Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Causal Inference
Theory and Applications in Enterprise Computing

Dr. Matthias Uflacker, Johannes Huegle, Christopher Schmidt

April 10, 2019

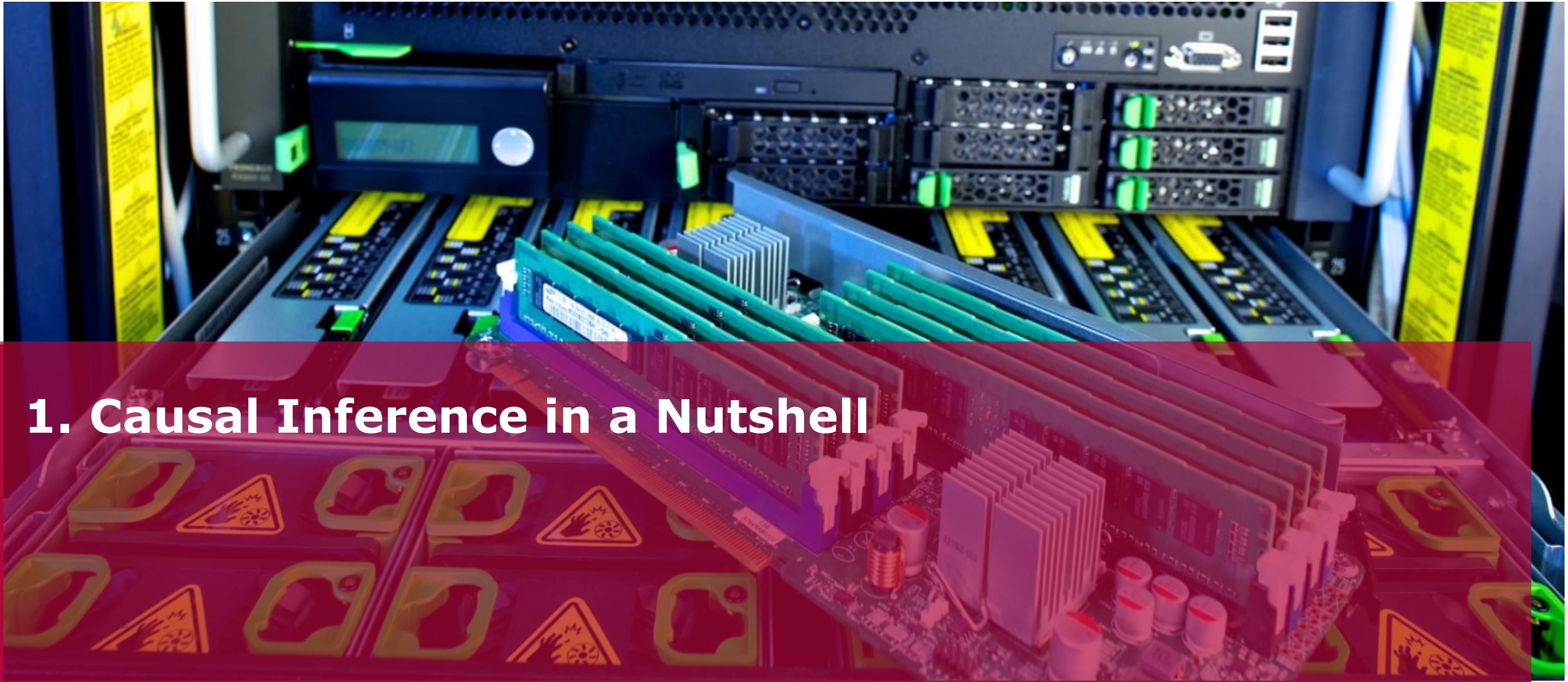
Agenda
April 10, 2019

1. **Causal Inference in a Nutshell**

2. **Causal Inference in Application**

3. **Introduction to Research Topics**

4. **Further Reading**



1. Causal Inference in a Nutshell

1. Causal Inference in a Nutshell
Recap: Summary

Traditional statistics, machine learning, etc.

- About associations
- Model the distribution of the data
- Predict given observations

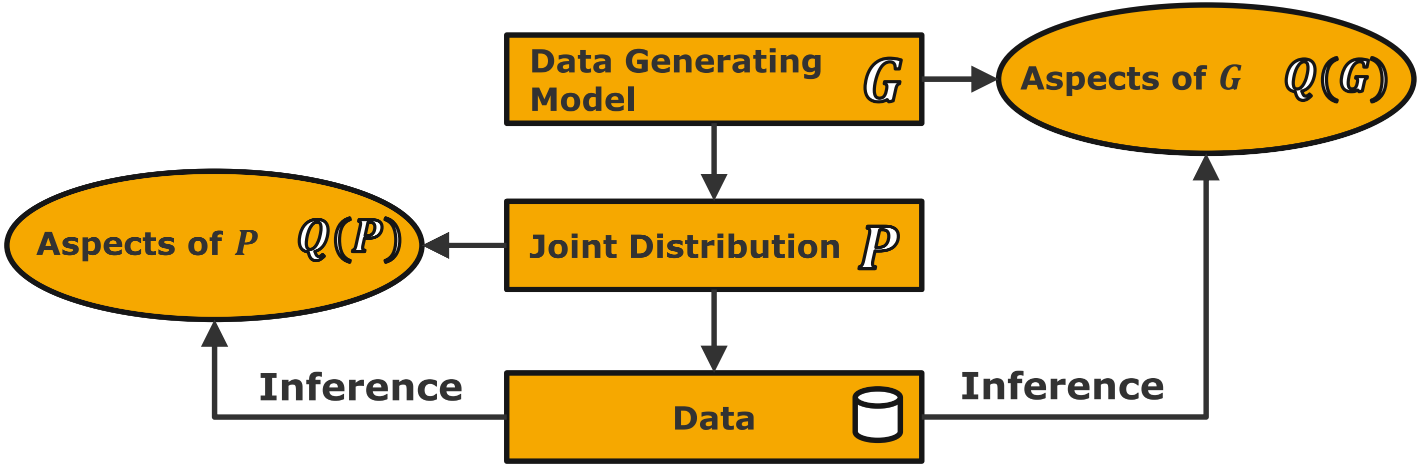
Causal Inference

- About causation
- Model the mechanism that generates the data
- Predict results of interventions

1. Causal Inference in a Nutshell
Recap: Concept

Traditional Statistical Inference Paradigm

Paradigm of Structural Causal Models



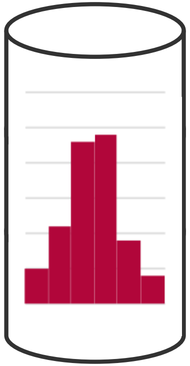
E.g., what is the sailors' probability of recovery when **we see** a treatment with lemons?

$$Q(P) = P(recovery|lemons)$$

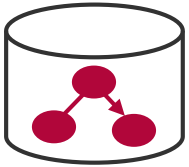
E.g., what is the sailors' probability of recovery if **we do** treat them with lemons?

$$Q(G) = P(recovery|do(lemons))$$

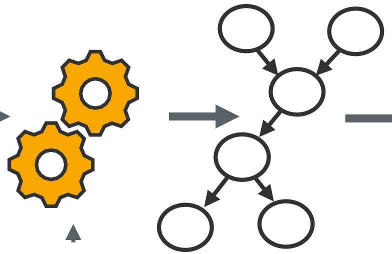
1. Causal Inference in a Nutshell
Recap: Inference Procedure



Observational Data



Background Knowledge



Causal Relationships

Causal Structure: *"What are the causal relationships in the system?"*

"How is lung cancer related to smoking and genetics?"

Probabilistic Inference

$$P(X_3 \mid X_1 = x_1, X_2 = x_2)$$
$$P(X_4 \mid X_2 = x_2)$$

Association: *"What is a certain probability if we find the system how it is?"*

"How likely do smoking people get lung cancer?"

Causal Inference

$$P(X_3 \mid do(X_1 = x_1), do(X_2 = x_2))$$
$$P(X_4 \mid do(X_2 = x_2))$$

Intervention: *"What is a certain probability if we manipulate the system?"*

"What if we ban cigarettes?

Functional Systems

$$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$$
$$f_2(x_3, x_4) = \quad \ldots$$

Counterfactuals: *"What if the system would have been different?"*

"What if I had not been smoking the past 2 years?"

Data | **Causal Structure Learning** | **Opportunities** | **Examples**

2. Causal Inference in Application

2. Causal Inference in Application
Causal Relationships (I/II)

"What are the principal structural properties of genetic control programs of the cell's biological processes?"

Causal Relationships

Probabilistic Inference

$P(X_3 | X_1 = x_1, X_2 = x_2)$

$P(X_4 | X_2 = x_2)$

Causal Inference

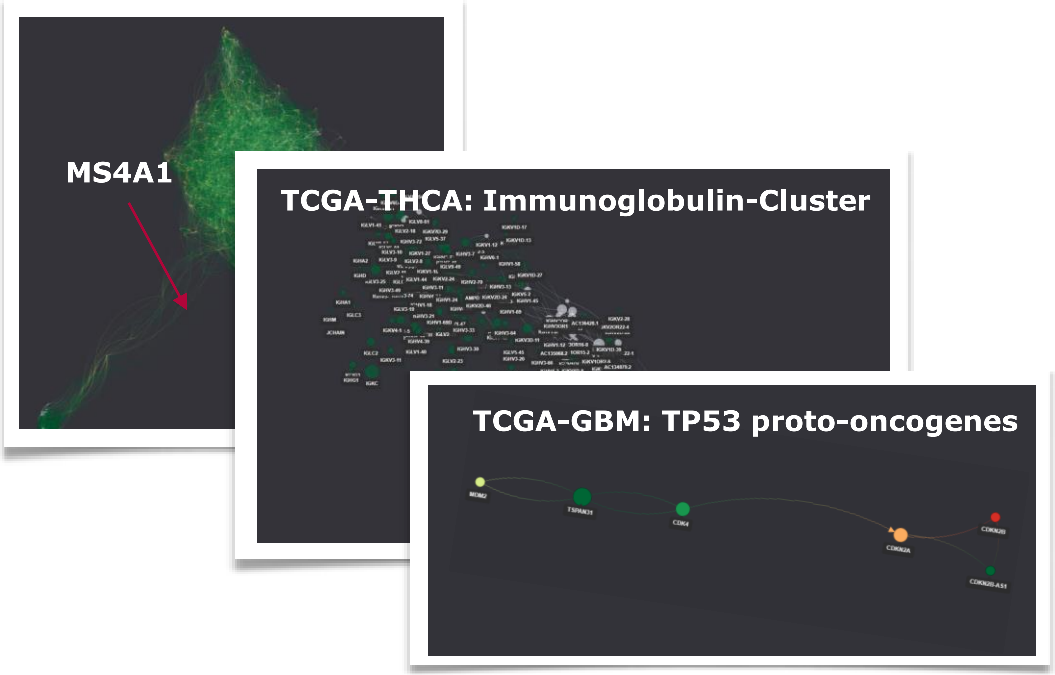
$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$

$P(X_4 | do(X_2 = x_2))$

Functional Systems

$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$

$f_2(x_3, x_4) = \quad \ldots$



2. Causal Inference in Application
Causal Relationships (II/II)

"What are causes or effects of errors in a complex automotive production process?"

Causal Relationships

Probabilistic Inference

$P(X_3 | X_1 = x_1, X_2 = x_2)$

$P(X_4 | X_2 = x_2)$

Causal Inference

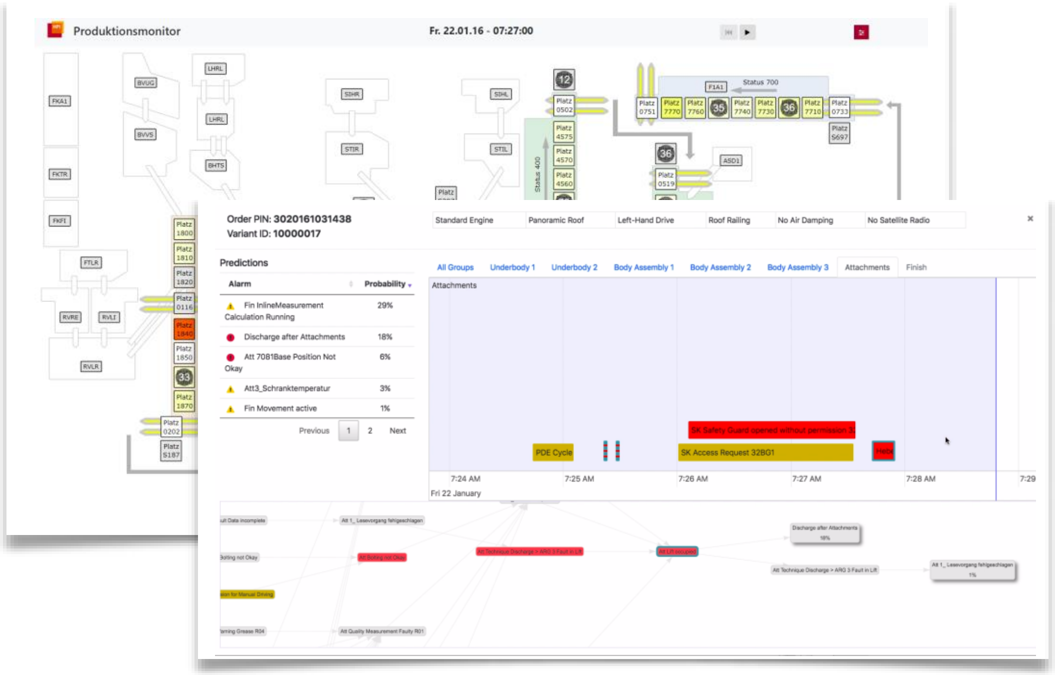
$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$

$P(X_4 | do(X_2 = x_2))$

Functional Systems

$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$

$f_2(x_3, x_4) = \quad \ldots$



Causal Inference
Theory and Applications in Enterprise Computing

Uflacker, Huegle, Schmidt

2. Causal Inference in Application
Probabilistic Inference (I/II)

"Given current error occurring in an automotive production process, what effect is likely?"

Causal Relationships

Probabilistic Inference

$P(X_3 | X_1 = x_1, X_2 = x_2)$

$P(X_4 | X_2 = x_2)$

Causal Inference

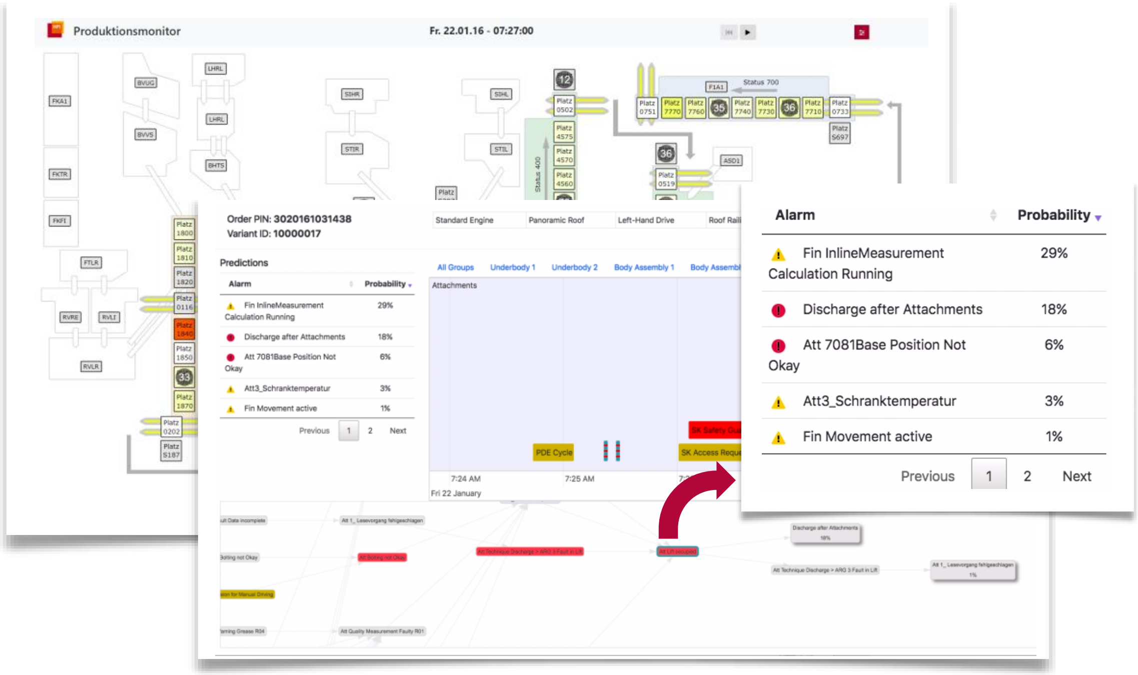
$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$

$P(X_4 | do(X_2 = x_2))$

Functional Systems

$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$

$f_2(x_3, x_4) = \quad \ldots$



2. Causal Inference in Application
Probabilistic Inference (II/II)

"How to leverage knowledge about likely effects given errors in the current production situation ?"

Causal Relationships

Probabilistic Inference

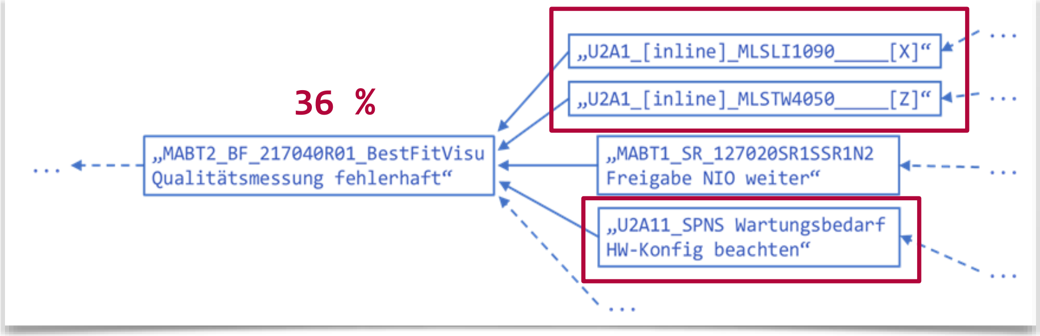
$P(X_3 | X_1 = x_1, X_2 = x_2)$
$P(X_4 | X_2 = x_2)$

Causal Inference

$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$
$P(X_4 | do(X_2 = x_2))$

Functional Systems

$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$
$f_2(x_3, x_4) = \quad ...$




```
if („U2A1_[inline]_MLSLI1090__[X]“ < LRL) and („U2A1_[inline]_MLSLI1090__[X]“ < LRL)
          then „MABT2_BF_217040R01_BestFitVisu Qualitätsmessung fehlerhaft“

if „MABT1_SR_127020SR1SSR1N2 Freigabe NIO weiter“
          then „MABT2_BF_217040R01_BestFitVisu Qualitätsmessung fehlerhaft“

if „U2A11_SPNS Wartungsbedarf HW-Konfig beachten“
          then „MABT2_BF_217040R01_BestFitVisu Qualitätsmessung fehlerhaft“
```

2. Causal Inference in Application
Causal Inference

“What is the causal effect behind the complex causal structures in a production process?”

Causal Relationships

Probabilistic Inference

$P(X_3 | X_1 = x_1, X_2 = x_2)$

$P(X_4 | X_2 = x_2)$

Causal Inference

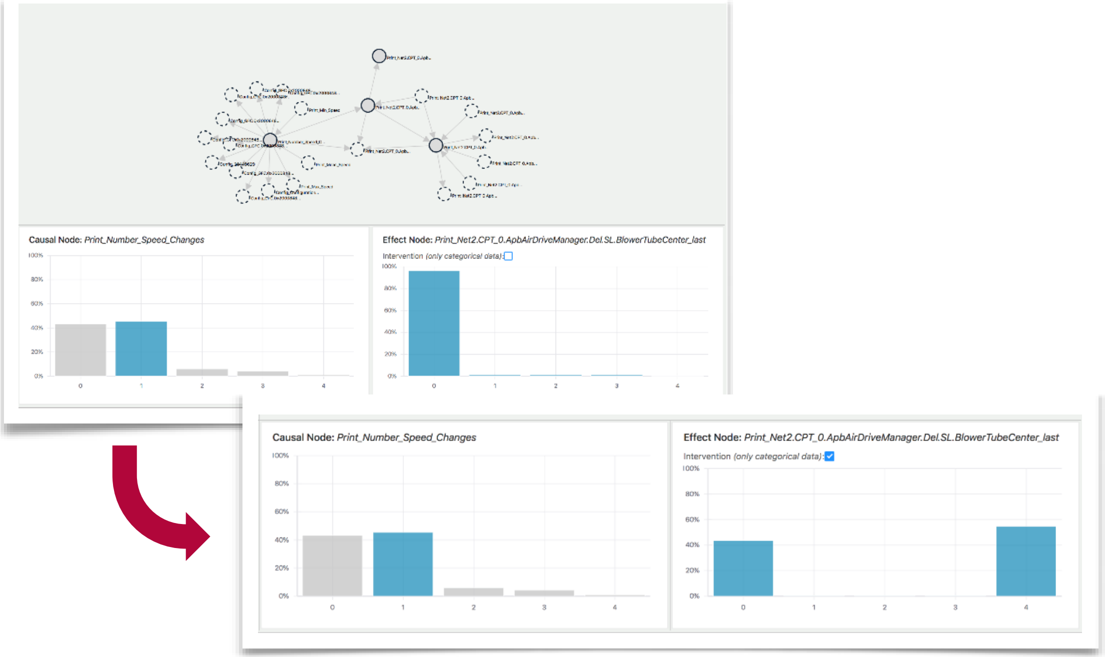
$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$

$P(X_4 | do(X_2 = x_2))$

Functional Systems

$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$

$f_2(x_3, x_4) = \quad \ldots$



2. Causal Inference in Application
Functional Systems

"What are the time lags within climate processes that generate local air pressures ?"

Causal Relationships

Probabilistic Inference

$P(X_3 | X_1 = x_1, X_2 = x_2)$

$P(X_4 | X_2 = x_2)$

Causal Inference

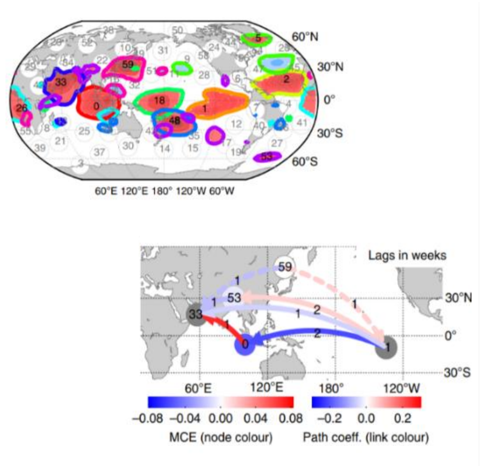
$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$

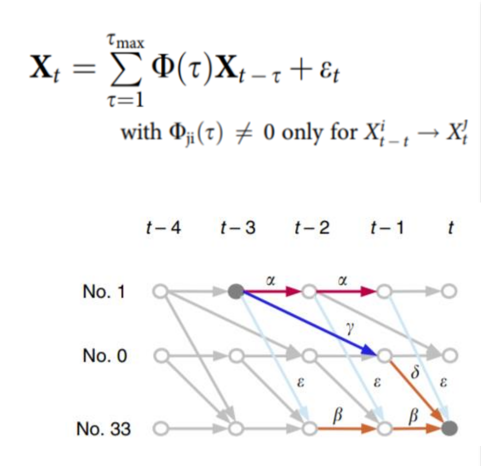
$P(X_4 | do(X_2 = x_2))$

Functional Systems

$f_1(x_1, x_2) = e^{\alpha x_1} + \beta x_2 + \gamma$

$f_2(x_3, x_4) = \quad \ldots$





Runge et. al. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems.

2. Causal Inference in Application
Lecture Example

Scope

- Mathematical concepts determine a conceptual causal inference procedure
- A simple example accompanies our lecture
  - will be extended when needed
  - you are invited to work in a personal notebook

Scenario: The causal relationships in a cooling house

Content

1. Introduction to R
2. Use Case
3. Causal Graphical Models
4. Conditional Independence Testing
5. Constraint-based Causal Structure Learning
6. Causal Inference on Causal Graphs
7. Further Opportunities of Causal Structures

2. Causal Inference in Application
Jupyter Notebook

Causal Inference - Theory and Applications

In our lecture Causal Inference - Theory and Applications, we look at the mathematical concepts that build the basis of causal inference.

Causal Inference in Application

We now look how these concepts are applied on observational data to derive causal relationships and how to use the do-operator to receive an estimation of the causal effect. In order to give you an overview on therelated procedure, this notebook gives a step by step approach in the context of a simple cooling house example.

Table of Contents

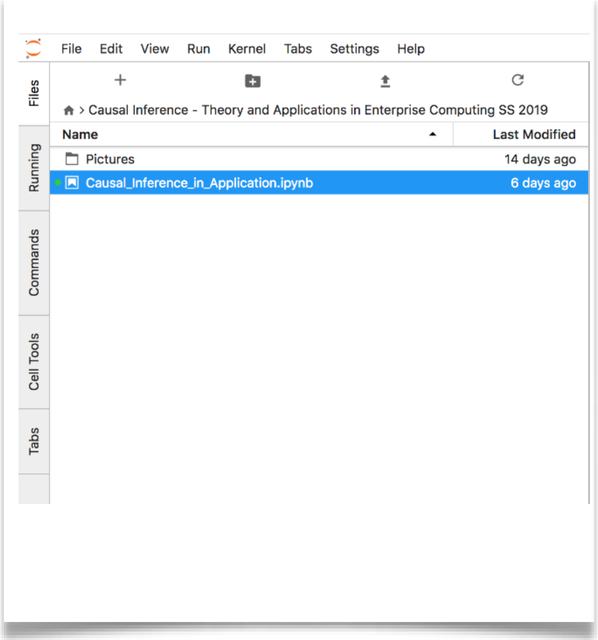
2. Causal Inference in Application
Access Information

System

http://vm-k8s-ctrl.eaalab.hpi.uni-potsdam.de:31157/

Procedure

1. Login via LDAP (standard HPI credentials)
2. Send email to christopher.schmidt@hpi.de
3. We copy you the Master Notebook into your user space for you to work with
4. Adapt and work in your own notebooks
5. Let us know if you require new packages or if anything does not work, as intended

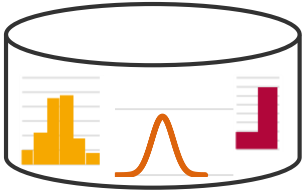


3. Introduction to Research Topics

3. Introduction to Research Topics
Overview on Topics

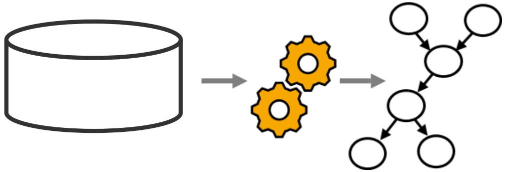
- ***Data, Distributions, Independence***

  Work on topics in the application of learnt techniques beyond the examples given in this lecture (e.g., heterogeneous data distributions)



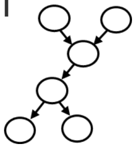
- ***Causal Structure-Learning***

  Work on topics in the context of performance improvements of causal structure learning algorithms (e.g., hardware acceleration)



- ***Applications Scenarios***

  Work on challenges and opportunities in the application of causal inference techniques on real-world data (e.g., industrial manufacturing)

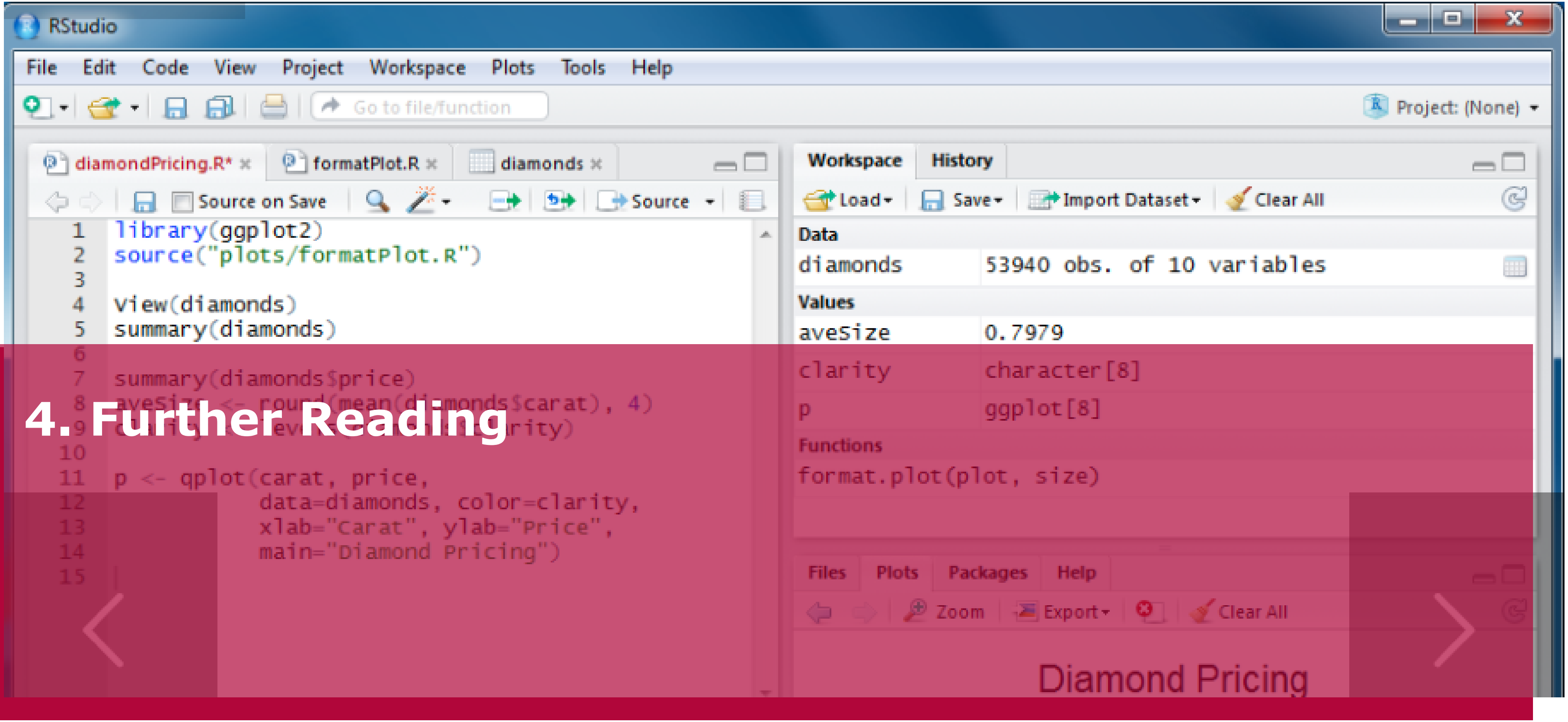


3. Introduction to Research Topics
Topic Application

- ***How to work on a topic?***
  1. Understand theoretic basis and your selected topic
  2. Work on implementation
  3. Present results
  4. Write scientific report in a review process

- ***How to apply for a topic?***
  - Build groups of around three students
  - Send prioritized list of top 3 topics to Johannes Huegle until: *Fri April 26, 11.59 PM*
  - Topic Assignments: *Tue April 30, 9:00 AM*

4. Further Reading



3. Further Reading
Programming

R

- Torfs et. al. (2014), A (very) short introduction to R.
- Venables et. al. (2018), An Introduction to R- Notes on R: A Programming Environment for Data Analysis and Graphics.
- Kalisch et. al. (2017), Package ‘pcalg’.
- Kalisch et. al. (2017), Causal Inference using Graphical Models with the Package pcalg, Journal of Statistical Software.
- Scutari (2007), Learning Bayesian Networks with the bnlearn R Package.

Thank you
for your attention!