USING NON-TRADITIONAL DATA TO BUILD DATA DISPOSITIONS

Leticia Perez[1], Sheri Johnson[2], and Pip Arnold[3]
[1]WestEd, Portland, Oregon, USA
[2]The Mount Vernon School, Atlanta, Georgia, USA
[3]Karekare Education Limited, Auckland, New Zealand
lperez2@wested.org

Working with non-traditional data such as personal photographs, teacher trainees identified themes and considered different dispositions that learners cultivate when working with data. By classifying and questioning elements within a set of shared photographs, teacher trainees were able to generate traditional data elements for analysis and reflect upon the role of identity, awareness, and curiosity during a statistical investigation. Additionally, technical tools were used to glean data from the photographs. By first estimating values, the teacher trainees compared machine-generated data with human evaluations. Throughout the process of developing a story about their data, teacher trainees were encouraged to take ethical considerations into account and identify issues that are important to consider when working with learners and data.

INTRODUCTION

As learners study statistics, they are developing dispositions for working with data. These dispositions involve learners in establishing identity, awareness, and curiosity. A data disposition of (a) *identity* involves considering human values, cultures, experiences, and beliefs; (b) *awareness* involves being aware of the implications of data use and its potential effects; and (c) *curiosity* involves being curious and using data to dig deeply into questions and coaxing attributes from existing data (Arnold et al., 2017). Data dispositions of identity, awareness, and curiosity can be used to develop and support learners in uncovering ethical questions and dilemmas often "hidden" within statistical investigations. When learners interrogate the data and ask questions with a sense of curiosity, they are developing data dispositions that can help improve interpretations, inform decision-making, and improve day-to-day operations. Thus, they are undertaking activities similar to what data scientists and statisticians do (Lee et al., 2022). At the school level, teacher trainees will be working with novice statisticians and data scientists; thus, we can use semi-structured activities to encourage learners to act and think like data scientists. For example, working with enormous amounts of data can occur in real-time, include non-traditional formats, and capture more information than might be realized by the data producer (Arnold, Perez, & Johnson, 2021; Franklin & Bargagliotti, 2020; Pangrazio & Selwyn, 2020). Data scientists are needed worldwide across industries to use and make sense of the data collected.

Traditional data are numerical or categorical and collected as counts, measurements, or defined categories. Nowadays, vast amounts of data are generated through digital devices, such as smartphones, smart-watches, and onboard computer systems, resulting in large quantities of secondary data. The data produced by these digital devices often include photographs, videos, sound bites, tweets, and the meta-data associated with them. These non-traditional data allow broader statistical investigations as additional variables, and information can be drawn directly from the data itself, for example, data in the form of a photograph (Bargagliotti et al., 2021). Future teachers must be prepared to help students finish secondary school statistically literate, that is, equipped with data dispositions and skills to work with traditional and non-traditional data sources. They need data dispositions that support and develop their awareness of and curiosity about the data. These dispositions prepare them to think about the implications of the data and be critical data consumers. In this paper, we discuss a data activity that involves using photographs as data sources and illustrates how this data activity can explicitly build data dispositions throughout the statistical problem-solving process (see Figure 1).

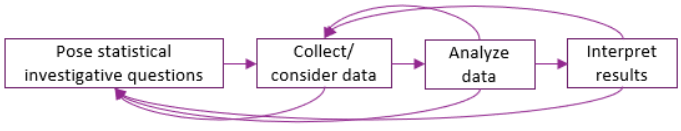

Figure 1. Statistical problem-solving process (Bargagliotti et al., 2020)

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

THE DATA ACTIVITY

The data activity took place within the context of a secondary teacher trainee program. The group was made up of 21 secondary math teacher trainees. The data activity began with a 20-minute introductory session one week, followed by a two-hour online workshop session the next week, after which the teacher trainees completed a feedback and reflection form. All authors were involved in the online workshop. The online workshop had a dual focus of engaging the teacher trainees in the data activity as adult learners as well as having the teacher trainees reflect and discuss the implications of using the activity with their students in the future. In this sense, the data activity positioned teacher trainees both as learners and teachers. The data activity (Table 1) pulled together elements of three extended teaching lessons that were developed by the authors to show how the context of *using photographs as data sources to tell stories about our favorite outdoor spaces* highlighted different elements of the statistical problem-solving process referenced in Figure 1 (Arnold, Johnson, & Perez, 2021). The teacher trainee experience went beyond sharing the available data lessons. It also included practitioner moves around data that are often invisible to novice educators or omitted in curricular materials. For example, in preparing for the activity, the authors chose to use ethics as a cross-cutting theme—teachers need to be deliberate about embedding the topic of ethical practices within statistical investigations for their students and not leave this inclusion to chance.

Table 1. Overview of the data activity

Action	Teacher Trainee Actions
0. Preparation (20 minutes previous week)	Using photographs on the teacher trainees' camera rolls as a prompt, teachers were invited to brainstorm a theme/topic of interest and possible statistical investigative questions. The data was *collected* through photos. Teachers, if they felt comfortable, shared a photo from their camera roll or elsewhere (e.g., royalty-free photos from the internet) for each of the identified themes. The group chose three themes: food, nature, and animals.
Between sessions	Three Google slide decks were created, one for each theme. Within each slide deck was one slide that included all submitted photos for a theme. The photos were reduced to the same height. See Figure 2 for photos related to the nature theme.
1. Ethics discussion	Consideration of ethics starts in the early stages of any statistical investigation. It should be considered in the problem stage (*formulate statistical investigative question*), and more deeply in the planning stage (*collect and consider data*). Time was spent exploring key ethical practices (see Table 2).
2. Getting more data from the participants	Using the photos as prompts, this step involves teacher trainees with exploring additional investigative questions and subsequent *survey questions* in order to collect additional data about the topic of interest. We interrogated both the investigative and survey questions by considering whether the teachers would be comfortable and happy sharing information and answering honestly.
3. Getting more data from the photographs	Using various digital and analog tools, the teacher trainees *gathered additional data* about their photos. This included (a) using a grid to find the percentage of the photo an animal takes up in animal photos, (b) using http://www.coolphptools.com/color_extract to find the five most frequent colors in food photos, and (c) finding the longitude and latitude of nature photos.
4. Data entry	Additional survey questions from action 2 were entered into a Google form, and additional data collection questions (*gather the data*) were generated in part 3. Teacher trainees entered their answers into the Google form.
5. Analyzing data and interpreting results	The data captured was uploaded into a Common Online Data Analysis Platform (CODAP) document that was shared with the class (https://codap.concord.org/). Individuals and groups discussed their themes, made *data visualizations,* and told a *data story* about what they found out about the class.

Ethical considerations should be at the forefront of our work with data and apply both to data we collect ourselves and to data that has been collected by others (Open Data Institute, 2021). Six ethical data practices (Arnold, 2022) were used during planning and during the activity to highlight the ways that ethics, human values, and decisions can impact a statistical investigation. Key ethical practices and considerations are organized in Table 2. The second and third columns show the teacher trainees' ideas about ethical practices and considerations for each of the six ethical data practices as the teachers discussed ethics in anticipation of doing the activity with their own students and what they should consider when collecting photos from their students.

Table 2. Teacher trainee ideas about ethical considerations for using their photos as a data source

Key ethical practices	Example of ethical practice within the lesson	Example of ethical consideration
Obtain informed and voluntary consent from participants.	Will people be happy sharing their photos with others?	Participation is voluntary.
Minimize the risk of harm to participants.	What if the students do not have a camera roll or a smartphone?	Option to source photos from the internet and awareness of permissions to use.
Respect participants' rights to withdraw at any time without giving a reason.	Are there consequences for students not sharing their photos?	Make involvement optional and allow students to withdraw their photos at any stage.
Respect and protect participants' privacy and confidentiality.	Will students have to identify which photo is theirs?	Students share photos with the teacher, who removes identifying data.
Avoid unnecessary deception to participants.	Are students aware of what their photos are used for?	Clearly explain what data will be used from the photos.
Be socially and culturally sensitive to participants' cultural and religious perspectives.	Will the photo theme be appropriate/accessible for all members of our class?	Check-in (anonymously) with the class about the appropriateness of topics. Would they be happy to provide a photo on the topic?

DATA DISPOSITIONS IN PLANNING FOR DATA COLLECTION

Curiosity

Curiosity sets the stage for statistical explorations. In this activity, we engaged teacher trainees' curiosity about themselves, their peers, and what is happening in the world around them. Statistical fluency and data dispositions can be developed by tapping into learners' desires for freedom of choice, voice, and representation of thinking. Using photos on teacher trainees' camera rolls as a stimulus, they brainstormed statistical investigative questions and designed how to collect data to answer these questions. This was accomplished by asking survey questions and by using technology to create data elements (Arnold, Johnson, & Perez, 2021; Arnold, Perez, & Johnson, 2021). The initial prompt to start the statistical problem-solving process was to have teacher trainees look at their camera rolls and consider what their photos could reveal about class members' interests. The teacher trainees worked in groups, and their curiosity kicked off mini data cycles as they scanned their camera rolls, chose photos, and then thought about possible themes from their group's collection. As a class, ideas were shared, and three themes were agreed upon. During this initial phase of theme identification, the teachers were primed to uncover ethical considerations involved with collecting data in the form of photos from their personal camera rolls. For example, one teacher trainee reflected, "There's a lot more information to a photo than one initially expects. We can each interpret a photo however we want." Choosing to focus on ethical practices as a cross-cutting theme provided an opportunity to develop dispositions of identity and awareness.

Identity and Awareness

When learners first start to think about ethical considerations it can be hard for them to put themselves in someone else's shoes. Taking time to build *personas* that highlight different identities and cultural perspectives that are relative to a task's context can be useful. For example, when considering food as a theme, the teacher trainees gained an awareness of others by responding to the prompt, "For whom and why might this topic make a person uncomfortable?" In the process of considering different personas that are impacted by the investigation, the teachers were confronted with a decision to continue, refine, or alter their focus. Alongside thinking about the topics to explore, learners should also discuss potential issues that might arise with the photos once they are submitted. This creates an opportunity for awareness of other cultures and how their interests and likes might vary from their own culture. Food that is considered a delicacy by one culture might be deemed inedible in other cultures—a topic that arose later with the teacher trainees in this activity. Learners come to respect and suspend judgment of the data (photos). By considering these personas or perspectives, learners are developing an awareness outside themselves and recognize that not everyone has the same experiences. This awareness can inform learners' decisions and considerations about topics for the data collection.

DATA DISPOSITIONS WHEN GETTING MORE DATA

Curiosity

The data activity highlights the role of maintaining a curious stance towards the data itself and the individuals the data are collected from. "Coaxing data" describes a process whereby various operations or procedures are applied to data to reveal attributes previously unnoticed. An infinity star cube (e.g., https://youtu.be/Wbw1GchCWmo) can help us imagine where the original dataset is acted upon to reveal attributes initially hidden to the observer. At first pass, the infinity cube appears to be made of interlocking cubes, but further manipulation reveals that the cubes are larger units hiding triangular prisms. As technological tools develop, we can process our data, in this case our cube, through sets of manipulations or "passes" to reveal new attributes. In this section, we will describe the series of passes teacher trainees took to coax numerical and categorical data from the photos.

Maintaining a sense of curiosity throughout the investigative cycle is a crucial data disposition. Many novices believe that after data are collected and variables are named and displayed, no changes can be made to the dataset. For example, the teacher trainees did not share all of their camera roll photos with each other. Instead, they made a series of data collection decisions during the preparation action (Action 0, Table 1). First, individuals were asked, "What types of photos do you have on your phone?" Then, the teacher trainees were placed into three small groups and were asked to first identify a shared photo theme. This theme identification represented a "first pass" of informal data collection. The teacher trainees then submitted a photo for each of the three identified themes: food, nature, and animals. In Action 2 (Table 1), "Getting more data from the participants," the teacher trainees sorted the photos for a particular theme around attributes visible within the photo, such as color amongst nature photos (Figure 2).

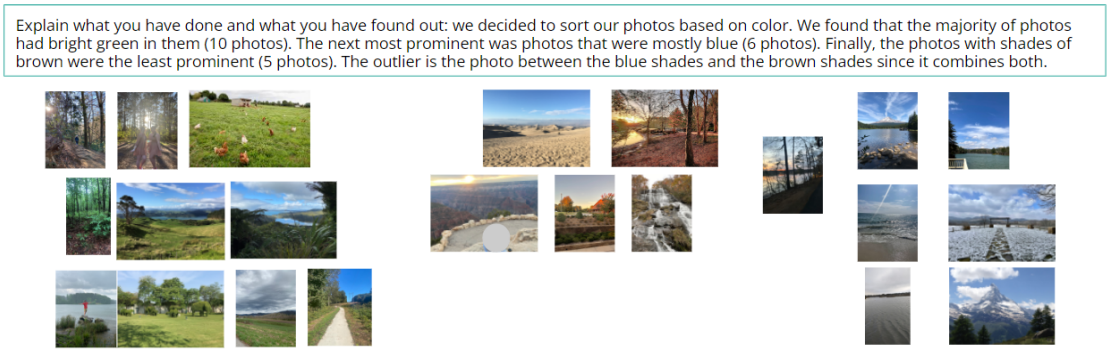

Figure 2. Example of nature photograph sort by teacher trainees

Sorting the photos by attributes visible in photos represents a "second pass" at data collection. The dialogue generated during the photo sort supported the development of survey questions. These

survey questions were subsequently answered by the photo owner, revealing additional information about the submitted photo, a "third pass" at data collection. This "third pass" resulted in additional data about the photos, including both categorical and numerical variables. Interestingly, teacher trainees noted that the number of themes and variables would likely differ across different teacher trainee or student populations. The resulting dataset includes data that were coaxed through a series of data collection passes that were driven by the learners' curiosity.

Maintaining a sense of curiosity about the photos and the data from the survey questions drove teacher trainees to consider a "fourth pass" at data generation by considering, "What else can we do, or what else would we want to know about the photographs?" In action 4 (Table 1), the wonderings that arose from the "fourth pass" led the group to further describe and quantify other attributes of the photos (e.g., landscape versus portrait orientation, pixel resolution, location of where the photo was taken), which were added as additional variables to the dataset.

Consider teacher trainees who sort nature photos based on color (see Figure 2). If prompted to reflect upon human factors that might lead to variation within photo descriptions, the teachers would likely indicate that visually impaired individuals may not be able to describe photos easily or consistently as either blue or green. An awareness of data accessibility around both data collection and representation offers yet another opportunity to incorporate ethics within the investigation. Are there other techniques we can employ to coax data that are either more reliable or at least offer a different technique other than human vision? For example, the "Image Color Extract PHP tool" (http://www.coolphptools.com/color_extract) lists the most frequent pixel colors for the photos. This information can inspire another set of investigative questions in which a class might look at the dominant pixel color of the photographs versus the dominant color identified by the group. The teacher trainees engaged in these activities displayed the disposition of curiosity through their multiple approaches to create additional variables from the photos. For example, a teacher trainee reflected, "I thought it was interesting how we sorted and analyzed each picture. We saw so many different ways [in this lesson] to look at one thing."

Identity and Awareness

"For the photographs submitted about animals, what additional information might be coaxed from our photos?" Posing this question could produce a variety of responses including counting the number of animals, describing the types of animals, whether the animals are domesticated or wild, picture orientation, or whether the animal was the subject of the photo. Additional quantitative elements could be collected by overlaying a 10x10 grid over the photograph to determine the percentage of the photo that includes the animal. Further classification of the animal as either a "pet" or "not a pet" provides an opportunity for learners to explore data issues around identity. Whether an animal is considered a pet is deeply connected to cultural norms and traditions that may not be shared by the classroom or a larger audience. Learners can engage in a discussion exploring these notions prior to proceeding with this classification. For example, teacher trainees posed the data collection question, "Would you eat this animal?", demonstrating an opportunity for the instructor to consider ethics within the lesson and build data dispositions that are reflective of an agile data scientist.

Although permissions to use the photo may have been granted, individuals may not recognize the impact of their consent with such artifacts. For example, in this task, teacher trainees were asked to see if any personal information was still attached to the photo. Often emails, GPS data, usernames, or camera/phone type accompany photos directly shared through a file transfer platform such as Google. The data we can coax from the photos may be exciting and interesting but comes with implications and costs that can cause harm, especially if published. Teacher trainees were directed to think back about the photo submitted of the woman holding her cat and consider, "What if the home address or GPS location were made available within the photo? How might we as a group help protect members of our class through this dataset?" A teacher trainee noted, "There are many ethical-related considerations to make when it comes to using photos as data. This may include considerations for learners' experiences, culture, and/or privacy."

Applying these technological approaches provides the opportunity to experiment with the dataset itself, expose ethical dilemmas, develop data dispositions, and possibly pose new investigative questions. This sequence leverages technology in ways that encourage the use of photos as a non-traditional data source to create datasets, as well as offers an approach to quantify underlying attributes

such as natural vs human made. We can then apply traditional statistical measures and representations to these datasets, affording learners the opportunity to build data dispositions within the context of an exploration rather than hypothetical situations.

IMPLICATIONS FOR THEORY AND PRACTICE

Learner-chosen topics for data exploration have the potential to engage learners deeply in statistical and data science practices. Educators need to pay attention to providing a range of experiences and cross-cutting themes, ensuring all learners have access to these experiences and build data dispositions over the course of their schooling. Whereas ethics was used as a cross-cutting theme in the session with the teacher trainees, the application of another cross-cutting theme such as Computational Thinking (CT) would likely surface a different set of data dispositions. For example, learners using the CT practice of decomposition would likely develop a disposition that tolerates ill-defined problems. Learners engaging in the CT practice of abstraction are seeking new ways of representing and communicating data-based relationships, an important disposition. In this activity, the teacher trainees had the opportunity to both reflect on their own identity and its connection to data collection. Additionally, they built an awareness of the beliefs and identities of others, thereby developing the data disposition of identity through engaging in the data investigation and learning about ethical practices. These approaches provide educators with the opportunity to develop culturally relevant lessons that center learner voice, choice, and ways of thinking about data. The teacher trainee feedback submitted at the end of the lesson, the source of many of the quotes in this paper, highlights the impact human beliefs, decisions, and approaches can have on data generation and subsequently the interpretations that are produced. This is a crucial aspect of data education typically omitted within K-12 settings. Interaction with non-traditional datasets generated collaboratively through photographs creates an opportunity for teachers to explicitly explore ethics and data dispositions in authentic contexts.

REFERENCES

Arnold, P. (2022). *Statistical investigations | Te tūhuratanga tauanga*. New Zealand Council for Educational Research.

Arnold, P., Johnson, S., & Perez, L. (2021, October 21). Using photographs as data sources to tell stories about our favorite outdoor spaces. *Statistics Teacher.* https://www.statisticsteacher.org/2021/10/21/using-photographs-as-data-sources/

Arnold, P., Krehbiel, A., Polman, J., Cramer, C., Vahey, P., Wright, T., Tally, B., Berg, S., Parris, J., Schultheis, L., Kolodner, J., & Hunter-Thomson, K. (2017, February). *Disposition/habit of minds/beliefs/capabilities—what we want to see* [Conference discussion session]. Data Science Education Technology (DSET) Conference. Berkeley, California.

Arnold, P., Perez, L., & Johnson, S. (2021). Using photographs as data sources to tell stories. *Harvard Data Science Review*, *3*(4). https://doi.org/10.1162/99608f92.f0a7df71

Bargagliotti, A., Arnold, P., & Franklin, C. (2021). GAISE II: Bringing data into classrooms. *Mathematics Teacher: Learning and Teaching PK-12, 114*(6), 424–435. https://doi.org/10.5951/MTLT.2020.0343

Bargagliotti, A., Franklin, C. A., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education*. American Statistical Association; National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Franklin, C., & Bargagliotti, A. (2020). Introducing GAISE II: A guideline for precollege statistics and data science education. *Harvard Data Science Review*, *2*(4). https://doi.org/10.1162/99608f92.246107bb

Lee, H., Mojica, G., Thrasher, E., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal, 21*(2), Article 3. https://doi.org/10.52041/serj.v21i2.41

Open Data Institute. (2021, June 28). *The Data Ethics Canvas: Methodology and training tools.* https://theodi.org/article/the-data-ethics-canvas-2021/#1563365825519-a247d445-ab2d

Pangrazio, L., & Selwyn, N. (2021). Towards a school-based 'critical data education.' *Pedagogy, Culture & Society, 29*(3), 431–448. https://doi.org/10.1080/14681366.2020.1747527