AP Statistics
Chapter 1 Data Analysis

What's the difference between categorical and quantitative variables?

For quantitative variables, what is the difference between a discrete and a continuous variable?

What is a distribution?

What is the difference between a frequency table, and a relative frequency table? When is it better to use relative frequency?

What is a two-way table? What is a marginal relative frequency?

What is a joint relative frequency?

What is a conditional relative frequency?

The Pew Research Center asked a random sample of 2024 adult cell phone owners from the United States which type of cell phone they own: iPhone, Android, or other (including non-smart phones). Here are the results, broken down by age category:

(a) What proportion of the sample use an iPhone?

	18–34	35–54	55+	Total
iPhone	169	171	127	467
Android	214	189	100	503
Other	134	277	643	1054
Total	517	637	870	2024

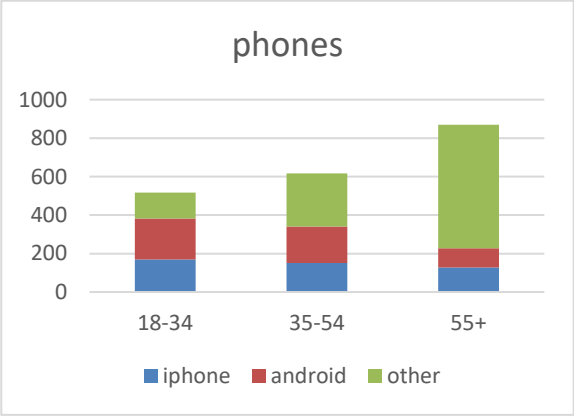
(b) What proportion of the sample use an iPhone and are 55+?

(c) What proportion of the 55+ people in the sample use an iPhone?

(d) What proportion of the iPhone users in the sample are 55+?

What does it mean for two variables to have an association?

How can you "see" an association between two categorical variables?

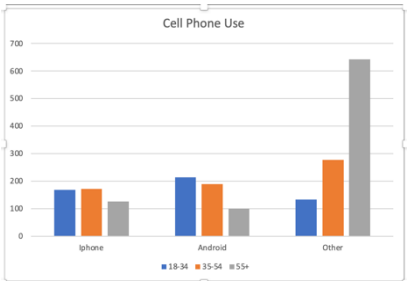


(e) Explain what it would mean if there was no association between age and cell phone type.

Some Definitions

DEFINITION Side-by side bar graph, Segmented bar graph, Mosaic plot

A **side-by-side bar graph** displays the distribution of a categorical variable for each value of another categorical variable. The bars are grouped together based on the values of one of the categorical variables and placed side by side.



A **segmented bar graph** displays the distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.

A **mosaic plot** is a modified segmented bar graph in which the width of each rectangle is proportional to the number of individuals in the corresponding category.

Briefly illustrate the following distribution shapes:

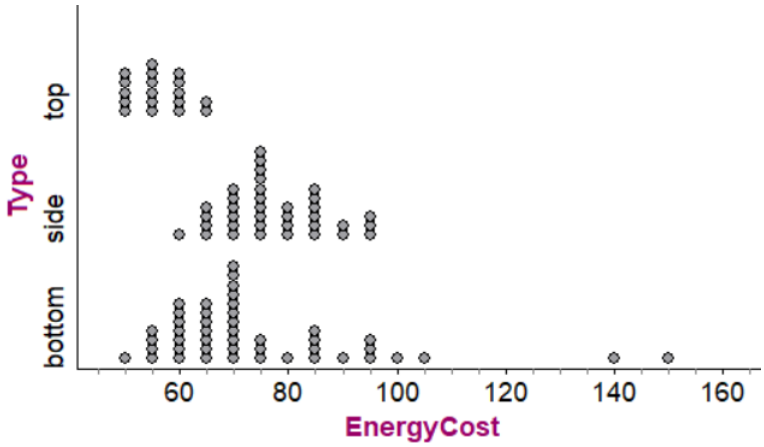
Symmetric | Skewed right | Skewed left

Unimodal (Single-peaked) | Bimodal (Double-peaked) | Uniform (no peaks)

How do you describe a distribution of a quantitative variable?

What are the 2 most important things to remember when you are asked to compare distributions?

How do the annual energy costs (in dollars) compare for refrigerators with top freezers, side freezers, and bottom freezers? The data below is from the May 2010 issue of *Consumer Reports*. Compare these distributions.



What is the most important thing to remember when making a stemplot?

A sample of 14-year-olds from the United Kingdom was randomly selected. Here are the heights of the students (in cm). Make a back-to-back stemplot and compare the distributions.

Male: 154, 157, 187, 163, 167, 159, 169, 162, 176, 177, 151, 175, 174, 165, 165, 183, 180
Female: 160, 169, 152, 167, 164, 163, 160, 163, 169, 157, 158, 153, 161, 165, 165, 159, 168, 153, 166, 158, 158, 166

The following table presents the PH (a scale on which a value of 7 is neutral and values below 7 are acidic) of water collected from rain and snow over a 6-month period in Allegheny County, PA. Make a dotplot to display the distribution of PH. Then, make a histogram. We will use technology to do this.

4.57	5.62	4.12	5.29	4.64	4.31	4.30	4.39	4.45	5.67
4.39	4.52	4.26	4.26	4.40	5.78	4.73	4.56	5.08	4.41
4.12	5.51	4.82	4.63	4.29	4.60				

How is a histogram different than a bar chart?

Why would we prefer a *relative* frequency histogram to a frequency histogram?

What is the difference between a statistic and a parameter?

What is a resistant measure? Is the mean a resistant measure of center?

How can you estimate the mean of a histogram or dotplot?

Is the median a resistant measure of center? Explain.

How do skewness and outliers affect the relationship between the mean and the median?

What is the range?

What are two problems with range as a measure of variability?

What does the standard deviation measure?

How do you calculate the standard deviation for a population? What about the variance?

How do you calculate the standard deviation for a sample?

What are some properties of the standard deviation?

A random sample of 5 students was asked how many minutes they spent doing HW the previous night. Here are their responses (in minutes): 0, 25, 30, 60, 90. Calculate and interpret the standard deviation.

What are quartiles?

What is the interquartile range (*IQR*)? Is the *IQR* a resistant measure of variability?

The table shows the number of runs the Cubs allowed to score during day games in two different types of weather. For each distribution, calculate the *IQR*.

Cloudy: 0 1 1 2 3 3 3 3 3 3 4 4 4 4 4 4 5 6 6 6 6 9 9 10 11 13 14

Sunny: 0 0 1 2 2 2 3 3 3 4 5 5 5 5 8 11 12 15

What is an outlier? How do you identify them? What is the *IQR* dance?

Are there any outliers in the runs allowed distributions from above? Justify.

What is the five-number summary? How is it displayed?

Draw parallel boxplots for Cubs cloudy/sunny data. Compare these distributions.

What are some weaknesses of boxplots?