Print

Designing an Effective Visualization: Part 1

For the Measles data, here is the practical question we are attempting to answer:

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

This is a high-dimension problem. There are data from all 50 states plus Washington D.C., and these data are across several years. In your initial investigations, you might explore many different graphs in order to answer your question.

The most basic graph for these data is a simple bar chart. Here's a bar chart for the average number of cases reported annually per 100,000 before and after 1963.

For this graph, we created a new variable, Vaccine B4/After. You learn how to create this variable in the "Data Preparation" lesson at the end of this module.

The graph shows that the overall incidence rate dropped substantially after 1963. This answers part of the question, "Was the measles rate reduced?"

But you lose information on how the measles rate is changing over time, from year to year.

More importantly, the data for all states are combined, so this graph doesn't address a key part of the question: across all states.

Here's another bar chart. This graph shows the annual cases reported over time. A reference line was added at 1963. This graph shows much more information than the simple bar chart.

From this graph, you learn that the reported incidence rate dropped after 1963. However, there was a bit of a transition. The incidence rate didn't drop entirely after the vaccine program was implemented. It took a few years.

This visualization tells you about the overall infection rate over time.

But, like the first bar chart, it doesn't tell you whether the infection rate dropped across all states.

Because you have time-ordered data from many years, you can use run charts to see whether the incidence rate drops for each state after 1963.

Run charts for three of the states are shown here. You can see a decline for these states shortly after 1963.

There are 50 states plus Washington D.C., so you need 51 individual graphs to see the data for all of the states. Viewers are unlikely to want to scroll down through a long series of run charts. Instead, you can use a trellis plot to create a matrix of run charts.

In this trellis plot, you can see run charts for all states across all years. The Y-axis scale is the same for all states, so you can see which states had a higher infection rate before 1963. For example, look at Utah and Vermont!

You can also see that some of the states, like Louisiana and Mississippi, had very low infection rates all along.

Because you can see the data for all states and all years, you can see the overall pattern in measles infections across the states. The infection rate approached zero for most states around 1970, shortly after the measles vaccine program started.

Is this a good visualization? It might answer our question, but do you find it easy to understand what it's telling you?

You have to look through and interpret 51 individual graphs. These graphs, and their axes, are very small. This graph is a bit hard to read. It's hard on the eyes, but it's also hard for our minds to make sense of so much information.

In the next video, we consider some other ways of graphing these data.

Close