Data Visualization

Data is everywhere

There are many varied sources of data. This may cause data and information overload. There is a need to cut through the noise and present key information in an effective way. Thus, for managers, a critical problem is how to utilize this data effectively in order to facilitate decision-making.

Data visualization

Data visualization is the practice of graphically representing data to help people see and understand patterns, insights, and other discoveries hidden inside data. In a business context, data needs to presented in a way to facilitate interpretation, analysis, learning, and action.

Common tools for data visualization

- Microsoft Excel
- Microsoft PowerBI
- Tableau
- R / Python

Why do managers need data visualization?

- It can add value at various points in a project – exploration as well as reporting
- A powerful tool for any manager working with data scientists
- Animate trends in the data
- Helps tell an engaging story
- Helps convince decision makers to act
- Helps address issues inherent in traditional business intelligence tools
    - These may be difficult to learn and use
    - Expertise required for preparing reports
    - Long turnaround time for report generation

Three key uses of data visualization

- Exploration of data
- Developing a deeper understanding of trends and phenomena encoded in the data
- Communication of findings and visualization of arguments

Data exploration

- Plotting data before analysis can help decide the appropriate analysis technique

- Prevent inclusion of outliers in summary statistics, that can cause problems

Deeper understanding of trends

- A typical dataset has many features
- Visualizing combinations of features (based on a certain hypothesis) can either provide supporting evidence or nudge the manager towards alternative explanations

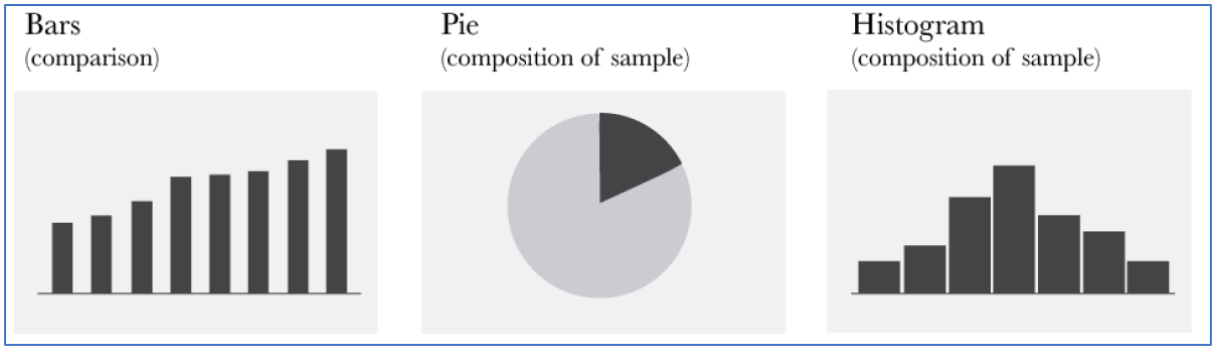
Communication of findings

- Effective visualizations take the brain's visual processing capabilities into account
- Present the key arguments while avoiding cognitive clutter
- Visualizations can help convince decision makers by helping narrate the key story

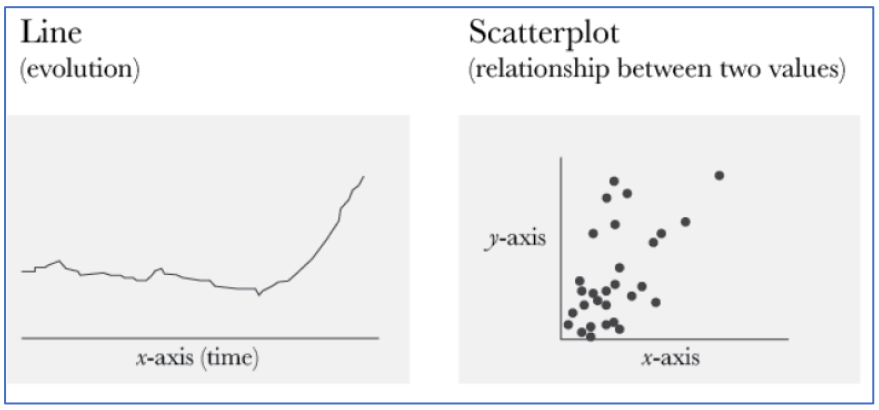
Choosing the right visuals

There are various visuals available, and choosing the right visual is important for communicating the right analysis based on the available data. Some examples are histogram, bar chart, pie chart, scatter plot, coxcomb chart, box and whisker plot, tree map, geographical map, heat map, cross-tabulation, timeline, Gantt chart, among others.

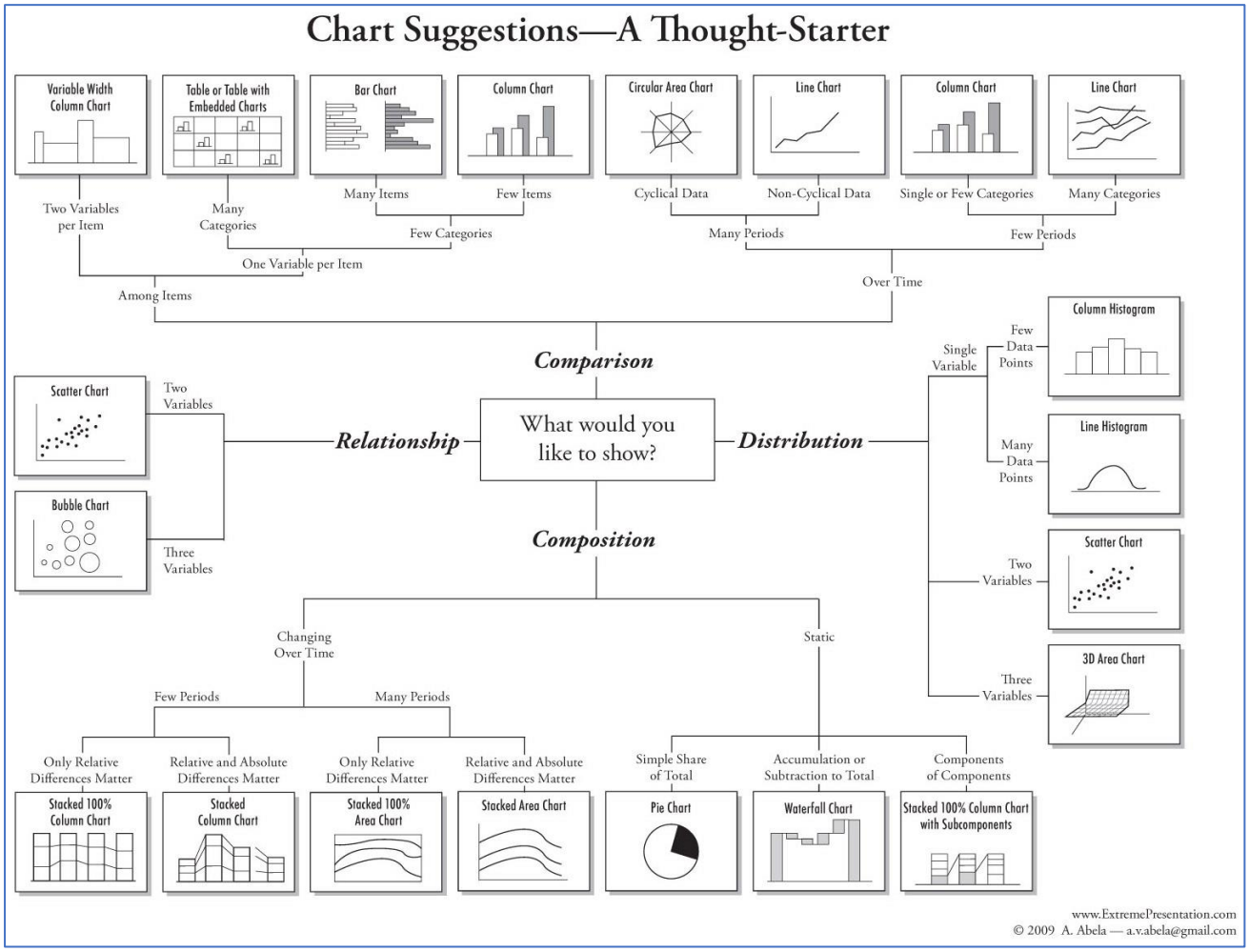
The two basic types of charts are – comparison and reduction. Comparison charts help in the quick comparison between categories, without any aggregation. Examples are bars, pie, and histogram.



Reduction charts help illustrate patterns based on aggregation or generalization of data. Examples are line and scatterplot.



Abela's chart type hierarchy is a useful tool to select the right chart or visual.



Human brain and data visualization

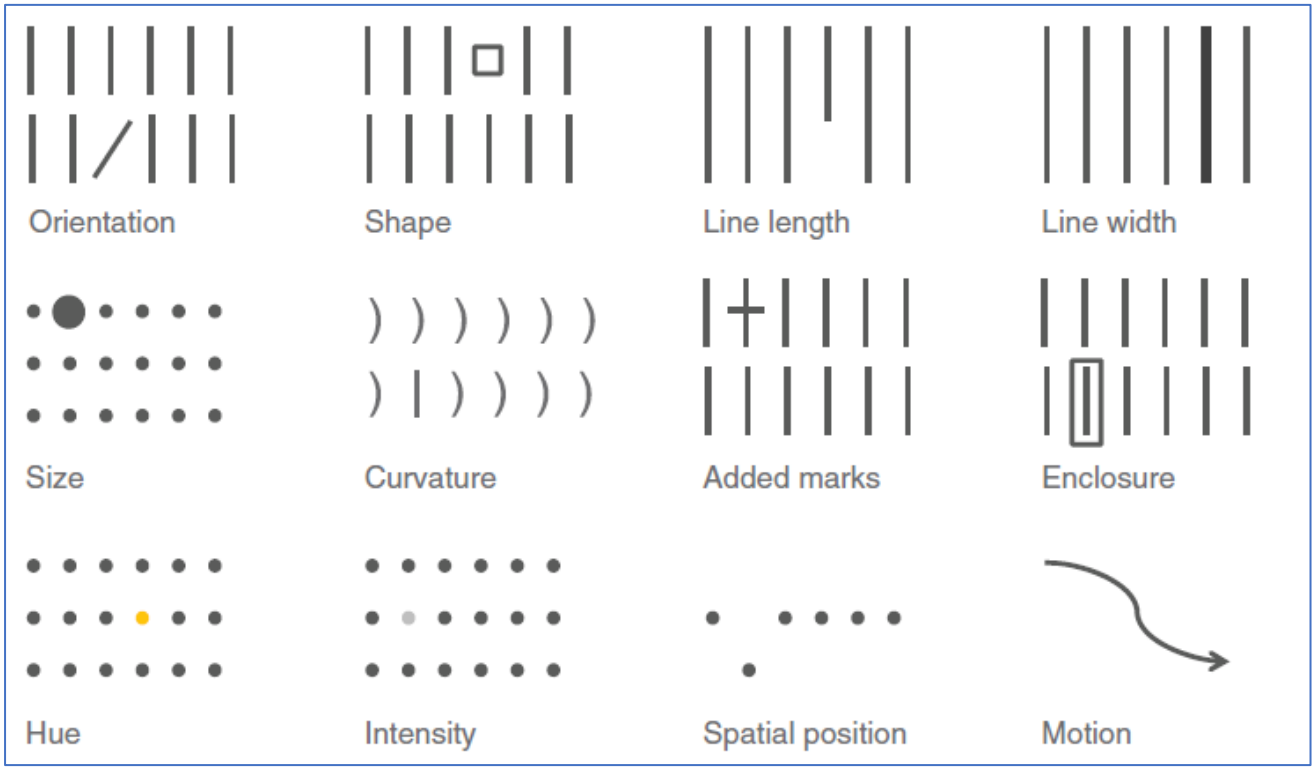
The human brain has its own strengths and limitations, but around 30-60% of its use is dedicated for visual processing. It helps detect, match, and make sense of patterns. Visualizations can be made more useful, engaging, and effective if we keep human brain attributes in mind while designing our visuals. We can achieve this using visual encoding.

Visual encoding

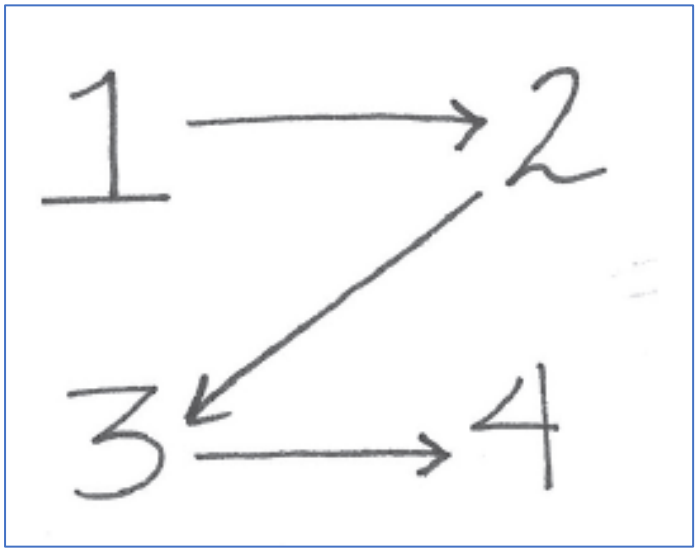
This refers to the assignment of various visuals attributes such as color and shape to different types of data. It translates data into visual code which can be decoded easily by humans. As a visual designer, we have to make the correct choice between various encoding options. The objective is to help clarify the data for the users.

Various pre-attentive attributes can be used in the right combination for effective encoding of visuals. Examples are size, color, and position on the page. These attributes can be leveraged to help direct audience attention to where we want them to focus. They can be used to create a

visual hierarchy of elements so that we lead the audience through the information we want to communicate in the way we want them to process it.



Position of elements on the page is also very important because, without other visual cues, most viewers will start at top left of visual and scan in a zig-zag manner across the page.



Cognitive load and clutter

Cognitive load refers to the amount of mental effort required to interpret information. It can be of three types – intrinsic, extraneous, and germane.

Intrinsic load refers to the amount of memory needed to understand and idea. Different tasks need different amounts of thought and attention.

Extraneous load refers to how information is presented. Extra brain power has to be expended to deal with bad data visualizations.

Germane load is about how we mentally organize into patterns and contextualize information for referring later.

Clutter

These are the visual elements which take up space but do not increase our understanding. These are the all the things we can remove from the data visualization while still preserving the key ideas. Reduction in clutter will reduce the cognitive load on the user.

Some types of visual clutter are:

- Irregular visual order
- Improper alignment
- Unnecessary white space
- Non-strategic use of contrast

Gestalt principles

These design principles help us evaluate how our brain perceives the world. They can help identify signal and noise in our visual.

The six gestalt principles are:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection