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
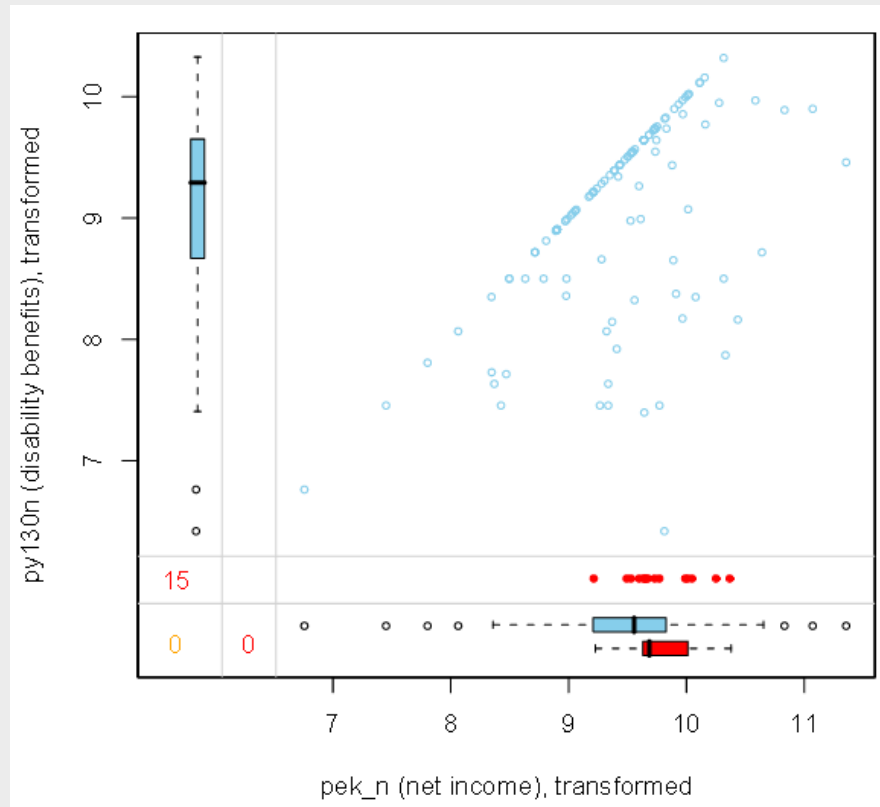
Data in Official Statistics

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);

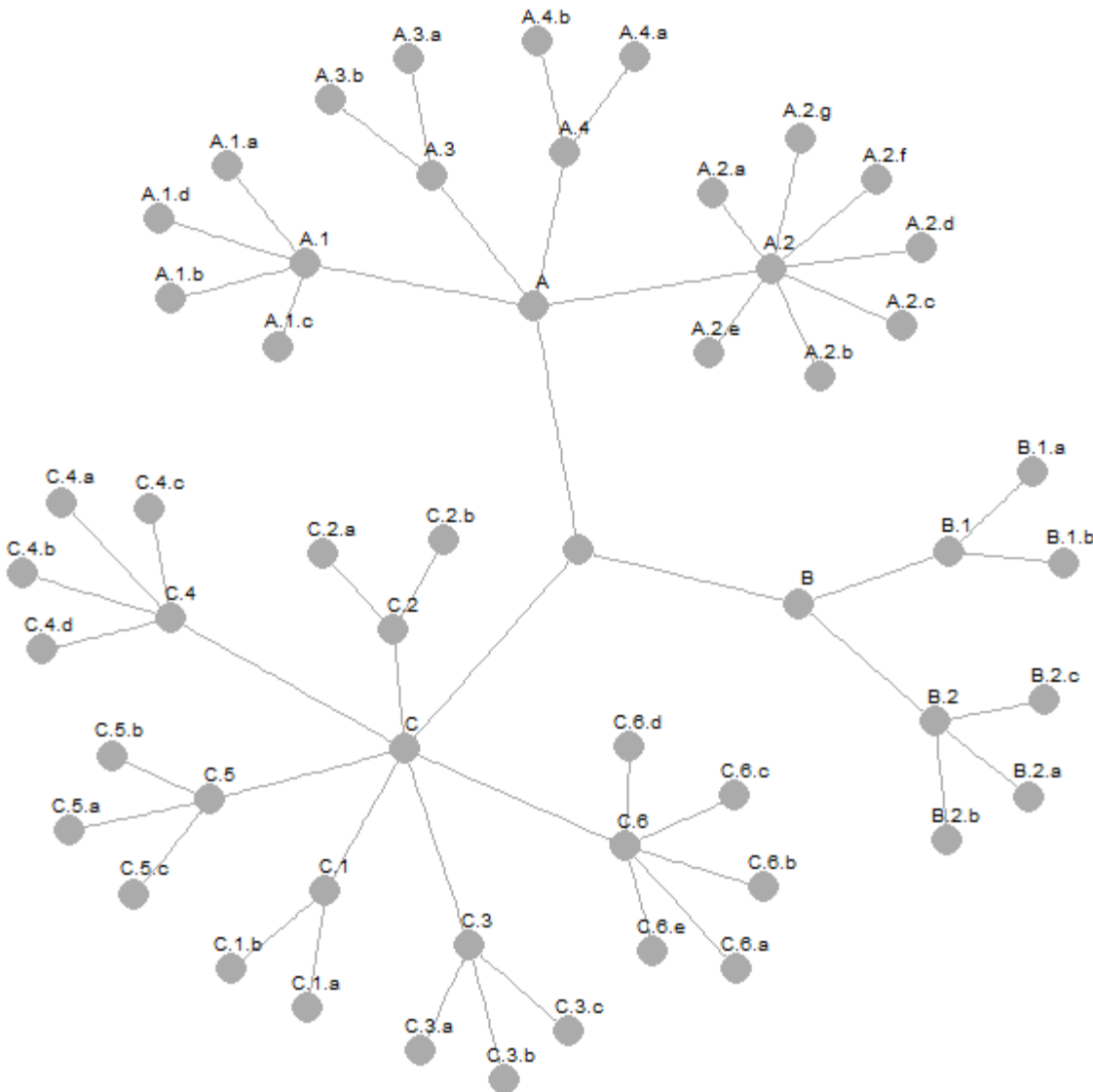- data often have a spatial component.

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

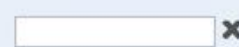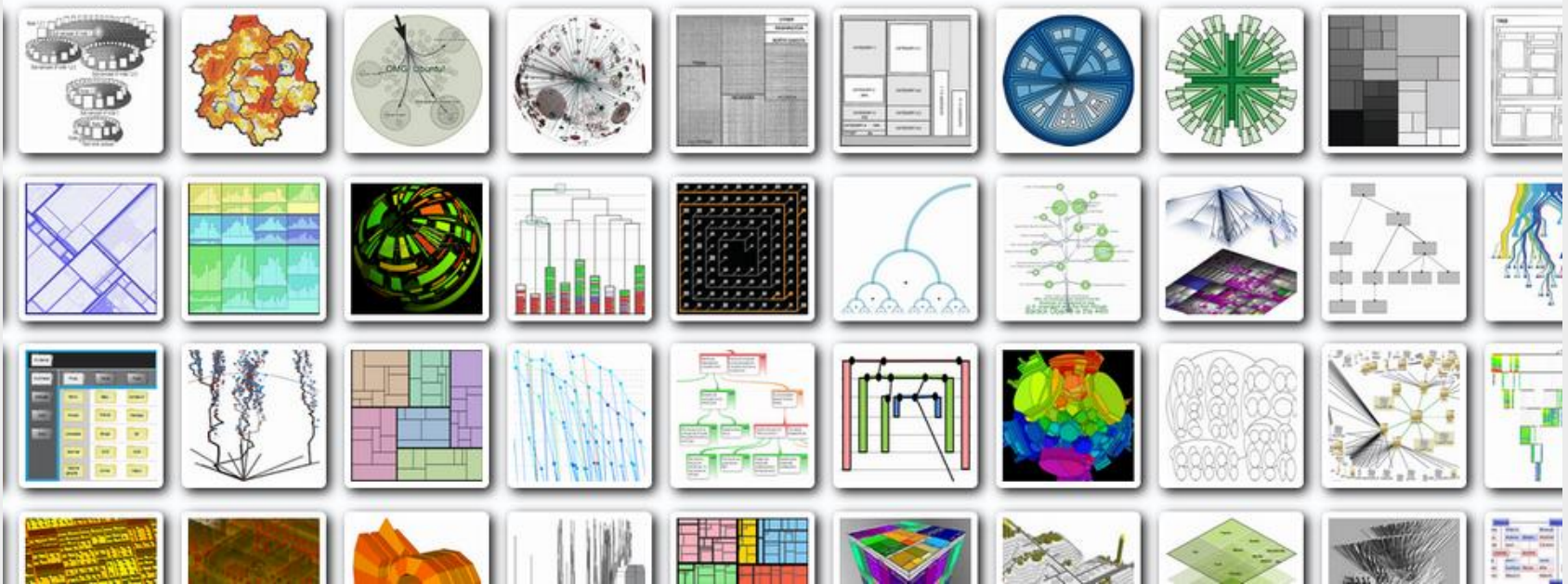treevis.net - A Visual Bibliography of Tree Visualization 2.0 by Hans-Jörg Schulz

Treemap

Shneiderman (1992)

Treemap

Shneiderman (1992)

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

Total (9)

Treemap

Shneiderman (1992)

A (3)	B (6)
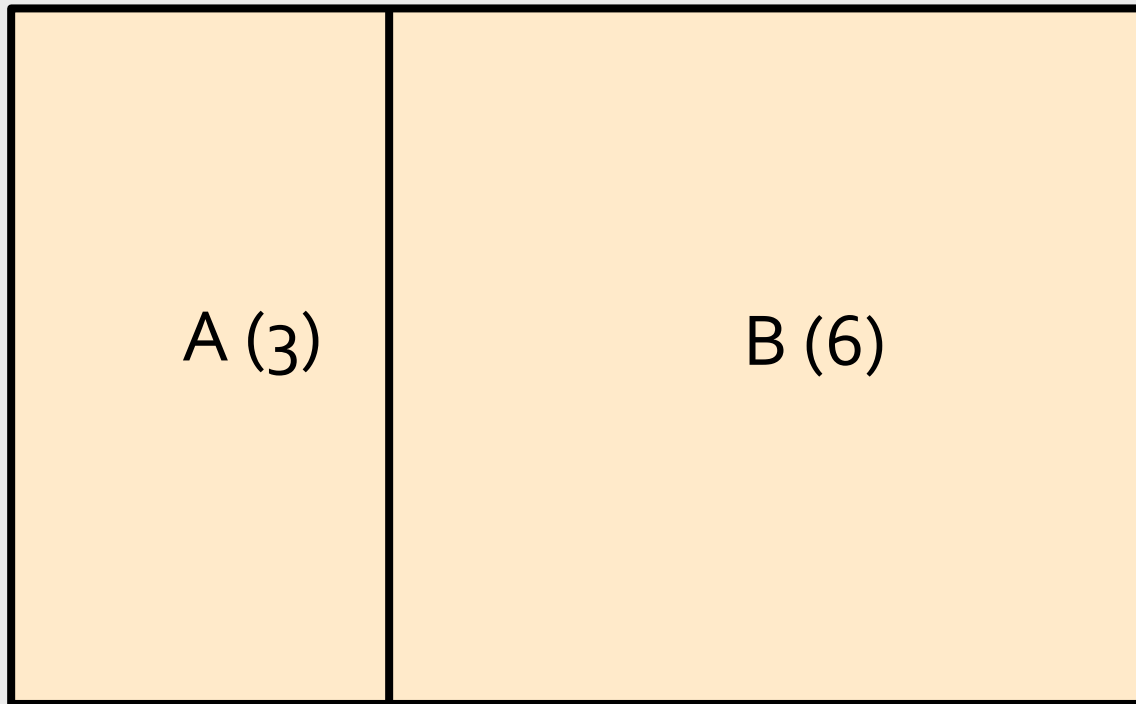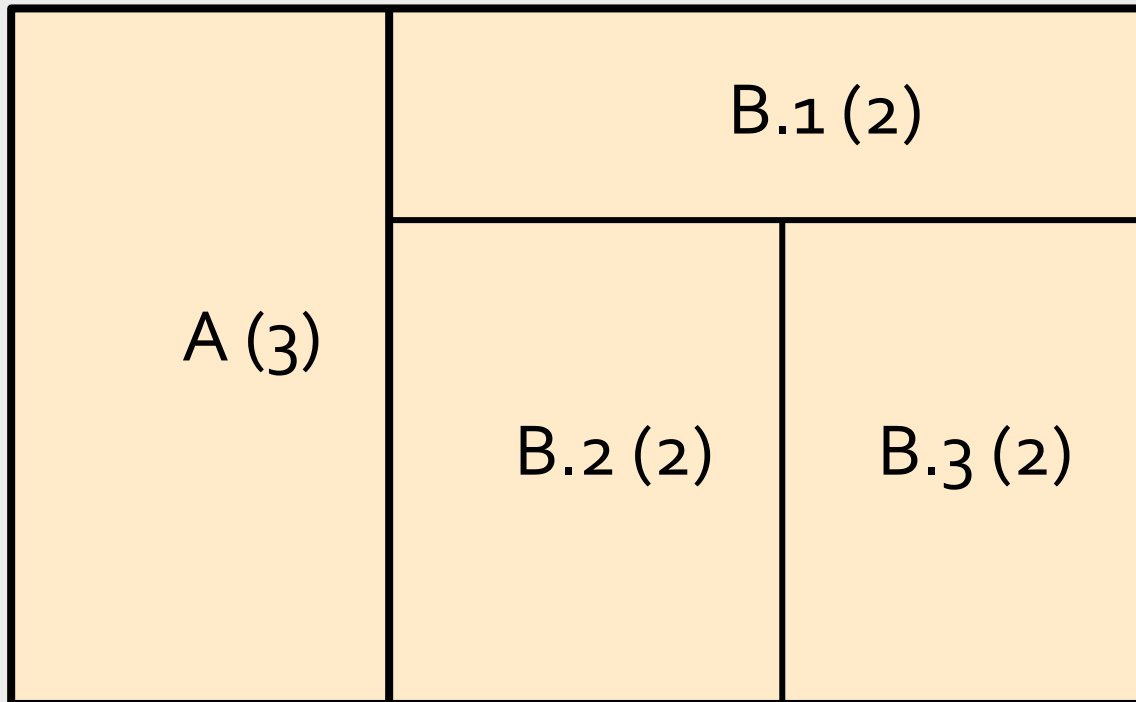
Class	Value
A	3
B.1	2
B.2	2
B.3.a	1
B.3.b	1

Treemap

Shneiderman (1992)

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

Treemap

Shneiderman (1992)
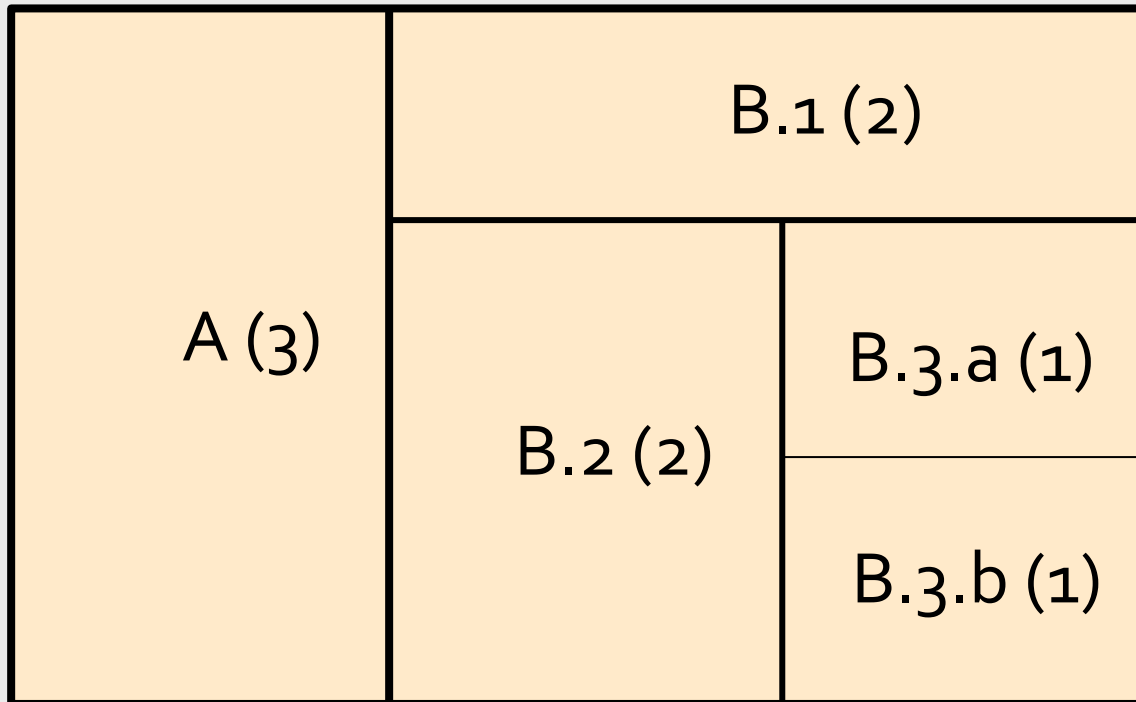
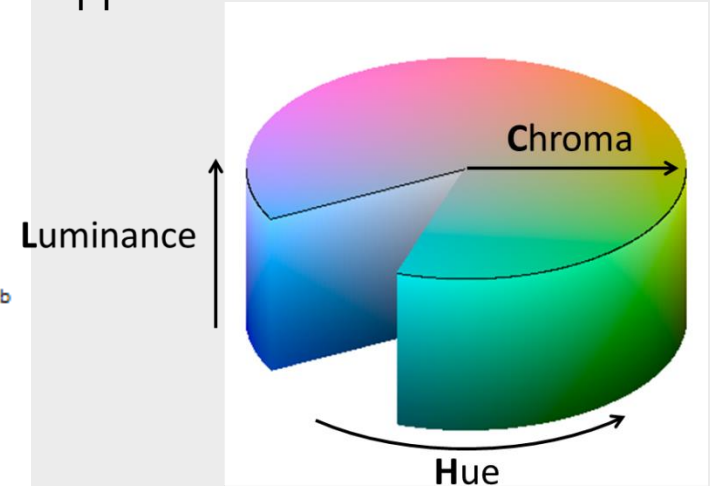| Class	Value
A	3
B.1	2
B.2	2
B.3.a	1
B.3.b	1

Treemap

Structural Business Statistics: aggregated by economic activity
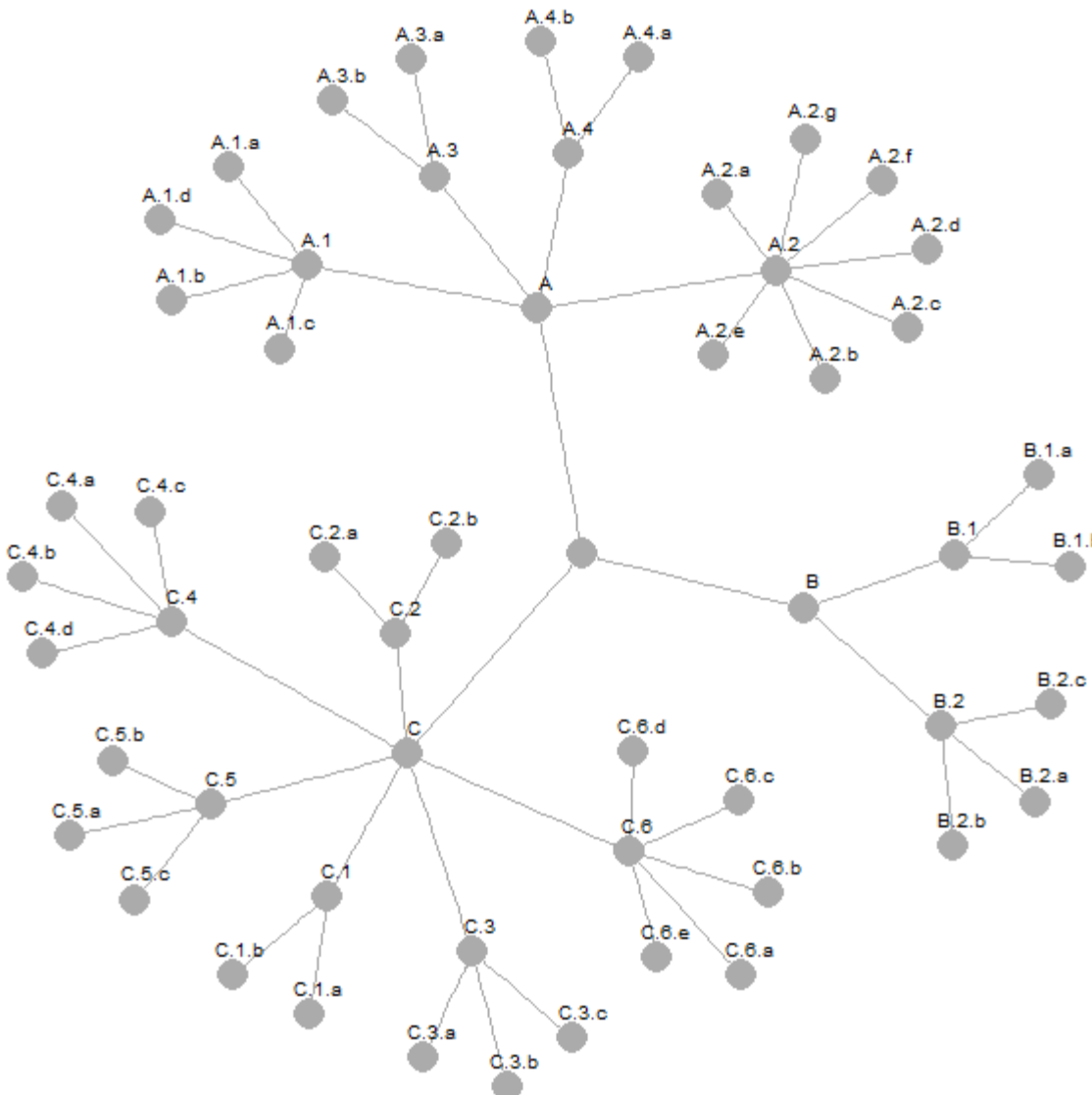
Tree Colors (Tennekes and De Jonge, 2014)

How to assign a color palette to a tree structure?

Approach:

- **Hue** resembles branches
- **Chroma** and **Luminance** discriminate hierarchical levels

Tree Colors

How to assign a color palette to a tree structure?

Approach:

- **Hue** resembles branches
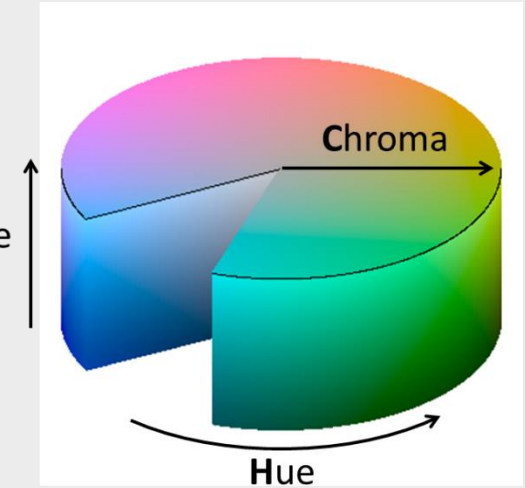- **Chroma** and **Luminance** discriminate hierarchical levels
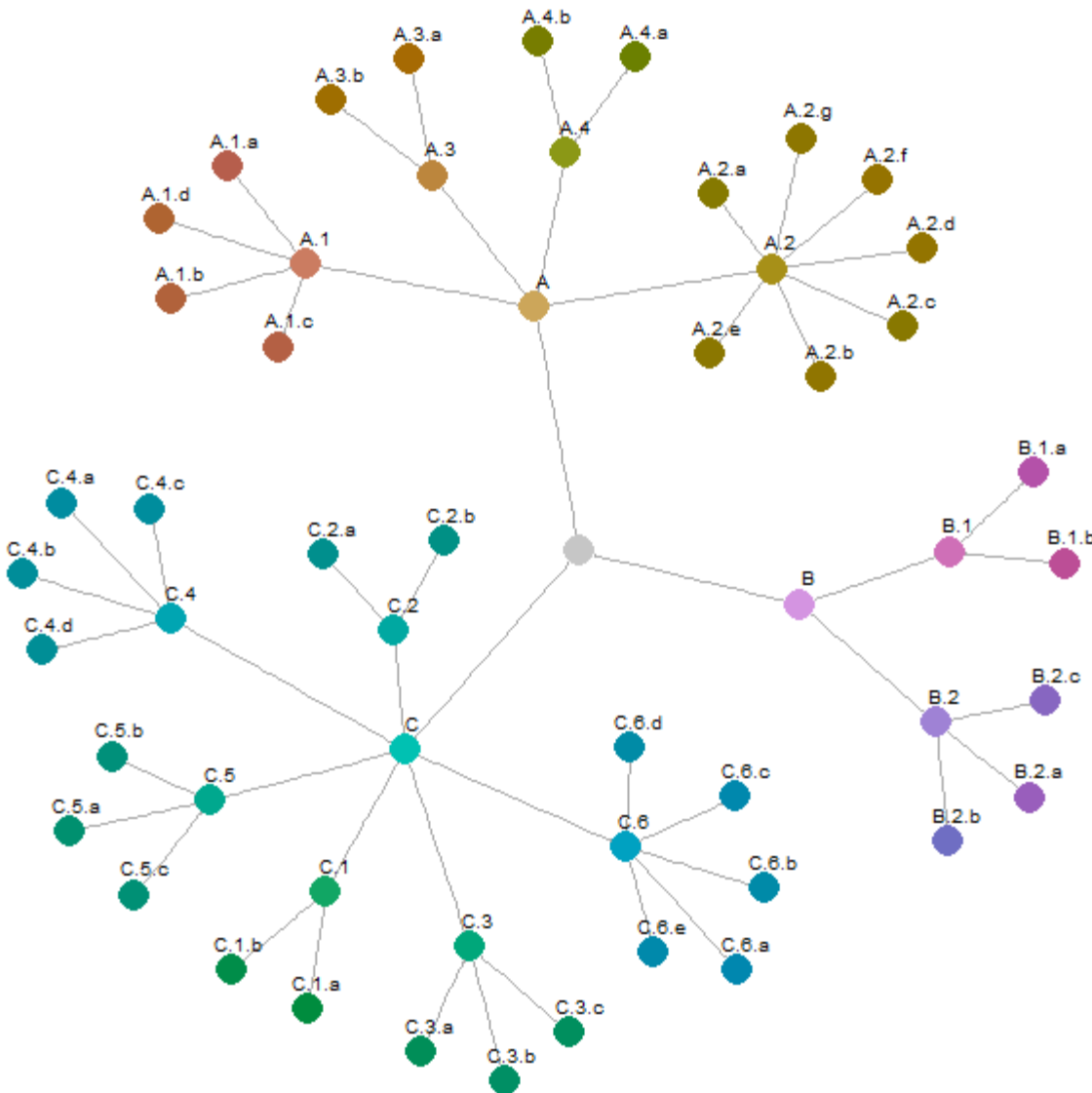
15

Tree Colors (Tennekes and De Jonge, 2014)
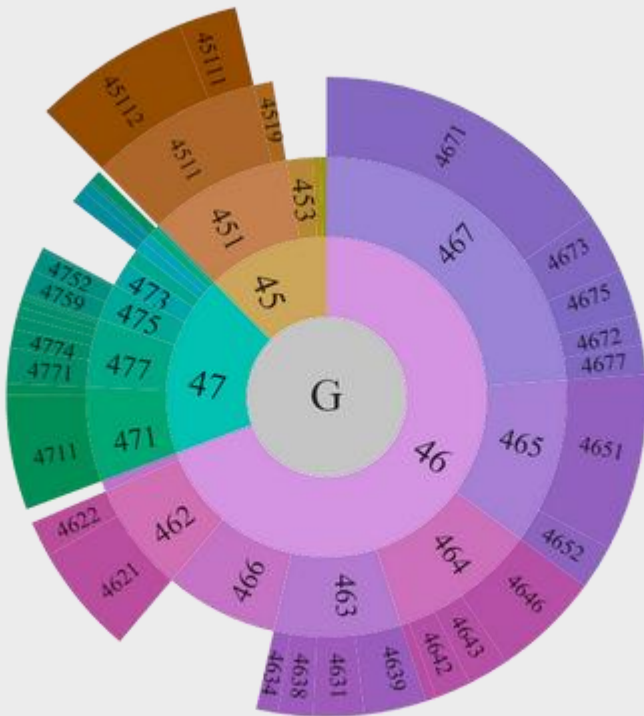
(a) Sunburst diagram

(b) Treemap

Net turnover for economic activity sector G, *Wholesale and retail tradec*

16

Large tabular data

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

	var1	var2	var3	var4	var5	var6	var7	var8
unit1								
unit2								
unit3								
…								
…								
unit10000								

Applications in Official Statistics:
- Large survey data
- Admin data
- Big data

17

Large tabular data

**Tableplot**: visual summary of a large data table

	var1 ↓	var2	var3	var4	var5	var6	var7	var8
unit1								
unit2								
unit3								
…								
…								
unit10000								

bin1 (brace spanning unit1 through the sixth row)

…
bin100

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,
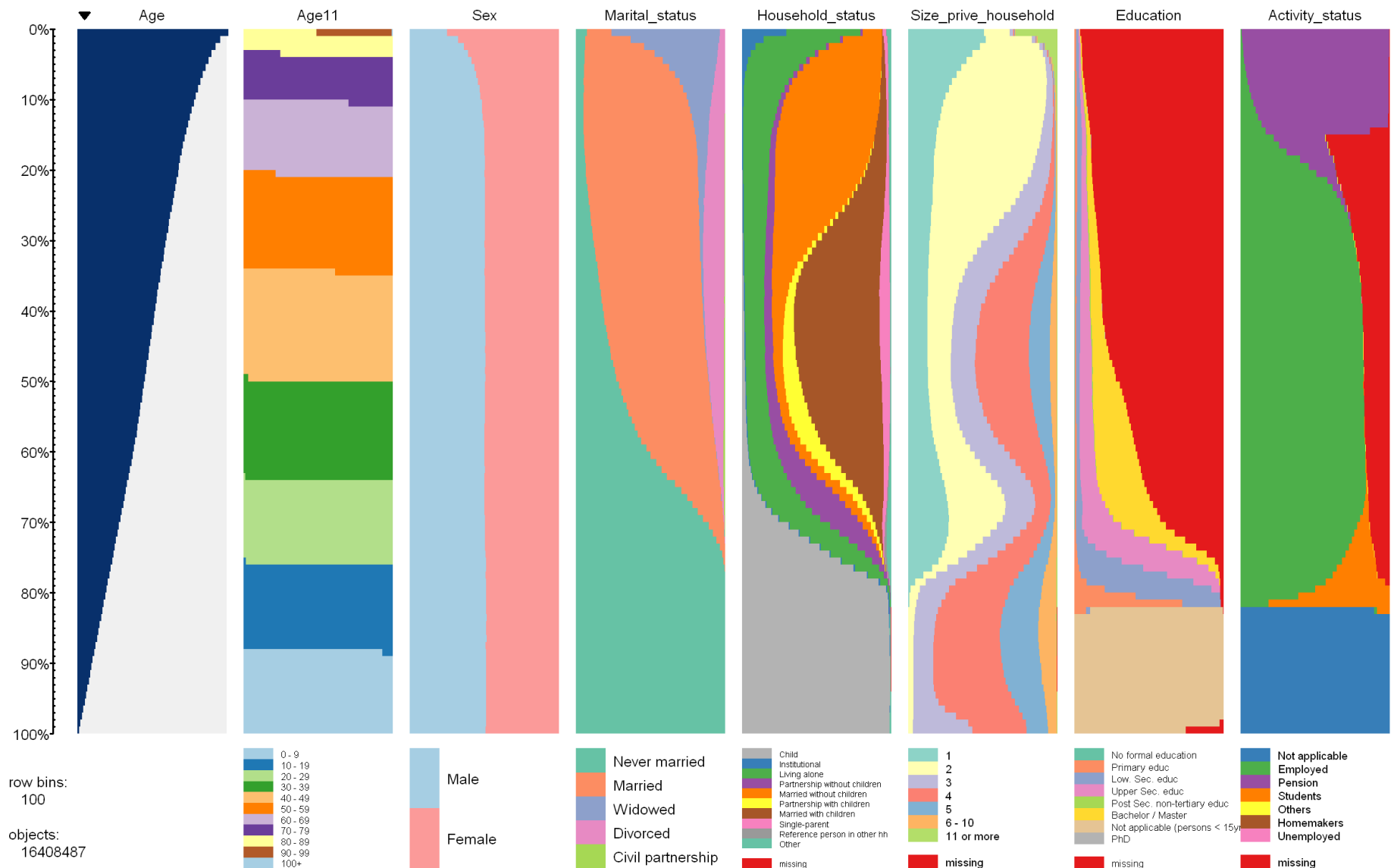   • for each categorical variable: calculate frequencies.
4. Plot it!   (see next slides…)

18
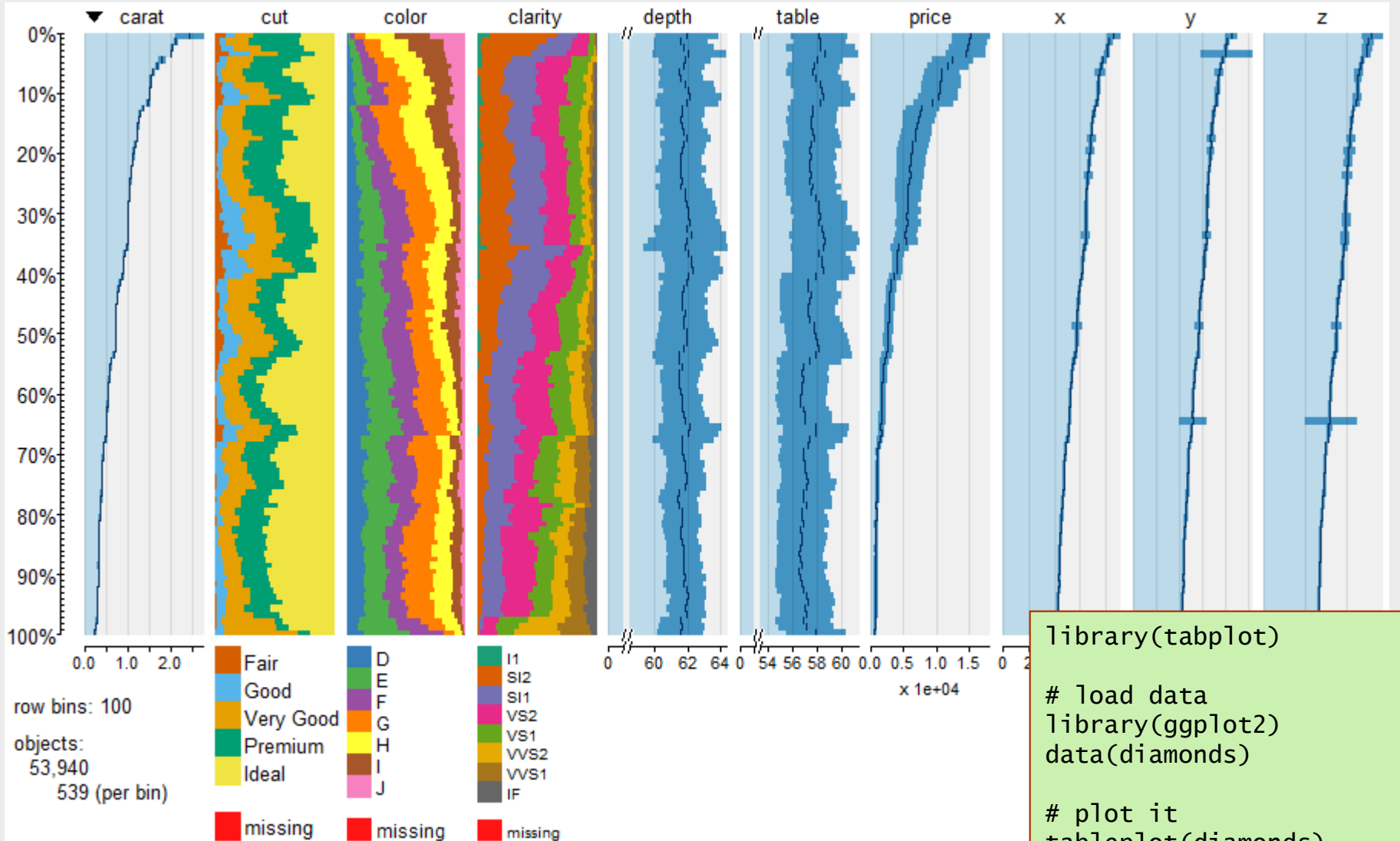
Tableplot

```
library(tabplot)

# load data
library(ggplot2)
data(diamonds)

# plot it
tableplot(diamonds)
```

Tableplot of the Dutch Virtual Census (test file, 2009)

Tableplot
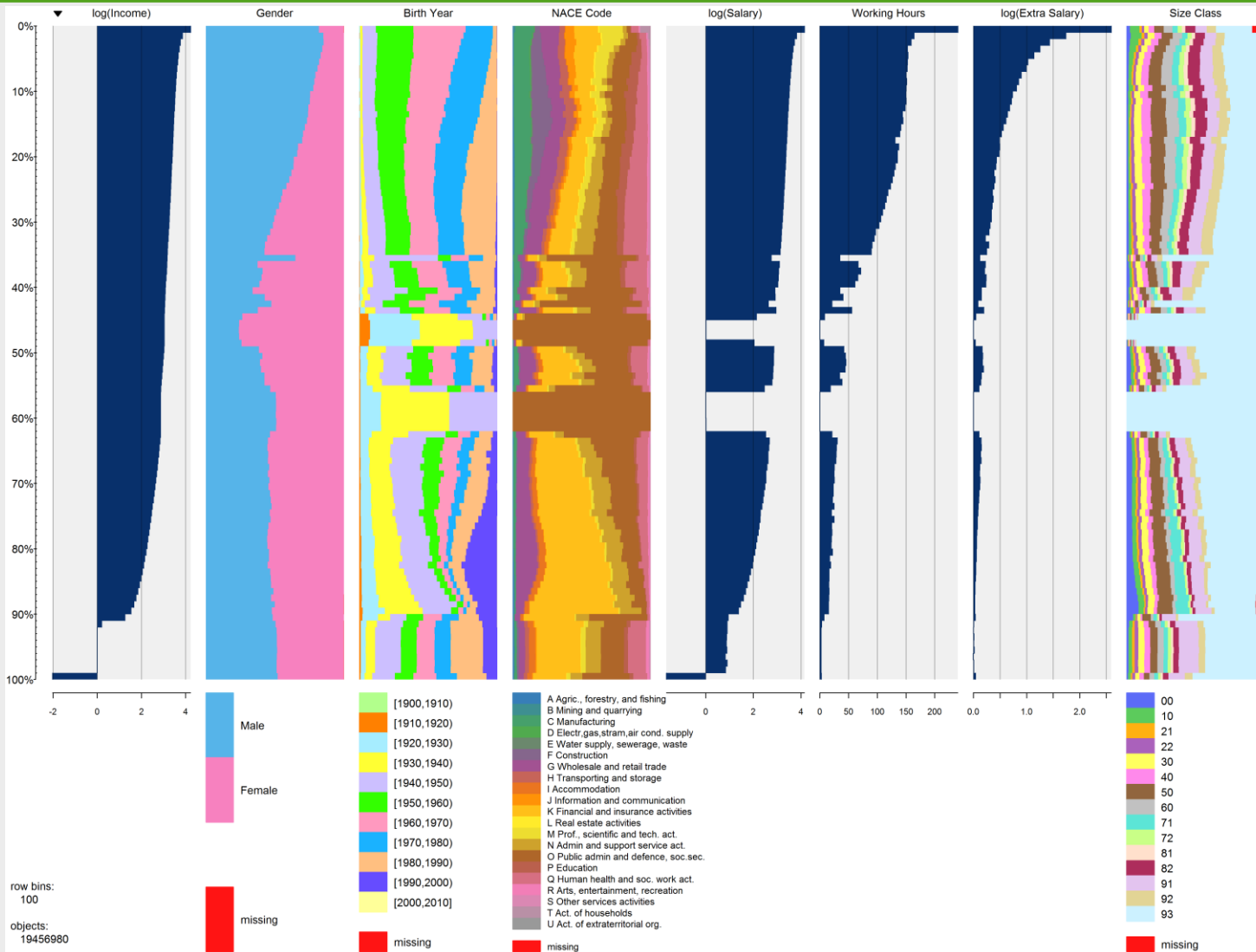
Tableplot of the Insurance Policy Record Administration (test file, October 2010)

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
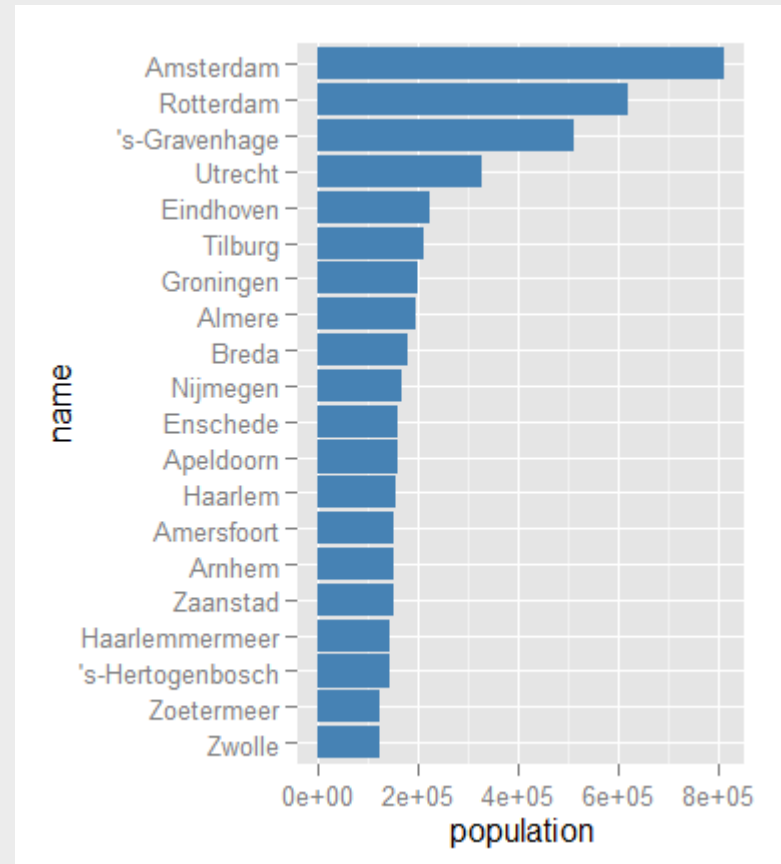
Geographic map   +   Theme   **23**

Thematic map
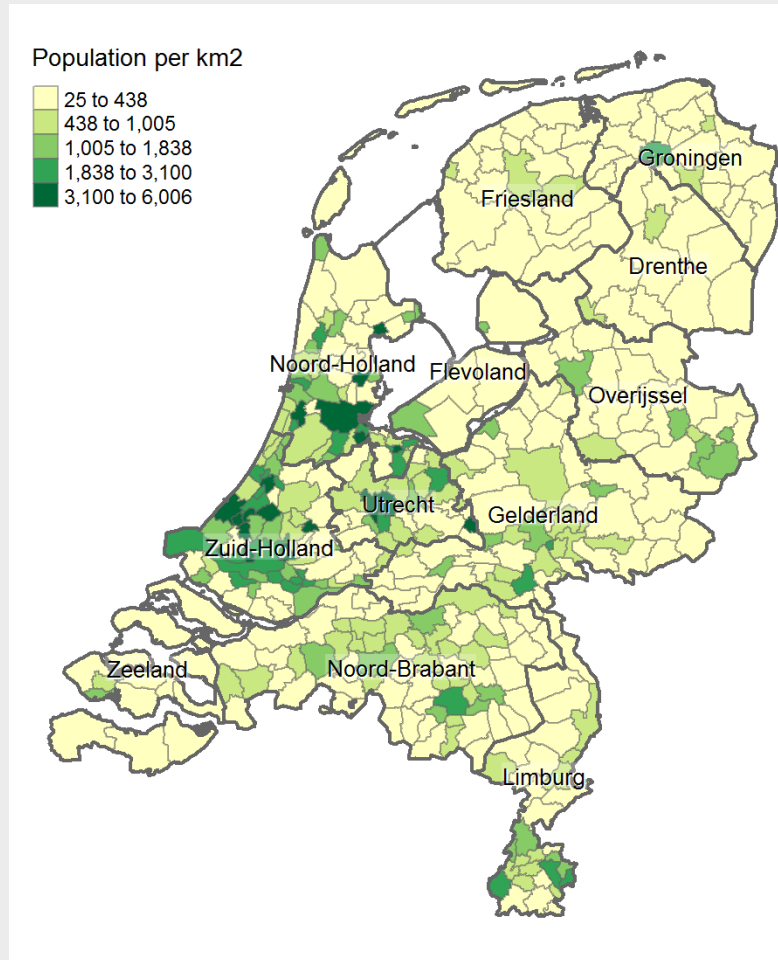
= Thematic map

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

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

```
tm_fill("blue")
```

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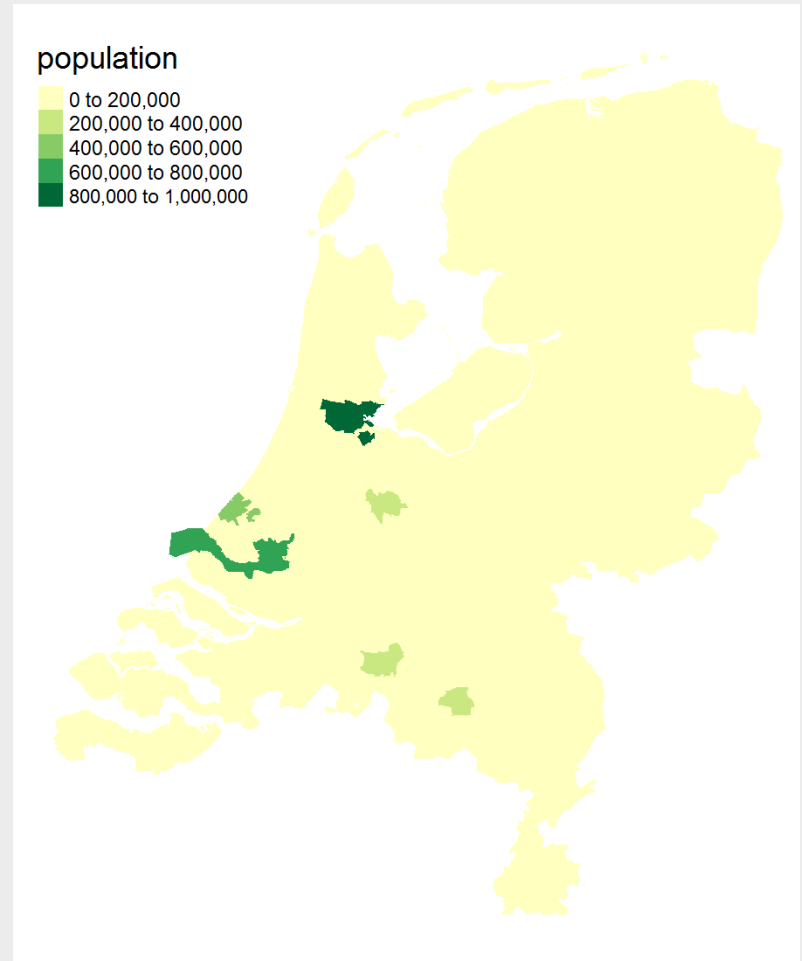
```
tm_shape(NLD_muni,
         projection="rd") +
```

```
tm_fill("population")
```


population

- 0 to 200,000
- 200,000 to 400,000
- 400,000 to 600,000
- 600,000 to 800,000
- 800,000 to 1,000,000

Building a thematic map

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

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


Population per km2

- 25 to 438
- 438 to 1,005
- 1,005 to 1,838
- 1,838 to 3,100
- 3,100 to 6,006

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) +
```


Population per km2

- 25 to 438
- 438 to 1,005
- 1,005 to 1,838
- 1,838 to 3,100
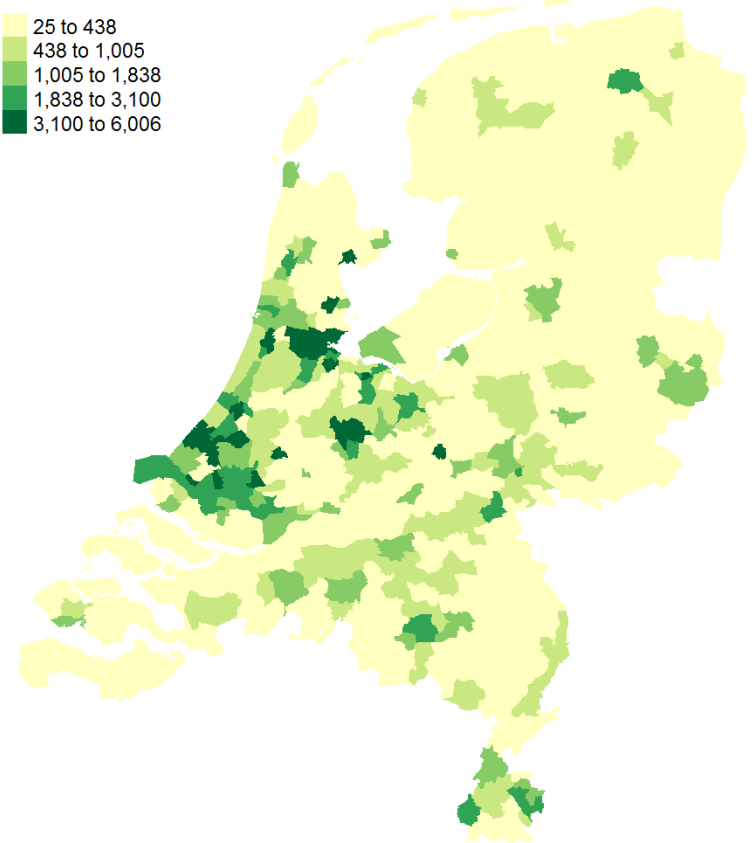- 3,100 to 6,006

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) +
```


Population per km2

- 25 to 438
- 438 to 1,005
- 1,005 to 1,838
- 1,838 to 3,100
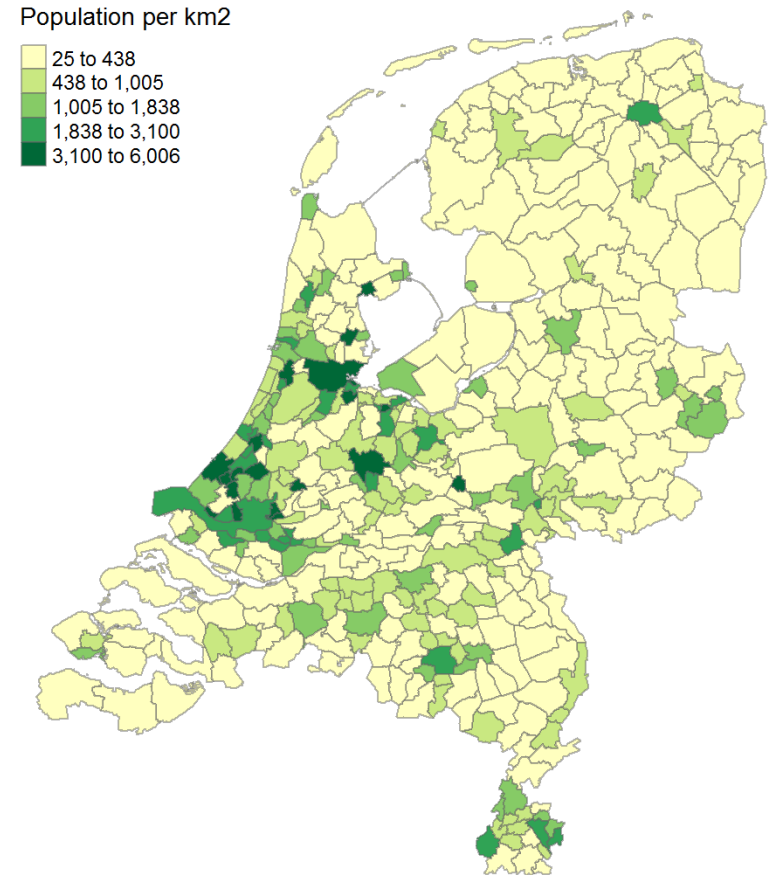- 3,100 to 6,006

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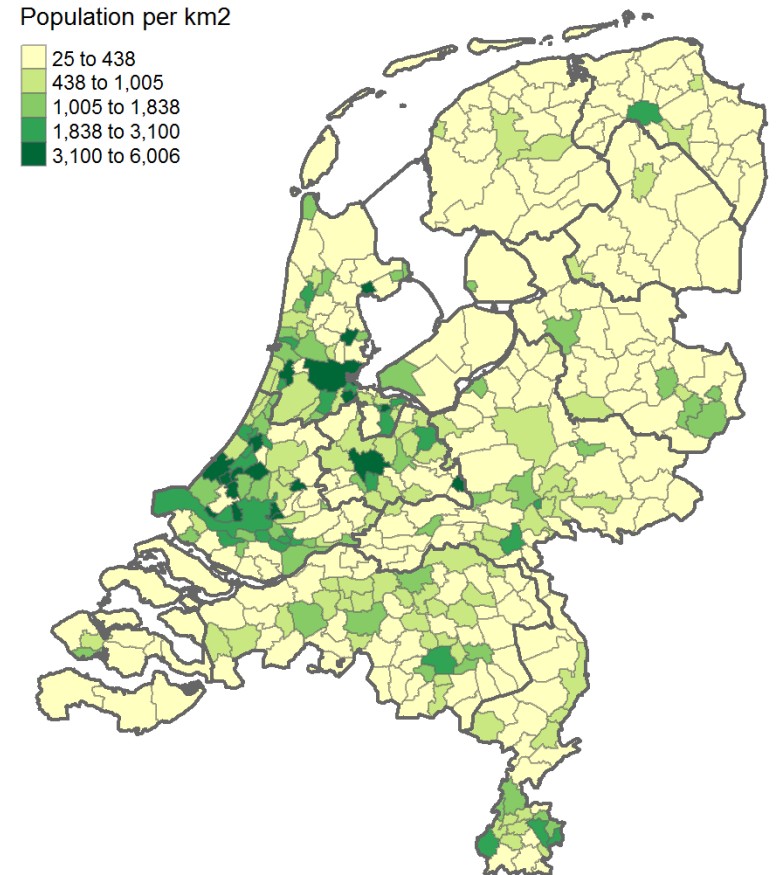
```
tm_shape(NLD_muni,
         projection="rd") +
```

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

```
tm_borders(alpha=.5) +
```
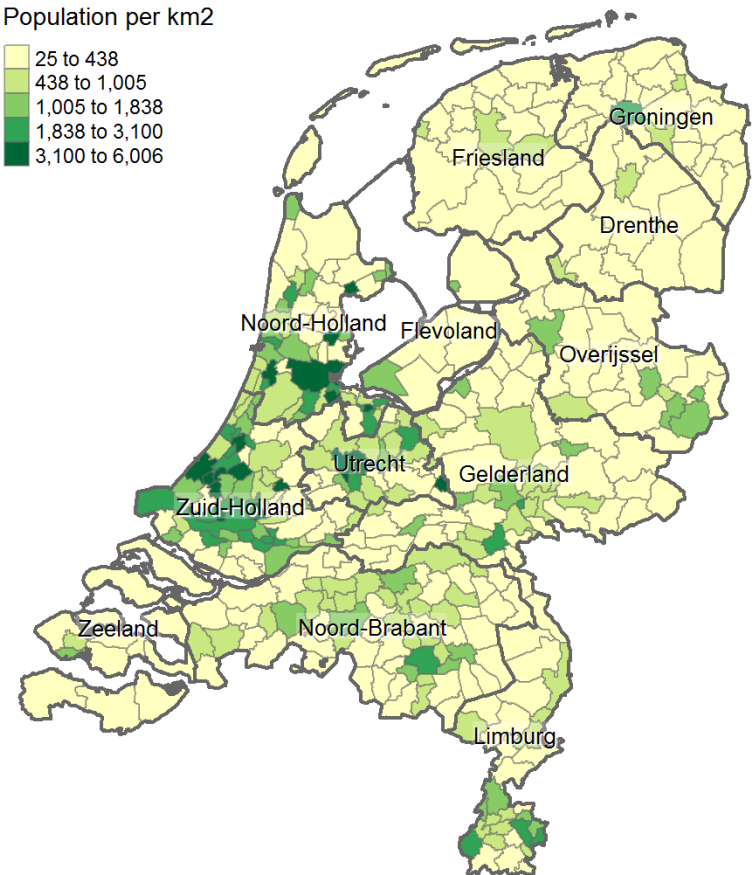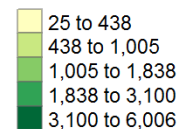
```
tm_shape(NLD_prov) +
```
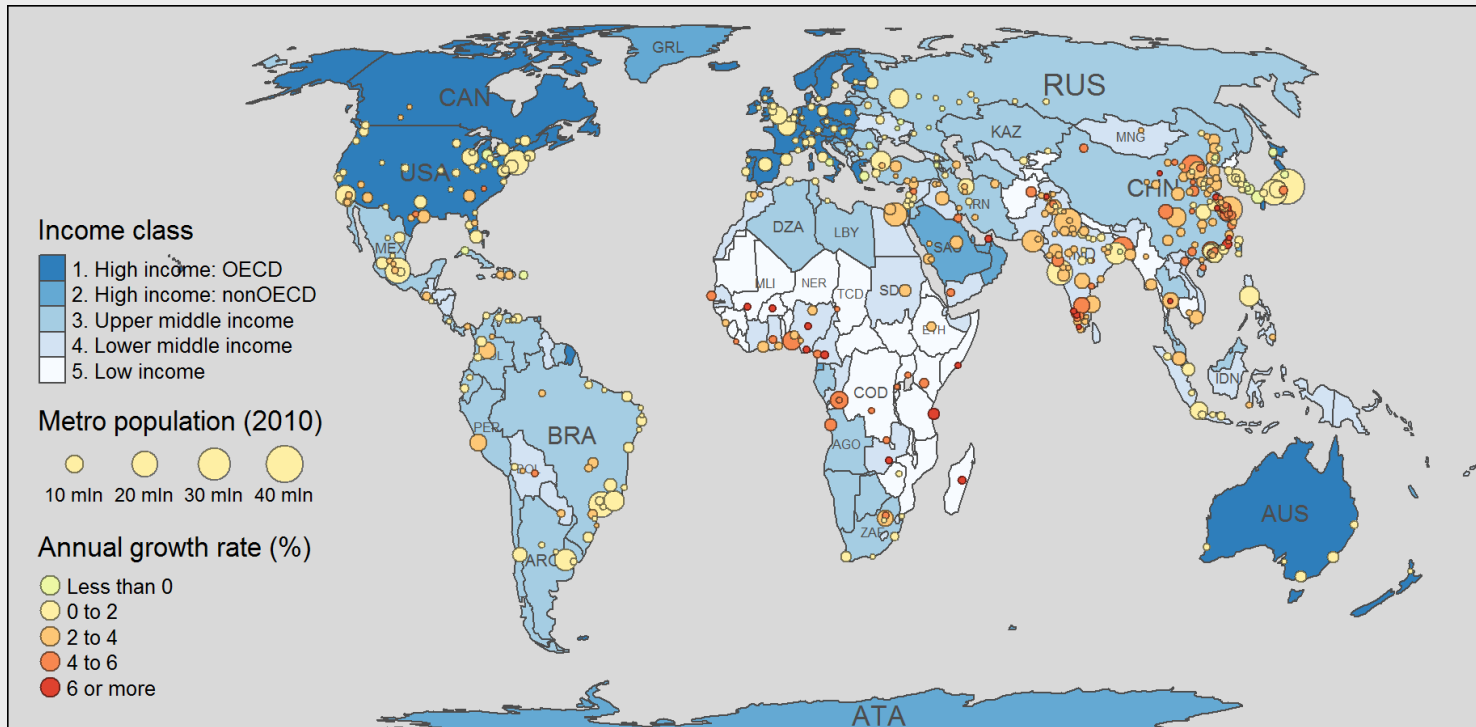
```
tm_borders(lwd=2) +
```

```
tm_text("name", size=.8, shadow=TRUE,
    bg.color="white", bg.alpha=.25)
```


Population per km2

- 25 to 438
- 438 to 1,005
- 1,005 to 1,838
- 1,838 to 3,100
- 3,100 to 6,006

Groningen
Friesland
Drenthe
Noord-Holland   Flevoland
Overijssel
Utrecht   Gelderland
Zuid-Holland
Zeeland   Noord-Brabant
Limburg

Choropleth + bubble map


```
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
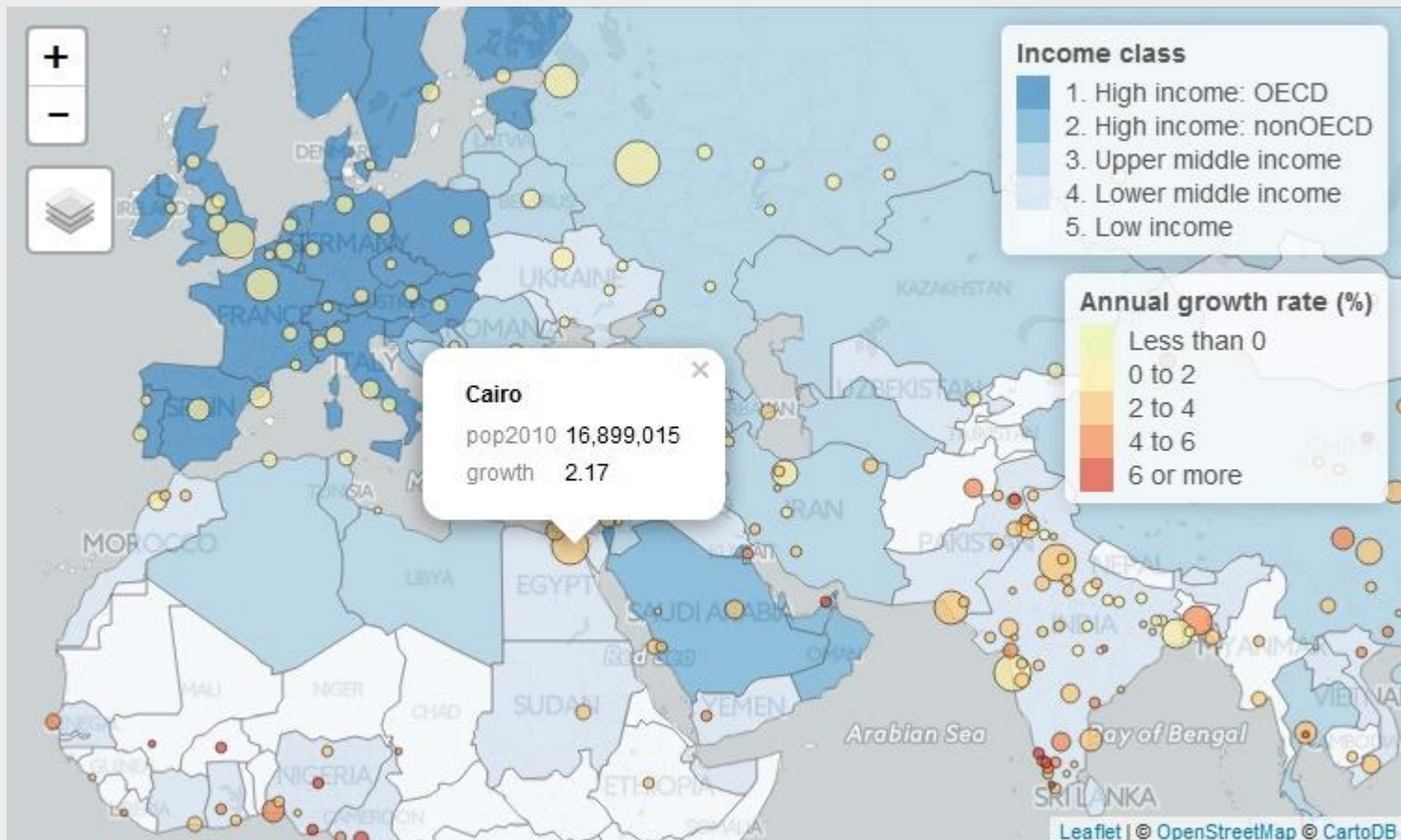

```
tmap_mode("view")
map1    # to which the previous plot has been assigned
```

35

US choropleth

2010 Adult Obesity by County, percent

Alaska

Hawaii

Data @ Unites States Department of Agriculture
Shape @ Unites States Census Bureau

	10 to 15
	15 to 20
	20 to 25
	25 to 30
	30 to 35
	35 to 40
	40 to 45
	45 to 50
	Missing

Crimes in Greater London
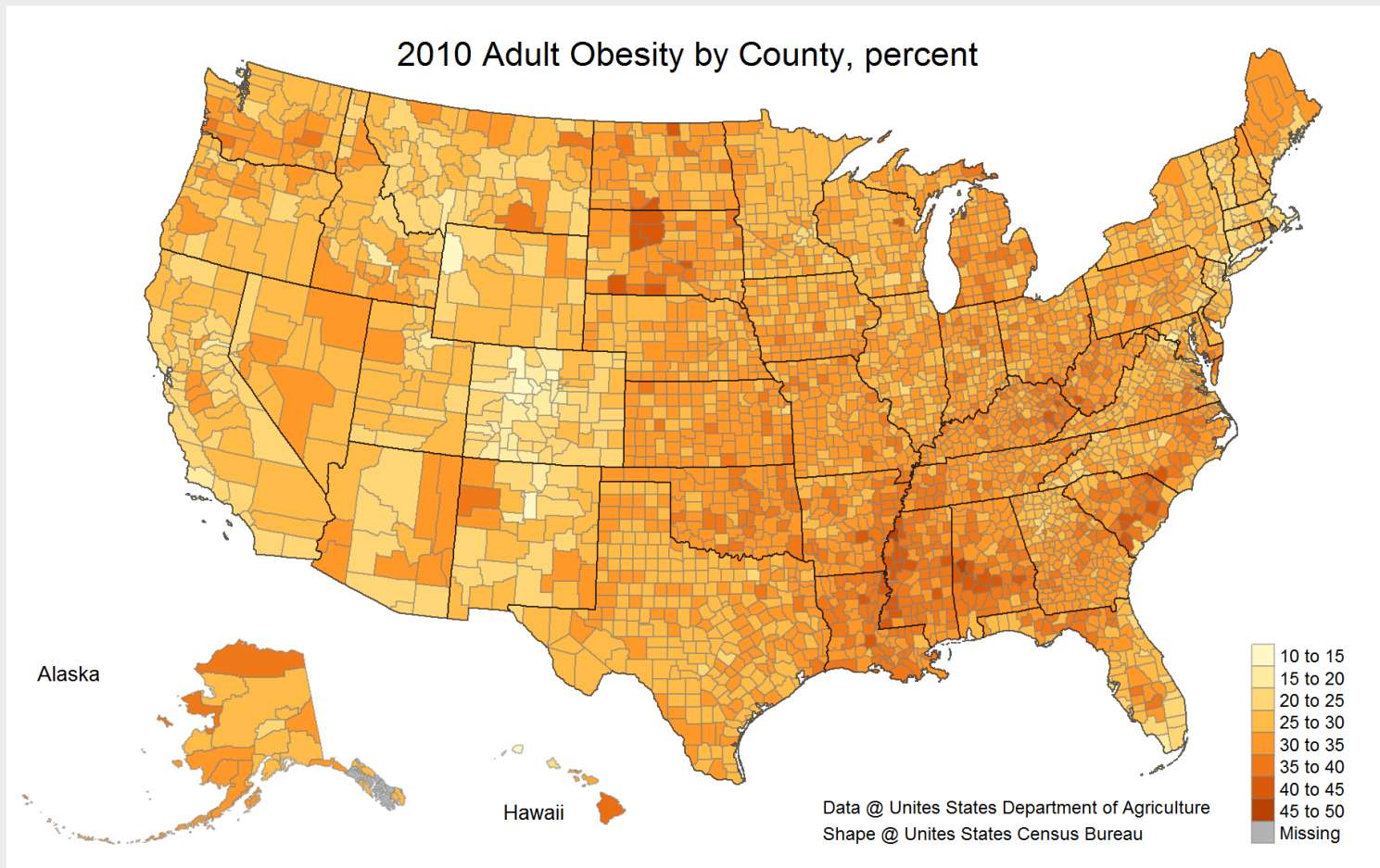
Crimes in Greater London
October 2015
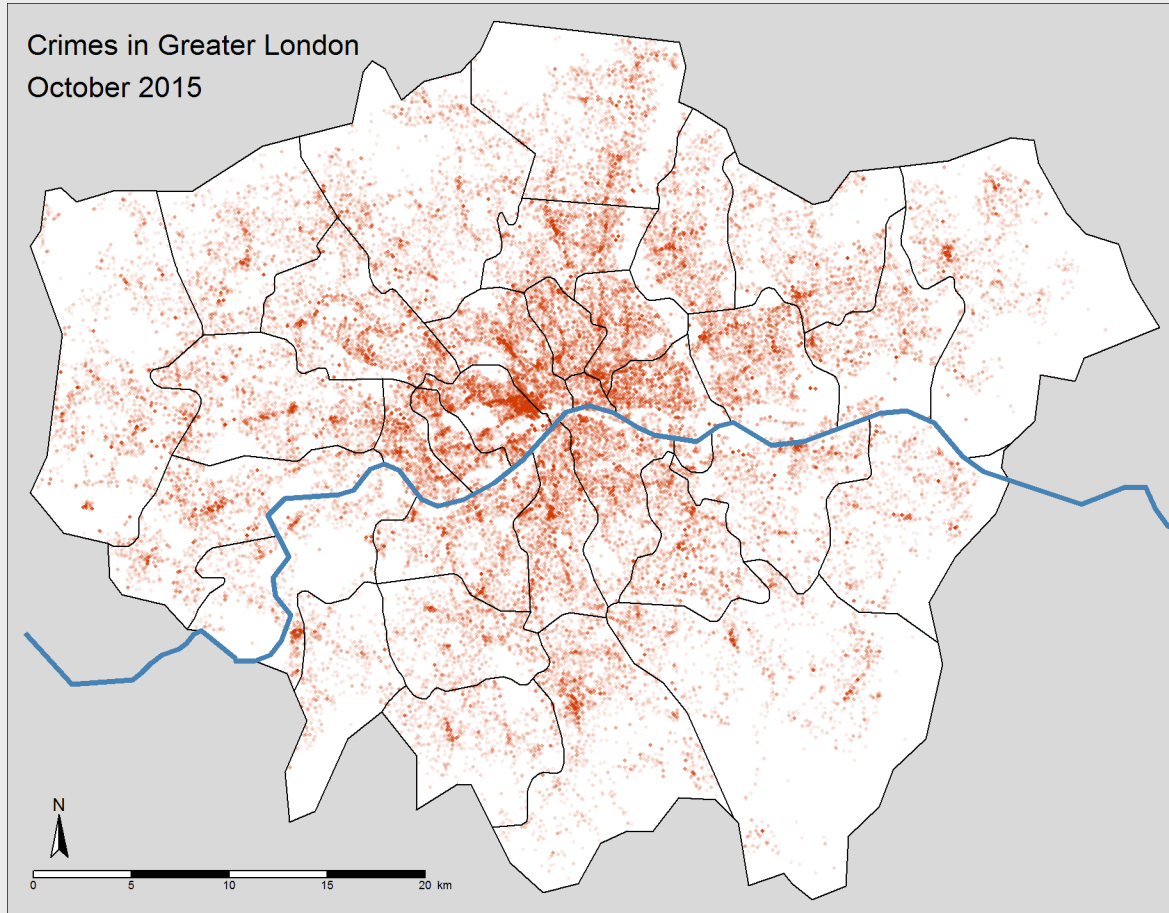
Dot map

Crimes in Greater London

Crimes in Greater London
October 2015

Crimes per km2
0 to 50
50 to 100
100 to 250
250 to 500
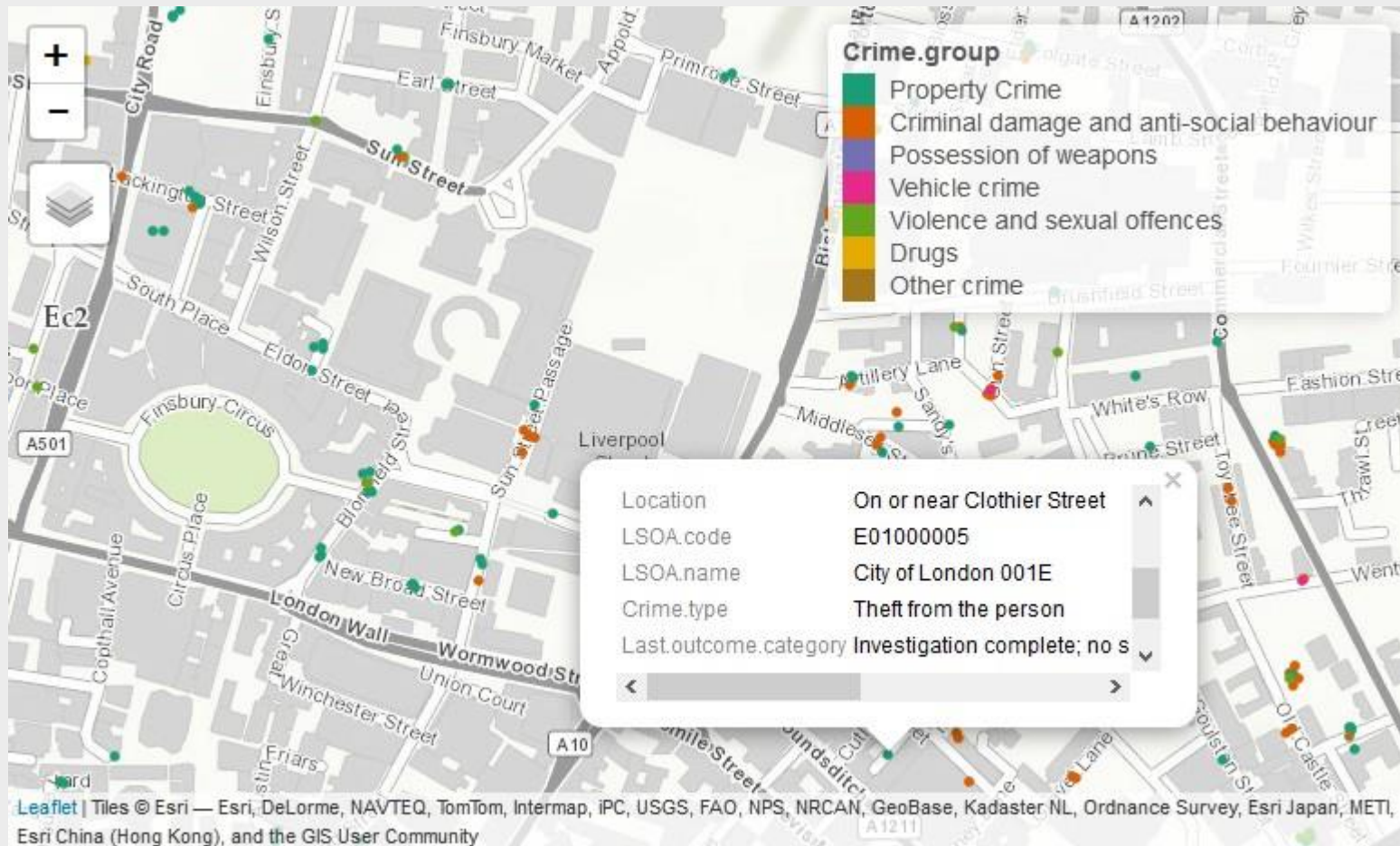500 to 1,000
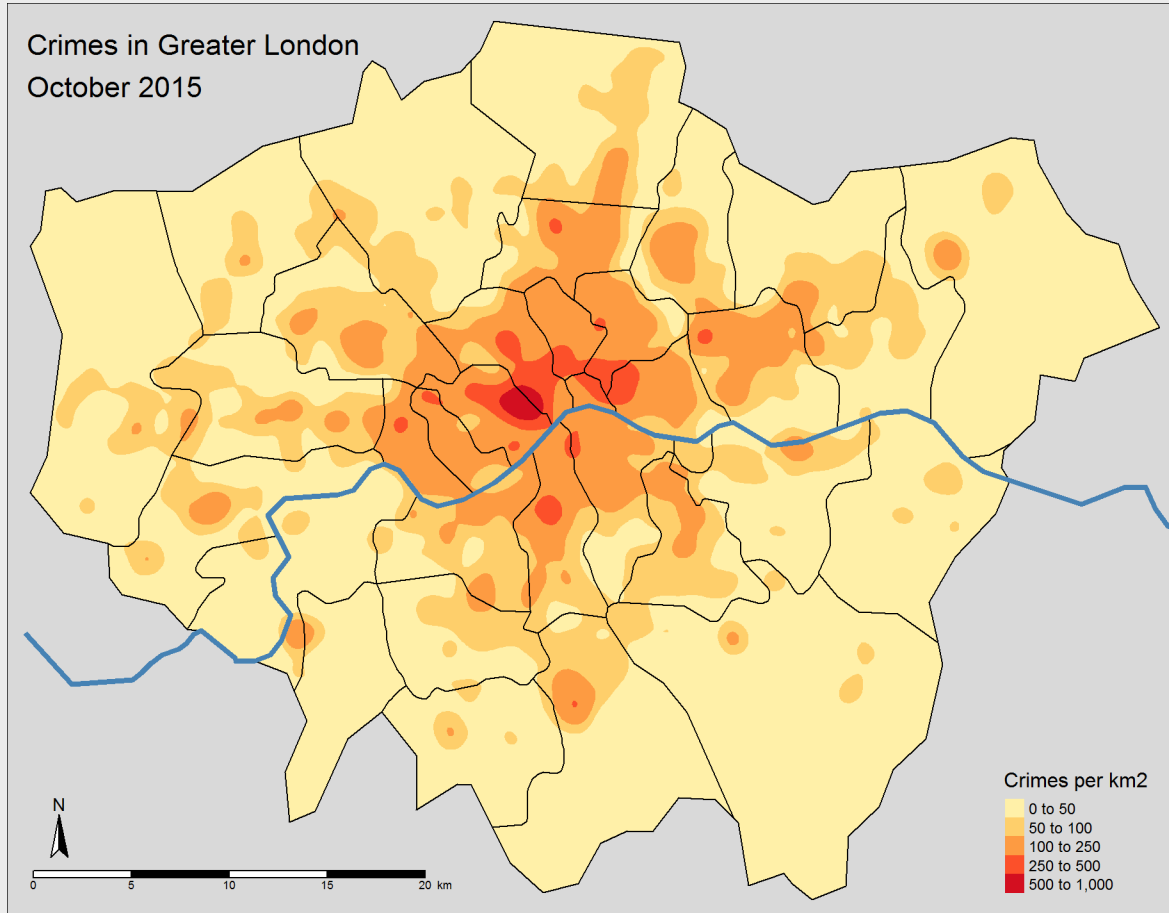
Dasymetric map

Crimes in the City of London

Small multiples

Crimes in the City of London

Interactive map

Interactive dot map

Prototype © Martijn Tennekes. Dots are evenly distributed across the residential area of each neighbourhood. Demographic data, Land use data   Leaflet | © CBS, labels: © CartoDB

Native Dutch

Prototype: http://research.cbs.nl/ColorDotMap
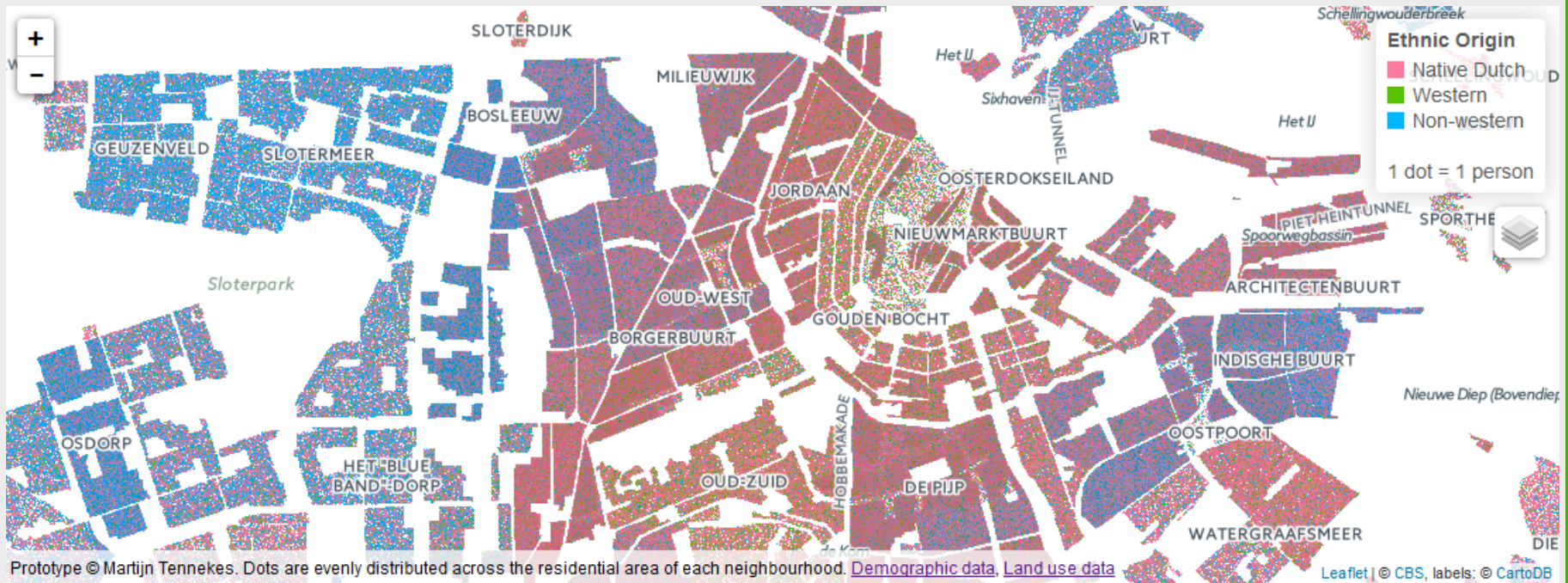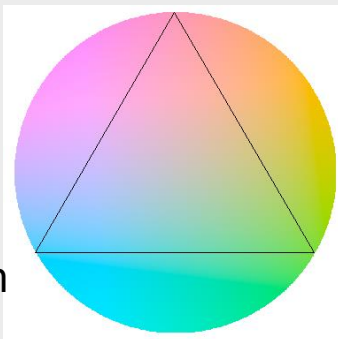
Non-western   Western

41

References

treemap
– CRAN version 2.4-1
– https://github.com/mtennekes/treemap
– Tennekes, M. Jonge, E. de (2014) Tree Colors: Color Schemes for Tree-Structured Data. IEEE Transactions on Visualization and Computer Graphics 20 (12), 2072 – 2081.
– Tennekes, M., Jonge, E. de (2011) Top-down data analysis with treemaps. Proceedings of the International Conference on Information Visualization Theory and Applications, IVAPP 2011, Algarve, Portugal.

tabplot
– CRAN version 1.3
– https://github.com/mtennekes/treemap
– Tennekes, M., Jonge, E. de (2013) On the exploration of high cardinality categorical data. Paper presented at the 2013 New Techniques and Technologies for Statistics (NTTS) conference, Brussels, Belgium.
– Tennekes, M., Jonge, E. de, Daas, P.J.H. (2013) Visualizing and Inspecting Large Datasets with Tableplots, Journal of Data Science 11 (1), 43-58.
– Tennekes, M., Jonge, E. de, Daas, P.J.H. (2011) Visual profiling of large statistical datasets. Paper presented at the 2011 New Techniques and Technologies for Statistics conference, Brussels, Belgium.

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