6. DAGs

ISS5096 || ECI

Jaewon ("Jay-one") Yoo
National Tsing Hua University

Where are we? Where are we going?

Up to now: Saw how to estimate the ATE with regression under selection on observables.

Now:

Where are we? Where are we going?

Up to now: Saw how to estimate the ATE with regression under selection on observables.

Now:

- How do we know if no unmeasured confounders holds? (i.e., What covariates do we need to condition on?)
    - One way, from the assumption itself: $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp D_i \mid \mathbf{X}_i$
        - Include covariates such that, conditional on them, the treatment assignment does not depend on the potential outcomes.
        - Somewhat circular
    - Another way: use DAGs and look at back-door paths.

Where are we? Where are we going?

Up to now: Saw how to estimate the ATE with regression under selection on observables.

Now:

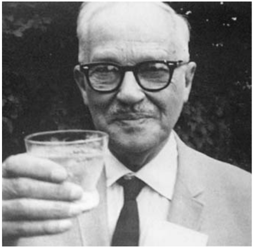
- How do we know if no unmeasured confounders holds? (i.e., What covariates do we need to condition on?)
  - One way, from the assumption itself: $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp D_i \mid \mathbf{X}_i$
    - Include covariates such that, conditional on them, the treatment assignment does not depend on the potential outcomes.
    - Somewhat circular
  - Another way: use DAGs and look at back-door paths.
- What if this assumption doesn't hold?
  - Sensitivity analysis: try to vary the amount of unmeasured confounding to see if it changes the effect.

Sources: https://www.redbubble.com/i/sticker/D-Ya-Like-Dags-by-salamincheese/27407958.EJUG5

p.s.: from an old British-American comedy film starring Brad Pitt

1/ DAGs

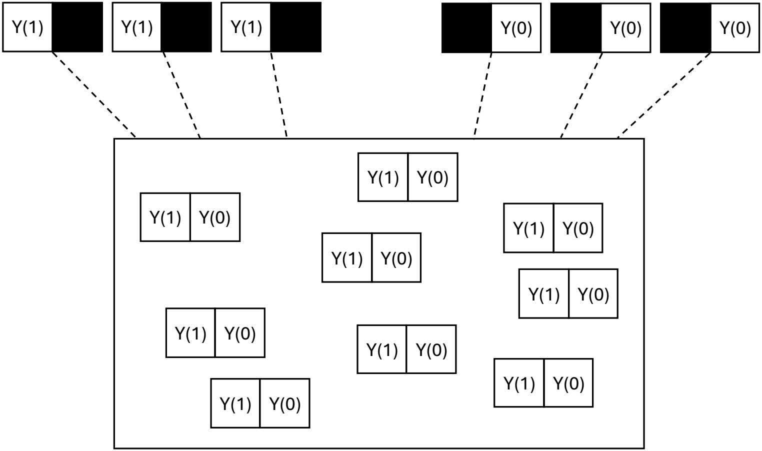
Neyman-Rubin Potential Outcomes Model



Jerzy Neyman (1894–1981)

Donald Rubin (1943–)

Neyman Urn Model



- Neyman-Rubin Potential Outcomes Model
- Defines causal effects as contrasts of POs! (Week 1)

Rubin CM: Average Treatment Effect (ATE)

Suppose we observe a population of 4 units:

i	D_i	Y_i	$Y_i(1)$	$Y_i(0)$	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_i(1)]$			1.5		
$\mathbb{E}[Y_i(0)]$				0.5	
$\mathbb{E}[Y_i(1) - Y_i(0)]$					1

- $\tau_{\text{ATE}} = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[\tau_i] = \frac{3+0+1+0}{4} = 1.$
- Why $\tau_{\text{ATE}} \neq \widehat{\tau}_{\text{diff}}$? When would they be equal?

Rubin CM: ATE on the Treated (ATT)

Again suppose we observe a population of 4 units:

i	D_i	Y_i	$Y_i(1)$	$Y_i(0)$	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_i(1) \vert D_i = 1]$			2		
$\mathbb{E}[Y_i(0) \vert D_i = 1]$				0.5	
$\mathbb{E}[Y_i(1) - Y_i(0) \vert D_i = 1]$					1.5

- $\tau_{\text{ATT}} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1] = \mathbb{E}[\tau_i \mid D_i = 1] = \frac{3+0}{2} = 1.5.$
- Why does $\tau_{\text{ATT}} \neq \tau_{\text{ATE}}$?

An Alternative Causal Model: Causal Graphs

- Did social scientists not do causal inference before Rubin? **No!**

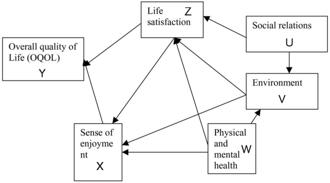
An Alternative Causal Model: Causal Graphs

- Did social scientists not do causal inference before Rubin? **No!**
- The old paradigm: structural equation modeling and path analysis

An Alternative Causal Model: Causal Graphs

- Did social scientists not do causal inference before Rubin? **No!**
- The old paradigm: structural equation modeling and path analysis

1. Postulate a causal mechanism and draw a corresponding path diagram



2. Translate it into a (typically linear) system of equations:

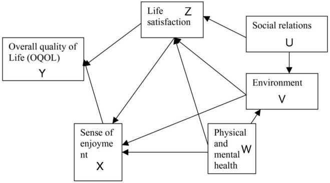
$$
\begin{aligned}
Y &= \alpha_0 + \alpha_1 X + \alpha_2 Z + \varepsilon_\alpha \\
X &= \beta_0 + \beta_1 Z + \beta_2 W + \beta_3 V + \varepsilon_\beta \quad \cdots
\end{aligned}
$$

3. Estimate α, β, etc. typically assuming normality and exogeneity

An Alternative Causal Model: Causal Graphs

- Did social scientists not do causal inference before Rubin? **No!**
- The old paradigm: structural equation modeling and path analysis

1. Postulate a causal mechanism and draw a corresponding path diagram



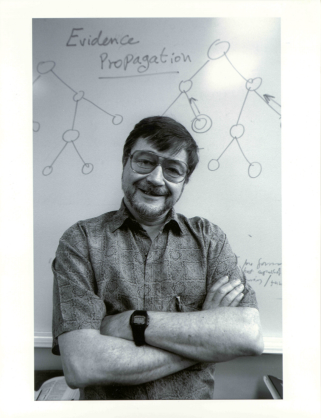
2. Translate it into a (typically linear) system of equations:

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 Z + \varepsilon_\alpha$$
$$X = \beta_0 + \beta_1 Z + \beta_2 W + \beta_3 V + \varepsilon_\beta \quad \cdots$$

3. Estimate α, β, etc. typically assuming normality and exogeneity

- Went out of fashion (until... Pearl's attack!):
  - Strong distributional/functional form assumptions
  - No language to distinguish causation from association

Pearl's Attack



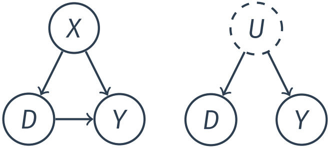
Judea Pearl (1936–) proposed a new causal inference framework based on nonparametric structural equation modeling (NPSEM)

- Originally a computer scientist
- Previous important work on artificial intelligence
- *Causality* (2000, Cambridge UP)
- Won the Turing Award in 2011 for his causal work

Pearl's framework builds on SEMs and revives it as a formal language of causality.

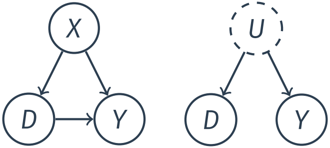
Directed Acyclic Graphs

- **Directed acyclic graphs** (DAGs) describe the causal structure of variables



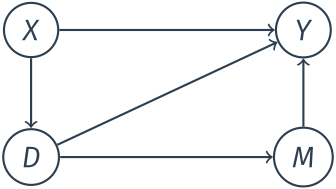
Directed Acyclic Graphs

- **Directed acyclic graphs** (DAGs) describe the causal structure of variables



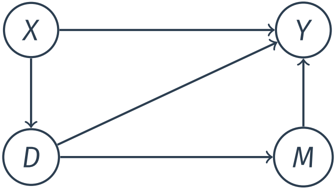
- **Nodes/vertices**: observed (solid) or unobserved (dashed) variables.
- **Edges**: arrows that encodes the presence or absence of a causal effect.
    - Arrow present = a direct causal effect: $Y_i(d) \neq Y_i(d')$ for some i and d.
    - Lack of an arrow = no causal effect: $Y_i(d) = Y_i(d')$ for all i and d.
    - Missing variables = no other common causes of any variables.
- **Directed**: each arrow implies a direction (causal ordering).
- **Acyclic**: no cycle: a variable cannot cause itself

DAG Terminologies



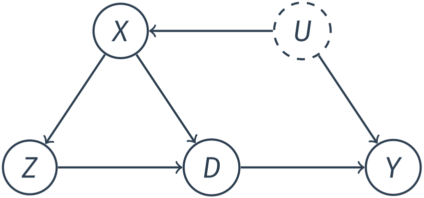
- **Path**: a sequence of edges that connect two nodes.
  - A **directed** or **causal** path is all in the same causal direction.
  - Non-causal path example: $D \leftarrow X \rightarrow Y$

DAG Terminologies



- **Path**: a sequence of edges that connect two nodes.
  - A **directed** or **causal** path is all in the same causal direction.
  - Non-causal path example: $D \leftarrow X \rightarrow Y$
- **Descendants**: nodes on a directed path away from some other node.
  - M is a descendant of D and X.
  - Ancestors is the reverse: X is an ancestor of M.
- **Parents**: immediate causes of a node.
  - D is the parent of Y and M.
  - **Children** are the reverse: M is a child of D.

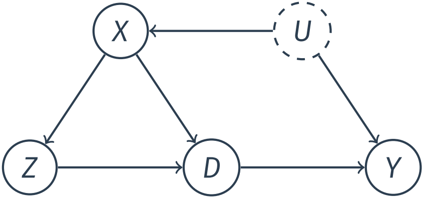
DAGs to Distributions



$$Y = f_y(D, U, \varepsilon_y)$$
$$D = f_d(Z, X, \varepsilon_d)$$
$$X = f_x(U, \varepsilon_x)$$
$$Z = f_z(X, \varepsilon_z)$$

- Causal DAGs equivalent to **nonparametric structural equation models**

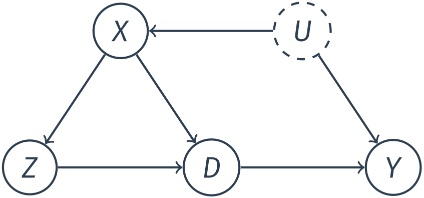
DAGs to Distributions



$$Y = f_y(D, U, \varepsilon_y)$$
$$D = f_d(Z, X, \varepsilon_d)$$
$$X = f_x(U, \varepsilon_x)$$
$$Z = f_z(X, \varepsilon_z)$$

- Causal DAGs equivalent to **nonparametric structural equation models**
- NPSEM have a **causal interpretation**, but are completely flexible.
  - No specification of a functional form or interactions, etc.
  - More standard linear SEM is a special case.

DAGs to Distributions

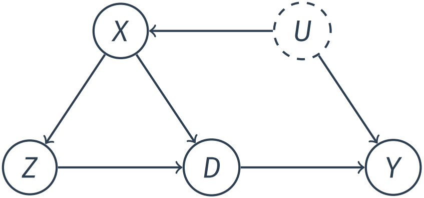


$$Y = f_y(D, U, \varepsilon_y)$$
$$D = f_d(Z, X, \varepsilon_d)$$
$$X = f_x(U, \varepsilon_x)$$
$$Z = f_z(X, \varepsilon_z)$$

- Causal DAGs equivalent to **nonparametric structural equation models**
- NPSEM have a **causal interpretation**, but are completely flexible.
  - No specification of a functional form or interactions, etc.
  - More standard linear SEM is a special case.
- Causal DAGs imply the following factorization:

$$\mathbb{P}(X_1, X_2, \ldots, X_J) = \prod_{j=1}^{J} \mathbb{P}(X_j \mid \text{pa}(X_j)) \quad \text{where} \quad \text{pa}(X_j) = \text{parents of } X_j$$

DAGs to Distributions



$$Y = f_y(D, U, \varepsilon_y)$$
$$D = f_d(Z, X, \varepsilon_d)$$
$$X = f_x(U, \varepsilon_x)$$
$$Z = f_z(X, \varepsilon_z)$$

- Causal DAGs equivalent to **nonparametric structural equation models**
- NPSEM have a **causal interpretation**, but are completely flexible.
  - No specification of a functional form or interactions, etc.
  - More standard linear SEM is a special case.
- Causal DAGs imply the following factorization:

$$\mathbb{P}(X_1, X_2, \ldots, X_J) = \prod_{j=1}^{J} \mathbb{P}(X_j \mid \mathrm{pa}(X_j)) \quad \text{where} \quad \mathrm{pa}(X_j) = \text{parents of } X_j$$

$$\mathbb{P}(D, U, X, Y, Z) = \mathbb{P}(Y|D, U)\mathbb{P}(D|X, Z)\mathbb{P}(Z|X)\mathbb{P}(X|U)\mathbb{P}(U)$$

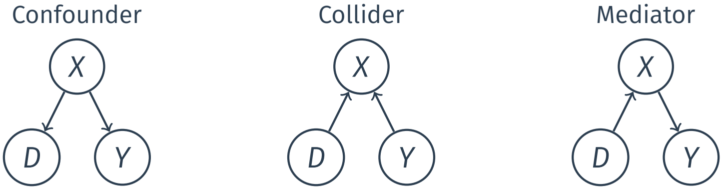
D-Separation

- Can we determine conditional independence from our causal DAG?
- Yes! To verify that $A \perp\!\!\!\perp B \mid C$ where each is a set of nodes:
    1. Find all paths from any vertex in A to any vertex in B.
    2. Check if each path is **blocked**.
    3. If all paths are blocked, then A is **d-separated** from B by C.

D-Separation

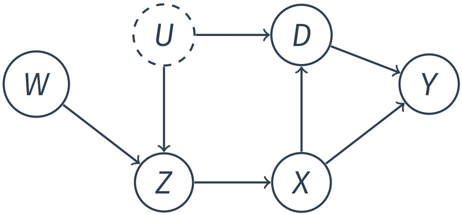
- Can we determine conditional independence from our causal DAG?
- Yes! To verify that $A \perp\!\!\!\perp B \mid C$ where each is a set of nodes:
    1. Find all paths from any vertex in A to any vertex in B.
    2. Check if each path is **blocked**.
    3. If all paths are blocked, then A is **d-separated** from B by C.
- A path is **blocked** conditional on C if:
    1. C includes a non-collider on that path **OR**
    2. Path includes a collider not in C and no descendant of any collider is in C.
- If A and B are d-separated, then we have $A \perp\!\!\!\perp B \mid C \rightsquigarrow$ if not, then d-connected.

Common Structures



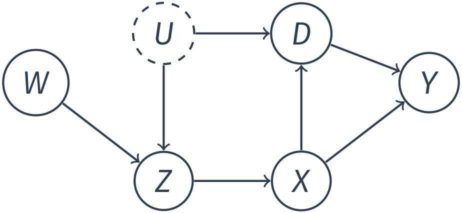
- **Confounder** (fork): Common cause of two variables.
  - D and Y unconditionally dependent, conditionally independent.
- **Collider** (inverted fork): common descendent of two variables.
  - D and Y unconditionally independent, conditionally dependent.
  - X “blocks” the relationship between them when not conditioned on.
  - Example: D (education); Y (experience); X (hired).
    - Now, X=1 with low D must mean high Y, vice versa.
- **Mediator** (chain): variable on the path from one variable to another.
  - D and Y unconditionally dependent.

D-Separation Example



- Are W and Y marginally independent of each other?
  - Blocked: $W \to Z \leftarrow U \to D \to Y \quad W \to Z \leftarrow U \to D \leftarrow X \to Y$
  - Unblocked: $W \to Z \to X \to Y \quad W \to Z \to X \to D \to Y$

D-Separation Example



- Are W and Y marginally independent of each other?
  - Blocked: $W \rightarrow Z \leftarrow U \rightarrow D \rightarrow Y \quad W \rightarrow Z \leftarrow U \rightarrow D \leftarrow X \rightarrow Y$
  - Unblocked: $W \rightarrow Z \rightarrow X \rightarrow Y \quad W \rightarrow Z \rightarrow X \rightarrow D \rightarrow Y$

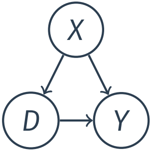
- Which variables should we condition on to make W and Y conditionally independent (d-separated)?
  - Block the unblocked paths without unblocking the blocked paths.
  - Conditioning on X would do this.
  - Conditioning on D and/or Z would unblock some of the blocked paths because they are colliders.

Backdoor Paths and Blocking Paths

- **Backdoor paths**: non-causal path from D to Y.
    - Would remain if we removed any arrows pointing out of D.

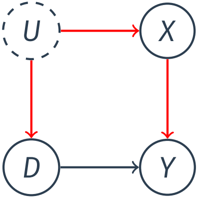
Backdoor Paths and Blocking Paths

- **Backdoor paths**: non-causal path from D to Y.
  - Would remain if we removed any arrows pointing out of D.
- Backdoor paths between D and Y $\rightsquigarrow$ common causes of D and Y:



- Here: backdoor path $D \leftarrow X \rightarrow Y$

Other Types of Confounding



- D: enrolling in a job training program.
- Y: getting a job.
- U: being motivated.
- X: number of job applications sent out.
- Big assumption here $\rightsquigarrow$ no arrow linking U to Y

Backdoor Criterion

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i$$

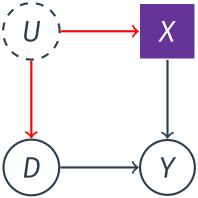
- Can we use a DAG to evaluate no unmeasured confounders?
- Holds if the **backdoor criterion** on a causal DAG is met:
  1. No vertex/node in $\mathbf{X}$ is a descend of D (**no post-treatment bias**), and
  2. $\mathbf{X}$ blocks all backdoor paths from D to Y.

Backdoor Criterion

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i \mid \mathbf{X}_i$$

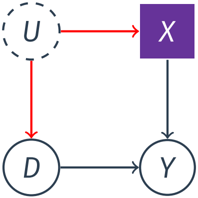
- Can we use a DAG to evaluate no unmeasured confounders?
- Holds if the **backdoor criterion** on a causal DAG is met:
  1. No vertex/node in $\mathbf{X}$ is a descend of D (**no post-treatment bias**), and
  2. $\mathbf{X}$ blocks all backdoor paths from D to Y.
- The backdoor criterion is pretty powerful. Tells us:
  1. If there is confounding given this DAG,
  2. If it is possible to remove the confounding, and
  3. What variables to condition on to eliminate the confounding.

Other Types of Confounding *Redux*



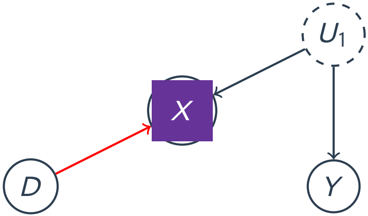
- D: enrolling in a job training program.
- Y: landing a job!
- U: being motivated.
- X: number of job applications sent out.
- Big assumption here $\rightsquigarrow$ no arrow linking U to Y

Other Types of Confounding *Redux*



- D: enrolling in a job training program.
- Y: landing a job!
- U: being motivated.
- X: number of job applications sent out.
- Big assumption here $\rightsquigarrow$ no arrow linking U to Y
- Conditioning on X blocks all backdoor paths.

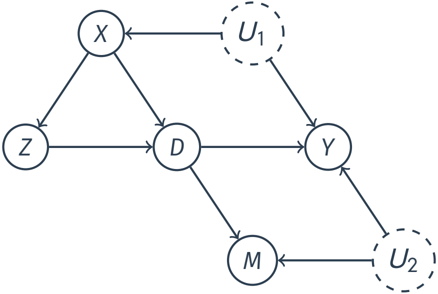
Why Not Condition on Descendents?



- No causal or statistical relationship between D and Y
- Conditioning on post-treatment variables opens up non-causal paths
  - $\rightsquigarrow$ Statistical relationship between D and Y conditional on X
  - But still no causal relationship $\rightsquigarrow$ selection bias.

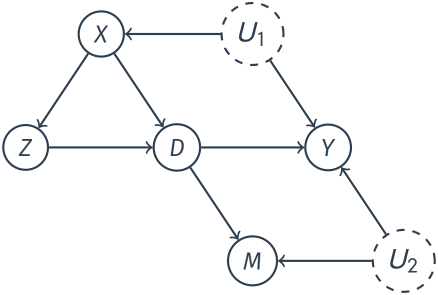
Usage Example of DAG

Assume that you've come out with a DAG based on your expertise:



Usage Example of DAG

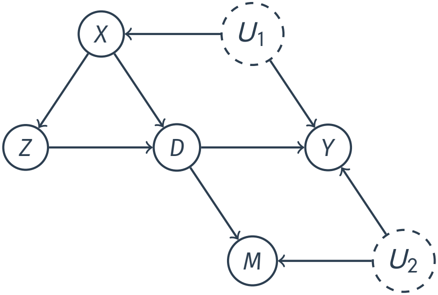
Assume that you've come out with a DAG based on your expertise:



- Suppose you want to identify a causal effect of D on Y.
    - In a nutshell, you want to block all the paths that yield statistical associations between D and Y.
    - Thus, you want to find a set of nodes **S** such that once we condition on **S**
        - **no unmeasured confounding** holds and
        - any descendant of D is not in S $\longrightarrow$ no post-treatment bias.

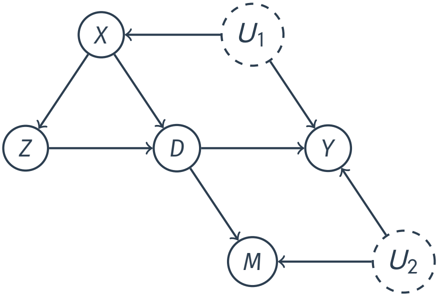
Usage Example of DAG

Assume that you've come out with a DAG based on your expertise:



- Suppose you want to identify a causal effect of D on Y.
  - In a nutshell, you want to block all the paths that yield statistical associations between D and Y.
  - Thus, you want to find a set of nodes **S** such that once we condition on **S**
    - **no unmeasured confounding** holds and
    - any descendant of D is not in S $\longrightarrow$ no post-treatment bias.
- $\rightsquigarrow$ Use backdoor criterion!

Back Door Criterion Example



1. List all of the **backdoor paths** between D and Y.
2. List all the possible set of nodes **S** that you can condition on.
3. List all the **S** such that **blocks** all the backdoor paths.
4. Among those **S**, drop the sets which include a descend of D.

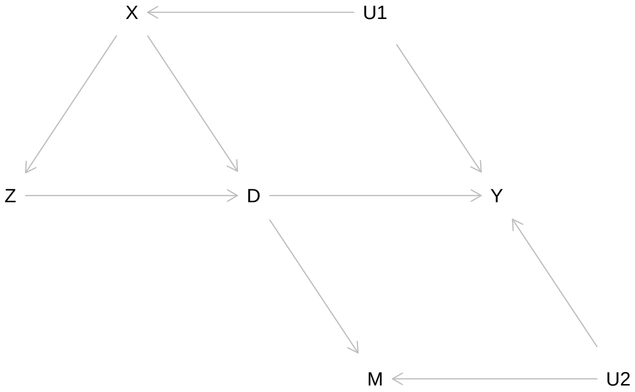
R Package: dagitty

- DAGitty: www.dagitty.net

```
> library(dagitty)
> g <- dagitty('dag {
      X [pos="1,-1.5"]
      Y [pos="4,0"]
      Z [pos="0,0"]
      M [pos="3,1.5"]
      D [pos="2,0"]
      U1 [pos="3,-1.5"]
      U2 [pos="5,1.5"]
      X -> Z -> D -> Y
      X -> D -> M
      M <- U2 -> Y
      X <- U1 -> Y
  }')
> latents(g) <- c("U1", "U2")
```

R Package: `dagitty`

```
> plot(g) # Visualize the DAG
```



R Package: dagitty

- Access parent/ancestor nodes:

```
> parents(g, "D") ## [1] "X" "Z"
[1] "X" "Z"

> ancestors(g, "D") ## [1] "D" "Z" "X" "U1"
[1] "D"  "Z"  "X"  "U1"
```

- Or children/descendent nodes:

```
> children(g, "D") ## [1] "M" "Y"
[1] "M" "Y"

> descendants(g, "D") ## [1] "D" "Y" "M"
[1] "D" "Y" "M"
```

R Package: dagitty

- Identify paths using paths():

```
> paths(g, "D", "Y")$paths
[1] "D -> M <- U2 -> Y"      "D -> Y"                   "D <- X <- U1 -> Y"
[4] "D <- Z <- X <- U1 -> Y"
```

- Extract causal path(s) by setting directed = T:

```
> paths(g, "D", "Y", directed = TRUE)$paths # only causal path(s)
[1] "D -> Y"
```

- Check whether two nodes are d-separated using dseparated():

```
> dseparated(g, "Z", "D", c("X")) # because of Z -> D
[1] FALSE

> dseparated(g, "Z", "M", c("D"))
[1] TRUE
```

R Package: `dagitty`

- Access the set of nodes **S** to condition on for **no unmeasured confounding** to hold:

```
> adjustmentSets(g, "D", "Y", type="minimal")
{ X }
```

- Caveat: `adjustmentSets` may include unobserved variables which we cannot actually condition on.

```
> S = adjustmentSets(g, "D", "Y", type="all")

> S[!grepl("U1|U2", S)]
{ X }      # X typically preferred for simplicity and max statistical power, unless
{ X, Z }   # Z is a known strong confounder and you have sufficient data.
```

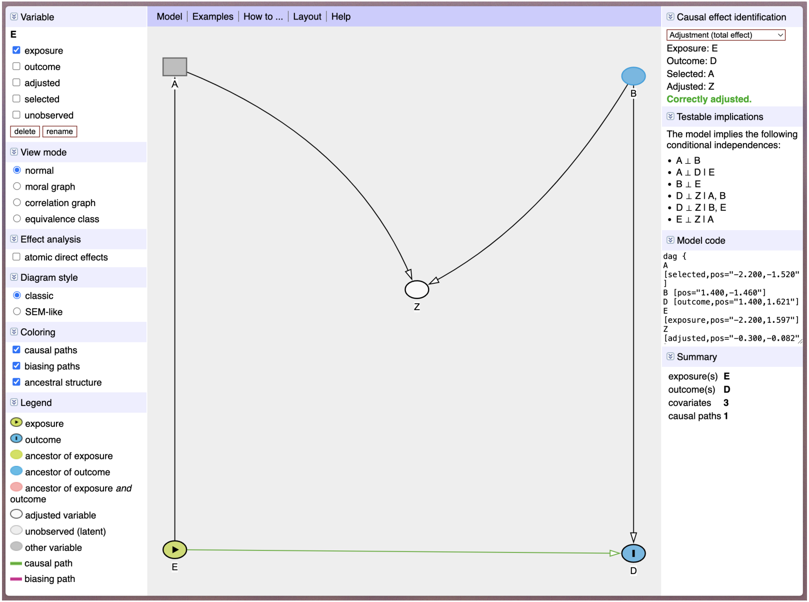
- Note that this implements a slightly more general criterion (sometimes it may contain descendants)

R Package: dagitty

- Full list of adjustment sets:

```
# S = adjustmentSets(g, "D", "Y", type="all")
> S
{ U1 }
{ U1, U2 }
{ M, U1, U2 }
{ X }
{ U1, X }
{ U2, X }
{ M, U2, X }
{ U1, U2, X }
{ M, U1, U2, X }
{ U1, Z }
{ U1, U2, Z }
{ M, U1, U2, Z }
{ X, Z }
{ U1, X, Z }
{ U2, X, Z }
{ M, U2, X, Z }
{ U1, U2, X, Z }
{ M, U1, U2, X, Z }
```

DAGitty (https://www.dagitty.net/)



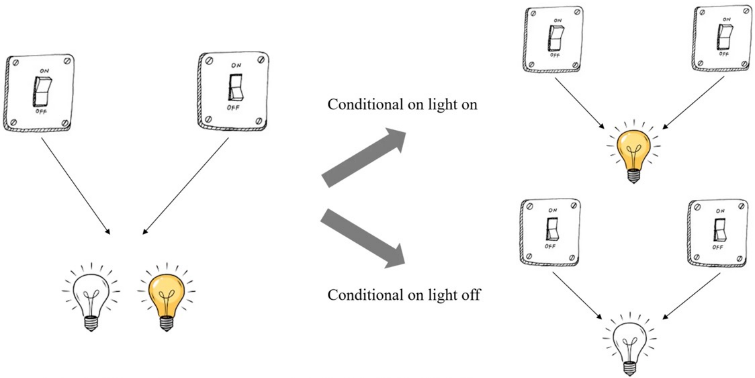
Onto the presentations & discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

Appendix

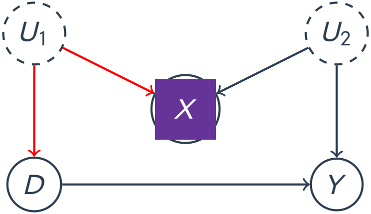
Conditioning on Collider

- Two independent switches that turn the light on/off:



Sources: Jiyong Park (UNC at Greensboro), Korea Summer Session on Causal Inference 2021

M-Bias



- Not all backdoor paths induce confounding:
- No conditioning: backdoor path blocked by the collider X.
- If we control for $X \rightsquigarrow$ this "opens up" the path and induces confounding!
  - Sometimes referred to as **M-bias** or **collider bias**.
- Controversial because of differing views on what to control for:
  - Rubin thinks that M-bias is a "mathematical curiosity" and we should control for all pretreatment variables.
  - Pearl and others think M-bias is a real threat.