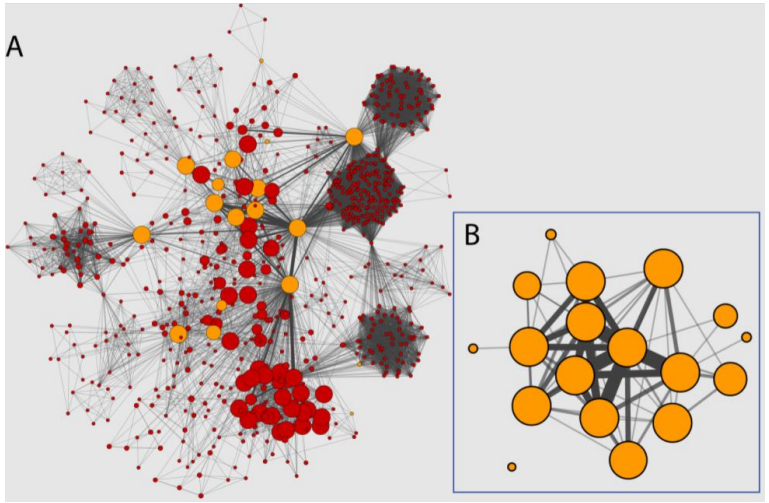
دانشگاه کردستان
University of Kurdistan
زانکۆی کوردستان

Biological Network Visualization

Sadegh Sulaimany



Biological network analysis course

Motivation

Information is the oil of the 21st century, and [visual] analytics is the combustion engine

Peter Sondergaard, Senior Vice President at Gartner

Agenda

- Importance of visualization
- Visual analytics
- Prerequisites
- Introducing and selecting the right chart
- Tools
- Drawbacks
- Study more

Importance of data visualization

- Anscombe's quartet
  - importance of graphing data before analyzing it
  - counter prevailing scientific opinions that numerical calculations were more exact than visualizations
  - numerical calculations may not be feasible due to disparate data types
  - Bioinformaticians are seeking to understand patterns through visualization

Graphs in Statistical Analysis

F. J. Anscombe

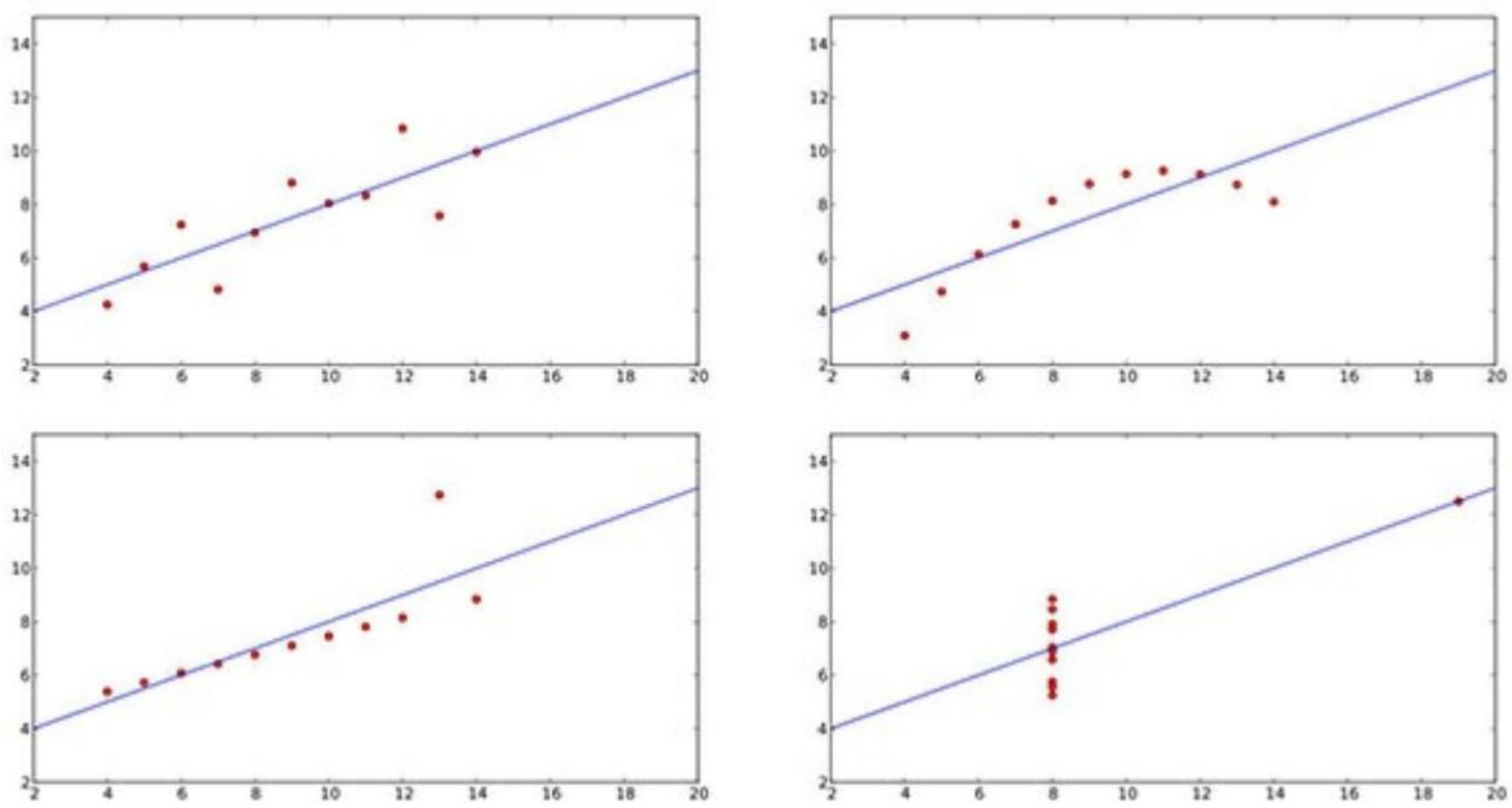
The American Statistician, Vol. 27, No. 1 (Feb., 1973), 17-21.

Stable URL:
http://links.jstor.org/sici?sici=0003-1305%28197302%2927%3A1%3C17%3AGISA%3E2.0.CO%3B2-J

The American Statistician is currently published by American Statistical Association.

Anscombe's quartet

- four datasets (each with 11 points) with nearly identical statistical properties
  - yet very different distributions and graphs



Hint

- Edward Tufte
  - pioneer in the field of data visualization
    - to communicate 4 or fewer numbers, use a sentence
    - for 4 - 20 numbers, consider a table
    - and for 8 or more numbers, consider a graph
  - Bioinformaticians may need to visualize data more than once during the complete data analytical iterative process
  - We will discuss visualization recommendations based on data types
  - and the software most commonly used to generate visualizations

Visual analytics

- Data visualization is a key skill set for bioinformatics and computational biology

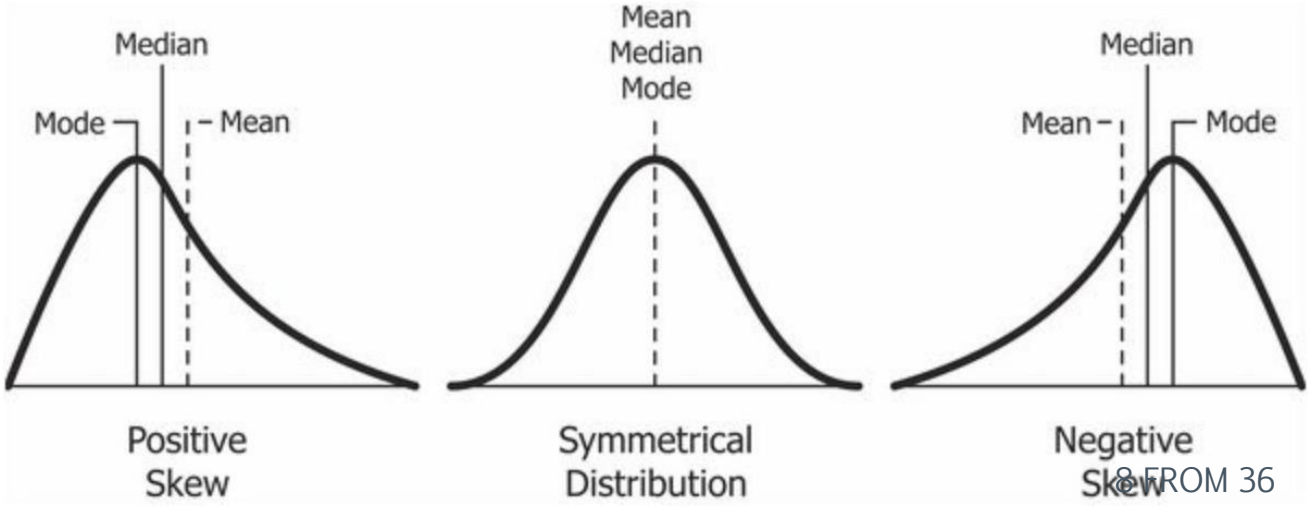
- visualization software is often promoted as **visual analytics**
  - use of sophisticated tools and processes to analyze datasets using **visual** representations of the data

- data visualization is considered an important skill set in the job market of most industries

- Data visualization
  - assists in the understanding and analysis of complex data by placing the data in a context that is easier to perceive

Concept mini-Review

- **Mean**
  - Add up the values in the dataset and then divide by the number of values that you added.
- **Median**
  - List the values of the data set in numerical order and identify which value appears in the middle of the list.
- **Mode**
  - Measure the values that occur most frequently.



Selecting the Right Chart Type

- There are four basic presentation types:
  - Comparison
  - Composition
  - Distribution
  - Relationship

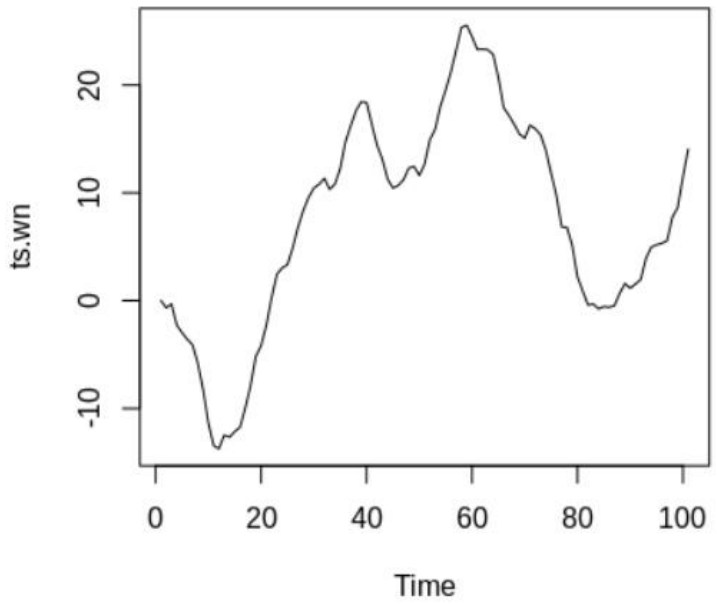
- Network visualization is interdisciplinary
  - Statistics
  - Social network
  - Biological network
  - …

Popular Charts

- Scatter Plot
- Histogram
- Bar & Stack Bar Chart
- Box Plot
- Pie chart
- Area Chart
- Heat Map
- Dendrogram
- Treemap
- Radar chart

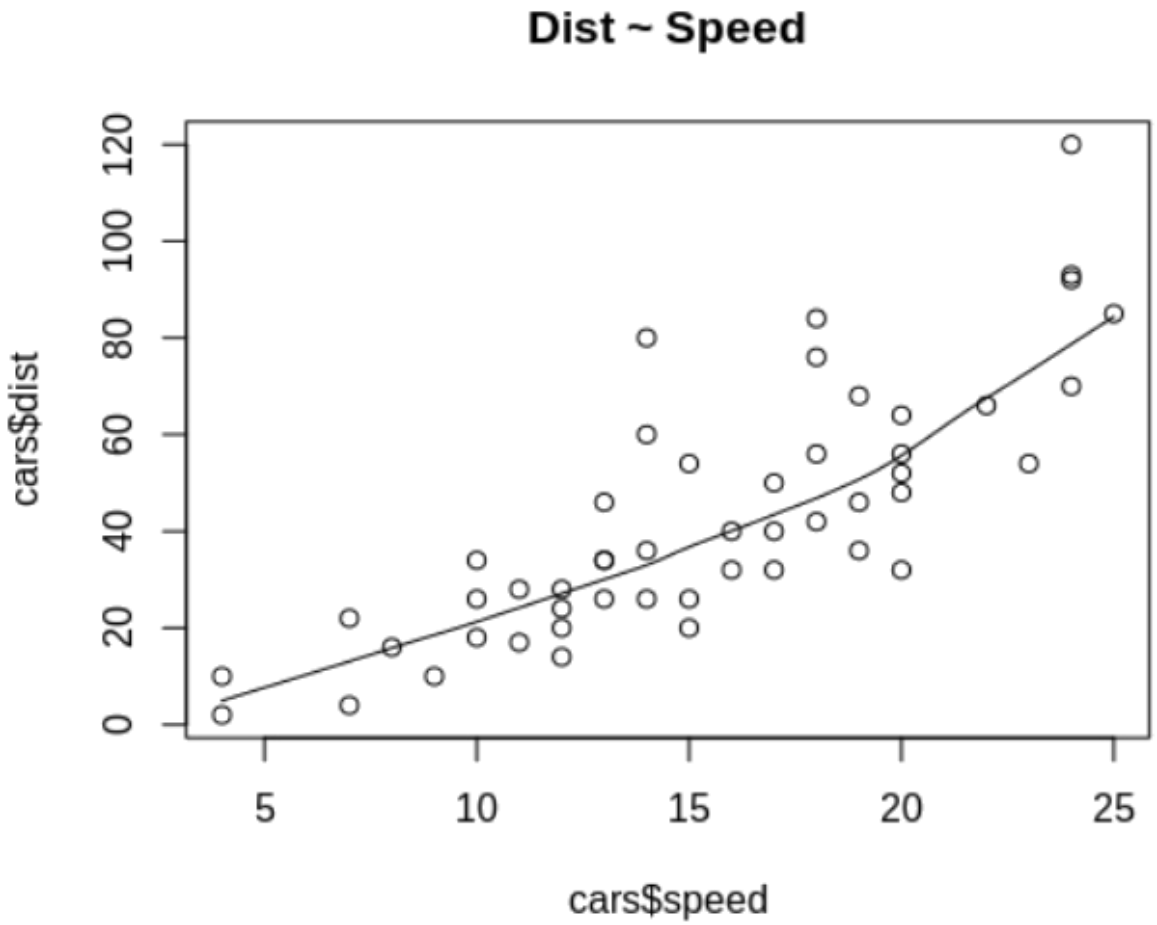
Area Chart

- show trends across a variable like time
- It is almost same as a line chart.
- we can use it for time series plots.
- We can use it alternatively to plot continuous variables and analyze the underlying trends
- Easy and popular



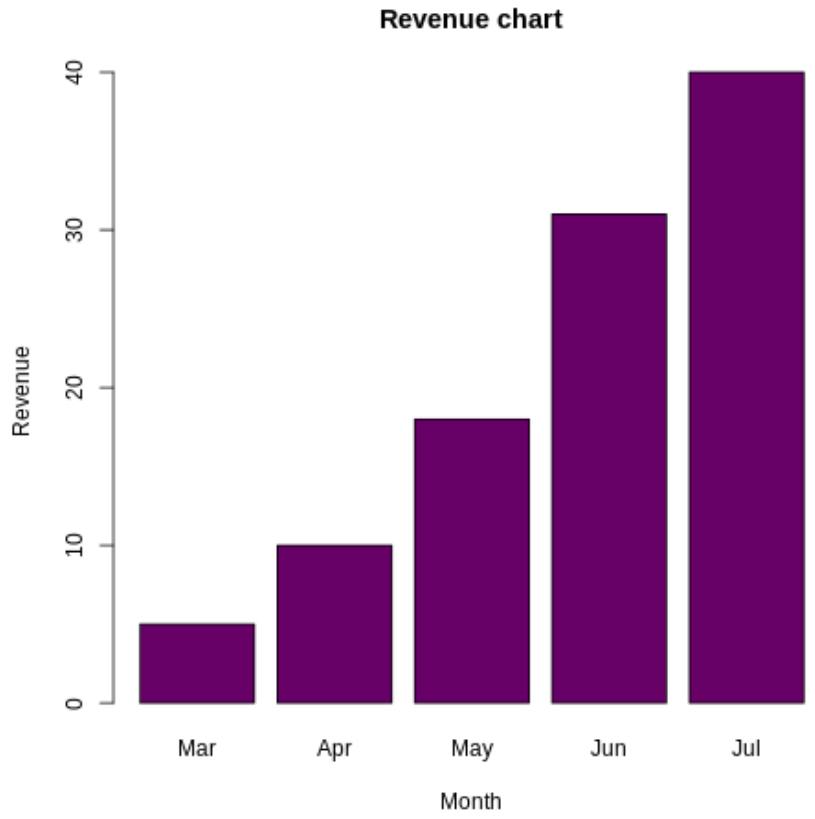
Scatter Plot

- To see the relationship between two continuous variables



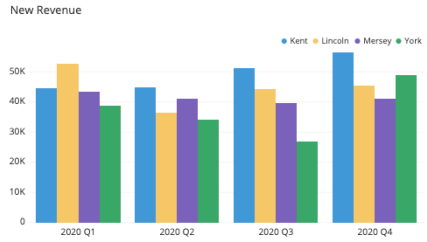
Bar Chart

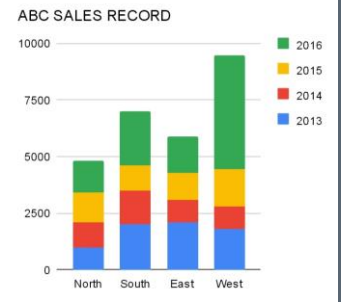
- to plot a categorical variable
  - compare the density, frequency,...



Bar chart details

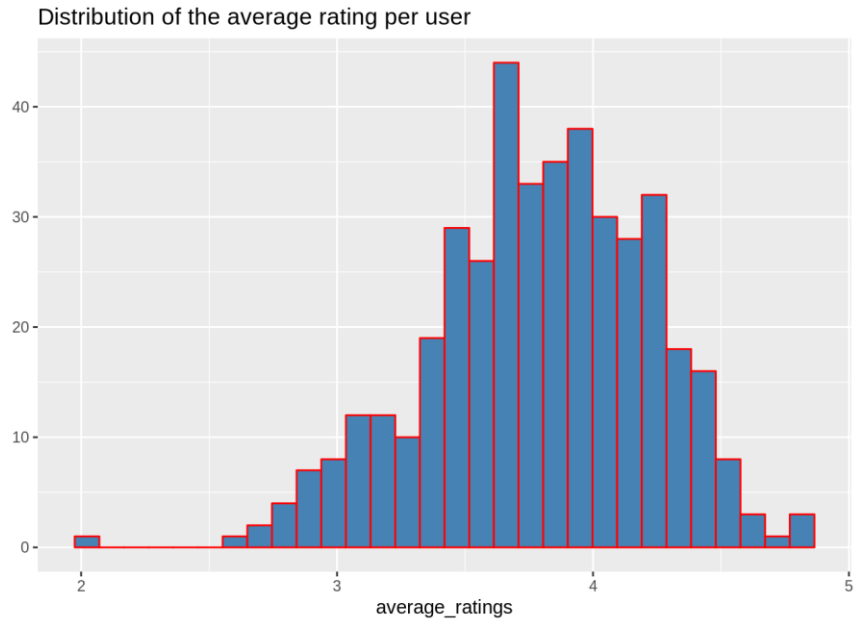
- Vertical and Horizontal
  - In order to effectively present data in data categories that possess long titles
- Grouped
  - show information about different sub-groups of the main categories
  - Each sub-group is represented by a separate bar
  - colored and shaded in a way that differentiates them from the other categories
- Stacked
  - similar to group bar charts
  - display information about the sub-groups that make up the different categories





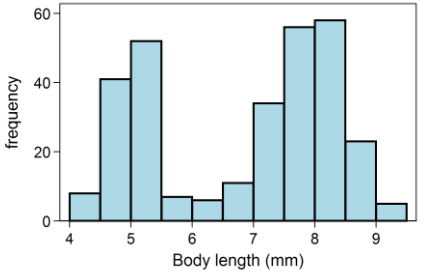
Histogram

- Represents A large amount of data, as well as the frequency of the data values
- Also, It helps to break the data into bins and shows the frequency distribution of these bins.
  - Thus, we can always change the bin size and see the effect it has on visualization.
- a special form of bar chart

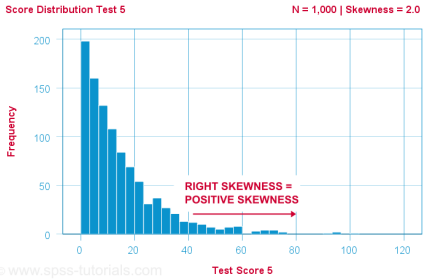


Histogram distributions

- Normal
  - the data on one side of the average is the same as another side. Normal distribution is
  - also known as Gaussian Distribution
- Bimodal Distribution
  - There are two peaks in a bimodal distribution
  - two normal distributions

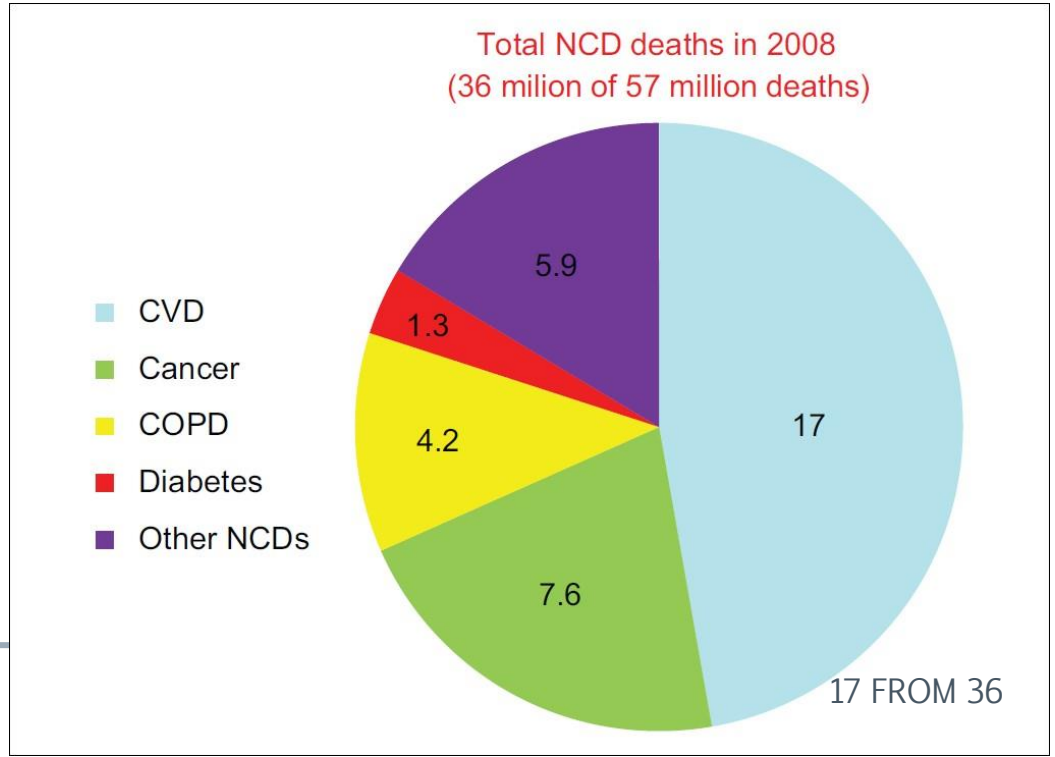


- Right Skewed Distribution
  - large number of data values occur on the left side with very few data points on the right side
- Left Skewed Distribution
  - Reverse of right skewed distribution
- Random distribution
  - does not possess a structured pattern
  - has several peaks in it



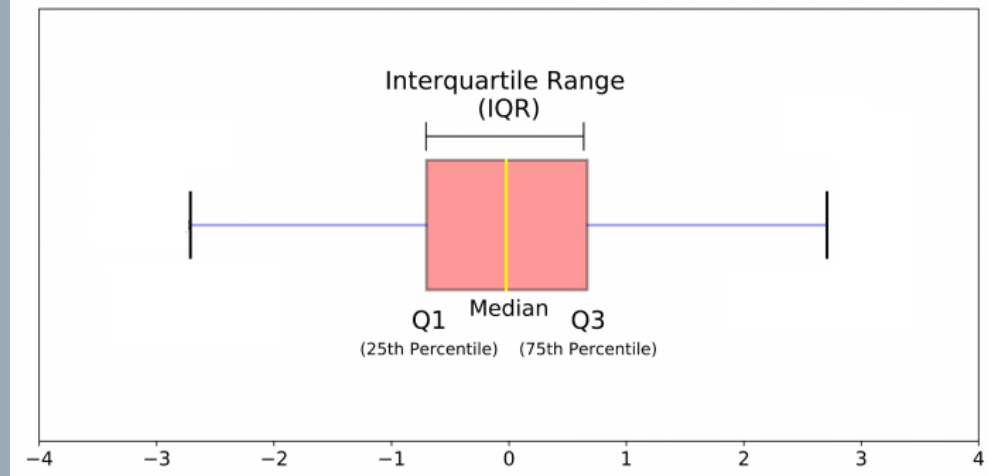
Pie Chart

- A circular statistical graphic
- is divided into slices to illustrate numerical proportion
- the arc length of each slice
  - is proportional to the quantity it represents
- when trying to work out the composition of something

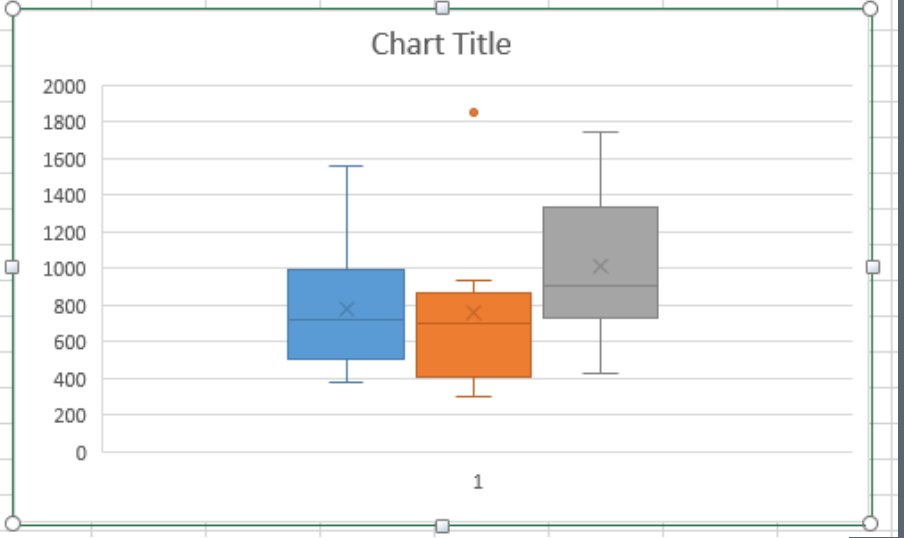


Box Plot

- graphically depicting groups of numerical data through their quartiles
  - › visualizing the spread of the data and detect outliers
- it shows five statistically significant numbers:
  - Minimum
  - 25th percentile
  - Median
  - 75th percentile
  - Maximum

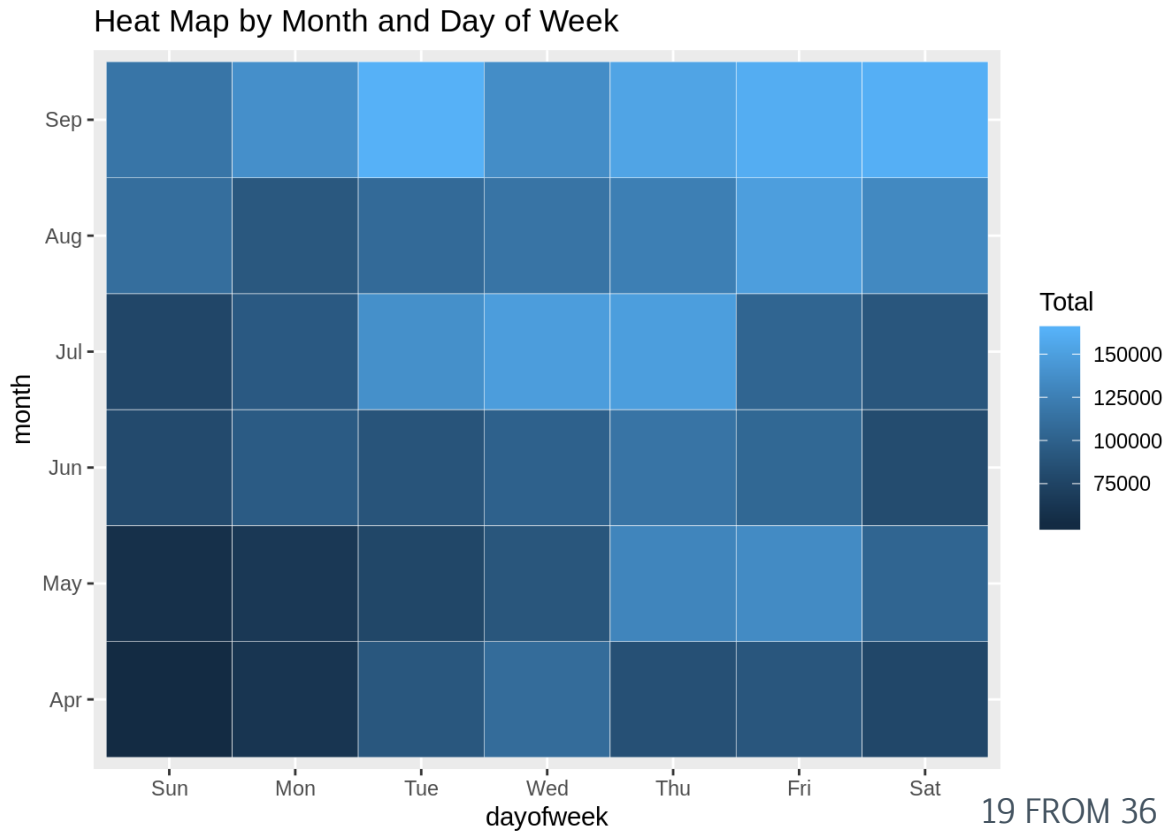


	A	B	C
1	Brand A	Brand B	Brand C
2	1020	840	1430
3	1560	940	1750
4	560	780	870
5	780	650	920
6	990	720	1300
7	670	430	890
8	510	1850	740
9	490	300	720
0	380	360	430
1	880	690	1050
2			
3			
4			
5			



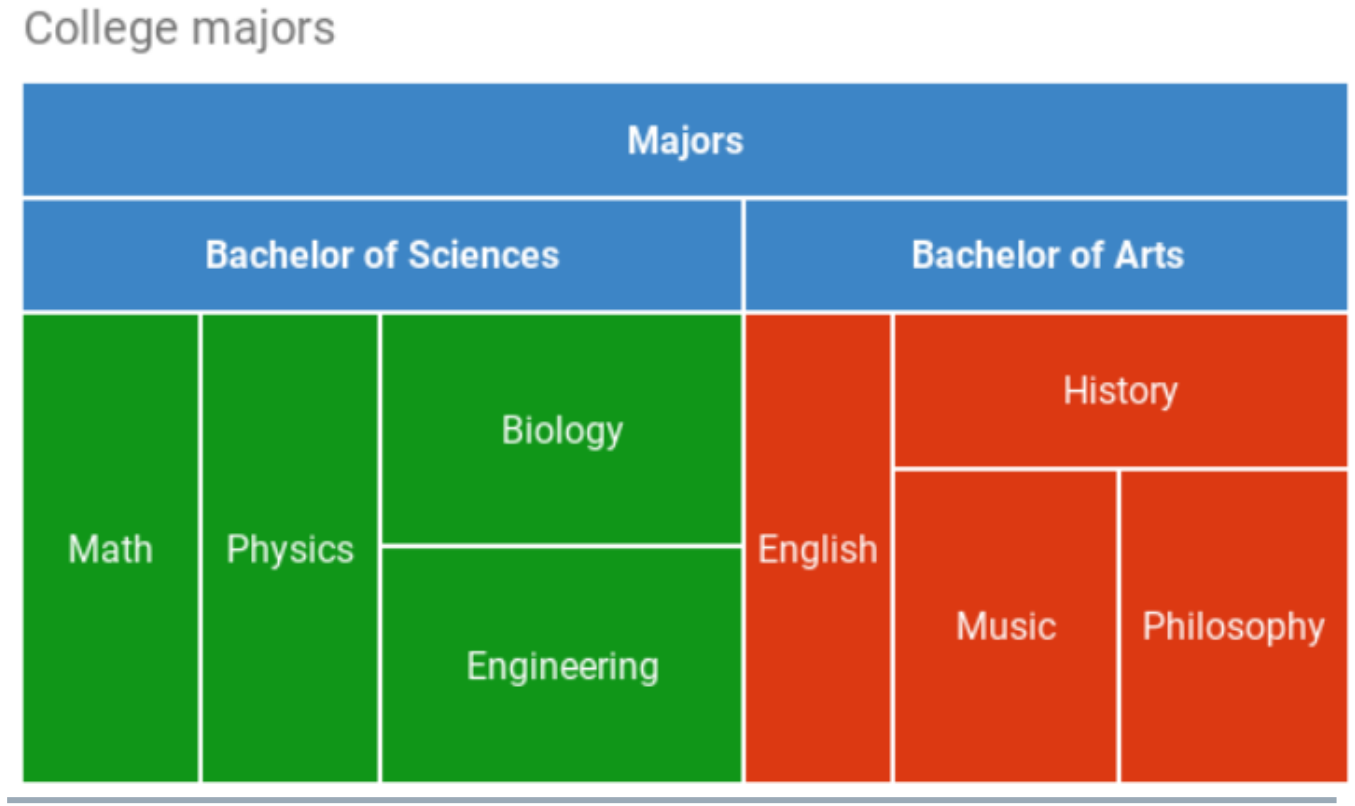
Heat Map

- We use it for the intensity of colors.
- It is also used to display a relationship between two or three or many variables in a two-dimensional image.
  - it allows us to explore two dimensions of the axis and the third dimension by an intensity of color



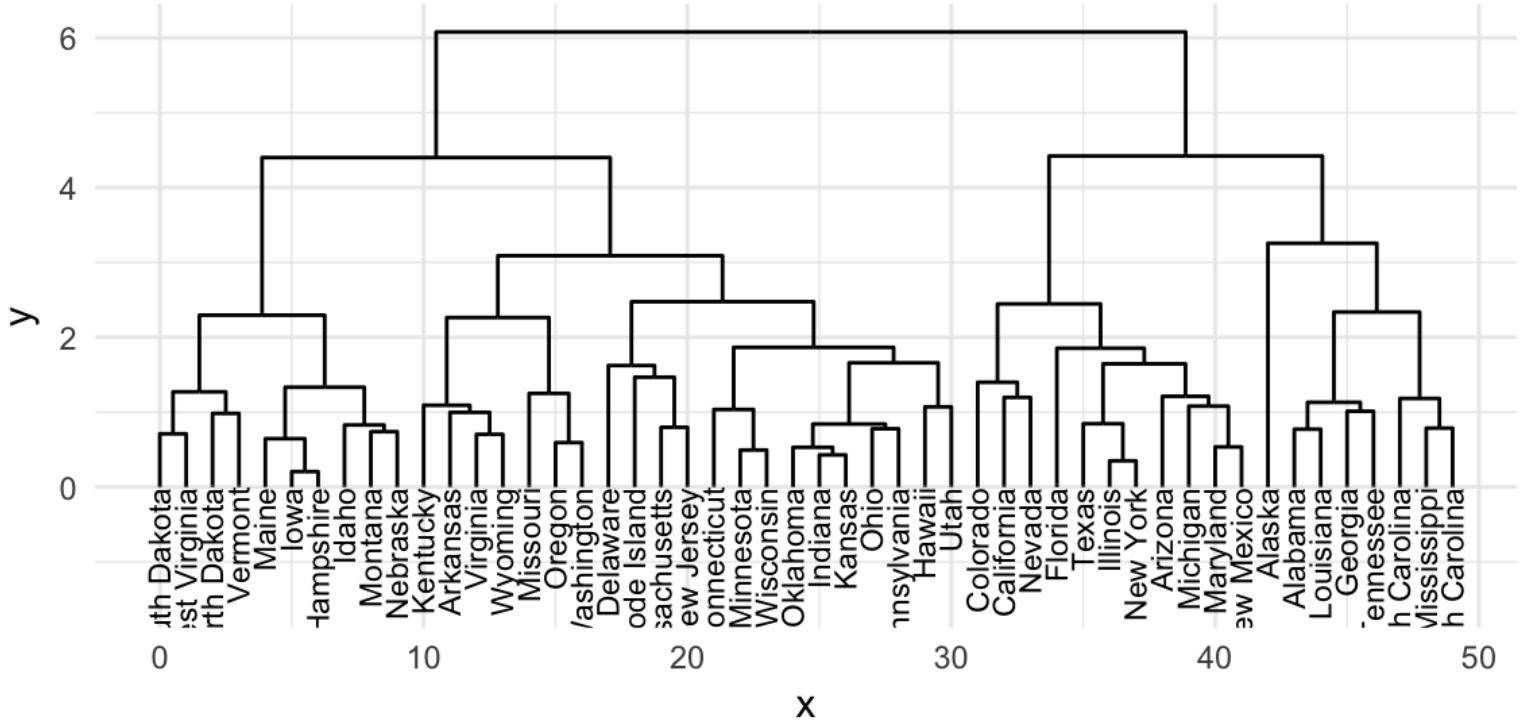
Treemap chart

- Useful for hierarchical data
  - each item is represented by a rectangular shape, where smaller rectangles represent the sub-groups



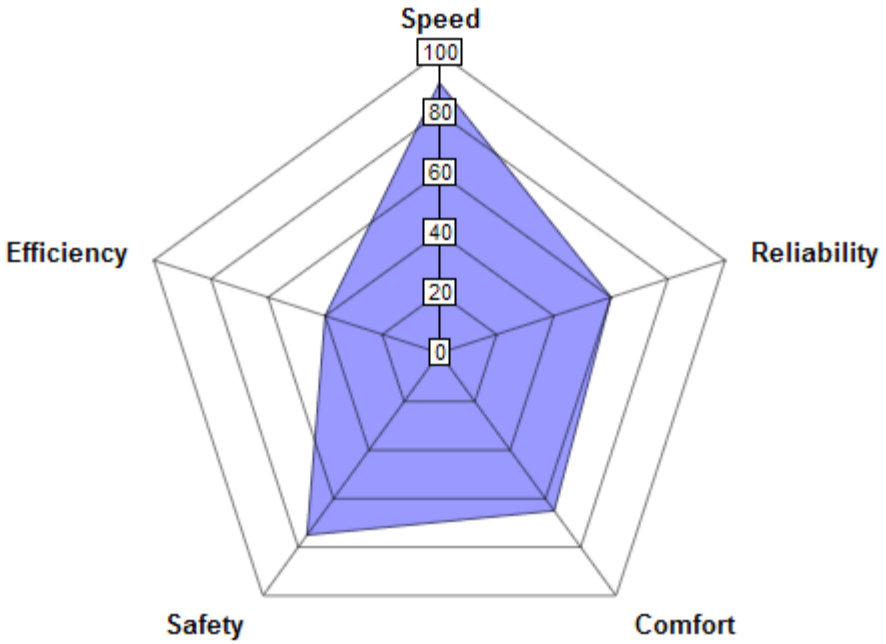
Dendrogram

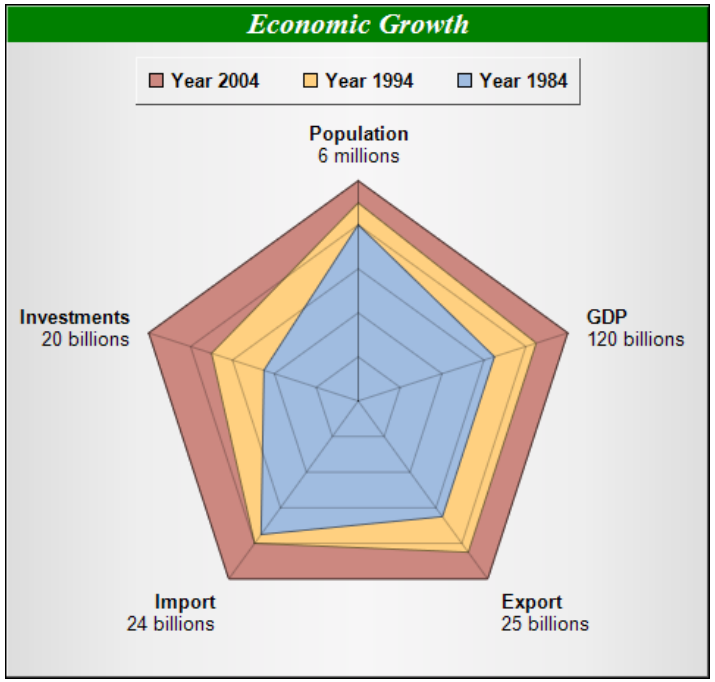
› suited for hierarchical graph visualization
  – including trees and hierarchies
  – Phylogeny analysis



Radar chart

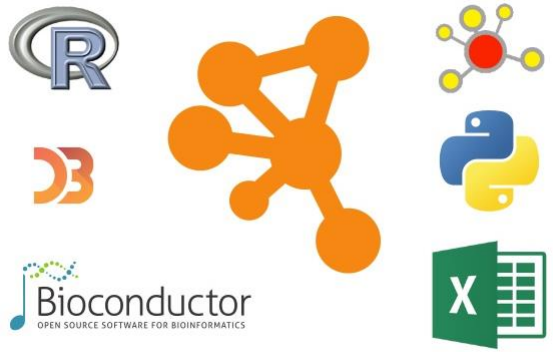
- displaying multivariate data
- in the form of a two-dimensional **chart**
- of three or more quantitative variables
- represented on axes starting from the same point





Visualization tools

- Plenty of
  - *in silico* visualization and analysis tools
    - provide user-friendly environment for the system biologist
    - Most of the tools are even equally applicable in other type of networks too
    - Web based and Desktop versions
    - Some general and some specialized
      - for example for GRN



Data Visualization Softwars

- Cytoscape
- Gephi
- Pajek
- Python
  - NetworkX
  - iGraph
- R
  - iGraph
  - ggplot2
- Netlogo
- UCINET
- NodeXL
- ...

Cytoscape

- Open source desktop software
  - complex network analysis and visualization tool in life sciences
- Cytoscape *core* distribution provides
  - a basic set of features for
    - data integration, analysis, and visualization
  - Many additional features are available as plugins
- Extendable open API based on Java™ technology
  - Most of the Apps are freely available from Cytoscape App Store.

  - Sample Intro
https://www.youtube.com/watch?v=iGpxX0Kd4Z0

Cytoscape

- most suitable for large-scale network analysis
  - since it can handle thousands of nodes and edges and still run smoothly
- supports directed, undirected, and weighted graphs
- comes along with powerful visual styles
  - allowing users to change the properties of nodes or edges
- Plenty of elegant layout algorithms for visualization
- Expression data
  - can be mapped as node color, label, border thickness, or border color
- Web version Cytoscape.js

Pajek

- make analysis and visualization of large scale networks
- supports abstraction by recursive factorization of a large network into smaller networks
- powerful visualization tools and implements some algorithms for analysis of large networks

- uses several data structures as:
  - (1) Network, The main object, which includes vertices (nodes) and lines (links).
  - (2) Permutation, Reordering of vertices.
  - (3) Vector, A set of values of vertices.
  - (4) Cluster, subset of vertices.
  - (5) Partition, Shows for each vertex to which cluster the vertex belongs.
  - (6) Hierarchy, Hierarchically ordered clusters and vertices.

Pajek

- conduct simplifications and transformations
  - like deleting loops, multiple edges, transforming arcs to edges, etc.
  - calculate components (strong, weak, biconnected, and symmetric components),
  - make decompositions (symmetric-acyclic, hierarchical clustering),
  - find paths (shortest path(s), all paths between two vertices),
  - calculate flows (maximum flow between two vertices),
  - make neighborhood analysis (*k*-neighbors)
  - ...

Gephi

- open-source and multiplatform software
- visualization and exploration software for all kinds of graphs and networks.
- Like Photoshop™ but for graph data,
  - the user interacts with the representation, manipulate the structures, shapes and colors to reveal hidden patterns
- help data analysts to make hypothesis, intuitively discover patterns, isolate structure singularities or faults during data sourcing.
- Many visualization features
  - Networks up to 100,000 nodes and 1,000,000 edges
  - Iterate through visualization using dynamic filtering
  - Rich tools for meaningful graph manipulation

Netlogo

- multiagent programmable modeling environment,
  - provides frameworks of computer codes of models/algorithms
  - suggestions for extending these models/algorithms
- It has built in implementation for
  - *Network algorithms*
    - Giant component model, Diffusion in a directed network, Preferential attachment, Team assembly
  - Computer algorithms
    - PSO, Cellular automata, Genetic algorithm
  - Bio and bio-inspired algorithms
    - Ant colony, Tumor model, Termites model, Wolf–Sheep Predation, Crystallization Basic (Chemistry)
  - Social algorithms
    - …
- Farsi help

NodeXL

  - https://nodexl.com
- gives data-driven marketers
  - access to powerful social media analysis features
    - including influencer identification,
    - brand evaluation,
    - social listening,
    - content analysis, ideation,
    - lead harvesting,
    - competitor social and campaign analysis,
    - automation, white-labeling,
    - and more!
- Integrate with excel
- Both Microsoft Excel and Google Sheets
  - also have many visualization options

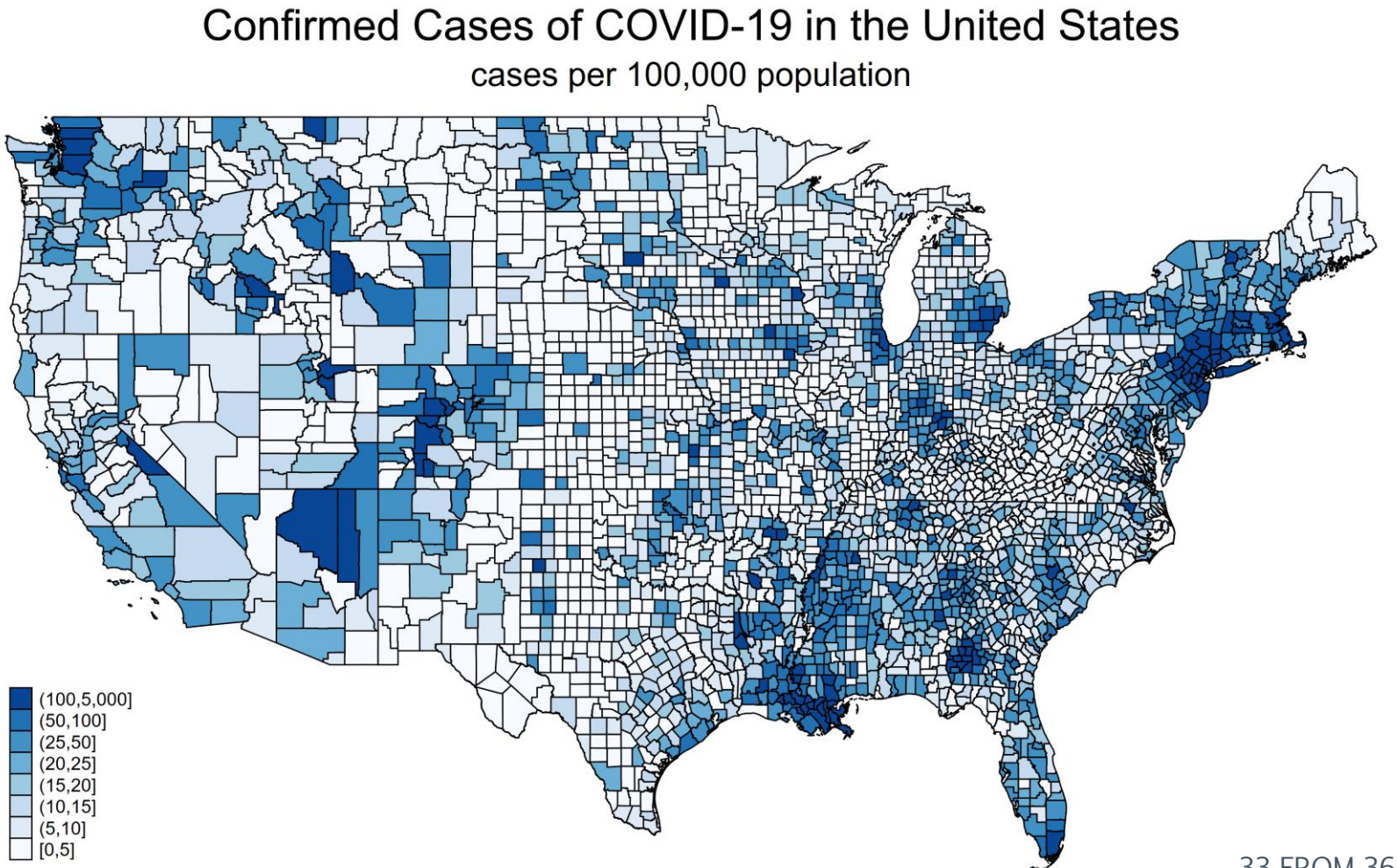
Challenges

- Data visualization tends to be like artwork
  - everyone has a favorite style and method
  - So there are very few standards

- Taking time to find or create a good visualization
  - may be tedious
    - but pays off in generating better science through understanding the data
    - discovering novel patterns which generate hypotheses
    - Better reporting of research findings

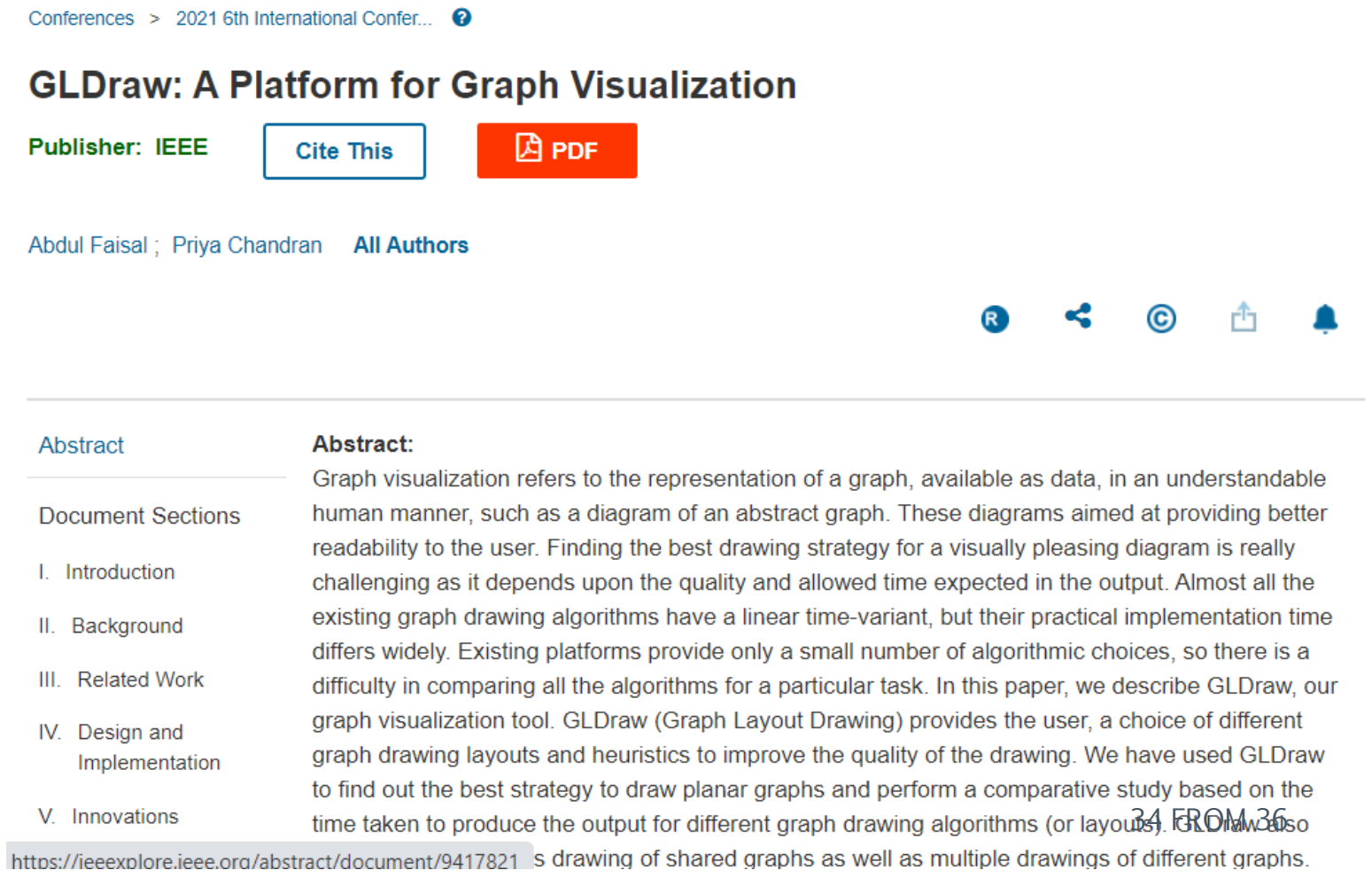
Example

› Choropleth map



Recent papers and trends

- Biological network visualization tool development
  - is itself a research field



Conferences > 2021 6th International Confer...

GLDraw: A Platform for Graph Visualization

Publisher: IEEE Cite This PDF

Abdul Faisal ; Priya Chandran **All Authors**

Abstract

Document Sections

I. Introduction

II. Background

III. Related Work

IV. Design and Implementation

V. Innovations

Abstract:

Graph visualization refers to the representation of a graph, available as data, in an understandable human manner, such as a diagram of an abstract graph. These diagrams aimed at providing better readability to the user. Finding the best drawing strategy for a visually pleasing diagram is really challenging as it depends upon the quality and allowed time expected in the output. Almost all the existing graph drawing algorithms have a linear time-variant, but their practical implementation time differs widely. Existing platforms provide only a small number of algorithmic choices, so there is a difficulty in comparing all the algorithms for a particular task. In this paper, we describe GLDraw, our graph visualization tool. GLDraw (Graph Layout Drawing) provides the user, a choice of different graph drawing layouts and heuristics to improve the quality of the drawing. We have used GLDraw to find out the best strategy to draw planar graphs and perform a comparative study based on the time taken to produce the output for different graph drawing algorithms (or layouts). GLDraw also

https://ieeexplore.ieee.org/abstract/document/9417821 s drawing of shared graphs as well as multiple drawings of different graphs.

Study more

REVIEW article

Front. Bioeng. Biotechnol., 31 January 2020 |
https://doi.org/10.3389/fbioe.2020.00034

Check for updates

A Guide to Conquer the Biological Network Era Using Graph Theory

Mikaela Koutrouli[1†], Evangelos Karatzas[1,2†], David Paez-Espino[3] and Georgios A. Pavlopoulos[1*]

[1]Institute for Fundamental Biomedical Research, BSRC "Alexander Fleming", Vari, Greece

[2]Department of Informatics and Telecommunications, University of Athens, Athens,

Question?

Bioinformation.ir

info@Bioinformation.ir