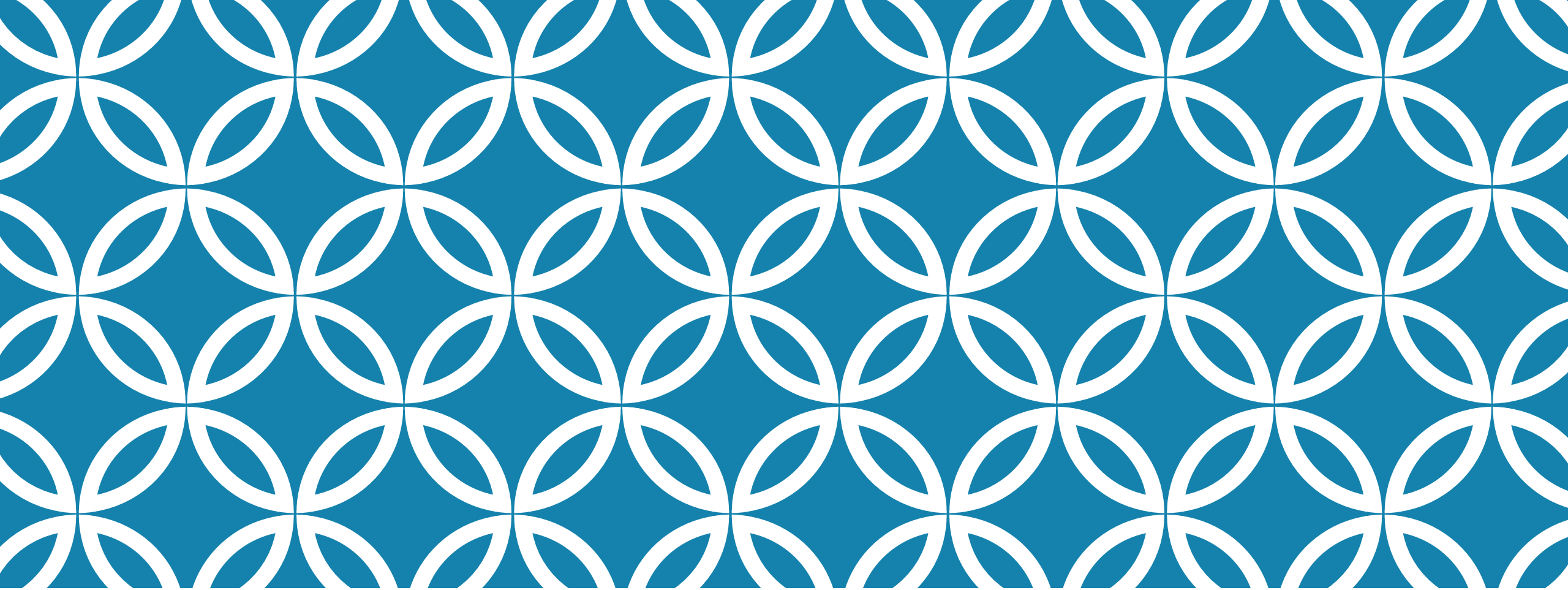


CHAPTER 1 INTRODUCTION TO VISUALIZATION

Dipesh Koirala

OUTLINE

- Introduction of visual perception
- Visual representation of data
- Data abstraction
- Visual encodings
- Use of color
- Perceptual issues
- Information overloads

LET'S BEGIN FROM HERE

What is data?

- Collection of records and their attributes
- An attribute is a characteristic of an object
- A collection of attributes describe an object

Attributes

Objects

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

LET'S BEGIN FROM HERE

Data comes from everywhere

But, they have different form

Hospital

Weather Station

Social Media

LET'S BEGIN FROM HERE

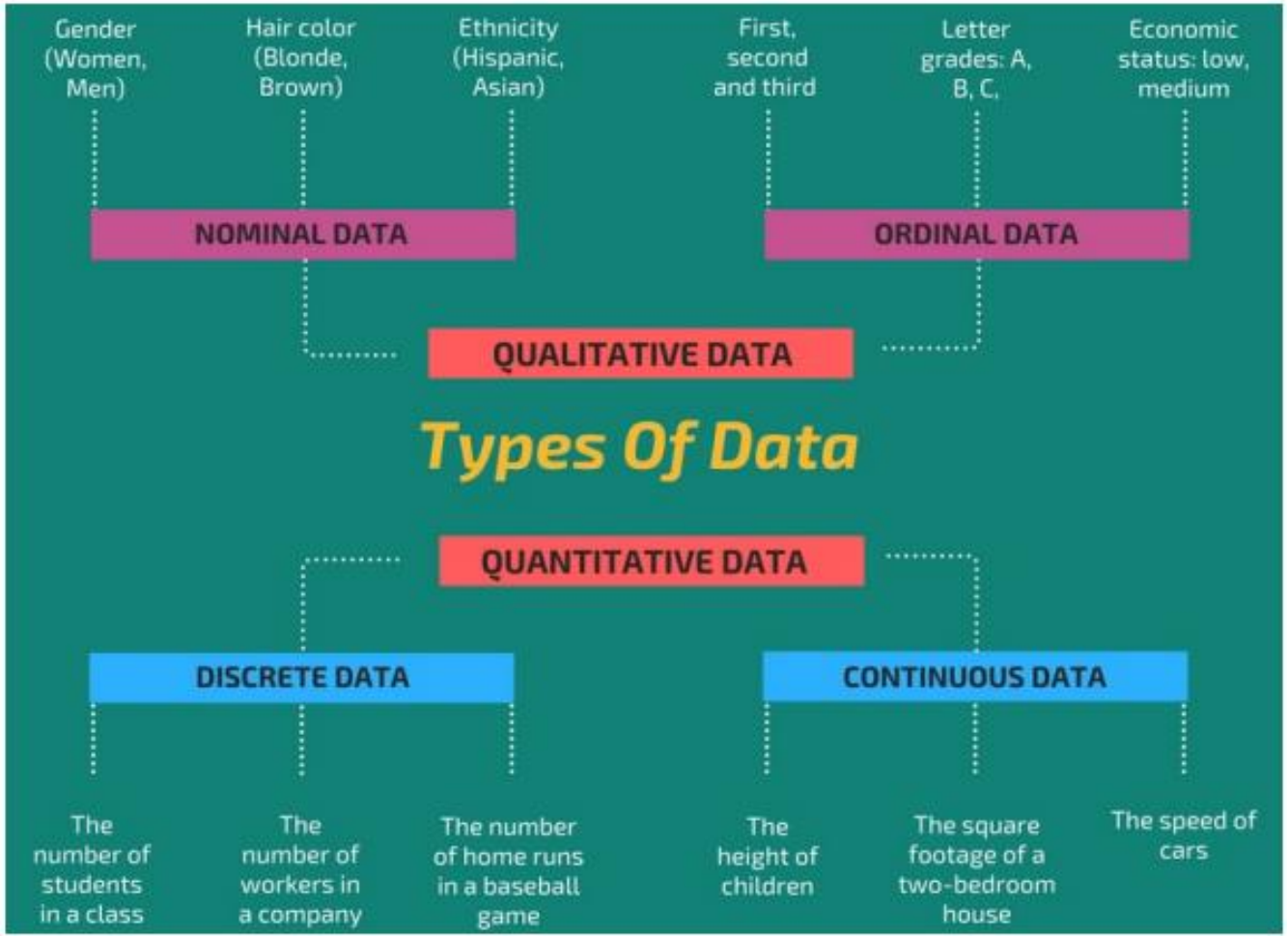
Types of Source of Data

- Record Data
  - Transactional Data
- Temporal Data
  - Time Series Data
  - Sequence Data
- Spatial & Spatial-Temporal Data
  - Spatial Data
  - Spatial-Temporal Data

- Graph Data
  - Transactional Data
- UnStructured Data
  - Twitter Status Message
  - Review, news article
- Semi-Structured Data
  - Paper Publications Data
  - XML format

LET'S BEGIN FROM HERE

Types of Source of Data



LET'S BEGIN FROM HERE

Some stats:

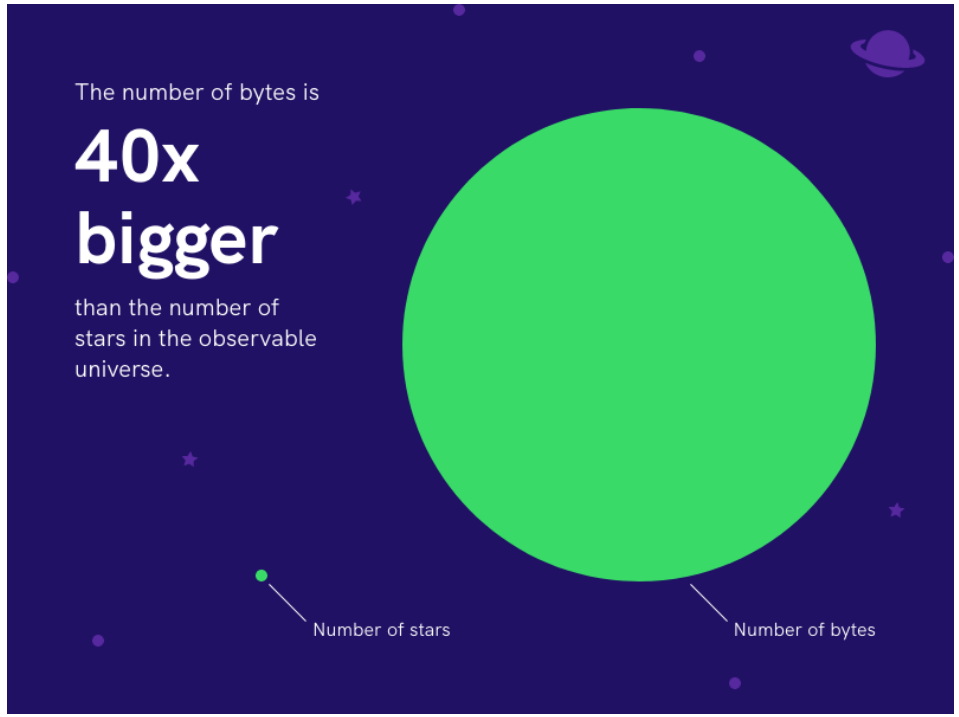
- The human brain can process an image in just **13 milliseconds** (Source: MIT)

- Human brains process visuals **60, 000 times faster** than they do text. (Source: University of Minnesota)

CONTINUE

Capture this

Facebook users post around 350 million photos in a day, which contributes 4 petabytes of data. (Raconteur). To help you grasp the magnitude of this number, the Milky Way Galaxy is home to approximately 200 billion stars. Assuming every individual star was a single byte, we would need 20,000 Milky Way Galaxies to match the number of stars with the number of data created by Facebook users each day.



Source: https://piktochart.com/blog/data-visualization-statistics/

No of bytes daily in facebook > Stars in Milky Way Galaxy

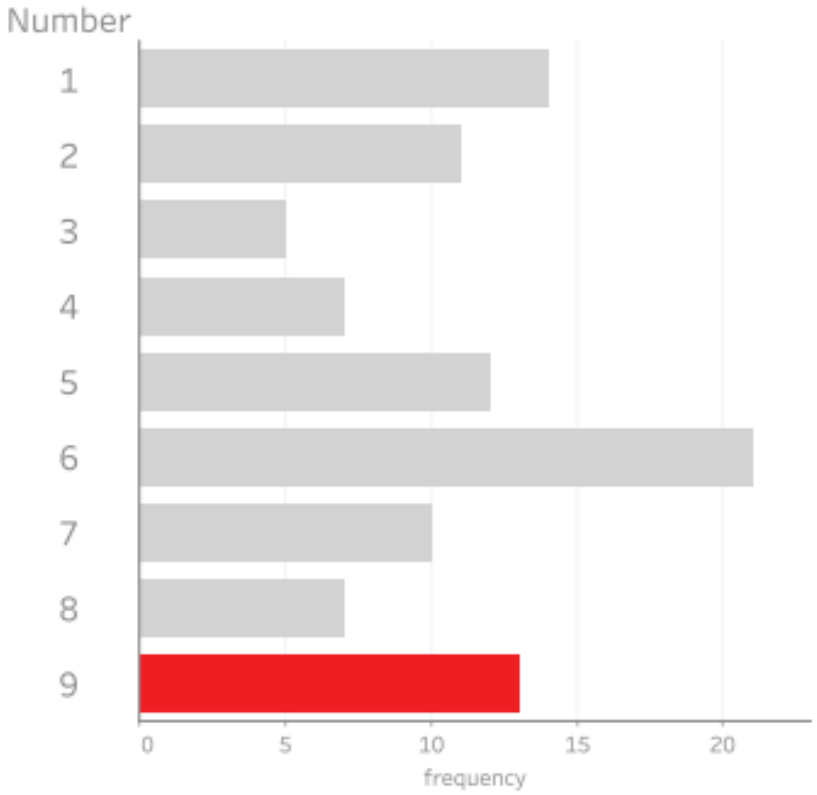
CONTINUE

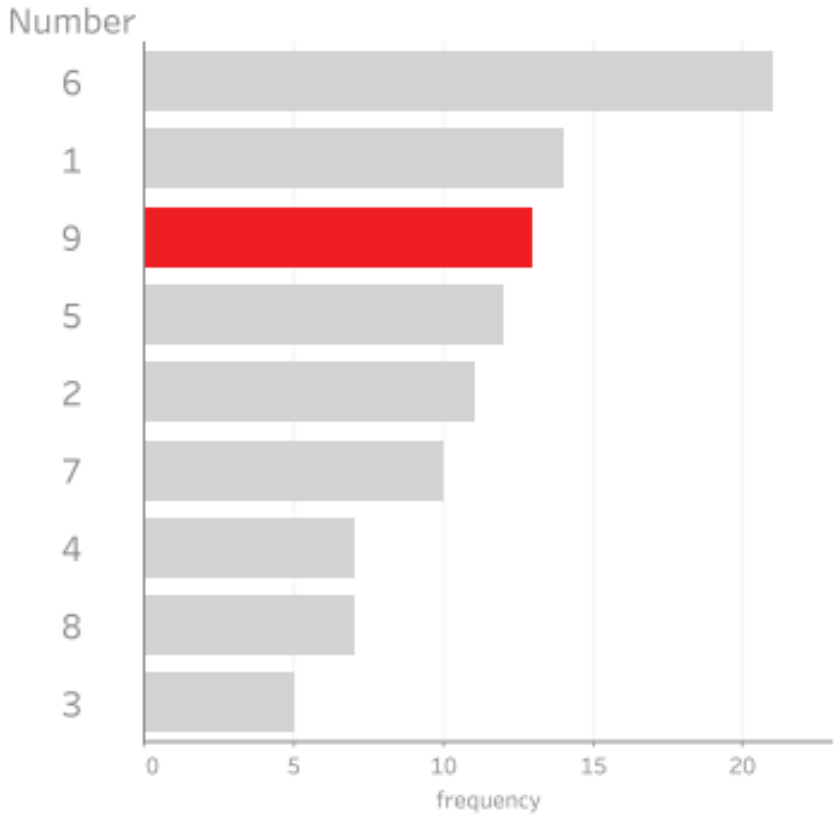
Why do we need data visualization?

2 2 5 6 7 1 1 6 9 1
9 1 7 5 5 5 6 2 5 9
4 5 2 9 6 9 7 6 4 6
8 1 5 7 8 5 6 6 6 7
7 2 3 6 8 9 1 7 9 1
3 8 6 8 4 5 6 9 4 5
4 9 9 2 3 7 1 9 1 2
3 7 8 1 6 1 5 6 1 6
5 6 6 8 6 6 9 1 2 6
3 2 4 2 6 9 4 2 7 1

CONTINUE

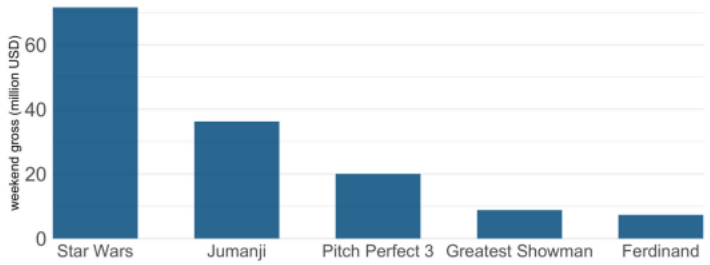
Why do we need data visualization?

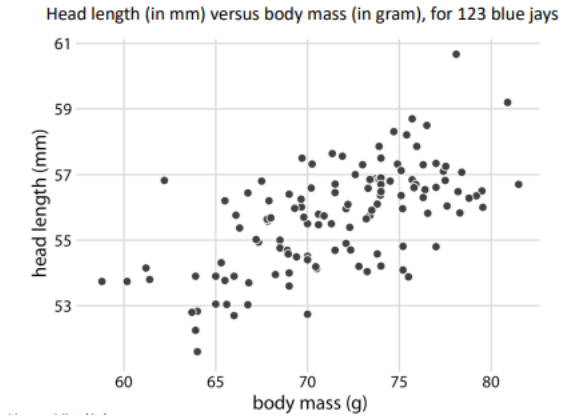


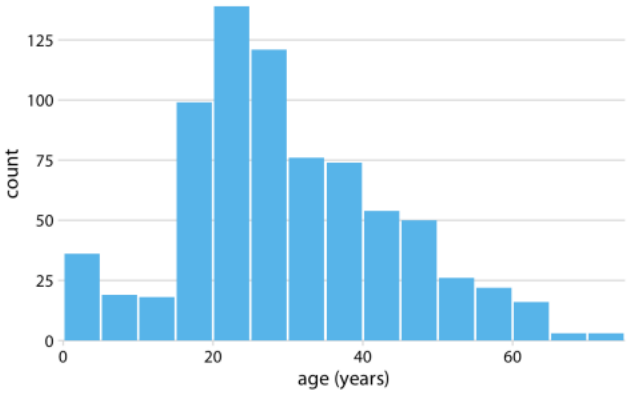


CONTINUE

- Data visualization is the creation and study of the **visual representation of data**
- Data visualization involves converting our data sources into visual representations
- These might be **charts, maps, graphs**







CONTINUE

What makes an effective data visualization?

- It has clear purpose and message

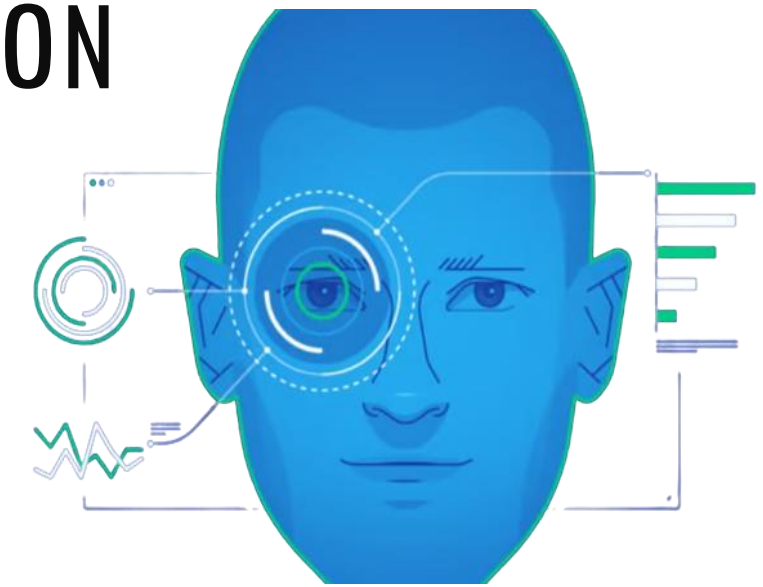
  Complements and enhances the text

- It is easy to interpret

  Note: A good figure is like a joke, if you have to explain it, it's not that good.

- It accurately reflects the data.

INTRODUCTION TO VISUAL PERCEPTION



- Visual perception is the process by which the *human brain interprets visual information* from the environment.
- Visual perception plays a crucial role in our everyday lives, as it allows us to make sense of the world around us and to interact with it effectively.

- An example of visual perception in action can be seen when **we look at a simple object, such as a ball**. When we see a ball, our eyes detect the light that is reflected off its surface and send this information to the brain.
- The brain then processes this information to create a visual **representation of the ball, including its size, shape, color, and texture.**

CONTINUE

Create 3D representation to interact with the environment.

- Visual perception also allows us to *perceive depth and distance.* For example, when we look at a mountain range, our brain uses cues such as perspective, shadow, and interposition to determine which peaks are closer and which are farther away.

Social Communication

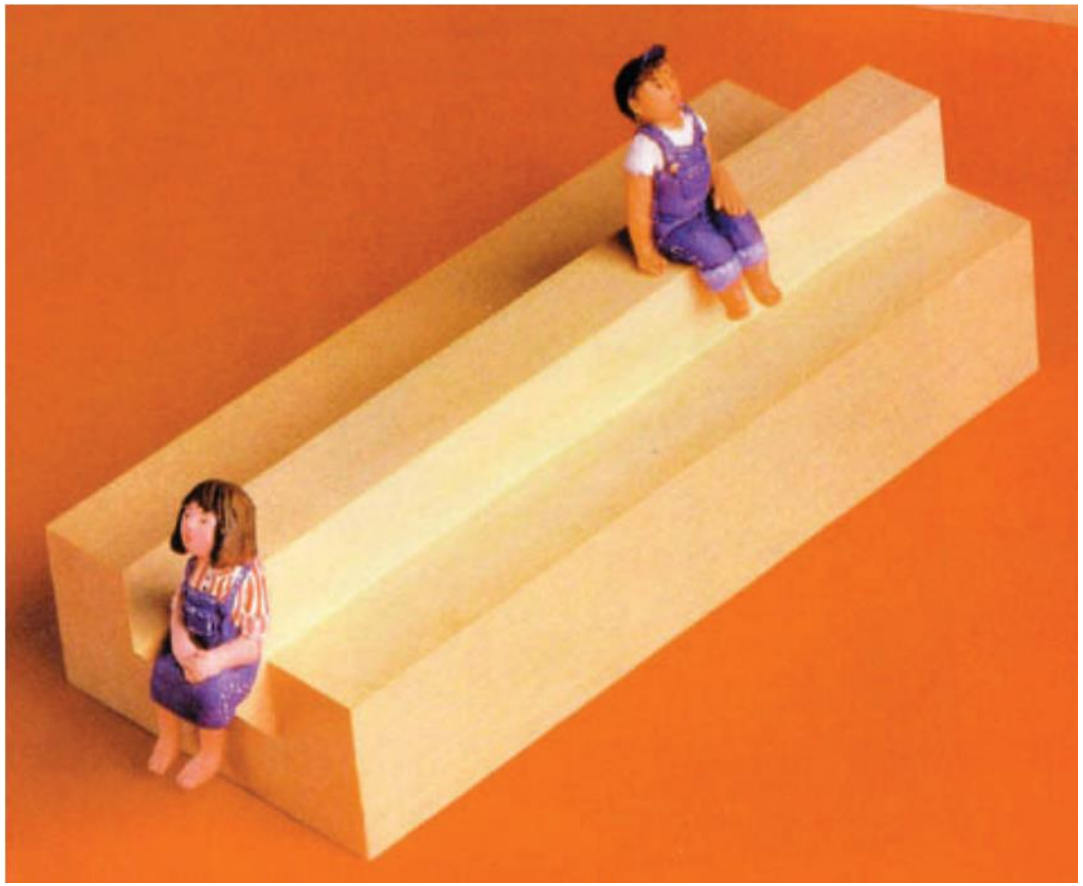
- *Facial expressions, body language*, and other visual cues are used to convey emotions, intentions, and social status, allowing us to communicate and interact with others effectively.

CONTINUE

- Perception is **studied to better control the representations of data** and eventually to utilize or exploit human perception.

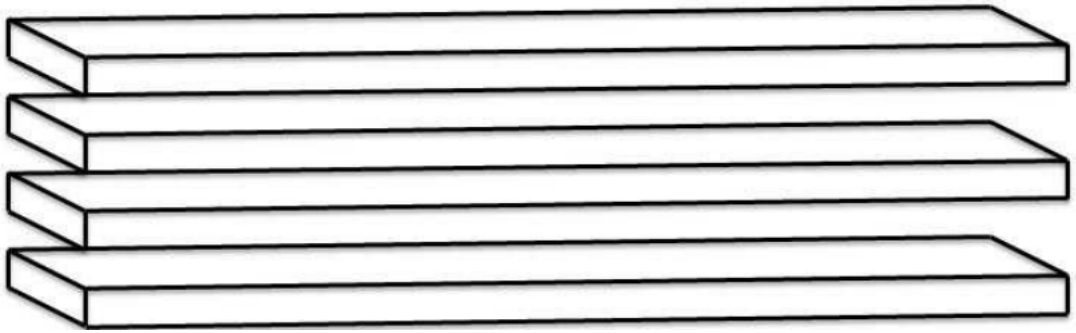
- Not paying attention to perception will lead to problems in visualization.

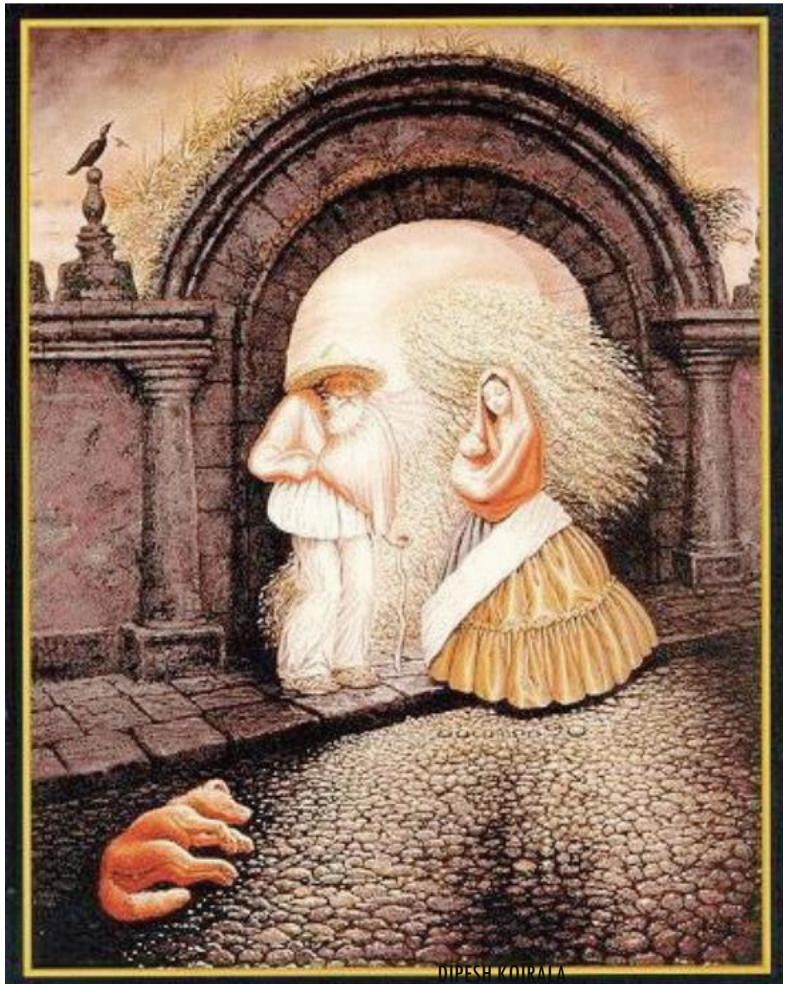
- *How can we be sure that the data we present is understood?*



CONTINUE

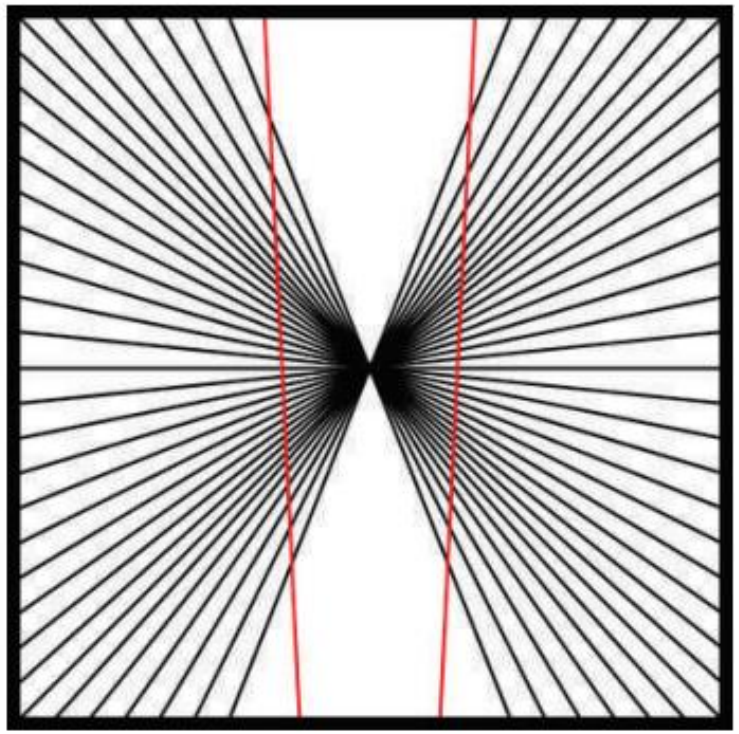
- *Visual representations of object are often misinterpreted* either because they do not match our perceptual system or they are intended to be misinterpreted.

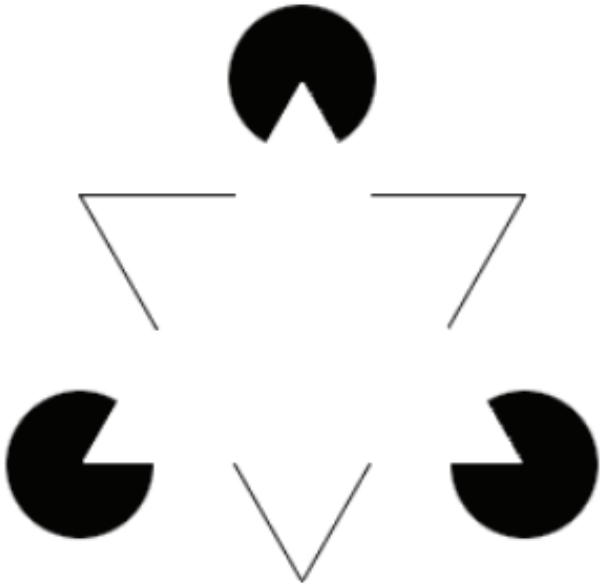




CONTINUE

Some illusions:

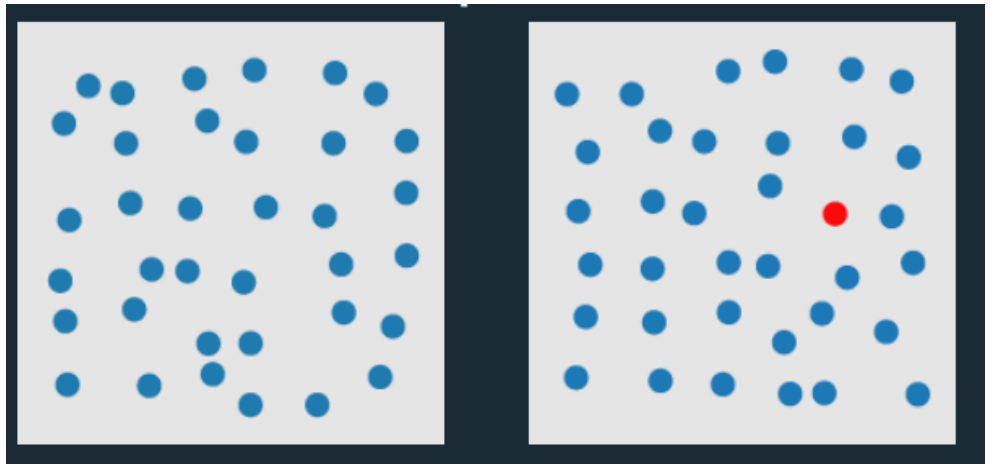


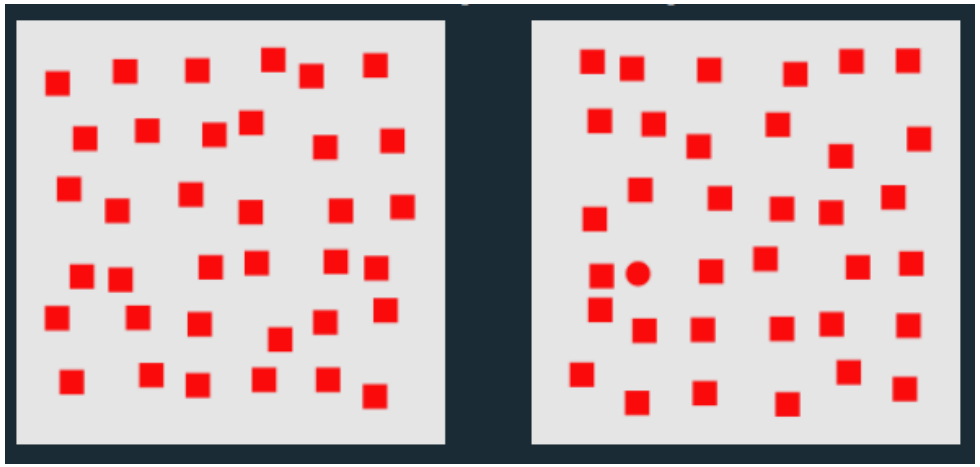


KNOWING OF PERCEPTUAL SYSTEM

Pre-Attentive Processing :

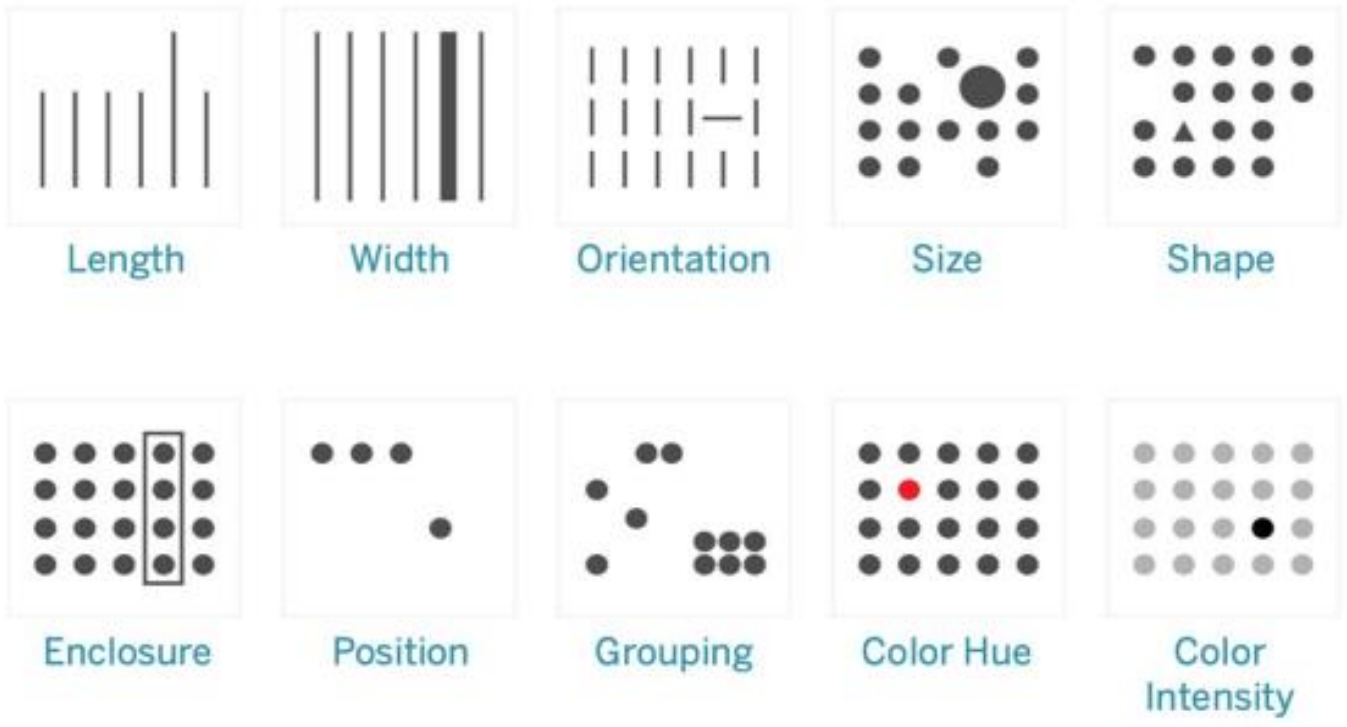
- How easy is it to spot some values from the rest?
- How immediately does our visual system perceive differences in a scene?





KNOWING OF PERCEPTUAL SYSTEM

Pre-Attentive Attributes



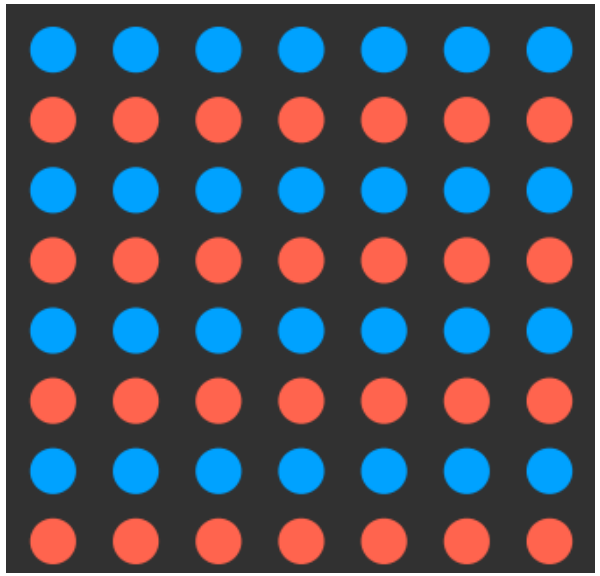
GESTALT PRINCIPLES

Principles or laws of human perception that describe how humans group similar elements, recognize patterns and simplify complex images when we perceive objects

1. Similarity
2. Proximity
3. Uniformed Connectedness
    a. Connection
    b. Enclosure
4. Continuity
5. Symmetry

GESTALT PRINCIPLES

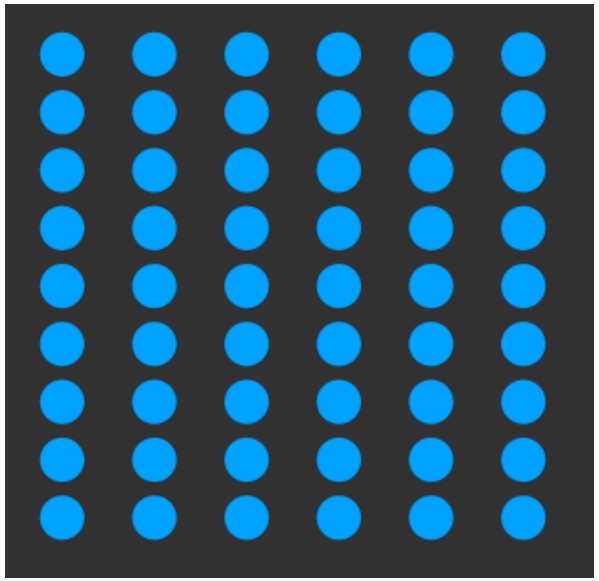
Similarity



Human nature to group like things together.

They can be grouped by color, shape or size

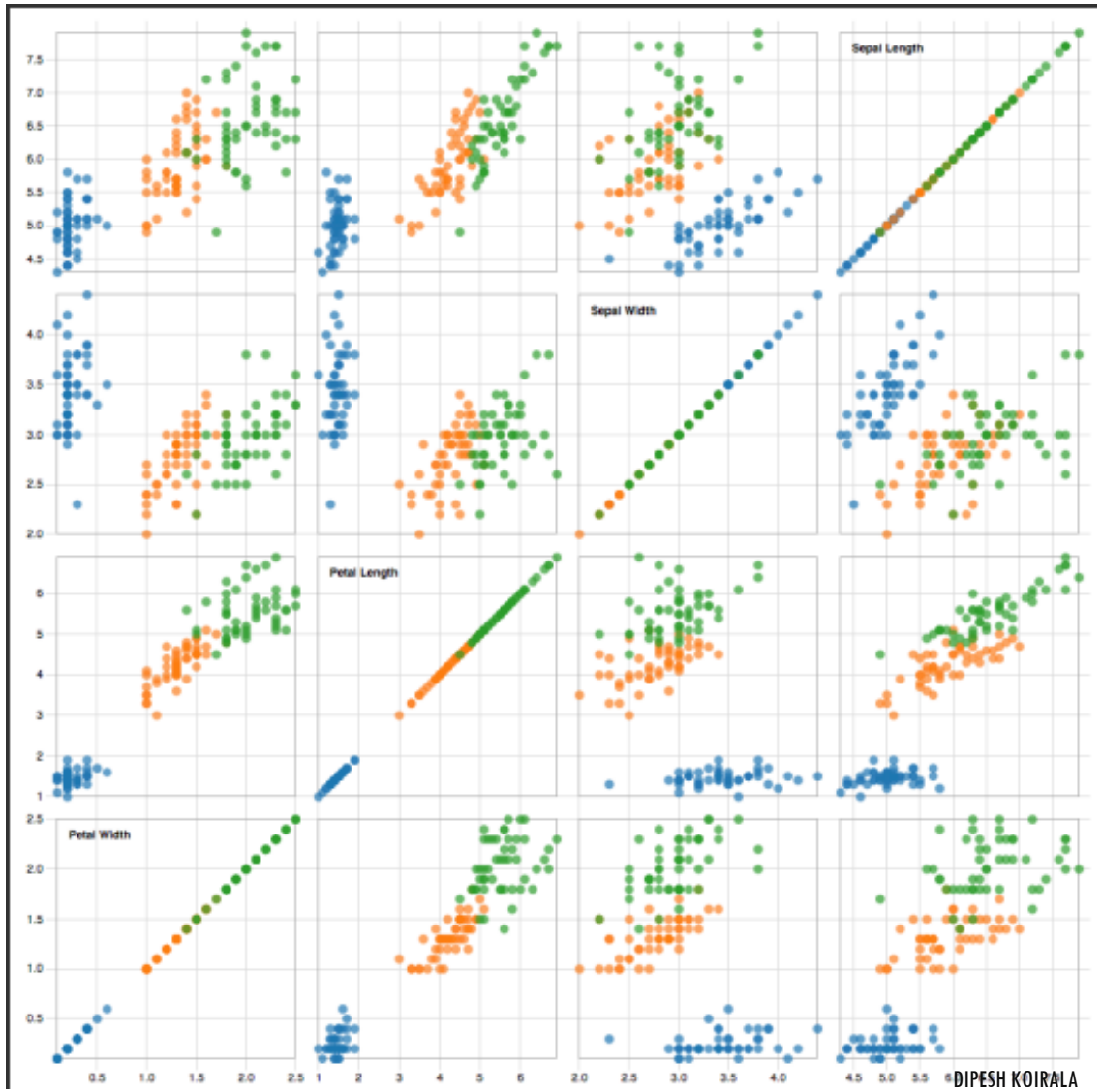
Proximity



Proximity refers to how close elements are to one another

GESTALT PRINCIPLES

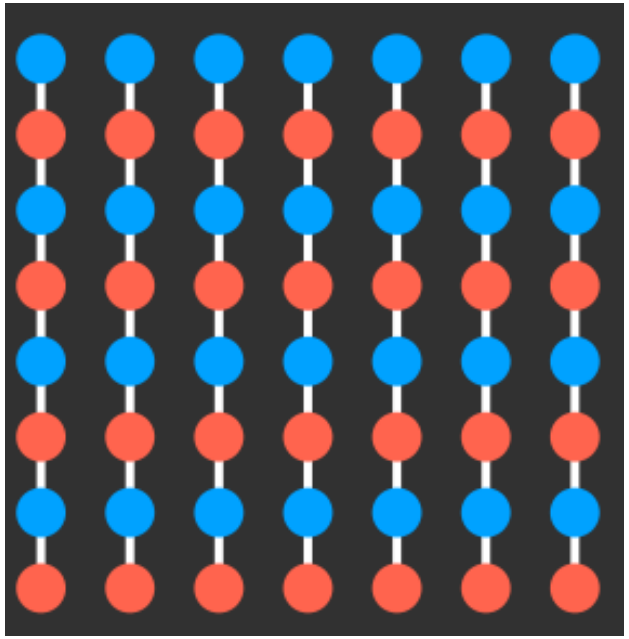
Proximity and Similarity

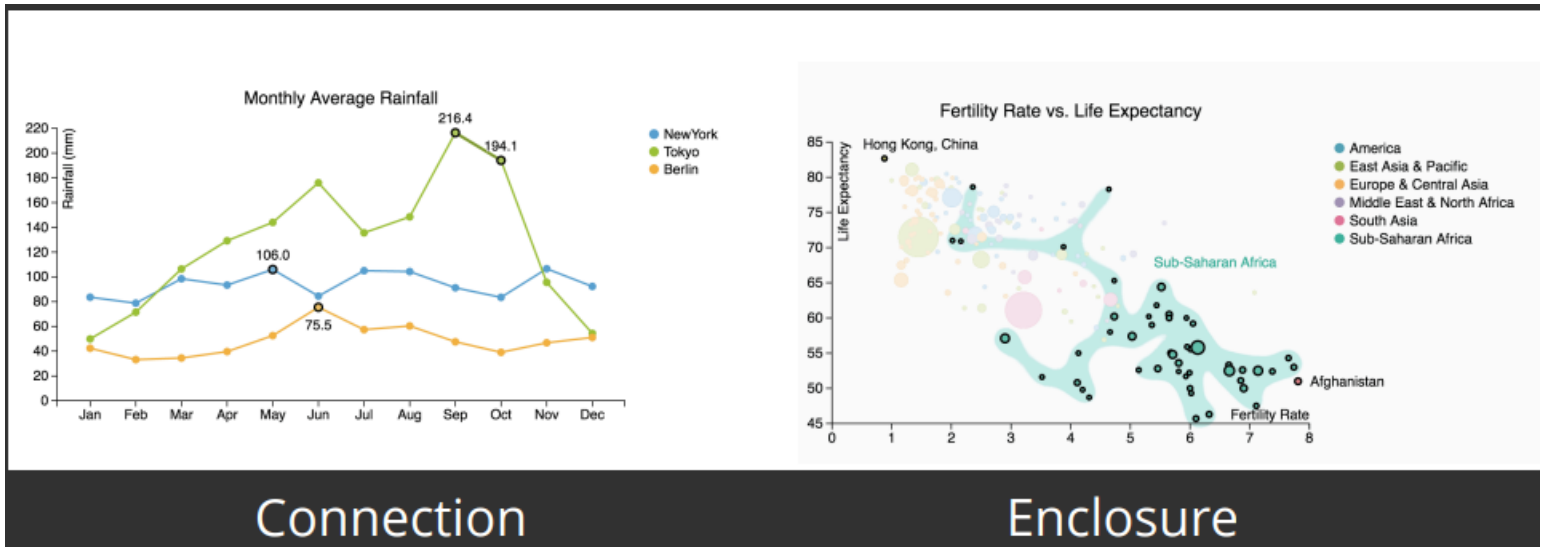


GESTALT PRINCIPLES

Connectedness

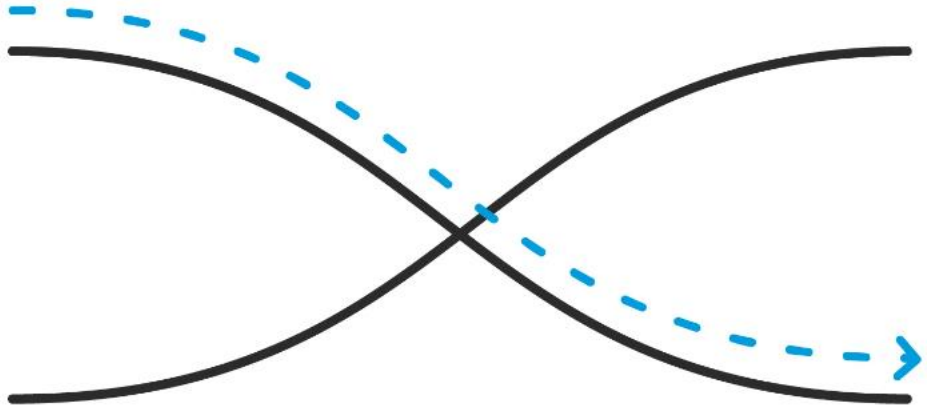
- Connectedness dominates proximity and similarity





GESTALT PRINCIPLES

Continuity



- Human eye will follow smoothest path when viewing lines regardless of how the lines were actually drawn.

Symmetry

[ ] { }

- Human eye tend to see visual elements as grouped when they are arranged.

GESTALT PRINCIPLES

Emergence

- Human process the world without thinking too much about understanding every small things around us.
- We perceive as group.

VISUAL REPRESENTATION OF DATA

- Visual representation of data involves **the use of graphical elements to present complex data sets** in a way that is easy to understand and interpret.
- Use of **graphs, shapes, color** that allow users to quickly grasp key insights from large amounts of data.

1. Simple Text

- When there is just a number or two to share,

  simple text can be a great way to communicate

VISUAL REPRESENTATION OF DATA

2. Tables

- Tables are great for just that-communicating to a mixed audience whose members will each look for **their particular row of interest.**
- If there is need to communicate multiple different units of measure, this is also typically easier with a table than a graph.

Variation: Heatmap

Student	Physics	Chemistry	Math	English
Peter	80	65	80	74
Samuel	90	85	72	65
Ruth	78	75	60	82
John	82	72	85	92
Mark	70	60	65	60
Andre	75	80	70	82

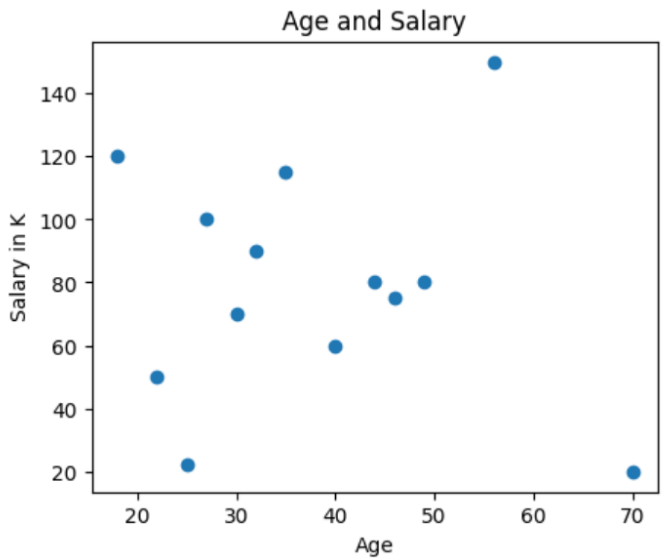
VISUAL REPRESENTATION OF DATA

3. Graphs

- Graphs *interact with our visual system*, which is faster at processing information.
- Four categories: Points, lines, bars and area.

Points : Scatterplot

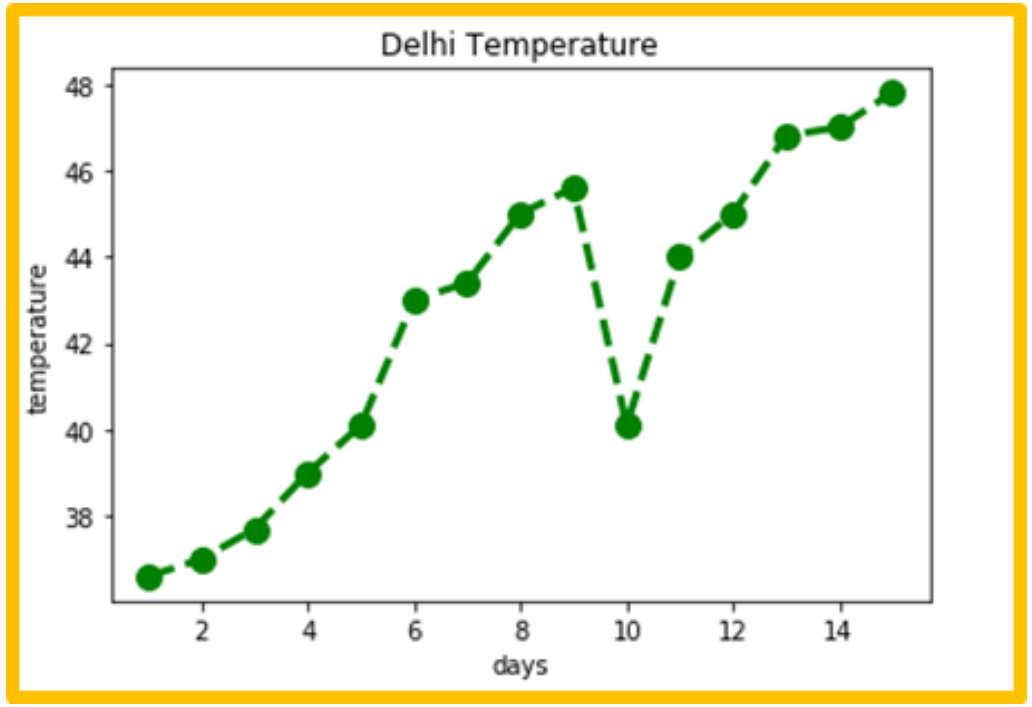
- Useful for showing the *relationship between two things*, because it allow to encode data simultaneously on horizontal x-axis and vertical y-axis.
- It allows seeing whether and what relationship exists.



VISUAL REPRESENTATION OF DATA

Lines: Line graph

- Line graphs are most commonly used to plot continuous data. The *continuous data is in some unit of time: days, months, quarters or years.*



VISUAL REPRESENTATION OF DATA

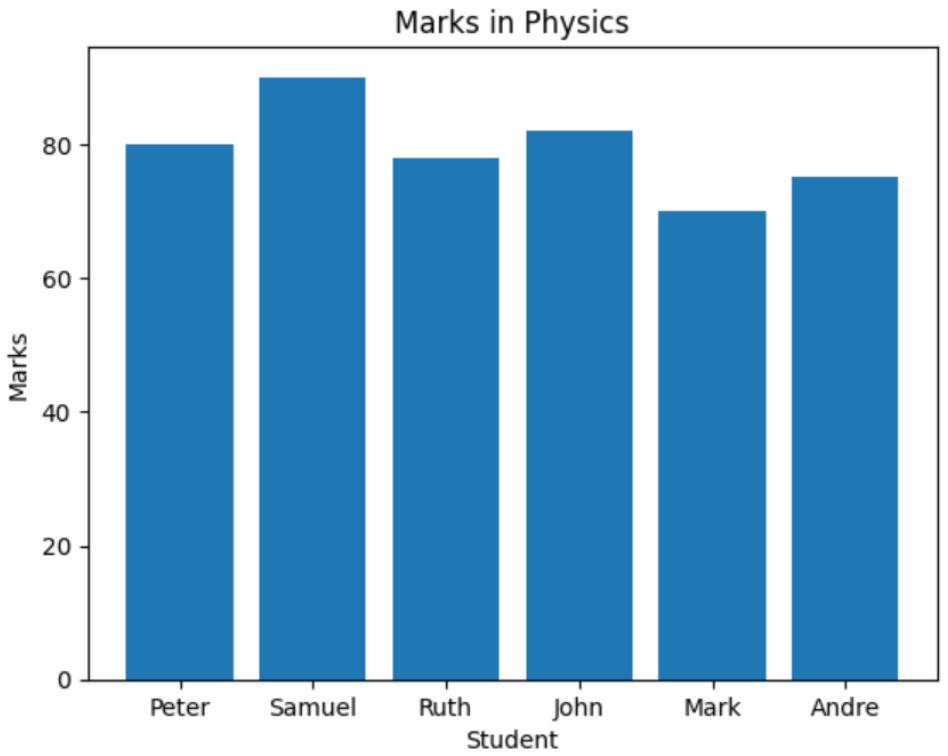
Bars: Vertical bar chart, stacked vertical bar chart, waterfall chart, horizontal

- A bar chart is used when we have *numerical data that splits nicely into different categories.*
- Able to quickly see trends within your data.

Variations: stacked vertical bar chart,

horizontal bar chart,

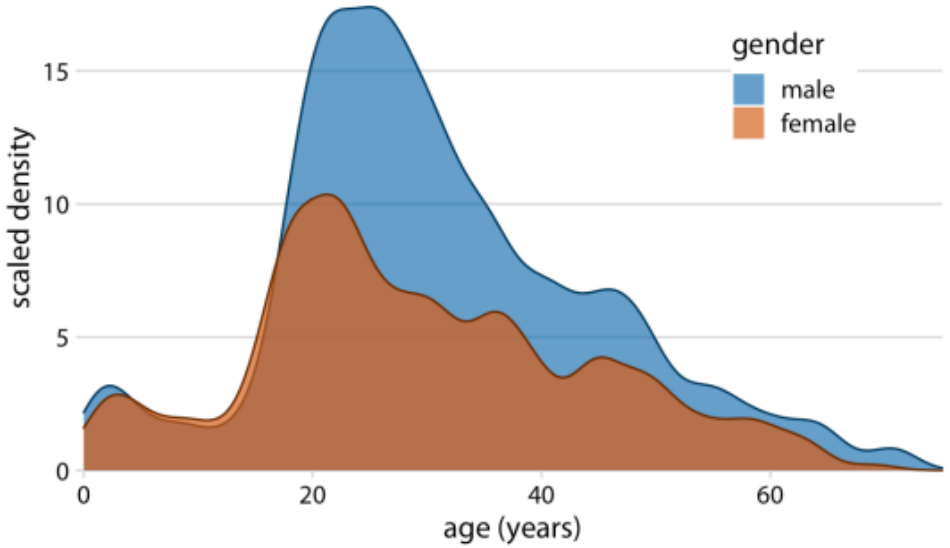
horizontal stacked bar chart,



VISUAL REPRESENTATION OF DATA

4. Areas

- To show **proportion or distribution** of certain categorical group.
- Used less compares to point line and bars.



DATA ABSTRACTION

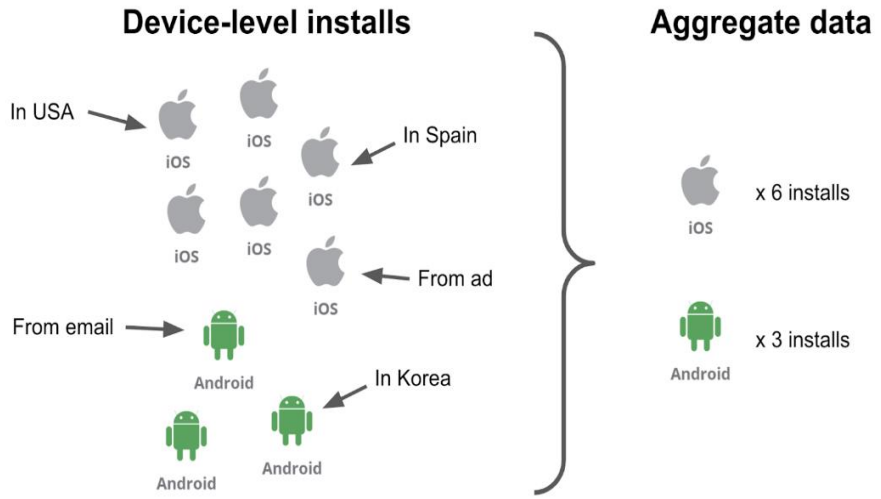
- Data abstraction is the process of simplifying complex data by *focusing on the most important features or characteristics*, while ignoring or hiding irrelevant details.
- *Involves reducing the amount of data* presented to the user to provide a clearer, more concise view of the underlying patterns or trends.

i. Aggregation
ii. Sampling
iii. Filtering
iv. Clustering
v. Dimensionality Reduction etc.

DATA ABSTRACTION

1. Aggregation

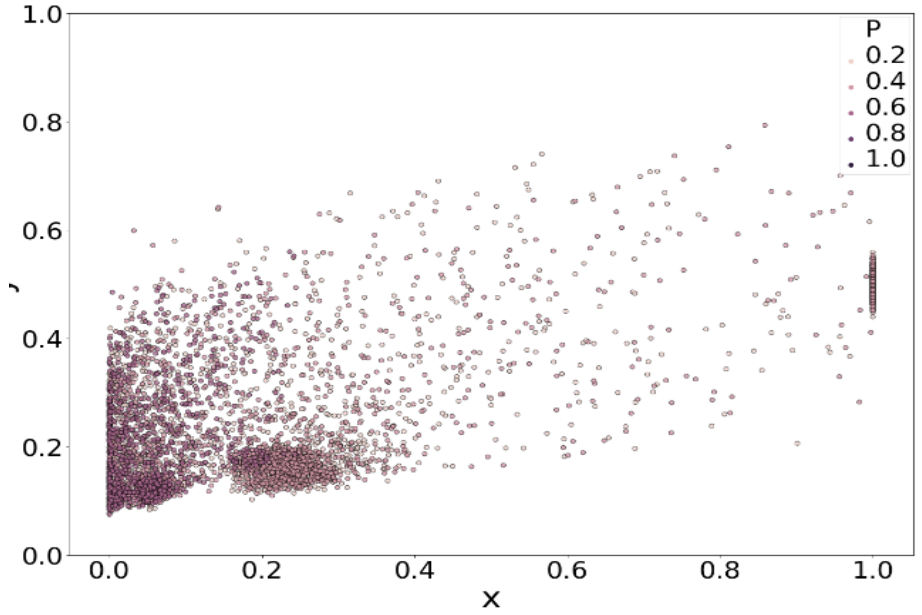
- Aggregation involves *combining multiple data points into a single summary value*. This is often done to provide a higher-level view of the data, or to make it easier to compare values across different groups.
- **E.g.:** A bar chart that shows the sales figures for different products can be aggregated by product category to provide a summary of the total sales for each category.



CONTINUE

2. Sampling

- Sampling *involves selecting a subset of the data for visualization*. This is often done to provide a representative view of the data, or to reduce the amount of data being presented.
- **E.g.:** A scatter plot that shows the relationship between a person's age and their income can be sampled to include only a subset of the population, such as individuals within a certain age range.



CONTINUE

3. Filtering

- Filtering involves selecting a *subset of the data based on certain criteria.*
- This is often done to focus on a specific aspect of the data, or to remove outliers or irrelevant data points.
- **E.g.:** A heat map that shows the temperature distribution of a region can be filtered to show only temperatures above a certain threshold, or to exclude data points from areas with inconsistent or unreliable temperature readings.

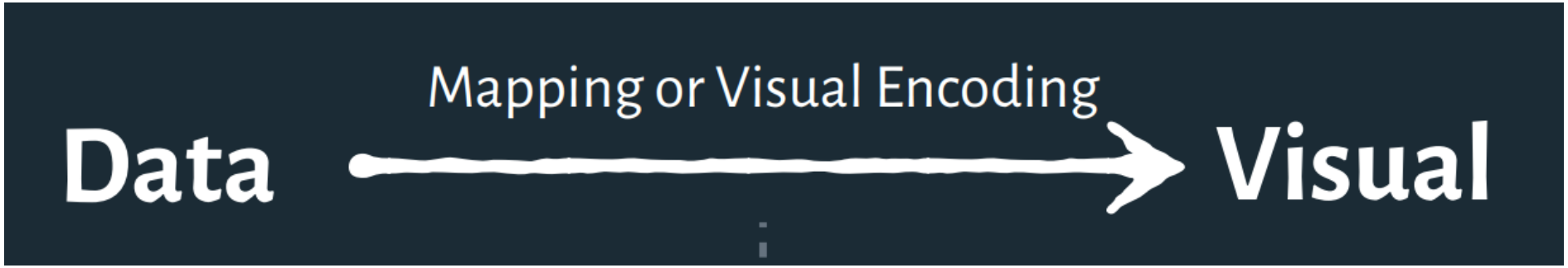
CONTINUE

4. Dimensionality Reduction

- Dimensionality reduction *involves reducing the number of variables* used to represent the data.
- This is often done to simplify the data and improve visual clarity.
- **E.g.:** A line graph that shows the trend of stock prices over time can be simplified by only showing the closing price of the stock at the end of each day, rather than including all of the intermediate price values throughout the day.

VISUAL ENCODINGS

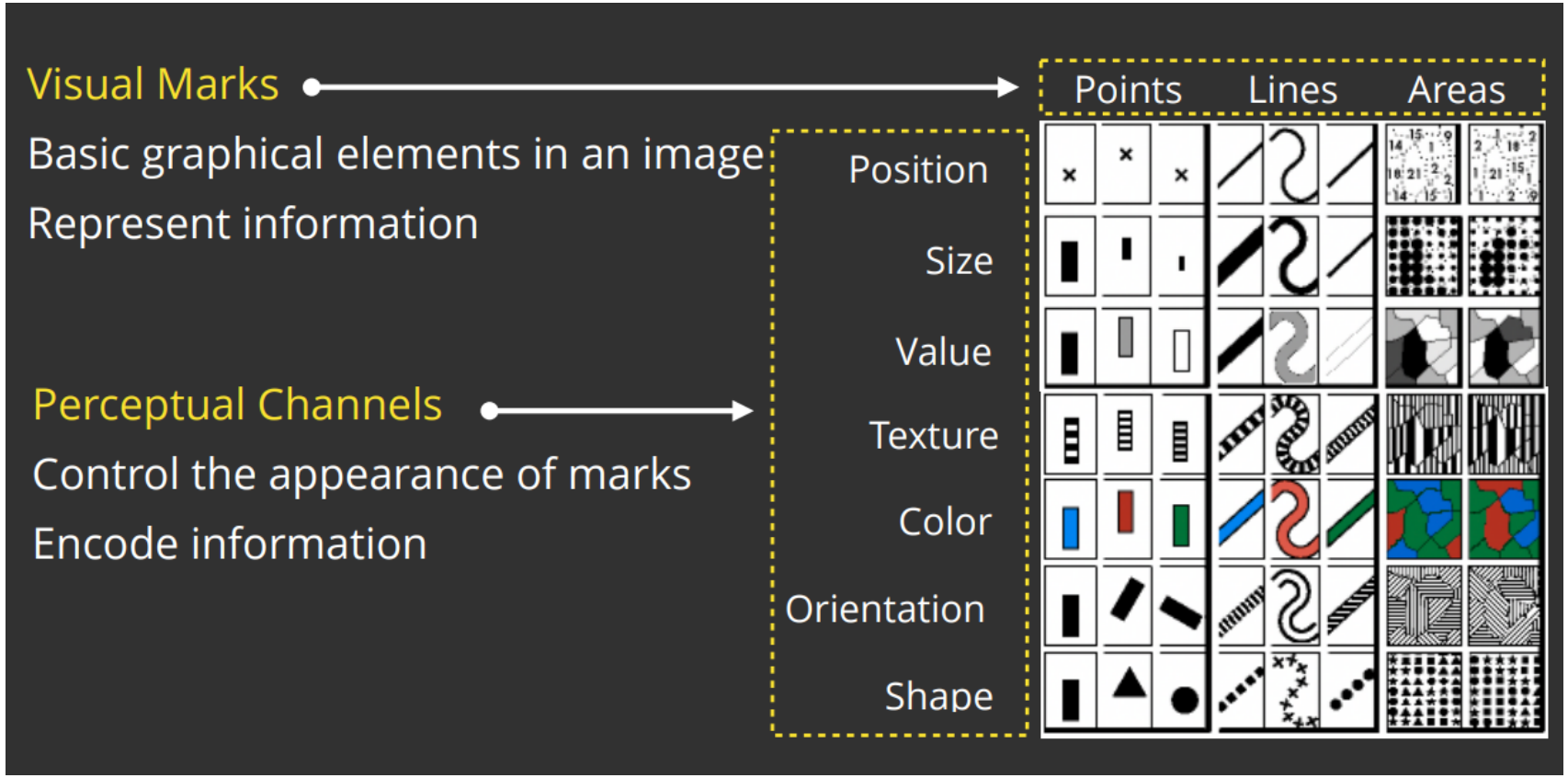
- Visual encodings is the *process of representing data values in visual display* using visual cues or symbols.
- Process of converting data to visual is known as mapping or visual encoding.
- This can include different types of shapes, colors, sizes, and textures that are used to represent different types of data.



VISUAL ENCODINGS

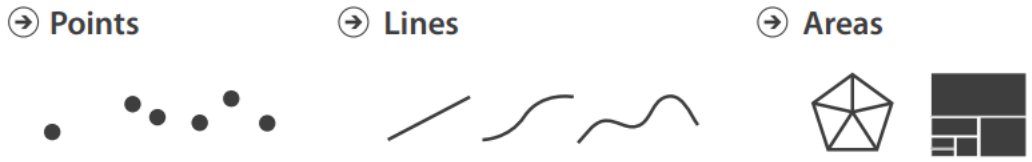
Visual encodings

Visual encodings can be described as combination of two aspects: graphical elements called marks and visual channels to control their appearance.

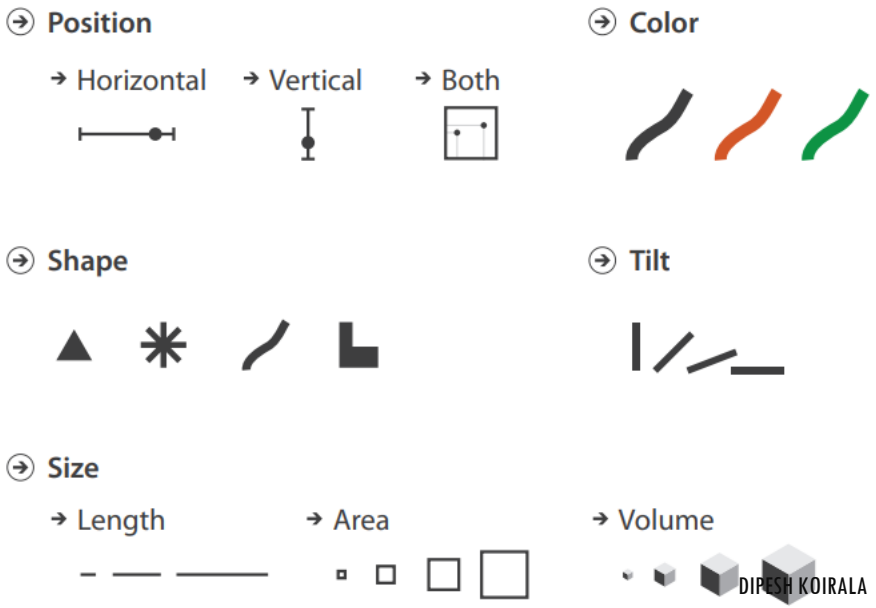


VISUAL ENCODINGS

- A **mark** is a basic graphical element in an image.

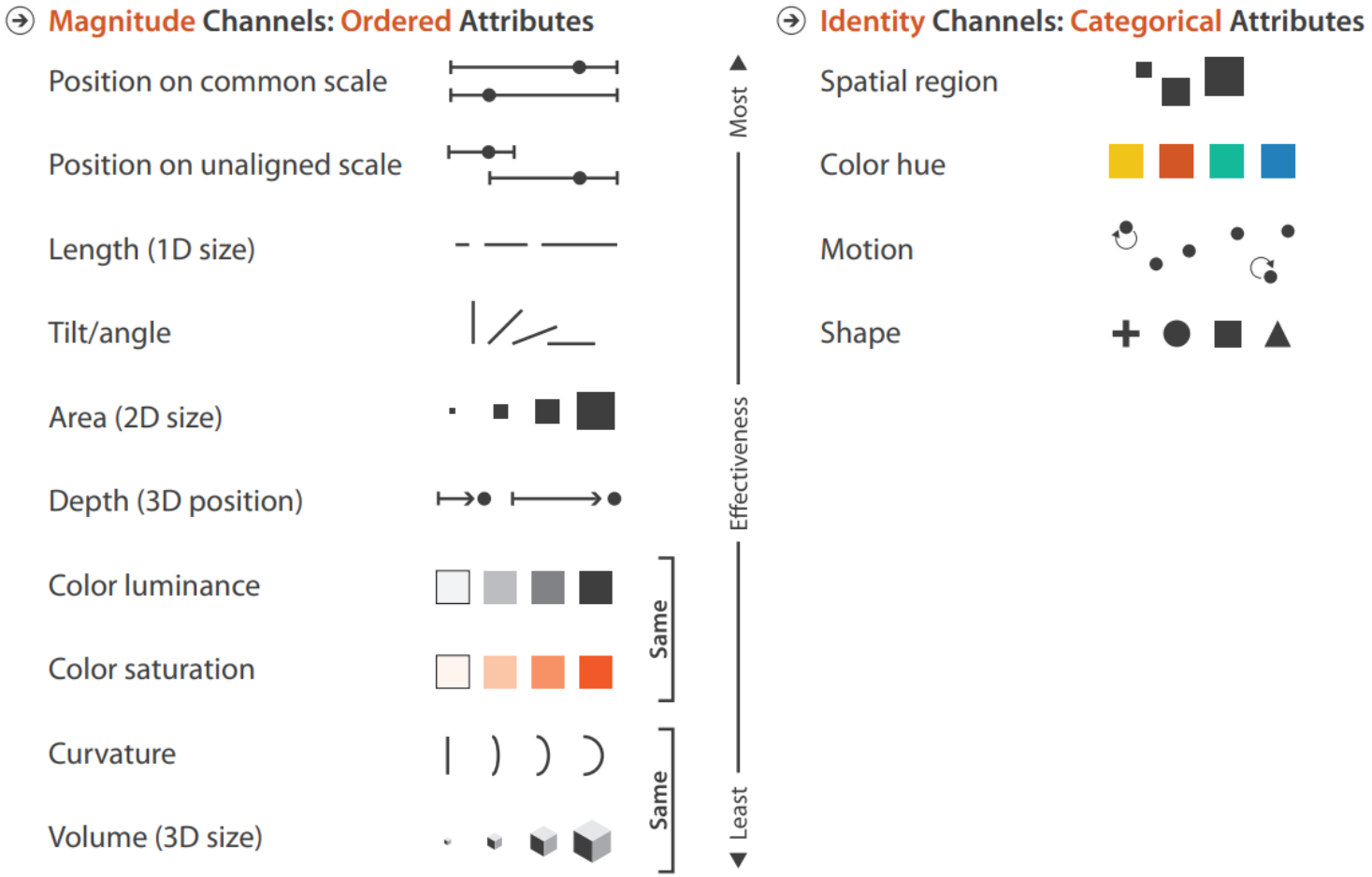


- A **visual channel** is a way to control the appearance of marks



VISUAL ENCODINGS

Magnitude Channels and Identity Channels



PRINCIPLES OF VISUAL ENCODINGS

Expressiveness Principle

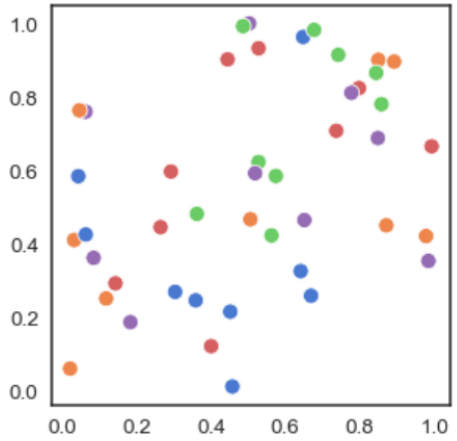
- Visual encodings should *express all of, and only,* the information in the dataset attributes.

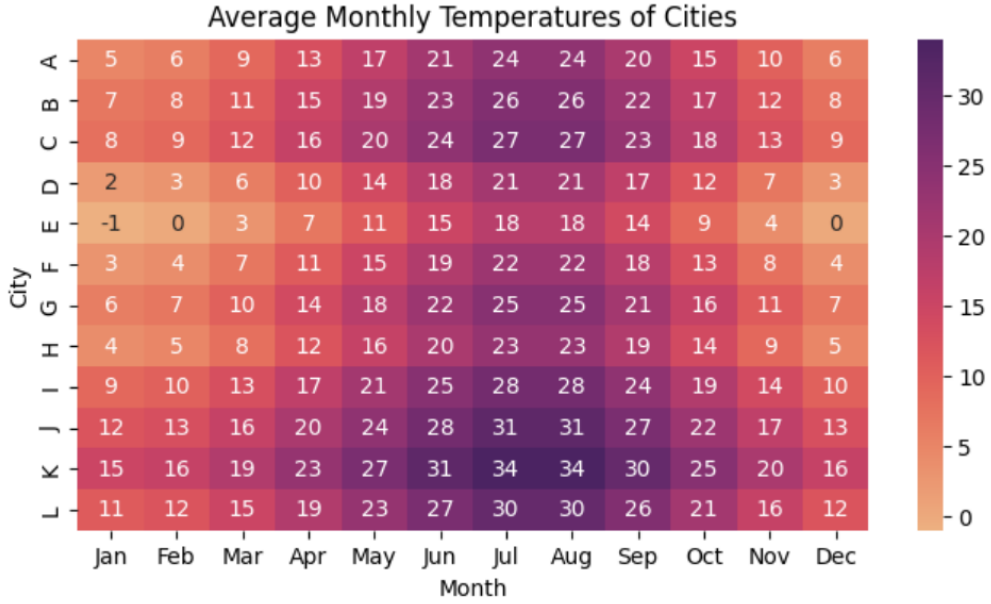
Effectiveness Principle

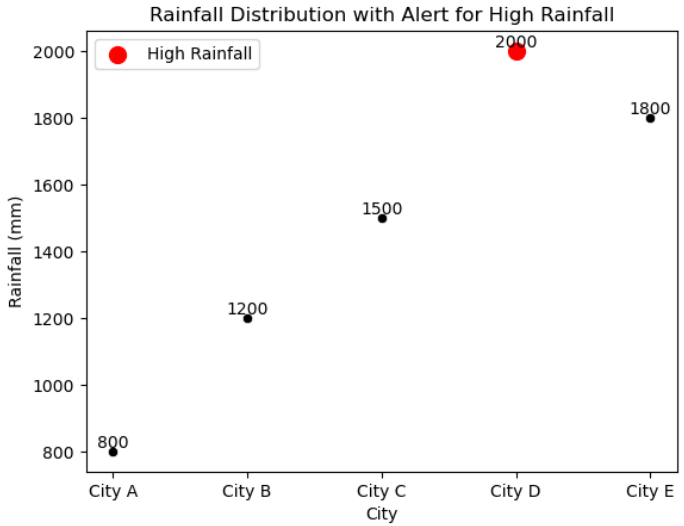
- A visualization is more effective than another one if the information conveyed *is more readily perceived.*

USE OF COLORS

Focus here







USE OF COLORS

Used to :

i. draw attention of the reader
ii. highlight portion of data
iii. distinguish between different categories

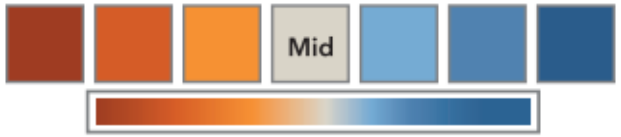
- Color should be used in data visualization in three primary ways: *sequential, diverging, and categorical.*
- In addition, there is often the need to highlight data or alert the reader of something important.

USE OF COLORS

The big book of dashboards

SEQUENTIAL
color is ordered from low to high

DIVERGING
two sequential colors with a neutral midpoint



CATEGORICAL
contrasting colors for individual comparison

HIGHLIGHT
color used to highlight something

ALERT
color used to alert or warn reader

PERCEPTUAL ISSUES

When using color in data visualization, there are some important considerations to keep in mind:

Color blindness:

- Approximately 8% of men and 0.5% of women have some form of color blindness, which can affect their ability to distinguish between certain colors.

Consistency:

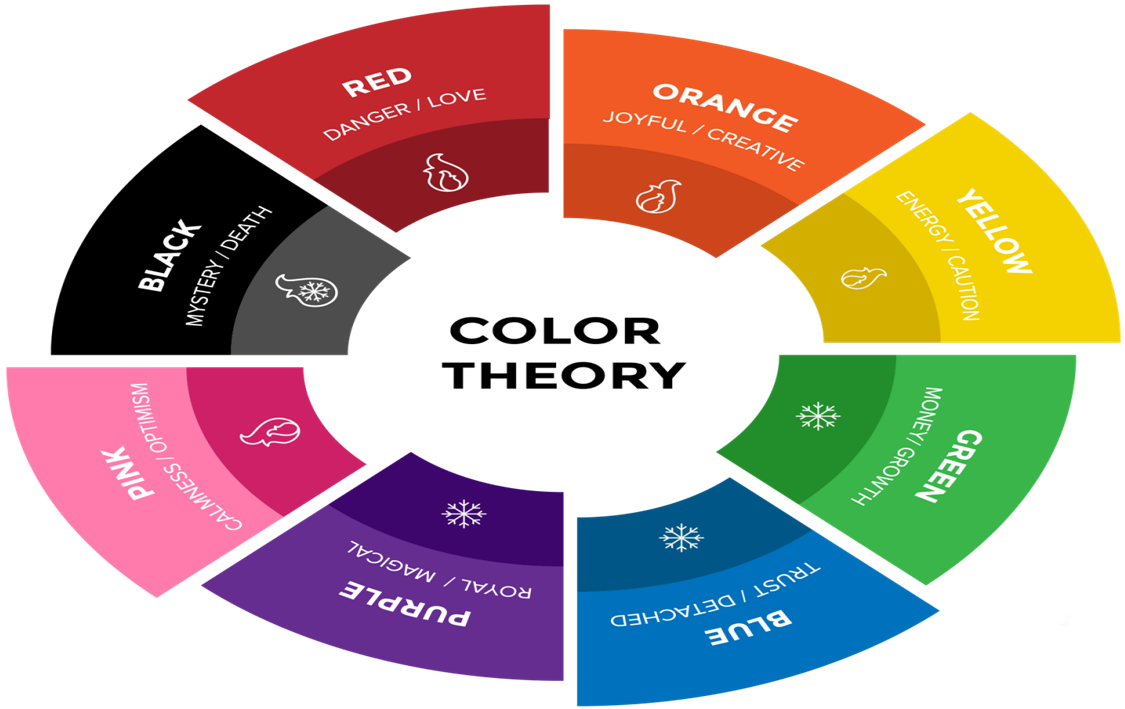
- Using consistent color schemes throughout a data visualization can make it easier for the viewer to understand the data and identify patterns or relationships.
- Using too many colors or changing color schemes can create .

Context:

- The use of color should always be considered in the context of the data being presented and the goals of the visualization.

USE OF COLORS

Context

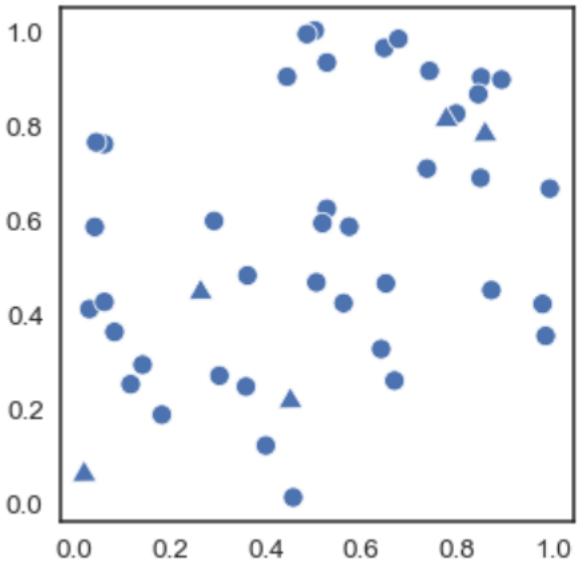


PERCEPTUAL ISSUES

- Perceptual issues can arise in visual representation when the visual cues used to represent data values *are not aligned with the way the human brain processes visual information*.

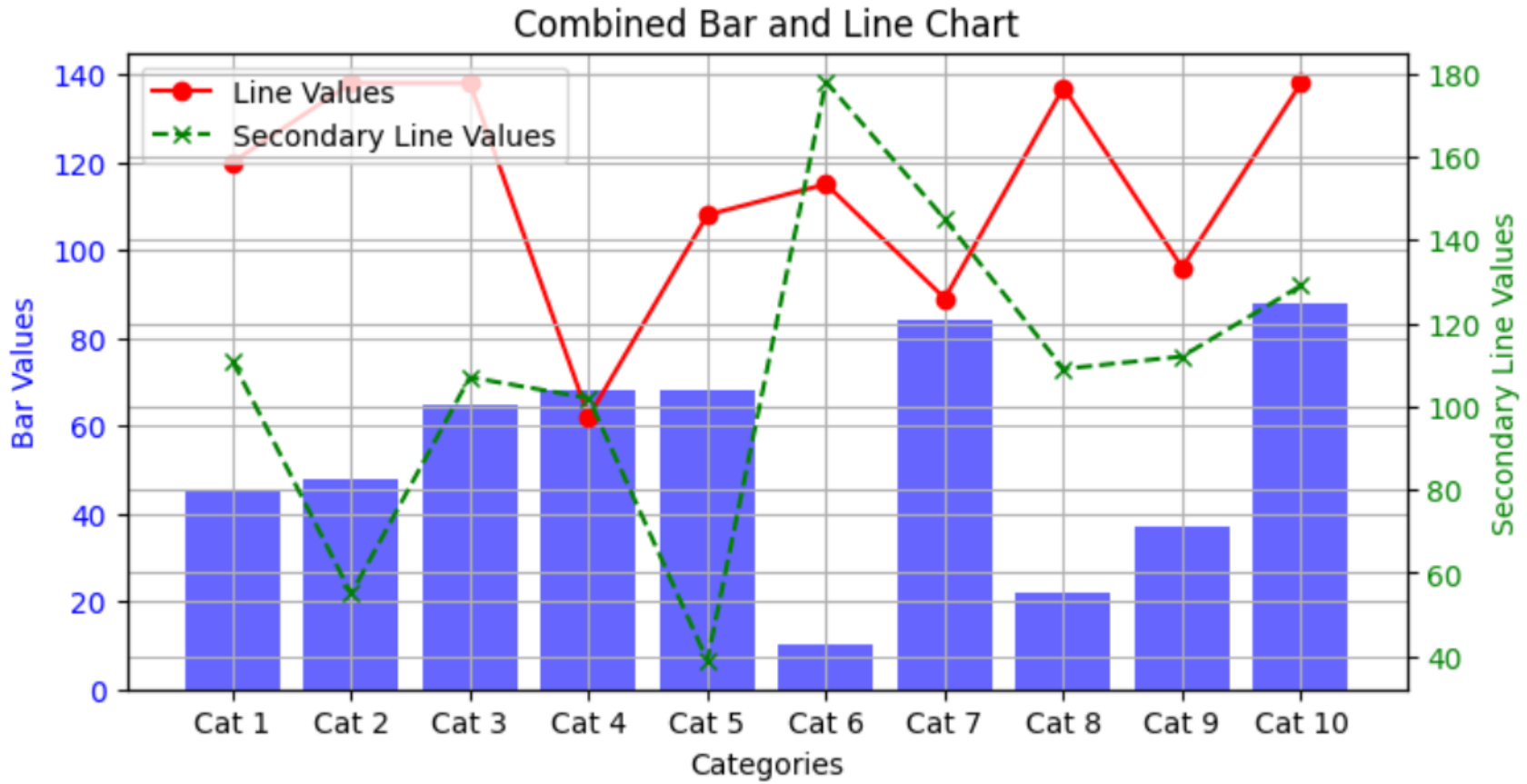
- To address perceptual issues, designers can use techniques such as:

  i. Gestalt Principles
  ii. Data scaling
  iii. Labeling



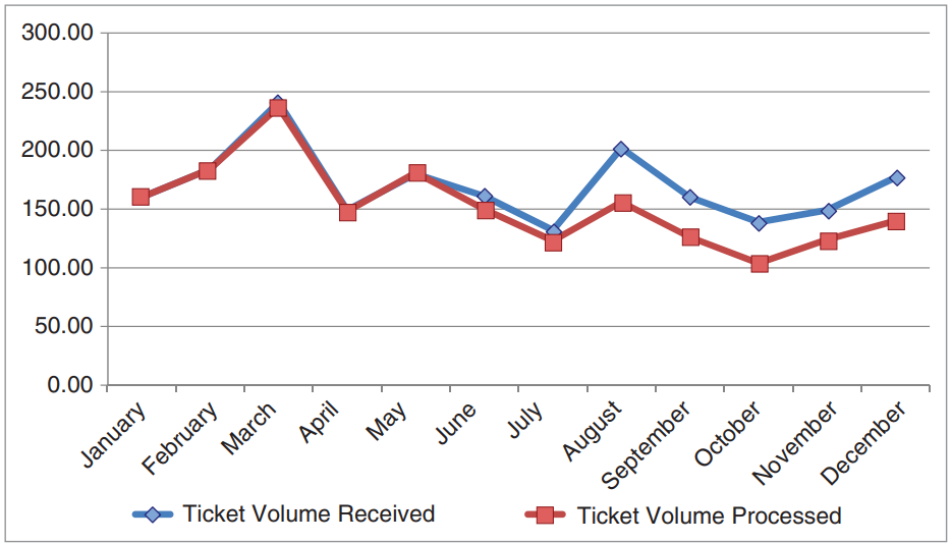
INFORMATION OVERLOADS

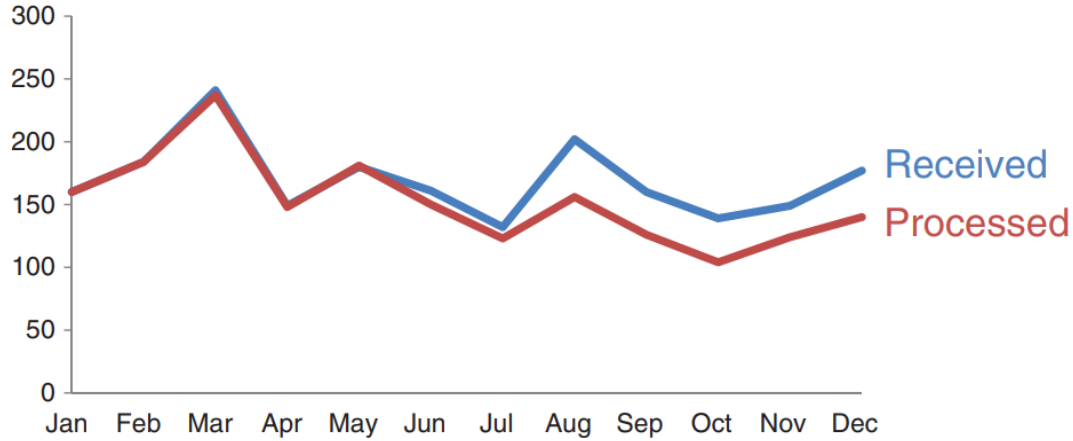
- Information overload occurs when a visual display presents *too much information* at once.



INFORMATION OVERLOAD

- Information overload occurs when a visual display presents *too much information at once*, making it difficult for users to effectively process and interpret the data.
- This can occur when the visual display *is too cluttered* or when too many variables are presented at once.





INFORMATION OVERLOAD

- When a visual display presents too much information at once, it can be difficult for users to *identify patterns and relationships*, make informed decisions, or *draw meaningful conclusions.*

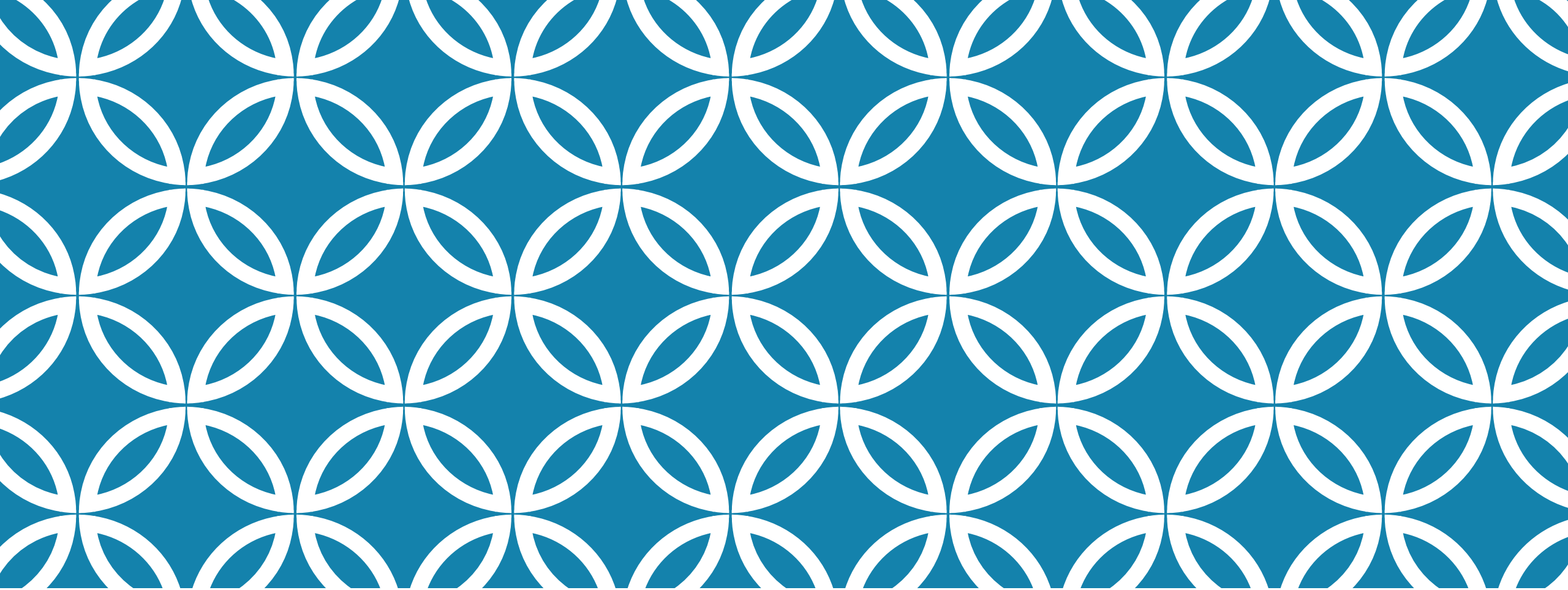
- Techniques to avoid information overload

    i. Data abstraction

    ii. Filtering

    iii. Hierarchical organization

    iv. Interactive visualizations



THANK YOU..