Agenda

Covariates for causal analysis

1. Administrative
2. Interacting variables in regression
3. Causal analysis in regression
4. Mediation, moderation, confounding, and collision
5. ***Hands on***: *Building indicator variables in R*

Slides are licensed under CC BY-NC-SA 4.0
Uploading to generative AI tools or use for training machine learning models is prohibited

Updated worksheet dates

- First worksheet due *Tuesday*, Sept 22 by midnight
- Peer assessments due that Friday, Sept 25

Predicting income: *interactions*

DUMMY VARIABLES

$$\log(Inc_i) \sim \text{Norm}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_1 M_i + \beta_2 A_i$$
$$\alpha, \beta_1, \beta_2 \sim \text{Norm}(0, 30)$$
$$\sigma \sim \text{Unif}(0, 50)$$

M_i
Indicator variable for male

A_i
Indicator variable for respondents at least 35 years old

	Mean	Std. dev	2.5%	97.5%
α	10.16	0.03	10.10	10.21
β_1	0.13	0.03	0.07	0.19
β_2	0.59	0.03	0.54	0.65
σ	0.81	0.01	0.79	0.83

β_1: $\exp(0.13) \approx 1.14$
(men make about 14% more than women, on average)

β_2: $\exp(0.59) \approx 1.81$
(people at least 35 years old make about 81% more than those under 35, on average)

INTERACTING DUMMIES

$$\log(Inc_i) \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_1 M_i + \beta_2 A_i + \beta_3 M_i A_i$$

$$\alpha, \beta_1, \beta_2 \sim \text{Norm}(0, 30)$$

$$\sigma \sim \text{Unif}(0, 50)$$

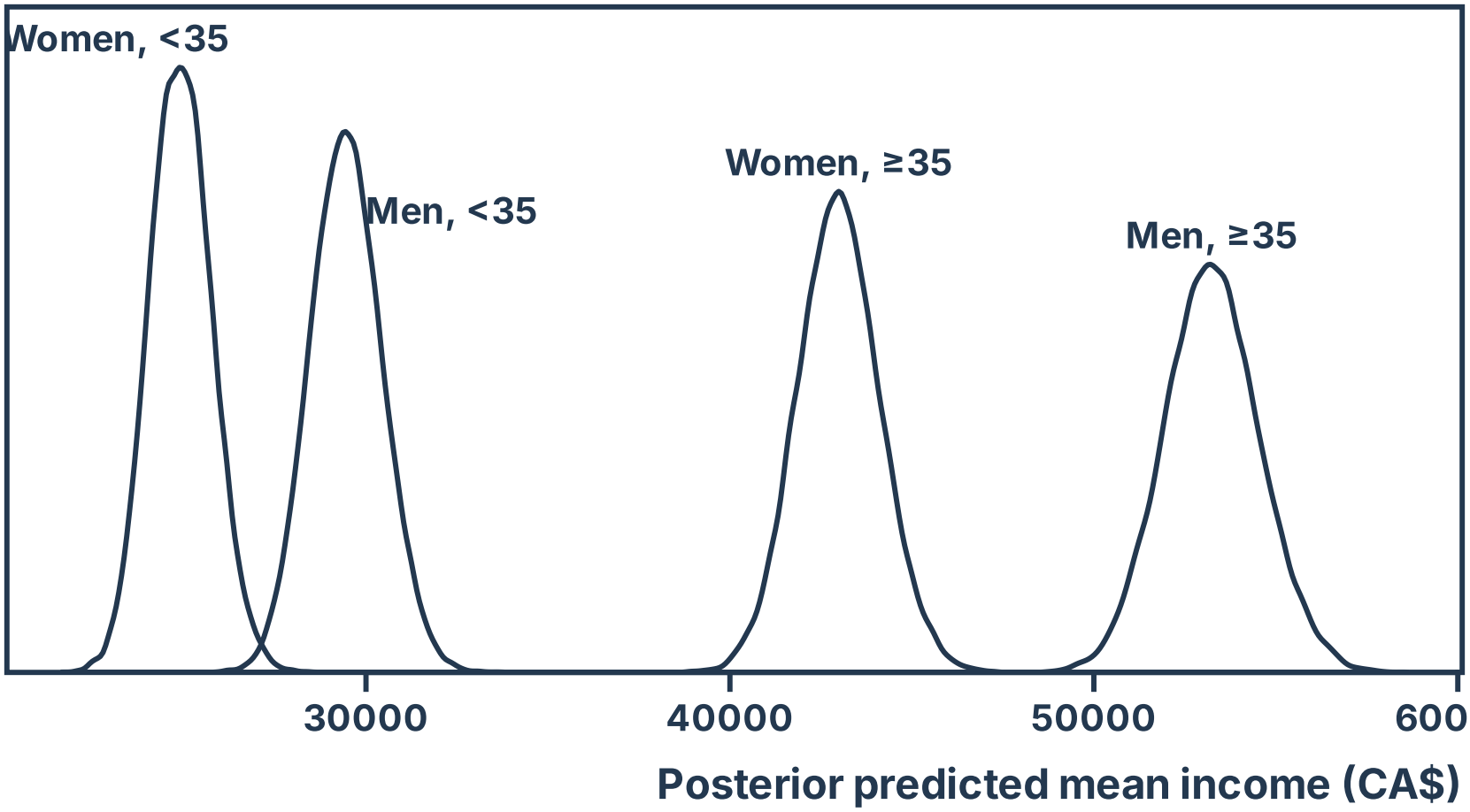
$M_i A_i$
Interaction between both indicators

	Mean	Std. dev	2.5%	97.5%
α	10.20	0.03	10.13	10.27
β_1	0.05	0.05	-0.04	0.14
β_2	0.52	0.04	0.44	0.61
β_3	0.13	0.06	0.02	0.25
σ	0.81	0.01	0.79	0.83

INTERACTING DUMMIES

$$\mu_i = \alpha + \beta_1 M_i + \beta_2 A_i + \beta_3 M_i A_i$$

$$\mu_{(<35,female)} = \alpha$$
$$\mu_{(<35,male)} = \alpha + \beta_1$$
$$\mu_{(\geq 35,female)} = \alpha + \beta_2$$
$$\mu_{(\geq 35,male)} = \alpha + \beta_1 + \beta_2 + \beta_3$$



INTERACTING DUMMIES

$$\mu_i = \alpha + \beta_1 M_i + \beta_2 A_i + \beta_3 M_i A_i$$

$$\begin{aligned}
\mu_{(<35, female)} &= \alpha \\
\mu_{(<35, male)} &= \alpha + \beta_1 \\
\mu_{(\geq 35, female)} &= \alpha + \beta_2 \\
\mu_{(\geq 35, male)} &= \alpha + \beta_1 + \beta_2 + \beta_3
\end{aligned}$$

	Mean	Exp(Mean)
α	10.20	26,853
β_1	0.05	1.05
β_2	0.52	1.69
β_3	0.13	1.14

Interpreting the interaction coefficient β_3

The pay boost for being over 35 years old (β_2) is about 14% greater for men than for women.

OR

The income advantage for men over women (β_1) is about 14% greater for respondents over 35 years old.

INTERACTING CONTINUOUS VARIABLES

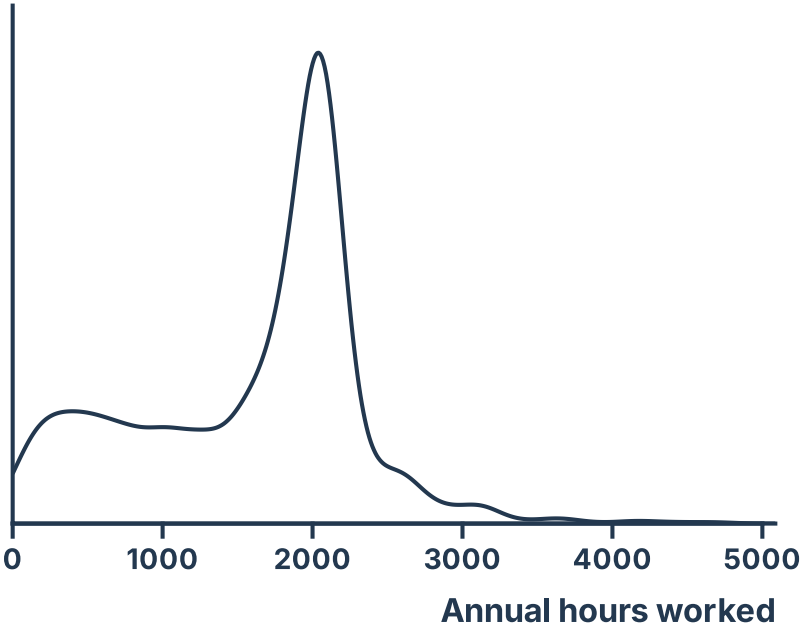
$$\log(Inc_i) \sim \text{Norm}(\mu_i, \sigma)$$

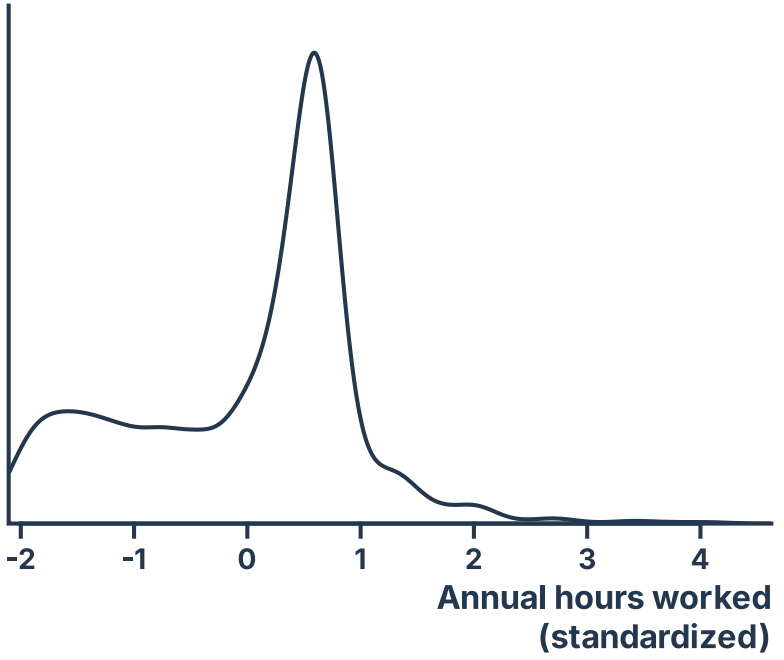
$$\mu_i = \alpha + \beta_1 Hours_i + \beta_2 Dues_i + \beta_3 Hours_i Dues_i$$

Hours worked (standardized)

Union and prof. assoc. dues (standardized)

Standardization: Transforming a variable X so that $\text{mean}(X) = 0$ and $\text{sd}(X) = 1$





$$\mu_i = \alpha + \beta_1 Hours_i + \beta_2 Dues_i + \beta_3 Hours_i Dues_i$$

	Mean	Exp(Mean)
α	10.58	39,549
β_1	0.40	1.49
β_2	0.20	1.22
β_3	-0.07	0.93

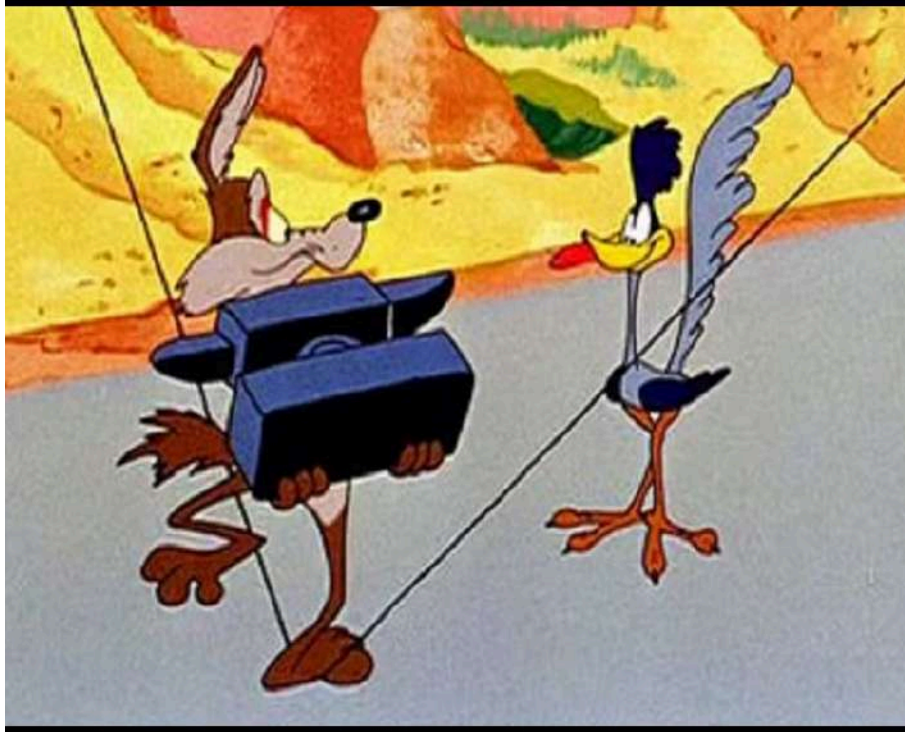
Interpreting the interaction coefficient β_3

The extra income associated with increased union and professional dues (β_2) is reduced by about 7% for every standard-deviation increase in hours worked.

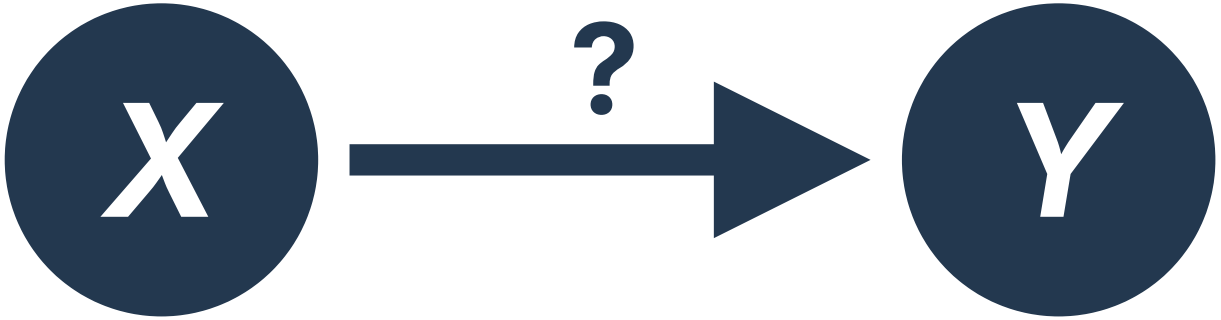
OR

The extra money made by working more hours (β_1) is reduced by about 7% for every standard-deviation increase in union dues paid.

Identifying cause & effect



CAUSAL ANALYSIS

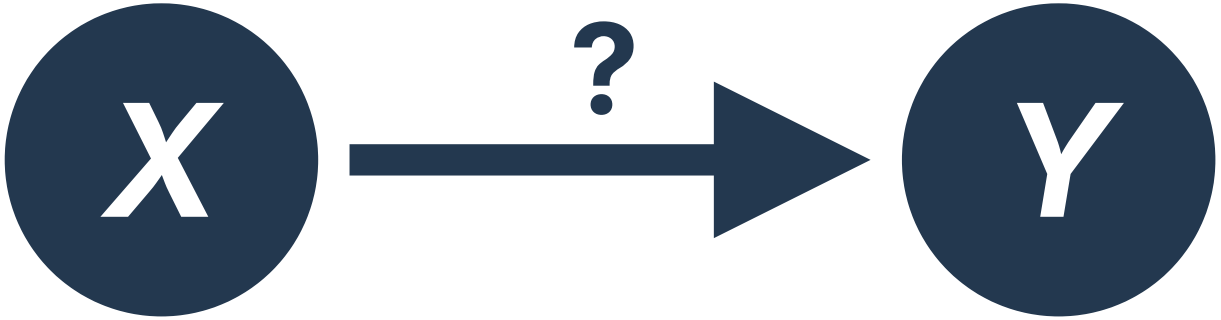


Causal question: Does a change in one variable (*X*) ***cause*** a change in another (*Y*)?

Regression only identifies statistical relationships, not causal relationships

To draw a "causal arrow" you need ***theory***

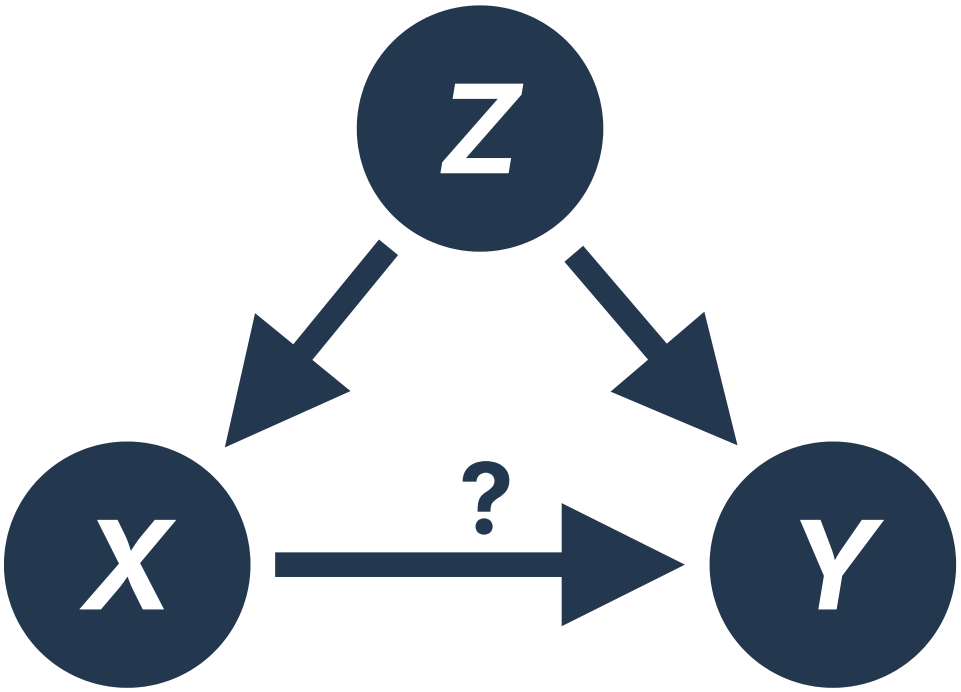
CAUSAL ANALYSIS



To establish a causal relationship you (usually) need

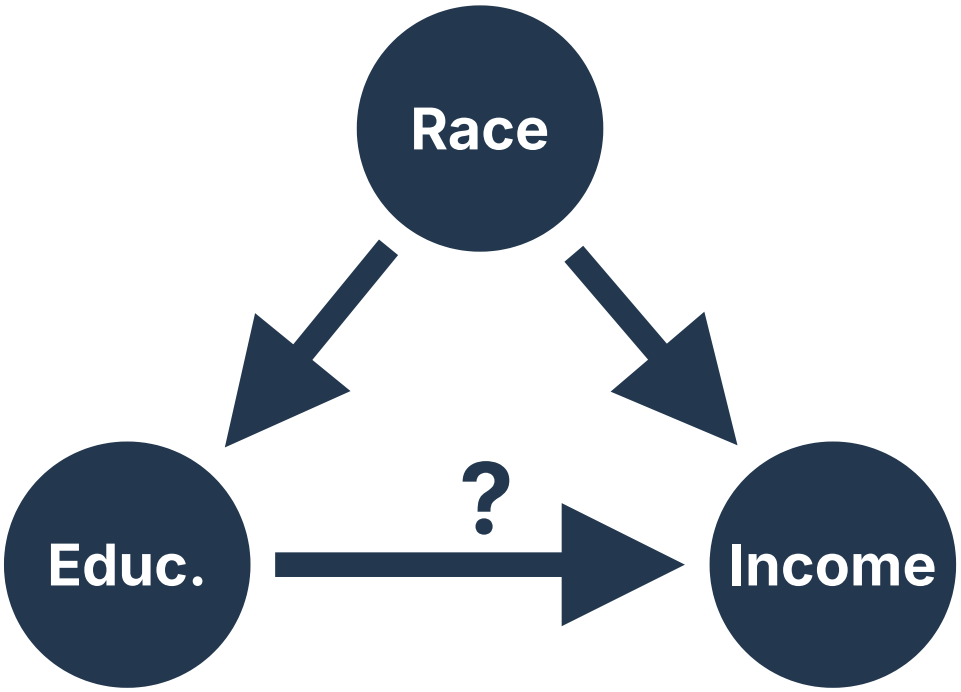
1. **Causal precedence**
   A theoretical reason to believe that changes in *X* *could* affect *Y* (e.g. precedes *Y* in time)
2. **Statistical association**
   An established statistical association between *X* and *Y* (e.g. a convincing coefficient estimate)
3. **No unaccounted-for confounders**
   No other variables, observed or otherwise, that *confound* the association between *X* and *Y*

CONFOUNDING VARIABLES



A variable *Z* is a ***confounder*** of the relationship between *X* and *Y* if *Z* is a causal influence on both *X* and *Y*

CONFOUNDING VARIABLES

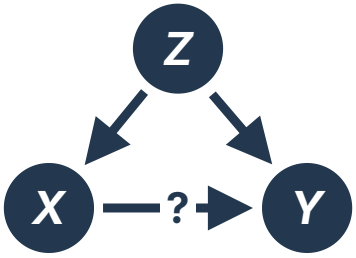


A variable *Z* is a ***confounder*** of the relationship between *X* and *Y* if *Z* is a causal influence on both *X* and *Y*

For example:
To establish a causal relationship between education and income, you need to account for race, which could affect both education and income

TYPES OF COVARIATES

Confounder

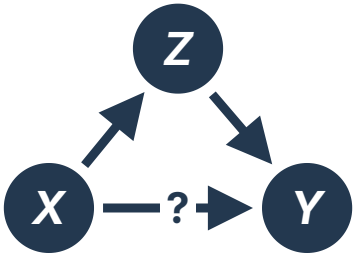


Z is a causal factor on both *X* and *Y*.

Must be "controlled for" to establish non-spurious relationship between *X* and *Y*.

E.g.:
Race confounds the relationship between education and income.

Mediator

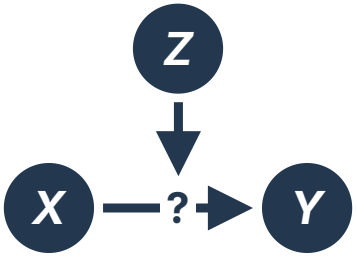


Z is influenced by *X* and influences *Y*.

Including as covariate elaborates on relationship between *X* and *Y*.

E.g.:
Occupation mediates the relationship between gender and income.

Moderator

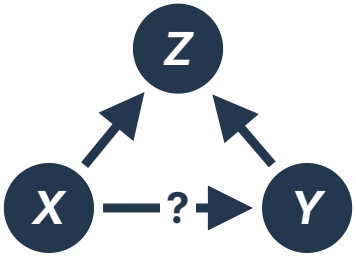


Z alters the relationship between *X* and *Y*.

Can be included as interaction variable to better describe the relationship between *X* and *Y*.

E.g.:
Marital status moderates the relationship between gender and income.

Collider

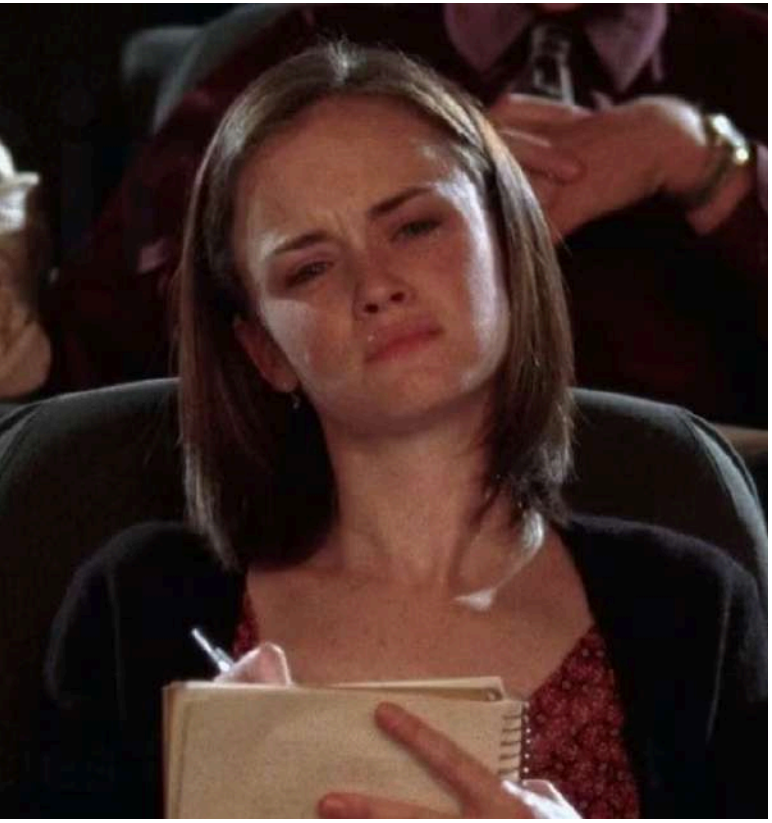


Z causally influenced by both *X* and *Y*.

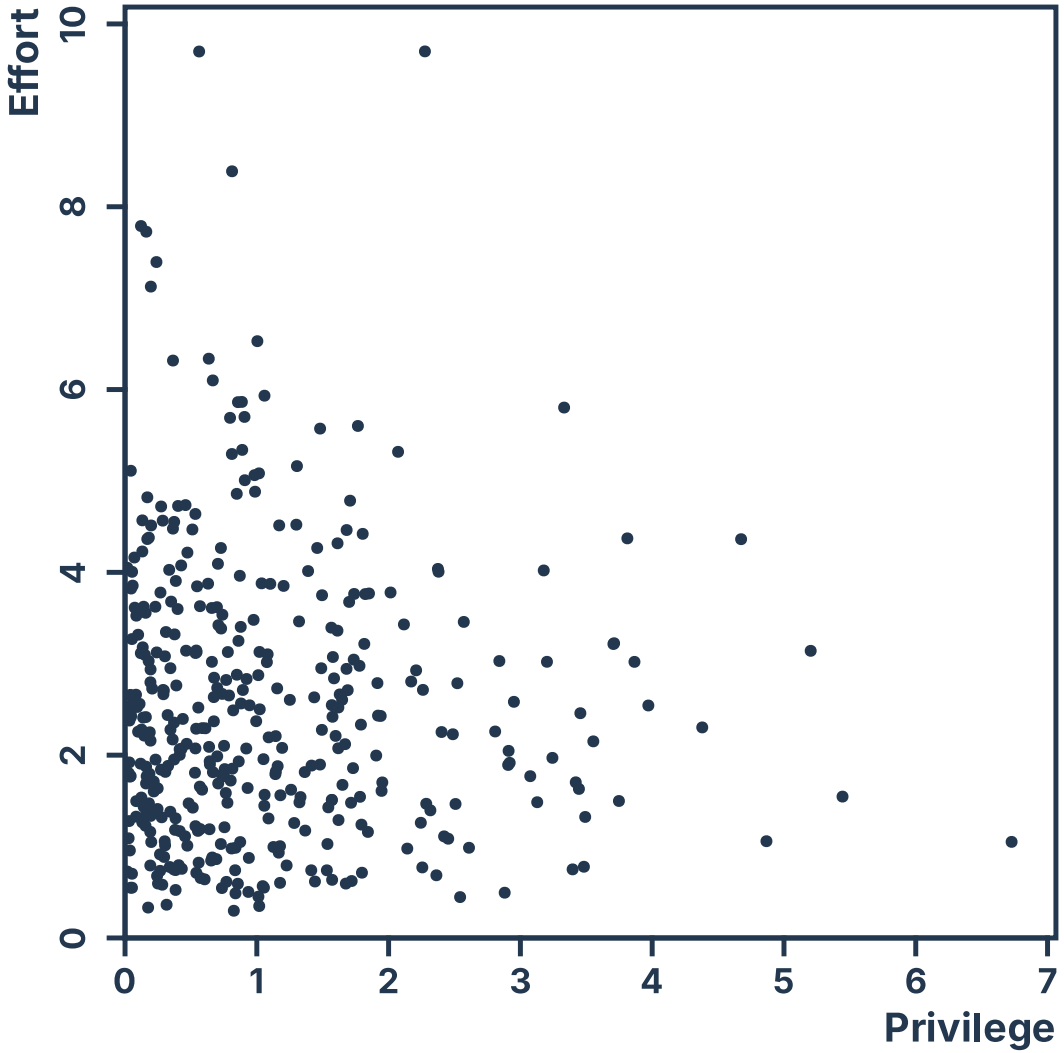
Must ***not*** be "controlled for" when establishing relationship between *X* and *Y*.

E.g.:
Income is a collider for the relationship between gender and occupation.

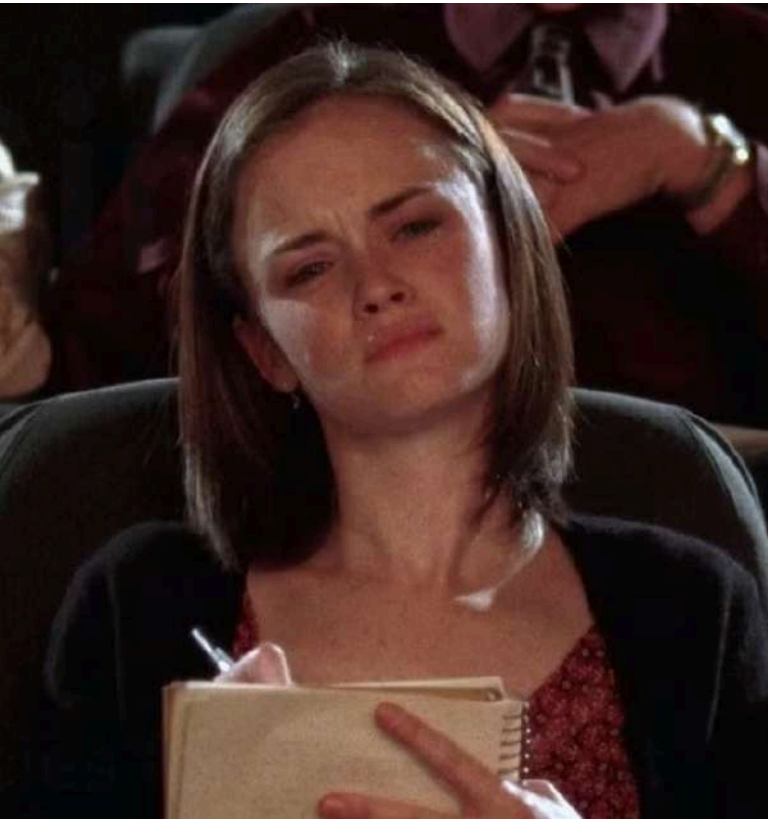
COLLIDER BIAS: AN EXAMPLE



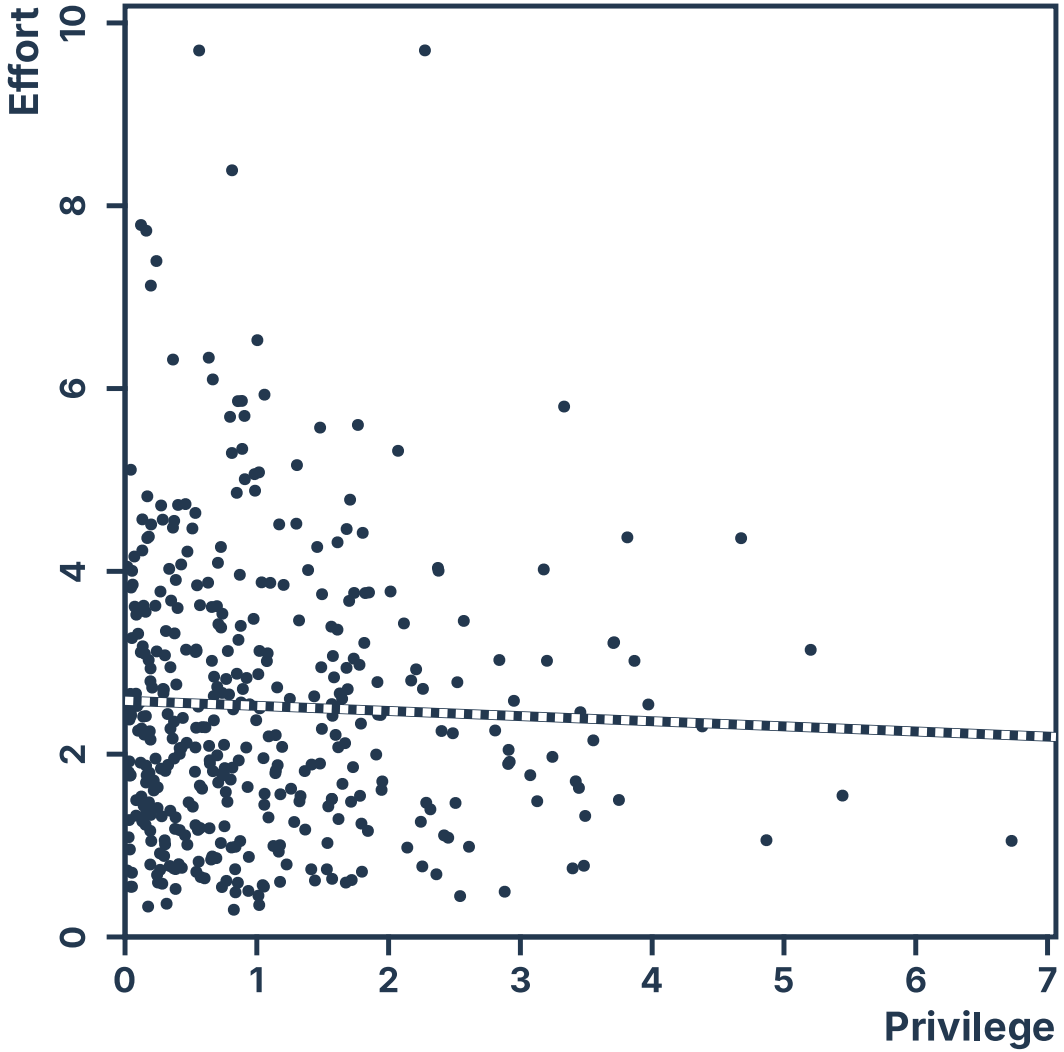
Do college students who come from privileged backgrounds dedicate less effort to studying?



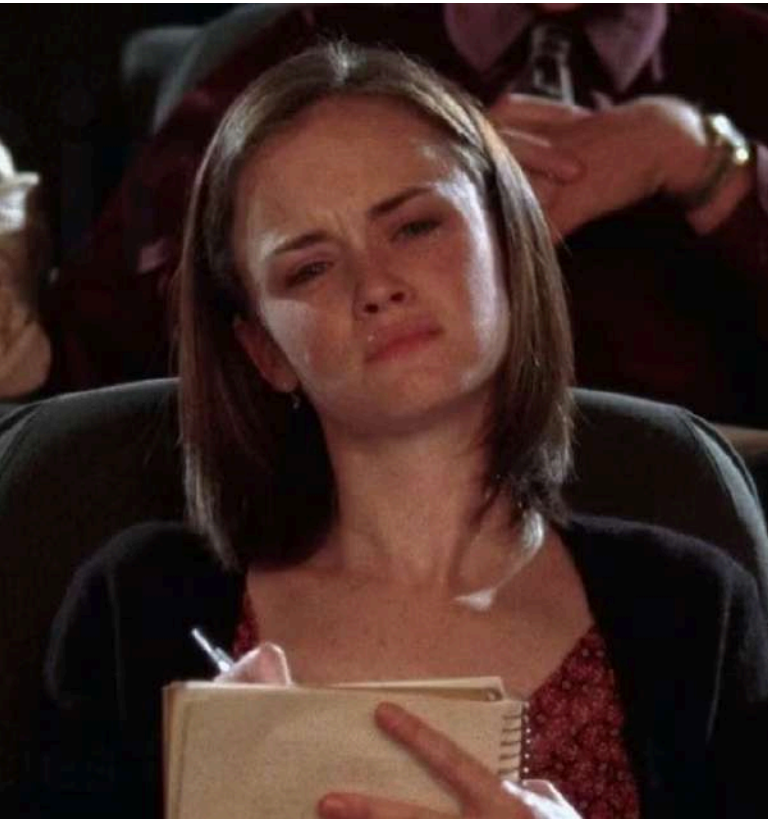
COLLIDER BIAS: AN EXAMPLE

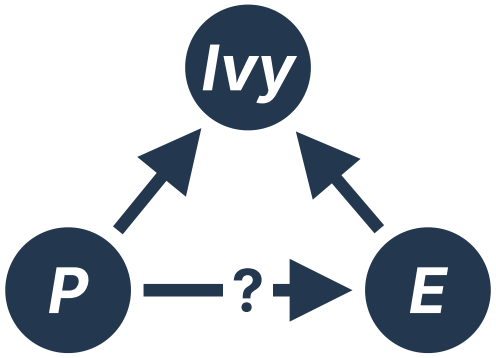


Do college students who come from privileged backgrounds dedicate less effort to studying?

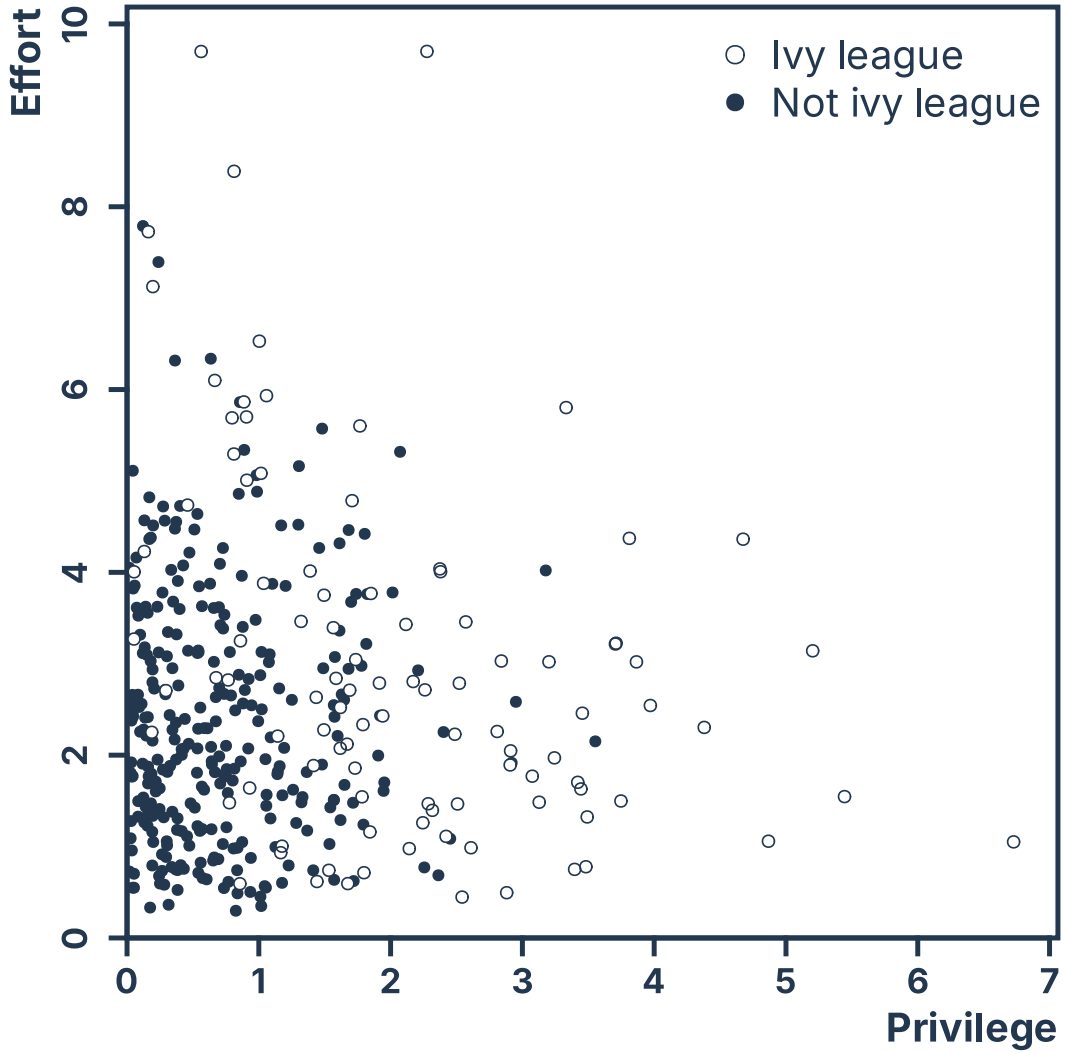


COLLIDER BIAS: AN EXAMPLE

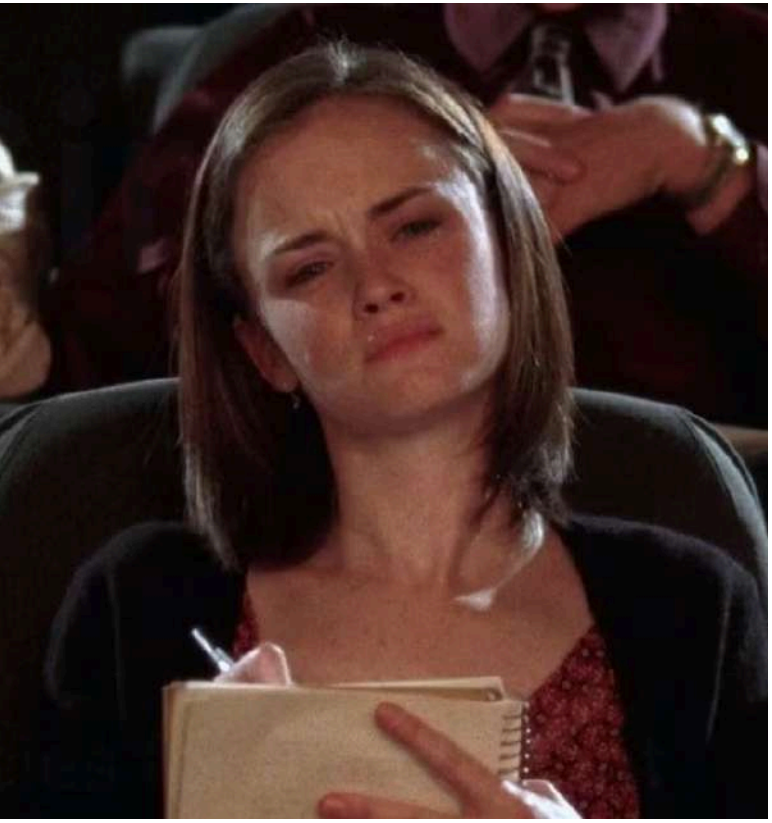


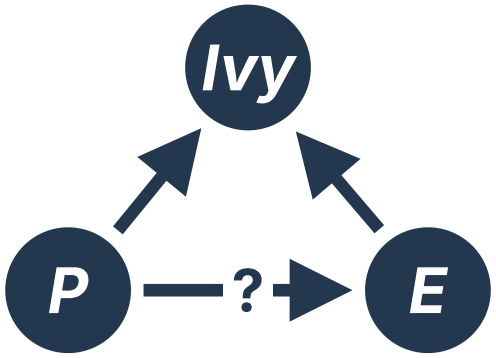


Do college students who come from privileged backgrounds dedicate less effort to studying?

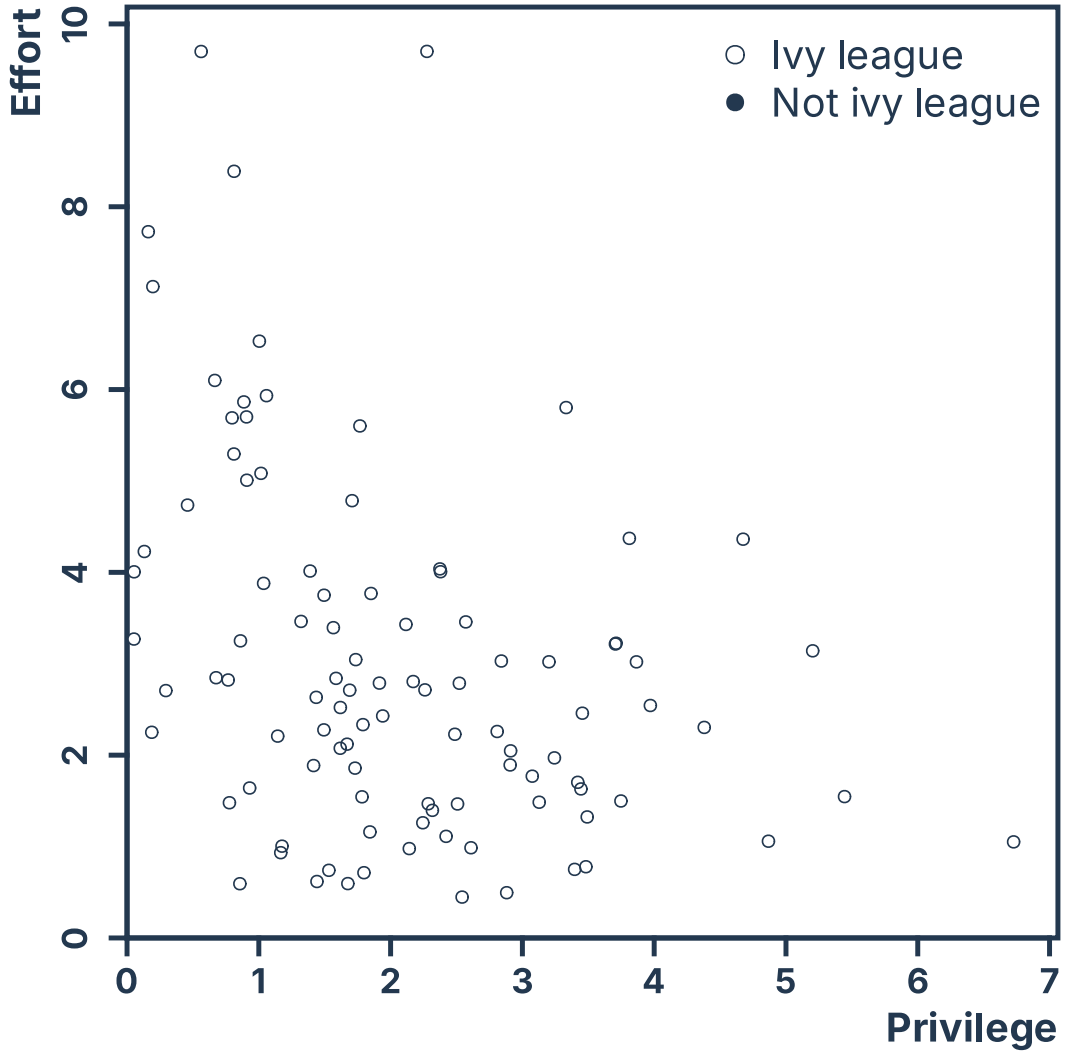


COLLIDER BIAS: AN EXAMPLE

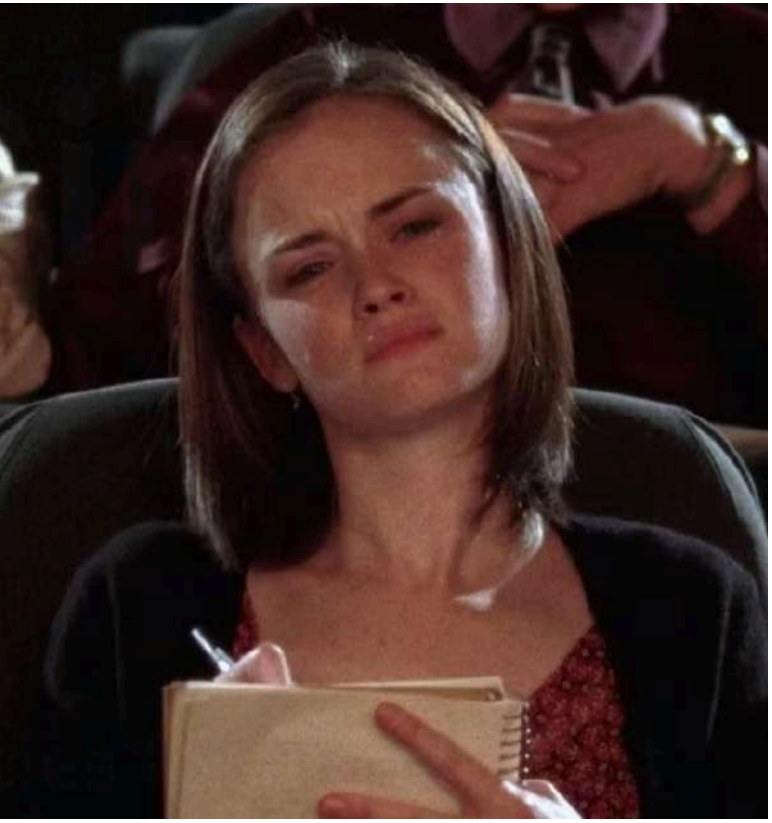


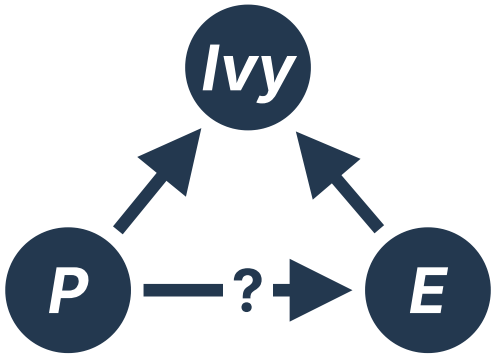


Do college students who come from privileged backgrounds dedicate less effort to studying?



COLLIDER BIAS: AN EXAMPLE





Do college students who come from privileged backgrounds dedicate less effort to studying?

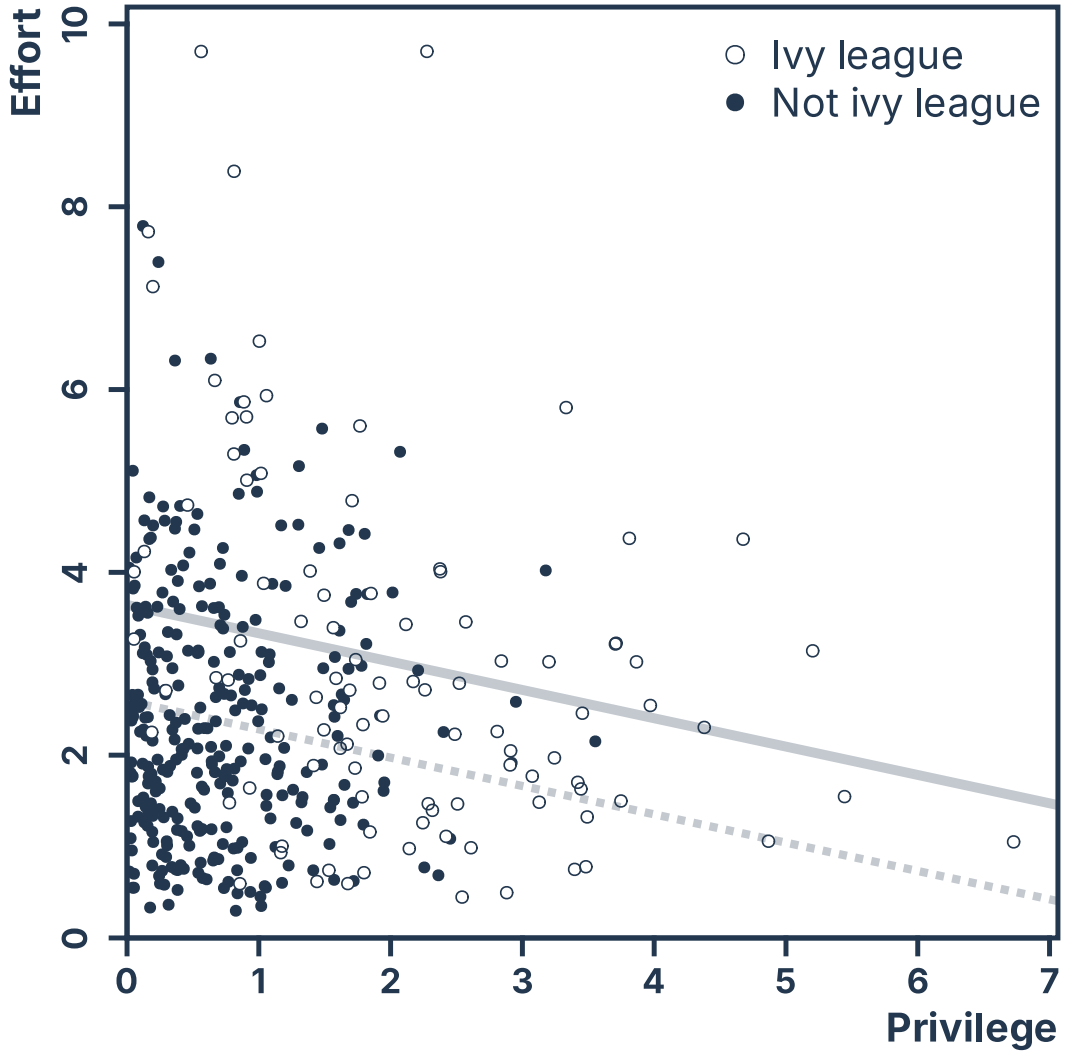
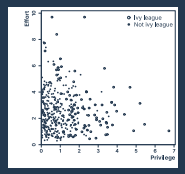


Image credit



Figures by Peter McMahan (source code)

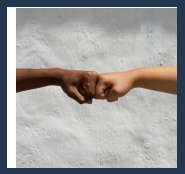
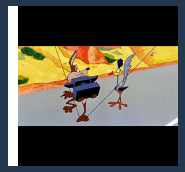
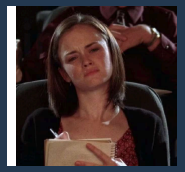


Photo by Arturo Añez on Unsplash



Road Runner and Wile E. Coyote © Warner Bros. Entertainment



Still from Gilmore Girls (2000)