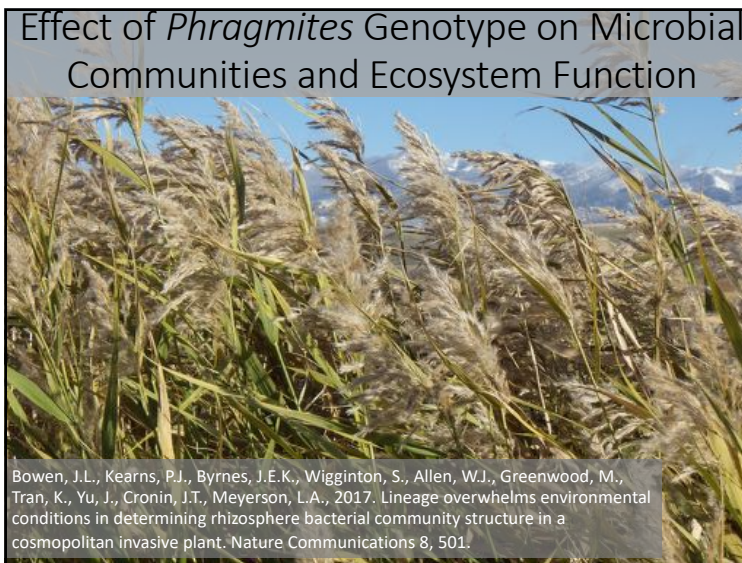
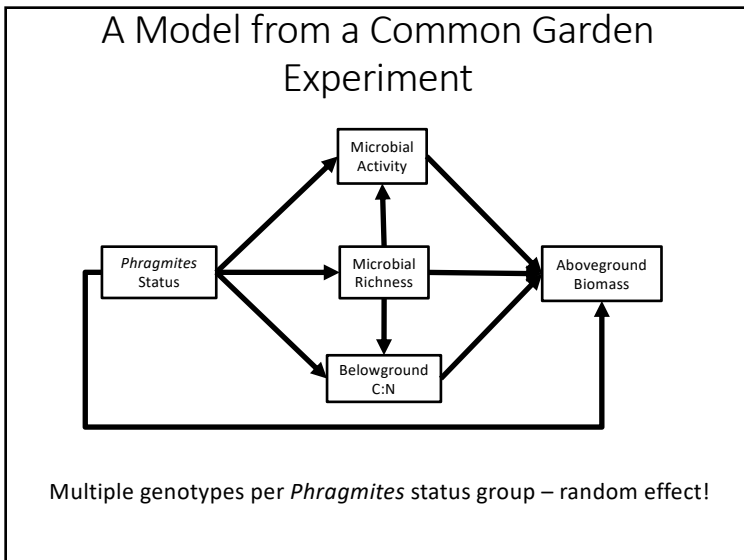
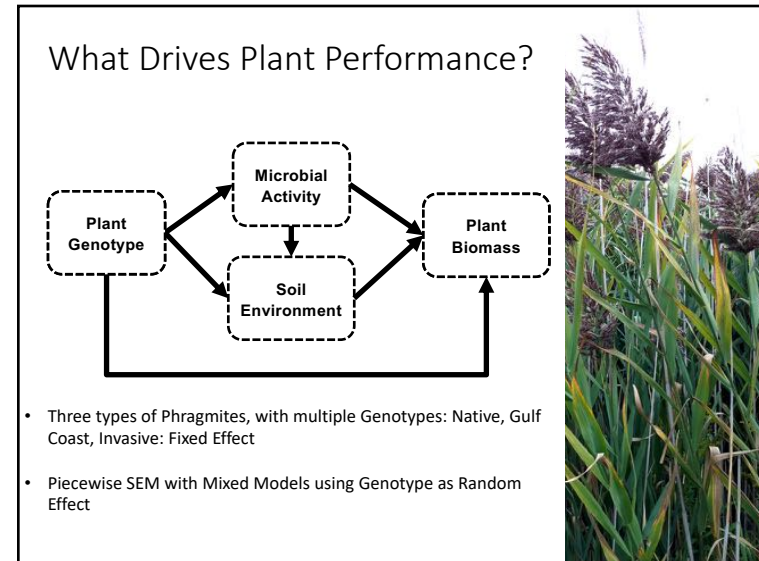
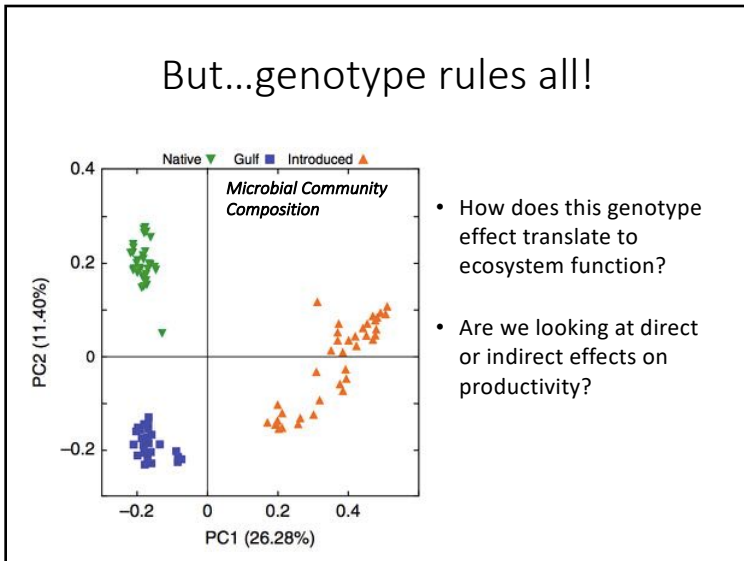


- ### Categorizing our SEMs
1. Categorical Predictors
  2. Multigroup Analysis and pooling data from multiple sources
  3. Multigroup analysis as model-wide interaction effect
  4. General technique of imposing & releasing constraints
  5. Multigroup Analysis in R



### Multiple *Phragmites* Genotypes Across the US

- Local environment should shape soil microbial communities
- These communities should regulate ecosystem function



### The Four Submodels

```

bowen <- read.csv("../data/bowen.csv")
library(nlme)

###
# A categorical model
###
div_mod <- lme(observed_otus ~ status,
              random =~ 1|Genotype,
              data = bowen, method = "ML")

activity_mod <- lme(RNA.DNA ~ status + observed_otus,
                  random =~ 1|Genotype,
                  data = bowen, method = "ML")

c_mod <- lme(below.C ~ observed_otus + status,
            random =~ 1|Genotype,
            data=bowen, method="ML")

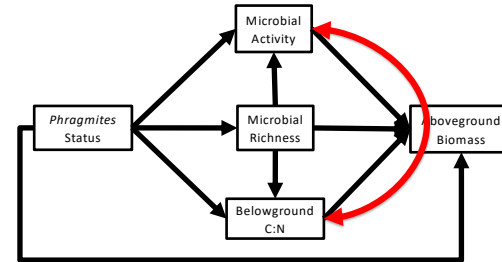
biomass_mod <- lme(abovebiomass_g ~ RNA.DNA + observed_otus + belowCN + status,
                 random =~ 1|Genotype,
                 data = bowen, method="ML")

method = "ML" for accurate estimates of fixed effects
    
```

### Build a pSEM

```
bowen_mod <- psem(
  div_mod,
  activity_mod,
  c_mod,
  biomass_mod,
  data = bowen
)
```

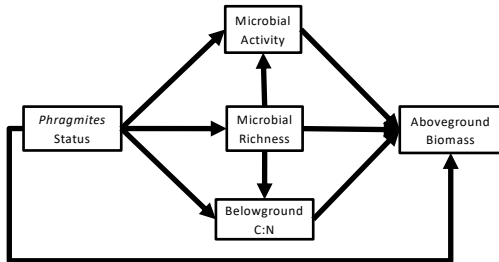
### Assessing D-Separation



```
> dSep(bowen_mod)

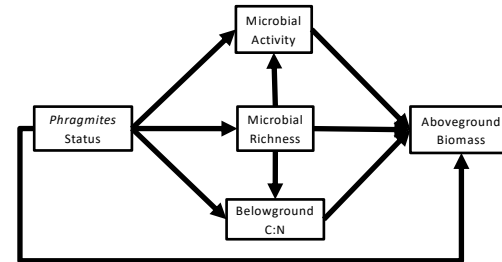
      Independ.Claim Estimate Std.Error DF Crit.Value P.Value
1 below.C ~ RNA.DNA + ... 1.867359 0.5073194 12 3.6808340.003144028
```

### Standardized Coefficients For Slopes, But...



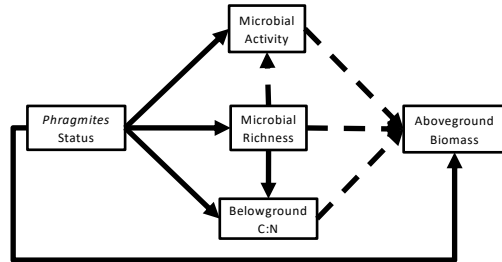
	Response	Predictor	Estimate	Std.Error	DF	Crit.Value	P.Value	Std.Estimate
1	RNA.DNA	observed_otus	0.0000	0.0000	58	1.8264	0.0729	0.1126
2	below.C	observed_otus	0.0008	0.0003	58	2.5503	0.0134	0.2765
3	abovebiomass_g	RNA.DNA	-1.7261	1.9200	56	-0.8990	0.3725	-0.1329
4	abovebiomass_g	observed_otus	-0.0001	0.0002	56	-0.6524	0.5168	-0.0801
5	abovebiomass_g	below.C	-0.0029	0.0747	56	-0.0392	0.9688	-0.0049

### Evaluating What Matters with LRTs



```
anova(bowen_mod)
```

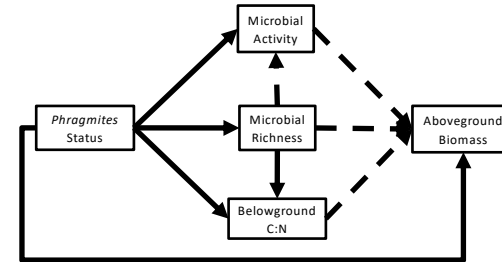
## OK, Some Death By LRT Tables



```

> anova(bowen_mod)
  Response      Predictor    Chisq Df Pr(>Chisq)
1 observed_otus      status  5.99687118  2 0.0498650168 *
2 RNA.DNA            status  9.00617069  2 0.0110747742 *
3 RNA.DNA observed_otus  3.52633809  1 0.0604010515
4 below.C observed_otus  6.87568700  1 0.0087376164 **
5 below.C            status 17.15849976  2 0.0001879659 ***
6 abovebiomass_g RNA.DNA  0.87954773  1 0.3483255823
7 abovebiomass_g observed_otus 0.46314090  1 0.4961597167
8 abovebiomass_g below.C  0.00167648  1 0.9673398711
9 abovebiomass_g status  9.80718892  2 0.0074198647 **
  
```

## emmeans Let's Us See Status Effects at Median of Covariates



```

library(emmeans)
lapply(bowen_mod[-length(bowen_mod)],
  emmeans, specs = ~status )
  
```

```

> lapply(bowen_mod[-length(bowen_mod)],
  emmeans, specs = ~status )
  
```

```

[[1]]
  status      emmean      SE df lower.CL upper.CL
introduced  2534 126.0 14      2264      2804
invasive    2530  53.7 12      2413      2647
native      2260 101.5 12      2039      2481
  
```

Degrees-of-freedom method: containment  
Confidence level used: 0.95

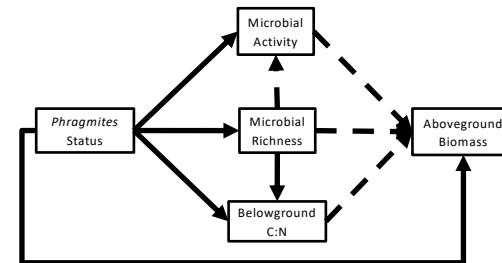
```

[[2]]
  status      emmean      SE df lower.CL upper.CL
introduced  0.731 0.0263 14      0.674      0.787
invasive    0.712 0.0117 12      0.686      0.737
native      0.784 0.0215 12      0.737      0.830
  
```

Degrees-of-freedom method: containment  
Confidence level used: 0.95

...

## Least Square Means for Posthoc Tests



```

generic_tukey <- function(x)
  emmeans(x, list(pairwise ~ status))
  
```

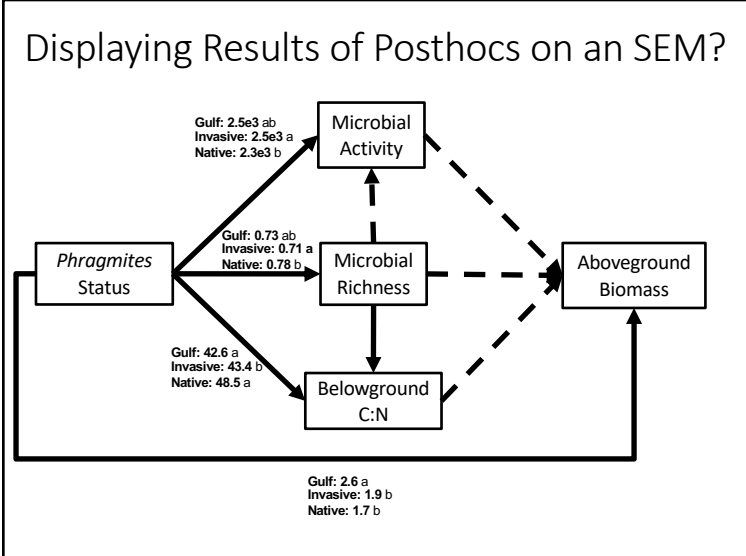
```
> lapply(bowen_mod[-length(bowen_mod)], generic_tukey)

[[1]]
$`emmeans of status`
  status      emmean      SE df lower.CL upper.CL
introduced 2533.885 126.03953 14 2263.557 2804.213
invasive   2530.027  53.73269 12 2412.953 2647.100
native     2259.921 101.46059 12 2038.857 2480.985

Confidence level used: 0.95

$`pairwise differences of status`
  contrast      estimate      SE df t.ratio p.value
introduced - invasive  3.858772 137.0152 12  0.028  0.9996
introduced - native  273.964182 161.8030 12  1.693  0.2475
invasive - native    270.105410 114.8105 12  2.353  0.0863

P value adjustment: tukey method for comparing a family of 3
estimates
...
```



### Conceptual Answers!

- It's all about that genotype!
- With the exception of biomass, invasive are different from Gulf and East Coast genotypes

- ### Future Directions
- Categorical endogenous variables will require implementation of multinomial logistic regression within piecewiseSEM (but you can implement by hand?)
  - Interaction Effects with categorical variables imply multigroup analysis

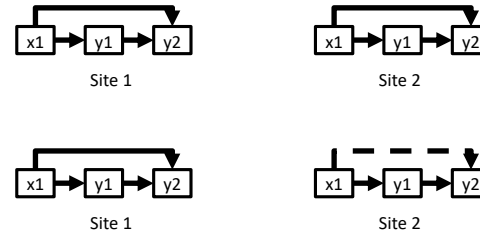
### Categorizing our SEMs

1. Categorical Predictors
2. Multigroup Analysis and pooling data from multiple sources
3. Multigroup analysis as model-wide interaction effect
4. General technique of imposing & releasing constraints
5. Multigroup Analysis in R

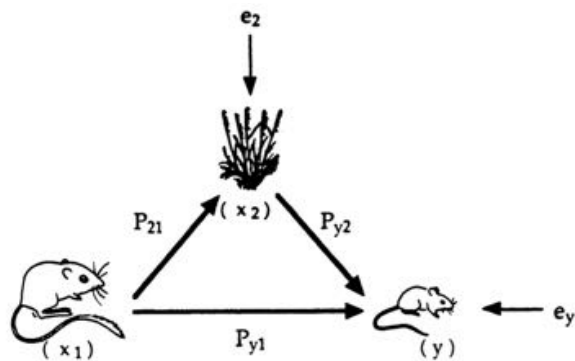
### Multigroup Analysis: Are Parameters The Same Across Groups?

Let's say you have data from two sites.

2 possibilities:

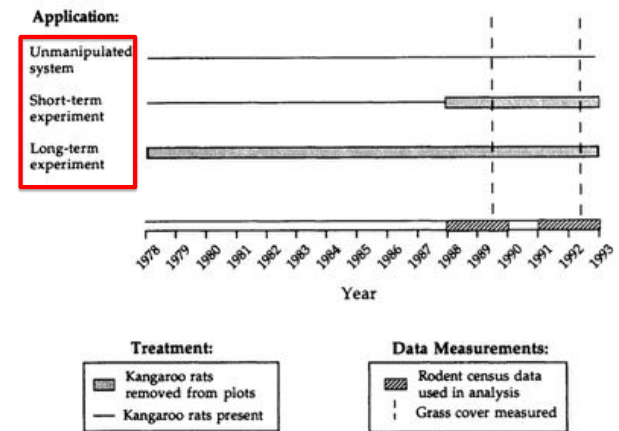


### Example: The Effects of Kangaroo Rats on Other Rodents



Smith et al. 1997 Am. Nat.

### Smith et al. – Results SHOULD Be the Same Regardless of Data Source



### Smith et al. – Path Analytic Results Wildly Vary due to Technique

**TABLE 3**  
CALCULATED PATH EQUATIONS OF THE EFFECT OF KANGAROO RATS ON OTHER RODENTS

Treatment	Path Equation	$e_y$
Harvest mice:		
A	-.303 = -.230 - .073	.85
B	-.710 = -.723 + .014	.66
C	-.760 = -.450 - .310	.57
Pocket mice:		
A	-.212 = -.150 - .062	.91
B	-.554 = -.561 + .006	.82
C	-.362 = .045 - .407	.67

### Grace and Pugesek: Multigroup Analysis Disagrees

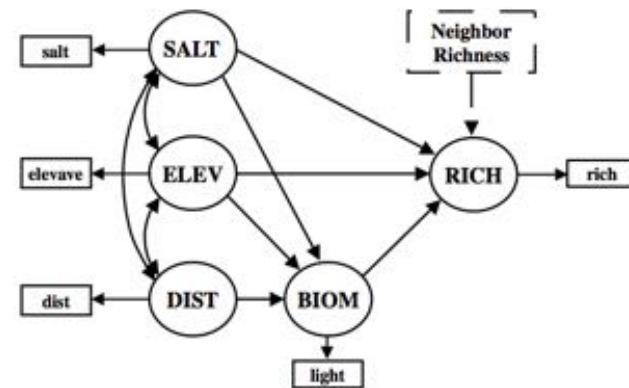
Harvest mice	Pocket mice
A. Multigroup goodness-of-fit statistic	
N for multigroup analysis = 38 $\chi^2$ with 9 df = 7.8400 (P = .5503)	N for multigroup analysis = 38 $\chi^2$ with 8 df = 9.6308 (P = .2919)
Group A $\chi^2$ = 3.2771	Group A $\chi^2$ = 4.6998
Group B $\chi^2$ = 1.7419	Group B $\chi^2$ = 2.2934
Group C $\chi^2$ = 2.8210	Group C $\chi^2$ = 2.6376

- Differences in std. coefs were due to differences in range of data.
- Indirect effects detected, largely due to data from long-term observations.
- Smith et al. replied that this *still* meant that Path Analysis could provide bogus results without the proper data. Oy.

### Categorizing our SEMs

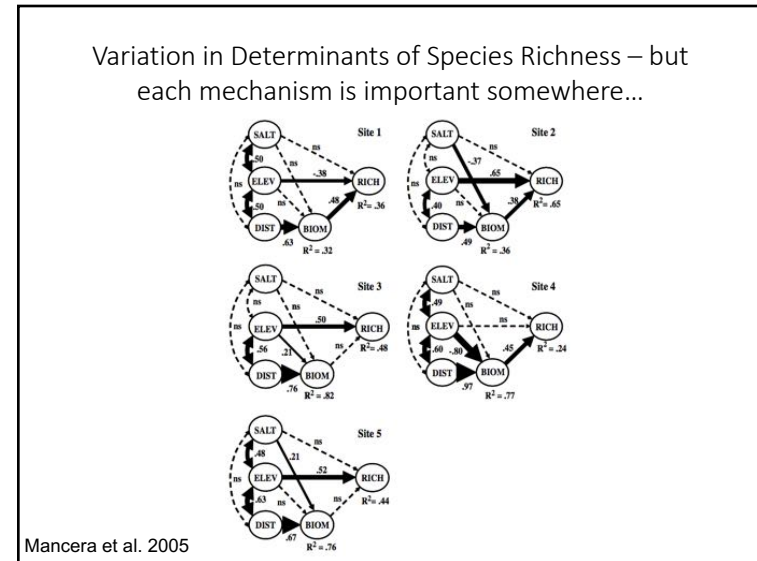
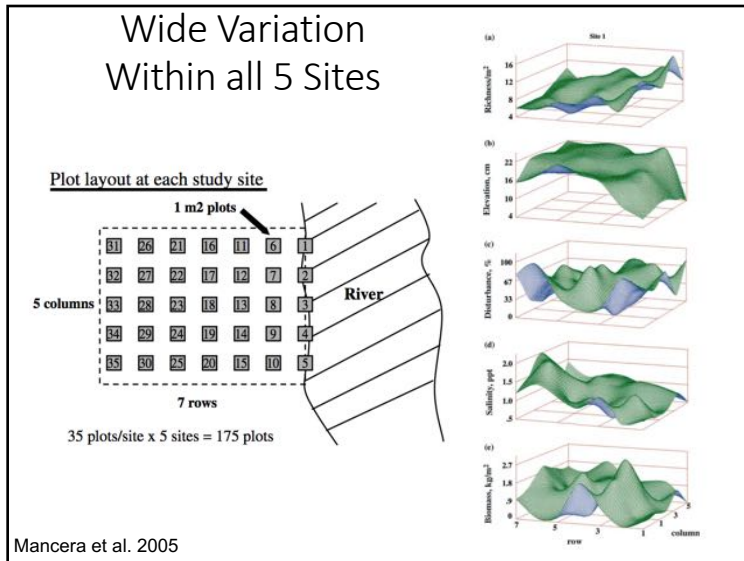
1. Categorical Predictors
2. Multigroup Analysis and pooling data from multiple sources
3. Multigroup analysis as model-wide interaction effect
4. General technique of imposing & releasing constraints
5. Multigroup Analysis in R

### Example: Is Richness Determined by the Same Factors at Different Sites?



Mancera et al. 2005





- ### Categorizing our SEMs
1. Categorical Predictors
  2. Multigroup Analysis and pooling data from multiple sources
  3. Multigroup analysis as model-wide interaction effect
  4. General technique of imposing & releasing constraints
  5. Multigroup Analysis in R





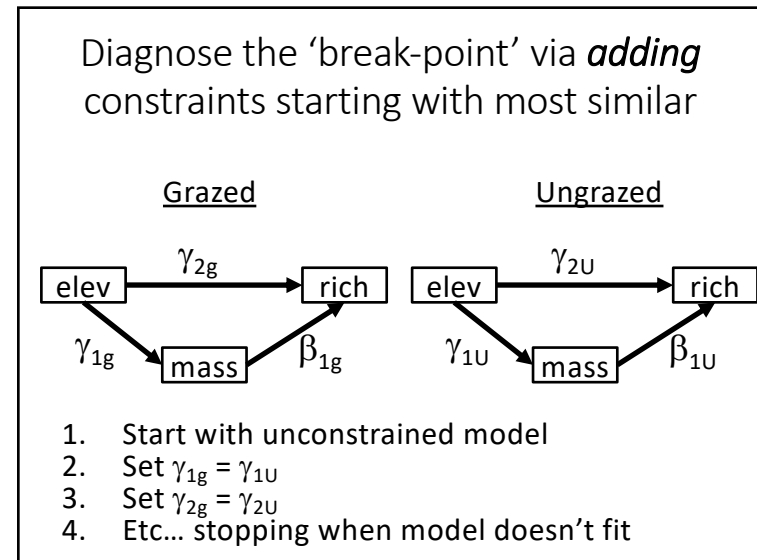
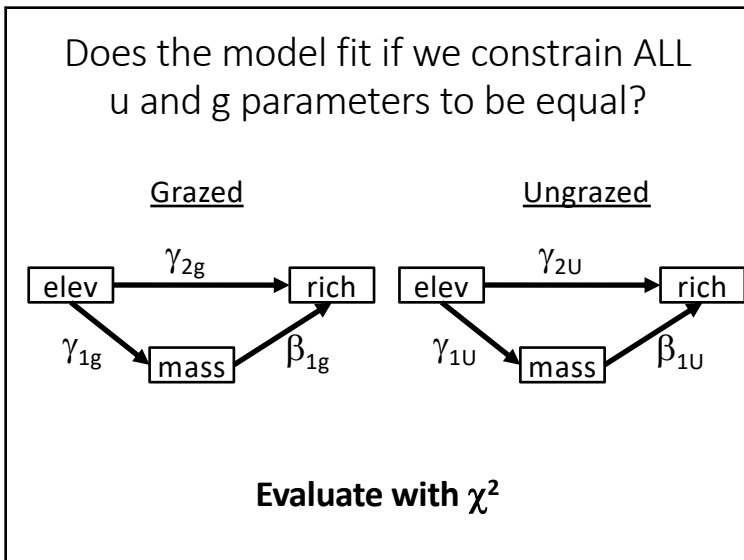
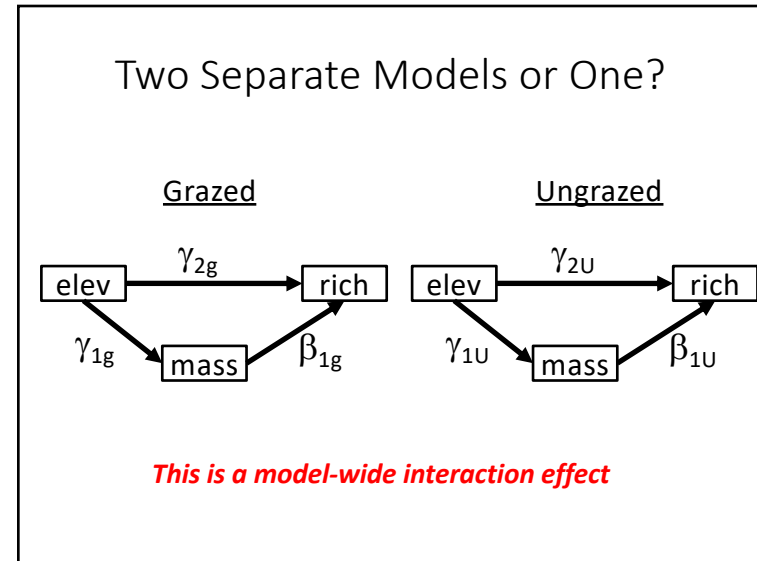
View of Data in FinnishMeadows.xls

```
> head(meadows)
  grazed elev stressmn dol par1 par2 par3 par4 par5 sol1 sol2
1      1  19.2   15.283 1.0  0    1    0    0    0    1    0
2      1  18.8   15.833 1.0  0    1    0    0    0    1    0
3      1  12.5   25.658 1.5  0    1    0    0    0    1    0
4      1  38.2   39.908 6.0  0    1    0    0    0    1    0
5      1  37.4   20.667 2.5  0    0    1    0    0    1    0
6      1  38.0   11.467 4.0  0    0    0    1    0    1    0
```

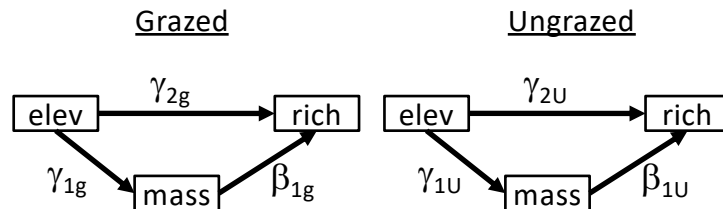
Data from 1-m<sup>2</sup> plots arrayed along an elevation gradient in each of several paired grazed and ungrazed meadows in SW Finland.

grazed = 0 is no, 1 is yes (this is our grouping variable)  
 elev = elevation of the plot above mean sea level  
 stressmn = flood stress index derived from long-term flooding records  
 dol = depth of litter layer in the plot  
 par1 - par5 = different parent materials  
 sol1 - sol5 = different soil types  
 mass, mass2, masslog = biomass in g/m<sup>2</sup>, square of biomass, and log biomass  
 rich, rich2, richlog = species richness per m<sup>2</sup>, square of richness and log richness

33



Diagnose the 'break-point' via *removing* constraints starting with most different

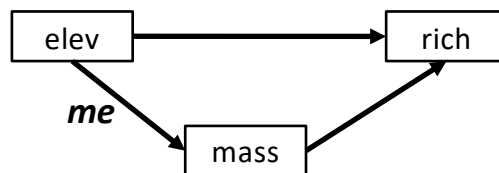


1. Start with constrained model
2. Set  $\gamma_{1g} \neq \gamma_{1U}$
3. Set  $\gamma_{2g} \neq \gamma_{2U}$
4. Etc... stopping when model fits

## Categorizing our SEMs

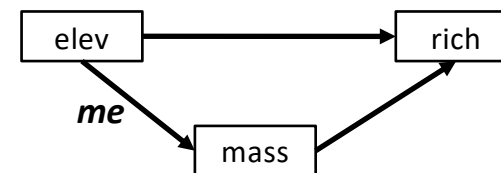
1. Categorical Predictors
2. Multigroup Analysis and pooling data from multiple sources
3. Multigroup analysis as model-wide interaction effect
4. General technique of imposing & releasing constraints
5. Multigroup Analysis in R
  - lavaan
  - piecewiseSEM

Start by Giving a Path a Name for Later Constraint Testing



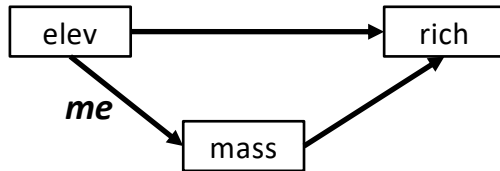
```
meadowModel<-'rich ~ elev + mass
mass ~ me*elev'
```

## Giving a Path a Name



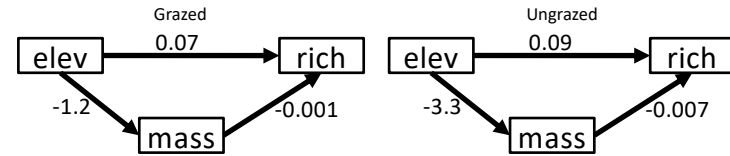
```
> meadowFit<-sem(meadowModel, data=meadows)
>
> coef(meadowFit)
rich-elev rich-mass      me rich~~rich mass~~mass
0.073    -0.003    -2.733    16.275  39563.789
```

### Giving a Path a Name



Regressions:		Estimate	Std.err	Z-value	P(> z )
rich ~					
elev		0.073	0.008	9.026	0.000
mass		-0.003	0.001	-2.988	0.003
mass ~					
elev	(me)	-2.733	0.372	-7.352	0.000

### Multiple Groups



```
> meadowFitFree<-sem(meadowModel, data=meadows,
group="grazed")
```

```
> coef(meadowFitFree)
rich-elev    rich-mass    me    rich--rich    mass--mass    rich-1
0.073        -0.001        -1.203    12.459        28057.591    7.169
mass-1    rich-elev.g2    rich-mass.g2    mass-elev.g2    rich--rich.g2    mass--mass.g2
260.855    0.088        -0.007        -3.274        14.384        43567.993
rich-1.g2    mass-1.g2
11.349        451.732
```

### meadowFitFree

```
lavaan (0.5-23.1097) converged normally after 53
iterations

Number of observations per group
1 165
0 189

Estimator ML
Minimum Function Test Statistic 0.000
Degrees of freedom 0

Chi-square for each group:
1 0.000
0 0.000
```

### summary(meadowFitFree)

```
Group 1 [1]:

Regressions:
rich ~
elev 0.073 0.010 7.232 0.000
mass -0.001 0.002 -0.424 0.672
mass ~
elev (me) -1.203 0.470 -2.559 0.010

Intercepts:
rich 7.169 0.708 10.126 0.000
mass 260.855 26.764 9.746 0.000

Variances:
rich 12.459 1.372
mass 28057.591 3089.039
```

### Testing Complete Constraints

```

> meadowFitEqual<-sem(meadowModel, data=meadows, group="grazed",
  group.equal=c("intercepts", "regressions"))

> anova(meadowFitFree, meadowFitEqual)
Chi Square Difference Test

          Df  AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
meadowFitFree  0 10019 10073  0.000
meadowFitEqual  4 10103 10142 92.116      92.116      4 < 2.2e-16
  
```

**Model Differs Between Treatments**

### Testing One Constraint

```

#constrain just the elev-mass relationship
meadowModel2<-'rich ~ elev + mass
  mass ~ c("me", "me")*elev'

meadowFitFree2<-sem(meadowModel2, data=meadows,
  group="grazed")
  
```

### Testing One Constraint

```

> coef(meadowFitFree2)
rich-elev  rich-mass      me  rich--rich  mass--mass  rich-l
  0.073    -0.001    -2.067    12.459    28632.459    7.169
mass-l    rich-elev.g2  rich-mass.g2  rich--rich.g2  mass--mass.g2  rich-l.g2
303.829    0.088      -0.007    14.384    44661.148    11.349
mass-l.g2
408.622
  
```

### Testing One Constraint

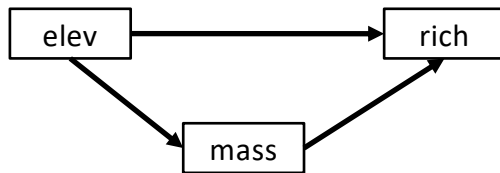
```

> anova(meadowFitFree, meadowFitFree2)
Chi Square Difference Test

          Df  AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
meadowFitFree  0 10019 10073  0.0000
meadowFitFree2  1 10025 10075  8.0301      8.0301      1 0.004601
  
```

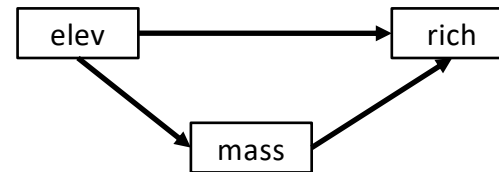
**elev -> mass relationship differs between treatments**

## Multigroup Exercise



1. Fit a free and constrained version of the meadow model with par2 as your group.
2. Evaluate a single constraint of your choice.

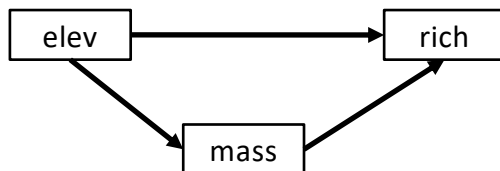
## Solution: The Fits



```
meadowFitFreePar2<-sem(meadowModel,
  data=meadows, group="par2")
```

```
meadowFitEqualPar2<-sem(meadowModel,
  data=meadows, group="par2",
  group.equal=c("intercepts",
  "regressions"))
```

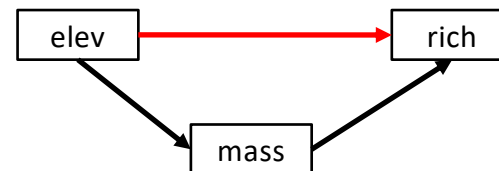
## Solution: Par2 Matters!



Chi Square Difference Test

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
meadowFitFreePar2	0	10072	10127	0.000			
meadowFitEqualPar2	4	10088	10127	23.511	23.511	4	0.0001001

## Solution: Testing elev->rich constraint



```
meadowModel3<-'rich ~ c("re", "re")*elev + mass
  mass ~ elev'
```

```
meadowFitFreePar2_3<-sem(meadowModel3, data=meadows,
  group="par2")
```

#does it matter?

```
anova(meadowFitFreePar2, meadowFitFreePar2_3)
```

### Solution: Elev-> Richness relationship varies by Par2

```

Chi Square Difference Test

      Df  AIC  BIC  Chisq  Chisq diff  Df  diff  Pr(>Chisq)
meadowFitFreePar2      0 10072 10127   0.000
meadowFitFreePar2_3    1 10080 10131 10.137    10.137      1 0.001453 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

### Next Steps?

- Maybe it's not just elev -> rich that differs!
- Need to test more constraints
- But...what if we want to test models with more constraints than free parameters?

### Testing One Release

```

meadowModelNoLabel<- 'rich ~ elev + mass
                    mass ~ elev'

meadowFitOneFree<-sem(meadowModelNoLabel, data=meadows,
                    group="grazed",
                    group.equal = c("regressions",
                                    "intercepts"),
                    group.partial = c("mass ~ elev",
                                       "mass ~ 1"))
    
```

### Testing One Release

```

lavaan (0.5-23.1097) converged normally after 40 iterations

Number of observations per group
1 165
0 189

Estimator ML
Minimum Function Test Statistic 65.917
Degrees of freedom 3
P-value (Chi-square) 0.000
    
```

**Need to release more!**



### Testing Two Releases

```

meadowFitTwoFree<-sem(meadowModelNoLabel, data=meadows,
  group="grazed",
  group.equal = c("regressions",
    "intercepts"),
  group.partial=c("mass ~ elev", "mass ~ 1",
    "rich ~ mass", "rich ~ 1"))
  
```

### Testing Two Releases

lavaan (0.5-23.1097) converged normally after 53 iterations

Number of observations per group	
1	165
0	189
Estimator	ML
Minimum Function Test Statistic	0.948
Degrees of freedom	1
P-value (Chi-square)	0.330

**Now we're good!**

### Fit Not Different Than Completely Unconstrained Model

```

> anova(meadowFitTwoFree, meadowFitFree)
Chi Square Difference Test
  
```

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
meadowFitFree	0	10019	10073	0.0000			
meadowFitTwoFree	1	10018	10068	0.9477	0.94767	1	0.3303

**Parsimony suggests retaining the constraint on elev -> rich**

### Care and Feeding of Multigroup Analysis

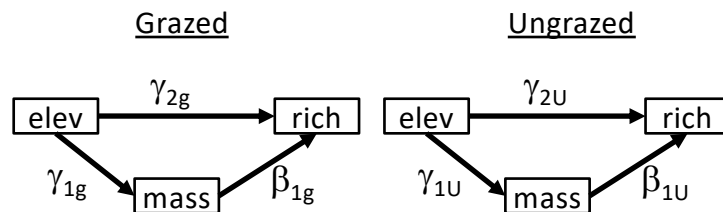
- Disparate results can be produced by different groups encompassing different ranges of variability.
- Variation in one group can reinforce weak patterns in another group.
- Need to have adequate sample size for each group!

What should I constrain? What should I test?

1. What are you interested in constraining?  
Typically just regression parameters.
2. Test a free v. constrained model.
3. Evaluate releasing key parameters or constraining key parameters, based on questions.

But what about piecewise...

A Simpler Paradigm?



*This is a model-wide interaction effect*

An Interactive Model  
(i.e., unconstrained)

```
rich_unconstrained <-
  lm(rich ~ elev*grazed +
      mass * grazed,
      data = meadows)

mass_unconstrained <-
  lm(mass ~ elev * grazed,
      data = meadows)
```

## Examining Constraints with LRTs

```
#Fit the SEM
unconstrained_int_mod <- psem(
  rich_unconstrained,
  mass_unconstrained,
  data=meadows
)

#Examine importance of model components
anova(unconstrained_int_mod)
```

## Death by LRT Table?

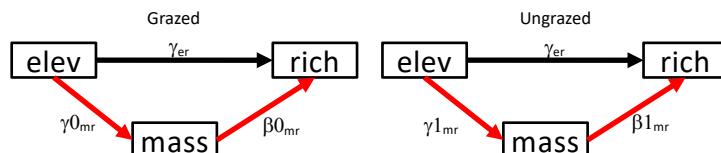
```
$rich
Anova Table (Type II tests)

Response: rich
      Sum Sq Df F value    Pr(>F)
elev    1556.7  1 113.4686 < 2.2e-16 ***
grazed   812.0  1  59.1898 1.488e-13 ***
mass     305.6  1  22.2723 3.433e-06 ***
elev:grazed 12.7  1  0.9290 0.335790
grazed:mass 126.3  1  9.2052 0.002595 **
Residuals 4774.3 348
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$mass
Anova Table (Type II tests)

Response: mass
      Sum Sq Df F value    Pr(>F)
elev   1416938  1 38.5521 1.510e-09 ***
grazed  854310  1 23.2441 2.131e-06 ***
elev:grazed 287419  1 7.8201 0.005452 **
Residuals 12863853 350
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

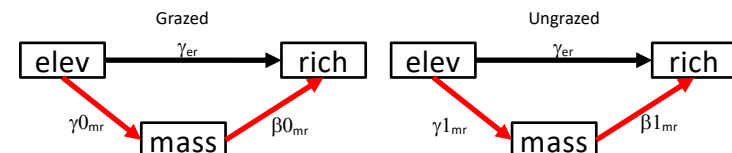
## Refit with Constraint



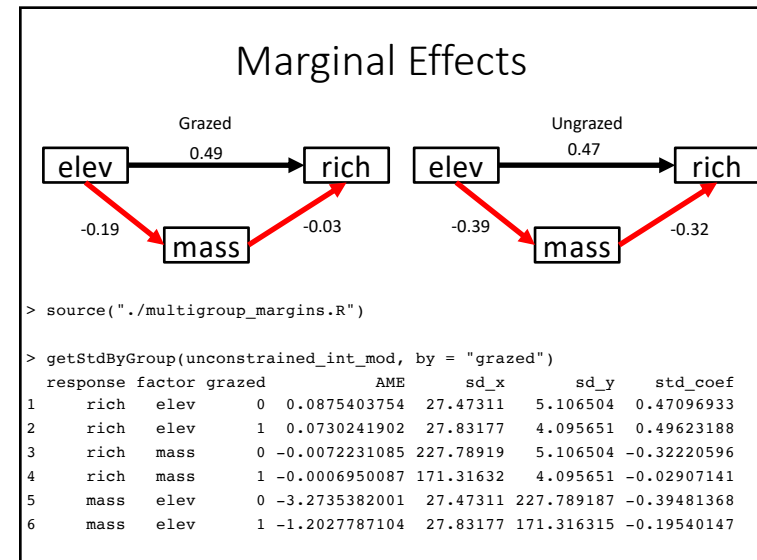
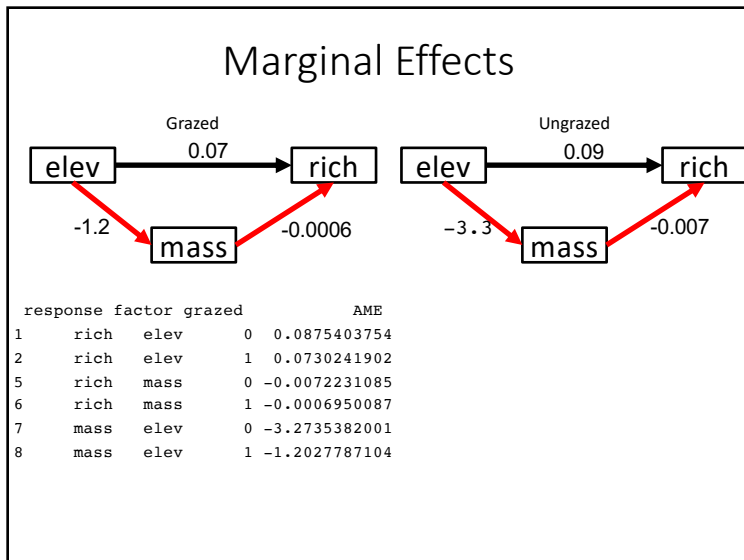
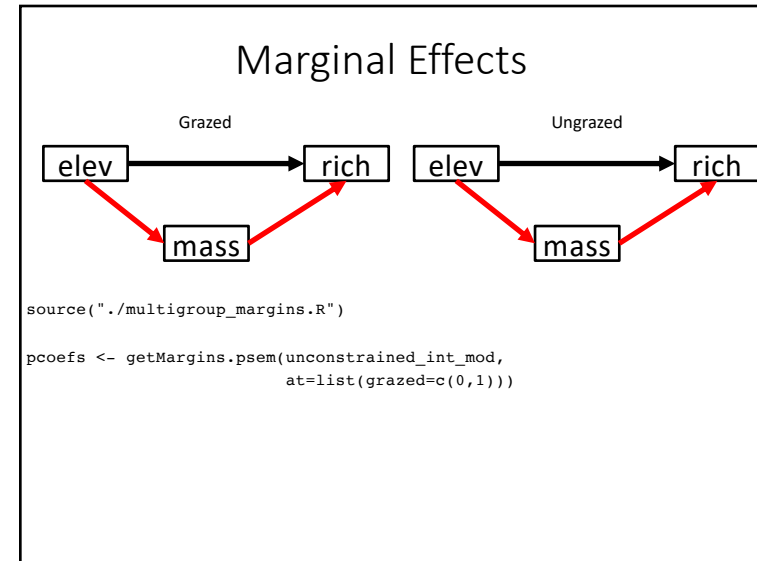
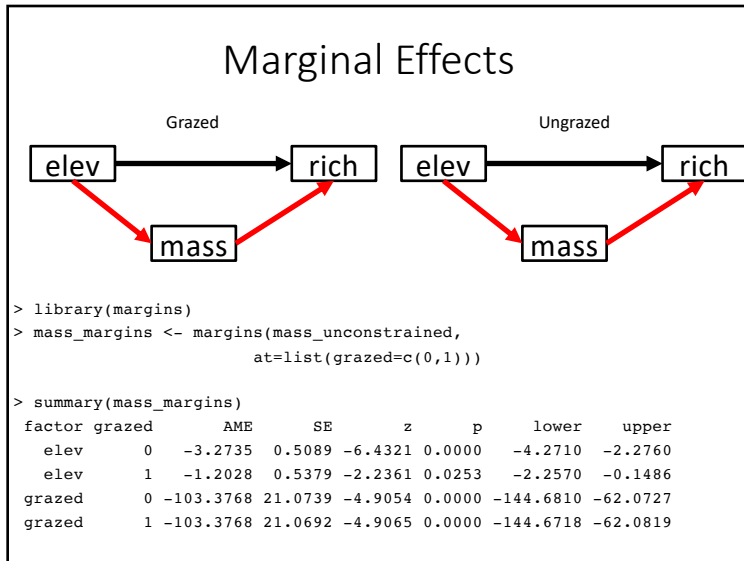
```
# One Constraint
rich_constrained <- lm(rich ~ elev +
  mass * grazed, data=meadows)

constrained_int_mod <- psem(
  rich_constrained,
  mass_unconstrained,
  data=meadows
)
```

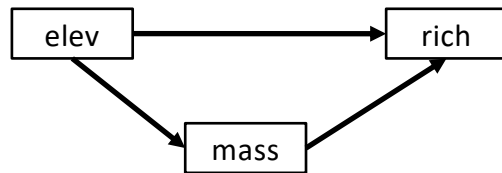
## How do we calculate $\beta$ ?



- $\beta_{xy} = b_{xy} \text{sd}(x)/\text{sd}(y)$
- BUT – now we have groups
- $\beta_{xy|g} = b_{xy|g} \text{sd}(x | g)/\text{sd}(y | g)$



### Multigroup Exercise



1. Fit a free and constrained version of the meadow model with par2 as your group.
2. Evaluate a single constraint of your choice.