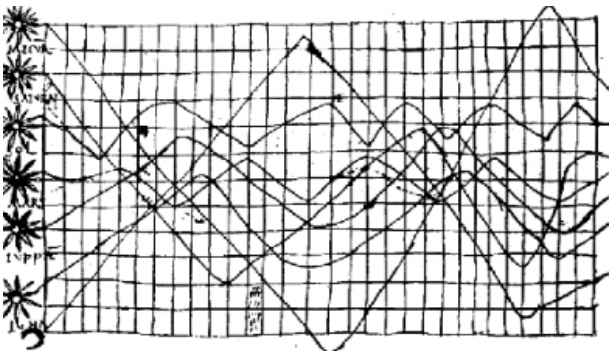
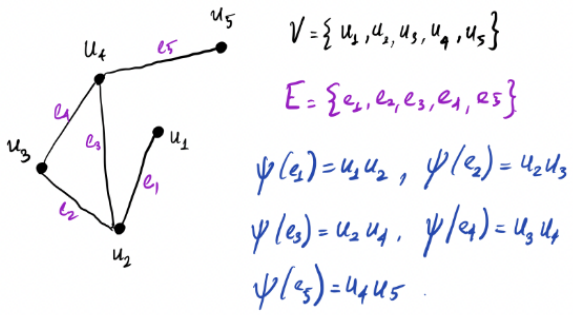
CAS CS 365

InClass Note 1

1. Data visualization
    a. Positions of Sun, Moon, and Planets throughout the year
    b. 
    c. Galileo's visualization of Sunspots
    d. 

2. What is a graph?
    a. A graph G is a triple (V_G, E_G, ψ_G) consisting of:
        - Set of vertices/nodes V(G)
        - Set of edges E(G), disjoint from V(G)
        - Incidence function ψ that associates with each edge a pair (not necessarily distinct) nodes:
            - Unordered pair - undirected graph
            - Ordered pair - directed graph

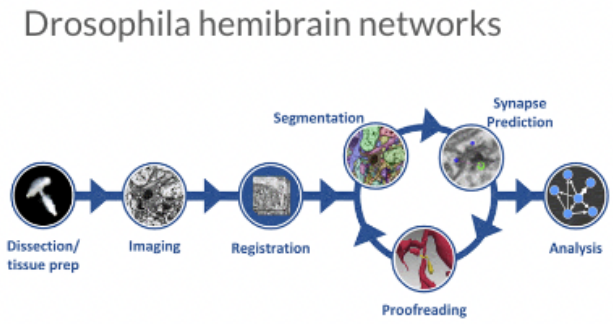
b. If ψ(e) = (u,v) = uv, u, v are endpoints of e that joins them. Nodes u and v are called adjacent nodes

c. 

d. In-degree: number of incoming edges

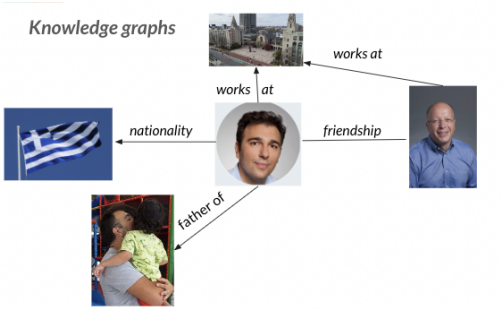
e. Out-degree: number of outcoming edges

3. Graph model a wide variety of datasets

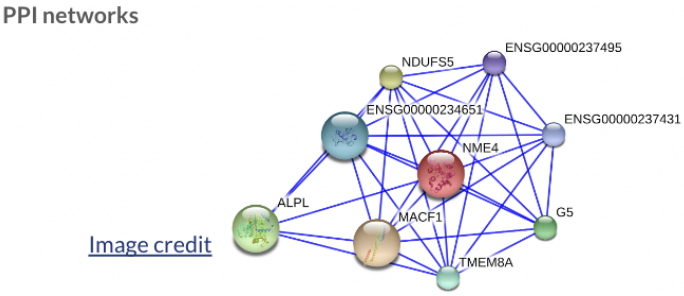
a. 

b. 

c.

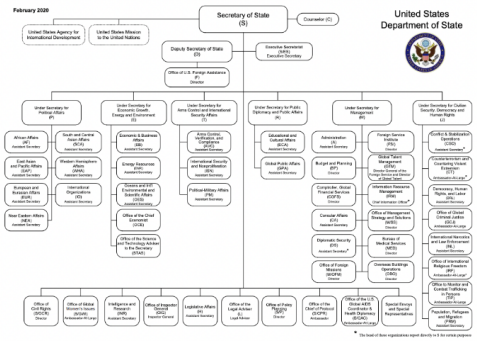


d.



4. Trees: connected graph with no cycles

a.

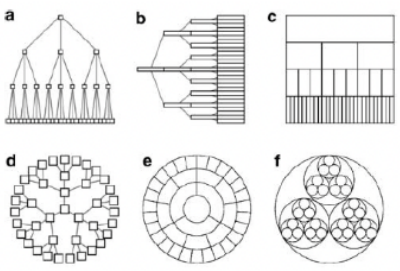


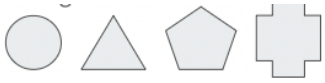
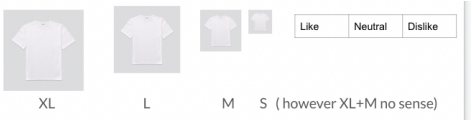
5. Remark

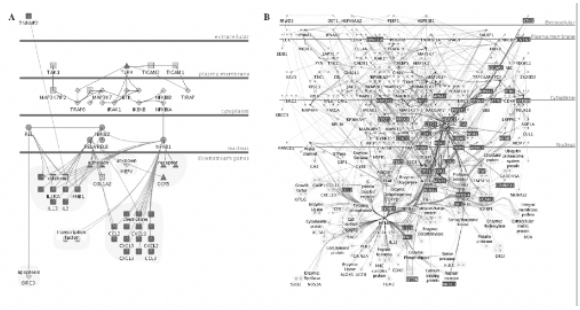
   a. Datasets can be visualized in a verity of ways

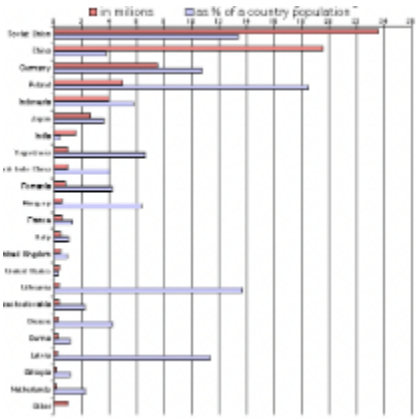
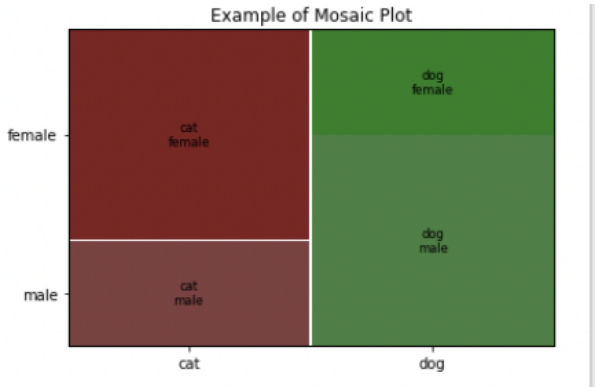
   b. For example, a complete 3-ary tree of depth 3 can be visualized as follows:

-



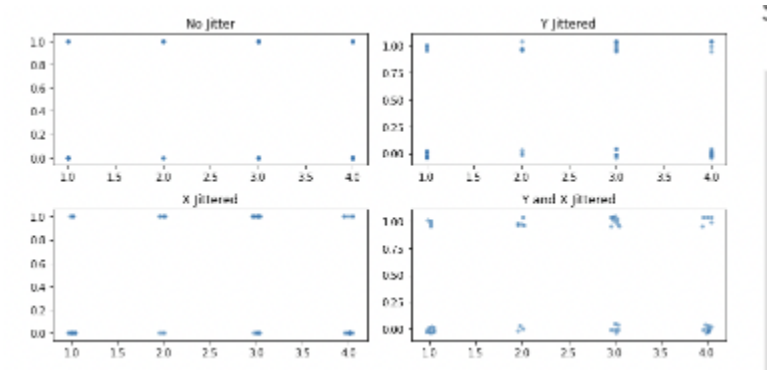
6. Dataset availability and types of attributes
    a. Static vs dynamic
        - Dynamic dataset is a stream of updates (consider a table with items/tuples can be updated, deleted)
    b. Attribute Types
        - Categorical
            -  , marital status, gender, religion etc.
        - Ordinal (sometimes also classified under quantitative)
            - 

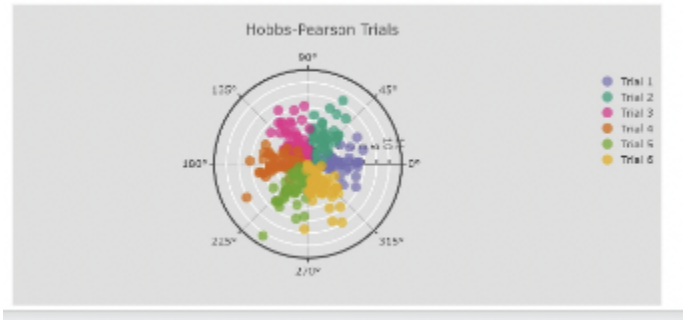
                - Quantitative data is information that can quantified: measured or counted, and thus be given a numerical value
7. Visualization
    a. Computed-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
    b. Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods
8. Visualizing graphs
    a. 

9. Mistakes to avoid when visualizing data
    a. A good display conveys information. The less information in the display, the worse it is!
10. Categorical data - Bar chart
    a. A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent
    b. 
11. Categorical data - Mosaic plots
    a. Graphical method for visualizing data from two or more categorical variables
    b. 

    c. They can be used to visually contrast distribution of empirical data
12. Quantitative data - scatterplot

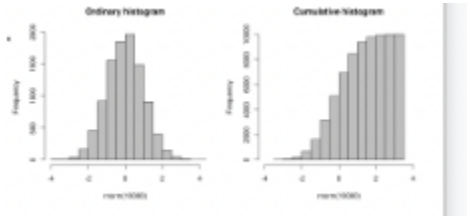
a. Some important types of plotting univariate continuous data

b. 

c. 

13. Quantitative data - Histogram

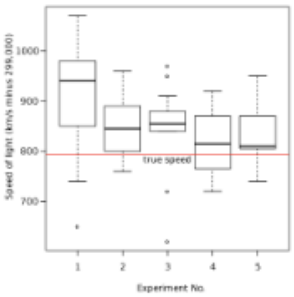
a. Great for visualizing a distribution

b. 

c. Weakness: detection of outliers

d. Choosing the number of bins, their width and the anchorpoint can have a major effect on the visualization, and hence on the information conveyed by a histogram

14. Quantitative data - Boxplot



a.

b. Box plot is a method for graphically demonstrating the locality, spread and skewness groups of numerical data through their quartiles

c. Great for detecting outliers and comparing distributions

d. Mixture of summary information as histogram and of individual points as dotplots