

„Geo Data Mining and Visual Analytics“

„Beyond Limits – Developments in Cadastral Domain“

Workshop, Zürich

19 March 2015

Susanne Bleisch

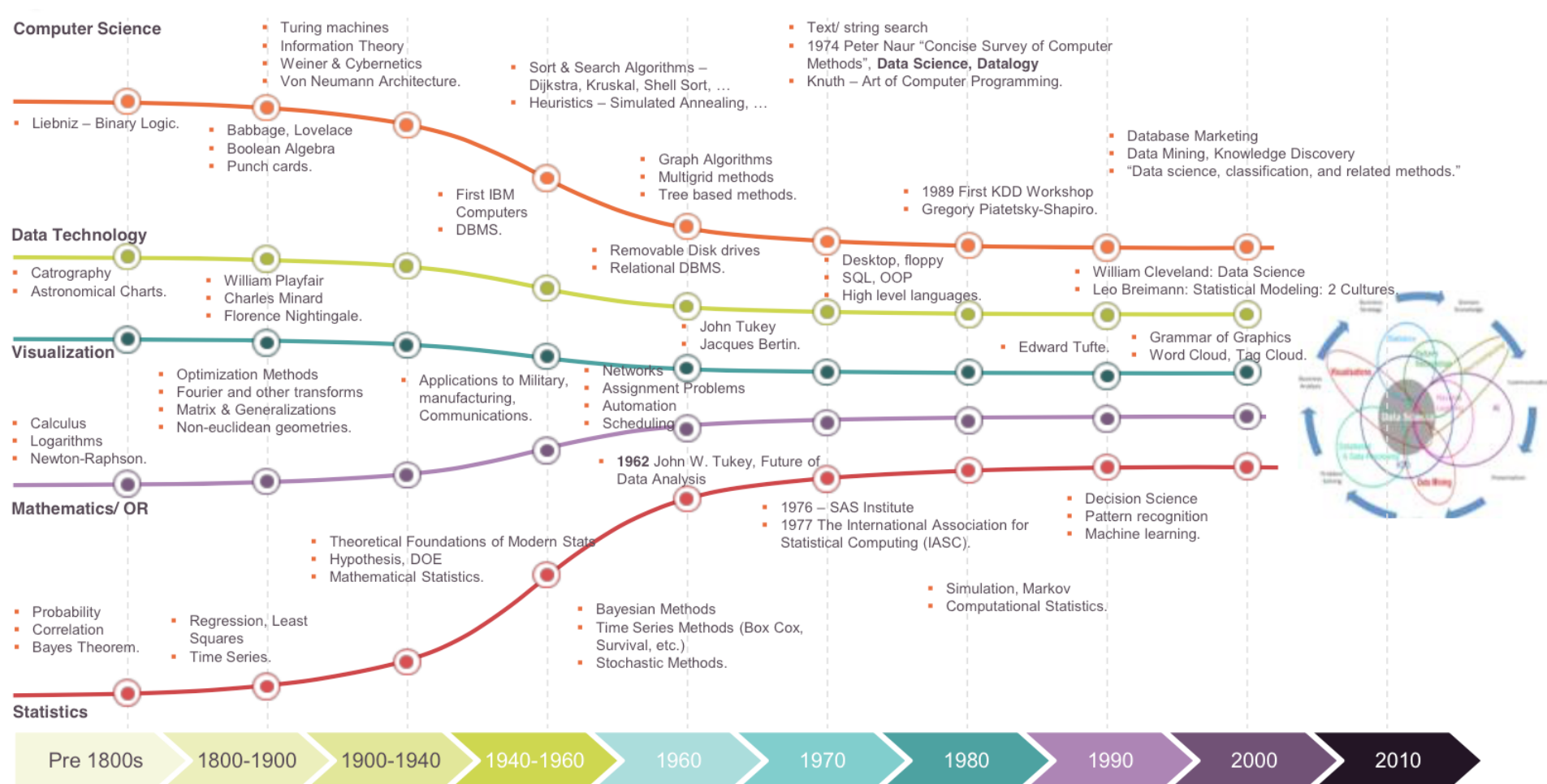
Institute of Geomatics Engineering

School of Architecture, Civil Engineering and Geomatics HABG

FHNW University of Applied Sciences and Arts Northwestern Switzerland

Define - Exemplify - Use

Data Science



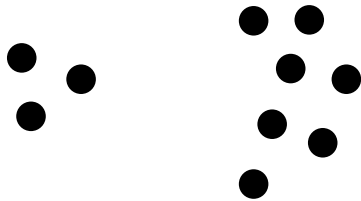
Statistics - Data mining - Machine Learning - Artificial Intelligence

(Geographic) Data Mining

- ◆ **Data- and computation-poor to data- and computation-rich environments**
- ◆ **Traditional (spatial) analytical methods developed for small and homogenous data sets**
- ◆ **From confirmatory (a priori hypotheses) to exploratory (gain insight)**
- ◆ **Find hidden information in the data: novel, valid, useful and understandable**

Typical data mining tasks

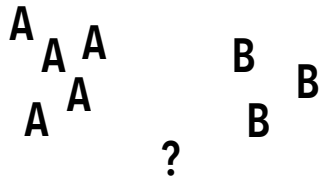
Clustering



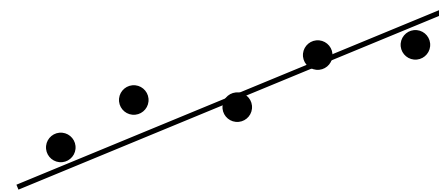
Outlier analysis



Classification



Modelling/Prediction



Association analysis

{wine, cheese} -> {bread}

Sequence analysis

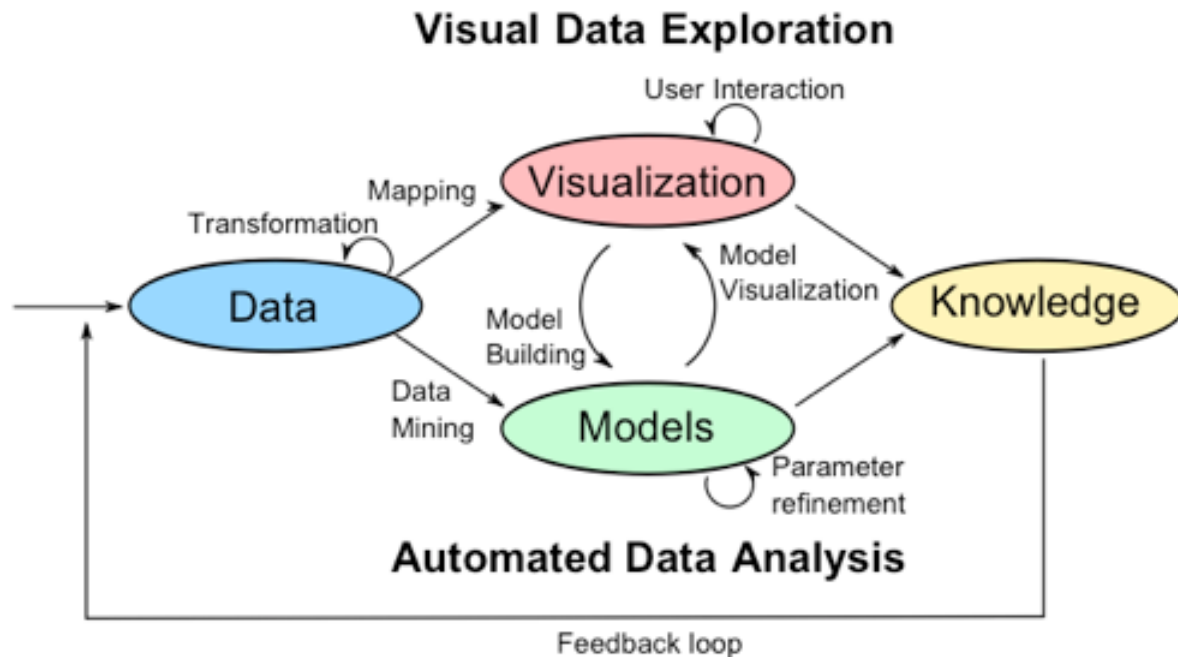


Visual Analytics

Humans in the loop

Visual Analytics is the science of analytical reasoning supported by interactive visual interfaces.

„... detect the expected and discover the unexpected...”

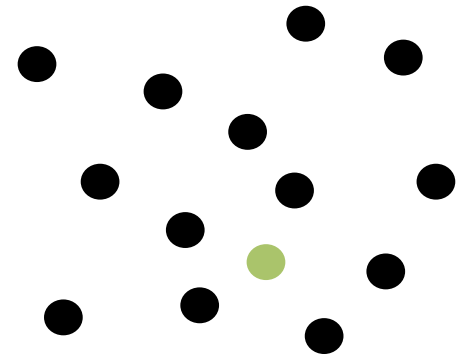


Thomas, J.J. & Cook, K.A., 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE Computer Society.

Figure: <http://www.visual-analytics.eu/faq/>

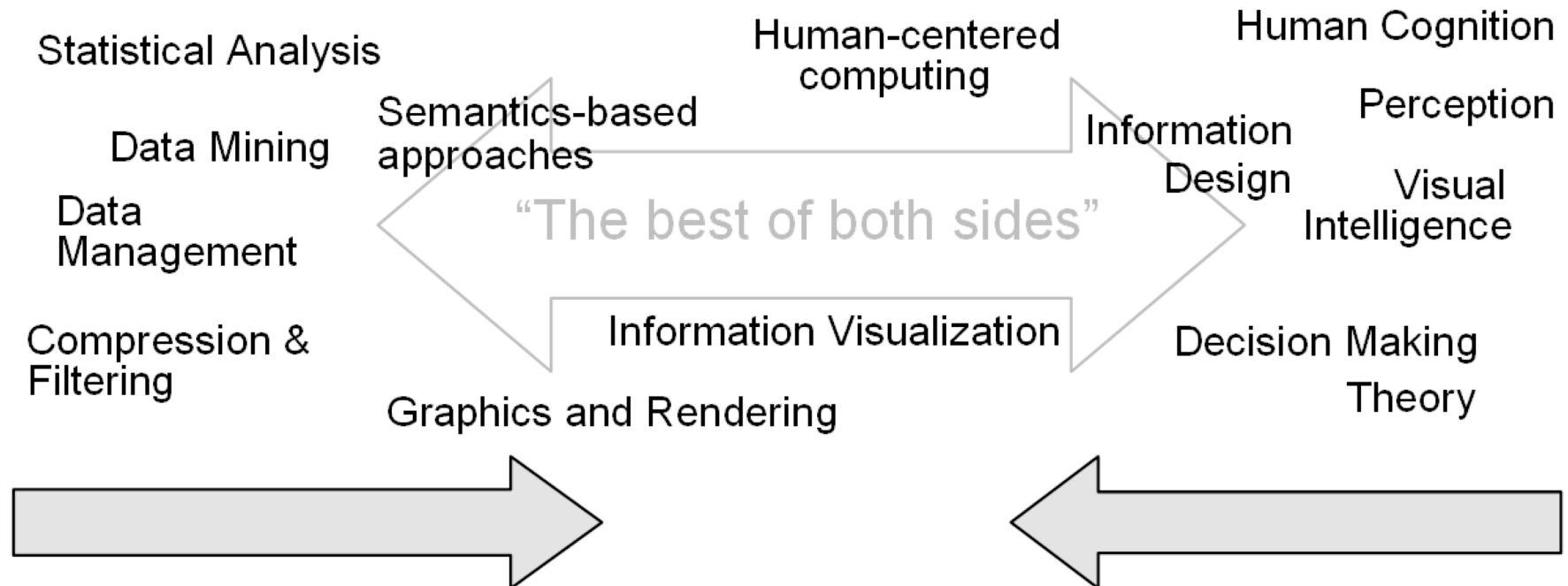
Visual Analytics

781	300	465	975	110	111	591	299
30	417	593	571	856	114	355	196
421	38	875	885	19	31	252	664
545	659	937	57	690	285	859	388
552	301	863	33	559	424	307	26
276	812	94	298	507	621	713	615
331	502	871	260	641	350	968	510
332	837	964	382	894	317	800	782
110	169	116	548	829	815	997	708
213	655	953	631	793	50	564	859
433	355	354	220	167	160	648	458
511	418	315	944	534	105	290	790
776	85	121	794	173	572	166	960
289	468	44	206	759	62	395	915
137	27	632	747	575	957	782	586
772	387	338	332	665	820	500	898



Machine

Human



Data

- ◆ Textual data: documents, news, email, tweets,...
- ◆ Databases and collections
- ◆ Image data: from satellite to user generated
- ◆ Sensor data
- ◆ Video data
- ◆ Networks, linked data
- ◆ Structured, semi-structured, unstructured
- ◆ ...

Challenges:

- ◆ Heterogeneous sources and types
- ◆ massive amounts but incomplete, inconsistent, time-varying and potentially inaccessible,...

Spatial is special

- ◆ **GI is uncertain, imprecise or vague.**
- ◆ **GI is correlated, i.e. non-random.**
- ◆ **GI is dynamic, it represents an ever-changing world.**
- ◆ **GI is scale dependent.**
- ◆ **GI is heterogeneous.**

Define - Exemplify - Use

reCAPTCHA

ZKVV 4

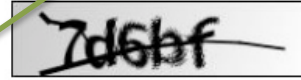
[Megaupload]



[Reddit]

944 531

[eBay]



[CNN]

RAE3

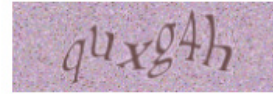
[Baidu]

3nc9z

[Authorize]

3-2 parks

[Recaptcha]



[Skyrock]

adv ses

[Captcha.net]

2CC EX

[NIH]



[Digg]

skthngomi

[Google]

trustothr

[Wikipedia]



[Slashdot]



[Blizzard]



I'm not a robot

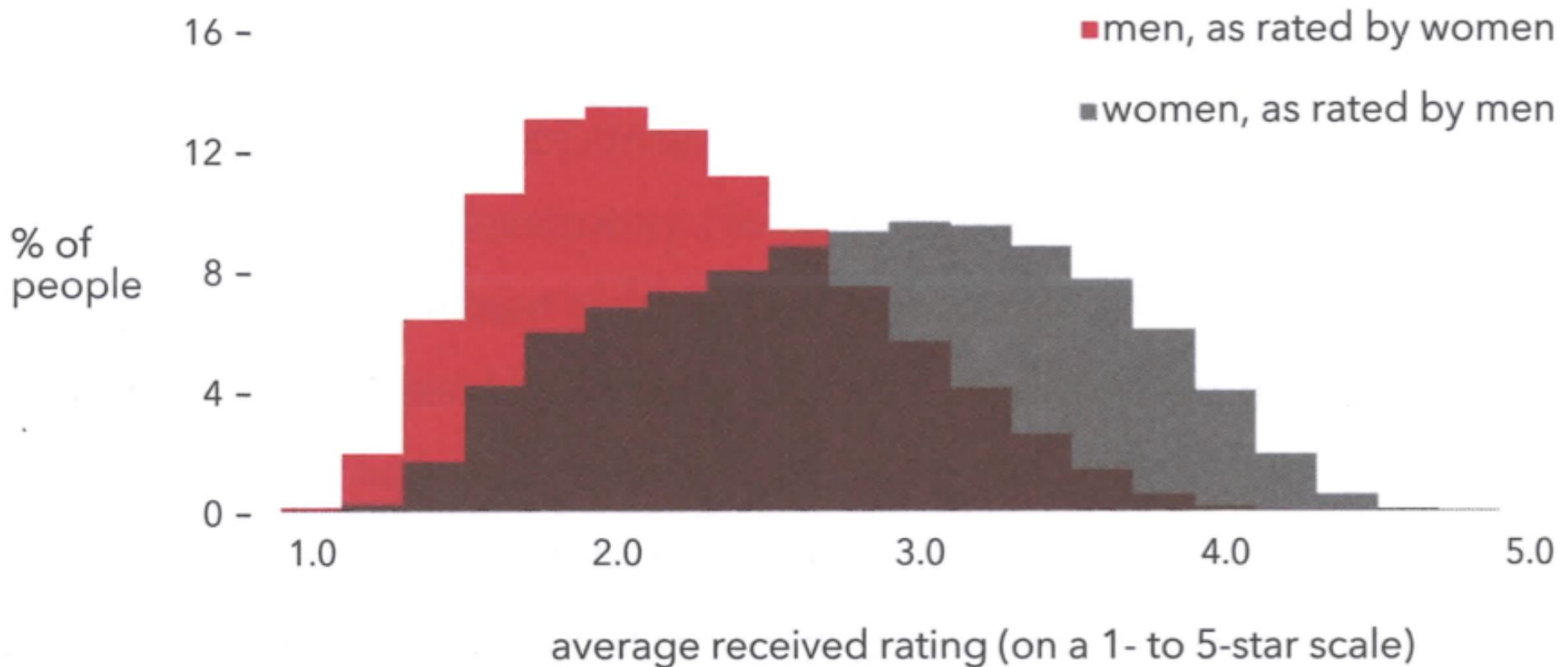


reCAPTCHA

[Privacy](#) - [Terms](#)

Information from small pieces of data

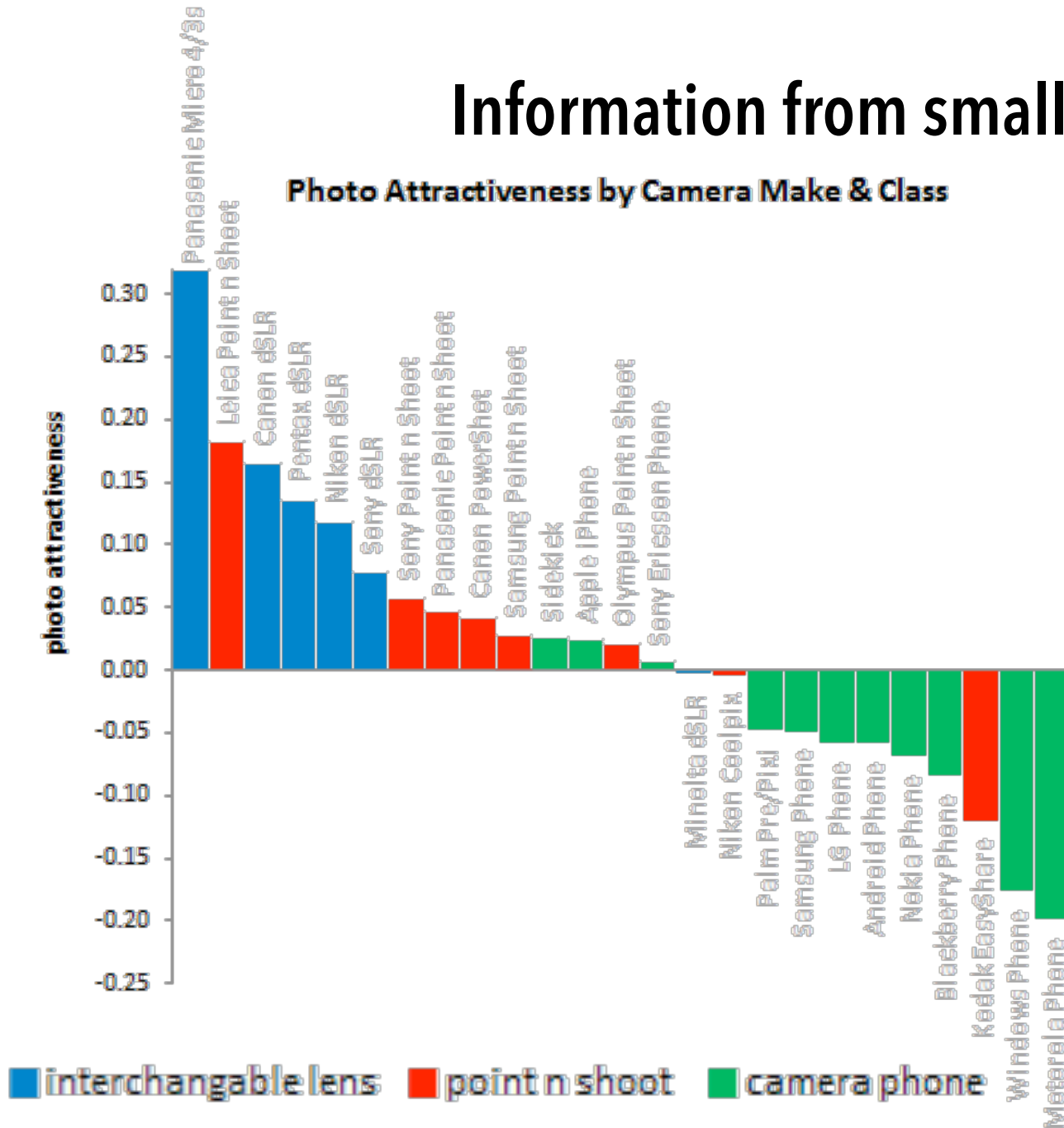
perception of male attractiveness vs. female attractiveness



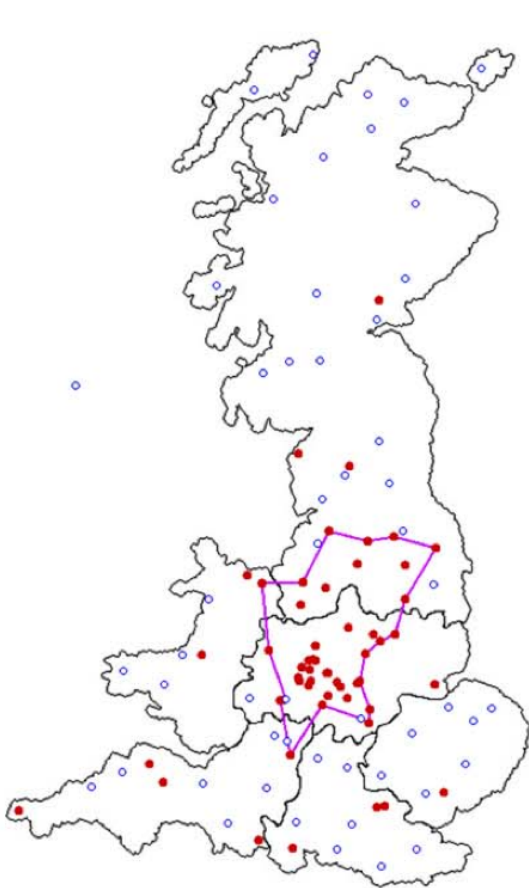
Rudder, C. (2014). *Dataclysm: Who We Are (When We Think No One's Looking)*. London, UK: Crown/Archetype.

Information from small pieces of data

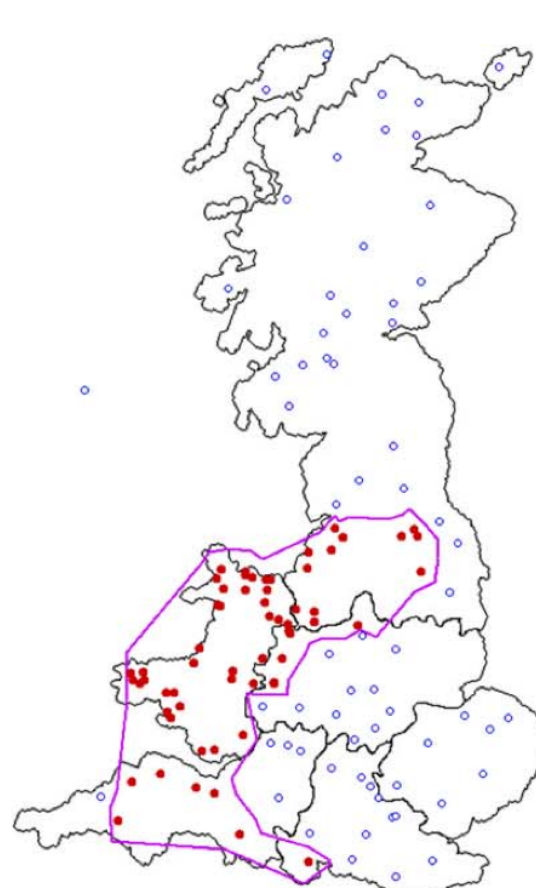
Photo Attractiveness by Camera Make & Class



Geographic information retrieval



Midlands

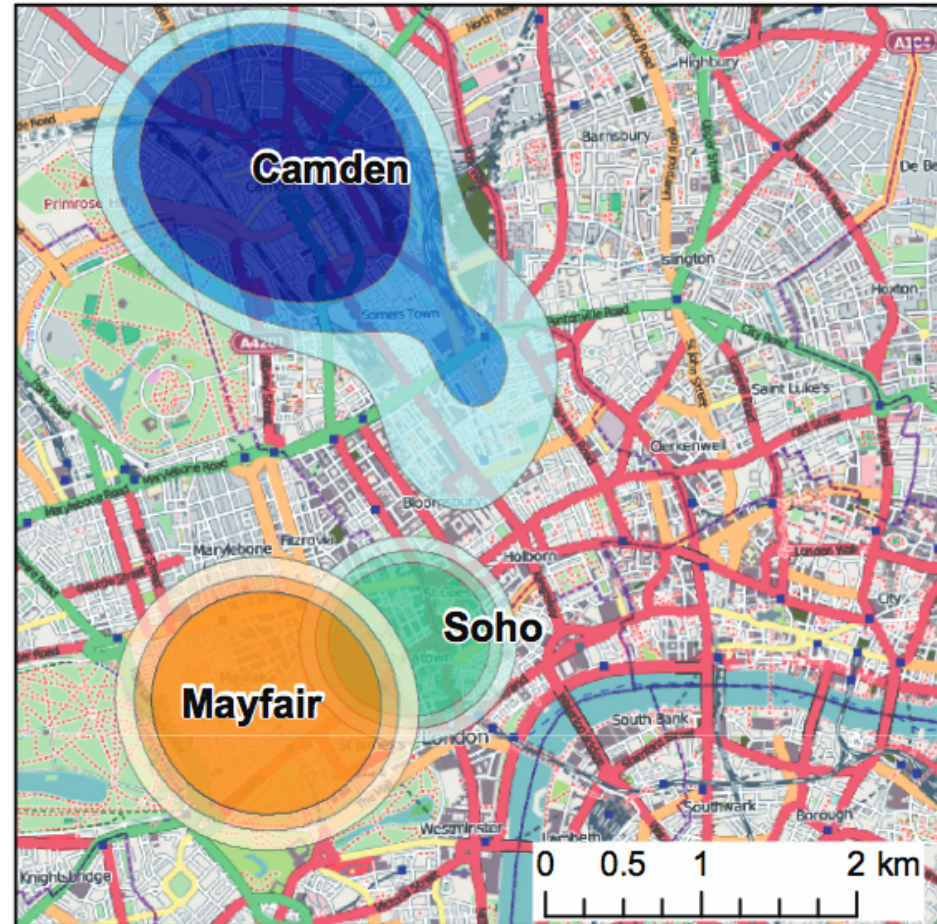
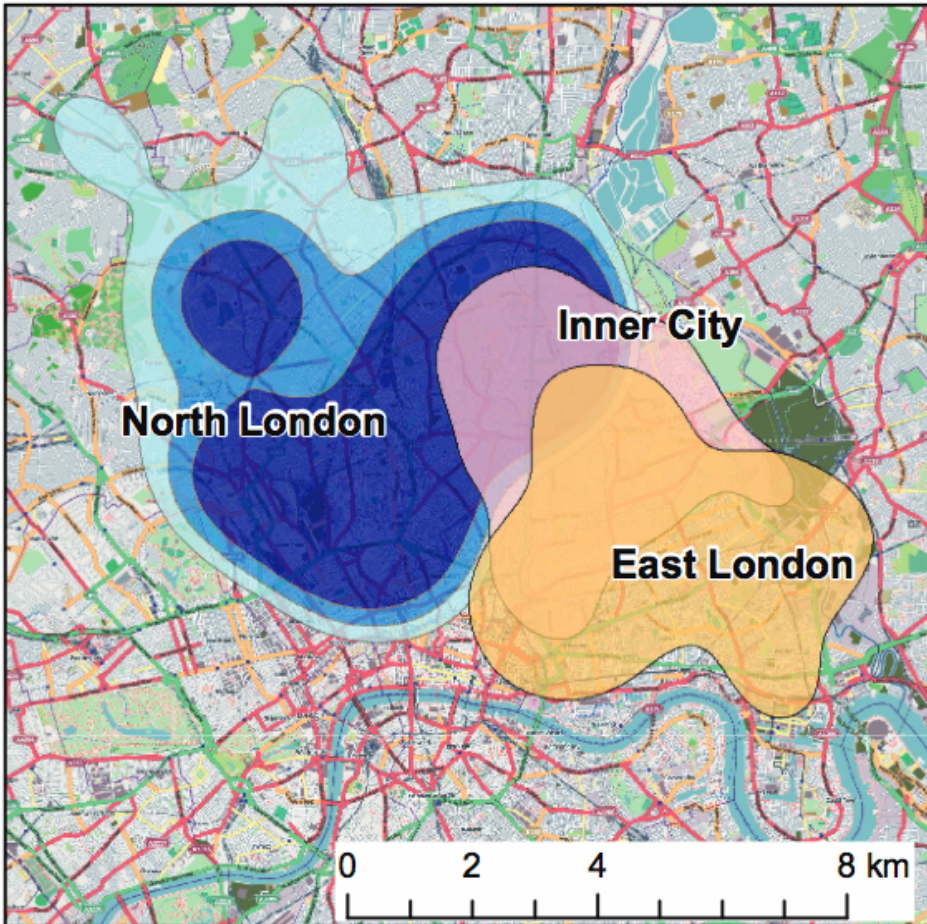


Wales

Arampatzis, A., van Kreveld, M., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P., Hideo, J. & Sanderson, M. (2006).

Web-based delineation of imprecise regions. *Computers, Environment and Urban Systems*, 30, 436–459.

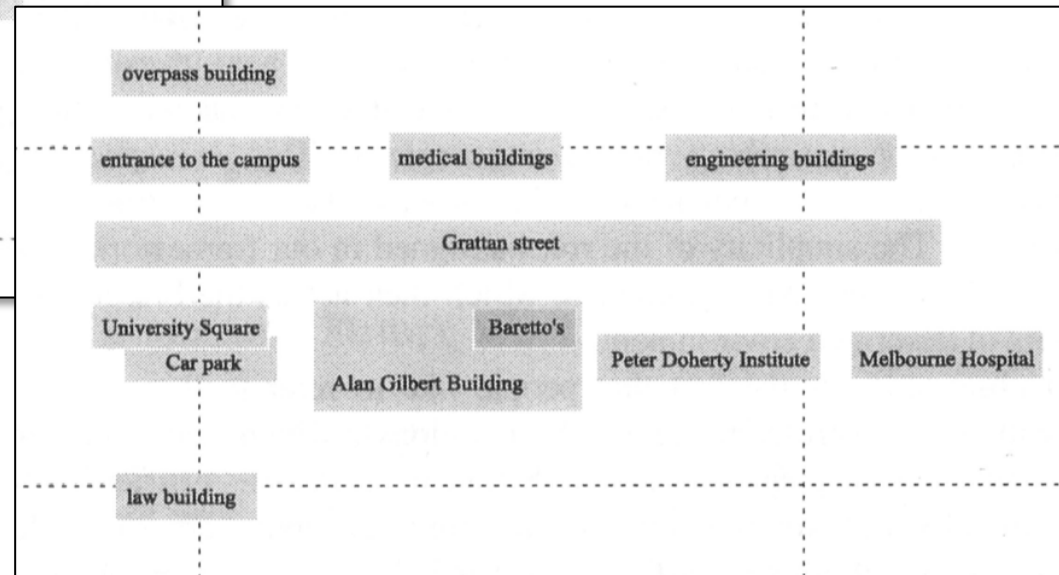
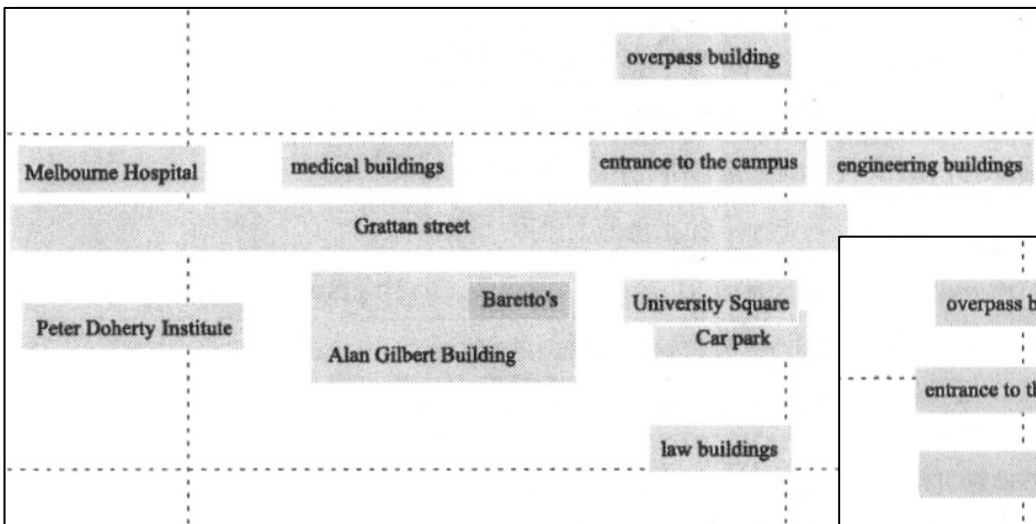
Geographic information retrieval

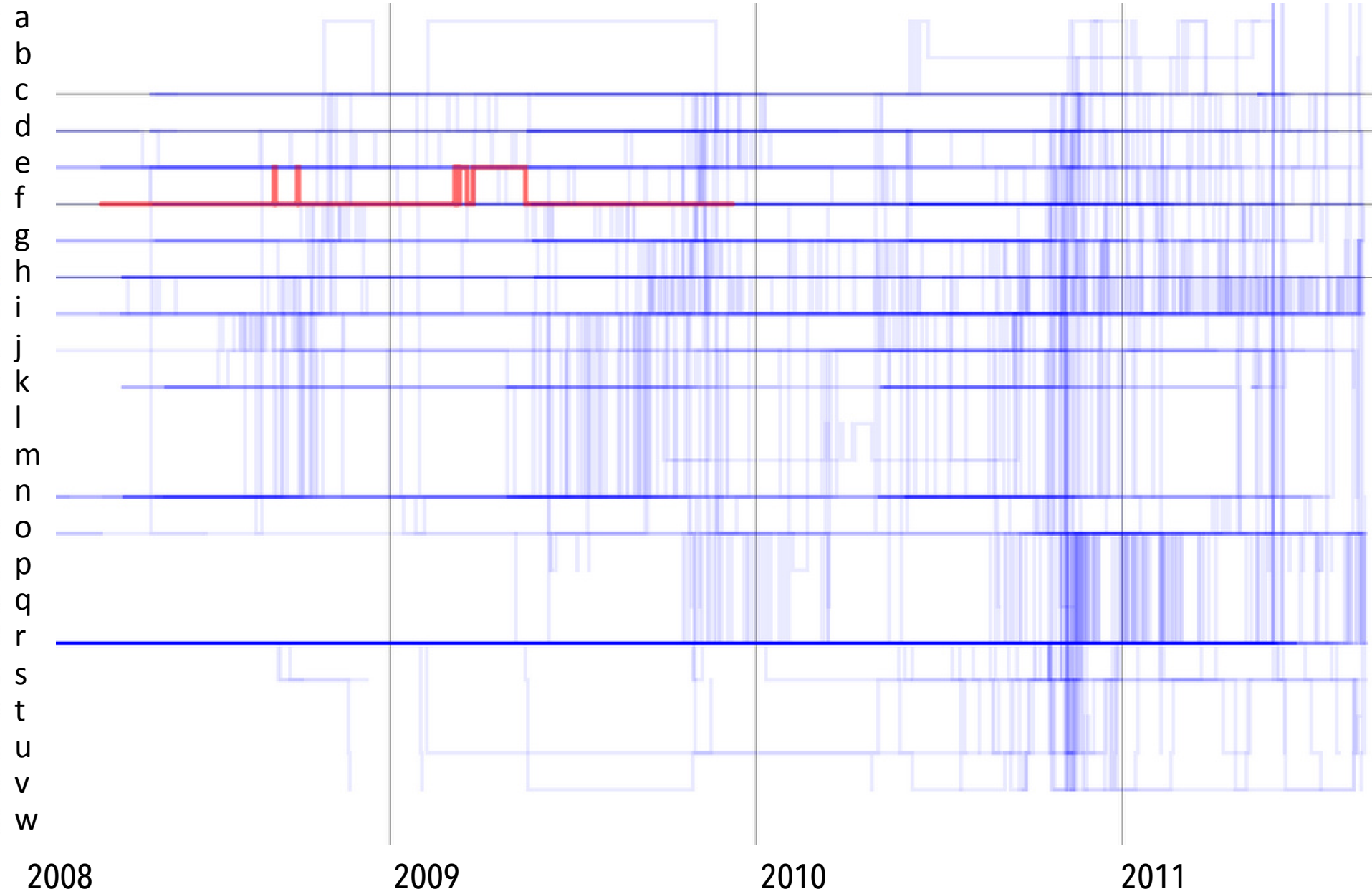


Hollenstein, Livia; Purves, Ross S (2010). Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*, 1(1):21-48.

Descriptions > maps

„We're sitting in Baretto's in the Alan Gilbert Building, across Grattan street is one of the medical buildings. Down the hill along Grattan street the new building being constructed is the Peter Doherty Institute and diagonally across the road (Royal Parade) is Melbourne Hospital. In the other direction..."





Event

causes

Event

Full moon

causes

Start of fish migration

State

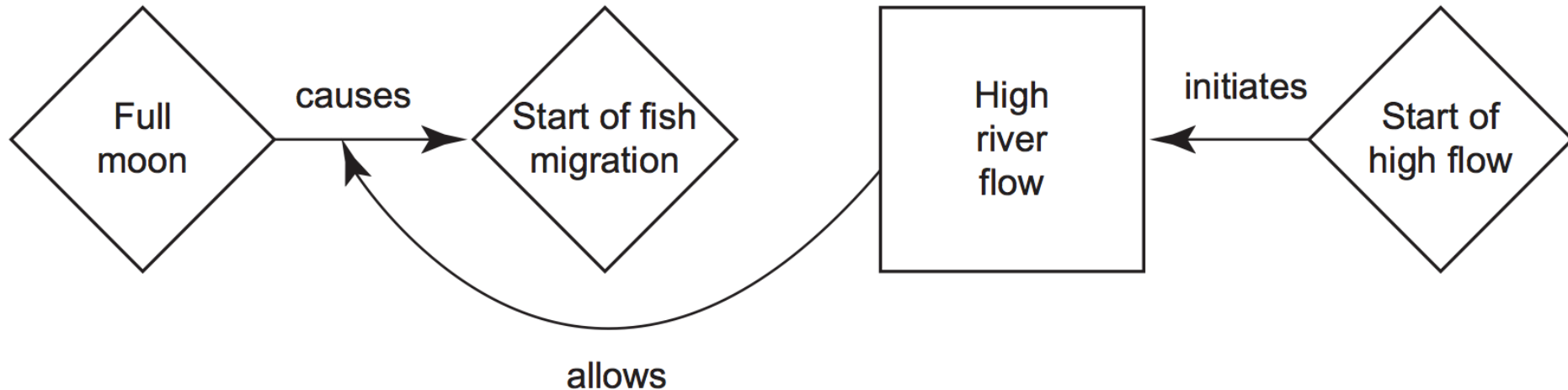
allows

Event

High river flow

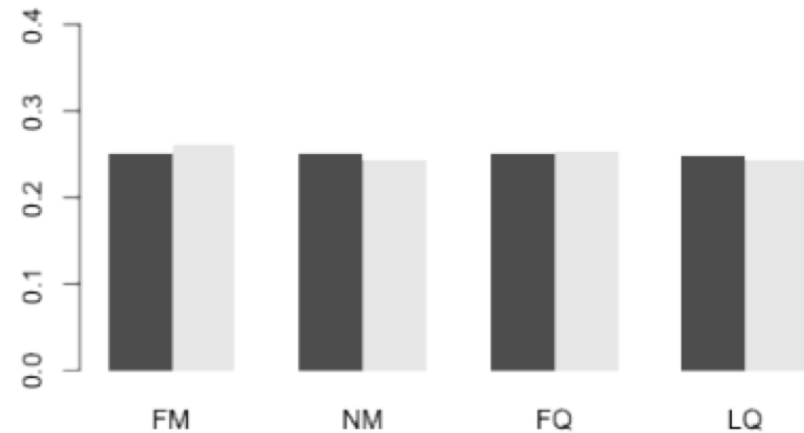
allows

Start of fish migration



Data Mining - Sequence analysis

Fish migration upstream

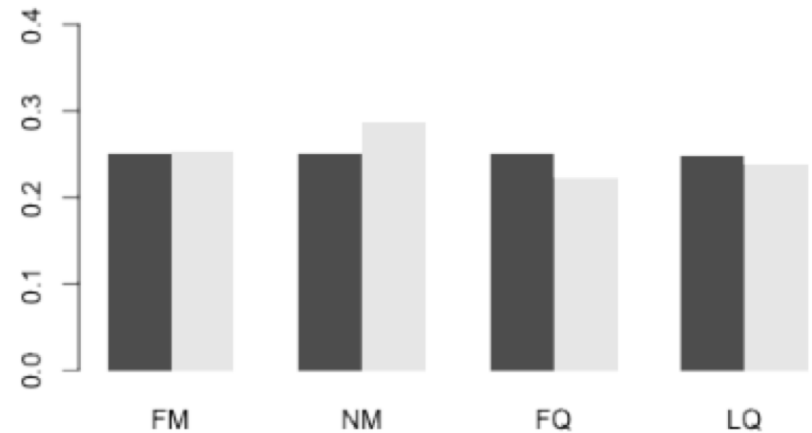


Moon phases

$$\chi^2 = 0.2439, df = 3, \\ p\text{-value} = 0.970$$

■ expected
■ observed

Fish migration downstream

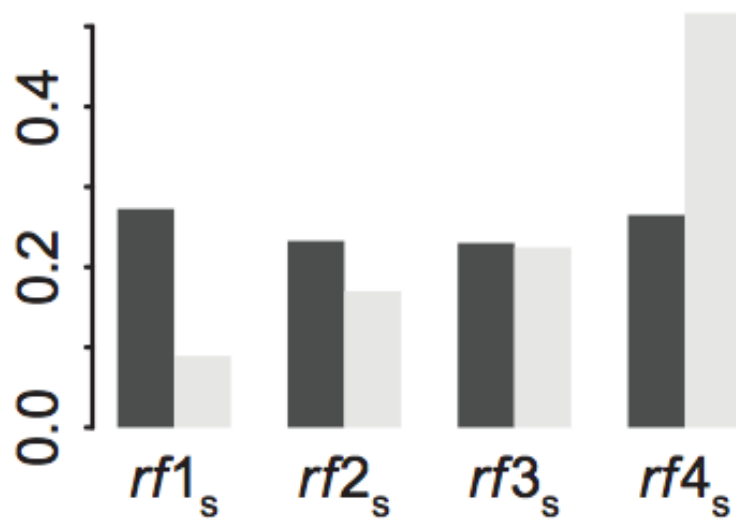


$$\chi^2 = 3.346, df = 3, \\ p\text{-value} = 0.341$$

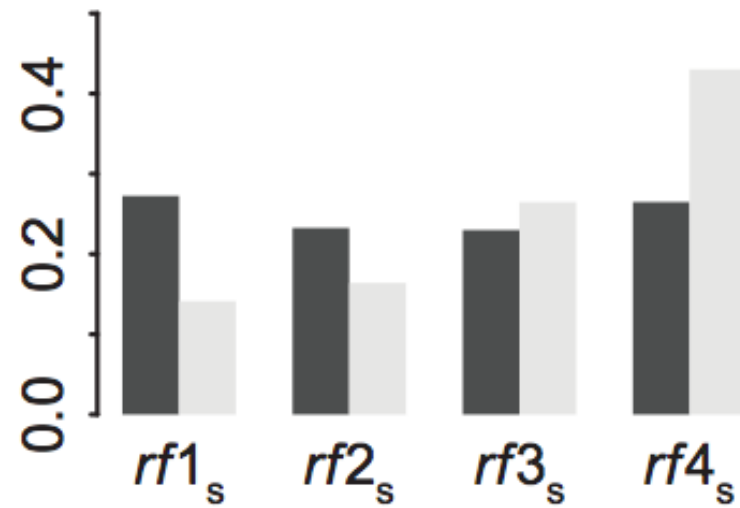
Data Mining - Association analysis (river flow)

Fish migration
upstream

Fish migration
downstream



$X^2 = 229.90$, $df = 3$,
 $P\text{-value} < 0.001$



$X^2 = 118.20$, $df = 3$,
 $P\text{-value} < 0.001$

■ expected
■ observed

Event

causes

Event

-> event sequences

	Subsequence	Support	Count	...			
1	(1)	0.98536210	1279	84	(7) - (7) - (7) - (1)	0.08089368	105
2	(7)	0.50693374	658	85	(7) - (1) - (5) - (1)	0.08012327	104
3	(7) - (1)	0.49152542	638	86	(2) - (7)	0.07935285	103
4	(5)	0.42604006	553	87	(7) - (5) - (7)	0.07935285	103
5	(5) - (1)	0.40369800	524	88	(2) - (5) - (1)	0.07858243	102
6	(1) - (1)	0.38212635	496	89	(5) - (1) - (7)	0.07858243	102
7	(5) - (5)	0.33281972	432	90	(5) - (2)	0.07858243	102
8	(10)	0.31432974	408	91	(1) - (5) - (7) - (5) - (1)	0.07781202	101
9	(5) - (5) - (1)	0.30739599	399	92	(4)	0.07704160	100
10	(10) - (1)	0.29583975	384	93	(10) - (5) - (5)	0.07704160	100
11	(2)	0.24653313	320	94	(2) - (7) - (1)	0.07627119	99
12	(2) - (1)	0.23959938	311	95	(5) - (5) - (10) - (1)	0.07550077	98
13	(5) - (7)	0.23497689	305	96	(5) - (2) - (1)	0.07473035	97
14	(7) - (5)	0.22804314	296	97	(7) - (7) - (1) - (1)	0.07473035	97
15	(5) - (7) - (1)	0.22187982	288	98	(7) - (5) - (7) - (1)	0.07395994	96
16	(7) - (7)	0.21263482	276	99	(4) - (1)	0.07318952	95
17	(7) - (5) - (1)	0.21109399	274	100	(5) - (1) - (5) - (5) - (1)	0.07318952	95
18	(1) - (5)	0.21032357	273	101	(5) - (1) - (7) - (1)	0.07318952	95
19	(1) - (7)	0.20801233	270	102	(9) - (7)	0.07318952	95
20	(7) - (7) - (1)	0.20801233	270	103	(2) - (1) - (1)	0.07241911	94
21	(1) - (7) - (1)	0.19491525	253	104	(5) - (7) - (7) - (5)	0.07241911	94
22	(5) - (7) - (5)	0.19029276	247	105	(10) - (5) - (5) - (1)	0.07241911	94
23	(1) - (5) - (1)	0.18489985	240	106	(1) - (3)	0.07164869	93
24	(5) - (7) - (5) - (1)	0.17796610	231	107	(5) - (5) - (1) - (5) - (1)	0.07087827	92
25	(7) - (1) - (1)	0.17642527	229	108	(1) - (3) - (1)	0.06933744	90
26	(5) - (5) - (5)	0.17334361	225	109	(7) - (1) - (10)	0.06933744	90
...				...			

Define - Exemplify - Use

Thank you!

**„Beyond Limits – Developments in Cadastral Domain“
Workshop, Zürich
19 March 2015**

Susanne Bleisch
susanne.bleisch@fhnw.ch