CHAPTER TWO: GRAPHS   Review September 23 ✎ Test October 3

For research to be valuable, it must be shared, and a graph can be an effective way to do so. The fundamental aspect of a good graph is that it makes the results clear at a glance. Graphs often fail to reach this standard due to poor labeling, being more complicated than necessary, or being an inappropriate type for the data. This chapter discusses circle graphs, time series plots, scatter plots, and bar graphs, including histograms which are used to show frequency distributions.

2-A   Types of Graphs   Monday • 9/19

time series plot • scatter plot • bar graph • pareto chart • circle graph

❶ Choose an appropriate type of graph for two variables.
❷ Make a circle graph.
❸ Make a two-variable time series plot, scatter plot, or bar graph.
❹ Make a time series plot, scatter plot, or bar graph with two independent variables.

2-B   Histograms   Wednesday • 9/21

distribution • class • frequency histogram • relative frequency histogram • uniform • bimodal • skewed

❶ Make a frequency histogram.
❷ Make a relative frequency histogram.
❸ Describe the shape of a distribution graphed in a histogram.

2-A   Types of Graphs

Different types of graphs are used for different types of data.

A TIME SERIES Plot is used for *x-y* data pairs when the independent variable *x* is time on a timeline and the graph represents a function (no two data points can have the same *x* value).

A SCATTER Plot is used for *x-y* data pairs that do not necessarily represent a function (two data points could have the same *x* value).

A BAR Graph is used to compare averages, totals, or frequencies of two or more categories.

In many cases, a bar graph is more clear if it is a PARETO Chart, in which the bars are displayed from largest to smallest.

When a bar graph has only one variable, making the total of all categories 100%, a CIRCLE Graph can be used if there is no advantage for a pareto chart.

A HISTOGRAM is a bar graph in which the *x*-axis has equal-width numerical categories that connect (e.g., 0-10, 10-20, 20-30) and the *y*-axis shows the frequency or proportion of data in each range.

❶ Choose an appropriate type of graph for two variables.

   1. Use the flowchart below.
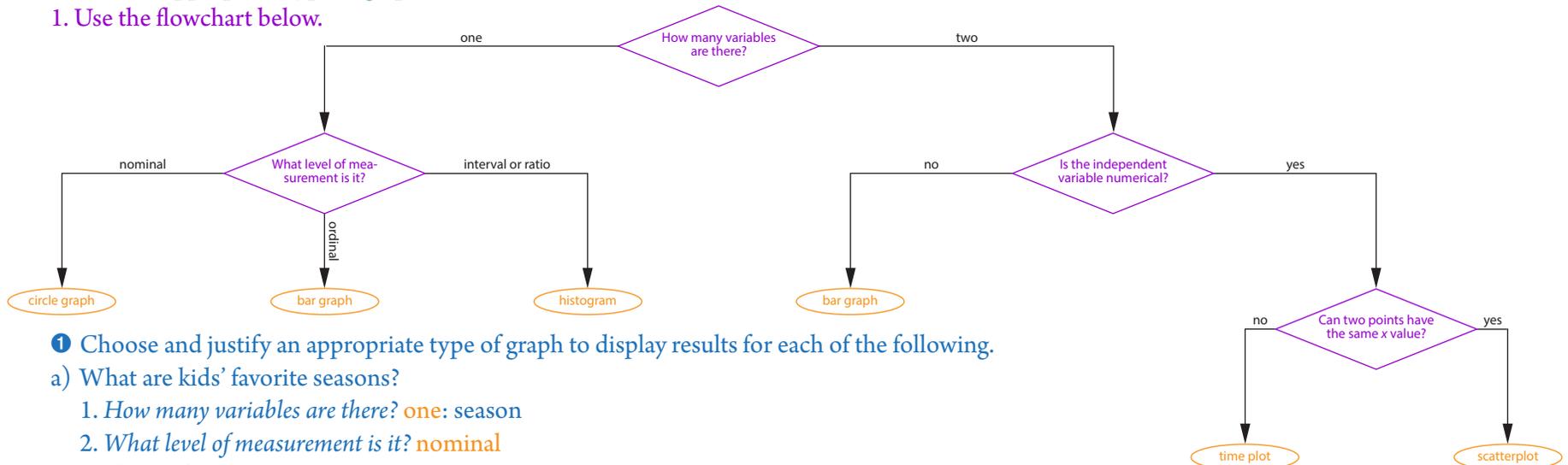
❶ Choose and justify an appropriate type of graph to display results for each of the following.

a) What are kids' favorite seasons?

   1. *How many variables are there?* one: season

   2. *What level of measurement is it?* nominal

   circle graph

b) How much do adults pay for food each month?

   1. *How many variables are there?* one: montly food bill

   2. *What level of measurement is it?* ratio

   histogram

c) How much rainfall has Scotts Valley had in each of the past 30 years?

   1. *How many variables are there?* two: the IV is year and the DV is amount of rainfall

   2. *Is the independent variable numerical?* yes

   3. *Can two points have the same x value?* no: there is only one rainfall total for each year

   time series plot

The most important aspect of a graph is that it makes the results clear at a glance. This can be achieved by using a type appropriate for the data, precisely labeling it, and keeping it neat and not unnecessarily complicated.

❷ Make a circle graph.

1. Multiply each proportion by 360°.
2. Use a protractor to measure out each segment.
3. Title the graph and label the slices.

❷ Out of 70 kids, 31 preferred chocolate, 26 preferred strawberry, and 13 preferred vanilla.

chocolate: $\frac{31}{70} \cdot 360° = 159°$ strawberry: $\frac{26}{70} \cdot 360° = 134°$ vanilla: $\frac{13}{70} \cdot 360° = 67°$
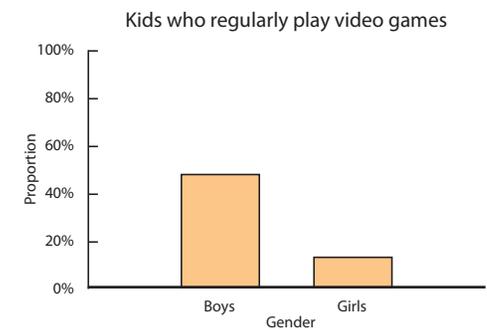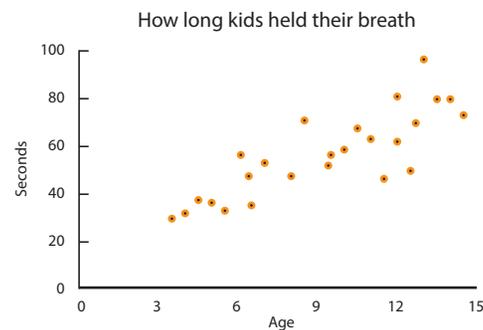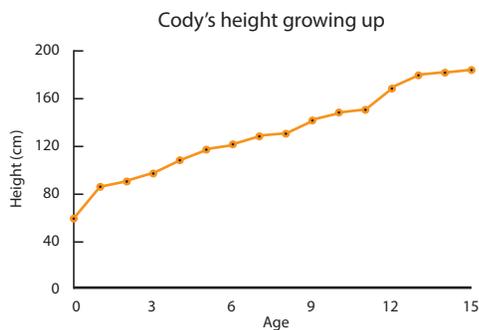
❸ Make a two-variable time series plot, scatter plot, or bar graph.

1. Label and scale the independent variable on the *x*-axis. For a bar graph, each level of the independent variable should be labeled and should be done in order from largest to smallest bar unless there is a specific reason to do otherwise.
2. Label and scale the dependent variable on the *y*-axis. The scale should represent an appropriate range for the variable, starting at zero unless there is a specific reason to do otherwise, and should be in natural, equal increments with consistent decimal places. If proportions are being compared, use percents to scale the *y*-axis.
3. For a bar chart, draw the bars. If proportions are being compared, divide to convert each bar's frequency to a percentage. For example, if 12 of the 25 boys surveyed regularly play video games, the height of the bar is $\frac{12}{25} = 48\%$.
   For a time series plot or scatter plot, plot the points.
   For a time series plot, connect the points.
4. Give the graph a title that clarifies its purpose.

❸ Make up realistic data and create each of the following:

a) a time series plot showing Cody's height through age 15
b) a scatter plot showing how long 25 kids of different ages could hold their breath
c) a bar chart comparing teenagers by gender on whether or not they regularly play video games

Kids' Favorite Ice Cream Flavors

Cody's height growing up

How long kids held their breath

Kids who regularly play video games

A single graph can display results from two independent variables by using a different color for each level of the second independent variable.

❹ Make a time series plot, scatter plot, or bar graph with two independent variables.

    1. Make a legend showing one color for each level of the second independent variable.

    2. Follow the directions for a two-variable graph (see ❸).

    3. For a time series plot or a scatter plot, plot multiple data sets on top of each other, one for each level of the second independent variable.

    For a bar graph, for each level of the first independent variable, plot a bar for each level of the second independent variable so that these bars are touching each other in a cluster. Do not connect the clusters.

    ❹ Add a second independent variable to each of the graphs in ❸.
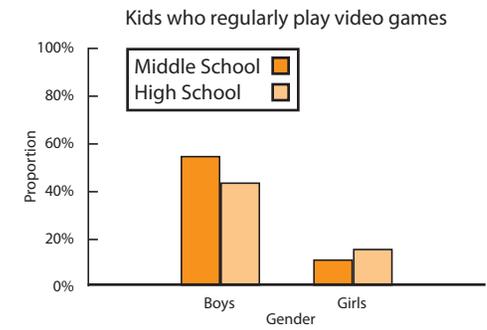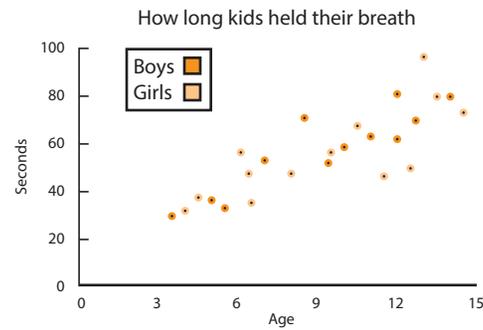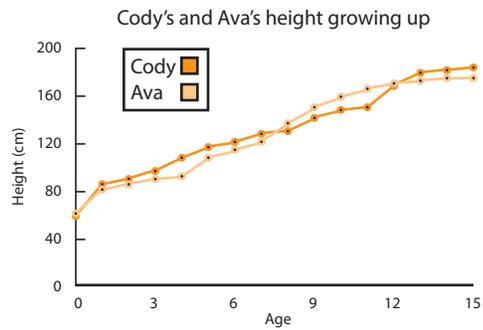
2-B   Histograms

A DISTRIBUTION states each possible category and how common each is.

A HISTOGRAM shows a distribution in which the categories are a series of connected numerical values or ranges called CLASSES. The x-axis is a numerical scale, and each bar represents one class. (In this chapter, only classes that are ranges, such as 0-10 and 10-20, will be used. In later chapters, discrete values, such as 0 and 1, will also be used.)

A Frequency Histogram is a histogram in which the y-axis represents the number of data values in each category.

❶   Make a frequency histogram.
     1. Decide on classes. They must all be the same size with no space between them.
     2. If the categories are values, label the bars. If the categories are ranges, label the boundaries of the bars.
     3. Count how many data points fall into each class. If a data point falls on a boundary, include it in the higher class.
     4. Draw and scale the y-axis, starting at zero.
     5. Draw a bar for each class representing how many data are in the class.
     6. Label the axes and title the graph.
     ❶ Make a frequency histogram for the following chapter three homework scores: 0, 11, 12, 16, 25, 25, 29, 30, 31, 34, 35, 35, 38, 38, 38, 39, 41, 41, 41, 41, 42, 42, 42, 43, 43, 43, 44, 45, 45, 45.

A RELATIVE Frequency Histogram is a histogram in which the y-axis represents proportion of the total rather than raw frequencies. Other than the scale and label on the y-axis, it is exactly the same as a frequency histogram.

❷   Make a relative frequency histogram.
     1. Do steps 1-3 for a frequency histogram (see ❶).
     4. Count the data to get the sample size n.
     5. Find the proportion of data in each class by dividing each class frequency by n.
     6. Draw the y-axis and scale it using percentages or decimals, starting with 0 and ending with a value equal to or slightly higher than the highest proportion in step 4.
     7. Draw a bar for each class representing the proportions in step 4.
     8. Label the axes and title the graph.
     ❷ Make a relative frequency histogram for the homework data in ❶.
     4. $n = 30$
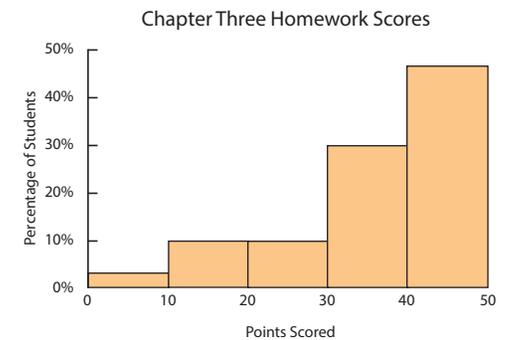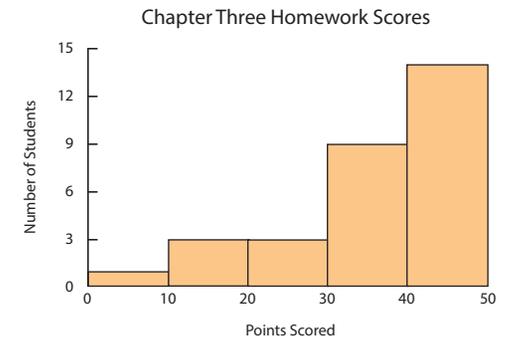   5. $\frac{1}{30} \approx 3\%$ $\frac{3}{30} = 10\%$ $\frac{9}{30} = 30\%$ $\frac{14}{30} \approx 47\%$

In some histograms, the bars are labeled instead of the boundaries. This is appropriate for discrete classes, which will be seen in later chapters.

A distribution can be various shapes. Some general shapes are of particular importance.

A UNIFORM Distribution has the same frequency of data in every class.

A NORMAL Distribution is symmetrical and mound-shaped.

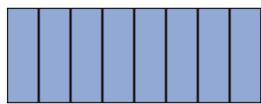A BIMODAL Distribution has two classes that are each higher than those around them.

A SKEWED Distribution has one or more data on only one side that are far away from the mean. The side (right or left) with the extreme data is the direction of the skew.

❸   Describe the shape of a distribution graphed in a histogram.

1. If all the bars are equal height, it is a uniform distribution.

2. If there are two separate peaks, it is a bimodal distribution.

3. If it is somewhat symmetrical and mound-shaped, it is approximately a normal distribution.

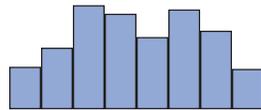4. If there is a longer tail on one side, it is skewed in that direction.
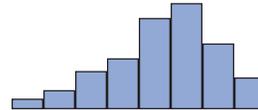
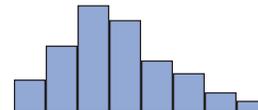❸   Describe the following distributions.

a)   b)   c)   d)   e)
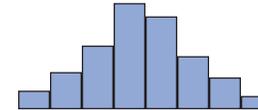
uniform   bimodal   skewed left   skewed right   approximately normal