

## Data visualisation

### i) Interpretation of data

- Identifying patterns and trends in the data

### ii) Relationship

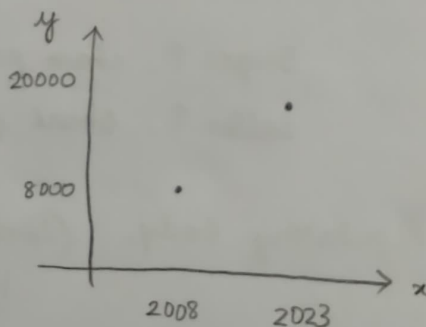
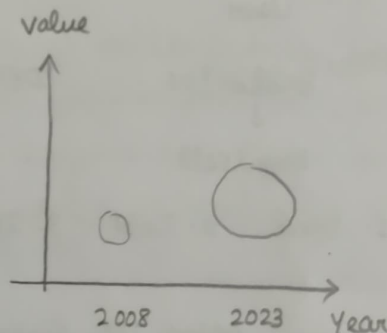
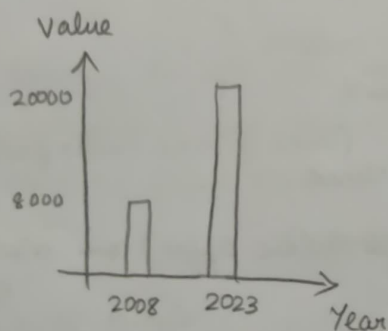
- Analysing the correlation in the data, ...

### iii) Outlier detection

- Using line graphs, ...

All the above are considered as the goals of data visualisation.

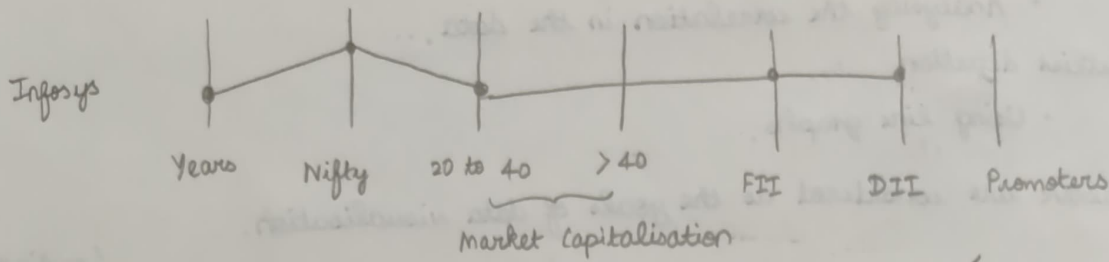
- Data has uncertainty, [Context of data]? — purpose of data visualisation; (audience)
- Data collection
  - i) Primary - the data can be highly trusted upon
  - ii) Secondary - retrieval of data from other authorised sources is preferred.
- Better to know about the 'What', 'When', 'Who' and 'Where' of the data.
- Python, Tableau, Griffy (network) are known data visualisation tools.
- Representation of data
  - Flowchart
  - Line chart
  - Pie chart
- Visual clues  $\rightarrow$  shapes, sizes (length), colour



Scatter plot

- Use color to add third variable.
- Use color and size for representation of 4th var.

- To represent more than 2 variables, use 'heatmap'. Changing the intensity of colours to represent higher values.
- Another alternative to heatmap is 'parallel coordinates' for representation of higher number of variables.



Parallel coordinates tell the nature of correlation (+ve, -ve, weak)

- Types of data → Categorical data

Represented using bar graphs, pie charts.

- Time series data (identification of trends, seasonality, cycles)

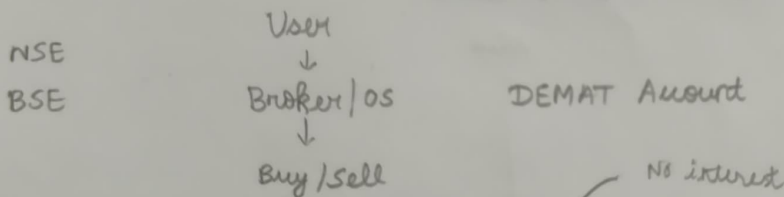
Using line graphs with Time as x-axis.

- Spatial data

Contour map

19/12

## Stock market



Capital: 1. Bank 2. Debt 3. IPO (Initial Public Offer) ← Merchant Banker

After IPO, listed in NSE

Buyer ↑, share price ↑

Seller ↑, share price ↓

SEBI - Regulatory body, (Quarters - documents)

Net Profit - E A I T

Nifty (50 companies)

large capital - moderate returns

15-18 %

→ Small capital → Risk and Returns ↑

→ Honesty, Good management, Promoters, EPS, PE, DER, ROI, ROE, Inventory, Future growth } Things to look before investment

20/12

NBFC < Bajaj  
L&T Finance

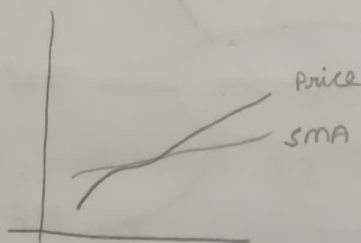
Net Profit → Retained earnings

→ Dividend ratio - 4.1% Infosys

PSU → PFC, REC, ONGC, BPCL } Govt. main stakeholder

Less capital appreciation, more dividends

ANALYSIS / Fundamental - Annual report - Intrinsic Value  
- Balance sheet - Financial ratios  
Technical - charts (MACD, SMA, Bollinger Band, RSI)



Trading / Intraday x, F&O x  
Short-term  
medium-term

LSTM, GRU  
Used

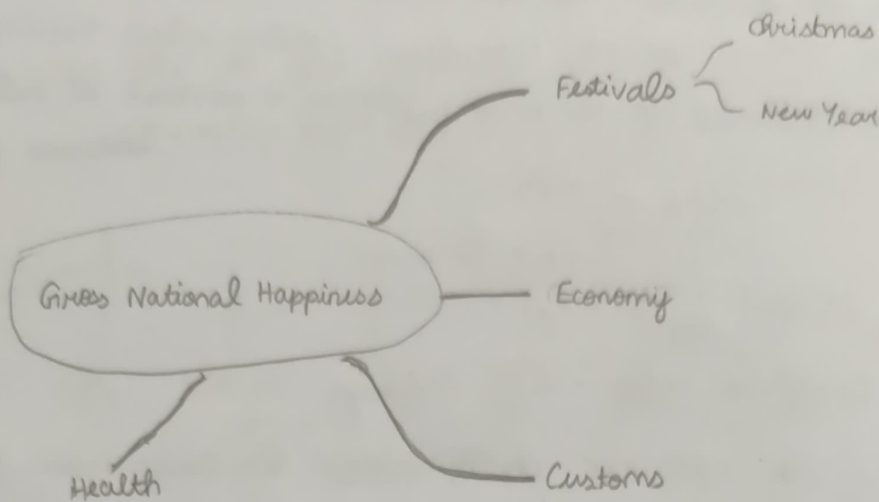
Trading view, com (KNN) -

"Portfolio mgmt"

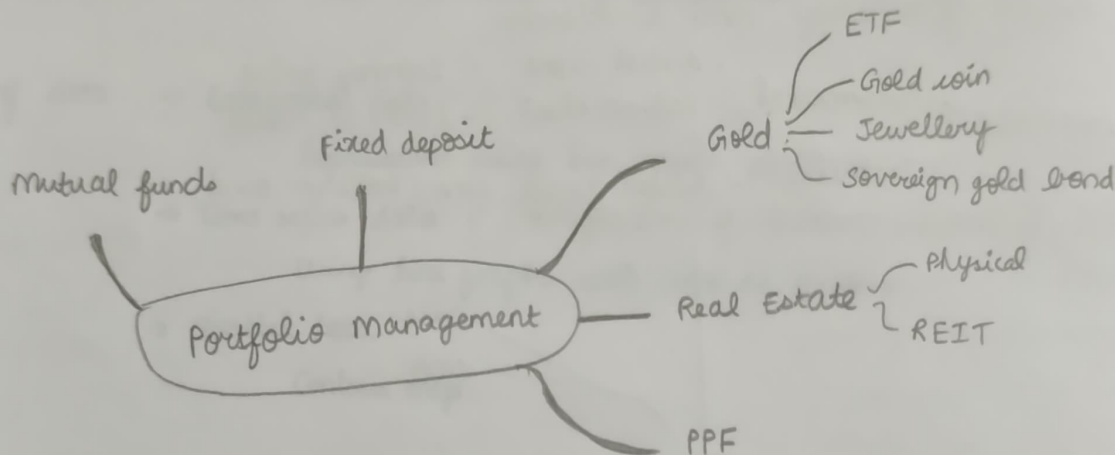
Mutual Funds,

## Mind Mapping

Ex1



Ex2



03/01

There are 4 basic visualisation types that you can use to present your data: i) Comparison, ii) Composition, iii) Distribution and Relationship. Discuss each type in detail and suggest atleast 2 chart types and an application for each visualisation type.

- i) column chart, dot plot // Sales
- ii) Waterfall chart, pie chart // Budget
- iii) Histogram, density plot (pdf) // Covid death rate
- iv) Scatterplot, parallel coordinates, Line chart //

05/01

## Data Cleaning

missing values  $\begin{cases} \text{Ignore and delete} \\ \text{Imputing} \end{cases}$

Categorical variable  
/  
Decision trees

- L NA, 0
- L mean, median, mode
- L Previous/next DP (time series)
- L Regression



10/01

The manager of a real estate wishes to monitor the performance of sales staff. She records the no. of properties sold by each of the 15 staff members for a randomly chosen period of time. What type of variable is "the no. of properties sold"? Justify.

Discrete, numeric value.

4M

Compare and contrast two techniques for visualising multivariate...

2D scatter plot, 11 coordinates

3M

If you have data on the yearly average temp. for Capetown International Airport from 2005 - 2020. If you are particularly interested in change of time, effective graphical display?

Line chart

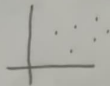
12/01

How do the visual cues, coordinate system, scale, context support data visualisation?

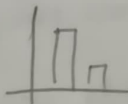
VISUALISATION COMPONENTS

Visual Cues

Position



Length



Angle

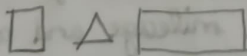


Rotation b/w vectors

Direction



Shapes



Area

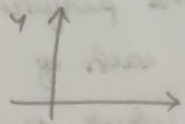


Colour saturation (Intensity)

Colour hue (Red, blue)

## Coordinate system

### Cartesian



### Polar

- radius
- angle

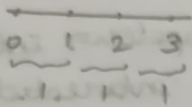


### Geographic

- Latitude - Longitude

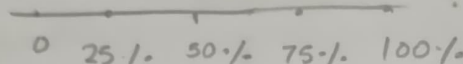
## Scale

### Linear

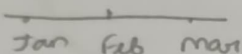


evenly spaced

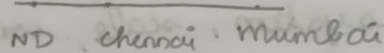
### Percentage



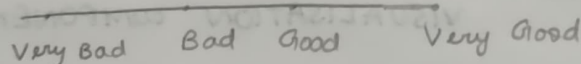
### Time



### Categorical



### Ordinal



## Context

- What
- When
- Where
- Who
- Why

19/01

The FORD 'focus' is a compact car introduced to NA for the model year 2000. The table shows the model year, mileage and asking price (in \$) for all ten used FORD focus automobiles advertised for sale on the website Seattle Times on Jan 31, 2010. Answer the following ques. based on your understanding of the data.

- What type of association would you expect to see b/w mileage and price?

Year	Mileage	Price
2007	25426	14595
2008	49223	13991
2008	49028	13991
2008	27690	11994
2008	36216	11980
2002	71646	10991
2007	41107	9671
2002	83454	8991
2007	49443	7998
2007	34179	7499

no association found.

(ii) Which one would be the explanatory variable?

Year and mileage

(iii) Which one would be the response variable?

Price

(iv) Create an appropriate graphical display to investigate this association.

Mileage  $\rightarrow$  x-axis

Price  $\rightarrow$  y-axis

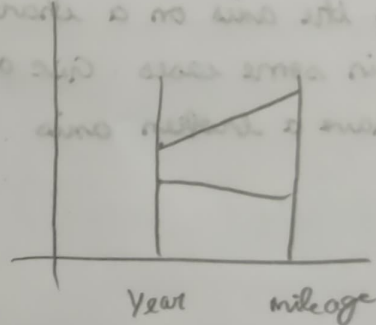
Different colors for different years.

Scatter plot

(v) What type of association b/w year and mileage.

Negative correlation.

// 11 coordinates also can be used.



(vi) What type of user interaction can be introduced?

Clicking on the coloured shape, year & price can be displayed.



23/01

It has been claimed that VODACOM has the highest market share among cellphone users in South Africa. A random sample of 250 cellphone users was asked which network they subscribe to. What type of data has been collected and which graphical technique will be the most appropriate to highlight the various market shares? Justify.

Sampling

Stratified

Probability

Non-probability

// Any sample randomly

// Certain samples have higher priority

Probability — Simple  
— Systematic

- Categorical data.

- Pie chart / Bar chart.

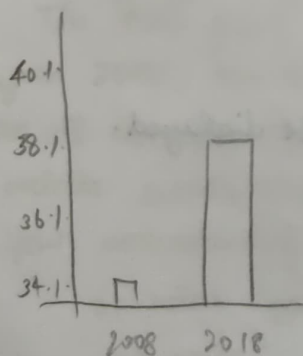
Ans: Simple random sampling method

24/01

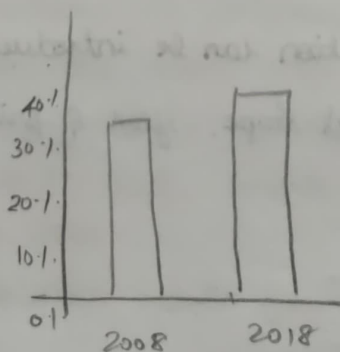
4M

- Suggest 4 qualitative/quantitative parameters to prepare a word cloud with all your classmates' names.
- Breaking the axis on a chart can be problematic, but it can also be useful in some cases. Give an example and which chart type should never have a broken axis.

30/01



Axis Broken



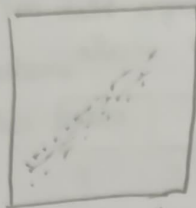
Actually change is less.



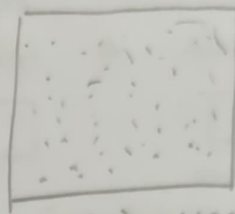
→ For each of the six scatter plots below, describe the strength of the relationship (example - weak, moderate / strong), the form of relationship (linear, non-linear) and the direction of the relationship (+ve, -ve, something else) and make note of any unusual features.



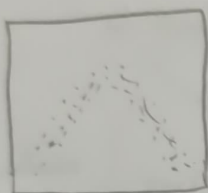
Strong (a) (NL)



Strong (b) (L)  
+ve



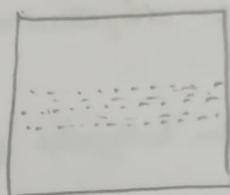
(c) weak



Strong (d) (NL)



Strong (e) (L)  
-ve



moderate (f) (L)  
SE

Unusual feature: (c) and (f)

31/01

Comment on the statement "vector graphics are preferred for data visualisation".

### Vector Graphics

- Store coordinates
- Less storage
- No blur
- Supports scalability
- Limited browser support

### Raster graphics

- store RGB intensity
- More storage
- Blur
- Not scalable
- All browsers support

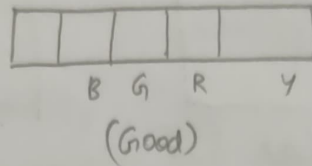
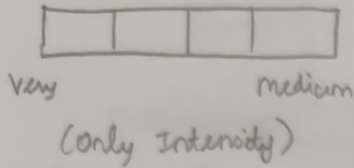
How does the type of audience affect your chart preparation?

Based on the audience, level of technicality.

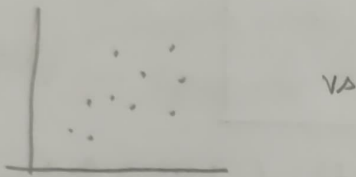
For the given 3 cases, which of the visualisation has better readability? Why?

→ Narrow color scale Vs

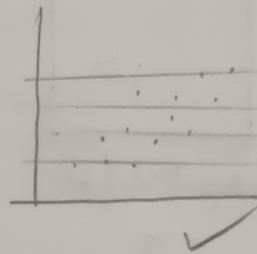
Wide color scale



→ Showing data points only



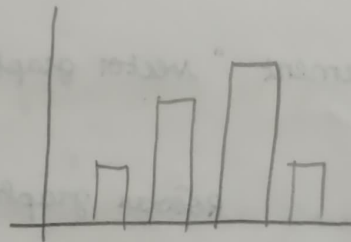
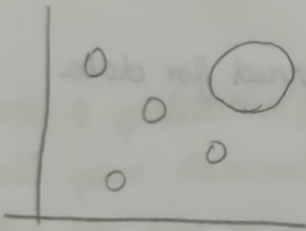
Additional visual elements



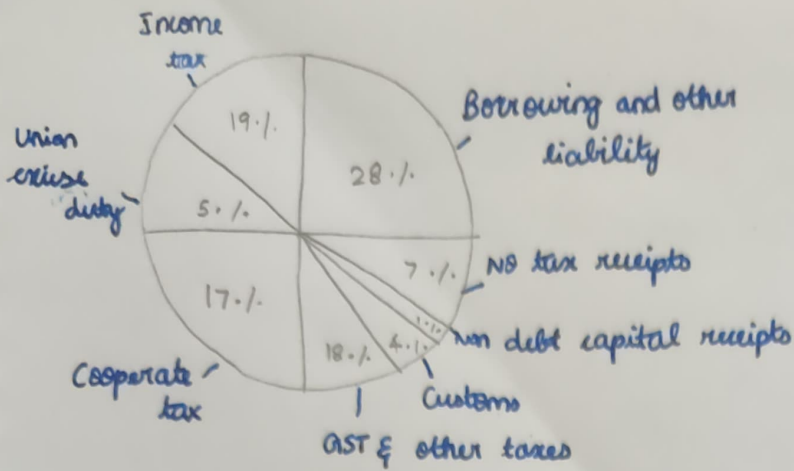
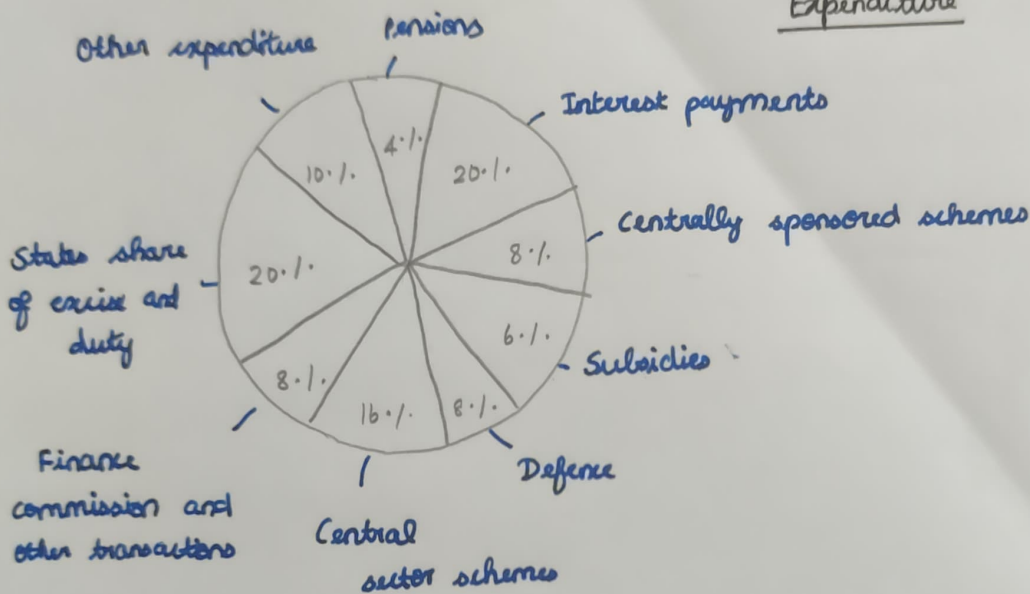
→ Using area as visual cue

Vs

Using length as visual cue



DPI - Density Per Inch

Expenditure

07/02

## 1) Match

- Descriptive → Optimisation techniques // Genetics, particles
- Predictive → Data mining techniques // Apriori
- Prescriptive → Machine learning algo. // Regression, CNN, LSTM

Give example of each.

## 2) Visualisation techniques and tools can be classified using 5 relevant models such as task, audience, target, medium and representation.

For the given 2 examples of visualisation applications

1. For the mobile phone sales data collected from an e-commerce website
2. Indian cricketers' performance since 2015.

Explain the 5 elts for these applications.