Agenda

- Intro to matplotlib and seaborn
- New Dataset : Tencent Games
- Anatomy of a Figure
- Univariate Charts
  - Categorical
    - BarChart
    - Pie-chart
    - Countplot
  - Continous
    - Histogram
    - KDE Histogram
    - Box-plot

- Matplotlib : Core library for plotting charts

- Seaborn : - The syntax is easy
  - lot of themes, colors and customization

- plotly / D3.js / Altair

Data ?

- Row and Colums

  Rows : Dat-point, Records, Samples

  Columns: Variables, features, attributes

Type of cols / features

① Continuous / Numerical

- They can take any value in range

Ex:

- Sales

② Categorical

- Discrete

Year	Genre
2008	Shooter
2009	...

- Repeated across Dataset

→ Ordinal

2008 < 2009

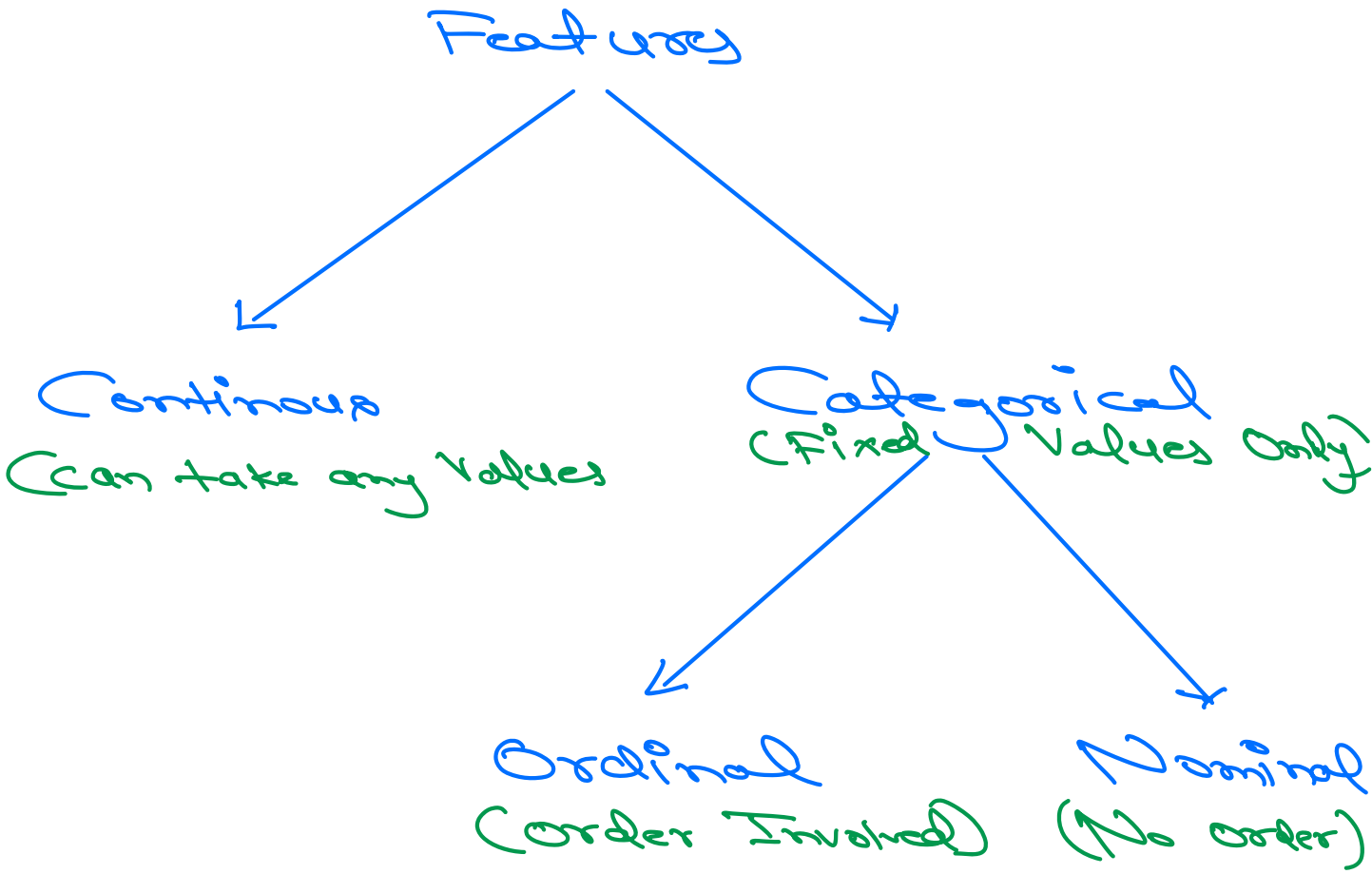
- Have some Order among Categories

Ex: Year

→ Nominal / Non-ordinal

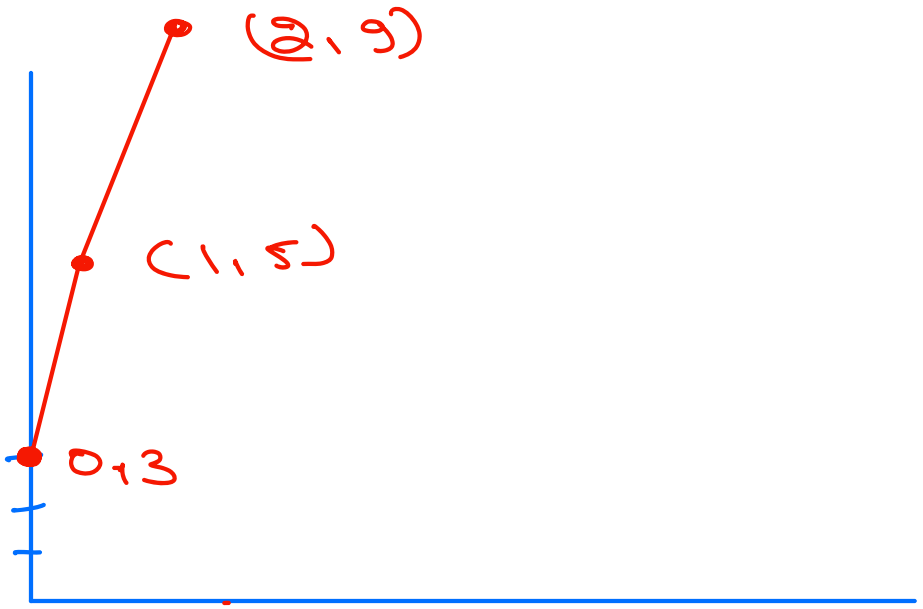
→ No Order among Categories

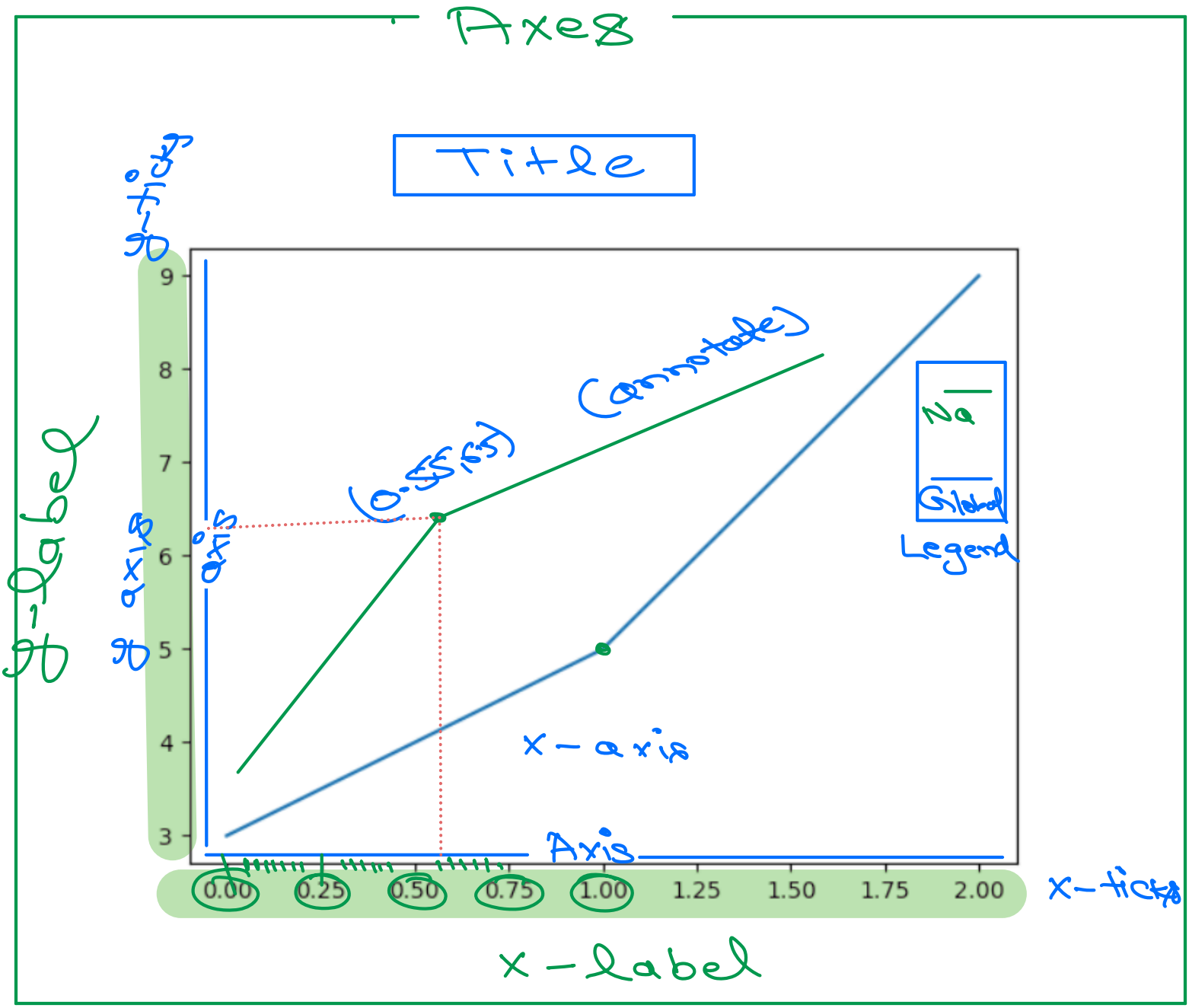
Ex: Genre



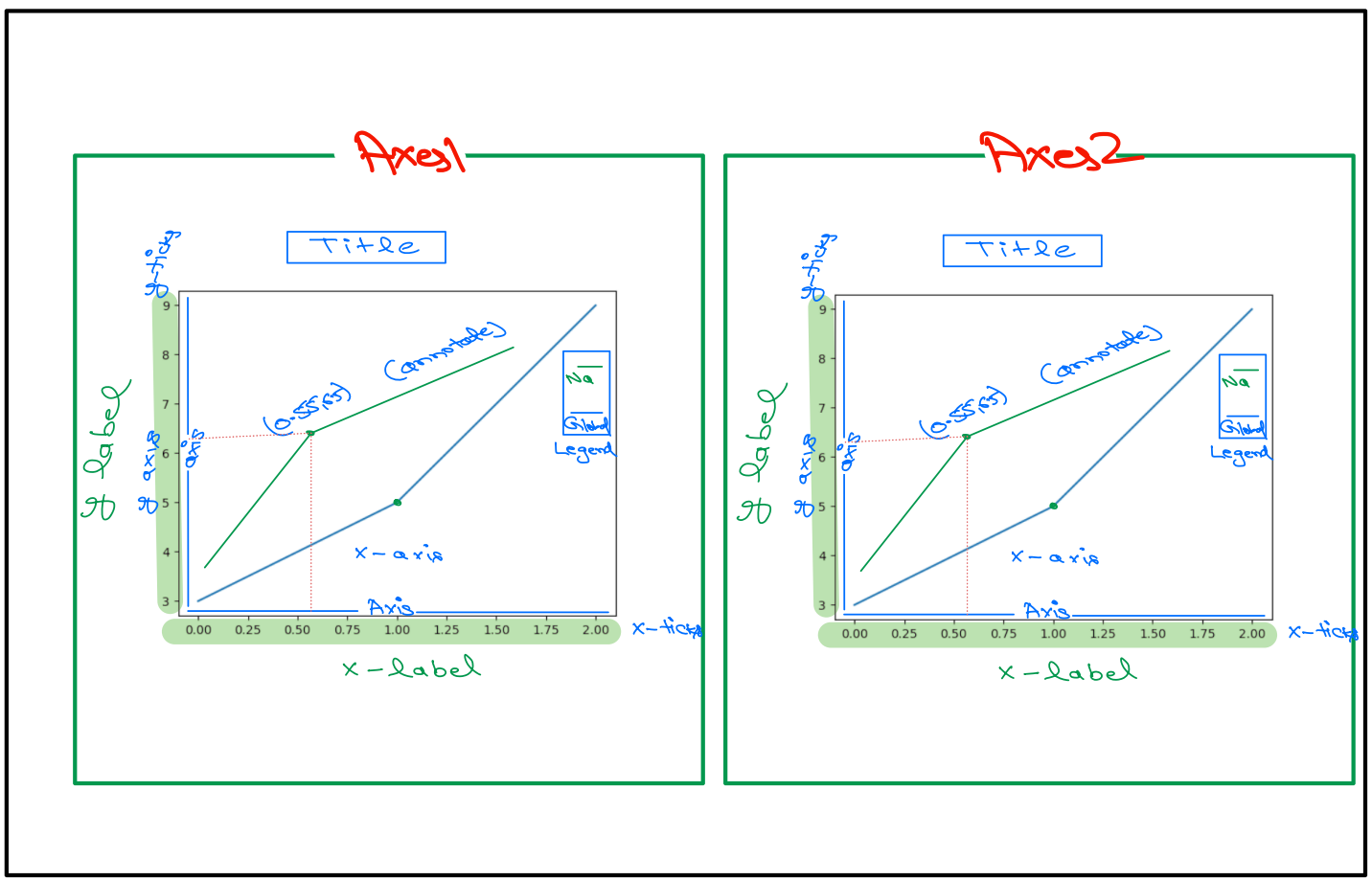
Ex :,

Gender → Nominal Categorical Var




Axes
Title
y-label
y-axis
y-ticks
(0.55,6.5)
(connected)
Na
Global
Legend
x-axis
Axis
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
x-ticks
x-label

- Title
- Axis
- ticks
- legend



Figure

- Figures
  - Sub figures (1)
  - Sub figures (2)

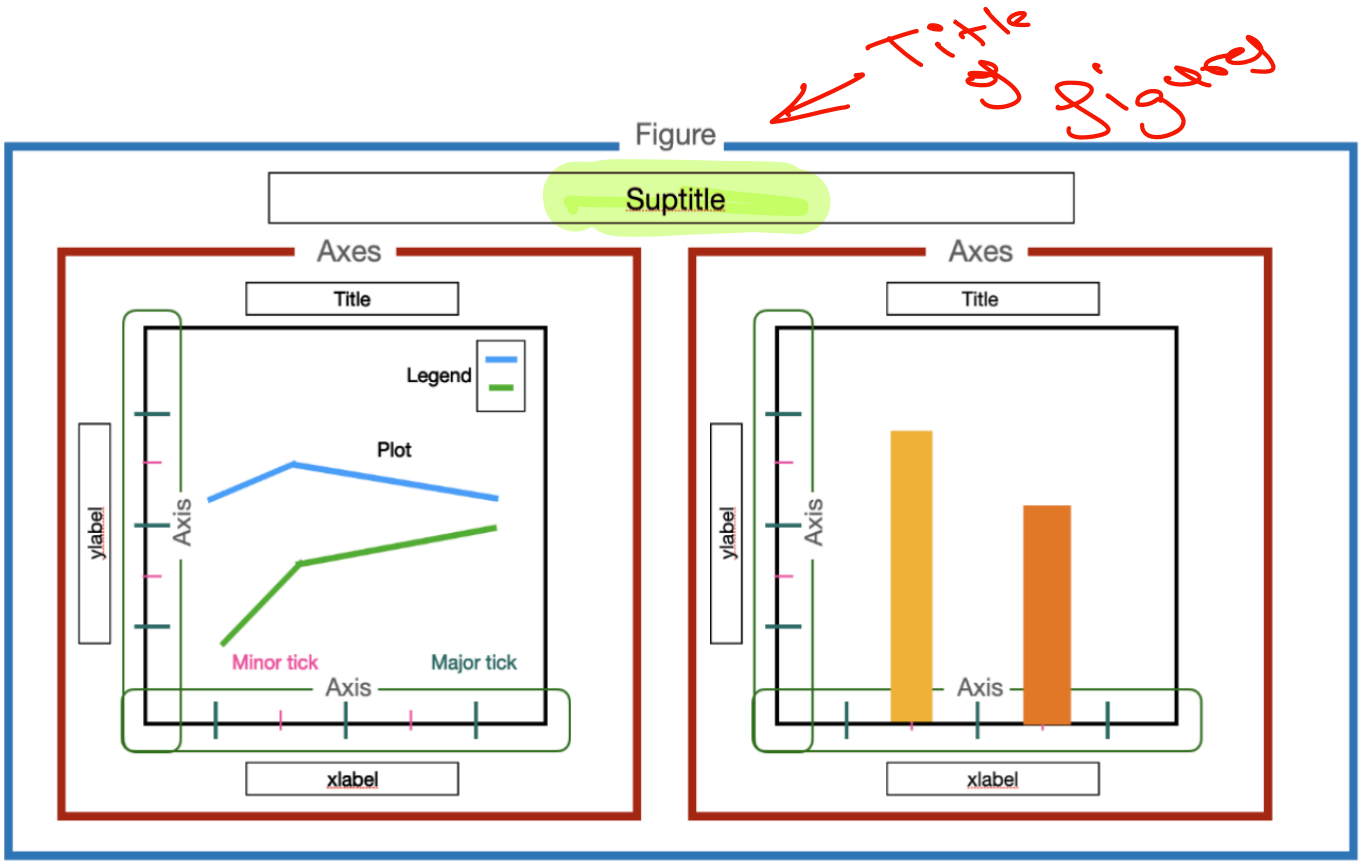
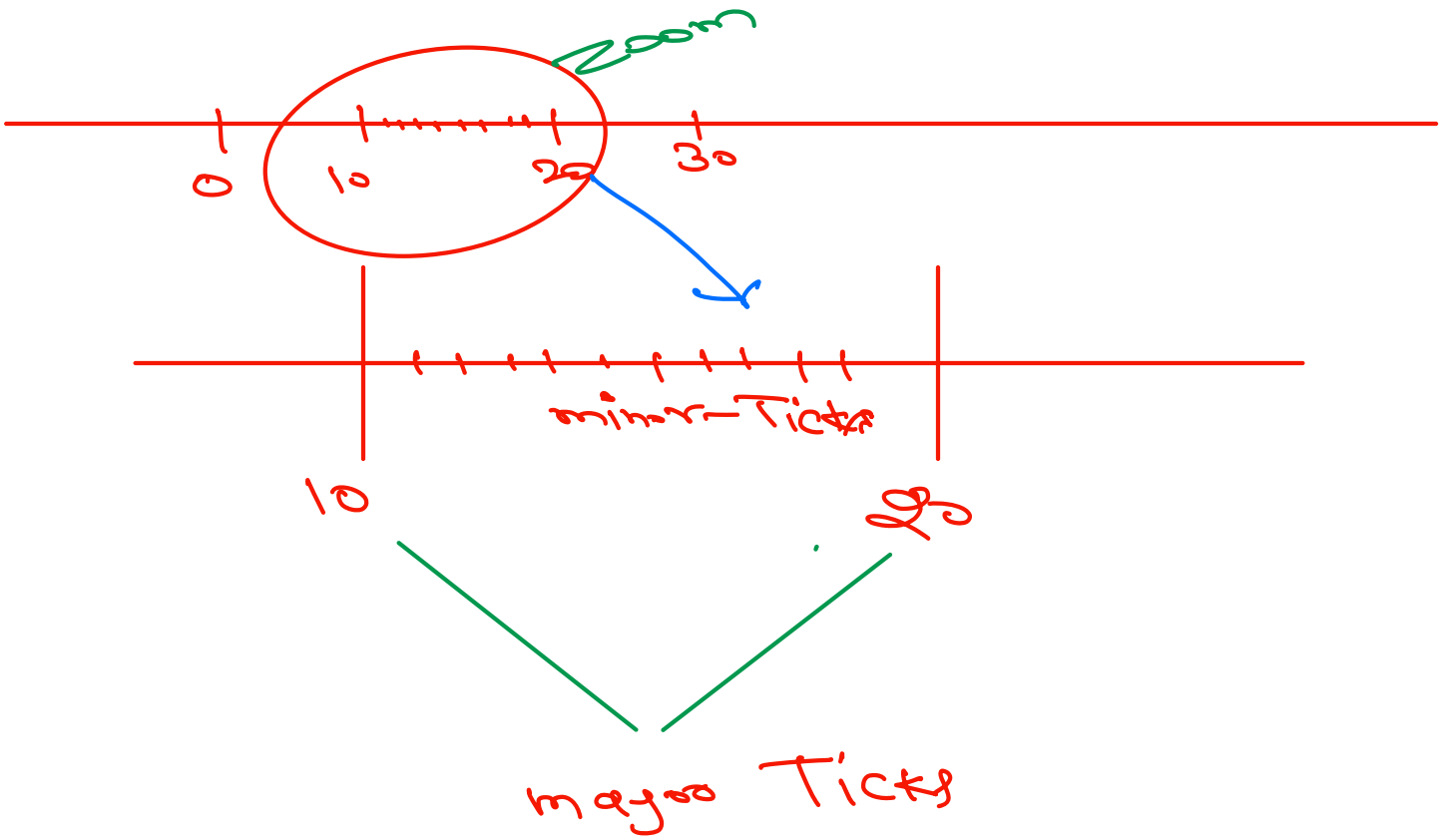


Figure
Suptitle
Axes
Title
Legend
Plot
ylabel
Axis
Minor tick
Major tick
Axis
xlabel
Axes
Title
ylabel
Axis
Axis
xlabel

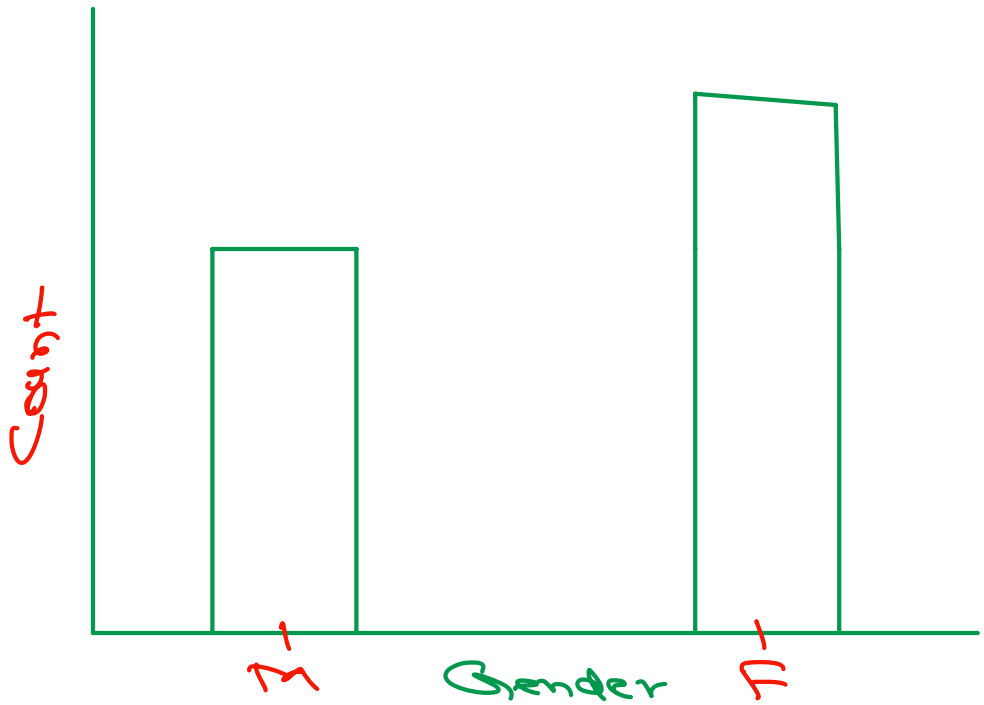

0
10
20
30
10
20

* Type of Analysis

- Univariate → Single Variable
  - → Frequency distributed, Histogram
    - Numerical
    - Categorical
- Bi-variate Analysis → 2 Vars
  - Analyzing Relationship b/w Variable
    - Numerical - Numerical
    - Numerical - Categorical
    - Categorical - Categorical
- Multi-variate → More than 2
  - 3
    - N - N - N
    - N - N C
    - N - C - C
    - C - C - C

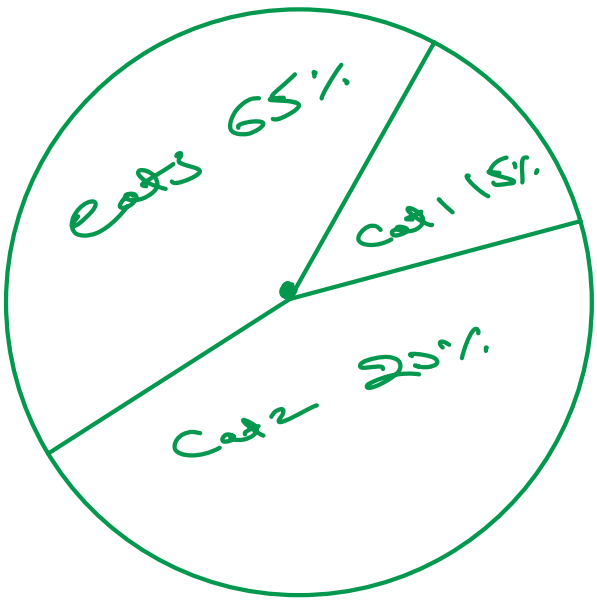
Bar-chart

- Categorical Uni-variate
- Frequency Plot

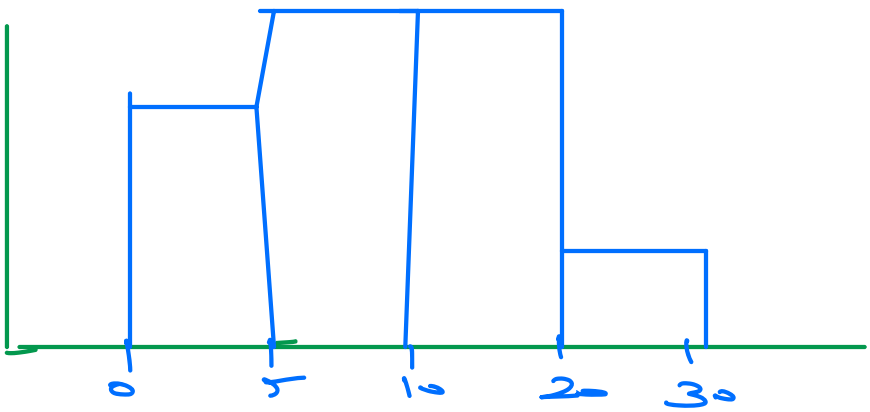


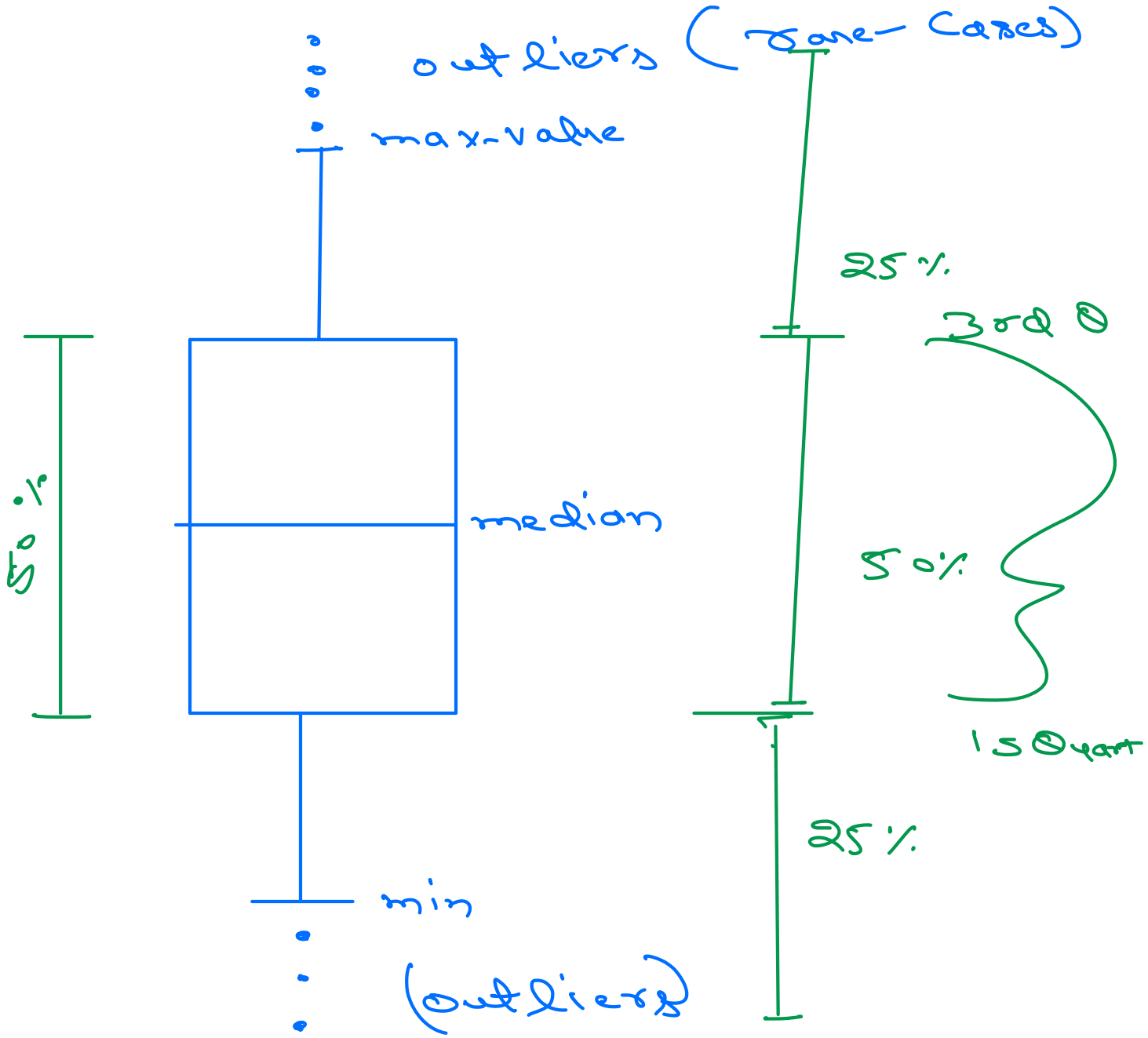
Pie-charts

- Categorical Uni-variate
- proportions / percentages / Ratios



- Continous Univariate
- Distributions




outliers
(rare cases)
max-value
25 %
3rd Q
50 %
median
50 %
1st Quart
25 %
min
(outliers)