

## **Data are everywhere – In every class and even outside of school**

Andee Rubin

TERC, Cambridge, U.S.A.

[andee\\_rubin@terc.edu](mailto:andee_rubin@terc.edu)

*This paper aims to expand our ideas of where to find data beyond the “usual places,” i.e. math, statistics and possibly science classes. It encourages readers to develop a “data lens” in order to see data everywhere. In the paper, I share the characteristics of a “data lens” and provide several examples of data activities for students of all ages, including museums, zoos, aquariums and after-school clubs.*

### INTRODUCTION

When I was growing up, “statistics” was an adult word, something people studied in college or beyond. I didn’t really know what it meant. And “data” was certainly not a topic of conversation at the dinner table. In retrospect, I see that my father, a men’s wear store owner, was constantly analyzing data (although I’m sure he didn’t think of it that way) – how many of those leather suits sold? How does business this year compare with past years? Is there a relationship between how the local football team is doing and sales? (This was not an idle question: many of the people who shopped at my father’s store were football players!). Little did I know then that I was surrounded by data – and that I would spend decades of my career figuring out how to get people of all ages to engage productively with data.

The point of my keynote-paper is to expand our ideas of where to look for data beyond math classrooms to other school subjects and to out-of-school contexts. I will share with you a bunch of examples from work I’ve done over the past 35 years in places as diverse as kindergarten classes, zoos and science centers. My hope is to help you develop a “data lens” to look at the world, so that you can see potential data experiences everywhere you look.

### A DATA LENS

What is a “data lens”? In a commentary paper I wrote in a special issue of the Journal of the Learning Sciences focused on data (Rubin, 2020), I identified five fundamental characteristics that are central to reasoning with data. While I made this list before “data science” and “AI” became everyday topics, these fundamentals haven’t changed and still provide the basis for making meaning of data.

1. Context – Unlike the quantities students sometimes deal with in math class, all data have context. They are not “naked numbers” and interpreting them means, inevitably, considering what was measured, how, by whom, etc. (See added #6 below.)
2. Variability – Data always exhibit some kind of variability – whether it’s multiple measures of the same thing under different conditions, a quantity that changes over time, or a measure on different individuals, part of dealing with data is understanding and making sense of variability.
3. Visualization – While it’s possible to make sense of data from a list or a table, it is much easier to see where data are most common, note unusual values, see relationships from an image. Data in non-academic contexts are often presented in the form of visual images, so learning both to create and interpret data visualizations is a key skill.
4. Aggregate – Children (and adults) often focus on individual data points when they first encounter data. But as they become more sophisticated in their approach to data, they learn to examine all of the data distribution at once, to see it in the aggregate. Measures of center (mean, median), and variability (variance, standard deviation) are among the ways we describe data distributions – these measures are one of the ways we deal with variability (#2).
5. Inference/generalizability – Any time we note a pattern or draw a conclusion from data, we need to also consider how well what we see would apply to other similar data – perhaps from a larger population or a different classroom or people of a different age. While this idea gets formalized in the process of “statistical inference,” young students can begin to develop a habit of mind to always ask: “Do we think this same pattern will show up with different data? How sure are we?”

In the past few years, I have added one more characteristic of data that will, I believe, become increasingly important in the “AI age.”

6. Measurement/data production. Data are not the same as the quantities we decide to measure. The measurement process, in my mind, is what turns “life” into “data” – and all of the decisions along the way have a major impact on just what the data capture – and what they don’t. Just consider asking a group of people how many years they have had their current job. A bunch of questions immediately arise: what if I am at the same organization in a different role? What if I changed from full time to part time? What if I took a sabbatical and came back? The answer to each of these questions will determine the number each person reports. And when we work with data that someone else has collected, it’s critical to ask all of the questions about how it was produced: where, by whom, how, why, when and where.

With this definition of a data lens, I will describe several examples of data that can engage people of all ages. In each case, I describe the activity and then, importantly, specify what aspects of data fluency the activity could support.

### ANIMALS MAKE GREAT DATA

What could be more engaging for people of all ages than animals? With a few moments’ thought, it’s easy to think of a variety of attributes we could record about individual animals. In the early 1990s, the developers of the Used Numbers data curriculum realized that they had access to a bunch of cats (their pets and some in shelters) that they could easily turn into data – and, thus, the “cat cards” were born. Each card has a picture of a cat on one side (Harmony was my cat back then) and data about the cat on the other. We chose a combination of numerical and qualitative measures and managed to get our cats to stand still long enough to measure their tails and their bodies. Here are the two sides of the card with my cat (long gone now), Harmony.



Figure 1. Front of Harmony’s Data Card.

Because each cat was on a separate card, elementary school children could move them around, sort them, make piles of similar cats or even “adopt” one of the cats to tell other students about. They could explore questions about, for example, the frequency of different eye or pad colors by making multiple piles of cards. They could find the oldest and youngest cats by lining up the cards in order. They could separate the cards into “males” and “females” and then compare patterns they found for each subgroup.

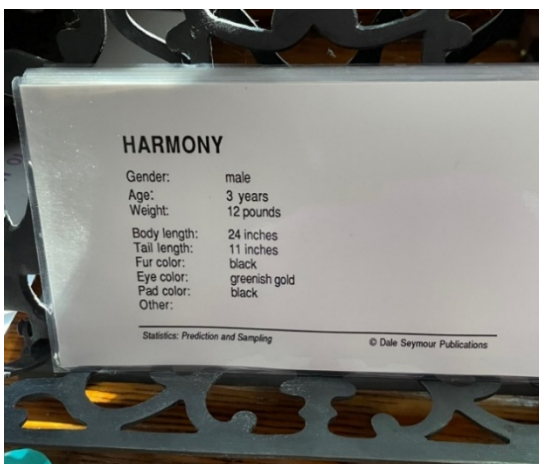


Figure 2. Back of Harmony’s Data Card.

While the physicality of the cards made them very accessible for young students, it also restricted the complexity of analyses that could be carried out. So, these cards were eventually expanded to a digital data set, allowing upper elementary students to explore the data more robustly, originally in TinkerPlots and later in CODAP. While the cat cards were designed to be used in school, the idea of cards with animal images and data was easily extendable to out-of-school settings in which families could engage together in data activities. In a project involving multiple zoos and aquariums, educators created animal cards for penguins, flamingos and goats in their zoo to support visitors’ engagement with the animals and to get them to work with data in an accessible way.



Figure 3. Goat Data Card from the Phoenix Zoo.

At the Phoenix Zoo, one of the popular exhibits contains several dozen goats. Each goat has a name and many were born at the Phoenix Zoo. Here is one example of the goat cards – this one for Abraham, who was born at the Phoenix Zoo in 1999, weighs 122 pounds and has black and white hair. Zoo educators sometimes used the cards with family groups before they visited the animals, as an introduction to what they would see. Families would naturally sort them into groups (there were two sizes of goats, one much smaller than the other) and children would often find a favorite goat and then watch out especially for him or her.

Let’s consider what data-related concepts students or families are learning from these activities. One is case-attribute structure; the individual cards are a physical embodiment of cases and each one explicitly lists the attributes associated with that case. Learners are immediately introduced to variability – the animals vary in terms of numerical variables such as weight, length, and age, as qualitative ones such as gender and color. Finally, these cards give learners an opportunity to sort – including to define their own categories.

Flamingo Observation Chart (s = seconds)					
15s	Aggression 	Foraging 	Preening 	Rest/Stand 	Walking 
30s	Aggression 	Foraging 	Preening 	Rest/Stand 	Walking 
45s	Aggression 	Foraging 	Preening 	Rest/Stand 	Walking 
60s	Aggression 	Foraging 	Preening 	Rest/Stand 	Walking 

Figure 4. Data Collection Sheet for Flamingo Behaviours.

Pictures of animals are fun – but animal behavior can be even more engaging – and another source of data. In the zoo and aquarium project, educators also designed several “Be An Animal Scientist” activities. The most extensive was tracking flamingo behavior. The zoo staff identified five distinguishable flamingo behaviors (aggression, walking, foraging, preening and resting/standing) and developed “flamingo behavior flashcards” with a photo of the flamingos carrying out that behavior and a cartoon depiction of each. They recruited participants by standing by the flamingo exhibit asking: “Who would like to be an animal scientist?” Once a family expressed interest, they were each given a data collection sheet, told to choose a particular flamingo to observe, and instructions to circle the behavior they saw every fifteen seconds for a minute – a total of 4 observations. Here is a copy of the data collection sheet, complete with human interpretations of the flamingo behaviors.

After the data were collected, zoo educators helped visitors add their data to a growing bar chart of flamingo behavior. After a few days of data collection, zoo staff started to wonder if there was a difference in the frequency of different behaviors in the morning vs. the afternoon, so they set up two different graphs, one for each part of the day and could then engage visitors in comparing them. As it turned out, the flamingos were more likely to forage at some times of day than others – a first hypothesis was that foraging took energy, so in the hottest part of the day, they were less likely to do so.

What were visitors learning about data in this brief encounter? One important learning is about the process and complexity of measurement. Visitors got a sense of how important it is to have a clear



measurement process (one observation every 15 seconds) and a shared definition of what to observe (accomplished in this situation with the behavior flash cards and the data collection sheet). Given the complexity of flamingo behavior, it seems that they would have also encountered ambiguity in measurement – did that flamingo just get aggressive or was that just a random peck? Through cooperatively creating the behavior graph (shown on the left), they learned about seeing the aggregate patterns in a dataset through the power of visualization.

Figure 5. A Zoo Visitor Adding Data to a Graph of Flamingo Behaviors.

### EXPLORING NON-TRADITIONAL VISUALISATIONS

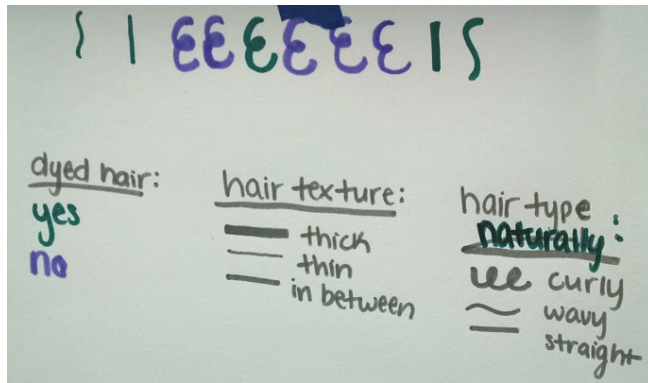
I've been inspired for several years by the book *Dear Data* (Lupi and Posavec, 2016), which tells the story of two data-loving graphic artists who got to know one another over the course of a year by trading postcards with data representations of aspects of their lives. Each week, they chose some part of their life to focus on (e.g. checking the time, saying thank you, sounds), collected data over the next seven days, and represented it on a postcard, which they sent to the other. Because they were graphic artists, their data representations are colorful and beautiful as well as informative. Key to making sure their data representations were understandable to their “data pen pal” was a clear legend that identified what each design aspect of the representation reflected about their data. One of my favorites (and one of the simplest!) is this one, representing the clothes in Georgia’s closet. In case you can’t read it, her text at the bottom brags about getting rid of four bags of clothes she hadn’t worn in years after drawing her representation.

This more artistic approach to data representation appeals to some students whose relationship with math has been less than ideal and who may, therefore, have some hesitation to engage with data. In the Data Clubs project ([www.terc.edu/dataclubs](http://www.terc.edu/dataclubs)), my colleagues and I tried this approach with middle school students to see how it would work. With a group of middle school girls, we focused on hair, a topic of considerable interest to them. As a group, we first brainstormed what hair attributes we would collect data on, then settled on three: Was it dyed? What texture was it? What hair type is it? We all recognized that we would have to clarify the details of these questions and decide on possible responses in order to have comparable data. We decided that the “dyed” question was, more specifically, “Has your hair EVER been dyed?” and the possible responses were just “yes” and “no.” The three possible “texture” responses were “thick,” “thin,” and “in between.” And the three choices for “hair type” were “straight,” “wavy,” and “curly,” with an additional specification that this was what the natural hair type – not chemically produced. With a quickly designed data sheet in hand, the girls in our group went down the hall to gather data from another group to add to their own data.



Figure 6. Data representation of clothes in Georgia’s closet from *Dear Data*.

It can be difficult to get youth to consider ways to represent data that aren't the traditional ways they have learned in school. For this Dear Data experiment to be successful, we knew that the attributes we chose would have to be categorical, so we had steered the original conversation about attributes in that direction. Because students haven't had as much experience graphing categorical variables, they are less likely to gravitate toward a conventional X-Y graph.



My favorite data representation produced by one of the participating girls is this one. Notice how she has indicated the three elements of the legend – color, line thickness and shape – clearly below the data representation. The elements of the representation look like hair – they are much more evocative than the dots we usually use on graphs. And it's easy to look at each image and read what it represents.

Figure 7. Non-traditional Data Representation of Hair Created by a Participant in Data Clubs.

What did the youth learn about working with data from this experience? In order to get appropriate data, participants had to carefully design their data collection methods, including how to pose the question and what possible responses to accept. But perhaps the biggest learning was how the grammar of data visualization works. When graphs are made automatically by software – or by rote in the classroom – the crucial mapping of the values of attributes to their visual representation in a graph is taken for granted. In this Dear Data-inspired activity, that mapping, laid out in the legend, is explicit and a central part of the activity.

#### READING DATA VISUALISATIONS FROM THE MEDIA

In a recent project (<https://www.terc.edu/data-science-in-museum-exhibits/>), I worked with museum educators to consider how they could engage their visitors with data science. One of the partner museums came up with a simple idea that turned particular types of data visualizations into a game they called Data Viz Quiz. The idea was to remove the labels in a set of visualizations, provide the labels separately, and ask visitors to guess which visualization matched which label. Here is one example of such an activity.

Common injuries for boys among popular high school sports  
Injuries per 10,000 competition plays

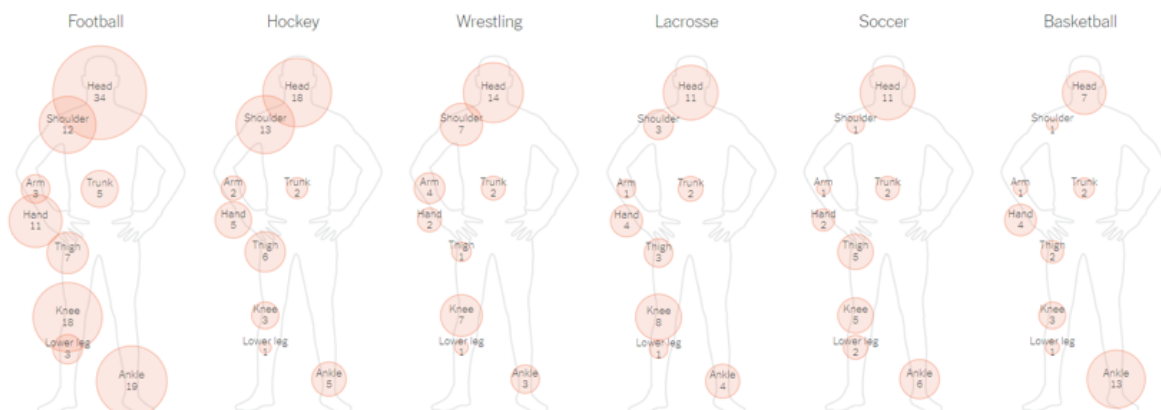


Figure 8. Data Representation of Injuries Sustained in Several Sports.

To get a sense of the experience visitors had, first look at the data representation without looking at the labels above each figure (perhaps cover them with your hand). Note that the data represented are common injuries for BOYS who are in HIGH SCHOOL and the units are “injuries per 10,000 competition plays.” The sports represented are (not in order!) wrestling, basketball, football, soccer, hockey and lacrosse. Which figure do you think goes with each sport? What information are you using to support your guess? Do you find yourself moving your body to simulate playing the sports to gauge the possibility of injury to different body parts? Now uncover the labels. What surprised you, if anything?

In our project, groups of visitors explored this visualization together and we observed them acting out the sports, pointing to parts of their bodies, telling stories about their sports experiences, and making data-based arguments about their guesses.

The learnings from this experience are similar to those in the Dear Data-inspired activities, but since visitors are interpreting a data visualization rather than creating one, they are learning to pay attention to – and to question – how data were measured (e.g. this data is just about boys) and how data values are represented.

### MORE SERIOUS TOPICS

While most of the topics described above are relatively light-hearted, the skills and dispositions both youth and adults develop are critical for living in our data-drenched world. Here are two last quick examples that are more closely related to contemporary issues. The first is a children’s book, Yumbo Gumbo (Dawson, 2024), that illustrates how data (specifically voting!) can help us make decisions that take multiple people’s opinions into account. Set in the Louisiana bayou, the book describes how a family decides what kind of gumbo (a stew made by Cajun families) to have for dinner. Multiple attempts at voting end up in ties, leading the family to come up with a creative solution that honors multiple preferences. If only our political leaders could do the same! This book is written for children ages 4 – 7 or so, but it contains important lessons for adults.

Finally, there are many examples of projects and programs that ask slightly older students (middle school and above) to engage with data about social conditions, health measures, poverty etc. on country, city or neighborhood levels. One of these is the Civic Data project, which provides middle school social studies students with UN-gathered data on the countries of North America and the Caribbean, including measures like happiness index, life expectancy, average years of schooling, per capita GNP. Students note inequities among the countries and investigate how each country’s history (including colonization) may have impacted their current status. Here, students are immediately faced with issues of aggregation, as the data are not about individual people, but rather, about whole countries. In this context, they need to be aware of the variability hidden in the data they work with. In all of the other examples in this paper, the data learners encounter is at the individual level and a challenge is to create and understand an aggregate measure. In the Civic Data example, the data are presented as an aggregate and a challenge is to appreciate the variability.

My final recommendation: take your data lens with you as you move through the world. You will see data everywhere!

### ACKNOWLEDGEMENTS

Work described here was funded by NSF under grants DRL-1742255, DRL-2215060 and DRL-2247675.

### REFERENCES

- Dawson, K. V. (2024). *Yumbo Gumbo*. (K. Crumpton, Illus.). Charlesbridge.
- Lupi, G., & Posavec, S. (2016). *Dear Data: A Friendship in 52 Weeks of Postcards*. Princeton Architectural Press.
- Rubin, A. (2020). Learning to Reason with Data: How Did We Get Here and What Do We Know? *Journal of the Learning Sciences*, 29(1), 154–164. <https://doi.org/10.1080/10508406.2019.1705665>
- Used Numbers (1989). Palo Alto, CA: Dale Seymour Publications.