Questions Module 2.3

What are some possible outcomes of exploratory data analysis?

- *a.* finding anomalies in your data
- *b.* seeing patterns in your data
- *c.* identifying potential relationships between variables
- *d.* generating interesting questions
- *e.* determining that you don't have enough data or the right data to answer your question
- *f.* all of the above

Incorrect.
The correct answer is **f**. We didn't talk about choice **e**, but this is a common outcome of exploratory data analysis.

A Pareto plot is not a bar chart.

- *a.* True
- *b.* False

Incorrect.
The correct answer is **b**. A Pareto plot is a sorted bar chart with a cumulative percent curve for identifying the top issues, problems, or opportunities.

In a problem-solving context, the problem that occurs most frequently is the one that should be solved first.

- *a.* True
- *b.* False

Incorrect.
The correct answer is **b**. It depends. Some problems might be costlier, more critical (for example, due to safety), easier to solve, or more important to the business or the customer (or any combination of these things). These factors must be considered when determining which problems to solve first.

You can use a packed bar chart to graph a variable that has hundreds of categories (or more).

- *a.* True
- *b.* False

Incorrect.
The correct answer is **a**. If you have a lot of categories and the distribution of frequencies for the categories is skewed, a packed bar chart can help you identify the top categories. If the frequency of occurrence is the same across all categories, the packed bar chart might not be very informative.

When you want to graphically display the relationship between two categorical variables, you should always use a mosaic plot.

- *a.* True
- *b.* False

Incorrect.
The correct answer is **b**. It depends on your data. If you have many levels of one or both categorical variables, you can create a mosaic plot, but it can be difficult to interpret. In this situation, a tree map might be better.

Which of the following dimensions can you add to a bubble plot?

- *a.* color
- *b.* size
- *c.* time
- *d.* movement
- *e.* all of the above

Incorrect.
The correct answer is **e**. You can add all of these dimensions to a bubble plot.

Which of the following can be used to create a geographic map?

- *a.* Zip codes or country codes
- *b.* city names
- *c.* latitudes and longitudes
- *d.* X and Y coordinates that map to positions on a part
- *e.* all of the above

Incorrect.
The correct answer is **e**. All of these can be used to create a geographic map.

It's time for you to apply what you've learned in this module thus far.

Consider this scenario. The file **Measles.jmp** in the course data folder includes information on the reported cases of the measles for each state in the United States per 100,000 residents since 1928. The data are from the 2010 US Census. The measles MMR vaccine (MMR – Measles, Mumps and Rubella) program was introduced in 1963. Before 1963, an estimated 3 to 4 million people per year got the measles.

Use these data and the exploratory tools you have learned in this module to answer the following question. Note that you are not limited to the exploratory tools that you learned in this module. There are many ways to use these data to answer this question, and we encourage you to explore both the data and the potential tools.

Was the measles vaccine program effective at reducing the rate of measles infections across all states?

One effective tool for visualizing these data is a heat map, with **State** plotted on the **Y** axis, **Year** plotted on the **X** axis, and the cells colored by **Yearly Incidence**. **Hint**: The answer is yes, the program was effective.

We revisit this scenario, and evaluate the effectiveness of different exploratory tools, in the next lesson.