Innovative Visual Tools for Data Editing

Martijn Tennekes Edwin de Jonge, Piet Daas

24-26 September 2012







Funded under Socio-economic Sciences & Humanities



Visual tools for data editing

Visual tools for data editing

Traditional tools

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

Visual tools for data editing

Traditional tools

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

- Hierarchical data
- Multivariate data
- BIG data

Visual tools for data editing

Traditional tools

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

- Hierarchical data
- Multivariate data
- BIG data

Visual tools for data editing

Traditional tools

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

- Hierarchical data
- Multivariate data
- BIG data

Visual tools for data editing

Traditional tools

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

- Hierarchical data
- Multivariate data
- BIG data

Visual tools for data editing

Traditional tools

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

Novel tools

- Treemap
- Tableplot

- Hierarchical data
- Multivariate data
- BIG data

Visual tools for data editing

Traditional tools

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

Novel tools

- Treemap
- Tableplot



- Hierarchical data
- Multivariate data
- BIG data

Visual tools for data editing

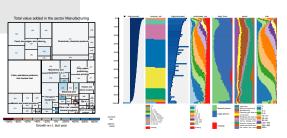
Traditional tools

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

Novel tools

- Treemap
- Tableplot

- Hierarchical data
- Multivariate data
- BIG data



Treemap (SBS data) Total value added

Treemap (SBS data)

Total value added

Manufacturing	Real estate, renting, business activities	Transportation, storage, communication		
	Construction	Electricity, gas, Mining and and water supply quarrying		
Wholesale and retail trade		Hotels and restaurants Dither community Health Health Property Pro		

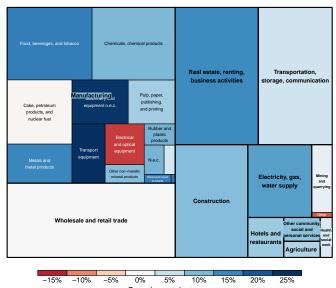
Treemap (SBS data)

Total value added

Food, beverages, and tob	Chemicals, chemical products		al products	Real estate, renting, business activities	Transportation,			
Coke, petroleum products, and nuclear fuel	products, and and printing		business activities	Storage, con	munication			
Metals and metal products	Transpo equipme			plastic products N.e.c.		Electricity, ga	ae Mining	
Wholesale and retail trade					Construction	water supply quarrying		
						Hotels and person restaurants	r community ocial and Health onal services and social work	

Treemap (SBS data)

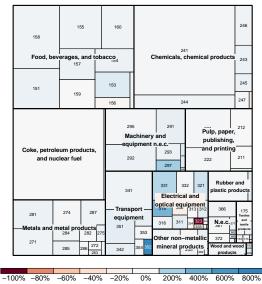
Total value added



Growth w.r.t. last year

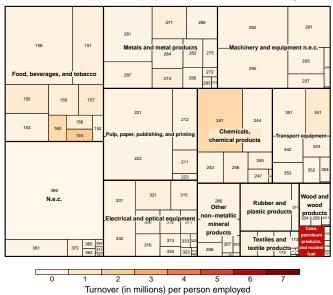
Treemap: zooming in

Total value added in the sector Manufacturing



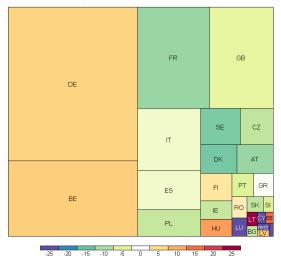
Treemap: densities

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

- 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

- 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

- 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

- 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

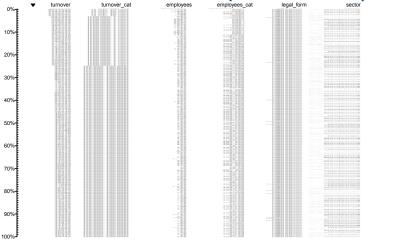
- Useful to explore hierarchical data
- Follows the top-down approach of data editing
- Implementation in R:
 - Package treemap
 - Interactive use is very limited
 - Can be used with other tools

- 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

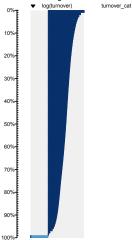
						/
turnover	turnover_cat	employees	employees_c	at legal_form	n sector	size
1758	[1,000, 10,000)	17	10-19	Legal person	Real est., renting, bus. act.	20-49
4	[10, 100)	3	2-4	Natural person	Wholesale and retail trade	2-4
84	[10, 100)	1	1	Natural person	Manufacturing	2-4
11778	[10,000, 100,000)	69	50-99	Legal 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

	00000	p. 00				2.CP
size	n sector	_cat legal_forr	s employees_	t employe	turnover_ca	▼ turnover
500+	Transp., storage, comm.	Legal person	500+	17172	[1 mln, 10 mln)	8124110
500+	Mining and quarrying	Legal person	500+	1865	[1 mln, 10 mln)	7375977
500+	Electricity, gas, water supply	Legal person	500+	1366	[1 mln, 10 mln)	5496033
200-499	Wholesale and retail trade	Legal person	200-499	431	[1 mln, 10 mln)	4421226
500+	Manufacturing	Legal person	500+	1036	[1 mln, 10 mln)	4386360
100-199	Electricity, gas, water supply	Legal person	100-199	136	[1 mln, 10 mln)	4275533
20-49	Wholesale and retail trade	Legal person	10-19	12	[1 mln, 10 mln)	4031636
500+	Manufacturing	Legal person	500+	9187	[1 mln, 10 mln)	4012386
500+	Wholesale and retail trade	Legal person	500+	1004	[1 mln, 10 mln)	3909439
200-499	Wholesale and retail trade	Legal person	200-499	213	[1 mln, 10 mln)	3133380
500+	Wholesale and retail trade	Legal person	500+	3747	[1 mln, 10 mln)	3095844
200-499	Wholesale and retail trade	Legal person	200-499	491	[1 mln, 10 mln)	2931738
500+	Manufacturing	Legal person	500+	3166	[1 mln, 10 mln)	2883939
500+	Manufacturing	Legal person	500+	4113	[1 mln, 10 mln)	2768831
500+	Manufacturing	Legal person	500+	1695	[1 mln, 10 mln)	2272912
50-99	Wholesale and retail trade	Legal person	200-499	229	[1 mln, 10 mln)	2250557
200-499	Wholesale and retail trade	Legal person	200-499	294	[1 mln, 10 mln)	2162875
500+	Real est., renting, bus. act.	Legal person	500+	4124	[1 mln, 10 mln)	2049469
500+	Electricity, gas, water supply	Legal person	500+	1309	[1 mln, 10 mln)	1952200
200-499	Manufacturing	Legal person	200-499	336	[1 mln, 10 mln)	1945146
200-499	Manufacturing	Legal person	200-499	415	[1 mln, 10 mln)	1840955
100-199	Wholesale and retail trade	Legal person	100-199	168	[1 mln, 10 mln)	1724782
500+	Transp., storage, comm.	Legal person	500+	2787	[1 mln, 10 mln)	1721028
200-499	Electricity, gas, water supply	Legal person	100-199	105	[1 mln, 10 mln)	1499374
500+	Wholesale and retail trade	Legal person	500+	943	[1 mln, 10 mln)	1406432
500+	Manufacturing	Legal person	500+	710	[1 mln, 10 mln)	1361787
500+	Wholesale and retail trade	Legal person	500+	5841	[1 mln, 10 mln)	1329278
500+	Manufacturing	Legal person	500+	5032	[1 mln, 10 mln)	1320454
20-49	Wholesale and retail trade	Legal person	20-49	30	[1 mln, 10 mln)	1318814
500+	Electricity, gas, water supply	Legal person	500+	2055	[1 mln, 10 mln)	1316711

100000	turnover_cat	employees	employees_cat	legal_form		
*633333			111000			1110
	14M4386627 HHI	- 2 1 2				
199222		1223				
162368	1100 200: 1 0001	-828				
103335		-555				
122222		* X 2 2	===415764			
142433		488	A 11 11 A 3.3.1		Post social and social and social	
111222		- 222	111000			
113333		- 2 2 3				1100
112923	11111 2000 1 11011	255	X1111, X31,11		884444	
145533	111112000 1 11111	-589	-H8158H		200000000000000000000000000000000000000	11800
189388		-080	\$11 HT 2 H H			
33338		125	\$11H(\$33H			211816
82355		-22	\$11H SHH			11113
88338	118:888: 188:888	288	118591			8 1111 8
25355	110-2005-1305-2001	- 23	\$1101980			
17233		-23	2110151111			2000
62333		200	1118:981			
83363		-86	1111151111			11110
85263		- 222	2110-200			\$111.3
15932		-89	1101580			51815
10221	110 000 100 000	288	\$HH-\$HH			1103
14333	110-000-100-000	238	\$1111.5111		REPORTED TO THE REAL PROPERTY.	\$118.3
38338		-28	X11013111		MARKET CONT. (1971)	
33388	110:000:100:000	788	211073000	1201111111	210000100100000	11113
33338		- 58	811013111		=: FEEDOLE 100	811013
3114.53	18:888: 188 888	2.65	\$11B1S1H	10011111111111	000 00 00 00 00 00 00 00 00 00 00 00 00	911815
17233		6.80	11112			111113
111333	110:000:100:000	193				211000
11223	110-000-100-000		1185			
11222		-25				
33938	118:888: 188.888		x.0815311			4 1 1 1 1 1 1
10019		0.02	01107200			1100
11233		700	THE SH			7183
111233		- 68	11818181			31810
18198		125	*#BESBB		=1000100000	
17616	10-000-100-000	- 23	2 853 88			
26566		199	1183			11112
111222		-53	2112111			411811
15440		- 88	1183381			
11633	110:000:100:000	- 5 8	2 - 615 9 9	1 08 11 11 11 11 11		2.1.010
11939		188	111815111			118
11753		28	- 812.01			
11103		68	110501			111873
11032		~28	11818.88			1193
11700	116:666 166:666		10000			1186
10133	10 000 100 000	- 55	11868		THE STATE OF STREET STATE OF STREET	
101133	110-000 100 000	7 0 2	2000000			



Tableplot: SBS data (unprocessed) v log(turnover) turnover_cat log(employees) employees_cat legal_form sector



. . . _

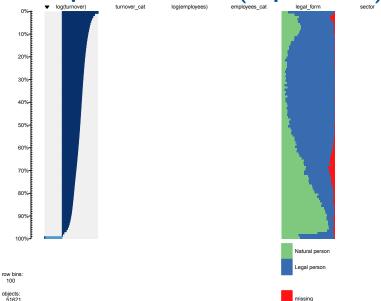
iegai_torn

size

row bins: 100

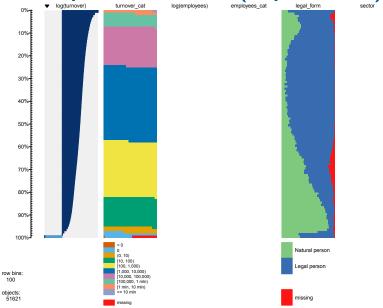
objects: 51621

size

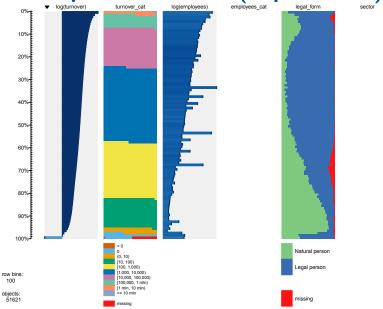


51621

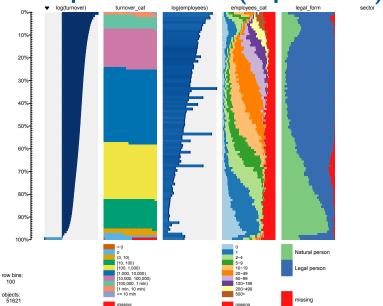
size



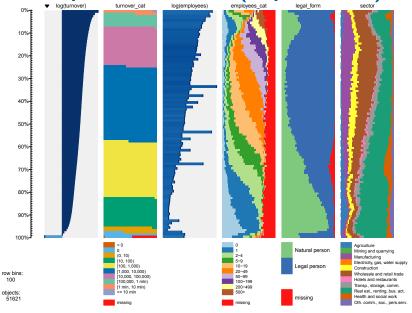
size

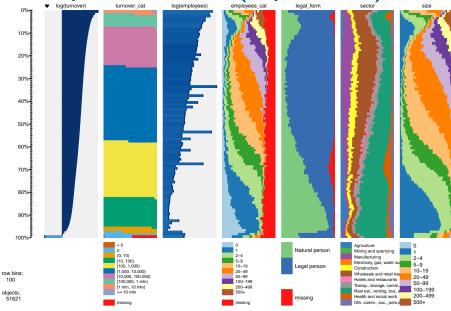


size

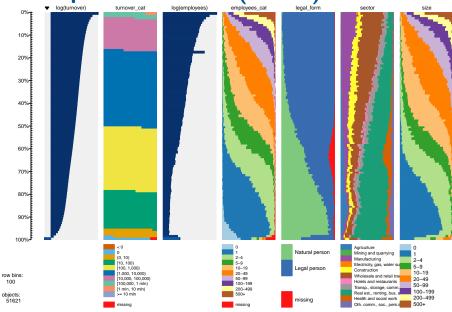


size

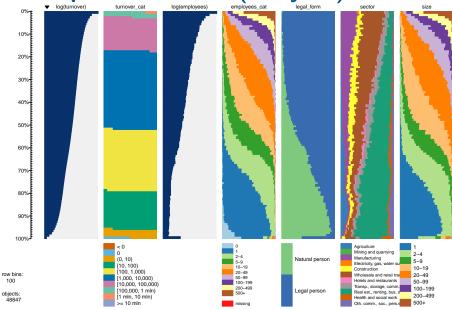




Tableplot: SBS data (edited)



Tableplot: SBS data (analysed)

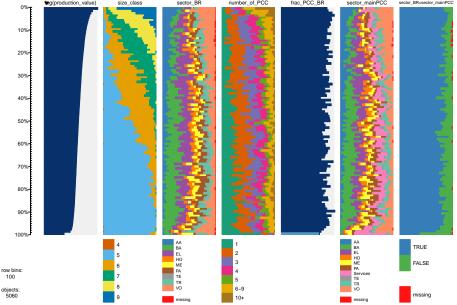


Tableplot: prodcom (large industries)

g(production_value)

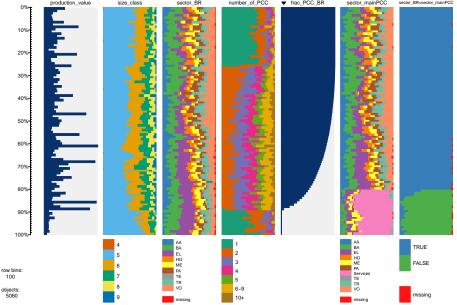
sector_BR

**

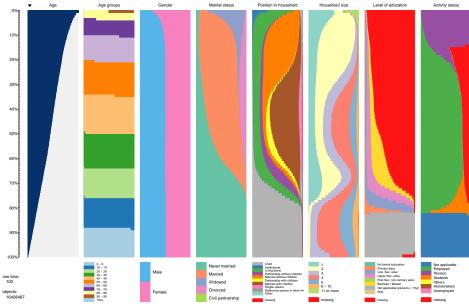


Tableplot: prodcom (large industries)

production_value size_class sector_BR number_of_PCC ▼ frac_PCC_BR sector_mainPCC sector_BR-sector_mainPCC sector_BR-



Tableplot: virtual census



Tableplot: policy record administration FTE 10% row bins objects:

missing

- 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

- 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

- 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

- 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

- 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

- 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/
 - tabplot: code.google.com/p/tableplot/
 - tabplotd3: code.google.com/p/tabplotd3/