


Data Visualization - 1

Start at 9:03 PM

⇒ Data viz → 3 classes

⇒ Why is data viz important →

- Exploratory → Exploration of Data) → EDA
- Explanatory → Storytelling

Q Why do we learn data viz in python?

- Quick Analysis
- Unstructured data → (Tableau, Excel → Structured)
- Easy and wide manipulation options

1) What is the science in data viz?

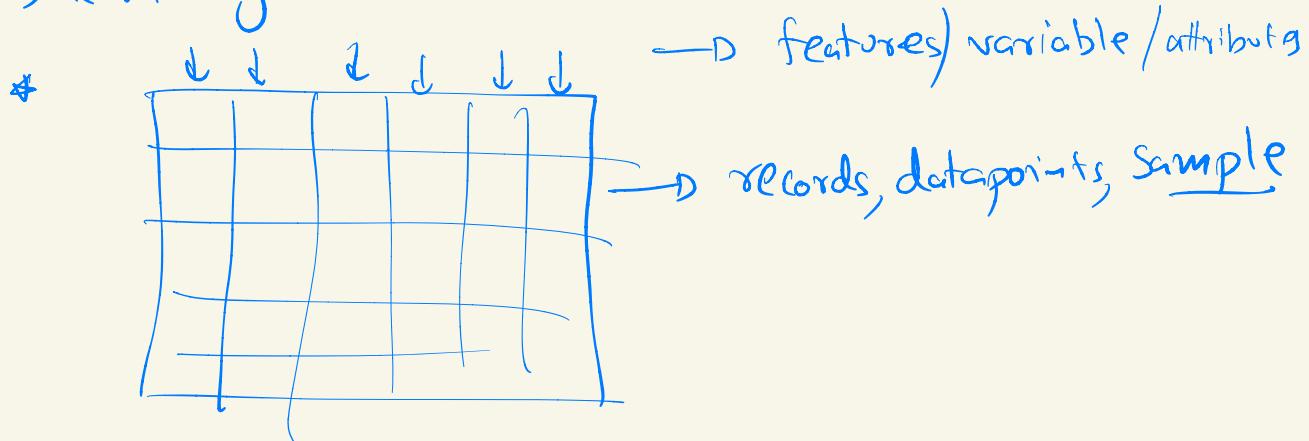
- Anatomy of chart →
- How to use the right plot

2) Art in data viz

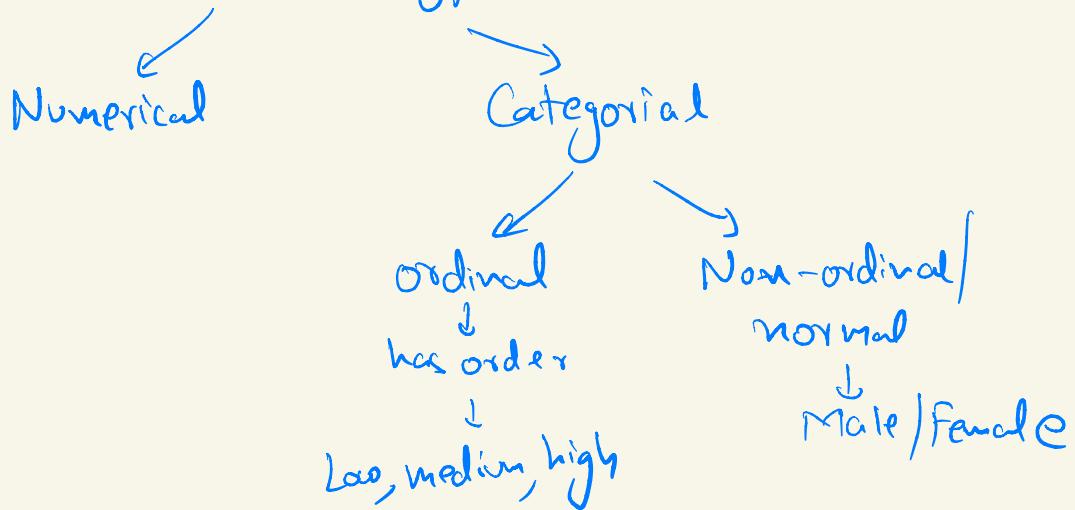
- Color, scale, labels
- Highlight

* Libraries → matplotlib
→ Seaborn → more Beautiful Version
of matplotlib

⇒ Terminologies



* Different data types →



* How to choose a right plot ??

(i) How many variables/features are involved

(ii) Variable is of what type \rightarrow Num/Categorical

(iii) \rightarrow Types of viz

1-variable \rightarrow Univariate

2-variable \rightarrow bivariate

3 or more \rightarrow multivariate

(iv) a) Univariate Data viz \rightarrow

Num \downarrow Cat \uparrow

b) Bivariate \rightarrow

Num - Num

Cat - Cat

Num - Cat

c) Multivariate (3) \rightarrow

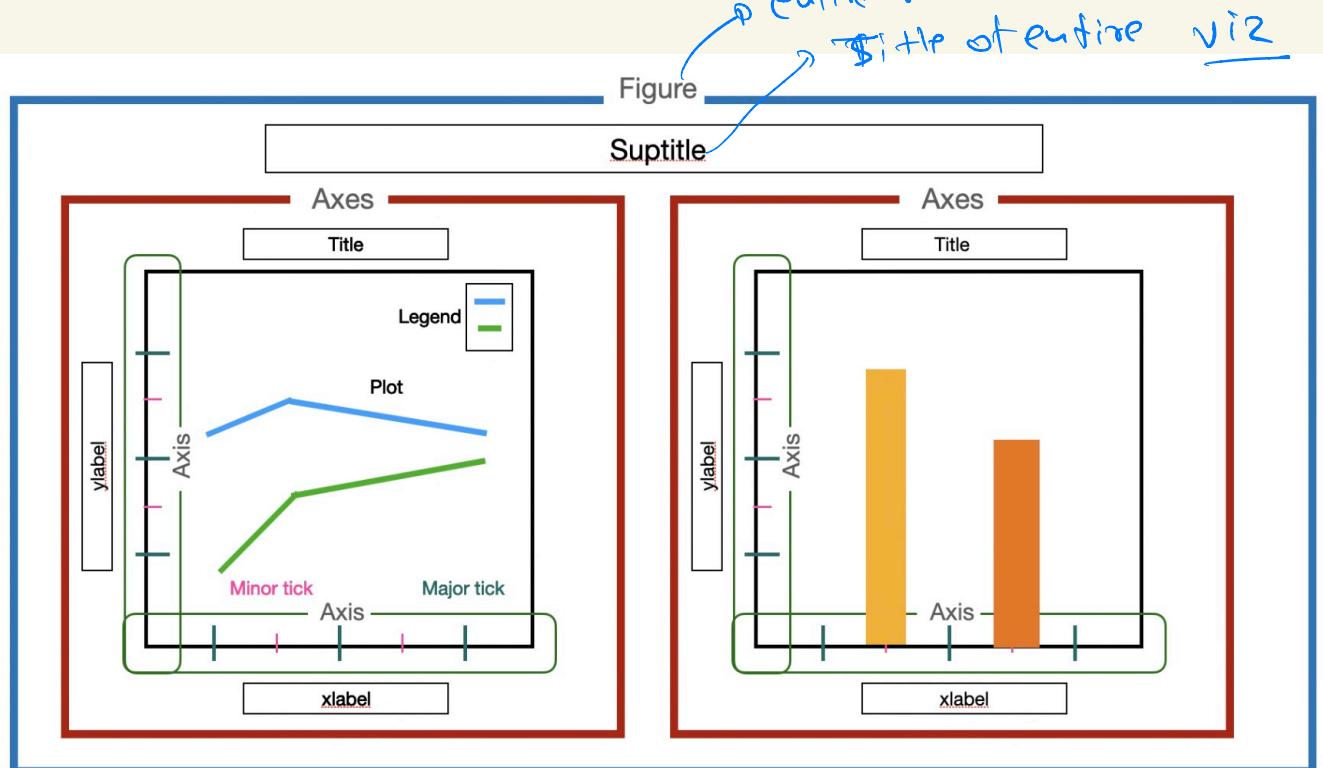
N - N - N

C - C - C

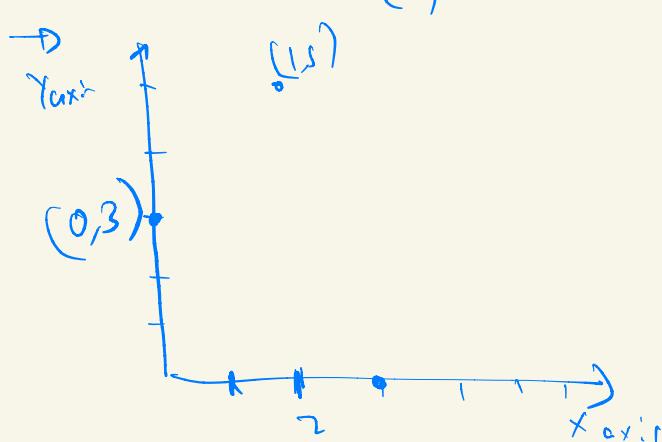
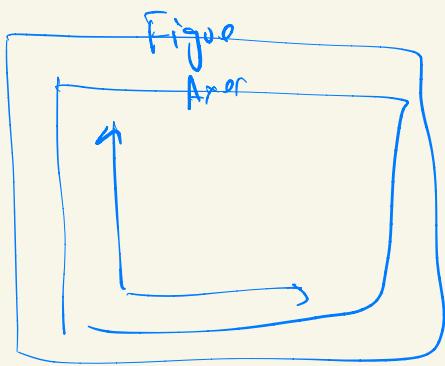
N - N - C

N - C - C

⇒ Anatomy of Matplotlib →



⇒



- $(0, 3)$
- $(1, 5)$
- $(2, 9)$

$$x\text{-values} = \{0, 1, 2\}$$

$$y\text{-values} = \{3, 5, 9\}$$

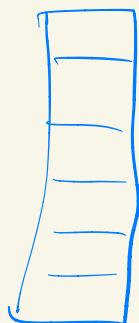
$$\begin{aligned} &x_1, y_1 \\ &x_2, y_2 \end{aligned}$$

\Rightarrow Univariate Analysis \rightarrow 1 variable

\rightarrow Categorical \rightarrow variable \rightarrow genre

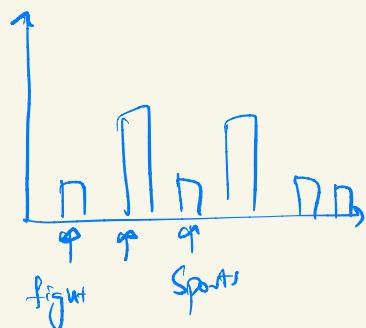
1) Count of each category. \Rightarrow Bar

2) % contribution for each category
 \hookrightarrow Pie



1) Bar \rightarrow all the categories \rightarrow

- Top 5 Categories
- Top 7



\Rightarrow matplotlib

↳ Complicated

• Take a lot of code

• Charts are not that great

↳ Seaborn

\Rightarrow Univariate \rightarrow Numerical

\Rightarrow From and To range \rightarrow Distribution

• Min, max, count, avg

\Rightarrow How is the data distributed

\Rightarrow Outlier in data \Rightarrow Is data skewed

\Rightarrow Special numbers \rightarrow min, max, range, ...

\Rightarrow Distribution

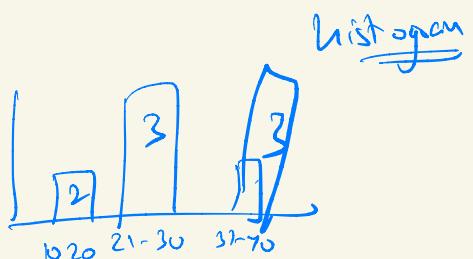
\hookrightarrow Histogram, Scatter

Bins \rightarrow divide your data into bin and frequency

Plot for

\Rightarrow 15, 20, 30, 24, 25, 36, 40, 38

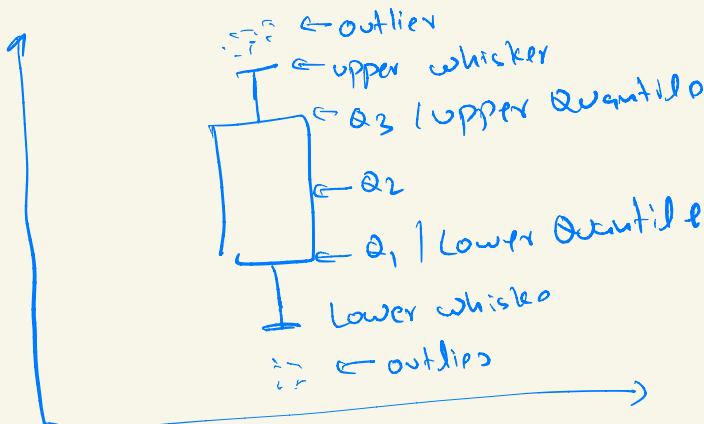
$\frac{18-20}{\downarrow} \quad \frac{21-30}{\downarrow} \quad \frac{31-40}{\downarrow}$
2 3 3



\Rightarrow • Popularity of video games over the years?

\downarrow
Histogram for the yearly data ~

\Rightarrow Box & whisker plot \rightarrow



Quantile

$Q_1 = 25$ percentile



25% value in data
is lower than this

$Q_3 = 75$ percentile



75% data is lower
than this

$Q_2 = 50$ percentile

Inter Quartile Range = $Q_3 - Q_1$

Upper whisker $\rightarrow Q_3 + 1.5 \times IQR$ | Max Score
Lower whisker $\rightarrow Q_1 - 1.5 \times IQR$ | Min Score

$\Rightarrow 1, 2, 3, 5, 3, 2, 1, 3, 7, 8, 9, 5, 7, 8, 300$

Max value $\rightarrow \underline{300}$ 9.5%

