Print

Designing an Effective Visualization: Part 2

For the Measles data, bar charts, run charts, and trellis plots tell part of the story, but they aren't the most effective visualizations in answering our question.

We have data for all of the states, so a geographic map is an option.

To account for the transition period evident in the previous graphs, the variable Vaccine was created. This variable is coded as follows: Before includes all of the data prior to 1963, Transition is all of the data from 1963 to 1966, and After is all of the data from 1967 to 2011.

Here, each state is colored by the yearly incidence per 100,000. By default, the states are colored using a diverging red-to-blue scale. The reddest states had the highest incidence rates, and the darkest blue states had the lowest.

You can see that all of the states in the After graph are dark blue. This means that all of the states, after the transition period, had very low incidence rates.

Diverging color scales can be difficult to interpret. You have to rely on the legend to make sense of the colors.

Here, a sequential white-to-blue scale is used instead. This sequential scale can be easier to interpret than a diverging scale. The darker the color, the higher the infection rate. You don't have to look at the legend to compare the rates for the different states.

You can also see that data for Kansas, in the transition period, is missing. You couldn't see this with the diverging scale, because the middle of the scale was gray.

This geographic map shows that the incidence rate after the transition period was much lower than before, and this seems to be true for all states.

However, a lot of information is buried in the maps. For example, each graph is an average of data over several years, so you can't see the year-to-year variation within a state.

You can create a map where the data are broken down into blocks of years to show more of the variation over time.

This graph gives you a sense of the overall pattern of measles infections over time, across the blocks of years. But is this the most effective visualization to address our question?

There is one core problem with this visualization: You can't see the very small states, like Rhode Island and Connecticut. So you can't see what is happening to the measles rates in these states.

It's also difficult to directly compare states from one map to the next. For example, does the measles rate change for Florida across the different time blocks? It's hard to tell.

Let's return to our original graph, the heat map. A heat map is an efficient visualization for these data, enabling you to see the measles rate for all states and all years. The darker the color, the higher the measles rate.

Adding a reference line for the vaccine at 1963 enables you to easily see the infection rate, both before and after the measles vaccine program started.

You can see states that were really bad before the program, and states that weren't that bad. You can also see states that are missing values for different years these cells are shaded gray.

Adding an annotation, along with a second reference line marking the end of the transition period in 1966, makes this graph much more informative.

Let's return to the question at hand. This heat map clearly communicates, in one graph, that the measles rate decreased dramatically for all states with the introduction of the measles vaccine program. Although, you can also see that some states were already very low.

But does the heat map provide too much information? Do you need to see the information for all states for all years? Would a simpler visualization suffice?

One element of this question, which we've completely ignored until now, is the intended audience. For a non-technical audience, the heat map might be overwhelming or difficult to interpret. The simple geographic map showing the three phases might suffice.

When you add annotations to this visualization, change the layout, and provide more descriptive labels, this geographic map might be more effective than the heat map for answering the question. This horizontal layout is particularly helpful. It conveys a sequence or chronology to the graphs, so you don't have to work as hard to interpret the visualization.

Ultimately, the effectiveness of your visualization in communicating your message is determined by your audience.

It's easy to get excited when you create a beautiful or fancy visualization, but the best graph is often the one that is the easiest for your audience to understand. If your audience can't see the message, then the visualization isn't effective.

Note that, throughout this lesson, we use graphical methods that were introduced in previous lessons. For information on how to create these graphs, use the Search tool in the side panel. You learn how to add animations and how to customize your graphs in upcoming videos.

Close