

## Evaluating the Effectiveness of a Visualization

This heat map was created to show the impact of the measles vaccine.

To evaluate the effectiveness of this visualization, first you should ask, "What is the practical question we are attempting to answer with this graph?"

The question should be well defined, and it should be aligned with your research question. The question should also be interesting or compelling. That is, there should be a good reason for creating the graph to begin with.

Let's say, for this discussion, that this is the question we are attempting to answer:

Was the rate of measles infections reduced, across all states, after the measles vaccine program started?

Is this question clear and well defined?

Note that this is a lot different from simply asking, "Was the measles program effective?"

This statement is too vague. How will you determine whether the program was effective? What is the operational definition for the word "effective?"

If you don't have a clear and concise statement of the practical question of interest, you won't collect the right data, and you won't be able to answer the question that you're really interested in.

For the purpose of this video, let's assume that our question is well defined.

For information about developing operational definitions, see the discussion in the first module of this course.

Next, you should ask, Do you have the right data?

The data you compile should be aligned with the practical question you are addressing. That is, the data should enable you to answer your question. If you can't answer the question with your data, then you don't have the right data.

Recall that our question is: Was the rate of measles infections reduced, across all states, after the measles vaccine program started?

To answer this question, you need to have measles data over time for all states.

What if the question was worded in terms of prevention or eradication of measles? You might use different data, different analyses, and different visualizations to answer this question.

Note that compiling the "right" data is a lot harder than it sounds. For this example, state-by-state data are available. But these data are aggregated and have been pre-summarized. This can limit how you use the data and the types of conclusions you can draw.

For example, you can't dig deeper into the question to understand the measles rate for urban versus rural populations, or for different age or demographic profiles.

Assuming that you DO have the right data, you might use a variety of exploratory methods to investigate your question. The types of methods you use depend on the question that you are asking and the types of variables that you have. You see examples of the different exploratory methods you might use to answer the measles question in the upcoming videos.

As you explore the data, you might identify a number of issues with the data itself. For example, you might need additional data, the data might not be in the right format for analysis, or the data might be messy. For example, you might be missing a lot of values or you might have typographical errors.

You learn about steps that you might need to take to prepare data for analysis in the last lesson of this module.

The third question you should ask when evaluating visualization is, Does the graph address your question in a clear and concise way; does it represent and accurately communicate what the data are telling you?"

The heat map shows data for all states, over time, colored by the number of reported cases per 100,000 residents. A reference line is drawn at 1963. It certainly looks like the measles rate dropped substantially, and this reduction took place shortly after 1963.

Alberto Cairo, a visualization specialist, professor, consultant, and author, has a unique perspective on interpreting visualizations.

Alberto warns that what you design is often not what people see, because we tend to project the things we want to see in our visualizations. We impose our own narratives, our own perspectives, but there can be other narratives that are just as reasonable.

According to Alberto, "A chart shows only what it shows, and nothing else."

You need to make sure that the graph does indeed represent what the data are telling you, and that you're not reading more into the graph than it actually says.

In the upcoming videos, we look at different visualizations for the Measles data. We evaluate the effectiveness of each of these visualizations, before ultimately evaluating the heat map.