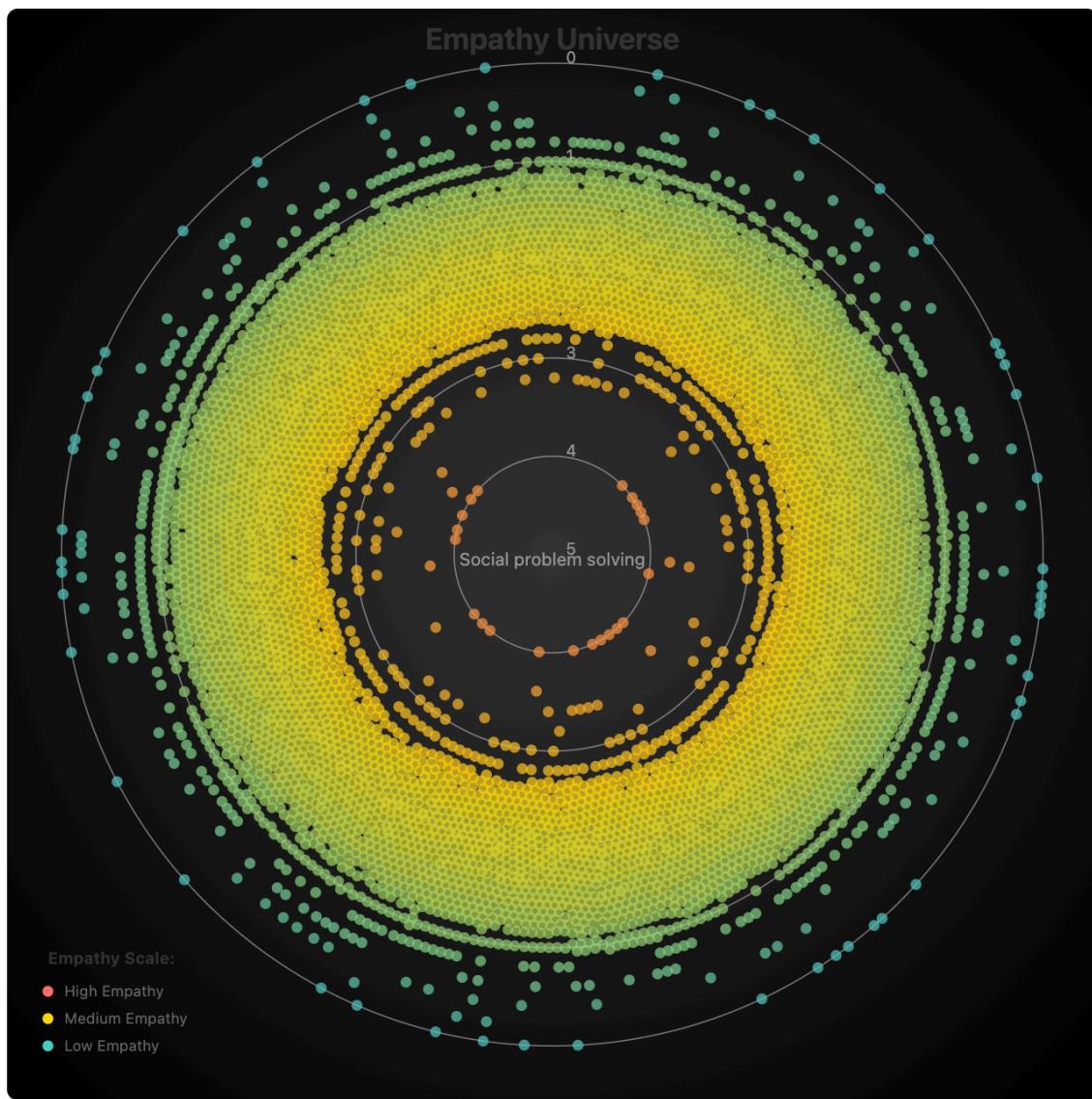
Each dataset was converted into csv files within the public>data folder. All data were cleaned and analyzed by R.

Functional D3 visualizations



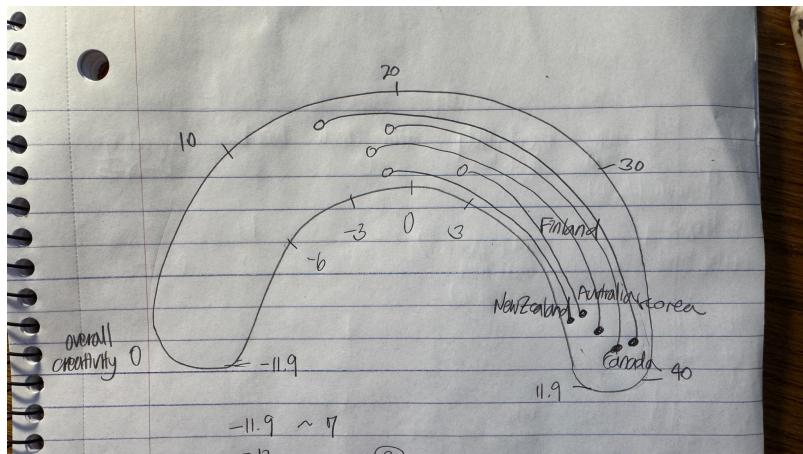
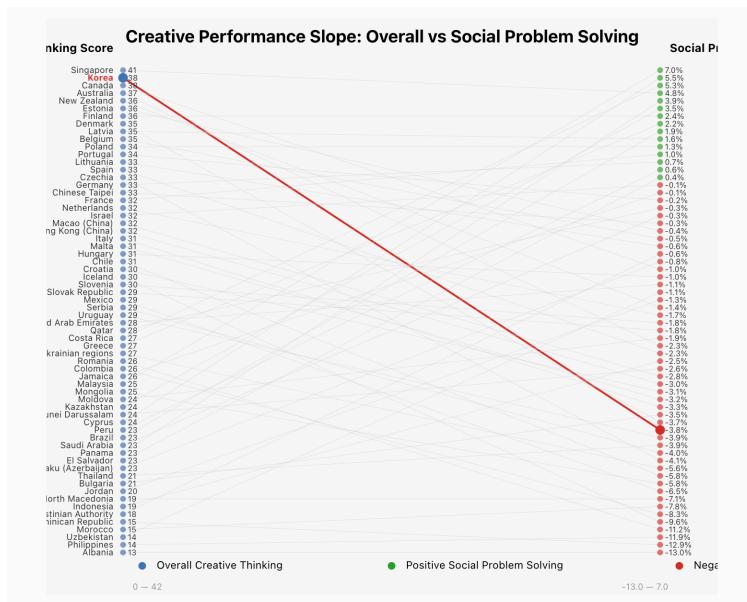
Power of Empathy on Self-Directed Learning

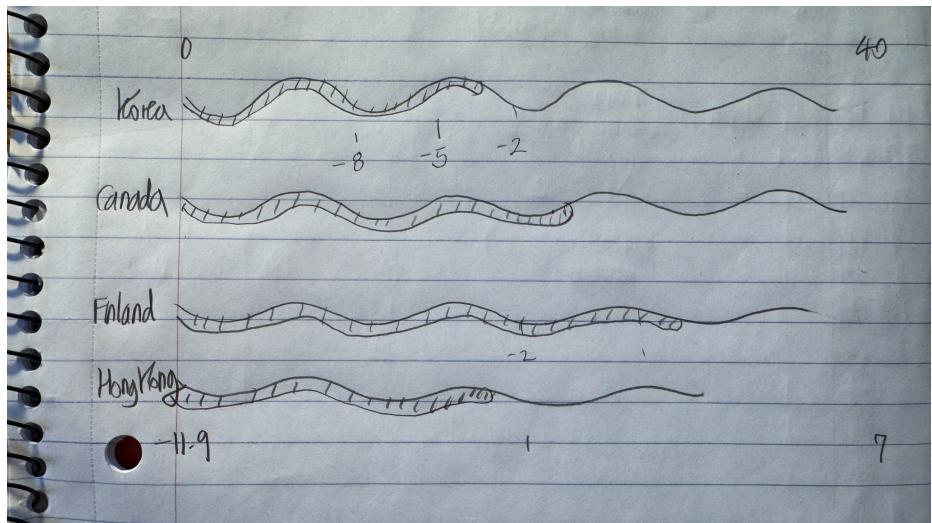




Additional visualization drafts

For the additional visualization, I am searching for an effective way to present data from the first graph. The main purpose is to highlight Korean students' relatively low performance on creativity related to social problem solving. For now, I have used a slope graph as a temporary measure. However, it may lead to distorted interpretation of the data as their social problem solving performances are ranked. Countries with the same score appear at different positions on the right side, which influences slopes. I am currently in the progress of exploring various forms of data visualizations formats to better represent this information.





Website structure and view design

ThinkMate

 HOME RATIONAL REGISTER GROUP SURVEY ANALYSIS

Empathy in Student: Unlocking creative solutions to social challenges

'Men don't know why they became unhappy': the toxic gender war dividing South Korea

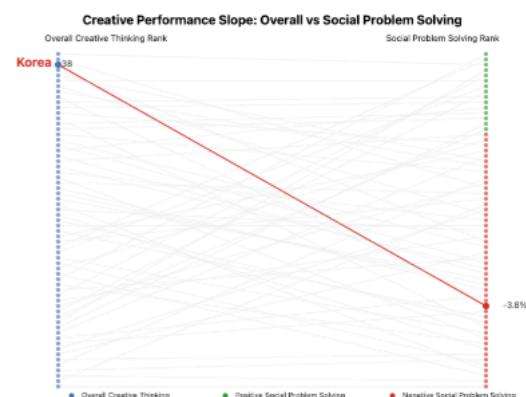
Gender, generation gap on full display in exit poll showing entrenched differences

South Korea's deep political divide

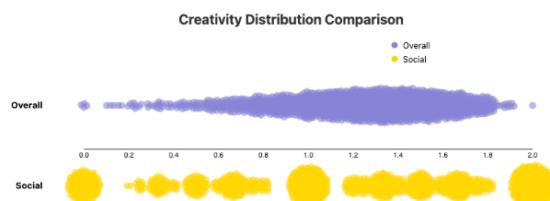
For cash-strapped South Koreans, the class conflict in "Squid Game" is deadly serious

S. Korea has 3rd highest social conflict index among OECD countries

What would these issues mean for our students?
They do think these problems are important, however, they do not think they should make a difference.
How would these problems impact on them? Would it be okay to let them ignore these issues?
As someone to guide our students, how can we solve them?



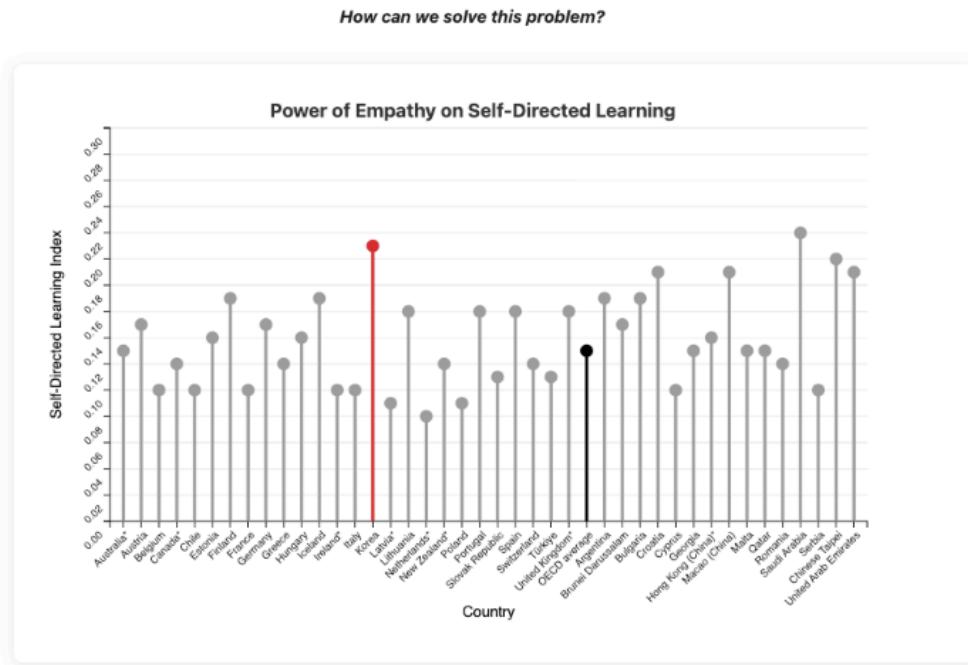
IT is really a problem, especially comparing between students.



We begin the discussion with several news reports on social conflicts in South Korea. At this point, the concepts of social problem solving creativity and possible solutions to these conflicts should be clearly explained as follows:

The social conflicts in South Korea have existed for a long time. People resonate with the necessity to solve the problems, however, they struggle to find good solutions.

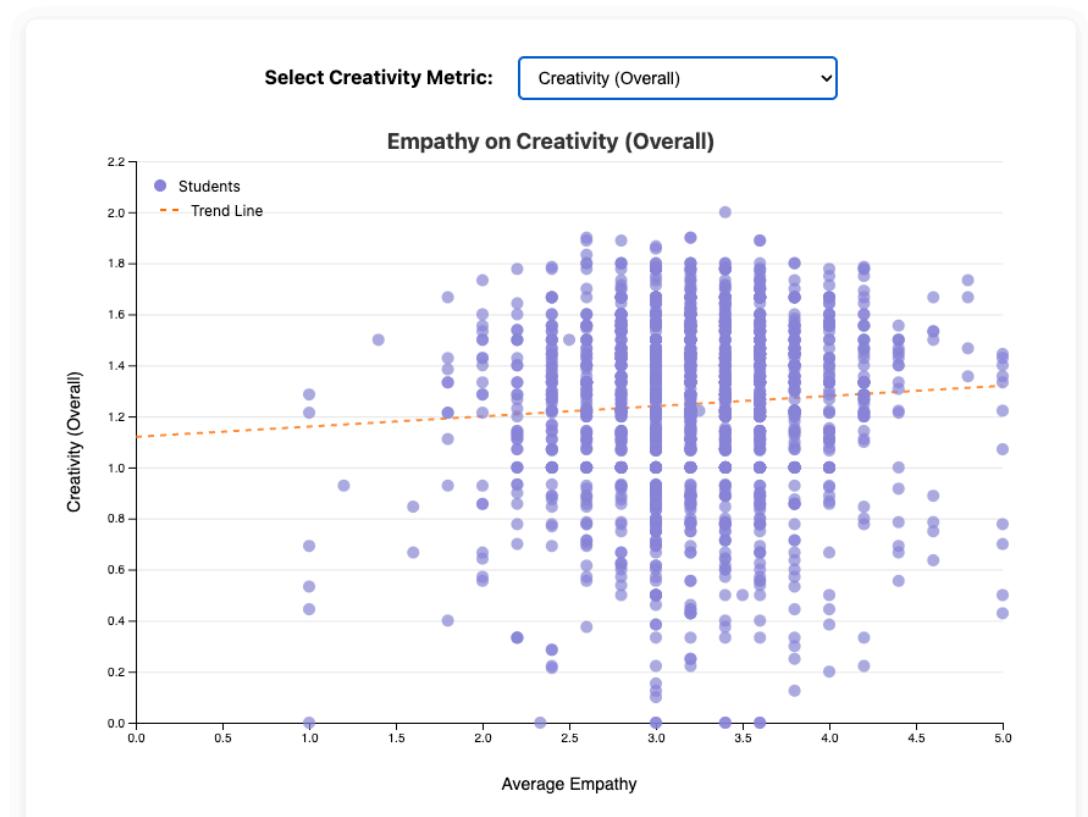
I am considering adding results from academic papers as supporting evidence.



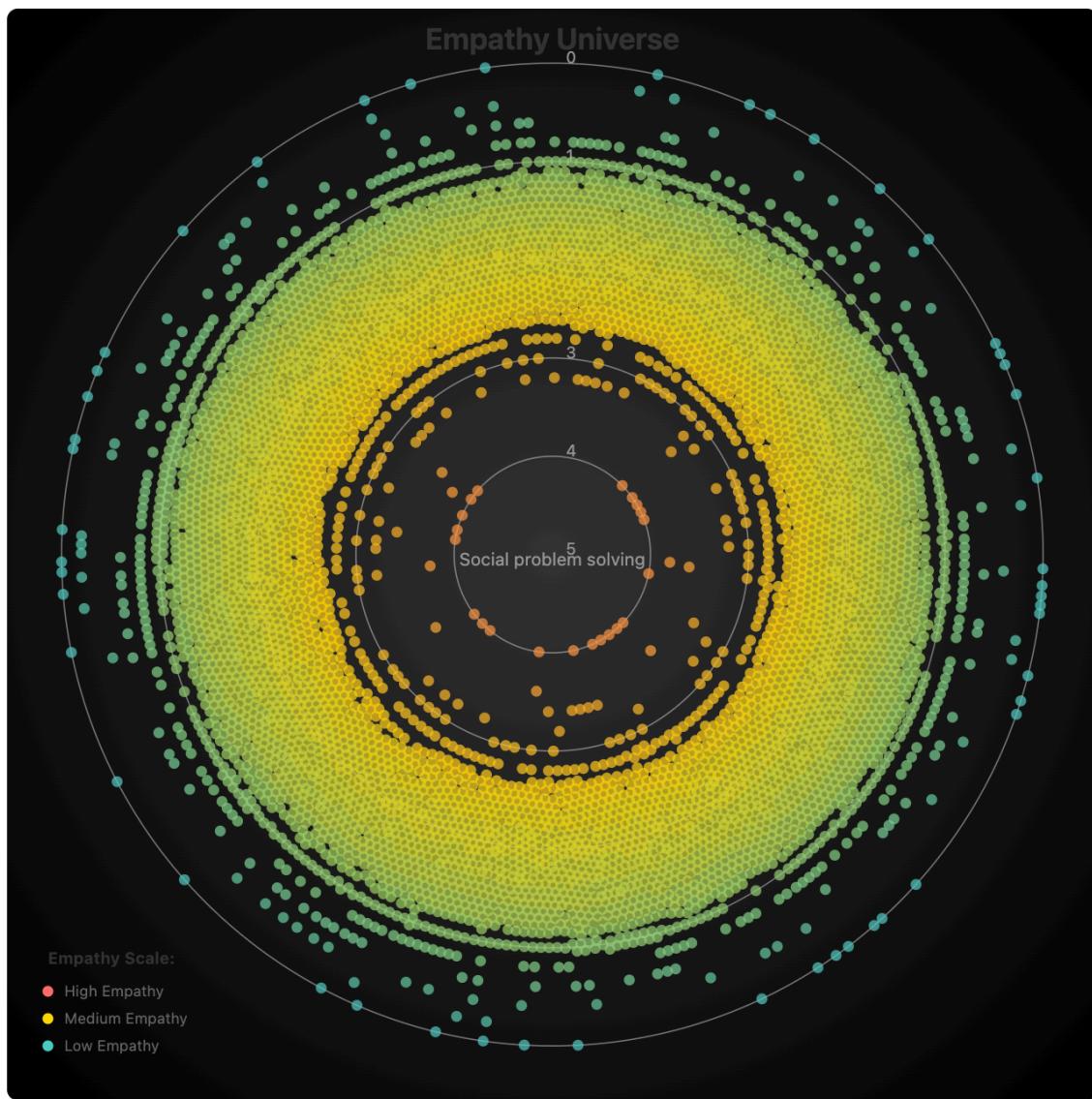
To suggest empathy as a solution, we present data on the relationship between empathy and self-directed learning. Self-directed learning is closely linked to willingness to solve social problems.

We can find hint in empathy. Chart above is about confidence in self-directed learning, and social and emotional skills. It shows change in the index of confidence in self-directed learning index with a one-unit increase in each of the social and emotional skills (SES) indices after accounting for students' and schools' socio-economic profile, and mathematics performance.

We can see that students who has higher empathy score tends to have higher confidence in self-directed learning index.



To illustrate a more direct relationship, we suggest a scatterplot that shows a strong association between Korean students' empathy and their performance in creativity. As the slopes for overall creativity and social problem solving creativity are different, the audience can easily observe the importance of empathy in social problem solving.



Reference

PISA 2022 RESULTS (VOLUME I) © OECD 2024

PISA 2022 RESULTS (VOLUME V) © OECD 2024

PISA 2022 RESULTS (VOLUME III) © OECD 2025