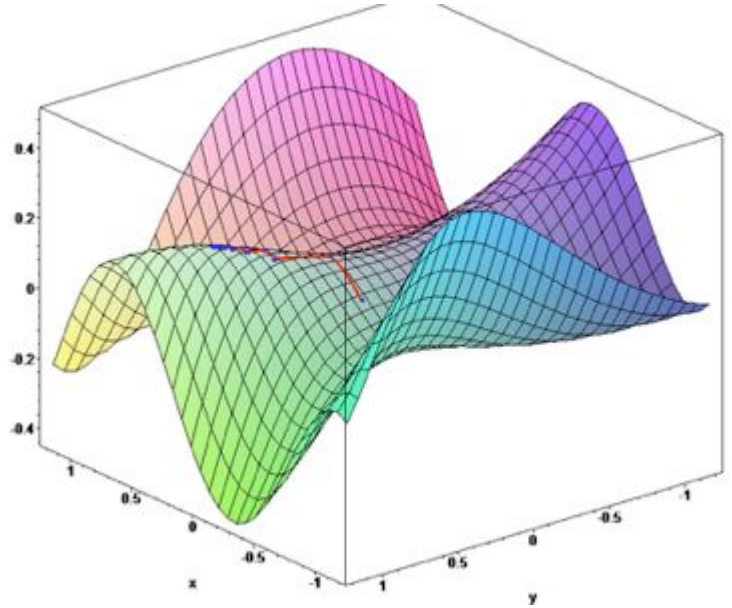



# Model Fitting: The Basic Concept

Created by Cara Brook
and Michelle Evans

Presented by Michelle Evans

E2M2
December 2022

# Model Fitting in Science

1. Define your research question
2. Formulate a hypothesis
3. Collect Data
4. Construct a model that demonstrates your hypothesis
5. Assess model fit: assuming our model is true, how likely are we to recover the observed data?
6. Optimize parameters behind the model to result in best model fit

# Model Fitting in Science

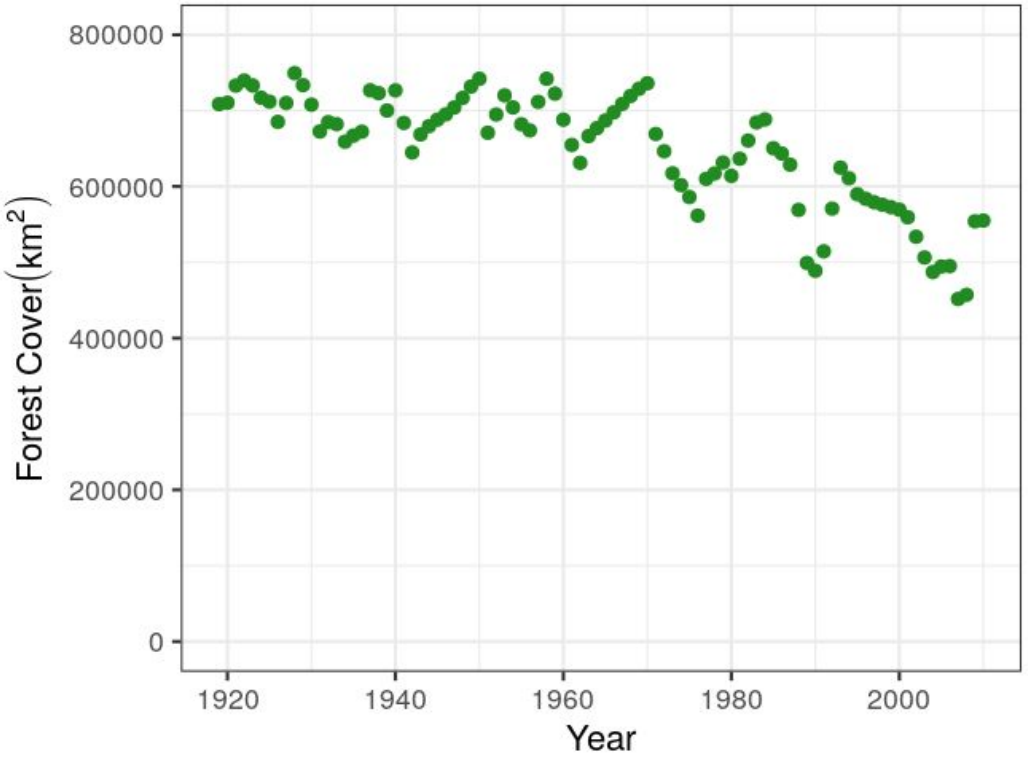
1. Define your research question
2. Formulate a hypothesis
3. Collect Data
4. Construct a model that demonstrates your hypothesis
5. Assess model fit: assuming our model is true, how likely are we to recover the observed data?
6. Optimize parameters behind the model to result in best model fit

**Statistical and Mechanistic**

# Statistical models are **data-driven**

Goal: find patterns and correlations in data

What is the trend in Madagascar's forest cover through time?



# 1. Construct a model that represents our hypothesis

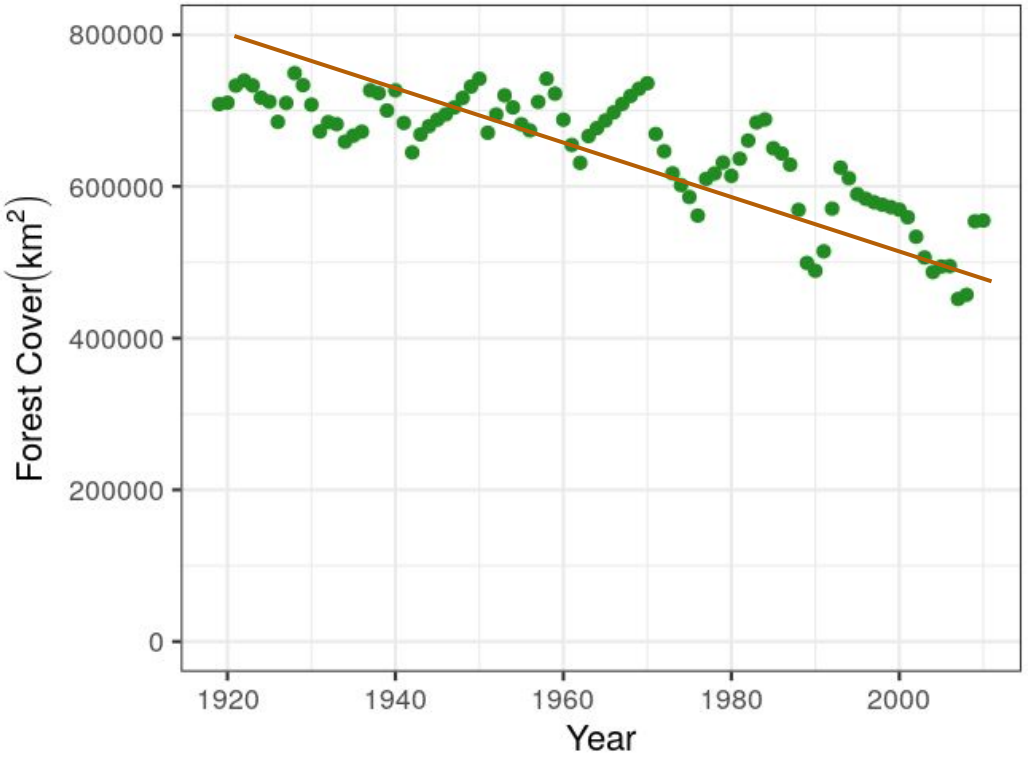
What is the trend in Madagascar's **forest cover** through **time**?

Forest = f(time)

Forest = slope*year + intercept

Y = mx +b

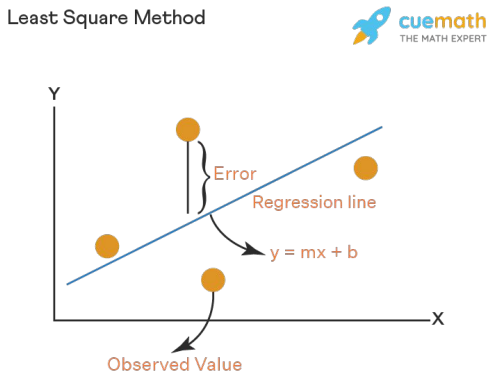
Linear regression

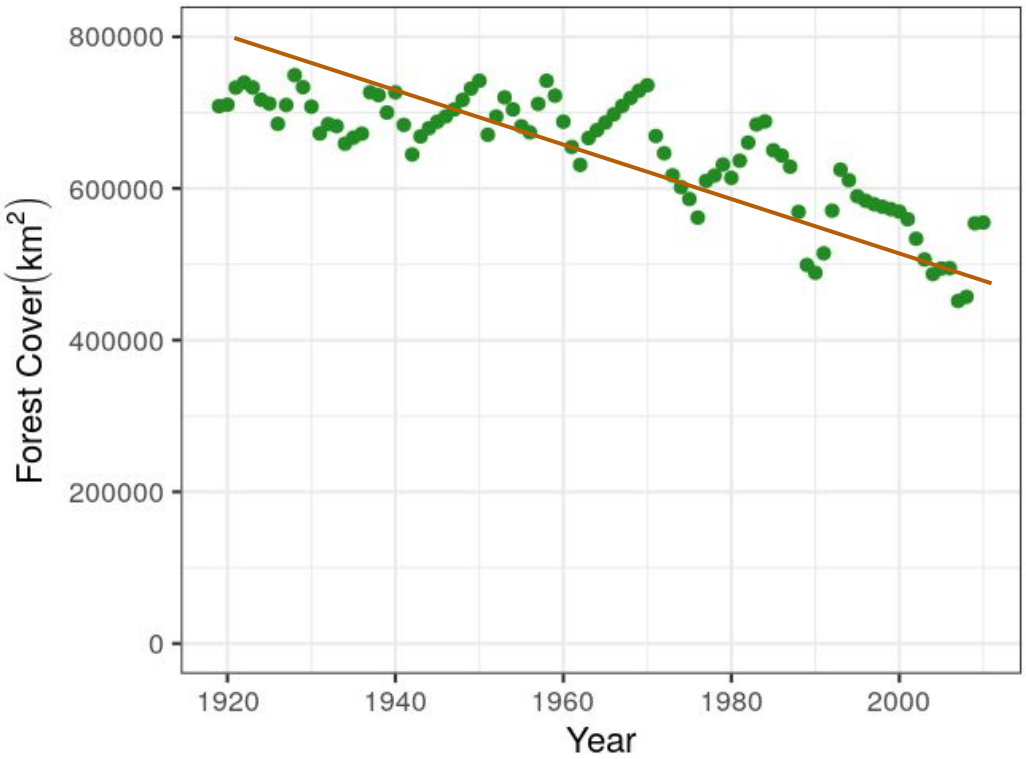


# 2. Assess model fit

Given our model (y= mx +b), how likely are we to recover the observed data?

Least squares = $\sum_i (data_i - prediction_i)^2$

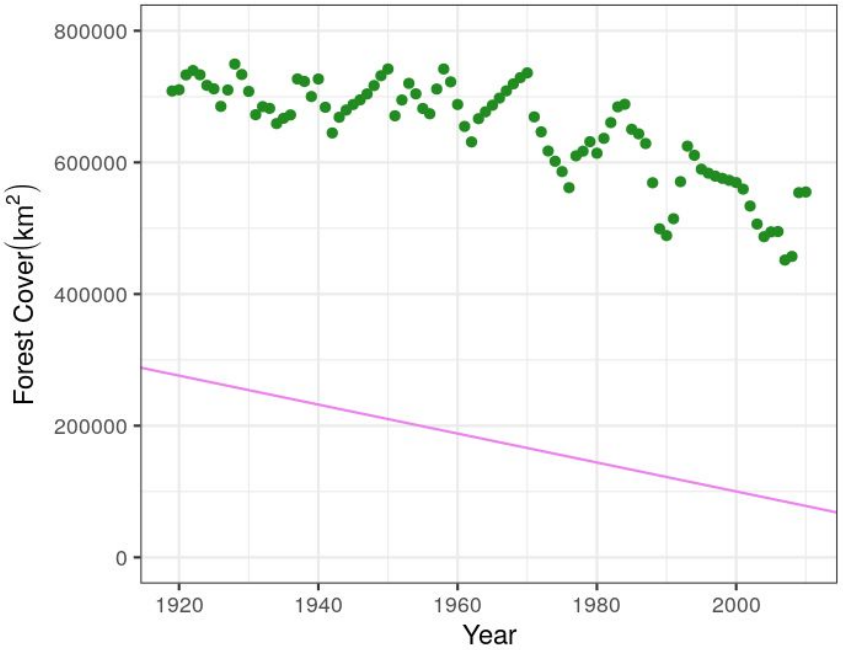


# 3. Optimize the parameters

## Optimize slope (m) and intercept (b)

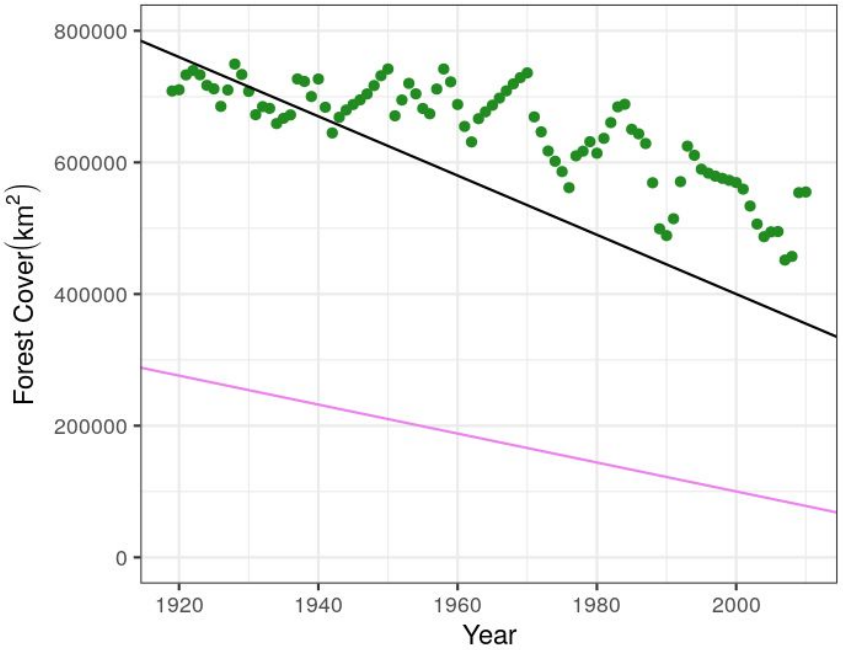
| Slope (m) | Intercept (b) |
|---|---|
| -2200 | 4.5e6 |
| | |
| | |
| | |



# 3. Optimize the parameters

## Optimize slope (m) and intercept (b)

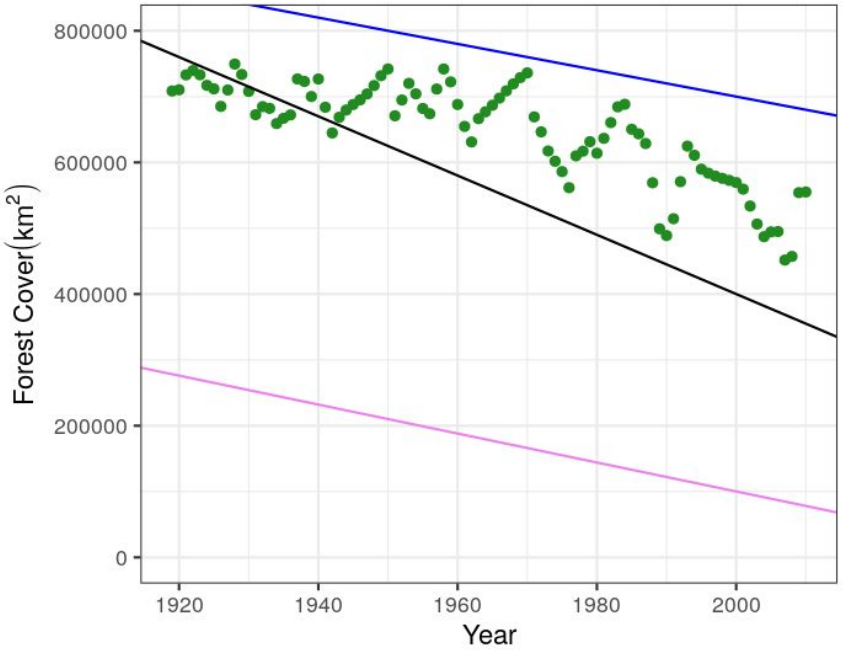
| Slope (m) | Intercept (b) |
|---|---|
| -2200 | 4.5e6 |
| -4500 | 9.4e6 |
| | |
| | |



# 3. Optimize the parameters

## Optimize slope (m) and intercept (b)

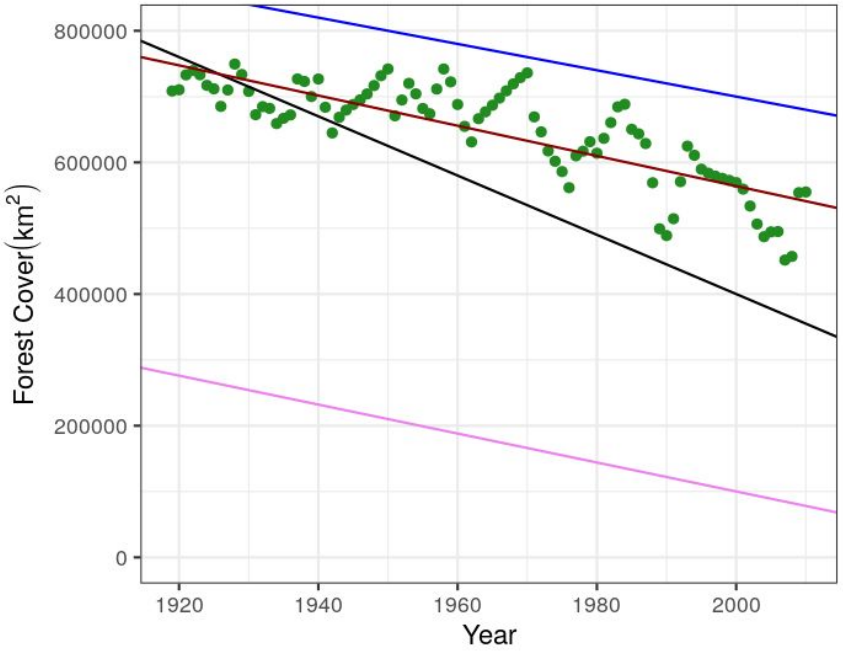
| Slope (m) | Intercept (b) |
|---|---|
| -2200 | 4.5e6 |
| -4500 | 9.4e6 |
| -2000 | 4.7e6 |
| | |



# 3. Optimize the parameters

## Optimize slope (m) and intercept (b)

| Slope (m) | Intercept (b) |
|---|---|
| -2200 | 4.5e6 |
| -4500 | 9.4e6 |
| -2000 | 4.7e6 |
| -2293 | 5.2e6 |



# What do we learn from this model?

**Model :**

y = mx +b

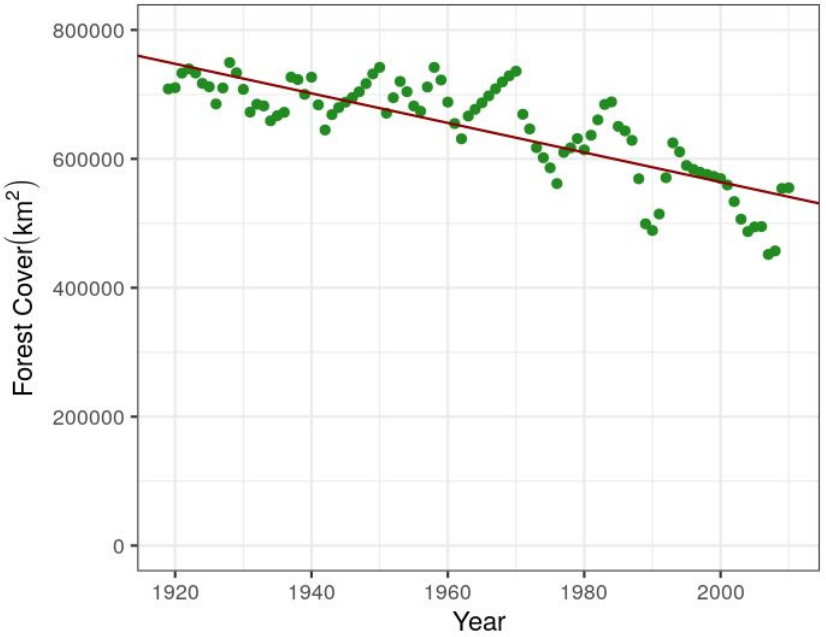
Forest = m*year + b

**Parameters:**

m = -2293

b = 5,200,000

The slope (m) is negative, so there is a **negative relationship** between time and forest cover.



# What do we learn from this model?

**Model :**

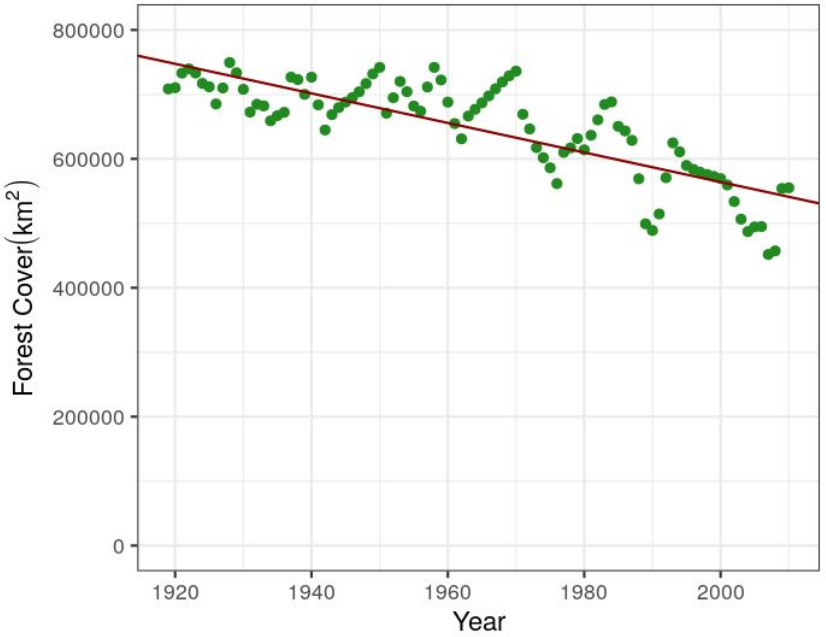
y = mx +b

Forest = m*year + b

**Parameters:**

m = -2293

b = 5,200,000




The slope (m) is negative, so there is a **negative relationship** between time and forest cover.

This model does not explain **causation**.

# Mechanistic modeling is **process-driven**

We want to understand what happened,
when it happened, and why it happened

Build a model that uses explicit **processes**
to recover the same outcomes (“**states**”) as
our data

# Mechanistic modeling is **process-driven**

We want to understand what happened, when it happened, and why it happened

Build a model that uses explicit **processes** to recover the same outcomes (“**states**”) as our data

What states are in our data?

# Mechanistic modeling is **process-driven**

We want to understand what happened, when it happened, and why it happened

Build a model that uses explicit **processes** to recover the same outcomes (“**states**”) as our data

What states are in our data?

Forest

Savanna

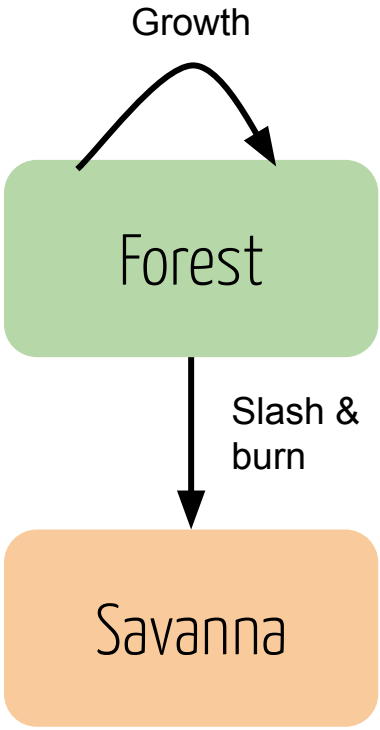
# Mechanistic modeling is **process-driven**

We want to understand what happened, when it happened, and why it happened

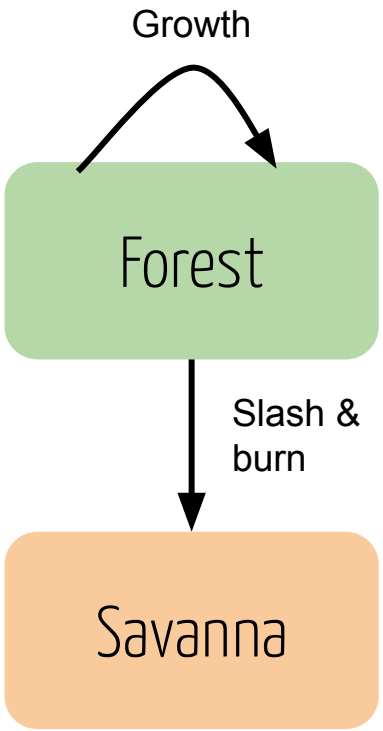
Build a model that uses explicit **processes** to recover the same outcomes (“**states**”) as our data

What states are in our data?

What processes are in our data?



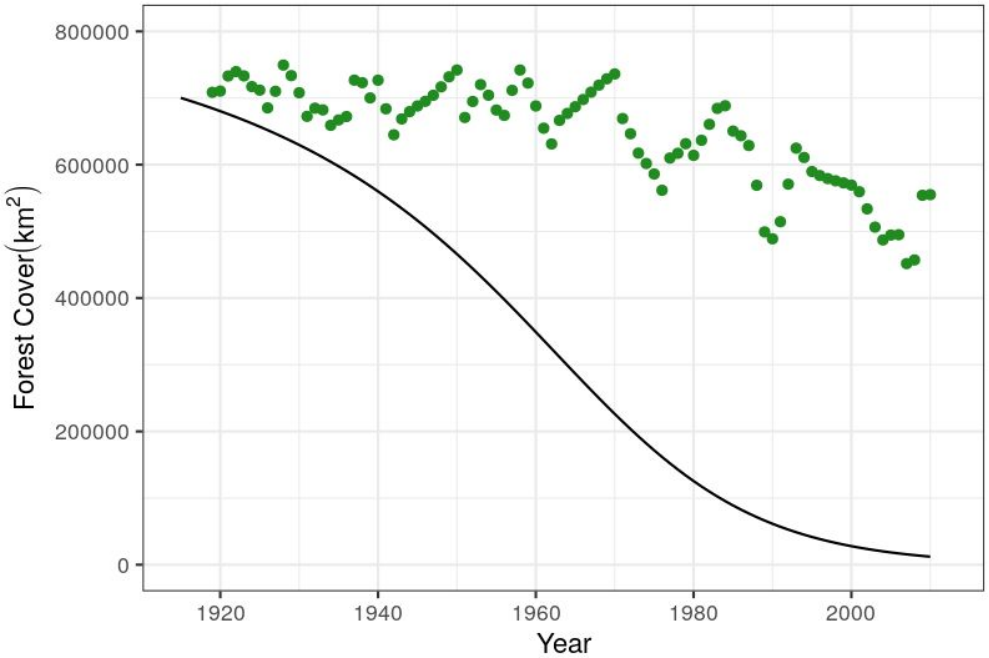
# 1. Construct a model




$$\frac{dF}{dt} = rF\frac{K-N}{K} - \gamma FS$$

$$\frac{dS}{dt} = \gamma FS\frac{K-N}{K}$$

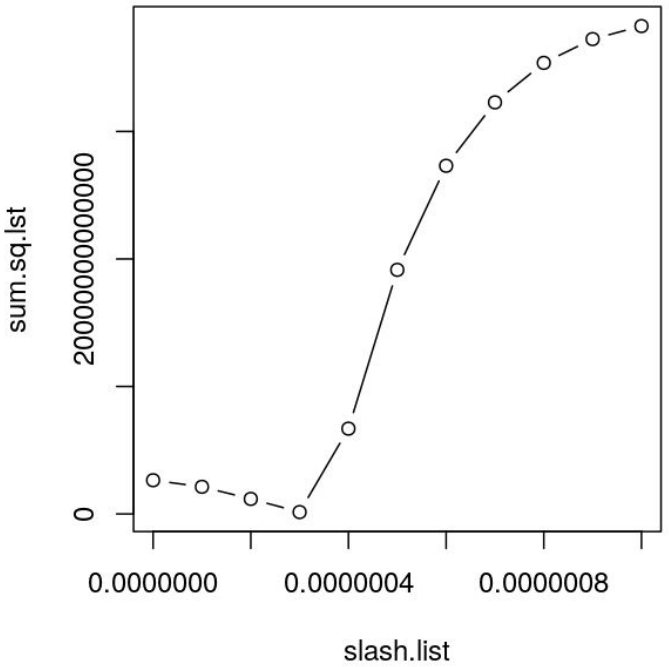
# 2. Assess model fit




Our model predicts forest would decline faster than the data do

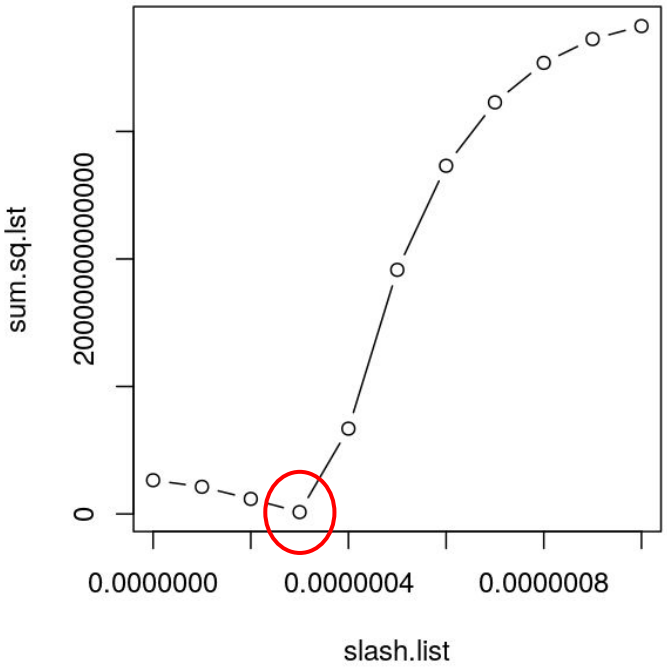
What does this suggest about our guess for the slash and burn rate?

# 3. Optimize the model




Identify the value for the slash and burn rate that minimizes the sum of least squares

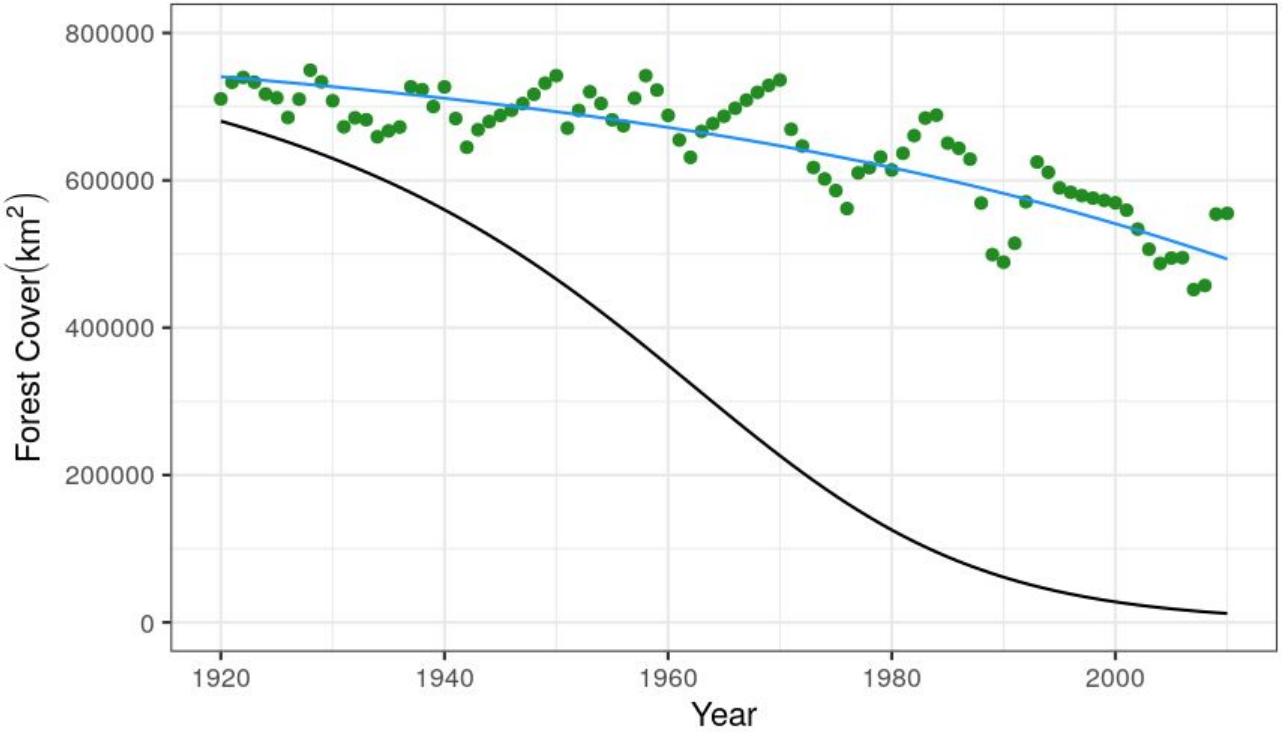
# 3. Optimize the model




Identify the value for the slash and burn rate that minimizes the sum of least squares

Fit the model with this optimized value...

# Does this optimal value result in a model that better matches the data?



# Whether fitting statistical or mechanistic models:

**Statistical:** identify patterns and correlations in data

**Mechanistic:** understand the processes (what, when, why) that resulted in the data

## Three steps

1. Construct a model that fits your hypothesis
2. Assess model fit to the data
3. Optimize parameters in the model that result in the best model fit