DATA VISUALIZATION

Bùi Tiến Lên

2023

Contents

Data Visualization

- Explainable Artificial Intelligence
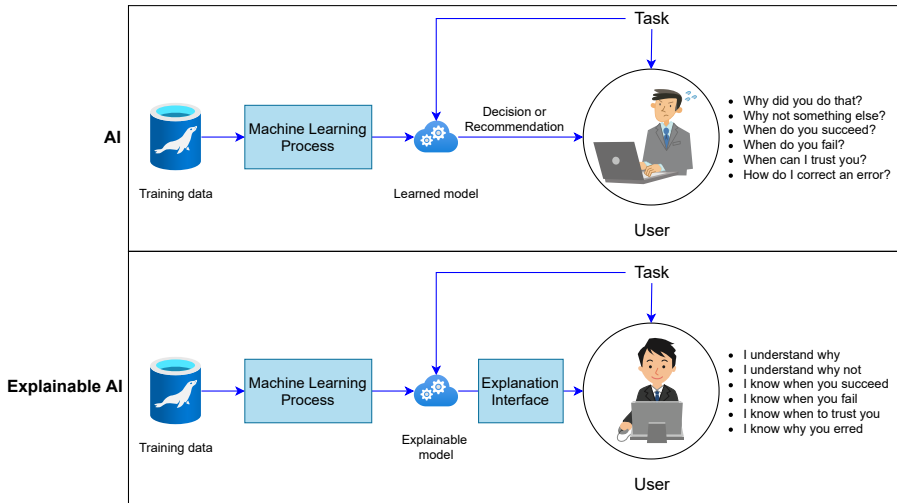- Data Visualization

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Explainable Artificial Intelligence

- **Explainable** could mean **interpretable**
- **Abstraction**: how much of an explanation do we need?
- The **Explainable AI** (XAI) program aims to create a suite of machine learning techniques that:
  - Produce more explainable models, while maintaining a high level of learning performance (prediction accuracy)
  - Enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

Data
Visualization

Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

The Explainable Artificial Intelligence (cont.)

AI

Training data → Machine Learning Process → Learned model → Decision or Recommendation → User

Task

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Explainable AI

Training data → Machine Learning Process → Explainable model → Explanation Interface → User

Task

- I understand why
- I understand why not
- I know when you succeed
- I know when you fail
- I know when to trust you
- I know why you erred

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Human vs. Model

The challenge is people don't understand the system and the system doesn't understand the people.

- We need models for the system to use to understand and explain things to the human.
- We also need models in the human's head about what the system does.

Understand each other

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze
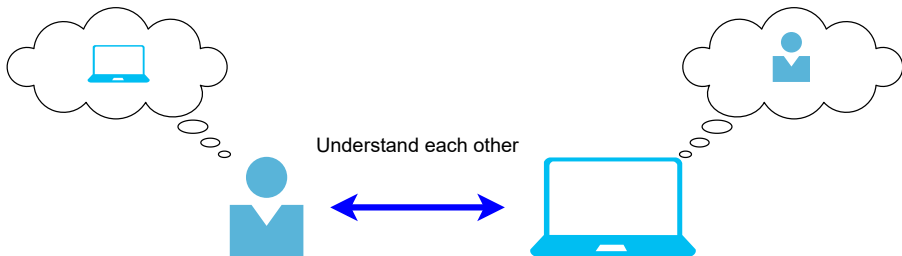
What is Visualization

Concept 1

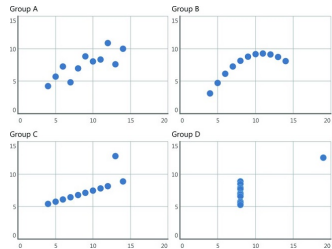**Visualization** is a computational process that generates visual representations of data.

- It offers a method to see the unseen.

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Why Do We Visualize Data?

- Table with four groups of numbers: What do they tell you?

Group A		Group B		Group C		Group D	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Why Do We Visualize Data? (cont.)

Consider the following table, which shows sales numbers for three categories, by quarter, over a four-year period. What trends can you see?

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	$463,988	$352,779	$338,169	$317,735	$320,875	$287,934	$319,537	$324,319
Office Supplies	$232,558	$290,055	$265,083	$246,946	$219,514	$202,412	$198,268	$279,679
Technology	$563,866	$244,045	$432,299	$461,616	$285,527	$353,237	$338,360	$420,018

Category	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	2016 Q3	2016 Q4
Furniture	$307,028	$273,836	$290,886	$397,912	$337,299	$245,445	$286,972	$313,878
Office Supplies	$207,363	$183,631	$191,405	$217,950	$241,281	$286,548	$217,198	$272,870
Technology	$333,002	$291,116	$356,243	$386,445	$386,387	$397,201	$359,656	$375,229

9

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Visualization Systems

Concept 2

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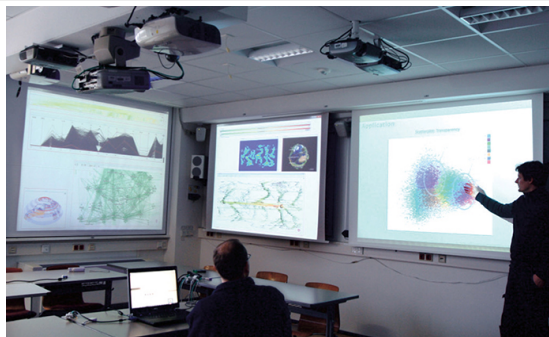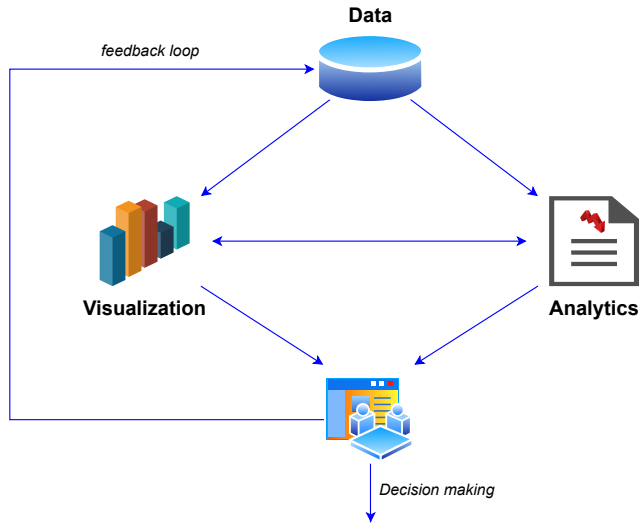
- Computer-based visualization systems

Data
Visualization

Explainable Artificial
Intelligence

Data Visualization

Visual Data
Analyze

Diagram of Components

Data

feedback loop

Visualization

Analytics

Decision making

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

**Visual Data
Analyze**

Why have a human in the loop?

- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods
- Vis allows people to analyze data when they don't know exactly what questions they need to ask in advance
- **Don't need** vis when fully automatic solution exists and is trusted

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Why have a computer in the loop?

- By using computers, we can build tools that allow people to explore or present large datasets that would be completely infeasible to draw by hand, thus opening up the possibility of seeing how datasets change over time.
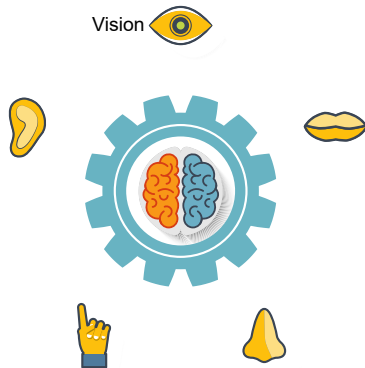

(a)   (b)

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze
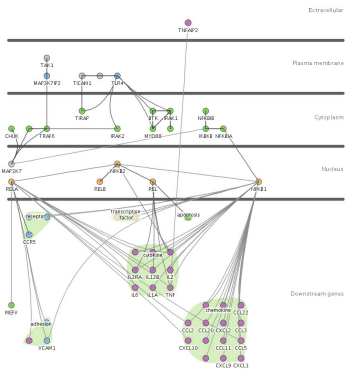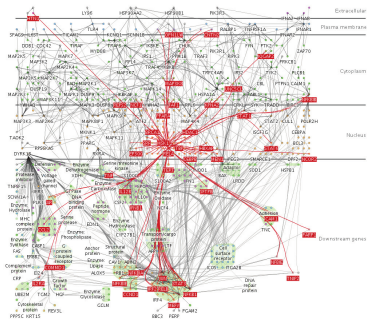
Why depend on vision?

- **Human visual system** is high-bandwidth channel to brain, overview possible due to background processing
- **Sound**: lower bandwidth and different semantics, overview not supported
- **Touch**: impoverished record/replay capacity, only very low-bandwidth communication thus far
- **Taste**
- **Smell**

Vision

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Why Use Interactivity?

- **Interactivity** is crucial for building vis tools that handle complexity.
- It allows us to actively take part in the visual data analysis.

Data
Visualization

Explainable Artificial
Intelligence
Data Visualization

**Visual Data
Analyze**

Why analyze?

Three-part analysis framework for a vis instance:

- **what** data is shown in the views
- **why** is the task being performed
- **how** is the vis idiom constructed in terms of design choices.

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze
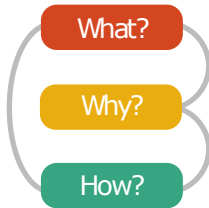
Why analyze? (cont.)

- Analyzing vis usage as chained sequences of instances, where the output of one instance is the input to another.

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

**Visual Data
Analyze**
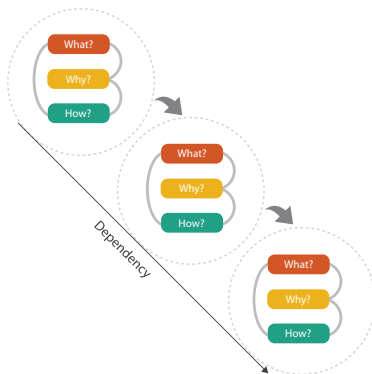
Four levels, three questions

- **Domain situation**
  - who are the target users?
- **Abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**, often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? **task abstraction**
- **Idiom (figure, chart, diagram)**
  - **how** is it shown?
    - visual encoding idiom: how to draw
    - interaction idiom: how to manipulate
- **Algorithm**
  - efficient computation

What?

Why?

How?

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Validation

- Different ways to get it wrong at each level

Domain situation
You misunderstood their needs

Data/task abstraction
You're showing them the wrong thing

Visual encoding/interaction idiom
The way you show it doesn't work

Algorithm
Your code is too slow

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Validation (cont.)

- Use methods from different fields at each level

Domain situation
Observe target users using existing tools

Data/task abstraction

Visual encoding/interaction idiom
Justify design with respect to alternatives

Algorithm
Measure system time/memory
Analyze computational complexity

Analyze results qualitatively
Measure human time with lab experiment (*lab study*)

Observe target users after deployment (*field study*)

Measure adoption

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Good Visualization

Good visualizations

- are based on good data
- represent the data accurately
- show the right amount of data
- attract user's attention

Data
Visualization
Explainable Artificial
Intelligence
Data Visualization

Visual Data
Analyze

Figure Validation

- **Ugly**: A figure that has aesthetic problems but otherwise is clear and informative
- **Bad**: A figure that has problems related to perception; it may be unclear, confusing, overly complicated, or deceiving
- **Wrong**: A figure that has problems related to mathematics; it is objectively incorrect
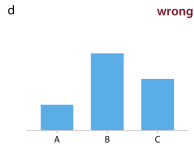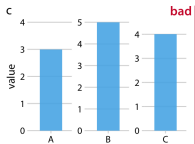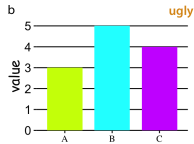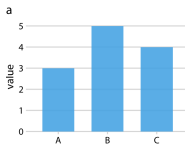
References

Goodfellow, I., Bengio, Y., and Courville, A. (2016).
Deep learning.
MIT press.

Munzner, T. (2014).
Visualization analysis and design.
CRC press.

Russell, S. and Norvig, P. (2016).
Artificial intelligence: a modern approach.
Pearson Education Limited.

Ward, M. O., Grinstein, G., and Keim, D. (2015).
Interactive data visualization: foundations, techniques, and applications.
CRC Press.