

How to Become the MacGyver of Data Visualizations

Tricia Aanderud, Zencos Consulting

ABSTRACT

If you don't understand what makes a good data visualization - then chances are you're doing it wrong. Many business people are given data to analyze and present when they often don't understand how to present their ideas visually. We are taught to think about data as numbers. We often fail to understand that numbers show causes and help others reason through issues. In this paper, we will review how data visualizations fail to understand what makes a good data visualization work.

INTRODUCTION

In the 1980s, there was a popular TV show about a character named MacGyver, who was especially adept at taking ordinary objects and rescuing himself from impossible situations. While some may argue his success involved the magic of television, I say his understanding of how things work on a basic level also contributed to his effectiveness. While working with data does not put you in many life and death situations, a good understanding of data visualization (datavis) basics can help you rescue yourself from an embarrassing situation and create effective data visualizations.

There are multiple ways to visualize data – everything from a table to a map. Many people use tables or even rows of data in a spreadsheet to impart information – but it is hard to understand patterns with that method. Our eyes can interpret patterns more quickly when offered as a visualization. This paper discusses datavis methodology for common chart types and suggests some alternate strategies.

DATA VISUALIZATION BASICS

No matter which datavis method you are using there are still a few rules that apply to all of them. In Roger Parker's book called *Looking Good in Print* he has many examples of how even professionals create ineffective advertisements, party invitations, and newsletters because of a failure to understand how people consume visual information.

He provided several makeovers to show what a difference a clean layout, a simple color change, or removing words made to the result. While he was often re-doing something that was terrible to view - he was careful to note that design was not about a *good or bad* result – it was about *effective* communication. A cardboard sign that read Yard Sale in a faint, small font was just not as effective as one with large black letters and a date. A careful person could see the smaller sign, while the re-made one could draw more attention to itself and capture more customers. Thus, the larger sign was more effective. These same ideal applies to datavis.

KNOW YOUR POINT AKA “WHAT ARE YOU TRYING TO COMMUNICATE?”

It seems silly to start with a statement like “Make sure you understand your message.” Why would someone assemble a datavis otherwise? The problem introduces itself when you mix a fancy datavis application with a mild-mannered data analyst. The result tends to be a datavis that shouts, “Hey look what I can do!” It is easy to find cool ways to display the data without considering if it leaves the audience with an ineffective message.

This is where data storytelling enters the picture. The datavis must answer a question, clarify a point, or reveal relationships within the data. After seeing the datavis, the user should have a takeaway. The takeaway can be as simple as an insight or as complex as a process improvement.

Definitely, analysts should be encouraged to find new ways to display data but the method should enhance their message and not focus on the software. Think of what your datavis is trying to communicate to your audience as you create it. It might help if you put your question on the top of graph and then explain to yourself how the datavis supports the point.

KNOW YOUR AUDIENCE

If someone is inexperienced with a bar chart then a box plot will really take some explaining. If the audience is confused about the datavis technique, they might miss your point completely. However, if your audience is willing to learn, it might be worth your time to educate them. In general, save your sophisticated datavis for an advanced crowd.

Also, consider how well the audience understands the underlying data. Those audience members more familiar with call center traffic and issues require less education about the data than someone who walks off the street. Those who are familiar expect issues about inadequate staffing or increased call volumes so they might be able to handle an advanced datavis because they understand the data and collection process better.

FOLLOW THE KISS PRINCIPLE

Probably you have heard the Keep It Simple Sweetie (KISS) principle stated hundreds of times – mainly because it is true. Keep your message and datavis simple by removing any unnecessary data and visual clutter. Your job is to direct their attention to what is important about the message.

Data visualization experts (Few and Tufte) remind us that the users should not be distracted by the presentation method; instead they should be focused on the numbers and the message. Your goal is to simplify the datavis so the users can see what you see.

USING LINE CHARTS EFFECTIVELY

Line charts allow you to see trends over time. They have a much more simple purpose than any other chart type. Variations of line charts include area charts and Pareto charts. There are some simple guidelines for producing this chart type:

- Keep the intervals in order. In the following example, there is a value for each month and year.
- Notice that the line connects each data point. It is easier to understand the trend when the points are connected.
- Indicate missing values. If you did not have data for the summer of 2013 then you would want to ensure the user understood the data was missing. Otherwise, your chart might take a huge leap forward and the user would draw the wrong conclusion.

Line charts use the X-axis for time series, such as year, month, hour, or even minute. Use the Y-axis for the value you want to plot. In the following example, you can see the arrival rate for consumer complaints by product. There is a line for each product. This datavis is showing that consumer complaints about Mortgages has decreased while Credit Reporting complaints doubled and kept going. The line chart makes following the trends easy.

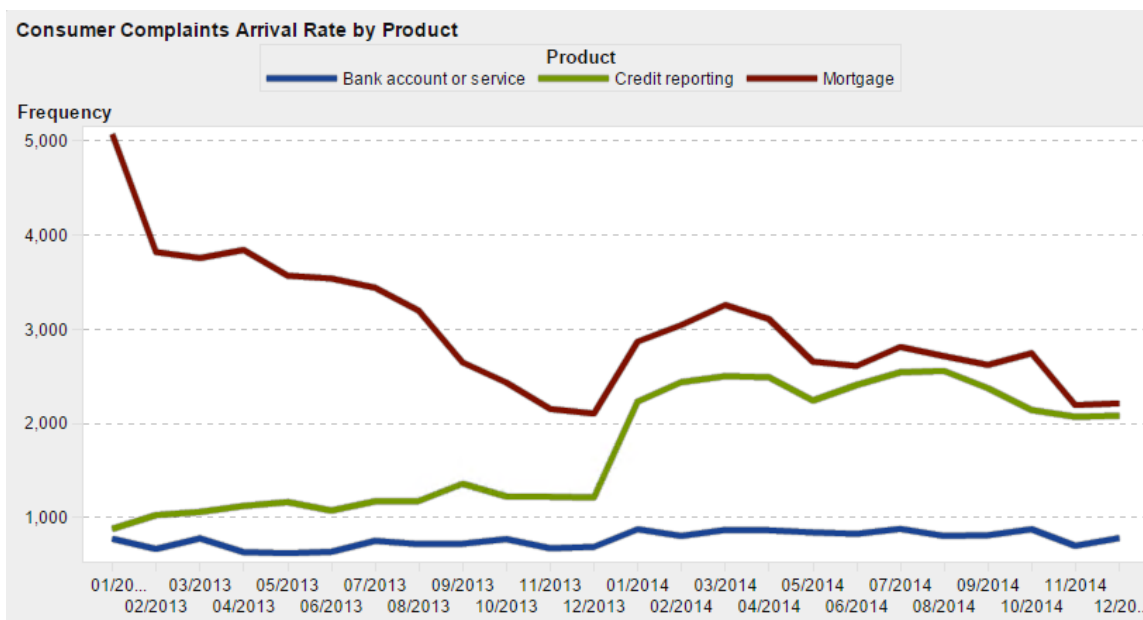


Figure 1 Simple line chart example

REMEMBER THAT KISS PRINCIPLE?

[According to Miller's Law, most people can keep about 5-7 items in their working memory at once.](#) When a chart becomes too busy or has too many lines, it is more difficult for the user to absorb the information. In the following chart, only 11 lines are showing but you will spend a lot more time studying it as compared to the chart above. One takeaway is that some products receive few complaints. However, if your message is "there's only a few products with issues" then use this chart to emphasize that point. If your point is to show the growth difference in the main areas, use the chart in Figure 1.

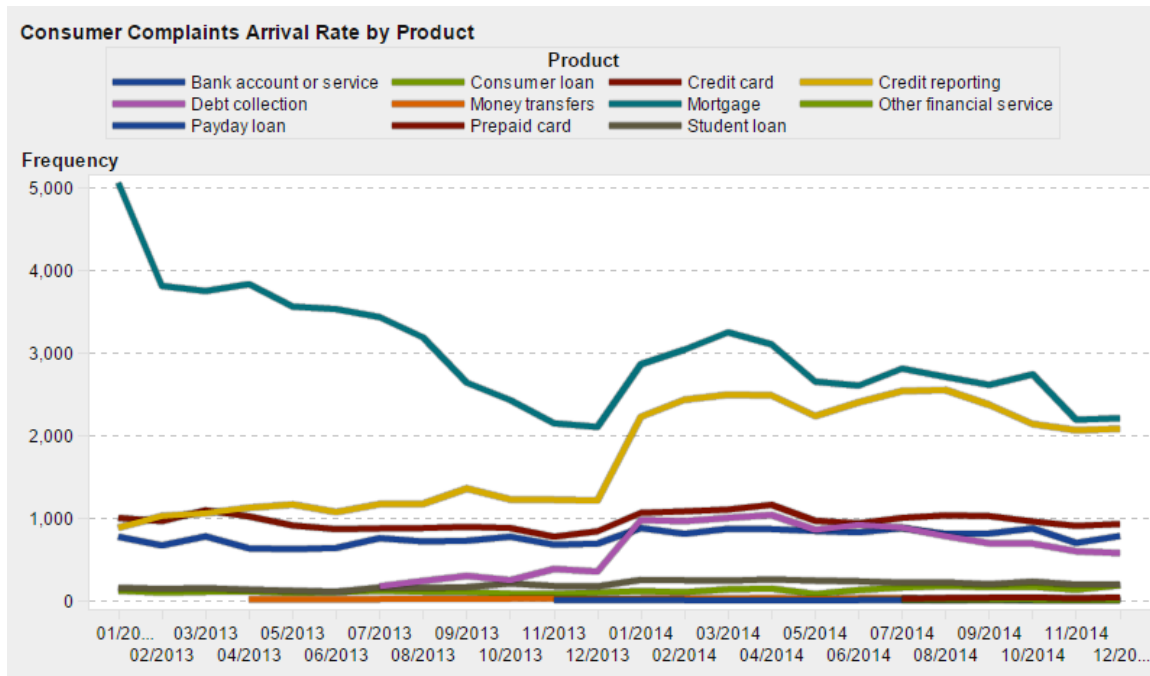


Figure 2 Not very KISSy

GENERALLY USE 0 AS Y-AXIS VALUE

If you need to infuse your chart with some drama, then play with the Y-axis value. Consider the following graphs and how much more dramatic the trend seems when we changed the Y-axis value. The reported product issues are arriving a dramatic pace indicating a product with many issues. When we place the y-axis back at 0 it is easier to understand the there is a flow to the arrival that may even be seasonal.

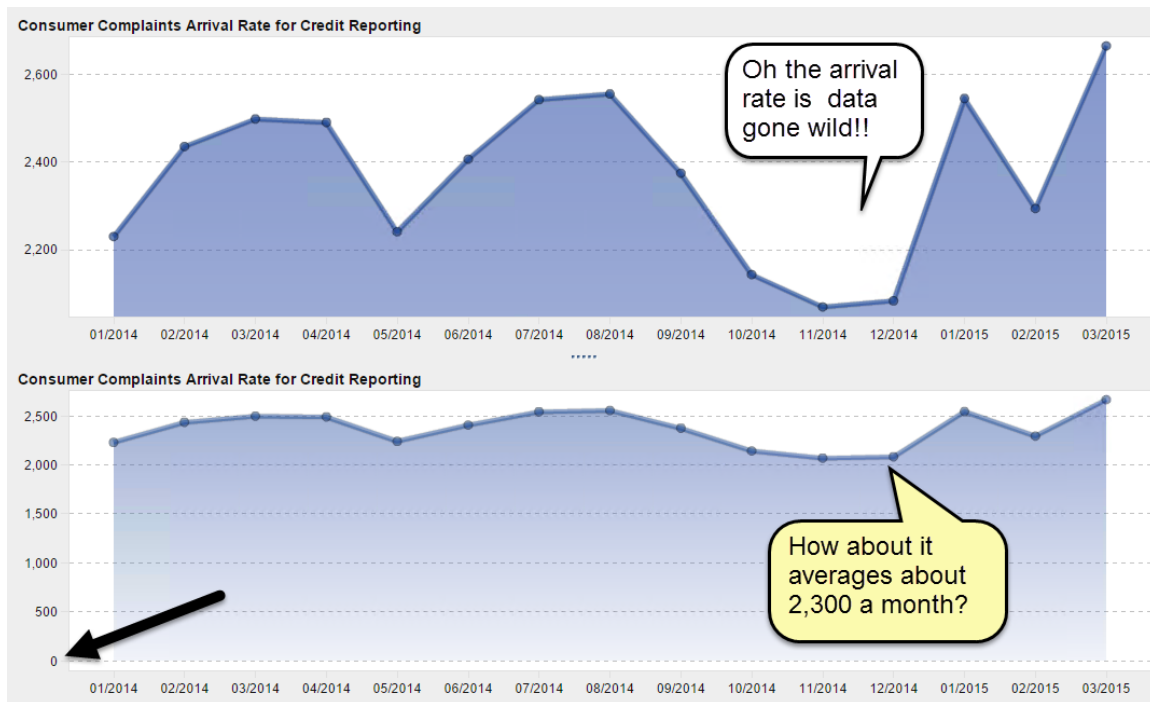


Figure 3 Area chart with and without the 0-start point

In *Show Me the Numbers*, Stephen Few suggested that a better way to handle this situation was to show the overall chart and then a second chart with a more focused trend line. You can imagine a case where a drop of 400 records might get a small business excited – especially when they are trying to staff a call center or plan production runs.

CAREFUL WITH STACKING AREA CHARTS

When you are working with stacked area charts, you can easily confuse users who don't understand your main point. The problem lies in how you want to emphasize the parts to the whole. In this example, the datavis shows an area chart grouped by complaint channel to help the user understand which channels drove the overall trend. The question was "Which channel contributed the most to the arrival rate in 2012?"

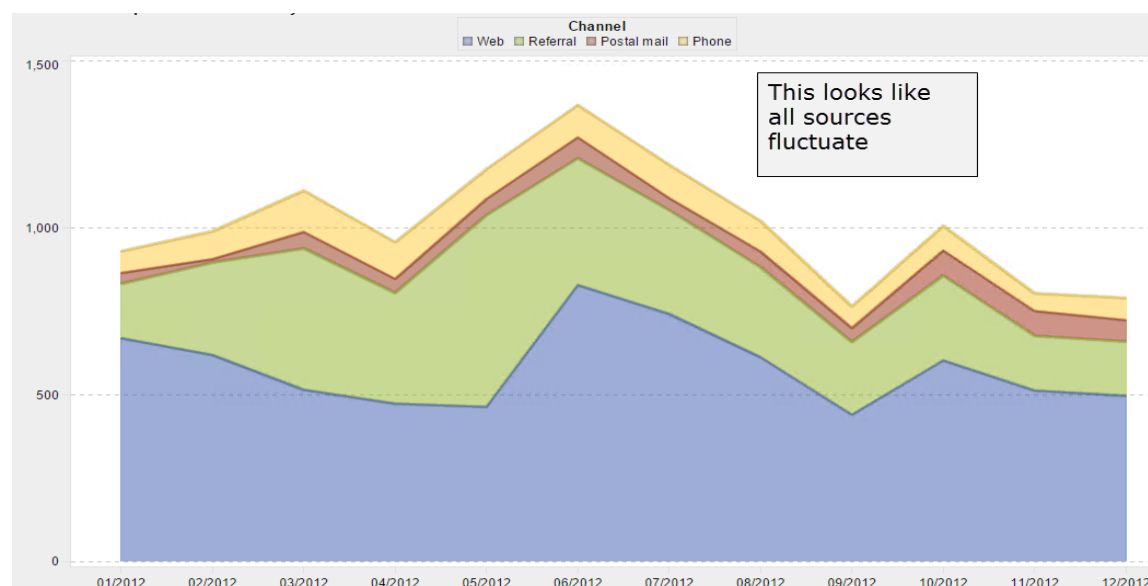


Figure 4 Parts to the whole or arrival rate fluctuates for all channels

What if the title was a more generic one – such as "Arrival Rate by Channel?" which causes the user to focus on the arrival rate fluctuations? While it appears that Phone and Postal mail had a lot of variation, it is not true. When you divide the channel into a trellis chart, a different story emerges. In this story, the Web and Referral channels contribute the most to the trending with the Web channel driving everything by the year-end.

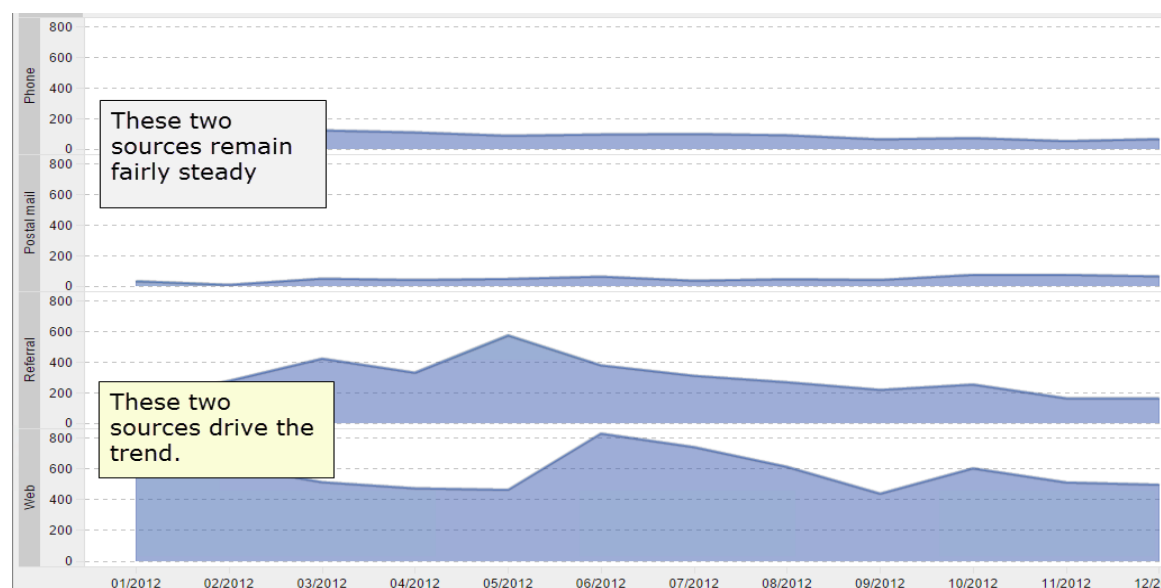


Figure 5 A trellis chart shows what really drives the trend

My point with this illustration is not that the stacked area chart is bad but instead the question is, "Was it effective?" This example is to help you understand how a datavis was mis-interpreted despite our best intentions.

USING PIE CHARTS EFFECTIVELY

Pie charts show the parts to the whole. Many data visualization experts do not advocate using pie charts because as Stephen Few says “they communicate information poorly.” If you want to use a pie chart, make sure you understand the guidelines for doing it correctly. Generally, a pie chart offers visual relief in a sea of text or boxes.

Here are the guidelines for how to use a pie chart to display your data.

- Parts to a whole equals 100% – always. If your datavis does not equal 100% - tell the user in a footnote.
- Limit to 4 or 5 categories but it's better when one category is significant percentage-wise
- Legends should be superfluous when pie chart is done correctly

A pie chart shows how each slice contributes to the entire pie. Each slice is a category and a user should quickly look at the chart and have an answer. This is why many datavis experts hate a pie chart! Their argument is that statement often times would work better than what Edward Tufte, in his *This Visual Display of Quantitative Information* book, calls a “dumb old pie chart”.

IS IT EFFECTIVE OR NOT?

In following figure, you can see of an example of each technique. Oftentimes datavis newbies try to do too much with a pie chart and it just goes wrong. The pie chart in the figure is simple and the user's takeaway should be that Netflix and Twitter are distractions or someone needs to spend more time doing her chores. There was some loss in detail with the text message but was it as effective? Imagine if the pie chart showed a Yes/No response to a survey question – which technique might be more effective then?

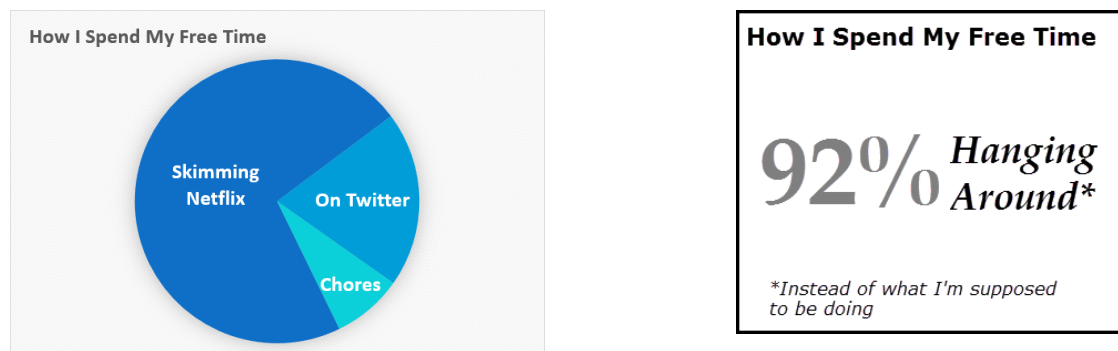


Figure 6 Pie charts versus a textual statement

LIMIT THE CATEGORIES TO FOCUS THE USER'S ATTENTION

In an earlier topic, you observed how the line chart with too many lines started to get confusing and quickly lost its point. When you have too many categorical values in a pie chart, you make the user's job 10x more difficult. The user may ask themselves “Is this a ranking?” or “Do these other categories really matter – why am I being shown this?” Notice how going back and forth between the colors and legend is a drag. With the following datavis, you can see why the horizontal chart becomes a better choice.

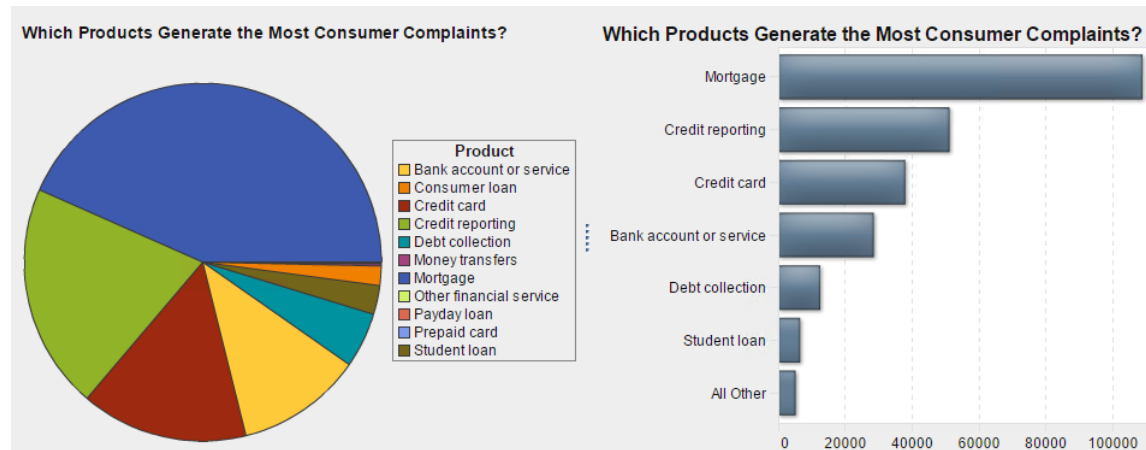


Figure 7 Pie charts versus a horizontal bar chart

HARDER TO COMPARE PIE CHARTS

In the following figure, the datavis compares complaint arrival by channel. You have 5 seconds to tell me the second most popular channel to initiate a complaint. Again – is this an effective way to display data? Would MacGyver use it?

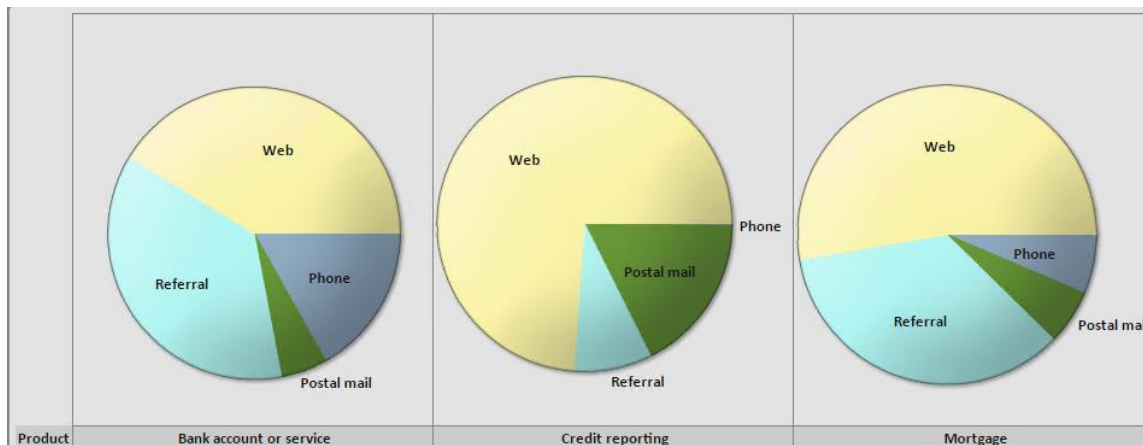


Figure 8 Pie charts used to compare categorical values

USING BAR CHARTS EFFECTIVELY

Bar charts provide more detailed information than line charts. This datavis type makes it easier to compare exact quantitative categories. There are two types of bar charts: vertical and horizontal. Vertical charts compare categories while horizontal charts work especially well for ranking.

When producing these charts, keep the following tips in mind:

- Your axis should start at 0 for this chart as well
- Careful when vertical bar chart categories exceed 10 – it can get overwhelming
- When using an *Other* category, ensure you keep it to a low percentage.

In the vertical bar chart, the X-axis is categorical data so no order is necessary. The Y-axis is the value that indicates the length of the bar. Some choose to sort the variables in descending order so the highest value is to the left. In the following figure, you can more easily see how the line chart allows the eye to see the trend while the bar chart shows a specific value. This is another occasion where you have to determine which is more effective in communicating your point.

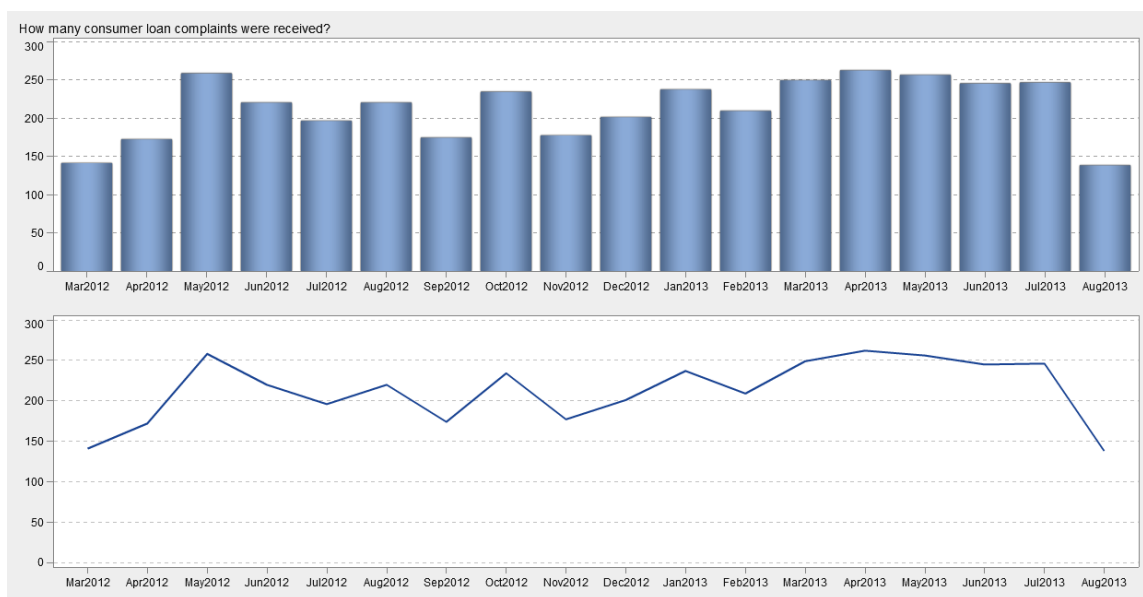


Figure 9 Showdown – Bar versus line chart

HELP THE USER VISUALLY

With bar charts, it is a little easier for a data analyst to turn into a wanna-be artist and let the creative juices flow. If you break away from the norm, then you must have a solid understanding of graphic design and data presentation skills.

Keep the following guidelines in mind as you produce your bar charts:

- Allow white space between the bars and keep it the same distance. Usually the software handles this task so it is a non-issue.
- Keep bars the same color when the data is a single category. Unless your whole package is using a theme for a particular category it usually only distracts the user.
- Avoid using patterns or anything unusual for the bars. Yes - it is distracting.

Here is a datavis makeover – you can see how much easier it is to read and understand the one on the right. In addition, the *Other* category was more appropriately handled. In the original the Fax and Email had such a small contribution it was almost nothing. Moreover, there was no value in having different colors for the categorical values.

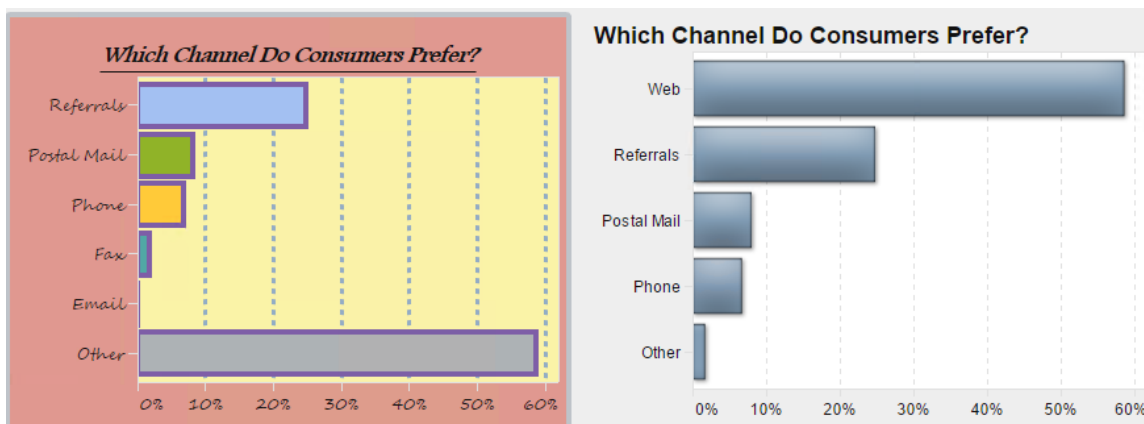


Figure 10 These guidelines apply to all charts

RESCUE YOUR LONG LABELS AND YOUR USER

Horizontal bar charts assist with making comparisons but are also useful if your labels are long. Notice in this example the difference in the labels. The slanted labels are difficult to read mainly because they are too long. By turning the chart on its side – the values are much easier to read.

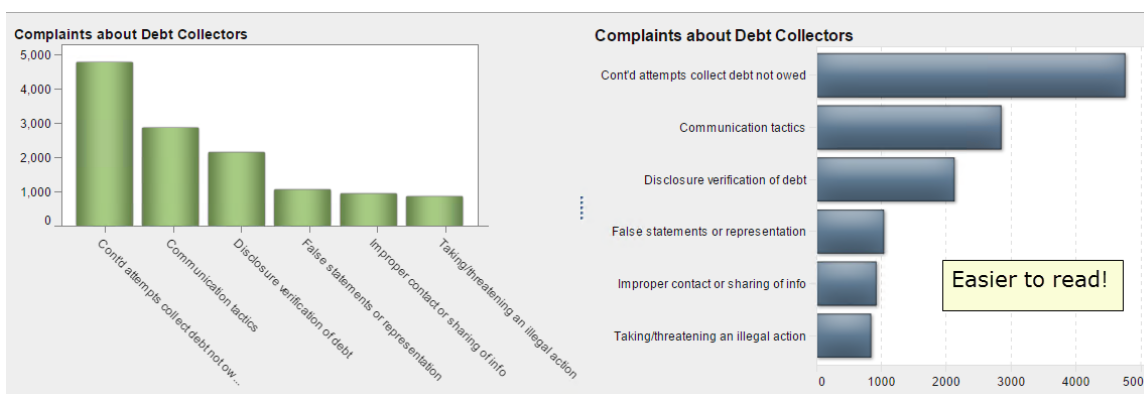


Figure 11 A sore neck should not be part of datavis

GROUPED VERSUS STACKED CHARTS

In an earlier topic, we talked about how a user could interpret a stacked area chart multiple ways and may miss your point. Then we tried to compare this same data with a pie chart and that resulted in an epic fail.

Now let us talk about where a bar chart really shines – comparing across groups. There are two ways to compare values with a bar chart: stacked and grouped (or clustered). You can decide which is most effective for your message.

Stacked charts reveal the whole and show how the parts contribute. In the following figure, you can see a stacked chart both horizontally and vertically. Notice that the percentages are sorted which ranks the values. Even with values not shown as percentages, you get a sense how many more complaints are about mortgages over the other categories.

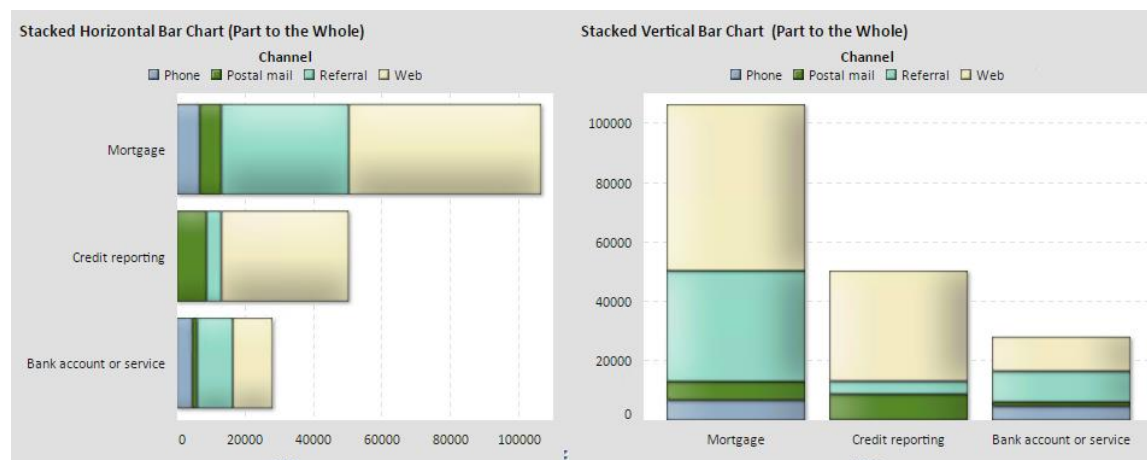


Figure 12 Parts to the whole seen both ways

Grouped charts are easier to compare across categories. Notice that the white space is between Product instead of Channel. Your eyes take the visual clue that those items are related within the grouping. This chart does give you a sense of overall counts but it does show the Web channel as the most popular contact method. What you also see is that almost no one uses Postal mail to complain about his or her bank account, but it is a popular method for the other categories.

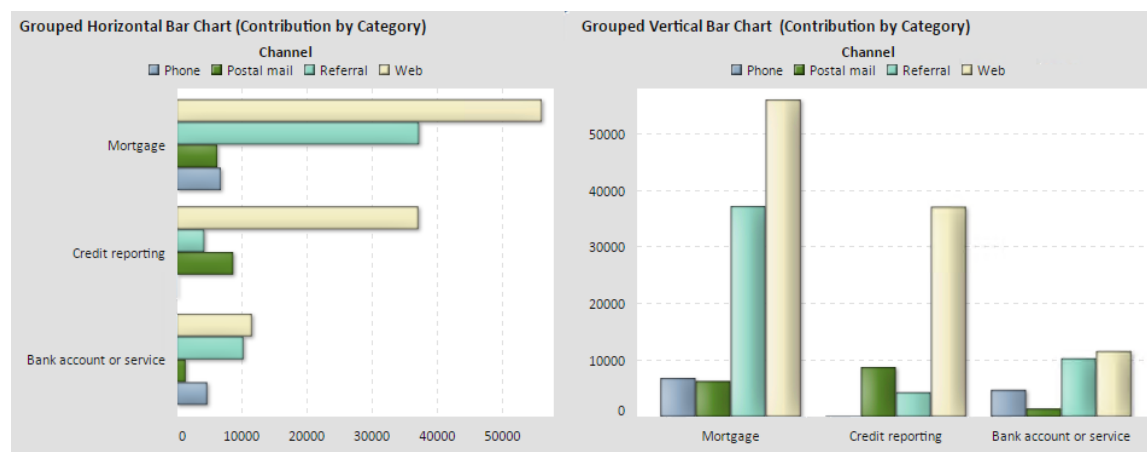


Figure 13 Comparing across categories

Take a moment to study the previous figure. In *Show Me the Numbers*, Stephen Few noted that most likely due to our cultural preferences, we tend to sort values as top to bottom or left to right. In the previous figure, the vertical bar chart values are not stored left to right, did you notice? Did it make you pause or want to correct it? The vertical bar chart sorting might have made you think the horizontal one was more effective.

USING GEO-SPATIAL CHARTS EFFECTIVELY

The most important rule for geospatial charts is that your story has to be about why the geography is important. If you want to show that your customers live close to your stores, then you have a good reason. However if your intent is simply to say here's how each state spent budget on a particular line item – it won't make sense if there is not a story to go with it. Oh, did I hear you mumble, it will not be very effective?

USING GEO COORDINATE MAPS TO GET TO THE EXACT POINT

Some items lend themselves to geospatial visualization especially well. For instance in the following figure, the markers indicate where tornados with an F5 strength (200 mph+ winds) occurred. A geo-coordinate map allows the datavis to show the exact location an event occurred. You can imagine users having a particular interest in where a tornado touched down. Moreover, it helps the user understand where in the country the event is most likely to occur.

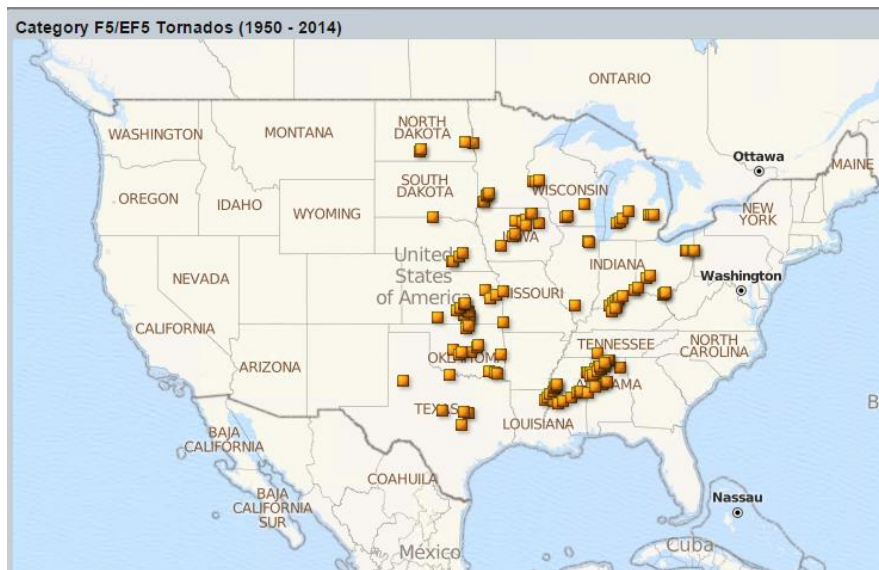


Figure 14 Geo Coordinate Maps Pinpoint Locations on the Map

USING A GEO REGIONAL MAP TO COMPARE REGIONAL AREAS

Using a Geo Region map, you can place a value over the entire region, such as a country or a state. In this datavis, you can see the associated property damage for the tornados. The darker the color the more damage the storm caused. The storm events appeared to have been intense in the southern states but surprisingly Kansas and Ohio had a more costly impact.

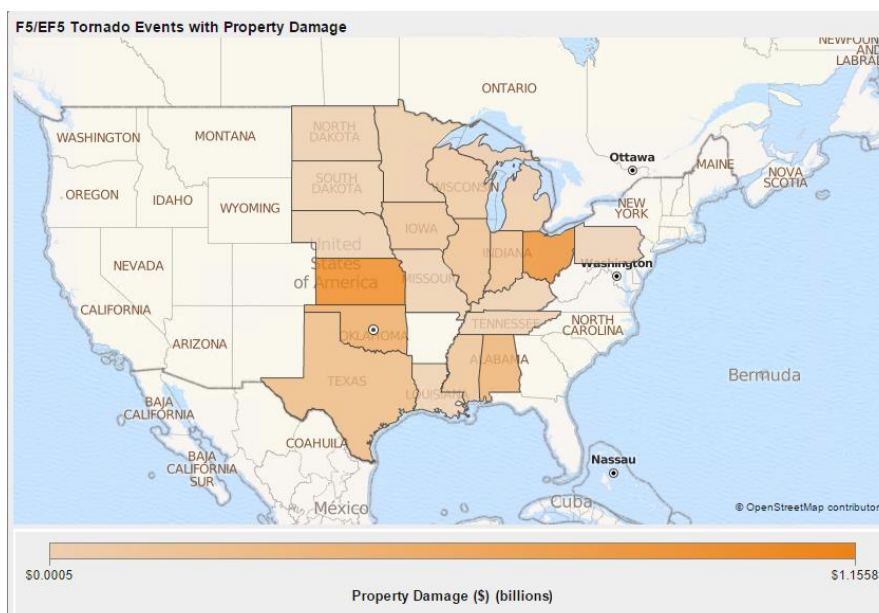


Figure 15 Geo Region maps color the areas for comparison

USING A GEO BUBBLE MAP TO COMBINE DATA

You may find yourself not wanting to compare the previous maps but instead want the data on a single chart, which is what a Geo Bubble Plot allows. The size of the bubble has the count of events while the color explains the estimated property damages. Now it is more apparent that Kansas endured almost as many events as Alabama but endured more damage.

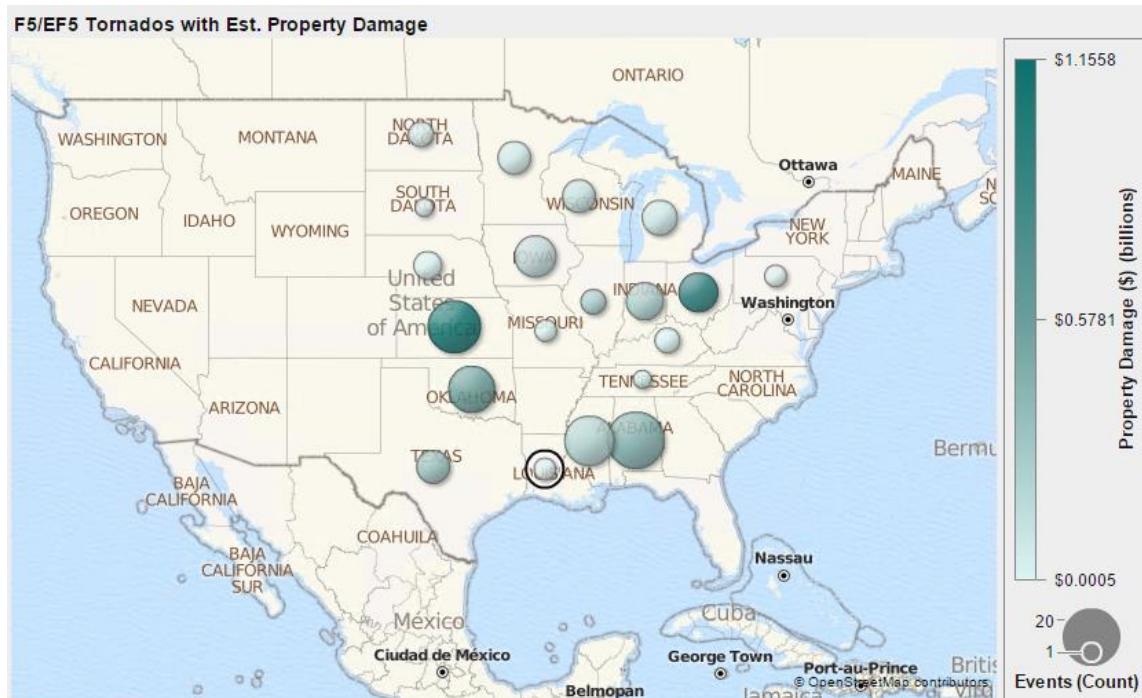


Figure 16 GeoBubble charts allows more data to be displayed at once

CONCLUSION

There are many methods for presenting data to users. The point of this paper is learning what works or is more effective and when to use the methods. Datavis is an iterative process. It is normal to go through many design cycles and revisions for a few charts while other charts will flow into your presentation with ease.

MacGyver was effective because he understood the basics; you now can be effective as well.

REFERENCES

- Bessler, L. 2013. "Data Visualization Tips and Techniques for Effective Communication" PharmaSUG. Available at: <http://www.lexjansen.com/pharmasug/2013/DG/PharmaSUG-2013-DG10.pdf>
- Few, S. 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten (2nd Edition)*. Burlingame, CA Analytics Press.
- Parker, R. 2006. *Looking Good In Print (6th Edition)*, Scottsdale, AZ, Paraglyph Press.
- Tufte, E. 2001. *The Visual Display of Quantitative Information (2nd Edition)*, Cheshire, Connecticut. Graphics Press
- Wong, D. 2010 *Wall Street Journal Guide to Information Graphics*, New York, NY: WW Norton and Company
- "The Magical Number Seven, Plus or Minus Two", Wikipedia. Available at: https://en.wikipedia.org/wiki/The_Magical_Number_Seven,_Plus_or_Minus_Two

ACKNOWLEDGMENTS

Thanks to everyone who assisted with the preparation and ideas in this paper.

RECOMMENDED READING

- McCandless, D. Information is Beautiful, Website: <http://www.informationisbeautiful.net/>

- Simon, P 2014. *The Visual Organization*, Hoboken, NJ, John Wiley and Sons, Inc.
- US Census Bureau. *Data Visualization Gallery*. Available at: <http://www.census.gov/dataviz/>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Tricia Aanderud
Enterprise: Zencos Consulting, Cary, NC
E-mail: taanderud@zencos.com
Web: <http://www.zencos.com>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.