Introduction to structural equation modeling and mixed models in R

Day 3 – Part 1: SEM

Oksana Buzhdygan

oksana.buzh@fu-berlin.de

About me

Dr. Oksana Buzhdygan

Contact | Publications

AG Tietjen - Theoretische Ökologie

Institut für Biologie
Freie Universität Berlin
Researcher

Address	Freie Universität Berlin Institut für Biologie Königin-Luise-Straße 2/4, Gartenhaus 14195 Berlin
Email	oksana.buzh@fu-berlin.de

Research Interests

- Impacts of environmental changes on biodiversity and ecosystem functioning
- Relationships between biodiversity, community structure, and ecosystem functioning in multitrophic systems
- Multitrophic interactions, energy flow, and nutrient cycle in food webs
- Ecological Network Analysis

Curriculum Vitae

- since 2021: Senior scientist, Theoretical Ecology, Freie Universität Berlin

https://www.bcp.fu-berlin.de/en/biologie/arbeitsgruppen/botanik/ag_tietjen/People/wissenschaftliche_mitarbeiter/Buzhdygan

Day 3 – Part 1

Outline

- Basics of SEM

    ✓ From regression to SEM

    ✓ SEM history. SEM in natural sciences.

    ✓ SEM workflow process. Where do I start?

    ✓ First impression of 'lavaan'

Day 3 – Part 1

Outline

- Basics of SEM

  ✓ **From regression to SEM**

  ✓ SEM history. SEM in natural sciences.

  ✓ SEM workflow process. Where do I start?

  ✓ First impression of 'lavaan'

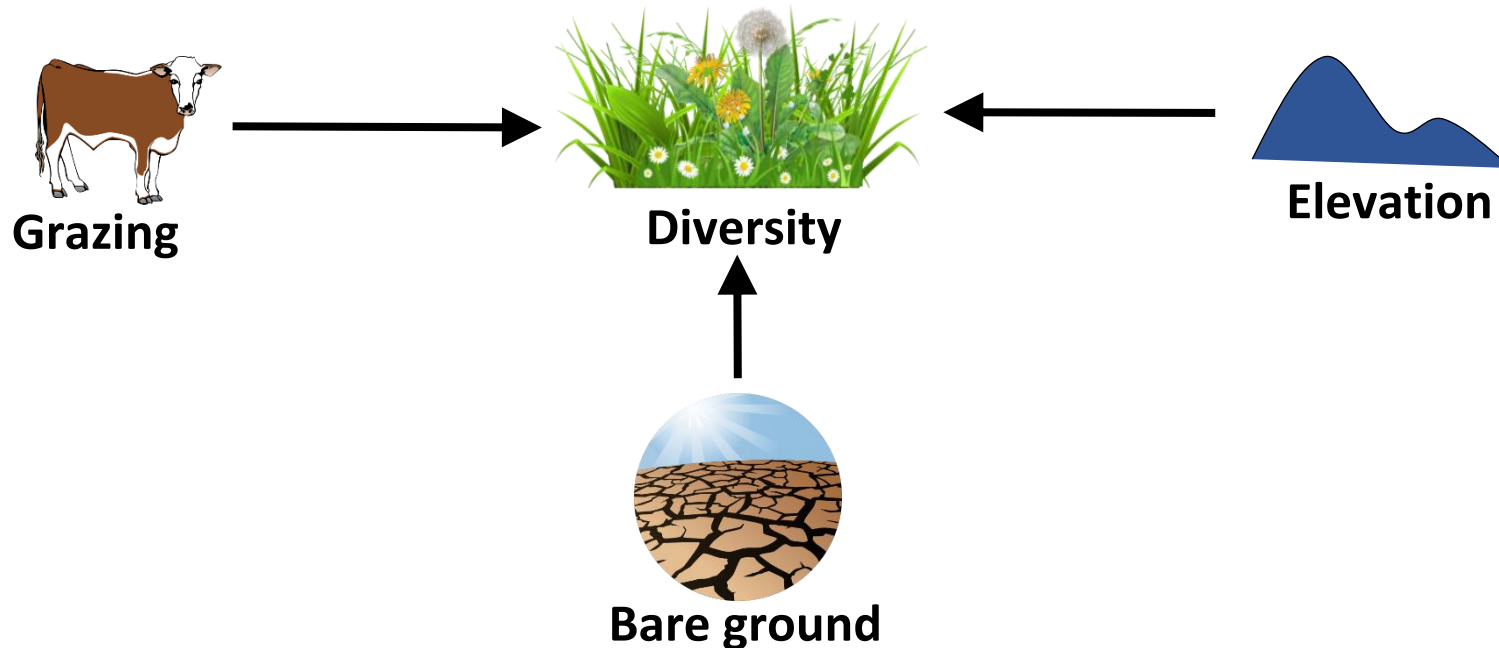# From regression to SEM

Aim of regression model:

- (How) does variable x impact variable y?

- Can we better predict values for variable y, if we account for variable x?

$$y = a + bx$$

From regression to SEM

Aim of regression model:

- (How) does variable x impact variable y?

- Can we better predict values for variable y, if we account for variable x?

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$$

Buzhdygan, *et al.* 2021 *PLosONE*

Grazing → **Diversity** ← **Elevation**
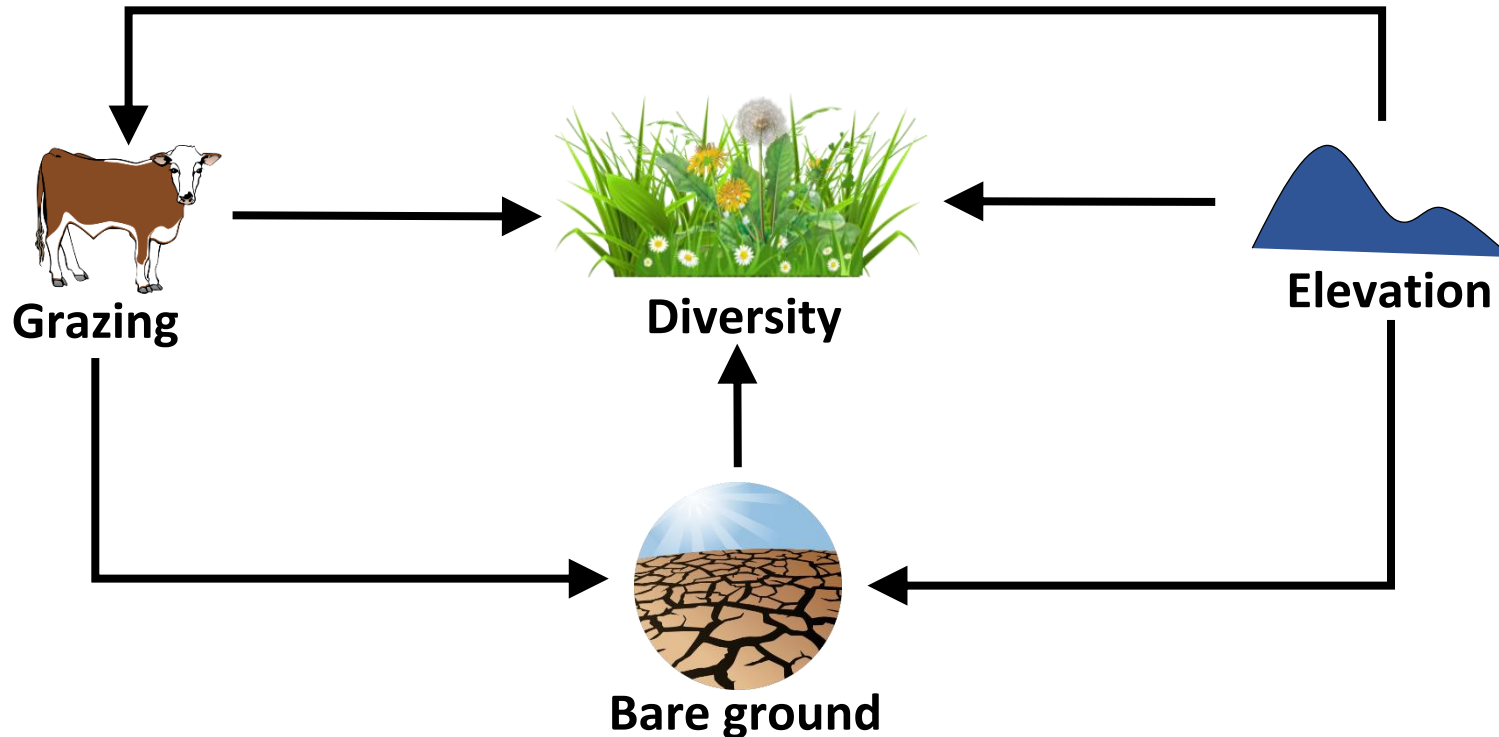
Bare ground → **Diversity**

Univariate relationships
- involve response variable explained by a set of predictors

From regression to SEM

SEM:

- Tests **systems of relationships** (multivariate) rather than a dependant variable and a set of predictors (univariate relationships)
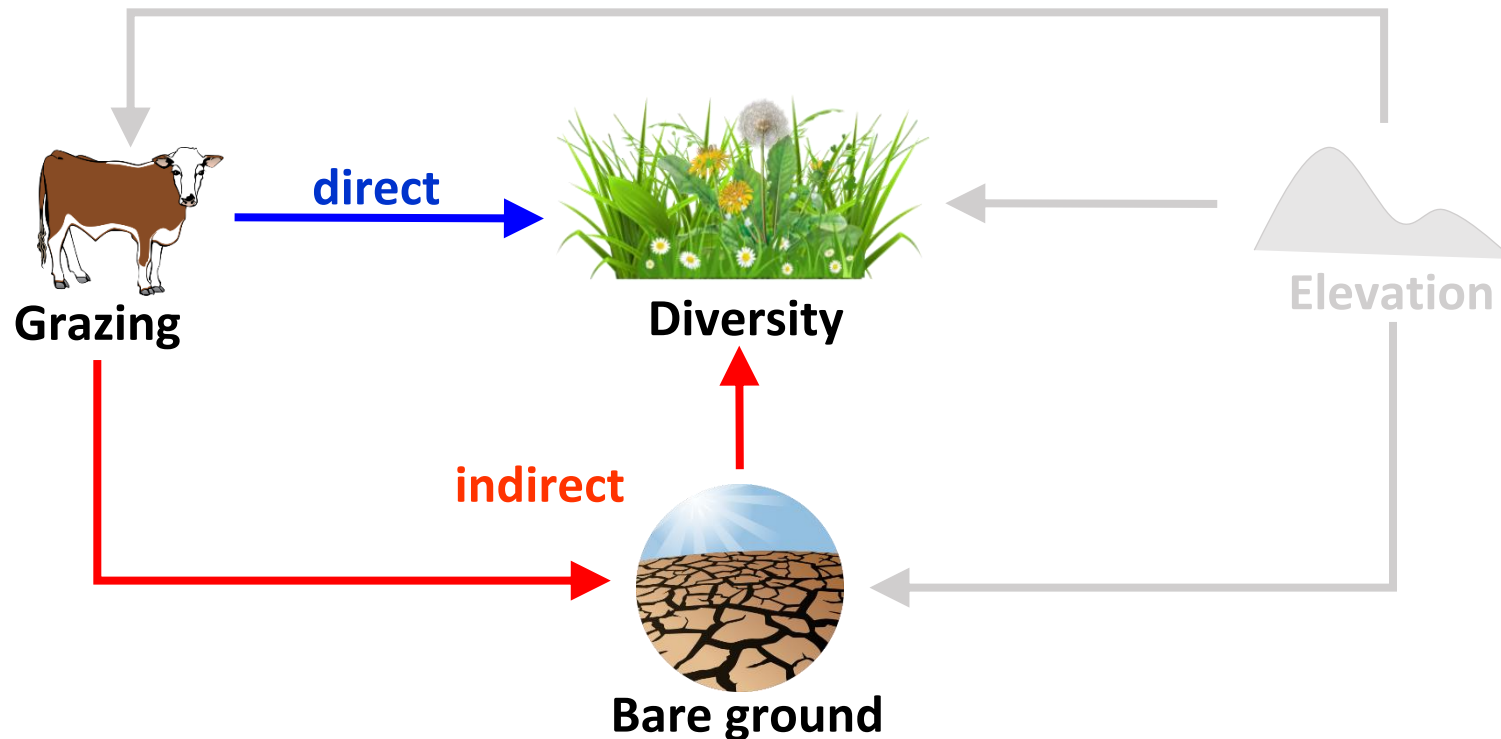

Buzhdygan, *et al.* 2021 *PLosONE*

Grazing

Diversity

Elevation

Bare ground

Multivariate relationships
- involve simultaneous influences and responses
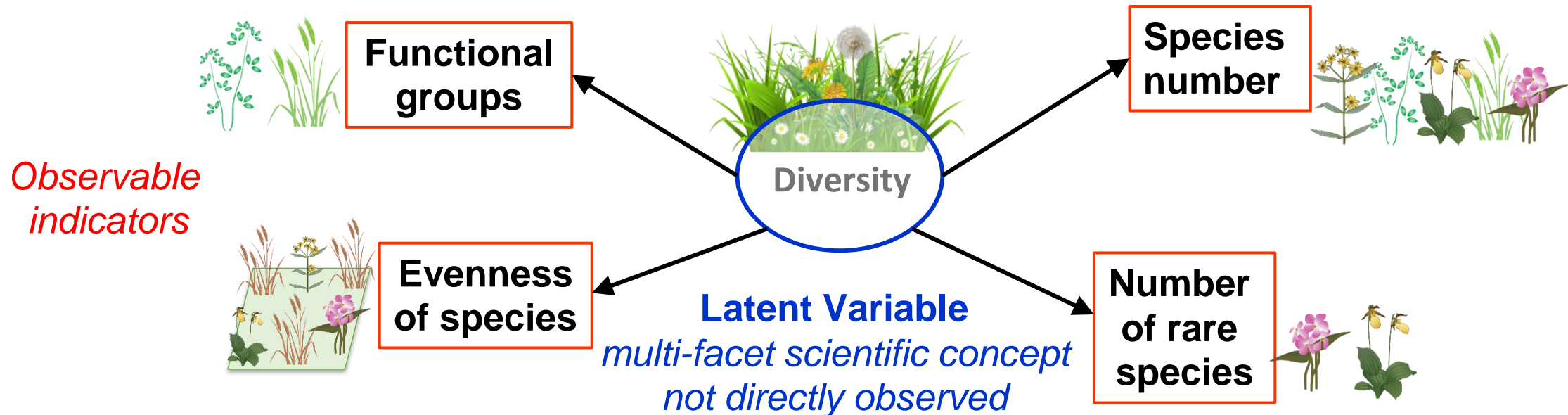
From regression to SEM

SEM:

- Tests **systems of relationships** (multivariate relationships)

- Allows testing **indirect and direct effects** of variables on other variables

From regression to SEM

SEM:

- Tests **systems of relationships** (multivariate relationships)

- Allows testing **indirect and direct effects** of variables on other variables

- Involves complex, multi-faceted **constructs,** approximated by observed indicators

Observable indicators

Functional groups

Species number

Diversity

Evenness of species

Number of rare species

Latent Variable
multi-facet scientific concept not directly observed

What is SEM?

Structural

There is hypothesized underlying **structure** to study system (a **cause and an effect**)…
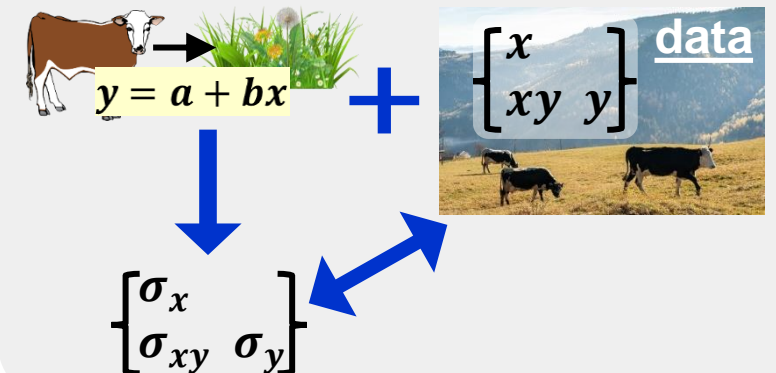
Grazing → Diversity

x → y

Equation

…that can be translated to a series of mathematical **equations**…

$$y = a + bx$$

Modelling

……which can be **modelled** **against data** to support or refuse the proposed **structure**

$y = a + bx$

$\begin{bmatrix} x \\ xy & y \end{bmatrix}$ data

$+$

$\begin{bmatrix} \sigma_x \\ \sigma_{xy} & \sigma_y \end{bmatrix}$

What is SEM?

- **not one** statistical method or technique

Statistical & mathematical tools

SEM framework

within the SEM framework

Barnes et al. (2018)
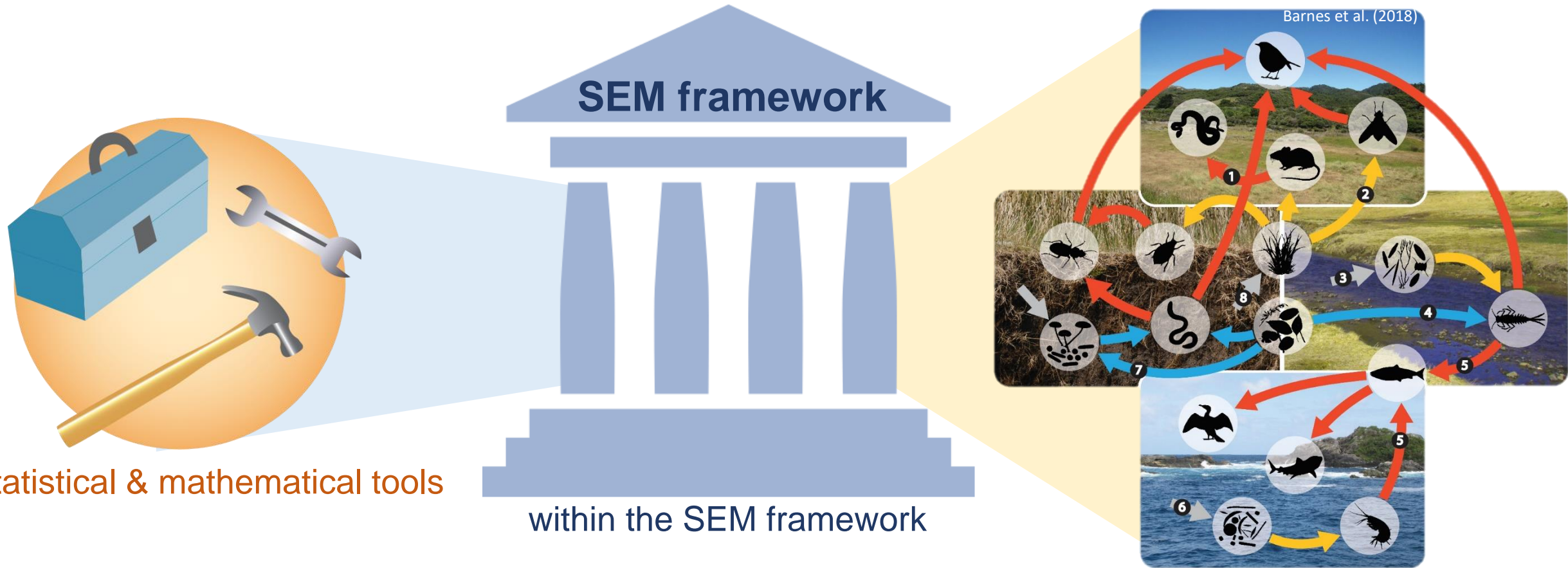
to understand multiple processes in complex systems

11

What is SEM?

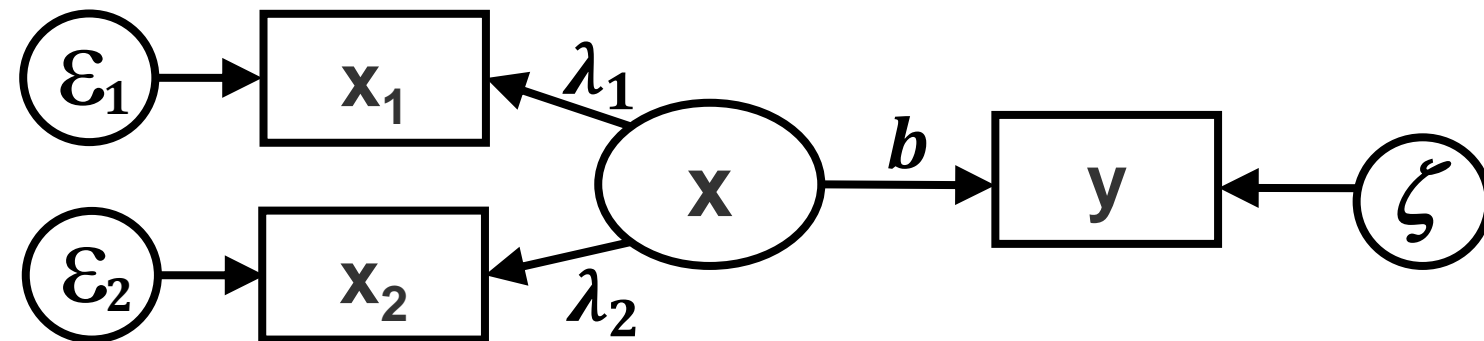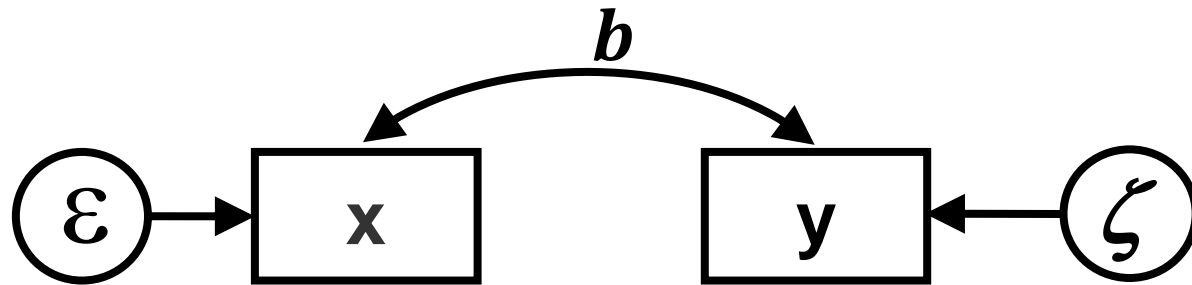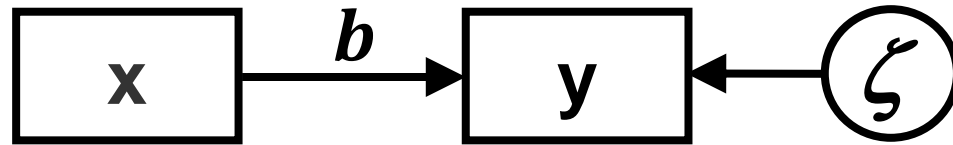- **not one** statistical method or technique

Integrates:
- Measurement theory
- Factor analysis
- Path analysis
- Correlation & Regression
- Simultaneous equations
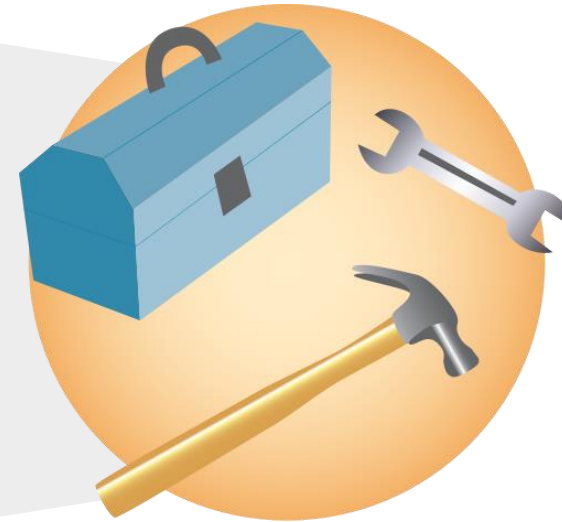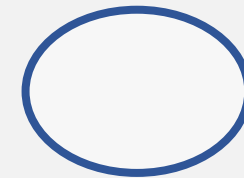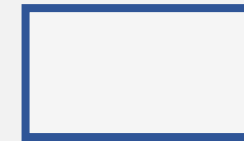
Statistical & mathematical tools

SEM is Graphical Modelling

Path Diagram Notations:

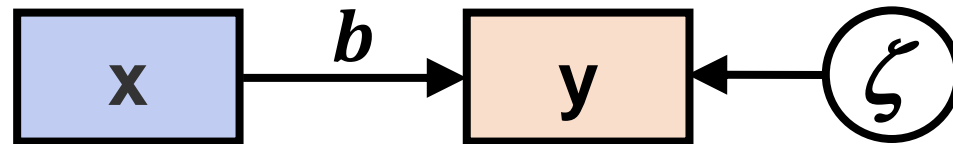Latent variable

Observed variable

Error variance

Regression

Covariance

SEM is Graphical Modelling

Equation form:

$$y = a + bx + \zeta$$

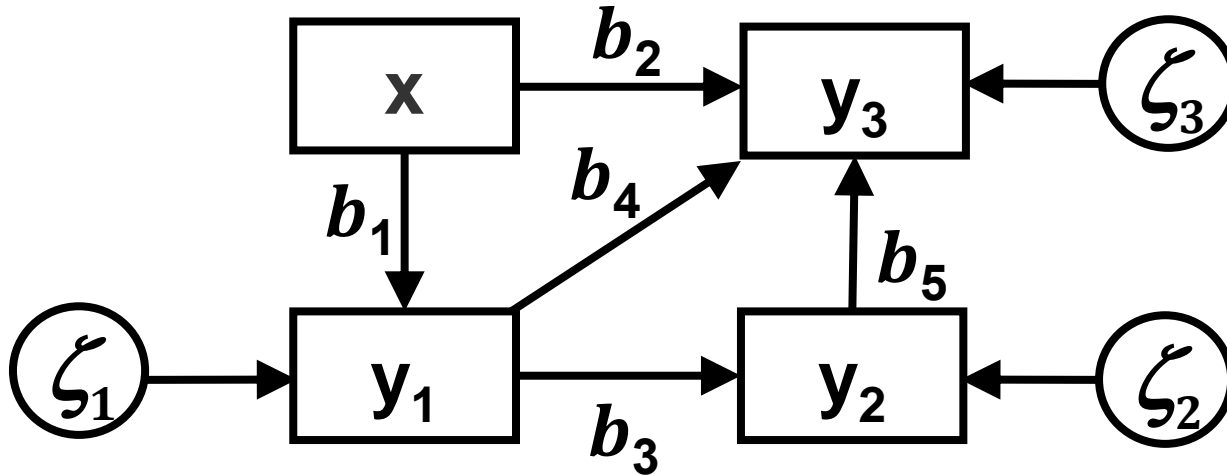**Graphical form (Path Diagram):**

Exogenous variable
*   have arrows directed only ***out of it***
    (i.e., no arrows going into it)

Endogenous variable
*   for which arrows are also directed ***into it***
*   can also have arrows directing out of it,
    but **must be predicted** at the same time

What is SEM?

Path Diagram:

Corresponding equations:

$$y_1 = b_1 x + \zeta_1$$
$$y_2 = b_3 y_1 + \zeta_2$$
$$y_3 = b_2 x + b_4 y_1 + b_5 y_2 + \zeta_3$$

SEM addresses
- multivariate relationships
  (simultaneous influences and responses)
- mechanical understanding
  (direct & indirect effects)

What is SEM?

Graphical form **with causality:**
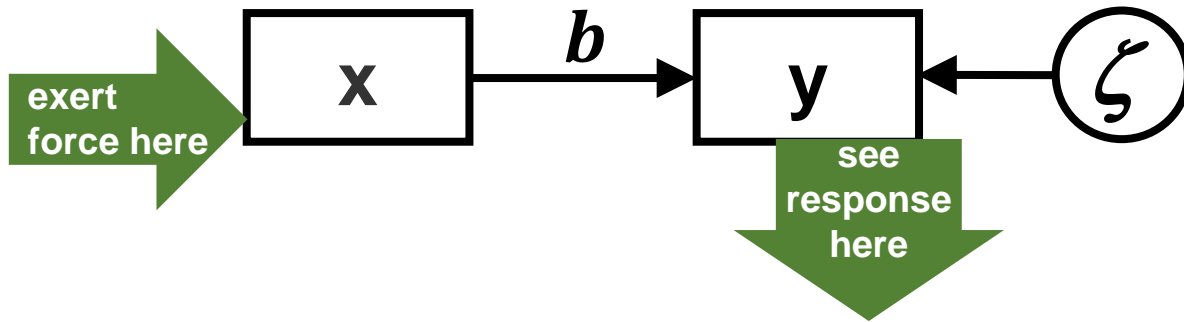
Cause-Effect Relations

x causes **y**
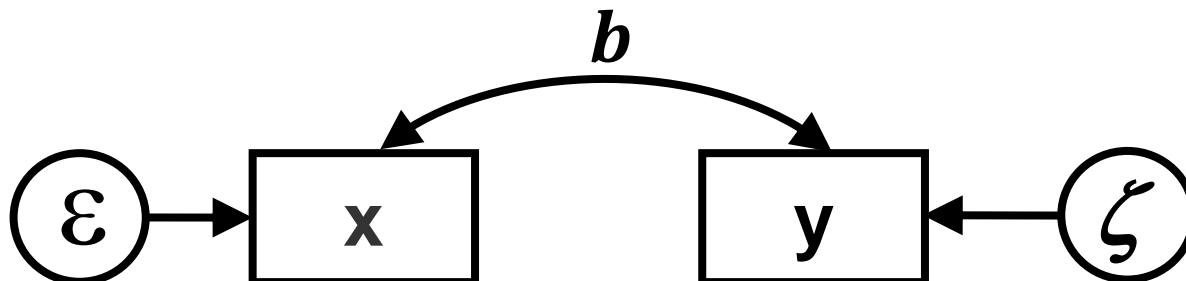
(if manipulation of **x** leads to a response in **y**)

"correlation does not imply causation"

R.A. Fisher

Graphical form **without causality:**

Correlation

y and **x** tend to be observed at the same time

Causality *vs* Correlation
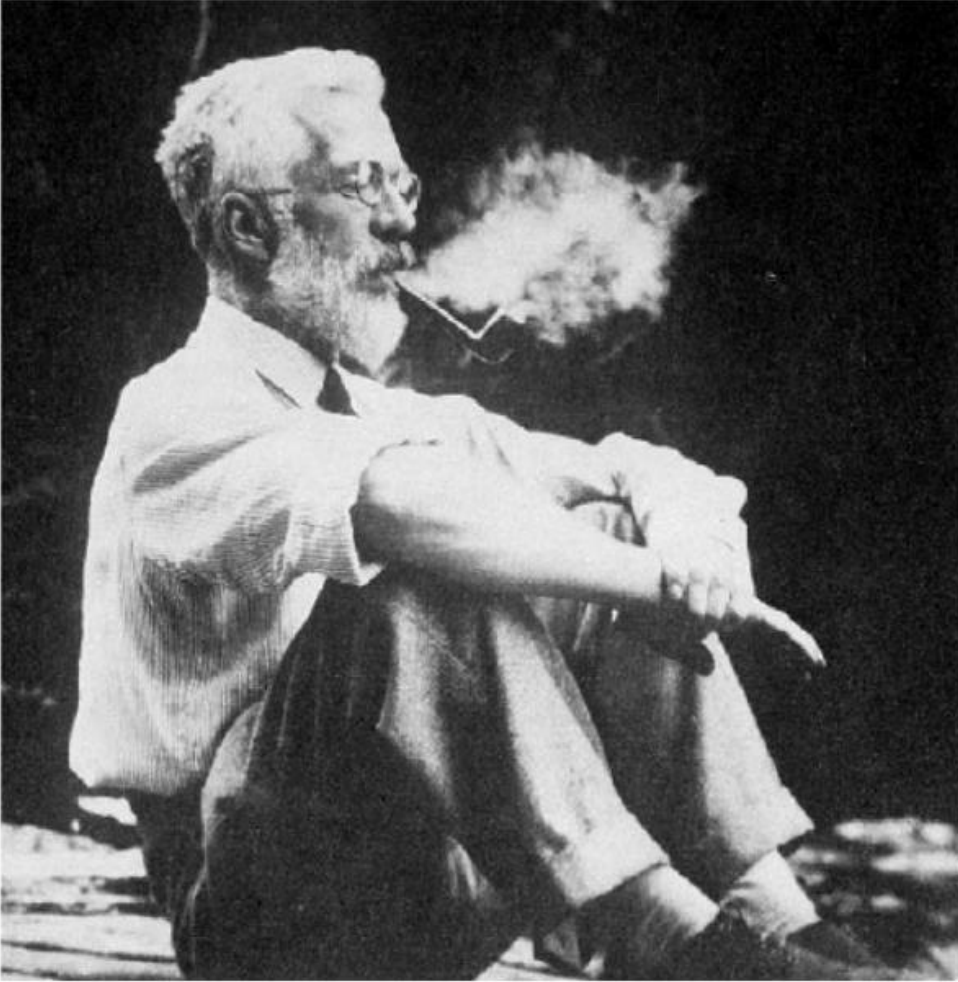
R.A. Fisher smoking a pipe, 1956.

Smoking ⟷ Lung Cancer

"correlation does not imply causation"
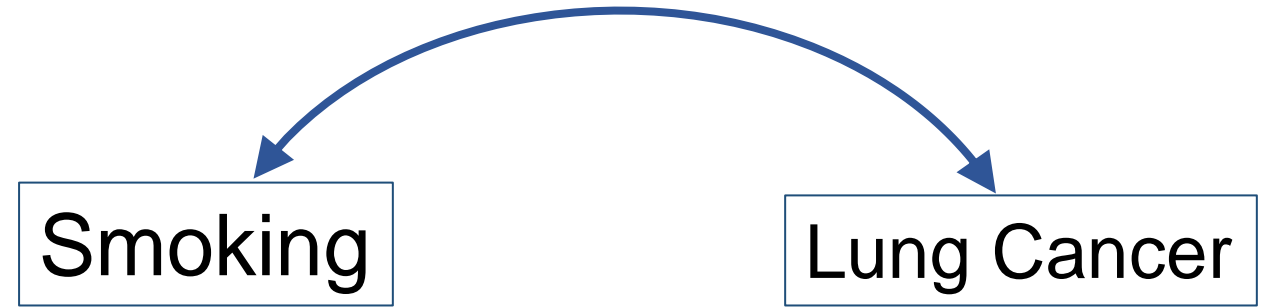
R.A. Fisher

Smoking → Lung Cancer

Causality *vs* Correlation

Ambulance cars tend to be observed in traffic jams

Ambulance ⟷ Car accidents in traffic jams

Ambulance ← Car accidents in traffic jams

Ambulance → Car accidents in traffic jams
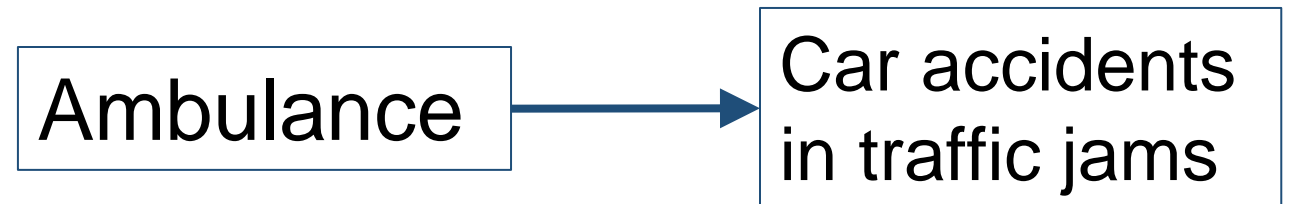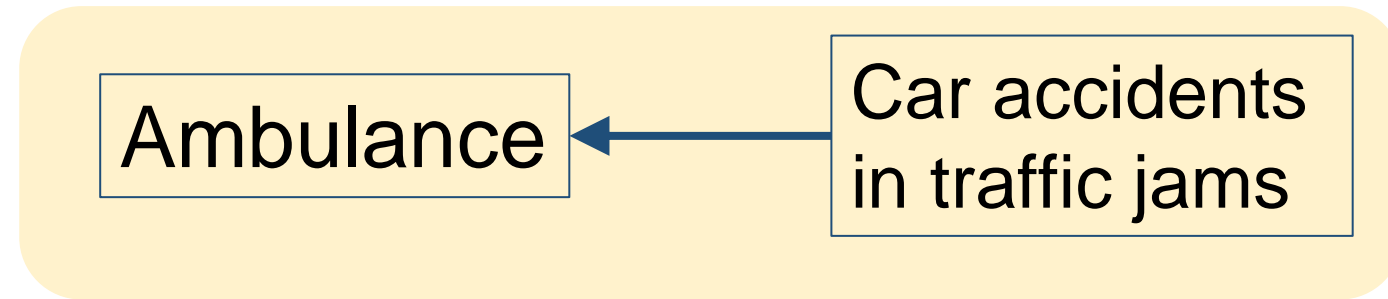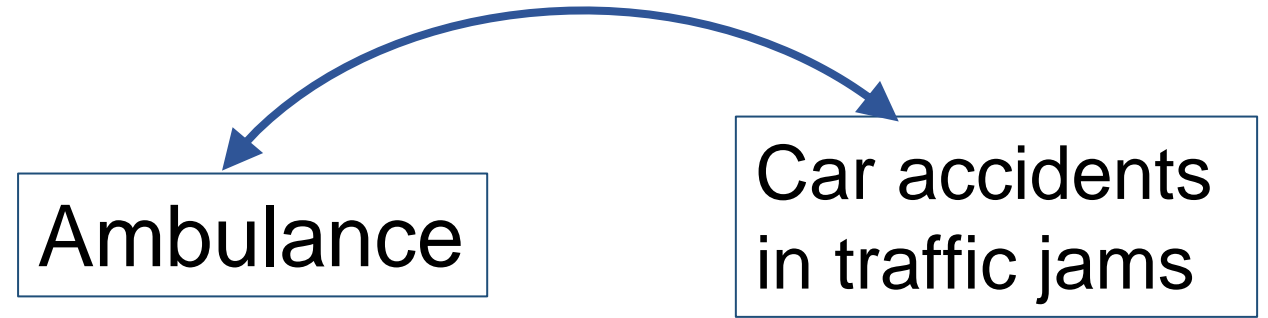
What is SEM?

SEM is not a method for discovering causes
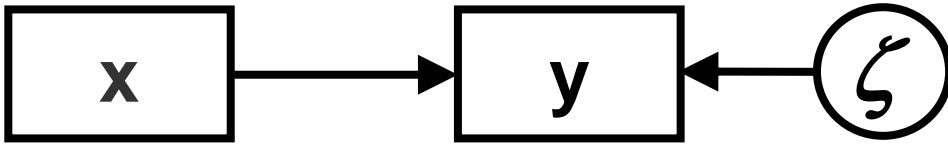
- SEM results is **not a proof of causal claims**.

- SEM relies upon the **causal assumptions made by us**, when building the model.

We assume that **x** causes **y** from:

- Research design
- Prior observation
- Prior statistical models
- Prior experimentation
- Logical arguments
- Some or all of the above

Not from SEM

!

The **credibility of the SEM** depends on the **credibility of the causal assumptions** made by the **researcher**

Day 3 – Part 1　　　Outline

- Basics of SEM

  - ✓ From regression to SEM

  - ✓ **SEM history. SEM in natural sciences.**

  - ✓ SEM workflow process. Where do I start?

  - ✓ First impression of 'lavaan'

SEM History

adapted from Grace, 2006

SEM in Ecology and Evolution

Jim B. Grace   www.structuralequations.org   **SEM adaptation to the needs of ecology and evolutionary biology**

- Grace (2010) Structural Equation Modeling for Observational Studies. *Journal of Wildlife Management*, 72:14-22
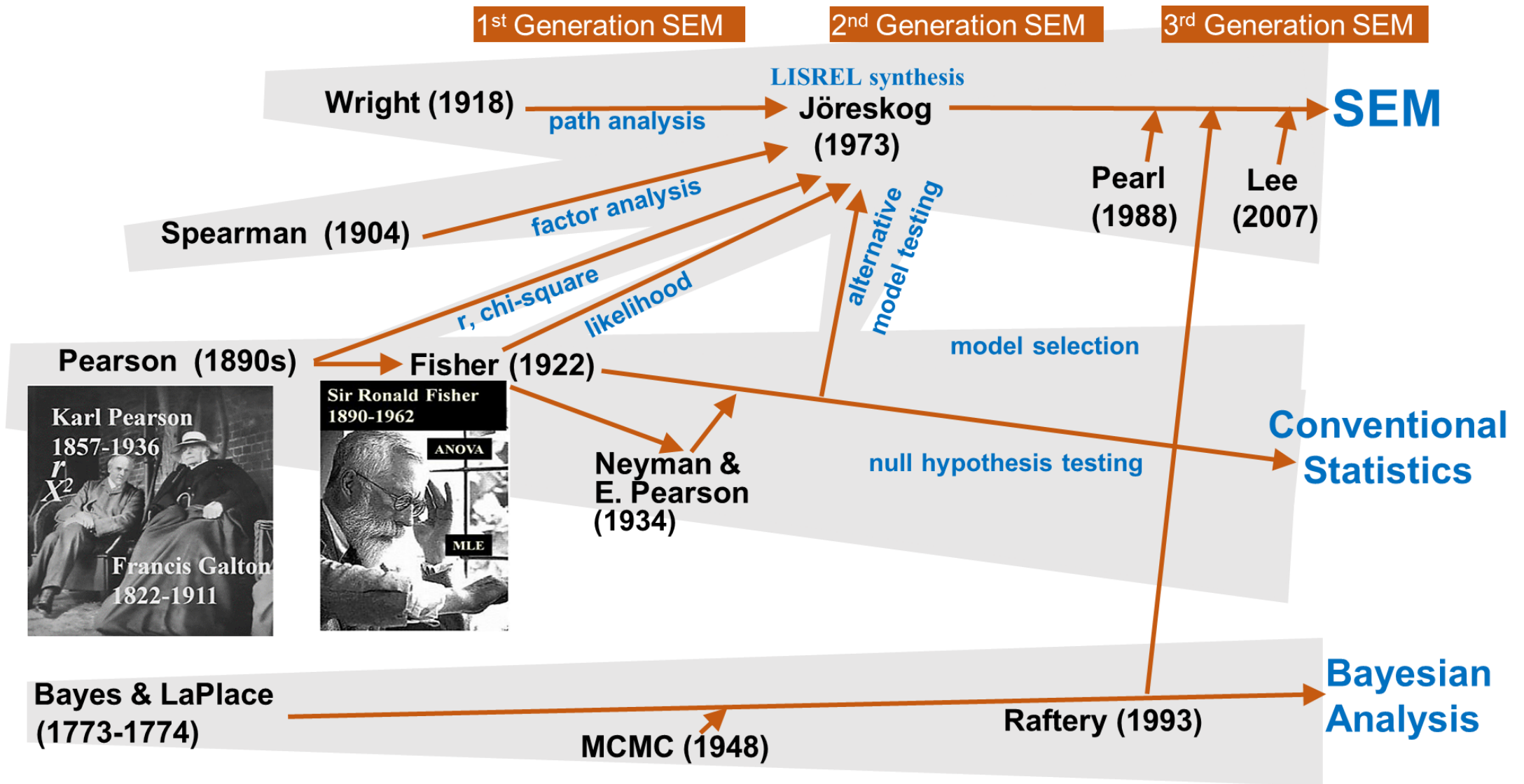- Grace et al. (2010) On the specification of structural equation models for ecological systems. *Ecological Monographs*, 80, 67-87.
- Grace, Bollen (2005) Interpreting the Results from Multiple Regression and Structural Equation Models. *Bulletin of the Ecological Society of America*, 86, 283-295.
- Grace (2015) Taking a systems approach to ecological systems. *Journal of Vegetation Science* 26, 1025-1027.

Jon Lefcheck   https://jslefche.github.io/sem_book   **Tools for SEM in R**: piecewiseSEM

Jarrett Byrnes   https://jebyrnes.github.io/semclass   **Tools for SEM in R**: *sem.additions* collaborate on *sem* and *lavaan*

Learn More about SEM

- Grace (2006) Structural Equation Modeling and Natural Systems. Cambridge Univ. Press.

- Shipley (2000) Cause and Correlation in Biology. Cambridge Univ. Press.
- Kline (2012) Principles and Practice of Structural Equation Modeling. (3rd Edition) Guilford Press.

- Bollen (1989) Structural Equations with Latent Variables. John Wiley and Sons.

- Hoyle (2012) Handbook of Structural Equation Modeling. Guilford Press.

Day 3 – Part 1   Outline

- Basics of SEM

  ✓ From regression to SEM

  ✓ SEM history. SEM in natural sciences.

  ✓ **SEM workflow process. Where do I start?**

  ✓ First impression of 'lavaan'

Where to start?

Theory

- The literature
- Natural history
- Exploratory analyses
- Logical arguments
- Available data

Build a Model

Collect Data

Revisit Theory

Confront Model with Data

Model & Data consistent

Estimate Parameters, Assess Model Fit

Model & Data inconsistent

fail = learning!

Interpret Model, Predict

Modify Model
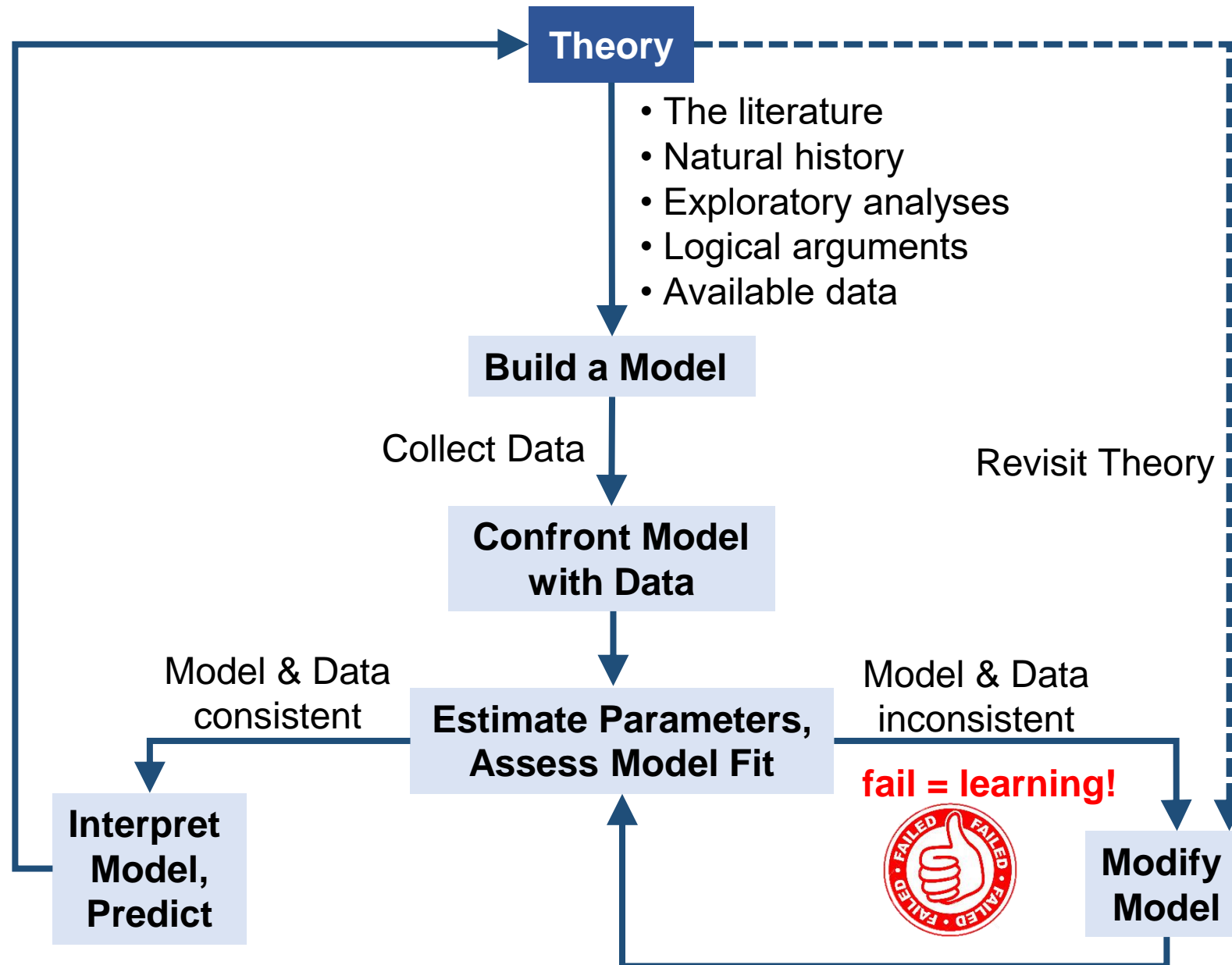
- multi-step process to build knowledge through sequential learning

- fail implies learning from your data and through revisiting theory

25

Where to start?

Theory

- The literature
- Natural history
- Exploratory analyses
- Logical arguments
- Available data

Build a Model

Collect Data

Confront Model with Data
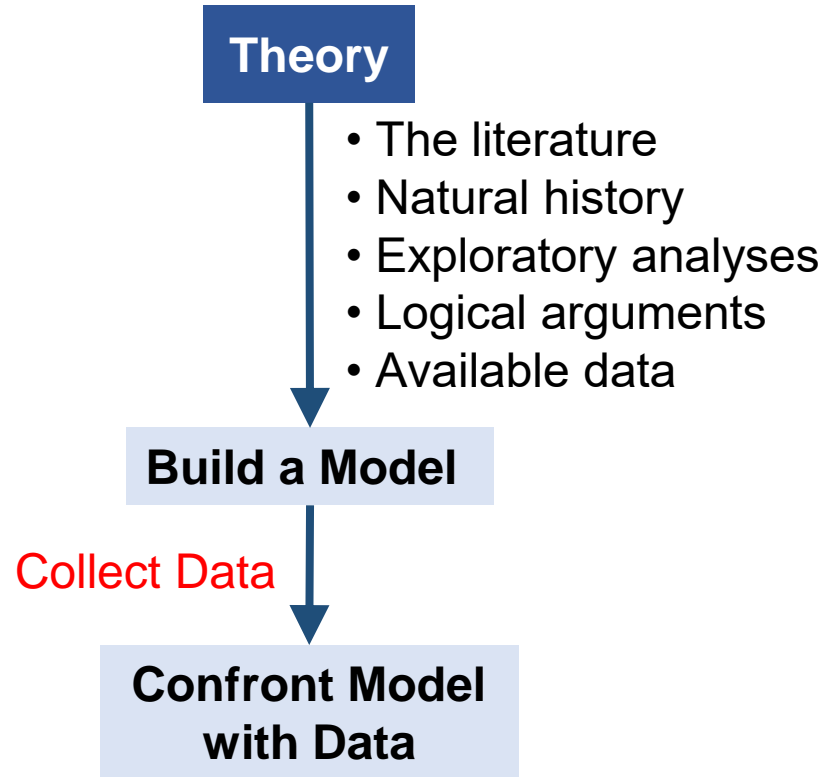
Buzhdygan, *et al.* 2020 *PLosONE*

Study area

Carpathians

Map: Google Earth

Collect Data and Parametrise Model

Elevation, m asl

Plant species number

Elevation → **Diversity**

Grazing → **Bare ground**
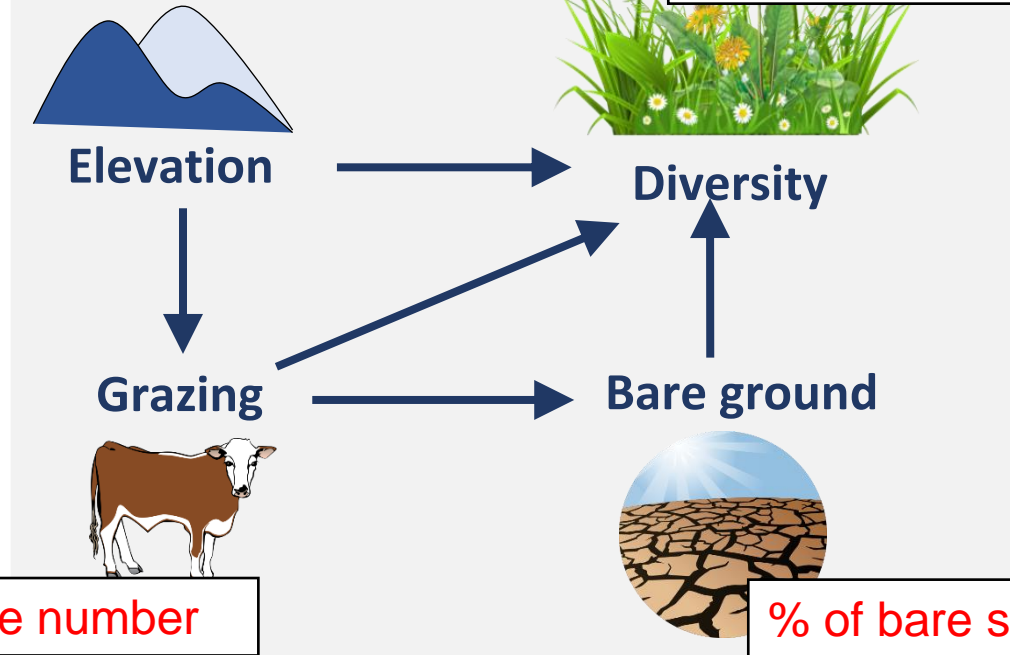
Cattle number

% of bare soil

Where to start?

Theory

- The literature
- Natural history
- Exploratory analyses
- Logical arguments
- Available data

Build a Model

Collect Data

Confront Model with Data
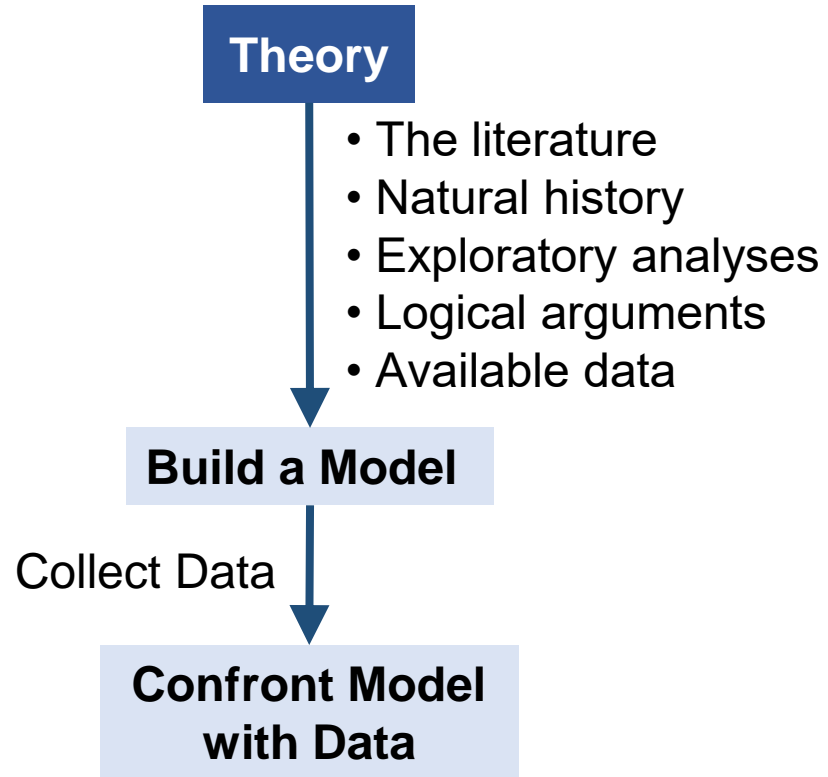
Buzhdygan, *et al.* 2020 *PLosONE*

Where to start?

```
data <- read.csv("Grassl_data.csv")
names(data)

# view the data
pairs(data)
```
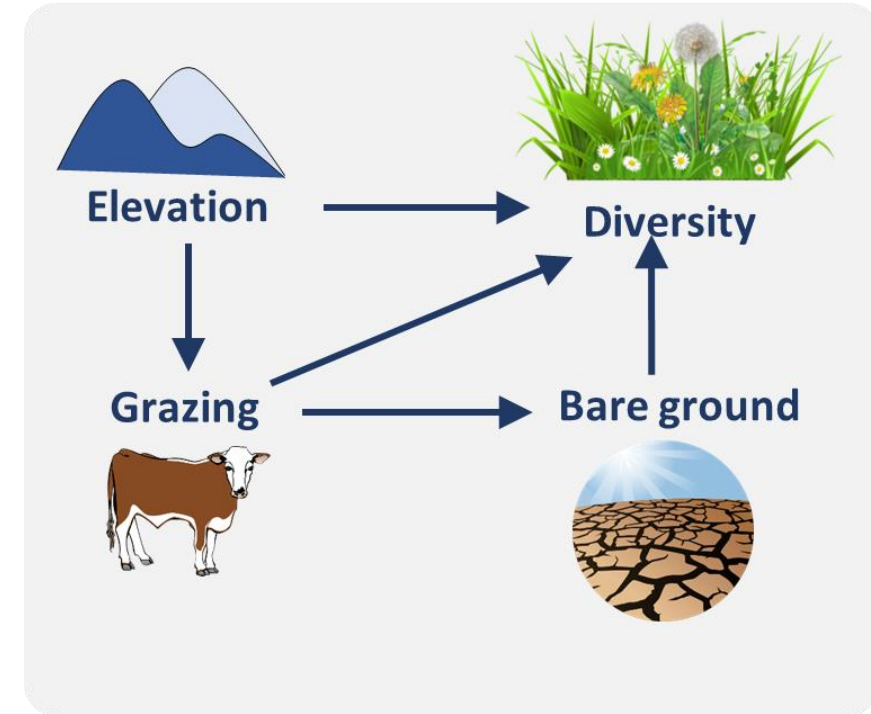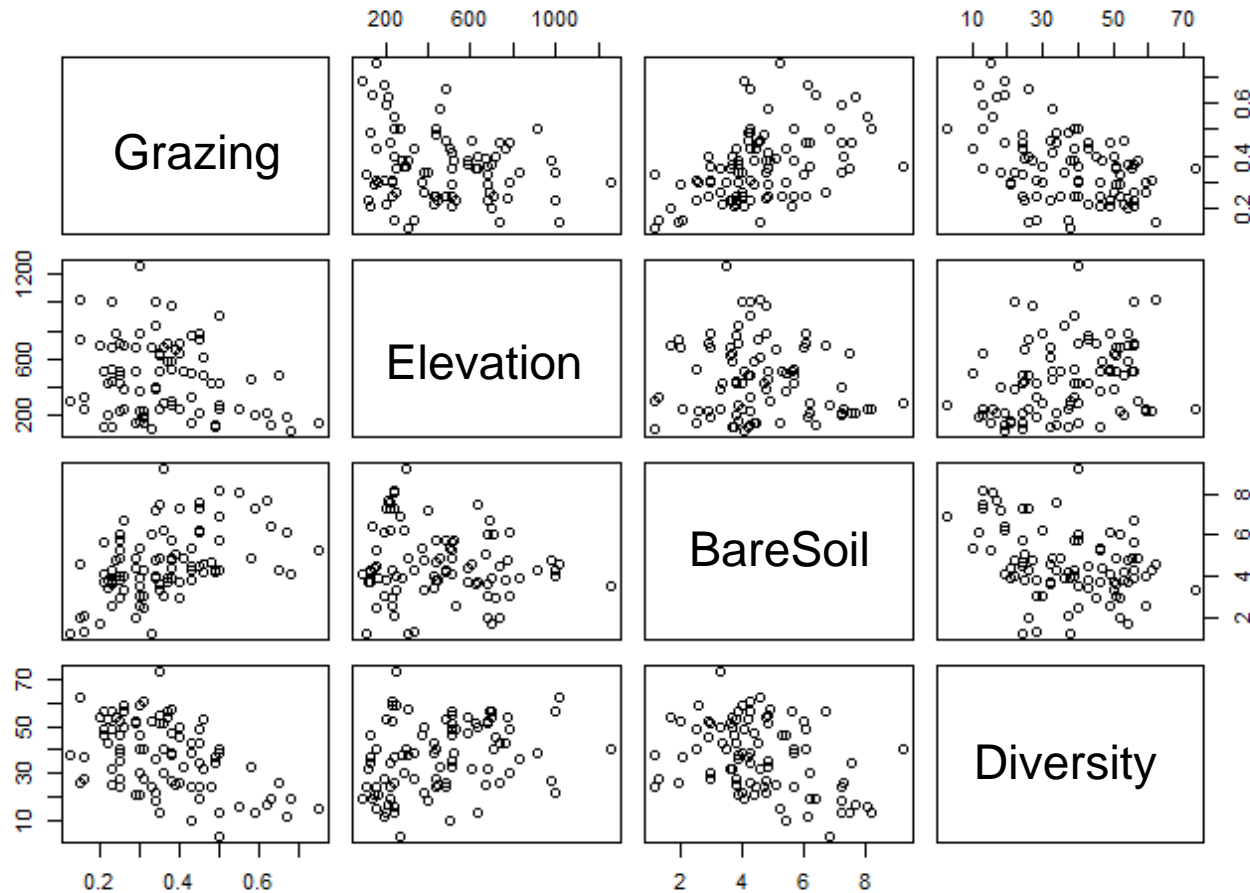

Buzhdygan, *et al.* 2020 *PLosONE*

Day 3 – Part 1   Outline

- Basics of SEM

  - ✓ From regression to SEM

  - ✓ SEM history. SEM in natural sciences.

  - ✓ SEM workflow process. Where do I start?
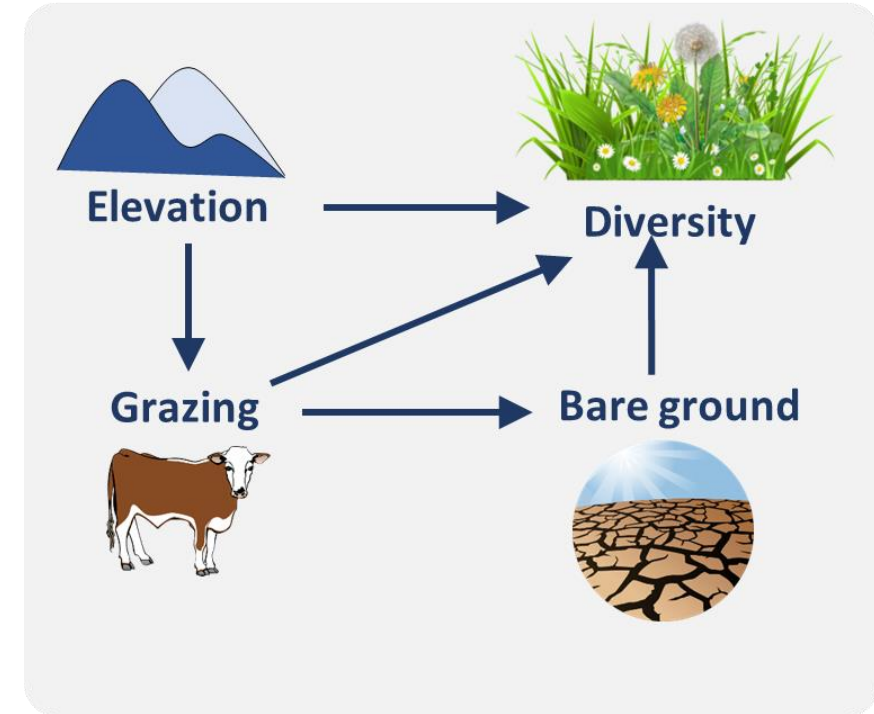
  - ✓ **First impression of 'lavaan'**

First impression of 'lavaan'

```
# Coding SEM
library(lavaan)

# Specify model structure

sem_mod <- '
    Grazing  ~  Elevation
    BareSoil  ~  Grazing
    Diversity ~ Elevation + Grazing + BareSoil
'
```


Buzhdygan, *et al.* 2020 *PLosONE*

First impression of 'lavaan'

```r
# Coding SEM
library(lavaan)

# Specify model structure

sem_mod <- '
    Grazing  ~  Elevation

    BareSoil  ~  Grazing

    Diversity ~ Elevation + Grazing + BareSoil
'
```
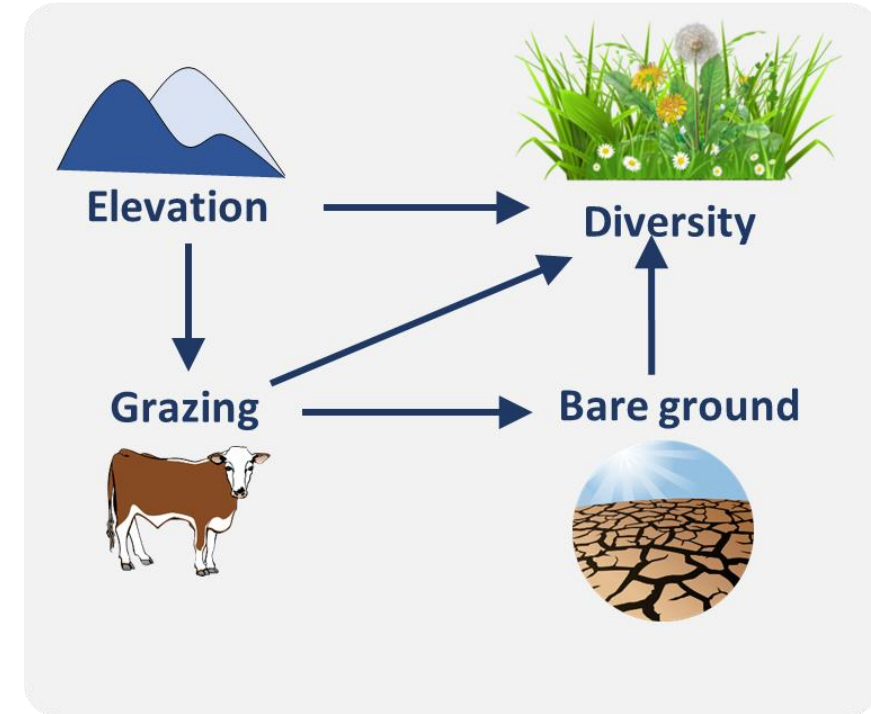
Specification operators in 'lavaan'

formula type	operator	meaning
Regression	~	"regressed on"
Correlation	~~	"correlated with"
Intercept	~ 1	"estimates intercept"
Latent variable	=~	"is measured by"
Composite	<~	"is caused by"

Path Diagram Notations:

Regression

Covariance

Observed variable

Error variance

Latent variable

Composite variable

First impression of 'lavaan'

What is lavaan?

https://lavaan.ugent.be

- Stands for **LA**tent **VA**riable **AN**alaysis

- Written by Yves Roseel in 2010

- Currently in version 6

- Uses R lm syntax

First impression of 'lavaan'

```r
# Coding SEM
library(lavaan)

# Specify model structure

sem_mod <- '
    Grazing  ~  Elevation
    BareSoil  ~  Grazing
    Diversity ~ Elevation + Grazing + BareSoil
`


# Estimate parameters, assess model fit
sem.fit <- sem(sem_mod, data=data)

# extract results
summary(sem.fit)
```

Theory

- The literature
- Natural history
- Exploratory analyses
- Logical arguments
- Available data

Build a Model

Collect Data

Confront Model with Data

Estimate Parameters, Assess Model Fit
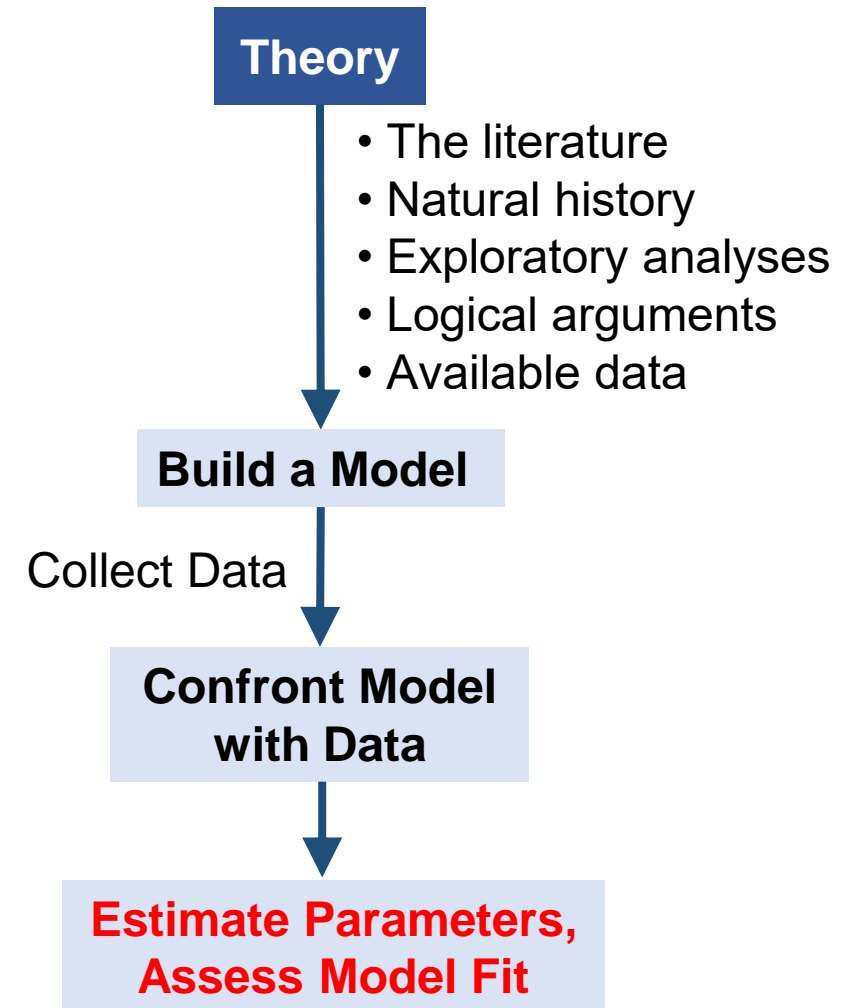
First impression of 'lavaan'

When you fit the model

```
# Error about data scales

Warning message:

In lav_data_full(data = data, group = group, cluster = cluster,  :

  lavaan WARNING: some observed variances are (at least) a factor 1000 times larger than
others; use varTable(fit) to investigate
```

First impression of 'lavaan'

```
# Call the model-implied covariance matrix


lavInspect(sem.fit, "obs")$cov


>

              Grazng      BareSl      Dvrsty         Elevtn
Grazing        0.017
BareSoil       0.102       2.685
Diversity     -0.904      -8.969      217.200
Elevation     -8.439     -55.722     1125.614      65289.346
```

- The covariance matrix is Ok, - there are no data problems.

- This is a likelihood algorithm problem – we can ignore the WARNING

- If you are worried about it, rescale data and see if answers change

First impression of 'lavaan'

```
# Check the data scales

varTable(sem.fit)


        name   idx nobs   type    exo  user    mean        var
1    Grazing    1   90 numeric    0     0     0.361      0.017
2   BareSoil    3   90 numeric    0     0     4.587      2.716
3  Diversity    4   90 numeric    0     0    37.022    219.640
4  Elevation    2   90 numeric    1     0   456.856  66022.934



 # Transform the data: recode vars to roughly same scale

 data$Diversity    <- data$Diversity/10

 data$Elevation <- data$Elevation/100


 # Repeat model estimation using transformed data
```

First impression of 'lavaan'

```
# extract results
summary(sem.fit)


lavaan 0.6-9 ended normally after 55 iterations


  Estimator                                      ML
  Optimization method                        NLMINB
  Number of model parameters                      8


  Number of observations                         90

Model Test User Model:


  Test statistic                              0.021
  Degrees of freedom                              1
  P-value (Chi-square)                        0.886
```

Assessed model fit

More soon!
(Day 3 part 3)

First impression of 'lavaan'

```
# Results from SEM model

> summary(sem.fit , standardize = T)

Parameter Estimates:
Regressions:
```

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
Grazing ~						
Elevation	-0.000	0.000	-2.475	0.013	-0.000	-0.252
BareSoil ~						
Grazing	5.963	1.161	5.136	0.000	5.963	0.476
Diversity ~						
Elevation	0.011	0.005	2.062	0.039	0.011	0.190
Grazing	-37.259	11.739	-3.174	0.002	-37.259	-0.331
BareSoil	-1.696	0.913	-1.856	0.063	-1.696	-0.189

Standardised coefficients for latent variables

Standardised coefficients for all variables

Raw unstandardized coefficients

standard errors

Wald statistic

Probability of a z this big by chance

40

First impression of 'lavaan'

```
# Results from SEM model

> summary(sem.fit , standardize = T)


Parameter Estimates:
Regressions:
```

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
Grazing ~						
Elevation	-0.000	0.000	-2.475	0.013	-0.000	-0.252
BareSoil ~						
Grazing	5.963	1.161	5.136	0.000	5.963	0.476
Diversity ~						
Elevation	0.011	0.005	2.062	0.039	0.011	0.190
Grazing	-37.259	11.739	-3.174	0.002	-37.259	-0.331
BareSoil	-1.696	0.913	-1.856	0.063	-1.696	-0.189

Standardised coefficients for latent variables

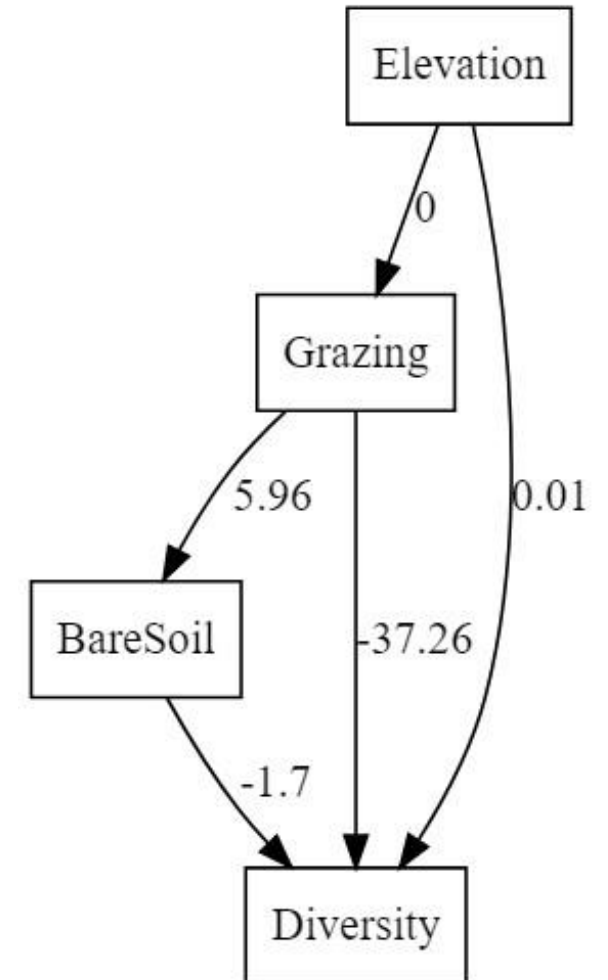Standardised coefficients for all variables

Raw unstandardized coefficients

standard errors

Wald statistic

Probability of a z this big by chance

```
library(lavaanPlot)
lavaanPlot(model = sem.fit,
coefs = TRUE)
```
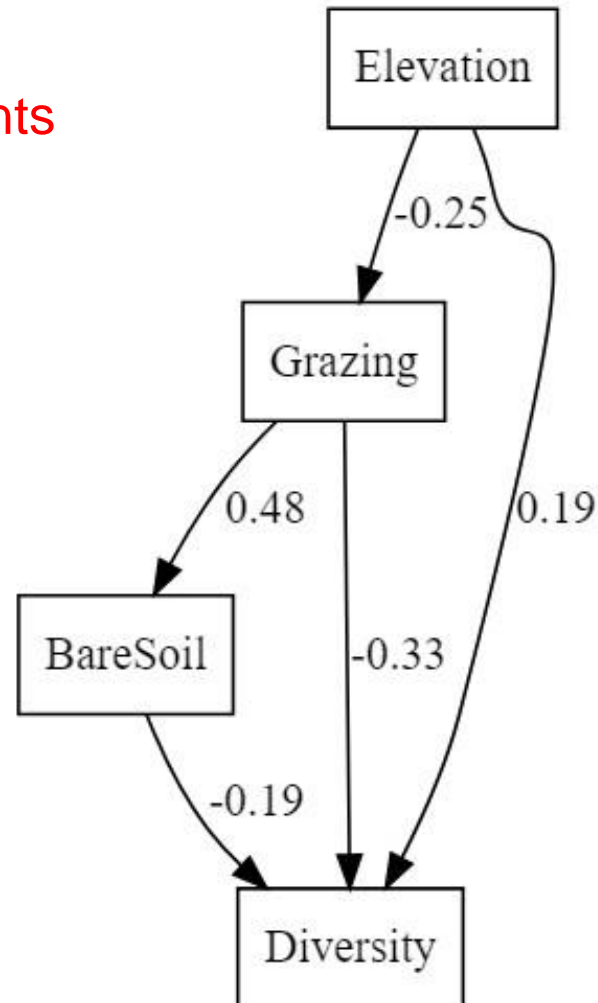

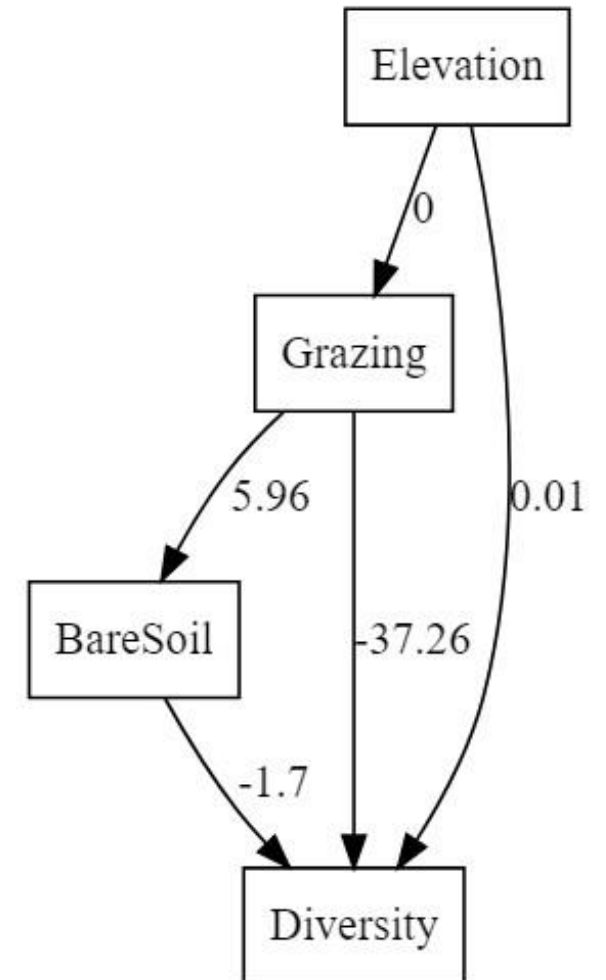
41

First impression of 'lavaan'


```
lavaanPlot(model = sem.fit,
coefs = TRUE, stand=TRUE)
```

```
lavaanPlot(model = sem.fit,
coefs = TRUE)
```

Standardised coefficients
• comparable across the entire model
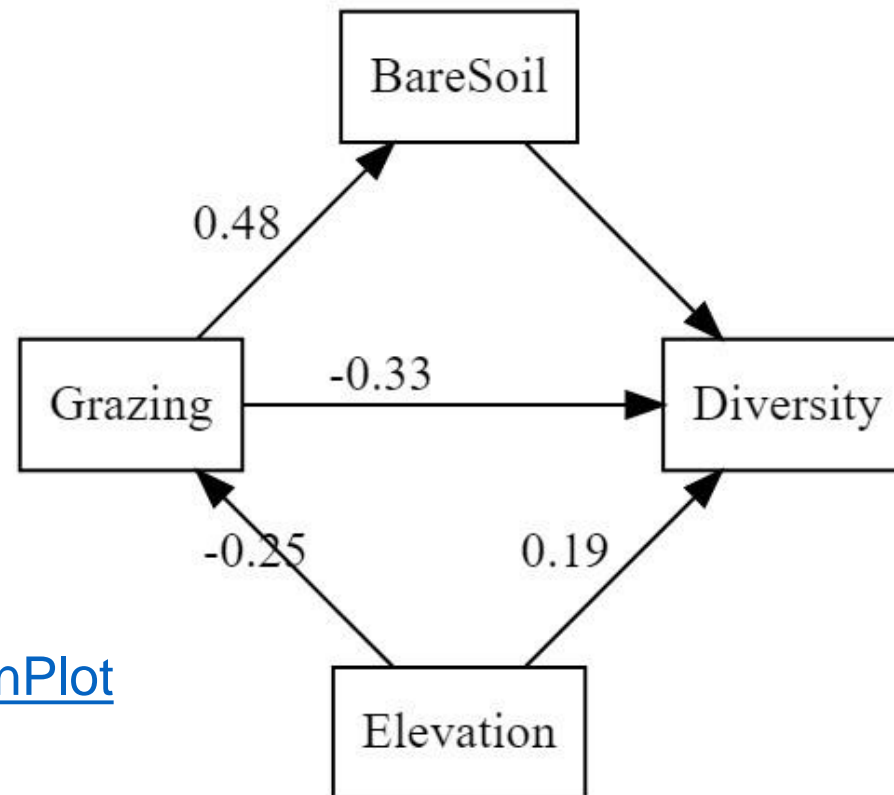
First impression of 'lavaan'

```
lavaanPlot(model = sem.fit,
           coefs = TRUE, stand=TRUE,
graph_options = list(layout = "circo"),
                     sig = 0.05)
```

changes layout

only shows coefficients p≤0.05

See more:
https://cran.r-project.org/web/packages/lavaanPlot

First impression of 'lavaan'

Characters in node labels

```
library(semPlot)
semPaths(sem.fit, what='std', nCharNodes=6, sizeMan=10,
         edge.label.cex=1.25, curvePivot = TRUE,
         fade=FALSE)
```
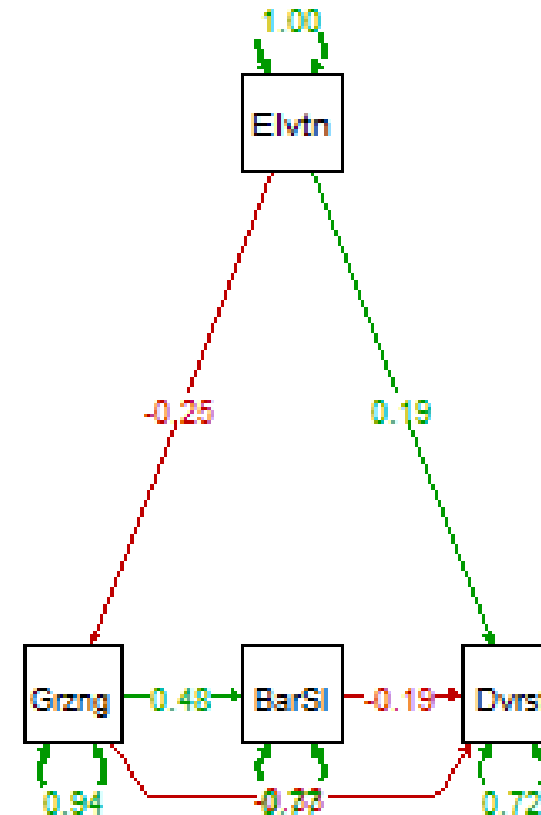
Curved links

No transparency
of links

See more: