

---

# Reasoning About Distributions

## Sticky Note Activity

In this activity, you are going to be introduced to the idea of distribution to represent individual observations of the same variable. We will use physical sticky notes to represent individual observations.

### Data For a Real Variable

1. You will be given a sticky note
2. Locate the **battery percentage remaining on your phone**
3. Record your percentage on your sticky note

### Summarizing the data

Hold up your sticky note

1. As a "class" what is our average percentage battery remaining?

*This will vary.*

#### Instructor Tip

*Phone battery percentage is a well-chosen variable. If you need a substitute, you can use minutes to get to school, hours of sleep, height. Avoid using variables that will not vary a lot like age.*

2. How variable is that percentage?

*This will vary. The important thing is that students will struggle to answer with any precision just by looking at the held-up sticky notes. They can guess a rough range but not pin down a center.*

### Creating a Distribution

Bring your sticky note to the axes on the board

1. As a "class" what is our average percentage battery remaining?

*Answers will vary, but students should now give a much more confident answer that is reasonably close to the actual class mean. They can visually identify the mean.*

2. How variable is that percentage?

This is an important question to start thinking about how to measure variability.

Instructor Tip

This is a key moment of the activity. The same two questions that students couldn't answer a minute ago are now easy. Make this explicit: "A minute ago, none of us could answer these confidently. What changed?" The answer you want is that organizing the data into a distribution lets us see patterns in the collection that are invisible when we look at individuals.

**Vocabulary**

This collection of points in a graph makes it easier to think of the data as an entity: a **distribution**. Distributions are characterized by: **Shape, Center, and Variability**.

## Sorting Histograms Activity

Your group will be given a stack of different graphs, representing different data sets. The variable being measured is indicated on the horizontal axis, and the vertical axis represents how many data values are in each bar. Working together as a group, sort the stack of graphs you are given into piles, *according to those that look the same or similar*.

Once you have agreed upon the groupings for the histograms, discuss and write down answers to the following questions for each grouping of histograms.

1. Describe what's similar about the graphs in each group.

This will vary. Some will look at the center, others will look at the variability, and others will do a combination of them. Typical groups: symmetric, right/left skewed, uniform, bimodal

2. Give the group a name that you think describes the general shape.

This will vary. Typical answers: the hill, the slide, the double hump

3. Pick one graph in each group that is the best example of that group.

This will vary.

4. If there were graphs that didn't fit easily into any group, try to determine why it was more difficult to place them in a group. What is different or confusing about them?

*This will vary.*

## Guess the Histogram Activity

In this activity, you are going to explore distributions of real data collected from students across the United States through Census at School (<https://ww2.amstat.org/censusatschool/>). Census at School is an international classroom project that engages students in grades 4–12 in statistical problem solving. Students complete a brief online survey, analyze their class census results, and compare their class with random samples of students in the United States and other countries.

The six variables you will explore are:

- **Age:** the student's age in years
- **Height:** the student's height in inches
- **Sleep Hours (Non-School Night):** hours of sleep on a weekend or non-school night
- **Home Occupants:** the number of people living in the student's home
- **Birth Month:** the month (1–12) in which the student was born
- **Text Messages Sent Yesterday:** the number of text messages the student sent on the previous day

1. Match every histogram to its associated variable

*Answers will vary. Pairs of groups usually agree on the easy ones (Birth Month is almost always identified correctly) and split on the harder ones.*

2. Pair with another group to talk about what pairs you made. Did you get the same things?

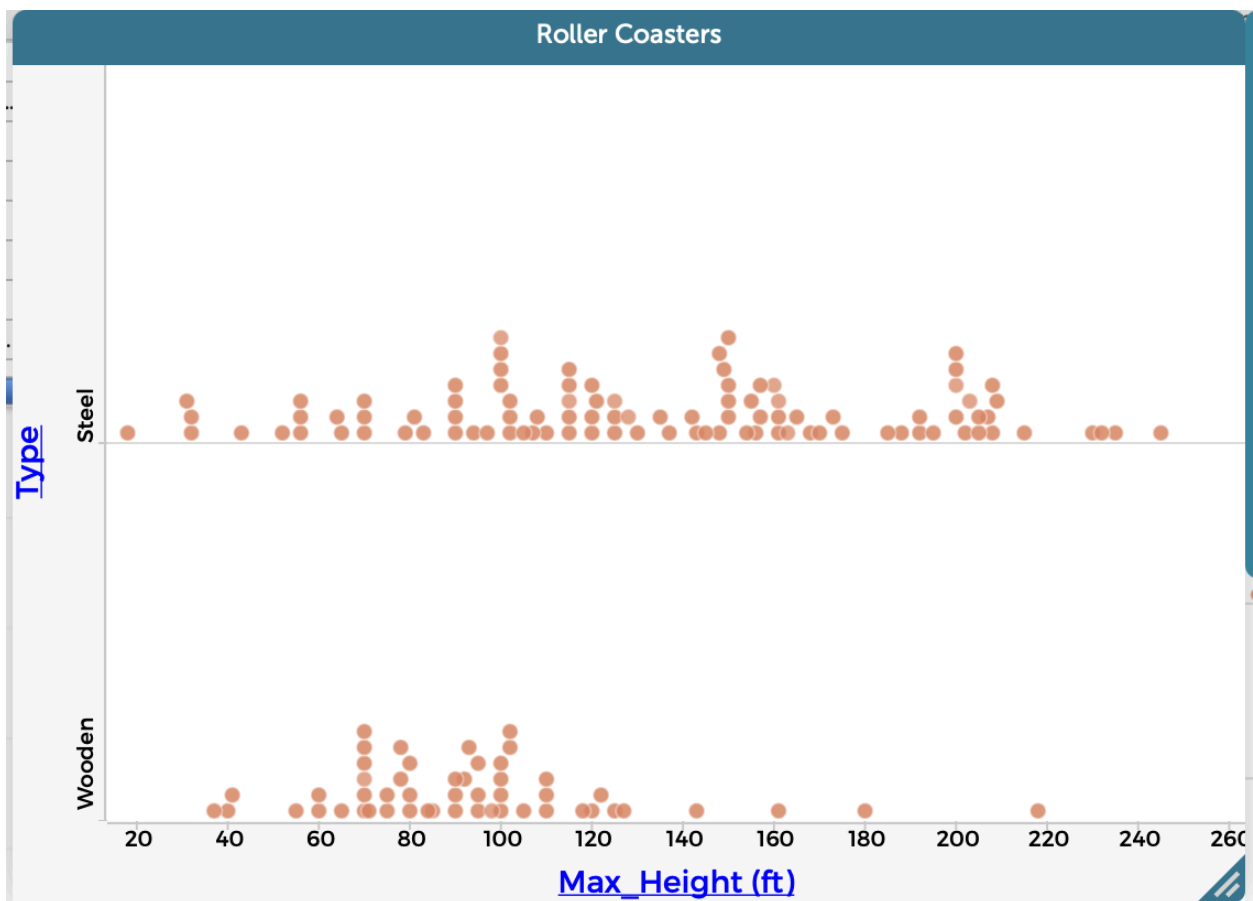
### Instructor Tip

*A good extension: ask "What would the histogram of school-night sleep hours look like compared to non-school-night?" or "What would the histogram of height look like for just 10th graders?"*

## Rollercoaster Activity

### Which rollercoaster should I try?

In this activity, you are going to compare two distributions of the maximum height (ft) reached by a sample of rollercoasters. The two types of rollercoasters featured are made of steel or wood. The data for this activity is from [CODAP](#).



1. What does each dot in the plot represent?

Each dot represents one rollercoaster, plotted at its maximum height in feet.

2. Describe the distribution of the maximum heights for the steel rollercoasters (hint: mention shape, center, and variability).

Answers will vary, but should include all three components. Expected observations: the distribution is right-skewed or roughly symmetric, with most coasters in a moderate height range and a tail of very tall coasters; the center (mean or median) is noticeably higher than for wooden coasters; the spread is wide, with heights ranging from small kiddie coasters up to several hundred feet.

3. Describe the distribution of the maximum heights for the wooden rollercoasters (hint: mention shape, center, and variability).

Answers will vary, but should include all three components. Expected observations: the distribution is right-skewed, the center is lower than for steel; the spread is narrower, wooden coasters don't reach the extreme heights that steel coasters do.

4. Which would you guess has the higher mean?

Steel rollercoasters.

5. Which has more variability?

Steel rollercoasters.

6. Which type of rollercoaster could you recommend for me if I don't like heights? Explain your answer using evidence from your descriptions of the rollercoasters above.

Wooden rollercoasters would be the better recommendation. Wooden coasters have a lower center AND less variability, which means the chance of ending up on an unexpectedly tall one is smaller. A strong student answer cites both features: "the wooden distribution has a lower center, and its tallest values don't come close to the tallest steel coasters, so you're less likely to get a surprise."