6. Share Data Through the Art of Visualization

Effective Data Visualizations :

1. The mcCandless method:
    1. Information : the data you are working with.
    2. Story : a clear and compelling narrative or concept.
    3. Goal : specific objective or function for the visual.
    4. Visual form : an effective use of metaphor or visual expression.

Visualization Notes :

- The x- axis is used to represent categories, time periods or other variables.
- The y -axis usually has a scale of values for variables.

rules about what makes a helpful data visualization:

- **Five-second rule:** A data visualization should be clear, effective, and convincing enough to be absorbed in five seconds or less.
- **Color contrast:** Graphs and charts should use a diverging color palette to show contrast between elements.
- **Conventions and expectations:** Visuals and their organization should align with audience expectations and cultural conventions. For example, if the majority of your audience associates green with a positive concept and red with a negative one, your visualization should reflect this.
- **Minimal labels:** Titles, axes, and annotations should use as few labels as it takes to make sense. Having too many labels makes your graph or chart too busy. It takes up too much space and prevents the labels from being shown clearly.

- **Correlation** in statistics is the measure of the degree to which two variables move in relationship to each other. An example of correlation is the idea that “As the temperature goes up, ice cream sales also go up.” It is important to remember that correlation doesn’t mean that one event causes another. But, it does indicate that they have a pattern with or a relationship to each other. If one variable goes up and the other variable also goes up, it is a positive correlation. If one variable goes up and the other variable goes down, it is a negative or inverse correlation. If one

variable goes up and the other variable stays about the same, there is no correlation.

- **Causation** refers to the idea that an event leads to a specific outcome. For example, when lightning strikes, we hear the thunder (sound wave) caused by the air heating and cooling from the lightning strike. Lightning causes thunder.

If there is a correlation between two variables, a pattern will be seen when the variables are plotted on a scatter plot.

There are three ways to describe the correlation between variables.

- Positive correlation: As x increases, y increases.
- Negative correlation: As x increases, y decreases.
- No correlation: As x increases, y stays about the same or has no clear pattern.

Line chart - A line chart is used to track changes over short and long periods of time. When smaller changes exist, line charts are better to use than bar graphs. Line charts can also be used to compare changes over the same period of time for more than one group.

Column charts - use size to contrast and compare two or more values, using height or lengths to represent the specific values.

Heatmap - Similar to bar charts, **heatmaps** also use color to compare categories in a data set. They are mainly used to show relationships between two variables and use a system of color-coding to represent different values.

pie chart - is a circular graph that is divided into segments representing proportions corresponding to the quantity it represents, especially when dealing with parts of a whole.

Scatter plots - show relationships between different variables. Scatter plots are typically used for two variables for a set of data, although additional variables can be displayed.

distribution graph - displays the spread of various outcomes in a dataset.

Horizontal bar: a bar chart functions similarly to a column chart, but is flipped horizontally.

Area: area charts allow you to track changes in value across multiple categories of data.

Combo: combo charts use multiple visual markers like columns and lines to showcase different aspects of the data in one visualization. The example below is a combo chart that has a column and line chart together.

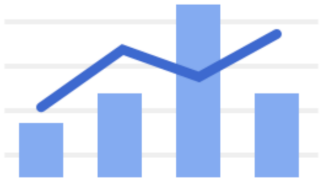


Chart selection :

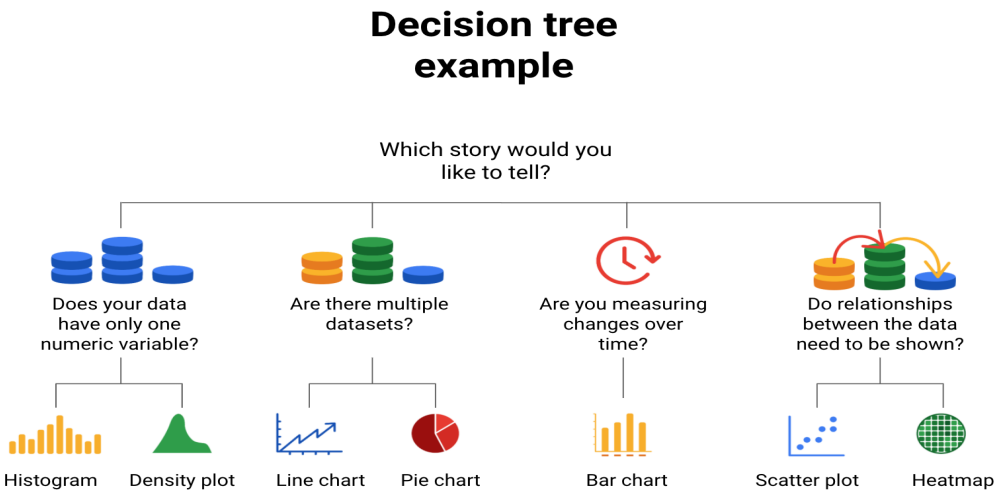
A **line chart** is ideal for highlighting trends over time.

A **bar chart** is ideal for comparing similar data side by side.

A **histogram** is ideal for comparing the distribution of two variables by individual grouping.

A **pie chart** is ideal for measuring data as a proportion of the whole.

A **scatter plot** is ideal for exploring potential relationships between two variables.



Initial hypothesis : -

- The theory you are trying to prove or disprove with data.

Design thinking for data visualization involves five phases:

1. **Empathize:** Thinking about the emotions and needs of the target audience for the data visualization
2. **Define:** Figuring out exactly what your audience needs from the data
3. **Ideate:** Generating ideas for data visualization
4. **Prototype:** Putting visualizations together for testing and feedback
5. **Test:** Showing prototype visualizations to people before stakeholders see them

Tableau :

- Dynamic visualization : visualization that is interactive or changes over the time.
- Tableau visualizations are dynamic.
- Tableau interprets the type of data in each column. The following icons, which are above in the column name, refer to how Tableau interprets the data type in the column:

#: Numeric data
Abc: String data
Globe: Geographic data

Calendar: Date data
Calendar with a clock: Date and time data

- In tableau if you want to create joins then after adding it to the area double click on it.
- Check the data types of the joining field, it should be compatible.

- **Data storytelling** means communicating the meaning of a dataset with visuals and a narrative that are customized for a particular audience.
- Storytelling parts
    1. Character :- people affected by your story.
    2. Setting :- what's happening and other info.
    3. Plot : creates tension in a situation.
    4. Big reveal : how data solves the problem.
    5. Aha moment : data driven recommendations.

- Windows use the snipping tool to take screenshots to include in presentations.
- Another option is to go to the dashboard option on top in tableau desktop app and select the copy image option. Then paste it in the presentation.

- **Data visualization** is the representation and presentation of data to help with understanding. You can use graphs, charts, word clouds, and other visual depictions to help your audience see and clearly understand your data.

Data Presentation Tips :

How to make Presentation :

- Include a title, subtitle, and date
- Use a logical sequence of slides
- Provide an agenda with a timeline
- Limit the amount of text on slides. Your audience should be able to scan each block of text on your slides within 5 seconds
- Start with the business task. Focus on the business task and frame the information in the context of the business task.

- Establish the initial hypothesis
- Show what business metrics you used
- Use visualizations
- Introduce the graphic by name
- Provide a title for each graph
- Go from the general to the specific
- Use speaker notes to help you remember talking points
- Include key takeaways