

Data Visualization Best Practices Webinar

IAP Data Analytics Track

[Introductory remarks]

Jessie Parker (JP): Many of you on the call are probably familiar with the Medicaid Innovation Accelerator Program, but I'll give a brief overview. It's a four-year commitment by CMS to support ongoing innovation in Medicaid programs through various forms of targeted technical assistance. IAP is funded by CMMI but is led by the Center for Medicaid and CHIP Services. In short, we support delivery system reform efforts with a goal of increasing the number of states moving towards delivery system reform across program priorities. Under the umbrella of working towards Medicaid delivery system reform, IAP is organized across program areas and functional areas. The IAP program areas of focus include: beneficiaries with complex needs and high costs, community innovation through long term services and support, physical and mental health integration and substance use disorders. In the IAP functional areas, which we see as critical levers for delivery system reform include data analytics, quality measurements, performance improvement, and value-based payment modeling and financial simulations.

Today's webinar is part of our data analytic functional area. The main goal of the IAP data analytics work is to expand on and improve data analytic capabilities to support Medicaid delivery system reform and other innovations. One of the ways in which we're supporting this goal is by offering quarterly state webinars, such as this one. Today's webinar will review techniques to improve your data visualization capabilities, which is an area I know is in very high demand and of great interest to many of you. And with that, I'll turn it over to Kathy.

Kathy Rowell (KR): Thank you. So, this is Kathy Rowell, and today what I hope to do is raise awareness about all of the elements that we have found and know to be absolutely crucial to creating great data visualization, dashboard reports and infographics. We've tried to select the high points, and a lot of the things that we see folks making mistakes on, so that we can show you what the best practices are around those.

In today's presentation we'll talk about:

- What we need to create great visual displays.
- The importance of understanding your data's lineage.
- Basic statistics, which, unfortunately, or fortunately, are required for this type of work.
- The lessons learned from visual and cognitive research.
- When we use a table versus when we use a graph to display data.
- A dashboard is a term that gets thrown around a lot, but how do we define it? What does it really mean?
- How do we create great dashboards?
- What are infographics? We talk a lot about them, but what are they?

Then we will chat a tiny bit about technology, because often we find that folks believe that technology alone is the solution, but we know that it isn't, and we want to chat about why.

Today we're going to talk about the visual intelligence and data visualization research and best practices. What do we need to know to create great visual displays? What we know is that it really takes three elements, we represent them here by the three-legged stool.

First, you absolutely have to know health care and basic statistics, so we need subject matter experts to understand the domain in which they're working to be able to communicate data and information.

Certainly, on the far right, we need great technology, which is available to us, to create really compelling data displays.

In the middle, we need to understand data visualization and visual intelligence. The middle piece is not intuitive; it's something that needs to be learned just like understanding statistics and health care, and understanding the technology that we work with. So, there is formal learning and information out there, and that's what we're going to talk about, and start on the path to today.

In the first section, Health, Healthcare and Basic Statistics, what we're interested in first is that folks understand their data lineage. What do we mean by that?

- The intent and purpose of the data.
- Why it was established.
- Why it was created.
- Why we couldn't correct it.
- How it may have changed over time.
- What we can and can't do with it.

Just because we have data doesn't mean that it always tells the story we need it to tell.

For example, in 1982, some of you may not have been born yet, but some of you may remember like I do, that the third-party payers all came together and agreed that they would create a universal bill for submission of claims data, so that insurance claims could be paid. The intent and purpose of this data was just that. It was an alignment with getting an insurance claim paid. What we see in this slide has been used for research. For example, the Dartmouth Atlas, with which you may be familiar, has used a lot of this data and information.

But it's not so great if we wanted to study clinical outcomes and clinical care. For that, we may have things like the Society of Thoracic Surgeons' Cardiac Data Base, which is the gold standard of risk adjusted clinical outcomes, designed by cardiac surgeons and cardiac specialists to capture data specific to those types of procedures and outcomes of interest or ronds(?) [7:58]. It's designed specifically for that focus. That could be where we would get some great clinical data and information, as compared to administrative data, which was designed to do something entirely different.

Then, as a lot of you work with survey data, we also have survey data. Survey data is designed to get at specific questions. However, one of its disadvantages is that what you can learn and know may depend on who's answering the survey, in regard to the questions and information of interest.

These are just three quick examples to show that it's not enough to say that I've got data; we really encourage people to understand why that data was established, why was it being captured, and what can it be used for. Because then it also helps inform how you can display it. We'll talk more about that.

We also of course want to think about the quality of the data you're working with.

- How is it defined?
- How is it documented?
- Is it timely, or is it so old that it's really not of much use?
- Is it complete? Are there data elements that you'd really like to use but not enough folks have reported them or there are not enough of them in the database to really be able to use them?
- And then a big one is, is this data ever audited? Is anyone ever going out and checking that folks are submitting the data in alignment with the data definitions, or the submission requirements?

These are just a few of the things you want to consider before you start to think about how to display and communicate your data.

How good is your data and what is its lineage? Because you really have to understand, why and for what purpose data is being captured, along with the quality so that you can understand what it can and cannot deliver when you're trying to create reports, dashboards, and infographics.

Basic statistics are required, and folks may say they've got a team of folks who understand all the statistics. But we find a couple of things: One, sometimes statistics are not used correctly, or they may be used correctly but they're displayed in a way that makes the story unclear. And two, we've got to be able to think about the new claim language in easy terms so that we can communicate them.

Let's look at a very simple example. If I were to ask, which would you use to describe this collective group's annual income, the mean or the median? You might take a look at this photo collage, and say well, on the diagonal you've got Warren Buffet and Bill Gates, we've got Oprah in the middle, and Mark Zuckerberg in the lower right-hand corner. They're all billionaires. Around them, we've got the teachers and the data analysts, and the janitors and the fire fighters, the rest of us, who are represented as the average income.

So, in this case we want to use the median because that helps us to hold out the outliers. If we were to use the average of these incomes, we all would look like we were multi-millionaires, and we might not be on this webinar. It seems like a self-evident thing, but let's take a few looks at how this might go awry.

Here's the medical patient's average length of stay, we've lined up all their days in the hospital, and with that divided we've got an average of 4.2 days average length of stay. But when we look at the median we see we've got some outliers here, some really long lengths of stay, so when we look at the middle most number, we see that it's three. So most frequently, we would observe a length of stay of about three.

I put this visual up, the Surgical Care Outcomes Assessment Program in the State of Washington, who's done some terrific work around surgical care, and there is displayed, a couple of graphs to provide information about some work they've done to reduce lengths of stay around colon resection surgery and gastric bypass surgery. When you first look at these you say, well, it looks like they both started at about the same length of stay, and they're coming down a little bit. I can see that in both graphs. You really wouldn't look much beyond that. But when you look closer, we can see that the data being displayed on the left hand side is the average length of stay, and the right-hand graph is the median length of stay. Additionally, we can see that in the colon resection graph, the y-axis starts at five, and on the gastric bypass graph it starts at zero. So what we can see when we look that closely is (A), we're looking at two different statistics but we put them next to one another.

Perhaps in the median colon resection length of stay we're using an average but perhaps you have some outliers, but we never would pick up all of that simply by looking at the display and not examining it in great depth. We would just say well, it looked like they're both the same and they're coming down about the same rate. We'll talk more about this slide, but these are some of the things we see happen all the time. So if we're going to put these two things next to each other we should probably use the same statistics and think about our axes. We'll tell you more about how this all works, but this is an example where we have two statistics telling us two potentially very different things.

The other thing that we always want to think about is compared with what every single thing we do in data analysis and data visualization is compared with what? If we fail to do this we really haven't

told the whole story. So if we were just to look at 10,000 cancer deaths or 25,000 surgeries, thousands for beneficiaries, immunizations, we'd say okay, those are numbers, but I don't really know what it's telling me about anything. It just lays flat on the page. Every single thing that we do in data analysis and data visualization has to have context. It needs to be compared with what. Very often we just see numbers like this put on a page. So some of the comparisons we want to think about are compared with a budget or a target or expected or similar programs that we might be working on. How are we doing this year compared to a previous year? What do our competitors look like? What is a benchmark in another state? Every single thing that we do is compared with what. Because if the viewers of your dashboards and reports' infographics can't answer so what, you need to go back to the drawing board, and comparisons in context help us understand the so what. Is it bigger, smaller, etc.? So it's always compared with what.

Let's chat about data visualization and visual intelligence. We talked about understanding your source data, history lineage, how good it is. Is it complete? Is it ever audited? Is it timely? All those things. We talked a tiny bit about __[00:15:50] statistics and really kicking the tires on using the right one, and how are we putting them on the page. So now let's talk about data visualization and visual intelligence. Some lessons learned from cognitive research and visual intelligence research. This is where we want to pause and say data visualization is not about an aesthetic, like perhaps graphic design. Data visualization is really based on hard research that happens and continues to happen about how our eyes work, how our brains work, how we take things in, how we turn it all around, what it is that we're really reading that we're seeing and understanding. And if we can understand how well that works, we'll create better visualizations so folks can see the message in our data.

One thing we know from visual research is that 70% of the way we take in all data and information in our world is through our eyes. Thirty percent is all our other senses – taste, smell, hearing, touch, but 70% of it is through our eyes. So we need to understand how our eyes work. Another thing we've learned is that we are pattern seekers as humans and there are gestalt principles of visual perception, which is a fancy way of saying how do we see, understand and seek patterns? These are really important, especially when we get to dashboards. As a summary, there is proximity so objects close together are perceived as a group. There is similarity; if they have similar attributes we perceive things as being in a group. If they are enclosed with some sort of boundary like a square or box or some line around them, we perceive items or data as a group. Closed or even open structures like parentheses make us understand this is grouped together. Continuity. If things are aligned together or appear to be continuation, we perceive them as a group and connection. So if we have dots on a graph and then we connect them with a line, we perceive that as a trend. That would be because of connection.

For example, here's proximity. So these are just nine blue circles on a page, but because we grouped them in threes on different parts of the page we perceive them as a group because they're near each other. Again this gets important when we think about dashboards and how we organize things because we want to put things that make sense in context together in different parts of our report.

Then there's continuity, so even though these three boxes are entirely separate, we can see that they are completely separate, surrounded by a box. We perceive that blue line as being connected in some way and going upwards, even though they really aren't connected. So if we think back to this graph we were looking at previously, even though there's no connection, we have placed them one next to the other, so proximity, we think these things go together. They're both surgeries, right? And continuity, we read left to right. We sort of miss all those nuances like the scale and what statistics are being displayed in the label. This is because of the way that we see, that we group things, and read left to right. We want to be aware of all these principles of how we pattern seek as well. There's lots of information and resources about this if you're curious to read more.

The other thing we learned from a lot of this is pre-attentive processing. That's the ability of a low level human visual system to rapidly identify certain basic visual properties. What does that mean? If I asked you to look at these numbers and count to five, take a few seconds and see how many fives you think are on the page. Now if I were to say how many fives are on the page, you would see it immediately because pre-attentively we're looking at what's different and we see that. When we're here, we're looking through every single line. We're searching for the fives. We're holding it in our short-term memory, which isn't all that great. We're trying to count and say how many have I got? Whereas here, pre-attentively we're saying hey, look here, this is different, we want to capture your attention. This is a really important concept again because we want to call out what's important in our displays but we also don't want to overwhelm. If everything's important, nothing's important and we get all confused.

One of the data gurus you may know, his name is Edward Tufte. He talks about the data-ink ratio. In part this is leveraging our pre-attentive attention. He says good graphics should only data-ink. Non-data-ink is to be deleted everywhere else possible. So we really don't want to exhaust folks with displays like this with lots of red, yellow and green where we're trying to figure out what do all of those things mean, and perhaps we have a magic key to see how to get from one spot to another. What we really want to do, especially in tables, is get rid of all this color and just show the data. I'll tell you more when we talk about tables, but the idea is this is a whole lot of non-data-ink that we could do better and in a different way. So let's get rid of all of this and let's just show the data. You'll see there's really very little non-data-ink here. There's one line under the month separating it, but we simply use light space to say this is January and a column of data, October, September, etc. We don't need lots of colors and non-data-ink. We just want the data to stand on its own.

The other thing about color specifically is we really want to use it to enhance and clarify a presentation. When we use a lot of it, it just obscures the message we're trying to get across. It muddles and confuses. Continue to think about pre-attentive processing when I ask you to look at this slide. This is simply a slide of overtime hours, but every single bar is in a different color. So pre-attentively we're trying to figure out what are all those different colors trying to tell us? Well, the different colors don't tell us anything, because all of these bars represent overtime hours. What we really should do is make them all one color because we want to draw your attention to the shape of the bar, not entertain you with different color bars. So we don't want to confuse folks, and people get really tired when they're trying to figure out what all those different colors mean, and they'll move on and away from the visual that you're trying to present to them.

Colors should really be used again pre-attentively to say this is different. I want you to see this in comparison to something else. I want to call your attention. So if we were just ranking results on different measures but we wanted to compare that to a benchmark, we could potentially say the benchmark is a different color and above it and use below it. But we wouldn't want to make all of these bars different colors and challenge people to try and understand what all those colors mean.

Fill in the blanks. Approximately 10% of all males and 1% of all females are color blind. The most frequent type of color blindness we see are where folks don't have red cones in their eyes which help differentiate between red, yellow and green. So where you and I may see red, yellow and green on the left-hand side, our color blind counterparts are seeing this on the right-hand side. This is actually protected under the Americans with Disabilities Act that we shouldn't be using regular green in combination because folks who don't have these red cones in their eyes can't see this on the left-hand side. There's lots of other ways we can do this and we'll show you.

Any questions? My colleague Janet will let us know.

Janet: There is one question. What color should we use instead of red, yellow, green?

Kathy: Great question. A lot of the software these days have color blind palettes, but what we use most frequently are blue and orange and some gradations of those colors.

Let's talk about tables and when to use a table. The tables are designed to perform a specific job. You use a table to look up individual values, compare individual values, display precise values and communicate more than one unit of measure. Let's think about this for a second. We look up values, we compare values, precise values, etc. What we see all too often, however, is that folks are trying to show us some sort of trend or pieces of information by applying color to their table. This is a report we got that was about six pages long and they were using all these different colors trying to draw our attention to some pattern in this. But that's not what tables are really good for and what we should be using them for. All we want to do is look up values, compare values, etc. So we want to get rid of all this color.

Now if there's something in here that you want to convey we could think about how we arrange the table. For example, these are different measures of care by clinicians, physicians to their patients, and it's how groups are doing on certain metrics compared to some national percentage, and then the variant. So instead of all that blue and orange, if you wanted us to see the ones where we needed to do work versus the ones where we were doing very well, you could just arrange the table in that way and get rid of all that non-data-ink. Because as a pattern seeker and pre-attentively, in my short-term memory I can't hold all those colors and the story that you're trying to tell me in my head. I just need a table that if you want me to see those things arranges the information in that way. If you wanted me to see the information by a region, which was included in that table with all the blue and orange, you could arrange it by region and show me by region which metrics are needed to be worked on compared to the national average. That's how we want to use tables.

You can always consider how you might arrange a table to make it easy for the viewer to see for information you want to convey. Sometimes, however, you just list the information and organize it alphabetically. If you're looking at a group of states maybe you want to list it alphabetically because I can look that up and I know where my state is and I can see the data information. So real take-home message here, there's a lot of information around tables. But the real take-home message here is really you can't use [\[00:28:31\]](#) color to convey patterns in tables. We'll show you how to do that better in graphs. So show the data-ink, get rid of the non-data-ink, and think about how you might arrange the data and information in the table to convey any message you think is important.

A really beautiful example of this is the Vietnam Veterans Memorial. If you've been to it you know the names are arranged in order by date of death or missing in action. What this does that listing alphabetically never could do is it provides so many more dimensions to this just list of names. First off, because they're by date of death or missing in action, as you walk into the monument the first granite slabs are smaller because there were fewer events. As you move through there are more and more casualties, and as you come out we deescalate from the war and the panels get smaller again because there are fewer casualties. In addition to the fact when you look up the name of someone you may know or a loved one, the names that are around that name are most often or usually the comrades who would have been in their platoon and died with them. So the arrangement of just simple names can convey so much more and I think a beautiful and elegant example is the Vietnam Veterans Memorial.

As we move towards talking about graphs, we have to talk first about chart junk. This is also a term coined by Edward Tufte. Essentially he said that chart junk does not achieve the goals of its propagators. The overwhelming fact of data graphics is that they stand or fall on their content

gracefully displayed. In my terms, essentially what he's saying is when we create graphs like this they're just a bunch of junk that don't do anyone anything, chart junk.

Let's talk about when to use a graph and how to create some really great graphs. What I'm going to show you are some things we see folks struggle with most frequently, the most frequent mistakes. Certainly there's way more to all of this but these are the things we see most frequently. So when do we want to use a graph? We want to use a graph to show patterns, trends, exceptions, relationships between multiple values. So you notice I'm saying words that I said tables weren't great at doing. Tables aren't great at seeing patterns and trends and exceptions. But graphs are really great at being able to help us see these things. Here's the fact of the matter. You really don't need anything very fancy. You really need about six very simple ways to encode your data. You can use a point. You can use bars, vertical or horizontal.

Let's talk about healthcare. Let's think about how we read a book. If you think about how we read a book, you open a book and see it has a table of contents and you turn the pages. They go from right to left. We read from top to bottom, left to right. If there's a word we don't know we might go look in the dictionary. We might turn over a corner. We might underline something of interest. That is how we read a hardcover or softcover book. If Amazon and Kindle had met then said, "Hey, we created this great new thing called the Kindle and it's going to revolutionize how you read a book – you start at the back, you read from the bottom to the top and right to left," would any of us ever have adopted the Kindle? No! What they did was they took our mental model and they studied it and they understood how we in this country have all learned to read a book, and that's what they translated into a Kindle. That's how we adopted it if it made sense to us. So we have to think about the same thing for our audiences. How do they think about the work they do, the decisions they need to make? What are their mental models?

And they need to frame it within the context of this scope, role and their decision or need. What do I mean by that? If I'm designing a data dashboard for a chief executive officer, he or she is responsible for everything in the organization, so I have to work with him or her to say okay, what is the summary information that you need to monitor on a page? What is that that you need to get on a page given your scope in this organization? What is your role? So what function do you play? "As a CEO, my function is to keep us on target and make sure my managers and directors are programmed forward." What decisions do you need to make? "As the CEO I need to know who to pick up the phone and call about all of these metrics that I'm responsible for."

There's a ton more to this but what I want to impart to you is it's important to start to think about mental models and to start having real conversations with people. Often the way I do this is I say "When you come in the office Monday morning, what do you think about first? Second? How do you use data to make a decision in the past? What is it you need to know and understand in order to make a decision?" That decision may just be I need to ask somebody else to go look at this in more depth. So mental model, how do I think about these things in the context of my scope, role, and decisions. Clearly managers are different than patients and their families, etc. So we need to really think about those things.

So we thought about those things. We understand some of the best practices of data viz. We understand the gestalt principles and some things about colors. How do we put it all together into dashboard? Clearly we want these well-organized, and again that's back to these principles of how we see things and how we group things. We want to think about the organization of our pages.

Details __00:51:08? that leads to summaries. This is probably the hardest part of a dashboard. What is it that you need to summarize on a page so that folks know where they should go, look, or who they

should talk to next? You're never going to get everything on that page so you've got to be thinking about how to summarize these things.

We want to highlight the unfavorable trends and exceptions. Remember, pre-attentively if everything's important, nothing's important; we want clear, concise displays and compared with what is everything.

This is an example of a CEO dashboard. I'm not going to go through every single piece of it; you'll receive these slides and you'll be able to study it. If we think about a hospital CEO, they're responsible for things like budget FTEs so we summarize are you on budget or off for FTEs. What does your revenue look like your sense is? Payer mix is extremely important so what does your payer mix look like through a previous period?

They're responsible for summary quality metrics so we've put the composite scores here; things like patient satisfaction - some other summary quality metrics compared to a target and national average.

They're spending a lot of money on electronic health records; how well are different departments using them. These are OE ratios about mortality and we simply use color to say, your OE, your observed and expected mortality rates are higher here than here; you want to take a look at what's going on. So you'll see not a lot of color, well organized, we've just used lines, points, and bars to get all of this on a page to help a CEO dashboard monitor what they need to. They might not pick up the phone about these two departments and how well they're doing on their electronic health record use, but they might pick up the phone to these groups because they know, and can see, that they're not doing so well in their compliance.

On a state prototype public dashboard, and again we went to Idaho and grabbed some data about diabetes rates, and around thinking about context we did a couple of things. First we just used a quadrant map where we use darker colors to show higher rates of diabetes. Here we used a simple table to say what is the population of the state and what is the overall rate. Then we were curious about these uninsured and how many primary care physicians there are per 100,000 folks in those states.

You can see here it's ranked and we just use quadrants to plot uninsured and diabetes rates and how many physicians and diabetes rates; all on one page, not a lot of colors, etc., just points and some boxes and narrative.

Very quickly, moving through to what's an infographic because lots of folks talk about those and how are they different from a dashboard. On a dashboard we are looking for exceptions and trends. With an infographic we are working to deliver some message to inform, persuade, teach, and move folks to actions. These you'll see in lots of newspapers or on web sites and they are often topic-focused. They can be interactive, but often they are just a poster story.

Here's one we pulled from a newspaper and it's, I think, very simple and very elegant. Essentially what you'll see here is we're looking at percentages uninsured by county in the United States from 2013 to 2015. You'll see that in 2013 essentially there are only 10 states with a percentage of residents who lacked health insurance lower than 9%. In most of these states the uninsured rate was higher than 9%.

Just using a small multiple of a map; cut these small multiples we have a map, we have a map, we have a map. We can see a timeline of what happened. The ACA was rolled out here and you'll see in the black, the states that are outlined in black, those are the states that expanded Medicaid under the ACA and you can now see the difference in uninsured rates.

In 2015 you'll see that Pennsylvania and Indiana are also now outlined in black, where they weren't before, and you'll see that their uninsured rates came down. This is an info graphic; it's made to just teach and inform. Perhaps move to action and perhaps persuade, but simply we've used a small multiple, it's like a timeline, and they've just used color saturation to tell a story of what's gone on around the ACA and the uninsured.

Finally I want to make this point because we see so many folks falling into this trap and that is, technology alone is not the solution. Technology will let you do things that are not best practices; they'll let you make 3-D flying pie charts in red, yellow, and green, all sorts of funkadelic stuff, but just because software lets you do it doesn't mean you should. You need to know those things we've been talking about today and best practices.

Additionally, simply giving someone the technology doesn't mean they're going to create beautiful data visualizations because if it was true I would have written the great American novel a long time ago. I'm great with words, but I'm never going to write a great novel; I would have to train and learn how to do that. Just because you've got some great piece of technology doesn't mean that you're necessarily going to have fantastic reports and dashboards and info-graphics; you need all three legs of these stools. You need to get the subject matter experts in the room, you need to get the folks who know the technology in the room, and you need to start to learn and embrace the best practices of data visualization and best practices. You're never necessarily going to find one person who can do all of these things; it's really about a team that comes together and knows how to talk with one another and think through mental models, scope role and decisions, what the technology can indeed do, what the best practices are to create data visualizations and dashboards.

I think I'm going to turn it back over to Jessie. You are going to take it from here and then we're going to answer some questions.

Jesse Parker (JP): I think we're ready to jump right into questions and then we'll do the formal wrap-up.

CR: If you've got questions fire away.

Janet: Yes, we've got questions.

How do you recommend use of color in a way that's readable when printed out in black and white? A dark blue and a dark red will look similar in black and white.

CR: That's a great question. You can do some things with saturation grayscale. If you remember where I showed you the count to five part that was in grayscale and then we used black, but you can do a few things. You could make something bold, you could potentially italicize it, you could enclose it like put a box around something you want to draw someone's attention to, you could asterisk it; all those things that we use in writing you could also use in your data visualizations and display. If it looks different it draws someone's attention to it. If it's italicized you go oh, that's different; if it's bolded you go oh, that's a little bit different; if it's called out with a box you go oh, I'm going to look there it's a little bit different. So you could try some of those techniques as well.

Janet: My superiors like pie charts in red, yellow, green; how do I get them to accept a change?

CR: We get this question all the time and here's what I say: pick one thing at a time. I also think it's about not just saying no we can't do that, I think it's in part explaining why. If you show them I can only do this with a pie chart, but look how much more I can give you with a bar and I can use my labels once and I can show you a variance and I can rank and I can show a trend. Once they see how much more they can get they will start to embrace it. We see that time and again. The red, yellow, green,

part of it working for state agencies is it really is under the Americans with Disabilities Act (ADA). You can do alternatives; you can do the blue or the orange.

Once folks start to see things that are more elegant and simple and not so exhausting, they begin to really embrace it and they can see the difference very quickly. Don't try to change everything at once; pick a few things and ease them into it.

Janet: What software or softwares do you primarily use for visualization?

CR: All the software out there is pretty general. We use any of them that our clients need to use. You can do a lot of cool things with Excel and some of the things I showed you were in Excel. We also use a piece of software called Tableau that we like and is very nice to use. Sometimes we use Click or Click Views, so there are a lot of tools out there.

Janet: Is there any visual intelligence research about interactive info-graphics?

CR: Yes, I'm certain that there is, but I admit that I am not fully expert in that; we don't do a ton of it. I would recommend that there are a couple of really great books out there about user interface design so you would want to explore some of those. For example, one of the things that we know is where people look on a page is really interesting. People first look on a page in the upper left, their eyes move over to the right and then down the right-hand column. The last place they look is in the left-hand corner of the page or computer screen so you wouldn't want to put something really important in the left-hand corner. If you think about something like Amazon, when you go in the first thing you see is the search bar, but then the next place if you look down on the right is where you do one click or you buy or you order. That's based on research about where folks look.

Rule of thumb: you want to put important things upper left, people move over to the right, down the right-hand column, and you don't want a lot of instruction. You want to use blue with an underline, which says it's a hyperlink; if you click here it takes you someplace else. I would encourage whoever asked that question is interested, there's lots of user experience interface design work books out there that you might want to explore.

Janet: If bar charts always start at zero, and I'm working with percentages, do they always end at 100%?

CR: No they don't. What you're trying to show is how long a bar chart is so you don't necessarily have to go to 100%. That's one of those things you need to think about. Sometimes you want it to go to 100% because so many of your bars are at 80 or 90%, but you don't have to go to 100%. You do need to start at zero, again, because you're showing how big something is.

Janet: Do line graphs also have to start at zero?

CR: We get folks to understand that bar graphs have to start at zero and what we start to see is that then folks think all their line graphs have to start at zero, but line graphs don't. Line graphs can start at a number closer to your starting point. The reason is that you're not showing how big something is, the size of something, what you're showing is how it has changed over time. If something starts at 70% you don't need a whole lot of white space under it with nothing in it; you could potentially start that line graph at 60% or 65% because what you're trying to show is the shape of how something has changed over time.

Janet: How do you make small multiple bar charts on Excel?

CR: Essentially, and here I would direct you to a couple of web sites, one is Peltiertech. If you want to send me a note a cathy@healthdataviz.com I can direct you. Essentially you make one bar chart and then copy it and line them up one right after the other. At places like Peltiertech there are tons of tutorials on how to do things like that. If you want to shoot me a note I'm happy to direct you to resources.

Janet: Do you have recommendations for flowcharts?

CR: Flowcharts are a different animal, if I am understanding the question correctly. I would say there's some software out there that does that. Excel has some that does that and those have come out of engineering. We don't do a ton of those so they're different than data visualization dashboards and reports so I don't know that I have a strong recommendation about what you might want to use for that.

Janet: If I want to show a trend over time, but I'm missing some data, what do I do? How do I display it?

CR: If you have data over time and you're missing some data points in that... you have January, February, and March, you don't have anything for April and then you have May, June, and July and April was not zero; you just have no data. You don't want to show that as zero, you show it as a gap. So you will have a line through January, February, and March, there will be a space, the gap, and then you'll pick up the line again. Visually that says there's a gap there and you could add a note, "no data imported," but you don't want to show that as a zero because zero is a value. If you are missing data in a trend graph, show it as a gap.

Janet: Any recommendations about keeping dashboards to one topic? They keep asking for the "ultimate" dashboard that has everything and every option.

CR: Go back to scope, role, and decisions; start there. You've got to figure out and you've got to talk to folks and say is your scope and role that you're responsible for a specific project or program? Or it could be the CEO and he or she is responsible for all of the projects. Again, the bugaboo is you've got to get out of the weeds and you've got to think about how to categorize those weeds, those details, into summary categories.

One of the things I didn't talk about on the webinar that we teach in our data visualization courses is thinking about how your data is categorized. Is it ordinal? Is it hierarchical? There are different categories of how data is categorized. Pretty much everything has categories. You might have five programs. Programs are a category so it might be you summarize those, but you've really got to spend a lot of time with people redirecting them to think about how to summarize; making them understand that we can always give you a report that gives you more details about that one program out of the five you're responsible for.

Before I ever go to a computer I get pencil and paper and I think about what's my data, what can I do with it, what are the categories, and how might I summarize it. If you just stick to your guns and you keep bringing people back around, they'll start to get it. Sometimes it takes me a full day, but I have yet to have anybody not really say to me, "Oh, I see I'm in the weeds and you're trying to get me to summarize." That's the work at hand in addition to scope, role, and decision. If you're not responsible for budget, you're just responsible for performance on projects, budget doesn't go on your dashboard; it might go on a separate report, but it doesn't go on here; it's not within your scope and role. You've got to figure that out with your client.

Janet: Do you have any additional tips or guidance on public-facing or consumer-focused dashboards or reports versus manager dashboards and reports?

CR: They are very different animals. I would almost say that consumer-facing doesn't really necessarily fit the pure definition of how I think about dashboards. If we think about dashboards for monitoring performance, etc. then that's very different than information that we're providing to consumers that may or may not be interactive. We might be splitting hairs about data definitions, but that's how I think about it.

The other thing I would say is, and we didn't talk about this today, but the big, big, big thing that I see that's a mistake on consumer-facing reports and portals is that we have this cursive knowledge about the work that we do and so we talk about it in terms that we understand. We've got to really think about plain language in how we talk to folks who don't live in the world of three letter acronyms and gobblede gook that we live in. We've just got to give them very simple graphs, in plain language, that helps them to understand what they're looking at whether it's which physician to choose or which hospital to go to or which insurance product they might want to buy; we don't want to overwhelm them and we want to use simple charts and simple language to get through to them the information that they need. That's my best advice: we have to step back from our expertise and we have to think about how they think about these things.

The other thing we didn't chat about much today, but it's about user testing. We create these things, but we never test them with folks; we roll them out and we point and click and say isn't this cool what we've created for you, look how it works; and we know how it works. Very simple user testing can be you create a few sample questions and you say can you tell me which physician has the lowest readmission rate, consumer. If we design things beautifully and simply, they should be able, without us pointing and clicking, to see that information and provide an answer. If they can't we need to go back to the drawing board.

That's true of all of our dashboards and reports; we should be able to write up some simple sentences, three or four, and then hand the software or the dashboard to our stakeholders, our customers, and they should be able to navigate and find the answers that we can see have provided to them.

There's a really great book called "It's Not Rocket Surgery" [actual title: "Rocket Surgery Made Easy"] by a gentleman named Steve Krug and we can make that available to you. He provides very simple low cost ways to do that type of user testing on the things that you design.

Janet: Is it ever okay to show two different independent variables on the same plain if they are at different scales, such as the graphs that have two different vertical axis on the left and right of the chart?

CR: That's one that I'm going to pass on right now; I would have to see it in more detail before I answered the question. Again, if you want to shoot me a note I'm happy to look at it. I'd want to look at that one to answer more completely.

Janet: Any recommendations on heat map or scattered map design?

CR: We use heat maps, sometimes, in our data visualizations and, essentially, heat maps just show by a darker color where something may be a higher rate or a lower rate and they're often in a table of information. What was the question again?

Janet: Do you have any recommendations about heat map design?

CR: Keep it simple, don't go crazy. We find that we use them when we've got a lot of rates over a lot, maybe, of time frames or places that we need to show. We will use them and they can be very effective that way. They can be effective, but don't go crazy with them. For example, if we wanted to know there are a bunch of clinics, there are a bunch of programs open on different days of the week, and when are we seeing the most traffic through those programs or clinics, we would use a heat map. That would allow us to show days of the week and hours and we would use a heat map that way. Again, if you have a specific example you want me to take a look at I'm happy to have you send it along to me.

Janet: Is there free software you would recommend for making charts?

CR: That's a good question. I would have to research that; I am not aware of any off the top of my head, but I'm certain there is some somewhere.

Janet: Do you have resources for creating color blind friendly color palettes?

CR: There's a website called ColorBrewer and we can share that link with you. Also, some of the software already has that option built into it and depending on what you're using we can show you that. ColorBrewer is a very good site for which colors to use, when and how, and we could add that to our resource list for you.

Janet: No more questions.

JP: Thank you, Cathy. There have been questions coming in via the chat on the side. We will be posting the slides on the IAP Data Analytics web site. Every registered participant will receive an email with the link when it's posted so you don't have to keep checking; we'll let you know. That wraps it up for us today.

I want to thank Cathy for her informative and helpful presentation and for taking all of these questions. Thanks to everyone on the call who took the time to join us on this webinar. Keep an eye out for more state webinars offered by the Data Analytics IAP group in the future. These will continue to be announced via the SOTA newsletter. Thanks again.