Learning Goal: For the distribution of a quantitative variable, describe the overall pattern (shape, center, and spread) and striking deviations from the pattern.

Specific Learning Objectives:

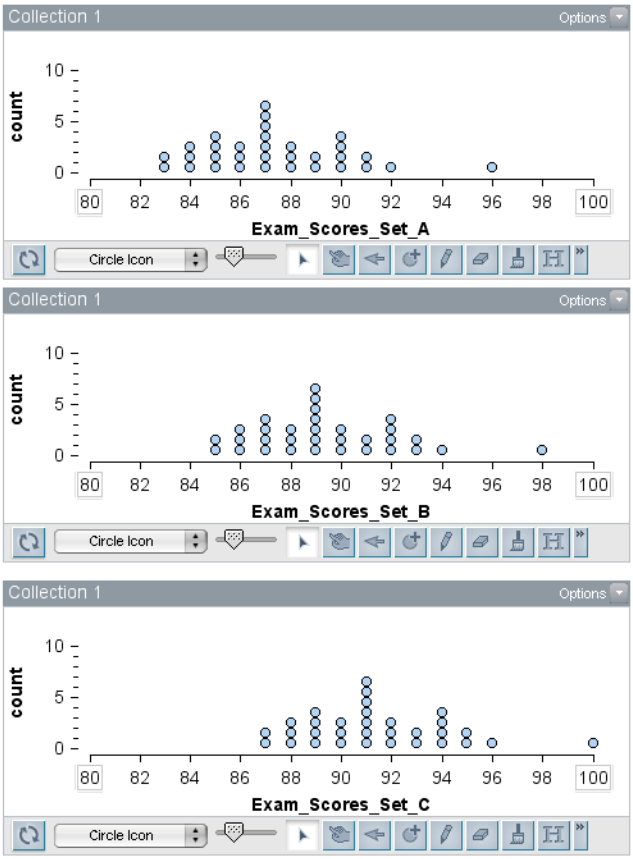
- Develop a way to describe and distinguish graphs of a quantitative variable.
- Identify reasonable explanations for what might explain the differences seen in different data sets.

Introduction: Statisticians collect data in order to study and analyze groups. They focus on the group's data in aggregate, instead of focusing on individuals. For a statistician creating and comparing graphs is often the first step in analyzing data.

In this activity you will compare and contrast the graphs of hypothetical sets of exam scores. In other words, these graphs show made-up data. These hypothetical data sets have been constructed to help you begin to "see" like a statistician. By comparing these data sets, you will begin to develop an informal understanding of the key features of a graph that statisticians use to describe data. We will develop these ideas more formally in future lessons.

During this activity do your best to describe what you see. Jot down notes to capture your thinking as you go. You can use your notes during our class discussion of this activity. This activity does not require you to remember anything or to apply previous knowledge.

1) Compare and contrast the 3 graphs shown at the right.



   a) How are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?

   b) For each graph, pick a single exam score to summarize the overall performance of the students. In other words, summarize each set of data with one number.
   Set A _____ Set B _____ Set C _____

   c) The average score for Set A is 87.4. What do you think the average score is for Set B and for Set C? (See if you can answer this without doing any calculations.)

   d) Which, if any, of the statements below is a reasonable explanation for the differences in the graphs? Why?

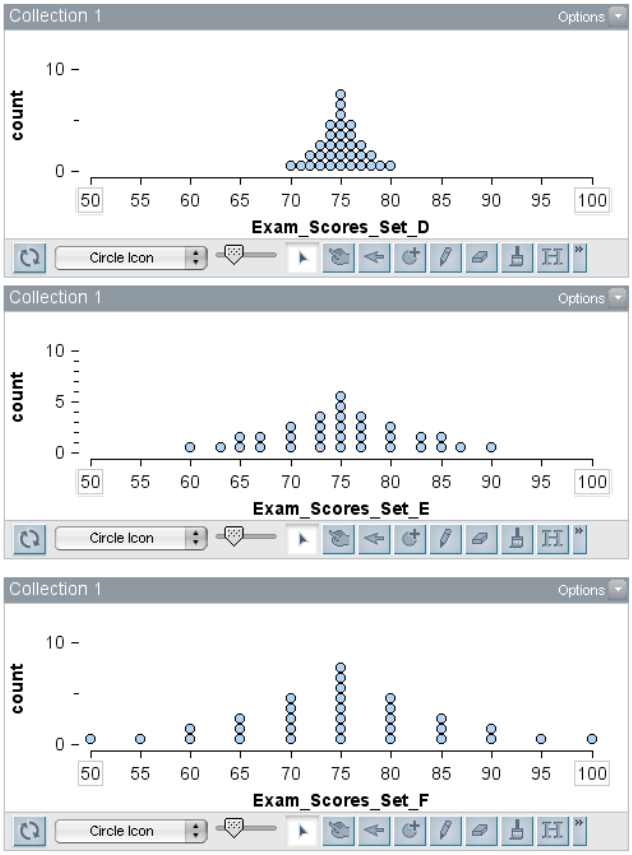
   - The graphs represent different classes. Different groups of students will obviously perform differently on an exam.

   - The graphs represent a single class after the teacher adjusted the grades. The teacher realized that some of the exam questions were not well written, so she adjusted the grades by adding points to the original exam scores.

   - There are no differences in the graphs. They look the same.

2) Compare and contrast the 3 graphs shown at the right.

a) How are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?



b) The average exam score is the same for each set of data (D, E and F). What do you think the average is?

c) Which of the statements below, if any, is a reasonable explanation for the differences in the graphs? Why?

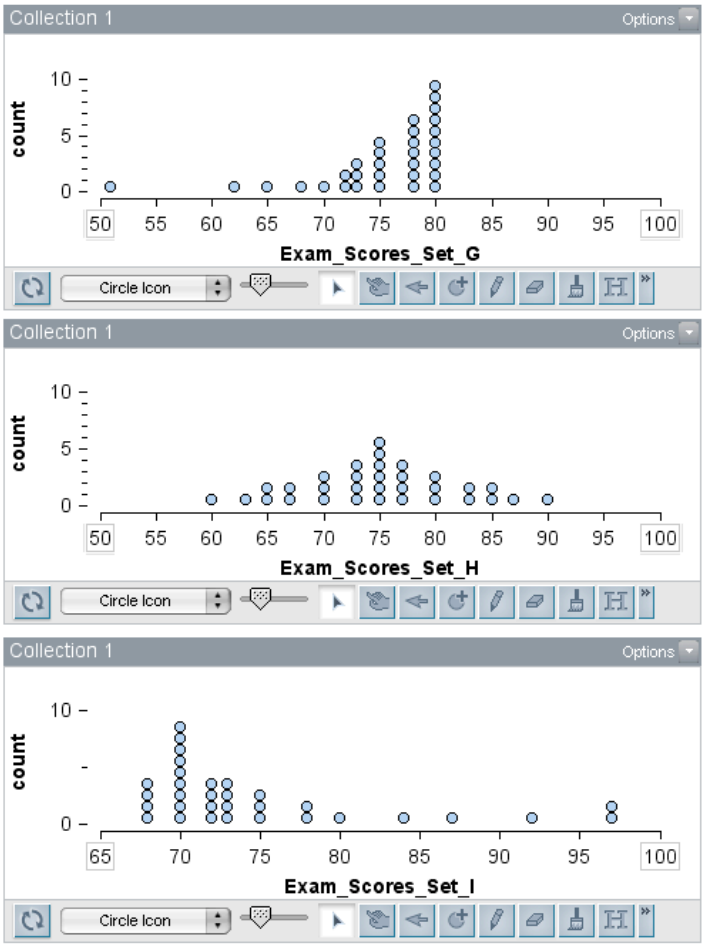
- The graphs represent a single class after the teacher adjusted the grades. The teacher realized that some of the exam questions were not well written, so she adjusted the grades by adding points to the original exam scores.

- The graphs represent 3 different classes with teachers that have different grading standards. One is an easy grader. One is a really hard grader.

- The differences could be explained by whether the teachers allowed the students to work together on the exam and how much time the students were given to finish the exam.

3) Compare and contrast the 3 graphs shown at the right.

a) In each of these 3 graphs, the average score is the same. The average score is 75. In what other ways are the graphs similar? How are they different? What is the most distinctive feature that distinguishes these three graphs from each other?



b) What might explain the differences in these graphs?

Distributions for Quantitative Data: Vocabulary

Individuals	Individuals are the objects described by a set of data. Individuals may be people, but they may also be animals or things.
Variable	A statistical variable is a characteristic that describes individuals in the data. A variable can take different values for different individuals.
Categorical variable	A categorical variable places an individual into one of several groups or categories. Example: a person's college major
Quantitative variable	A quantitative variable takes a number value, which you can add or average. Example: number of people in household
Distribution of a variable	A statistical distribution is an arrangement of the values of a variable (often a graph) showing their observed frequency of occurrence. Here is an other definition: "a representation that shows the possible values of a variable and how often the variable takes those values."
Describing the distribution of a quantitative variable	To describe a distribution, describe the shape, center, spread and outliers. Shape: describe the overall trends in the data. Typical ways to describe shape include symmetric, left or right skew, uniform. Center: give a single number that represents the data; a typical value or average (We will make this more precise in Topic 2.2) Spread: give a single number that measures how much the data varies. Range (maximum value minus minimum value) is one way to measure spread. (We will learn other ways to measure spread in Topic 2.3 and 2.4) Outliers: unusual data values
Shape	Symmetric: The right and left sides of the graph are mirror images of each other Skewed to the right: The graph has more spread in the upper half than the lower half. There may be outliers to the right, so the graph has a longer tail to the right. Skewed to the left: The graph has more spread in the lower half than the upper half. There may be outliers to the left, so the graph has a longer tail to the left. Uniform: The graph is shaped like a rectangle. Each value occurs with the same frequency.