1. Mention and explain three uses of clustering in data visualization.

Clustering is the process of making a group of abstract objects into classes of similar objects.

Points to Remember

- A cluster of data objects can be treated as one group.
- While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups.
- The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

Applications of Cluster Analysis

- Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.
- Clustering can also help marketers discover distinct groups in their customer base. And they can characterize their customer groups based on the purchasing patterns.
- In the field of biology, it can be used to derive plant and animal taxonomies, categorize genes with similar functionalities and gain insight into structures inherent to populations.
- Clustering also helps in identification of areas of similar land use in an earth observation database. It also helps in the identification of groups of houses in a city according to house type, value, and geographic location.
- Clustering also helps in classifying documents on the web for information discovery.
- Clustering is also used in outlier detection applications such as detection of credit card fraud.
- As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

Requirements of Clustering in Data Mining

The following points throw light on why clustering is required in data mining −

- Scalability − We need highly scalable clustering algorithms to deal with large databases.
- Ability to deal with different kinds of attributes − Algorithms should be capable to be applied on any kind of data such as interval-based (numerical) data, categorical, and binary data.
- Discovery of clusters with attribute shape − The clustering algorithm should be capable of detecting clusters of arbitrary shape. They should not be bounded to only distance measures that tend to find spherical cluster of small sizes.
- High dimensionality − The clustering algorithm should not only be able to handle low-dimensional data but also the high dimensional space.

- Ability to deal with noisy data − Databases contain noisy, missing or erroneous data. Some algorithms are sensitive to such data and may lead to poor quality clusters.
- Interpretability − The clustering results should be interpretable, comprehensible, and usable.

Ref: https://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm

2. Mention and explain three uses of force-directed algorithms in data visualization.

A Force-Directed Graph, or Force-Based Graph, is a type of layout commonly used in a variety of application areas: network visualization, large graph visualization, knowledge representation, system management, or mesh visualization.

It is used to visualize the connections between objects in a network. By grouping the objects connected to each other in a natural way, a Force-Directed Graph is visually interesting and also makes it possible to discover subtle relationships between groups.

These graphs have the particularity of being drawn by algorithms called "force-directed graph drawing algorithms" or "force-directed graph drawing algorithms".

Good-quality results

At least for graphs of medium size (up to 50–500 vertices), the results obtained have usually very good results based on the following criteria: uniform edge length, uniform vertex distribution and showing symmetry. This last criterion is among the most important ones and is hard to achieve with any other type of algorithm.

Flexibility

Force-directed algorithms can be easily adapted and extended to fulfill additional aesthetic criteria. This makes them the most versatile class of graph drawing algorithms. Examples of existing extensions include the ones for directed graphs, 3D graph drawing,[6] cluster graph drawing, constrained graph drawing, and dynamic graph drawing.

Intuitive

Since they are based on physical analogies of common objects, like springs, the behavior of the algorithms is relatively easy to predict and understand. This is not the case with other types of graph-drawing algorithms.

Simplicity

Typical force-directed algorithms are simple and can be implemented in a few lines of code. Other classes of graph-drawing algorithms, like the ones for orthogonal layouts, are usually much more involved.

Interactivity

Another advantage of this class of algorithm is the interactive aspect. By drawing the intermediate stages of the graph, the user can follow how the graph evolves, seeing it unfold from a tangled mess into a good-looking configuration. In some interactive graph drawing tools, the user can pull one or more nodes out of their equilibrium state and watch them migrate back into position. This makes them a preferred choice for dynamic and online graph-drawing systems.

Ref: https://www.toucantoco.com/en/glossary/force-directed-graph.html#:~:text=A%20Force%2DDirected%20Graph%2C%20or,between%20objects%20in%20a%20network. https://en.wikipedia.org/wiki/Force-directed_graph_drawing#Advantages

3. What are stack (or stacked) graphs? What is their main use? How would you apply them for text visualizations?

This type of visualisation depicts items stacked one on top (column) of the other or side-by-side (bar), differentiated by coloured bars or strips. Items are "stacked" in this type of graph allowing the user to add up the underlying data points. Stacked graphs should be used when the sum of the values is as important as the individual items.
Uses:
A stacked graph is useful for looking at changes in, for example, expenditures added up over time, across several products or services.
Stacked graphs are commonly used on bars, to show multiple values for individual categories, or lines, to show multiple values over time. Thus, stacked graphs must always work with positive values.
Stacked line graphs often show how quantities have changed over time.

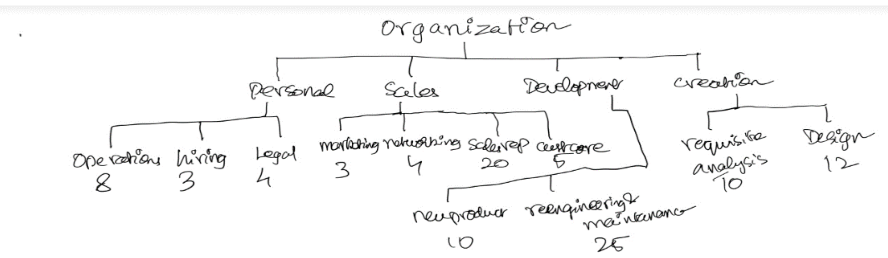
theme river
- each one of these layer is how a topic changed over time
- process a doc along timeline and compile different topics along something
- no of times a doc/ word used in social discussion in a duration
- size of the flow is the count, flow itself(diff color) is a theme/ sub/particular partition of dataset(cluster)
- for eg 10 clusters than change over time

4. An analyst is organising a data set for managing the numbers of employees of the various departments of the company. The following text describes the data:

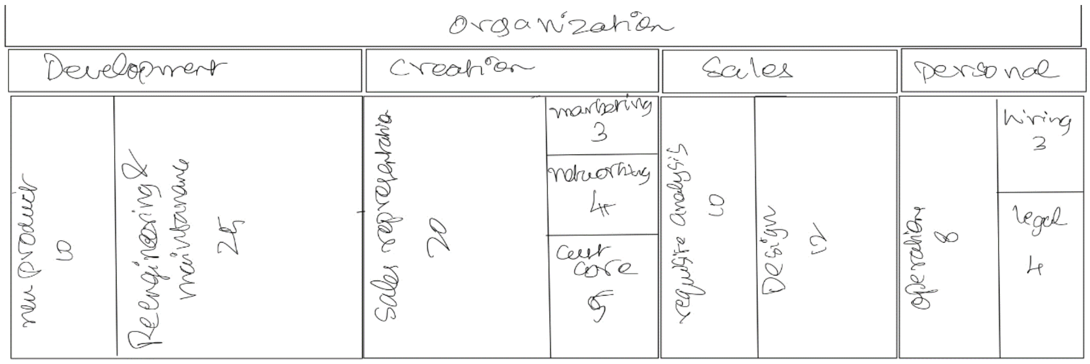
Spread Sheet:

Personal			Sales				Development		Creation	
Operations	Hiring	Legal	Marketing	Networking	Sales Representations	Customer Care	New Product	Reengineering and maintainance	Requisites analysis	Design
8	3	4	3	4	20	5	10	25	10	12

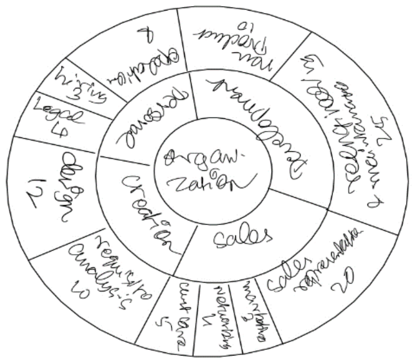
Node Link Tree:



Tree Map:



Sunburst:



5. **An analyst is designing a visualization support real state agents to keep control of contents of houses they rent. She has built a fake data set to test alternatives. The following text describes the data:**

6. **Draw a comparative table of 5 multidimensional visualization techniques. What are they useful for? What type of observations do they provide? What is their input? What are the advantages and disadvantages? What applications could they support?**

   Sunburst, force-based graph, tree map, node link, Spreadsheet, stack graph

   Sunburst :

   Input : Hierarchical dataset

   Usefulness : Each level of the hierarchy is represented by one ring or circle with the innermost circle as the top of the hierarchy. A sunburst chart without any hierarchical data (one level of categories), looks similar to a doughnut chart. However, a sunburst chart with multiple levels of categories shows how the outer rings relate to the inner rings. The sunburst chart is most effective at showing how one ring is broken into its contributing pieces, while another type of hierarchical chart, the treemap chart, is ideal for comparing relative sizes.

   Observations : Sunburst charts shows the hierarchy of the categories, thereby giving a very clear understanding of the lineage.

   Advantage : It presents the entire hierarchical data in a single-screen snapshot, which makes it a preferred choice over a tree view.

Disadvantage :The radial structure limits the number of nested levels that can be easily understood from a static picture.
Understanding angular readings is difficult for human eye.
https://www.fusioncharts.com/resources/chart-primers/sunburst-chart

Tree map :
Input : tree data structure
Useful for representing hierarchies. Eg Phylogenic trees
Observations provided : relationships between various points. Parent child relationship. Evolutionary details.
Application : Treemaps are often used for sales data, as they capture relative sizes of data categories, allowing for quick perception of the items that are large contributors to each category. Color can identify items that are underperforming (or overperforming) compared to their siblings from the same category.
Disadvantage : In a treemap, as we go down the hierarchical levels, the space available to plot decrease dramatically. This sets a limitation to the number of hierarchical levels that can be displayed at once. As the categories delve deeper, they become harder to read. This is good for comparing macro-level data and giving viewers a sense of how many sub-categories there are. But it isn't really effective when you want to drill down into those sub categories.
Use a treemap to display hierarchical information presented using a tree structure. You can, for instance, use a treemap to show the folder structure of a system. You can also use it to study patterns and occurrences in a large dataset since the color, and the size dimensions are correlated.
Advantage of tree Map :

Force-based graph : Proximity is based on strength. It's a graph so it's relationship based.
Input : Graph
Useful : The algorithm is able to identify some predefined regular substructures like chains, stars, cycles, and parallel structures and arrange them in an optimized manner.
Advantage : Suppose 2 text documents are completely different, then the distance will be too large. But the distance is measured in such a way that it puts a limit on the largest distance so that extremely large distances can be represented effectively.
Disadvantage : Projection technique is fast but lacks precision.
Application : network visualization, large graph visualization, knowledge representation, system management, or mesh visualization. It is used to visualize the connections between objects in a network.

Node-link
Input : Graph
Use : A node link diagram is an excellent construct for the many use cases where the key questions are around identifying and understanding how and whether entities are connected. Node link diagrams are extremely useful in use cases such as intelligence analysis (e.g., one person is an associate of a suspect or known criminal), fraud detection (e.g., the same social security number was used by different people), and many others.

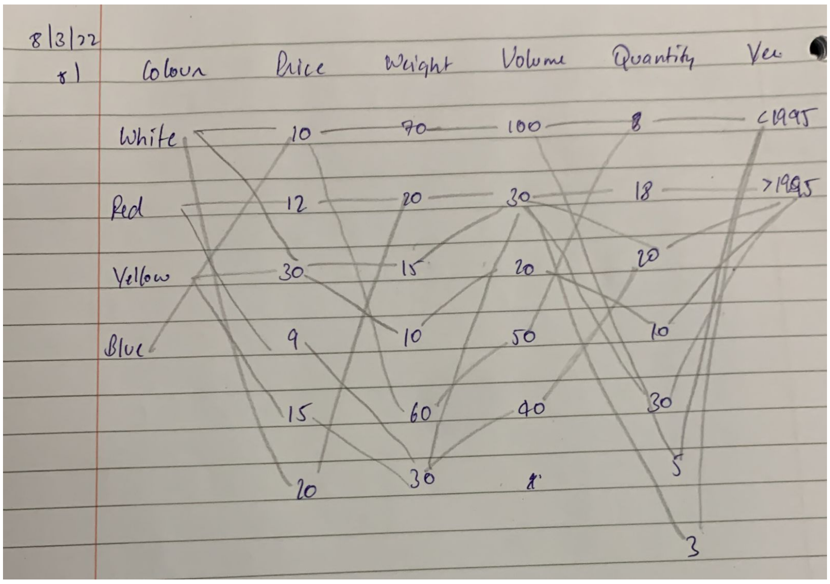
Advantage : significantly better performance for the two network connectivity tasks- identifying the number of clusters and finding a path between two nodes-than the other two representations.
Disadvantage : In an ideal situation, nodes are automatically placed into positions that allow the user to simply group them visually into clusters on the basis of their proximity and thereby deliver new knowledge about network structure, which would not be possible without such representation. Unfortunately, in most cases, this is not possible - many network datasets are so complex that their graph visualization looks like a "hairball", where the nodes are no longer able to be positioned according to the proximity rule.

7. **Visualization Task: Suppose you want to produce a visualization that summarizes and helps as a guide to a web-based course. The course is composed by series of linked web pages plus external links and references. Pages can be classes, tasks or additional information. The user must understand the general structure of the course, verify what pages he/she visited and for how long, and quickly find references given for each class. Describe the elements of your visualization and discuss the possible interactive functionalities the user would have in this context. Draw a schema for your visualization. Justify your choices of visual elements. What would change if the visualization were designed to support the teacher in improving the course?**
Yet to do

8. **Visualization Task: Given the Data set in Table 1, draw the parallel coordinates visualization for it. What can you conclude about the data from the visualization?**

Table 1: Products

Color	**Price**	**Weight**	**Volume**	**Quantity**	**Year Fab.**
White	10	70	100	8	Before 1995
Red	12	20	30	18	After 1995
White	30	15	30	20	After 1995
Yellow	30	10	20	10	After 1995
Blue	10	60	50	8	Before 1995
Red	9	30	40	20	After 1995
Yellow	15	30	30	30	After 1995
Blue	10	70	100	5	Before 1995
White	20	20	30	3	Before 1995



9. What is the distinction you make between Information Visualization (InfoVis) and Scientific Visualization (SciVis)? Mention two common points and five distinct features between them.

Scivis applications mainly focus on so-called physical data, which has an inherent spatial placement, such as the flow of water in a 3D container or a medical scan of a patient limb (see Chapter 10). In such cases, the user already has a mental image of what the flow container or the limb looks like.

In contrast, many infovis applications attempt to help users form a mental image about data that has no inherent physical, or spatial, placement. Information has no "innate shape and color" [Koike 93], so its visualization has a purely abstract character. computer file systems, databases, documents from archives, and stock exchange courses

•Information visualization covers areas such as visual reasoning, visual data modeling, visual programming, visual information retrieval and browsing, visualization of program execution, visual languages, visual interface design, and spatial reasoning

Scivis applications mainly focus on measurements and simulations of real-world data

•Infovis data is discrete and cannot be used for interpolation

Scivis data is continuous and more suitable for interpolation

•Scivis data types are scalar, vector or tensors

Infovis data types are nominal, numerical, categorical, relationship

•The geometry of the visual depends on the phenomenon. For eg when we analyse a wing of an airplane, we know what the design of a wing looks like and the visual will be same.

On the other hand, in case of infovis the geometry completely depends on the designer. He can design any visual as per the requirement.

•engineering and computational fluid mechanics and mathematics to medical and Earth sciences

data range from generic graphs and trees to database tables, text, and computer software

10. What differs InfoVis and SciVis regarding data types?

	Scivis	Infovis
Data domain	spatial $\subset \mathbb{R}^n$	abstract, non-spatial
Attribute types	numeric $\subset \mathbb{R}^m$	any data types
Data points	samples of attributes over domain	tuples of attributes without spatial location
Cells	support interpolation	describe relations
Interpolation	piecewise continuous	can be inexistent

11. What is the distinctions between Direct Volume Rendering and Surface based Rendering?

Surface rendering relies on an assumption about the underlying structures you are visualizing from the data. In other words, you are estimating or making assumptions about what's underneath the surface based on its structure.

In volume rendering, assessing that underlying structure is part of the visualization process, and there is no such assumption. Instead, the nature of the data at each voxel is analyzed. Based on that analysis, colors and opacities are assigned, calculations are made, and the structures are visualized based on optical behavior of the components.

12. What are the distinctions of 2D reconstruction and 3D reconstructions and in what circumstance each one would be applied?

2d reconstruction - we are reconstructing surfaces. It is embedded in 3d but the elements of visualization are surfaces

3d - for some reasons, after reconstruction if we need volume, go for 3d reconstruction. Output of 3d reconstruction is a mesh of planar elements. It can model the inside of the surface together with the outside. It is especially useful in cases where we want to study the variation of density

13. What are the main strategies to visualize hierarchies? What distinguishes them?

Tree Diagram

The idea behind the Tree Diagram hierarchical data visualization is to showcase relations of chart objects against each other in ranked order — it starts with a single value residing on top and all other values spiralling from it and their relations with each other.

It is often used in organizations to determine superiors and subordinates in a management system, showcase seniority in family trees, display the hierarchy of items according to their value like in Maslow's pyramid of needs, to fast-forward decision making and track the cause of a certain problem and its possible effects.

Treemap

Treemaps not only show value hierarchy by splitting the whole area into smaller rectangle pieces but also show value relations by obtaining rectangles of different sizes within each split category.

Visualizing hierarchical data using Treemaps is a great way to show items relations to the whole and to each other in a single system.

Sunburst Diagram

Here hierarchical data format is distributed in a way, quite similar to the Treemap, but instead of a rectangle, a whole circle is divided into separate categories that take up space in accordance with its value.
Sunburn Diagrams have multiple levels, which are using a hierarchical data structure and are placed one below the other in correspondence to each other's values.

Circle Packing

Circle Packing displays a whole area with a single circle and the biggest circles representing the main categories, with each smaller circle inside representing a value in proportion to the parent and the entire circle accordingly.
This visualized hierarchical data structure is best suited to display a large area with multiple elements making it up, and their relations to the whole and to each other.
REF: https://insightwhale.medium.com/how-to-show-hierarchy-with-data-visualization-526fb45ee4c2

14. Give an example of an application and a task or question for which a hierarchy is a natural application. Give an example of an application and a task or question for which a hierarchy can be used, but is non-native.

Native:

There is a hierarchical nature in every government organization of a country like parliament, army, etc. For example, in Army, we have Field Marshals -> Generals -> Lieutenant Generals -> Major General -> Brigadier -> Colonel -> Lieutenant Colonel -> Major -> Captain -> Lieutenant. When we have two entities under same group, they cannot interact with each other as one is not dependent on another. Here, hierarchy is a natural application.

Non-Native:

Documents in an organization gets transferred from one department to another. This flow of documents through various departments can be represented by hierarchy but it is non-native because, the documents can go back to the departments from where it came.

15. Mention and explain three different types of layout for a tree, as well as their advantages and disadvantages.

Similarity trees:

They organize data objects on the visual plane emphasizing their levels of similarity with high capability of detecting and separating groups and subgroups of objects.

Advantages:

ability to decrease point clutter;
high precision;
consistent view of the data set during focusing, offering a very intuitive way to view the general structure of the data set as well as to drill down to groups and subgroups of interest.

Disadvantages:

their computational cost
presence of virtual nodes that utilize too much of the visual space.

Classification Trees:

Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

Advantages:

1. Compared to other algorithms decision trees requires less effort for data preparation during pre-processing.
2. A decision tree does not require normalization of data.
3. A decision tree does not require scaling of data as well.

4. Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent.
5. A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders.

Disadvantage:

1. A small change in the data can cause a large change in the structure of the decision tree causing instability.
2. For a Decision tree sometimes calculation can go far more complex compared to other algorithms.
3. Decision tree often involves higher time to train the model.
4. Decision tree training is relatively expensive as the complexity and time has taken are more.
5. The Decision Tree algorithm is inadequate for applying regression and predicting continuous values.

Tree maps:

Treemaps display hierarchical (tree-structured) data as a set of nested rectangles. Each branch of the tree is given a rectangle, which is then tiled with smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension of the data.[1] Often the leaf nodes are colored to show a separate dimension of the data.

Advantages:

When the color and size dimensions are correlated in some way with the tree structure, one can often easily see patterns that would be difficult to spot in other ways.

A second advantage of treemaps is that, by construction, they make efficient use of space.

Disadvantage:

Size Distortion: The more pixels you use to show the hierarchy, the more the size distorts. You quickly run into an issue that you can only optimize for size comparisons at one level of the hierarchy at a time.

Requires Interactivity: Treemaps are ill-suited for print. Unless you have only a few data points, the labels quickly disappear or become unreadable.

###THIS IS THE ANSWER#####

Rooted tree layout:

Here, all children nodes of the same parent have the same y-coordinate. Their position on the x-axis is used to reflect a certain ordering. Here, the order in which files appear in a listing of their containing folder is used. Moreover, the placement of the children on the x-axis is done so that the horizontal extent of the entire subtree of a given parent node is centered with respect to the parent's x-coordinate. The appearance of the nodes reflects several attributes. The node glyphs and color show the file level in the hierarchy: a green square for the root, blue rings for the first-level folders, yellow balls for the second-level folders, and red squares for the remaining folders and files. The node glyph sizes reflect the size of the corresponding files or folders. However, these sizes range from a few tens of bytes for the smallest ones to over a hundred megabytes for the root folder.

Radial tree layout:

A similar assessment of the file hierarchy structure can be obtained using the so-called radial tree layout shown in Figure 11.6. Here, the root is placed at the center of the image and its children are distributed in clockwise order along a circle centered at the root. Nodes on deeper levels are laid out on correspondingly wider circles. The (x, y) coordinate system used for the layout in Figure 11.5 is now replaced with a polar ρ, α coordinate system, hence the name radial. The advantage of the radial layout is that it always has a one-to-one aspect ratio since the entire picture always fits in a circle. Moreover, more space is allocated, relatively speaking, to the deeper levels in the tree than in the previous layout. This is visible if we compare the space allocated to the leaves of the tests folder, drawn in red in Figures 11.5 and 11.6. In contrast, there is less space allocated to the upper levels of the tree. This can create problems when these nodes need more space to be drawn, as in the case of the first level of our example, whose nodes have large icons and display textual annotations.

Bubble tree layout:

In contrast to the radial layout, where a subtree occupies a pie sector of the entire layout, in this new layout a subtree always occupies an entire circle centered at the subtree's root. The entire layout can be seen as a placement of circles inside other larger circles; hence this layout is known as bubble, or balloon, layout [Boardman 00, Grivet et al. 04]. In contrast to the rooted and radial layout, the bubble layout offers a better visual separation of the subtrees. Additionally, this layout also keeps the tight aspect ratio of the radial layout. A difference between the two previous layouts is that edges have now considerably different lengths—subtrees with fewer children will have shorter edges, since their bubbles will have smaller radii. This makes the visual size of the subtrees reflect their number of children more strongly than for the rooted and radial layouts.

Cone tree layout:

A related tree layout to the bubble layout is the cone layout (see Figure 11.8). Here, the nodes are arranged in three dimensions, rather than two dimensions as in the three layouts discussed so far. Whereas a subtree was laid out in a circle in the bubble layout, it is now laid out as a cone with the subtree root placed at the apex [Robertson et al. 91]. The advantage of the cone layout is that its 2D rendering may be more compact than other layouts, when viewed from an optimal angle. Also, the visual separation of different subtrees may be easier when interactively rotating the viewpoint around the cone layout. Nevertheless, the cone layout inherits the typical problems of most 3D visualizations of discrete entities: occlusion, the chance of "getting lost" in the 3D space during interactive viewing, and the potential confusion created by the foreshortening of the perspective projection.

Treemaps:

are a different layout for tree structures that use virtually every pixel of display space to convey information. The basic idea behind the treemap layout is simple: Every subtree is represented by a rectangle that is partitioned into smaller rectangles which correspond to its children.

Squarified treemaps. Both the aspect ratio and the nonleaf node visibility problems are addressed by a variant of the treemap layout shown in Figure 11.10. Let us first consider the layout. The principle of partitioning the rectangle of a subtree root into smaller rectangles corresponding to its children, sized to reflect the children's sizes, stays the same. However, instead of using the simple horizontal-vertical alternating slicing, we use now a more involved technique that tries to create cells of near-unity aspect ratio. The implementation of this technique, called squarified treemap layout, is described in detail by Bruls et al. [Bruls et al. 00], and can be easily reimplemented by the interested reader.

Cushion treemaps. The second design element present in the tree visualization shown in Figure 11.10 is the use of shading to reflect the hierarchical structure. Since only the tree leaves are explicitly rendered, we shall use their surface to show structural information on their position in the tree. Figure 11.10 shows a set of convex "bumps." Given their appearance, these bumps are called shaded cushions in the literature [van Wijk and van de Wetering 99]. If we look carefully, we notice that these bumps occur at several levels: Each treemap rectangle, which corresponds to a tree leaf, displays a small-scale bump. However, larger bumps are also visible in the shading signal or, to be more precise, in the luminance discontinuities of the shading.

16. What are the main strategies for visualizing networks? What are the main applications? What would be the role of centrality measures in these applications? What can be done to improve clutter problems in network visualisations? What can be done to support visual analysis of large networks?

Main strategies : similarities correlations, precedence, transformation

Main Applications : Communications in social networks, network traffic, word semantic represent in graph by having historic data. Routes - there are different paths in data so we use to analyse the topography (connectivity)

Role of centrality measure :

Eg : If you want to identify critical node in electrical web and maintain in order to eliminate any future disturbance to your network, centrality measure is important.

Improve clutter :

Group similar nodes together

Remove edges : only draw important data points

Node and edge aggregation.

Support large network vis : Use force directed graph

18. What insights can be drawn when visualizing text using: force-directed graphs, word trees and similarity trees? What is the purpose of word clouds?

Force - Based :

Used to visualise connections between objects in a network. Used for large graph visualization as it shrinks the graph. Detailing topics in the cluster.

Word trees:

Word trees show a pre-selected word(s) and how it is connected to other words in text-based data through a visual branching structure. Unlike word clouds, word trees visually display the connection of words in the dataset, providing some context to their use. Words that show up more frequently in combination with the pre-selected word(s) are displayed in larger font size. The visualisation allows users to choose whether they are interested in connections preceding a word or following a word.

Purpose of word cloud:

Word Clouds display the most prominent or frequent words in a body of text. Typically, a Word Cloud will ignore the most common words in the language (“a”, “an”, “the” etc). The remaining words are displayed in a “cloud” with the font size of the word (and-or the colouring of the characters in the word) depicting the relative frequency of occurrence of each target word in the source material. Word clouds helps in showcasing written data in a visual manner. It provides fast insights into the more relevant keywords in the data, summarises large volumes of text and reveal trends and find patterns in the data.

Similarity tree :

Intuitive way to visualise the general structure of dataset as well as to drill down groups and sub groups of interest. Supports feature analysis of selection.

19. What is the role of data summarization in InfoVis? Exemplify data summarization for two different types of data.

Summarising techniques like pattern detection, classification and clustering, dimensionality reduction, and aggregation can be used to compress large multi-dimensional datasets into smaller datasets which still retains the principal characteristics of the original data. Summarizing large datasets before visualizing the data, could help in more effective visualizations and support more efficient and accurate visual analysis.

1. The bivariate relationship between two categorical variables is summarized using a contingency table.
2. Quantitative data are grouped into classes. The lower limit of a class equals the smallest value within that class.
3. To summarize the relationship between a quantitative variable and a categorical variable, we calculate summary statistics for the quantitative variable for each level of the categorical variable

The term Data Summarization refers to presenting the summary of generated data in an easily comprehensible and informative manner.

We summarize data to "simplify" the data and quickly identify what looks "normal" and what looks odd. The distribution of a variable shows what values the variable takes and how often the variable takes these values.

The two most useful ways of describing the distribution of data are:

1. The typical: This describes the center–or middle–of the data. This way of describing the center is also called a "measure of central tendency".
2. The spread of the values around the center: This describes how densely the data is distributed around the center. This is also called a "measure of dispersion".

These two ways of describing the data are also referred to as descriptive statistics.
Types of data:

1. Categorical or Nominal
2. Ordinal
3. Continuous

20. (a) For the NJ-tree below, which of the Distance matrices CANNOT be the one that generated it? Why?
(b) Draw an NJ-tree for the answer in 20a
Yet to complete