Visualisation of the desktop files with Degree-of-Interest Tree, Treemap 4.0 and Divide & Conquer Treemap.

Introduction and scope

Information visualisation promotes understanding and supports humans to analyse large, complex data sets (Bennett et al. 2007; Moere & Purchase 2011). Initially, academic research focused on performing advanced data exploration, pursuing unbiased data replication and comprehension (Moere & Purchase 2011). However, the focus shifted from highly skilled users in the area of research to encompass larger user audience, non-expert users (NEU) (Moere & Purchase 2011; Quispel, Maes & Schilperoord 2016). Discourse surrounding information visualisation is moving towards user-centric challenges as researchers agree that creation of aesthetically pleasing graphs has a practical aim of revealing hidden patterns, evoke memorization and optimizes readability of the visualisations regardless of viewer's prior knowledge (Bennett et al. 2007).

The seven aesthetics parameters defined by Purchase (Purchase 2002) have become one of the imperative attempts to define the aesthetics of a graph and have been backed up by the study that found correlation between aesthetics and task abandonment, erroneous response time and motivation to engage with tasks (Cawthon & Moere 2007; Liang et al. 2015). However, the seven parameters as in most cases it is impossible to satisfy all the criterions nor do they carry same weights, the priority changes based on the research purpose and dataset itself – semantic issues (Bennett et al. 2007).

This paper attempts to construct the non-expert users' (NEU) experience from both graph construction (creator) and exploration (viewer). It will firstly, draw out the strengths and drawbacks of the tools used for data visualisation. The dataset for this paper was limited to the author's desktop file system (213 nodes) backed up by the rationale that using a common and easily accessible data allows exploration of the potential of graph visualisation tools in hands of NEU. In addition, by using relatively small dataset we can focus on application and exploration versus how tools perform with large datasets.

Degree-of-Interest Tree (DOITree)

Techniques used to visualize hierarchical structure are classified in two categories: connection and enclosure. Degree-of-Interest Tree (DOITree) is an example of connection approach, which focuses on the relationship structure between nodes (Liang et al. 2015). As this paper focuses on exploring the nodes, DOITree is used as an introduction as it would be the easiest to understand from NEU's perspective. DOITree is an improved version of Space Tree where expansion view can be applied to multiple branches at a same time but there is still an opportunity to improve it to gain better insight during navigation (Nguyen, Simoff & Huang 2014).

Treemap 4.0

On the other side of connection approach is enclosure – a method of visualising large data sets with attributed properties while ensuring that all nodes and their sub-hierarchies are located within the parent node, ultimately in super node (Liang et al. 2015). This technique

allows to display the data within a compact area and Treemap 4.0 is an example of enclosure approach visualisation, where it divides the area into a nested sequence of rectangles and the core lies in the algorithm used to create those rectangles (Baehrecke et al. 2004; Bederson, Shneiderman & Wattenberg 2005).

The improved algorithm in Treemap 4.0 not only focuses on creating rectangles with lowest ratio but improves on creating layouts that preserve the order of the dataset. Another significant modification in Treemap 4.0 is availability of Strip treemaps, where it orders rectangles in strips of varied thickness, allowing increase in readability and stability: it takes approximately 3 and 4 times less to complete a task with Ordered and Strip treemaps respectively than original Squarified treemap (Bederson, Shneiderman & Wattenberg 2005).

One of the main advantage of Treemap 4.0 would be the usability: a control panel that allows to easily modify the look of the graph and a filtering feature that allows to find and highlight data of interest based on quantitative measurements (Baehrecke et al. 2004).

Divide & Conquer Treemap

Alternative to Treemap 4.0 is Divide & Conquer Treemap (D&CTreemap) where it moves away from rectangles to include other geometric shapes to unlock the user's capability to recognize orientation, shape and differentiate between patterns, thus increasing user experience and awareness during data exploration (Liang et al. 2015). The shortcoming of the D&CTreemap is the need to understand the code to make changes, thus limiting the full potential of the tool to experienced users.

Desktop files visualisation

Desktop files became the choice for visualisation as:

1. It is easily accessible for anyone in possession of personal computer, allowing a large population to explore the visualisation tools and make discoveries based on the dataset.
2. DOITree and Treemap 4.0 had built-in features that would turn directory into visualisation, while D&C Treemap was provided with additional tool that would turn directory into correct xml format. This is important as all three tools read different file formats and would require understanding of the code to create a dataset.
3. Similarly to the study on personal websites, desktop files contain person's digital print and contrary to carefully constructed identity claim on the website, desktop can unveil intrinsic behavior with exploration of hidden patterns (Vazire & Gosling 2004; Youyou, Kosinski & Stillwell 2015). As desktop files are personal and intimate, this paper observes the author's files to give commentary to the patterns.

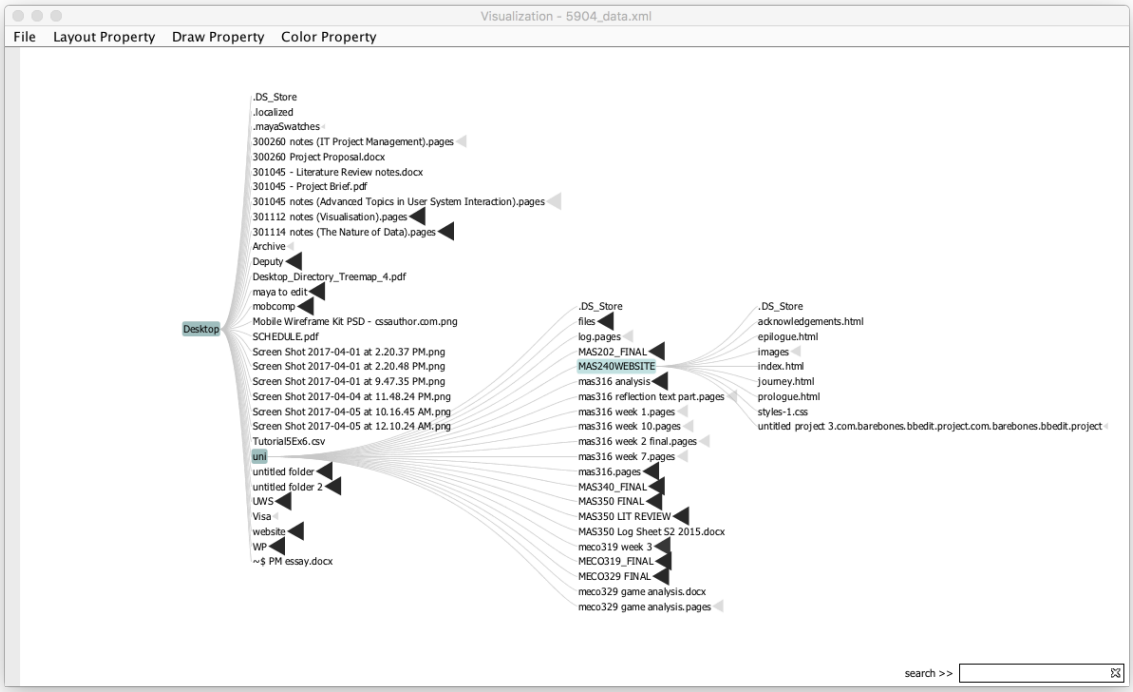

Figure 1 Desktop visualisation via DOITree

From Fig. 1, where the hierarchy of Desktop file system is presented with no significant modifications, we already can make some assumptions about the owner of the desktop:

1. Although owner has a relatively organised desktop, there are files that have not yet been moved to the correct folder: e.g. unit papers should go into UWS folder, there are screenshots that go right after super node with no sub-hierarchy. In 'uni' folder (from owner's undergraduate studies), although the files are mostly in their folders, there should have been organised in respective unit folders, if the owner was a conscientious individual. The author indeed strives for order but is not naturally organised individual and the desktop can easily betray the identity claims the author may make.
2. The files directly located on desktop are the files author currently works on and are within immediate reach.

These first assumptions already can assist strangers to get insight on the personality, the treemaps can not only assist in defining the owner but also the purpose of the folders and give background on the rationale behind them.

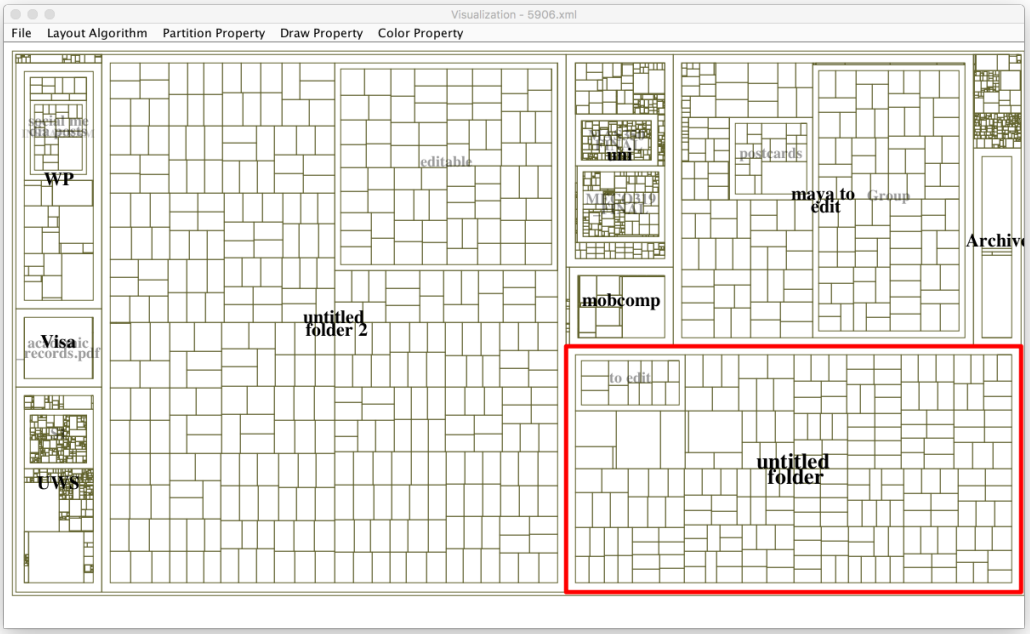

Figure 2 D&CTreemap visualisation with no edits

As D&CTreemap calculates the weight of each file, the size of the rectangle/triangle reflects the actual file size. From the unmodified version (Fig. 2), we can already conclude that 'untitled folder', 'untitled folder 2' and 'maya to edit' folders are the biggest in size. To further investigate those folders, Fig. 3 accentuates those folders by utilising 30 degree angular layout and colours in the file types. We can now see that those folders consist of images. Viewers may conclude that the individual is a visual person: they might like to take pictures, work with imagery, draw images. Indeed, the three folders in question contain the photography projects the individual works on for social media.

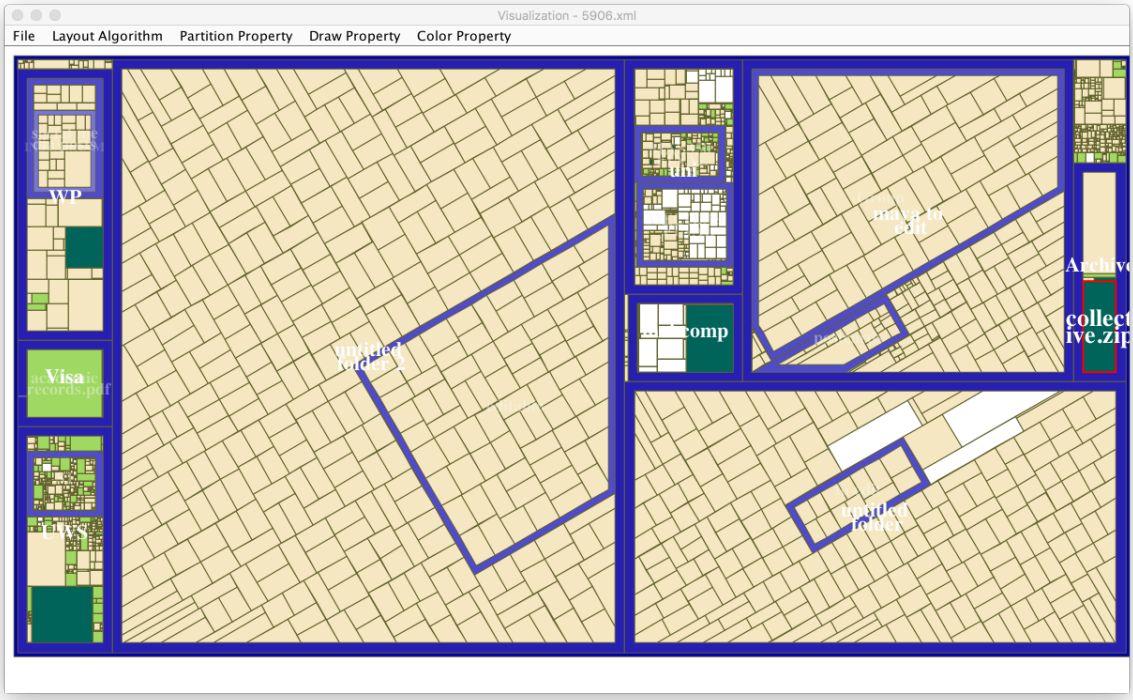

Figure 3 D&CTreemap with angled folders with images (modified)

The Figure 3 can be reproduced by Treemap 4.0 as well: Figure 4 illustrates same parameters: where the size of rectangles are created by the file size and green colour represents images.

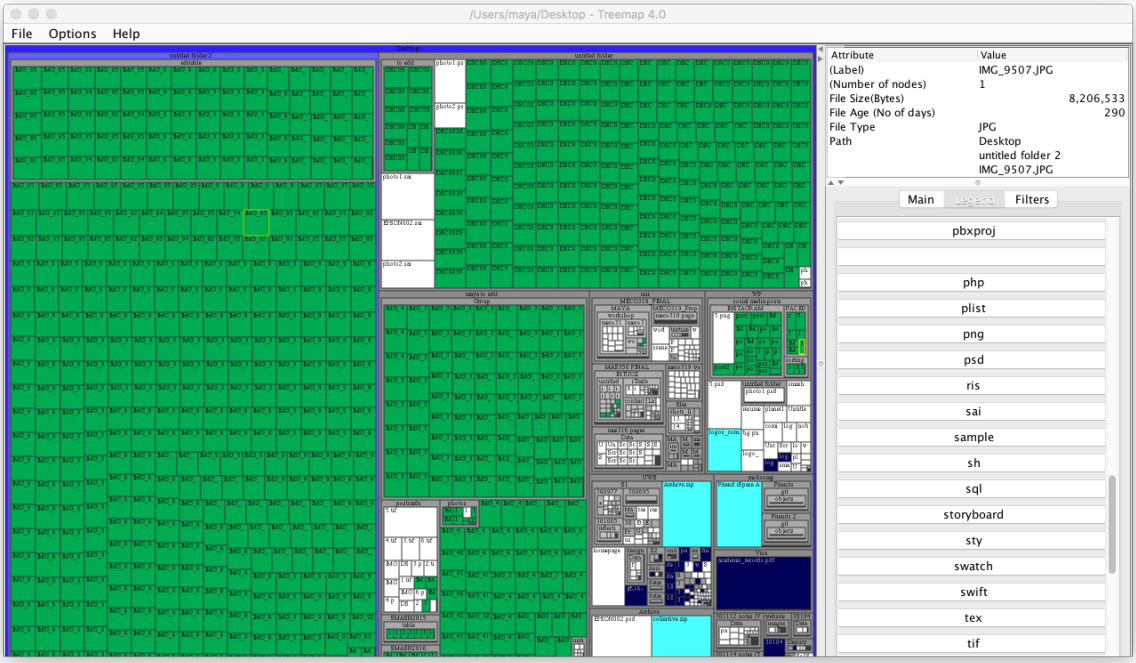

Figure 4 Treemap 4.0 visualisation, reproducing Fig. 3

Alternatively to D&CTreemap, the size of the rectangles in Treemap 4.0 can represent a variety of things. In Figure 5, the size of rectangles is dependent on file age. Here, we can observe the folder 'uni' is oldest, followed by 'UWS' folder. This can be explained as 'uni' folder was created in 2011, when the author started undergraduate degree, while 'UWS' folder was created in 2016, the year author started the current masters degree.

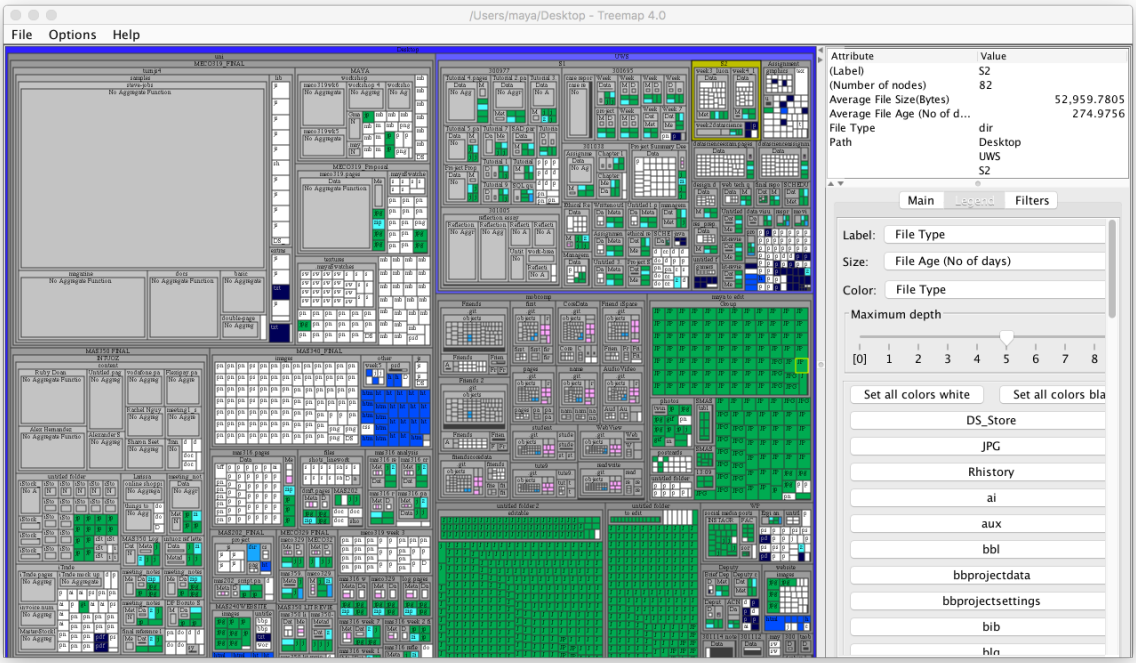

Figure 5 Treemap 4.0 with file age affecting rectangles size

Finally, the unique feature for Treemap 4.0 is filtering capability: Figure 6 filters out all the JPG files (hiding all the other nodes). The filtering can be done on file type, file size and file age, allowing NEU to explore and adjust the visualisation to match their purpose of the data exploration.

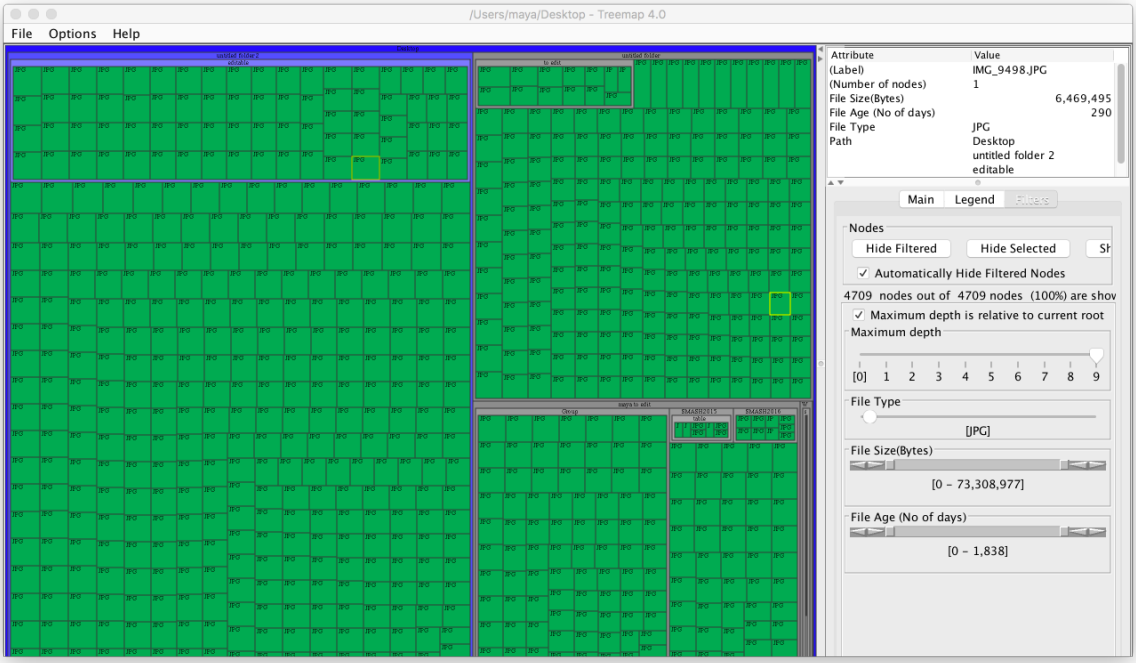

Figure 6 Treemap 4.0 filtering out all files except JPG files

Discussion

The variety of visualisations allows discovery of patterns and interesting properties. This paper follows the lead of design challenges that shifted to user-centric approaches and allows some perspective on NEU application of visualisation tools and limitation of usability of such tools (Gough et al. 2014):

1. For users with no coding background, the data sets would be limited as they would need to convert the data they possess into program-readable format, challenging for no coding background users.
2. The absence of standardized file format that would be applicable to all tools further challenge the accuracy of the visualisation creation for NEU.
3. These challenges were overcome with provided tools that allowed file creation based on the path names of file systems, however this would not be a solution if dataset was a different hierarchical data.

Usability of tools, demand on prior knowledge and education problem are the challenges that information visualisation still faces at this stage of development (Chen 2005).

Conclusion

Deploying guided visualisations and educating general population to explore and understand the process of graph visualisation will raise public awareness and allow NEU to critically assess the information they are being presented with. Ultimately, visualisations should save time on data exploration and not complicate the understanding of data (Gough et al. 2014) and while this paper didn't focus on big data sets, it still allowed to discover hidden properties and encourage NEU to collaborate in visualisation development.

References

Baehrecke, EH, Dang, N, Babaria, K & Shneiderman, B 2004, 'Visualization and analysis of microarray and gene ontology data with treemaps', *BMC bioinformatics*, vol. 5, no. 1, p. 84.

Bederson, BB, Shneiderman, B & Wattenberg, M 2005, *Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies (2001)*.

Bennett, C, Ryall, J, Spalteholz, L & Gooch, A 2007, 'The aesthetics of graph visualization', *Computational Aesthetics*, vol. 2007, pp. 57-64.

Cawthon, N & Moere, AV 2007, 'The effect of aesthetic on the usability of data visualization', in *Information Visualization, 2007. IV'07. 11th International Conference*, pp. 637-48.

Chen, C 2005, 'Top 10 unsolved information visualization problems', *IEEE computer graphics and applications*, vol. 25, no. 4, pp. 12-6.

Gough, P, Ho, X, Dunn, K & Bednarz, T 2014, 'Art and chartjunk: a guide for NEUVis', in *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*, p. 171.

Liang, J, Nguyen, QV, Simoff, S & Huang, ML 2015, 'Divide and Conquer treemaps: Visualizing large trees with various shapes', *Journal of Visual Languages & Computing*, vol. 31, pp. 104-27.

Moere, AV & Purchase, H 2011, 'On the role of design in information visualization', *Information Visualization*, vol. 10, no. 4, pp. 356-71.

Nguyen, QV, Simoff, S & Huang, ML 2014, 'Using visual cues on DOITree for visualizing large hierarchical data', in *Information Visualisation (IV), 2014 18th International Conference on*, pp. 1-6.

Purchase, HC 2002, 'Metrics for graph drawing aesthetics', *Journal of Visual Languages & Computing*, vol. 13, no. 5, pp. 501-16.

Quispel, A, Maes, A & Schilperoord, J 2016, 'Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use', *Information Visualization*, vol. 15, no. 3, pp. 238-52.

Vazire, S & Gosling, SD 2004, 'e-Perceptions: personality impressions based on personal websites', *Journal of personality and social psychology*, vol. 87, no. 1, p. 123.

Youyou, W, Kosinski, M & Stillwell, D 2015, 'Computer-based personality judgments are more accurate than those made by humans', *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. 1036-40.