Causal Inference
Theory and Applications in Enterprise Computing

Christopher Hagedorn, Johannes Huegle, Dr. Michael Perscheid

April 29, 2020

Hasso Plattner Institut
Digital Engineering • Universität Potsdam

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

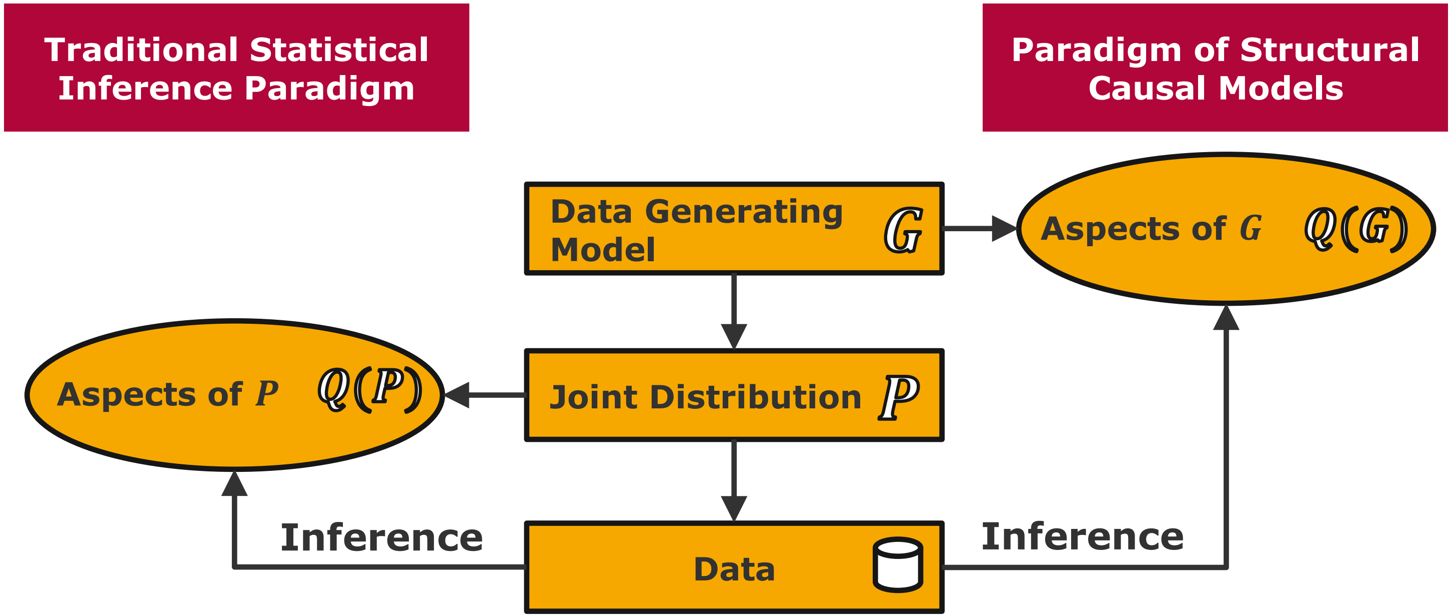
Traditional statistics, machine learning, etc.

- 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

1. Recap: Causal Inference in a Nutshell
Concept



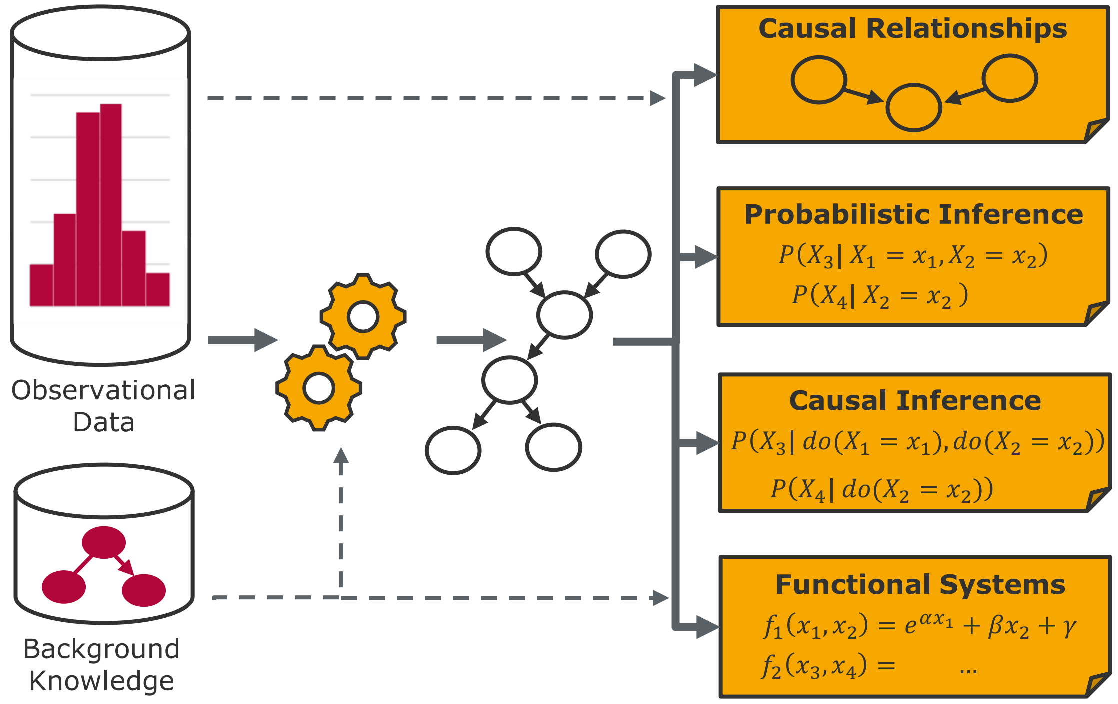
E.g., what is the sailors' probability of recovery when **we see** a treatment with lemons?

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

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

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

1. Recap: Causal Inference in a Nutshell
Inference Procedure



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

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

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

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

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

HPI Hasso Plattner Institut

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

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

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

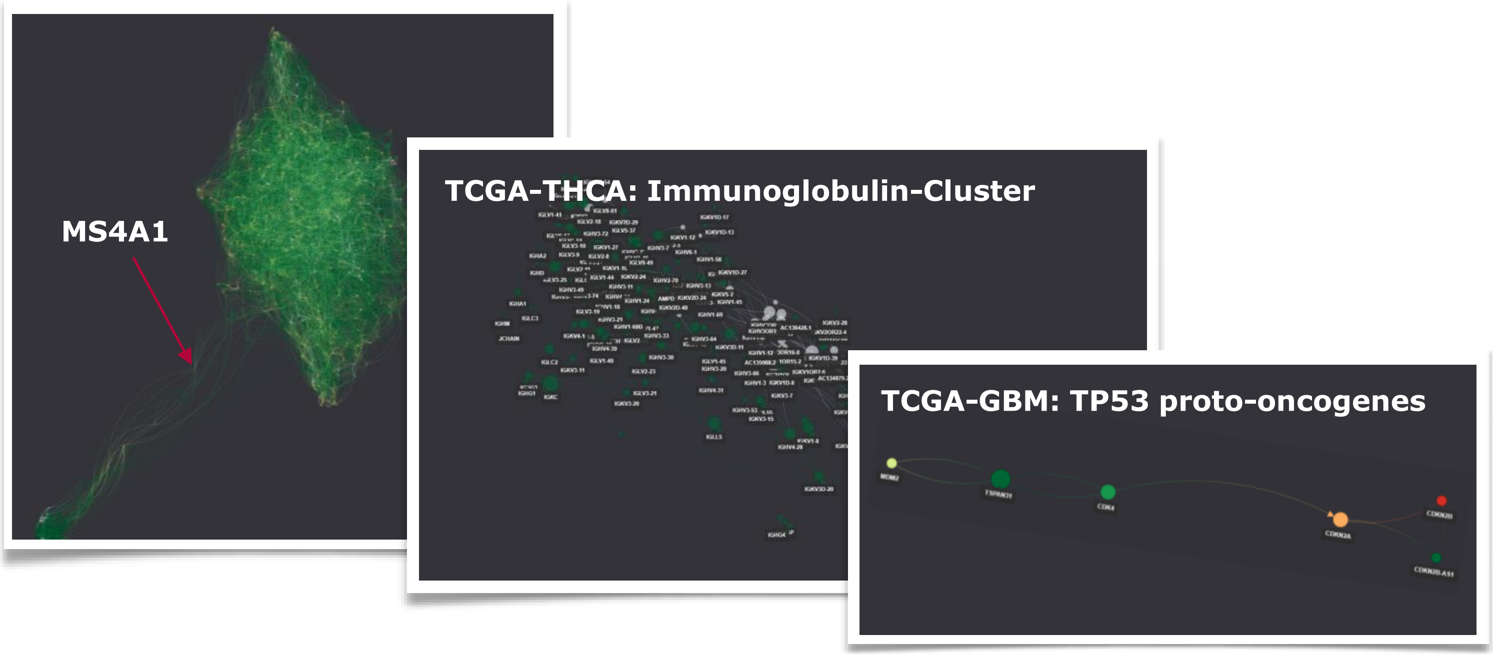
Functional Systems

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

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

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

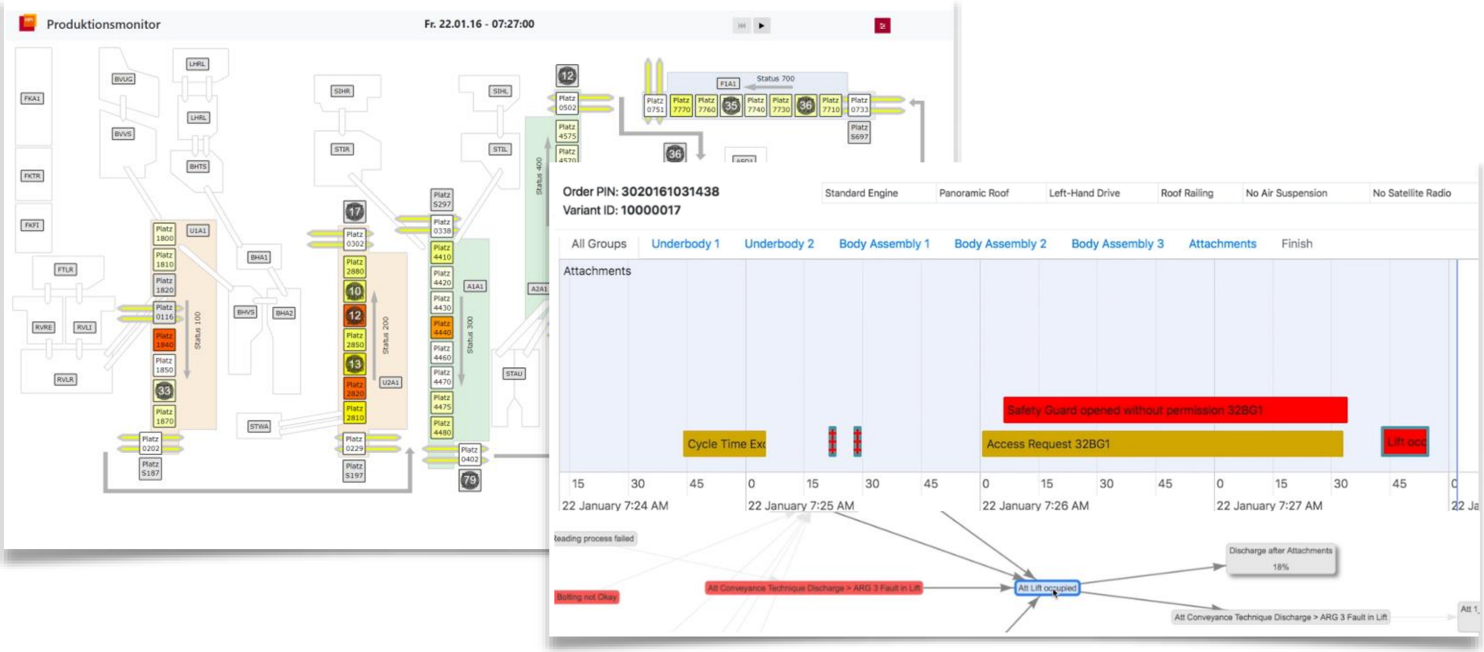
Causal Inference
$P(X_3 | do(X_1 = x_1), do(X_2 = x_2))$
$P(X_4 | do(X_2 = x_2))$

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

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

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

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

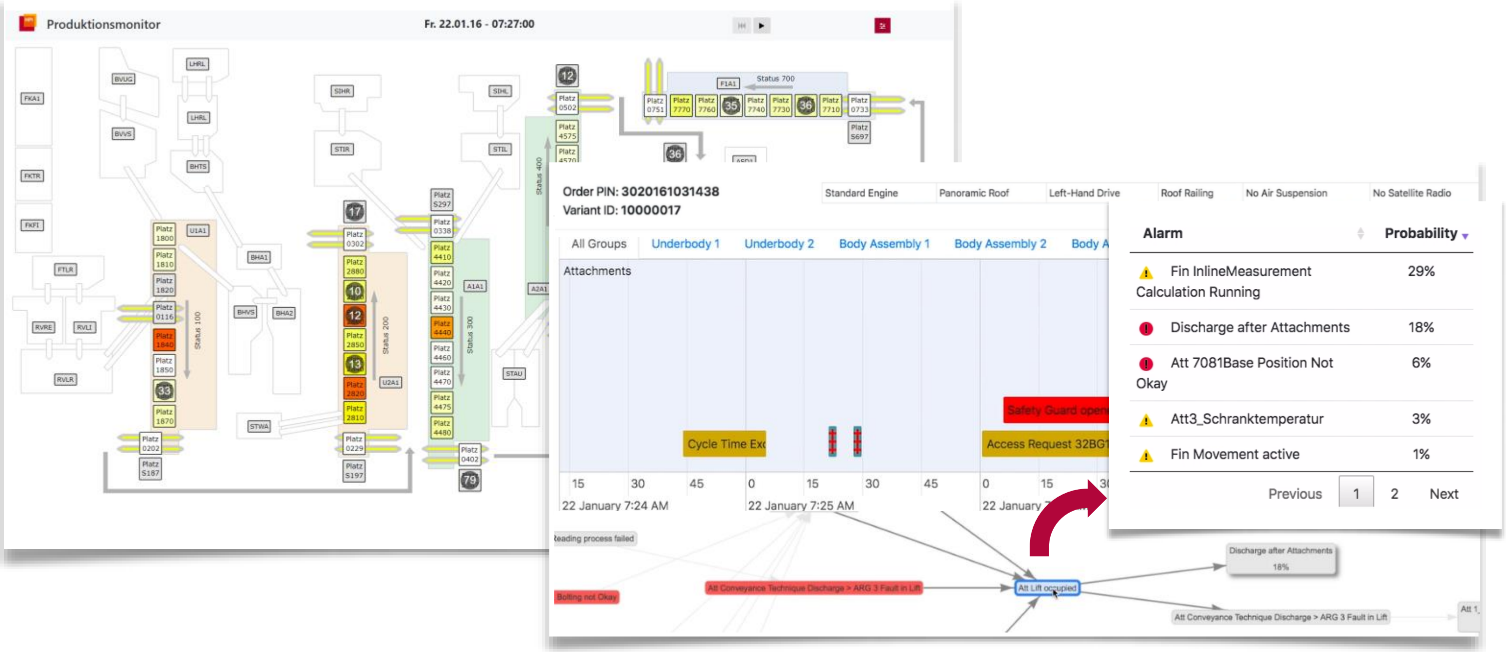
Functional Systems

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

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

2. Causal Inference in 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

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

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

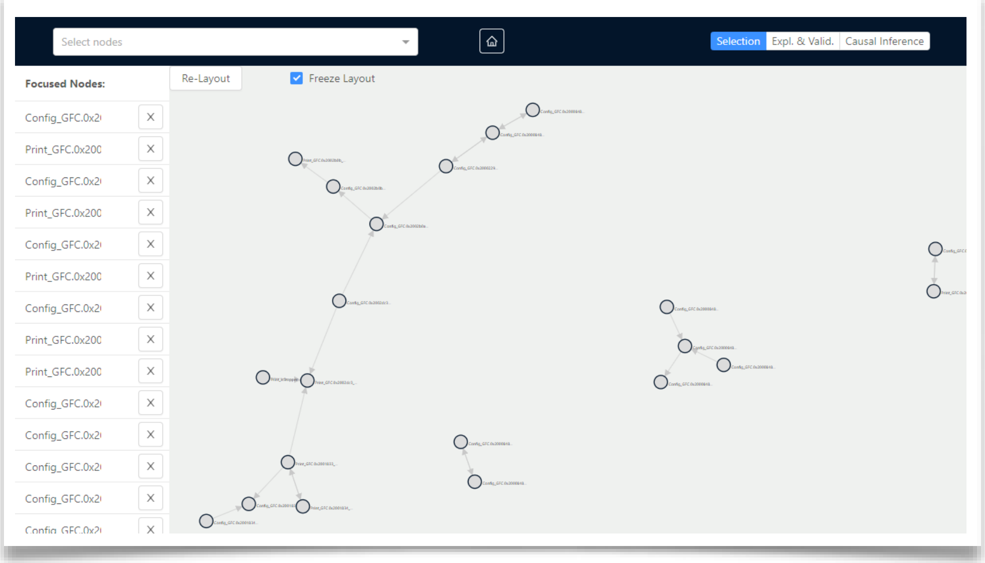
Functional Systems

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

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

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?"



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

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

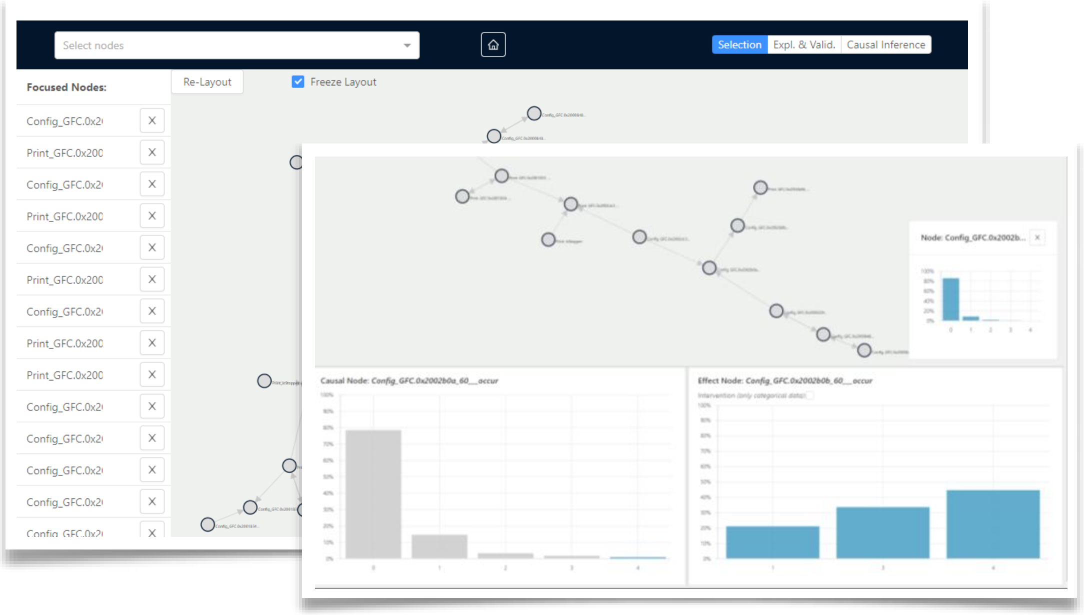
$f_2(x_3, x_4) = \quad \ldots$

Causal Inference
Theory and Applications in Enterprise Computing

Hagedorn, Huegle, 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?"



Causal Relationships

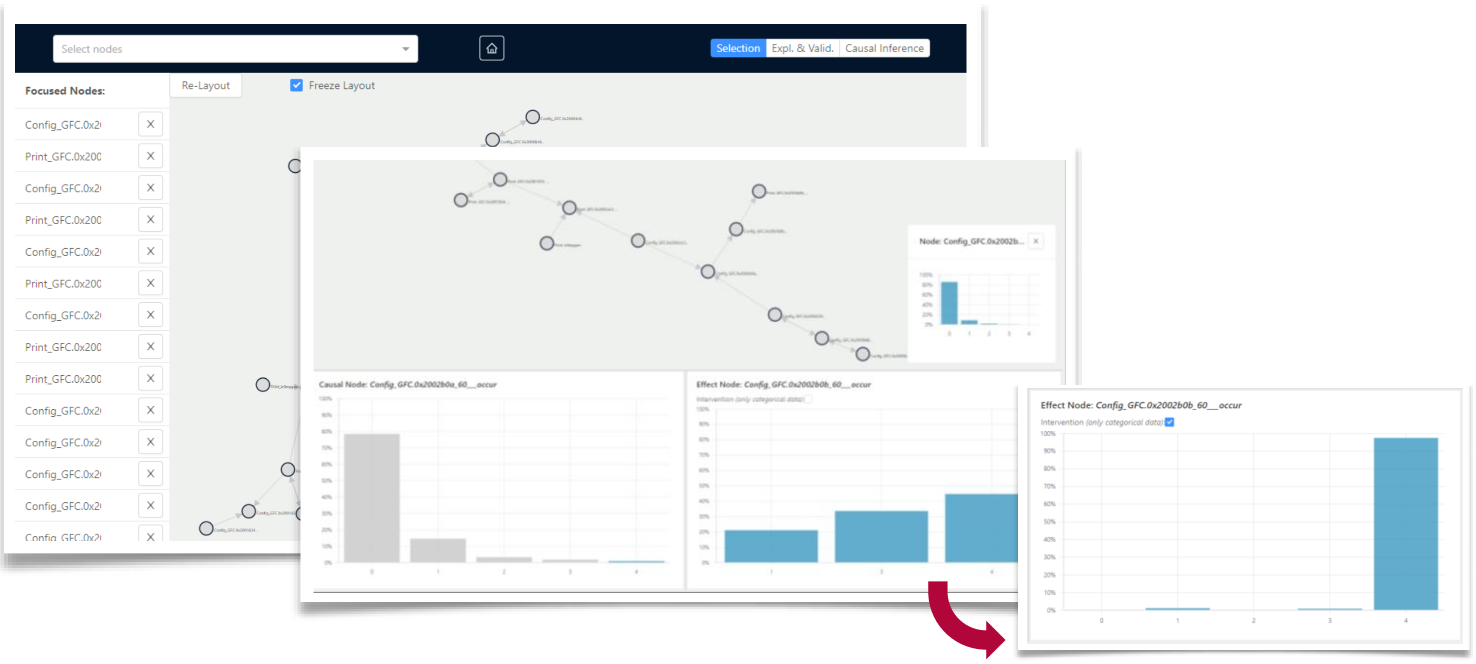
Probabilistic Inference
$P(X_3 | X_1 = x_1, X_2 = x_2)$
$P(X_4 | X_2 = x_2)$

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 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

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

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

2. Causal Inference in Applications
Lecture Scenario - Overview

Hasso Plattner Institut

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?

Causal Relationships

Probabilistic Inference

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

$P(X_4 | X_2 = x_2)$

Causal Inference

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

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

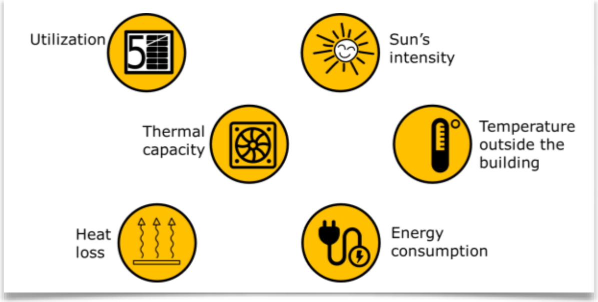
Functional Systems

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

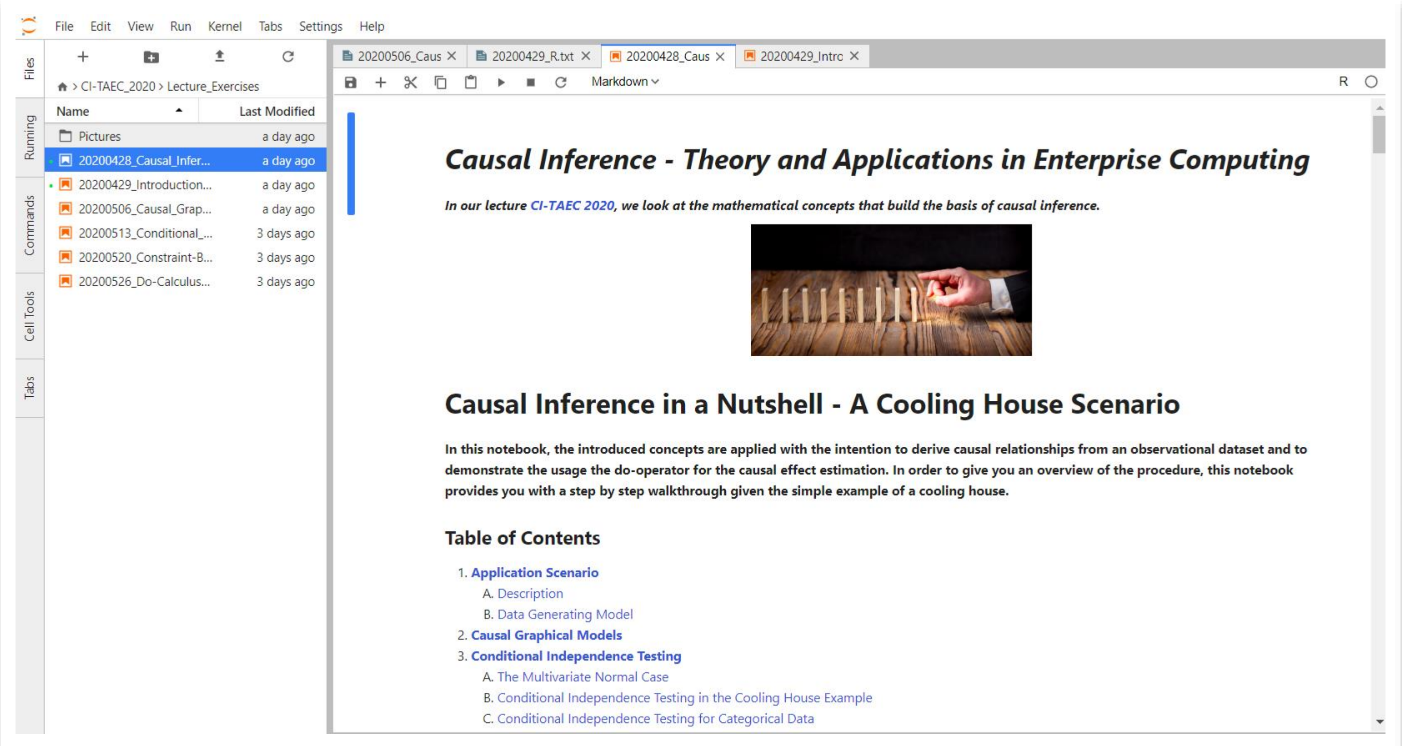
$f_2(x_3, x_4) = \quad ...$

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



2. Causal Inference in Applications
Lecture Scenario – Jupyter Notebook



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

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

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

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

2. Jupyter Lab

2. Jupyter Lab
Access Information

System

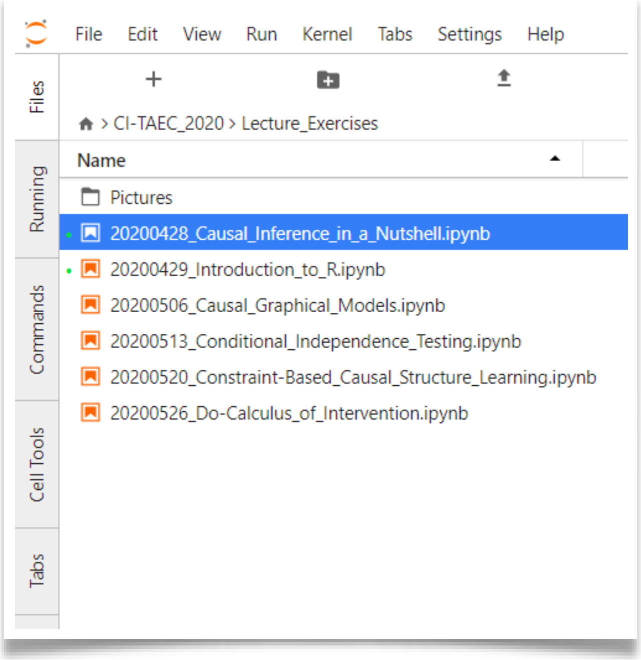
- Link was provided via email

Access

- Login via LDAP (standard HPI credentials)

Exercises (Wednesdays)

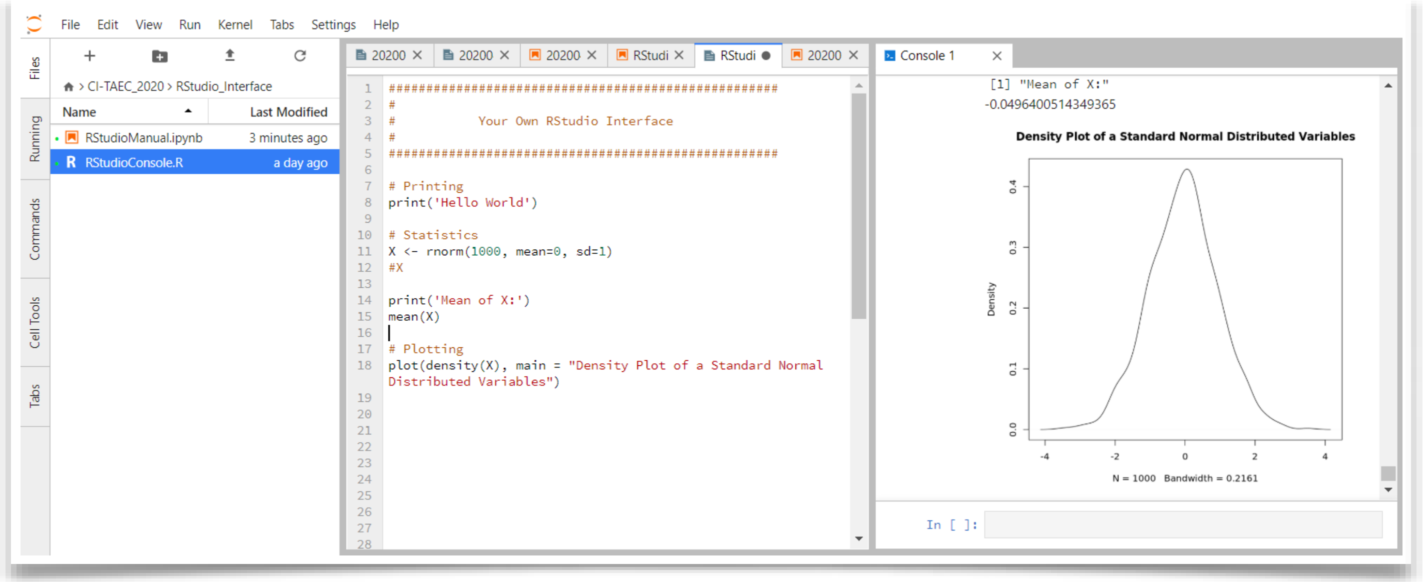
1. We copy currently relevant notebooks including exercises into your own user space
2. Adapt and work on the exercises in your own notebooks
3. Together, we discuss challenges, ideas and solution proposals
4. A solution is provided in your Jupyter Lab file system afterwards



2. Jupyter Lab
RStudio

Replicated RStudio Environment

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



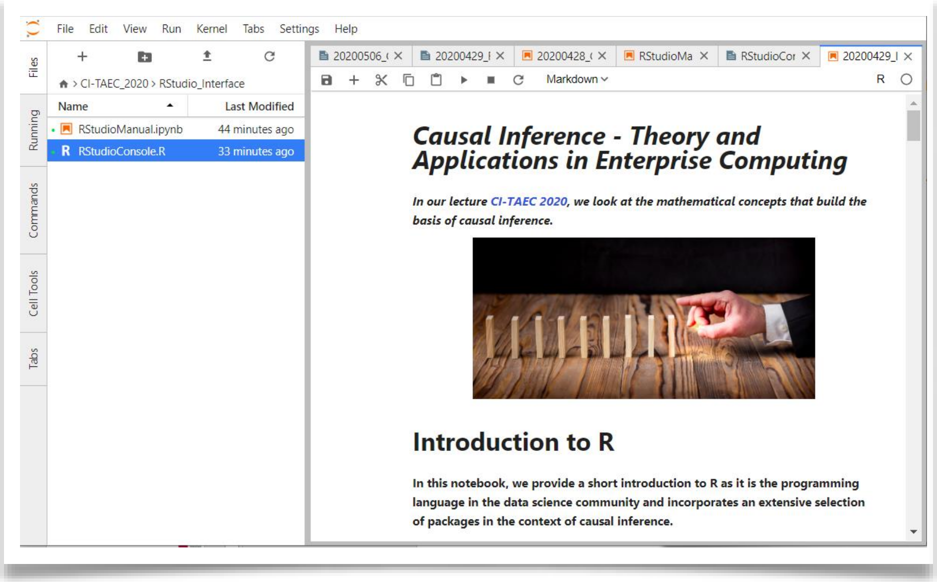
2. Jupyter Lab
Introduction to R

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



Thank you
for your attention!