

Data Visualization

Assignment - I

$\frac{10}{10}$ 4

K. Gayatri

VTU 21743

Slot: SHL6

Explain how human Perceptual Processing models and Gestalt principles influence the effectiveness of data visualization. Discuss with suitable examples how visualization maximize information overload and as Gibson's Affordance theory, data abstraction and appropriate dataset representation.

Information:

Data visualization is one of the most powerful tools in modern communication because it transforms raw data into meaningful interpretable patterns. Its effectiveness depends on how well it aligns with human Perception and Cognition.

By applying perceptual principles like Gestalt Principle and concepts like Gibson's Affordance theory, data abstraction and proper dataset representation, designers can create clear, intuitive and userfriendly visualization that reduce confusion and enhance understanding.

Human Perceptual Processing models in Visualization:

Human Perceptual models explain how people quickly interpret visual data by recognizing patterns, colors, shapes and contrasts faster than text.

* Pre-attentive Processing:

The brain instantly notices features like color, size or orientation. Example: a red bar in chart of blue bar always immediately attention.

* Working memory Limits:

Since, human can only process few chunks of information at once (7 ± 2 rule) visualization should summarize data.

Gestalt principles and Data visualization:

Gestalt psychology explains how humans naturally perceive and grouping. These principles are essential

in visualization design because they help determine how users interpret graphs, charts or dashboards.

i) Proximity:

Elements that are close together are perceived as belonging to the same group.

ii) Similarity:

Objects with similar shapes, color or sizes are seen as part of the same category.

iii) Continuity:

The human eye prefers continuous lines and curves, Line chart are effective because viewers naturally follow trends along a smooth path.

iv) Closure:

Humans tend to fill in gaps to perceive a complete shape.

v) Figure - Ground:

People distinguish between foreground (focus) and background (context). Designers use this by ensuring important data stands out from gridlines or background elements.

Minimizing information overload and maximizing clarity

1. Gibson's Affordance Theory:

Gibson's Affordance Theory suggests that objects have inherent properties that indicates their possible uses. In visualization, this translates to intuitive design where the user instantly understands how to interact with (or) interpret the chart.

determine
dashboards.

2. Data Abstraction:

Raw data is often too large and complex for direct visualization. Data abstraction helps by reducing complexity while preserving meaning.

3. Appropriate Dataset Representation:

Choosing the right representation for the dataset is crucial. The wrong visualization may confuse the audience, even if the data is accurate.

* Time-Series data → Best shown with line charts

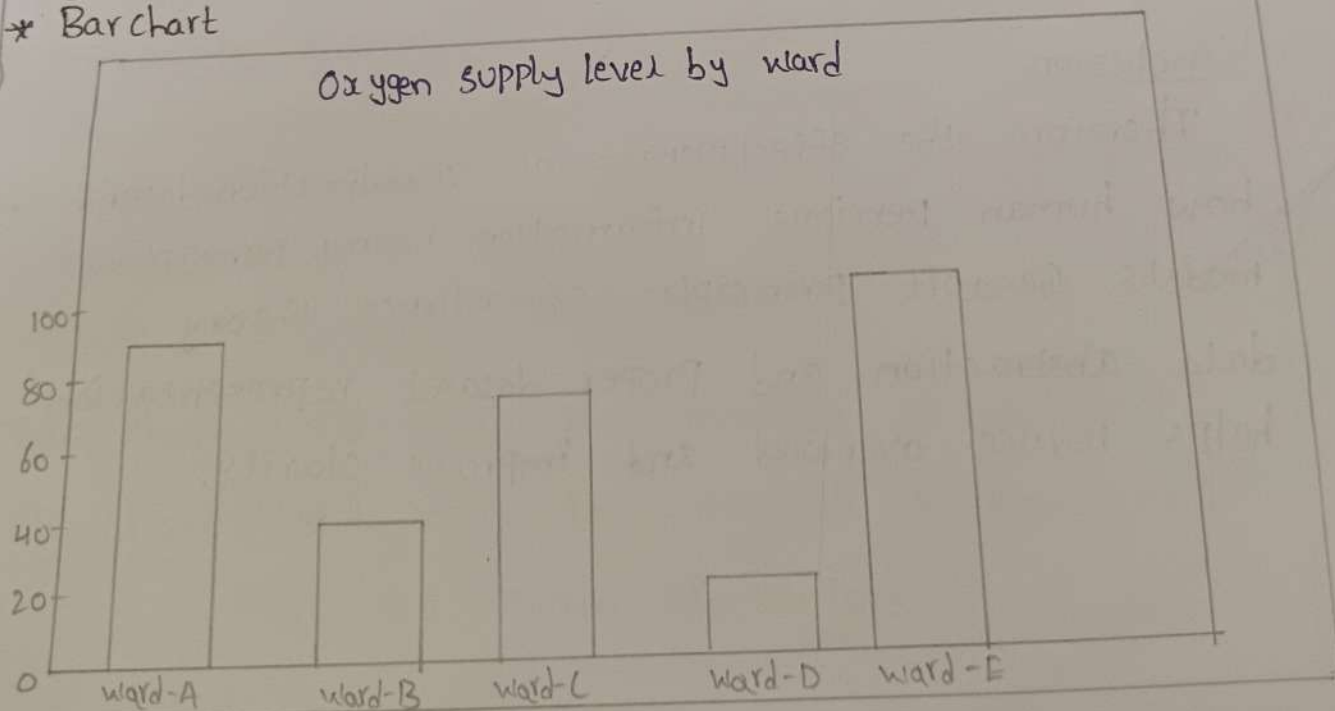
* Categorical Comparison → Piecharts are less effective than Bar charts.

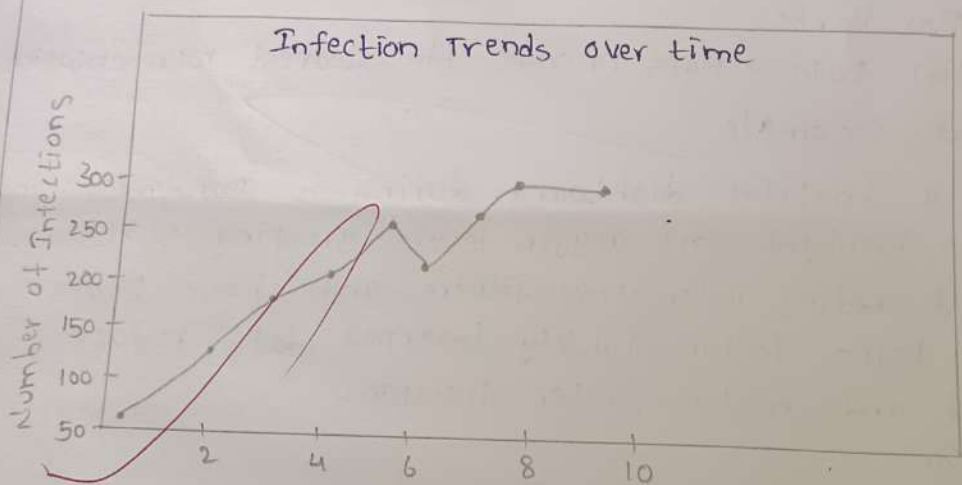
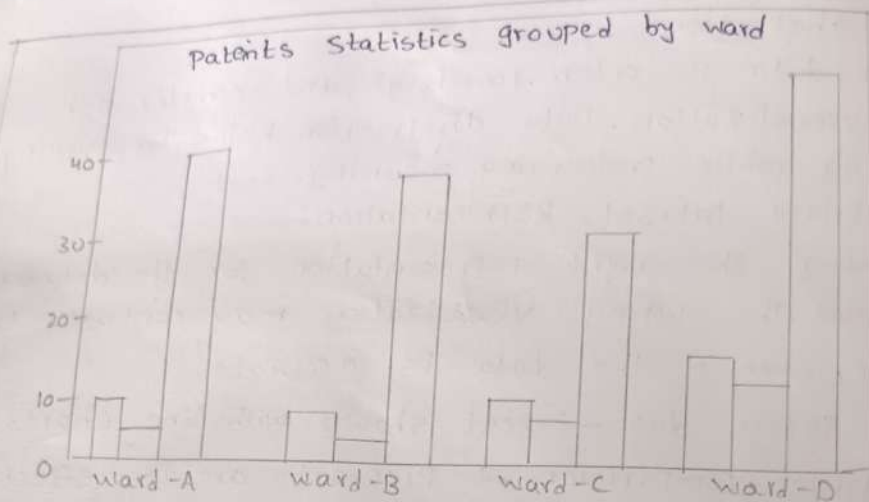
* Spatial Data → Maps provide the clearest representation.

Practical Example

In a hospital dashboard during a pandemic, using color to highlight low oxygen levels grouping stats by ward, adding interactive filters and chose proper Visuals helps doctor quickly interpret data, reduce over load and make faster decision.

* Bar chart





Conclusion!

Therefore, the effectiveness of visualization depends on how human perceive information using Perceptual models, Gestalt principles, affordance theory, data abstraction and Proper dataset representation helps reduce overload and improve clarity.

with the help of suitable datasets, compare and contrast different visualization techniques used in univariate, Bivariate and Multivariate analysis. Explain how the choice of visualization depends on the type of data and the number of variables being analyzed. Provide at least one practical example for each analysis type.

Introduction:

Data analysis is often categorized into univariate, Bivariate and multivariate analysis depending on the number of variables studied visualization. Plays a vital role in each type of analysis because it helps in identifying patterns, relationships and anomalies. The choice of visualization depends on:

1. Nature of data: Categorical vs Continuous.
2. Number of variable: One, two or more than two.
3. Objective of analysis: Distribution, Comparison or relationship

* Univariate Analysis:

Univariate analysis involves analyzing a single variable at a time to understand its distribution, central tendency and spread.

Suitable visualization Techniques:

- * Bar chart: Used when the variable is categorical
- * Pie chart: Used when the variable is partitioned into the form of sectors.

Examples:

- 1) Indian census (male and female categorized by bar chart)
- 2) Monthly Expenditure (pie chart) (variable: food, saving etc.)

Bivariate Analysis:

Bivariate Analysis deals with the relationship between two variables. The aim is to identify correlation, trends or difference between them.

Suitable Visualization Techniques:

- * Scatterplot with Fit line: Used for two continuous variables.
- * Side by side Box Plot: Used one variable is categorical and the other is continuous.
- * Grouped Bar chart: Used for two continuous variables.

Examples:

Consider a dataset of students study hours (continuous) and exam score (continuous)

- 1) A Scatterplot with a regression line can reveal whether more study hours lead to higher marks.

Alternately, with salary vs gender, a bar plot would reveal differences in distribution across categories.

Multivariate Analysis:

Multivariate Analysis involves three or more variables simultaneously to uncover complex relationships.

Suitable Visualization Techniques:

- * Heatmap:

Used for showing correlation matrices among multiple continuous variables.

Bubble Chart:

Similar to a scatterplot but with a third variable represented by bubble size or color.

* Pair Plot:

Displays scatterplot for all pairs of variables in a dataset.

* 3D Plots:

Used when analyzing multiple dimensions simultaneously.

Example:

Consider a dataset of patients in a hospital with variables: age (continuous), blood pressure (continuous) and cholesterol level (continuous)

- 1) A heatmap can show correlations between age, blood pressure and cholesterol
- 2) A bubble chart could plot age vs cholesterol while using bubble size to represent blood pressure.

Conclusion:

Therefore, univariate, bivariate and multivariate visualization serve different purposes, distributions, relationships and complex interactions. Choosing the right chart type based on data type and helps turn raw data into insights, supporting better decisions.