# Distributions, Trends, and Contradictions: A Case Study in Sensemaking With Interactive Data Visualizations

Vasiliki Laina, Tufts University, Vasiliki.Laina@tufts.edu
Michelle Wilkerson, University of California, Berkeley, mwilkers@berkeley.edu

**Abstract:** Data visualizations are transforming how information is communicated. Educators should understand and support how youth reason about these complex representational artifacts. We analyze an interview in which young learners interpreted an interactive visualization as describing both relative distribution and absolute trends, and resolved conflicts related to this dual interpretation by re-aligning mathematical and personal knowledge. This case reflects a need for attention to multidimensional data and narrative design in research on data visualization literacy.

**Keywords**: data visualization, case study, technology, data analysis

## Introduction and background

Novel, narratively driven visualizations of data are transforming how knowledge is communicated in popular media and many professional disciplines. Such visualizations are designed to foreground certain "stories" through data, or to allow viewers to pursue certain lines of inquiry through those data (Segel & Heer, 2010). Sometimes the intentions or emphases of a given visualization, however, might not connect to learners' interests or prior knowledge. Understanding and supporting how learners work with such visualizations—how they find, explore, and critically interpret patterns in terms of the systems they describe—is important as educators seek to integrate data visualization into the K-12 curriculum, and to prepare data literate learners.

Making sense of data requires learners to construct and coordinate understandings of evident or implicit patterns in data, and of the situation for which the data were collected. There is evidence that students can productively engage in this back-and-forth, with varying levels of success. Lehrer and Schauble (2004) showed that, while making sense of a representation, students leveraged contextual knowledge to explain patterns they noticed, often supplementing their own observations of the actual phenomenon. Ben-Zvi and Aridor-Berger (2016) described a pair of students gradually developing connections between their understanding of context and data during a statistics exercise. Schwartz and Martin (2004) argued that when students coordinate between data, representations, and mathematics, they "…go through a process of invention, noticing, and revision that helps them develop insight into the relation between representations and the quantities they represent" (p. 138).

We posit that narrative visualizations—designed to emphasize certain stories or lines of inquiry through complex, multidimensional datasets—can introduce additional complexities to this coordination process. Here, we present a case study that illustrates this complexity. While working to make sense of a visualization illustrating fuel consumption trends, a pair of middle-grade students developed a double interpretation of the data, one focusing on distribution (emphasized in the visualization), and one focusing on trends over time (driven by their knowledge of global energy trends). These two interpretations created logical conflicts that emerged in the girls' interactions with one another and with the visualizations. They overcame these conflicts by re-aligning their interpretations of both the data and the situation. Since many visualizations feature both distributions and change over time, and since we found evidence of such conflations in other interviews, we argue that this paper sheds light on a number of specific themes worthy of further study.

## Methods

The DataSketch project studies how middle school youth think and learn about data visualization via interviews, classroom studies, and design activities. Here, we focus on one approximately 45 minute interview with a pair of girls, Aphrodite (6th grade) and Stryker (8th grade; participants' chosen pseudonyms). Both were members of an all-girls after school science club where interviews were conducted, and knew each other prior to the study.

During the interview, participants were asked to look at two interactive data visualizations. For the purposes of this paper, we focus on Aphrodite and Stryker's work with one visualization, described in detail below. The interview was semi-structured, based on an eight-question protocol designed to probe how learners understood and navigated connections between data presented, the intended message of the designers, and learners' understandings of the situation represented. Questions included "What is the message you think this is trying to send?", or "What would you expect this [visualization] to look like in the future?". We used video and screen capture to record participants' interactions with one another, the interviewer, and the visualization.

## The visualization

The visualization we used concerns the total primary energy consumption of different fuels in the UK each decade from 1970 to 2010 (Figure 1; EvoEnergy, 2011). It is an interactive tree with circle "leaves" of different colors and sizes. To the right of the tree is a legend with fuel types and corresponding colors; to the left, there is a list of years. When the user hovers over a fuel type, a corresponding percent value is shown on the tree. When a user hovers over the years, circles on the tree change colors to reflect the relative distribution of fuel types for that year. The size of the tree, and position of "leaves", stay the same. Fuel consumption is always presented as a relative percent of energy consumption; and no information about total consumption is provided.

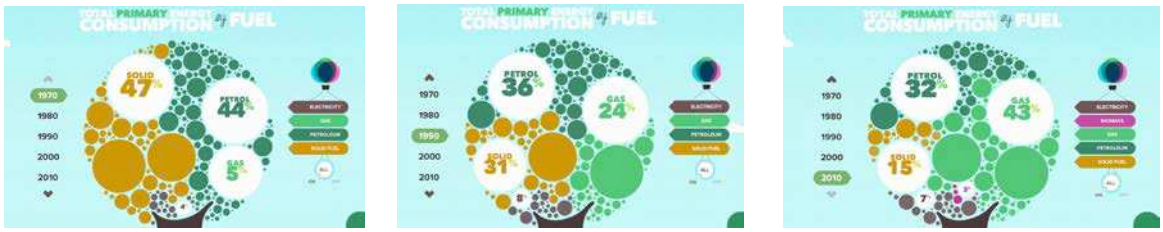


Figure 1. Snapshots of the visualization displaying data for 1970 (left), 1990 (middle) and 2010 (right).

We selected this visualization because some features are relatively conventional (area as percentage), while others are unexpected (apparent meaninglessness of circle size). There are also multiple ways to "follow" data: Using years on the left to view distributions over time, or fuels on the right to focus on only one fuel.

## Case selection and analysis

We selected this case as the focus of this short report for three reasons. First, it serves as a particularly explicit example of a phenomenon we have observed in other interviews, whereby learners conflate descriptions of distributions with descriptions of trends over time. Second, the participants were talkative and comfortable working with the data, which makes more explicit and available for study the nature of such conflation. Third, novel interactive visualizations of the sort popular in media and science communication *often* involve data about distributions and trends over time, and as such are likely to trigger similar questions and conflicts. Our analysis identified themes in learners' treatment of data, interpretations, and ways of resolving conflict (Aronson, 1995).

# Findings

We report two main findings. First, we will show the girls interpreted data in the visualization in two different ways, as describing (i) distribution and (ii) trends over time. This was evidenced in the girls' references to percentages as describing *relative consumption of fuel*, or change in percentages over time as describing change in *actual consumption of fuel*. We call these usages 'relative measure' and 'absolute quantity', respectively.

These two interpretations can be in conflict. Imagine 5 red marbles are added to a collection of 10 red and 10 blue marbles. The percent distribution of this collection would change from 50% red and 50% blue to 60% red and 40% blue. An 'absolute quantity' interpretation would emphasize the decrease in percentage of blue marbles from 50% to 40%, perhaps suggesting there are fewer blue marbles in the collection. A 'relative measure' interpretation would emphasize that the blue marbles account for a smaller part of the collection.

Next, we will argue that the ways in which participants shifted between interpreting the data as relative measure or actual consumption, and the ways they addressed conflicts that emerged from these different usages, were driven by their negotiations between contextual and mathematical knowledge.

## Multiple interpretations of the same data

When introduced to the visualization, Aphrodite and Stryker quickly began attending to and noting changes in the percentages of different types of fuels over the years available in the visualization. When offered paper and markers, they re-arranged the data, grouping by fuel type rather than year and calculating change (Figure 2).

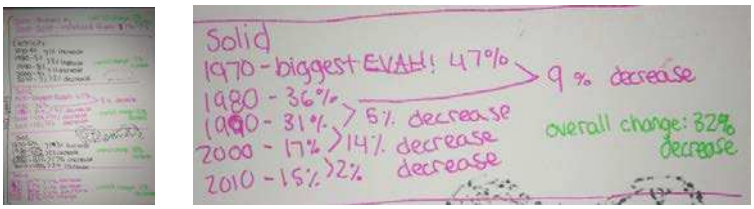


Figure 2. The girls' written work grouped percentages by fuel (left), and computed change over time (right).

Throughout the rest of the interview, the girls treated percentages sometimes as describing relative measures and other times as describing absolute quantities. For example, when she noticed the percent electricity consumption featured in the visualization decreased from 2000 to 2010, Aphrodite was initially surprised and exclaimed "It goes down! Ha!" This implies she was expecting the data to reflect a known trend—a rise in total consumption from 2000 to 2010. Stryker, however, pointed out that since a new fuel type, biomass, was introduced in the simulation "Everything should go down," which implies that she was attending to the distribution of fuels and how the total (100%) is distributed among its parts. These interpretations became more explicit later, when Stryker suggested that in 2020 the visualization would feature "green energy" as a new fuel source:

| | |
|---|---|
| Stryker: | We're adding in another variable. Which [inaudible] everyone will go down some, to make room for green energy. |
| Aphrodite: | True. |
| Stryker: | But I think biomass will still go up. |

Here Stryker again attended to percent as a relative measure; suggesting everything should 'make room' for the new variable. But soon after this, while still discussing how the tree might look in 2020, Aphrodite extrapolated a pattern without attending to that distribution, suggesting a usage of percent as an absolute quantity:

| | |
|---|---|
| Aphrodite: | Solid… How much [inaudible]? That's been going down down down down [looking on the screen, browsing along the years]. |
| Stryker: | I think we could put solid at 14. |
| Aphrodite: | 14? I feel like it should be a little lower than that. |
| Stryker: | But... [inaudible] |
| Aphrodite: | 13, 12. Cause it's been going down a lot. |

While the excerpts above suggest that the two different participants leveraged two different treatments, there was evidence throughout the interview that both participants moved back and forth between both interpretations of the same data. This is illustrated clearly in the next excerpt:

| | |
|---|---|
| Interviewer: | This one [refers to biomass] was the… the most stable one? |
| Stryker: | Yeah. |
| Aphrodite: | 2%... 2% increase. |
| Stryker: | Even.. No one really use it, but, it stayed steady. |

Stryker's last utterance combined elements of both relative measure and absolute quantity. On the one hand, she claimed that 'no one really' uses biomass, thus attending to its low percentage in the energy distribution. On the other, she claimed that a 2% increase from 2000 to 2010 meant that biomass 'stayed steady', thus treating these numbers as representing quantities and 2% implying an insignificant change in its absolute consumption.

## Re-aligning content knowledge to interpretation of data

These two interpretations of data introduced tensions: between what the girls argued the data showed, their own expectations, and the visualizations' emphases. This tension was reflected in the girls' early reorganization of the data in the visualization, and persisted throughout. The girls stated often that they were concerned about the environment. They described decreases in percentages as 'good' and increases as 'bad', suggested the tree should be heart shaped to encourage users to "love the Earth", and described the visualization's underlying message as "The Earth is gonna die soon." These concerns were not supported by the data, given the girls' dual interpretation. When asked what they notice in general about the visualization, Stryker and Aphrodite noted:

| | |
|---|---|
| Stryker: | So, these... Most of them sadly... [finding total change on the paper] Em... Two of the biggest ones, petrol and solid decreased over the… how long is it? |
| Aphrodite: | Overall it's gas that's been up the most. [inaudible] And… down the most… solid. So gas went up the most and solid went down the most. |

Given the participants' environmental concerns, it seems Stryker was preparing to say "most [fuel consumption] sadly *increased*", the use of the word 'sadly' highlighting the environmental consequences of fuel consumption on Earth. When this increase (presumed about absolute quantities) wasn't reflected in the data (intended as relative measure), Stryker paused and changed what she observed about the data, which Aphrodite re-voiced.

This realignment also happened when the interviewer explicitly highlighted a mathematical conflict that emerged from the girls' dual interpretation of the percentages as relative measures or absolute quantities. At some point Aphrodite argued that the (absolute) consumption of petrol did not change "compared to the others", even while both girls argued that car use had increased through the years. The interviewer followed up:

> Interviewer: So, if they use their cars... if we have more cars, actually, than in 1970s, and we use petrol mostly in cars, why do you think it didn't change much? The consumption?
>
> Stryker: There are new ways to power cars now.

Stryker invoked her personal content knowledge—alternative sources of fuel for cars—to justify an inconsistency in their interpretation of the data available, that petrol stayed stable despite increases in car use. Both of these examples illustrate adjustments that the girls made to their descriptions of data, and to their descriptions of the context in which those data were collected, to better align the two throughout the interview.

## Discussion and conclusion

Narrative visualizations, designed to be visually pleasing and to emphasize stories with data, are an important part of how information is communicated. These visualizations introduce complexities to the interpretation process of learners: they involve complex, multivariate data, explicitly organized to emphasize a certain point of view or path for investigation. Thus in addition to coordinating understandings of data and context as has been explored in the literature, dealing with narrative visualizations requires learners to coordinate *multiple* possible patterns embedded within the data, and with designers' often explicit intentions.

In this paper, we explored how one pair of middle school learners, Stryker and Aphrodite, navigated this territory. By closely examining the students' discussion, interactions with the visualization and the way they decided to reorganize the data, we found that both participants interpreted the data provided in an interactive visualization of energy consumption by different fuel types in two potentially contradictory ways: as 'relative measures' describing distribution in terms of percentages, and as 'absolute quantities' describing total increases or decreases in energy consumption over time. These dual interpretations created conflicts between the participants' understanding, the data, the representation, and their knowledge of actual energy consumption trends. They addressed these conflicts by realigning interpretations fluidly—reorganizing and manipulating data presented, adjusting interpretations of those data, and adjusting justifications for patterns observed. In addition to the issues of distribution, trend, and conflict emphasized in our analysis, we note that certain interpretations might have been triggered by the way the visualization was designed. For example, the size of the tree stayed the same, which could have led the girls to assume that the total energy consumption had remained unchanged.

This case represents a first step in work on how learners reason with complex, narratively organized data. It details specific themes—the conflation of distribution and trend, and the adjustment of data and interpretation—as possible starting points in this investigation, and highlights ways these issues may manifest.

## References


Abelson R.P. (2004). *Statistics as Principled Argument,* Hillsdale, NJ: Lawrence Erlbaum

Aronson, J. (1995). A pragmatic view of thematic analysis. *The Qualitative Report, 2*(1), 1-3.

Ben-Zvi, D., & Aridor, K. (2016). Children's wonder how to wander between data and context. In D. Ben-Zvi and K. Makar. (Eds.), *The Teaching and Learning of Statistics: International Perspectives* (pp. 24-36). Cham, Switzerland: Springer.

EvoEnergy (2011). [Data sourced from 2011 factsheet by UK National Statistics.] The interactive U.K. energy consumption guide. Retrieved from http://www.evoenergy.co.uk/uk-energy-guide/

Lehrer, R. & Schauble, L. (2004) Modelling natural variation through distribution. *American Educational Research Journal,* 41(3), p. 635-679.

Schwartz, D., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition & Instruction, 22*(2), 129-184.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualizations and Computer Graphics, 16*(6), p. 1139-1148.


## Acknowledgments


This work is funded by the National Science Foundation Grant IIS-1350282. We thank the youth participants and collaborators at the after school program in this project for their time and partnership.