School of Information Systems

SMU SINGAPORE MANAGEMENT UNIVERSITY

Lesson 5
Visualising and Analysing Multidimensional Data

Instructor: Dr. Kam Tin Seong
Associate Professor of Information System (Practice)
School of Information Systems
Singapore Management University

What will you learn from this lesson?

- Understand the characteristics of multidimensional data
- Visual analytics techniques and tools for visualising and analysing multidimensional continuous data
- Visual analytics techniques and tools for visualising and analysing multidimensional categorical data
- Sensing both categorical and continuous multidimensional data
- Multidimensional data analysis best practices

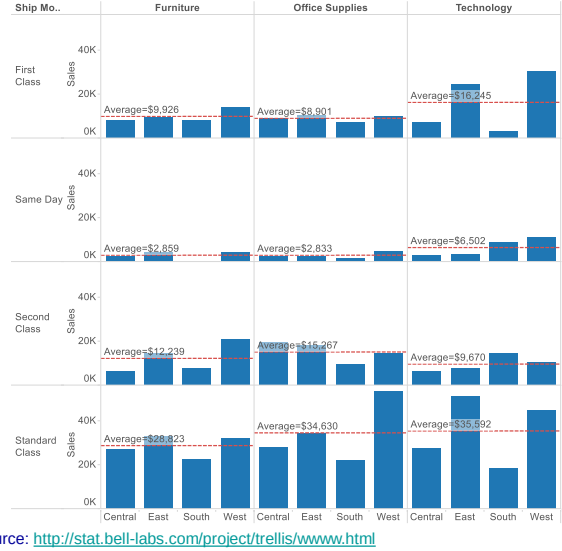
School of Information Systems
SMU SINGAPORE MANAGEMENT UNIVERSITY
2

Visual analytics techniques of multidimensional continuous data

- Trellis display
- Ternary
- Glyphs
- Parallel coordinates
- Heatmap
- Table Lens
- Tableplots

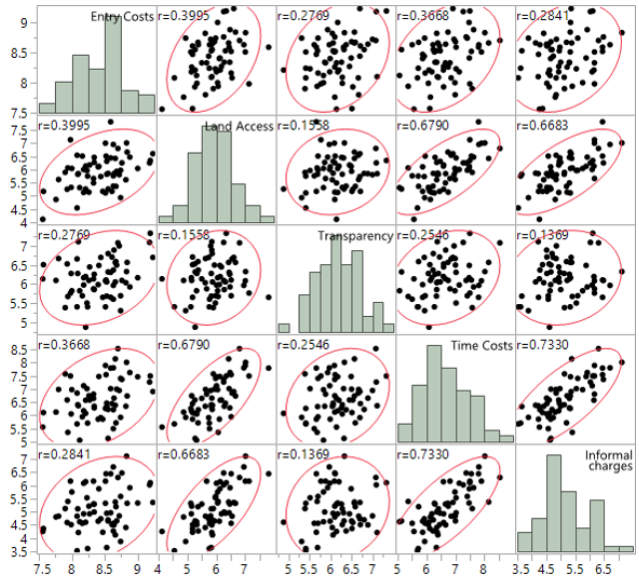
Trellis Display: Basic principles

- More popularly know as small multiples



Source: http://stat.bell-labs.com/project/trellis/wwww.html

Interactive Trellis



School of Information Systems

SMU Singapore Management University

Trellis Display I

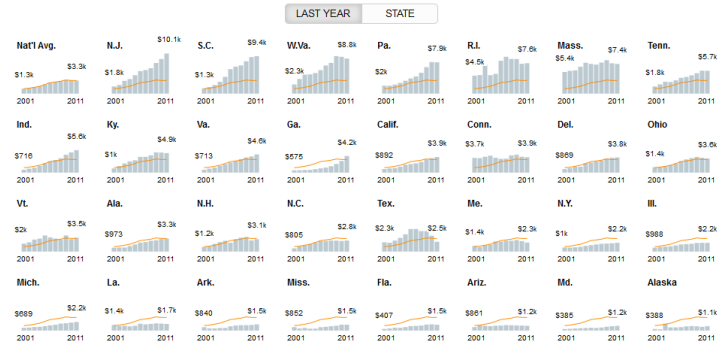
Treatment Tracker
The Doctors and Services in Medicare Part B

Ambulances for Dialysis Patients on Rise

by *Eric Sagara* and *Charles Ornstein*, *ProPublica*
June 12, 2014

Tweet 40 Like 0

New Jersey leads the nation in average annual Medicare spending on ambulance services per dialysis patient, billing for unusually large numbers of non-emergency ambulance rides, according to a our analysis of Medicare payment data. Several ambulance providers said they've heard of providers who sign up patients who don't need the service — a form of fraud. The charts below show spending by state from 2001 to 2011, compared to national averages. Sort by the most-recent year or by state. **Related story:** Medicare Taken For a Ride By Ambulance Companies in New Jersey.

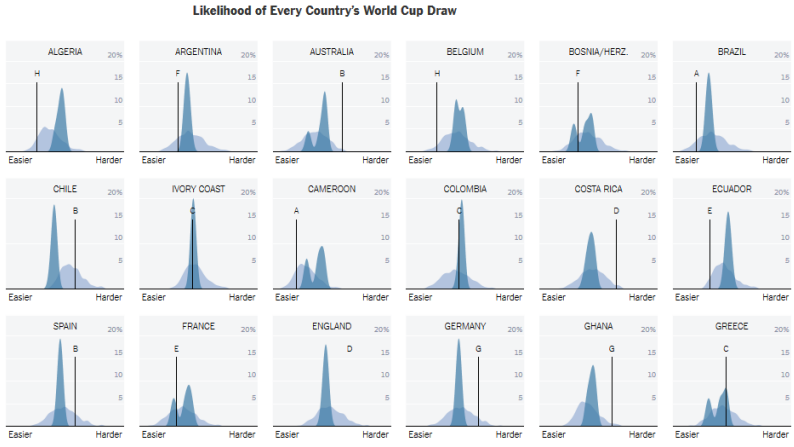


Source: http://www.propublica.org/article/medicare-taken-for-a-ride-by-ambulance-companies-in-new-jersey

School of Information Systems

SMU Singapore Management University

Trellis Display II



Source: http://www.nytimes.com/interactive/2014/06/03/upshot/world-cup-draw-simulation.html

Beyond bivariate data

- Three continuous variables

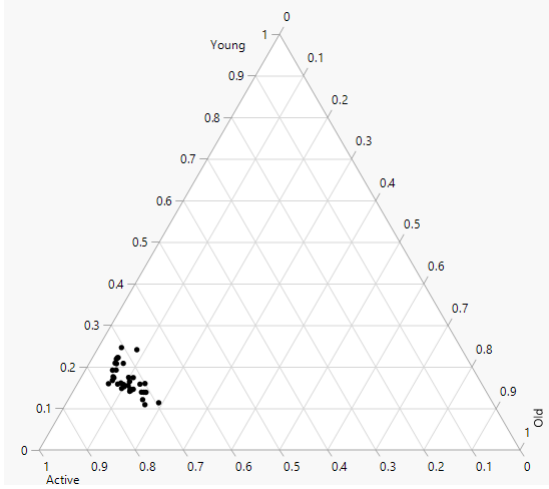
CENSUS OF POPULATION 2010
Table A3 Resident Population by Planning Area, Age Group and Sex

Number

Planning Area	Total	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65 & Over
	Total														
Total	3,771,721	194,432	215,675	244,302	263,750	247,190	272,639	298,687	320,024	309,441	323,459	303,044	248,696	191,995	338,387
Ang Mo Kio	179,297	7,967	8,424	9,335	10,457	10,656	13,400	14,502	14,510	13,525	14,862	14,605	13,785	11,868	21,401
Bedok	294,519	13,230	15,018	17,489	20,083	20,156	21,265	21,707	22,751	22,018	24,615	24,632	21,891	18,018	31,646
Bishan	91,298	3,941	4,759	5,691	7,188	6,338	5,930	6,010	7,354	7,102	8,041	8,245	6,725	5,141	8,833
Bukit Batok	144,198	7,187	8,516	9,738	10,427	10,625	11,332	10,941	11,829	12,259	13,049	12,433	9,911	6,487	9,464
Bukit Merah	157,122	8,049	6,892	6,894	7,479	7,940	10,865	13,871	13,495	11,796	11,650	12,057	11,352	10,782	24,000
Bukit Panjang	128,734	7,106	8,334	9,324	10,224	9,734	9,102	9,488	11,171	10,469	11,528	10,784	7,907	5,316	8,247
Bukit Timah	70,314	3,115	4,456	4,681	5,238	4,613	3,816	4,096	5,689	6,071	6,156	5,726	4,841	4,178	7,638
Changi	2,155	169	192	159	139	94	111	187	252	223	175	122	81	73	178
Choa Chu Kang	173,291	9,232	11,823	15,195	15,362	11,637	11,893	12,541	14,483	16,675	16,919	13,299	9,112	5,855	9,265
Clementi	91,874	4,235	4,567	4,713	4,732	5,133	7,093	7,818	7,831	7,201	7,331	6,882	7,064	6,648	10,626
Downtown Core	3,722	144	122	140	137	161	287	315	326	357	295	295	265	262	616

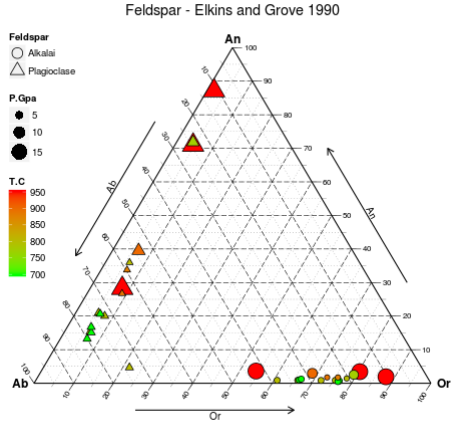
Planning Zone	Young	Active	Old
Outram	2260	13788	3811
Downtown Core	406	2700	616
Rochor	1899	11341	2424
Queenstown	13777	69670	15055
Bukit Merah	21835	111287	24000
Toa Payoh	17327	88788	18538
Kallang	13861	71438	14260
Marine Parade	7573	33116	6629
Novena	7389	33164	6087
Geylang	17657	88237	14796
Ang Mo Kio	25726	132170	21401
Others	634	3322	528
Clementi	13515	67733	10626
Bukit Timah	12252	50424	7638
Bedok	45737	217136	31646

Ternary chart



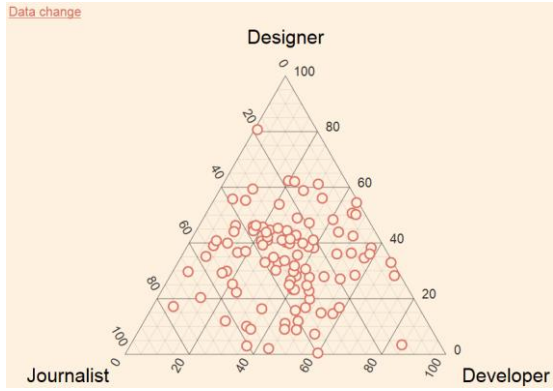
Ternary Chart

- ggtern: ternary diagrams in R (http://www.ggtern.com/)



D3.js based Ternary Chart

- Ternary Plot in D3.js (http://bl.ocks.org/tomgp/7674234), (https://gist.github.com/widged/5780720), (http://bl.ocks.org/tomgp/7766353)



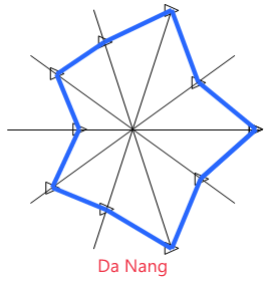
Multidimensional continuous data

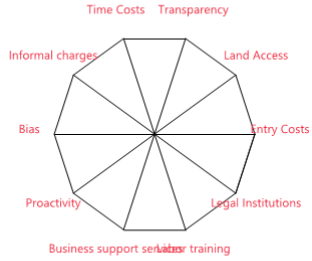
- Multivariate data
- More than three variables

Province	Entry Costs	Land Access	Transparency	Time Costs	Informal charges	Bias	Proactivity	Business support services	Labor training	Legal Institutions	PCI 2015	PCI 2015 ranking	PCI tier
Da Nang	9.19	6.35	7.33	7.5	6.11	4.77	6.17	6.06	7.62	6.46	68.34	1	Excellent
Dong Thap	8.7	6.81	7.08	8.54	6.31	6.69	7.04	5.94	5.71	7.44	66.39	2	Excellent
Lao Cai	8.64	5.85	6.89	6.66	4.51	6.44	6	6.2	5.84	6.33	62.32	5	Excellent
Quang Ninh	9.18	6.26	7.09	7.27	6.03	4.69	5.31	5.92	7.19	5.93	65.75	3	Excellent
Vinh Phuc	8.41	6.13	6.49	7.48	5.88	5.65	5.78	5.26	6.65	6.28	62.56	4	Excellent
Ben Tre	8.59	7.82	5.66	7.8	6.45	5.16	4.88	5.51	5.51	6.33	60.1	12	High
HCMC	7.57	5.18	6.51	6.11	4.37	4.28	4.19	7	6.89	5.04	61.36	6	High
Kien Giang	8.66	6.43	6.57	8.15	5.38	4.56	5.78	5.4	5.2	7.62	60.31	11	High
Long An	8.76	6.26	6.1	7.37	6.48	5.83	5.48	5.16	5.88	6.48	60.86	9	High
Quang Nam	8.52	6.52	6.11	7.55	6.45	4.16	5.13	5.77	5.76	6.8	61.06	8	High
Thai Nguyen	8.84	6.67	6.57	6.74	5.33	4.6	5.09	4.82	7.14	5.69	61.21	7	High
Thanh Hoa	8.44	5.74	6.7	5.92	4.74	3.44	4.32	6.06	6.82	5.83	60.74	10	High
Ca Mau	7.9	5.9	5.39	7.58	5.13	4.98	4.72	5.69	4.45	5.35	54.4	59	Low
Cao Bang	8.47	5.18	5.66	5.86	4.23	4.82	3.92	5.3	5.87	4.83	54.44	58	Low
Hung Yen	8.21	5.27	4.88	6.08	4.61	4.21	4.2	5.18	6.72	5.87	55.1	56	Low
Lang Son	8.12	5.32	5.83	5.06	4.63	5.4	3.32	5.51	5.55	5.19	54.61	57	Low
An Giang	8.68	5.94	6.5	7	4.66	4.41	4.47	5.67	5.14	6.18	57.61	39	Mid-high
Bac Giang	8	6.05	5.83	6.98	5.76	4.7	4.71	5.69	5.65	4.64	57.61	40	Mid-high
Bac Lieu	7.98	7.14	5.77	7.5	6.17	7.29	5.97	4.97	4.66	7.09	58.44	33	Mid-high
Bac Ninh	8.1	5.85	7.11	6.37	4.22	3.5	5.07	5.35	6.82	5.38	59.91	13	Mid-high
Binh Dinh	9	6.05	6.17	7.47	5.34	4.85	4.87	5.23	6.1	5.56	59.23	20	Mid-high
Binh Duong	8.08	6.55	6.66	6.85	4.84	4.47	5.58	5.2	5.76	6.1	58.89	25	Mid-high
Binh Thuan	8.21	6.14	6.2	6.58	6.15	5.81	4.46	5.68	5.47	4.75	58.83	26	Mid-high
BRVT	8.09	5.83	6.28	6.41	5.12	5.34	4.38	5.56	6.49	5.38	59.51	18	Mid-high
Can Tho	8.18	6.01	5.98	7.13	6.09	3.9	4.32	5.89	5.94	6.62	59.81	14	Mid-high

Glyphs

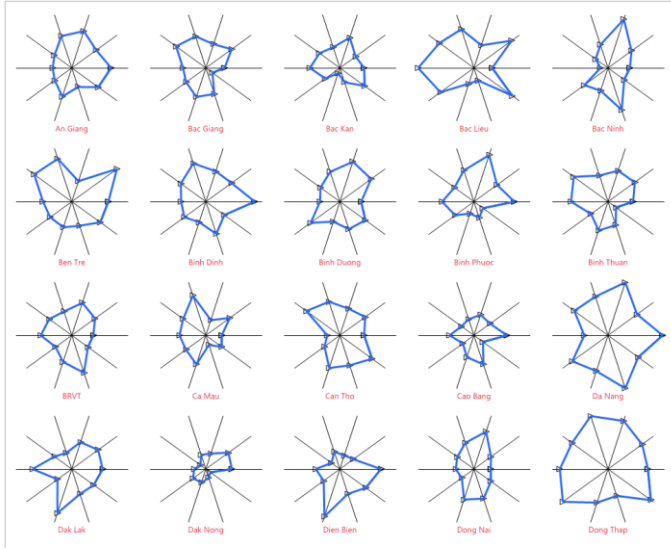
- Also known as radar chart, star chart and spider chart



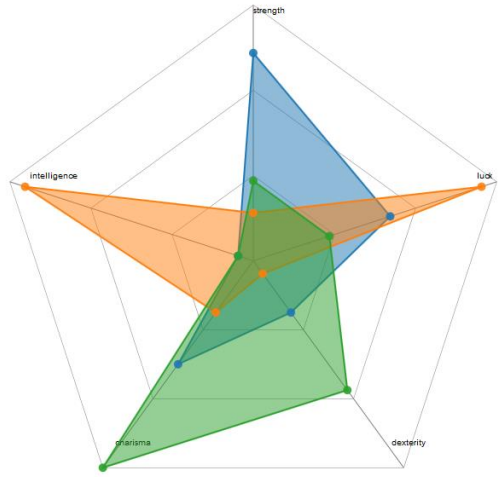


Source: http://en.wikipedia.org/wiki/Radar_chart and
http://blog.scottlogic.com/2011/09/23/a-critique-of-radar-charts.html

Multiple Glyphs

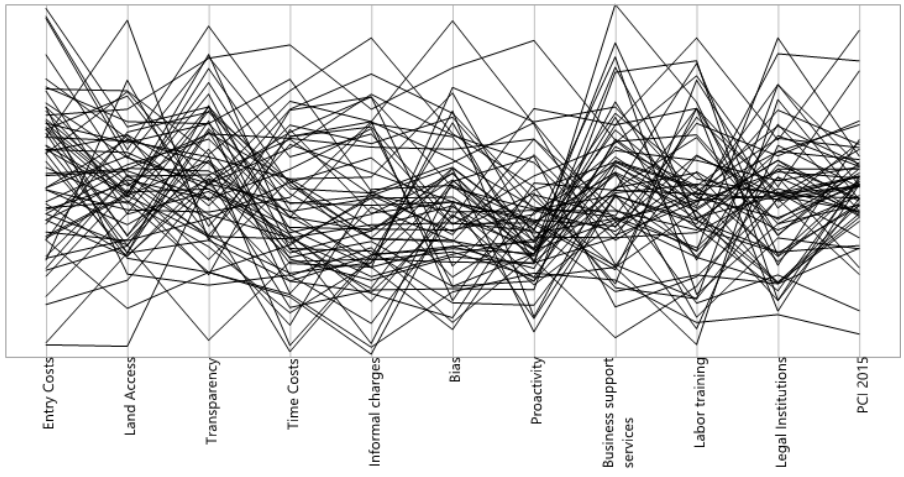


Radar Chart in D3.js

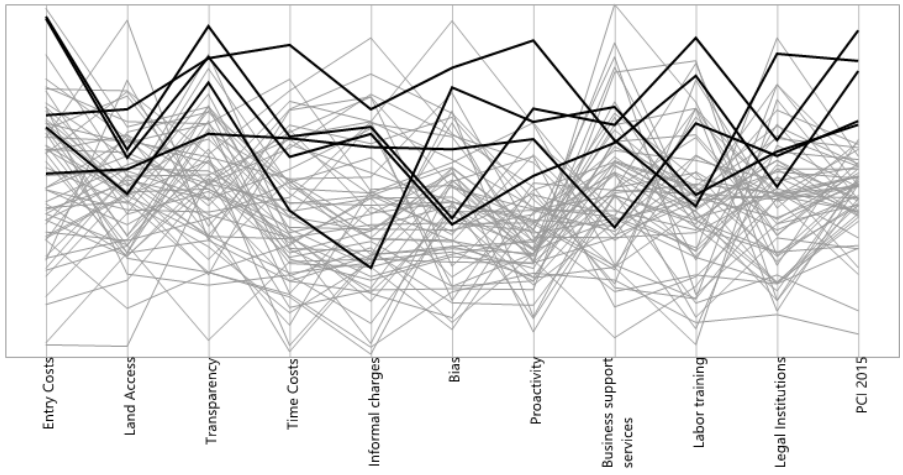


Source: http://graves.cl/radar-chart-d3/ and http://bl.ocks.org/nbremer/6506614

Sensing Multidimensional Data with Parallel Coordinates



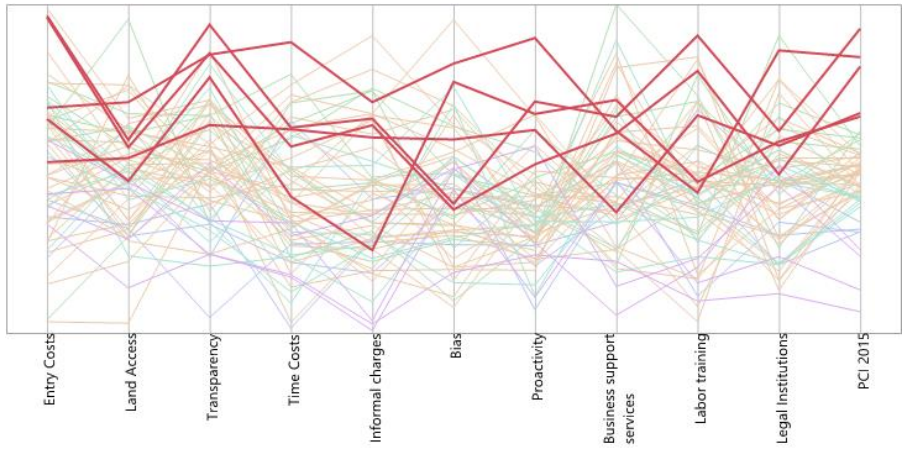
Parallel Coordinates: Brushing



School of Information Systems

SMU Singapore Management University

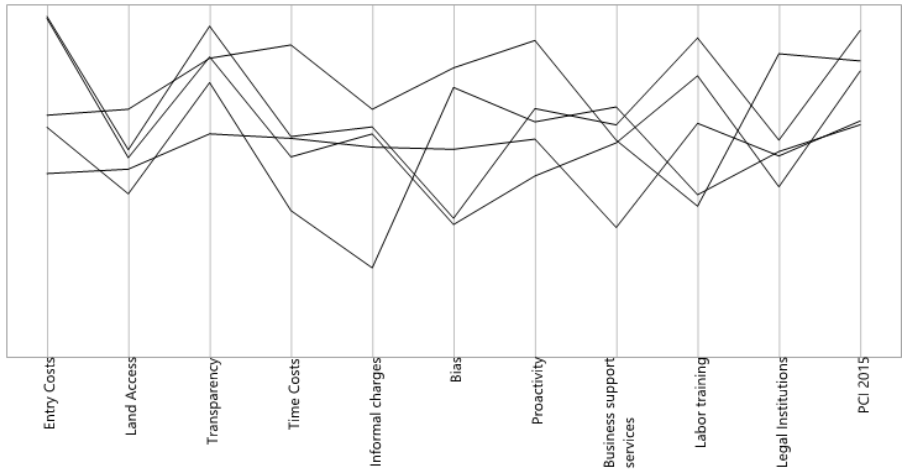
Parallel Coordinates: Colour and Brushing



School of Information Systems

SMU Singapore Management University

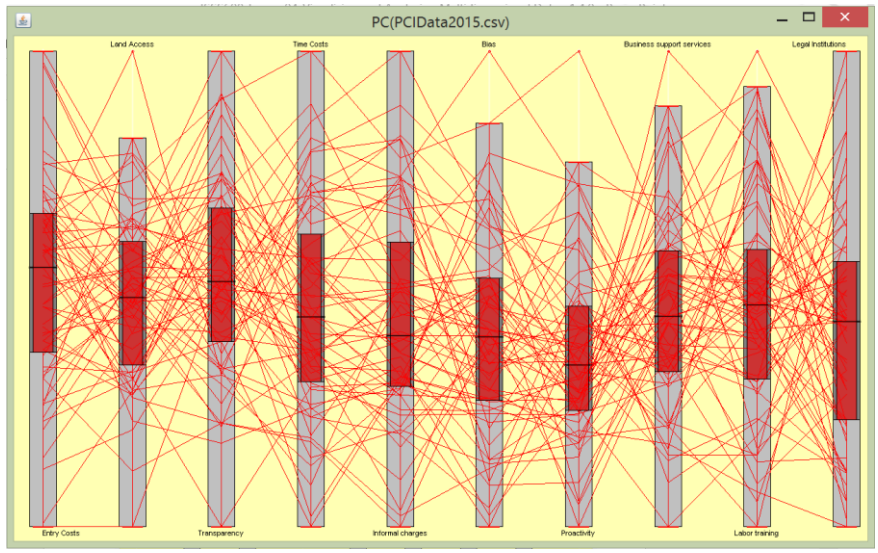
Parallel Coordinates: Filtering



School of Information Systems

SMU Singapore Management University

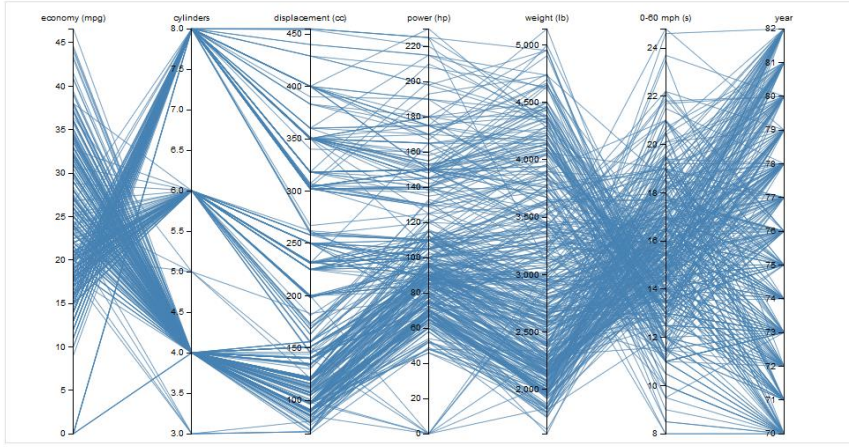
Parallel Coordinates and boxplot



School of Information Systems

SMU Singapore Management University

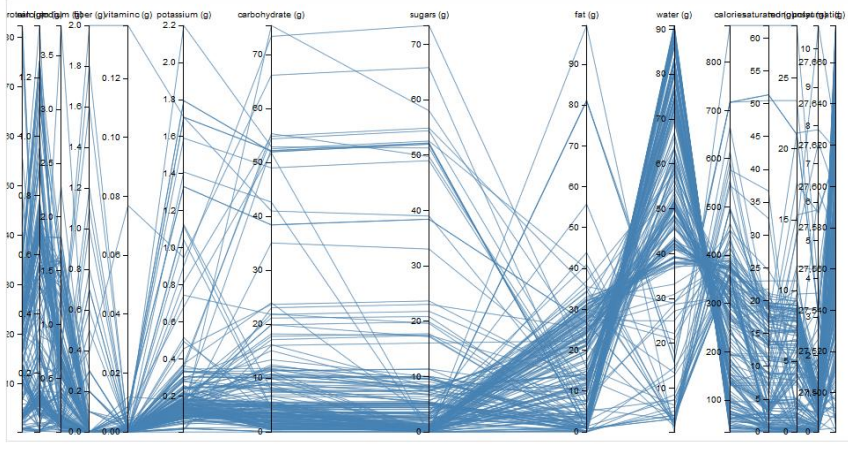
Parallel Coordinates in d3.js



School of Information Systems

SMU Singapore Management University

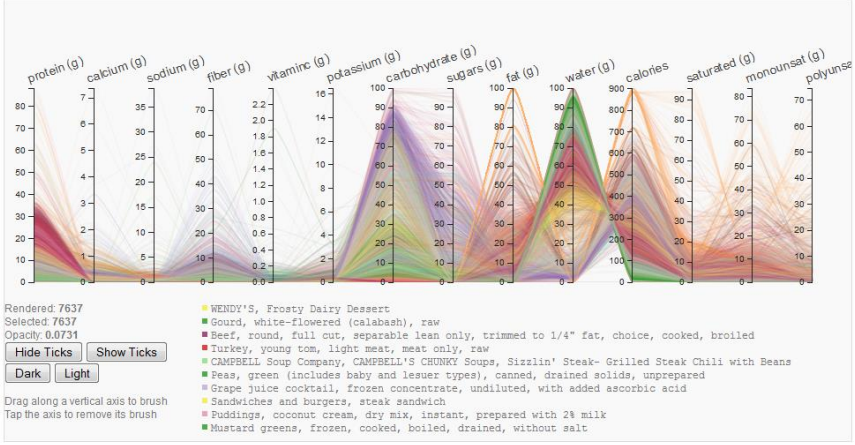
Parallel Coordinates with Fisheye



School of Information Systems

SMU Singapore Management University

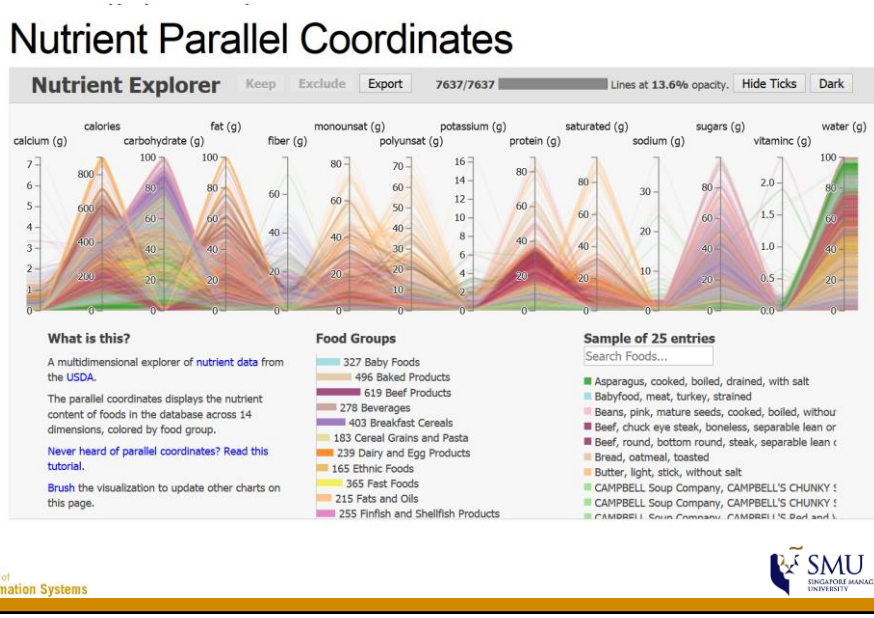
Parallel Coordinates - Shuffled Rendering



School of Information Systems

SMU Singapore Management University

Data Visualisation Application



School of Information Systems

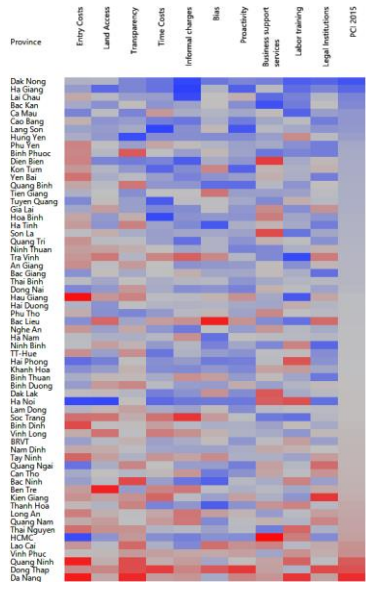
SMU Singapore Management University

Sensing Multidimensional Data with Heatmap

School of Information Systems

SMU Singapore Management University

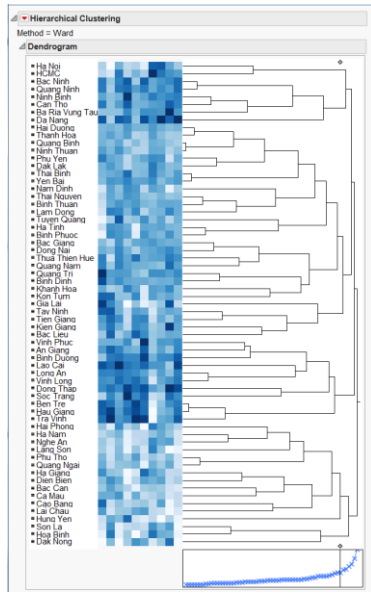
Heatmap: Sorting



School of Information Systems

SMU Singapore Management University

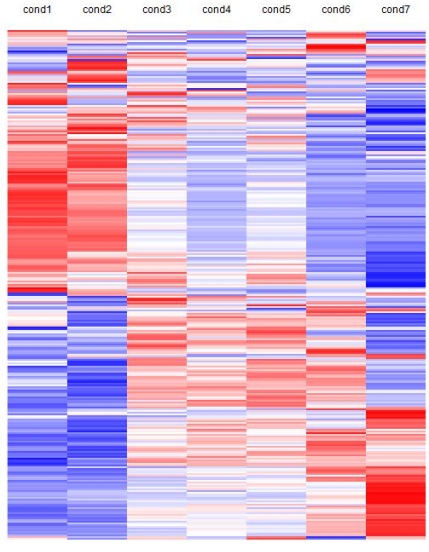
Heatmap + Dendrogram



School of Information Systems

SMU Singapore Management University

Heatmap library in D3.js

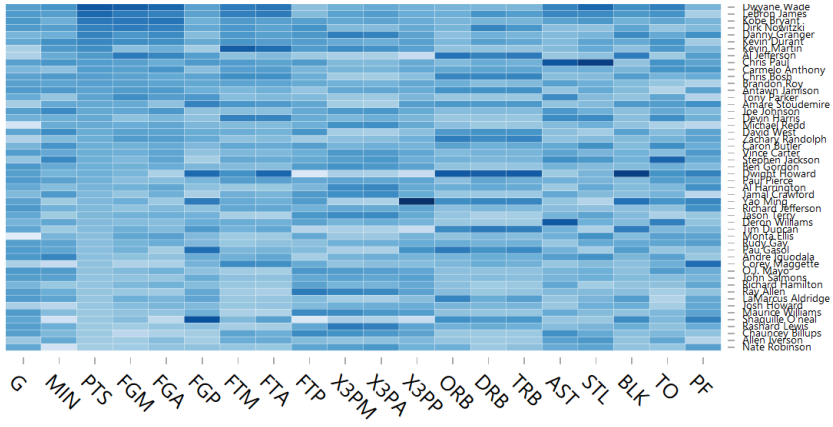


Source: http://blog.nextgenetics.net/?e=44

School of Information Systems

SMU Singapore Management University

d3heatmap: An interactive heat maps



School of Information Systems

SMU Singapore Management University

Multidimensional Categorical Data

- The Titanic Disaster

	Class	Age	Sex	Survived
1	First	Adult	Male	Yes
2	First	Adult	Male	Yes
3	First	Adult	Male	Yes
4	First	Adult	Male	Yes
5	First	Adult	Male	Yes
6	First	Adult	Male	Yes
7	First	Adult	Male	Yes
8	First	Adult	Male	Yes
9	First	Adult	Male	Yes
10	First	Adult	Male	Yes
11	First	Adult	Male	Yes
12	First	Adult	Male	Yes
13	First	Adult	Male	Yes
14	First	Adult	Male	Yes
15	First	Adult	Male	Yes
16	First	Adult	Male	Yes
17	First	Adult	Male	Yes
18	First	Adult	Male	Yes
19	First	Adult	Male	Yes
20	First	Adult	Male	Yes
21	First	Adult	Male	Yes
22	First	Adult	Male	Yes
23	First	Adult	Male	Yes
24	First	Adult	Male	Yes
25	First	Adult	Male	Yes

School of Information Systems

SMU Singapore Management University

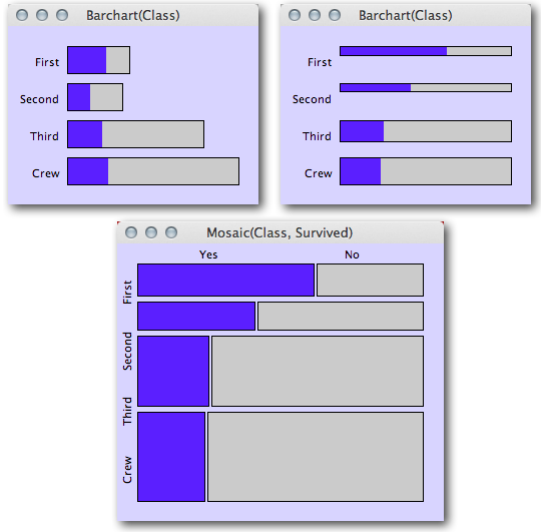
The Titanic Disaster - the tragedy

- The general rule
  - Women and children first

Source: http://en.wikipedia.org/wiki/Sinking_of_the_RMS_Titanic

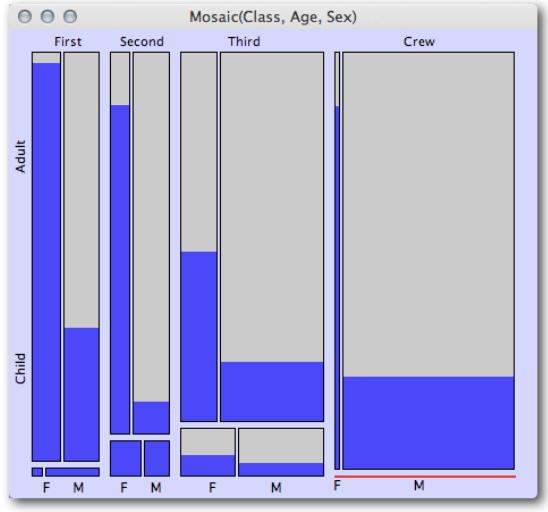
Mosaic plots

- Univariate and bivariate mosaic plot



Mosaic Plots

- Three variables

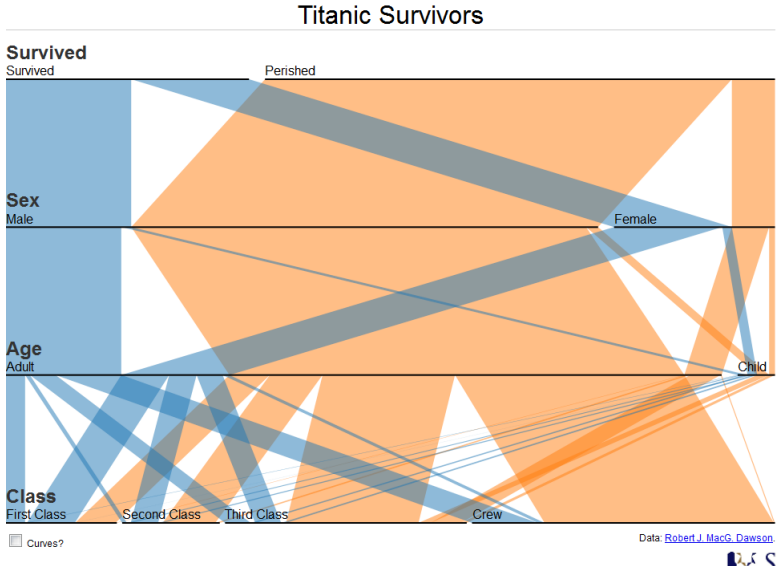


Source: http://www.theusrus.de/blog/understanding-mosaic-plots/

School of Information Systems

SMU Singapore Management University

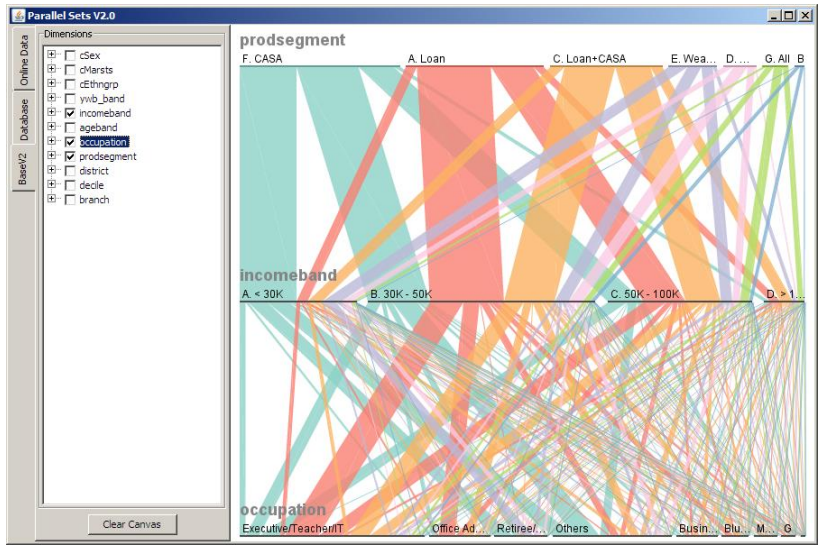
Parallel Sets – d3.js



School of Information Systems

SMU Singapore Management University

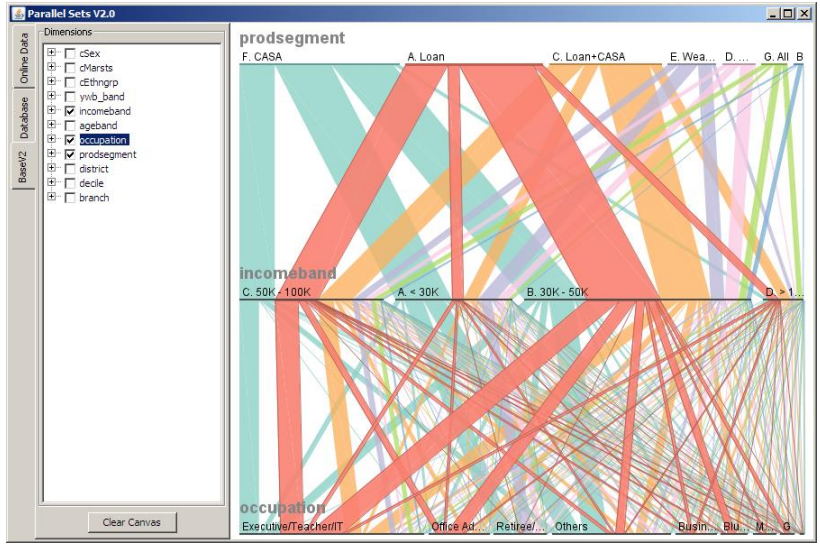
Visualising Multidimensional Categorical Data with Parallel Sets



School of Information Systems

SMU Singapore Management University

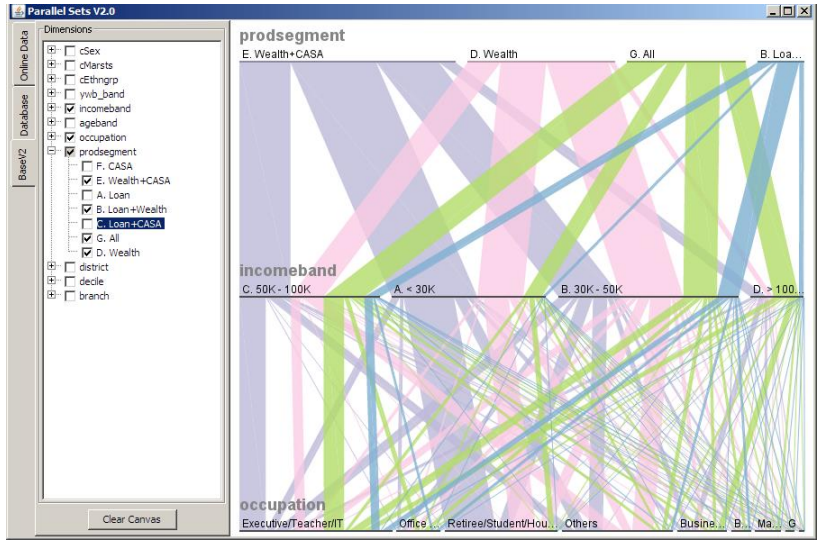
Exploring Multiudimensional Categorical Data with Highlighting Technique



School of Information Systems

SMU Singapore Management University

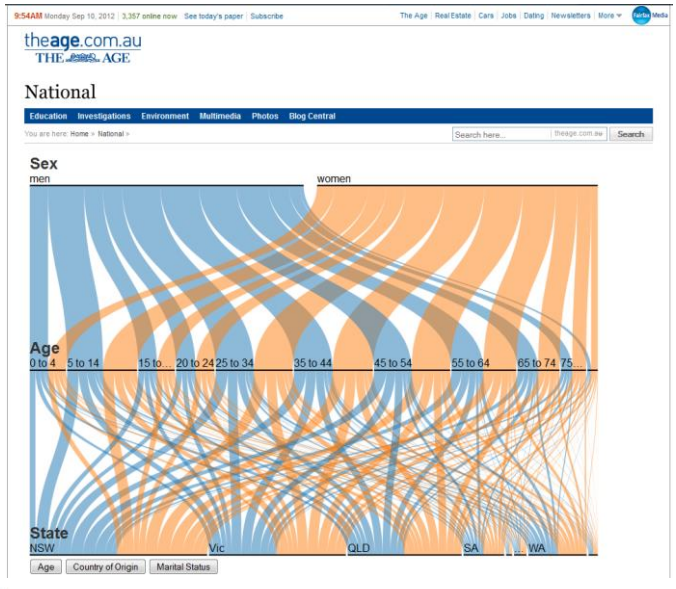
Exploring Multidimensional Categorical Data with Filtering Technique



School of Information Systems

SMU Singapore Management University

Data Visualisation Application – Parallel Sets

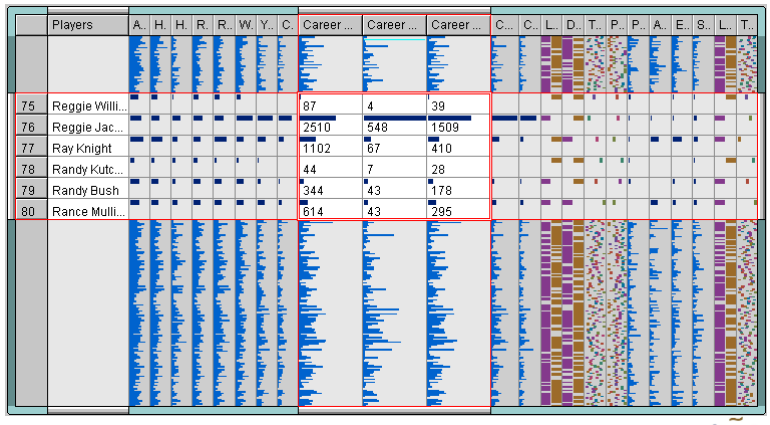


School of Information Systems

SMU Singapore Management University

TableLens

- Provides a structured graphical representation that supports browsing of the values for hundreds of cases and tens of variables on typical workstation display.



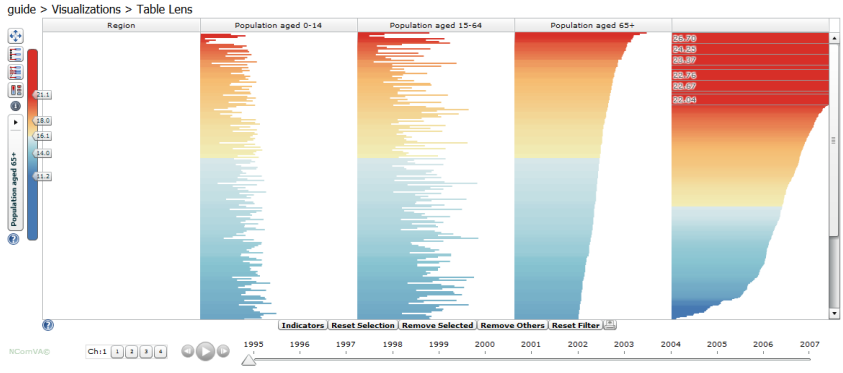
School of Information Systems

SMU Singapore Management University

TableLens

Table Lens

guide > Visualizations > Table Lens



Source: http://ncva.itn.liu.se/education-geovisual-analytics/table-lens?l=en and http://www.ncomva.se/guide/?chapter=Visualizations&section=Table%20Lens#_General

School of Information Systems

SMU Singapore Management University

TableLens in D3.js

- DataComb: An interface for combing through tabular datasets

Source: http://www.bytemuse.com/post/data-comb-visualization/#demo

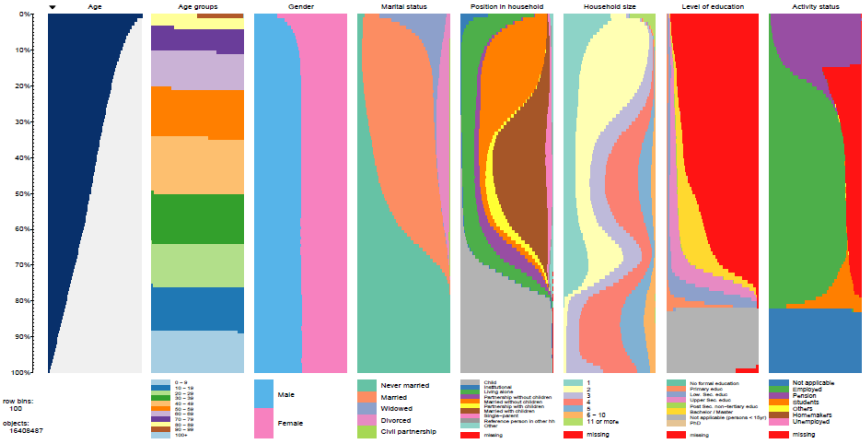
School of Information Systems

SMU Singapore Management University

41

Tableplots

- Multidimensional continuous data



Source: Visualizing and Inspecting Large Datasets with Tableplots (www.jds-online.com/file_download/379/jds-1108.pdf)

School of Information Systems

SMU Singapore Management University

42

References

- Chernoff face (https://en.wikipedia.org/wiki/Chernoff_face)
- Chernoff face in d3.js (https://github.com/gnarmis/chernoff-faces) and (http://bl.ocks.org/larskotthoff/2011590)
- Radar Chart (http://en.wikipedia.org/wiki/Radar_chart)
- Ternary plot (https://en.wikipedia.org/wiki/Ternary_plot), (http://csmres.jmu.edu/geollab/Fichter/SedRx/readternary.html) and (http://wvaughan.org/ternaryplots.html)
- ggtern 1.0.3.1 on CRAN
- Ternary Plot in D3.js (http://bl.ocks.org/tomgp/7674234), (https://gist.github.com/widged/5780720), (http://bl.ocks.org/tomgp/7766353)
- Radar Chart (http://en.wikipedia.org/wiki/Radar_chart)