E^2M^2: Ecological and Epidemiological Modeling in Madagascar

Data and Models

Centre ValBio

Ranomafana National Park, Madagascar

December 2022

Thanks to our sponsors!

Lecture contributions from:

Tanjona Ramiadantsoa

Steve Bellan

Goals for this lecture

- To explain what we’re doing here

Goals for this lecture

- To explain what we’re doing here
- To define “science”

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”
- To define “models”

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”
- To define “models”
- To introduce many different types of models

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”
- To define “models”
- To introduce many different types of models
  - Statistical

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”
- To define “models”
- To introduce many different types of models
  - Statistical
  - Mathematical

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”
- To define “models”
- To introduce many different types of models
  - Statistical
  - Mathematical
- To introduce the “E” in E^2M^2

Goals for this lecture

- To explain what we’re doing here
- To define “science”
- To define “data”
- To define “models”
- To introduce many different types of models
  - Statistical
  - Mathematical
- To introduce the “E” in E^2M^2
  - Ecology

Goals for this lecture

- To explain what we’re doing here
- To define ”science”
- To define ”data”
- To define “models”
- To introduce many different types of models
  - Statistical
  - Mathematical
- To introduce the “E” in E^2M^2
  - Ecology
  - Epidemiology

E2M2.org

Friday: R Bootcamp
- Intro to R Studio
- Exploring and Visualizing Data in R
- For-loops, Functions, and If-Else Statements

Saturday: Travel

Sunday: "Getting Started with the Basics"
- Data and Models
- Student introductions & presentations
- Linear regression & simple statistics
- Linear regression tutorial

Monday: "Using Models with Data"
- Model-Guided Study Design
- Study design tutorial
- Intro to Compartmental Models & Differential Equations
- Building mechanistic models in R
- Refining research questions for modeling
- Defining a model world

Tuesday: "Applying Simple Models"
- Dynamical Fever
- Introduction to Mixed Modeling
- Reading a Research Paper

Wednesday: "Refining Your Work"
- Model Fitting in Practice – the Basic Concept
- Epidemic Cards
- Model Fitting with Epidemic Cards
- Model Selection and Comparison
- Model Selection Tutorial
- Model Telephone

Thursday: "Bells and Whistle"
- Introduction to Phylogenetic Modeling • Tutorial
- **Intoduction to Spatial Visualization and Plotting** • Tutorial
- Modeling Insights from the Metapopulation Game • Tutorial
- Introduction to Network Modeling • Tutorial
- Research Snapshots

Friday: "Putting it All in Perspective"
- The Life Cycle of a Modeling Project
- C4C Student Presentations

Saturday: Travel

January: "Sharing Your Work"
- Final student presentations

- **Programming**
- **Data**
- **Models**
- **Research Development**

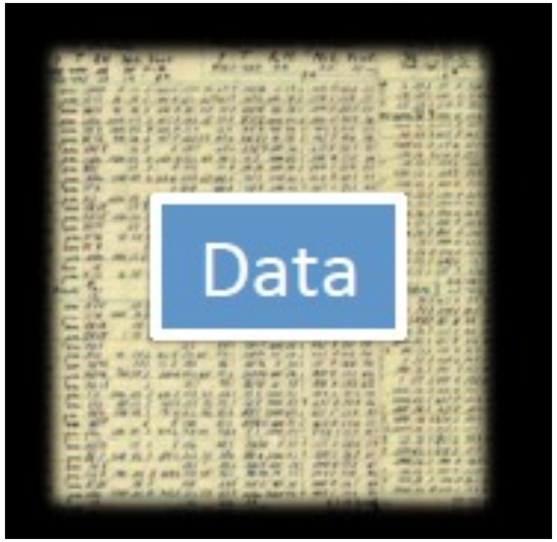
What is science?

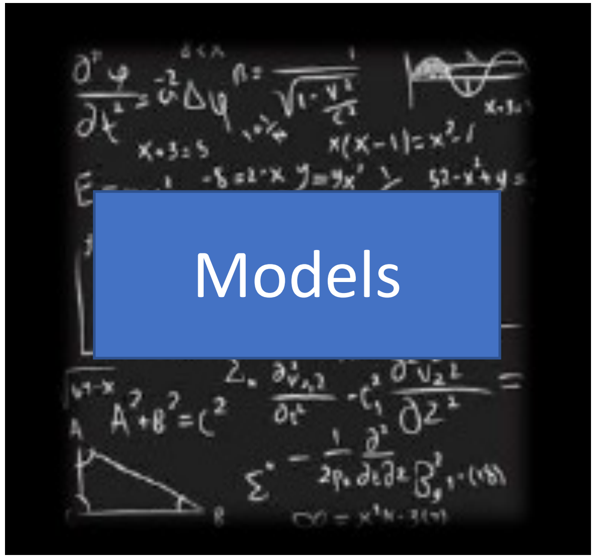
the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles based on these facts.

– *Academic Press Dictionary of Science & Technology*

Observations and Laws and Principles

Data and Models





Data and Models

- What is data?

Data and Models

- What is data?
  - Backbone of science

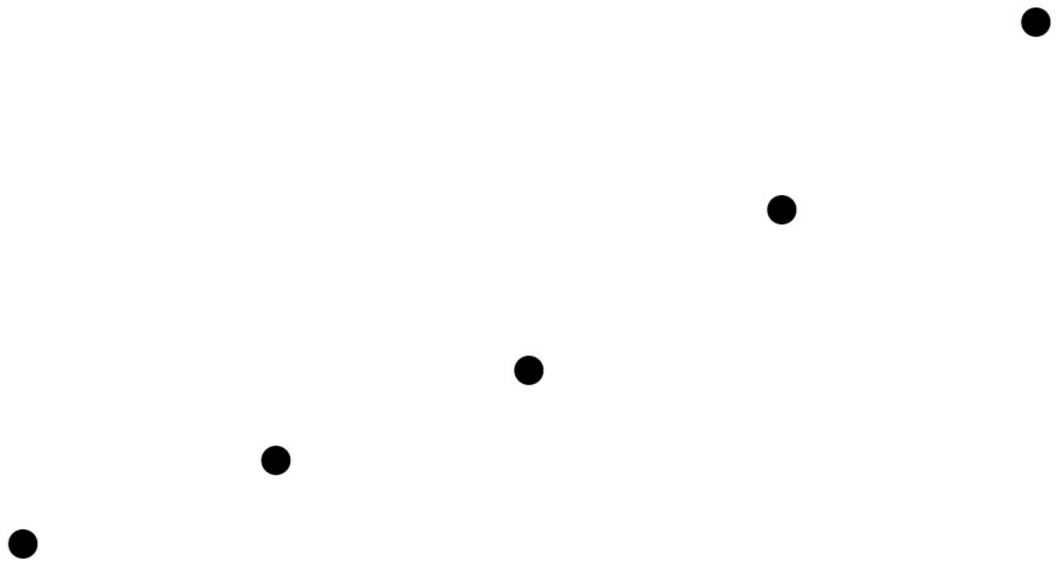
What is science?

the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles

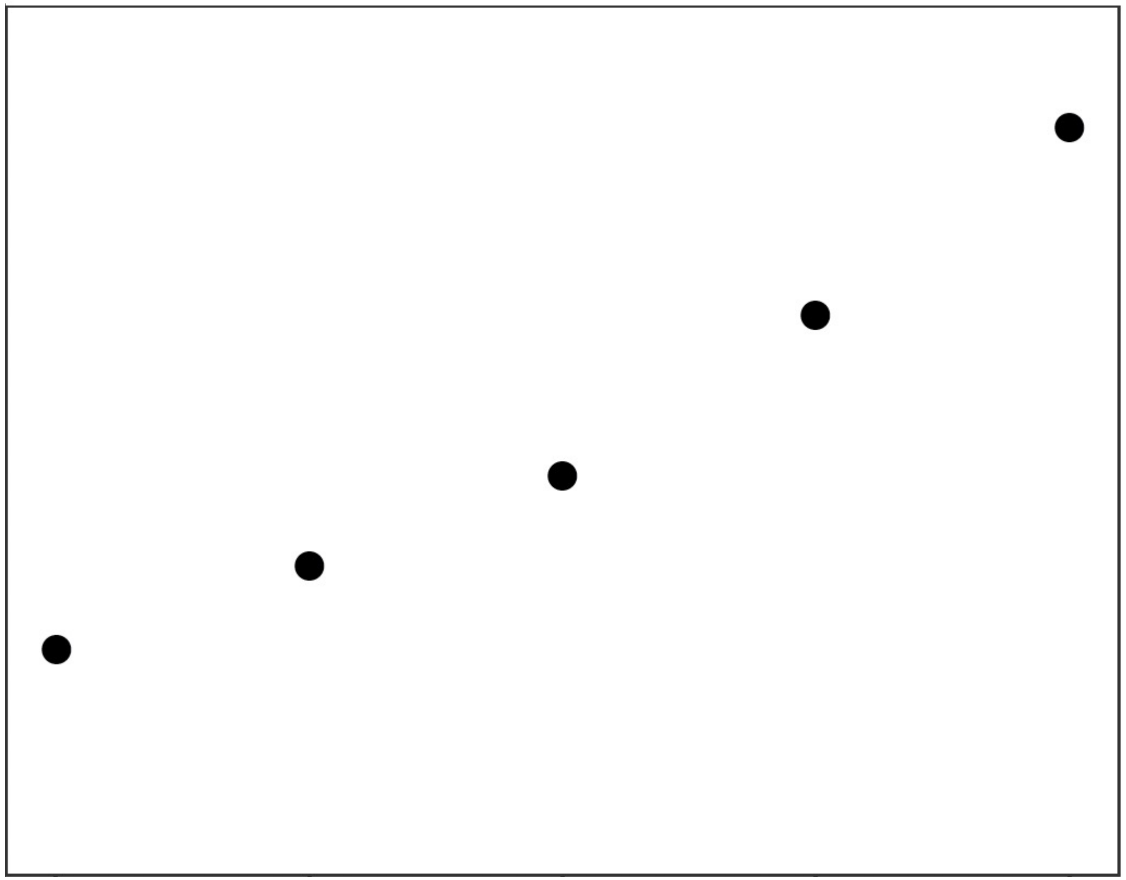
Data and Models

- What are data?
  - Evidence to support a claim

Are these data?

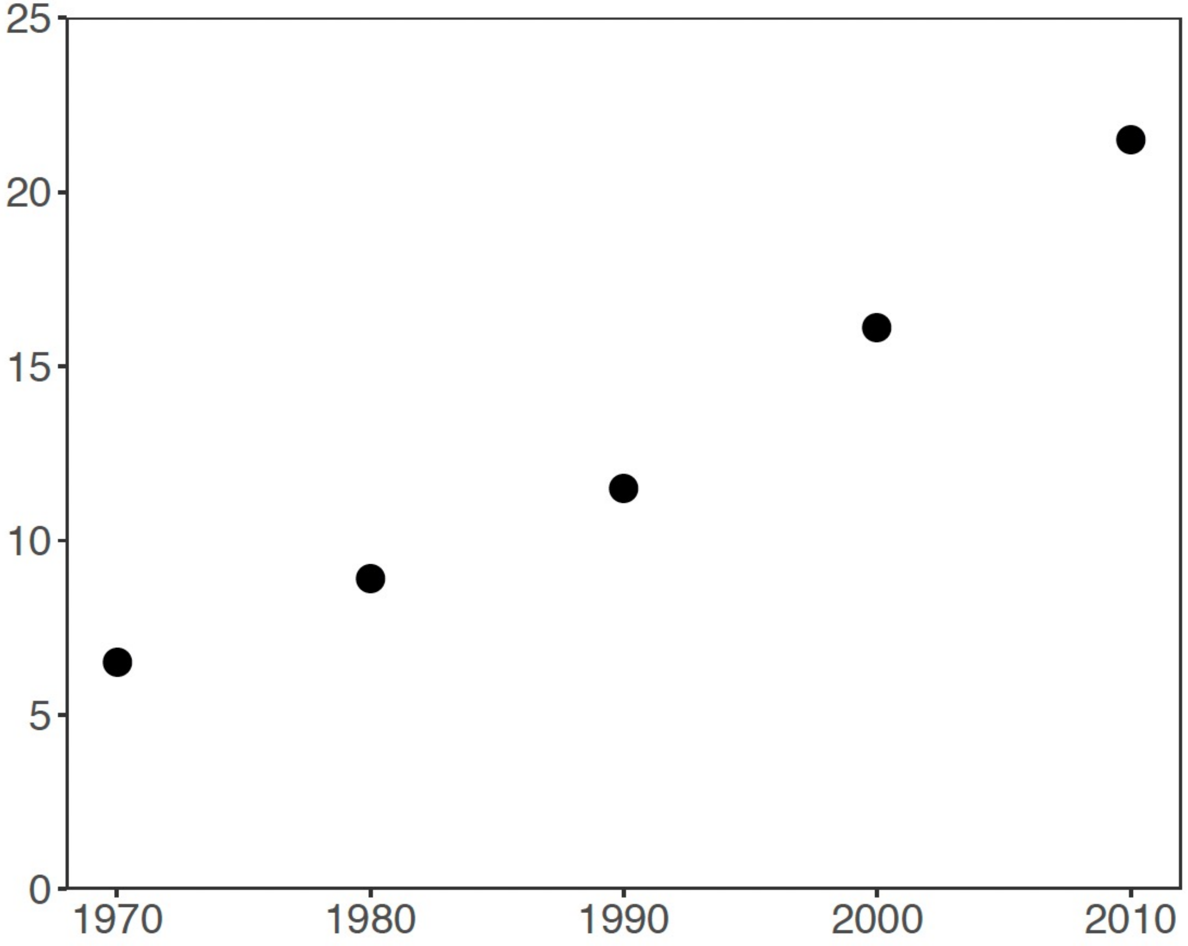


Are these data?

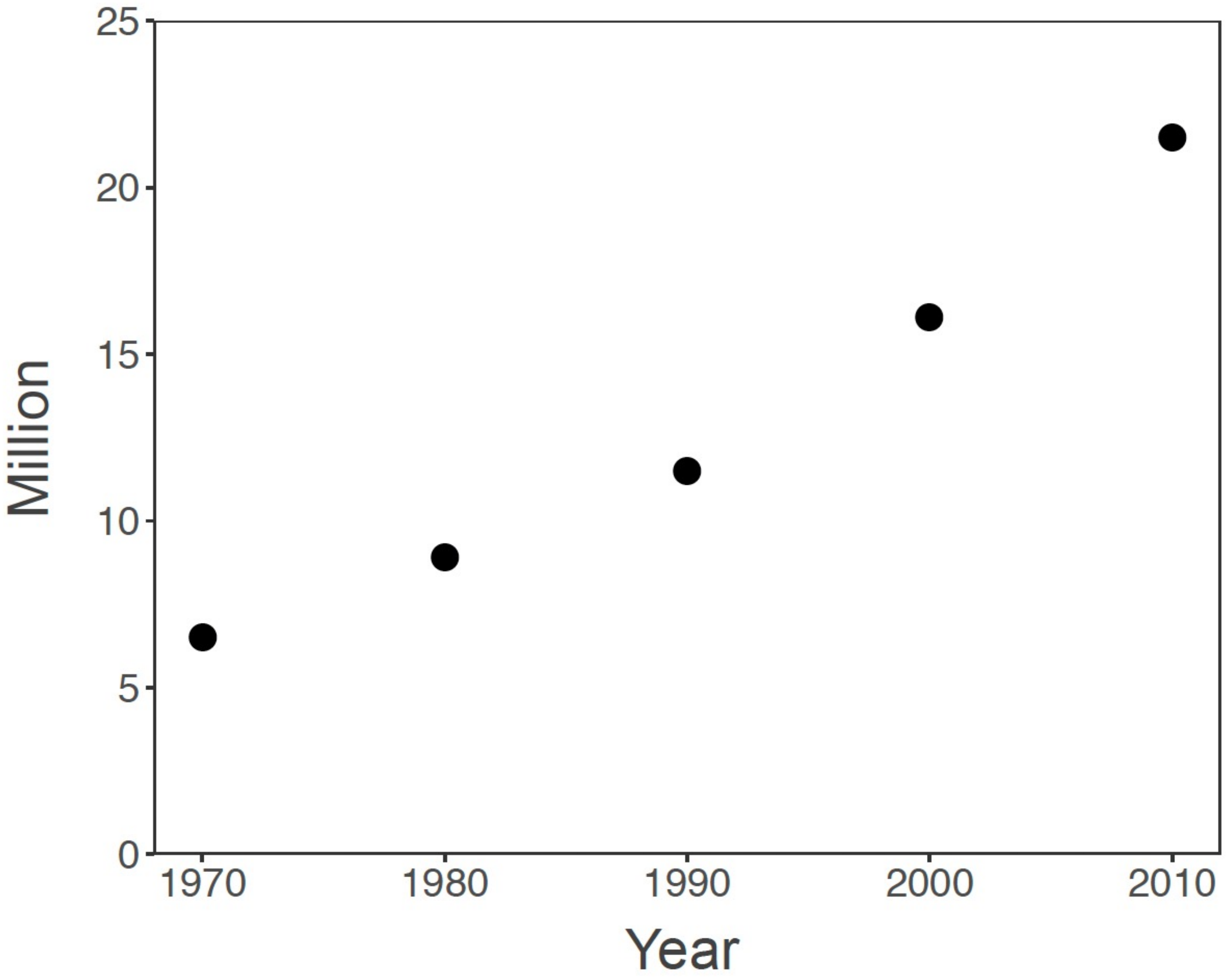


What do we need to make these data?

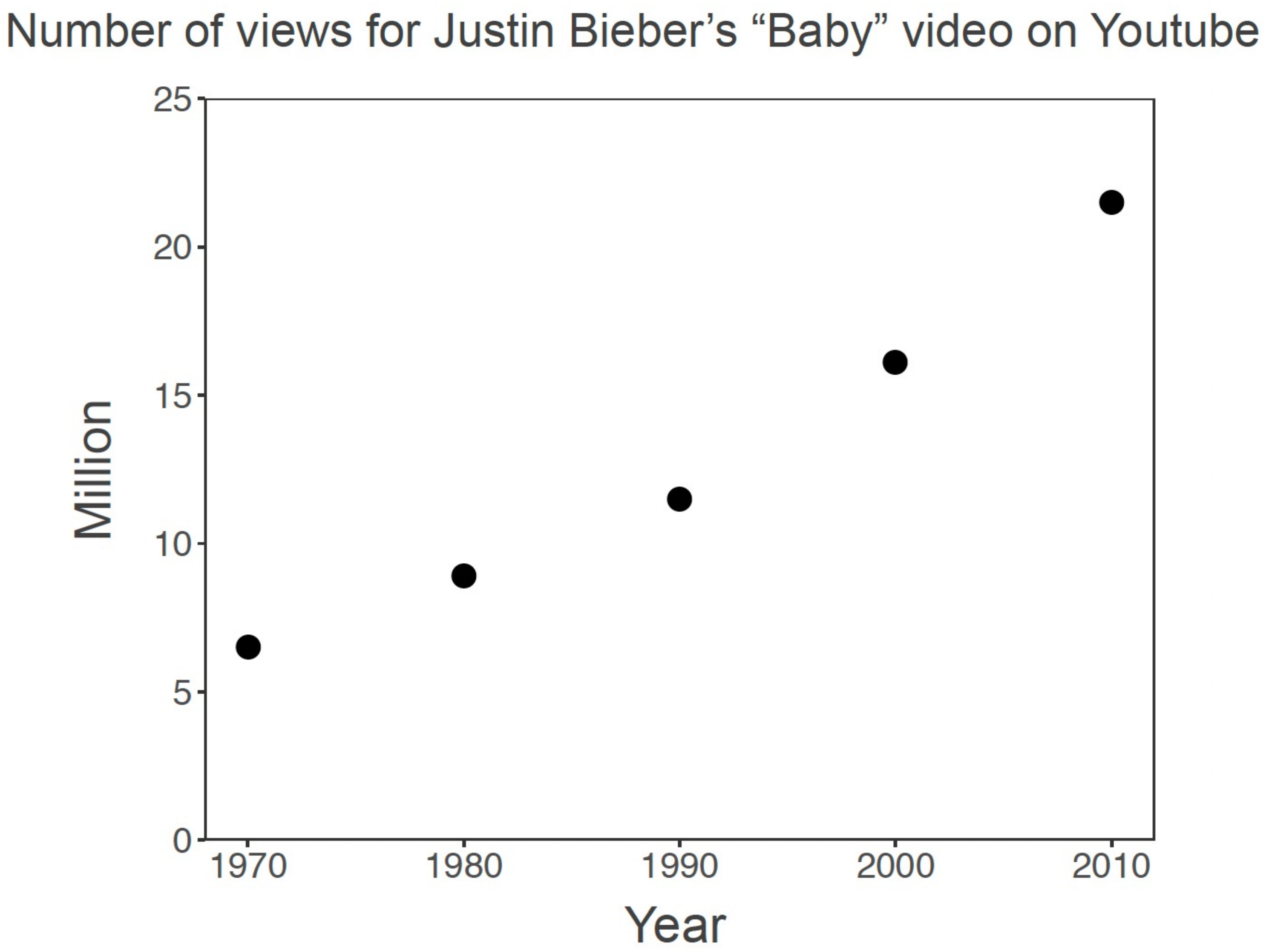
Are these data?



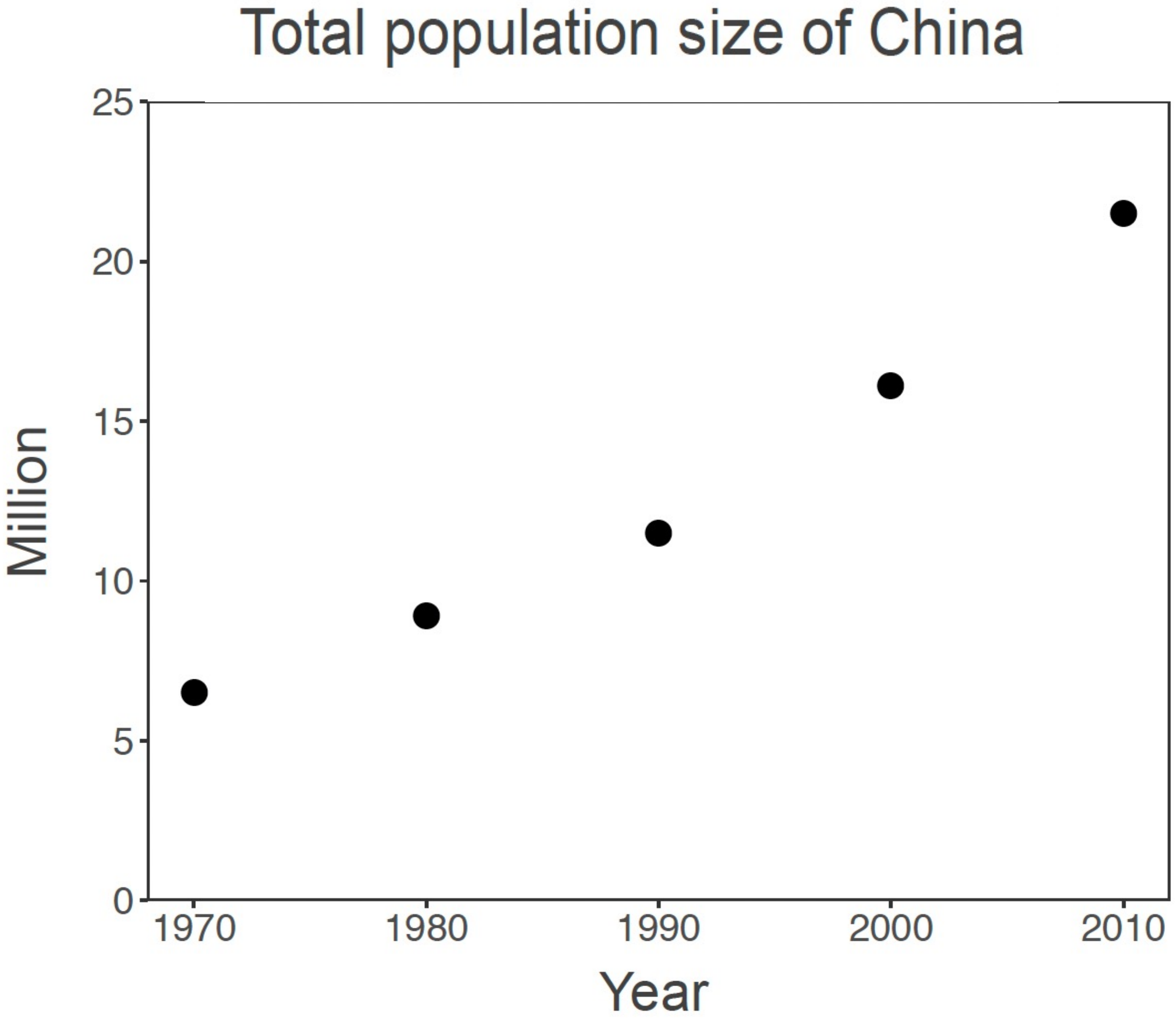
Are these data?



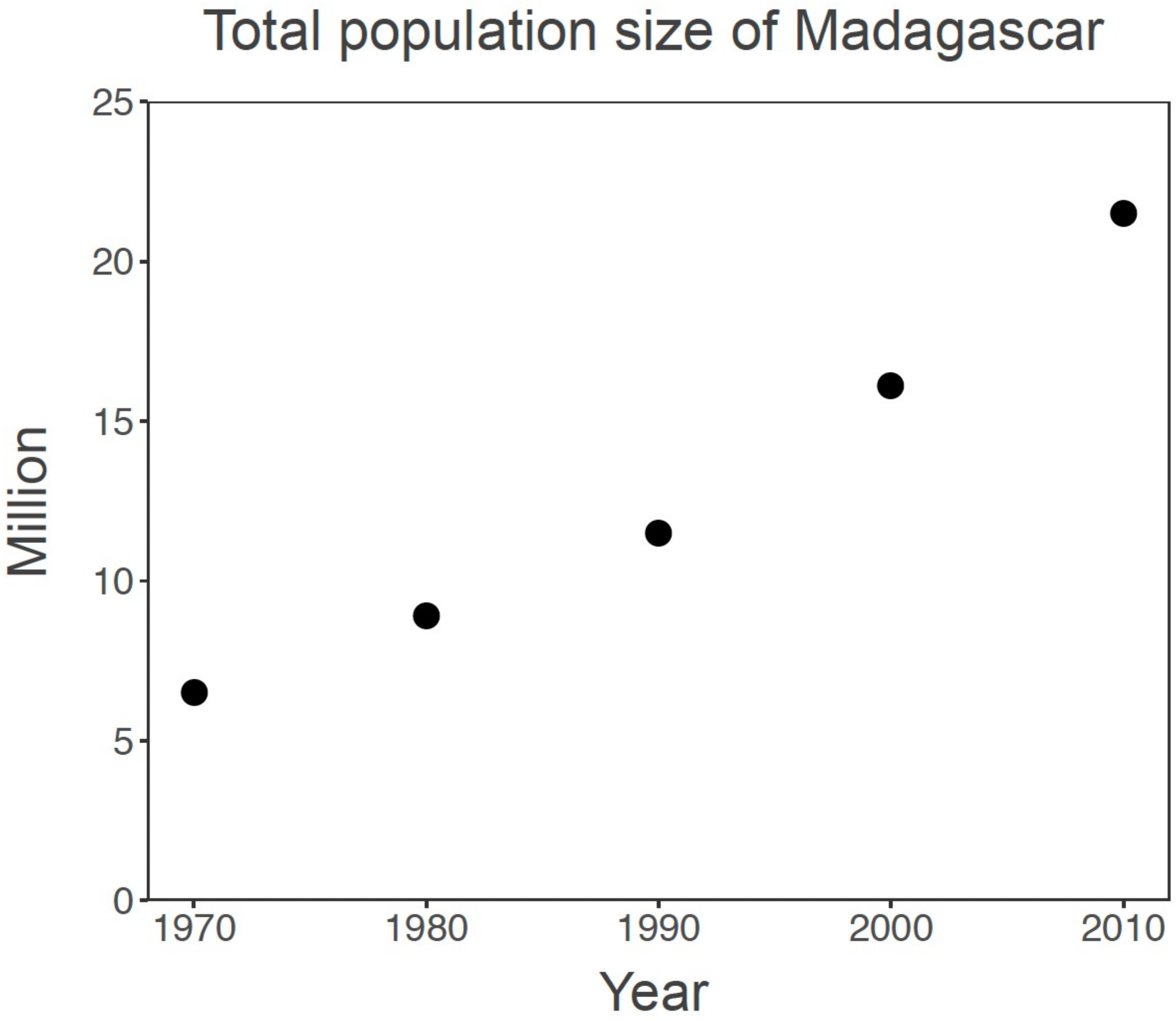
Are these data?



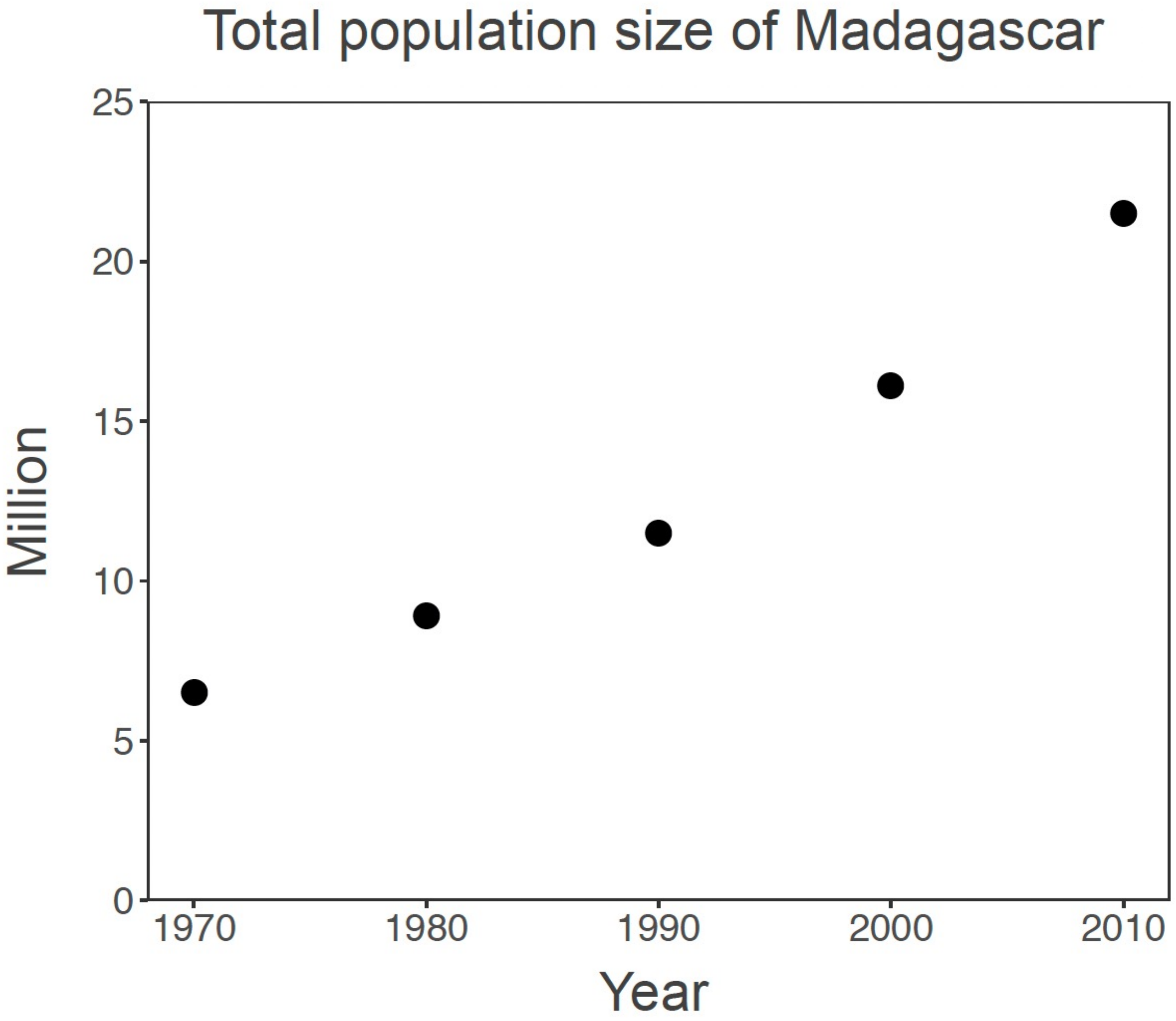
Are these data?



Are these data?



Are these data?



Source: World Bank (accessed 2017)

What are data?

- A relationship between at least two variables
  - x: explanatory, control, driver, independent variable(s)
  - y: response, dependent variable(s)

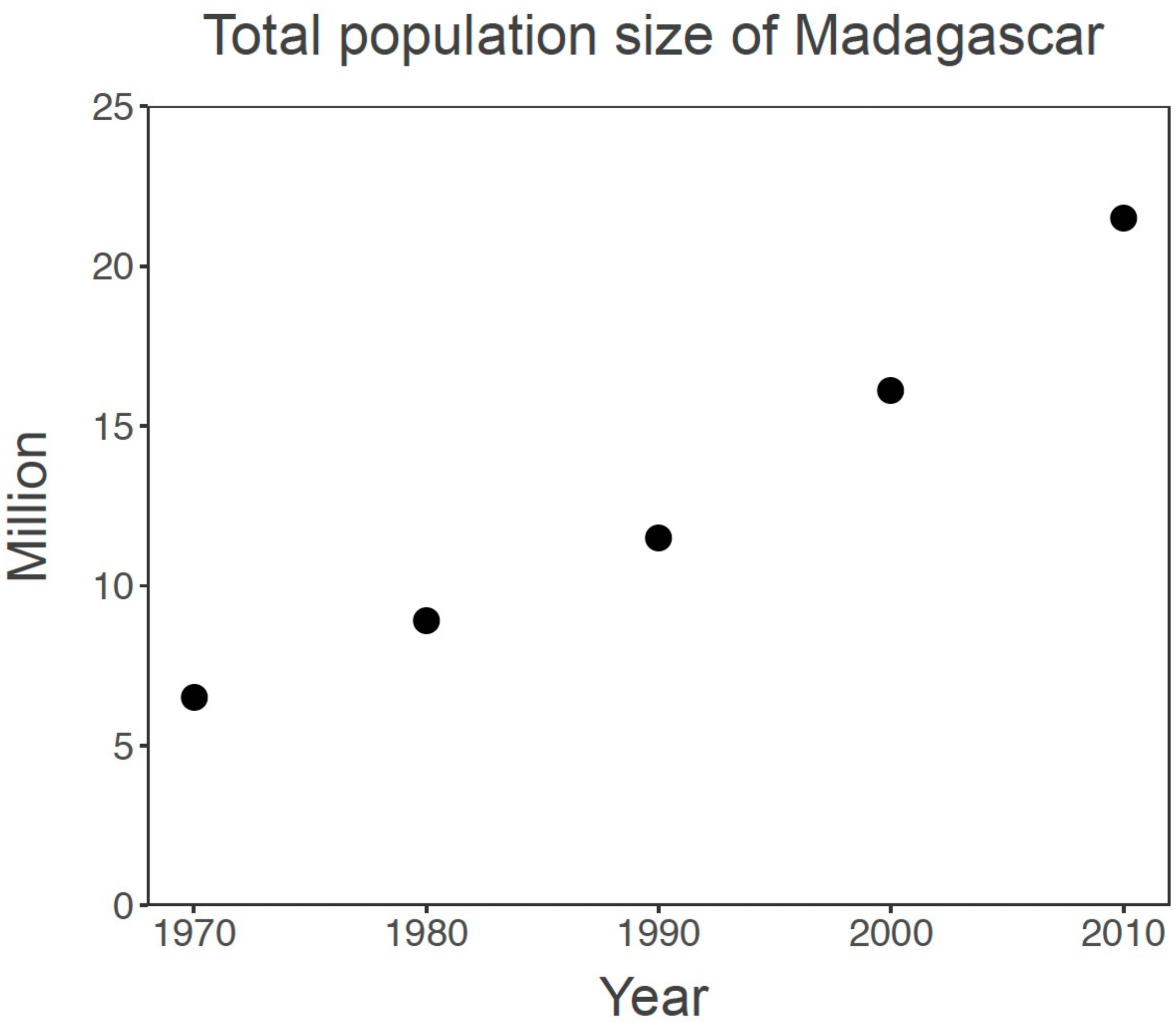
What are data?

- A relationship between at least two variables
  - x: explanatory, control, driver, independent variable(s)
  - y: response, dependent variable(s)
- x and y should be clearly defined
  - with respect to the **question!**

What are data?

- A relationship between at least two variables
  - x: explanatory, control, driver, independent variable(s)
  - y: response, dependent variable(s)
- x and y should be clearly defined
  - with respect to the **question!**
- Backbone of science
- **Evidence** to support a **claim**

Data provide evidence to support a claim.

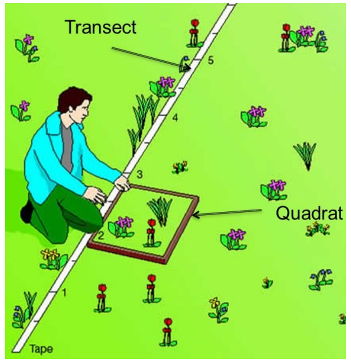


CLAIM: The population size of Madagascar has increased throughout the past 50 years

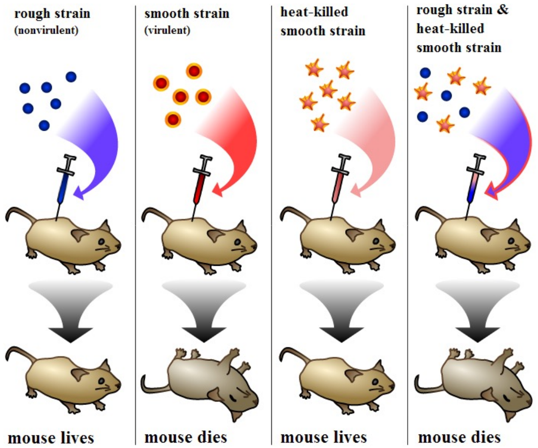
Source: World Bank (accessed 2017)

Data: Sources of x and y

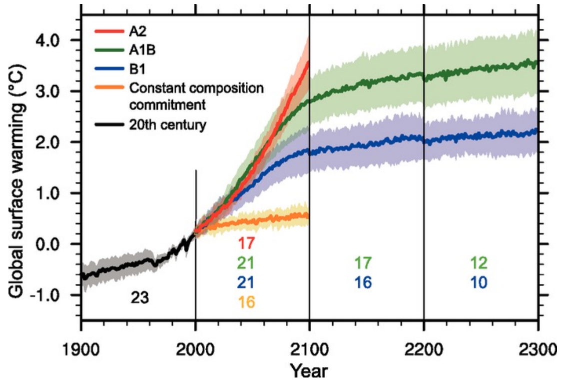
Observational



Experimental



Simulated



Empirical data

Data: Types

Numerical

Categorical

Data: Types

Numerical

- A variable is numerical when you can transform it with mathematical operation
- Examples?

Categorical

Data: Types

Numerical

- A variable is numerical when you can transform it with mathematical operation
- Examples:
- Integer, real number, multi-dimensional number

Categorical

Data: Types

Numerical

- A variable is numerical when you can transform it with mathematical operation
- Examples:
- Integer, real number, multi-dimensional number

Categorical

- A variable is categorical when it is not numerical but a categorical can be numerical?
- Examples?

Data: Types

Numerical

- A variable is numerical when you can transform it with mathematical operation
- Examples:
- Integer, real number, multi-dimensional number

Categorical

- A variable is categorical when it is not numerical but a categorical can be numerical?
- Examples:
- Colors, (blood) types, species name

Data: Things to consider

Data: Things to consider

- Data acquisition
  - Impossible, example?

Data: Things to consider

- Data acquisition
  - Impossible, example?
  - Theoretically possible but practically unfeasible, examples?

Data: Things to consider

- Data acquisition
  - Impossible, example?
  - Theoretically possible but practically unfeasible, examples?
- Data quality and quantity
  - In practice there is always a trade-off

Data: Things to consider

- Data acquisition
  - Impossible, example?
  - Theoretically possible but practically unfeasible, examples?
- Data quality and quantity
  - In practice there is always a trade-off
    - Example: monetary cost, human effort -> power analysis, sampling design etc.

Data: Things to consider

- Data acquisition
  - Impossible, example?
  - Theoretically possible but practically unfeasible, examples?
- Data quality and quantity
  - In practice there is always a trade-off
    - Example: monetary cost, human effort -> power analysis, sampling design etc.
- Reproducibility

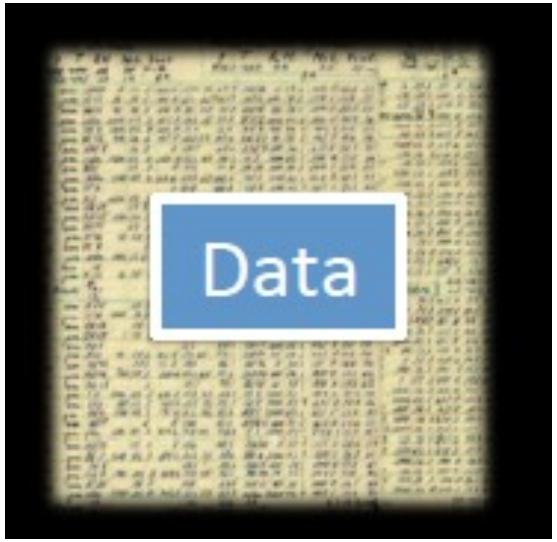
Data: Things to consider

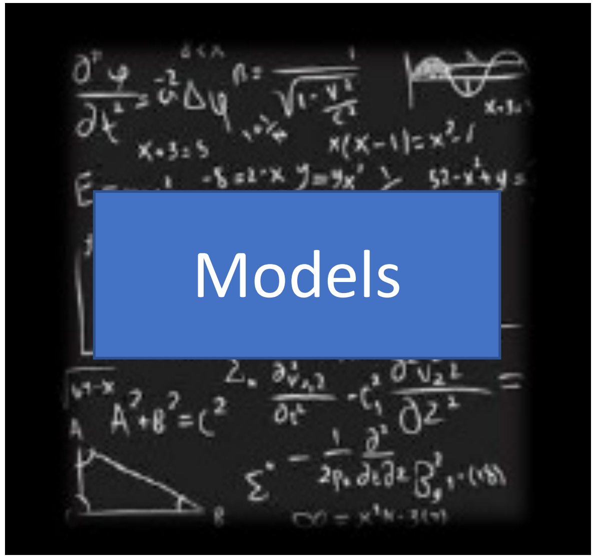
- Data acquisition
  - Impossible, example?
  - Theoretically possible but practically unfeasible, examples?
- Data quality and quantity
  - In practice there is always a trade-off
    - Example: monetary cost, human effort -> power analysis, sampling design etc.
- Reproducibility
- Measurement errors

Data: Things to consider

- Data acquisition
  - Impossible, example?
  - Theoretically possible but practically unfeasible, examples?
- Data quality and quantity
  - In practice there is always a trade-off
    - Example: monetary cost, human effort -> power analysis, sampling design etc.
- Reproducibility
- Measurement errors
  - Examples?

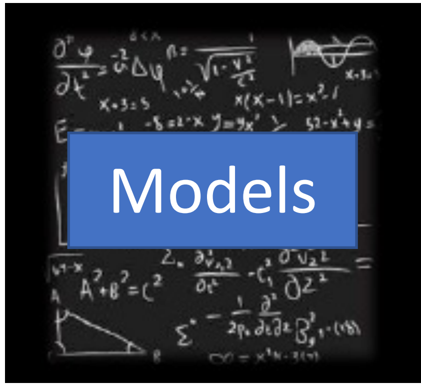
Data and Models





Data and Models

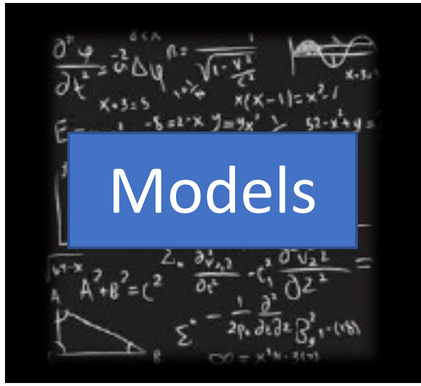
- What is a model?



What is science?

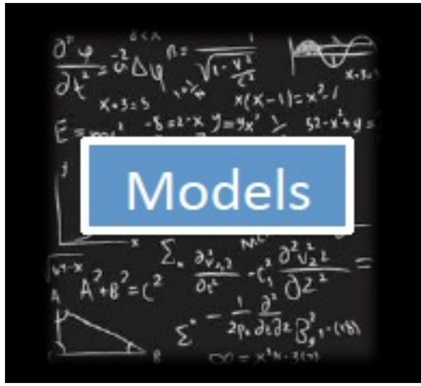
the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles

Laws and Principles



- A theory = a declaration to explain a phenomenon
  - Logical and falsifiable
- A model = an abstract representation of a phenomenon
- A hypothesis = a testable declaration that is derived from a theory

Theory, Models, Hypotheses

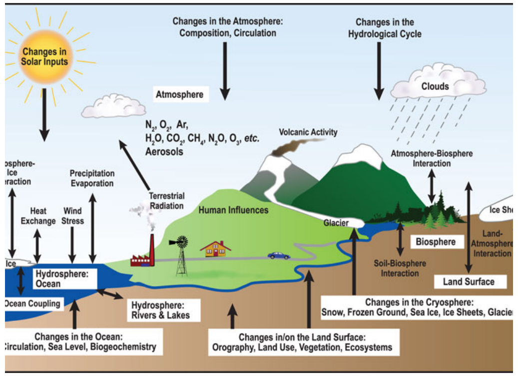


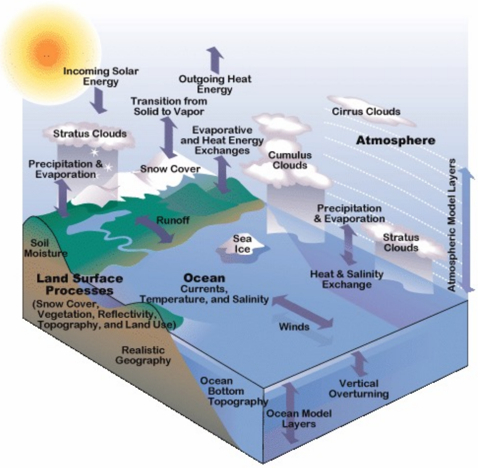
Theory | General

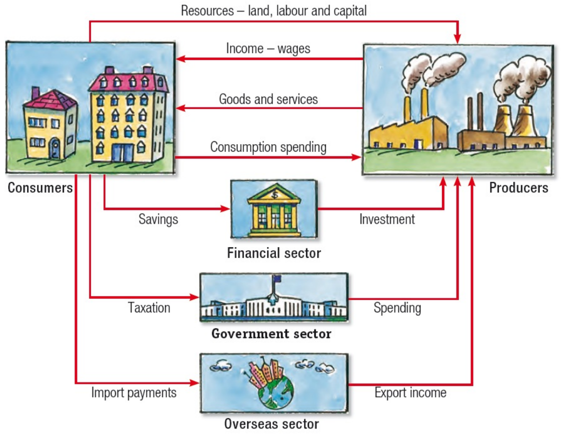
Model

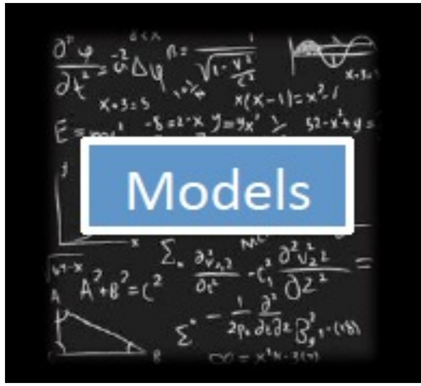
Hypothesis | Specific

Models

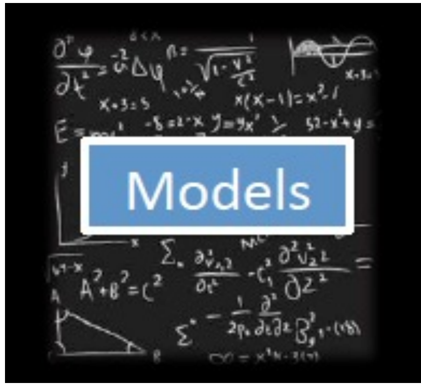

Changes in Solar Inputs
Changes in the Atmosphere: Composition, Circulation
Changes in the Hydrological Cycle
Atmosphere
Clouds
N_2, O_2, Ar, H_2O, CO_2, CH_4, N_2O, O_3, etc. Aerosols
Volcanic Activity
Atmosphere-Biosphere Interaction
Precipitation Evaporation
Terrestrial Radiation
Heat Exchange
Wind Stress
Human Influences
Glacier
Biosphere
Hydrosphere: Ocean
Soil-Biosphere Interaction
Land Surface
Ocean Coupling
Hydrosphere: Rivers & Lakes
Changes in the Cryosphere: Snow, Frozen Ground, Sea Ice, Ice Sheets, Glacie
Changes in the Ocean: Circulation, Sea Level, Biogeochemistry
Changes in/on the Land Surface: Orography, Land Use, Vegetation, Ecosystems


Incoming Solar Energy
Outgoing Heat Energy
Transition from Solid to Vapor
Cirrus Clouds
Stratus Clouds
Evaporative and Heat Energy Exchanges
Atmosphere
Cumulus Clouds
Precipitation & Evaporation
Snow Cover
Runoff
Precipitation & Evaporation
Sea Ice
Stratus Clouds
Soil Moisture
Atmospheric Model Layers
Land Surface Processes
(Snow Cover, Vegetation, Reflectivity, Topography, and Land Use)
Ocean
Currents, Temperature, and Salinity
Heat & Salinity Exchange
Winds
Realistic Geography
Ocean Bottom Topography
Vertical Overturning
Ocean Model Layers


Resources – land, labour and capital
Income – wages
Goods and services
Consumption spending
Consumers
Producers
Savings
Investment
Financial sector
Taxation
Spending
Government sector
Import payments
Export income
Overseas sector

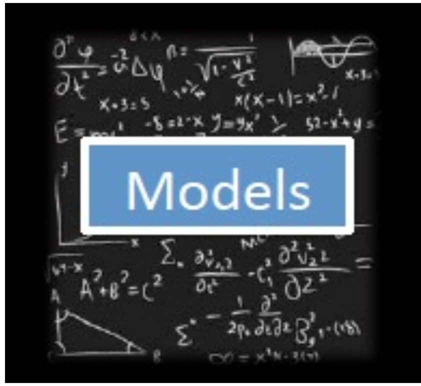


- When you make a model,
  you include the
  elements that you feel are most important
  to explain a phenomenon.

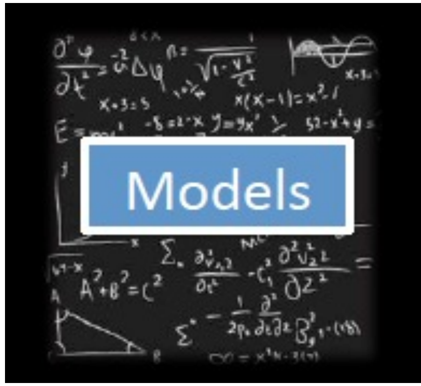


- When you make a model,
  you include the
  elements that you feel are most important
  to explain a phenomenon.
- Generally, we try to make models
  that can reproduce real-world data

- When you make a model,
  you include the
  elements that you feel are most important
  to explain a phenomenon.
- Generally, we try to make models
  that can reproduce real-world data
- In E^2M^2, we distinguish between
  statistical and mechanistic models

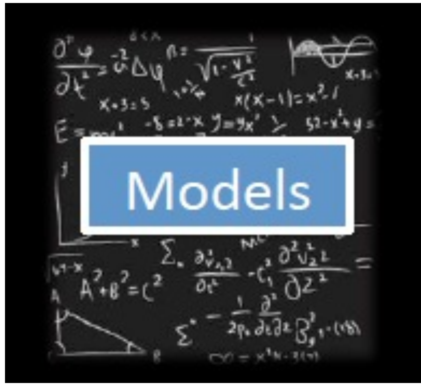


Statistical vs. Mathematical Model



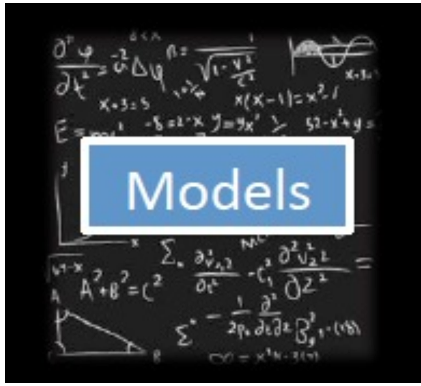
The choice depends on the research question!

Statistical Models



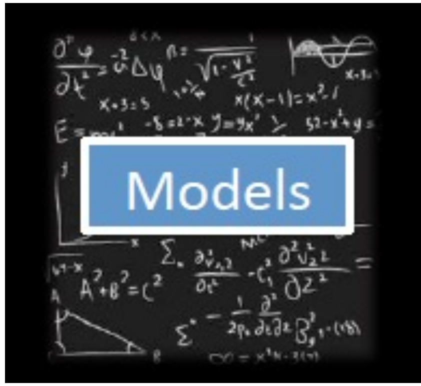
- Goal: To rigorously assess the strength of relationship between x and y
  - Find a significant relationship using a p-value as a measure of relationship strength
  - **Statistical models can demonstrate correlations.**

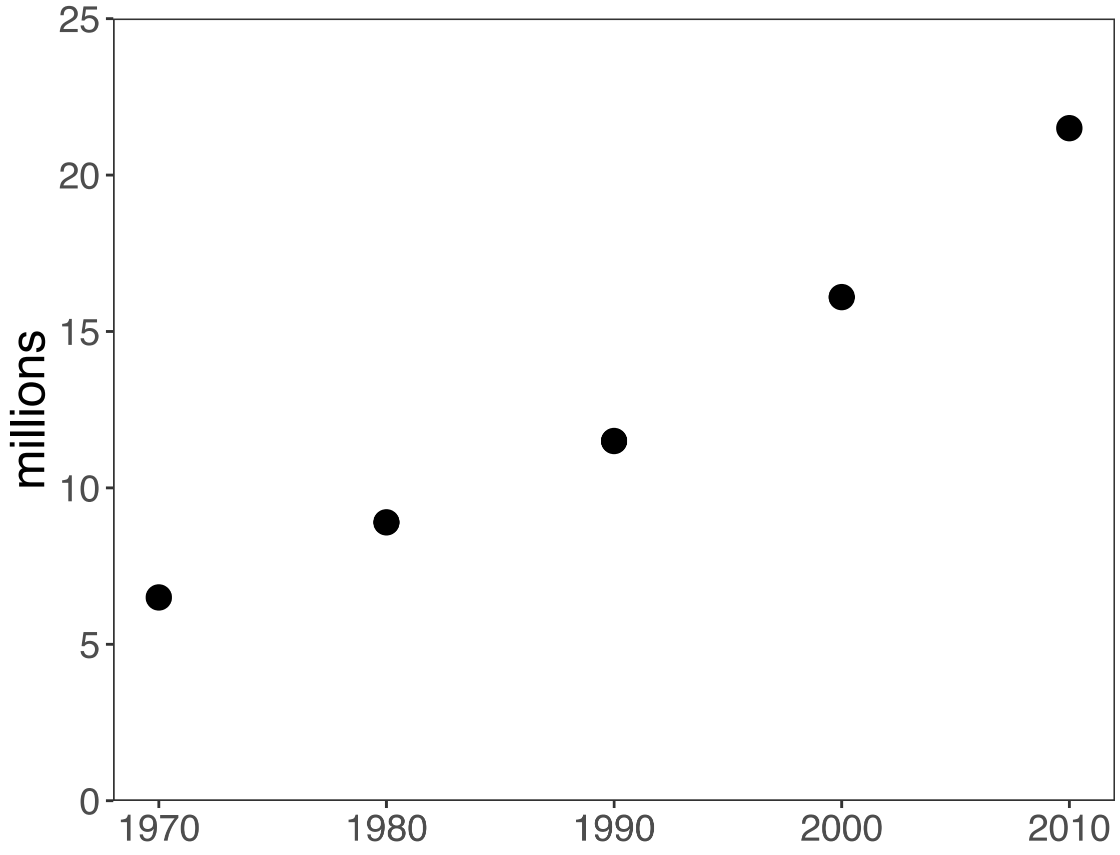
Statistical Models



- Goal: To rigorously **assess** the strength of relationship between x and y (describe patterns)
  - Find a significant relationship using a p-value as a measure of relationship strength
  - **Statistical models can demonstrate correlations.**
- Steps:
  1. Formulate a research question
  2. Formulate a hypothesis
  3. Develop a model to demonstrate your hypothesis.
  4. Collect **data (required!!!)**
  5. Evaluate hypothesis with appropriate statistical tools
     - t-test, Chi-square, ANOVA
     - Ordination (PCA)
     - Regression (LM, GLM, GLMM, GAM)

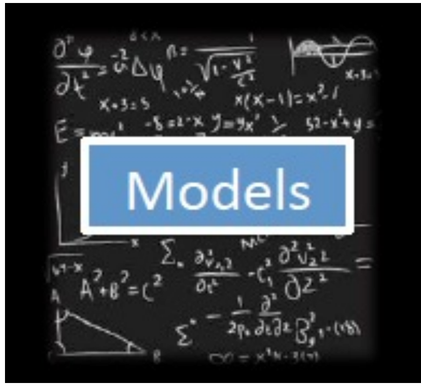
1. Example Question: What is the trajectory Malagasy population size through time?



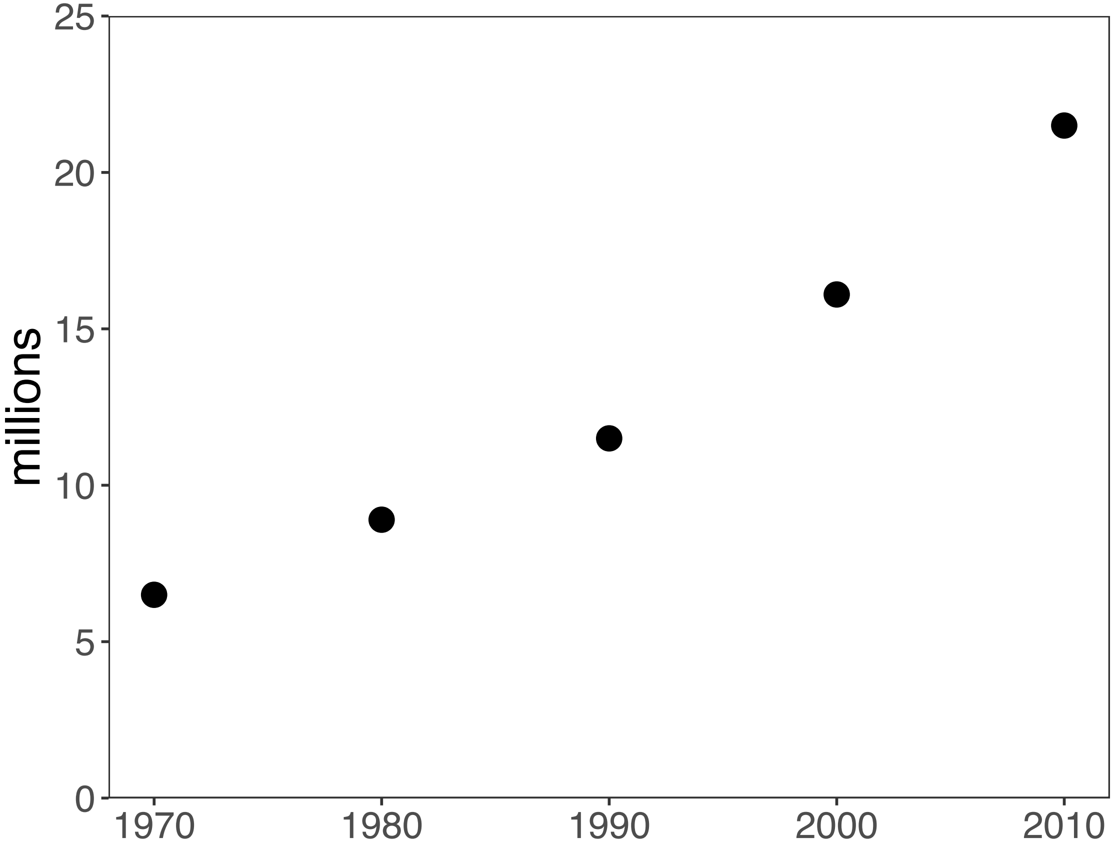


Source: World Bank

1. Example Question: What is the trajectory of Malagasy population size through time?

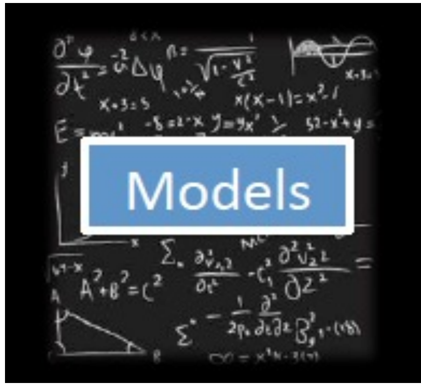


2. Hypothesis: Malagasy population size increases with time



Source: World Bank

1. Example Question: What is the trajectory of Malagasy population size through time?

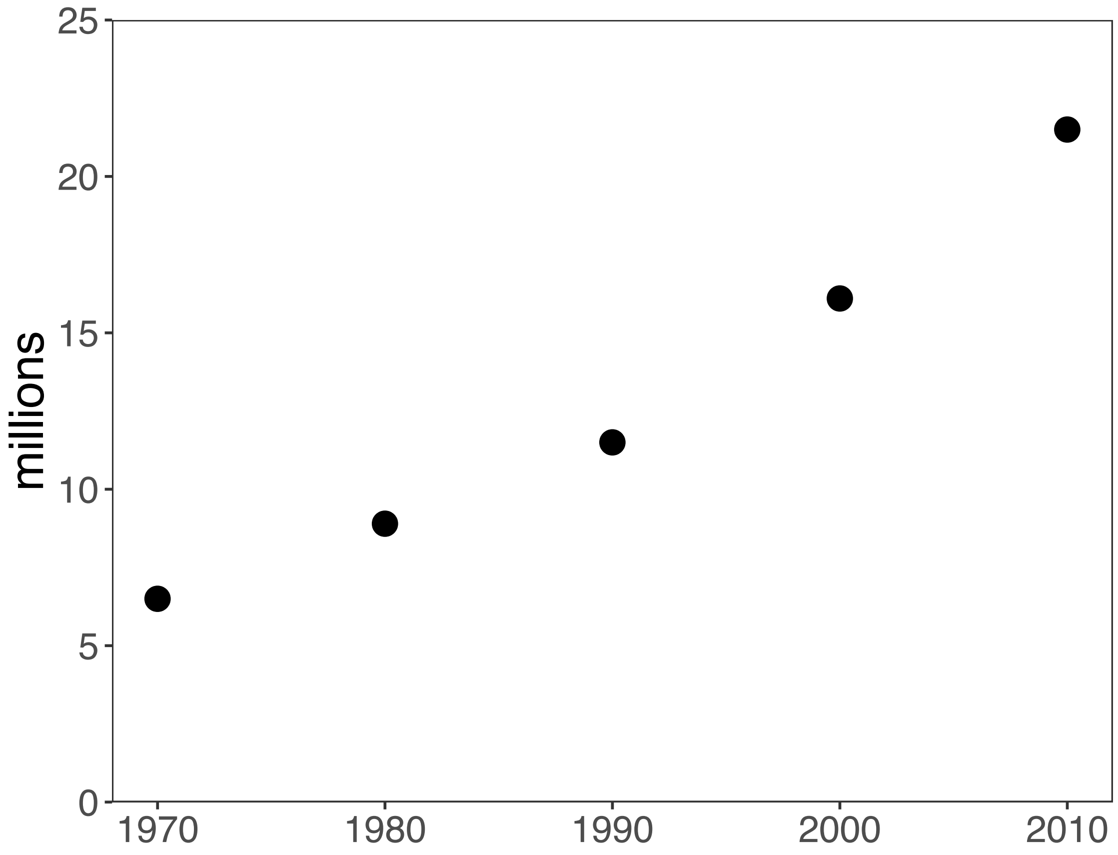


2. Hypothesis: Malagasy population size increases with time

3. Statistical Model:

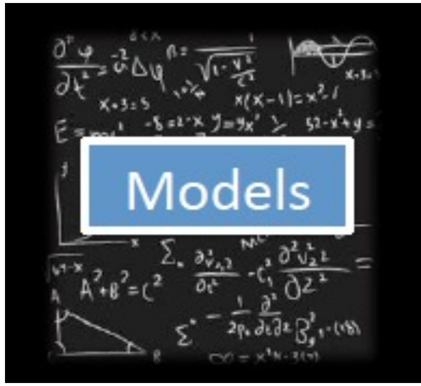
$$y = mx + b$$

Linear Regression



Source: World Bank

1. Example Question: What is the trajectory of Malagasy population size through time?



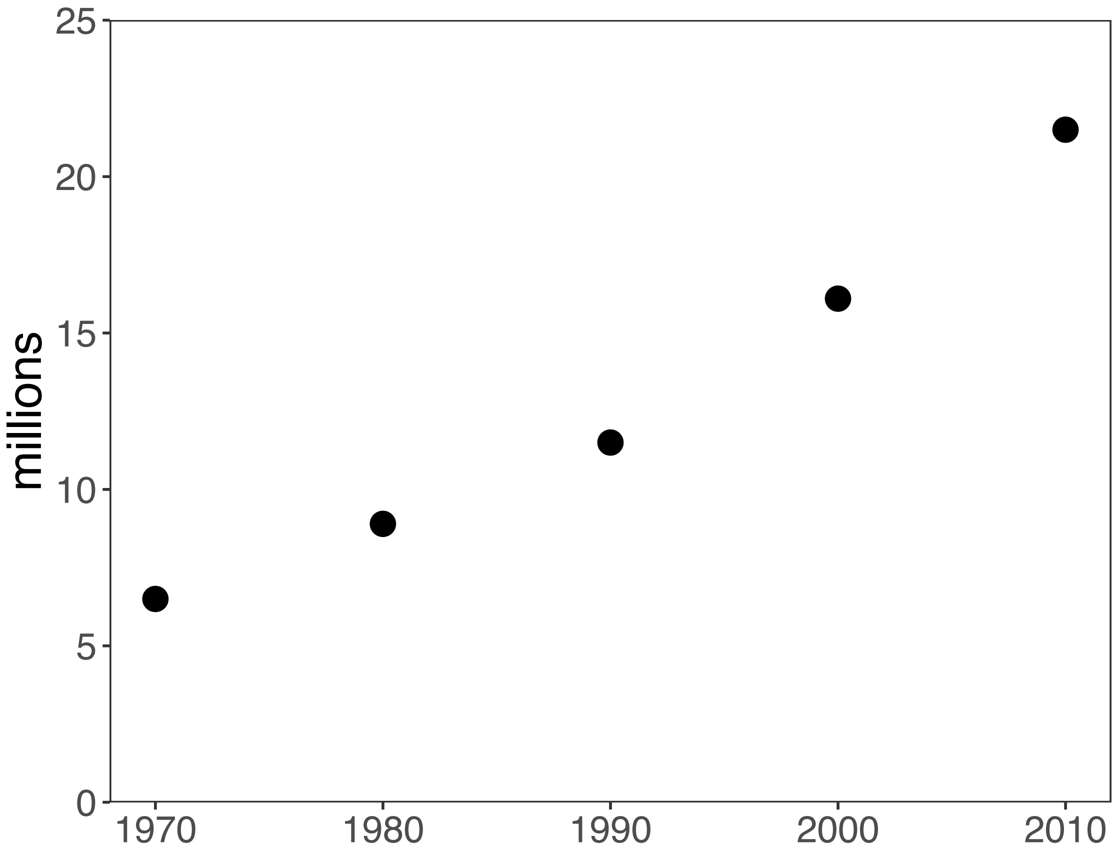
2. Hypothesis: Malagasy population size increases with time

3. Statistical Model:

$$y = mx + b$$

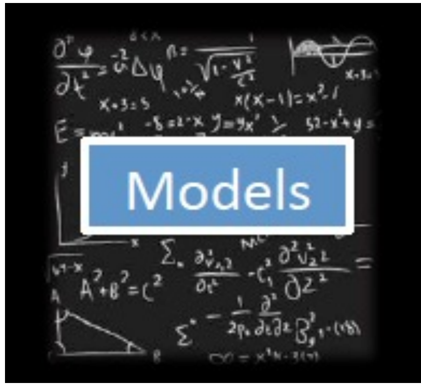
Linear Regression

4. Data:



Source: World Bank

1. Example Question: What is the trajectory of Malagasy population size through time?

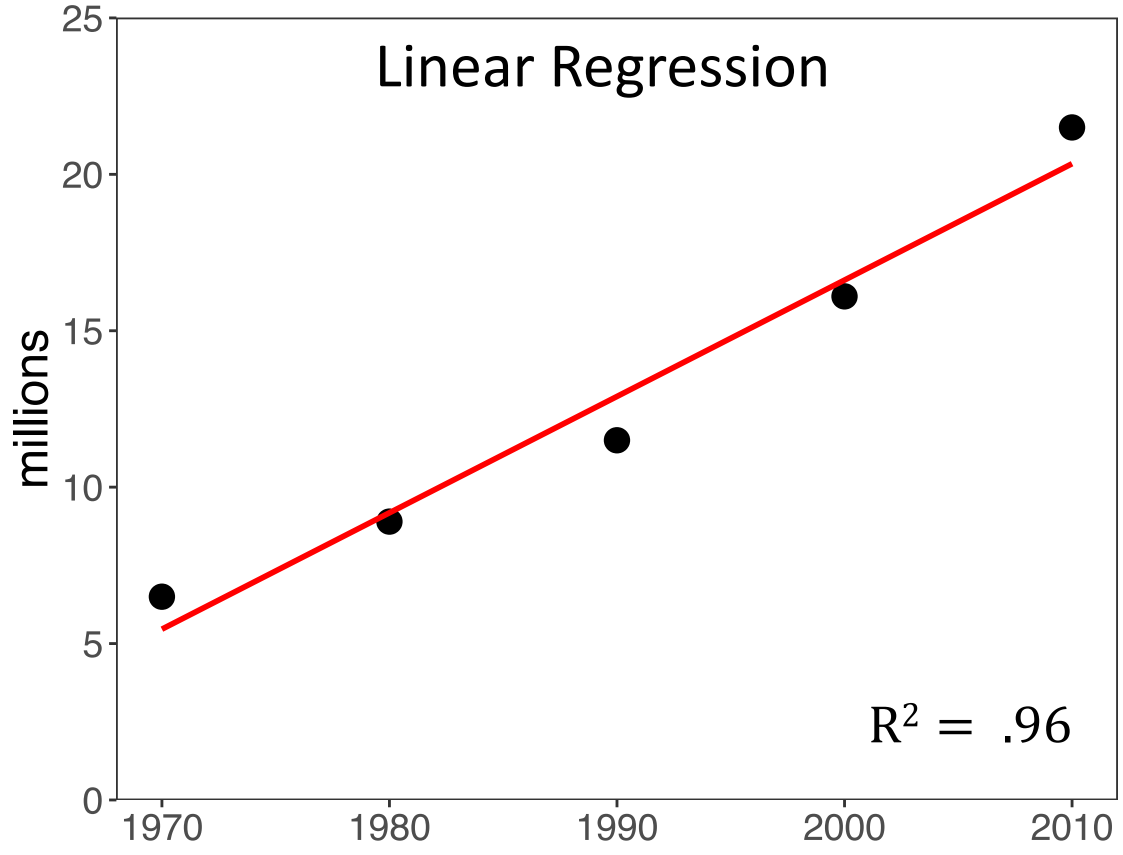


2. Hypothesis: Malagasy population size increases with time

3. Statistical Model:

$$y = mx + b$$

4. Data:



5. Evaluation

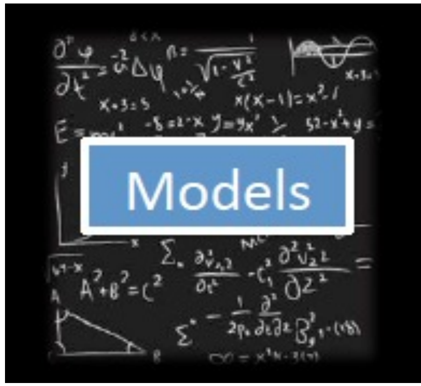
$$m = .372 \text{ million}$$

$$p = .003$$

What can we conclude from this fitted model?

Source: World Bank

1. Example Question: What is the trajectory of Malagasy population size through time?



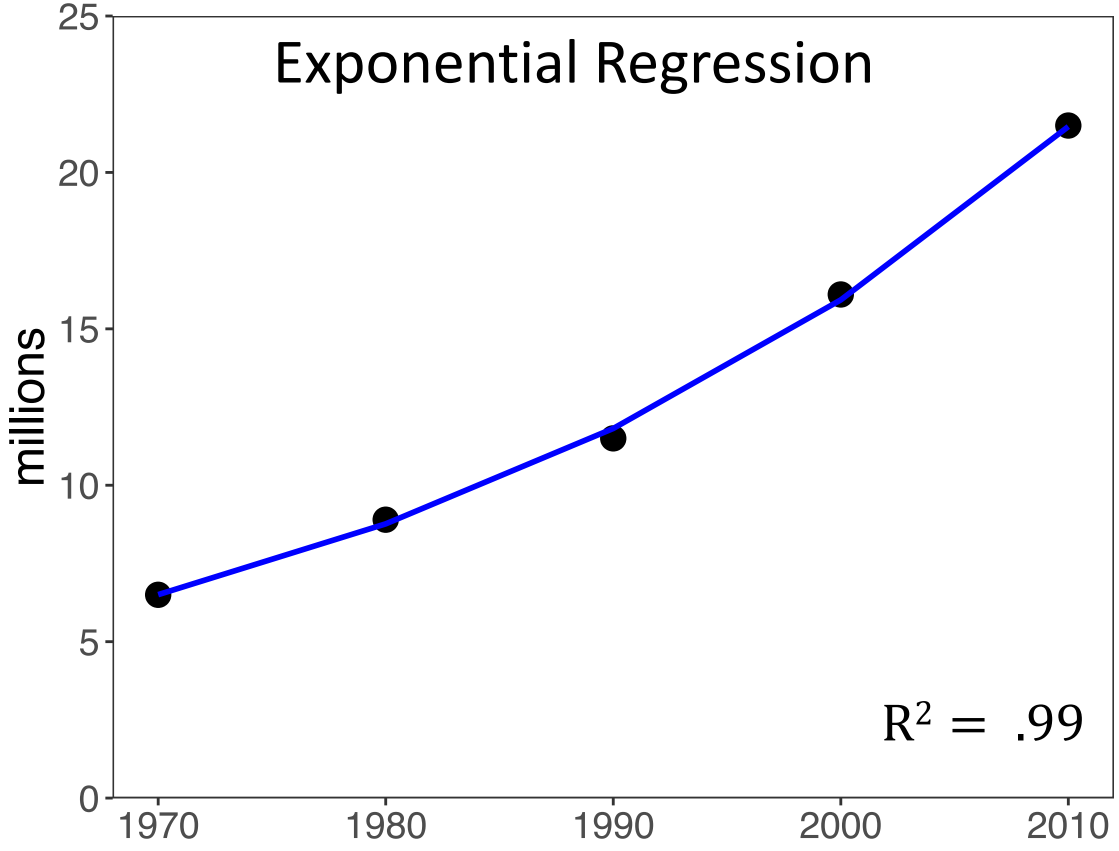
7. Adapt your model and re-evaluate:

$$y = e^{mx+b}$$

Exponential Regression

$m = 0.029$ mil.

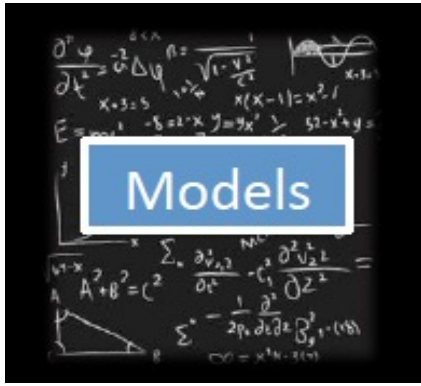
$p < .001$



What can we conclude from this fitted model?

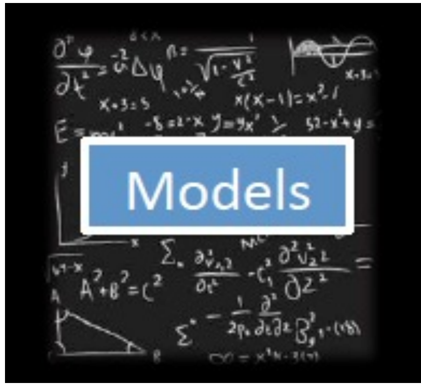
Source: World Bank

Statistical Models: Beware!



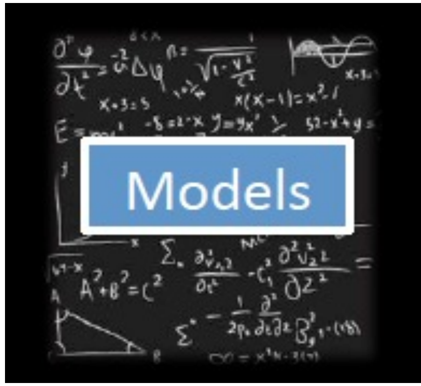
- Statistical models and tests are based on specific assumptions
  - data normally distributed
  - y and y independent
  - etc.

Statistical Models: Beware!



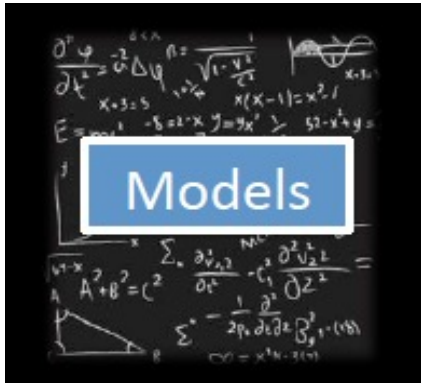
- Statistical models and tests are based on specific assumptions
    - data normally distributed
    - y and y independent
    - etc.
- Assessing a model means you need to make sure the assumptions are not violated.

Statistical Models: Beware!



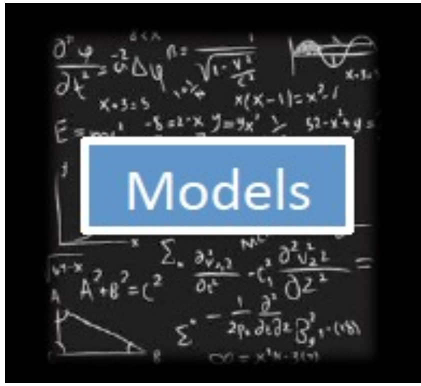
- Statistical models and tests are based on specific assumptions
  - data normally distributed
  - y and y independent
  - etc.
- Assessing a model means you need to make sure the assumptions are not violated.
- There are so many statistical models…

Statistical vs. Mathematical Model



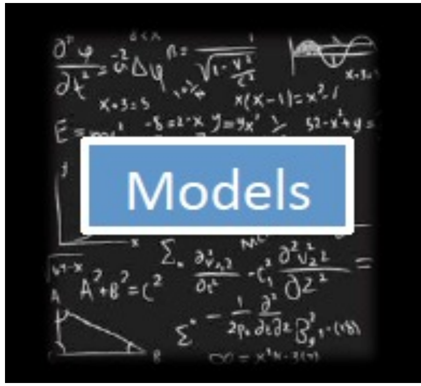
The choice depends on the research question!

Mechanistic Models



- Goal: To **demonstrate the processes** that underlie a relationship between x and y
  - Find a significant relationship using a p-value as a measure of relationship strength
  - **Mechanistic models can demonstrate causation.**
- Steps:
  1. Formulate a research question
  2. Formulate a hypothesis
  3. Develop a model to demonstrate your hypothesis.
  4. Collect **data** (for certain questions)
  5. Evaluate the extent to which your model-simulated data matches that from the real world.

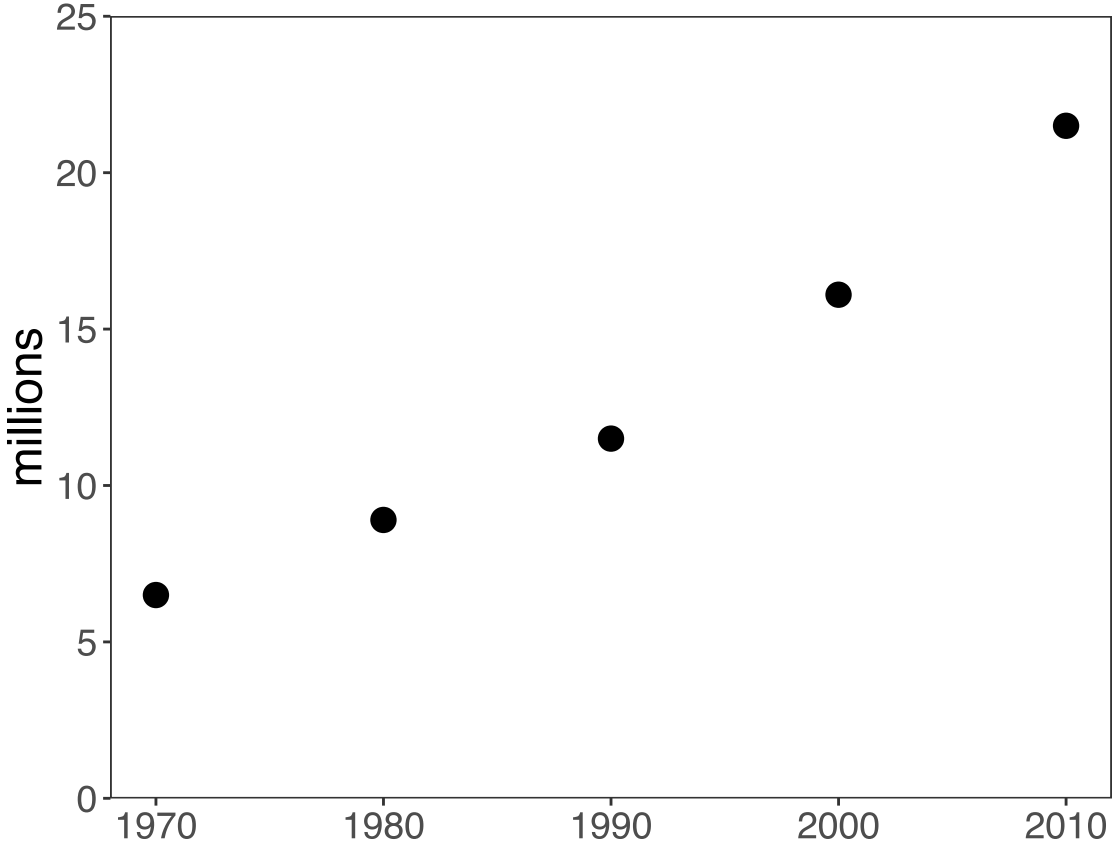
1. Example Question: How does Malagasy population size change with time?



2. Hypothesis: Malagasy population size increases because people are having children.

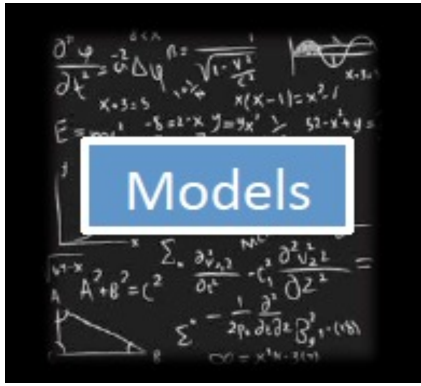
Can you think of an alternative hypothesis?

4. Data:



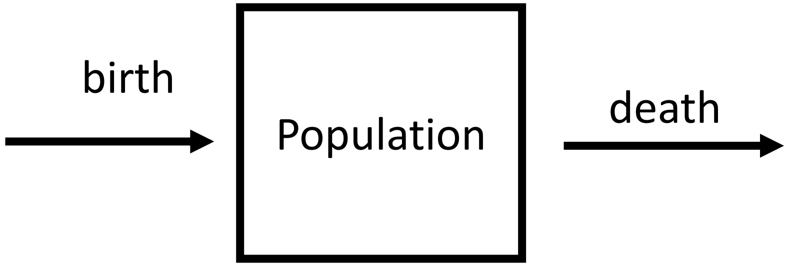
Source: World Bank

1. Example Question: How does Malagasy population size change with time?



2. Hypothesis: Malagasy population size increases because people are having children.

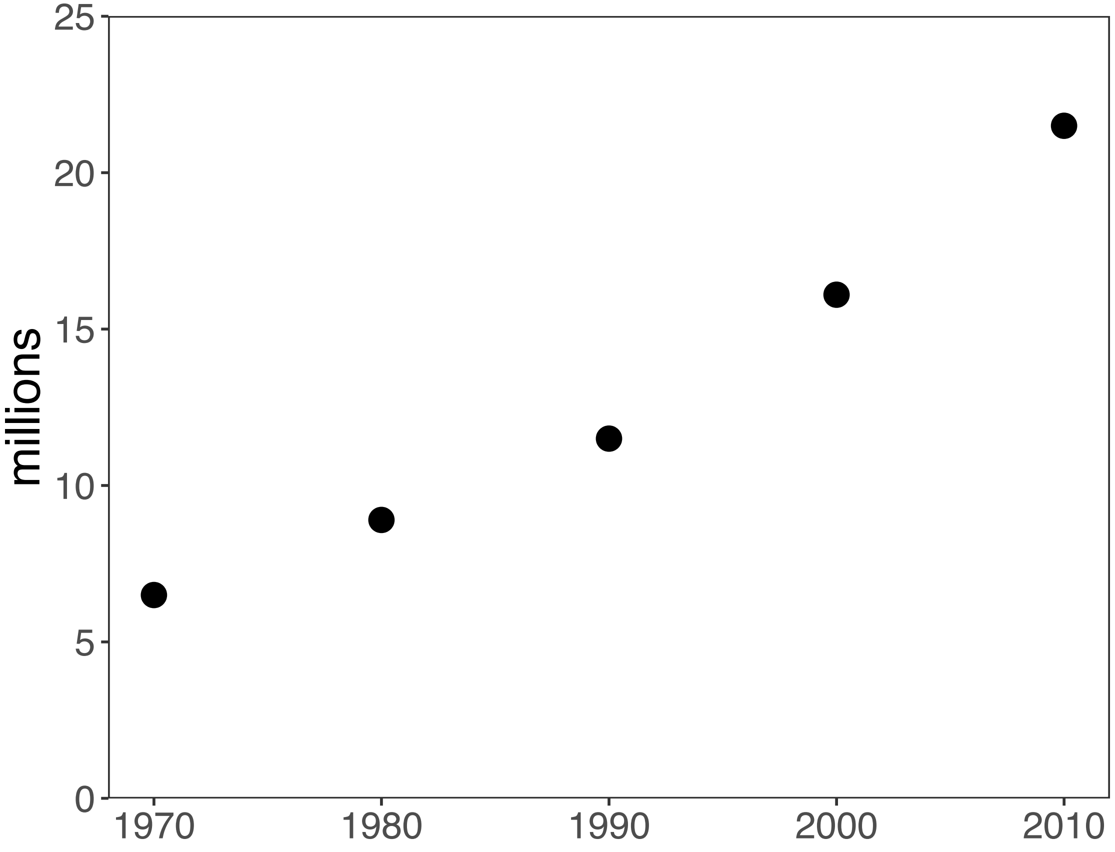
3. Mechanistic Model:



$$P_{t+1} = P_t + b * P_t \ - d * P_t$$

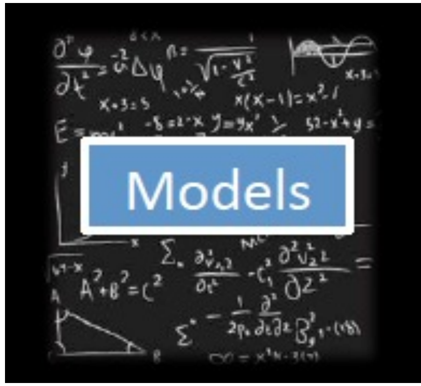
$$P_{t+1} = P_t + r * P_t$$

4. Data:



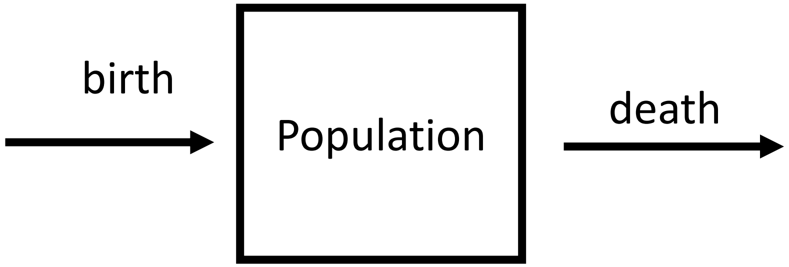
Source: World Bank

1. Example Question: How does Malagasy population size change with time?

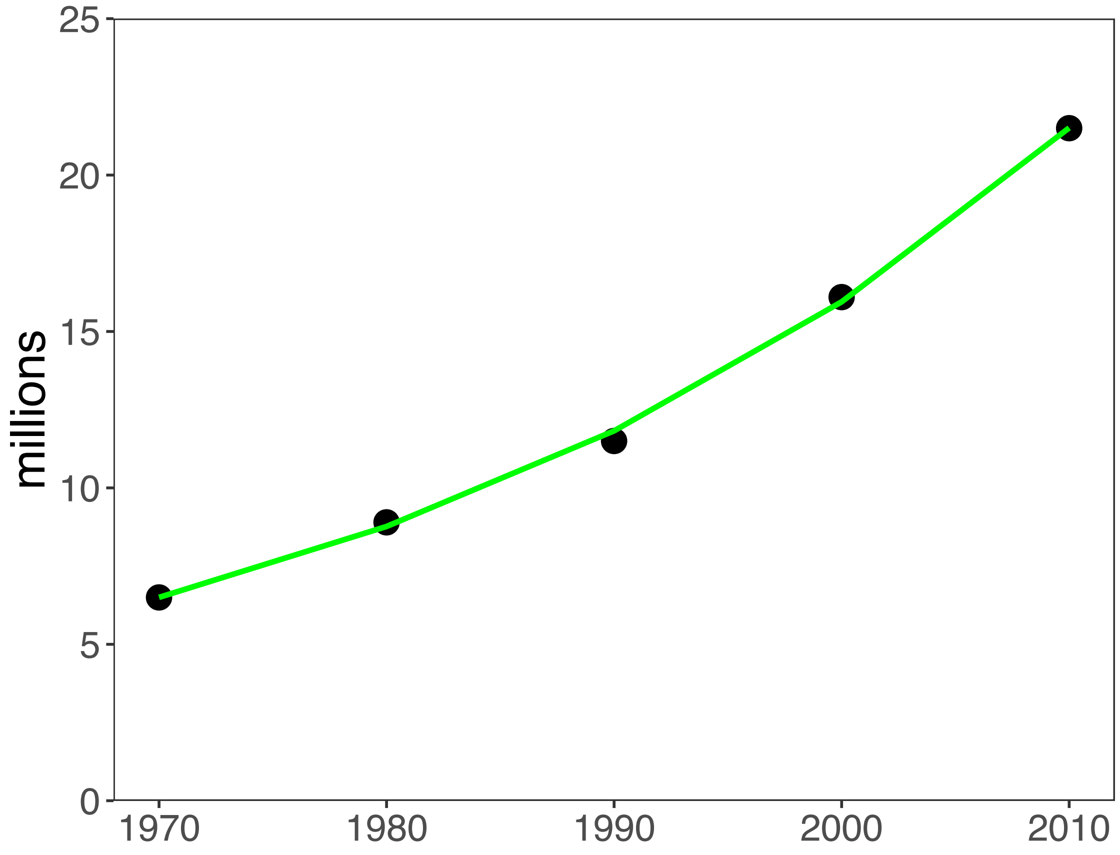


2. Hypothesis: Malagasy population size increases because people are having children.

3. Mechanistic Model:



4. Data:

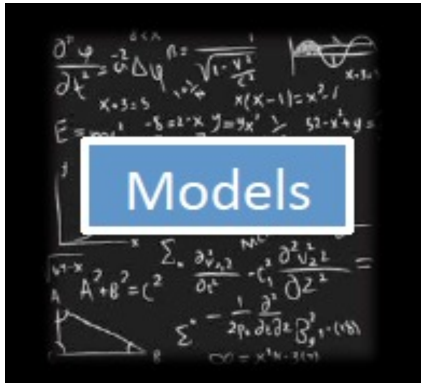


5. Evaluation:

$r =$ .349/person/yr

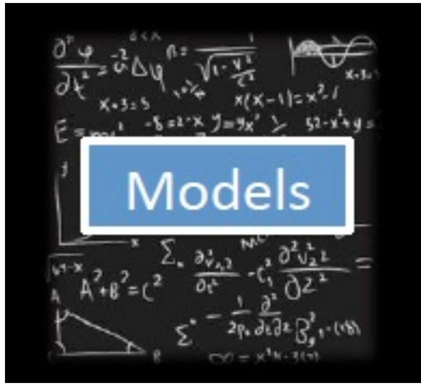
What can we conclude from this fitted model?

Mechanistic Models: Beware!



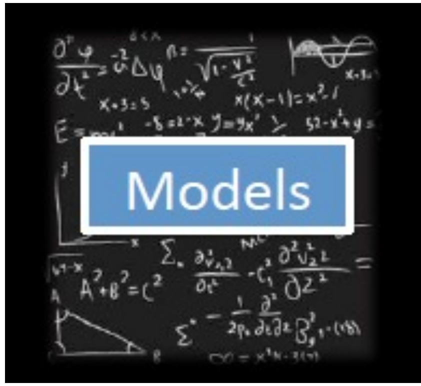
- Parameters used in the mechanistic models sometimes are not measurable!

Mechanistic Models: Beware!



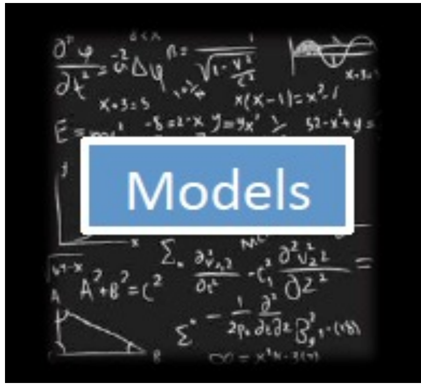
- Parameters used in the mechanistic models sometimes are not measurable!
- Simulations can be computationally intensive

Mechanistic Models: Beware!



- Parameters used in the mechanistic models sometimes are not measurable!
- Simulations can be computationally intensive
- Advances in computational power often inspire development of more complex models which are not necessarily better

Mechanistic Models: Beware!

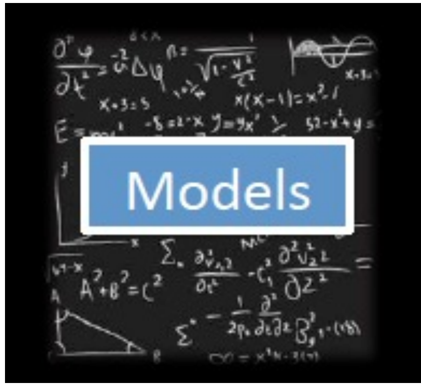


- Parameters used in the mechanistic models sometimes are not measurable!
- Simulations can be computationally intensive
- Advances in computational power often inspire development of more complex models which are not necessarily better

“All models are wrong but some are useful…”

-George Box

Mechanistic Models: Beware!



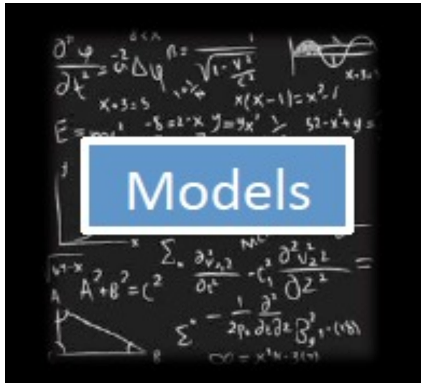
- Parameters used in the mechanistic models sometimes are not measurable!
- Simulations can be computationally intensive
- Advances in computational power often inspire development of more complex models which are not necessarily better

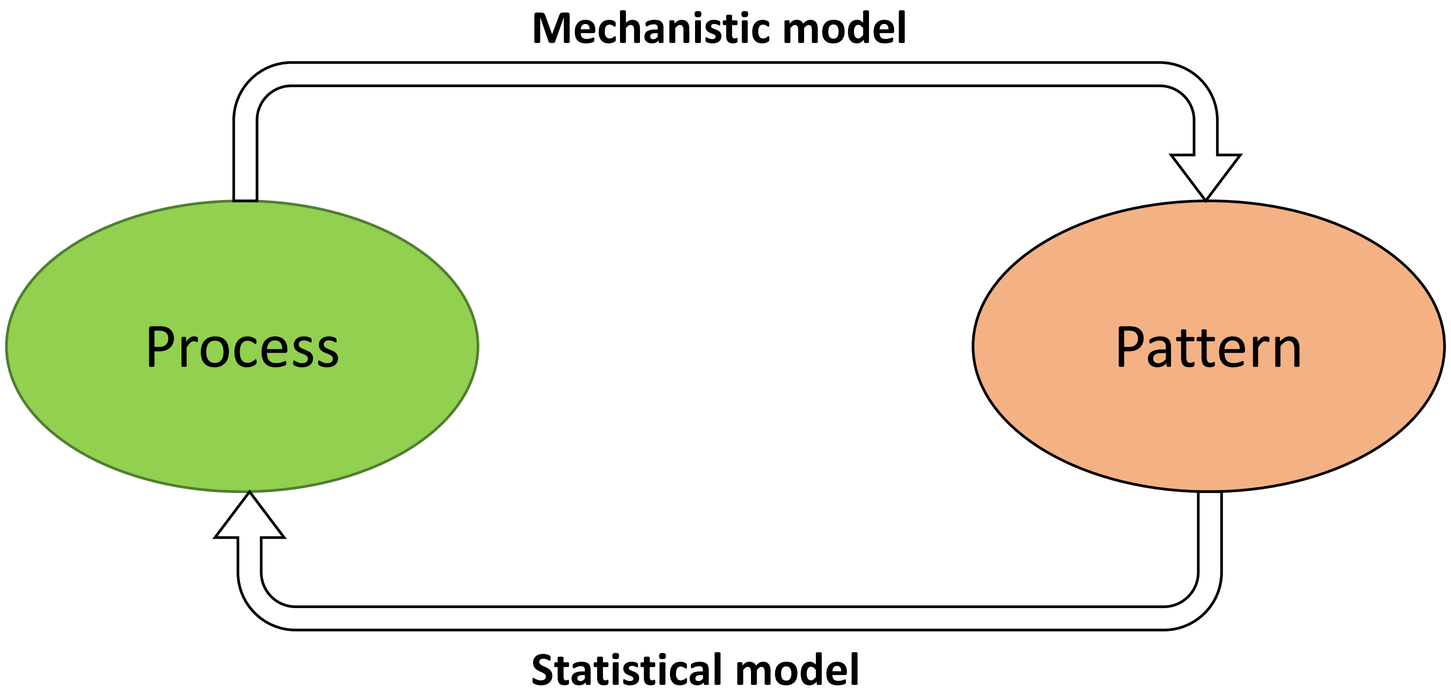
“All models are wrong but some are useful…”

-George Box

We use models to both **predict** and **explain**.

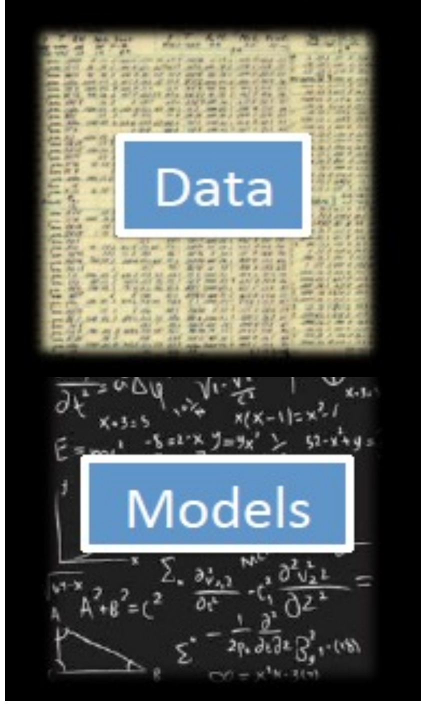
It is ideal when statistical and mechanistic models meet:





A Tool for E^2M^2

- Computer power keeps increasing
- Language/software
  - Fortran, C, C++
  - Julia, Java, Python
  - Matlab, Maple, Mathematica,
  - SAS, SPSS, Stata
- Specific programs
  - Vortex, RAMAS, NetLogo for IBM
  - NicheMapper for physiology, iLand for forest dynamics
  - MaxEnt for species distribution modeling
  - Zonation for reserve selection etc…
- The compromise: R---very powerful for
  - Visualization
  - Data formatting and sorting
  - Statistical analyses
  - Simulation (mechanistic model)



Goals for this lecture

- To explain what we’re doing here
- To define ”science”
- To define ”data”
- To define “models”
- To introduce many different types of models
  - Statistical
  - Mathematical
- To introduce the “E” in E^2M^2
  - Ecology
  - Epidemiology

Goals for this lecture

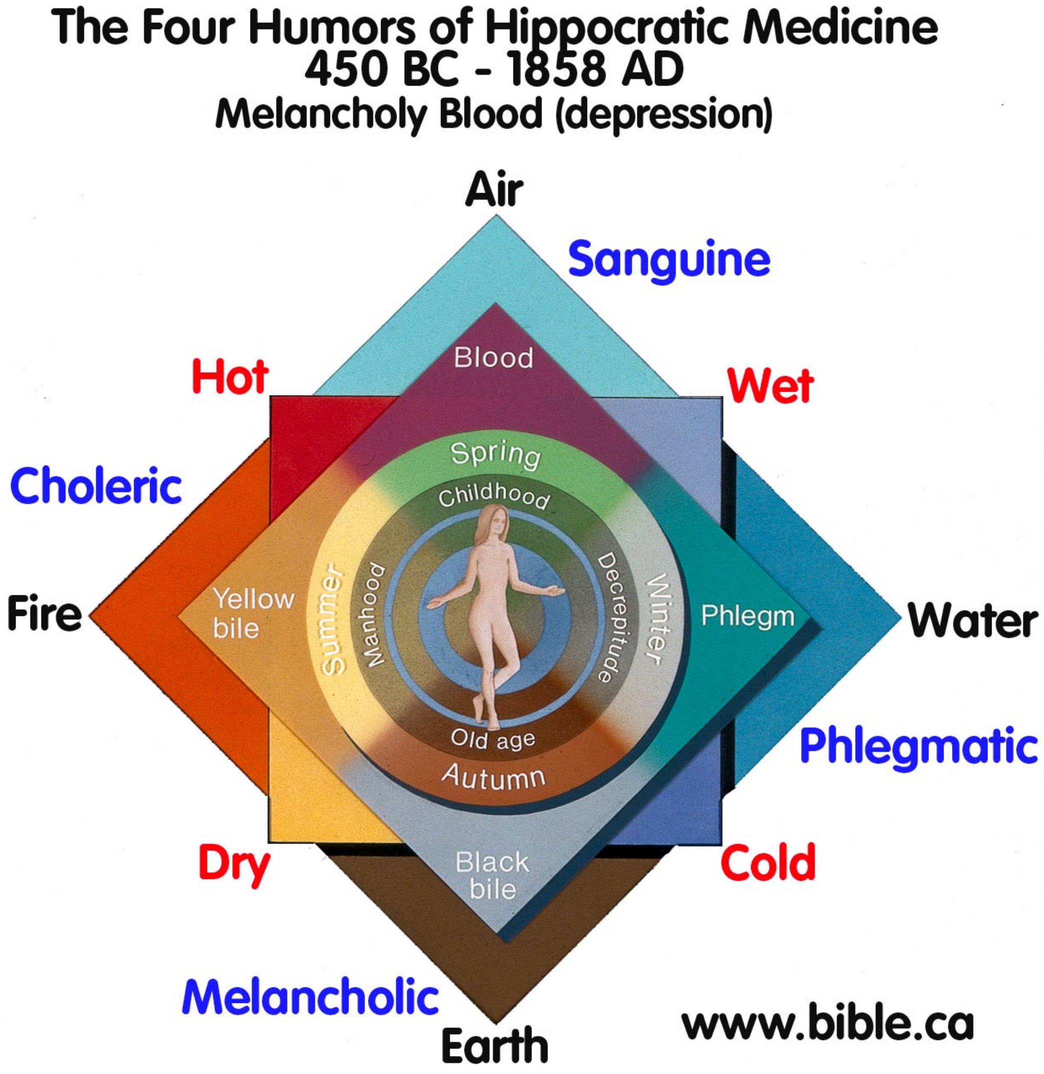
- To explain what we're doing here
- To define "science"
- To define "data"
- To define "models"
- To introduce many different types of models
  - Statistical
  - Mathematical
- To introduce the "E" in E^2M^2
  - Ecology
  - Epidemiology

What is Epidemiology?

- “the study of **what** is on the people”
  - coined in 1802 to describe diseases in the Spanish population
- Emphasis on the study and analysis of the distribution and determinants of health and disease (“risk factors”)

Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)

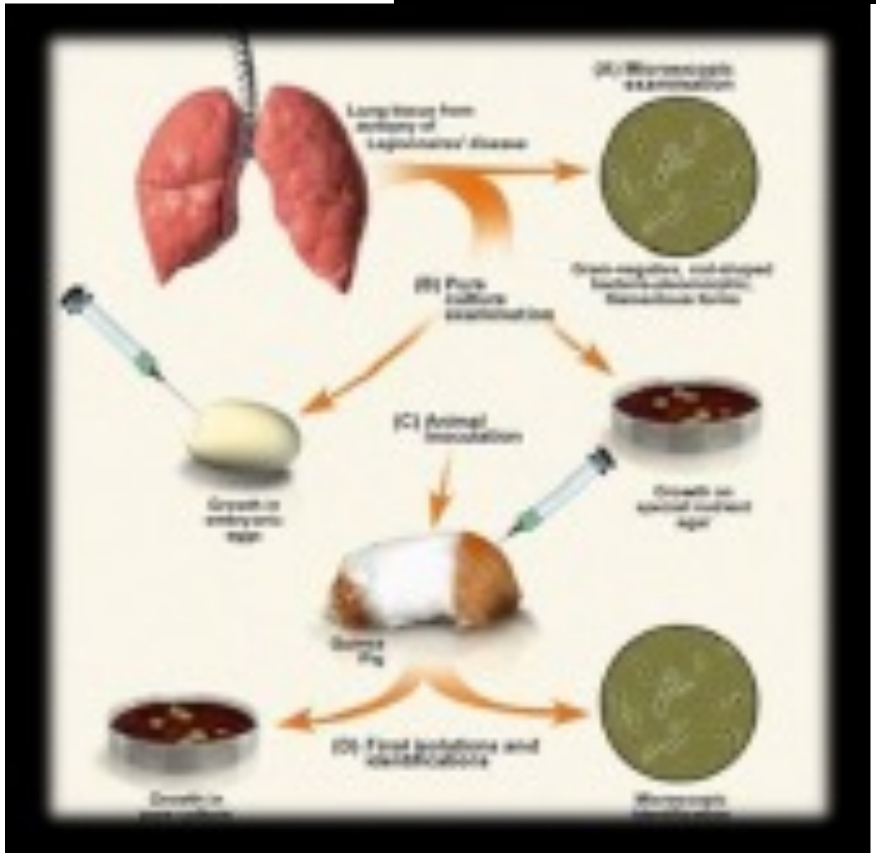


Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)
2. Miasmatic theory of disease (1500s)
   - Sickness results from emanations of 'bad air'

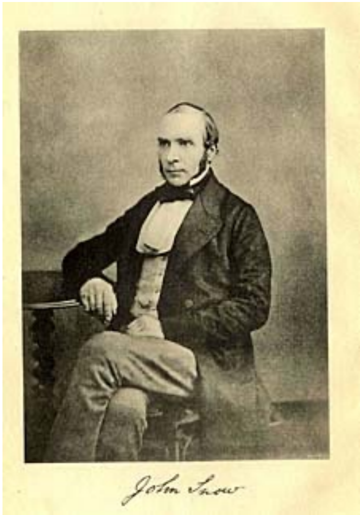
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)
2. Miasmatic theory of disease (1500s)
   - Sickness results from emanations of ‘bad air’
3. Germ theory of disease
   - Leeuwenhoek’s microscope (1675)
   - Koch’s postulates (1890)



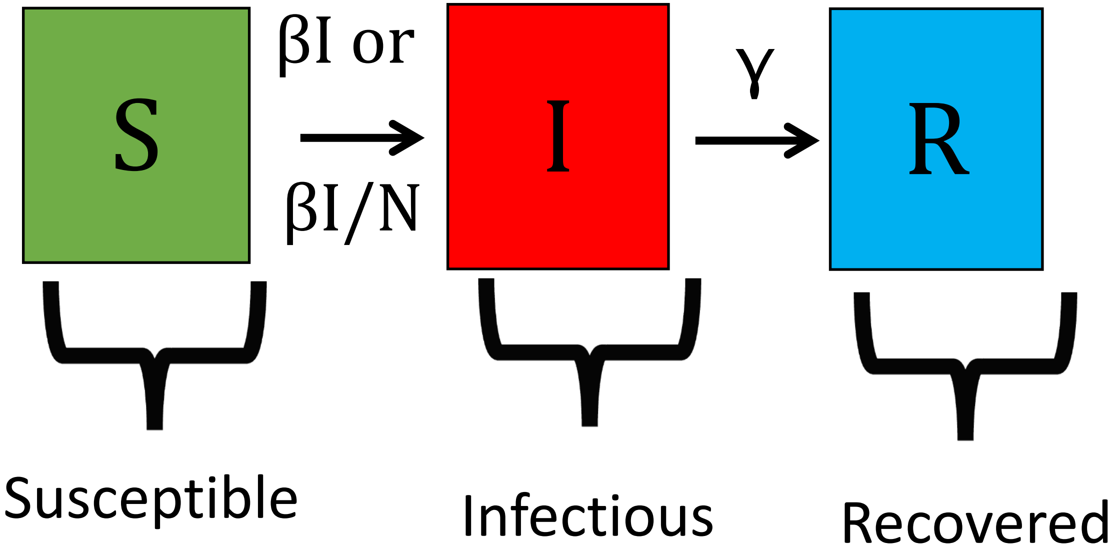
Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)
2. Miasmatic theory of disease (1500s)
   - Sickness results from emanations of 'bad air'
3. Germ theory of disease
   - Leeuwenhoek's microscope (1675)
   - Koch's postulates (1890)
4. Classical epidemiology
   - John Snow and London cholera (1854)



Models in Epidemiology

1. Sickness caused by an imbalance in the four humors (Hippocrates)
2. Miasmatic theory of disease (1500s)
   - Sickness results from emanations of ‘bad air’
3. Germ theory of disease
   - Leeuwenhoek’s microscope (1675)
   - Koch’s postulates (1890)
4. Classical epidemiology
   - John Snow and London cholera (1854)
5. Mathematical epidemiology
   - Kermack and McKendrik (1927)



What is Ecology?

- The study of the interactions of organisms and their environment
  - Coined in 1866 by German scientist Ernst Haeckel
  - Nile crocodiles opening mouths for sandpipers (Herodotus)
- Emphasis on explaining dynamical processes in nature

Models in Ecology

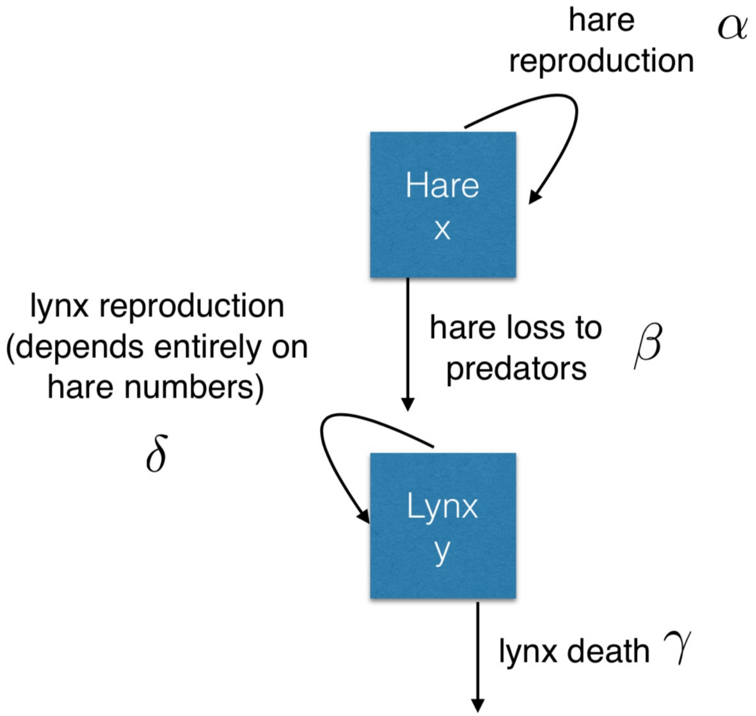
1. Origins in plant biology and succession
    - Clements and Gleason (1910s-20s)

Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)
2. Food Chains
   - Charles Elton (1920s)

Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)
2. Food Chains
   - Charles Elton (1920s)
3. Predator-Prey Dynamics
   - Lotka-Voltera (1920s)

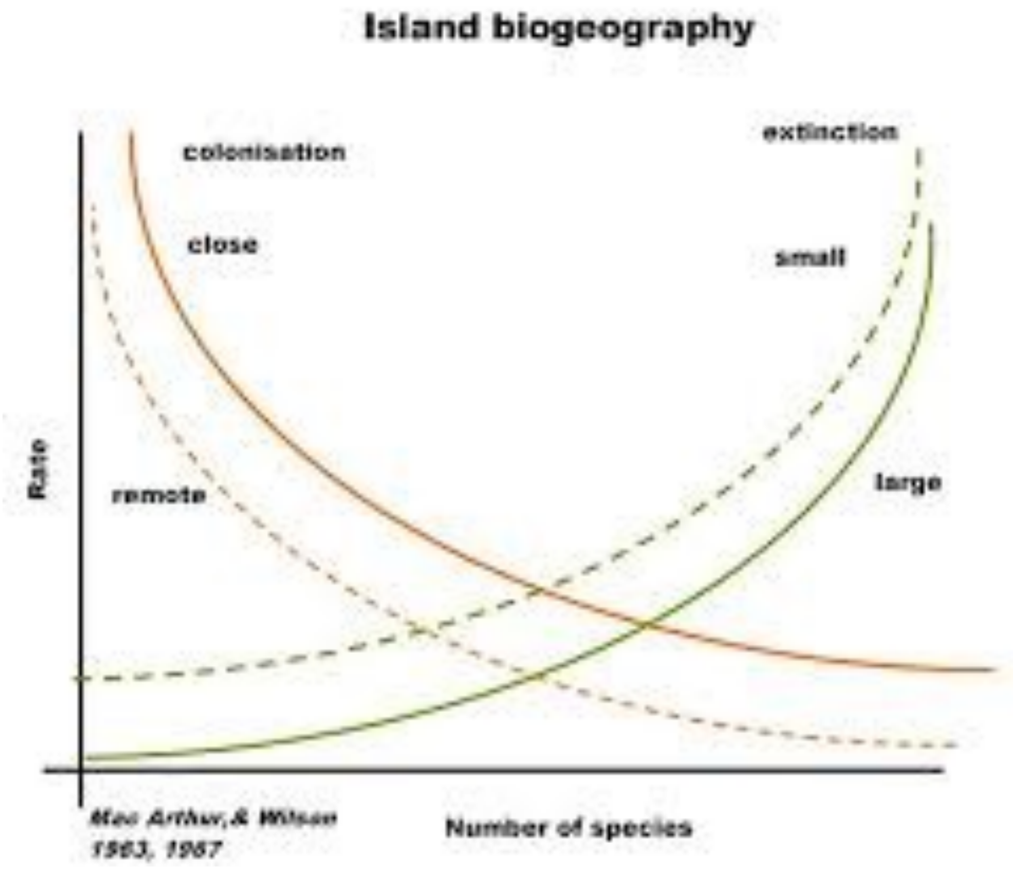


$$\frac{dx}{dt} = x(\alpha - \beta y)$$

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

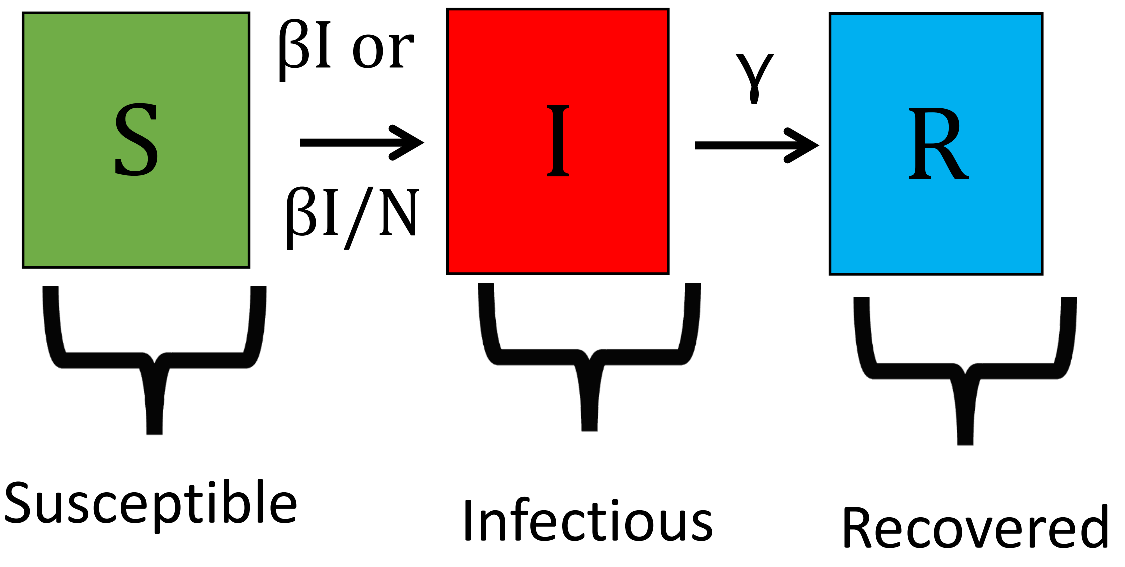
Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)
2. Food Chains
   - Charles Elton (1920s)
3. Predator-Prey Dynamics
   - Lotka-Voltera (1920s)
4. Mathematical Ecology
   - Robert MacArthur (1950s)
   - Island biogeography



Models in Ecology

1. Origins in plant biology and succession
   - Clements and Gleason (1910s-20s)
2. Food Chains
   - Charles Elton (1920s)
3. Predator-Prey Dynamics
   - Lotka-Voltera (1920s)
4. Mathematical Ecology
   - Robert MacArthur (1950s)
   - Island biogeography
5. Disease Ecology
   - Anderson and May (1980s)
   - Island biogeography



Friday: R Bootcamp
- Intro to R Studio
- Exploring and Visualizing Data in R
- For-loops, Functions, and If-Else Statements

Saturday: Travel

Sunday: "Getting Started with the Basics"
- Data and Models
- Student introductions & presentations
- Linear regression & simple statistics
- Linear regression tutorial

Monday: "Using Models with Data"
- Model-Guided Study Design
- Study design tutorial
- Intro to Compartmental Models & Differential Equations
- Building mechanistic models in R
- Refining research questions for modeling
- Defining a model world

Tuesday: "Applying Simple Models"
- Dynamical Fever
- Introduction to Mixed Modeling
- Reading a Research Paper

Wednesday: "Refining Your Work"
- Model Fitting in Practice – the Basic Concept
- Epidemic Cards
- Model Fitting with Epidemic Cards
- Model Selection and Comparison
- Model Selection Tutorial
- Model Telephone

Thursday: "Bells and Whistle"
- Introduction to Phylogenetic Modeling • Tutorial
- **Intoduction to Spatial Visualization and Plotting** • Tutorial
- Modeling Insights from the Metapopulation Game • Tutorial
- Introduction to Network Modeling • Tutorial
- Research Snapshots

Friday: "Putting it All in Perspective"
- The Life Cycle of a Modeling Project
- C4C Student Presentations

Saturday: Travel

January: "Sharing Your Work"
- Final student presentations

- **Programming**
- **Data**
- **Models**
- **Research Development**

Misaotra!

ICI3D International Clinics on Infectious Disease, Dynamics, & Data

MMED: ***Clinic on the Meaningful Modeling of Epidemiological Data***

May-June 2023, Cape Town, South Africa

DAIDD: ***Clinic on Dynamical Approaches to Infectious Disease Data***

December 2023, Virtual

South African Center for Epidemiological Modeling and Analysis (SACEMA), Director

Dr. Juliet Pulliam

University of Stellenbosch

www.ici3d.org