



Exploratory Data Analysis

Visualization-based analysis

Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

A picture is worth a thousand words...

- Ancient Chinese proverb

A picture is worth a thousand words...

- Ancient Chinese proverb

"A visualization is worth much more than thousands of data points".

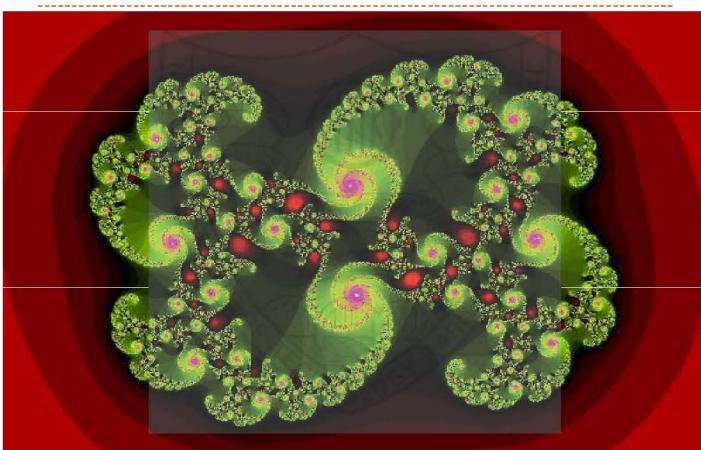
- In the era of Big Data

An Illustration..



" It looks like a swirl. There are smaller swirls at the edges. It has different shades of red at the outside, and is mostly green at the inside. The smaller swirls have purple highlights. The green has also different shades. Each small swirl is composed of even smaller ones. The swirls go clockwise. Inside the object, there are also red highlights. Those have different shades of red also. The green shades vary in a fan, while the purple ones are more uni-color. The green shades get darker towards the outside of the fan..... "

Data visualized ...



Another Illustration..



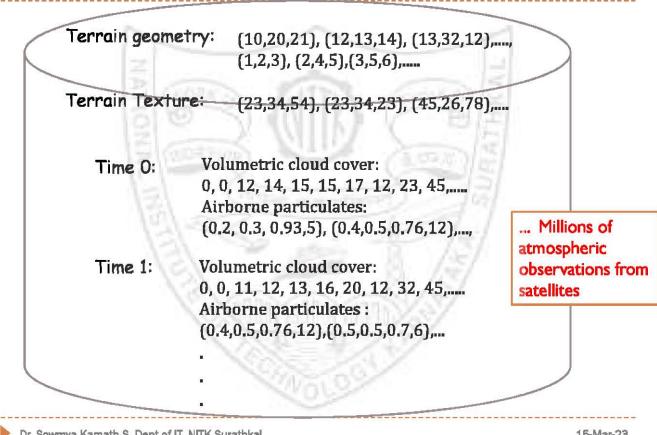
Terrain geometry: (10,20,21), (12,13,14), (13,32,12),...
(1,2,3), (2,4,5),(3,5,6),...

Terrain Texture: (23,34,54), (23,34,23), (45,26,78),...

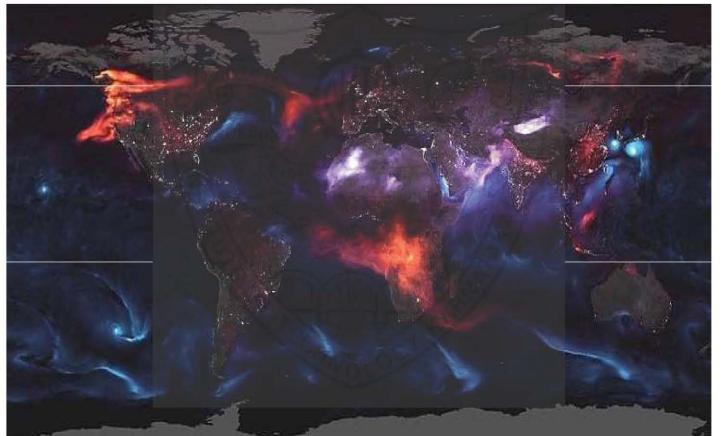
Time 0: Volumetric cloud cover:
0, 0, 12, 14, 15, 15, 17, 12, 23, 45,...
Airborne particulates:
(0.2, 0.3, 0.93,5), (0.4,0.5,0.76,12),...

Time 1: Volumetric cloud cover:
0, 0, 11, 12, 13, 16, 20, 12, 32, 45,...
Airborne particulates :
(0.4,0.5,0.76,12),(0.5,0.5,0.7,6),...

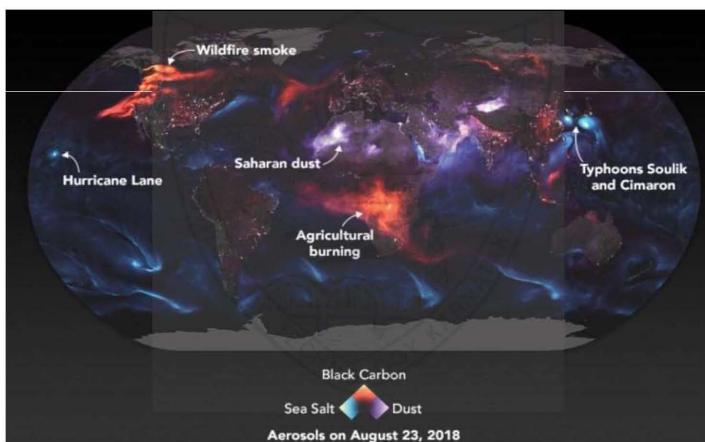
Another Illustration...



Data visualized....



Data visualized....



Data visualized....



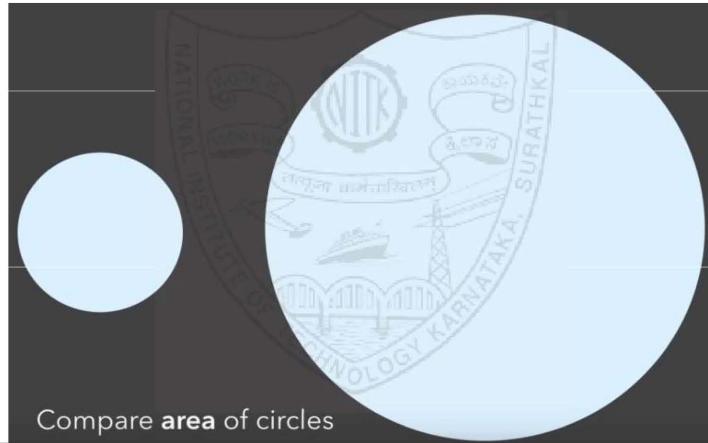
Defining Visualization...

“... finding the **artificial representation** that best supports our **natural means of perception**.”

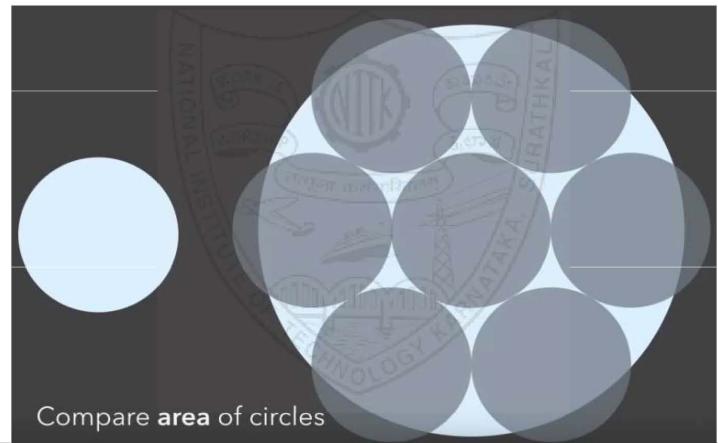
The use of computer-supported, **interactive**, visual representations of **abstract data** to amplify cognition...

Why is InfoVis important?

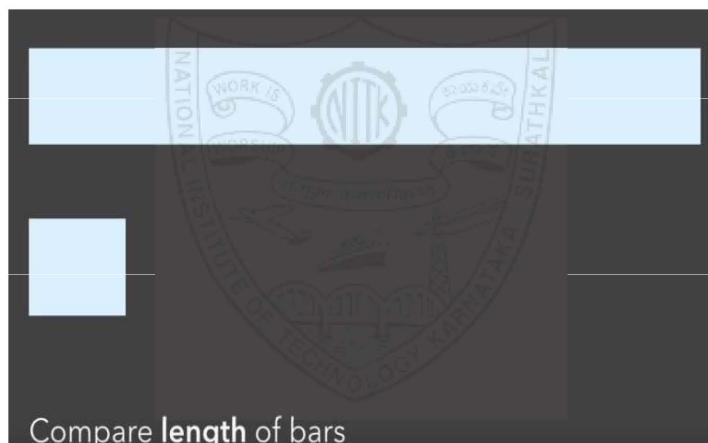
Why is InfoVis important?



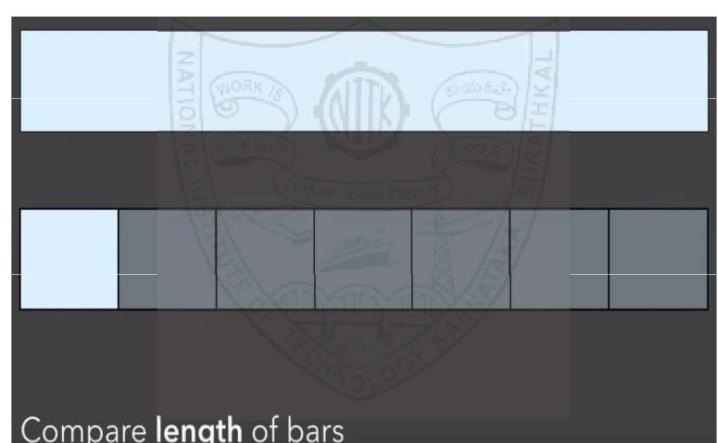
Why is InfoVis important?



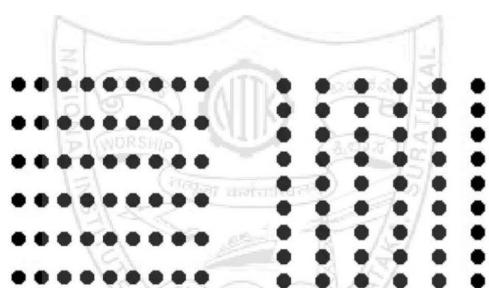
Why is InfoVis important?



Why is InfoVis important?



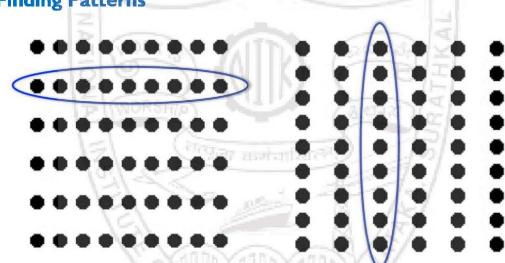
Why is InfoVis important?



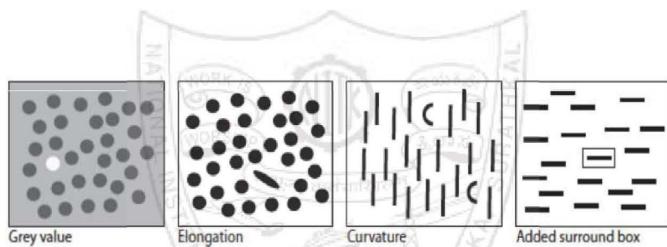
Why is InfoVis important?



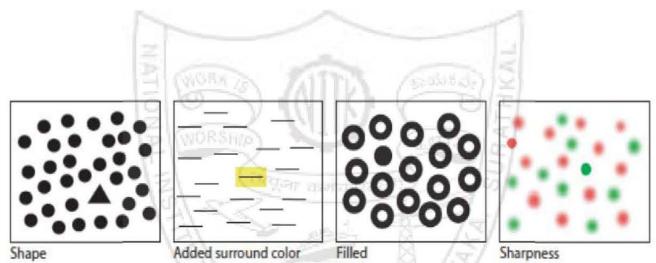
Finding Patterns



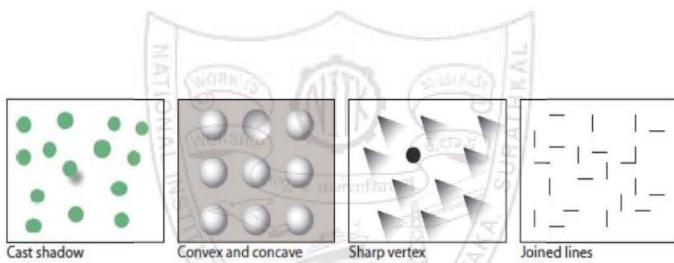
Finding Patterns



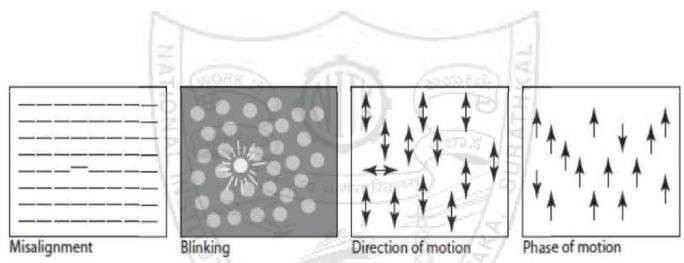
Finding Patterns



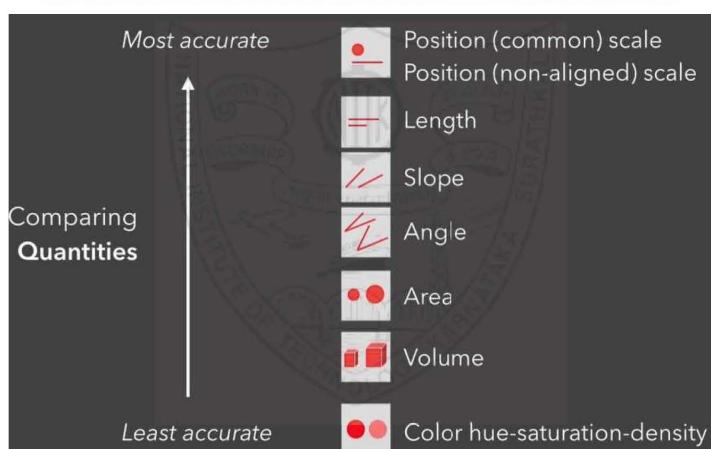
Finding Patterns



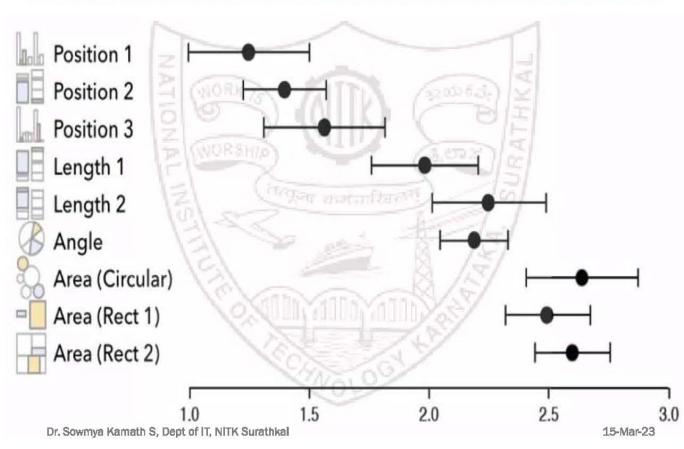
Finding Patterns



Ranking Visual Decodings



Accuracy of Visual Decoding



Data visualization in Data Science

Good old data....



	Set A		Set B		Set C		Set D	
	X	Y	X	Y	X	Y	X	Y
0	10	8.04	10	9.14	10	7.46	8	6.58
1	8	6.95	8	8.14	8	6.77	8	5.76
2	13	7.58	13	8.74	13	12.74	8	7.71
3	9	8.81	9	8.77	9	7.11	8	8.84
4	11	8.33	11	9.26	11	7.81	8	8.47
5	14	9.96	14	8.10	14	8.84	8	7.04
6	6	7.24	6	6.13	6	6.08	8	5.25
7	4	4.26	4	3.10	4	5.39	19	12.50
8	12	10.84	12	9.13	12	8.15	8	5.56
9	7	4.82	7	7.26	7	6.42	8	7.91
10	5	5.68	5	4.74	5	5.73	8	6.89
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
std	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
corr.	0.82		0.82		0.82		0.82	
lin. reg.	y = 3.00 + 0.500x		y = 3.00 + 0.500x		y = 3.00 + 0.500x		y = 3.00 + 0.500x	

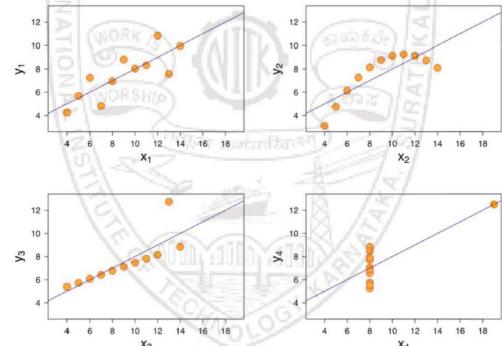
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



15-Mar-23

Data visualization in Data Science

Helps to see data in context



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

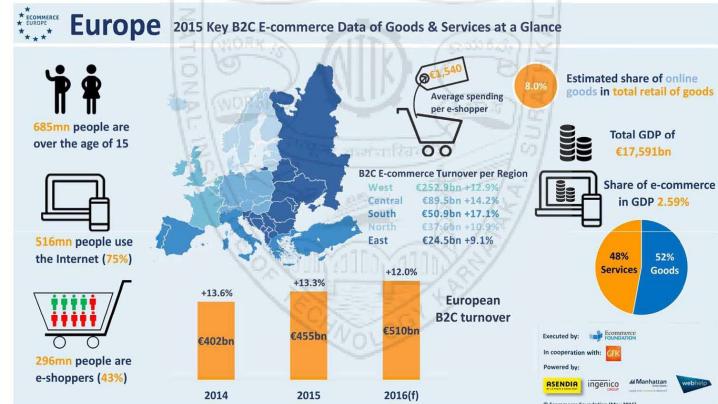
Data visualization in Data Science

Helps in reasoning and decision making



Data visualization in Data Science

Helps communicate information better



© Ecommerce Foundation (May 2016)

Modern Data Visualization Techniques

- ▶ Counts and densities
 - Tallys, Dot plots, Stem and leaf plots, Bar charts, Line/area charts
- ▶ Proportions
 - Divided bar charts, Pie charts
- ▶ Batches
 - Scatter plots, Box plots
- ▶ Pixel oriented
- ▶ Aggregation & Level of Details (LOD)
- ▶ Distortion
- ▶ Clutter Reduction
- ▶ Query based Visualization

► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

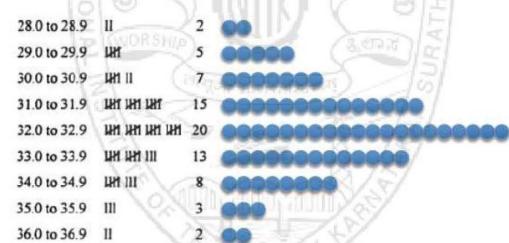


15-Mar-23

Modern Data Visualization Techniques

- ▶ Counts and densities

- Dot plots



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

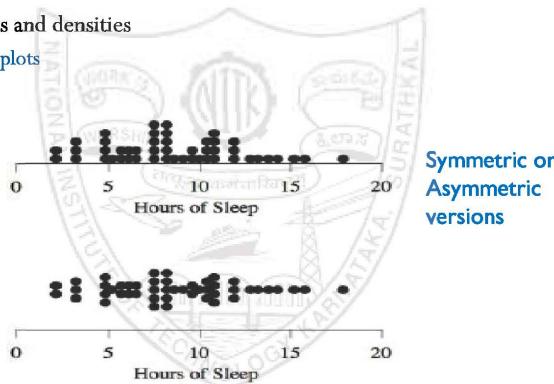
15-Mar-23



Modern Data Visualization Techniques



- Counts and densities
- Dot plots



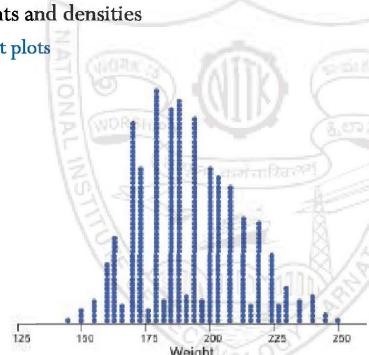
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques



- Counts and densities
- Dot plots



To reveal granularity

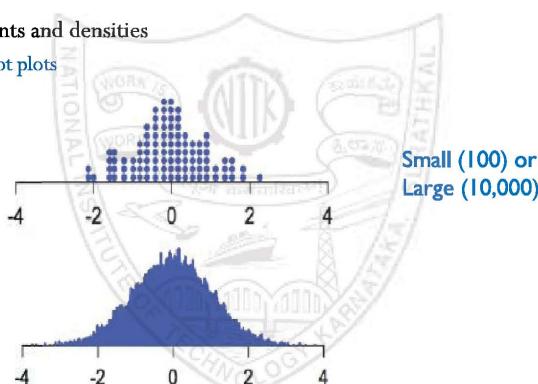
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques



- Counts and densities
- Dot plots



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques



- Counts and densities
- Dot plots

3D dot plot showing enormous range in density
not as well represented in heatmap

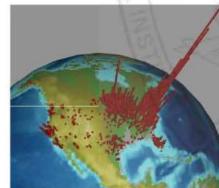


Fig. 8. Dot plot presenting Lyme disease from 2003 to 2005.

Same data from CDC map



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques



- Counts and densities
- Stem and Leaf plots

28.0 to 28.9	II	2	28	36
29.0 to 29.9	III	5	29	13378
30.0 to 30.9	III II	7	30	2224589
31.0 to 31.9	III III III	15	31	111224666677889
32.0 to 32.9	III III III III	20	32	00112223335567888999
33.0 to 33.9	III III III	13	33	0222344466899
34.0 to 34.9	III III III	8	34	11344668
35.0 to 35.9	III	3	35	333
36.0 to 36.9	II	2	36	48

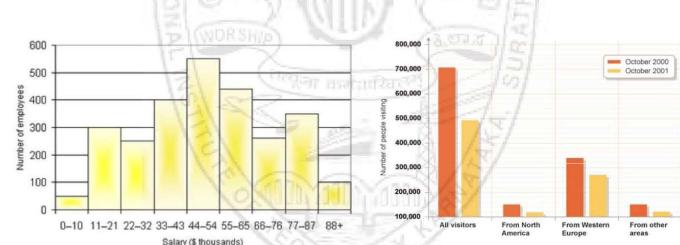
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques



- Counts and densities
- Bar charts

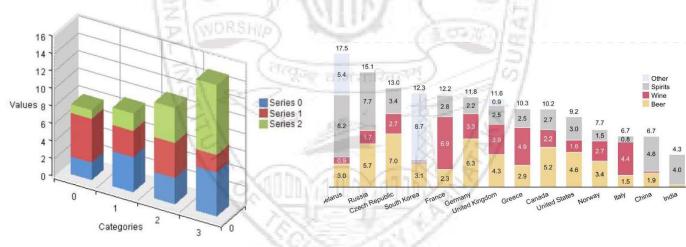


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- Counts and densities
- Bar charts

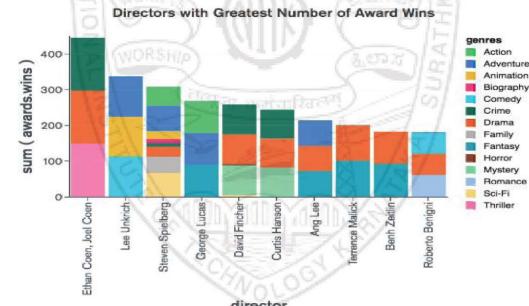


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- Counts and densities
- Bar charts



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- Pixel-oriented Visualization
- Represent each attribute value as a single colored pixel
- Map the range of possible attribute values to a fixed color map
- Advantage:**
 - Maximizes the amount of information represented at one time without any overlap

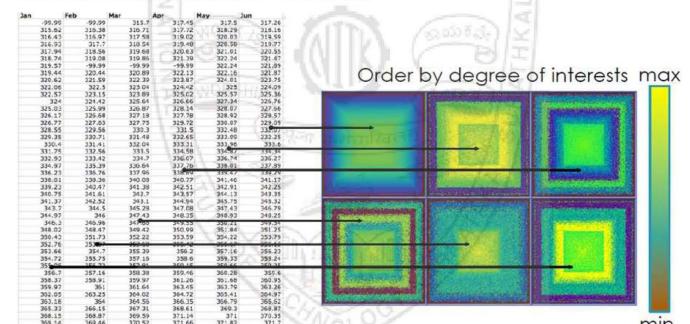
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

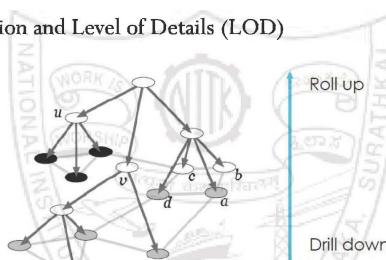
Pixel-oriented Visualization

Database visualization (10,000 items, 6 dimensions)



Modern Data Visualization Techniques

- Aggregation and Level of Details (LOD)



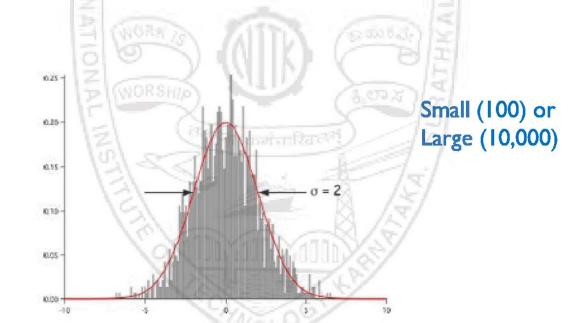
Building a tree for aggregating data items in either a bottom-up or top-down approach

► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- Aggregation and Level of Details (LOD)

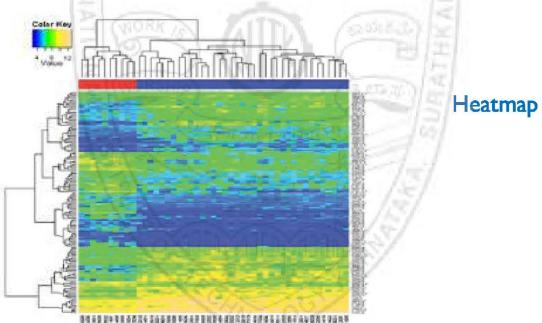


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- ▶ Aggregation and Level of Details (LOD)

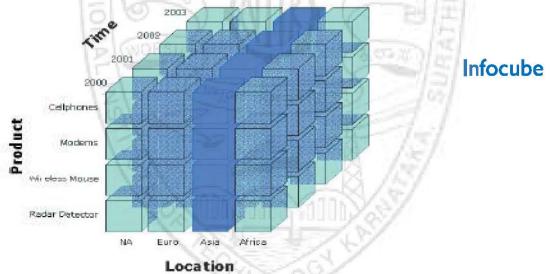


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- ▶ Aggregation and Level of Details (LOD)

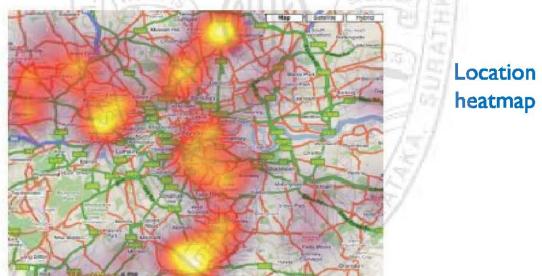


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- ▶ Aggregation and Level of Details (LOD)

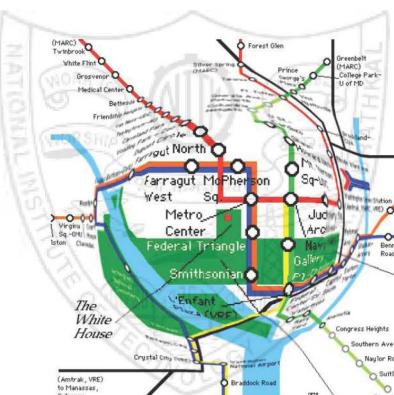


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- ▶ Distortion

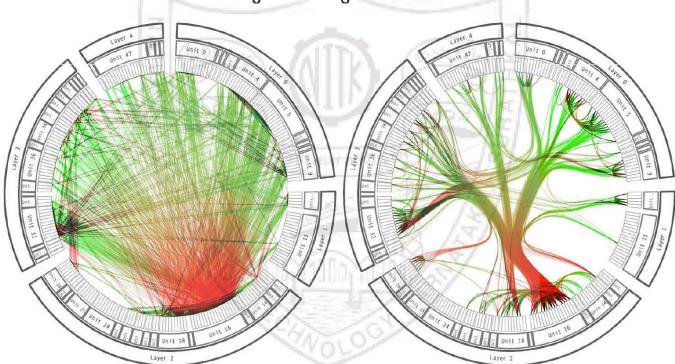


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- ▶ Clutter Reduction – Edge Bundling

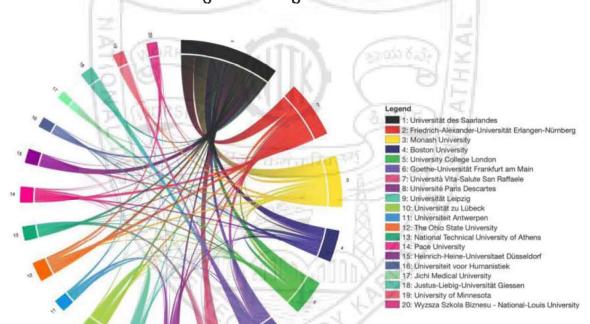


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

- ▶ Clutter Reduction – Edge Bundling

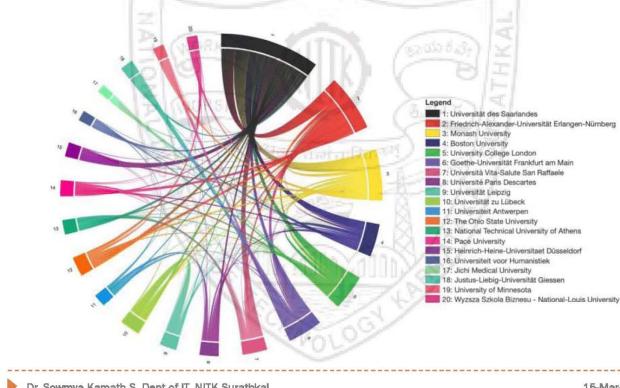


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Modern Data Visualization Techniques

Clutter Reduction – Edge Bundling

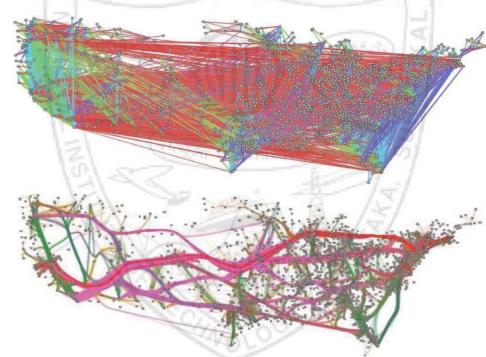


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



Modern Data Visualization Techniques

Clutter Reduction – Edge Bundling

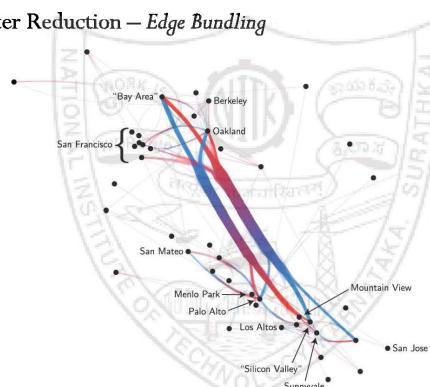


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



Modern Data Visualization Techniques

Clutter Reduction – Edge Bundling



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



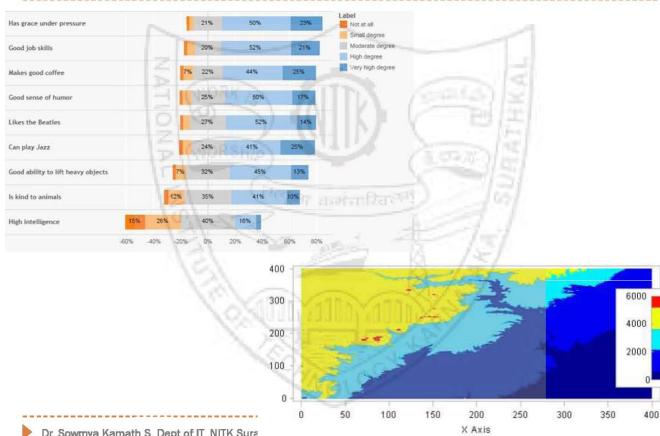
Examples: Visualizing 1D Numerical Data



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



Examples: Visualizing Ordinal data

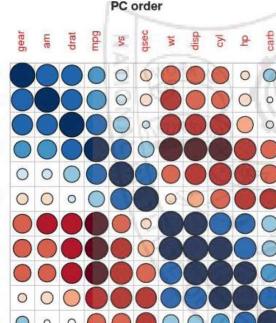


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



Examples: Visualizing Categorical data

PC order



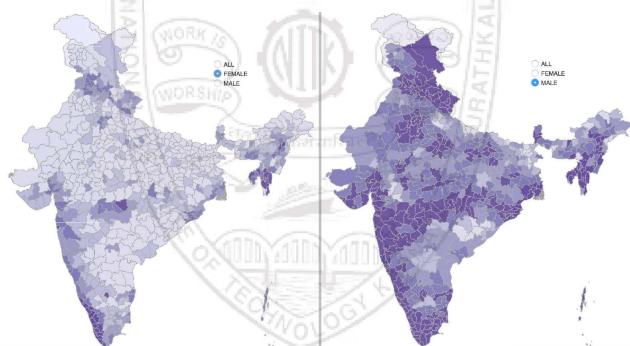
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



15-Mar-23

Examples: Visualizing 2D data

Chloropleth maps: show the level of variability within a region.



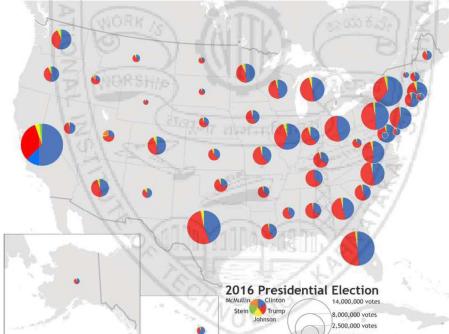
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



15-Mar-23

Examples: Visualizing 2D data

Proportional Symbol Maps: uses map symbols that vary in size to represent a quantitative variable.



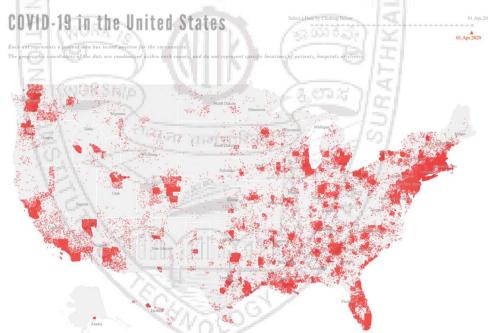
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



15-Mar-23

Examples: Visualizing 2D data

Dot Density Maps: rely on a visual scatter to show spatial patterns, especially variances in density



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23



Examples: Visualizing 2D data

Population of Asia and Oceania in 2018



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

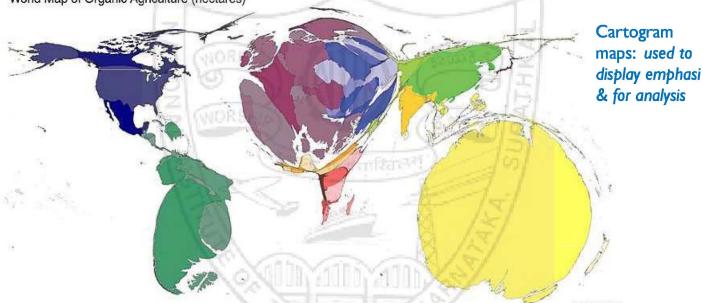


Cartogram maps: used to display emphasis & for analysis

15-Mar-23

Examples: Visualizing 2D data

World Map of Organic Agriculture (hectares)



Cartogram maps: used to display emphasis & for analysis

Area cartogram of the world with each country rescaled in proportion to the hectares of certified organic farming

► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

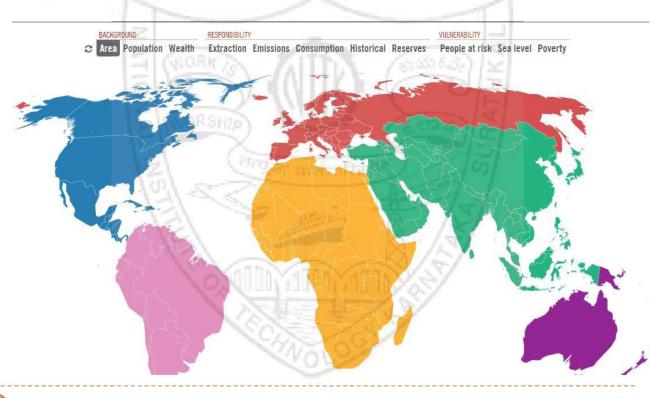
15-Mar-23



Examples: Visualizing 2D data

Cartogram maps

<https://www.carbonmap.org/>



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

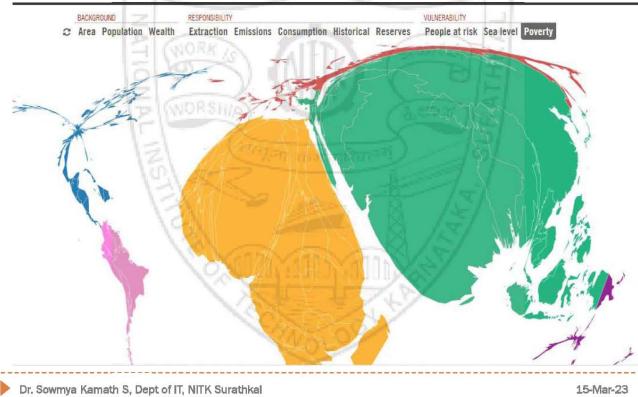


15-Mar-23

Examples: Visualizing 2D data

Cartogram maps

<https://www.carbonmap.org/>

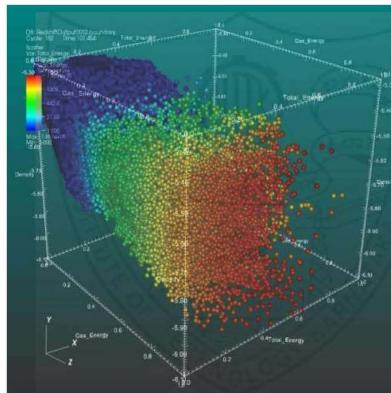


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal



Examples: Visualizing 3D data

3d Scatter plot:
allows the
visualization of
multivariate data



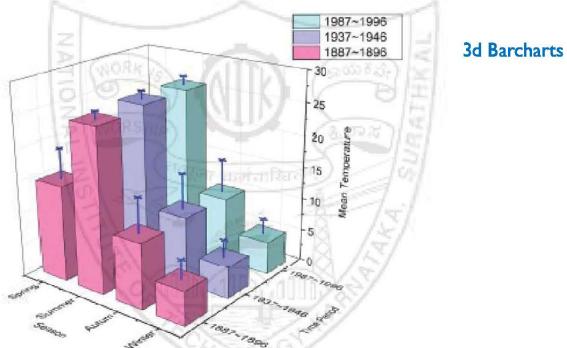
► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Examples: Visualizing 3D data



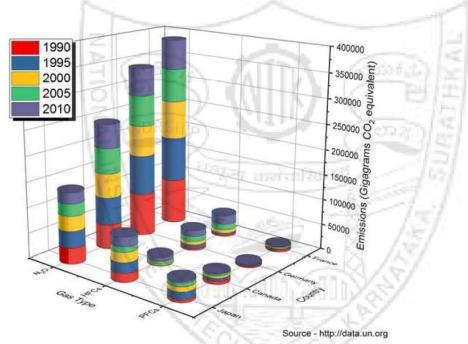
3d Barcharts



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

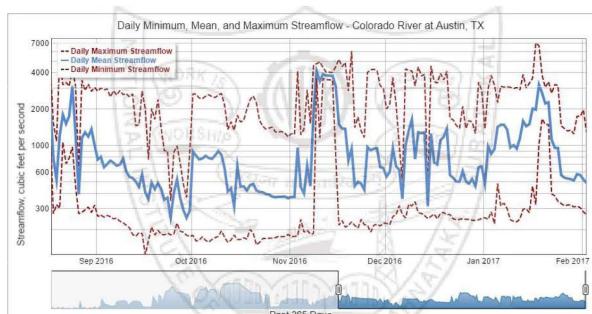
3d Stacked Barcharts



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Examples: Visualizing Temporal data

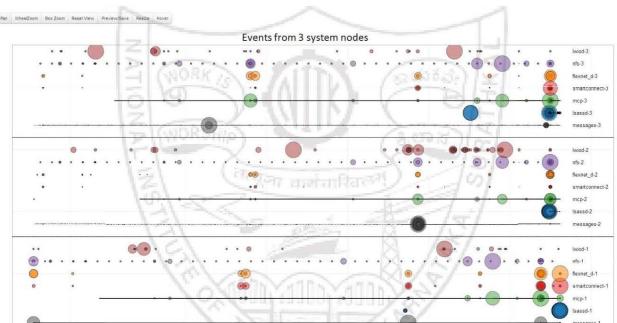


Time series: a plot of sequences taken at successive equally spaced points in time.

► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Examples: Visualizing Temporal data

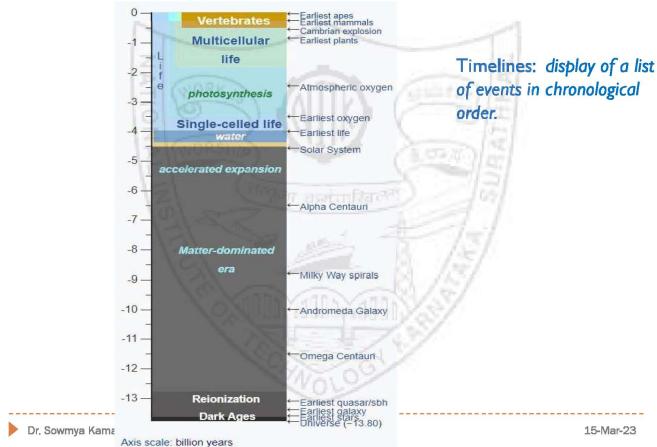


Timelines: display of a list of events in chronological order.

► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

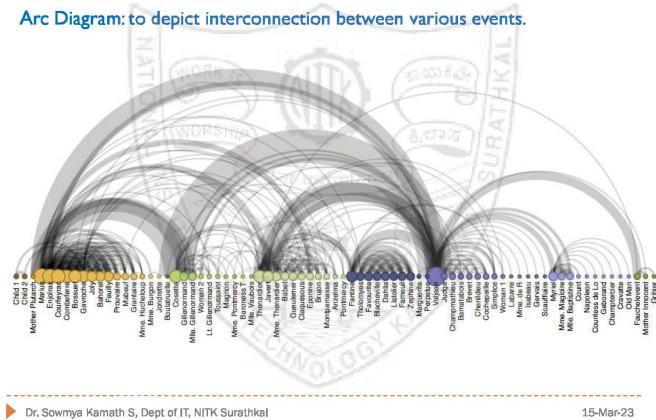
15-Mar-23

Examples: Visualizing Temporal data

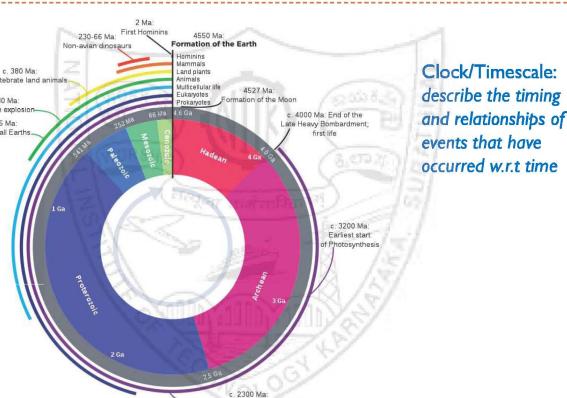


Examples: Visualizing Temporal data

Arc Diagram: to depict interconnection between various events.



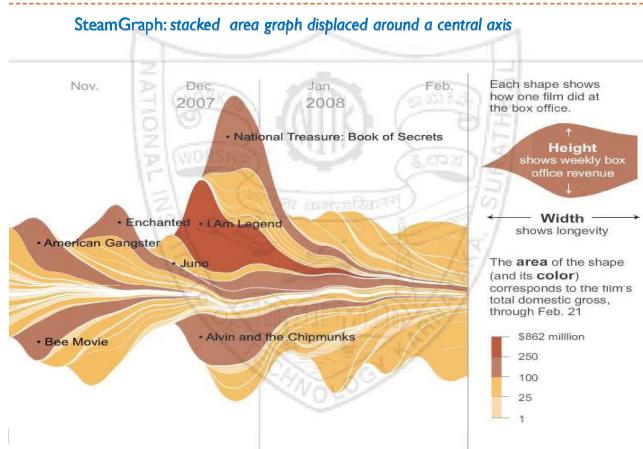
Examples: Visualizing Temporal data



Examples: Visualizing Temporal data

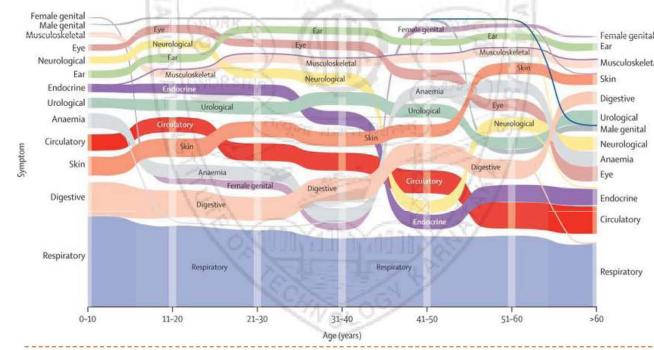


Examples: Visualizing Temporal data



Examples: Visualizing Temporal data

Alluvial Diagram: represent changes in network structure over time

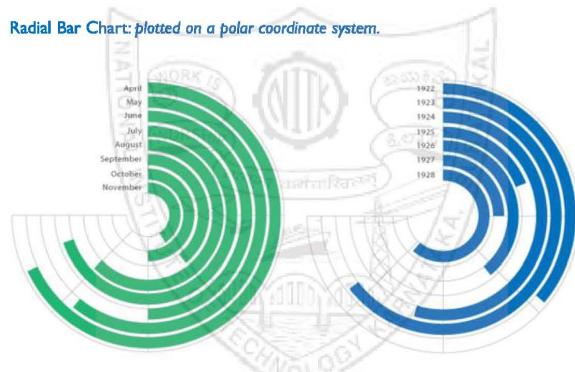


Examples: Visualizing n-Dimensional data



Examples: Visualizing n-Dimensional data

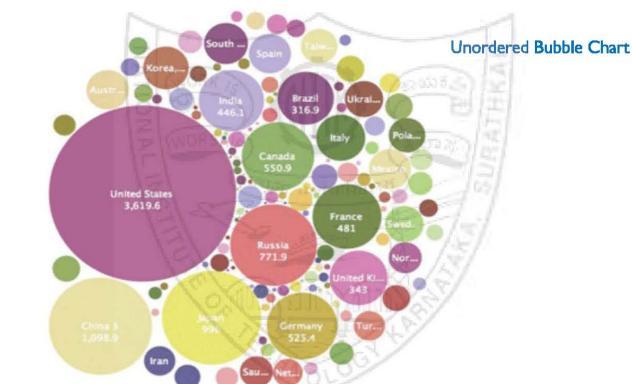
Radial Bar Chart: plotted on a polar coordinate system.



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Examples: Visualizing n-dimensional data



► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

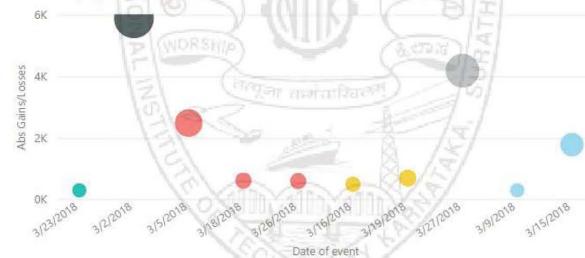
15-Mar-23

Examples: Visualizing n-dimensional data

Ordered Bubble Chart

Abs Gains/Losses and Abs Gains/Losses by Business Group and Date of event

Business Group: A, B, C, D, E, F

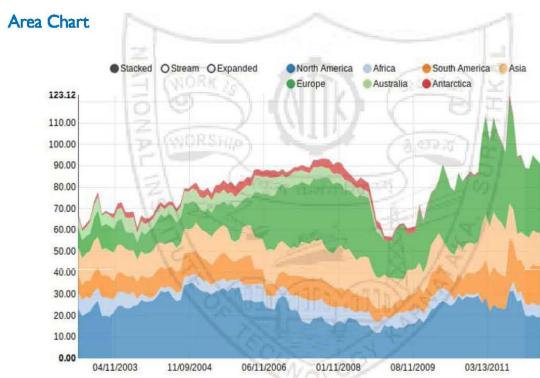


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Examples: Visualizing n-dimensional data

Area Chart

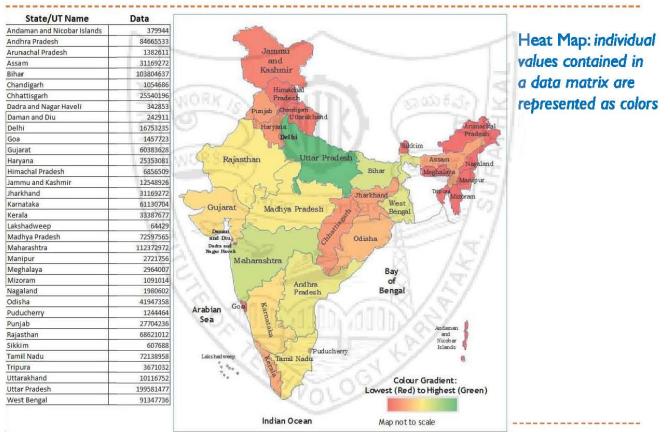


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

Examples: Visualizing n-dimensional data

Heat Map: individual values contained in a data matrix are represented as colors

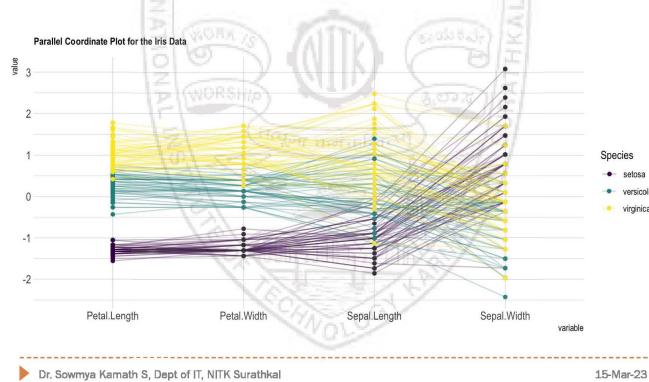


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

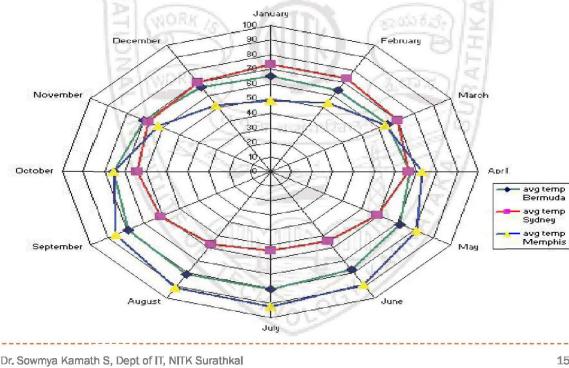
Examples: Visualizing n-dimensional data

Parallel Coordinates: used for visualizing high-dimensional geometry & analyzing multivariate data.



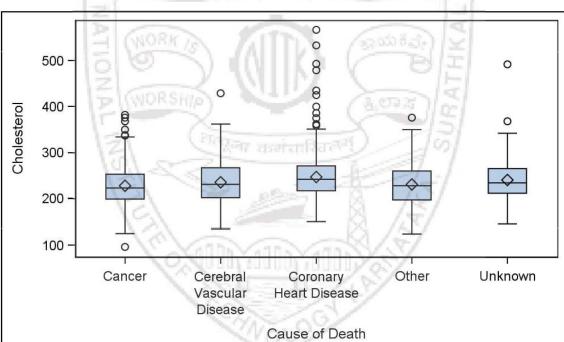
Examples: Visualizing n-dimensional data

Radar/Spider Chart: used for displaying multivariate data in the form of a 2-d chart of three or more quantitative variables represented on axes starting from the same point.



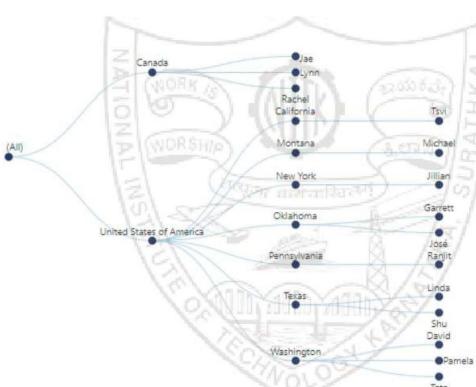
Examples: Visualizing n-dimensional data

Box & Whisker plot: graphically depicting groups of numerical data through their quartiles.



Examples: Visualizing Hierarchical data

Simple tree visualization

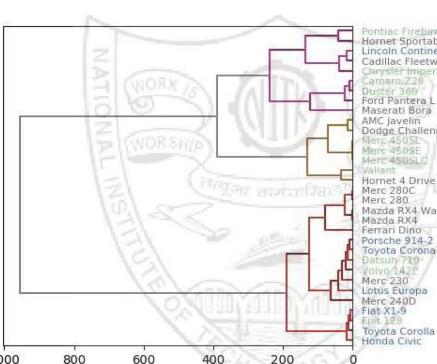


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

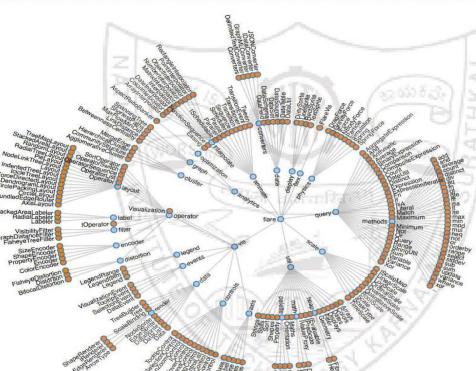
Examples: Visualizing Hierarchical data

Dendrogram: diagram representing a tree



Examples: Visualizing Hierarchical data

Radial tree: diagram representing a tree

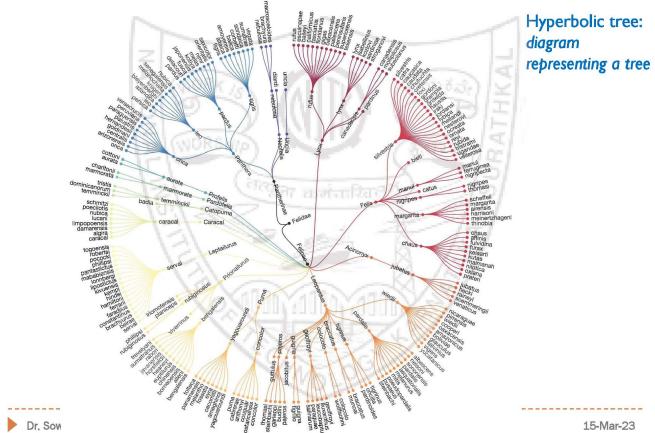


► Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

15-Mar-23

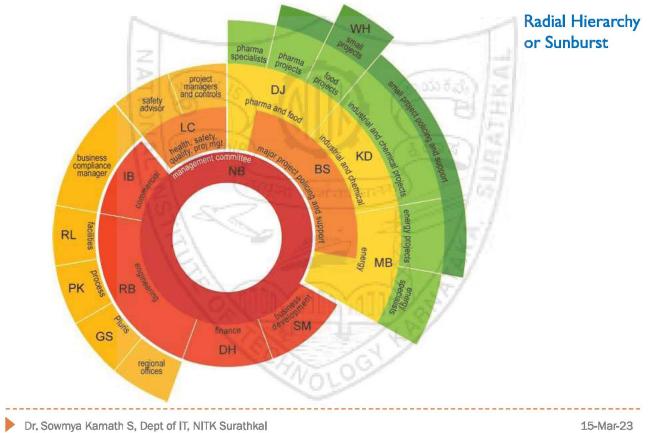
15-Mar-23

Examples: Visualizing Hierarchical data



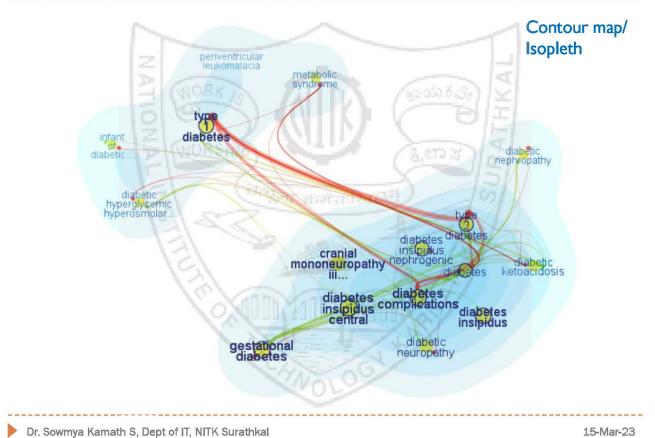
Hyperbolic tree: diagram representing a tree

Examples: Visualizing Hierarchical data



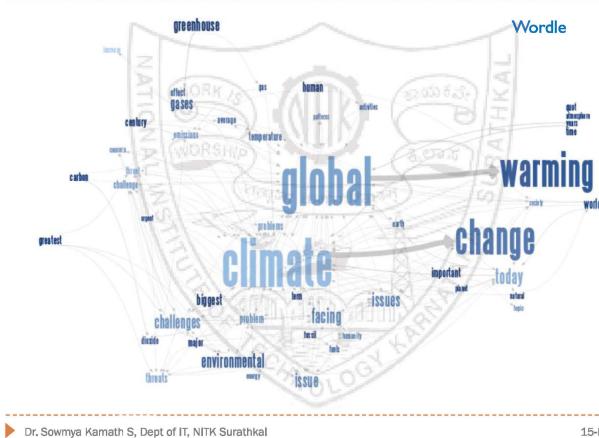
Radial Hierarchy or Sunburst

Examples: Visualizing Unstructured Data



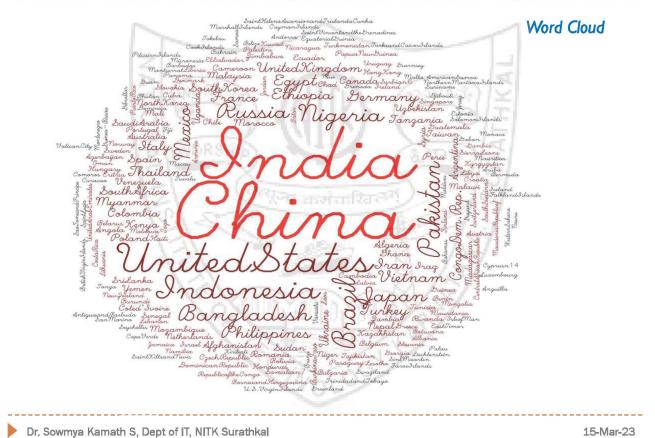
1 Contour map/ Isopleth

Examples: Visualizing Unstructured Data



15-Mar-23

Examples: Visualizing Unstructured Data



Word Cloud

Tools and Technologies

- ▶ R Packages like ggplot2, highcharter ...
 - ▶ D3.js
 - ▶ Javascript, and CSS with HTML ..
 - ▶ SVG (Scalable Vector Graphics)
 - ▶ Many free tools - <http://selection.datavisualization.ch/>

Dr. Sewmira Kamath S, Dept. of IT, NITK Surathkal

More information...



► Websites

- ▶ <http://crossingjs.org/>
- ▶ <http://d3js.org>
- ▶ <http://code.google.com/apis/ajax/playground/>
- ▶ <http://www.visualcomplexity.com/>
- ▶ <http://www.visualizationdepot.com/2009/06/50-great-examples-of-data-visualization>