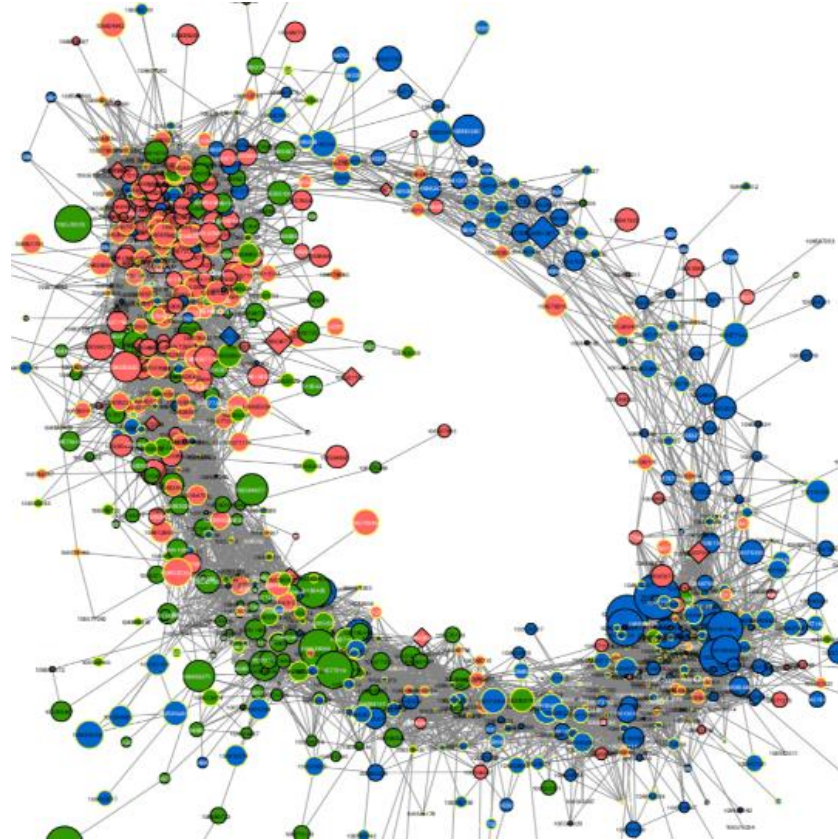# Path analysis

Some fun reading!

Sarah Cubaynes
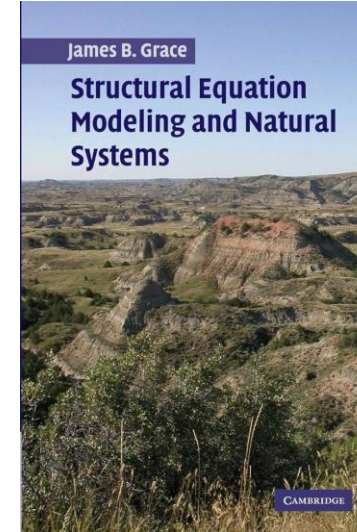

Kate Layton-Matthews

Jon Lefcheck

`https://jslefche.github.io/sem_book/`

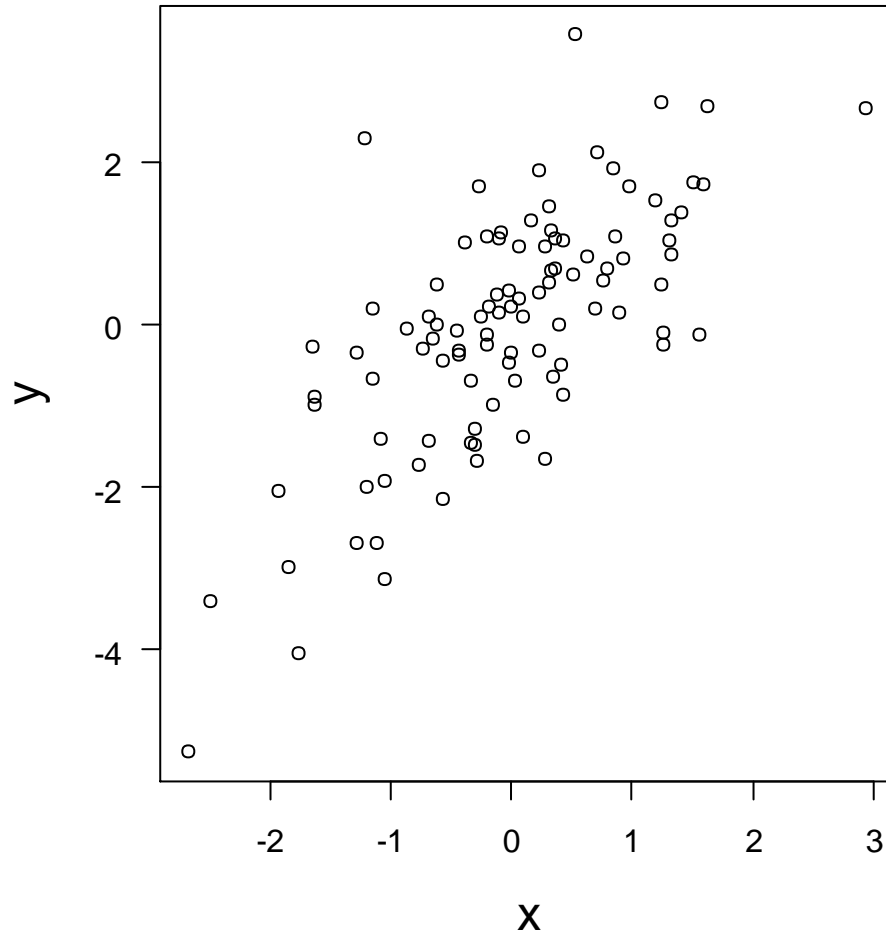**James Grace**

Grace **(2006)** *Cambridge University Press*

Cubaynes et al. **(2014)** *Journal of Animal Ecology*

Layton-Matthews et al. (**2019**) *Journal of Animal Ecology*
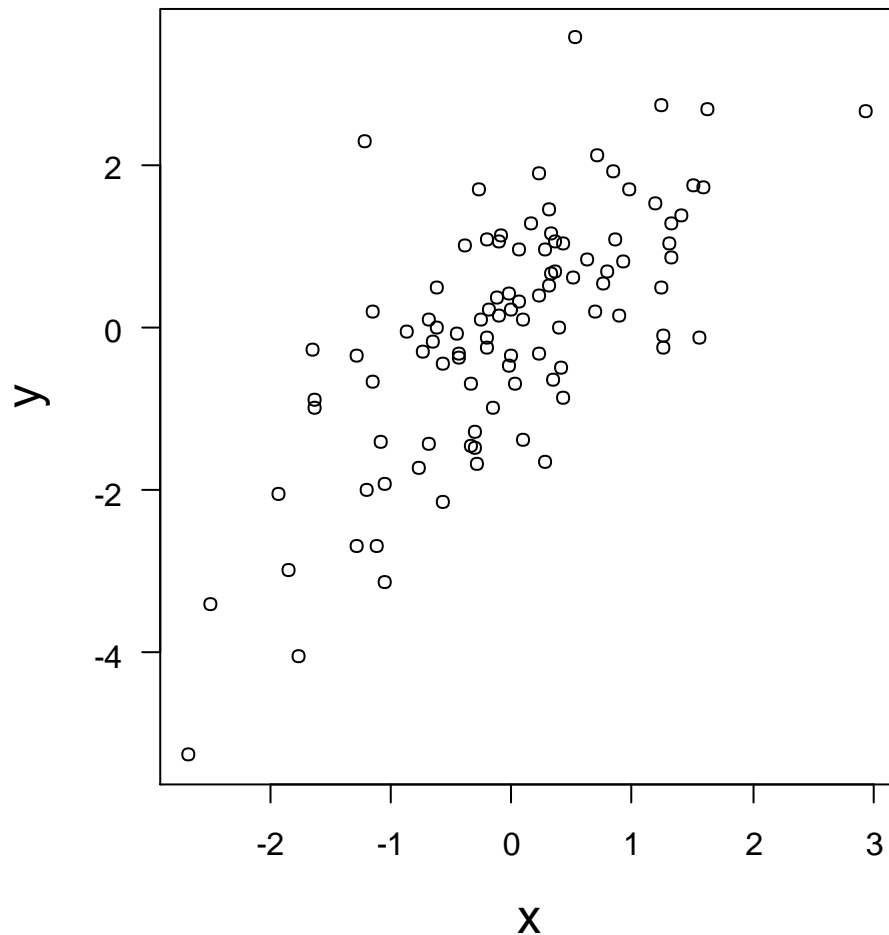
Simulating data (n = 100)


```
summary(lm(y ~ x, dat = d))
```

```
x ~ normal(0,1)
y ~ normal(x, 1)
```

Grace **(2006)**

Variance-covariance matrix

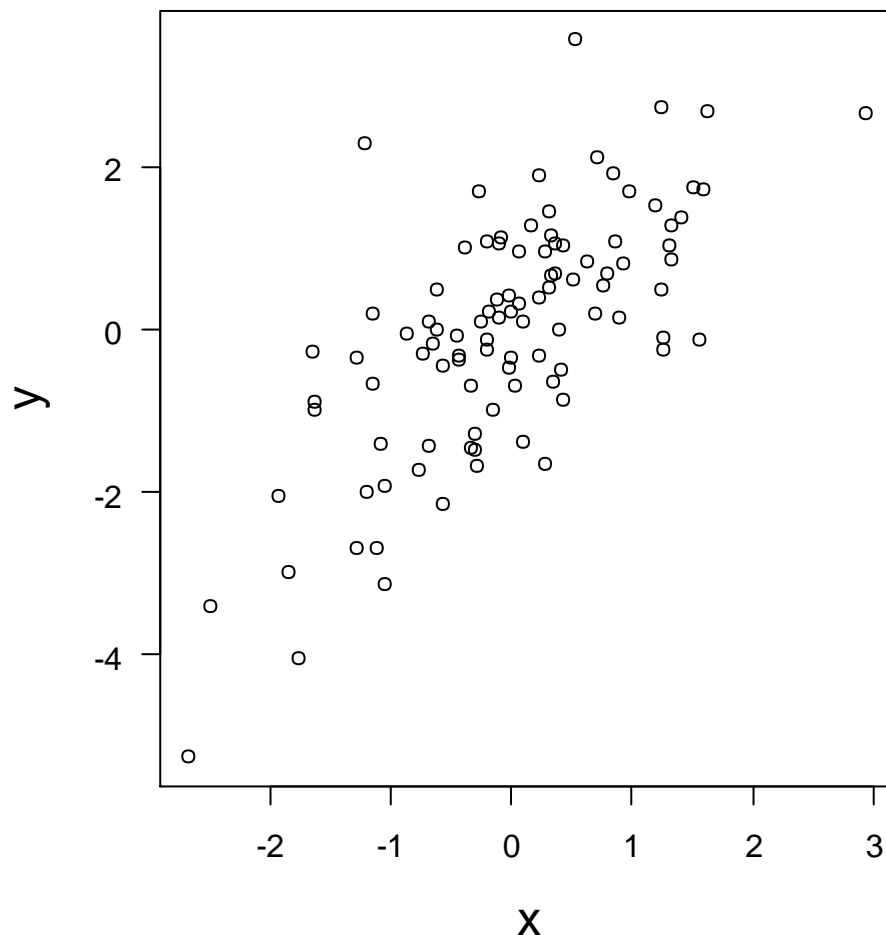$$\Sigma_{yx} = \begin{bmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{bmatrix}$$

$$\Sigma_{yx} = \begin{bmatrix} 2.247 & 1.035 \\ 1.035 & 0.946 \end{bmatrix}$$

Grace **(2006)**

Variance of y

`var(x)`

$$\Sigma_{yx} = \begin{bmatrix} \boldsymbol{\sigma_y^2} & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{bmatrix}$$
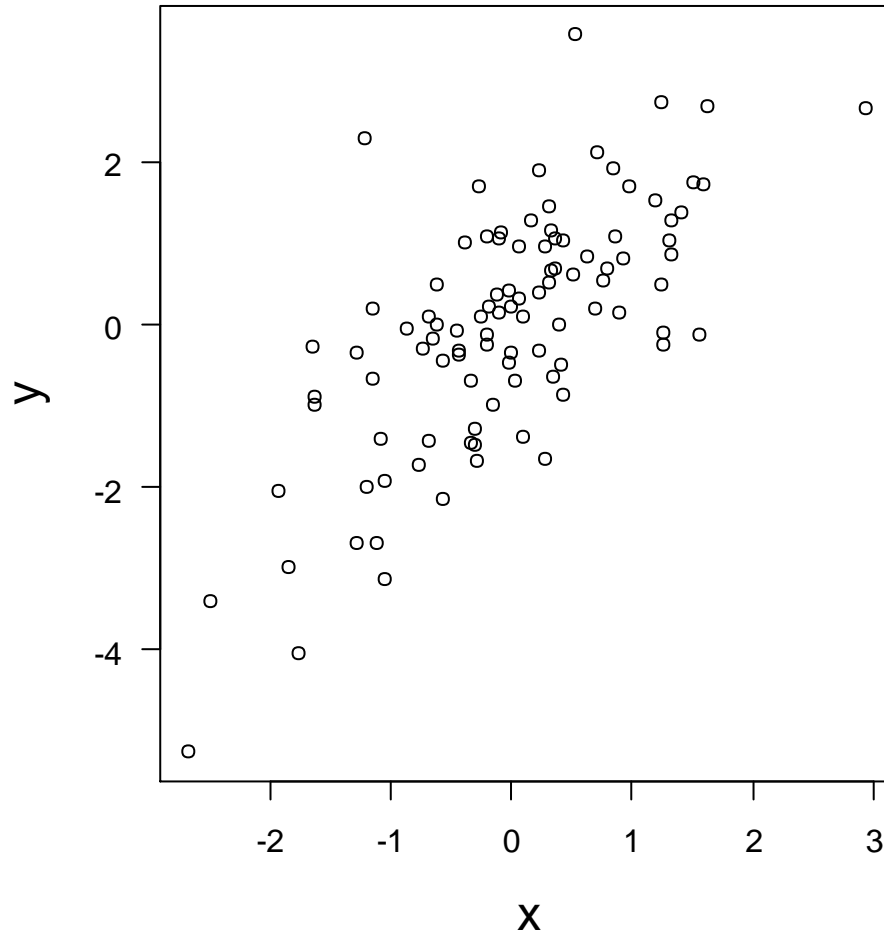
Variance of y

$$\sigma_y^2 = \frac{\sum(y_i - \bar{y})^2}{n - 1}$$

Grace **(2006)**, eqs. 3.1 & 3.2

Variance of x

`var(x)`

$$\Sigma_{yx} = \begin{bmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \boldsymbol{\sigma_x^2} \end{bmatrix}$$
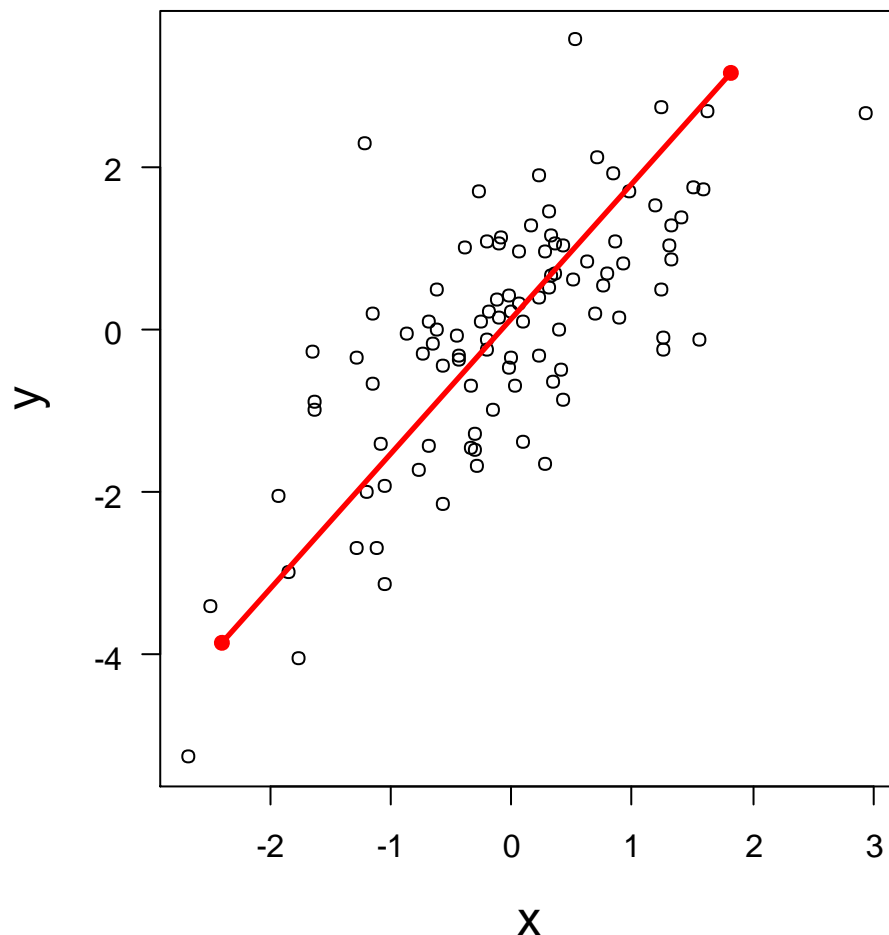
Variance of x

$$\sigma_x^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

Grace **(2006)**, eqs. 3.1 & 3.2
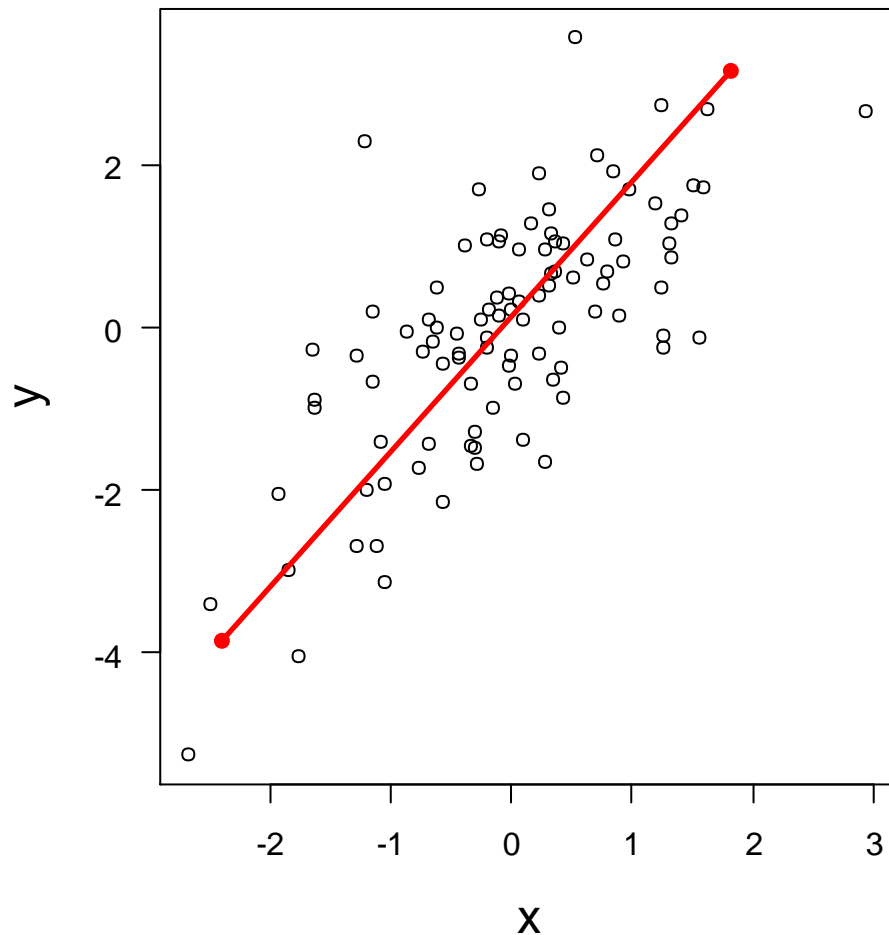
Covariance

$$\mathtt{cov(x,y)}$$

$$\Sigma_{yx} = \begin{bmatrix} \sigma_y^2 & \boldsymbol{\sigma_{yx}} \\ \boldsymbol{\sigma_{yx}} & \sigma_x^2 \end{bmatrix}$$

Covariance

$$\sigma_{xy} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{n-1}$$

Correlation

`cor(x,y)`

Covariance

$$\sigma_x \sigma_y \rho = \sigma_{xy} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{n-1}$$
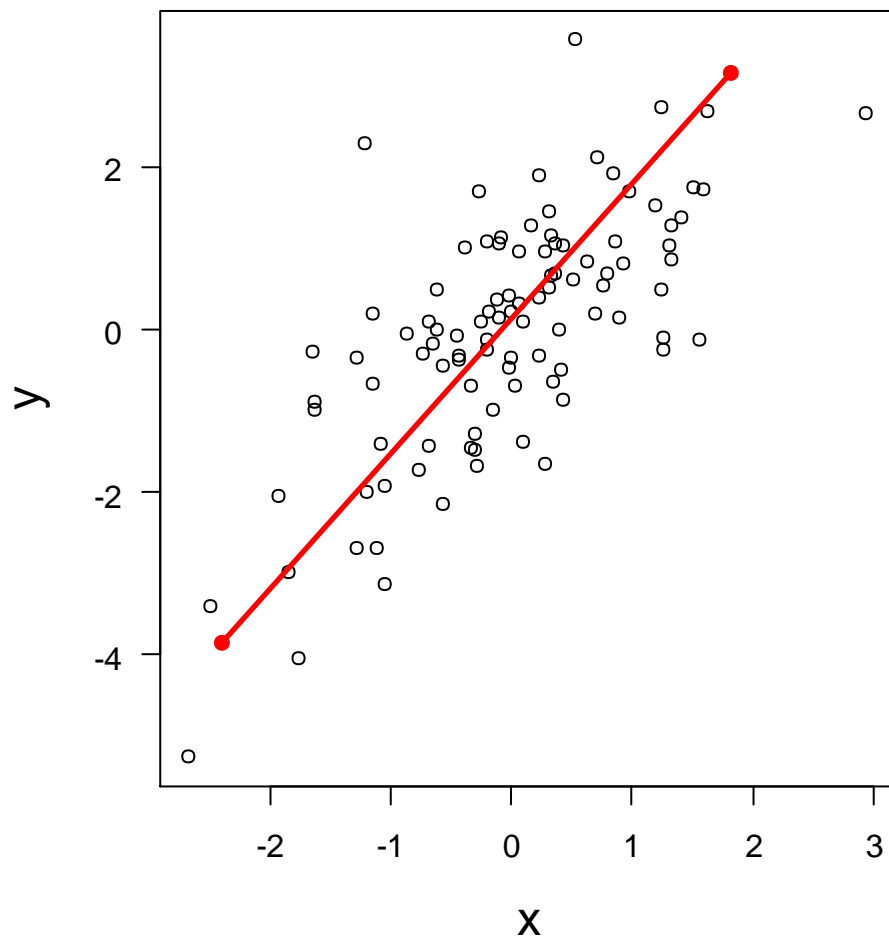
Correlation

$$r = \rho = \frac{\sigma_x \sigma_y \rho}{\sigma_x \sigma_y}$$
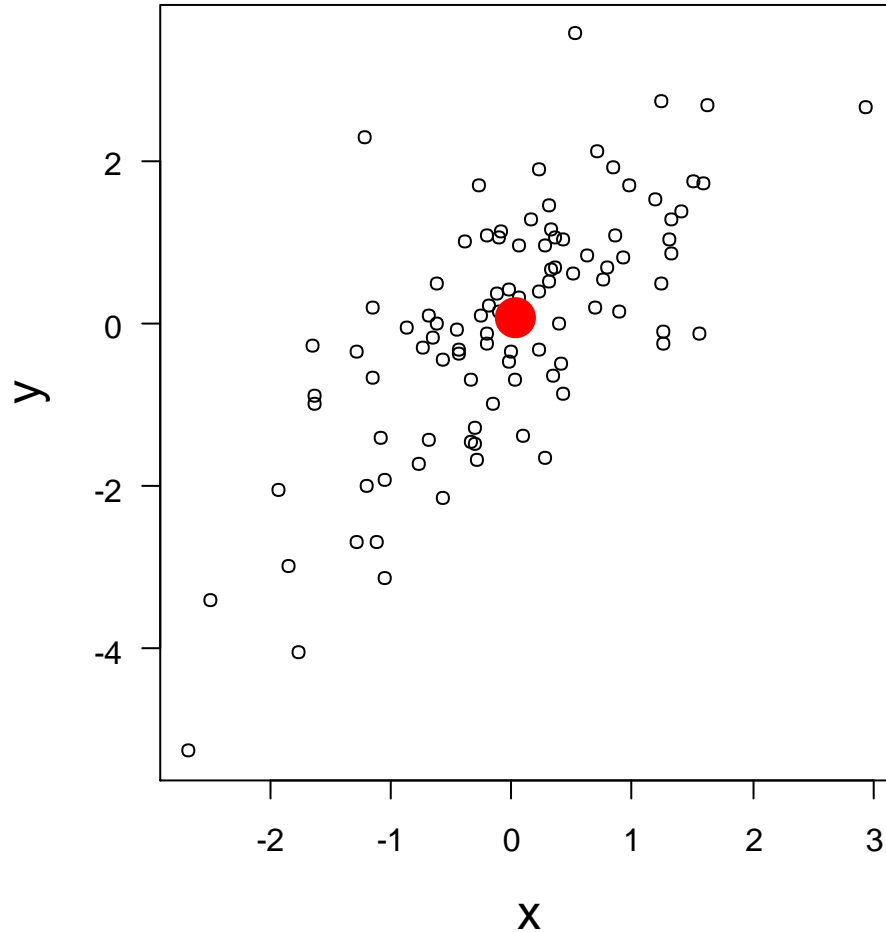
Slope

`lm(y ~ x)`

Covariance

$$\sigma_x \sigma_y \rho = \sigma_{xy} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{n - 1}$$

Slope

$$\beta_1 = \frac{\sigma_x \sigma_y \rho}{\sigma_x^2}$$
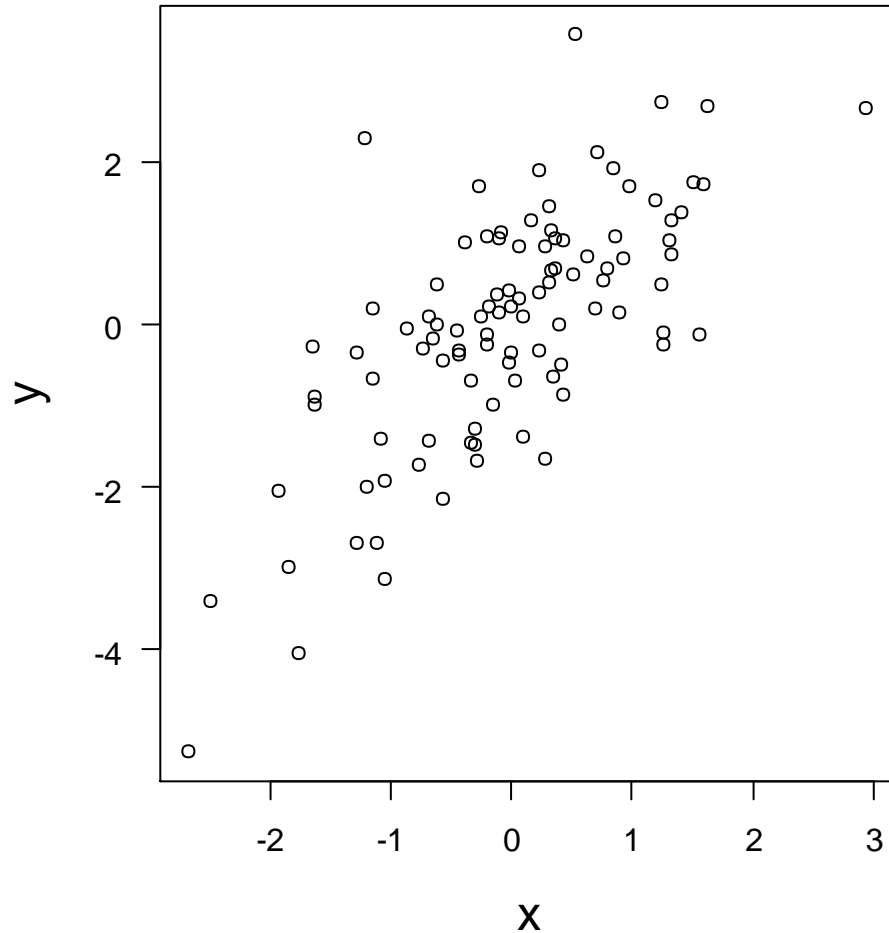
Intercept

`lm(y ~ x)`

Slope

$$\beta_1 = \frac{\sigma_x \sigma_y \rho}{\sigma_x^2}$$

Intercept

$$\beta_0 = \bar{y} - \beta_1 \times \bar{x}$$

An easy-peasy linear model

R code

```
lm(y ~ x)
```
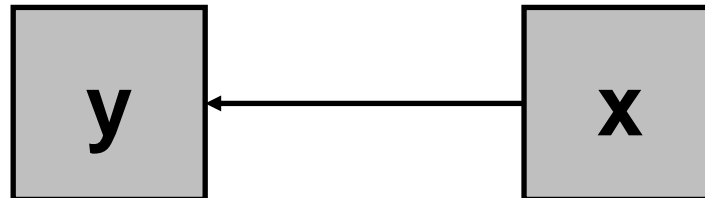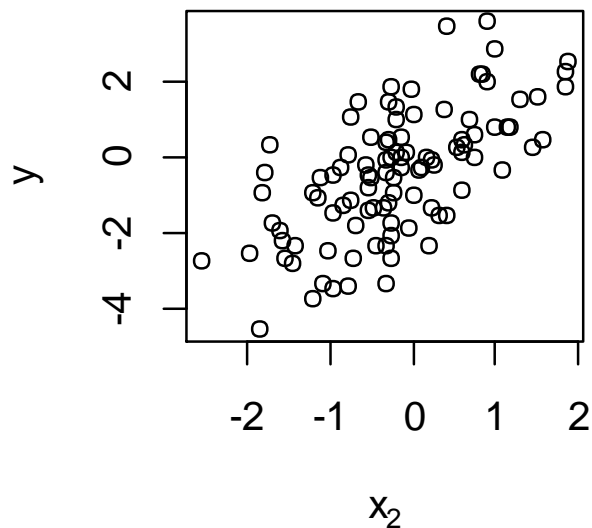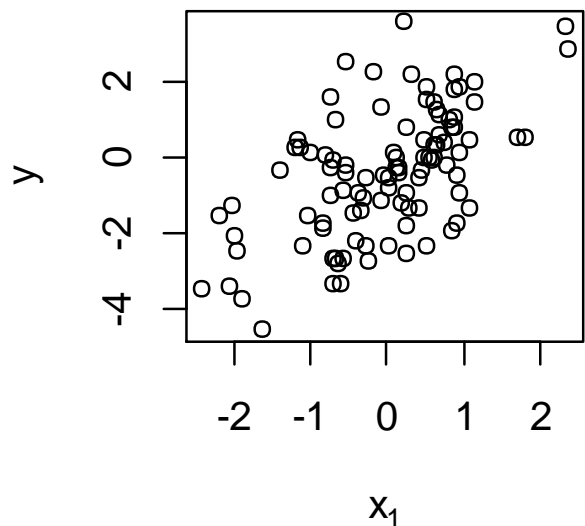
Model

$$y_i \sim \text{normal}(\beta_0 + \beta_1 x_i, \sigma_y^2)$$

Causal diagram
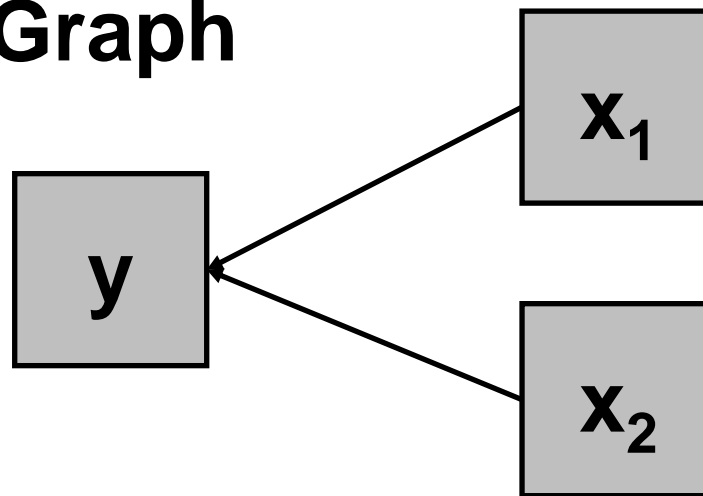
Expanding to two explanatory variables

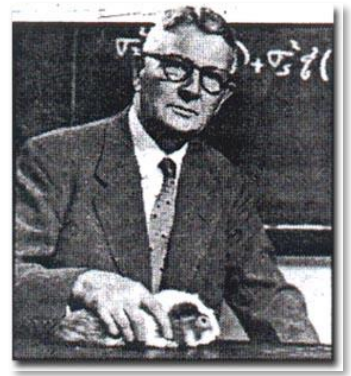**R code**

```
lm(y ~ x1 + x2)
```

Model

$$y_i \sim \text{normal}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \sigma_y^2)$$
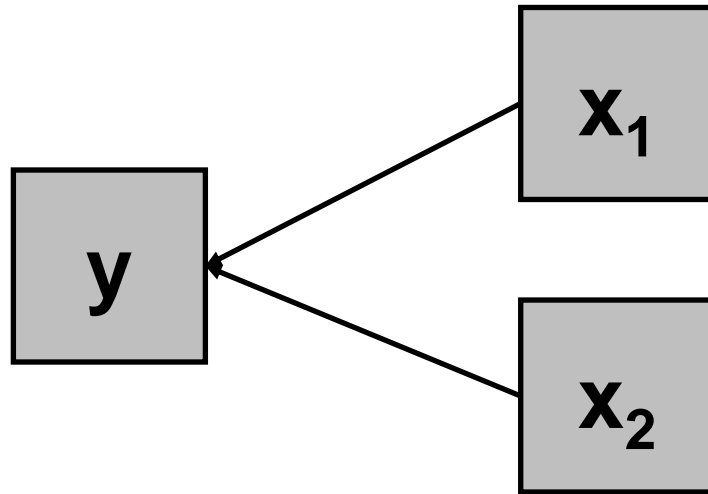
Graph

Sewall Wright's genius breakthrough

Multiple regression

Path analysis
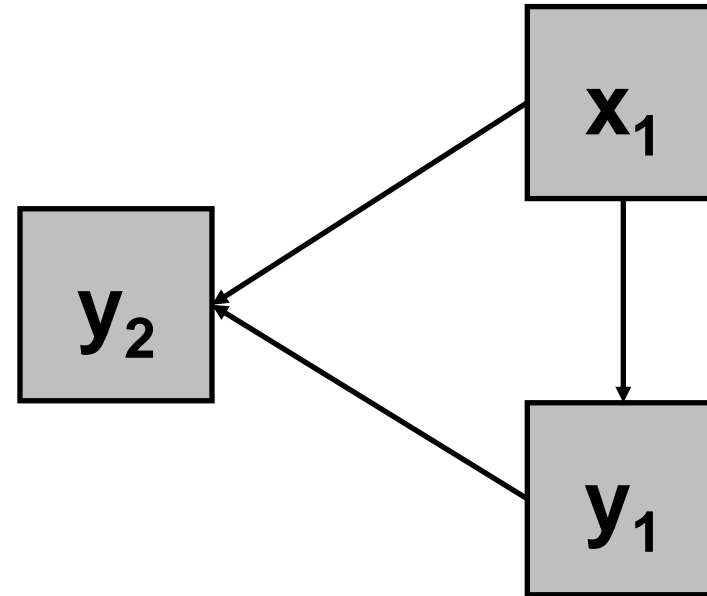
Wright **(1921; 1934)**

Barbara Burks: nature vs. nurture

ON THE INADEQUACY OF THE PARTIAL AND MULTIPLE CORRELATION TECHNIQUE

BARBARA STODDARD BURKS

Stanford University

PART I. IN THE STUDY OF CAUSATION

Logical considerations lead to the conclusion that the techniques of partial and multiple correlation are fraught with dangers that seriously restrict their applicability. In fact their attempted use in (a) isolating the causes which operate upon observed effects, and (b) defining the extent to which two measures involve common factors unique to themselves, often result in intepretations that are misleading and even untrue excepting in a few special types of situation. Only issues arising in the first field (*i.e.*, causation) will be discussed at this time. Consideration of the second field will be left for a subsequent paper (Part II).

Burks **(1926)**

Barbara Burks: collider bias

Other unmeasured causes

Social status

Parental intelligence

Child's intelligence

ON THE INADEQUACY OF THE PARTIAL AND MULTIPLE CORRELATION TECHNIQUE
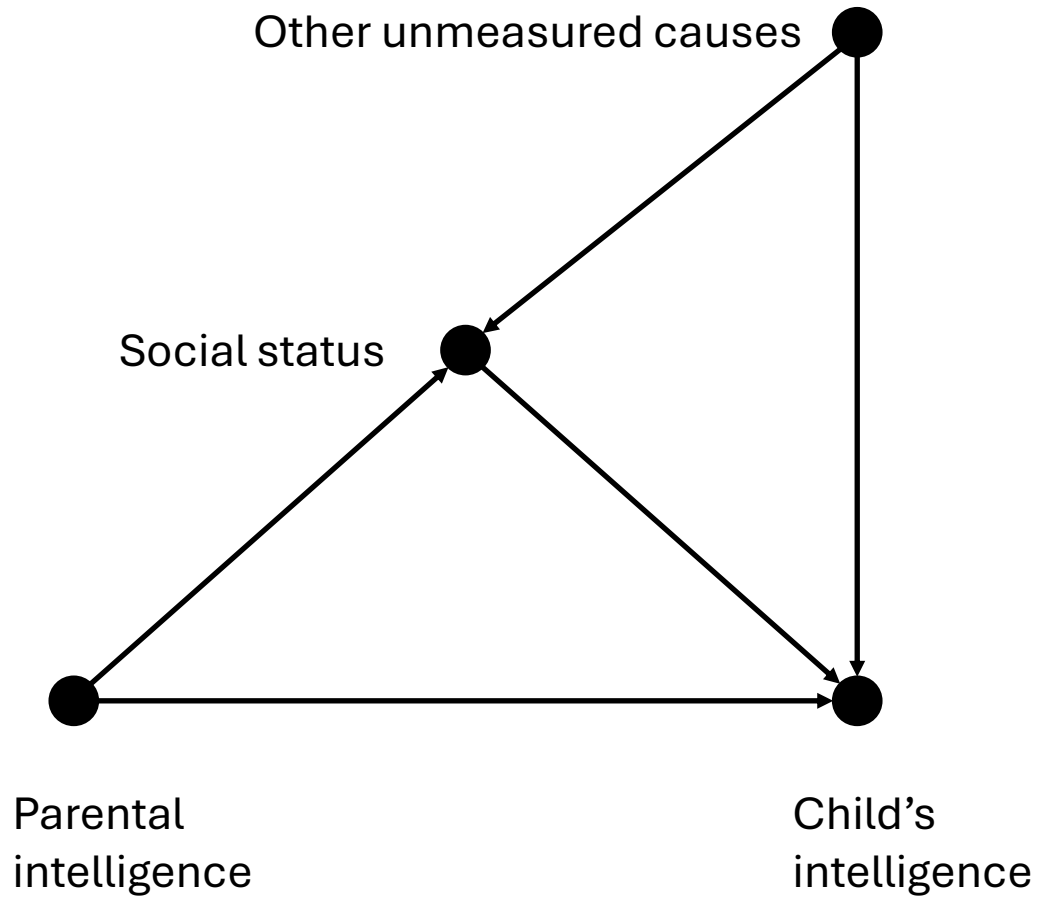
BARBARA STODDARD BURKS

Stanford University

PART I. IN THE STUDY OF CAUSATION

Logical considerations lead to the conclusion that the techniques of partial and multiple correlation are fraught with dangers that seriously restrict their applicability. In fact their attempted use in (a) isolating the causes which operate upon observed effects, and (b) defining the extent to which two measures involve common factors unique to themselves, often result in intepretations that are misleading and even untrue excepting in a few special types of situation. Only issues arising in the first field (i.e., causation) will be discussed at this time. Consideration of the second field will be left for a subsequent paper (Part II).

Burks **(1926);** Pearl **(2001)** Figure 9.2

Simulating data (n = 100)
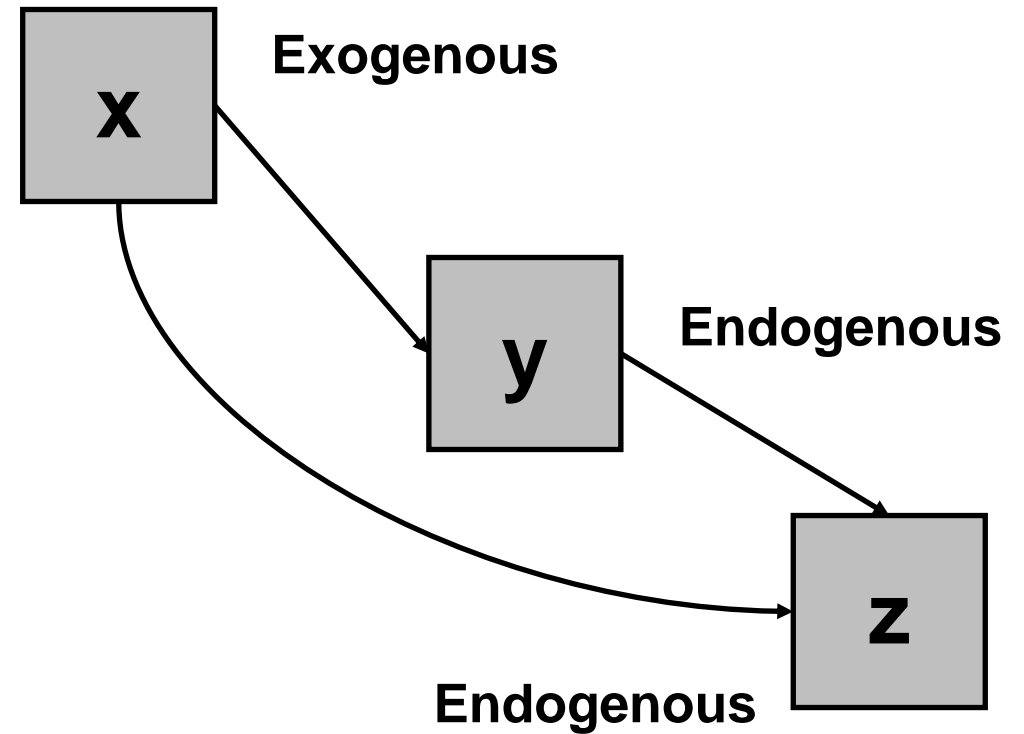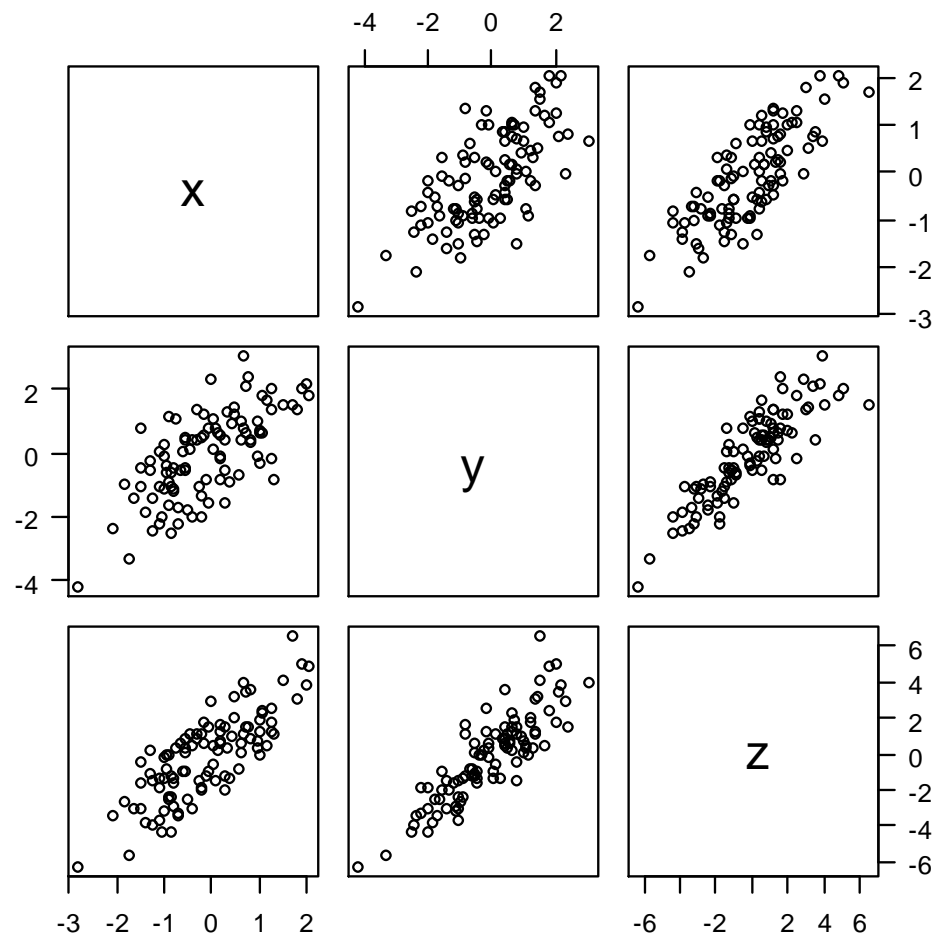

```
x ~ normal(0,1)
y ~ normal(x, 1)
z ~ normal(x + z, 1)
```


Exogenous

Endogenous

Endogenous

Simulating data (n = 100)


```
x ~ normal(0,1)
y ~ normal(x, 1)
z ~ normal(x + z, 1)
```

Simulating data (n = 100)


```
x ~ normal(0,1)
y ~ normal(x, 1)
z ~ normal(x + z, 1)
```


Exogenous

x

Endogenous

z

Mediator

y

Endogenous
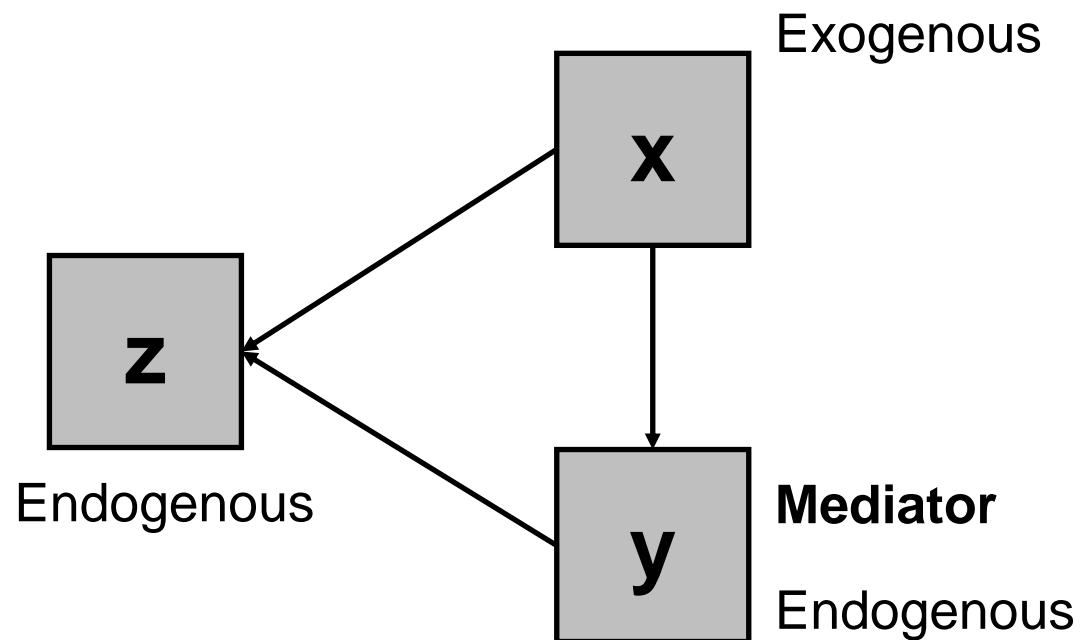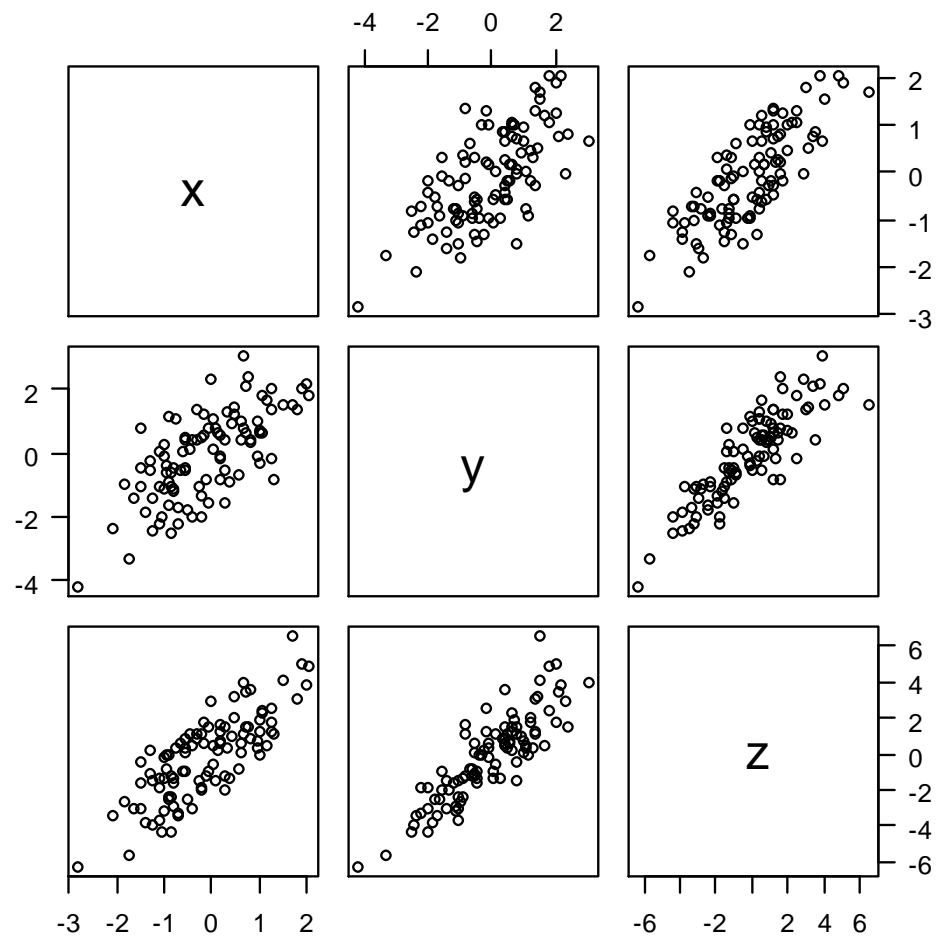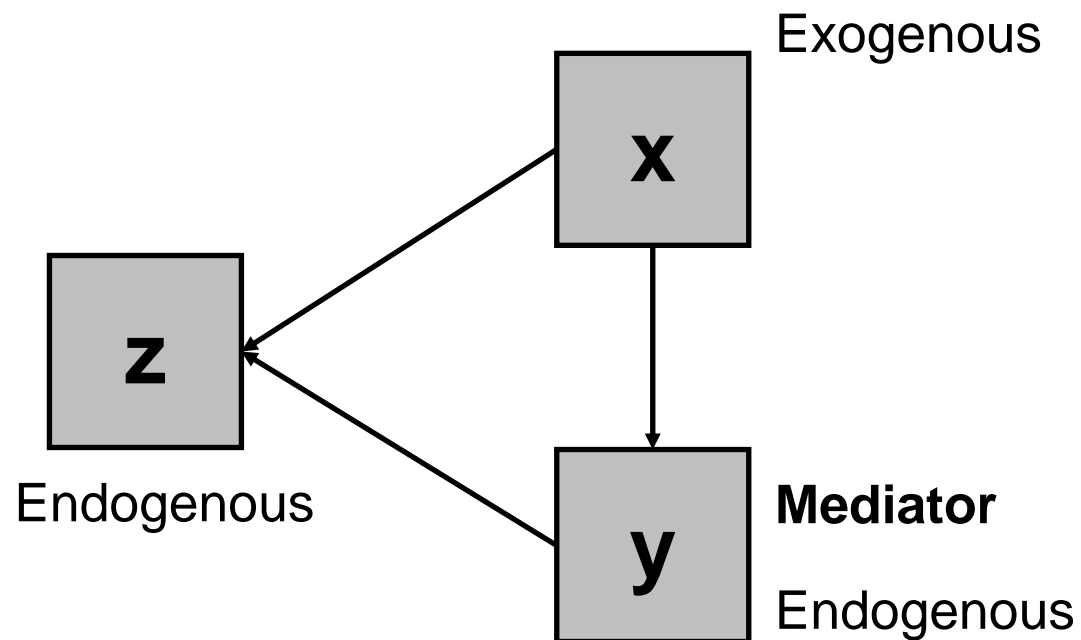
Variance-covariance and correlation matrices
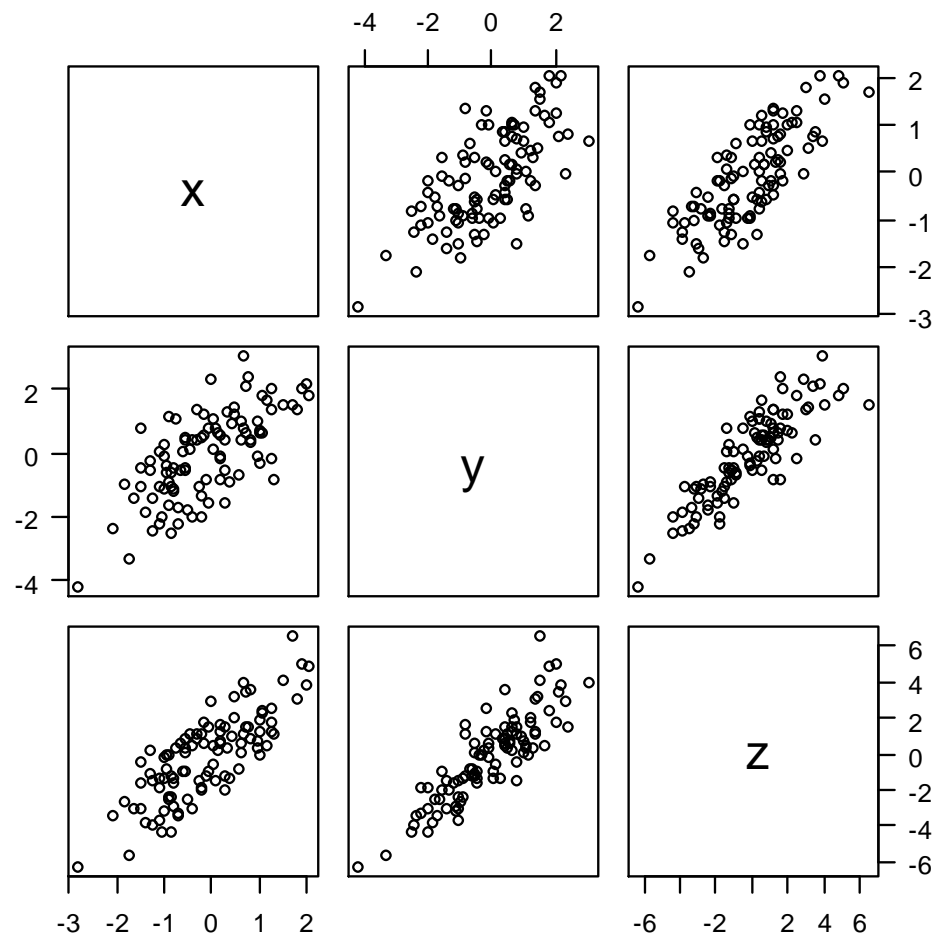
Variance-covariance matrix

$$\Sigma_{xyz} = \begin{bmatrix} 0.992 & 0.925 & 1.884 \\ 0.925 & 1.869 & 2.798 \\ 1.884 & 2.798 & 5.625 \end{bmatrix}$$

Correlation matrix

$$R_{xyz} = \begin{bmatrix} 1 & 0.679 & 0.798 \\ 0.679 & 1 & 0.863 \\ 0.798 & 0.863 & 1 \end{bmatrix}$$

Let's fit a model

```
104  path1 <- psem(
105    lm(y ~ x, data = d),
106    lm(z ~ y + x, data = d),
107    data = d
108  )
109  summary(path1)
```


Variance-covariance matrix

$$\Sigma_{xyz} = \begin{bmatrix} 0.992 & 0.925 & 1.884 \\ 0.925 & 1.869 & 2.798 \\ 1.884 & 2.798 & 5.625 \end{bmatrix}$$

Correlation matrix

$$R_{xyz} = \begin{bmatrix} 1 & 0.679 & 0.798 \\ 0.679 & 1 & 0.863 \\ 0.798 & 0.863 & 1 \end{bmatrix}$$

The relationship between y ~ x

$$\gamma_1 = \frac{\Sigma_{xy}}{\sigma_x^2} = (R_{12})\frac{\sigma_y}{\sigma_x}$$

Variance-covariance matrix

$$\Sigma_{xyz} = \begin{bmatrix} 0.992 & 0.925 & 1.884 \\ 0.925 & 1.869 & 2.798 \\ 1.884 & 2.798 & 5.625 \end{bmatrix}$$

Correlation matrix

$$R_{xyz} = \begin{bmatrix} 1 & 0.679 & 0.798 \\ 0.679 & 1 & 0.863 \\ 0.798 & 0.863 & 1 \end{bmatrix}$$

The relationship between z ~ x|y

$$\gamma_2 = \frac{R_{xz} - (R_{xy} \times R_{yz})}{1 - R_{xy}^2} \times \frac{\sigma_z}{\sigma_x}$$

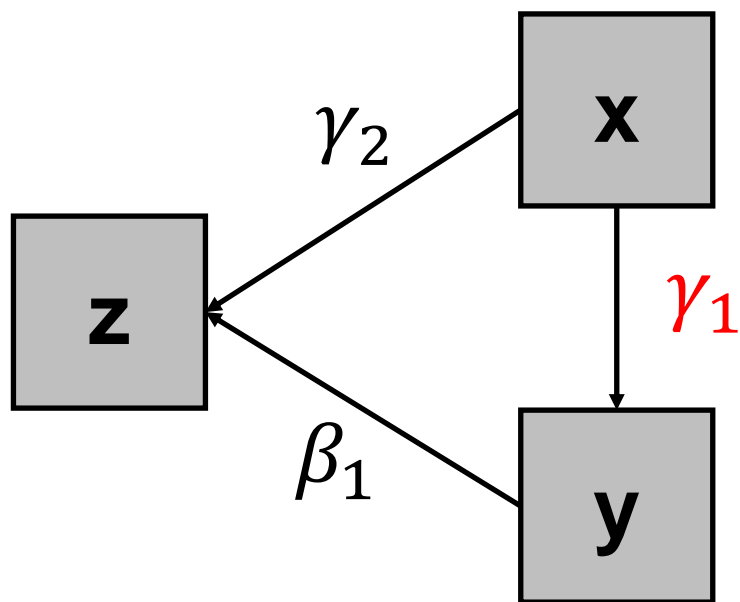**Variance-covariance matrix**

$$\Sigma_{xyz} = \begin{bmatrix} 0.992 & 0.925 & 1.884 \\ 0.925 & 1.869 & 2.798 \\ 1.884 & 2.798 & 5.625 \end{bmatrix}$$

Correlation matrix

$$R_{xyz} = \begin{bmatrix} 1 & 0.679 & 0.798 \\ 0.679 & 1 & 0.863 \\ 0.798 & 0.863 & 1 \end{bmatrix}$$

The relationship between z ~ y|x

$$\beta_1 = \frac{R_{yz} - (R_{xy} \times R_{xz})}{1 - R_{xy}^2} \times \frac{\sigma_z}{\sigma_x}$$
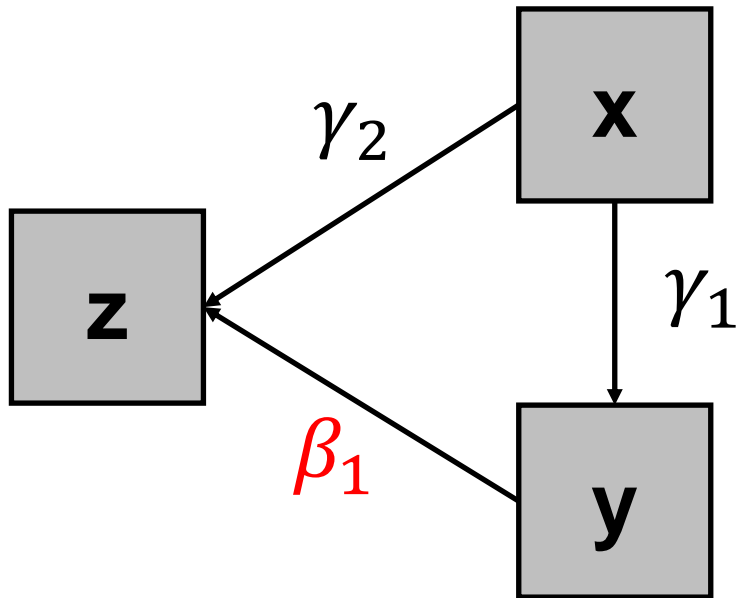
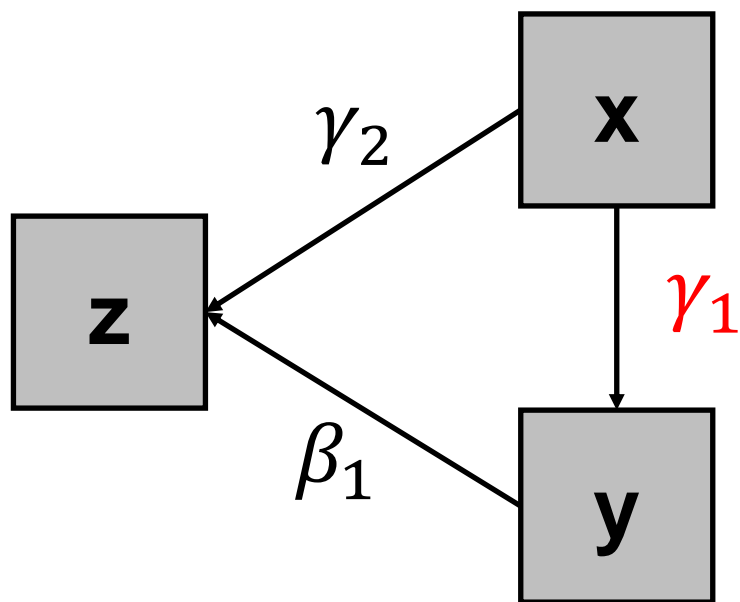**Variance-covariance matrix**
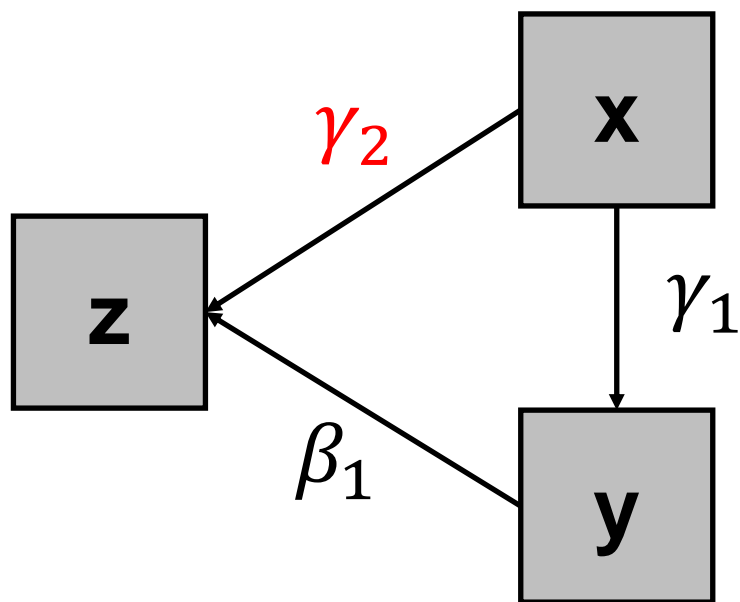
$$\Sigma_{xyz} = \begin{bmatrix} 0.992 & 0.925 & 1.884 \\ 0.925 & 1.869 & 2.798 \\ 1.884 & 2.798 & 5.625 \end{bmatrix}$$

Correlation matrix

$$R_{xyz} = \begin{bmatrix} 1 & 0.679 & 0.798 \\ 0.679 & 1 & 0.863 \\ 0.798 & 0.863 & 1 \end{bmatrix}$$

This was 100 years ago…

They just figured it out… by hand.

One variable can influence another through a third!

We have always '**known**' that when we discuss and teach principles of wildlife management and ecology, **we can also formally model it.**

Our first example

Available browse (**v**; vegetation) as a function of ungulate (**u**) and predator (**w**; wolf) abundance.

| wolves (w) | $\xrightarrow{\alpha_1}$ | ungulates (u) | $\xrightarrow{\beta_1}$ | vegetation (v) |

Lecture2b_wolf_rsf.R
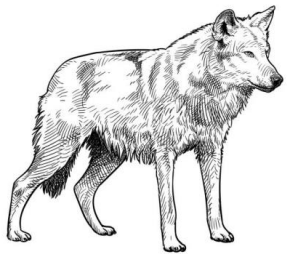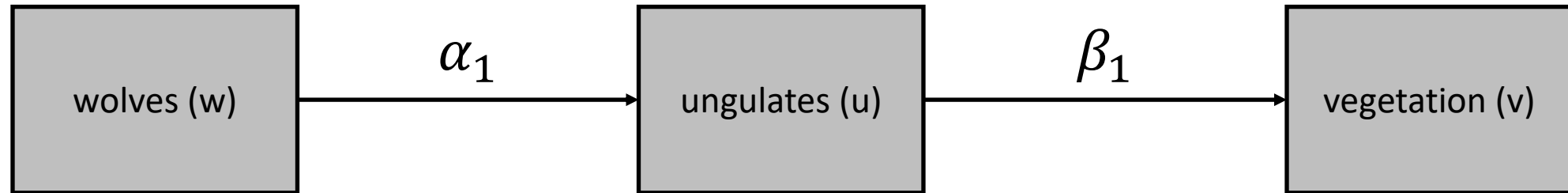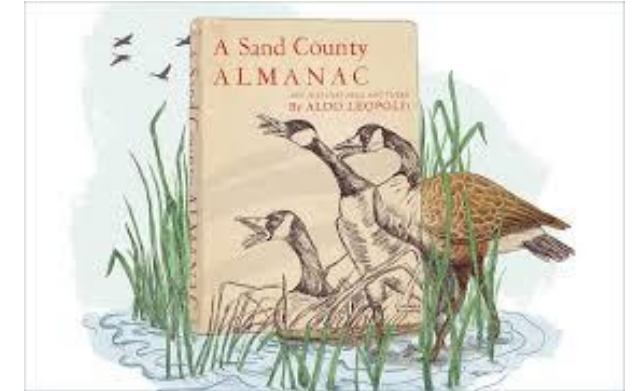
First, we'll simulate random variation in wolf density (*w*)

$$w_i \sim \text{lognormal}(0.35, 0.75)$$
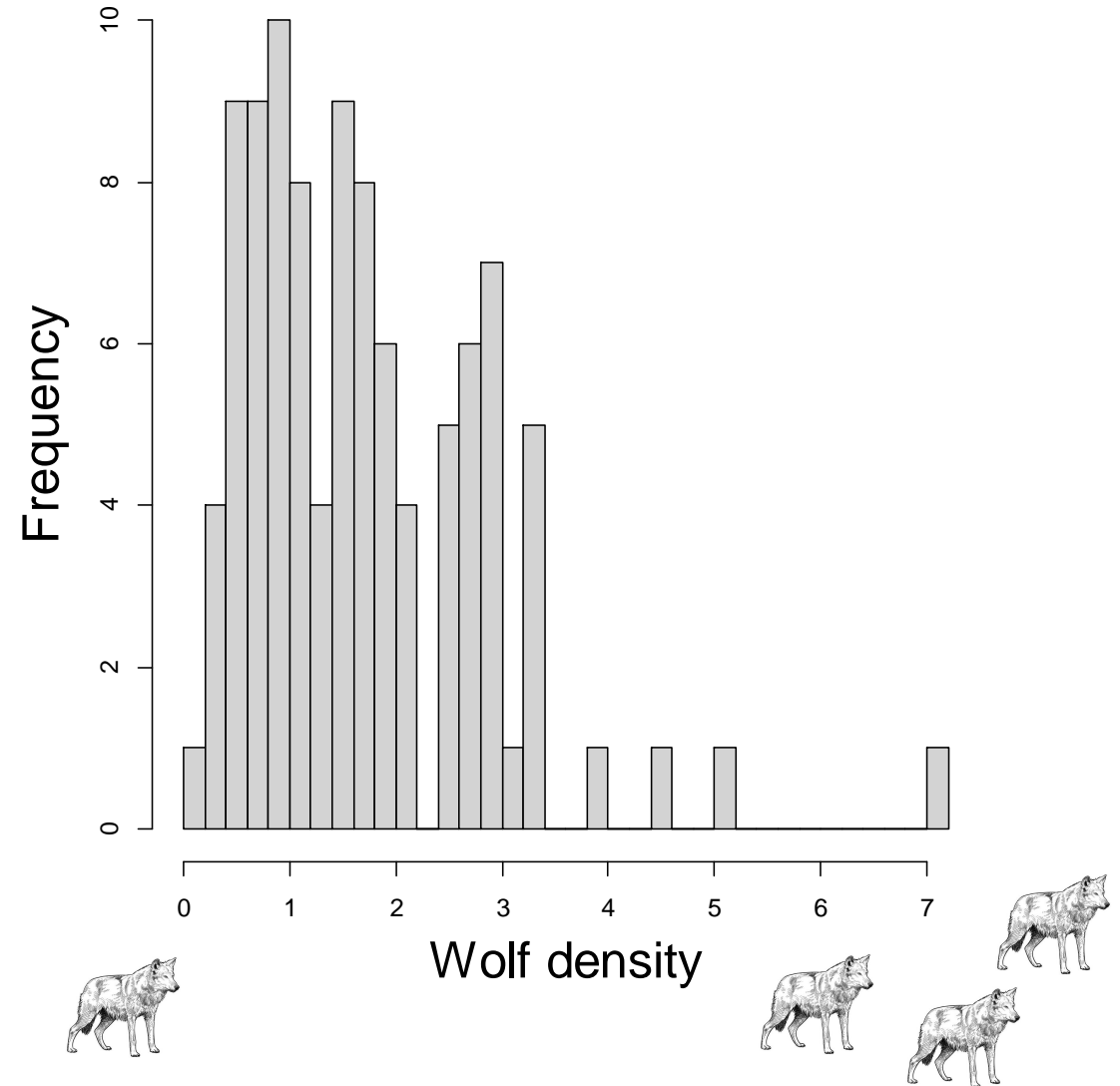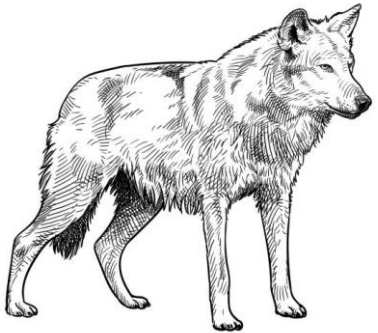
Second, we'll simulate ungulate density (*u*) as a function of *w*

$$u_i \sim \text{lognormal}(\alpha_0 + \alpha_1 w_i, \sigma_v^2)$$
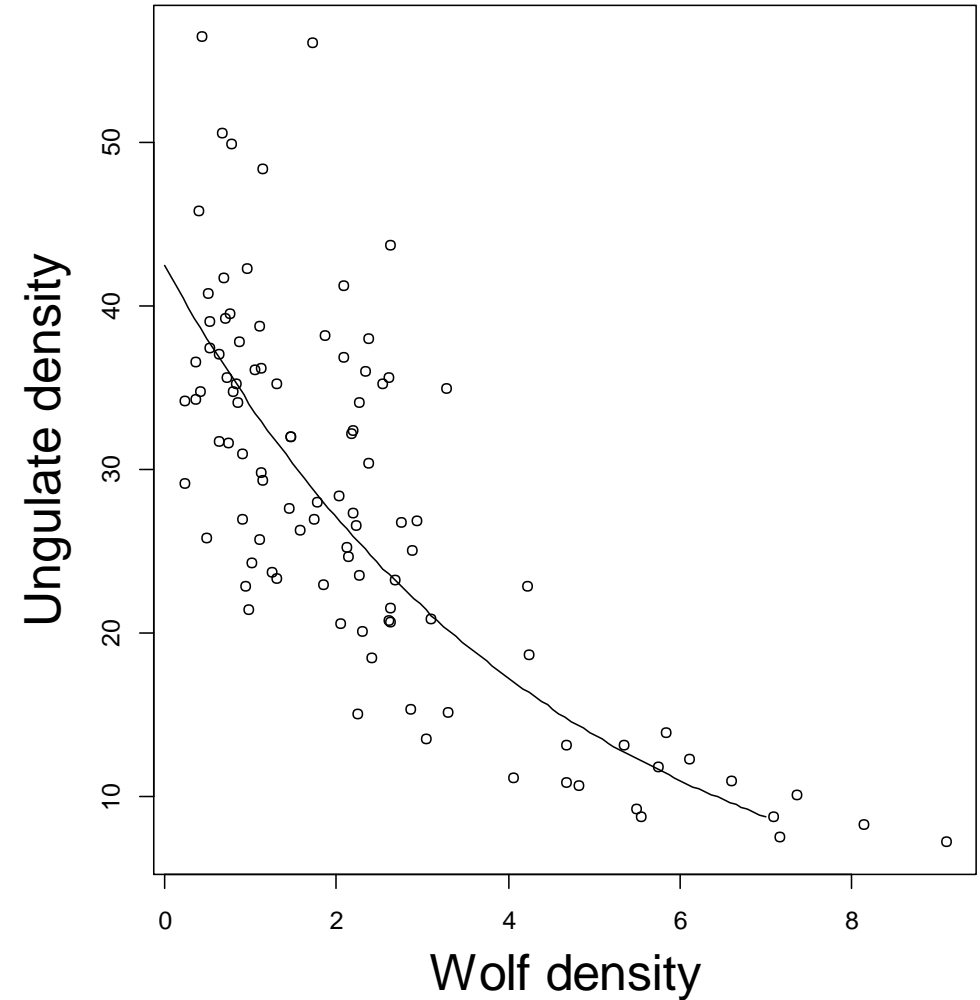
Third, we'll simulate browse density (*v*) as a function of *u*

$$v_i \sim \text{lognormal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$
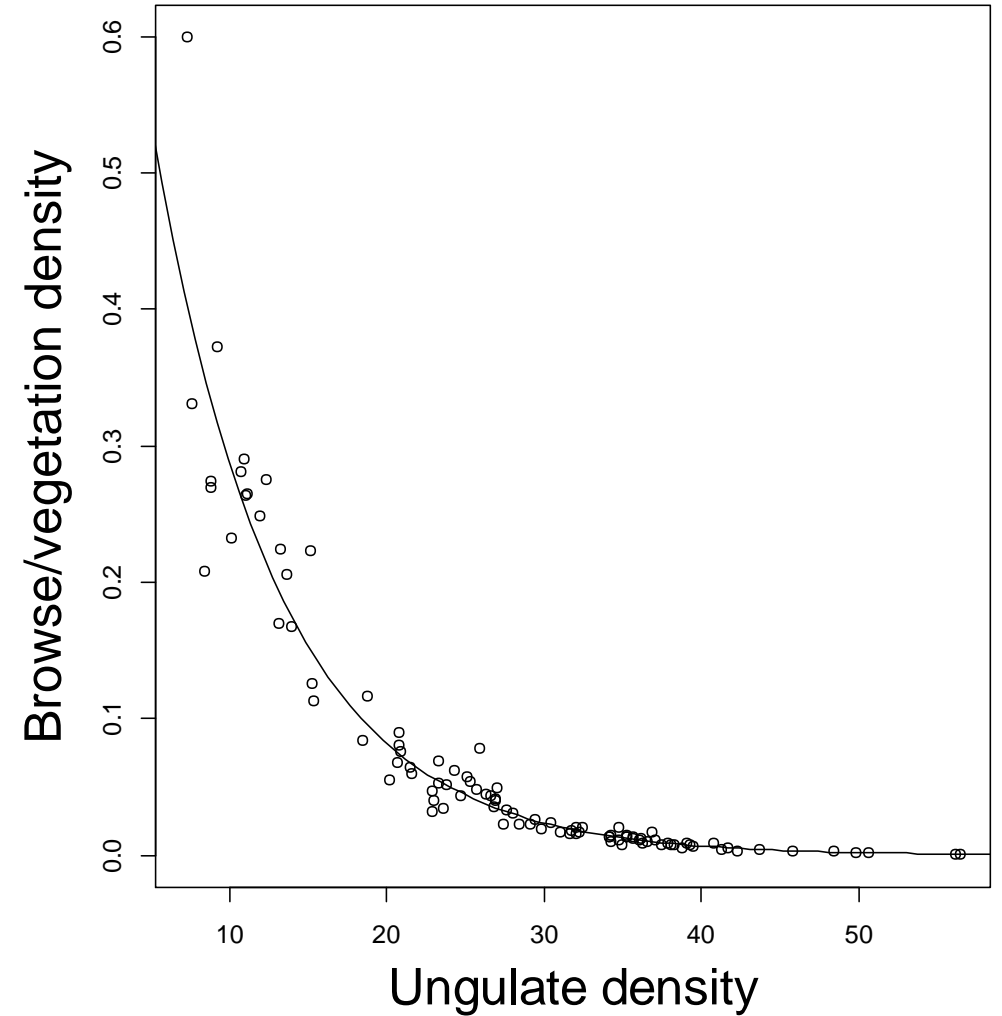
We'll use simple linear models to analyze the data

1. Identity link functions (i.e., normal distributions) in R
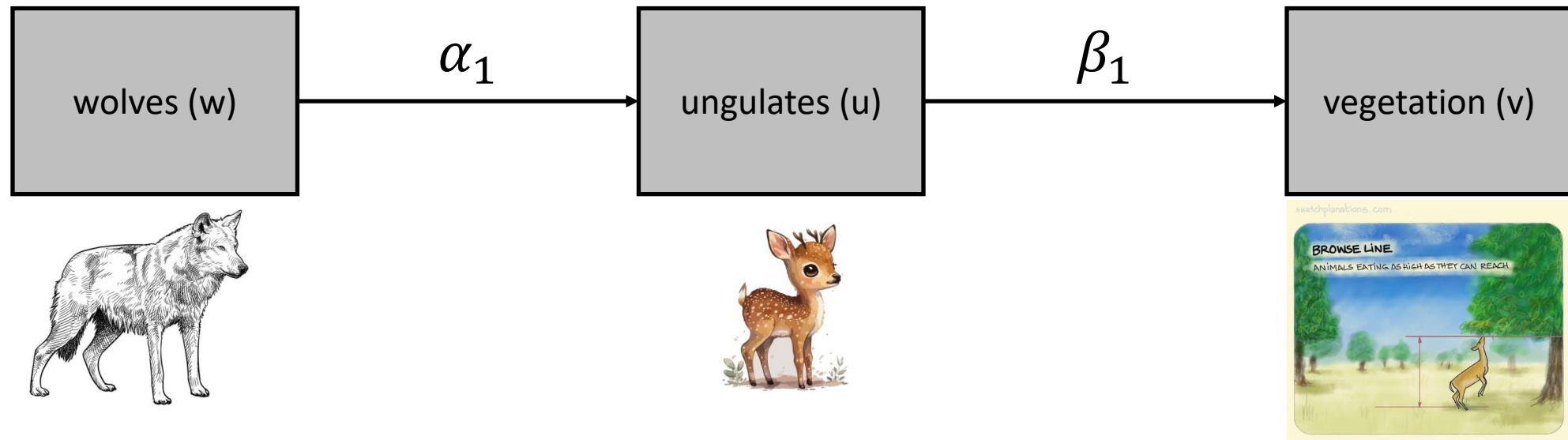
First, we'll use 'identity' link functions in R [lm()]

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$
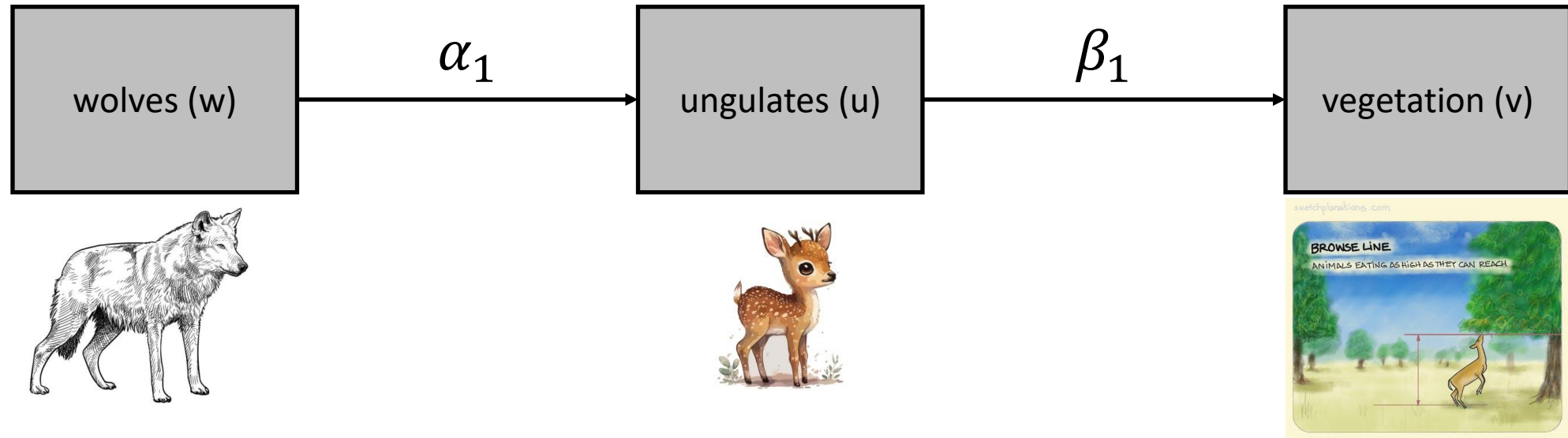
$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$

It really is this simple

Something to think about

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$

$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$

`lm(u ~ w)`

| wolves (w) | α_1 | ungulates (u) |
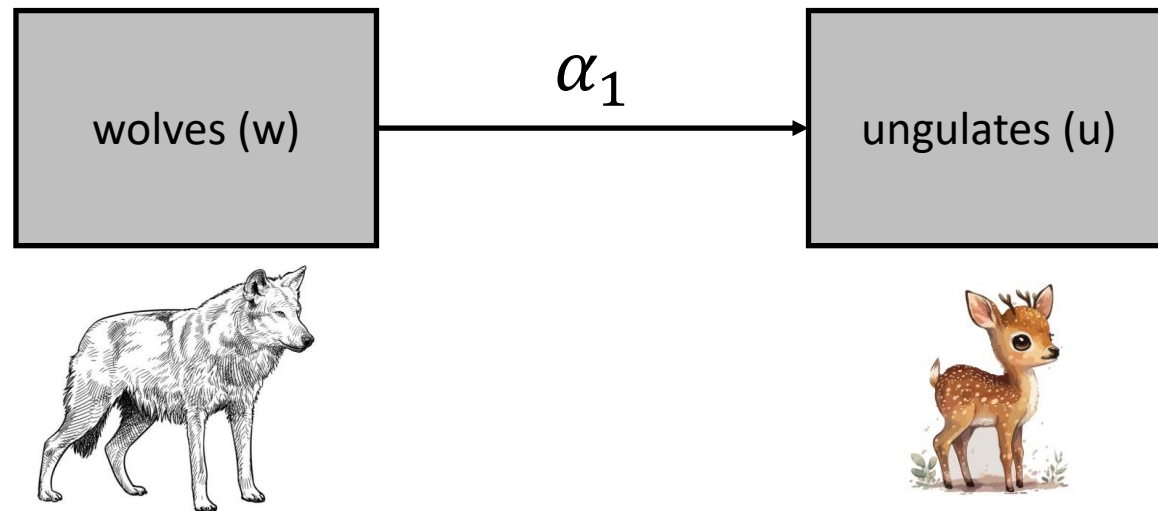
It's easy to calculate the effect of wolves on ungulate (α₁)!

Something to think about

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$

$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$


```
lm(v ~ u)
```

It's also easy to calculate the effect of ungulates on vegetation (β_1)

Something to think about (we'll discuss this shortly!)

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$

$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$
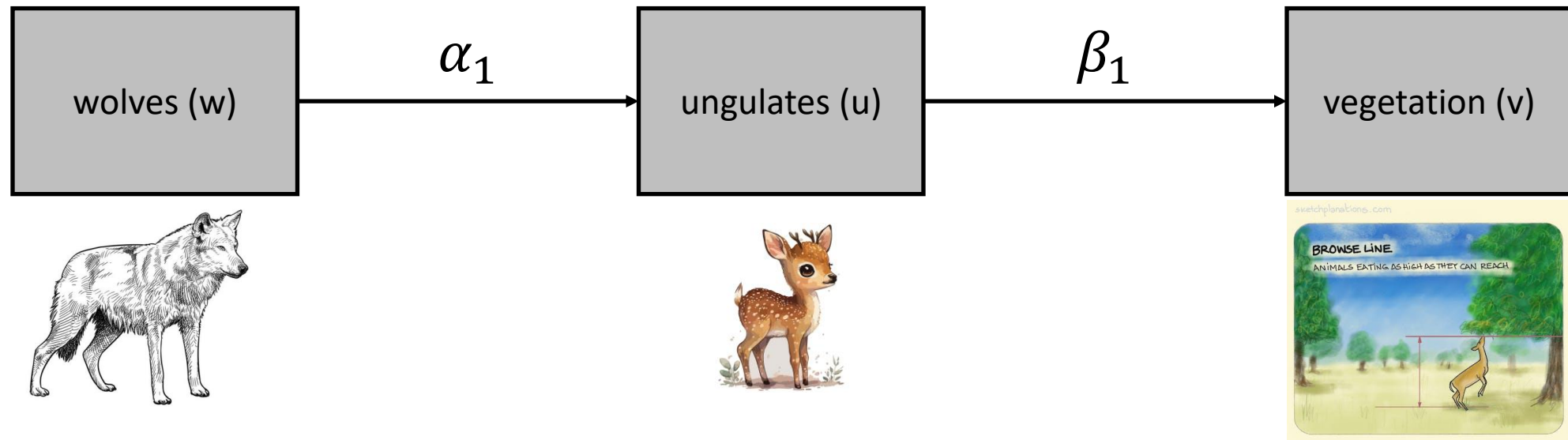
How would we calculate the effect of wolves on browse?!

How would we calculate the effect of wolves on browse?!
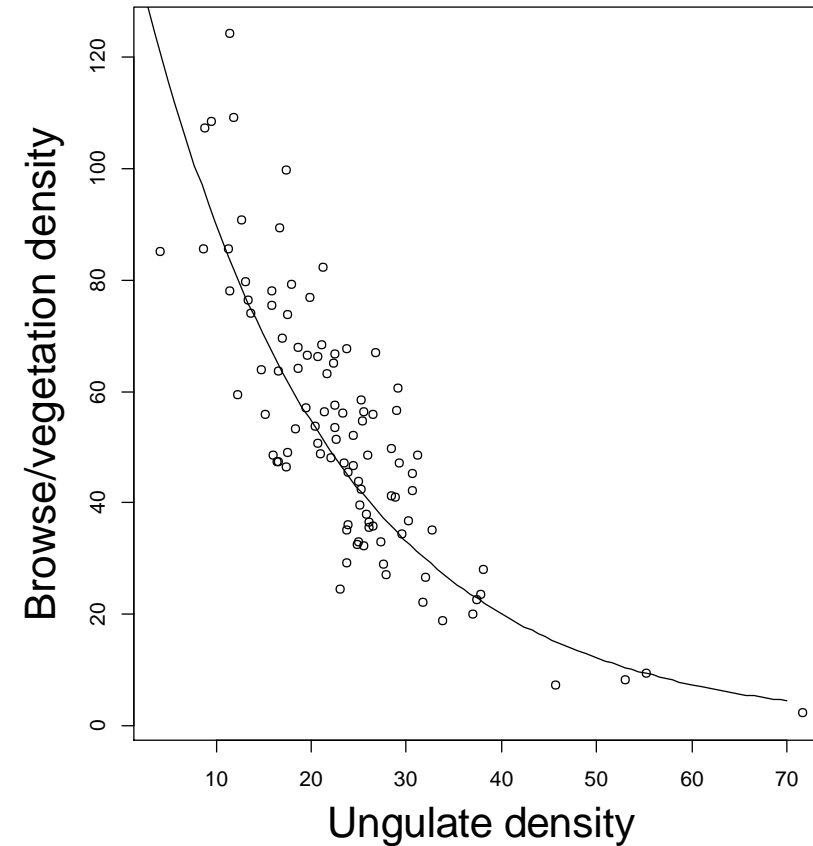
Talk to your neighbor

Let's imagine there is a hypothetical 101st study site with a current density of 2 wolves per 'unit'

1. What is the current ML estimate of ungulates and browse?

2. What would happen to ungulate density if wolves were extirpated?

3. What would happen to browse if wolves were extirpated?
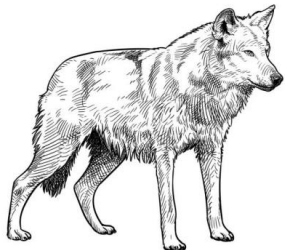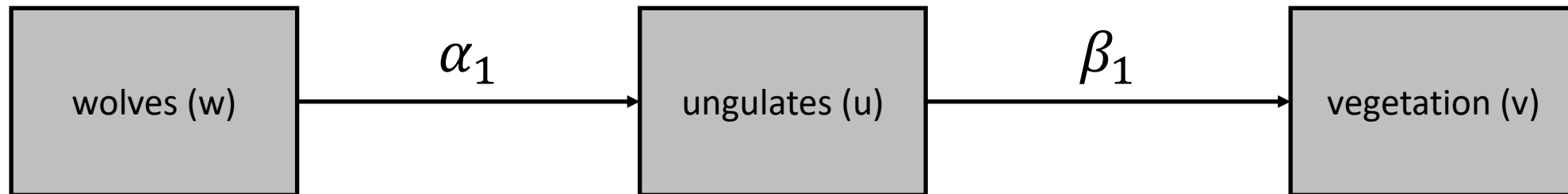
Let's visualize an indirect effect!

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2) \qquad v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$
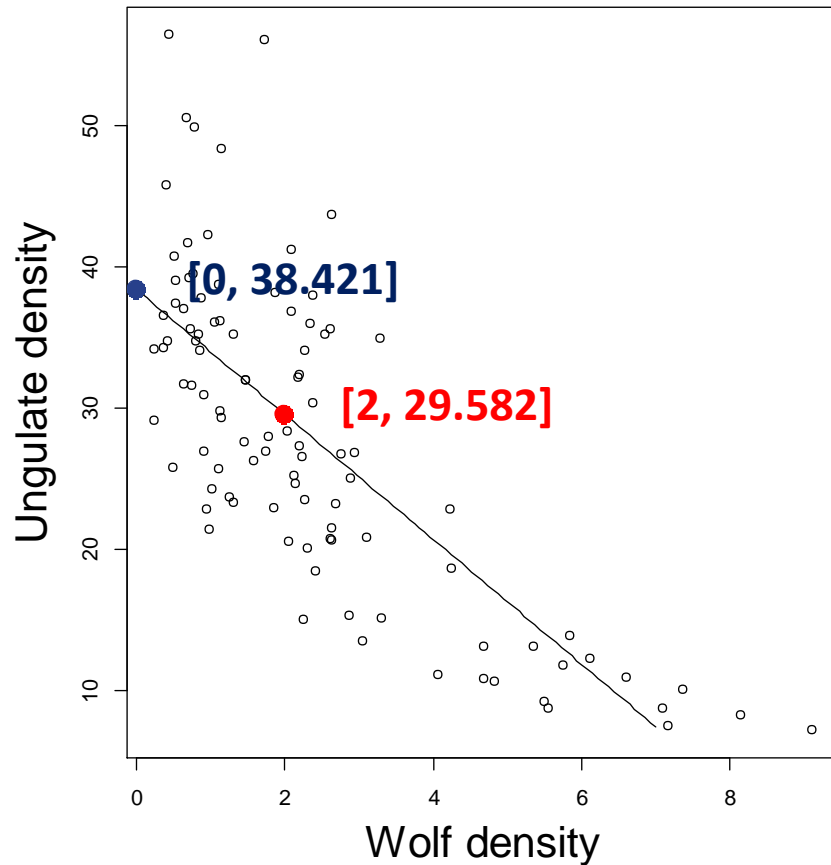
We would gain about 8.83 ungulates, going from 29.582 to 38.421.

$$\alpha_0 = 38.4211$$
$$\alpha_1 = -4.4194$$

What would happen if we went from 2 wolves to 0?

Let's visualize an indirect effect!

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2) \qquad v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$
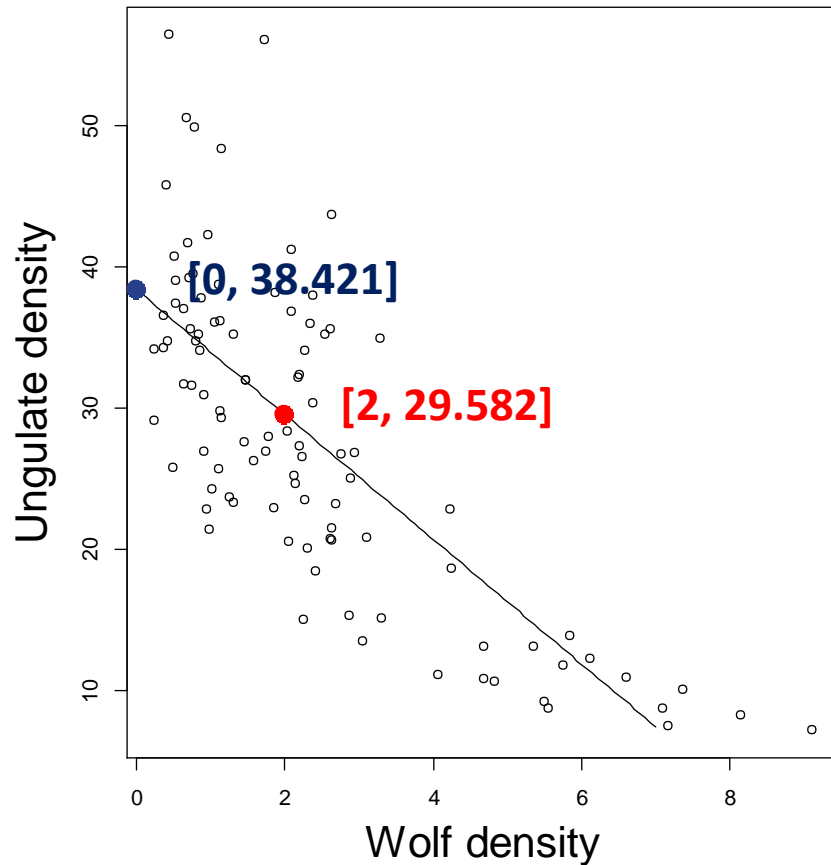
*We would gain about 8.83 ungulates, going from 29.582 to 38.421.*

$$\alpha_0 = 38.4211$$

$$\alpha_1 = -4.4194$$

How would that 8.83 gain in ungulates affect browse?

Let's visualize an indirect effect!
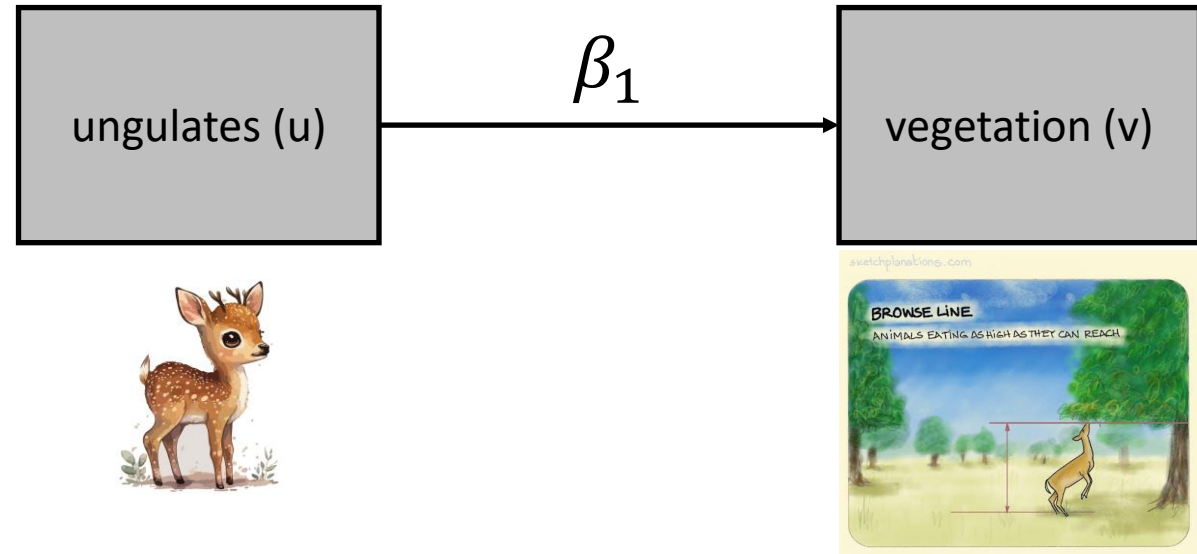
$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2) \qquad v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$
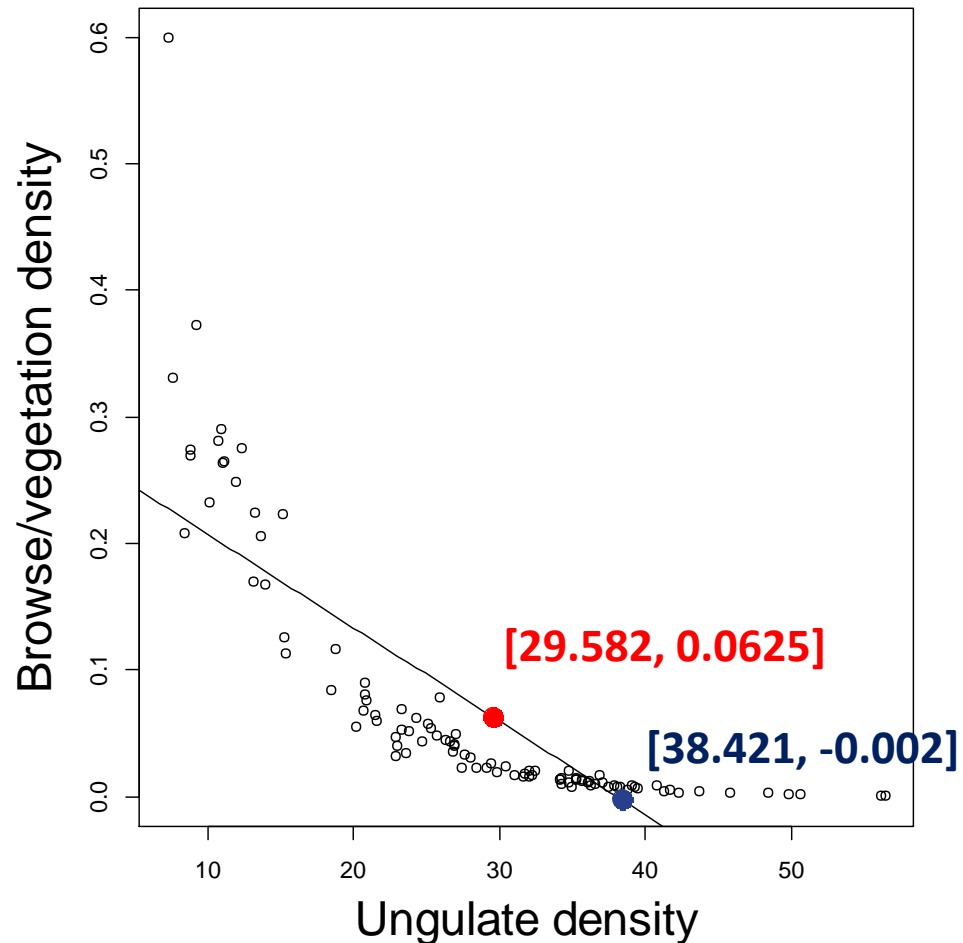
If we gain about 8.83 ungulates, we lose 'browse'.

$$\beta_0 = 0.2809$$

$$\beta_1 = -0.00738$$

Let's visualize an indirect effect!

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$
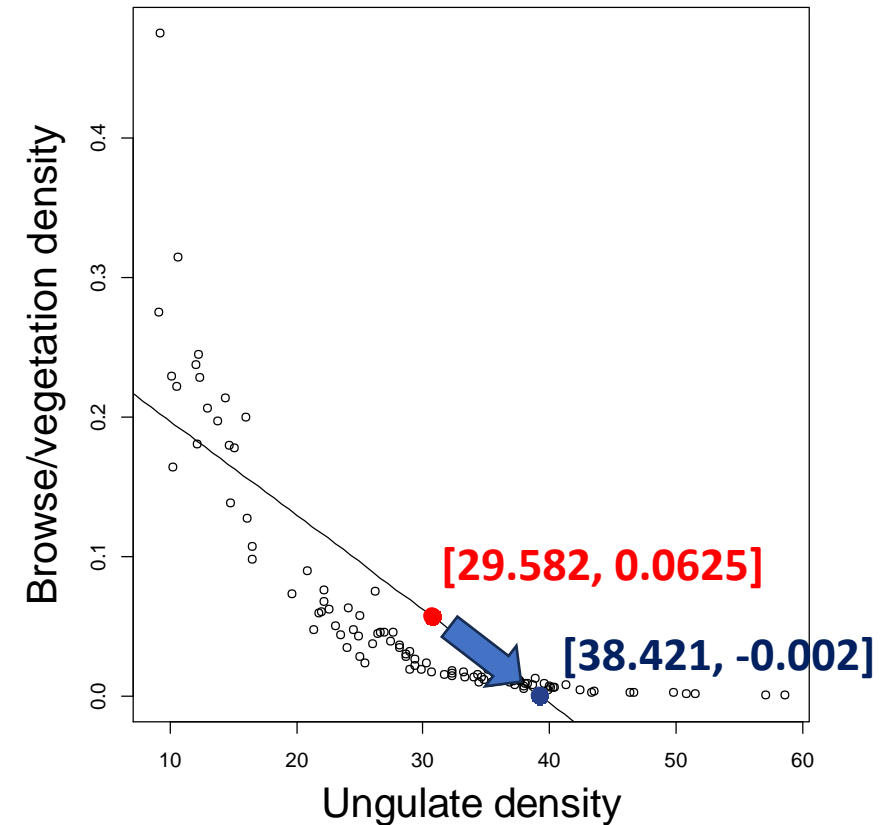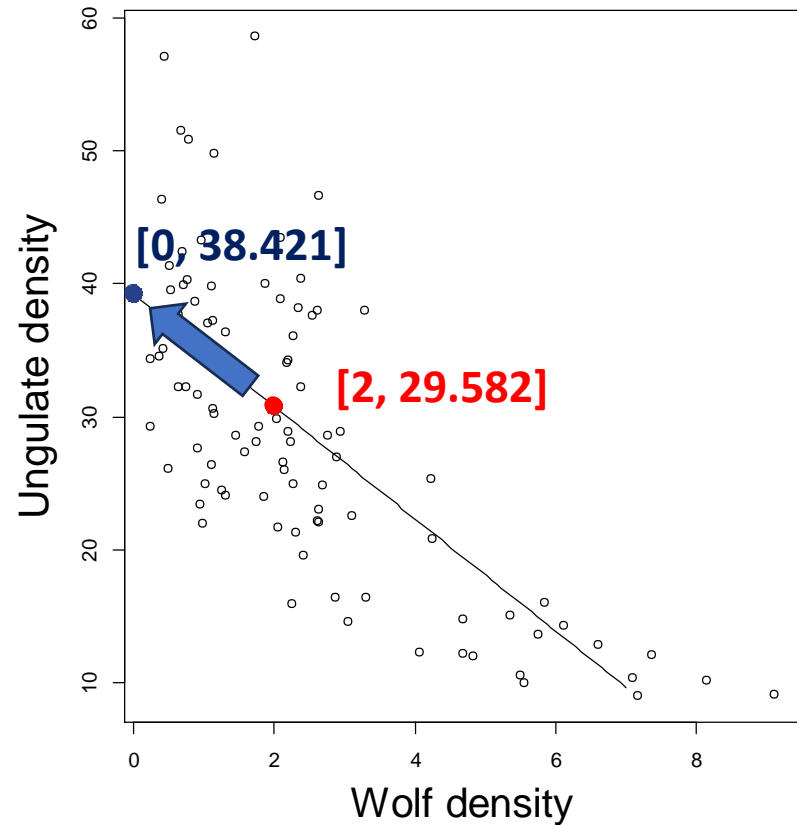
$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$

Losing wolves -> more deer; more deer -> less browse

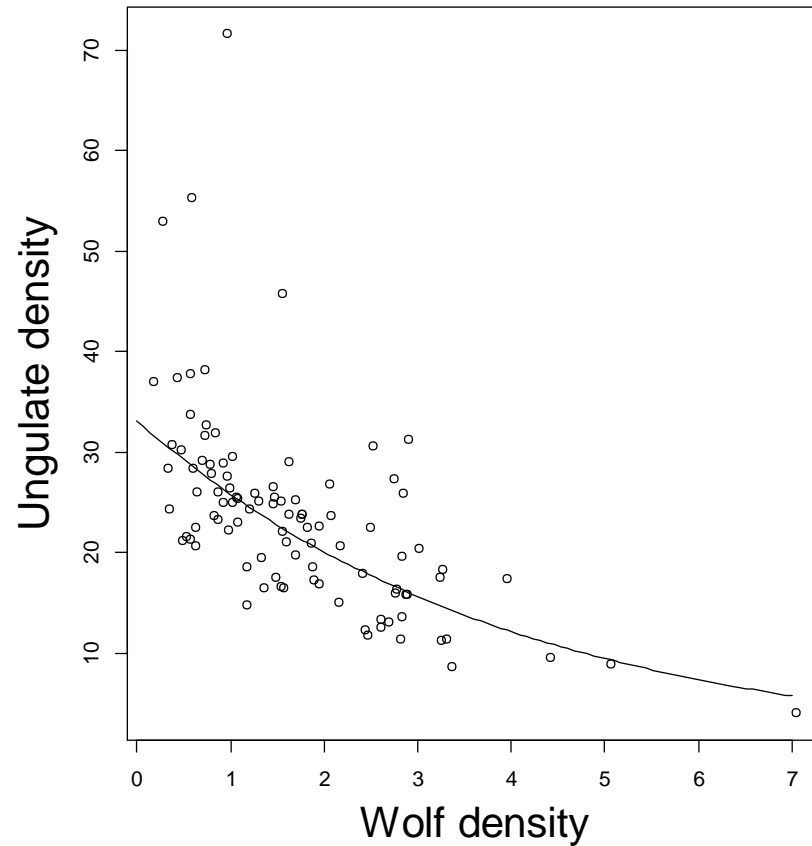# Let's visualize an indirect effect!

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2) \qquad v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$

Left plot: Ungulate density vs Wolf density, with points [0, 38.421] and [2, 29.582]

Right plot: Browse/vegetation density vs Ungulate density, with points [29.582, 0.0625] and [38.421, -0.002]
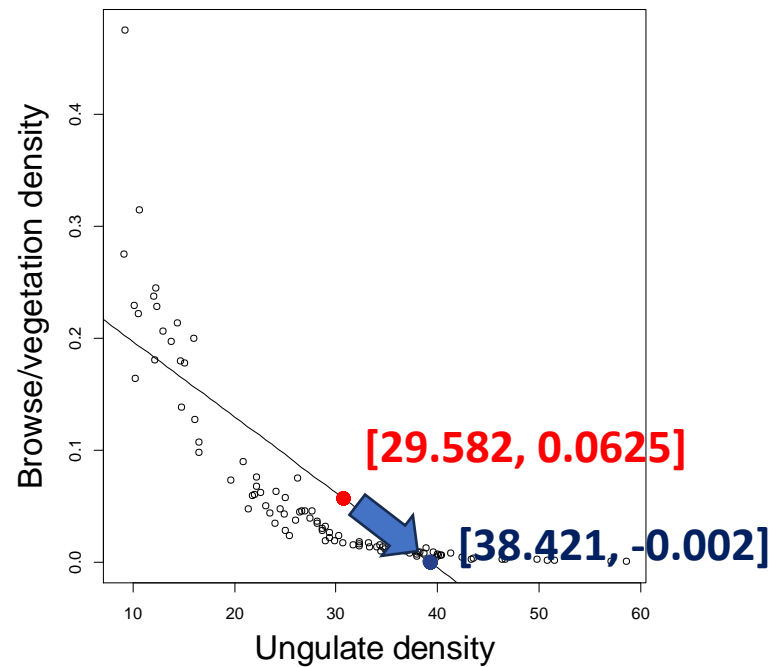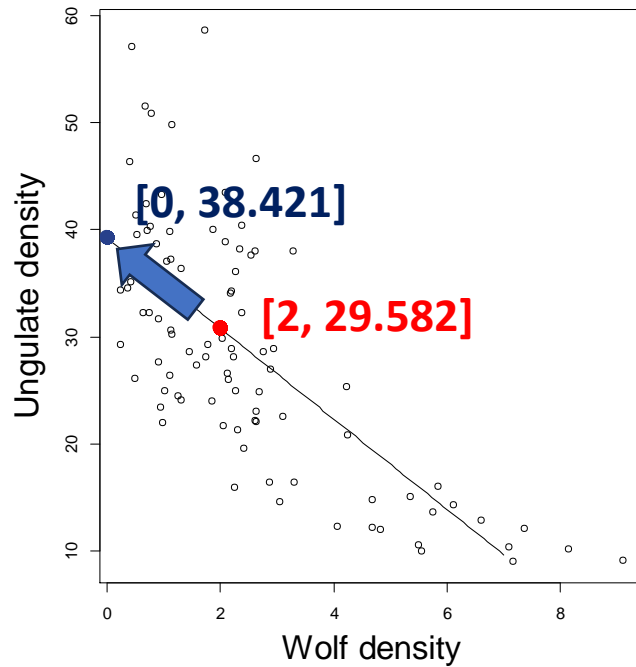
$$\beta_1 = -0.00738$$
$$\alpha_1 = -4.4194$$

$$-0.06524 = \boxed{-0.002 - 0.0625}$$

Change in browse

Let's visualize an indirect effect!

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2) \qquad v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$

$$\beta_1 = -0.00738$$

$$\alpha_1 = -4.4194$$

[0, 38.421]

[2, 29.582]

[29.582, 0.0625]

[38.421, -0.002]

$$-0.06524 = \boxed{-0.002 - 0.0625} = \alpha_1(\beta_1)(-2)$$

Change in browse

Take-home 1: indirect effects aren't as 'horrifying' as they seem

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$
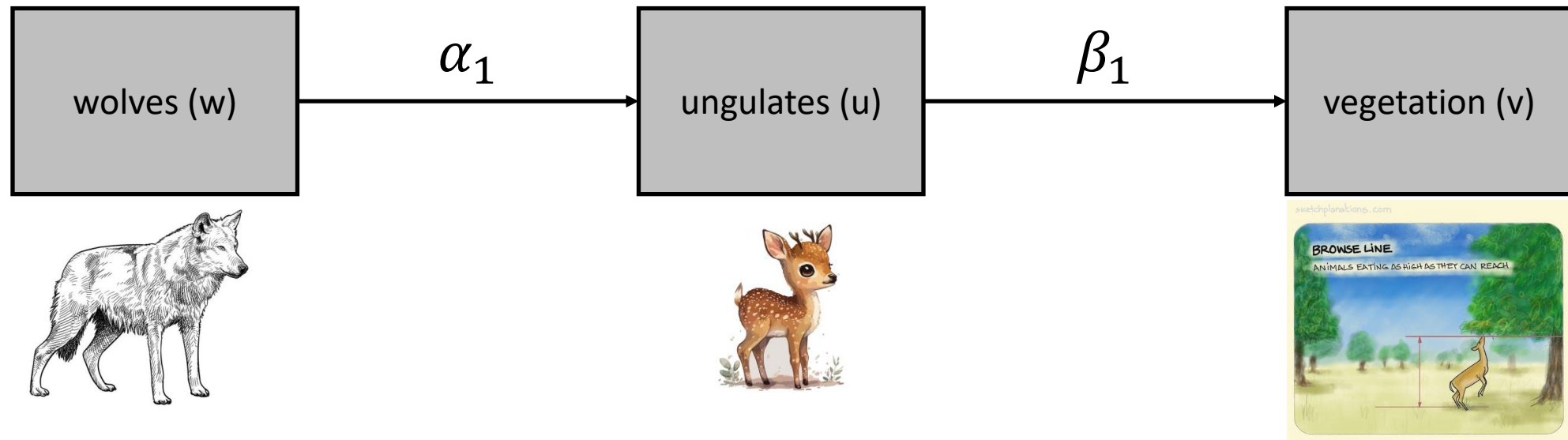
$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$

i.e., if identity link, then effect of wolves on veg = α₁β₁

$$E(v_i|w_i) = [\boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{\alpha_0})] + \alpha_1\beta_1 w_i$$

slope $= \alpha_1\beta_1$

intercept $=$
what value of veg would be
given number of deer if
there were no wolves
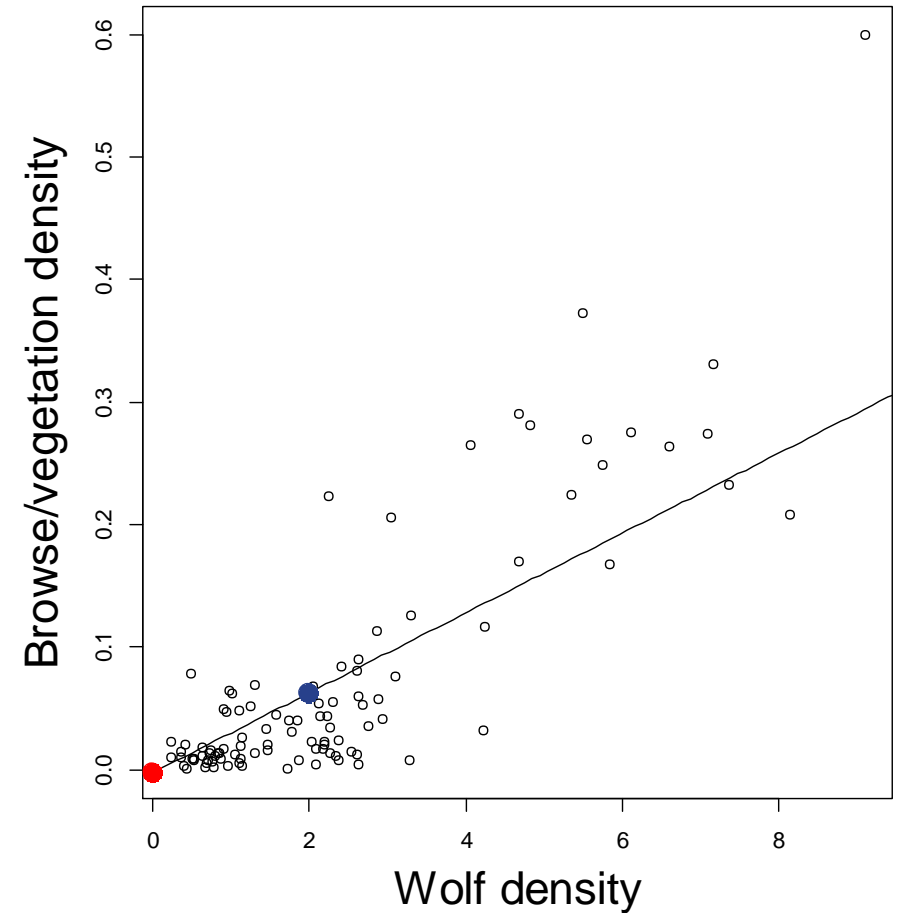
Take-home 1: this is easy?!

$$u_i \sim \text{normal}(\alpha_0 + \alpha_1 w_i, \sigma_u^2)$$

$$v_i \sim \text{normal}(\beta_0 + \beta_1 u_i, \sigma_v^2)$$
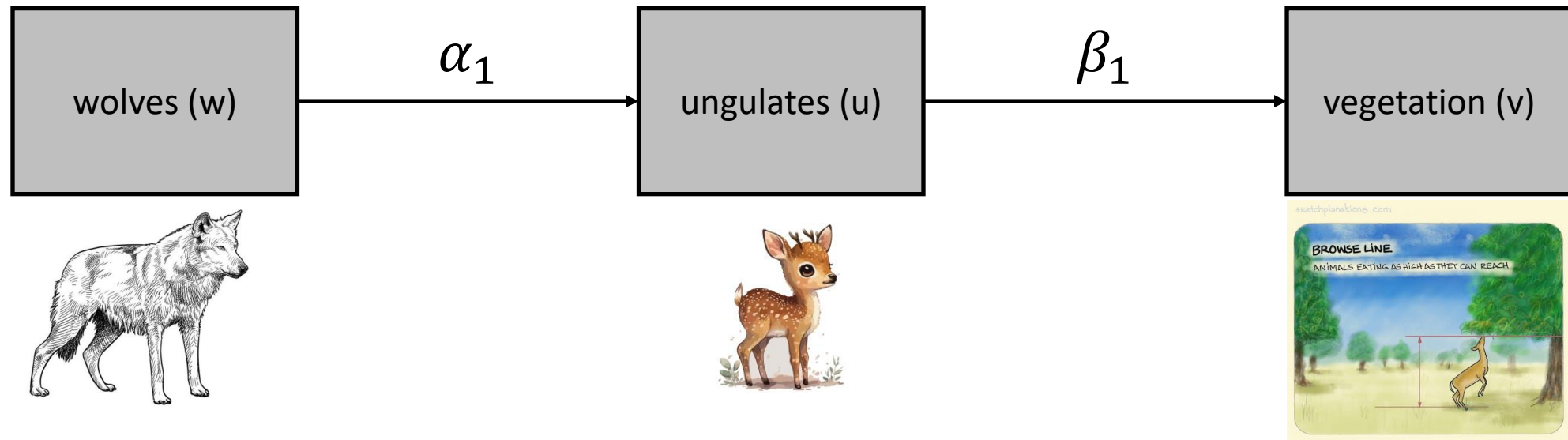
i.e., we're more or less 'stapling' two linear models together

Our first analysis: wolf resource selection in Bow Valley

HUMAN ACTIVITY MEDIATES A TROPHIC CASCADE
CAUSED BY WOLVES

MARK HEBBLEWHITE,[1,7] CLIFFORD A. WHITE,[2,3] CLIFFORD G. NIETVELT,[1] JOHN A. MCKENZIE,[4,5]
TOMAS E. HURD,[2] JOHN M. FRYXELL,[4] SUZANNE E. BAYLEY,[1] AND PAUL C. PAQUET[6]

Lecture2c_wolf_rsf.R

Hebblewhite **(2000)**

Collider bias: wolf resource selection in Bow Valley

- *The black line is the direct effect of elevation holding deer abundance constant at the mean.
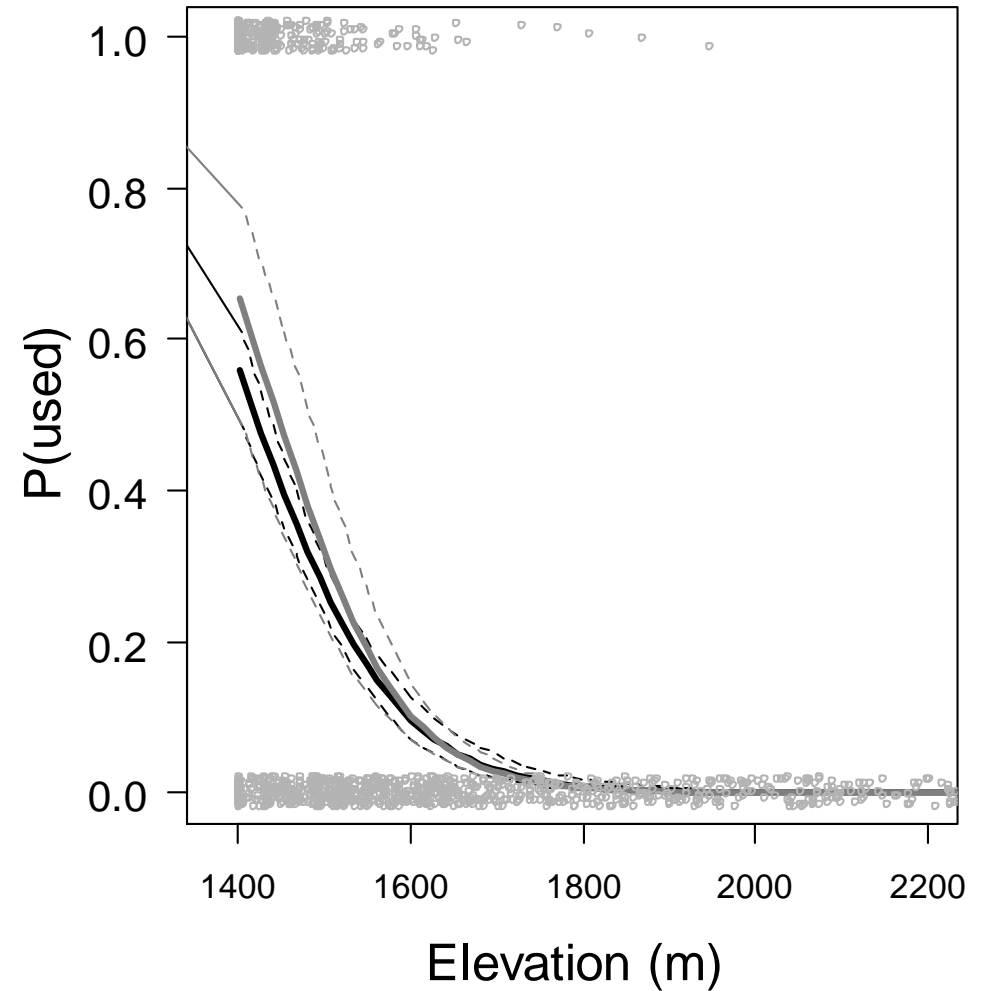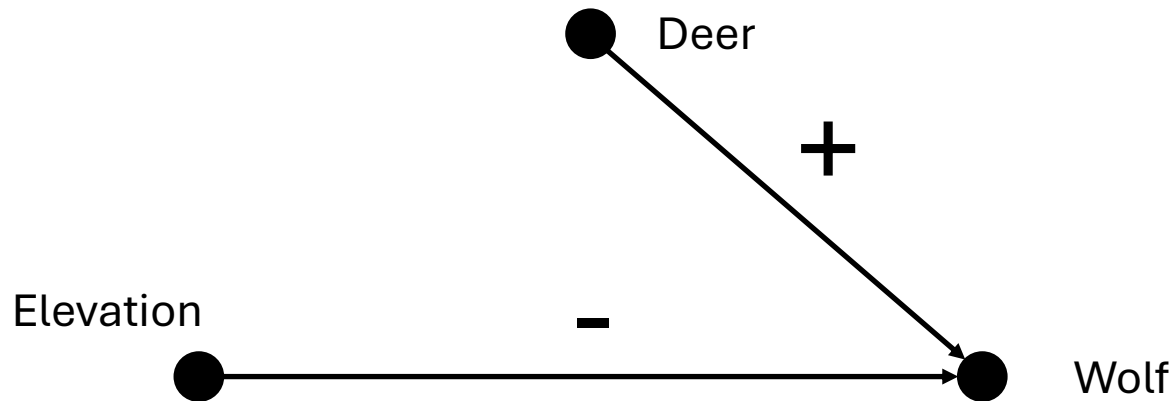
*Deer abundance will not remain constant as elevation changes

- The black line is the direct effect of elevation holding deer abundance constant at the mean.

- The grey line is the total effect of elevation via formally acknowledging that increases in elevation will reduce deer abundance.
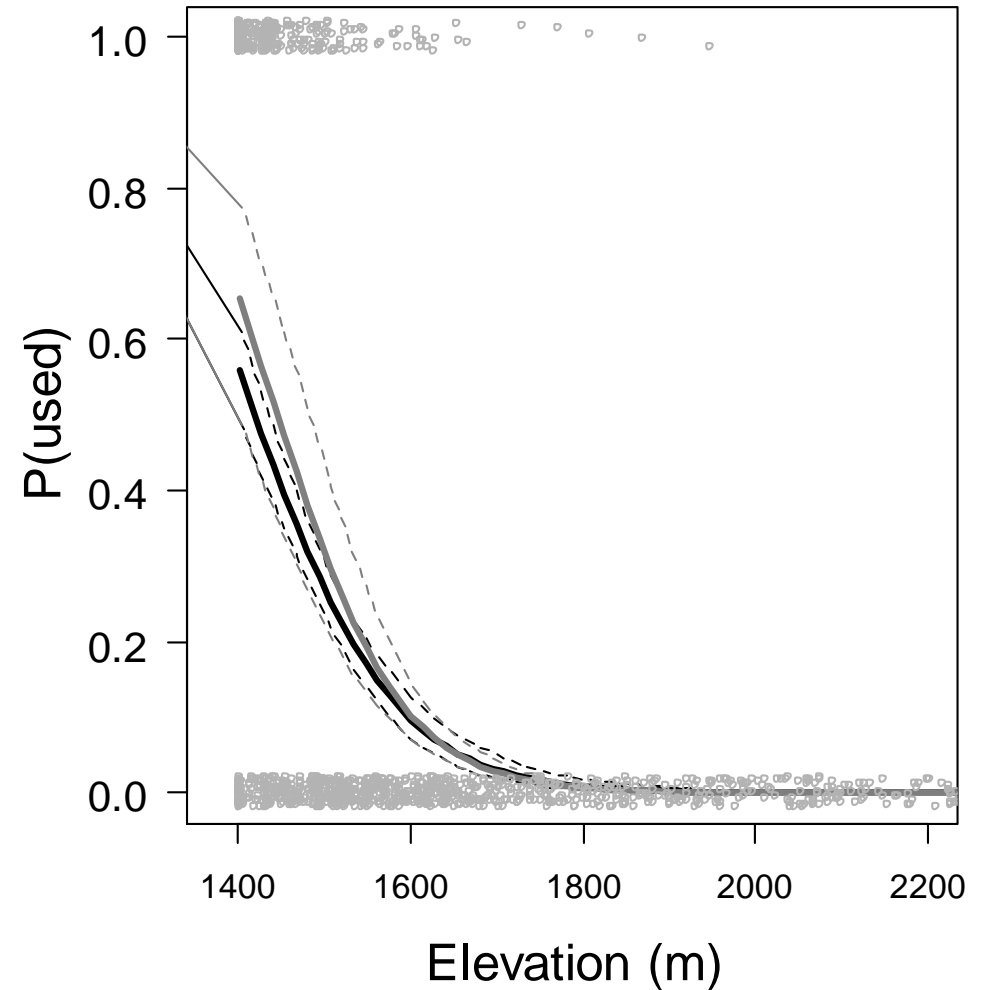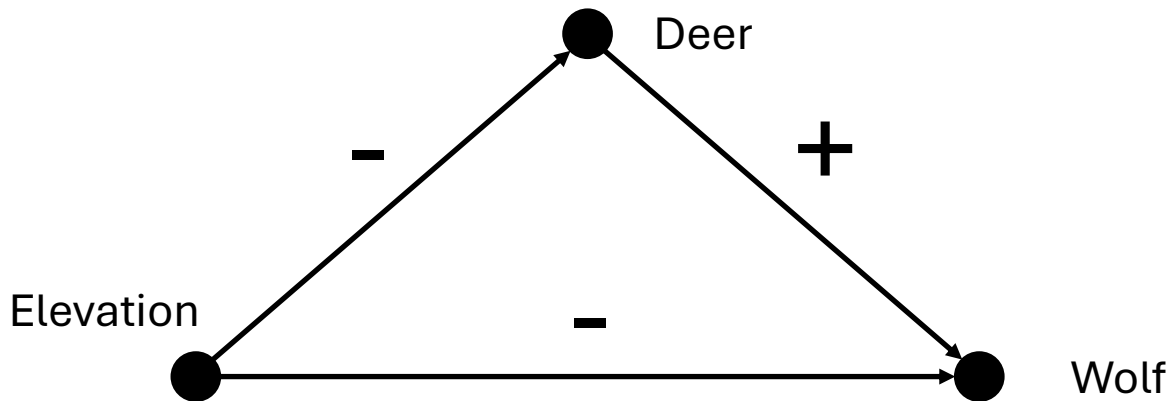
Note that a path analysis and simple glm(used ~ elevation) are similar.

- The grey line is the total effect of elevation via formally acknowledging that increases in elevation will reduce deer abundance.

- The red line is the stand-alone effect of elevation from a model w/o deer
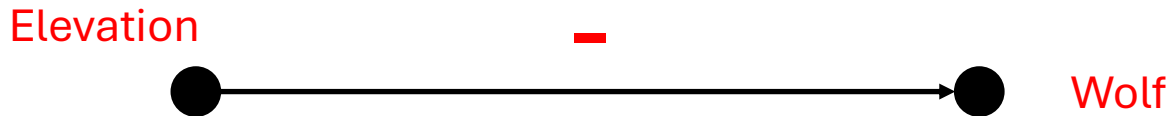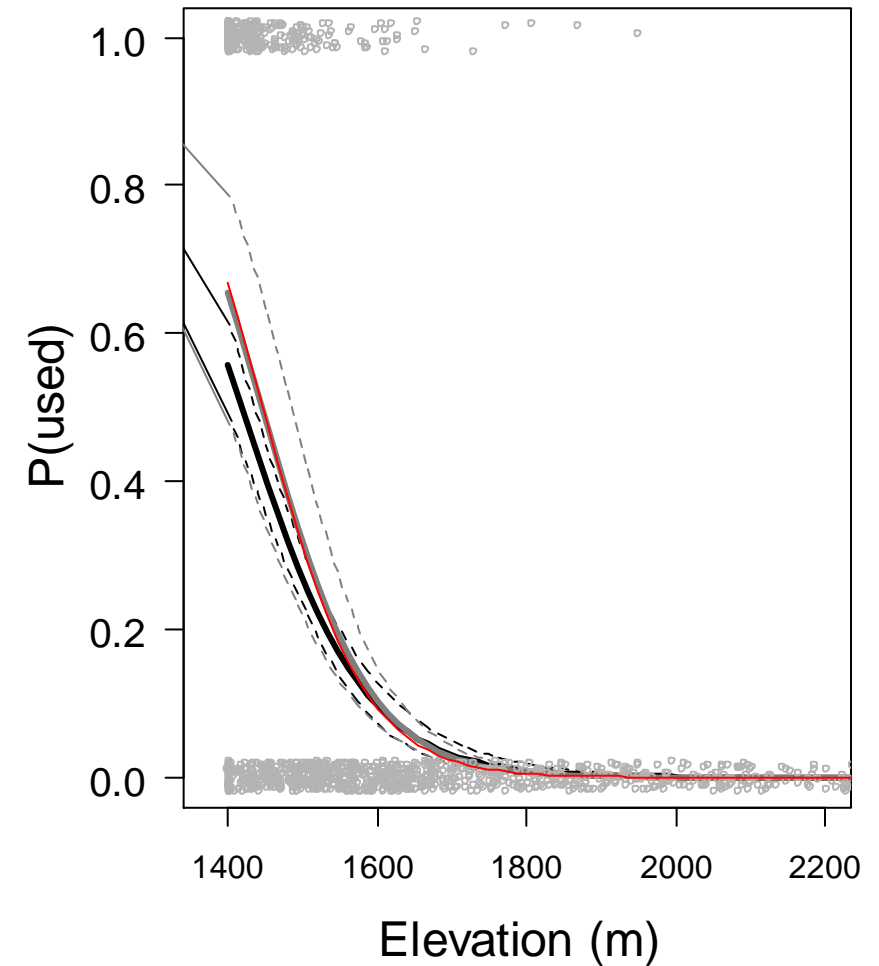
Why is this important?

- AIC (deer + elevation) = 892.93
- AIC (elevation) = 937.86

Why is this important?

Canopy

-0.2

0.4

Fire
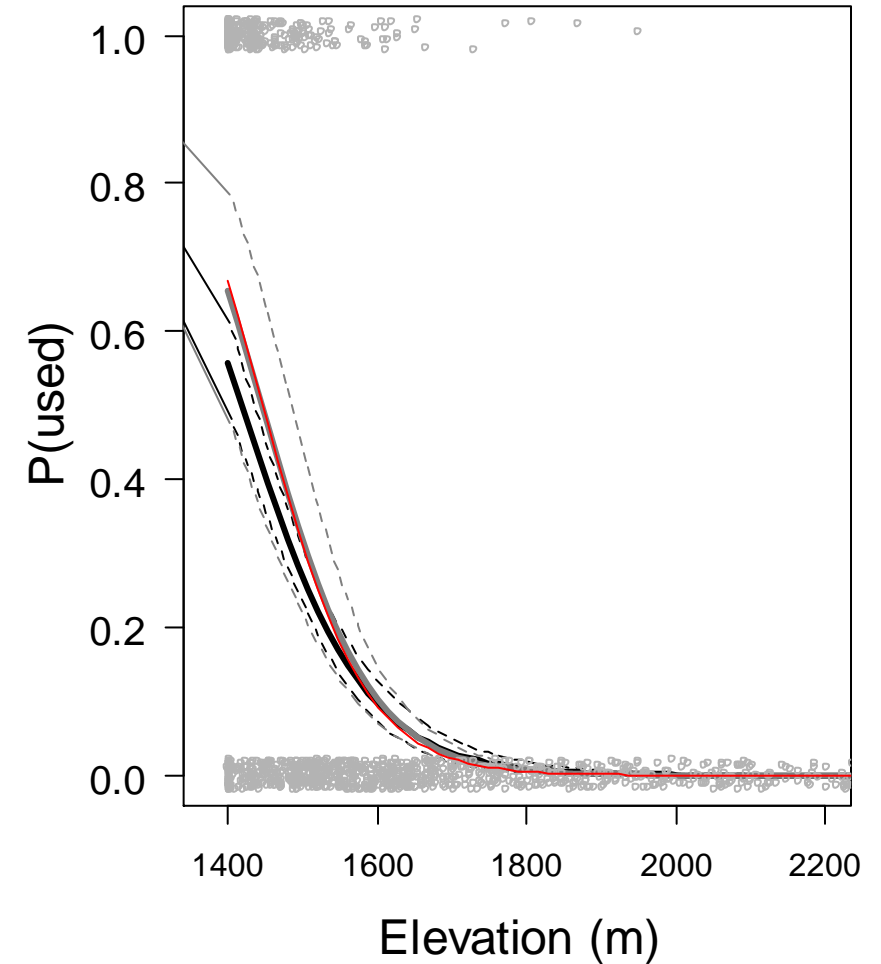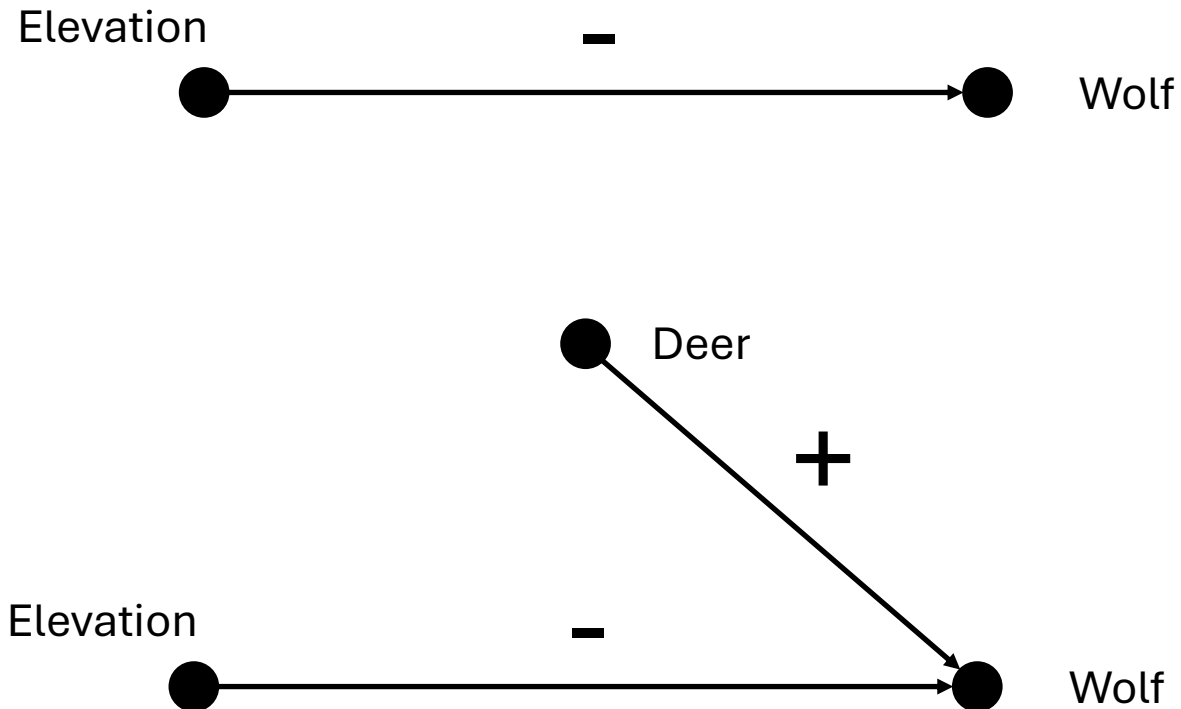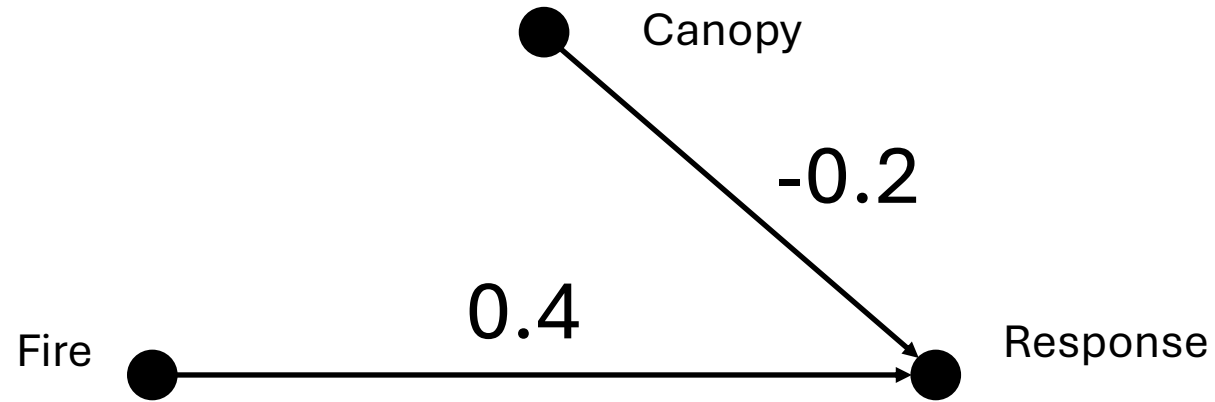
Response

Prescribed burns were associated with an increase in our ecological response variable of interest ($\beta = 0.4$; $p < 0.01$). Canopy cover negatively effected our response variable of interest (-0.2, $p < 0.01$).

Imagine you're interested in the impact of a management action

Why is this important?

Canopy

-1

-0.2

0.4
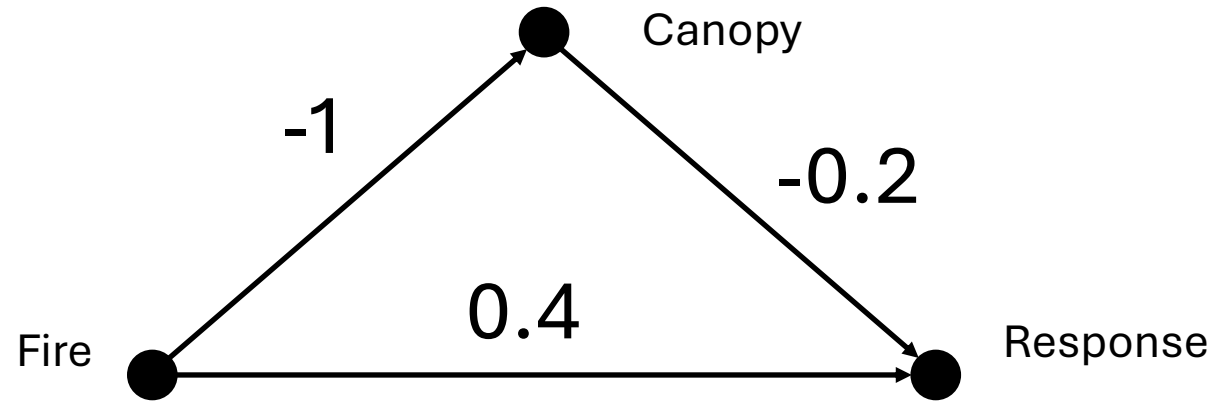
Fire   Response

Prescribed burns had a positive direct effect on our ecological response variable of interest ($\beta = 0.4$; $p < 0.01$). Burns also reduced canopy cover ($\beta = -1$; $p < 0.01$), which negatively effected our response variable of interest (-0.2, $p < 0.01$) **leading to a total increase of 0.6.**

The total effect is 50% greater than the direct effect!

Why is this important?
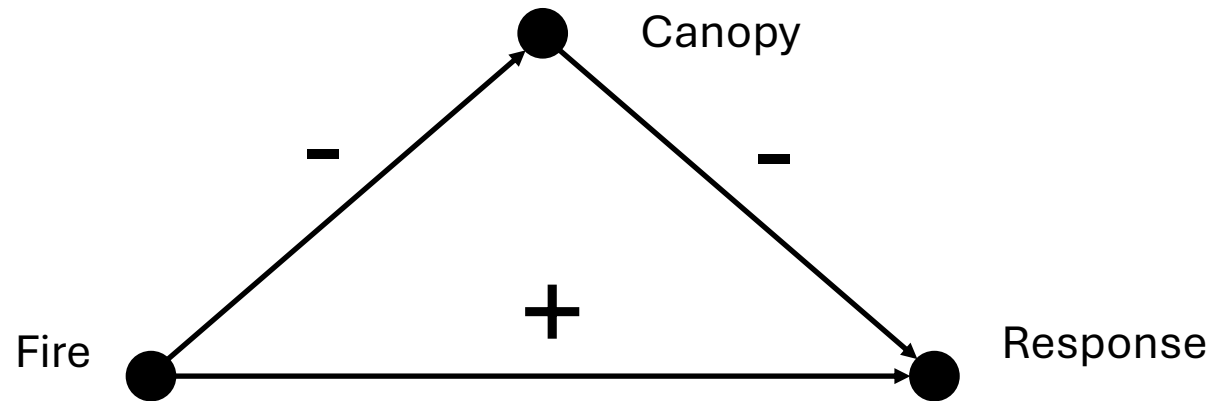
Inference can change substantially via causal hypotheses

When we 'fix' independent variables to estimate effects
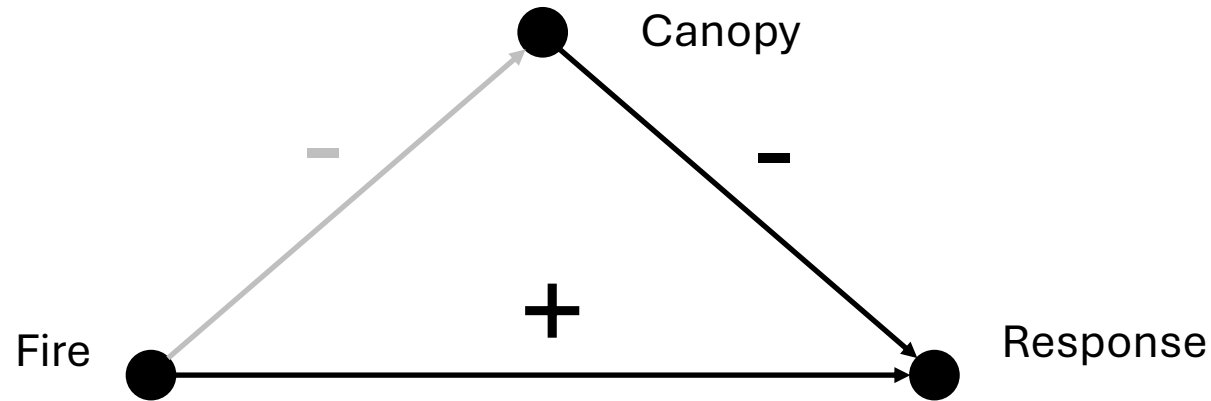
We ignore the fact that changes in one covariate can affect another

This can lead to absurdity!
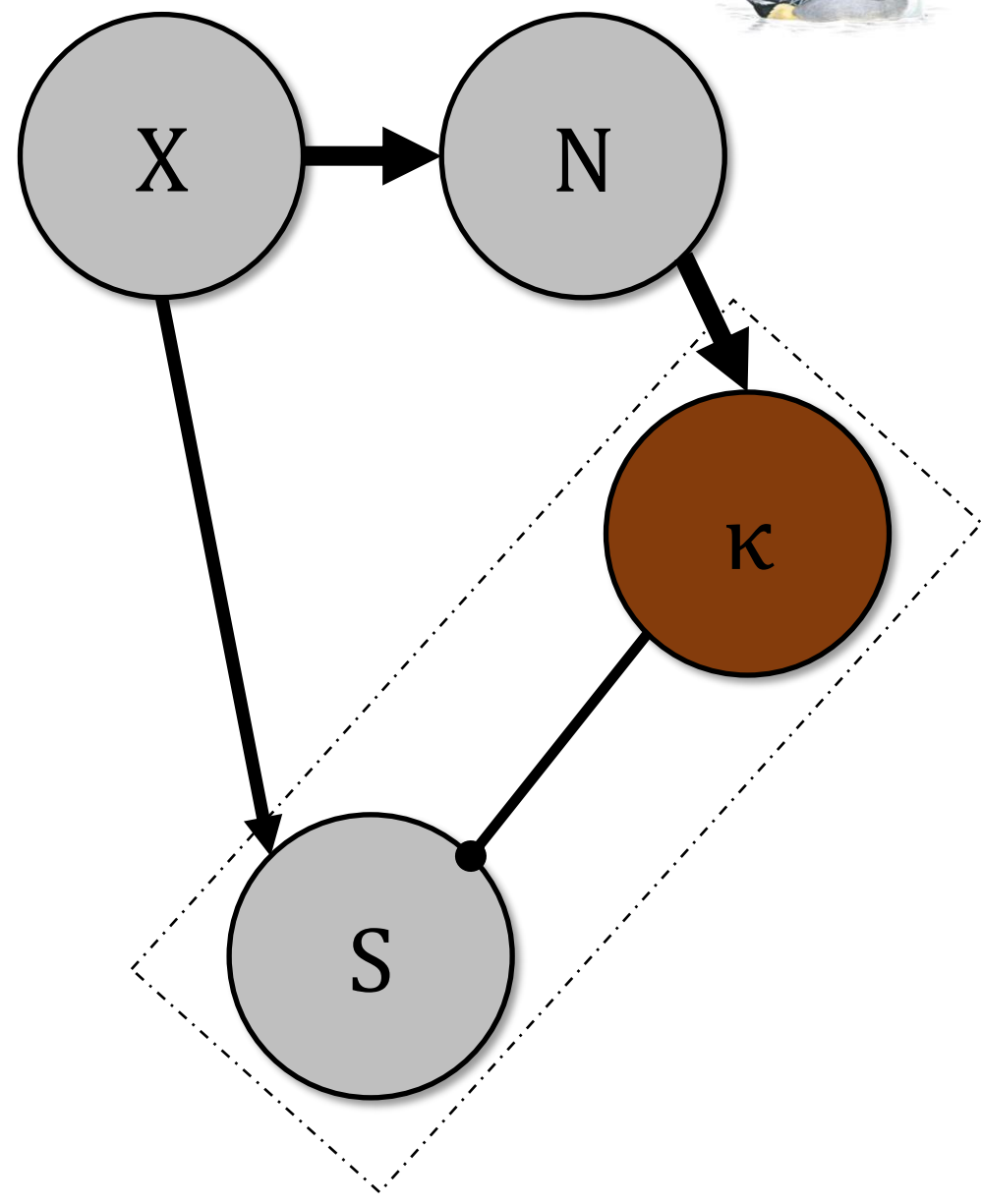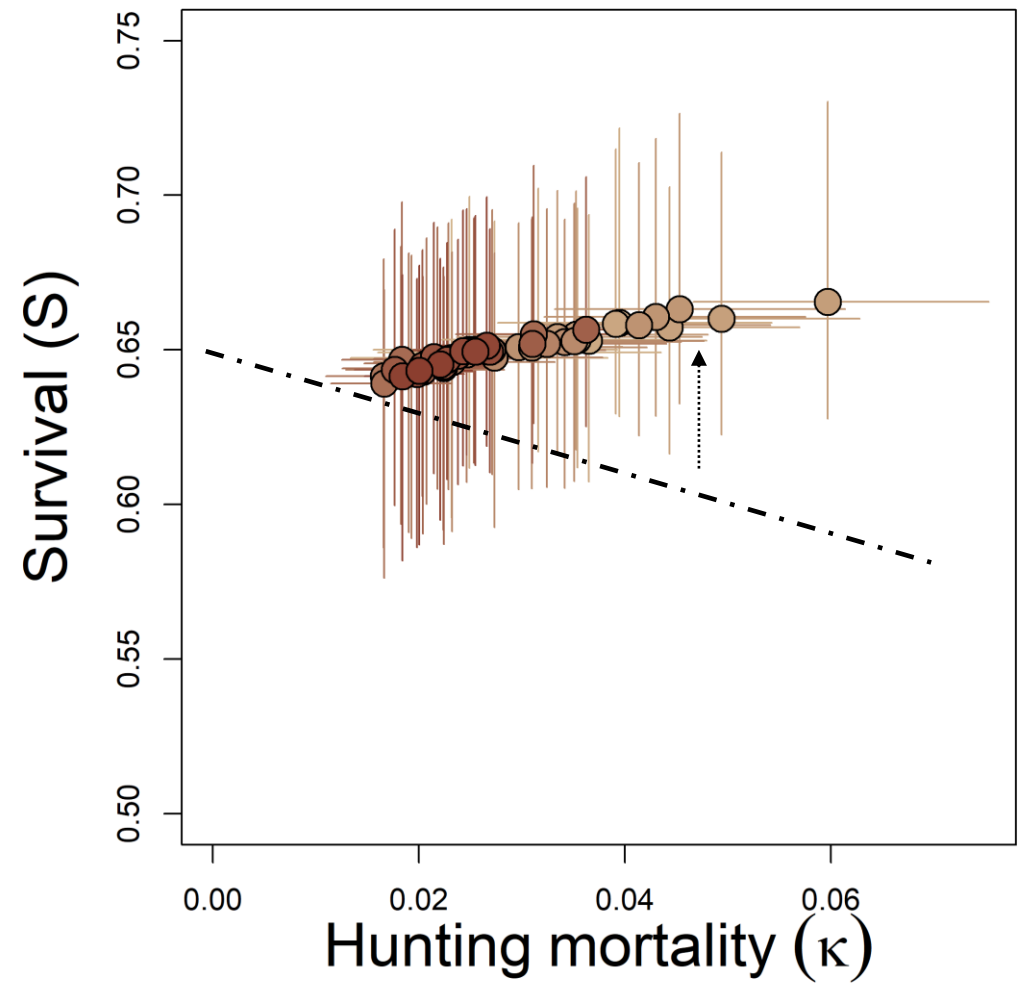
It's more likely that it will lead to smaller yet important changes
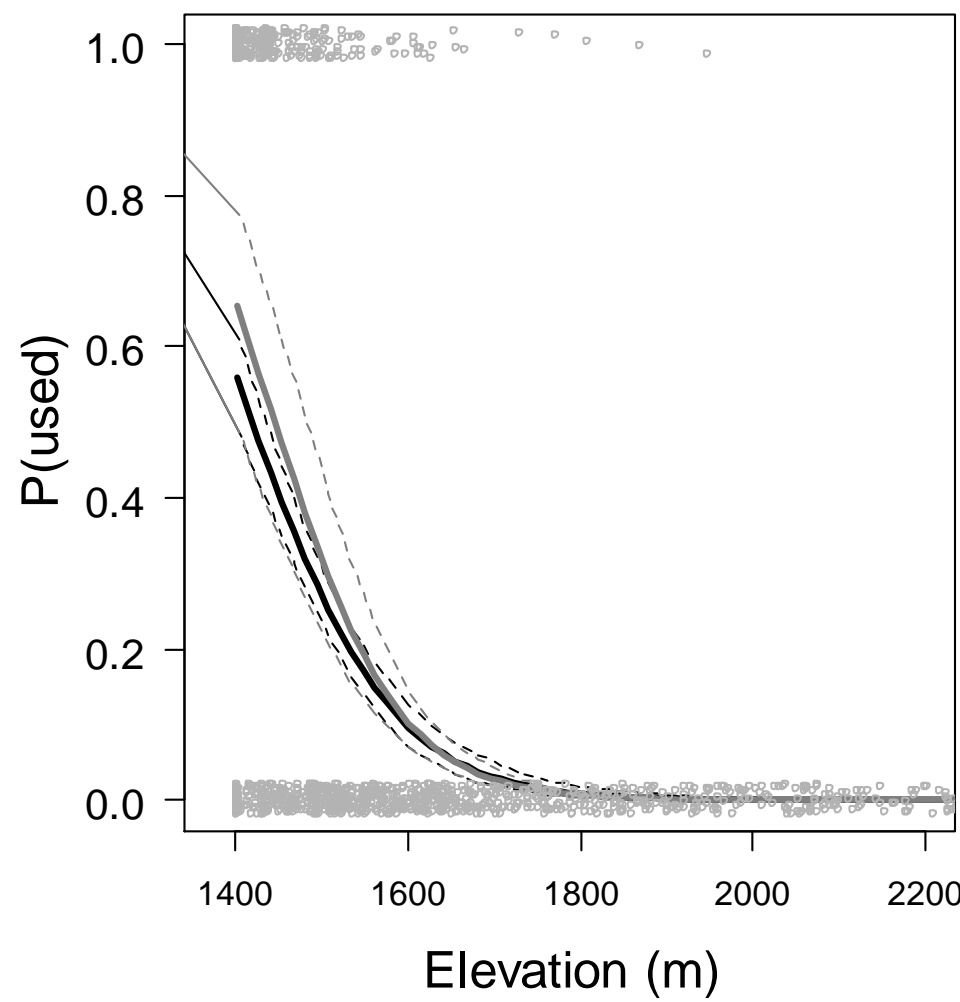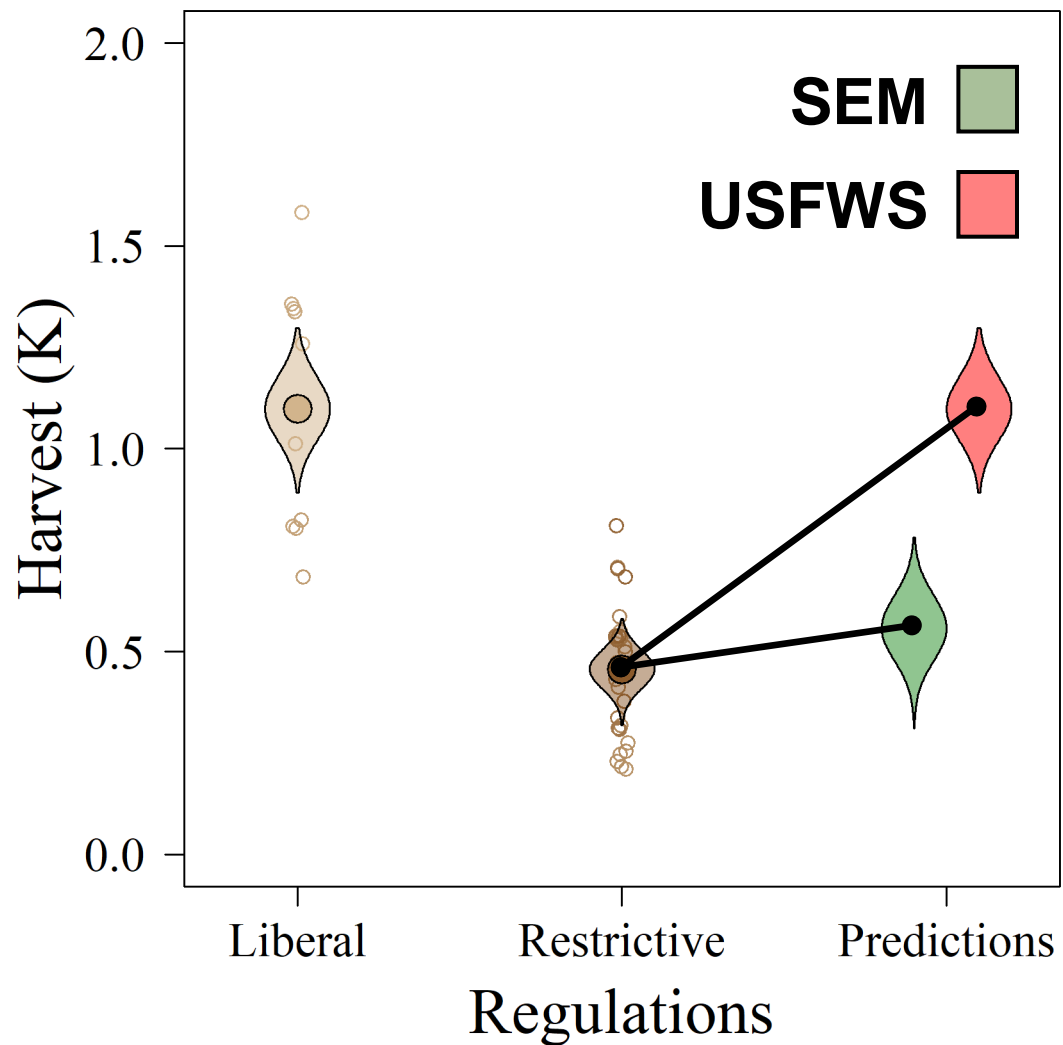
We can build substantially more complex models.

Density-dependence produces spurious relationships among demographic parameters in a harvested species

Thomas V. Riecke[1,2,3] | Madeleine G. Lohman[1,2] | Benjamin S. Sedinger[1,2,4] | Todd W. Arnold[5] | Cliff L. Feldheim[6] | David N. Koons[7] | Frank C. Rohwer[8] | Michael Schaub[3] | Perry J. Williams[2] | James S. Sedinger[2]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, Nevada, USA; [2]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Nevada, USA; [3]Swiss Ornithological Institute, Sempach, Switzerland; [4]University of Wisconsin-Stevens Point, Stevens Point, Wisconsin, USA; [5]Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St. Paul, Minnesota, USA; [6]California Trout, San Francisco, California, USA; [7]Fish, Wildlife, and Conservation Biology & Graduate Degree Program in Ecology, Colorado State University, Ft. Collins, Colorado, USA and [8]Delta Waterfowl Foundation, Bismarck, North Dakota, USA

RIECKE ET AL.   Journal of Animal Ecology   2265

$$\eta_t = (1 - \kappa_t)(1 - e^{-h_{\eta,t}})$$
$$s_t = 1 - \eta_t - \kappa_{t+1}$$
$$\kappa_t = 1 - e^{-h_{\kappa,t}}$$

FIGURE 2   A directed acyclic graph demonstrating the relationships among abundance (**n**), ponds (**p**; blue), fecundity (ξ), hunting mortality hazard rate (**h_κ**), natural mortality hazard rate (**h_η**), survival (**s**) and the number of duck hunters (**H**; brown) for blue-winged teal breeding in the North American Prairie Pothole Region across the annual cycle (1973–3016). Solid arrows represent estimated directional relationships, and dashed arrows represent processes leading to changes in population abundance.

The same assumptions inherent to linear models apply, plus…

The importance of causal diagrams

Journal of Animal Ecology

RESEARCH ARTICLE

Density-dependence produces spurious relationships among demographic parameters in a harvested species

Thomas V. Riecke[1,2,3] | Madeleine G. Lohman[1,2] | Benjamin S. Sedinger[1,2,4] | Todd W. Arnold[5] | Cliff L. Feldheim[6] | David N. Koons[7] | Frank C. Rohwer[8] | Michael Schaub[3] | Perry J. Williams[2] | James S. Sedinger[2]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, Nevada, USA; [2]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Nevada, USA; [3]Swiss Ornithological Institute, Sempach, Switzerland; [4]University of Wisconsin–Stevens Point, Stevens Point, Wisconsin, USA; [5]Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St. Paul, Minnesota, USA; [6]California Trout, San Francisco, California, USA; [7]Fish, Wildlife, and Conservation Biology & Graduate Degree Program in Ecology, Colorado State University, Ft. Collins, Colorado, USA and [8]Delta Waterfowl Foundation, Bismarck, North Dakota, USA

RIECKE ET AL. | Journal of Animal Ecology | 2265

$$\eta_t = (1 - \kappa_t)(1 - e^{-h_{\eta,t}})$$
$$s_t = 1 - \eta_t - \kappa_{t+1}$$
$$\kappa_t = 1 - e^{-h_{\kappa,t}}$$
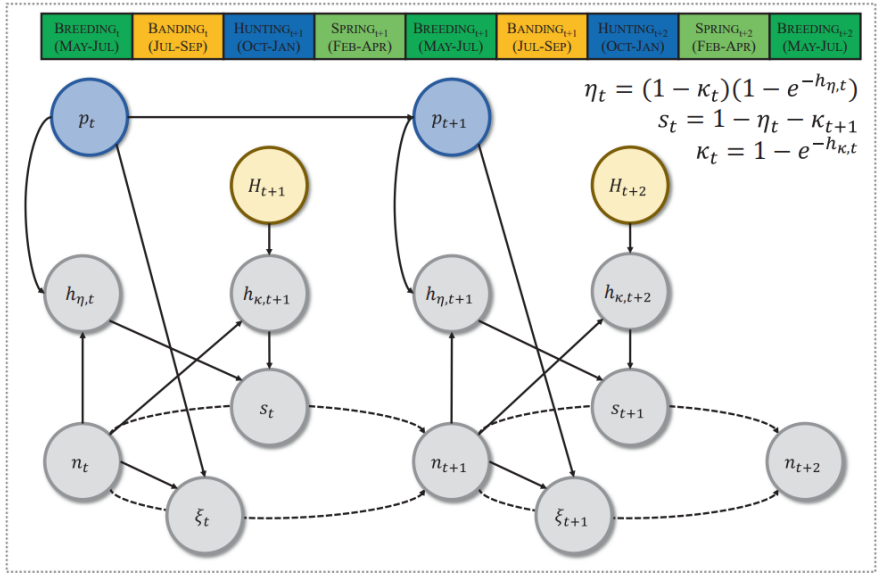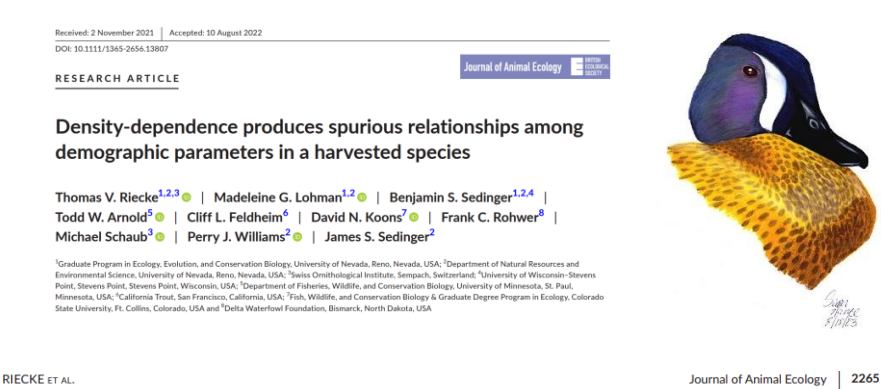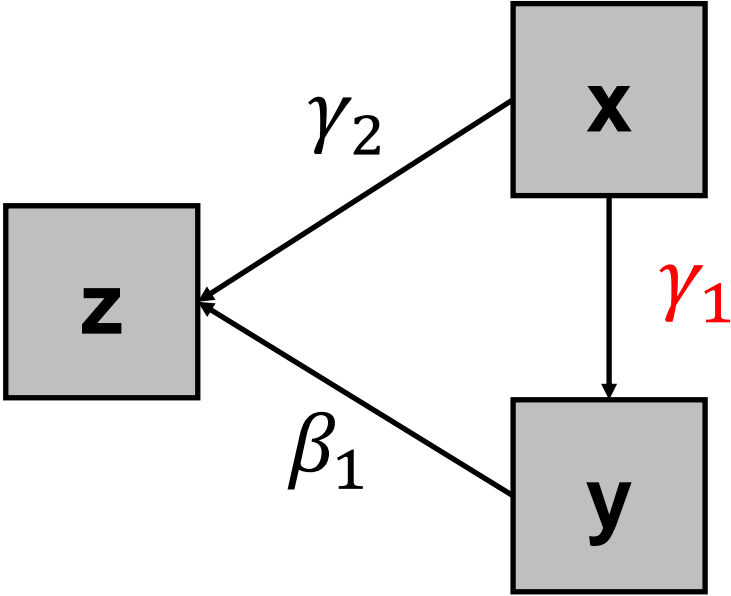
FIGURE 2 A directed acyclic graph demonstrating the relationships among abundance (*n*), ponds (*p*; blue), fecundity (*ξ*), hunting mortality hazard rate (*h_κ*), natural mortality hazard rate (*h_η*), survival (*s*) and the number of duck hunters (*H*; brown) for blue-winged teal breeding in the North American Prairie Pothole Region across the annual cycle (1973–3016). Solid arrows represent estimated directional relationships, and dashed arrows represent processes leading to changes in population abundance.

```
104   path1 <- psem(
105      lm(y ~ x, data = d),
106      lm(z ~ y + x, data = d),
107      data = d
108   )
109   summary(path1)
```

Next up, latent variables…