WORKING WITH DATA

Highlights of Related Research

By Clifford Konold and Traci L. Higgins

DRAFT VERSION: 1/31/01

Theme 1: **Forming a Statistical Question**

Theme 2: **Differentiating Data from the Real Situation**

Theme 3: **Creating and Interpreting Data Displays**

Theme 4: **Representing Non-occurring Events and Data Values of Zero**

Theme 5: **Viewing Data as an Aggregate**

Theme 6: **Summarizing Data with Averages**

Theme 7: **Comparing Groups**

Theme 8: **Relating Data Back to the Real Situation**

This chapter is to appear in S.J. Russell, D. Schifter, & V. Bastable, (in press), *Developing mathematical ideas: Collecting, representing, and analyzing data*. Parsippany, NJ: Dale Seymour Publications, a module of the K-6 professional-development curriculum, *Developing Mathematical Ideas*. The writing of this chapter was supported by the National Science Foundation under Grant Nos. REC-9725228, ESI-9730683, ESI-9731064, and ESI-9818946. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Introduction

We can conceptualize data investigations as involving a four-stage process: 1) ask a question, 2) collect data, 3) analyze data, 4) form and communicate conclusions. Real research, however, seldom proceeds in this orderly fashion. One reason is that conscientious researchers often find themselves retracing their steps. While writing the report, they think of another analysis to do and perhaps return to the study site to collect more data. But research does not proceed linearly for a more profound reason, which is that these research phases are not independent components.

Experienced researchers look forward from the beginning. Although they don’t analyze data before collecting them, they imagine doing so and make guesses about what they will find. They develop and refine their questions and decide what data to collect by thinking ahead to the kind of conclusions they would like to make, the statistical methods they can use, and to their intended audience. Experienced researchers also look backwards. When it’s time to analyze the data, they do so from the perspective of their original question, testing the intuitions they started with against what the data reveal. And typically their questions evolve and change as they discover unanticipated results in the data. In these respects, data analysis is like a give-and-take conversation between the hunches researchers have about some phenomenon and what the data have to say about those hunches.

It’s important to keep this more complex picture of data analysis in mind as we consider what both the Working with Data Casebook and the research literature tell us about students’ statistical thinking. Simplistic views can lead to the use of recipe approaches to reasoning with data and to the treatment of data as numbers only, stripped of context and practical importance. Conversely, staying grounded in the data and attentive to what they have to say keeps the tools of data analysis — the collecting, graphing, and averaging — in their appropriate, supportive role.

Although there is considerable research on the reasoning of college students, there is relatively little on how younger students reason and learn about data. Because data analysis has only recently become an integral part of the pre-college curriculum in the United States, we have limited practical experience with what works and what doesn’t. Accordingly, we draw heavily in this chapter on what we, as researchers, have learned from the episodes in the Working with Data Casebook, connecting our observations when we can to published findings. In our opinion, the reflections of these teachers and their descriptions of students’ thinking is one of the richest source of information to date on children’s reasoning about data and on how children’s thinking evolves during instruction.

Theme 1. Forming a Statistical Question

Turning observations into data involves an explicit process of abstraction. In this process, we transform a question about the real world into a statistical question, one we can answer with data. Young students begin to struggle productively with this process as they discover, often while designing surveys, how difficult it is to pose a question that different people interpret in the same way.

A data investigation usually begins with a question about the real world. For example, students at a K-8 school believed that the water from fountains on the third floor was better than water from the floors below (Rosebery, Warren, & Conant, 1992). A combined class of seventh and eighth grade bilingual students decided to see whether there was a difference in the taste of water from different floors.

Coming up with an interesting question is often the first step in a data investigation. However, before collecting data we must transform our initial question, which is often too general, into a more specific, statistical question, one that we can answer with data. In the above example we might reformulate the question as "In a blind taste test of water samples from each of the three floors, which sample will most students prefer? The statistical question allows us to develop measurement instruments and data collection procedures that we can use to collect the data. Rosebery et al. (1992) do not provide details of the study, but we can presume that students made a number of decisions before collecting data: Who would they use as tasters? How many should they test? Should tasters drink directly from fountains or from cups? Should the same students taste all three water samples? How should tasters indicate their preference? Such decisions are part of the process of making a general question a statistical one that can be answered with data.

Elementary students draw on their own experiences as they learn how to formulate statistical questions. By thinking about how they themselves would answer a proposed survey question, for example, they quickly discover not only the range of possible responses, but that there are multiple interpretations of a question, and that the wording of the question matters. In the words of a second grader, "Everyone has to understand your question. If they don't understand your question, everyone will be answering any old way" (case 10, pg. 41).

In case 8, Nadia challenged her fifth graders to re-think the wording of a question so that it would not be interpreted in "any old way." The students initially proposed to include on their survey the question: "Do you speak more than one language?" (p. 35). Nadia asked:

> How do we know when someone speaks another language? For example, is knowing how to say, "Where is the bathroom?" in French speaking French?

One student responded, “No. We mean speaking fluently.” This, in turn, raised the issue of what is meant by “speaking fluently.” The students resolved the problem through more discussion.

We carefully shape questions not only so that people interpret them in the same way but also to get the information we are interested in. It is easy to become so engrossed in formulating a more precise question that we lose track of what we wanted to know in the first place. Case 10 describes a pair of second graders, Nadia and Keith, who were interested in finding out from fellow students, “How many states have you visited?” (p. 42). They quickly realized that “visited” could be interpreted in many different ways. Nadia offered further criteria for defining a visit:

> [A] visit could only count if you were going to that state for a specific purpose other than simply driving through to reach another destination. Airports could not count. If you stayed with a friend from out of state it could only count if you really, really wanted to see them and you stayed with them for more than a day. (p. 42)

Despite these criteria, they phrased the question in their final survey as: “How many states have you ever set foot in?” (p. 42). They apparently adopted this wording at Keith’s prompting because, phrased this way, the question seemed clear. Nadia was not satisfied. She thought the phrase “set foot in” missed the point. She wanted to know whether students had traveled *to*, rather than *through*, a state. In transforming a general question to a statistical one, the challenge is not only finding a wording that people will interpret consistently but also making sure the statistical question gets at what you wanted to know in the first place.

Theme 2. Differentiating Data from the Real Situation

Data are records of things we observe happening in the world. As part of the process of creating data, students learn to differentiate the data from the events they observe. Students come to view data as objects in their own right that can be analyzed and queried in ways that the events they observed could not be.

In formulating questions and learning to collect and analyze data to answer them, students learn to walk two fine lines. They learn to see the data they create as in many ways separate from the real-world events they observe, while treating the numbers they generate as meaningful referents to those events. Distinguishing between data and the event entails a particularly delicate balance, because taken to the extreme, it can lead to reasoning about data as numbers only, stripped of the context that gives them meaning (Moore 1992).

“Creating data” may seem an odd phrasing. However, data are not lying around like melons on the ground to gather up and cart off to the table. Turning observations into data involves an explicit process of abstraction. Lehrer and Romberg (1996) claim that the “very idea of data entails a separation between the world and a representation of that world” (p. 70). In reasoning about data, students construct a model of some situation or event which, like any model, is only “a partial representation” (Hancock, Kaput, & Goldsmith, 1992, p. 339).

In creating data, we must consider what aspects of a situation we are most interested in and make sure that we explicitly record that information. Furthermore, we must record that information so that later when we or others look at the records, their meaning is clear. Like labeling a file “Recent articles,” it is easy when recording data to overlook things that are obvious at the time but will not be later. Students discover this as they work with data they have created. Fifth graders in a study by Hancock, Kaput, and Goldsmith (1992) collected data to determine which of three cafeteria meals students most liked. The student researchers conducted surveys in the cafeteria on different days, asking students whether they had “bought,” “brought,” or skipped lunch on that day (“none”). Their plan was to determine menu preferences by comparing the number of students buying versus bringing lunch on a particular day. Their rationale was that since daily menus were published in advance, many students would decide whether or not to bring a lunch on a particular day depending on whether they liked what was offered. As they began analyzing their data, the student researchers discovered that they had failed to record what meal was served. On the day they recorded the data, a mark under a column on the survey had a clear meaning, a meaning that was gone once they had forgotten the menu of the day.

When interpreting data, we must consider what information the data provide about the real-world event as well as the information they do not provide. The casebook includes numerous examples where younger children do not distinguish between the data and the situation they observed when they recorded the data. In case 21, Barbara gave each of her kindergarten students a bag of M&Ms to count. The class created a line plot using stick-on notes on which each student had recorded the number of M&Ms in his or her package (see p. 102). The teacher asked, “What can you tell from this graph?”

Faruh: We eat M&Ms.

Rocky: Joy has the most.

Desmond: We know how many I ate.

Tammy: Andrea’s is the most…because hers is a bigger number. (p. 102)

The students associated names with values even though the graph they were interpreting did not show who had counted each bag. Students were basing their interpretations on their memories of counting and eating M&Ms rather than on the data they had abstracted from that event. We see similar examples throughout the cases where data serve merely as pointers to the more complex event. In forming conclusions, these students draw without awareness on their memories of the event as well as on the objectified data. As a way to help her students begin to distinguish the information in the coded data from what they knew from observing the events, the teacher in this episode suggested to her students that they “pretend that the principal walks into our room and looks at this chart, what would he know from this chart?”

While we see some students treat data as if they provide more information than they explicitly do, we also see instances when students take a restrictive view of data, as if once collected they can use data to answer only the question they initially asked. Lehrer and Romberg (1996) worked with a group of fifth graders to design a survey of student interests. Among other things, the survey asked students to list their favorite school subject and their favorite winter sport and to estimate hours spent watching TV. After collecting the surveys, the instructors asked the student researchers to come up with questions that they could “ask about the data.” To these fifth graders, this request was ridiculous. Questions, they countered, could be posed to people, but certainly not to data. The instructors prompted the students with various examples, such as “Which is the least favorite school subject?” The students successfully used these examples to generate a list of similar questions. But they needed further assistance to see that they could answer many of these by analyzing the data they already had. Their initial impulse was to conduct another survey using these new questions.

As Hancock, Kaput and Goldsmith (1992) pointed out, once recorded, data become objects in their own right, objects which we can manipulate and query quite independently from the observations from which we abstracted them. Students can manipulate and organize data by stacking, grouping, and ordering — things they often couldn’t do, or do easily, to real events.

Because students can reorganize the data in a number of different ways, they can pose and answer questions that may not have occurred to them before collecting the data. For example, kindergarten and first grade students worked with data gathered from a school lunch count (case 1). While gathering data, students became concerned that there were only 15 pieces of data, but a total of 18 students in their class. The three missing students turned out to be absent that day. One student was intrigued that they had gotten an attendance count from data they had gathered about lunches. She wondered how this was possible:

> Can the clothespins [markers used to record answers to the survey questions] tell only one thing? If the clothespins tell us how many school lunches and how many home lunches there are, can they tell us how many are at school—I mean at the same time? (p. 3)

This student understood that the sum of the “yes” and “no” counts would equal the number of students in the class that day. But she struggled with the idea that “fifteen could stand for how many in school as well as how many students were getting lunch” (p. 3). We see her struggle as the beginning of the discovery that once recorded, data have a life of their own, and that in examining them, new questions may arise which the data can answer.

Sometimes we can operate on data much as we can the corresponding real-world events. Exploring these instances can help students develop confidence that by operating on data they can get additional information about the real events they observed. In case 32, Ana suggested that the way to decide whether third graders were taller than fourth graders was to have each group of students lay down end to end and measure them to determine their total height. The teacher asked whether they could use the height data they had already collected to figure out what would happen if they performed Ana’s physical experiment. She reminded the students that Celia earlier had claimed that the total of all the fourth graders’ recorded heights would be greater than the third graders’. Ana countered, “But how can you be sure. If we did it, we would know.” Leah explains, “There’s an easier way to do it. You can just add up all the numbers, and see which one was bigger. That’d be easier than having to go out and measure it” (p. 168).

Theme 3. Creating and Interpreting Data Displays

Different types of displays highlight different aspects of data. Younger students tend to make plots that allow them to identify and answer questions about individual data points. As they gain more experience with data, they begin using representations to answer questions about the data as a whole — how they are distributed and where they cluster.

In the form we first collect them, data are usually pretty useless. A stack of completed questionnaires is like a messy room in need of a good cleaning. To find what we want, we must organize the information. How we organize data depends on what we want to know.

Young students with little prompting construct a variety of spatial arrangements that serve to organize categorical data. Rosemary’s first graders (case 11) divided their papers into columns, rows, or quadrants and drew pictures or symbols to represent various types of games students played during recess (see Figure 1). When working with categorical data, children readily clump like responses together and from these clumps figure out which responses are more or less popular.

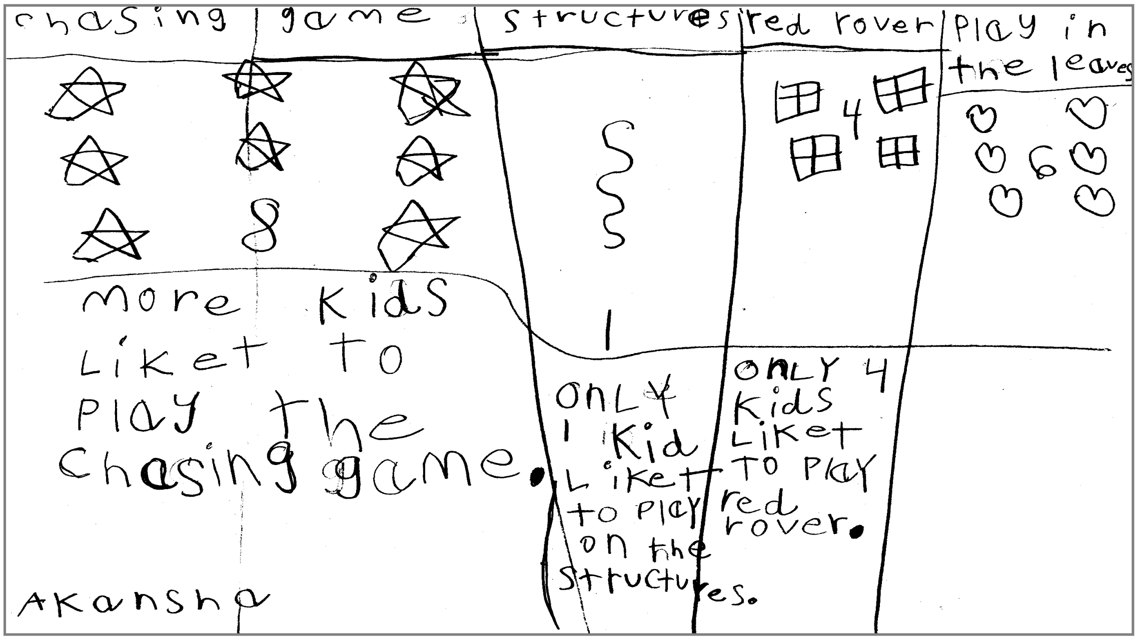


Figure 1. A student's representation of categorical data.

We also see young children use a variety of representations to spatially organize numerical data. One group of second graders in case 18 represented the number of teeth lost by classmates in tabular form ordering the data according to number of lost teeth (see Figure 2). The representation has considerable detail including, for each case, the student's name, a pictograph showing number of teeth lost, and the corresponding numeral written in two or three different locations. This representation would be useful for looking up how many teeth a particular student had lost or for quickly determining who had lost the most or least number of teeth. However, for other purposes, such as describing characteristics of the class as a whole, their representation would not be as useful.

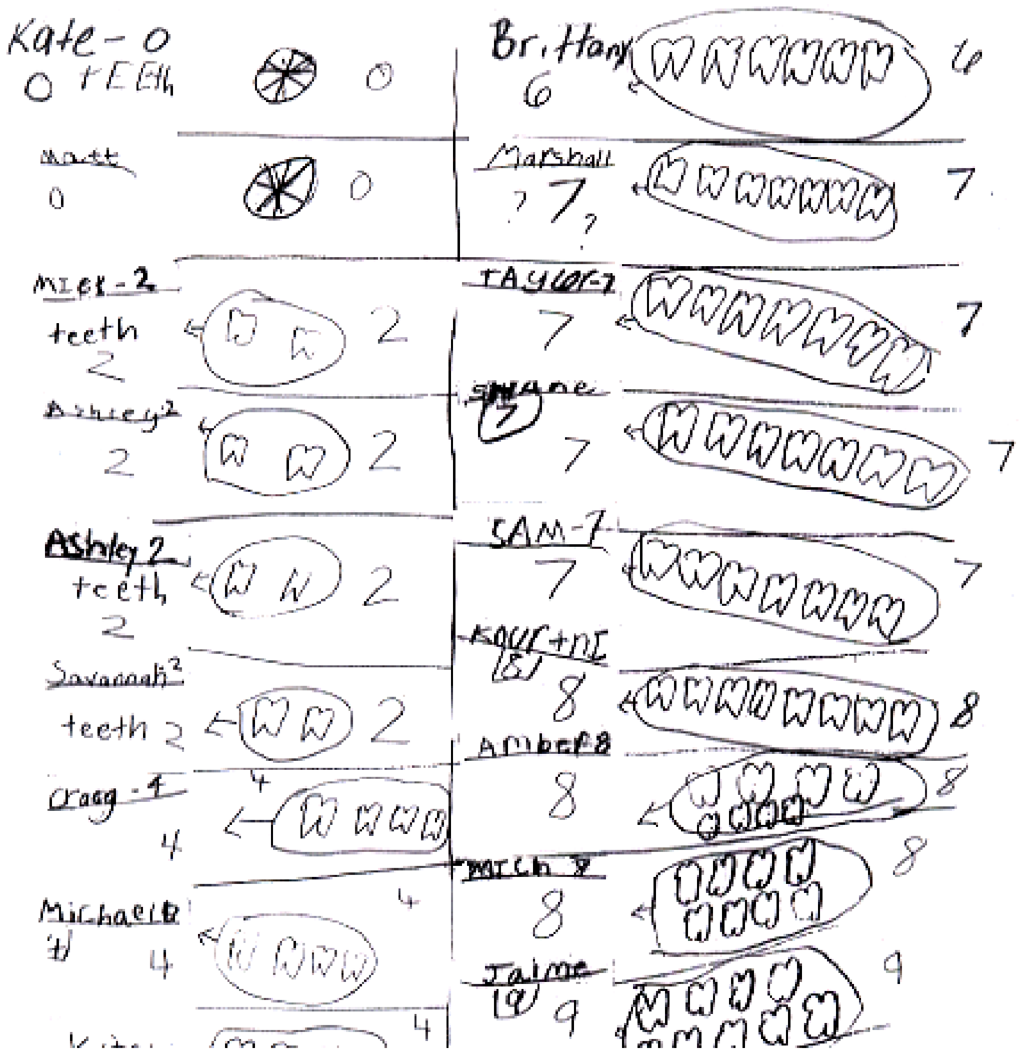


Figure 2. A tabular representation which identifies individual students and the number of teeth they had lost..

Another group of students in case 18 represented the same data with pictographs only, drawing faces with toothy grins. A student noted one of the disadvantages of these iconic representations: “It’s hard to draw teeth” (p. 82). Pictographs can include detail that takes time to render and seems to convey no useful information. It is important to notice, however, that even in making pictographs, students are abstracting elements from the observed events. For example, in representing the number of teeth lost, the teeth drawn in the mouth stood for missing teeth, and their location in the drawn mouth appeared unrelated to the location of gaps in students' mouths. Thus, pictographs are often a way for young students to begin to abstract or simplify information in the process of coding events.

With pictographs, students form explicit links between the data and the event, which may help them reason about the data in their appropriate context. It would therefore be a mistake to rush students into using more abstract forms. Furthermore, the level of abstraction appropriate for a particular representation depends on the questions students have. Given that younger students are drawn to question about who has the most and where they personally fall within a range of values, it makes sense that their representations make it easy to read off individual values and to identify who they belong to.

Different displays serve different purposes. Therefore, we should choose our particular display by considering our objective, the type of question we have and the audience we want to communicate to. Thus, one *type* of representation is not inherently superior to another. Graphs are not better than tables; bar graphs are not better than pictographs. Sconiers (1999) describes a project undertaken by kindergartners who were frequently asking their teacher for help tying shoes. The class reasoned that if they all knew which of them could tie shoes, then those who didn't know could get help from those who did. After conducting a survey, they posted a list of names of those who could tie shoes. Had the class not been trying to solve a specific problem, they might have made a graph showing how many students could and could not tie shoes, a useless plot for their purposes. The list worked.

Even decisions about how big to make a graph, whether to labeled axes or provide titles should depend on our purposes and should not be made according to a fixed list of "graph dos and don'ts." Suppose students wanted to quickly make a line plot to help them see how the data were distributed. It would be unnecessary in this case to fuss with the display or label the axes; doing so would be squandering time that could go into thinking about their question. On the other hand, if these same students made a graph to communicate their findings to the whole class, then labeling the axes and taking care to make the display easily readable would be critical to achieving their goal.

Although there is not a hierarchy of graph quality, some representations are harder than others to learn to interpret (Bright & Friel, 1998). Roth and Bowen (1994) described how as we represent numerical data with maps, lists, graphs, and equations, we move from concrete to increasingly abstract statistical representations. As we move along this continuum, information about individual data values becomes increasingly aggregated and obscured.

Case 18 provides a good example of plots showing different levels of aggregation. In Figure 2 above we can identify individual students and the teeth they lost. In

Figure 3, we still can identify individual teeth, but not the identities of the student to whom they belong. Finally in Figure 4, individual teeth are no longer evident.

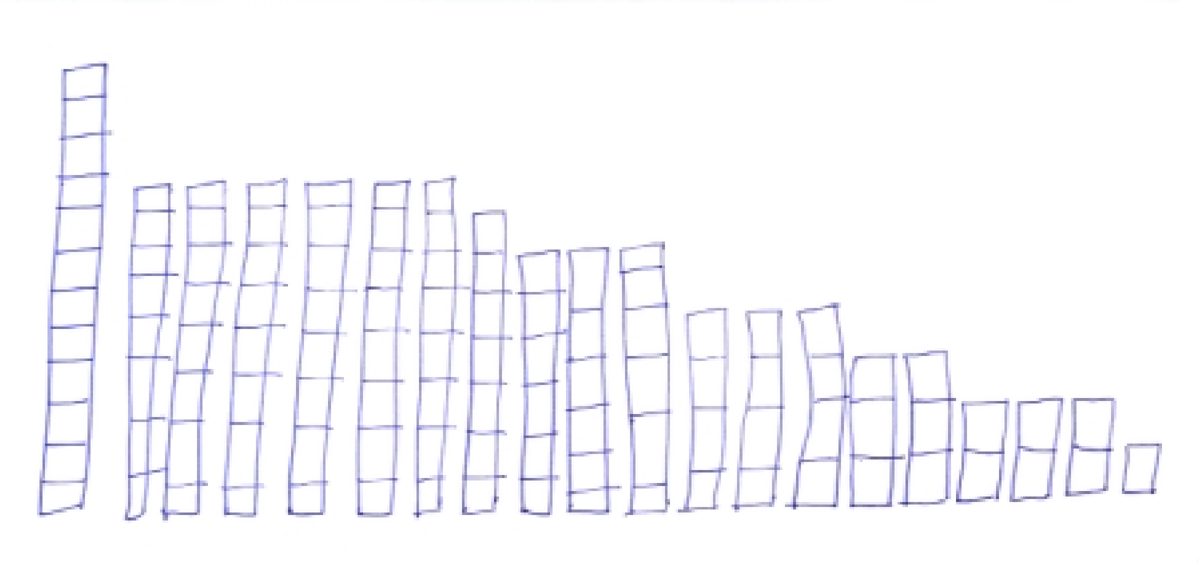


Figure 3. In this student's graph, each surveyed individual is represented by a stack of cubes showing the number of teeth lost.

Figure 4. In this student's graph, individual responses are aggregated to indicate the number of students who lost various numbers of teeth.

The individual cases "disappear" into larger aggregates. By increasing the level of aggregation, we can:

> perceive ever more general features of the data at the expense of being able to identify individual data values. It is easy to forget, however, the

> learning required to interpret the more abstract statistical plots. As a result, [educators] often encourage students to use plots and summaries before they sufficiently understand them and, by doing so, effectively pull the rug from beneath them. (Feldman, Konold, & Coulter, 2000, p. 119)

As Bright and Friel (1998) pointed out, graphs use axes and other display elements in a variety of ways. We can see some of these differences in the two plots that fifth grade students located in a textbook while searching for graphs they could use to compare two sets of data (case 5). With a quick look, the graphs seem to represent data in exactly the same way (see Figures 5 and 6). Look closer. Figure 5 is a "case value plot." As the name suggests, case value plots display the value of each element in the data set. In this instance, the bar lengths show the cost (value) of each costume item (case).

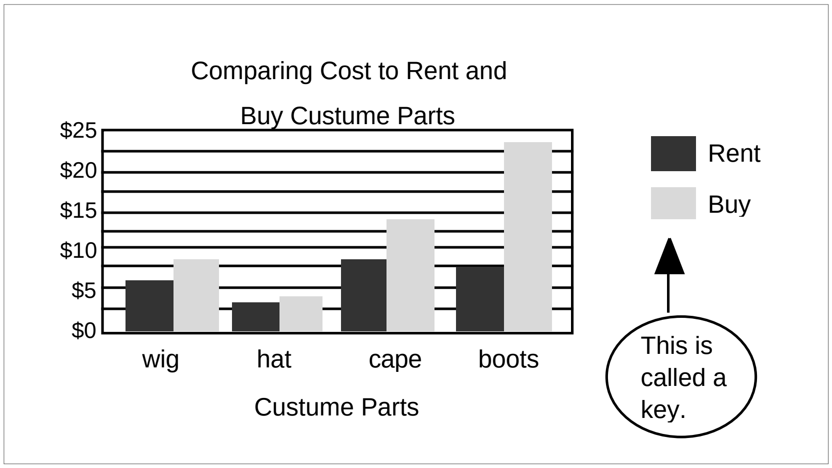
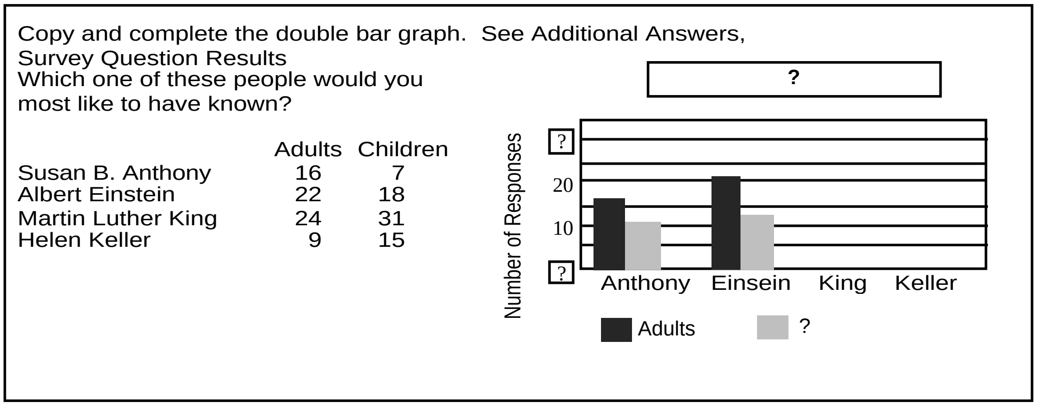


Figure 5. A case-value plot of the cost of renting vs. buying various costume parts.



Copy and complete the double bar graph. See Additional Answers,
Survey Question Results
Which one of these people would you most like to have known?

	Adults	Children
Susan B. Anthony	16	7
Albert Einstein	22	18
Martin Luther King	24	31
Helen Keller	9	15

Figure 6. A frequency bar graph of the number of adults and children who named various personalities as people they would like to have known.

The graph in Figure 6 shows the number of respondents who chose various celebrities as the person they would most like to have known. This is a *frequency* bar graph. In this type of graph, a bar’s height is not the value of an individual case, but rather the number (frequency) of cases (respondents) that all have a particular value. In this instance, the left-most bar shows that 16 respondents (cases) selected Susan B. Anthony.

We can see the difference between these two types of graphs by comparing Tim’s plot in Figure 7 to Kenny’s plot in Figure 8 (see case 19). Tim made what is sometimes referred to as a “line plot.” He represented family size along the horizontal axis, and each X stands for one family. The height of the column of X’s above a particular location shows the number of families of that size. Thus this is a frequency graph, and if Tim were to replace the stacks of X’s with bars, he would have a frequency bar graph much like the celebrity graph in Figure 6.

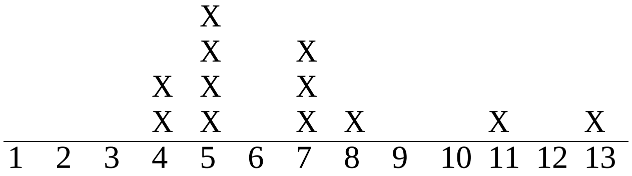


Figure 7. Tim’s line plot showing the frequency of different family sizes. The numbers along the horizontal axis are the various family sizes; the 3 X’s over the 7 show 3 families of size 7 in his sample.

Kenny, on the other hand, represented each individual family in his sample along the horizontal axis (see Figure 8). The height of the column of X’s above each family (case) shows the size of that family. Family number 1, for example, has 12 individual members. This is a case value plot, and if Kenny replaced stacks of X’s with bars, his graph would look much like the costume graph in Figure 5. Kenny’s plot provides ready information about the relative sizes of various families, whereas Tim’s plot provides ready information about how frequently various family sizes occur.

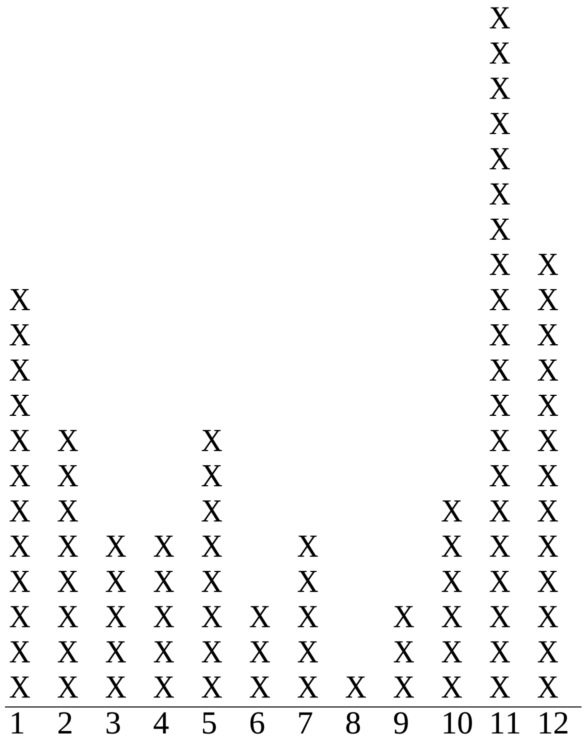


Figure 8. Kenny's case value plot showing the size of 12 families. Each number on the horizontal axis is a family and probably indicates the order in which Kenny obtained the data; the 5 X's over the 7 show that family #7 had 5 members.

As students begin attending to frequency as well as case value they sometimes struggle to distinguish between them. In case 18, one group of students used pairs of bars to separately represent the case value and that value's frequency (see Figure 9). According to the students, the first stack on the right represented 2 people (frequency) and the next stack of 2, moving left, represented the number of teeth lost by those students (value). The third stack from the right represented 3 people (frequency), and the fourth stack represented the number of teeth (value) lost by those three people (see line 544, pg. 93). Another group of students created the frequency graph shown in Figure 4 above (case 18). In doing so, they struggled with how to coordinate values and frequencies. One of them exclaimed, "Oh, now I get it." when he realized that the numbers along one axis could represent the number of teeth lost (value) while the number of faces drawn along the other axis could represent the number of people who lost that many teeth (frequency).

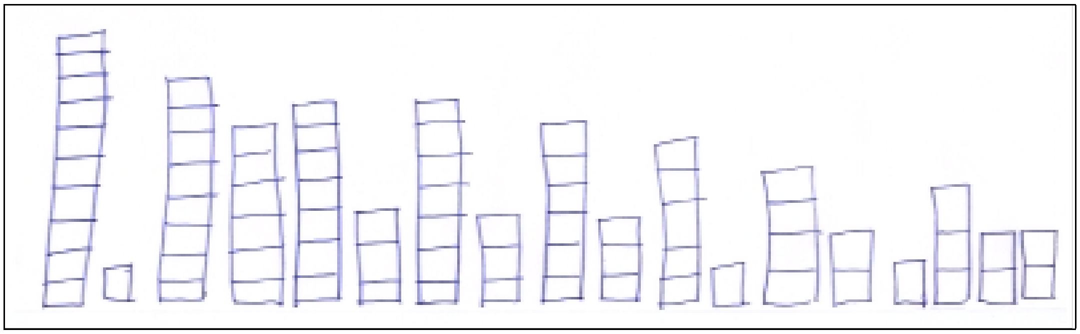


Figure 9. A student representation in which number of people and number of teeth are represented by separate stacks of cubes.

We have stressed that in deciding how to organize and represent data in graphs and tables, it is critical to consider what information will address your question. These considerations tend to get backgrounded, however, if students are focused primarily on applying learned conventions. Roth and McGinn (1997) pointed out how "In schools… students make graphs for the purpose of making graphs" (p. 95). Students are well practiced, therefore, in setting aside their own intentions and purposes to get down to the business of producing "good graphs."

Deciding on plot scales and on what data they should include in their plots poses a number of interesting challenges to students. Many students maintain that plot scales should not extend beyond the range of observed values, while others argue that the scales should extend to include values that could have occurred or far enough, at least, to make a pleasant boundary.

In case 4, two fourth graders considered what scale to use in plotting data on family size. Their scale initially went from 4 to 18, the actual range of the data. The teacher prompted them to consider what range they might need if they collected more data. One student replied:

> I think 2 could happen. You could have one kid and just one parent. Maybe their father died or something. (p. 16)

After discussing what made sense in this instance, the students redrew their scale to allow for the possibility of smaller families. Of course, there is no single correct scale for their plot. What is significant in this episode is that the students came to perceive the scale range as a choice that hinged on the particular data and on their question.

In later grades, students confront additional scaling decision: Should they group numeric scores into larger interval sizes (e.g., show the frequency of all values 0-4, 5-10 … with one bar each)? How big should they make the x and y axes relative to each other (e.g., should the bars in a frequency graph be tall and skinny or short and fat)? Choices about how to scale data affect how the data appear. As students gain more experience with scaling decisions, they come to see that there is no ideal scale that will make the data appear as they “really” are. Thus it is best to try out several alternative plots and scales and learn what we can from each. When it comes time to summarize results for others, we select those representations that do the best job of telling the story sharply and fairly.

Theme 4. Representing Non-occurring Events and Data Values of Zero

As in other areas of mathematics, zero poses special challenges to students. Regarding zero as synonymous with "nothing," some students argue against including in their representations either values of zero or non-occurring (zero-frequency) events. They eventually come to regard zeros as any other quantity and understand that whether to include them or not depends on what they want to know.

A community planner needs to know the size and composition of families in a certain neighborhood. Looking at census information, she notes that many families have no children. These 0-children families are part of the information she needs, no less important than 6-children families.

Students hold strong and differing views about whether to include data values of zeros in their data plots and summaries. de Lange, Burrill, Romberg, and van Reeuwijk (1993) describe high-school students struggling with just this context "Do we have to include these [0-children] families or not?" (p 66-67). In case 20, Nadia's fifth graders recorded the number of instruments each student in the class played. In summarizing the data, some students suggested that the number of instruments ranged from 0 to 3. Rick disagreed.

> Rick: But 0 shouldn't be there! If they play no instruments they shouldn't be there. This is only for finding out how many play instruments and if you don't you shouldn't be there.
>
> James: I disagree …I think that the graph has to show the zero children because then it is like we only have 15 children in the class. We interviewed all 19 and if we get rid of 4 of them we don't show the whole class. (p. 101)

The above episode focused on values of zero. Students also wonder whether to represent possible outcomes that never occurred (occurred with 0 frequency). In case 9, third-grader Crissy asked her classmates to name their favorite kind of math. She recorded responses by putting a corresponding math symbol (+, x, ÷) next to each student's name. Looking at her final results, she noted:

> No one even picked subtraction. I could just write the subtraction sign at the end. I could leave it out, leave it blank because nobody likes it best. (p. 37)

Whether or not to represent frequencies or values of zeros depends, of course, on the particular questions being investigated. Do Nadia's students in case 20 want to look at the distribution of instrument playing in the whole class, or do they want to investigate whether those students who do play instruments play more than one?

With experience, students learn that omitting or including frequencies of zero can drastically affect how they perceive those displays. For example, Maura's students in case 17 plotted the number of years each of their families had lived in town. The plot (see Figure 10) shows a few families living there over 16 years — 37 years in one case. More significantly, it shows two distinct clusters, with one group of families having lived in town between 0 and 6 years, and the other between 10 and 14 years.

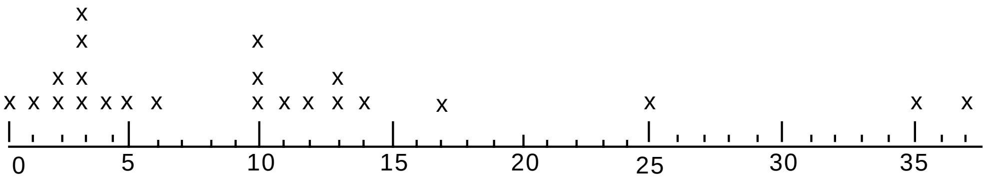


Figure 10. Line plot of the number of years students' families have lived in their town. The continuous scale highlights gaps between data values.

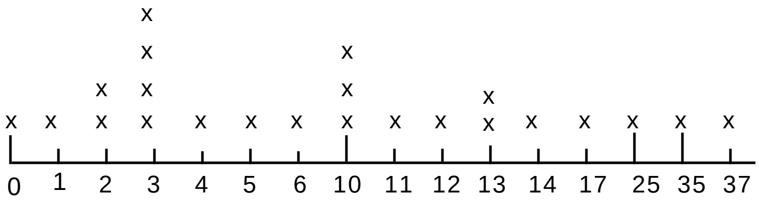


Figure 11. The same data displayed Figure 10 now plotted on an axis that eliminates non-occurring values. The gaps in the data are harder to see.

Seeing these two separate clumps may raise interesting questions about factors affected town growth. Indeed, an economic recession had hit the area beginning about 1989, which more recently had abated. Note how the plot appears in Figure 11 when we omit non-occurring values. Extreme values above 16 no longer stand out, and the two clusters are hard to see.

Theme 5. Viewing Data as an Aggregate

Younger students tend to view data as a collection of unique individuals. As the questions they explore change, they come to see data as an aggregate, a distribution of values with emergent features, such as center, spread, and shape, which are not evident in any of the individual cases.

David Moore (1990) has suggested that there are five core ideas of statistics. Topping his list is the awareness that variation is everywhere. "Individuals are variable; repeated measurements on the same individual are variable" (p. 135). The idea that individuals vary is apparent even to young students. "Just by looking around the room," students in a combined third and fourth grade class could see that their heights varied and that not every fourth grader was taller than every third grader (case 32, p. 165). Their classmates come in a variety of heights, hair colors, and temperaments. Their local weather varies not only from season to season, but day to day, and sometimes from one minute to the next. If students know nothing else when they begin collecting data, they know that they'll get a variety of values.

Variability among individuals is obvious to students. What is not so obvious is how to quantify variability in a group or to perceive and characterize the group as a whole when individuals in that group are so different from one another.

In early experiences with data, students tend to focus on describing individual data points, or clusters of similar individuals. Case 6 described an activity in which kindergarten students reported their favorite color. As the teacher recorded the information on the board, students spontaneously commented on which color was ahead — the modal value. However, the next day when the teacher asked, "What does this graph tell us?", they replied:

> We know what everyone's favorite color is.
>
> My favorite color is red.
>
> We learned English and Chinese colors.
>
> My shirt is blue. (p. 28)

The teacher wondered why it was so obvious to her from the graph that blue was the favorite color and why her students "did not seem to pull the individual pieces of information together to share ideas about the data as a whole" (p. 28).

When looking at data, most elementary students attend to characteristics of individuals — they tend to see the trees rather than the forest. Students make a conceptual leap when they switch from seeing data as an amalgam of unique individuals to seeing them as an aggregate, a group with emergent properties.

These emergent properties may not be evident in any individual member. For example, examine the frequency distribution of the bedtimes of a sample of third and fourth graders described in case 24 (Figure 12). The distribution is mound shaped, with lots of bedtimes at or near 9:00. Moving away from 9:00 in either direction, we tend to find fewer and fewer bedtimes. This mounded shape is a characteristic of the bedtimes of the *group* of students as a whole and not of any of the individual bedtimes that make up the group. You could never guess the *shape* of this distribution by knowing the bedtime of a single student.

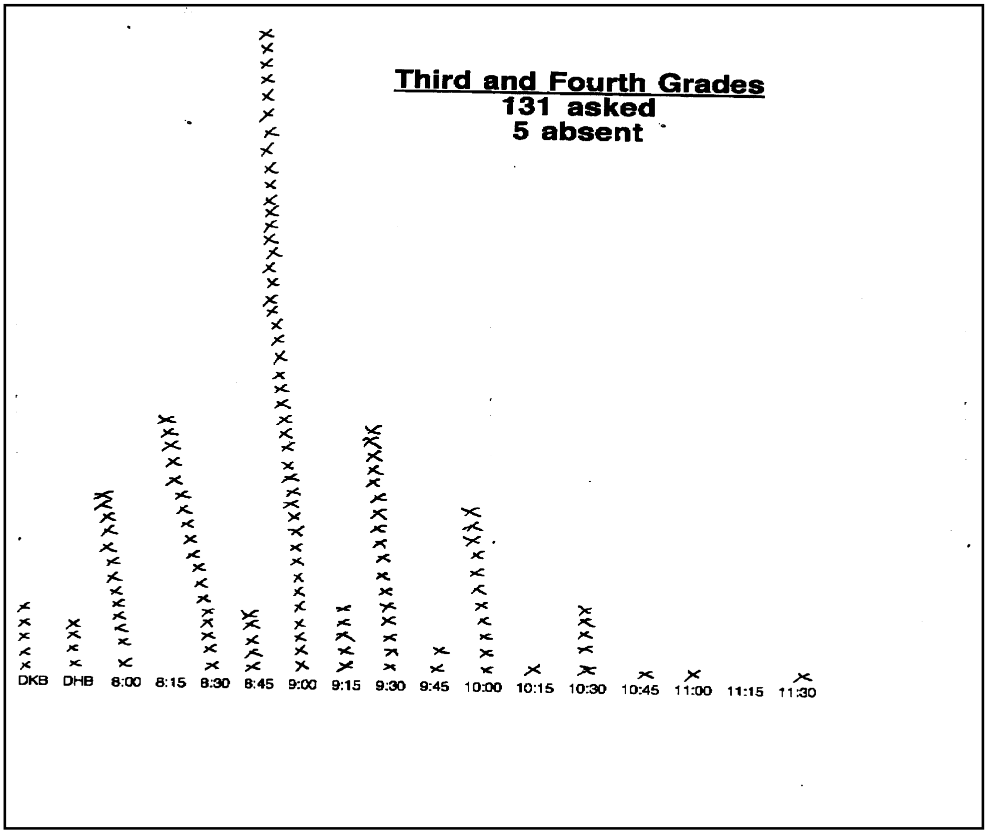


Figure 12. A line plot of third and fourth graders' bedtimes.

We can characterize a number of other features of this distribution: the range, the center of the distribution, that it is a bit stretched out (skewed) to the right, that bedtimes on the hour and half hour are more common than on the quarter hours. Depending on the question we have about children's bedtimes, one or more of these group features could provide important insights.

The challenge for students in reasoning about the aggregate is evident in the report of Hancock, Kaput and Goldsmith (1992). They worked in an after-school setting with small groups of students of ages 8-15 to design statistical questions, collect data, and explore those data using the software tool Tabletop. Tabletop displays each individual data point on the screen and allows students to attach identifying labels to these data points. Students consistently wanted to keep a label showing the identity of each data point (e.g., the names of individual students). Worried that this was preventing students from noticing features of the group, Hancock and colleagues had students remove these labels. However, students showed an uncanny ability to remember which data point belonged to which individual and continued to draw on this information in interpreting the data. The teacher in case 6 observes a similar pattern in her Kindergarten students who “seemed to attend to names on the chart and the information that was recorded about each person” (p. 28).

Hancock and colleagues (1992) reported that although they explicitly encouraged students to use distributional terms such as “cluster” and “range” to characterize data, “students often focused on individual cases and sometimes had difficulty looking beyond the particulars of a single case to a generalized picture of the group” (p. 354). The researchers characterized this individual-based analysis as resulting in blow-by-blow descriptions of results: “This person said ‘yes’ to Question 1 and to Question 2, but this person said ‘yes’ for Question 1 but she didn’t say ‘yes’ for Question 2…” (p. 354). We see similar responses throughout the cases, especially in the earlier grades. For example, when asked what they learned from their survey about who liked their vacation, one kindergartner simply replied: “11 people said yes, 2 people said no and 8 people said something else. That makes 21” (case 12, p. 55).

When students focus exclusively on information about individuals, they are unlikely to notice characteristics of the group as a whole. However, coming to see data as an aggregate is most probably prompted in the first place by the questions students ask about the group. Questions that draw attention to aggregate features include, “What is our class’ favorite color?” and “How do our heights compare to the heights of fourth graders?” Until students ask these sorts of questions, they have no particular reason to think about data as an aggregate or to describe group features such as center and spread.

Theme 6: Summarizing Data with Averages

Young students bring to the classroom a rich set of intuitions and expectations about typicality. Very young students rely on the mode almost exclusively as a descriptor of center. In later years, students use a wider variety of averages to summarize data. They prefer averages that are actual data values, lie in the center of the range, and are close to the modal value. Using as an average a range of values in the center of the distribution often allows them to satisfy all these criteria. Learning to use the median and mean as meaningful indicators of center is a challenge even for older students.

There are many types of averages: mean, median and mode are commonly encountered ones. But there are many other ways to characterize the center of distributions. For example, economists sometimes describe average growth rates with “geometric” means.

By second or third grade most children have heard the word *average.* Their ideas about average are based on everyday meanings that draw on qualitative, rather than quantitative, notions of typicality. For example, in case 25 the teacher recounts a discussion about averages that occurred in her second grade classroom. One student described average as “not the best, not that great, but okay.” Other students offered similar notions, describing average as “normal,” “regular,” or “what most people are.” Students develop more quantitative notions of average as they begin to use them to describe and compare sets of data.

Judging from the cases, the ideal average that many students have in mind as they reason about numerical data is an actual value in the data set which is also the most frequently occurring (the mode), positioned midway between the two extremes both in terms of value (the midrange) and order (the median), and not too far from all the other values (see case 31, in particular). In symmetric, mound-shaped distributions with lots of data, one can often find a value that has all, or nearly all, of these properties. But with many of the small data sets students explore, all of these conditions are seldom satisfied. When students have to start giving up criteria of ideal averages, most of them hang tenaciously onto average as mode. Thus, teachers in the casebook describe their students as heading “straight for the mode” and as considering it “the end-all way to describe what’s typical in a set of data” (p. 75; p. 112).

In their research of student understandings of averages, Mokros and Russell (1995) found that most students in grade 4, and even a few in grades 6 and 8, used the mode in situations where other indicators of center would be more representative. For example, the researchers asked students to argue for a certain allowance based on a graph (p. 24) that showed a modal allowance of $2.00 and an arithmetic average (mean) of $3.27 (the distribution is skewed towards higher values). Most students, however, considered only the modal value of $2.00 in making their argument. In the words of Mokros and Russell,

> Even when there was strong motivation to see the higher numbers as more representative (e.g., they could help one argue for a higher allowance), they did not make an argument based on representativeness. According to these students, $2.00 was the only number that mattered—at least mathematically—in the distribution. (p. 28).

The mean (arithmetic average) is strikingly distinct from the ideal average students in the casebook seem to envision. We get some insights as to why from the research of Strauss and Bichler (1988). As part of their study of student understandings of the mean, they described seven fundamental properties of the mean. Of those properties, the only one that is clearly among those students regard as important in an average is that the mean is located between (though not necessarily midway between) the extreme values. Two of the properties — that it is not necessary for the mean to a) equal one of the values in the data set, and b) have any counterpart in physical reality — are, in fact, reasons students give for dismissing a mean, such as 2.3 children per family, as a useful average.

Students are quite clear in voicing their objections. In case 30, third grade students tested how far they could blow a Styrofoam cylinder. One student objected to using the mean of his two attempts as a measure of his performance, pointing out: “I didn’t get that [the mean value] as one of my distances. It wouldn’t be true. It’s a lie!” (p. 158). In determining the average length of names of students in their fifth grade class, one group of students computed a mean of 13.2 then rounded it down to 13 “because we can’t have a point something [fractional part]” (case 33, p. 171).

Case 33 illustrates the problems that can arise when students try to use the mean without understanding what it represents. The students’ task was to determine the typical name length of students in their fifth grade class. Kara and her partner decided they would try to compute an average based on the number of letters in the names of students in the class list the teacher had passed out. Kara offered: “I think we need to add something and then, what, multiply something?” (case 33, p. 172-173). After they finally remembered that it was add and *divide*, Kara summed the values and then divided not by the total number of values (20), but by the number of unique values (8). This gave her an average that was nearly twice the length of the longest student name. But she registered no alarm. The mean to her seemed to be simply the result of a computation; it did not need to make sense.

Although the add-and-divide algorithm is relatively simple to execute, research indicates that many students who are familiar with the algorithm have not developed the conceptual underpinnings that allow them to meaningfully interpret or apply the mean (Gal, Rothschild, & Wagner, 1990; Mokros & Russell, 1995; Watson & Moritz, 1999). This finding is not limited to elementary grade students; similar results have been documented with high school and college students (Cai, 1998; Pollatsek, Lima, & Well, 1981).

Researchers who have explored students’ use and understanding of means are generally recommending that we place less emphasis in the elementary grades on teaching the mean (Mokros & Russell, 1995). In recent elementary curricula, use of the median seems to have replaced the mean as an objective of early statistics instruction. However, pushing students to compute and use the median before they have a sense for why and when it might be useful risks promoting mechanism over meaning. As one teacher observed, “I had focused so much of their prior work on finding the median that they were no longer looking at the whole picture” (case 28, p. 142).

An alternative notion of average that many students spontaneously adopt is what one third grader called the “middle clump” (case 26). Fourth grade students in case 4 made a line plot display of the number of people in their families. Below the plot they wrote this wonderful summary, which included descriptions of spread, center, and a value of special interest:

> One person has 18 in her family. The range of the data: 4-18. Most typical number of people in the family is 5 or 6. (p. 18)

As in this case, a middle clump is typically a cluster of values in the heart of the distribution that includes all, or most, of the ideal features of averages listed earlier. The clump of 5-6 in the distribution of family sizes includes the mode, the median, and is near most of the data: two-thirds of the cases lie in the interval 4-7. In describing a distribution, statisticians often specify values for both center and spread. They might summarize this distribution of family size by saying that its median is 6 and the middle 50% of the data (the interquartile range) is between 5 and 9. The middle clump potentially serves a somewhat similar purpose for students, letting them express at the same time what’s average and how spread out the data are.

Students’ Interpretations of Average

To explore further how students think about and use averages, it helps to distinguish between the types of averages students use (modes, medians, midranges, etc.) and the meanings they give to those averages. Konold and Pollatsek (1999) suggested several possible ways to interpret an average, including average as a data reducer, as a fair share, and as a typical score. While all of these interpretations are useful in certain contexts, some interpretations are more conducive than others to viewing an average as representative of a group of data.

An average is interpreted as a *fair share* when we imagine redistributing a quantity among individuals so that in the end each has the same amount. The fair share interpretation appears to be what Trudy in case 31 was originally considering as an average of the 4 heights in their group. She described getting an average by taking inches off the taller heights and adding them to the shorter heights, “because then you could even all the heights out.” She opted not to do this and instead to use the add-and-divide algorithm “to just make it simple” (p. 163). Most students are introduced to averages in contexts where a total is evenly redistributed. Konold and Pollastek (1999) pointed out, however, that from this perspective an average is not necessarily viewed as representative of the set of original values.

Average interpreted as a *typical* score includes ideas related to the majority, mode, median, and midrange. Teachers in the cases often pose questions to students hoping to illicit this interpretation of a “typical” or “representative” value: The teacher posing a question such as “How tall is a typical fourth grader?” is presumably thinking of a value that is representative of the entire group.

However, many of the students’ responses suggest that they believe a typical value is a characteristic of a *particular* case, or set of cases, in the distribution. The teachers in case 27 asked her third graders, “What would you say is the average height of kids in our room?” One student volunteered, “It’s me. I think I am average” (p. 134). She did not seem to be focused on an average as a characteristic of the group but rather on a characteristic of a person: “I’m average.” Other students in the class gave similar responses: “Sam is average, and I’m average too” (p. 135). To claim that Sam is of “typical” or “average” height is to characterize him, not necessarily the group as a whole. We do not know whether these students would consider Sam's height to be a good characterization of the whole group. The use of averages to describe particular individuals rather than the group is supported by common usage where we frequently speak of the “average” or “typical” student.

Mokros and Russell (1995) described some students as using a “reasonable” approach in arriving at representative values for averages. These students drew on both everyday experience and informal judgments of where the data seemed centered to come up an average that made sense to them. For example, a fourth grader’s real-life and mathematical sense comes through in her explanation of the distribution of allowances she constructed to reflect an average of $1.50.

> Well, just as they get higher, sometimes they should get lower. And you said the typical allowance is about $1.50, so some kids can get $1.50. And if it were $1.75 that would be pretty close and so would [$1.25], because that’s around it…If the typical [allowance] is $1.50, you’re not going to really go above $5.00 for any kid. If I got $5.00, it would be good…And you know that when you run around with a lot of kids, most of them are like $1.50 or $1.75 or $1.25 or $1.00, something like that. (p. 30)

Students who used this approach relied on intuitions that averages are roughly in the center. They often treated an average not as a precise location but as an around-about sort of thing.

In analyzing the written reflections in notebooks kept by her third graders, Suzanne (case 27) offered an analysis that fits remarkably well with averages as intuitive estimates, or round-abouts.

> I felt more certain than ever that the beginning of understanding of average starts with this sense that it “feel right” that Will had mentioned. I noticed that “feeling right” seemed to be associated

> with a tendency towards the center of the data. When children were pressed to explain why they had chosen a particular value as an average, they began to analyze the data to look for reasons, and they sometimes sounded like they were talking about traditional methods for finding average: median, mode, and sometimes, midpoint of the range. However, that's not where they started. Rather they started with a general idea that the average is typical and in the center of the data set. (p. 137)

Assuming that learning to use averages meaningfully requires integrating formal approaches with these more intuitive ones, many researchers have stressed that we should encourage students to draw on their intuitions and informal methods of summarizing data, and that in many situations what students come up with as descriptors of average are perfectly adequate summaries (Bakker, 1999; Cobb, 1999; Mokros & Russell, 1995). Noss, Pozzi, and Hoyles (1999) reported the use of informal notions of averages among practicing nurses. When the nurses they studied wanted to find a baseline systolic blood pressure for an individual across time, they did not compute a mean or median. Among the methods used was to visualize an imaginary line roughly in the middle charted data. One nurse explained, "When I'm talking with another member of staff or a doctor I'd say we'd be talking about averages in terms of what's the middle line" (p. 15). It would be unnecessary and a waste of time for nurses to compute exact averages when monitoring such vital information on a minute to minute basis.

As we discussed earlier, an interpretation of average which appears to fit well with students' informal ideas of average is "middle clump." Students in Maura's combined third and fourth grade class (case 17) used a clump to summarize the number of years students' families had lived in town. Fighting her desire to point them toward the median, Maura let them proceed. They summarized the data with the statement that "almost half" of the values were between 0 and 6. Maura reflected:

> That first big clump clearly needed to be part of it [the summary] in the kids' eyes, and the fact that it also contained that mode at 3 didn't hurt either.... It seemed to carry some weight of significance, and as I thought about it, I realized that it did for me also. This was a meaningful statement to make about our data.... (p. 78-79)

Cobb (1999) described how the seventh graders in their teaching experiment began reasoning about data sets as wholes once they were able to perceive and talk about the "hills" in the line plots they were examining. Clusters, hills, or middle clumps, may not only serve as descriptors that are often good enough for the task at hand; they may also give students experience working with ideas that will help them eventually construct meaningful interpretations of measures of typicality such as means and medians.

Theme 7. Comparing Groups

Questions about if and how two groups differ motivate students to look at distributions in different ways, focusing on features of the group as a whole rather than individuals in the group. However, many students who seem to readily use averages to describe a single group do not use them to compare groups. Using an average of one group to compare it to another requires viewing that average as representative of the whole group.

Throughout chapter 5 of the casebook we find evidence that although comparison tasks are challenging to students, these tasks seem to motivate students to begin focusing on the data as aggregate. Reflecting on their project comparing bedtimes across several grade levels, the teacher was pleased that her third grade students had

> arrived at a place where they were able to describe the shape of the data and to look at the features that it made most sense to examine. And yet I'm not sure exactly how they got there. This was the first time that I was asking the class to compare data sets. Is there something in setting up a comparison task that makes it inherently more interesting? Do more features jump out at you when you're comparing because the presence of a feature on one graph shows up the absence of that same feature on another graph? (case 24, p. 119)

As to why comparison problems pose a challenge to many students, the teacher in case 32 noted that comparing groups can be

> a little puzzling to kids at this age [grades 3 and 4] — how can you talk about the group, after all, as something somewhat separate from the individuals in the group? (p. 165)

As long as students are working with single groups, it is not clear why they would need to summarize the data with something like averages. But when faced with the problem of comparing groups, averages potentially become useful. However, research has demonstrated that students do not initially view averages as useful tools for comparing groups, presumably because they have not yet adopted a view of data as an aggregate and therefore of averages as ways to characterize groups. This includes students who appear to know how to compute means (Hancock, Kaput & Goldsmith, 1992; Watson & Moritz, 1999; Jones, Thornton, Langrall, Mooney, Perry, & Putt, 1999).

Gal, Rothschild and Wagner (1990) gave students in grades 3, 6, and 9 several pairs of line plots that portrayed data from one of two contexts. In one cover story, the plots showed the results of a frog leaping contest between two teams, with x's on the graphs representing the distances jumped by individual frogs of each team. The students' task was to use the data to decide which team won the contest. Only half of the sixth and ninth grade students who knew how to compute means went on to use means to compare the two groups.

This finding is not limited to the use of means. Bright and Friel (1998) questioned eighth grade students about a stem-and leaf plot that showed the heights of 28 fifth-grade students. They then showed them a stem-and-leaf plot that included these data along with the heights of 23 basketball players. The plot is shown in Figure 13 (from Bright and Friel, 1998, p. 81). Heights of basketball players were indicated as they are here in bold type.

10	
11	
12	
13	8 8 8 9
14	1 2 4 7 7 7
15	0 0 1 1 1 1 2 2 2 2 3 3 5 6 6 7 8
16	
17	1
18	**0 3 5**
19	**0 2 5 7 8 8**
20	**0 0 2 3 5 5 5 5 7**
21	**0 0 0 5**
22	**0**
23	

Figure 13. Heights of students and basketball players (**bold**). The row headed by 13 (the stem) contains four cases (leaves), three students each of 138 centimeters, and a fourth student of 139 centimeters.

Asked about the “typical height” in the single distribution of the fifth grade students, two of four students who were interviewed specified a middle clump (e.g., 147-151 cm). But shown the plot with both distributions, these students could not generalize their method to determine “How much taller are basketball players than students?” The students who did make comparisons, compared selected individuals from each group (e.g., pointed out that the tallest student was shorter than the shortest basketball player). In the words of Bright and Friel (1998, p. 80), some of these students could

> describe a ‘typical’ student or basketball player, but they did not make the inference that the ‘typical difference’ in heights could be represented by the ‘difference in typicals.’

Konold, Pollatsek, Well, and Gagnon (1997) reported similar results from their study of high school seniors who had just completed a yearlong course in probability and statistics. During the course, the students had frequently used medians and means to compare groups. But during a post-course interview, where they were free to use whatever methods they chose, they seldom used medians, means, or percentages when comparing two groups. These students made most of their decisions about group difference by comparing the numbers of individuals in each group within narrow slices of the range (e.g., “More of the A students have curfews than don’t. Therefore, students with curfews get better grades than those without curfews.”).

Similarly, in case 32 we see students using slices of the data to compare groups. One student argued that the fourth graders were taller than the third graders because “the third grade line plot has 5 x’s [at 51]. The fourth grade has none at [51]” (p. 165). While this method of comparison gives some useful information when the groups are of equal size, it can be quite misleading when the groups are of different sizes (e.g., suppose in the curfew example that there were twice as many students overall with curfews as without them).

Cobb (1999) and colleagues reported similar findings from their middle school teaching experiment. They had originally designed their curricula to support students using medians to compare groups. However, they found that the students rarely used medians for this purpose but rather tended to compare slices across the groups as described above. Cobb (1999) described a critical episode during an investigation of traffic speed before and after a police speed trap. During a group discussion, one student compared center “hills” of the two distributions to argue that the speed trap successfully slowed traffic:

> If you look at the graphs and look at them like hills, then for the before group, the speeds are spread out and more than 55, and if you look at the after graph, then more people are bunched up close to the speed limit [50 mph], which means that the majority of the people slowed down close to the speed limit. (p. 19)

This was the first occasion during class discussion that a student had “described a data set in global, qualitative terms by referring to its shape” (p. 19). Other students adopted this terminology, and comparison of “hills” became a standard way to describe and compare groups. As they progressed to comparing data sets of different sizes, they began talking not just about hill location, but about the number of cases in the hills relative to group size. We see similar forms of reasoning in Georgia’s class (case 24) in which students used middle clumps (or modes) to describe an increasing trend in bedtimes across the grades.

Theme 8. Relating Data Back to the Real Situation

During all phases of data analysis it is critical that students not lose sight of the questions they are pursuing and of the real world events from which the data come. These connections are easier to maintain when students work with data from familiar contexts and use representations they understand.

In theme 2 we stressed the importance in data analysis of seeing data as related to, but not identical with, real events — as models of those events. It is equally critical that, once they have organized and represented data, they interpret the data by relating them back to the real-world observations and the questions that motivated the investigation in the first place.

Cobb (1999) reported that the seventh grade students they interviewed before instruction often viewed working with data as “doing something with the numbers” (p. 12). In summarizing student responses, Cobb concluded that it is “doubtful whether most of the students were actually analyzing data, in that the numbers they manipulated did not appear to signify measures of attributes of a situation about which a decision was to be made” (p. 13). Early in the subsequent teaching experiment the researchers saw the same tendency as students began reasoning about a set of data concerning how long two brands of batteries lasted. Students were interpreting the graph in Figure 14 showing hours of use of “Always Ready” and “Tough Cell” batteries.

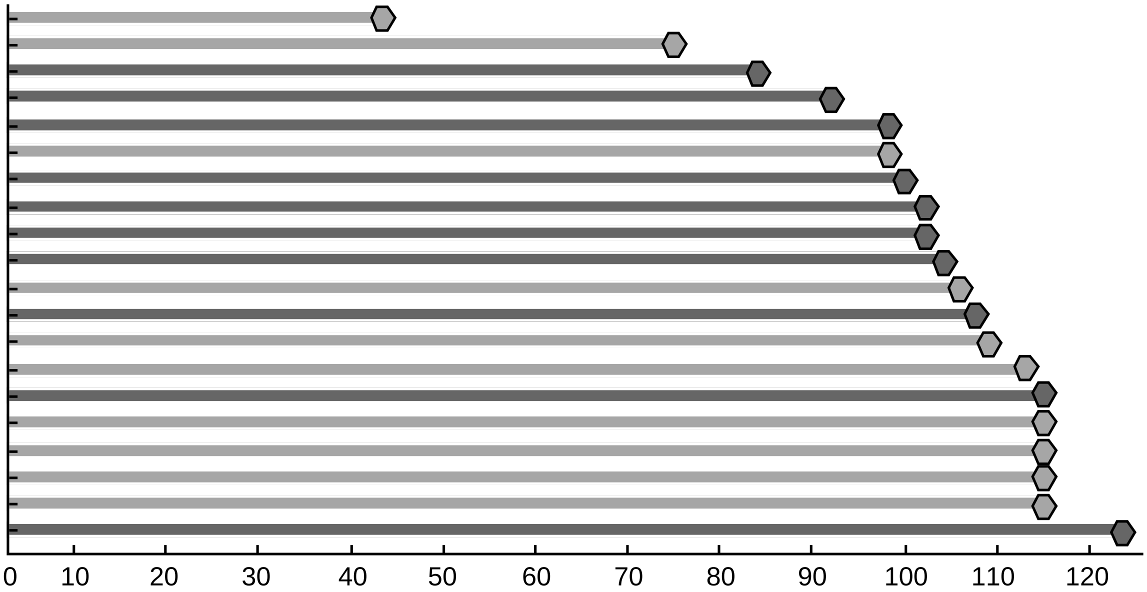


Figure 14. Case-value plot of hours of use of “Always Ready” (light gray) and “Tough Cell” (dark gray) batteries; adapted from Cobb (1999).

The two brands of batteries appeared as green and pink bars on the computer screen the students were viewing. In Figure 14, as well as in the class dialogue below, we have changed these colors to light and dark gray, respectively. During the first day working with this display, the students’ referred mostly to numbers and colors. Noticing this, the teacher began to encourage them to talk instead about *batteries* .

Casey: And I was saying, see like there’s 7 [light gray] that last longer.

Teacher: OK, the [light gray] are the Always Ready, so let’s make sure we keep up with which is which. OK?

Casey: OK, the Always Ready are more consistent with the 7 right there, and then 7 of the Tough ones are like further back, I was just saying ‘cause like 7 out of ten of the [light gray] were the longest, and like…

Ken: Good point.

Janice: I understand.

Teacher: You understand? OK, Janice, I’m not sure I do, so could you say it for me?

Janice: She’s saying that out of 10 of the batteries that lasted the longest, 7 of them are [light gray], and that’s the

> most number, so the Always Ready batteries are better because more of those batteries lasted longer. (p. 14-15)

Cobb (1999) concluded that a critical step in these students learning to reason about data was coming to expect that the statements and claims regarding various plots should extend beyond mere numbers by making reference to a specific real-world situation.

In most of the activities described in the cases, students were collecting their own data. Yet we still frequently see students talking about numbers only. This problem recurs as students begin learning how to describe general features of the data, when they can again lose sight of what those general features tell them about the real situation. For example, third grade students wrote summaries describing a line plot of daily temperatures they had collected in February (case 28). One student wrote:

> At first very spread out. Then it gets more bunched up. (p. 144)

Many of the summaries, like this one, did not connect the data to the context. Concerning the student who wrote this summary, the teacher wondered:

> Did he know it wasn't just a clump of x's, but a representation of a real thing, which was indicating a predominance of a certain temperature on the high side of the range of temperatures for the month? … I wondered how to help him see that what he noticed about how the data *looked* implied something significant about what the temperature was like in February. (p. 144)

Feldman, Konold, and Coulter (2000) cited several examples from various data-intensive science projects of what happens when students are given data about phenomena far removed from their experience and with no clear questions in mind. They concluded that:

> Nearly every problem associated with…keeping them engaged in analysis ultimately stems from students not making, or losing, the connection between the data they have and a real-world question. This being the case, the solution to most of the problems can be found in focusing on how to make and maintain these connections. (p. 127)