# Treemaps and the Visual Comparison of Hierarchical Multi-Attribute Data

Kent Wittenburg
Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, 8th Floor
Cambridge, MA 02129 USA
wittenburg@merl.com

Tommaso Turchi
Brunel University London
Kingston Ln, Uxbridge, Middlesex
UB8 3PH, United Kingdom
tommaso.turchi@brunel.ac.uk

## ABSTRACT
Treemaps have the desirable property of presenting overviews along with details of data and thus are of interest in visualizations of multi-attribute tabular data with attribute hierarchies. However, the original treemap algorithms and most subsequent variations are hampered in making parallel structures in a hierarchical data structure visually comparable. Structurally parallel elements are not aligned, making it difficult to compare them visually. We propose a method that allows for proportional and non-proportional subdivisions of subtrees while preserving visual alignment of parallel structures. We extend the framework so that other types of data visualizations can be placed within the graphical areas of a treemap to allow for the visual comparison of a broad collection of data types including temporal data.

## CCS Concepts
• **Human-centered computing~Visualization techniques**

## Keywords
Treemaps; hierarchical multi-attribute data; multidimensional data; information visualization; visual comparison

## 1. INTRODUCTION
Visual comparison of hierarchical multi-attribute data is a task found in many application domains including business sales data [5], organizational performance evaluation [8], and patent landscaping, just to name a few. For example, among the tasks in business applications of patent landscaping are the following [9]:

1. Determine a company's points of IP vulnerability, i.e., in what jurisdictions, product components, or classes of technology is a company at risk?
2. Determine a company's points of IP strength, i.e., in what jurisdictions, product components, or classes of technology might a company be able to counterattack?

These tasks entail synthesizing comparisons along a number of different attributes present in patent data sets. Examples include filing trends, categorizations of patents by product component or technology, jurisdictions of patent filings, company business strength, litigation histories, and meaures of individual patent strength. Showing overviews that highlight points of comparative differences, whether positive or negative, enable users to drill down to attribute-specific details to develop business strategies.

Prior work has not provided a solution for structurally parallel visual comparison that can use proportional representations at intermediate levels of an attribute hierarchy. In this work we present proportional treemap layouts for multi-attribute hierarchical data. Our solution uses proportionality separately in the horizontal and vertical dimensions. We extend the framework to treat graphical areas of a treemap layout as canvases on which to draw non-proportional data visualizations that are appropriate for temporal data types, among others. We also note the limits of embedded proportional visualizations if visual comparison through length is to be retained across the entire tree.

## 2. RELATED WORK
Multi-attribute data visualization methods have long benefited from the application of methods from on-line analytical processing (OLAP) in the database world [1]. Dynamic attribute hierarchies can be used to hierarchically organize and then "slice and dice" tabular data. In principle, many more attributes than two or three can be incorporated into the visualizations. Two types of basic layouts have been proposed for such dynamic hierarchies for tabular data: table-based layouts and proportionally-based layouts.

Among the table-based approaches, Kehrer et al. utilized a small-multiples approach [5]. They give examples of all three types of comparative visualization noted by Gleicher et al. [3]: "juxtaposition (showing different objects separately), superposition (overlaying objects in the same space), and explicit encoding of relationships.…" The generalized algebra of Kehrer et al. offers a theoretical foundation for a tabular approach to comparative analytics. However, it does not sum or aggregate to provide comparative overviews at the higher levels.

Classic treemaps are a solution that reveals aggregate information at the higher levels of a hierarchical tree proportional to the sizes of its descendants, thus providing an overview with connection to detail [6]. The use of treemaps to visualize dynamic attribute hierarchies was introduced in [2], and a large volume of work has followed. An difficulty, however, is that most treemap layouts place parallel structural elements at irregular positions. Juxtaposition and/or alignment, basic to some forms of visual comparison, are not easily achieved, and visual searching is needed to locate structurally parallel cells for comparison [4][8].

Vliegen et al. [8] solve the juxtaposition/alignment problem by means of what they call a matrix layout. The effect of a matrix layout is to align the areas of comparable nodes with the use of empty space. For parallel structures in the tree to be comparable,


AVI '16, June 07-10, 2016, Bari, Italy
© 2016 ACM. ISBN 978-1-4503-4131-8/16/06…$15.00
DOI: http://dx.doi.org/10.1145/2909132.2909286

they must be normalized, i.e., have the same number of children in the same order, where some of those children are empty cells. The authors demonstrate how the matrix layout can simulate a barchart or be embedded in a myriad number of other shapes where the proportional information is at the leaf level of the hiearchical structure. The approach, while solving the visual comparison problem of treemaps at one level, apparently gives up a defining property of treemaps, namely, proportionality all the way up (and down).

Vliegen et al. [8] develop an important theoretical concept called the Uniform Density Property, which, to paraphrase, is the property that the ratio between measure and area remains constant throughout the tree. Measure represents the value associated with a node and area is the graphical area taken by a node. If this property holds, nodes are comparable throughout the tree through their relative area sizes. We adapt the concept to length (rather than necessarily overall area) in the work that follows.

In [6], proportional representations and linearly scaled representations are each independently used in embedded layouts for hierarchical multi-attribute data. Our goal for visual comparison is to combine these methods within a single visualization. The hope is to take advantage of treemaps in presenting data overviews along with the benefits of table-based approaches with their larger palette of visualizations for comparison purposes.

## 3. Hierarchical multi-attribute comparison maps

We now present treemap layout variants that embed linearly scaled 2D visualizations such as histograms. Our goal is to maximize visual comparison of hierarchical multi-attribute structures. We make use of (row-skipping) alignment of parallel structures and show how proportional and non-proportional layouts can be combined. We investigate the implications of the Uniform Density Property when proportional and non-proportional layouts are embedded within one another.

Figure 2 shows an abstract example of our approach. The overall effect is a rectilinear layout that is a by-product of making structurally parallel elements aligned, though not always adjacent. Unlike a table, the cells are of unequal size. Proportional layout is in evidence at the top level through height differences in the vertical dimension for the values of Dimension A. In the horizontal dimension, some attribute rows are proportionally presented (Dimensions C and D), others non-proportionally (Dimension B).

A high-level summary of our layout algorithm is in Figure 3. The algorithm traverses the attribute tree top-down, left-to-right. It can be utilized for space-filling approaches where the graphical area is fixed in both the horizontal and vertical dimensions. It can also be adapted through the use of constant-height rows for display areas where the width of the display area is fixed and the overall height of the display grows with the number of rows.

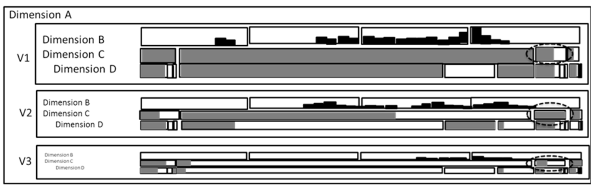


**Figure 2. Abstract example of a hierarchical multi-attribute comparison map.**

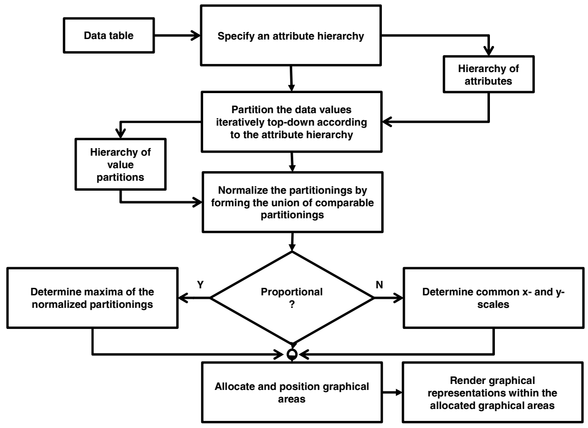


**Figure 3. Overview of algorithm.**

Table 1 shows the data that we will use in the next set of examples to illustrate the algorithm. It is taken from the domain of patent landscaping. We will use binned data for File_Date in our examples and will also at times truncate the IPC class codes. IPC_Class refers to a domain classification of patent assets in the International Patent Classification system.

Figure 4 shows an example attribute hierarchy over the data in Table 1. Reflecting its significance in many business tasks, the attribute Owner, normally a company, is picked as the top level partitioning attribute, followed by File_Date and Country (of filing). Country is further partitioned by the IPC_Class attribute.

A hierarchical data partitioning of the Table 1 data according to Figure 4 is shown in Figure 5. We have found it useful to think of the hierarchical structure as composed of attributes (shown as

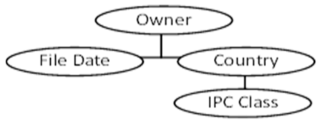


**Figure 4: Example attribute hierarchy over the data.**

**Table 1: Tabular data used in examples**

| Owner | Country | File_Date | IPC_Class |
|---|---|---|---|
| Company A | US | 6/18/2008 | H05H13 |
| Company A | EP | 1/30/1998 | A61N5 |
| Company A | EP | 1/30/1998 | A61N5 |
| Company A | EP | 1/30/1998 | A61N5 |
| Company A | JP | 8/28/1997 | A61N5 |
| Company A | JP | 10/4/2002 | A61N5 |
| Company A | JP | 1/27/2003 | A61N5 |
| Company A | JP | 4/14/2003 | A61N5 |
| Company A | JP | 5/13/2011 | A61N5 |
| Company B | JP | 4/2/1998 | G12B13 |
| Company B | JP | 4/2/1998 | G12B13 |
| Company B | JP | 5/28/1997 | A61N5 |
| Company B | JP | 11/12/1997 | A61N5 |
| Company B | JP | 2/29/2000 | A61N5 |
| Company B | JP | 4/30/2002 | A61N5 |

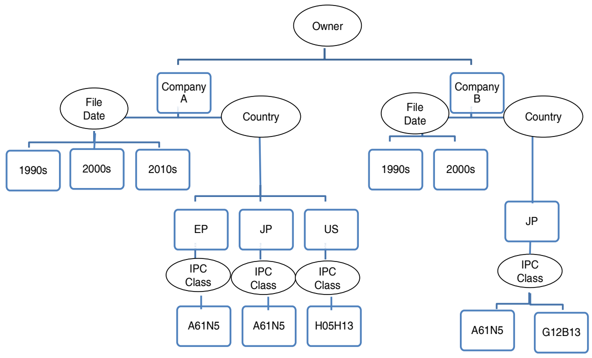


**Figure 5. Hierarchical data partitioning based on Table 1 data and Figure 4 attribute hierarchy.**

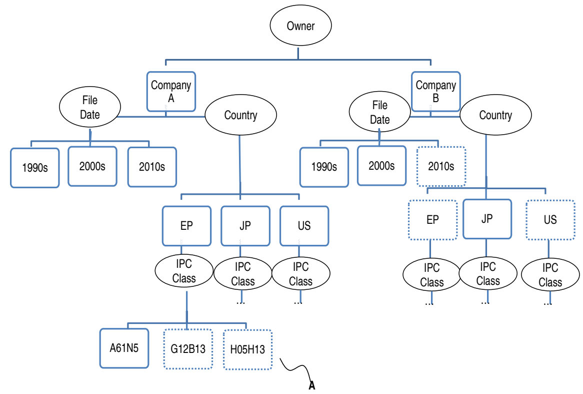


**Figure 6. Normalized tree. All parallel structures have the same partitioning.**

circles) and value partitions (shown as rectangles). Parallel, and thus comparable, areas of the tree are those that share a common attribute path such as Owner/*/Country/*/IPC_Class.

In Figure 6, we see a normalization of the hierarchical data structure of Figure 5, following [8]. The partitioning at the bottom left (A) should be understood to be repeated at all nodes labelled IPC_Class. All parallel value structures have the same set of partitions and the same ordering within the partitioning although many partitions in this example are empty.

As summarized in Figure 3, the layout algorithm walks the tree top-down and allocates graphical areas of the tree depending on the type of layout chosen for each level and the graphical dimension (horizontal or vertical) for the placement of sibling nodes. There is a choice point depending on whether the graphical areas are set to be proportional or non-proportional. If they are proportional, the allocation of graphical areas must be relative to a maximal measure for each of the parallel partitions in the normalized partitioning. For non-proportional levels, a common x- and y-scale must be found that incorporates all values found in all parallel structures. All parallel structures will thus utilize the same layout and/or scale for parallel value partitions.

Figure 7 shows an example where the top level (Owner) is set to be proportional in the vertical dimension; the two branches at level 2 (File Date and Country) are set to be non-proportional and proportional, respectively, in the horizontal dimension; and level 3 (IPC_Class) is rendered non-proportionally as a horizontal histogram. In this case, the Uniform Density Property holds for all proportionally allocated cells in the sense that area and length (in horizontal or vertical dimensions) is comparable. Even though one must skip rows to find comparable parallel attributes, it seems the overall result is a reasonable solution to the problem and points the way toward visualizations of more complex data.

Figure 8 shows other layouts related to our approach. In Figure 8(A), a simple embedded treemap is shown. It is hard to compare

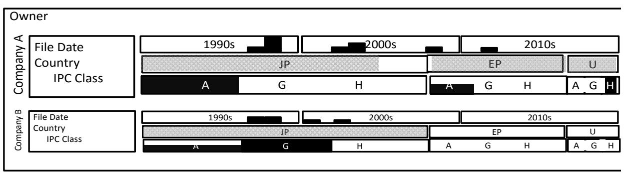


**Figure 7. Mixed proportional/non-proportional layout.**

internal value distributions with this layout. Figure 8(B) shows a vertical stacking at the top level. The top level proportions are easier to compare given the use of length rather than area, but it is still hard to compare internal structurally parallel values. Figure 8(C) shows the merits of normalizing proportional views. Here it is easy to compare internal distributions of Country values; comparison of top-level Owner values can still be conveyed by relative row height as shown in C or by additional rows as in B and in Figure 9. Figure 8(D) illustrates the limits of embedding normalized proportional layouts. Consider the horizontal length of the cells for the "H05H13" classification. In Company_A/US/H05H13, the unit size/length represents the value of 1 in the data. This length should be consistent for all other comparable instances of Owner/*/Country/H05H13, empty or not. But note that some lengths are not the same in sibling cases, indicated in the "Areas unequal" callout, even though their cooresponding measures are the same. Lengths are only comparable in the vertical dimension at level 3 in this example, not in the horizontal. Once a horizontal or vertical dimension has been used for normalized proportional layout allocation, all further descendent nodes should avoid proportional layouts to maintain the Uniform Density Property. The addition of empty nodes causes this problem.

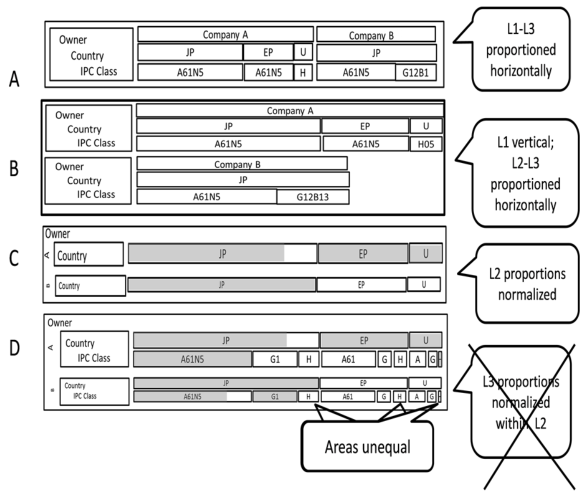


**Figure 8. More layout variants and the limits of proportional partitioning.**

Given this embedded proportionality constraint, it behooves a designer to consider non-proportional layouts for descendants of proportional layouts, as we have done in Figures 7 and 9. Some visualizations are inherently more invariant to differences in horizontal scale such as histograms or line graphs and thus would be good choices here.

We close with a more substantive example of our methods in the domain of patent landscaping in Figure 9. Color mapping is added to what we have discussed so far for additional expressiveness. A color map for the "grouping" attribute ranges in saturation of green and orange to represent positive or negative deviations from the mean of structurally parallel values. Neutral color represents that the values are close to the mean. Small multiples in the form of line graphs are used to represent distributions of patent strength scores in a node of the structure that has already been proportionally divided. Proportional visualizations within the structure help to provide important overviews and, not to be underestimated, afford more graphical area to be allocated to more important data.

A visual patent landscape, of which this would be only a small part, can give an overview in several dimensions at once in which partitioning of the data through attribute hierarchies can afford meaningful comparisons relevant to tasks. A user can be naturally drawn to investigate relevant areas of detail given this visualization. For example, the areas circled in red are important comparable information structures in the data. It is notable that Company A shows weakness (orange color) in the domain groupings within the top red circle. A patent professional would want to investigate where other companies showed IP strength relative to these areas of weakness. This is easily done through this visualization and associated interactivity for getting details on demand in the areas of interest. See [9] for more.

# 4. CONCLUSION

In this short paper we have presented novel methods for comparative visualization of hierarchical multi-attribute data. We have explored the space of combining proportional visualizations (treemaps) with non-proportional visualizations (e.g., 2D histograms and line charts) in order to try to take advantage of the strengths of each. Treemaps afford overviews in which relative area is representative of attribute value distributions across all levels of the hierarchy. Non-proportional visualizations afford comparison across common x- and y-scales and can be invariant to scale. As we have shown, it is possible to mix the two types of visualizations in a single structure, but there are limits to the mixing. Proportional division must be done only once per graphical dimension if parallel structures are to remain visually comparable. We suggested how non-proportional visualizations can be embedded into proportional ones and still provide comparative visual power. Empirical evaluations showing effectiveness of our proposed approach is left for future work.

# 5. ACKNOWLEGEMENTS

Our great thanks to Alessio Malizia of Brunel University London for his support and participation in this project.

# 6. REFERENCES


[1] Chaudhuri, S., and Dayal, U. 1997. An overview of data warehousing and OLAP technology. *SIGMOD Rec.* 26, 1 (March 1997), 65-74.

[2] Chintalapani, G., Plaisant, C., and Shneiderman, B. 2004. Extending the utility of treemaps with flexible hierarchy. In *Information Visualisation. IV 2004. Proceedings, Eighth International Conference on* (July 2004), 335-344.

[3] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., and Roberts, J. C. 2011. Visual comparison for information visualization. *Information Visualization*, 10, 4 (October 2011), 289-309.

[4] Guerra-Gomez, J., Pack, M. L., Plaisant, C., and Shneiderman, B. 2013. Visualizing Change over Time Using Dynamic Hierarchies: TreeVersity2 and the StemView. *Visualization and Computer Graphics, IEEE Transactions on*, 19, 12 (Dec. 2013), 2566-2575.

[5] Kehrer, J., Piringer, H., Berger, W., and Groller, M. E. 2013. A Model for Structure-Based Comparison of Many Categories in Small-Multiple Displays. *Visualization and Computer Graphics, IEEE Transactions on*, 19, 12 (Dec. 2013), 2287-2296.

[6] Shneiderman, B., 1992. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11,1, 92-99.

[7] Sifer, M. User interfaces for the exploration of hierarchical multi-dimensional data. 2006. In *Visual Analytics Science And Technology, 2006 IEEE Symposium on,* 175-182.

[8] Vliegen, R., van Wijk, J. J., and van der Linden, E. 2006. Visualizing Business Data with Generalized Treemaps. *Visualization and Computer Graphics, IEEE Transactions on*, 12, 5 (September 2006), 789-796.

[9] Wittenburg, K., and Pekhteryev, G. 2015. Multi-Dimensional Comparative Visualization for Patent Landscaping. In *Proceedings of BusinessVis Workshop*, IEEE VisWeek, Chicago, IL.


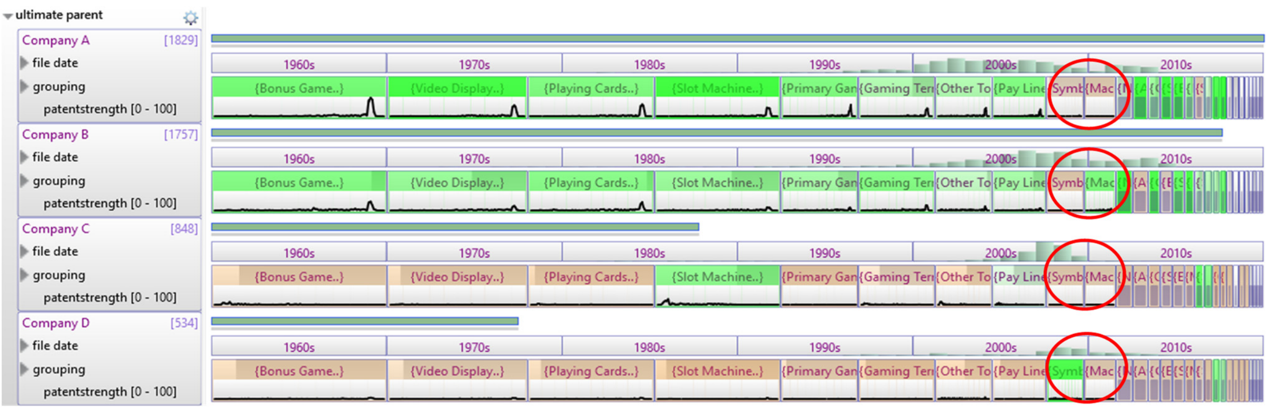


**Figure 9. An application of our methods in patent landscaping. File date indicates filing trends per company. Grouping is an attribute representing domain topics for the patents that has been derived from topic clustering. Groups are further broken down by patent-strength scores. The areas circled show comparable strengths and weaknesses across the landscape.**