# MDS651
# Unit 2 – Visual Representations

Dipesh Koirala

# Outline

- Visualization reference model
- Visual mapping
- Visual analytics
- Design of visualization applications

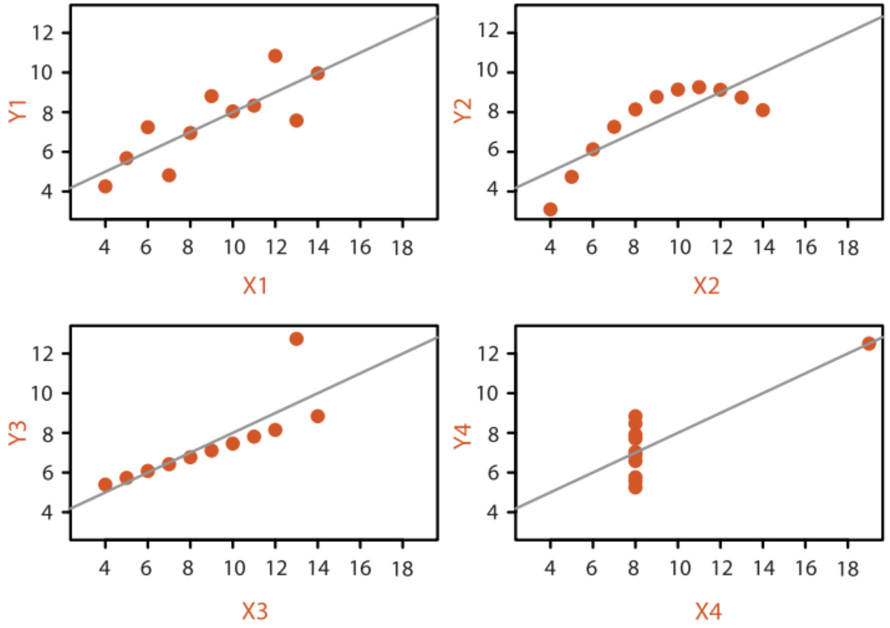
# Visualization

- Visualization is the process of starting with data and generating visuals from the data.

**Anscombe's Quartet: Raw Data**

| | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| | X | Y | X | Y | X | Y | X | Y |
| | 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| | 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| | 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| | 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 7.11 | 8.0 | 8.84 |
| | 11.0 | 8.33 | 11.0 | 9.26 | 11.0 | 7.81 | 8.0 | 8.47 |
| | 14.0 | 9.96 | 14.0 | 8.10 | 14.0 | 8.84 | 8.0 | 7.04 |
| | 6.0 | 7.24 | 6.0 | 6.13 | 6.0 | 6.08 | 8.0 | 5.25 |
| | 4.0 | 4.26 | 4.0 | 3.10 | 4.0 | 5.39 | 19.0 | 12.50 |
| | 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15 | 8.0 | 5.56 |
| | 7.0 | 4.82 | 7.0 | 7.26 | 7.0 | 6.42 | 8.0 | 7.91 |
| | 5.0 | 5.68 | 5.0 | 4.74 | 5.0 | 5.73 | 8.0 | 6.89 |
| Mean | 9.0 | 7.5 | 9.0 | 7.5 | 9.0 | 7.5 | 9.0 | 7.5 |
| Variance | 10.0 | 3.75 | 10.0 | 3.75 | 10.0 | 3.75 | 10.0 | 3.75 |
| Correlation | 0.816 | | 0.816 | | 0.816 | | 0.816 | |

Difficult to understand, **then visualize..**



# Visualization reference model

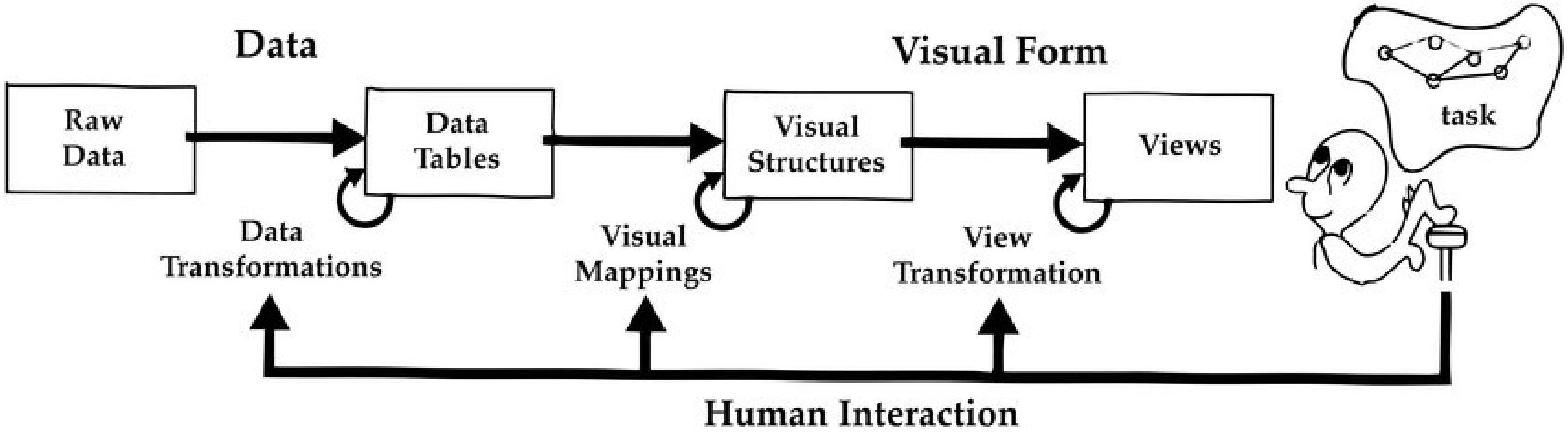
- Visualizations can be thought as adjustable mappings from data to visual form to the human perceiver.

- A visualization reference model is a conceptual framework that breaks down the process of creating visualizations into a series of steps or stages, helping to understand how data is transformed into visual representations.

- It provides *structured approach and various stages* for transforming raw data into insightful visualizations.

# Visualization reference model

- Card, Mackinlay, and Shneiderman's (CMS) Information Visualization Reference Model (1999)
- CMS model is a **framework** that illustrates the visualization process as a **pipeline** of transformations.




**Raw Data:** idiosyncratic formats
**Data Tables:** relations (cases by variables) + metadata
**Visual Structures:** spatial substrates + marks + graphical properties
**Views:** graphical parameters (position, scaling, clipping, ...)

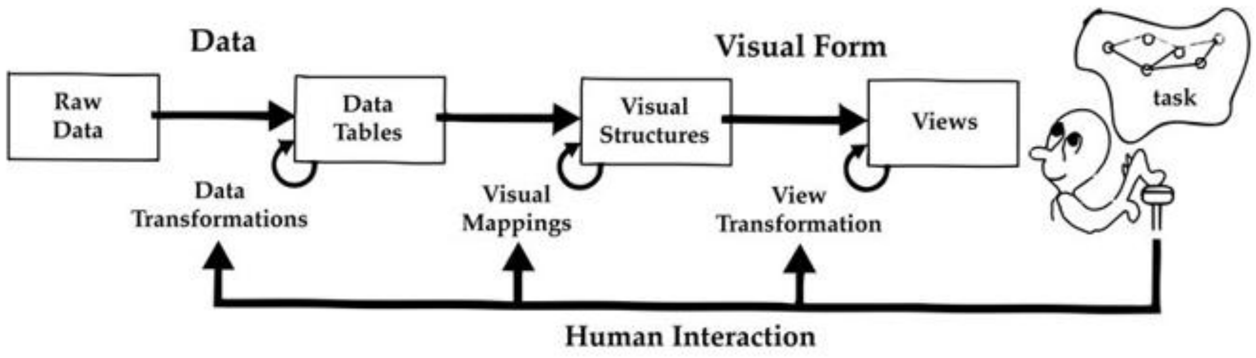
# Visualization reference model

- **From** :‘Readings in Information Visualization: Using Vision to Think’

# Visualization reference model

- In Figure, arrows flow from Raw Data on the left to the human, indicating a series of data transformations.
- Each arrow might indicate multiple chained transformations.

- Arrows flow from the human at the right into the transformations themselves, indicating the adjustment of these transformations by user-operated controls.



# Visualization reference model

## Raw Data

- The initial, unorganized data collected from various sources.
- Raw Data comes in many forms, from spreadsheets to the text of novels.
- The usual strategy is to transform this data into a relation or set of relations that are more structured and thus easier to map to visual forms.

**E.g.,** A marketing company wants to create a visualization that shows the sales performance of a new products. The raw data stage would involve collecting sales data from various sources.

# Visualization reference model

## Data Transformation

- The starting point is to process the raw data into something usable by the visualization system. i.e structured form.
  - Made sure data are mapped to fundamental data types.
  - Data Cleaning : Handling missing values, errors in input, inconsistencies.
  - Data Aggregation : Summarizing data (e.g., calculating averages, sums).
  - Large data may require sampling, filtering, aggregation, or partitioning

- The output of this stage is often referred to as **Data Tables.**

# Visualization reference model

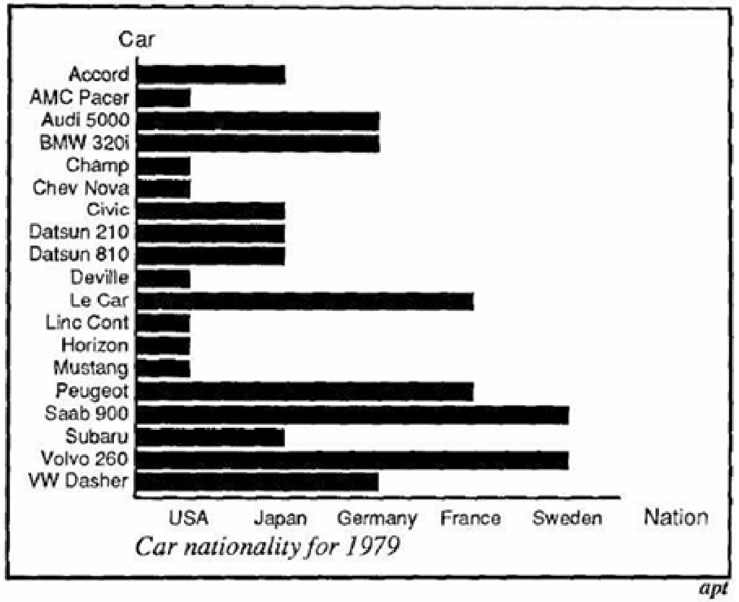
## Visual Mappings

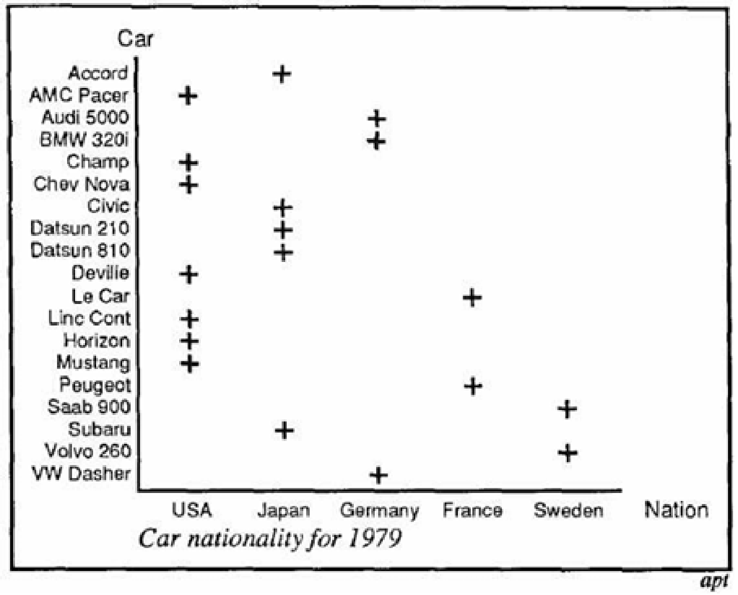
- processed data attributes are translated into visual elements and their properties (visual channel)
  - Spatial Substrate: Determining the overall layout and dimensions of the visualization (e.g., axes in a scatter plot, regions on a map).
  - Marks: Choosing the fundamental graphical primitives to represent data items (e.g., points, lines, areas).
  - Visual Channels: Assigning data attributes to visual characteristics of the marks (e.g., color, size, shape, orientation, texture).
- The output of this stage is a Visual Structure.

# Visualization reference model

## Visual Mappings

- It is easy to simply develop a nonsense visualization, or one that conveys the wrong information.



# Visualization reference model

## Visual Transformation

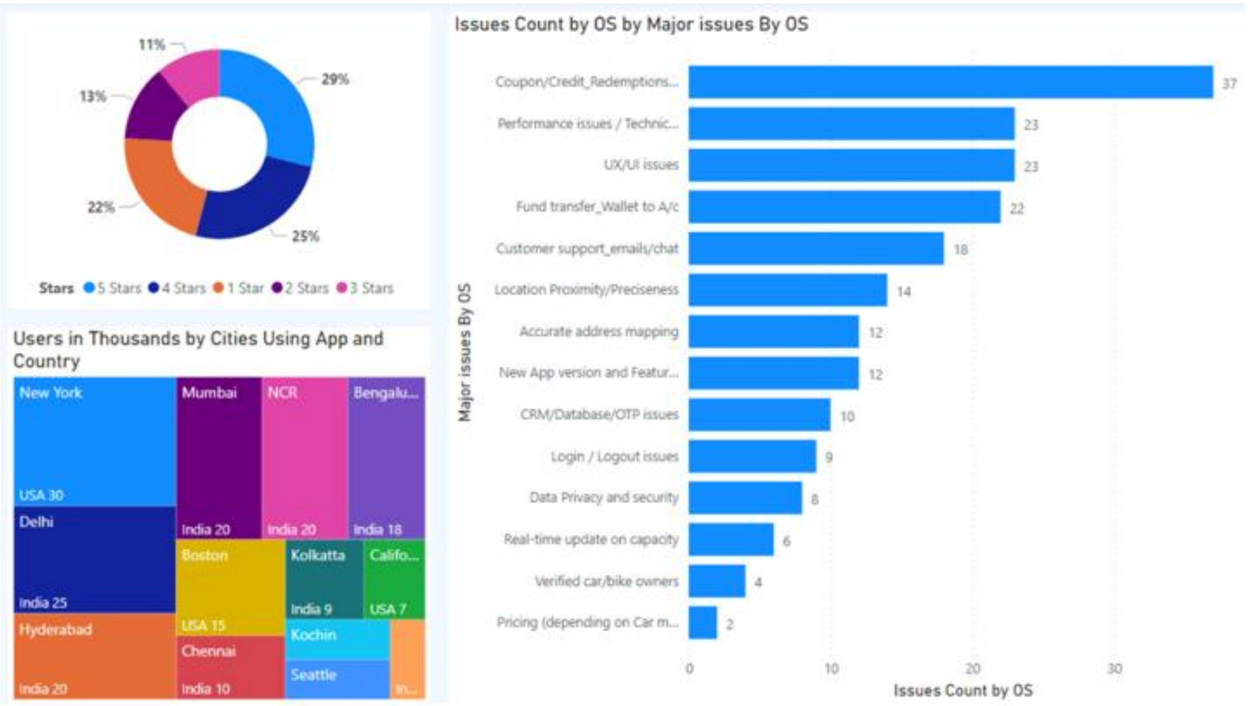
- The final stage involves mapping from geometry data to the image.
- We need to select the viewing parameters, shading technique if 3D, device transformations (for display, printers, ...).
  - Rendering: Displaying the visual structure on a screen or other medium.
  - Interaction: Providing controls for users to manipulate the view (e.g., zooming, panning, filtering, selecting, reordering).
- The output of this stage is a View.

# Visualization reference model

1. Raw Data
2. → Data Transformation (leading to Data Tables)
3. → Visual Mappings (leading to Visual Structures)
4. → View Transformation (leading to Views/Images)
5. ↔ Interaction (loops back to any stage)

# Visual Mappings

- **Visual mapping** (also called aesthetic mapping) refers to the process of systematically translating data variables into visual properties of a graphic.
- Visual mapping transform Data tables into Visual Structures, structures that combine spatial substrates, marks and graphical properties.



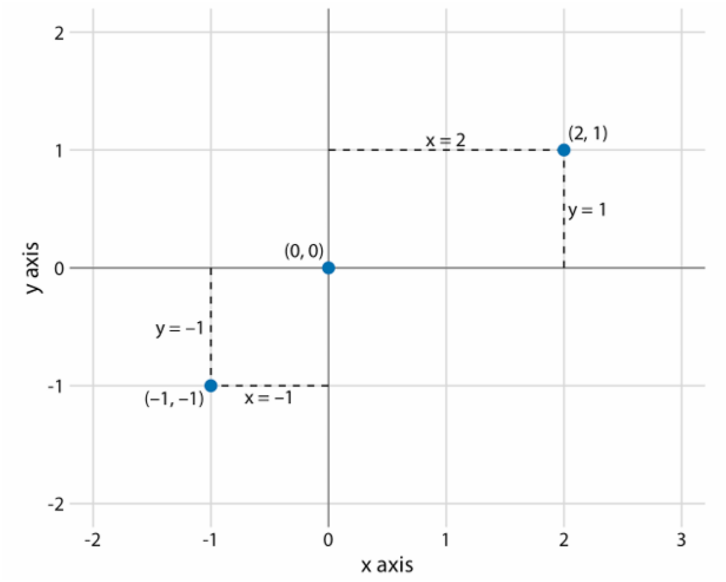
# Spatial Substrate

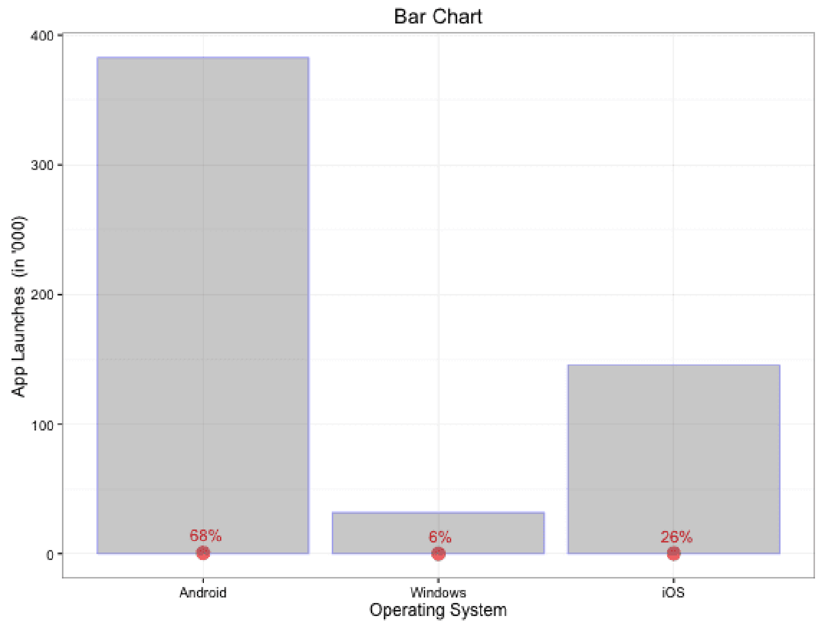
- a spatial substrate refers to **the underlying space or layouts** upon which visual elements are arranged to represent data
- it's the underlying canvas or coordinate system upon which graphical elements like points, lines, and bars are placed to encode information.
- makes foundation of a visualization, providing the context and structure for how data is presented spatially.
- space may be physical or conceptual

# Spatial Substrate

## Cartesian coordinate System:

- is a substrate that is defined by perpendicular X and Y axes (and Z in 3D). Data values are mapped to positions along the axes.
- E.g., Bar plots, line charts, scatter plot, histograms

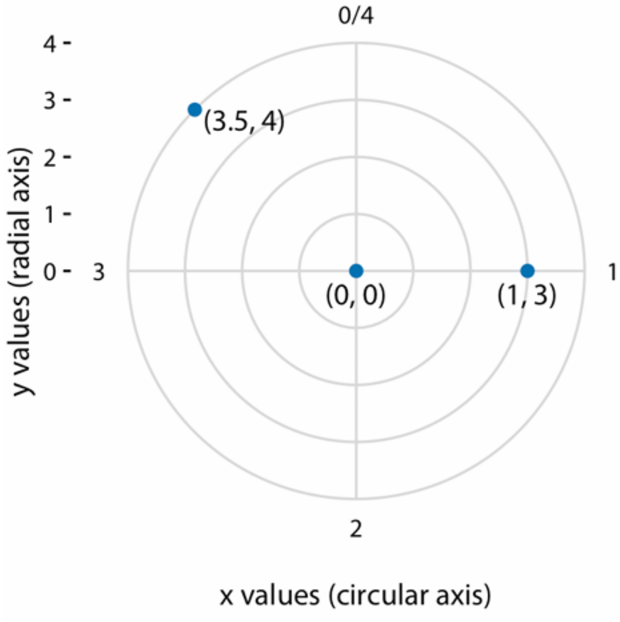


# Spatial Substrate

## Polar Coordinate System:

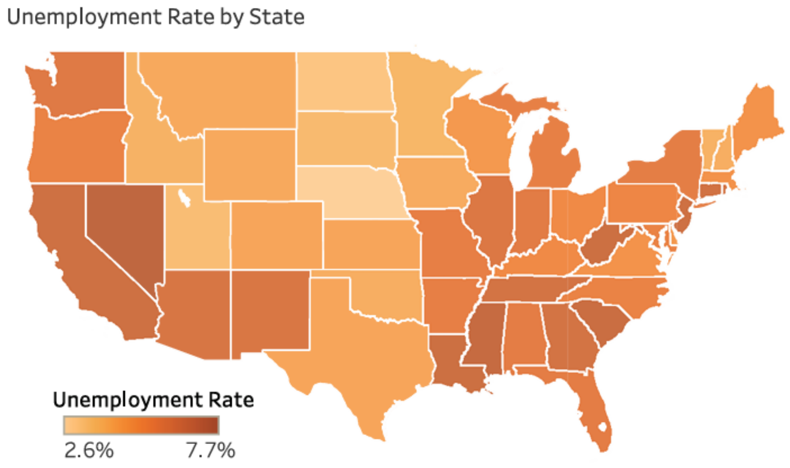
- Instead of X and Y axes, this system uses an **angle** and a **radius** to define positions from a central point (the origin).
- One variable is mapped to the angle (e.g., categories), and another variable is mapped to the radius (e.g., quantity, frequency).
- **E.g.,** Pie charts, Radial Bar Charts



# Spatial Substrate

## Geographic Coordinate System:

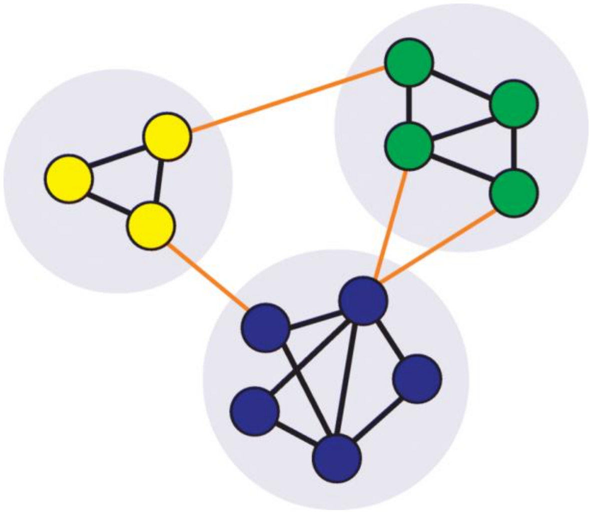
- is a specialized spatial substrate where positions correspond to real-world locations on the Earth's surface.
- It uses **latitude** and **longitude** (and sometimes altitude) to define points. Often, these spherical coordinates are projected onto a flat 2D plane using various map projections.
- Data values are associated with specific geographic points, regions, or lines.
- **E.g.,** Choropleth maps, point maps

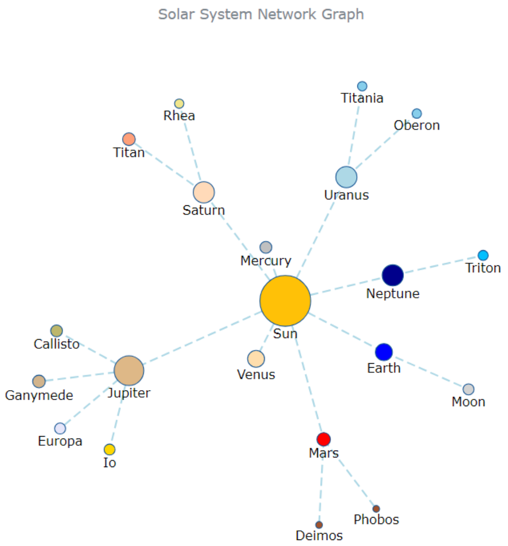


# Spatial Substrate

## Network/Graph Layouts:

- Is the substrate where space is formed by **nodes (entities)** and **edges (connections).**
- The axes is not present or have no fixed meaning and the position is relative.
- **E.g.,** Network graph for social networks, biological networks, transportation networks.



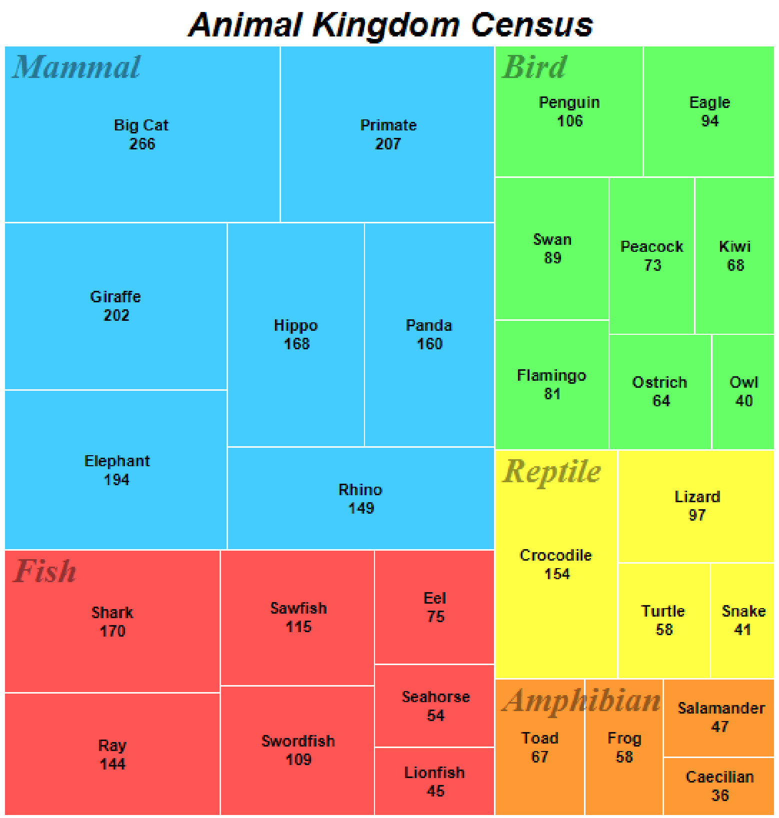


# Spatial Substrate

## Hierarchical Layouts

- is substrate that organize data based on hierarchical (parent-child) relationships.

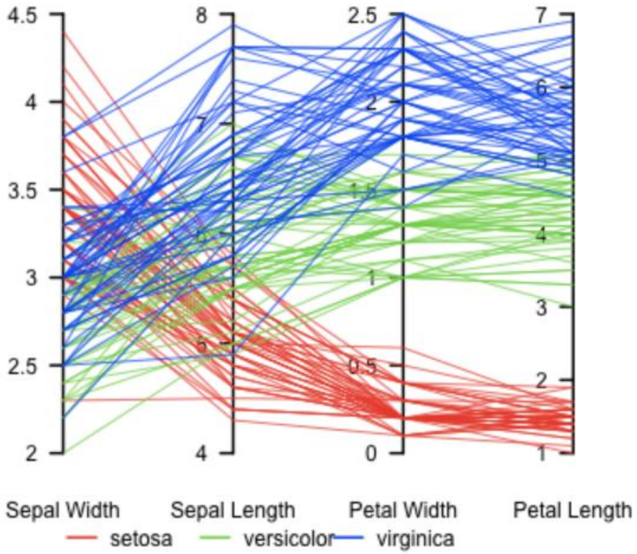
- The size of rectangular or angular segments typically encodes a quantitative value, while color can represent another attribute.
- **E.g.,** Tree maps



# Spatial Substrate

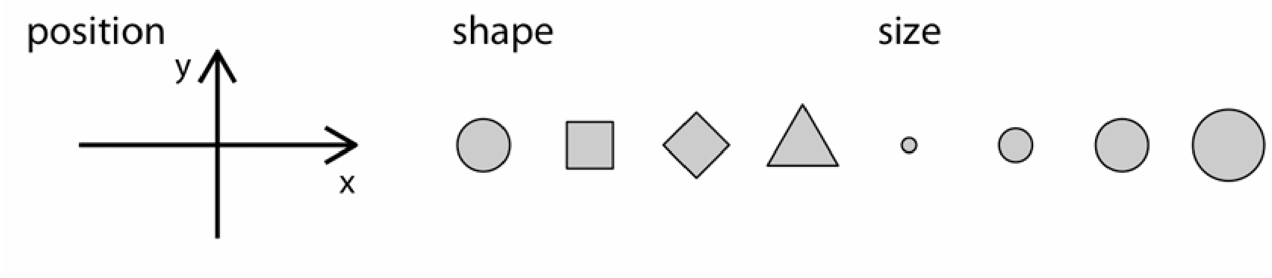
## Parallel Coordinates

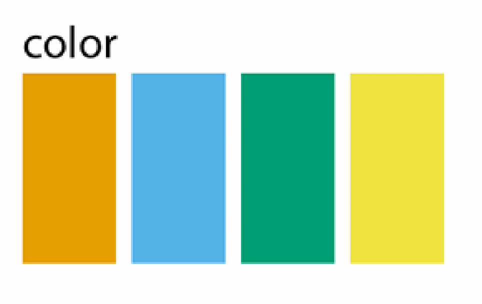
- is a unique substrate for high-dimensional data. Instead of orthogonal axes meeting at a point, axes are **parallel** to each other. Each data point is represented as a polyline that crosses each axis at a position corresponding to its value for that variable.
- Used to explore relationships and patterns across many variables simultaneously.
- **E.g.,** Parallel Coordinate Plot

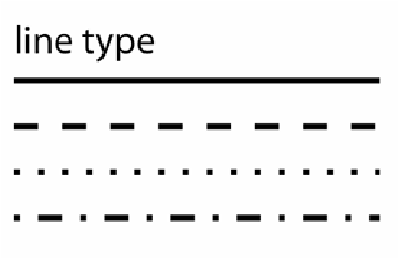


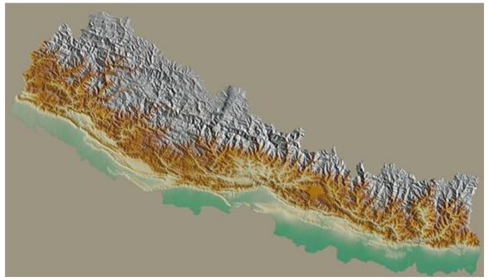
# Visual Mapping

- Position Mapping
- Shape Mapping
- Size Mapping
- Color Mapping
- Orientation Mapping
- Texture Mapping



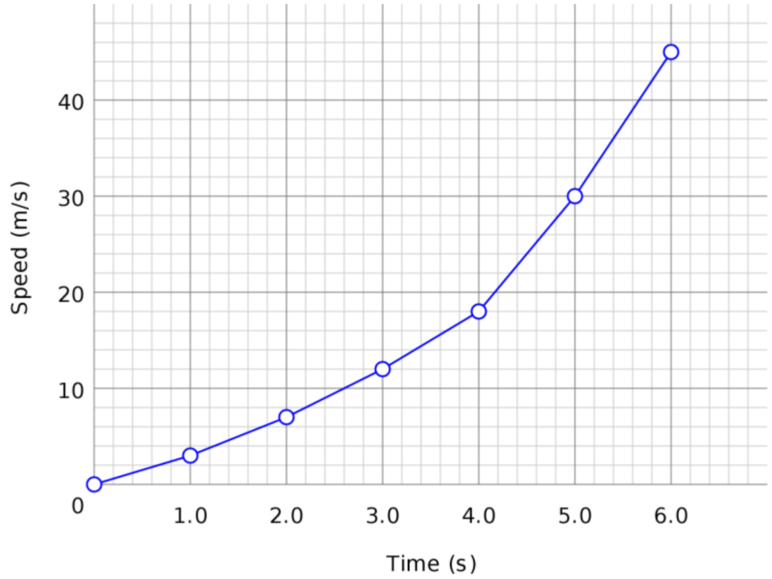



# Visual Mapping

- Visual Substrate: Position mapping
- Graphical Elements: Fundamental graphical primitives to represent data items (e.g., points, lines, areas).
- Graphical Properties: color mapping, size mapping, shape mapping, texture mapping

# Visual Mapping

- **Example:**




the graphical elements are points and lines

Here the spatial substrate is two dimensional

graphical properties are the size of points, length of lines, the orientation of the lines and the blue color.

# Visual Analytics

- is process of uncovering patterns, trends, and relationships that might otherwise remain hidden, by transforming raw data to data visualization.

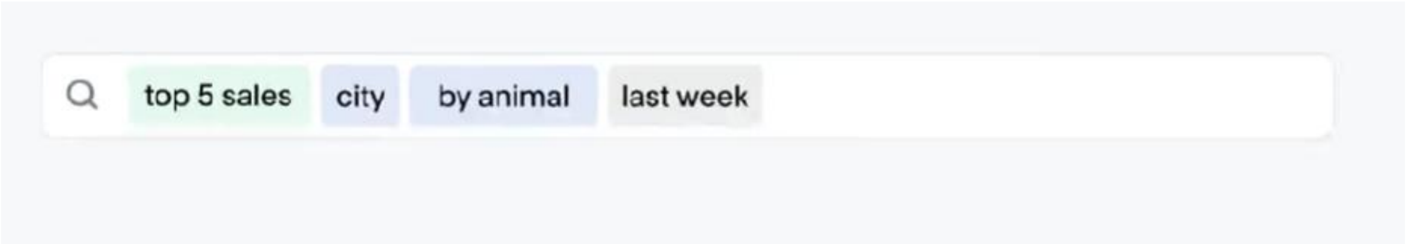
- is an interdisciplinary field that combines techniques from **data visualization, interactive interfaces, and analytical reasoning** to explore and understand complex datasets.

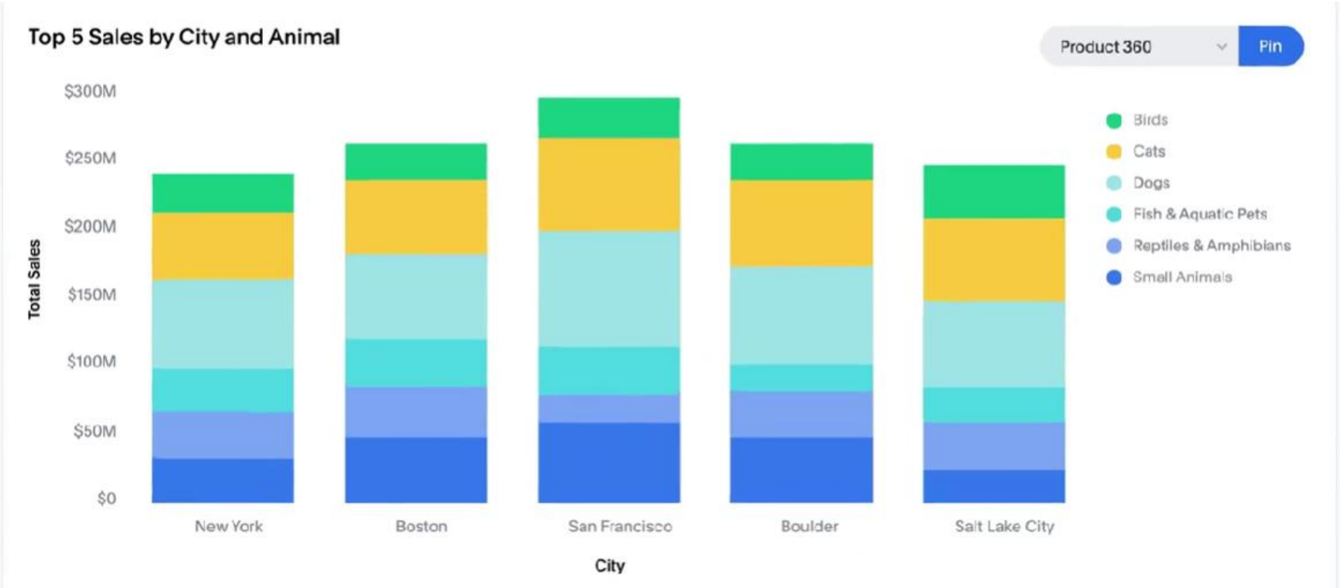
- Asking “what” and “why”
- Data visualizations present views of data that answer “what,” such as, “What are our sales and profits, for different regions, and different months or years?”

- Enhanced understanding
  - By presenting data in graphical formats, visual analytics helps to grasp trends, patterns, and insights more quickly than traditional data analysis methods.

# Visual Analytics

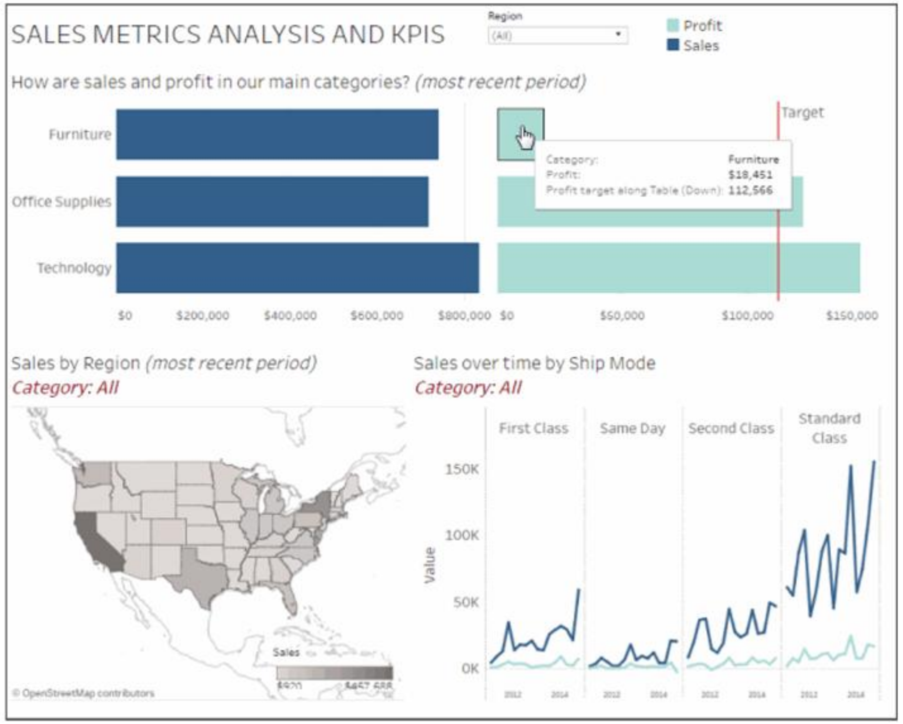
- Enhanced Understanding



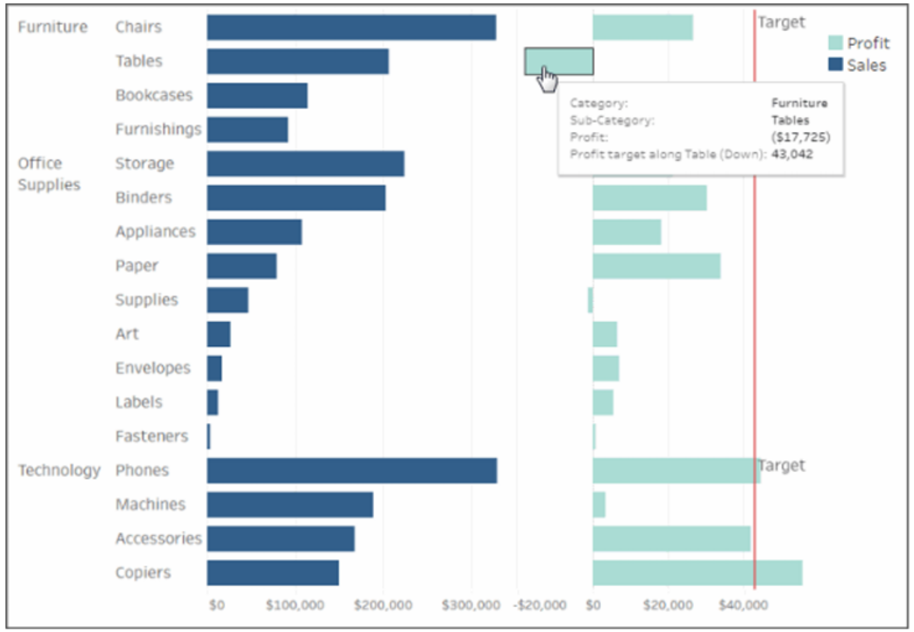
# Visual Analytics

- The **next step is asking "why."**
- Visual analytics can help to explore and find answers , handling both 'what' and 'why'
- **E.g.,** Let's start with the interactive sales report dashboard. It is a data visualization that answers a set of questions about sales and profits, for different regions.



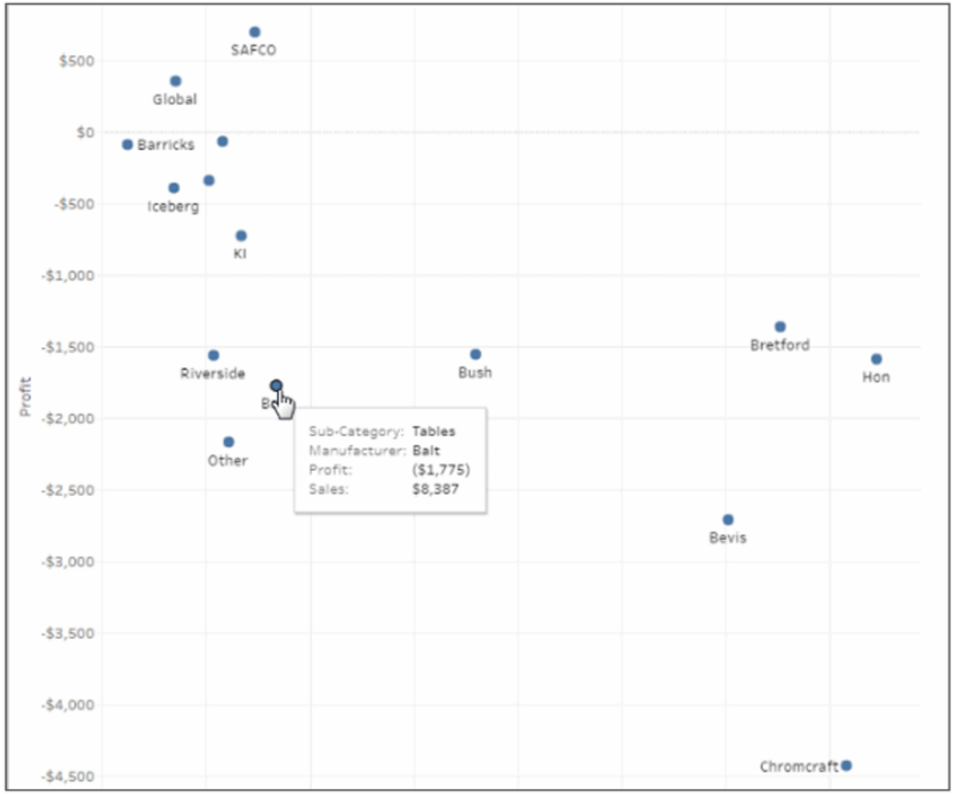
# Visual Analytics

- Furniture is not hitting the profit target, but looking at different regions isn't showing why.
- To explore why, a new view is created that shows sales and profits for each type of product included in Furniture.



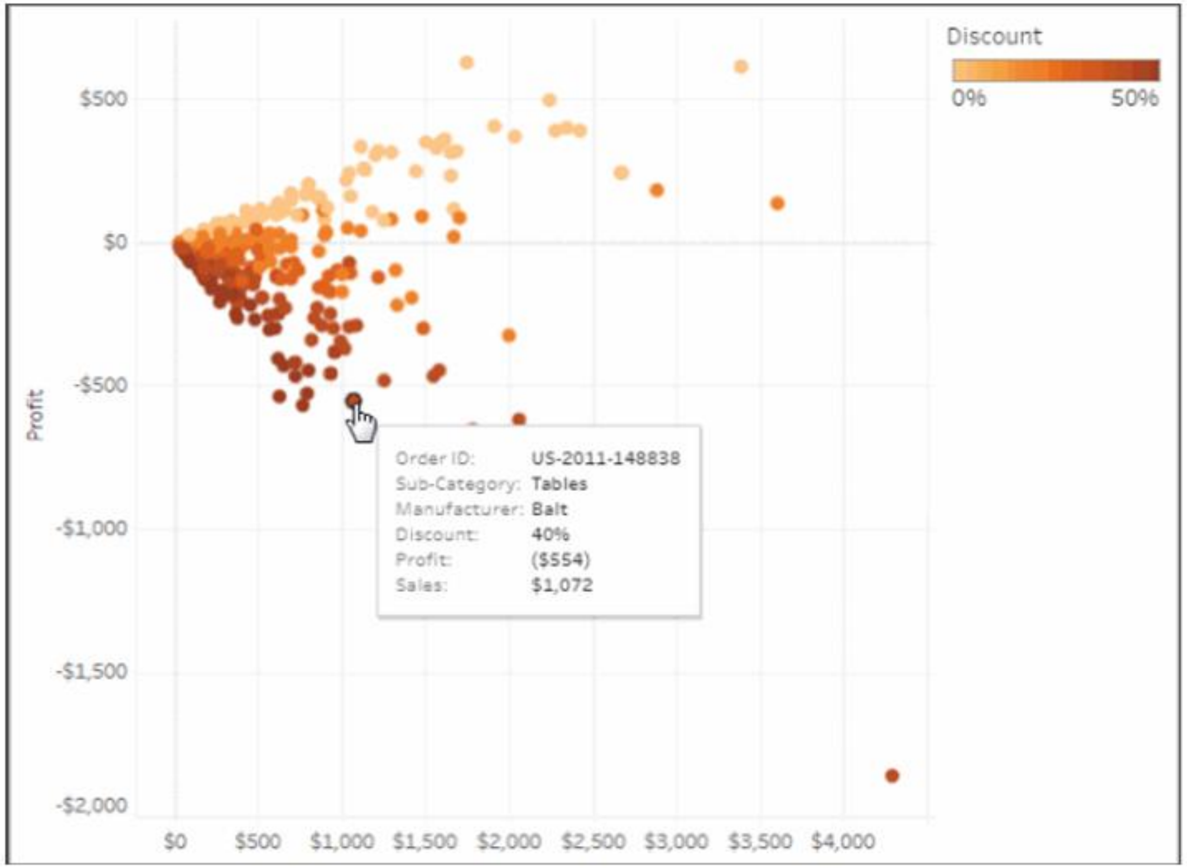
# Visual Analytics

- Now it is seen that **tables are not profitable.**
- To answer why, look at sales and profits for table manufacturers only.



# Visual Analytics

- Why are so many different brands of tables losing money? Discounts are often applied to tables. Change the view to see what level of discounts have been applied to every table sale order.



# Visual Analytics

- It looks like selling tables at a discount is creating lost profits.
- Source: [752750_core_why_visual_analytics_whitepaper_0.pdf]

# Visual Analytics

- So, what’s next?

# Visual Analytics

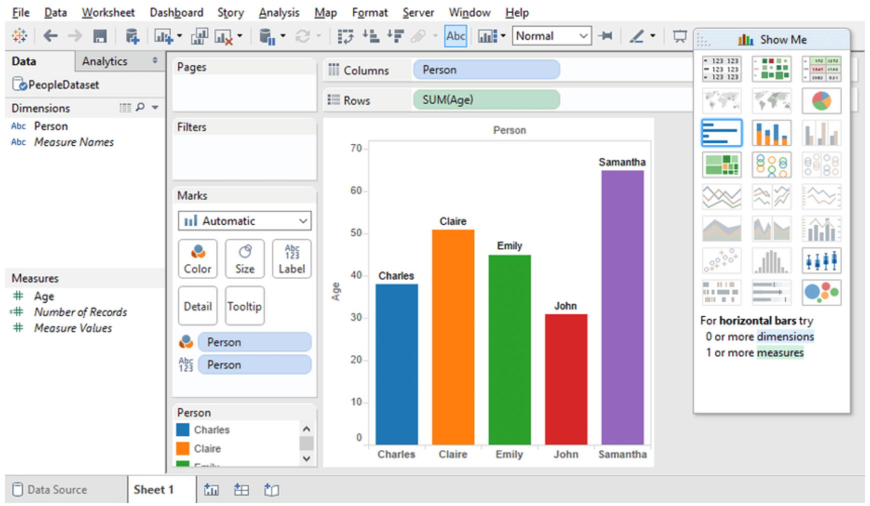
- Example:
- Pandemic Response During the COVID-19 pandemic, visual analytics played a crucial role in tracking the spread of the virus, analyzing vaccination rates, and predicting outbreak hotspots.
- Urban Planning and Smart Cities like Singapore and Barcelona use visual analytics to optimize traffic flow, resource management, and improve public services.


# Design of visualization applications

- The design of visualization applications involves *creating user interfaces that enable users to interact* with and explore data through visualizations.

- The goal of designing visualization applications is *to make the process of exploring and analyzing data as easy and intuitive as possible.*

- A well-designed visualization application should provide users with clear and informative visualizations, as well as tools to explore and manipulate the data.



# Design of visualization applications

**Key Consideration:**

- Data loading
- Visualization options
- Interactivity: features such as *zooming, panning, filtering, and sorting.*
- Customization: options for *changing colors, fonts, and other visual elements.*
- Collaboration: The application should enable users to collaborate and share their visualizations with others. This might include features such as *shared dashboards, commenting, and sharing options.*

# Design of visualization applications

**Some Applications:**

- Tableau
- Power BI
- D3.js
- Python Libraries

# References

- https://www.interaction-design.org/literature/topics/visual-mapping
- 20 Types of Charts And Graphs For Data Visualization
- Visual Analytics

# End of Unit 2

Thank you