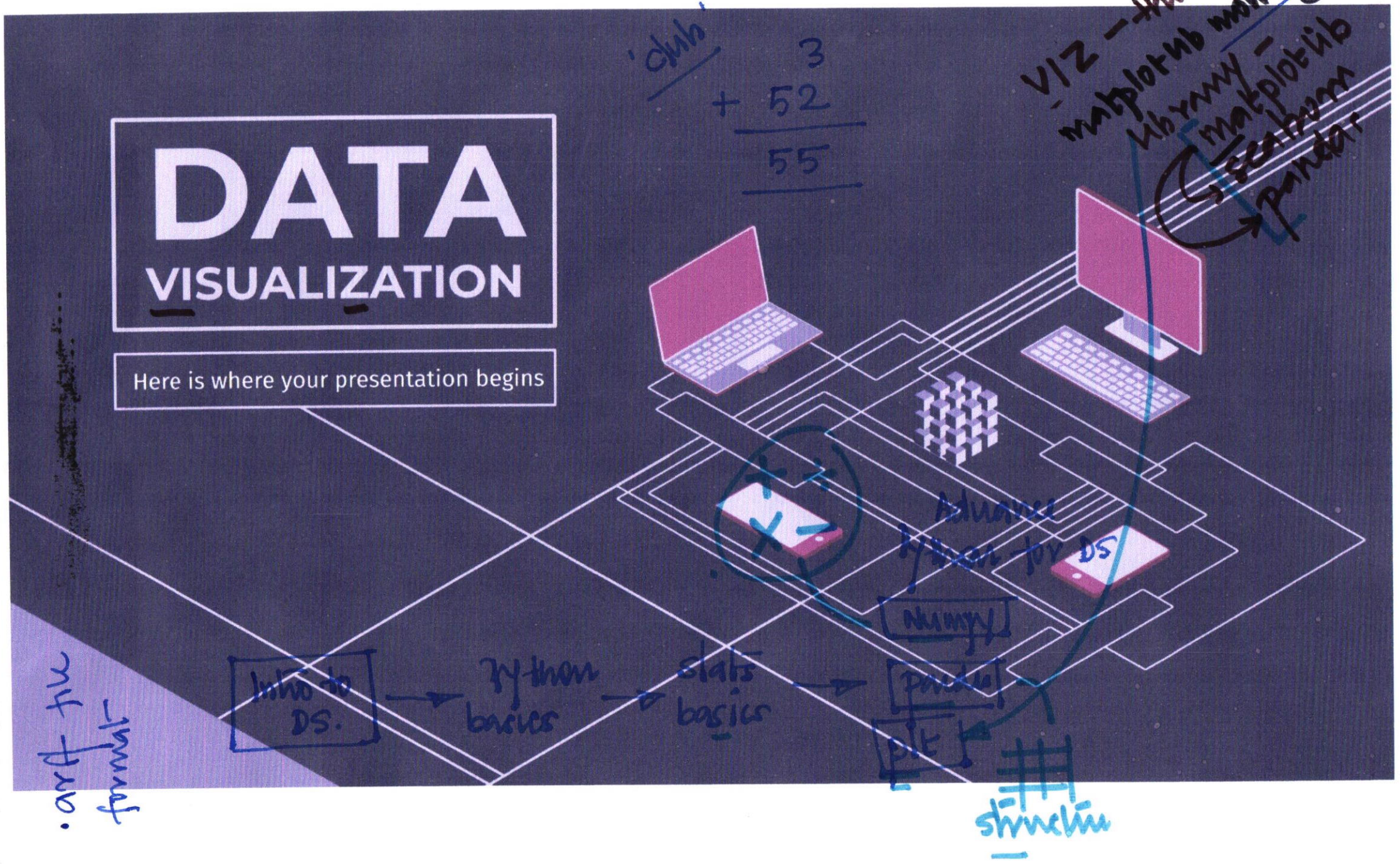

DATA
VISUALIZATION
Here is where your presentation begins

Term Context

- **Introduction to Data Visualization** ← You are here...
- Data Visualization using Matplotlib ← matplotlib monologue
- Hands on Pandas for Rapid Visualization
- Seaborn for Data Visualization

INSAID

Agenda

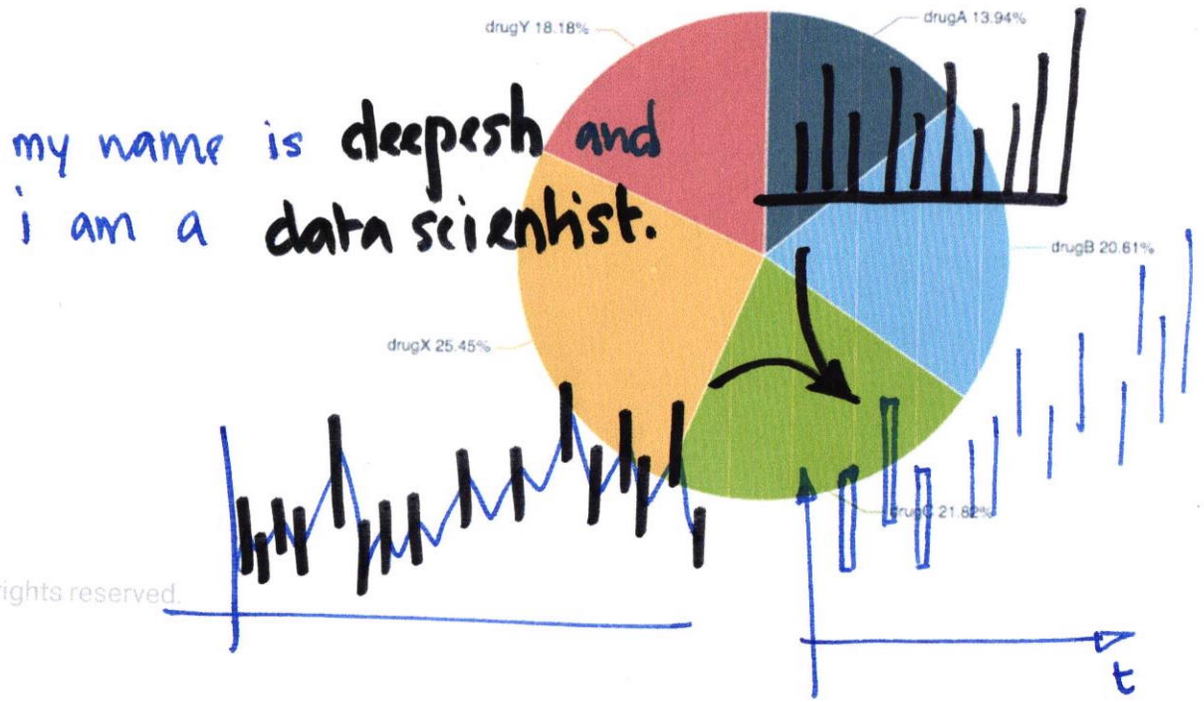
1. Data Visualization
2. Need of Data Visualization
3. Importance of Data Visualization
4. Types of Visualization
5. Univariate Plots
6. Bivariate Plots
7. Multivariate Plots
8. Advantages of Visualization

Data Visualization

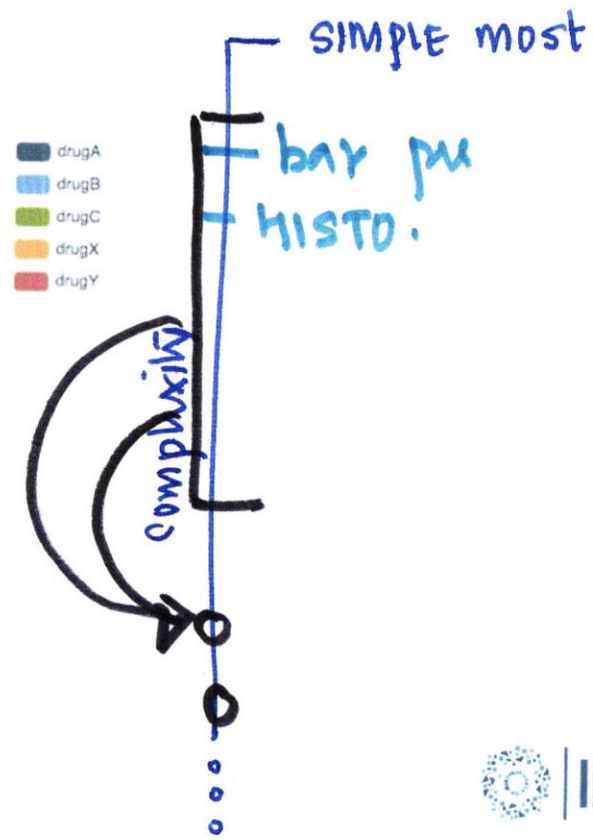
colours

storytelling

- It is the practice of **translating information** into a **visual context**, such as a map or graph.
- To make data **easier** for the human brain to **understand** and pull **insights** from.
- **Main Goal:** Identify patterns, trends and outliers in large data sets easily.



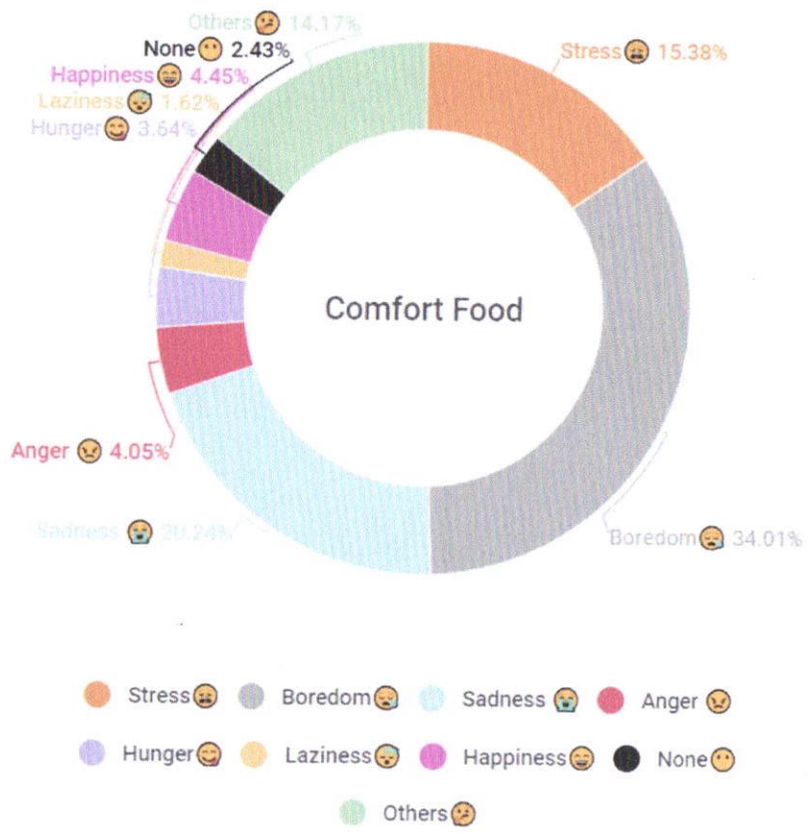
my name is deepesh and
i am a data scientist.



INSAID

Need of Data Visualization

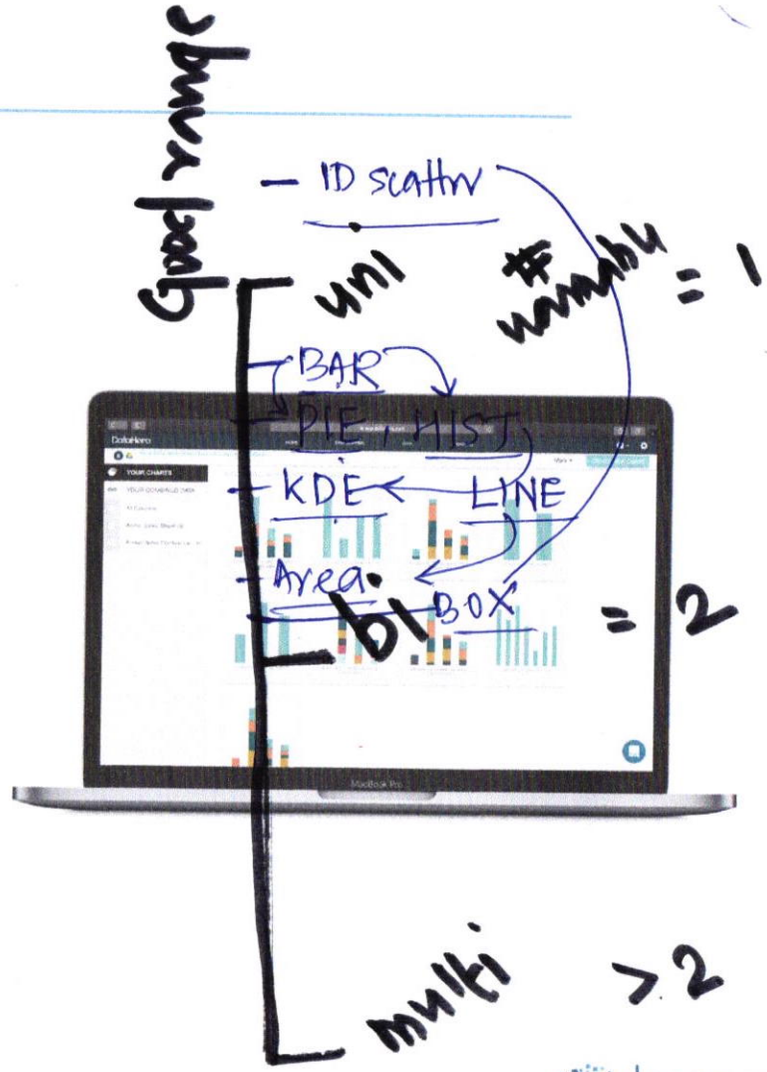
- Helps humans to **inspect** and understand better.
- Provides better **visual input** and **faster to process**.
- Easier to **find** patterns and **digest** the data.
- Develop **insights** to solve the problems.
- **For example:** College student's Reason for eating comfort food.



INSAID

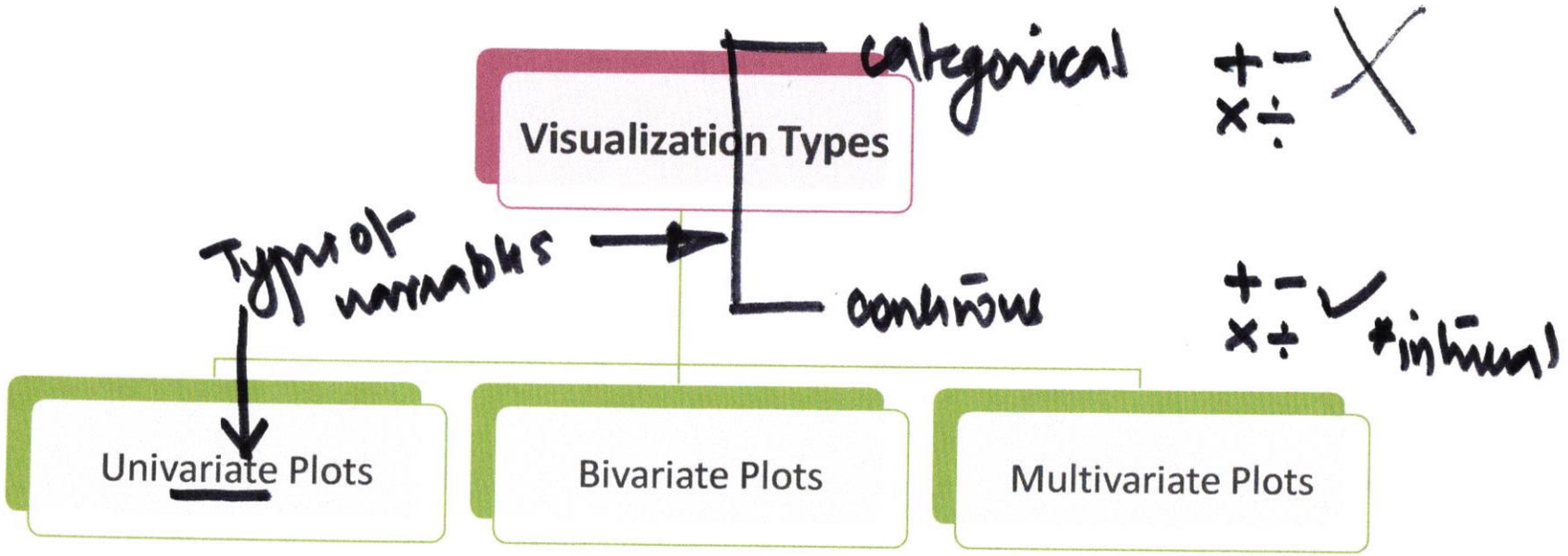
Importance of Data Visualization

- Analyze the data in a better way.
- Faster decision making.
- Making sense of complicated data.



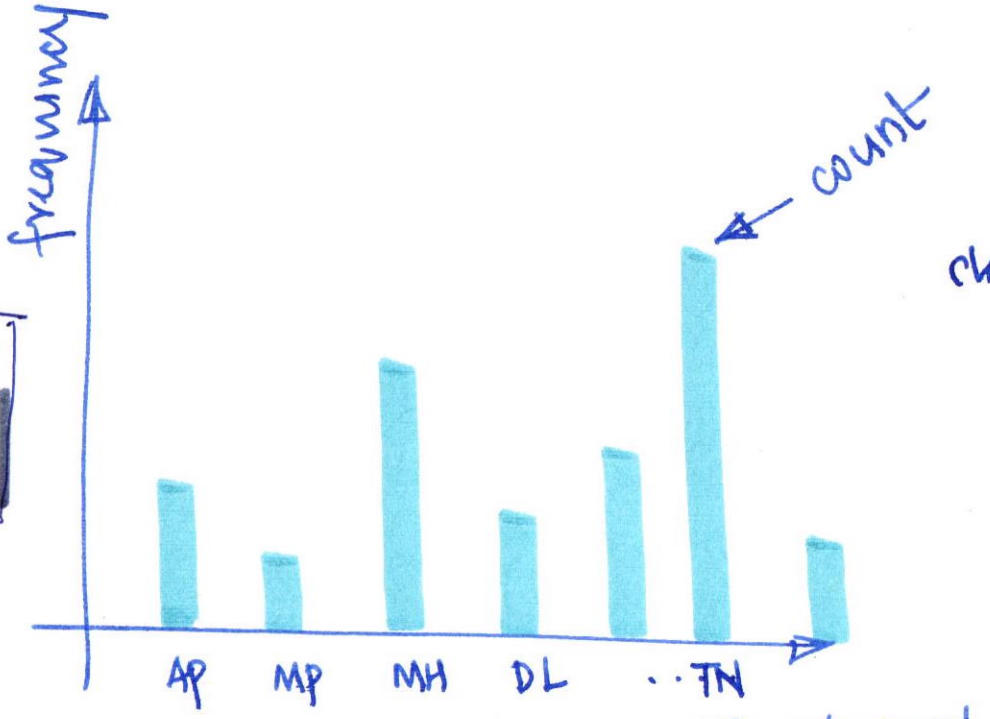
INSAID

Types of Visualization



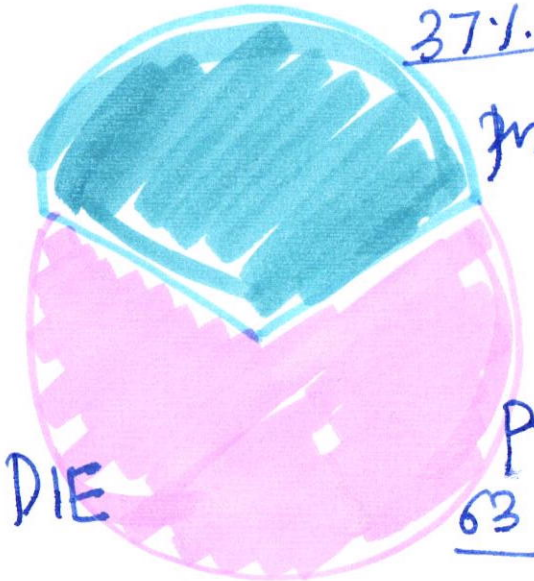
UNI variati
single.

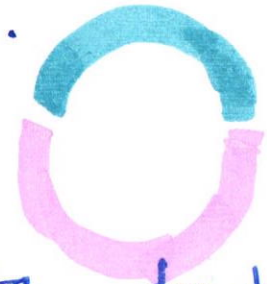
BAR



The categorical ~~COLUMN~~.
variable

PIE

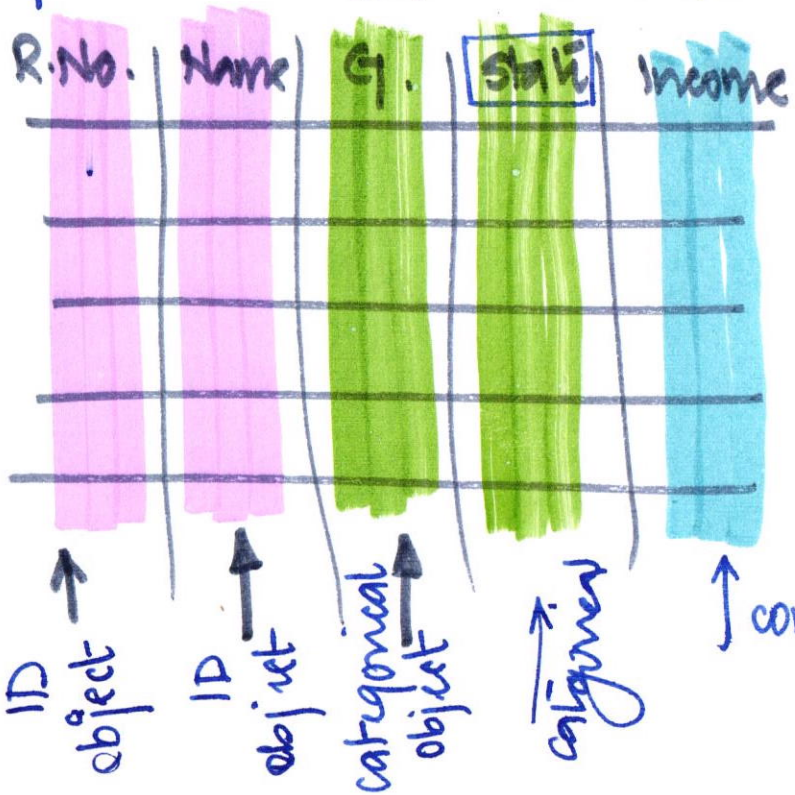




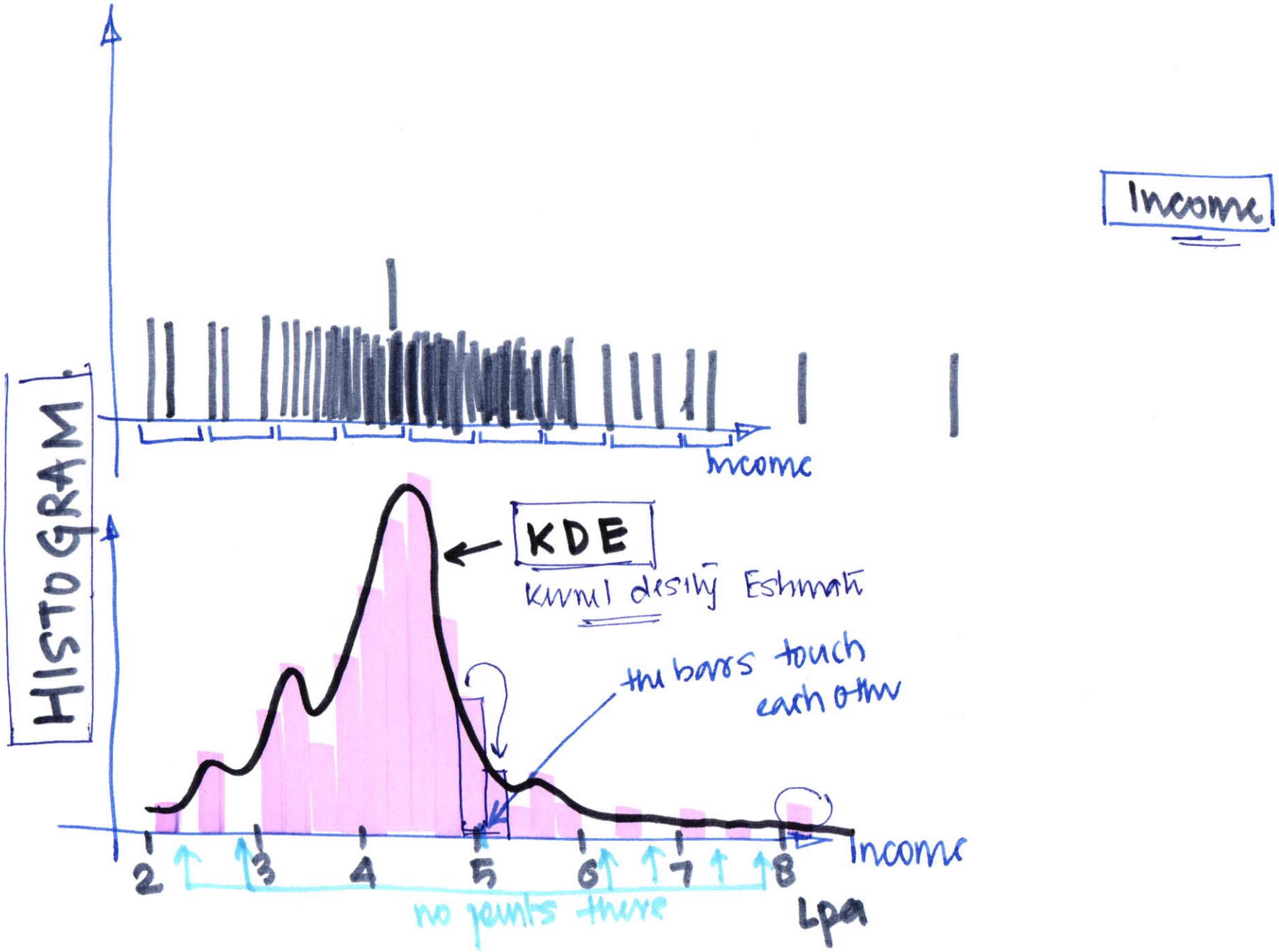
pasta donut

class_df

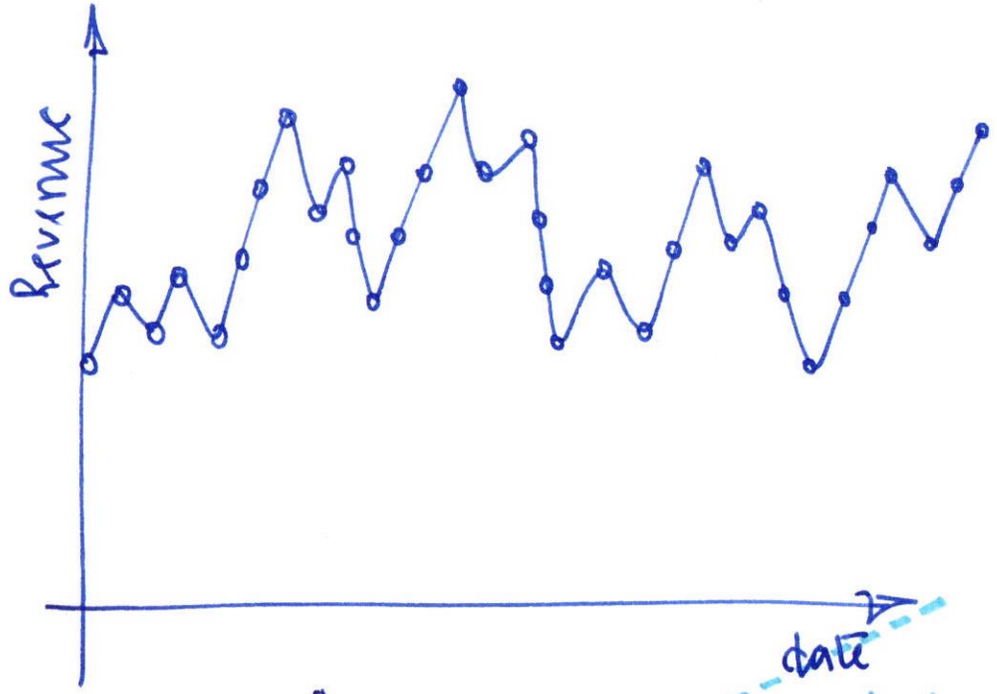
class students




```
class_df['state'].value_counts()
AP  —
MP  —
DL  —
MH  —
```


Income
HISTOGRAM
Income
KDE
Kernel density Estimate
the bars touch each other
2
3
4
5
6
7
8
Income
no points there
Lpa

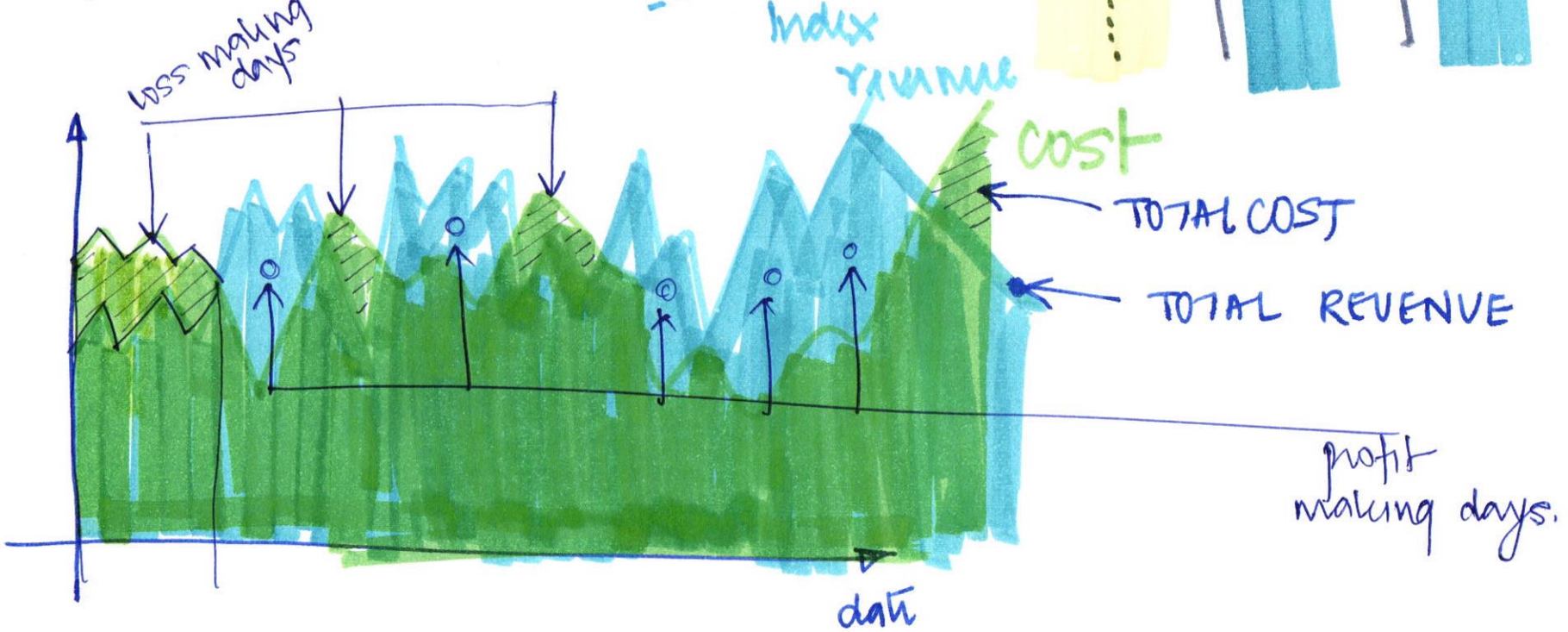
LINE

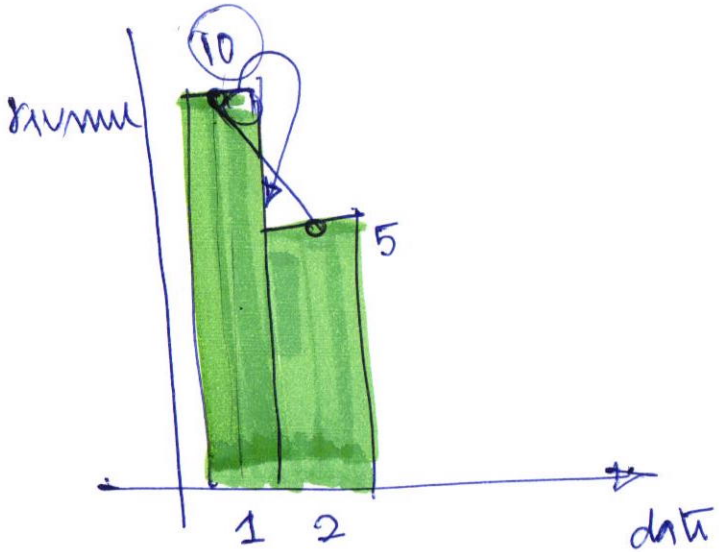


Big bazar.
dataset

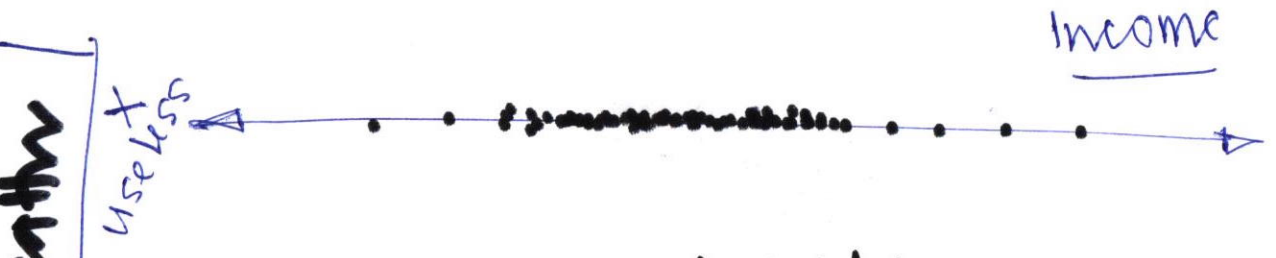
	date	Revenue	Total cost
01	JAN 21		
02	JAN 21		
03	JAN 21		
	⋮		

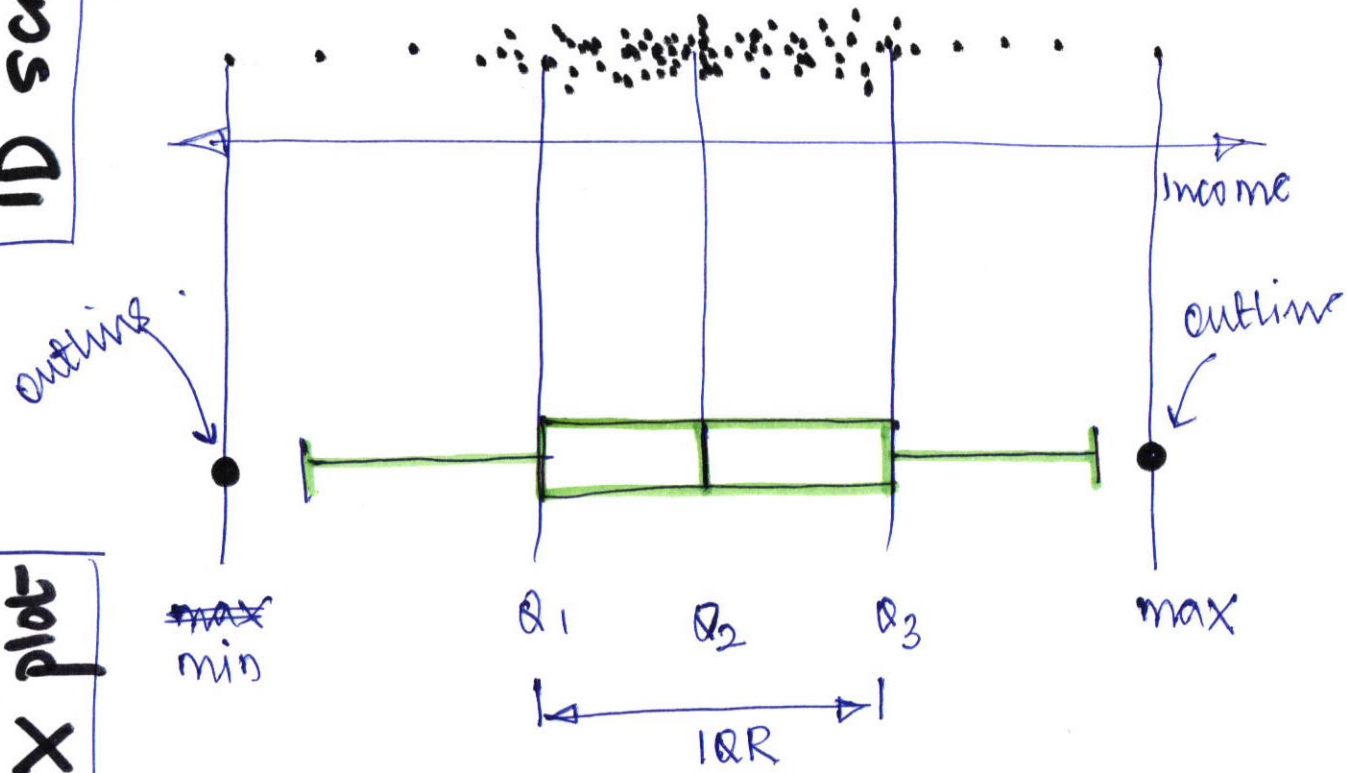
AREA





1D scatter

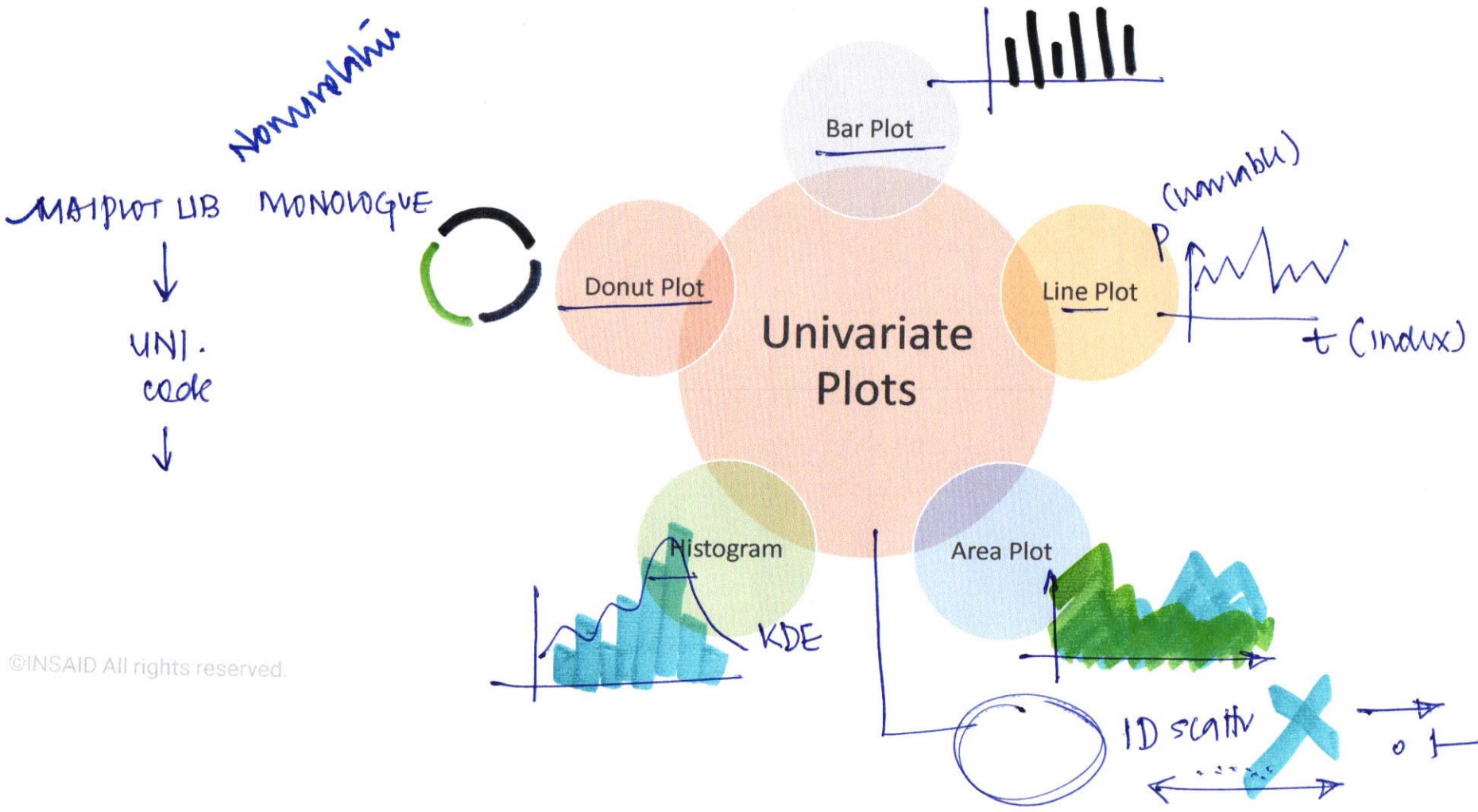




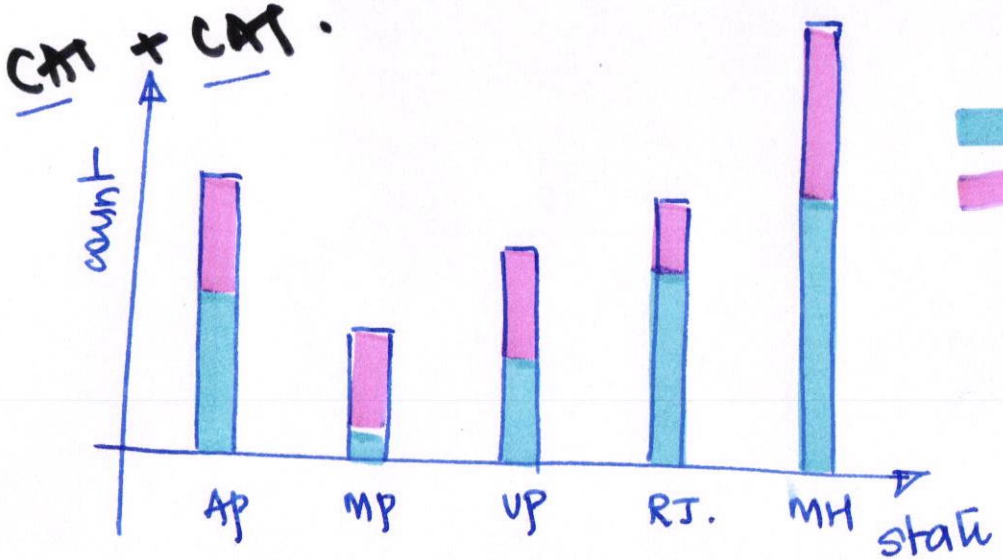
uni variate BOX plot

Univariate Plots

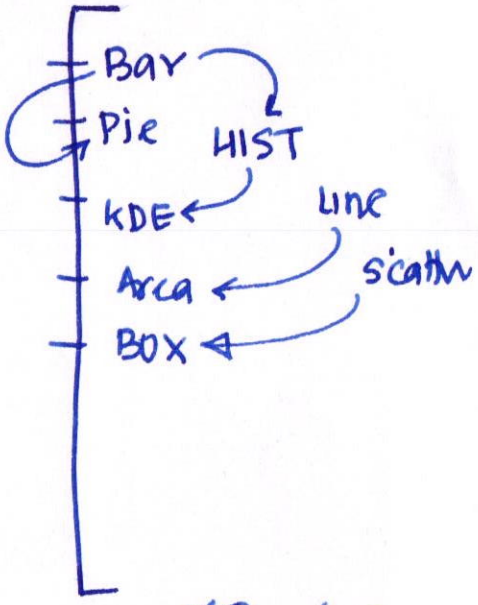
- Describe a type of data which consists of observations on only a single characteristic or attribute.

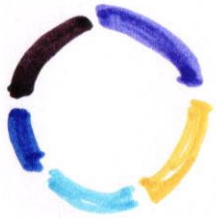


CAT + CAT.



stacked BAR CHART.





TOTAL

% M

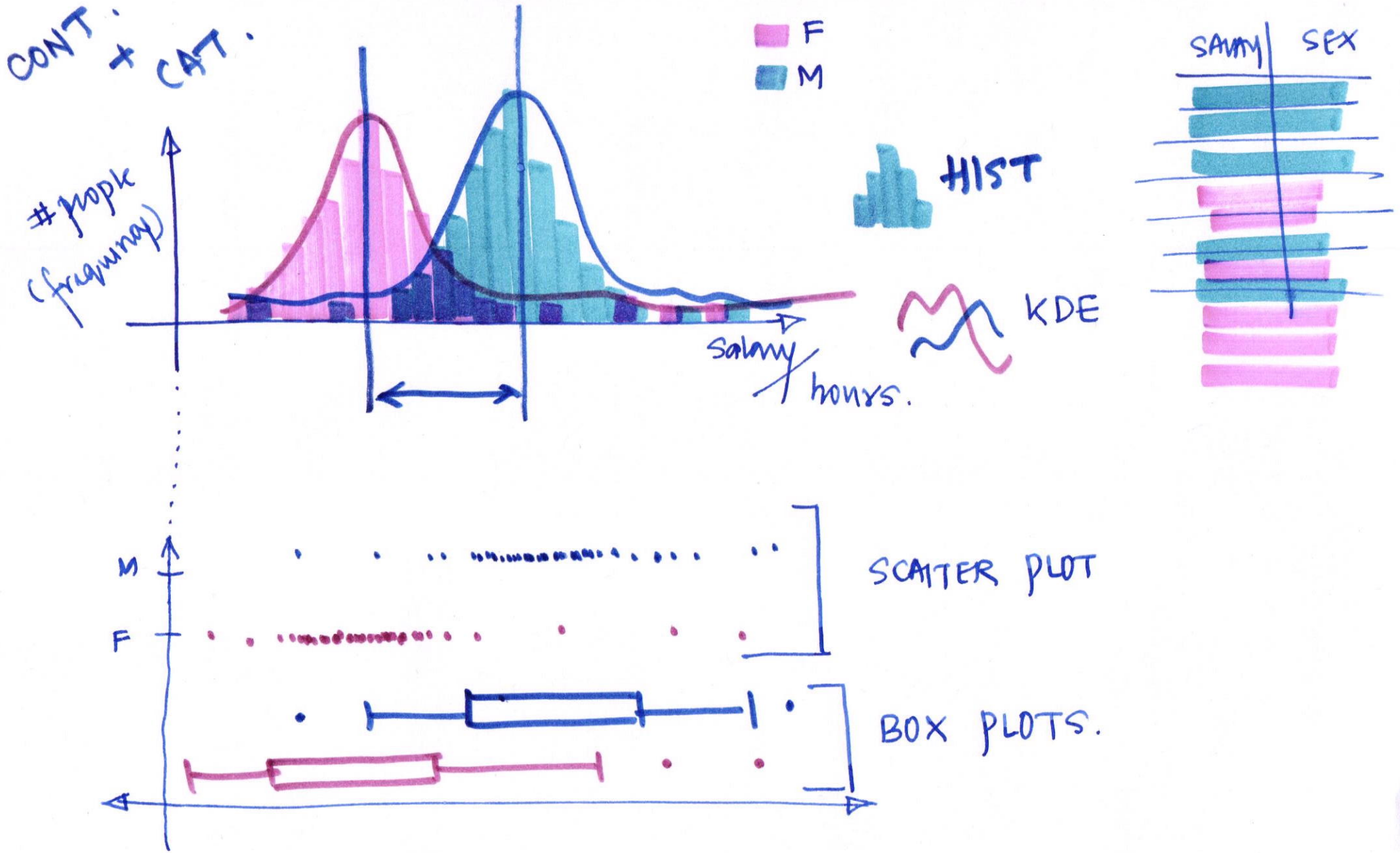
% F

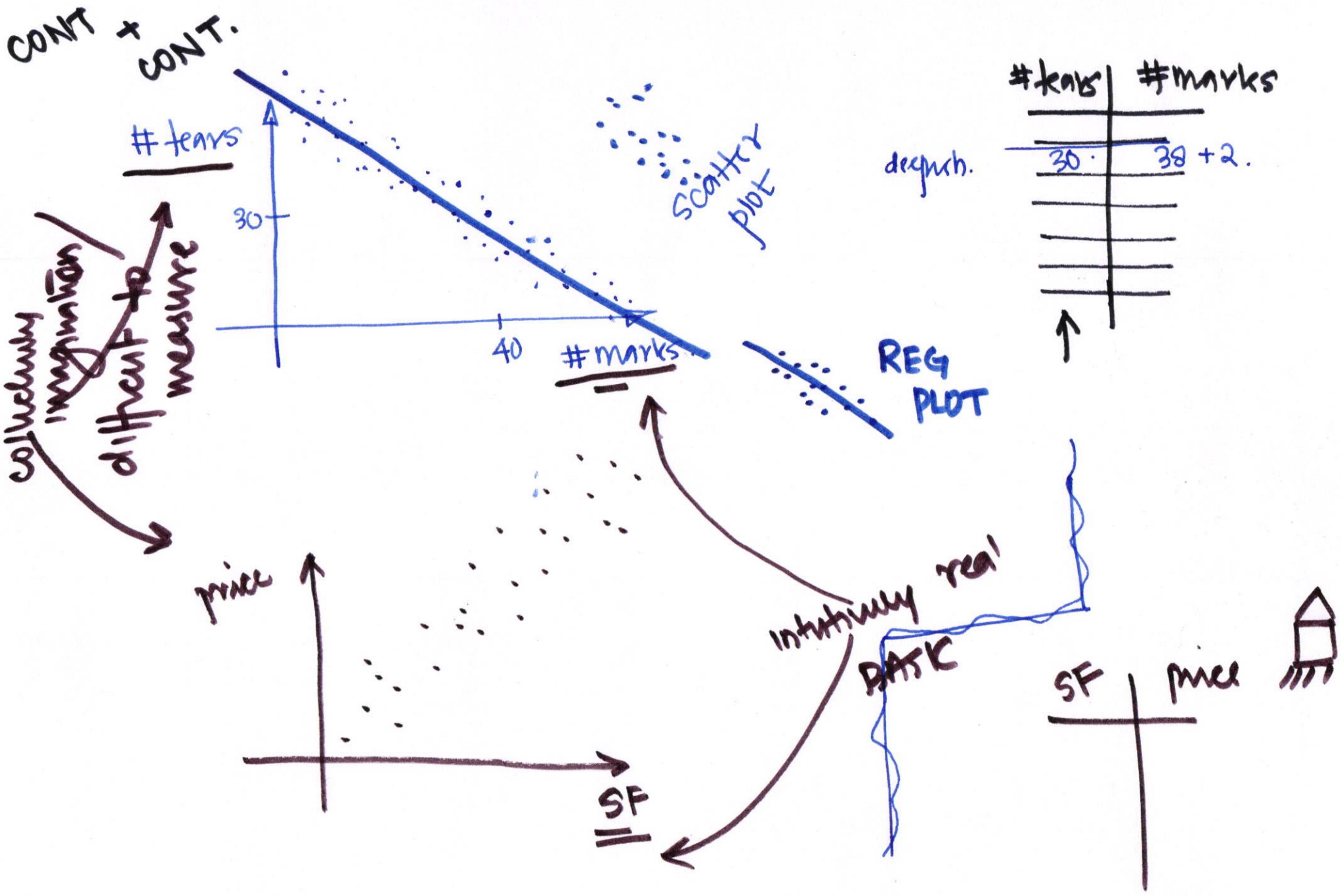
X_1	X_2
CAT	CAT
CAT	CONT.
CONT	CAT.
CONT.	CONT.

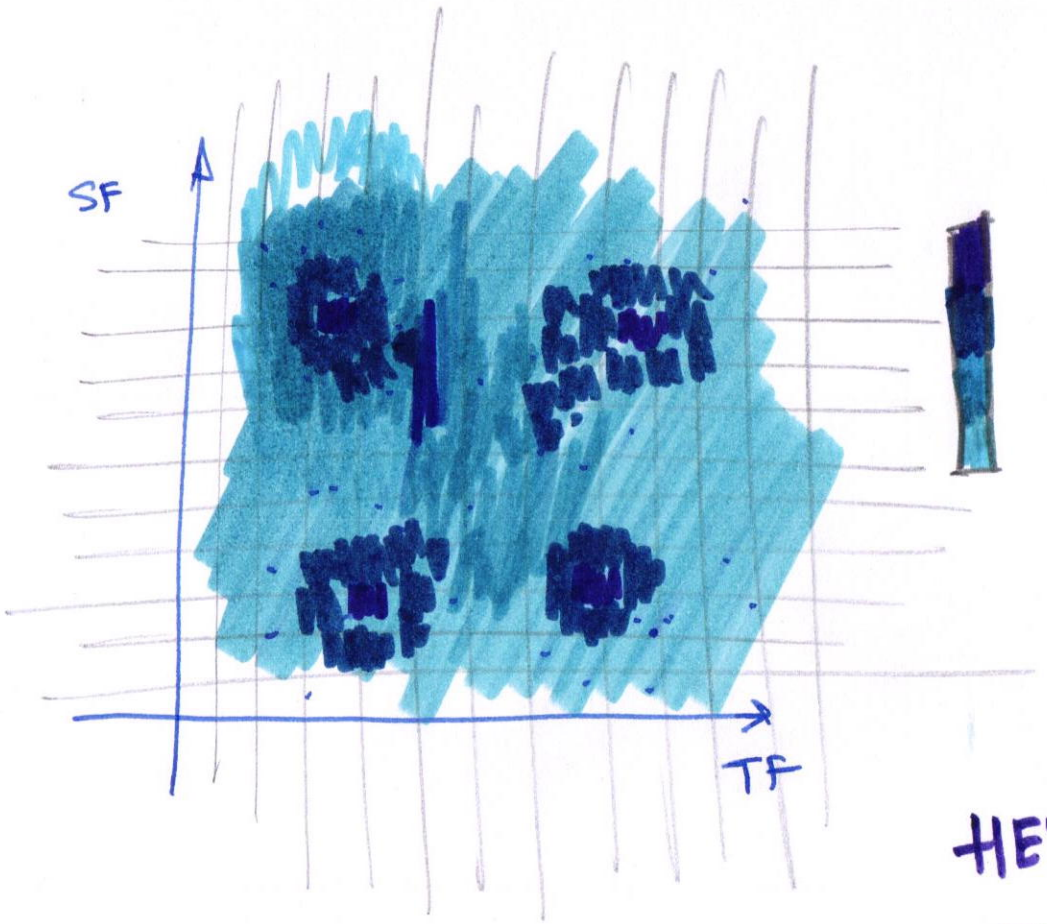
identities] × Random Variable
categorical
continuous
date-time] × Random variable

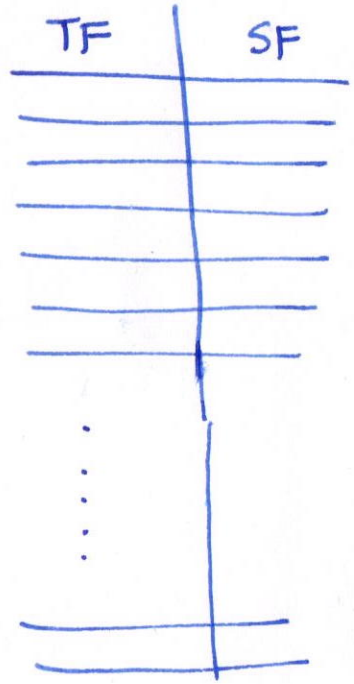
SEX	STATE
M	
M	
F	
F	

```
df[df['SEX'] == 'M'].plot.pie()
```

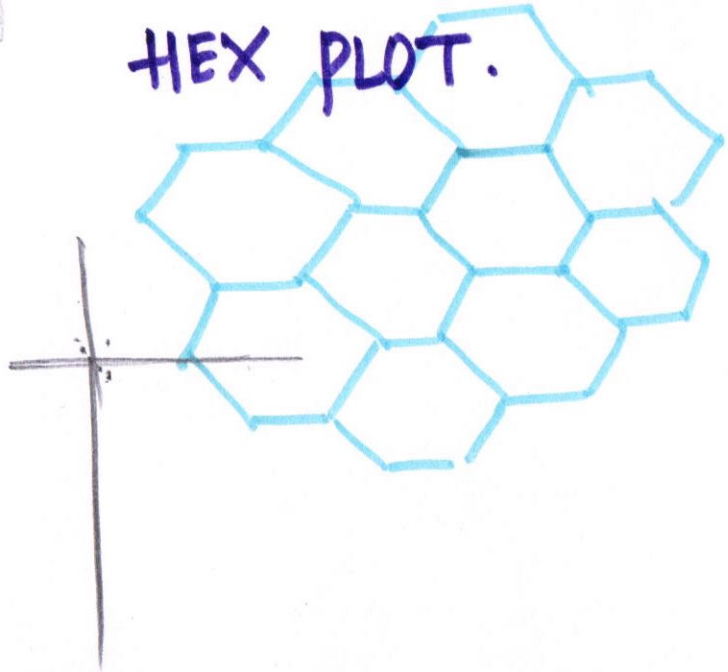

CONT. + CAT.
F
M
people (frequency)
Salary
hours.
HIST
KDE
SALARY
SEX
M
F
SCATTER PLOT
BOX PLOTS.

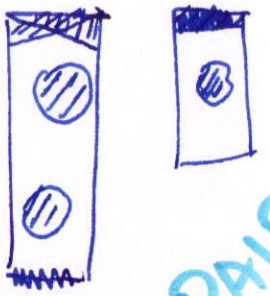

CONT + CONT.
#tears
30
40
#marks
scatty plot
REG PLOT
#tears
#marks
depth.
30
38 +2.
collectively
imagination
difficult to
measure
price
SF
intuitively real
DATA
SF
price





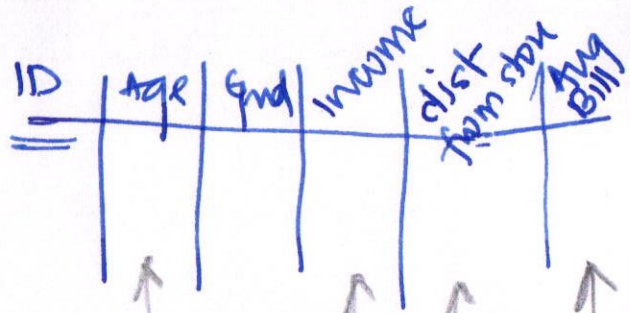
HEX PLOT.



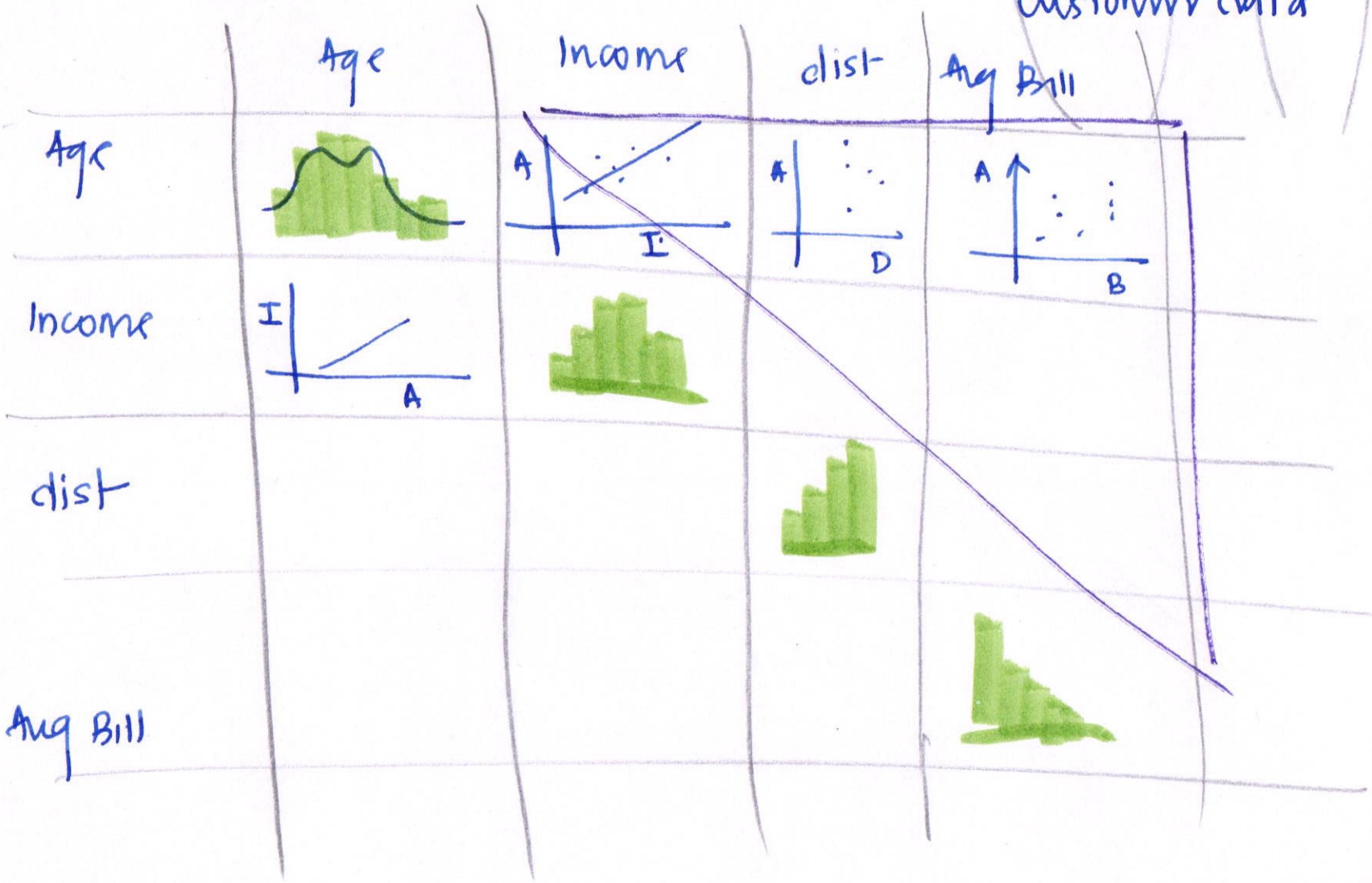


PAIR PLOT

$$nC_2 = \frac{n^2 - n}{2}$$

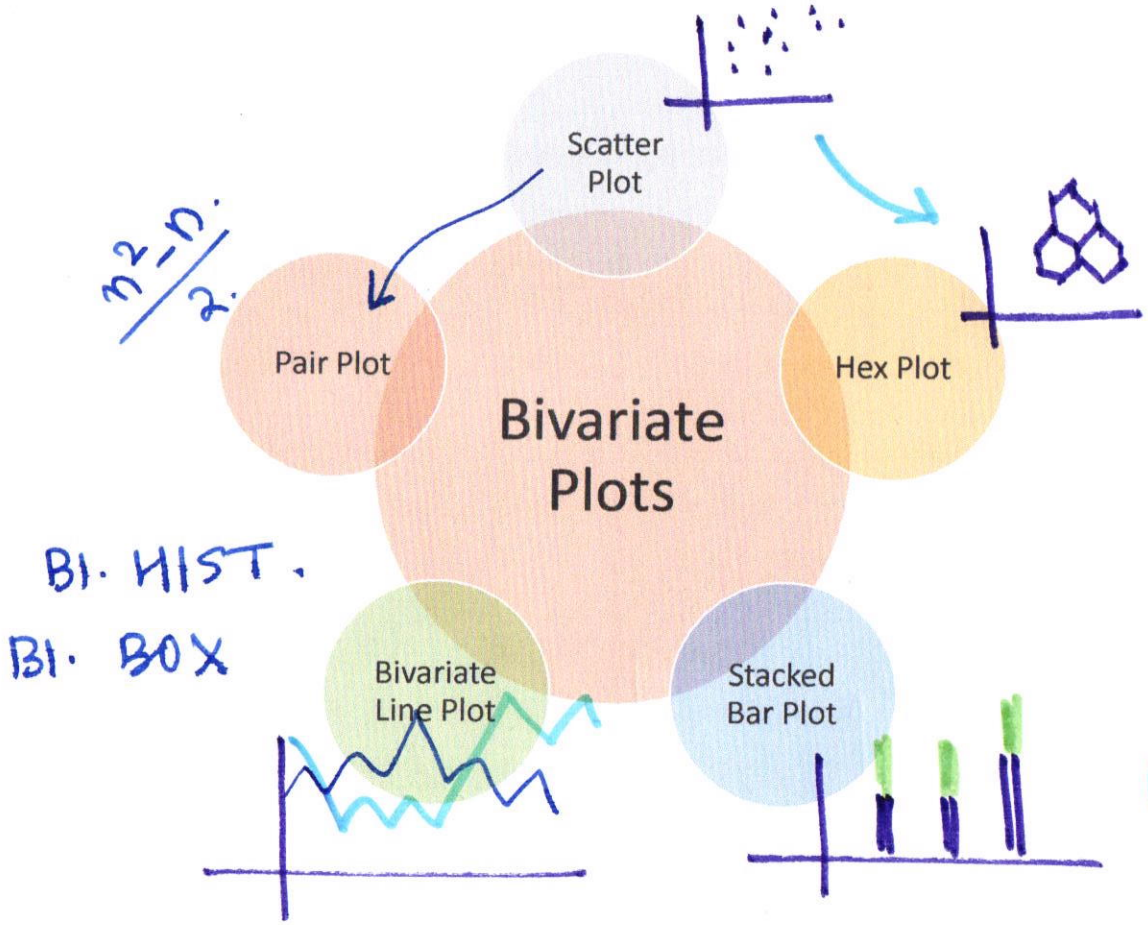


customer data

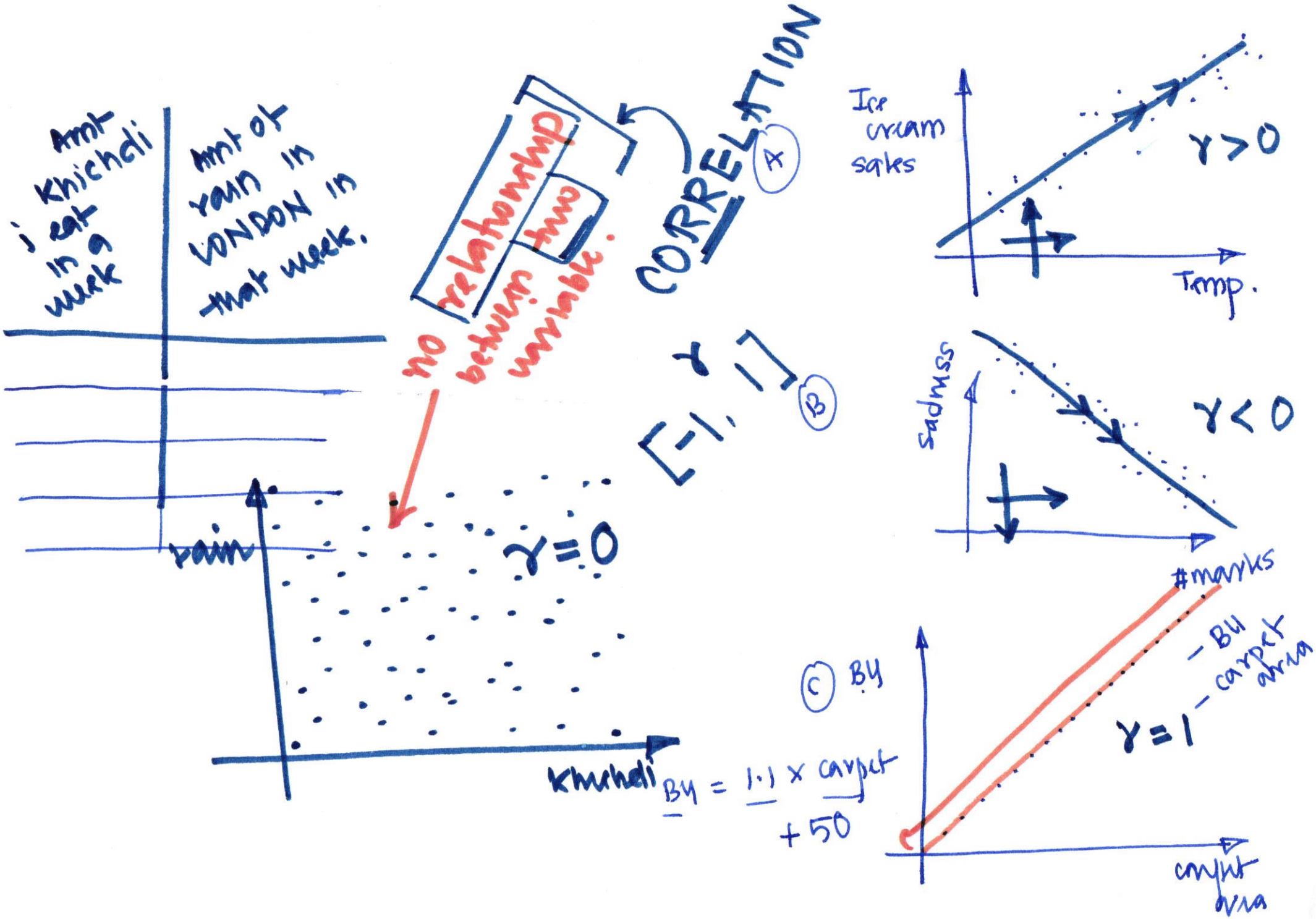


Bivariate Plots

- Describe a type of data which consists of observations on two characteristics or attributes.



INSAID


Amt Khichdi I eat in a week
Amt of rain in LONDON in that week.
no relationship between two variable.
CORRELATION
A
B
γ
$[-1, 1]$
rain
khichdi
$\gamma = 0$
Ice cream sales
Temp.
$\gamma > 0$
Sadness
$\gamma < 0$
#marks
carpet area
$\gamma = 1$
C
BU
$BU = 1.1 \times \text{carpet} + 50$

$$\text{correlation} = \frac{\sum (x-\mu_x)(y-\mu_y)}{\sqrt{\sum (x-\mu_x)^2} \cdot \sqrt{\sum (y-\mu_y)^2}}$$

	Age	Income	dist	Avg Bill
Age	1	.6	0.1	.3
Income	.6	1		
dist	0.1		1	
Avg Bill	.3			1

correlation matrix

HEAT MAP.

sns.heatmap(df.corr())

(decent).

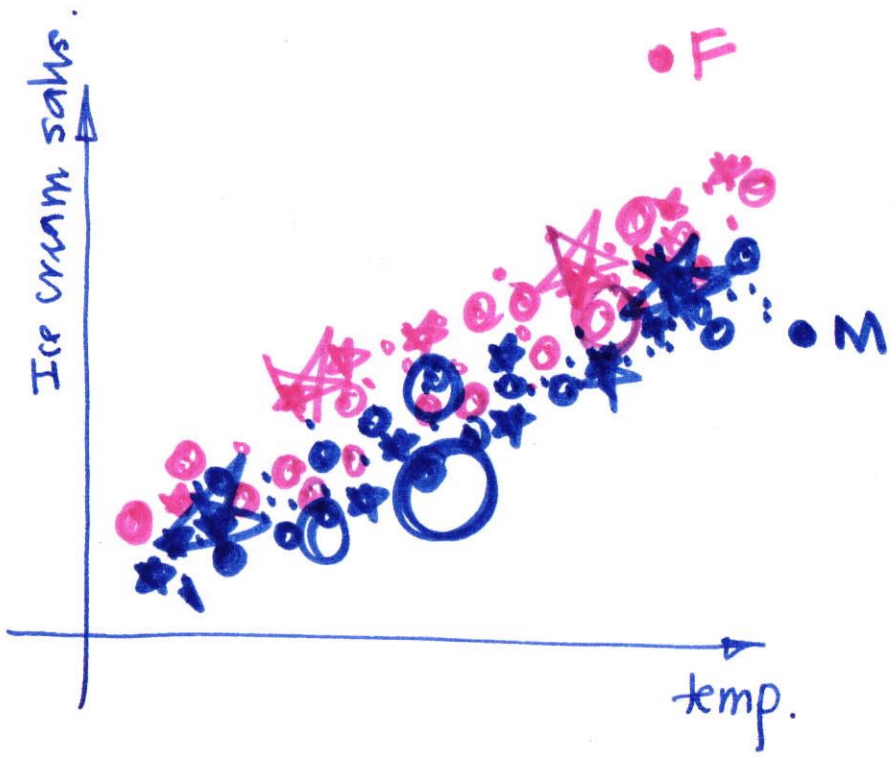
DATA SOURCE:

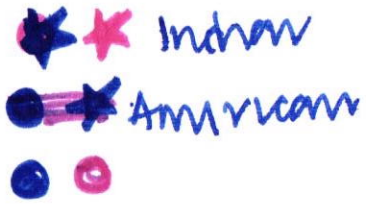
"UCI. Machine Learning Repository."

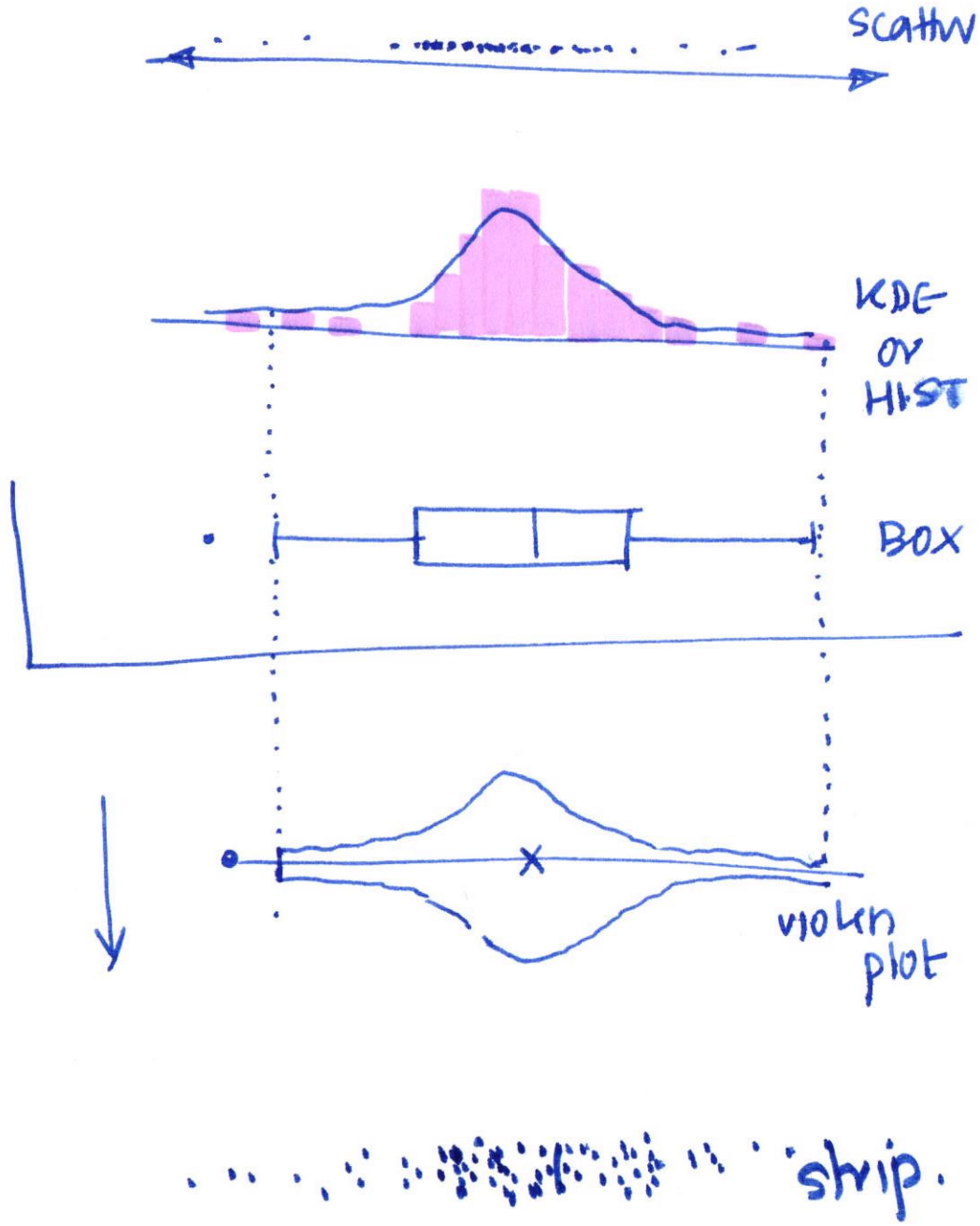
(Google it)

MULTI VARIATE SCATTER PLOT.

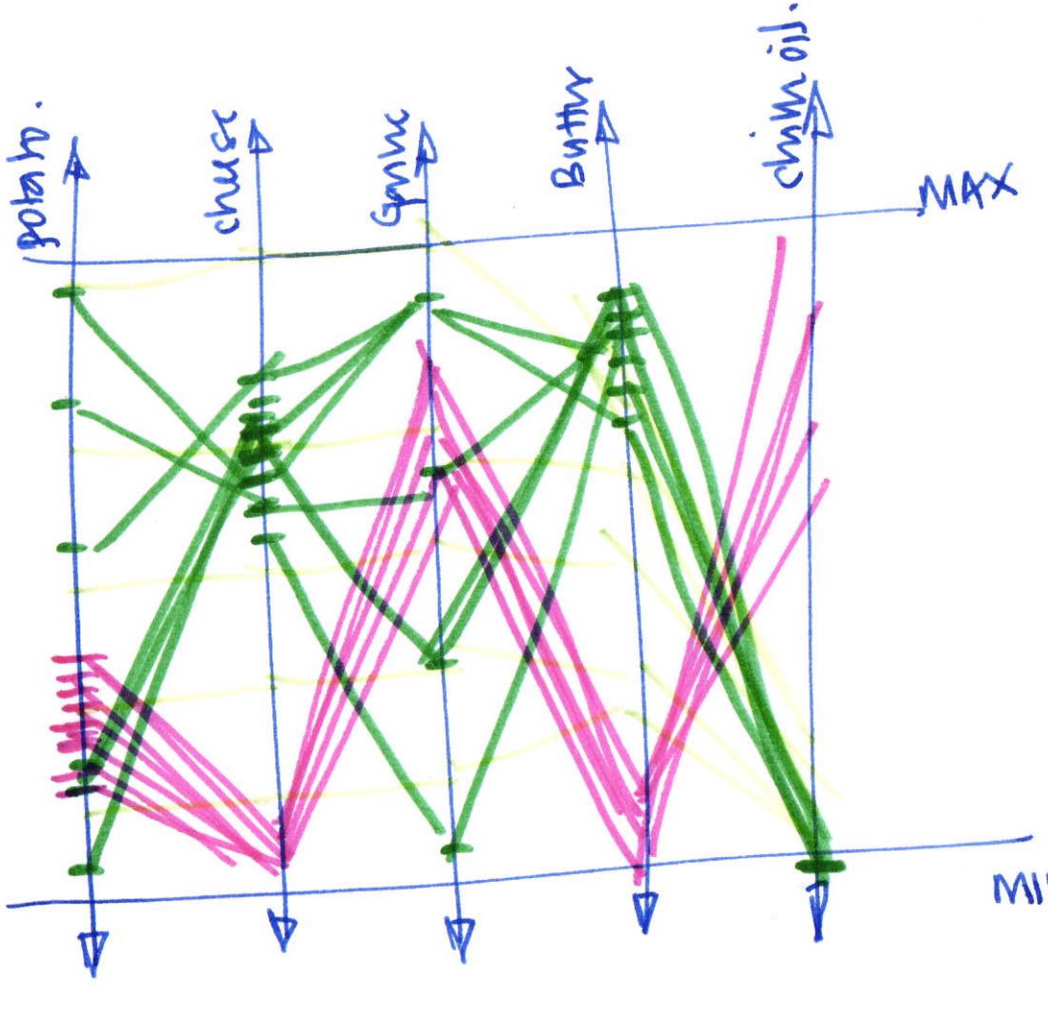
3 variable.







strip.

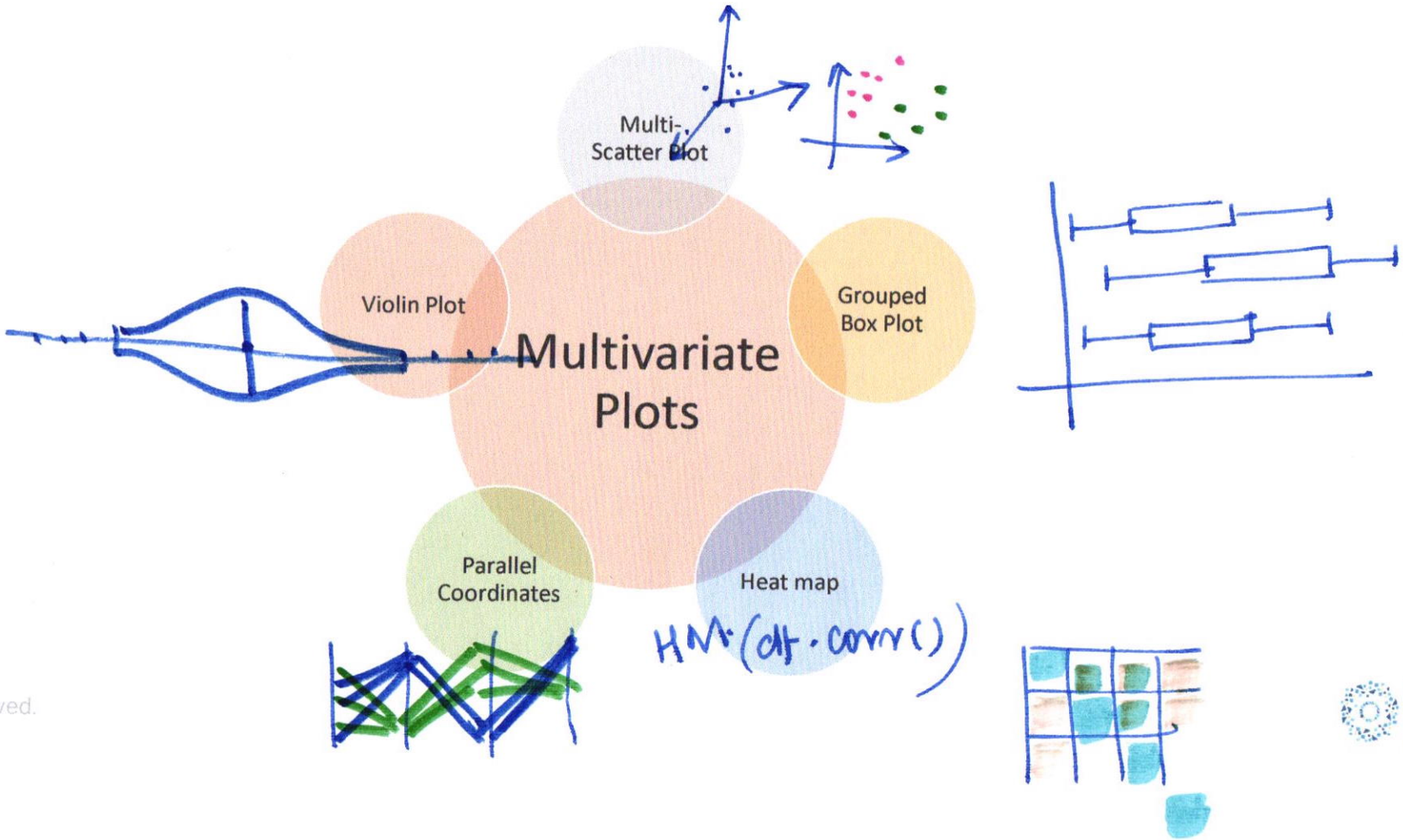


Parallel Coordinate Plot

separability.
'classification'

Multivariate Plots

- Describe a type of data which consists of observations on more than two characteristics or attributes.



INSAID