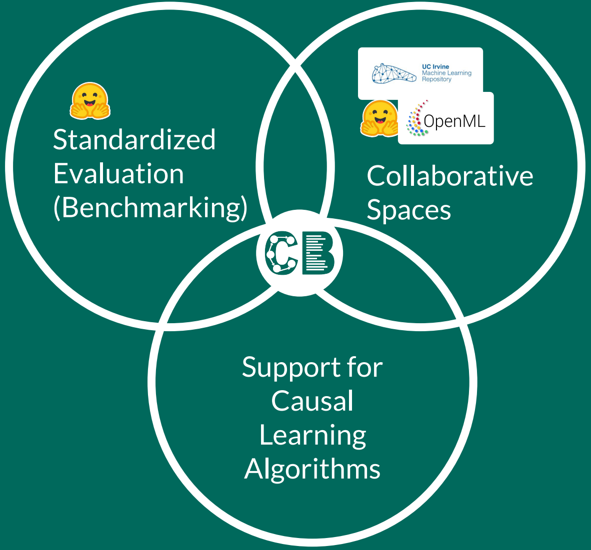
CausalBench: Causal Learning Research Streamlined

Ahmet Kapkiç *
Pratanu Mandal *
Abhinav Gorantla *
Shu Wan
Ertuğrul Çoban
Dr. Paras Sheth
Dr. Huan Liu *
Dr. K. Selçuk Candan

* In-person presenters



tutorial.causalbench.org

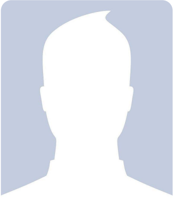
August 3-7, 2025
KDD2025

This research is funded by NSF Grant 2311716, "CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration", and NSF Grants #2230748, "PIRE: Building Decarbonization via AI-empowered District Heat Pump Systems", #2412115, "PIPP Phase II: Analysis and Prediction of Pandemic Expansion (APPEX)" and USACE #GR40695, "Designing nature to enhance resilience of built infrastructure in western US landscapes".

The Team

Ahmet Kapkiç
Ph.D. Student

Pratanu Mandal
Ph.D. Student



Abhinav Gorantla
M.S. Student

Shu Wan
Ph.D. Student

Ertuğrul Çoban
Ph.D. Student

Dr. Paras Sheth
(recent graduate – congrats!)

Dr. Huan Liu
Regents Professor
Arizona State University

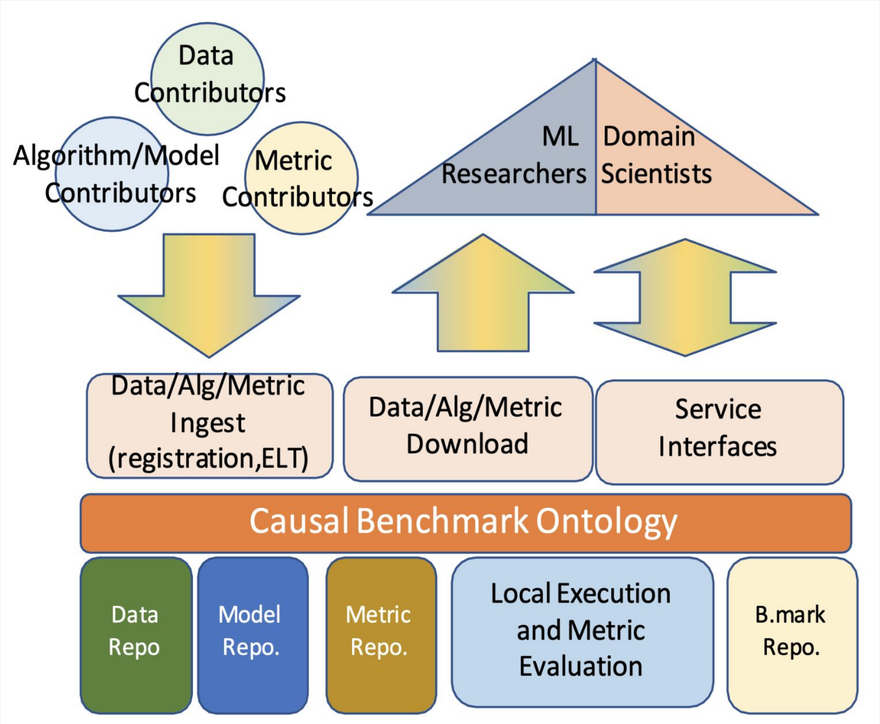
Co-Principal Investigator

Dr. K. Selçuk Candan
Professor
Arizona State University

Principal Investigator

* In-person presenters

What is CausalBench?

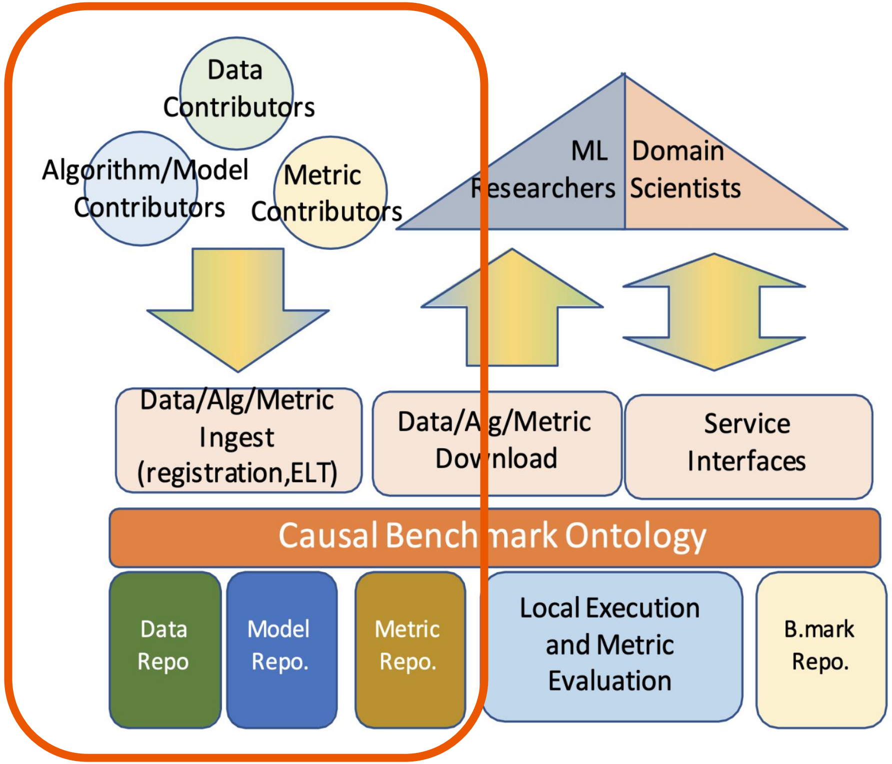


- **CausalBench** is a benchmarking platform for [Causal Learning research](#).

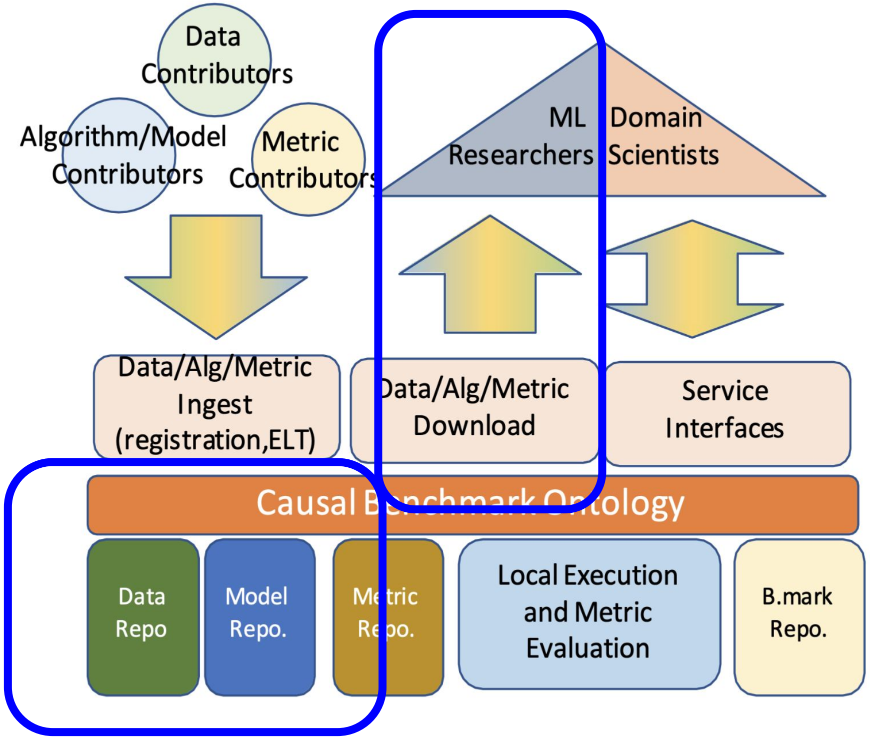
- Goals:
  - Promoting universal adoption of standard datasets, metrics and procedures for causal learning.
  - Facilitating collaboration.
  - Trustable and reproducible benchmarking.
  - Fair and flexible comparison of models.

CausalBench: Use Scenario #1

Data, model, metric contributor

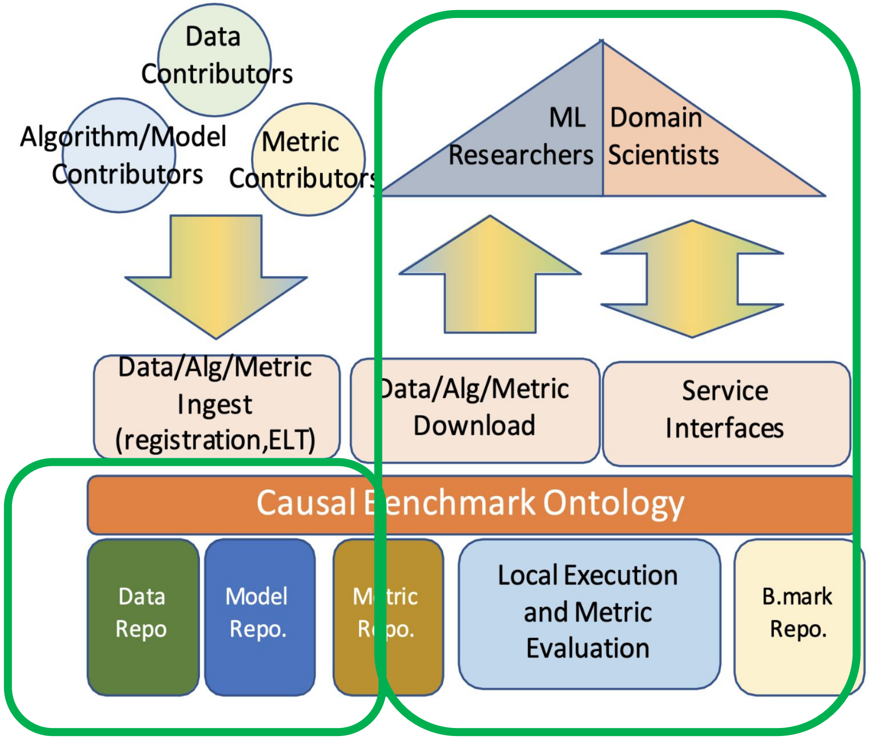


CausalBench: Use Scenario #2



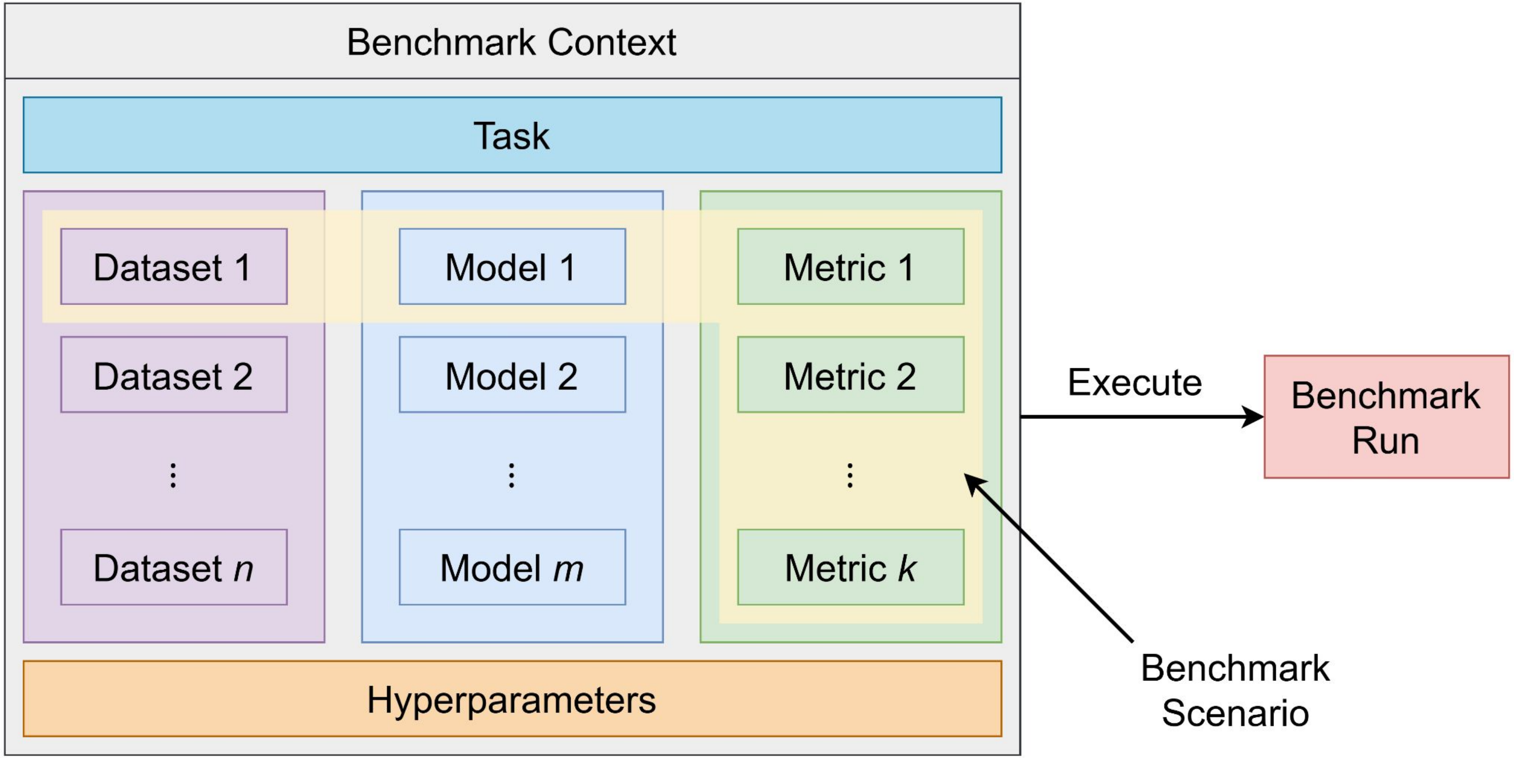
Data, Model, Metric Explorer

CausalBench: Use Scenario #3

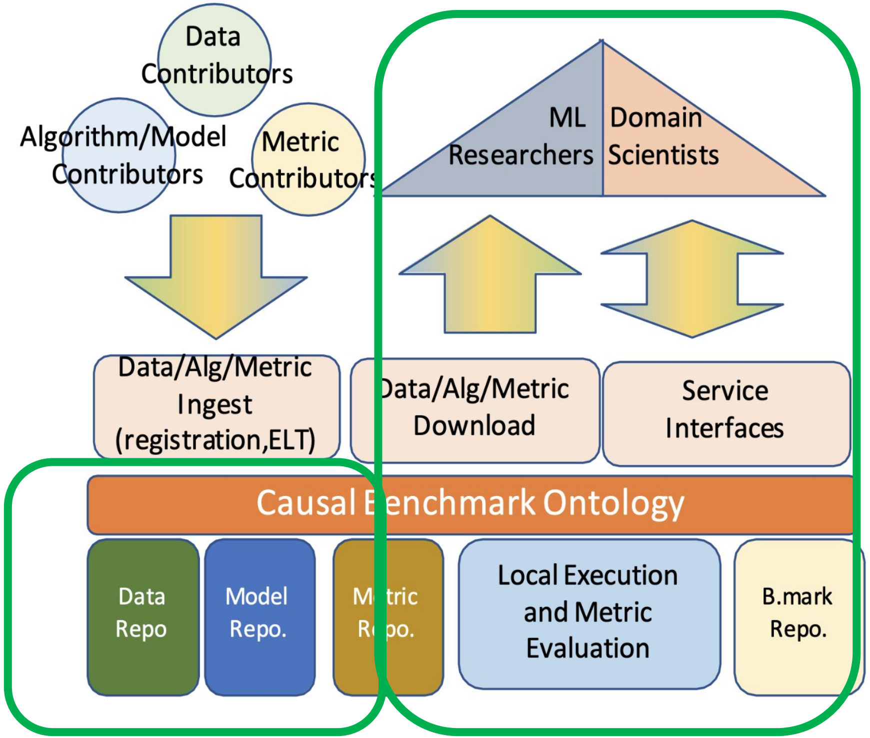


Benchmark executor

What is a Benchmark?

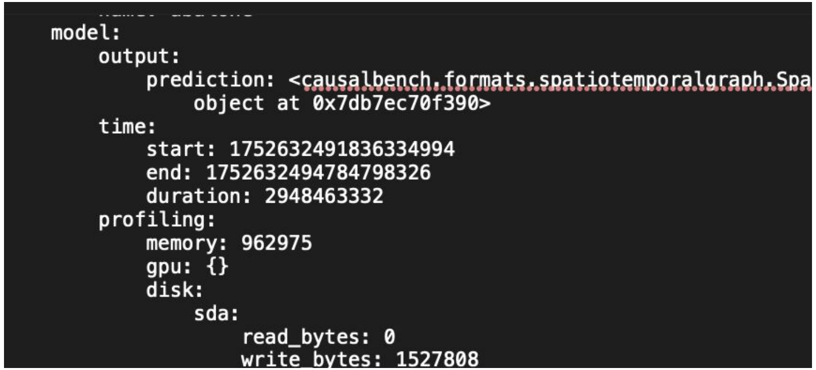


CausalBench: Use Scenario #3

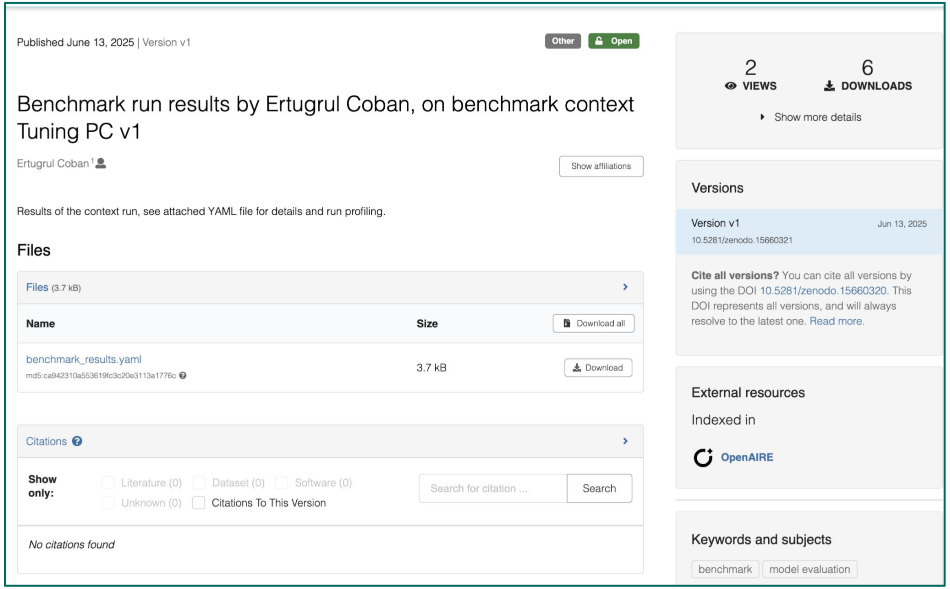


Sample benchmark output

- Includes
  - Model
  - Dataset
  - Hardware/software profiling
  - Accuracy metrics

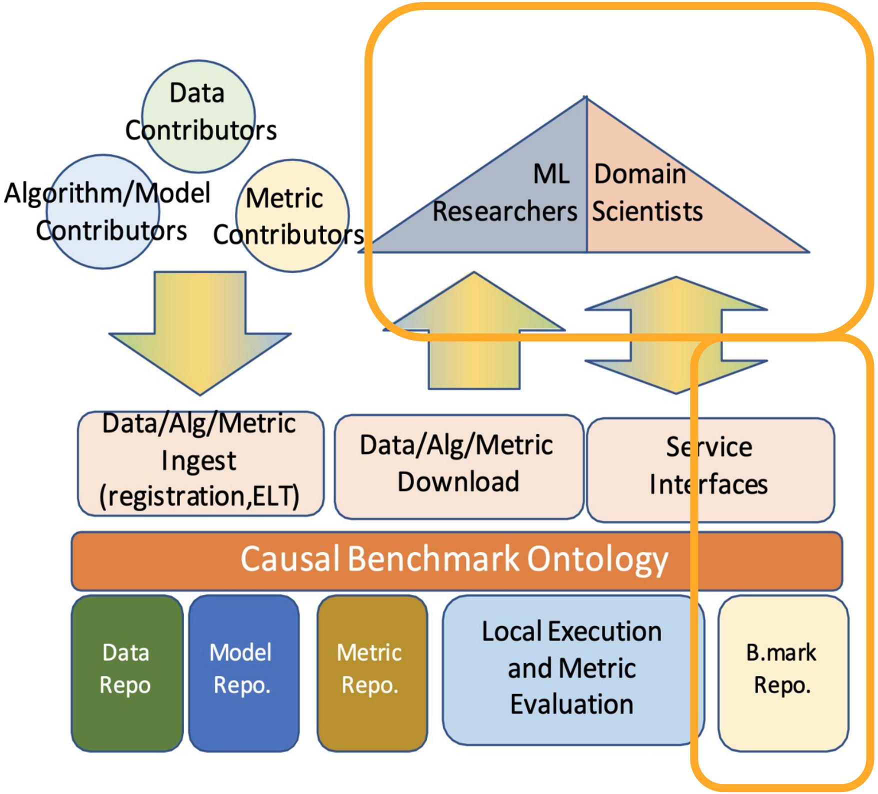


Uploaded and stored in CausalBench



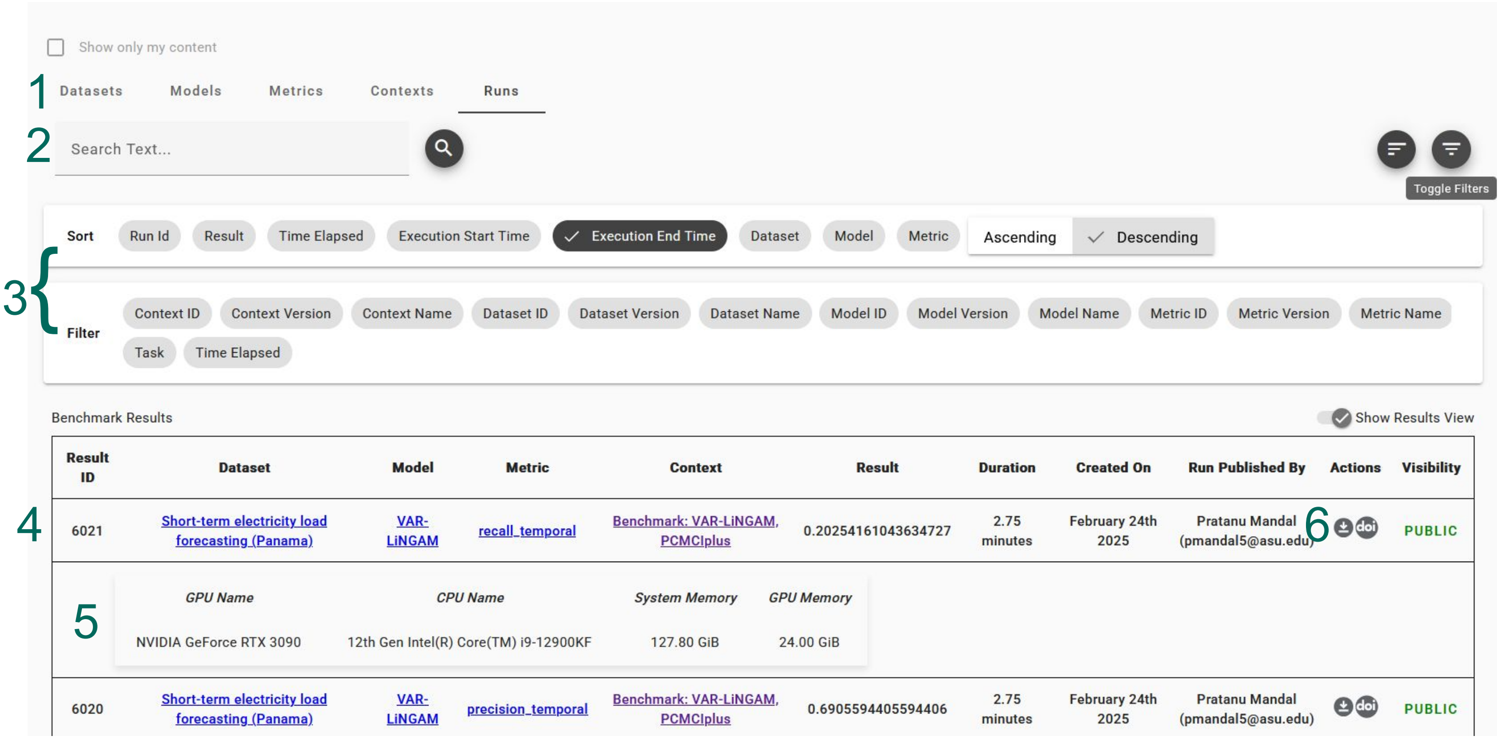
Permanently indexed (and citable) in Zenodo

CausalBench: Use Scenario #4



Benchmark explorer

CausalBench: Exploring benchmark results



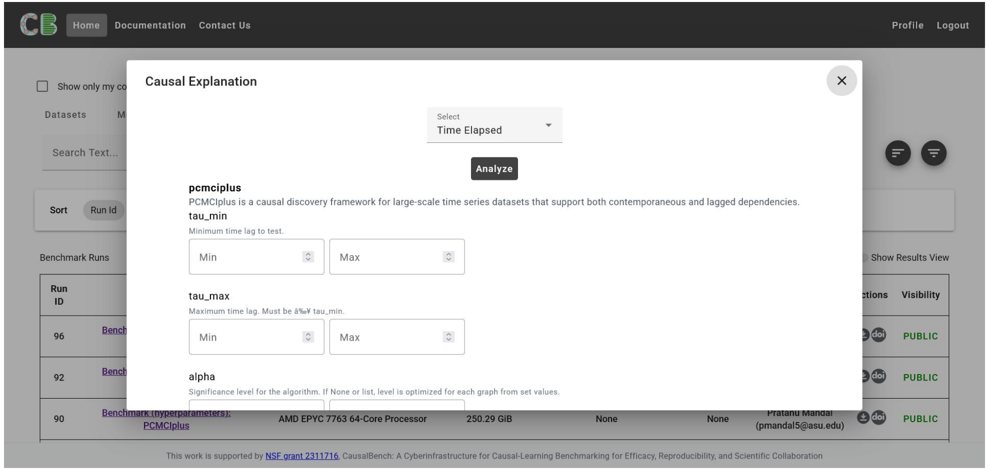
1. Repository selector
2. Search Function
3. Filter/Sort
4. Detail overview
5. On-demand details
6. Download/Cite

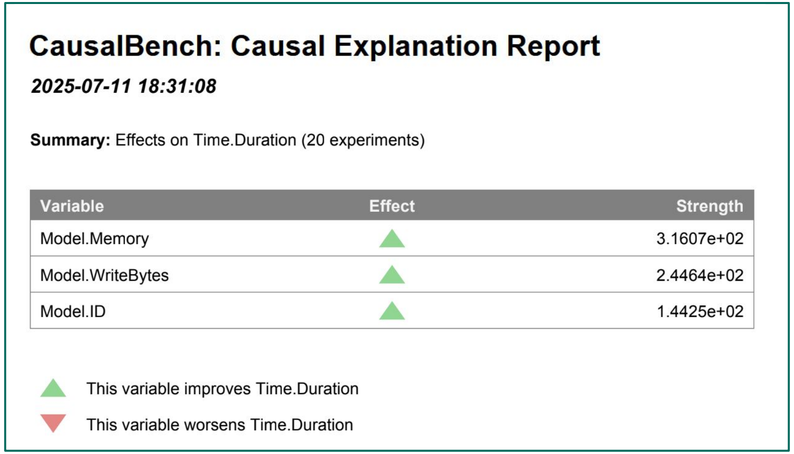
CausalBench: Explaining benchmark results

Result ID	Dataset	Model	Metric	Context	Result	Duration	Created On	Run Published By	Actions	Visibility
630	time_sim	VAR-LiNGAM	accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.9375	8.31 seconds	February 16th 2025	Abhinav Gorantla (agorant2@asu.edu)	doi	PUBLIC
640	Short-term electricity load forecasting (Panama)	VAR-LiNGAM	accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.568359375	4.67 minutes	February 16th 2025	Abhinav Gorantla (agorant2@asu.edu)	doi	PUBLIC

- **Sample question:**
  - Why does VAR-LiNGAM have better accuracy with time_sim but lower training time in this benchmark?
    - Did the hyperparameters play a role?
    - Could it be because of the dataset size?
    - Is there something else?
- These questions can be answered by ***generating explanations*** using CausalBench.

CausalBench: Explaining benchmark results





CausalBench: Causal Explanation Report

2025-07-11 18:31:08

Summary: Effects on Time.Duration (20 experiments)

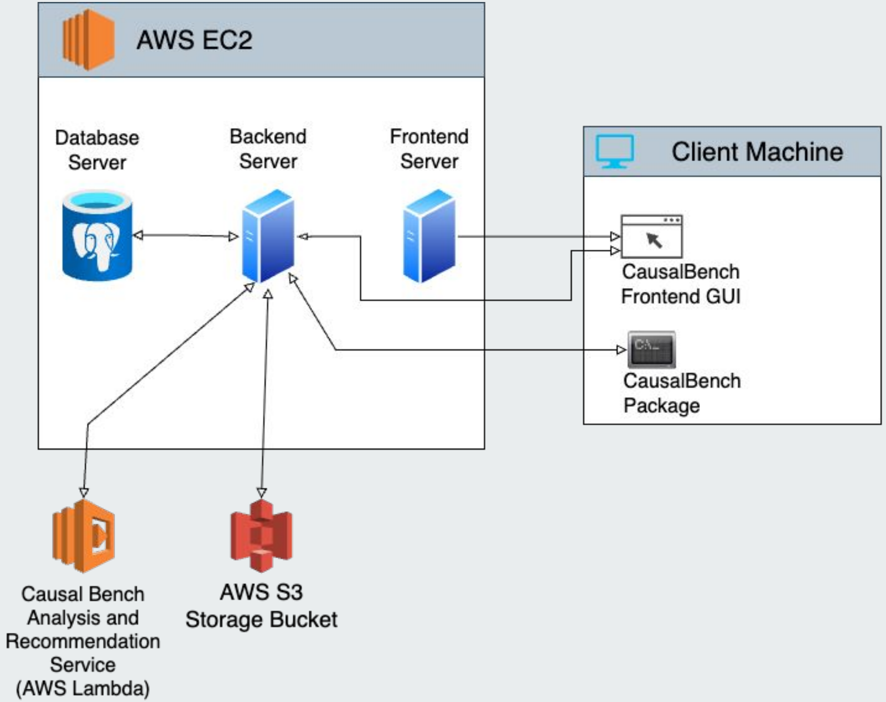
Variable	Effect	Strength
Model.Memory	▲	3.1607e+02
Model.WriteBytes	▲	2.4464e+02
Model.ID	▲	1.4425e+02

▲ This variable improves Time.Duration

▼ This variable worsens Time.Duration

- **Sample question:**
  - Why does VAR-LiNGAM have better accuracy with time_sim but lower training time in this benchmark?
    - Did the hyperparameters play a role?
    - Could it be because of the dataset size?
    - Is there something else?
- These questions can be answered by ***generating explanations*** using CausalBench.

Components of CausalBench



- CausalBench contains three components:
  - The python package handles the process of benchmarking.
  - The web backend receives the results from the python package and publishes it to zenodo.
  - The web frontend provides users with a GUI to browse benchmark runs, datasets, models, metrics and contexts already published to causalbench.org.

Agenda for today's Hands-on Tutorial

tutorial.causalbench.org

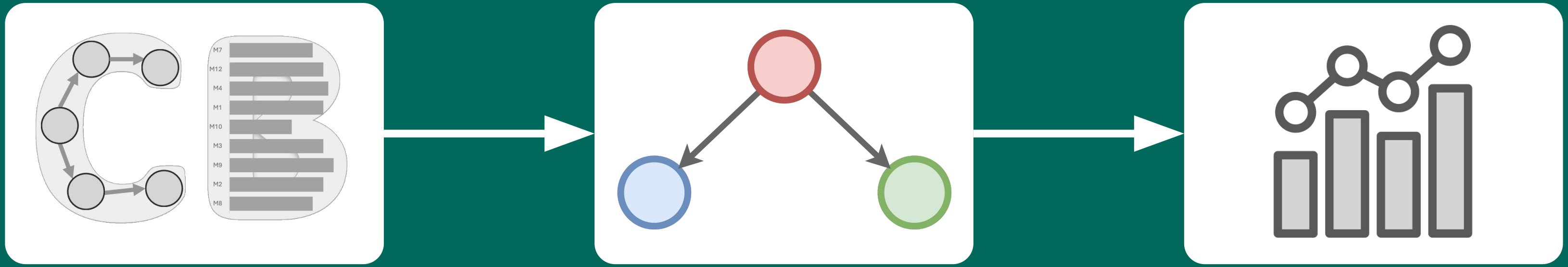
08:00-08:05	Introduction to the Tutorial
08:05-08:25	Introduction to **CausalBench**
08:25-08:55	Introduction to Causality and Causal Learning
08:55-09:30	Delve into the **CasualBench** framework to create and execute benchmarks
09:30-10:00	Coffee break
10:00-10:10	Shorter introduction to **CausalBench**
10:10-10:35	Explore published benchmarks and reproduce experiments
10:35-10:50	Gain further insights using Causal Explanation and Recommendations
10:50-11:00	**CausalBench**: What's Next?

End of Deck 1

Any Questions?

CausalBench: Causal Learning Research Streamlined

Understanding Causality



CausalBench **Causality** Benchmarking

tutorial.causalbench.org

August 3–7, 2025 KDD2025

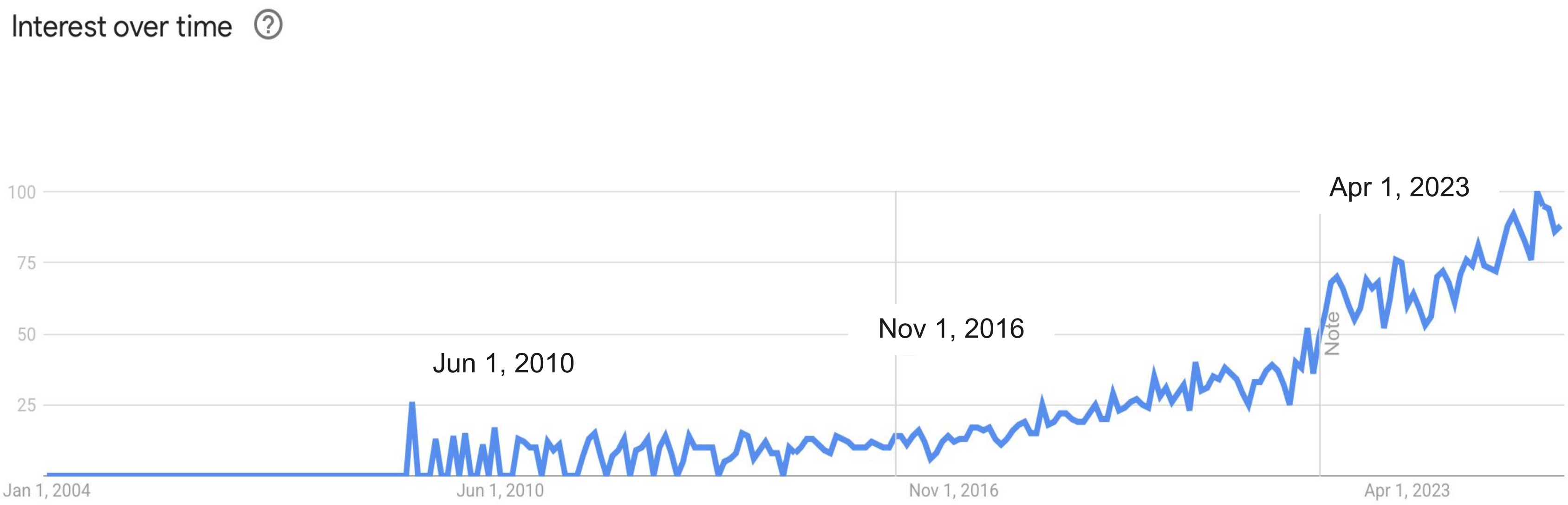
This research is funded by NSF Grant 2311716, "CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration", and NSF Grants #2230748, "PIRE: Building Decarbonization via AI-empowered District Heat Pump Systems", #2412115, "PIPP Phase II: Analysis and Prediction of Pandemic Expansion (APPEX)" and USACE #GR40695, "Designing nature to enhance resilience of built infrastructure in western US landscapes".

Causal Learning: Why it matters?

1. What is Causal Learning?
2. Why does Causal Learning matter?
3. Two Tasks in Causal Learning.

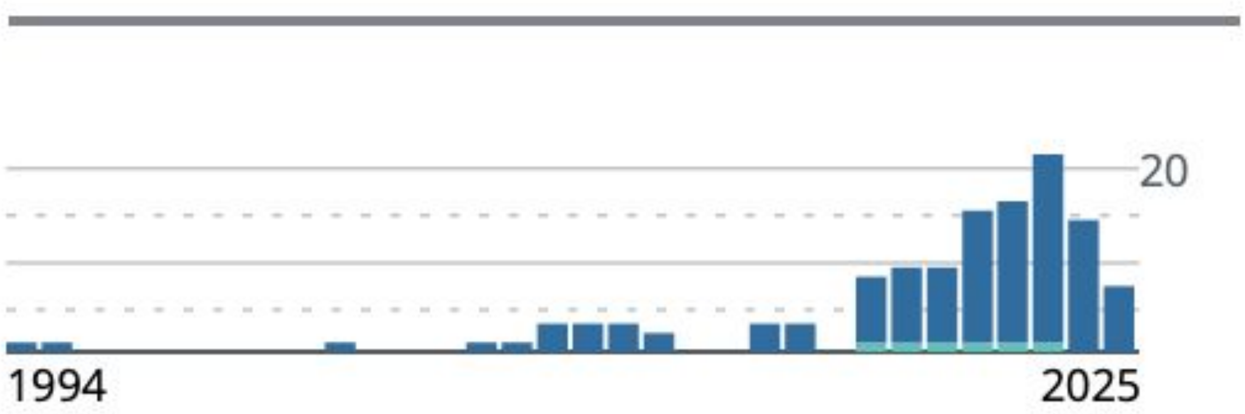
Interests in “Causal Learning”

Google Trends for the term “Causal Learning”



Interests in “Causal Learning”

of papers published at KDD with the term “causal” in the title



What is Causal Learning?

Causal Learning answers the question of “**Why**” and describe the relationship between

- a **cause** (an action, event, or condition), and
- its **effect** (an outcome that results from it).

How do **socioeconomic status, nutrition, study time,** and, **sleep**, influence student **GPA**?

What’s the effect of the **vaccine** on a patient’s **health?**

Two Tasks in Causal Learning

How do **socioeconomic status, nutrition, study time,** and, **sleep**, influence student **GPA**?

Causal Discovery

We don't know what causes what. We want to uncover the structure — who influences whom.

What's the effect of the **vaccine** on a patient's **health?**

Two Tasks in Causal Learning

How do **socioeconomic status, nutrition, study time,** and, **sleep**, influence student **GPA**?

Causal Discovery

We don’t know what causes what. We want to uncover the structure — who influences whom.

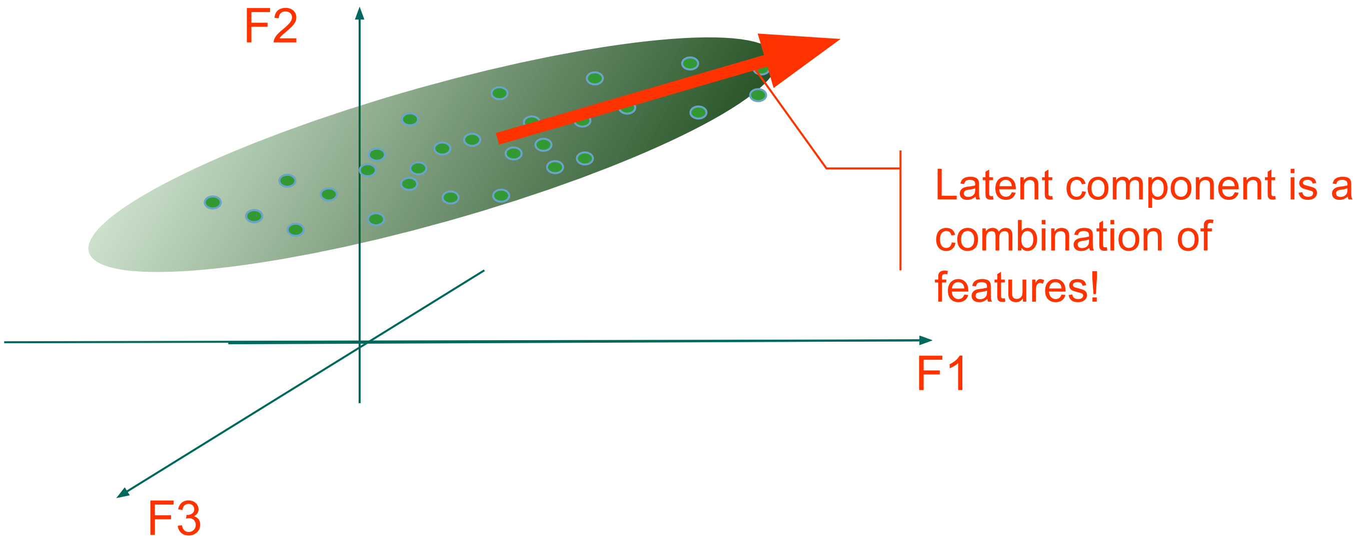
What’s the effect of the **vaccine** on a patient’s **health?**

Causal Effect Estimation

Knowing cause and effect, want to estimate how much effect one variable has on another.

So…why does causal learning matter?

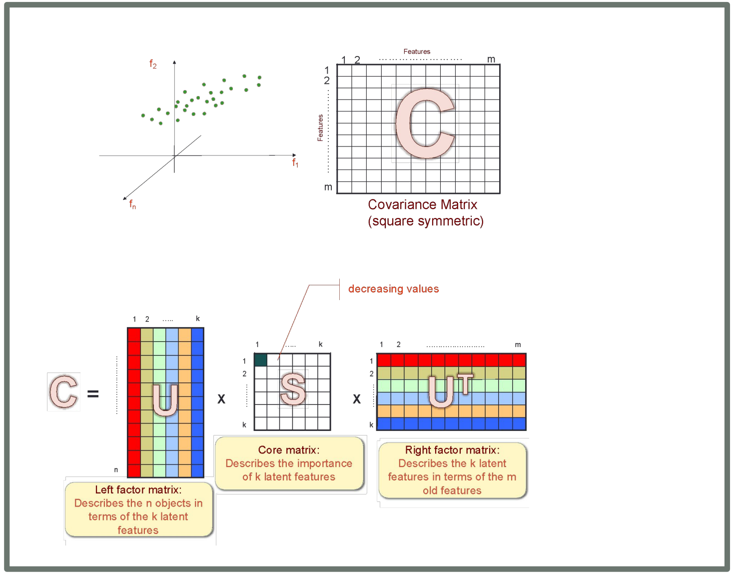
- Traditional data analysis and retrieval is based on statistical/probabilistic cues underlying the data
  - e.g. dimensionality reduction often relies on identifying and eliminating redundancies in terms of correlation or covariance



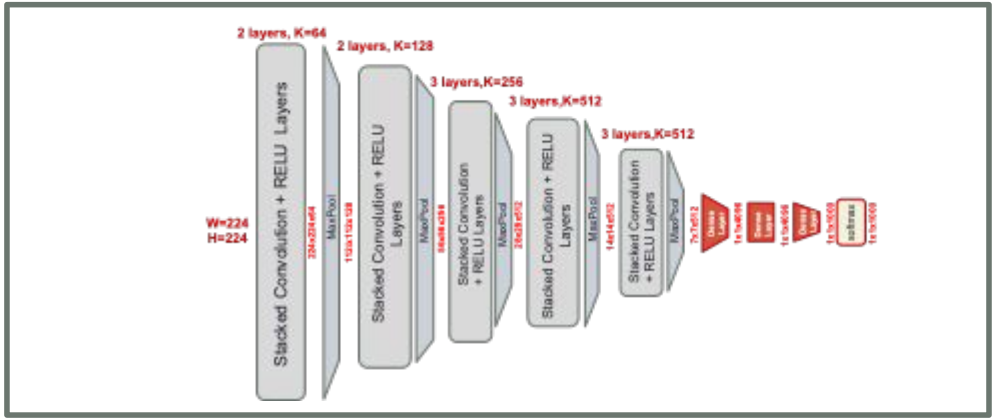
So...why does causal learning matter?

- Examples range from simple matrix decomposition (e.g., PCA) to more complex DNNs

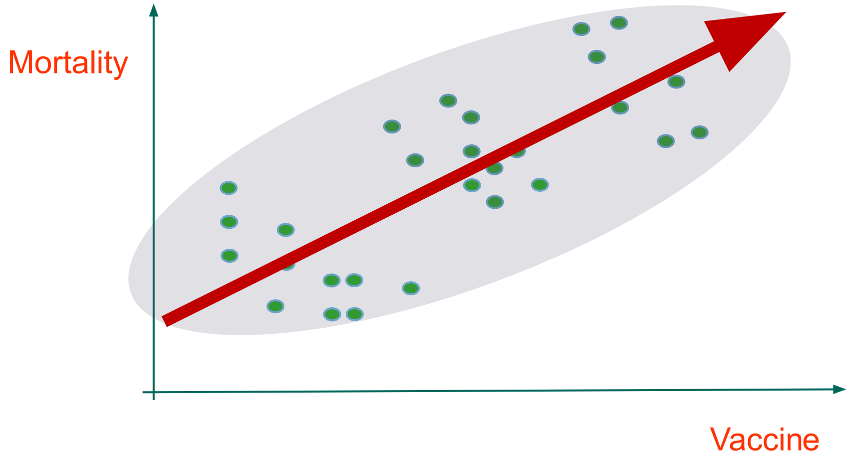
Principal Component Analysis



DNN (VGG16)



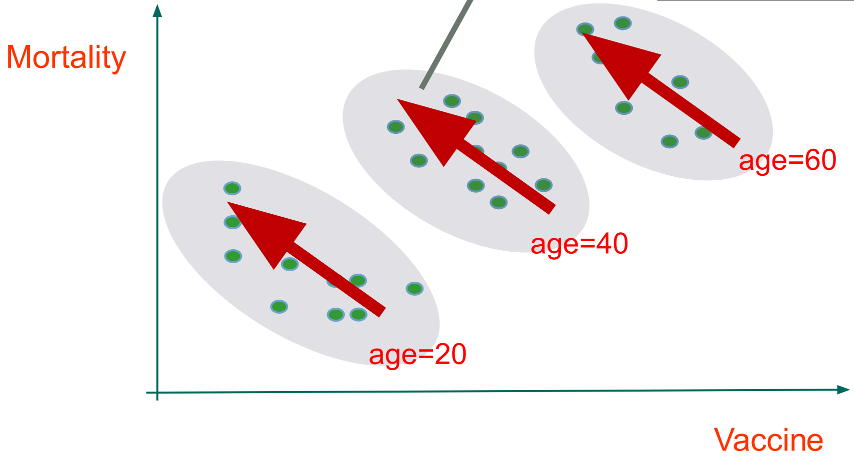
Problem: this approach does not always make sense!

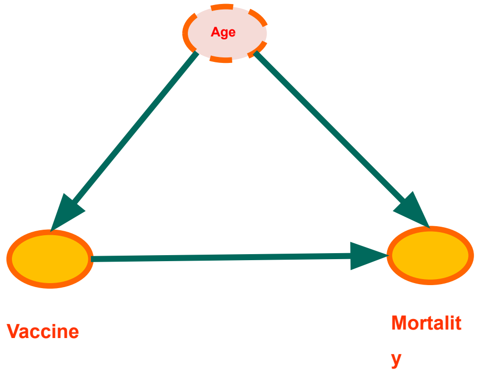


e.g. Simpson's paradox

Problem: this approach does not always make sense!

- “Age” is an (unobserved) confounding variable





- Data analysis without accounting for confounding variables will result in wrong conclusions…

Key questions..

- Q1: Can we obtain causal knowledge (discover the causal graph) from observations and answer causal queries?
    - Can we analyze observations to discover underlying causally-meaningful patterns and relationships between input parameters, key events/interventions, and outcomes?

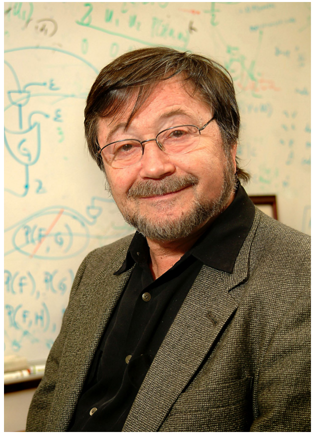
- Q2: Can we compute the probability distribution of Y after we intervene on X – denoted as $P(Y \mid do(X = x))$?

- Q3: If we are given a-priori causal knowledge, can we leverage this in our data analysis or in explaining our results?
    - Can we support causally-informed explanations and root-cause analysis?
    - Can we support what-if analysis and optimize for different outcomes?
    - Can we transfer knowledge and models across causally-similar systems?
    - Can we make causally-robust predictions and recommendations?

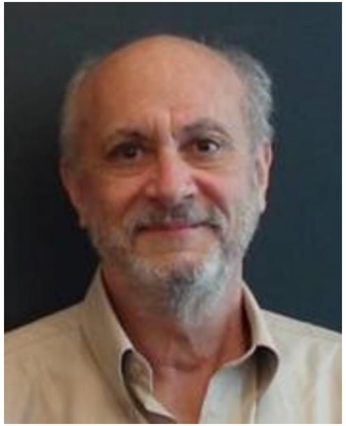
Causal Model Frameworks

A **Causal Model Framework** helps us

- represent how variables influence each other
- make predictions under interventions, not just observations
- go beyond correlation to answer “why” and “what if” questions



Judea Pearl



Donald Rubin

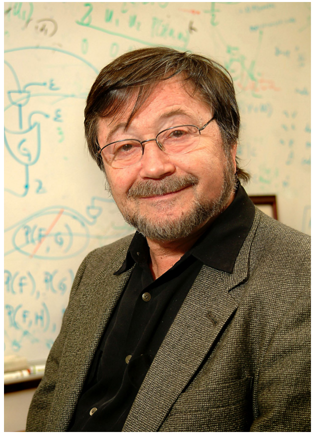
Causal Model Frameworks

A **Causal Model Framework** helps us

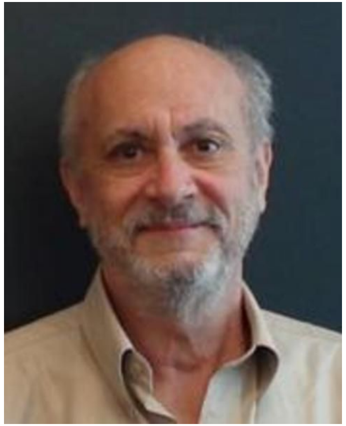
- represent how variables influence each other
- make predictions under interventions, not just observations
- go beyond correlation to answer "why" and "what if" questions

There is no single causal model — different frameworks suit different goals:

- Pearl's Causal Model: Popular in computer science.
- Rubin's Causal Model: Popular in statistics, econometrics.
- and more…



Judea Pearl



Donald Rubin

Causal Algorithms

Causal Discovery

- Constraint-based: **PC**[1], FCI[2]
- Score-based: **GES**[3], FGES[4]
- Functional: **LiNGAM**[5], ANM[6]
- Optimization-based: NOTEARS[7], DAG-GNN[8]
- Temporal: **PCMCI+**[9], **VAR-LiNGAM**[10]

Highlighted algorithms are supported by CausalBench out-of-box.

Causal Effect Estimation

- Regression-based: **Linear regression**[11], GLMs[12]
- Matching: Propensity score[13], Mahalanobis[14]
- IPW (Inverse Probability Weighting)[15]
- Meta-learners[16]: S-Learner, T-Learner, X-Learner
- Causal Forests[17]
- DML (Double Machine Learning)[18]

Highlighted algorithm is used for causal explanation in CausalBench.

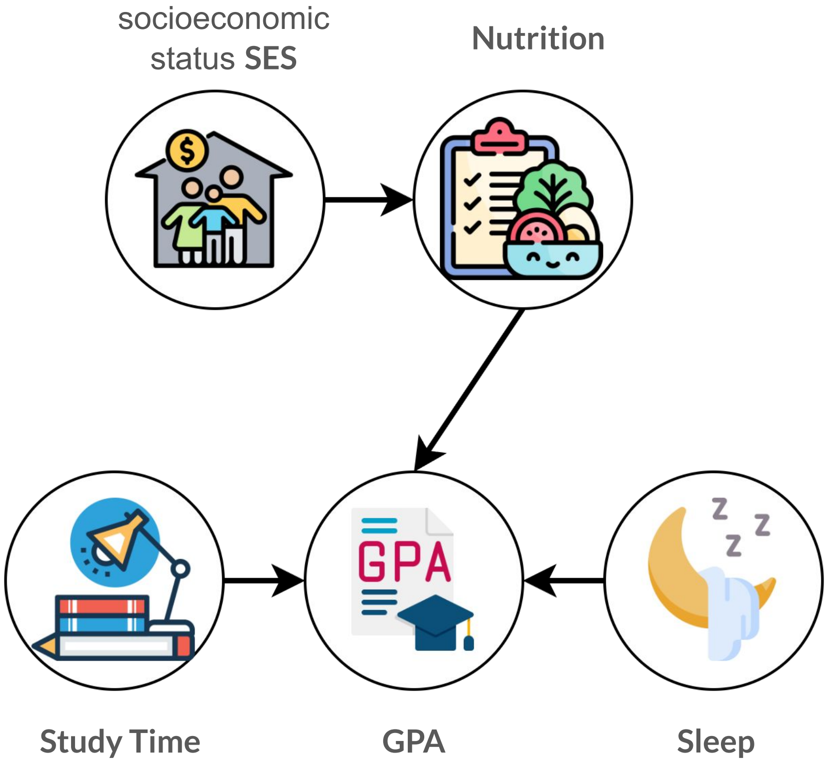
Basics of Causal Graphs

1. What's a Causal Graph?
2. Causal Graph and Data Dependencies
3. D-Separation

Causal graph: nodes and edges

We use a Causal Graph G = (V, E) to describe the causal relationships between variables.

A common assumption is that causal graphs are acyclic.

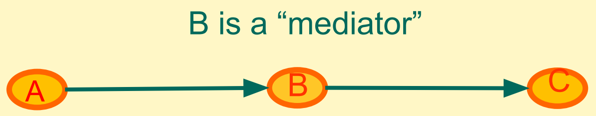


Key concepts - Mediator/Chain

Real World Example

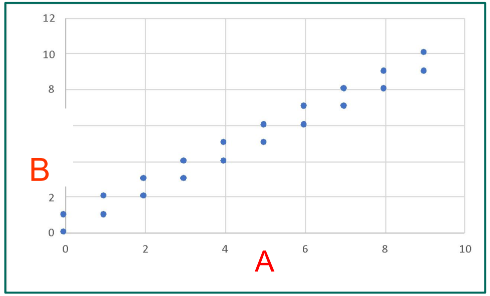
A: Sleep time
B: Wake-up time
C: Arrival time at work

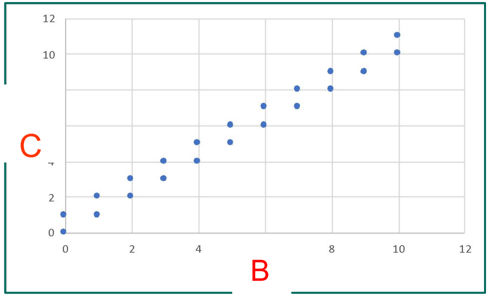
Graph Structure

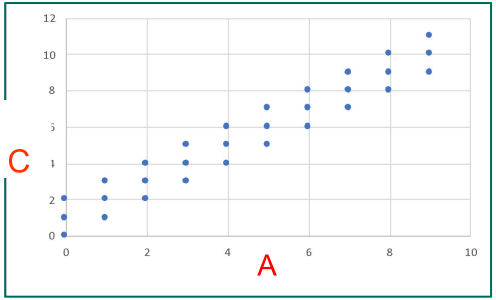


Data (in)dependencies

- A and B are dependent
- B and C are dependent
- A and C are dependent





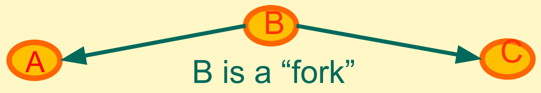


Key concepts - Fork/Common Cause

Real World Example

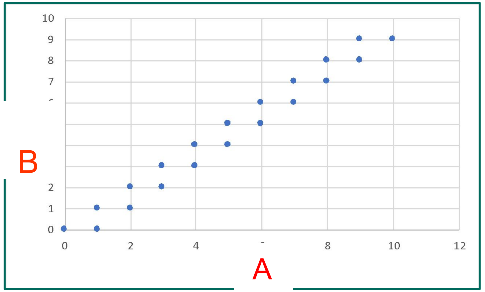
A: Vaccine
B: Age
C: Mortality

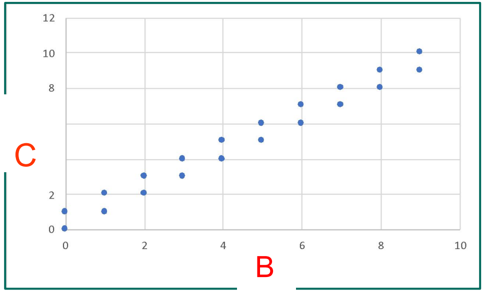
Graph Structure

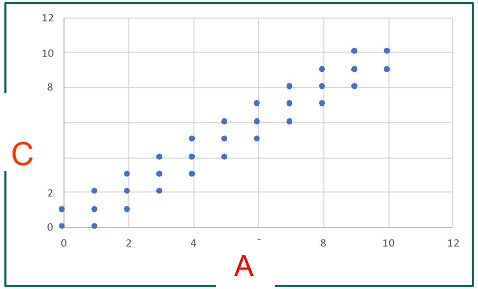


Data (in)dependencies

- A and B are dependent
- B and C are dependent
- A and C are dependent





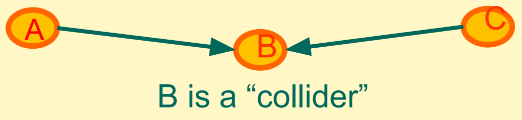


Key concepts - Collider/V-structure/Common Effect

Real World Example

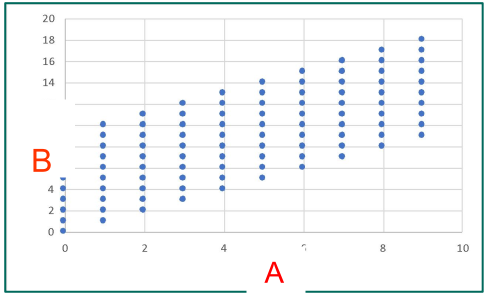
A: Good Looking
B: Award
C: Acting Ability

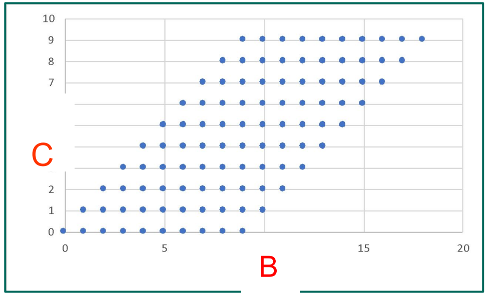
Graph Structure

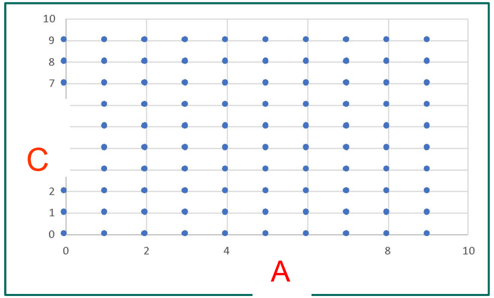


Data (in)dependencies

- A and B are dependent
- B and C are dependent
- A and C are **independent**







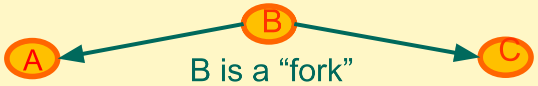
Summary - (in)dependencies

Graph Structure

Chain

B is a “mediator”

Fork



Collider

B is a “collider”

Data (in)dependencies

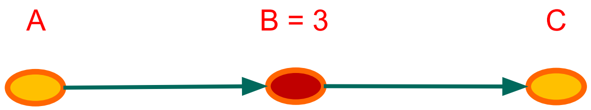
- A and B are dependent
- B and C are dependent
- A and C are dependent

- A and B are dependent
- B and C are dependent
- A and C are dependent

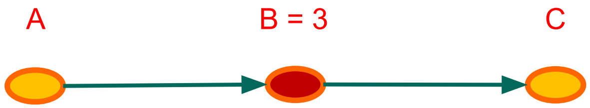
- A and B are dependent
- B and C are dependent
- A and C are **independent**

Conditioning and Conditional Independence

- Conditioning: set a variable to a fixed value. P(A, C | B = 3)

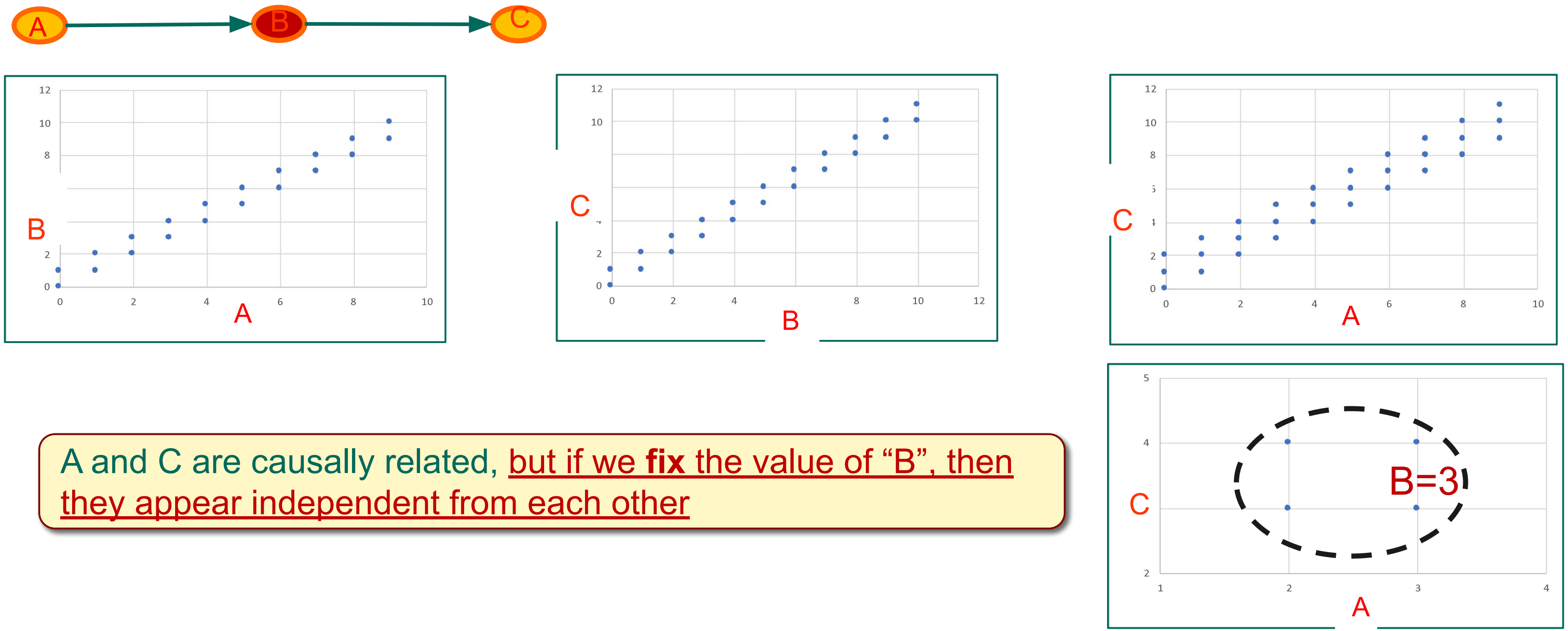


- Conditional Independence: Two variables C and A are conditionally independent given B
  P(A, C | B = 3) = P(A | B = 3)P(C | B = 3)?



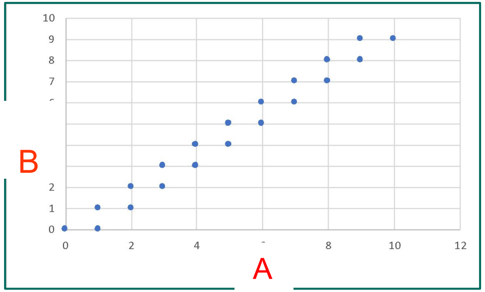
Key concepts - Mediator/Chain

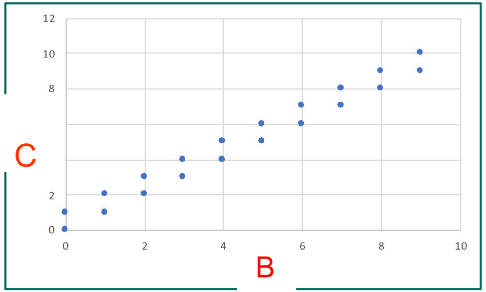
B is a “mediator”

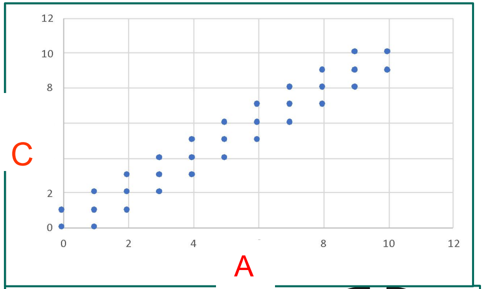


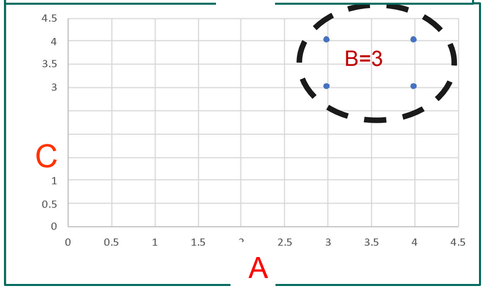
A and C are causally related, <u>but if we **fix** the value of “B”, then they appear independent from each other</u>

Key concepts - Fork



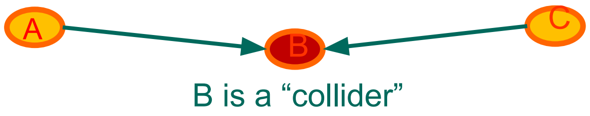


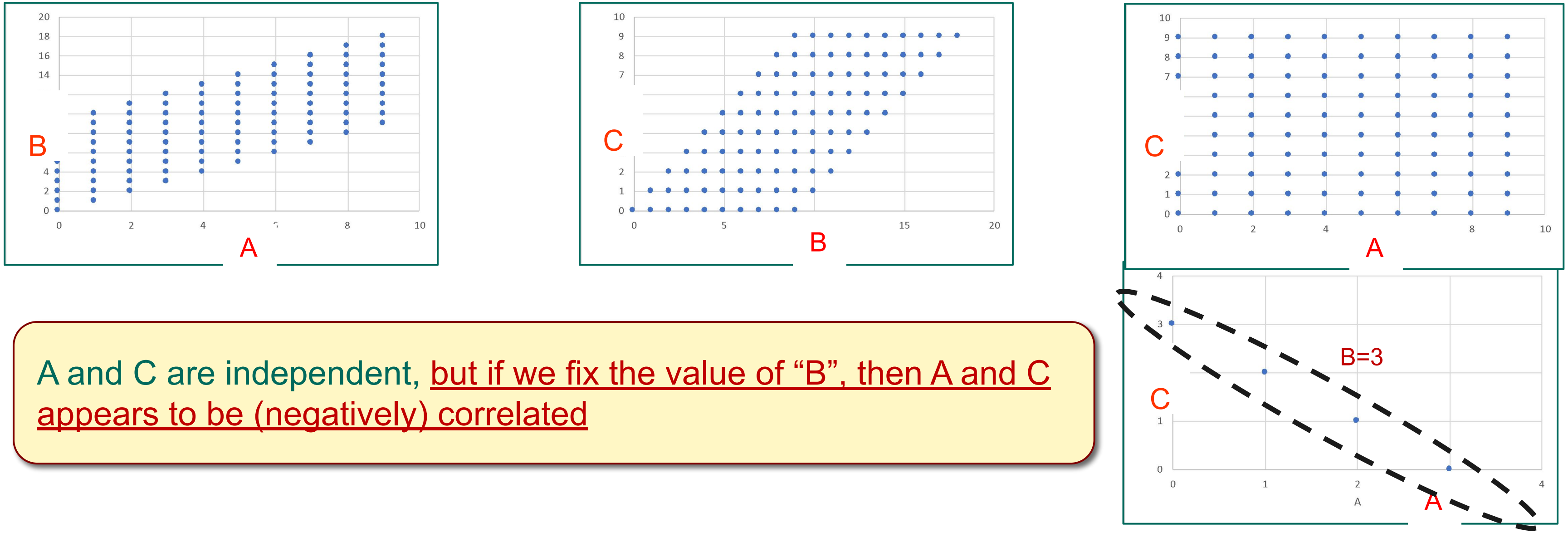




A and C are correlated, <u>but if we fix the value of “B”, then the correlation disappears</u>

Key concepts - Collider





A and C are independent, <u>but if we fix the value of "B", then A and C appears to be (negatively) correlated</u>

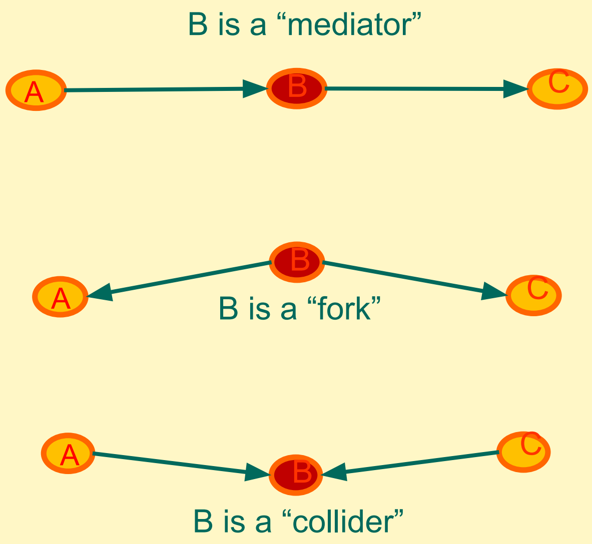
Summary - (in)dependencies

Graph Structure

Chain

Fork

Collider



Data (in)dependencies
When Conditioned on ***B***

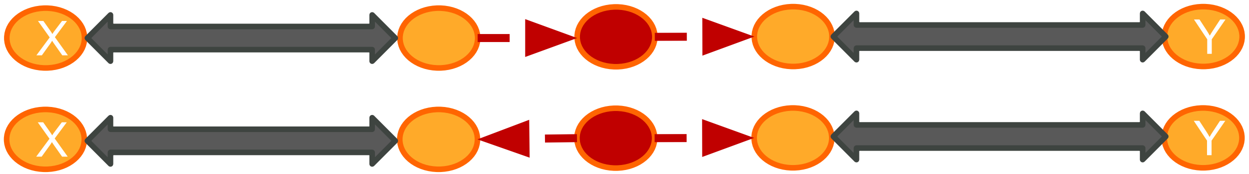
- A and B are dependent
- B and C are dependent
- A and C are **independent**

- A and B are dependent
- B and C are dependent
- A and C are **independent**

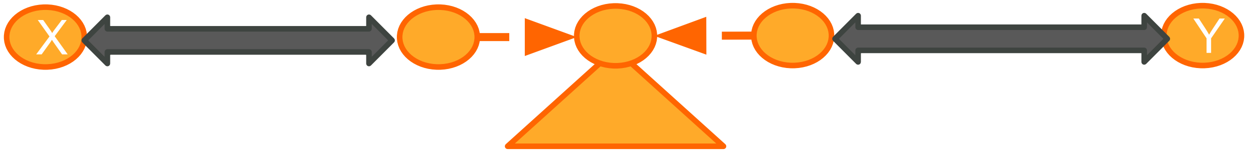
- A and B are dependent
- B and C are dependent
- A and C are dependent

Key concepts - Causal blocking

- A path in the causal graph is blocked[a] if
  - the path contains a **chain** or a **fork** that has been conditioned,



  - the path contains a **collider** such that the collision node and its descendants have not been conditioned

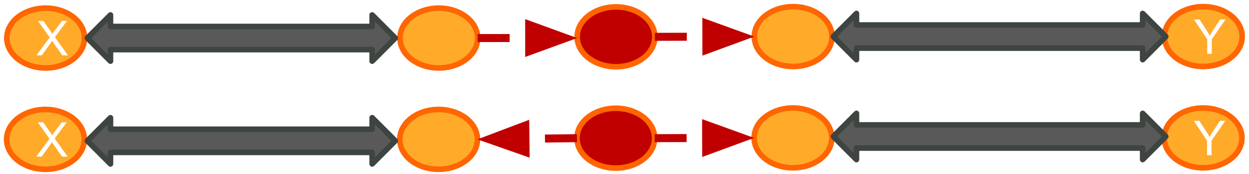


[a] Pearl, J. 2009a. Causal inference in statistics: An overview. Statist. Surv., 3: 96–146.

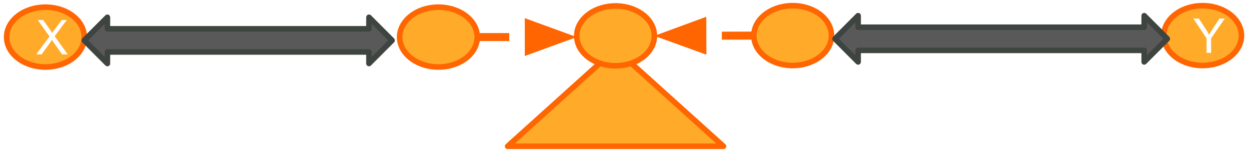
Key concepts - Causal blocking

- A path in the causal graph is blocked[a] if

Conditioning erases evidence of the underlying causal relationships



  - the path contains a **collider** such that the collision node and its descendants have not been conditioned

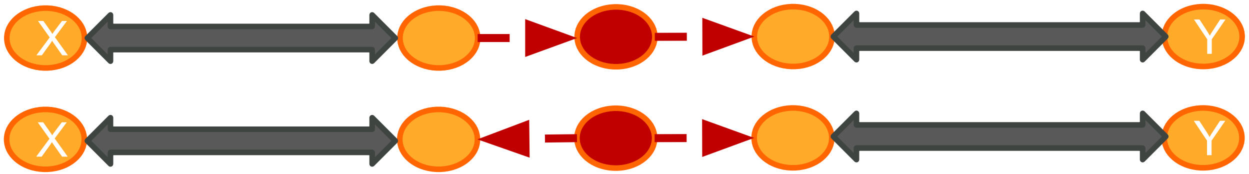


[a] Pearl, J. 2009a. Causal inference in statistics: An overview. Statist. Surv., 3: 96–146.

Key concepts - Causal blocking

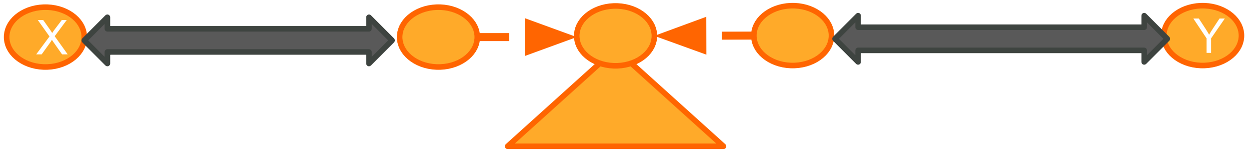
- A path in the causal graph is <u>blocked</u>[a] if

Conditioning erases evidence of the underlying causal relationships



Conditioning unblocks the path and introduces spurious correlations

  - the path contains a **collider** such that the collision node and its descendants <u>have not been conditioned</u>

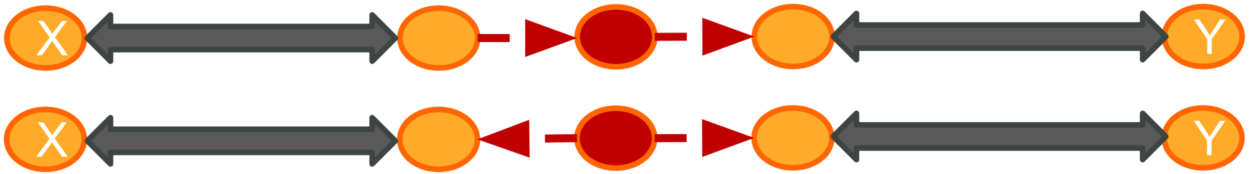


[a] Pearl, J. 2009a. Causal inference in statistics: An overview. Statist. Surv., 3: 96–146.

Key concepts - Causal blocking

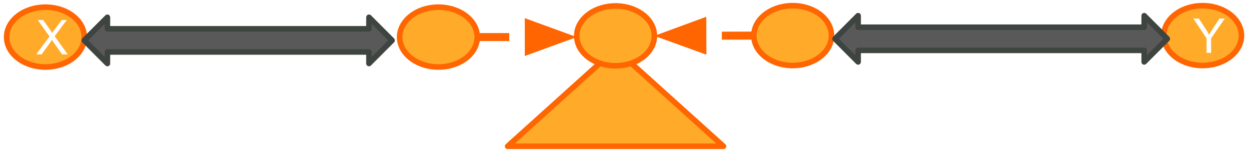
- A path in the causal graph is blocked[a] if

Conditioning erases evidence of the underlying causal relationships



Conditioning unblocks the path and introduces spurious correlations

  - the path contains a **collider** such that the collision node and its descendants have not been conditioned



Let **X, Y, Z** be three sets of nodes in a causal graph **G**.
X and **Y** are ***d-separated*** given **Z**, if all path from **X** to **Y** through **Z** are ***blocked***.

Task: Causal Discovery

1. What is Causal Discovery?
2. Common Assumptions
3. Markov Equivalence Class
4. Model: PC Algorithm
5. Metrics

Where do causal graphs come from ?

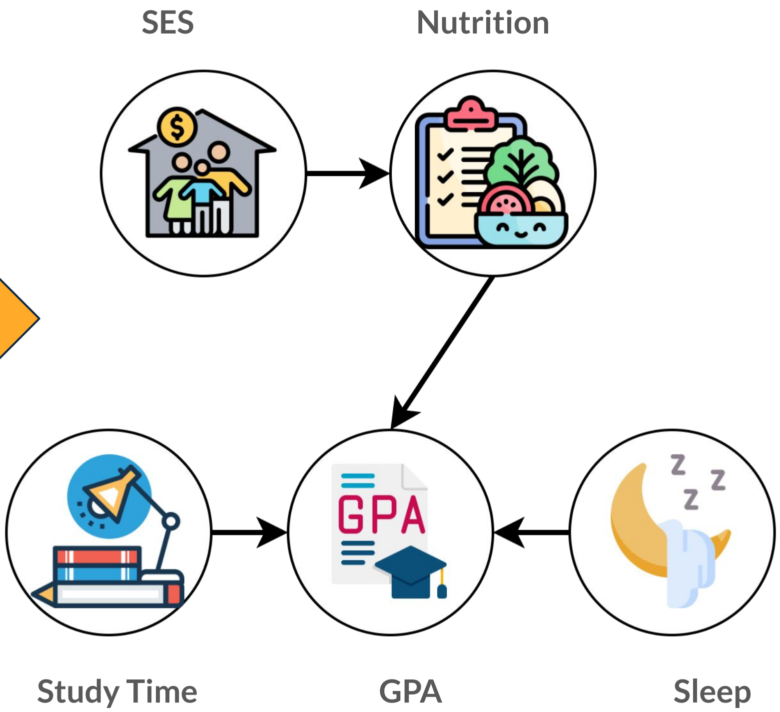
So, causal graph is a useful tool– but, where does it come from?

- Option #1: Expert provided - Rare, Scarce
- Option #2: Learned from observations - Causal Discovery!!!

Causal Discovery

SES	Nutrition	Study	Sleep	GPA
...	...	...	...	...

Observations



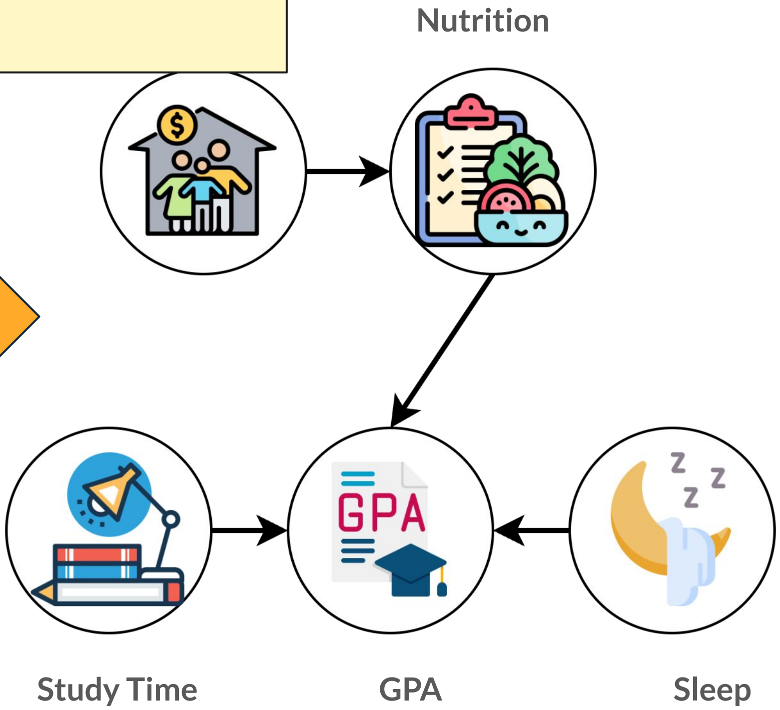
49

Causal Discovery

If observations are temporal, rely on whether one variable can be used to predict the other.

SES	Nutrition	Study	Sleep	
…	…	…	…	…

Observations



50

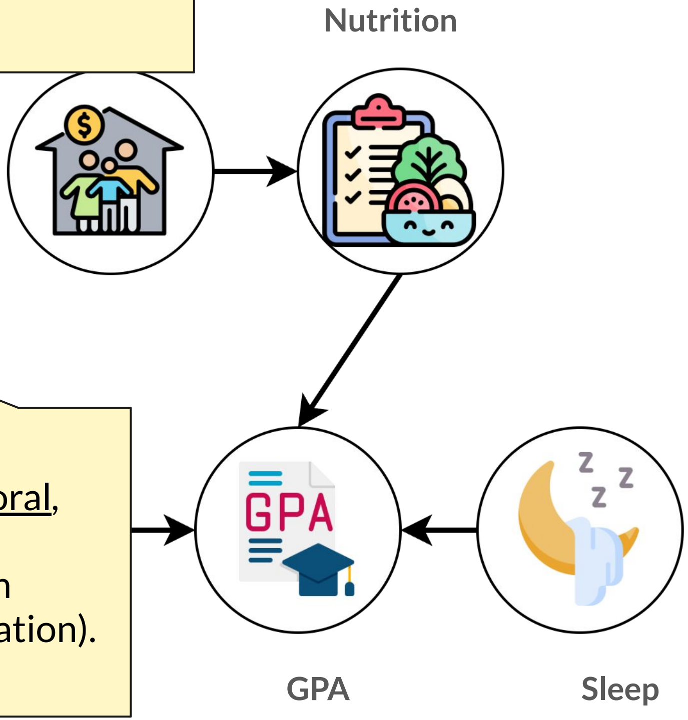
Causal Discovery

SES	Nutrition	Study	Sleep	
…	…	…	…	…

Observations

If observations are temporal, rely on whether one variable can be used to predict the other.

If observations are not temporal, consider what statistical independences result from conditioning variables (d-seperation).



51

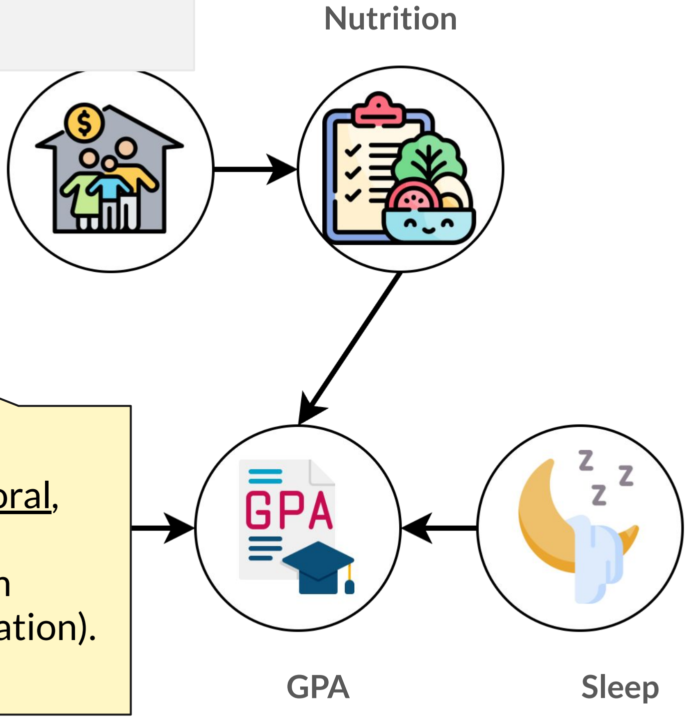
Causal Discovery

If observations are temporal, rely on whether one variable can be used to predict the other.

SES	Nutrition	Study	Sleep	G
…	…	…	…	…

Observations

If observations are not temporal, consider what statistical independences result from conditioning variables (d-seperation).

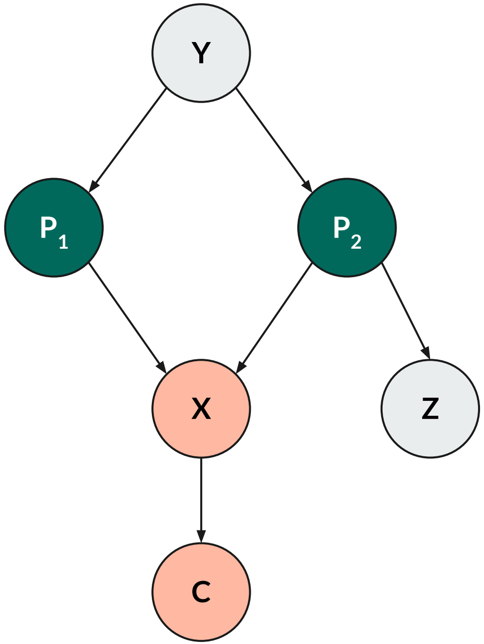


52

Common Assumptions

- Markov Condition
  - A variable is independent of non-descendents, given its parents.
  - All conditional independencies in the graph are reflected in the dataset.

$$X \perp_G Y|Z \Rightarrow X \perp_D Y|Z$$



$P(X \mid P_1, P_2, Y, Z) = P(X \mid P_1, P_2)$

Common Assumptions

- Markov Condition
  - A variable is independent of non-descendents, given its parents.
  - All conditional independencies in the graph are reflected in the dataset.

$$X \perp_G Y|Z \Rightarrow X \perp_D Y|Z$$

- Faithfulness
  - All conditional independencies in the data (D) are reflected in the graph structure.
  - Conditional independence = d-separation.

$$X \perp_G Y|Z \Leftarrow X \perp_D Y|Z$$

Common Assumptions

- Markov Condition
  - A variable is independent of non-descendents, given its parents.
  - All conditional independencies in graph reflected in dataset.

$$X \perp_G Y|Z \Rightarrow X \perp_D Y|Z$$

- Faithfulness
  - All conditional independencies in the data (D) are reflected in the graph structure.
  - Conditional independence = d-separation.

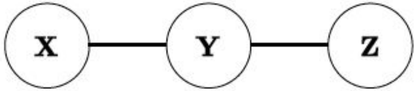
$$X \perp_G Y|Z \Leftarrow X \perp_D Y|Z$$

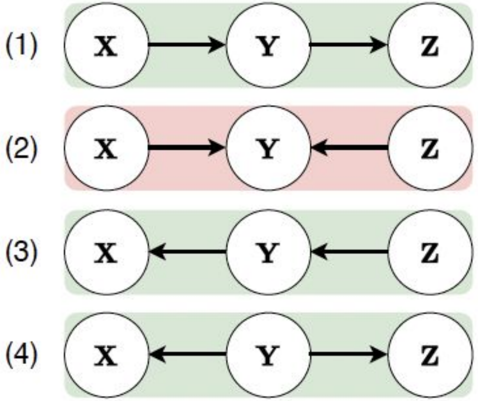
- Sufficiency
  - All common causes are included among the observations.
  - No missing variables, no latent variables.

Warning! - conditional independence tests may not always be sufficient to distinguish causal graphs

- In the example on the right,
  - Causal graphs #1, #3, and #4 have the same conditional independence structure
    - X dependent on Z
    - X independent from Z, only given Y
  - #2 has a different conditional independence structure
    - X independent from Z
    - X dependent on Z, given Y

- #1, #3, and #4 are said to be in the same Markov Equivalence Class

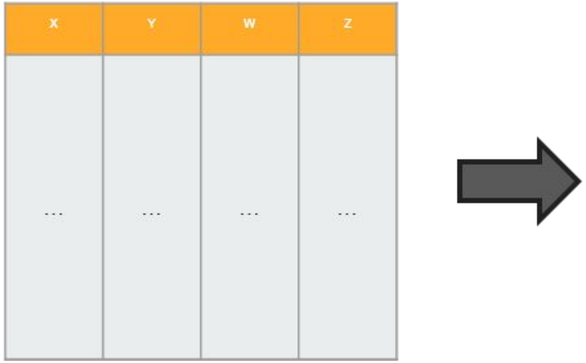




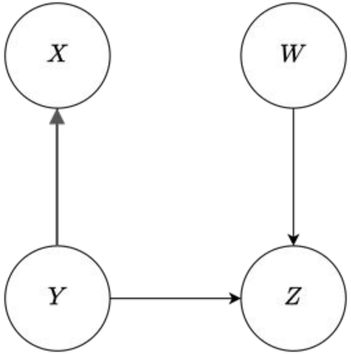
Models: Peter-Clark (PC) Algorithm

Key idea:

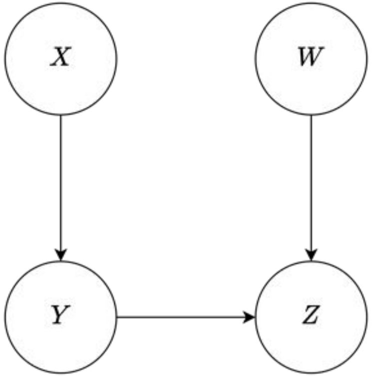
- Uses **conditional independence tests** to infer graph structure.
    - If variables X and Y are conditionally independent given <u>any</u> set of variables (excluding X and Y), there cannot be a direct causal edge between X and Y.



Observations



Discovered graph

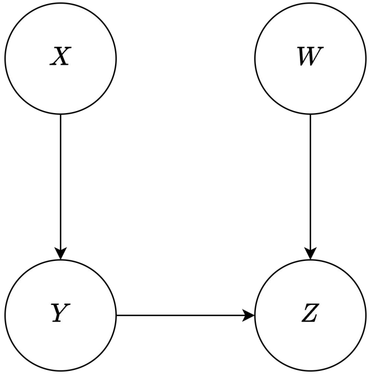


Ground truth graph

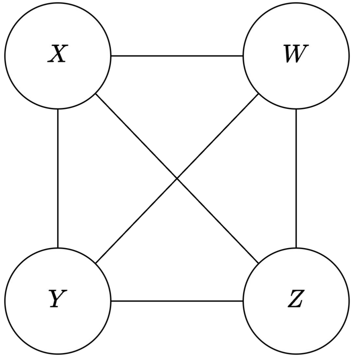
Models: Peter-Clark (PC) Algorithm

Steps:

1. Start with a fully connected undirected graph.



Ground truth

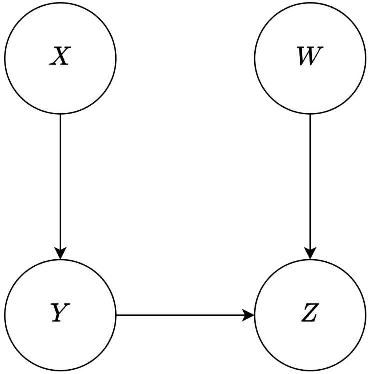


Undirected graph

Models: Peter-Clark (PC) Algorithm

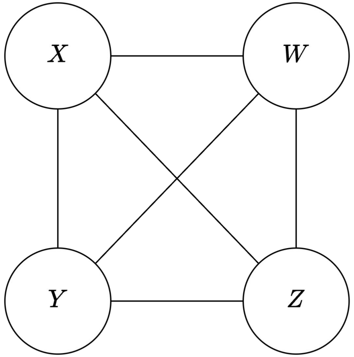
Steps:

1. Start with a fully connected undirected graph.
2. Consider conditioning set sizes m=1,2, ...
   a. For each edge X->Y
   b. Check if there is a set S of size m that renders X and Y statistically independent
   c. If such a set, S, is found, then remove the edge from the graph



Ground truth

$W \perp X$
$X \perp Z | Y$
$Y \perp W$

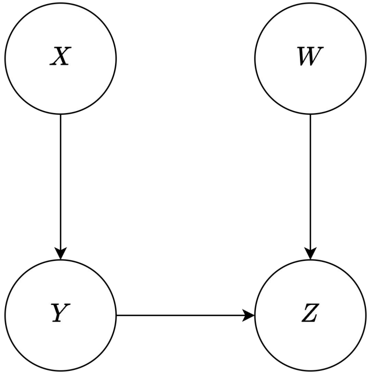


Undirected graph

Models: Peter-Clark (PC) Algorithm

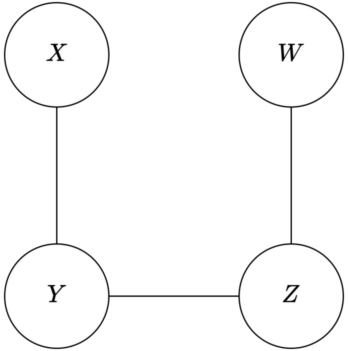
Steps:

1. Start with a fully connected undirected graph.
2. Consider conditioning set sizes m=1,2, ...
   a. For each edge X->Y
   b. Check if there is a set S of size m that renders X and Y statistically independent
   c. If such a set, S, is found, then remove the edge from the graph



Ground truth

$W \perp X$
$X \perp Z | Y$
$Y \perp W$



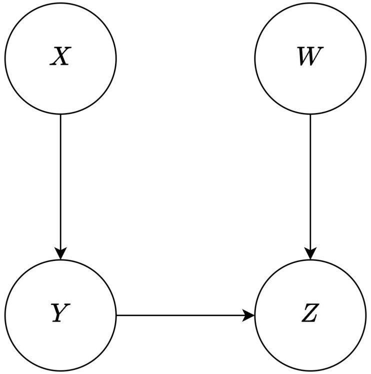
Undirected graph
Remove edge per step 2

Models: Peter-Clark (PC) Algorithm

Steps:

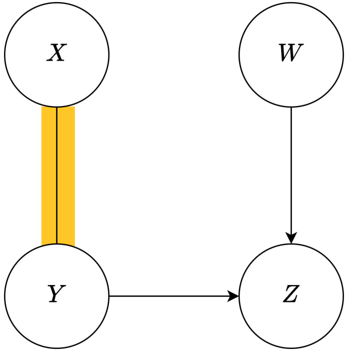
1. Start with a fully connected undirected graph.
2. Consider conditioning set sizes m=1,2, ...
   a. For each edge X->Y
   b. Check if there is a set S of size m that renders X and Y statistically independent
   c. If such a set, S, is found, then remove the edge from the graph
3. Orient remaining edges based on collider rules
   a. For each pair of non-neighbors, X and Y, with a common neighbor, Z
      i. If Z is not in the separator set for X and Y, then we must have X -> Z <- Y

Can only discover up to a markov equivalent class (MEC)



Ground truth

$W \perp X$
$X \perp Z|Y$
$Y \perp W$



Directed graph

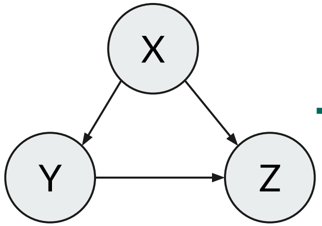
Edge orientation per step 3

Metrics - Graph Structure

- A Ground Truth Graph is required.

- Uses **adjacency matrix** to represent Causal Graphs.
  - Ground Truth Adjacency Matrix **A**,
  - Predicted adjacency matrix **A***.

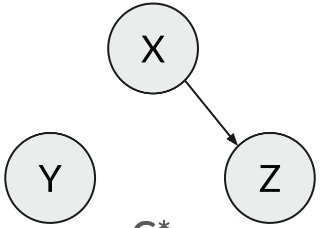
- Compare edges in the two causal graphs:
  - Precision, Recall, F1, and more...
  - SHD (Structural Hamming Distance) = # insertions + # deletions + # flips.
    - In the example, SHD(A, A*) = 2



G

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

A



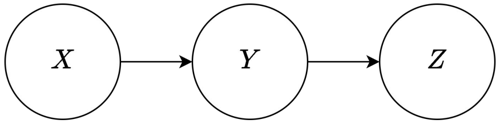
G*

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

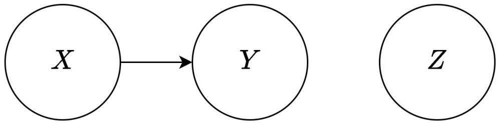
A*

Metrics - Intervention behavior

- Intervention
  - like randomized experiments; not conditioning.
  - remove incoming edge, denoted by **do(X)**.
- **intervention distribution**
  - $P(Y_j \mid do(X_i))$
- Structural Intervention Distance (SID)
  - count how many node pairs (i, j) exist where G* would produce a **different intervention distribution** than G.



G



G*

- Example:
  - (X,Y), (Y,X), (Z,X), (Z,Y): no difference
  - (X,Z), (Y,Z): different (Y -> Z not in G*)
  - SID = 2

Other Causal Discovery Models

Score-Based

- Search for the best-fitting graph by optimizing a scoring function like BIC or likelihood.
- GES[3], FGES[4]

Functional Form-Based

- Assume specific functional forms (e.g., additive noise) to infer causal direction.
- LiNGAM[5], ANM[6]

Optimization-Based

- Frame structure learning as a continuous optimization problem over graphs with acyclicity constraints.
- NOTEARS[7], DAG-GNN[8]

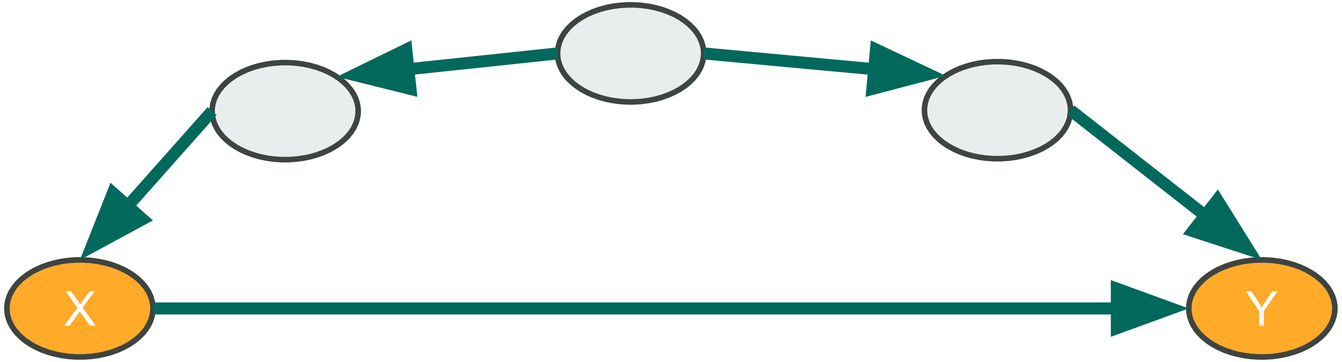
Temporal Data

- Extend causal discovery to time series by accounting for time lags and autocorrelation.
- PCMCI+[9], VAR-LiNGAM[10]

Task: Causal Effect Estimation

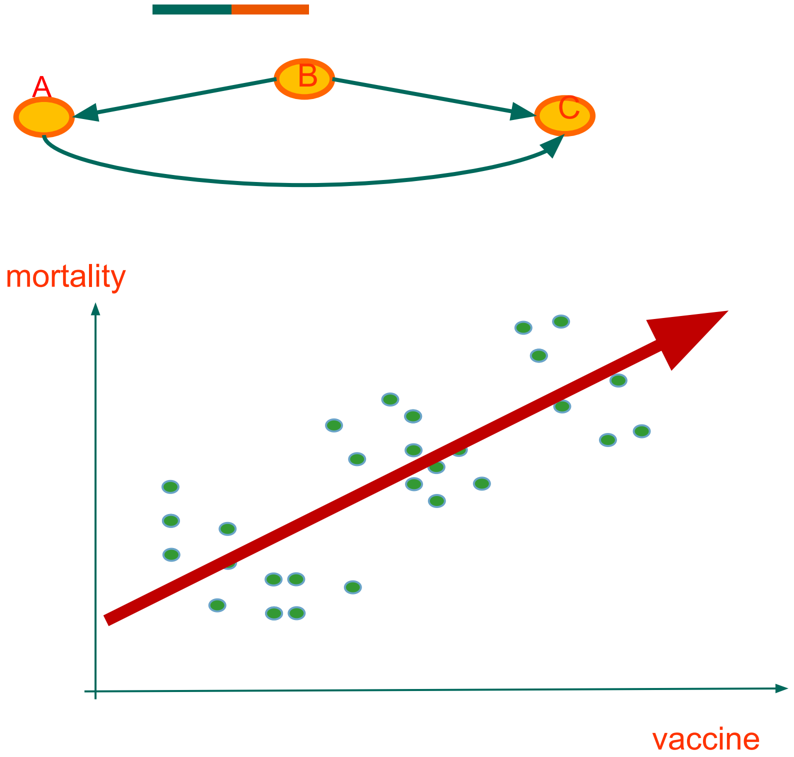
1. What is Causal Effect Estimation?
2. Backdoor Adjustment
3. Treatment Effect
4. Model: S-Learner
5. Model: T-Learner
6. Metrics

Causal Effect Estimation

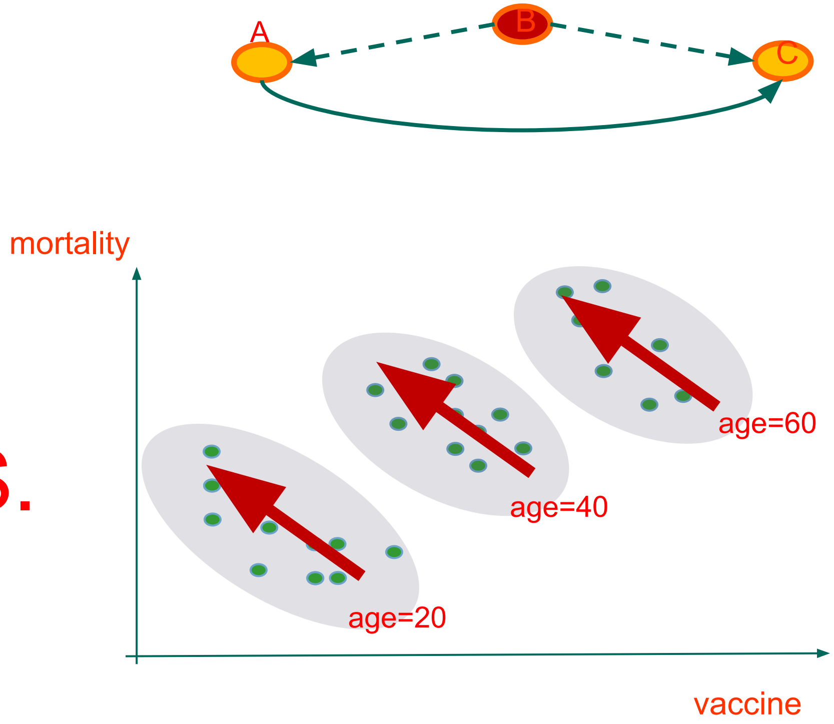


If I apply a particular treatment on X, what would its effect be on Y?

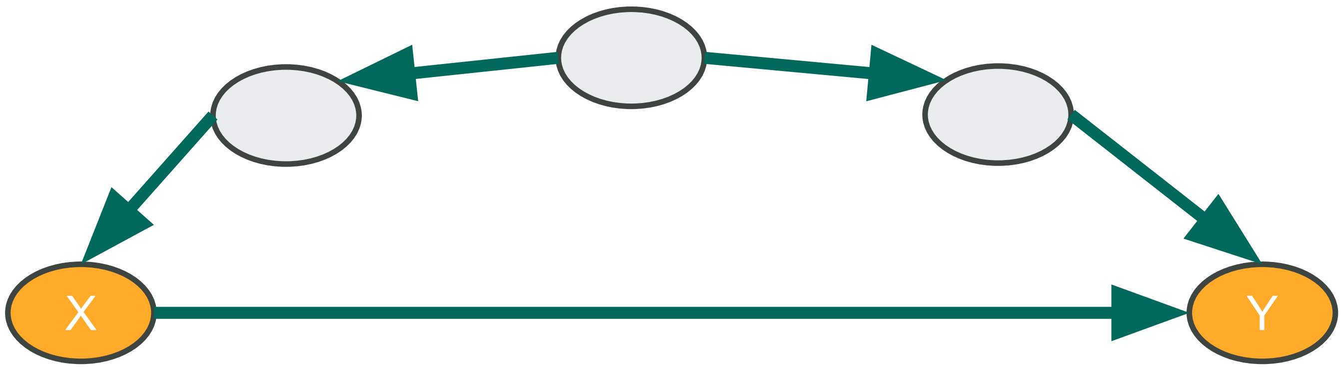
Eliminating confounding effects through conditioning…



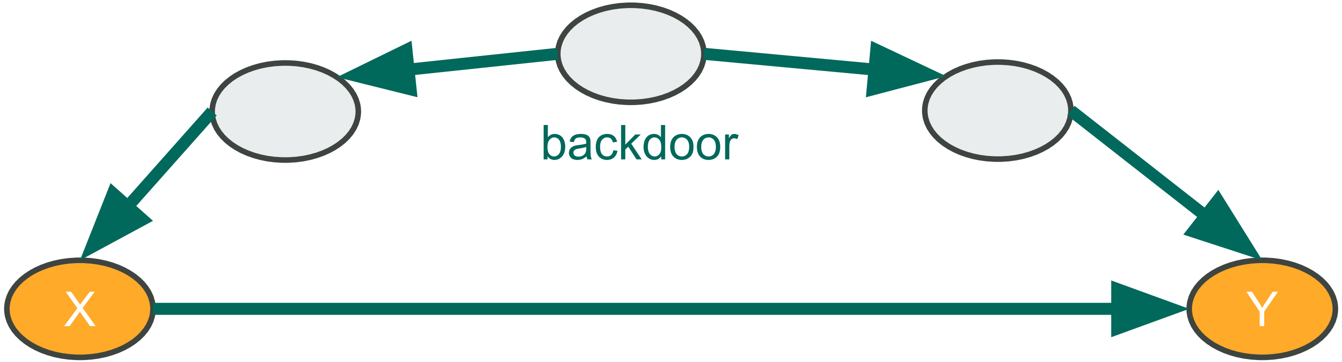
VS.



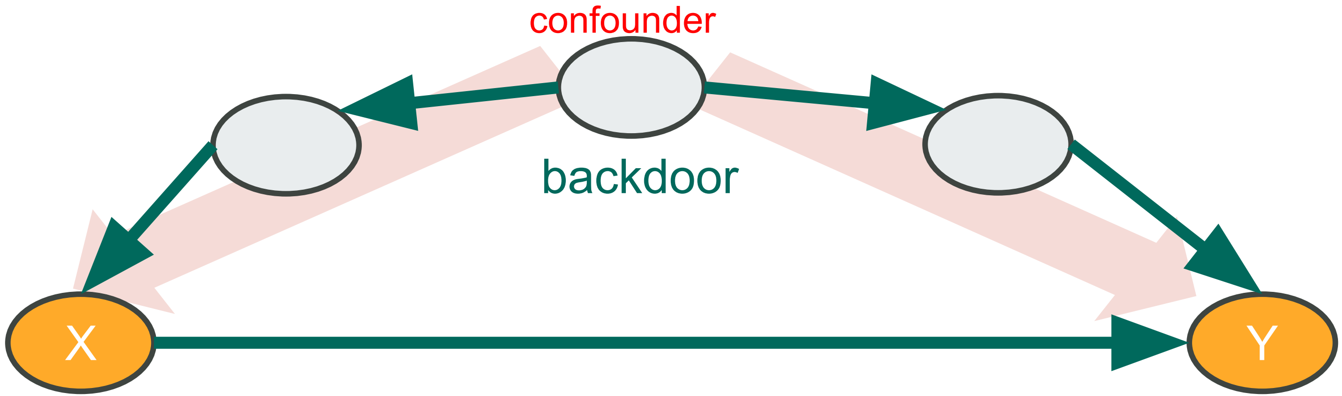
More generally...



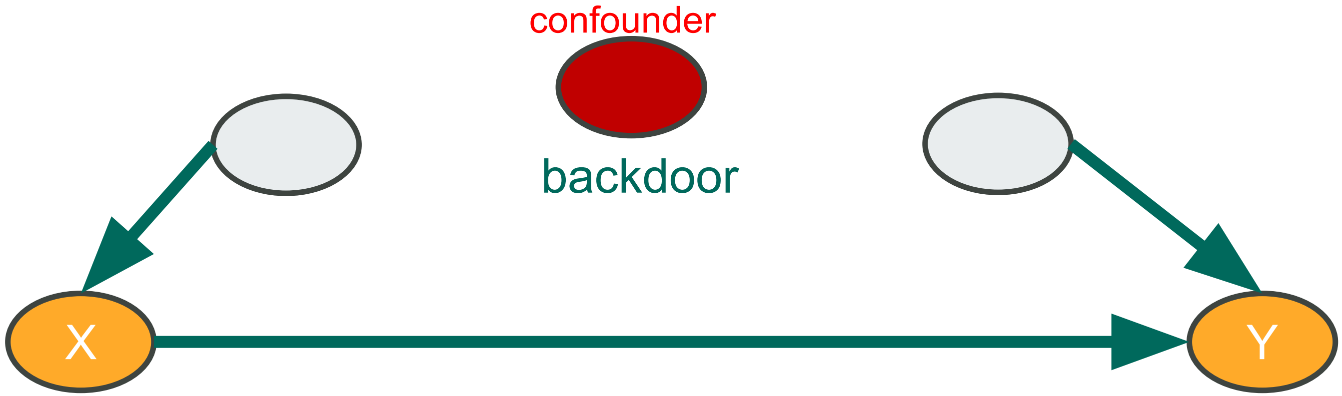
More generally...



More generally...



More generally...



Key Concept: Treatment Effect

Individual Treatment Effect (ITE)

- The effect of treatment on a single unit
  - $ITE_i = Y_i(1) - Y_i(0)$
- ITE(Emily) = 0 - 1 = -1

Customer	Y	$Y(0)$	$Y(1)$	T	X
Emily	1	0	1	1	0
Michael	0	0	1	0	0
Olivia	0	0	1	0	1
David	1	0	1	1	0
Sophia	0	0	1	0	0
James	1	1	1	1	1
Charlotte	1	1	1	0	1
Ethan	0	0	0	1	0
Ava	0	0	1	0	0
Benjamin	1	0	1	1	0

Key Concept: Treatment Effect

Individual Treatment Effect (ITE)

- The effect of treatment on a single unit
  - $ITE_i = Y_i(1) - Y_i(0)$
- ITE(Emily) = 0 - 1 = -1

Conditional Average Treatment Effect (CATE)

- The average effect given a subgroup or covariates X:
  - CATE(X) = E[Y(1) - Y(0) | X]
- CATE(X = 0) = [(1-0)*6 + (0-0)]/7 = 0.85

Customer	Y	$Y(0)$	$Y(1)$	T	X
Emily	1	0	1	1	0
Michael	0	0	1	0	0
Olivia	0	0	1	0	1
David	1	0	1	1	0
Sophia	0	0	1	0	0
James	1	1	1	1	1
Charlotte	1	1	1	0	1
Ethan	0	0	0	1	0
Ava	0	0	1	0	0
Benjamin	1	0	1	1	0

Key Concept: Treatment Effect

Individual Treatment Effect (ITE)

- The effect of treatment on a single unit
  - $ITE_i = Y_i(1) - Y_i(0)$
- ITE(Emily) = 0 - 1 = -1

Conditional Average Treatment Effect (CATE)

- The average effect given a subgroup or covariates X:
  - CATE(X) = E[Y(1) - Y(0) | X]
- CATE(X = 0) = [(1-0)*6 + (0-0)]/7 = 0.85

Average Treatment Effect (ATE)

- The overall average effect across the population:
  - ATE = E[Y(1) - Y(0)]
- ATE = [(1 − 0) * 7 + (1 − 1) * 2 + (0 − 0)] / 10 = 0.3

Customer	Y	$Y(0)$	$Y(1)$	T	X
Emily	1	0	1	1	0
Michael	0	0	1	0	0
Olivia	0	0	1	0	1
David	1	0	1	1	0
Sophia	0	0	1	0	0
James	1	1	1	1	1
Charlotte	1	1	1	0	1
Ethan	0	0	0	1	0
Ava	0	0	1	0	0
Benjamin	1	0	1	1	0

Common Assumptions for Causal Inference

- **Stable Unit Treatment Value Assumption (SUTVA)**
  - No interference among units.
    - Violation: social network, treated units may more likely interact with other treated units. Spillover effect.
  - Treatment gives the same outcome under the same conditions.
    - Also known as *consistency*.
- Positivity
- Unconfoundedness (also known as Ignorability)

Common Assumptions for Causal Inference

- Stable Unit Treatment Value Assumption (SUTVA)
    - No interference among units.
        - Violation: social network, treated units may more likely interact with other treated units. Spillover effect.
    - Treatment gives the same outcome under the same conditions.
        - Also known as *consistency*.
- **Positivity**
    - Have samples in both control and treated groups.
    - $0 < P(T = 1|X) < 1$
    - Violation: all samples are assigned to a single group.
- Unconfoundedness (also known as Ignorability)

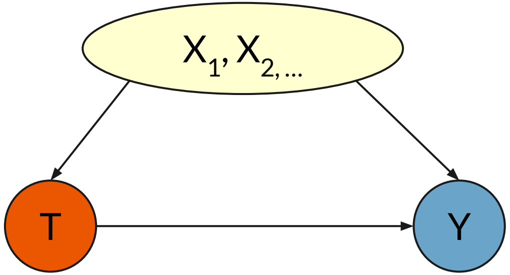
Common Assumptions for Causal Inference

- Stable Unit Treatment Value Assumption (SUTVA)
  - No interference among units.
    - Violation: social network, treated units may more likely interact with other treated units. Spillover effect.
  - Treatment gives the same outcome under the same conditions.
    - Also known as *consistency*.
- Positivity
  - Have samples in both control and treated groups.
  - $0 < P(T = 1|X) < 1$
  - Violation: all samples are assigned to a single group.
- **Unconfoundedness**
  - Also known as Ignorability
  - All confounders are observed. No unmeasured confounders.
  - Latent confounders can't be controlled directly, leave the backdoor path open.

Models: Meta Learner - T-Learner

Key idea:

- Control for confounders (X_1, X_2, ...) to block the backdoor path.
- Separate datasets to control (T=1) and treated (T=0) groups, and fit a corresponding model.
- Full flexibility on the choice of models



- T: binary treatment, vaccine
- Y: outcome, mortality rate
- $X = \{X_1, X_{2,\dots}\}$: confounders, such as age, health, etc.

Models: Meta Learner - T-Learner

Steps

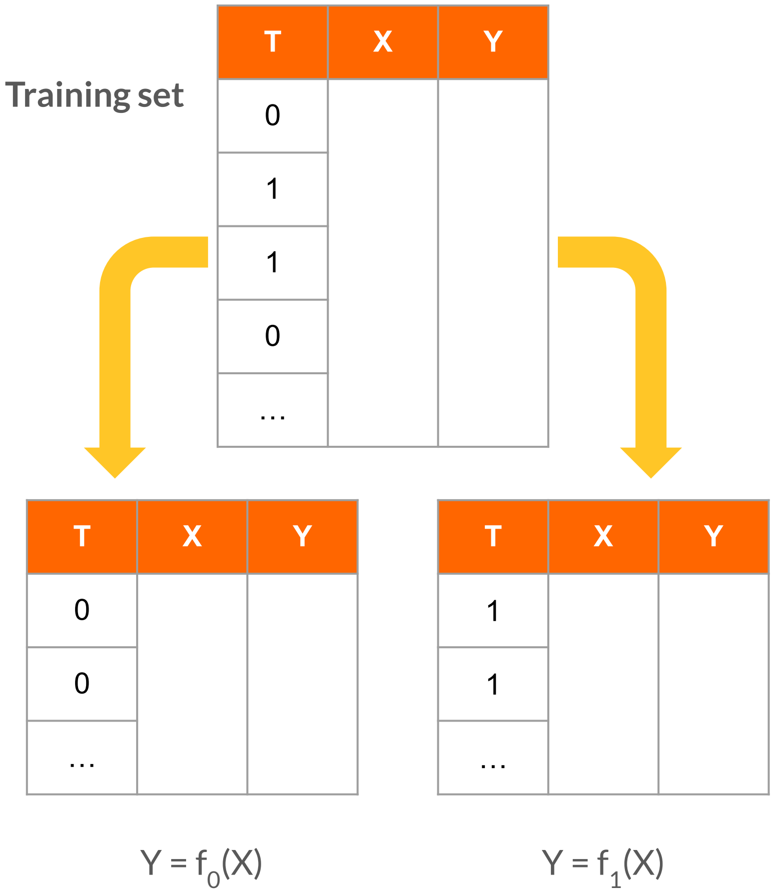
1. Choose any regression model f (e.g., linear model, random forest, neural net).

T	X	Y
0		
1		
1		
0		
…		

Models: Meta Learner - T-Learner

Steps

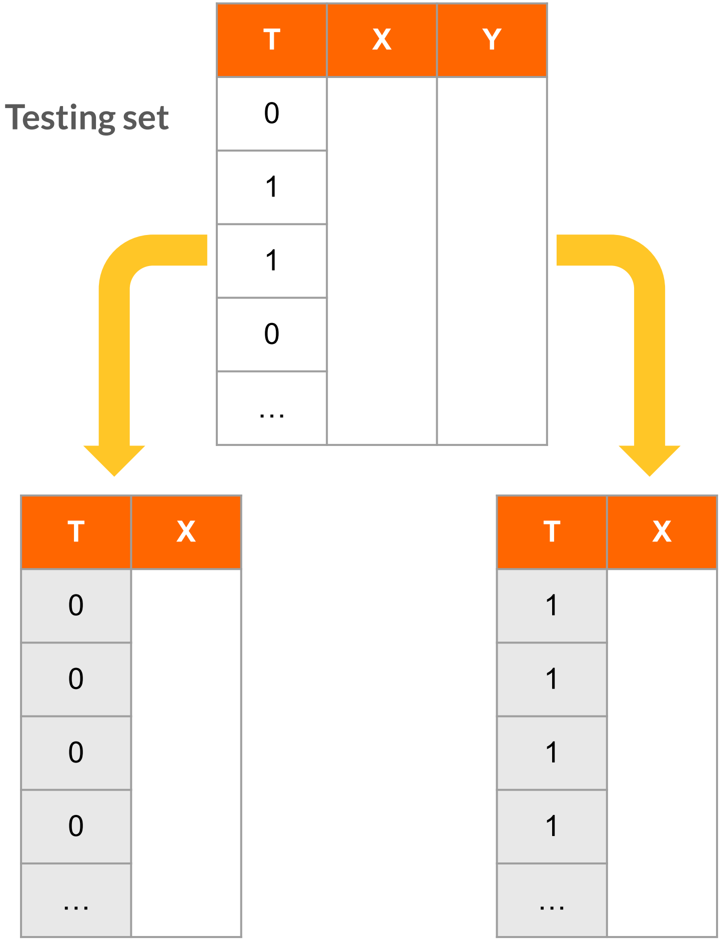
1. Choose any regression model f (e.g., linear model, random forest, neural net).
2. Split training data by treatment T.
   a. Fit $Y_1 = f_1(X)$ on samples with T =1
   b. Fit $Y_0 = f_0(X)$ on samples with T =0



Models: Meta Learner - T-Learner

Steps

1. Choose any regression model f (e.g., linear model, random forest, neural net).
2. Split training data by treatment T.
   a. Fit $Y_1 = f_1(X)$ on samples with T =1
   b. Fit $Y_0 = f_0(X)$ on samples with T =0
3. On the test set
   a. Create two copies of testing dataset with the same confounders X.
   b. Set T = 1 in one copy, T = 0 in the other.
   c. Use Y_1 to predict T=1, and Y_0 to predict T= 0
4. Estimate:
   a. $ATE = E_X[\, f_1(T=1, X) - f_0(T=0, X)]$
   b. $CATE(X) = E[\, f_1(T=1, X) - f_0(T=0, X)]$
   c. $ITE = f_1(T=1, X_i) - f_0(T=0, X_i)$



Metrics

ATE, CATE

- Compare with ground truth directly (synthetic data).
- On real-world data, we often assume the test set is **balanced** across treatment groups to approximate these comparisons.

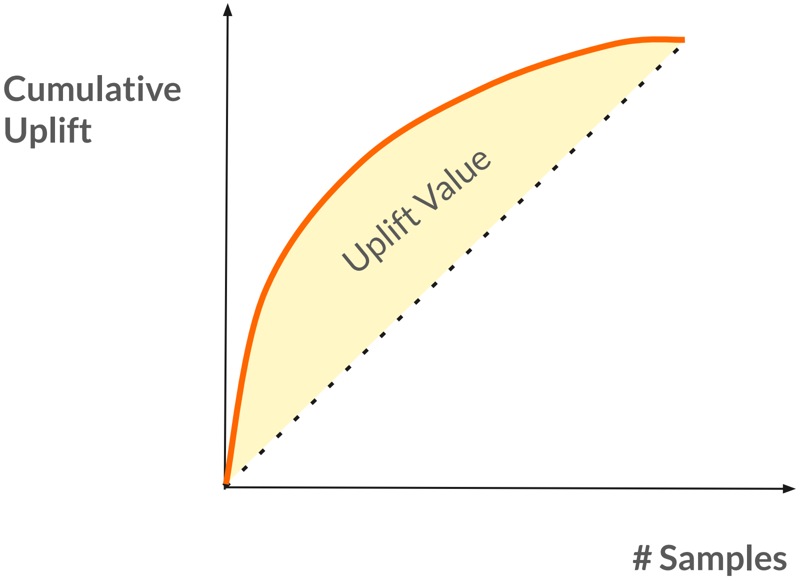
ITE

- Can't observe both treated and untreated outcomes for the same individual.
- Direct evaluation of ITE is impossible on real-world data.

Metrics

Uplift Curve

- Rank individuals by predicted ITE, from highest to lowest.
- Partition the dataset into percentiles (e.g., deciles) by rank (g_1, g_2, …).
- In each group, compute the observed uplift:
  - Uplift = E[Y | T = 1, g] - E[Y | T = 0, g]
- Plot cumulative uplift (Y-axis) vs. % of population targeted (X-axis) → Qini curve.
- Uplift Value:
  - Area between model's Uplift curve and random targeting line.
  - Standardized (by sample count) uplift curve is called Qini curve.



Other Causal Effect Models

Matching [13]

- Compare samples in treated and control groups with similar characteristics.
- Mimics a randomized experiment by balancing covariates.

IPW (Inverse Probability Weighting) [15]

- Reweights individuals to create a balanced virtual population.
- Especially useful when treated and untreated groups are very different.

Meta-Learners [16]

- Use machine learning to estimate outcomes under treatment and control separately.
- Adaptable to flexible models and heterogeneous effects.

Causal Forests [17]

- Tree-based method that learns how treatment effects vary across individuals.
- Automatically finds subgroups with different effects.

Recap and Resources

Causal Discovery

- Constraint-based: PC, FCI
- Score-based: GES, FGES
- Functional: LiNGAM, ANM
- Optimization-based: NOTEARS, DAG-GNN
- Temporal: PCMCI+, VAR-LiNGAM

Causal Effect Estimation

- Regression-based: Linear regression, GLMs
- Matching: Propensity score, Mahalanobis
- IPW (Inverse Probability Weighting)
- Meta-learners: S-Learner, T-Learner, X-Learner, R-Learner
- Causal Forests
- DML (Double Machine Learning)

DoWhy

CAUSAL-LEARN
Python package for causal discovery

EconML

References

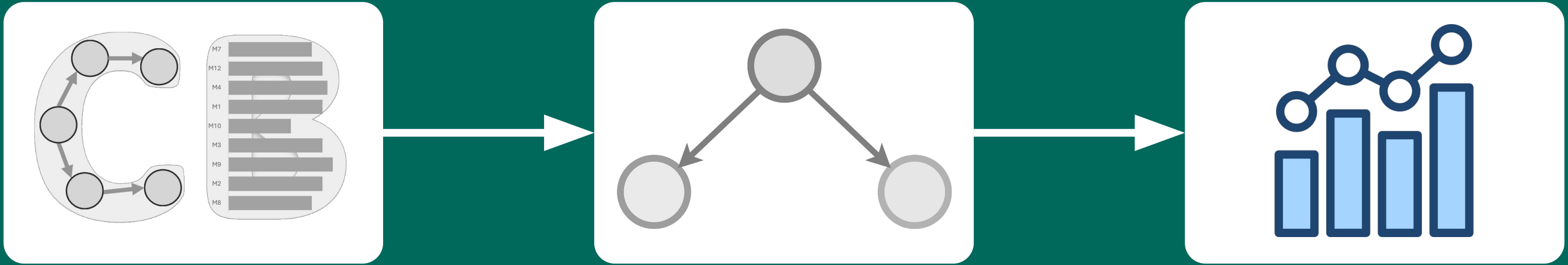
1. Spirtes, Peter, and Clark Glymour. "An algorithm for fast recovery of sparse causal graphs." *Social science computer review* 9.1 (1991): 62-72.
2. Spirtes, Peter. "An anytime algorithm for causal inference." *International Workshop on Artificial Intelligence and Statistics*. PMLR, 2001.
3. Chickering, David Maxwell. "Optimal structure identification with greedy search." *Journal of machine learning research* 3.Nov (2002): 507-554.
4. Ramsey, Joseph, et al. "A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images." *International journal of data science and analytics* 3.2 (2017): 121-129.
5. Shimizu, Shohei, et al. "A linear non-Gaussian acyclic model for causal discovery." *Journal of Machine Learning Research* 7.10 (2006).
6. Peters, Jonas, et al. "Causal discovery with continuous additive noise models." *The Journal of Machine Learning Research* 15.1 (2014): 2009-2053.
7. Zheng, Xun, et al. "Dags with no tears: Continuous optimization for structure learning." *Advances in neural information processing systems* 31 (2018).
8. Yu, Yue, et al. "DAG-GNN: DAG structure learning with graph neural networks." *International conference on machine learning*. PMLR, 2019.
9. Runge, Jakob. "Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets." *Conference on uncertainty in artificial intelligence*. Pmlr, 2020.
10. Hyvärinen, Aapo, et al. "Estimation of a structural vector autoregression model using non-Gaussianity." *Journal of Machine Learning Research* 11.5 (2010).
11. Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
12. Nelder, John Ashworth, and Robert WM Wedderburn. "Generalized linear models." *Journal of the Royal Statistical Society Series A: Statistics in Society* 135.3 (1972): 370-384.
13. Holland, Paul W. "Statistics and causal inference." *Journal of the American statistical Association* 81.396 (1986): 945-960.
14. Rubin, Donald B. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66.5 (1974): 688.
15. Horvitz, Daniel G., and Donovan J. Thompson. "A generalization of sampling without replacement from a finite universe." *Journal of the American statistical Association* 47.260 (1952): 663-685.
16. Künzel, Sören R., et al. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the national academy of sciences* 116.10 (2019): 4156-4165.
17. Wager, Stefan, and Susan Athey. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113.523 (2018): 1228-1242.
18. Chernozhukov, Victor, et al. "Double/debiased machine learning for treatment and structural parameters." (2018): C1-C68.

End of Deck 2

Any Questions?

CausalBench: Causal Learning Research Streamlined

Hands-On Benchmarking



CausalBench | Causality | **Benchmarking**

tutorial.causalbench.org

August 3–7, 2025
KDD2025

This research is funded by NSF Grant 2311716, "CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration", and NSF Grants #2230748, "PIRE: Building Decarbonization via AI-empowered District Heat Pump Systems", #2412115, "PIPP Phase II: Analysis and Prediction of Pandemic Expansion (APPEX)" and USACE #GR40695, "Designing nature to enhance resilience of built infrastructure in western US landscapes".

Installing CausalBench Python Package

Getting started

- https://tutorial.causalbench.org/
- Google Colab Notebook (Jupyter)

Prerequisites

- Python (>= 3.10)
- pip

```
$ pip install causalbench-asu
```

Additional requirements for this tutorial

- gcastle

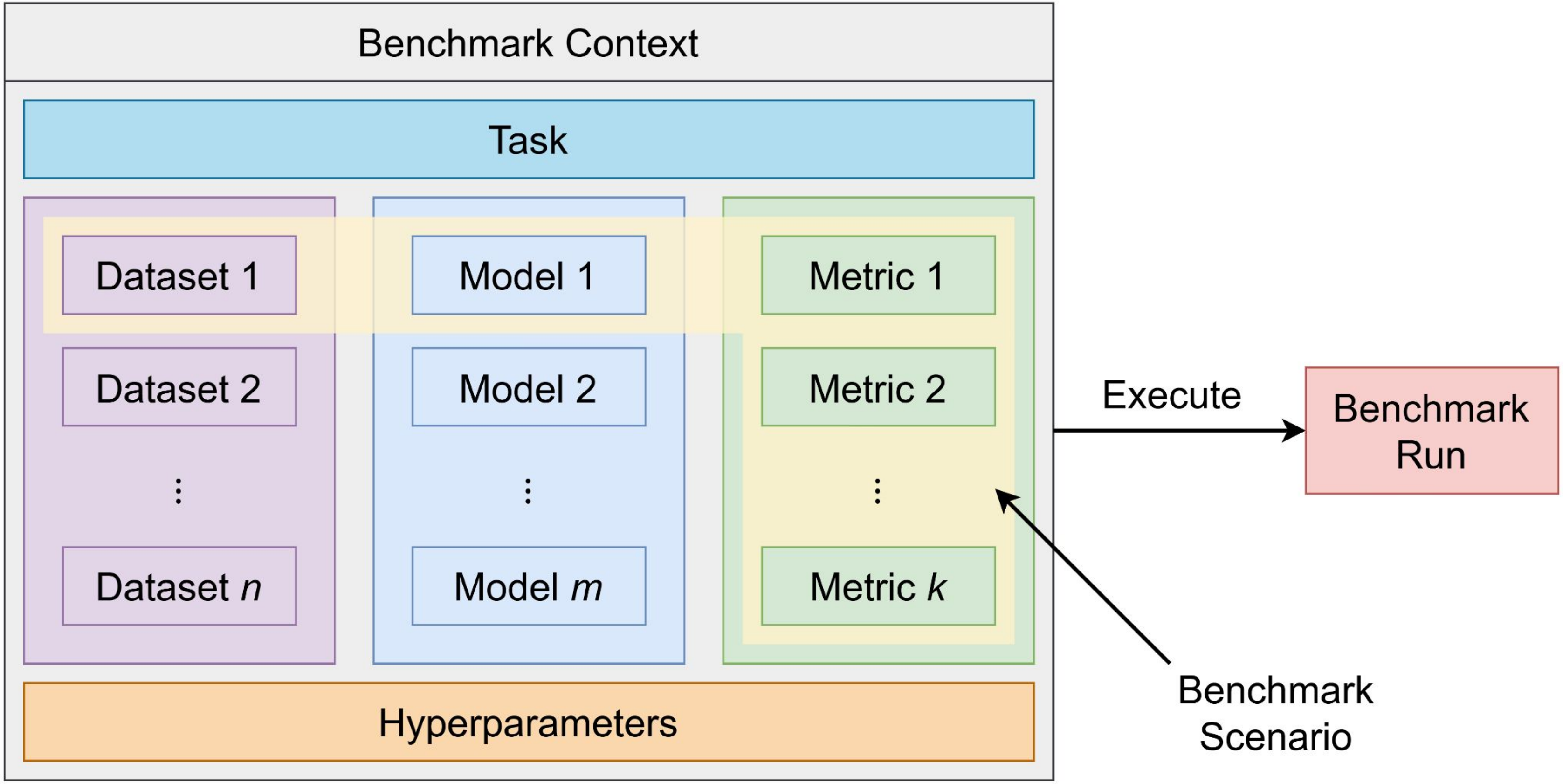
Using CausalBench Python Package

Next Steps

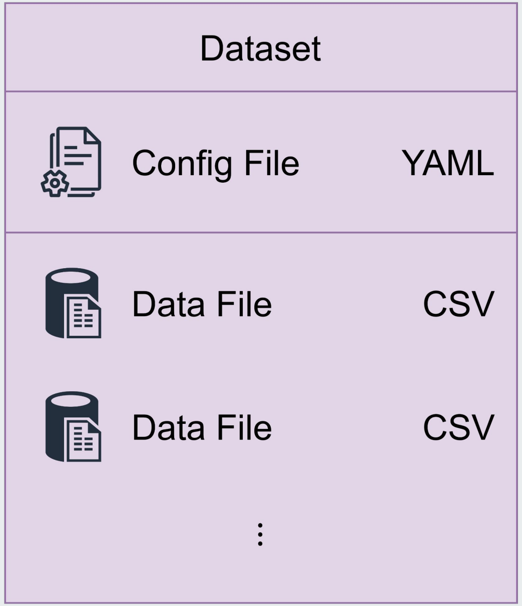
- Create an account on https://causalbench.org/
- Use credentials for first use of **CausalBench** Python package

```
Credentials required
Email: user@example.com
Password: 🔑
```

CausalBench: Modules



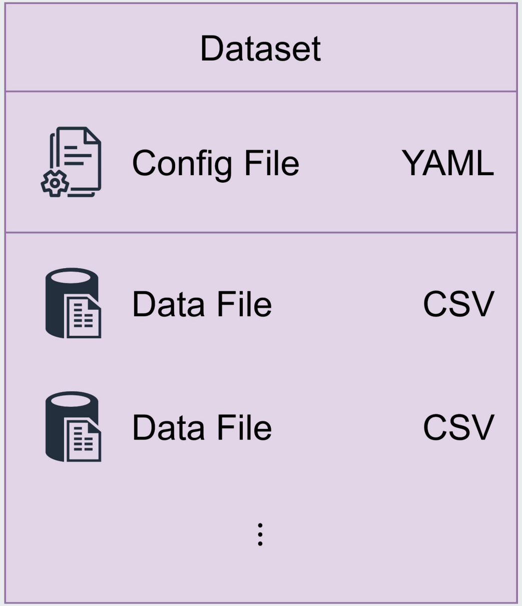
CausalBench: Modules (Dataset)



Config file:

- Metadata
  - name, description, and URL
- Names and metadata of data files
- Structure of data files
  - Number of rows
  - Columns — number, name, data type, description
  - Index — time, space, etc.

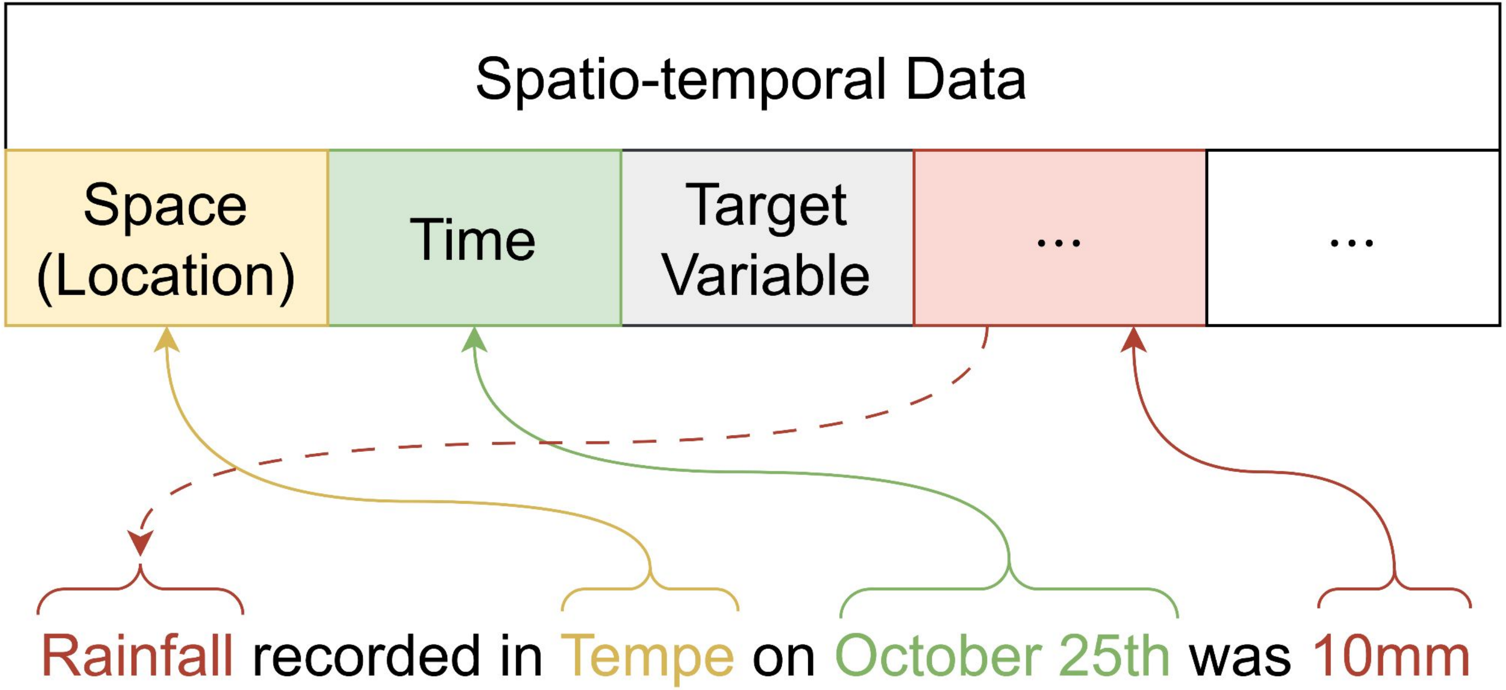
CausalBench: Modules (Dataset)



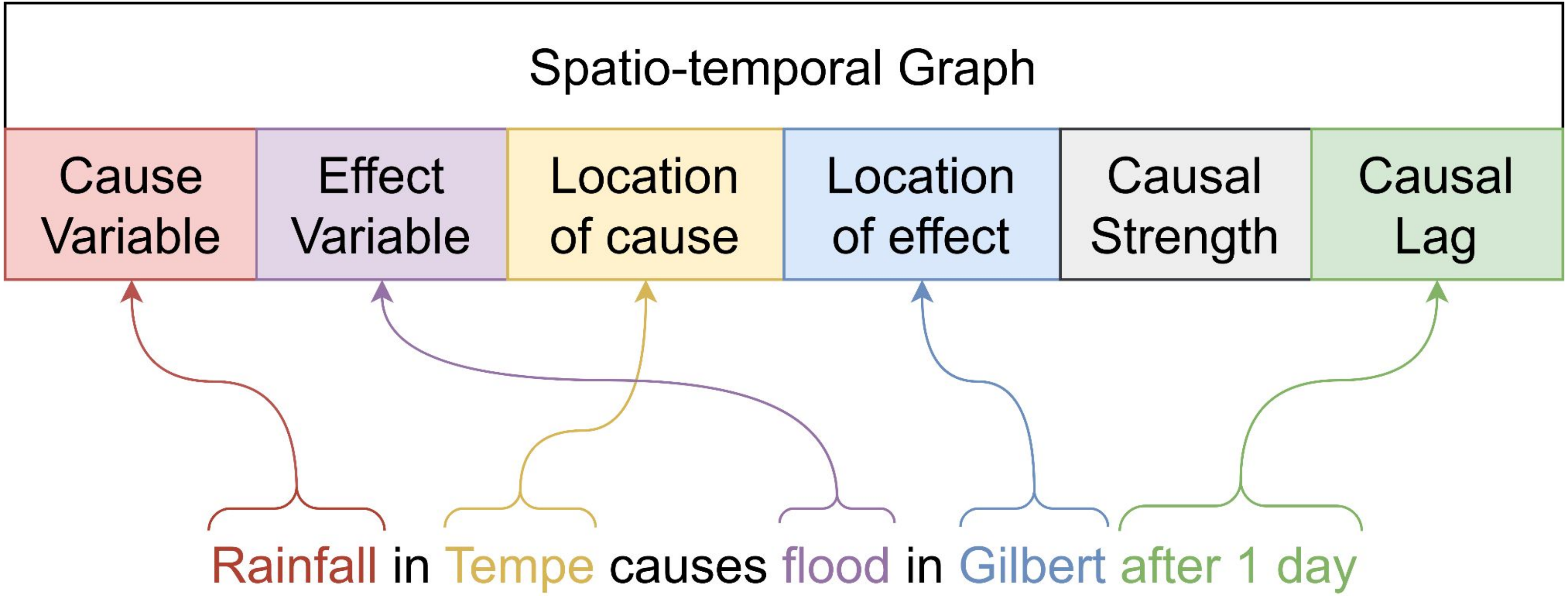
Data file:

- Tabular data
- Data formats
  - Spatio-temporal Data
  - Spatio-temporal Graph
- Helper functions
  - Static tabular data → Spatio-temporal data
  - Static adjacency matrix → Spatio-temporal graph

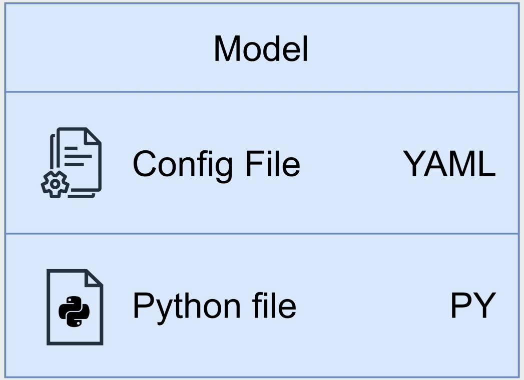
CausalBench: Modules (Dataset) – Data Formats



CausalBench: Modules (Dataset) – Data Formats



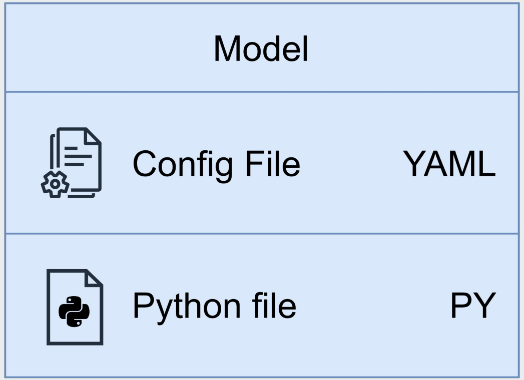
CausalBench: Modules (Model)



Config file:

- Metadata
  - name, description, and URL
- Name and metadata of Python file
- Task
  - Causal Discovery
  - Causal Inference, etc.
- Hyperparameters
  - data type, description, and default value

CausalBench: Modules (Model)



Python file:

- Function to take accept inputs and provide outputs
    - Comply with function signature specified by task

CausalBench: Modules (Metric)

Metric	
Config File	YAML
Python file	PY

Same structure as model

Config file:

- Metadata
  - name, description, and URL
- Name and metadata of Python file
- Task
  - Causal Discovery
  - Causal Inference, etc.
- Hyperparameters
  - data type, description, and default value

CausalBench: Modules (Metric)

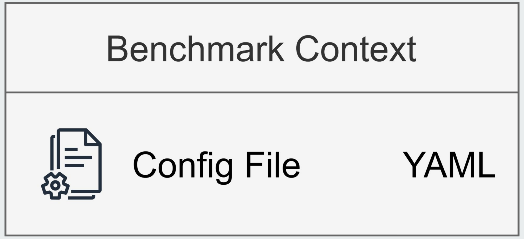
Metric	
Config File	YAML
Python file	PY

Same structure as model

Python file:

- Function to take accept inputs and provide outputs
    - Comply with function signature specified by task

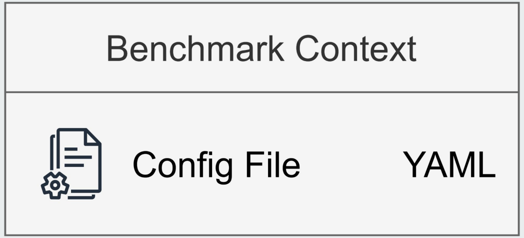
CausalBench: Modules (Benchmark Context)



Config file:

- Metadata
  - name, description, and URL
- Task
  - Causal Discovery
  - Causal Inference, etc.
- Datasets
  - Dataset IDs and versions
  - Data files to task mapping

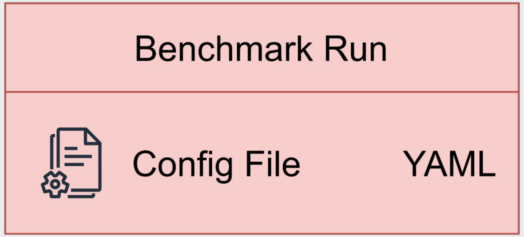
CausalBench: Modules (Benchmark Context)



Config file:

- Models
  - Model IDs and versions
  - Model hyperparameters (if not using default values)
- Metrics
  - Metric IDs and versions
  - Metric hyperparameters (if not using default values)

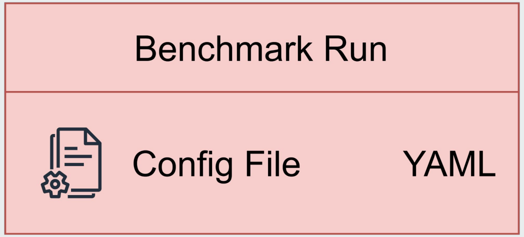
CausalBench: Modules (Benchmark Run)



Config file:

- Reference to Benchmark Context
  - Benchmark Context ID and version
- Platform information
  - Operating System
  - CPU
  - GPU
  - Memory
  - Disk

CausalBench: Modules (Benchmark Run)



Config file:

- Scenarios
  - Consists of 1 dataset, 1 model, and multiple metrics
  - Dataset
    - ID and version
  - Model and Metrics
    - IDs and versions
  - Output / results
  - Profiling information
    - Execution time
    - Hardware utilization
    - Software packages

End of Deck 3

Any Questions?

Agenda for today's Hands-on Tutorial

tutorial.causalbench.org

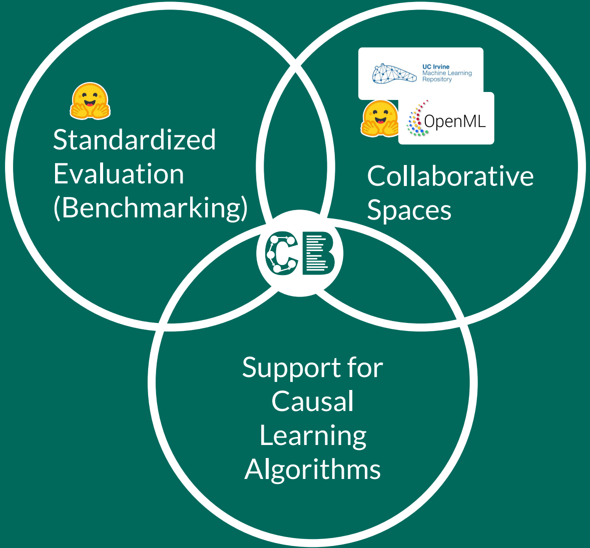
See you back at 10am!

08:00-08:05	~~Introduction to the Tutorial~~
08:05-08:25	~~Introduction to **CausalBench**~~
08:25-08:55	~~Introduction to Causality and Causal Learning~~
08:55-09:30	~~Delve into the **CasualBench** framework to create and execute benchmarks~~
09:30-10:00	**Coffee break**
10:00-10:10	Shorter introduction to **CausalBench**
10:10-10:35	Explore published benchmarks and reproduce experiments
10:35-10:50	Gain further insights using Causal Explanation and Recommendations
10:50-11:00	**CausalBench**: What's Next?

CausalBench: Causal Learning Research Streamlined

Ahmet Kapkiç *
Pratanu Mandal *
Abhinav Gorantla *
Shu Wan
Ertuğrul Çoban
Dr. Paras Sheth
Dr. Huan Liu *
Dr. K. Selçuk Candan

* In-person presenters

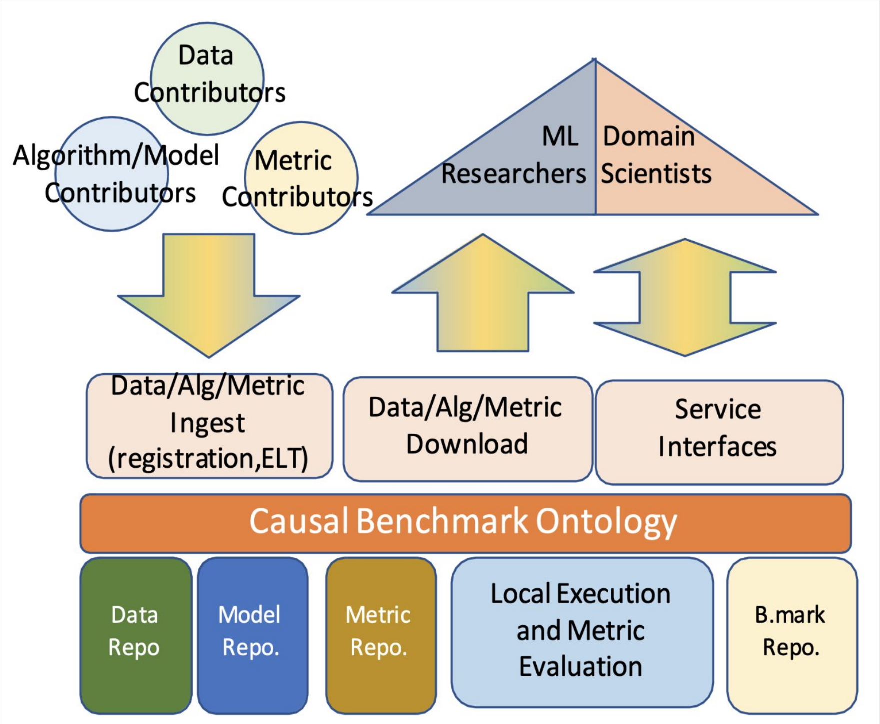


tutorial.causalbench.org

August 3–7, 2025
KDD2025

This research is funded by NSF Grant 2311716, "CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration", and NSF Grants #2230748, "PIRE: Building Decarbonization via AI-empowered District Heat Pump Systems", #2412115, "PIPP Phase II: Analysis and Prediction of Pandemic Expansion (APPEX)" and USACE #GR40695, "Designing nature to enhance resilience of built infrastructure in western US landscapes".

What is CausalBench?

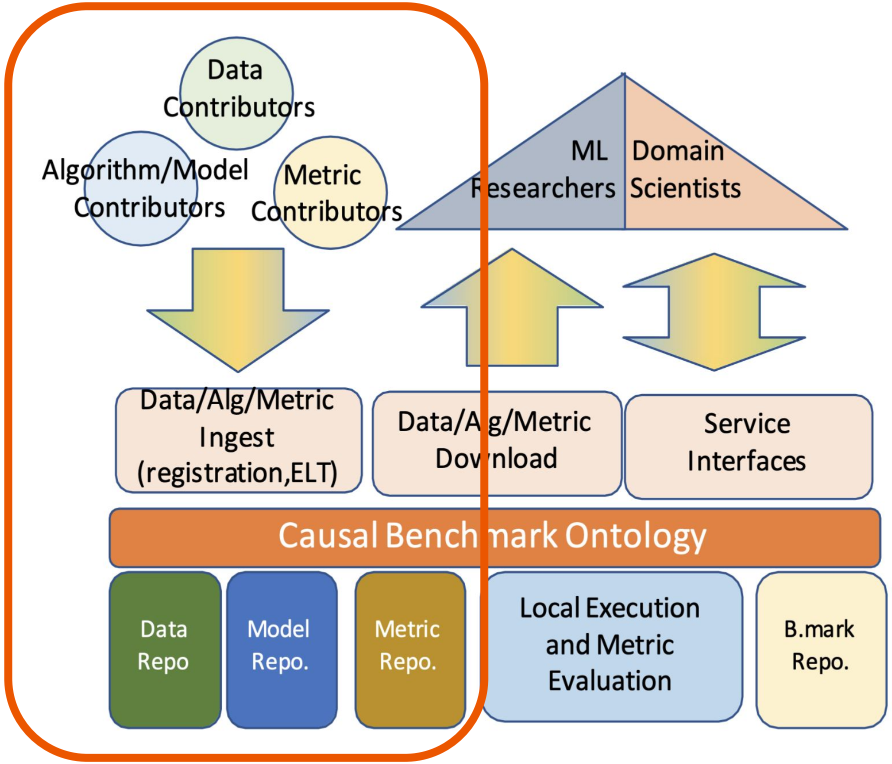


- **CausalBench** is a benchmarking platform for [Causal Learning research].

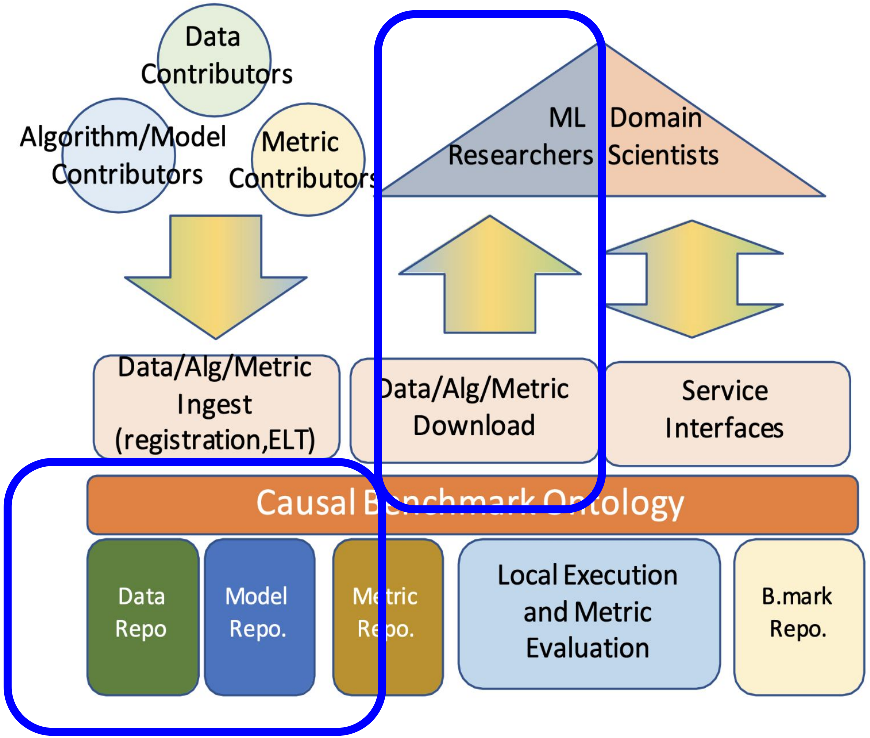
- Goals:
  - Promoting universal adoption of standard datasets, metrics and procedures for causal learning.
  - Facilitating collaboration.
  - Trustable and reproducible benchmarking.
  - Fair and flexible comparison of models.

CausalBench: Use Scenario #1

Data, model, metric contributor

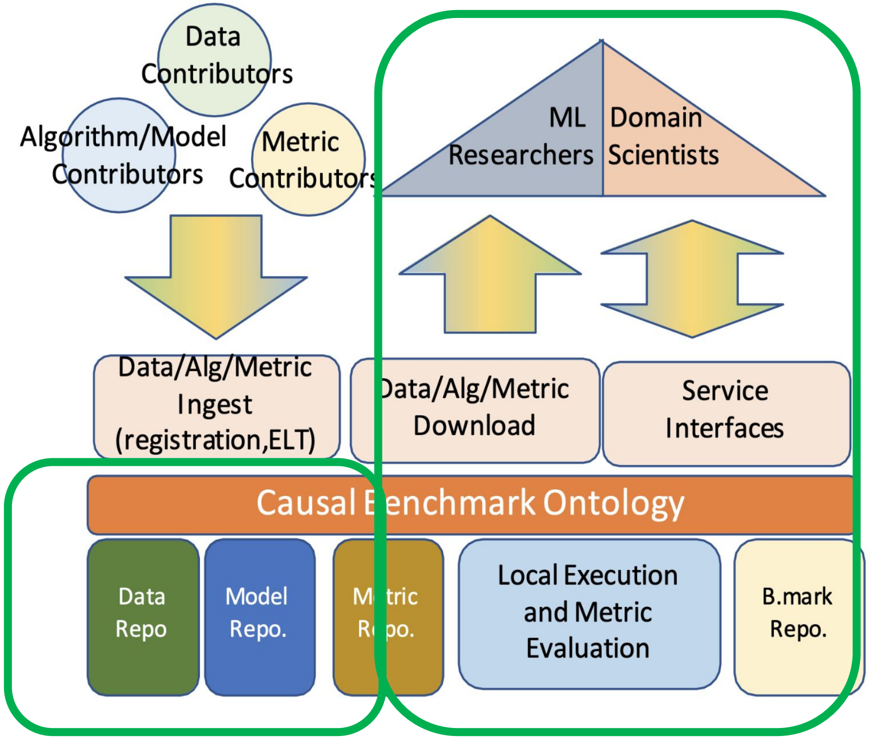


CausalBench: Use Scenario #2



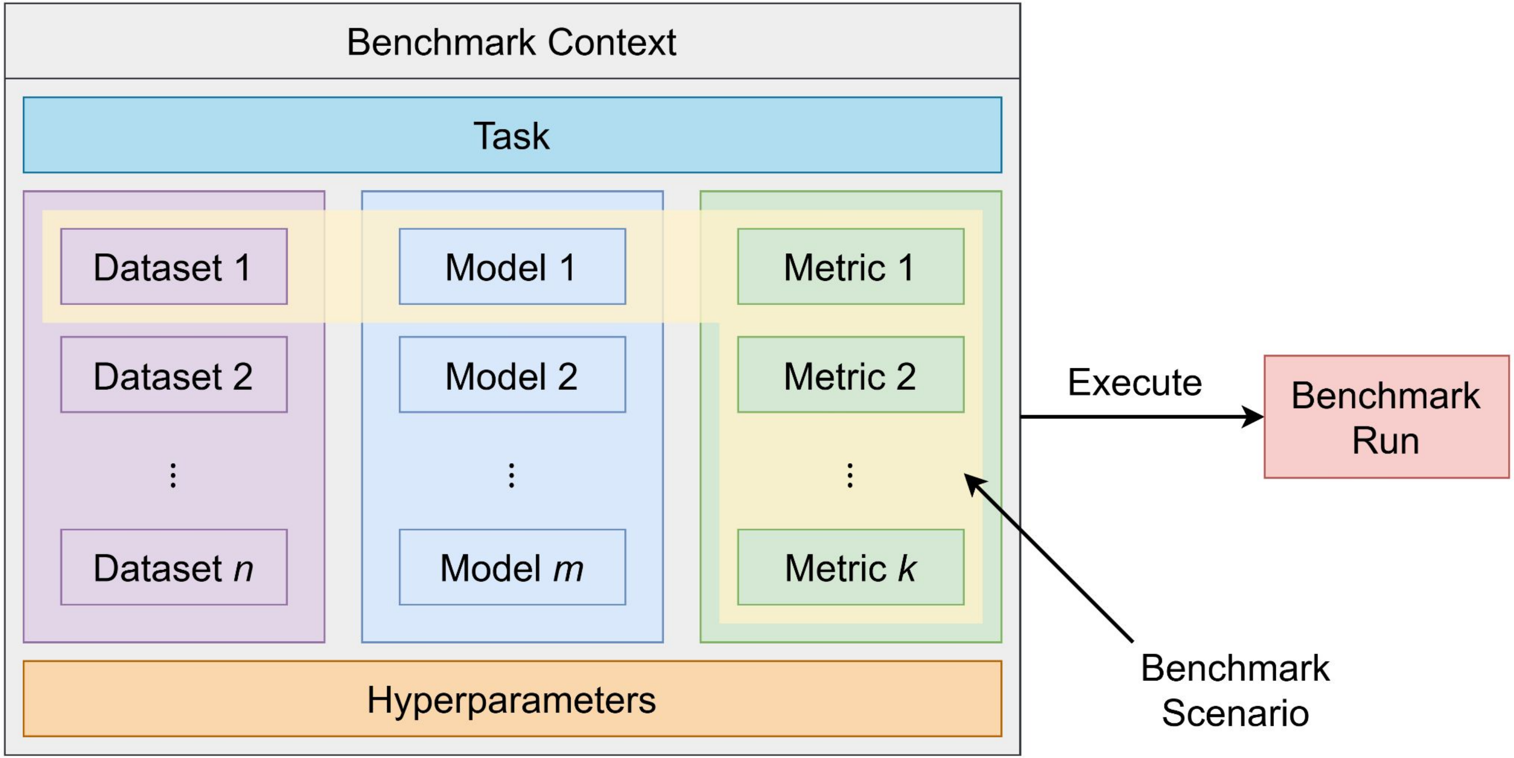
Data, Model, Metric Explorer

CausalBench: Use Scenario #3

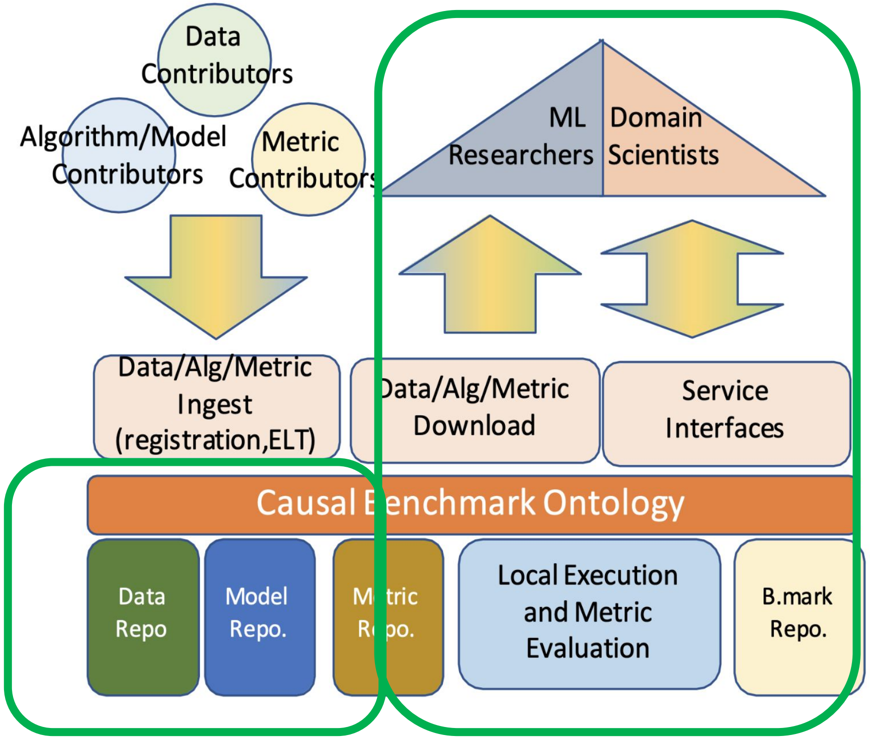


Benchmark executor

What is a Benchmark?



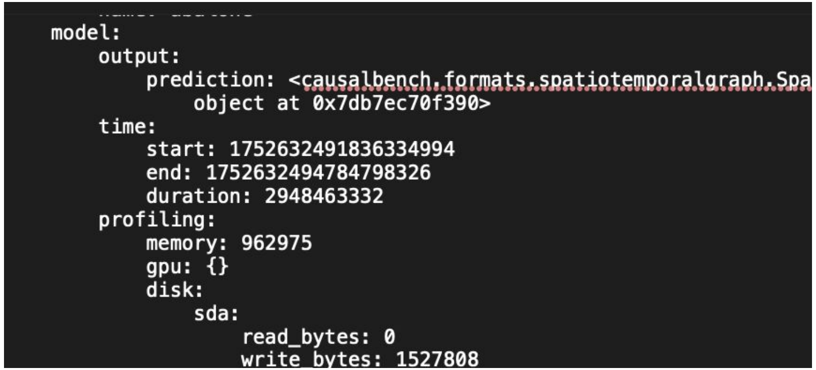
CausalBench: Use Scenario #3



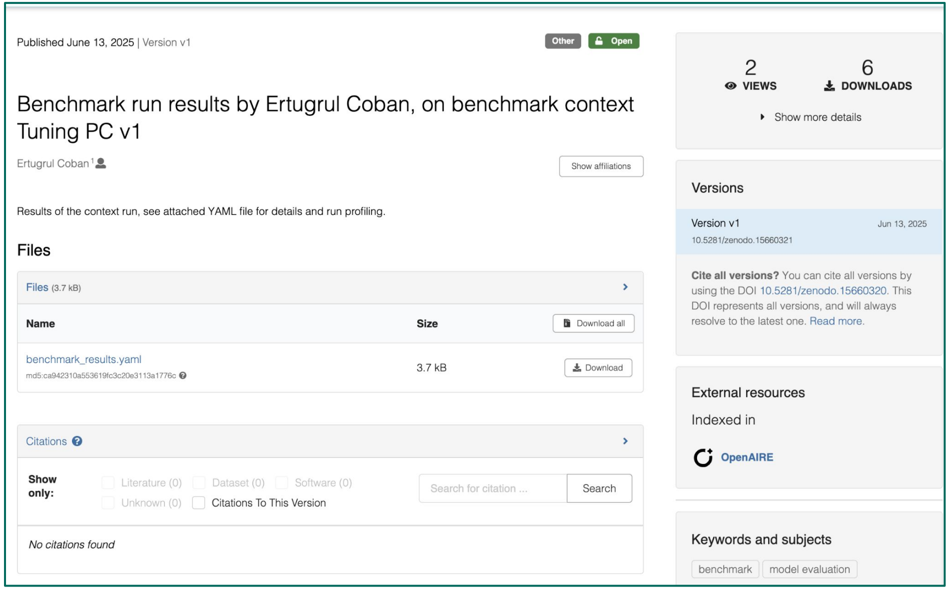
Benchmark executor

Sample benchmark output

- Includes
  - Model
  - Dataset
  - Hardware/software profiling
  - Accuracy metrics

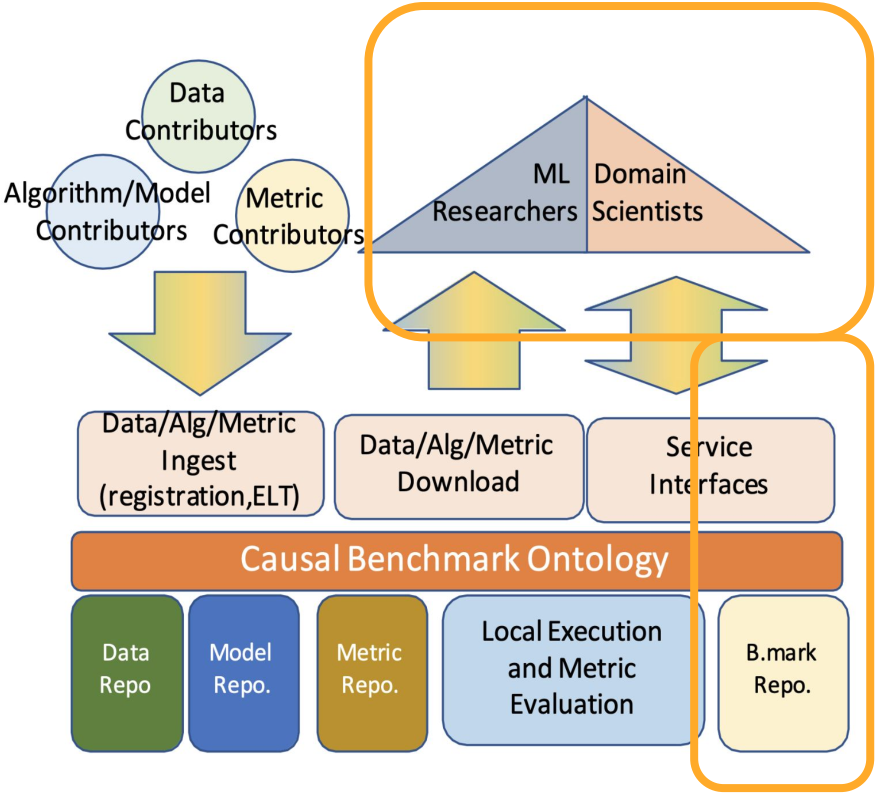


Uploaded and stored in CausalBench



Permanently indexed (and citable) in Zenodo

CausalBench: Use Scenario #4



Benchmark explorer

CausalBench: Explaining benchmark results

Result ID	Dataset	Model	Metric	Context	Result	Duration	Created On	Run Published By	Actions	Visibility
630	time_sim	VAR-LiNGAM	accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.9375	8.31 seconds	February 16th 2025	Abhinav Gorantla (agorant2@asu.edu)	doi	PUBLIC
640	Short-term electricity load forecasting (Panama)	VAR-LiNGAM	accuracy_temporal	Benchmark: VAR-LiNGAM, PCMCIplus	0.568359375	4.67 minutes	February 16th 2025	Abhinav Gorantla (agorant2@asu.edu)	doi	PUBLIC

- **Sample question:**
  - Why does VAR-LiNGAM have better accuracy with time_sim but lower training time in this benchmark?
    - Did the hyperparameters play a role?
    - Could it be because of the dataset size?
    - Is there something else?
- These questions can be answered by ***generating explanations*** using CausalBench.

What is Causal Learning?

Causal Learning answers the question of “**Why**” and describe the relationship between

- a **cause** (an action, event, or condition), and
- its **effect** (an outcome that results from it).

How do **socioeconomic status, nutrition, study time,** and, **sleep**, influence student **GPA**?

What’s the effect of the **vaccine** on a patient’s **health?**

Two Tasks in Causal Learning

How do **socioeconomic status, nutrition, study time,** and, **sleep**, influence student **GPA**?

Causal Discovery

We don't know what causes what. We want to uncover the structure — who influences whom.

What's the effect of the **vaccine** on a patient's **health?**

Two Tasks in Causal Learning

How do **socioeconomic status, nutrition, study time,** and, **sleep**, influence student **GPA**?

What's the effect of the **vaccine** on a patient's **health?**

Causal Discovery

We don't know what causes what. We want to uncover the structure — who influences whom.

Causal Effect Estimation

Knowing cause and effect, want to estimate how much effect one variable has on another.

Quick Recap: Hands-On Benchmarking

Benchmarking using CausalBench

- Static Causal Discovery
- Create and publish:
  - Dataset
  - Model
  - Metric
  - Benchmark Context
- Execute the Benchmark Context
- Publish generate Benchmark Run
- Explore the published modules on the CausalBench website

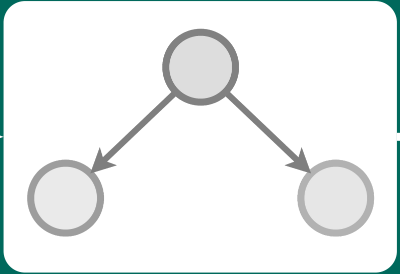
Agenda for today’s Hands-on Tutorial

tutorial.causalbench.org

08:00-08:05	~~Introduction to the Tutorial~~
08:05-08:25	~~Introduction to CausalBench~~
08:25-08:55	~~Introduction to Causality and Causal Learning~~
08:55-09:30	~~Delve into the CasualBench framework to create and execute benchmarks~~
09:30-10:00	~~Coffee break~~
10:00-10:10	Shorter introduction to CausalBench
10:10-10:35	Explore published benchmarks and reproduce experiments
10:35-10:50	Gain further insights using Causal Explanation and Recommendations
10:50-11:00	CausalBench: What's Next?

CausalBench: Causal Learning Research Streamlined

Hands-On Benchmarking



CausalBench → Causality → **Benchmarking**

tutorial.causalbench.org

August 3-7, 2025 KDD2025

This research is funded by NSF Grant 2311716, "CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration", and NSF Grants #2230748, "PIRE: Building Decarbonization via AI-empowered District Heat Pump Systems", #2412115, "PIPP Phase II: Analysis and Prediction of Pandemic Expansion (APPEX)" and USACE #GR40695, "Designing nature to enhance resilience of built infrastructure in western US landscapes".

Quick Recap: Installing CausalBench Python Package

Getting started

- https://tutorial.causalbench.org/
- Google Colab Notebook (Jupyter)

Prerequisites

- Python (>= 3.10)
- pip

```
$ pip install causalbench-asu
```

Additional requirements for this tutorial

- gcastle

Quick Recap: Using CausalBench Python Package

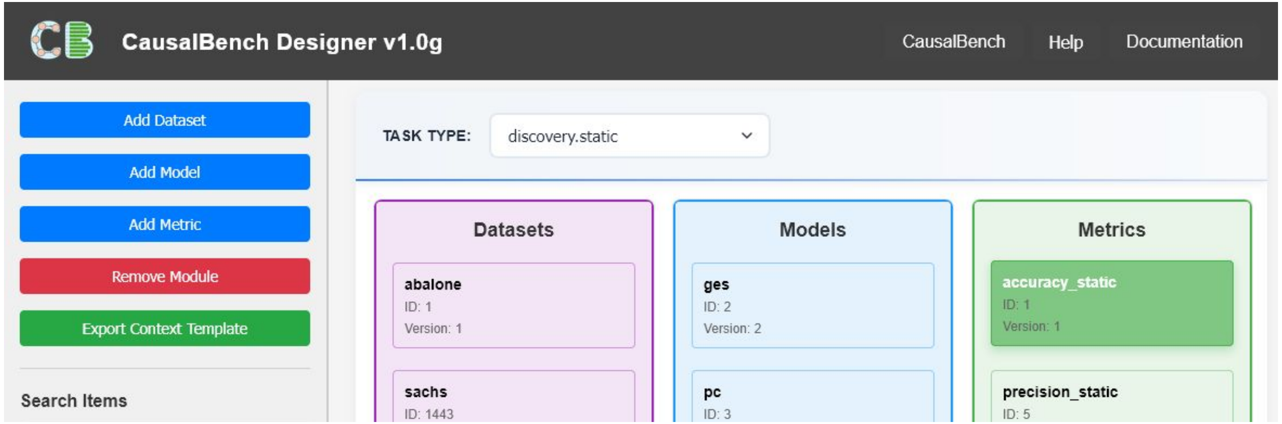
Next Steps

- Create an account on https://causalbench.org/
- Use credentials for first use of **CausalBench** Python package

```
Credentials required
Email: user@example.com
Password: 🔑
```

CausalBench Designer

- https://designer.causalbench.org/
- Design a Benchmark Context using GUI
- Zero coding environment



CausalBench Designer

Let's design a Benchmark Context!

- Task:
  - Static Causal Discovery
- Datasets:
  - Abalone
  - Sachs
- Models:
  - GES
  - PC

- Metrics:
  - Accuracy
  - Precision
  - Recall
  - F1-score
  - Structural Hamming Distance (SHD)

CausalBench: Reproducibility

How can we reproduce someone else's benchmark?

- Benchmark Context:
  - Module ID: 19
  - Version: 1

Fetch → Execute → Publish

CausalBench: Transparency

Provenance of public Benchmark Runs

Complete
- Should store any information collected during benchmarking

Available
- Anyone should be able to access and cite

Permanent
- Should always be available once made public
- Cannot be retracted

Immutable
- Prevent changes
- Original state is always preserved

CausalBench: Transparency

Zenodo

- Public Benchmark Runs are published to Zenodo
- All results and profiling information are recorded
- Digital Object Identifier (DOI) is assigned
- An immutable URL is generated
- Can be cited for future research

CausalBench: Explanation

- Rich corpus of benchmark data at our disposal
- What causal relationships can we extract from the benchmark runs?
- Potential questions:
  - How does CPU affect the execution time?
  - How does a hyperparameter affect the F1 score?

    ...
- Causal Explanation attempts to answer such questions

CausalBench: Explanation

- Causal graph from domain knowledge
- Causal effect estimation — estimate strengths of edges for a target
- Rank causes by the magnitude of their causal strength on the target

d $\rightarrow$ Dataset

m $\rightarrow$ Model

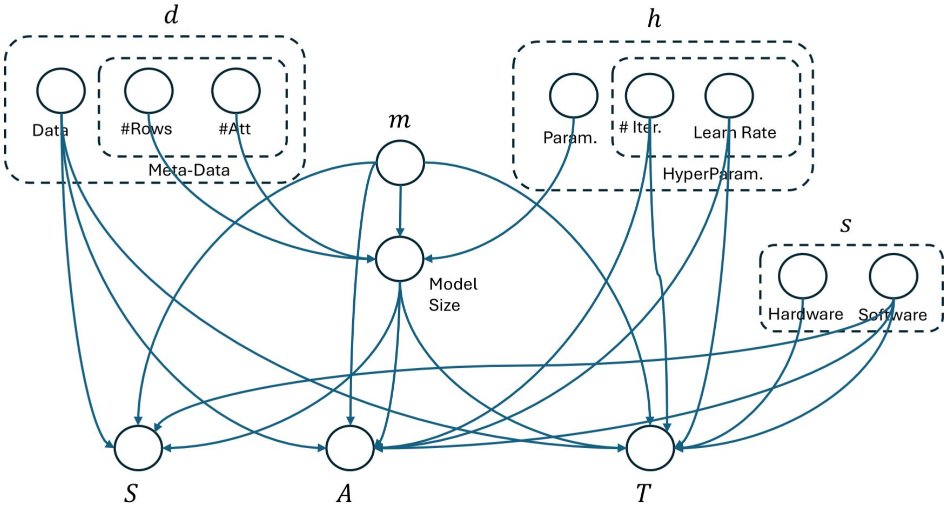
h $\rightarrow$ Hyperparameter

s $\rightarrow$ System configuration

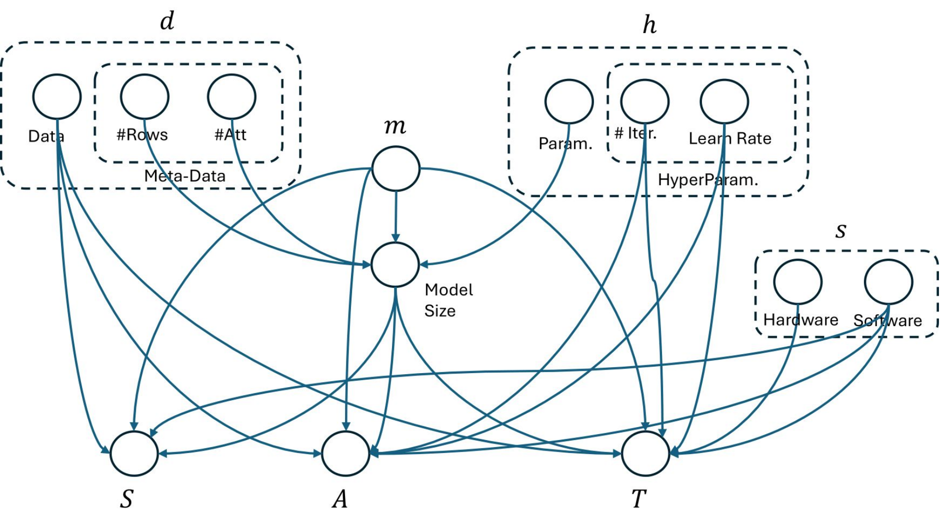
S $\rightarrow$ System profiling data

A $\rightarrow$ Accuracy metric

T $\rightarrow$ Execution time



CausalBench: Explanation



Benchmark Runs

Explanation

Run ID	Context Name	CPU Name	System Memory	GPU Name	GPU Memory	Run Published By
92	Benchmark (hyperparameters): PCMCIplus	13th Gen Intel(R) Core(TM) i7-13620H	46.67 GiB	NVIDIA GeForce RTX 4070 Laptop GPU	8.00 GiB	Ahmet Kapkic (akapkic@asu.edu)
90	Benchmark (hyperparameters): PCMCIplus	AMD EPYC 7763 64-Core Processor	250.29 GiB	None	None	Pratanu Mandal (pmandal5@asu.edu)
89	Benchmark (hyperparameters): PCMCIplus	Apple M3	16.00 GiB	None	None	Abhinav Gorantla (agorant2@asu.edu)
63	Benchmark (hyperparameters): PCMCIplus	12th Gen Intel(R) Core(TM) i9-12900KF	127.80 GiB	NVIDIA GeForce RTX 3090	24.00 GiB	Pratanu Mandal (pmandal5@asu.edu)

CausalBench: Causal Explanation Report

2025-08-04 01:53:46

Summary: Effects on Time.Duration (4000 experiments)

Variable	Effect	Strength
Model.ReadBytes	▲	679.4513
Model.GPUMemoryIdle	▼	-250.6410
Model.WriteBytes	▲	75.8604
Model.Memory	▲	28.4219
Model.GPUMemoryPeak	▲	8.5581
HP.max_conds_dim	▲	4.5975
HW.CPUSingleCore	▲	4.0273
HW.GPUScore	▲	4.0273
HW.StorageTotal	▲	4.0273
HW.CPUMultiCore	▲	4.0195
HW.MemoryTotal	▲	4.0156
HP.alpha	▲	1.7737

▲ This variable improves Time.Duration

▼ This variable worsens Time.Duration

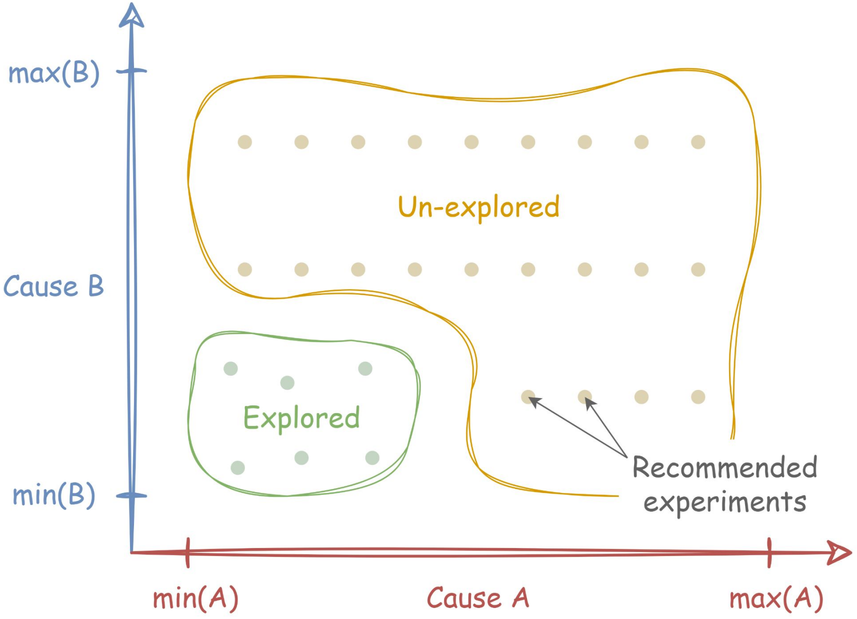
— This variable has no effect on Time.Duration

CausalBench: Recommendation

- Can we recommend more causally meaningful experiments?
- Causal Explanation provides key insights — exploit this!
- Key Idea:
    - Particular cause has more effect on a target
    - Further exploration of this cause might yield more granular insights
    - Recommend more granular experiments for this cause

CausalBench: Recommendation

- Causally informed space filling strategy
- Causes that have larger impacts on the target are more finely experimented
- Avoid new experiments that are close to existing benchmark runs



Effect of Cause A is greater than Cause B

CausalBench: Recommendation

CausalBench: Causal Explanation Report

2025-08-04 01:53:46

Summary: Effects on Time.Duration (4000 experiments)

Variable	Effect	Strength
Model.ReadBytes	▲	679.4513
Model.GPUMemoryIdle	▼	-250.6410
Model.WriteBytes	▲	75.8604
Model.Memory	▲	28.4219
Model.GPUMemoryPeak	▲	8.5581
HP.max_conds_dim	▲	4.5975
HW.CPUSingleCore	▲	4.0273
HW.GPUScore	▲	4.0273
HW.StorageTotal	▲	4.0273
HW.CPUMultiCore	▲	4.0195
HW.MemoryTotal	▲	4.0156

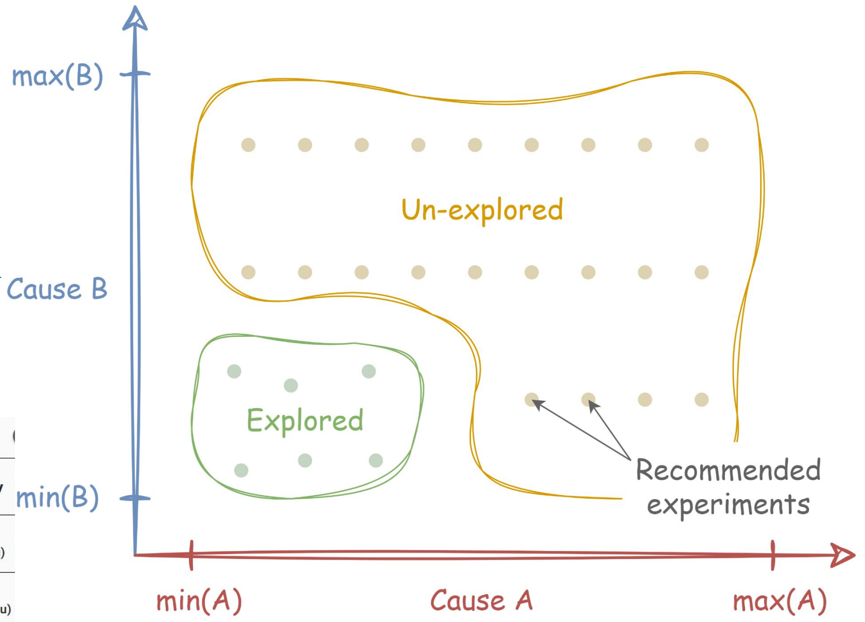
Benchmark Runs

Explanation

Run ID	Context Name	CPU Name	System Memory	GPU Name	GPU Memory	Run Published By
92	Benchmark (hyperparameters): PCMCIplus	13th Gen Intel(R) Core(TM) i7-13620H	46.67 GiB	NVIDIA GeForce RTX 4070 Laptop GPU	8.00 GiB	Ahmet Kapkic (akapkic@asu.edu)
90	Benchmark (hyperparameters): PCMCIplus	AMD EPYC 7763 64-Core Processor	250.29 GiB	None	None	Pratanu Mandal (pmandal5@asu.edu)
89	Benchmark (hyperparameters): PCMCIplus	Apple M3	16.00 GiB	None	None	Abhinav Gorantla (agorant2@asu.edu)
63	Benchmark (hyperparameters): PCMCIplus	12th Gen Intel(R) Core(TM) i9-12900KF	127.80 GiB	NVIDIA GeForce RTX 3090	24.00 GiB	Pratanu Mandal (pmandal5@asu.edu)

Recommendations:

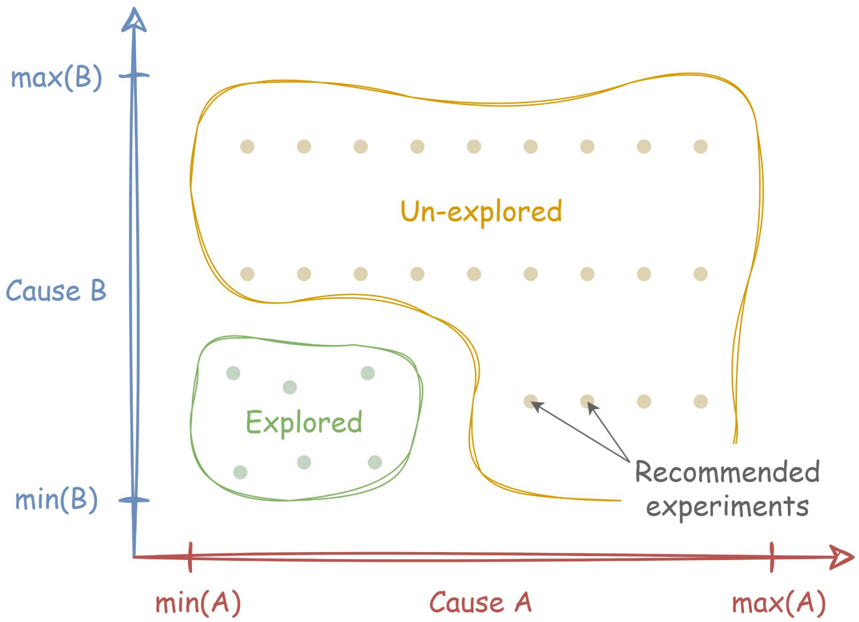
Additional Hyperparameter settings to consider for your experiments:
[HP.max_conds_dim, HP.alpha]: [(1.0, 0.355), (13.25, 0.355), (25.5, 0.355), (37.75, 0.355), (5
0.355)]



CausalBench: Recommendation - Hands On

Let **CausalBench** to explain and recommend some benchmarks for you!

- ***Filter by***
  - Context ID: 10
  - Context Version: 2
- ***Explanation***
  - By: Time Elapsed
  - PCMCIPlus:
    - alpha.Min: 0.01
    - alpha.Max: 0.8
    - max_conds_dim.min: 1
    - max_conds_dim.min: 30
- ***Analyze!***



End of Deck 4

Any Questions?

CausalBench: What's Next?

Going public

- Open Source
- Workshops
- Creating a research community

…

Conclusions

- Benchmarking: Problems
- Causality
- CausalBench
- Benchmarking:
  - Analyzing
  - Improving

Thank you!

Any Questions?

CausalBench	KDD Tutorial	Docs/Github
(Application)	(Usage)	(Contribution)

Further questions? Feedbacks? Want to use **CausalBench**?

support@causalbench.org