On the visualization of hierarchical, tabular and spatial data in R

Martijn Tennekes
Visualization of data: why?

- Exploration: what’s in the data?
- Analysis: what does the data tell you?
- Communication: how to let the data speak?
- Publication: how to make the data attractive and insightful for a broad audience?
Standard visualization methods

Scatter plot, line chart, bar chart, histogram, boxplot, etc.

Especially useful for small datasets, i.e.
– up to 1000 units,
– at most 3 variables (most plots are uni- or bivariate),
– preferably without missing values.

R:
– **base** graphics: useful for quick plots
– **ggplot2**: elegant plotting system
Data in Official Statistics

In the **real world** of Official Statistics:
- large data, millions of units, dozens of variables;
- missing values are very common;
- data often have a hierarchical structure (e.g., classification of goods or jobs);
- data often have a spatial component.
Missing values

**VIM** package (M. Templ et al.): standard plot types extended with missing values, e.g.
Hierarchical data

Applications within Official Statistics:

• Economic activity
• Goods
• Jobs
• Regions
Overview of tree visualizations

treevis.net - A Visual Bibliography of Tree Visualization 2.0 by Hans-Jörg Schulz

<table>
<thead>
<tr>
<th>Dimensionality</th>
<th>Representation</th>
<th>Alignment</th>
<th>Fulltext Search</th>
<th>Techniques Shown</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>All</td>
<td>All</td>
<td></td>
<td>276</td>
</tr>
</tbody>
</table>

![Image of tree visualizations](image-url)
Treemap

Shneiderman (1992)
Treemap

Shneiderman (1992)

Total (9)

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>B.1</td>
<td>2</td>
</tr>
<tr>
<td>B.2</td>
<td>2</td>
</tr>
<tr>
<td>B.3.a</td>
<td>1</td>
</tr>
<tr>
<td>B.3.b</td>
<td>1</td>
</tr>
</tbody>
</table>
Treemap

Shneiderman (1992)

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>B.1</td>
<td>2</td>
</tr>
<tr>
<td>B.2</td>
<td>2</td>
</tr>
<tr>
<td>B.3.a</td>
<td>1</td>
</tr>
<tr>
<td>B.3.b</td>
<td>1</td>
</tr>
</tbody>
</table>
Treemap

Shneiderman (1992)

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>B.1</td>
<td>2</td>
</tr>
<tr>
<td>B.2</td>
<td>2</td>
</tr>
<tr>
<td>B.3.a</td>
<td>1</td>
</tr>
<tr>
<td>B.3.b</td>
<td>1</td>
</tr>
</tbody>
</table>

R-package **treemap**
Treemap

Shneiderman (1992)

Class | Value
--- | ---
A | 3
B.1 | 2
B.2 | 2
B.3.a | 1
B.3.b | 1

R-package treemap
Treemap

Structural Business Statistics: aggregated by economic activity
How to assign a color palette to a tree structure?

**Approach:**

- **Hue** resembles branches
- **Chroma** and **Luminance** discriminate hierarchical levels
Tree Colors (Tennekes and De Jonge, 2014)

How to assign a color palette to a tree structure?

Approach:

- **Hue** resembles branches
- **Chroma** and **Luminance** discriminate hierarchical levels
Tree Colors (Tennekes and De Jonge, 2014)

(a) Sunburst diagram

(b) Treemap

Net turnover for economic activity sector G, *Wholesale and retail trade*
**Large tabular data**

Number of variables: around 5 – 20
Number of units: 10,000 - billions

<table>
<thead>
<tr>
<th></th>
<th>var1</th>
<th>var2</th>
<th>var3</th>
<th>var4</th>
<th>var5</th>
<th>var6</th>
<th>var7</th>
<th>var8</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit10000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Applications in Official Statistics:
- Large survey data
- Admin data
- Big data
**Tableplot**: visual summary of a large data table

1. Sort the data according to the values of a key variable (say var1).
2. Group the data into, say, 100 equally sized bins.
3. Per bin, do
   - for each numeric variable: calculate mean and sd,
   - for each categorical variable: calculate frequencies.
4. Plot it! (see next slides...)
library(tabplot)

# load data
library(ggplot2)
data(diamonds)

# plot it
library(tabplot)
tableplot(diamonds)
Tableplot of the Dutch Virtual Census (test file, 2009)
Tableplot of the Insurance Policy Record Administration (test file, October 2010)
Spatial data

Spatial data in Official Statistics:
• Regional statistics (e.g. NUTS areas, municipalities)
• Exploration of spatial distributions
• Specific GIS publications, e.g. land use.

R-package **tmap**

Thematic maps in R
Thematic map

Geographic map + Theme
Thematic map
Building a thematic map
Building a thematic map

tm_shape(NLD_muni, projection="rd") +
Building a thematic map

tm_shape(NLD_muni, projection="rd") +
tm_fill()
Building a thematic map

```r
tm_shape(NLD_muni, projection="rd") +
tm_fill("blue")
```
Building a thematic map

tm_shape(NLD_muni, projection="rd") +

tm_fill("population")
tm_shape(NLD_muni, projection="rd") +

tm_fill("population", convert2density=TRUE, style="kmeans", title="Population per km2") +
Building a thematic map

tm_shape(NLD_muni, 
    projection="rd") +

tm_fill("population", 
    convert2density=TRUE, 
    style="kmeans", 
    title="Population per km2") +

tmBorders(alpha=.5) +
Building a thematic map

```r
tm_shape(NLD_muni, projection="rd") +

tm_fill("population",
  convert2density=TRUE,
  style="kmeans",
  title="Population per km2") +

tm_borders(alpha=.5) +

tm_shape(NLD_prov) +

tmBorders(lwd=2) +
```
Building a thematic map

```
tm_shape(NLD_muni, projection="rd") +

  tm_fill("population", convert2density=TRUE, style="kmeans", title="Population per km2") +

  tm_borders(alpha=.5) +

  tm_shape(NLD_prov) +

  tm_borders(lwd=2) +

  tm_text("name", size=.8, shadow=TRUE, bg.color="white", bg.alpha=.25)
```
tm_shape(world) +
  tm_fill("income_grp", palette="-Blues", title="Income class") +
  tm_borders() + tm_text("iso_a3", size="AREA") +
  tm_shape(metro) +
  tm_bubbles("x2010", col = "growth", border.col = "black", border.alpha = .5, style="fixed", breaks=c(-Inf, 0, 2, 4, 6, Inf), palette="-RdYlBu", title.size="Metro population (2010)", title.col="Annual growth rate (\%)") +
  tm_format_World(bg.color = "gray80")
Interactive map

tmap_mode("view")
map1  # to which the previous plot has been assigned
US choropleth

2010 Adult Obesity by County, percent

Data © United States Department of Agriculture
Shape © United States Census Bureau
Crimes in Greater London

Dot map
Crimes in Greater London

Dasymetric map
### Crimes in the City of London

#### Small multiples

<table>
<thead>
<tr>
<th>Category</th>
<th>Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-social behaviour</td>
<td><img src="image1.png" alt="Map of Anti-social behaviour" /></td>
</tr>
<tr>
<td>Bicycle theft</td>
<td><img src="image2.png" alt="Map of Bicycle theft" /></td>
</tr>
<tr>
<td>Burglary</td>
<td><img src="image3.png" alt="Map of Burglary" /></td>
</tr>
<tr>
<td>Criminal damage and arson</td>
<td><img src="image4.png" alt="Map of Criminal damage and arson" /></td>
</tr>
<tr>
<td>Drugs</td>
<td><img src="image5.png" alt="Map of Drugs" /></td>
</tr>
<tr>
<td>Other crime</td>
<td><img src="image6.png" alt="Map of Other crime" /></td>
</tr>
<tr>
<td>Other theft</td>
<td><img src="image7.png" alt="Map of Other theft" /></td>
</tr>
<tr>
<td>Possession of weapons</td>
<td><img src="image8.png" alt="Map of Possession of weapons" /></td>
</tr>
<tr>
<td>Public order</td>
<td><img src="image9.png" alt="Map of Public order" /></td>
</tr>
<tr>
<td>Robbery</td>
<td><img src="image10.png" alt="Map of Robbery" /></td>
</tr>
<tr>
<td>Shoplifting</td>
<td><img src="image11.png" alt="Map of Shoplifting" /></td>
</tr>
<tr>
<td>Theft from the person</td>
<td><img src="image12.png" alt="Map of Theft from the person" /></td>
</tr>
<tr>
<td>Vehicle crime</td>
<td><img src="image13.png" alt="Map of Vehicle crime" /></td>
</tr>
<tr>
<td>Violence and sexual offences</td>
<td><img src="image14.png" alt="Map of Violence and sexual offences" /></td>
</tr>
</tbody>
</table>
Crimes in the City of London

Interactive map
Interactive dot map

Prototype: http://research.cbs.nl/ColorDotMap
References

**treemap**
- CRAN version 2.4-1
- [https://github.com/mtennekes/treemap](https://github.com/mtennekes/treemap)

**tabplot**
- CRAN version 1.3
- [https://github.com/mtennekes/treemap](https://github.com/mtennekes/treemap)

**tmap**
- CRAN version 1.4
- [https://github.com/mtennekes/tmap](https://github.com/mtennekes/tmap) (with many links on the home page)
- Paper in review process ...