Artificial Intelligence and Data Analytics for Engineers (AIDAE)

Lecture 10
July, 10th

Andrés Posada

Today’s Lecturer

Anas Abdelrazeq

Marco Kemmerling

Vladimir Samsonov

Artificial Intelligence and Data Analytics for Engineers
Overview Lectures 1 – 4

1 — **Introduction to Data Analytics and Artificial Intelligence in Engineering**: Organizational matters (e.g. exam, exercises, dates). Goals, Challenges, Obstacles, and Processes.

2 — **Introduction into the primary programming language of the lecture, Python**: Syntax, libraries, IDEs etc. Why is Python the *lingua franca* of the Data Scientist?

3 — **Data Preparation**: Cleansing and Transformation. How do real world data sets look like and why is cleaning and transformation an integral part of a Data Scientist's workflow?

4 — **Data Integration**: Architectures, Challenges, and Approaches. How can you integrate various data sources into an overarching consolidating schema and why is this important?

Artificial Intelligence and Data Analytics for Engineers
Overview Lectures 5 – 8

✓ 5	**Data Representation**: Feature Extraction and Selection. How to pick relevant features for the task at hand. Manual vs automatic methods. What is the curse of dimensionality?
✓ 6	**Data-Driven Learning**: Supervised (Classification, Regression) methods and algorithms. What is an artificial neural net? What methods are there for evaluation of your model?
✓ 7	**Data-Driven Learning**: Unsupervised (Clustering) methods and algorithms. How can machines learn without labels? What methods are there for evaluation of your model?
✓ 8	**Environment-Driven Learning:** Reinforcement Learning

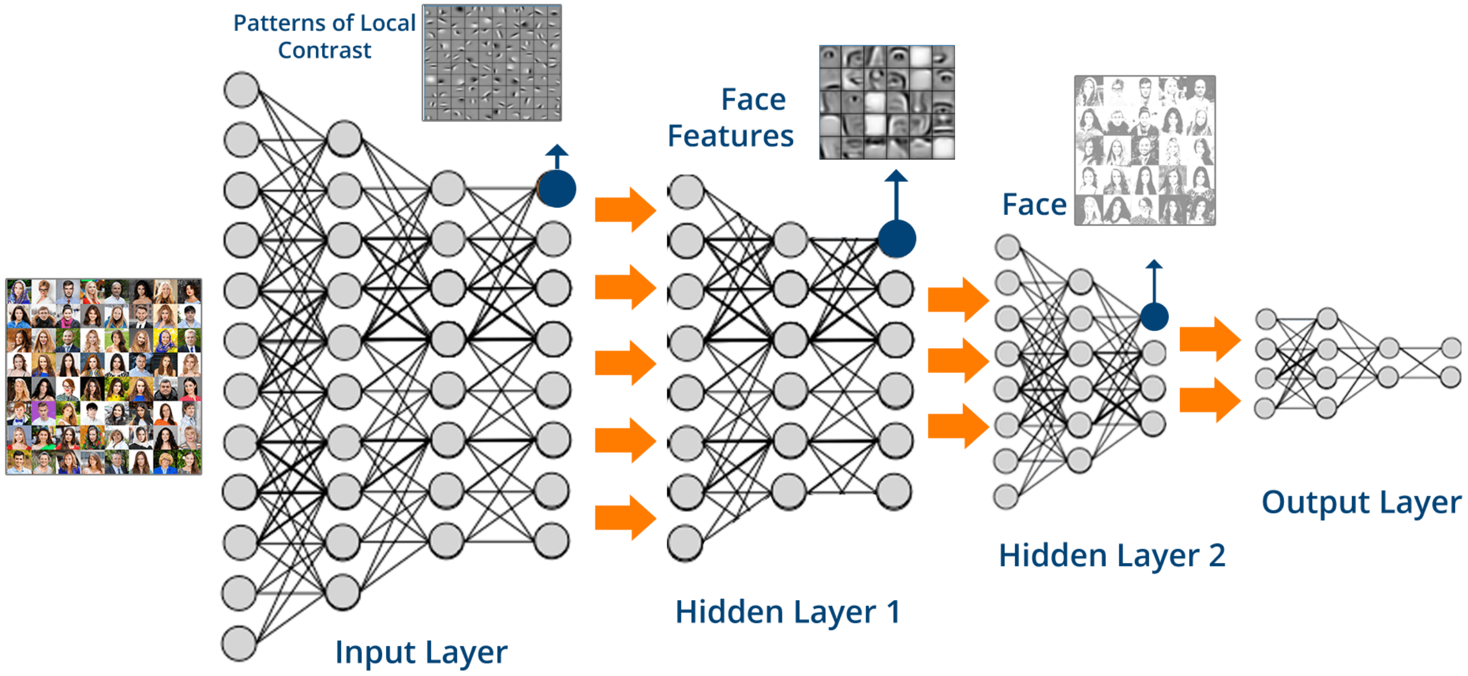
Artificial Intelligence and Data Analytics for Engineers
Overview Lectures 5 – 8

9	**Data-Driven Learning: Artificial Neural Networks**
10	**State-of-the-Art Methods**: Deep Neural Networks (e.g. GANs, CNNs, Restricted Boltzman Machines).
11	**Visual Analytics**: Why is an interactive exploratory approach to data analytics used in the industry and what is meant by "Overview first, filter, details-on-demand"?

Recap Lecture 10

What is Deep Learning?

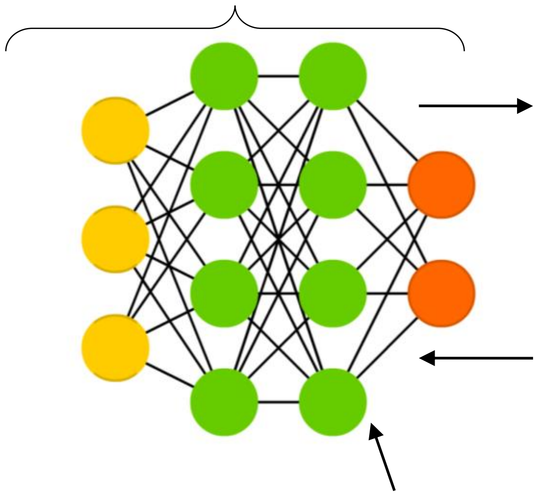
! Deep learning is a subset of machine learning algorithms, in which multi-layered neural networks are used to model the dependencies between input data and a desired output.



Source: Nvidia deep learning

Anatomy a deep neural network

Architecture:
Deep Neural networks are composed of multiple interconnected layers (groups of units). Its architecture refers to the way the units and layers are connected.



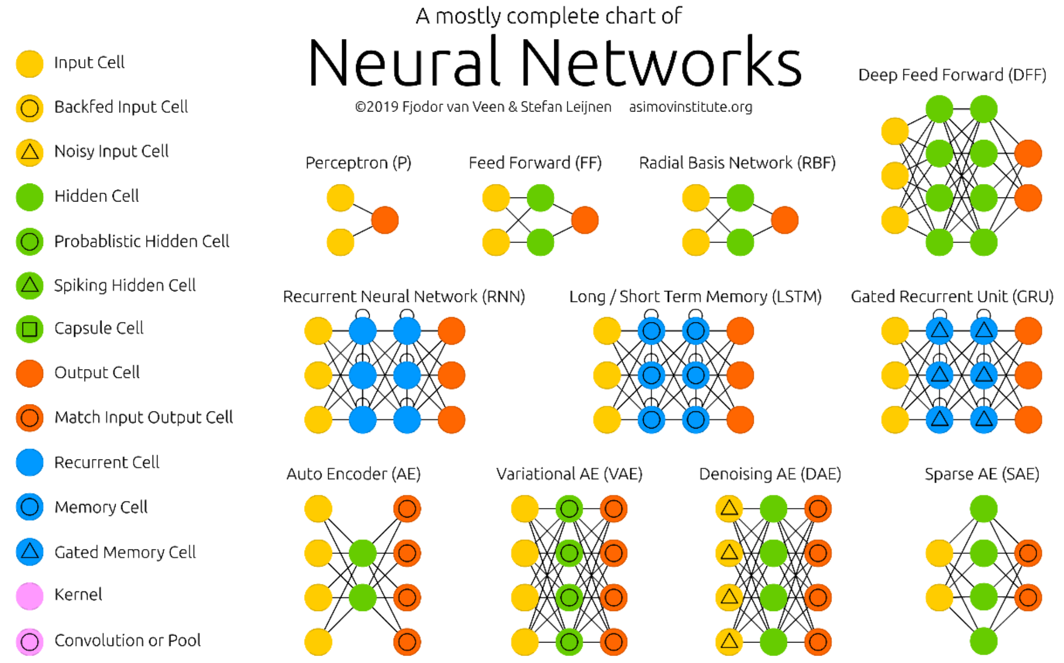
In the **Forward pass**, the input values are used and the operations of the units are computed systematically to obtain the output values of the network.

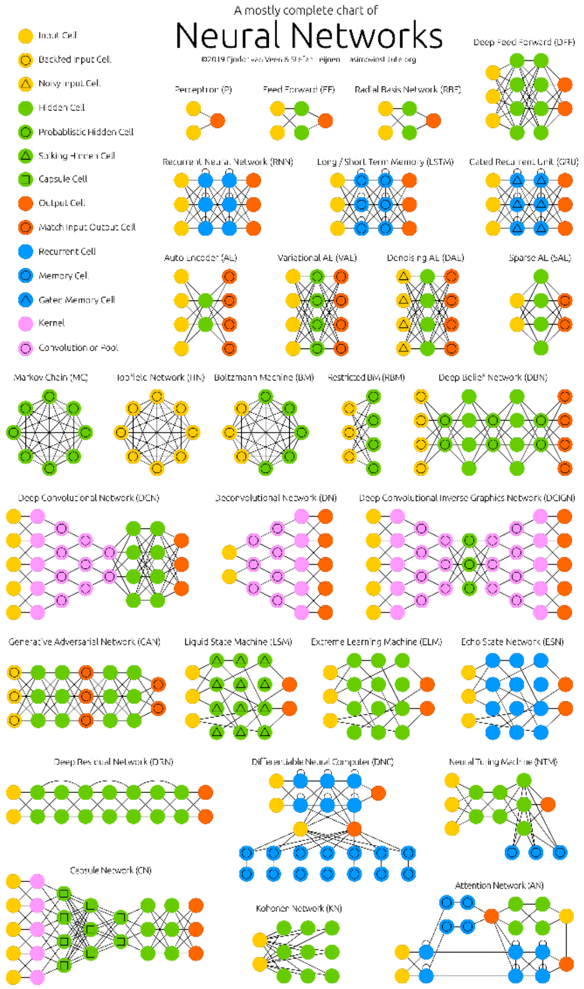
In the **Backward pass**, the obtained output and the desired output are used by the **Loss function** to obtain the errors/gradients and back propagate them through the network. This is latter used to modify the parameters/weights of the units.

Neurons/Units: basic computational units, it receives inputs and computes certain output based on internal parameters and a predefined function.

Some types of Deep neural networks

Depending on the type of units and architectures of the neural networks, they can be used for different tasks, such as object detection, segmentation, classification, and translation.





http://www.asimovinstitute.org/wp-content/uploads/2019/04/NeuralNetworkZo19High.png

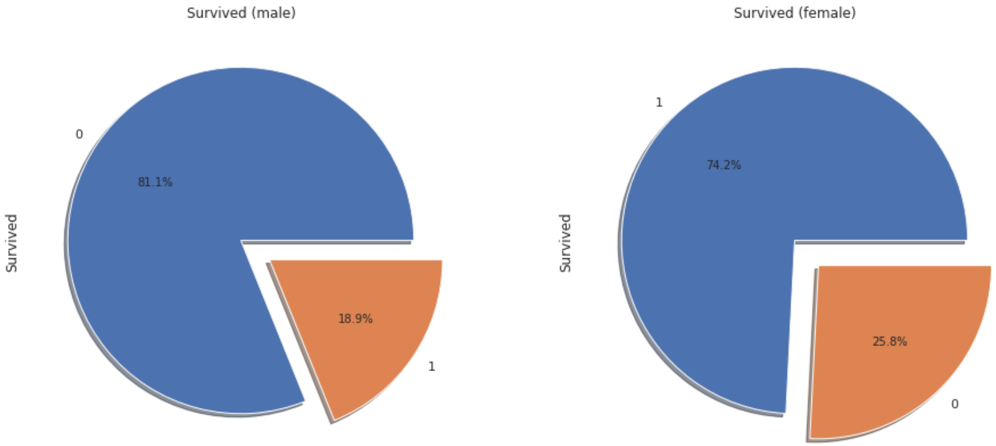
Lecture 11: Visual Analytics

Visual analytics

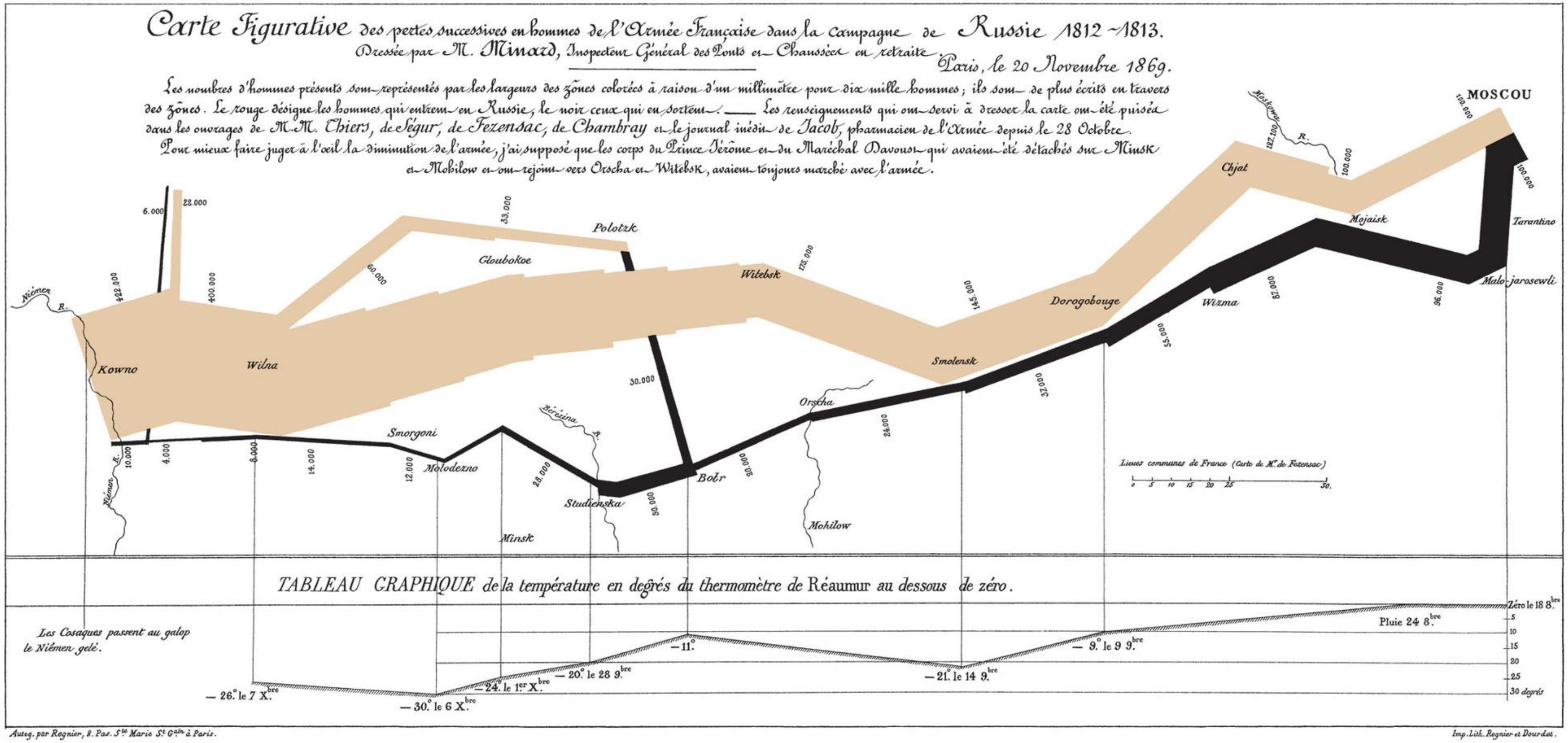
Visualization
Is any technique for creating visual representations (images, diagrams, or animations) from data, in order to communicate a message

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

→ Process
Clean
Visualize →



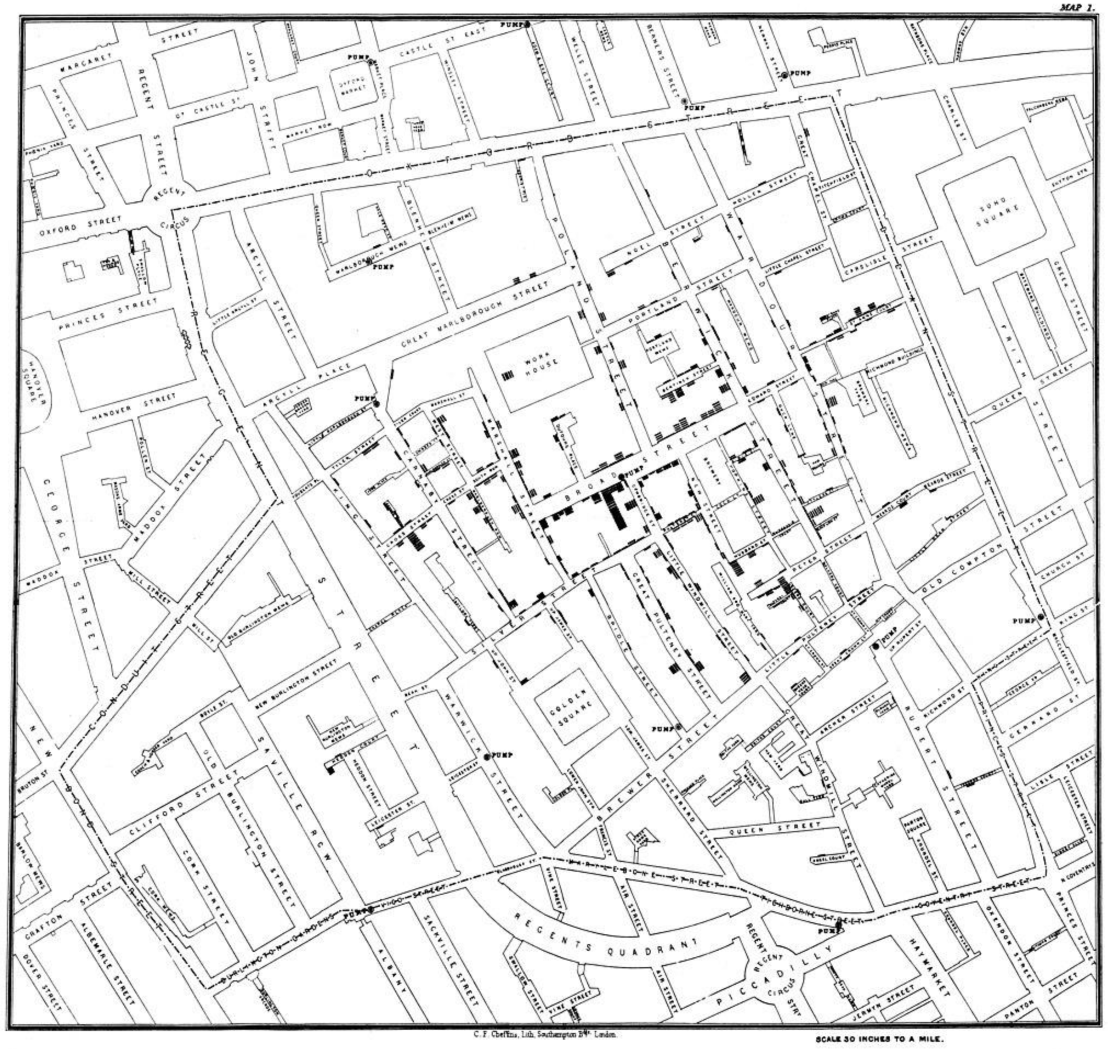
Charles Minard's 1869 chart showing the number of men in Napoleon’s 1812 Russian campaign.



Visual analytics

The John Snow Cholera map:

It uses small bar graphs on city blocks to mark the number of cholera deaths at each household in a London neighbourhood.

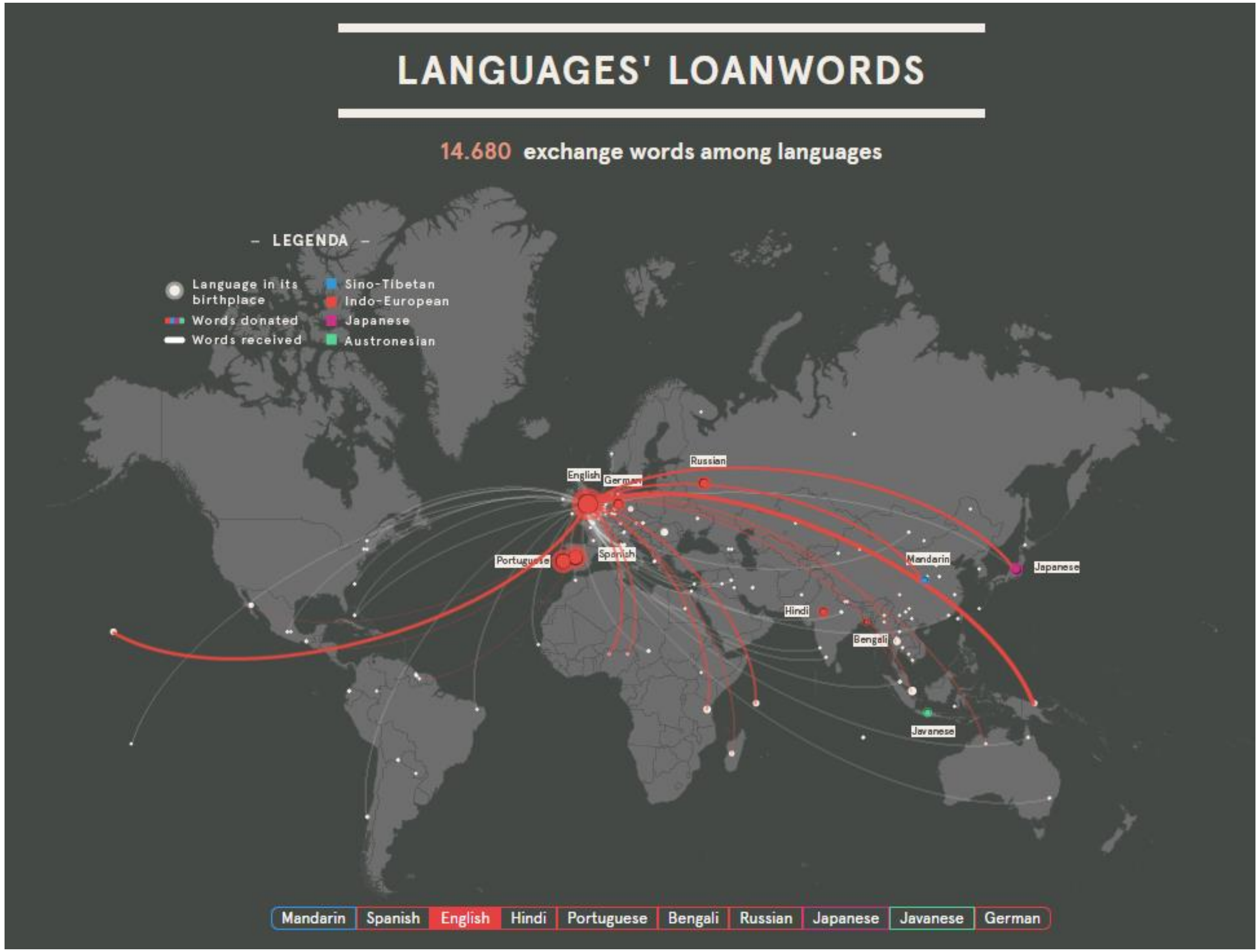


Visual analytics

Living languages in the world.

Visualization of current linguistic situation of the world.

By: puff puff Team



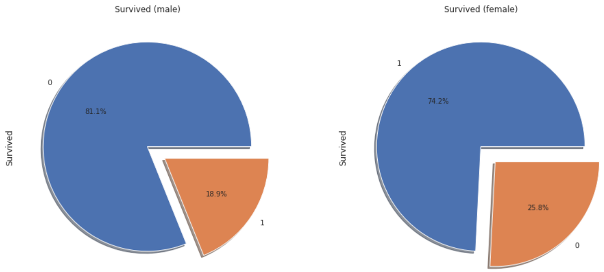
Visual analytics

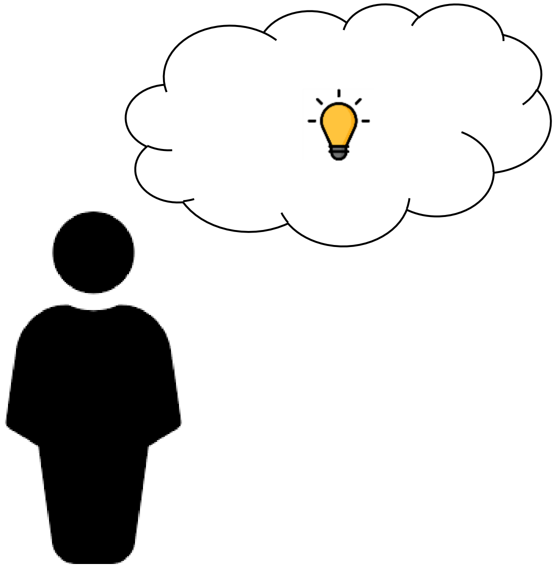
Why Visualization?

- Humans are visual creatures.
- It's an efficient way of communicating concrete and abstract ideas.

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Process
Clean
Visualize





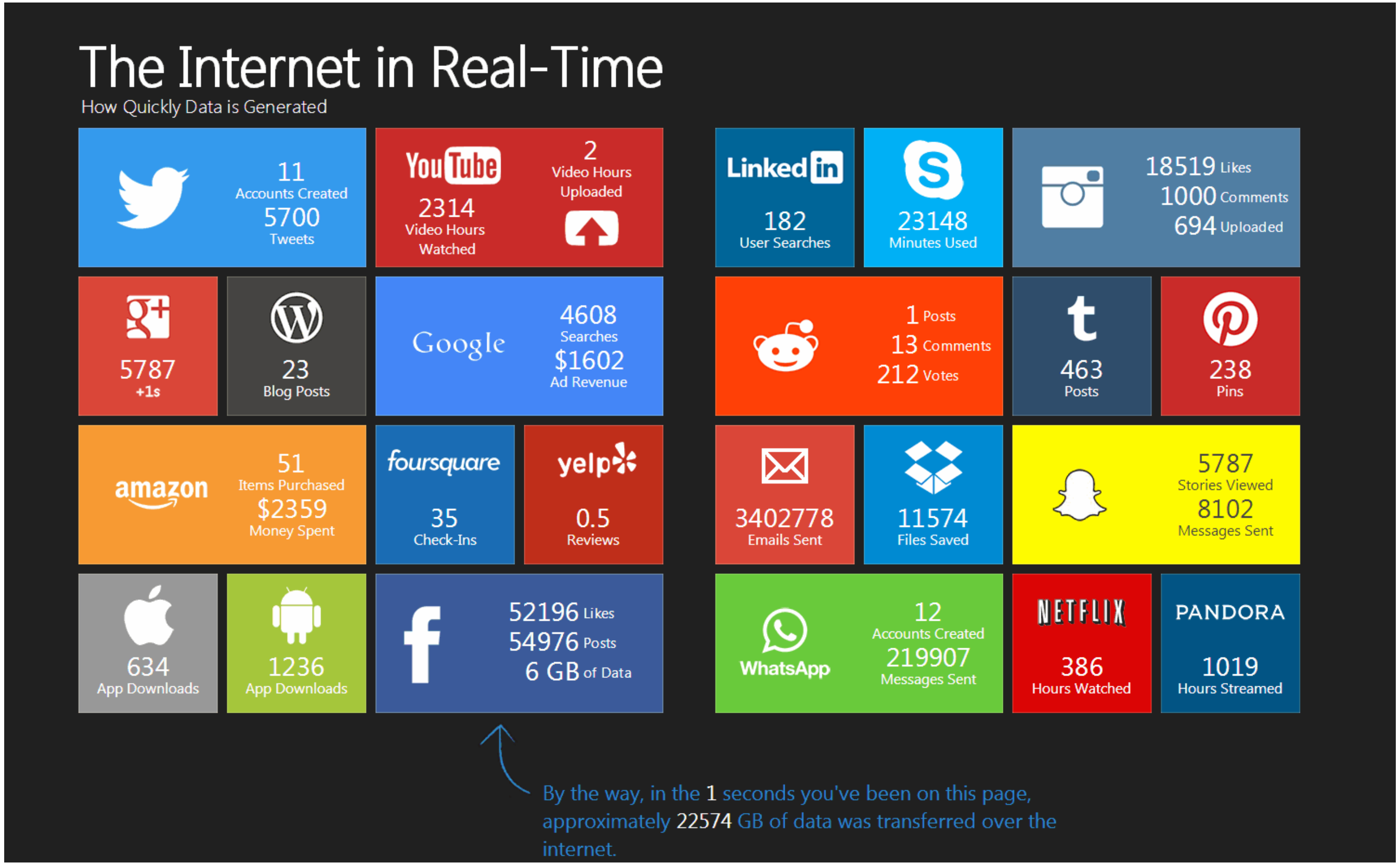
?

Is that all?
What has changed in the world?

Visual analytics

DATA

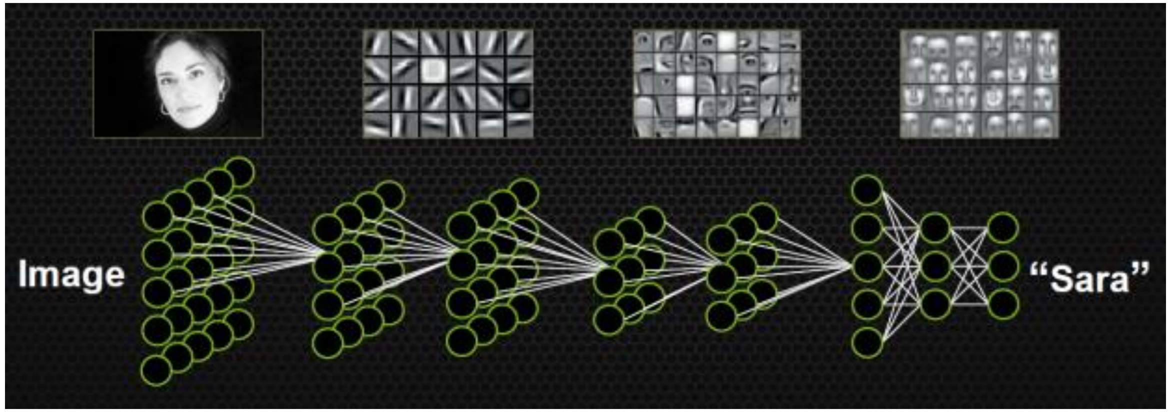
- Our ability to collect and store massive amounts of data has increased exponentially.
- Our ability to analyse it and obtain relevant insights has not.



https://visual.ly/community/infographic/how/internet-real-time

Algorithms

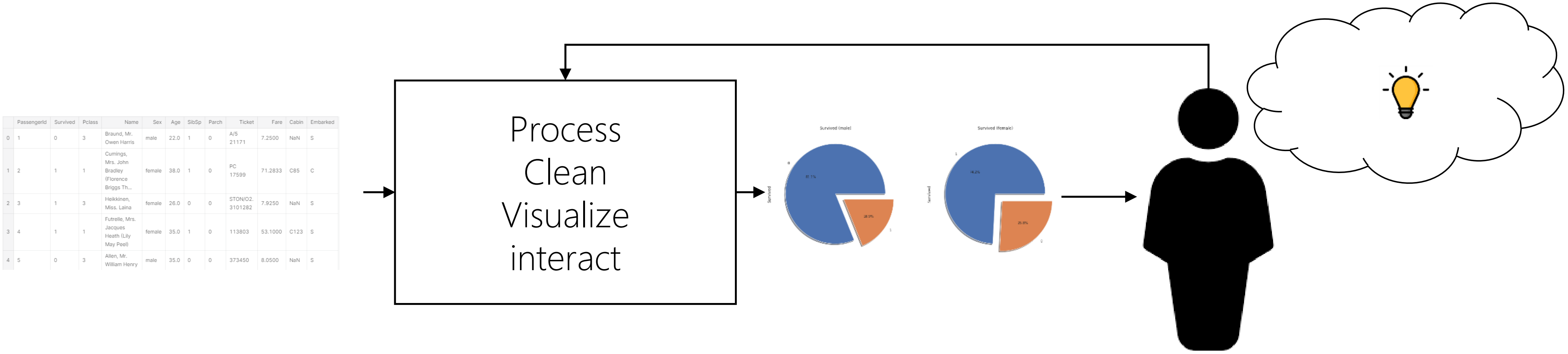
- New black box models currently outperform traditional white box models.
- The pervasive use of new analytic technologies generate questions about trust and accountability within different stakeholders.



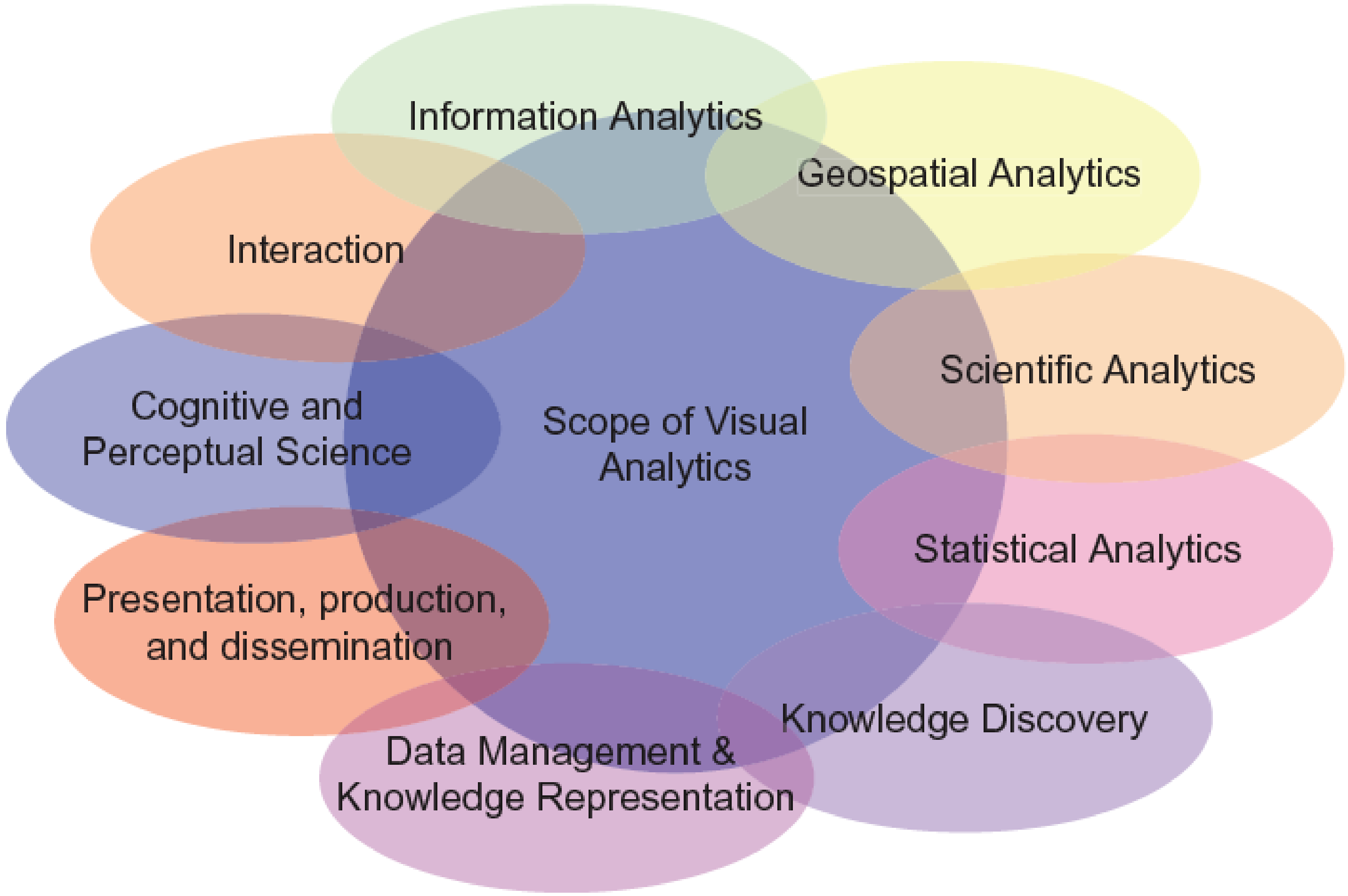
Visual Analytics

Visual analytics

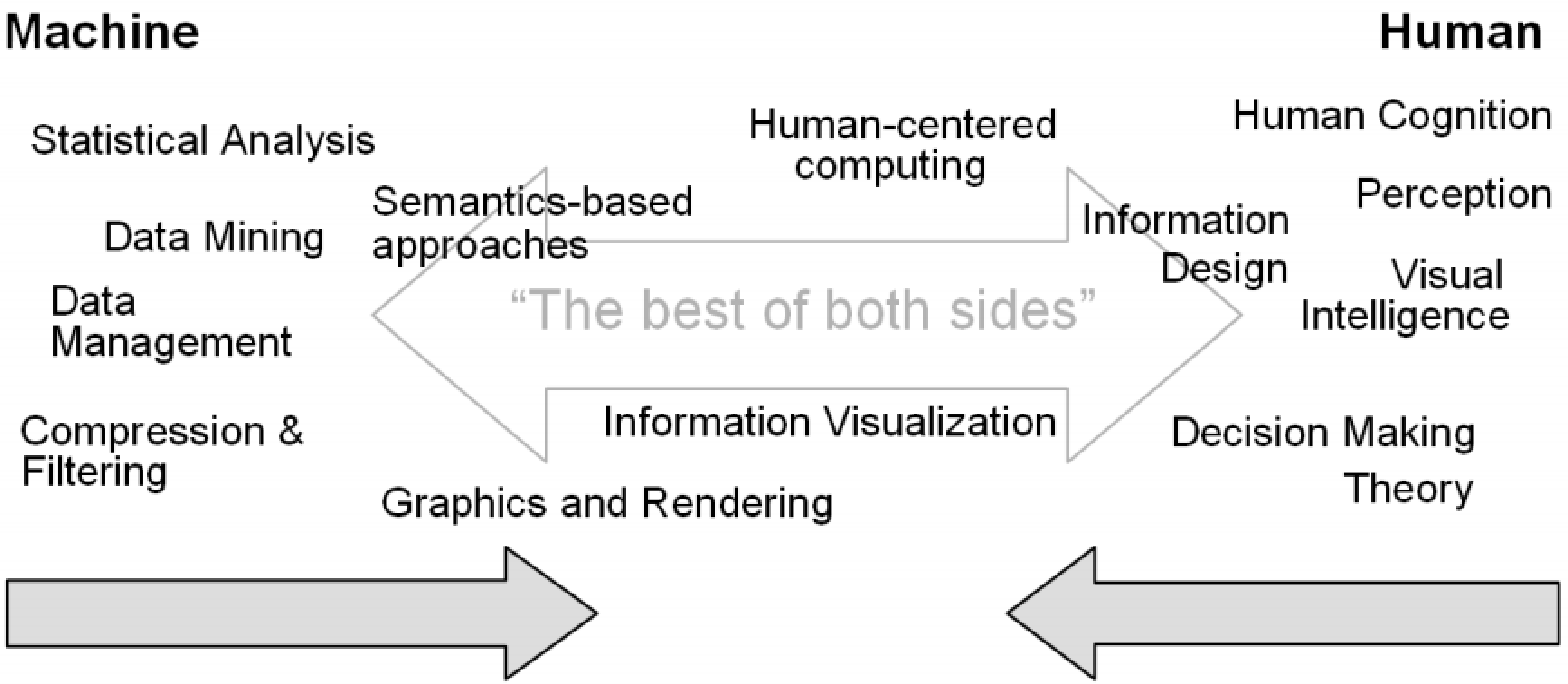
Visual analytics
Refers to the science of analytical reasoning supported by interactive visual interfaces.



Visual analytics



Visual analytics



Visual Analytics: Definition, Process, and Challenges

Visual analytics

Visual analytics

The goal of visual analytics is the creation of tools and techniques to enable people to:

– Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data.

– Detect the expected and discover the unexpected.

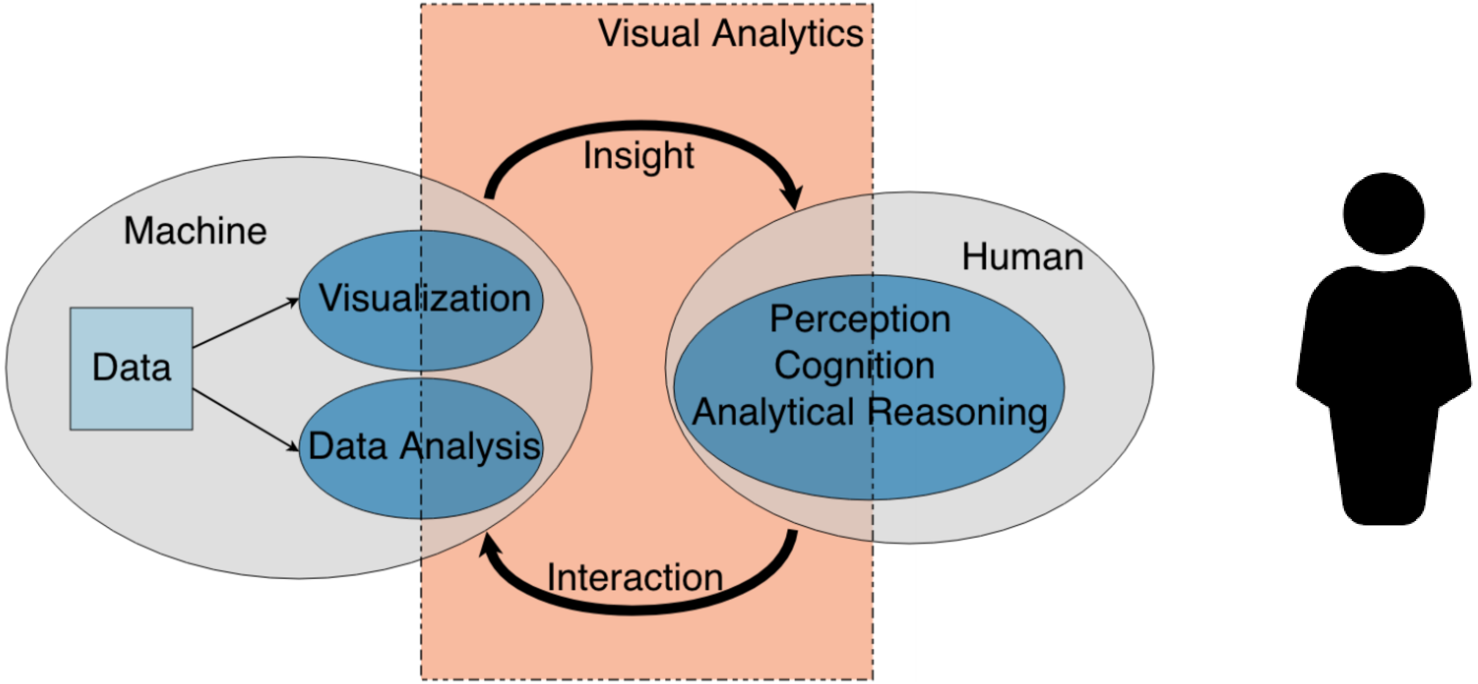
– Provide timely, defensible, and understandable assessments.

– Communicate assessment effectively for action

Visual Analytics: Definition, Process, and Challenges

Visual analytics

Visual analytics
Is a process that combines our visual intelligence and analysis techniques with interactive visualization technologies to get relevant information/insights about data.



Visual Analytics: A Comprehensive Overview

Visual analytics

Visual analytics
Is a process that combines our visual intelligence and analysis techniques with interactive visualization technologies to get relevant information/insights about data.

Step 1 Preprocess (clean, transform, integrate) the data in order to prepare it for further processing.

Step 2 Apply algorithmic analysis methods to the data.

Step 3 Visualize the (processed) data with appropriate visualization techniques.

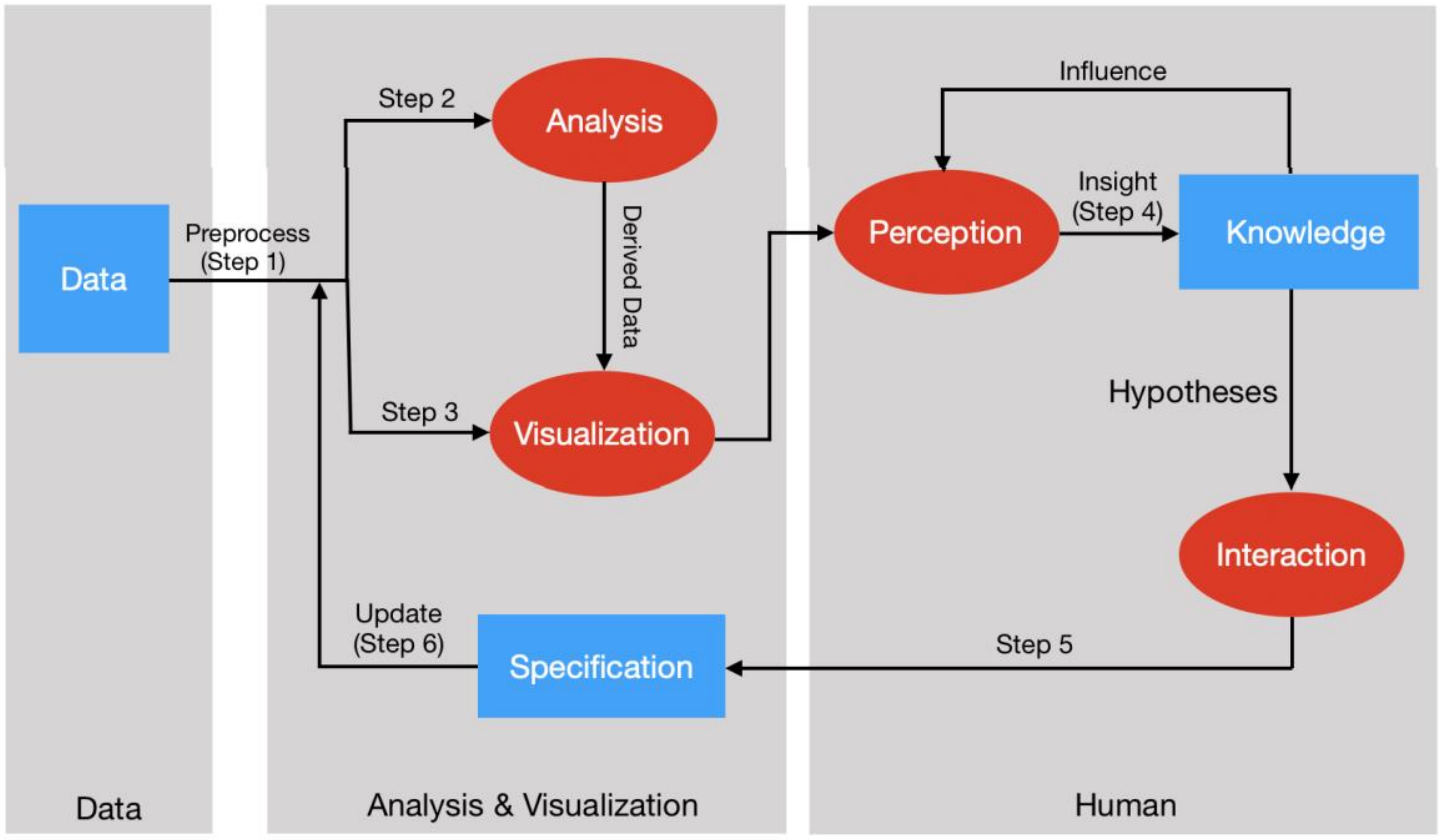
Step 4 Users generate insightful knowledge through human perception, cognition, and reasoning activities.

Step 5 Users make new hypotheses and integrate the newly generated knowledge into the analysis and visualization through interactions.

Step 6 Regenerate an updated visualization based on the interactions to reflect the user's understanding of the data.

Visual Analytics: A Comprehensive Overview

Visual analytics



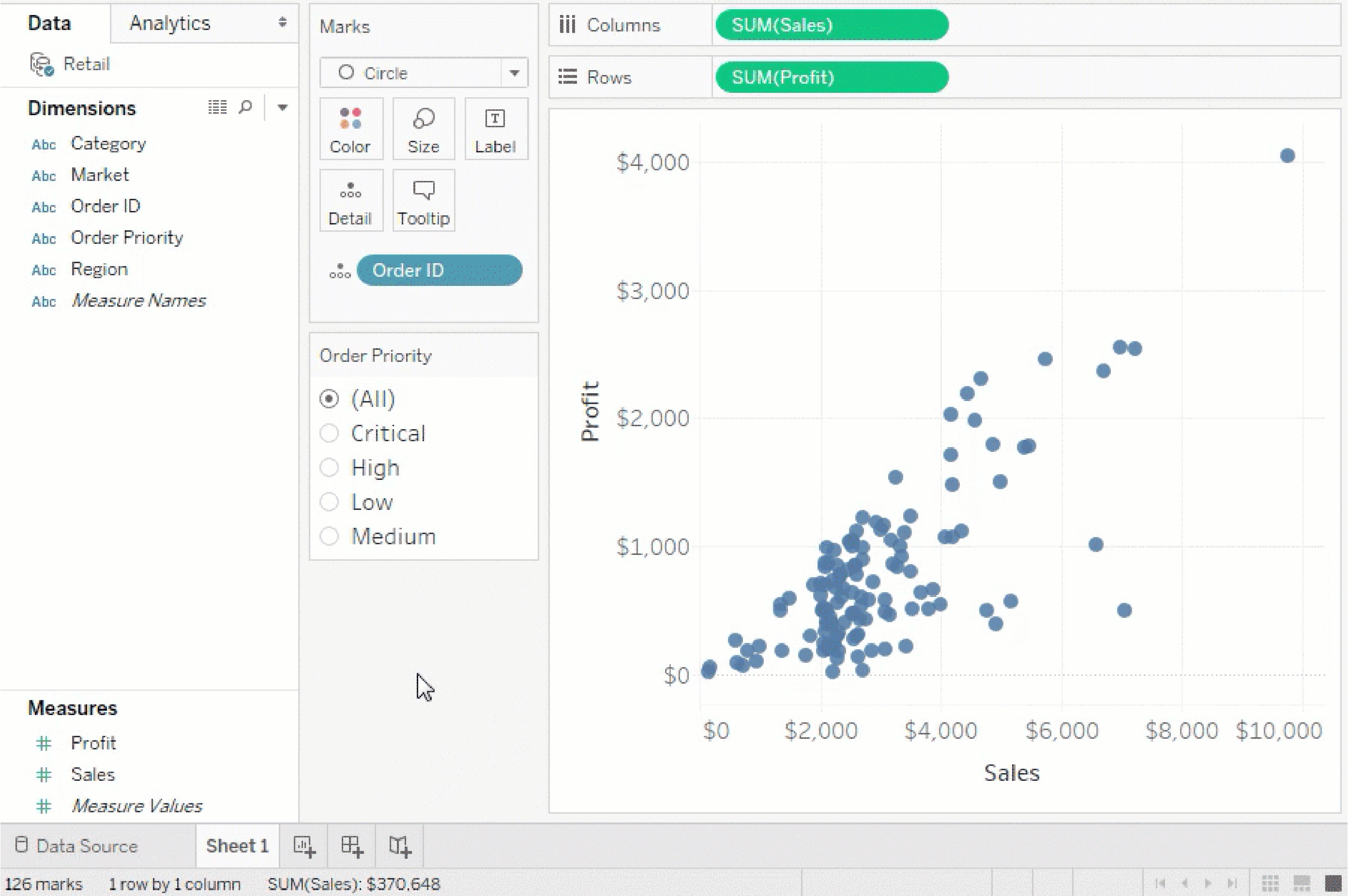
Visual Analytics: A Comprehensive Overview

? How does this look in practice?

Visual analytics

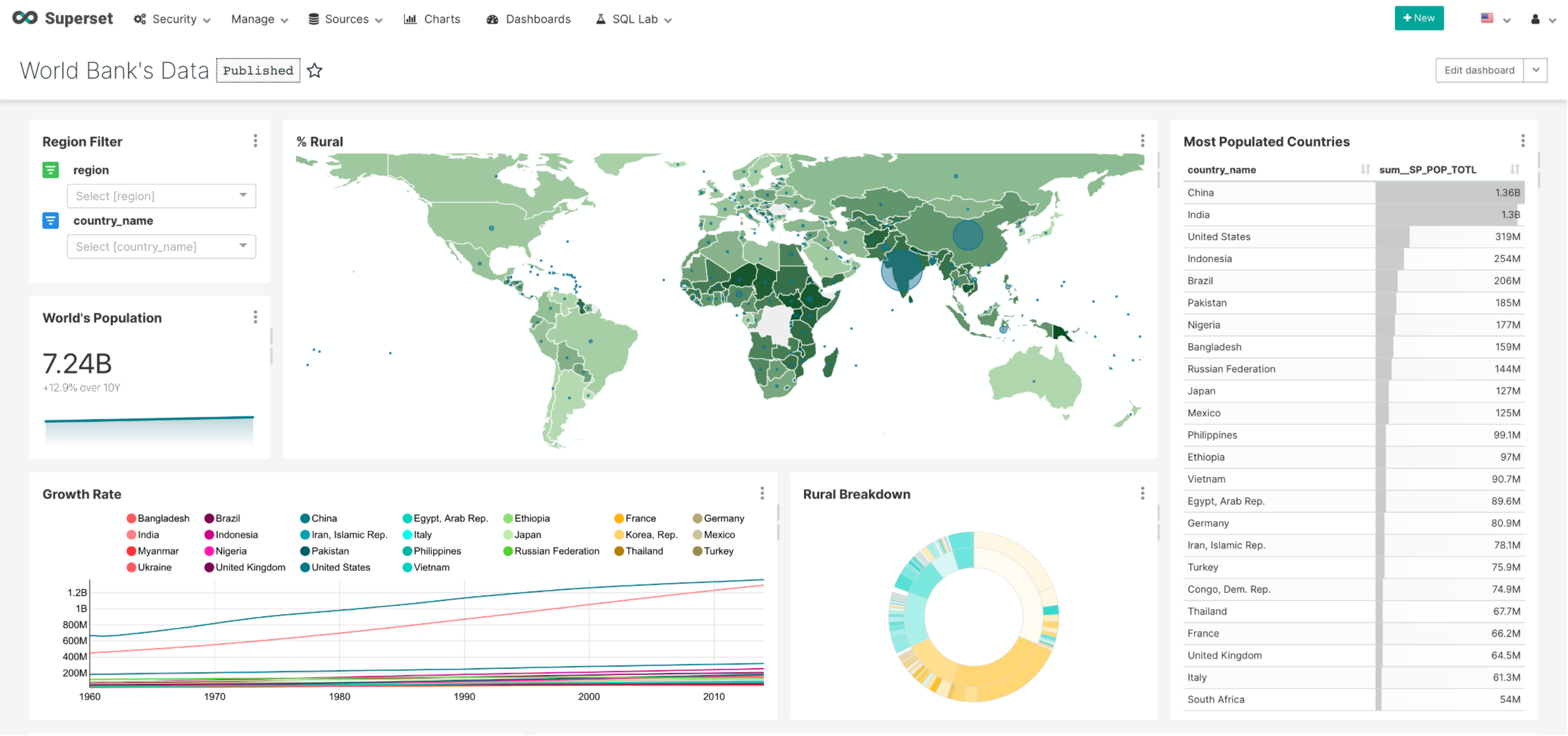
Visual analytics

Many visual analytics tools are also known as Business Intelligence tools.



Visual analytics

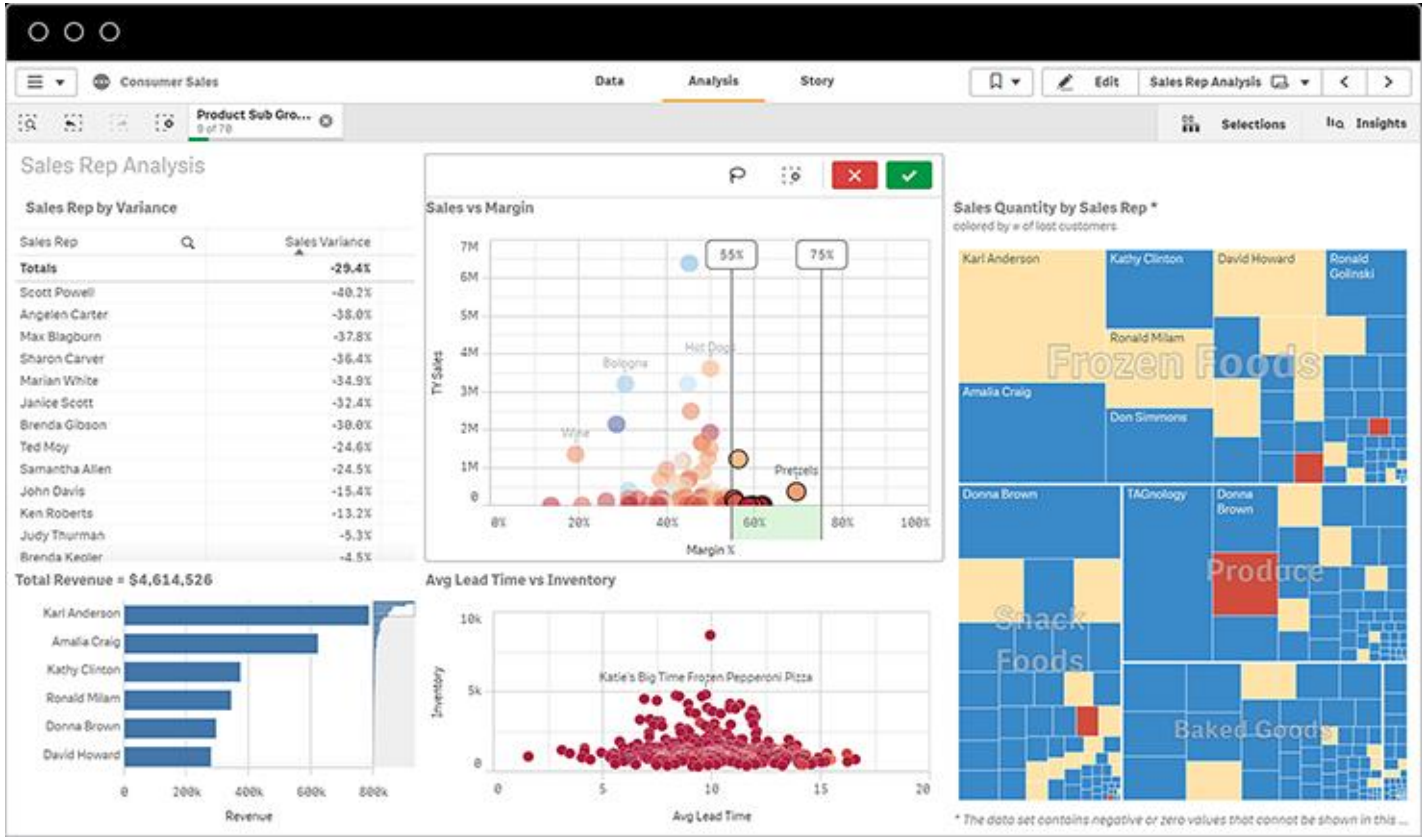
Many visual analytics tools are also known as Business Intelligence tools.



Visual analytics

Many visual analytics tools are also known as Business Intelligence tools.

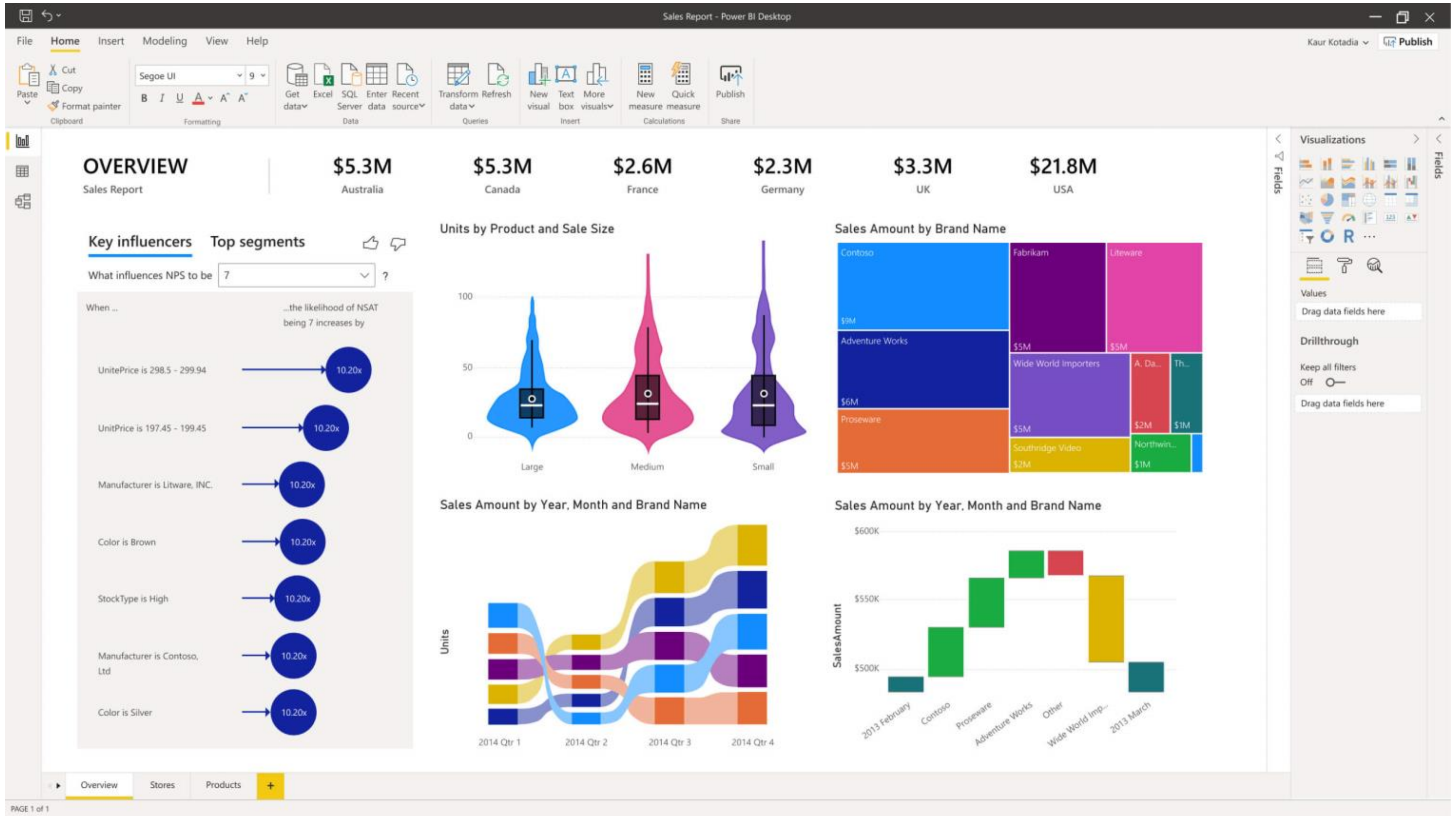
Qlik



Visual analytics

Many visual analytics tools are also known as Business Intelligence tools.

Power BI

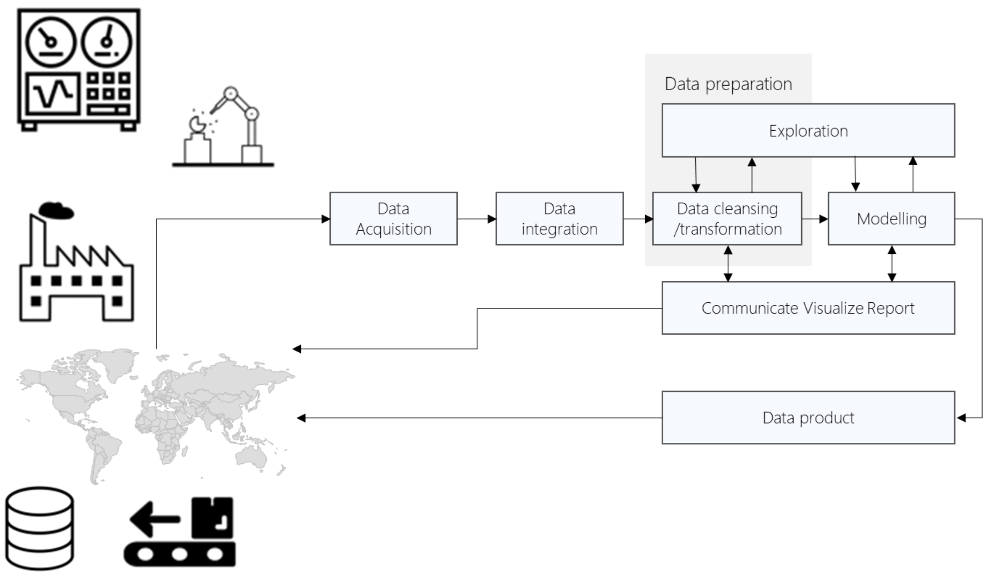


Visual analytics in the Data science pipeline

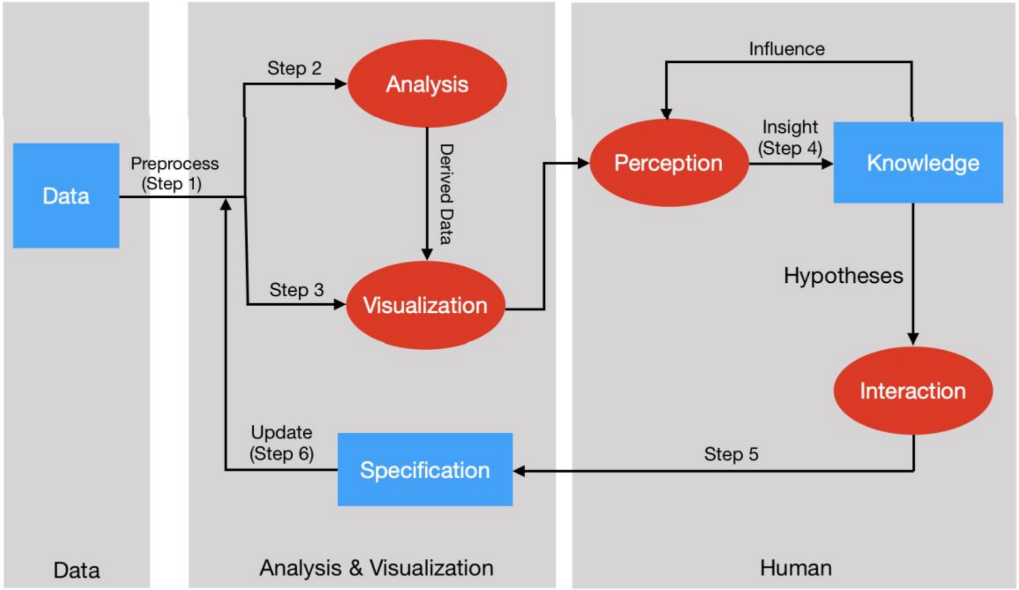
Visual analytics

Visual analytics

Recently, algorithmic data analysis has been combined with interactive visualizations as a mean to involve human judgement in the data analysis process.



?

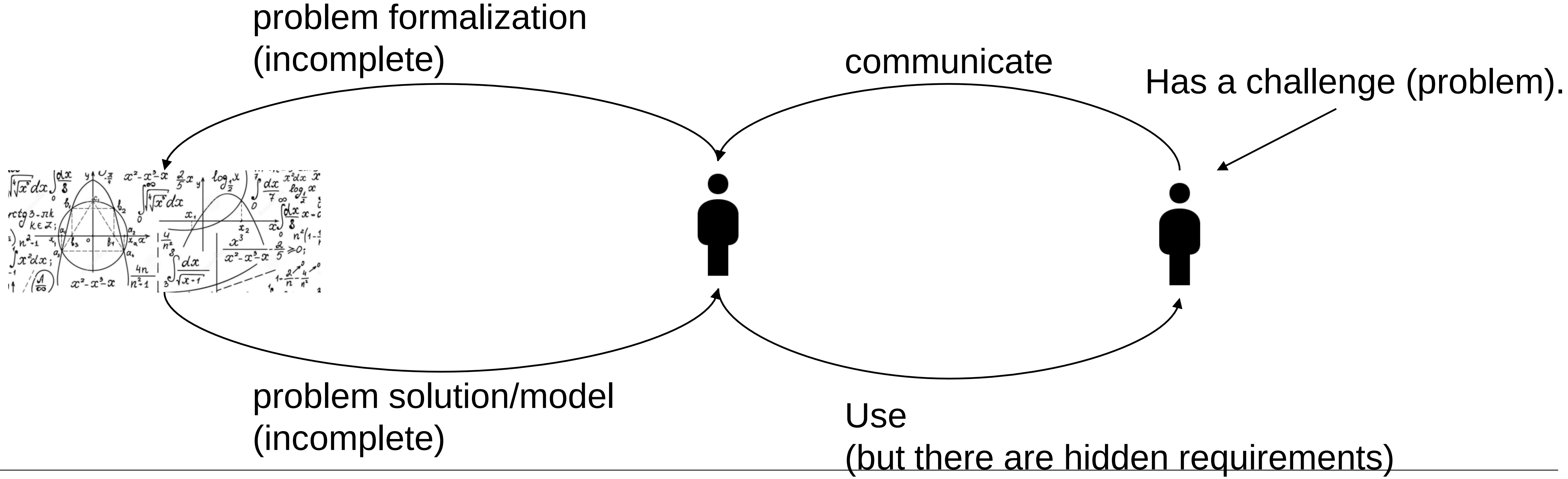


? Why visualizations if we have statistics?

Visual analytics

Incomplete problem formulation

Explanations/Visualizations of data and analytic models are needed in the first place because of the incompleteness of the problem formulation of the methods used for analysis.



Visual analytics

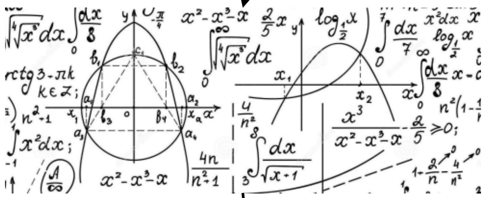
Incomplete problem formulation

Humans have a lot of prior knowledge about the context of a problem.
Complex problems tend to be simplified, leaving aside valuable information.

problem formalization
(incomplete)

communicate

Has a challenge (problem).

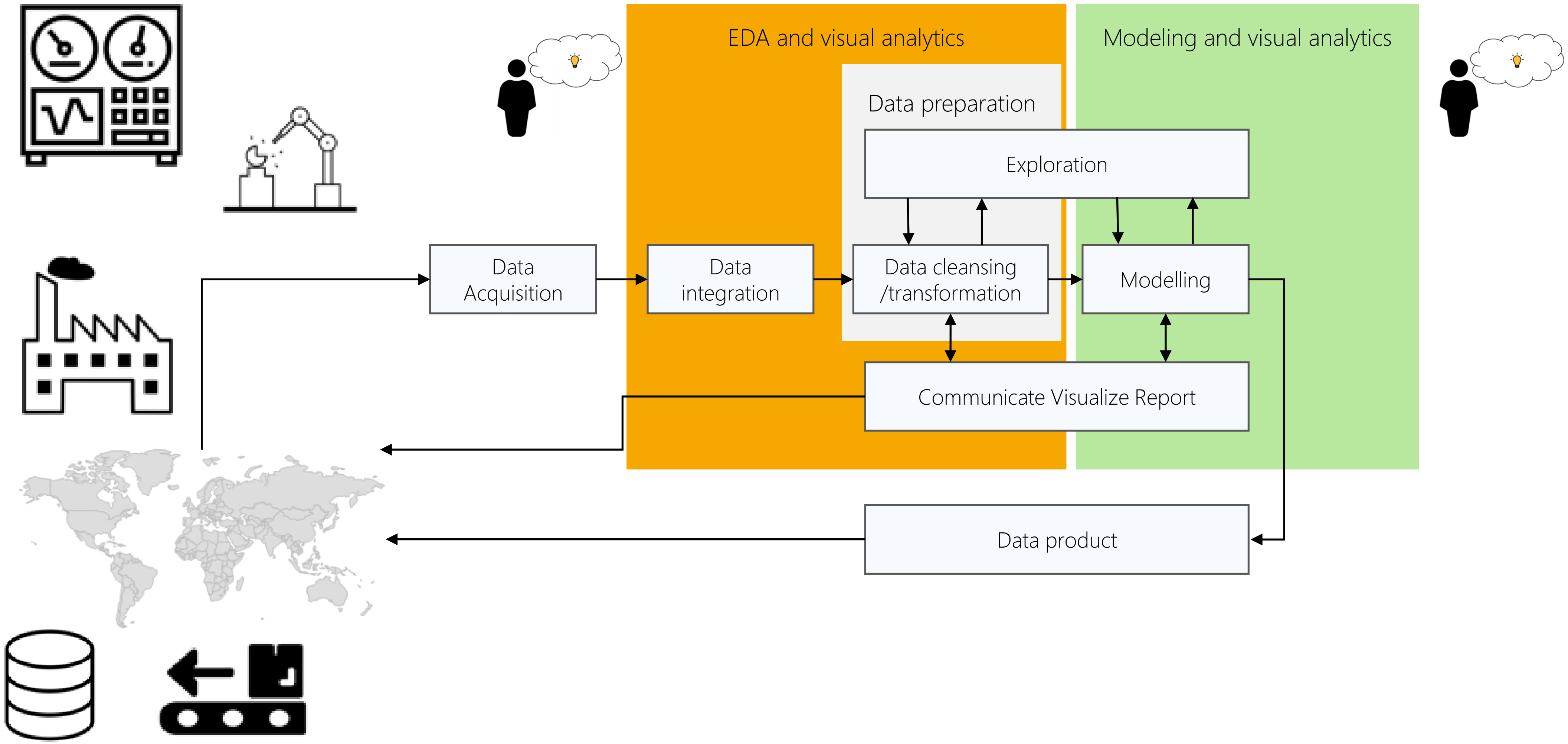


Business knowledge
Common sense
Years of experience

problem solution/model
(incomplete)

Use
(but there are hidden requirements)

Visual analytics



Visual analytics

EDA and visual analytics

Focus:

- Visual representations of raw data.
- Understanding underlying information of the data.

Modeling and visual analytics

Focus:

- Visual representation of model and predictions.
- Understanding the working mechanisms of the models.

EDA and visual analytics

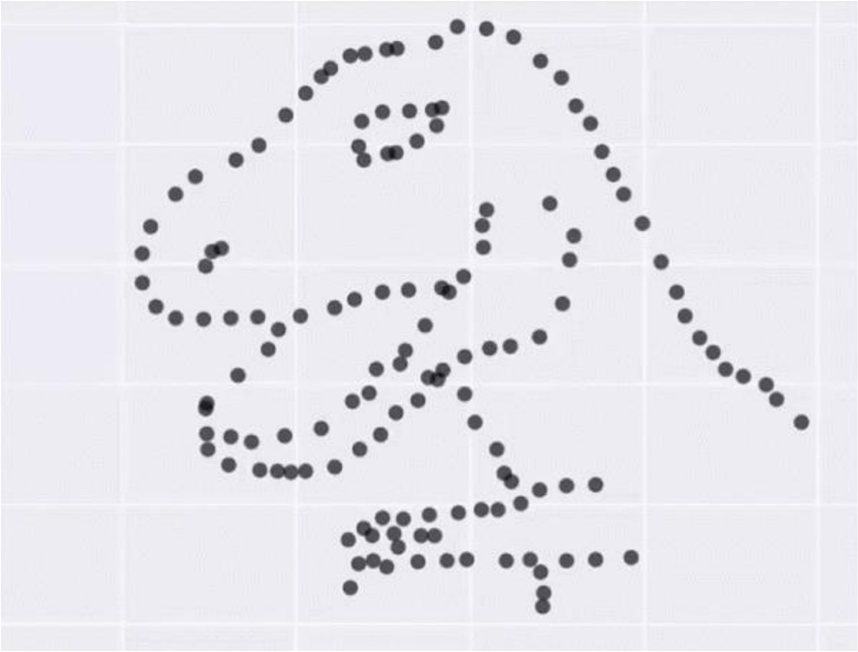
Visual analytics

EDA and visual analytics

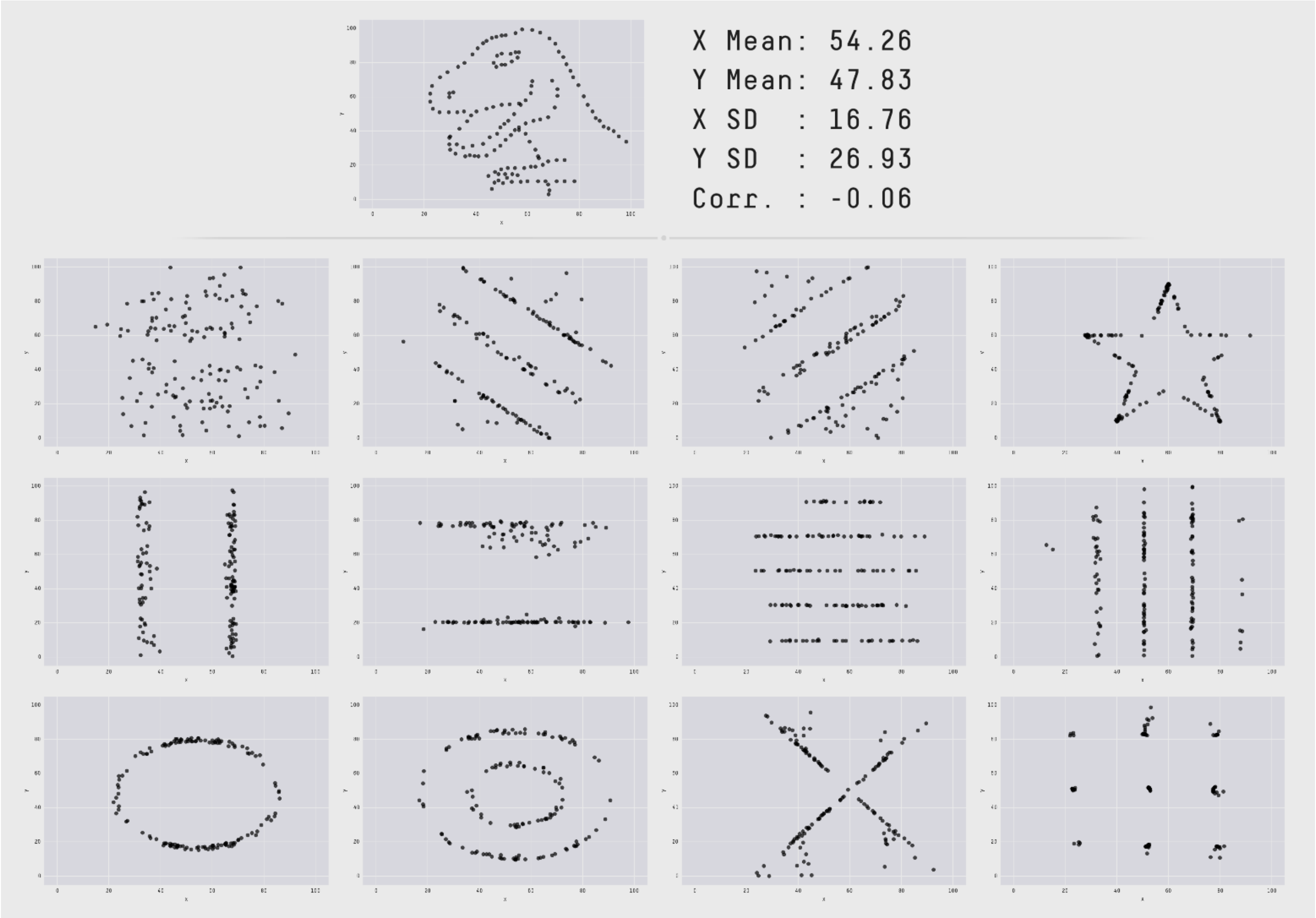
"…make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding."

F. J. Anscombe, 1973

```
X Mean: 54.26
Y Mean: 47.83
X SD  : 16.76
Y SD  : 26.93
Corr. : -0.06
```



Visual analytics



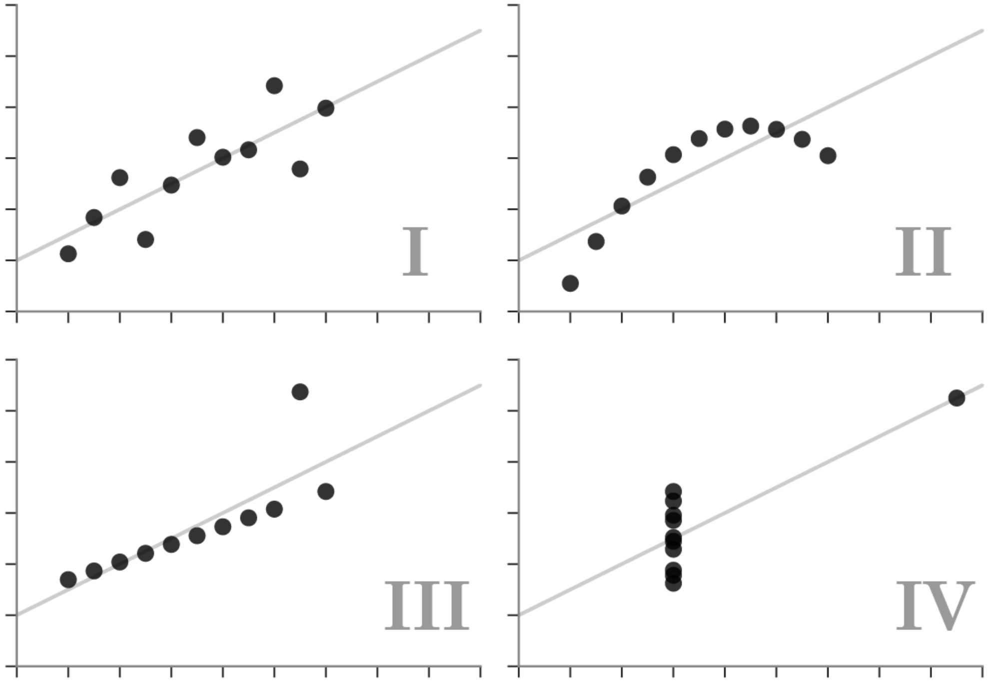
Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing

Visual analytics

- Visual context also has it's limits:

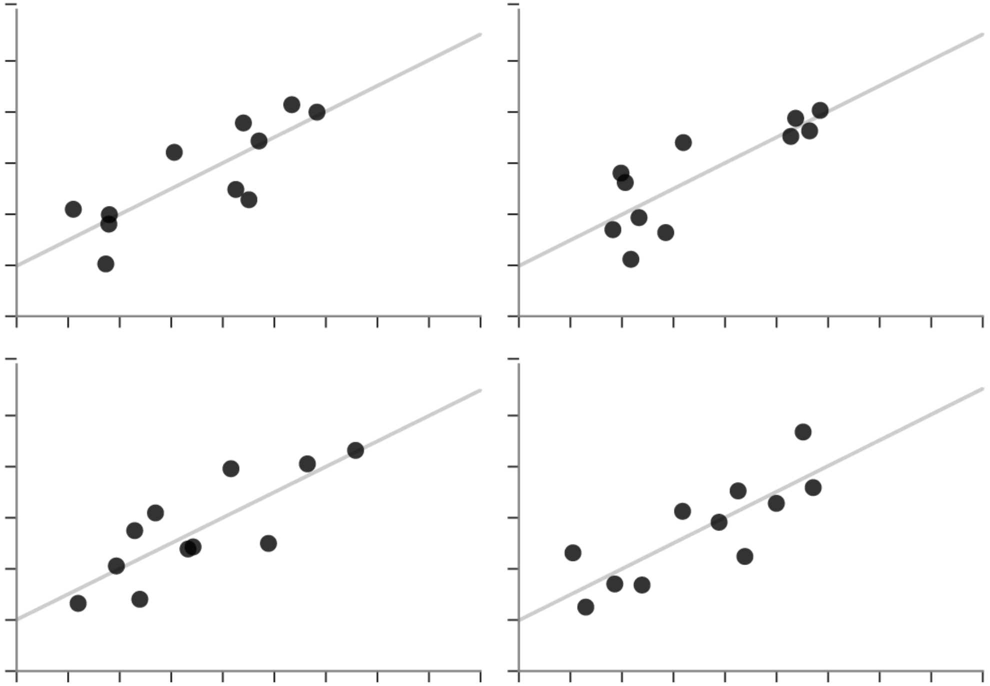
Anscombe's Quartet

Each dataset has the same summary statistics (mean, standard deviation, correlation), and the datasets are *clearly different*, and *visually distinct*.



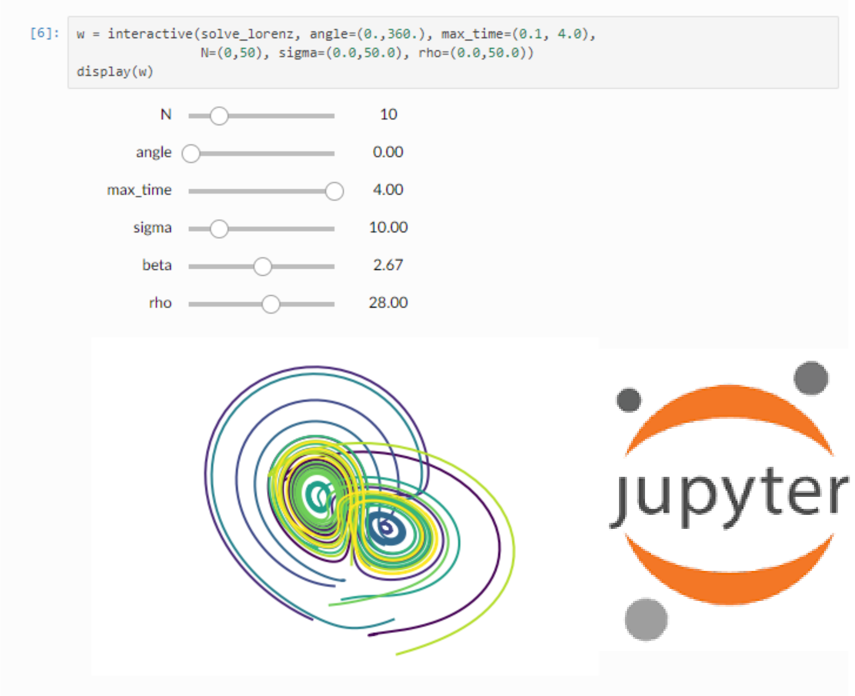
Unstructured Quartet

Each dataset here also has the same summary statistics. However, they are not *clearly different* or *visually distinct*.



Visual analytics

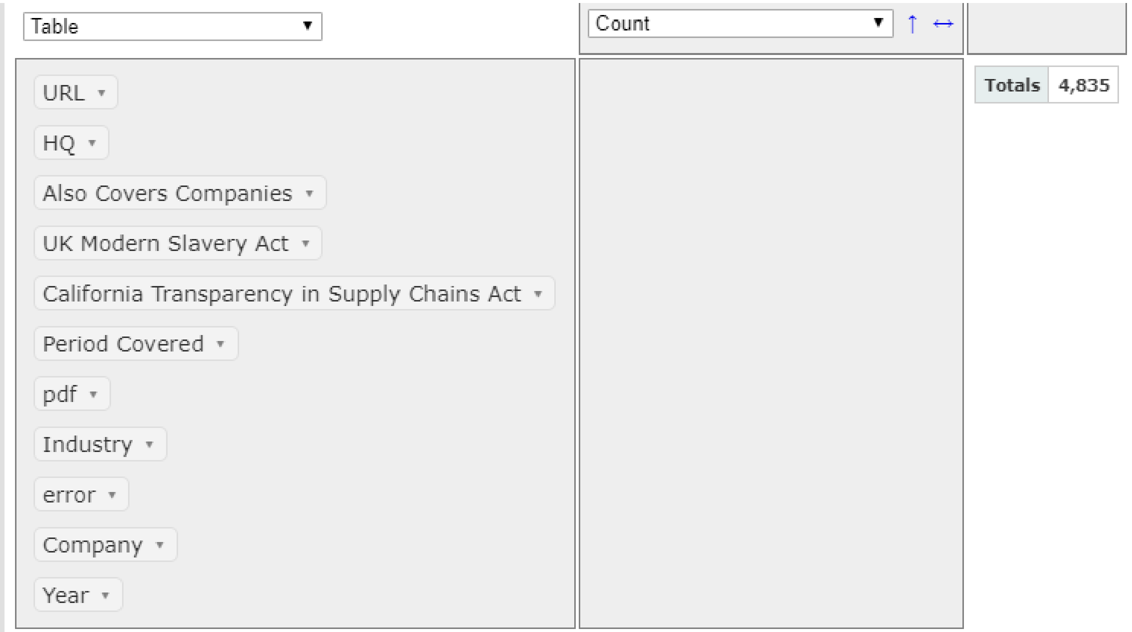
EDA and visual analytics
Explore, visualize and interact



Visual analytics

Example pivottablejs

```
from pivottablejs import pivot_ui
pivot_ui(df, outfile_path='pivottablejs.html')
HTML('pivottablejs.html')
```



Visual analytics

Example qgrid

import qgrid
qgrid.show_grid(df)

In [2]:
```python
import qgrid
qgrid_widget = qgrid.show_grid(df, show_toolbar=True)
qgrid_widget
```

Add Row | Remove Row

index	A	B	C	D	E
0	2013-01-01	-2.0773	washington	foo	✓
1	2013-01-02	-0.55236	adams	bar	
2	2013-01-03	-1.46247	washington	buzz	
3	2013-01-04	0.85559	madison	bippity	
4	2013-01-05	-1.66466	lincoln	boppity	
5	2013-01-06	0.85659	jefferson	foo	✓
6	2013-01-07	1.23665	hamilton	foo	✓
7	2013-01-08	-0.36515	roosevelt	bar	
8	2013-01-09	0.89955	kennedy	zoo	

In [3]:
```python
qgrid_widget.get_changed_df()
```

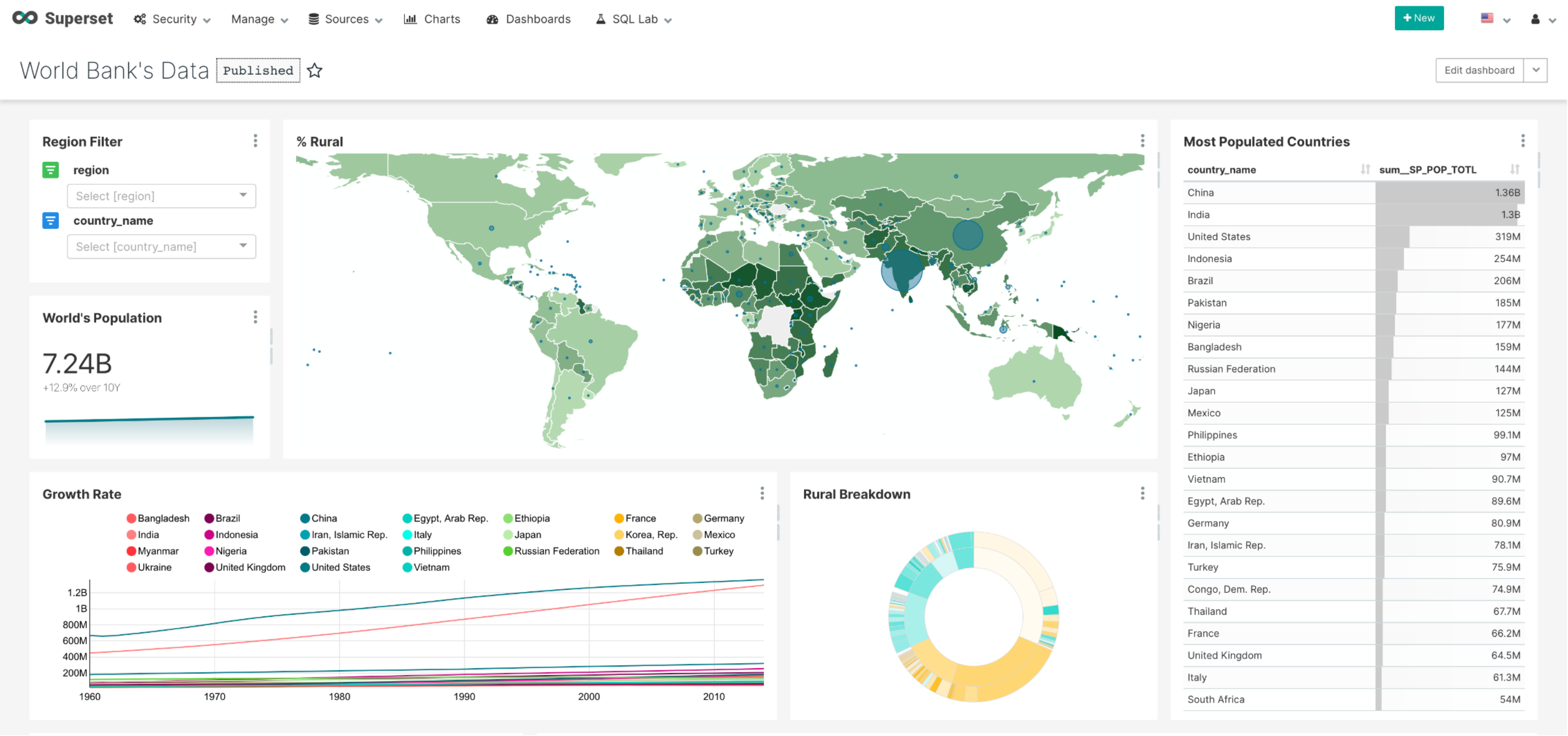
Out[3]:

	A	B	C	D	E
0	2013-01-01	-2.077295	washington	foo	True
1	2013-01-02	-0.552359	adams	bar	False
2	2013-01-03	-1.462471	washington	buzz	False
3	2013-01-04	0.855593	madison	bippity	False
4	2013-01-05	-1.664660	lincoln	boppity	False
5	2013-01-06	0.856594	jefferson	foo	True
6	2013-01-07	1.236655	hamilton	foo	True
7	2013-01-08	-0.365152	roosevelt	bar	False
8	2013-01-09	0.899548	kennedy	zoo	False

Visual analytics

Or, export your data tables and use a prebuilt alternative

Superset



IfU IMA RWTH AACHEN UNIVERSITY

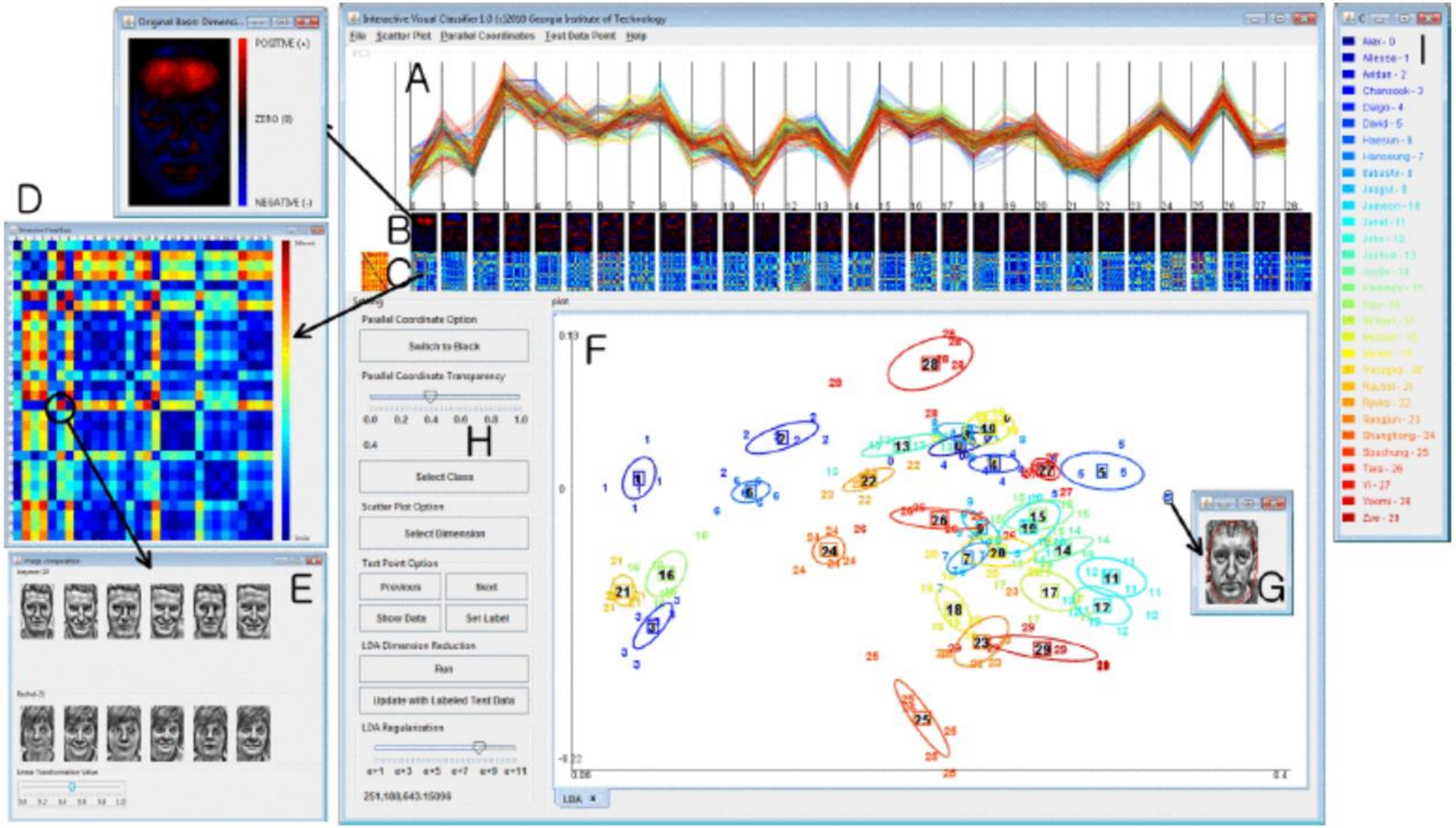
Applications and challenges

Visual analytics

Visual analytics
Visual analytics is an application oriented discipline driven by practical requirements in important domains.

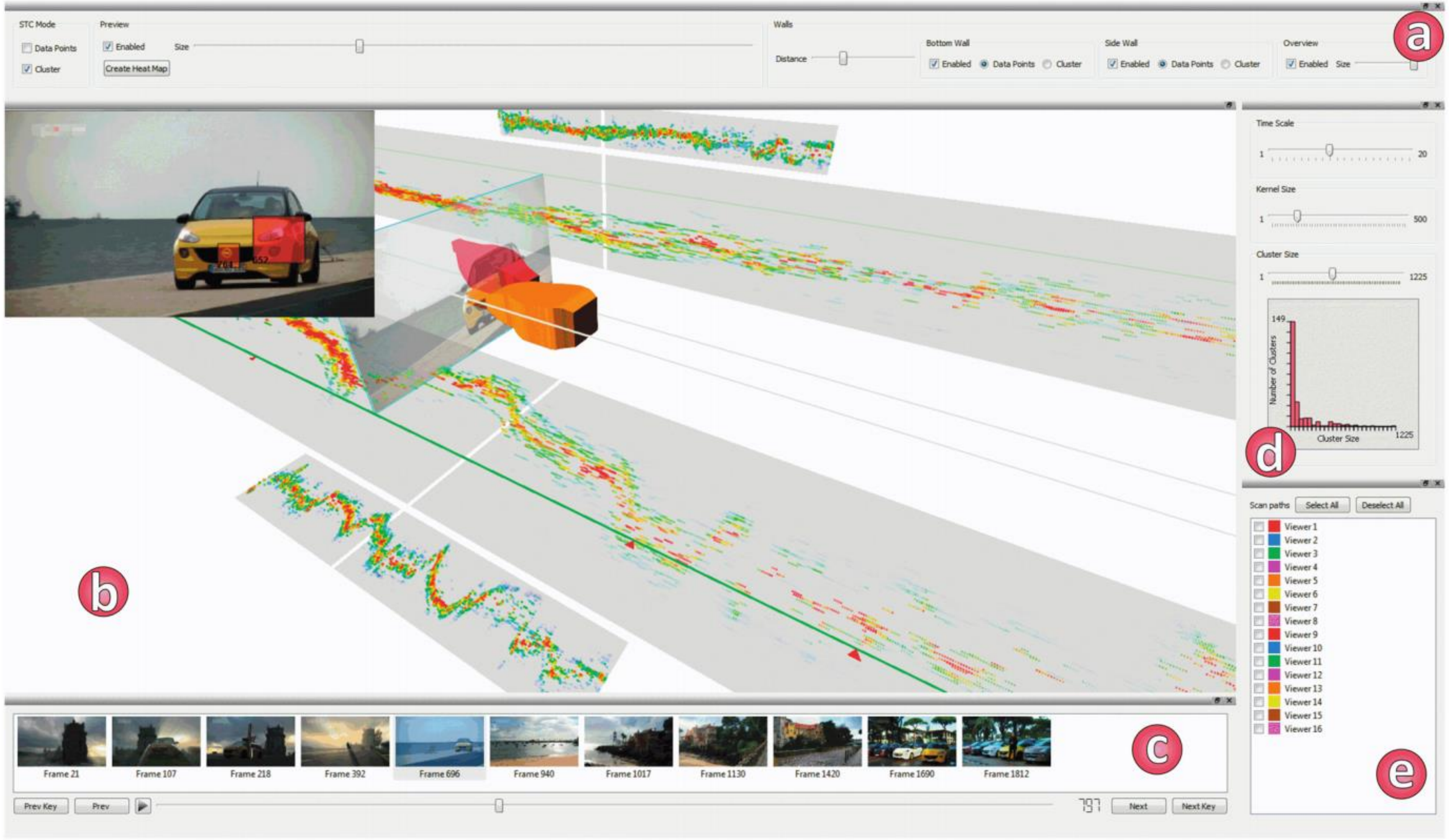
- Speed up development.
- Improve efficiency with the discovery of hidden information.
- Enables the understanding of high volumes of data.
- Enables the understanding of highly heterogeneous information sources (Ex: Financial analysis)
- User friendly interactive visualizations often help in high stack decision making processes (government)

Visual analytics



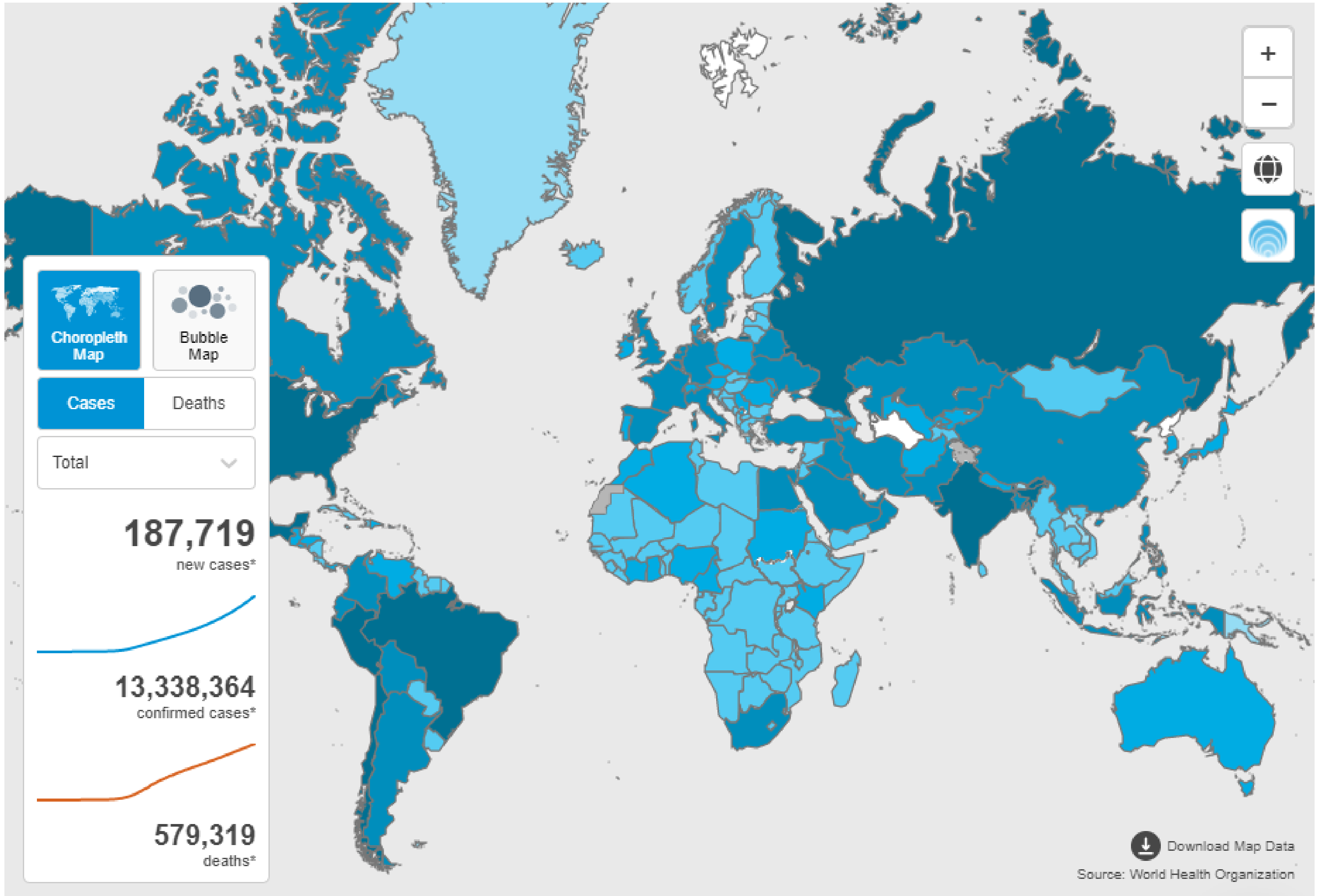
iVisClassifier: An interactive visual analytics system for classification based on supervised dimension reduction

Visual analytics

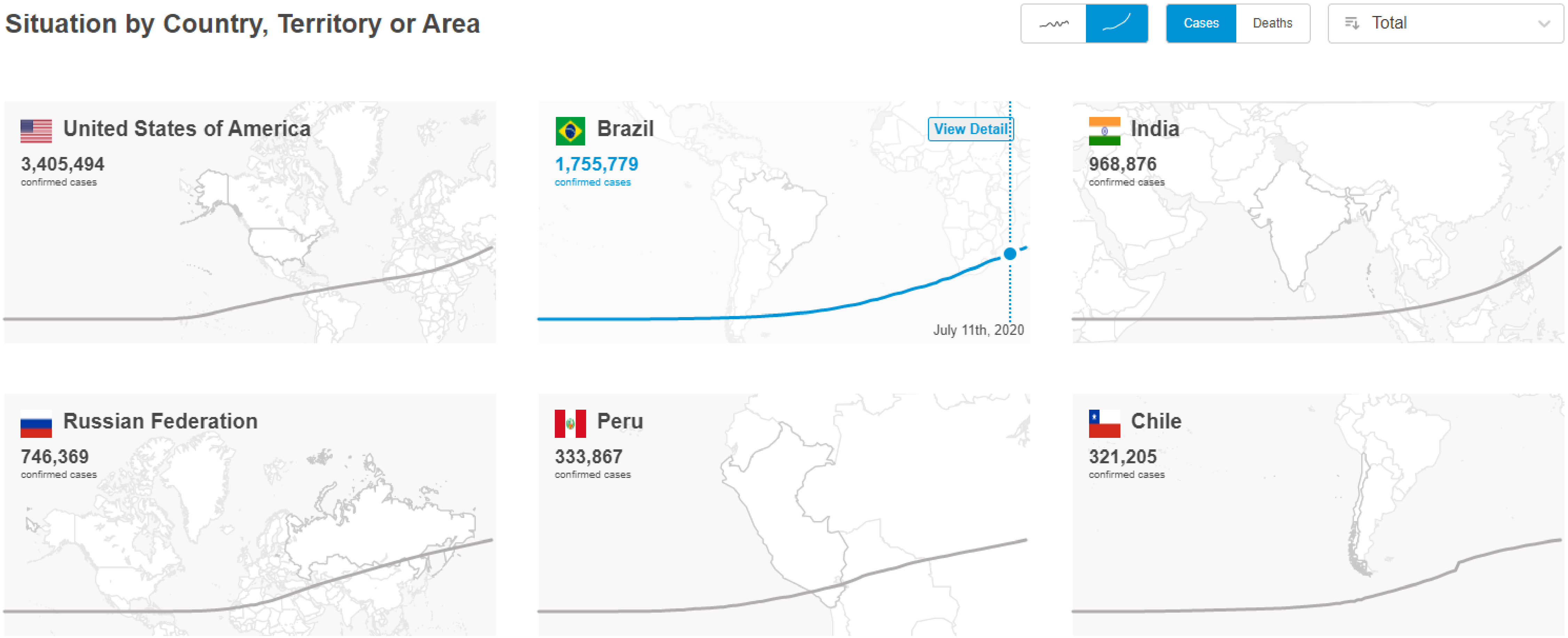


multi-dimensional-to-3D visual-analytics application for eye-tracking data

Visual analytics



Visual analytics



Visual analytics

Challenge: Scalability with Data Volumes and Data Dimensionality

- How to represent high dimensional data?
- Data streams details vs bandwidth
- Curse of dimensionality?

Visual analytics

Challenge: Quality of Data and Graphical Representation

- How does noise affect the data visualizations?
- Which representations to use? (avoid misleading representations)
- How to represent data quality and confidence ?

Visual analytics

Challenge: Visual Representation and Level of Detail

- How to select the correct details to visualize?
- Which patterns are relevant and which ones are not?
- Global vs local importance?

Visual analytics

Challenge: User Interfaces, and Interaction Styles and Metaphors

- How to help the user focus in the task?
- Complexity vs simplicity?
- Limited interaction

Challenge: Display Devices

- Device dependent workflows?
- Limited resolution?
- Perception vs device limitations?

Visual analytics

Challenge: Evaluation

- How to evaluate the goodness of a process?
- Evaluation involving human factors?

Visual analytics

Challenge: Infrastructure

- Big data vs responsiveness?
- Real time analysis vs complexity?
- Synchronous vs asynchronous response?

Modeling and visual analytics

Visual analytics

Democratization of AI.

Great power

Stanford
Artificial Intelligence Laboratory

RWTH AACHEN UNIVERSITY

fast.ai

Frameworks

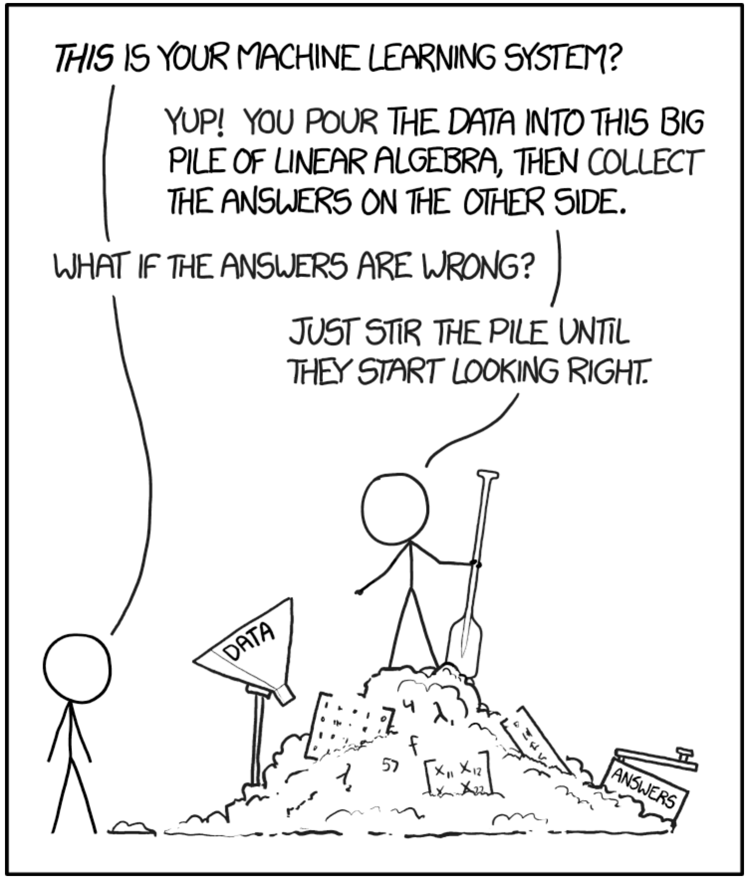
K + TensorFlow   PYTORCH   scikit learn   Amazon SageMaker   AI solutions

Diabetes detection

Self-driving cars

Visual analytics

Deep learning algorithms do what you tell them, not what you want them to do.



Visual analytics

Legislation

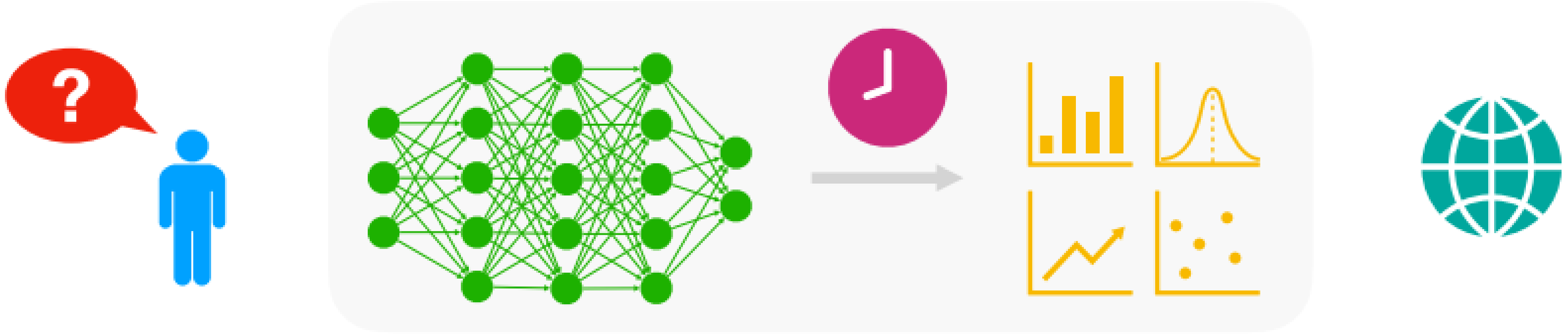
- GDPR (articles 13-15):
  - if there is an “**automated decision-making**”,
  - Then “**meaningful information about the logic involved**” must be provided.

XAI

- Human-AI systems:
  - Interactions, Evaluation, mental models, explanation goodness…..
- Visualization:
  - Visual analytics, knowledge discovery,
- Interpretability:
  - Explaining data.
  - Post training explanations.
  - Building inherently interpretable models.

Visual analytics

Modeling and visual analytics
Because of the internal complexity of models, the underlying decision making processes for why these models are achieving such performance are challenging and sometimes mystifying to interpret.



Visual analytics

Modeling and visual analytics
Because of the internal complexity of models, the underlying decision making processes for why these models are achieving such performance are challenging and sometimes mystifying to interpret.

§4 WHY

Why would one want to use visualization in deep learning?

Interpretability & Explainability
Debugging & Improving Models
Comparing & Selecting Models
Teaching Deep Learning Concepts

§6 WHAT

What data, features, and relationships in deep learning can be visualized?

Computational Graph & Network Architecture
Learned Model Parameters
Individual Computational Units
Neurons In High-dimensional Space
Aggregated Information

§8 WHEN

When in the deep learning process is visualization used?

During Training
After Training

Visual analytics

Modeling and visual analytics
Because of the internal complexity of models, the underlying decision making processes for why these models are achieving such performance are challenging and sometimes mystifying to interpret.

§5 WHO

Who would use and benefit from visualizing deep learning?

Model Developers & Builders
Model Users
Non-experts

§7 HOW

How can we visualize deep learning data, features, and relationships?

Node-link Diagrams for Network Architecture
Dimensionality Reduction & Scatter Plots
Line Charts for Temporal Metrics
Instance-based Analysis & Exploration
Interactive Experimentation
Algorithms for Attribution & Feature Visualization

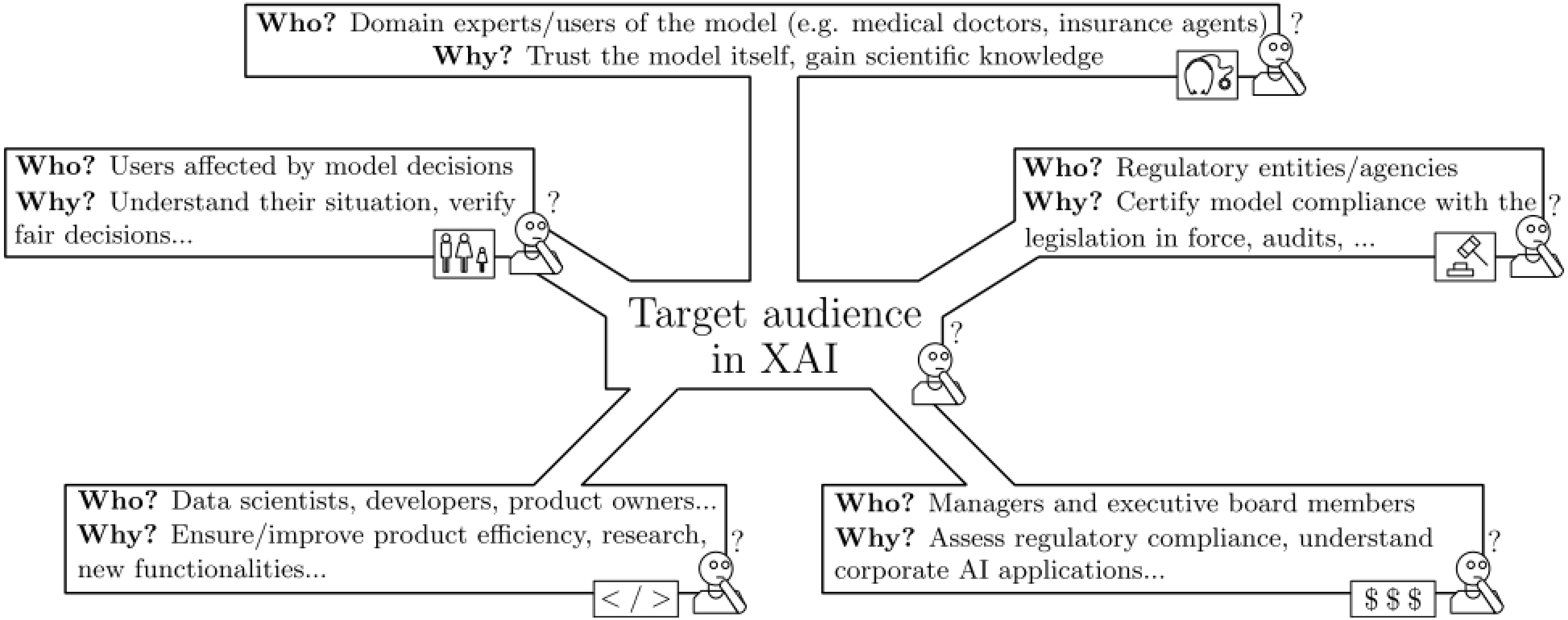
§9 WHERE

Where has deep learning visualization been used?

Application Domains & Models
A Vibrant Research Community

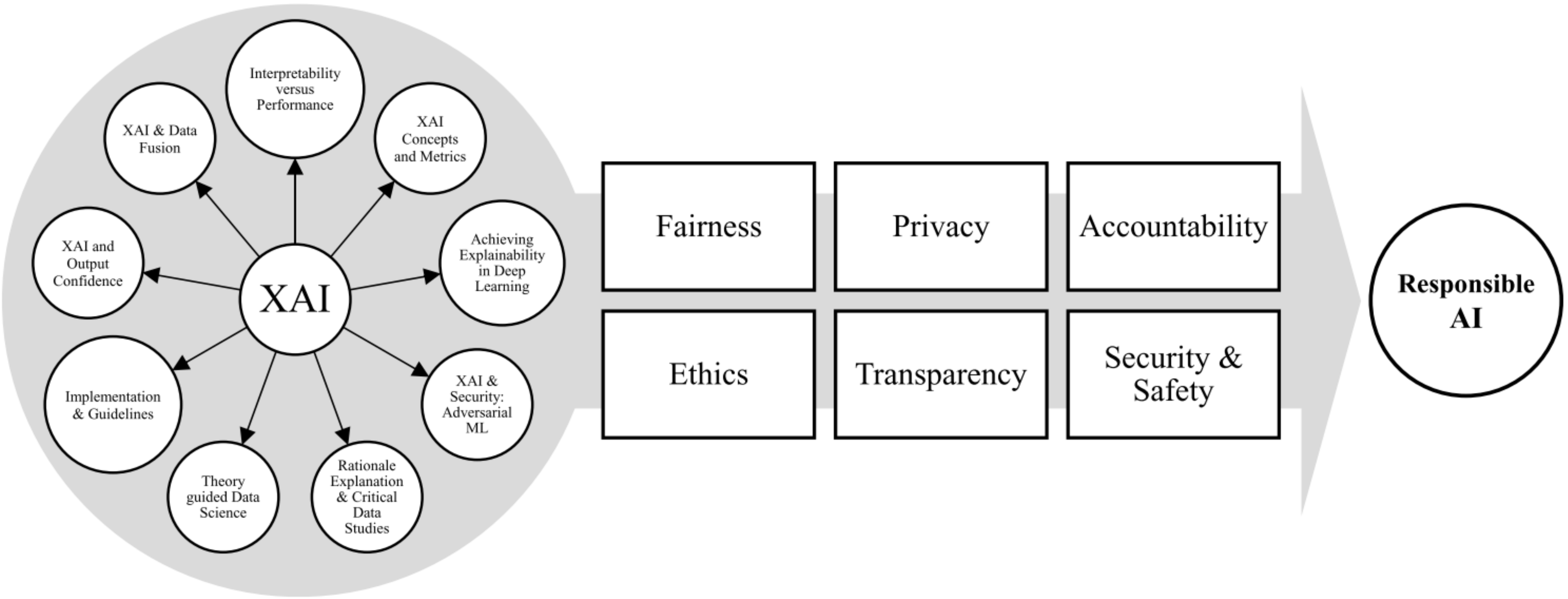
Visual analytics

Target audience



Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI

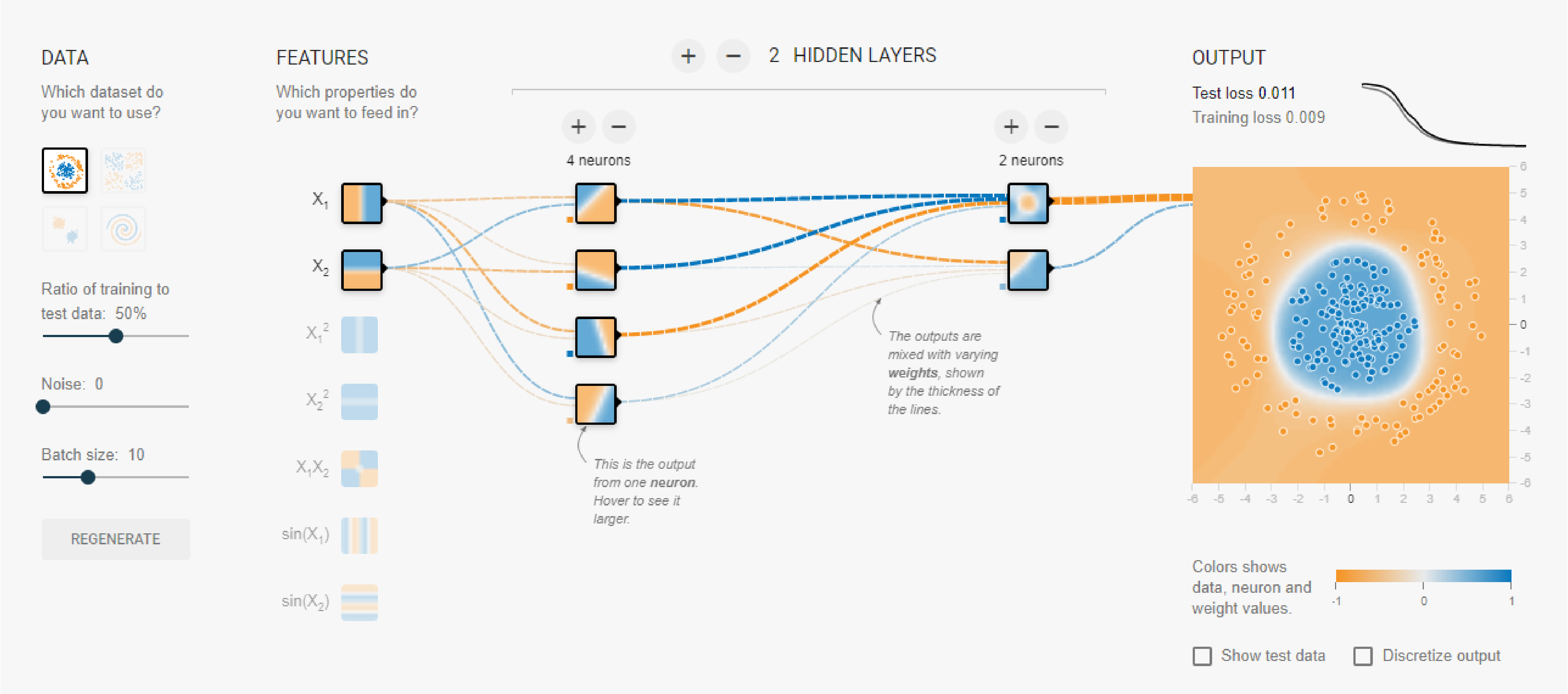
Target audience



Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI

Visual analytics

Example: Tensorflow playground



Example: Activation atlas

Visual analytics

To use machine learning responsibly, we need to ensure that:

1. Our **values** are aligned.
2. Our **knowledge** is reflected.

How? By using **interpretability** methods.

Explaining data

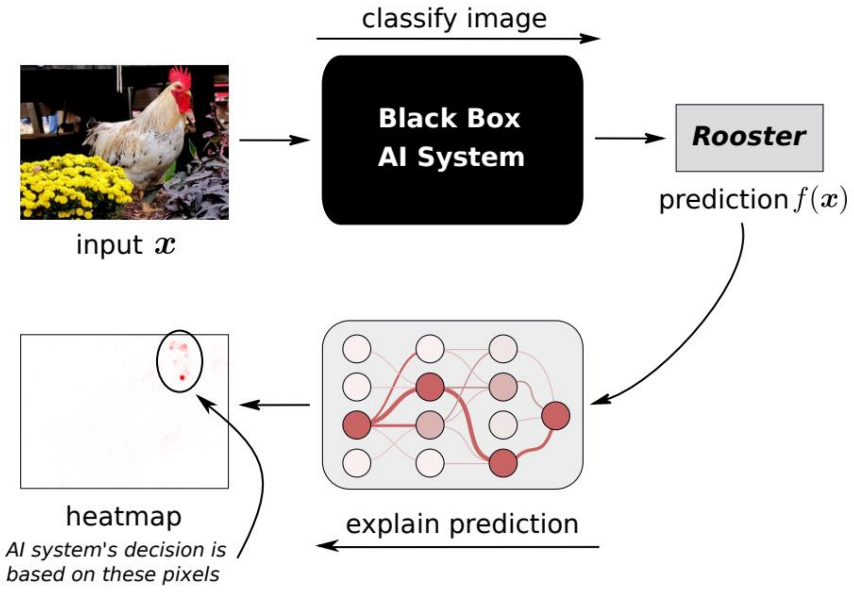
$$\underset{E}{\operatorname{argmax}}\ Q(\mathbf{Explanation}|\mathbf{Human}, \mathbf{Data}, \mathbf{Task})$$

Post-training explanations

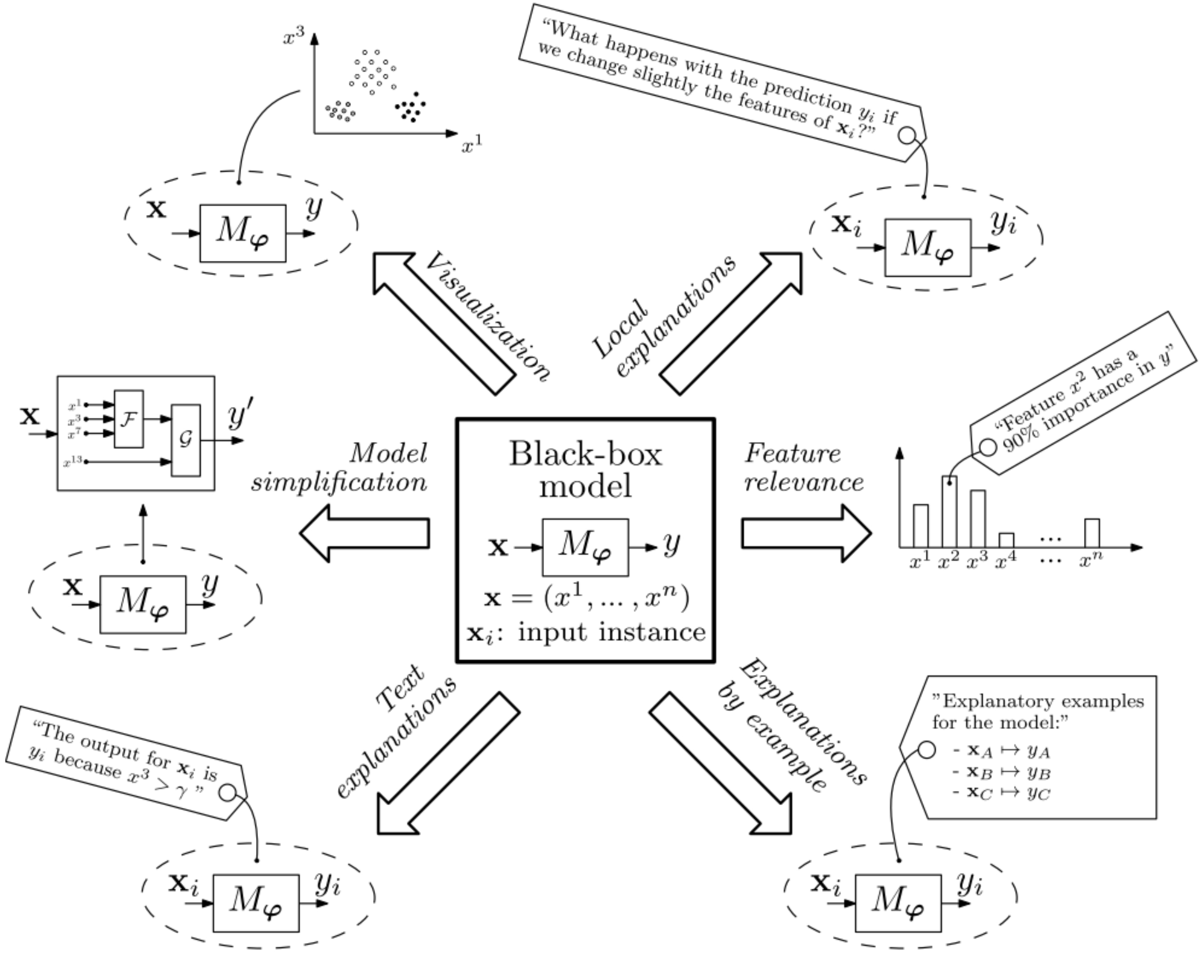
$$\underset{E}{\operatorname{argmax}}\ Q(\mathbf{Explanation}|\mathbf{Model}, \mathbf{Human}, \mathbf{Data}, \mathbf{Task})$$

Building inherently interpretable models

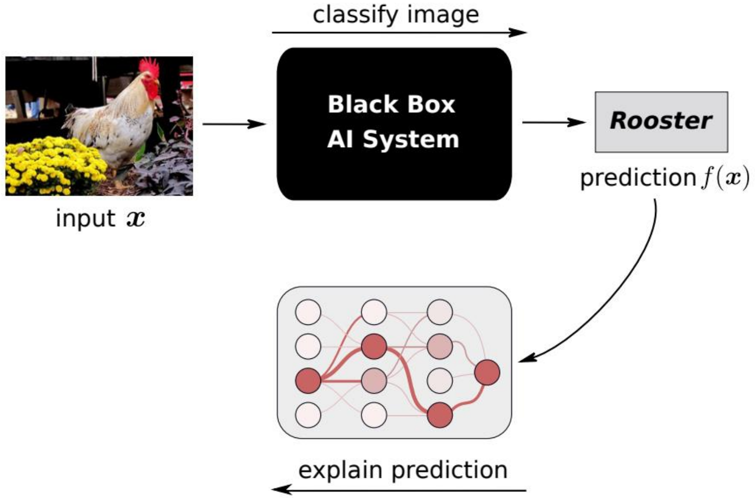
$$\underset{E,M}{\operatorname{argmax}}\ Q(\mathbf{Explanation}, \mathbf{Model}|\mathbf{Human}, \mathbf{Data}, \mathbf{Task})$$



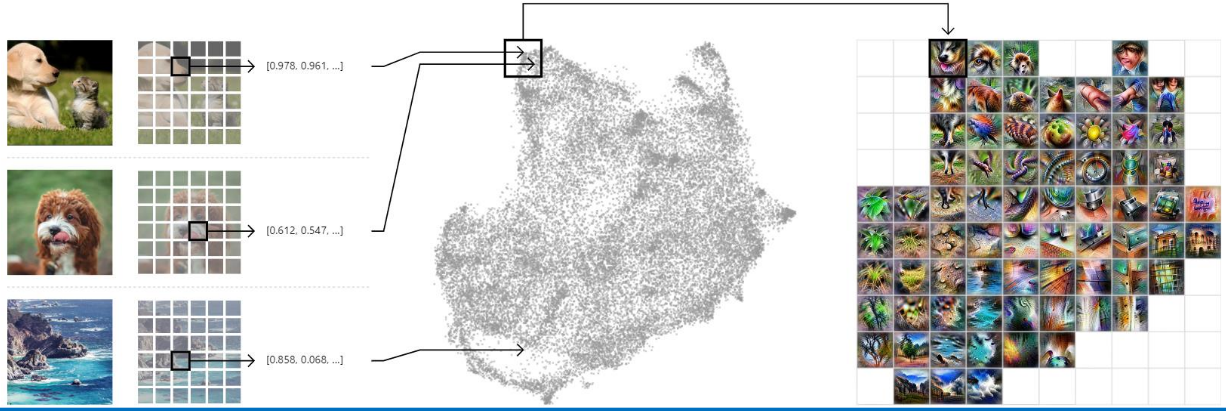
Example: Activation atlas



Visual analytics



1



Unit activations

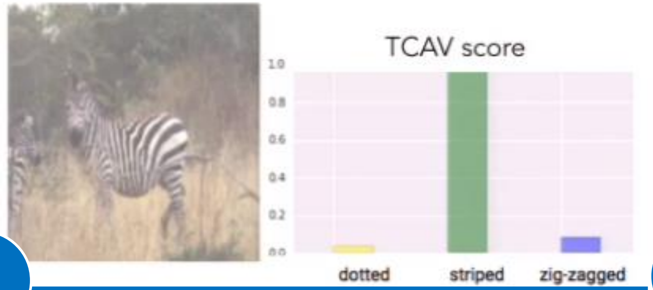
2

Saliency maps

Screw

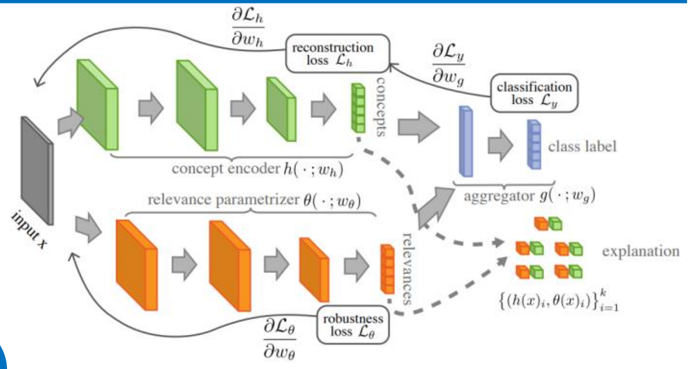
3

Class prototypes



4

Concept testing



5

Self-Explaining Neural Networks

Visual analytics

- Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications, and Bibliography for Explainable AI
- (Evaluation of explanations)

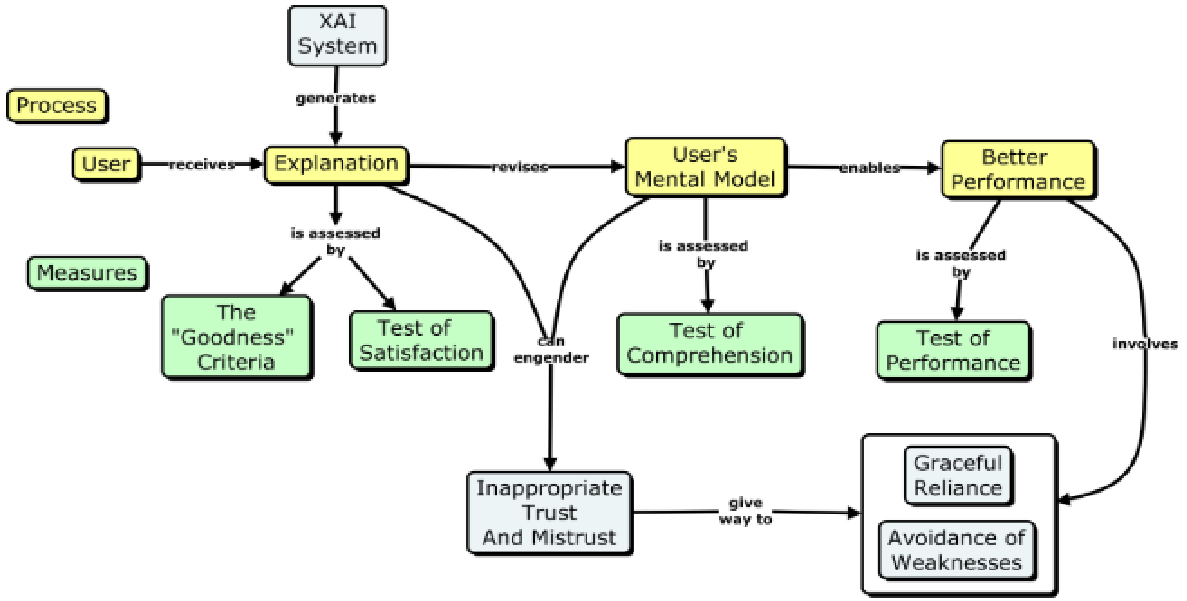


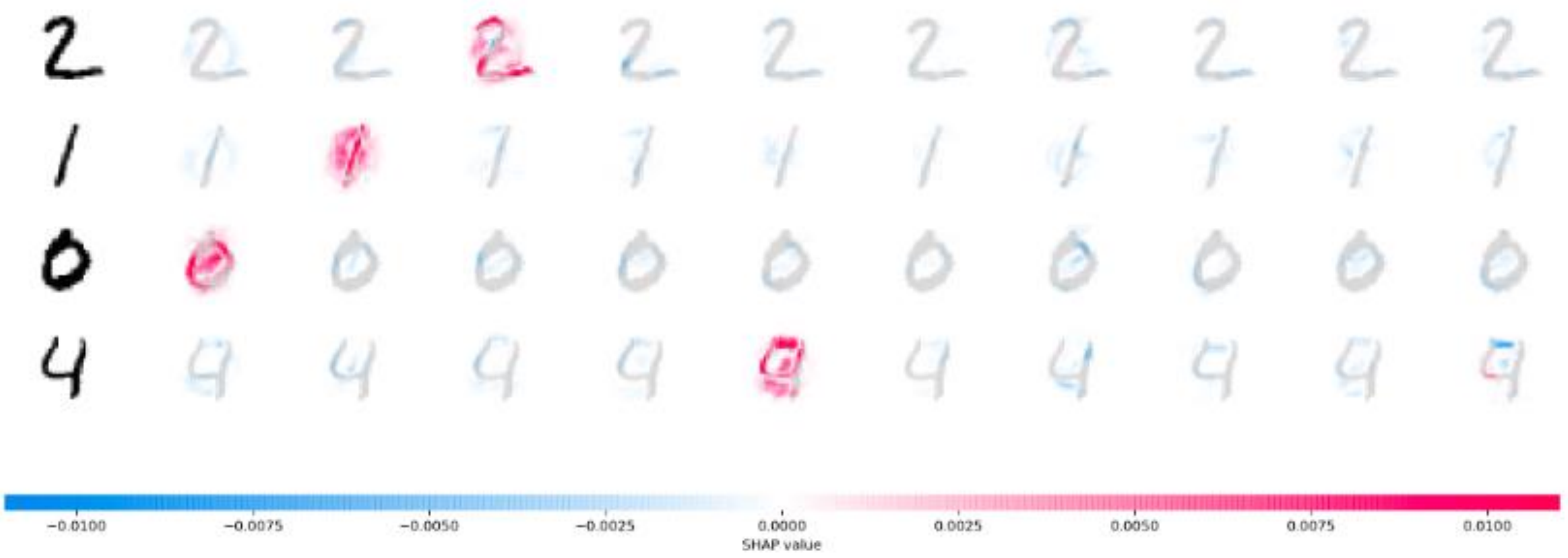
Figure 8.1. DARPA's XAI Evaluation Framework. Explanations should induce better mental models and performance, which produce appropriate trust in the system.

?

But how? And when?

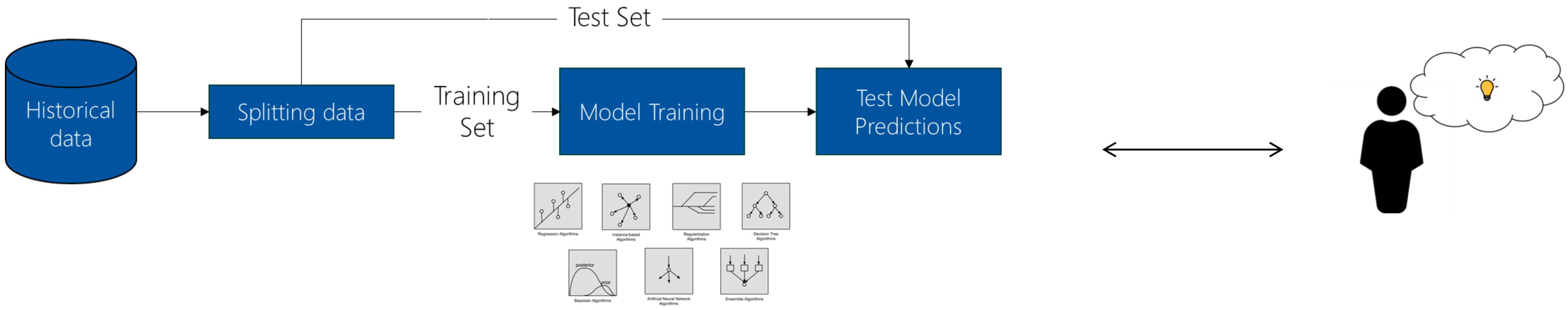
Visual analytics

An example of attribution methods in visual analytics



Visual analytics

An example of attribution methods in visual analytics



Visual analytics

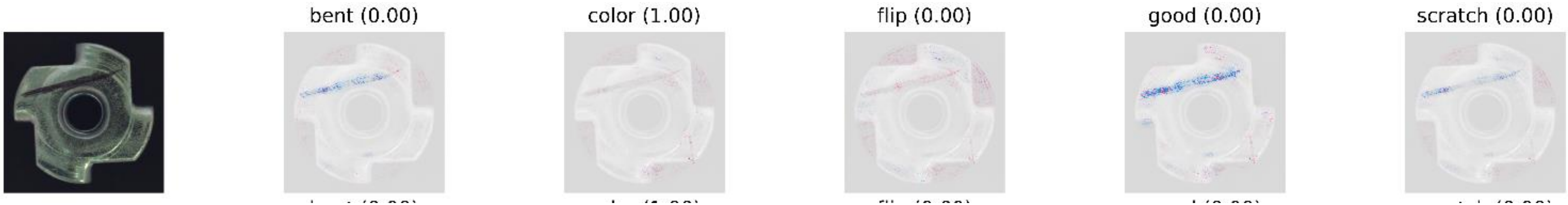
An example of attribution methods in visual analytics

- Interactive explanation-training processes can help understand what the models are learning

Visual analytics

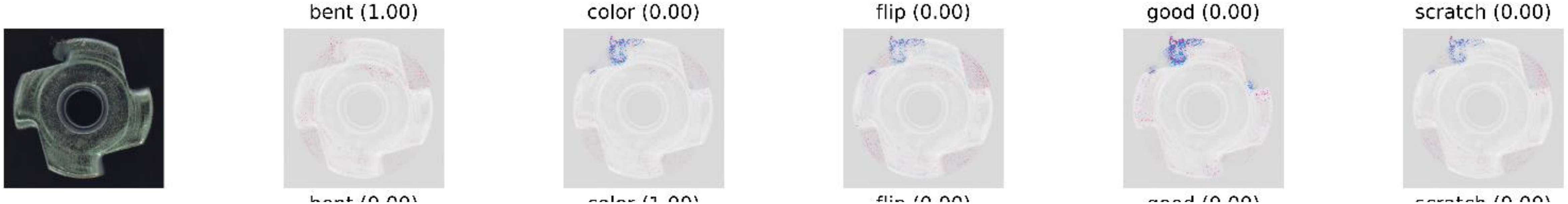
An example of attribution methods in visual analytics

- Bad explanations with high accuracy can show bias in datasets or mislabeling issues.



An example of attribution methods in visual analytics

- Correct model explanations can also generate productive insights for experts.



Visual analytics

An example of attribution methods in visual analytics

- Similar methods exist for other data types

y=sci.med (probability **0.996**, score **5.826**) top features

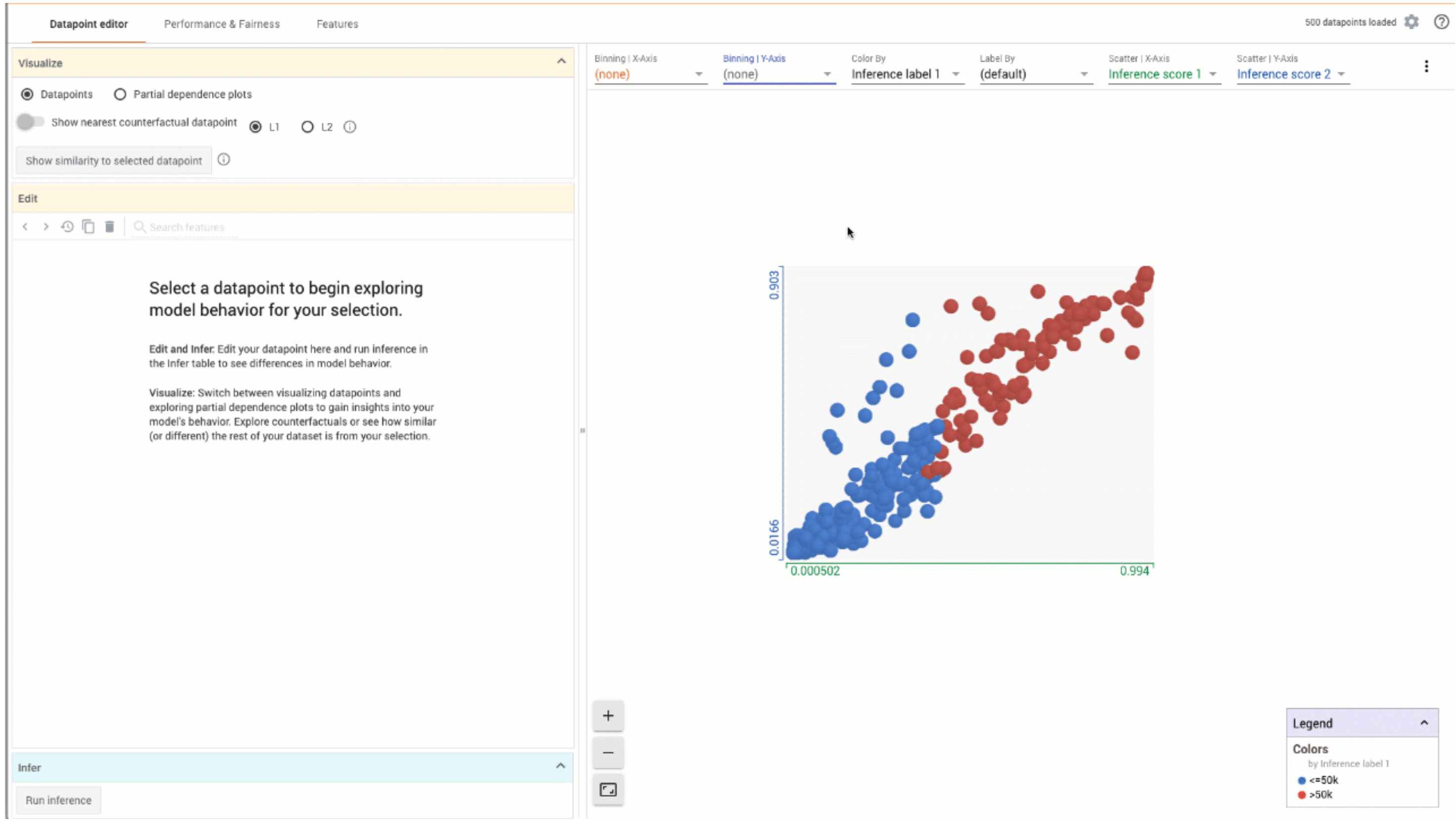
Contribution?	Feature
+5.929	Highlighted in text (sum)
-0.103	<BIAS>

as i recall from my bout with kidney stones, there isn't any medication that can do anything about them except relieve the pain. either they pass, or they have to be broken up with sound, or they have to be extracted surgically. when i was in, the x-ray tech happened to mention that she'd had kidney stones and children, and the childbirth hurt less.

Visual analytics

An example of interactive visualization methods to probe models

- What if?



https://colab.research.google.com/github/PAIR-code/what-if-tool/blob/master/WIT_Smile_Detector.ipynb

Further Reading Material

Cui, Wenqiang. "Visual analytics: a comprehensive overview." *IEEE Access* 7 (2019): 81555-81573.

Keim, Daniel, et al. "Visual analytics: Definition, process, and challenges." *Information visualization*. Springer, Berlin, Heidelberg, 2008. 154-175.

Hohman, Fred, et al. "Visual analytics in deep learning: An interrogative survey for the next frontiers." *IEEE transactions on visualization and computer graphics* 25.8 (2018): 2674-2693.

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion* 58 (2020): 82-115.

Thank you for your attention!

Lecture Team AIDAE
aidae@ima-ifu.rwth-aachen.de