# Data Visualization

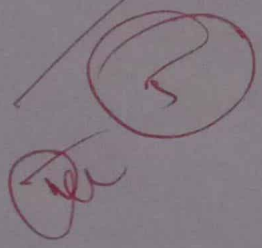
## Assignment - I

Name : V. Harshita

Vtu No : VTU 25115

Course Code : 10212CA214 (Data visualization) & Title

Registration No: 23UECL0063



Explain various concepts and Process of data visualization

Definition

Data visualization is the graphical representation of information and data using visual elements like charts, graphs, maps and infographics. It helps in transforming complex datasets into visual forms, making it easier to analyze.

Importance

1) Simplifies complexity: Summarizes large datasets into visual formats.
2) Pattern Recognition: Identifies trends, correlations and outliers.
3) Better decision making: Makes insights accessible to decision-makers.

Key concepts

Data types

- Categorical data: Represent categories (eg country names product types.
- Ordinal data: Data with a meaningful order (eg: rankings)
- Quantitative data: Numerical data (eg) sales no., Population.

2) Visual variables (Design Principles).

It uses visual elements known as visual variables (or) encoding channels to represent data attributes.

Position (x and y coordinates): very accurate

Length (Bar / Length / Height)

Angle (Bubble chart, Pie, Slices)

Color (Hue & Saturation) categorical vs gradient data representation.

Area (Bubble Chart size)

Types of visualization

Bar charts - Comparing categories

Line charts - Showing trends time

Pie charts - Parts of whole.

Scatter plots - Correlation between variables

Heat maps - Density and intensity of data points

Tree maps - Hierarchical data representation

Important Principles

1) Gestalt Principle:
   How humans naturally group visual elements
   (proximity, similarity)

2) Visual Hierarchy:
   Emphasize important data with size, color

3) Data-Ink Ratio (Tufte):
   Remove unnecessary decorations to focus on data

4) Pre-attentive Processing:
   Use, color, shape and size to highlight insights instantly.

5) Data Storytelling:
   Combine visuals and narrative for better understanding

Best Practices

1) Choose the right chart type
2) Avoid clutter
3) Label clearly
4) Highlight key insight

3) Explain detail about Bivariate analysis with suitable charts and plots

Definition

- Bivariate analysis studies the relationship between two variables.
- Helps us understand if there is a correlation, pattern or association between them.

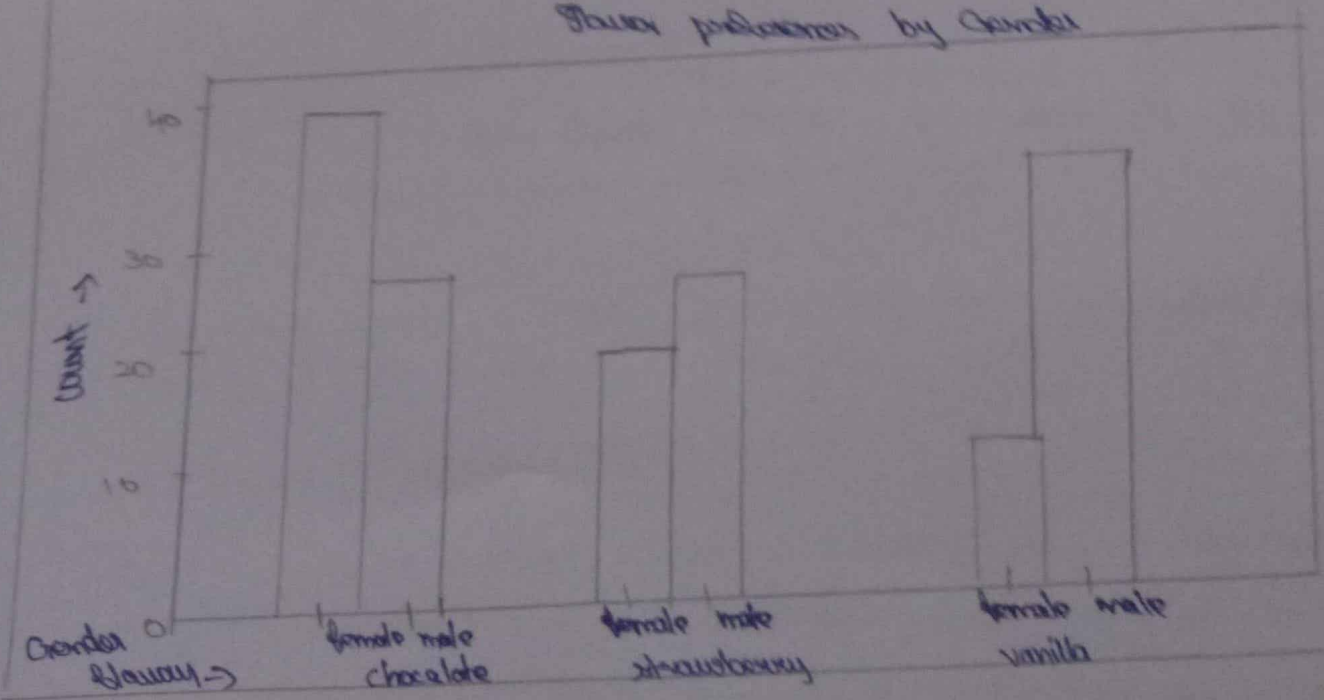
Variables can be

1) Quantitative vs Quantitative
2) Categorical vs Categorical
3) Quantitative vs Categorical

Types of Bivariate analysis

Quantitative vs Quantitative

⇒ Both variables are numerical (eg) height, weight)



Scatter plot: Shows relationship between two continuous variables

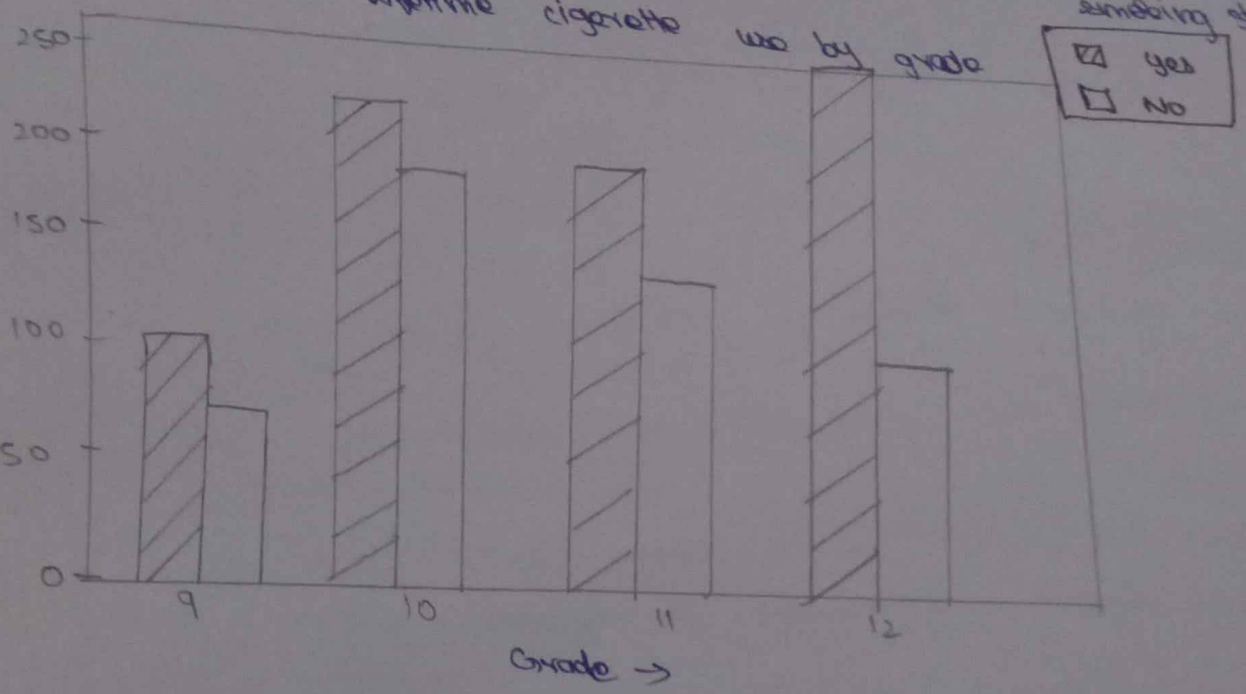
Ex: plot weight vs height → reverse correlation

Line plot: Group when data is time-based

(eg) sales over month.

Categorical Vs Categorical

- Both variables represent categories (eg; Gender & smoking status)




Stacked Bar chart: Shows distribution of one categorical variables across another.

Contingency Table: Frequency of combination of (cross tabulation) categories.

Interpretation:

Check for association (or) independence

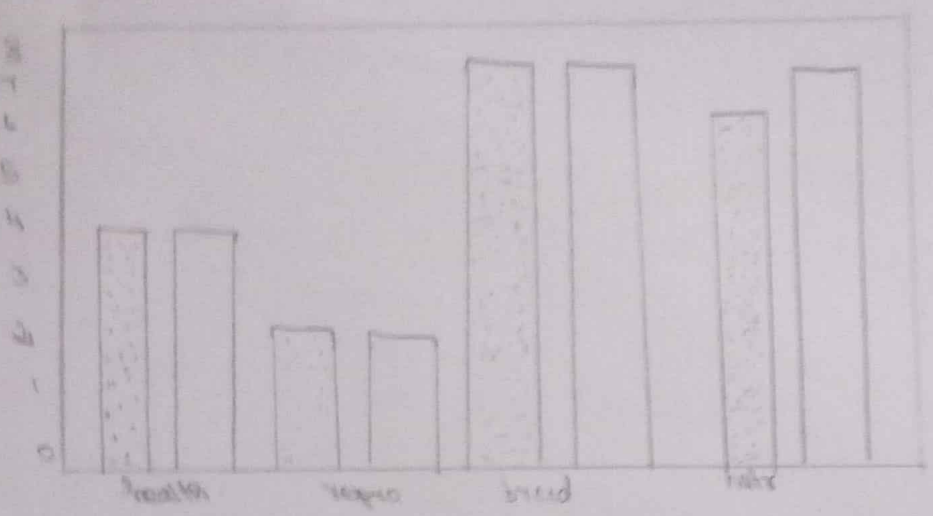
(eg via chi-square test)

1) Quantitative vs Categorical

One variable is numeric, the other is categorical

eg: Find score vs gender.

mean ranking in each category.




Box Plot: Show distribution of numeric variables across categories.

Violin Plot: Similar to box plot but shows the full distribution shape.

Bar Chart (with mean or median)

Display average of quantitative variable per category.

Interpretation:

mean, median and spread of quantitative data between categories.

Examples of Bivariate Analysis:

Data Example: Height vs weight, Gender vs Smoker status, salary vs department.

Type: Quantitative vs quantitative.
Categorical vs categorical.

# Data Visualization Assignment

(C10212CA214)

Name: V. HARSHITA

VTU NO: VTU 29115

Reg no: 23UECL0063

FACULTY NAME: DR. R.T. THIVYA LAKSHMI

DATE: 15-10-25

5

# Data Visualization Assignment

(C10212CA214)

Name: V. HARSHITA

VTU NO: VTU 28115

REG NO: 23UECL0063

FACULTY NAME: DR. R.T. PHIVYA LAKSHMI

DATE: 15-10-25

5

1) Steps in Designing Visualizations - Problems

Designing an effective visualization requires careful planning and execution.

Step 1: Define the Purpose and Audience

Before creating any visualization, you must understand why you are creating it and who will view it.

Problem:

If purpose is unclear or audience is not properly considered, visualization may fail to communicate effectively.

Step 2: Gather and Understand data

You need to collect the relevant data and access its quality, completeness & accuracy.

Problem:

Data can be messy, incomplete or contain errors. Poor-quality data leads to misleading visualizations

Step 3: Choose the Right Visualization Type

Selecting the appropriate chart or graph is crucial. Common types are bar charts, line graphs, scatter plots and heat maps.

Problem:

Using the wrong type of visualization can confuse the viewer.

Step 4: Design the Visual Elements

This step involves choosing colors, scales, labels, legends and layout. Visual hierarchy and clarity are important.

Problem:

Overloading the visualization with too many colors, labels or data points can overwhelm users.

Step 5: Add Context and Annotations

Providing Titles, axis labels, units, and helpful explanations helps viewers interpret data correctly.

Problem:

Lack of context causes confusion.

3) Visualization techniques for Geospatial Data

- Geospatial data involves location information such as coordinates, addresses, or regions.
- Visualizing such data helps us understand spatial patterns, trends and relationships.

a) Choropleth Maps.

* Represent data using, shades or colors within predefined boundaries based on a data value
* Colors usually represents ranges
* Use case: Showing poverty rates by country.
* Limitation: Can obscure data variability within regions.

b) Heat Maps.

* Show intensity or concentration of data points in an area using colors.
* Example: Visualizing cellular hotspots or WiFi signal strength in a city
* Benefit: Quickly highlights in clusters.
* Limitation: Can be less precise for exact location.

c) Point (Dot) Maps

* Plot individual data points as a dots on a map.

* Good for showing distribution and density

* Challenge: Too many points can clutter maps

d) Flow Maps

* Show movement (or) flow between places with lines (or) arrows, where width represents volume.

* Helps understand connectivity and movement patterns.

e) Cartogram

* Distort map areas based on data values instead of geographic size.

* Visually striking but can confuse geographic orientation.

f) 3D Surface Maps

* Represent elevation (or) terrain data using 3D models (or) color gradients.

* Useful in geography, urban planning (or) environmental science.

# 1) Regression and Model Selection

Regression analysis is a fundamental statistical tool in data science used to model and analyze relationships between variables.

## What is Regression?

It predicts (or) explains the value of a dependent variable (response) based on (or) more independent variables (predictors).

Example: Predicting house prices based on size, location, number of rooms

## Types of Regression

### Linear Regression

Models a straight-line relationship.

### Multiple Regression

Uses multiple predictors

### Logistic Regression

Used for binary outcomes.

(Eg) Yes / No.

Importance of Regression

* It is strong and known for relationships.
* Allows prediction on new data.
* Assists in identifying significant predictors

Model Selection

Choosing the best model involves balancing complexity using key criteria

Goodness of Fit

Measures like ($R^2$) or adjusted $R^2$ show how well the model explains data variation. Higher values (closer to 1).

Overfitting vs Underfitting

* Overfitting occurs when a model fits the training data too closely, capturing noise and failing to generalise.
* Underfitting happens when the model is too simple to capture underlying patterns.

Techniques

* Cross - Validation (splitting data)
* Information Criteria (Akaike (AIC and BIC)
* Stepwise Selection (adding or removing predictors)

In visualization Regression and Model Selection are powerful making sense of data and making predictions.