

# Innovative Visual Tools for Data Editing

Martijn Tennekes  
Edwin de Jonge, Piet Daas

24-26 September 2012



Funded under Socio-economic Sciences & Humanities



# Introduction

Visual tools for data editing

# Introduction

Visual tools for data editing

## Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- ...

# Introduction

Visual tools for data editing

## Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- ...

## Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data

# Introduction

Visual tools for data editing

## Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- ...

## Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data

# Introduction

Visual tools for data editing

## Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- ...

## Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data

# Introduction

Visual tools for data editing

## Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- ...

## Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data

# Introduction

Visual tools for data editing

## Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- ...

## Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data

## Novel tools

- Treemap
- Tableplot



# Introduction

## Visual tools for data editing

### Traditional tools

- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- . . .

### Novel tools

- Treemap
- Tableplot

### Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data



# Introduction

## Visual tools for data editing

### Traditional tools

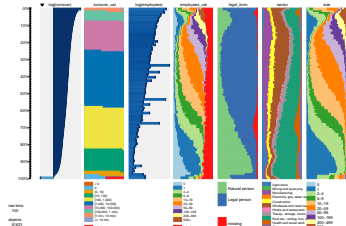
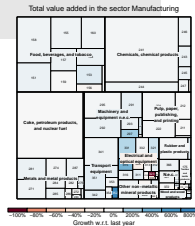
- Spreadsheets
- Frequency and cross tables
- Bar charts
- Scatter plots
- . . .

### Not always suitable for

- Hierarchical data
- Multivariate data
- BIG data

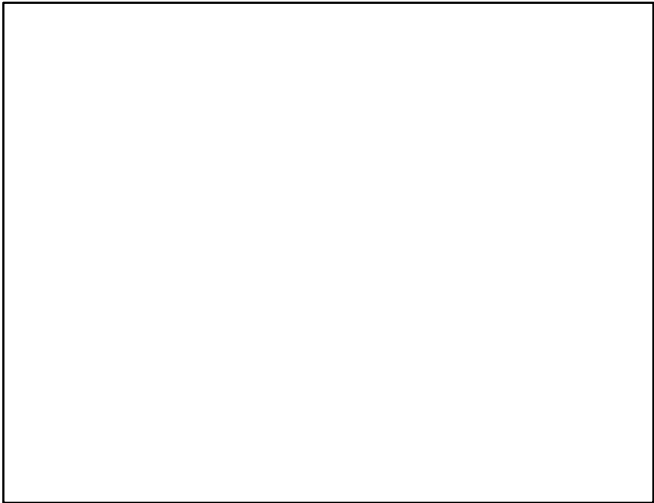
### Novel tools

- Treemap
- Tableplot



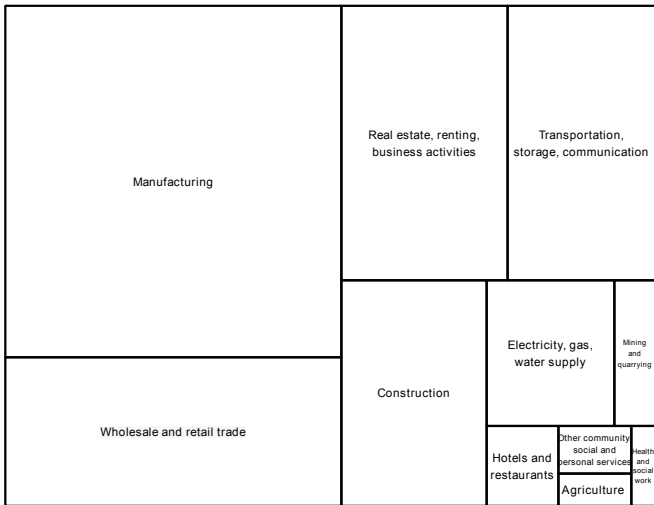
# Treemap (SBS data)

Total value added



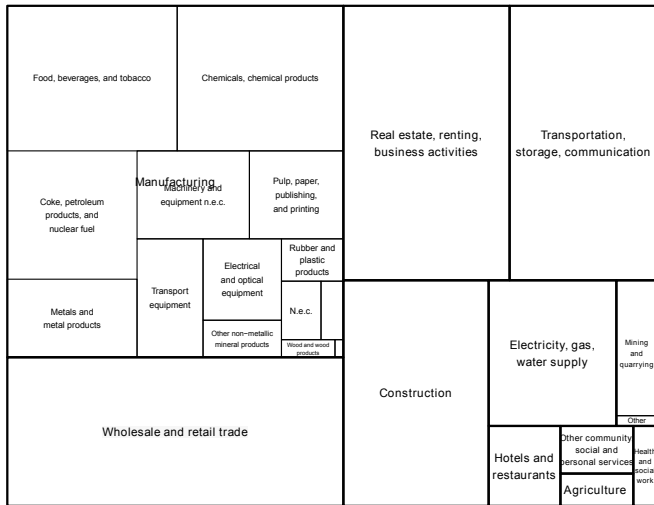
# Treemap (SBS data)

Total value added



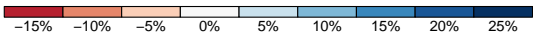
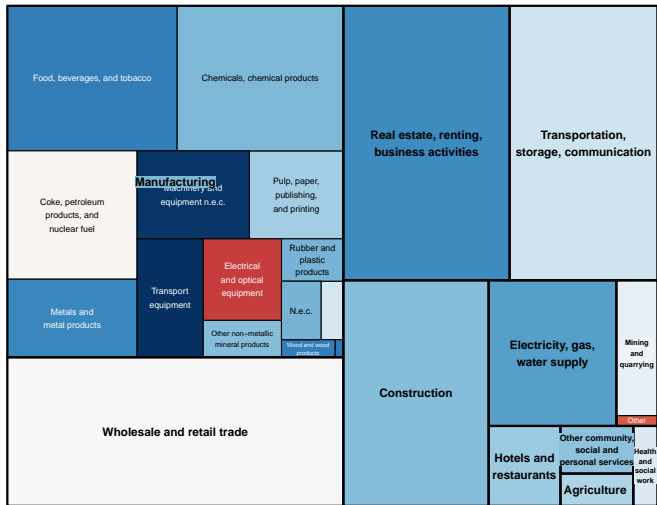
# Treemap (SBS data)

Total value added



# Treemap (SBS data)

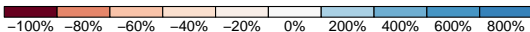
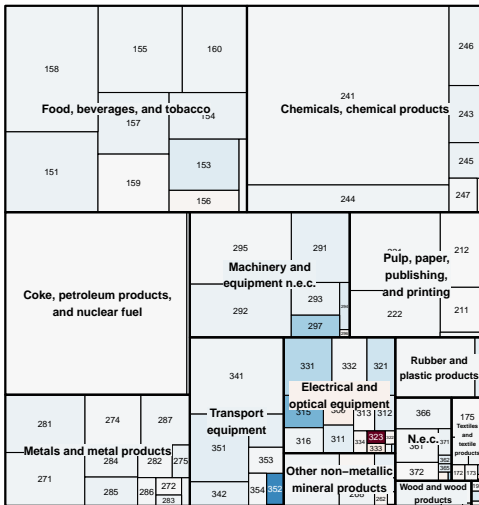
Total value added



Growth w.r.t. last year

# Treemap: zooming in

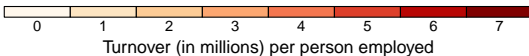
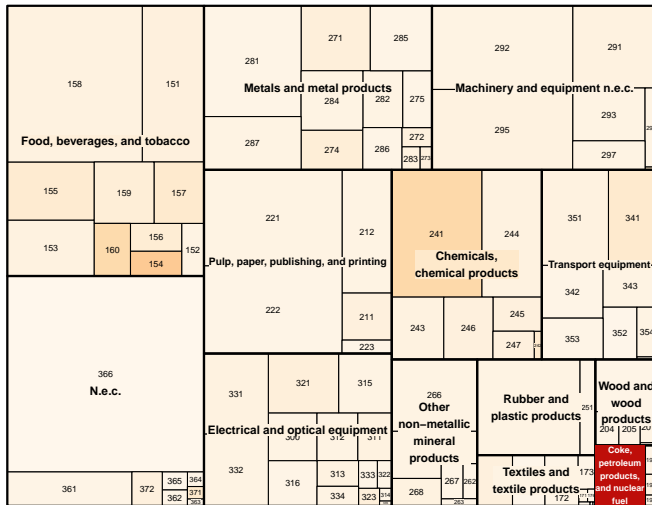
Total value added in the sector Manufacturing



Growth w.r.t. last year

# Treemap: densities

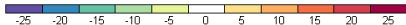
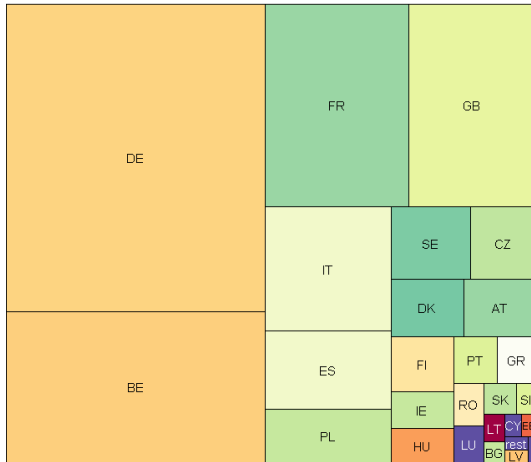
Number of persons employed in the sector Manufacturing





# Treemap: trade statistics

Export from NL to EU countries



Relative difference in % with import XX from NL

# Treemap

- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
  - Package *treemap*
  - Interactive use is very limited
  - Can be used with other tools

# Treemap

- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
  - Package *treemap*
  - Interactive use is very limited
  - Can be used with other tools

# Treemap

- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
  - Package *treemap*
  - Interactive use is very limited
  - Can be used with other tools

# Treemap

- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
  - Package *treemap*
  - Interactive use is very limited
  - Can be used with other tools

# Treemap

- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
  - Package *treemap*
  - Interactive use is very limited
  - Can be used with other tools

# Treemap

- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
  - Package *treemap*
  - Interactive use is very limited
  - Can be used with other tools

# Tableplot: SBS data (unprocessed)

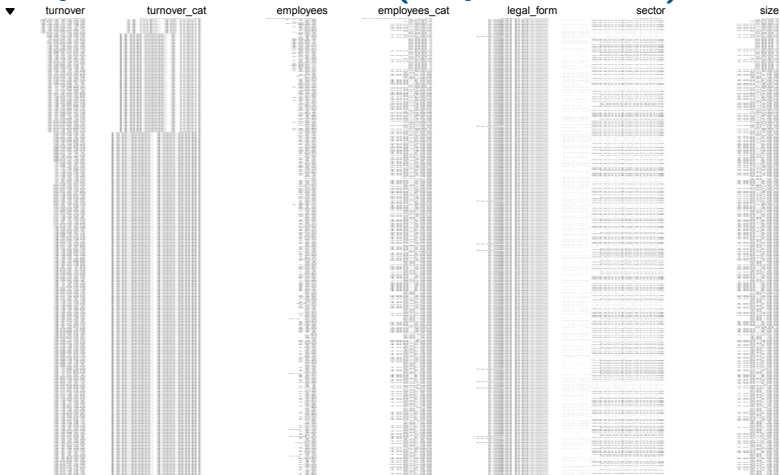
turnover	turnover_cat	employees	employees_cat	legal_form	sector	size
1758	[1,000, 10,000)	17	10-19	Legal person	Real est., renting, bus. act.	20-49
4	[10, 100)	3	2-4	Natural person	Wholesale and retail trade	2-4
84	[10, 100)	1	1	Natural person	Manufacturing	2-4
11778	[10,000, 100,000)	69	50-99	Legal person	Wholesale and retail trade	50-99
10	[10, 100)	7	5-9	Legal person	Health and social work	5-9
438	[100, 1,000)	4	5-9	Legal person	Wholesale and retail trade	2-4
134	[100, 1,000)	3	1	Natural person	Manufacturing	2-4
1394	[1,000, 10,000)	20	10-19	Legal person	Real est., renting, bus. act.	10-19
164	[100, 1,000)	3	2-4	Natural person	Real est., renting, bus. act.	2-4
21700	[10,000, 100,000)	76	50-99	Natural person	Wholesale and retail trade	50-99
872	[100, 1,000)	14	10-19	Legal person	Manufacturing	10-19
158	[100, 1,000)	4	2-4	Legal person	Real est., renting, bus. act.	2-4
57120	[10,000, 100,000)	211	200-499	Legal person	Manufacturing	100-199
9	[10, 100)	0	1	Natural person	Wholesale and retail trade	1
99296	[10,000, 100,000)	127	100-199	Legal person	Manufacturing	50-99
695	[100, 1,000)	5	2-4	Natural person	Real est., renting, bus. act.	2-4
55618	[10,000, 100,000)	362	200-499	Legal person	Manufacturing	200-499
1113	[1,000, 10,000)	17	10-19	Legal person	Hotels and restaurants	10-19
7478	[1,000, 10,000)	3	5-9	Legal person	Wholesale and retail trade	10-19
8662	[1,000, 10,000)	75	50-99	Legal person	Transp., storage, comm.	100-199
8633	[1,000, 10,000)	36.08	20-49	Legal person	Manufacturing	20-49
1028	[1,000, 10,000)	1	5-9	Legal person	Wholesale and retail trade	20-49
183	[100, 1,000)	0.2	2-4	Natural person	Wholesale and retail trade	2-4
25	[10, 100)	1.6	1	Natural person	Real est., renting, bus. act.	1
7182	[1,000, 10,000)	82	50-99	Legal person	Construction	10-19
2005	[1,000, 10,000)	3	0	Legal person	Real est., renting, bus. act.	20-49
10398	[10,000, 100,000)	8	10-19	Legal person	Real est., renting, bus. act.	100-199
2245	[1,000, 10,000)	34	20-49	Legal person	Real est., renting, bus. act.	20-49
4903	[1,000, 10,000)	18	10-19	Legal person	Wholesale and retail trade	20-49
68	[10, 100)	-1	1	Legal person	Real est., renting, bus. act.	2-4
...	...	...	...	...	...	...



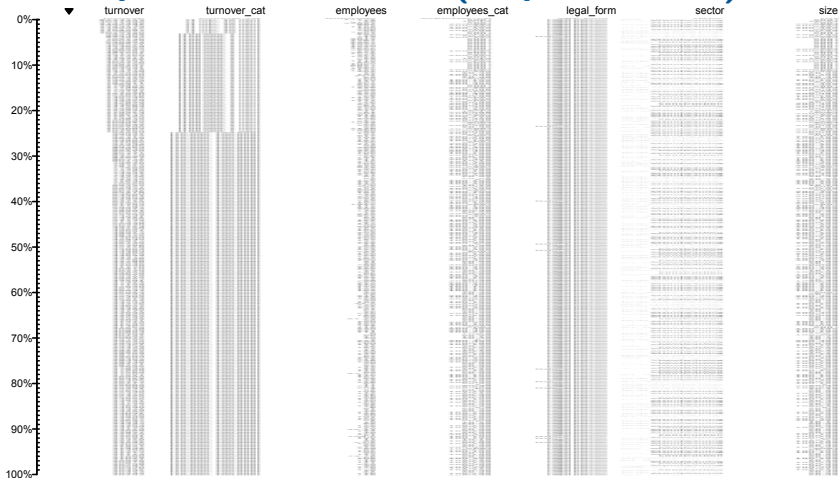
# Tableplot: SBS data (unprocessed)

▼ turnover	turnover_cat	employees	employees_cat	legal_form	sector	size
8124110	[1 min, 10 mln]	17172	500+	Legal person	Transp., storage, comm.	500+
7375977	[1 min, 10 mln]	1865	500+	Legal person	Mining and quarrying	500+
5496033	[1 min, 10 mln]	1366	500+	Legal person	Electricity, gas, water supply	500+
4421226	[1 min, 10 mln]	431	200-499	Legal person	Wholesale and retail trade	200-499
4386360	[1 min, 10 mln]	1036	500+	Legal person	Manufacturing	500+
4275533	[1 min, 10 mln]	136	100-199	Legal person	Electricity, gas, water supply	100-199
4031636	[1 min, 10 mln]	12	10-19	Legal person	Wholesale and retail trade	20-49
4012386	[1 min, 10 mln]	9187	500+	Legal person	Manufacturing	500+
3909439	[1 min, 10 mln]	1004	500+	Legal person	Wholesale and retail trade	500+
3133380	[1 min, 10 mln]	213	200-499	Legal person	Wholesale and retail trade	200-499
3095844	[1 min, 10 mln]	3747	500+	Legal person	Wholesale and retail trade	500+
2931738	[1 min, 10 mln]	491	200-499	Legal person	Wholesale and retail trade	200-499
2883939	[1 min, 10 mln]	3166	500+	Legal person	Manufacturing	500+
2768831	[1 min, 10 mln]	4113	500+	Legal person	Manufacturing	500+
2272912	[1 min, 10 mln]	1695	500+	Legal person	Manufacturing	500+
2250557	[1 min, 10 mln]	229	200-499	Legal person	Wholesale and retail trade	50-99
2162875	[1 min, 10 mln]	294	200-499	Legal person	Wholesale and retail trade	200-499
2049469	[1 min, 10 mln]	4124	500+	Legal person	Real est., renting, bus. act.	500+
1952200	[1 min, 10 mln]	1309	500+	Legal person	Electricity, gas, water supply	500+
1945146	[1 min, 10 mln]	336	200-499	Legal person	Manufacturing	200-499
1840955	[1 min, 10 mln]	415	200-499	Legal person	Manufacturing	200-499
1724782	[1 min, 10 mln]	168	100-199	Legal person	Wholesale and retail trade	100-199
1721028	[1 min, 10 mln]	2787	500+	Legal person	Transp., storage, comm.	500+
1499374	[1 min, 10 mln]	105	100-199	Legal person	Electricity, gas, water supply	200-499
1406432	[1 min, 10 mln]	943	500+	Legal person	Wholesale and retail trade	500+
1361787	[1 min, 10 mln]	710	500+	Legal person	Manufacturing	500+
1329278	[1 min, 10 mln]	5841	500+	Legal person	Wholesale and retail trade	500+
1320454	[1 min, 10 mln]	5032	500+	Legal person	Manufacturing	500+
1318814	[1 min, 10 mln]	30	20-49	Legal person	Wholesale and retail trade	20-49
1316711	[1 min, 10 mln]	2055	500+	Legal person	Electricity, gas, water supply	500+
...	...	...	...	...	...	...

# Tableplot: SBS data (unprocessed)



# Tableplot: SBS data (unprocessed)



row bins:  
100

objects:  
51621

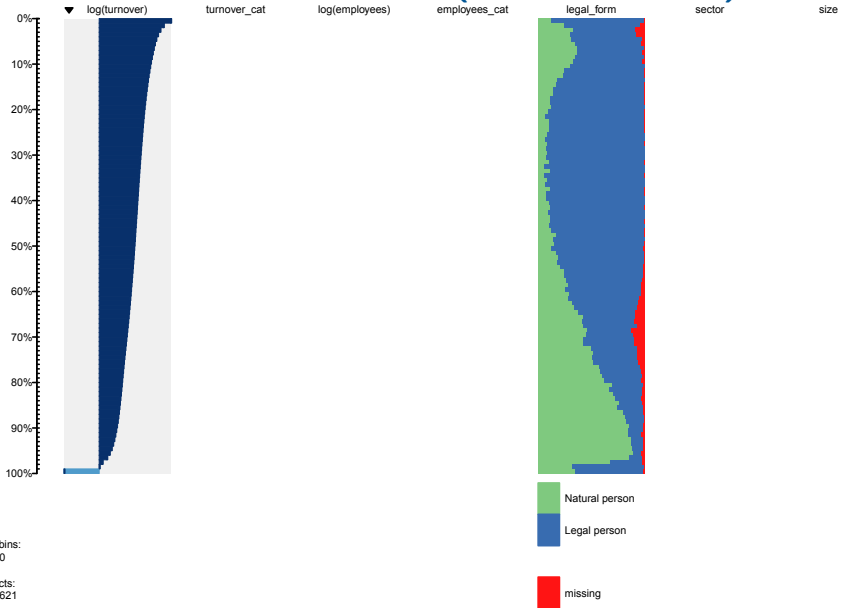
# Tableplot: SBS data (unprocessed)



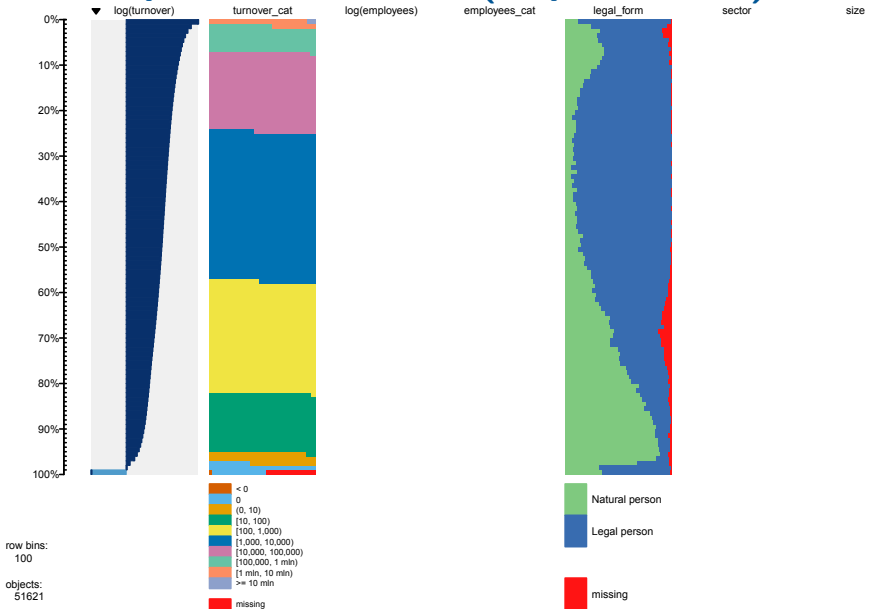
row bins:  
100

objects:  
51621

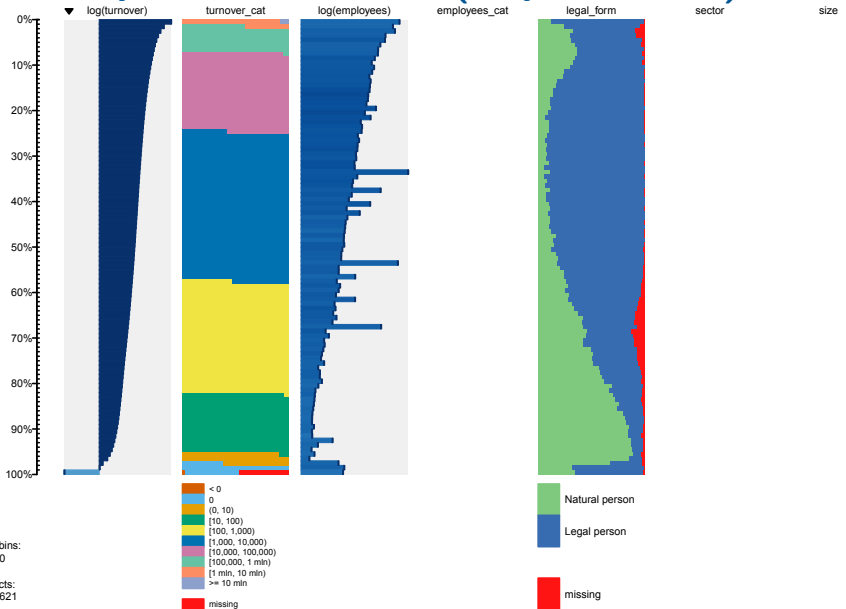
# Tableplot: SBS data (unprocessed)



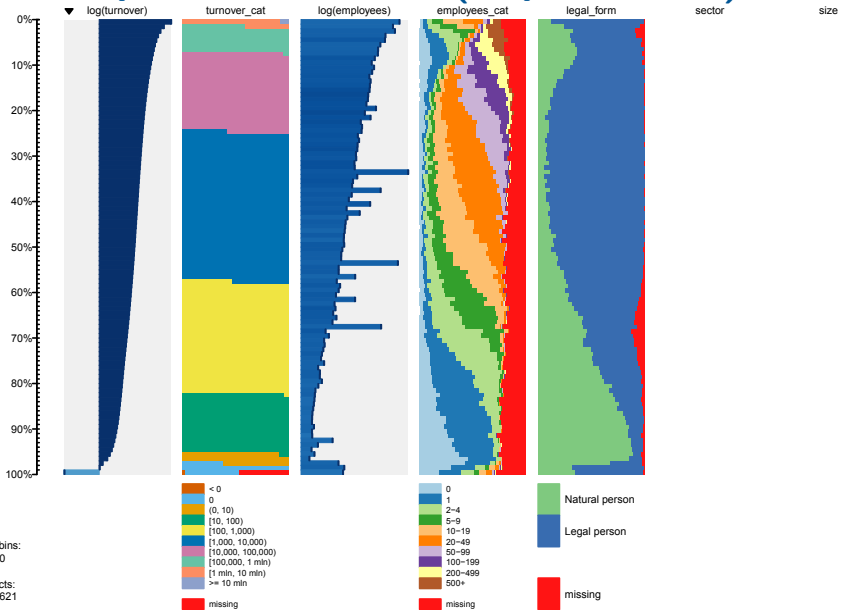
# Tableplot: SBS data (unprocessed)



# Tableplot: SBS data (unprocessed)



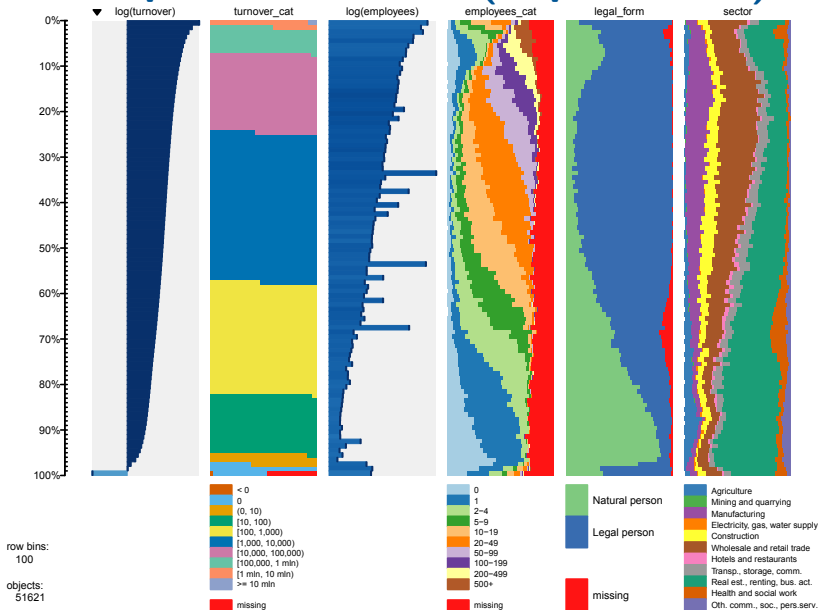
# Tableplot: SBS data (unprocessed)



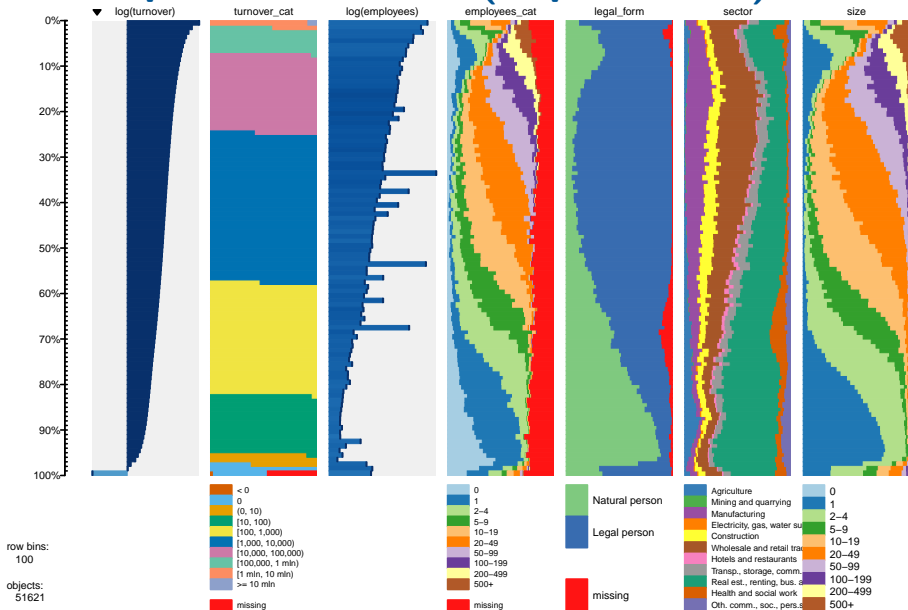


# Tableplot: SBS data (unprocessed)

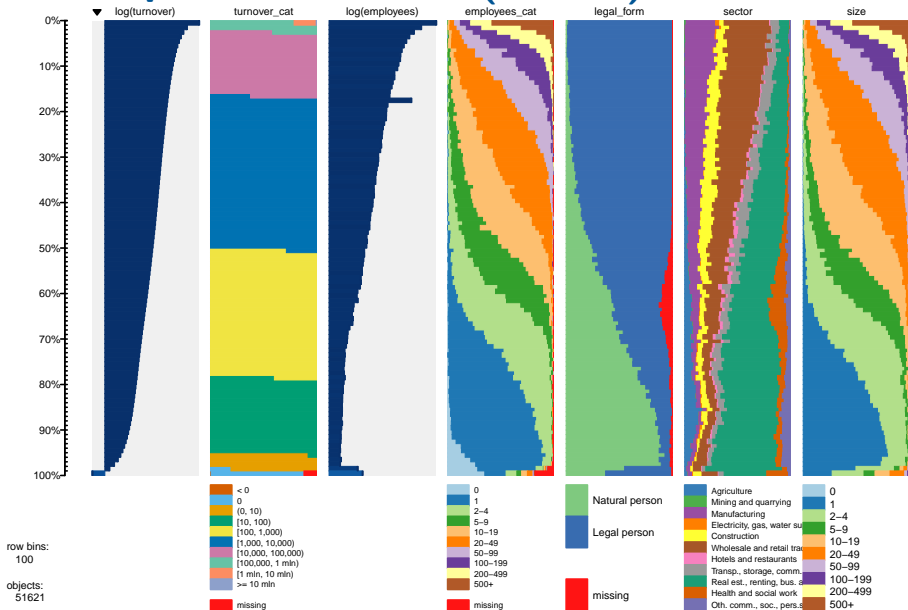
size



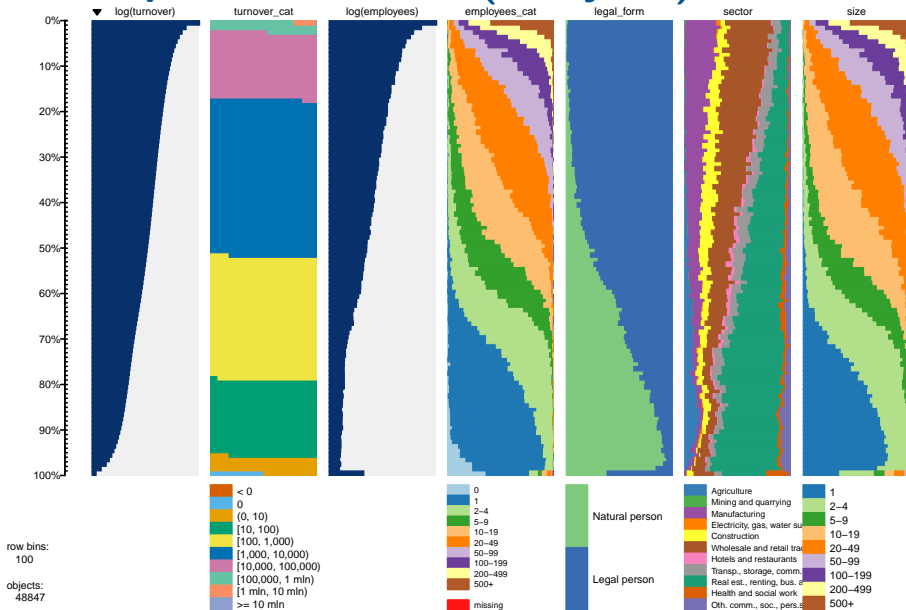
# Tableplot: SBS data (unprocessed)



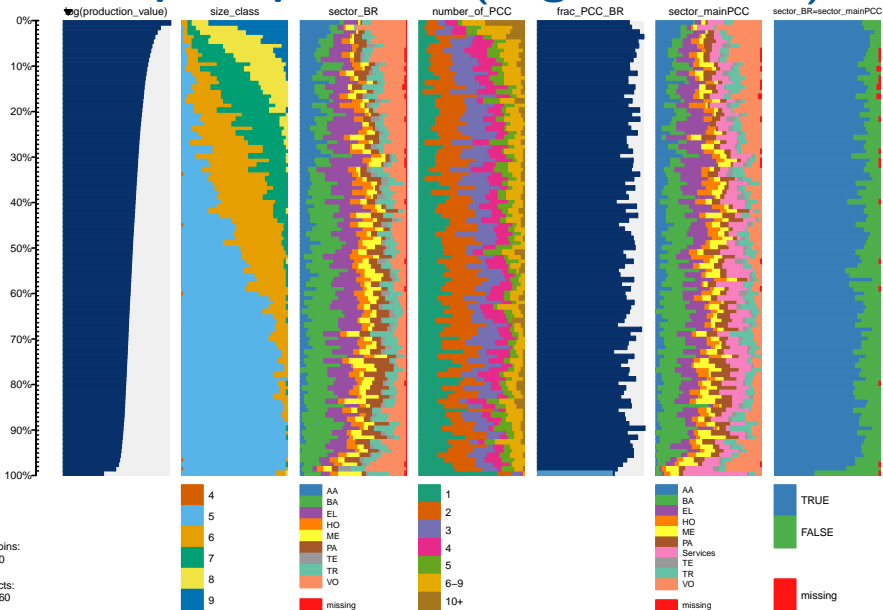
# Tableplot: SBS data (edited)



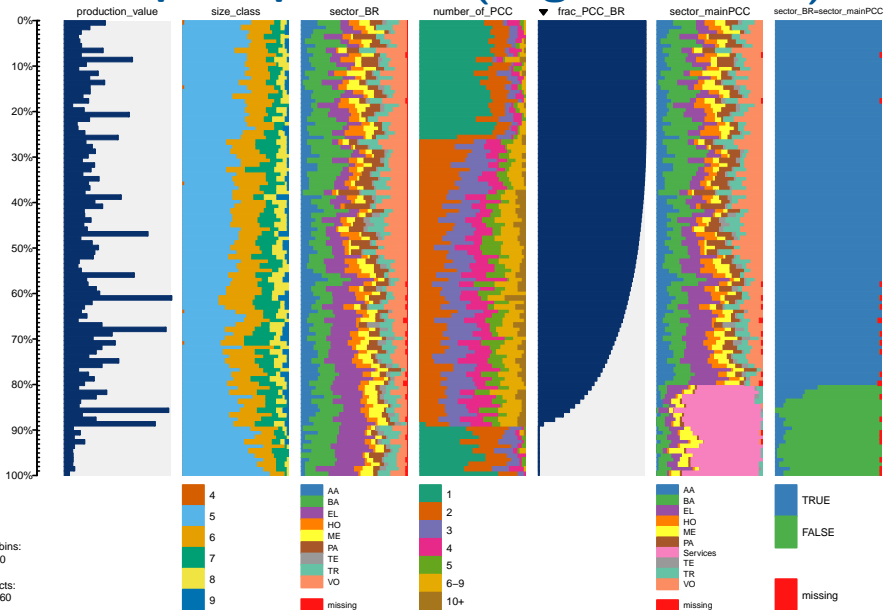
# Tableplot: SBS data (analysed)



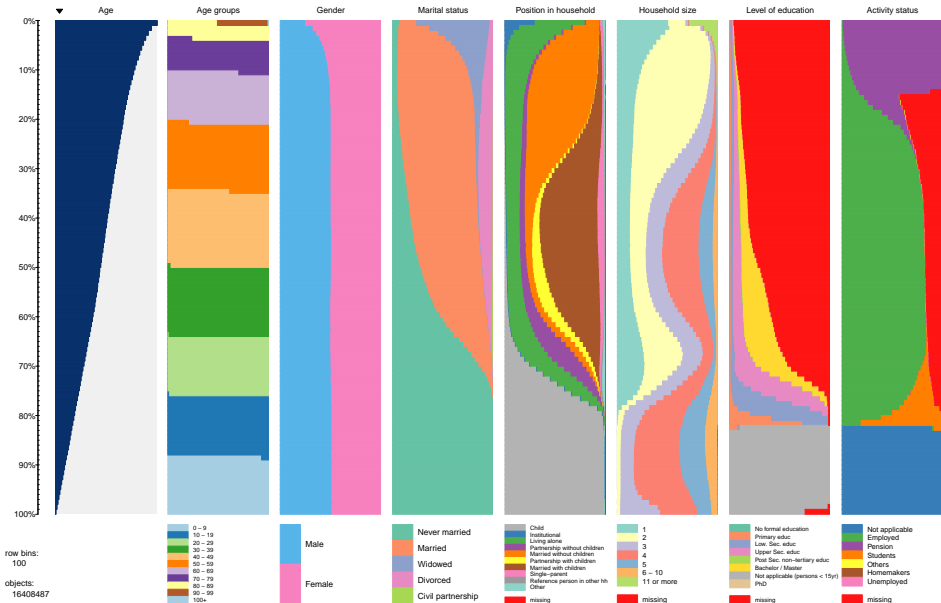
# Tableplot: prodcom (large industries)



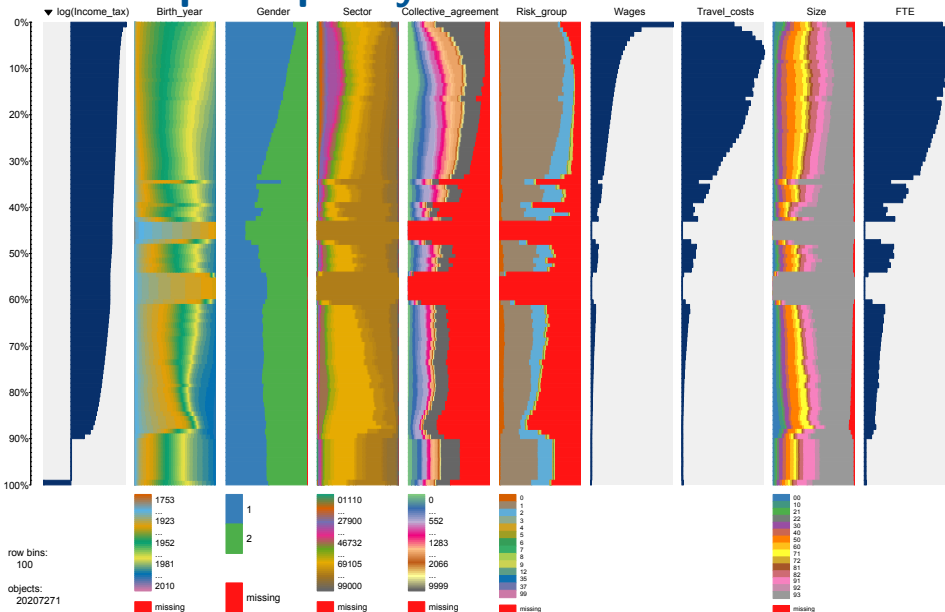
# Tableplot: prodcom (large industries)



# Tableplot: virtual census



# Tableplot: policy record administration





# Tableplot

- Useful to explore BIG multivariate data
  - Strange data patterns
  - Selectivity of missing values
  - Correlation between variables
- Implementation in R:
  - Package *tabplot*
  - Old graphical user interface: package *tabplotGTK*
  - Interactive web-based interface: package *tabplotd3*

# Tableplot

- Useful to explore BIG multivariate data
  - Strange data patterns
  - Selectivity of missing values
  - Correlation between variables
- Implementation in R:
  - Package *tabplot*
  - Old graphical user interface: package *tabplotGTK*
  - Interactive web-based interface: package *tabplotd3*

# Tableplot

- Useful to explore BIG multivariate data
  - Strange data patterns
  - Selectivity of missing values
  - Correlation between variables
- Implementation in R:
  - Package *tabplot*
  - Old graphical user interface: package *tabplotGTK*
  - Interactive web-based interface: package *tabplotd3*

# Tableplot

- Useful to explore BIG multivariate data
  - Strange data patterns
  - Selectivity of missing values
  - Correlation between variables
- Implementation in R:
  - Package *tabplot*
  - Old graphical user interface: package *tabplotGTK*
  - Interactive web-based interface: package *tabplotd3*

# Tableplot

- Useful to explore BIG multivariate data
  - Strange data patterns
  - Selectivity of missing values
  - Correlation between variables
- Implementation in R:
  - Package *tabplot*
  - Old graphical user interface: package *tabplotGTK*
  - Interactive web-based interface: package *tabplotd3*

# Tableplot

- Useful to explore BIG multivariate data
  - Strange data patterns
  - Selectivity of missing values
  - Correlation between variables
- Implementation in R:
  - Package *tabplot*
  - Old graphical user interface: package *tabplotGTK*
  - Interactive web-based interface: package *tabplotd3*

# References

- 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.
- Tennekes, M., Jonge, E. de, Daas, P.J.H. (2013) Visualizing and Inspecting Large Datasets with Tableplots, Forthcoming in Journal of Data Science.
- All R-packages are available on CRAN. Development sites:
  - *treemap*: [code.google.com/p/treemap-package/](http://code.google.com/p/treemap-package/)
  - *tabplot*: [code.google.com/p/tableplot/](http://code.google.com/p/tableplot/)
  - *tabplotd3*: [code.google.com/p/tabplotd3/](http://code.google.com/p/tabplotd3/)