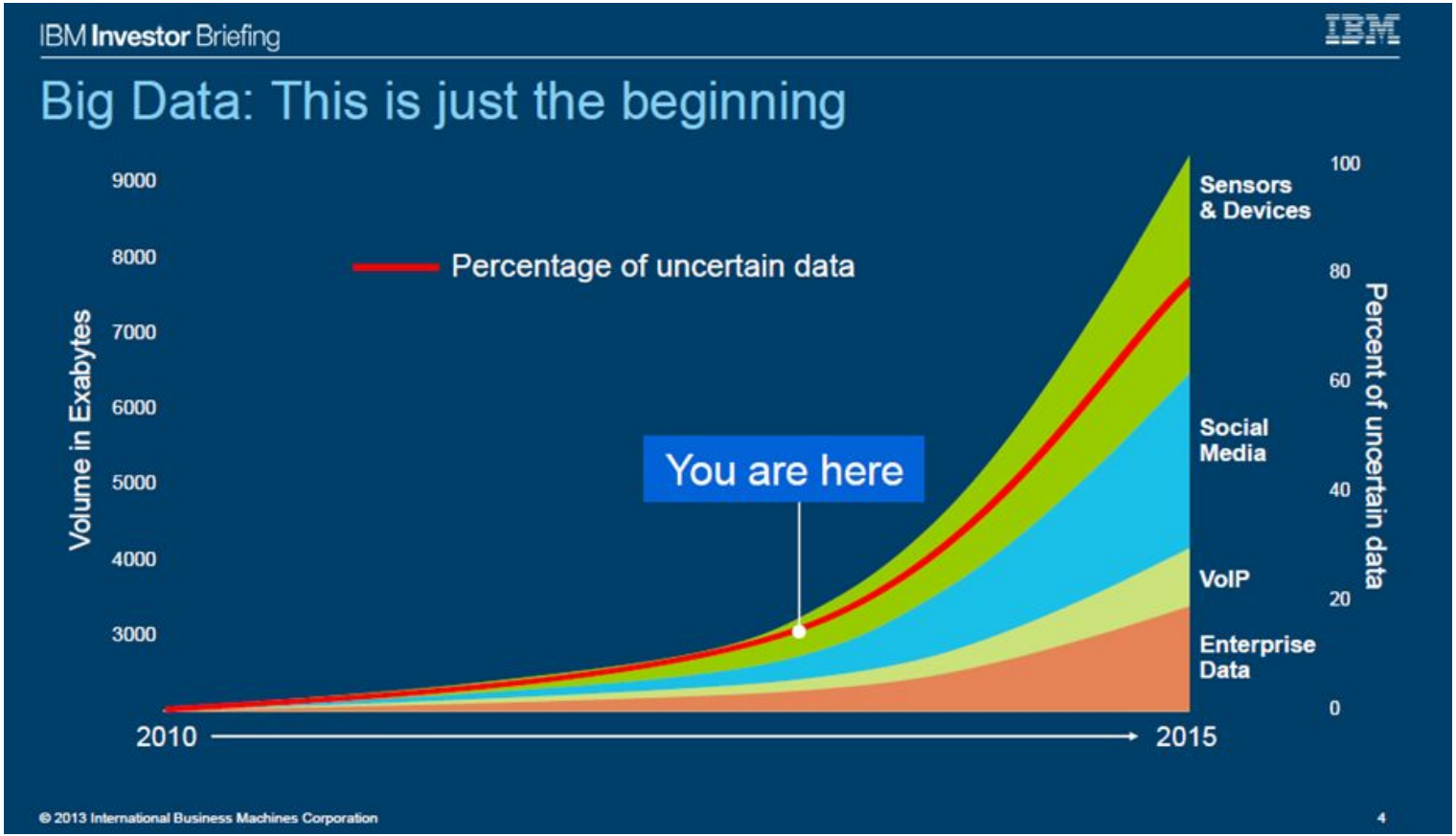
Data Visualization

Gina Lucia Muñoz Salas

Universidad Católica San Pablo

Centro de Investigación e Innovación en Ciencia Computación

Data growth


IBM Investor Briefing
IBM
Big Data: This is just the beginning
Volume in Exabytes
9000
8000
7000
6000
5000
4000
3000
Percentage of uncertain data
You are here
Sensors & Devices
Social Media
VoIP
Enterprise Data
100
80
60
40
20
0
Percent of uncertain data
2010
2015
© 2013 International Business Machines Corporation
4

Data growth

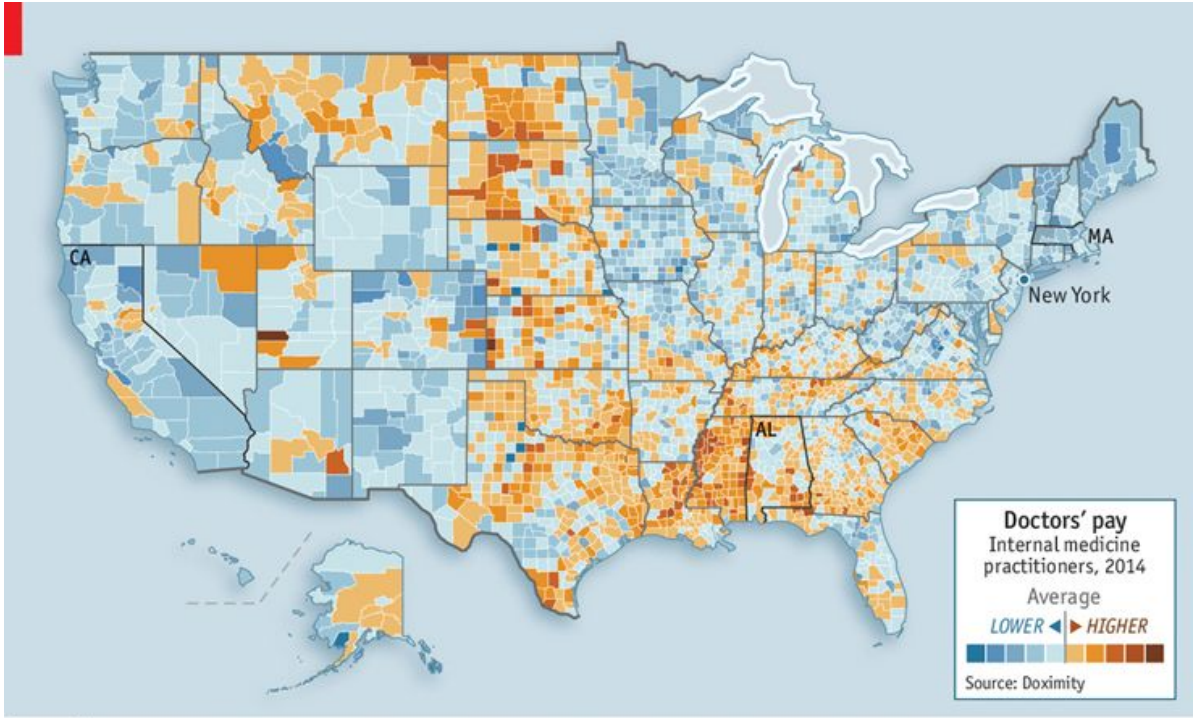
- 40,000 search queries are performed per second (on Google alone), which makes it 3.46 million searches per day and 1.2 trillion every year.
- Every minute Facebook users send roughly 31.25 million messages and watch 2.77 million videos.
- In just 5 years the number of smart connected devices in the world will be more than 50 billion – all of which will create data that can be shared, collected and analyzed

How to deal with this volumen of data?

Month	Force	Neighbourh	Burglary	Robbery	Vehicle crim	Violent crim	Anti-social b	Criminal dan	Shoplifting	Other theft	Drugs	Public disord	Other crime
2013-05	Cheshire Constabulary	N020	4	0	0	6	12	0	3	2	0	0	0
2013-05	Cheshire Constabulary	N021	22	3	4	19	100	12	20	17	3	2	0
2013-05	Cheshire Constabulary	N022	16	1	8	12	60	7	2	7	3	1	0
2013-05	Cheshire Constabulary	N023	16	0	8	10	56	10	5	5	1	2	1
2013-05	Cheshire Constabulary	N024	3	0	2	31	76	8	33	25	10	20	11
2013-05	Cheshire Constabulary	N025	18	0	15	10	102	15	16	19	1	2	2
2013-05	Cheshire Constabulary	N008	13	0	5	12	41	7	2	15	1	1	0
2013-05	Cheshire Constabulary	N009	5	0	2	10	53	11	5	10	3	1	0
2013-05	Cheshire Constabulary	N006	1	0	1	1	18	0	2	2	0	1	0
2013-05	Cheshire Constabulary	N007	2	0	5	22	73	16	6	12	6	6	1
2013-05	Cheshire Constabulary	N004	26	3	22	79	416	100	70	91	12	12	8
2013-05	Cheshire Constabulary	N005	3	0	4	2	23	2	3	4	0	0	0
2013-05	Cheshire Constabulary	N002	14	1	6	20	90	18	8	21	6	4	1
2013-05	Cheshire Constabulary	N003	2	0	4	11	53	8	6	13	2	3	1
2013-05	Cheshire Constabulary	N001	21	0	16	26	105	23	14	14	5	4	1
2013-05	Cheshire Constabulary	N048	8	1	11	3	30	11	3	16	1	0	0
2013-05	Cheshire Constabulary	N049	24	1	21	56	261	60	57	45	7	7	7
2013-05	Cheshire Constabulary	N042	12	0	7	43	254	20	71	39	16	10	0
2013-05	Cheshire Constabulary	N043	14	0	8	17	78	18	2	7	9	6	0
2013-05	Cheshire Constabulary	N040	29	0	55	21	73	15	13	8	2	1	1
2013-05	Cheshire Constabulary	N041	9	0	4	18	40	10	2	8	1	0	0
2013-05	Cheshire Constabulary	N046	12	0	3	9	45	13	3	9	7	3	1
2013-05	Cheshire Constabulary	N047	7	0	5	7	45	11	1	10	2	2	0
2013-05	Cheshire Constabulary	N044	9	0	23	19	78	24	17	40	8	5	0
2013-05	Cheshire Constabulary	N045	6	0	9	7	44	14	6	18	3	3	1
2013-05	Cheshire Constabulary	N033	8	0	5	8	36	8	12	10	1	0	1
2013-05	Cheshire Constabulary	N026	11	0	8	4	51	9	1	11	1	1	3
2013-05	Cheshire Constabulary	N050	9	0	6	5	80	11	4	6	2	3	1
2013-05	Cheshire Constabulary	N027	10	0	13	16	81	19	15	19	5	4	1
2013-05	Cheshire Constabulary	N028	2	1	4	7	26	7	4	12	3	0	0
2013-05	Cheshire Constabulary	N029	13	0	6	7	55	9	3	7	5	1	1
2013-05	Cheshire Constabulary	N039	10	0	5	2	37	2	9	4	6	2	2
2013-05	Cheshire Constabulary	N032	12	1	14	11	47	13	12	6	2	3	0
2013-05	Cheshire Constabulary	N019	6	0	4	8	61	5	12	15	4	3	0
2013-05	Cheshire Constabulary	N018	7	0	5	14	39	5	8	17	1	1	1
2013-05	Cheshire Constabulary	N031	5	1	6	7	16	2	0	5	0	0	0
2013-05	Cheshire Constabulary	N030	1	0	7	5	18	2	0	3	1	2	2
2013-05	Cheshire Constabulary	N037	17	0	9	32	182	41	34	25	8	11	2
2013-05	Cheshire Constabulary	N036	11	0	5	23	84	22	17	6	1	2	1
2013-05	Cheshire Constabulary	N035	4	0	5	2	11	1	0	2	1	0	0
2013-05	Cheshire Constabulary	N034	11	0	13	14	74	16	9	17	0	6	2

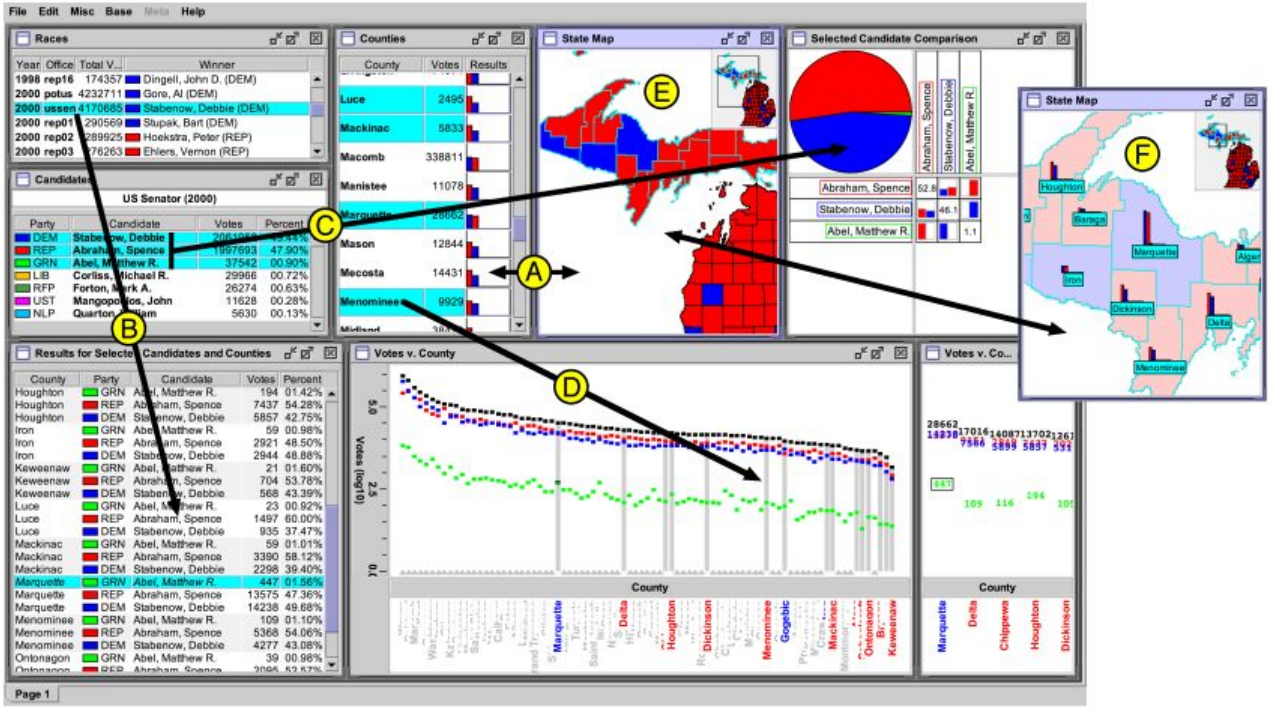
Data Visualization

Visualization is the information communication using graphic representations.



- Visualization can easily represent information
- Helps in knowledge discovery process and decisión making

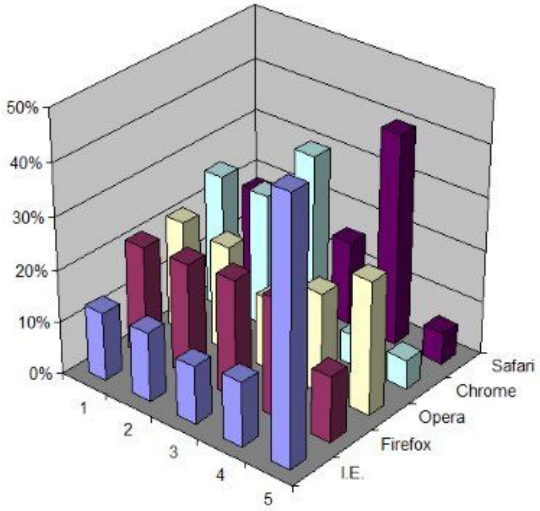
Data Visualization



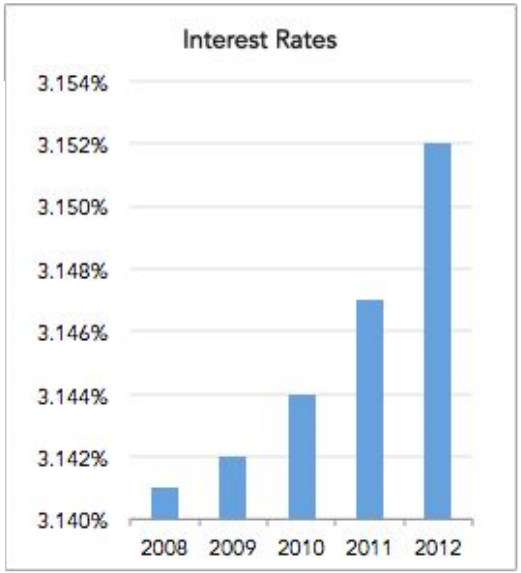
Weaver, C. E.. 2004. Building highly-coordinated visualizations in Improvise. Proceedings of the IEEE Information Visualization Conference: 159-166.

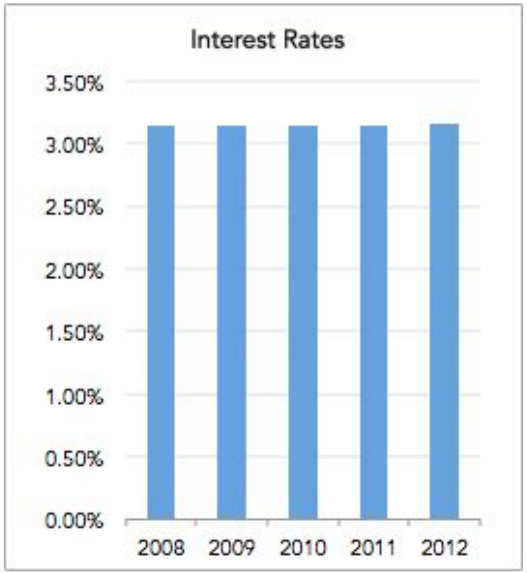
Accurate representation

- Simplicity
- Design
- Scale

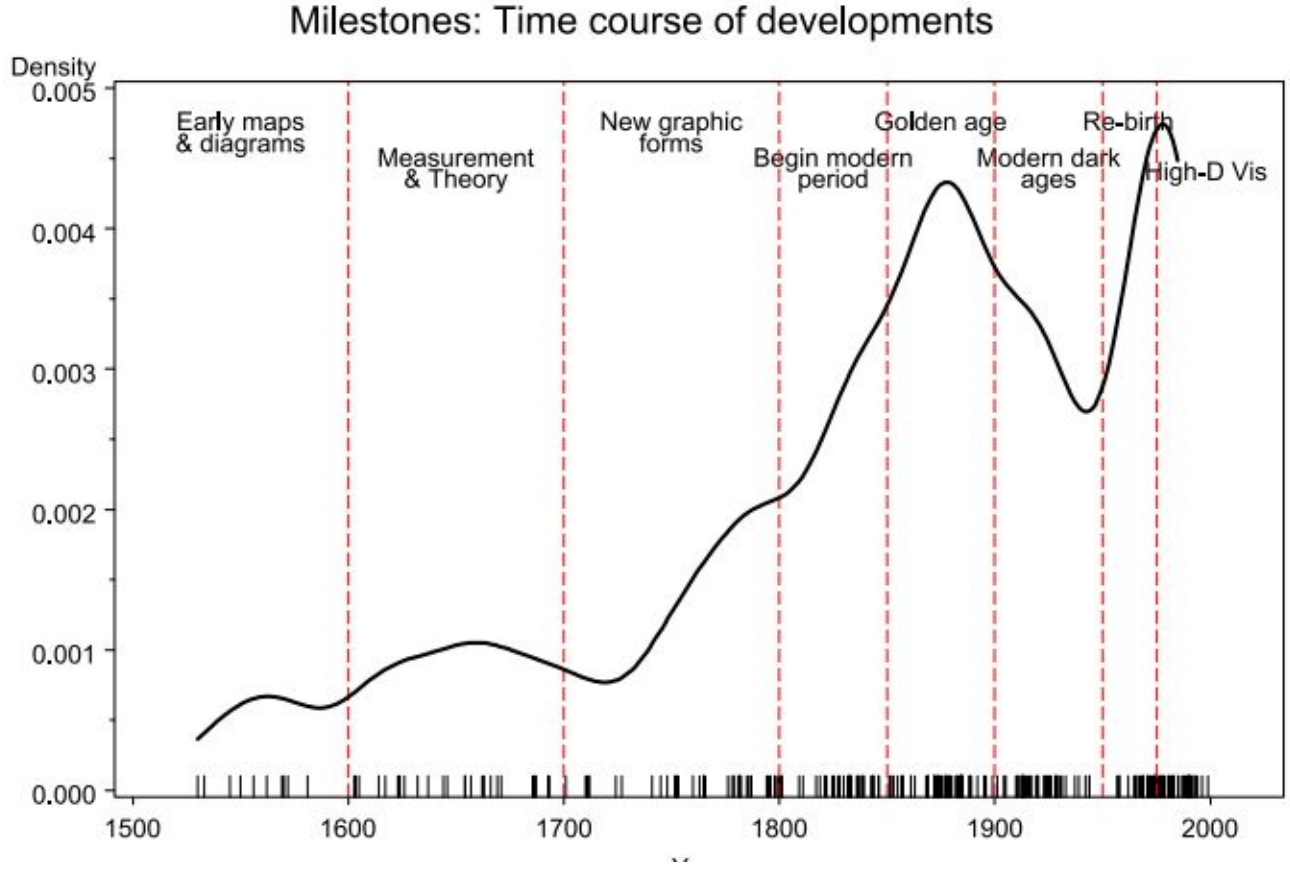


Same Data, Different Y-Axis

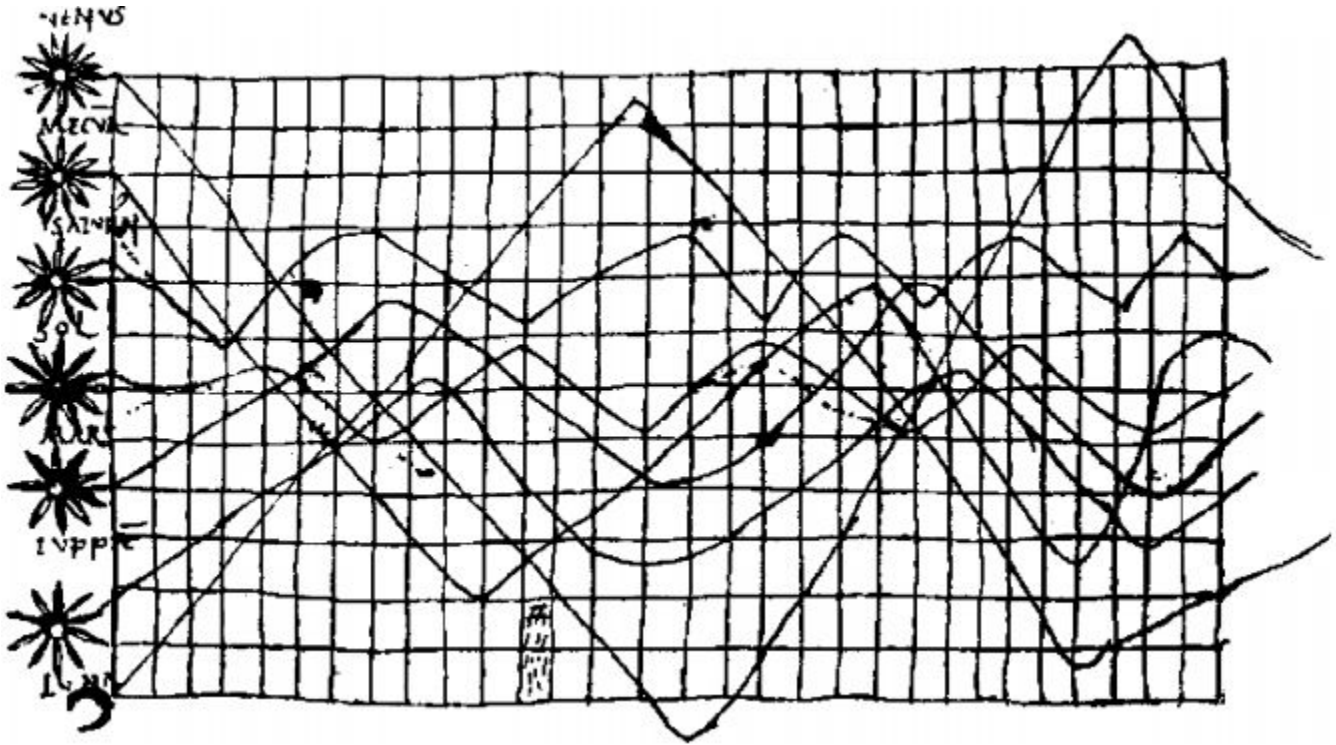




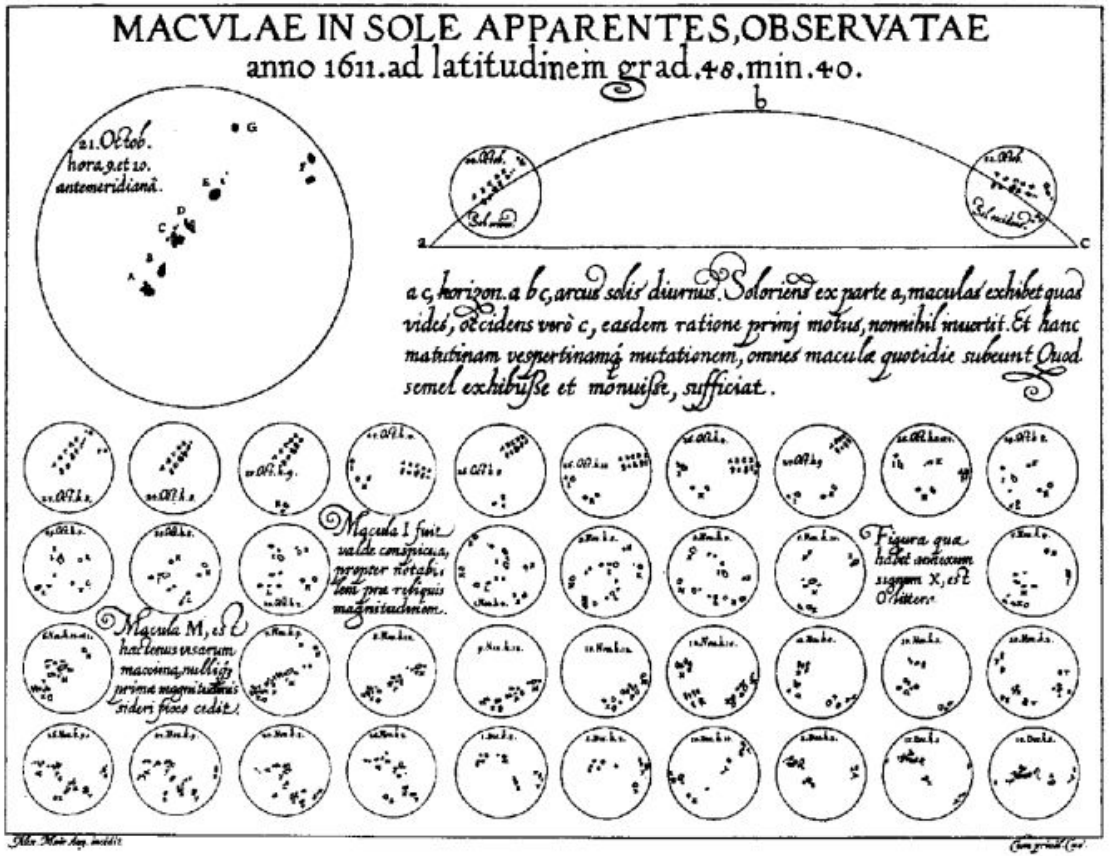
Visualization history- Milestones Project



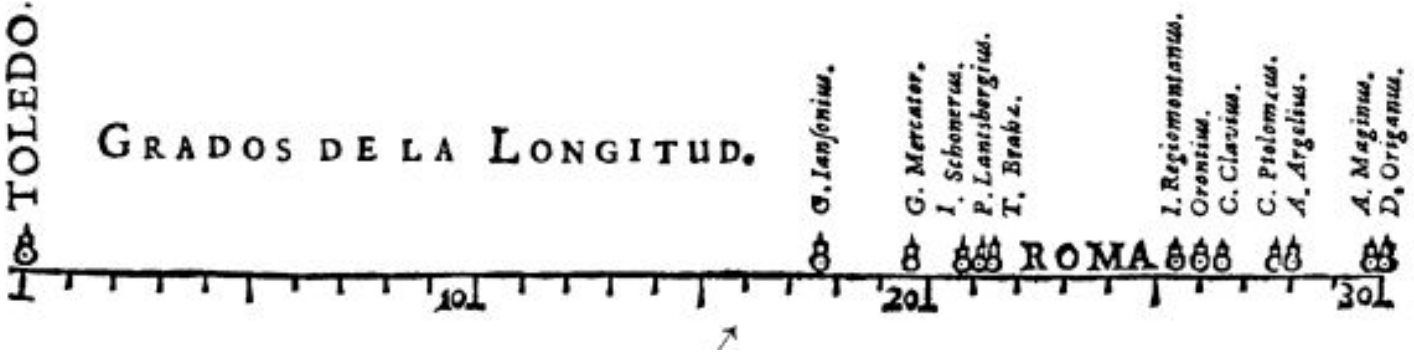
Visualization history



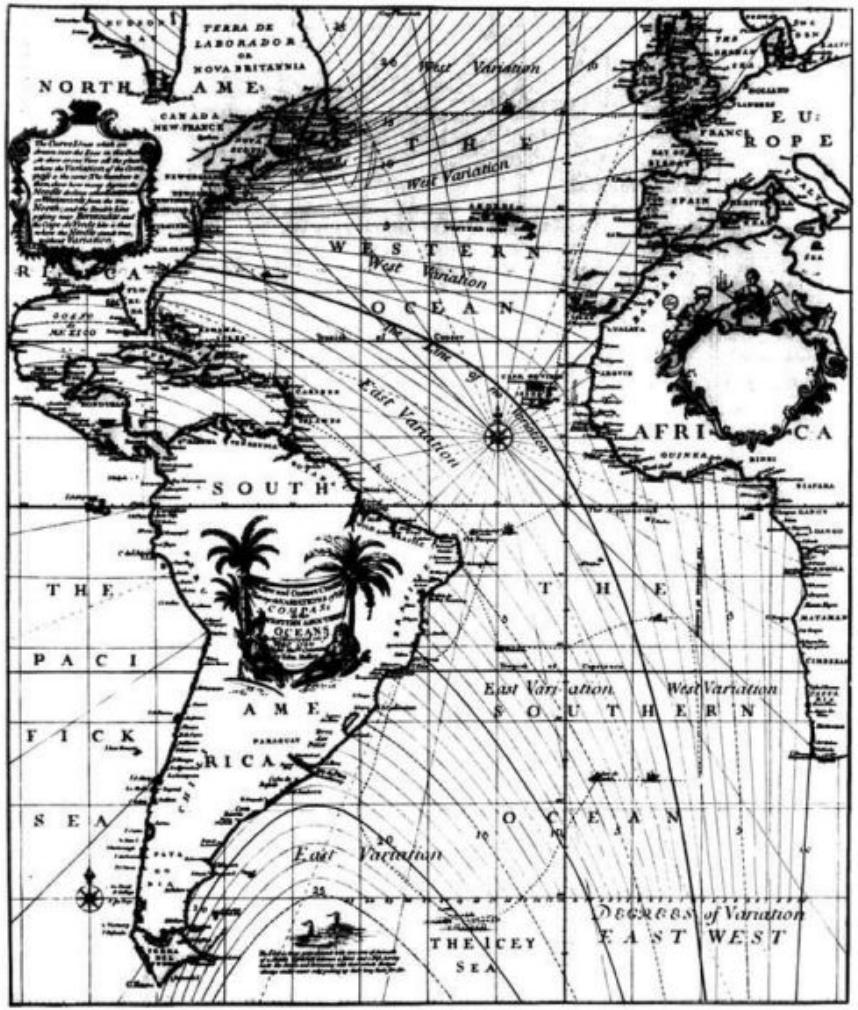
Planetary movements shown as cyclic inclinations over time, by an unknown astronomer, appearing in a10th-century appendix to commentaries by A.T. Macrobius on Cicero's In Somnium Sciponis.
Source: Funkhouser (1936)



Scheiner's 1626 representation of the changes in sunspots over time. Source: Scheiner (1626–1630)

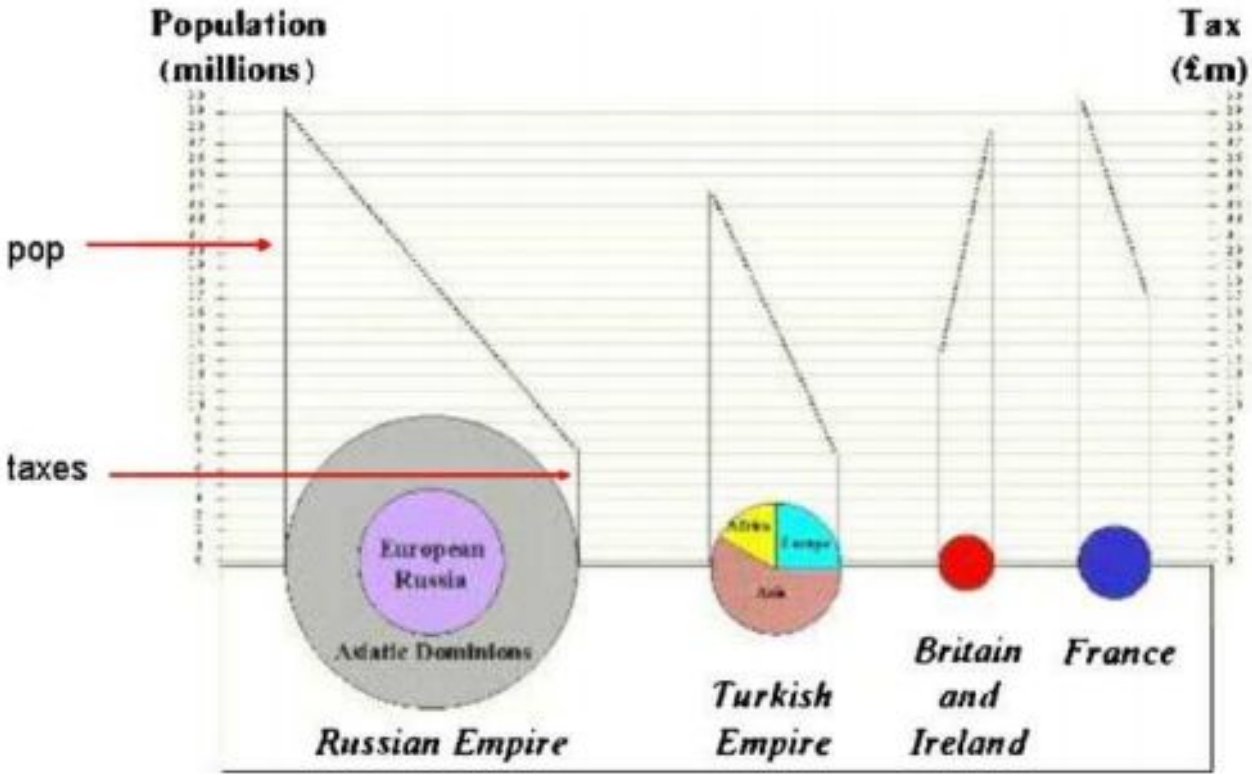


Langren's 1644 graph of determinations of the distance, in longitude, from Toledo to Rome. he correct distance is 16° 30' . Source: Tute (1997)

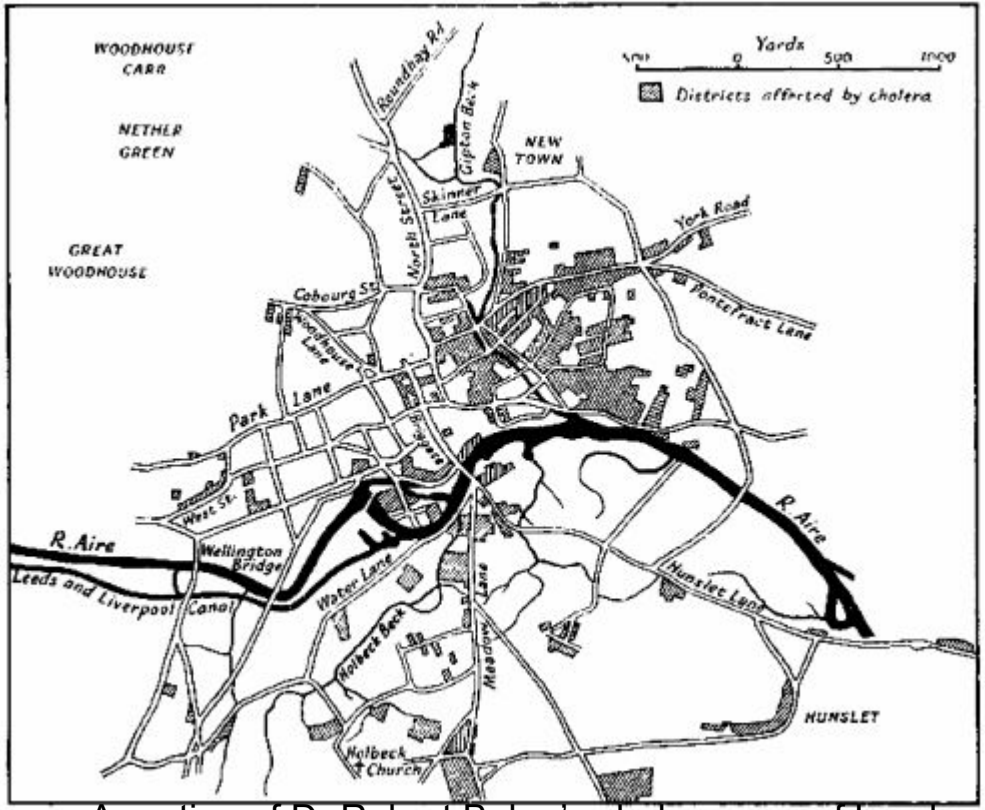


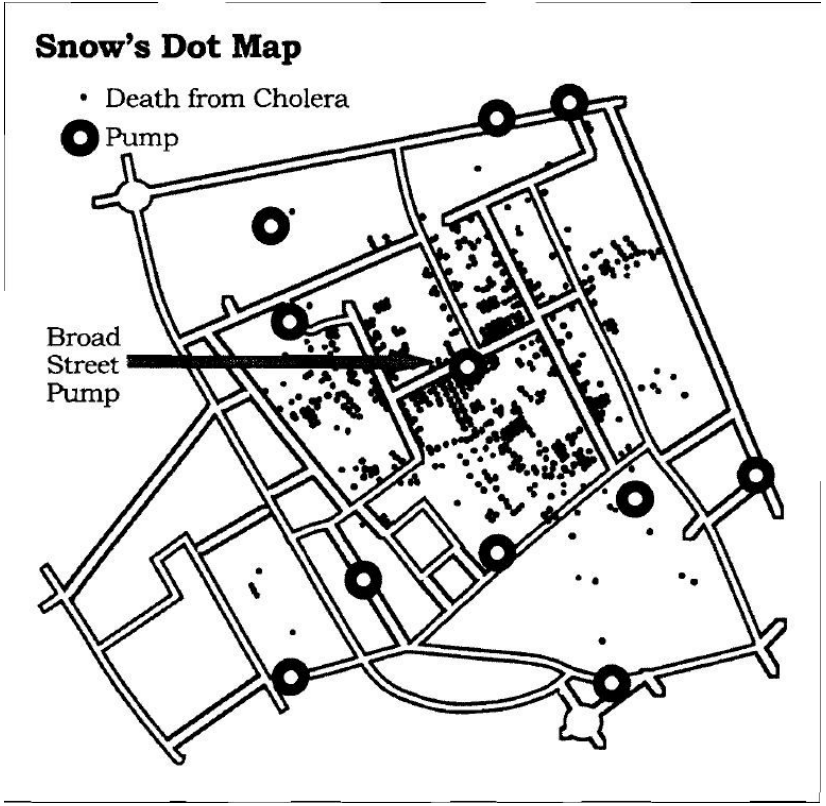
Map with isogons – lines of equal magnetic declination

A portion of Edmund Halley's New and Correct Sea Chart Shewing the Variations in the Compass in the Western and Southern Ocean,1701 . Source: Halley (1701), image from Palsky (1996)

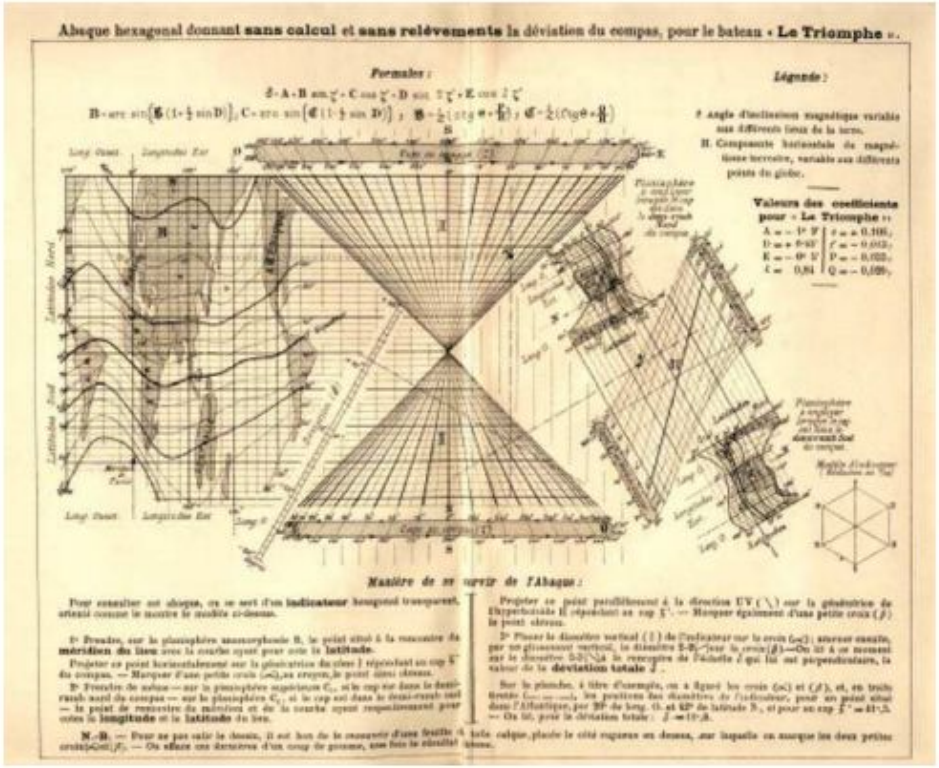


Redrawn version of a portion of Playfair's 1801 pie-circle-line chart, comparing population and taxes in several nation

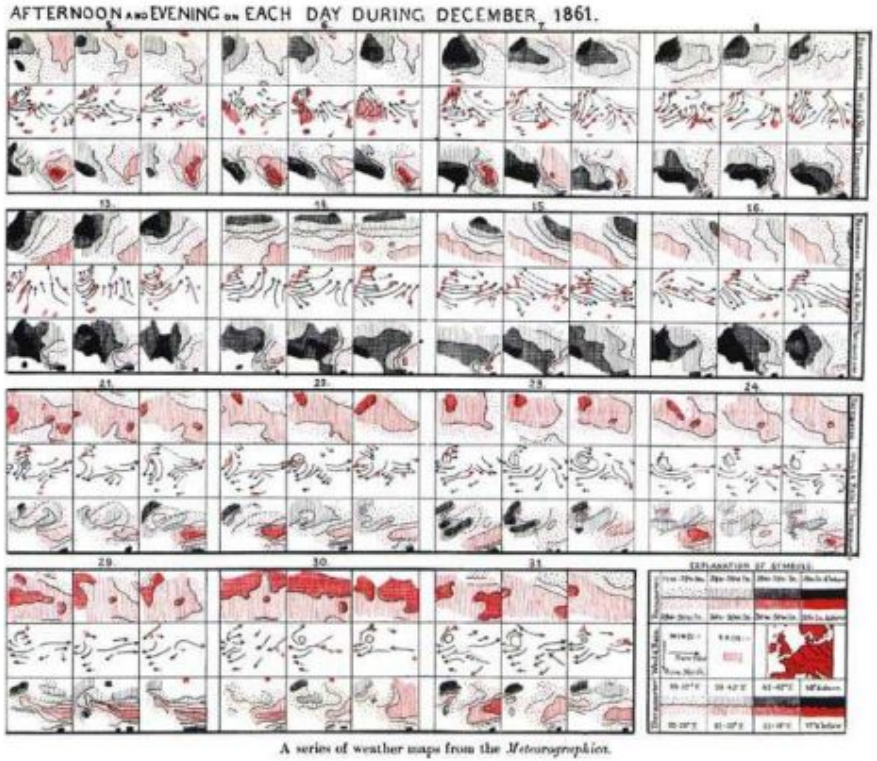




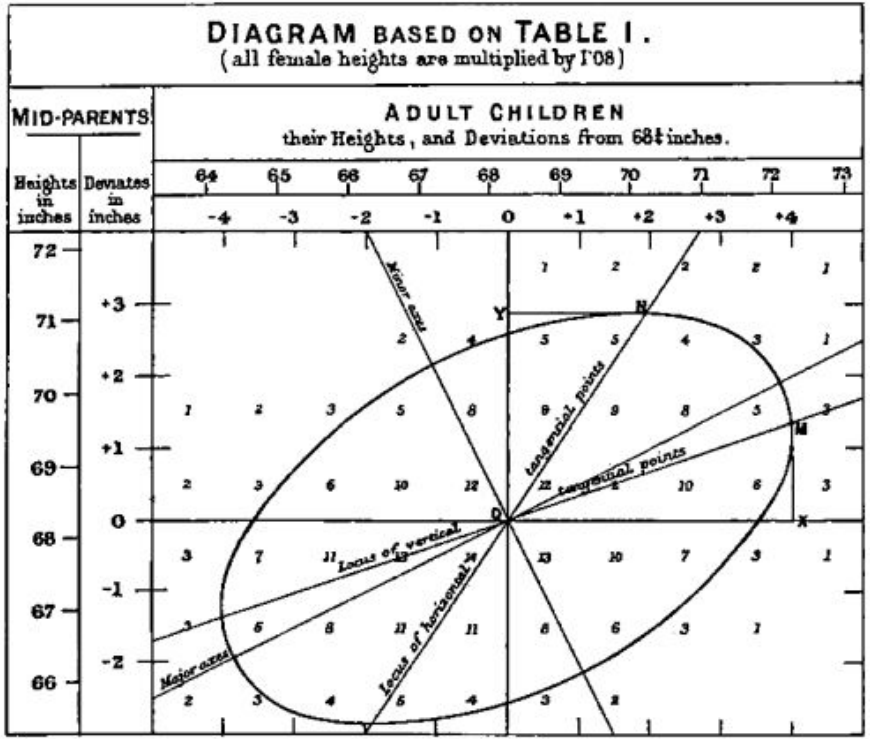
A portion of Dr Robert Baker's cholera map of Leeds, 1833, showing the districts affected by cholera. Source: Gilbert (1958)



Lallemand's 1885 L' abaque du bateau "Le Triomphe", allowing determination of magnetic deviation at sea without calculation. Source: courtesy Mme Marie-Noëlle Maisonneuve, Les fonds anciens de la bibliothèque de l'École des Mines de Pari

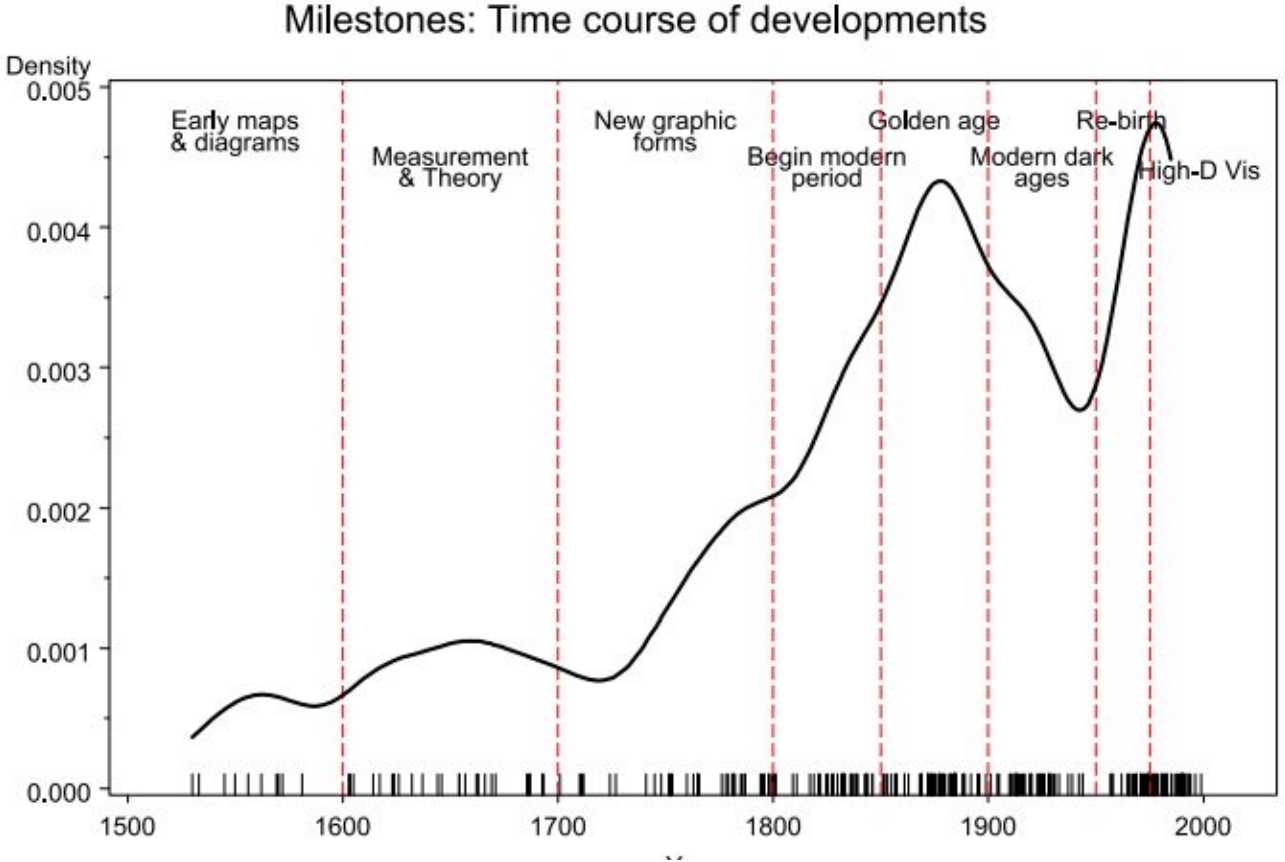


One page of Galton's 1863 multivariate weather chart of Europe showing barometric pressure, wind direction, rain and temperature for the month of December .1861 Source: Pearson (1914–1930)



Galton's smoothed correlation diagram for the data on heights of parents and children, showing one ellipse of equal frequency. Source: (Galton, 1886, Plate X)

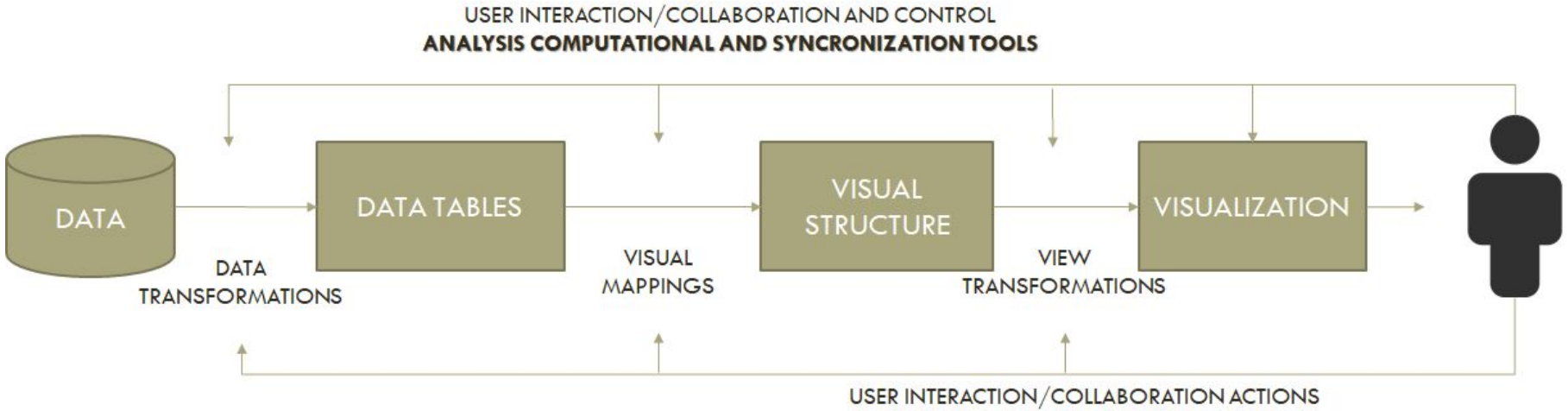
Visualization history- Milestones Project



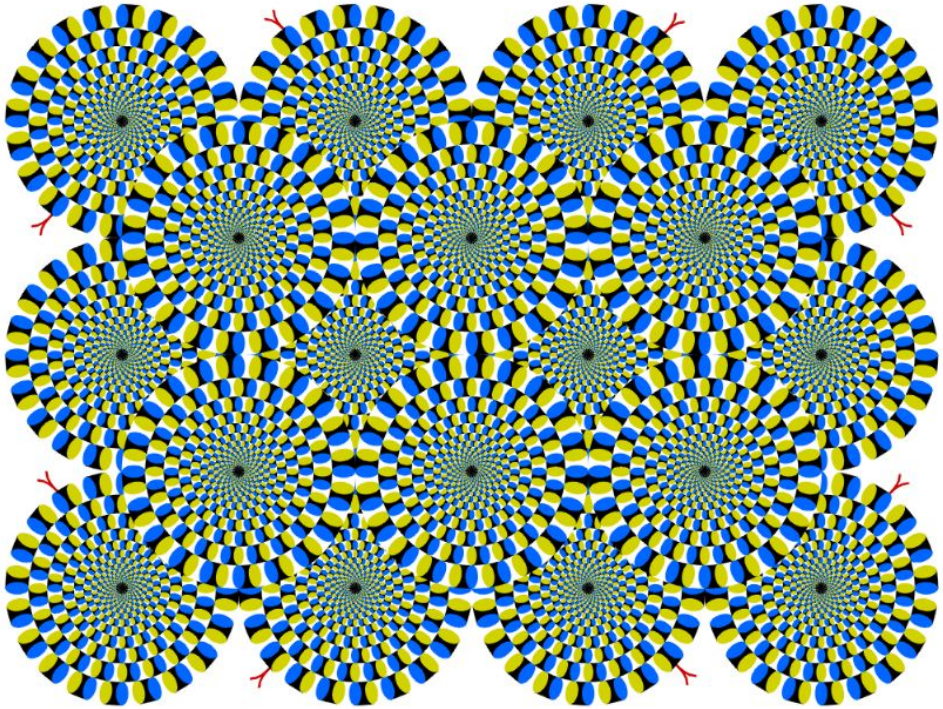
1975–present: High-D, Interactive 1.2.8 and Dynamic Data Visualization

- Highly interactive statistical computing systems.
- New paradigms of direct manipulation for visual data analysis.
- New methods for visualizing high-dimensional data.
- The invention (or re-invention) of graphical techniques for discrete and categorical data.
- The application of visualization methods to an ever-expanding array of substantive problems and data structures.
- Substantially increased attention to the cognitive and perceptual aspects of data display.

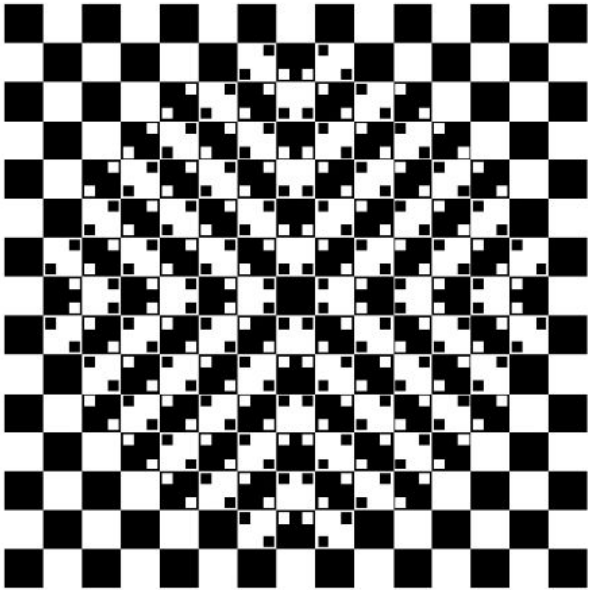
Visualization process



Perception

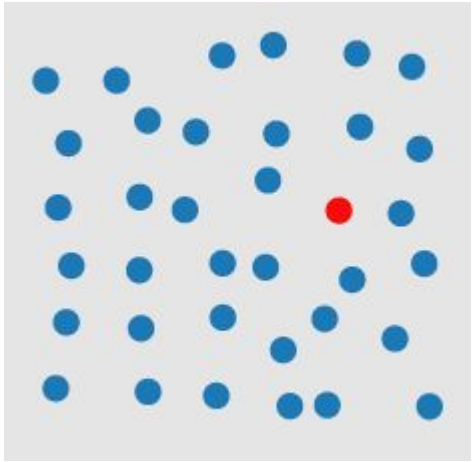


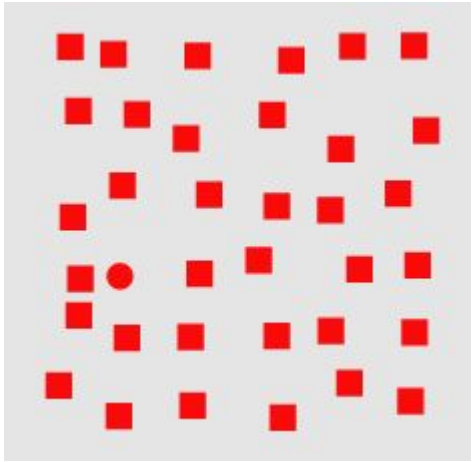
- Visual human system limitations are critical for visualization.



Perception

Pre attentive process





Perception

Perception

Gestalt principles of form perception

- Proximity
- Similarity
- Continuity
- Symmetry
- Closure
- Background
- Front plane
- Size

Visualizations

A Tour through the Visualization Zoo

A survey of powerful visualization techniques, from the obvious to the obscure

https://queue.acm.org/detail.cfm?searchterm=BSE&id=1805128

Data Visualization

Gina Lucia Muñoz Salas

Universidad Católica San Pablo

Centro de Investigación e Innovación en Ciencia Computación