TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Laura Bassi Centre of Expertise
Centre for Visual Analytics Science & Technology

Visual Analytics of Large Matrix Data

Bilal Alsallakh – bilal@cvast.tuwien.ac.at

Large Matrix Data

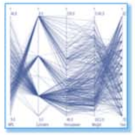
$$F : E \times C \rightarrow \mathbb{R}^{+} : (e_i, c_j) \rightarrow f_{ij}\ ,\ m = |C| \ll n = |E|$$

- E is a large set (thousands) of entities such as movies. The entities can have ***attributes*** such as genres.
- C is a small set (tens) of columns that represent categories, classes, labels or tags.
- F is a bivariate function that defines how the entities E are related to the columns C.

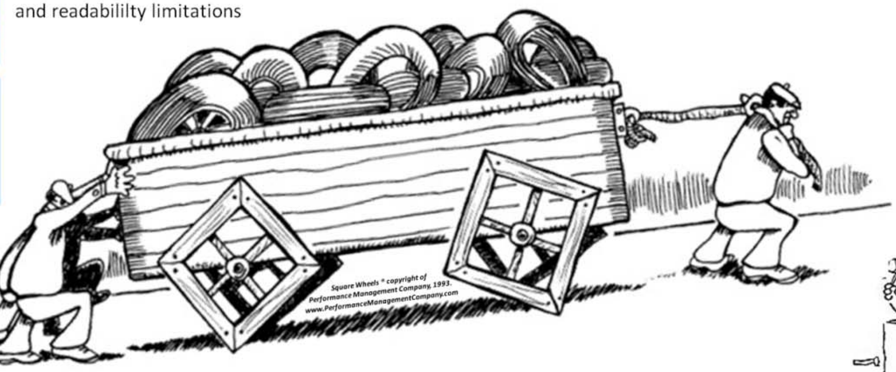
Example: a large 3706 x 21 contingency table

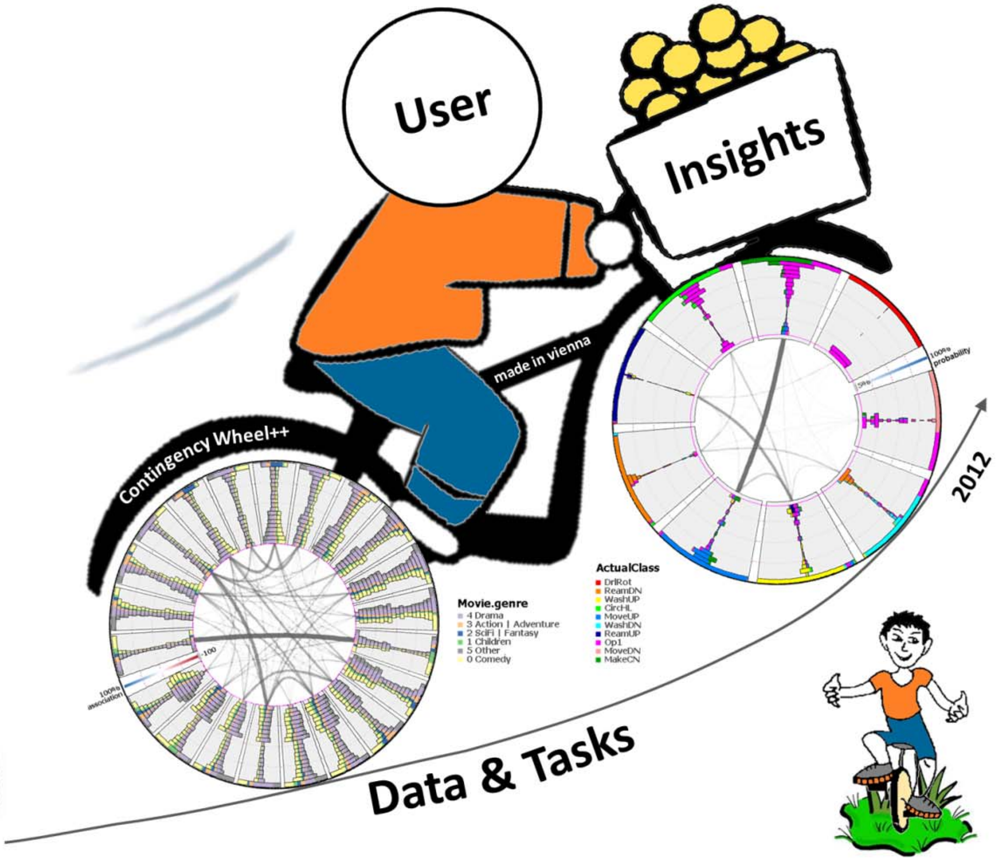
	C occupations / E movies	K-12 student	self-employed	scientist	executive	writer	homemaker	academic/educator	programmer	technician/eng.	other	clerical/admin	sales/marketing	college/grad stud.	lawyer	farmer	unemployed	artist	tradesman	customer service	retired	doctor/health care	f_{i+}
1	One Flew Over the Cuckoo	25	79	47	196	96	23	191	107	109	195	59	81	200	39	3	20	86	22	33	42	72	1725
2	James and the Giant Peach	29	19	10	42	39	10	35	29	33	69	19	20	75	9	1	7	35	4	10	7	23	525
3	My Fair Lady (1964)	19	30	15	62	43	17	81	32	34	75	27	25	66	14	0	8	38	1	8	15	26	636
4	Erin Brockovich (2000)	41	57	26	165	59	18	126	75	101	131	38	81	188	30	1	13	49	9	18	39	50	1315
5	Bug's Life, A (1998)	72	73	49	159	80	32	103	117	145	215	52	94	247	23	4	25	78	18	33	19	65	1703
:	: : :	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		:
3703	Broken Vessels (1998)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
3704	White Boys (1999)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
3705	One Little Indian (1973)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3706	Five Wives, Three Secretar	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Data collected by GroupLens Research www.grouplens.org f_{+j}	23290	46021	22951	105425	60397	11345	85351	57214	72816	130499	31623	49109	131032	20563	2706	14904	50068	12086	21850	13754	37205	f_{++}

State of the Art



Both **heat maps** and **Parallel Coordinates** are applicable, but might impose scalability and readabililty limitations





Research Questions

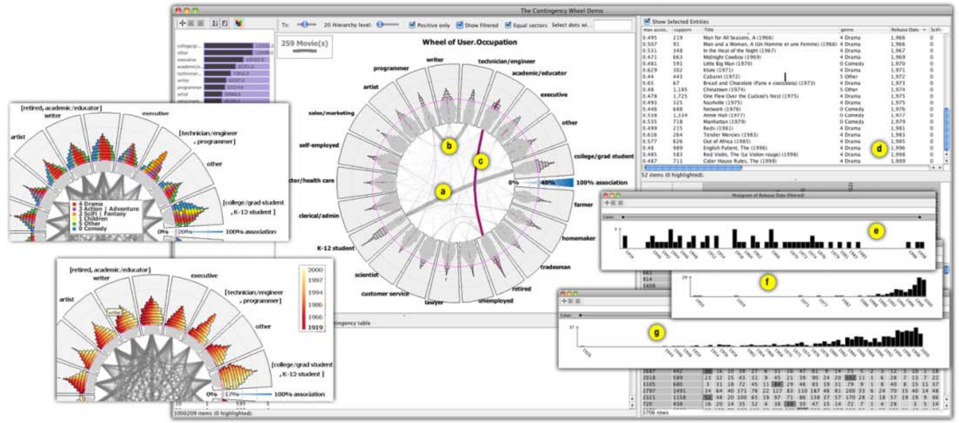
Main hypothesis: Visual Analytics provides new insights into patterns difficult to find via automated methods

- **Q1**: How to analyze the relations between the row entities and the columns in the light of the entity attributes?
- **Q2**: How to analyze the similarity between columns based on their relations with the rows?
- **Q3**: How to analyze the changes of these relations and similarities over time, and relate them the to attributes?

Visual Analytics Approach

The proposed approach consists of three tightly integrated components:

- **Automated methods** to compute important information in the data (high associations and similarities).
- **Visual methods** to depict the associations and similarities using family representations (colored histograms, arcs).
- An **exploration environment** to enable easy access to all available data via a multi-level overview+detail interface.



Video and online demos: http://www.cvast.tuwien.ac.at/wheel