

SCHOOL OF Management Studies

**ASSIGNMENT COVER SHEET**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STUDENT DETAILS** | | | | | | | | | | | | | | | | | |
| Student name: | | | | Sanket Shinde | | | | | | Student ID number: | | | | 20097530 | | | |
|  | | | | | | | | | | | | | | | | | |
| **UNIT AND TUTORIAL DETAILS** | | | | | | | | | | | | | | | | | |
| Unit name: | | | Visualisation | | | | | | | | | Unit number: | | | COMP7016 | | |
| Tutorial group: | | | | | Q3 2022 | | | | | | Tutorial day and time: | | | | | Tuesday 19:00 to 21:00 | |
| Lecturer or Tutor name: | | | | | | | Ms Zhonglin Qu | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | |
| **ASSIGNMENT DETAILS** | | | | | | | | | | | | | | | | | |
| Title: | | Assignment 1 Report – Hierarchical Data Visualisation. | | | | | | | | | | | | | | | |
| Length: | | 1560words | | | | | Due date: | 03/08/2022 | | | | | Date submitted: | | | | 03/08/2022 |
| Home campus (where you are enrolled): | | | | | | | | | Parramatta City Campus | | | | | | | | |
|  | | | | | | | | |  | | | | | | | | |
| **DECLARATION** | | | | | | | | | | | | | | | | | |
| \* | I hold a copy of this assignment if the original is lost or damaged. | | | | | | | | | | | | | | | | |
| \* | I hereby certify that no part of this assignment or product has been copied from any other student’s work or from any other source except where due acknowledgement is made in the assignment. | | | | | | | | | | | | | | | | |
| \* | I hereby certify that no part of this assignment or product has been submitted by me in another (previous or current) assessment, except where appropriately referenced, and with prior permission from the Lecturer / Tutor / Unit Coordinator for this unit. | | | | | | | | | | | | | | | | |
| \* | No part of the assignment/product has been written/produced for me by any other person except where collaboration has been authorised by the Lecturer / Tutor /Unit Coordinator concerned. | | | | | | | | | | | | | | | | |
| \* | I am aware that this work will be reproduced and submitted to plagiarism detection software programs for the purpose of detecting possible plagiarism **(which may retain a copy on its database for future plagiarism checking).** | | | | | | | | | | | | | | | | |
| **Student’s signature:** | | | | | | Sanket Shinde | | | | | | | | | | | |
| **Note:** An examiner or lecturer/tutor has the right to not mark this assignment if the above declaration has not been signed. | | | | | | | | | | | | | | | | | |

ARO 00380 08/15

ASSIGNMENT 1  
REPORT  
Hierarchical Data Visualisation

Introduction:

We live in a world where every second billion of data is created, shared and interpreted across the globe. Such data is easy when represented in graphical form, as the world-famous quote by Fred Barnard says, “one picture is worth than thousand words”. Humans tend to consume the knowledge effectively when represented in graphical form instead of pages and a long list of data which require dedicated time to sit and understand the same. We come across various data the moment we start our day and do our daily chores. Eventually, it might come in different forms and sizes. So it is essential to have techniques and methods to design the visualisation of such data. These techniques help in formatting the data which will further create the ease to interpret and implement in knowledge form.

Heitzman (2019) defined Data visualisation as the process of accumulating knowledge or stats which can further be transformed into graphs and maps, that provide good perception and a better understanding of the data. It is frequently associated with any graphic element that displays information in any industry or subject (Kelleher & Wagener 2011). Companies with huge and complicated data depend on data visualisation for its representation in the most acknowledgeable way (Qin et al. 2020, Bayoumi et al. 2018 and Moer & Purchase 2011). Bayoumi (2018) and Loos (2018) highlighted that this help in human cognition and enable good decision-making.

A few techniques that help in achieving successful visualisation are to implement the chart or graph in its straightforward form which justifies by providing necessary information. Always keep in mind the design or style of information is built with the intention of visualisation. It should justify the objective of visualisation. Also, the colour used in visualisations must validate the information which needs to be addressed (Kelleher & Wagener 2011).

Moreover, the most of data are in a hierarchical form such as an organisational chart of corporations, Classification of animals or plants, etc. Such data have large elements and relationships of elements. For this, graphs are used to explain their relationship with each other. Graphs consist of nodes that are connected through edges.

The visualisation of hierarchical data can be categorized into two types which are the enclosure approach and the connection approach. The enclosures approach uses area and area segregation for data that is hierarchically arranged, whereas the connection method concentrates on the connections between the nodes. The enclosure approach is efficient for large hierarchical data sets with assigned properties (Liang 2015). Researchers such as Bennett(2007), Eades(2010), Hu, Purchase and Ware proposed the standard for aesthetics which enhances readability and so perception. Some of the aesthetics are minimising overlapping, edge length and edge bends. Bundling of associated nodes for large data sets, maintaining the symmetry of the graph and even spacing of vertices are also considered standards to follow for creating visualisations of hierarchical data.

This report will discuss the technical feature of one enclosure-based approach featuring the Divide and Conquer Treemap and one connection-based approach which is DOI Treemap. Further will analyse the visualisation generated using dummy data sets for both mentioned approaches.

Divide and Conquer Treemap:

Treemap utilises the most of available space to map hierarchical anatomy in the rectangular expanse. The development of the Treemap was done by Johnson and Shneiderman in the year 1991 especially to confront the problem of fewer visualisation methods for the File system. The size of the rectangular area that the Treemap algorithm recursively encloses tree nodes depends on the number of child nodes to be partitioned or the weight assigned to a particular node attribute, such as file size for the file system. Enhancing this approach Liang, one of the Data visualisation researchers, proposed Divide and Conquer Treemap in 2013 to confront the demand for other shapes in treemaps. D&C Treemap expanded on the treemap idea by adding confinement, which encloses child nodes in the parent node and encodes value-using region in a distinctively formed container. It allows for freedom in the geometries of the container and containment.

The first step in constructing D&C Treemap is calculating the weight of every node which is related to its attribute. Then, by repetitively dividing the area into subareas depending upon its hierarchical branch. Also, subarea size will again relate to the node’s attribute. The position of each node is determined within its immediate vicinity. The geographic centre of the area is often where the vertex is located. Finally adding graphical attributes enhances the presentation of visualisation (Liang et al. 2015).

As the name suggests, the D&C Treemap algorithm consists of three sub-algorithms namely initialize, divide and conquer. The first side vertex for the starting point is estimated by the property of the polygon. The largest angle among the vertex decides the angle of the convex polygon whereas, the negative value of the concave angle is used for the concave polygon. Divide algorithm sort all nodes after dividing them into two sets, arrange that in ascending order as per their weights and assign them to the first set of data till the approximate half weight is achieved. Conquer algorithm constructs polygon with polygon proportionate to their weights. This procedure repeats till the last node in the hierarchy.

The D&C Treemap allows illustrating hierarchical data in any shape for container area and containment child. This help in the comparison of value between nodes. Also, it enhances visual cues making it a great tool for the visualisation of hierarchical data.

Diagram, schematic

Description automatically generated

Figure: D&C Treemap with 3000+ nodes for java file

The above D&C Treemap shows a visualisation of a Java file system which more than 3000 nodes. The size allocation is as per the procedure mentioned earlier. Big areas represent nodes with Large weight i.e file size. Though some areas are hallowed, most areas have lots of small areas within them explaining the child nodes with containment of parent nodes.

Degree of Interest(DOI) Treemap:

This connection-based approach was proposed by two Xerox PARC employees Dave and Nation in 2002. It is a sample visualisation that displays or illustrates focus and context (Budui et. Al 2006). As the name suggests, it calculates the degree of interest of the user and calculates the space around to display relevant data rather than complete data. The tool allows the interaction, expanding or shrinking nodes as the algorithm optimises screen or space usage (Nguyen, Simoff & Huang 2014).

DOI Treemap mainly uses focus and context techniques. It helps in choosing which portion to show at any instant (Dave and Nation, 2002). Also, DOI Treemap overcome the problem of display bounds by combining all techniques of focus and context along with some new methods ensuring trees stay within the bounded limit. Expanded computations of users' degree of interest estimates are combined with logical filtering to omit nodes with a low degree of interest, geometric scaling of node size following degree of interest to be able to hold various levels of information, semantic scaling of the contents of the nodes with node size, clustered representation of large unexpanded branches of the tree, and animated transitions, all of which are intended to speed up the user's understanding of changes in the tree (Dave and Nation, 2002).

There is a finite amount of display space. For the tree to remain inside its designated resource, it must always be limited. This necessitates techniques for keeping track of and modifying the tree visualisation. The tree visualisation must frequently be compressed to keep within its bounding box. However, if the area is underused, the tree might be profitably enlarged to make use of the extra room. The anticipated Degree of interest for each node by the users governs both compression and expansion.

DOI Trees forecast the user's dynamically changing interest in the content and the context of the tree's nodes to adjust the display. The most intriguing nodes are selected for visualisation, geometrically enlarged, semantically enlarged, and shown concerning other nodes that have been carefully selected to be pertinent. Nodes that are likely to be outside of the user's region of interest are aggregated, compressed, or eliminated. Access points to pertinent data on the internet or in applications like email are provided by the nodes in the display. DOI trees push the display to fit inside a constrained space even though they chose additional nodes display to fill that display.

Diagram, schematic

Description automatically generated

Figure: Shopping Categories in Australia using DOI Treemap

The above categorical classification of Shopping stores in Australia is the perfect example of a good example of DOI Treemap. At first main types of stores are displayed. Selecting the type first will then only open the next varieties sub-types. In the above DOI Treemap, clicking on Antiques open up different types of antique stores showing subtypes. This is extended 6 levels which are easily adjusted on screen helping users to gather information at ease. Also, the cues in front of types and sub-types provide an idea of how more data is present with that type or subtype.

Conclusion:

There are many different techniques and tools for Visualisation, and almost all have their unique advantage over others. In this report, we have seen the details of D&C Treemap which is based on the Enclosure approach and DOI Treemap which is based on the Connection Approach. Hierarchical data when visualised in D&C Treemap provides better information on relationships and structures, while DOI Treemap provides better user-friendly controllability. Hence it is important to understand the nature of data and then choose the best technique which will justify and provide better insights from its visualisation.

References:

Bayoumi, S., Alghamdi, K., Alqusair, D. and Alfutamani, A., 2018, April. Visualization of Fire Accidents in Saudi Arabia. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE.

Bennett, C, Ryall, J, Spalteholz, L & Gooch, A 2007, ‘The aesthetics of graph visualization’, *Computional Aesthetics in Graphics, Visualization, and Imaging,* pp. 57-64, viewed on 24 April 2022.

Budiu, R., Pirolli, P. and Fleetwood, M., 2006, May. Navigation in degree of interest trees. in *Proceedings of the working conference on Advanced visual interfaces* (pp. 457-462).

Eades, P, Gutwenger, C, Hong, SH % Mutzel, P 2010, ‘Graph drawing algorithms’, in *Algorithms and theory of computation handbook,* Chapman & Hall/CRC, Boca Ranton, Florida, pp 21-25.

Heitzman, A 2019, ‘Data visualization: what it is, why it’s important & How to use it for SEO’, SEJ, 29 January 2019, viewed 27 July 2022.

Johnson, B && Shneiderman, B 1991, ‘Tree-maps: a space filling approach to the visualization of hierarchical information structures’, *Visualization, 1991. Visualization ’91, Proceedings., IEEE Conference on*, IEEE, pp 284-297.

Kelleher, C and Wagener, T 2011. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software,* 26(6), pp.822-827.

Liang, J, Nguyen, QV, Simoff, S & Huang, ML 2015, ‘Divide and conquer treemaps: visualizing large trees with various shapes’, *Journal of Visual Languages and Computing*, vol. 31, pp 104-127.

Moere, A.V. and Purchase, H., 2011. On the role of design in information visualization*. Information Visualization*, 10(4), pp.356-371.

Nguyen, Q.V. and Huang, M.L., 2005. EncCon: an approach to constructing interactive visualization of large hierarchical data. *Information Visualization*, 4(1), pp.1-21.

Nguyen, Q.V., Simoff, S. and Huang, M.L., 2014, July. Using visual cues on doitree for visualizing large hierarchical data. In *2014 18th International Conference on Information Visualisation* (pp. 1-6). IEEE.

Shneiderman, B., 1992. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG),* 11(1), pp.92-99.