



**HAL**  
open science

## Pre-service mathematics teachers' experiences of acquiring and organizing image-based data

Sibel Kazak

► **To cite this version:**

Sibel Kazak. Pre-service mathematics teachers' experiences of acquiring and organizing image-based data. Twelfth Congress of the European Society for Research in Mathematics Education (CERME12), Feb 2022, Bozen-Bolzano, Italy. hal-03751833v2

**HAL Id: hal-03751833**

**<https://hal.science/hal-03751833v2>**

Submitted on 19 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Preservice mathematics teachers' experiences of acquiring and organizing image-based data

Sibel Kazak

Pamukkale University, Turkey; skazak@pau.edu.tr

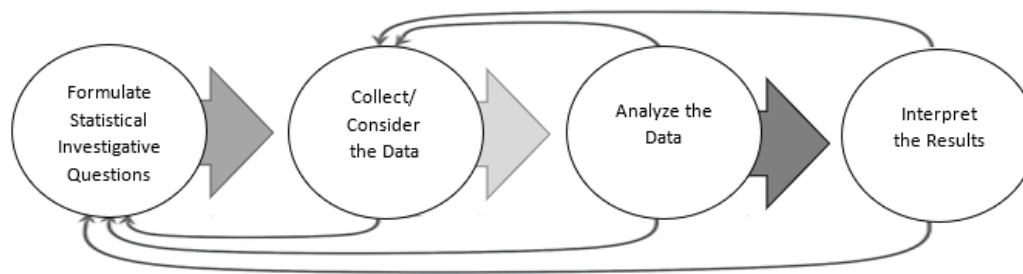
*There is a growing interest in engagement with non-traditional data generated from a variety of sources, such as sensory devices, smartphones/watches, and social media tools. This paper reports on experiences of a group of pre-service mathematics teachers in acquiring and organizing image-based data to analyze categorical data as part of a course assignment. To identify the actions pre-service teachers took to transform observations of butterflies' photos to cases to answer statistical questions related to categorical data, the data from retrospective interviews and artifacts generated during the data exploration and organization were analyzed. The findings suggest that data handling and structuring became an important component of the statistical investigation before formulating statistical questions. While the use of hierarchical tables in structuring data by hand appeared to be intuitive, it could be challenging to use such a structure already built in a data organizing tool.*

*Keywords: Statistical inquiry, data acquiring, data organizing, image-based data, categorical data.*

## **Introduction**

Understanding and reasoning with data become increasingly essential part of our everyday lives as we need to handle information related to global issues on health, environment and so on, such as the COVID19 pandemic and wild fires due to extreme weather conditions. The nature of data also evolves with the advanced digital technologies. Large quantities and different forms of data are generated from a variety of sources, such as sensory devices, smartphones/watches, social media tools etc., every day. To analyze and make predictions from these data require new skills. Therefore, an important goal of statistics education is to develop necessary data skills starting from early school years.

Statistics is typically taught as part of school mathematics curriculum and there has been a shift from simply computing numerical and graphical representations of data to an inquiry-based approach (Watson, Jones, & Pratt, 2013). In this approach to the teaching of statistics, the focus is on the statistical problem-solving process (Franklin, Kader, Mewborn, Moreno, et al., 2007; Bargagliotti, Franklin, Arnold, Gould, et al., 2020). According to *the Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II): A Framework for Statistics and Data Science Education*, the purpose of this process is “to collect and analyze data to answer statistical investigative questions” (Bargagliotti et al., 2020, p. 13). As seen in Figure 1, it has four components that are interlinked: (1) Formulate statistical investigative questions, (2) Collect/consider the data, (3) Analyze the data, and (4) Interpret the results.



**Figure 1: Statistical problem-solving process (reproduced from Bargagliotti et al., 2020)**

Traditional statistical education focuses on data from random samples and making inferences from a sample to an unknown population. In today's digital age, however, data come in various forms (big, messy, unstructured data, repurposed data, image/text/sound-based data etc.) and do not necessarily comply with the structure required for statistical inference in the traditional sense (Gould, Bargagliotti, & Johnson, 2017). Along with the emerging field of data science, statistics education tends to focus more on the engagement with such non-traditional data sets.

Image-based data, i.e., the use of photographs as data, becomes increasingly common in daily encounters. The GAISE II report (Bargagliotti et al., 2020) acknowledges the use of this kind of non-traditional data as part of the statistical problem-solving process and helping students to make sense of these non-traditional data. As an example, Bargagliotti et al. use data from Dollar Street website as part of the Gapminder Foundation project ([www.gapminder.org/dollar-street](http://www.gapminder.org/dollar-street)). The Dollar Street website displays the world as a street ordered by monthly income per person in the family from different countries and continents, and currently uses 43685 photos of 422 families in 66 different countries. Using these image-based data, students are expected to investigate "How are people's concepts of family and living spaces similar or different across the world?" (Bargagliotti et al., 2020, p. 63). The guidelines in the GAISE II report offers some instructional ideas with regard to the use of these image-based data through each four components of the statistical problem-solving process mentioned above. Bargagliotti et al. (2020) suggest that this kind of non-traditional, multivariate data sets requires a great amount of data exploration time to make observations and wonderings to analyze data. Engagement with other types of non-traditional (large and complex) data from secondary data sources, such as participatory sensing data (Gould et al., 2017) and public data sets (Wilkerson, Lanouette, & Shareff, 2021), also appears to require more emphasis on considering data and data preparation phases in statistical investigation process. As seen in these studies, such existing multivariate data sets are constructed and made publicly available by others with a particular purpose and investigators often repurpose them to explore new questions. To do so, data handling and structuring becomes an important part of statistical investigations.

The Collect/Consider the Data component of the statistical problem-solving process involves recording/acquiring, measuring and organizing the data to answer statistical questions with the acknowledgment of variability in data (Bargagliotti et al., 2020). Konold, Finzer, and Kreetong (2017) call attention to this important process of transforming observations into data and focus on table format as a typical means of data recording and structuring either by hand or using software. Two common table formats in statistics are case-data table and summary table as called by Konold et al. (2017). While case-data tables are mainly used for collecting and storing raw data, summary

tables usually include only some of the information gathered and are organized in a way to compare groups or detect trends in the data. The data analysis software, such as TinkerPlots (Konold & Miller, 2011) and CODAP (<https://codap.concord.org/>), uses attribute-based structure for recording data where each column includes a variable and each row holds one observation, called a ‘case’, as described by Konold et al. (2017). However, entering the information collected to organize the data in such a format can be challenging for the learners. Konold et al. point out the need for thinking of data based on attributes which requires considering properties of the observations along with their values case by case. Their study revealed that students and adults tended to create two types of case tables, flat and hierarchical, when they were asked to construct data sheets to record and organize the data by hand from the pictures showing traffic along two road segments (including information about the vehicle type, speed, direction etc.) in a given time and date. Compared to the flat tables (attributes as columns and cases as rows) as seen in most data analysis software, the hierarchical tables include the cases at more than one level created by nested structure. This type of hierarchical structure of observations is available in CODAP and allows exploration of multilevel data sets.

Given the scarce of research on how students/adults come to record and organize data, especially with non-traditional image-based data, as part of a statistical problem-solving process, the aim of this paper is to present a case of three pre-service mathematics teachers who chose to use photos of butterflies as data to analyze categorical data. More specifically, the following question will be investigated in this exploratory study: What actions did pre-service teachers take to transform observations of butterflies’ photos to cases to answer statistical questions related to categorical data?

## **Method**

As part of 3<sup>rd</sup> year course, Teaching of Statistics and Probability, in the mathematics education program, the pre-service teachers were given a group assignment to design a statistical investigation activity that is aligned with the learning outcomes in the middle school mathematics curriculum. Two of the groups were expected to focus on analyzing categorical data and comparing two or three groups using existing data sets available in publicly available internet sources. One of these groups, including Arya, Yelda and Hale (pseudonyms), chose image-based data using “Kelebek-Türk” (Butterflies-Turk) group’s website (<https://www.kelebek-turk.com/>) while the other group designed their data investigation activity involving data collected through survey questions on readings books.

Due to the uniqueness of transforming image-based data into a statistical analysis of categorical data with group comparisons, an interview was conducted with these three pre-service teachers to gain insights into their approaches when acquiring and organizing information from photographs of butterflies. Other artifacts, such as the documents created during the data exploration and organization, were also collected for analysis to answer the research question. Recordings of group interview and the student artifacts were analyzed using the perspectives on collect data/consider data suggested by Bargagliotti et al. (2020) and Konold et al. (2017).

## **Findings**

Preservice teachers began with considering possible data appropriate for analyzing categorical data with comparing groups, which was given in the course assignment. Due to the difficulty in finding at least three categorical variables in data sets using publicly available data sources, such as Turkish

Statistical Institution website, they stated that some of the data explored during the course, such as the ladybug data (Bargagliotti et al., 2020, p. 32), inspired them to consider similar data about other animals. Yelda came up with the idea of searching butterfly data on the internet and found the group of Kelebek-Türk that provides photographs and observational data provided by their members from different age and occupation groups. After reading the “About us” information on the website, the pre-service teachers got their initial information about how the data were collected. Then they decided to use some of these photographs to construct their own observations involving categorical data.

According to the Kelebek-Türk website, there are 416 types of butterflies under nine different families in Turkey and their members have photographed 393 of them. As seen in Figure 2, there are nine butterfly families with different number of types, such as the Family Hesperidae has 43 types, observed in Turkey and Figure 3 shows a sample of photos of different butterfly types belonging to the Family Hesperidae.

Ana Sayfa » Türler

Ülkemizde 9 aile altında toplam 416 tür bulunmaktadır [Bkz]. Bu sayı çeşitli çalışmalar kapsamında değişebilir. Bunlardan 393 tanesi üyelerimiz tarafından fotoğflanmıştır. Ailelere ait bilgilere aşağıdan ulaşabilirsiniz.

**Argynnidæ Ailesi**  
"Fırcaayaklı Kelebekler" olarak bilinirler, genellikle göze çarpan parlak renkli, orta ve büyük boy kelebeklerdir. Çok bilinen Atalanta, göçmen olan Dilan Kelebesi gibi yaygın ve dağılımı geniş türler bu ailedendir. Yürümek için dört bacaklarını kullanırlar, ön bacakları kısalmıştır. Tırtıllar çoğunlukla renkli, dikenli ve çeşitlidir, bazılarını toplu halde ipek ağlar içinde yaşar. Koza biki saplarına veya bazen kayalara asılır.  
Argynnidæ ailesine ait ülkemizde 55 tür bulunur, bunlardan 52 tanesi üyelerimizce fotoğflanmıştır.

**Danaidæ Ailesi**  
"Kral Kelebekleri" olarak bilinirler, çoğunlukla tropikal alanlarda bulunur. Büyük, güçlü, parlak renkli, göçmen kelebeklerdir. Ülkemizdeki tek temsilcisi Sultan'dır. Tırtıl beslediği bitüdeki zehirli maddeleri sindirdiğinden dolayı pupa ve enişin kelebek yem olarak tercih edilmez.  
Danaidæ ailesine ait ülkemizde 1 tür bulunur, bu tür üyelerimizce fotoğflanmıştır.

**Hesperidæ Ailesi**  
"Zıpzıplar" olarak bilinirler. Yapısal olarak diğer kelebeklerden kolayca ayırt edilir. Kafaları ve gözleri inidir. Antenleri birbirinden ayrıktır. Bütün türlerin tipik bir uçuşu vardır, bazılarını yavaş ve çok hızlı uçar.  
Hesperidæ ailesine ait ülkemizde 43 tür bulunur, bunlardan 41 tanesi üyelerimizce fotoğflanmıştır.

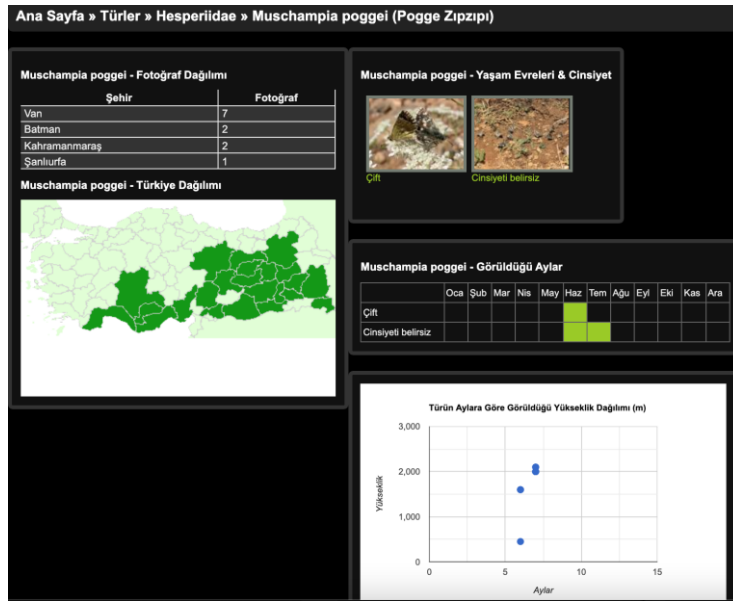
Tüm Türler	Fotoğraflı	Fotoğrafsız	Belirsizler
Argynnidæ Ailesi (55)			
Danaidæ Ailesi (1)			
Hesperidæ Ailesi (43)			
Libytheidæ Ailesi (1)			
Lycaenidæ Ailesi (182)			
Papilionidæ Ailesi (13)			
Pieridæ Ailesi (39)			
Riodinidæ Ailesi (1)			
Satyridæ Ailesi (81)			

Figure 2: Screenshot of the website displaying the information about all nine different butterfly families

Ana Sayfa » Türler » Hesperidæ Ailesi


**Figure 3: Screenshot of the website displaying the photographs of the butterfly types belonging to the Family Hesperiiidae**

Following the exploration of the information given in these pages, the pre-service teachers decided to focus on two families with a similar number of butterfly types (the Family Hesperiiidae with 43 types and the Family Pieridae with 39 types) for comparison purposes given in the assignment. To collect information about these two families, they then began to click on each photo as seen in Figure 3. For example, Figure 4 shows various information about *Muschampia poggei* from the Family Hesperiiidae, including numerical (frequencies of photographs by cities), geographic (distributions of butterflies across Turkey), photographic (captured images of butterflies with/without identified gender), tabular (months they were seen) and graphical (distribution of altitudes seen by months) representations. From this information, they selected to focus on the image of butterfly with unidentified gender (since not all types had gender information) to generate attributes, such as wing color and wing appearance which were considered as distinguishing characters of butterflies, and the seasons the butterflies seen, which involved grouping of months data given on the website.



**Figure 4: Screenshot of the website displaying various information about *Muschampia poggei* from the Family Hesperiiidae**

After this extensive exploration on the website, the pre-service teachers began to make an ordered list for each type of butterflies in both families, including the name of the butterfly, its photo, and the season it is seen, on a Word document. As part of the course assignment, they needed to use CODAP to analyze the data to answer questions considered in the Formulate statistical investigative questions phase of the statistical problem-solving process in Figure 1. Three questions emerged during this exploration: 1) What colors are the wings of the Family Hesperiiidae and the Family Pieridae? 2) How do the appearances of the wings of the Family Hesperiiidae and the Family Pieridae look like? 3) What seasons do the Family Hesperiiidae and the Family Pieridae tend to be seen? Then they constructed a

table, as seen in Figure 5, displaying the family name, type name, wing color, wing appearance, and season seen in columns respectively and case values in rows.

Kelebek Aileleri	Kelebek türleri	Kanat Rengi	Kanat görünümü	Yaşadıkları Mevsimler
Hesperiidae (Zıpzıplar ailesi)	Pyrgus cinaerae (Güzel Zıpzıp)	Siyah	Desenli	Yaz
	Pyrgus serratulae (Zeytuni zıpzıp)	Kahve	Desenli	İlkbahar-Yaz
	Eogenes alcides (Alsides Zıpzıpı)	Beyaz	Düz	Yaz
	Pyrgus sidae (Sarıbandlı Zıpzıp)	Siyah	Desenli	İlkbahar-yaz
	Pyrgus jupei (Kafkasya Zıpzıpı)	Kahve	Desenli	Yaz

Figure 5: Part of the table created for organizing the data

Organizing the data as shown in Figure 5 made entering them into CODAP table relatively easier due to its case-based structure. One difficulty the pre-service teachers encountered in this process was having more than one value for season variable for some cases, such as *Pyrgus serratulae* seen in both spring and summer. Therefore, they chose to create two different data sets, one including 70 cases of butterflies in two families with family names, wing color and wing appearance variables (Figure 6, table on the left) and the other including 120 cases with season variable for two families (Figure 6, table on the right). Due to the repeated cases for each season a butterfly seen, the number of cases was increased in the second data set. The pre-service teachers decided to separate multiple seasons since they wanted to answer their third question: What seasons do the Family Hesperidae and the Family Pieridae tend to be seen? In either of the data sets, they also did not include type name which was displayed in the table in Figure 5 as they considered that having the type names would not allow to see two categories (the families) which were compared in the data analysis.

Yeni Veri Seti			
Durumlar (70 cases)			
in- dex	Kelebek Aileleri	Kelebeklerin Kanat Renkleri	Kelebeklerin Kanat Görünümü
1	Hesperiidae (Zıpzıplar ailesi)	Siyah	Desenli
2	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
3	Hesperiidae (Zıpzıplar ailesi)	Beyaz	Düz
4	Hesperiidae (Zıpzıplar ailesi)	Siyah	Desenli
5	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
6	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
7	Hesperiidae (Zıpzıplar ailesi)	Boz	Desenli
8	Hesperiidae (Zıpzıplar ailesi)	Turuncu	Desenli
9	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
10	Hesperiidae (Zıpzıplar ailesi)	Boz	Desenli
11	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
12	Hesperiidae (Zıpzıplar ailesi)	Boz	Desenli
13	Hesperiidae (Zıpzıplar ailesi)	Siyah	Düz
14	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
15	Hesperiidae (Zıpzıplar ailesi)	Beyaz	Desenli
16	Hesperiidae (Zıpzıplar ailesi)	Boz	Desenli
17	Hesperiidae (Zıpzıplar ailesi)	Boz	Düz
18	Hesperiidae (Zıpzıplar ailesi)	Siyah	Desenli
19	Hesperiidae (Zıpzıplar ailesi)	Kahve	Desenli
20	Hesperiidae (Zıpzıplar ailesi)	Boz	Düz

Yeni Veri Seti			
Durumlar (120 cases)			
in- dex	Kelebek Aileleri	Kelebeklerin Uçtuğu Mevsimler	
1	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
2	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
3	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
4	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
5	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
6	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
7	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
8	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
9	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
10	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
11	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
12	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
13	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
14	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
15	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
16	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
17	Hesperiidae (Zıpzıplar ailesi)	İlkbahar	
18	Hesperiidae (Zıpzıplar ailesi)	Yaz	
19	Hesperiidae (Zıpzıplar ailesi)	Yaz	
20	Hesperiidae (Zıpzıplar ailesi)	Yaz	

Figure 6: The data set tables created in CODAP

## Concluding remarks

This paper presented insights from a group of pre-service teachers' experience of acquiring and organizing image-based data to analyze categorical data by comparing two groups as part of a course assignment. As pointed out by Bargagliotti et al. (2020), the pre-service teachers spent a great deal of time to explore the non-traditional data involving photographs and other types of data collected by others for specific purposes as mentioned in the Kelebek-Türk website. In accord with the findings of Gould et al. (2017) and Wilkerson et al. (2021), data handling and structuring became an important component of the statistical investigation before formulating statistical questions. They needed to consider attributes/variables and criteria for identifying categoric values for them, such as assigning values, plain or pattern, for wing appearance based on the dots or lines seen in the photos of butterflies. The pre-service teachers did not seem to have any issue with preparing the data for the analysis in CODAP as they were able to consider the data as attribute-based. This could be due to their prior experiences with using CODAP for data analysis during the Statistics course taken in the previous semester. However, transforming the initial table created by hand (Figure 5) into a single data set table in CODAP was challenging for them. Since their initial table included subsets of data within the season variable, the flat table did not work. Instead, they needed to create hierarchical data sets, which is possible in CODAP, but this feature was not used in the course before. It was also evident in the work of Konold et al. (2017) that the use of hierarchical tables in structuring data by hand was more intuitive, but it could be challenging to use such a structure that is already built in CODAP environment.

## 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.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.
- Gould, R., Bargagliotti, A., & Johnson, T. (2017). An analysis of secondary teachers' reasoning with participatory sensing data. *Statistics Education Research Journal*, 16(2), 305-334.
- Konold, C., & Miller, C. D. (2011). *TinkerPlots 2.0: Dynamic data exploration*. Emeryville, CA: Key Curriculum.
- Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal*, 16(2).
- Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2021). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, 1-19. <https://doi.org/10.1080/10986065.2021.1922838>