DATA IN SEARCH OF A CONTEXT: AN ICEBREAKER ACTIVITY

Alexander White, Layla Guyot and Amanda Walker
Texas State University, San Marcos, TX
whiteale@txstate.edu

Context is a key component to statistical reasoning (Cobb & Moore, 1997) and influences data collection, analysis and interpretation. However, students may be challenged by the interplay between data analysis and the context. In our introductory statistics courses, we use an icebreaker activity to give students a chance to introduce themselves and explore data with an unknown context. As the context is slowly revealed, we find that students learn about what information is provided by the data separate from the context. We present how students engaged in the activity, by creating graphs of their own and identifying features of data representations that help them reveal the context. We discuss how concepts introduced in the icebreaker activity can be developed in later classes.

BACKGROUND

Context is a key component to statistical reasoning. As opposed to mathematics where context may obscure structure, during data analysis we hope context will provide meaning (Cobb & Moore, 1997). The Guidelines for Assessment and Instruction in Statistics Education recommend that students learn to follow a statistical process consisting of formulating the question, collecting data, then analyzing the data and interpreting the results with the question in mind (Aliaga et al., 2005; Franklin et al., 2007). The question and its context drive the process and influence decisions about how to collect, summarize and analyze the data. However, in the authors' experience students may have difficulty managing the interplay between data analysis and the context. In some cases, students presuppose the answer to a question before analyzing the data or ignore the results of the data when it conflicts with the desired or intuitive answer. According to Konold and Higgins (2003, p. 195), students should learn to see data as separate in some ways from the real-world event they observed. They must organize the information in a way that provides insight into the question. Then based on how the data was collected and the properties it exhibits, students need to learn what the data tell them about the original statistical question.

We present an icebreaker activity that we have used as an introduction to statistics at the secondary and post-secondary levels but could be adapted for lower grades as well. In our classes, we use this activity on the first day of class to give students a chance to learn each other's names and ease their way into the study of statistics. To highlight the importance of context, in the activity students are presented data with an unknown context. They graph the data and explore what clues the data may give about the context it came from. As the context is slowly revealed, students learn about what information is provided by the data and what is provided by the context.

ICEBREAKER ACTIVITY

Below we present the different steps of the activity from the set up, data collection, graphing the data, matching the data and revealing the context.

Set up

Prior to the activity, the instructor must prepare sheets of paper, each one with a question printed with a large font size. The questions are taped to the back of the students. In that way, students cannot see the question taped to their back, but when they turn around their classmates can see and answer the question. To carry out the activity, students are provided with tape and large pieces of paper to graph their data.

Careful selection of questions can lead to significant discussions about the statistical properties of data. All questions should require a numerical response. To help the activity "break the ice", the questions should be interesting to students and reveal interesting pieces of information to help the students learn about each other. To make the follow up discussion richer, one should consider the statistical properties of the variables involved. Choosing variables with very different scales encourages students to think about the information the center and spread provide about a distribution.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

We have also found it helpful to include variables that are categorical but coded with numerical responses. When students collect and attempt to graph the data from the categorical variables, students will experience the difference in behavior exhibited by categorical and quantitative variables and realize why they require different statistical techniques. In Table 1, we list 10 examples of questions and the rationale for including them in the activity.

Table 1: Example Questions for Activity

Question	Rationale
What year was Leonard Da Vinci born?	Variation due to lack of knowledge
How old are you?	Small variation with possible outliers
How many cars does your immediate family own?	Discrete variable
How much did you spend on your last haircut?	Student interest, outliers
How many fingers and toes do you have?	Variation due to difference in interpretation
Do you have pets? 1 = Yes, 2 = No	Categorical Variable
What is the lowest temperature in degrees Fahrenheit you have ever experienced?	Allow for negative numbers
About how many miles is it from here to the farthest city you have ever visited?	Large variation with outliers
How many friends do you have on Instagram®? (answer 0 if you don't use Instagram)	Student interest
How many fish are in the ocean?	No true answer, scale issues

Data Collection

At the beginning of the activity, students are separated into groups of approximately 10 students. It may be a good idea to designate a few helpers that aid with management of the process and do not collect data. Each group forms a circle in a different part of the room with students facing each other. For discussion purposes, we consider a class with 2 groups of ten students. The instructor attaches a question unto the back of each student ensuring that the students do not see their question. Each group is given the same list of 10 questions. Students, one at time, introduce themselves to the group, then turn around. Other students in the group answer the question on the back of the student who turned around, while this student collects data on a sheet of paper. Only answers containing a numerical value are accepted. Numbers should be enunciated in a standard form (for example, read 1452 as one thousand four hundred and fifty-two) and without any units. Once each member of the group has collected the data, the instructor and helpers collect the questions from the back of the students and students return to their seat with their data.

Graphing the data

Once the students have returned to their seats, they are asked to graph their data. Students typically ask for more guidance, but we do not provide any. Eventually, we want students to choose graphs that are appropriate to the statistical properties of the variable and help satisfy the intended purpose. Although no particular type of graph is inherently better than another for univariate data, statistics textbooks typically present the same set of graphs: pie chart, bar chart, histogram, box plot, stem-and-leaf plot and dot plot (e.g., Starnes, Yates, & Moore, 2012). One goal of this activity is to see what kind of graphs students naturally choose. Research with younger students indicates that they view data as pointers to the more complex event of measuring the variable with a particular individual and show great reluctance to let go of any information attached to the event like the name of the respondent (Konold & Higgins, 2003).

We find that approximately 1/3 of our students include an index number indicating the original order of the response and create a scatter or time series plot (left graph in Figure 1). Another 1/3 of students present graphs familiar to them from past experiences like a pie or frequency bar chart (right graph in Figure 1). The final 1/3 is a diverse set of graphs. For example, one student produced a graph we might classify as an empirical distribution function. However, the orientation is nonstandard, and the relative frequency axis is not labeled. (Figure 2).

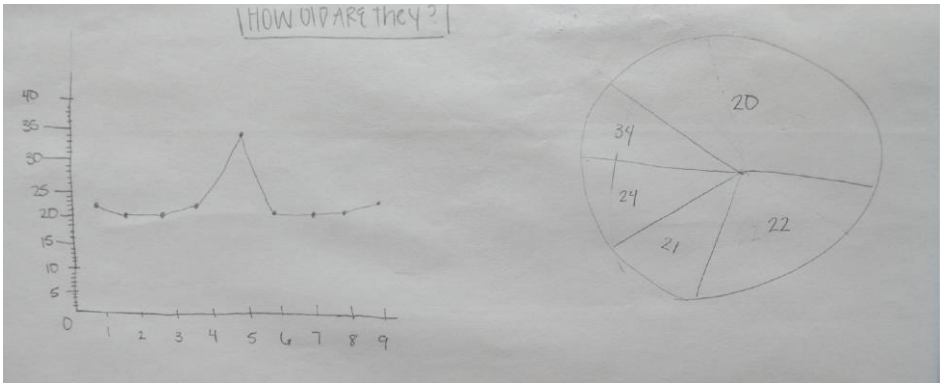

Figure 1: A student's graph for the question, "How old are you?"

Once the students have made a graph they are given the nearly impossible task: "Guess what the question on your back might be." To facilitate this task, we ask students to discuss what properties of the data and features of their graph help them determine what the question might be. Once all the students have made the first graph, we introduce the dot plot, and ask each student, if they haven't already, to make a dot plot of their data. If needed, this is a good time to have a class discussion of scale and note that the distance between tick marks may need to be greater than 1. In case of an extreme outlier, the data may need to have a broken scale.

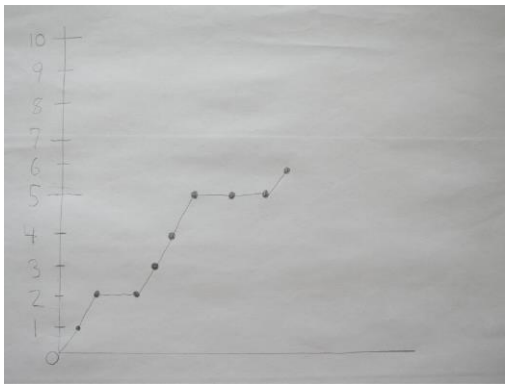

Figure 2: A student's graph for the question, "How many cars does your immediate family own?"

Matching the data

Once students have completed their graphs, one group of students places their graphs on a wall of the classroom. The second group is then asked to match their graphs to the ones on the wall. The class engages in a discussion about how, despite the fact that they do not know the question and that data on the wall is from an entirely different sample of students, it is still possible to match the graphs. In their discussion about creating useful representations of data, Konold and Higgins (2003) point out that we want students to move from concrete representations (John spent $10 on his last haircut) to abstract representations which use aggregates (50% of the students spent between $10 and $15 dollars on their last haircut). This activity provides the opportunity to discuss which aggregate features of the data can be used to match graphs.

Finally, the original questions are revealed, and the students match the question to its pair of graphs. The activity ends with a class discussion about what the distribution tells us about the students in the class as a whole. For example, what the distribution of the price of the last hair cut does tell us about the students in this class.

Follow up

Concepts that are touched on in the icebreaker activity can then be further developed in later classes. In the development of univariate graphical representations, aggregate features that were used

in matching the graphs of the two groups can be highlighted, shifting the focus from the specific values observed to the distribution of the whole data set. In a histogram, the original order, the actual data value, and the link between the data point and the individual are all lost. However, this loss of information allows us to focus on the data set as a whole made up of smaller aggregate chunks which can make informal inference easier (Rubin, Hammerman, & Konold, 2006).

Careful selection of the initial questions can lead to discussion about the causes of variation. As early as 6th grade, students are asked to understand that a statistical question anticipates variability in the data (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). In inference, understanding the source of that variation and possible bias is important in determining the appropriate statistical method of analysis to use and the possible valid interpretations of the results of that analysis. Important causes include measurement error, sample variability and response variability due to misunderstanding of the question. For example, with the question asking for the year that Leonardo Da Vinci was born, we expect there to be variety in the answers because it is unlikely that all the students know it is 1452 and hence will have to "guess". On the other hand, we expect the amount spent on the last hair cut to vary because of variation in the population. Another surprising example is the question about the number of fingers and toes that has provided interesting responses. When one of the authors first included this question he expected no variation, predicting students would all answer 20. However, that has not been the case. Each time we have used the icebreaker activity, at least one student has answered 18 (excluding the thumbs) or 10 (as in I have 10 fingers and I have 10 toes). This is a striking example of common occurrence in statistical surveys, and how respondents can interpret a question differently. When this happens, the results represent two or more samples depending on the interpretation making inference impossible. Or in the words of a second grader from a study by Russell, Schifter, and Bastable (2002, Case 6), "Everyone has to understand your question. If they don't understand your question, everyone will be answering any old way." The follow up discussion of this activity can use this example to illustrate why it is important to pilot test questions to ensure as much clarity in wording as possible.

CONCLUSION

The icebreaker activity presented in this paper is a fun way for statistics students to get to know one another while providing the opportunity to discuss several important issues in statistics. The setting allows students to reflect on what information is provided by the data, by the context and by aggregate notions of descriptive statistics including center and spread.

REFERENCES

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., . . . Utts, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College report*.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801-823.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*.

Konold, C., & Higgins, T. L. (2003). Reasoning About Data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*.

Rubin, A., Hammerman, J., & Konold, C. (2006). *Exploring informal inference with interactive visualization software.* In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS-7), Salvador Bahia, Brazil*.

Russell, S. J., Schifter, D., & Bastable, V. (2002). *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour.

Starnes, D. S., Yates, D., & Moore, D. S. (2012). *The Practice of Statistics* (4 ed.). New York: Macmillan.