Data Visualisation- Approach & Techniques (Part One)

Data Science Lecture Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly. **(FBV: Mutual Respect.)**

- No question is daft or silly - **ask them!**

- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. Moderators are going to be answering questions as the session progresses as well.

- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Open Classes. You can submit these questions here: **Open Class Questions**

CoGrammar

Data Science Lecture Housekeeping cont.

- For all **non-academic questions**, please submit a query:
  www.hyperiondev.com/support

- Report a **safeguarding** incident:
  www.hyperiondev.com/safeguardreporting

- We would love your **feedback** on lectures: **Feedback on Lectures**

Lecture Objectives

- **Learning and understanding data visualisation.**

- **Introduction to Python packages for Data Science.**

The Data Science Pipeline

★ **Part of Data Science, is to understand the data we are working with.**

★ **For larger chunks of data or larger datasets, we would like to visualise it.**

★ **Therefore the process of cleaning and preparing our data is crucial to Data Science.**

The Data Science Pipeline
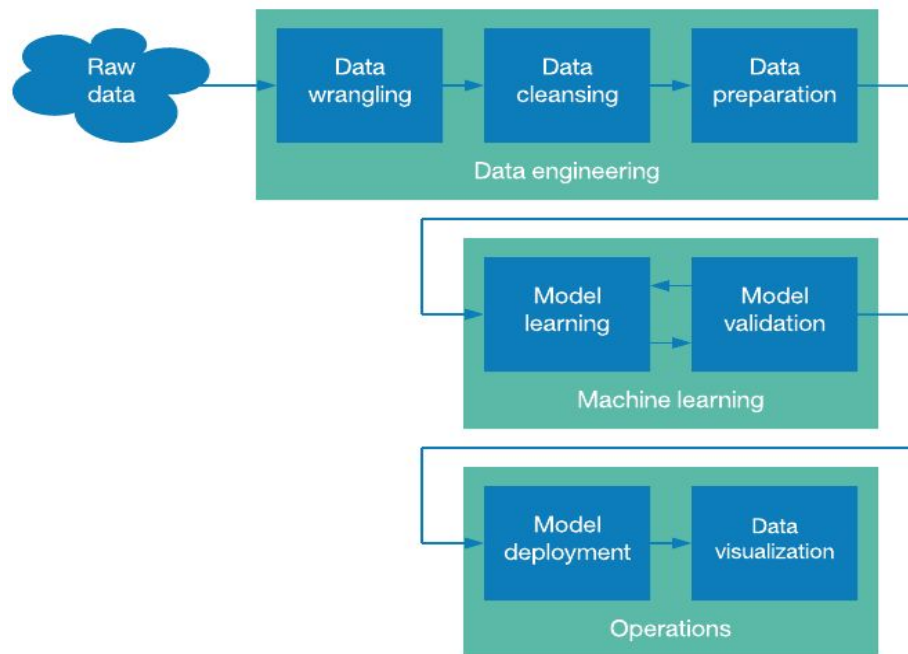
Approaching Data Visualisation

★ **Understand your dataset :**
  ○ **Useful since it would help in our analysis and also to potentially mitigate some inconsistencies.**

★ **Determine what you would like to find about our dataset :**
  ○ **Ask questions about your dataset, pull the data and answer the questions through visualisation.**

★ **Process and clean your dataset :**
  ○ **Missing / unstructured data will likely not yield anything useful.**

Approaching Data Visualisation

★ **Create data visualisations :**
  ○ **Time to answer the questions that we asked previously. Depending on what we asked, we'll need an appropriate visualisation.**

★ **Refine your visualisation :**
  ○ **Graphs should be easy to read on first glance.**

★ **Note down our findings:**
  ○ **At this stage, we can start analysing our graphs and finding conclusions.**

Types of Data

★ **Discrete Data :**
  ○ **Can take specific values, with an infinite range.**
  ○ **E.g : [1, 2.5, 3.1, 4, 5.9, ...]**

★ **Categorical Data :**
  ○ **Can take specific values, with a limited range.**
  ○ **E.g : [Dog, Cat, Hamster, Fish]**

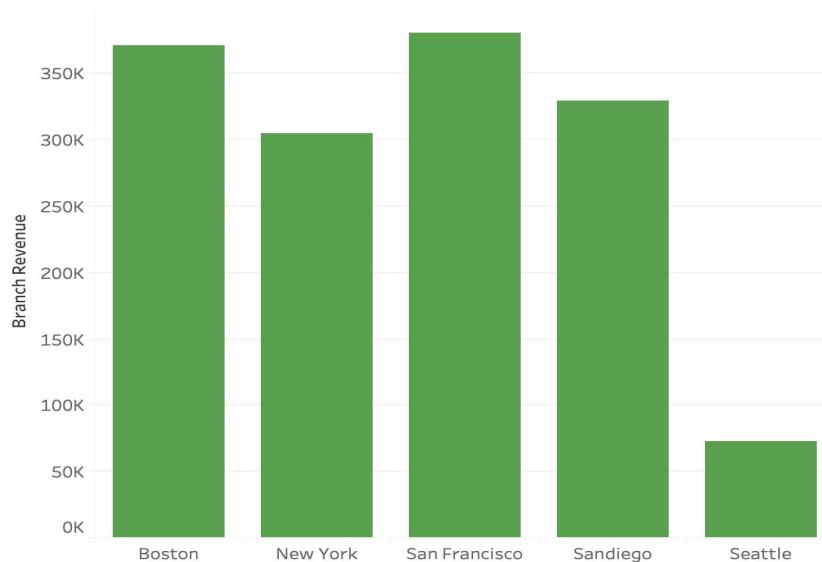★ **Continuous Data:**
  ○ **Available values in a spectrum. Meaning that there are an infinite number of values**
  ○ **E.g : Temperature or Distance.**

Types of Data

★ **Time Series Data :**
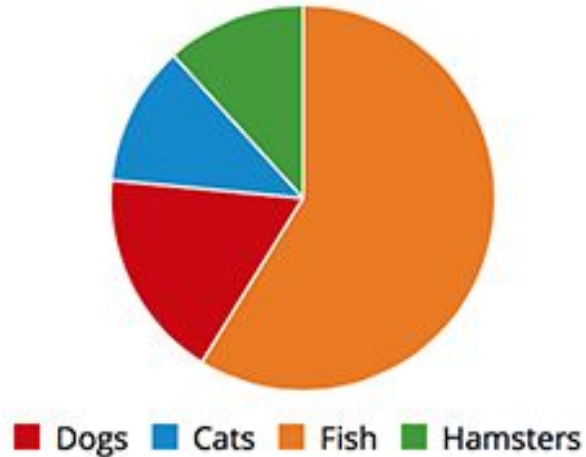 ○ **Data changes along with some form of time-related progression**

Pie Chart

**Good for plotting data that is : Categorical vs. Discrete
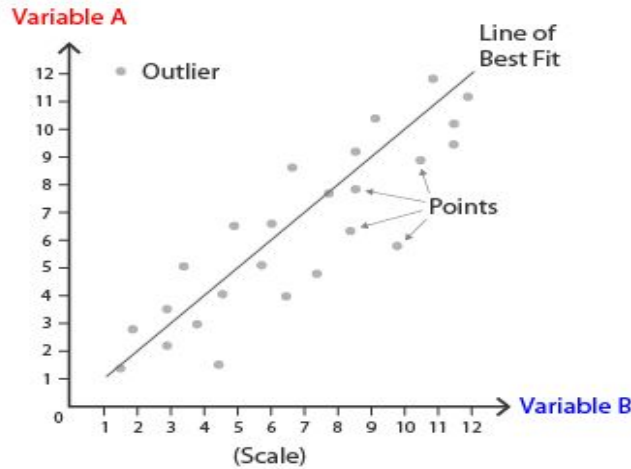Also great for getting a sense of proportions.**
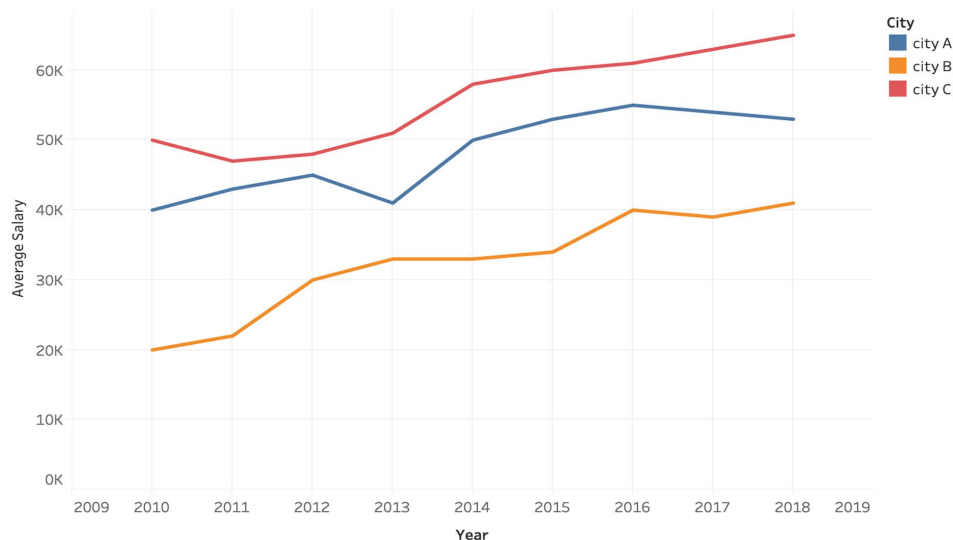

■ Dogs ■ Cats ■ Fish ■ Hamsters

Scatterplots

Great for plotting data that is : Discrete vs. Discrete or Continuous vs. Continuous. Can be useful for finding relationships between variables.

Line Graphs

Good for plotting that is : Discrete / Continuous vs. Discrete / Continuous or comparing with a Time Series.

Let's Breathe

Let's take a small break before moving on to the next topic.

More on Data Types

★ **Usually we tend to work with two main categories : Categorical & Continuous.**

★ **Categorical variables are also known as discrete or qualitative variables.**
   ○ **Examples consist of : Age Group, Gender or Race.**

   ○ **They can be even further divided into nominal, ordinal of dichotomous variables.**

Nominal Variables

Has two or more categories, which do not have a specific or predefined order.

For example, properties could be classified as houses, condos or bungalows. Therefore, the variable that holds the property type is a nominal variable.

For example, the state or province a person lives in would also be a nominal variable.

Ordinal Variables

They are nominal variables, but the categories can be ordered or ranked.

For example, if you were asked to rate your satisfaction with a course, your responses could be : "Completely satisfied", "Mostly satisfied" or "Very dissatisfied".

Dichotomous Variables

Are nominal variables which has only two categories or levels.

For example, we could use a dichotomous variable to describe whether an individual is a pensioner or not. In cases like this, the categories would be "True" or "False".

More on Data Types

★ **Continuous variables can take on infinitely uncountable numerical values, including integers and floating points.**

★ **They are also known as quantitative variables.**

★ **They can be further categorised as either interval or ratio variables.**

Interval Variables

Are "variables for which their central characteristic is that they can be measured along a continuum and they have a numerical value." (Laerd, n.d)

For example : Temperature measured in degrees Celsius or Fahrenheit.

Ratio Variables

Are interval variables, but a 0 value means there is none of that particular variable.

For example, "weight" is a ratio variable because if a variable measuring the sugar stock at a bakery was equal to 0 kgs it would mean there is no sugar in stock.

Temperature (measured in degrees Celsius) would not be a ratio variable because 0° C does not mean that there is no temperature.

CoGrammar

Q & A SECTION

Please use this time to ask any questions relating to the topic, should you have any.