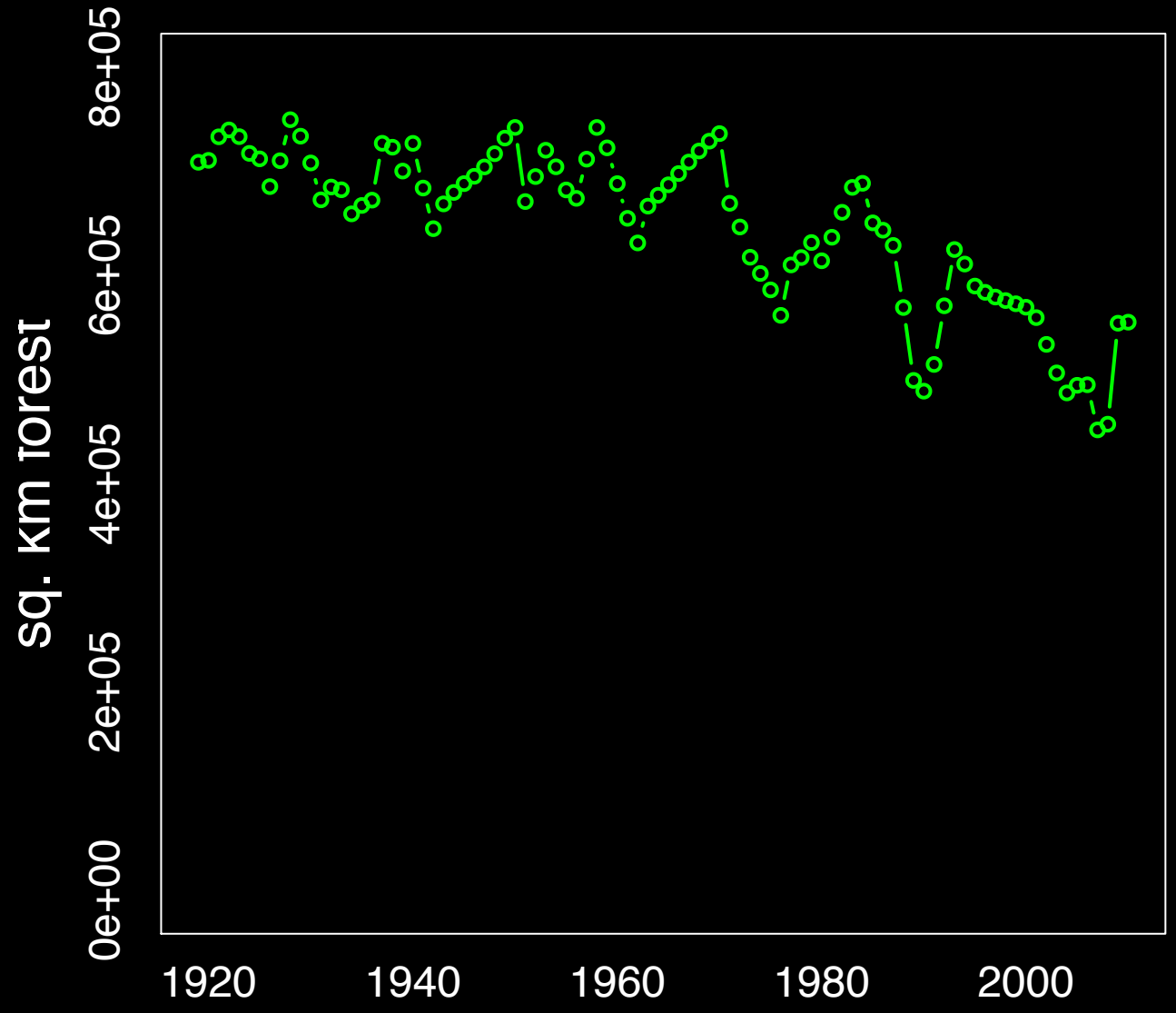
# Model Fitting: The Basic Concept

Cara Brook

$E^2M^2$, Centre ValBio

Ranomafana National Park, Madagascar

# What happened to Madagascar's forest?



# How to fit a model to data

1. Build a model that reproduces the data.

# How to fit a model to data

1. Build a model that reproduces your data.
2. Use any statistical tool (i.e. maximum likelihood, least squares) to ask, assuming our model is true, how likely are we to recover the observed data?

# How to fit a model to data

1. Build a model that reproduces your data.
2. Use any statistical tool (i.e. maximum likelihood, least squares) to ask, assuming our model is true, how likely are we to recover the observed data?
3. Optimize the parameters behind the model to make it most likely to recover the data.

# How to fit a model to data

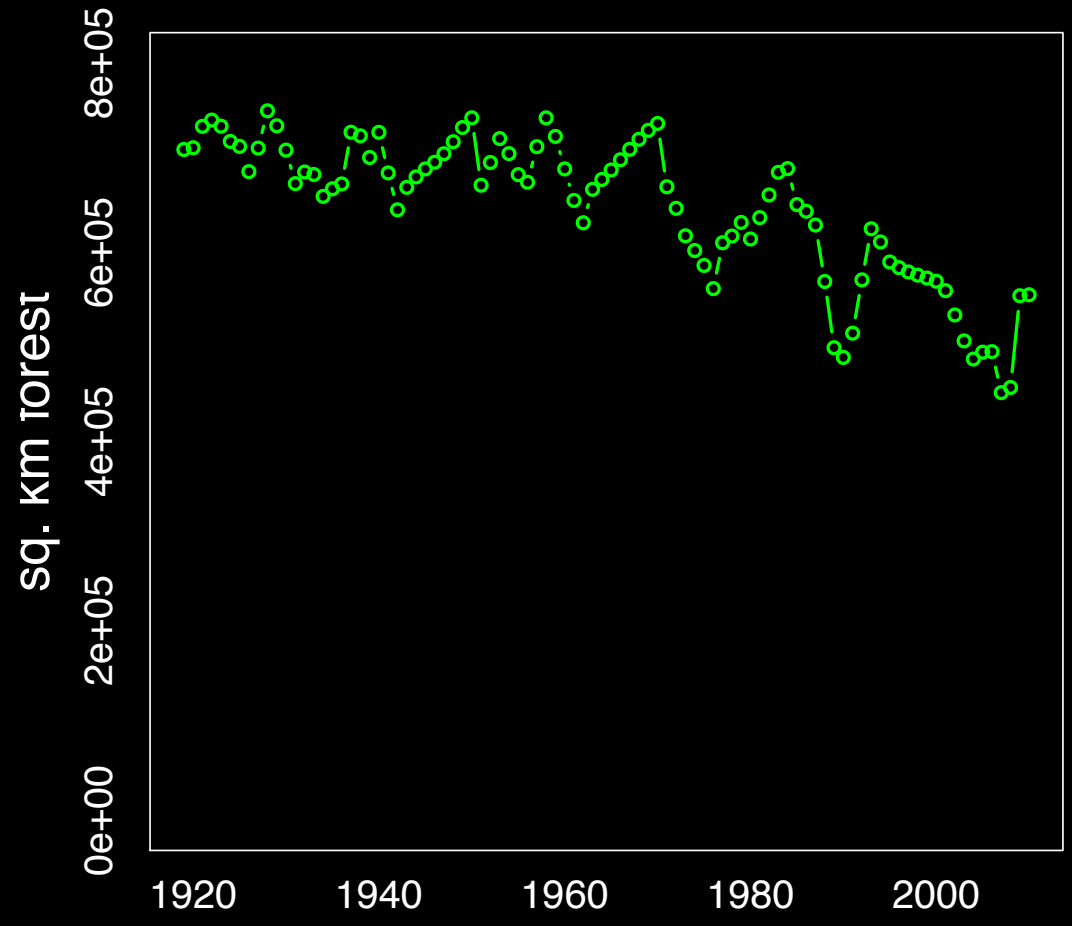
1. Build a model that reproduces your data.
2. Use any statistical tool (i.e. maximum likelihood, least squares) to ask, assuming our model is true, how likely are we to recover the observed data?
3. Optimize the parameters behind the model to make it most likely to recover the data.
4. If need be, restructure your model to better match your data.

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations

# Traditional statistics is data-driven…

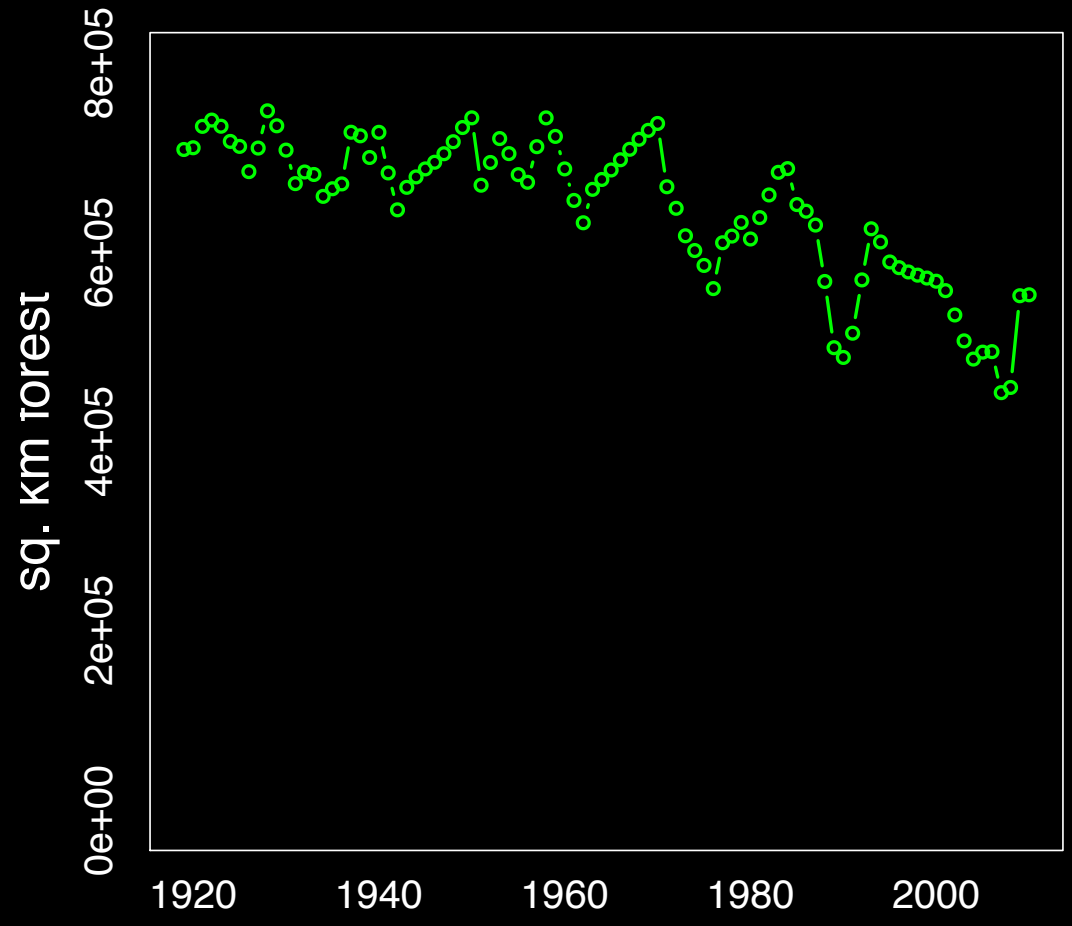
- We might ask questions about these data:
- Goal: find correlations




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

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations

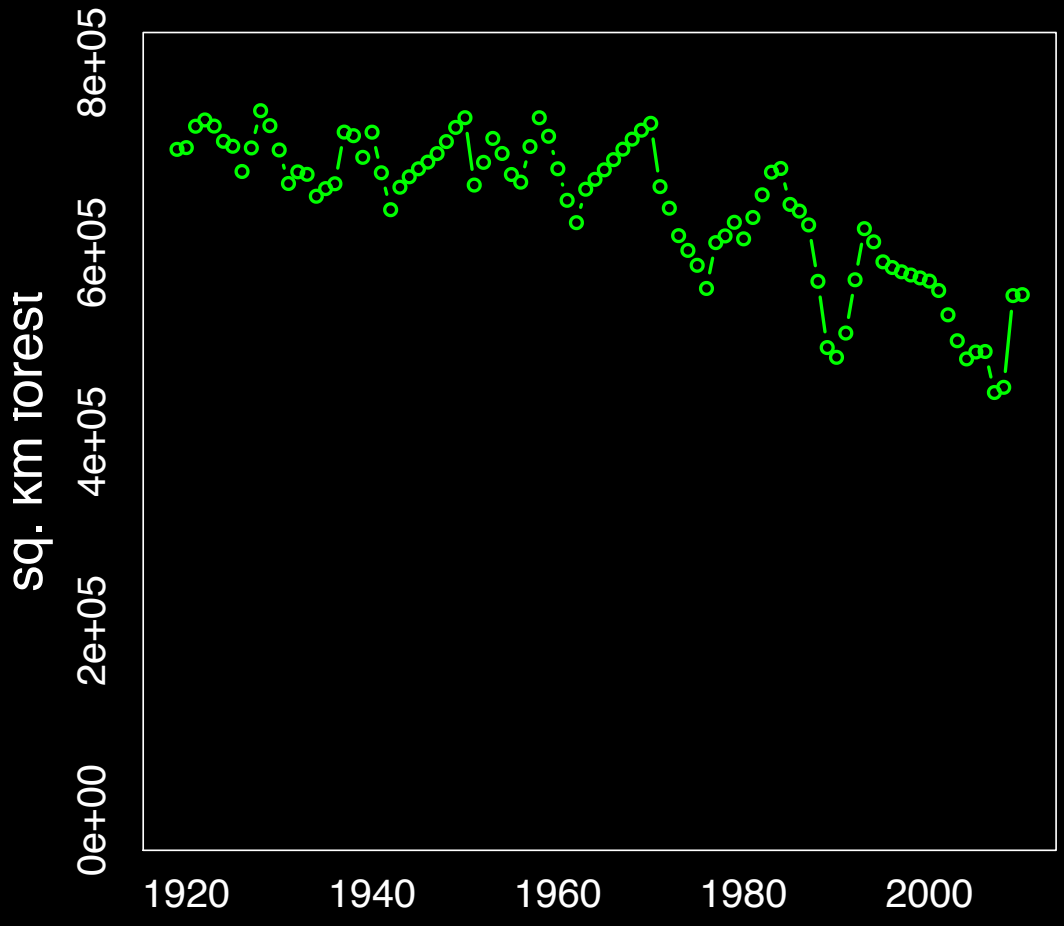



*What is the trend in Madagascar's forest cover through time?*

*Can fit a linear regression.*

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations



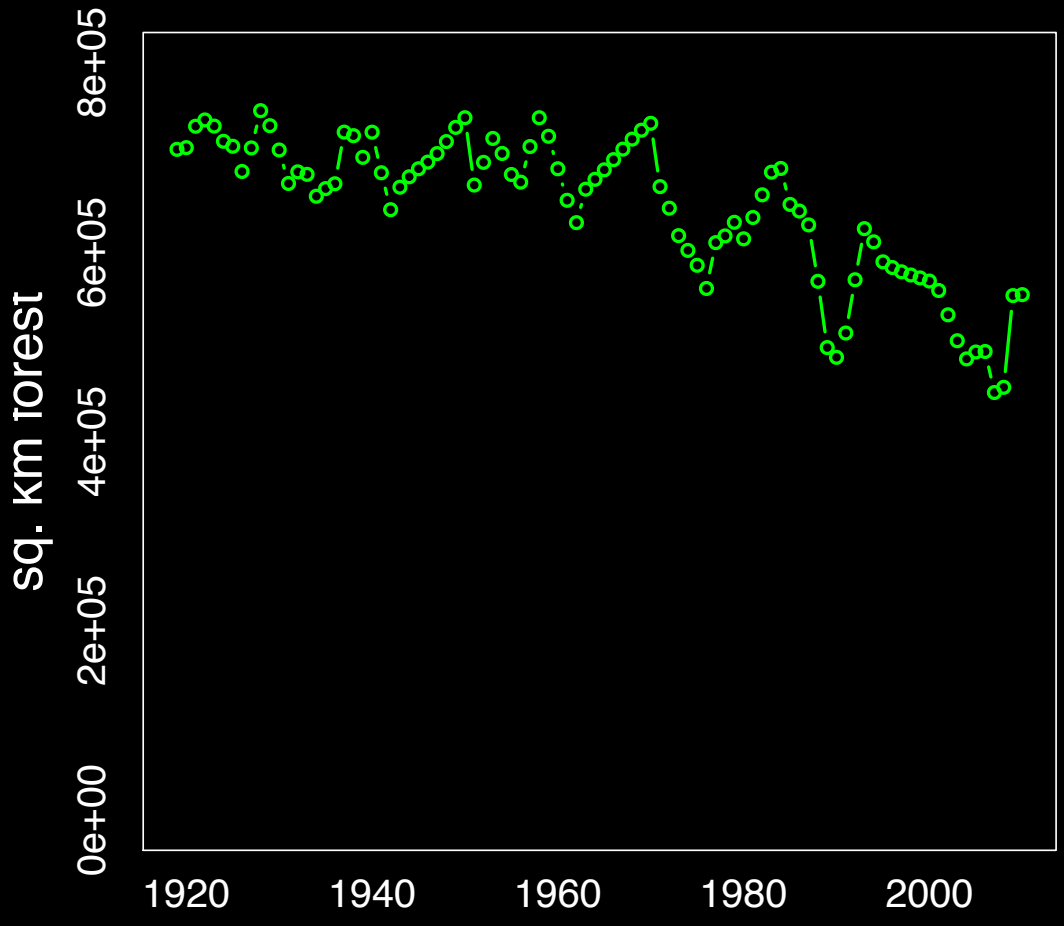

1. Build a model that reproduces your data.

$y = mx + b$

$km^2\ forest = slope*year + y.int$

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations



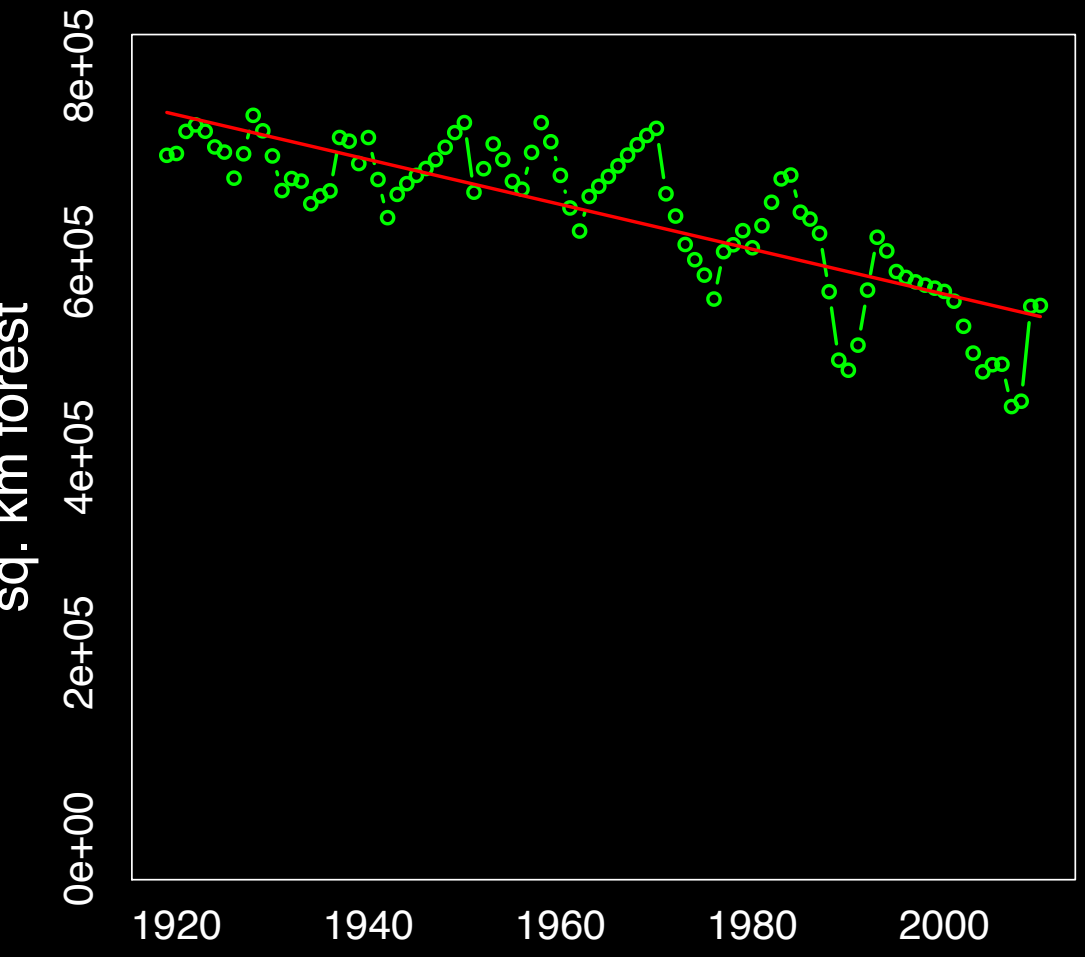

1. Build a model that reproduces your data.

$y = mx + b$

$km^2\ forest = slope*year + y.int$

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations




$y = mx + b$

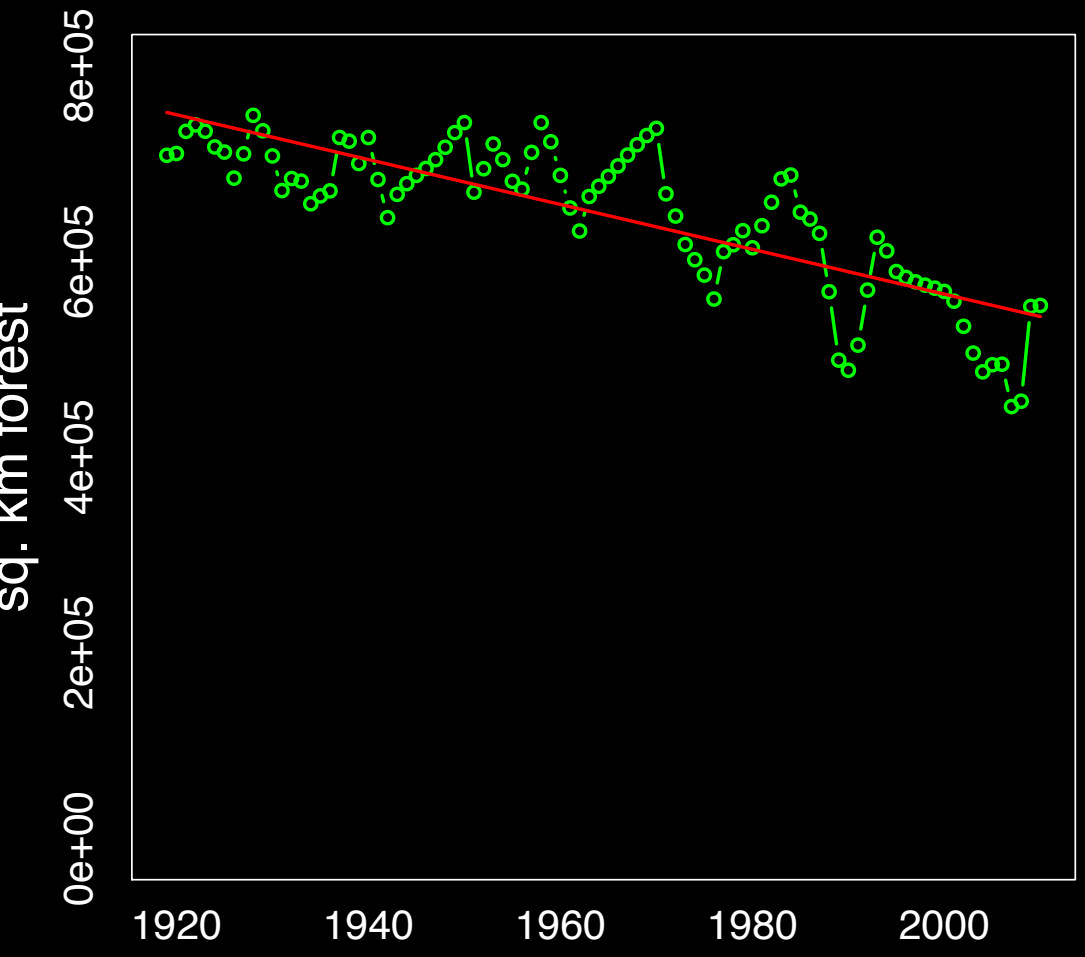
m = -2293

b = 5152515.5

1. Build a model that reproduces your data.
2. Use any statistical tool (i.e. maximum likelihood, least squares) to ask, assuming our model is true, how likely are we to recover the observed data?
3. Optimize the parameters behind the model to make it most likely to recover the data.

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations




$y = mx + b$

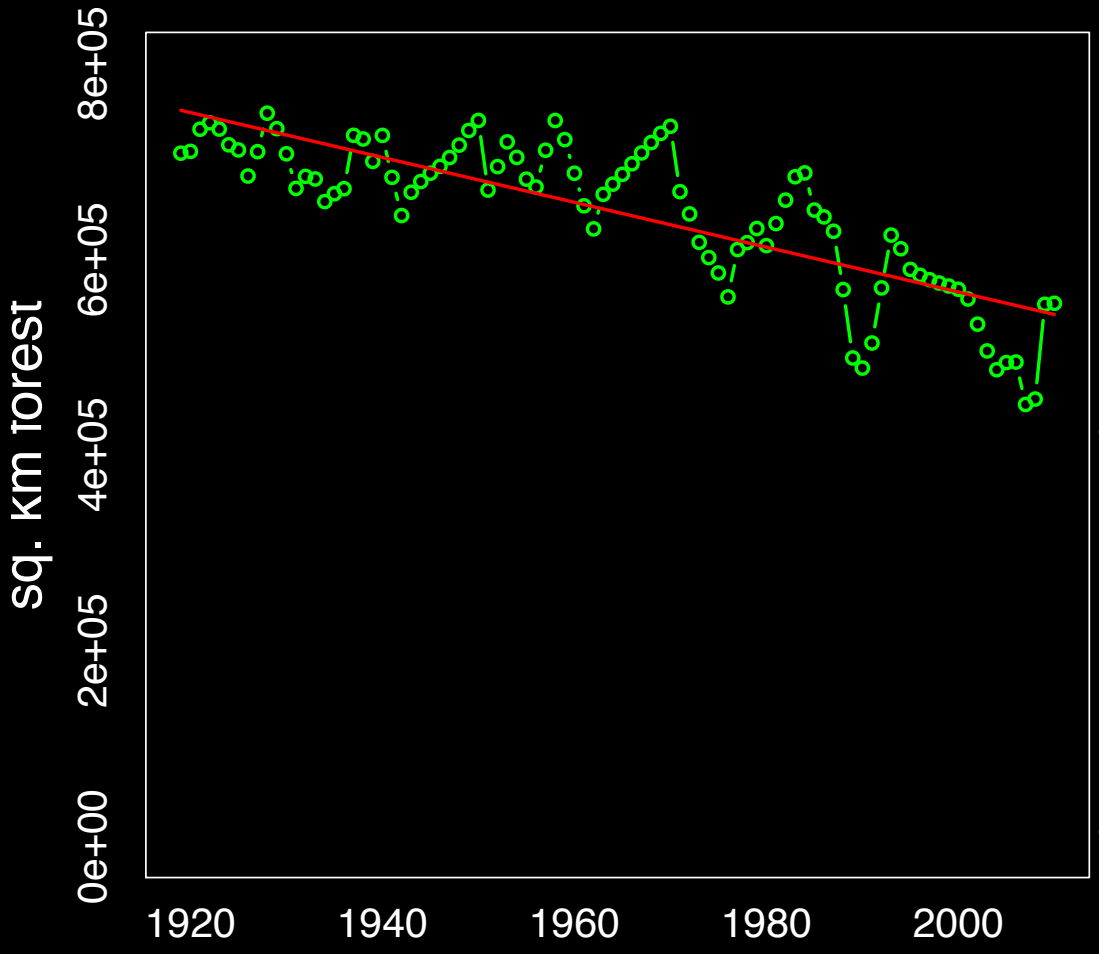
m = -2293

b = 5152515.5

- *This tells us that the time trend in forest cover is negative (declining) and that there was ~5*10^6 sq. km of forest in 1920.*

# Traditional statistics is data-driven…

- We might ask questions about these data:
- Goal: find correlations




$y = mx + b$

m = -2293

b = 5152515.5

- *This tells us that the time trend in forest cover is negative (declining) and that there was ~5*10^6 sq. km of forest in 1920.*
- *But we know nothing about causation.*

# Mechanistic modeling is process-driven…

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

# Mechanistic modeling is process-driven…

- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns

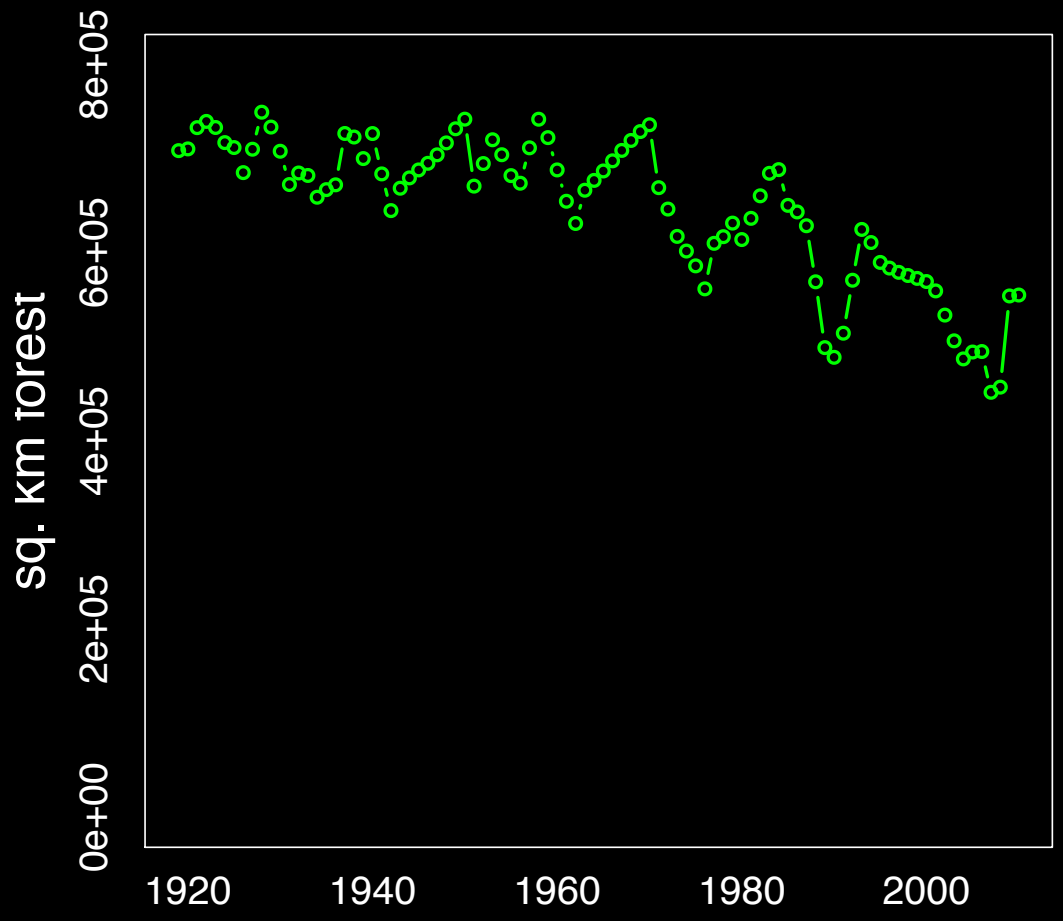
# Mechanistic modeling is process-driven…

- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns
- We start by building a model that uses explicit processes to recover the same outcomes ("states") as our data

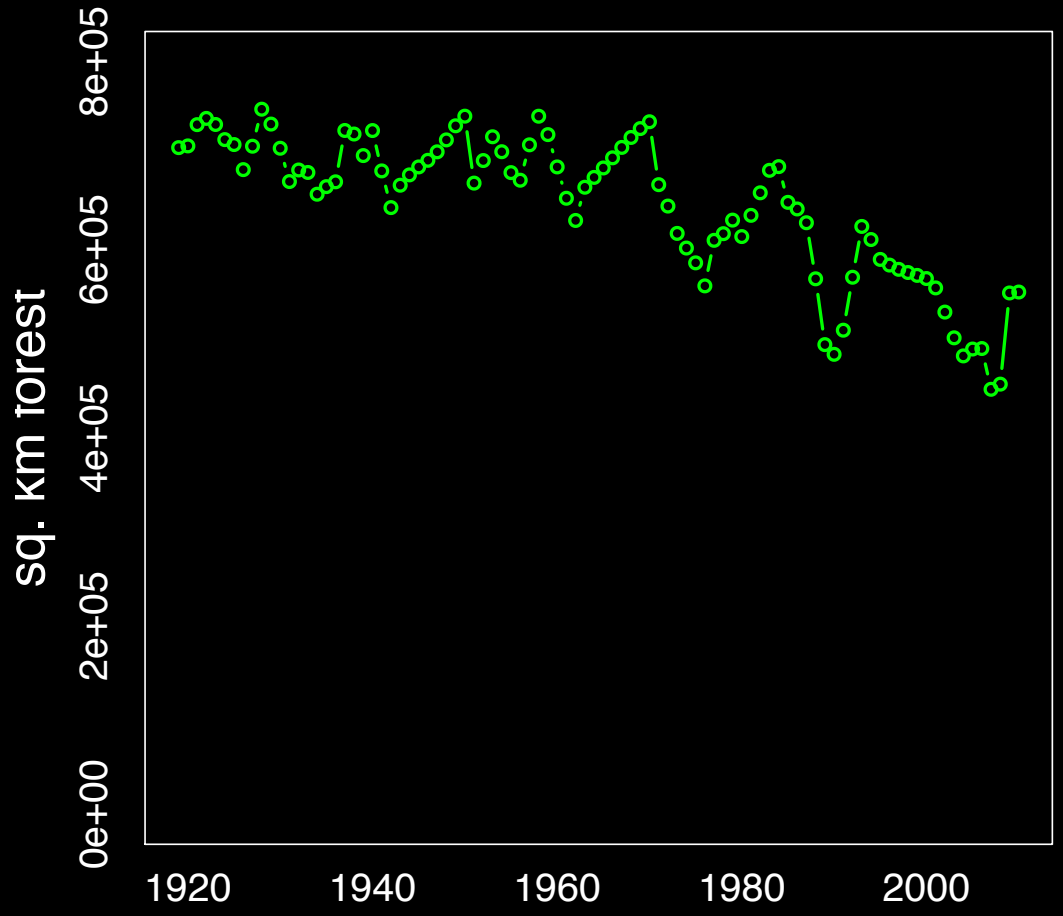
# Mechanistic modeling is process-driven…

- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns
- We start by building a model that uses explicit processes to recover the same outcomes ("states") as our data
- *What state variables are captured in our data?*

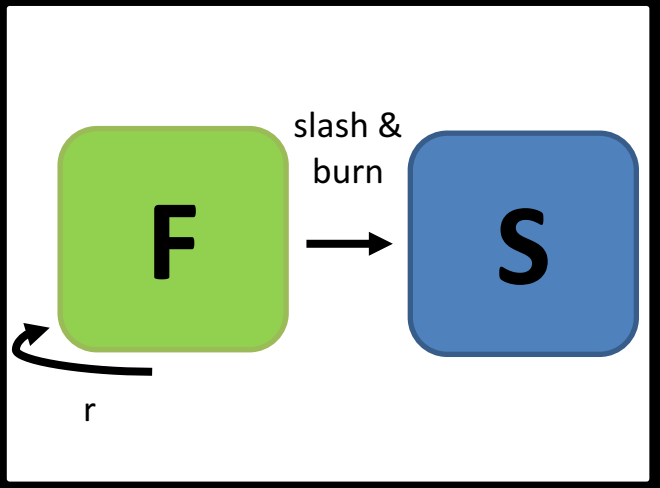
# These data give us forest over time…



# These data give us forest over time…




What *processes* contribute to the "forest" state in our system?



# How to fit a model to data

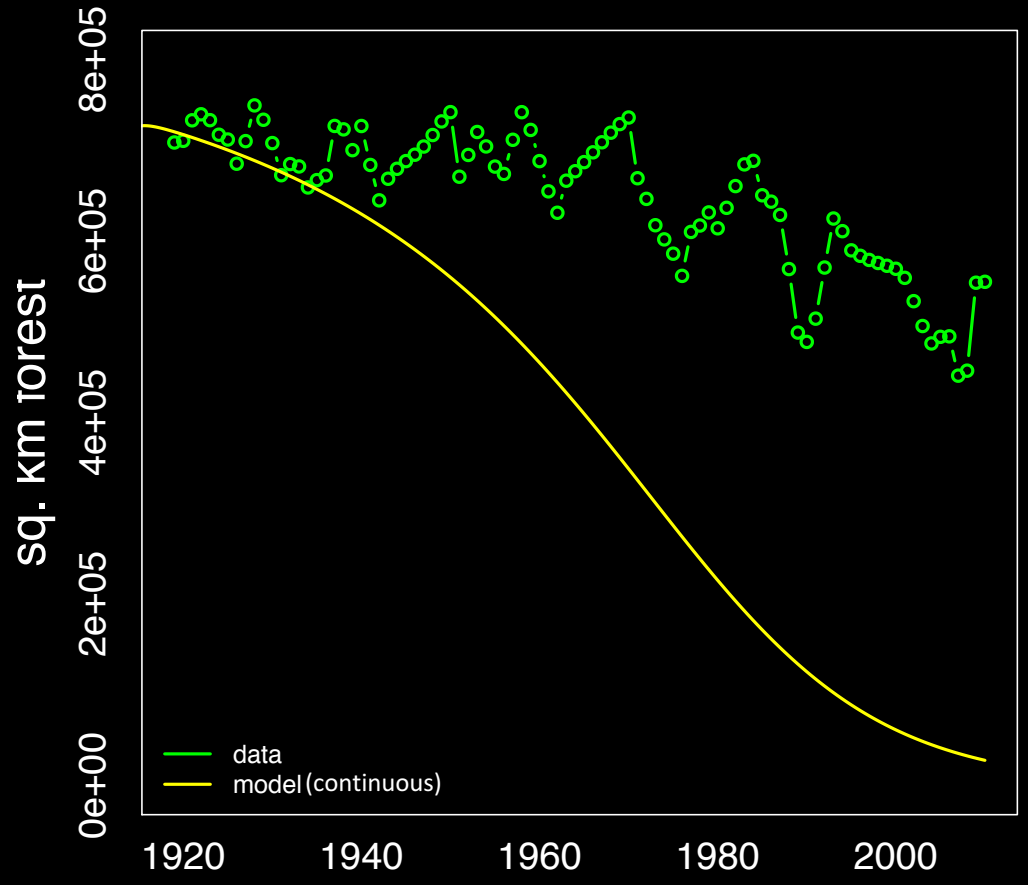
1. Build a model that uses explicit processes to recover the same states as the data.

   Ft+1 = Ft - slash*Ft*St

   Can be discrete or continuous . Let's try it!

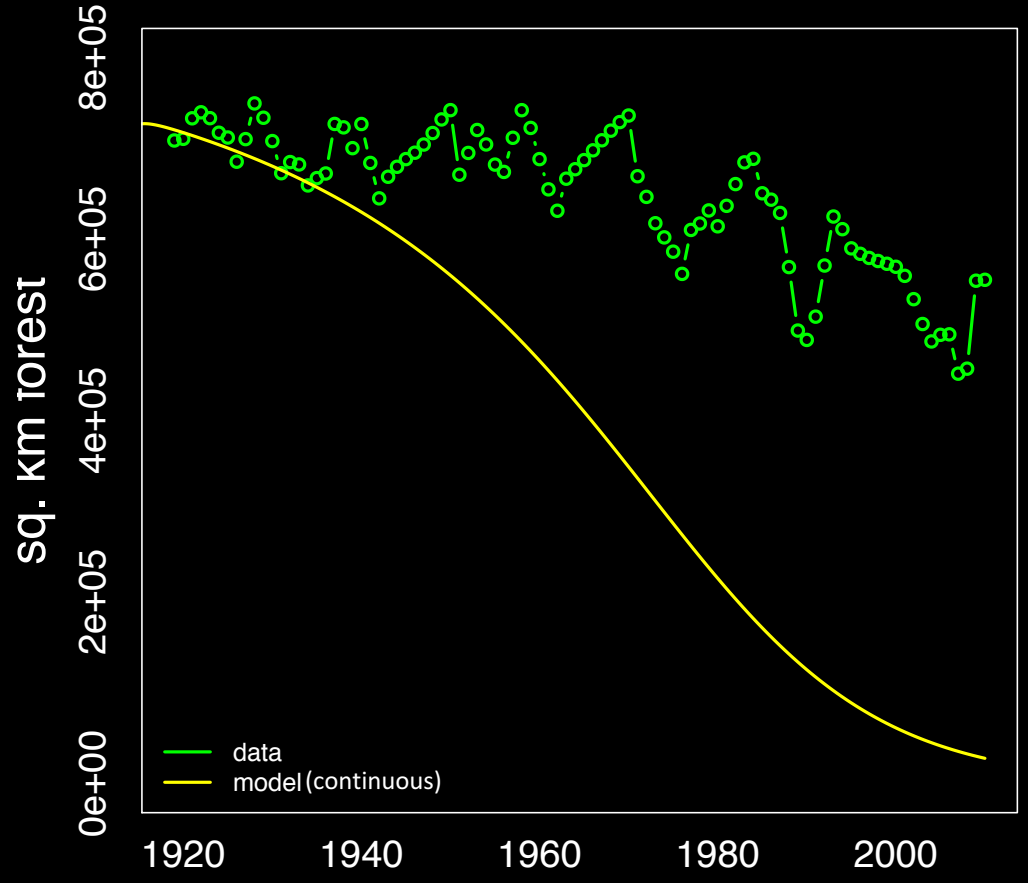
# How to fit a model to data

1. Build a mechanistic model that uses explicit processes to recover the same states as the data.



# How to fit a model to data

- Using least squares we ask, assuming our model is true, how likely are we to recover the observed data?

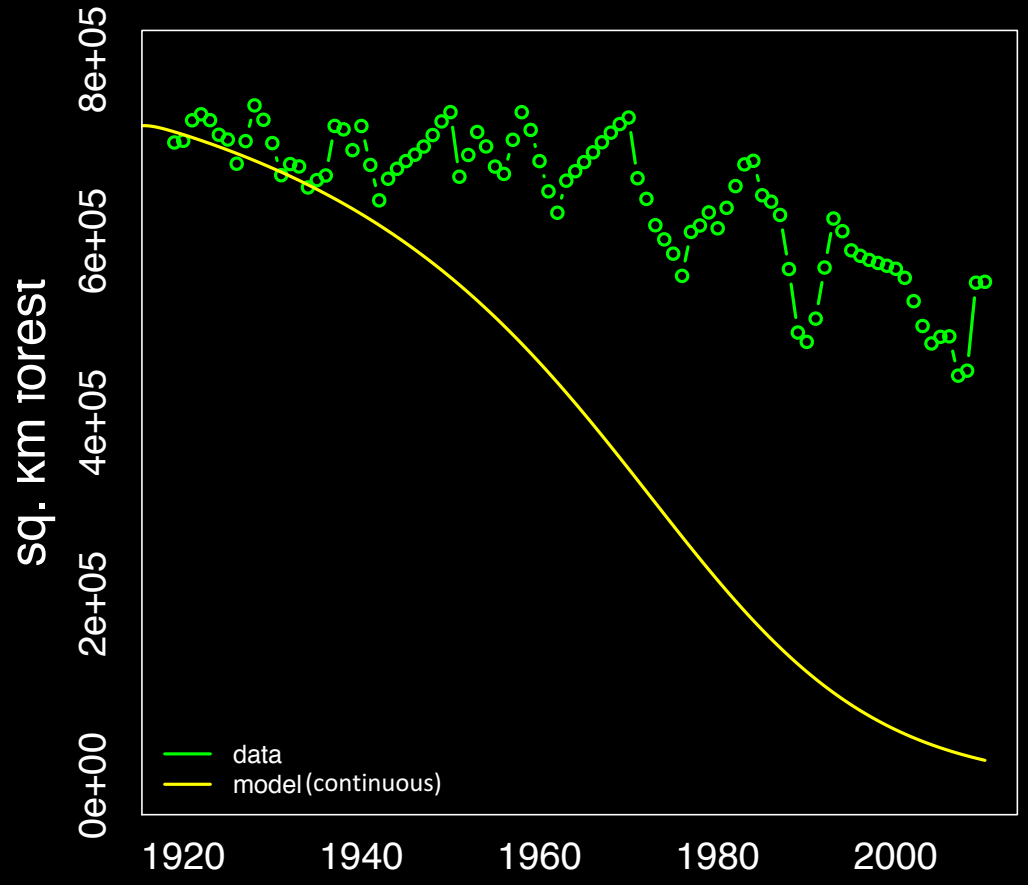



Model has the right trajectory but forest declines faster than in the data.

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

# How to fit a model to data

- Using least squares we ask, assuming our model is true, how likely are we to recover the observed data?




Model has the right trajectory but forest declines faster than in the data.

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

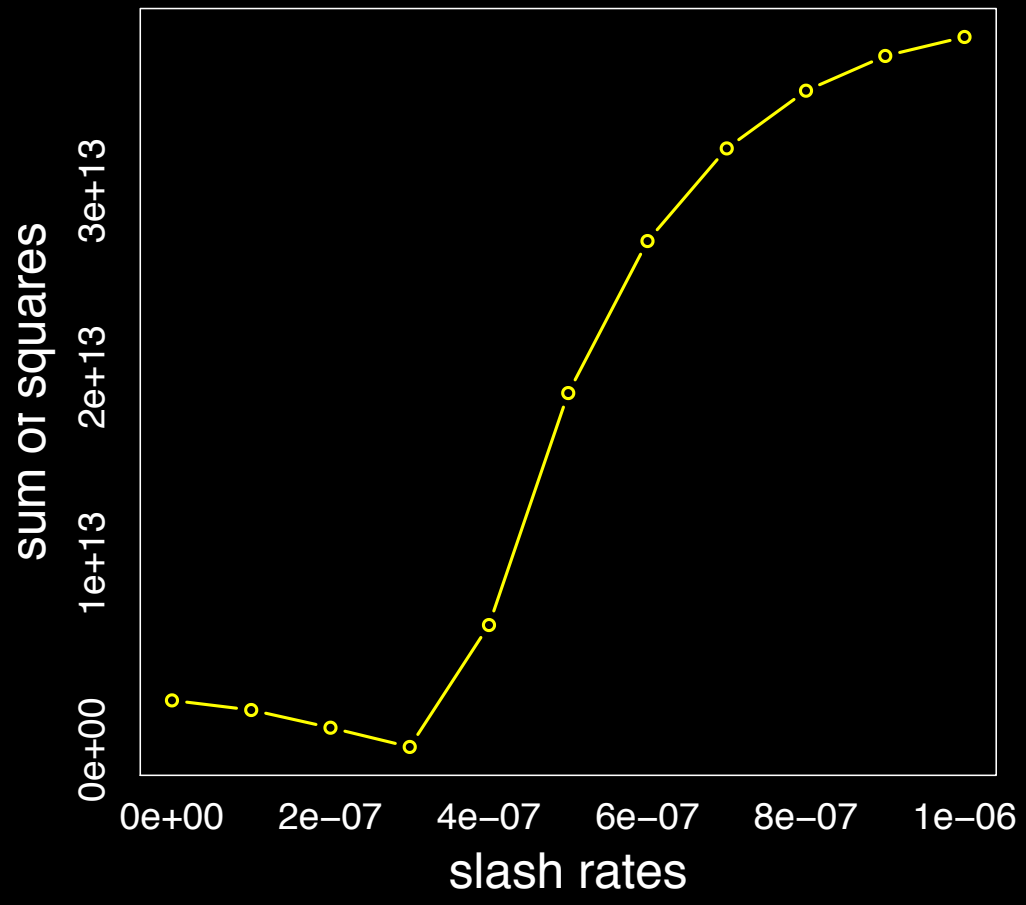
sum.sq = 3.914575e+13

Can we make that smaller?

# How to fit a model to data

2. Optimize the parameters behind the processes to make the model most likely to recover the data.

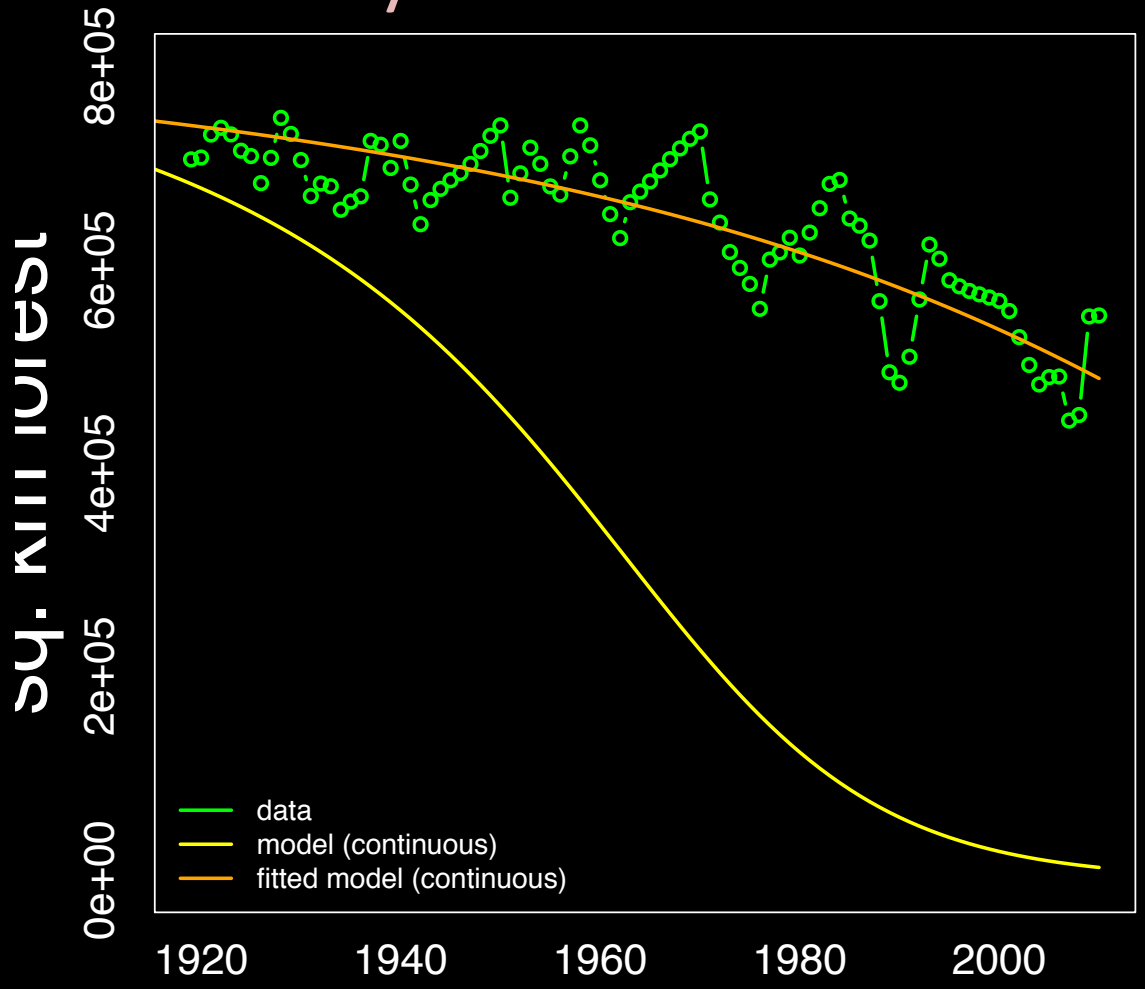
slash.list[which.min(sm.sq)] =
3e-7



# How to fit a model to data

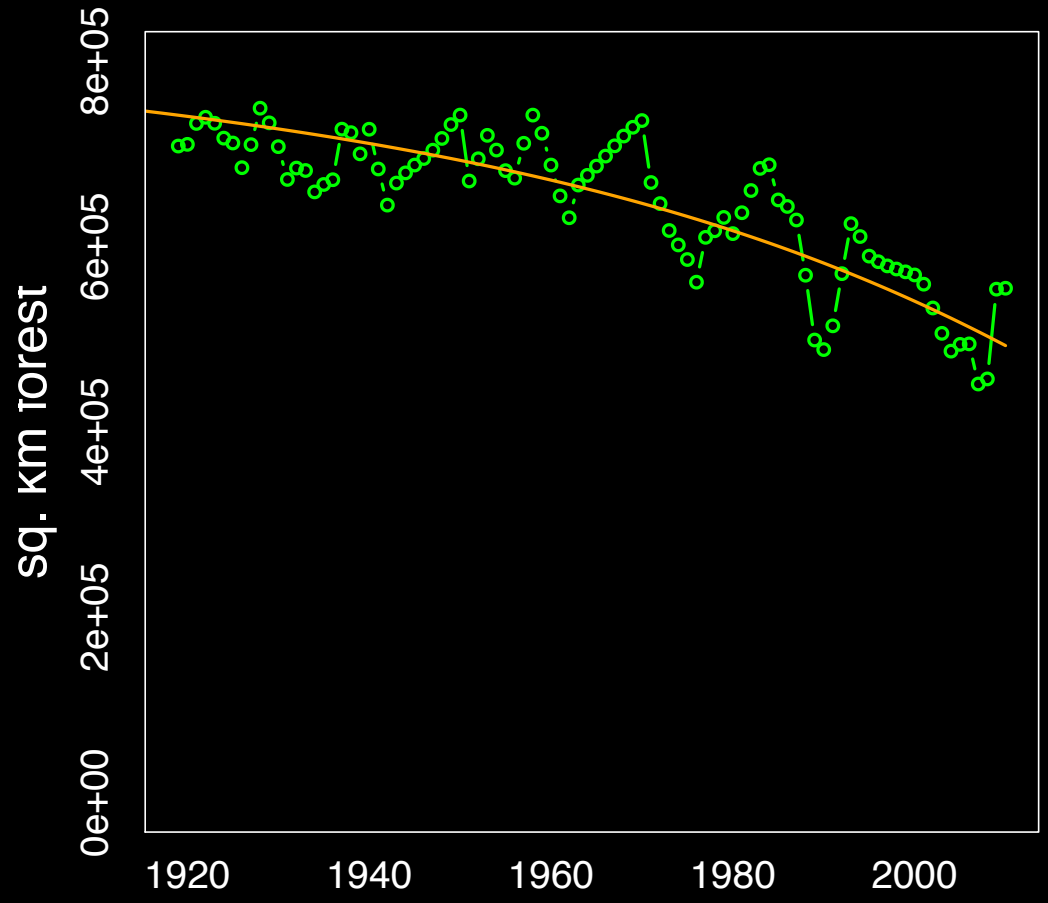
2. Optimize the parameters behind the processes to make the model most likely to recover the data.

New slash fits better!



# How to fit a model to data

4. If need be, restructure your model to better match your data.




We are good!

# Mechanistic modeling is process-driven…

- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns
- We start by building a model that uses explicit processes to recover the same outcomes ("states") as our data

# Mechanistic modeling is process-driven…

- Estimate: time series of state variables/parameters of interest ➔ DATA
- Inference: Build model to recapture data. Fit to optimize parameters and "infer" the process underlying the data
- Model assessment: Assess plausibility or model comparison
- End goal: explain observed patterns or predict