Application of Treemaps to Business Statistics Analysis

Martijn Tennekes
Statistics Netherlands (CBS)

Edwin de Jonge
Statistics Netherlands (CBS)

Introduction

Statistics Netherlands (CBS) produces statistics about the Dutch economy.

Production stages:
- data collection (questionnaires and administrative)
- data editing
- data analysis
- output

Traditional approach of data editing and analysis:
- analysis of data at level of enterprises
- tables and spreadsheets only

Top-down approach [1, 4]:
- analysis of aggregated data (e.g. by sector), and only in case of unexpected outcome: zoom in
- usage of plots

We think treemaps are very effective for this top-down approach.

Algorithm: Ordered treemap algorithm [2]
Color scales: ColorBrewer [3]
Implemented: R-package `treemap`

We apply several treemap methods to business statistics data: comparison treemaps, density treemaps and small multiples (forestmap). Further, we introduce a method to visualize confidence intervals.

Comparison treemaps

Goal: Detect disruptive or unexpected changes in time.

Sizes: aggregated variable *y* at period *t*
Colors: growth of *y* w.r.t. period *t-1*

Figure 1 shows the value added (at factor cost) of the active enterprises in The Netherlands, aggregated by economic sector and subsector (NACE Rev. 1.1.).

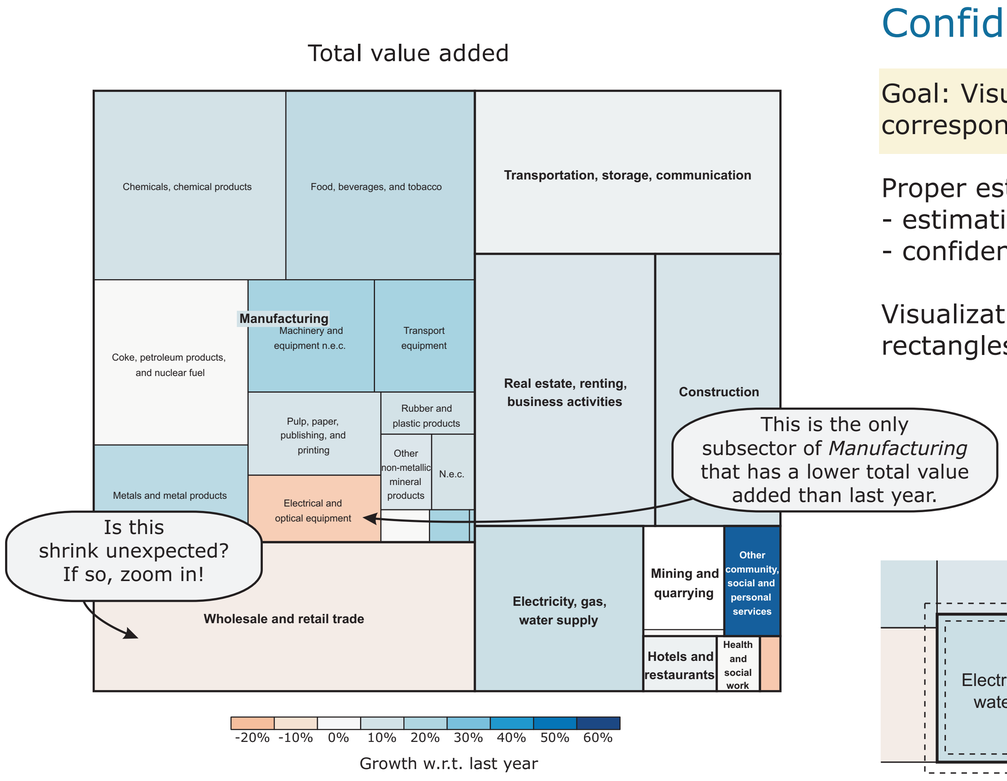

Figure 1: Inspect changes in value added w.r.t. last year

Density treemaps

Goal: Analyze the relationship between two variables.

Sizes: aggregated variable *y*
Colors: density parameter *x / y*

In Figure 2, a density treemap is shown.

By this treemap, analysts can intuitively observe how turnover is related to the number of persons employed.

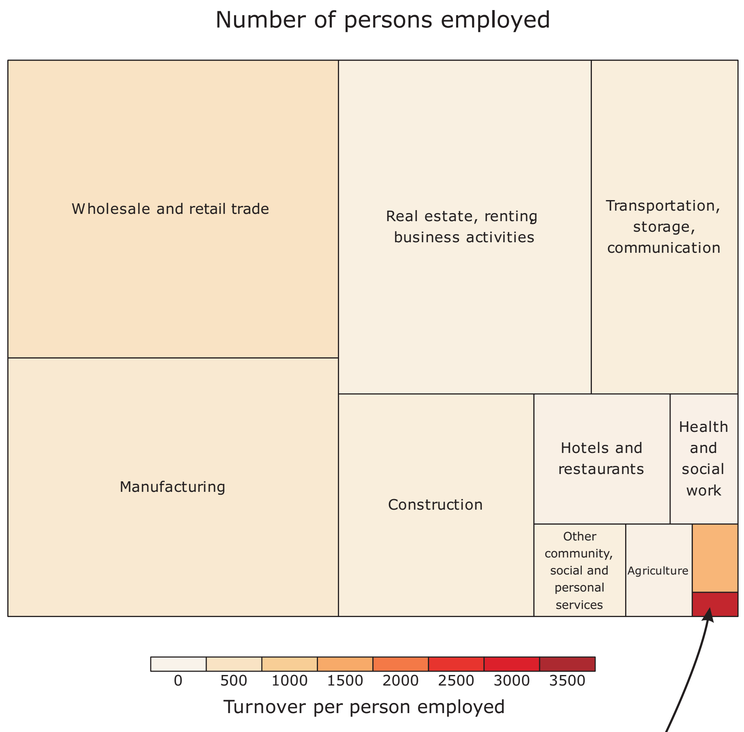

Figure 2: Highlight economic sectors with high turnover/person

Confidence intervals

Goal: Visualize the confidence interval along with the corresponding estimation of a parameter.

Proper estimations of the Dutch economy:
- estimations
- confidence intervals

Visualization of lower and upper bound: dashed rectangles. See Figure 3.

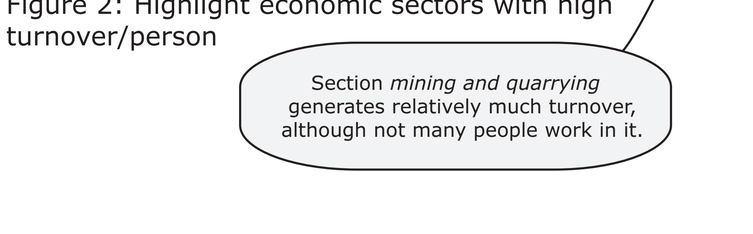

Figure 3: Show uncertainty of value added

Small multiples: forestmap

Goal: Analyze multiple variables at once.

Sizes: aggregated variable *y*
Colors: identifiable (such that they are visually linked)

Multiple comparison or density treemaps are also possible.

In Figure 4 we show four variables for the subsectors of the sector manufacturing.

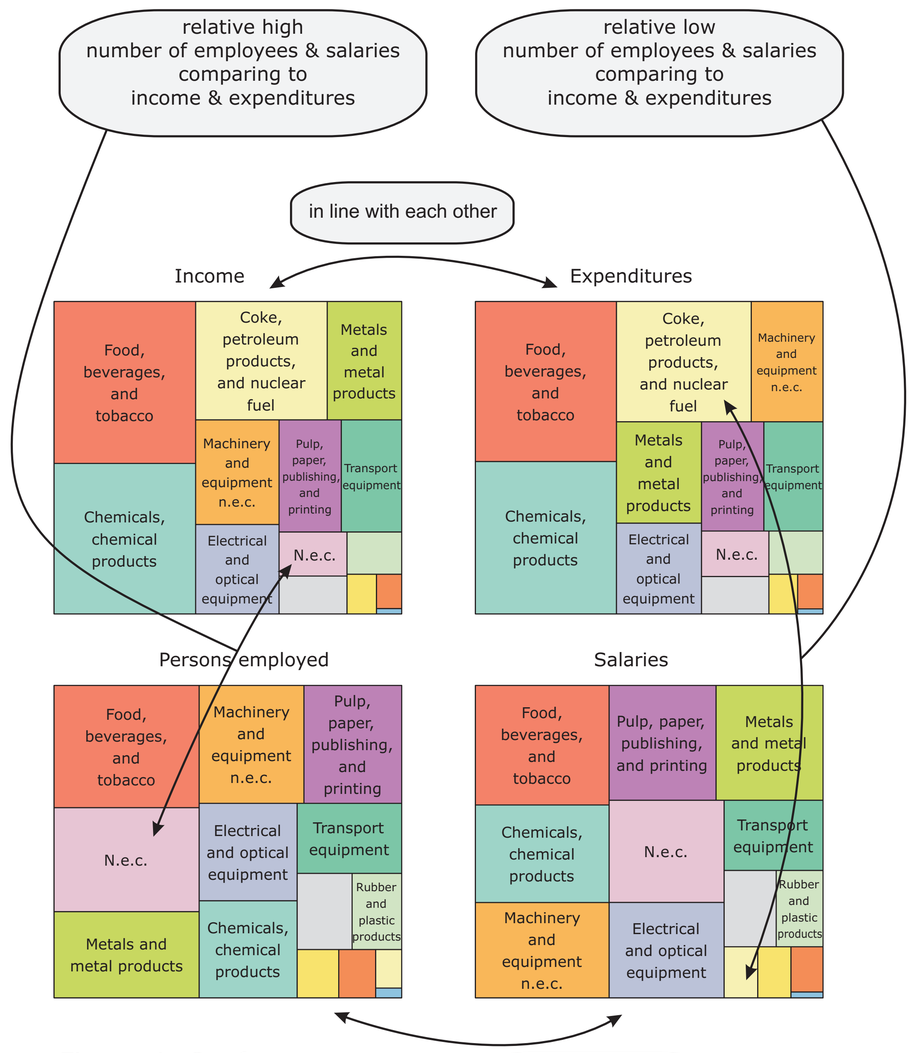

Figure 4: Confront income, expenditures, employment and salaries

Future work

- Further development, especially regarding interaction
- Evaluation 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 12th International Blaise Users Conference, pages 118–125, 2009.

Statistics Netherlands

Correspondence to:

Martijn Tennekes

E-mail m.tennekes@cbs.nl
Phone +31 455706372
Fax +31 455706244

Statistics Netherlands
Division of Methodology and Quality

P.O. Box 4481
6401 CZ, Heerlen
The Netherlands