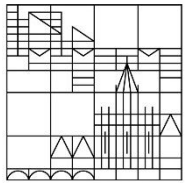
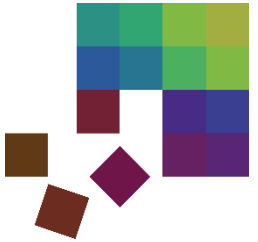


University of Konstanz
Data Analysis and Visualization Group

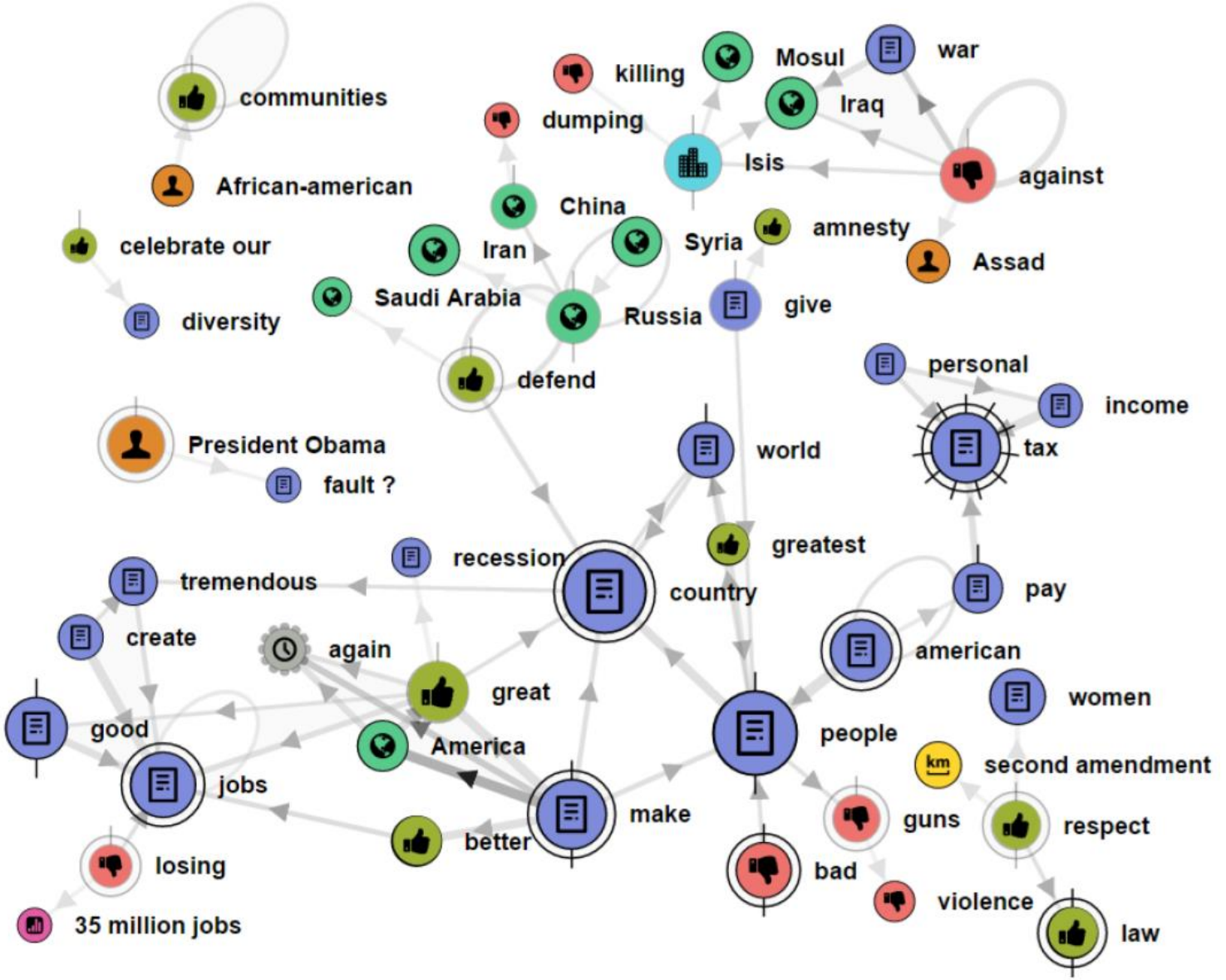


Visual Analytics for Linguistics

Raphael Buchmüller

Computational Linguistics Fall School 2024

Presidential Debate Visualization
Visual Analytics for Linguistics



NEREx (2017)

by Mennatallah El-Assady et al.

Mennatallah El-Assady, Rita Sevastjanova, Bela Gipp, Daniel Keim, and Christopher Collins. *NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations*. Computer Graphics Forum, vol. 36, no. 3, pp. 213-225, 2017.

Simple Text Visualization Pipeline

Below is a partial transcript of the exchange between Trump and Harris while the debate was ongoing.

DAVID MUIR: Tonight, the high-stakes showdown here in Philadelphia between Vice President Kamala Harris and former president Donald Trump. Their first face-to-face meeting in this presidential election. Their first face-to-face meeting ever.

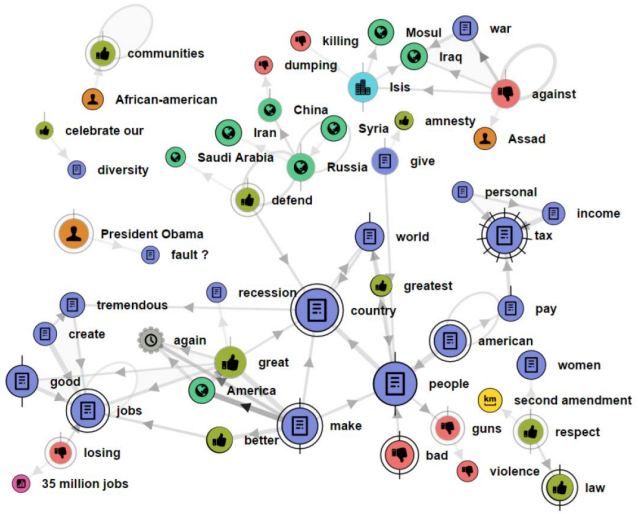
LINSEY DAVIS: A historic race for president upended just weeks ago. President Biden withdrawing after his last debate. Donald Trump is now up against a new opponent.

DAVID MUIR: The candidates separated by the smallest of margins. Essentially tied in the polls nationally. And in the key battlegrounds, including right here in Pennsylvania, all still very much in play. The ABC News Presidential Debate starts right now.

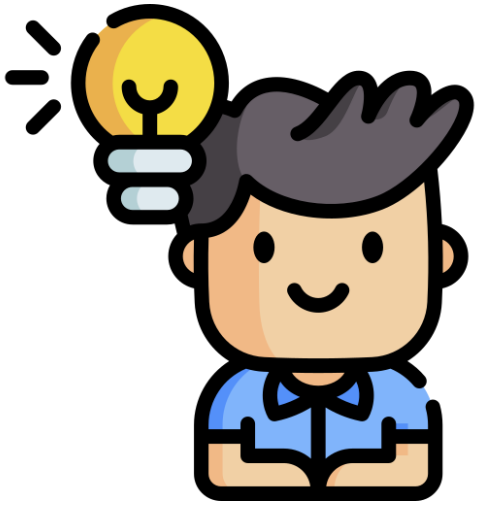
DAVID MUIR: Good evening, I'm David Muir. And thank you for joining us for tonight's ABC News Presidential Debate. We want to welcome viewers watching on ABC and around the world tonight. Vice President Kamala Harris and President Donald Trump are just moments away from taking the stage in this unprecedented race for president.

LINSEY DAVIS: And I'm Linsey Davis. Tonight's meeting could be the most consequential event of their campaigns, with Election Day now less than two months away. For Vice President Kamala Harris, this is her first debate since President Biden withdrew from the race on July 21st. Of course, that decision followed his debate against President Donald Trump in June. Since then, this race has taken on an entirely new dynamic.

Text Data



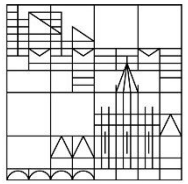
Visualization

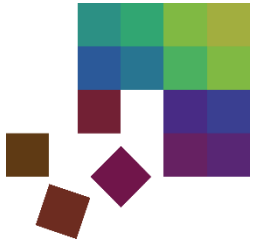


Knowledge

Whats the Schedule?

	Theory (1:30-3pm)	Practice (3:30-5pm)
Monday	Domain and Design	Design your Approach
Tuesday	Text Processing	Process your Text
Wednesday	Visualization Foundations	Implementation 1
Thursday	Text Visualization	Implementation 2
Friday	Projects, Experiences and Discussion	Present your Approach



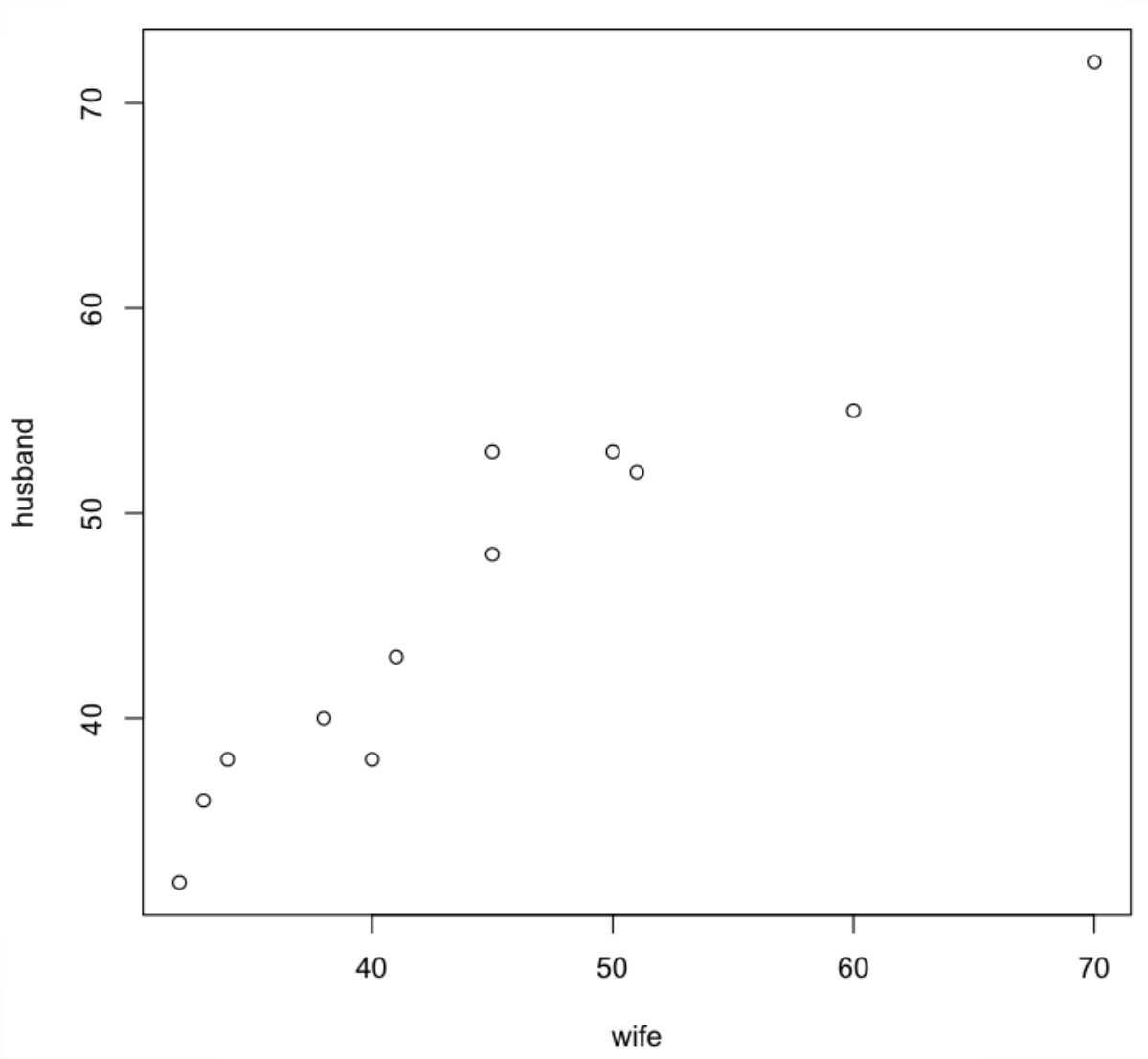


Why Visualization?

Wife (age)	Husband (age)
45	53
60	55
32	32
33	36
45	48
70	72
50	53
34	38
51	52
40	38
41	43
38	40

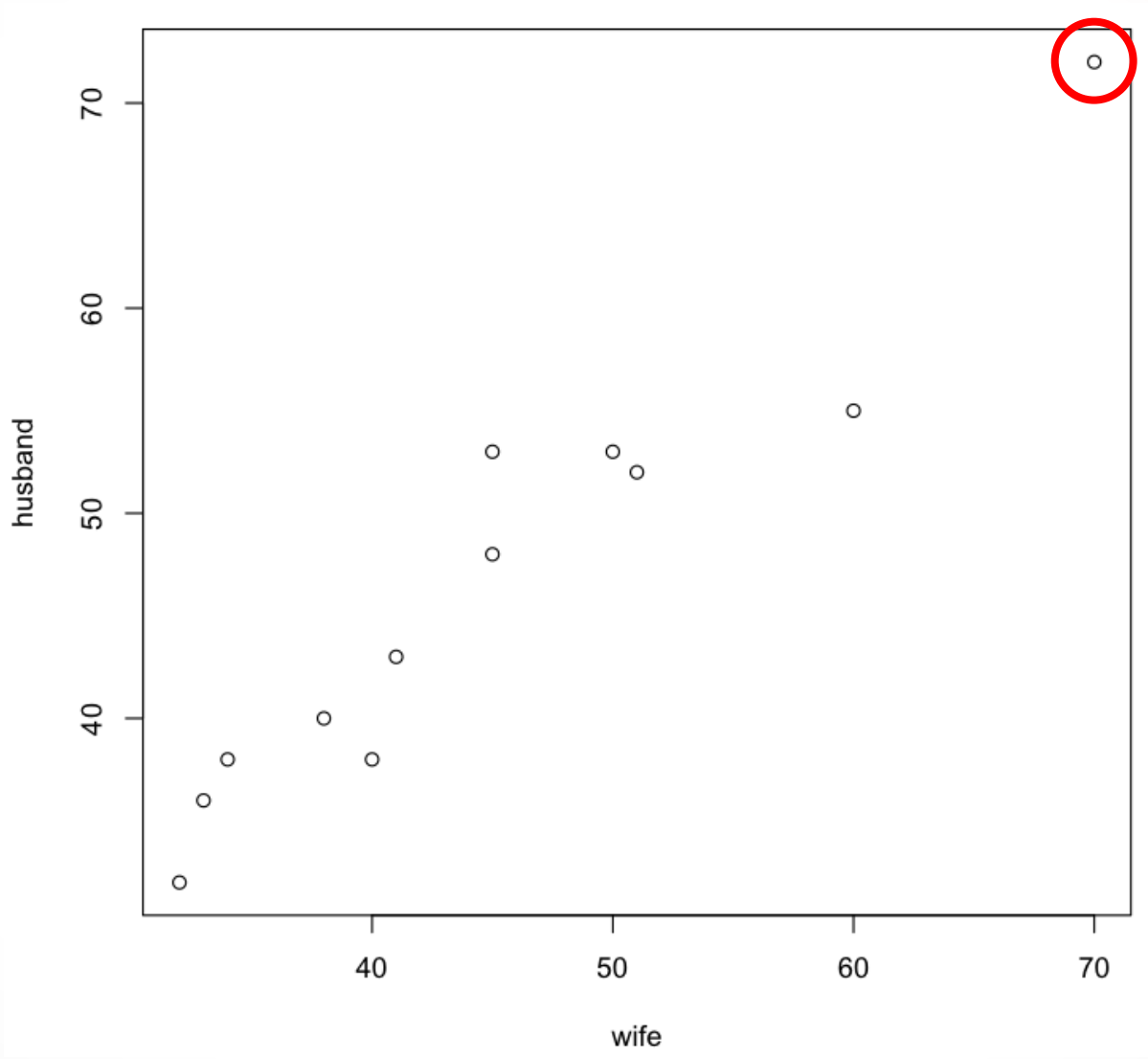
Presentation of Information

Data table showing the age of married couple.



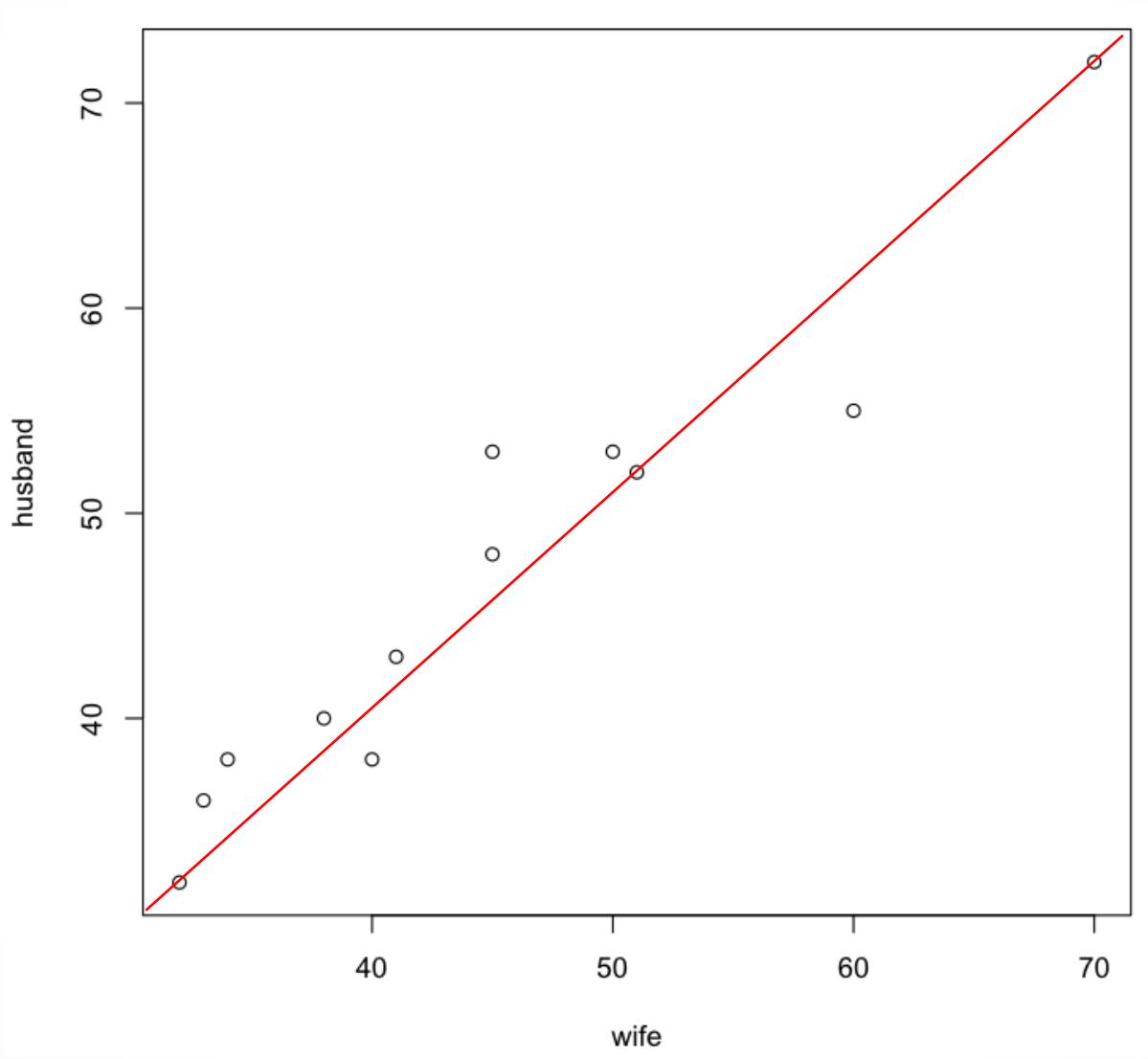
Presentation of Information

Scatterplot showing the age of married couple.



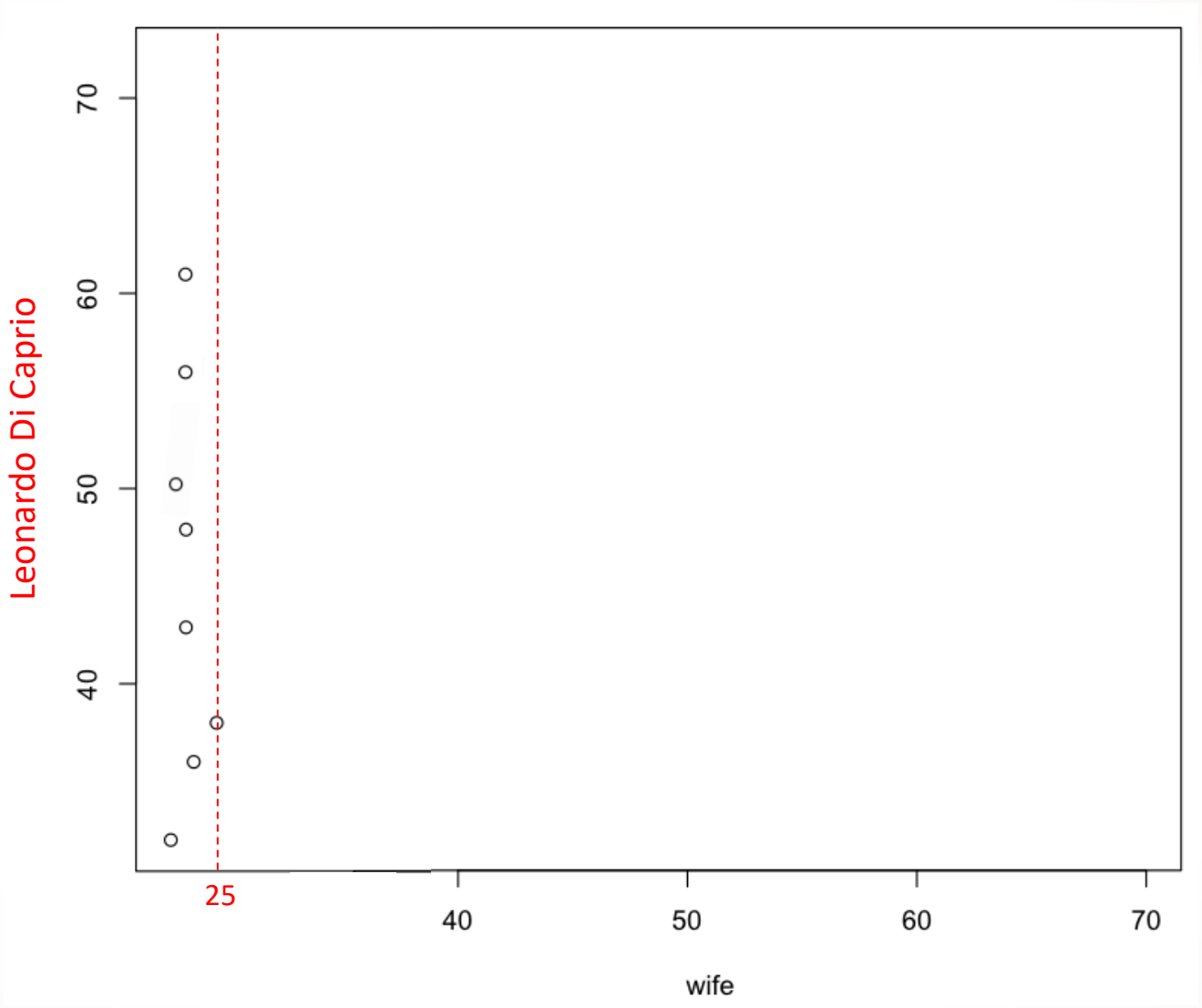
Presentation of Information

Scatterplot showing the age of married couple.



Presentation of Information

Scatterplot showing the age of married couple.



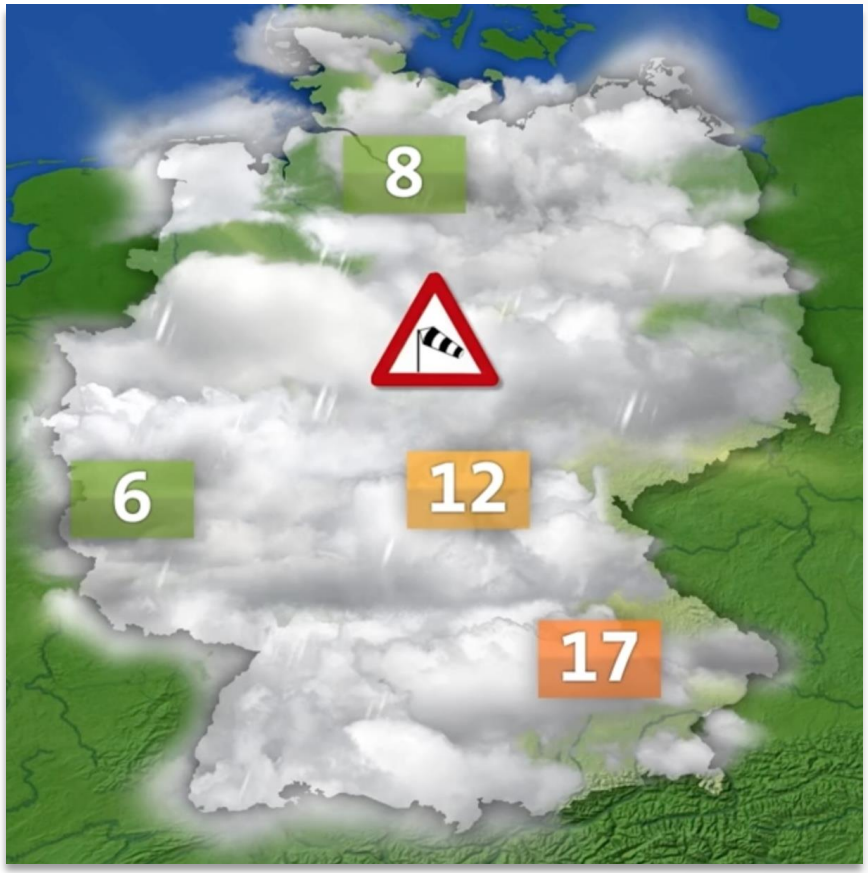
Presentation of Information

Scatterplot showing the age of married couple.

Who is looking at the data?

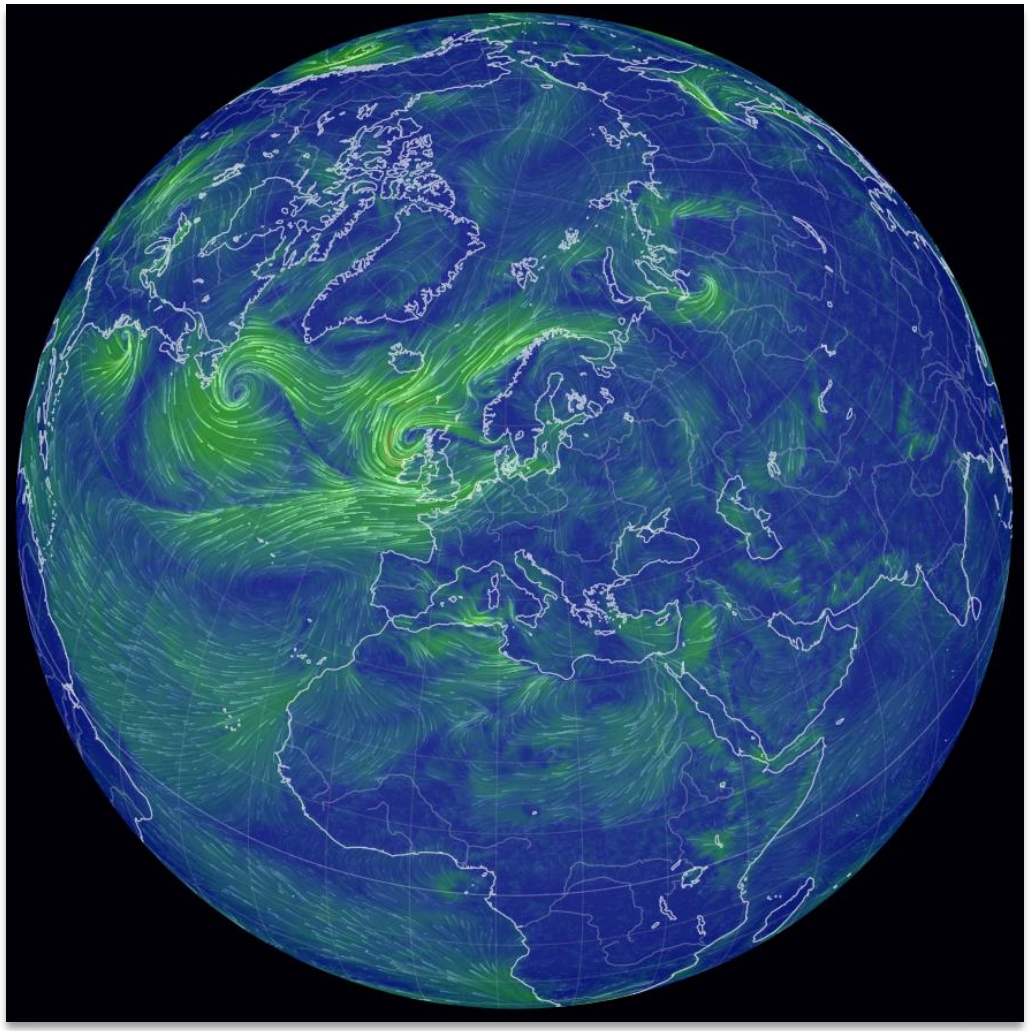
Weather forecast – temperature and wind information (1990)

https://www.tagesschau.de/multimedia/video/video659688.html

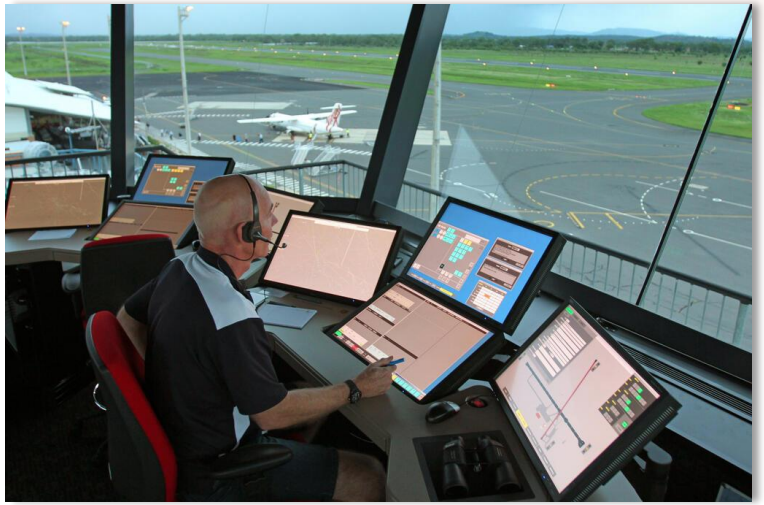


Weather forecast – temperature and wind information (2022)

https://wetter.tagesschau.de/

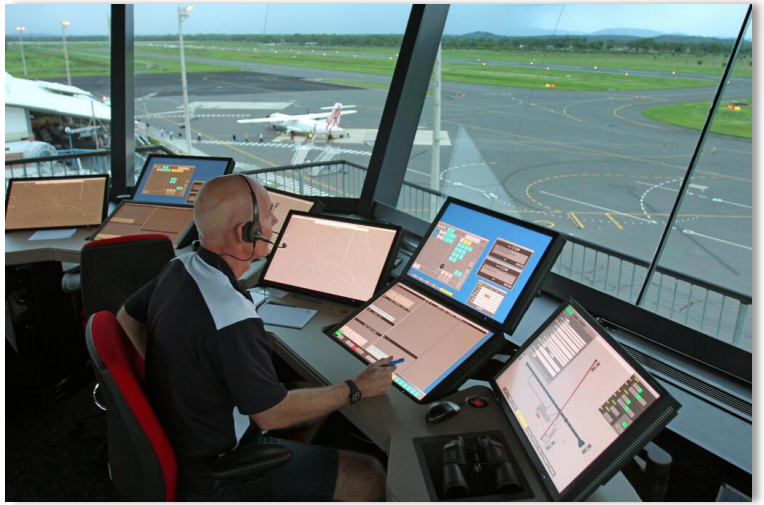


Weather forecast – temperature and wind information (2024)



Who is looking at the data?

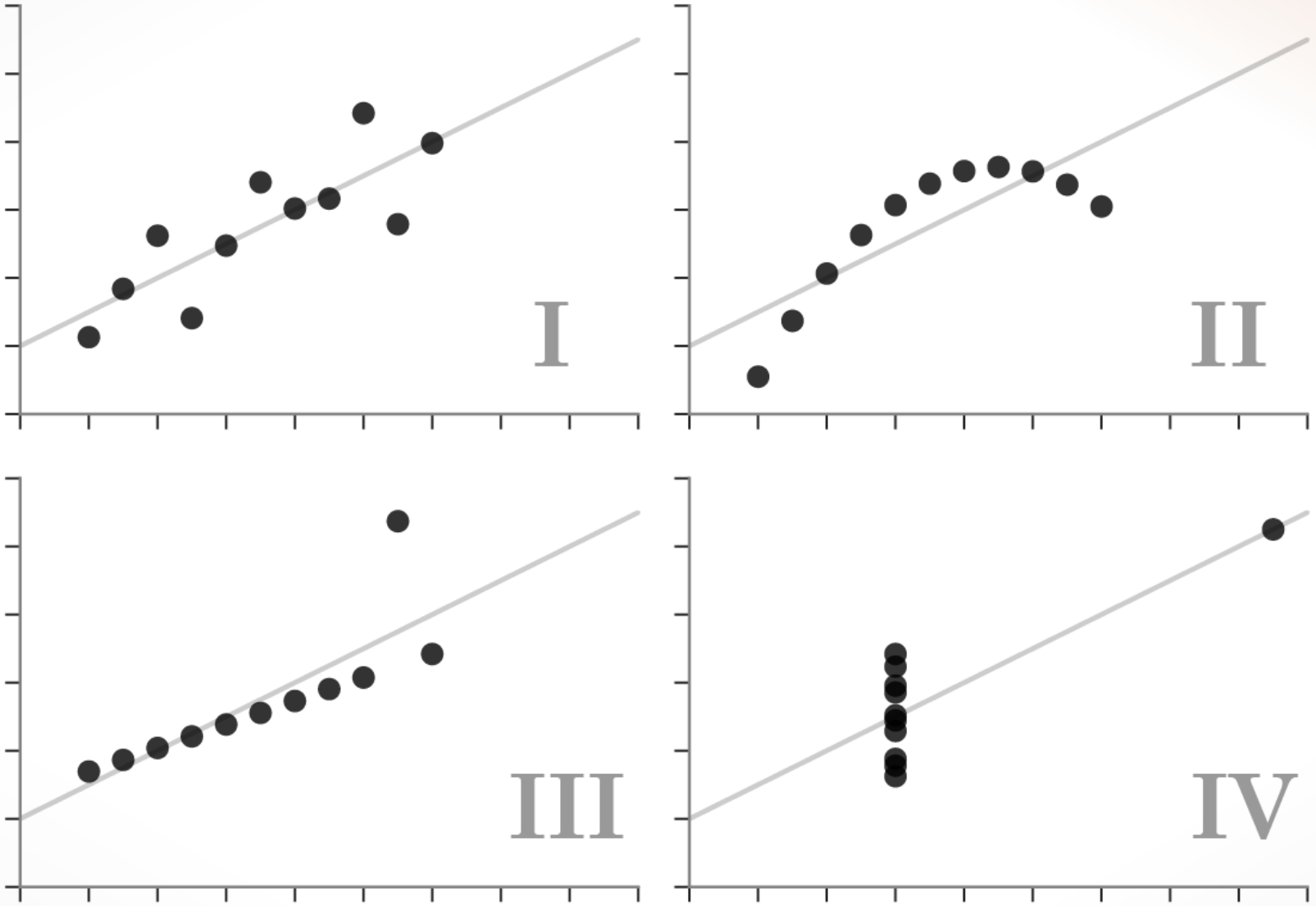
https://www.openinventor. https://grafana.com/blog/2021/09/27/how-to-visualize-real-time-data-from-an-iot-smart-home-weather-station-with-grafana-dashboards/com/



Who is looking at the data?

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

Why should we use visualizations?



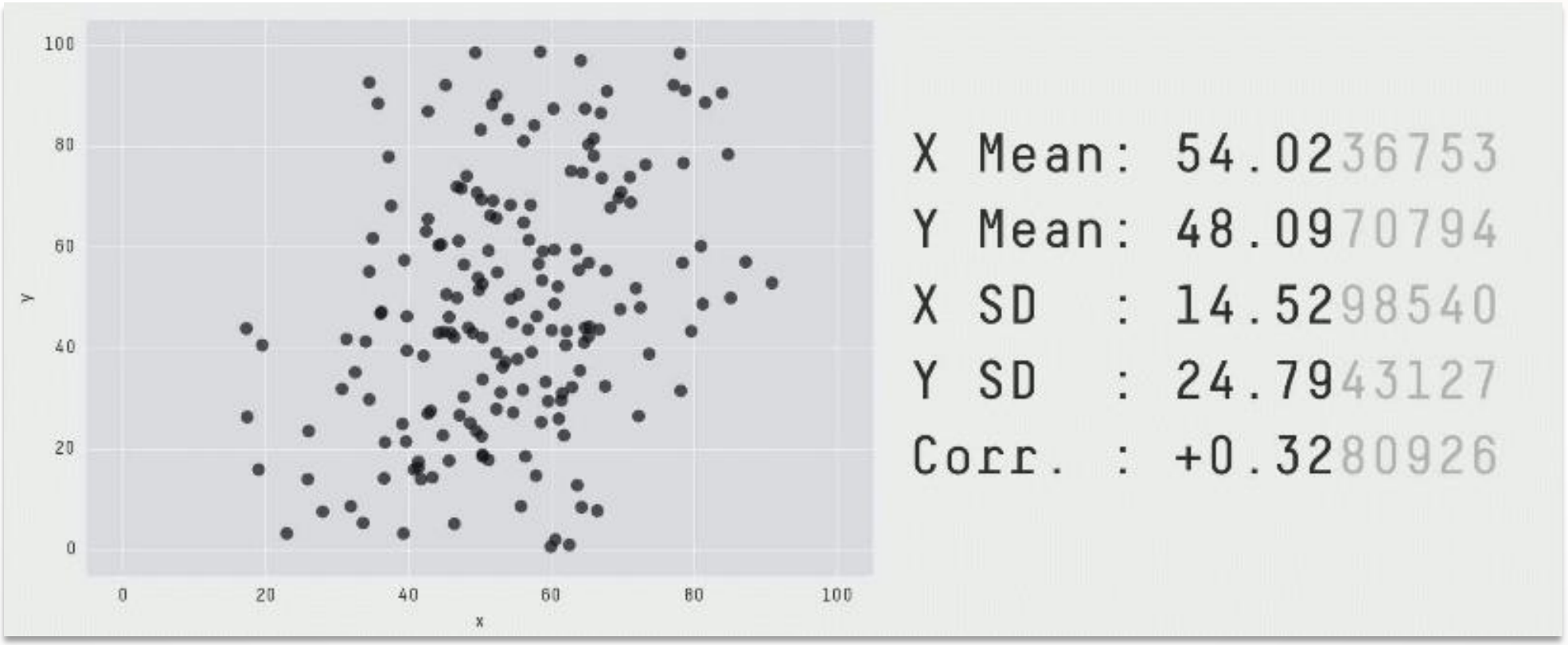
Why visualizations are necessary

Anscombe‘s Quartett: Each dataset has the same summary statistics.

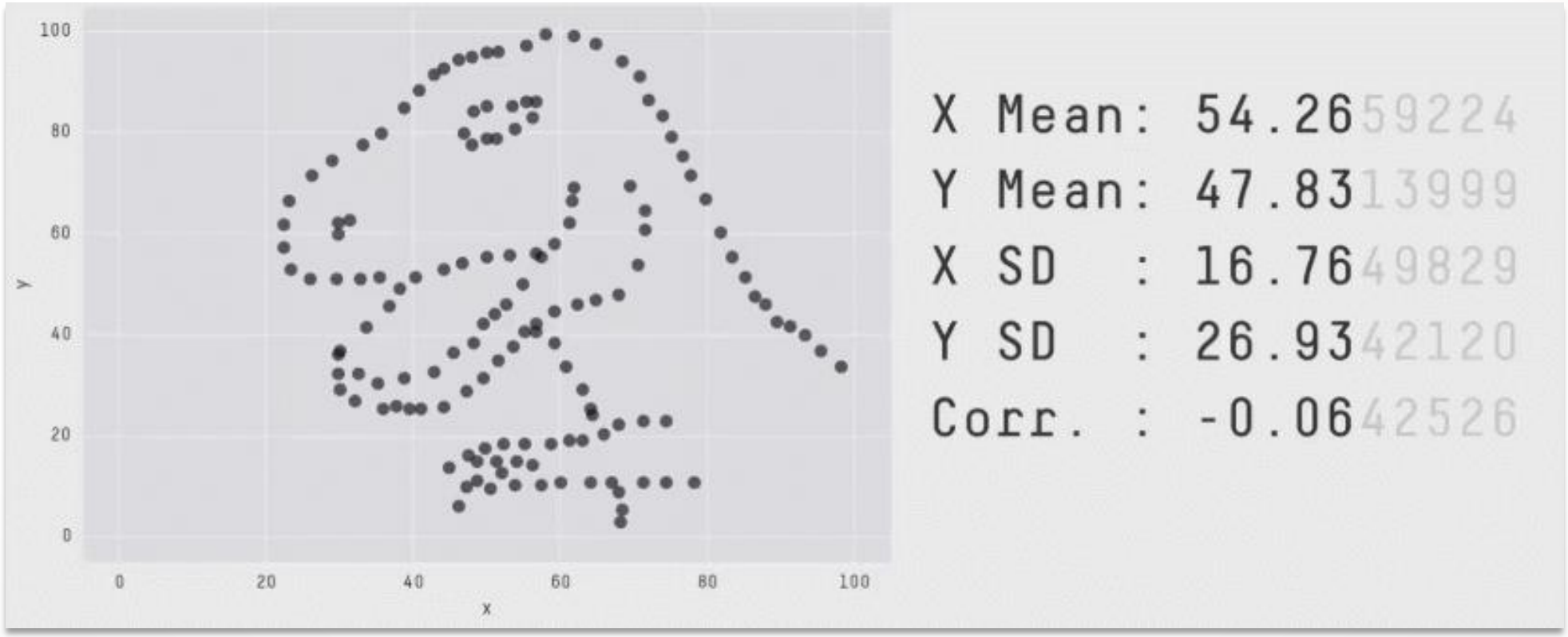
" ...make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding."

F. J. Anscombe

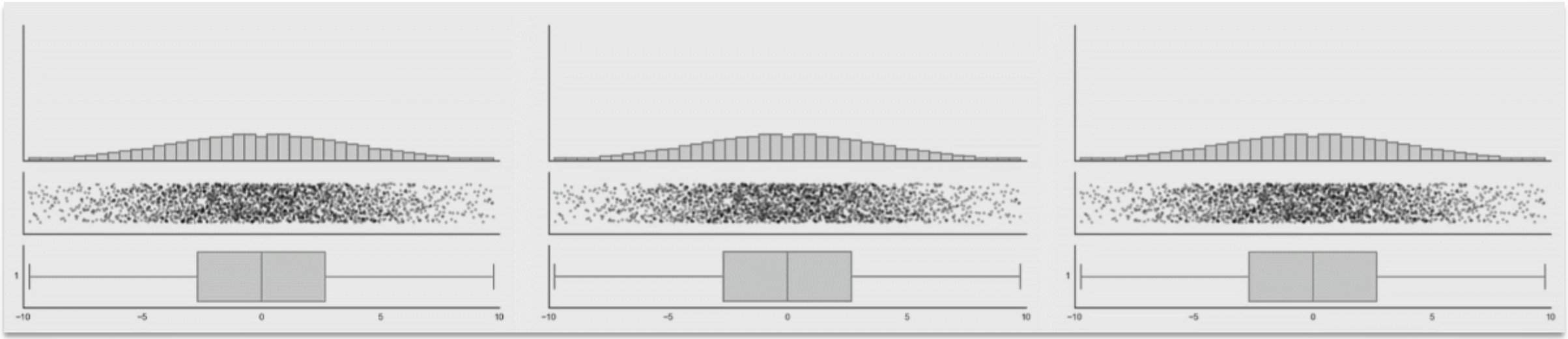
Graphs in Statistical Analysis, The American Statistician, 1973



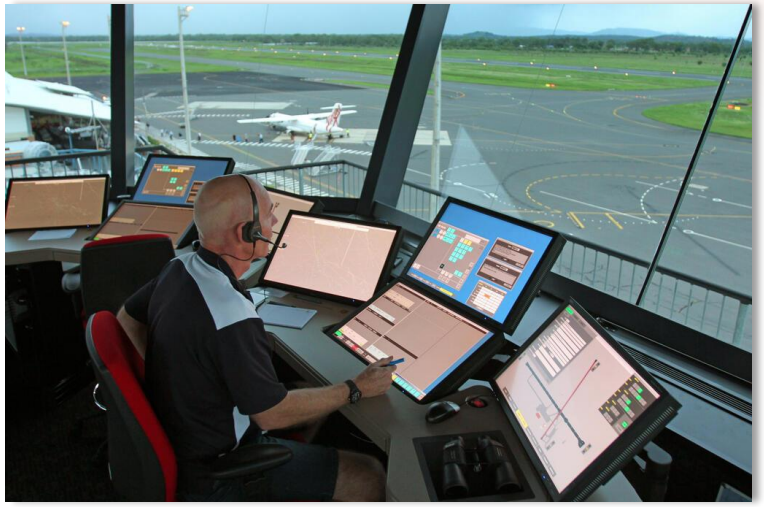
Identical statistics with varied appearance



Identical statistics with varied appearance

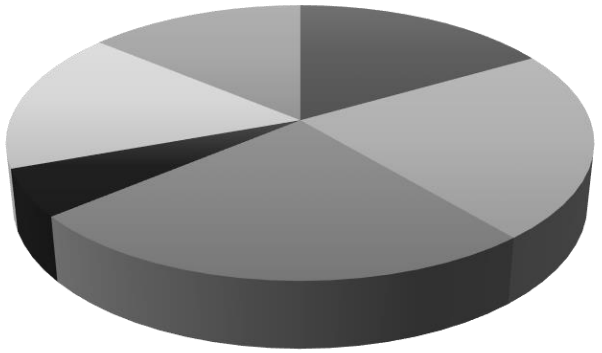


Identical statistics with varied appearance

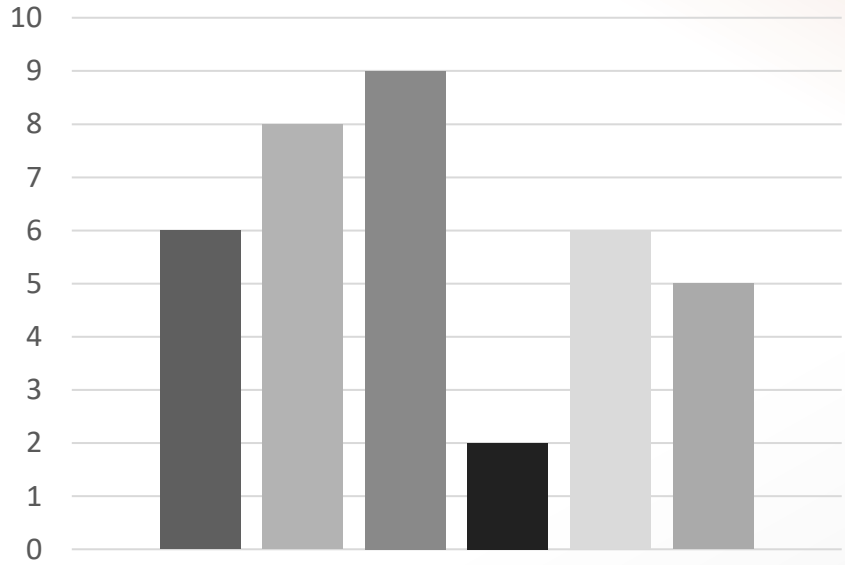


Who is looking at the data?

A

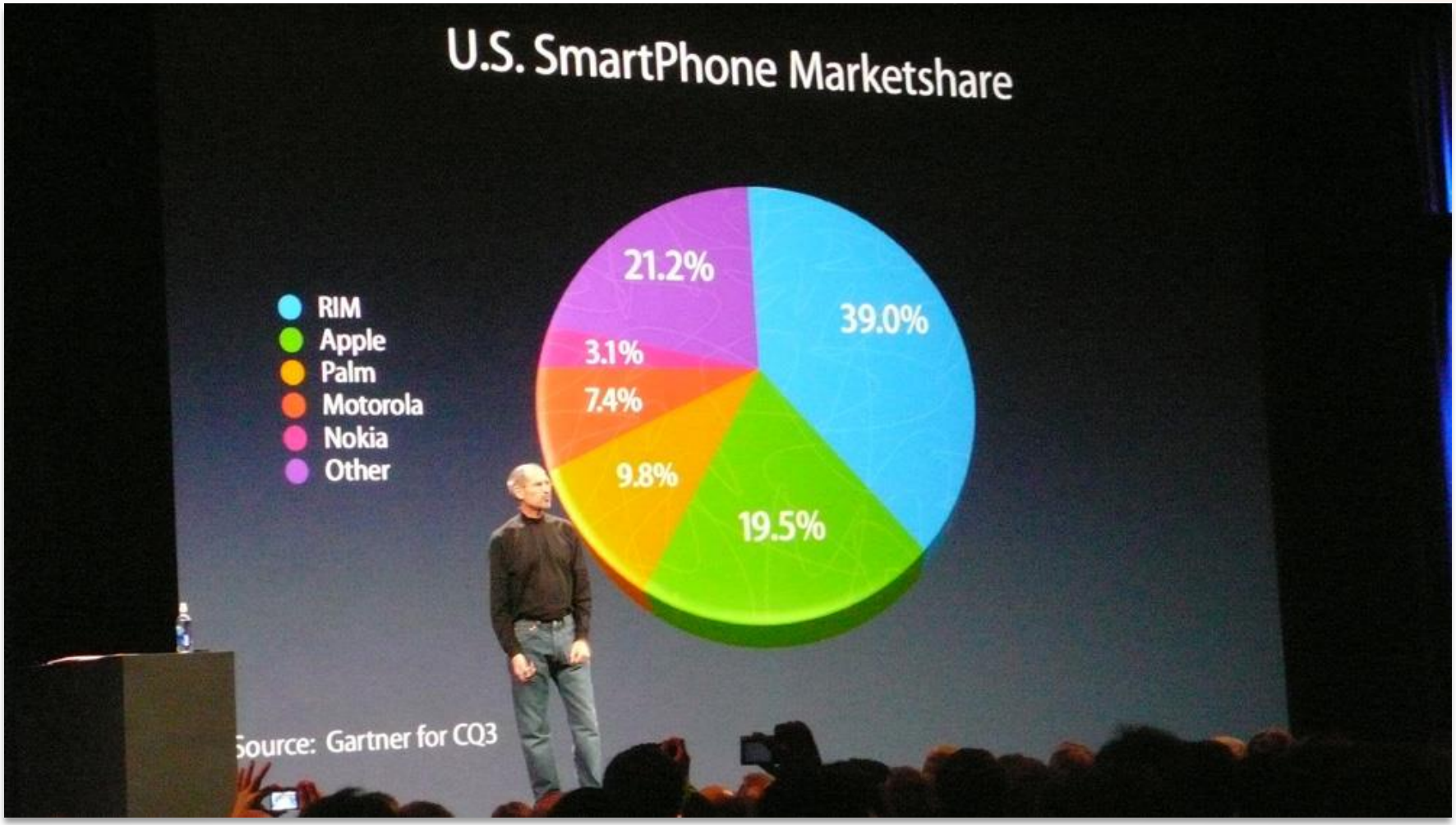


B



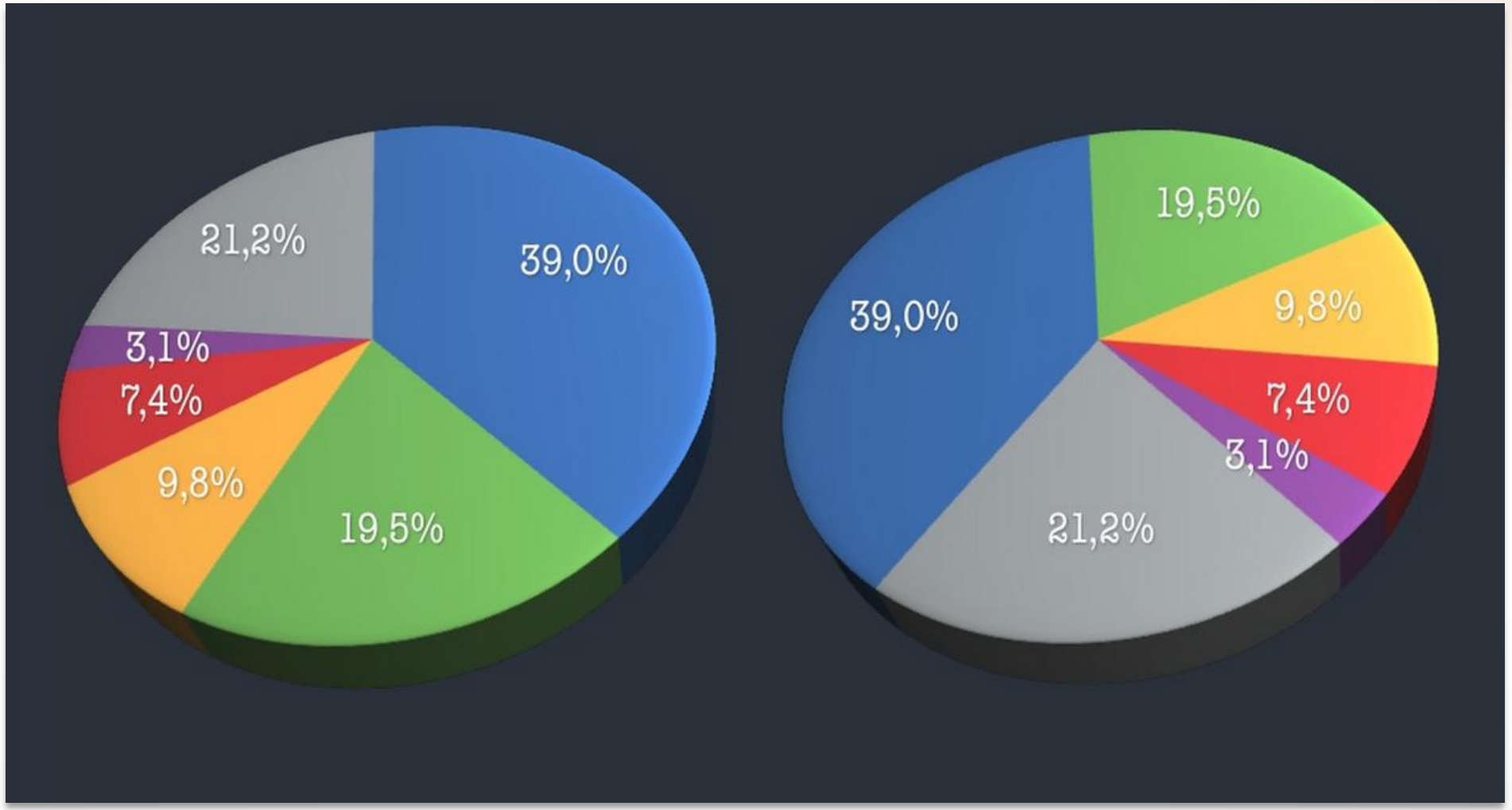
C

Which visualization is the best?



Apple keynote presentation – why 3d is just for marketing purposes

http://stavreas.com/scaling-is-a-race/



Apple keynote presentation – why 3d is just for marketing purposes

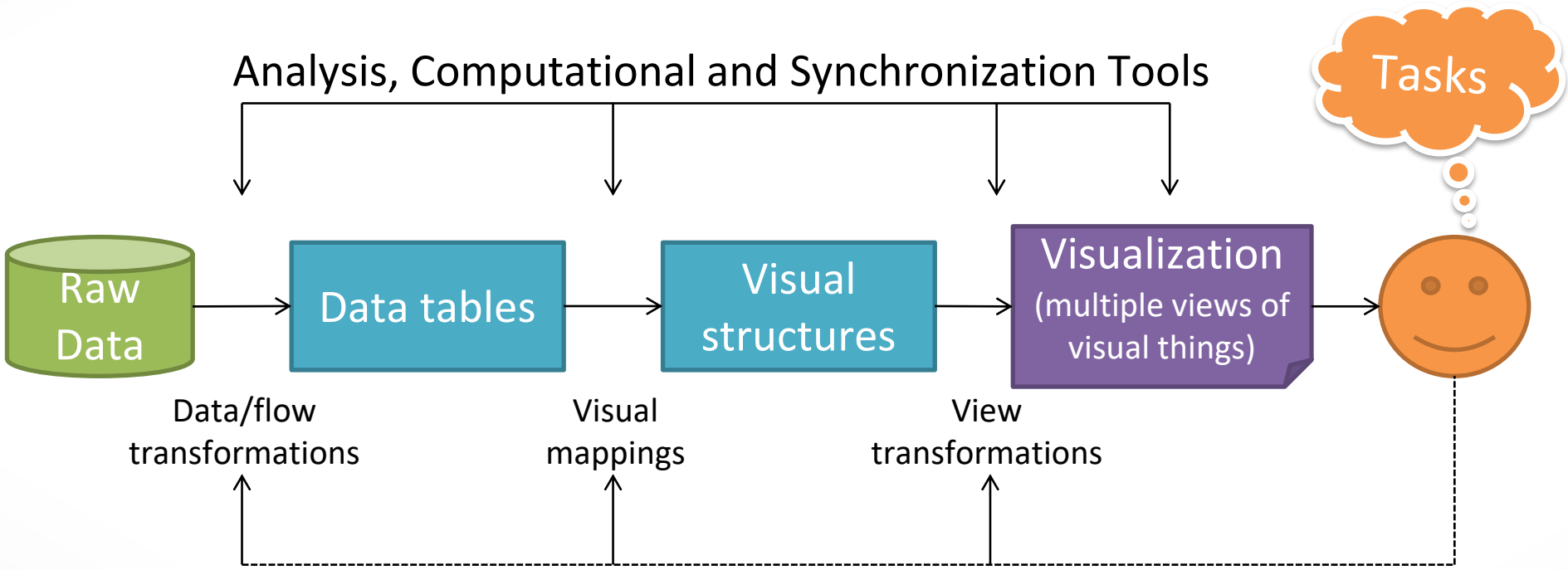
http://stavreas.com/scaling-is-a-race/

The Visualization Process

- Analysis of
  - The type of data.
  - The information the viewer hopes to extract.
- Preprocess the data.
- Define a mapping from the data to the display.
- Provide interactive controls if necessary.

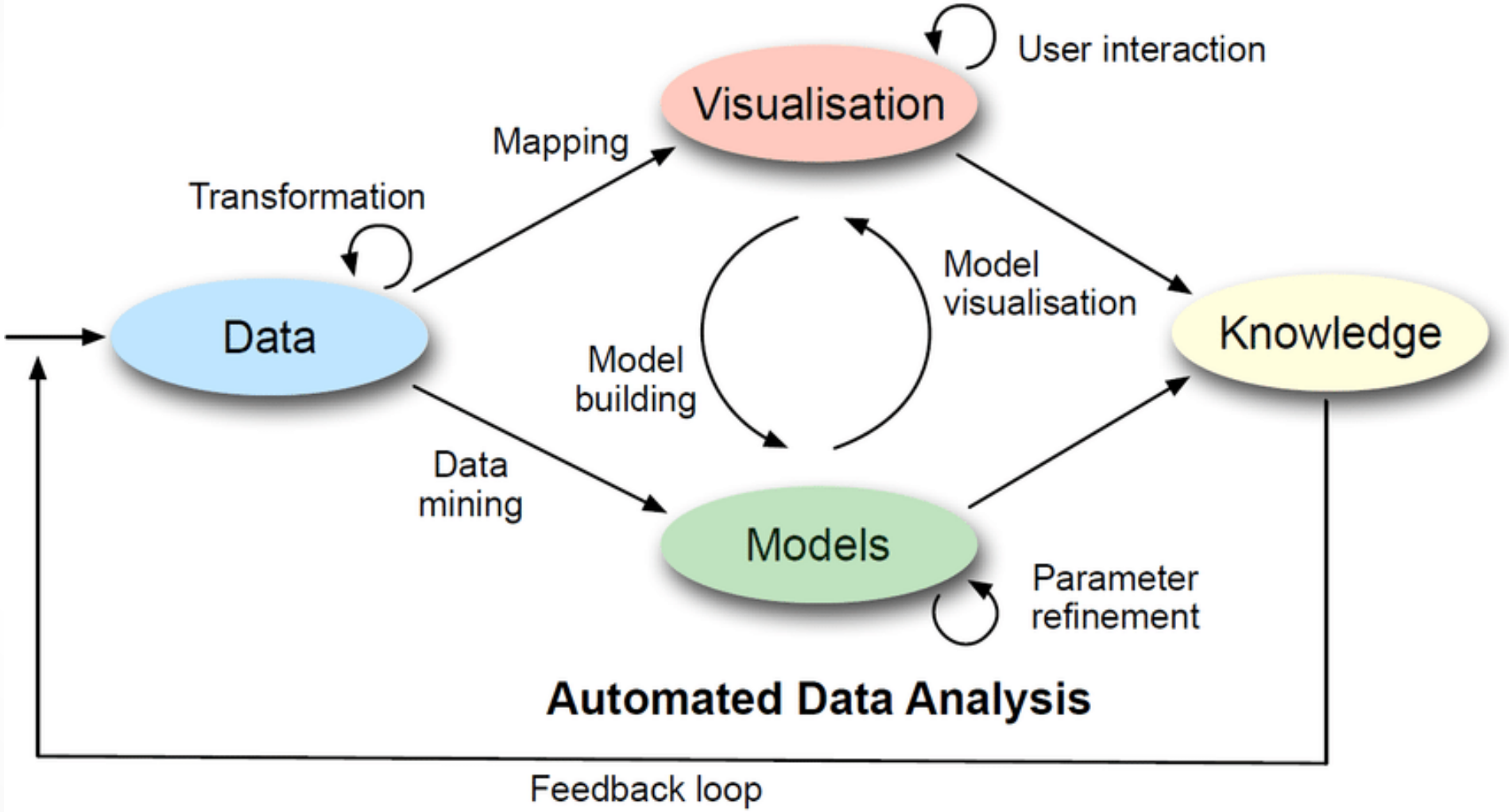
The Visualization Process

- Visualization is often part of a larger process:
  - Explorative data analysis.
  - Knowledge discovery.
  - Visual analytics.
- Visualization and analysis go, thereby, hand in hand.
- The process of starting with data and generating an image, a visualization, or a model is described as a **pipeline**.



Information Visualization Reference Model

Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, 1999.



Visual Analytics Pipleline

Interdisciplinary Field

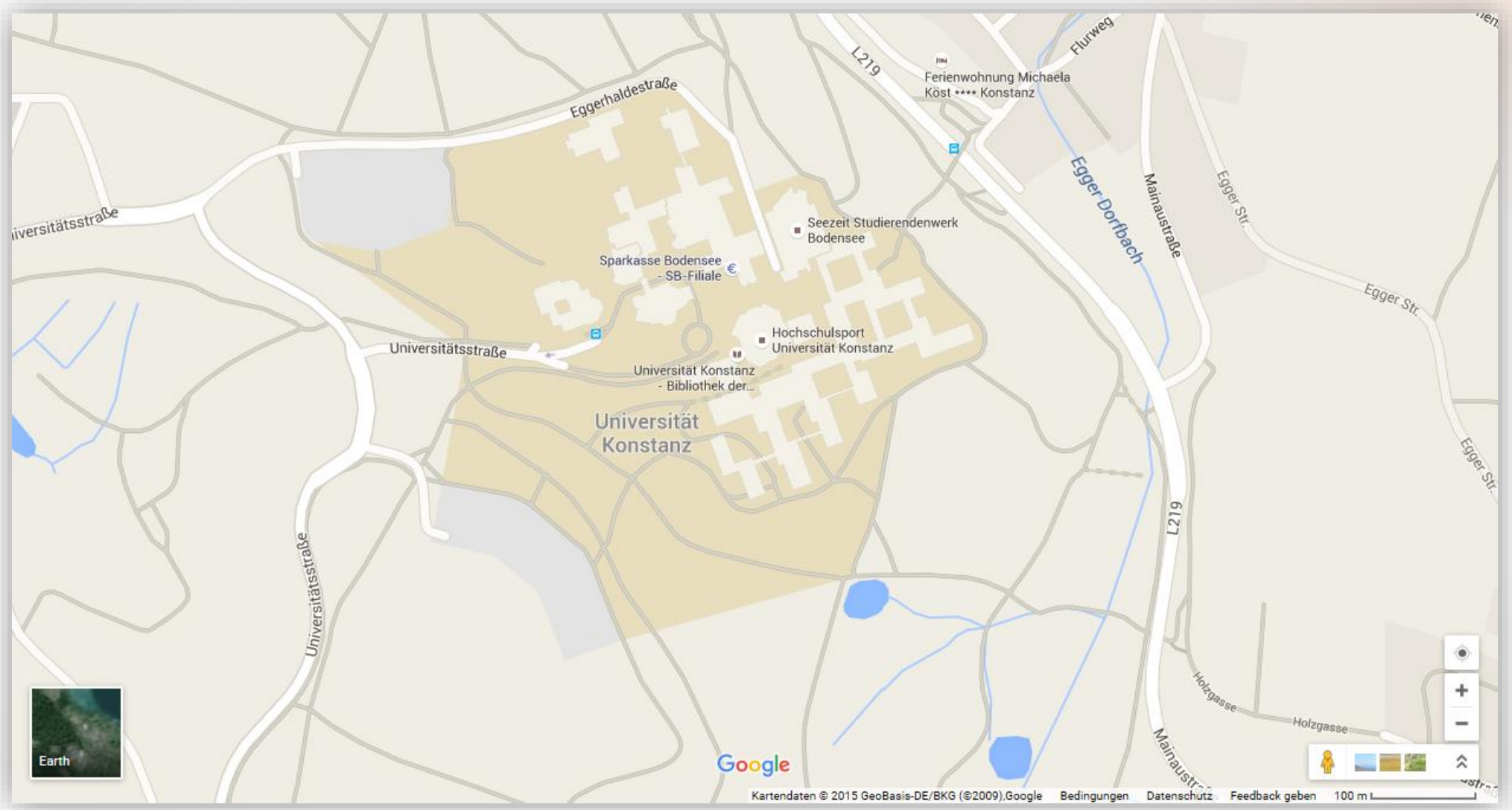
- Human-computer-interaction
- Perceptual psychology
- Databases
- Statistics
- Models
- …

Visualization in Everyday Life

- Table in newspaper.
- A train and subway map.
- A map of a region.
- A weather chart.
- A graph of stock market activities.
- A 3D image of your injured knee.
- A highway sign indicating a curve.
- …

Visualization in Everyday Life

https://www.reddit.com/r/funny/comments/1d1the/are_you_sure/



Visualization in Everyday Life

Why should we use visualizations?

Early Visualizations

Early visualizations came about out of necessity:

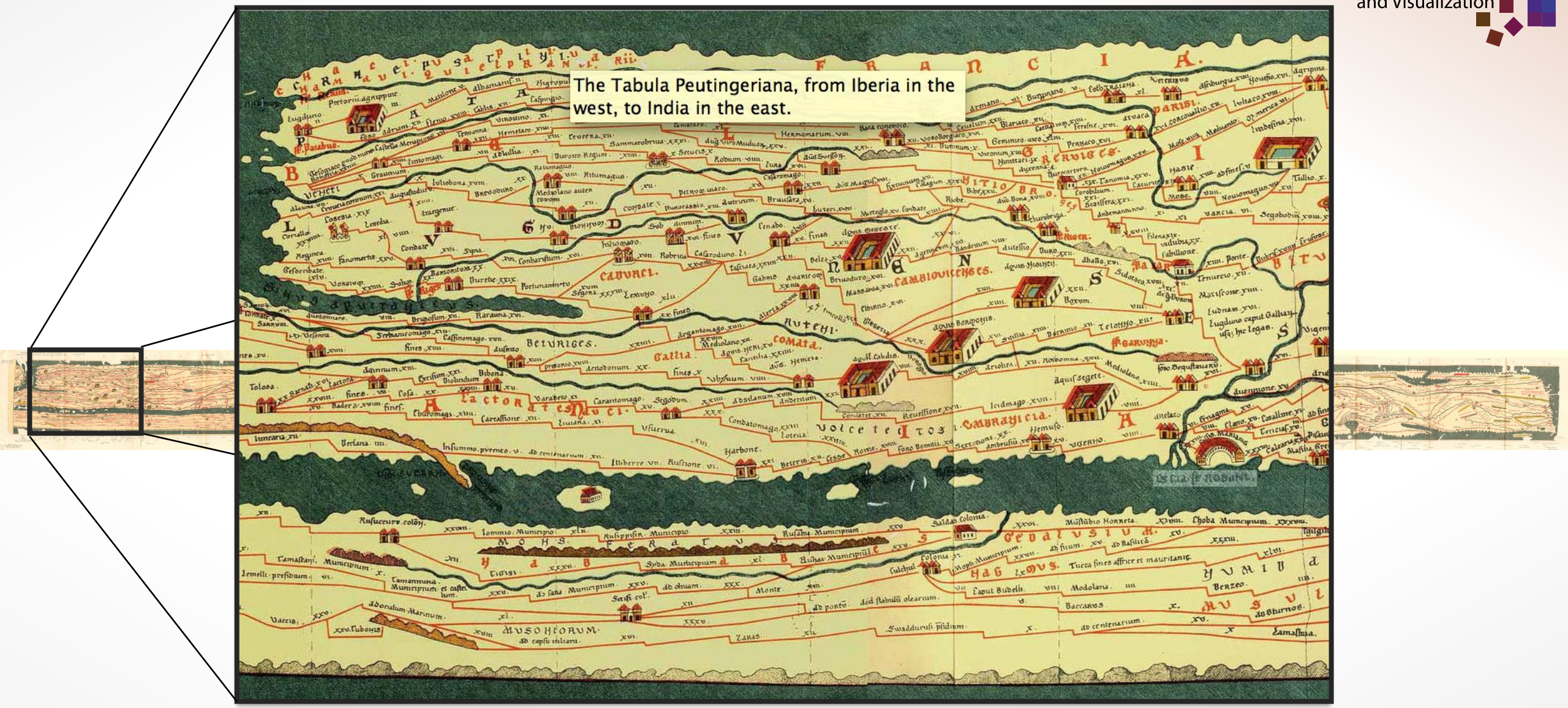
- For travel
- Commerce
- Religion
- Communication
- …

Kish Tablet - early graphical writing (3500 BC)

https://en.wikipedia.org/wiki/Kish_tablet

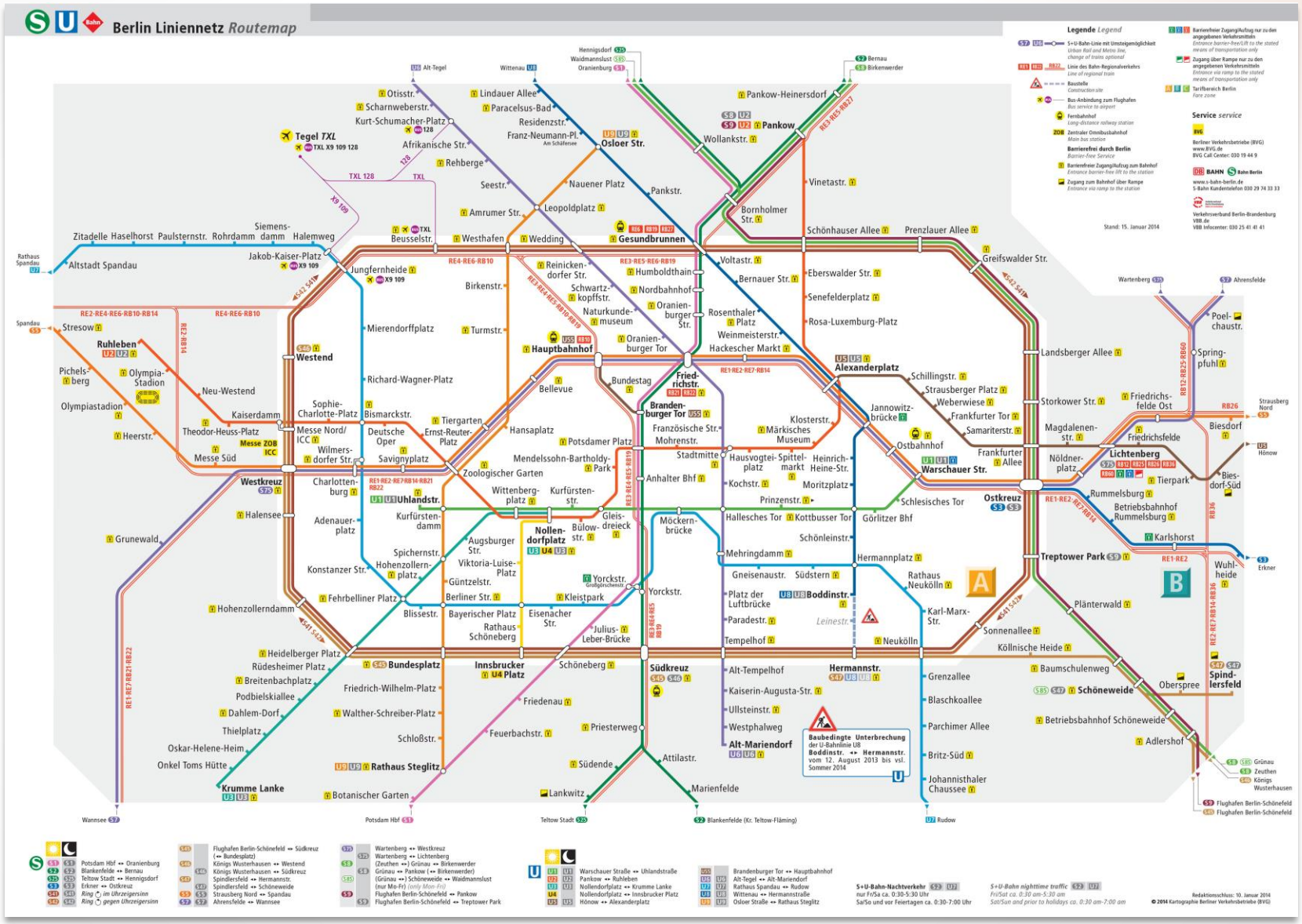
Peutinger Map - scrolls measuring approximately 34 cm high by 6,74 m (1265)

https://en.wikipedia.org/wiki/Tabula_Peutingeriana

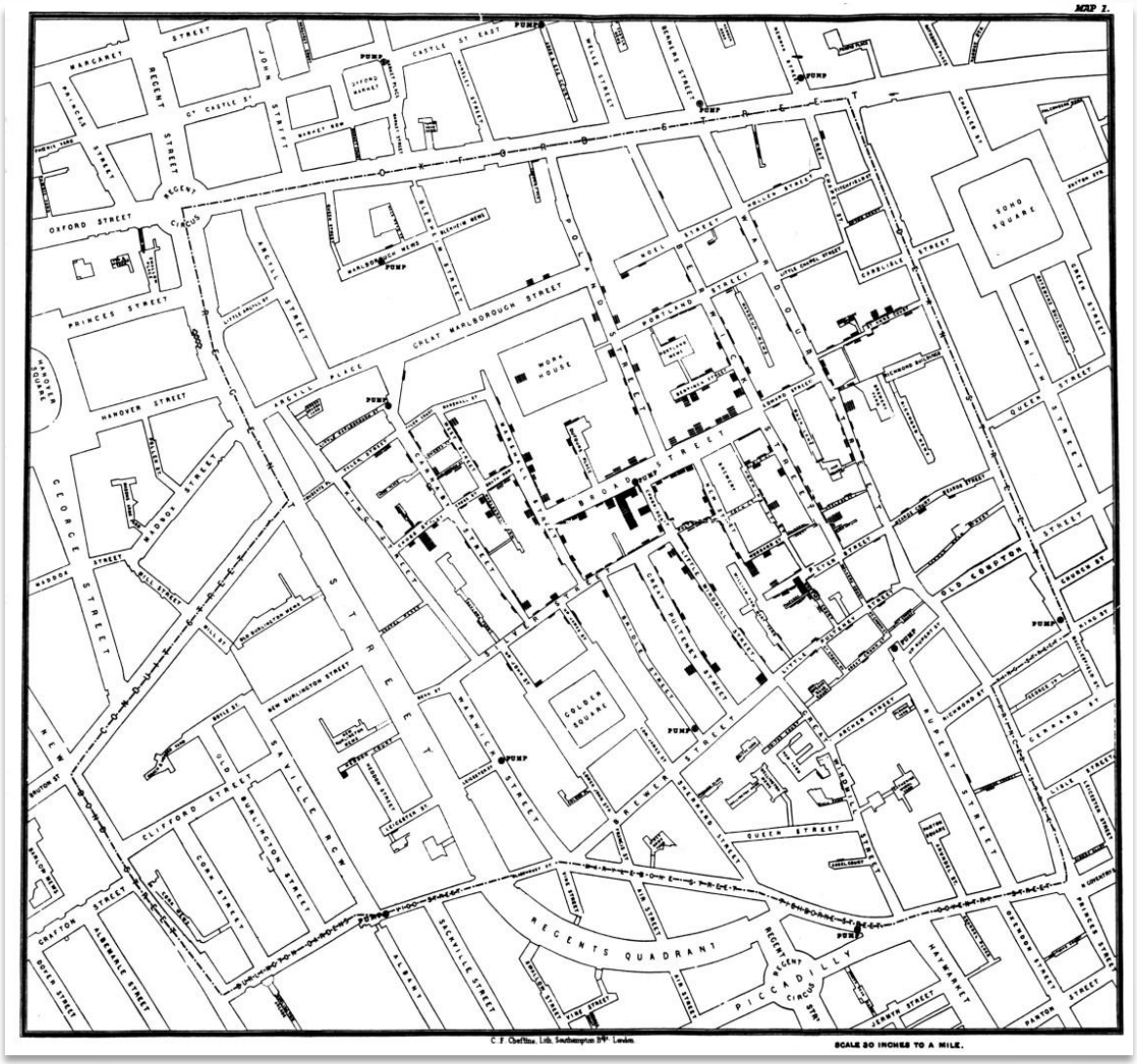


Peutinger Map - scrolls measuring approximately 34 cm high by 6,74 m (1265)

https://en.wikipedia.org/wiki/Tabula_Peutingeriana

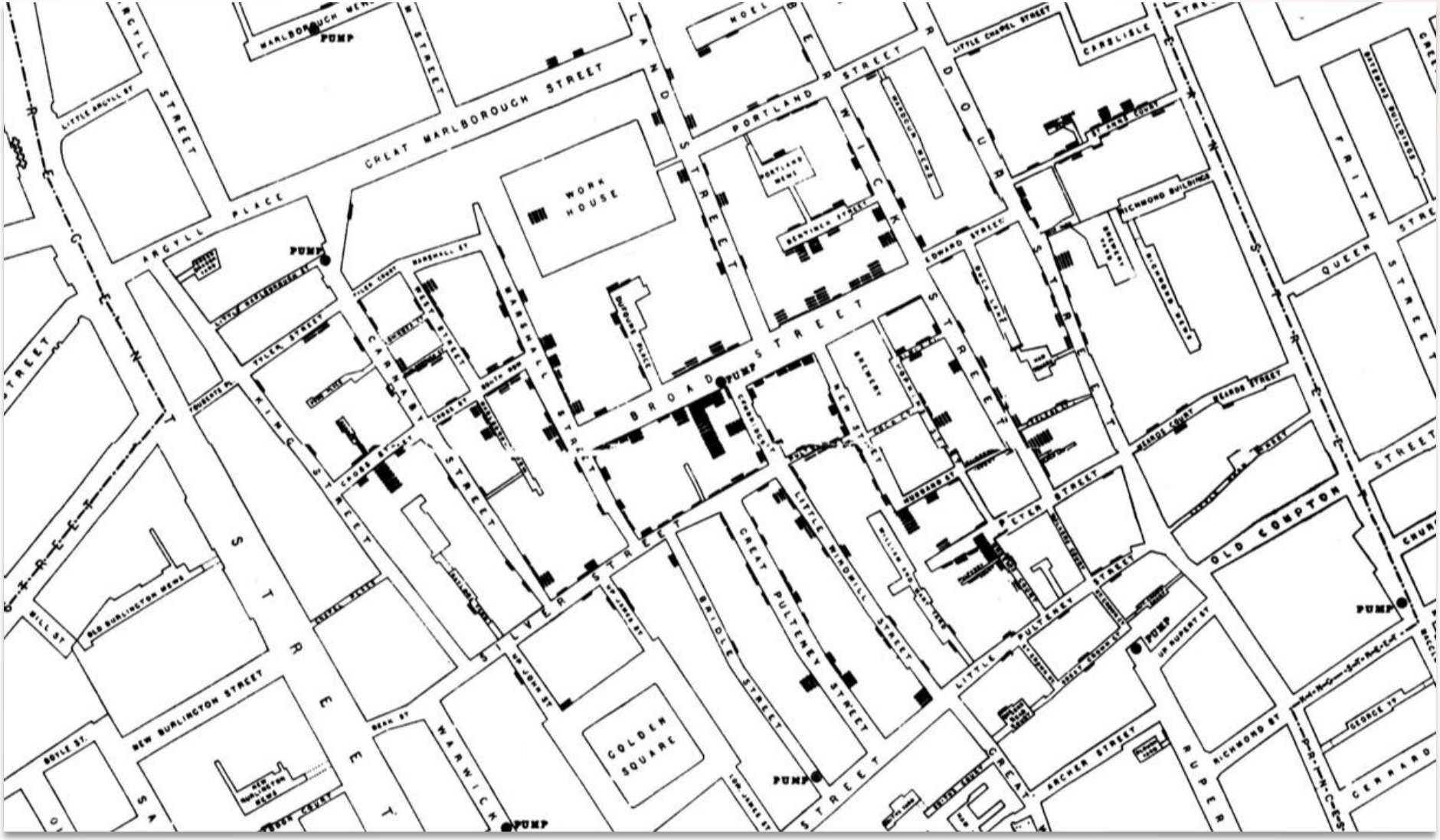


Berlin Subway (2022)



Deaths from Cholera in London (1663)

https://en.wikipedia.org/wiki/John_Snow



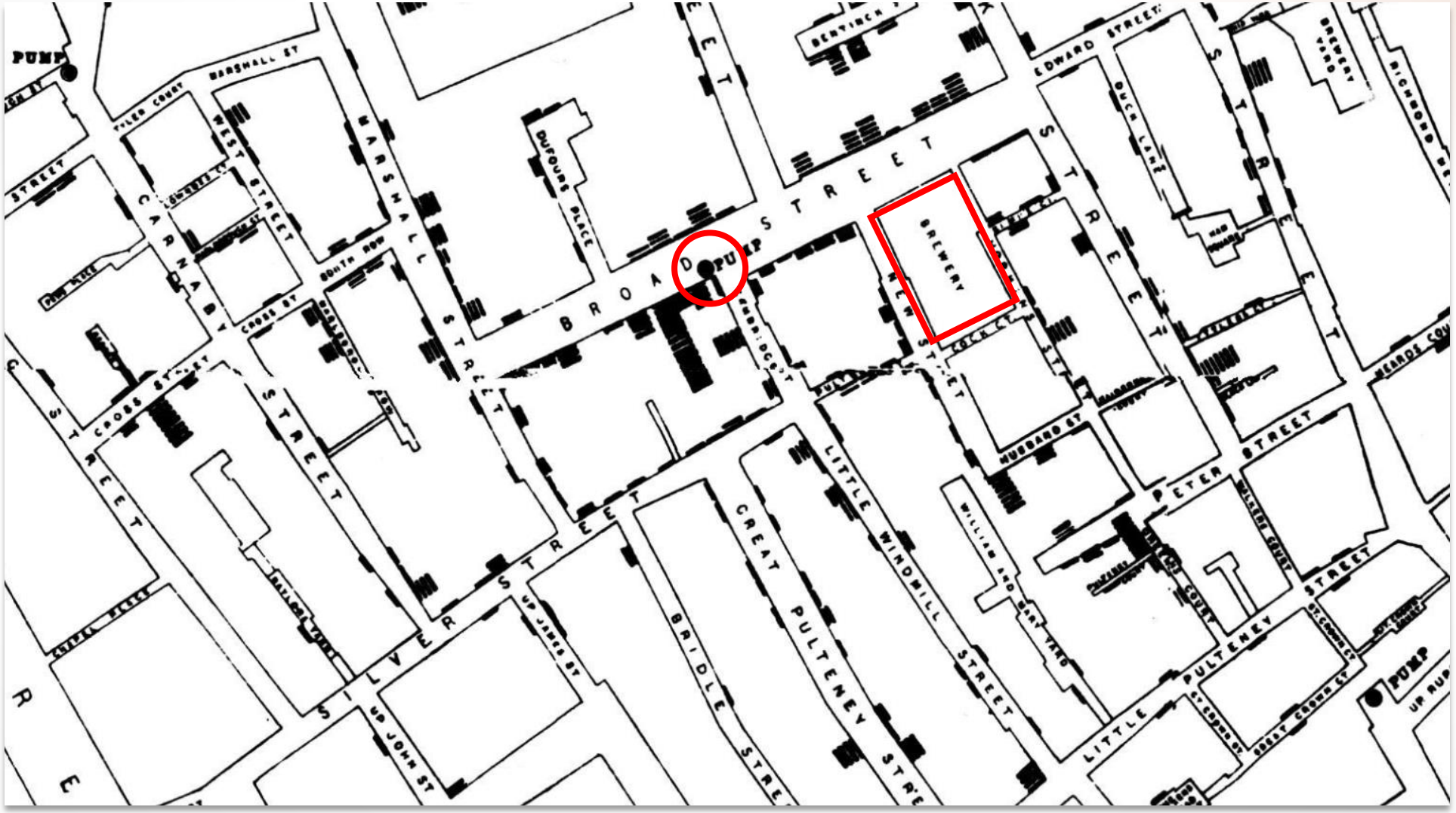
Deaths from Cholera in London (1663)

https://en.wikipedia.org/wiki/John_Snow

Deaths from Cholera in London (1663)

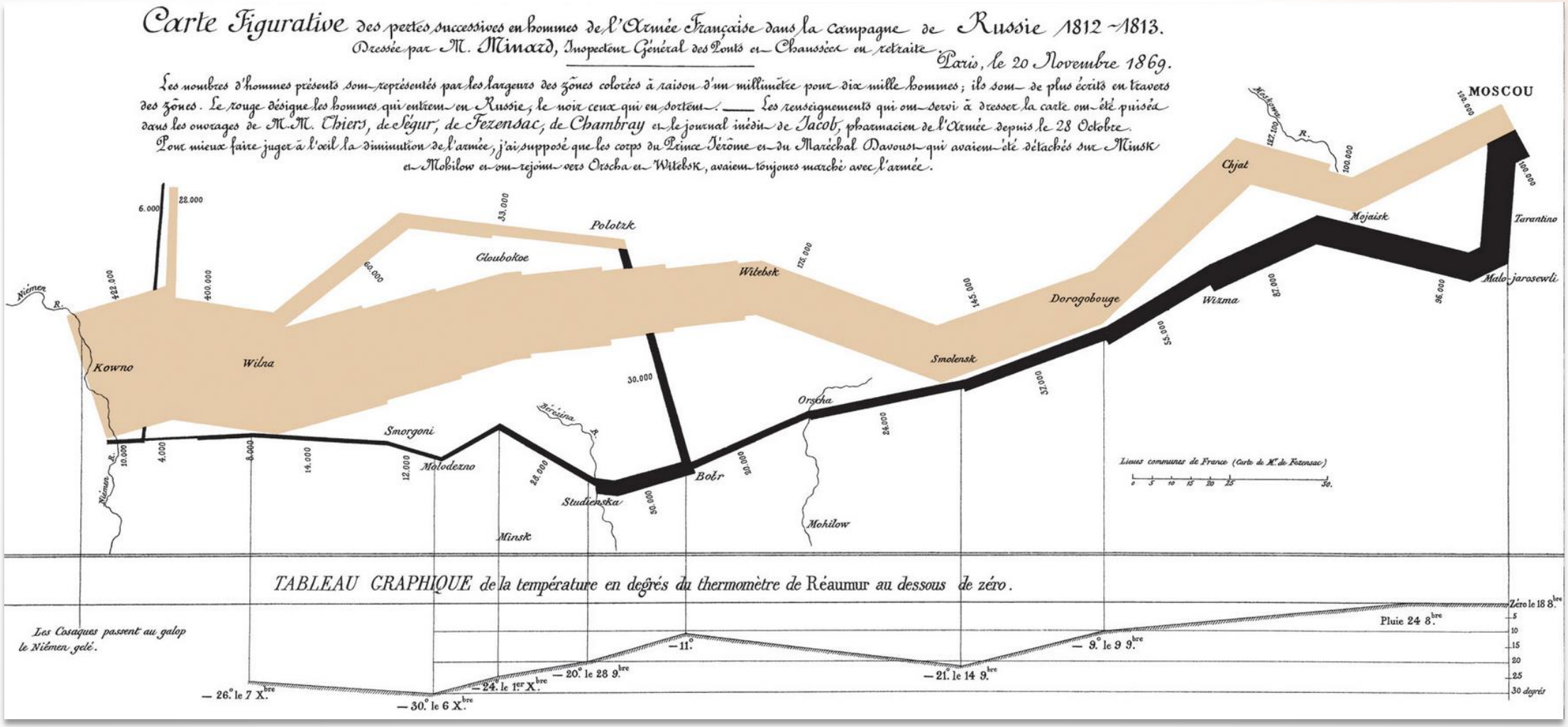
https://en.wikipedia.org/wiki/John_Snow

Deaths from Cholera in London (1663)



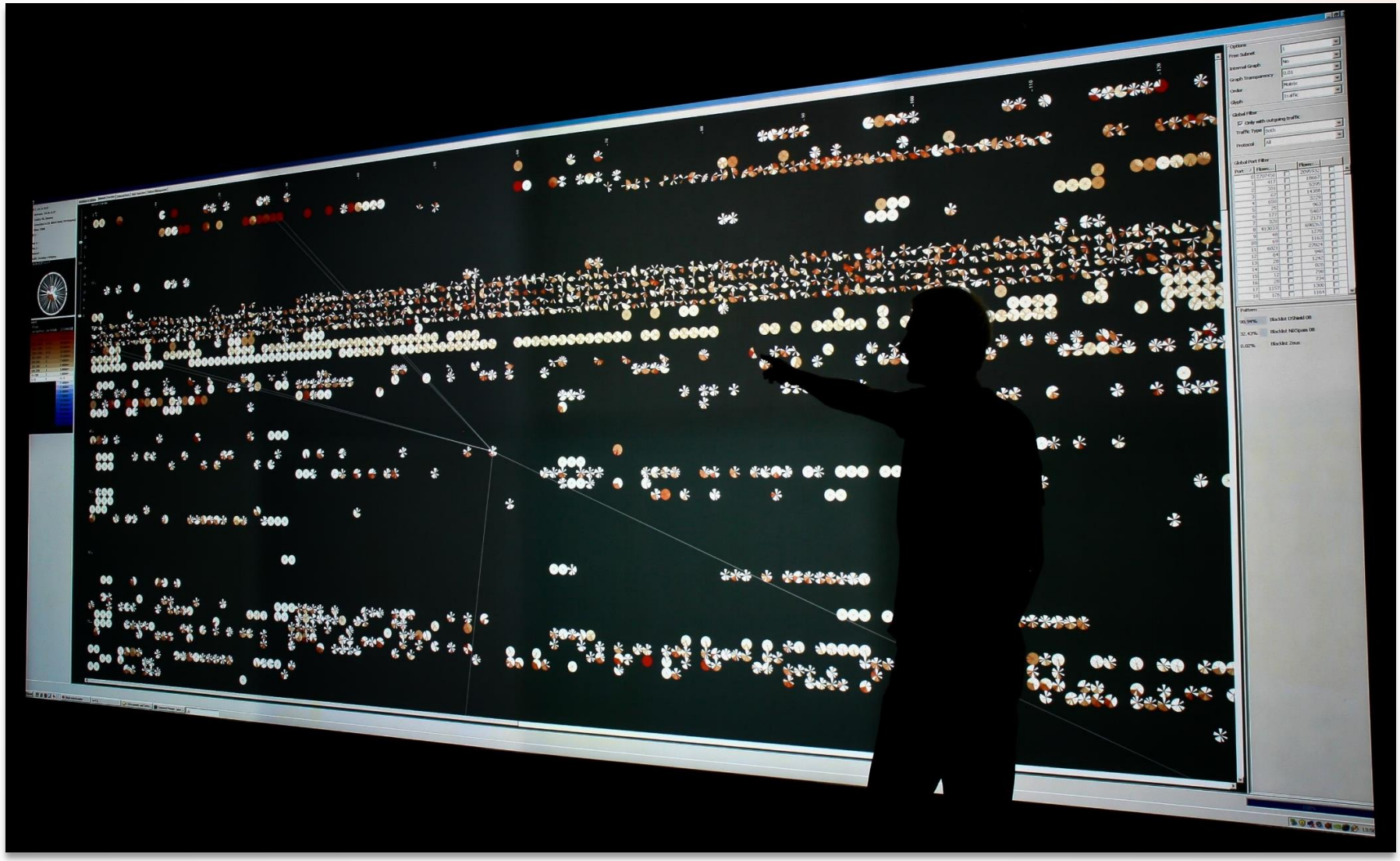
Deaths from Cholera in London (1663)

https://en.wikipedia.org/wiki/John_Snow

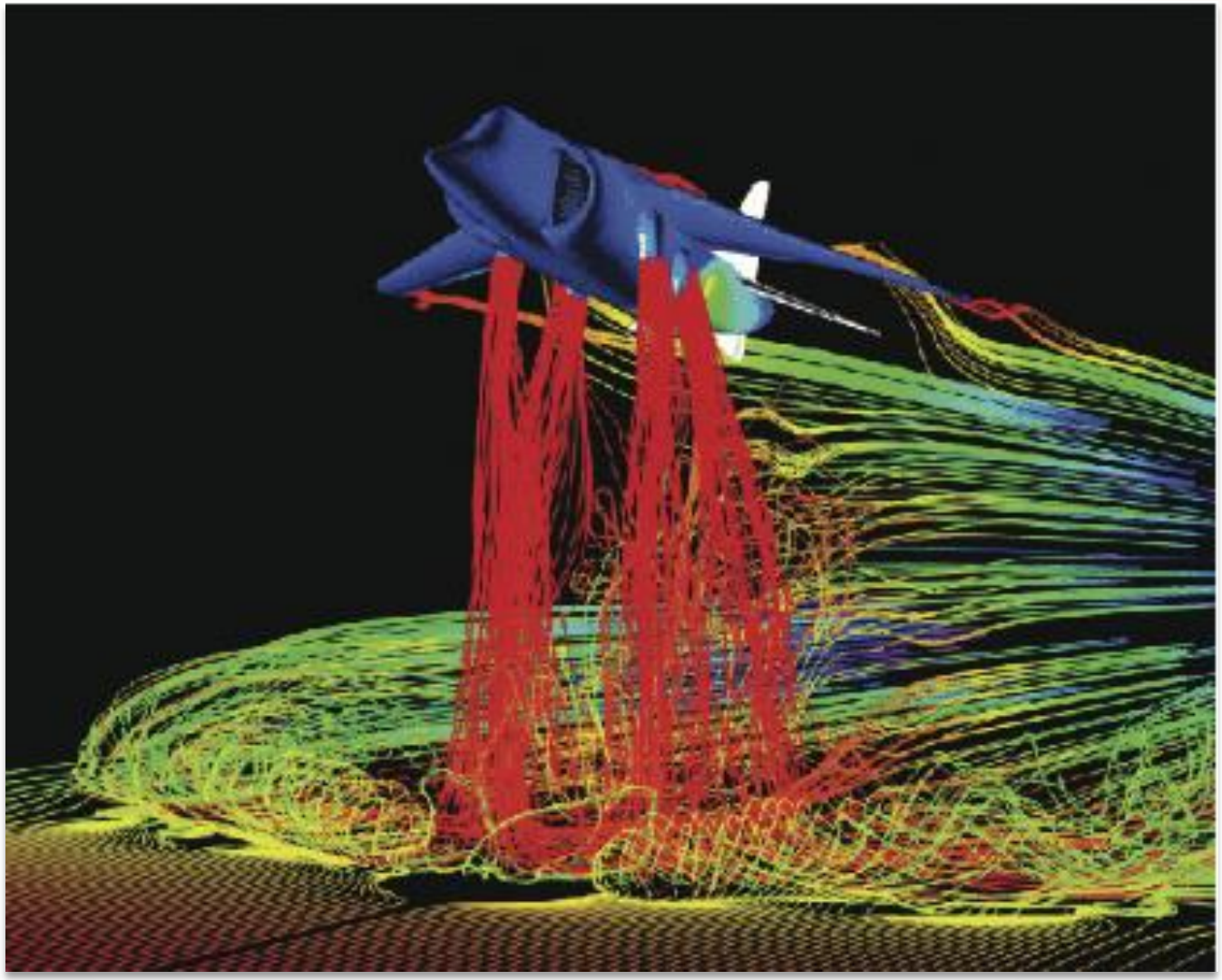


Minard's Map (1869)

https://en.wikipedia.org/wiki/Charles_Joseph_Minard



Network traffic – thousands of devices visualized on a Powerwall display (today)



Aircraft flight simulation – harrier hovering over a runway (today)

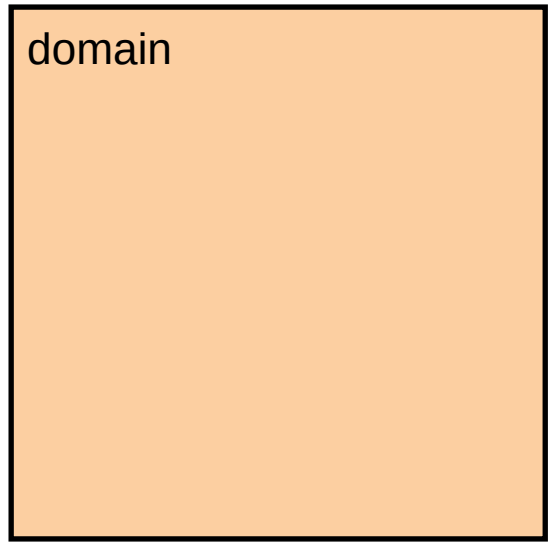
https://www.sciencephoto.com/media/350803/view/flight-simulation-of-a-harrier-jump-jet

How to design visualizations?

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: Four levels, three questions

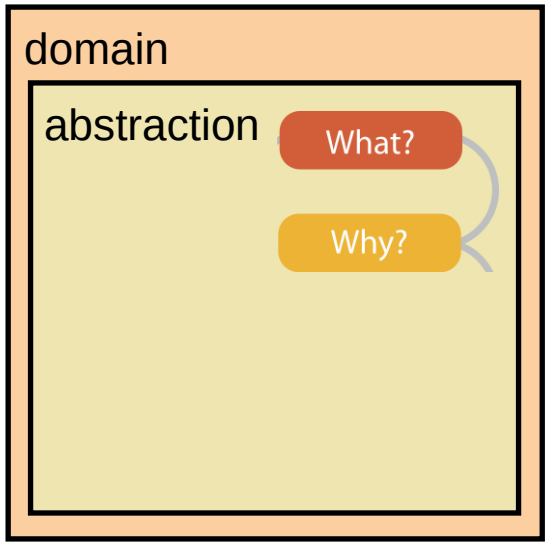
- *domain situation*
  –who are the target users?



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: Four levels, three questions

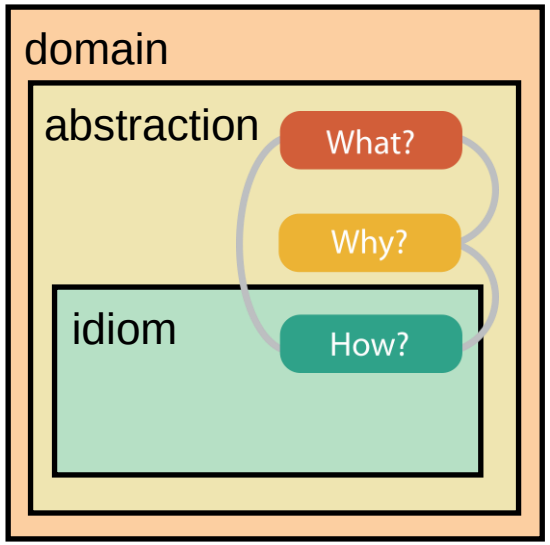
- *domain situation*
  - –who are the target users?
- *abstraction*
  - –translate from specifics of domain to vocabulary of vis
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction



[A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: Four levels, three questions

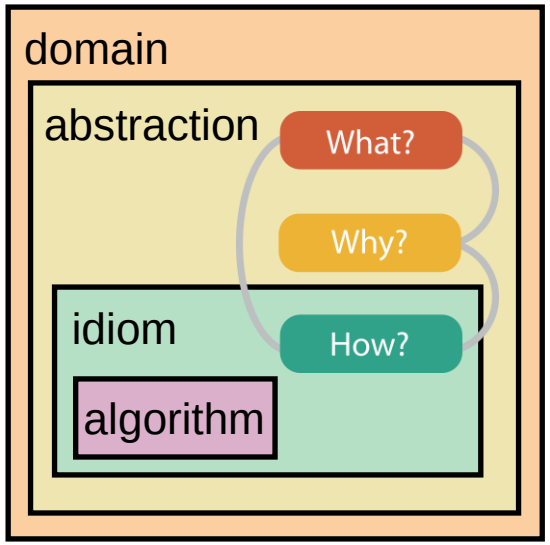
- *domain situation*
  - –who are the target users?
- *abstraction*
  - –translate from specifics of domain to vocabulary of vis
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
- *idiom*
  - –**how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate



[A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: Four levels, three questions

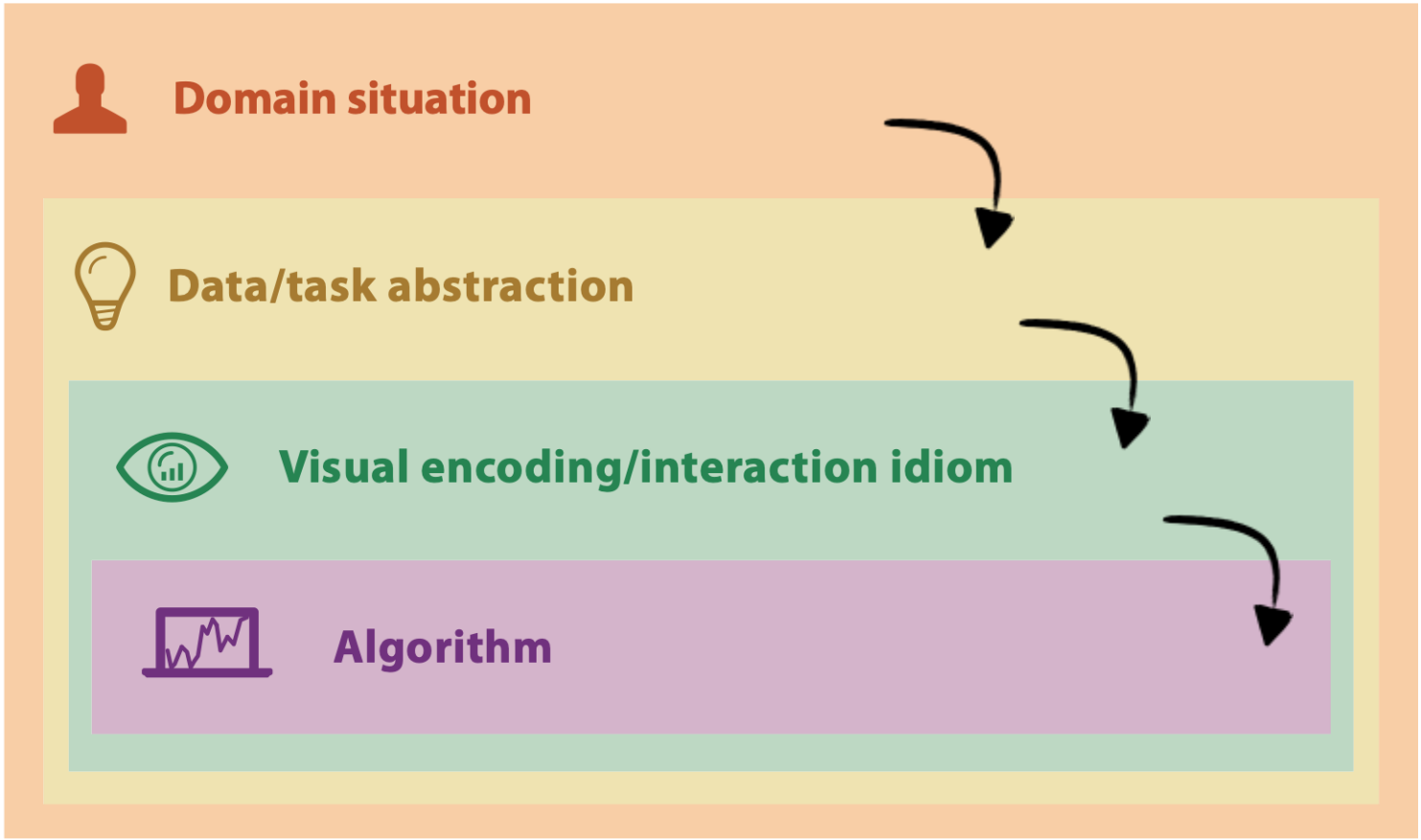
- *domain situation*
  - –who are the target users?
- *abstraction*
  - –translate from specifics of domain to vocabulary of vis
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
- *idiom*
  - –**how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate
- *algorithm*
  - –efficient computation



[A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Nested model

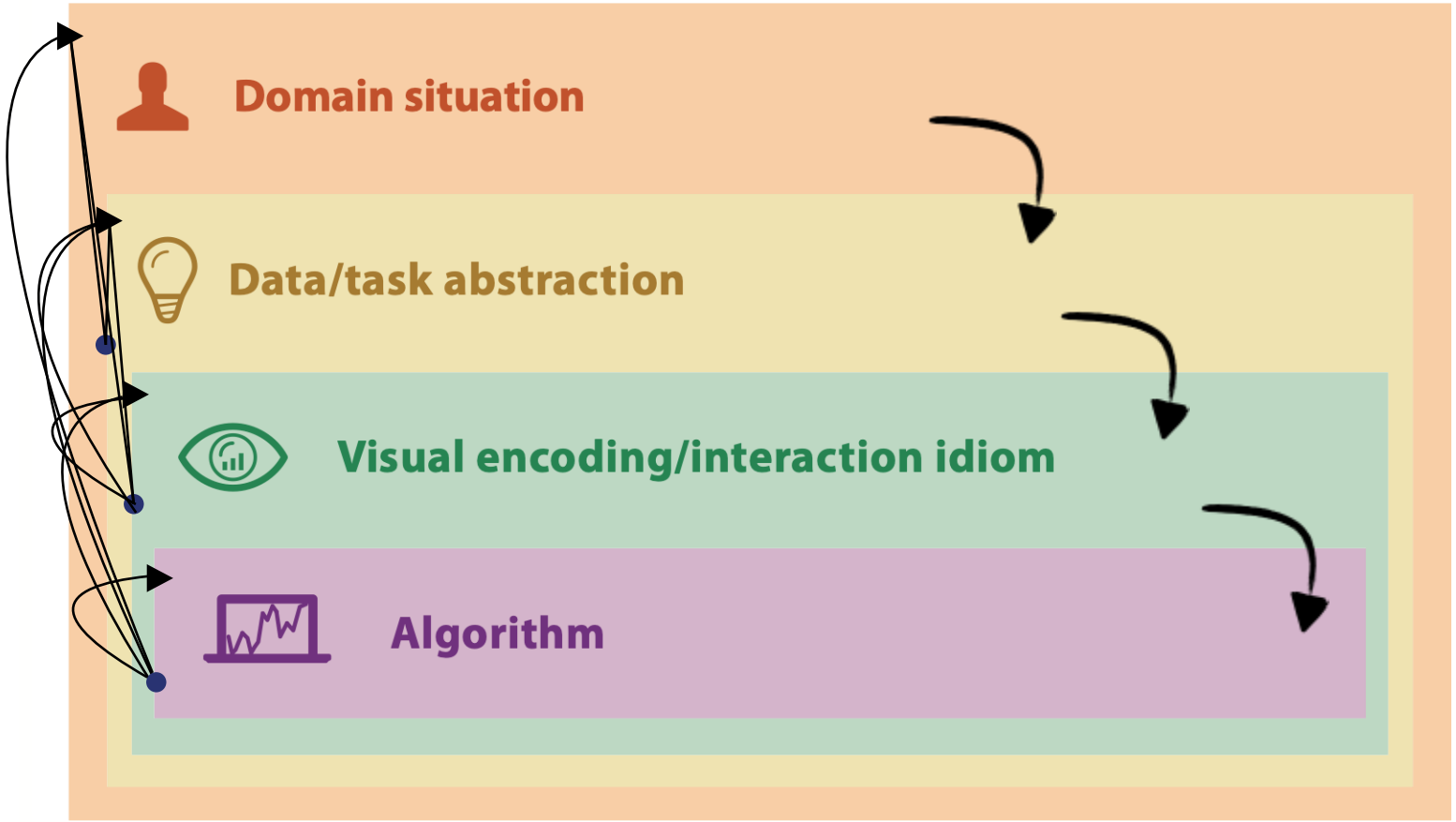
- downstream: cascading effects



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Nested model

- downstream: cascading effects
- upstream: iterative refinement



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

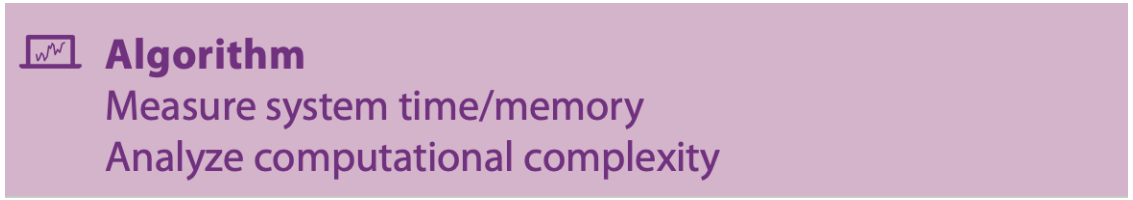
Why is validation difficult?

- different ways to get it wrong at each level

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- solution: use methods from different fields at each level

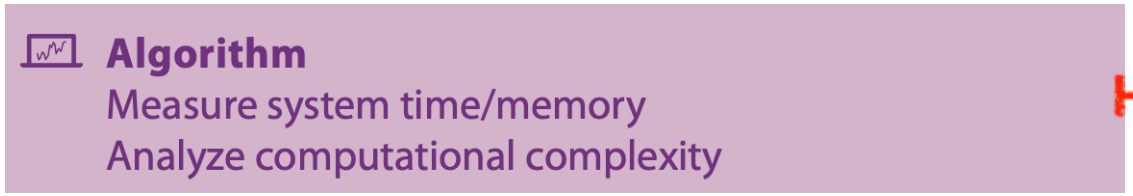


[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- solution: use methods from different fields at each level

computer science

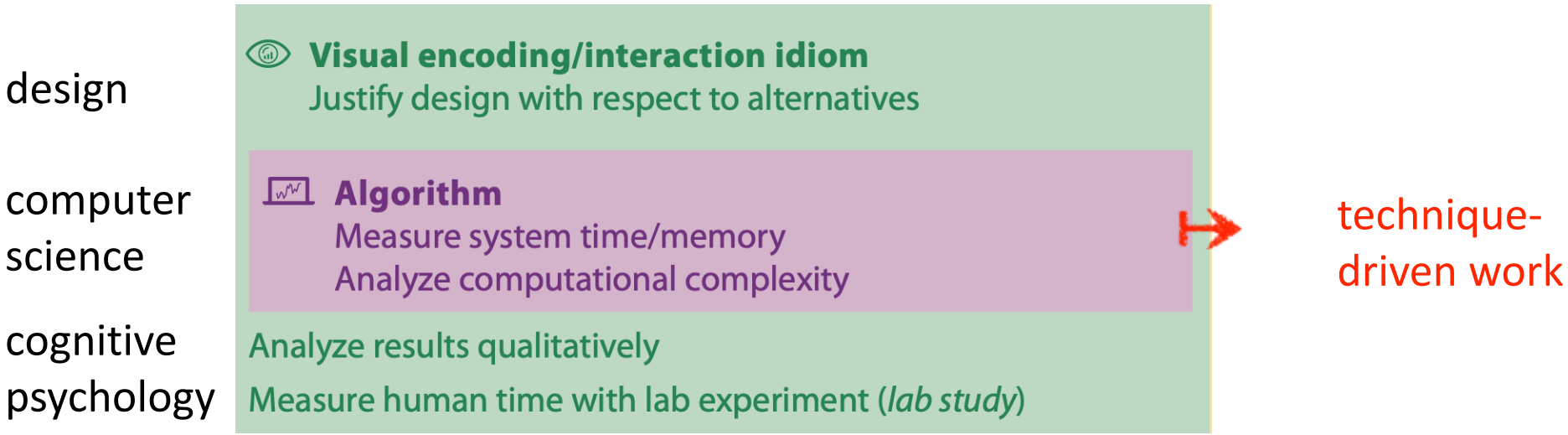


technique-driven work

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- solution: use methods from different fields at each level



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Why is validation difficult?

- solution: use methods from different fields at each level

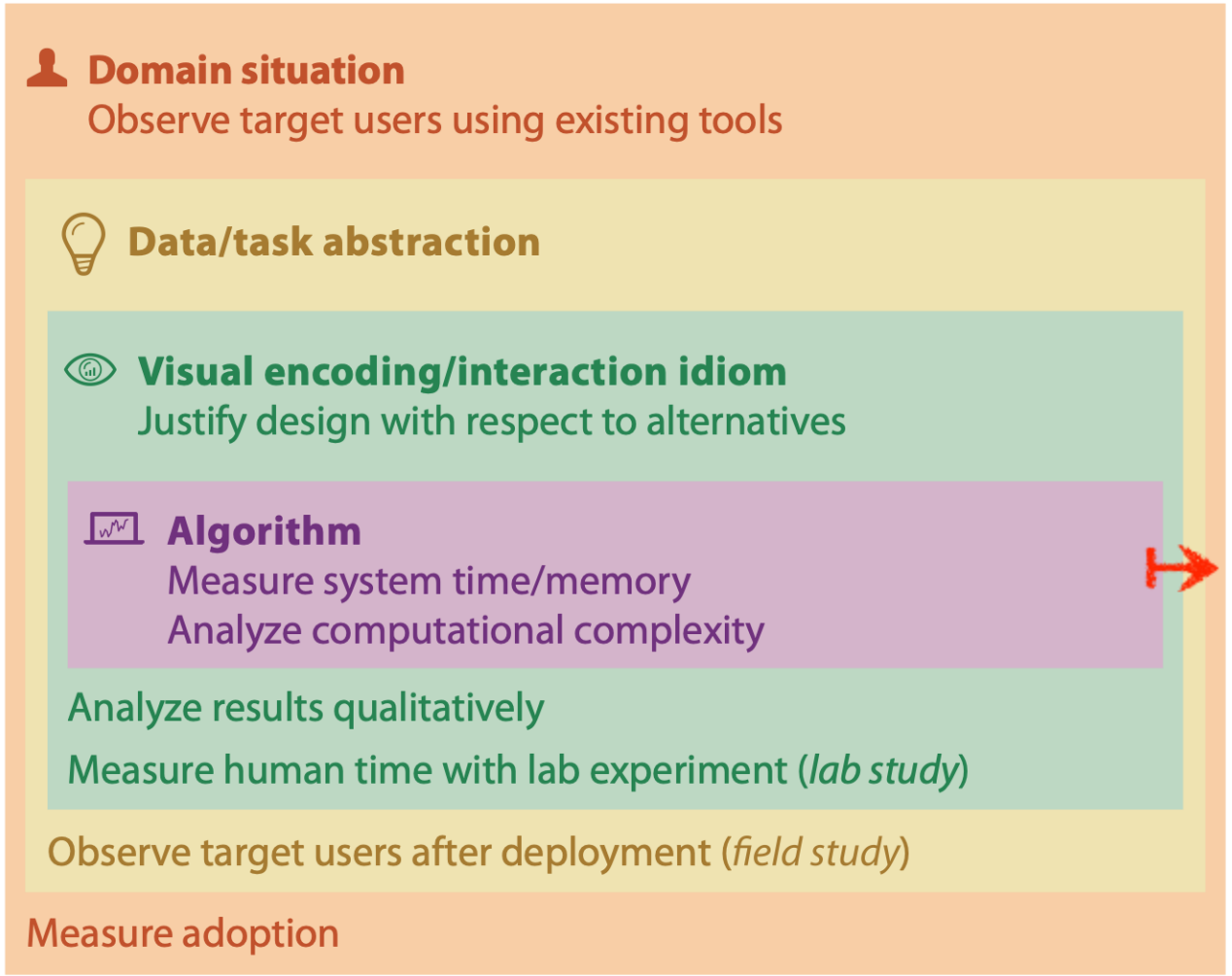
anthropology/ ethnography

design

computer science

cognitive psychology

anthropology/ ethnography

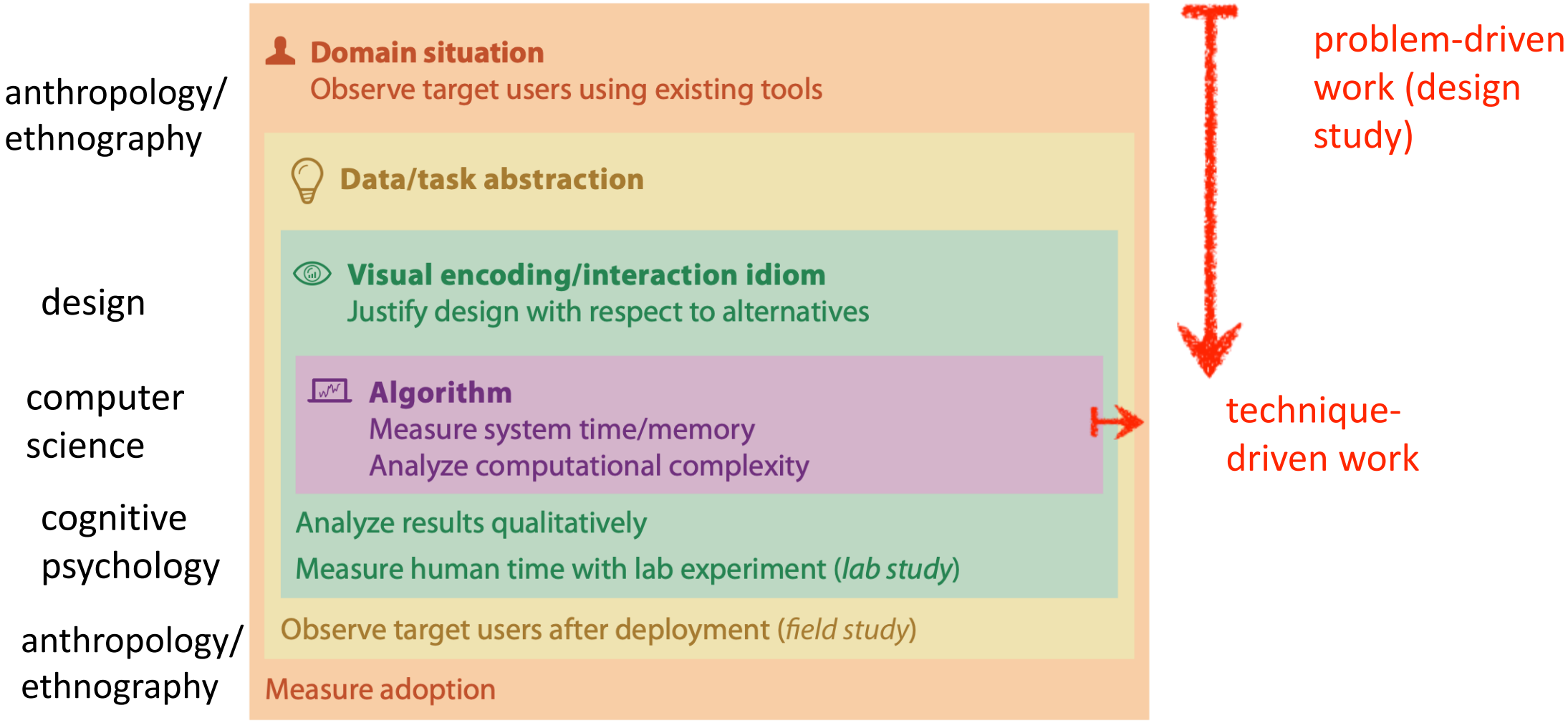


technique-driven work

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

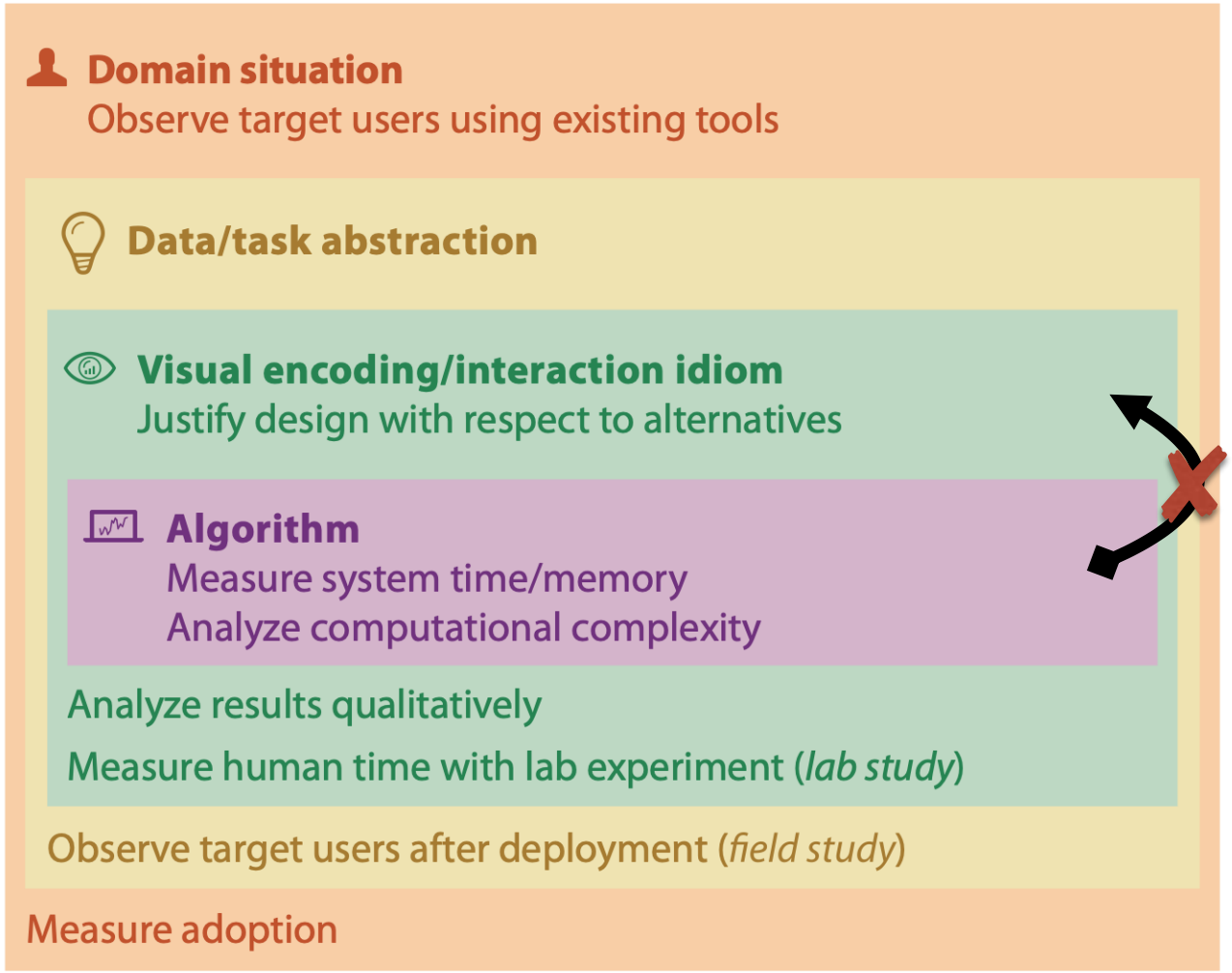
Why is validation difficult?

- solution: use methods from different fields at each level



[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

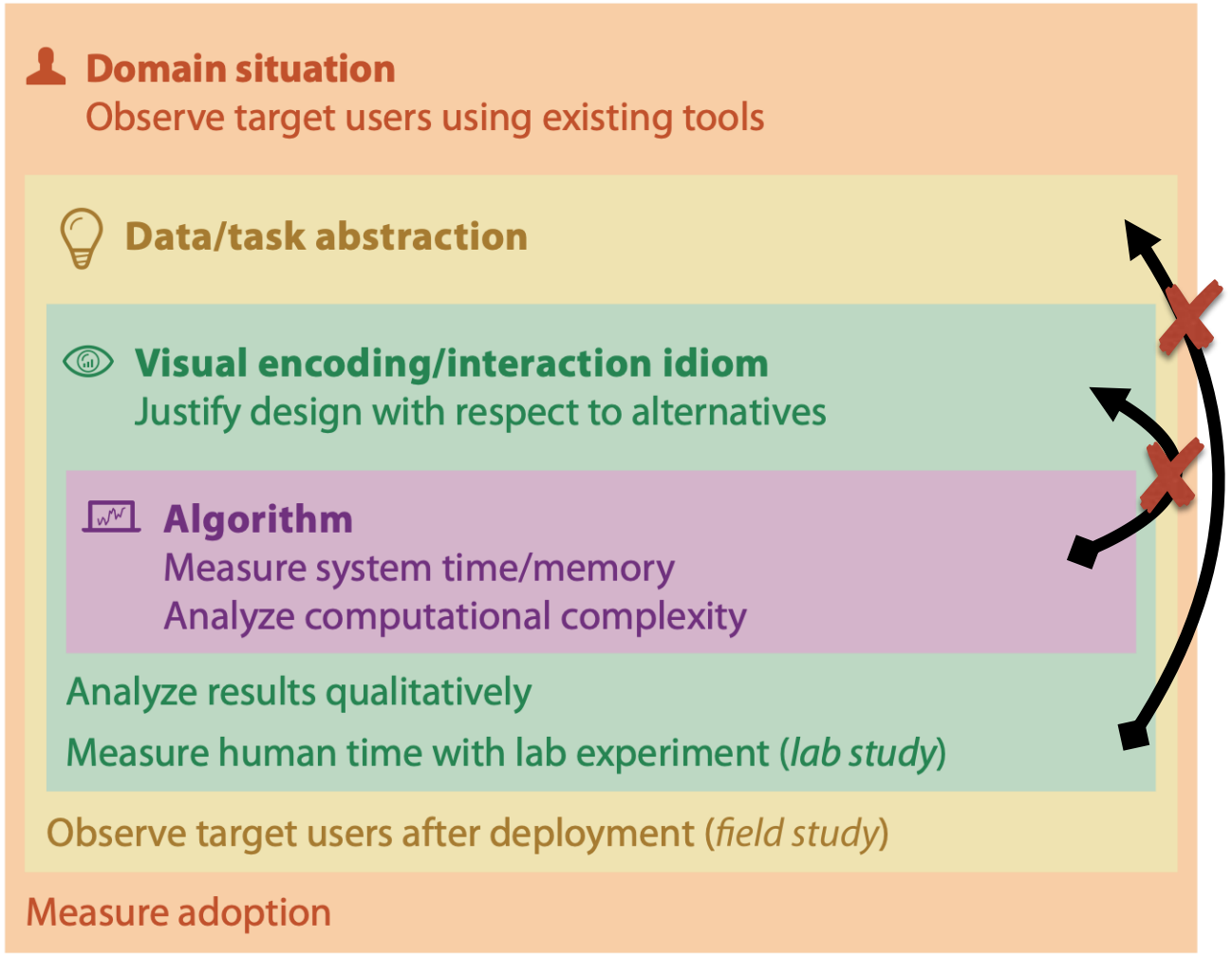
Avoid mismatches



computational benchmarks do not confirm idiom design

[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Avoid mismatches

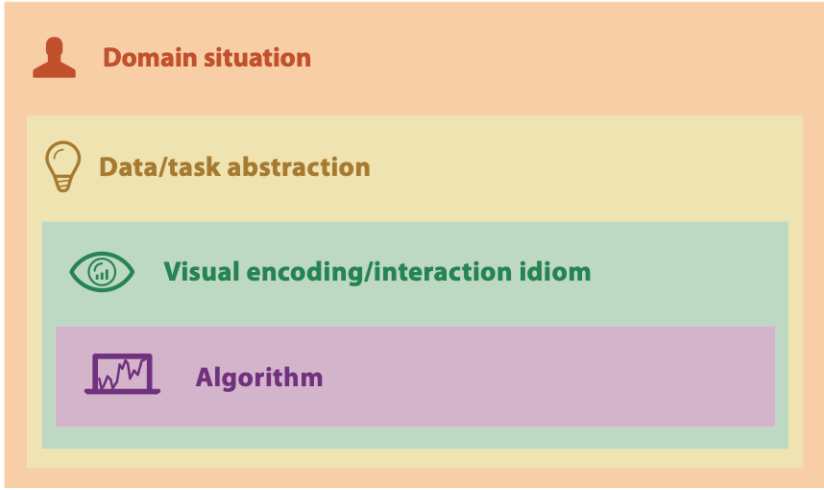


lab studies do not confirm task abstraction

computational benchmarks do not confirm idiom design

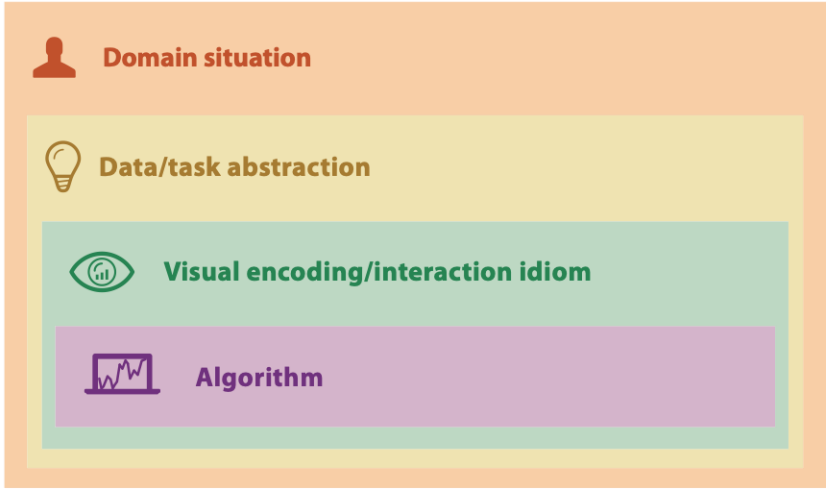
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

From domain to abstraction



From domain to abstraction

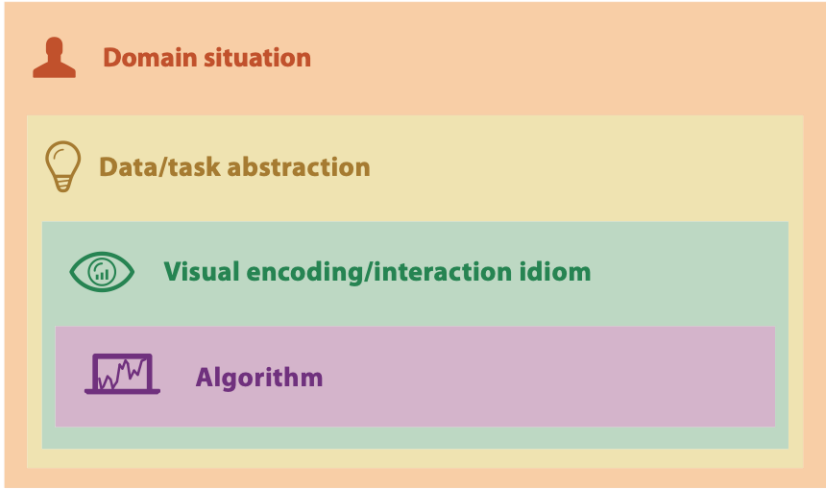
- domain characterization: details of application domain

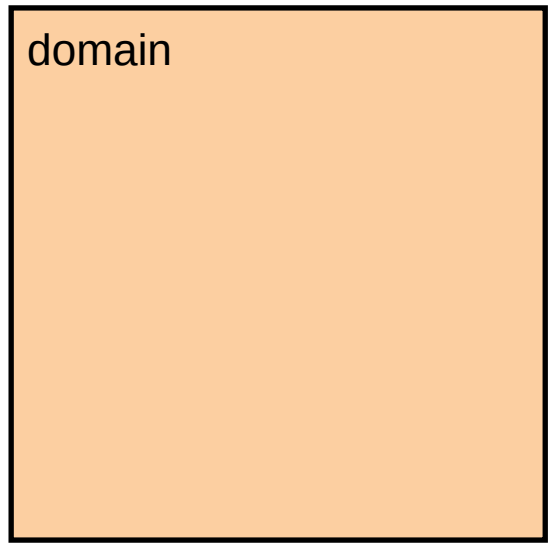


domain

From domain to abstraction

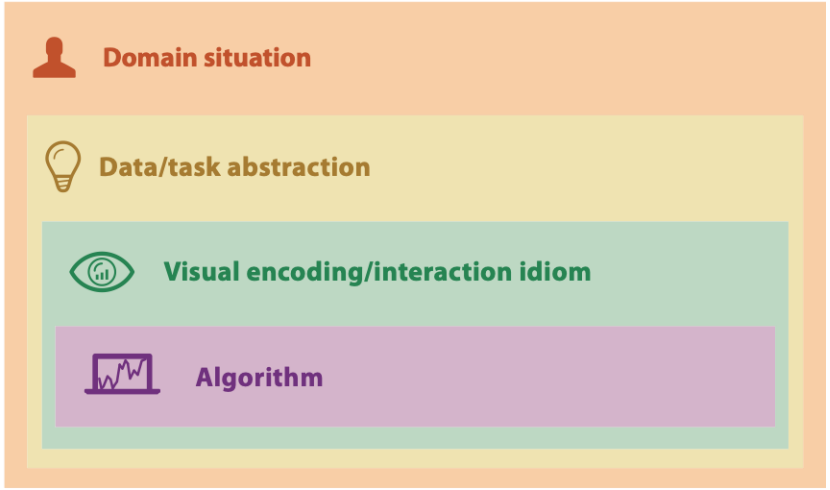
- domain characterization:
  details of application domain
  - group of users, target domain, their questions & data
    - varies wildly by domain
    - must be specific enough to get traction

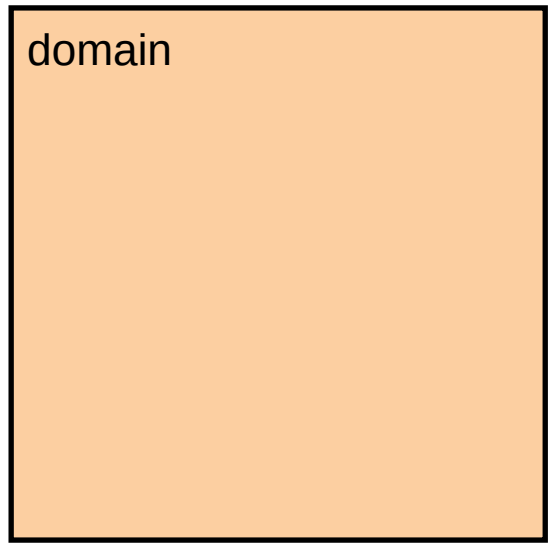




From domain to abstraction

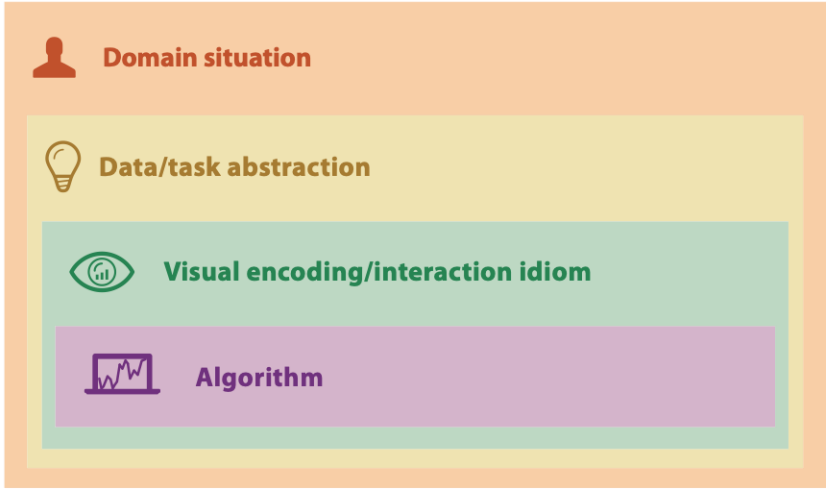
- domain characterization:
  details of application domain
  - group of users, target domain, their questions & data
    - varies wildly by domain
    - must be specific enough to get traction
  - domain questions/problems
    - break down into simpler abstract tasks

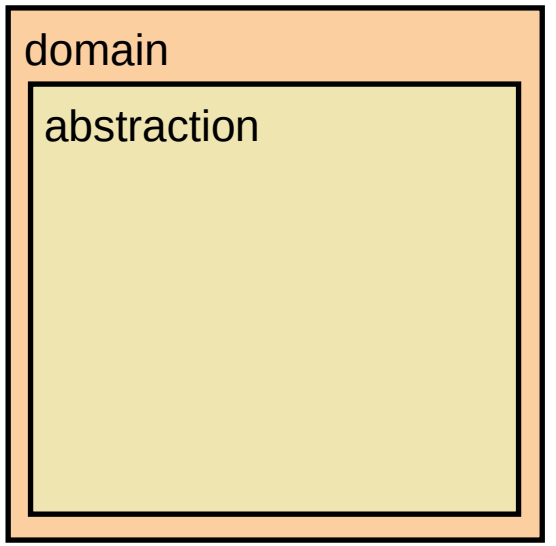




From domain to abstraction

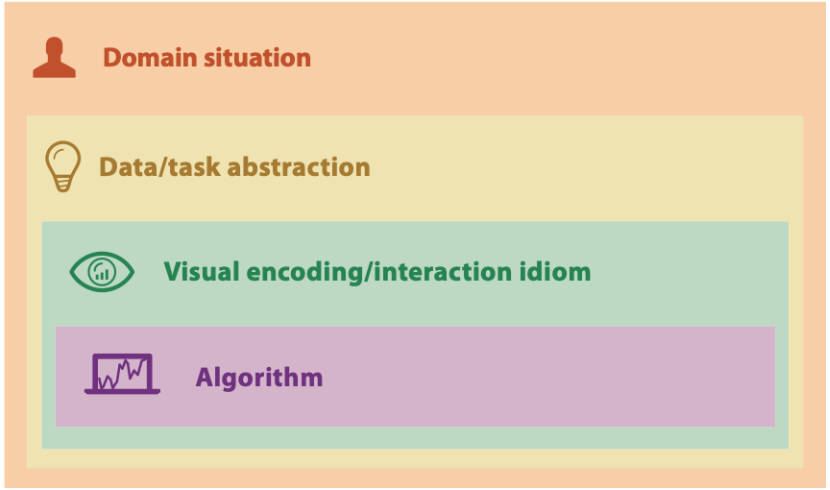
- domain characterization:
  details of application domain
  - group of users, target domain, their questions & data
    - varies wildly by domain
    - must be specific enough to get traction
  - domain questions/problems
    - break down into simpler abstract tasks
- abstraction: data & task
  - map *what* and *why* into generalized terms

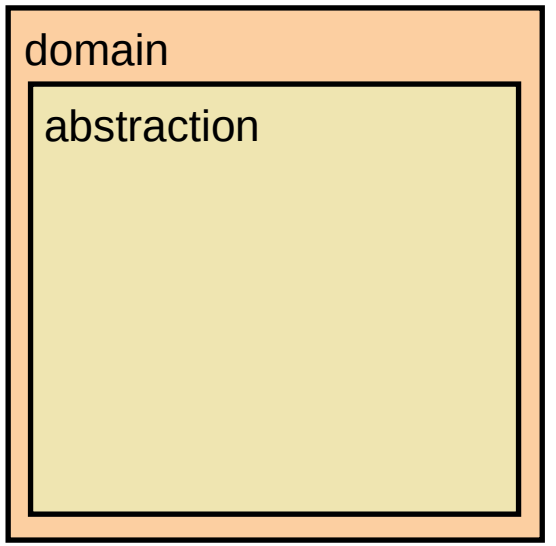




From domain to abstraction

- domain characterization:
  details of application domain
  - group of users, target domain, their questions & data
    - varies wildly by domain
    - must be specific enough to get traction
  - domain questions/problems
    - break down into simpler abstract tasks
- abstraction: data & task
  - map *what* and *why* into generalized terms
    - identify tasks that users wish to perform, or already do
    - find data types that will support those tasks
      - possibly transform /derive if need be





Task abstraction: Actions and targets

- very high-level pattern

- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Task abstraction: Actions and targets

- very high-level pattern

- actions
  - analyze
    - high-level choices
  - search
    - find a known/unknown item
  - query
    - find out about characteristics of item

- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Task abstraction: Actions and targets

- very high-level pattern

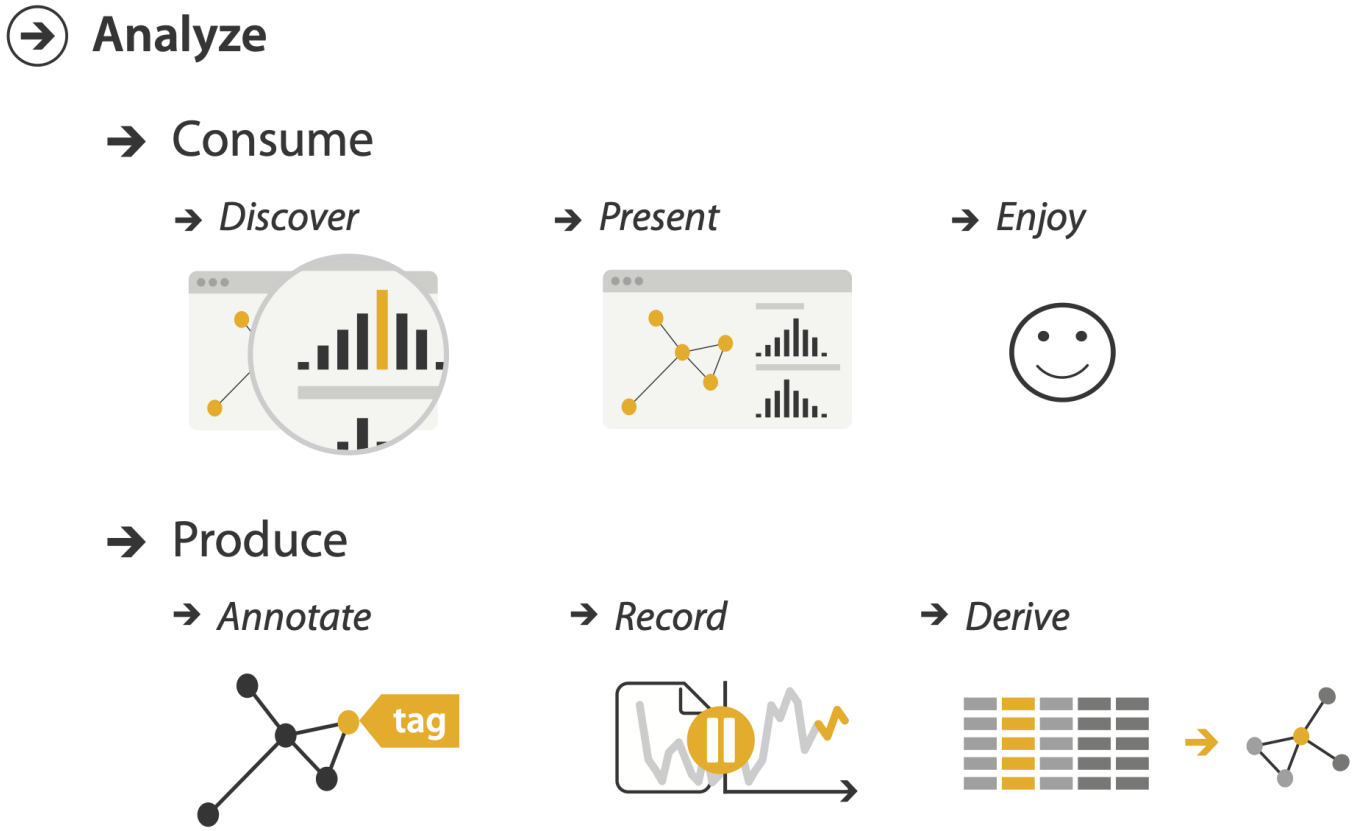
- actions
  - analyze
    - high-level choices
  - search
    - find a known/unknown item
  - query
    - find out about characteristics of item

- targets
  - what is being acted on

- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*

Actions: Analyze

- consume
  - discover vs present
    - classic split
    - aka explore vs explain
  - enjoy
    - newcomer
    - aka casual, social
- produce
  - annotate, record
  - derive
    - crucial design choice



Actions: Search

Actions: Search

- what does user know?
  - –target, location

Search

	Target known	Target unknown
Location known	 *Lookup*	 *Browse*
Location unknown	 *Locate*	 *Explore*

Actions: Search

- what does user know?
  - –target, location
- lookup
  - –ex: word in dictionary
    - alphabetical order

Search

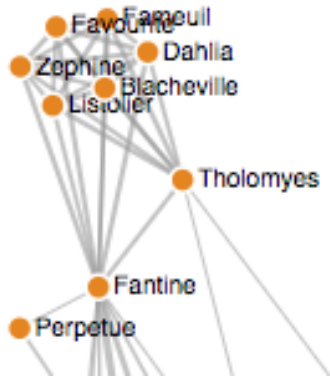
	Target known	Target unknown
Location known	 *Lookup*	 *Browse*
Location unknown	 *Locate*	 *Explore*

Actions: Search

- what does user know?
  - –target, location
- lookup
  - –ex: word in dictionary
    - alphabetical order
- locate
  - –ex: keys in your house
  - –ex: node in network

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*



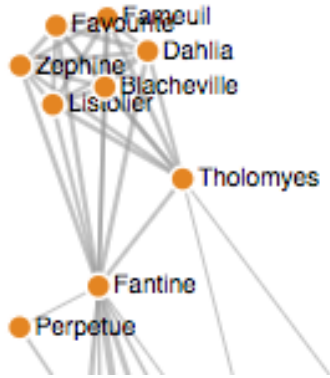
https://bl.ocks.org/heybignick/3faf257bbbbc7743bb72310d03b86ee8

Actions: Search

- what does user know?
  - –target, location
- lookup
  - –ex: word in dictionary
    - alphabetical order
- locate
  - –ex: keys in your house
  - –ex: node in network
- browse
  - –ex: books in bookstore

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*



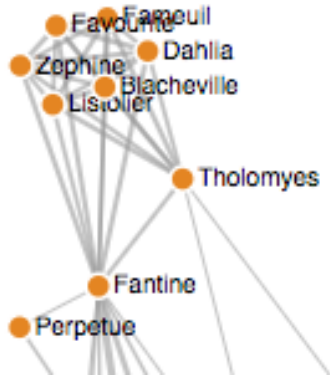
https://bl.ocks.org/heybignick/3faf257bbbbc7743bb72310d03b86ee8

Actions: Search

- what does user know?
  - –target, location
- lookup
  - –ex: word in dictionary
    - alphabetical order
- locate
  - –ex: keys in your house
  - –ex: node in network
- browse
  - –ex: books in bookstore
- explore
  - –ex: find cool neighborhood in new city

Search

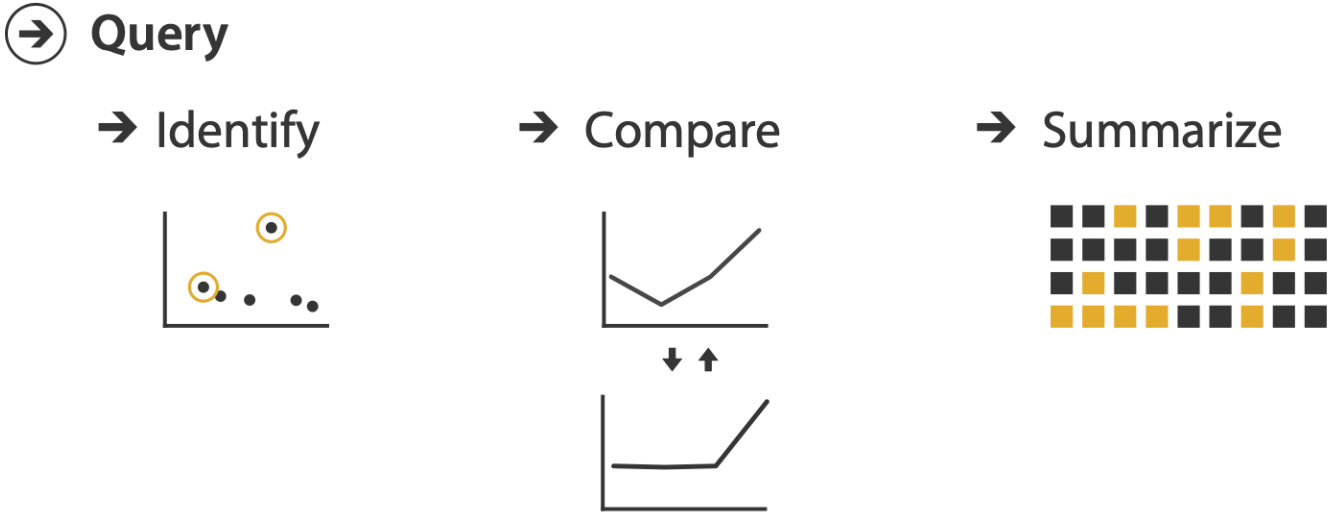
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*



https://bl.ocks.org/heybignick/3faf257bbbbc7743bb72310d03b86ee8

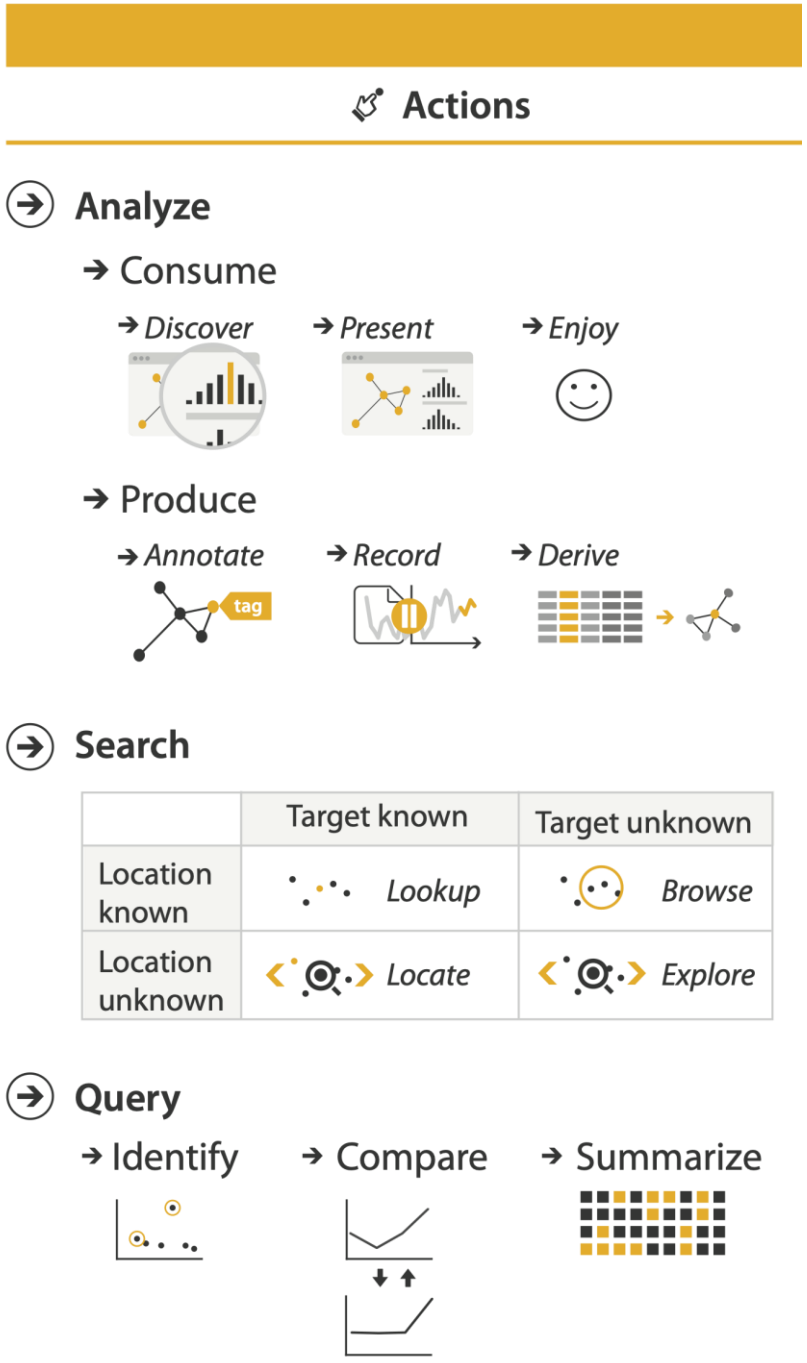
Actions: Query

- how much of the data matters?
  - one: identify
  - some: compare
  - all: summarize



Actions

- independent choices for each of these three levels
  - analyze, search, query
  - mix and match



Task abstraction: Targets

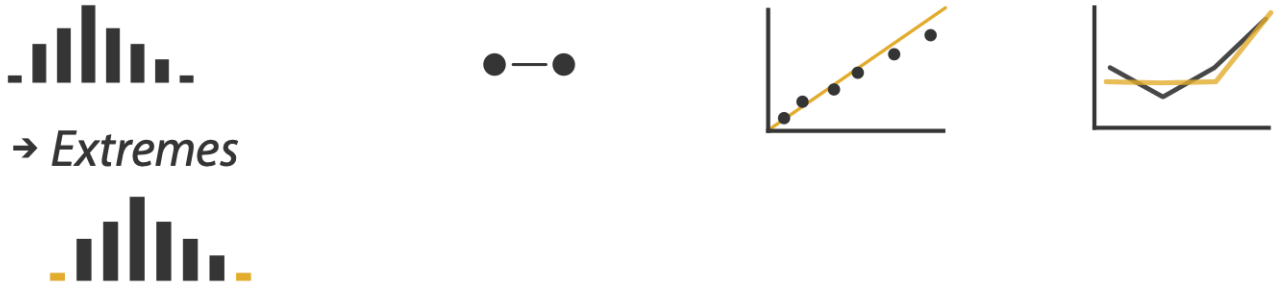
Task abstraction: Targets

- All Data
  - Trends
  - Outliers
  - Features

Task abstraction: Targets

- **All Data**
  - Trends
  - Outliers
  - Features

- **Attributes**
  - One
    - *Distribution*
    - *Extremes*
  - Many
    - *Dependency*
    - *Correlation*
    - *Similarity*



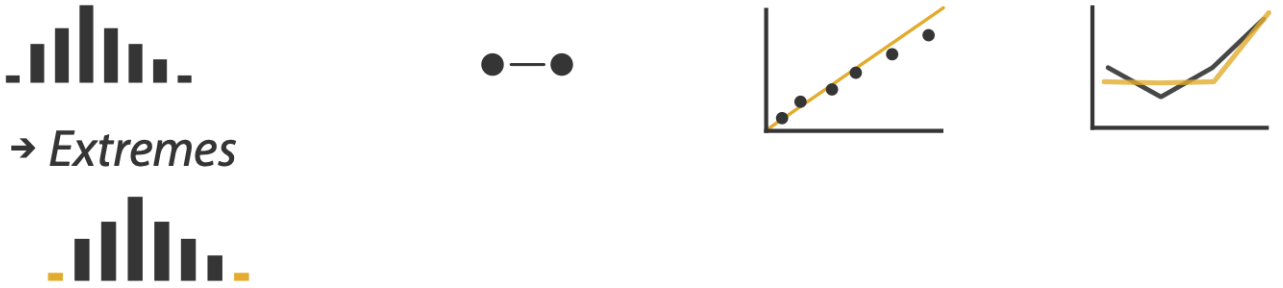
Task abstraction: Targets

All Data

- Trends
- Outliers
- Features

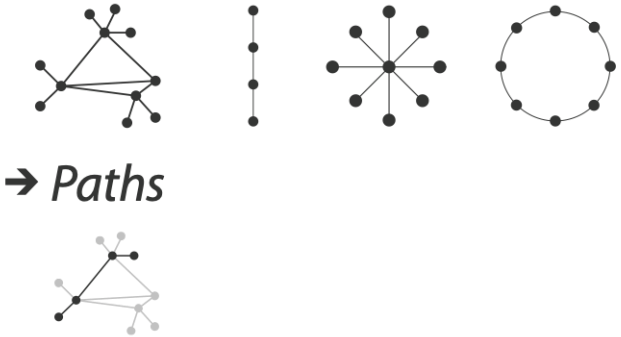
Attributes

- One
  - *Distribution*
  - *Extremes*
- Many
  - *Dependency*
  - *Correlation*
  - *Similarity*



Network Data

- Topology
- *Paths*



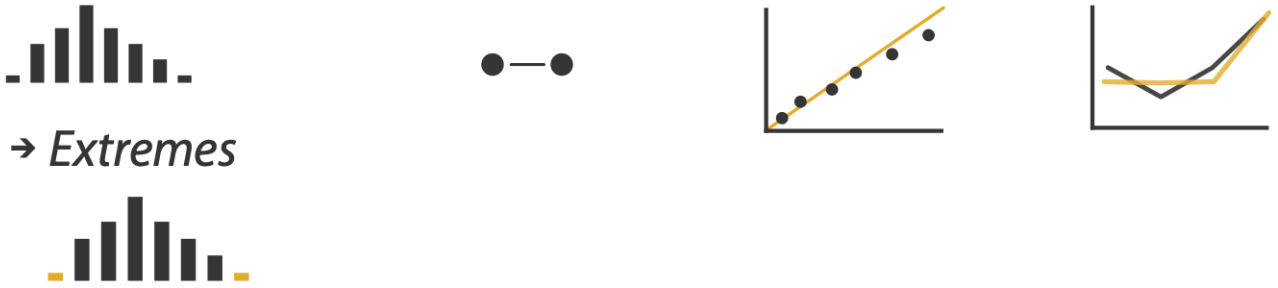
Task abstraction: Targets

All Data

- Trends
- Outliers
- Features

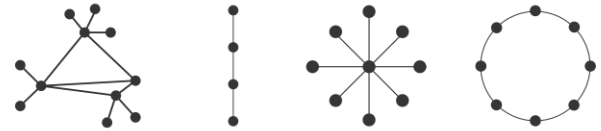
Attributes

- One
  - *Distribution*
  - *Extremes*
- Many
  - *Dependency*
  - *Correlation*
  - *Similarity*



Network Data

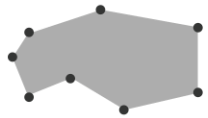
- Topology



- *Paths*

Spatial Data

- Shape



Summary: Goals of Visualization

- Presentation
- Confirmation
- Exploration

Summary: Goals of Visualization

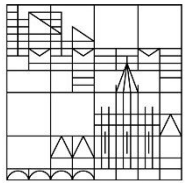
- Presentation
  - Starting point: facts to be presented are fixed a priori.
  - Process: choice of appropriate presentation techniques
  - Result: high-quality visualization of the data to present facts
- Confirmation
- Exploration

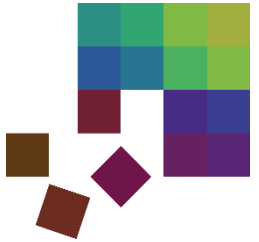
Summary: Goals of Visualization

- Presentation
- Confirmation
  - Starting point: hypotheses about the data
  - Process: goal-oriented examination of the hypotheses
  - Result: visualization of data to confirm or reject the hypotheses
- Exploration

Summary: Goals of Visualization

- Presentation
- Confirmation
- Exploration
  - Starting point: no hypotheses about the data
  - Process: interactive, usually undirected search for structures, trends
  - Result: visualization of data to lead to hypotheses about the data





Break

Thanks for listening

Practice: Design your own visualization!

Get into teams of three and start the discussion!

Todos

- *Select a text dataset of your choice*
  - *You could use your own text data set of interest or look into the github repo*

[A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Datasets

- *Script of the Presidential Debate between Trump and Harris from the 11.09.2024 as retrieved from CNN (https://www.cnn.com/2024/06/27/politics/read-biden-trump-debate-rush-transcript/index.html)*
- *Reports from the Russian-Ukrainian War from the 17.08 – 17.09.2024 (https://acleddata.com/about-acled/)*
- *The complete text of Harry Potter by JK Rowling*

[A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Practice: Design your own visualization!

Get into teams of three and start the discussion!

Todos

- *Select a text dataset of your choice*
  - *You could use your own text data set of interest or look into the github repo*
- *Choose your domain situation*
  - Who could be the target user of your visualization?
  - What is the persons motivation?
- *How could your Visual Encoding look like?*
  - *Sketch the design! You can use Pen and Paper, Powerpoint or Excalidraw (https://excalidraw.com/)*
- What are your text processing steps?
  - List your NLP tasks.

We‘ll have a final discussion on your visualization designs.

[A Multi-Level Typology of Abstract Visualization Tasks. Brehmer and Munzner. IEEE TVCG 19(12):2376-2385, 2013 (Proc. InfoVis 2013).]
[A Nested Model of Visualization Design and Validation. Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]