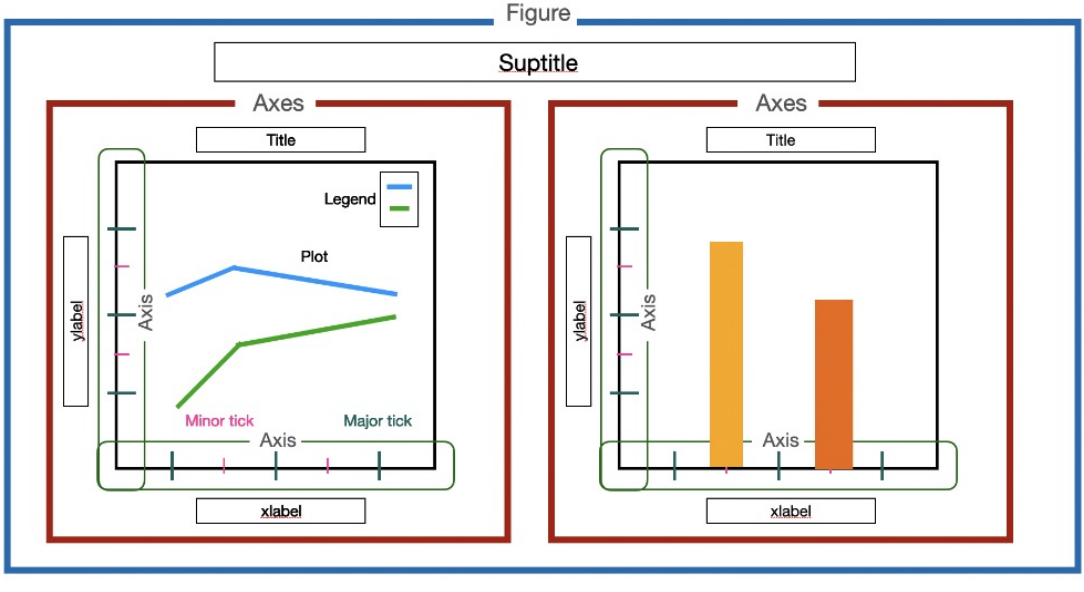
DATA VISUALISATION

→ Exploratory → EDA }

→ Explanatory → Showcase Insights.

DV → Art → how?

→ Science → why?



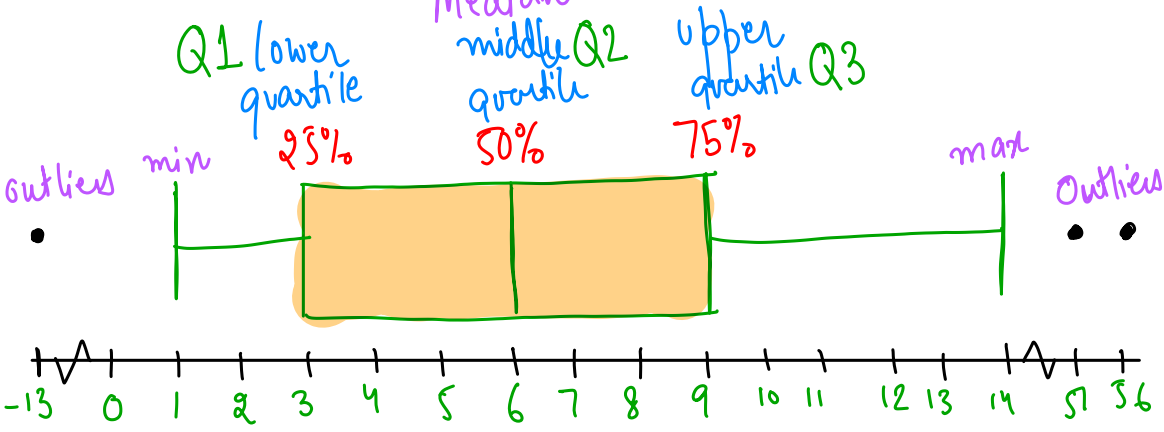
* Univariate
  - C - Barplot with plt, Countplot (sns), Piecharts (proportions)
  - N - Histograms, KDE, Box, VIOLIN
* Bivariate
  - CC -
  - CN -
  - NN -
* Multivariate
  - NNN -
  - CNN -
  - CCN -
  - CCC -

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



IQR ⇒ Inter Quartile Range

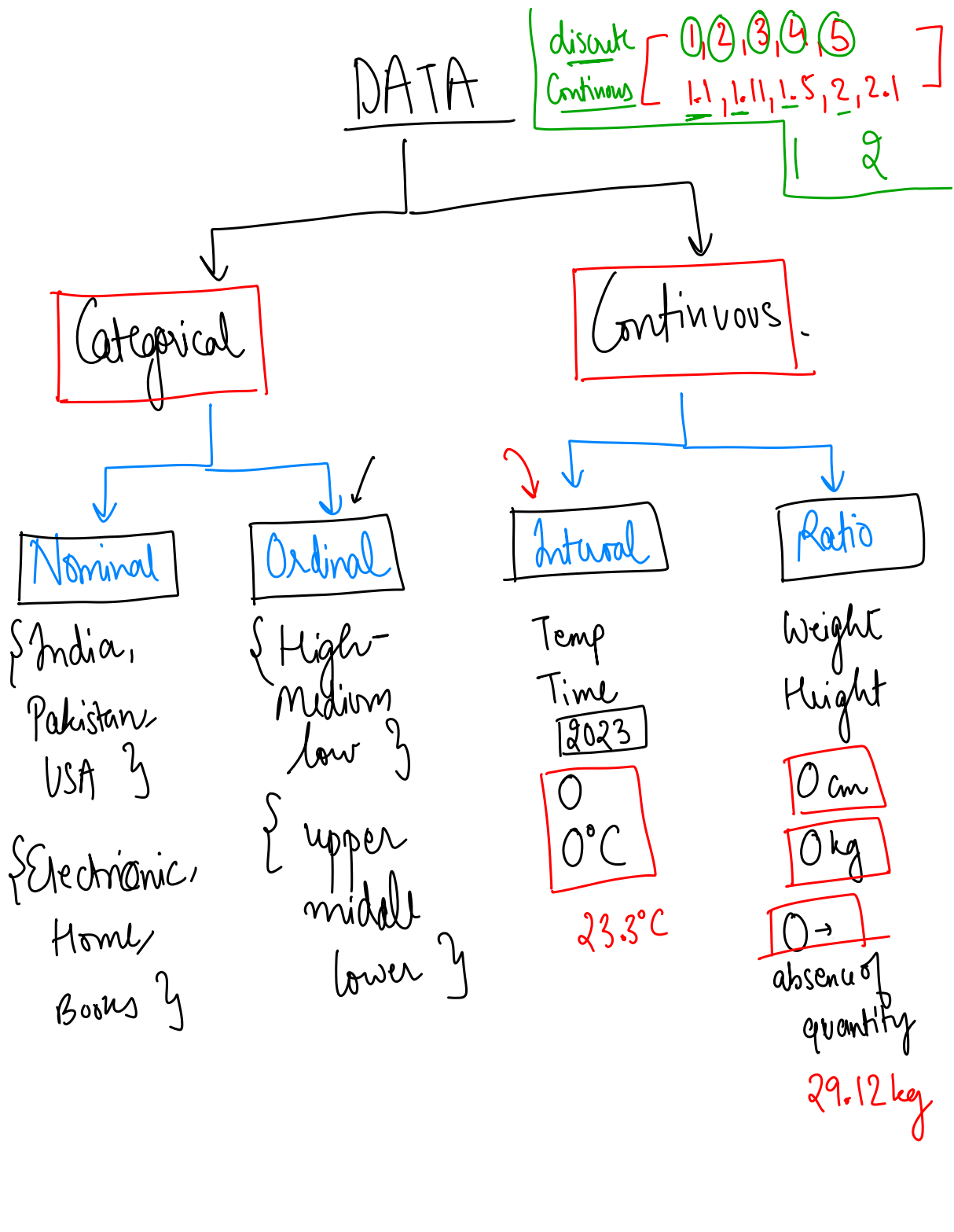
⇒ Q3 − Q1 ⇒ 6

Q3 = 9
Q1 = 3

IQR = 6

MAX = Q3 + 1.5 IQR ⇒ 9 + 1.5(6) = 18

MIN = Q1 − 1.5 IQR ⇒ 3 − 1.5(6) = −6


DATA
discrete
Continuous
1, 2, 3, 4, 5
1.1, 1.11, 1.5, 2, 2.1
1
2
Categorical
Continuous
Nominal
Ordinal
Interval
Ratio
{India, Pakistan, USA}
{Electronic, Home, Books}
{High-Medium-low}
{upper middle lower}
Temp
Time
2023
0
0°C
23.3°C
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 = $\frac{316}{7}$ = 45.14

→ Better Agg. ⇒ Median

Clipping
Capping
% percentile based
Impute
DBSCAN