Causality

Alexander Franks

Casal Inference.

Linear models

$$Y_{n\times 1} = X_{n\times p} \beta + \varepsilon$$ unknown population parameter.

$$\varepsilon \sim N(0, \sigma^2 I)$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (\text{OLS})$$

$$E[\hat{\beta} \mid X] = \beta \quad (\text{unbiased})$$

$$Var(\hat{\beta} \mid X) = \sigma^2 (X^T X)^{-1}$$

$$y = \underline{\beta_0} + \underline{\beta_1} X_1 + \underline{\beta_2} X_2 + \ldots \underline{\beta_p} X_p + \varepsilon$$

"For each unit increase in X_j, y is associated with a β_j increase."

causes?

A Motivating Example

HEALTH > NUTRITION & DIET

7 Science-Backed Health Benefits of Drinking Red Wine

Yep, moderate red wine consumption is healthy—and here's the proof.

By Ashley Zlatopolsky | Updated on November 5, 2022

Fact checked by Emily Peterson

A Motivating Example

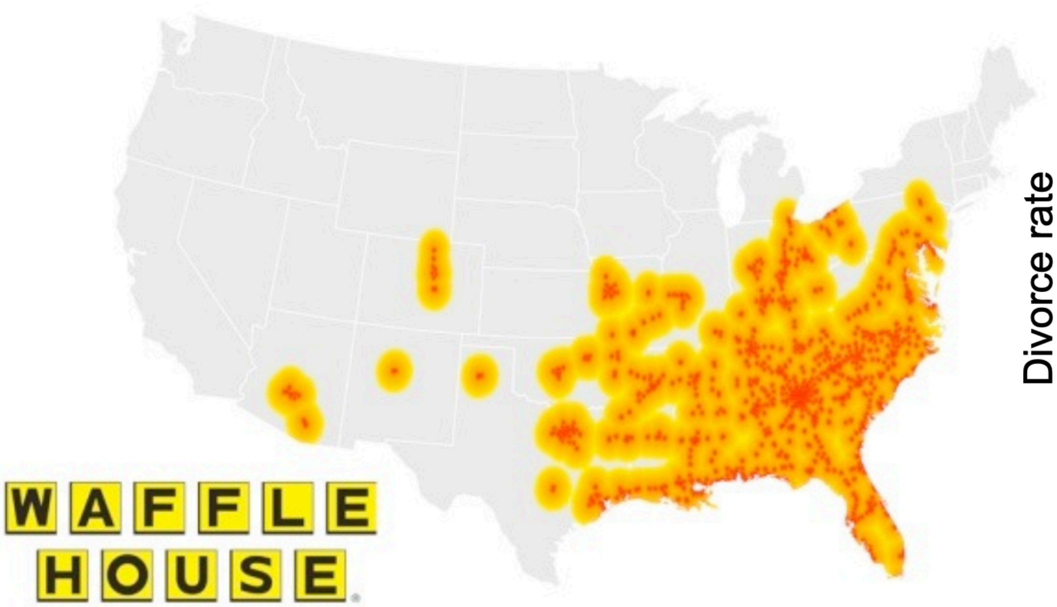
The New York Times

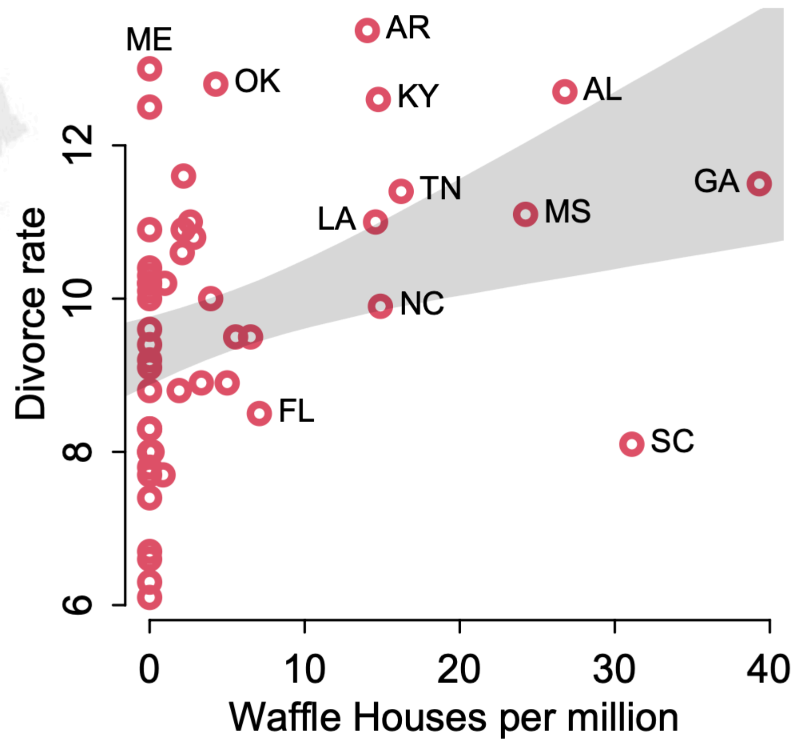
Even a Little Alcohol Can Harm Your Health

Recent research makes it clear that any amount of drinking can be detrimental. Here's why you may want to cut down on your consumption beyond Dry January.

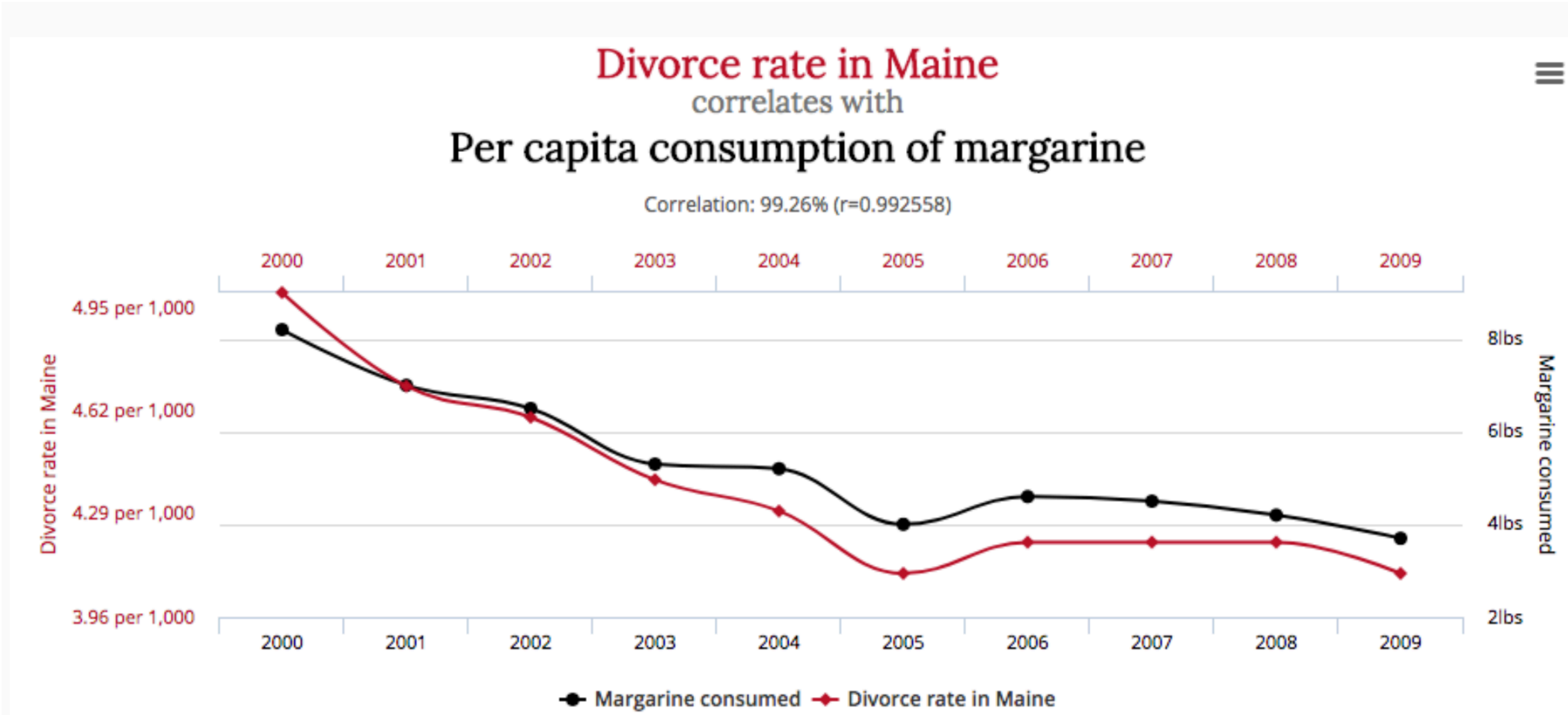
Correlation and Causality

Does Waffle House cause divorce?





Correlation



See https://www.tylervigen.com/spurious-correlations

Correlation and Causality

Question

What added assumptions might imply causation?

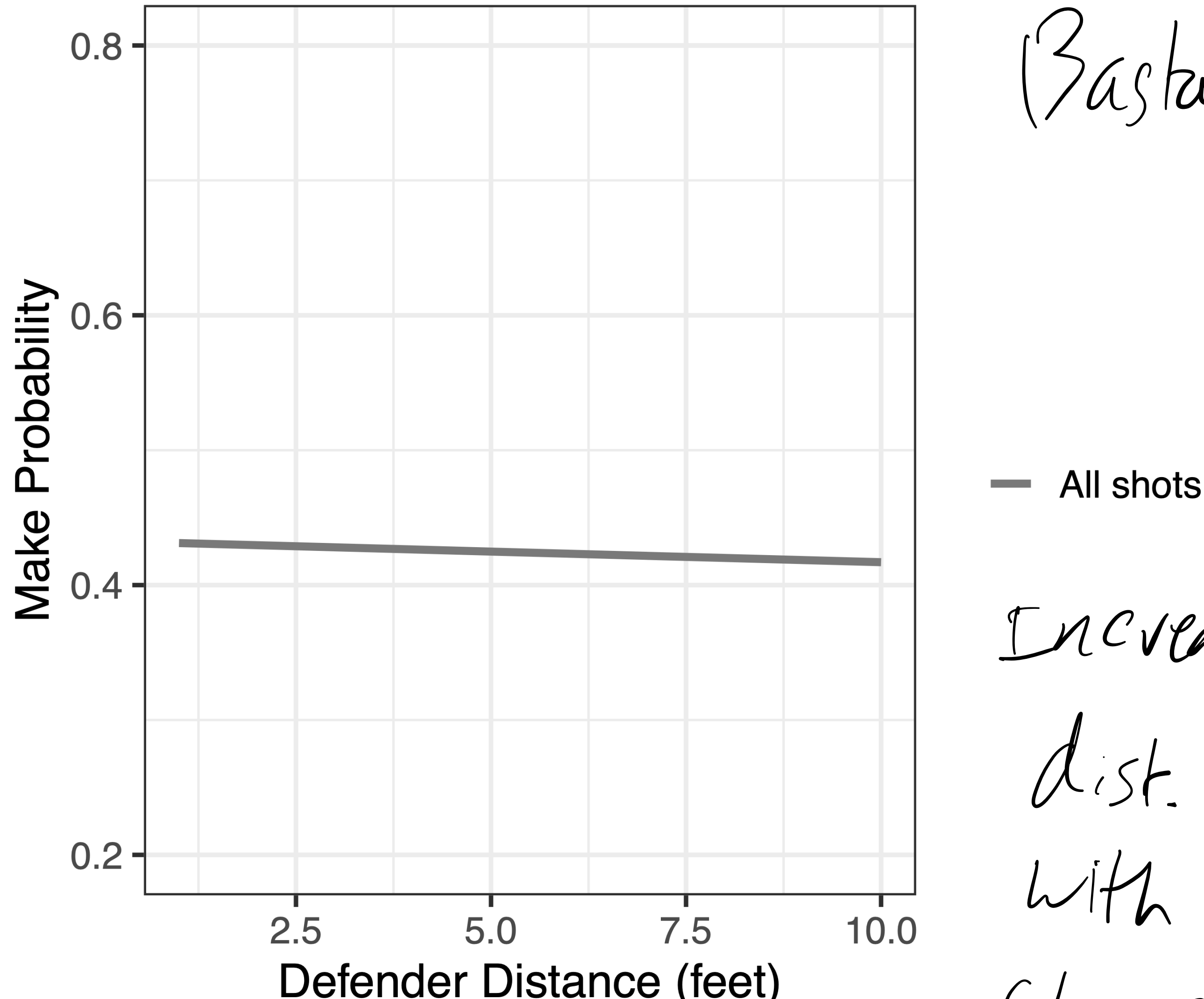
Berkeley Gender Discrimination Example

Category	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	44%	4,321	35%

Berkeley Gender Discrimination Example

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	**825**	62%	108	82%
B	585	63%	**560**	63%	25	68%
C	918	35%	325	37%	**593**	34%
D	792	34%	417	33%	375	35%
E	584	25%	191	28%	**393**	24%
F	714	6%	373	6%	341	7%
Total	4526	39%	2691	45%	1835	30%

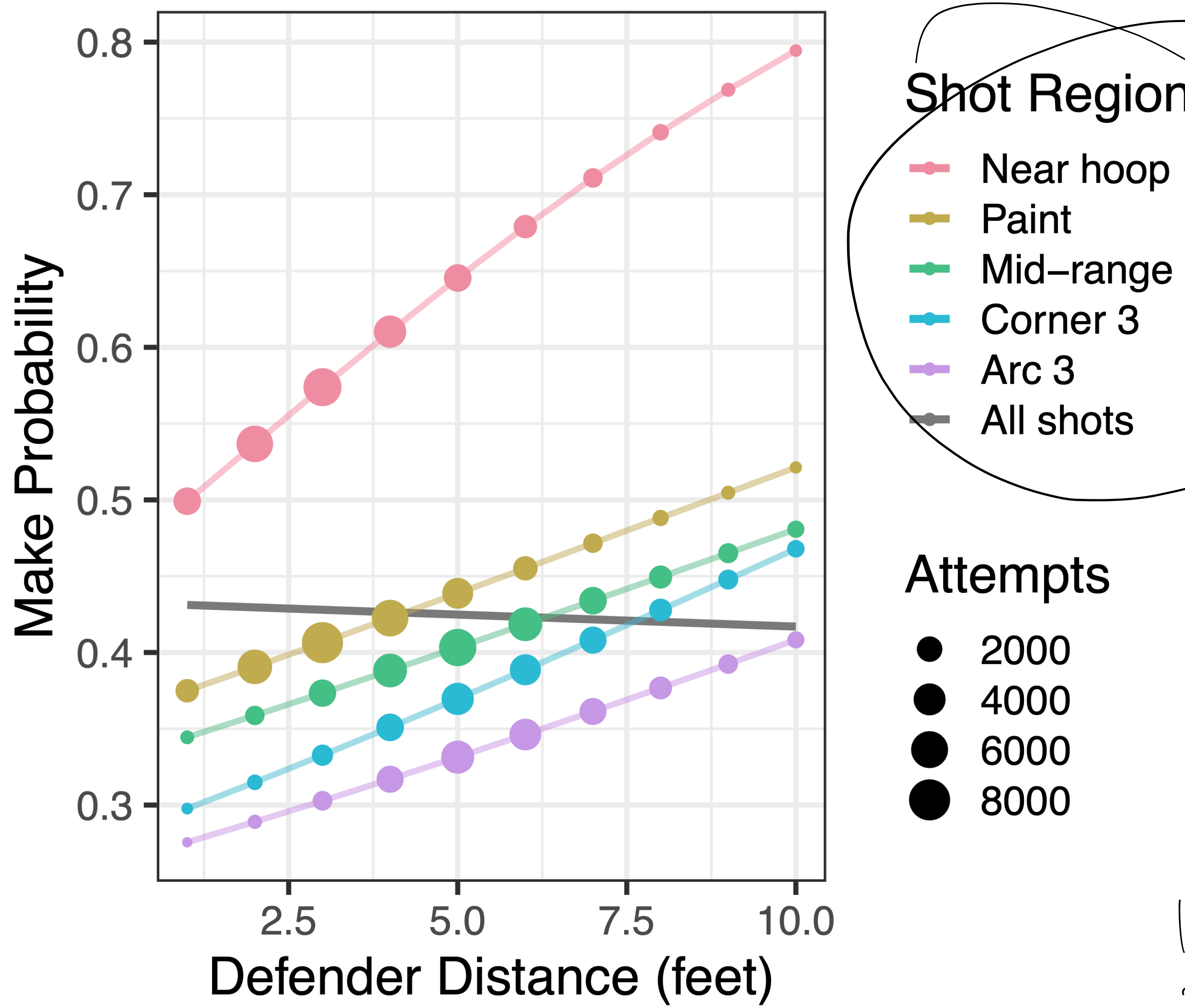
Simpson’s Paradox



Bastatball

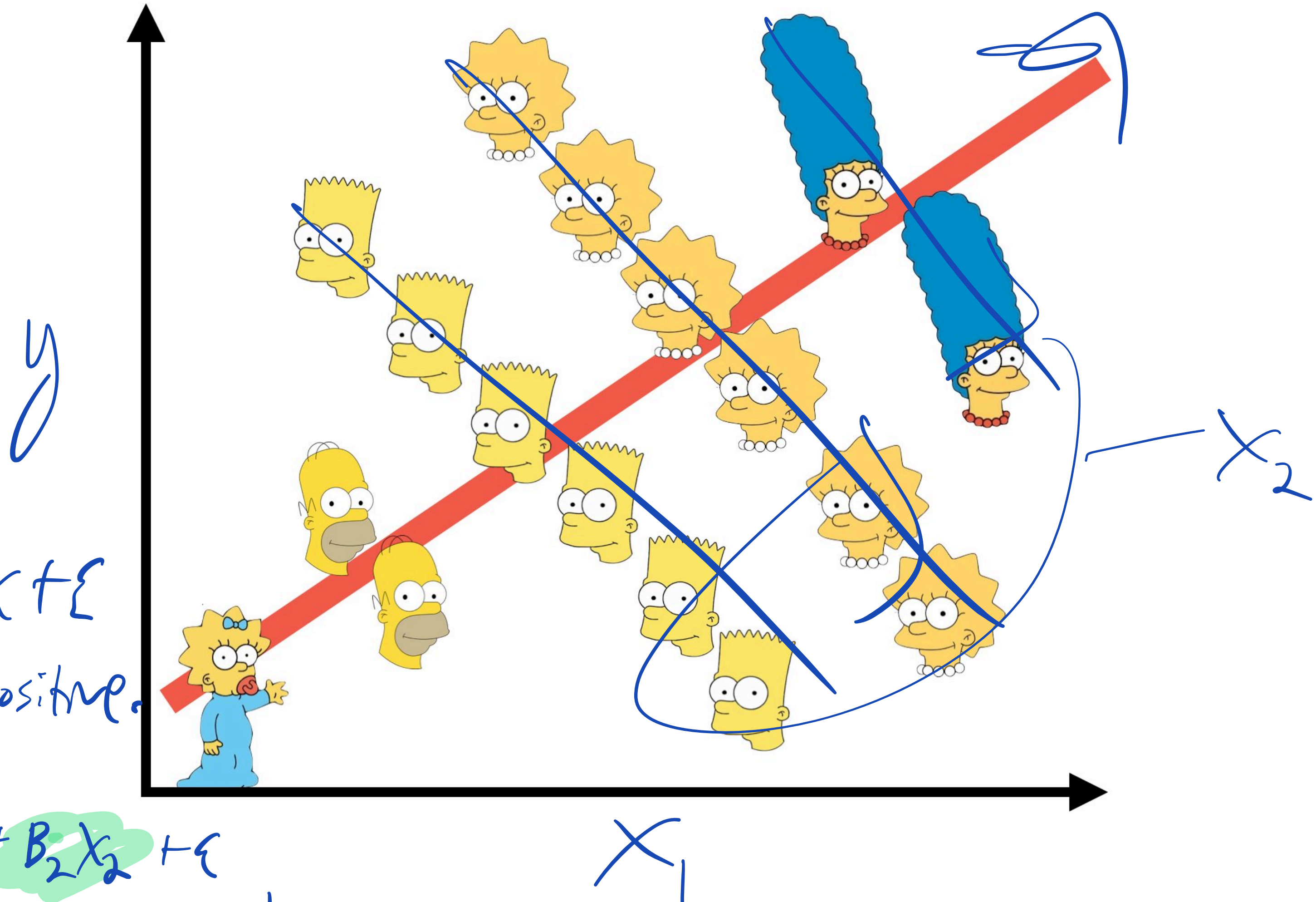
Increasing def. dist. associated with no change in make prob.

Simpson’s Paradox



Effects make prob f Defender Distance.

Simpson's Paradox



$y = B_0 + B_1 X + \varepsilon$

$\hat{B}_1$ is positive.

$y = B_0 + B_1 X_1 + B_2 X_2 + \varepsilon$

$\hat{B}_1$ is ative!

Remember.

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Let $X = (Z, U)$

"treatment"

unmeasured / unobserved.

Truth: $y = Z\beta_1 + U\beta_2 + \varepsilon.$

Can only fit: $y = Z\tilde{\beta}_1 + \varepsilon$

Question: How does $\tilde{\beta}_1$ compare to $\hat{\tilde{\beta}}_1$?

$$((Z, U)^T (Z\ U))^{-1} Z^T y =$$

nez

Omitted Confounders

Suppose the true model is

$$y = Z\beta + U\gamma + \epsilon, \epsilon\ N(0, \sigma^2)$$

but we don't observe U, and thus run `lm(y ~ Z)`. Consider $\hat{\beta}$ inferred from this regression.

...

Then

$$E(\hat{\beta}) = \overbrace{\boldsymbol{\beta}}^{\text{Truth}} + \overbrace{(Z^\top Z)^{-1} Z^\top U\gamma}^{\text{Bias}}$$

$Z^T U = 0$ in Randomized Experiments

Bias is large if $Z^T U$ is large. (Z & U are related)

Randomized vs Natural Experiments

Consider Z to be a "treatment".

- Some designed experiments make X orthogonal to Z, that is, $Z^\top U = 0$
- Randomized experiments ensure that $E[Z^\top U] = 0$
- In natural experiments ("observational studies") we have no direct control over Z

Causality

Association

Predictive questions:

- What will happen?
- How should I react?
- Speculation
- e.g. `predict.lm`

Causal questions

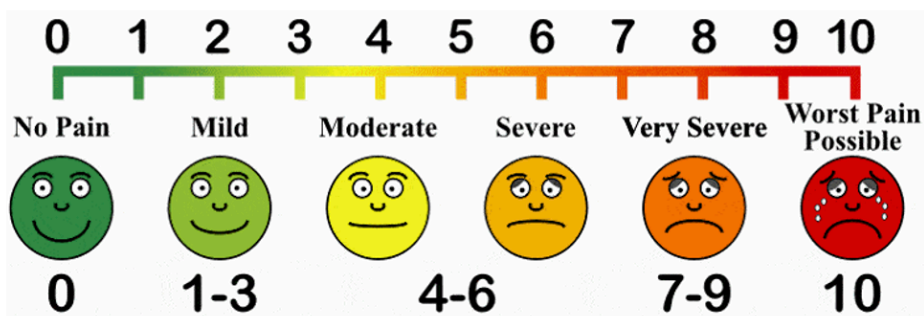
- What *would* happen?
- How should I intervene?
- Participation
- `???`

Causal Inference

Binary treatment Z

Z=0 Z=1

Observed outcome Y.



Potential Outcomes

Example:

Y(0)	Y(1)
6	3
5	5
7	9
9	6
4	2

Unit

- $\underline{Y_i(z)}$ is the pain level for person i given treatment level z.
- $Y(z)$ are called "potential outcomes".

Y(1) is pain w/ treat.
Y(0) is pain w/out treat.

Causal Inference

Example:

Y(0)	Y(1)
6	3
5	5
7	9
9	6
4	2

Treatment effects are contrasts between potential outcomes.

$$\tau_i^{ITE} = Y_i(1) - Y_i(0)$$

individual treatment effect.

$$\tau^{ATE} = E[Y_i(1) - Y_i(0)]$$

Average Treatment Effect.

Causal Inference

Example:

Y(0)	Y(1)
6	?
?	5
?	9
?	6
4	?
$\overline{Y(0)^{obs}}$	$\overline{Y(1)^{obs}}$

"Fundamental problem of causal inference" (Holland, 1986)

"Missing Data Problem."

Only one potential outcome is observed!

$$Y^{obs} = ZY(1) + (1 - Z)Y(0)$$

$$ATE \neq \overline{Y(1)}_{obs} - \overline{Y(0)}_{obs}$$

in general! (confounders!)

Observational causal inference

There are two distinct ways we can estimate average treatment effects from observational data, **assuming we measure a sufficient set of "control variables"**

confounders

- Inverse-Propensity Weighting (IPW) approach
- Regression Adjustment

IPW

Most common when Z, the treatment, is binary, $Z \in \{0, 1\}$.

Measure sufficient set of "control variables", X, to adjust for confounding.

$$\widehat{ATE}_{IPW} = \frac{1}{n} \sum^{n} \frac{Y_i(1) Z_i}{e(X_i)} - \frac{1}{n} \sum^{n} \frac{Y_i(0)(1 - Z_i)}{1 - e(X_i)}$$

$e(X) = Pr(Z = 1 \mid X)$, "Probability of being treated given X, confounders".

$$= \sum^n w_i Y_i(1) Z_i - \sum (1-w_i) Y_i(0)(1-Z_i)$$

$w_i = \frac{1}{e(X_i)}$ is my "weight"

$$\hat{ATE}_{IPW2} = \frac{\sum^n \frac{Y_i(1) Z_i}{e(X_i)}}{\sum^n 1/e(X_i)} - \frac{\sum^n \frac{Y_i(0)(1-Z_i)}{1-e(X_i)}}{\sum^n 1/(1-e(X_i))}$$

Different Perspective

Regression Adjustment.

Regression adjustment

Fork ("confounder")

$$\begin{tikzcd}
& X \arrow[dl] \arrow[dr] & \\
Z \arrow[rr, dotted] & & Y
\end{tikzcd}$$

Z: Ice Cream

Y: Drownings.

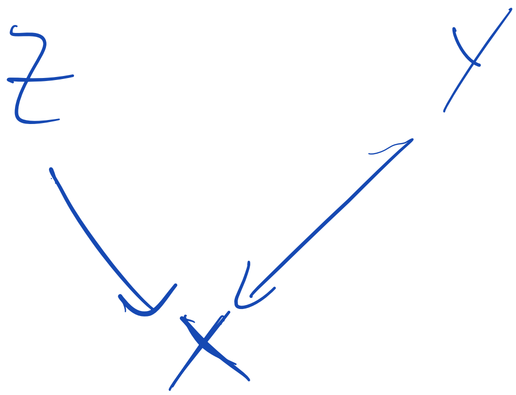
X: Temperatures

lm(Y ~ Z + X)

coef of Z should be close to 0.

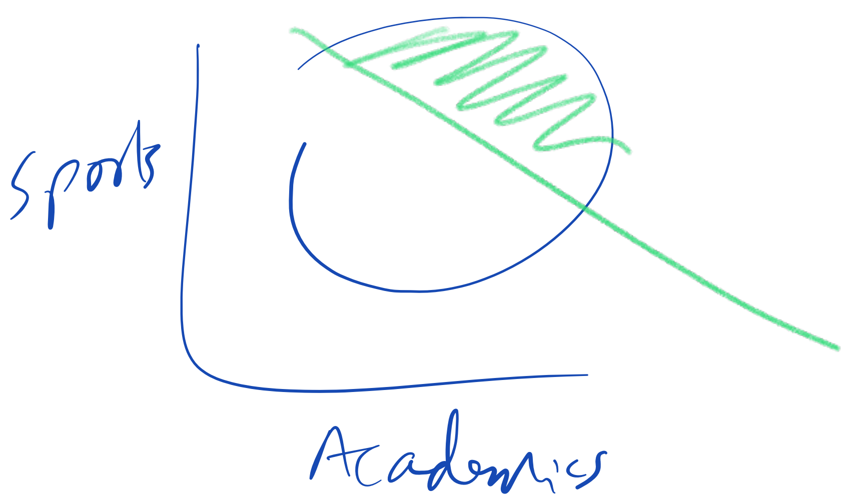
Collider

$lm(y \sim z)$ "no effect"



$lm(y \sim z + x)$

"spurious association between z and y"



Directed Acyclic Graphs

- We can try to represent all variables in our data with a directed acyclic graph (DAG)
- Variables are connected by arrows pointing in one direction
- Cycles are not allowed: cannot follow a path through the graph and return to your starting point.

Interventions

- A causal effect is a manipulation of the generative model, an intervention.
- $p(Y|do(Z))$ means the distribution of Y when we intervene ("do") Z
- This implies deleting all arrows into Z and "simulating" Y
- Interested in the *average causal effect* (ACE): $E[Y|do(Z)]$

Important graphical structures

- Forks: $Z \leftarrow X \rightarrow Y$
- Colliders: $Z \rightarrow X \leftarrow Y$
- Pipes: $Z \rightarrow X \rightarrow Y$

Variables in a graph.

$Z \longrightarrow Y$

Z "causes" Y.

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Z & & Y \end{tikzcd}$$

"confounder"

X causes both Z & Y.

$$\begin{tikzcd} Z \arrow[dr] & & Y \arrow[dl] \\ & X & \end{tikzcd}$$

Z and Y cause X.

Forks

$$Z \leftarrow X \rightarrow Y$$

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Z & & Y \end{tikzcd}$$

no arrow from Z to Y.

- We say X is a "common cause" of Z and Y.
- Z and Y are associated, that is $Z \not\perp\!\!\!\perp Y$ "not independent"
- Once "stratified" by X, $Z \perp\!\!\!\perp Y$

X is a "common cause"

$$\begin{tikzcd} & X \text{ (weather)} \arrow[dl] \arrow[dr] & \\ \text{(ice cream) } Z & & Y \text{ (drowning)} \end{tikzcd}$$

Forks - Simple Example

```r
rbern <- \(n, prob=0.5) rbinom(n, size=1, prob=prob)
n <- 1000
X <- rbern(n)
Z <- rbern(n, (1-X)*0.1 + X*0.9)
Y <- rbern(n , (1-X)*0.1 + X*0.9 )

table(Z, Y)
```

coin flip model.

$$\Pr(Z=1 \mid X) = \begin{cases} 0.1 & \text{if } X=0 \\ 0.9 & \text{if } X=1 \end{cases}$$

```
   Y
Z     0   1
  0 418  95
  1  82 405
```

```r
cor(Z, Y)
```

```
[1] 0.6462185
```

$$\Pr(Y=1 \mid X) = \quad //$$

Forks - Simple Example

- If we stratify on Z then X and Y are no longer dependent

```r
table(Z, Y, X)
```

```
, , X = 0

   Y
Z     0   1
  0 412  46
  1  39   4

, , X = 1

   Y
Z     0   1
  0   6  49
  1  43 401
```

controlled for X by stratifying ("subsetting")

```r
cor(Z[X==0], Y[X==0])
```

```
[1] -0.006928076
```

```r
cor(Z[X==1], Y[X==1])
```

```
[1] 0.01288529
```

Forks

X is not just 0 or 1

- For continuous confounders, use models to "stratify"
- In general, learn a different functional relationship between Y and Z for all possible x.
  - e.g. $Y = f_x(Z) + \epsilon$.
- Simple modeling assumption: $Y = \beta_0 + \underbrace{\beta_1 X}_{\text{controlling for } X.} + \beta_2 Z + \epsilon$

Forks - Simple Example

```
library(rethinking)
data("WaffleDivorce")

coef(summary(lm(Divorce ~ Marriage, data=WaffleDivorce)))
```

```
             Estimate Std. Error  t value     Pr(>|t|)
(Intercept) 6.0840440 1.31337263 4.632382 2.784458e-05
Marriage    0.1791765 0.06418477 2.791573 7.506903e-03
```

Z : marriage rate

Y : divorce rate

higher marriage rate ⇒ higher divorce rate?

Forks - Simple Example

```
coef(summary(lm(Divorce ~ Marriage + MedianAgeMarriage, data=WaffleDivorce)
```

```
                    Estimate Std. Error    t value     Pr(>|t|)
(Intercept)       36.8766478 7.66104269  4.8135286 1.575629e-05
Marriage          -0.0568648 0.08053035 -0.7061288 4.835938e-01
MedianAgeMarriage -0.9996495 0.24593051 -4.0647642 1.816080e-04
```

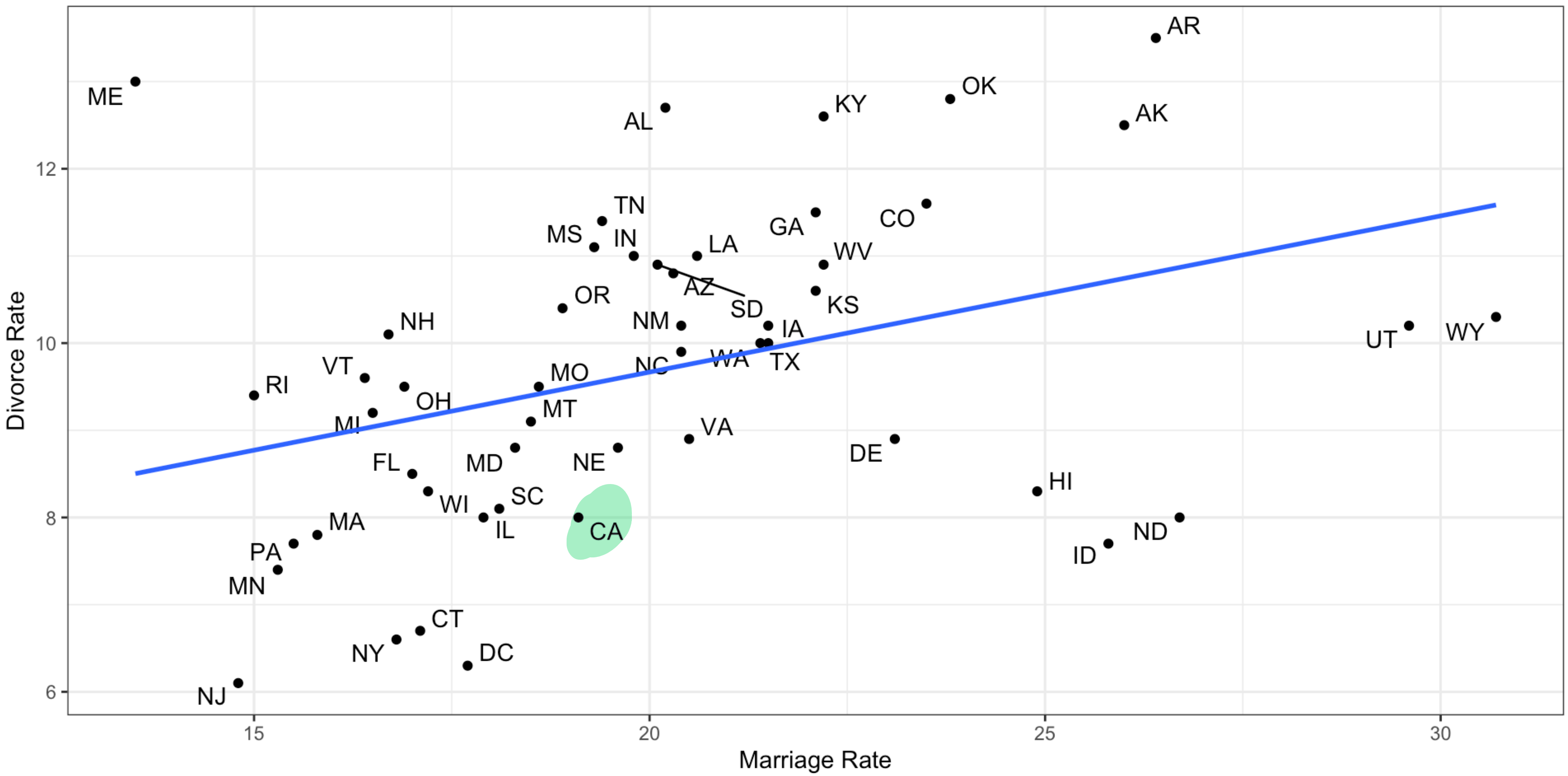
No longer significant.

significant association between Age at Marriage & Divorce.

Forks - Simple Example

```r
WaffleDivorce %>%
  ggplot(aes(x=Marriage, y=Divorce)) +
  geom_point() +
  ggrepel::geom_text_repel(aes(label=Loc)) + theme_bw() +
  geom_smooth(method="lm", se=FALSE) +
  xlab("Marriage Rate") +
  ylab("Divorce Rate")
```

Forks - Simple Example



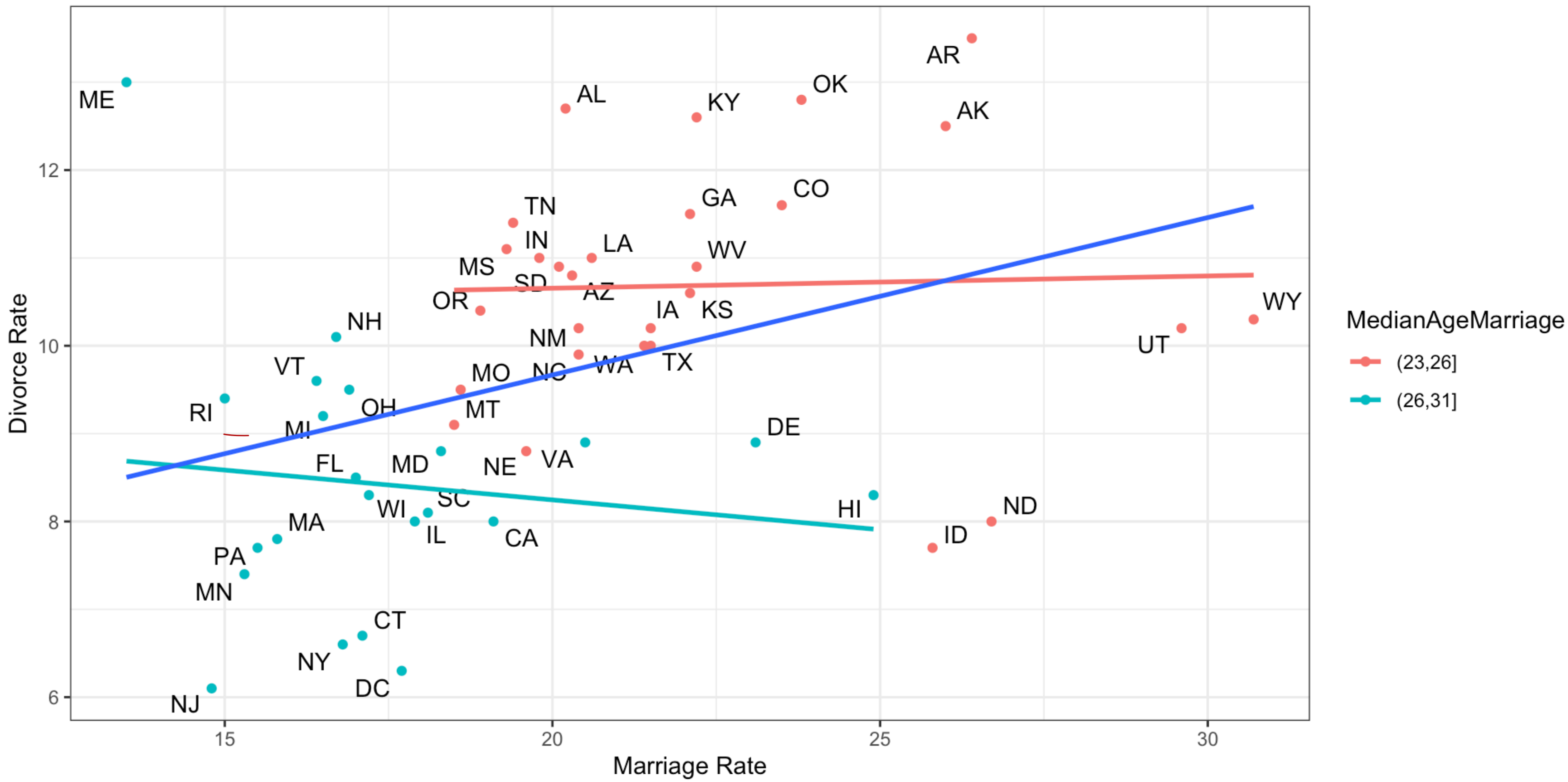
Forks - Simple Example

Divide age into 3 categories.

```
WaffleDivorce %>%
  mutate(MedianAgeMarriage = cut(MedianAgeMarriage, breaks=c(23, 26, 31))) %>%
  ggplot(aes(x=Marriage, y=Divorce)) +
  geom_point(aes(col=MedianAgeMarriage)) +
  ggrepel::geom_text_repel(aes(label=Loc)) + theme_bw() +
  geom_smooth(aes(col=MedianAgeMarriage), method="lm", se=FALSE) +
  geom_smooth(, method="lm", se=FALSE) +
  xlab("Marriage Rate") +
  ylab("Divorce Rate")
```

Divorce ~ marriage + Age

Forks - Simple Example



A different question

> **Note**
>
> What if instead, we want to estimate the causal effect of age on divorce rate? Do we include marriage rate as a control?

Pipes

"direct effect"

$$Z \to X \to Y$$

" indirect effect"

- Assume Z and Y are associated
- The influence of Z on Y is transmitted through X
- Given X, Z and Y might be independent

lm(y ~ x + z) or lm(y ~ z)?

coef of 0 for z does not mean z doesn't cause y.

total effect: direct + indirect,

Pipes

$Z \to X \to Y$

```r
n <- 1000
Z <- rbern(n)
X <- rbern(n, (1-Z)*0.1 + Z*0.9)
Y <- rbern(n , (1-X)*0.1 + X*0.9)

table(Z, Y)
```

```
   Y
Z     0   1
  0 414  86
  1  98 402
```

```r
cor(Z, Y)
```

```
[1] 0.6321821
```

Pipes

```
1  table(Z, Y, X)
```

```
, , X = 0

   Y
Z     0   1
  0 411  34
  1  44   5

, , X = 1

   Y
Z     0   1
  0   3  52
  1  54 397
```

```
1  cor.test(Z[X==0], Y[X==0])
```

```

	Pearson's product-moment correlation

data:  Z[X == 0] and Y[X == 0]
t = 0.6306, df = 492, p-value = 0.5286
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
 -0.05995414  0.11634843
sample estimates:
       cor
0.02841815
```

$$Z \rightarrow X \leftarrow Y$$

"X is a collider"

Controlling for X is bad!

Colliders

- Colliders: $Z \to X \leftarrow Y$
- Assume Z and Y are independent
- But Z and Y both influence X ("collide" at X)
- Once stratified by ~~Z~~, ~~X~~ and Y associated

condition

X, Z

$$Z \perp\!\!\!\perp Y$$

$$Z \not\perp\!\!\!\perp Y \text{ given } X.$$

Colliders

```
n <- 1000
Z <- rbern( n , 0.5 )
Y <- rbern( n , 0.5 )
X <- rbern( n , ifelse(Z+Y>0, 0.9, 0.2) )

cor(Z, Y)
```

Coin Flips

$$p(X=1) \begin{cases} .9 & Z+Y>0 \\ .2 & \text{else} \end{cases}$$

```
[1] 0.02810281
```

correct causal result: no cause.

```
cor(Z[X==0], Y[X==0])
```

```
[1] 0.4032796
```

```
cor(Z[X==1], Y[X==1])
```

```
[1] -0.2951158
```

incorrect causal Result!

Significant correlations within subgroups.

Colliders

```r
set.seed(100)
n <- 1000
Z <- rnorm(n)
Y <- rnorm(n)
X <- rbern(n , ifelse(Z+Y > 0, 0.9, 0.2) )

summary(lm(Y ~ Z))$coefficients
```

```
               Estimate Std. Error   t value  Pr(>|t|)
(Intercept) 0.003921337 0.03103895 0.1263360 0.8994914
Z           0.013414659 0.03012877 0.4452441 0.6562399
```

No association.

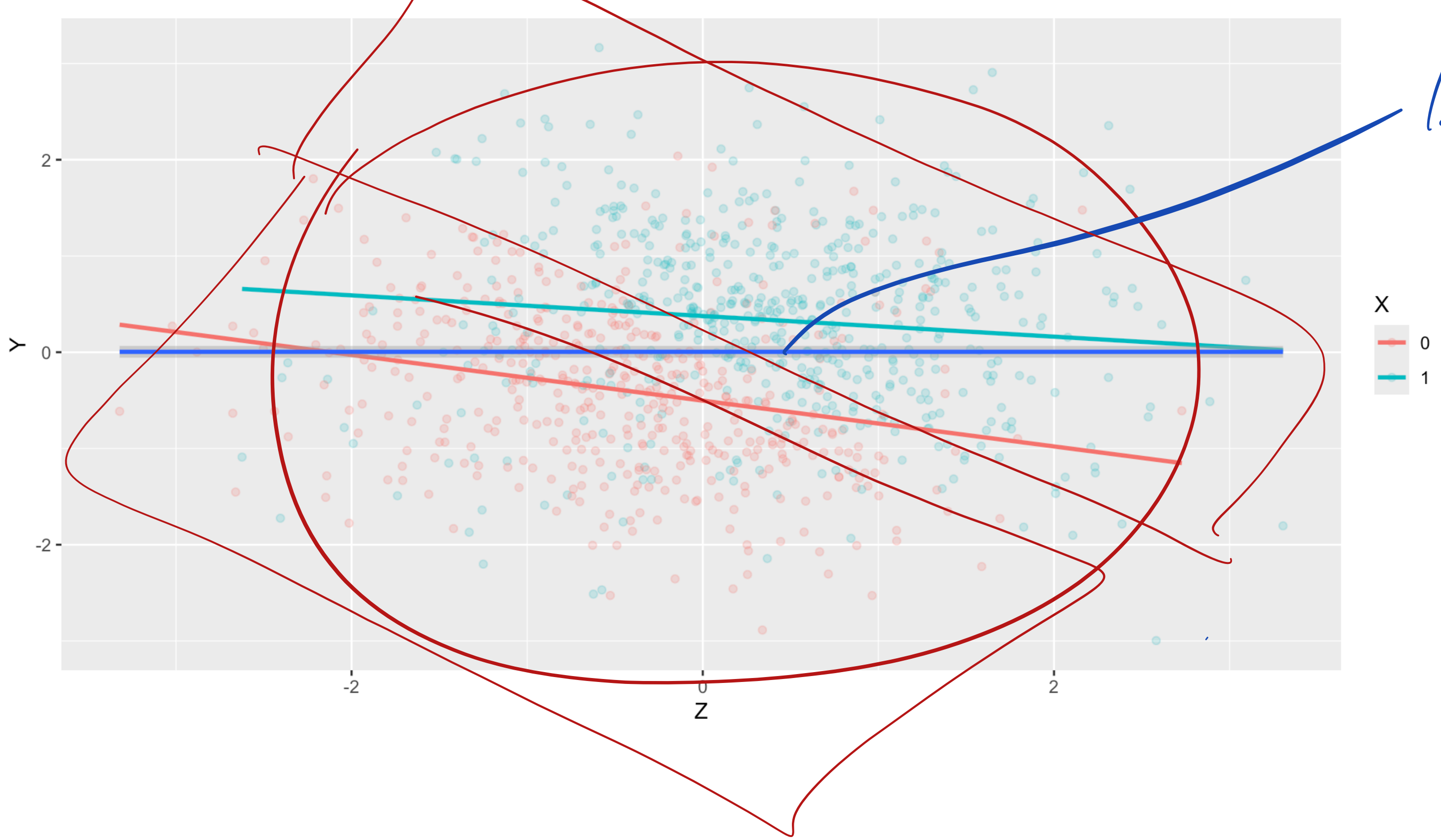
```r
summary(lm(Y ~ Z + X))$coefficients
```

```
              Estimate Std. Error    t value     Pr(>|t|)
(Intercept) -0.4692849 0.04462326 -10.516598 1.329936e-24
Z           -0.1640961 0.03048996  -5.381971 9.185200e-08
X            0.8689585 0.06310520  13.769998 1.273281e-39
```

Significant association. Bad!

Colliders

```
tibble(Z=Z, Y=Y, X=factor(X)) |>
  ggplot(aes(x=Z, y=Y)) +
  geom_point(alpha=0.2, aes(col=X)) +
  geom_smooth(method="lm", se=FALSE, aes(col=X)) +
  geom_smooth()
```



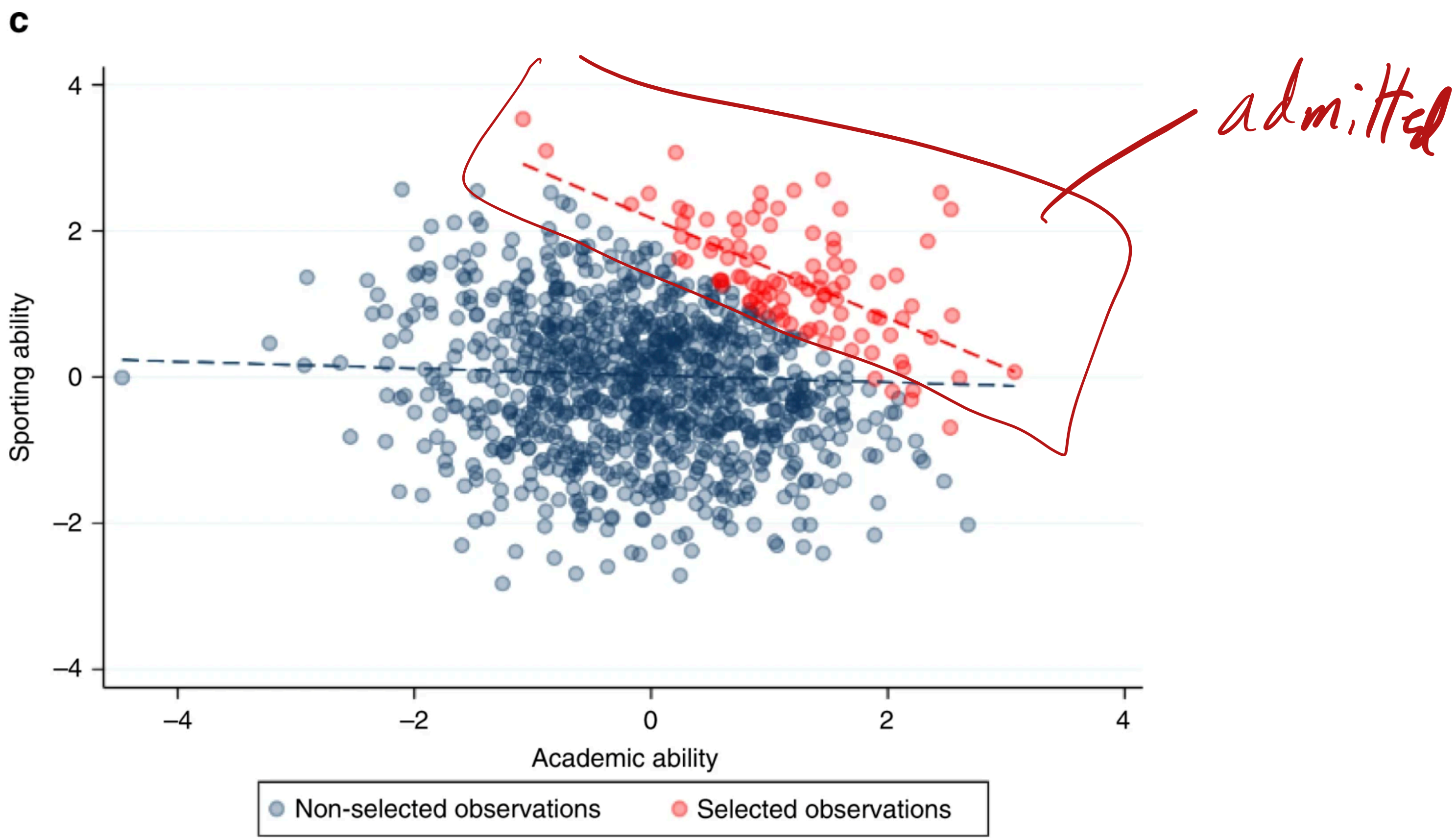
lm(y~Z)

⇒ 0

association.

Collider Example

Admissions at elite colleges.



Confounding bias

- Observed data regression of T on Y fails because the distribution of U varies in the two treatment arms

- We try to condition on as many *observed* confounders as possible to mitigate potential confounding bias

- Commonly assumed that there are “no unobserved confounders” (NUC) but this is unverifiable

Confounding bias

- Observed data regression of T on Y fails because the distribution of U varies in the two treatment arms

- We try to condition on as many *observed* confounders as possible to mitigate potential confounding bias

- Commonly assumed that there are “no unobserved confounders” (NUC) but this is unverifiable

- Sensitivity analysis is a tool for assessing the impacts of violations of this assumption

A Motivating Example

HEALTH > NUTRITION & DIET

7 Science-Backed Health Benefits of Drinking Red Wine

Yep, moderate red wine consumption is healthy—and here's the proof.

By Ashley Zlatopolsky | Updated on November 5, 2022

Fact checked by Emily Peterson

A Motivating Example

The New York Times

Even a Little Alcohol Can Harm Your Health

Recent research makes it clear that any amount of drinking can be detrimental. Here's why you may want to cut down on your consumption beyond Dry January.

The Effects of Light Alcohol Consumption

- Observational data from the National Health and Nutrition Examination Study (NHANES) on alcohol consumption.
- Light alcohol consumption is positively correlated with blood levels of HDL ("good cholesterol")

glass of wine
per night

The Effects of Light Alcohol Consumption

- Observational data from the National Health and Nutrition Examination Study (NHANES) on alcohol consumption.
- Light alcohol consumption is positively correlated with blood levels of HDL (“good cholesterol”)

HDL and alcohol consumption

```r
summary(lm(Y[, "HDL"] ~ drinking + X))
```

1 glass a night or not

confounder ("forks")

- Age
- Education.

What am I missing?

```
Call:
lm(formula = Y[, "HDL"] ~ drinking + X)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0855 -0.6127 -0.0512  0.6389  4.2383 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.225550   0.091105   2.476 0.013412 *  
drinking    0.597399   0.091917   6.499 1.11e-10 ***
Xage        0.006409   0.001452   4.415 1.09e-05 ***
Xgender     0.689557   0.049426  13.951  < 2e-16 ***
```

HDL and alcohol consumption

```r
summary(lm(Y[, "HDL"] ~ drinking + X))
```

```
Call:
lm(formula = Y[, "HDL"] ~ drinking + X)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0855 -0.6127 -0.0512  0.6389  4.2383

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.225550   0.091105   2.476 0.013412 *
drinking    0.597399   0.091917   6.499 1.11e-10 ***
Xage        0.006409   0.001452   4.415 1.09e-05 ***
Xgender     0.689557   0.049426  13.951  < 2e-16 ***
```

Didn't measure important confounders.

What must be true for this correlation to be non-causal?

Blood mercury and alcohol consumption

```
1 summary(lm(Y[, "Methylmercury"] ~ drinking + X))
```

```
Call:
lm(formula = Y[, "Methylmercury"] ~ drinking + X)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.3570 -0.7363 -0.0728  0.6242  4.1127 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.442044   0.096385   4.586 4.91e-06 ***
drinking     0.364096   0.097244   3.744 0.000188 ***
Xage         0.008186   0.001536   5.330 1.14e-07 ***
Xgender     -0.062664   0.052290  -1.198 0.230966    
```

But: we know alcohol can't increase mercury levels!

Strong association.

Blood mercury and alcohol consumption

```r
summary(lm(Y[, "Methylmercury"] ~ drinking + X))
```

```
Call:
lm(formula = Y[, "Methylmercury"] ~ drinking + X)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.3570 -0.7363 -0.0728  0.6242  4.1127 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.442044   0.096385   4.586 4.91e-06 ***
drinking     0.364096   0.097244   3.744 0.000188 ***
Xage         0.008186   0.001536   5.330 1.14e-07 ***
Xgender     -0.062664   0.052290  -1.198 0.230966    
```

Residual Correlation

```
hdl_fit <- lm(Y[, "HDL"] ~ drinking + X)
mercury_fit <- lm(Y[, "Methylmercury"] ~ drinking + X)

cor.test(hdl_fit$residuals, mercury_fit$residuals)
```

Residual Correlation

```r
hdl_fit <- lm(Y[, "HDL"] ~ drinking + X)
mercury_fit <- lm(Y[, "Methylmercury"] ~ drinking + X)

cor.test(hdl_fit$residuals, mercury_fit$residuals)
```

Residual Correlation

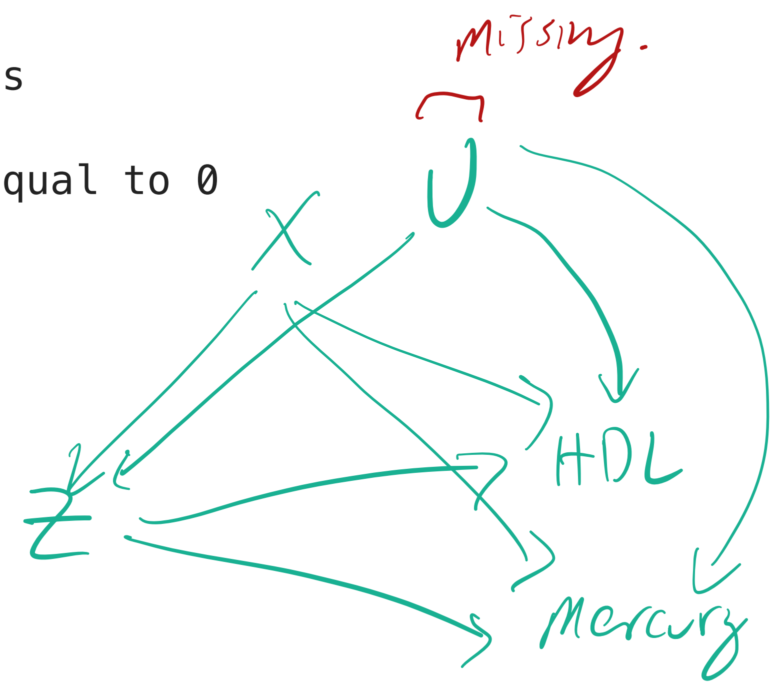
```r
hdl_fit <- lm(Y[, "HDL"] ~ drinking + X)
mercury_fit <- lm(Y[, "Methylmercury"] ~ drinking + X)

cor.test(hdl_fit$residuals, mercury_fit$residuals)
```

```
	Pearson's product-moment correlation

data:  hdl_fit$residuals and mercury_fit$residuals
t = 3.7569, df = 1437, p-value = 0.0001789
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.04718758 0.14953581
sample estimates:
      cor 
0.0986225 
```

Significant
But small
correlation.



Residual Correlation

```r
hdl_fit <- lm(Y[, "HDL"] ~ drinking + X)
mercury_fit <- lm(Y[, "Methylmercury"] ~ drinking + X)

cor.test(hdl_fit$residuals, mercury_fit$residuals)
```

```
    Pearson's product-moment correlation

data:  hdl_fit$residuals and mercury_fit$residuals
t = 3.7569, df = 1437, p-value = 0.0001789
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.04718758 0.14953581
sample estimates:
      cor
0.0986225
```

Residual correlation might be indicative of confounding bias

Sensitivity Analysis

- NUC unlikely to hold exactly. What then?
- Calibrate assumptions about confounding to explore range of causal effects that are plausible
- Robustness: quantify how “strong” confounding has to be to nullify causal effect estimates

Sensitivity Analysis

- NUC unlikely to hold exactly. What then?
- Calibrate assumptions about confounding to explore range of causal effects that are plausible
- Robustness: quantify how “strong” confounding has to be to nullify causal effect estimates
- Well established methods for single outcome analyses

Multi-outcome Sensitivity Analysis

- If we measure multiple outcomes, is there prior knowledge that we can leverage to strengthen causal conclusions?

Multi-outcome Sensitivity Analysis

- If we measure multiple outcomes, is there prior knowledge that we can leverage to strengthen causal conclusions?

- What might residual correlation in multi-outcome models mean for potential for confounding?

Multi-outcome Sensitivity Analysis

- If we measure multiple outcomes, is there prior knowledge that we can leverage to strengthen causal conclusions?

- What might residual correlation in multi-outcome models mean for potential for confounding?

- How do results change when we assume a priori that certain outcomes cannot be affected by treatments?
  - Null control outcomes (e.g. alcohol consumption should not increase mercury levels)