# Selecting the Right Data Visualization in a Data Science Pipeline for Data Exploration and Data Interpretation

## Introduction

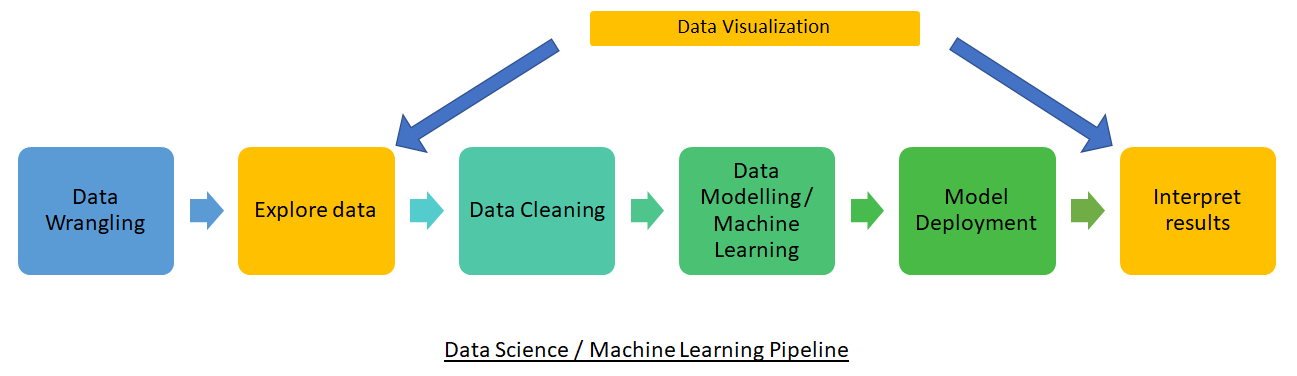
The result of many Data Science projects is to allow good business decision making, and an easily and accurately interpretable Data Visualization is key to that. Choosing the wrong visualization can lead to incorrect business decisions, impacting company bottom line.

As Data Scientists, it important we understand the science of human cognitive perception before choosing a visualization.

In this article I will explain the scientific basis and human psychology of why we find certain Data visualizations far easier to read, and why some are harder. I will also cover different use cases of data visualizations and share which visualization are better for which use cases. I will cover only concepts that are used in everyday practice.

## Data Science Pipeline and Data Visualization

Let us look at a typical Data Science project pipeline that consists of Data wrangling, Data exploration, Data Cleaning, Data modelling, Model Deployment and Result interpretation.



Data Visualization is used twice in the pipeline. Once to understand the source data, so that the correct model can be created. The second time to interpret the result, so that the correct business decisions can be made. If we chose the incorrect visualization in either of these steps, we could end up making an incorrect business decision, and the company could potentially lose a lot of money.

## Quantitative vs Categorical vs Ordinal

The human mind perceives different visualizations differently depending on if the data being plotted is Quantitative, Categorical or Ordinal. Quantitative data will always be a number that can be measured. Categorical or Nominal data is classified without a natural order or rank, whereas ordinal data has a predetermined or natural order.

## Human Cognitive Perception Accuracy Scale for Quantitative Data

Based on various experiments as stated in [1], then refined by [2], the human visual cognizance decodes visual information in the following order of accuracy

Table

Description automatically generated

Visual cues that are higher up in the above table are better, and allow for more accurate interpretation of data being represented.

We make certain observations

1. Position is the best cue to use for all types of variables: quantitative, ordinal and nominal
2. Length, Angle, Slope, Area and volume, each represent quantity, hence can be used for quantitative variables. They are at the bottom on the table for ordinal and nominal types of variables and should not be used since they are bad representations of category

Now we will look at different visualizations that can be used, and relate them to the above perception accuracy table. The below text and images are from [3]

## Charts

Let us look at the table below to see which charts should be used based on the type of variables.

Table

Description automatically generated

Or to look at the above table in images

Graphical user interface, application, PowerPoint

Description automatically generated

References

[1] [CLEVELAND, W. S., AND MCGILL, R.](https://www.jstor.org/stable/2288400) Graphical perception: Theory, experimentation and application to the development of graphical methods. Journal of the American Statistical Association, 79(387) 1984.

[2] [J. Mackinlay](https://research.tableau.com/sites/default/files/p110-mackinlay.pdf), Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5(2), 1986.

[3] MCS UIUC Class on Data Visualization by Prof. John C. Hart.

[4] Visualization Analysis and Design