NOTES

WEEK 1

Introduction to Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. This involves interaction with visual representation of data with the aim of changing the course of action. It is required in many areas such as security, healthcare, finance, environment, education, etc. It is needed where the decision makers need to come up with decisions fast.

Visualization is the pictorial representation of data. Interaction is the method to alter the visual representation based on the user requirement in real time.

Elements of Visual Analytics

There are a total of 4 key elements to visual analytics which are:

- Analytical reasoning
- Visual representation and interaction
- Data representation and transformation
- Production, presentation and dissemination

The process of visual analytics goes like:

1. Data representation and transformation
2. Visual representation and interaction
3. Analytical reasoning
4. Production, presentation and dissemination

Analytical Reasoning

It is the method to obtain deep insights. It involves one to understand the past and present situation quickly, ascertain or predict to a certain extent the alternative futures based on present and past data, monitor emerging events including uncertain or unexpected events, and any other such tasks which can affect the decision making process. The role of Analytical reasoning is as follows:

- Increased resources
- Reduced search
- Enhanced recognition of patterns
- Perceptual inference
- Perceptual monitoring
- Manipulable medium

Principles of depicting Information

- Appropriateness principle
    - Should provide neither too less nor too much information for the task at hand.
- Naturalness principle
    - Properties of visual representation should match information being represented.

- Matching principle
    - Should present affordances suggestive of appropriate action.
- Principle of congruence
    - Represent important concepts in the domain of interest.
- Principle of apprehension
    - Structure and content should be readily and accurately perceived and comprehended.

Interactions in Visualization

- Filtering – Modifying data transformation through interaction
- Visual mapping – Modifying visual representation through interaction
- Navigation – Moving through data space through interaction
- Human info discourse – Analytical process through interaction

Data representations and transformations

Data representation is a structured form of data suitable for computer based transformation. Certain characteristics of data representation are:

- Data types – numeric vs non-numeric
- Levels of structure – structured vs unstructured
- Geospatial – georeferenced numeric vs non-numeric
- Temporal – data may change over time

When the original data is not appropriate for visualization then it needs to be converted into different representations which can be then visualized.

Production, presentation and Dissemination

Production is the process of summarizing the results obtained through the analytical processes. Presentation is when all the results produced by the "Production" process are packaged together to form a contextualized artifact meaningful to the target audience. Dissemination is the process of effective sharing of the packaged presentation efficiently to all the relevant parties.

Human Visual System

Visualization does not just relate to visual sensory inputs, well most of the time it does, but it also refers towards all the other 5 senses of visual, auditory, tactile, olfactory and taste.

Data Driven vs. Concept Driven stages

Data driven stage is also known as the template scheme wherein from the given data we try to find out a known template. Concept driven stage is as the name suggests, a concept driven process where one starts with a given concept and then try to make sense of the data based on the concept.

Perception of Design

The basic Gestalt psychology of how humans group together similar entities, recognise patterns. This leads us to the key principles in visualization:

- Similarity
  - Entities have similar visual attributes such as shape and color are perceived as the same cluster or group and have similar functions.
- Proximity
  - Entities which are close together can be seen as a single group and it overrides the similarity concept.

- Common region
  - Entities which are surrounded by a same closed region can be seen as belonging to the same group despite differences in visuals.
- Closure
  - An entity whose part is missing can actually be completed mentally.
- Continuity
  - Entities that are arranged in a line or a curve are more perceived than those which are not.
- Connection
  - Connected entities are perceived as groups and override the proximity and similarity concept.

WEEK 2

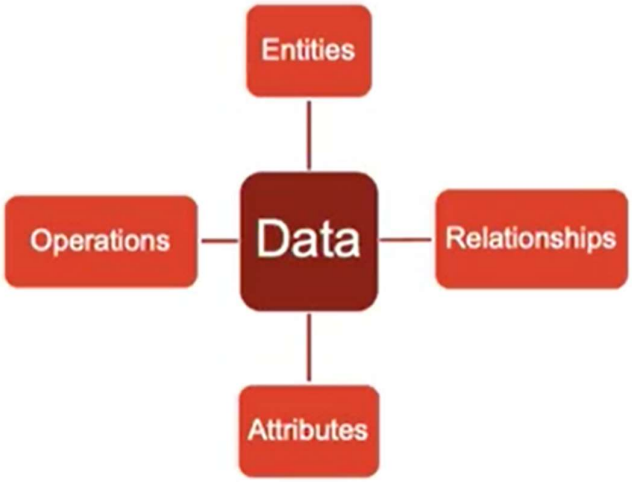
Data

There are different types of data such as:

- Categorical
  - Takes fixed number of values. These are of 2 types:
    - Nominal
      - No numeric value
    - Ordinal
      - Categorical data that is converted into numeric
- Interval
  - Can go to negative
- Ratio
  - Cannot go to negative
- Qualitative
  - Measures of types
  - May be represented as a name, symbol or number code.
  - Data about categorical variables
- Quantitative
  - Measures of values or counts
  - Expressed as numbers
  - Data about numeric variables
- Structured
  - Organised information in the database.
  - Can be fit in a spreadsheet

    - Easier to handle
- Unstructured
    - Freeform information
    - Cannot be fit into spreadsheet
    - More difficult to handle
    - We can visualize these using ML (Machine Learning) or NLP (Natural Language Processing)
- Metadata
    - It is a derived data
    - It describes another data

Data Models



- Entities
    - Objects of interest/value
    - These can be single (Eg: People, hurricane, etc.)
    - These can also be a group (Eg: a school of fish, etc.)
- Relationships
    - Structures that relate entities
    - Can be:
        - Structural and physical
        - Conceptual
        - Casual
        - Temporal
- Attributes
    - A property of some entity
    - Eg: Color of a fruit, Temperature of water, Duration of trip, etc.
    - It is either a dimension or a quality
        - Dimension
            - Can be scalar (1D, Eg: weight of a person)
            - Can be vector (2D, Eg: direction of travel)
            - Can be tensor (3D, Eg: direction + shear force)
            - Can be a field of scalar/vector/tensor (4D, Eg: gravitational field)
        - Quality

- There are four levels of measurements
    - Nominal – labelling purpose
    - Ordinal – can be ordered in a sequence
    - Interval – able to derive gap between values
    - Ratio – full expression of real number
- Category data (Nominal scale) is represented as string or enum data type in programming
- Integer data (Ordinal scale) is represented as int or long data type
- Real number data (Interval & ratio scale) is represented as float or double data type.

- Operations
    - Processes performed on data are called operations.
    - There are different types of operations such as:
        - Mathematical operations
        - Merging lists
        - Inverting a value
        - Creating an entity
        - Deleting an entity
        - Transforming a data
        - Forming a data
        - Splitting an entry

Color Characteristics

- Perception
    - Physical Detection
    - Psychological Perception

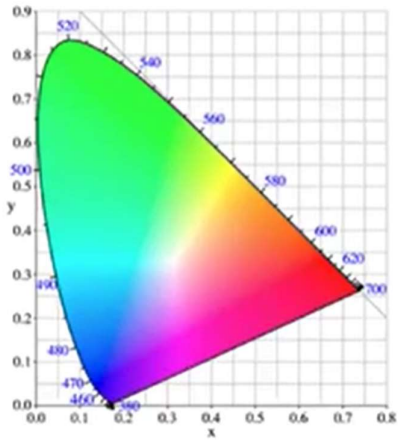
Psychological (Visual) variable	1st order Physical variable	2nd order Physical variable
Brightness	light intensity	wavelength, adaptation of eye
Hue	wavelength	spectrum structure, peripheral light intensity and wavelength
Vividness /Saturation	Spectrum structure	peripheral light
Contrast	Intensity, wavelength, peripheral	

Color spaces and systems

The Commission Internationale de l'Eclairage (CIE) system

CIE-XYZ

X: non-negative CIE RGB value,
Y: luminance,
Z: equivalent to Blue

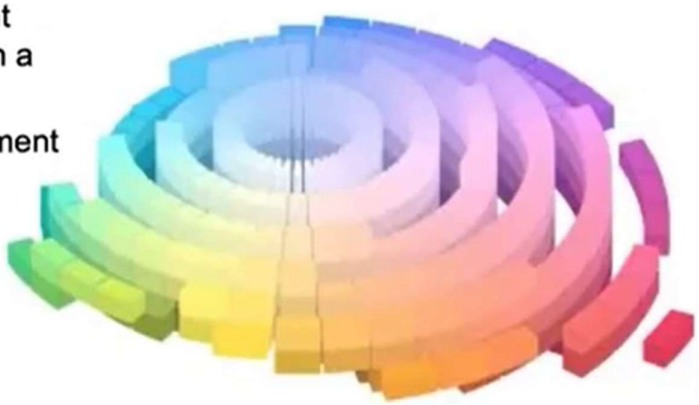


$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z}$$
$$z = \frac{Z}{X+Y+Z} = 1 - x - y$$

Wavelength defines the hue.
From the white point outwards to the coloured point defines the chroma.

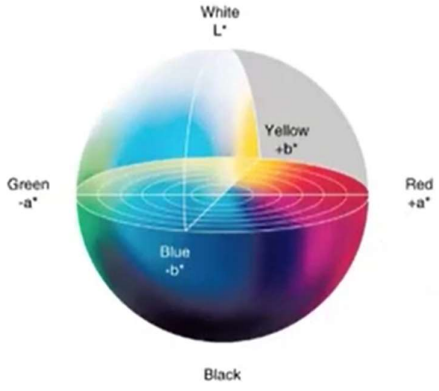
Munsell colour system (1905 by Albert Munsell)

- provides a set of standard color chips designed to represent equal perceptual spacing in a three-dimensional mesh
- provide a physical embodiment of a uniform color space



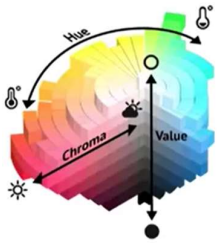
CIELAB colour space

- Based on opponent colour model
- Less uniform in colour axes, but useful for predicting small differences in colour



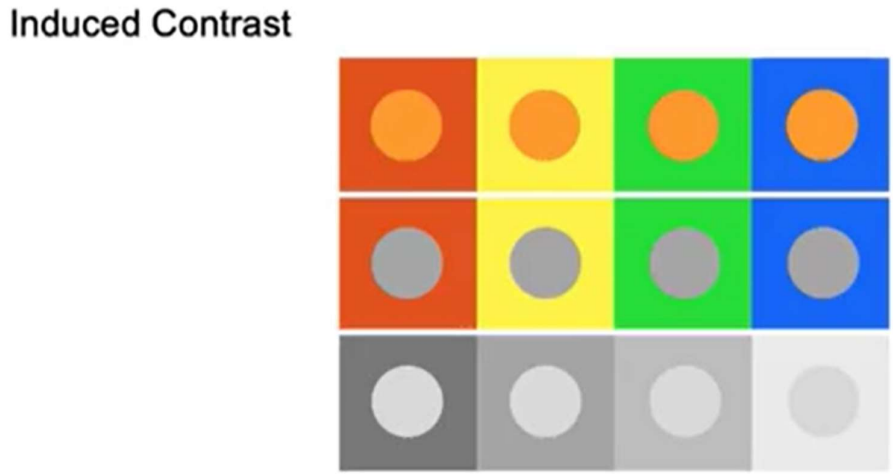
Color components

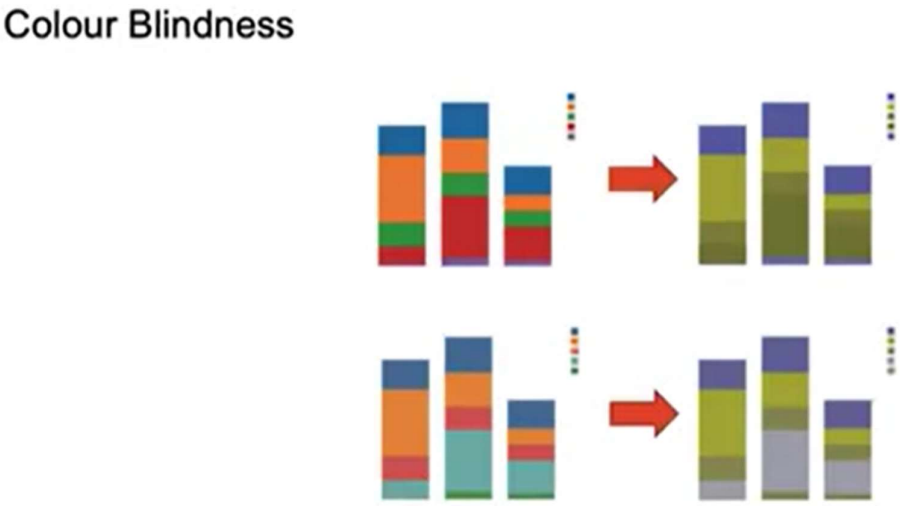
- Hue: Wavelength
- Chroma (Saturation): Amount of white
- Brightness (Value): Light intensity



HSV stands for Hue, Saturation, Value. HLS stands for Hue, Lightness, Saturation.

Color Perception





Applications of Colour in Visualizations

- Colour mapping in 3D visualization (Volume visualization)
- Cartography

WEEK 3

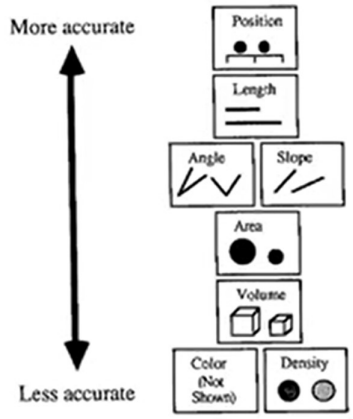
Introduction

Jacques Bertin was a French cartographer who took into consideration the following:

- Printable on a white paper.
- Everything should be visible at a glance.
- Should have a reading distance of a book or an atlas.
- People should be able to view the map under normal and constant lighting.
- The maps should be readily available to the people.

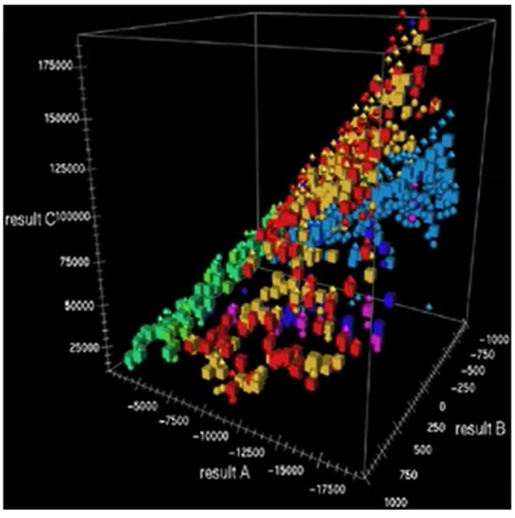
Semiology of graphics is an approach of graphic representation. The functions of graphic representations are as follows:

- Recording information
- Communicating information
- Processing information



Visualizations

A graph can have multiple dimensions. For example, for the below image, it has 6 dimensions: X, Y, Z as the position dimensions, Color, Shape and Size.



Current Visualization Production

1. Drawing applications (Graphic Designers)
    a. Photoshop, Illustrator, etc.
2. Graphic Libraries (Graphic programmers)
    a. OpenGL, WebGL, etc.
3. Visualization libraries
    a. D3, Matplotlib, etc.
4. Integrated Environment
    a. Tableau, PowerBI, etc.

Semiology of Graphs

In semiology of graphics, the plane is continuous, homogenous and a 2 dimensional space.

Implantation of a plane is when you put something on the plane to represent data. This can be done by:

- Point
    - Represents a location on a plane without any distance.
    - Example: position of an airplane.
- Line
    - Signifies a phenomenon on plane with measurable length but no area.
    - Example: Boundary of a continent.
- Area
    - Signifies something on the plane with measurable size.
    - Example: Island, lake on a map.

Sign-fields of a plane are the resemblance, order and proportional in graphics. These sign-fields are transcribed by visual variables having the same signifying properties.

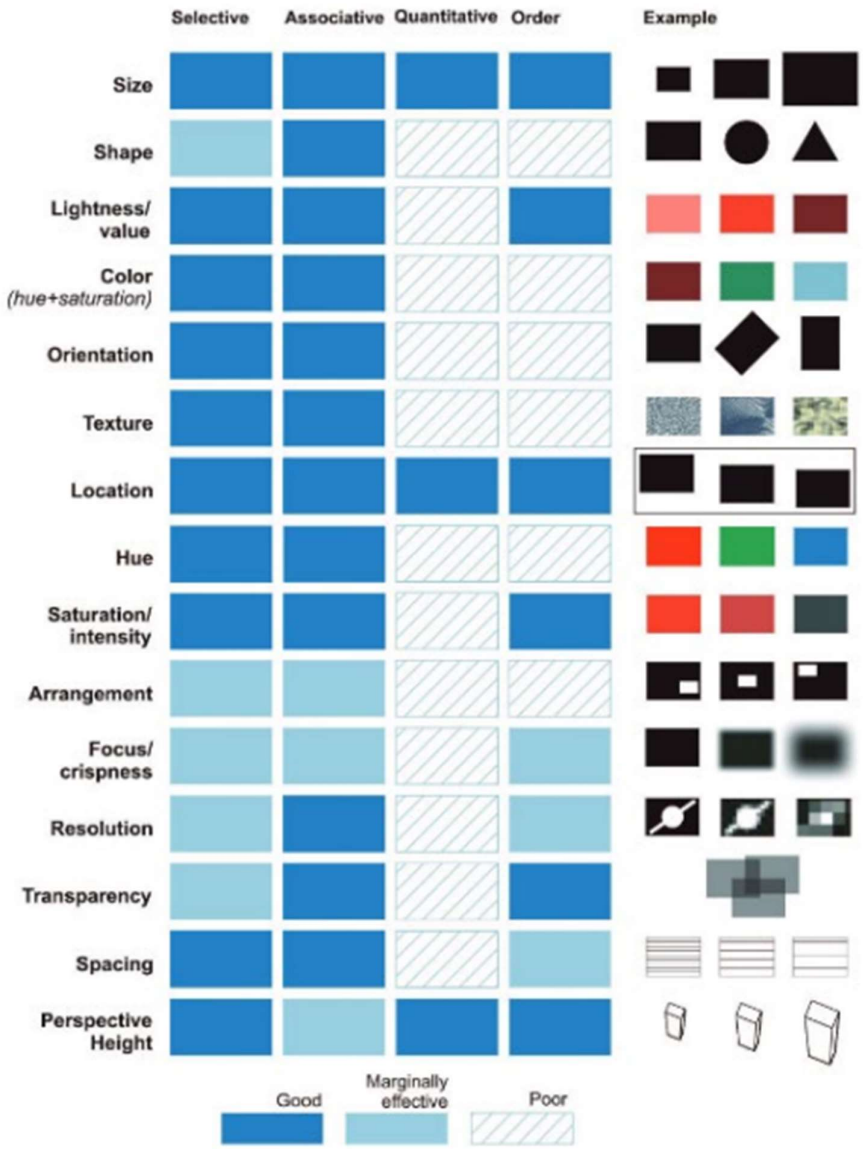
Visual Variables

Visual (Retinal) variables are necessary in all graphic problems involving three or more components. These are used when 2 dimensions are already in use as the X & Y coordinates. The visual variables are shown as:

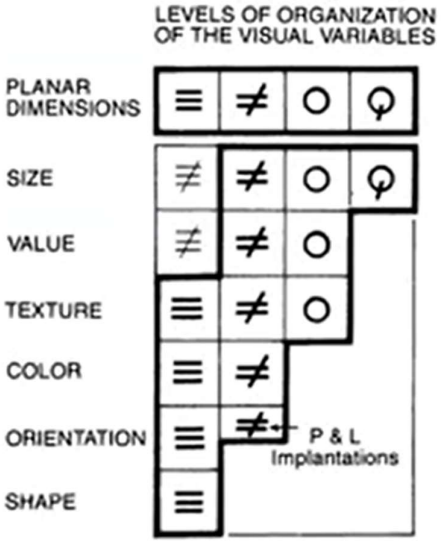
There are 6 identified visual variables which are:

- Size
    - Changes in height (length)
    - Changes in area of the sign
    - Changes in the number of equal signs (repetition)
- Shape
    - A mark with a fixed size can have infinite number of shapes
- Location

- Orientation
    - Various orientations of line or line patterns
    - These range from vertical to horizontal in a distinct direction
- Color
    - Changes in coloured sensation at equal value
- Texture (Pattern)
    - Changes in fineness or coarseness of the constituents of an area of given value
- Value (Lightness)
    - The various degrees between black and white
    - Changes from light to dark or vice-versa
- Saturation (Intensity)
- Focus (Crispness)
- Resolution
- Transparency, etc.



Levels of Organisation

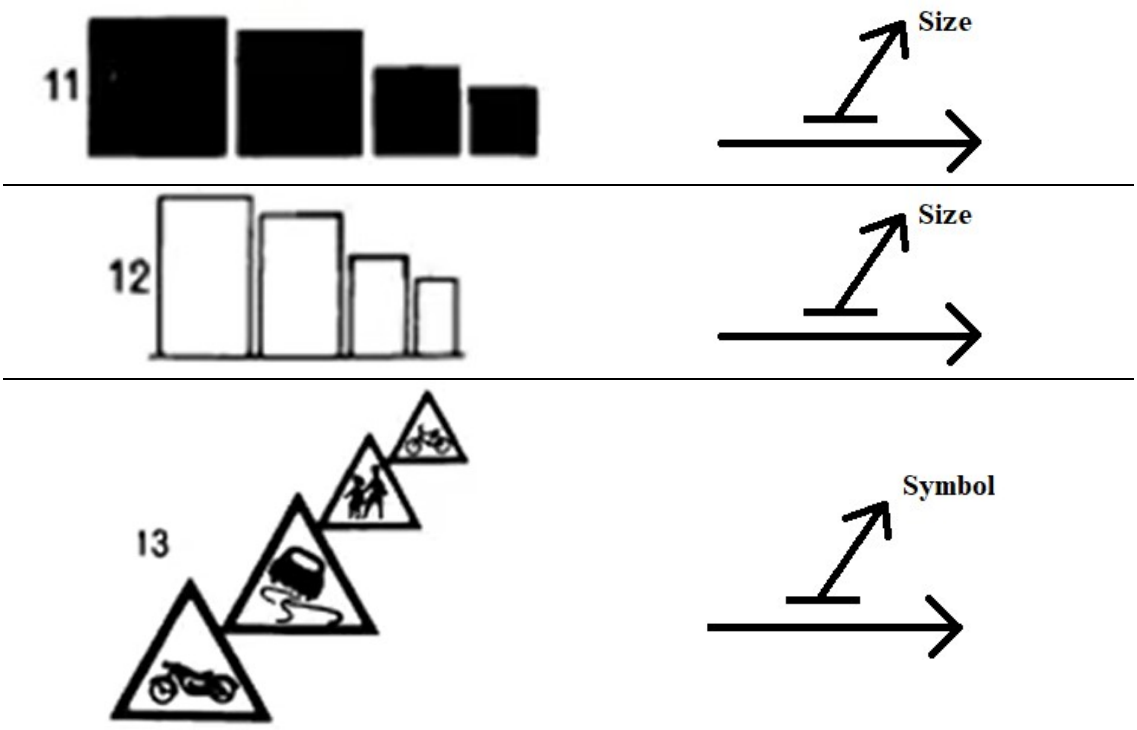


Note: PLA = Point, Line, Area

- Associative perception (=)
  - Any visual representation can be perceived as similar.
  - This is represented as ≡
  - One can use Shape (PLA), Orientation (PLA), Texture (PLA) and Color (PLA) to represent this perception.
- Selective perception (≠)
  - Any visual representation can be perceived as different.
  - One can use Orientation (PL), Texture (PLA), Value (PLA), Size (PLA) and Color (PLA) to represent this perception.
- Ordered (O)
  - Any visual representation can be perceived in a sequence.
  - One can use Texture, Value (PLA) and Size (PLA) to represent this perception.
- Quantitative (Q)
  - Any visual representation can be perceived as proportional to each other.
  - Once can use only Size (PLA) to represent this perception.

Rectilinear Elevation

Usually the visual representation would be in 2 dimensions, that is the X and Y axis. The use of rectilinear elevation is to denote something popping out of the 2 dimensions, or essentially the other dimensions of the visual representation. A few examples of it is given below:



Circular Construction

For circular representations such as Pie chart, we use the above construction. The inner arrow with the dash represents cumulative frequency. The outer arrow represents the value denoted in the Pie chart.

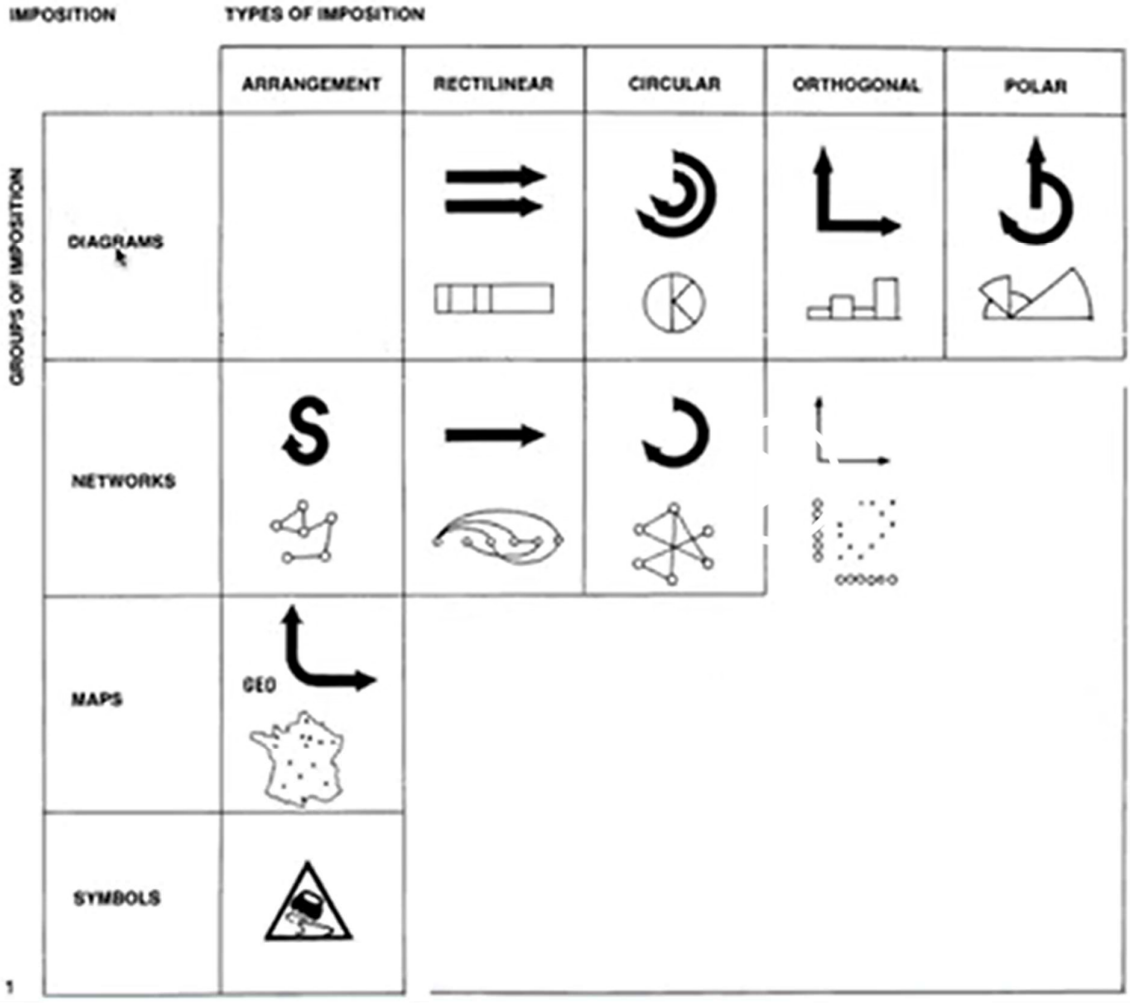
Polar Construction

These are used when we have something in the circular axis but we are using a radial direction. The circular arrow represents the circular axis of the visual representation whereas the straight arrow represents the outward axis, the other dimension.

Circular Elevation

These are used when the data points lined up on the circle. The slant arrow represents the visual variable which can be shape, size, color, etc. and the circular arrow denotes the circular plane.

Summary of the impositions



Homogenous Dimension

This arrow represents the values to be in 1 dimension, meaning that the values are homogenous.

Heterogeneous Dimension

This arrow represents the values to be in more than 1 dimension, meaning that the values are heterogeneous.

Homogenous Dimension with Heterogeneous sub-divisions

This arrow represents the values are repeated n times in the dimension.

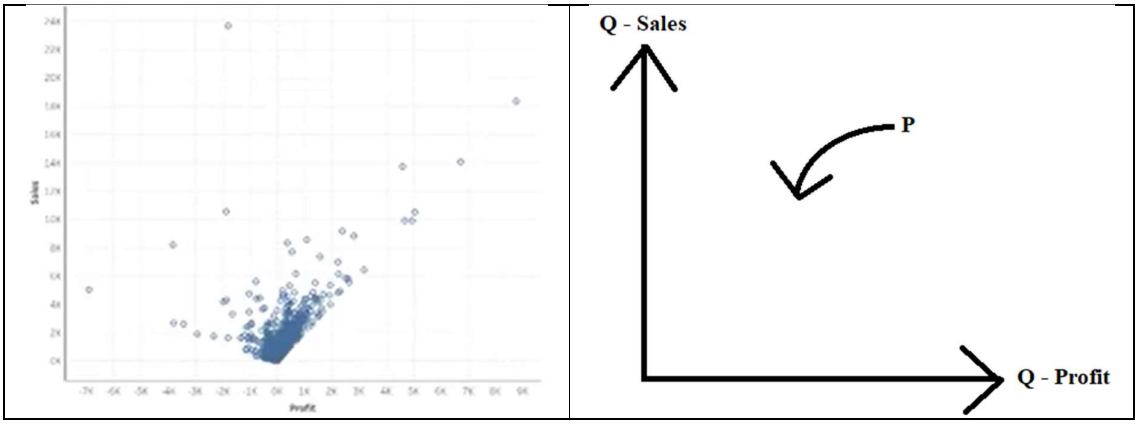
Cumulated Dimension

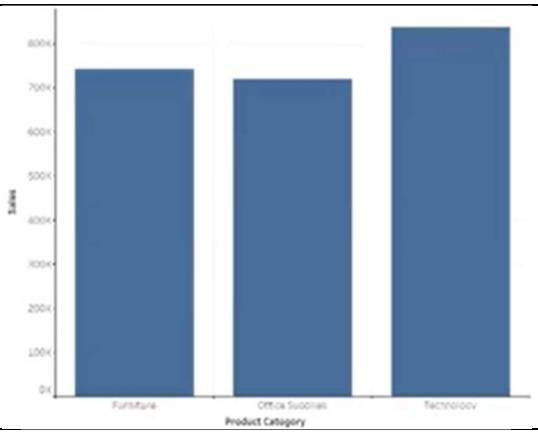
This arrow represents that the quantities are stacked together.

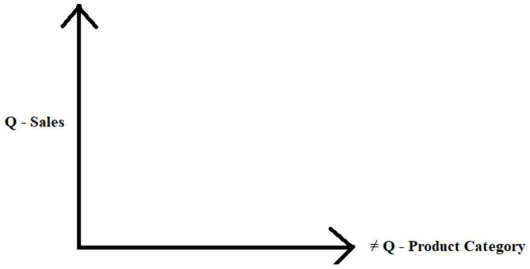
Arrangement, Tree (Network)

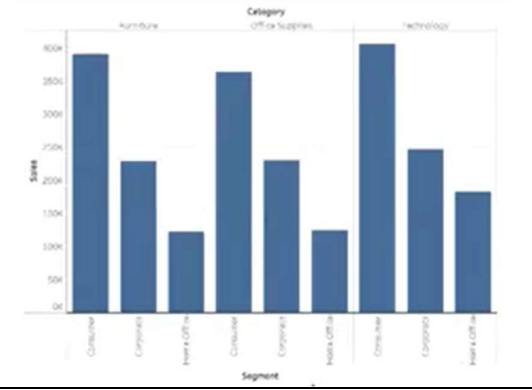
These are used when axes are neither linear nor circular. The second one with the horizontal line depicts when dimension is also observed to be in horizontal order or a hierarchy.

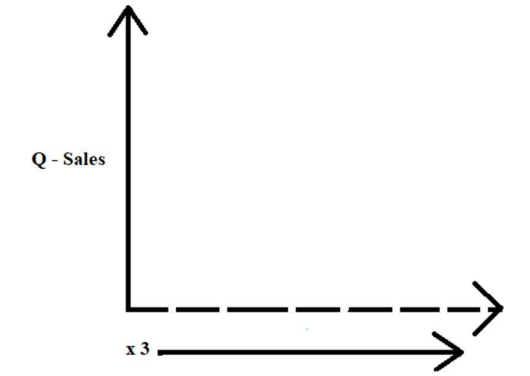
Examples for the Visual representations

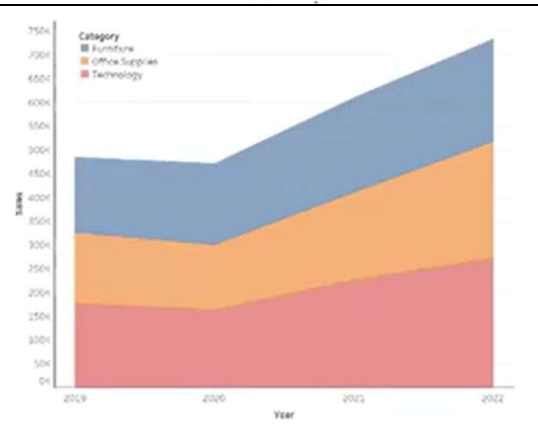


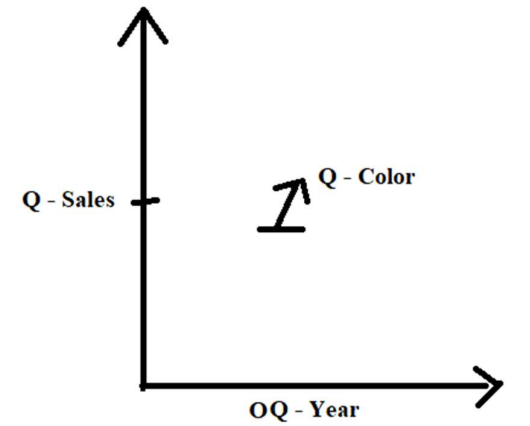

Sales
Product Category

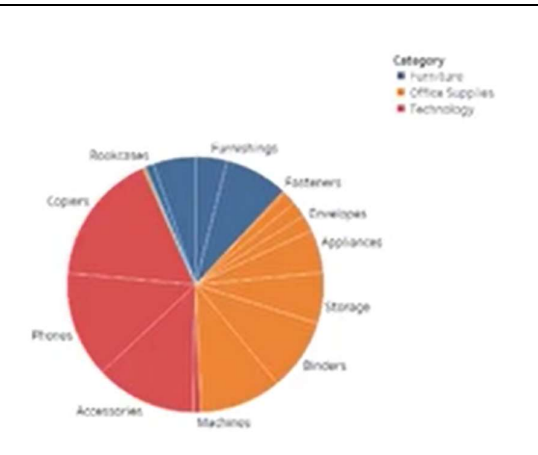

Q - Sales
≠ Q - Product Category

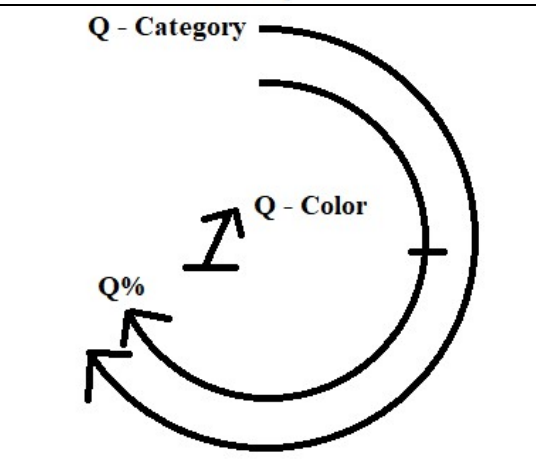

Category
Sales
Segment

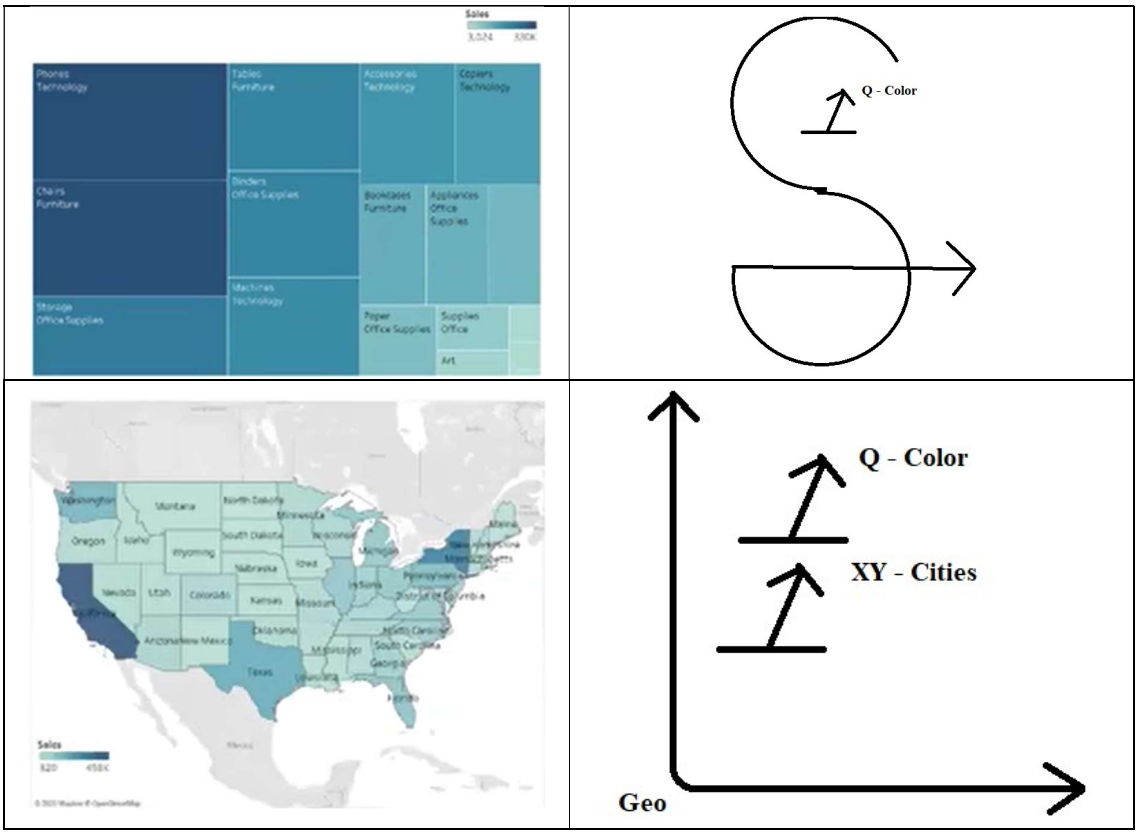

Q - Sales
x 3


Category
Furniture
Office Supplies
Technology
Sales
Year


Q - Sales
Q - Color
OQ - Year


Category
Furniture
Office Supplies
Technology
Bookcases
Furnishings
Fasteners
Envelopes
Appliances
Storage
Binders
Machines
Accessories
Phones
Copiers


Q - Category
Q - Color
Q%



WEEK 4

Purpose of Human-Computer Interaction

- Advancing human capabilities
    - The progress of capabilities had the following order:
        - Calculation (Eg: Abacus)
        - Simple memorization (Eg: Cameras)
        - Associative memory (Eg: Flashcards)
        - Language transition (Eg: Google translate)
        - Problem solving (Eg: Mathematica, Wolfram)
        - Automatic proofing (Eg: Plagiarism checkers)
        - Enhancing human cognitive capabilities
        - Requires understanding of human activities
- Advancing human communication
    - This is normally combined with the advances of the network technologies
    - Provides new media for human-human communication (Computer mediated communication). Eg: Translators, Chat-bots.
    - Unlike non-computer mediated communication, various services can be provided to enhance communication
- Advancing assistance capabilities
    - Many audio-visual products can be interconnected with networks
    - Perceptual user interfaces allow computers to sense our activities in order to provide us with adaptive services.
    - Eg: Smartwatches like apple watch that allows us to access the internet anytime.

- Improving our experience
    - With computer assisted artificial systems we try and improve our own experience.
    - These are characterized by not only their functional capabilities but also by how they can improve our physical/mental experiences.

Types of Interactions

Interactions occur between:

- Human and Human (HHI)
    - Eg: Human physical communication
- Human and Computer (HCI)
    - Eg: Touchscreens
- Human and Machine (HMI)
    - Eg: Driving a car, Elevators, etc.
- Human and Information (HII)
- Human and all (HXI)

Modality of Interaction

Interaction over here is the human interaction with artificial objects or computer systems (including both physical and virtual) in an ordered manner. Artificial system is a set of components, each of which interacts with the other in an ordered manner.

Modality of Interaction (Multimedia oriented)

There are basically 2 types of representation media which are language (Eg: Text, audio, voice, etc.) and non-language (Eg: other media). Following is a more detailed information:

- Language
    - There are 3 types of languages:
        - Natural Language – What we use for day to day communication
        - Artificial Language – Conveys information in a very simple and accurate manner. Suited for describing complex information in an orderly manner using common grammar and syntax. Eg: C, C++, C#, Java, Python, etc.
    - Eg: Text, audio (voice)
- Non-Language
    - Its purpose is to support or control the language media.
    - This type of media can convey information which cannot be easily expressed in language media.
    - There are a total of 9 types of non-language media:
        - Action
            - It is the expression by posture and movement of the human body.
            - Actions might have different meanings depending on the recipient's culture.
            - Eg: Facial expression.

- Peripheral language
    - Features associated with the conversational language. This includes the tempo, volume, accent, etc.
    - Through this the feeling, and even the personality can be conveyed.
- Silence
    - Eg: Pause in conversations, Intentional ignoring, etc.
- Time
    - This is a temporal factor which has a significant influence on communication.
    - The factor can be used in 2 different ways such as:
        - Timing of interruption in conversation
        - Timing to induce sleepiness
- Chroma
    - It is the color in the environment and this works on the human perception.
    - It can be used to control communication and action.
    - It subconsciously affects the perception of the recipient.
- Human body
    - Characteristics of human body expressed in terms of age, gender, physique, skin color, etc.
    - This is typically based on the information recipients cultural background.
- Body contact
    - This includes real body contact.
- Eye
    - This includes eye contact and eye expressions.
    - This is under the complete control of the information provider.
    - Facial expressions represent many emotional states of the users, supported by the gestures, it is difficult to control by the information provider consciously.
    - Eye contact and facial expressions can be used to control the timing of the speaker/listener change, monitor responses, display opinion or emotion and even express the attitude towards the recipient.
- Territorial Space
    - Interaction distances between people.
    - This is a spatial arrangement between participants which could be culturally biased.
    - In human communication, people are very sensitive to their territorial space.
    - Territorial space can be violated by:
        - Territorial pollution, Eg: verbal or physical
        - Unsolicited space use
        - Overtaking the territory
    - Personal space defines the comfortable interaction space.
    - Artificial systems can modify the personal space.
        - Eg: mobile phone usage in public space, etc.

- A few spaces:
    - Intimate distance: <45cm
        - Eg: Kids, Girlfriend, Wife
    - Individual distance: 45cm-1.2m
        - Eg: Individually interested parties
    - Social distance: 1.2m-3.6m
        - Eg: For group discussions
    - Public distance: >3.6m
        - Eg: Public speaking

- Body Language
    - There are multiple types of body languages:
        - Linguistic Body language
            - Associated with an apparent message.
            - Eg: Sign language, common gestures, etc.
        - Illustrative Body language
            - Assists with the conveyance of messages.
            - Eg: gesture to emphasize the point, etc.
        - Negotiation Body Language
            - Eg: Gesture for invitation, Patting on the shoulder, etc.
        - Ceremonial Body Language
            - Eg: Handshakes, bows, etc.
        - Body language to regulate verbal communication
            - Eg: nodding, touching chin, arm-crossing, etc.
        - Adapter Body language
            - Eg: yawning, scratching, etc.
        - Synchrony
            - Unconsciously imitate the person's body language when you agree with that person.
    - Eg: Facial expression, Body language, etc.

Human Information Processing Model

In cognitive psychology humans are treated as an information processing unit and the cognitive process is modelled as a series of information processes in this information processing unit. It is used to understand human cognitive processes such as inference, problem solving, memory and learning.

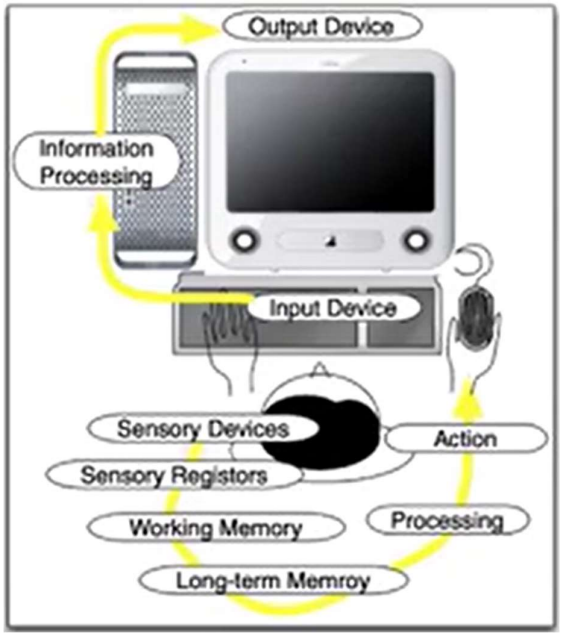
A computer has an input, a central processing unit and an output. Similarly, the human body has the 5 sensory devices for input, the brain for memory and central processing unit and our body including speech as the output.

There are very limited cons of the HIP model which is basically very little biologically equivalent explanations. It does have its own advantages such as being easily simulated and can even be used to evaluate functions and performances of the human interaction.

Human Memory Unit

- Sensory Register
  - Very short time storage for the 5 sensory devices.
- Working Memory
  - Temporary information storage.
  - Limited capacity and function.
- Long term Memory
  - Declarative memory
    - Episodic memory
    - Semantic memory
  - Procedural memory

Human Computer Interaction Modelling using Human Information Processing model



We have 2 information processing units: the computer and the human. We also have 2 sides which are connected by Input-Output devices. There are different models under this:

- D. Norman's Gulf Model
  - Sensing State of an Artificial System
  - Interpret the sensing results
  - Evaluate the interpretation against its own intention
  - (Re)-Set a new intention
  - Select a new input
  - Execute the new input
  - Execution by the Artificial System
  - This model is useful to analyse/explain various difficulties, ineffectiveness and breakdown.
  - Gulf prevents a seamless coupling of 2 execution systems
  - There are 2 gulfs: Execution (H to C) and Externalisation (C to H)

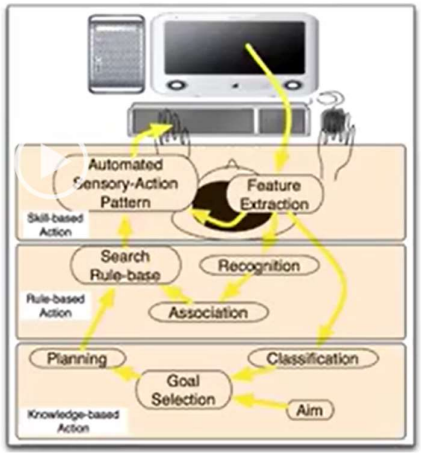
- Rasmussen's Decision Making Control Model

Three levels of hierarchical control model

1. Skill-based
2. Rule-based
3. Knowledge-based

Similar to "Subsumption Architecture" (Brooks@MIT)often used in Robot-Control.

- fast but low-quality low-level processing
- slow but high-quality high-level processing



- Minsky's Society of Mind: Agent based Model
  - Treats human minds as a group of inter-related agents
  - Provides many interesting possibilities to build the model of mind but has not been applied to practical models.
- Ecological Model
  - Closely related to Rasmussen's Decision Making Control Model.
  - Puts its focus on the environment rather than the human as an information processing unit.
  - Analyse how the environment offers relevant information to humans in order to assist any interaction.
  - It focuses on how well we can retrieve relevant information from the environment.
  - It does not analyse absolute physical characteristics the environment has but analyses the following:
    - How relevant information can be generated based on human needs.
    - How humans obtain such information to their actions.
    - How humans apply such information to their actions.

Affordance

A person will obtain pieces of information from the environment in order to assist its action plan. Affordance rich and affordance non-rich makes big difference in the base of interaction. Many pieces of affordance are available in a familiar environment but very limited in an unfamiliar environment.

Dynamic Touch

An action that the subject actively applies to the objects in the environment in order to obtain an invariant. For eg: Touching, Shaking, Hitting, etc. Humans go from gentle to rough naturally. That is, from micro-activity to action. An accumulation of micro-activities leads to finding the invariant.

Social Interaction Model

Human society consists of many small and large groups of people. Such groups try to achieve common goals, solve common problems, and assure trust. These activities are carried out through various interactions within the group.

A group, is when the members have apparent roles to achieve a common goal. A community is a gathering of participants who joined based on their own will. For eg: A circle, alumni, volunteer group, etc. Traditionally it indicates a geographical locality or association. Advances in the information networks led to the creation of Network community.

There is a clear boundary to differentiate between members and non-members. Through background knowledge and interests, the participants build strong associations. There are formal and tacit rules and regulations.

A common factor in both, group and community is that both are usually formed based on different objectives but both entities involve strong awareness of other members through background/tacit knowledge.

Awareness is closely related to tacit knowledge along with various pieces of background information in the environment. Appropriate awareness will add positive information towards successful completion of tasks and an awareness-rich community is more likely to succeed.

WEEK 5

Operating on Visual Representations

Visualization of Visual Analytic should facilitate the manipulation of visually represented data. A few examples of operations would be: zooming in/out, selecting, filtering, highlighting, etc. There are different types of interactions in visualization of data, a few being:

- Selection and Manipulation
    - Directly working on the visual representation of the data.
    - Path Tracing
        - Interaction with the visualization may require continuous tracing/steering of a visual cue.
        - The performance of the tracing act is influenced by the following:
            - Width of the path
            - Difficulty in the type of motor control you need to carry out
            - The Velocity
    - Control Compatibility
        - If the control required for the interaction is not compatible with what you already know then it might take time to learn the interaction.
        - We need to make sure that the interaction is very intuitive and familiar to the wider audience.
        - To do this, we need to use good metaphors.
- Exploration and Navigation
    - Understanding and walking through a visually represented space.
    - Visually sift through massive and complex data to reach something through changing visual appearances via interactions.

- Moving around in the visually represented data space means to change the viewing location, orientation, direction in, or on the visual landscape through interactions.
- Changing visual appearance by changing the focus of visual attention by:
  - Spatial Scaling
    - Changing the level of details within the visually represented space.
  - Structural Scaling
    - Switch between different levels of hierarchy writing complex data.
  - Temporal Scaling
    - Different time scale.
- Typically uses "Spatial navigation metaphors" to move around inside the visualized space.
  - Landscape: visually represent data points on 2D or 3D space like a map.
  - Reference point provisions.
- Dimensional mismatch is when there are different dimensions between visualization space and interaction space or has a different physical visualization space compared to the interaction space.

Rules of User Interfaces

- Hick-Hymen Law
  - It states that the time taken to reach the decision is proportional to the number of available choices.
  - Its equation is given as:

$$RT = c + k * \log_2 b$$

  - Where RT is the Reaction Time, c & k are constants and b is the number of choices.
  - Time taken formula is given by:

$$TST = (bt + k + c) * \frac{\log_2 N}{\log_2 b}$$

  - Where TST is the Total Search Time, k & c are constants, t is the time taken for each selection, b is the number of selections, d is the depth of the nest and N is the total number of selections.
- Fitt's Law
  - It states that the time taken to reach the target using a pointing device depends on the size and distance to the target.
  - The formula for it is given by:

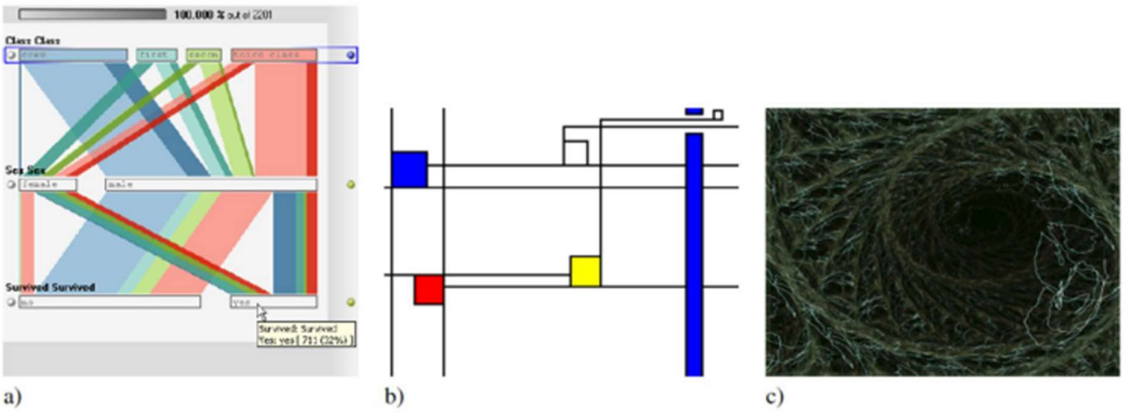
$$ID = \log_2 \frac{2A}{W}$$

  - Where ID represents the index representing the difficulty of pointing, A is the distance to the object and W is the width of the object.
- Updated Fitt's Law
  - It states that the ID is proportional to the sum of the time taken to recognize the pointer, distance between the target and pointer and for action.

Types of Visualizations

There are many types of visualization, a few of them given below:

- Readable/Recognizable
- Not Readable but Recognizable
- Neither Readable nor Recognizable



Benefits of Visualization

- Increased resources
  - Offload work to the perceptual system
  - Use it as an External memory
- Using simple visualization reduce search space
  - Grouping information
  - High data intensity
- Enhance recognition
  - Recognition instead of recall
  - Abstraction and aggregation
- It is a Manipulable medium

To make a better visualization we need to make sure the benefits of the visualization is observed. It needs to be able to make the people know how to sense data and evaluate the visualizations.

Evaluation

A visually inspectable representation of non-visual data. It is also defined as the minimal set of requirements for any visualization as based on data, produces an image and is readable and recognizable. There are many benefits of visualization:

- Increased resources
- Reduced search
- Enhanced recognition
- Manipulable medium

Evaluation is an assessment of visualizations to see if benefits have been provided. A good visualization is not pretty pictures. To make a good visualization we need to know how people make sense of data, and how to evaluate visualizations.

Gestalt principles of similarity, proximity, common region, closure, continuity and connection can be used for evaluating the visualizations. There are multiple ways of evaluating and a few are:

- Interview
    - It is a qualitative technique of gathering data about users by directly talking to them. It is usually done one to one and has either a structured format or an open ended format. The problem for this type of evaluation is the unstructured nature of the data can be misinterpreted.
- Questionnaire
    - It is a qualitative technique but where results can be quantified. In this type of evaluation, the questions asked are simple, clear and concise. The questions can again be either general or open-ended or closed or scale based.
- Analytical Inspection
    - The benefits of this evaluation is that it generates results quickly with low cost and can be used early in the design phase. It is typically based on some expert's ideas. For testing, we usually use the cognitive walkthrough wherein one starts with a task analysis hat specifies the sequence of steps or actions required by a user to accomplish a task and then work through the steps.
- Empirical Evaluation
    - It involves the usability tests which uses thinking aloud or eye tracking. This gives a lot of information and hence helps us design the visualization according to their thinking process.
    - Another way is through a controlled experiment where a comparison between multiple visualizations takes place

Aspects of a High Quality Graph

- Efficiency – The running time of algorithms should be reasonably fast.
- Elegance – Algorithms should be easy to understand and easy to code resulting in the final drawing to be beautiful.
- Effectiveness – Graph viewers should understand the underlying data quickly and correctly. The effectiveness is portrayed by maximised symmetry, minimised edge crosses and maximised angular resolution.

The graph must be easy to understand for the real users. It is assumed that graphs are more effective when drawn conforming to some predefined criteria such as:

- Maximised symmetry
- Minimized edge crosses
- Maximises angular resolution

WEEK 6

How to construct a compelling story?

1. What to visualize?
    - It is more important than how to visualize sometimes.
    - In statistical analysis we get to know about garbage in garbage out. Meaning that if in your statistical analysis you input garbage values, you will get garbage outputs for your results.
2. Why do you want to visualize?
    - To understand or explain data
        i. Organise or collate data and explain the trend/features of the data.
        ii. Not to instigate the users towards a particular opinion but make the data speak about itself.
    - To obtain a new hypothesis or insight
        i. Using interactive visualization make the user come up with a new insight regarding the data.
        ii. Allows the user to see the data from different perspective.
3. Set KGI and KPIs
    - Decide the key goal indicator (KGI)
        i. It will usually quantitatively evaluate the object or concept to be visualized.
    - Decide the key performance indicator (KPI)
        i. It will usually be derived from KGI to explain what needs to be done to achieve the KGI.
        ii. Example: KGI – Increase in Sales, KPI – Number of Customers, etc.

Modelling

There are different types of variables in modelling. These are:

- Objective variables
    - A variable to be explained using a model.
    - Example: Sales of bottles of water.
- Explanatory variables
    - These are variables to explain the Objective variables.
    - Example: Temperature of the day.
- Instrumental variables
    - These are explanatory variables but those which can be manipulated.
    - Example: The price of the bottle of water.

Selecting variables is an important step and the following needs to be kept in mind:

- It is important to differentiate between Instrumental and non-instrumental variables.
- Once we identify the KGI, one can translate that into Objective variables and KPI as Explanatory variables.
- Among these explanatory variables one needs to identify which one is instrumental and which one not.
- To introduce interaction in the visualization, one needs to focus on allowing the user to manipulate the instrumental variables.

- The KPI must be directly related to the KGI. One needs to start with minimum necessary KPIs and one should not increase the number of KPIs. The KPI should be clearly defined. They also need to be categorised based on data types such as quantitative, qualitative, etc.

If you do not find a KPI, you build them. One can do this by any one the following methods:

- Combining data
  - Combine multiple data to create a new KPI.
  - For example: BMI from Weight/Height^2
- Converting data
  - Metric conversion (changing the number to strings or something else, etc.)
  - To dimensionless data (ratio or something)
  - Normalisation of data (fit the distribution)
- Using comparative metrics
  - Choose which data is appropriate by checking with the question and understanding which one of the data is better for answering the KPI.
  - Compare against different targets.
  - Check for temporal (time) comparison for the same data type.
- Using summary statistics
  - Using mean (average), median or mode.

WEEK 7

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is driven by statistical features of datasets. EDA has a number of different tasks or processes. A few of them are:

- Testing assumptions
- Selecting models
- Validating regression models
- Selecting estimator
- Identifying relationships
- Detecting outliers

When one focuses on EDA, one puts less focus on visual or artistic design and puts more focus on statistic part of the visualization to assist the viewers to pay more attention to the data. EDA usually has a main message to convey to the audience. This is typically used in presentations, papers, hackathons, etc.

Types of Exploratory Data Analysis

- Temporal Comparison
  - This is where one displays how indices change over a period of time.
  - In this we visualize data, show the trend and expect to find some kind of pattern through it.
  - A few common visualizations which can be used for Temporal Comparisons are:

Basic		Line graph
Emphasize the amount		Area graph
Change of composition		Stacked Area graph
Change of % composition		% Stacked Area graph
A pair comparison		Slope chart
Compare multiple trends/changes		Spark line

- Attributes Comparison
    - This is where we compare indices for each attribute.
    - In this we visualize or comparing the size (large or small) or a visual aspect (good or bad).
    - A few common visualizations which can be used for the Attributes comparisons are:

Basic		Bar graph
Show multiple indices		Colored Bar graph
Combine multiple indices		Bullet chart
Overview of multiple indices		Tree map
Show actual values of indices		Score card

- Ranking Analysis
    - This is where we rank the attributes based on index.
    - In this we visualize and compare the ranking and take care of the order of the ranking.
    - A few common visualizations which can be used for the Ranking analysis are:

Basic		Bar graph
Temporal change		Bump chart

- Composition Analysis
    - This is where we compare the composition of the attributes. As in in which attribute can we find more elements in the dataset.
    - In this we visualize it and compare usually through ordering it.
    - A few common visualizations which can be used for the Composition Analysis are:

Small number of attributes		Pi chart
Many attributes		Tree map
Compare composition + amount		Stacked bar chart
Compare composition		% Stacked bar chart

- Distribution Analysis
  - This is where we compare the distribution of the attributes.
  - In this we visualize and compare the dataset and view the clusters, the spread/distribution of the data points.
  - A few common visualizations which can be used for the Distribution Analysis are:

Basic		Scatter plot
Overall Distribution		Histogram
Distribution of each attributes		Box plot

- Variance Analysis
  - This is where we compare the distribution as compared to the average.
  - In this we visualize and compare the datasets and view the clusters the spread/distribution of the data points.
- Correlation Analysis
  - This is where we compare the distribution of attributes with respect to the average.
  - In this we visualize and compare the dataset with positive or negative correlation.
  - A few common visualizations which can be used for the Correlation Analysis are:

Two attributes		Scatter plot
Three attributes		Bubble chart

- Geographic Analysis
  - This is where we analyse the location or arrangement of attributes based on geo-referenced information.
  - A few common visualizations which can be used for the Geographic Analysis are:

Compare an index among regions		Thematic map
Compare multiple indices among regions		Symbol map

Statistics in Data

We can get multiple types of statistical features from a dataset. These statistical features include but are not limited to:

- Average
- Range (minimum and maximum)
- Median
- Variance
- Standard Deviation
- Quartiles
- Skewness
- Kurtosis

Usually for these types of statistical features we use the following visualizations:

- Scatter plots
- Histograms
- Probability plots
    - If we have 2 attributes and we calculate the probability for each domain, we can plot the probability vs probability.
- Spaghetti plots
    - It is usually drawn to visualize possible flows through the system.
- Residual plots
    - It is plotted by calculating the residual of data points from the average.
- Box plots
- Block plots
- Bi-plots

Multidimensional Statistical features

Two approaches to handle this types of statistical feature are by:

- Having a coordinated multi-view visualization.
- Spatialization of the feature.

WEEK 8 → No Class (Cancelled due to public holiday)

WEEK 9

Design and Representation

- Selecting appropriate colours
    - Use colour where you wish to highlight.
    - Limit the number of colours on screen.
    - Prevent exaggeration with a low chroma (intensity of the colour).
    - Pick the right colour combination or just one colour so you do not lose focus with 2 separate hues.
    - Use of colour can give different attention.
    - Utilize perception associated with the colour.
    - Do not assign different meaning to the same colour.
    - Consider the visually impaired.

    - Put meaning background of the visualizations.
- Selecting appropriate decorative visual features
    - Remove unnecessary attention que.
    - Remove unnecessary border lines.
    - Avoid unnecessary visual variables (decoration).
    - Avoid unnecessary decoration.
    - Consider the appropriate use of thickness compared to the white spaces in the visualization.
    - Do not take away the attention from the graph.
    - Do not abuse labelling.
    - Avoid 3D modelling
- Bar Chart
    - It is an Explicit reference point
    - Explicit reference point 0.
    - Reduce the amount of comparison.
    - Use a line graph for visualizing trend
    - For visualizing a trend, you may omit the reference point.
    - Angles texts are harder to eat, so did u eat it?
    - Detailed comparison is a bit difficult with a pie chart.
    - Comparing the size or are of the size/area of glyph is difficult.
    - Trend of composition also appreciates the use offline graph.
- Line Chart
    - Use a Line graph for visualizing a trend.
    - Can omit the reference line if the goal is to just visualize the trend.
    - Do not put angled text, they are harder to read.
    - If there is no continuity, do not use line graph.
    - Trend or composition also appreciates the use of line graph.
- Pie Chart
    - It is difficult to show detailed comparison with Pi chart.
- Glyphs
    - Comparing value with the size/area of Glyph is difficult.

WEEK 10

Relationships are often described in the form of a graph. Graphs are often visualized to show hierarchical structure or many relationships. There are several methods to show hierarchy in the form of grouping such as tree map, which is ideal to describe tree structures, Partition diagrams which is used to describe data with size or nested relationships, Sankey chart to describe nested information wherein we show how data/information flows through the hierarchy. Graphs are mathematical and an abstract concept for which flowchart, organisational chart, UML are examples for.

Relationships always involve entities, and it is their relationship which we try to visualize.

Graphs

The components of the graph are vertices and edges. The vertices are also called as nodes and the edges and the binary relationship between nodes. The steps to construct a graph is given below:

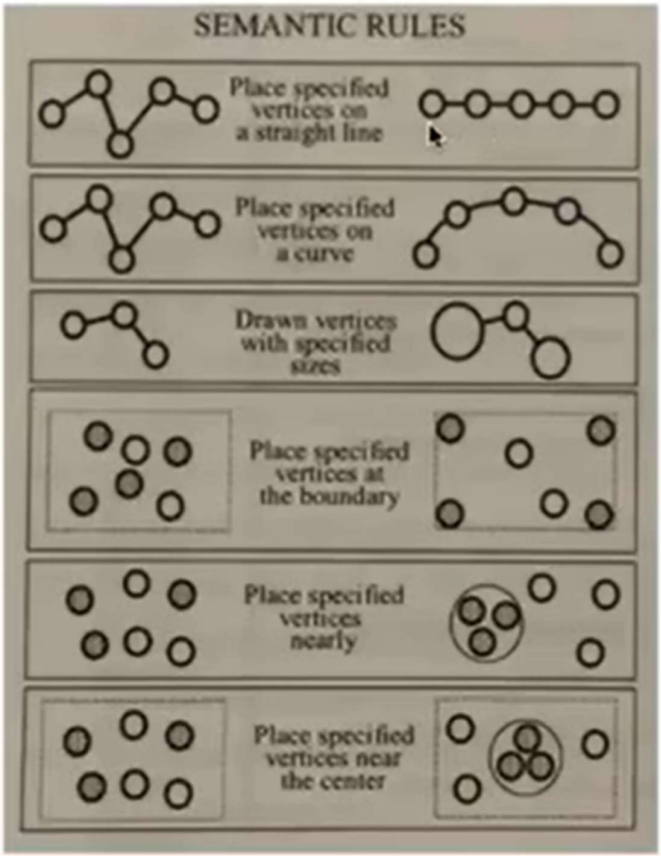
- Get the data.
- Represent the data in the form of nodes and edges.

A graph drawing is a straight-line drawing if every edge is a straight-line segment. A graph can have multiple appearances such as:

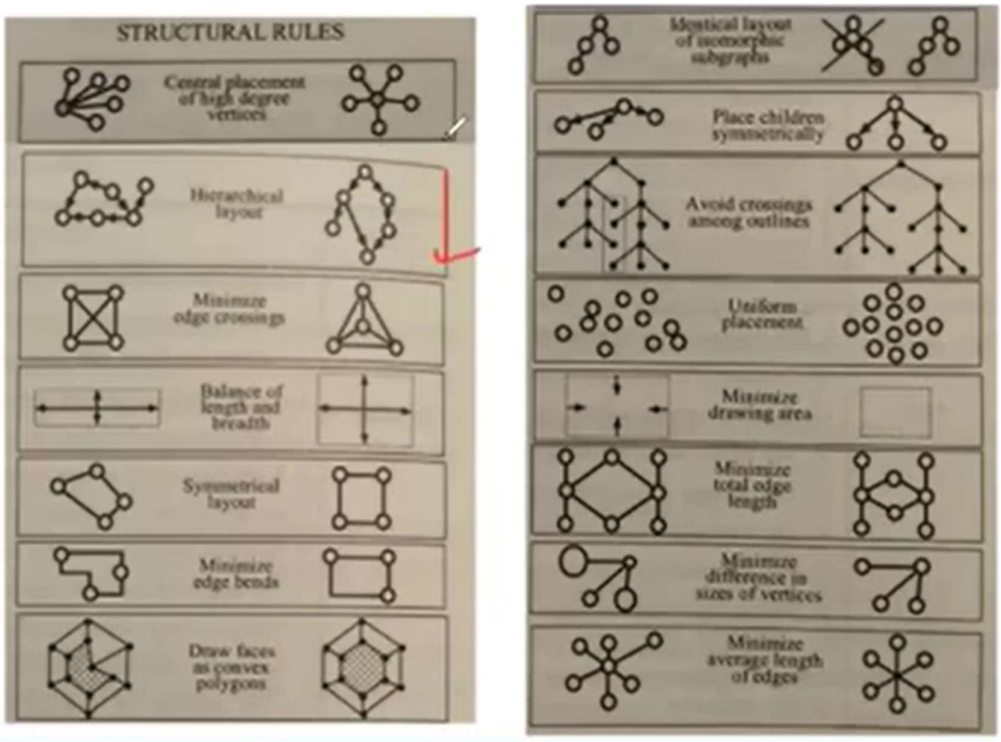
- Planar Graphs
    - A graph is planar if it can be drawn without its edges intersecting.
- Non-Planar Graphs
    - A graph is non-planar if it has intersecting edges.

A good graph is one which is easy to understand and is easy to remember. There are certain rules which needs to be followed to form a good graph. These rules are of 2 types:

- Semantic rules
    - These are driven from the nature or attribute of edges and vertices.



- Structural rules
    - These are derived directly from the graph theoretical features.



The following are the classification criteria of rules:

- Whether the solution to the rule can be obtained uniquely or not.
- Whether the rule is topological (specifying only the placement relationship between elements).
- Whether the rule is shape oriented (specifying the placement relationship between elements and the direction).
- Whether the rule is metric (specifying the placement relationship, direction and distances).
- Whether the rule applies globally to the whole drawing or locally only to the part of the diagram.
- Whether the rule is hierarchical, flat or both.

Graph layout algorithms

- Tutte's barycentre algorithm
  - Input → graph G = (V, E)
  - Output →A straight line drawing P
  - Steps for the algorithm:
    - Choose a subset A of V
    - Choose a location P(a) for each vertex such that a belongs to A.
    - For all u belonging to V-A, place u at the barycentre of its graph theoretic neighbours.
- Force directed layout
  - Mathematical curiosity turned towards visual data mining.
  - It is based on VLSI placement/routine method.
  - Peter Eades' method is one of the most used algorithms.
    - Input → Undirected graph G, the number of iterations is M

        - Output → Drawing of the undirected graph.
        - Steps for the algorithm:
            - The vertices of G are placed in random positions.
            - The following is repeated M times:
                - The forces working at each vertex is computed.
                - Each vertex is moved by c4 * the force working at that vertex.
            - G is drawn.
        - Use springs to represent the edge.
        - The spring has its own length.
        - Use an inverse square law repulsive force between non-adjacent vertices.
    - Some of its features are:
        - Adjacency vertices are placed near each other. (closeness)
        - Vertices are placed to not be too close to each other (smallest separation)
        - Edge lengths are fixed. (Fixed edge lengths)
        - Symmetry is clearly shown. (symmetry)
        - Within the given drawing frame, vertices are uniformly distributed. (Uniform distribution)
        - The shape of the drawing is adapted to the drawing frame. (Adaptation to the frame)
        - Edge crossings are minimised. (Edge crossing minimisation)

WEEK 11

Spatialisation

Spatialisation is the process of finding space or the position of the data points. It is typically used to visualise the similarities of data entries. An example of this is clustering. In multidimensional scaling as spatialization, the main goals are:

- Observed target objects.
- Observed data could be:
    - Categorical
    - Ordinal
    - Numerical
- Forms a vector in (n-dimension)
- Use MSD to find 2D-3D location.

Measures of observation

- Classificatory concept
    - Just to distinguish one from others.
- Comparative concept
    - Provides logical/ordered structure of relationships.
    - Measurement is done by evaluation of the value.
    - Empirical relations are defined as equivalence and magnitude relationships.
    - Equivalence has 2 types:
        - Symmetric
        - Transitive
    - Magnitude has 2 types:
        - Asymmetric

        - Transitive
- Quantitative concept
    - Subjects will have numerical attributes.

Multidimensional Scaling

High dimensional data being scaled down to low dimensional data so that the data can be used is called multidimensional data. It should have given some similarity measures. It describes an object as a point in the multidimensional space and used to calculate the placement of data points so that the distance between the points corresponds to the observed similarity.