



Causal Inference Theory and Applications in Enterprise Computing

Christopher Hagedorn, Johannes Huegle, Dr. Michael Perscheid

April 29, 2020

Agenda

April 29, 2020

- **Recap: Causal Inference in a Nutshell**

- **Causal Inference in Applications**
 1. Research Insights
 2. Lecture Scenario

- **Jupyter Lab**
 1. Access Information
 2. Introduction to R
 3. Replicated RStudio



1. Recap: Causal Inference in a Nutshell

1. Recap: Causal Inference in a Nutshell

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**

Causal Inference
Theory and Applications
in Enterprise Computing

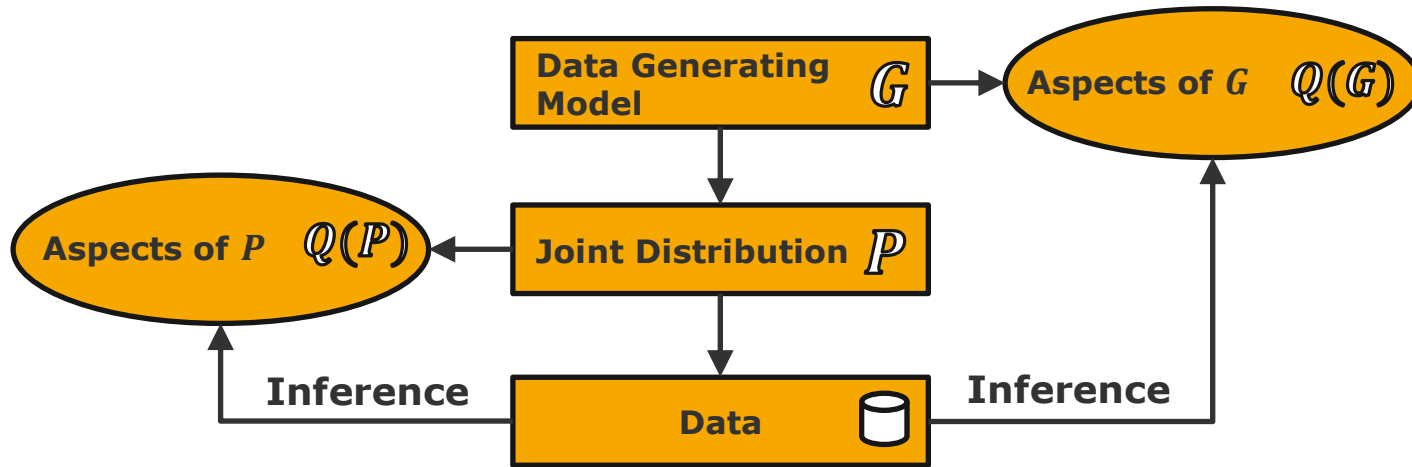
Hagedorn, Huegle,
Perscheid

1. Recap: Causal Inference in a Nutshell

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?

$$Q(P) = P(\text{recovery}|\text{lemons})$$

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

$$Q(G) = P(\text{recovery}|\text{do}(\text{lemons}))$$

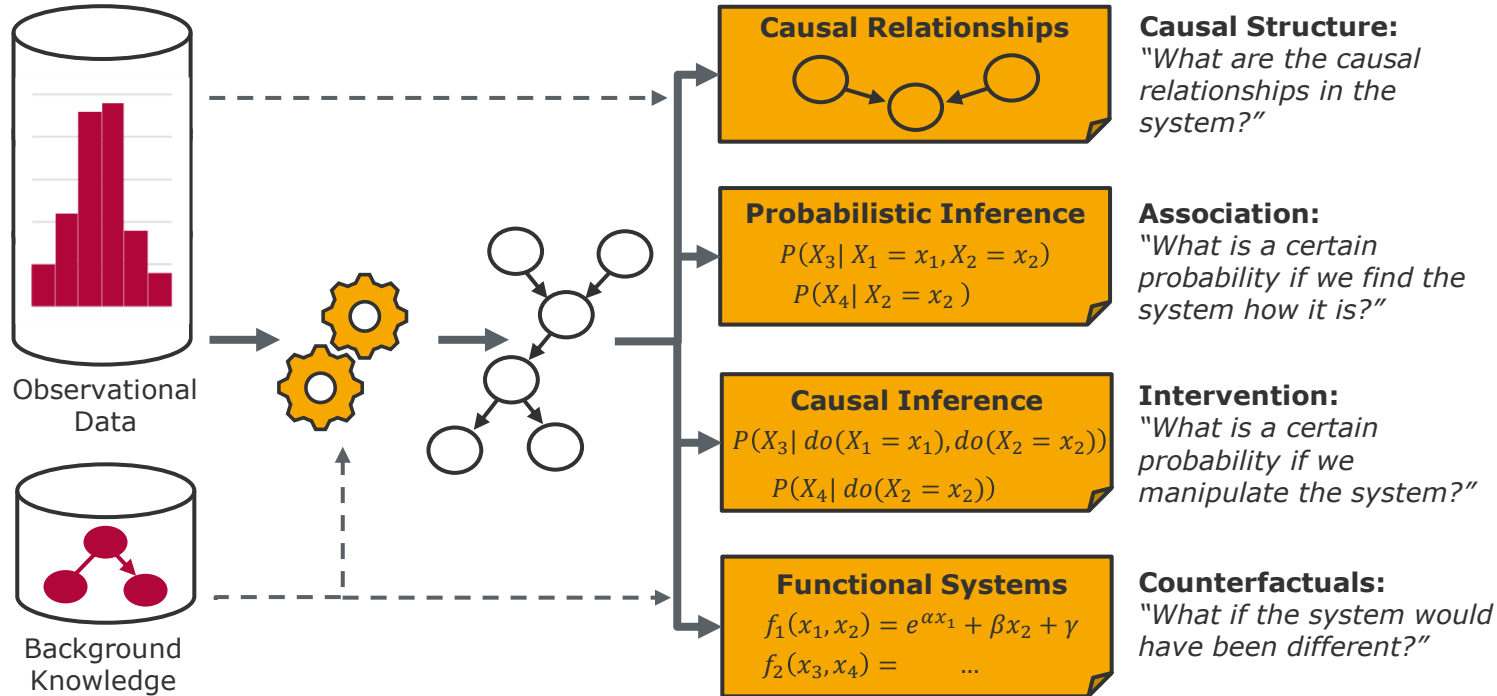
Causal Inference
Theory and Applications
in Enterprise Computing

Hagedorn, Huegle,
Perscheid

Slide 5

1. Recap: Causal Inference in a Nutshell

Inference Procedure



Causal Inference
Theory and Applications
in Enterprise Computing

Hagedorn, Huegle,
Perscheid

Slide 6

Data

Causal Structure Learning

Opportunities

Select nodes



Re-Layout

Freeze Layout

Focused Nodes:

id ×

CVP ×

LVV ×

PCWP ×

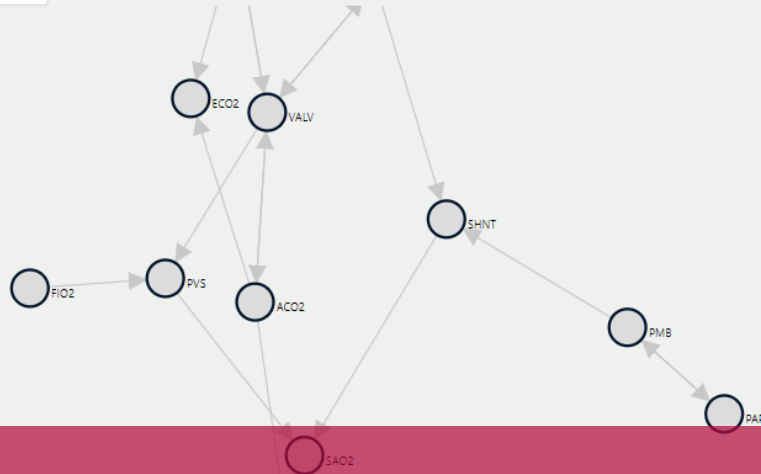
HIST ×

LVF ×

TPR ×

CCHL ×

BP ×



2. Causal Inference in Applications



2. Causal Inference in Applications

Research Insights - Topics

Data-Driven Causal Inference

Concepts and Methods

Research Objectives:

- **Improvement of flexibility** of CSL to address real-world settings, e.g., entropy-based CI tests
- **Improvement of applicability** to real-world setting, e.g., through the implementation of an evaluation pipeline

Hardware-Acceleration

Research Objectives:

- **Improvement of performance** of CSL through parallel execution on multi-core CPUs or GPUs
- **Improvement of scalability** of GPU-accelerated CSL, e.g., executing on multiple GPUs or overcoming on-chip memory limits

Transfer to application and validation
together with cooperation partner, e.g., case studies in real-world setting

Evaluation Pipeline for Causal Structure Learning

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

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

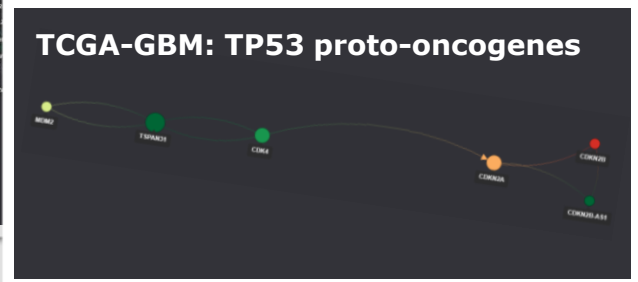
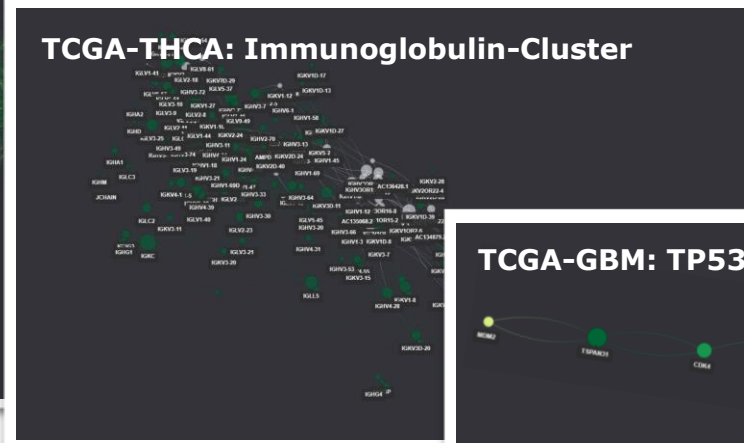
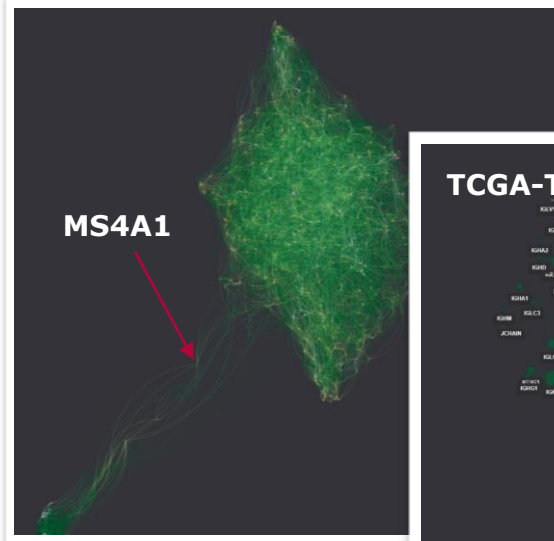
Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle,
Perscheid

2. Causal Inference in Applications

Research Insights - Genetics

"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

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

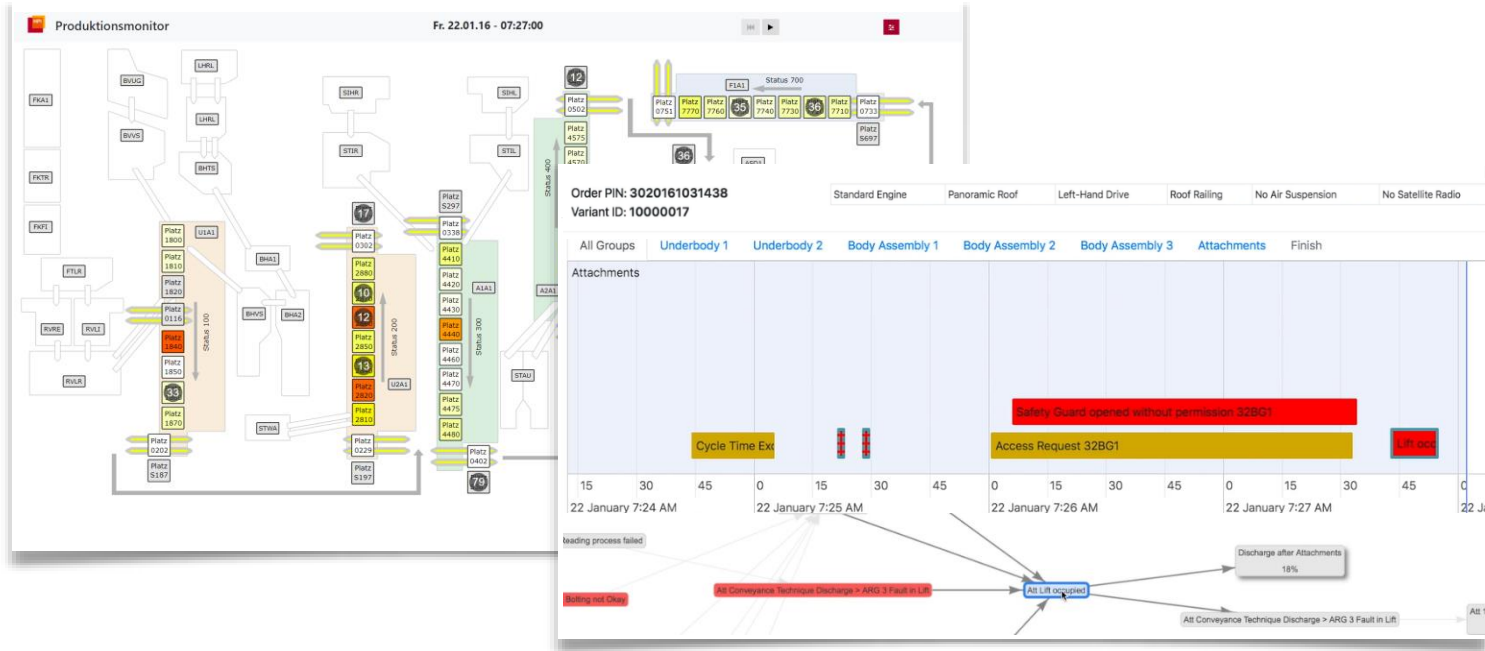
Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle, Perscheid

2. Causal Inference in Applications

Research Insights – Automotive Manufacturing (I/II)

"What are critical causal relationships between failures and quality deviations within a body shop assembly line?"



Causal Relationships

Probabilistic Inference

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

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

Causal Inference

$$P(X_3 \text{ do}(X_1 = x_1), \text{do}(X_2 = x_2))$$

$$P(X_4 | \text{do}(X_2 = x_2))$$

Functional Systems

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

$$f_2(x_3, x_4) = \dots$$

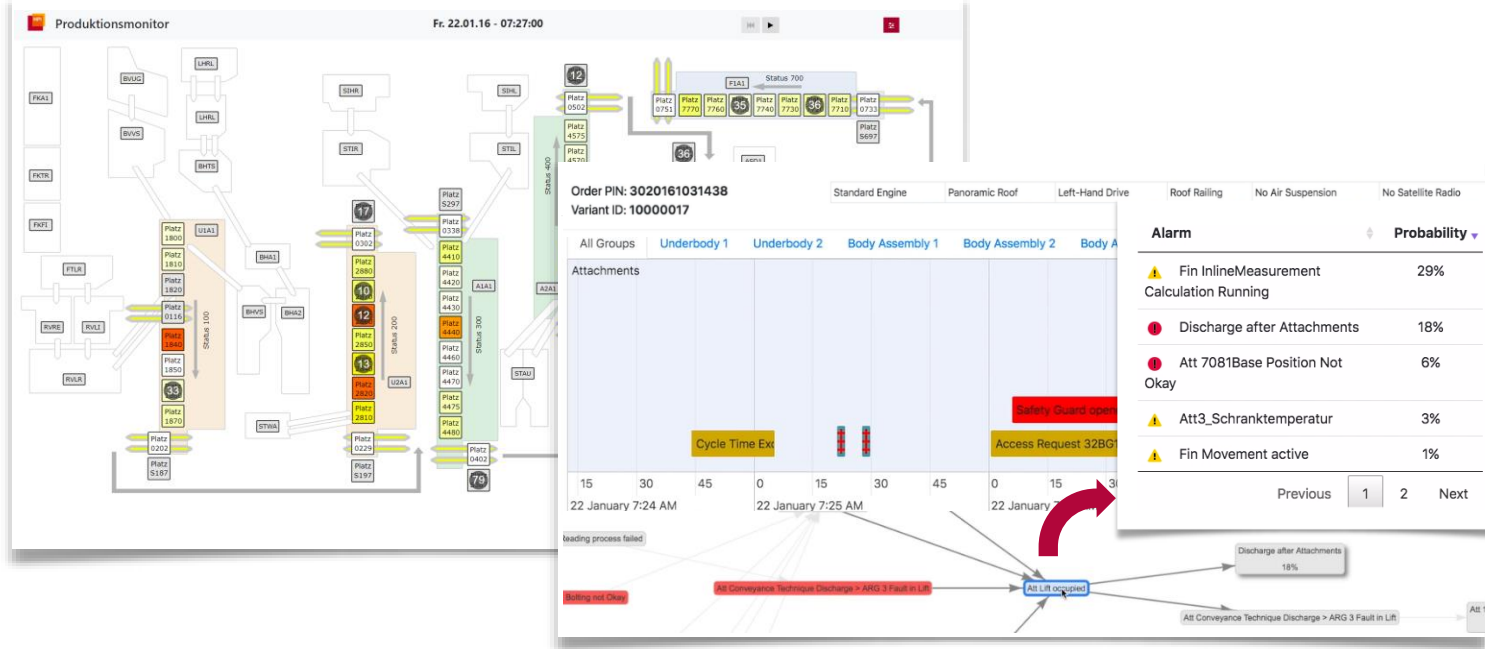
Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle, Perscheid

2. Causal Inference in Applications

Research Insights – Automotive Manufacturing (II/II)

"Given the knowledge about critical causal relationships what is the probability of failures and quality deviation in the current 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

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

$$f_2(x_3, x_4) = \dots$$

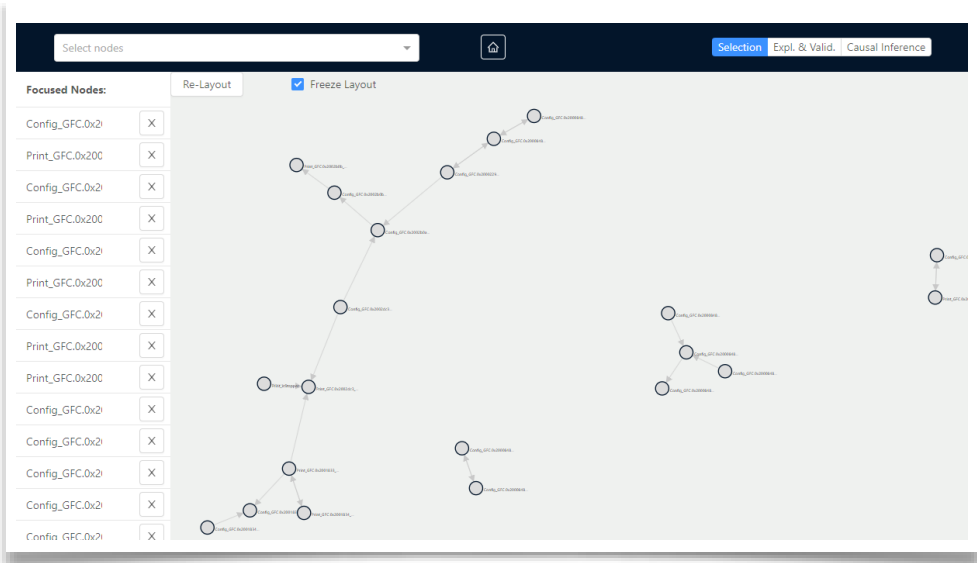
Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle, Perscheid

2. Causal Inference in Applications

Research Insights - Mechanical Engineering (I/III)

"How are configurations of a printing press and manual adjustments causally related to the quality of the printing output and stopper events?"



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

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

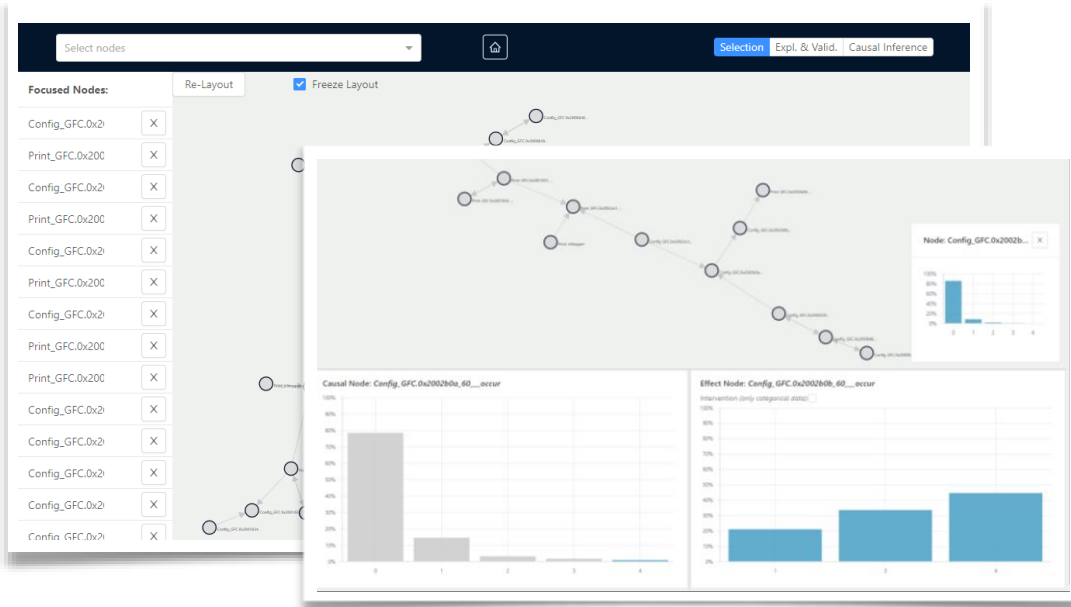
**Causal Inference
Theory and Applications
in Enterprise Computing**

Hagedorn, Huegler,
Perscheid

2. Causal Inference in Applications

Research Insights - Mechanical Engineering (II/III)

"What are the observed distribution characteristics of printing output and stopper events given a specific configuration of a printing press?"



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

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

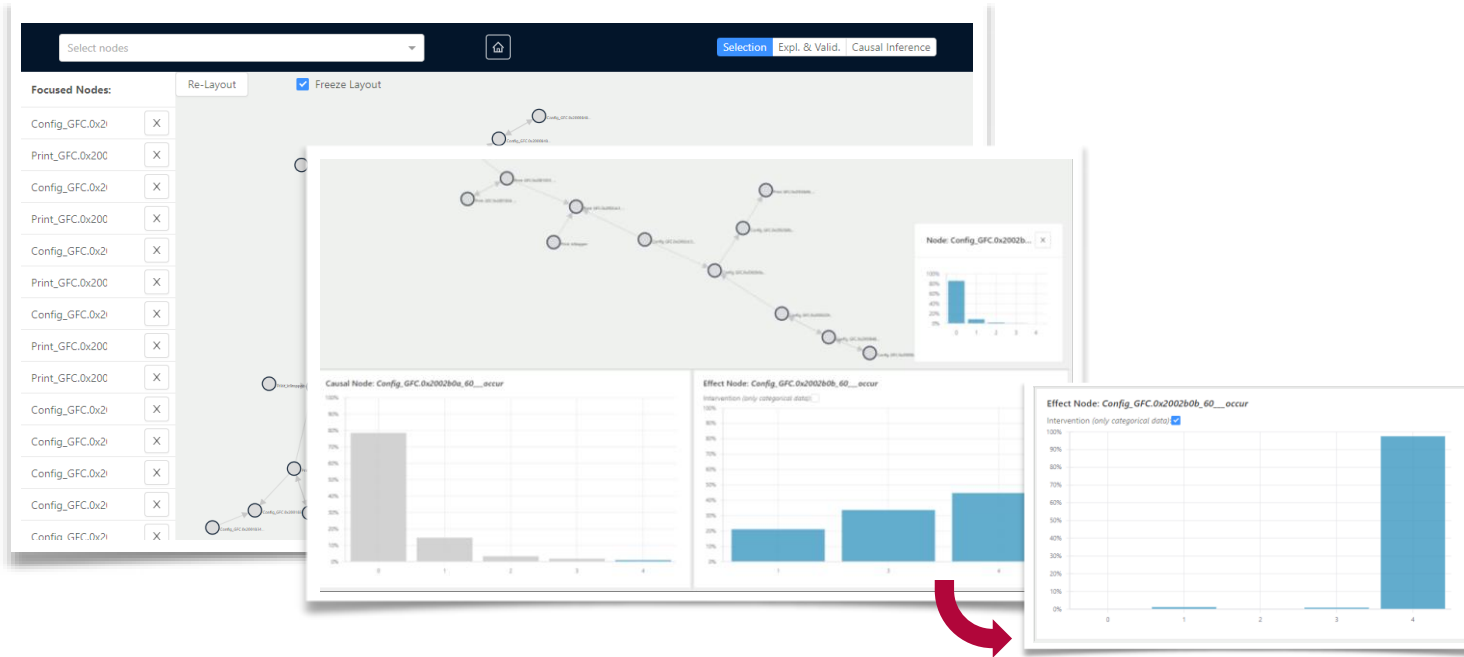
**Causal Inference
Theory and Applications
in Enterprise Computing**

Hagedorn, Huegle,
Perscheid

2. Causal Inference in Applications

Research Insights - Mechanical Engineering (III/III)

"What is the direct causal effect of an intervention on printing press configuration to the quality of the printing output?"



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

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

Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegler, Perscheid

2. Causal Inference in Applications

Lecture Scenario - Overview

A Cooling House

- Step by step causal inference walkthrough given the simple example of a cooling house
- Solutions to ubiquitous questions of causal inference in application scenarios:
 1. What are the causal relationships between the variables our system?
 2. How to derive these causal relationships from observational data?
 3. What are causal effects in our system?
 4. How to estimate the effect of interventions?
 5. What are omnipresent challenges of causal inference in application scenarios?



Probabilistic Inference

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

Causal Inference

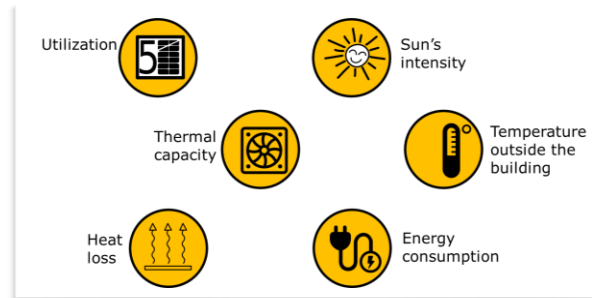
$$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$$
$$P(X_4 | do(X_2 = x_2))$$

Functional Systems

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

Variables Defining our Energy System

- V_1 Utilization of the cooling house
- V_2 Sun's intensity
- V_3 Temperature outside the building
- V_4 Thermal capacity of the cooling house
- V_5 Heat loss
- V_6 Energy consumption



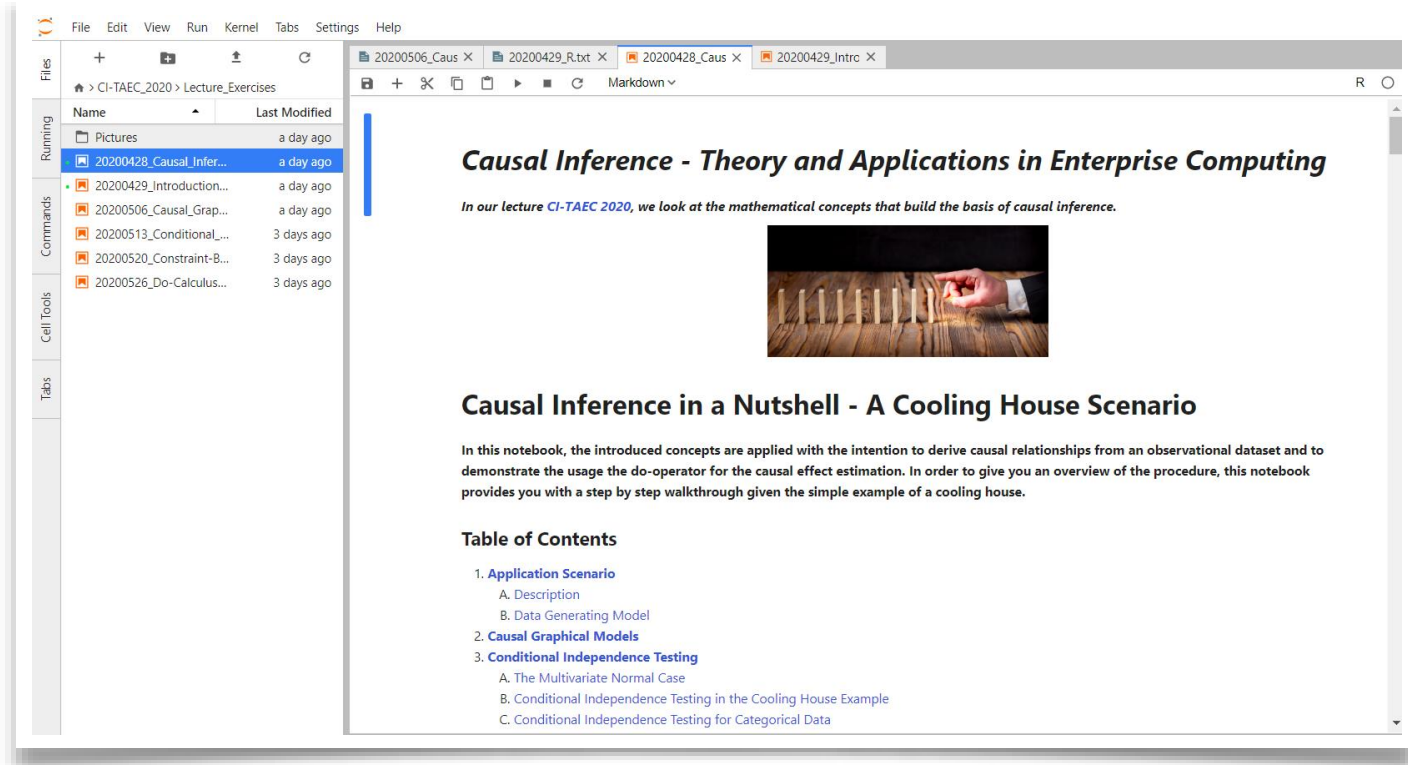
Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle, Perscheid

Slide 15

2. Causal Inference in Applications


Lecture Scenario – Jupyter Notebook



The screenshot shows a Jupyter Notebook interface. On the left, there is a file explorer showing a directory structure with files like '20200428_Causal_Infer...', '20200429_Introduction...', '20200506_Causal_Grap...', '20200513_Conditional...', '20200520_Constraint-B...', and '20200526_Do-Calculus...'. The main notebook area displays the following content:

Causal Inference - Theory and Applications in Enterprise Computing

In our lecture CI-TAEC 2020, we look at the mathematical concepts that build the basis of causal inference.



Causal Inference in a Nutshell - A Cooling House Scenario

In this notebook, the introduced concepts are applied with the intention to derive causal relationships from an observational dataset and to demonstrate the usage the do-operator for the causal effect estimation. In order to give you an overview of the procedure, this notebook provides you with a step by step walkthrough given the simple example of a cooling house.

Table of Contents

1. **Application Scenario**
 - A. Description
 - B. Data Generating Model
2. **Causal Graphical Models**
3. **Conditional Independence Testing**
 - A. The Multivariate Normal Case
 - B. Conditional Independence Testing in the Cooling House Example
 - C. Conditional Independence Testing for Categorical Data



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

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

Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle, Perscheid



2. Jupyter Lab

2. Jupyter Lab

Access Information

System

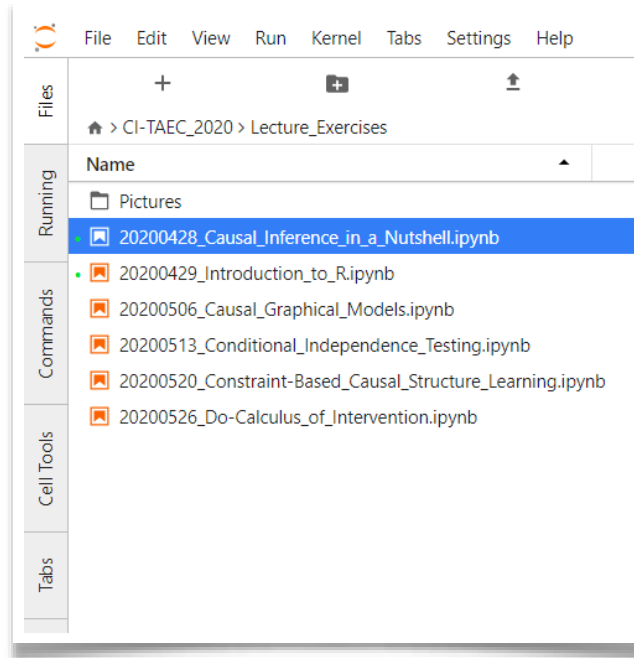
- Link was provided via email

Access

- Login via LDAP (standard HPI credentials)

Exercises (Wednesdays)

1. We copy currently relevant notebooks including exercises into your own user space
2. Adapt and work on the exercises in your own notebooks
3. Together, we discuss challenges, ideas and solution proposals
4. A solution is provided in your Jupyter Lab file system afterwards



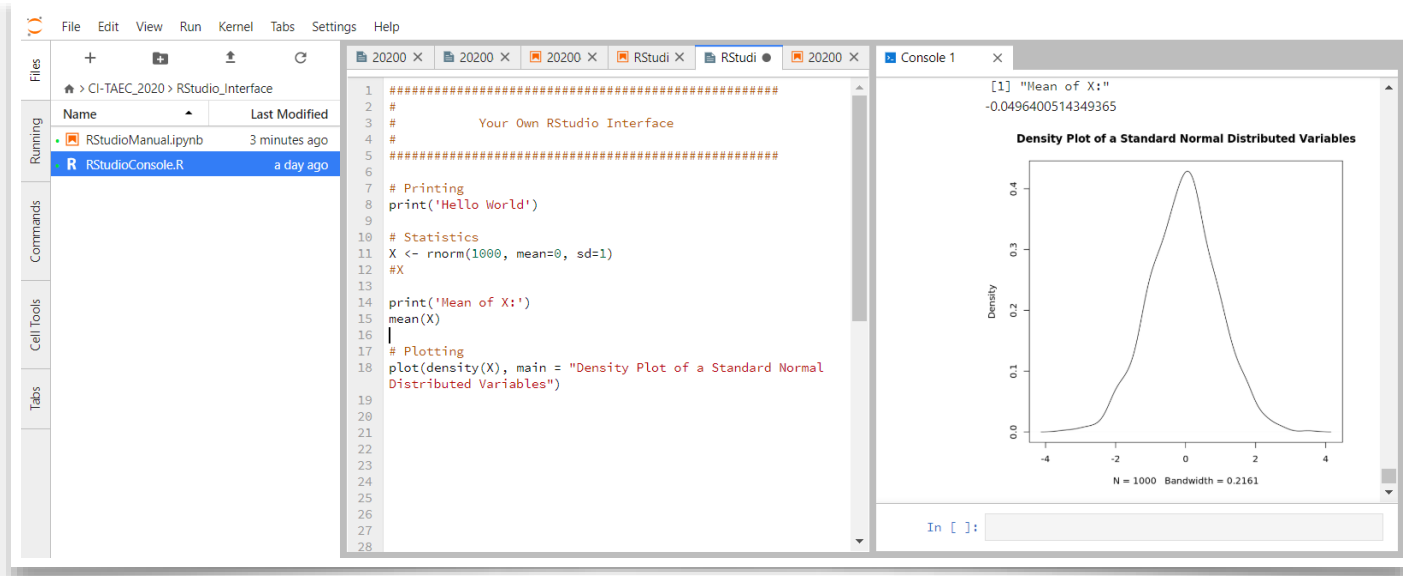
Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle,
Perscheid

2. Jupyter Lab RStudio

Replicated RStudio Environment

- Take the opportunity to strengthen your R programming skills in your own environment
- Let us know if you require new packages or if anything does not work, as intended



The screenshot displays the RStudio environment with the following components:

- Files Panel:** Shows a directory structure for 'CI-TAEC_2020 > RStudio_Interface' with files like 'RStudioManual.ipynb' and 'RStudioConsole.R'.
- Source Editor:** Contains R code for generating a standard normal distribution and plotting its density. The code includes comments and function calls like `rnorm()`, `print()`, and `plot(density(X))`.
- Console:** Shows the output of the code execution, including the mean of the generated data and the title of the plot.
- Plot:** A density plot titled 'Density Plot of a Standard Normal Distributed Variables' showing a bell-shaped curve centered at 0. The x-axis ranges from -4 to 4, and the y-axis (Density) ranges from 0.0 to 0.4. The plot includes the text 'N = 1000 Bandwidth = 0.2161'.

**Causal Inference
Theory and Applications
in Enterprise Computing**

Hagedorn, Huegle,
Perscheid

Slide 19

2. Jupyter Lab

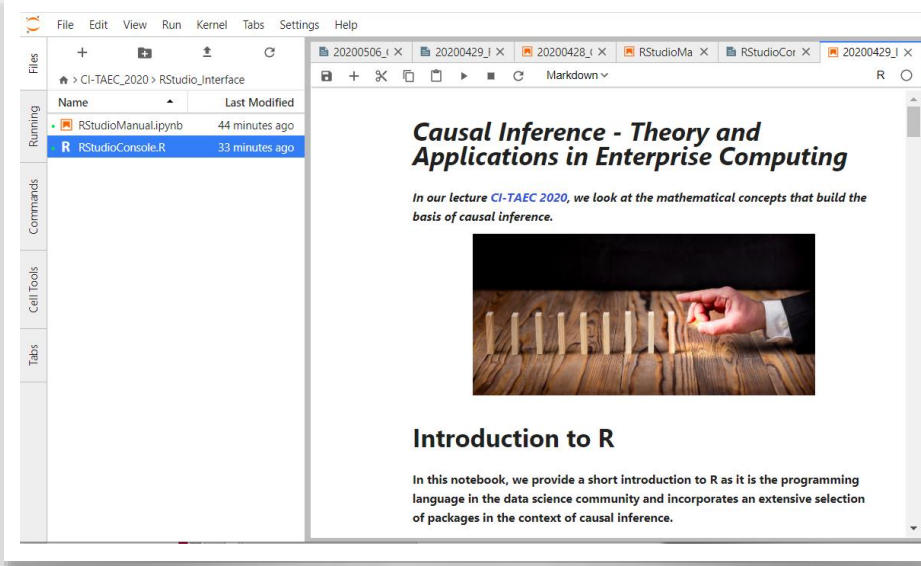
Introduction to R

What is R?

- Free Software under the terms of GNU General Public License
- R provides a wide variety of statistical and graphical techniques, see [CRAN](#)

Table of Contents


- Getting Started
- The Basics
- Exercises
- Further Reading



The screenshot shows a Jupyter Lab window with a file explorer on the left and a notebook in the center. The file explorer shows a directory structure with files like 'RStudioManual.ipynb' and 'RStudioConsole.R'. The notebook content includes a title, a subtitle, a paragraph of text, an image of a hand tipping dominoes, and another title 'Introduction to R' with a short introductory paragraph.

Causal Inference - Theory and Applications in Enterprise Computing

In our lecture CI-TAEC 2020, we look at the mathematical concepts that build the basis of causal inference.



Introduction to R

In this notebook, we provide a short introduction to R as it is the programming language in the data science community and incorporates an extensive selection of packages in the context of causal inference.

Causal Inference Theory and Applications in Enterprise Computing

Hagedorn, Huegle,
Perscheid

Slide 20

Thank you
for your attention!