Paper 4836-2020

Making Data Actionable: From Concept to Reality

Dr. Joseph W. Perez, NC Department of Health & Human Services

ABSTRACT

Anyone can make a pretty bar graph, but can you make sound decisions based on what your development staff currently provides? How do you turn a flashy concept into an actionable visualization? Can you see the end result? Will your ideas become reality? Do you have the vision and drive to figure out how to get from today to tomorrow before it becomes yesterday?

American mathematician John Tukey once said, "The greatest value of a picture is when it forces us to notice what we never expected to see." What value do you see in your data? And what ideas do you have when you see it? Learn how you can capitalize on your ideas and turn them into reality by blending the internal with the external, leveraging them into a cohesive strategy for both the short term AND the long term. In this session, see the five "Stages of the Spectrum" in action while discovering the difference between impact and influence, and how that plays into making data actionable.

INTRODUCTION

On May 25, 1961, President John F. Kennedy, in an address to the US Congress, issued a challenge to the entire country when he said, "I believe that this nation should commit itself to achieving the goal, before this decade is out, of landing a man on the moon and returning him safely to Earth." July of 2019 marked the 50th anniversary of the day America met that challenge, just eight years after the President threw down the gauntlet. But this monumental feat could not have been accomplished without using data that was made actionable.

There were innumerable logistical issues to tackle, such as:

- timing the launch when the relative positions of Earth & moon were optimal;
- choosing a landing site that met all the right conditions;
- supplying the needs of the astronauts;
- calculating fuel consumption parameters;
- determining out how to get them there and bring them home safely

At this point, it is important to remember that the microprocessor would not even be invented until 1971; so, the technology needed to conquer all these insurmountable obstacles simply did not exist. Performing a Google search would not provide any help either because neither the Internet (nor, as a result, "Google") had yet to be conceived or implemented. The scientists and other individuals involved had no choice other than to improvise solutions "on the fly" as they came to them.

The above factors exponentially increased the value of having actionable data available during the Apollo program to give NASA the insight needed to answer the incessant questions, make those difficult decisions, and ultimately, meet JFK's challenge.

Techopedia defines actionable insight as "information that can be acted upon or information that gives enough insight into the future, that the actions that should be taken become clear for decision makers."[1] Stated in an alternate way, meaningful data that is useful for:

- making a decision,
- answering a question, or
- solving a problem.

Note that these are the three essential tenets or characteristics of actionable data that will surface multiple times in this paper. It distinguishes data that can improve the overall situation from data that serves as nothing more than fancy window dressing or interesting trivia; i.e., totally useless from a practical standpoint.

In data storytelling, with every data visualization, there will always be a starting point. The developer or designer typically commences with an idea or a concept that exists only in his or her mind. As the end goal, this person will typically envision a completed data visualization containing one or more graphs. This leads to the question: How does one get from concept to reality? The author proposes to view this "***data storytelling journey***" as an imaginary horizontal line (from left to right), or spectrum, along which he suggests five stages for best practices in getting from the idea (on the left) to the completed visualization (on the right). Those five stages comprise the outline for this paper, illustrated as follows:

Figure 1. The Five Stages of the Spectrum

The way the above illustration depicts these stages from left to right across the screen might cause the reader to assume the author intends for them to occur sequentially. While that may be the tendency under normal circumstances, it is also possible for them to share concurrency or even overlap to a certain extent. In any case, it is essential to approach these stages in the light of a holistic strategy and with a positive mindset. Reason: this process should be an enjoyable, positive activity for the developer, data scientist, decision maker, business intelligence specialist, or other analytics professional with a passion for actionable data.

That same mindset should lead the professional observer to the following conclusion: Before a concept can become reality, it needs to be visualized, AND it must stand up to scrutiny.

FIRST STAGE: CONCEPTION (START THE RACE)

Each of the five stages will have a name ("Conception"), an imperative ("Start the Race"), and a key word. For the first stage, Conception, the key word is **ARTICULATE**. When commencing an endeavor (such as the aforementioned "***data storytelling journey***") that has an end goal in mind, viewing it as a race is an appropriate metaphor, especially considering that both have a starting point, a progression, and a finish line indicating completion. While there may or may not be an element of competition, there must still be a good beginning, characterized by having to "articulate" the need. This is where the concept or idea is conceived.

KNOW THE AUDIENCE

In articulating the need, it is important to understand the audience. These are the individuals who are going to consuming the data (often referred to as key stakeholders, users, customers, clients, business partners, or some other term). Visualization expert Tricia Aanderud has recently said, "Your audience determines the depth of your data communication AND the presentation method."[2] For example, it would not make any sense to articulate the need to a room full of developers in the same way as to a group of managers; nor to the end users in the same way as to BI developers, data scientists, DBAs, or anyone else. Since each of these groups has differing needs, agendas, and viewpoints, one must consider these factors and vary his or her approach to accommodate them.

KNOW THE DATA

Equally important is to know the data. How can a data analytics professional articulate anything without knowing the underlying data? Understanding the source, the variables, the units in which it is measured, whether or not it has been aggregated (and how) is an essential starting point. In fact, the most important characteristic to be articulated is the purpose for the data. That is, one must answer the question, "What am I trying to show?"

Dr. Andrew Abela, Chairman of the Department of Business & Economics at the Catholic University of America in Washington, DC, created a graphics tool[3] that masterfully articulates the need and sets the pace with four categories that he identified for data storytelling. Although designed more than a decade ago, this tool has stood the test of time. To use these four categories, the question becomes, "Am I showing comparison, distribution, composition, or relationship?" The answer to that question determines the type of visualization that would be the most appropriate for the story the analytics professional is trying to tell with their data. This, in turn, leads to even more questions that guide the analytics professional to refine their approach along the way in a hierarchical methodology.

Category Being Shown	Question to be Asked	Key
Comparison	Am I comparing among ***items*** or over ***time***?	Behavior
Distribution	Among ***how many*** variables is this being distributed?	Dimension
Composition	Does it ***change*** over time or is it ***static*** over time?	Structure
Relationship	How many variables are being ***related***?	Class

Figure 2. The Four Categories of Data

Abela's "Chart Chooser" (as shown in Figure 3) allows the analytics professional to ask these questions after considering what they're trying to show. This allows them to proceed along a type of decision tree to arrive at the appropriate type of graph that is considered "best practice" for the situation in question.

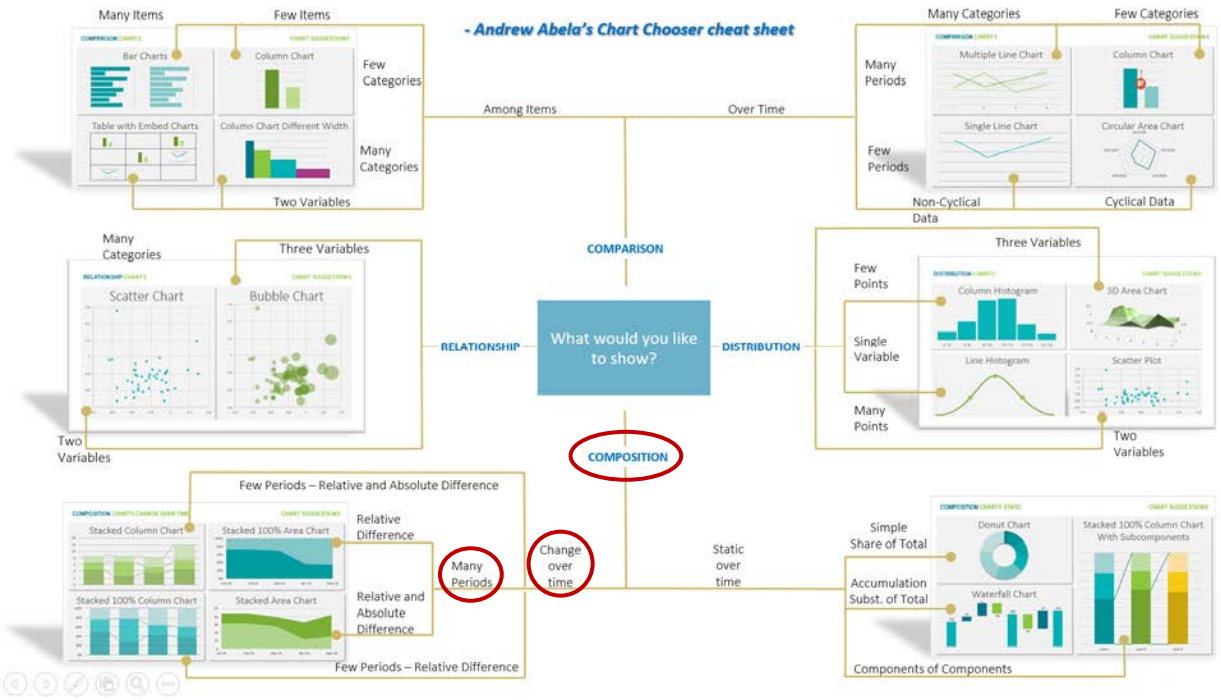

Figure 3. Abela's "Chart Chooser"

The above example shows how one might take "Composition" as an example and answer the questions in the decision tree to inform their decision regarding the type of graph that is the most appropriate for the given situation.

SECOND STAGE: INCEPTION (MAKE THE CASE)

After starting the race, the next stage involves making the case, with "**ADVOCATE**" as the key word. The analytics professional must make the case for expressing the concept visually. Two reasons for this imperative stand out, as shown by Mike Parkinson, CEO and founder of Billion Dollar Graphics, who advocated that humans are affected by graphics in two ways: cognitively and emotionally.

COGNITIVELY

Parkinson said, "Graphics expedite and increase our level of communication. They increase comprehension, recollection, and retention. Visual clues help us decode text and attract attention to information or direct attention increasing the likelihood that the audience will remember."[4]

EMOTIONALLY

Parkinson also said, "Pictures enhance or affect emotions and attitudes. Graphics engage our imagination and heighten our creative thinking by stimulating other areas of our brain (which in turn leads to a more profound and accurate understanding of the presented material)."[4]

To summarize Parkinson's assertion (about graphics affecting us both cognitively and emotionally), graphics get ahold of both our brains and our hearts. Any educational endeavor that manages to touch both the brain and the heart is bound to be successful.

FURTHER PROOF

British data journalist David McCandless said in his 2010 TED talk ("The Beauty of Data Visualization")[5] that sight, by far, has the highest bandwidth of any of the five senses. "About 80% of the information we take in is by eye," he said. Additional research in the years since McCandless' statement suggests that the number is likely closer to 90%. Two European research professors, Robin Hogarth and Emre Soyer, teamed up on an experiment in which three groups of economic professionals were asked the same question regarding a specific dataset, the results of which appear to bring credence to McCandless's TED talk claim. The first group of economists was given the data and an analysis of the data. 72% of them gave an incorrect answer. A second group was given the data, the analysis, and a graph. Although they did better than the first group, 61% of those economists were incorrect. The study instrument provided only a graph to the third group (no data, no analysis), and a mere 3% of them answered the question incorrectly.[6]

Figure 4. Results of Soyer/Hogarth study

These results "make the case" for visual representation of data and as such, should be advocated for an effective business intelligence and analytics reporting strategy.

THIRD STAGE: PERCEPTION (MOTIVATE THE BASE)

Having started the race and made the case, it is time to motivate the base. The key word is **EDUCATE**. Progress in the data storytelling journey is not likely if either the target audience or the stakeholders are not on board, so they must be educated; i.e., be brought to realization that they do indeed have a stake in this process. These individuals are not likely to take ownership or provide the necessary backing (financially or otherwise) if they do not understand what is being presented, or if they cannot see the value for their own agenda(s), or if the visualization does not help them to make a decision, answer a question, or solve a problem.

HOW TO ACCOMPLISH: FORM VS. FUNCTION

Understanding basic principles (and adhering to them) is foundational to any educational initiative and is certainly a major consideration in the data storytelling journey. Quite often, comments are heard such as: "He's all talk & no action," or "She's all fluff & no substance," or "This project is nothing but form w/o function." This begs the question, "Does form follow function?" – to which the answer is, "Yes."

This may lead the reader to ask for a definition of form and function. Function is what is being done with the data, whereas form is how to express functional outcomes in the best way. It brings up the age-old argument of "Is it pretty?" (form) vs. "Is it right?" (function)

There are some best practices or "cardinal rules" that should be kept in mind to ensure function gets the right priority over form in data storytelling. The following subsections include examples of pitfalls to avoid and how to deal with them.

PIE CHARTS

Many have a "love-hate" relationship with pie charts. Their use is typically praised enthusiastically when properly done and vilified vehemently when disaster strikes. When used correctly, pie charts are an effective way of showing the parts of a whole and how those parts relate proportionately to the whole. Unfortunately, not everyone understands this concept, as shown in the following example.

- The numbers in the pie chart add up to 193 (which is greater than 100). It is impossible for parts of a whole to add up to a number greater than that whole.
- The slices are disproportional (i.e., 70 is greater than 60, but the 60 and 63 slices appear larger than the 70 slice).

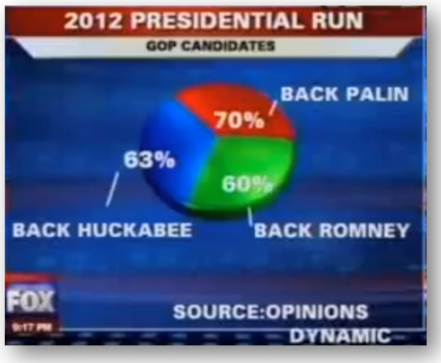

Figure 5. screen shot from 2012 news report with inaccurate pie chart

For the purpose of full disclosure, the survey questions used to assemble the news report on which the above graph is based allowed for multiple choices. That is, respondents were able to indicate support for more than one presidential candidate at a time. Although this clearly explains why the percentages add up to a number greater than 100, it does not justify the improper use of a pie chart.

Therefore, the above chart in Figure 5 can safely be judged to be misleading. In this case, a bar graph would be better suited to present the data accurately, as shown in Figure 6 at right. It is the same data as was shown in Figure 5 above but presented in a way that is not misleading.

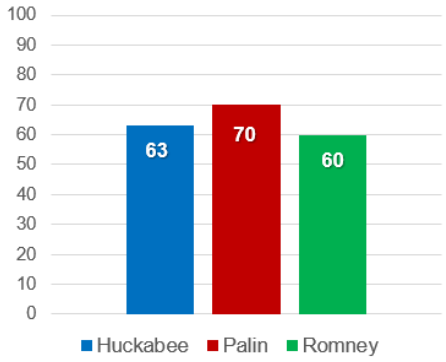

Figure 6. Bar graph to illustrate the correct use of the pie chart data

LINE GRAPHS AND BAR CHARTS

CONSISTENCY OF SCALE

For line graphs and bar charts, a best practice is to ensure the vertical axis has a consistent scale. In figure 7 below, the sample line chart on the left shows how the number of people aged 25 & older as a percent of total college enrollment has changed over time. A cursory glance might lead to the assumption of a huge increase between 1973 and 1974; however, a closer look will show that the vertical axis is scaled from 28 to 33 percent, thus making the change between those two years appear to be more drastic than it really is. Therefore, this line chart can safely be judged as misleading.

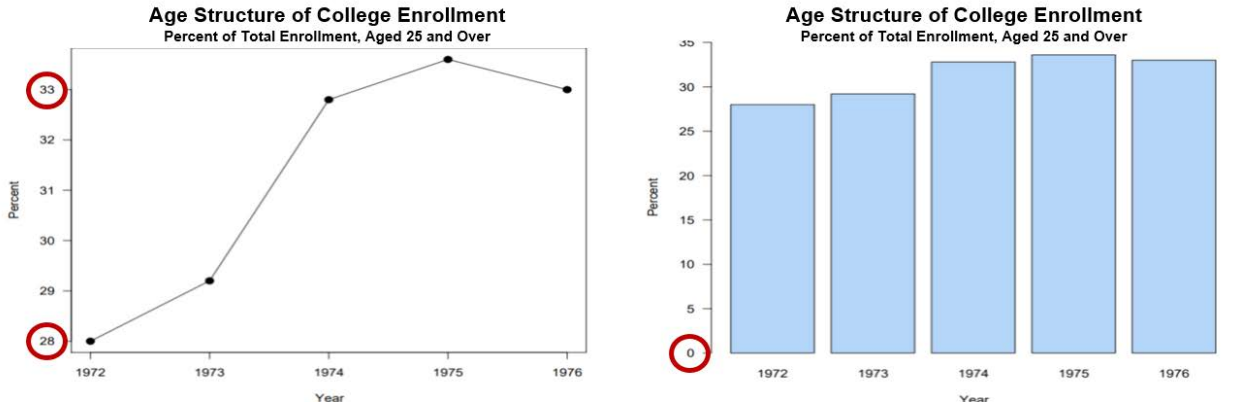

Figure 7. Line and bar charts with inconsistent and consistent Y-axis scale

By contrast, the bar graph on the right (Figure 7, above) shows the same data, but with the vertical axis scaled to start from zero. This graph shows a more realistic depiction of the percentage trend over that same five-year period; therefore, it can be judged to be **not** misleading.

Why is this done? Many organizations purposefully represent their graphics in a particular way with no regard for statistical accuracy. In fact, specific misrepresentation of data occurs in many cases for self-serving purposes; i.e., the only interest is to make the graph "look good" and convey the message they want it to convey, rather than letting the data speak for itself. Data analytics professionals should hold themselves to a higher standard of accuracy and integrity. That said, it does not necessarily mean that the Y axis scale should always be forced to start at zero in every case. Sometimes when multiple datasets having multiple series are being compared (especially those in which the numbers are very close together), the only way one CAN see a difference is to zoom in on the scale. However, it is best to employ consistent methodologies across the board and avoid making changes over time appear to be more drastic than they really are.

PROPER LABELING

Not only should the Y axis be scaled consistently, but it should also be labeled properly. As shown in Figure 8, the chart at right is missing the label on the Y axis. Although this may not seem important, a closer look reveals that the last bar on the right appears to be several times larger than the first one on the right. However, upon comparing the data labels, it is evident that 2062 is not several times larger than 1371 (it is not even double). The absence of the label for the Y axis accentuates the fact that the bars are disproportionate to the quantities they represent. As with the other "wrong" charts, this one is clearly misleading.

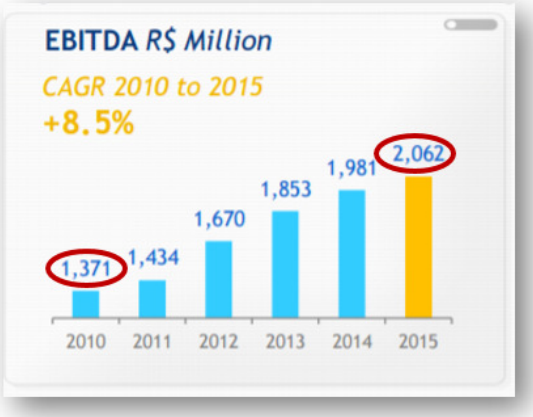

Figure 8. Bar chart with missing label on Y axis

OTHER BEST PRACTICES FOR INFLUENCING PERCEPTION

Chronology: When plotting chronological elements, they should be depicted from left (the past) to right (the present and future) along the X axis.

Quantity: When measuring quantities and other numerical items, plot them along the Y axis, with the smaller number at the bottom and the larger numbers at the top.

Color: Colors should be used consistently, remembering that there are some universally accepted constants, the most familiar ones of which are that RED means stop, danger, decline; whereas GREEN means go, safety, growth. Although there is a rising view that a more muted palate should be employed (for accessibility concerns), the overriding principle still applies. According to Spencer Sobczak, Enterprise Account Executive at Theia, "Contrasting colors differentiate between groups, while different shades of the SAME color can communicate the strength or weakness of a value."[7]

Shapes: Graphical elements used in visualizations should be kept proportional to each other. Master statistician and Yale professor Edward Tufte advocated not exceeding what he called the "Lie Factor." It states that the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented. As a mathematical equation, it is often expressed like this: the size of the effect shown in the graphic divided by the size of the effect shown in the data should never be greater than ONE.

To illustrate, in one of the iterative product releases for a new iPad model several years ago, Apple touted its greater battery capacity over its predecessor. This was no empty claim, for a 17% increase in battery capacity is significant. However, the graph in Figure 9 at right is misleading. The battery icon on the right, representing the capacity of the new iPad, appears to be three or four times larger than the one on the left representing the predecessor, disproportionately more than 17%.

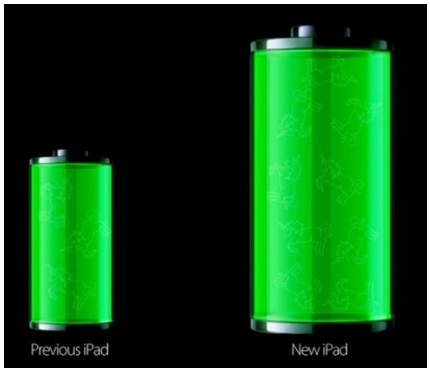

Figure 9. Apple visualization that exceeds the "Lie Factor"

FOURTH STAGE: INSPECTION (GIVE IT SPACE)

Being totally transparent helps avoid the pitfalls of bias and unethical use of data upon inspection. When analytical professionals have done their due diligence, they will neither be bothered nor intimidated by questions such as these:

- Did you ensure data quality?
- Where did you get your data?
- How do you know it's right?

IMPACT VS. INFLUENCE

As the data behind a visualization project is held up to inspection, it is important to be aware of the difference between impact and influence, and how that plays into making data actionable.

IMPACT

Impact is typically recognized as being an effect brought on by EXTERNAL factors that PUSH on the status quo. In a 2015 Harvard Business Review study, the following was found: "Nearly 85% of a company's performance is dependent upon external factors. Yet… many companies don't know where to look to determine which external drivers are affecting business performance."[8] Those who make it a priority to identify those factors are PUSHED into acting upon them, thus making an impact in their organization.

INFLUENCE

By contrast, influence is generally known to be the power to affect from INTERNAL factors that PULL on the status quo. These are factors that an organization can control internally. Such factors include data structure / organization, the frequency of data collection and assimilation into the data warehouse, and the cultural climate in the organization. Recognizing and dealing with these factors will tend to PULL the organization's strategy in the right direction from the INSIDE.

To put it bluntly, an idea must be VISUALIZED, then SCRUTINIZED before it can be REALIZED. In other words, the organization that has given the data some "space" and ensured it can stand up to inspection is well on its way to turning concepts into reality with actionable data.

A COMBINATION FOR SUCCESS

It is possible to measure success in this area by capitalizing on this knowledge and turn ideas into reality by blending the external with the internal, leveraging them both into a cohesive strategy to meet short-term needs AND provide long-term benefit.

FIFTH STAGE: DIRECTION (PUT IT INTO PLACE)

The key word here is **EFFECTUATE**; that is, ensuring the right direction for the solution to be correctly put into place. When data has been scrubbed, correct visualizations/charts have been chosen, and data validity/integrity has been assured, the project is ready for implementation in what the author of this paper sees as a four-fold iterative cycle.

The cycle starts by **DEPLOYING** the solution, using whatever organizational protocols and/or methods already in place (whether it be waterfall, agile, Kanban, etc.).

As the stakeholders start using this solution, they will **DISCOVER** issues and report them to the appropriate individuals as issues, defects, problems, or whatever nomenclature in use, using the existing organizational issue reporting mechanism.

The analytics professional will then analyze those issues and **DISCERN** the right course of action to address them, taking whatever functional, technical, and business requirements made available by the stakeholders into account.

Finally, the analytics professional will assign the appropriate individual or team (in which they may be included themselves) to begin **DEVELOPING** the remediation for those issues, ensuring all requirements, specifications, constraints, and deadlines are observed and met.

At this point, the cycle may start all over again (**Deploy**, **Discover**, **Discern**, and **Develop**) with the goal being to improve upon the solution with each iterative cycle. This approach ensures an effective direction because it leads to progress.

CONCLUSION

In the data visualization journey from concept to reality, as the visualizations are developed and moved along the five stages of the spectrum, it is imperative that the analytics professional ensure best practices are followed. American mathematician John Tukey once said, "The greatest value of a picture is when it forces us to notice what we never expected to see." The question can be asked: What ideas are generated when the underlying data is seen for what it really is? To stimulate these ideas, remember that graphs and data visualizations…

- are only as good as the data they portray,
- should be no more complex than the data they portray, and
- should never paint a distorted picture of the values they portray.

With actionable data, NASA exceeded President Kennedy's monumental challenge more than fifty years ago and not only sent astronauts to the moon successfully, but also returned them to earth safely. Moreover, they repeated this process five more times over the next three years. The author of this paper encourages the reader to turn ideas into reality by making data actionable.

CONTACT INFORMATION

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

Dr. Joseph W. Perez ("Joe")
jwperez@gmail.com
https://www.linkedin.com/in/jwperez
https://sessionize.com/joe-perez

REFERENCES

1. Techopedia Definition. "Actionable Insight." Available at https://www.techopedia.com/definition/31721/actionable-insight.
2. D'Agord, Jaime, "How You Communicate with Data: Be More Effective & Influential." Zencos. April 22, 2019. Available at https://www.zencos.com/blog/identify-audience-message-methods-data.
3. Abela, Andrew. "Chart Suggestions – A Thought Starter." January 2009. Available at https://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf.
4. Larson, Troy. "Powerful Facts About Visual Communication." November 2011. Available at https://www.mindjet.com/blog/2011/11/powerful-facts-about-visual-communication/.
5. McCandless, David. "The Beauty of Data Visualization." TEDGLobal 2010. July 2010. Available at https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization.
6. Ziliak, Stephen, Visualizing Uncertainty: On Soyer's and Hogarth's 'The Illusion of Predictability' (February 2, 2012). International Journal of Forecasting, Forthcoming. Available at SSRN: https://ssrn.com/abstract=1998149.
7. Sobszak, Spencer, "The Importance of Color in Data Visualization and Dashboard Design." February, 2019. Available at https://storybydata.com/datacated-challenge/the-importance-of-color-in-data-visualization-and-dashboard-design/.
8. Staff writer. "The Top 10 External Factors That Impact Forecast Accuracy." January 2016. Available at https://www.prevedere.com/the-top-10-external-factors-that-impact-forecast-accuracy/.