Graphs for Analysis

→ ① Univariate (only one feature)
→ ② Bivariate (Two feature)
→ ③ Multivariate. (More than Two)
corr.

1. Categorical feature.

→ Frequency Distribution Table.

Ex- we have survey of 200 people and ask about their favorite type of vacation.

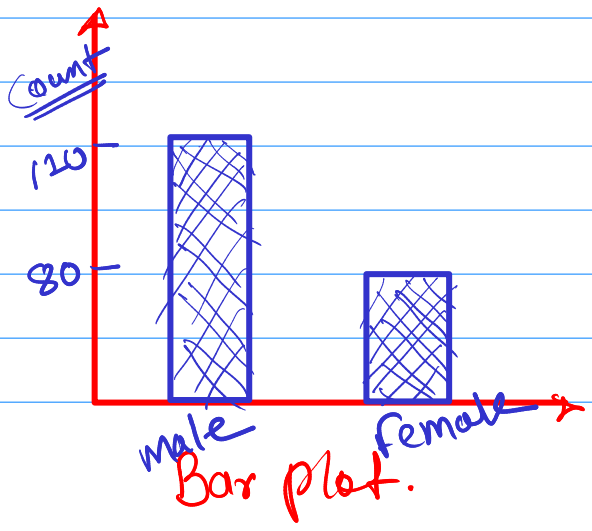
Type of Vac.	Frequency
Beach	60
City	40
Adventure	30
Nature	35
Cruise	20
other.	15
	200

Category

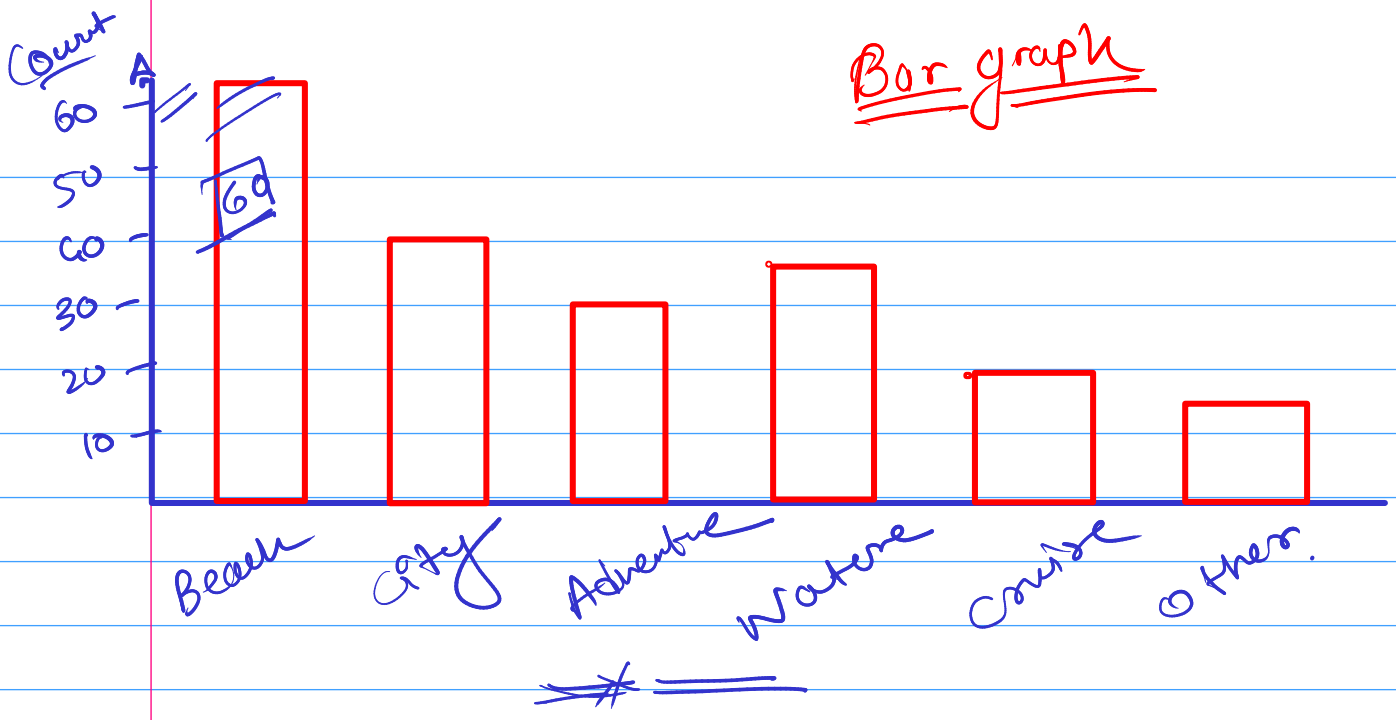
Category Gender	Freq.
Male	120
Female	80

Category

Frequency = count



Bar plot.

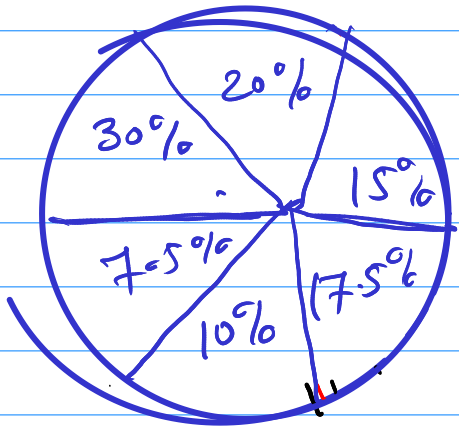


Bar graph

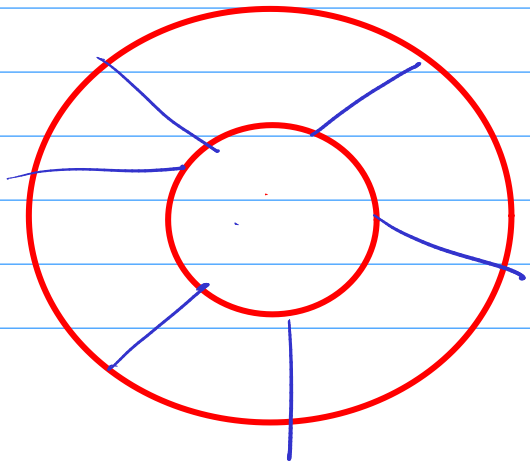
Proportion or percentage

(B) Relative Frequency.

Vacation	Frequency	Relative Freq.	
Beach	60	$60/200 = 0.3$	30%
city	40	$40/200 = 0.2$	20%
Adv.	30	0.15	15%
Nature	35	0.175	17.5%
Cruise	20	0.1	10%
other	15	0.075	7.5%
		1	100 %



pie chart

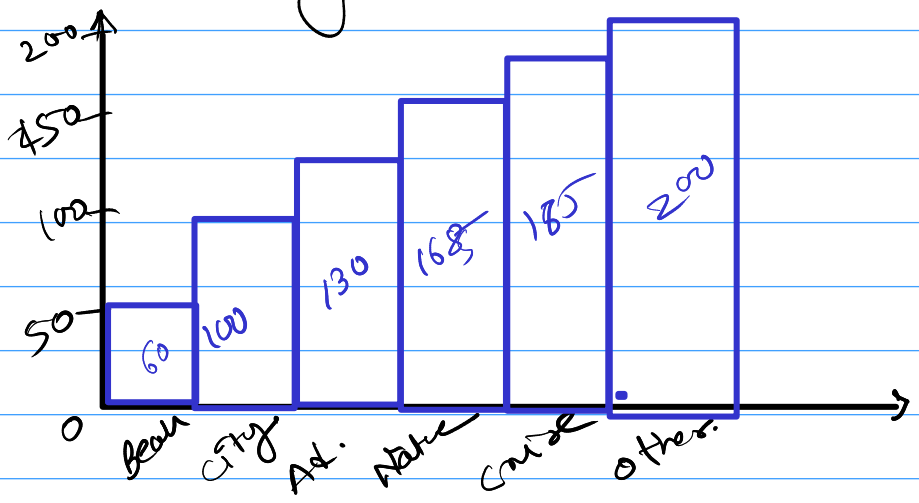


Donut chart

C. Cummulative Frequency

Vacation	Freq	Cum Freq
Beach	60	60
City	40	100
Adv.	30	130
Nature	35	165
Cruise	20	185
other	15	200

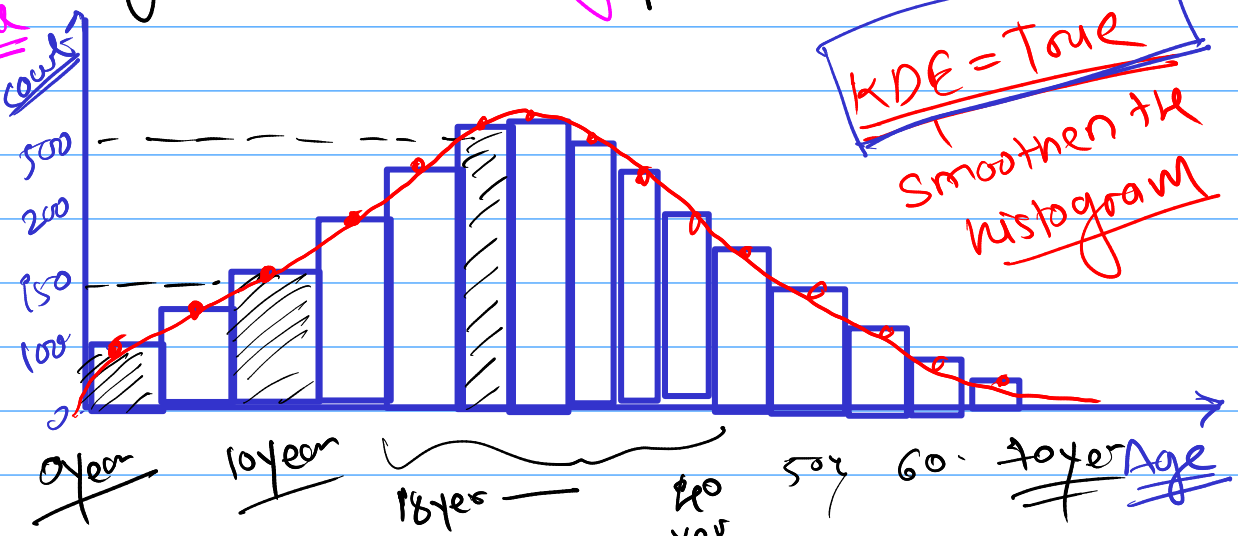
Waterflow chart.



2. Numerical Feature

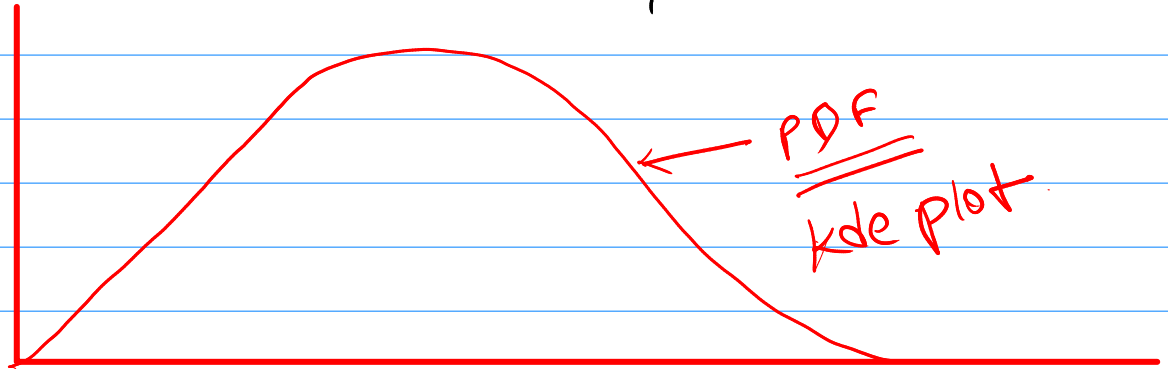
Age → Histogram → Frequency Distribution

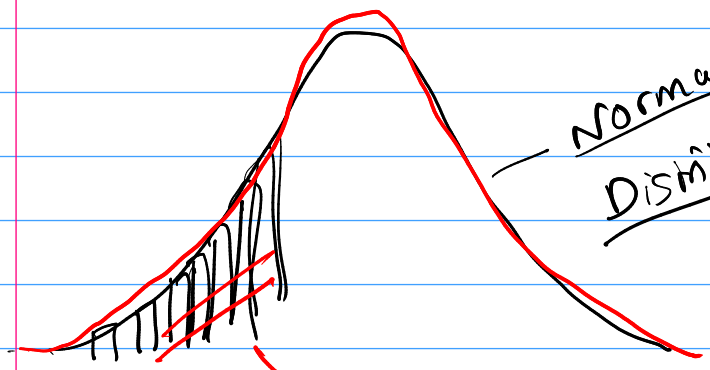
Age
18
19
20
25
30
35
02
07



KDE = True
Smoothen the histogram

KDE
Kernel density Estimator





Normal Distribution Type

weight Age

Gender
male = 60
fe = 60

sns.histogram (kde=True)

sns.kdeplot

Numerical convert into Categorical.

Age

1000 age

Age-group

Category group

0-10
11-20
21-30
31-40
41-50
51-60
61-70

Pegung

100 → 150

Age-group

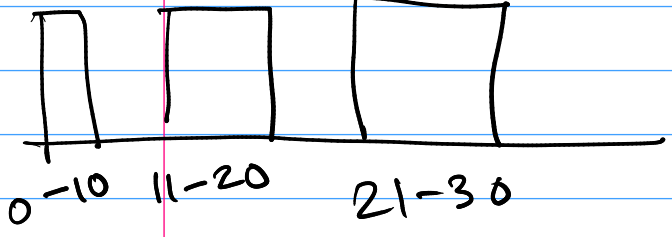
BAR

Child (0-18)
Teen (18-26)
Adult (25-40)
Senior (41-65)
Old age (66+)

Barplots

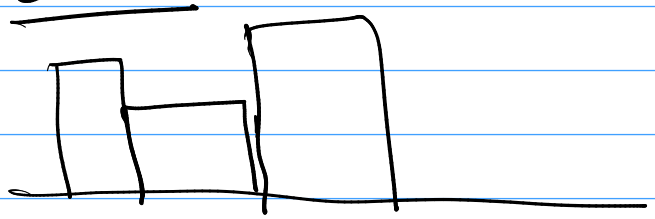
Bins → BIN

0–10
11–20
21–30



Bar

0–10
10–20
20–30



Histogram

2. Bivariate

① Categorical vs Categorical

② Cat vs Num / Num vs Cat

③ Num vs Num

1. Categorical Vs Categorical. ⇒ Cross Tab / Contingency Table.

Survived \ Coach	GN	SL	3AC	2AC	1AC
Death	31	29	71	83	02
Alive	120	80	06	98	78

2. Categorical Vs Num / Num Vs Cat

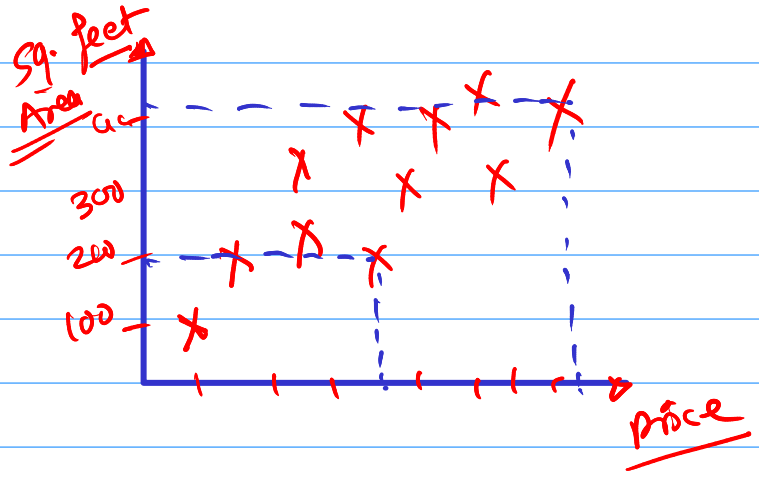
eg = | Gender | Age |
|---|---|
| M, F } 200 | 200 |

Step1 → numerical convert into categorical.

Step2 → Contigency Table.

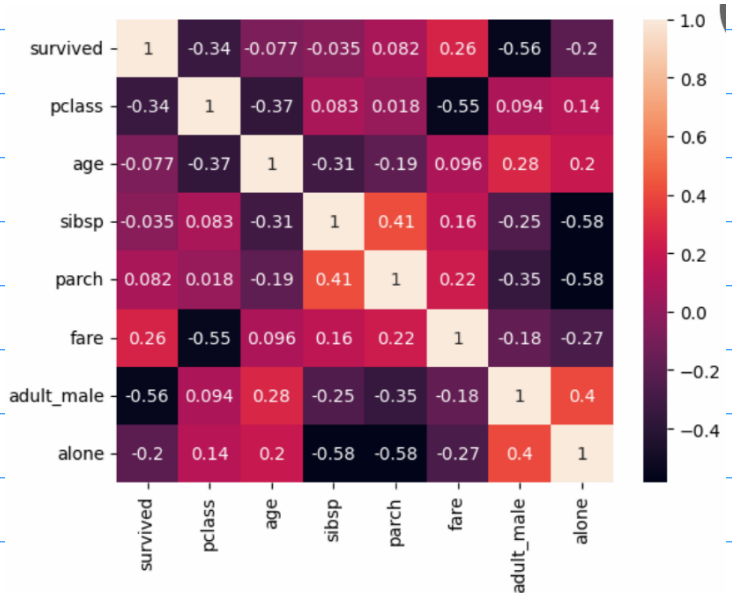
Age	0-10	11-20	21-30
male	32	41	110
female	15	18	120

3. Num Vs Num → Scatter plot

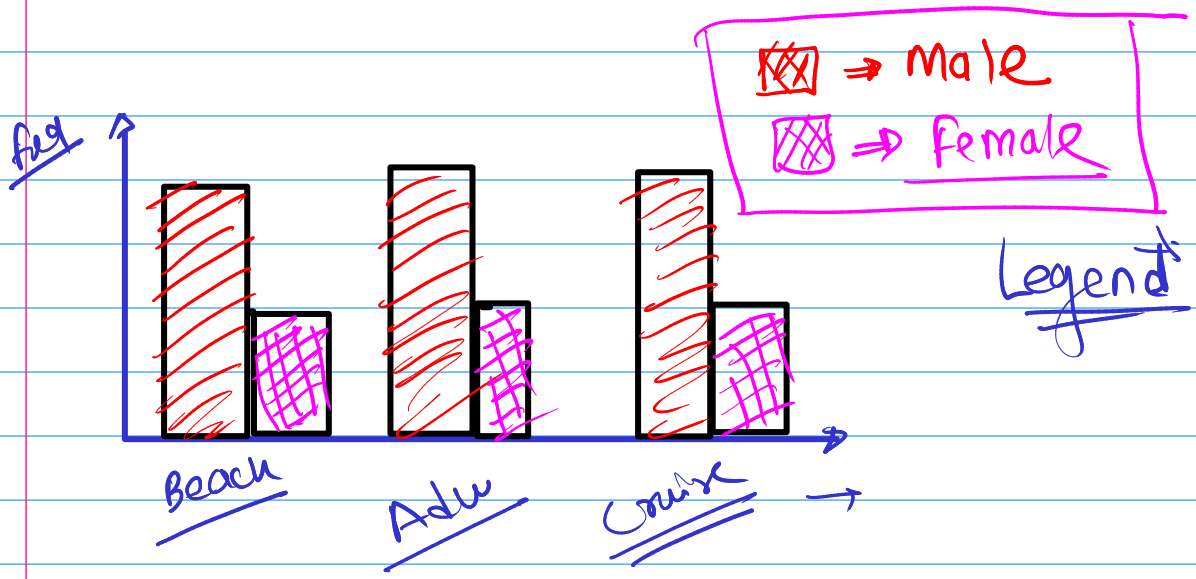


3. Multivariate Analysis

A. Heatmap → Correlation of Numerical Data only



2. BAR / PIE / Histogram → Hue Parameter



x = Vacation, y = count, hue = Gender

Box plot

Outlier finding