A Visualization Primer

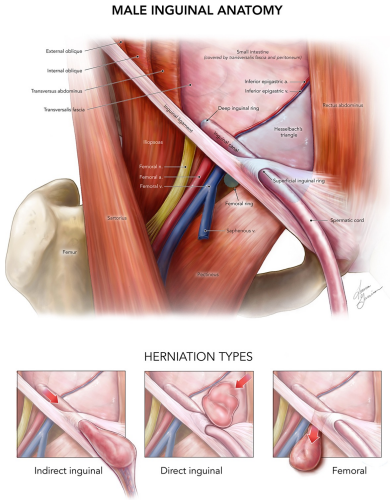
Laura Garrison, University of Bergen

laura.garrison@uib.no

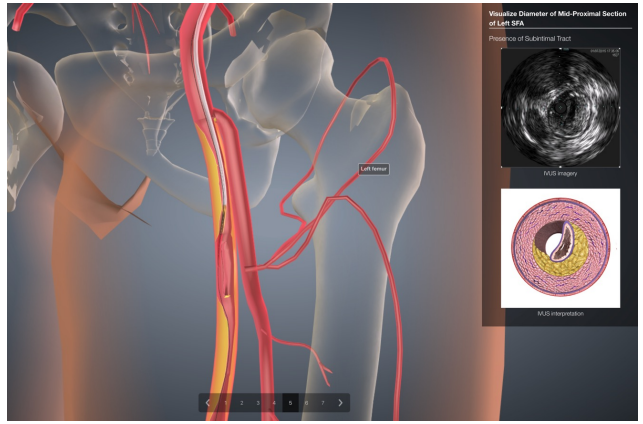
ICTP Workshop 2023

Credit: The Oatmeal

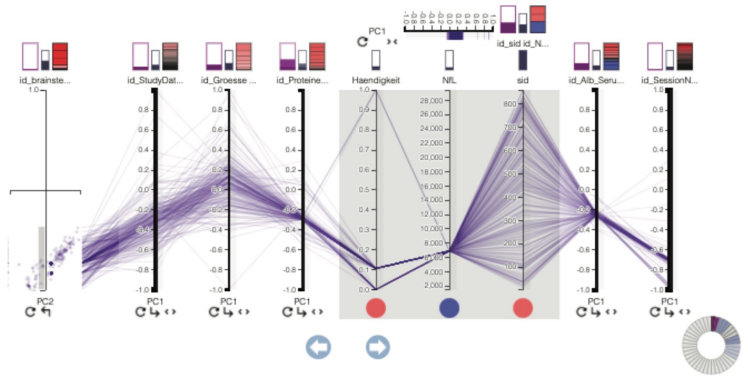
BA **Anatomy & Physiology**
Minors in Art, Chemistry
(NMU, 2009)



MS **Biomedical Visualization**
(UIC, 2012)



Medical & Health Tech Start-ups *(2012-18)*



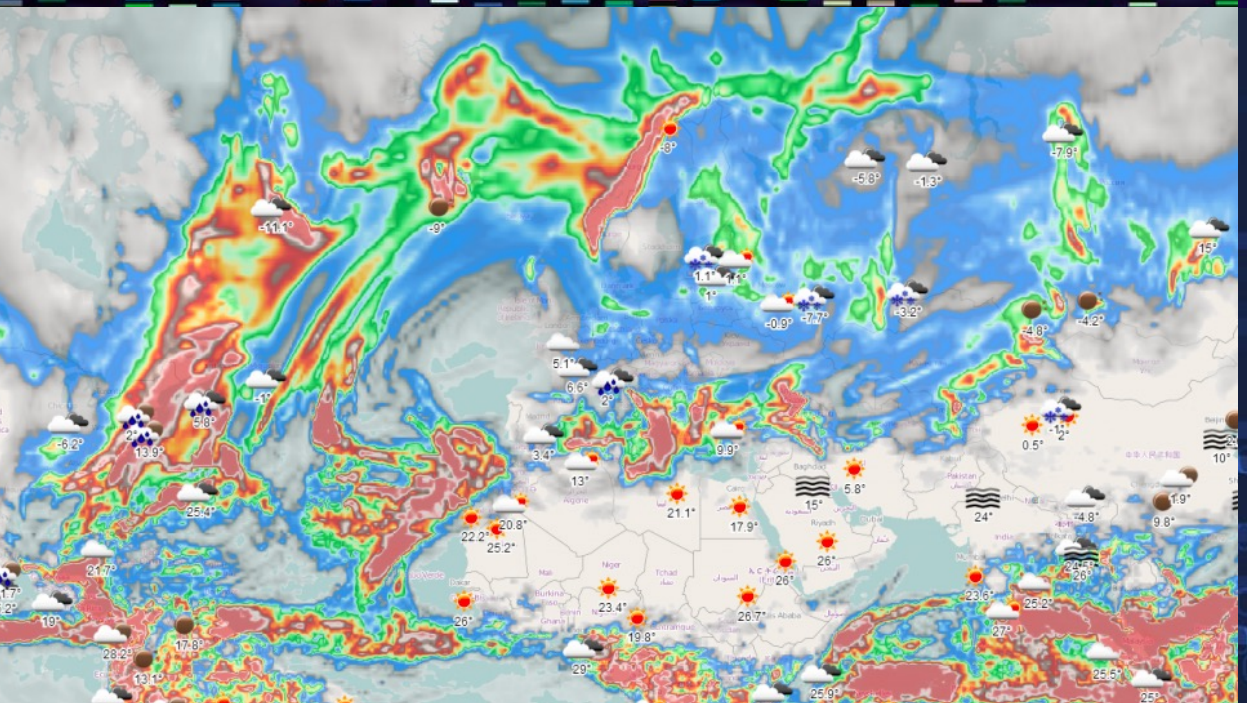
PhD **Visualization**
(UiB, 2022)

Assoc. Prof. in **Visualization**
(UiB, 2023-)

bouvet

CMMIV

(2023)



Data Science

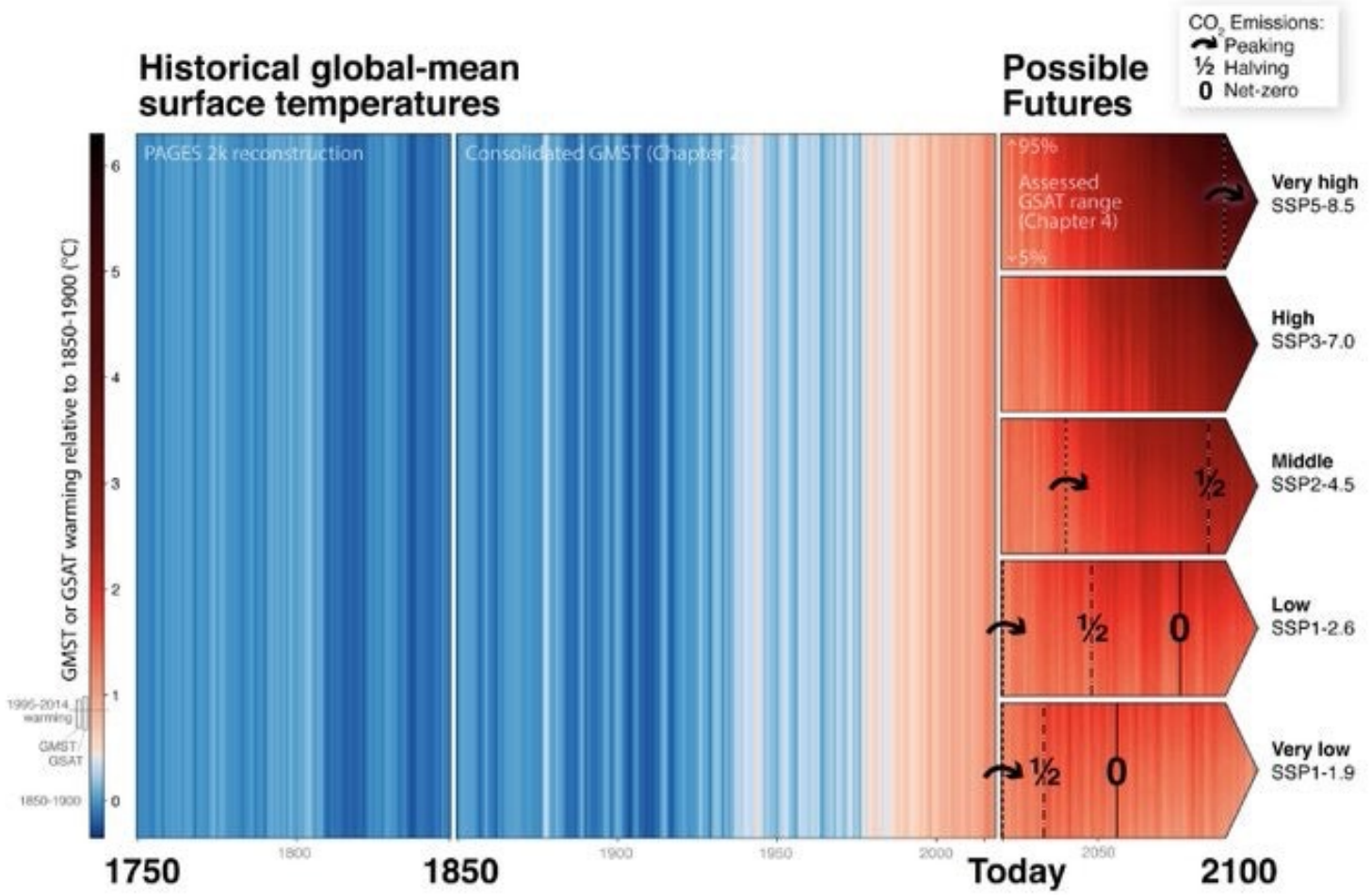
- Extract knowledge and insights from data, often using advanced analytics methods from:
  - Mathematics
  - Statistics
  - Algorithms
  - Machine learning

Src: Kiranshastry - Flaticon

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64-73.

Visualization

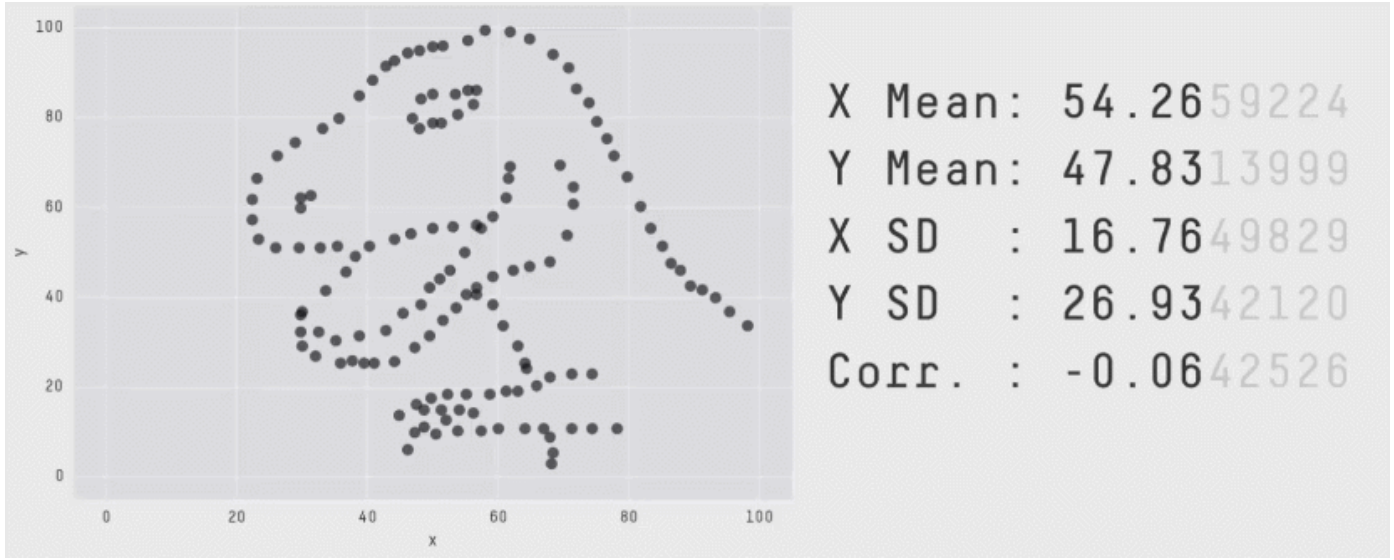
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.



Source: IPCC

Why visualize?

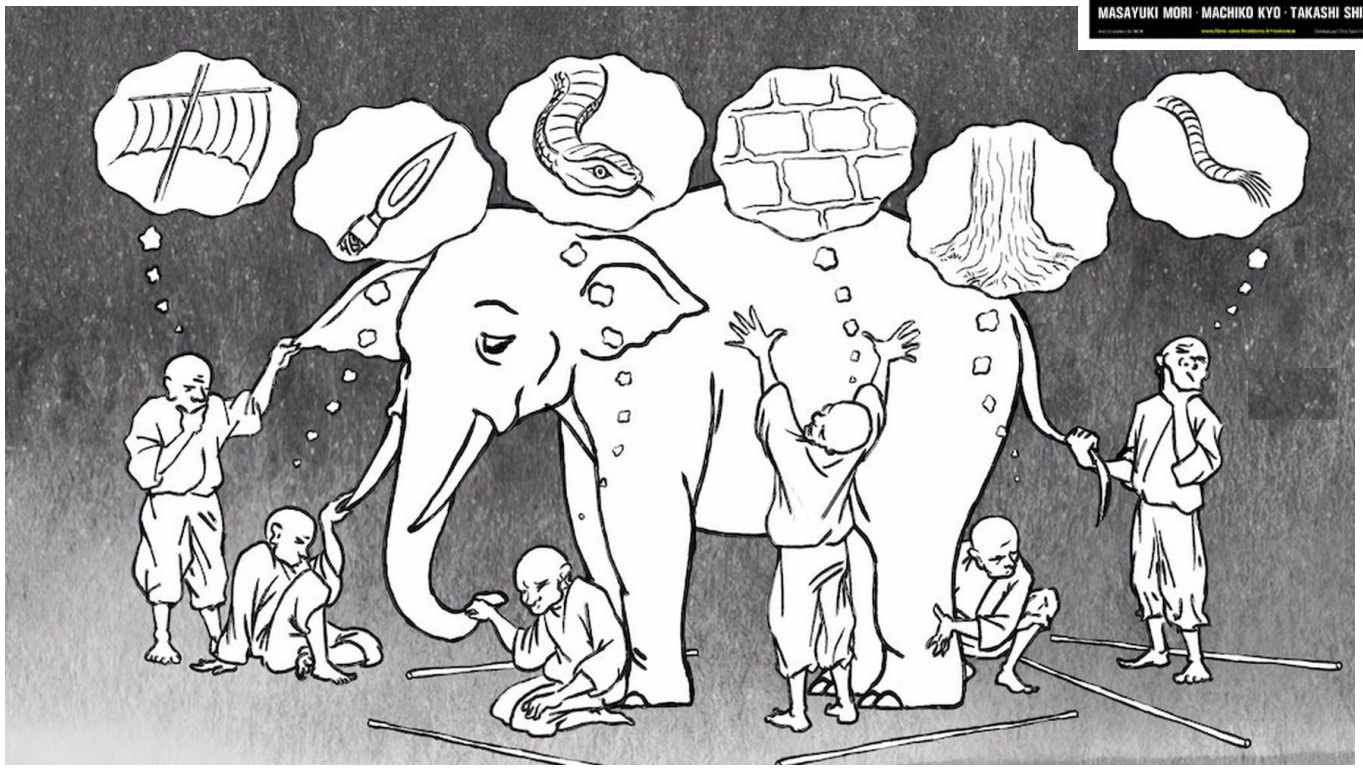
- "visual representation"
  - replace cognition with perception
- "representations of datasets"
  - details matter, summaries can lose information



Matejka, J., & Fitzmaurice, G. (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In Proceedings of the 2017 CHI conference on human factors in computing systems (pp. 1290-1294).

Why visualize?

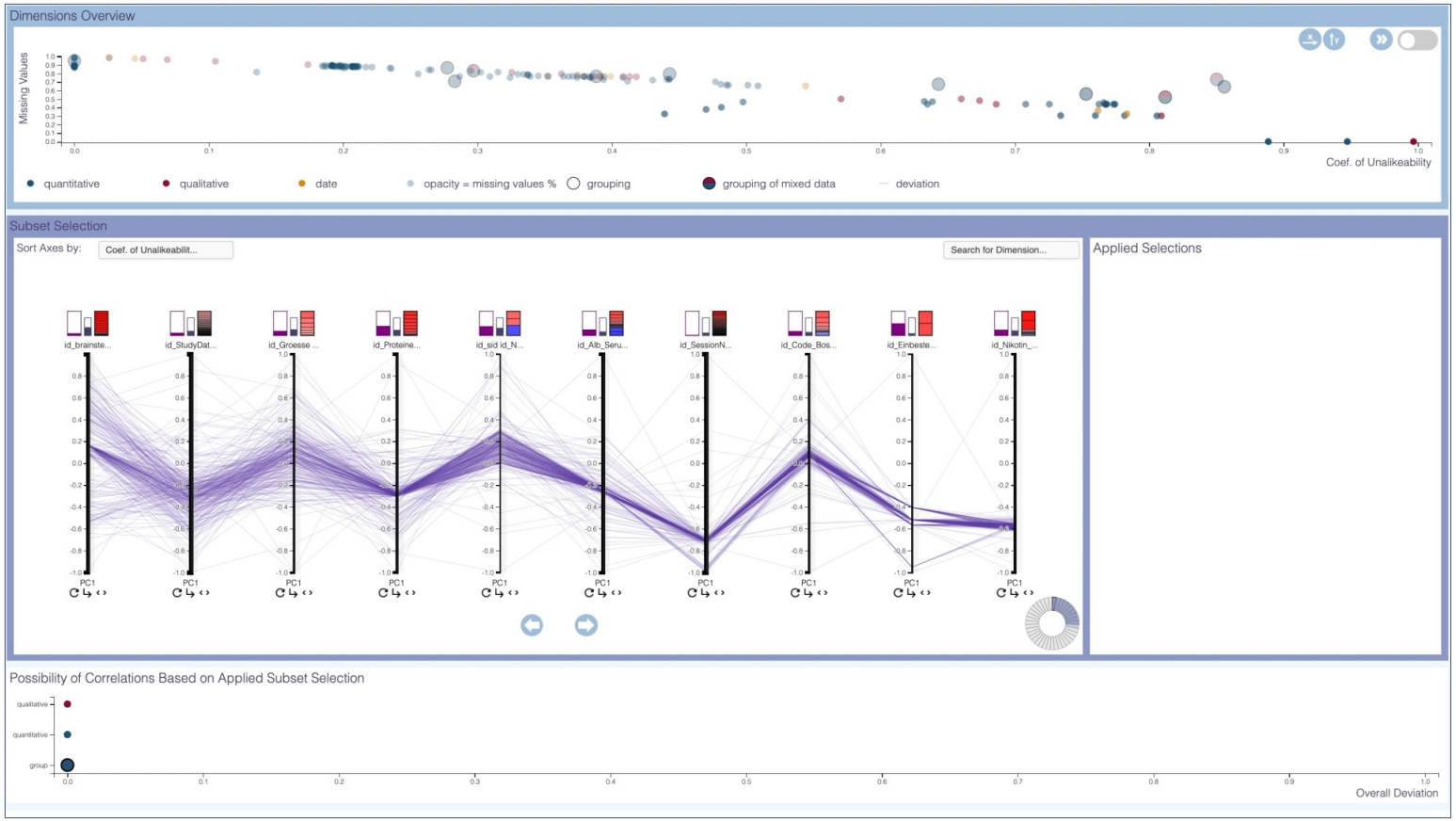
- Rashomon Effect
  - Different models, parameters, representations, etc. can tell different stories
- Visualization can help us spot and understand reasons for these differences



Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3), 199-231.

Why visualize?

- Visualization is not necessary if there is a trustworthy automatic solution
- **Augment** human capabilities, not replace
- Many analysis problems are **ill-specified**

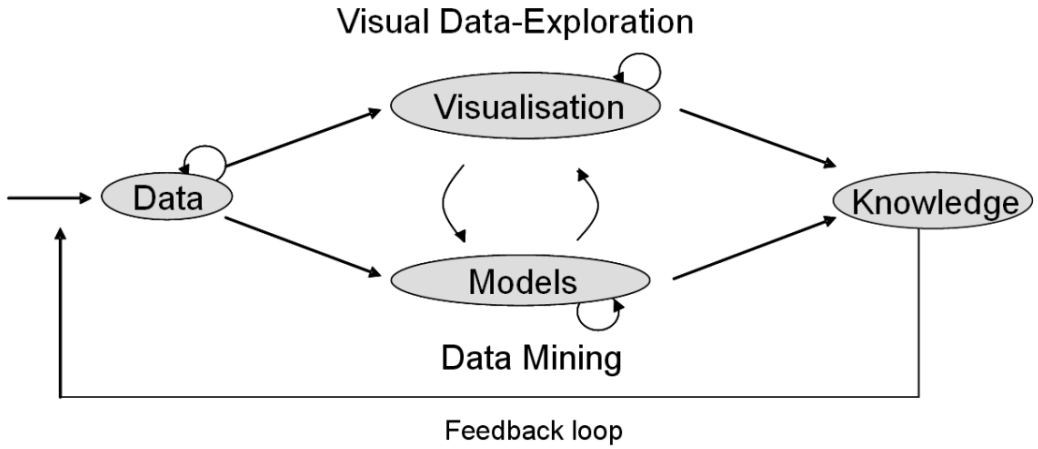


Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 1

Garrison, L., Müller, J., Schreiber, S., Oeltze-Jafra, S., Hauser, H., & Bruckner, S. (2021). Dimlift: Interactive hierarchical data exploration through dimensional bundling. IEEE Transactions on Visualization and Computer Graphics, 27(6), 2908-2922.

Why visualize?

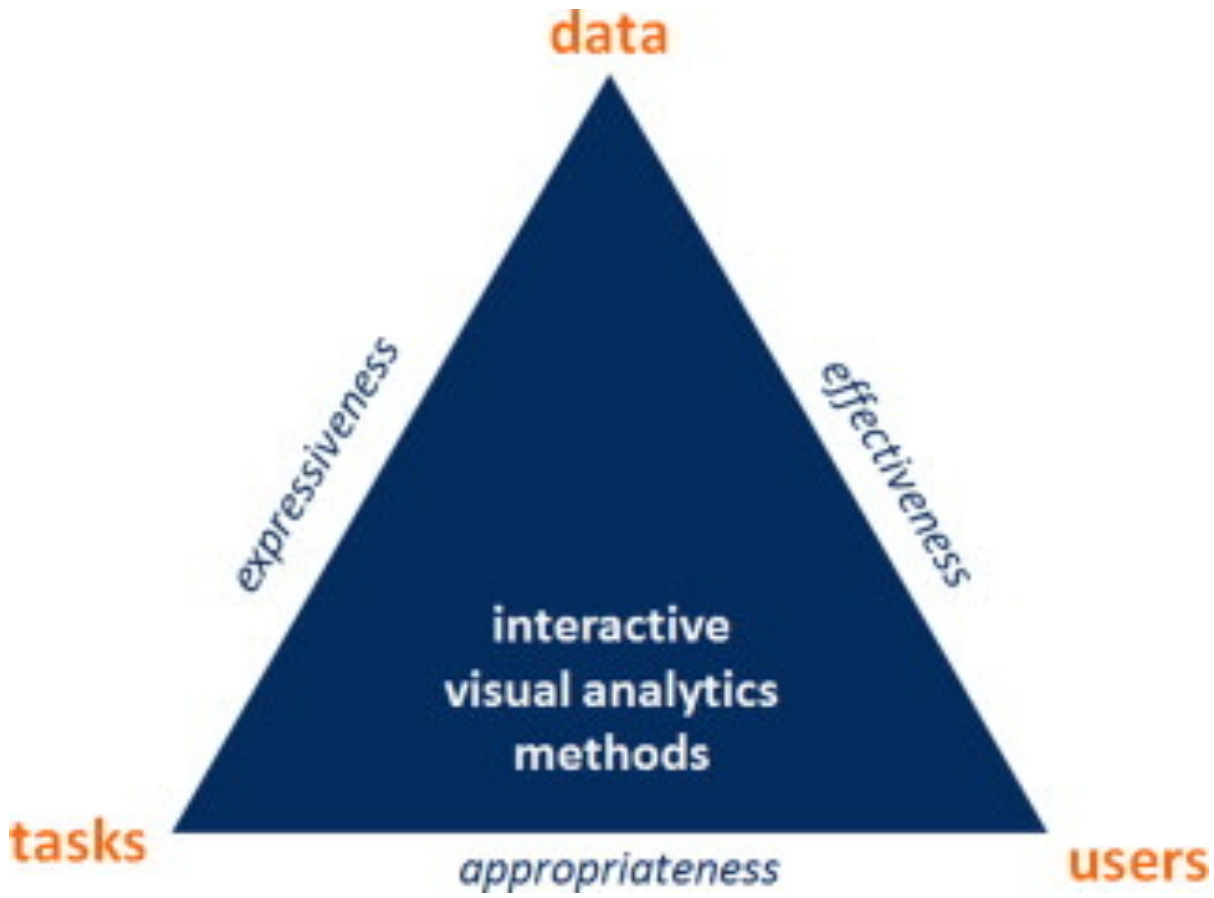
- Visualization integral **throughout** analysis pipeline to help steer a path of inquiry
- Help answer and form new questions:
  - *What do my data look like?*
  - *What are the requirments for developing a more complex model?*
  - *What if I remove/adjust this parameter?*
  - *Do I trust/can I verify the data/model?*



Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In Information visualization (pp. 154-175). Springer, Berlin, Heidelberg.

Visualization in YOUR workflow

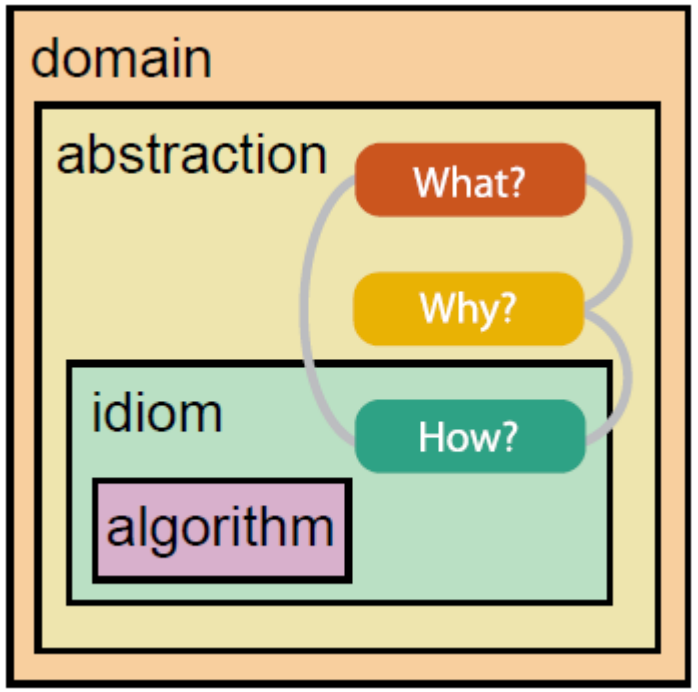
- Good visualization is possible for anyone
- Consider:
  - Data
  - User(s)
  - Task(s)



Miksch, Silvia, and Wolfgang Aigner. "A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data." Computers & Graphics 38 (2014): 286-290.

Visual Analysis Framework

- Domain
  - Who are you visualizing for? Yourself?
- Abstraction
  - What is shown (data abstraction)
  - Why showing (task abstraction)
- Idiom
  - How is it being shown
    - visual encoding idiom (how do you draw the picture)
    - interaction idiom (how do you manipulate the picture)
- Algorithm
  - efficient computation to show the picture



The visualization process

What - why - how

What?

Datasets

Data Types

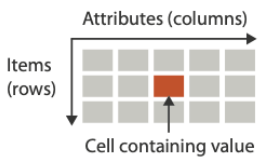
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

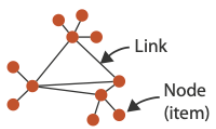
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

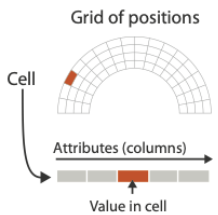
- Tables



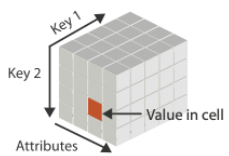
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

What?

Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*

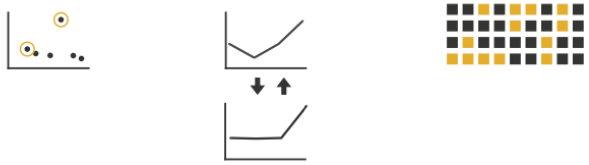
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

Search

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

Query

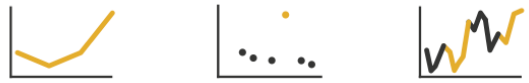
- Identify
- Compare
- Summarize



Targets

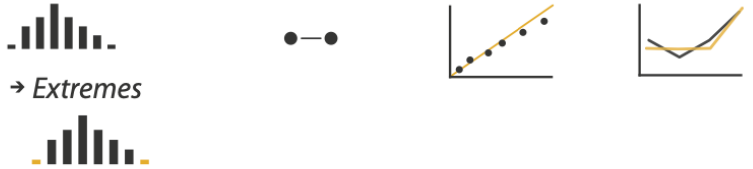
All Data

- Trends
- Outliers
- Features



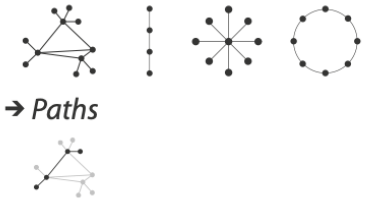
Attributes

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



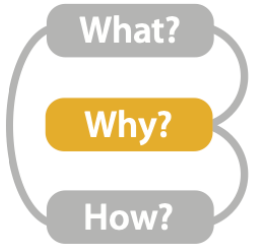
Network Data

- Topology
- *Paths*



Spatial Data

- Shape



What?

Why?

How?

Encode

Arrange

- Express
- Separate
- Order
- Align
- Use

Map

from categorical and ordered attributes

- Color
  - *Hue*
  - *Saturation*
  - *Luminance*
- Size, Angle, Curvature, ...
- Shape
- Motion

Direction, Rate, Frequency, ...

Manipulate

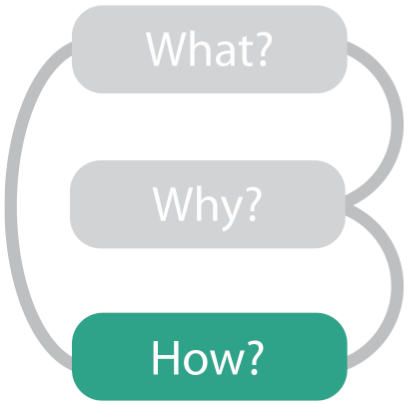
- Change
- Select
- Navigate

Facet

- Juxtapose
- Partition
- Superimpose

Reduce

- Filter
- Aggregate
- Embed



The “how” impacts interpretation

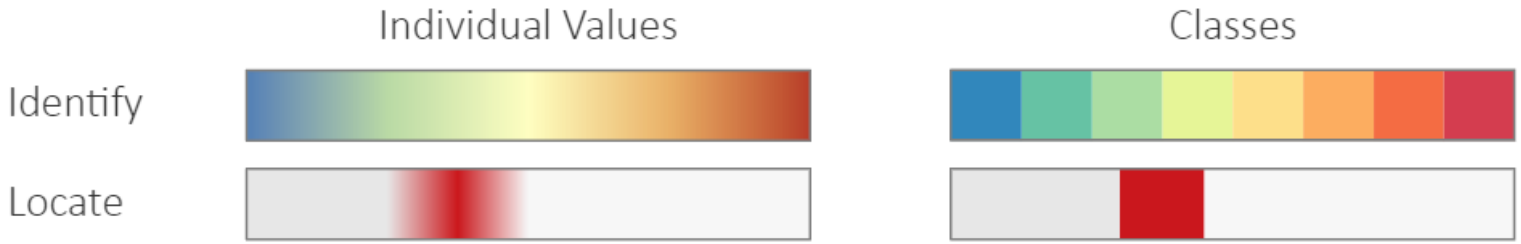
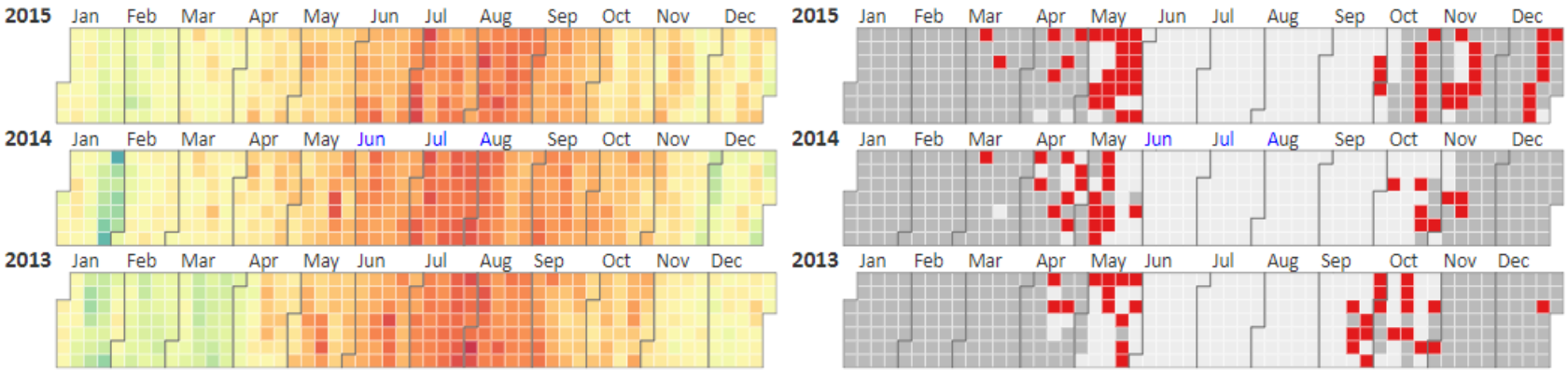
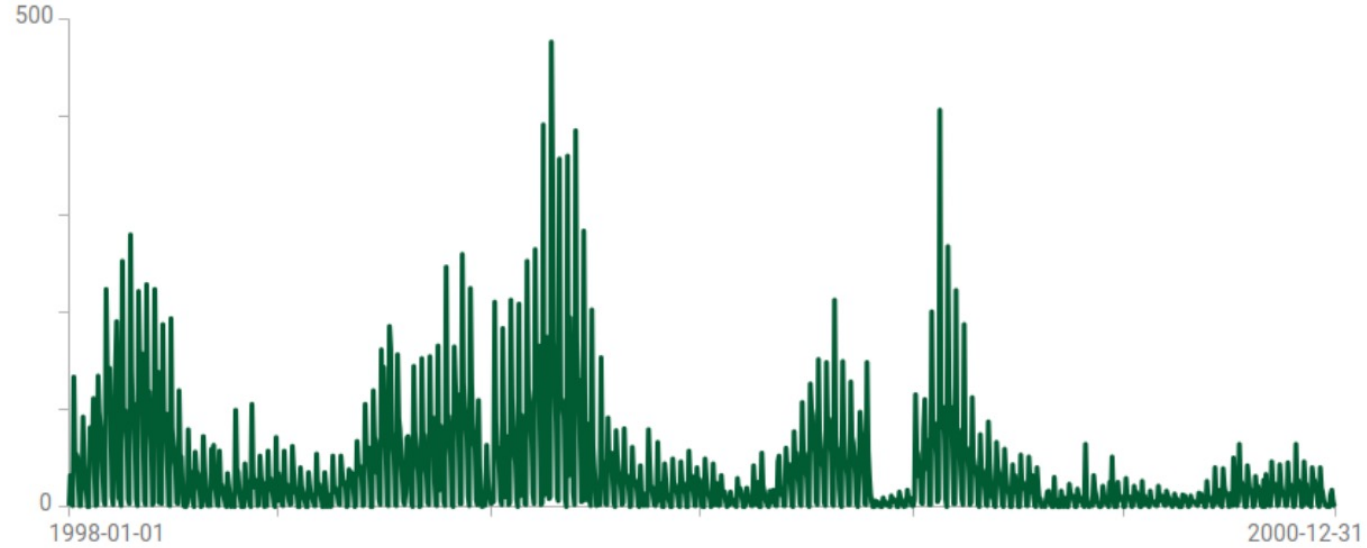


Figure 3.4 Color maps for identifying and locating values and classes.

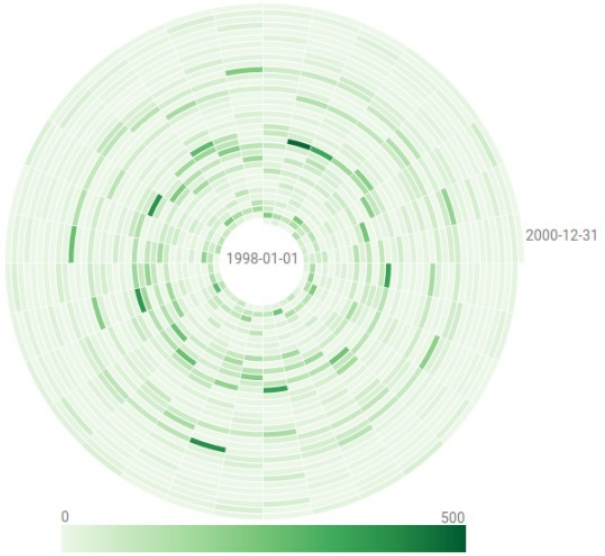


(a) Color coding for identification tasks. (b) Color coding for location tasks.

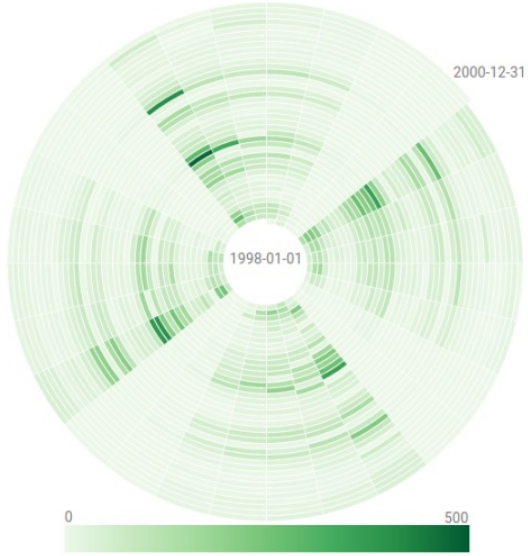
Figure 3.5 Applying the color maps from Figure 3.4 to temperature data. Adapted from bl.ocks.org/mbostock/4063318.



(a) Line plot.



(b) Spiral plot (cycle length 32 days).



(c) Spiral plot (cycle length 28 days).

The language of visualization

“Marks” – represent items or links

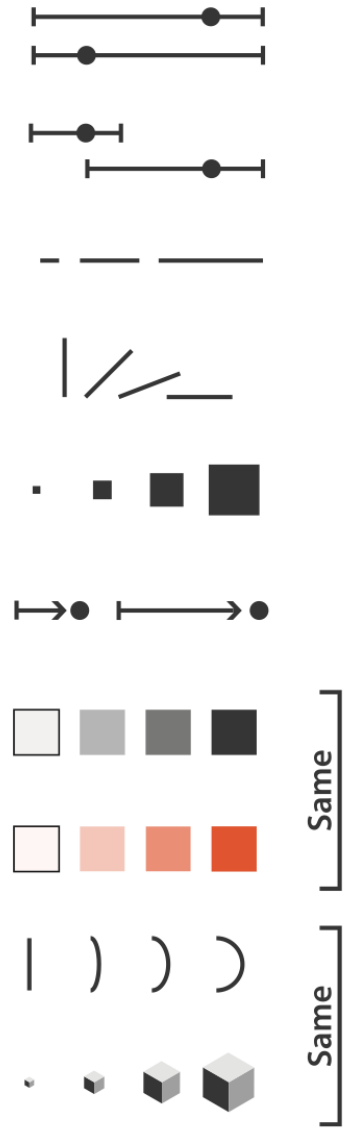
Points

Lines

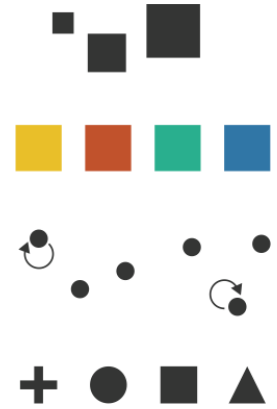
Interlocking Areas

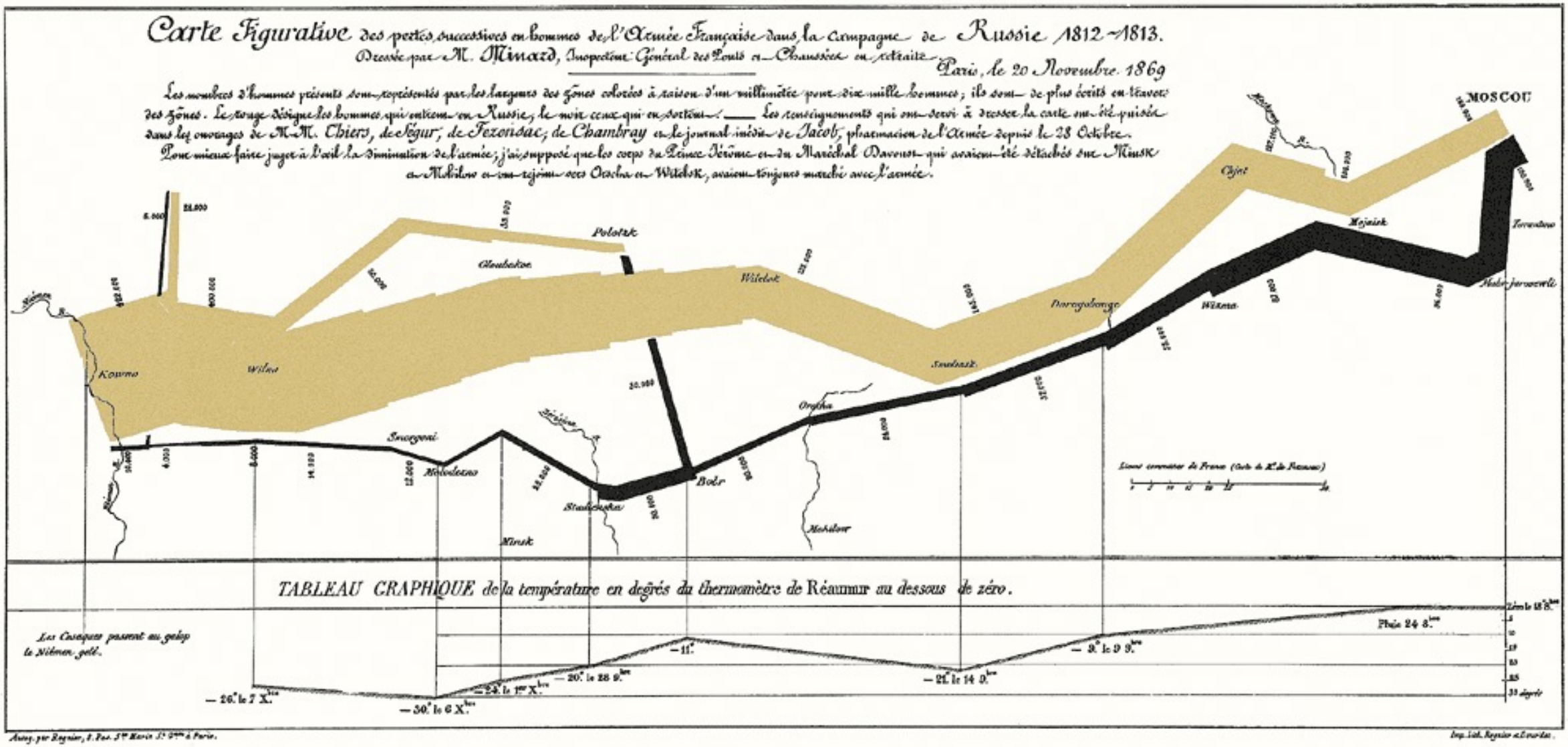
“Channels”

Magnitude Channels: Ordered Attributes

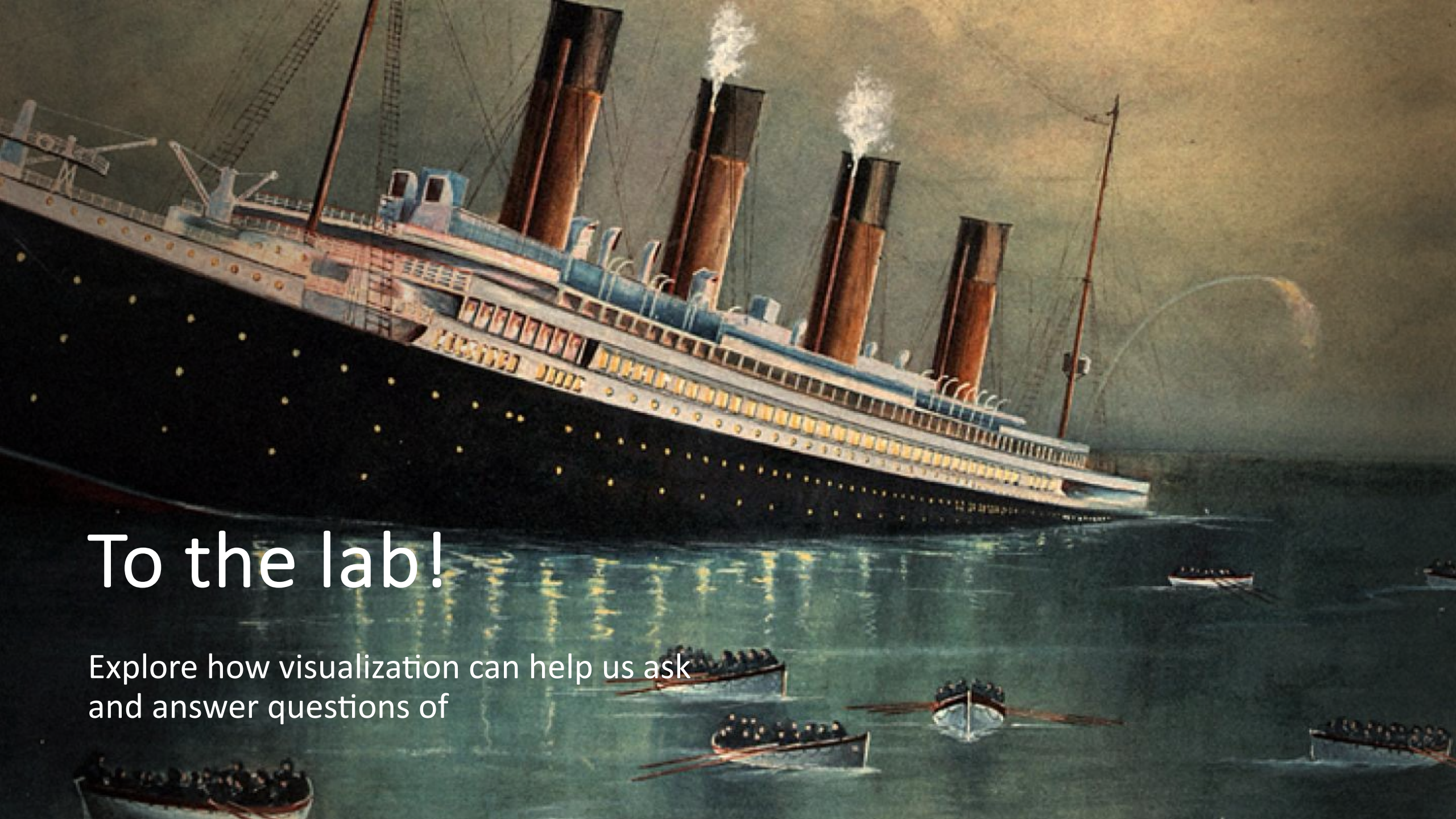


Identity Channels: Categorical Attributes



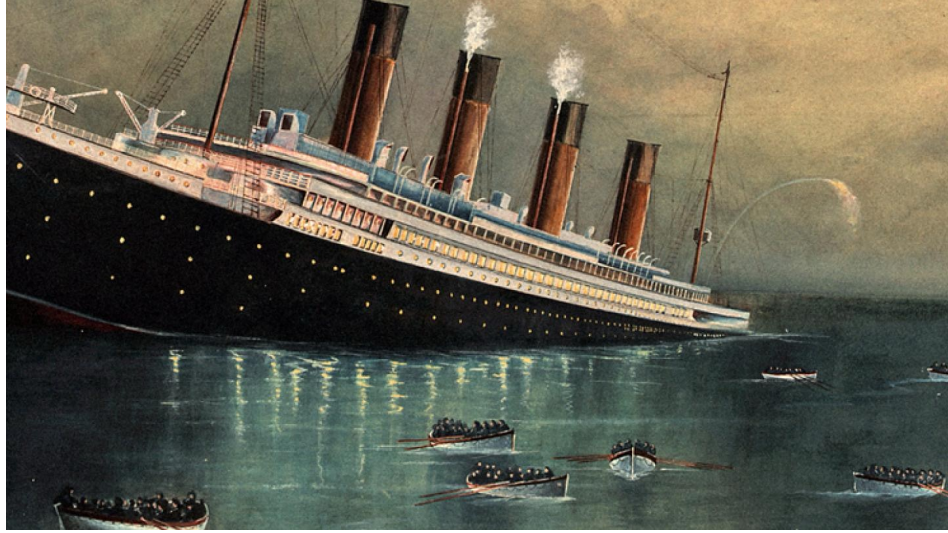


Figurative map of the successive losses in men of the French Army in the Russian campaign 1812–1813, by Charles Minard (1869)



To the lab!

Explore how visualization can help us ask and answer questions of



VARIABLE DESCRIPTIONS

Pclass	Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
survival	Survival (0 = No; 1 = Yes)
name	Name
sex	Sex
age	Age
sibsp	Number of Siblings/Spouses Aboard
parch	Number of Parents/Children Aboard
ticket	Ticket Number
fare	Passenger Fare (British pound)
cabin	Cabin
embarked	Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

matplotlib

pandas