DATA IN SEARCH OF A CONTEXT: AN ONLINE ICEBREAKER ACTIVITY

Alexander White and Layla Guyot
Texas State University, San Marcos, TX
The University of Texas at Austin
whiteale@txstate.edu

Because context plays a critical part in data collection, analysis, and interpretation, we offer an icebreaker activity during which students explore data with an unknown context. We previously developed this activity for small, in-person, introductory statistics courses (White et al., 2018), and we adapted the instructions and materials for online instruction. Surprisingly, we observed more robust conversations of critical concepts during the online implementation. We report how students engaged in the activity, by creating graphs of their own and identifying features of data representations that help them reveal the context. Finally, 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 follow a statistical process: formulate the question, collect data, analyze the data, and interpret the results with the question in mind (Aliaga et al., 2005; Franklin et al., 2007). In practice, this process is complex and typically involves several iterations. In the authors' experience, some students have difficulty managing the interplay between data analysis and context. Students may 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. In this paper, we present an icebreaker activity where students collect data, create graphs, and then try to determine which graphs presented by other students match their question. The activity was originally created for face-to-face post-secondary introduction to statistics courses (White et al., 2018). Here we discuss modifications for an online environment and discuss the similarities and differences in observed student behavior between the two modalities of teaching.

MOTIVATIONS

With the onset of COVID-19 and the move to remote teaching, many instructors taught online for the first time and were forced to quickly learn online teaching tools and develop pedagogical strategies for synchronous and asynchronous teaching (Burgess & Sievertsen, 2020). To combat the isolation common in synchronous online environments (Dorn et al., 2020), we found it is important to quickly create connections in the classroom and to teach students norms of interacting online via shared documents and discussions in breakout rooms. Through some simple changes to the original activity, we found that the online icebreaker provided some unique opportunities to create the connections and set the norms for the rest of the semester as a low stakes and fun task. Given the new knowledge learned during the pandemic, we predict that the number of online courses will increase considerably compared to pre-pandemic levels and hope this activity will be of use moving forward.

ICEBREAKER ACTIVITY

Goals

The principal pedagogical goal of the activity is to provide a fun setting where students must attend to characteristics of a data set to make decisions. The structure and sequencing of the activity requires that students rely more on numerical data features than contextual clues. In addition, careful selection of questions can provide a springboard for discussions about different sources of variation. From a researcher perspective, the freedom given to the students provides insights into which graphs students choose, which features of the data set students attend to, and why.

Participants

The second author taught an online section of 90 students in an introductory data science course. The course has an introduction to statistics prerequisite. This activity can be implemented with

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

students at any level (from elementary to post-secondary) and does not involve previous knowledge of statistics but requires some experience with online communication tools.

Main Modifications to the Original Activity

In the original activity, sheets of paper, each one with a question printed with a large font size, were taped to the back of the students (White et al., 2018). Each student introduced themselves briefly to their classmates, turned around, and collected data by asking other students to answer the question on their back without revealing the question. Each student then created a graph of their data. Students were split into two groups, each with the same set of questions. Then students tried to match their graphs across the two groups. Finally, the students were asked to match the questions with the graphs. Some features of the activity are lost in our online version. Here, the students know the questions in their group but then try to match their graphs to those from other groups. By working in groups, the graphs represent the consensus of the group, rather than a single students' thinking.

Set Up Before Class

Prior to the activity, the instructor must choose a set of questions that students will answer (see some examples in Table 1), prepare clear instructions, and prepare online documents that students can work on collaboratively to create their graphs. When we implemented the activity for our 90 students, the instructor created 9 breakout rooms and preassigned ten students to each room. Each breakout room is assigned two questions: A and B. The A question is common to three breakout rooms (BR; see Table 1). The B question is unique to each room. So, a total of 12 questions were needed. To help the activity "break the ice," the questions should be interesting to students and reveal interesting pieces of information for them to learn about each other but not be too sensitive that some students will be reticent to answer. Because in the final steps of the activity, students will attempt to match graphs of their data sets, all questions should require only numerical answers. To make the follow up discussion richer, one should consider the statistical properties of the variables involved. The examples in Table 1 include variables with only two possible values (3B), a large mean (7B), and with likely outliers (3A). Choosing variables with very different scales encourages students to think about the information center and spread provide about a distribution. Finally, careful selection of questions can lead to significant discussions about the sources of variation in data. For example, the cause of variation for questions 1A, 1B, and 2A are quite different. The differences in age are due to the natural variation in the population of students. On the other hand, the year Napoleon was born, 1769, was unknown to many students, so the variation came from their guessing the year. Finally, for the question about the number of fingers and toes, no variation is expected.

Set Up During Class

After a brief introduction and explanation of the activity, students are given access to online repositories that contain detailed instructions and allow students to post their graphs to share with classmates. In our implementation, we used Google Drive. Students are sent to breakout rooms and start reviewing the detailed instructions (which also had some general icebreaker questions such as "Have you done anything fun during the break?", "If you had to eat one meal every day for the rest of your life what would it be?", "The zombie apocalypse is coming, who are 3 people you want on your team?").

In the instructions, students are also asked to assume a group role. To foster active discussions online with students who do not yet know each other, we decided to assign specific roles to each student (see Table 2).

While students get to know each other and become familiar with the instructions and the online tools, the instructor visits each room and posts questions A and B in the chat. This time also gives an opportunity for the instructor to answer quick questions, make sure that students know how to share their screens, and encourage students to keep their video on with their microphone unmuted and to stay engaged.

Table 1. Examples of questions for the activity

BR	Questions
1	A. How many fingers and toes do you have? B. What year was Napoleon born?
2	A. How old are you? B. What is the lowest temperature in °F you have ever experienced?
3	A. How much did you spend on your last haircut? B. Do you have pets? Report 1 for Yes, 2 for No
4	A. How many fingers and toes do you have? B. How many languages do you speak?
5	A. How old are you? B. How many fish are there in Lake Austin?
6	A. How much did you spend on your last haircut? B. About how many students are enrolled at UT Austin this semester?
7	A. How many fingers and toes do you have? B. About how many miles is it from here to the farthest place you have ever visited?
8	A. How old are you? B. What is the average rent for a one-bedroom apartment in your neighborhood?
9	A. How much did you spend on your last haircut? B. How many friends/followers do you have on your favorite social media platform? (Answer 0 if you don't have social media)

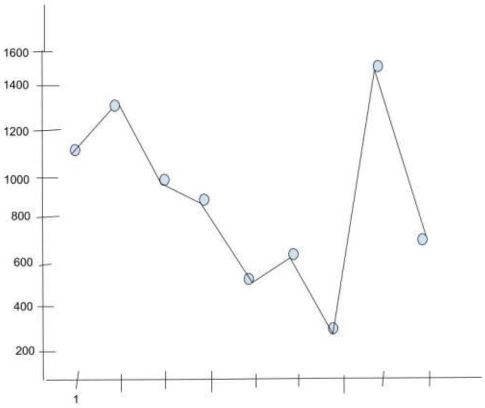
Table 2. List and description of roles for each participant of each breakout room

Role	Number	Task
Facilitator	1	The facilitator reads the questions, keeps track of the time and makes sure everyone stays on task.
Scribe	2: A & B	Each scribe records the data reported by each participant when answering the question and posts the data (just the numbers, no unit) in the chat.
Artist	2: A & B	Each artist creates a graph to represent the data for a question.
Matchmaker	2: A & B	Matchmaker A matches their graph for question A with graphs from two other breakout rooms.
Judge	1	The judge ensures that a conclusion is reached and that (almost) everyone agrees. Posts conclusion in the chat for the reporters to use when reporting back to the main room.
Reporter	2: A & B	Reporter A reports to the main room on how their room decided which graphs matched their question A. Reporter B reports to the main room on how their room decided which questions align with graphs from set B.

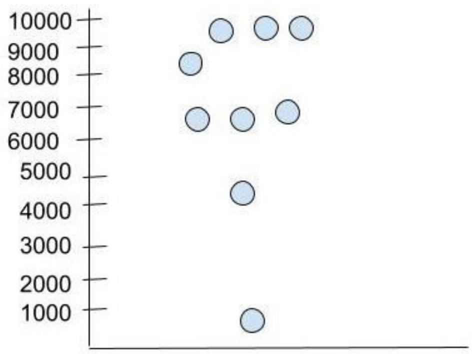
Data Collection and Display

The *facilitator* reads both questions, and the *scribes* gather answers for their respective questions. Students are encouraged to share anecdotes about their answer to the questions. Each *artist* then uses a Drawing template in Google Drive to sketch a graph of the data for their question. Some students ask for more guidance on what graph to create, but we do not provide any. Eventually we want students to choose graphs that are appropriate to the statistical properties of a variable and help to satisfy the intended purpose. Although no type of graph is inherently better than another for univariate data, statistics textbooks typically present the same set of graphs: histogram, box plot, stem-and-leaf plot, dot plot, pie chart, and bar chart (e.g., Starnes et al., 2012). One goal of this activity is to see what kind of graphs students naturally choose. In the original face-to-face activity students created graphs with paper and markers. Despite using a computer for this online version, the types of graphs created were very similar.

Research with younger students indicates that they consider 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 such as the name of the respondent (Konold & Higgins, 2003). Still, research with older students indicates that students think that case-value charts, where the height of a bar represents a value rather than the frequency, best display the distribution of a quantitative variable (Kaplan et al., 2014). Some groups in our class also showed a reluctance to aggregate data. For example, in Figure 1(a), the students choose to represent the data with a scatterplot where the horizontal axis is an index representing each of the 10 students in the breakout room. Many groups produced standard plots, histograms, pie charts, and bar charts. However, other students used nonstandard plots, as shown in Figure 1(b), which displays some form of a dot plot, though it is not clear what values are represented on the horizontal axis. As we progress through the semester, we use these initial examples to compare and contrast the graphs with standard statistical plots, emphasizing key features and how the standard plots differ from the graphs students initially made.



a. Graph for B9:
How many friends/followers do you have on your favorite social media platform?



b. Graph for B7:
About how many miles is it from here to the farthest place you have ever visited?

Figure 1. Examples of graphs created by students

Graphs posted by artists to share with the rest of the class are labeled Ak and Bk, for each question respectively, where k is the breakout room number. Through the shared folder on Google Drive, each class member has access to all 18 graphs created by the 9 groups.

Matching

There are two kinds of matching tasks. First, led by *Matchmaker A*, the breakout room tries to determine which two graphs among the other eight A-graphs represent the same question they had. Secondly, the instructor reveals the entire list of nine questions from the second set of questions, B, in random order. The group then must come to a consensus on which graph B matches their question B. The *Judge* is the final arbiter if there is a disagreement. Finally, the students return to the main room and the *reporters* share the choices and reasoning from each room, reporting any disagreement. For the A-graphs, the class engages in a discussion about how, despite the fact that the data is from two entirely different samples of students, it is still possible to match the graphs with the goal of providing the foundation for a discussion of probability distributions in later classes.

Table 3 shows the features students attended to when trying to match the A-graphs. Because students expected a unique answer for the question "How many fingers and toes do you have?", they were looking for a graph that reported only one possible outcome. (However, in previous implementations, some students reported different values.) The question, "How old are you?" had about the same center as the question "How much did you spend on your last haircut?" (Most students were between 19–21 years old.) However, responses to "How old are you?" displayed some small variation and responses to "How much did you spend on your last haircut?" generated a wider range of values. Students paid attention to variability when matching their graphs and looked for a graph representing more possible values to match with the latter question.

Table 3. Features used to match graphs for questions from set A

Question	Features
A1, A4, A7	Single bar, single line
A2, A5, A8	Frequencies, Range, Type of graph
A3, A6, A9	Use of histogram, range of data

In addition to the range of values, when matching questions from set B to the graphs, students considered unique features of the data. For example, for B8 ("What is the average rent for a one-bedroom apartment in your neighborhood?"), students recognized numbers that "make sense for rent" and for B9 ("How many friends/followers do you have on your favorite social media platform?"), students noted that there were no repeated values because that outcome is pretty unique for each student.

Class Discussion

This activity provides the opportunity to discuss which aggregate features of the data can be used to match 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 that use aggregates (50% of the students spent between $10 and $15 dollars on their last haircut). Kaplan et al. (2018) reported that students were much more likely to use the shape and the center when describing a histogram in a STEM context. However, in our matching task, when *reporters* described the rationale for finding a match from their breakout rooms, they mostly referred to the "ranges" of numbers to describe spread (min and max values), and to a lesser extent to the center. This may be an indication that this task can help students focus on the variability of a dataset. In addition, some groups paid attention to the type of graphs with the idea that some graph types are more appropriate with some specific types of variables.

Follow Up

Concepts that are touched on in the icebreaker activity can then be further developed in later classes. Arnold and Franklin (2021) provide a framework for question posing (investigative) and question asking (interrogative) and examine what makes a good statistical question. Table 1 includes examples of poor statistical questions (1A, 1B) and better statistical questions (2B, 3A, 4B). Comparison of the questions can introduce the discussion of how to generate a good investigative question. Additionally, in the development of univariate graphical representations, aggregate features that were used in matching the graphs 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 et al., 2006).

DISCUSSION

The icebreaker activity presented in this paper is a fun way for statistics students to get to know one another and learn how to communicate and collaborate during an online course. Using an online platform such as Google Drive allows student to collaborate within and between breakout rooms. At the same time, the activity provides a starting point to discuss several important issues for representing data. The activity is flexible and can be adapted to different levels and settings. In smaller classes, one can reduce the number of breakout rooms and increase the number of questions used in each room. The activity can be more challenging in large face-to-face classrooms. 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. Pfannkuch and Wild (2000) provided a framework for the foundations of statistical reasoning with four components: consideration of variation, transnumeration, building and reasoning from models, and integrating the statistical and contextual. We argue that this simple activity provides an entry to three of the four components. Creation of the graphs from the raw data involves transnumeration: the forming of data representations to better understand the data. Students consider the variation of the data, namely the range, when trying to match the graphs from A. Finally, students integrate statistical features and context when matching questions to the graphs.

REFERENCES

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., & Witmer, J (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College report.* American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/2005gaisecollege_full.pdf

Arnold, P., & Franklin, C. (2021). What makes a good statistical question? *Journal of Statistics and Data Science Education*, *29*(1), 122–130, https://doi.org/10.1080/26939169.2021.1877582

Burgess, S., & Sievertsen, H. H. (2020, April 1). *Schools, skills, and learning: The impact of COVID-19 on education.* VoxEu.org. https://voxeu.org/article/impact-covid-19-education

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823. https://doi.org/10.1080/00029890.1997.11990723

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020, June 1). *COVID-19 and student learning in the United States: The hurt could last a lifetime*. McKinsey & Company. https://www.mckinsey.com/industries/education/our-insights/covid-19-and-student-learning-in-the-united-states-the-hurt-could-last-a-lifetime

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*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/gaise/gaiseprek-12_full.pdf

Kaplan, J. J., Gabrosek, J. G., Curtiss, P., & Malone, C. (2014). Investigating student understanding of histograms. *Journal of Statistics Education, 22*(2). https://doi.org/10.1080/10691898.2014.11889701

Kaplan, J. J., Lyford, A., & Jennings, J. K. (2018). Effects of question stem on student descriptions of histograms. *Statistics Education Research Journal*, *17(1)*, 85–102. https://doi.org/10.52041/serj.v17i1.177

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). National Council of Teachers of Mathematics.

Pfannkuch, M., & Wild, C. J. (2000). Statistical thinking and statistical practice: Themes gleaned from professional statisticians, *Statistical Science, 15(2)*, 132–152.

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS7).* ISI/IASE. https://iase-web.org/documents/papers/icots7/2D3_RUBI.pdf

Starnes, D. S., Yates, D., & Moore, D. S. (2012). *The practice of statistics* (4th ed.). Macmillan.

White, A., Guyot, L., & Walker, A. (2018) Data in search of a context: An icebreaker activity. 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*. ISI/IASE. https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_9K1.pdf