

# Poster: Application of Treemaps to Business Statistics Analysis

Martijn Tennekes\*  
Statistics Netherlands

Edwin de Jonge†  
Statistics Netherlands

## ABSTRACT

Statistics Netherlands produces statistics about the economic activity in The Netherlands. These statistics are based on survey data and administrative sources, such as value added tax (VAT). A recent trend in the production of official statistics is to use a top-down analysis, which means that analysts first analyze high level aggregated data, and then zoom in on a more detailed level when necessary. In this paper, we discuss how treemap visualizations can be used for this top-down approach. We use comparison treemaps, density treemaps, and small multiples. Finally, we introduce a method to visualize confidence intervals in treemaps.

**Keywords:** Treemaps, Official Statistics

**Index terms:** H.5.0 [Information Interfaces and Presentation]: General; J.1 [Administrative Data Processing] (Government)


## 1 INTRODUCTION

In order to publish statistics of the total economy in The Netherlands, Statistics Netherlands (CBS) collects survey data from Dutch companies. Topics that appear on these questionnaires are total turnover, number of persons employed, total purchases, financial result, etcetera.

Before survey data can be used to make estimations of the economy in The Netherlands, this data has to be analyzed and edited. Usually, there are many data errors and inconsistencies: for instance, when the wages and salaries are not in line with the number of persons employed.

Traditionally, data analysts correct the data of the enterprises one by one using tables and spreadsheets. A better, more efficient way is to use a top-down approach [1, 4]: start with analyzing aggregated data per economic sector or enterprise size (based on the number of persons employed), and in case of an unexpected outcome, zoom to the specific subsector, and then to the specific enterprise(s) that cause(s) the unexpected outcome.

In this paper we explore several applications of treemap visualizations [5] for analysis of business statistics. We think these visualizations will make the top-down approach more effective and transparent.

All treemaps in this paper are constructed in . We implemented the ordered treemap algorithm [2]. For the coloring of the rectangles, we make use of color scales from ColorBrewer [3]. We use real, anonymized Dutch business statistics data from 2006 and 2007.

## 2 COMPARISON TREEMAPS

Statistical data analysts are interested in disruptive or unexpected changes in time. These changes can be real events, but often are indicators for data errors. Both cases are of interest: are changes taking place in one industry? Is it a big or small effect compared

\*e-mail: m.tennekes@cbs.nl

†e-mail: e.dejonge@cbs.nl

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

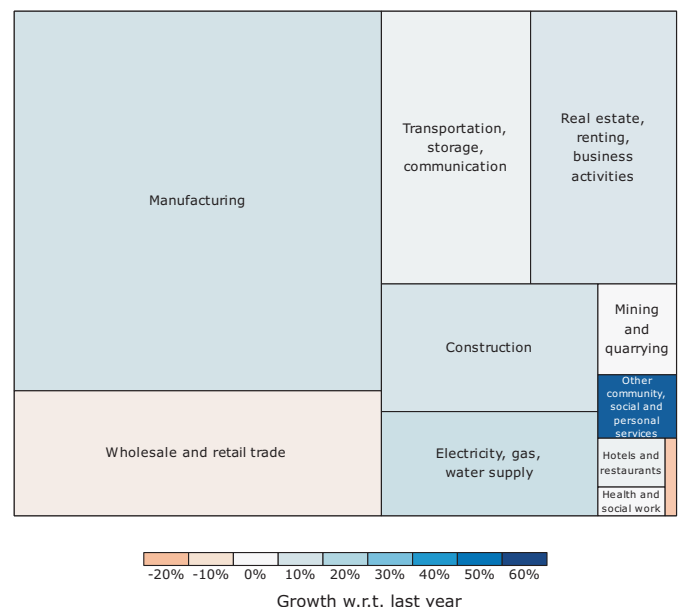
to other industries? These questions can be quickly assessed using comparison treemaps.

Figure 1 shows the value added (at factor cost) of the active enterprises in The Netherlands. The sizes of the rectangles correspond to the total value added of the different sectors. We use a color scale to indicate the growth (or shrink) with respect to the previous year. White is used for values that didn't change, blue for increasing and red for decreasing values.

This treemap shows only one aggregation level. It is also possible to divide the sectors into subsectors.

By this treemap, data analysts can quickly judge whether the data seem to be correct. If for instance the loss in the sector *wholesale and retail trade* is unexpected, data analysts can zoom in on this sector to find out which subsectors or companies cause this loss.

Figure 1: Total value added aggregated by sector



## 3 DENSITY TREEMAPS

In the treemaps we discussed so far, we used the colors to indicate changes in time. Colors can also be used to introduce a second variable. Of course, this can be done by mapping the second variable to a color scale. A more natural way is to use density colors. In this section, we discuss the usage of density colors in treemaps.

An example of a density treemap is shown in Figure 2. The main variable, the total number of persons employed per sector, determines the sizes of the rectangles. The colors indicate how much turnover is generated per person employed. The darker the color, the higher this amount.

By this treemap, analysts can intuitively observe how turnover is related to the number of persons employed. For instance, they can easily observe that although only a very small part of the people employed are working in the sector at the bottom right (*mining and quarrying*), this sector generates relatively much turnover.

Figure 3: Subsectors of *manufacturing*

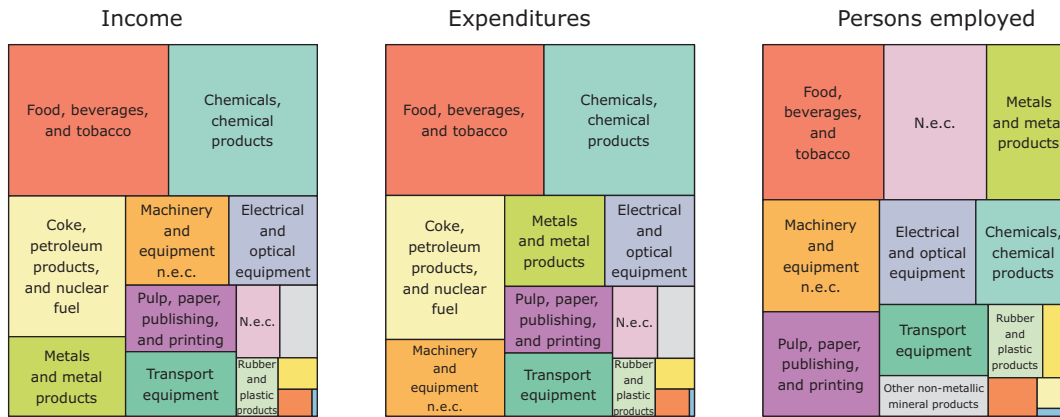
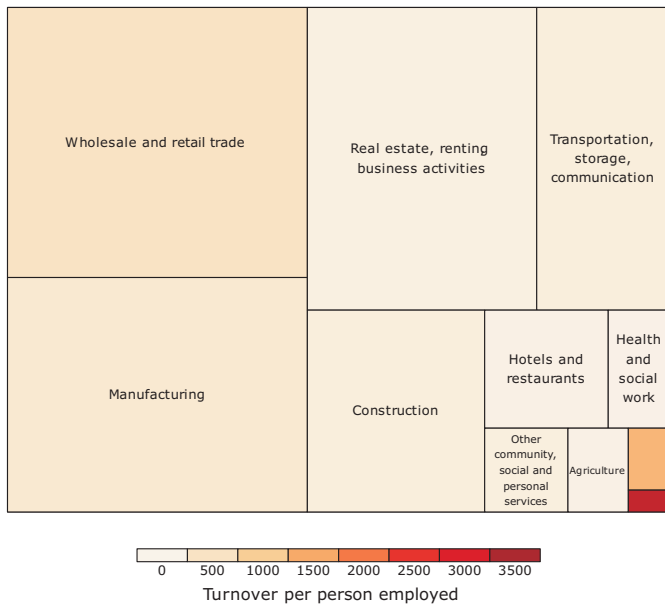


Figure 2: Number of persons employed aggregated by sector



Many other quantitative variables can be displayed by density treemaps, for instance the number of persons employed versus the personnel costs.

#### 4 VISUALIZING CONFIDENCE INTERVALS

The visualized data are estimations from a sample of Dutch enterprises. For proper analysis, confidence intervals are at least as important as the estimators themselves. An estimation with a very narrow confidence interval is more reliable than an estimation with a very wide confidence interval.

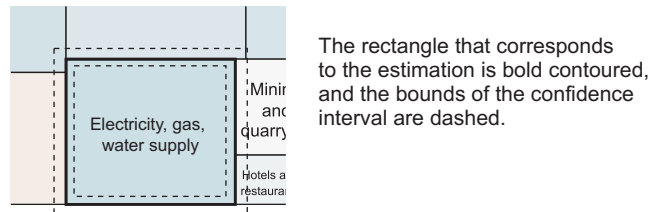
We plot an estimation rectangle with a bold contour and a lower and upper bound rectangle unfilled with a dashed contour. This is illustrated in Figure 4.

We decided to plot the confidence rectangles of only one estimation rectangle (that is, the selected one) at a time, since plotting the confidence of all estimation rectangles results in a visual clutter.

#### 5 SMALL MULTIPLES

Usually, data analysts focus on multiple variables at a time, because many variables are strongly related to each other. Small multiple

Figure 4: Visualization of a confidence interval



treemaps can be very effective for this job.

We use a color scale with identifiable colors to link the treemaps. In this way, analysts are able to compare the proportions of the rectangles of one treemap with those of another.

In Figure 3 we show three variables for the subsectors of the sector *manufacturing*. Analysts can easily observe that the total income has the same division into subsectors as the total expenditure, which is what they would expect. However, the division of the number of employees among the subsectors is different.

## 6 CONCLUSIONS AND FUTURE WORK

In this paper, we applied treemaps to the analysis of business statistics data. We applied comparison treemaps, density treemaps, small multiple treemaps, and we visualized confidence intervals.

For future research, we would like to further develop our methods, especially regarding interaction. Moreover, we like to evaluate our proposed methods by data analysts.

## REFERENCES

- [1] F. Aelen and R. Smit. Towards an efficient data editing strategy for economic statistics at statistics netherlands. European Establishment Statistics Workshop, 7-9 September 2009, Stockholm, Sweden.
- [2] B. B. Bederson, B. Shneiderman, and M. Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Trans. Graph.*, 21(4):833–854, 2002.
- [3] C. A. Brewer, G. W. Hatchard, and M. A. Harrower. Colorbrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30(1):5–32, 2003.
- [4] W. Hacking. Macro-selection and micro-editing: a prototype. In *IBUC 2009 12<sup>th</sup> International Blaise Users Conference*, pages 118–125, 2009.
- [5] B. Shneiderman. Tree visualization with treemaps. A 2D space-filling approach. *ACM Trans. Graph.*, 11(1):92–99, 1992.