A DATA PRACTITIONER'S GUIDE TO TELLING STORIES WITH DATA VISUALIZATIONS

How to create data visualizations that tell the story hidden in your data

THE NEED FOR DATA STORIES

- *"Data stories combine visualizations with narrative flow. This combination can breach the barriers between people and data, engaging the former and delving deeper into the latter.*
  (James Richardson, Research Director at Gartner)"
- *"An ancient Jewish folktale dating back to the eleventh century captures the essence of why we must tell stories with our insights. Truth, naked and cold, had been turned away from every door in the village. Her nakedness frightened the people. When Parable found her, she was huddled in a corner, shivering and hungry. Taking pity on her, Parable gathered her up and took her home. There, she dressed Truth in story, warmed her and sent her out again. Clothed in story, Truth knocked again at the villagers' doors and was readily welcomed into the people's houses. They invited her to eat at their table and warm herself by their fire.*
  (Simmons, A. 2009. The Story Factor: Secrets of Influence from the Art of Storytelling. New York: Basic Books.)"

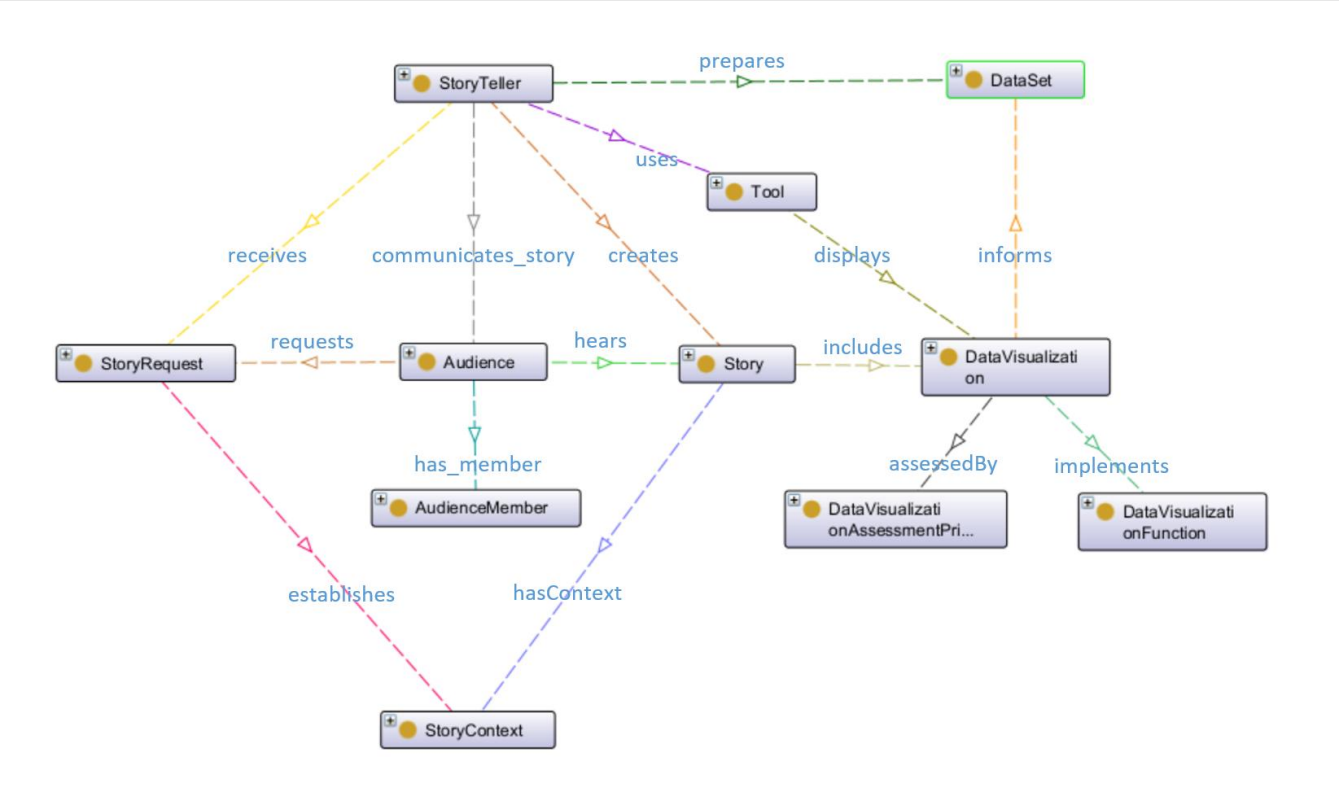
  - Dykes, Brent. Effective Data Storytelling (p. 119). Wiley. Kindle Edition.

THE NEED FOR A DATA PRACTITIONER TO BE A STORYTELLER

- Beth Stackpole in an MIT Sloan School 2020 post lays out the need for a Data Practitioner to be a storyteller to fill the gap between mastery of all aspects of data wrangling and the ability to convey contextually relevant information addressing an "audience's" need for knowledge.
- A James Cook University blog post (author unknown) puts it this way:
  - Without data storytelling, businesses are unlikely to have the 'a-ha' moments they need and want. Data storytelling helps the data feel alive and ensures that the message it's conveying is meaningful and relevant. Critical information may be lost without data storytelling and – importantly – data not encased in story form may not inspire action and change.
  - Data storytelling can also be a way to standardize the communication of data, to ensure that it can reach as broad an audience as possible.
  - In sum, data storytelling is beneficial because it:
    - Provides new perspectives.
    - Ensures that key findings are as accessible as possible, to as many people as possible.
    - Assists others with interpreting otherwise complex information.
    - Can help inspire emotions and ultimately meaningful action.

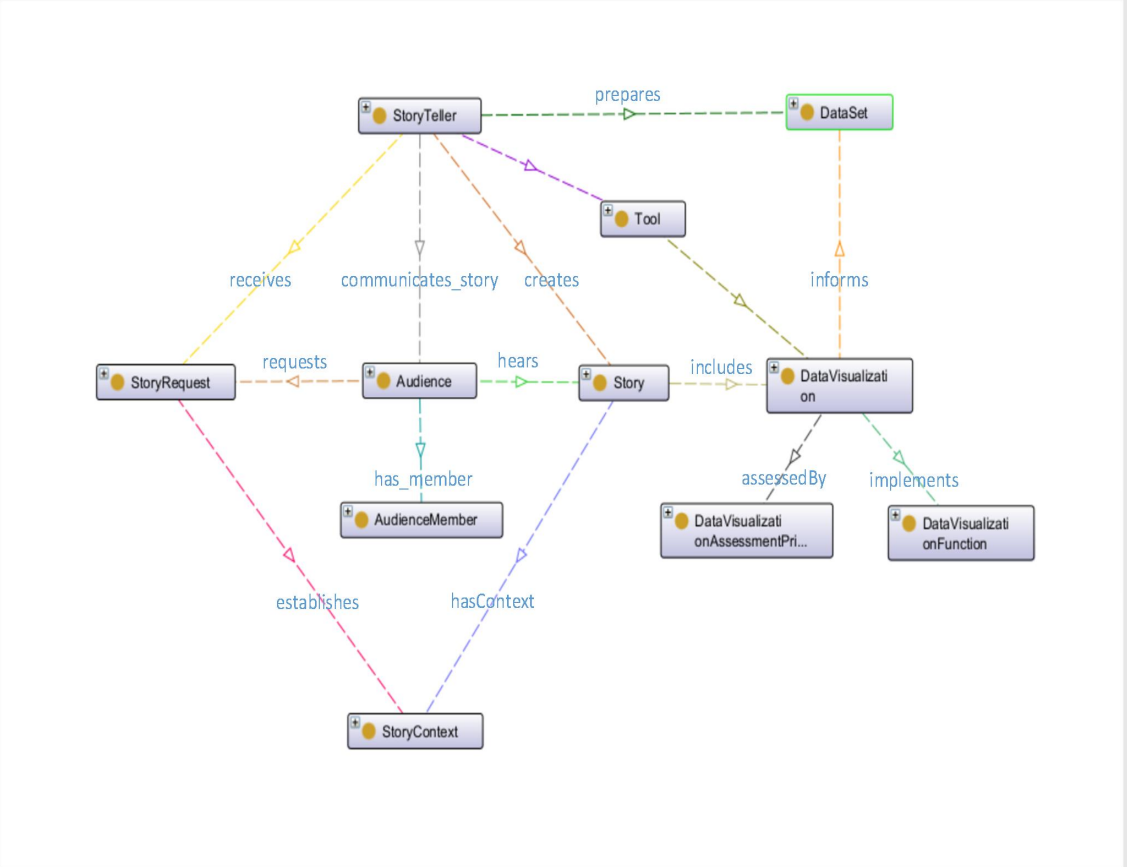
DATA VISUALIZATION MODEL

This model, created in Protégé, defines the Data Practitioner's role as a "Storyteller" and depicts the principal classes and relationships involved in understanding requirements and the selection and use of suitable tools to generate Data Visualizations that inform the meaning of data comprising elements of the story.



THE AUDIENCE

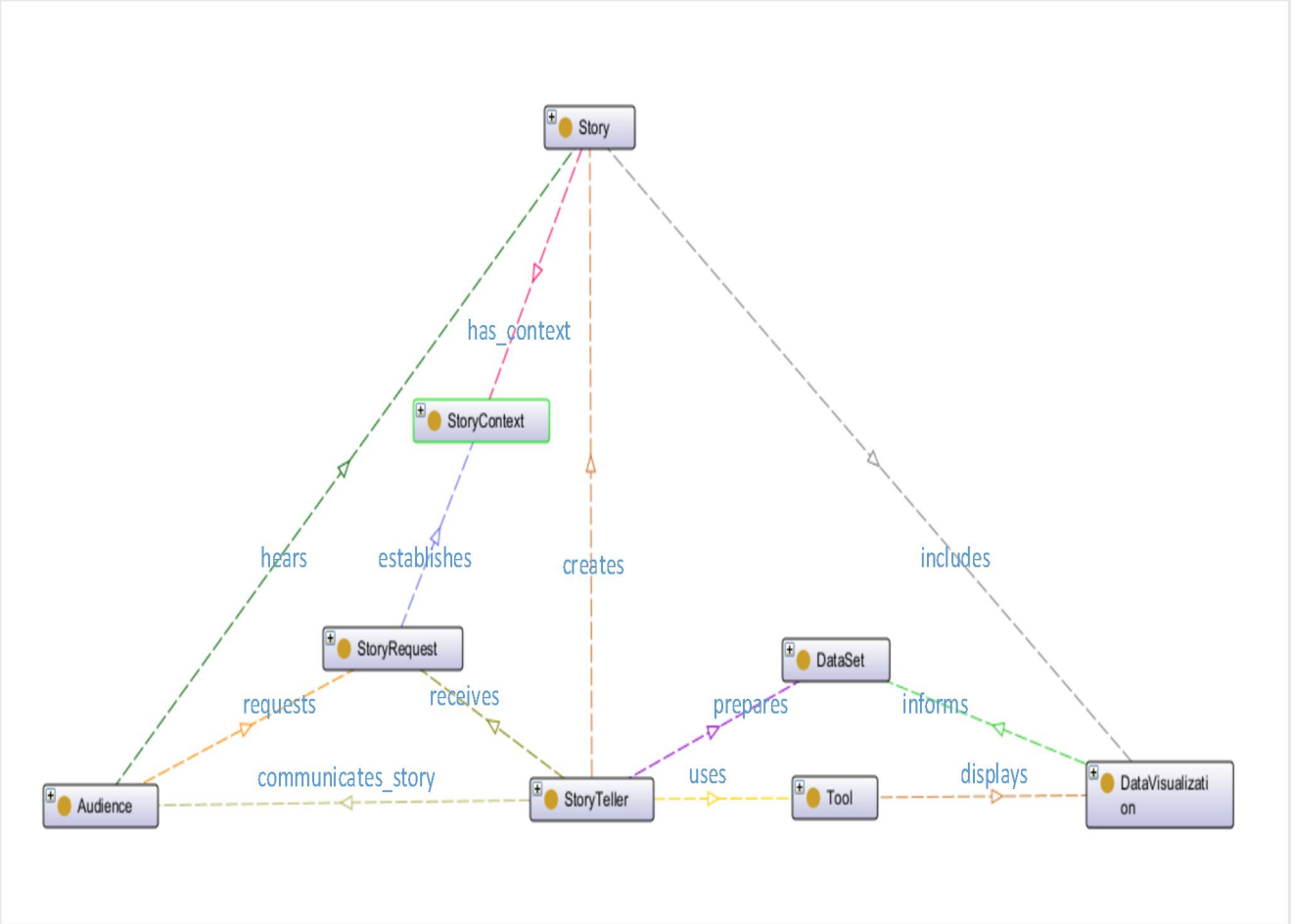
- **WHO –** Know the audience for the story!
  - A specific group of individuals or incumbents of specific roles
  - Readers of a publication, blog or other media
- **WHAT –** Know the required content of the story!
  - Know what they need to know, and what they want to hear - Often not the same.
- **WHY –** Understand the story context!
  - Understand the big picture as well as the underlying, often political issues.
- **WHEN –** Know the story time requirement!
  - Immediately - On a normal repeated cycle - When it's convenient
- **HOW –** Understand the right story modality!
  - Written report - Presentation - E-mail - Blog Post - YouTube - Research Paper - Other



THE STORYTELLER

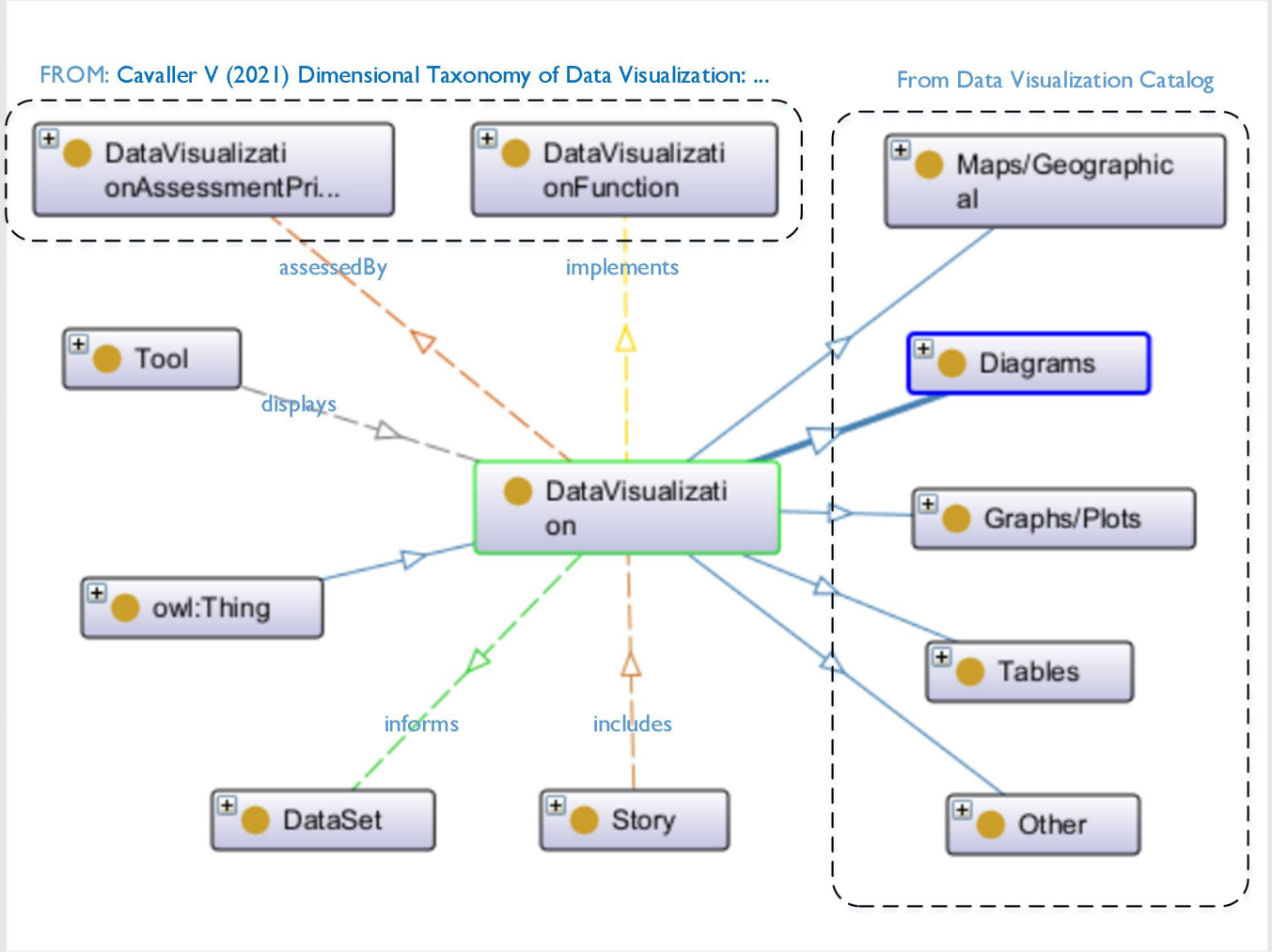
Having received a story request and understood its context the storyteller must:

- Identify and prepare required data-set(s)
- Select appropriate tools and create needed data visualizations
- Identify appropriate delivery modalities for the target audience
- Create and communicate the story



THE DATA VISUALIZATION

- Data Visualizations inform (give shape and substance to) the underlying data.
- They must be faithful in their representation of the data.
  - Omit nothing that is in the data
  - Include nothing that isn't in the data



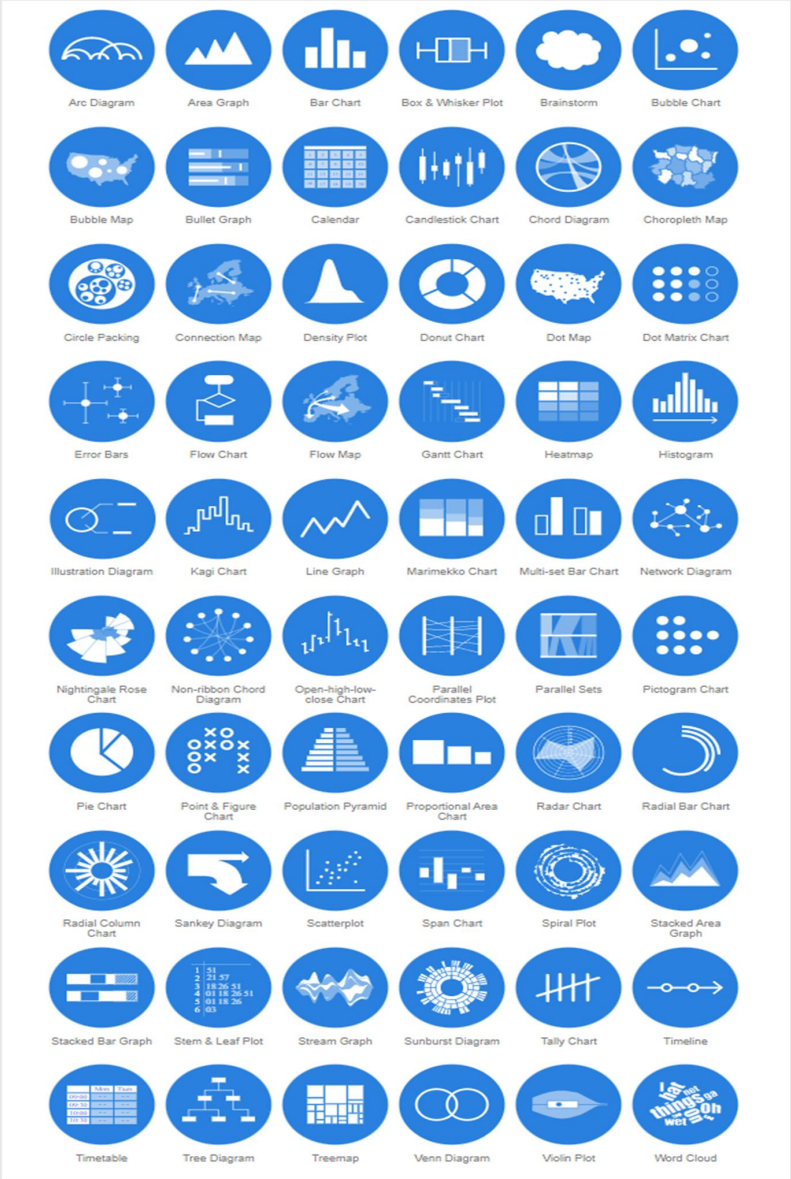
DATA VISUALIZATION QUALITY FEATURES

- Fidelity - Above all else, a Data Visualization must be a faithful representation of the underlying data, not leaving out what is in the data and not adding anything that is not.
- Simplicity - A quality Data Visualization should be as simple as possible, but not too simple (paraphrasing A. Einstein's description of the best solution to a problem). Leave out irrelevant embellishments.
- Utility - A quality Data Visualization must be fit for purpose as defined by the viewer's needs and not encumbered by extraneous information, as interesting as that might be.
- Saliency - A quality Data Visualization focuses the viewer's attention on the most important facts and conclusions, supported as necessary by secondary, supporting information and details
- Efficacy - A quality Data Visualization accomplishes the purpose for which it was designed, its context. Knowing that context is key to success.
- Uniformity - A quality Data Visualization or set of graphics adheres to a predetermined presentation theme.
- Amity - A quality Data Visualization is user friendly - it is easy for the viewer to grasp the message with the least required expenditure of mental effort.

VISUALIZATIONS IDENTIFIED IN THE DATA VISUALIZATION CATALOG

HTTPS://DATAVIZCATALOGUE.COM/INDEX.HTML

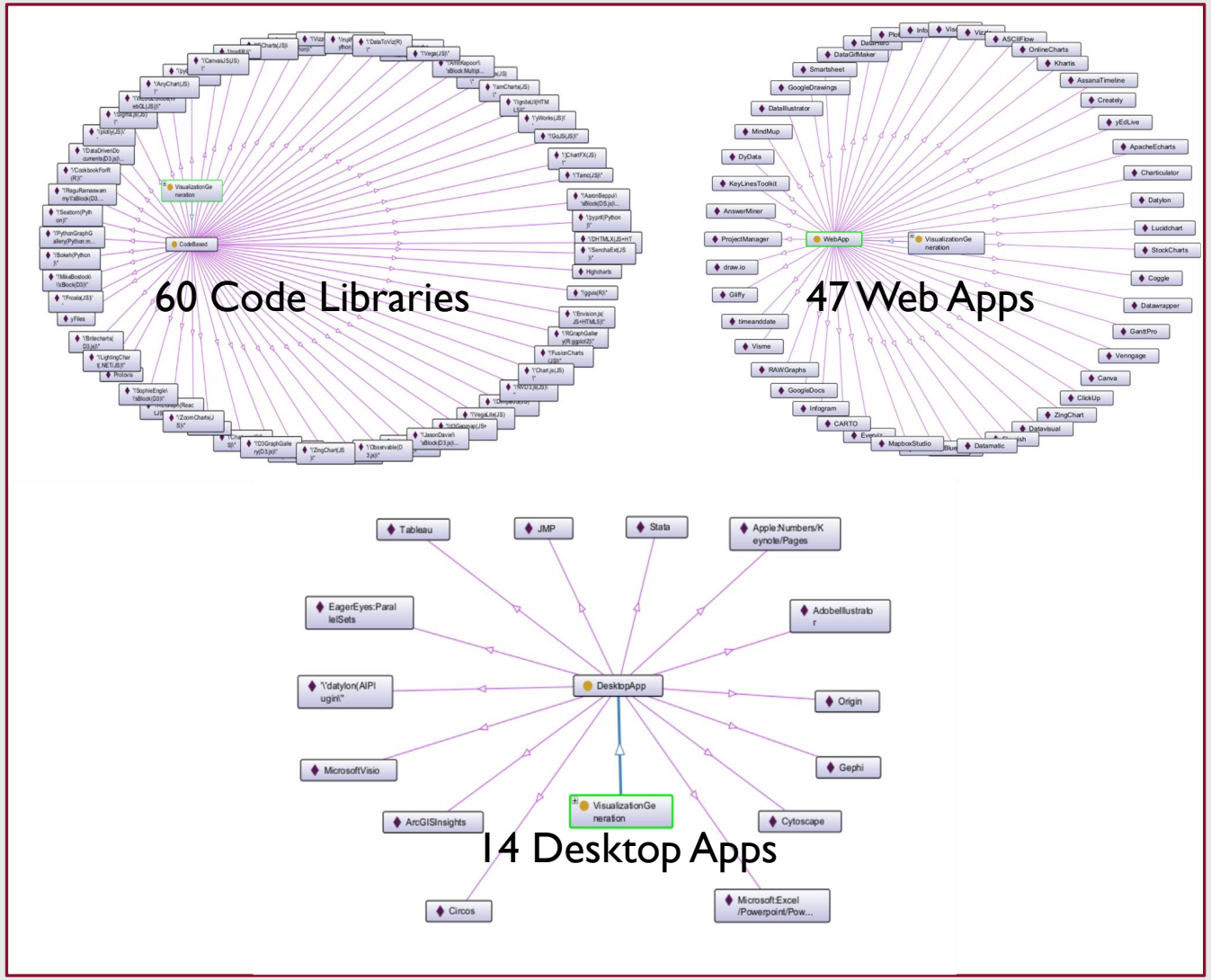
The data visualization catalog identifies and classifies 60 different types with multiple sub-types of many of them. You will find that in most of your work as a data practitioner a small sub-set of these is more than adequate. Your familiarity with the broader set of graphic types enables you to adapt their use when circumstances demand it.



DATA VISUALIZATION TOOLS

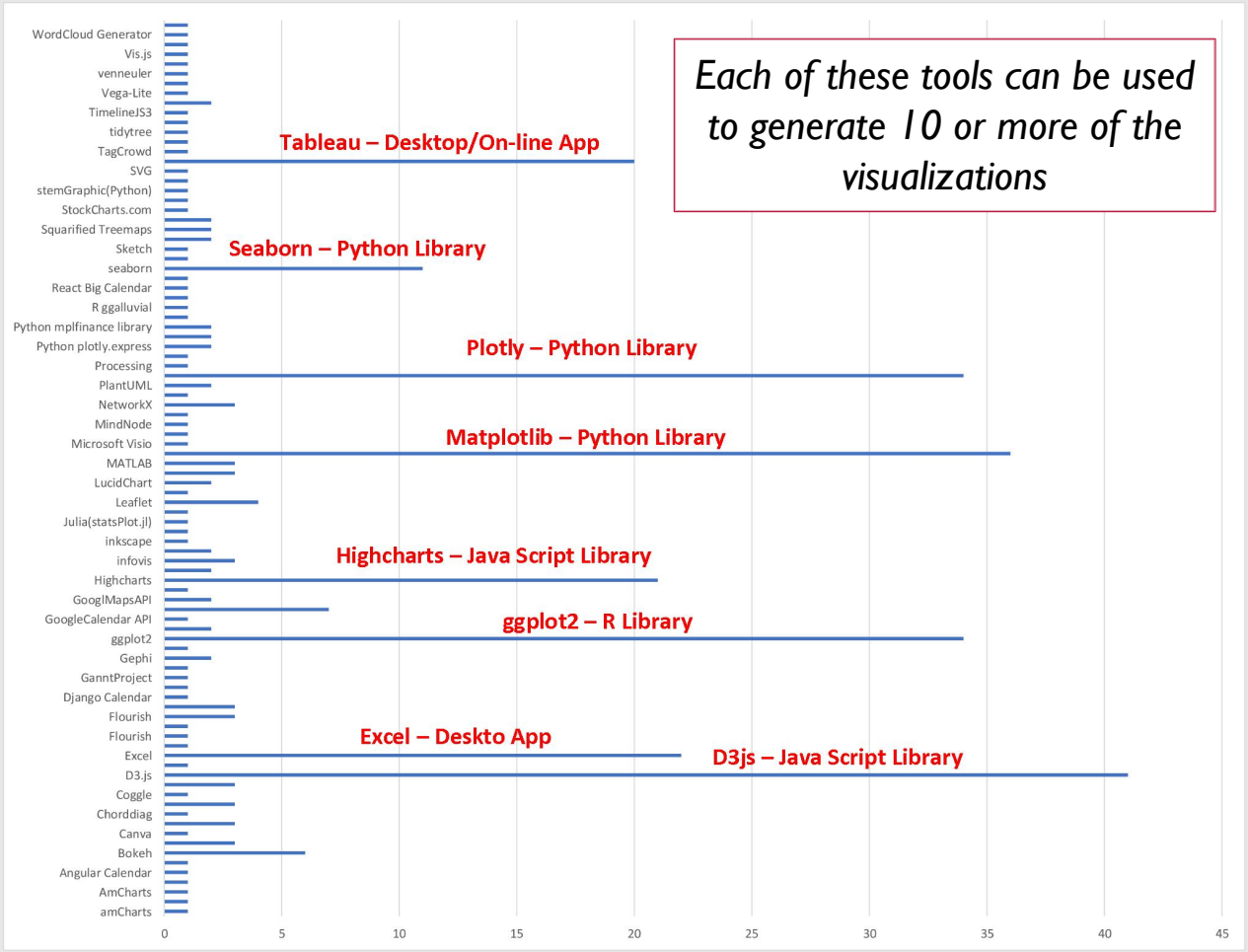
An even larger number of code libraries, web apps, and desktop apps are available. More than 120 are identified in the catalog.

Only a small number of these are sufficiently comprehensive for you to make them part of your core competency.



COMPREHENSIVE DATA VISUALIZATION TOOLS & LIBRARIES

Analysis of the Data Visualization Catalog inventory of 60 visualizations and 122 tools enables us to identify a combination of eight comprehensive applications and libraries. Specialized tools are available for geographic data and visualizations, and infrequently used one-of-a-kind visualizations.



TUFTE'S PRINCIPLES

A SET OF PRINCIPLES FORMULATED AND ESPOUSED BY EDWARD R. TUFTE IN HIS SEMINAL WORK, "THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION"*

*TUFTE, EDWARD R: THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION, 2ND EDITION; GRAPHICS PRESS LLC, 2007

PRINCIPLES OF GRAPHICAL EXCELLENCE

- Graphical Excellence is the well-designed presentation of interesting data – a matter of substance, of statistics, and of design.
- Graphical Excellence consists of complex ideas communicated with clarity, precision, and efficiency
- Graphical Excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
- Graphical Excellence is nearly always multivariate; and, graphical excellence requires telling the truth about the data

PRINCIPLES OF UNIFORMITY OF GRAPHICS

- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented
  - Lie Factor = (size of effect shown in graphic) / (size of effect in data)
- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.
  - Write out explanations of the data on the graphic itself.
  - Label important events in the data
- Show data variation, not design variation

PRINCIPLES OF DATA VARIATION AND CONTEXT PRESERVATION

- In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
- Avoid using areas (two dimensional images and graphics) to display one-dimensional data
- Graphics must not quote data out of context

PRINCIPLES OF GRAPHICAL DESIGN

- Above all else show the data.
- Maximize the data-ink ratio, within reason
  - Data-Ink Ratio = Share of "ink" devoted to the data
- "Erase" non-data-ink, within reason
- "Erase" redundant data-ink, within reason
- Revise and edit
  - Eliminate "chart-junk"
  - Mobilize every graphical element, perhaps several times over, to show the data
  - Maximize data density and the size of the data matrix, within reason
  - Graphics can be shrunk way down

PRINCIPLES OF GRAPHIC DESIGN AESTHETICS

- Have a properly chosen format and design
- Use words, numbers and graphic elements together in concert
- Reflect a balance, a proportion, a sense of relevant scale
- Display an accessible complexity of detail
- Have a narrative quality, a story to tell about the data
- Pay attention to technical details and production to produce professional high-integrity displays of statistical information
- Avoid content-free decoration, including chart-junk
- Data Visualizations are paragraphs about data and should be treated as such

CAVALLER'S PRINCIPLES

CAVALLER V (2021) DIMENSIONAL TAXONOMY OF DATA VISUALIZATION: A PROPOSAL FROM COMMUNICATION SCIENCES TACKLING COMPLEXITY; (FRONT. RES. METR. ANAL 6:643533, DOL: 10.3389/FRMA, 2021.643533

CAVALLER'S COMMUNICATION THEORETICAL FRAMEWORK FOR DATA VISUALIZATION

- "Data visualization constitutes a process of communication, the efficiency of which is conditioned by the actions that these elements imply: the selection of the content, the formal representation of the information, the encoding and setup of the visualization, the graphical design appropriate to the context, the adaptation to the medium, and the observation of user preferences. Furthermore, understanding the completion of these actions as a critical success factor, they must be undertaken considering their interconnection which plays a critical role and can be expressed by means of the following practical questions:

  1) What content does the sender want to communicate and to what degree of abstraction

  2) In which form? Which functionalities from which tools are appropriate for the graphical representation to be integrated in the pursued channel?

  3) Once content and form are defined, what specification must be applied to the setup of both data and graphical representation in order to adapt to each other?

  4) What are the approach mode and the graphical design suitable to the context? What properties does the visualization have to meet depending on the target or audience?

  5) What characteristics must the visualization contemplate in order to make it efficient according to the media where it is projected? What are the levels of communication efficiency that must be achieved?

  6) What requirements must be observed from the user's experience in order to improve understanding of the topic?"

CAVALIER'S PRINCIPLES FOR EVALUATING DATA VISUALIZATIONS

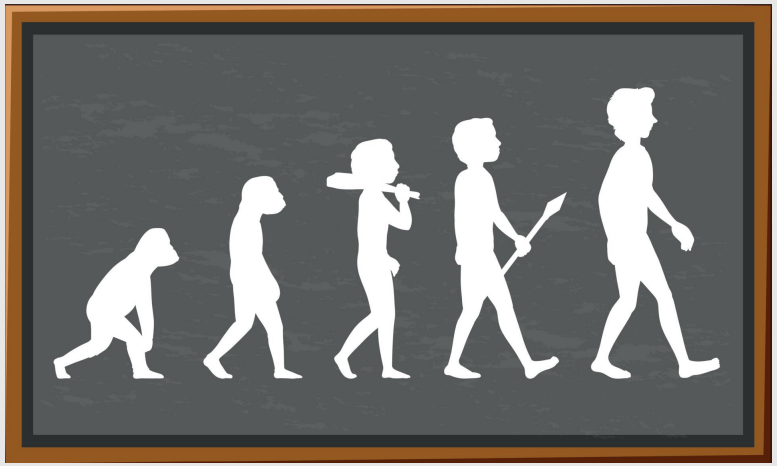
1. "The principle of analytical interest states that data visualization is right in so far as it keeps scientific rigor, order, and method in the quantitative procedures."
2. "The principle of functional or pragmatic interest states that data visualization is right in so far as the graphical representation has a practical utility and added value over other communicative forms facilitating their comprehension."
3. "The principle of managerial interest states that data visualization is right in so far as it is able to package data message and graphic representation in a singular configuration that promotes the understanding of a meaningful communication."
4. "The principle of interest for efficacy states that data visualization is right in so far as, taking into account the professional, social and cultural context and target; it produces the intended communicative result by a suitable design."
5. "The principle of interest for efficiency states that data visualization is right in so far as it achieves the communication goals by the optimal means of communication with maximum benefits and minimal use of resources."
6. "The principle of appraisal interest states that data visualization is right in so far as it receives a positive assessment from the user in terms of usability and of other factors related to H–M interaction."

DATA VISUALIZATION IMPROVEMENT TECHNIQUES

The literature on Data Visualizations and Graphic Design abounds with examples of weak and confusing graphics combined with improved versions as examples of WHAT NOT TO DO.

Here we will provide 10 essential techniques, examples of WHAT TO DO to improve data visualizations.

PRECONSCIOUS VISION IS A PRODUCT OF EVOLUTION



- Over many millennia our senses evolved to address our basic needs
  - Survival (avoiding predators and environmental threats)
  - Sustenance (finding and gathering food and hunting prey)
  - Procreation (attracting and securing a mate).

- Preconscious visual attributes include*

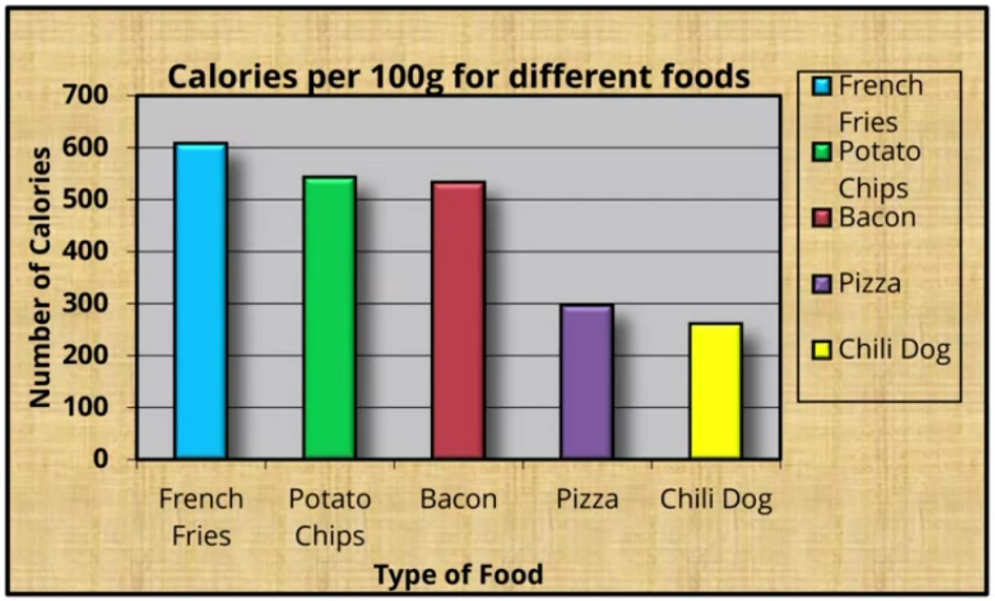
Orientation	Shape	Line Length	Line Thickness
Size	Curvature	Added Marks	Enclosure
Hue & Contrast	Intensity	Spatial Position	Motion

* Nusbaumer Knaflic, Cole, Story Telling With Data, pg. 105 (John Wiley & Sons, 2015)

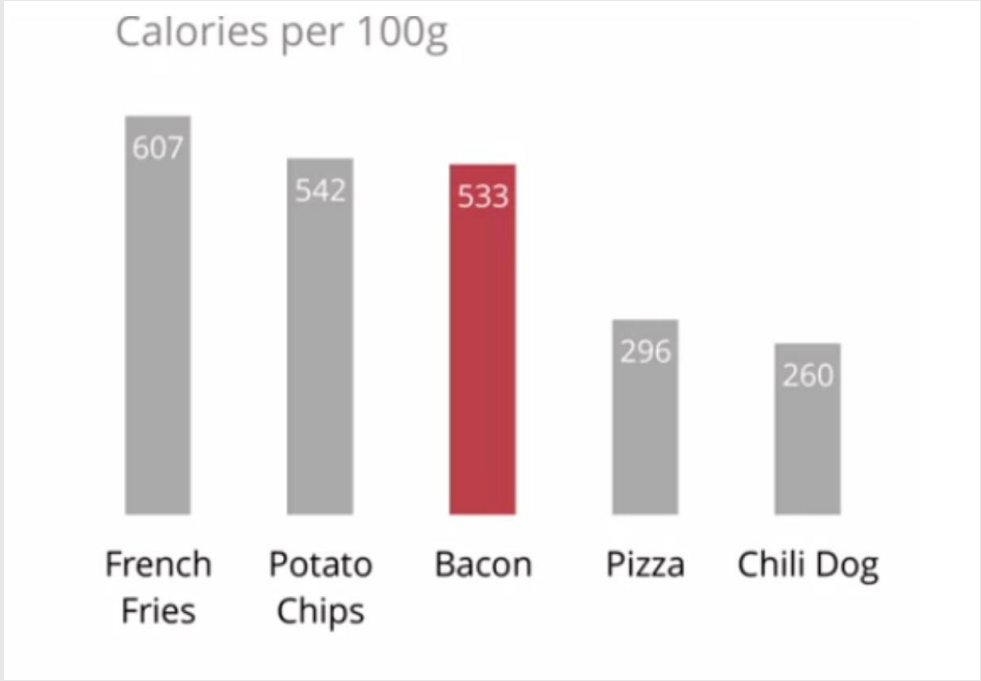
TECHNIQUE 1 – "KISS (KEEP IT SIMPLE STUPID)"

A significant contribution of the works of Tufte* and others is their emphasis on removing "chart junk" and maximizing the amount of ink devoted to the story told by the data.

This chart abounds in chart junk including the frame, the background colors, the grid and scale, the bar colors and 3D effect, and the key.



This "clean" chart is clear and tells the story!



Example charts from: https://www.codeconquest.com/blog/chart-junk-how-to-avoid-it/

*Tufte, Edward R. – "The Visual Display of Quantitative Information" 2nd ed. (Graphics Press 2007)

TECHNIQUE 2 – "INCLUDE A VISUAL KICKER"

A visual kicker is a short, punchy phrase or sentence that delivers the main message of the graphic. Put it at the top and use boldface and or color to leverage preconscious vision.

This data presentation is confusing and lacks a message. The key is at the bottom of the graphic in small type and the colors lack contrast.

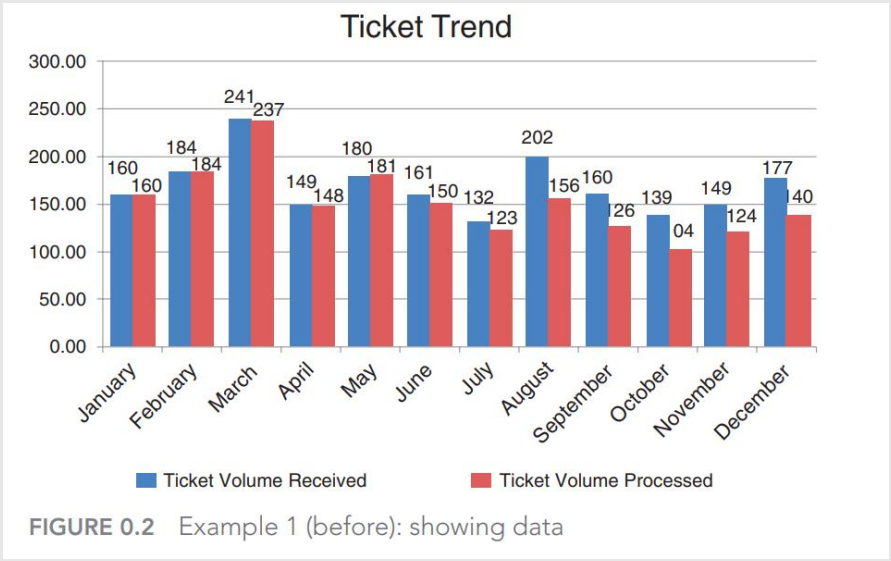


FIGURE 0.2 Example 1 (before): showing data

The visual kicker (Please approve…) is at the top of the graphic where we look first and is backed up by text telling the story. The colors of the data curves have contrast and are well labeled.

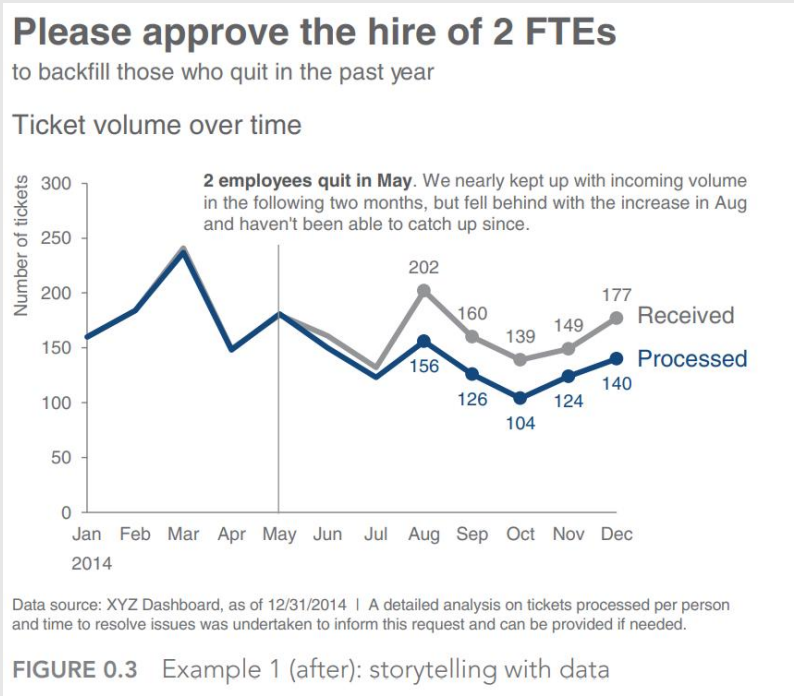


FIGURE 0.3 Example 1 (after): storytelling with data

* Nusbaumer Knaflic, Cole, Story Telling With Data, pg. 4 (John Wiley & Sons, 2015)

HOWEVER:
WHAT IS THE CONTEXT OF THIS DATA VISUALIZATION?

Context 1 - We need justification for hiring two individuals and increasing our FTE costs.

- If management has requested (implicitly or explicitly) this justification the visualization is fit for purpose.
- The underlying data on issue closure time is referenced in the footnote, but not displayed

Context 2 - The added delay in closing issues is an acceptable service level considering the achieved 2 FTE cost saving.

- If management wants to defend increased issue closure delay based on the 2 FTE savings this is the wrong visualization.
- The underlying data contains the issue closure times
- A visualization of average closure time by month would be required

TECHNIQUE 3 – "GO WITH THE FLOW"

When your graphic shows variation of a value (e.g., cost) with time or another variable take advantage of motion, a preconscious attribute, to "go with the flow".

It is hard to get the message of this image. There is no kicker and preconscious vision draws our attention to the colors of the bars

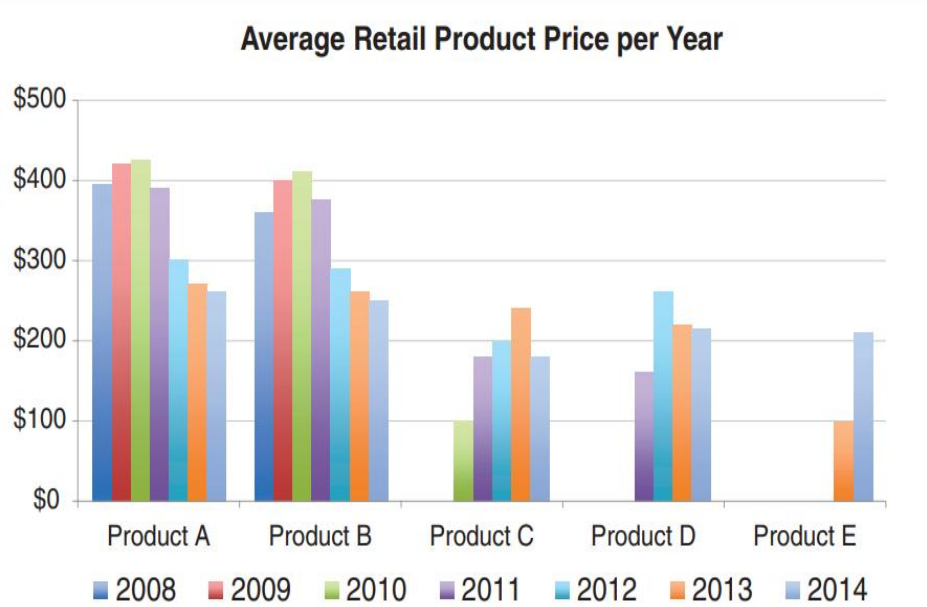


FIGURE 0.6 Example 3 (before): showing data

The message is clearly stated in the kicker and the flow immediately identifies time varying data.

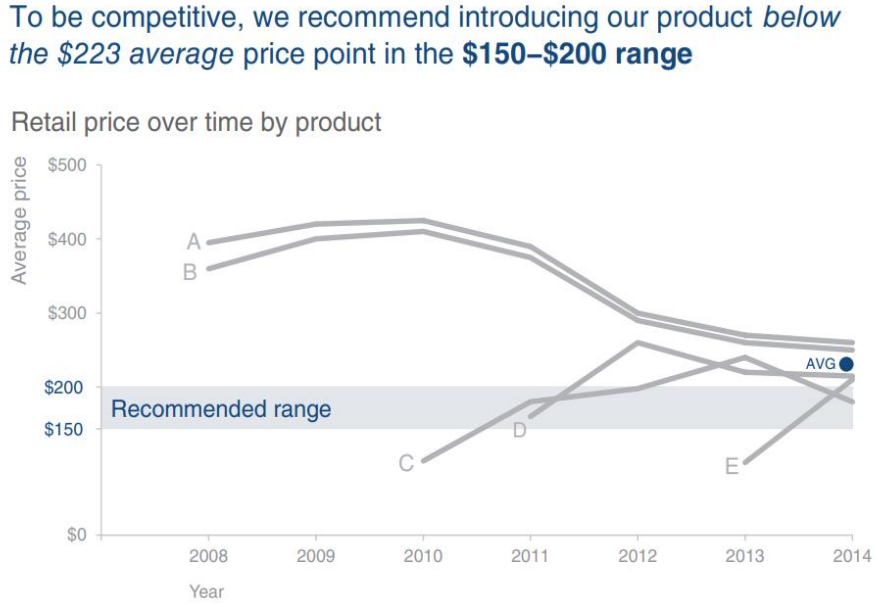
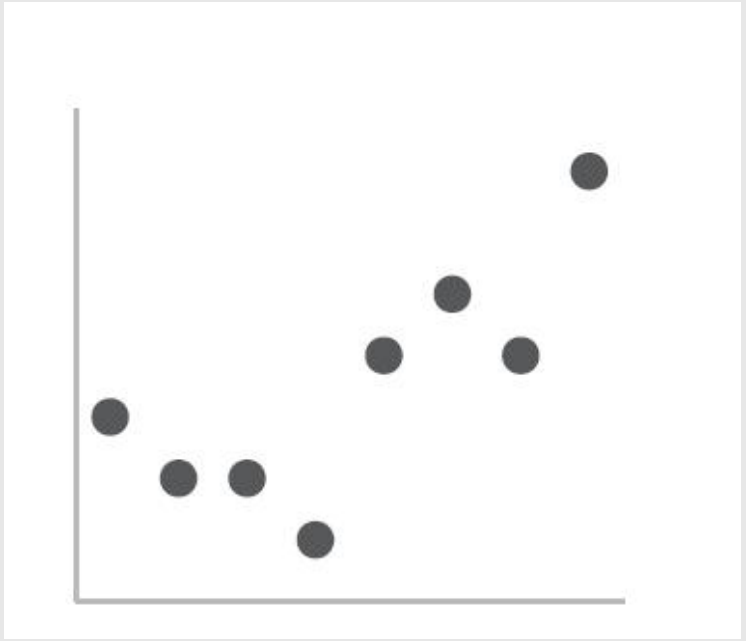


FIGURE 0.7 Example 3 (after): storytelling with data

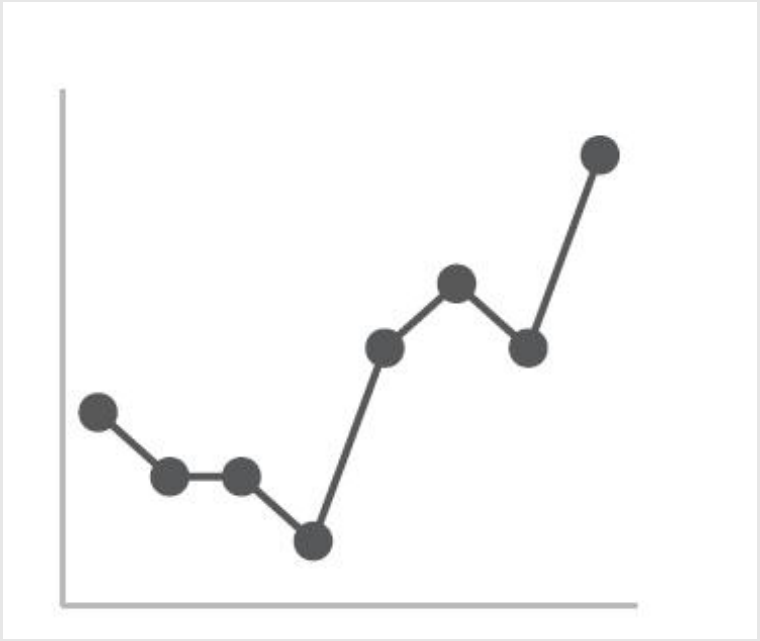
TECHNIQUE 4 – "SHOW THE WHOLE"

Gestalt* - a structure, arrangement, or pattern of physical, biological, or psychological phenomena so integrated as to constitute a functional unit with properties not derivable by summation of its parts.

Here are the parts. The relationship is unclear and there is no story being told.



Here is the whole, the beginning of a story to be told can be seen in this data series. CAUTION: Interpolation between data points is only valid for continuous functions.



* https://www.merriam-webster.com/dictionary/gestalt#

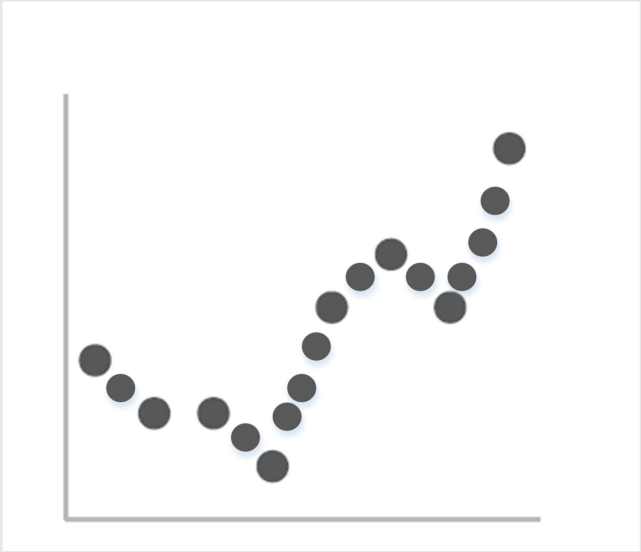
For a summary of Gestalt Principles of Visual Perception see Knaflic (Op. cit.) p74

TECHNIQUE 5 – "CONNECT THE DOTS"

This example demonstrate the gestalt principle of proximity. The primal need to recognize groups as potentially hostile drives a preconscious recognition of grouping.

You see columns on the left and rows on the right because of the vertical proximity of the dots on the left and horizontal proximity on the right.

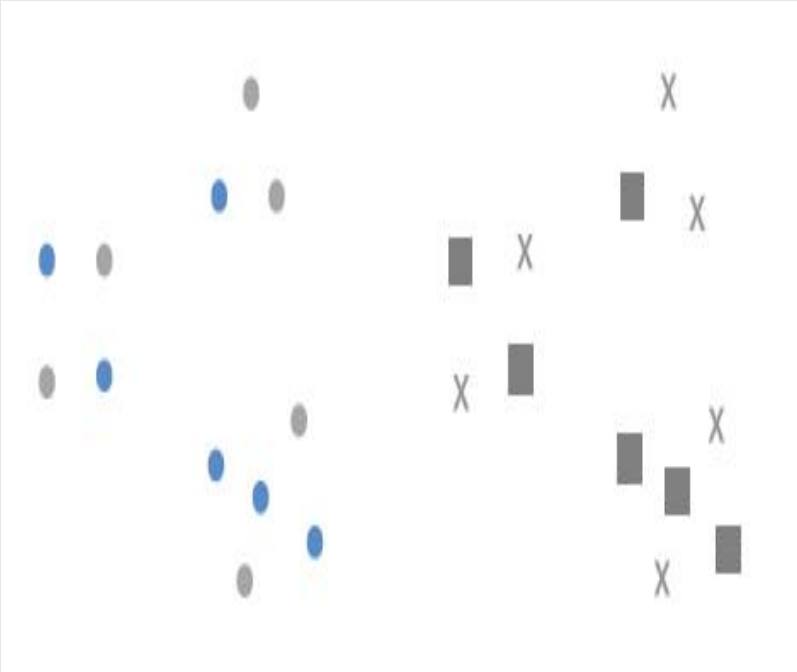
With added dots the series is evident. CAUTION; never be tempted to add points to a graphic that do not reflect what is in the underlying data.



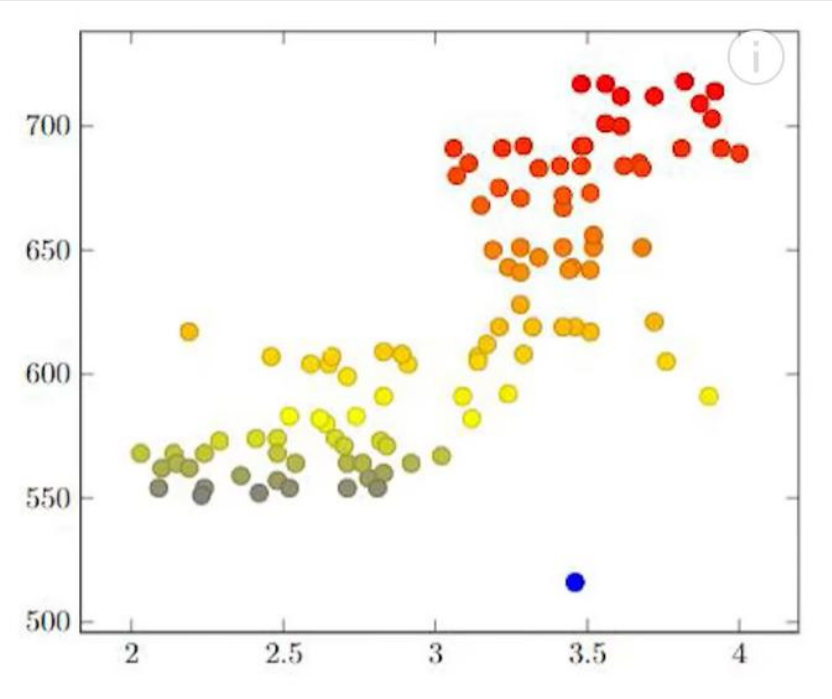
TECHNIQUE 6 – IDENTIFY MEMBERS OF A GROUP

The gestalt principle of similarity, that things that look alike are part of a group, evidences our preconscious process to recognize a group and assess it as hostile or friendly.

You naturally see the blue ovals on the left and the gray squares on the right as signifying members of groups having some manner of relationship.



Here we see the use of color to differentiate the red, orange, yellow and gray groups, and the lonely blue outlier

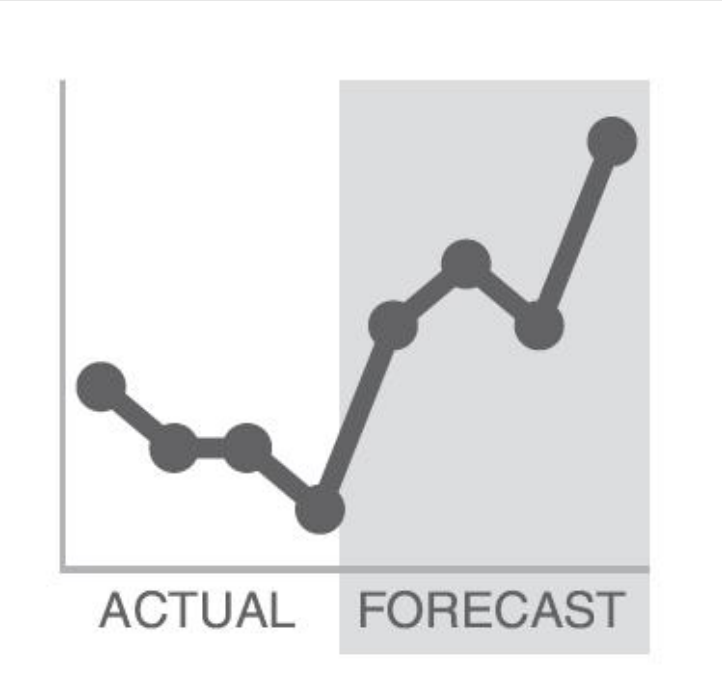


Example on the left if from Naflic (Op. Cit.) p76;
The illustration on the right is from https://www.youtube.com/watch?v=5jmIHOWpEg0

TECHNIQUE 7 – CREATE GOOD FENCES

Good Fences Make Good Neighbors applies to the Gestalt principle of enclosure and evidences our preconscious fear of hostile encroachment. If you are in you're a friend, if not watch out.

In the illustration below differentiation of the forecast from actual results is achieved by the gray box enclosing forecast data points.



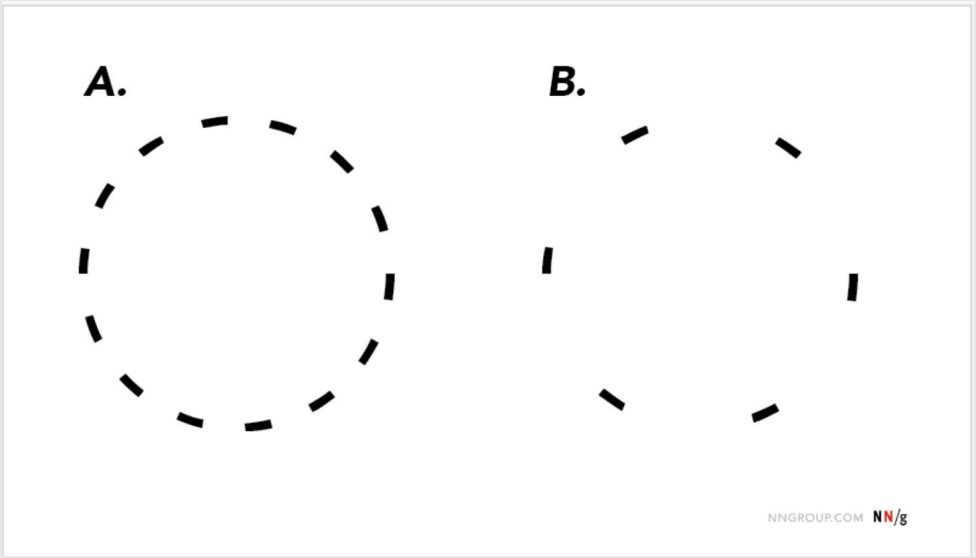
The familiar Gartner Magic Quadrant establishes group membership by placement in the four quadrants of the graphic.

TECHNIQUE 8 – LEAD THE VIEWER

The Gestalt principle of closure, evidencing our preconscious need to identify threats based on partial information, allows an audience to understand a message based on a sufficient, but not all, data.

Is that a Dalmatian or a bunch of spots?

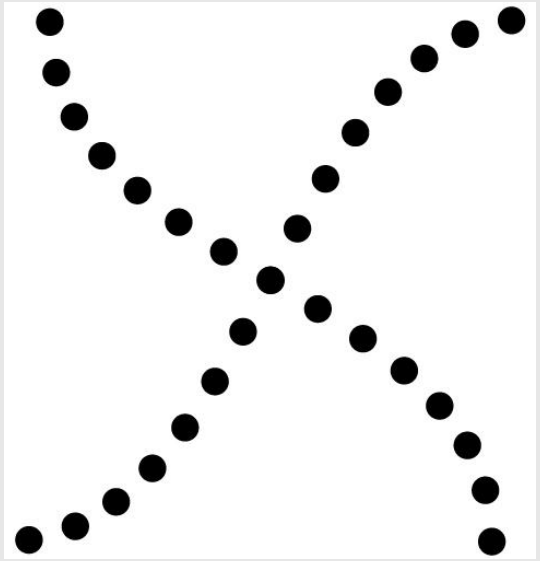
How much information is enough? The illustration on the left is a circle to most people; the one on the right is probably a circle, but maybe not.



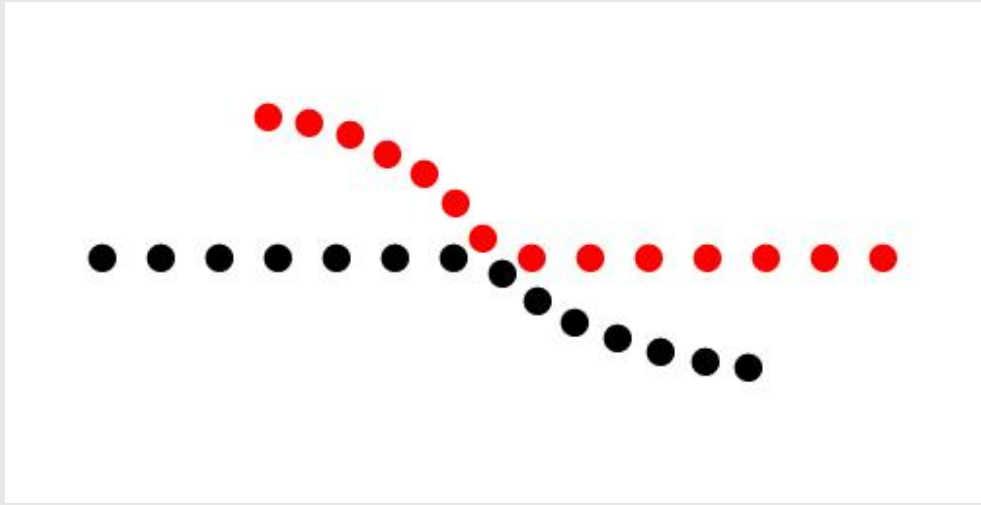
TECHNIQUE 9 – GET IN LINE

The Gestalt principle of continuity, evidencing our preconscious vision derived from the need to lead a game animal to hunt effectively, leads a viewer to follow and lead graphic elements on a path.

When viewing these two intersecting paths we are comfortable in assuming continuity of motion along each path across the intersection.



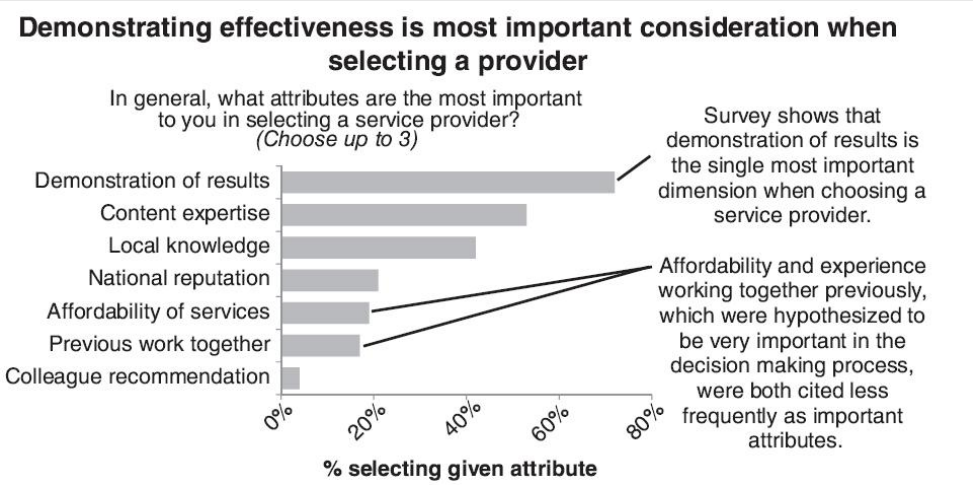
This image introduces dissonance since our unconscious assumption of continuity doesn't expect the sudden change in direction of the two paths at the intersection



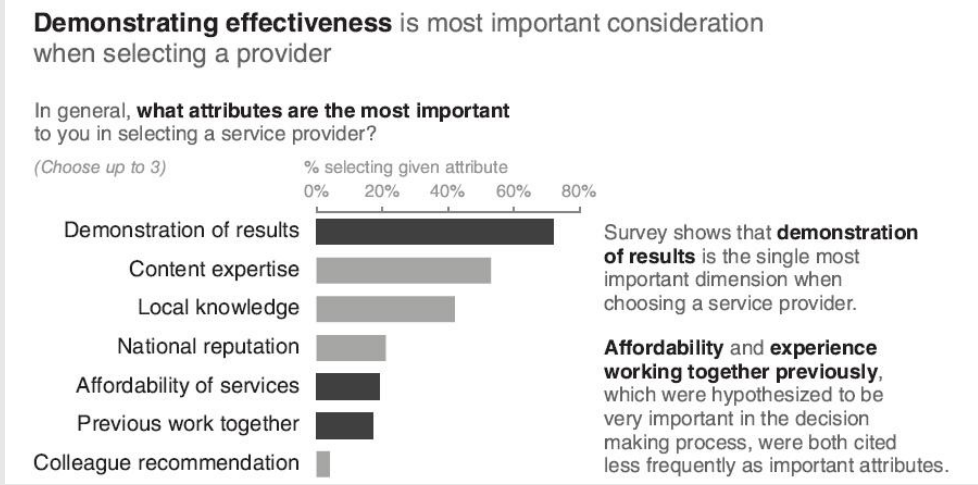
TECHNIQUE 10 – BE BOLD AND/OR COLORFUL, SELECTIVELY

Recent Research has incorporated saliency into the domain of gestalt principles. Visual saliency, used selectively, is effective in drawing the viewers attention to the most relevant information.

This chart lacks saliency – everything is gray, and the viewer must read all the text to get the message. The scale is also placed awkwardly making relative comparisons difficult



Selective use of black bars and bold-faced type introduces saliency, drawing the viewer's attention to what is important. The re-positioned and redrawn scale facilitates comparisons



Bottom-up processing is driven by the strongest force on earth, fear of the unknown. It is independent of any personal goals and in any new environment, the most salient objects will always draw our attention first.

https://abcofcolors.xyz/page/saliency-in uxdesign#:~:text=Saliency%20shows%20how%20different%20an,attention%20is%20distributed%20in%20space

QUALITY DATA VISUALIZATIONS MUST BE UNIVERSALLY ACCESSIBLE

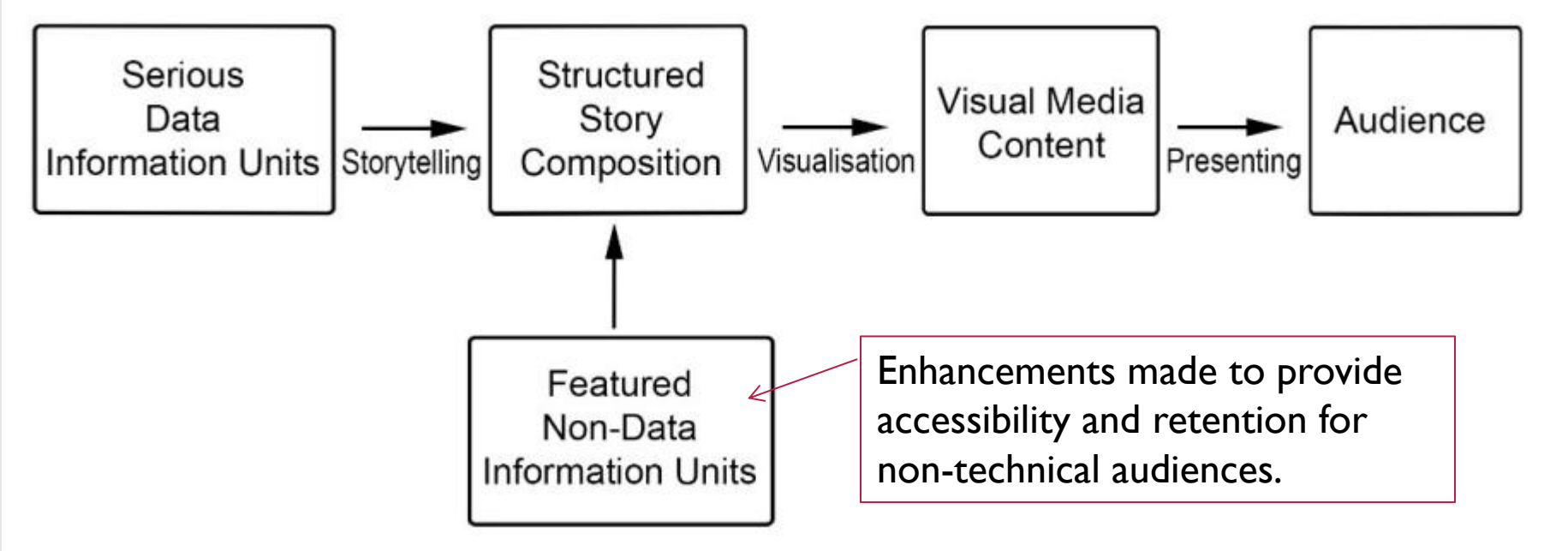
- In choosing colors, line weights, font sizes and visualization layout, consideration must be given to the needs of vision impaired and colorblind individuals.
  - Compliance is mandated by the American With Disabilities Act (ADA)
- Key references describing standards and techniques of accessibility design.
  - Okabe, Masataka and Kei Ito: *Color Universal Design (CUD) - How to make figures and presentations that are friendly to Colorblind people*; https://jfly.uni-koeln.de/color/
  - W3C Web Accessibility Initiative (WAI)
  - Wilke, Claus O.: *Fundamentals of Data Visualization (Chap 4 and Chap 19)*

A CONTRARIAN STORY TELLING APPROACH

While Tufte and others emphasize simplicity and the removal of clutter, an alternative contrarian school typified by Zhang et al.[1] emphasizes using pictorial enhancements as mnemonic devices to enhance audience receptivity and content retention.

1 - Zhang, Y.; Reynolds, M.; Lugmayr, A.; Damjanov, K.; Hassan, G.M. A Visual Data Storytelling Framework. Informatics 2022, 9, 73. https://doi.org/10.3390/informatics9040073

ZHANG ET AL. - VISUAL ENHANCEMENT PROCEDURE



Information Unit Definition

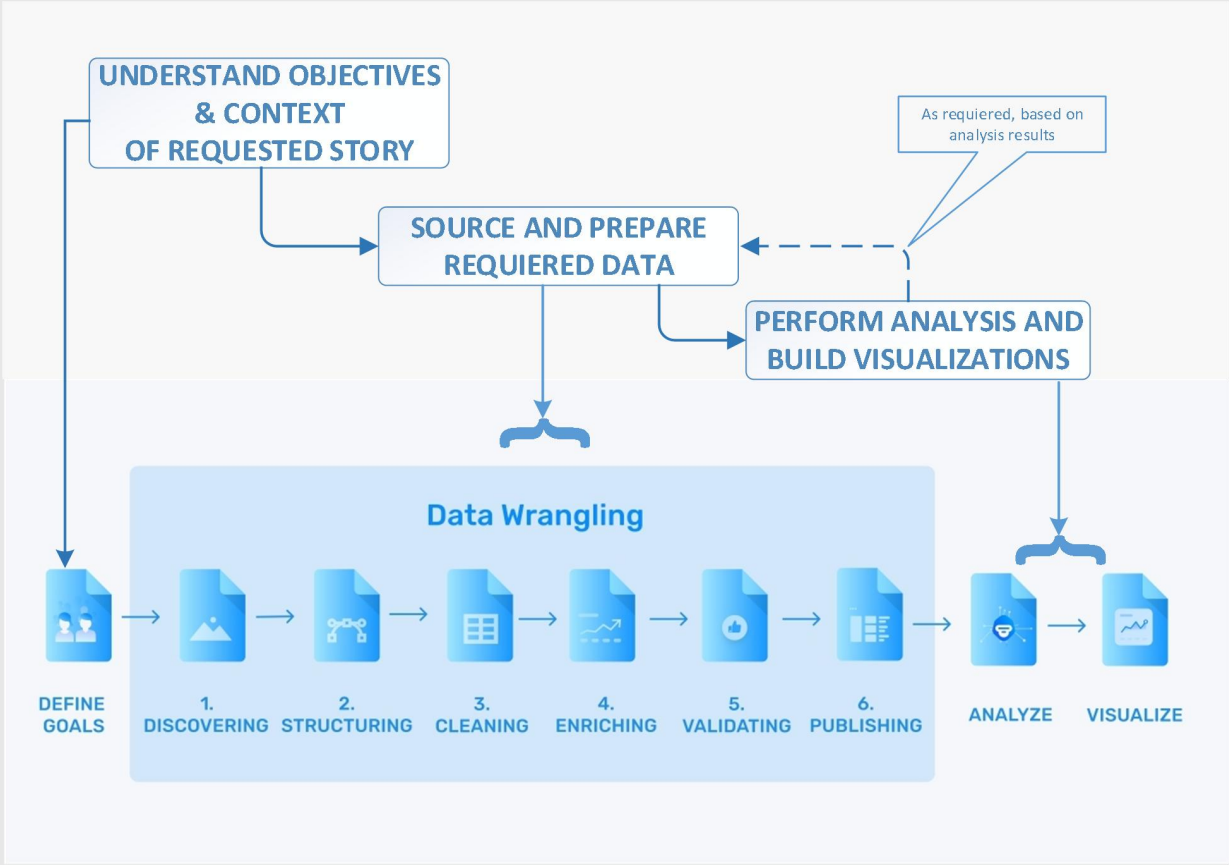
- Very basic meaning/information that is communicated in a human visual communication context.
- A single unit/element that is processed and encoded throughout the visual data storytelling process.
- A loose equivalent of an information bit (which is commonly referred to in the telecommunication relevant domain) in a human visual communication context.

THE PROCESS

These are the steps required to produce quality data visualizations.

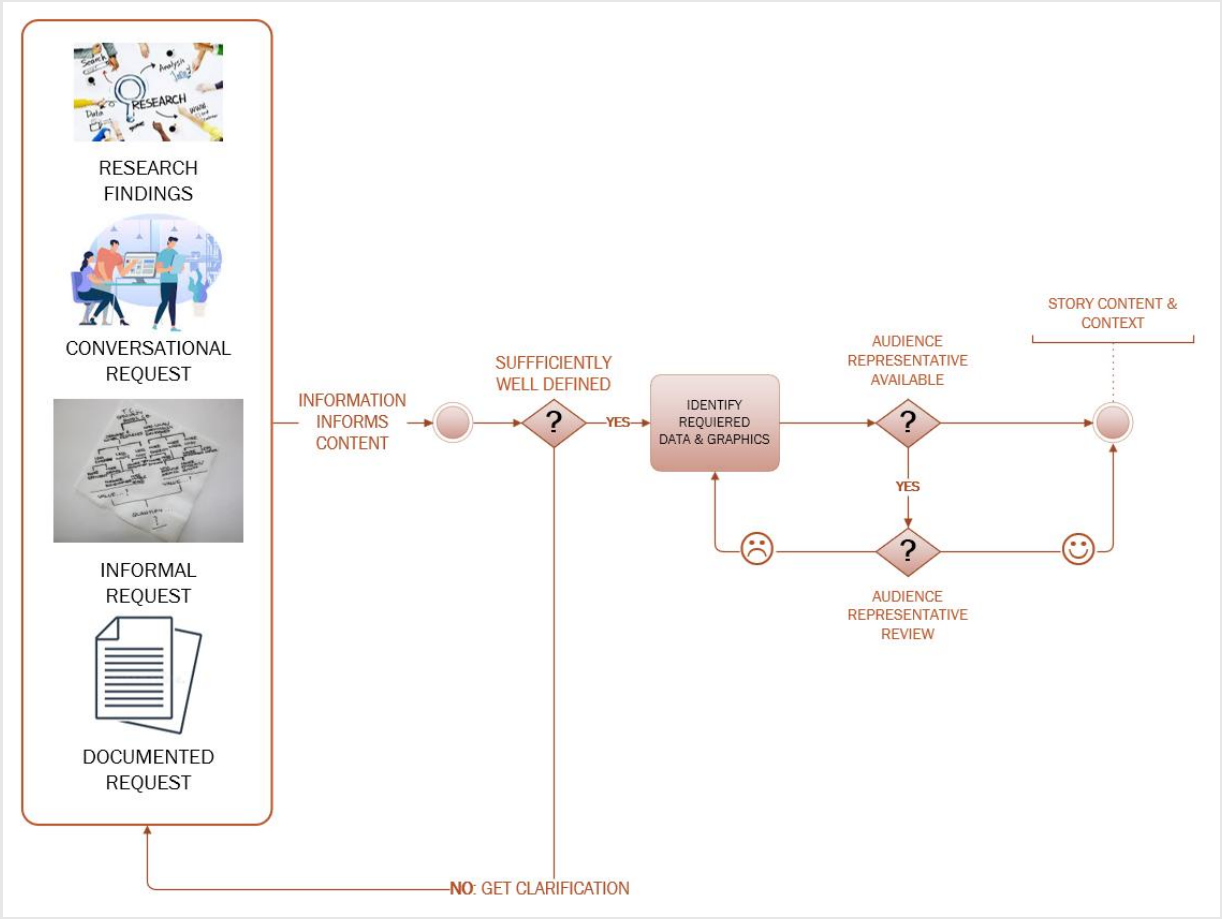
STORY DEVELOPMENT PROCESS

- The overall process begins with an understanding of goals, the content and context of the story, followed by Data Wrangling, to acquire, analyze and prepare the required data and analysis to develop results that will be the basis of the Data Visualizations that tell the story.



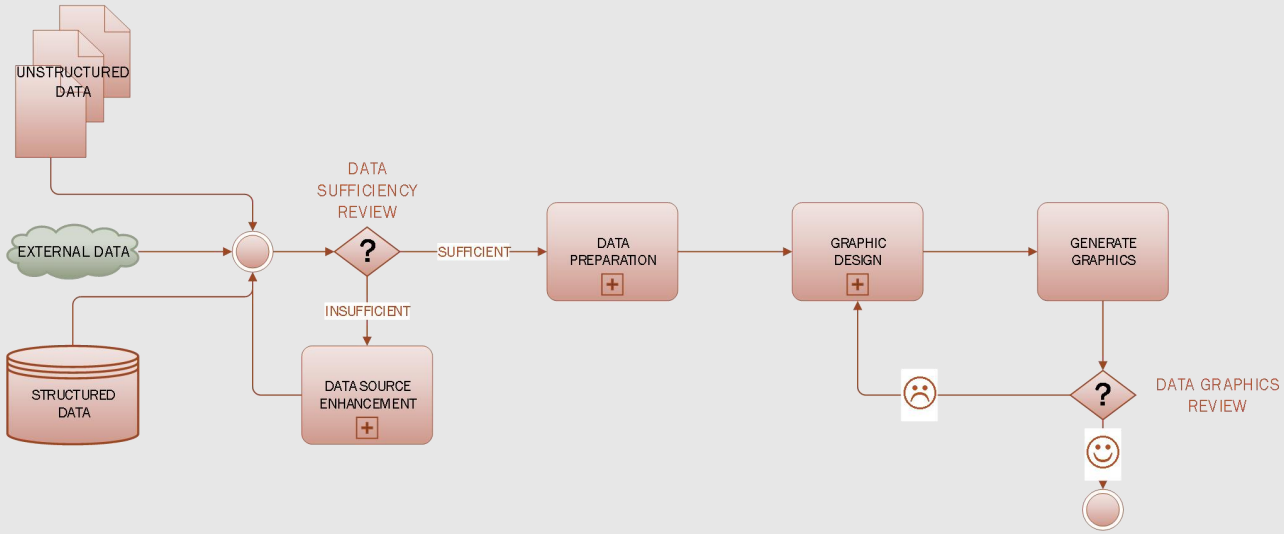
KNOW YOUR AUDIENCE AND THE STORY REQUEST CONTEXT

- Every story has an intended audience, even a scientific paper generated by a lone researcher reporting their research findings.
- Knowing your audience (KYA) and the context of their story request (whether explicit or implicit) is an essential prerequisite to creating a the story
  - Audience demographics
  - Most important audience members
  - Most important member interests and agendas
  - Time constraints



DAT SOURCES AND VISUALIZATIONS GO HAND-IN-HAND

- Given the content and context of a requested story you must find and source the required data from many types of data sources.
  - Internal company data
  - External publicly available data
  - Experimentally generated data
- Having sourced the required data, generate mock-ups of the visualizations you will need for initial review.
- Having passed the initial review undertake the necessary data wrangling required for generating your perfected data visualizations.



PRESENTATION QUALITY FEATURE REVIEW

- Developing high quality Data Visualizations that tell the right story is a challenging task.
- Undertake Quality Feature reviews as an ongoing part of your development process.
- Engage your colleagues (classmates and others) as objective reviewers.
- When possible, engage members of your intended audience as reviewers throughout the development cycle.