Exam Preparation 2018 Interactive Data Visualization

Q. What is data visualisation?

Q.What is data visualization technique?

Ans: Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

Q. Outline and discuss three benefits of data visualisation over alternative means of interpretation

1. Absorb more information easily. ...
2. Discover relationships & patterns between business & operational activities. ...
3. Identify emerging trends faster. ...
4. Directly interact with data.
5. Quick Access to Relevant Business Insights
6. Predictive Sales Analysis
7. Drill-Down Sales Analysis
8. Easy Comprehension of Data
9. Customized Data-Visualization

Con-

Visualizations, like words, can be used to lie, mislead, or distort the truth.

Q. What is the purpose of data visualization? And what is data visualization kind of same

Ans: Data visualization is a general term that describes any effort to help people understand the significance of **data** by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based **data** can be exposed and recognized easier with **data visualization** software.

Why interactive data visualization?

- Variety of perspectives
- Dimensionality of information
- Representing multidimensional datasets fairly in compare with static images is notoriously difficult.
- A fixed image is ideal when alternate views are neither needed nor desired, and required when publishing to a static medium, such as print.

Q. Why do we use data visualization?Why is **data visualization** important?

Ans: Because of the way the human brain processes information, using charts or graphs to **visualize** large amounts of complex **data** is easier than poring over spreadsheets or reports. ... **Data visualization** can also: Identify areas that need attention or improvement.

Q. What are Scalar Vector Graphics?

Ans: - Scalable vector graphics (SVG) is a text-based graphics language that illustrates images with text.

Benefits:

- SVG files are lightweight and present top-notch graphics in print, as well as on the Web
- In addition, SVG supports animation and scripting.
- As a result, it is ideally suited for data-driven, interactive and personalized graphics.
- The SVG is an open standard specification that the World Wide Web Consortium (W3C) has been developing since 1999.

Some of the advantages of SVG images include:

- Compact compared to bitmapped graphics like JPEG and GIF files
- Can be searched, scripted, indexed and compressed
- Can be linked to various parts of a graphic
- Scalable
- Independent of resolution, so the image can be scaled up or down to match the display of all sizes on all kinds of Web devices
- Every attribute and every element in SVG files can be animated
- Image quality remains intact even if the images are resized or zoomed

Svg allows three types of graphis objects

- Vector graphics
- Bitmap images
- Text

** <svg width =”200” height=”200” viewBox=”0 0 100 100”>

The whole SVG canvas here is 200px by 200px in size. However, the viewBox attribute defines the portion of that canvas to display. These 200x200 pixels display an area that starts at user unit (0,0) and spans 100x100 user units to the right and to the bottom.

**

<svg width="200" height="250" version="1.1" xmlns="http://www.w3.org/2000/svg">

-size of the canvas

<rect x="10" y="10" width="30" height="30" stroke="black" fill="transparent" stroke-width="5"/>

That will draw a rectange

<rect x="60" y="10" rx="10" ry="10" width="30" height="30" stroke="black" fill="transparent" stroke-width="5"/>

-that will draw a rectangle with round shape

<circle cx="25" cy="75" r="20" stroke="red" fill="transparent" stroke-width="5"/>

That will draw a circle

<ellipse cx="75" cy="75" rx="20" ry="5" stroke="red" fill="transparent" stroke-width="5"/>

That will draw a cllipse

<line x1="10" x2="50" y1="110" y2="150" stroke="orange" stroke-width="5"/>

That will draw a line

<polyline points="60 110 65 120 70 115 75 130 80 125 85 140 90 135 95 150 100 145"

stroke="orange" fill="transparent" stroke-width="5"/>

that will draw a zik zak line

<polygon points="50 160 55 180 70 180 60 190 65 205 50 195 35 205 40 190 30 180 45 180"

stroke="green" fill="transparent" stroke-width="5"/>

that will draw a star

<path d="M20,230 Q40,205 50,230 T90,230" fill="none" stroke="blue" stroke-width="5"/>

That will draw a wave like line

</svg>

Closed path is -Z

Move – M

Line – L

There are an infinite number of Bezier curves, but only two simple ones are available in path elements: a cubic one, called with C, and a quadratic one, called with Q.

"Close Path" command, called with Z.

There are three commands that draw lines. The most generic is the "Line To" command, called with L. L takes two parameters—x and y coordinates—and draws a line from the current position to a new position.

There are two abbreviated forms for drawing horizontal and vertical lines. H draws a horizontal line, and V draws a vertical line. Both commands only take one argument since they only move in one direction.

here are five line commands for <path> nodes. As the name suggests, each one just draws a straight line between two points. The first command is the "Move To" or M,

The shape of a path element is defined by one attribute: d . The " d " attribute contains a series of commands and parameters used by those commands.

Each of the commands is instantiated (for example, creating a class, naming

The other type of curved line you can create using SVG is the arc, called with A. Arcs are sections of circles or ellipses

Q. D3 stands for data-driven documents. Explain what is meant by this..

Ans: **- D3**.js (**data**-**driven documents**) is an open source JavaScript library

- that allows users to apply prebuilt **data** visualizations to their **data**. ...
- It binds **data** to a **Document** Object Model (DOM), and changes the DOM **based** on the **data**, creating the visualization.
- It is the successor to the earlier Protovis framework.

Q. What is D3.js used for?

-With the massive amount of data being generated today,

- communicating this information is getting difficult.
- Visual representations of data are the most effective means of conveying meaningful information
- and D3 provides a great deal of ease and flexibility to create these data visualizations.
- It is dynamic, intuitive and needs minimum amount of effort.

What D3 Doesn't Do

- D3 doesn't generate predefined or "canned" visualizations for you.
- D3 doesn't even try to support older browsers
- D3 doesn't hide your original data. Because D3 code is executed on the client side, the data you want visualized must be sent to the client.
- If your data can't be shared, then don't use D3.

Advantages of D3

- D3.js is a Javascript library. So, it can be used with any JS framework of your choice like Angular.js, React.js or Ember.js.
- D3 focuses on data, so it is the most appropriate and specialized tool for data visualizations.
- D3 is open-source. So you can work with the source code and add your own features.
- It works with web standards so you don't need any other technology or plugin other than a browser to make use of D3.
- D3 works with web standards like HTML, CSS and SVG, there is no new learning or debugging tool required to work on D3.
- Since D3 is lightweight, and works directly with web standards, it is extremely fast and works well with large datasets.

- Since data visualizations are visual representations, it is convenient to use SVG to render visualizations using D3.

Q: What is a Word Cloud Anyway?
-Word clouds are a method for visually presenting text data.
-They are popular for text analysis because they make it easy to spot word frequencies.
-The more frequent the word is used, the larger and bolder it is displayed.
-Word Clouds Add Clarity
-Word clouds can identify trends and patterns that would otherwise be unclear or difficult to see in a tabular format.
-Frequently used keywords stand out better in a word cloud
-Word clouds are easy to understand
-A picture after all, is worth 1,000 words.
-Word clouds are easy to share

Who Is A Word Cloud For?

Word clouds are not just for researchers; marketers are using them to convey customer sentiments, and employers are using them to communicate internally with employees.

The Benefits of Using a Word Cloud

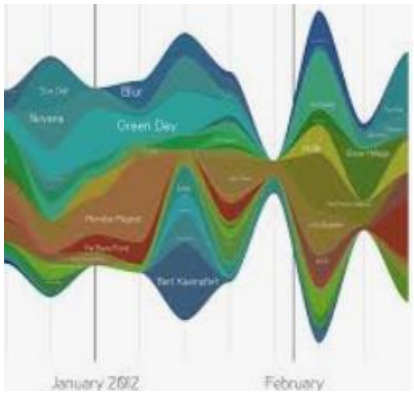
Word clouds can dazzle your audience with what might be otherwise viewed as every-day, information. Not only are they attention grabbing, they're also easy to use and inexpensive. But best of all, they are easy to understand and share.

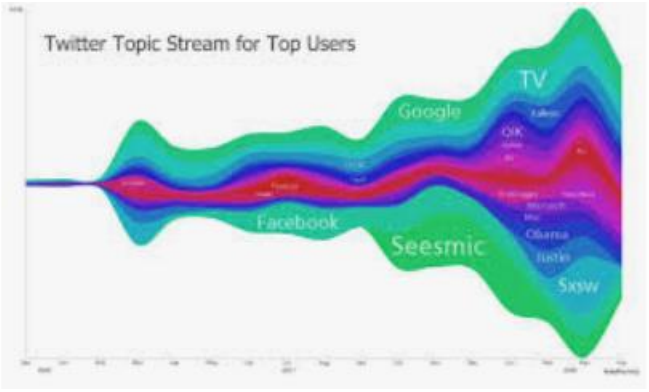
The Cons

Size isn't everything. Although the Word Cloud is designed to make words stand out according to their size based on their frequency of occurrence, other factors can affect the visual 'decoding' of the data from the observer's perspective. For example, the length of the word and the white space around the glyphs (letters) can make it look more or less important relative to others in the cloud. This can mislead your interpretation.

Q. What's the best interpolation?
Interpolation modifies the representation of data. Experiment with this drawing code to see how the different interpolation settings show different information than other interpolators

streamgraph
The streamgraph is a sublime piece of information visualization that represents variation and change, like the boxplot.
Ultimately, a streamgraph is a variant of what's known as a stacked chart.
They are often used to show changes of different categories over time when there are many categories AND these categories start and stop at different times. The height of each individual stream shape shows how the value of that stream has changed over time





Streamgraph, stacked bar charts, and stacked area charts are fundamentally the same thing and rely on the stack layout to format your dataset to draw it.

Stream graphs might be used to show how much money films have made when those films have not been released at exactly the same time. This would tell us something about the seasonality of cinema going

line-drawing

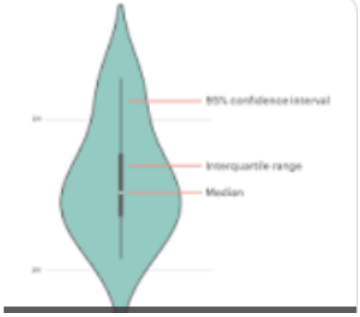
The line-drawing code produces a cluttered line chart. D3 provides d3.line and d3.area generators to draw lines or areas

You use d3.line when you want to draw most shapes and lines, whether filled or unfilled, or closed or open.
You should use d3.area when you want to draw a shape where the bottom of the shape can be calculated based on the top of the shape as you're drawing it.

Line graphs can also be used to compare changes over the same period of time for more than one group. Pie **charts** are best to **use** when you are trying to compare parts of a whole. They do not show changes over time. Bar **graphs** are used to compare things between different groups or to track changes over time.

Drawing violin plots

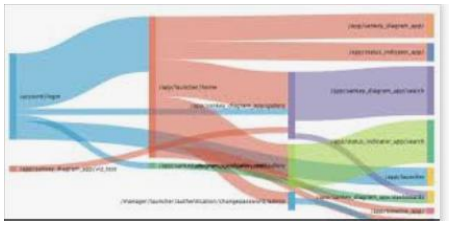
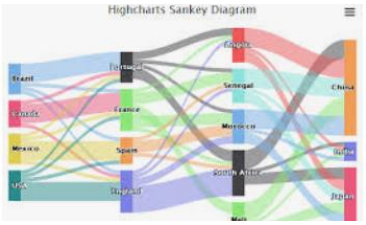
A violin plot is a mirrored curved area that bulges where many datapoints exist and tapers where few exist.
They're commonly seen in medical diagrams dealing with dosage and efficacy but also used more generally to show distributions,

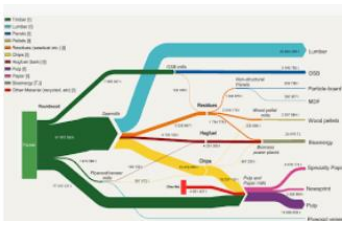


boxplot that only shows sample points, the violin plot encodes the entire distribution.

Bar chart: In many ways, the stacked bar chart is much more readable than the streamgraph.
It presents the same information. There's a reason why bar charts, line charts, and pie charts are the standard chart types found in your spreadsheet.

Sankey diagram
Sankey diagrams consist of two types of objects: nodes and edges. In this case, the nodes are the web pages or events, and the edges are the traffic between them. This differs from the hierarchical data you worked with before, because nodes can have many overlapping connections (figure 5.16) to show event flow or user flow from one part of your website to another.

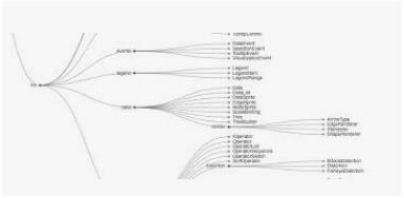
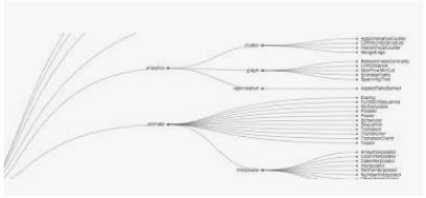
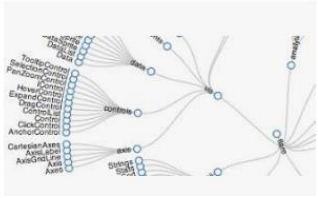


With a Sankey diagram like this at your disposal, you can track the flow of goods, visitors, or anything else through your organization, website, or other system

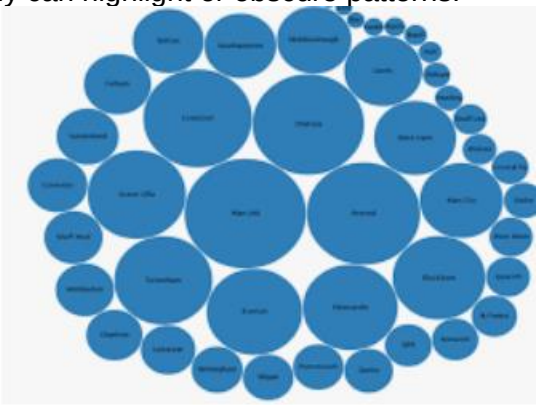
Hierarchical data
This is any data that maps the parent to child relationships, exists in every system: people in family trees, business org charts, and even categories like the food pyramid.

If you only show the numerical patterns, then the best charts you can use are the ones that encode using length, which means bar charts and line charts.

But if you only make decisions based on what you can measure with bar charts and line charts, then you hamstring your organization. A classic example of the value of hierarchical data visualization comes from data dashboards with their innumerable filters.

Circle Pack
This kind of quick overview gives us a sense of the hierarchical pattern in the wins and losses. It also highlights one of the issues we need to address when deploying hierarchical data visualization, the order of the hierarchy can highlight or obscure patterns.



When to use circle packing
Circle packs don’t use space efficiently—

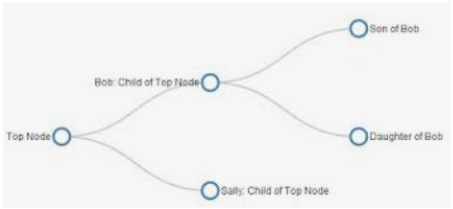
There's also a big difference between the way you see a circle when it's being used to enclose other circles versus when it's floating there on its own.

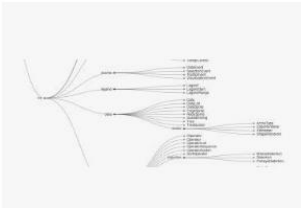
With that in mind, you should use circle packing when you're trying to focus on the thing at the bottom of the circle pack, the leaf nodes, and how they're sorted by the various categories you've nested them by, which will be all the other circles you see.
Circle Packing actually reveals hierarchal structure better than a Treemap.

Because circles are so bad at encoding numerical value with their radius.

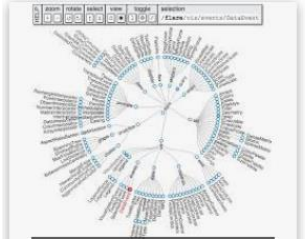
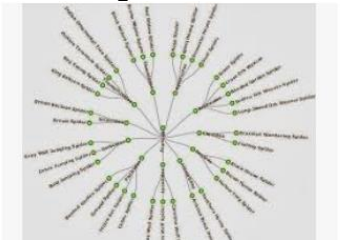
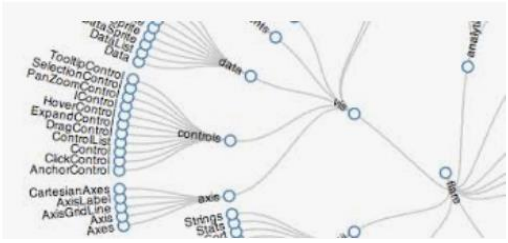
Trees
Another way to show hierarchical data is to lay it out like a family tree, with the parent nodes connected to the child nodes in a dendrogram

Treemapping is a data visualization technique that is used to display hierarchical data using nested rectangles; the treemap chart is created based on this technique of data visualization.

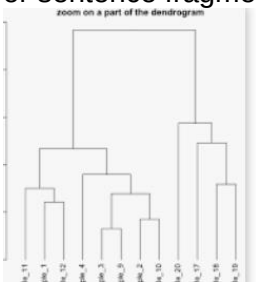


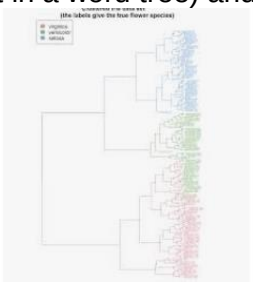
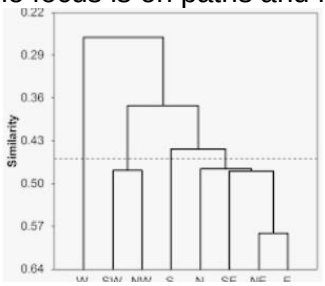


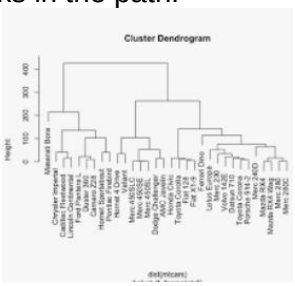
Radial tree diagrams
We have other choices besides drawing our tree from top to bottom and left to right. If we tie the position of each node to an angle, we can draw our tree diagrams in a radial pattern.

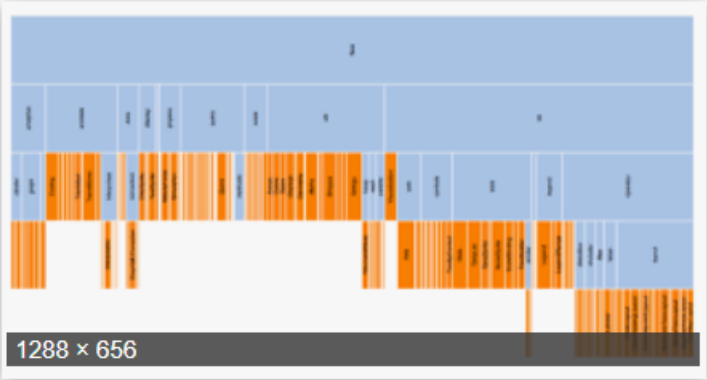
When to use dendrograms
As a general rule, you should use d3.tree
Dendrograms should be used when each parent and child is of the same type (like a word or sentence fragment in a word tree) and the focus is on paths and forks in the path.



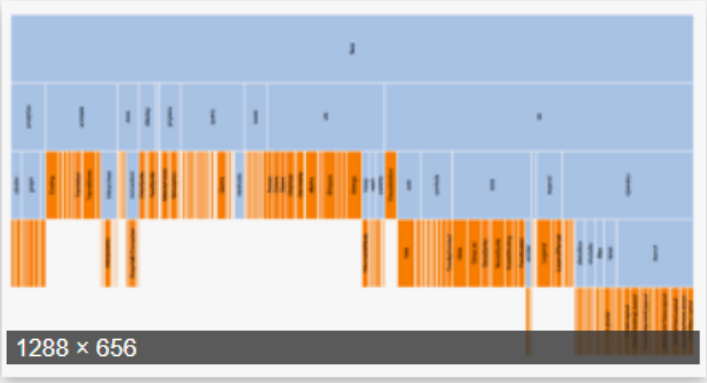


Partition



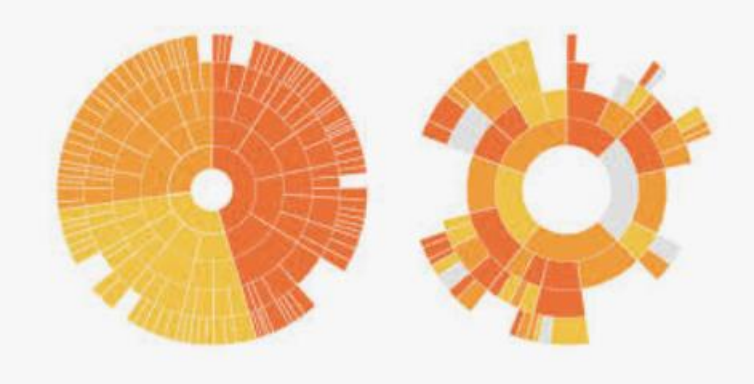
Partition Layer Chart Icicle Diagram ...

Icicle

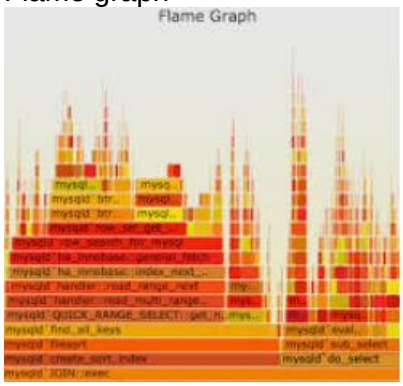


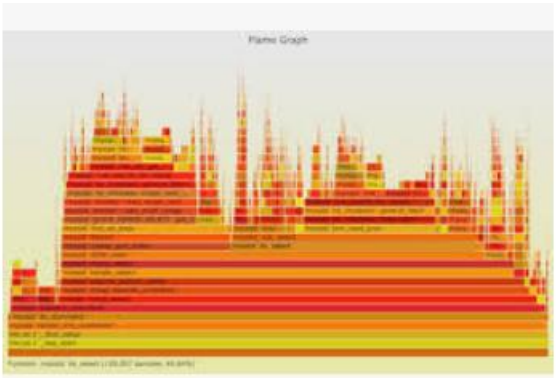
Partition Layer Chart Icicle Diagram ...

Sunburst



Flame graph

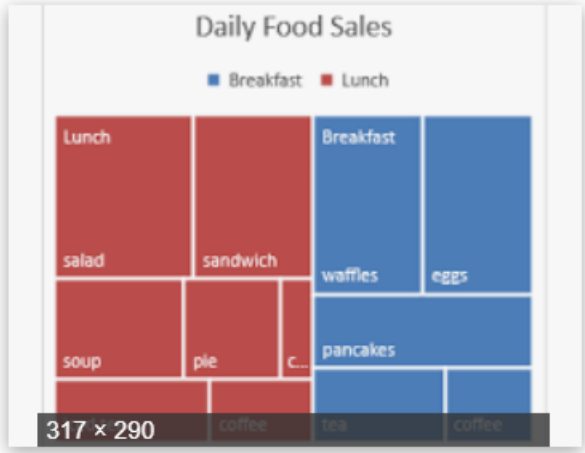




When to use the partition layout
In comparison to the dendrogram and the circle pack, the partition layout has great data-to-ink ratio. Literally no space is wasted on links, and the value of each node is encoded in the length of the node, facilitating good ability on the part of the reader to evaluate the numerical difference between the nodes. It's great for use in applications where you need to give your readers the ability to quickly and effectively measure the values encoded in the nodes.

The best use case for partition layouts is the kind of data that drives the flame graph, where the accretion of time or processing power is exactly what you want to emphasize to software developers looking to optimize their code.

Treemaps
The last hierarchical data visualization method we'll look at is the treemap, which was developed to show stock performance while at the same time showing parts of the market into which those stocks were categorized. Because of this serious business oriented pedigree, the treemap is a well-received way of showing hierarchical data. The treemap is a mix of circle packing and partition, using rectangles to represent nodes and enclosing those rectangles within their parent rectangles. Unlike circle packing, it has the benefit of using rectilinear shapes on a rectilinear screen, so we don't see a lot of wasted space like we do with circle packing.



Create a treemap chart in Office ...

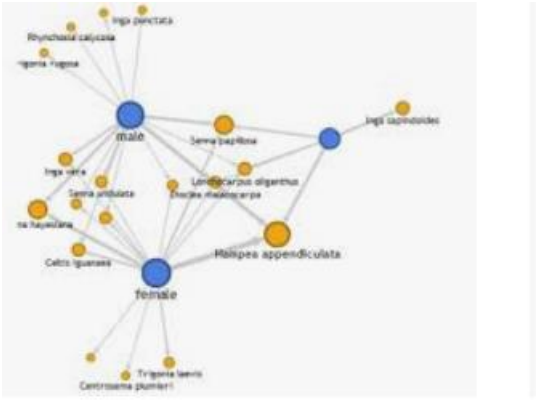
Treemap – The R Graph Gallery

When to use treemaps
Unlike the length of rectangles, people have a hard time evaluating the area of rectangles and understanding the value mapped to that area, so treemaps aren't going to be as effective as icicle charts for allowing precise comparison of values.

They're a good kind of chart for hierarchical data that's numerical and that you want to compare the rough value and aggregated value across categories.

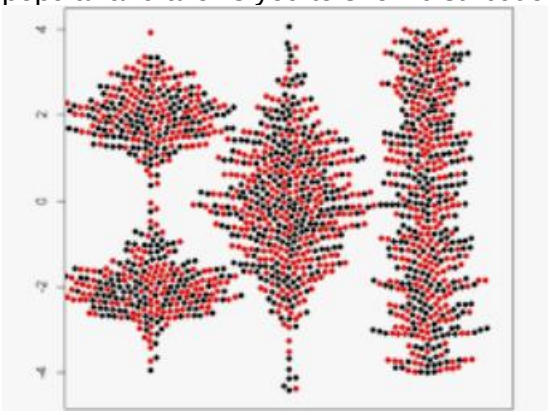
Another example is demographic data, where each leaf node represents items that vary numerically, like counties or census blocks, and for which you might want to see the breakdown by demographics aggregated by their hierarchical parents.

Network Visualisation
Network analysis and network visualization are more common now with the growth of online social networks like Twitter and Facebook, as well as social media and linked data, all of which are commonly represented with network structures. They represent systems more accurately than the traditional flat data seen in more common data visualizations.



Static network diagrams
Network data is different from hierarchical data. Networks present the possibility of any-to-many connections, like the Sankey layout seen previously, whereas in hierarchical data a node can have many children but only one parent, like the tree and pack layouts. A network doesn't have to be a social network. This format can represent many different structures, such as transportation networks and linked open data. We will look at four common forms for representing networks: as data, as adjacency matrices, as arc diagrams, and using force-directed network diagrams.

Beeswarm plot
The result of your simulation this time tries to arrange each node in a way that they're laid out along a shared x-axis but positioned to show their value. This beeswarm plot, is pretty popular and allows you to show distributions while maintaining individual sample points.



Geospatial Information Visualised
One of the most common categories of data you'll encounter is geospatial data. This can come in the form of administrative regions like states or counties, points that represent cities or the location of a person when sending a tweet, or satellite imagery of the surface of the earth.

Geospatial Mapping with D3 | Pluralsight

Q.Appraise the suitability of the visualisation technique for the web

Ans: The appraisals are structured around the intended audience, the intended message, the knowledge required to understand visualisations, any unintentional message that may be associated with visual types, any missing information that may be needed to understand the visualisations, and some potentially useful software to reproduce these visuals.

Results of appraisal Individual's perception and comprehension of visualisations can be affected by various factors such as cognitive functions, emotions and feelings, as well as gender.

Perceptual tasks (what is asked of the individuals to perceive) may also affect the accuracy, the consistency and the perception of visual representation of the benefit-risk balance.

Understanding the limitations of long-term memory and short-term memory may help in designing more effective visualisations.

Perceived risks are also affected by individual's dread or fear of certain risk event; and therefore may serve as an "early-warning" system during the decision-making process.

Visualisations may be used throughout the whole benefit-risk process to enhance the transparency and the clarity of tasks and decisions to be made.

The naïve stages in benefit-risk process involve planning, evidence gathering and data preparation, analysis, exploration, decision and dissemination.

A different set of visualisations may be more helpful at different stages of the process, but obviously these visualisations are propagated and revisited at later stages. Whilst dynamic and interactive visualisations may appear superior to their static counterparts, the same design principles still apply.

There may be a need for multiple visualisations to communicate a message successfully, and the dashboard design principles may be adopted to achieve an efficient layout for communicating benefits and risks.

Visual representations of benefits and risks are not only limited to graphics, but also include the use of emotive language (verbal labels) and numerical presentations. Whilst the use of verbal labels may trigger emotive response, they should not be presented without numerical representations. Numerical literacy plays huge role in visual perception of benefits and risks, and is associated with better comprehension.

There is no one visualisation type that triumphs in all situations, therefore alternatives should be offered to account for the variation in stakeholders' characteristics and preferences

Q. The basic functions of most interactive visualization tools have changed little since 1996, when Ben Shneiderman of the University of Maryland first proposed a "Visual Information Seeking Mantra": overview first, zoom and filter, then details on demand. Would you agree or disagree with this mantra? Articulate your rationale and provide examples where appropriate.

Ans- from notes

This design pattern is found in most interactive visualizations today. The combination of functions is successful, because it makes the data accessible to different

audiences, from those who are merely browsing or exploring the dataset to those who approach the visualization with a specific question in search of an answer. An interactive visualization that offers an overview of the data alongside tools for "drilling down" into the details may successfully fulfill many roles at once

Everything points to the conclusion that the phrase 'the language of art' is more than a loose metaphor, that even to describe the visible world in images we need a developed system of schemata.

The success of direct-manipulation interfaces is indicative of the power of using computers in a more visual or graphic manner. A picture is often cited to be worth a thousand words and, for some (but not all) tasks, it is clear that a visual presentation-such as a map or photograph-is dramatically easier to use than is a textual description or a spoken report. As computer speed and display resolution increase, information visualization and graphical interfaces are likely to have an expanding role. If a map of the United States is displayed, then it should be possible to point rapidly at one of 1000 cities to get tourist information. Of course, a foreigner who knows a city's name (for example, New Orleans), but not its location, may do better with a scrolling alphabetical list.

Users can scan, recognize, and recall images rapidly, and can detect changes in size, color, shape, movement, or texture.

There are many visual design guidelines but the basic principle might be: summarized as the Visual Information Seeking Mantra: Overview first, zoom and filter, then details-on-demand

Overview: Gain an overview of the entire collection. Overview strategies include zoomed out views of each data type to see the entire collection plus an adjoining detail view.

Another popular approach is the fisheye strategy (Furnas, 1986) which haa been applied most commonly for network browsing . The fisheye distortion magnifies one or more areas of the display, but zoom factors in prototypes are limited to about 5. Although query language facilities made it difficult to gain an overview of a collection, information visualization interfaces support some overview strategy, or should. Adequate overview strategies are a useful criteria to look for. Along with an overview plus detail (also called context plus focus) view there is a need for navigation tools to pan or scroll through the collection.

Zoom: Zoom in on items of interest. Users typically have an interest in some portion of a collection, and they need tools to enable them to control the zoom focus and the zoom factor. Smooth zooming helps users preserve their sense of position and context. Zooming could be on one dimension at a time by moving the zoombar controls or by adjusting the size of the field-of -view box. A very satisfying way to zoom in is by pointing to a location and issuing a zooming command, usually by clicking on a mouse button for as long as the user wishes. Zooming in one dimension has proven useful in starfield displays.

Filter: filter out uninteresting items. Dynamic queries applied to the items in the collection is one of the key ideas in information visualization. By allowing users to control the contents of the display, users can quickly focus on their interests by eliminating unwanted items. Sliders, buttons, or other control widgets coupled to rapid display