Between Seeing and Measuring: The Human Aspect of Defining Variables with Image-Based Data

ProDaBi Colloquium, 04/06/25

Sibel Kazak
Department of Mathematics and Science Education

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

1. Data
2. Statistics Education in Pre-K–12
3. Data Investigation Cycle
4. Prior Research on Non-Traditional Data
5. Study Overview
6. Results
7. Implications

What count as DATA?

> A collection of numbers or other pieces of information to which meaning has been attached.
>
> (Utts, 1996, p. 16)

Traditional data

- Numerical
- Categorical
- Tabular format with rows and columns

Non-traditional data

- Potentially non-numerical
- Text/Sound/Image/ Video
- Unstructured
- Repurposed

STATISTICS EDUCATION IN PRE-K-12

DATA and STATISTICAL LITERACY

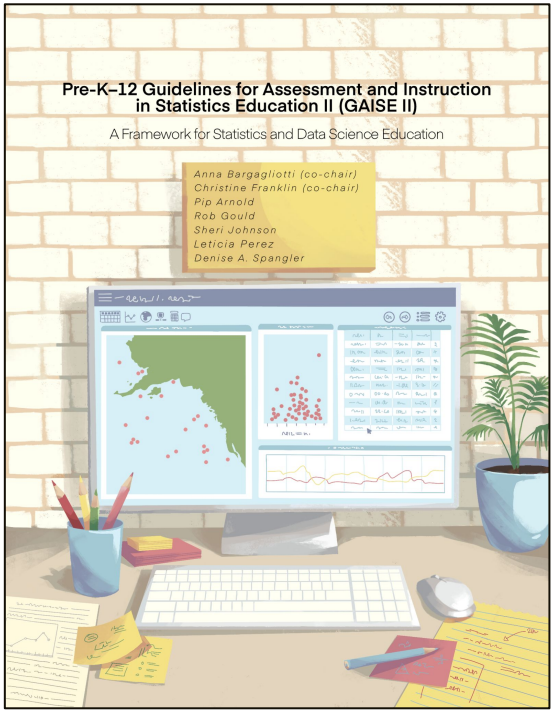
Questioning and interrogation of the data

Different data/variable types

Multivariable thinking

Use of technology

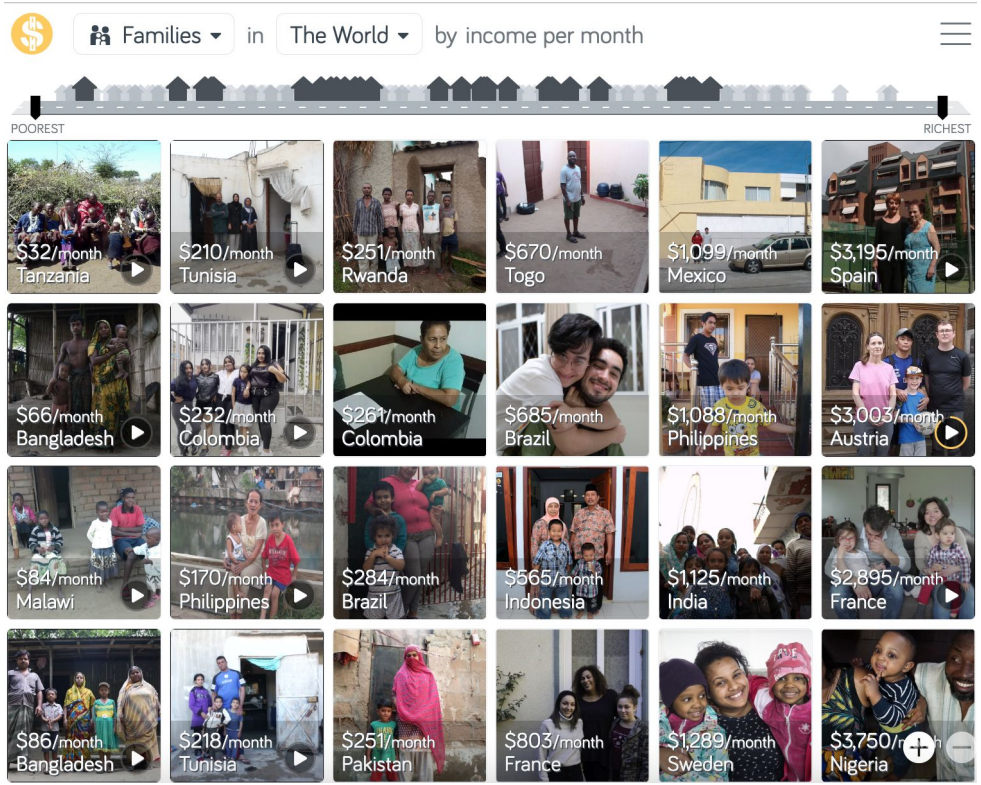
Statistical problem solving process



GAISE II Report (Bargagliotti et al., 2020)

DOLLAR STREET WEBSITE

- The world as a street ordered by monthly income in the household
- Photographs of families and their environments from 460 homes in 66 countries around the world (Africa, America, Asia, Europe)



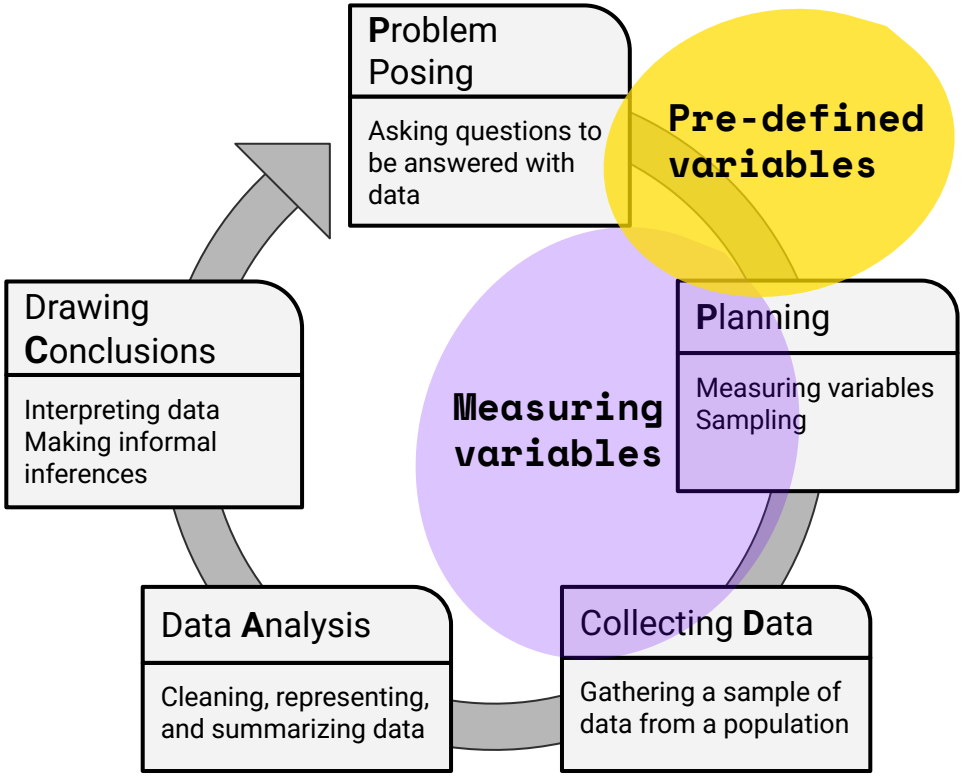
Turning these **cases into data** to explore statistically investigative questions

https://www.gapminder.org/dollar-street

A TRADITIONAL DATA INVESTIGATION CYCLE

➔ PPDAC cycle

(Wild & Pfannkuch, 1999)

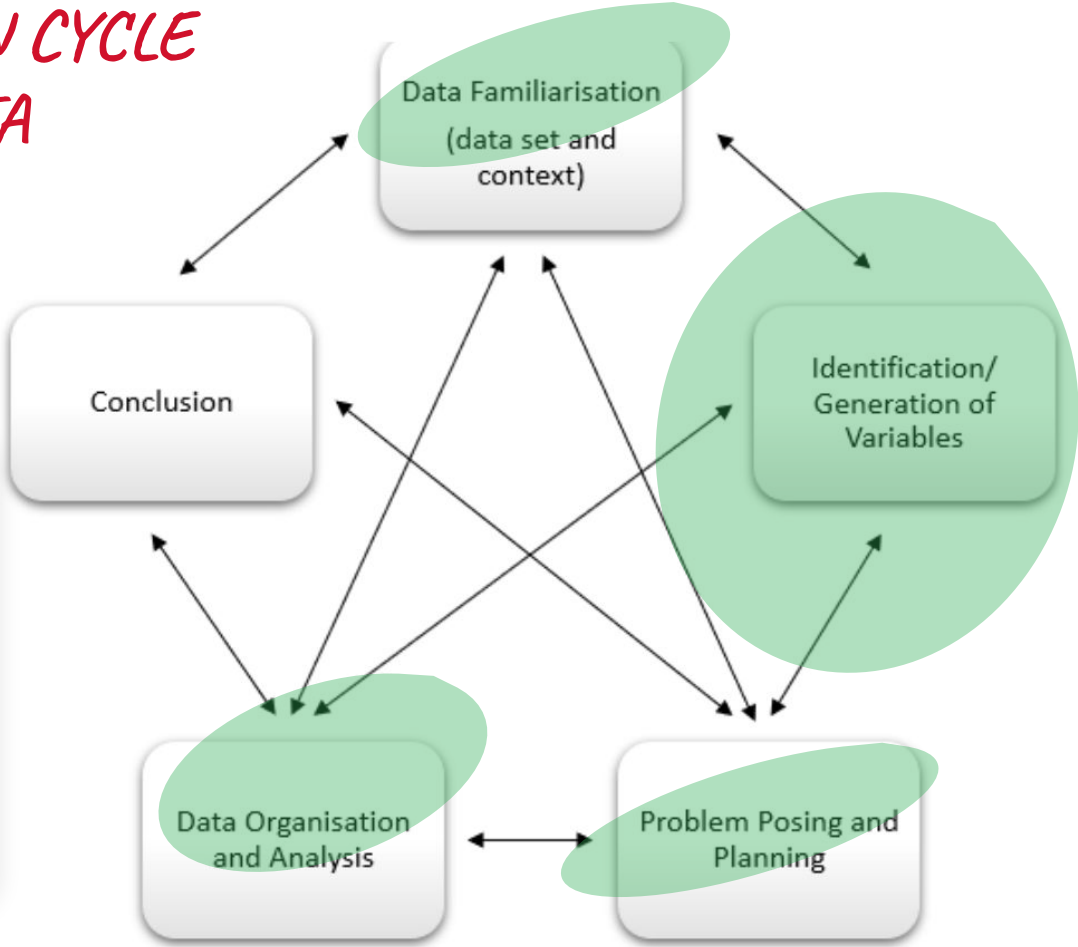


STATISTICAL INVESTIGATION CYCLE FOR IMAGE-BASED DATA

→ **Data-ing** process

“making **observations** from photos to describe a quantity/quality and organizing those observations into **variables** to explore statistical questions with the recognition of variability in the data before analyzing data”

(Kazak, 2025, ZDM)



(Kazak, Fielding, & Zapata-Cardona, 2022)

Prior Research on Non-Traditional Data

Undergraduate students' **creation of features from text data** while classifying headlines as clickbait or news: Distinctions emerging between **human-perceivable** features and **computer-detectable** features based on structural or rule-based text elements.

(Horton et al., 2023)

Secondary teachers' decision rules to **classify grayscale photos** as light/dark or high/low contrast using statistical summaries of pixel value distributions: Shift from **visual and descriptive reasoning** to using aggregate measures to bridge **human-judged features** with **computer-detectable patterns** in the data.

(Fergusson & Pfannkuch, 2024)

7–8-year-old Brazilian children's use of **images from the Dollar Street website** to explore global living conditions: From **comparing, classifying**, and **identifying** patterns in the images (people's pets) to discussing **perceived attributes**, such as grouping birds by type or by whether they were kept in cages.

(Buehring & Grando, 2023)

RESEARCH QUESTIONS

- What is the nature of the data-ing process when pairs of pre-service mathematics teachers work with Dollar Street images?
- How do pairs of pre- service mathematics teachers come to identify and generate variables in data-ing?

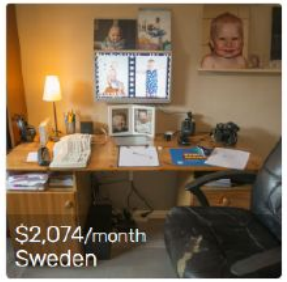
PARTICIPANTS

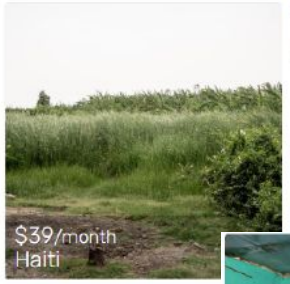
- 3 pairs of volunteering pre-service mathematics teachers (ages 21–23, females)
- Completed a two-semester Introduction to Probability and Statistics course

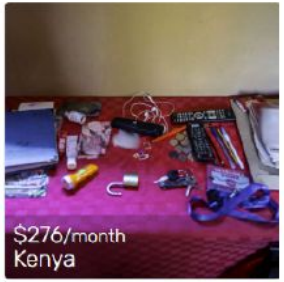
TASK-BASED INTERVIEW

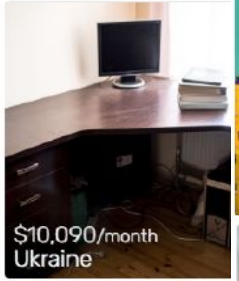
- Pair work, online platform (about 90 minutes)
- Selected Dollar Street photographs of the Work Areas of the participating families with different income levels (ranging from $30/month to $10,090/month) from different countries and continents

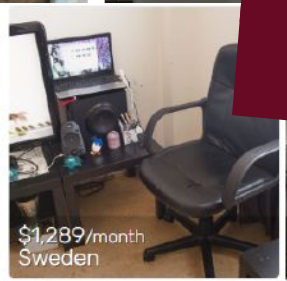
Work Areas in Dollar Street

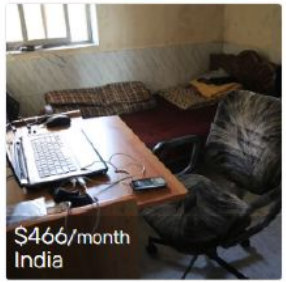


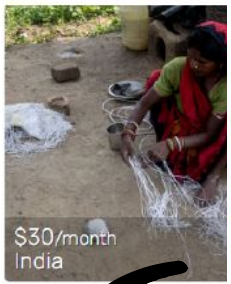


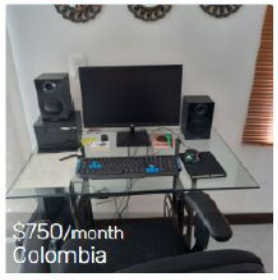


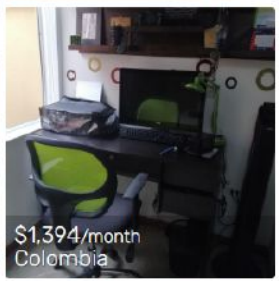


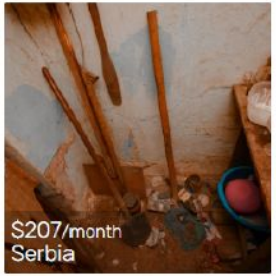


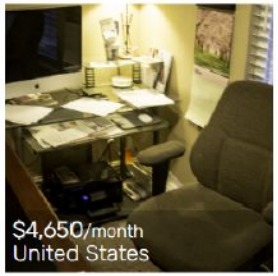












TASK PROCEDURE

- “What do you notice after examining the photos of work areas given on this page? What do you wonder about?”
- “Can you pose an investigative question that can be answered with these photos? Can you think of another question?”
- “How would you arrange the photos to answer this question? Can you explain why you did that?”

DATA ANALYSIS

- what variables the pair defined during the whole interview session
- how the variables were identified and generated
- what actions, related to data-ing, each pair took when identifying and generating variables during data familiarization, question posing, and data organization
- how these actions occurred during data-ing
- what characteristics of these variables were identified and generated during data-ing

(Kazak, 2025, ZDM)

RESULTS

➔ Actions emerging in the data-ing process across the three pairs

- Observing
- Interpreting
- Conjecturing
- Inferring
- Comparing
- Grouping
- Ordering
- Questioning/ question posing
- Relating variables

- Categorizing variables
- Measuring variables

➔ Two types of identifying variables

observational variables based on visual judgement/ metadata

inferential variables based on personal interpretation of image-based data

inferring and **grouping** to **categorize** the work areas as 'home office' and 'workplace'

P3.1: Maybe we can infer something. Some of them are home offices. Some workplaces are outside the home. I think this is the house of someone in Colombia ($654/month).

P3.2: That [in Colombia] might not even be a home. For example, South Africa is a house, but Colombia, with a 654 income, is not a house.

P3.1: Hmm. I think that one is a house [India $466/month]

P3.2: It is a house.

P3.1: This [India $466/month] works from home, for example.

P3.2: I think you call it home because of those things (a bed or couch) in the back, right?

P3.2: Uh-huh.

When the pair was prompted to pose a statistical question (*question posing*)

measuring variables

P3.1: I am thinking about how measurement will be done when the data is a photograph. Let us say I write a question about a word called comfort. Maybe comfort is something that can be measured with a photograph. However, I am thinking about what we can measure with a photograph so that I can pose a question.

P3.2: Maybe not the comfort but the quality of the materials in the work area. When we say quality, can we fully capture what it means? For example, when we look at the work area in Colombia, India, or Haiti, is it the quality of the material? How exactly can we specify it?

Identifying variables

Is there a relationship between the monthly income of the same occupational groups and the working conditions (quality of materials in the environment, tidiness of the setting, etc.)?

P3.1: You have said work setting. Let us go back to that. Let us use the word comfort, though comfort...

P3.2: But the better working conditions at work are also **very subjective**. For example, working in an outdoor or garden is perhaps more comfortable for me than being stuck in a room.

P3.1: Yes, and we cannot fully capture the working conditions with a photo. I mean, can we know how many hours they work?

P3.2: Then let us say the same occupational groups. As the income of the same occupational groups increases, do they tend to work in better conditions, or can it be associated with their work?

P3.1: What does better condition mean?

P3.2: Let us write that question first. Now, the same occupational groups. We can describe it as, for example, better quality materials. Or we can say more functional tools like computers are getting better. We can categorize it as such. Because when we do different occupations, it is not very useful. However, **when I classify the same occupations, comfort means something**.

IMPLICATIONS

- Working with image-based data is an emerging area in K–12 education settings
- The complex process: Human aspects of developing and negotiating variables based on personal interpretation of image-based data and the related challenges due to the subjectivity of such experiences
- Providing opportunities for pre-service teachers to engage with image-based data to enhance their understanding of the data-ing process perceived beyond traditional data investigations during their teacher preparation programs

REFERENCES

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) report II*. American Statistical Association and National Council of Teachers of Mathematics.

Buehring, R. S., & Grando, R. C. (2023). Reading and writing the world with children: Statistical thinking and multivariate data. *Statistics Education Research Journal*, *22*(2), 6.

Fergusson, A., & Pfannkuch, M. (2024). Using grayscale photos to introduce high school statistics teachers to reasoning with digital image data. *Journal of Statistics and Data Science Education*, *32*(4), 345–360.

Horton, N. J., Chao, J., Palmer, P., & Finzer, W. (2023). How learners produce data from text in classifying clickbait. Teaching Statistics, 45, S93–S103.

Kazak, S., Fielding, J., & Zapata-Cardona, L. (2022). *Investigation cycle for analysing image-based data: perspectives from three contexts*. Paper presented at the Eleventh International Conference on Teaching Statistics (ICOTS11), Rosario, Argentina.

Kazak, S. (2025). Data-ing process with image-based data: variable identification and generation. *ZDM Mathematics Education*, *57*(1), 61–74.

Utts, J. M. (1996). *Seeing through statistics*. Duxbury Press.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical inquiry. *International Statistical Review*, *67*(3), 223–248.

Thank You for Your Attention

Sibel Kazak
skazak@metu.edu.tr

ProDaBi Colloquium, 04/06/25