DATA VISUALISATION
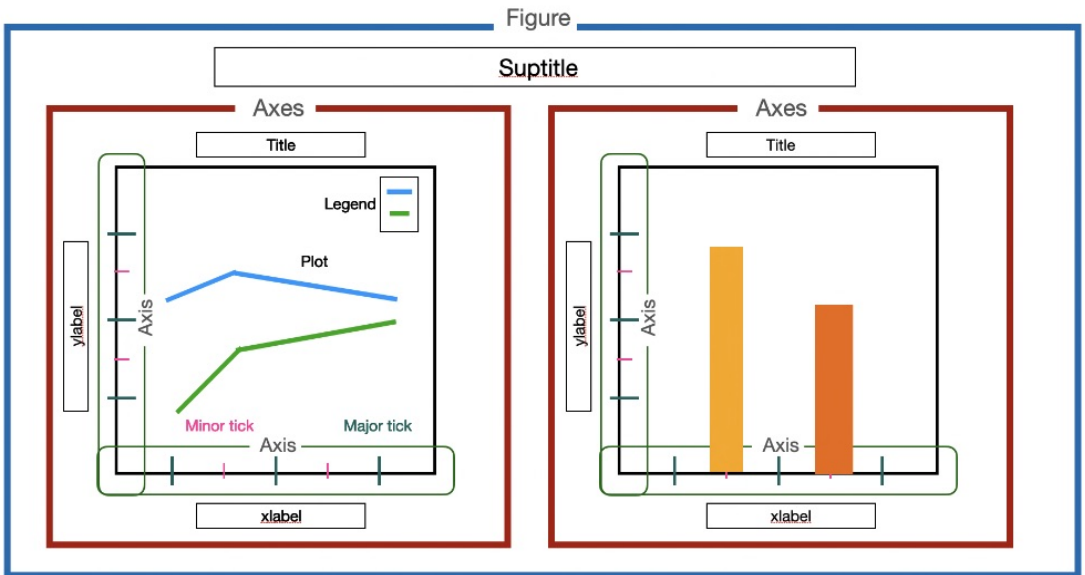
→ Exploratory → EDA ?

→ Explanatory → Showcase Insights.

DV → Art → how?

DV → Science → why?

C → Categorical   N → Numerical/Continuous
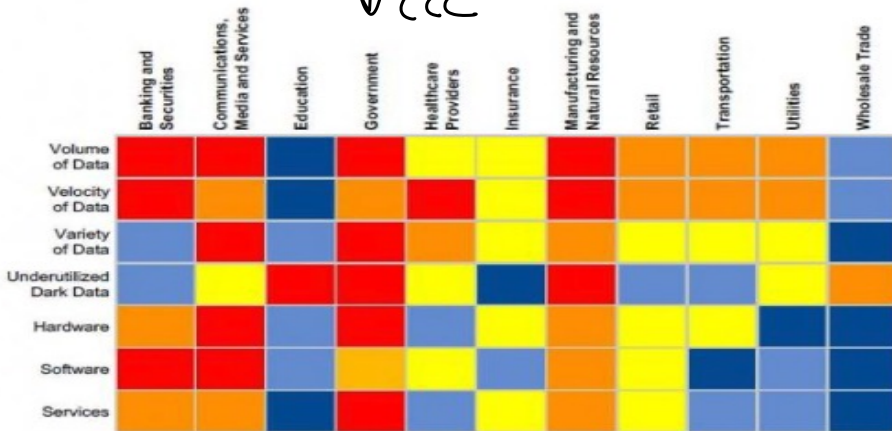
* Univariate
- **C** — Barplot with plt, Countplot (sns), Piechart
- **N** — Histogram, KDE Plot, Boxplot, Violin Plot
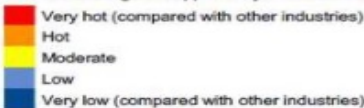
* Bivariate
- **CC** — Stacked bar plot, dodged barplot, Countplot with Hue,
- **CN** — Boxplot, Violin plot, barplot (sns).
- **NN** — Line plot, Scatter plot, heatmaps, joint plot, pair plot

* Multivariate
- **NNN** — Scatter with size
- **CNN** — Scatter with hue, lineplot with hue, joint plot with hue
- **CCN** — Boxplot with hue, Violin plot with hue,
- **CCC** — [   ]
       ↓ CCC

	Banking and Securities	Communications, Media and Services	Education	Government	Healthcare Providers	Insurance	Manufacturing and Natural Resources	Retail	Transportation	Utilities	Wholesale Trade
Volume of Data											
Velocity of Data											
Variety of Data											
Underutilized Dark Data											
Hardware											
Software											
Services											

Potential big data opportunity on each dimension is:
- Very hot (compared with other industries)
- Hot
- Moderate
- Low
- Very low (compared with other industries)

BOX AND WHISKERS PLOT.

Data: 14, 51, 1, 2, 6, 56, 7, 8, 9, 3, 4, -13, 6

100
0-25
25-50
50-75
75-100

Order: $\boxed{-13}$, 1, 2, 3, 4, 6, $\boxed{6}$ 7, 8, 9, 14, $\boxed{51, 56}$

Q1 lower quartile
25%

Median middle Q2 quartile
50%

upper quartile Q3
75%

outliers   min   max   Outliers.

-13 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 51 56
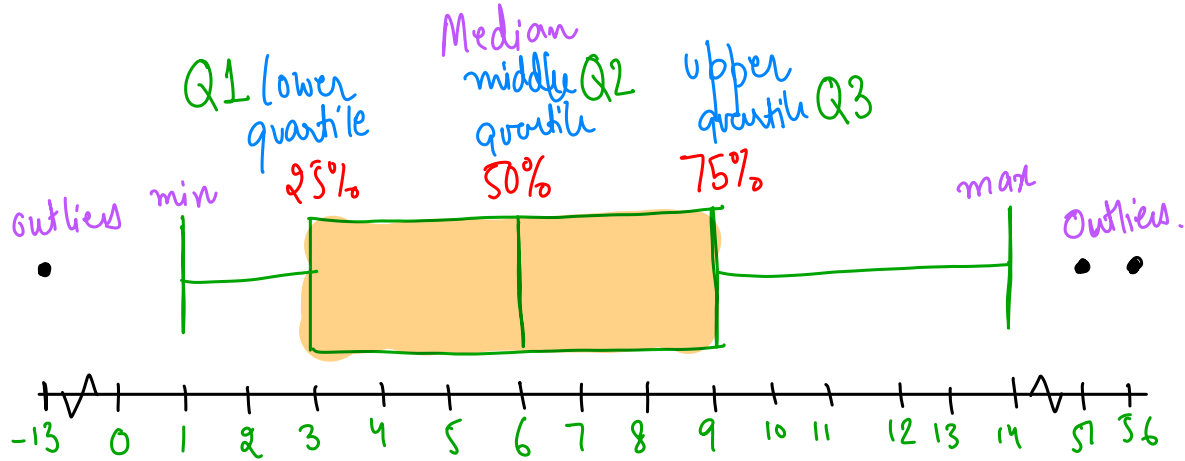
IQR ⇒ Inter Quartile Range

⇒ Q3 - Q1 ⇒ 6

$Q3 = 9$
$Q1 = 3$

IQR = 6

MAX = Q3 + $\boxed{1.5}$ IQR ⇒ $9 + 1.5(6) = 18$

MIN = Q1 - $\boxed{1.5}$ IQR ⇒ $3 - 1.5(6) = -6$

DATA

discrete [1, 2, 3, 4, 5]

Continuous [1.1, 1.11, 1.5, 2, 2.1]

1   2

Categorical

Nominal

{ India, Pakistan, USA }

{ Electronic, Home, Books }

Ordinal

{ High, Medium, low }

{ upper, middle, lower }

Continuous

Interval

Temp
Time
2023

| O |
| 0°C |

23.3°C

Ratio

Weight
Height

| 0 cm |
| 0 kg |
| 0 → |

absence of quantity

29.12 kg

→ (Avg) distance

Not a good metric

Uber Cars.

Outliers

1, 2, 3, 4, 6, | 100, 200

Avg $= \dfrac{316}{\cancel{5} 7} = 45.14$

→ Better Agg. ⇒ Median

Clipping
Capping
% percentile based
Impute

DBSCAN

Distribution   Mean

1 5, 5, 5, 5, 5, 9   ⇒ 5

1, 3, 3, 4, 5, 6, 9, 9   ⇒ 5