OVERVIEW

This is an extremely open-ended activity with no "correct" answer (except for Step 1). For that reason, it is a valuable example of true data analysis. Statisticians usually make guesses based on data but never know what the "correct" answer is. In this case, we created these data ourselves, but randomly so that they are realistic. Because of this, we know that the break-in occurred at exactly 2:30 A.M. (case 195). You should not mention this to the students initially, but after they've argued their conclusions they can use it to evaluate how close they came.

Activity Time: Two to three class periods

Objectives

- Identify important (or unusual) aspects of data.
- Justify conclusions based on data.
- Represent data with graphs.
- Summarize variability among data values with some measure of center.
- Judge when a shift in a group average is probably due to chance and when it is an indicator or real change.
- Evaluate how well different graphical representations of the same data show important aspects of the data.

Common Core Standards Addressed

Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predications.

Grade 7, Statistics and Probability Standard 2

Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

Grade 7, Statistics and Probability Standard 4

Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

Grade 8, Statistics and Probability Standard 1

Prerequisites

None.

Materials

- The Yo-Yo Mystery worksheet (one copy per student)
- **Yo-Yo Mystery.tp**

LESSON PLAN

Introduction (10 minutes)

Hand out The Yo-Yo Mystery worksheet.

The Information From the Police Report section sets up a scenario about a robbery at a yo-yo factory. Students are challenged to pinpoint the time period when the robbery most likely happened and compare this with the time periods for when the prime suspect has an alibi. You may want to read this section aloud to the class and agree on the facts before going on.

Students should quickly understand that they want to find the point when the machine began performing at a slower average rate. The catch, however, is that the machine doesn't make the same number of yo-yos each time period. The machine's productivity fluctuates, and this "natural" fluctuation makes it harder to spot the precise moment when the average drop in production occurred. One general approach is to get some sense of how much this production varies ordinarily (before the break-in) and then use this as a guide to decide when values seem to be occurring reliably below this range. To complicate matters, the range during which productivity seems to drop off is right on the border of the suspect's alibi, so students will need to make a judgment call about his guilt.

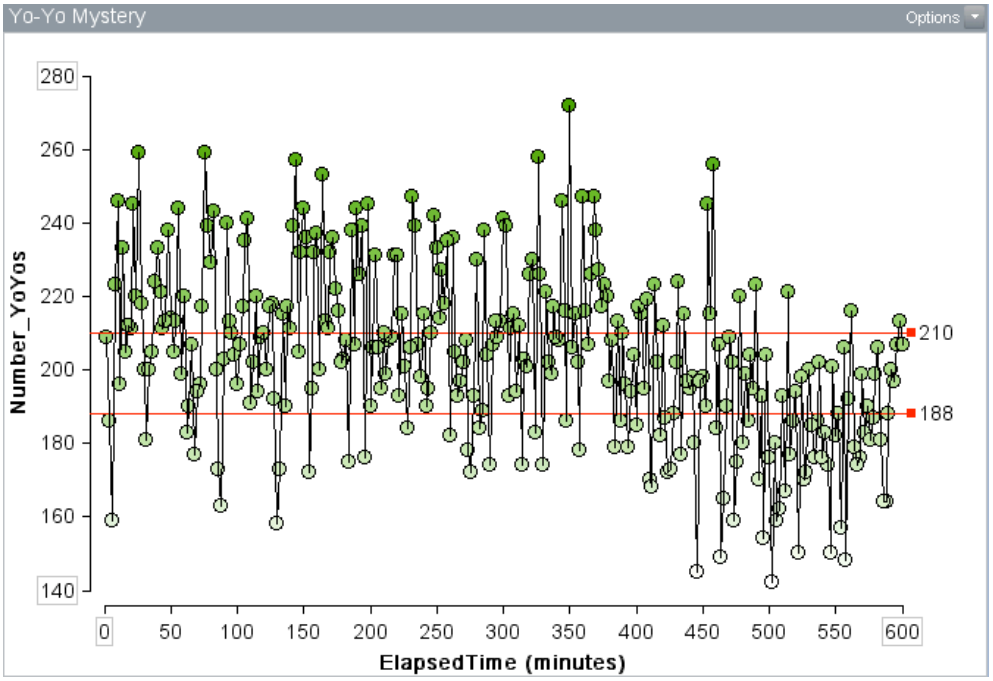
Because of its open-ended nature, this activity could take several hours. Once students get into the data, they may not want to stop. At the minimum, you should allow one full class period for data exploration and part of another class period for writing conclusions. If you have the time, you could make it an extended project, culminating in group presentations.

Plot and Investigate

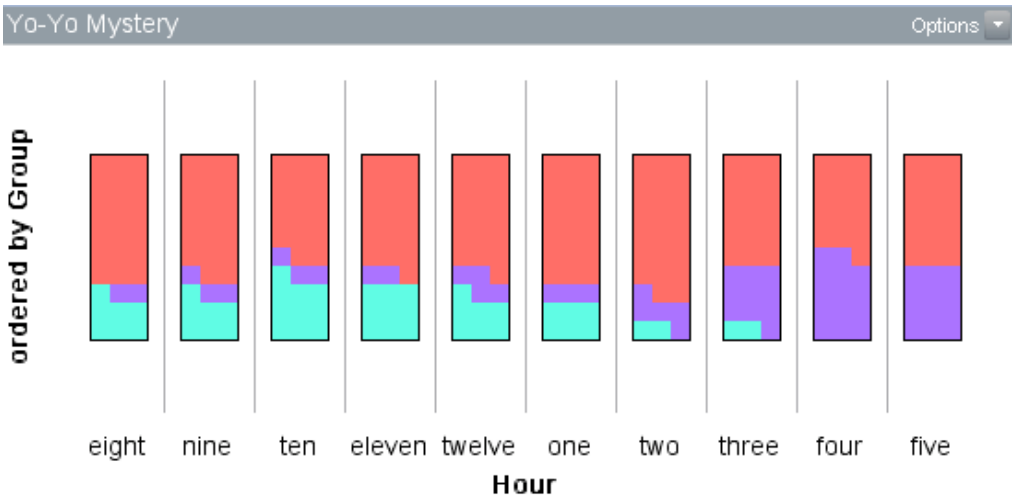
Have students move to computers and open the document **Yo-Yo Mystery.tp**.

For Steps 3–5, students might first try a time series graph (especially if they have done the activity Men's 100-Meter Dash at the Olympics, or Men and Women at the Olympics). However, the machine's productivity fluctuates so much during normal operation that it is hard to notice a time period when the number of yo-yos dropped drastically. The production does seem lower sometime after 350 minutes, but no one case icon stands out.

Students might try using two horizontal reference lines, placing one at the known average of 210 (as shown on the next page) and then eyeballing where the other one might go to fit the data after production decreased.

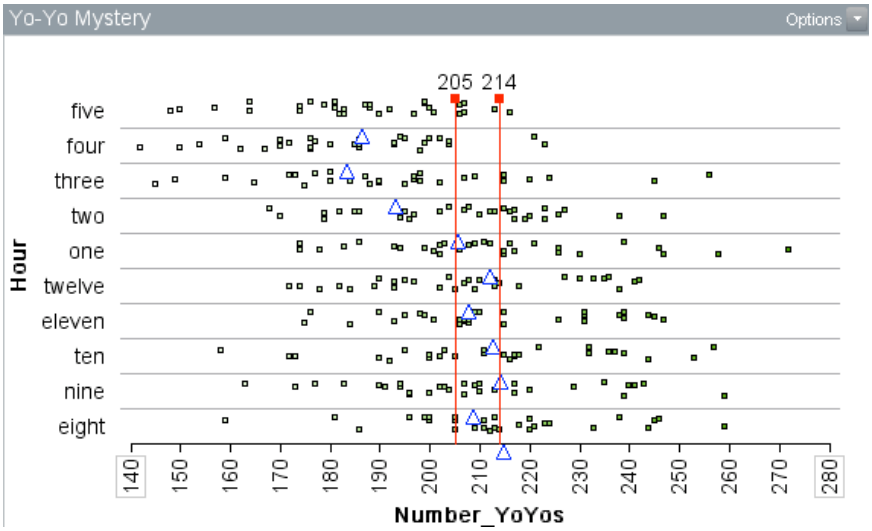


The graph below implies that the break-in happened during the 2:00 A.M. hour or the 3:00 A.M. hour. The cases are separated by *Hour,* with fused rectangular icons, and colored and ordered by *Group*. Notice that every hour up until the 2:00 A.M. hour has six or more "high" production periods. But beginning in the 2:00 A.M. hour the number of high production periods drops noticeably, with a corresponding increase in the number of low periods. So from this graph the 1:00 A.M., 2:00 A.M., and 3:00 A.M. hours seem likely periods in which the break-in occurred.

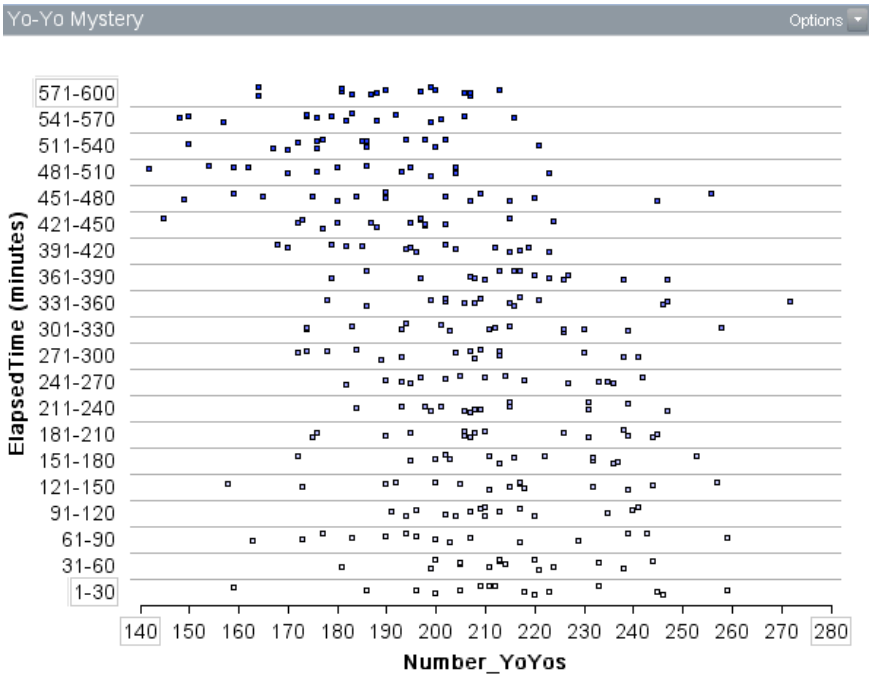


Students may feel that there is not enough information from the graph above and probe further.

In the following graph, the hours up to 2:00 A.M. seem to be fluctuating within the band indicated by the reference lines, between 205 and 214 minutes. But production in the 3:00 A.M. hour was lower on average than anything previously seen, and production continues at a lower rate in the 4:00 A.M. and 5:00 A.M. hours. This suggests that the suspect could be guilty.



But it's possible that the break-in occurred, say, 15 minutes before 3:00 A.M. If it did, the drop in average production would not be very apparent when looking at bin sizes of one hour, because during the majority of time in the 2:00 A.M. period the production would have been at the normal rate. The graph below shows half-hour intervals of *ElapsedTime*. (Set the bin width to 30.) There appears to be a drop from the previous average in the interval 391–420 minutes, or between 2:31 A.M. and 3:00 A.M. This suggests that the suspect is not guilty. (The document used to create the data has the break-in set to occur at exactly 2:30 A.M.)



Extensions (optional)

1. As mentioned in the overview, this activity could be enriched by conducting a mock trial in which students present their arguments and graphs before a "jury" of people unfamiliar with the project. This would help further develop students' communication and presentation skills, and provide additional motivation for choosing graphs that help tell their story fairly and persuasively.

2. You may want to hold a discussion about the possible consequences of trying a suspect without being entirely sure of his or her guilt. As part of this you could discuss common legal terms such as *burden of proof* or *reasonable doubt*. Collaborating with your school's social studies teacher, you could even use this activity as a way to motivate learning more about the foundations of your country's justice system.

3. You might use the document **Yo-Yo Master.tp** to create alternative data sets for students to explore on their own time. The document is set up so that you can change the break-in time and produce a new random sample of data. The document also includes a plot and instructions for a classroom demonstration that you and your students may find illuminating.

ANSWERS

2. a. 201 yo-yos were made during the previous two-minute period.

   b. This measurement was made during the fifth hour on the clock, or during the 5:00 A.M. hour.

   c. This measurement was made 548 minutes after 8:00 P.M., specifically at 5:08 A.M. It might help students to make a chart that they can use to quickly translate between elapsed time in minutes and the time in hours and minutes. From such a chart they would see that at 5:00 A.M., 540 minutes had gone by. 548 would be 8 minutes later, or at 5:08 A.M.

4. Answers will vary depending on the plots and statistics used. Students must explain how the plot supports their answers. As a broad guidepost for answers, students should find that the break-in happened sometime between 2:00 A.M. and 4:00 A.M. Students should try to refine their answer by narrowing the range of data from the nearest two hours, to the nearest hour, to the nearest half-hour, and so on.

5. Answers will vary depending on the answers to Step 4. For those students who feel that the best they can say is that the crime happened somewhere between 2:00 A.M. and 4:00 A.M., a reasonable conclusion would be that the data don't provide strong enough evidence either to rule the suspect out or to convict him. If they can narrow the break-in time down to 2:00 A.M. to 3:00 A.M., then they should conclude that the suspect could not have done it. Some students might validly point out, however, that given these data, the break-in may have in fact occurred after 3:00 A.M., but that by chance the production output in the prior period happened to be lower than average.