

Visual Data Analytics Visualization

Dr. Johannes Kehrer – Siemens Technology, Munich

Disclaimer



These lecture slides and materials are for personal use by students only and are intended for educational purposes only. Distribution or reproduction of the material for any other purposes is prohibited.

The views and opinions expressed by any individual during the course of these lecturers are their own and do not necessarily represent the views, opinions, or positions of Siemens or the Technical University of Munich.

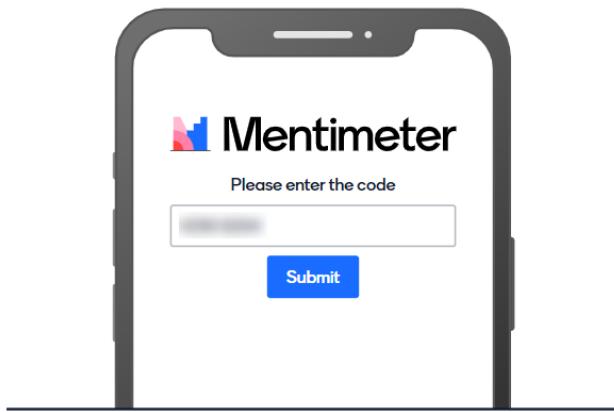
Although reasonable efforts have been made to ensure the information contained in these lecture slides is current and accurate, we do not assume any responsibility for the content of the posted material.

These slides also contain links to websites of third parties. As the content of these websites is not under our control, we cannot assume any liability for such external content.

Any comments/questions?

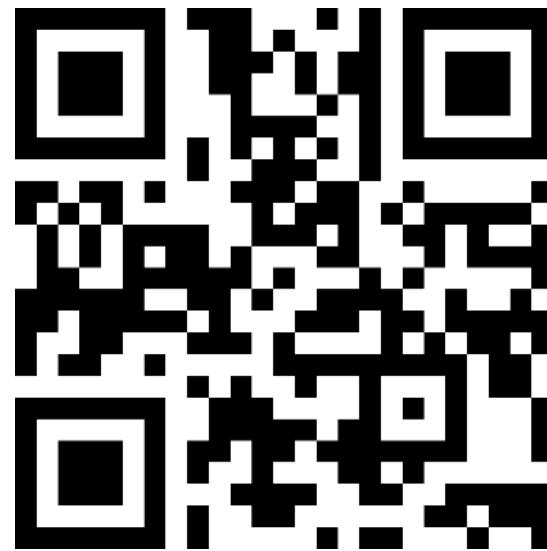
Go to

www.menti.com



Enter the code

54 11 67 0



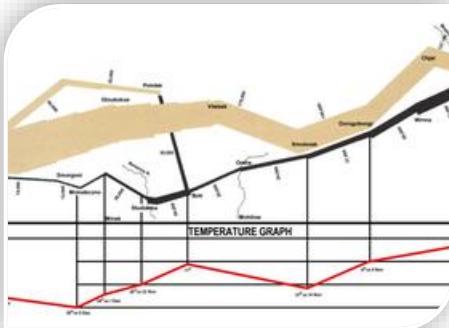
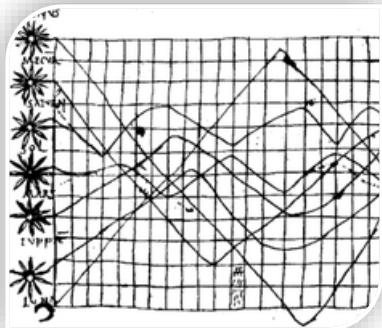
Or use QR code

Today's lecture

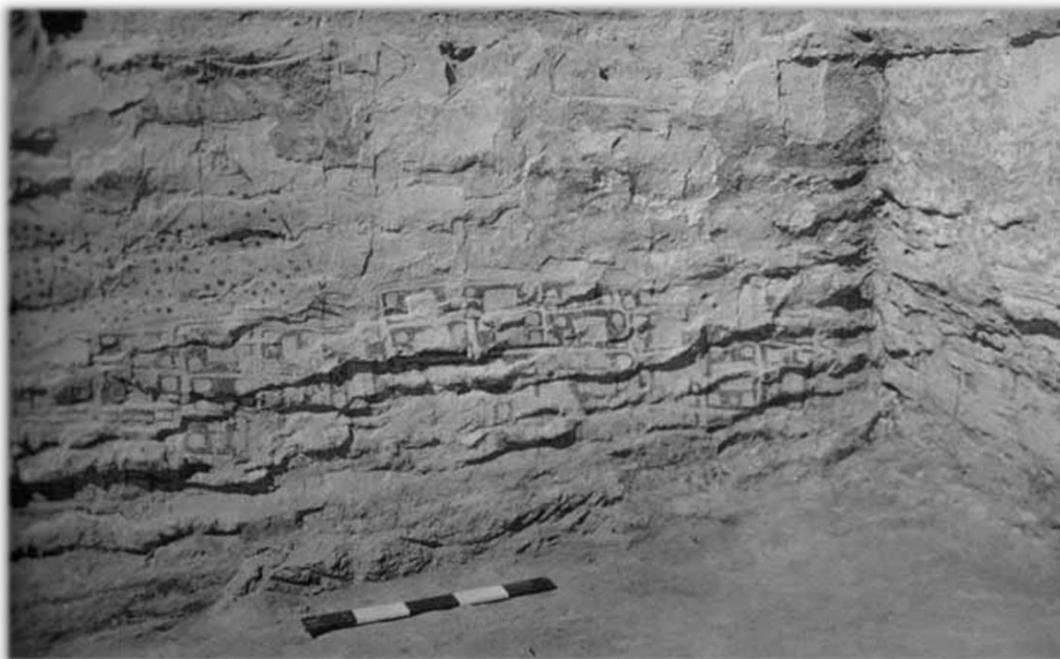
- History of visualization
- Definitions, goals and major areas

History

Techniques for finding visual representations of (abstract) data are not new!



History



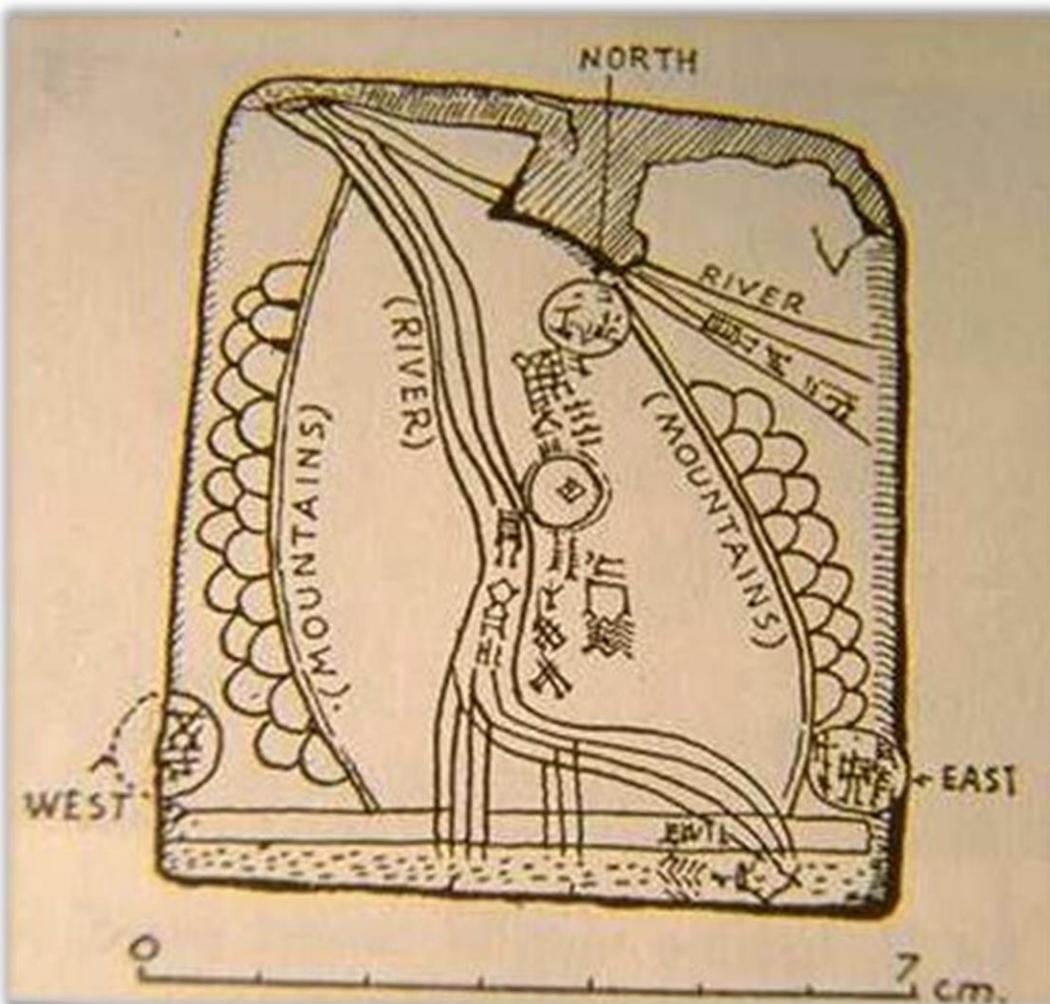
(Oldest) Map of Catal Hyük,
Turkey 6200 BC

Excavation



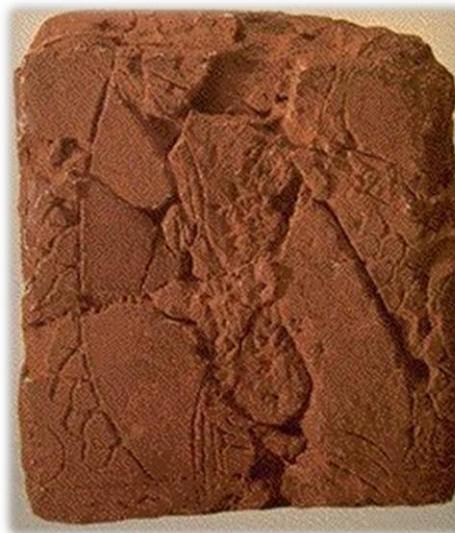
Reconstruction

History

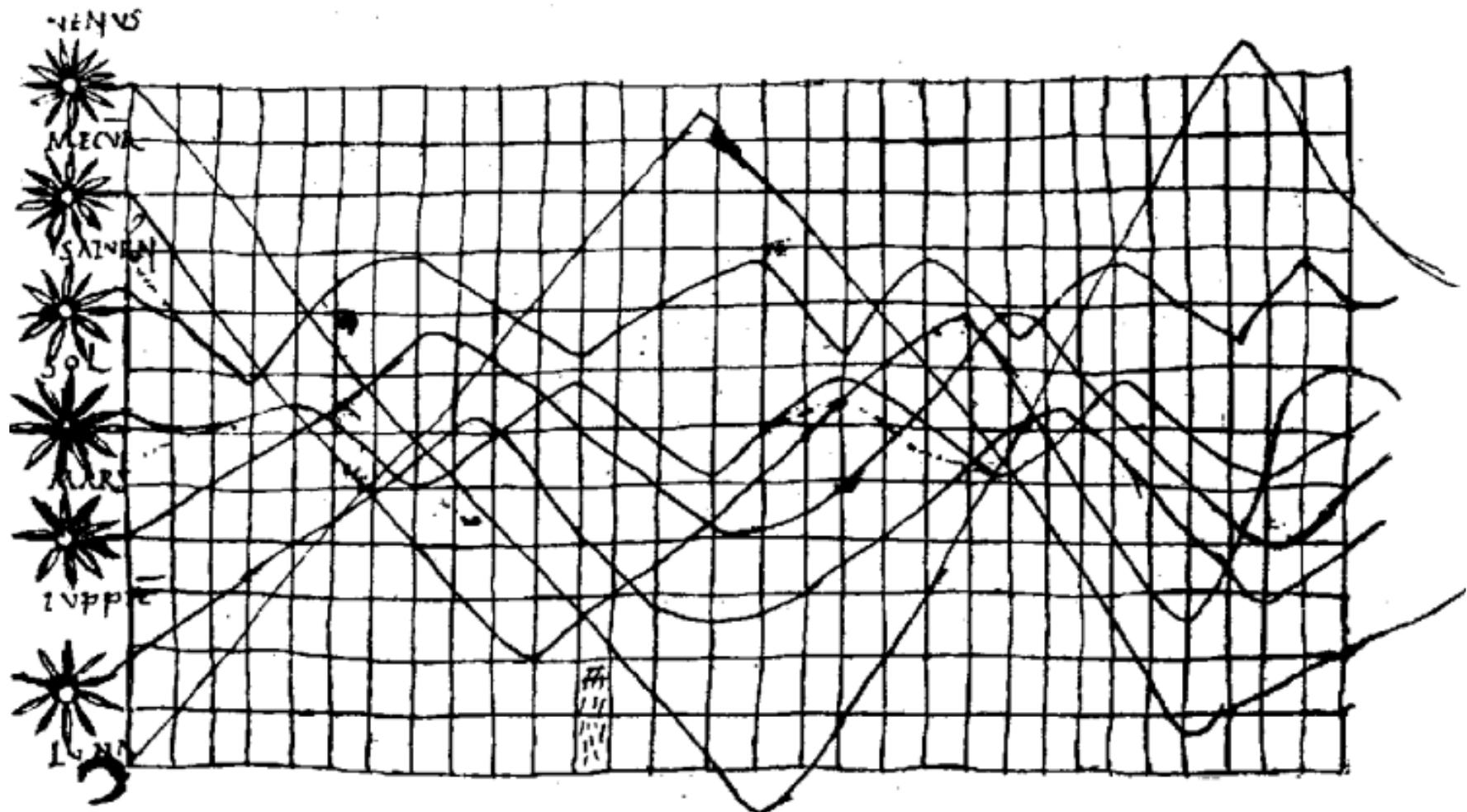


Map on clay of Ga-Sur, Iraq
2500 BC

Map on clay
Reconstruction



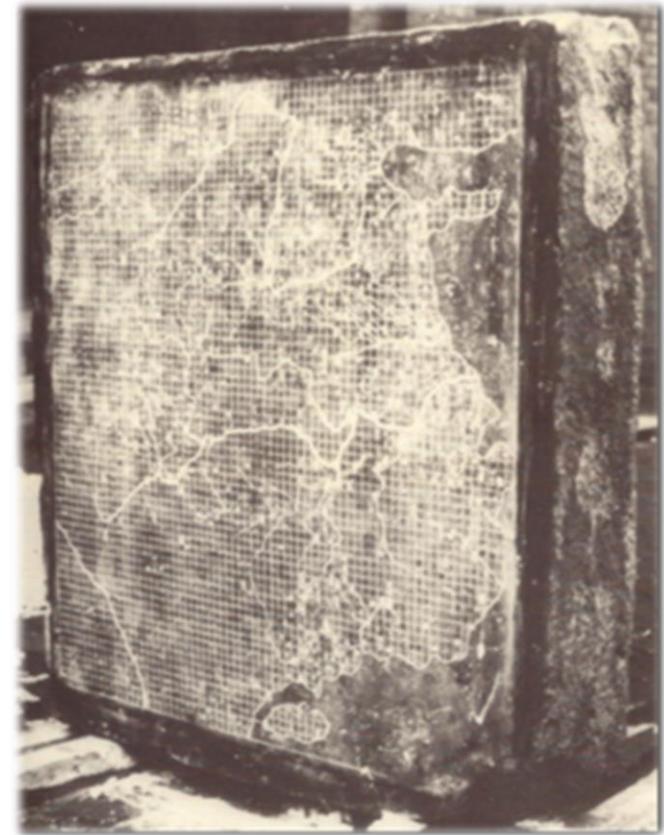
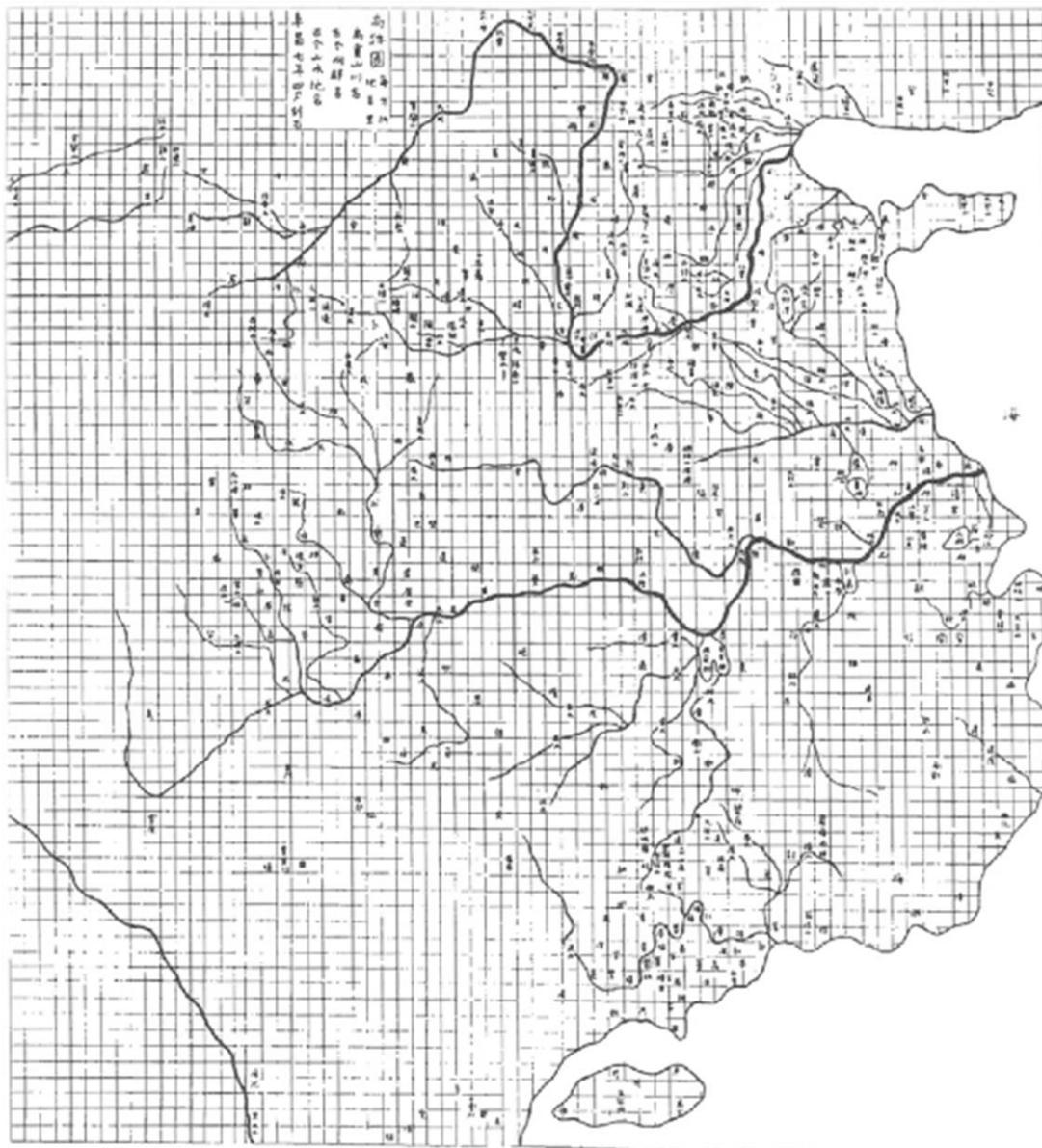
History



Inclinations of planetary orbits as function of time

10th century

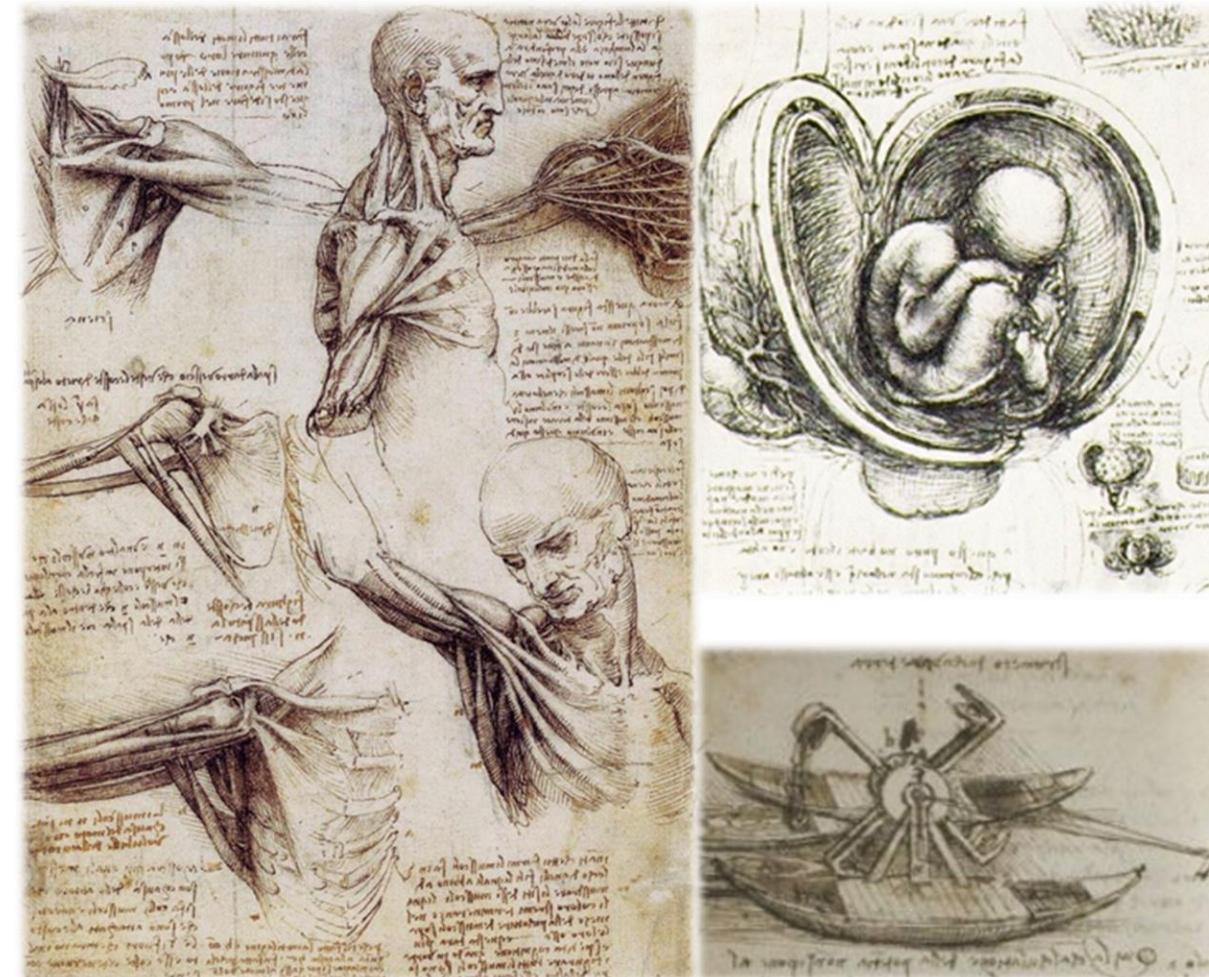
History



Map of the Tracks of
Yü the Great, China (11th century)

History

SIEMENS
Ingenuity for life

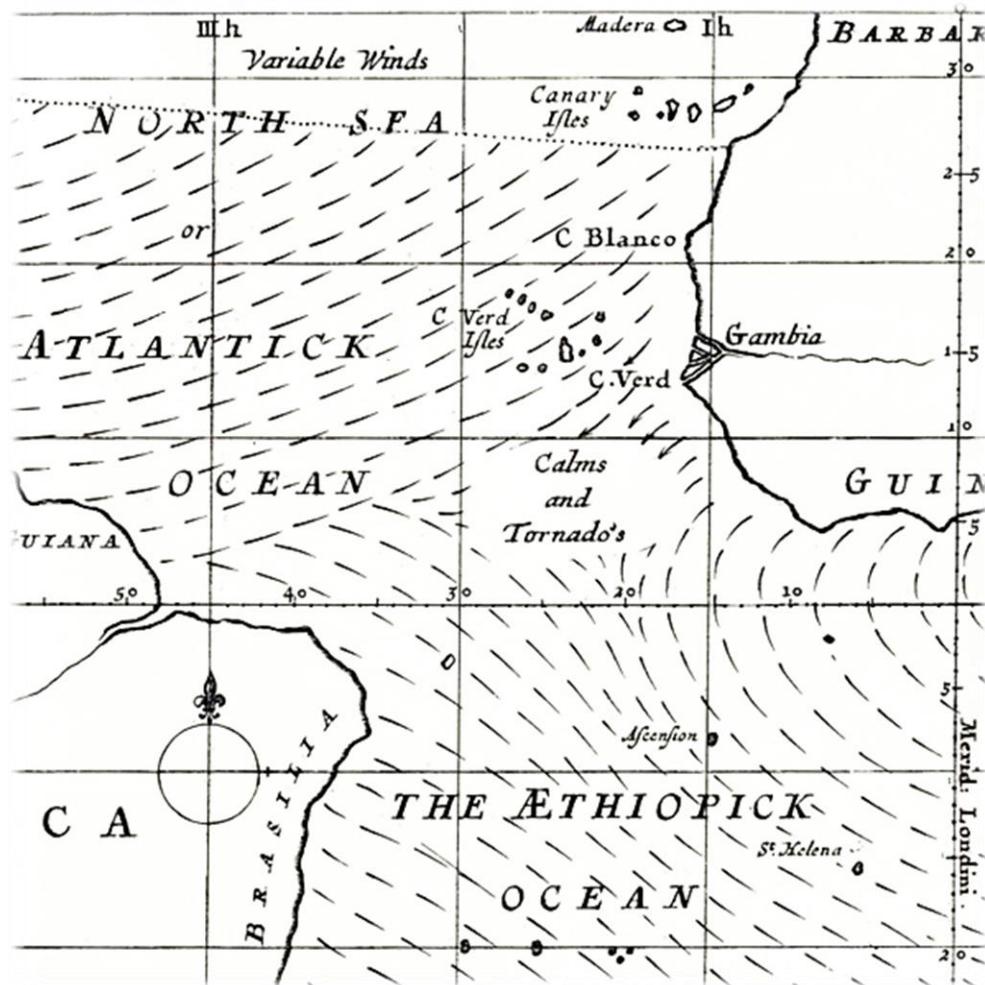


Illustrations by Leonardo DaVinci (1452-1519)

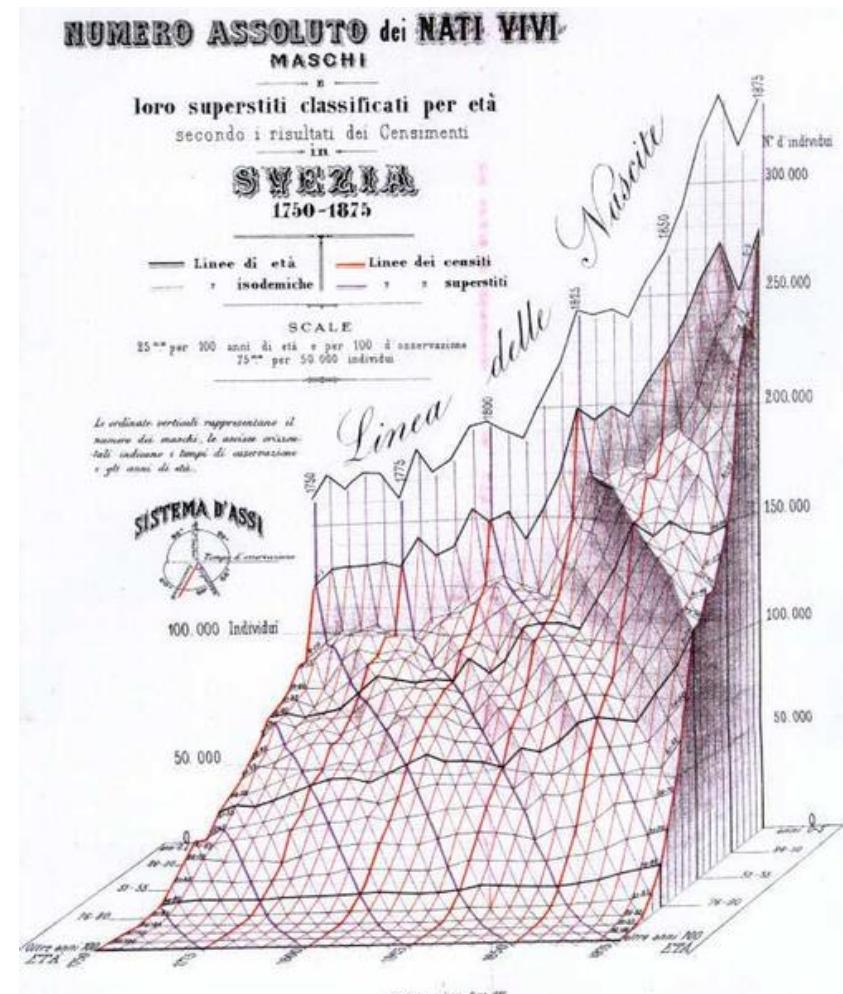


William Curtis (1746-1799)

History

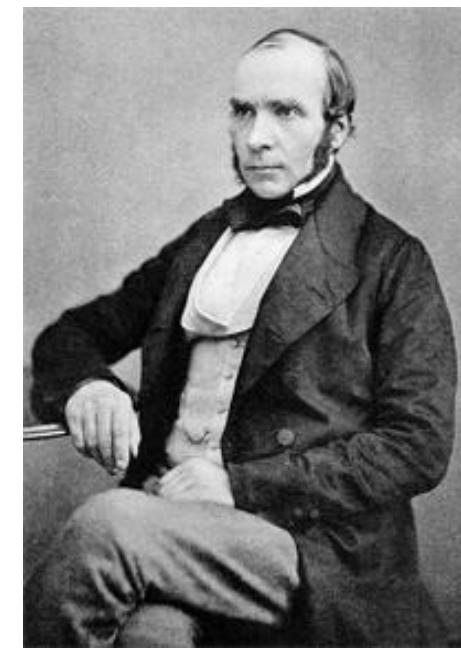
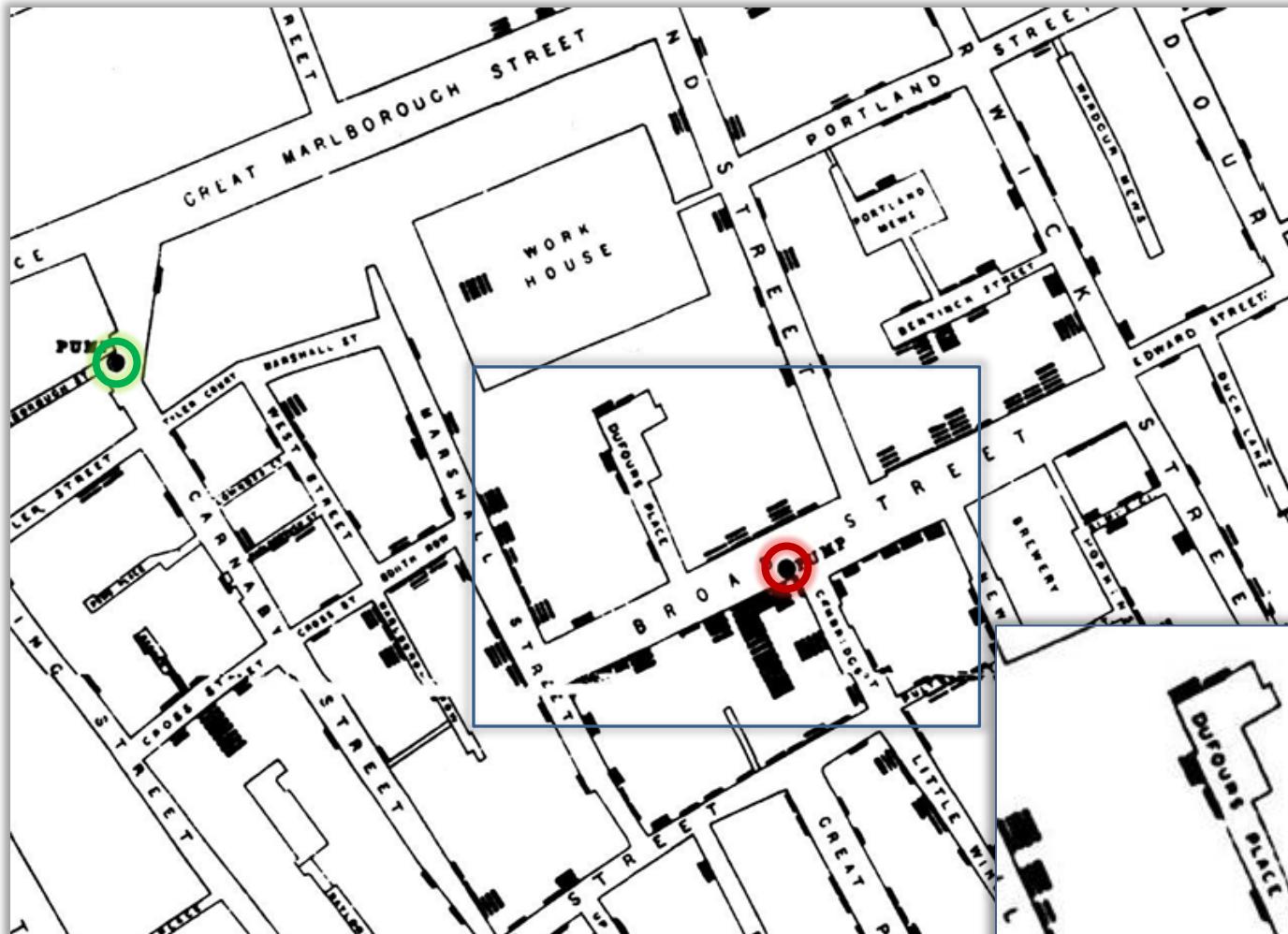


Vector visualization (1686)



Height field (1879)

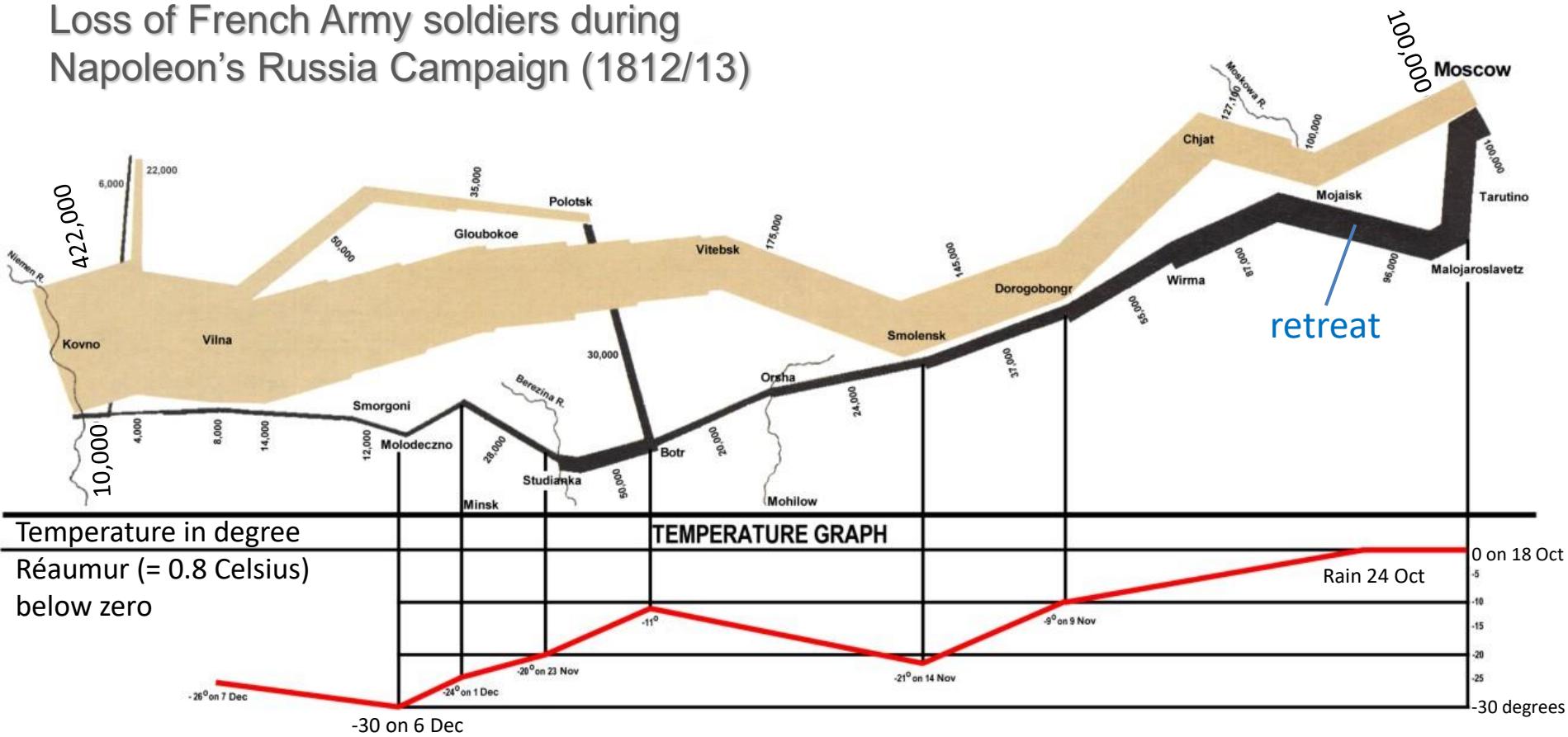
History



Cholera outbreak in Soho, London
John Snow (1854)

History

Loss of French Army soldiers during Napoleon's Russia Campaign (1812/13)

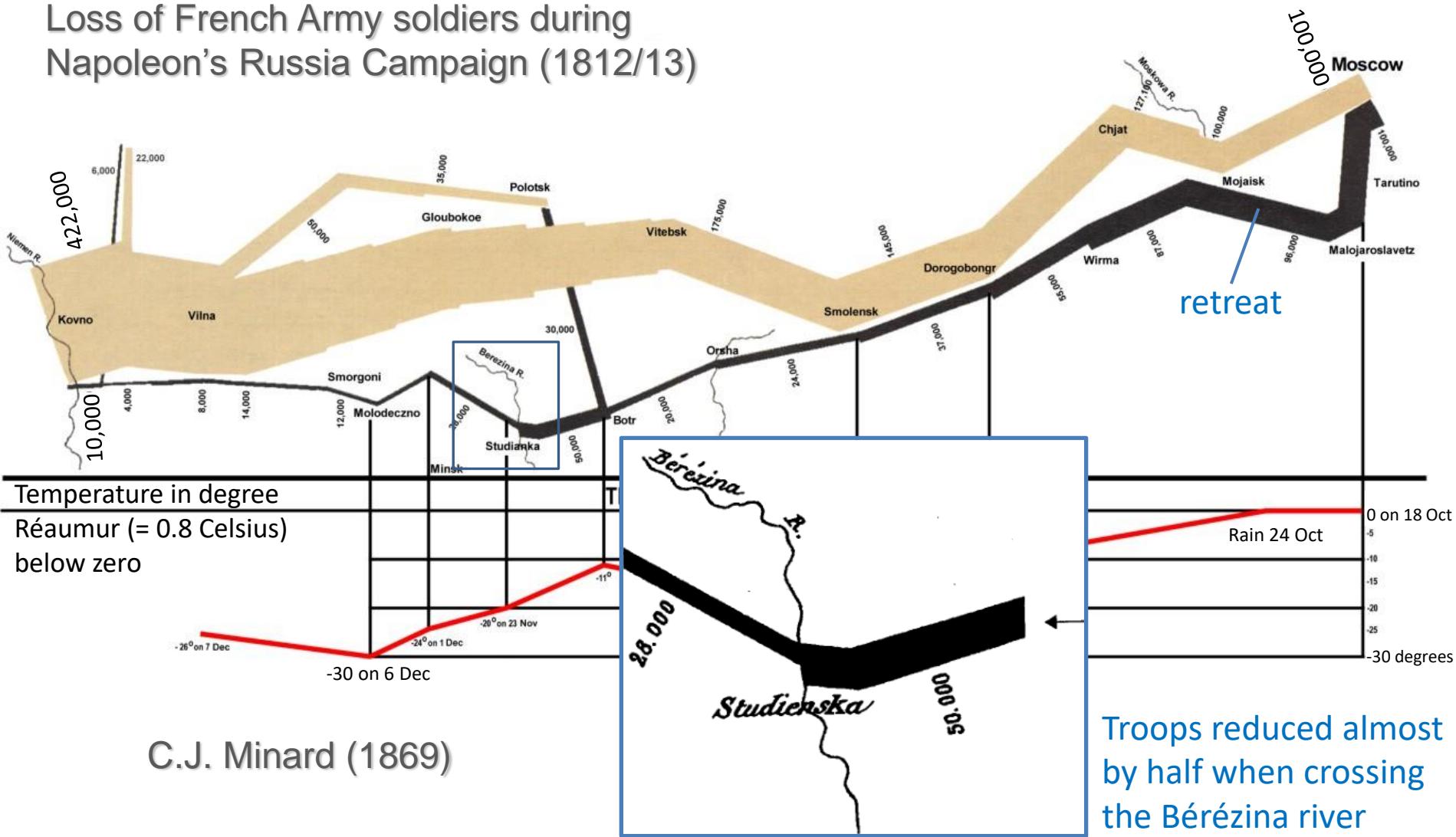


C.J. Minard (1869)

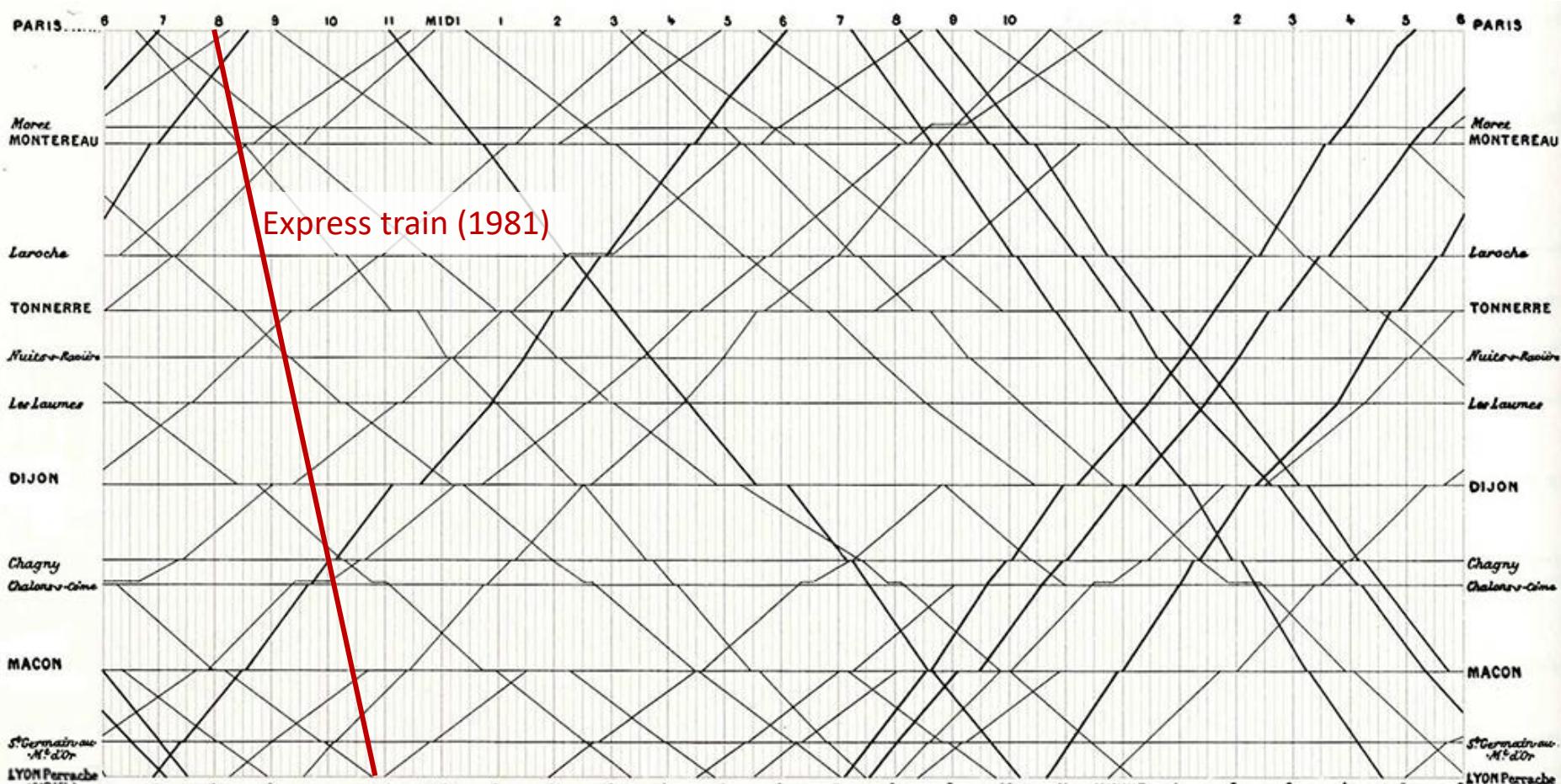
Map encodes size of army (width of band), location & direction of movement, split and reunion of troops, and temperature during retreat

History

Loss of French Army soldiers during
Napoleon's Russia Campaign (1812/13)



History



Train schedule between Paris and Lyon

E.J. Marey (1880)

History of Modern Visualization

“The purpose of computing is
insight, not numbers”

(Hamming 1962)

- 1987 US NSF Advisory Panel on Graphics and Image Processing
 - Computer experiments allow access to new worlds
 - Real experiments are too expensive, too dangerous, etc.
 - Arbitrary large, small time scales, and spatial dimensions

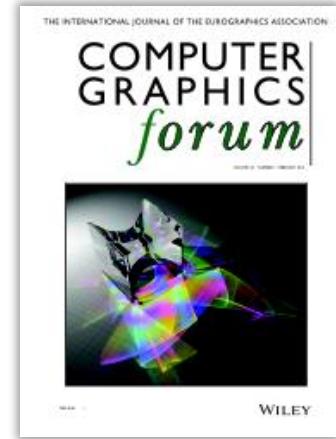
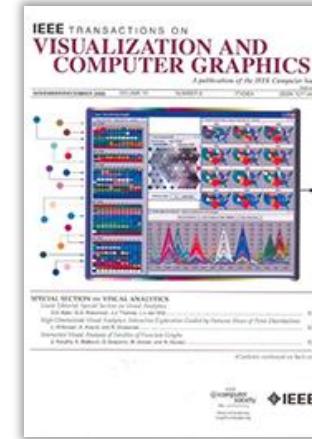


History of Modern Visualization

- Direct implications
 - The data flood from supercomputer simulations can only be dealt with visually
 - Needs a visualization specialist and an interdisciplinary team
 - New developments in hard/software, nets, etc. are necessary
- Advantages in the long term will be
 - Faster insight
 - Faster product – development cycles
 - Stronger position in global competition
- **Suggestion:** spend lots of money to support scientific visualization

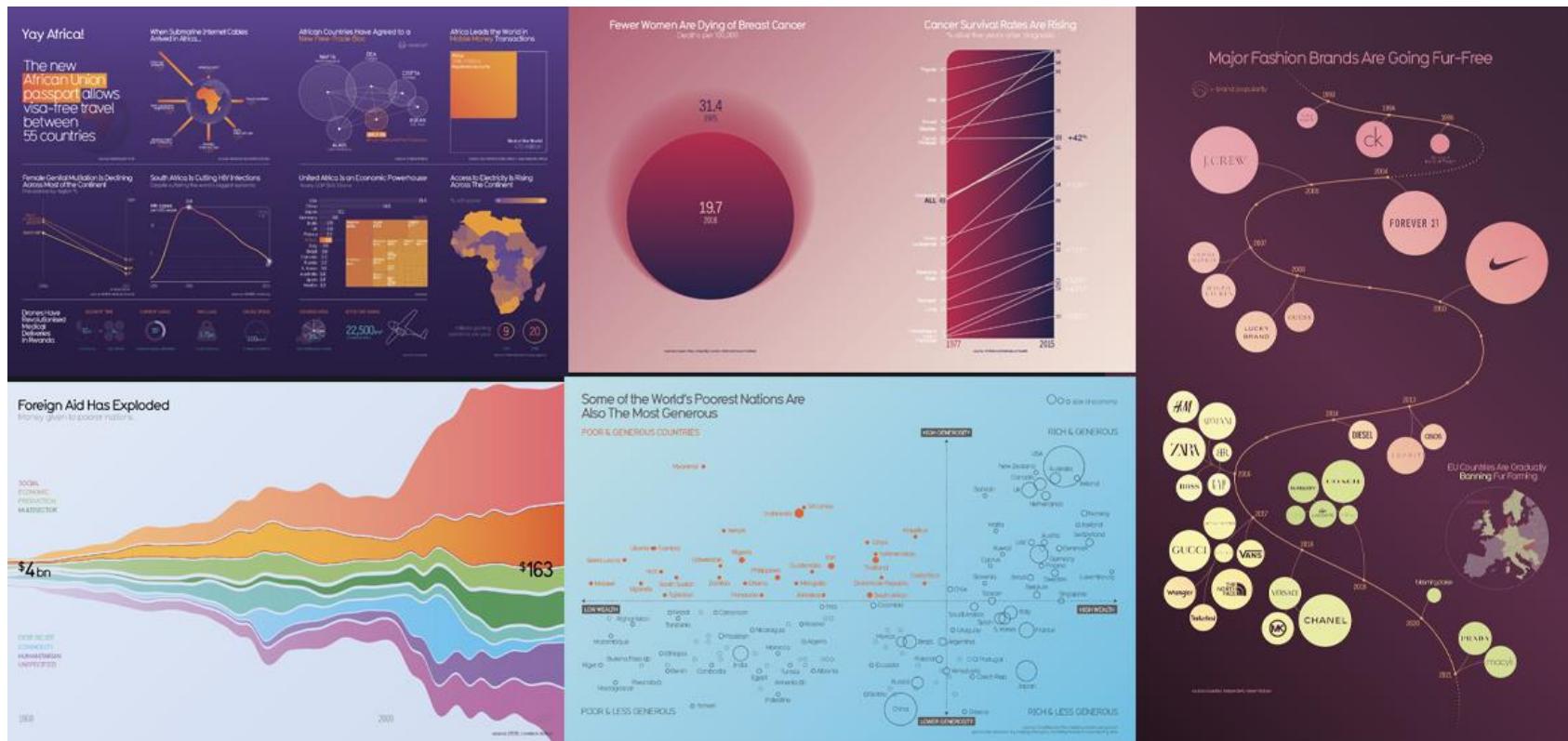
History of Modern Visualization

- First visualization conference: 1990
- Conferences:
 - IEEE SciVis,
 - IEEE InfoVis, IEEE VAST
 - IEEE PacificVis, EuroVis, etc.
- Journals
 - IEEE Transactions on Visualization and Computer Graphics
 - Computer Graphics Forum
 - Computers & Graphics, etc.



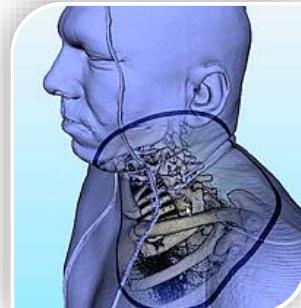
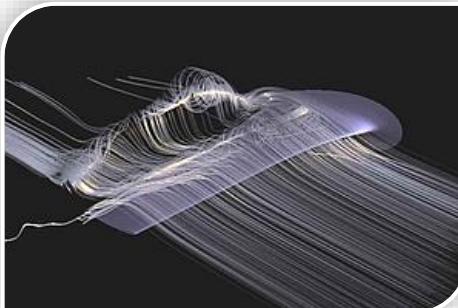
History of Modern Visualization

- Infographics
- Data journalism (NY Times, Die Zeit, etc.)
- Web-based visualizations (D3.js, Vega, ChartJS, etc.)



Visualization

Definitions, goals, and major areas

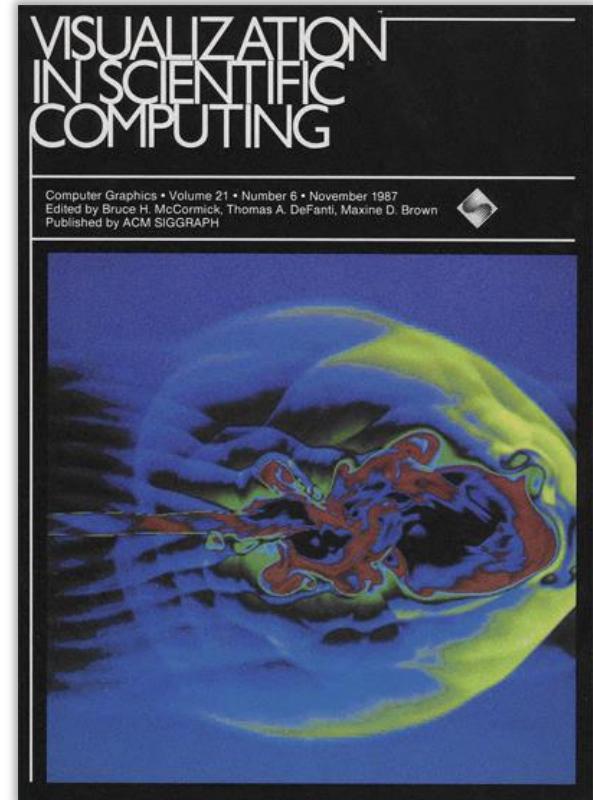


Visualization – Definitions

- McCormick, DeFanti, Brown 1987:

“Visualization is a method of computing. It **transforms the symbolic into the geometric**, enabling researchers to observe their simulations and computations. Visualization offers a method for **seeing the unseen**. It enriches the process of scientific discovery and fosters profound and unexpected **insights**.

[...] It studies those mechanisms in **humans and computers** which allow them in concert to perceive, use and communicate visual information.”



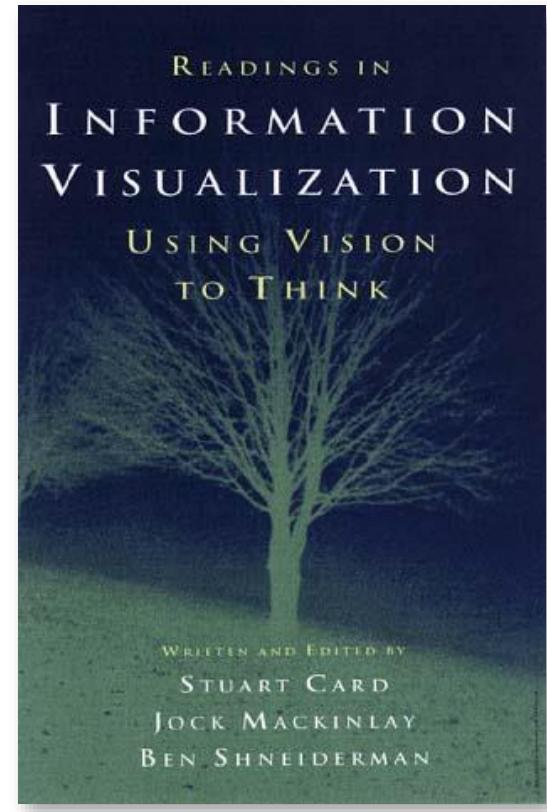
Visualization – Definitions

- Card, MacKinlay, Shneiderman 1999:

“[Information / Scientific] Visualization ...

The use of **computer-supported, interactive, visual representations** of [abstract/scientific] **data** to **amplify cognition**.”

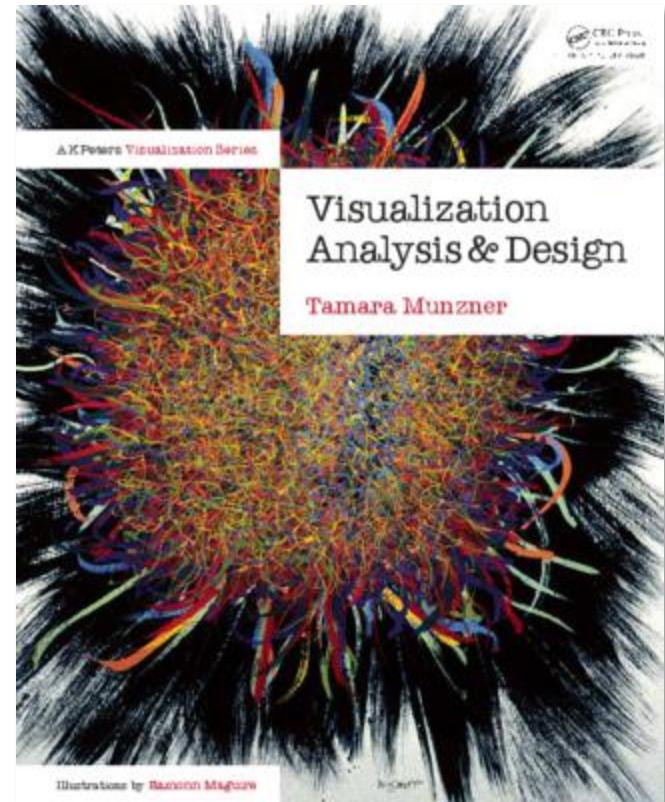
“The purpose of visualization is **insight**,
not pictures.” [B. Shneiderman]



Visualization – Definitions

- Munzner 2014:

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”

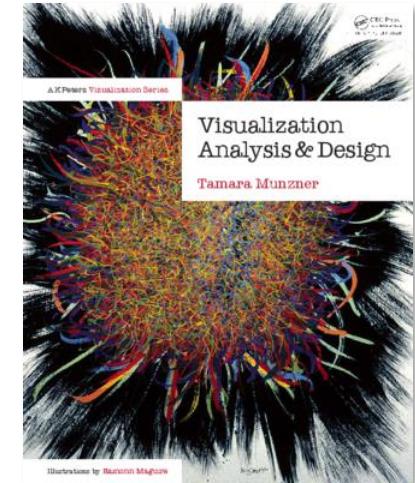


Visualization – Definitions

- Munzner 2014:

“Computer-based visualization systems provide visual representations of **datasets** designed to help **people** carry out tasks more effectively”

- Why have a **human** in the loop?
 - No need for vis when fully automatic solution exists that **can be trusted**, e.g.,
 - if question can be answered by a compact, precise query
 - if a decision can be automated (e.g., stock market trading)

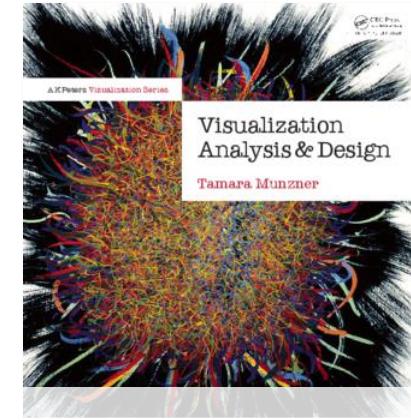


Visualization – Definitions

- Munzner 2014:

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

4 datasets with identical statistics



Identical statistics

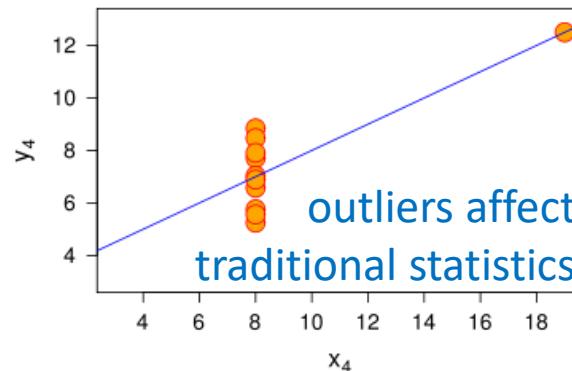
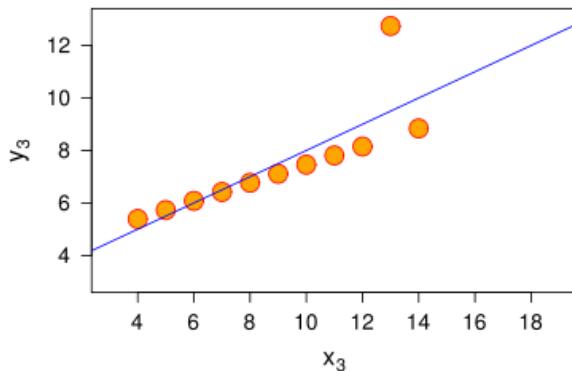
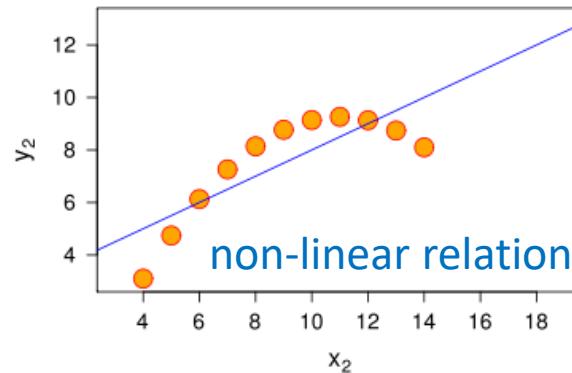
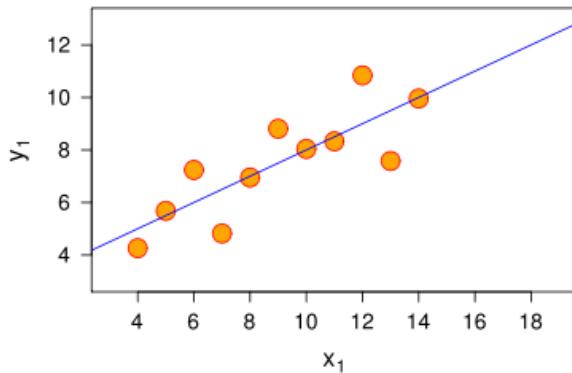
x mean	9.0
y mean	7.5
x variance	10.0
y variance	3.75
x/y correlation	0.82
regression line	$y = 3 + 0.5x$

[Anscombe, 1973]

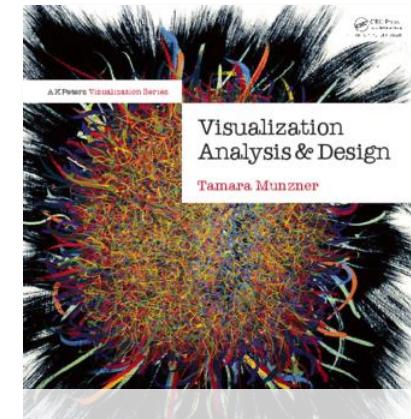
Visualization – Definitions

- Munzner 2014:

Statistics don't always tell you the full story!



4 datasets with identical statistics



Identical statistics

x mean	9.0
y mean	7.5
x variance	10.0
y variance	3.75
x/y correlation	0.82
regression line	$y = 3 + 0.5x$

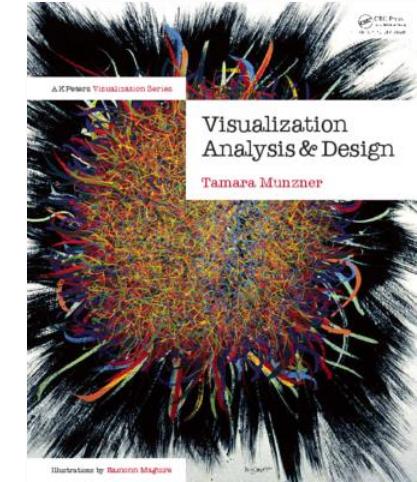
[Anscombe, 1973]

Visualization – Definitions

- Munzner 2014:

“Computer-based visualization systems provide visual representations of **datasets** designed to help **people** carry out tasks more effectively”

- Why have a human in the loop?
 - Statistics only work for well-specified problems
 - Many analysis problems are ill-defined
 - Don't know exactly what questions to ask in advance
 - Detect the expected & discover the unexpected [Thomas & Cook]

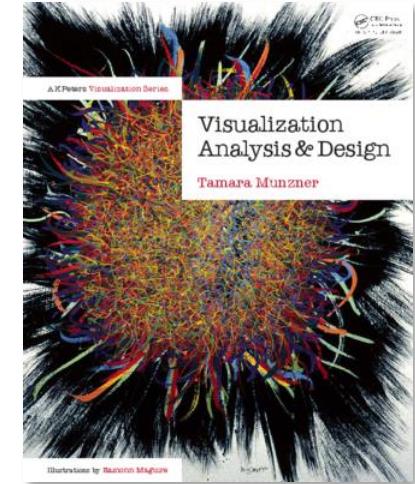


“Visualization is suitable when there is a need to **augment** human capabilities **rather than replace** people with computational decision-making methods.”

Visualization – Definitions

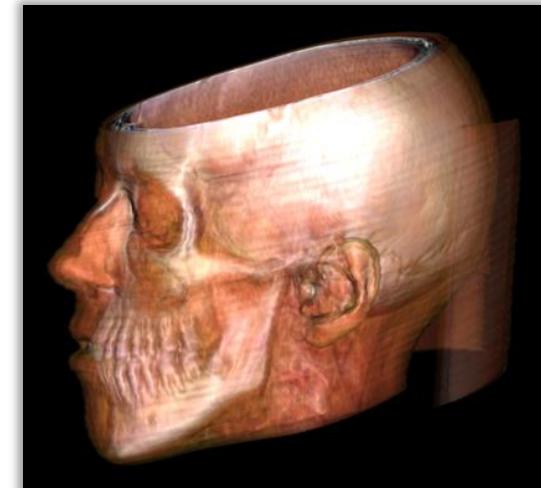
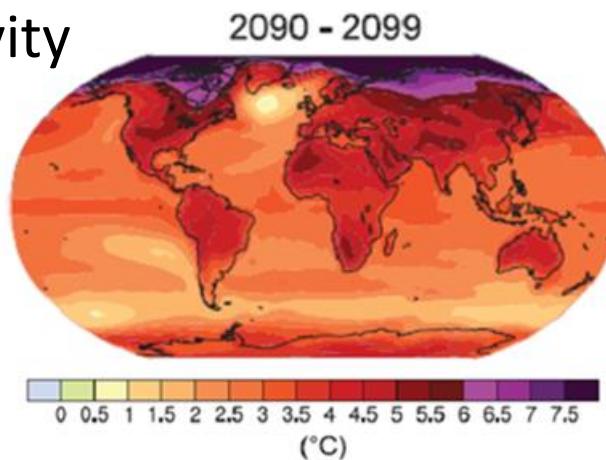
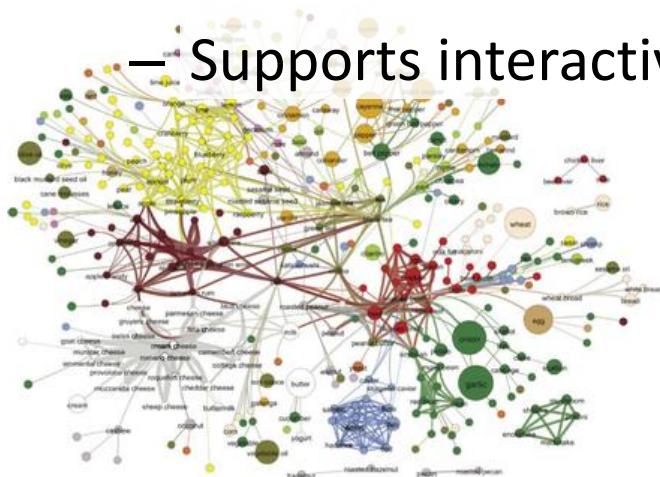
- Munzner 2014:

“**Computer-based** visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”



- Why have a **computer** in the loop?

- Large datasets are infeasible to draw by hand
- Goes beyond human capacities / patience
- Supports interactivity





Search Task

- Spot the differences



Comparison Task

Visualization – Definitions

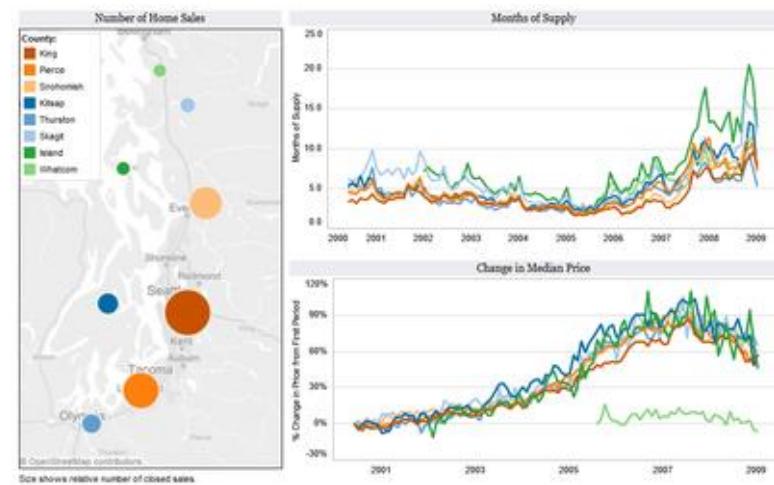
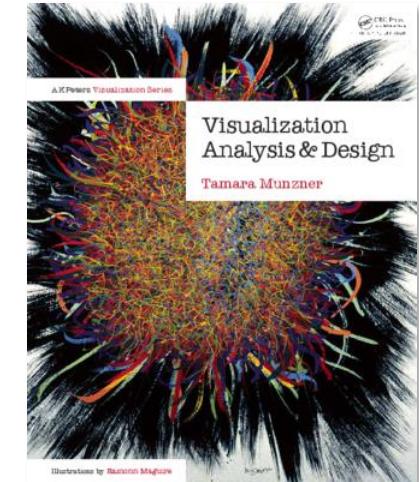
- Munzner 2014:

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”

- Why use **interactivity**?
 - Handle data complexity
 - A single static view can show only one aspect of data

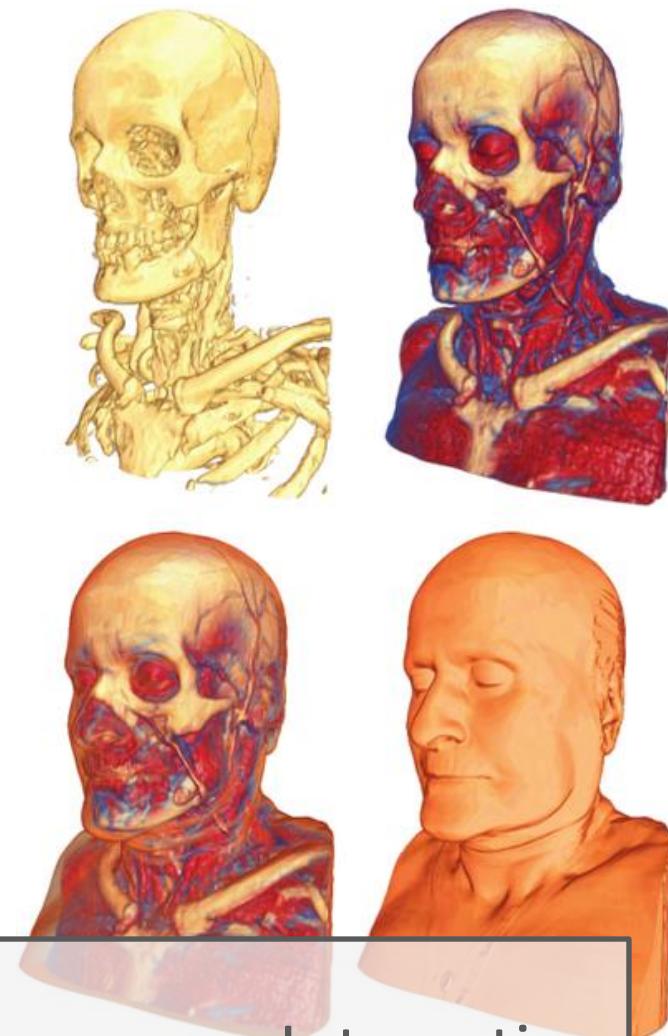
“Overview first, zoom and filter, then details-on-demand.”

[B. Shneiderman]



www.tableau.com/stories/gallery/real-estate-prices

Visualization – Definitions



Interaction

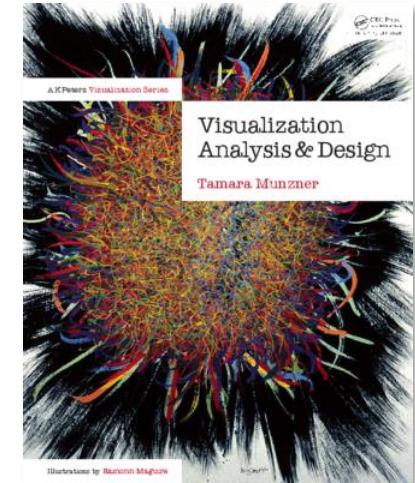
The screenshot displays the VolumeShop application interface. The main workspace on the right shows a 3D rendering of a human head with a brown, textured surface. To the left, the 'Resources' panel lists various materials: Volume (256,256,230), Skin (256,256,1), Bone (256,256,1), Muscles (256,256,1), Vessels (256,256,1), Grey (256,256,1), Orange (256,256,1), Black (256,256,1), Gold (282,282,1), Metal (256,256,1), Purple (256,256,1), Escher (256,256,1), Glossy (256,256,1), and Stipples (256,256,1). Below the resources is the 'Environments' panel, which includes a 'Style Transfer Function Editor' with a preview showing a checkered pattern with colored nodes.

Visualization – Definitions

- Munzner 2014:

“Computer-based visualization systems provide **visual** representations of datasets designed to help people carry out tasks more effectively”

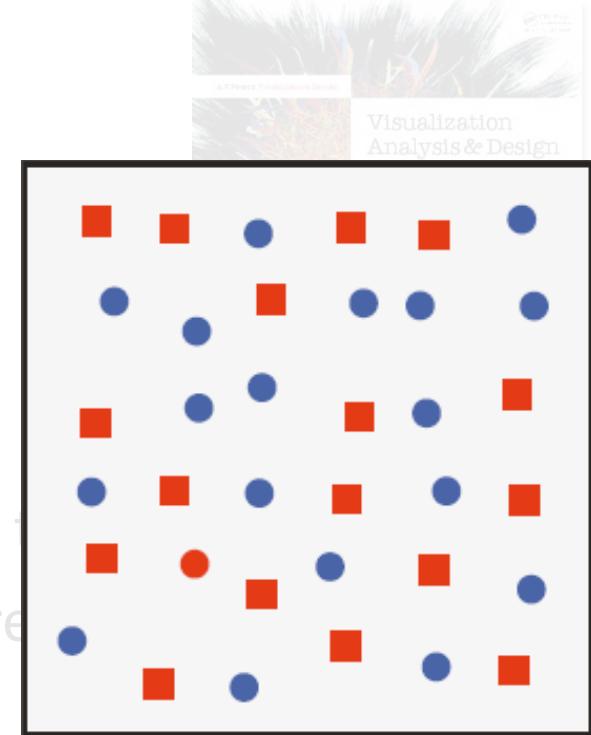
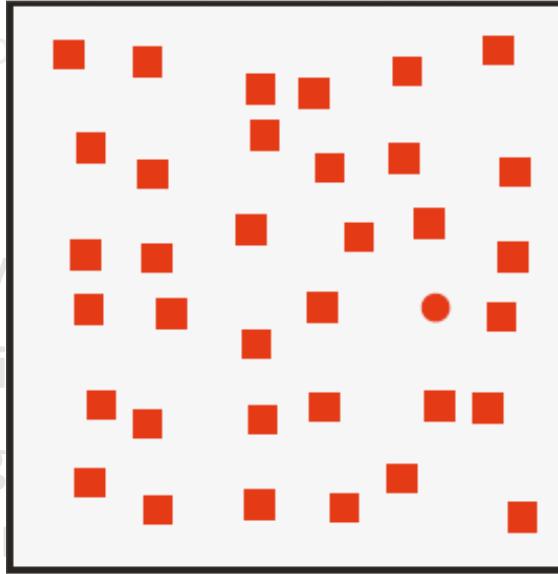
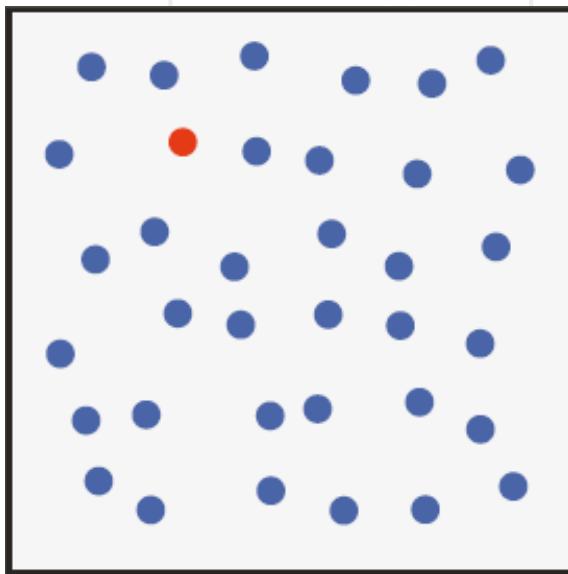
- Why depend on **vision**?
 - Visual system is high-bandwidth channel to brain
 - Detect interesting visual structures and relationships (e.g., anomalies, patterns, or trends)
 - Sequential vs. parallel processing (popout)



Visualization – Definitions

- Munzner 2014:

“Computer-based visualization systems provide **visual** representations of datasets



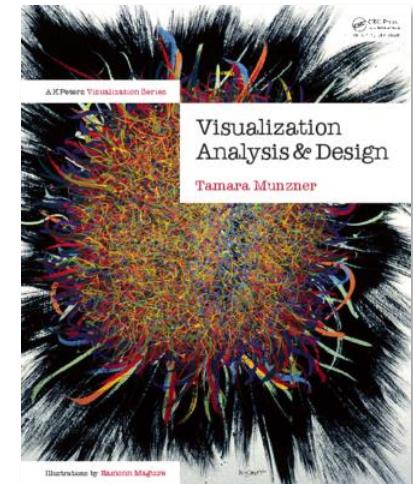
- Sequential vs. parallel processing (popout)

Visualization – Definitions

- Munzner 2014:

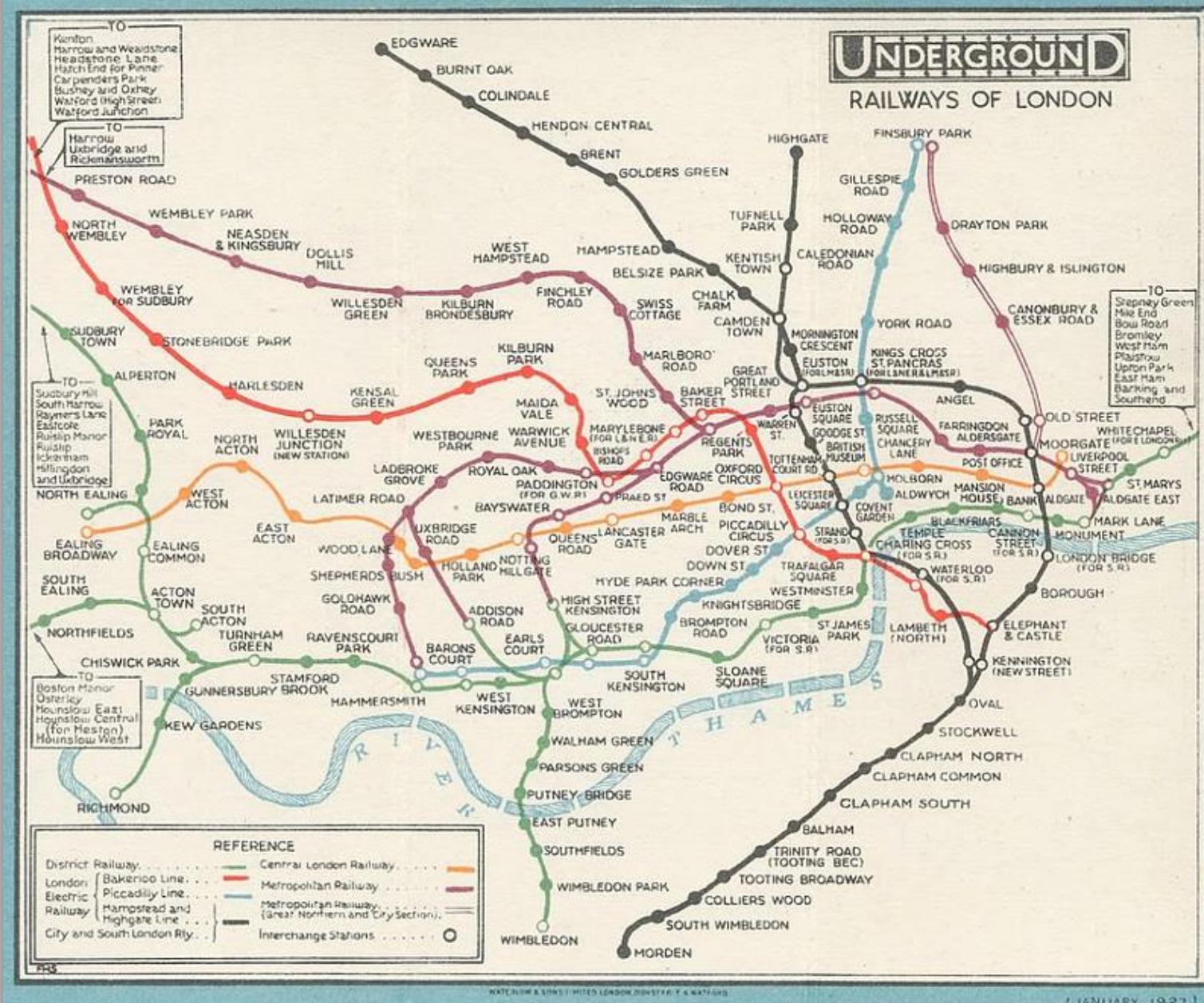
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out **tasks** more **effectively**”

- Why focus on **tasks & effectiveness?**
 - What problem do we want to solve?



UNDERGROUND

RAILWAYS OF LONDON

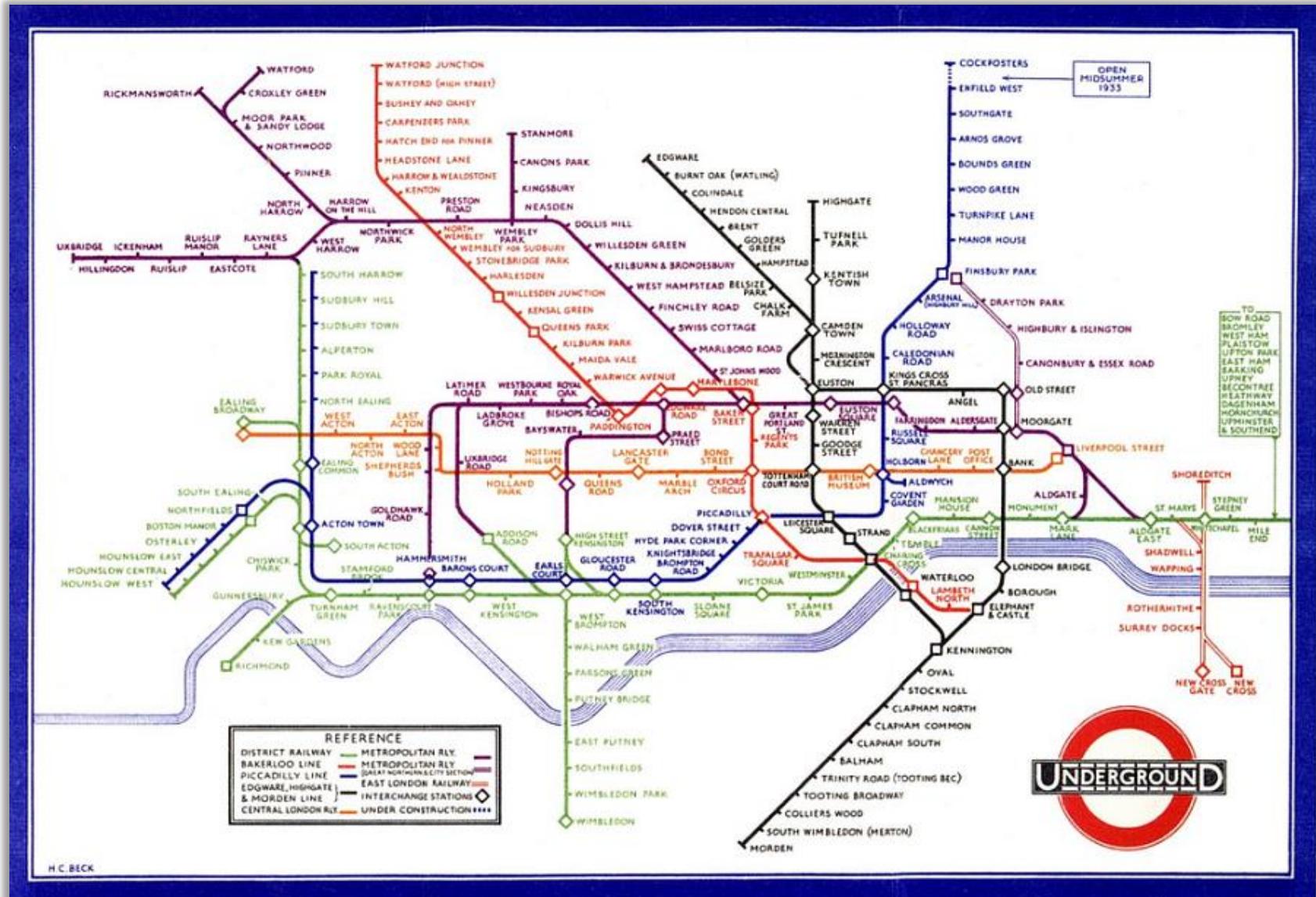


London underground map
Fred Stingemore (1927)

Which station is closest to me?

Visualization – Definitions

SIEMENS
Ingenuity for life



Redesign by Harry Beck (1933)

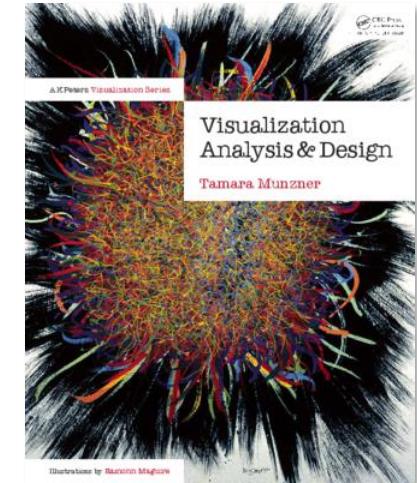
What's the fastest path from A to B?

Visualization – Definitions

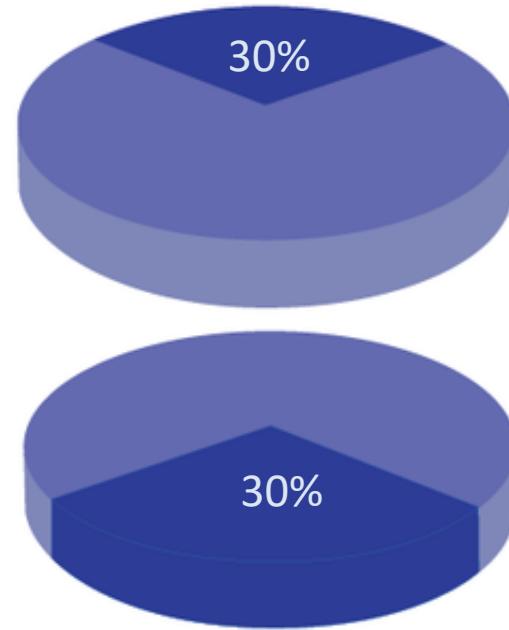
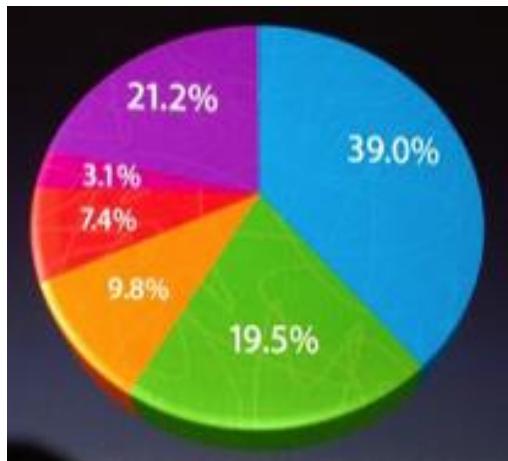
- Munzner 2014:

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out **tasks** more **effectively**”

- Why focus on tasks & effectiveness?
 - What problem do we want to solve?
 - A tool serving one task can be poorly suited for another one
 - Most possibilities are ineffective
 - Representation should be correct, accurate, and truthful



Visualization – Definitions



Since when is 19.5% bigger than 21.2%?

- Representation should be correct, accurate, and **truthful**

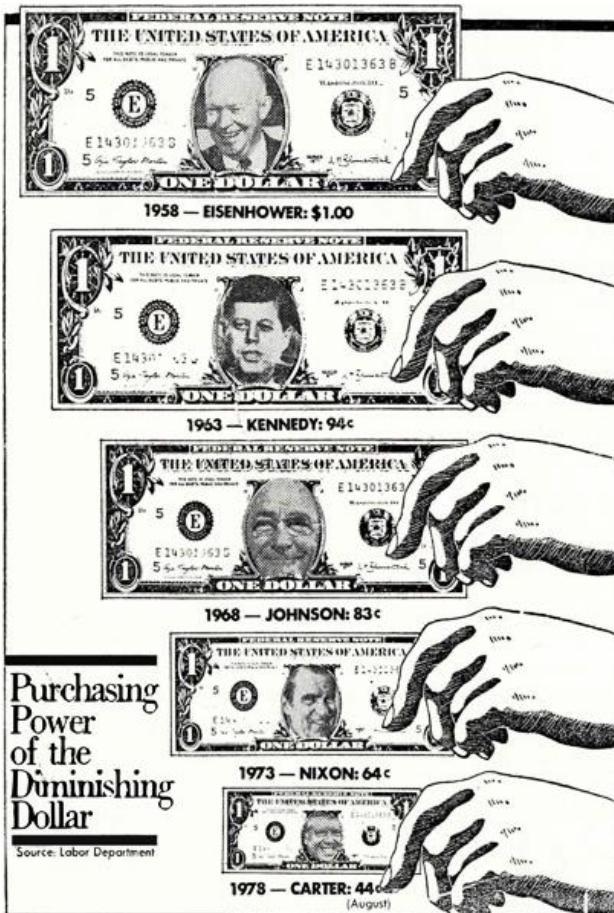
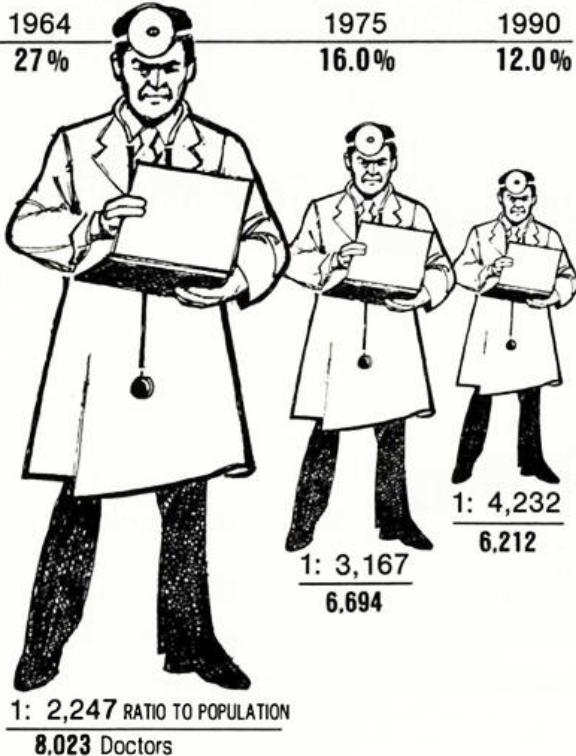
Visualization – Definitions

SIEMENS
Ingenuity for life

THE SHRINKING FAMILY DOCTOR In California

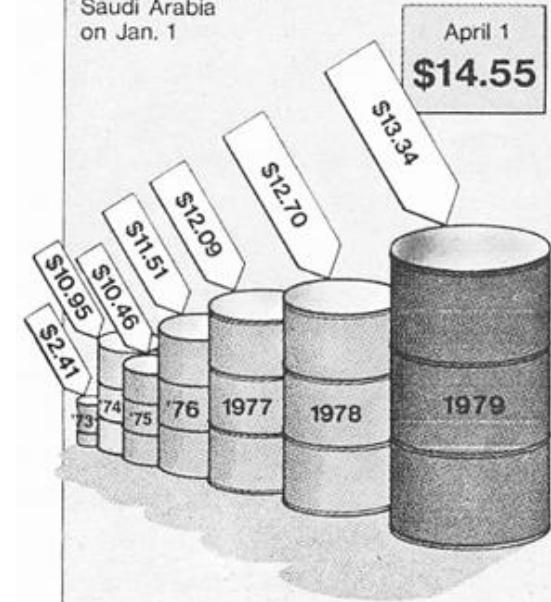
Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



IN THE BARREL...

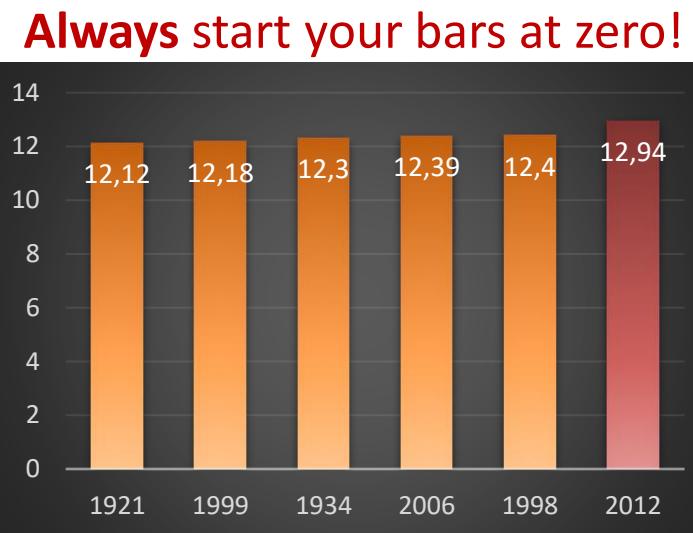
Price per bbl. of
light crude, leaving
Saudi Arabia
on Jan. 1



- Representation should be correct, accurate, and **truthful**

$$\text{Lie factor} = \frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$

Visualization – Definitions



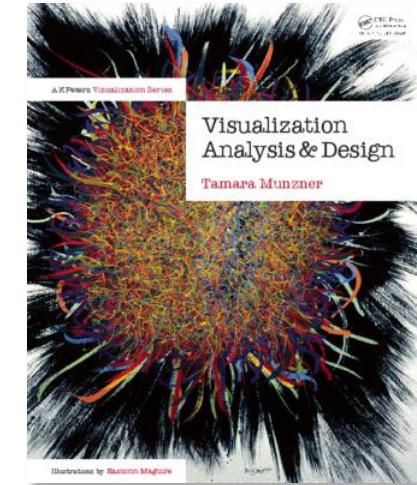
- Representation should be correct, accurate, and **truthful**

Visualization – Definitions

- Munzner 2014:

“Visualization designers must take into account three very different kinds of resource limitations: those of **computers**, of **humans**, and of **displays**”

- Computational limits
 - Processing time / system memory
- Display limits
 - Number of pixels
 - **Information density:** ratio of used space vs. unused whitespace
- Human limits
 - Perception, attention and memory (e.g., change blindness)



Visualization – Definitions



- Human limits
 - Perception, attention and memory (e.g., change blindness)

Visualization – Definitions



- Human limits
 - Perception, attention and memory (e.g., change blindness)

Visualization – Definitions



- Human limits
 - Perception, attention and memory (e.g., change blindness)

Visualization – Goals

Visualization is good for

– **Visual exploration**

- find unknown/unexpected
- generate new hypotheses

Nothing is known
about the data

– **Visual analysis** (confirmative vis.)

- confirm or reject hypotheses
- information drill-down

There are hypotheses

– **Presentation**

- effective/efficient communication of results

“Everything” is known

Visualization – Major Areas

- Major areas

- Volume Visualization
- Flow Visualization

} Scientific Visualization

Inherent spatial reference

3D

-
- Information Visualization
 - Visual Analytics

nD

Usually no spatial reference

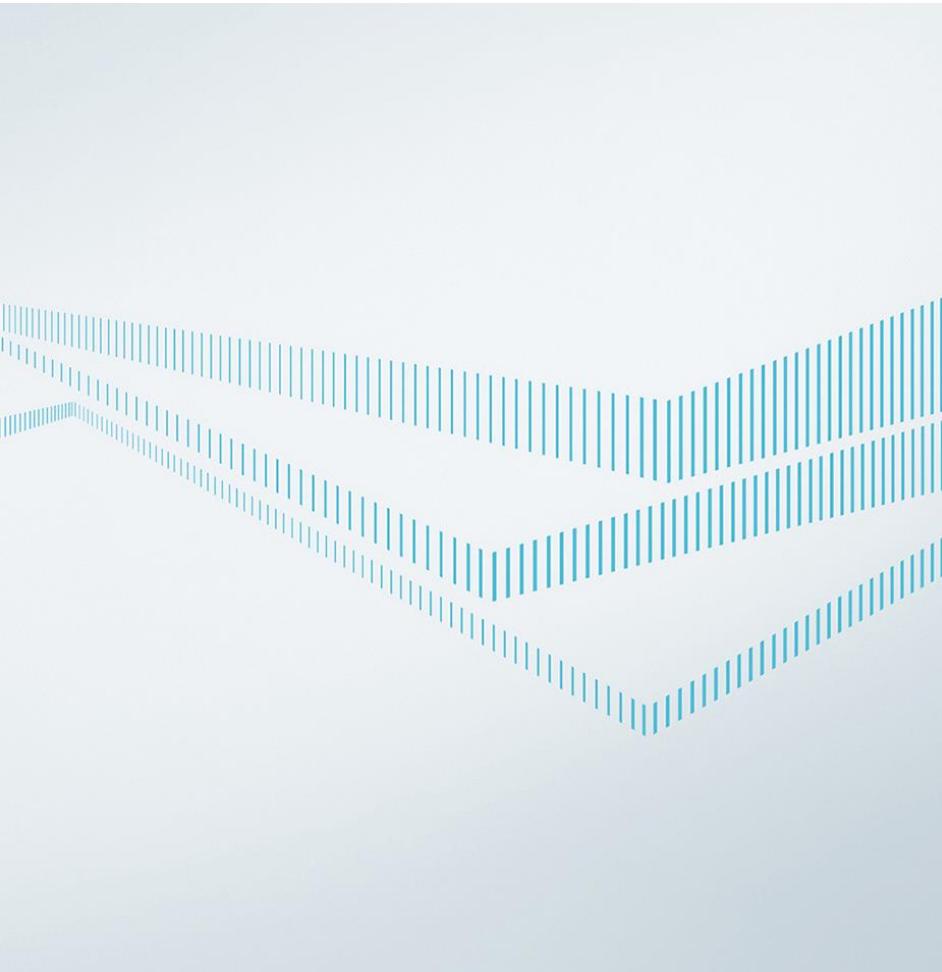
Visualization – Major Areas

- Nowadays, borders between SciVis / InfoVis / Visual Analytics are not that well defined any more
 - Information visualization may also include spatial data (e.g., geographic data)
 - Need for integration of abstract data in spatial data

Acknowledgements

- M. Eduard Gröller
- Helwig Hauser
- Torsten Möller
- Tamara Munzner
- many more

Contact information



Dr. Johannes Kehrer

Siemens Technology
T RTA BAM IBI-DE
Otto-Hahn-Ring 6
81739 München, Deutschland

E-mail:
kehrer.johannes@siemens.com

Internet
siemens.com/innovation