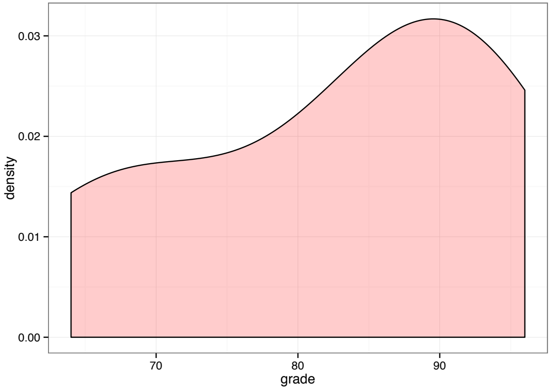
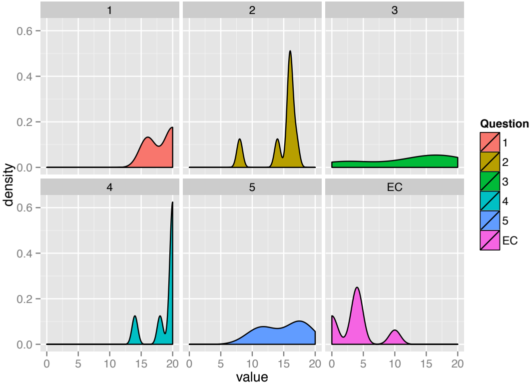
You all Did Fine

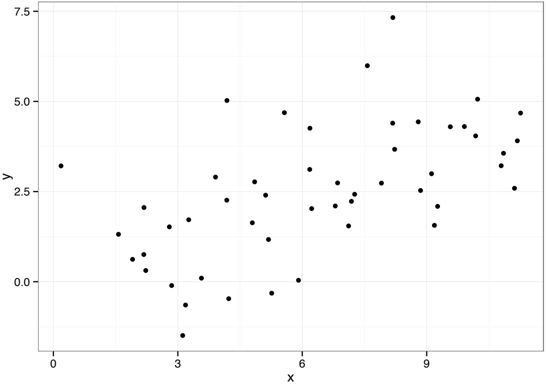


Main Issues were Power & Confidence Intervals of Non-Normal Regression

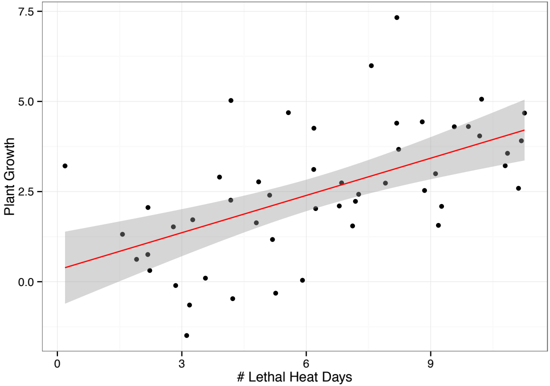


Observational Study Design

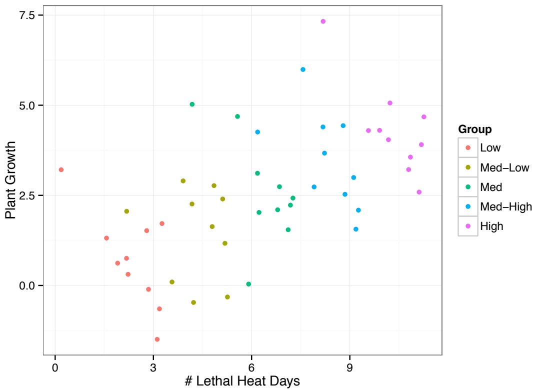
Problem: What if An Observed Relationship Doesn't Make Sense?



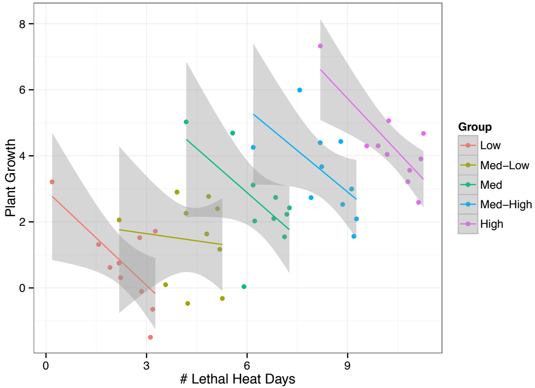
Problem: What if An Observed Relationship Doesn't Make Sense



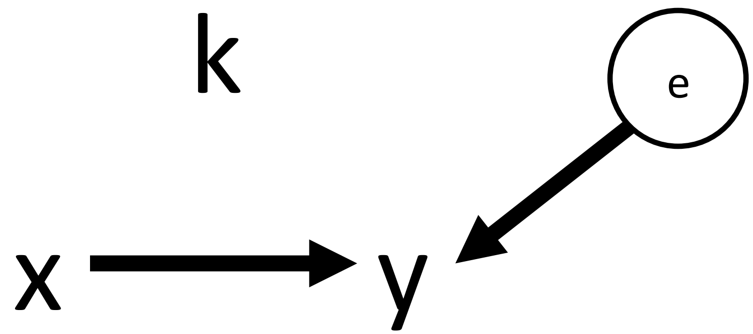
Covariates can Change Results



Simpson's Paradox



How will including k change B_{xy}?



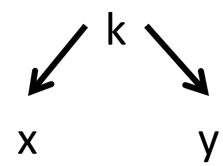
$$y = a + B_{xy}x + e$$

k

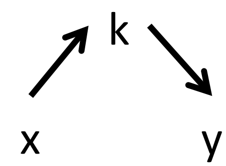
$x \longrightarrow y$

B_{xy} not changed by k

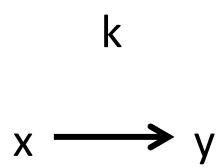
Full Explanation



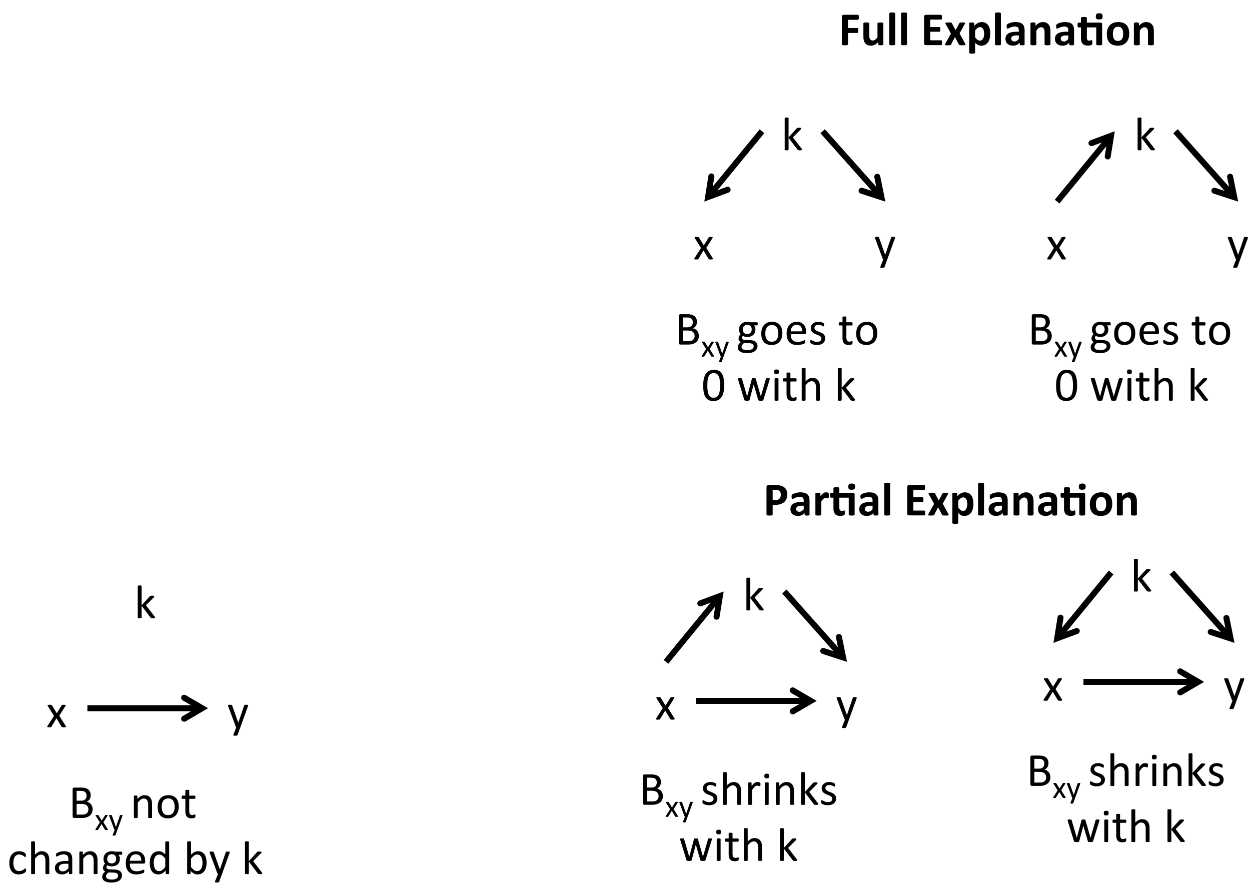
B_{xy} goes to 0 with k

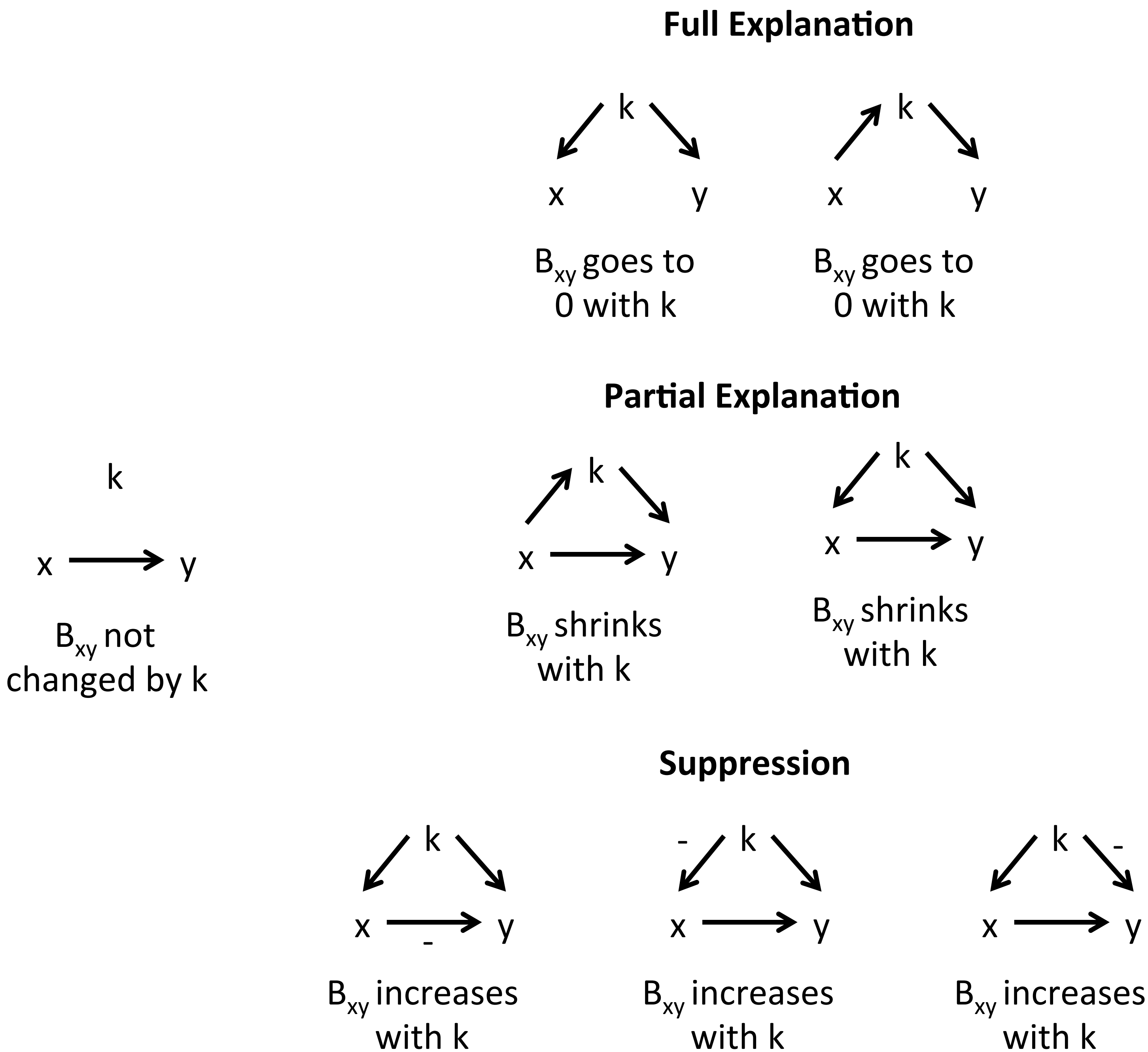


B_{xy} goes to 0 with k

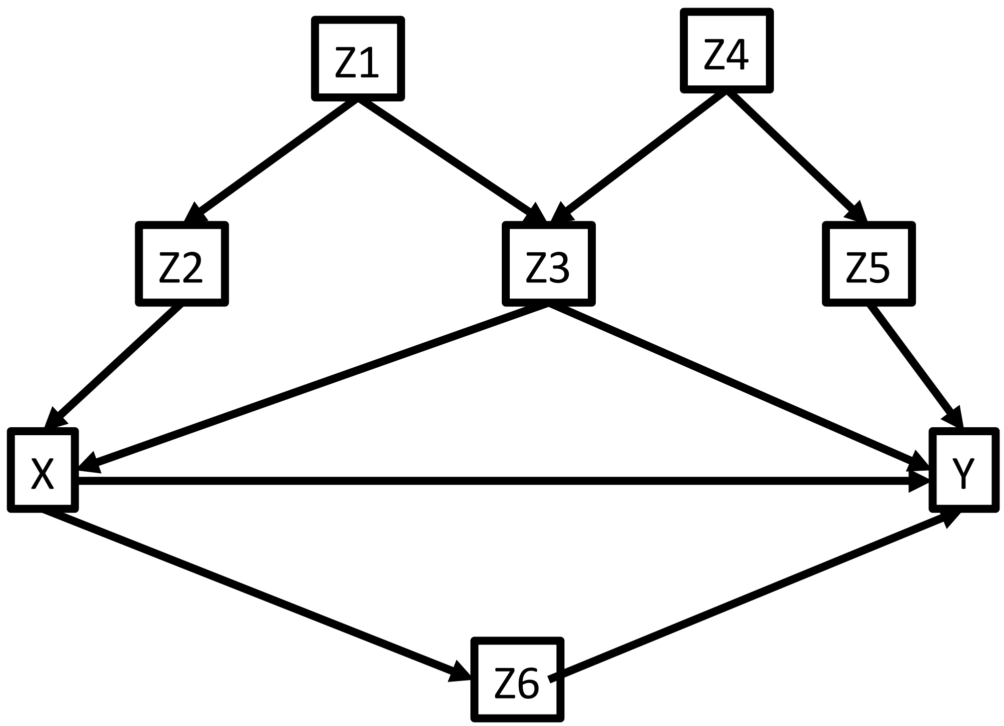


B_{xy} not changed by k


Full Explanation
k
x
y
B_{xy} goes to 0 with k
k
x
y
B_{xy} goes to 0 with k
Partial Explanation
k
x
y
B_{xy} not changed by k
k
x
y
B_{xy} shrinks with k
k
x
y
B_{xy} shrinks with k

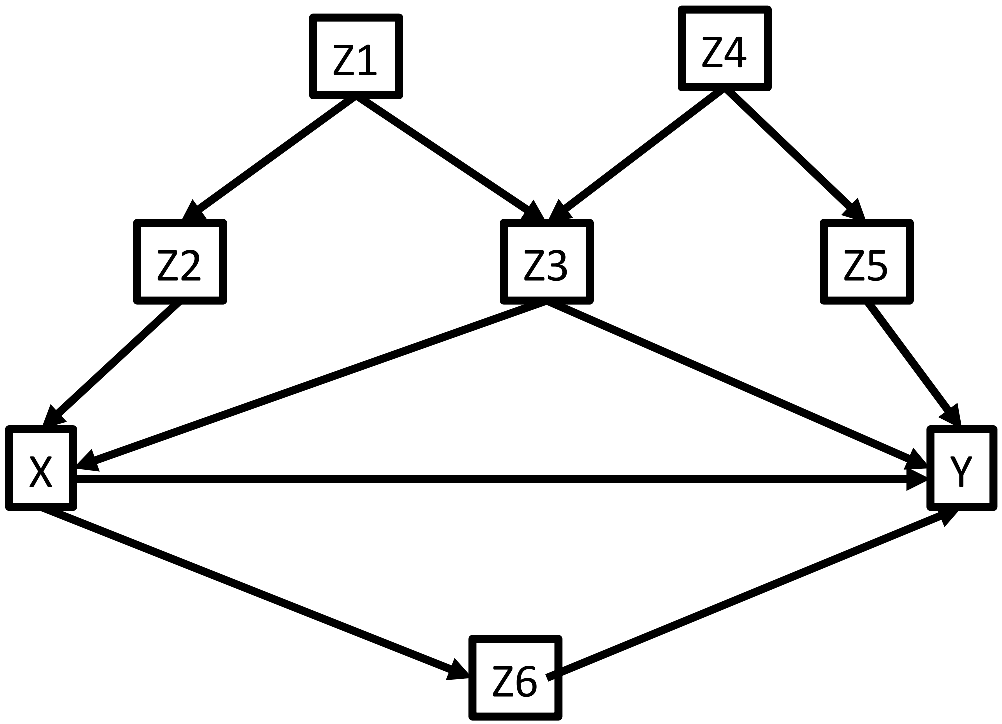

Full Explanation
k
x
y
B xy goes to 0 with k
k
x
y
B xy goes to 0 with k
Partial Explanation
k
x
y
B xy not changed by k
k
x
y
B xy shrinks with k
k
x
y
B xy shrinks with k
Suppression
k
x
y
-
B xy increases with k
-
k
x
y
B xy increases with k
k
-
x
y
B xy increases with k

What do we Control For?



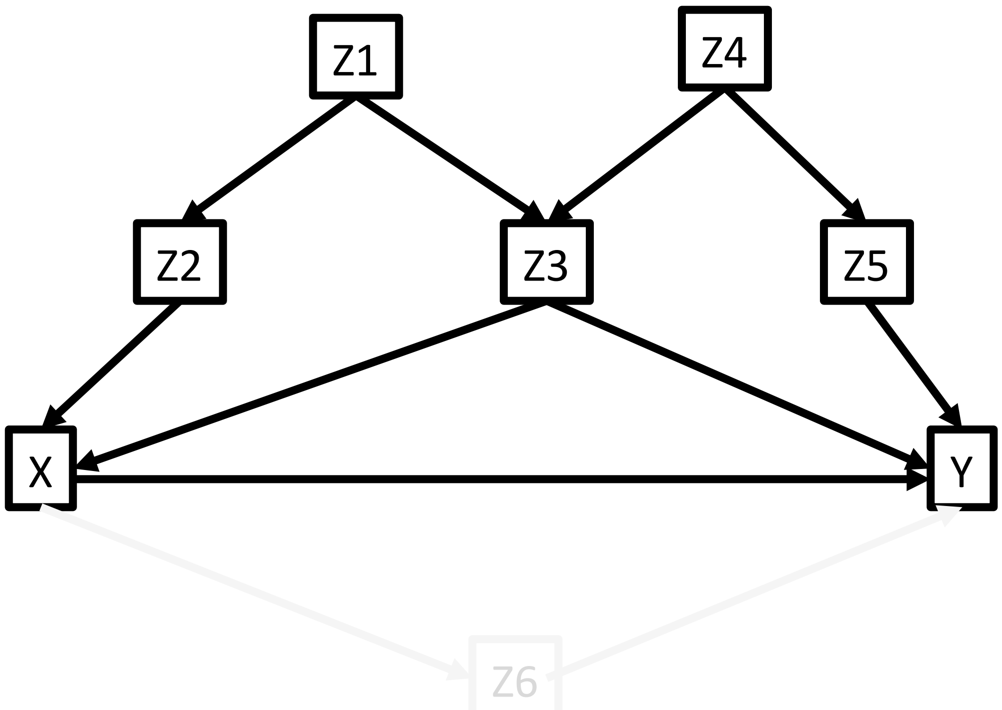
What do we Control For?

1) No node in our control set is a descendant of X.



What do we Control For?

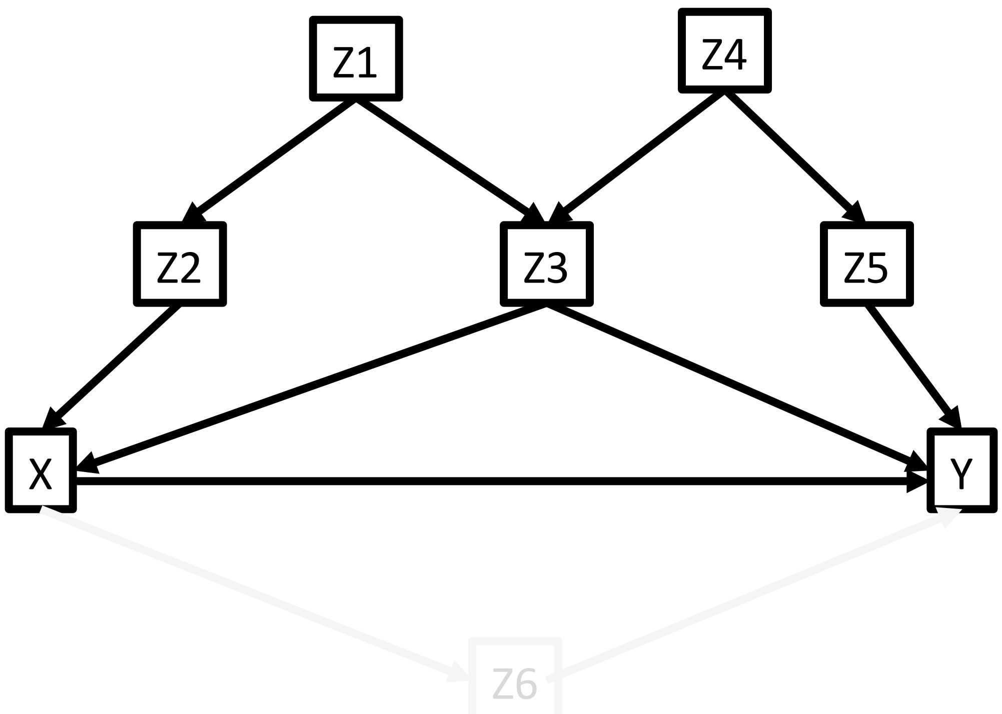
1) No node in our control set
is a descendant of X.



What do we Control For?

1) No node in our control set is a descendant of X.

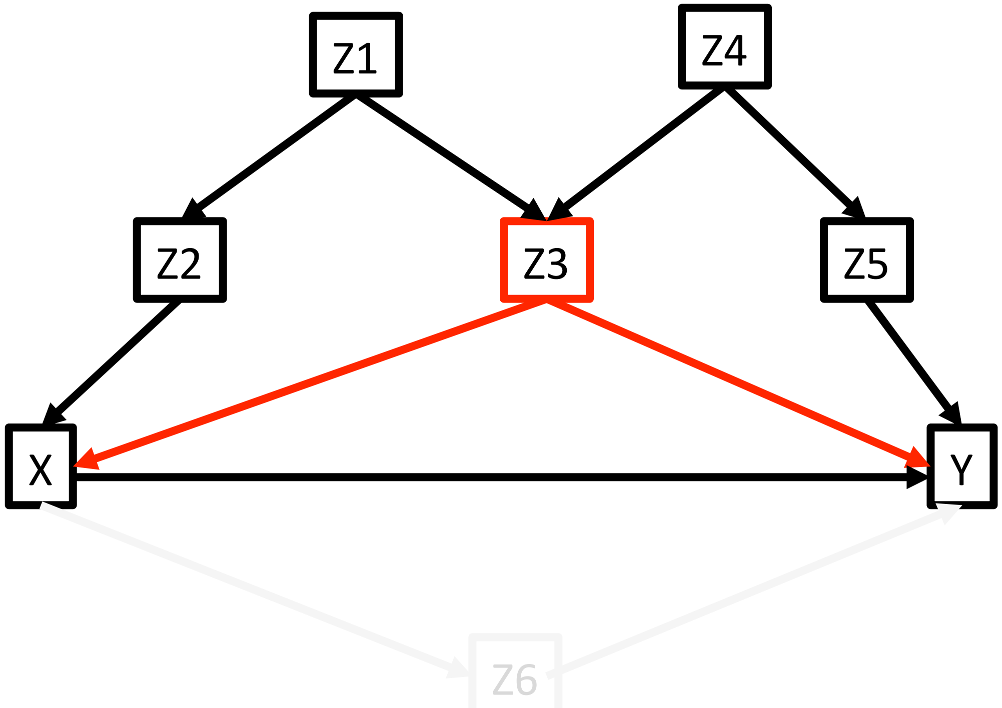
2) Z blocks every path between X and Y that contains an arrow into X.



Back-Door Criterion *sensu* Pearl

1) No node in our control set is a descendant of X.

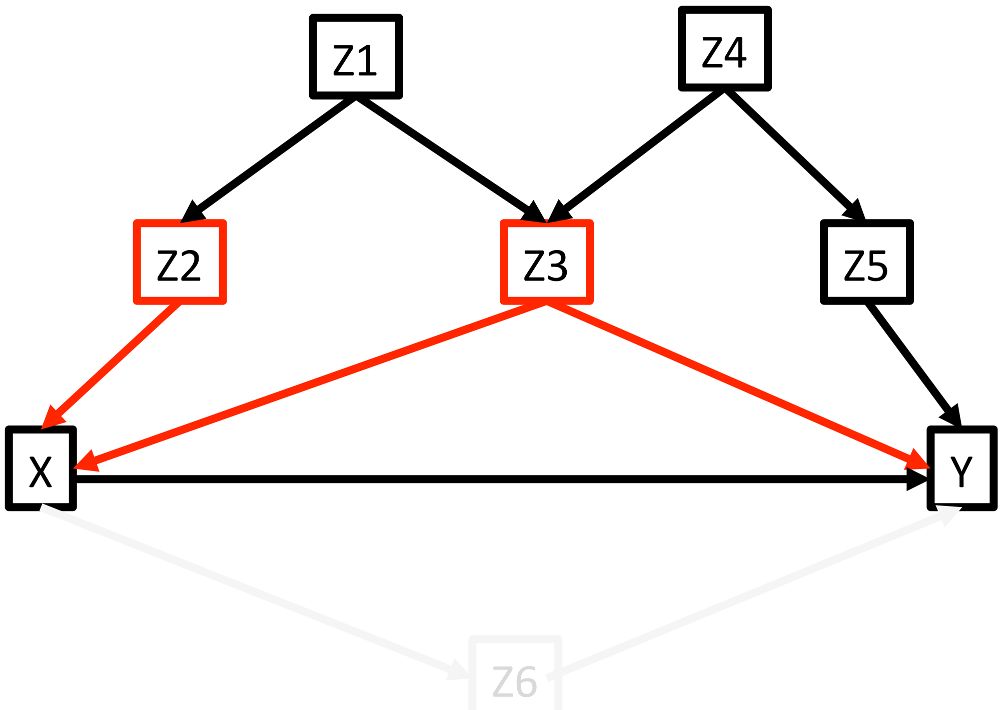
2) Z blocks every path between X and Y that contains an arrow into X.



Back-Door Criterion *sensu* Pearl

1) No node in our control set is a descendant of X.

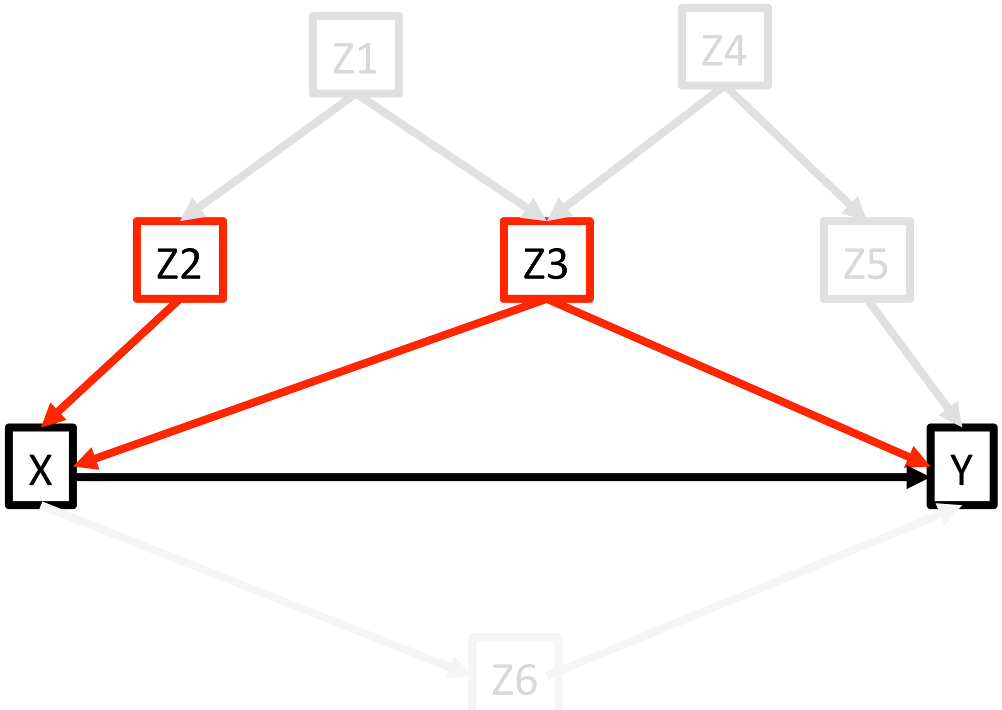
2) Z blocks every path between X and Y that contains an arrow into X.



Back-Door Criterion *sensu* Pearl

1) No node in our control set is a descendant of X.

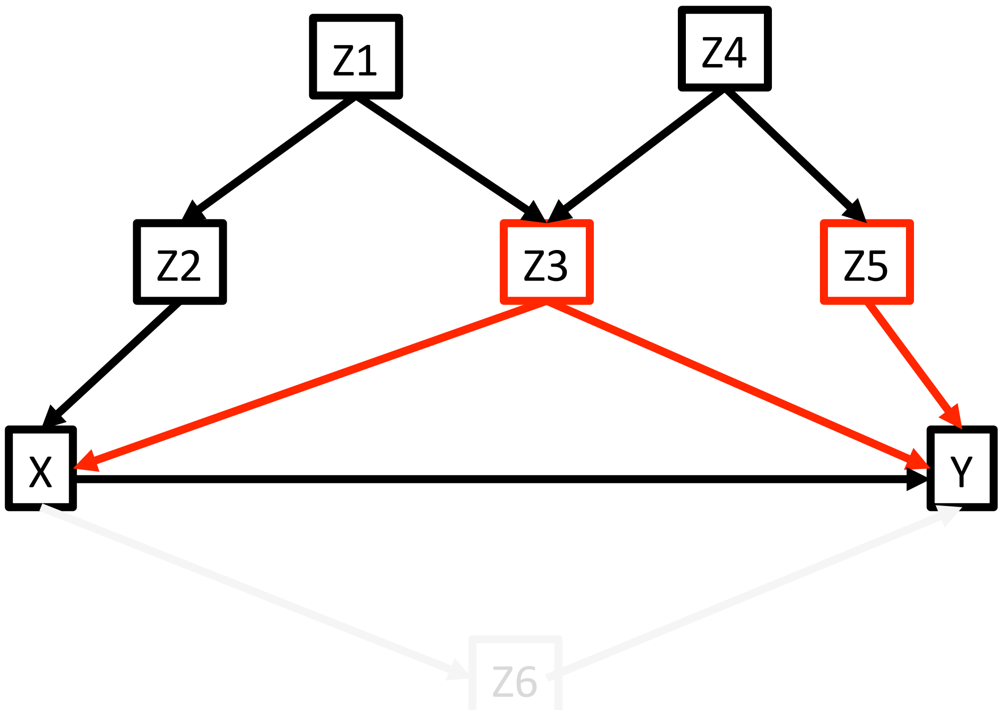
2) Z blocks every path between X and Y that contains an arrow into X.



Back-Door Criterion *sensu* Pearl

1) No node in our control set is a descendant of X.

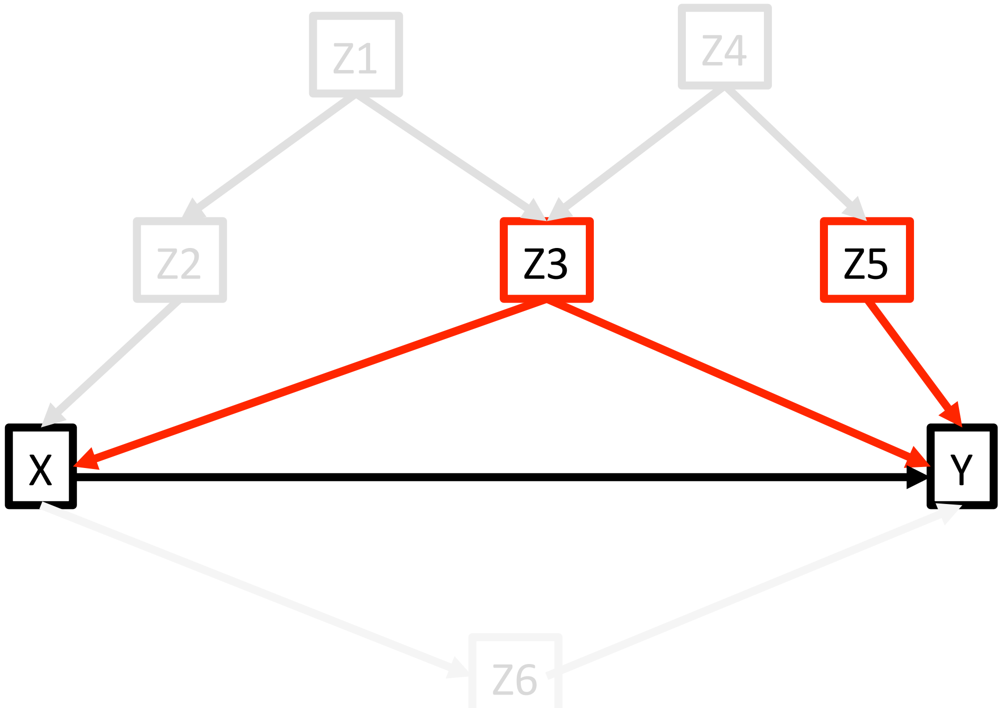
2) Z blocks every path between X and Y that contains an arrow into X.



Back-Door Criterion *sensu* Pearl

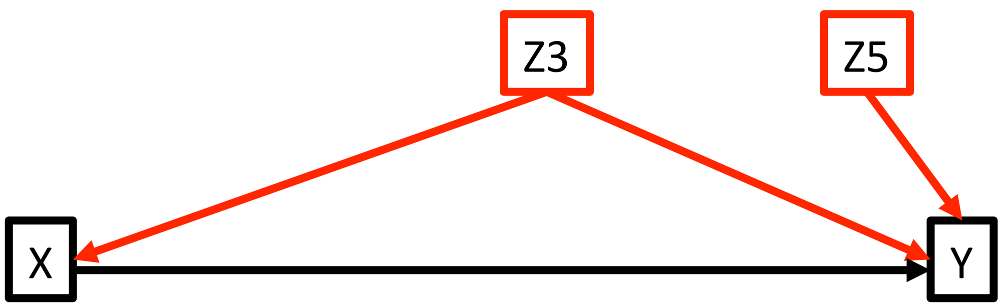
1) No node in our control set is a descendant of X.

2) Z blocks every path between X and Y that contains an arrow into X.



How Do We Account for Covariates?

1. Include control variables, but be exercise care with interpretation

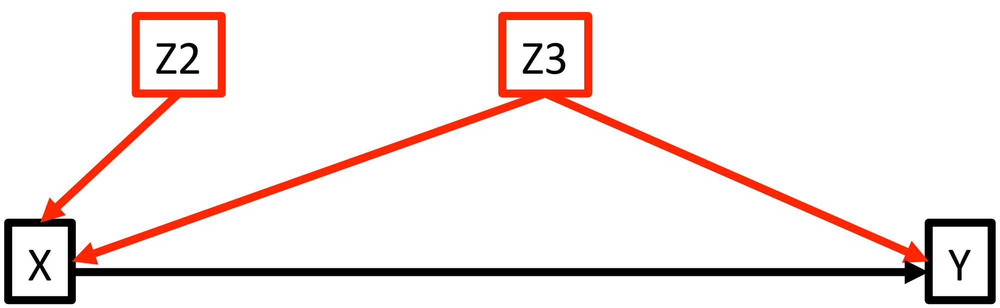


Z3 -> X Path is the covariance between them, accounted for in calculation of coefficients in multiple linear regression

How Do We Account for Covariates?

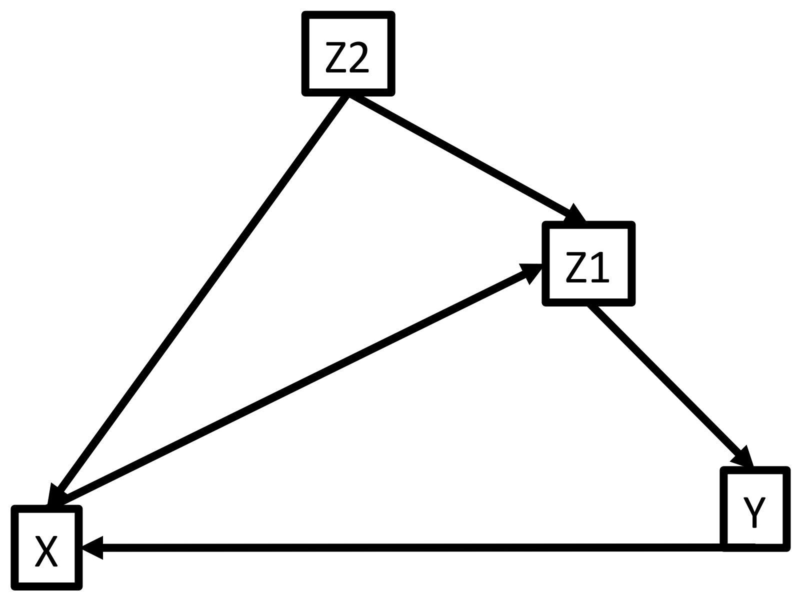
1. Include control variables, but be exercise care with interpretation

2. Take residuals of predictor with respect to relevant variables in control set

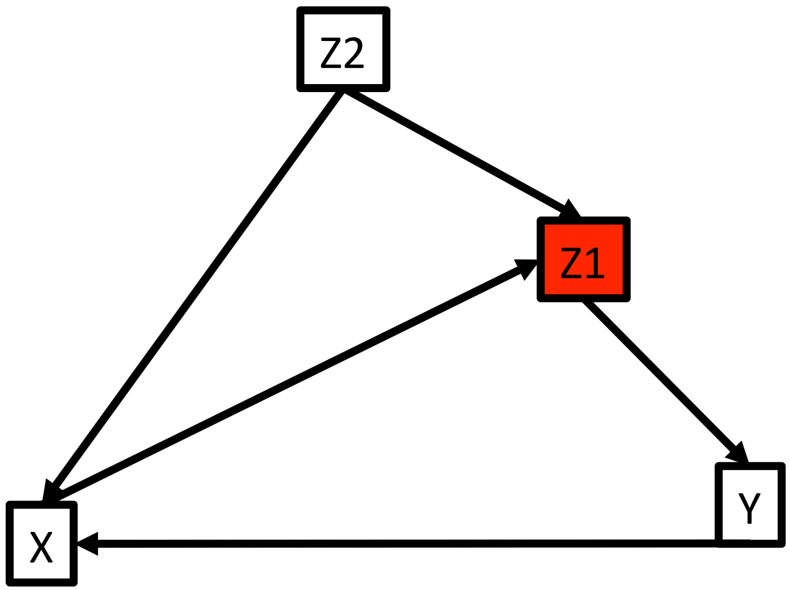


Residuals with respect to Z2 may be helpful

You Try…

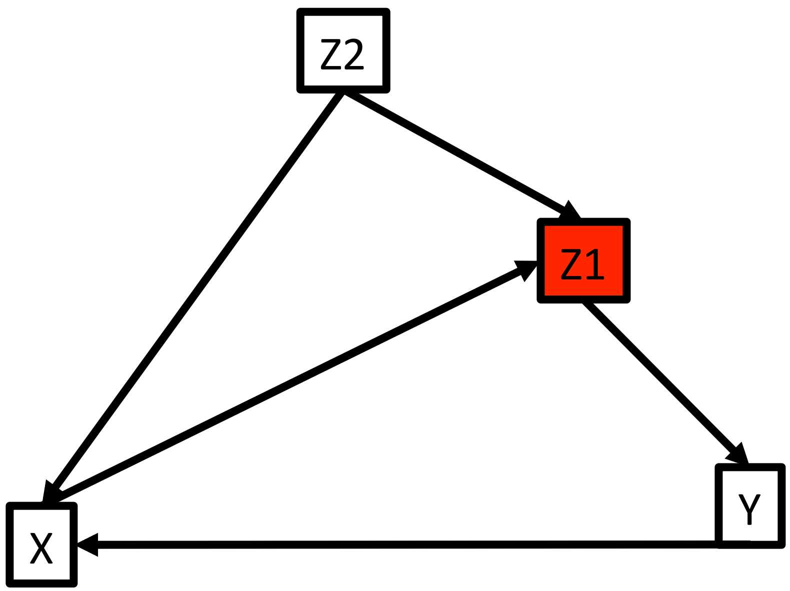


You Try...



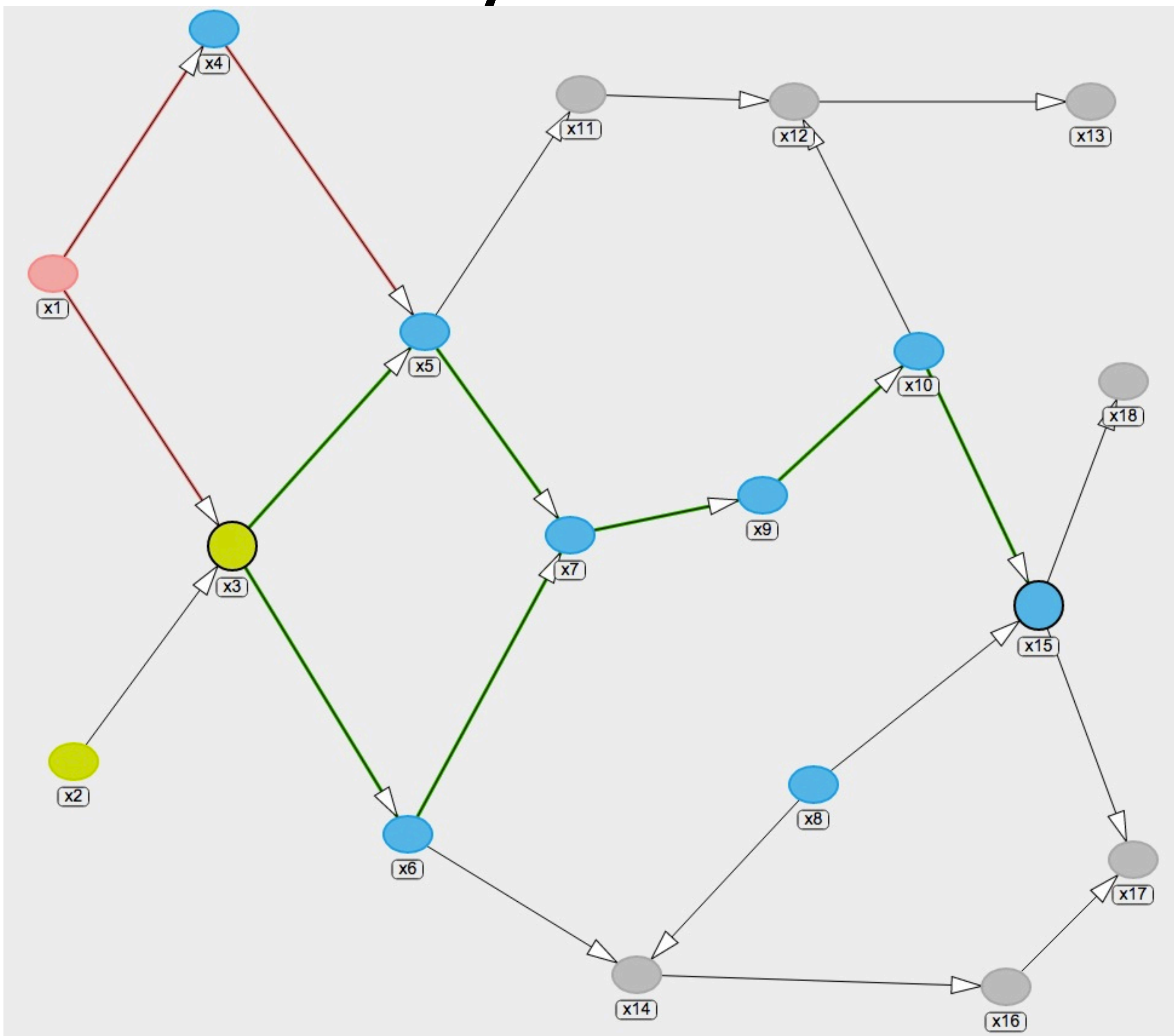
Backdoor going through Z2 goes through Z1

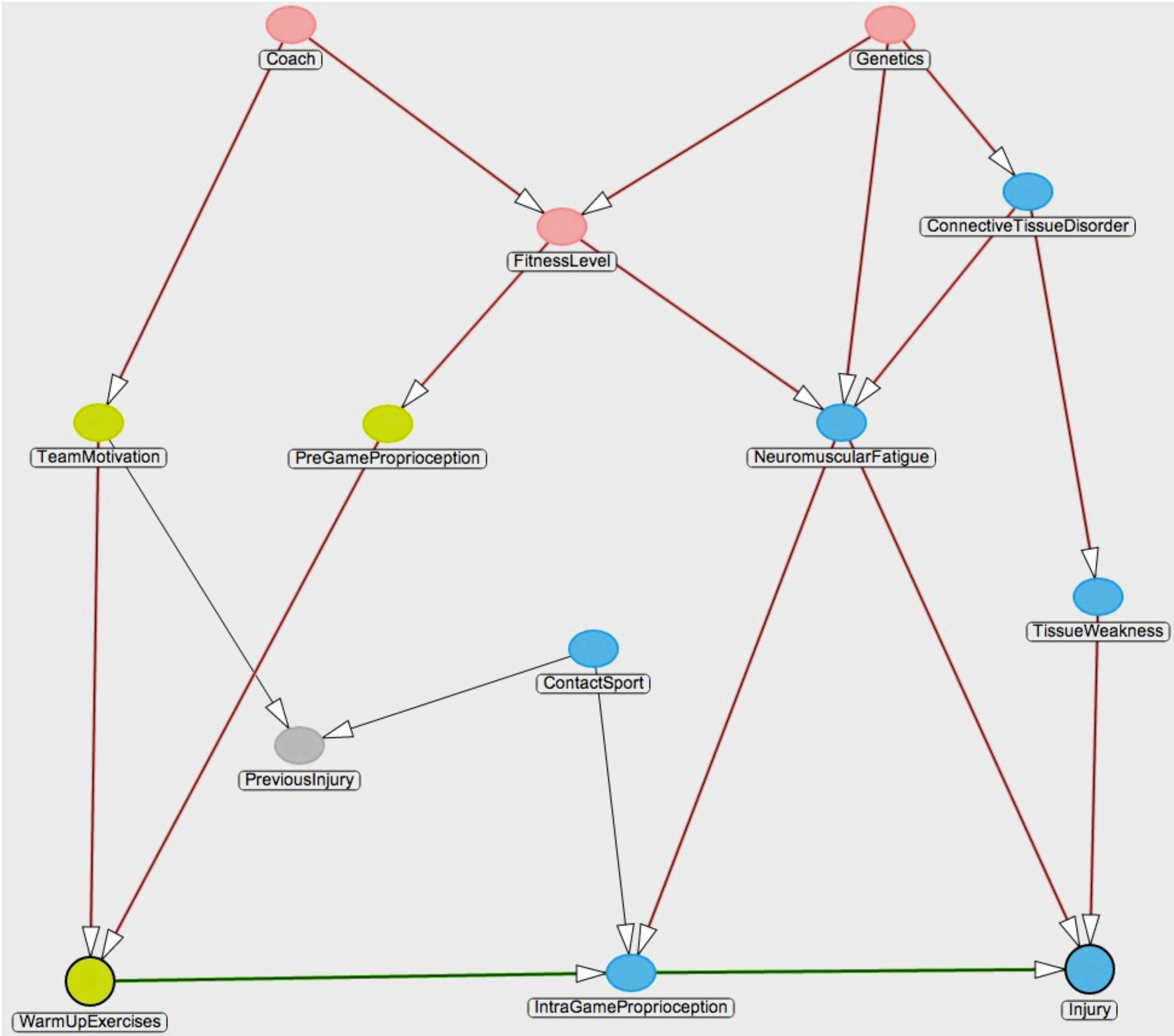
You Try…



Backdoor going through Z2 goes through Z1

You Try…x3 -> x5



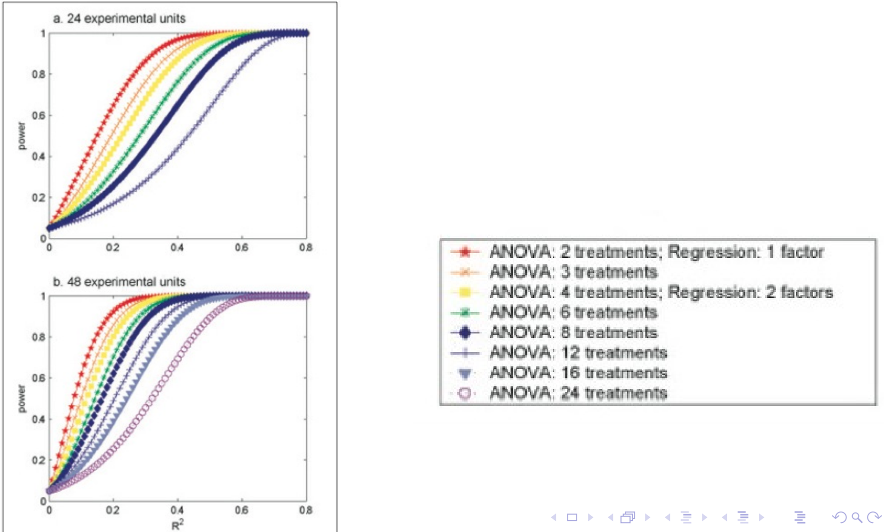

Coach
Genetics
ConnectiveTissueDisorder
FitnessLevel
TeamMotivation
PreGameProprioception
NeuromuscularFatigue
TissueWeakness
ContactSport
PreviousInjury
WarmUpExercises
IntraGameProprioception
Injury

ANOVA v. Regression for Experiments

Regression Design and ANOVA Design have the Same Model

- $Y = BX + e$ underlies both
- F-Test for both examines variation explained
- BUT Regression has fewer parameters to sample size

For Linear Relationships, More Power from Regression



A Simulation Approach to ANOVA and Regression Power

```
getY <- function(x) rnorm(length(x), x , 10)

#two approaches
x<-1:24
xAnova<-rep(seq(1,24,length.out=6),4)
```

A Simulation Approach to ANOVA and Regression Power

```
powFunc <- function(predictor, n.sims=500, a=F, fun=getY){
  pvec <- sapply(1:n.sims, function(i) {
    y <-fun(predictor)

    #run either a regression or categorical model
    if(a){
      alm <- lm(y~I(factor(predictor)))
    }else{
      alm <- lm(y~predictor)
    }

    #get p from an f test
    anova(alm)[1,5]
  } )

  #power
  1 - sum(pvec > 0.05)/n.sims
}
```

Yes, Regression More Powerful

```r
set.seed(100)
powFunc(x)

# [1] 0.914

powFunc(xAnova, a=T)

# [1] 0.712
```

What if the Relationship is Nonlinear

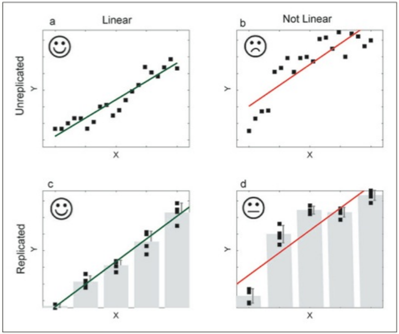
```r
getYSat <- function(x) rnorm(length(x), -2/x, 0.7)
#
powFunc(x, fun=getYSat)

# [1] 0.39

powFunc(xAnova, a=T, fun=getYSat)

# [1] 0.918
```

Replicated Regression or Other Options



Nonlinear Least Squares an option, GLM if Hetereoskedasticity exists