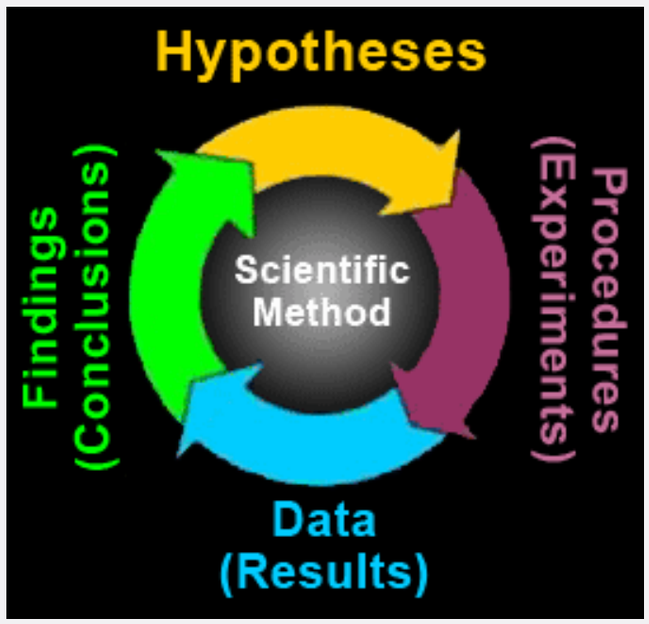
PSC 202
SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

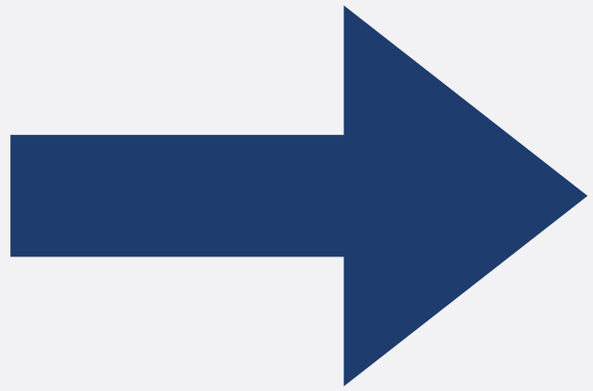
MULTIPLE REGRESSION, PART 1

WHERE WE ARE

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- Use data to evaluate hypotheses
- Reassess explanation



WHERE WE ARE

Sleep (X)  GPA (Y)

- Fundamental problem of causal inference
  - Causal effect of sleep: GPA if you sleep a lot - GPA if you sleep little
  - We don’t observe you in both states, only in one (either you sleep a lot, or not)

ZERO-ORDER RELATIONSHIP

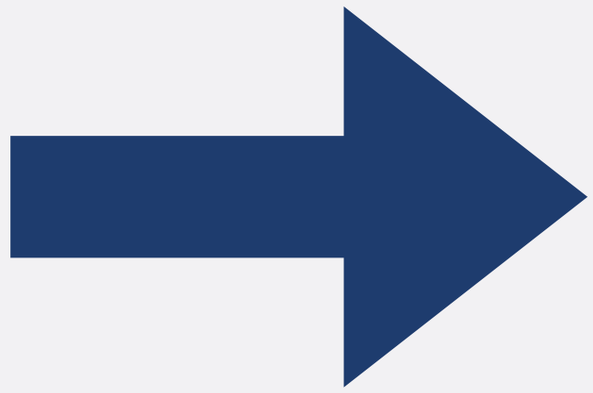
	How much sleep?	
	More Than 7 Hours/Night	7 Or Fewer Hours/Night
Average Gpa	3.61	3.45
	(55)	(78)

0.16

- Frequency in parentheses

PROBLEM

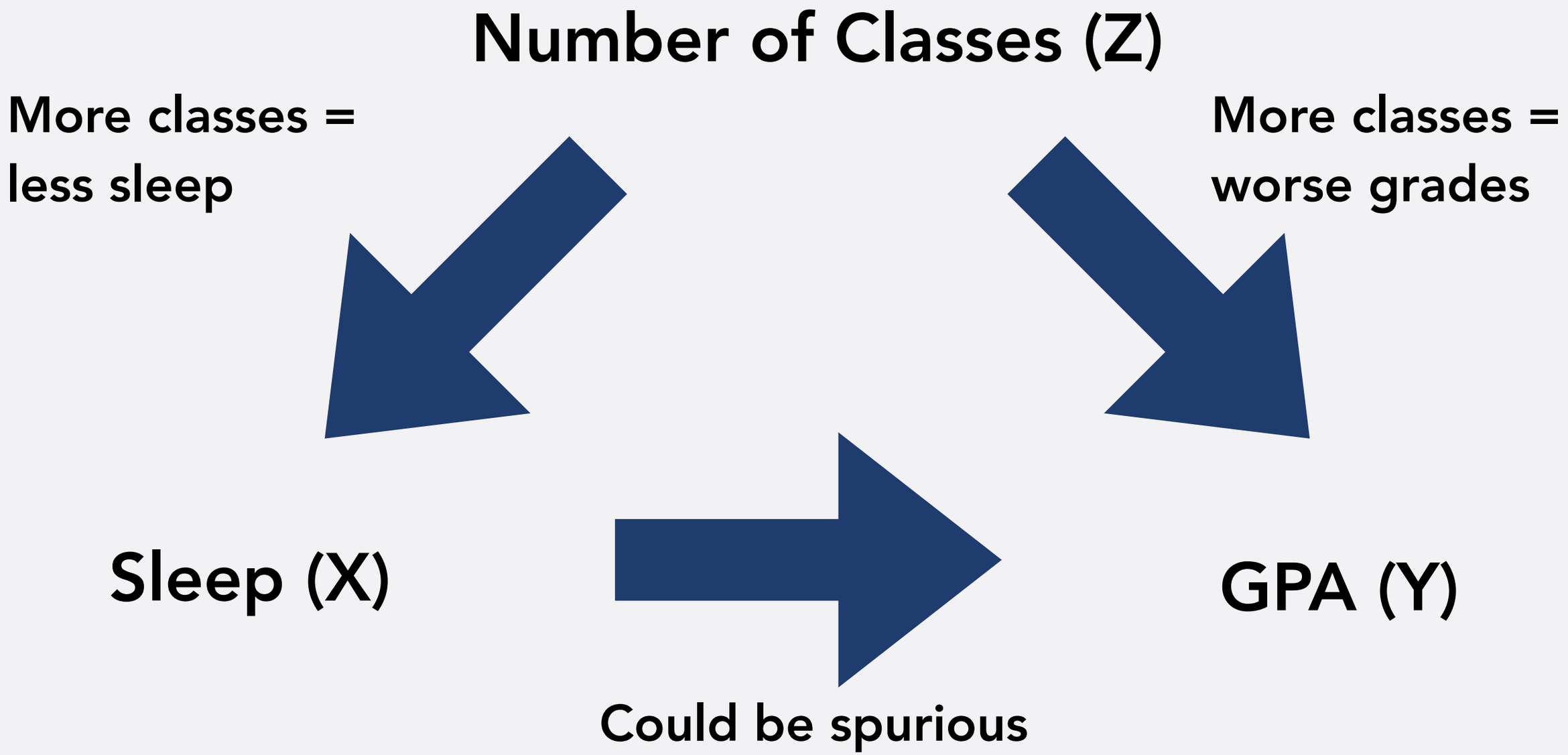
Sleep (X)



GPA (Y)

- Fundamental problem of causal inference
  - We can compute GPA of people who sleep a lot - GPA of people who sleep little
  - But: People who sleep a lot might be different from people who don’t in many other ways
  - And these other differences might affect GPA as well

GPA



- Does sleep still have effect on GPA when controlling for number of classes?

CONTROLLED EFFECTS

	5 Or Fewer Classes		6 Or More Classes	
Sleep	More Than 7 Hours/Night	7 Or Fewer Hours/Night	More Than 7 Hours/Night	7 Or Fewer Hours/Night
Average Gpa	3.53 (40)	3.45 (53)	3.62 (17)	3.45 (27)
	0.08		0.17	

- Idea: Look at relation between sleep and GPA among people who take many classes and (separately) among people who take few classes

CONTROLLED EFFECTS

	5 Or Fewer Classes		6 Or More Classes	
Sleep	More Than 7 Hours/Night	7 Or Fewer Hours/Night	More Than 7 Hours/Night	7 Or Fewer Hours/Night
Average Gpa	3.53 (40)	3.45 (53)	3.62 (17)	3.45 (27)
	0.08		0.17	

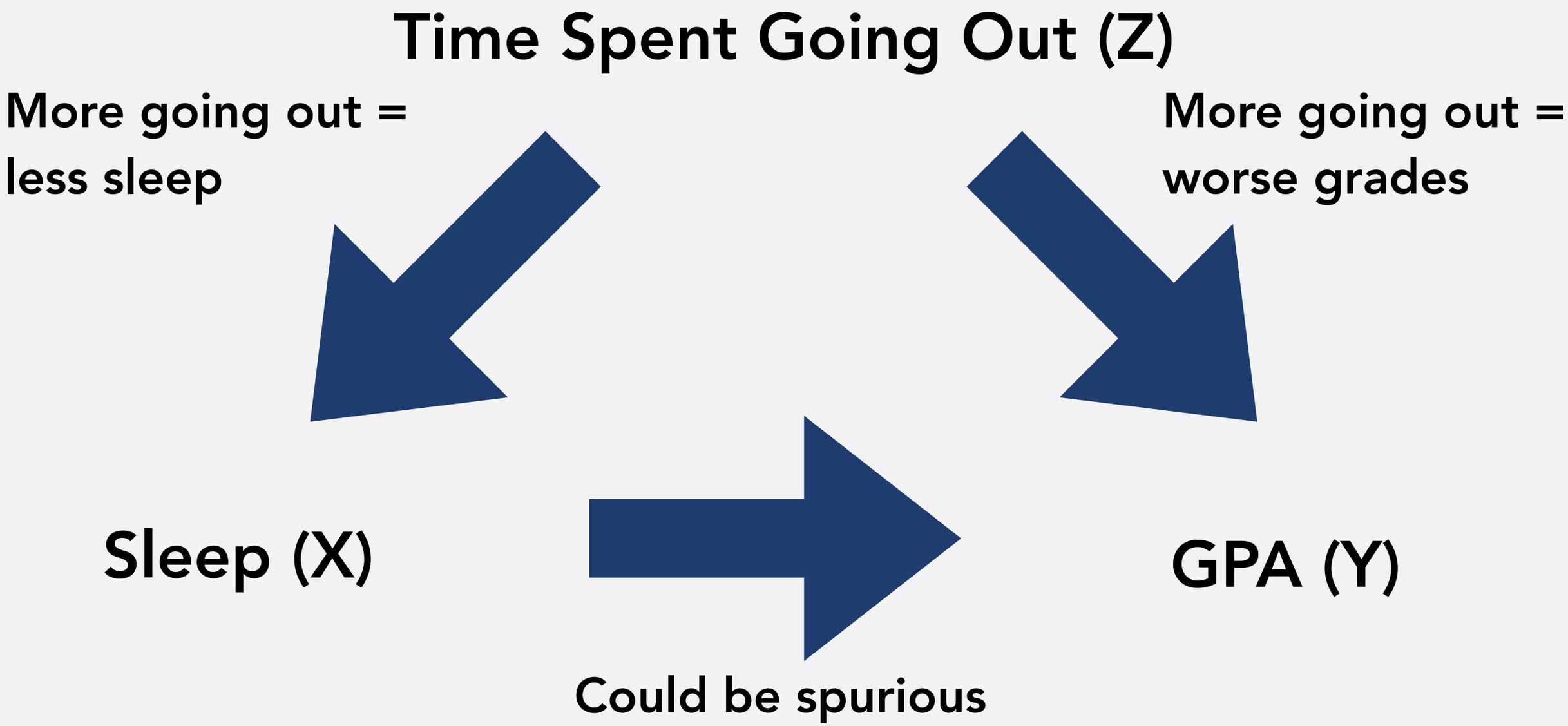
- Still differences in GPA between more/less sleep when controlling for number of classes.

CONTROLLED EFFECTS

	5 Or Fewer Classes		6 Or More Classes	
Sleep	More Than 7 Hours/Night	7 Or Fewer Hours/Night	More Than 7 Hours/Night	7 Or Fewer Hours/Night
Average Gpa	3.53 (40)	3.45 (53)	3.62 (17)	3.45 (27)
	0.08		0.17	

- More nuance: Interactive relation. Sleep matters a lot for GPA among people who take many classes, sleep matters less for GPA among people who take fewer classes

GPA



- Does sleep still have effect on GPA when controlling for how often students go out?

CONTROLLED EFFECTS

Sleep	Go Out 3 Nights Per Week Or Less		Go Out More Than 3 Nights A Week	
	More Than 7 Hours/Night	7 Or Fewer Hours/Night	More Than 7 Hours/Night	7 Or Fewer Hours/Night
Average Gpa	3.59	3.50	3.66	3.44
	(34)	0.09	(48)	(24) 0.22 (34)

- Idea: Look at relation between sleep and GPA among people who go out less and (separately) among people who go out more

CONTROLLED EFFECTS

	Go Out 3 Nights Per Week Or Less			Go Out More Than 3 Nights A Week		
Sleep	More Than 7 Hours/Night	7 Or Fewer Hours/Night		More Than 7 Hours/Night	7 Or Fewer Hours/Night	
Average Gpa	3.59	3.50		3.66	3.44	
	(34)	0.09	(48)	(24)	0.22	(34)

- Sleep still matters for GPA after controlling for how often people go out

HURDLES TO CAUSALITY

- Is there a credible causal mechanism that connects X to Y?
- Can we rule out the possibility that Y could cause X?
- Is there covariation between X and Y?
- Have we controlled for all confounding variables (Z) that might make the association between X and Y spurious?

EXAMPLE

	Sleep More	Sleep Less
Gpa		

2 INDEPENDENT VARIABLES

	More Going Out		Less Going Out	
	Sleep More	Sleep Less	Sleep More	Sleep Less
Gpa				

3 INDEPENDENT VARIABLES

	More Classes				Fewer Classes			
	More Going Out		Less Going Out		More Going Out		Less Going Out	
	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less
Gpa								

4 INDEPENDENT VARIABLES

	Male								Female							
	More Classes				Fewer Classes				More Classes				Fewer Classes			
	More Going Out		Less Going Out		More Going Out		Less Going Out		More Going Out		Less Going Out		More Going Out		Less Going Out	
	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less	Sleep More	Sleep Less
Gpa																

- As we control for more potential confounders, table gets increasingly unwieldy (and few/no observations in some cells)

ANOTHER ISSUE

- What if a control is interval-level?
  - e.g. sleep or number of classes
  - I divided into more/less sleep and more/fewer classes, but that's not ideal
  - But if I didn't do this, unwieldy table with many cells where there are no observations

TODAY

- **Multiple regression**
- **Extends bivariate regression to incorporate not just one, but many independent variables**

CONTROLLED COMPARISON

- Logic of controlled comparison:
  - Separate students with more classes and fewer classes
  - Among students w/ more classes: What is the effect of sleep on GPA?
  - Among students w/ fewer classes: What is the effect of sleep on GPA?
- Gives us the partial effects of sleep, holding number of classes constant

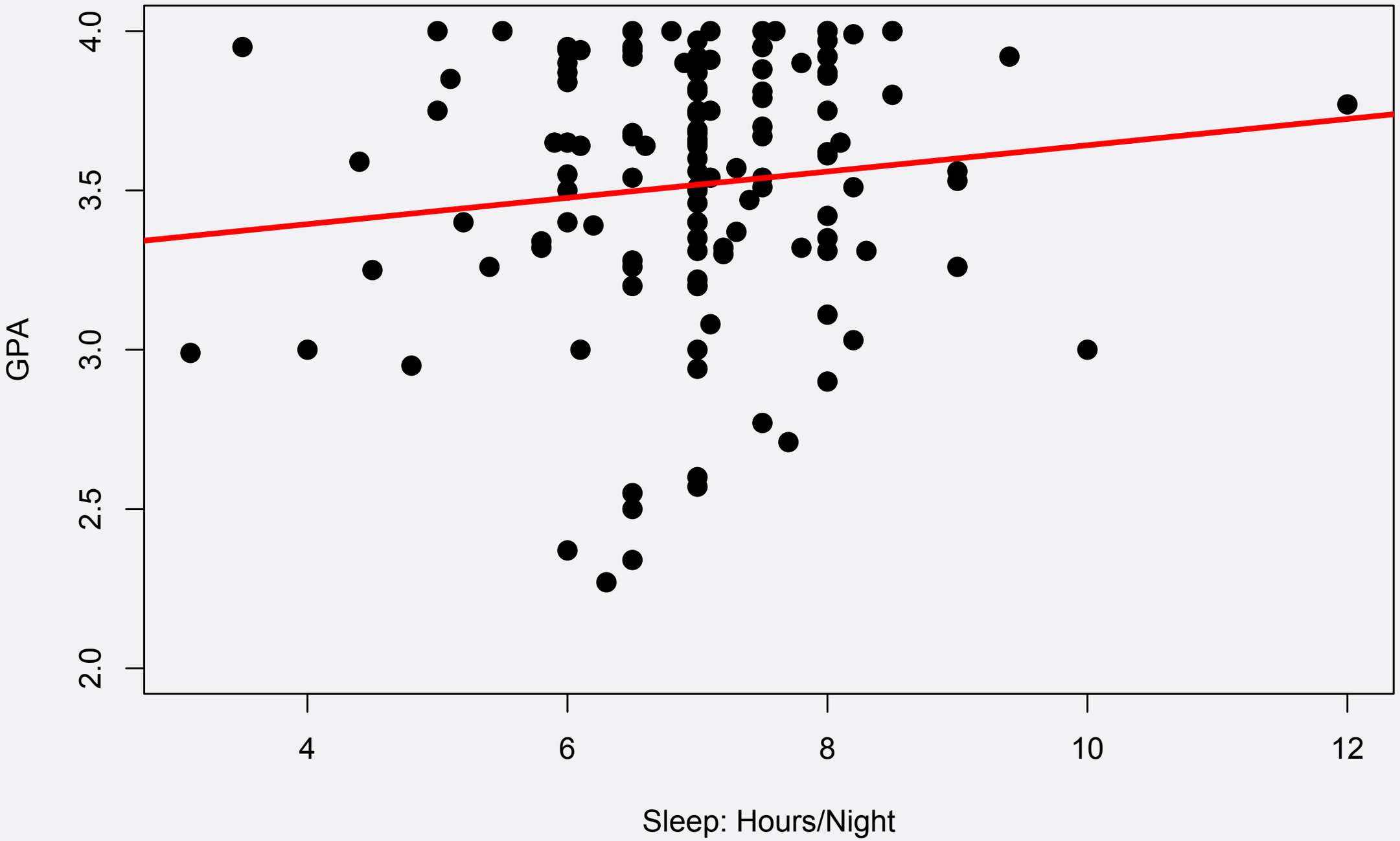
MULTIPLE REGRESSION

- Multiple regression does something similar
- Can estimate the effect of two variables on dependent variable (GPA)
  - Gives the partial effect of sleep, holding number of classes constant
  - And: gives the partial effect of number of classes, holding sleep constant

MULTIPLE REGRESSION

- Can include more than 2 independent variables
  - e.g. sleep, number of classes, how often going out
  - Gives the partial effect of sleep, holding constant classes and going out
  - Gives the partial effect of classes, holding constant sleep and going out
  - Gives the partial effect of going out, holding constant sleep and classes

BIVARIATE REGRESSION



BIVARIATE REGRESSION

	Coefficient	Standard Error	T-Value
Intercept	3.23	0.23	14.08
Sleep: Hours/ Night	0.04	0.03	1.28

R^2: 0.012

LINEAR REGRESSION

- $y = a + b_1 * x_1$
- y: GPA
- x_1: hours of sleep/night

LINEAR REGRESSION

- Let's add number of classes as a second control
- $y = a + b_1*x_1 + b_2*x_2$
- y: GPA
- x_1: hours of sleep/night
- x_2: number of classes

EFFECT OF SLEEP

	Coefficient	Standard Error	T-Value
Intercept	3.28	0.36	9.14
Sleep: Hours/ Night	0.04	0.03	1.27
Number Of Classes	-0.01	0.05	-0.19

R^2: 0.013

EFFECT OF SLEEP

- Coefficient: 0.04 (SE 0.03, t-value 1.27)
- Interpretation: For every additional our of sleep, GPA increases by 0.04 points, *holding all other variables constant*
- We cannot reject H_0, so effect of sleep on GPA is not statistically significant at the 5% level

EFFECT OF CLASSES

	Coefficient	Standard Error	T-Value
Intercept	3.28	0.36	9.14
Sleep: Hours/ Night	0.04	0.03	1.27
Number Of Classes	-0.01	0.05	-0.19

R^2: 0.013

EFFECT OF AGE

- Coefficient: -0.01 (SE 0.05, t-value -0.19)
- Interpretation: For every additional class taken, GPA decreases by 0.01 points, *holding all other variables constant*
- We cannot reject H_0, so effect of number of classes on GPA is *not* significant at the 5% level

INTERCEPT

	Coefficient	Standard Error	T-Value
Intercept	3.28	0.36	9.14
Sleep: Hours/ Night	0.04	0.03	1.27
Number Of Classes	-0.01	0.05	-0.19

R^2: 0.013

INTERCEPT

- Intercept: 3.28
- Gives expected GPA when *both* sleep=0 *and* number of classes=0

INTERCEPT

	Coefficient	Standard Error	T-Value
Intercept	3.28	0.36	9.14
Sleep: Hours/ Night	0.04	0.03	1.27
Number Of Classes	-0.01	0.05	-0.19

R^2: 0.013

LINEAR REGRESSION

- Let's add days/week going out as another control
- $y = a + b_1*x_1 + b_2*x_2 + b_3*x_3$
  - y: GPA
  - x_1: hours of sleep/night
  - x_2: number of classes
  - x_3: days/week going out

INTERCEPT

	Coefficient	Standard Error	T-Value
Intercept	3.17	0.39	8.03
Sleep: Hours/ Night	0.04	0.03	1.13
Number Of Classes	0.01	0.05	0.19
Going Out: Days/Week	0.02	0.03	0.53

R^2: 0.013

CETERIS PARIBUS

- Linear regression allows us to estimate the effect of several independent variables on the dependent variable
- Gives us the effect of an independent variable on the dependent variable, *holding all other variables constant*
  - "ceteris paribus"
- We can assess the effect of the variables independently of each other

CETERIS PARIBUS

- GPA = 3.17 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out
- What is the predicted evaluation for a person who sleeps 8 hours, takes 5 classes, and goes out 3 nights a week?

CETERIS PARIBUS

- GPA = 3.17 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out
- What is the predicted evaluation for a person who sleeps 8 hours, takes 5 classes, and goes out 3 nights a week?
- GPA = 3.17 + 0.04*8 + 0.01*5 + 0.02*3 = 3.6

CETERIS PARIBUS

- GPA = 3.17 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out
- What is that person slept 4 hours instead?

CETERIS PARIBUS

- GPA = 3.17 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out
- What is that person slept 4 hours instead?
- GPA = 3.17 + 0.04*4 + 0.01*5 + 0.02*3 = 3.44

CETERIS PARIBUS

- GPA = 3.17 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out
- And what if they went out 6 nights/week?

CETERIS PARIBUS

- GPA = 3.17 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out
- And what if they went out 6 nights/week?
- GPA = 3.17 + 0.04*4 + 0.01*5 + 0.02*6 = 3.5

CETERIS PARIBUS

- Of course, going out more might also lead to sleeping less
- But the linear regression estimates the effects of sleep and going out *independently of each other*

LINEAR REGRESSION

- So far: The independent variables were interval-level
  - Hours of sleep, number of classes, number of nights going out
- What if independent variable is nominal or ordinal-level?
  - e.g. effect of gender

DUMMY VARIABLE REGRESSION

- Nominal or ordinal-level independent variable can easily be incorporated in linear regression
- $y = a + b_1*x_1 + b_2*x_2 + b_3*x_3 + b_4*x_4$
  - x_3=0 if gender=female
  - x_3=1 if gender=male
- Same idea as before, but x_4 can only take values of 0 or 1
- "Dummy variable"
  - 0/1

REGRESSION

	Coefficient	Standard Error	T-Value
Intercept	3.16	0.40	7.95
Sleep: Hours/ Night	0.04	0.03	1.10
Number Of Classes	0.01	0.05	0.22
Going Out: Days/Week	0.02	0.03	0.49
Gender (Male)	0.04	0.08	0.53

R^2: 0.14

EFFECT OF GENDER

- Coefficient: 0.04 (SE 0.08, t-value 0.53)
  - Where female is coded 0 and male coded 1
- Interpretation: If someone is male, their GPA is expected to be 0.04 points higher than if someone is female, *holding all other variables constant*
- However, we do not reject H_0, so effect of gender on GPA is not significant at the 5% level

EFFECT OF GENDER

- GPA = 3.16 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out + 0.04*Gender
  - female: gender=0
  - male: gender=1

EFFECT OF GENDER

- GPA = 3.16 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out + 0.04*Gender
  - female: gender=0
  - male: gender=1
- What is the predicted GPA for a student who sleeps 8 hours, takes 5 classes, goes out 2 mights, and who is male?

EFFECT OF GENDER

- GPA = 3.16 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out + 0.04*Gender
  - female: gender=0
  - male: gender=1
- What is the predicted GPA for a student who sleeps 8 hours, takes 5 classes, goes out 2 mights, and who is male?
- GPA = 3.16 + 0.04*8 + 0.01*5 + 0.02*2 + 0.04*1 = 3.61

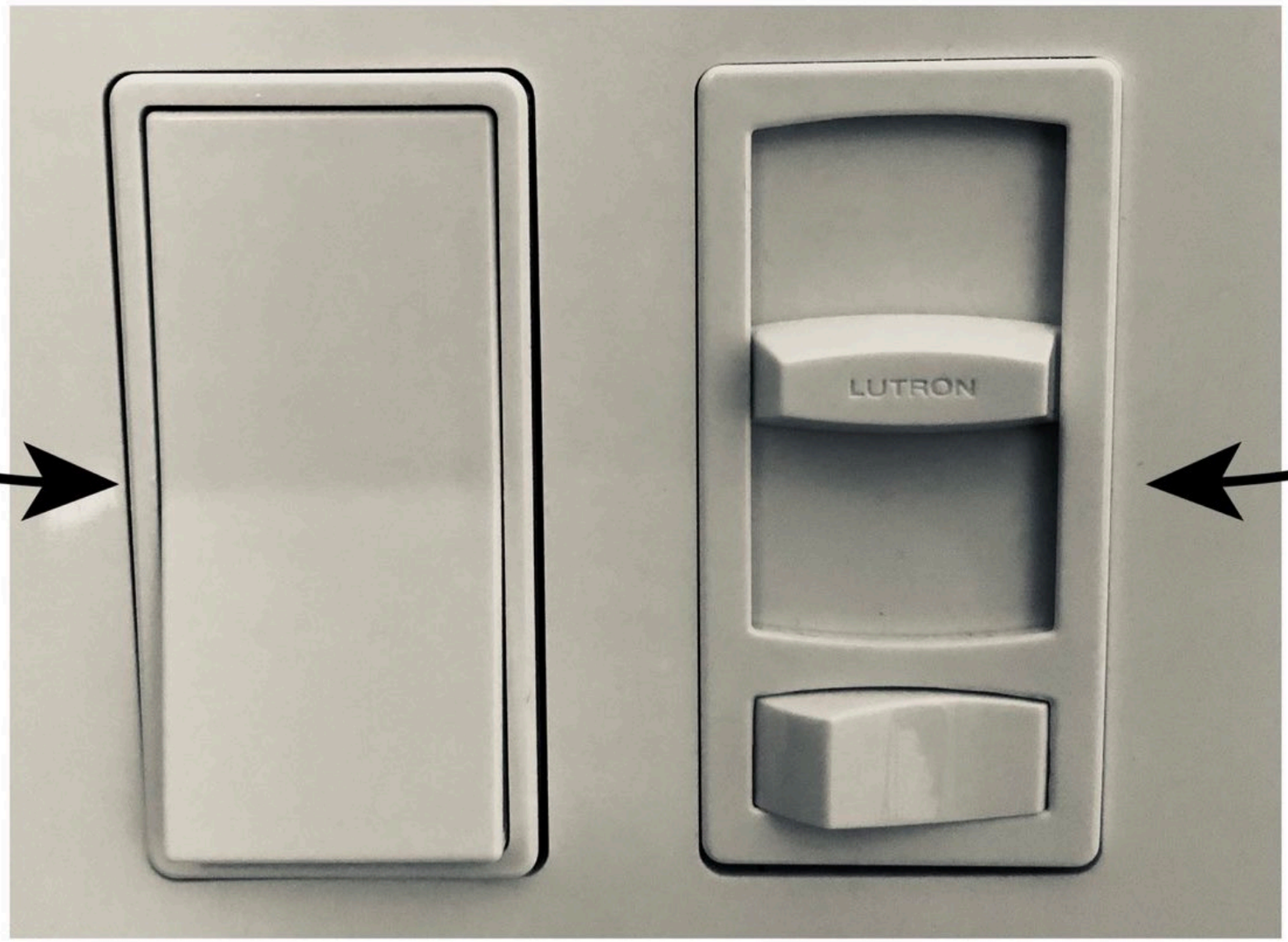
EFFECT OF GENDER

- GPA = 3.16 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out + 0.04*Gender
  - female: gender=0
  - male: gender=1
- What is the predicted GPA for a student who sleeps 8 hours, takes 5 classes, goes out 2 mights, and who is female?

EFFECT OF GENDER

- GPA = 3.16 + 0.04*Sleep + 0.01*Classes + 0.02*Going Out + 0.04*Gender
  - female: gender=0
  - male: gender=1
- What is the predicted GPA for a student who sleeps 8 hours, takes 5 classes, goes out 2 mights, and who is female?
- GPA = 3.16 + 0.04*8 + 0.01*5 + 0.02*2 + 0.04*0 = 3.57

SLIDERS AND SWITCHES



NOTE

- All independent variables we tried do not show any significant effect on GPA
  - And the R-squared is very low
  - Interesting: I tried some other variables, and it seems like sleep during finals week has a significant effect
    - Every additional hour of sleep per night during finals week is associated with increase in GPA of 0.07.

WHAT THIS ALLOWS US TO DO

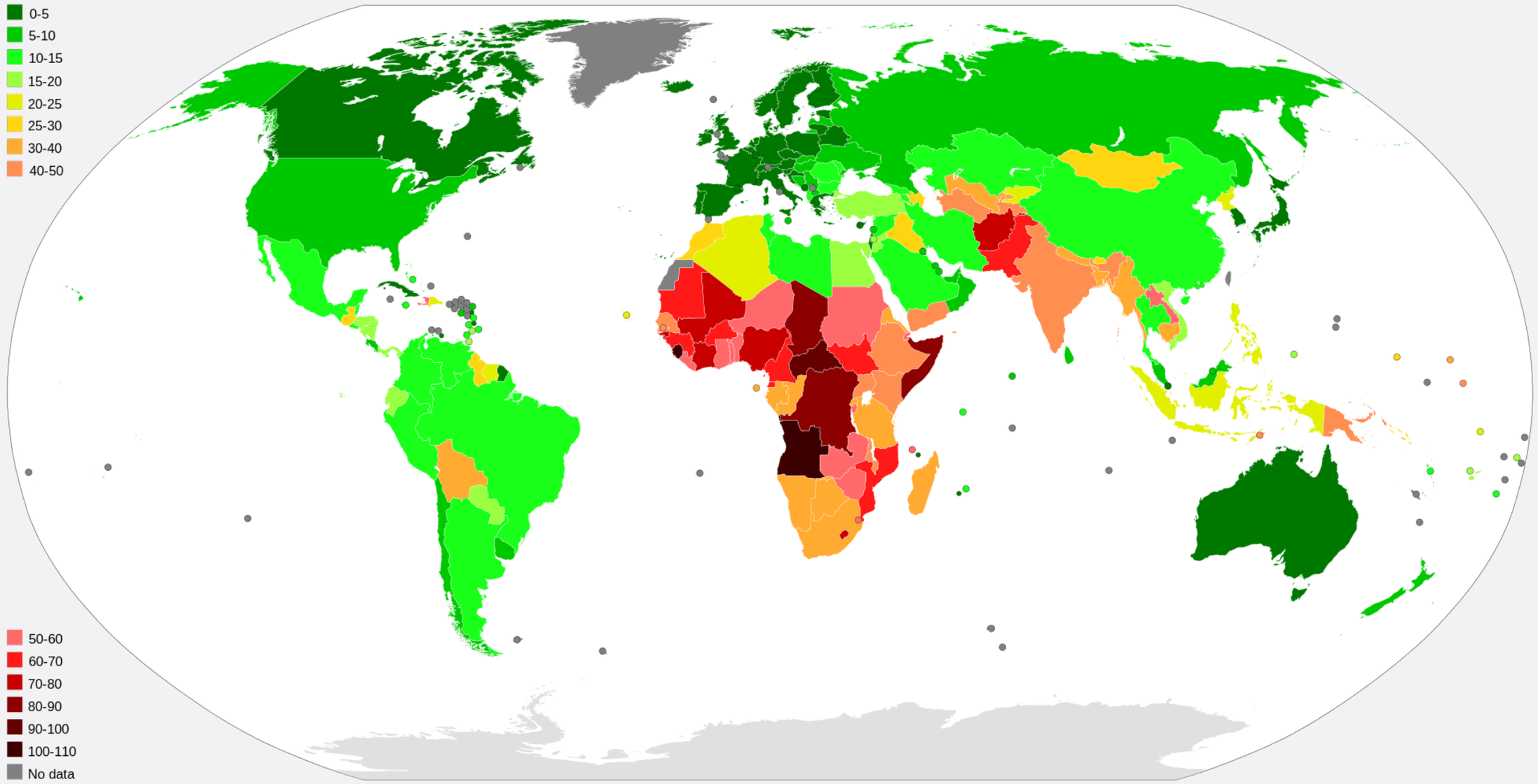
- **Multiple regression is a tool that allows us to tackle the fourth hurdle to causality**
  - **Have we controlled for *all* confounding variables (Z) that might make the association between X and Y spurious?**
  - **We can now estimate effect of X on Y controlling for all confounders we can think of (Z_1, Z_2, etc.)**

WHAT THIS ALLOWS US TO DO

- If we have not one theory about what influences Y, but many theories, we can test which one's have an effect on Y and which don't

HOW IS THIS USEFUL?

- What causes high infant mortality rates?



- Infant mortality rates (Death under 1 year of age per 1,000 live births)