Visual Data Science: Improving Science through visual reasoning

Torsten Möller
Visualization and Data Analysis
University of Vienna

Overview

- Data Science is all about modelling

- The three types of modelling

  - Computational modelling

  - Statistical modelling

  - Empirical modelling

- Challenges of Visual Data Science

- Conclusions

What is data science?

- Dhar 2013: "Data Science is the study of the generalizable extraction of knowledge from data."

What is data science?

- Dhar 2013: "Data Science is the study of the generalizable extraction of knowledge from data."

- Data Science is the study of exploration, abstraction, and communication of complex systems through models from data.

What is data science?

- Dhar 2013: "Data Science is the study **of the generalizable extraction of knowledge** from data."

- Data Science is the study of **exploration, abstraction, and communication** of complex systems through models from data.

What is data science?

- Dhar 2013: "Data Science is the study **of the generalizable extraction of knowledge** from data."

- Data Science is the study of exploration, abstraction, and communication **of complex systems** through models from data.

What is data science?

- Dhar 2013: "Data Science is the study **of the generalizable extraction of knowledge** from data."

- Data Science is the study of exploration, abstraction, and communication of complex systems **through models** from data.

Data Science
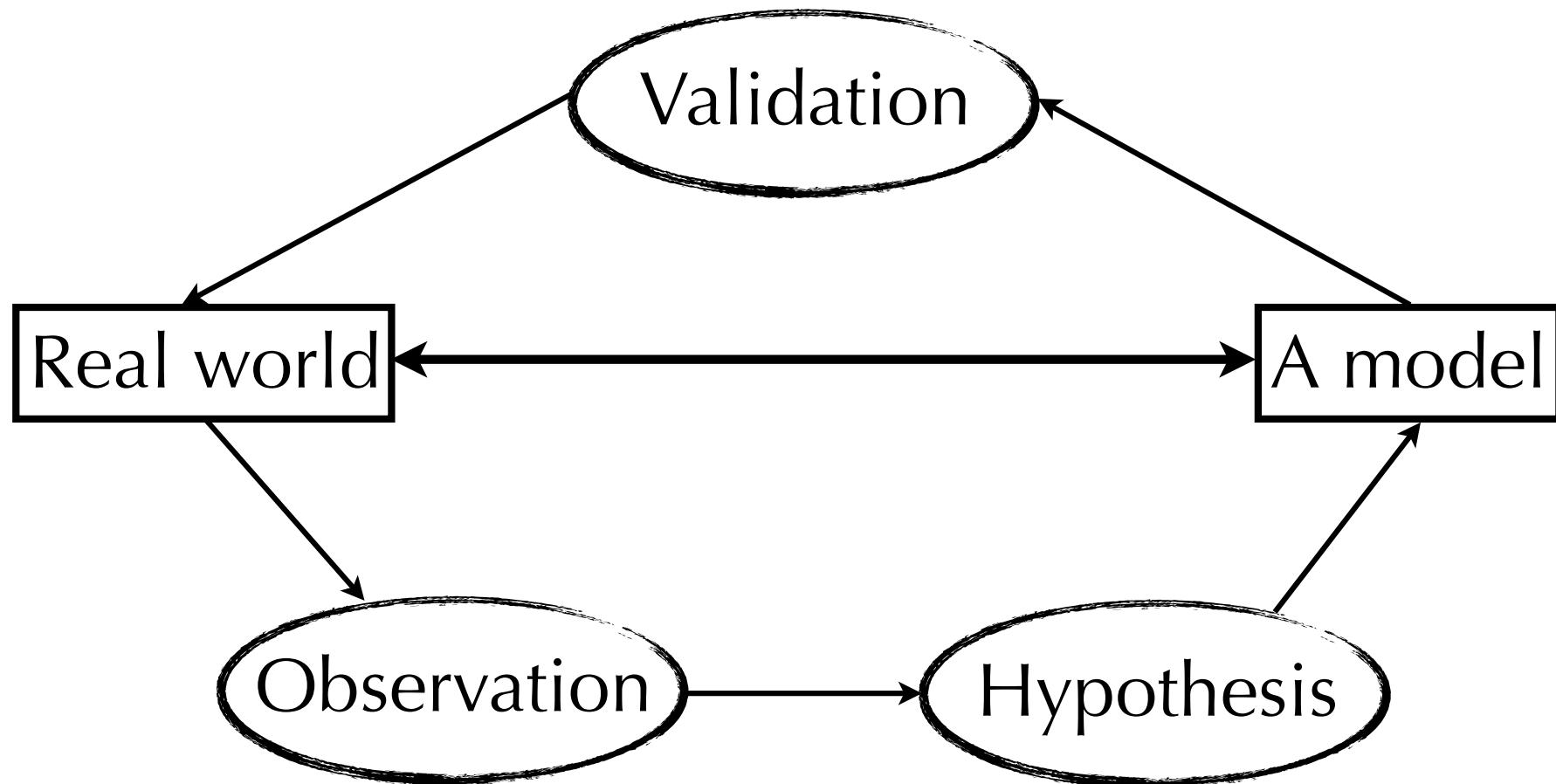
- Jeff Leek: "The key word in 'Data Science' is not Data, it is Science"

  "The issue is that the hype around big data/data science will flame out (it already is) if data science is only about "data" and not about "science". The long term impact of data science will be measured by the scientific questions we can answer with the data."
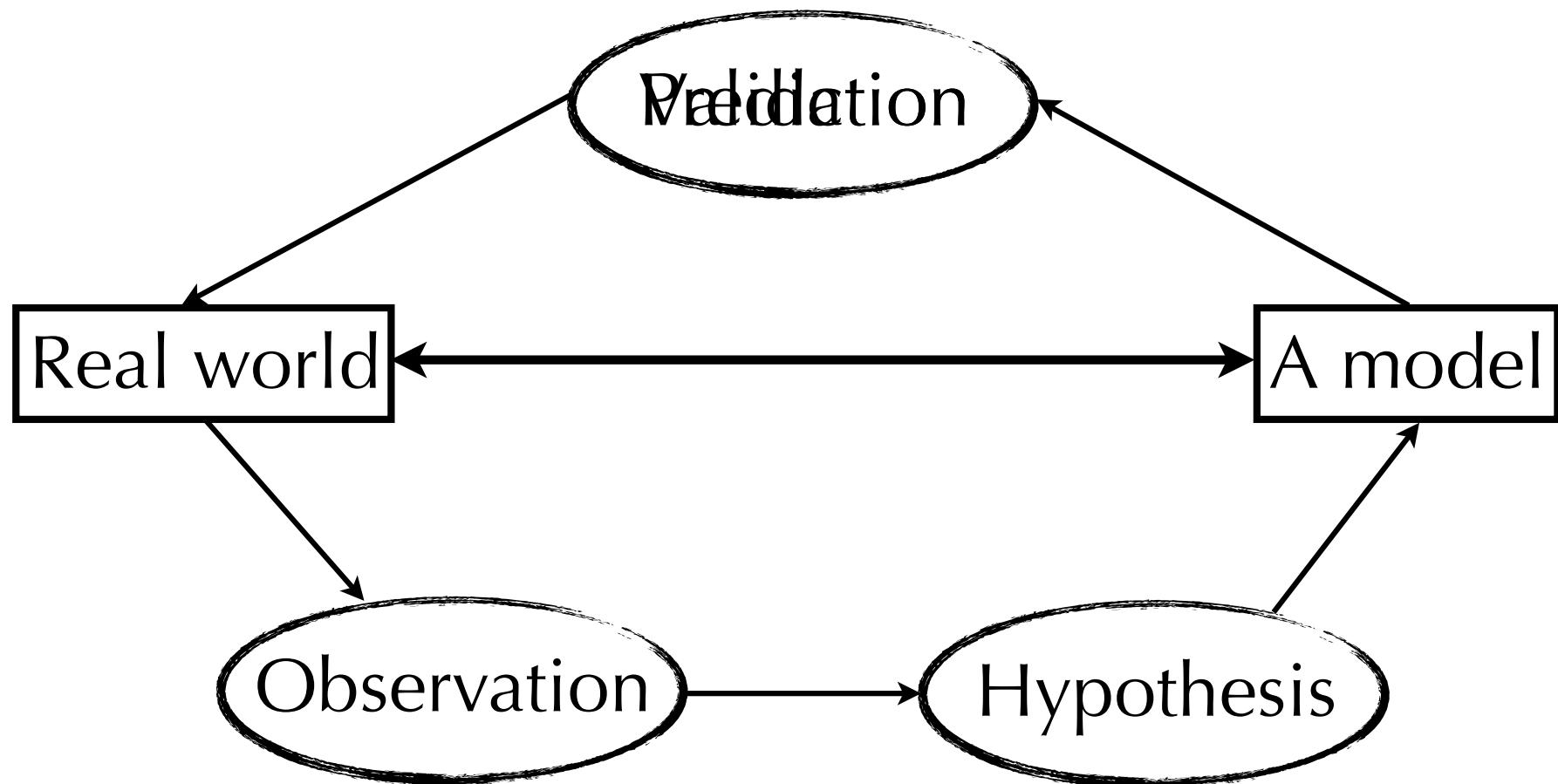
Overview
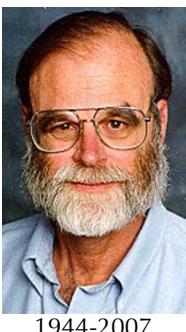
- Data Science is all about modelling

- The three types of modelling

  - Computational modelling

  - Statistical modelling

  - Empirical modelling

- Challenges of Visual Data Science

- Conclusions

Scientific Method

Torsten Möller
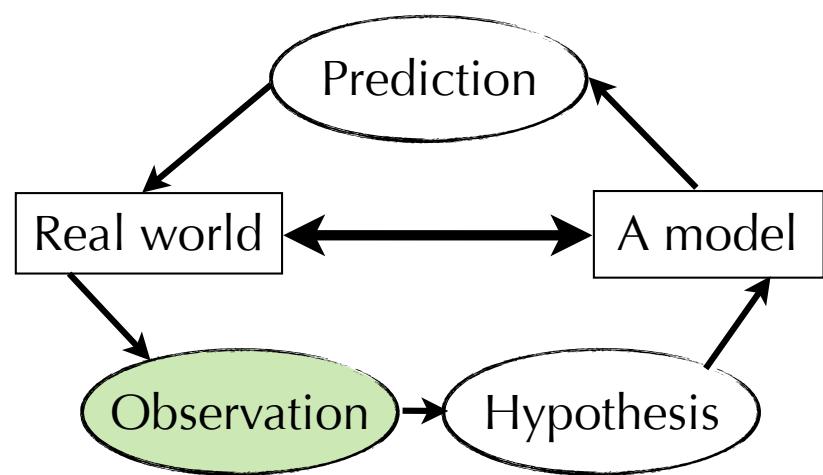
# Validation ➤ Prediction

4 Paradigms of Science

1944-2007
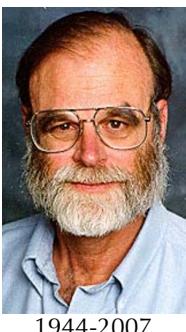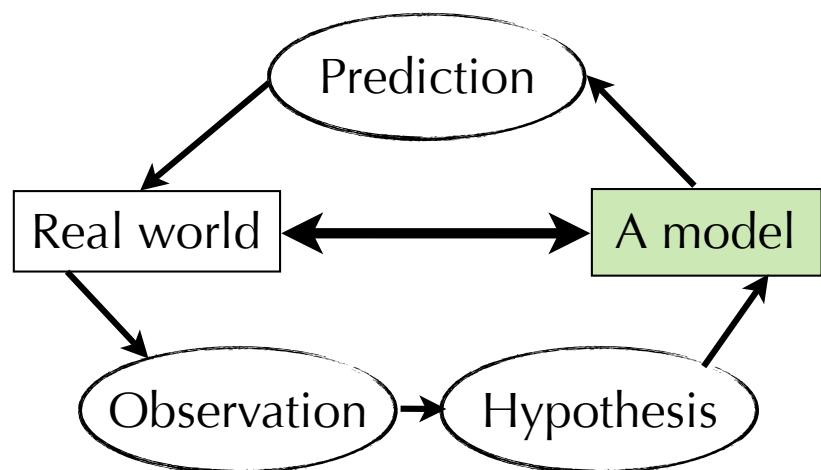
- empirical: observe, then derive

4 Paradigms of Science

1944-2007
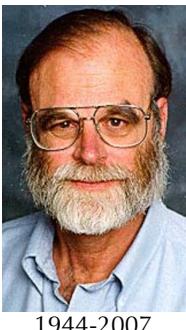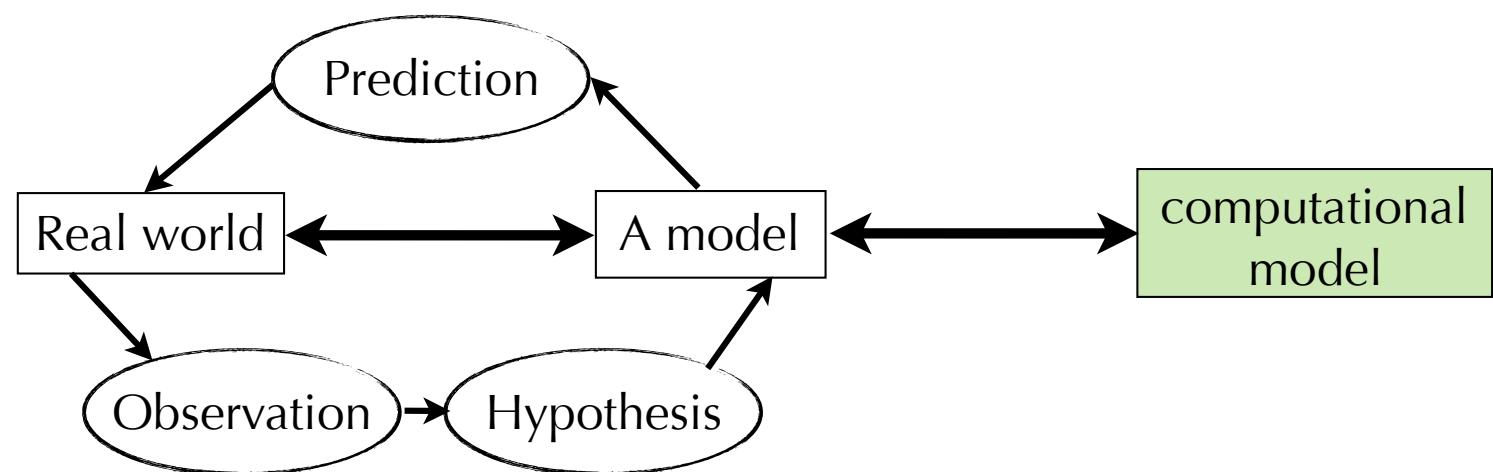
- empirical: observe, then derive

- predictive: derive, then observe

4 Paradigms of Science

1944-2007
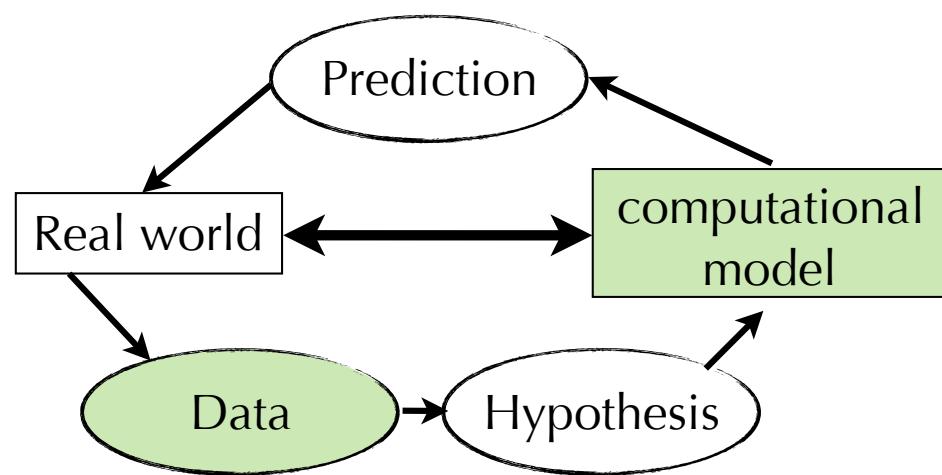
- empirical: observe, then derive

- predictive: derive, then observe

- computational: simulate

4 Paradigms of Science

1944-2007

- empirical: observe, then derive

- predictive: derive, then observe

- computational: simulate

- data-driven: measure
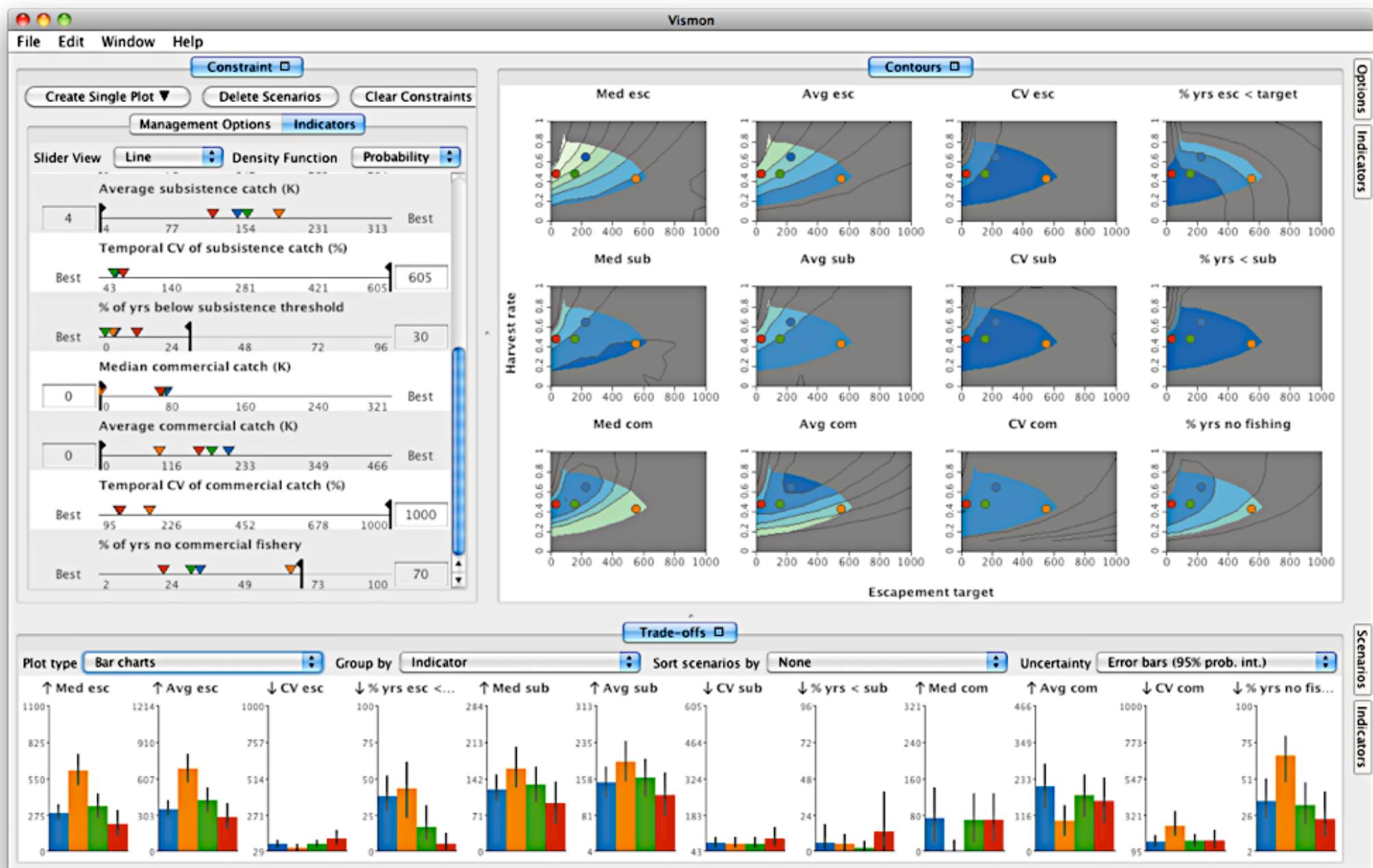
Three types of modelling

- computational: the simulation of discretized mathematical models (computational science)

- statistical: data-driven — extracting statistical models from data

- empirical: simple, often linear models

Computational Modelling

- (almost) every discipline has these models

- Examples:

  - Navier-Stokes, Maxwell, etc.

  - Population Dynamics

- computational science: experimentation through simulation of discretized models

Vismon: Fisheries Science

Roles - old days

Recommendations

Objectives

Scientists

Managers

Goal

Data-driven decision making
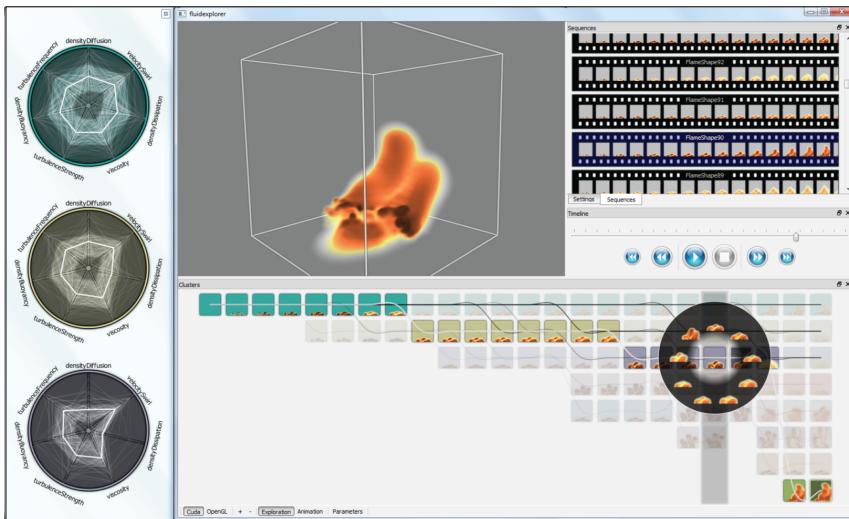
Simulation worksheet
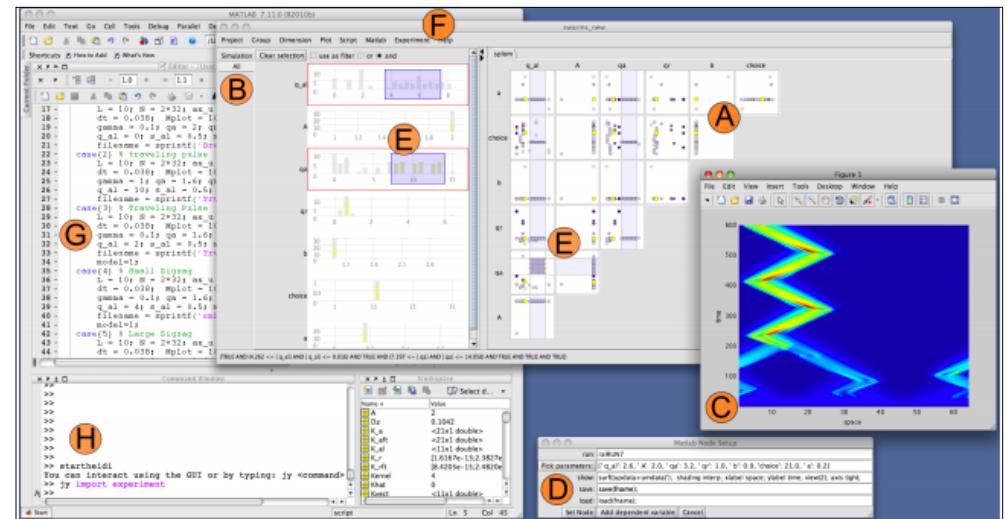
Scientists
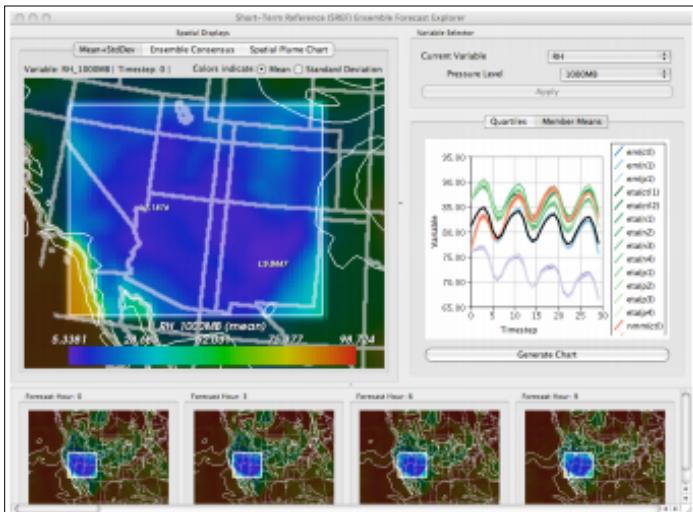
Managers

Goal:
move data-driven decision making from scientists to managers
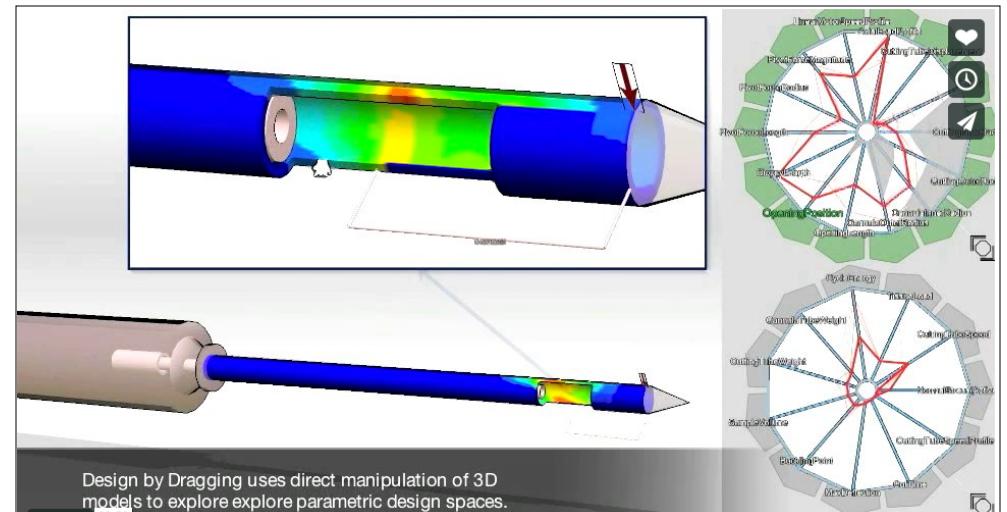
[Bruckner & Möller 2010]

[Bergner et al. 2013]
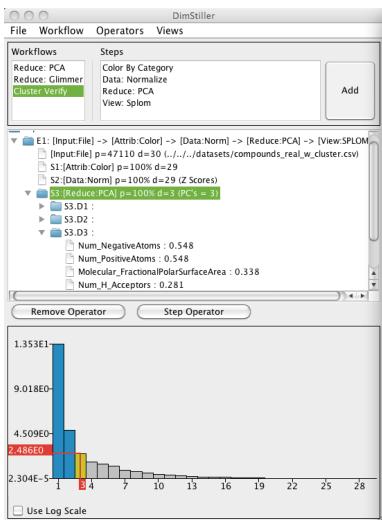
[Potter et al. 2009]
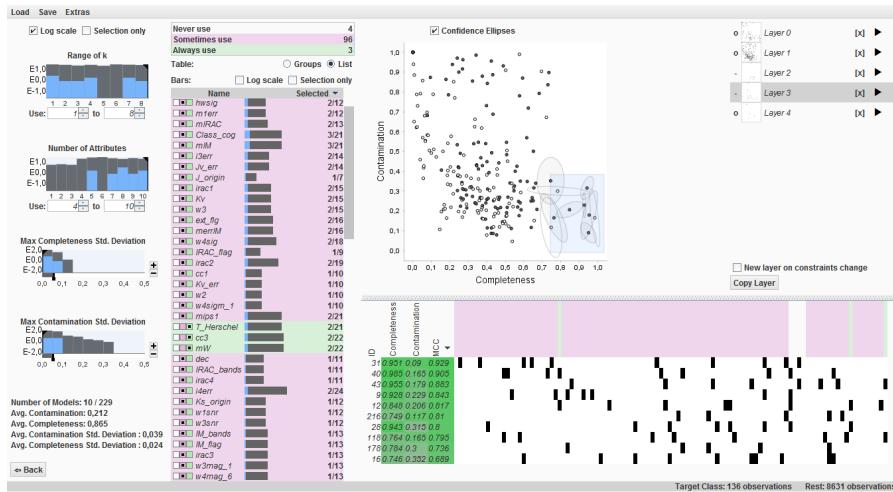
[Coffey et al. 2013]

# Statistical Modeling
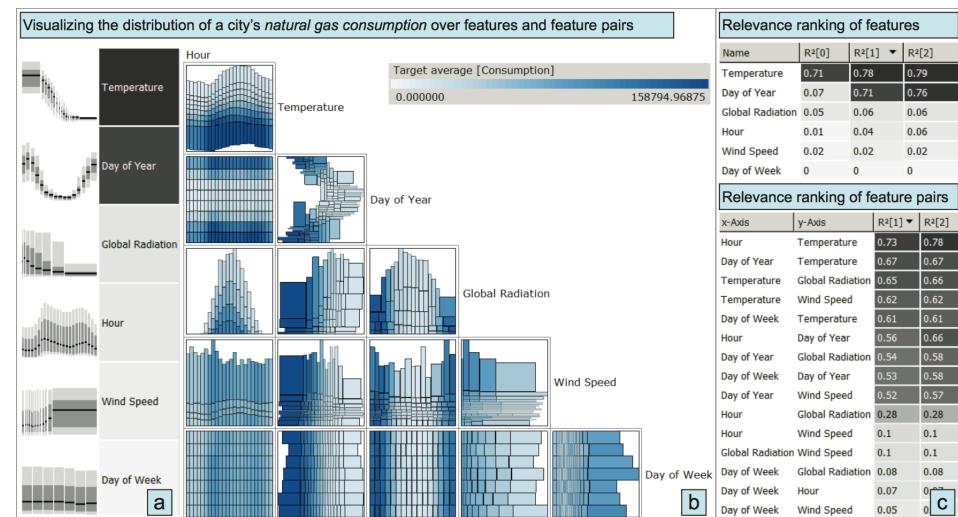
- "Mainstream" understanding of Data Science

- Classical (machine learning) approaches:

  - Clustering

  - Classification

  - Regression

  - (dimensionality reduction, outlier detection, etc)
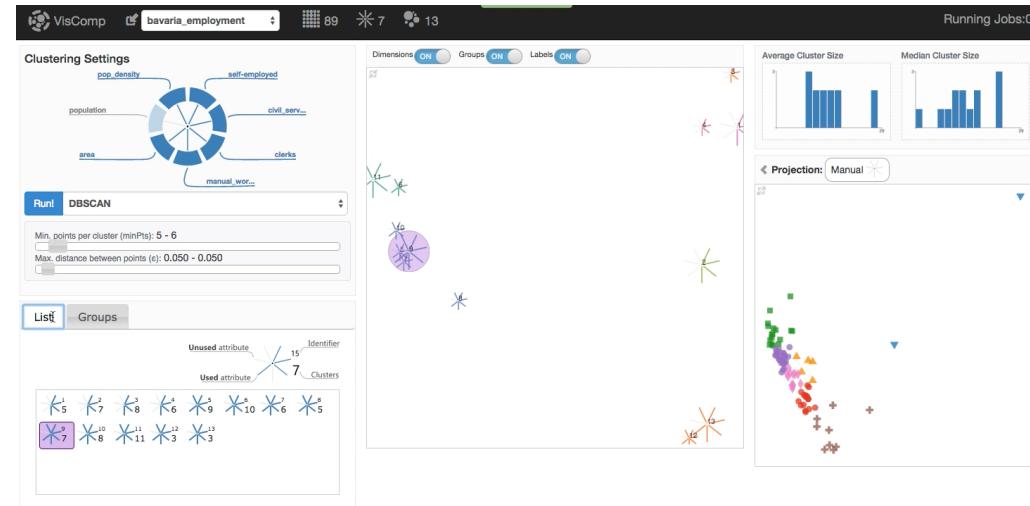
Dim reduction — [Ingram et al. 2010]

Regression — [Mühlbacher & Piringer 2013]

Classification — [Linhardt et al. 2016?]

**Clustering** — [Sedlmair et al. 2016?]
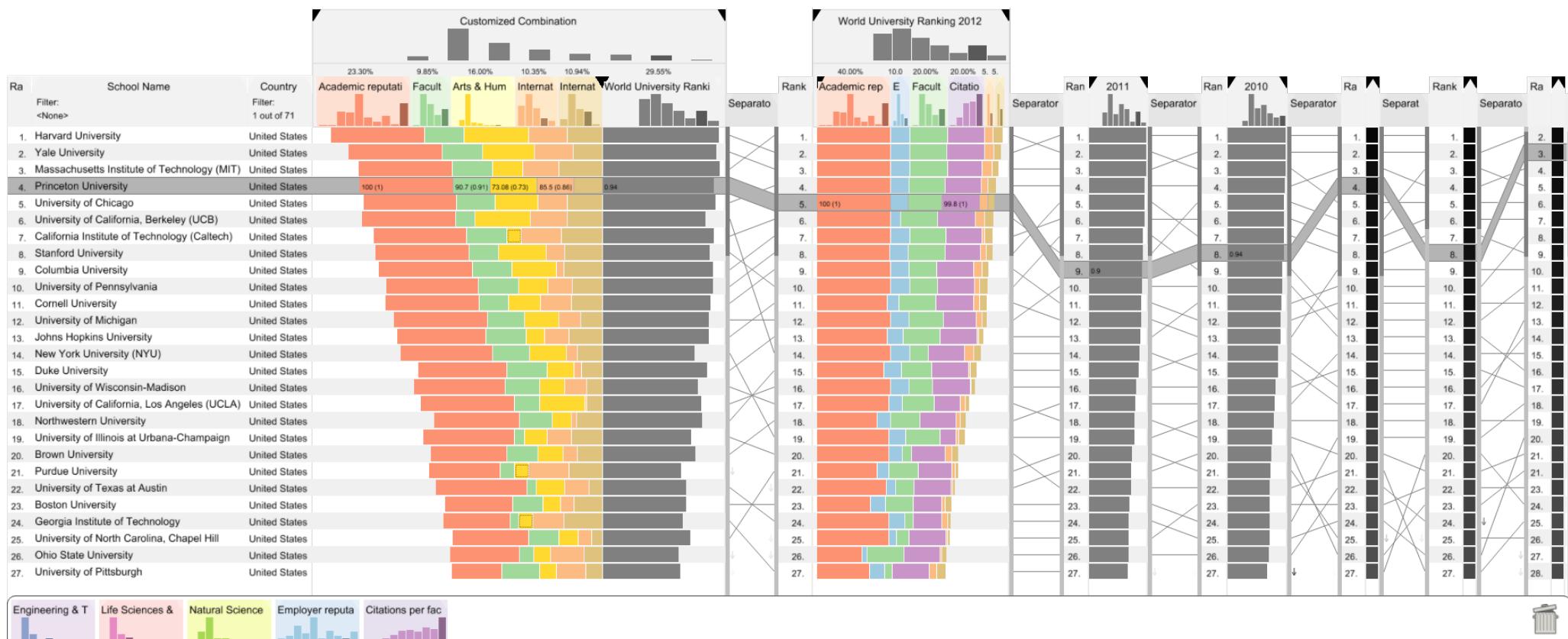
Empirical Modeling
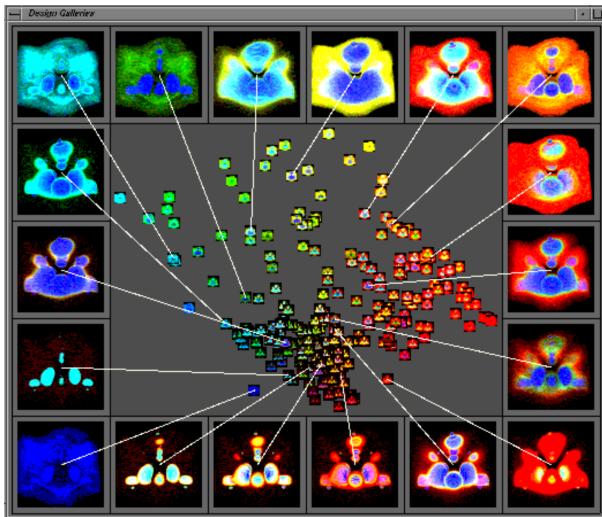
- often no explicit modelling or only simple models, e.g.

  - linear models

  - weighted averages etc.

- examples: spreadsheets, rankings

LineUp: Gratzl et al. 2013
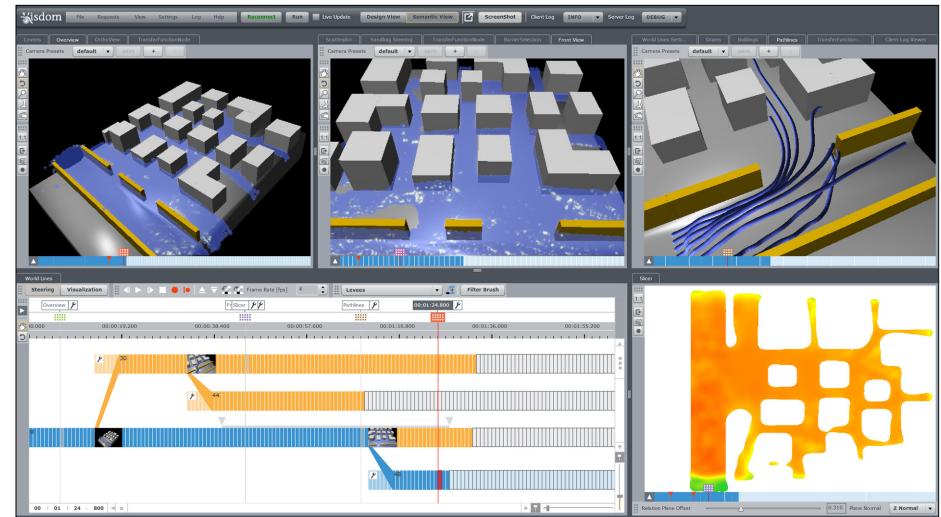
LineUp: Gratzl et al. 2013
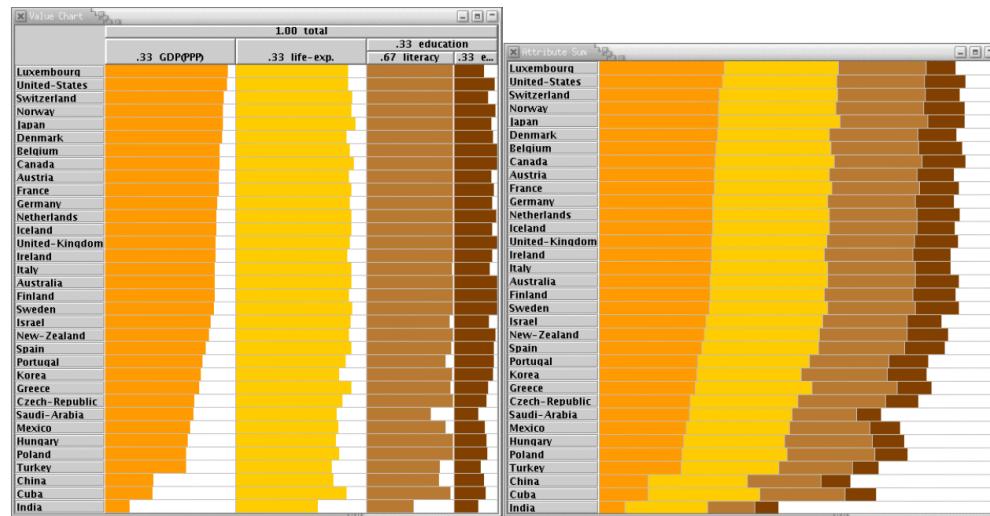
Design Galleries — [Marks et al. 1997]
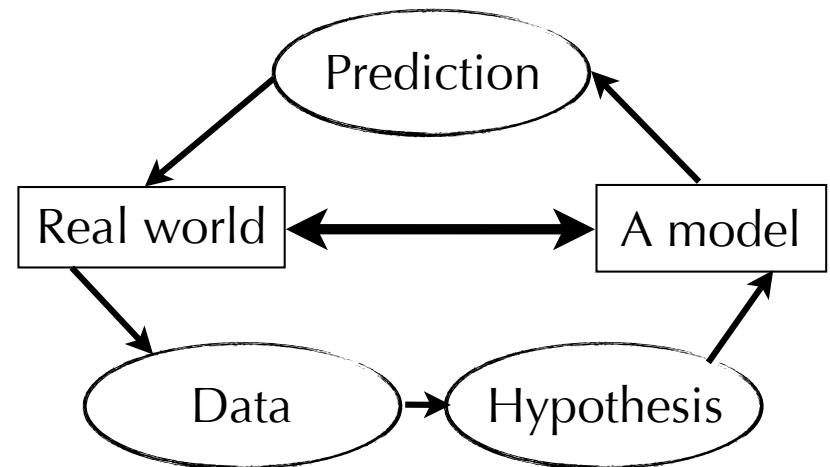
World Lines — [Waser et al. 2010]

ValueCharts — [Carenini et al. 2004]

Not just Labcoat Science

- valid for business, engineering, public policy

- general data analysis approach

Overview

- Data Science is all about modelling

- The three types of modelling

  - Computational modelling

  - Statistical modelling

  - Empirical modelling

- **Challenges of Visual Data Science**

- **Conclusions**

What is visual data science?

- **Visual** Data Science **is helping users** explore, abstract, and communicate complex systems through models from data.

Acting upon models

Data → Models (predictions) → Decisions

Building vs. Using

```
┌──────────┐        ┌──────────────┐        ┌─────────────┐
│   Data   │ ─────> │    Models    │ ─────> │  Decisions  │
└──────────┘        │ (predictions)│        └─────────────┘
                    └──────────────┘
```
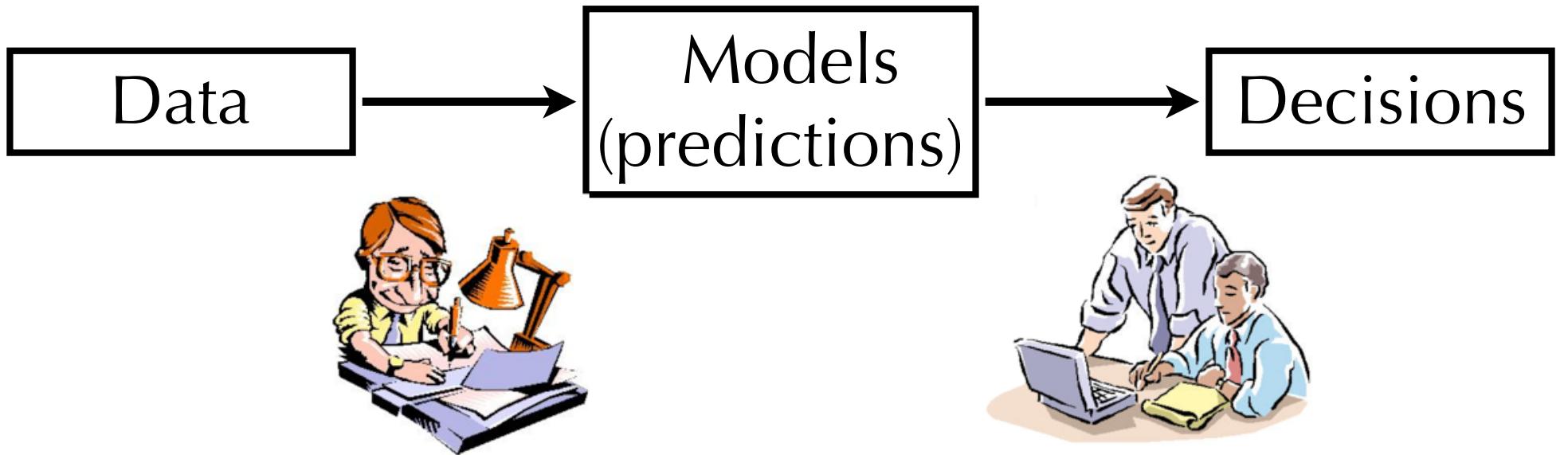
- building models
  - computational experts
  - bioinformaticians

- using models
  - decision makers
  - domain experts
  - biologists

Building vs. Using

```
┌──────────┐        ┌──────────────┐        ┌──────────────┐
│   Data   │───────▶│    Models    │───────▶│  Decisions   │
└──────────┘        │ (predictions)│        └──────────────┘
                    └──────────────┘
```
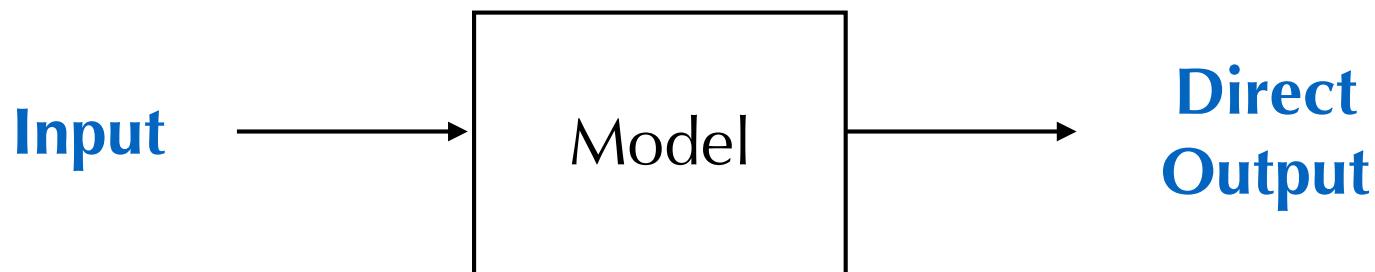
- building models
  - validation
  - uncertainty

- using models
  - trust
  - tradeoffs + risks

A modern microscope
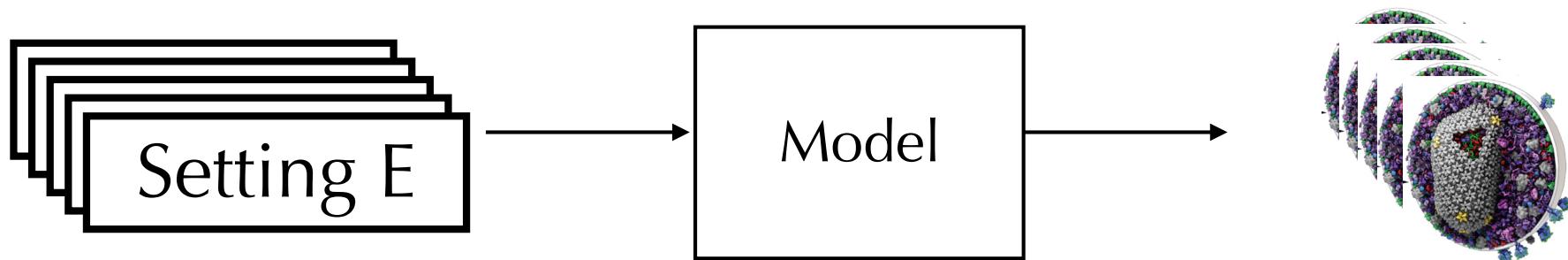
| Data	→	Models (predictions)	→	Decisions

- making difficult algorithmic solutions accessible to a broad audience: enable model users to become model builders

What is a model?

Input → **Model** → **Direct Output**

- has input parameters

- creates outputs

- it's really "just" an algorithm

What is a model?

Setting E → Model →

- paradigm shift:

  - from single input/output exploration to input ranges and ensemble outputs

Supporting the user

Data $\longrightarrow$ Models (predictions) $\longrightarrow$ Decisions

- hypothesis creation

- uncertainty / risk analysis

- sensitivity analysis / model uncertainty

- decision making / sense making

Conclusions
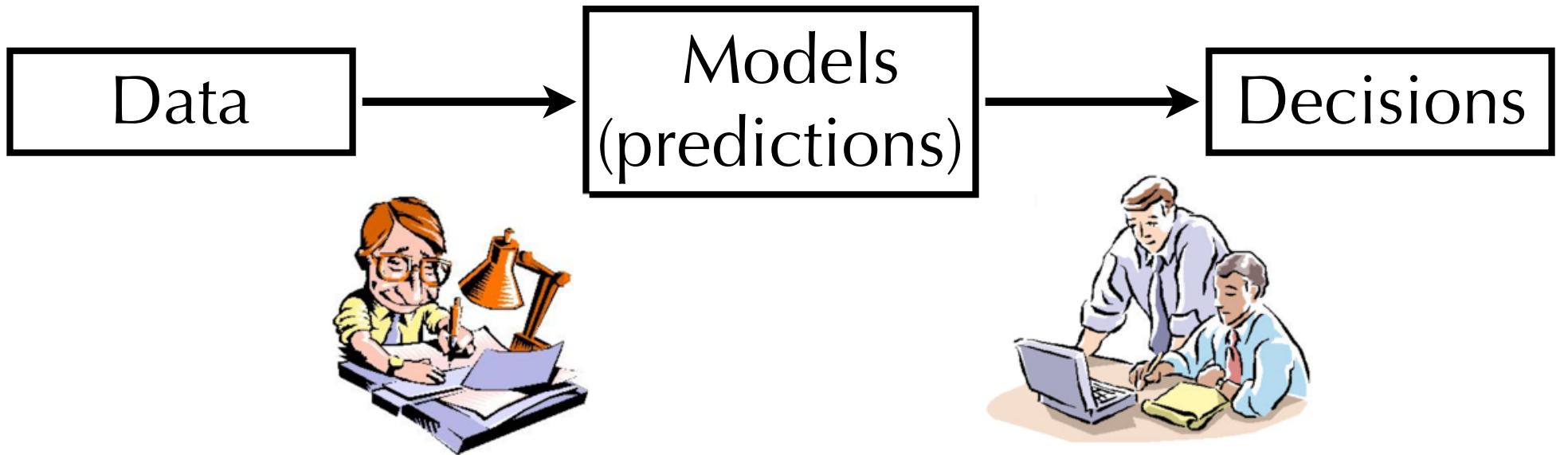
Torsten Möller

What is visual data science?

- **Visual** Data Science **is helping users** explore, abstract, and communicate complex systems **through models** from data.

Three types of modelling

- computational

- statistical

- empirical

A modern microscope

- making difficult algorithmic solutions accessible to a broad audience: enable model users to become model builders

Visual   Data Science

=

=
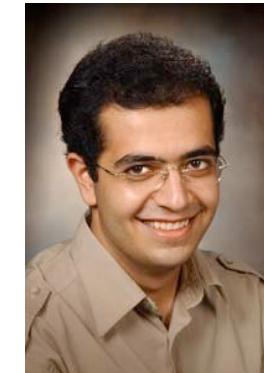
Making modelling techniques accessible to a broad set of users without requiring a PhD in Stats/ML.

Acknowledgments

Steven Bergner
FINCAD

Maryam Booshehrian
Muprime Tech

Stephen Ingram
Coho Data

Tom Torsney-Weir
U of Vienna

Hamid Younesy
SFU

Lorenz Linhardt
U of Vienna

Stefan Bruckner
U of Bergen

Tamara Munzner
UBC

Melanie Tory
U Victoria

Harald Piringer
VRVis

Michael Sedlmair
U of Vienna

Patrick Wolf
Software Dev

Questions?

http://vda.cs.univie.ac.at

torsten.moeller@univie.ac.at