Simplifying Your Data Visualizations

By Braden Tabisula, MBA, RHIA, CHDA

Editor's Note: This is the first installment of the new Advancing Analytics department, which will feature articles that provide best practices and practical resources for HIM professionals who are involved in or pursuing data analytics roles in healthcare.

Have you ever sat down at your desk with a cup of hot coffee, opened up your email inbox, clicked on an attachment, and—lo and behold—it is another document with some colorful and overly detailed charts?

They may have been designed in Excel, or, if you are hip and your organization has invested some money in visualization tools, you could be staring at a beautiful dashboard with fancy graphics designed in Tableau or Power BI. Regardless of the tool used, have you ever wondered, "What the heck am I looking at?"

It is commendable if this has never happened to you. Maybe you are the designer of these visualizations, in which case a lack of confusion would make sense. However, many healthcare professionals have had those instances where they really could not tell what they were looking at, at first glance, when it comes to data analytics visualizations.

It is one thing when you look at an article or an emailed document and have the time to sit with it for a while to digest the visualization. However, it really is inefficient when you are presenting your visualization, live, and you are pressed for time because of an already-packed agenda. There is no time to waste; you need to get your message across—and quickly. This article addresses a few guidelines that will help your audience interpret your data analytics visualizations quickly and avoid the deer-in-the-headlights look when you present your visualized data.

What are Visualizations?

When it comes to discussing basic visualizations, there are a few to choose from: bar charts, column charts, line graphs, scatter plots, and the dreaded pie chart—just to name a few. If no one has told you by now to avoid using pie charts at all costs, please let this be the one tidbit you take away from your reading today.

Pie charts are looked down upon in the data analytics world because when there are more than two variables to graph, the pie "slices" are not easily interpretable by the human eye. Say, for example, you drop the values of 33 percent, 30 percent, and 37 percent in a pie chart. You will get very similar-looking angles. Now, instead, try putting this data into a bar or column chart, and viewers can easily see the "highest" column or "longest" bar. So, unless you have just two variables to put into a graph, and they are extremely different values, then please use another means of visualizing your data. Another extreme case of the pie chart is when you have 20 categories represented in the pie as tiny little slivers—so tiny that you would not be very happy after Thanksgiving dinner if that size pie was served. Consumers of this type of pie chart would also be left unsatisfied.

Another distinction that should be addressed during this introduction is between bar and column charts. You may think they are the same, and really, they are—with only one major difference. A column chart will have vertical stacks to display the data, while a bar chart will have horizontal stacks. There are advantages to using one or the other when creating visualizations, which will be explained below.

Sometimes visualizations can also be very effective with nothing more than plain old text. When providing a statistic, use size to distinguish the percentage you are reporting, and provide what that statistic is representing. A passage like: "25 percent of all patients…" could be very powerful when the 25 percent is bolded and has a bigger font than the rest of the text.

Using the Gestalt Principles of Design

A set of principles called the Gestalt Principles should be used when designing a visualization. These principles are based on theories of visual perception first developed by German psychologists in the 1920s, and attempt to explain how people have a tendency to organize visual elements into groups under certain conditions.[1] In reality, when you make a graph, you should think like a designer. The Gestalt Principles are listed below along with how to use each when creating a data visualization.

Proximity

When objects on a chart are close to each other, they speak to the fact that they might be related. Imagine a scatter plot with dots all along the quadrant or quadrants. When you see dots clustered tightly together, they give the mind a clue that these clusters have something in common.

Similarity

Different sizes, shapes, or colors can be used in a graph to have the mind understand that there is a relationship among those of the same size, same color, or same shape. If you had to provide a graphic of training scores across five different departments who took the training assessment, using a different shape or color to represent scores pertaining to a department would be a great use of the similarity principle.

Continuity

This is a powerful principle alone. Continuity gives you control of where you want the viewer to look. Connecting two points with a line shows your audience how they are related. Apart from that, don't be afraid to insert an arrow to point out for the audience the data that is most important.

Closure

Another simple way to encapsulate someone's attention would be to draw a circle or box around the data points you want your audience to look at. Draw their attention to a group of hospitals in the graphic, for example. Do not be afraid of grabbing your drawing tool, changing the outline to red, and placing a box or circle around the most important part of your visualization.

Figure/Ground

This principle is subtle but powerful when used correctly. You can really make your chart stand out by creating a distinction between what is your background and your foreground. Use contrast wisely so that your audience is not confused by the background or thinks it is part of the data. Do not use colors for the background that are too similar to your bars or dots. This also confuses the audience.

Keep Your Graphs Clean

Let's take a look at some common data elements one should consider adding or removing in a visualization. Lines are not really important in visualizations, unless they are connecting data points. The grid lines on a chart do not provide much help in deciphering the data. You see high points and low points, so let the graph tell you that Group 1 is not like Groups 2 and 3. There are few instances where your reader would trace the data along these lines to argue that a bar is at 71 or 72 on a scale of 100. Those are minute details that the mind assumes are the same. Now, if a one- or two-point difference is a big deal then certainly point it out. Generally though, lines become more distracting than helpful.

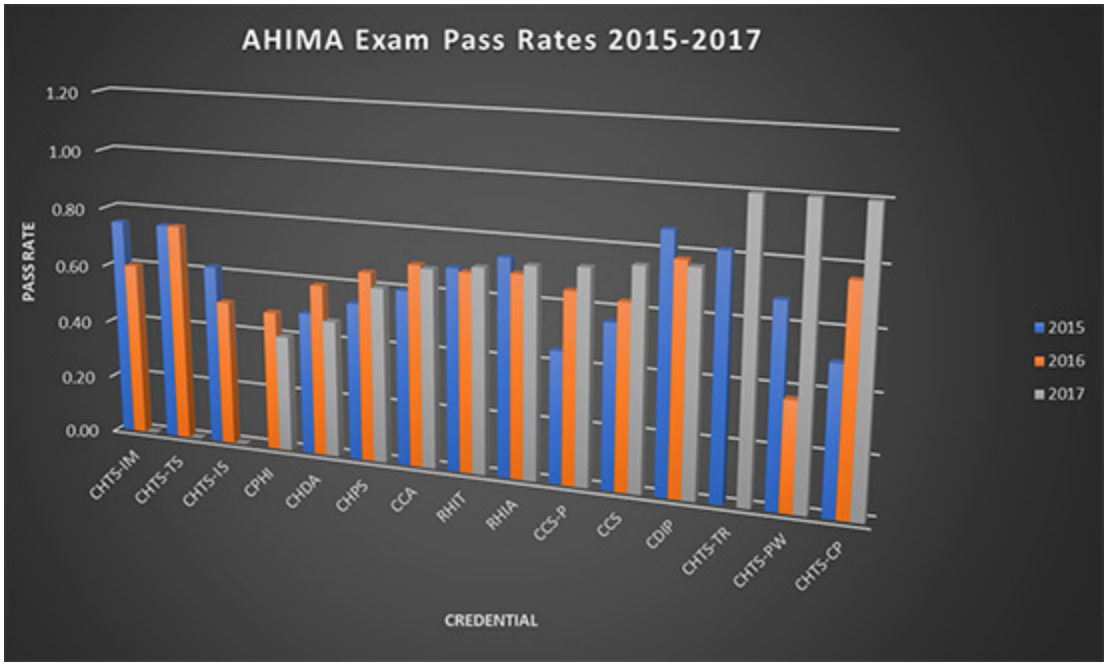
Remove any background color and generic borders that come with the charts—they do not need to be there. Background colors other than white disrupt the interpretation of the data being presented. The same goes for chart borders. You also do not need data markers and labels on every data point in a line graph, or at the top of every column or bar. Use these when they are appropriate to point out the precise value. Otherwise this is distracting, especially if you have the axes on a chart.

Pay close attention to the labels for your axes. Utilize a horizontal bar chart when you have group titles or names. It is odd when you are looking at a column chart and the names are at the bottom of the graph at the 45 degree angle. Do not make

your audience crane their necks to read. Instead, flip the chart on its side, create a horizontal bar chart and have the names of the groups along the left which can be easily read without strain.

Legends are typically placed off to the side or at the bottom of charts. It is a bit difficult to go back and forth from the legend to the data. Instead, place the data label next to the data, and make them the same color. So, if the trend line for your inpatient census is orange, create a label "Inpatient" with the font color set to orange and place it at the end of the line. Do the same for your other trend line and associate the color of the line with the same color of the label.

Do Not Present the Data This Way

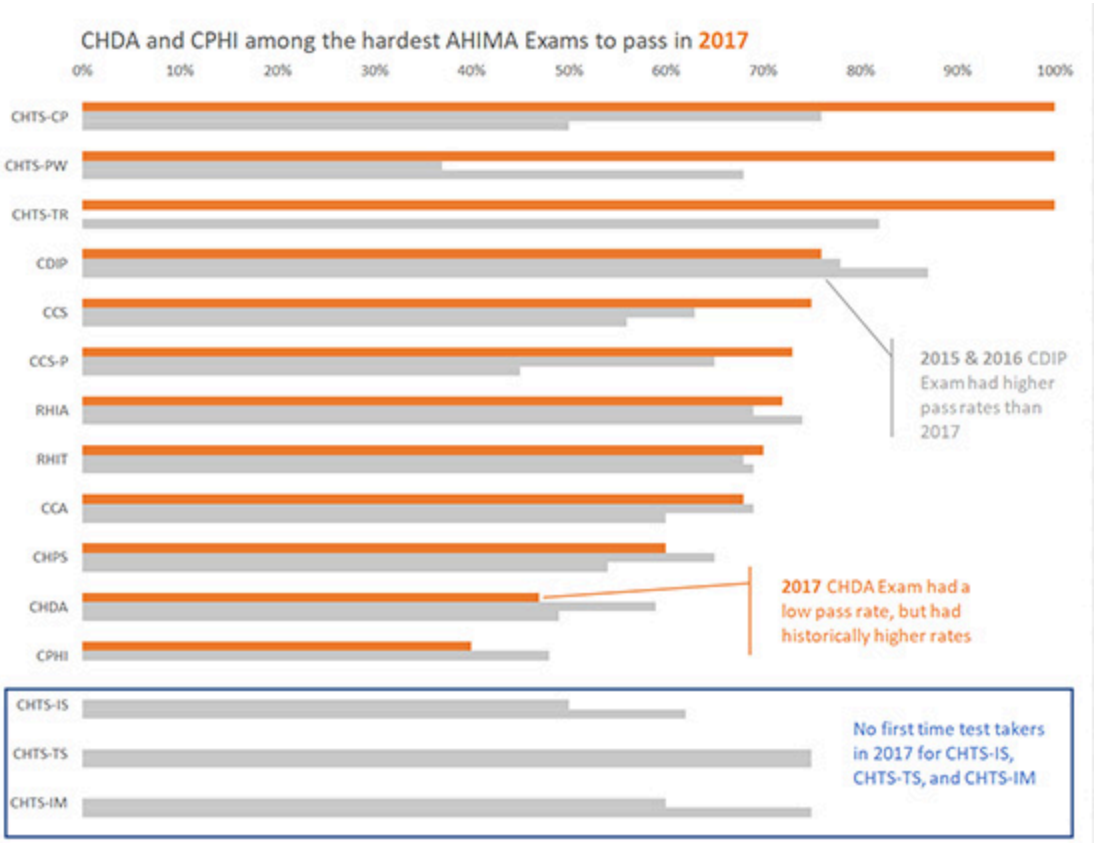


Here is a list of what's wrong with this chart. When creating a data visualization, do not:

1. Use the 3D chart layout—it makes the graph difficult to interpret due to the depth of the bars.
2. Put a black background on the chart—it competes with the data you are trying to present.
3. Leave the 1.20 line for pass rate since the data does not have a value greater than 100 percent.
4. Use too many colors; this chart does and its bars have competing colors.
5. Use gridlines when they can be avoided; they are cause for further distraction from the data and are not that meaningful.
6. Leave angular axis labels, like the credential labels here; they are very difficult to read.
7. Leave decimals points when you want to present percentages; in this case the Pass Rate is formatted incorrectly (i.e., "0.40" should be "40%").
8. Leave axes unorganized—the pass rates are in order, so why is there no order or organization to the credentials?

Source: AHIMA. "Certification Activity Exam Pass Rates." February 13, 2018. www.ahima.org/certification/chda.

Do Present the Data This Way



Using the same data as the "do not" chart, this chart does follow visualization best practices. When creating a data visualization, do:

1. Flip the axes when you have categorical data labels like credentials so they can be read without neck strain.
2. Remove the three different colors from the bars to highlight 2017 only.
3. Decide what you want to focus on. Using the similarity principle, this chart uses orange for 2017 in the title as well as for the bars related to 2017. The bars now focus the reader's attention to that year while still leaving the other bars available for interpretation.
4. Change Pass Rate to percentage and remove trailing zeros.
5. Use the proximity principle on the bars so that the reader understands that closer bars (either orange or gray) relate to a credential.
6. Utilize the continuity principle; even without the Y-axis line the reader can still see a continuous 0% starting point for all bars.
7. Enclose data that are related; CHTS-IS, CHTS-TS, and CHTS-IM are enclosed to highlight that there were no test takers in 2017.
8. Take advantage of white space and add notes on interesting observations—this requires less cognitive processing from readers.
9. Remove labels when appropriate; in this case, readers understand that labels on the left are credentials through AHIMA.
10. Order your axes; the 2017 rates are organized from highest to lowest pass rate, or could also organize alphabetically.
11. Move the X-axis to the top so the reader immediately understands what the bars represent instead of looking for the axis at the bottom.

Source: AHIMA. "Certification Activity Exam Pass Rates." February 13, 2018. www.ahima.org/certification/chda.

Visualizations Turn Data into Information

There is so much to learn about data visualizations, and this article just scratches the surface. A great book for all data analysts and data visualizers out there is *Storytelling with Data* by Cole Nussbaumer Knaflic, a former data analyst with Google.[2] She provides so much insight and is an expert in visualization.

Overall, visualizations put the data into depictions that give us an idea of what our data look like for the purpose of comparing and trending. One could argue that you need all the details, and if you do, put the data in a table. Whatever method you choose to present your data, keep these principles and guidelines in mind. Always think about the design of your visualization, and what your audience sees first. You want to have the audience understand what you are showing them sooner rather than later.

Notes

1. Spokane Falls Community College. "The Gestalt Principles." Graphic design tutorials. http://graphicdesign.spokanefalls.edu/tutorials/process/gestaltprinciples/gestaltprinc.htm.
2. Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: John Wiley & Sons (2015).

Braden Tabisula (btabisula@llu.edu) is assistant professor and program director of the health informatics master's program at Loma Linda University.
