DAV - II

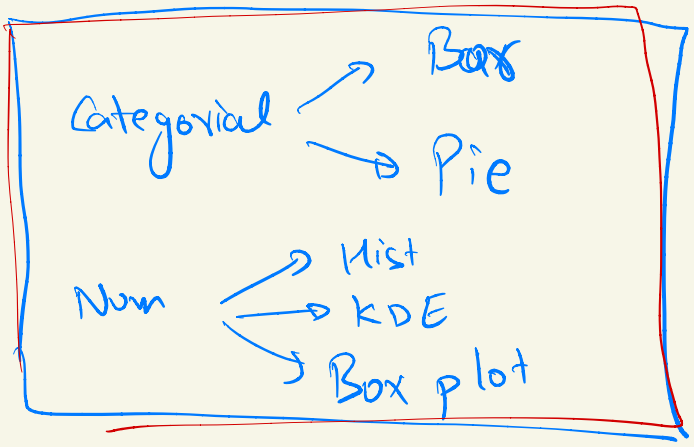
Start at 9:03pm

=> Agenda ->
- Revision
- Bivariate Analysis
- Subplots

=> Revision ->
- Univate -> 1 variable
- Bivariate ->
- multivariate ->

=> Univariate Data viz

1 variable



⇒ Bivariate → 2 variable

→ Num - Num → Cat - Cat → Cat - Num

⇒ Num - Num / Continuous - Continuous

Q what question can we ask?

Sales, year ↓ ↓

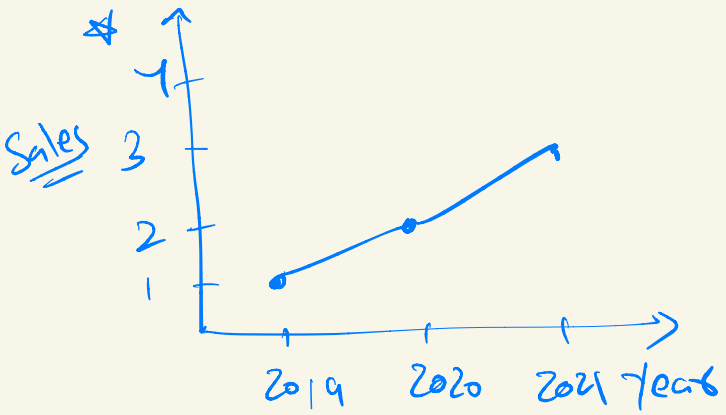
i) How does sales vary over years

ii) How are features associate
↳ Correlation →
→ Year ↑ ~ ↑ ↓

iii)

(i) → Sales vs Year →

Q what chart should we use ??
↳ line



① Dataviz →

Date / Time ⇒ go to x axis

Num-Num → line
→ Scatter

⇒ Cat - Cat →

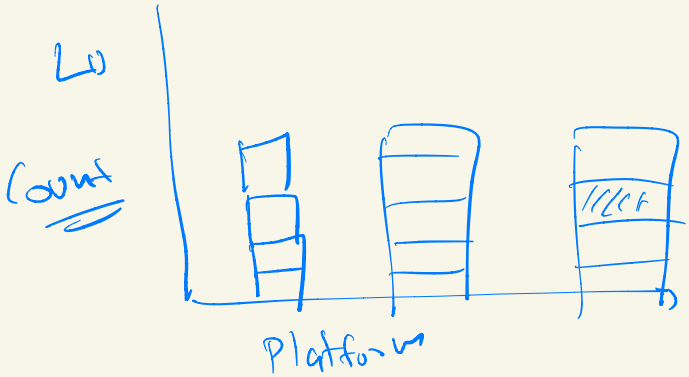
⇒ • Publisher, Platform

Q. what types of questions can be asked?

→ preferred platfor for publisher.

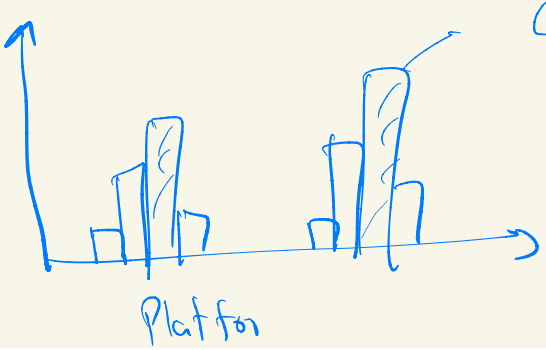
→ distribution of publisher for top 3 platforms

⇒ Distribution (count) of one wrt other category.



Color → publi"

Stacked bar



Color → publishe

Dodged bar

⇒ Sunburst →

⇒ Stacked bar →

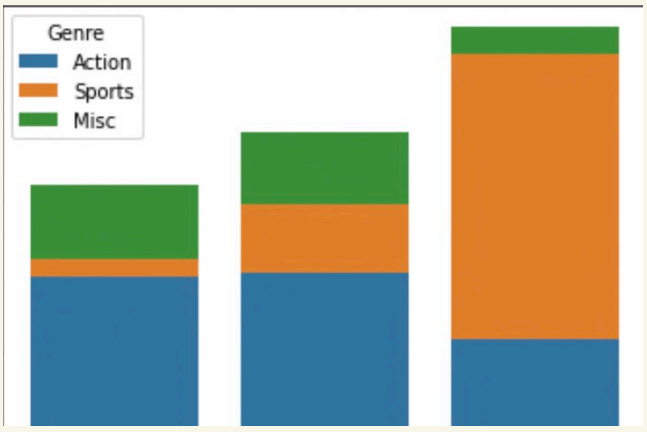
Stacked

→ it is not readable

↳ If you don't label for data

dodge

→ for comparision



→ Helps only with total

⇒ Total is more important

2 Cat - Cat

→ Dodge Bar chart

→ Stacked Bar

⇒ Cat - Num →

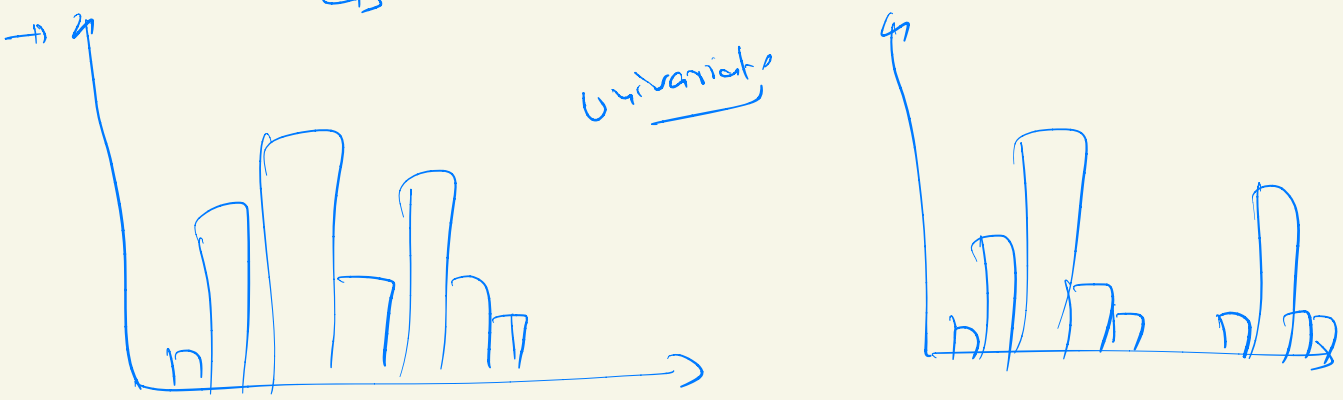
* Publisher vs Sales →

Q → what question can be asked??

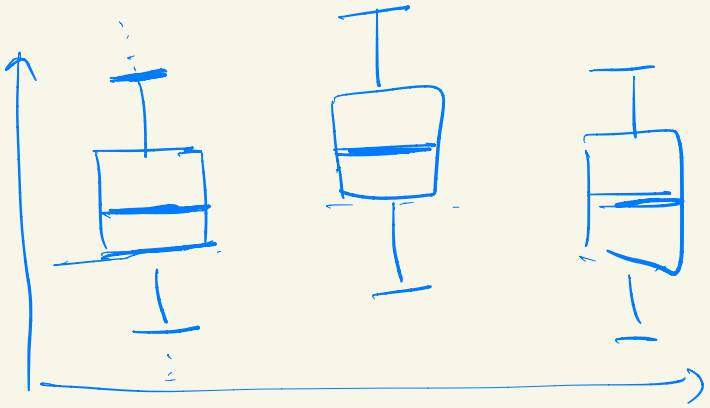
1) what is avg/sum sales for every publisher
2) Sales distribution for top 3 publishers.

⇒ Distribution of a numerical value.

↳



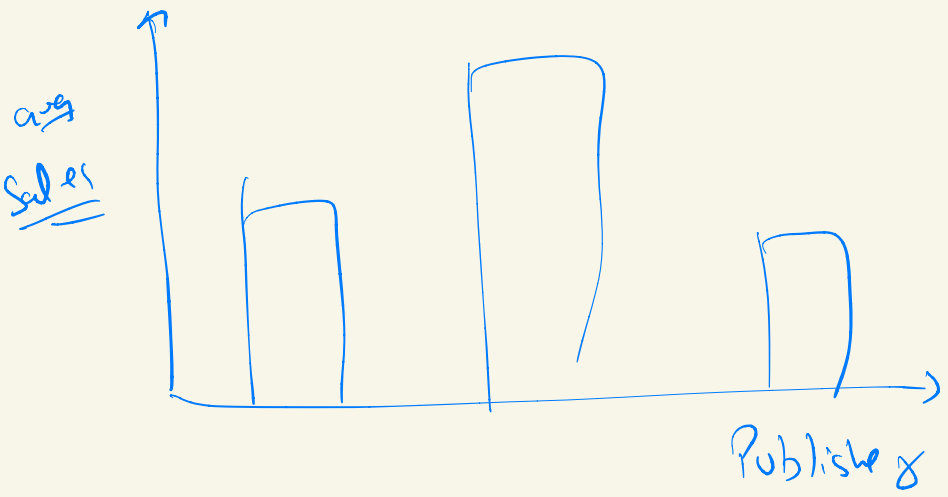
⇒ Box whiskers -



⇒ Q Can we use bar plot as well for num – Cat variables.

⇒ Sum of Sales | avg sales | median Sales

bar chart



Doubts

→ Salary of DS →

→ 30, 35, 25, 28, 40, 42, 38, 39, 300

→ avg → 31.5 ~ → 72

median → $\frac{35+38}{2} = 36.5$ → 38

25, 28, 30, 35, 38, 39, 40, 42, 300

percentile → % value lower or equal to the given val.

25 → $\frac{1}{9} \times 100$

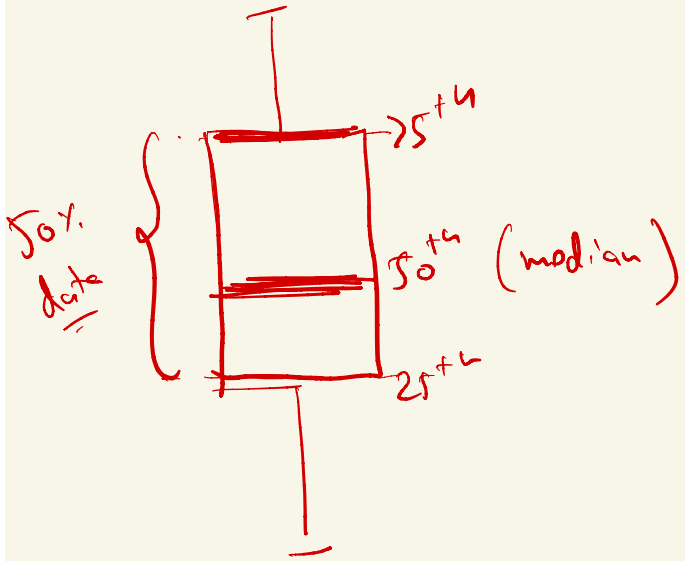
28 → $\frac{2}{9} \times 100$

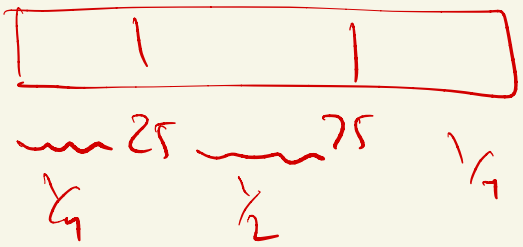
30 → $\frac{3}{9} \times 100$

. . . . 300 → 100%

=> • Distribution viz →

- lower value
- median
- outlier
- upper value.
- 75th percentile 25 percentile.



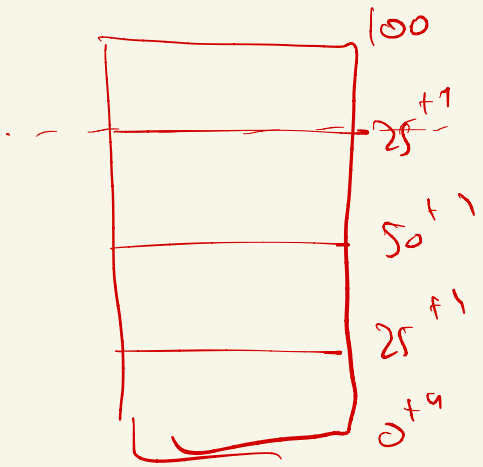


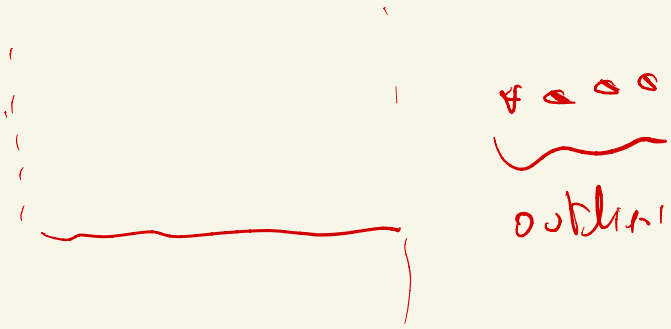
outliers $25^{th} + 1.5\ IQR$

$\rightarrow$.

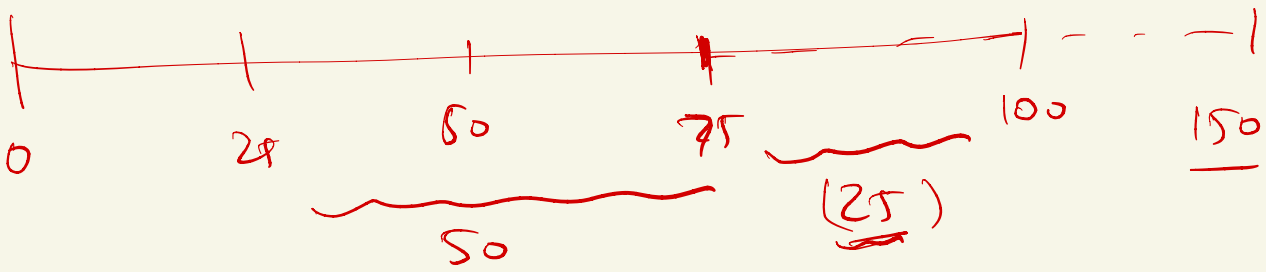
$\Rightarrow$ $25^{th} + 1.5\left(25^{th} - 25^{th}\right)$

anything above this is an outlier





=)



$75 + (50 \times 1.5)$

75