

# Agenda

- Intro to matplotlib and seaborn
- New Dataset : Tencent Games
- Anatomy of a Figure
- Univariate Charts
  - Categorical
    - ⇒ Bar Chart
    - ⇒ Pie-chart
    - ⇒ Countplot
  - Continuous
    - ⇒ 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 Columns

Row : 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

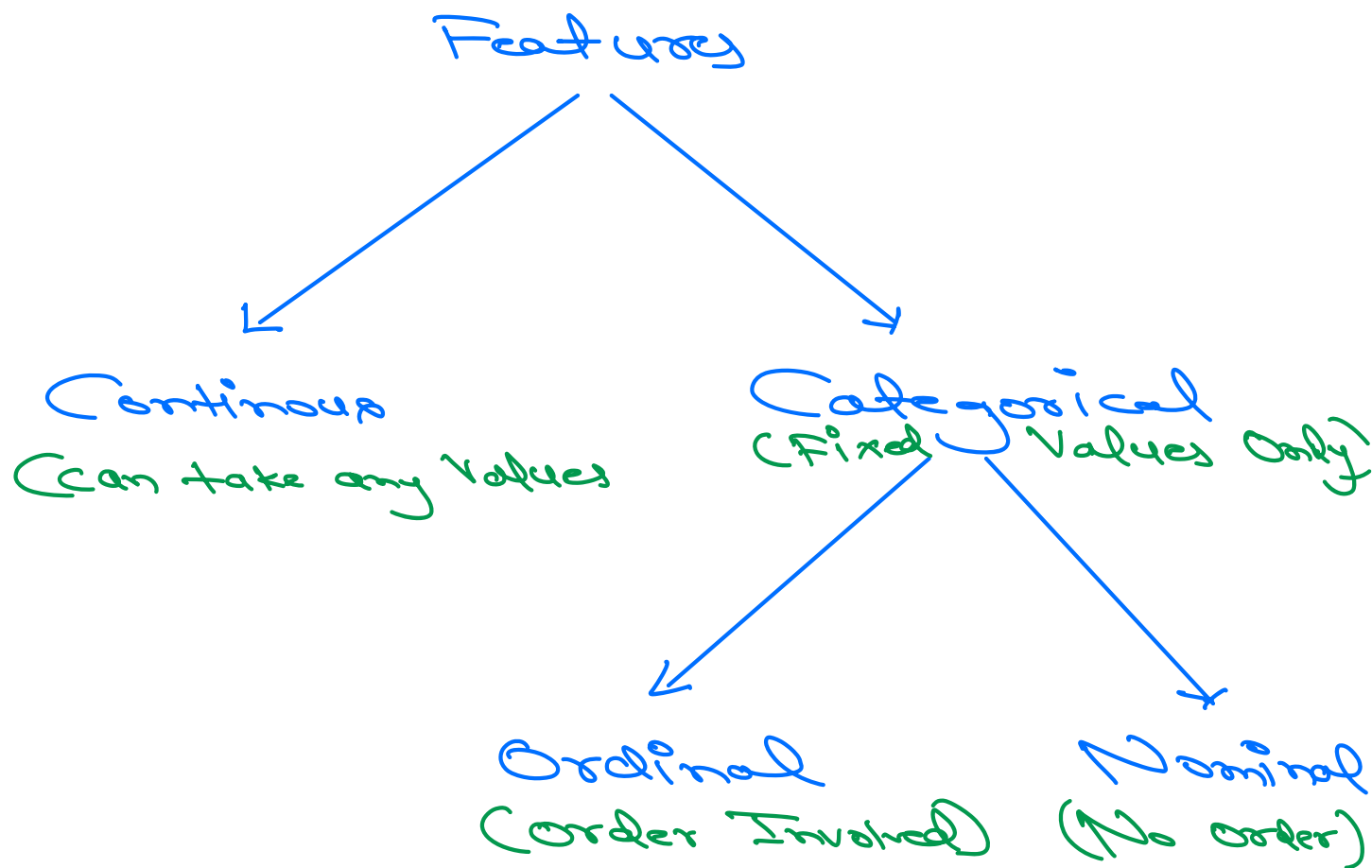
→ Have some Order among Categories

Ex: Year

→ Nominal/Non-ordinal

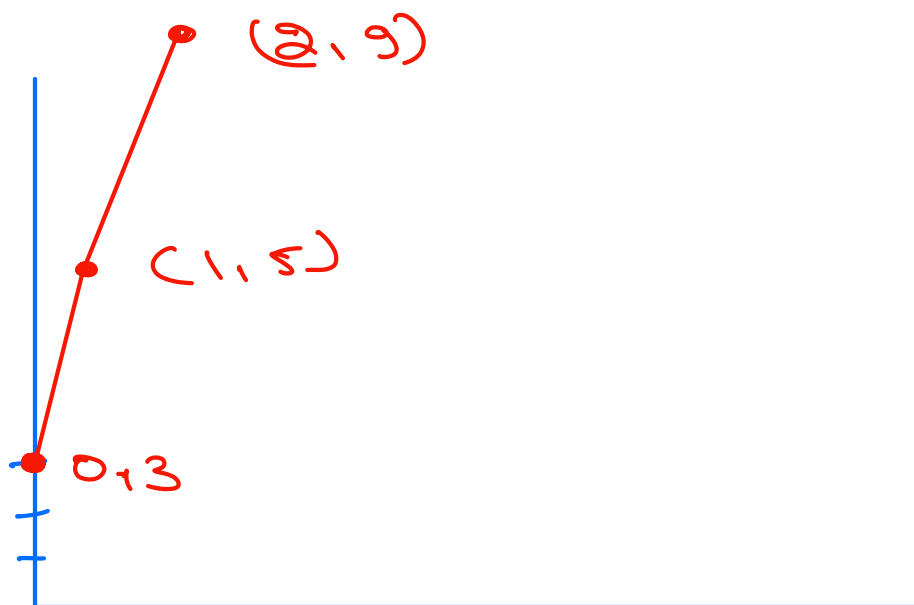
→ No Order among Categories

Ex: Genre



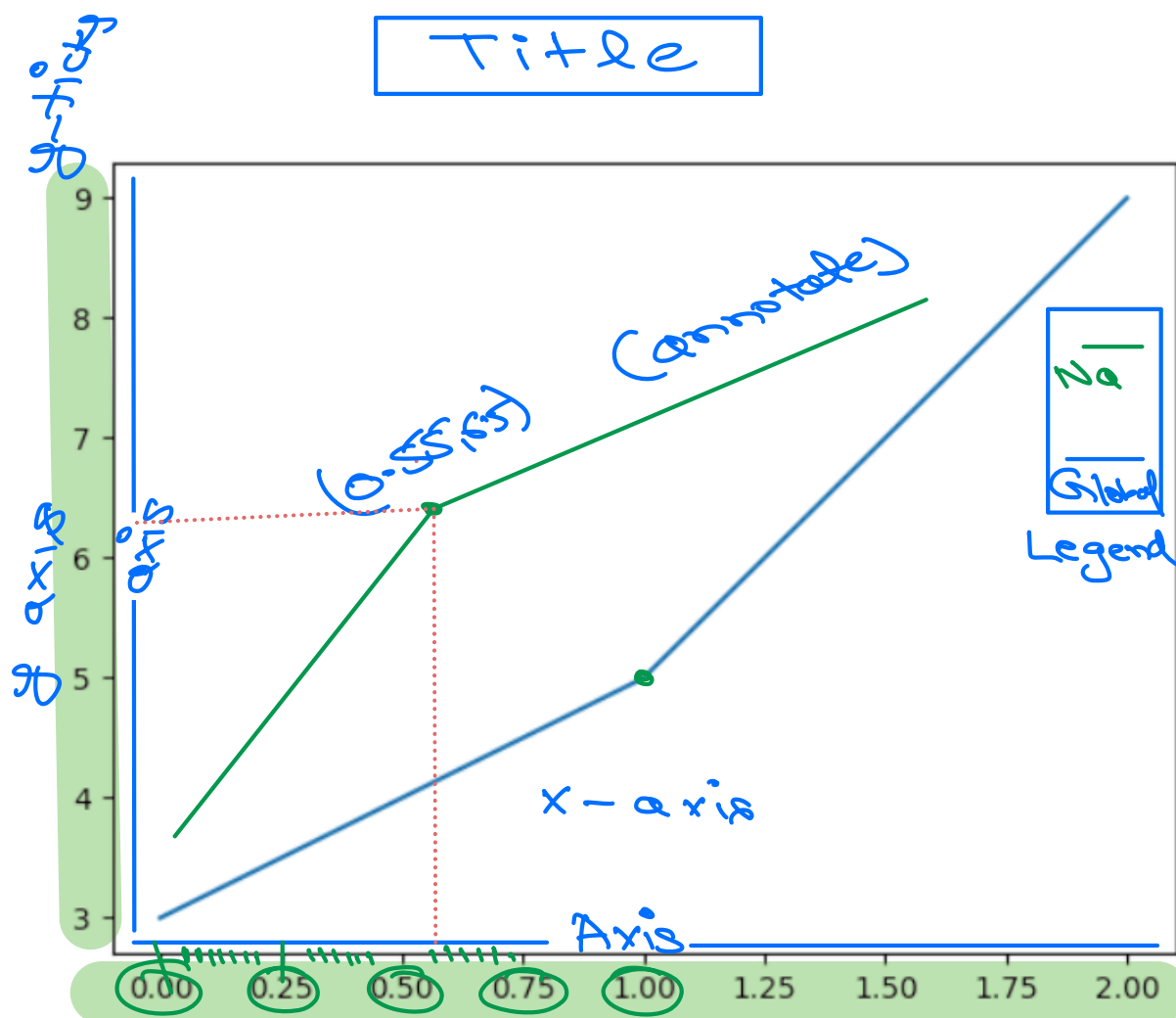
Ex ;

Gender  $\rightarrow$  Nominal Categorical Var



Axes

Title



x-axis

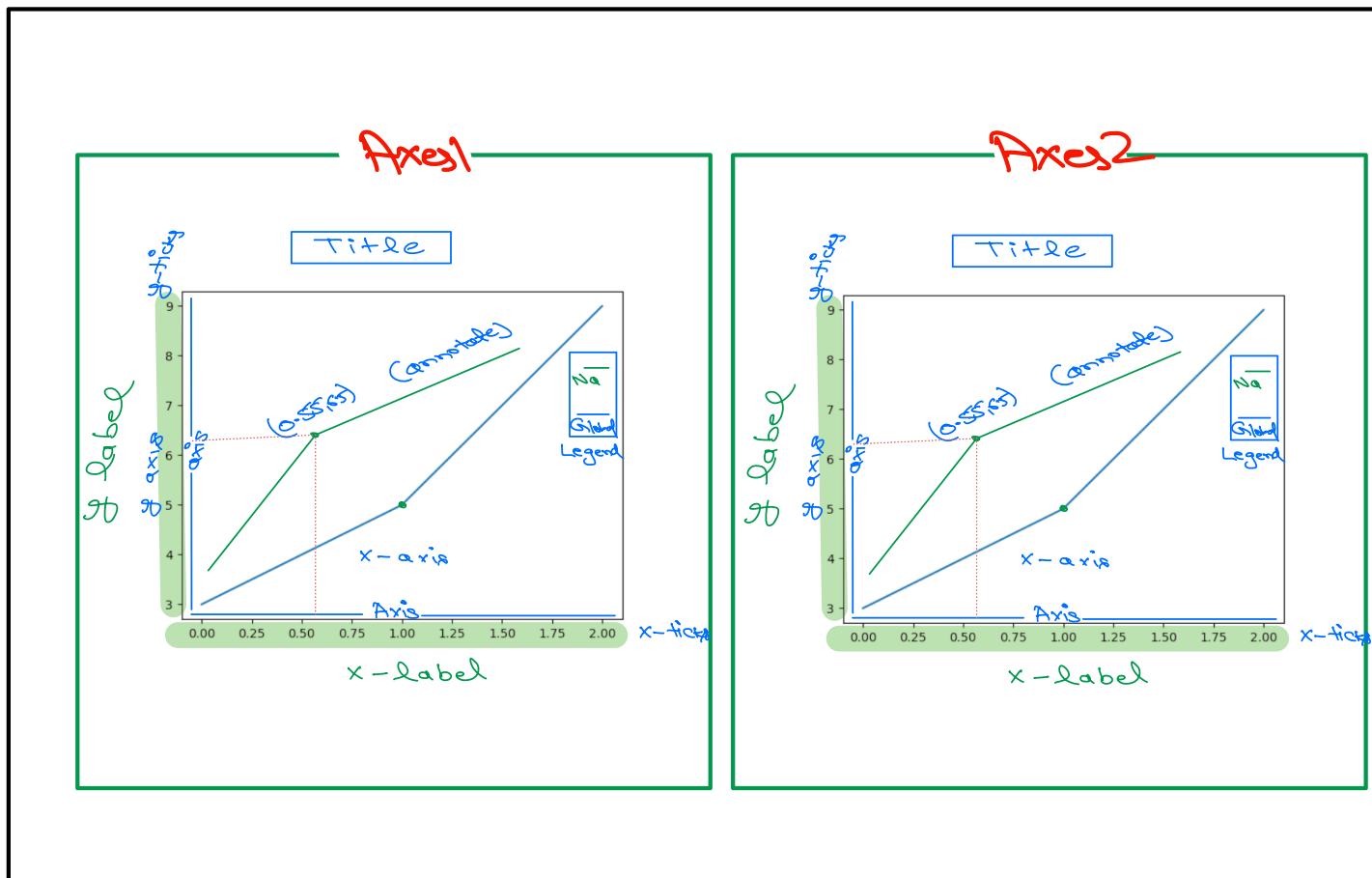
Axis

x-label

x-ticks

Legend

- Title
- Axis
- ticks
- legend



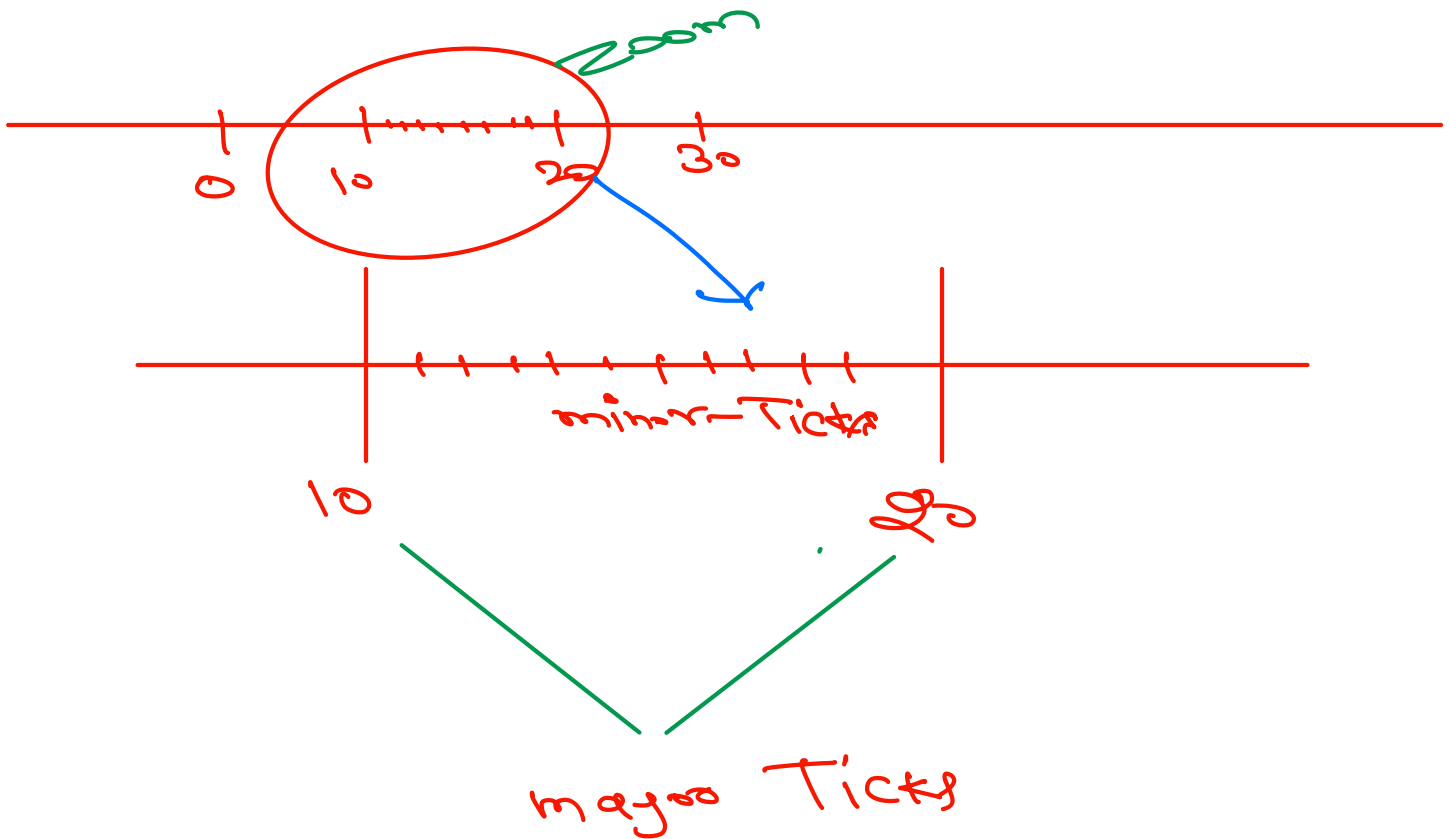
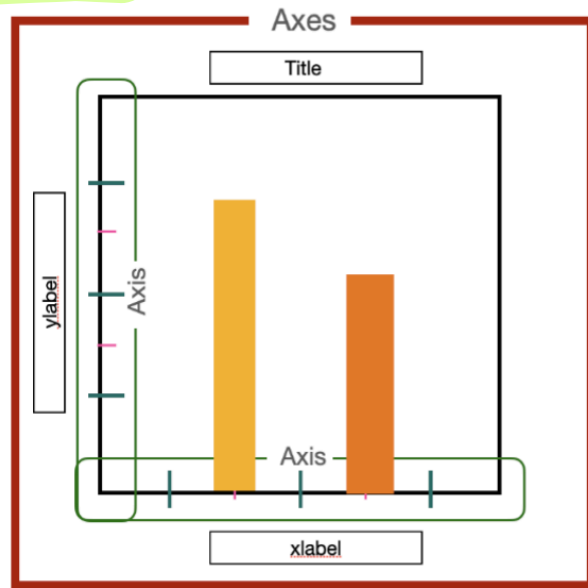
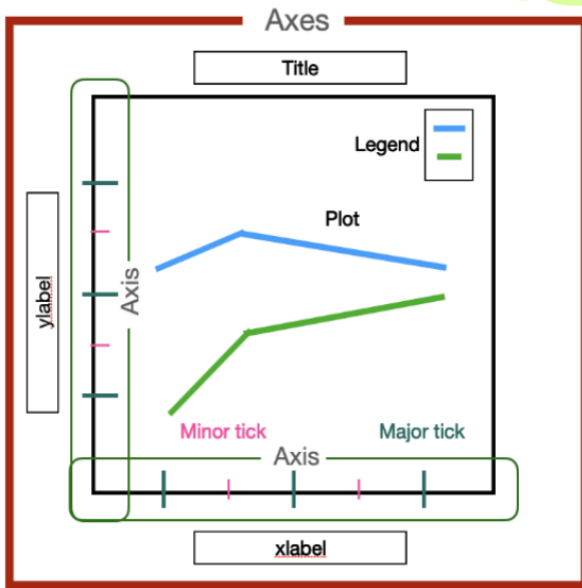
Figure

Figure < Sub-figure ①  
Sub-figure ②

← Title of figures

Figure

Suptitle



## \* Type of Analysis

1 Univariate  $\Rightarrow$  Single Variable

$\Rightarrow$  Frequency, Histogram  
distributed

$\Rightarrow$  Numerical

$\Rightarrow$  Categorical

2 Bi-variate Analysis  $\Rightarrow$  2 Variables

Analyzing Relationship

b/w Variables

$\Rightarrow$  Numerical - Numerical

$\Rightarrow$  Numerical - Categorical

$\Rightarrow$  Categorical - Categorical

3 Multi-variate  $\Rightarrow$  More than 2

3

$\Rightarrow$  N - N - N

$\Rightarrow$  N - N - C

$\Rightarrow$  N - C - C

$\Rightarrow$  C - C - C



# Bar-chart

- o Categorical Uni-variate
- o Frequency plot

