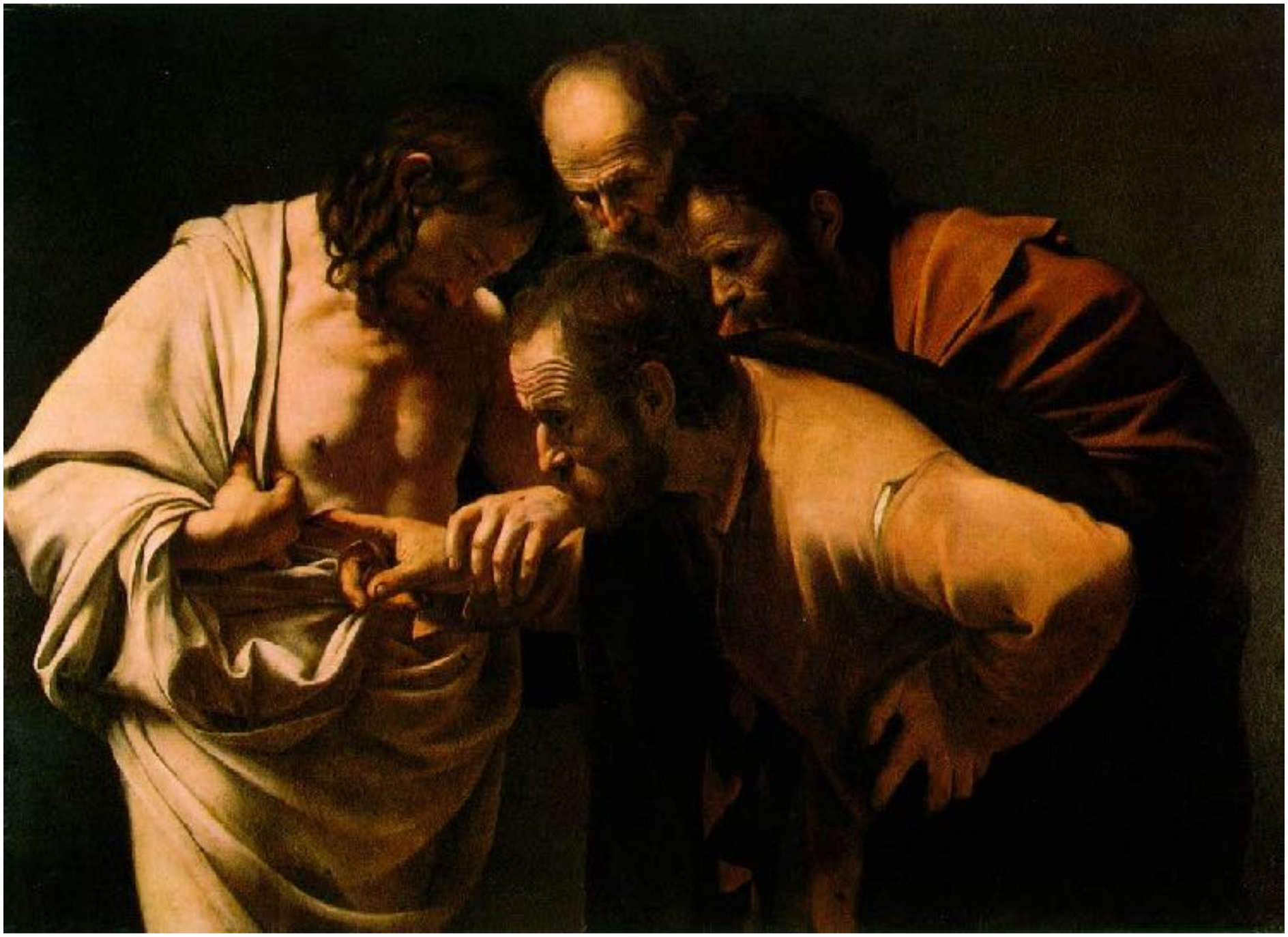
Tanjona Ramiadantsoa

A history of your new skills

E2M2 2020
Valbio Ranomafana
January 10th, 2020

What is a scientist?



i Steps in a modeling project

1. *Development of the study concept and question*
2. *Literature review*
3. *Data collection*
4. *Construction of model framework*
5. *Model analyses and selection*
6. *Model validation*
7. *Manuscript writing and submission*

ii Literature review

- Who has tried to answer this before and how did they do it?
  - Empirical studies
  - Modeling studies
- What are these studies short-comings?
- Are there already parameter estimates or data sets to help you answer

vi Manuscript writing and submission

- What are the main results that provide the answer to my question?
  - 1 to 3 graphs
  - 1 to 3 tables
- What is the journal that best fits my study?
  - Scope, audience, impact factor, math focus
- How do I present my manuscript?
  - Introduction: set the stage to your question
  - Methodology: describe explicitly all steps for replicability
  - Results: clear and concise
  - Discussion: explain how your study improves previous knowledge

ii Literature review

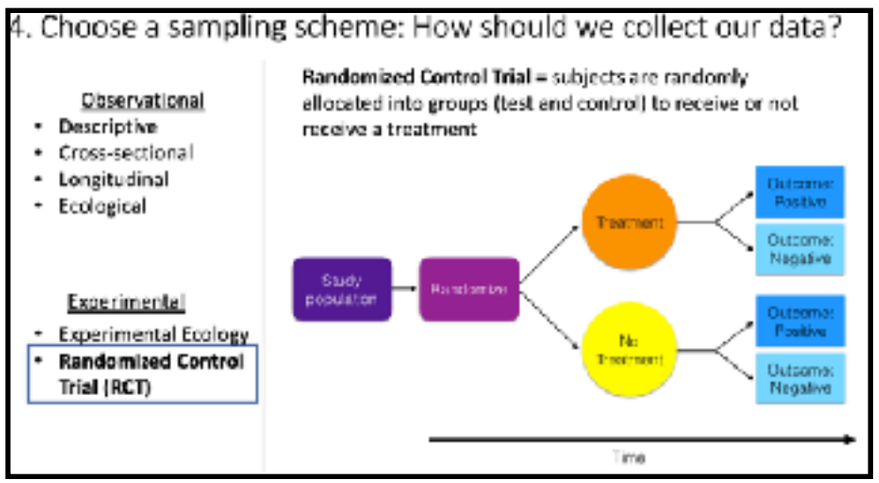
- Who has tried to answer this before and how did they do it?
  - Empirical studies
  - Modeling studies
- What are these studies short-comings?
- Are there already parameter estimates or data sets to help you answer

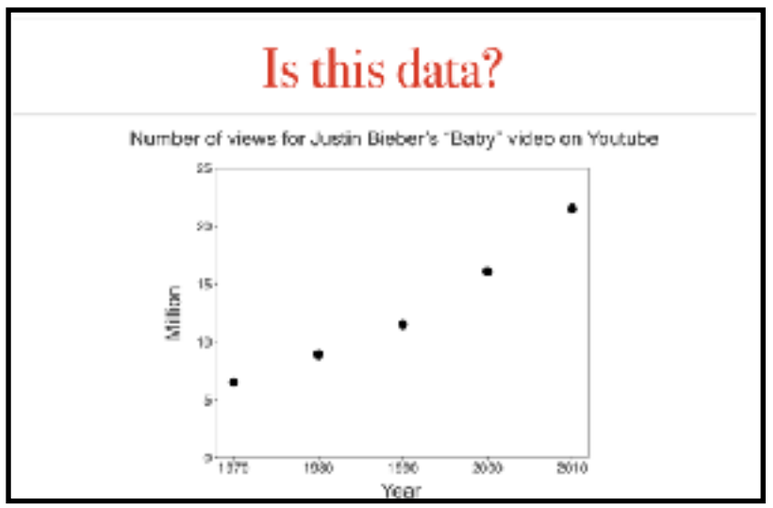
To help you prepare, try to answer the next few questions after reading:

- What are the context and key points of the paper?
- Identify the problem that inspired the authors to write the article.
- Do you agree with the methods, the results, and the discussion?
- Where do the authors use models to address their questions? What type of models can you identify?
- What do you like and what do you dislike about the paper?
- What did you not understand?
- What have you learned from the paper?
- And come up with three questions of your own.

Study Design

1. Define the **research question**: What are we trying to answer?
2. Define the **sample type**: What data do we need to answer our question?
3. Identify a **system**: Where can we collect our data?
4. Choose a **sampling scheme**: How should we collect our data?
5. Acknowledge **limitations**: What can we actually infer from our data?
6. Outline a **data organization plan**: How should we organize our data?
7. Be **flexible**: How can we prepare for potential/unanticipated challenges?





Take homes

- Research question and hypothesis are more important that models
- Any data needs context: the X and Y should be clear
- Models are rigorous tools to assess how the data support the claim
  - There are figuratively an infinite number of models
  - Statistical model works with question starting with **what**
  - Mechanistic model **generates data** and works with question starting with **how**

Two broad classes of models

Statistical

Correlative

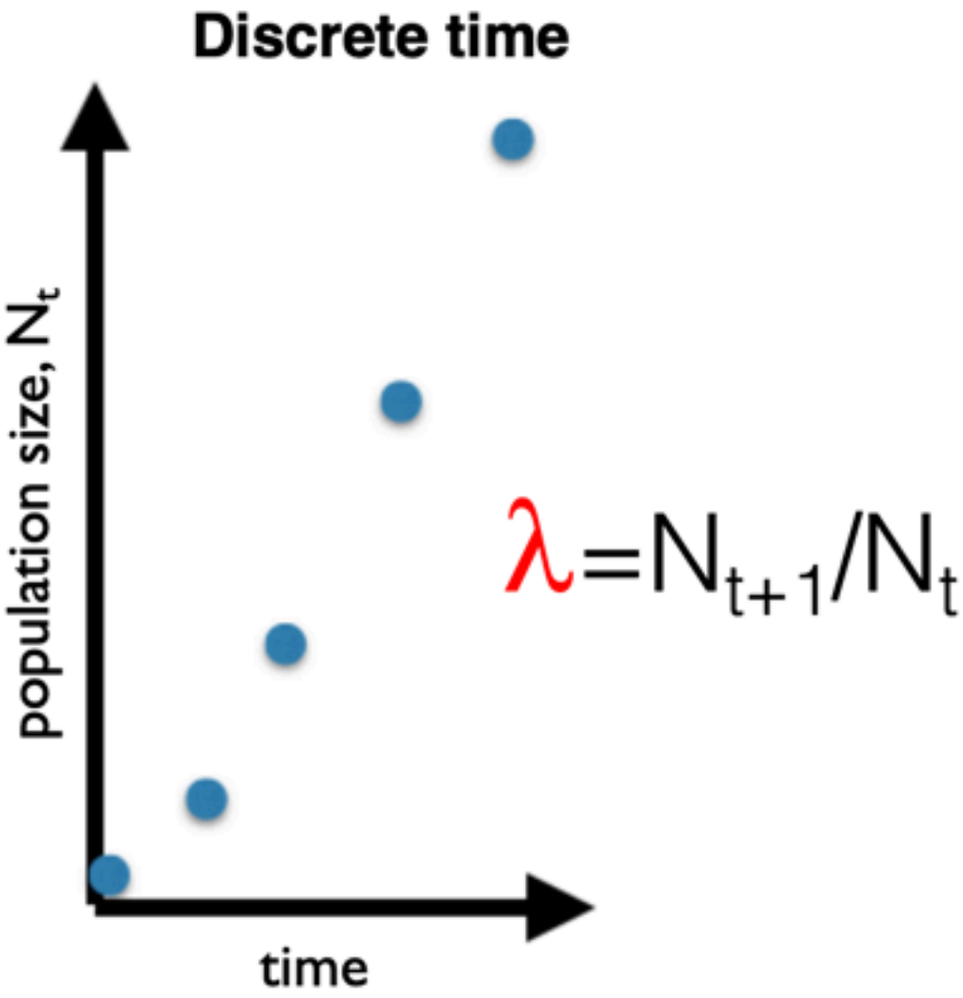
Mechanistic

Causative

Mechanistic (mathematical) model

- **Process-driven** with question starting with **How**
- You generate data with simulation

Discrete vs. continuous

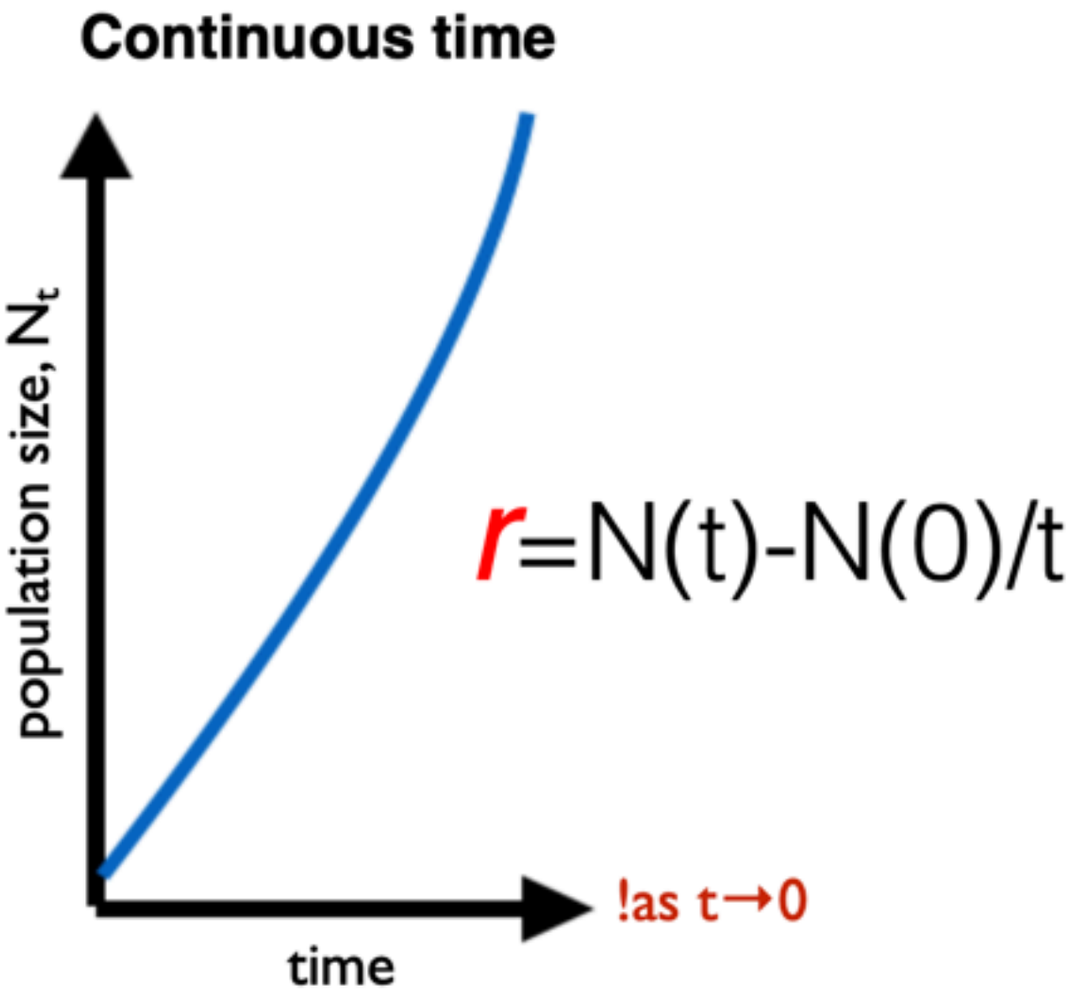


$N_1 = \lambda N_0$

$N_2 = \lambda[\lambda N_0] = \lambda^2 N_0$

$N_3 = \lambda^3 N_0$

$N_t = \lambda^t N_0$

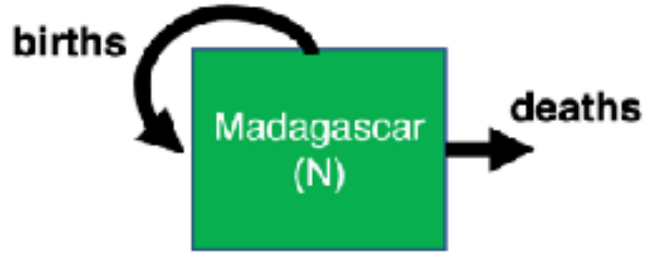


$dN(t)/dt = rN(t)$

$N(t) = N(0)e^{rt}$

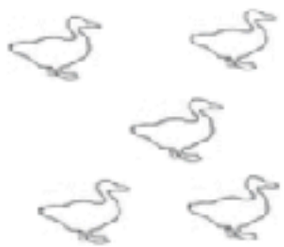
Deterministic vs. stochastic

The basic population model



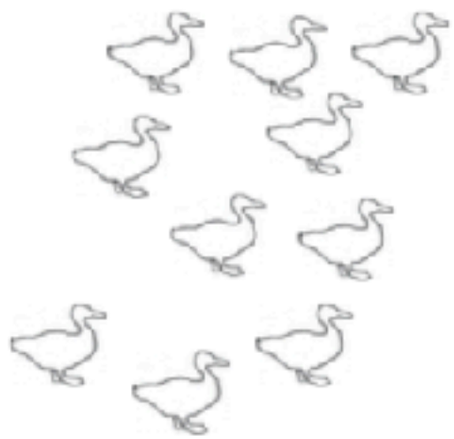
starting population

probability of death = 0.5



if deterministic "always the same"

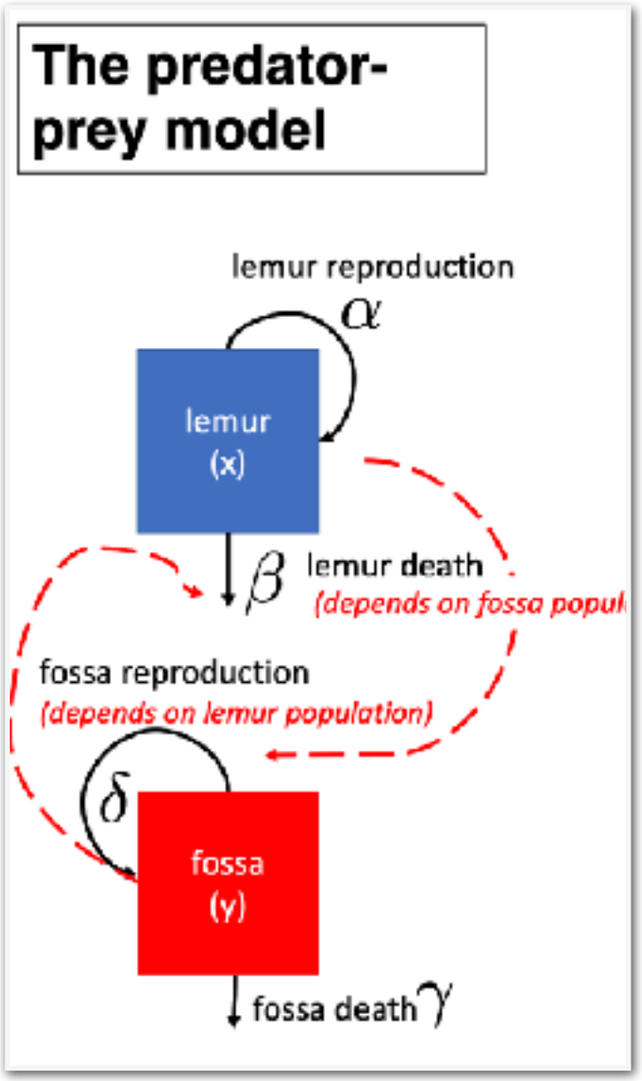
starting population

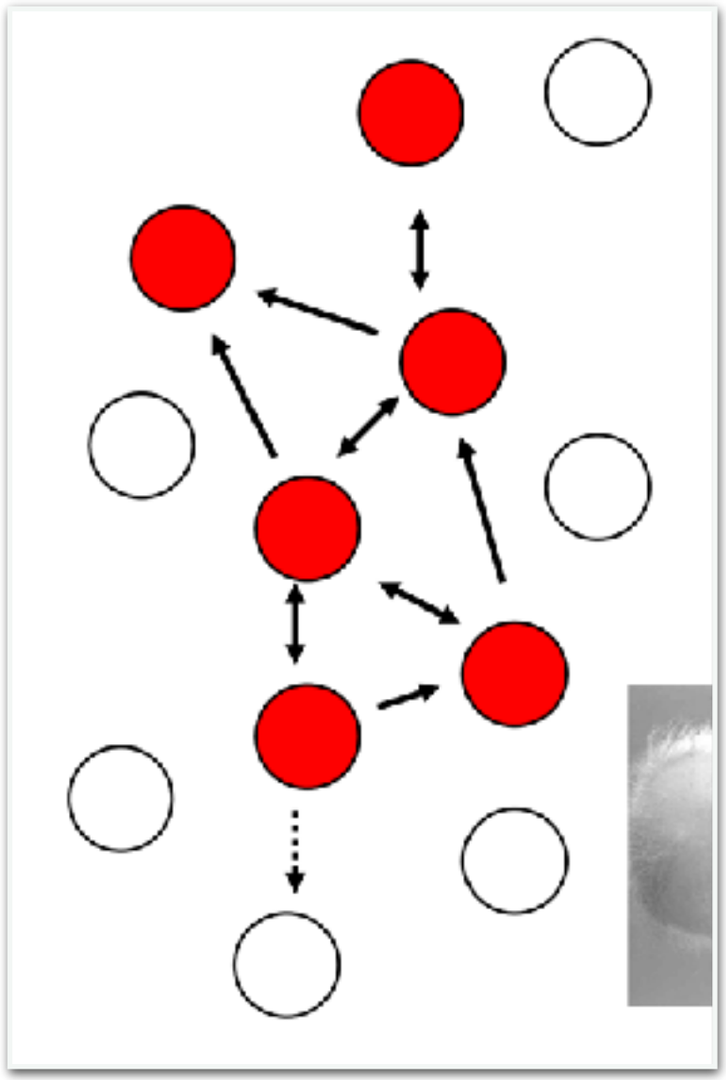


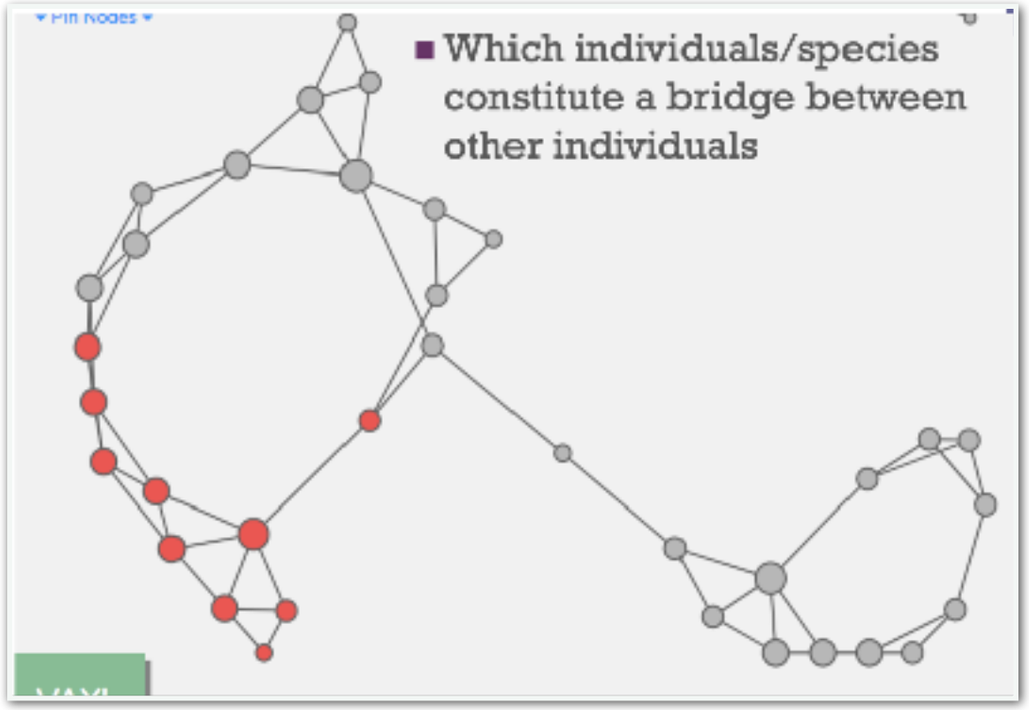
probability of death = 0.5

if stochastic? "up to chance"

Non-spatial vs. spatial vs. network







Compartment models

The predator-prey model

lemur reproduction α

lemur (x)

β lemur death *(depends on fossa population)*

fossa reproduction *(depends on lemur population)*

δ

fossa (y)

fossa death γ

Compartmental models (Mechanistic Models)

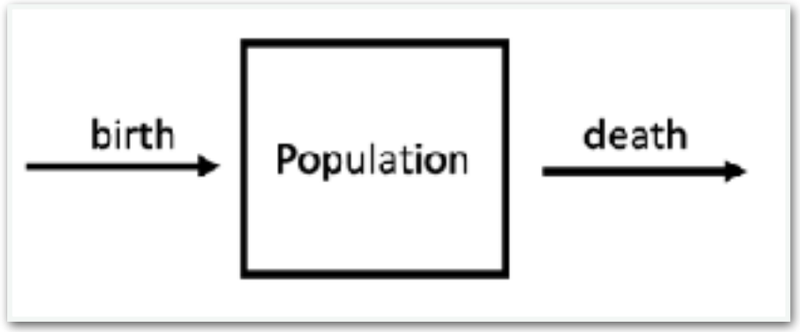
1. Populations are divided into compartments
2. Individuals within a compartment are homogenously mixed
3. Compartments and transition rates are determined by biological systems
4. Rates of transferring between compartments are expressed mathematically

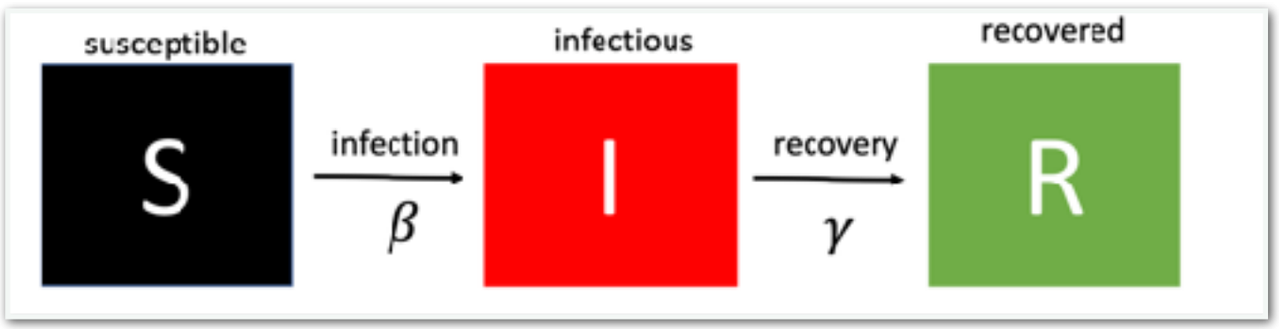
$$\frac{dx}{dt} = x(\alpha - \beta y)$$

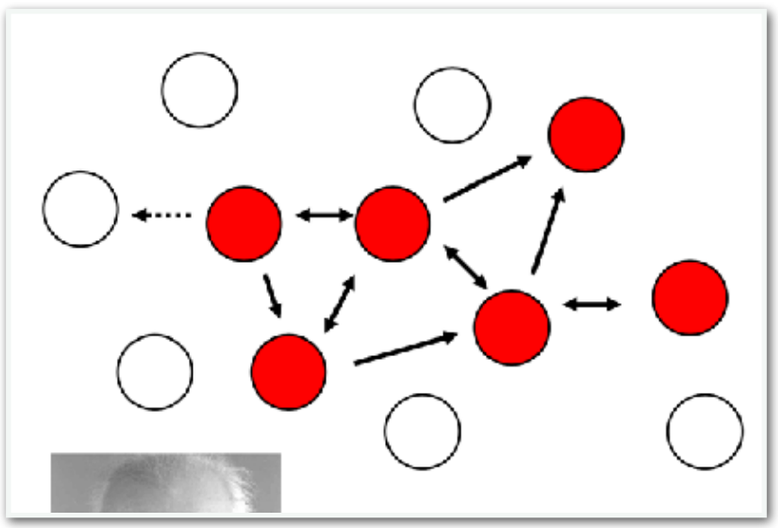
$$\frac{dy}{dt} = y(\delta x - \gamma)$$

SOME ASSUMPTIONS

- the **lemur** has an unlimited food supply
- the **lemur** only dies from being eaten by fossa
- the **fossa** is totally dependent on a single prey species (the lemur) as its only food supply



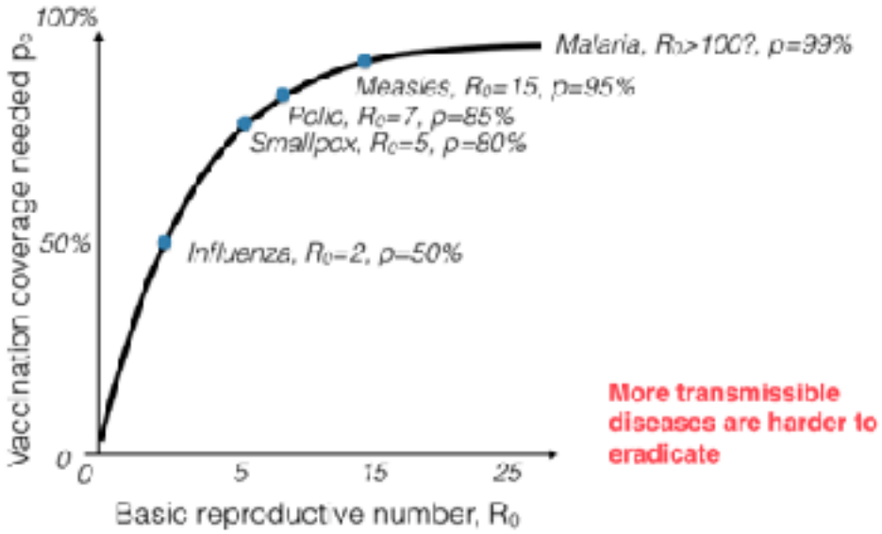




Insights

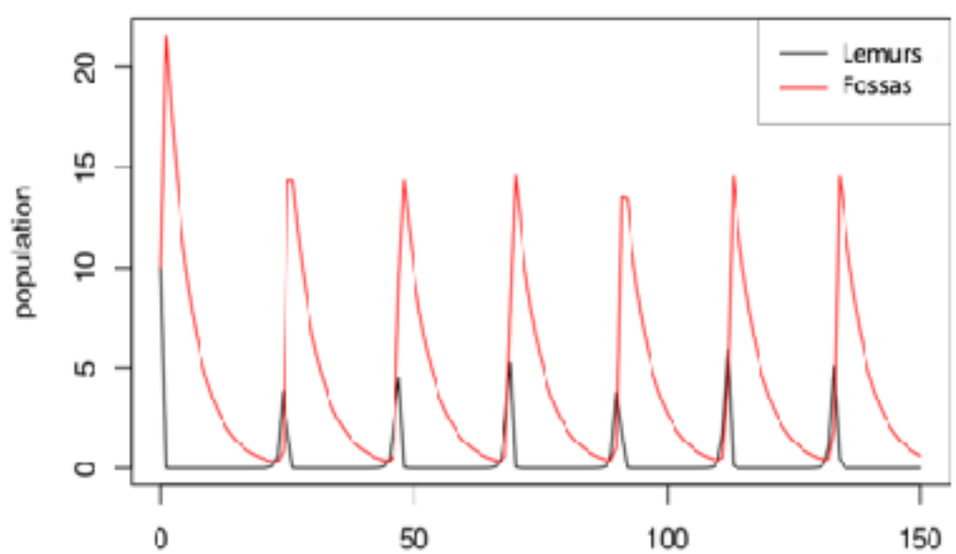
The SIR model : eradication

$$p_c = 1 - \frac{1}{R_0}$$

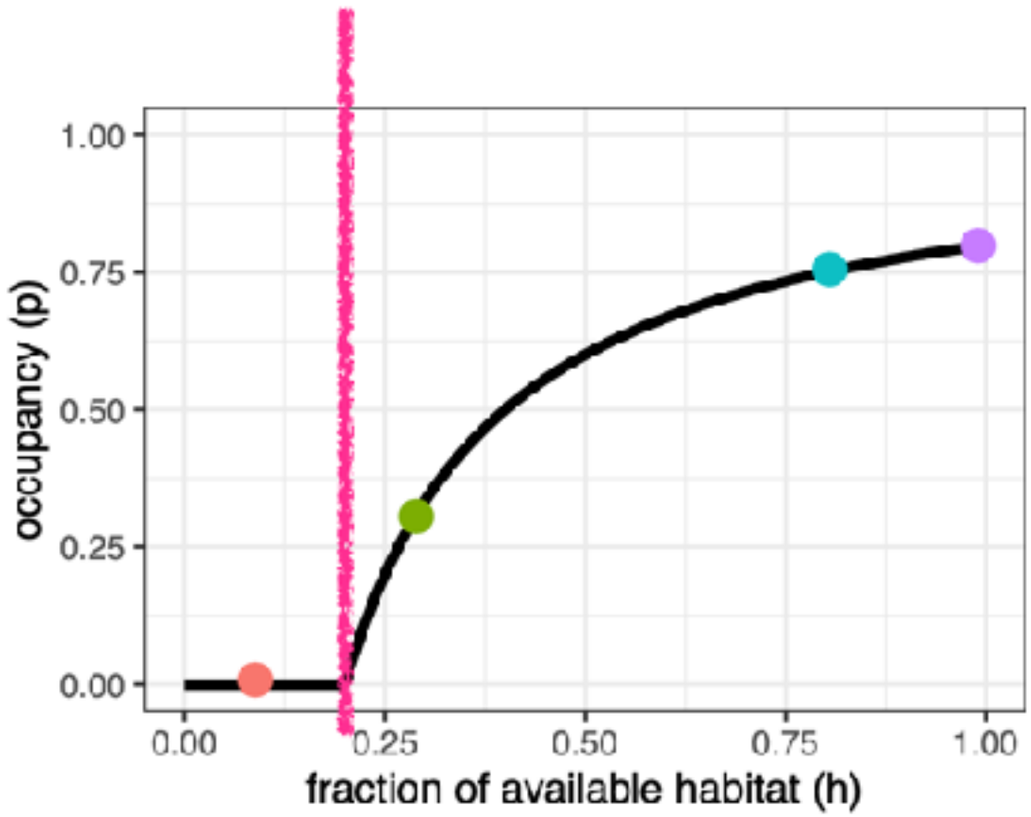


More transmissible diseases are harder to eradicate

The predator-prey model



Extinction threshold = the root



Extinction threshold: $h = \frac{e}{c}$

Two broad classes of models

Statistical

Correlative

Mechanistic

Causative

Statistical model

- **Data-driven** with question starting with **What**
- Test patterns in data using predefined functions

Statistical model

- **Data-driven** with question starting with **What**
- Test patterns in data

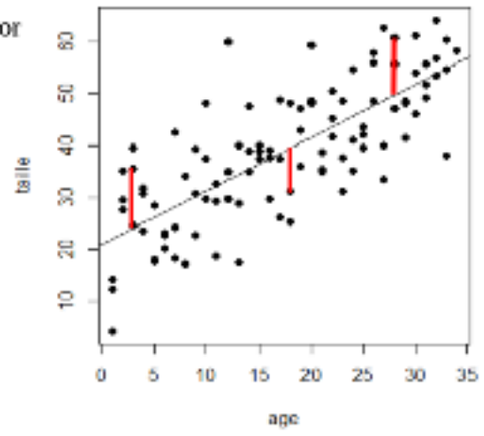
POSITIVE PROOF OF GLOBAL WARMING

Correlation does not imply causation

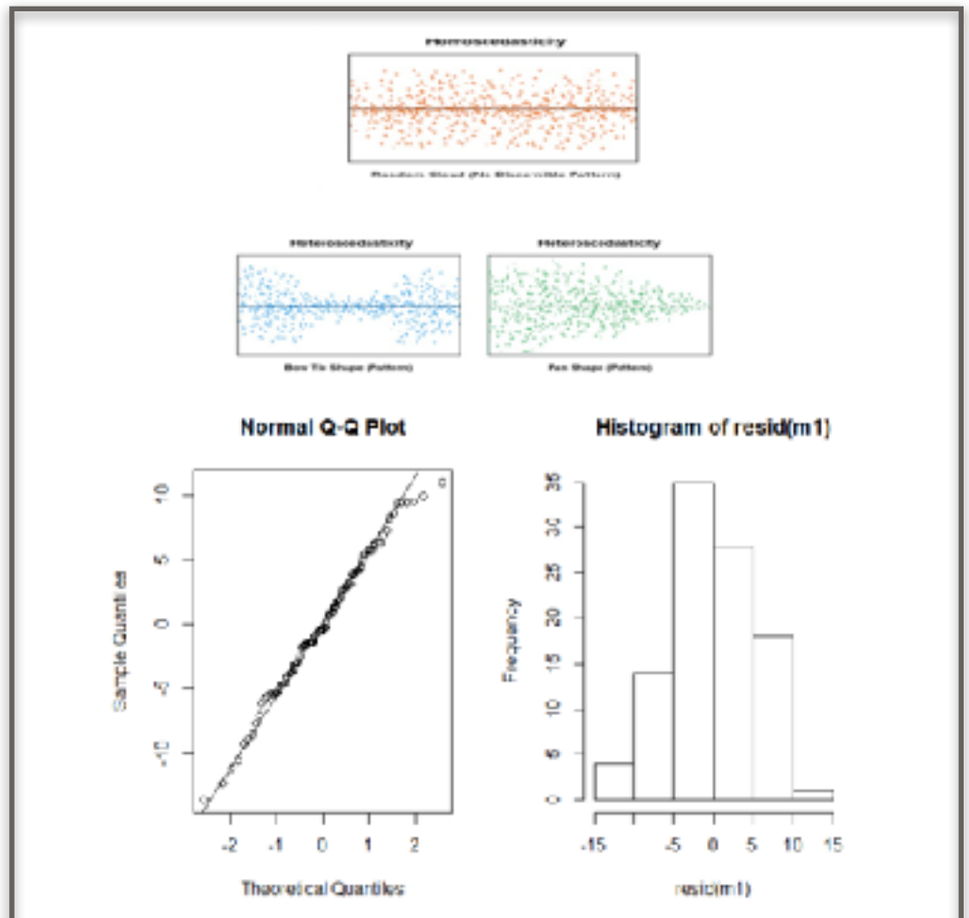
Univariate linear model: simple linear regression

- Quantify the relationship between the response variable and each explanatory variable
- Linear relationship: $y = a + bx + \epsilon$
  - y: response variable, x: explanatory variable
  - a: intercept, b: slope, ϵ: Error or residual
- Minimize the error

Taille=20+1.15 *Age(Months)+Error



Model validation



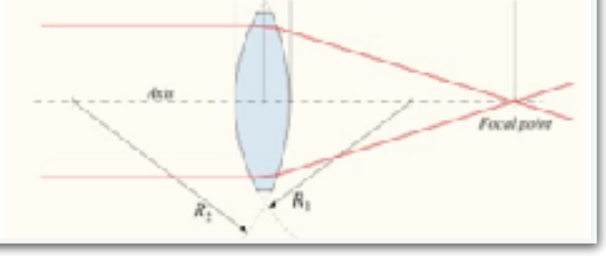
Multilinear model

```
Call:
lm(formula = taille ~ age + sexe + GIparasites + malaria, data = lemur.

Residuals:
     Min       1Q   Median       3Q      Max
-12.3696  -4.2168   0.0111   3.8715   9.9466

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 24.37448    1.40044  17.405  < 2e-16 ***
age          0.87527    0.05423  16.141  < 2e-16 ***
sexeMale    10.20143    1.04410   9.771 5.11e-16 ***
GIparasites -0.30170    0.02601 -11.598  < 2e-16 ***
malariaOui  -0.10413    1.04603  -0.100    0.921
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.203 on 95 degrees of freedom
Multiple R-squared:  0.8463,    Adjusted R-squared:  0.8399
F-statistic: 130.8 on 4 and 95 DF,  p-value: < 2.2e-16
```

Generalized linear model

- Extend the linear model framework by using a linear predictor and a link function
- link function: describe the relationship betweeen the linear combination of the explanatory variables and the mean of the response variable
- Rcommand: glm(**response_variable~ explanatory_variable,family=** family_distribution)

Most common family function :
Gaussian : Identity
Binomial : logit
Poisson : log
Neg binomial : log



Why use GLMMs?

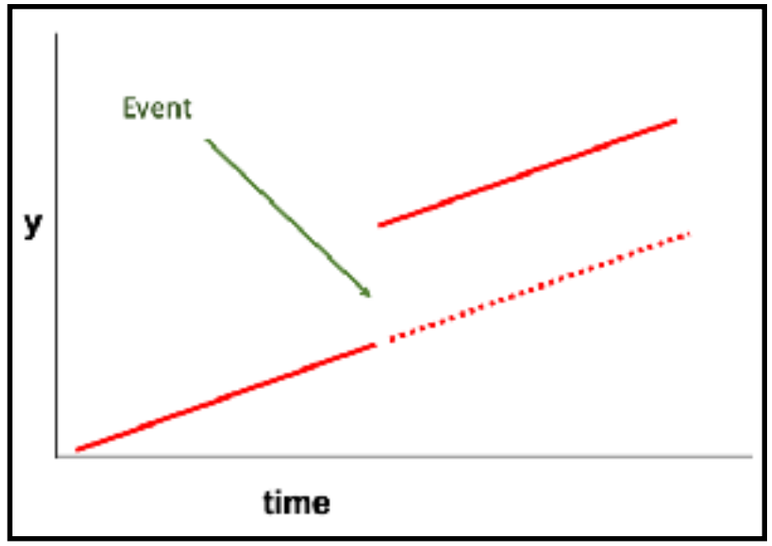
Generalized linear mixed models include both fixed effects and random effects in order to allow for:

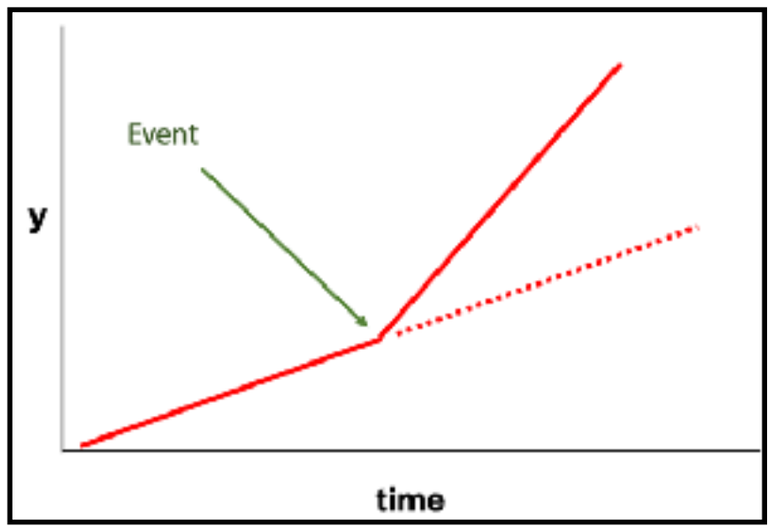
- Repeated measures
- Temporal correlation
- Spatial correlation
- Heterogeneity
- Nested data

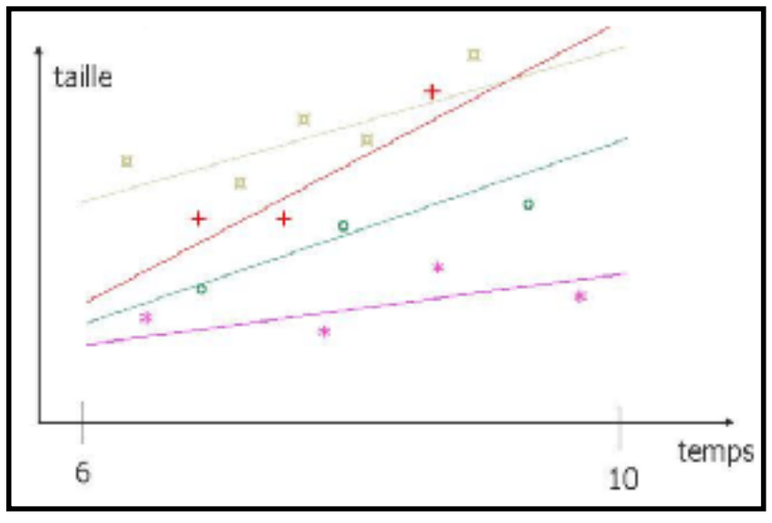
$$y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

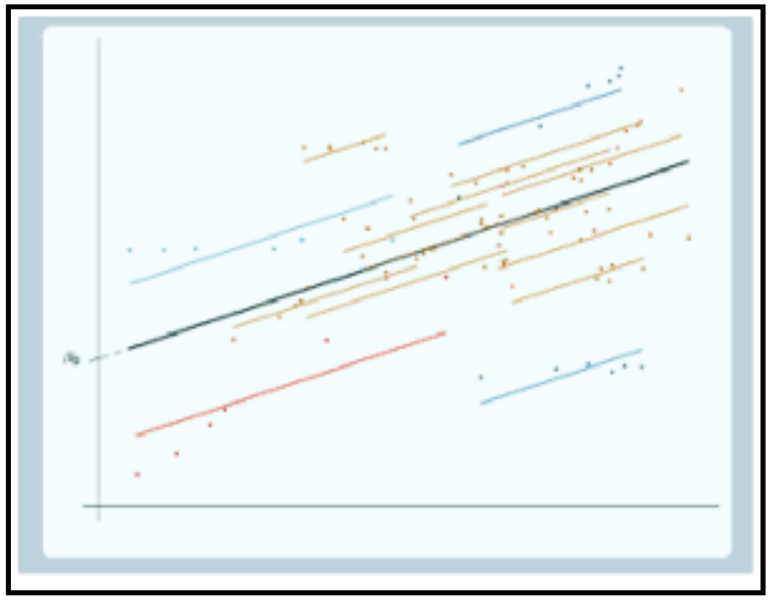
Fixed Effects *Random Effects*

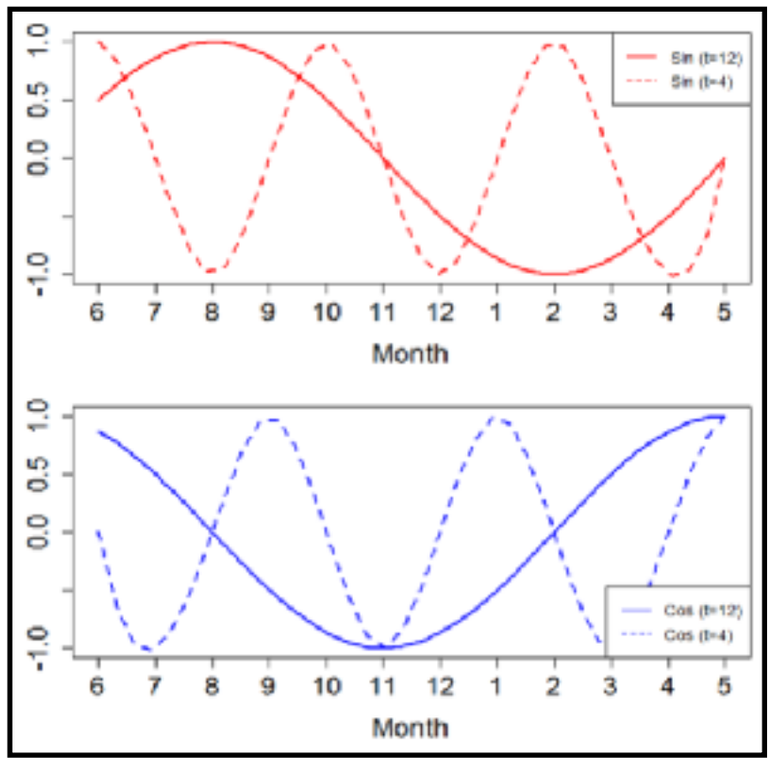
The R function to fit a generalized linear mixed model is glmer() which uses the form
fitted.model <- glmer(formula, family="model family", data=data.frame)











Network analyses

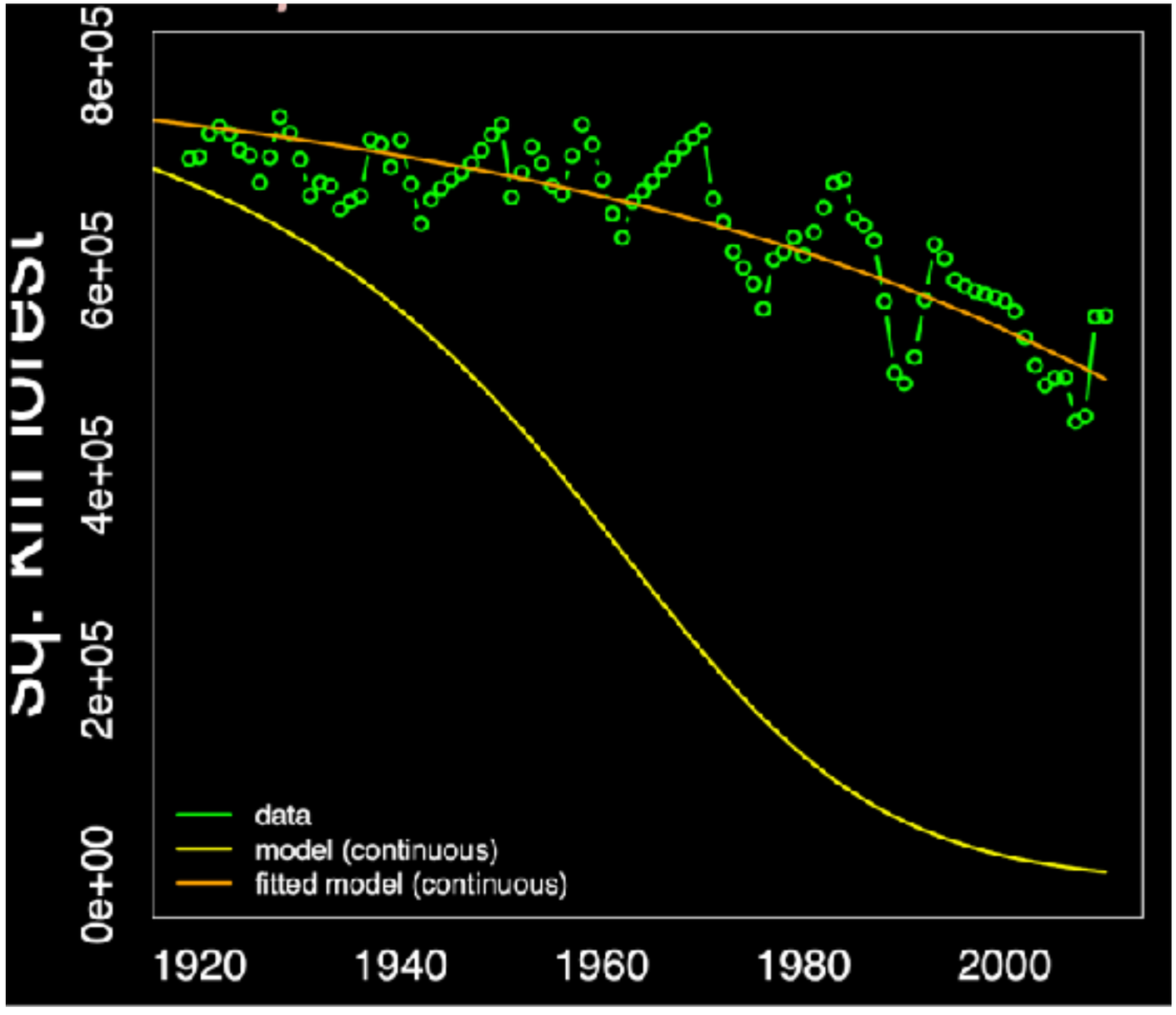
MRQAP (Quadratic Assignment Procedure)

- Multiple Regression Quadratic Assignment Procedure
  - Basically logistic regression analysis applied to matrix data.
  - Is your response variable linked to explanatory variable 1 while controlling for all other variables?

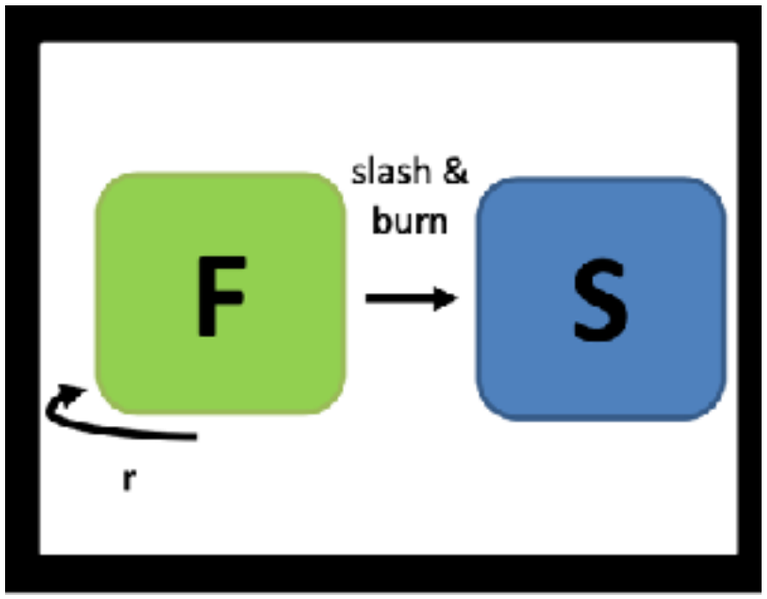
Fit mechanistic models to data

Fit mechanistic model to data

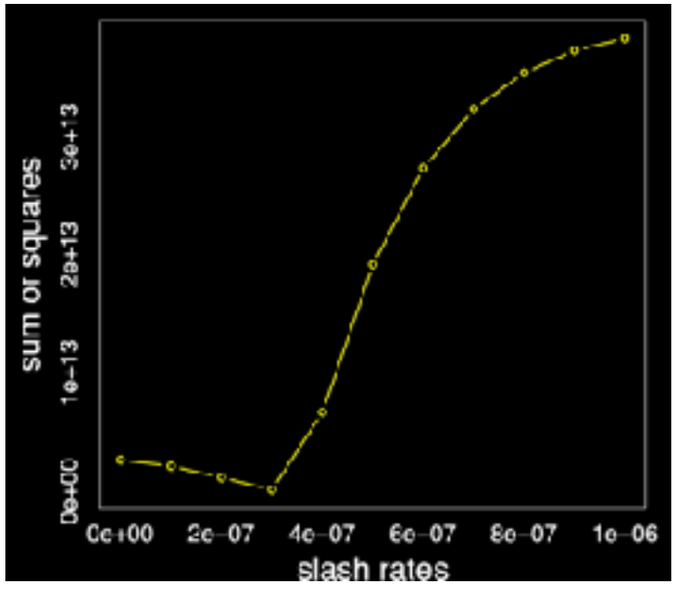
Combines the best of mechanistic and statistical method



Mechanistic model



Minimize sum of squares



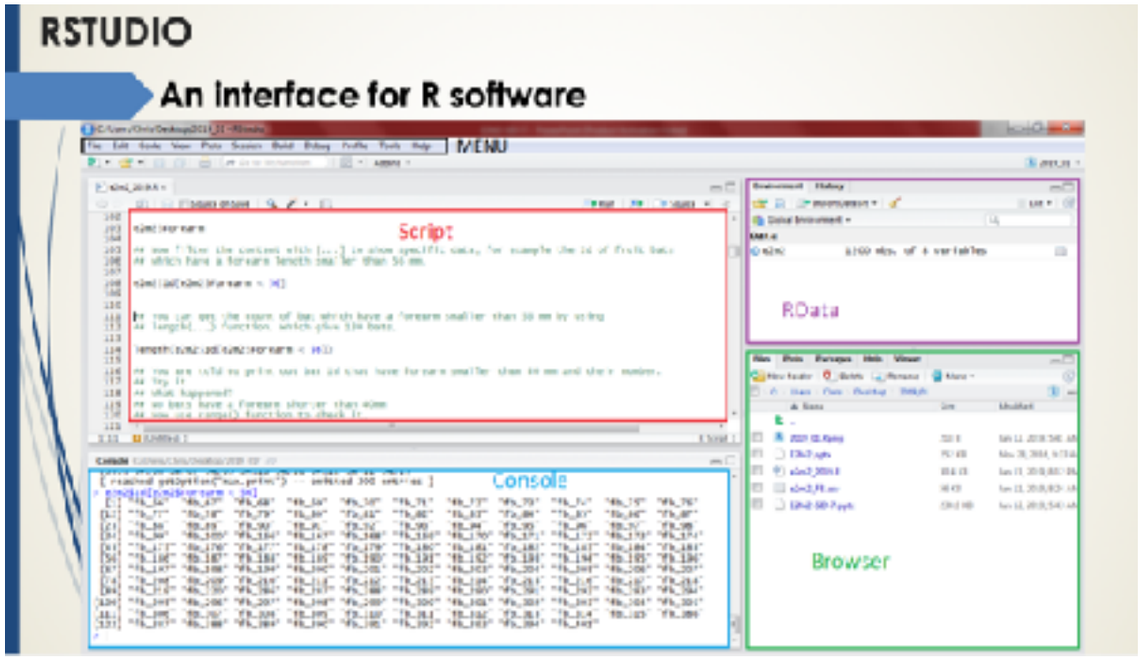
Model selection

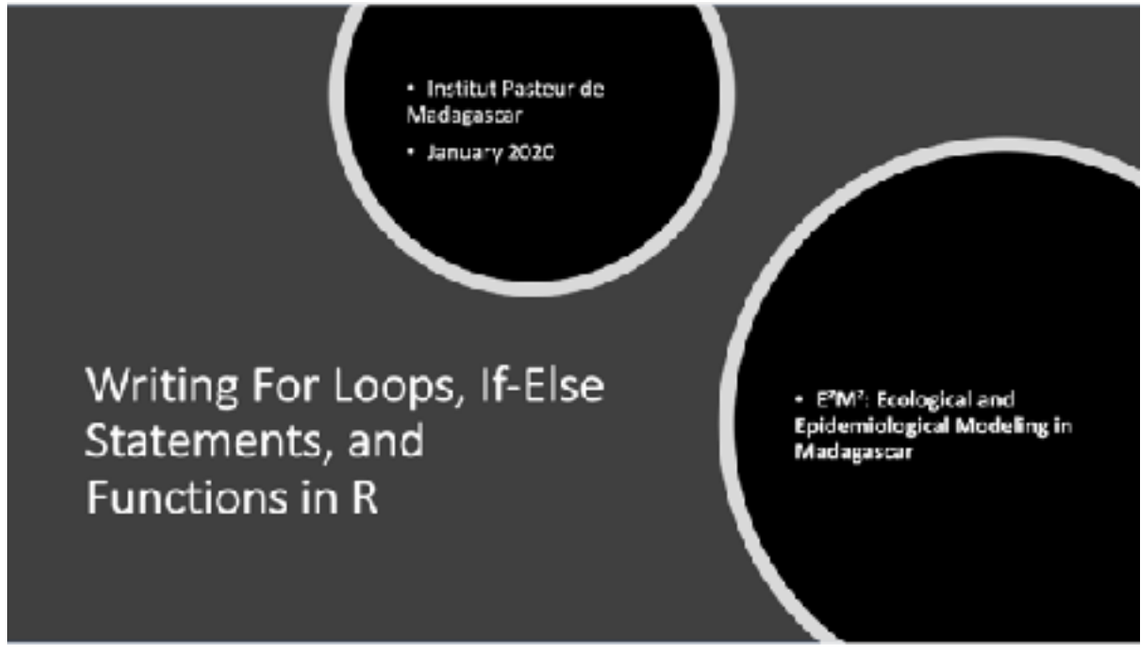
- Some methods
  - R^2, adj-R^2
  - Least square
  - Log likelihood
  - AIC
  - BIC
  - RMSE
  - …
- There is no consensus among statisticians…it is an art

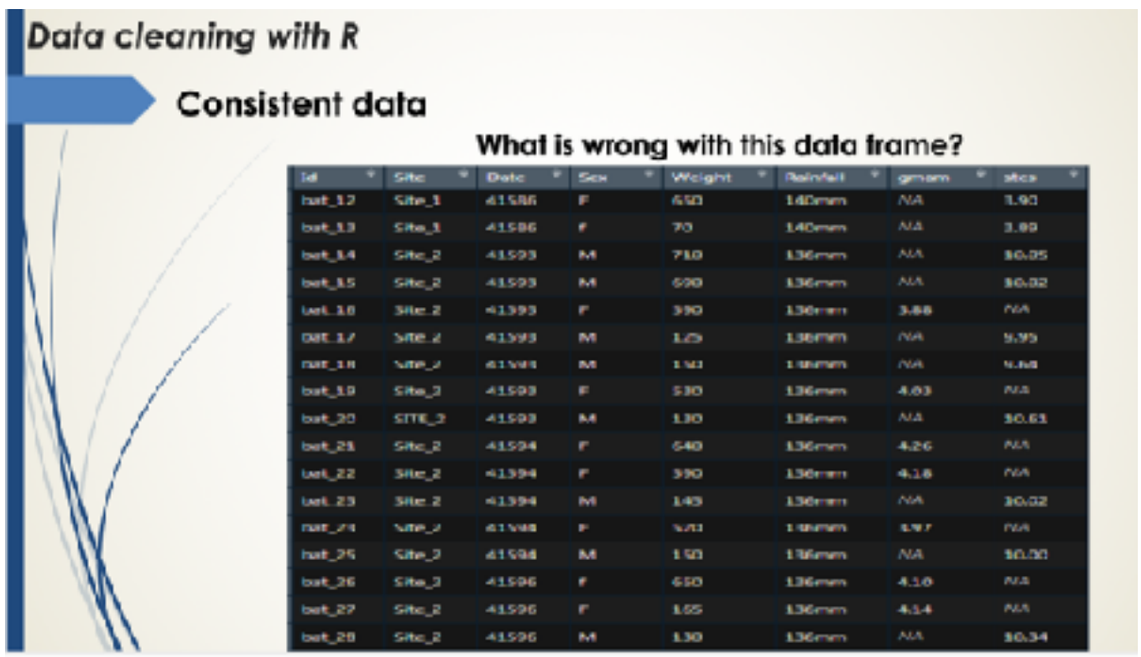
You can always fit a model,
and select the best model
But it is just the best based on what you tried

More on quantitative skills

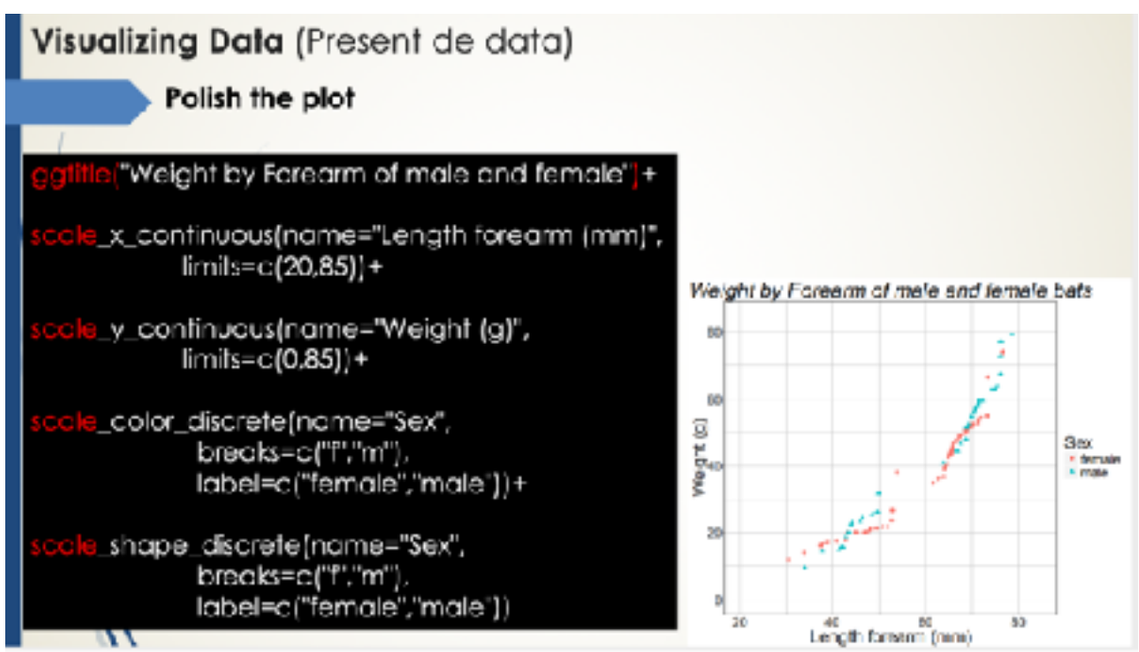
Basic R





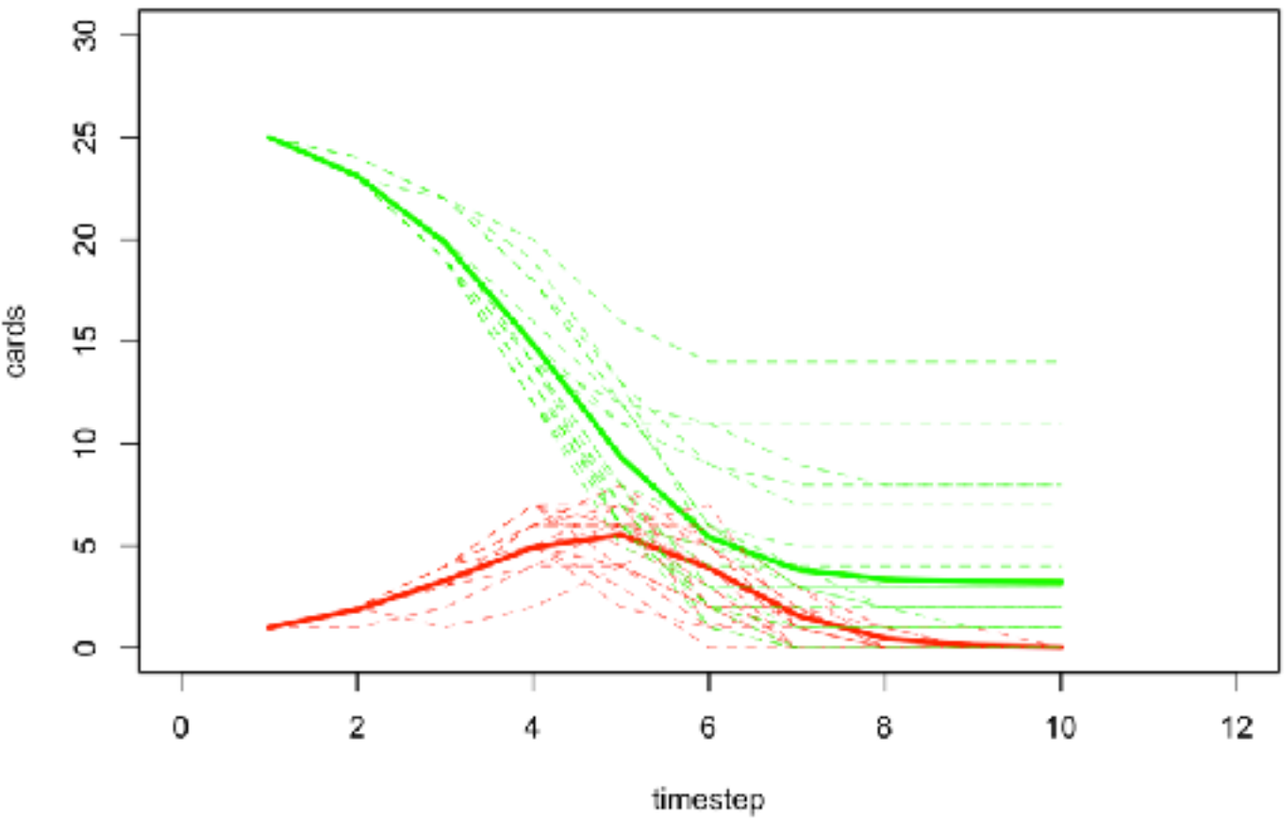


Id	Site	Date	Sex	Weight	Rainfall	gmam	skin
bat_12	Site_1	41586	F	650	140mm	NA	3.90
bat_13	Site_1	41586	F	70	140mm	NA	3.89
bat_14	Site_2	41593	M	710	136mm	NA	30.05
bat_15	Site_2	41593	M	690	136mm	NA	30.02
bat_16	Site_2	41393	F	390	136mm	3.88	NA
bat_17	Site_2	41393	M	125	136mm	NA	5.95
bat_18	Site_2	41593	M	140	136mm	NA	4.88
bat_19	Site_2	41593	F	530	136mm	4.03	NA
bat_20	SITE_2	41593	M	130	136mm	NA	30.63
bat_21	Site_2	41594	F	640	136mm	4.26	NA
bat_22	Site_2	41394	F	390	136mm	4.18	NA
bat_23	Site_2	41394	M	140	136mm	NA	30.02
bat_24	Site_2	41588	F	570	136mm	3.97	NA
bat_25	Site_2	41598	M	150	136mm	NA	30.00
bat_26	Site_2	41596	F	650	136mm	4.16	NA
bat_27	Site_2	41596	F	155	136mm	4.14	NA
bat_28	Site_2	41596	M	130	136mm	NA	30.34



More advanced R

Model fitting

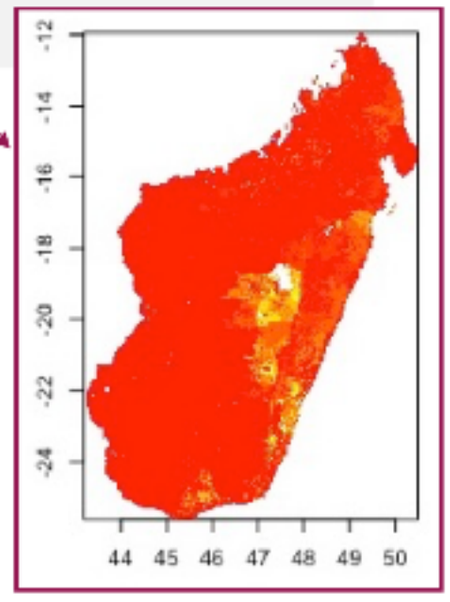


Model construction
Forward simulation in time
Optimization
Precise visualization

Spatial visualization

```
par(mfrow=c(1,3))
image(mdg_preg, col = blue2red(10))
image(log(mdg_preg+1), col = blue2red(10))
image(mdg_preg, zlim = c(0,10))
```

Plotting



Solving differential equations?

Useful math

Greek letters

Α α	Β β	Γ γ	Δ δ	Ε ε	Ζ ζ	Η η	Θ θ
alpha	beta	gamma	delta	epsilon	zeta	eta	theta
a	b	g	d	e	z	ē	th
[a/ɑ]	[b]	[g]	[d]	[e]	[zd/dz]	[ɛː]	[tʰ]
Ι ι	Κ κ	Λ λ	Μ μ	Ν ν	Ξ ξ	Ο ο	Π π
iota	kappa	lambda	mu	nu	xi	omikron	pi
i	k	l	m	n	ks/x	o	p
[i/ɪ]	[k]	[l]	[m]	[n]	[ks]	[o]	[p]
Ρ ρ	Σ σ/ς	Τ τ	Υ υ	Φ φ	Χ χ	Ψ ψ	Ω ω
rho	sigma	tau	upsilon	phi	chi	psi	omega
rh	s	t	y/y	ph	khkh	ps	ō
[r]	[s/z]	[t]	[y/ʏ]	[pʰ]	[kʰ]	[ps]	[ɔː]

Eigenvectors and eigenvalues

$$M = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 4 & 1 \\ 0 & -2 & 1 \end{bmatrix} \quad v_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \quad v_2 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \quad v_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

- $M \cdot v_1 = 3v_1$
- $M \cdot v_2 = 2v_2$
- $M \cdot v_3 = 1v_3$

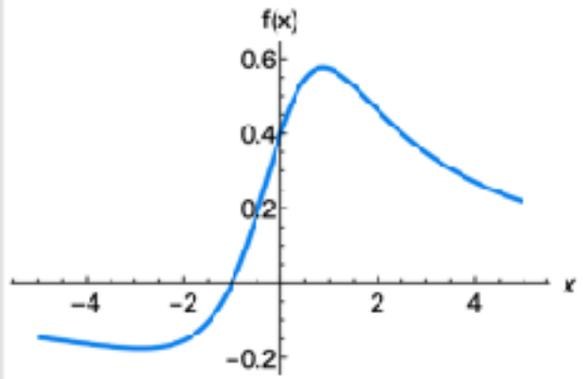
if $V = [v_1\ v_2\ v_3] = \begin{bmatrix} -1 & -2 & 1 \\ -1 & -1 & 0 \\ 1 & 2 & 0 \end{bmatrix}$

and $\Lambda = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ then $M = V^{-1}\Lambda V$

Function properties

- Intercept / root(s)
- Positive / negative value
- Maximum / minimum value
- Increasing / Decreasing / Constant
- Concave / Convexe
- Asymptotic

$$f(x) = \frac{x+1}{x^2+2.5}$$



GLMM Network

LM, GLM

Advanced R

Compartment model

Basic R

The new you

Research questions

Basic math

Steps in a modeling project

1. *Development of the study concept and question*
2. *Literature review*
3. *Data collection*
4. *Construction of model framework*
5. *Model analyses and selection*
6. *Model validation*
7. *Manuscript writing and submission*

You are now well equipped

Any aminareo ny baolina