

## Visualization of Complex Statistical Information in Maps

Matthias Templ\*

Beat Hulliger†

Alexander Kowarik‡

### Abstract

The checkerplot is a new type of graphics that combines geographical information and plots. Hereby, traditional plots like barplots or polygon lines are visualized in geographical order on a grid. The checkerplots can be seen as a mixture between thematic maps and the grid-representation in trellis plots.

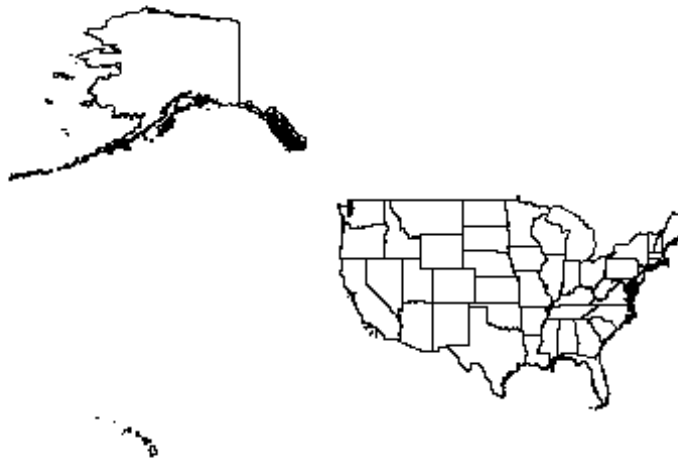
**Key Words:** Indicators on Regional Level, Visualization, Software R

### 1. Traditional Methods

Traditional methods either present statistical information on areas either in a thematic map, or by presenting the information in bar charts, or in trellis graphics where the regions often ordered in alphabetic order on a grid.

For the latter two possibilities, surely the geographical information is almost lost and therefore these methods of presenting geographical information is not optimal.

But there exists also problems to present complex statistical informations in thematic maps, because of small regions, see e.g. Figure 1. Here readability and comparability is lost whenever statistical information is presented for the small states in the east.



**Figure 1:** A possible representation of a map of the US states.

---

\*Statistics Austria, Guglgasse 11, 1110 Vienna

†Institute for Competitiveness and Communication, FHNW Switzerland

‡Statistics Austria, Guglgasse 11, 1110 Vienna

## 2. The Checkerplot

A checkerplot, a stylized map in rectangular grid-like form (like a checker board), in which geographical position and neighborhood are approximately (as best as possible) maintained. Therefore, for all geographical entities an equally sized plotting region exists on a grid.

The checkerplot combines all advantages of traditional statistical graphs and representation in thematic maps:

- All areas have equal space for presenting (even complex) statistical information.
- The geographical position of the areas are approximately maintained.
- The common baseline of the plots in a row (or column) helps comparison over rows (or over columns).

The areas are optimally arranged in that grid to preserve the geographical information, for which a complex linear programming problem has to be solved. For the mathematical description of the problem we refer to Templ et al. [2012].

## 3. Examples

### 3.1 US Health Insurance Coverage

In the United States of America, health insurance is provided by government sponsored entities as well as by private insurances.

Figure 2 shows a checkerplot of not covered health insurance rates in 2008 and 2009. The data presented on an  $10 \times 10$  grid. On this chosen grid, the states are arranged optimally. As expected, Alaska can be found in the upper left, on the west- to south boarder of US, one can find California, Arizona, New Mexico, Texas, . . . and also all other states should be found on their approximately position. Comparisons over rows and columns and finding a geographical trend can easily be done.

### 3.2 Unemployment Rates in Europe

Figure 3 shows a checkerplot that optimally assigns each European Union member state (including Switzerland and Norway) into a  $8 \times 7$  grid. For each country the unemployment rate (based on year 2008) is presented with breakdown by gender. In addition, confidence intervals of the unemployment rate are drawn optimally. Since real values for the variances were not available, faked values for the confidence intervals are plotted. It is easy to see, that the Scandinavian countries, which are located in the north of Europe, are at the top of the grid, Portugal, the most westerly country of Europe, can be found on the lower left beside Spain. In summary, the geographical order is well preserved. Additionally, the flags in the labels of each area help the user to rapidly recognise a particular region he or she is looking for.

## 4. The Implementation of the Checkerplot: The sparkTable Package

The checkerplot is available in the `sparkTable` R-package [Kowarik et al., 2012] with it's utility function `optimal_grid_allocation()` to solve the linear program.

Using the facilities of this package, which is an implementation of *the grammar of graphics* book by Wilkinson [2005], allows a flexible implementation of Wickham [2009]

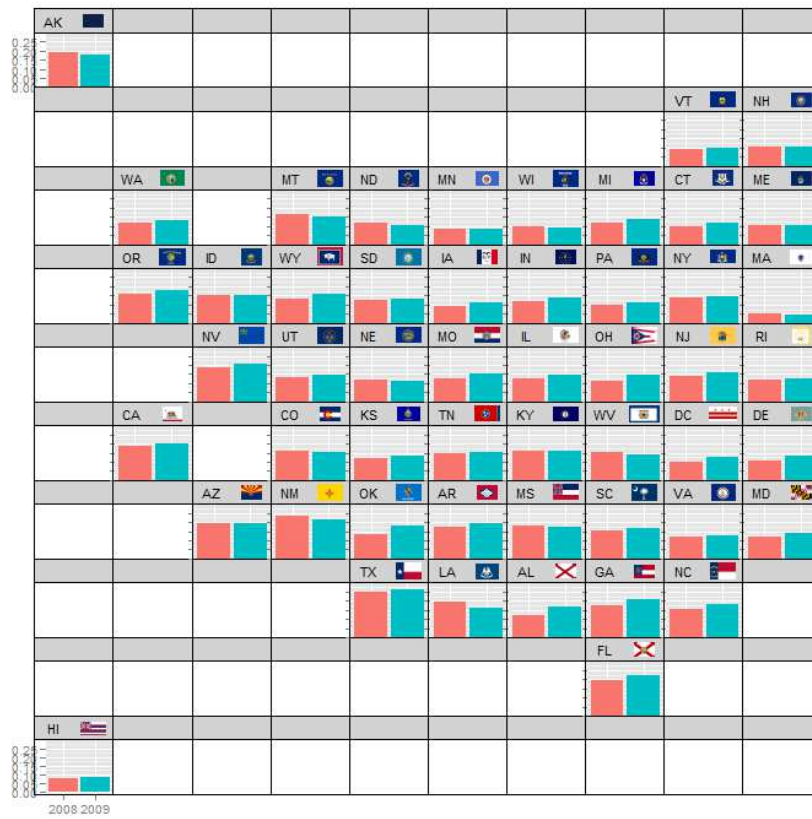


Figure 2: US Government health insurance - not covered, 2008 compared to 2009.

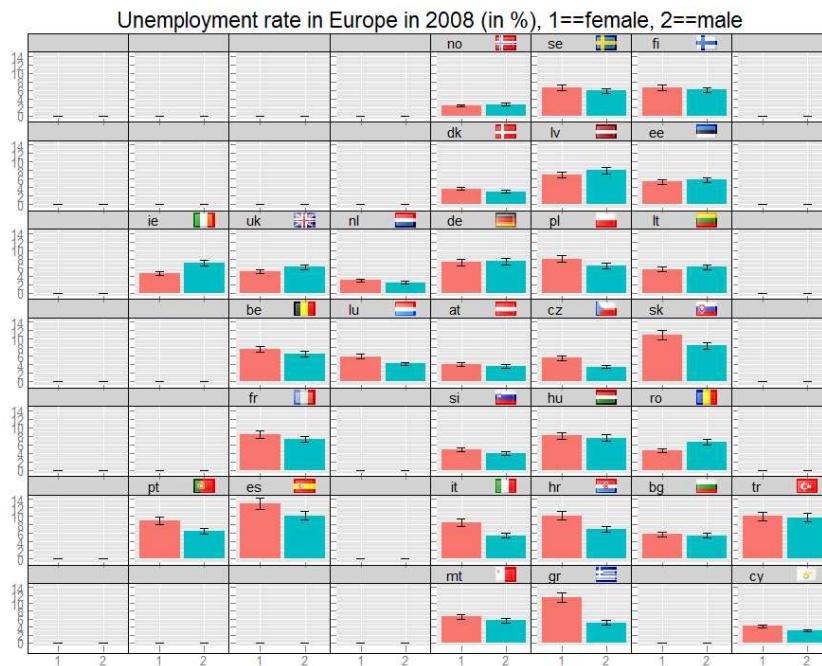


Figure 3: Unemployment rate of men and females in Europe. Faked error bars are plotted.

developed enhancements to the grammar of graphics book, the *layered grammar of graphics*. The grammar tells us that a statistical graphic is a mapping from data to aesthetic attributes (colour, shape, size) of geometric objects (points, lines, bars). The plot may also contain statistical transformations of the data and is drawn on a specific coordinate system. Faceting can be used to generate the same plot for different subsets of the dataset. It is the combination of these independent components that make up a graphics.

**Explicit Parameters for the Corresponding Checkerplot Functions** First, the optimal arrangement in the grid has to be determined. In function `optimal_grid_allocation` (see Listing 1), the data (coordinates), the number of grid columns and rows can be specified amongst others.

```
optimal_grid_allocation <- function(data, grid.cols=NULL, grid.rows=NULL,
  addGrid=0, plot=FALSE){ ... }
```

Listing 1: Function arguments of the optimal arrangement function.

This order should be provided to the `checkerplot()` function for which the function arguments are shown in Listing 2.

```
checkerplot <- function(data, cols=5, rows=5, geom="line", title=NULL,
  title.size=20, label.size=11, ymin=NULL, ymax=NULL
  ,
  img=NULL, aes_geom=NULL, formatter=NULL, margin_
  yaxis=0,
  opts=NULL, ...){
  ...
}
```

Listing 2: Function arguments of the checkerplot function.

To display the corresponding data, the data itself must be provided as a matrix within function argument `data` with the following structure

- Column 1: x-values
- Column 2: y-values
- Column 3: labels
- Column 4: order

The function arguments `cols` and `rows` correspond to the desired columns and rows in the grid. `geom` defines the kind of the plot such as 'line', 'bar' or 'point'. `title`, `title.size` and `label.size` are self-explaining arguments. `ymin` and `ymax` define the floor and cap of the *y*-axis, if not provided it is automatically derived from the data. Function argument `img` has to be a vector with the names of the flags of the underlying areas. If not provided, no flags are put on the graphic. `formatter` and `margin_yaxis` allows to format the values of the *y*-axis and to adjust plot margins of each grid. With `aes_geom` and `opts` the aesthetics of the corresponding geom `ggplot2` object and the settings in the `ggplot2` plots can be changed [for details, have a look at Wickham, 2009].

In the current implementation of the plot, the images have to be provided in the *pnm* file format. The R package `pixmap` [Bivand et al., 2011] is used for this task.

## 5. Conclusion

The checkerplot combines the advantages of thematic maps and of grid-based Trellis plots. It approximates the underlying geography on a low resolution grid. In that grid complex information can be visualized while having a maximum of readability (same space for each region) and comparability (over rows, columns or in the plain).

The implementation in R makes the application possible for any user.

## References

- R. Bivand, F. Leisch, and M. Maechler. *pixmap: Bitmap Images (“Pixel Maps”)*, 2011. URL <http://CRAN.R-project.org/package=pixmap>. R package version 0.4-11.
- A. Kowarik, B. Meindl, and M. Templ. *sparkTable: Sparklines and graphical tables for tex and html*, 2012. URL <http://CRAN.R-project.org/package=gridExtra>. R package version 0.9.3.
- M. Templ, B. Hulliger, A. Kowarik, and K. Fürst. Combining geographical information and traditional plots: the checkerplot. *International Journal of Geographical Information Science*, 2012. doi: 10.1080/13658816.2012.684386. URL <http://www.tandfonline.com/doi/abs/10.1080/13658816.2012.684386>. in press, online available.
- H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL <http://had.co.nz/ggplot2/book>.
- L. Wilkinson. *The Grammar of Graphics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. ISBN 0387245448.