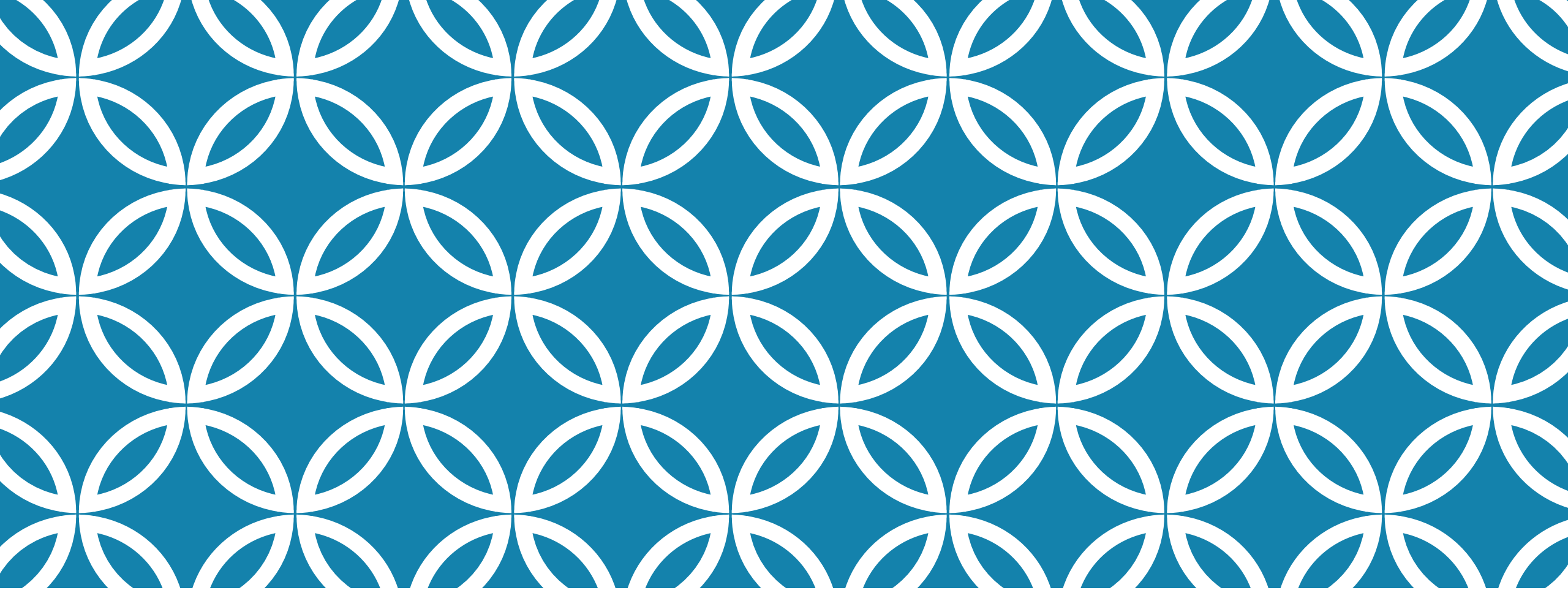


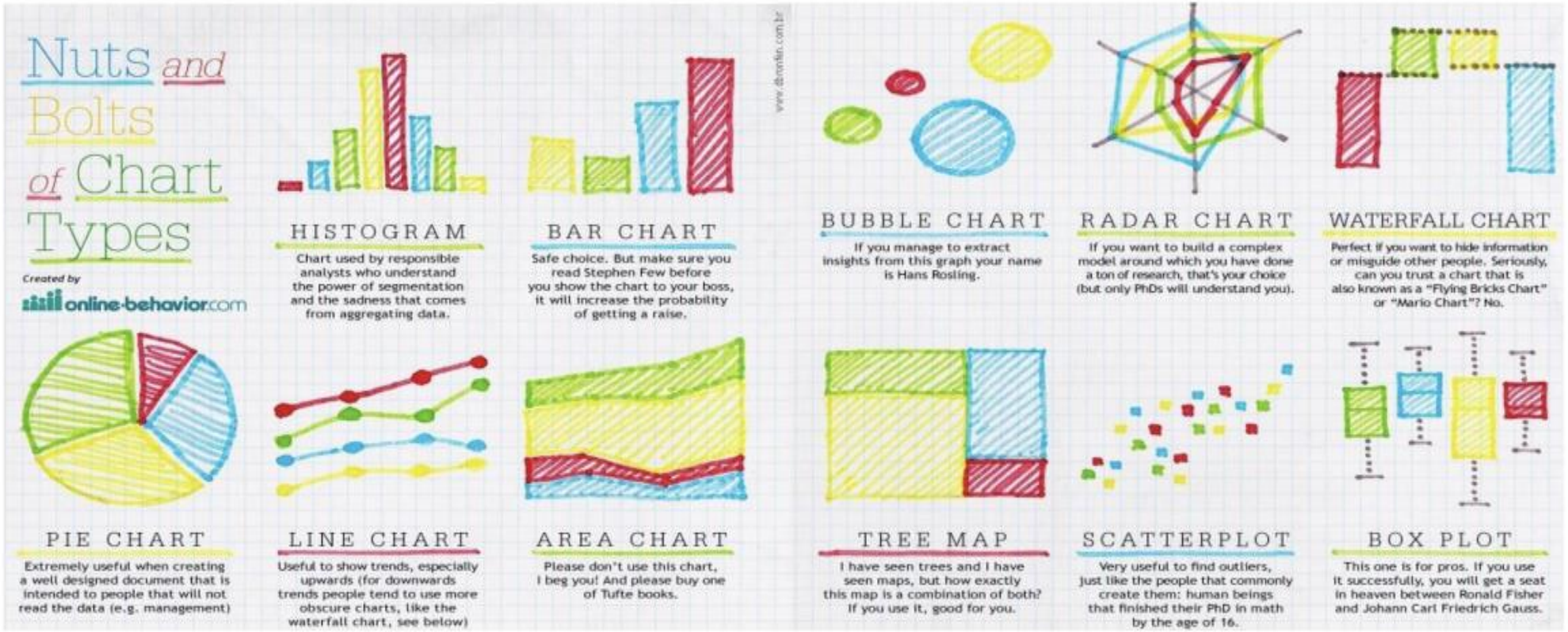
UNIT-3: ATTRIBUTE DATA VISUALIZATION

Dipesh Koirala

OUTLINE

- Visualization of one, two and multi-dimensional data, Tabular data, quantitative values (scatter plot)
- Separate, Order and align (Bar, staked bar, dots and line charts),
- Tree data, Displaying Hierarchical structures,
- Graph data, Rules for graph drawing and labeling
- Time series data, Characteristics of time data, Visualization time series data, Mapping of time

VISUAL REPRESENTATION



ATTRIBUTE DATA VISUALIZATION

- Is a way of *representing attributes of data through* visual elements such as graphs, charts and diagrams.

- Useful for displaying data that *has attributes or characteristics* such as categories, time periods and numerical values.
- Attribute data visualization is an essential tool for displaying data in a graphical format that enables easy interpretation and analysis.

- It can be categorical, numerical or time-series data.

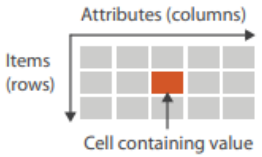
ATTRIBUTE DATA VISUALIZATION

Dataset Types

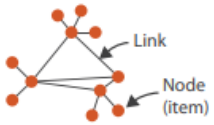
- A dataset is any collection of information *that is the target of analysis*.
- The four basic dataset types are tables, networks, fields, and geometry.

Dataset Types

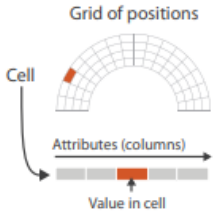
→ Tables



→ Networks

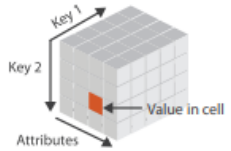


→ Fields (Continuous)



→ Geometry (Spatial)

→ *Multidimensional Table*



→ *Trees*

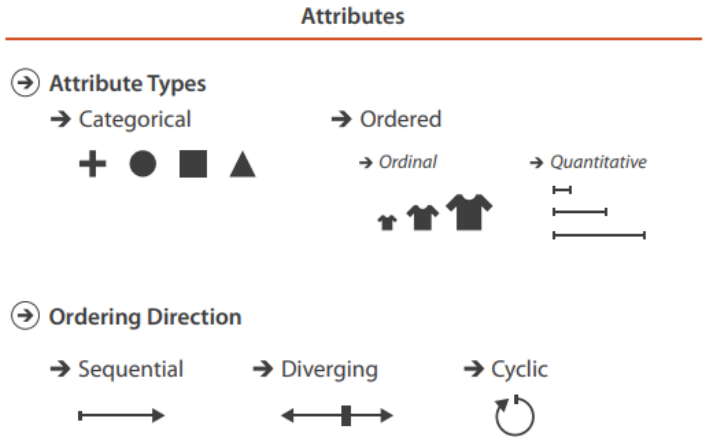
Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

ATTRIBUTE DATA VISUALIZATION

Data Types

- An item is an *individual entity* that is discrete, such as a row in a simple table or a node.
- An attribute is some specific property that can be measured, observed, or logged.
- For e.g., attributes could be *salary, price, number of sales, protein expression levels, or temperature.*



A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box		7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/6/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/6/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

attribute

item

cell

VISUALIZATION OF ONE, TWO AND MULTIDIMENSIONAL DATA

One Dimensional Data

- Given a one-dimensional sequence of univariate data or only one value per data item.
- E.g., temperature, age, sales of category etc.

```
# One-Dimensional Data

category = pd.Series(['A','B','A','A','B','A','C','A','B','A','C','A','B','A'])
```

```
[ ]  # One-Dimesnsional Data

     data = {
         'category': ['Ferrari','BMW','Buggati','Audi','Ford'],
         'sales': [200,250,350,100,80]
     }
```

VISUALIZATION OF ONE, TWO AND MULTIDIMENSIONAL DATA

One Dimensional Data

- Visualization Techniques:
  - Categorical: Bar Plots, Count Plots, Pie Charts (for few categories)
  - Continuous: Histograms, Density Plots, Box Plots
- These plots help to understand distribution, patterns, count within a single variable.

VISUALIZATION OF ONE, TWO AND MULTIDIMENSIONAL DATA

2-Dimensional Data

- When there is two variables present in data.
- E.g., height vs. weight, age vs. income, month vs. sales

- Visualization techniques:

  Scatter plots, bar plots, line plots

- These plot reveal relationships, correlations and outliers between two variables.

	carat	price
0	0.23	326
1	0.21	326
2	0.23	327
3	0.29	334
4	0.31	335
5	0.24	336
6	0.24	336
7	0.26	337
8	0.22	337
9	0.23	338

VISUALIZATION OF ONE, TWO AND MULTIDIMENSIONAL DATA

Multi-Dimensional Data

- When there is more than two variables present in data.
- **E.g.,** Temperature, humidity, wind speed and pressure

- Visualization Techniques:

  Bubble Chart

  Heat Map

- These plots allow to explore complex interactions among multiple variables.

TABULAR DATA

- Tabular data is data that is organized in rows and columns in a table format.

```
df.head()
```

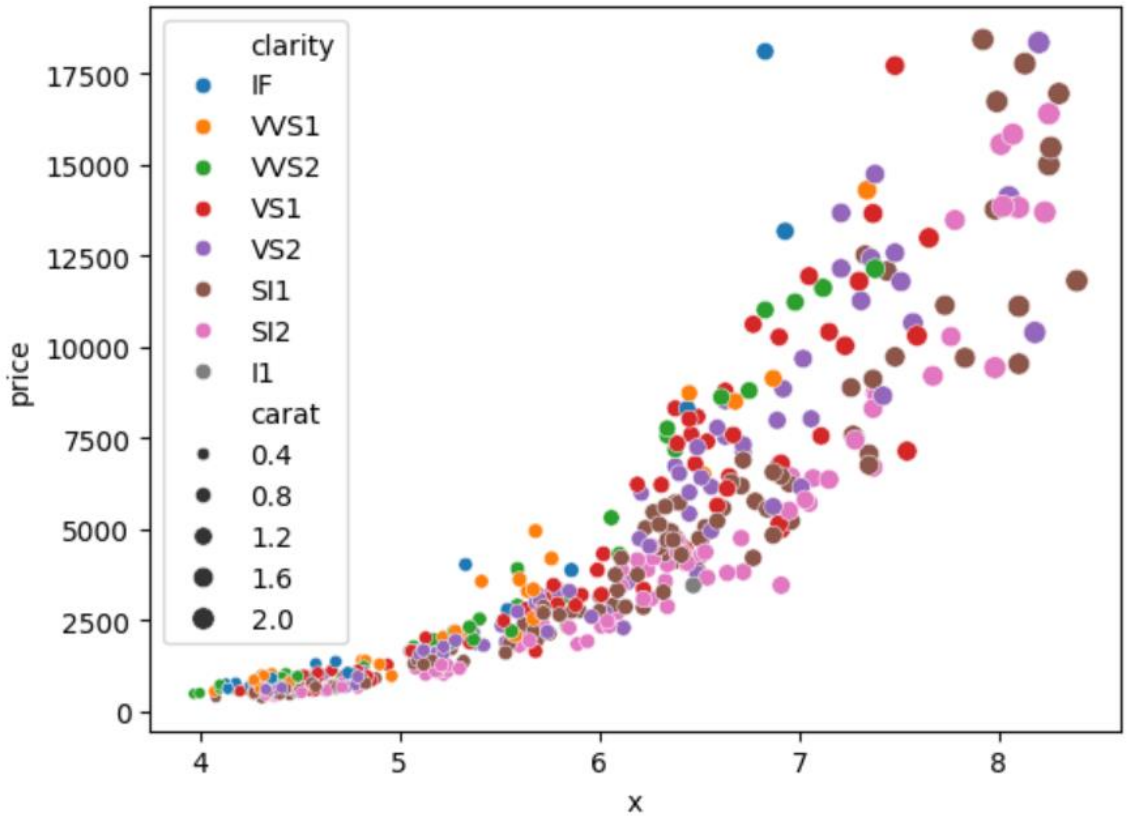
	carat	cut	color	clarity	depth	table	price	x	y	z
7980	0.84	Ideal	G	VS1	61.8	56.0	4325	6.02	6.11	3.75
36902	0.40	Ideal	F	VS1	61.4	57.0	960	4.74	4.77	2.92
1187	0.72	Ideal	F	VS2	61.8	59.0	2931	5.71	5.74	3.54
29546	0.38	Very Good	G	VS2	61.6	59.0	705	4.65	4.70	2.88
24980	1.75	Premium	D	SI2	60.6	60.0	13485	7.78	7.74	4.70

- Different types of plots can be applied in different attributes of column data.

QUANTITATIVE VALUES (SCATTER PLOT)

- Quantitative values are numerical values that can be *measured and analyzed mathematically.*

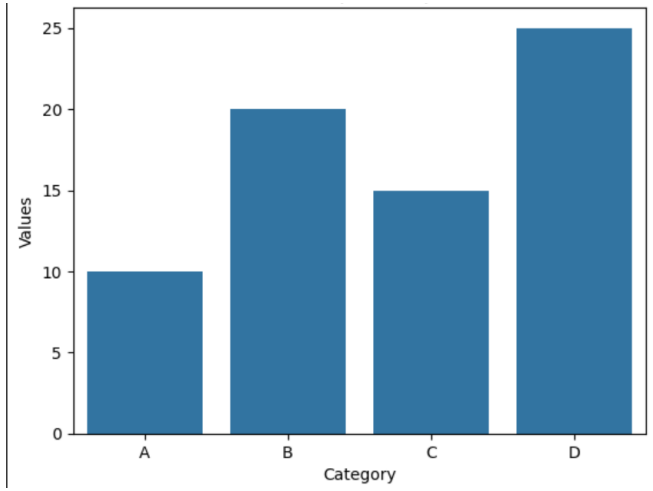
- In data visualization, *scatter plots are a useful way to visualize relationships* between two quantitative values.

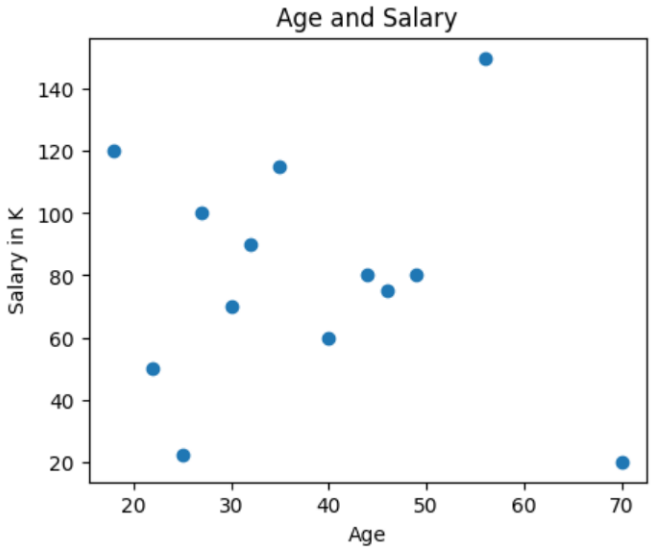


SEPARATE, ORDER AND ALIGN

Separate

- Practice of *distinguishing different data* groups or categories from each other.
- This can be done through the use of *color, shape, size, or spatial separation*.

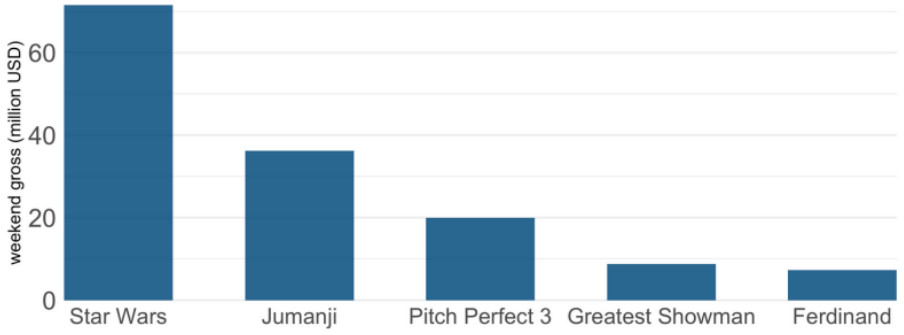


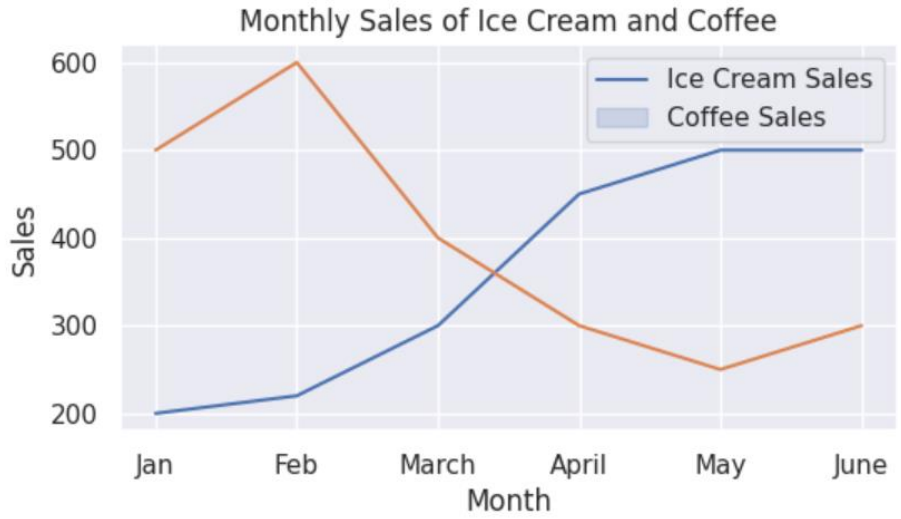


SEPARATE, ORDER AND ALIGN

Order

- Involves *arranging data in a specific sequence* to convey a meaningful pattern or relationship.
- This can be done in ascending or descending order, or according to some other logical progression (*e.g., chronological, alphabetical, or by magnitude*).

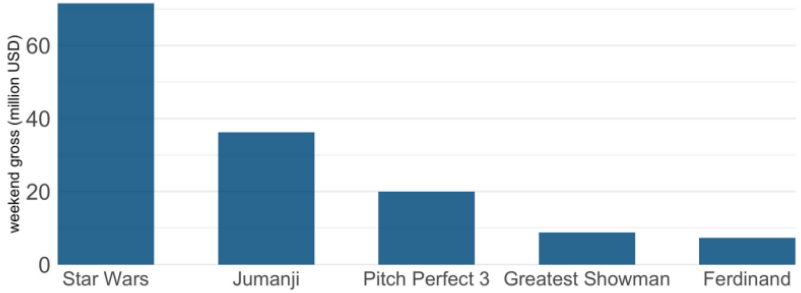


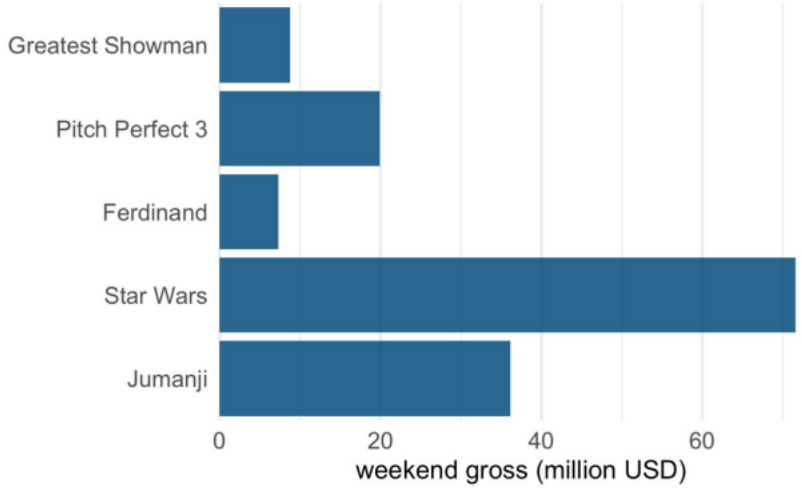


SEPARATE, ORDER AND ALIGN

Align

- Placing data points in a way that *they form a straight line, either horizontally or vertically.*
- Alignment helps in making comparisons and recognizing patterns or trends within the data more easily.

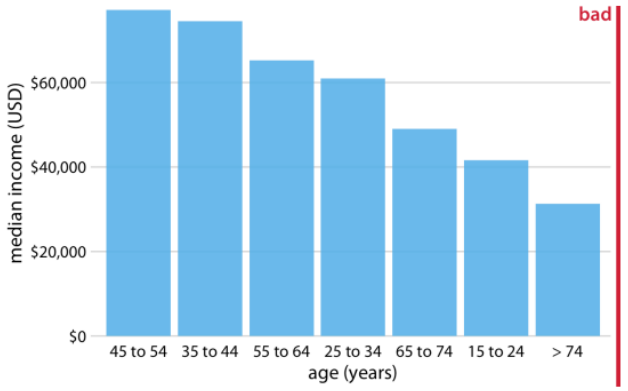


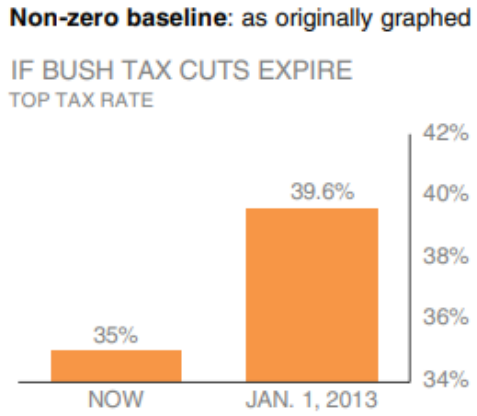


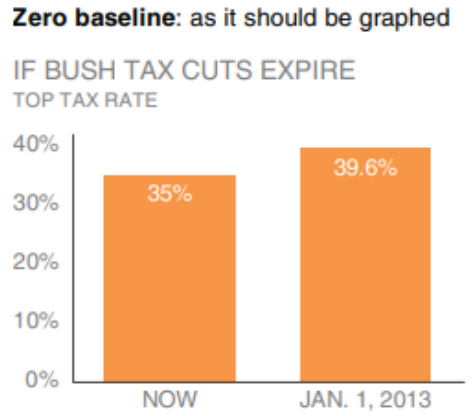
SEPARATE, ORDER AND ALIGN

- Separate helps to *distinguish different groups* of data.
- Order helps to *show progression or ranking* within the data.
- Align helps to *facilitate comparison* by placing data points along a common axis.

Identify Problem Here:







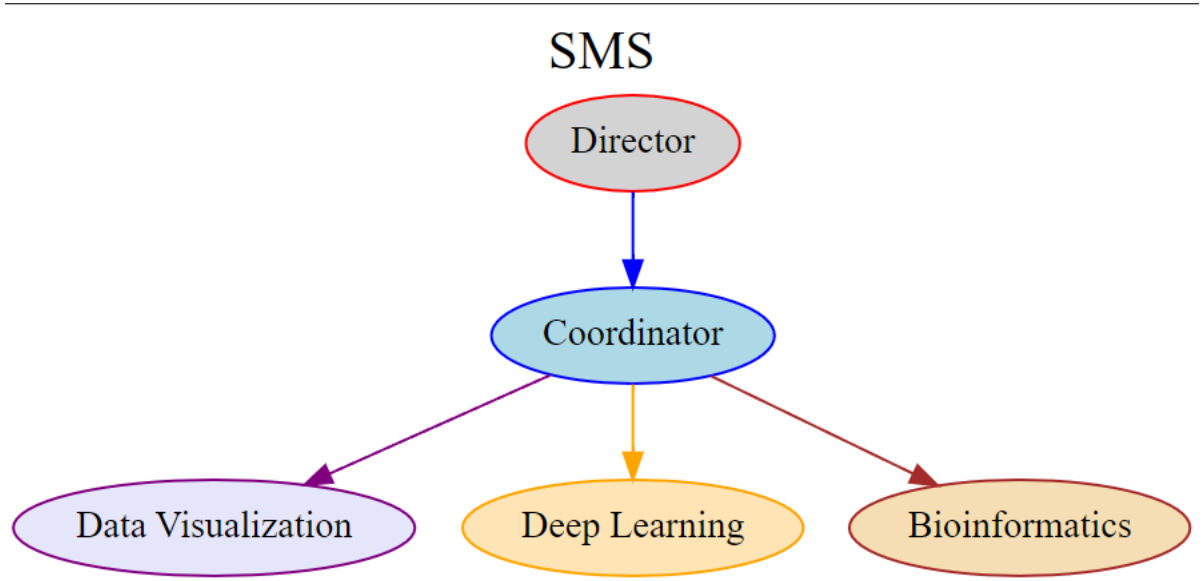
TREE DATA, DISPLAYING HIERARCHICAL STRUCTURES

- Important application of visualization is also the *conveying of relational information*, e.g., how data items are related to each other. These interrelationships can take many forms:
  - part/subpart, parent/child or other hierarchical relation
  - connectedness, such as cities connected by roads or computer connected by networks.

- *Tress or hierarchies* are one of the most common structures to hold relational information.

TREE DATA, DISPLAYING HIERARCHICAL STRUCTURES

- The most common visual encoding idiom for tree and network data is with *node-link diagrams.*
- Nodes are drawn *as point marks* and the links connecting them are drawn *as line marks.*
- This idiom uses connection marks to indicate the relationships between items.
- In addition to the connection marks, it uses *vertical spatial position channel* to show the depth in the tree.
- The horizontal spatial position of a node does not directly encode any attributes.



GRAPH DATA, RULES FOR GRAPH DRAWING AND LABELING

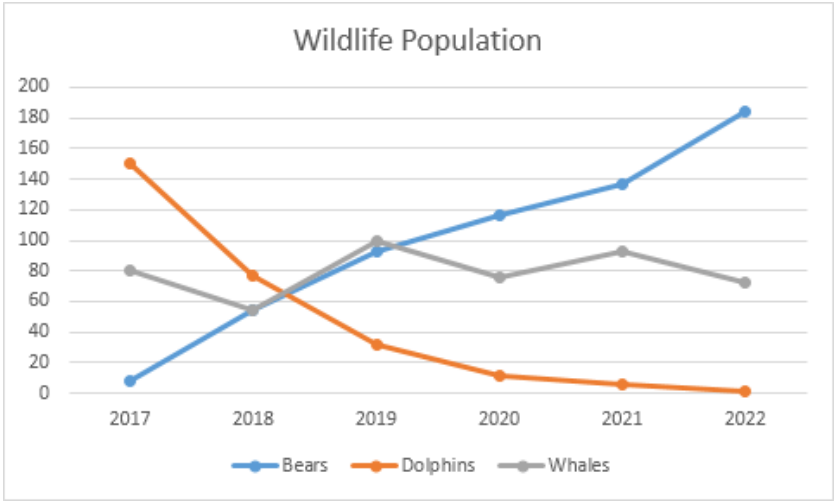
- Graph consists of *nodes and edges (connections)*.
- Graphs can be directed or undirected.
- Used in social networks, transportation networks, representation of map.

Rules for Graph Drawing and Labeling:

1. **Node Placement:** Place nodes to minimize edge crossings.
2. **Edge Drawing:** Use straight lines or smooth curves for edges.
3. **Labeling:** Clearly label nodes and edges. Use consistent font sizes and styles.
4. **Color Coding:** Use colors to differentiate between different types of nodes or edges.
5. **Hierarchy:** If representing a hierarchy, ensure the levels are clearly distinguishable

TIME SERIES DATA

- Time series data is a sequence of data points *collected or recorded at regular time intervals.*
- This type of data is crucial for analyzing trends, patterns, and changes over time.

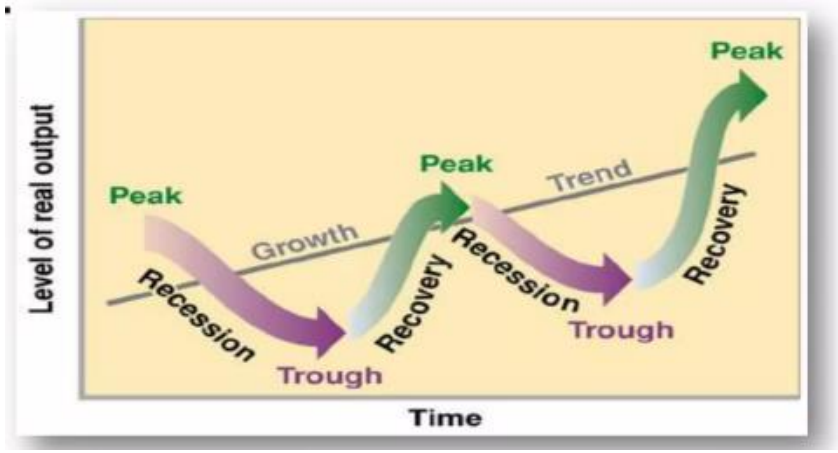


- Timestamped Data Points – data point is associated with a specific timestamp
- Continuous or Discrete Time: senor data recorded every second vs monthly sales data
- Univariate vs. Multivariate:

CHARACTERISTICS OF TIME DATA

Characteristics:

- **Trend**: The long-term movement or direction in the data.
- **Seasonality**: Regular, repeating patterns or cycles in the data.
- **Noise**: Random variations or fluctuations that do not follow a pattern.
- **Cyclic Patterns**: Long-term oscillations that are not of a fixed period.
- **Stationarity:** Mean, variance doesn't change over time

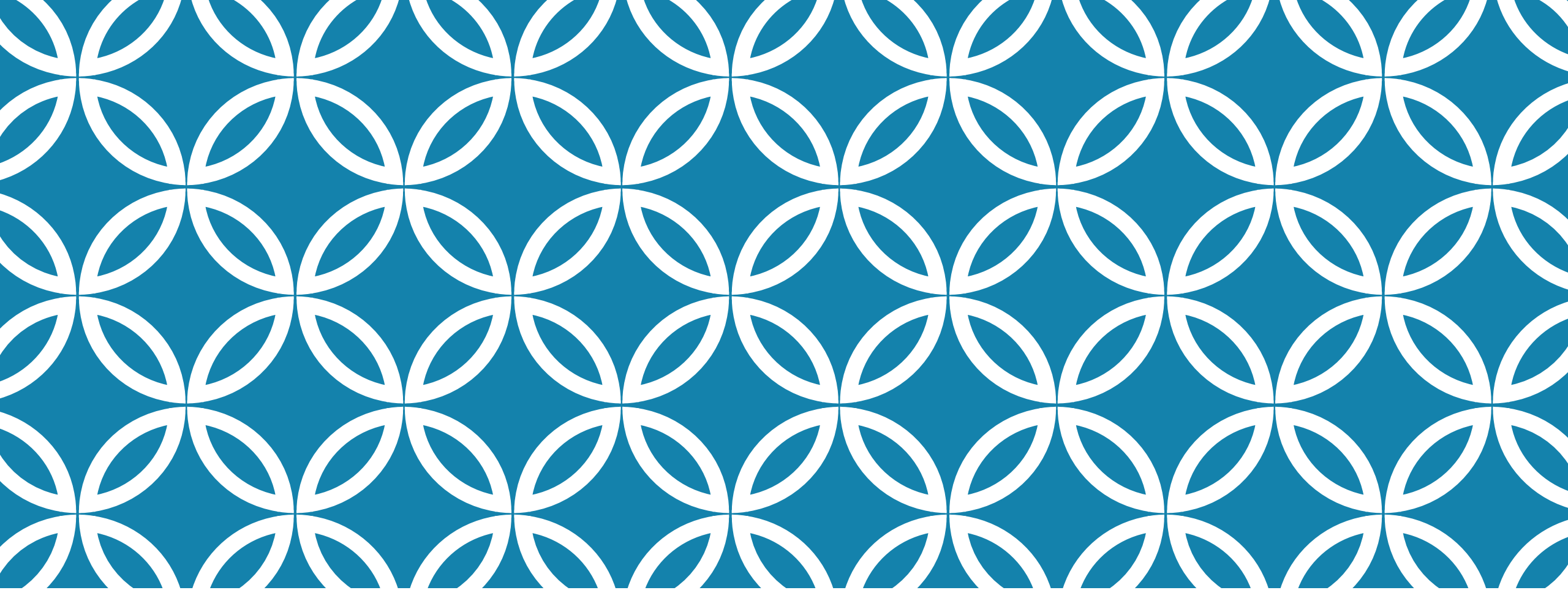


VISUALIZATION OF TIME SERIES DATA

Visualization of Time Series Data

- Line charts
- Heat maps
- Moving Averages
- Autocorrelation Plots

E.g., Stock Market Trends, Weather Data, Economic Data, Sales Data



THANK YOU..

Dipesh Koirala