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

## Introduction

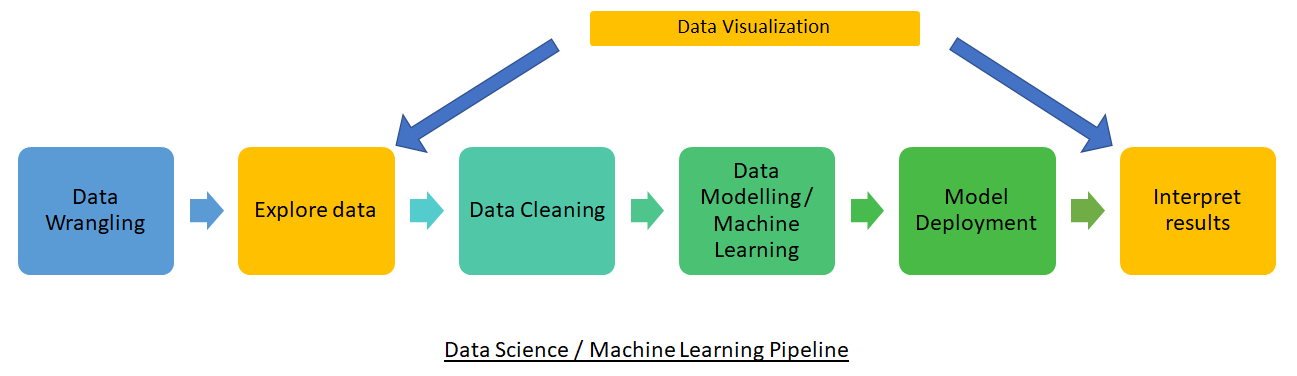
The result of many Data Science projects is to allow good business decision making, and an easily and accurately interpretable Data Visualization is key to that. Choosing the wrong visualization can lead to incorrect business decisions, impacting company bottom line.

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.



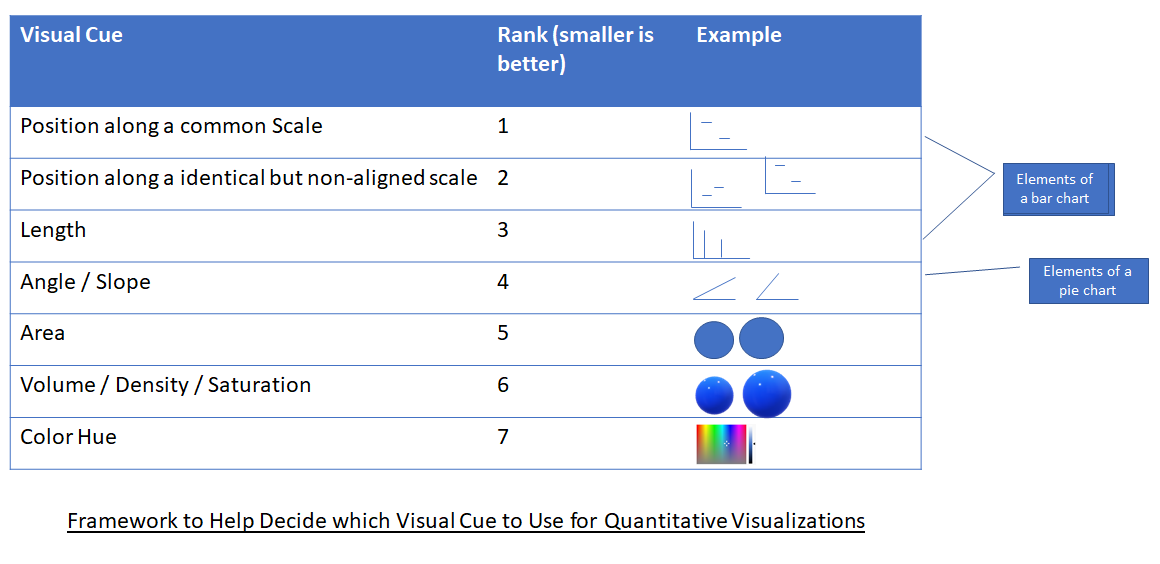
Data Visualization is used twice in the pipeline. Once to understand the source data, so that the correct model can be created. The second time to interpret the result, so that the correct business decisions can be made. If we chose the incorrect visualization in either of these steps, we could end up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

## Continuous vs Categorical vs Ordinal

The human mind perceives different visualizations accurately depending on if the data being plotted is Continuous, or Categorical or Ordinal. A diagram below is to refresh the reader’s memory on what they are.

## Human Cognitive Perception Accuracy Scale for Quantitative Data

Based on various experiments as stated in [1], the human visual cognizance decoded visual information in the following order of accuracy



According to [1], visual cues that are higher up in the above table are better to use to allow for more accurate interpretation of the Quantitative data being represented visually.

Different

ColorforLabeling(NominalCodes): nominal information coding. A nominal code does not have to be orderable; it simply must be remembered and recognized. Color can be extremely effective when we wish to make it easy for someone to classify visual symbols into separate categories giving the objects distinctive colors is often the best solution.

References

[1] Cleveland, William & Mcgill, Ron. (1985). Graphical Perception and Graphical Methods for Analyzing Scientific Data. Science

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

[3] Visualization Analysis and Design