Unit-5

Visualization techniques for multivariate Data

Spatial attributes are an integral part of representation, which take advantage of the power of GIS to determine locational properties of heritage data we organize the presentations based on the graphical primitive used in the rendering, namely points, lines, or regions, followed by techniques that combine two or more of these types of primitives.

Meaning of Univariate, Bivariate & Multivariate Analysis of Data :-

Univariate Analysis - In univariate analysis, one variable is analysed at a time.

Bivariate Analysis - In bivariate analysis two variables are analysed together and examined for any possible association between them.

Multivariate Analysis - In multivariate analysis, the concern is to analyze more than two variables at a time.

→ The type of statistical techniques used for analysing univariate and bivariate data depends upon the level of measurements of the questions pertaining to those variables. Further, the data analysis could be of two types, namely, descriptive and inferential

Point Based Visualization Techniques:

point plots can be defined to display individual records or summary records, and can be structured by various projection techniques.

point plots are introduced as visualizations that project records from an n-dimensional data space to an arbitary k-dimensional display space, such that data records map to k-dimensional points.

Scatterplot and Scatter matrices : (scout)

Scatterplots are one of the earliest and most widely used visualization techniques used in data analysis. Both 2D and 3D scatterplots can be found in most packages designed to support data and information analysis.

→ As the dimensionality of the data increases, the choices for visual analysis consist of

1. Dimension subsetting
2. Dimension reduction
3. Dimension embedding
4. Multiple Displays

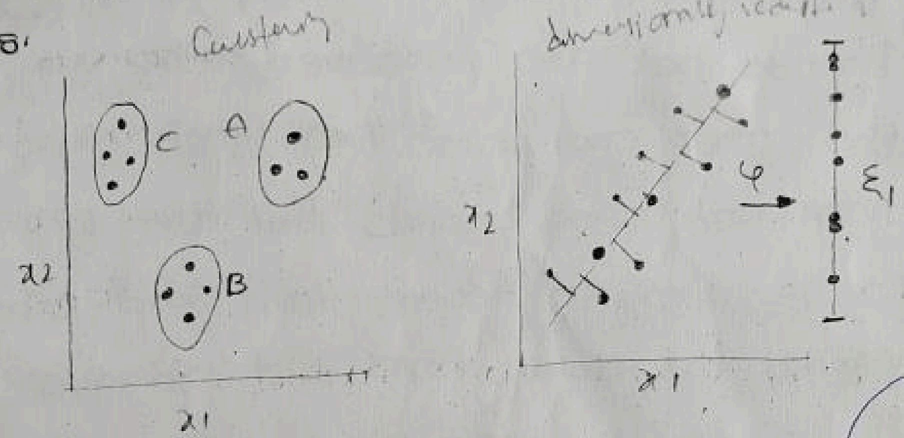
Dimension sub setting:

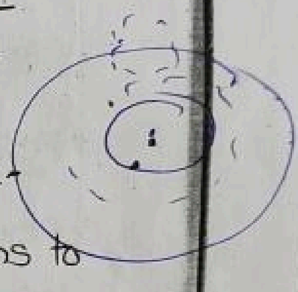
Dimension sub setting - allowing the user to select a subset of the dimensions to display, or to develop algorithms to find the dimensions containing the most useful information for the task at hand.

Dimensionality Reduction:

Dimension reduction - Using techniques such as

...nciple component analysis or multidimensional scaling to transform the high-dimensional data to data of lower dimension, while attempting to preserve as best as possible the relationships among the data points.





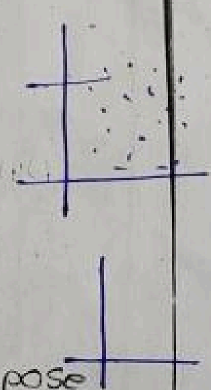
Dimension Embedding & Multiple Display:-

Dimension embedding - mapping dimensions to other graphical attributes besides position, such as colour, size, and shape (though there are limits to how many dimensions can be included this way).

Multiple displays - showing, either superimposed or juxtaposed, several plots, each of which contains some of the dimensions

↳ placing two or more things side by side — compare, contrast, or create an inter

Multi Dimensional SCALING (MDS):-



Multidimensional scaling is a means of visualizing the level of similarity of individual cases of a dataset.

From a non-technical point of view, the purpose of multidimensional scaling (MDS) is to provide a visual representation of the pattern of proximities (i.e., similarities or distances) among a set of objects... 15/6/17

For example, given a matrix of perceived similarities between various brands of air freshe[ner] MDS plots the brands on a map such that those brands that are perceived to be very similar to each other are placed near each other on the map, and those brands that are perceived to be very different from each other are placed far away from each other on the map.

FORCE BASED METHODS:

Any techniques for projection high-dimensional points into 2D or 3D display space have been developed.

→ the key goal is to attempt to maintain the N-dimensional features and characteristics of the data through the projection process, e.g., relationships that exist in the original data must also exist after projection.

→ this, however, is not always possible, especially as the dimensionality of the data increases. The projection may also unintentionally introduce artifacts that may appear in the visualization and are not present in the data

COMBINATION TECHNIQUES :-

In addition to the techniques based on points, lines, or regions, there are a number of hybrid techniques that combine features of two or more of these classes.

→ We describe two of the more popular techniques of this type : glyphs and dense pixel displays.

→ In the context of data and information visualization, a glyphs is a visual representation of a piece of data or information where a graphical entity and its attributes are controlled by one or more data attributes.

→ As an example, the width and height of a box could be controlled by a student's score on the midterm and final exam for a course, while the color could be associated with the gender of the student.

GLYPHS :-

A glyph is any marker, such as an arrow of similar marking, used to specify part of a visualization.

→ this is a representation to visualize data where the data set is presented as a collection of visual objects - these visual objects are collectively called a Glyph.

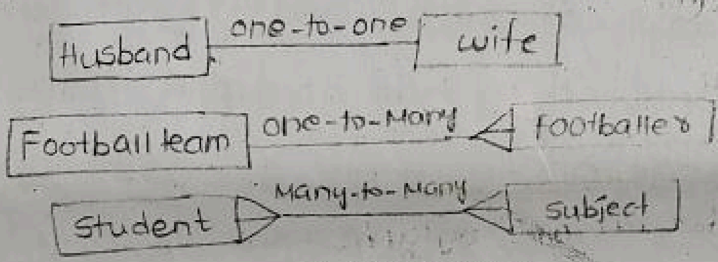
(pictures in ppt 16 pg : 3)

MAPPING :-

→ A wide range of possible mappings for data glyphs are discussed, including:

* one-to-one mappings :- where each data attribute maps to a distinct and different graphical attribute
* one-to-many mappings :- where redundant mappings are used to improve the accuracy and ease with which a user can interpret data values, and
* Many-to-one-mappings :- where several or all data attributes map to a common type of graphical attribute, separated in space, orientation, or other transformation.

glyphs



DIFFERENT TYPES OF MAPPING :-

→ one-to-one mappings are often designed to take advantage of the user's domain knowledge, using intuitive pairings of data to graphical attributes to ease the learning process. Examples include mapping color to temperature, and flow direction to line orientation.

→ Redundant mappings can be useful in situations where the number of data dimensions is low and the desire is to reduce the possibility of misinterpretation. For example, one might map population to both size and colour to ease analysis for color-impaired users, and to facilitate comparison of two populations with similar values.

any to-one mapping are best used in situations
where it is important to not only compare
values of the same dimension for separate
records, but also to compare different
dimensions for the same record.

TYPES OF GLYPHS :-

the following list contains a subset of glyphs
that have been proposed in the literature or are
in common use. some are automized to a
particular application, such as visualizing fluid
flow, while others are more general purpose

*. profiles - height and color of bars;
*. stars - length of evenly spaced rays emanating from center;
*. Anderson/metroglyphs - length of rays;
*. stick figures - length, angle, color of limbs;
*. trees - length, thickness, angles of branches; branch structure.
*. derived from analyzing relations between dimensions:
*. autoglyph - color of boxes;
*. boxes - height, width, depth of first box; height of successive
*. boxes;

Examples of multivariate glyphs :-

Faces - size and position of eyes, nose, mouth; curvature of mouth; angle of eyebrows; &

arrows - length, width, taper, and color of base and head;

polygons - conveying local deformation in a vector field via orientation and shape changes;

dashtubes - texture and opacity to convey vector field data;

ORDERING STRATEGIES :-

Once a glyph design is chosen, there are N! different dimension orderings that can be used in the mapping. Which ones are likely to reveal the most interesting features? several ordering strategies can be imagined.

→ Dimensions could be ordered according to their correlation, so that similar dimensions are mapped adjacent to each other. This can help reveal general trends, as well as expose some outliers.

→ Dimensions can be mapped in such a way as to promote symmetrically shaped glyphs, which can be easier to perceive and remember. shapes that are less symmetric than their neighbours will also stand out.

→ Dimensions can be sorted according to their values in one record. For example, if the data represents a multivariate time series, sorting the dimensions based on the first record can veying which dimensional relationships were maintained versus those that changed significantly.

→ Dimensions can be manually sorted based on the user's knowledge of the domain. Thus, semantically similar dimensions can be grouped or used for symmetric glyph features, which can simplify the interpretation.

GLYPH BASED LAYOUT STRATEGIES:-

A final important consideration in designing a glyph-based visualization is the layout of the glyphs on the screen. There are 3 general classes of layout strategies:

Uniform: glyphs are scaled and positioned with equal space between them to occupy the entire screen. This strategy eliminates overlaps, while making efficient use of the screen space. Different orderings of records can expose different data features.

Data-driven: data values are used to control the positioning of the glyphs. Two approaches are possible. In the first, two data dimensions (or three for 3D display) are chosen to set the locations.

Structure-driven: If the data has an implicit or explicit structure to it, such as cyclic or hierarchical, this can be used to control the positioning.

$\frac{1}{2}$ $\frac{2}{4}$

DENSE PIXEL DISPLAYS :- 10×10 . 18×18

Dense pixel displays (also known as pixel-oriented techniques) are a hybrid between point-based and region-based methods.

the displays make maximal use of the screen space, allowing data sets with millions of values to be shown on a single screen.

→ each data value will control the color for a single pixel; changing the color map used can potentially reveal new features of the data. Picture PPIS Pg: 14

VISUALIZATION TECHNIQUES FOR TREES, GRAPHS, AND NETWORKS

While most of the visualization techniques discussed thus far focus on the display of data values and their attributes, another important application of visualizaton is the conveying of relational information, e.g., how data items or records are related to each other. these interrelationships can take many forms:

- part / subpart, parent / child, or other hierarchical relation;
- connectedness, such as cities connected by roads or computers connected by networks;
- derived from, as in a sequence of steps or stages;
- shared classification;
- similarities in values;
- similarities in attributes (e.g., spatial, temporal)

UNIT-4

Text and Document Visualization

Levels of Text Representation:-

we define 3 levels of text representation:

1) Lexical
2) syntatic
3) semantic

Eg:- Medium blog is a 2-gram
"A medium blog post" - 4-gram
write on Medium - 3-gram

Lexical:-

→ the lexical level is concerned with transforming a string of characters into sequence of atomic entities called tokens.

a series of characters of length N

→ Tokens can include characters, character n-gram, words, word stems, lexemes, phrases, or word n-grams, all with associated attributes.

↳ it is a part of a word responsible for its lexical meaning

→ Many types of rules can be used to extract tokens, the most common of which are finite state machines defined by regular expressions.

Syntatic level:-

→ the syntactical level deals with identifying and tagging each token's function.

→ we assign various tags, such as sentence position (or) whether a word is a noun, expletive, adjective. (or) conjuction

→ Tokens can also have attributes such as whether they are singular or plural, or

their proximity to other tokens.
→ Richer tags include date, money, place, person, organization and time.
→ the process of extracting these annotations (taggings) is called name entity recognition (NER).

Semantic level:-

→ the semantic level encompasses the extraction of meaning and relationships between pieces of knowledge derived from the structures identified in the syntatical level.
→ the goal of this level to define an analytic ~~ops~~ interpretation of the full text within a specific context, or even independent of context

2) The vector space Model:-

vector space models are algebraic ... of ... that are often used to represent text as vector of identifiers.

→ Where are you from? What is your age?
Where are you going? How old are you?

Same words, Different meaning

Different words same meaning

The are numerous instances we may decide to employ a vector spaced model, for instance:

1. Information Filtering
2. Information Retrieval
3. Machine Translation
4. chatbots

In general, vector space models allow us to represent words and documents as vectors.

1) Word By Word method

2) Word By Document method

(1) Word By Word method :-

→ performing this task involves first creating a co-occurence matrix

→ we want to go from our co-occurence matrix to a vector representation

co-occurence ⟶ vector representation

1) Word By Word Method :-

this design counts the no. of times words occur within a certain distance 'k'.

k=2

eg: I like simple data

I prefer simple raw data

	simple	raw	like	I
Data	2	1	1	0

In the word by word design, the cooccurrence matrix is b/w 1 and N entries

2) Word By Doc Method :-

the no of times words from the vocabulary appear in documents that belong to certain categories

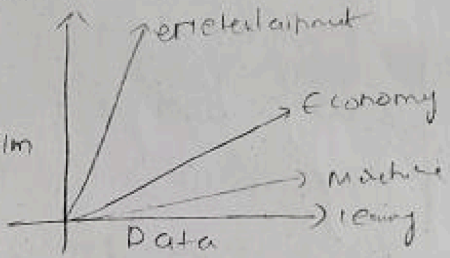
	Entertainment	economy	machine learning
Data	500	6620	9320
Film	7000	4000	1000

repres..

a) Vector representation :-

→ using these vector representations. we can now represent our text (or) documents in vector space.

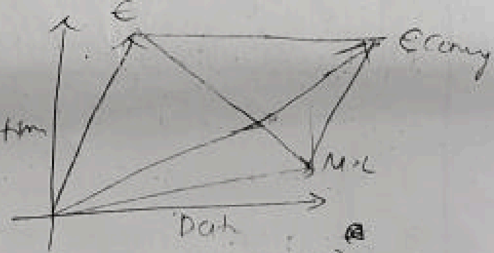
→ this is perfect because in vector space we can determine the relationships b/w types of documents, such as their similarity.



b) Euclidean distance :-

→ A similarity metric we may use to determine how far apart 2 vectors are from one another is the Euclidean distance (length of st line that connects 2 vectors.

$$Eu(A,B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2}$$



Euc(E, Ec) = 6515.75

Euc(E, ML) = 1086675

Euc(ML, EC) = 403.609 (remaining in ppt) 17

single document visualizing :-

Here we present several visualizations of a single text document

1) Tag cloud
2) word tree
3) Text arc
4) arc diagrams
5) literature fingerprinting

2) Tag cloud:- word clouds /text clouds

→ A tag cloud visualization generated by the free service tagcloud.com

→ the font size & darkness are proportional to the frequency of the word in the document

→ Tag clouds also known as text clouds or word clouds are layout of raw tokens, colored and sized by their frequency within a single document

Text clouds & their variations, such as Wordle are examples of visualizations that use only term frequency vectors and some layout algorithm to create the visualizations

3) Text arc:-

A Text arc visualization that uses the full text of Alice in wonderland.

→ words that occur evenly throughout the document are positioned in the centre of the display.

→ while words that appear only in specific sections are located closer to the circumference.

Word Tree:- The word Tree visualization is a visual representation of both term frequency, as well as their context. size is used to represent the term or phrase frequency. The root of the tree is a user-specified word or phrase of interest, & the branches represent the various contexts in which the word or phrase is

ARC diagrams:-

→ Arc diagrams are a visualization focused on displaying repetition in text or any sequence

→ Repeated subsequences are identified and connected by semicircular arcs.

→ the thickness of the arcs represents the length of the subsequence and height of the arcs represents the ~~diagra~~ distane b/w the subsequences

literature fingerprinting:-

→ It is a method of visualizing features used to characterize text.

→ Instead of calculating just one feature value or vector for the whole text, we calculate a sequence of feature values per text and present them to the user as a characteristic fingerprint of the document.

→ This allows the user to "look inside the document and analyze the development of the values across the text.

4) Document collect visualizations:-

→ In most cases of document collection visualizations, the goal is to place similar documents close to each other and dissimilar ones far apart

→ we compute similarity b/w pairs of documents and determine a layout

We present several document collection visualizations, such as self organizing maps, cluster maps and themescapes.

→ The common approaches are graph spring layouts multidimensional scaling, clustering

i) self organization maps:-

A self-organizing map is a unsupervised learning algorithm using a collection of typically 20 nodes, where documents will be located.

→ Each node has an associated vector of the same dimensionality as the input vectors used to train the map.

ii) themescapes:-

Themescapes are summaries of corpora using abstract 3D landscapes in which height and colour are used to represent density of similar document. Eg: Pg 18

iii) Document cards:-

Document cards are a compact visualization that represents the document's key semantics as a mixture of images and import key. terms, similar to cards in a top trumps game.

→ The key terms are extracted using an advanced text mining approach based on an automatic extraction of document structure

→ the images and their captions are extracted using a graphical heuristic, and the captions are used for a semi semantic image weighting.

Interaction concepts:- (see in ppt)

Interaction operators

1) selection operator → user controls for identifying object such as highlighting, deleting & modify

2) filtering → user controls of reducing the size of the data being mapped

3) Reconfiguring " → user controls for changing the way data is mapped to graphical entities & attributes

4) Encoding " → user controls for changing the G.E, such as point size (or) line color

5) connection " → user controls for linking different view.

6) abstraction " → user controls for ~~linking different~~ modifying the level of detail

7) Navigation " → user controls for altering the position of the camera and for scaling the view

1) Navigation operators:-

→ Navigation is used to search for a subset of data to be viewed, the orientation of this view and the level of detail (LOD)

→ Navigation operators can work in absolute or relative coordinates within their particular space.

→ the user can control the step size b/w views, with the trade-off being smoothness versus the number of projections that need to be inspect.

...tion operators:-

...n selection, the user isolates a subset of the display components, which will then be subjected to some other applications such as highlighting, deleting, masking or moving to the centre of focus.

→ selection can be articulated in many different ways. the user may click on entities, paint over a selection of entities

filtering operators:-

→ filtering as the name implies, reduces the volume of data to be visualized by setting constraints specifying the data to be preserved or removed.

→ the distinction b/w filtering & selection followed by deletion or masking is a subtle, but important point.

→ filtering in general, is most often done in an indirect manner, eg: the filter specification is not performed on the data visualization itself but via a seperate interface.

→ selection is most often done in a direct manner, by indicating objects on the visualization.

Reconfiguring operators:-

→ Reconfiguring the data within a particular visualization can often be used to expose features or cope with complexity or scale.

→ By reorganizing the data, say by filetering some dimensions and reordering those that remain, different ways "views are provided to the user.

Encoding operators:-

→An given data set can be used to generate countless different visualizations.

→ recoding can provide the user library of possible different types of visualizations.

→features of the data that are difficult or impossible to see with one such mapping might become quite appearent in another.

→other forms of encoding operators include those that modify the color map used, the size of graphical entities and their shape.

Connection operators:-

→ a frequent use for selection operations is to link the selected data in one view to the corresponding data in other views.

→ while other forms of connection between sub windows of an application exist, such as when opening a new data file, linked selection is probably the most common form modern visualization tools.

Abstraction/Elaboration operators :-

-> one of the most popular techniques of this type is using distortion operators.

-> while some researchers classify distortion as a visualization technique, it is actually a transformation that can be applied to any type of visualization.

UNIT-5

1) Modern integrated visualization systems:-

Tableau is the commercial software package initially developed by "pat Hanrahan" and his students at stanford.

→ It is designed with more modern interactions to aid in data analysis.

→ Tableau allows users to import data from a wide variety formats.

→ It provides standard visualizations, such as graphs, scatterplots, barcharts and piecharts, as well as maps

→ Tableau recognises data types such as geography. and automatically handles the first steps in the generation of visualizations.

→ It also supports the export of annotated presentation visualization in PdF format and provides interactive dashboards that automatically update from data sources, as well as web publishing tools.

advantages	Disadvantages
→ high performance	→ poor versioning
→ mobile-friendly	→ NO automatic refreshing of reports
→ extensive customer resources	→ Need manual efforts
→ Excellect mobile support	→ Not a comprehensive solution

Data characteristics:-

The characteristics of the data being visualized can have profound influence on the effectiveness of the visualization technique and must be considered in the evaluation process

Type:- floating point, numbers, strings - All combinations has to be tested.

size:- Tests should cover normal sizes & extreme sizes as well

Dimensionality:- It is generally useful to test a visualization with all possible subsets of dimensions

structure:- Tables, grids, Hierarchy - all have to tested

Range:- Testing should involve exercising the entire range of possible values, including the extremes of the range.

Goal of Data visualization.

→ A successful visualization is one that efficiently and accurately conveys the desired information to the targeted audience.

→ purpose of visualization (exploration, confirmation, presentation).

→ A visualization may be ineffective for a number of reasons.

→ It might be too confusing or complex to be interpreted by the intended audience.

→ some of the data may have been distorted, occluded or lost during the mapping process.

→ other signs of deficient visualizations are the lack of support for view modification or color map control.

→ Even aesthetics can influence the success of a visualization;

→ A visualy unappealing presentation can affect an audience's willingness to look at the images.

6) Explain in steps in data visualizaton in terms of intuitiveness, views, colors, labels.
3

~~Intuitive map~~ steps in designing data visualization:-

creating a visualization involves deciding how to map the data fields to graphical attributes.

→ selecting and implementing methods for modifying views, and choosing how much data to visualize.

→ Additional information regarding the data being shown (e.g., labels) and the mapping (e.g., a color key) are also essential to facilitate interpretation, and must be integrated into the visualization.

...uitive mappings from data to visualizations :-

→ To create the most effective visualization for a particular application, it is critical to consider the semantics of the data and the context of the typical user.

→ By selecting data-to-graphics mappings that cater to the user's domain-specific model.

→ the interpretation of the resulting image will be greatly facilitated.

→ Mapping spatial data attributes, such as longitude and latitude, to screen position is perhaps the most common and intuitive mapping found in visualizations.

selecting and modifying views:-

→ the key to developing an effective visualization is to be able to anticipate the types of views and view modifications that will be one of most use to the typical user and then provide intuitive controls for setting and customizing the views.

→ scrolling and zooming operations are needed if the entire data set cannot be presented at the resolution desired by the user

→ colormap control is almost always desirable, minimally supporting a set of different pallettes.

→ Mapping control allows users to switch b/w ... visualizing the same data.

Keys, labels and legends:- intuitive

→ A common problem with many visualizations is insufficient information is provided to the user to allow unambigous and accurate interpretation.

→ If symbols are being used, a key must be provided either along the border of the display (or) within a separate widget.

→ The use of grid and tick marks can be both a boon and a curse to the visualization.

Using color with care:- intuitive aesthetic

→ Color can add significant visual appeal to a visualization, But can also significantly decrease the effectiveness of the communication proces.

→ One of the most frequently misused parameters in visualization design is that of color.

→ Selecting a wrong color map can lead to ineffective or misleading visualizations.

→ Also, since color perception is context-dependent.

The importance of Aesthetics:-

→ Once we have ensured that our designed visualization conveys the desired information to the user, the final step is to assess the aesthetics (form) of the results.

→ There are many guidelines for attractive visualization design that can be drawn from the art and graphic design communities. These include focus, Balance, simplicity.

problems in Designing effective visualizations?

Misleading visualizations:-

i) Data scrubbing:
Raw data can often be very rough in form, and the temptation when creating a visualization is to remove some of the roughness.
→ outliers removal is a common tatic in this situation.

ii) unbalanced scaling:-
scaling is a powerful tool in visualization,
→ since careful selection of scale factors can reveal patterns and structures not visible in unscaled views.

iii) Range distortion:-
→ viewers often have an expectation about the ranges for a particular data dimension.
→ By setting this range to be significantly different from this expectation, the user may be deceived into misinterpretation.

iv) Abusing Dimensionality:-
mapping a scalar value to a graphical attribute such as volume can dramatically increase the likelihood of erroneous interpretation.

Visual non-sense- comparing apple & oranges:-
Visualizations are designed to convey information, and it is important that the information be meaningful.

→ visualizations are often created by combining
sets from different sources.

→ However, it is easy to combine unrelated components
into a single visualization and identify what seems
to be structure.

→ for eg: plotting stock market values against
occurences of sunspots.

Raw versus derived data :-

→ In some visualizations, it is common practice
to throw out all of the raw data and only show
the smooth approximation derived from that data.

→ This forces the viewer to trust that the
approximation is an accurate portrayal of the data.

→ It is best to show both the raw data and the
fitted model first, and to allow one or the
other to be deemphasized or filtered out on
demand.

Absolute versuse relative judgment :-

→ Bounding boxes, grids and ticks marks are
all execellent tools for converting an absolute
judgment task to one that depends more on
relative judgement.

→ By comparing the length or position of a
graphical entity against a quantified structure,
users can more rapidly determine the
approximate value relative to the known levels.

→ Humans have a fairly limited ability to make
absolute judgments of visual stimuli

Comparing and evaluating visualization Techniques:-

User Tasks:-

To perform a valid assessment of a particular visualization technique, or to compare two or more techniques, it is important to identify the specific actions or tasks that one wishes to accomplish with the assistance of visualization.

i) identify - to recognize an object on the characteristics presented, such as finding a fracture in an x-ray.

ii) locate - to establish the position of an object is distinct ~~or~~ dif such as determining the location of maximal stress in structural analysis.

iii) distinguish - to determine that an object is distinct or different from another.

iv) categorize - to classify objects into distinct types.

v) cluster - to group similar objects based on some relationship.

vi) rank - to place a group of objects in an order, such as numerical or chronological.

vii) compare - to examine the similarities and differences b/w two or more objects.

viii) associate - to draw a relationship b/w two or more objects, such as linking temp and location in weather map.

ix) correlate - To find a causal or reciprocal relationship b/w two or more objects such as determining the relationship b/w interest rates and economic growth.

2) User characteristics :-

Besides the tasks to be performed, the effectiveness of a visualization technique is tightly associated with the users of the visualization.

→ user can be classified based on their knowledge and skills.

i) familiarity with domain : How much expertise does the user have the domain of the data being explored? Has she studied the field for a long time, or is she relatively new to the field?

ii) familiarity with task : How much experience has the user had in performing the desired task? Is she an expert or somewhere in between?

iii) familiarity with data : Has the user examined this data previously and formed a reasonable mental model of its contents, or is this her first exposure to it? Is it similar to other data sets she has examined?

iv) familiarity with the visualization technique : Is this the user's first attempt to interpret the data using this particular kind of visualization, or she spent considerable time using the technique?

4) Familiarity with the visualization environment: Has the user employed the particular tool in the past, or is it her first exposure?

5) <u>Visualization characteristics</u> :-

Computational performance: How quickly can the visualization be generated, using data sets of various sizes?

Memory performance: How much computer memory is required to generate the visualization?

Data limitations: What are the upper and lower bounds for the size and complexity of the data that can be visualized with this technique?

Degree of complexity: What is the normal learning curve for the technique? How many parameters does the user need to set in order to generate views?

Degree of usability: How easy is it to perform the task? How intuitive is the interpretation of the visualization? How intuitive are the controls for interactions?

Degree of accuracy: How frequently is the user successful or unsuccessful in performing the desired task with this technique?

7981808340

6) Visualization systems:-

i) systems based on Data type

ii) system based on Analysis type

i) system based on datatype - scientific

openDX, which was formerly marketed as IBM visualization Data Explorer, is an extensible visualization environment primarily used for the analysis of scientific and engineering data.

→ the modules fall into several distinct classes, including:-

* import and export - modules to load and save data in different formats;

*. flow control - modules to create loops and conditional execution;

*. realization - modules to map the data to renderable entities, such as isosurfaces, grids and streamlines;

*. rendering - modules to control display attributes, such as lighting cameras, and clipping;

* transformation - functions to apply to the data, such as filtering, mathematical functions and sorting.

*. interactor - widgets such as file selectors, menus, dials/ sliders and button boxes.

Systems based on multivariate data:-

XmdvTool [280], developed at Worcester polytechnic Institue (WPI) by Matthew Ward, Elke Rundensteiner software package that integrates five common methods for multivariate data visualization into a single exploration application.

→ this tool includes standard scatterplot matrices scatterplots of star glyphs, parallel coordinates, dimensional stacking, and pixel-oriented techniques.

→ these visualizations are linked together using a simple selection and highlighting mechanism called an N-dimensional brush, which defines a hyperbox in the data space.

→ selected data in one view are also selected in other views and the resulting selection can be highlighted, masked, deleted or analyzed separate from the rest of the data

ii) System based on Analysis type - statistics

the system continues to evolve, with many others contributing to its development.

→ It supports a number of different visualizations including scatterplot matrices, bar charts, graphs, and parallel coordinates.

→ for each visualization, a control panel specific to that view is shown; clicking on any visualization exposes its control panel.

the when
and

→ color is used to link data between multiple views, and the user has a wide range of options for controlling the colors assigned to graphical entities.

→ the user starts by selecting a data dimension to control the color; an interactive histogram can then be used to adjust the ranges of values assigned to each color.

system based on analysis type - spatio Temporal:-

Macrofocus has produced a number of powerful interactive tools for visually exploring data and information.

→ one such tool is Infoscope, which links geographic views with several other visual and text representations of information.

7) Text Analysis and visualization:-

Jigsaw is a text visualization tool that was developed by John stasko and his students at the Georgia institute of Technology.

→ this system explores entities (people, places, dates, money) and the relationships between them.

→ Jigsaw uses several different views to present information to the user including calendar, list, graph, scatterplot, text, and time line views.

Each view is presented as a separate window that updates automatically with the results of various queries.

→ This tool is especially useful in data exploration involving large collections of data and information.

→ applications include military intelligence, law enforcement, or journalism, among many others.

8) Structures for Evaluating visualizations :-

Many forms of evaluation for interactive systems have been developed within the field of human - computer interaction,

Most of which are applicable and have been applied to data and information visualization systems.

1. Usability Tests
2. Expert Reviews
3. field Tests
4. Case study

1) Usability tests :-

these evaluations concentrate on "the Five E's"

effective
efficient
engaging
error tolerant and ease of learning

These tests are usually carried out by observing users attempting to perform tasks, and noting the types of difficulties they are having, the features they commonly use, and their level of comfort/ satisfaction with the tool.

2) Expert review:-

while some forms of evaluation depend on having a significant number of participants, expert reviews can generally be carried out with a small number of qualified reviewers.

-> These evaluators may be experts in visualization, or they may be domain experts who can thoroughly test the applicability of a tool for a set of specific application tasks.

Field Tests & usecases:-

unlike usability tests, which are often carried out in a controlled environment over a short period of time to better enable measurement.

-> field tests are performed in the natural environment of the typical user and may last for weeks or months.

Benchmarking procedures:-

Benchmarking is a formal procedure for evaluating the performance of some object or set of objects.

→ Benchmarks can be quantiative or qualitative
→ Qualiatative benchmarks generally involve the use of human subject, while quantitative can be done either with or without the use of human subjects.

1) formulate a hypothesis: A bench mark requires a specific statement about one or more attributes of the object being assesed.

2) Design the experiments: The key to designing benchmark experiments is to create tests that vary only a single attribute at a time.

3) Execute the experiments: there are many ways a well-designed experiment can be executed so that little in the way of reliable conclusions will results.

9) Toolkits :-

→ the visualization Toolkit is an open source toolkit for building 3D visualizations that include computer graphics, modelling, imaging and both scientific and information visualization.

4/10 — analysing

Toolkit, 436

UNIT-IV

1) Interaction operands and space :- techniques for interaction

1) screen space (pixels)
2) Data value space (multivariate data values)
3) Data structure space (components of Data organization)
4) Attribute space (components of graphical entities)
5) object space (3D surfaces)
6) visualization structure space

User interaction plays an integral part in the effective visualization of data and information.

→ An interaction operand is the section of space upon which an interactive operator is applied.

→ To determine the result of an interactive operation, one needs to know within what space the interaction is to take place.

1) Screen space :-

Navigation in screen space typically consists of actions such as panning, zooming, and rotation.

→ In each case, no new data is used.

→ the process consists of pixel-level operations such as transformation, sampling and replication

2) Data Value Space (Multivariate Data Values) :-

Data value space selection is similar to a database query in that the user specifies a range of data values for one or more data dimensions.

Navigating in data values space involves using the data values as a mechanism for view specification.

3) Data structure space (components of Data organization)

→ Data can be structured in a number of ways such as lists, tables, grids, hierarchies and graphs.

→ for each structure, one can develop interaction mechanisms to indicate what portions of the structure will be manipulated, and how this manipulation will be manifested.

→ Navigation in data structure space involves moving the view specification along the structure.

4) Attribute space (components of Graphical entities)

→ Navigation in attribute space is similar to that in data value space.

→ panning involves shifting the range of the values of interest, while zooming can be accomplished by either scaling the attributes or enlarging the range of values of interest.

→ Remapping is often done in attribute space.

5) object space (3D surfaces):-

In these displays, the data is mapped to a geometric object, and this object can undergo interactions and transformations.

→ Navigation in object space often consists of moving around objects and observing the surfaces on which the data is mapped.

→ the system should support global views of the object space as well as close-up views.

6) visualization structure space:-

→ A visualization consists of a structure that is relatively independent of the value attributes, and structures of data.

→ Navigation in visualization structure space might include moving through pages in a spreadsheet-style visualization tool (or) zooming in on an individual plot in a scatterplot matrix.

2) A unified Framework :- ①

In computer systems, a framework is often a layered structure indicating what kind of programs ~~capr or~~ should be built and how they would interrelate.

→ for each interaction operator to be applied to a specified space/operands, several parameters are required. some of these may be constants for a given system.

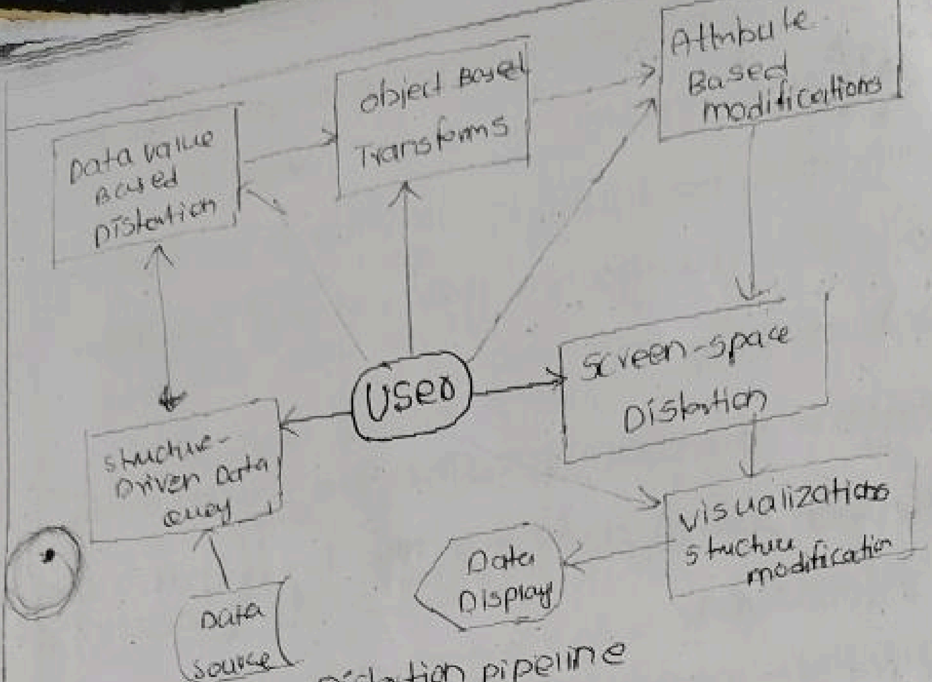
1) focus
2) Extents
3) transformation
4) Blender

Focus:- The location within the space
centre of the area of user interest.
→ There may be multiple simultaneous foci,
though for navigation this usually requires
multiple display windows

ii) Extents:- The range within the space (can be
multidimensional) defining the boundaries of
the interaction.
→ The metric used for specifying the range is
specific to the space;
→ In screen space this could be in pixels,
while in structure-space this might be the
number of rows in a table or inks in a graph.

iii) Transformation:-
The function applied to the entities within the
extents, generally a function of distance
or offset from the focus.
→ The shape of the transformation might
also depend on the type of information
being affected.

iv) Blender:-
How to handle parts of space touched by
more than one interaction.



Distortion pipeline

3) Interaction Control :-
At each stage of the pipeline, the user requires mechanisms to control the type, location, and level of each interaction as he or she navigates within both the data space and the visualization.

focus selection :-
selection is most readily accomplished via direct manipulation tools,
eg: using a mouse or other selection device to indicate the focus location.

Extent selection :-
specifying the extents for an interaction is generally dependent on the type of interaction and the space in which the interaction is being applied

Interaction type selection :- Give the many types of interaction possible, and the variety of spaces in which they may be applied, a reasonable interface for this task would be a pair of menus.

Interaction level selection:- the degree of interaction is an important control parameter that can be specified by a single value.

Blender type selection:- If more than one interaction can be simultaneously viewed and manipulated, there must be some mechanism for selecting a strategy for mixing regions of space affected by more than one interaction.

3rd unit

Relationships

Relationships can be simple or complex: unidirectional or bi-directional, Non-weighted or weighted, certain or uncertain.

1) **Displaying Hierarchical structures:-**
(trees, Graphs, networks)

i) Trees or Hierarchies :-

Trees or hierarchies are one of the most common structures to hold relational information.

we can divide these techniques into two classes of algorithms:

Hierarchies
→ Space filling
→ Non-space filling

i) Space filling :- As the name implies, space-filling techniques make maximal use of the display space.

→ ~~This is~~ the two most common approaches to generate space-filling hierarchies are rectangular and radial layouts.

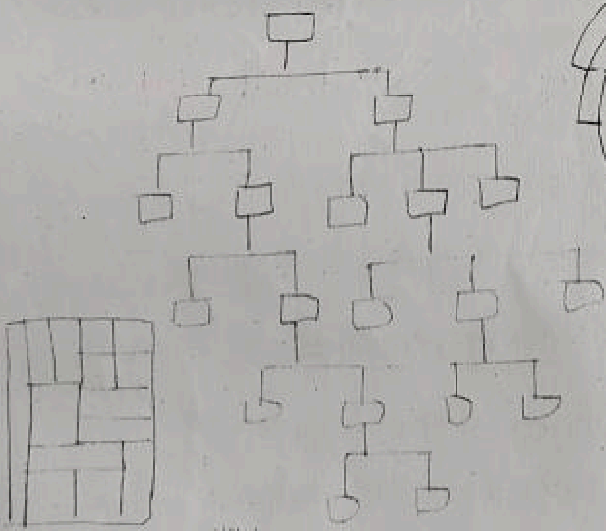
→ Treemaps and their many variants are the most popular form of rectangular space-filling layout.

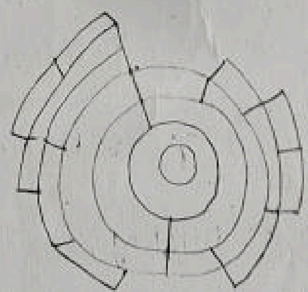
→ In the basic treemap, a rectangle is recursively divided into slices, based on the populations of the subtrees at given level.

including squarified treemaps (to reduce the occurrence of long, thin rectangles) and nested treemaps (to emphasize the hierarchical structure).

→ Radial space-filling hierarchy visualizations, sometimes referred to as sunburst displays, have the root of the hierarchy in the centre of the display and use nested rings to convey the layers of the hierarchy.

→ ~~coto~~ symbols and other markings may also be embedded in the rectangular or circular segments to communicate other data features.





a sample hierarchy and the corresponding

Non space filling methods:-

The most common representation used to visualize tree or hierarchical relationships is a node-link diagram.

when designing an algorithm for drawing any node-link diagram (not just trees) one must consider three categories of often-contradictory guidelines: drawing conventions, constraints, and aesthetics.

→ conventions may include ~~requiring~~ restricting edges to be either a single straight line, a series of rectilinear lines, polygonal lines, or curves.

→ constraints may include requiring a particular node to be at the centre of the display, or that a group of nodes be located close to ~~close to~~ each other (or) from top to bottom (or) left to right.

→ Aesthetics, however, often have significant impact on the interpretability of a tree drawing. some typical rules are:

* Minimize line crossing
* Maintain a pleasing aspect ratio
* minimize the total area of the drawing

rules for non space filling methods.

→ minimize the total length of the edges

→ " " no of bends in the edges

→ " " no. of distinct angles or curvature used

→ strive for a symmetric structure.