Depicting Uncertainty in 2.5D Treemaps

Daniel Limberger
Hasso Plattner Institute,
Faculty of Digital Engineering,
University of Potsdam, Germany

Matthias Trapp
Hasso Plattner Institute,
Faculty of Digital Engineering,
University of Potsdam, Germany

Jürgen Döllner
Hasso Plattner Institute,
Faculty of Digital Engineering,
University of Potsdam, Germany

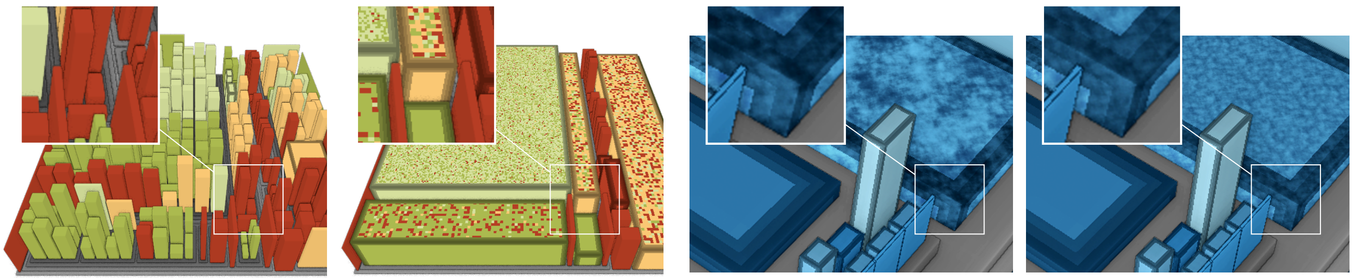

Figure 1: Left: a software map depicting development activity on code units, without and with color weaving and nesting-level contouring applied for the rendering of aggregates. Right: utilization of Perlin noise of low and high frequencies depending on the underlying data distribution. The nesting-level contouring denotes the maximum depth of the collapsed data sub-tree.

ABSTRACT

A truthful and unbiased display of data using information visualization requires detecting and communicating uncertainty. Uncertainty is often inherent in data or is introduced by data processing and visualization (e.g., visual display of accumulated data) but frequently not accounted for. This paper discusses the suitability of advanced visual variables such as sketchiness, noise, nesting-level contouring, and color weaving for communicating uncertainty.

ACM Reference Format:
Daniel Limberger, Matthias Trapp, and Jürgen Döllner. 2020. Depicting Uncertainty in 2.5D Treemaps. In *The 13th International Symposium on Visual Information Communication and Interaction (VINCI 2020), December 8–10, 2020, Eindhoven, Netherlands.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3430036.3432753

1 INTRODUCTION

The expressiveness of any visualization is subject to and limited by the visual variables used to encode data (attributes). With careful consideration, multiple visual variables can be superimposed to one another, without significantly reducing their effectiveness. We use recently introduced visual variables [9] and discuss their potential for the representation of uncertainty in 2.5D treemaps. A truthful and unbiased depiction of the underlying data is relevant, especially when visual analytics tools become regularly used communication artifacts. Haber and McNabb differentiate between visualization of uncertainty and uncertainty of visualization [5]. The first covers uncertainties that inherent to the quality of data, correctness of measurements, trustworthiness of assumptions, as well as probability of trend forecasts. The latter is inherent to accumulated or sampled data, e.g., when using level-of-detail techniques. A visual aggregate is a specialized node representation that can be used to resolve clutter, reduce the cognitive load, or emphasize regions-of-interest, as recently shown for *software maps* [8]. We differentiate between the (1) *attribute uncertainty* and (2) *distribution uncertainty*. The first can be associated directly to mapped attributes and is depicted using the additional visual variables on leaf nodes. For the latter, the same techniques can be applied on aggregates, which, however, introduces information loss and increases uncertainty by means of (1) obfuscating whether or not to explore and investigate further and (2) hindering identification of relevant nodes and outliers. *Color weaving* [6], *sketchiness* [1, 7], *nesting-level contouring* [8], *perlin noise* [2, 3], and *chart-based rendering and labeling* can be used to mitigate these issues.

VINCI 2020, December 8–10, 2020, Eindhoven, Netherlands

The definitive Version of Record was published in *The 13th International Symposium on Visual Information Communication and Interaction (VINCI 2020), December 8–10, 2020, Eindhoven, Netherlands*, https://doi.org/10.1145/3430036.3432753.

2 ATTRIBUTE UNCERTAINTY

In order to allow for additional depiction of uncertainty, we use sketchiness [1, 7] or Perlin noise [2, 3] in addition to node size, color, and height. Both approaches can depict data with inherent quality information or uncertainty data available, such as sampling rate, currentness, error rate, and relevance. Sketchiness, however, seems to be better suited to convey multiple distinguishable degrees of uncertainty. When applied to aggregates, we superimposed nesting-level contours for both techniques. Height and color mapping can still be used, e.g., for encoding the existence of outliers using aggregation operators [8].

Sketchiness can be configured in terms of style (pattern, outline, stroke thickness) and intensity. Increased uncertainty is mapped to increased sketchiness intensity using a 5-point Likert scale (none, minor, moderate, high, and severe/major). Specifics of that mapping depend on the actual task and we assume uncertainty to play a secondary role during analysis. If uncertainty is the primary concern of a task, it should be mapped directly to color or height instead.

When using Perlin Noise, the depth or level of aggregated nodes can be mapped to noise frequency. This is also useful when different levels of a treemap have different semantics, e.g., code units within a file system, budgets, or team composition. The amplitude of the noise is used to encode the degree of uncertainty and can be mapped directly as well as continuously.

3 DISTRIBUTION UNCERTAINTY

For the visual encoding of data distribution three techniques can be taken into account: (1) Perlin Noise with weighted octaves, (2) color weaving, and (3) chart glyphs [10]. The use of Perlin Noise (Figure 1) takes advantage of combining multiple octaves with varying degrees of intensity, thereby resulting in low or high frequent components that can be used to indicate data distribution as well as existence of outliers. Alternatively, chart-depicting glyph accurately encode distribution in a well known manner, e.g., pie charts or box plots (Figure 2). Especially for small aggregates, tiny glyphs can be placed on top of aggregates, e.g., next to their labels.

The uncertainty display of aggregates needs to adhere to *visual summary* (convey information about underlying data), *discriminability* (distinguishable presentation of aggregates and data items), and *interpretability* (remain correctly interpretable within the visual mapping) [4]. Since color weaving [6] is based on color mixing with a given color map, it especially supports interpretability and further improves *fidelity* (aggregates may lie about their underlying data). It can mitigate the uncertainty of the visualization by explicitly depicting it. This in turn allows to guide the user during interactive exploration, e.g., whether or not to investigate certain nodes, identify relevant nodes, detect outlier, or prevent data to be accumulated in misleading ways. Color weaving can also reflect the underlying data with respect to other attributes, resulting in unweighted and weighted depictions (Figure 3).

The patch size specifies the resolution for any mapping. For distribution mapping, for example, a bigger patch size (e.g., 16×16) allows for a more accurate distribution depiction. Furthermore, we adjust the mapping to always account for single occurrences within a value range in order to not lie about the underlying data (note that this marginally distorts the distribution display). Another important parameter is the overall grid scale. For large treemaps with deep nesting levels, the grid scale should be adjusted per hierarchy level. Our implementation also allows for weighted data distribution display (Figure 3) intended for tasks where qualitative color scales with more color classes are used. This way, we can emphasize specific classes explicitly.

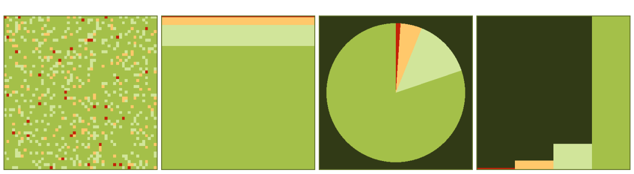

Figure 2: Distribution representations to color weaving (left) using glyphs charts, rendered onto the top faces of aggregates: stacked chart, pie chart, and bar chart.

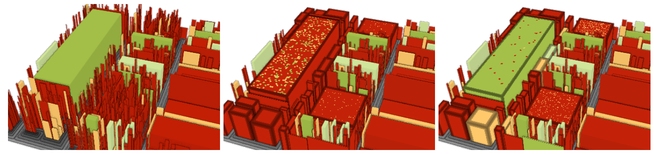

Figure 3: 2.5D Treemap using, unweighted (middle) and weighted (right) color weaving for the visual display of uncertainty of accumulated software system data.

4 DISCUSSION & FUTURE WORK

We found that both Perlin noise and color weaving can be combined with nesting-level contours without limitations. Furthermore, the discriminability of aggregates can be increased and preexisting shading, contouring, shadowing, or highlighting remains unaffected. There are edge cases, however, that require additional adjustment w.r.t. the number of noise octaves, patch size, or the glyph charts placement and scale. In terms of visual software analytics, we explored different data mappings for data item uncertainty, e.g., developer participation, developer activity, as well as data distribution uncertainty for common software metrics and their evolution. Future work could investigate whether color weaving (1) accurately conveys distribution data and (2) outperforms well-known plotting approaches or single-color aggregation strategies. The results would enable automated selection of the best-fit visualization for a given uncertainty and screen estate.

REFERENCES

[1] Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, and Jean-Daniel Fekete. 2012. Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty. *IEEE TVCG* 18 (2012), 2769–2778.

[2] Alexandre Coninx, Georges-Pierre Bonneau, Jacques Droulez, and Guillaume Thibault. 2011. Visualization of uncertain scalar data fields using color scales and perceptually adapted noise. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV)*. 59–66.

[3] Suzana Djurcilov, Kwansik Kim, Pierre Lermusiaux, and Alex Pang. 2002. Visualizing Scalar Volumetric Data with Uncertainty. *Computers and Graphics* 26 (2002), 239–248.

[4] Niklas Elmqvist and Jean-Daniel Fekete. 2010. Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. *IEEE TVCG* 16, 3 (2010), 439–454.

[5] Robert B. Haber and David A. McNabb. 1990. Visualization idioms: a conceptual model for scientific visualization systems.

[6] Haleh Hagh-Shenas, Sunghee Kim, Victoria Interrante, and Christopher G. Healey. 2007. Weaving Versus Blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color. *IEEE TVCG* 13, 6 (2007), 1270–1277.

[7] Daniel Limberger, Carolin Fiedler, Sebastian Hahn, Matthias Trapp, and Jürgen Döllner. 2016. Evaluation of Sketchiness as a Visual Variable for 2.5 D Treemaps. In *Proc. IEEE iV*. 183–189.

[8] Daniel Limberger, Willy Scheibel, Sebastian Hahn, and Jürgen Döllner. 2017. Reducing Visual Complexity in Software Maps using Importance-based Aggregation of Nodes. In *Proc. IVAPP*. INSTICC, SciTePress, 176–185.

[9] Daniel Limberger, Willy Scheibel, Matthias Trapp, and Jürgen Döllner. 2019. Advanced Visual Metaphors and Techniques for Software Maps. In *Proc. ACM VINCI*. 8.

[10] A. G. Soares, D. H. dos Santos, C. L. Barbosa, A. S. Goncalves, C. G. dos Santos, B. S. Meiguins, and E. T. Miranda. 2018. Visualizing Multidimensional Data in Treemaps with Adaptive Glyphs. In *Proc. IEEE iV*. 58–63.