DATA VISUALIZATION

UNIT-1

FOUNDATIONS FOR AN APPLIED SCIENCE OF DATA VISUALIZATION

- Contents

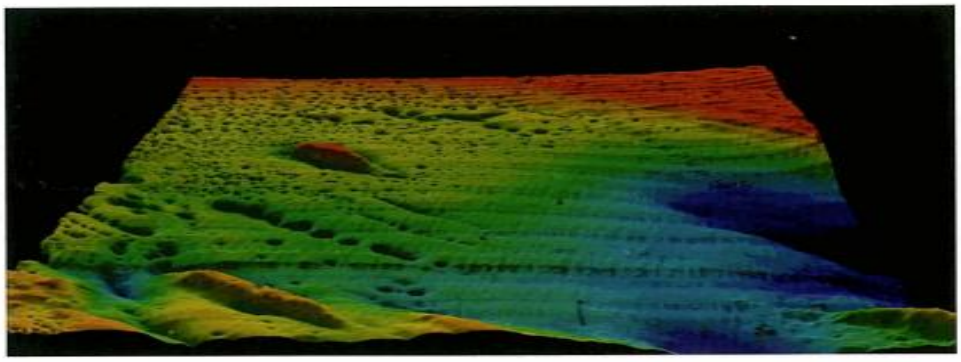
Foundations for an Applied Science of data visualization:

- Visualization Stages
- The Study of Arbitrary Conventional Symbols:
    - Gibson's Affordance theory,
- A Model of Perceptual Processing:
    - Stage 1: Parallel Processing to Extract Low-Level Properties of the Visual Scene, Stage 2: Pattern Perception, Stage 3: Sequential Goal-Directed Processing
- Types of Data:
    - Entities, Relationships, Attributes of entities or Relationships, data dimensions, Operations considered as data.

- Visualization---Why?
  - Acquire more information through vision
  - Provide high bandwidth channel from the computer and human
  - Constructing a visual image in the mind
  - An external artifact supporting decision making.

Major Advantage:

Sheer quantity of information that can be rapidly interpreted if it is presented



.1 Passamoquoddy Bay visualization. *Data courtesy of the Canadian Hydrographic Service.*

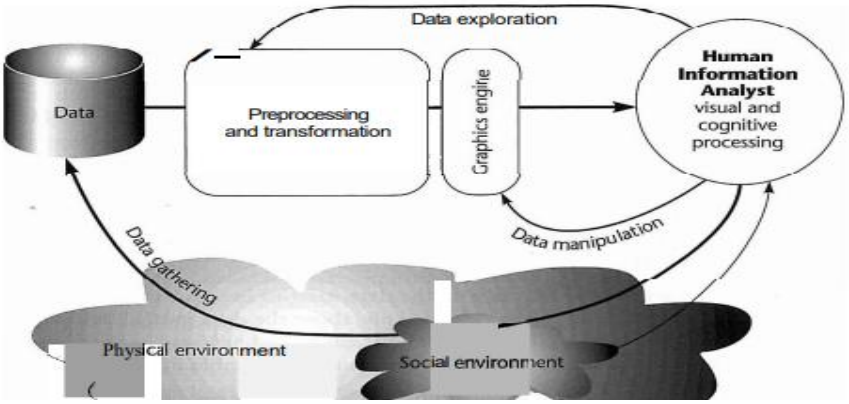
- A pattern of features called pockmarks can immediately be seen, and it is easy to see how they form lines.

The Passamoquoddy Bay image highlights a number of the advantages of visualization:

- Visualization provides an ability to comprehend huge amounts of data
- Visualization allows the perception of emergent properties that were not anticipated
- Visualization often enables problems with the data itself to become immediately apparent.
- Visualization facilitates understanding of both large-scale and small-scale features of the data.
- It is valuable in allowing the perception of patterns linking local features.
- Visualization facilitates hypothesis formation.

Visualization Stages

- The process of data visualization includes four basic stages, combined in a number of feedback loops.



- Visualization stages:-The four stages consist of:
- The collection and storage of data itself
- The pre-processing designed to transform the data into something we can understand
- The display hardware and the graphics algorithms that produce an image on the screen
- The human perceptual and cognitive system (the perceiver)

Visualization stages- **Gathering data**

- The longest feedback loop involves gathering data.
- A data seeker, such as a scientist or a stock-market analyst, may choose to gather more data to follow up on an interesting lead.
- The physical environment -- source of data
- The social environment determines in subtle and complex ways what is collected and how it is interpreted
- Visualization stages- **preprocessing**
- Another loop controls the computational preprocessing that takes place prior to visualization.
- The analyst may feel that if the data is subjected to a certain transformation prior to visualization, it can be persuaded to give up its meaning.
- Visualization stages- **Visualization**
- The visualization process itself may be highly interactive
- 3D data visualization, the scientist may fly to a different vantage point to better understand the emerging structures
- Visualization stages- **Perception**
- The emphasis is on data, perception, and the various tasks to which visualization may be applied.
- The computer is treated as universal too, for producing interactive graphics.
- How best to transform the data that people can understand for optimal decision making

The Study of Arbitrary Conventional Symbols

- Gibson's Affordance theory

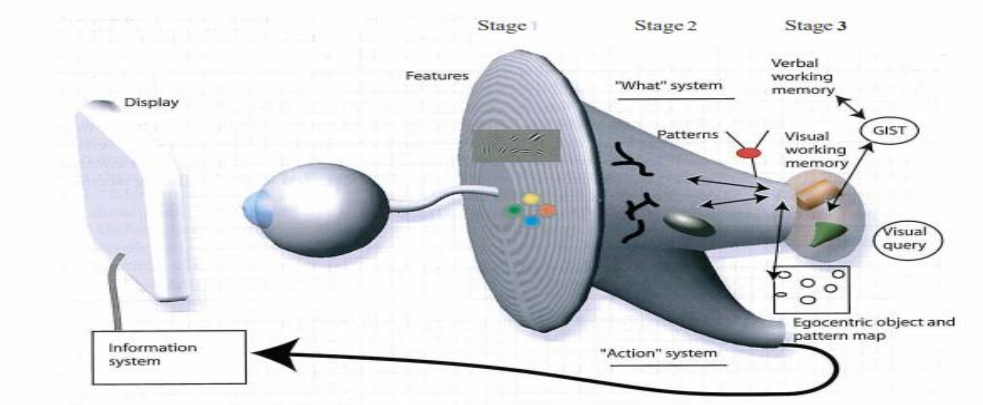
According to Gibson, affordances are physical properties of the environment that we directly perceive.
Many theorists, unlike Gibson, think of perception as a very active process:
The brain deduces certain things about the environment based on the available sensory evidence.

Limitations
There are three problems with Gibson's direct perception in developing a theory of visualization.
The first problem is that even if perception of the environment is direct, it is clear that visualization of data through computer graphics is very indirect. Many layers of processing between the data and its representation which are invisible

- Second, there are no clear physical affordances in any graphical user interface.
  - To say that a screen button "affords" pressing in the same way as a flat surface affords walking
  - the use of buttons is arbitrary-when pressed do interesting things which are indirect
- Third, Gibson's rejection of visual mechanisms is a problem
  - To reject the importance of understanding visual mechanisms would be to reject a tremendous proportion of vision research as irrelevant

Model of Perceptual Processing

- In Stage 1, information is processed in parallel to extract basic features of the environment.
- In Stage 2, active processes of pattern perception pull out structures and segment the visual scene into regions of different color, texture, and motion patterns.
- In Stage 3, the information is reduced to only a few objects held in visual working memory by active mechanisms of attention to form the basis of visual thinking.



Stage 1: Parallel Processing to Extract Low-Level Properties of the Visual Scene

- Visual information is first processed by large arrays of neurons in the eye and in the primary visual cortex at the back of the brain.
- Selective tuning of information for a particular kind of information such as the orientation of edges or the color of a patch of light.
- Important characteristics of Stage 1 processing include:
  - Rapid parallel processing
  - Extraction of features,
  - orientation, color, texture, and movement patterns
  - Transitory nature of information, which is briefly held in an iconic store
  - Bottom-up, data-driven model of processing

Stage 2: Pattern Perception

- At the second stage, rapid active processes divide the visual field into regions and simple patterns.
- The pattern-finding stage of visual processing is extremely flexible, influenced both by the massive amount of information available from Stage 1 parallel processing and by the top-down action of attention driven by visual queries.
- Marr (1982) called this stage of processing the 2-1/2D sketch.
- Triesman (1985) called it a feature map.
- Rensink (2002) called it a proto-object flux to emphasize its dynamic nature.
- There is increasing evidence that tasks involving eye-hand coordination and locomotion may be processed in pathways distinct from those involved in object recognition.
- This is the two-visual system hypothesis: one system for locomotion and action, called the **"action system,"** and another for symbolic object manipulation, called the **"what system."**
- Important characteristics of Stage 2 processing include:
  - Slow serial processing Involvement of both working memory and long-term memory
  - More emphasis on arbitrary aspects of symbols
  - In a state of flux, a combination of bottom-up feature processing and top-down attentional mechanisms
  - Different pathways for object recognition and visually guided motion
  -

Stage 3: Sequential Goal-Directed Processing

- At the highest level of perception are the objects held in visual working memory by the demands of active attention.
- At this level, only a few objects can be held at a time; they are constructed from the available patterns providing answers to the **visual queries**.
- The **visual object identification** process interfaces with the **verbal linguistic** subsystems of the brain so that words can be connected to images.

- The perception-for-action subsystem interfaces with the motor systems that control muscle movements.
-

Types of Data- Entities, Relationships

- Goal – transform data into perceptually visual format.
- to make generalised statements-know about the existing types of data
- Example: colors –stock market analysis, textures for geological maps
- Define broad categories of data
  - Scalar-continuous height maps
  - Vector – continuous flow
  - Category - Color coding
  - Motion coding –highlighting selected data

Types of Data- Entities, Relationships

- Bertin (1977) has suggested that there are two fundamental forms of data: **data values and data structures**.
- **Entities** are the visualized objects
- **Relations** define the structures and patterns that relate entities to one another.
- Concepts are related to relational database-ERModel

Definitions

- Entities-objects of interest
  - Example: School of fish, group of things, people, hurricanes
- Relationships- form the structures that relate entities
  Structural and Physical-house of component parts
  Conceptual - store and customers
  Casual – one event causes another
  Temporal – time interval between the events

Attributes of Entities or Relationships

- Entities and Relationships have attributes.
- Attribute it is property –cannot be thought independently
- Examples:
  - Color - Apple
  - Temperature – Water
  - Duration – Journey
  - Salary – Employee
- Attribute Quality
- The quality of data is the taxonomy of number scales defined by the statistician S.S. Stevens (1946).

According to Stevens, there are four levels of measurement: nominal, ordinal, interval, and ratio scales.

- Nominal: This is the labeling function.
  Example: Fruit can be classified into apples, oranges, bananas..
  No ordered sequence
- Ordinal: used for ordering things in a sequence
  Example: rank some group of things (films, political candidates, computers) in order of preference.
- Interval: to derive the gap between data values.
  Example: The time of departure and the time of arrival of an aircraft are defined on an interval scale.
- Ratio: The full expressive power of a real number.
  Example: statements such as "Object A is twice as large as object B."

Only three levels of measurement are used:

- Category data: Nominal class.
- Integer data: Ordinal class in that it is discrete and ordered.
- Real-number data: This combines the properties of interval and ratio
- Attribute Dimensions: 1 D, 2D, 3 D, . . . Scales:
- An attribute of an entity can have multiple dimensions.
  - single scalar quantity - weight of a person.
  - vector quantity - the direction in which that person is traveling.

Operations Considered as Data

- Mathematical operations on **numbers-multiplication, division**, and so on
- **Merging** two lists to create a longer list
- **Inverting** a value to create its opposite
- **Bringing** an entity or relationship into existence (such as the mean of a set of numbers)
- **Deleting** an entity or relationship (a marriage breaks up)
- **Transforming** an entity in some way (the chrysalis turns into a butterfly)
- **Forming** a new object out of other objects (a pie is baked from apples and pastry)
- **Splitting** a single entity into its component parts (a machine is disassembled)