Bayesian Inference

Saras Windecker
James Kelleher, Kevin Newman, David Uribe

Code of conduct

Wifi

Shared doc

Github

Question: why are you interested in learning about Bayesian inference?

Why Bayes

Flexibility in building models > can test any hypothesis, not just the null

Principled way of building your model > think about what you know & the implications of what you think you know

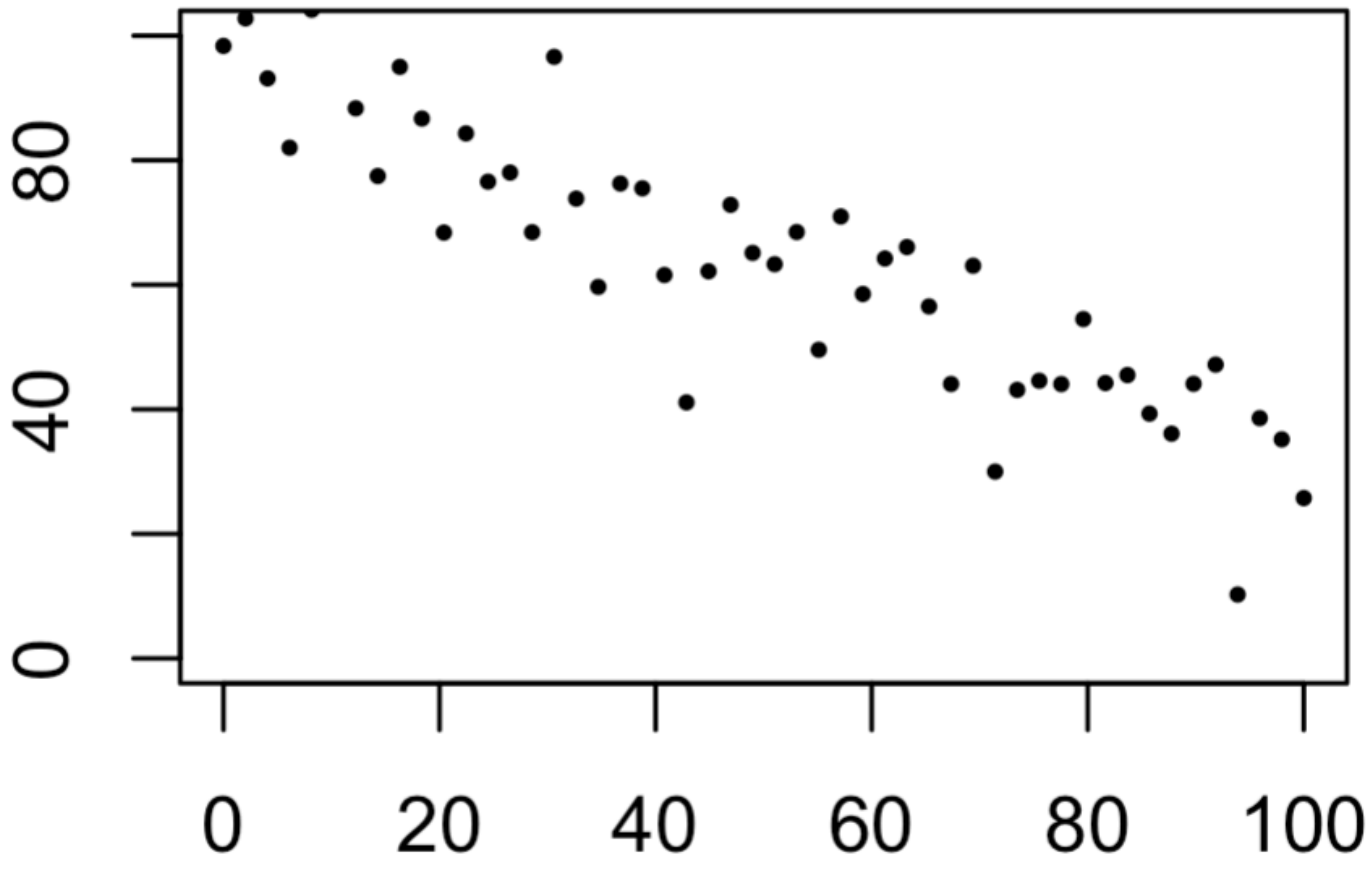
Your inferences will fail often, giving you a way to diagnose the issue

Scales up, same tools and workflow for range of model processes > from t test to bayesian network analysis

More subjective > you make more choices.

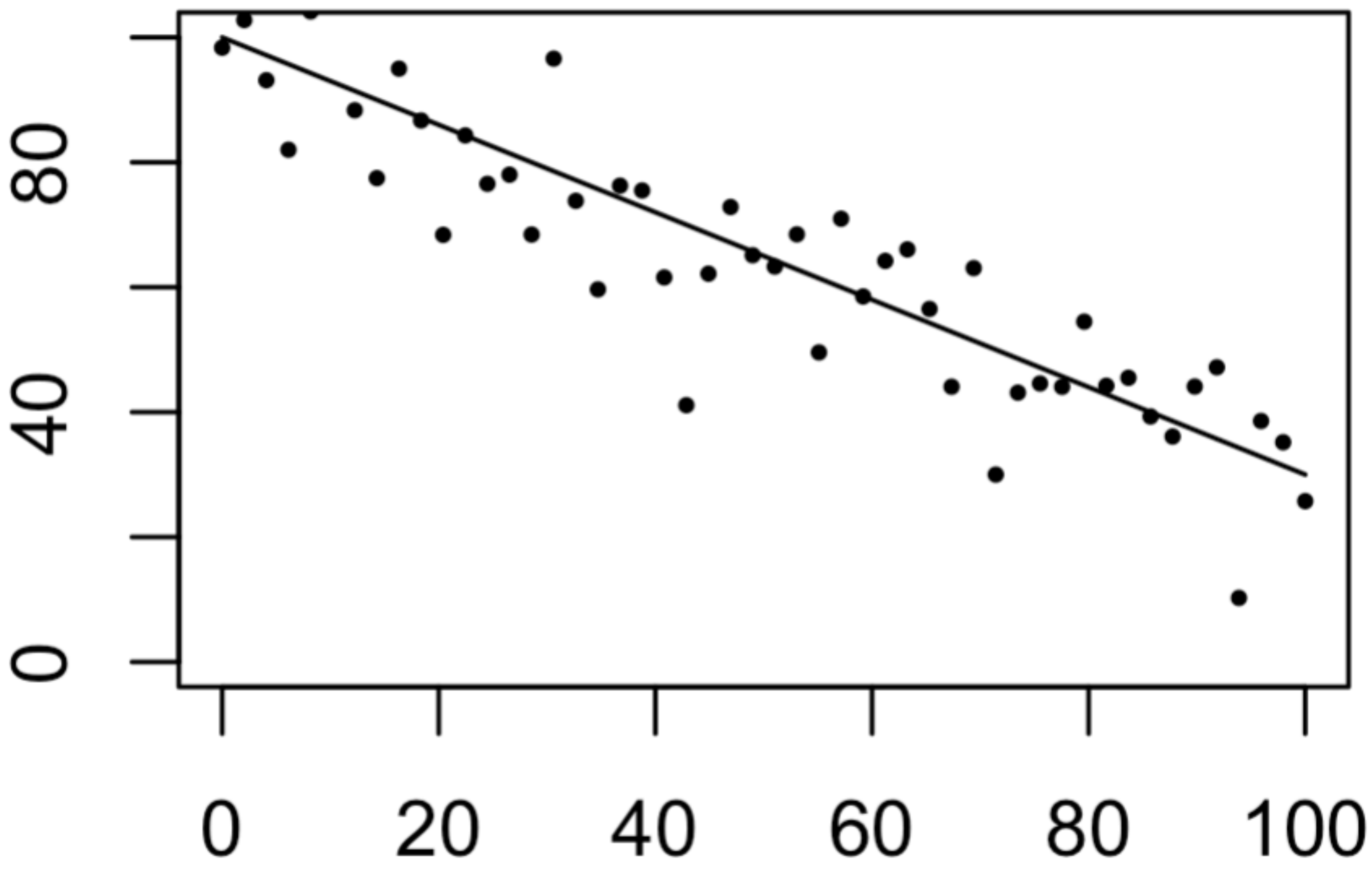
What is *not* a reason to choose Bayes?

There are no
Frequentist models or
Bayesian models



$$y = a + bx + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$



$$y = a + bx + \epsilon$$

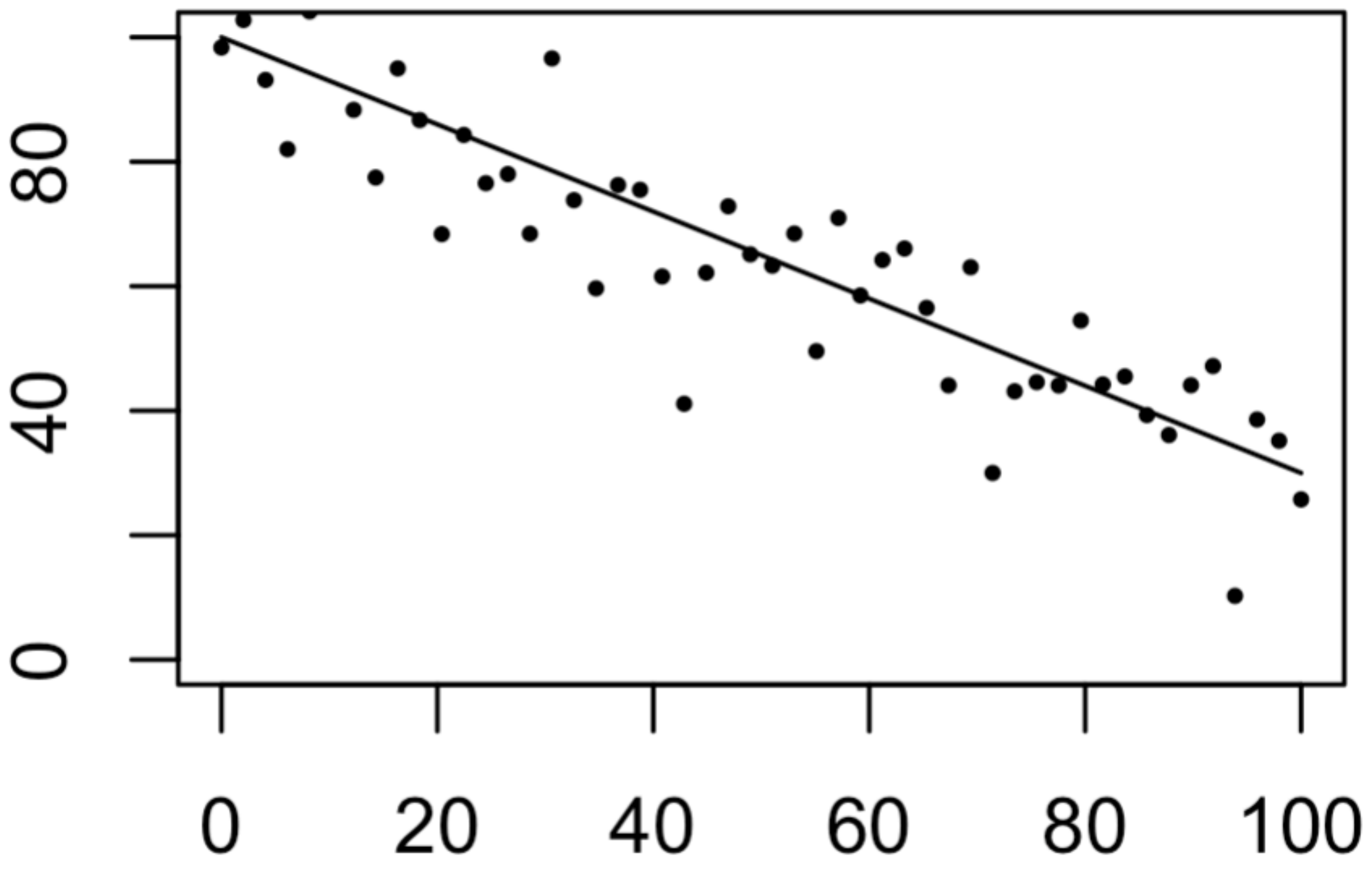
$$\epsilon \sim N(0, \sigma^2)$$

MLE

$$b = -0.67$$

Posterior mean

$$b = -0.67$$



Statistical models exist independent of the method of estimating parameters

Statistical models exist independent of the method of estimating parameters

There are no Frequentist models or Bayesian models

Statistical models exist independent of the method of estimating parameters

There are no Frequentist models or Bayesian models

May choose to analyse a model in a Bayesian way

I. Thinking about models

Exercise 1: Research workflow

Conceptual model

Prepare question

Experimental design

Data collection

Write and build model

Create model outputs

Conceptual model

Prepare question

Experimental design

Data collection

Write and build model

Create model outputs

== Using a model to understand our data.

Conceptual model

Prepare question

Experimental design

Data collection

Write and build model

Create model outputs

Conceptual model

Prepare question / Write model

Experimental design

Data collection

Build model

Create model outputs

Conceptual model

Prepare question / Write model

Experimental design

Data collection

Build model

Create model outputs

We want to collect data to test our model.

Using a model to understand our data is asking:

"What is the probability of this data, given a certain model?"

Using a model to understand our data is asking:

"What is the probability of this data, given a certain model?"

Collecting data to test our model is asking:

"What is the probability of this model, given observed data?"

A well-specified research question defines a statistical model.

AND

A statistical model underpins a research question.

Thinking about models

Model

Model-defined data

Model

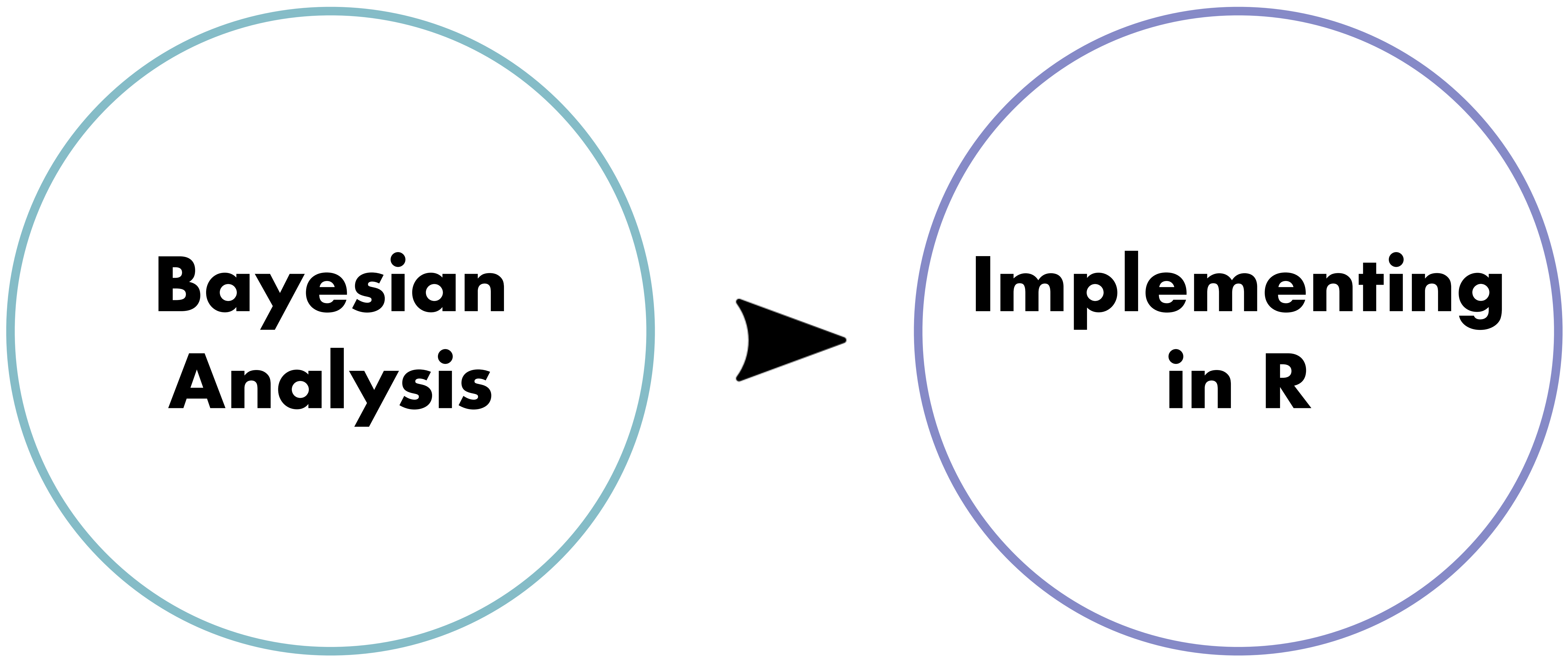
Experiment

Model-defined data

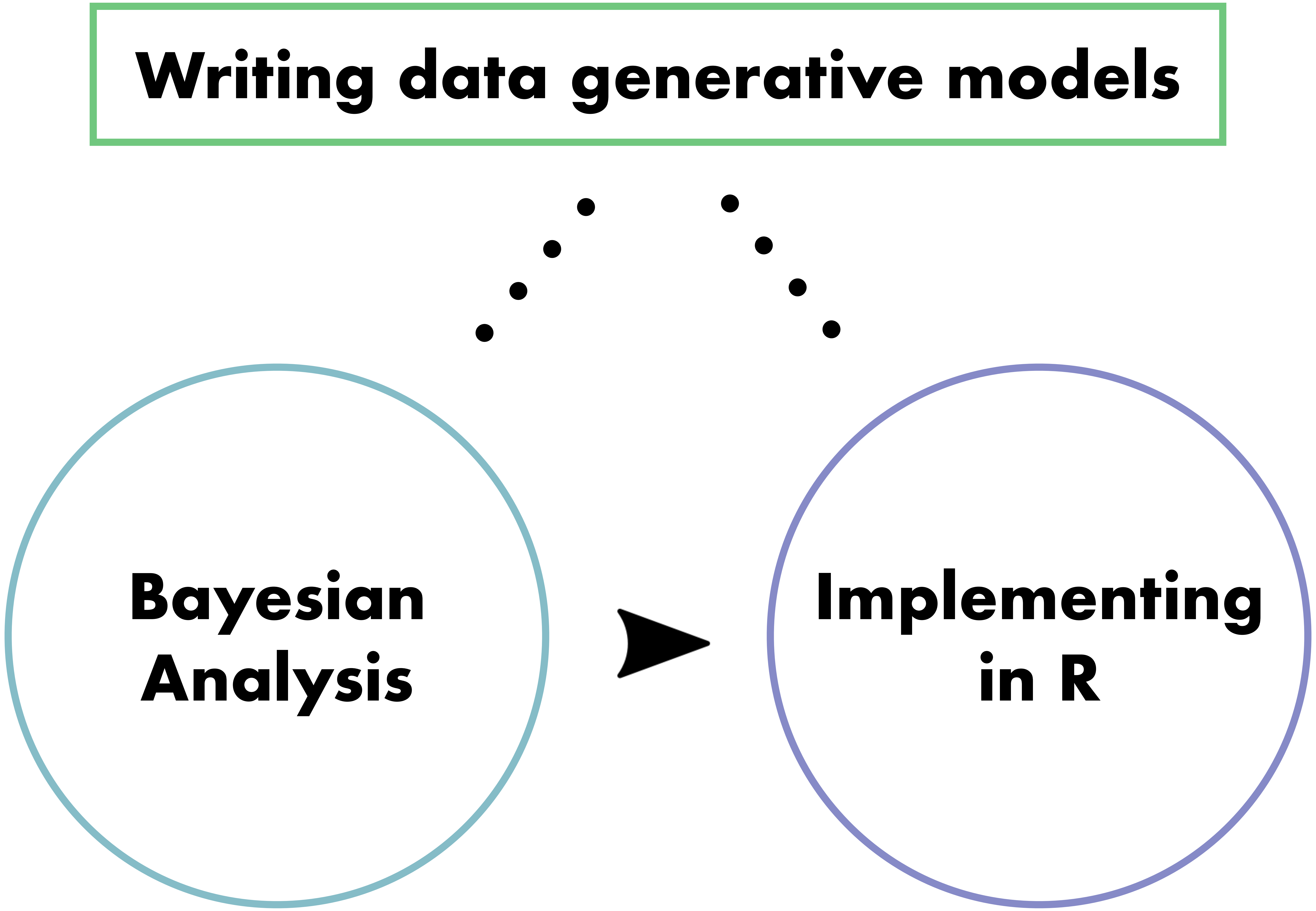
?=

Collected data

Workshop aims



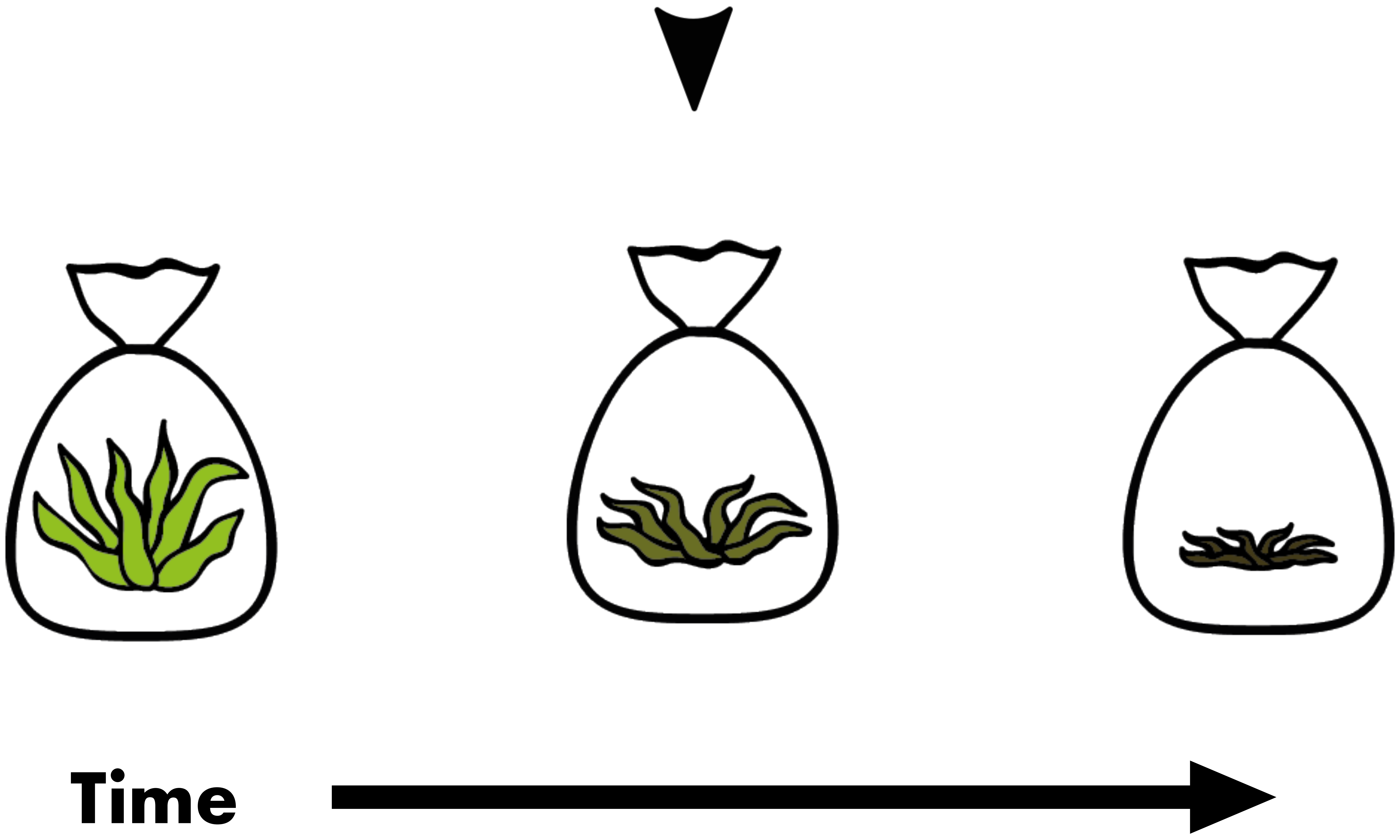
Workshop aims



II. Writing data generative models & simulating data

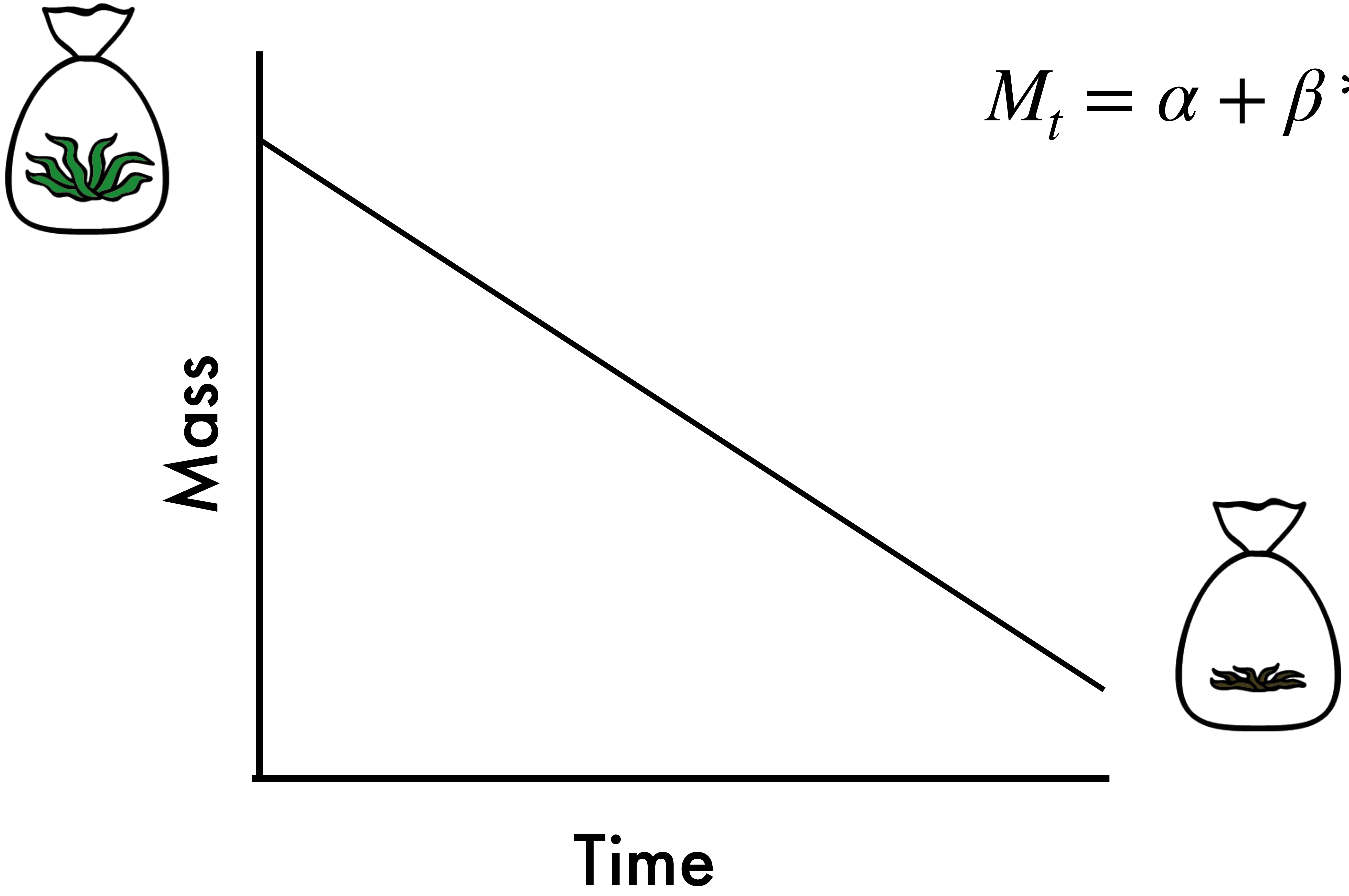
What is species' mass loss over time?

What is species' mass loss over time?



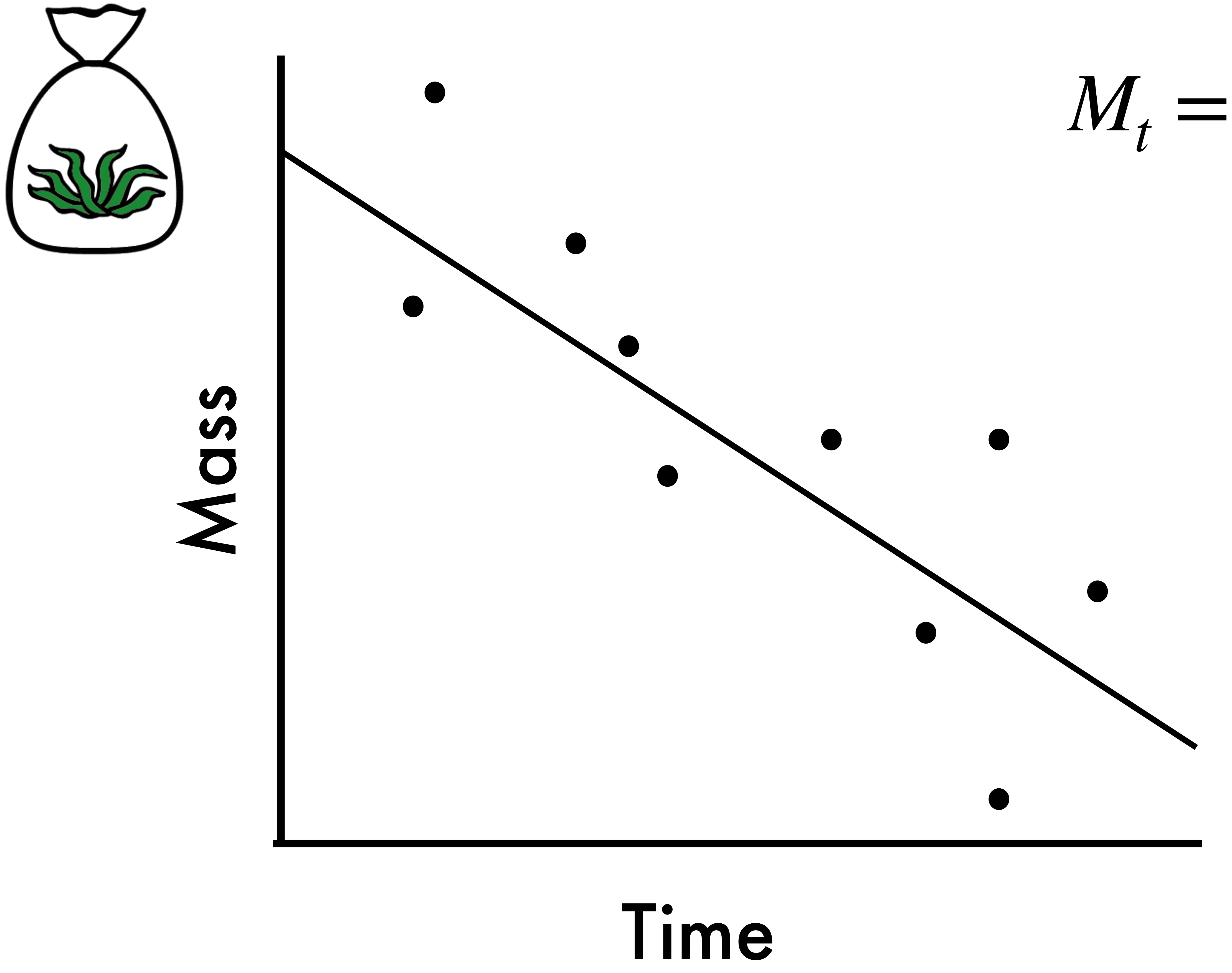
What is species' mass loss over time?

$$M_t = \alpha + \beta * time$$

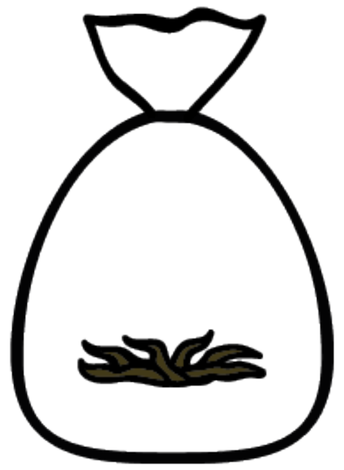


Writing data generative models

What is species' mass loss over time?

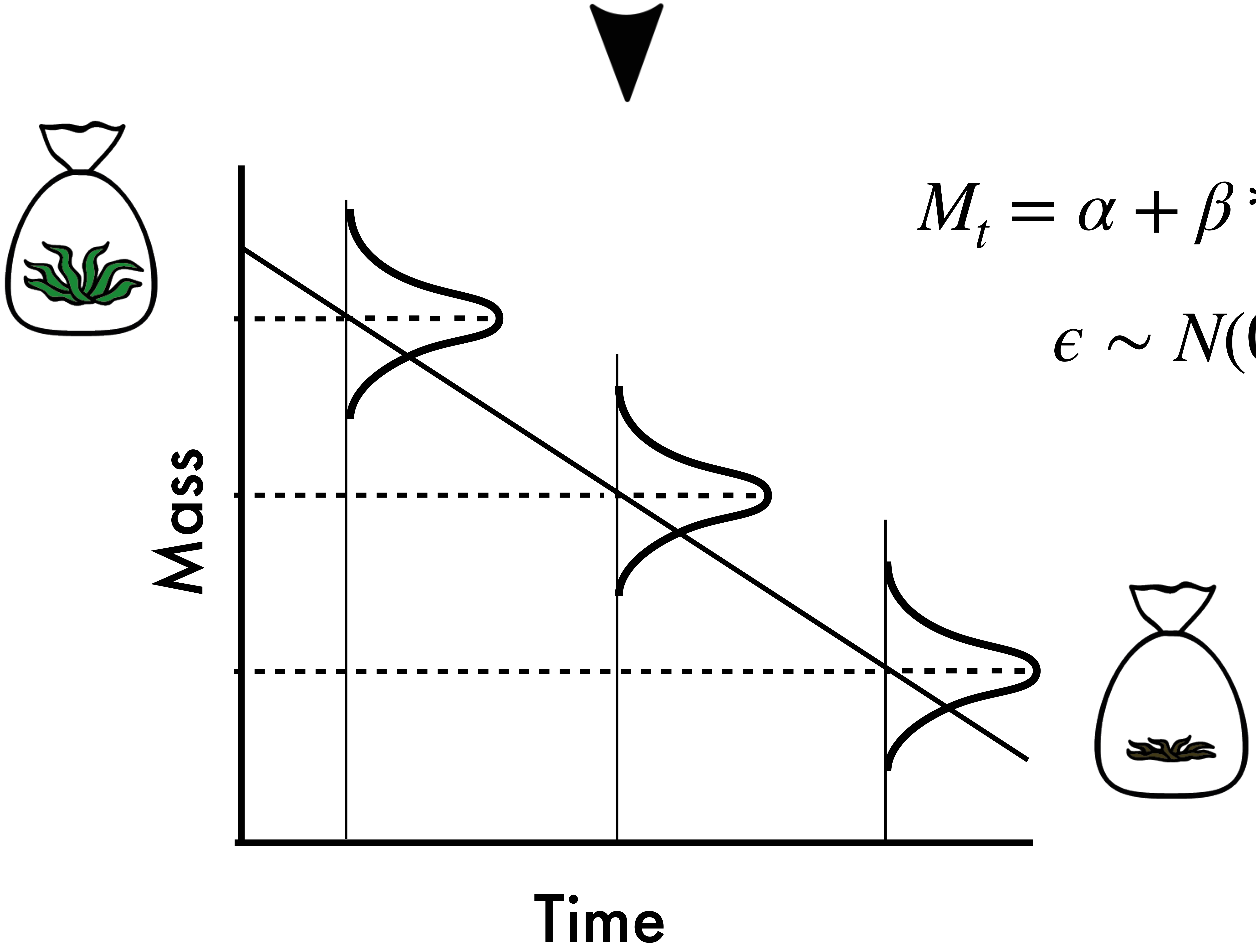


$$M_t = \alpha + \beta * time + \epsilon$$



Writing data generative models

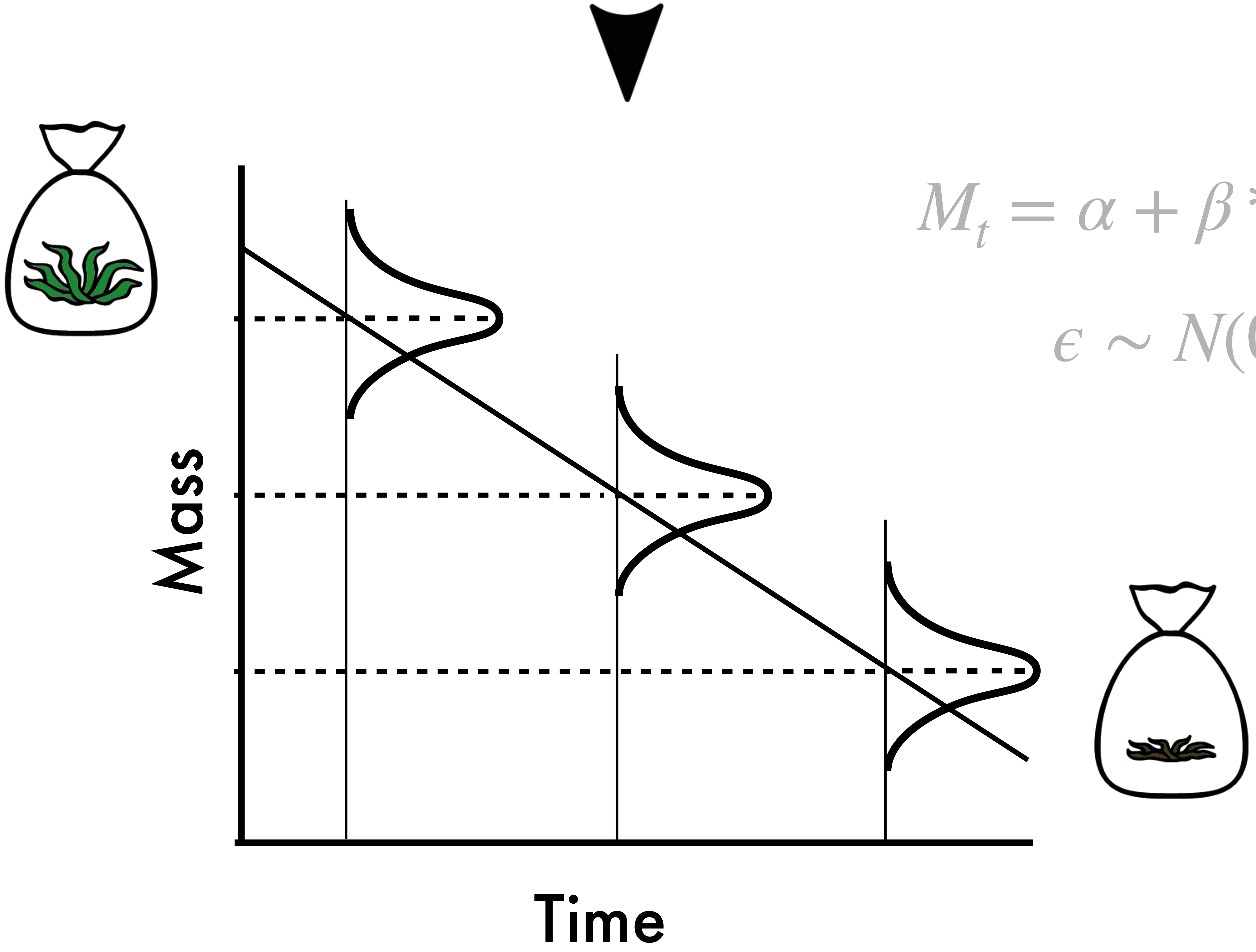
What is species' mass loss over time?



$$M_t = \alpha + \beta * time + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

What is species' mass loss over time?



$$M_t = \alpha + \beta * time + \epsilon$$

$$\epsilon \sim N(0,\sigma^2)$$

$$M_t \sim N(\mu_t, \sigma^2)$$

$$\mu_t = \alpha + \beta * time$$

Exercise 2: Write a data generative linear model

Model

Model-defined data

Model

Model-defined data

$$M_t \sim N(\mu_t, \sigma^2)$$

$$\mu_t = \alpha + \beta * time$$

R practical: simulating data

R/1-simulating_data.R

III. Thinking about probability

Using a model to understand our data is asking:

"What is the probability of this data, given a certain model?"

Using a model to understand our data is asking:

"What is the probability of this data, given a certain model?"

p(data | parameter)

Using a model to understand our data is asking:

"What is the probability of this data, given a certain model?"

$p(\text{data} \mid \text{parameter})$

Collecting data to test our model is asking:

"What is the probability of this model, given observed data?"

Using a model to understand our data is asking:

"What is the probability of this data, given a certain model?"

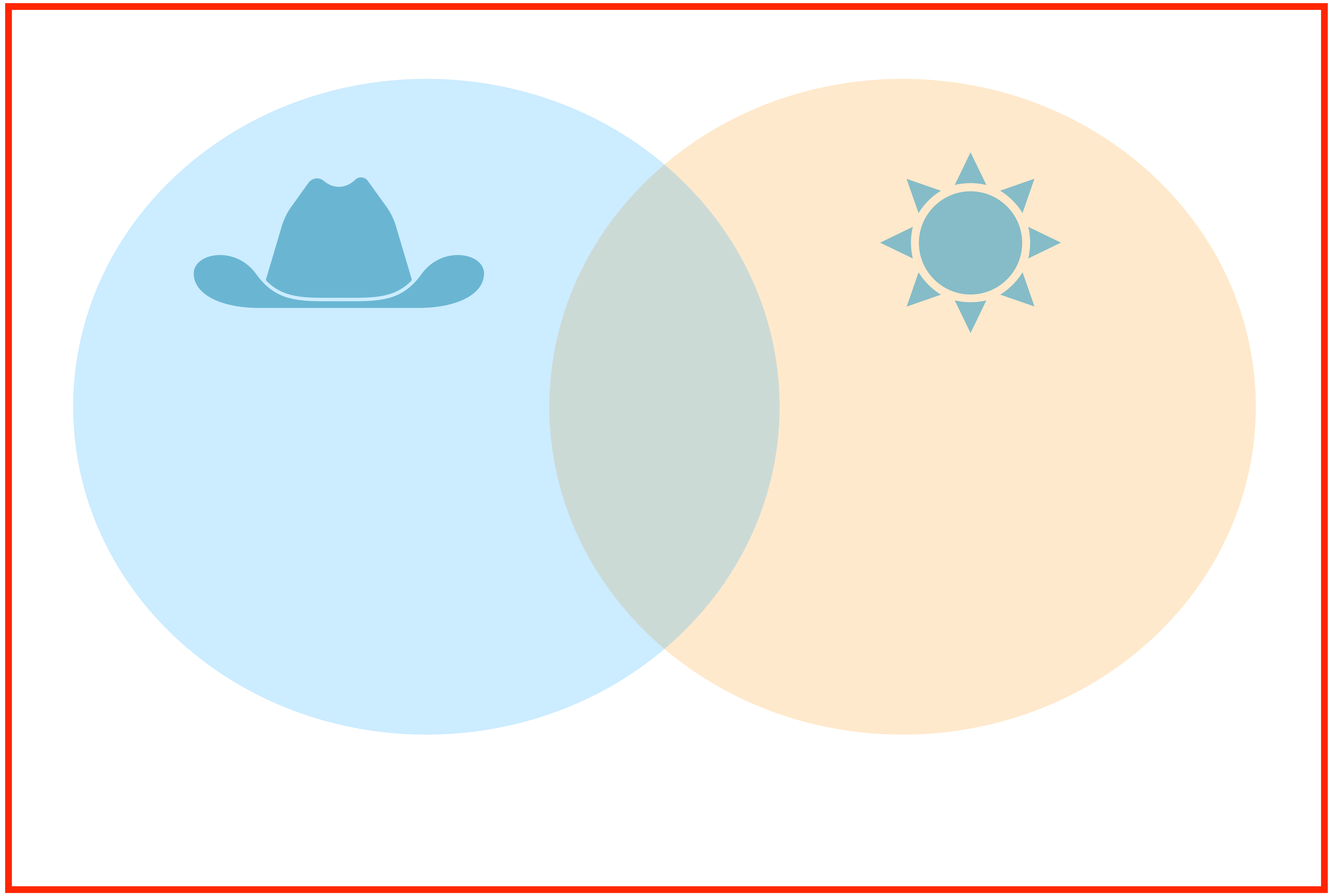
$p(\text{data} \mid \text{parameter})$

Collecting data to test our model is asking:

"What is the probability of this model, given observed data?"

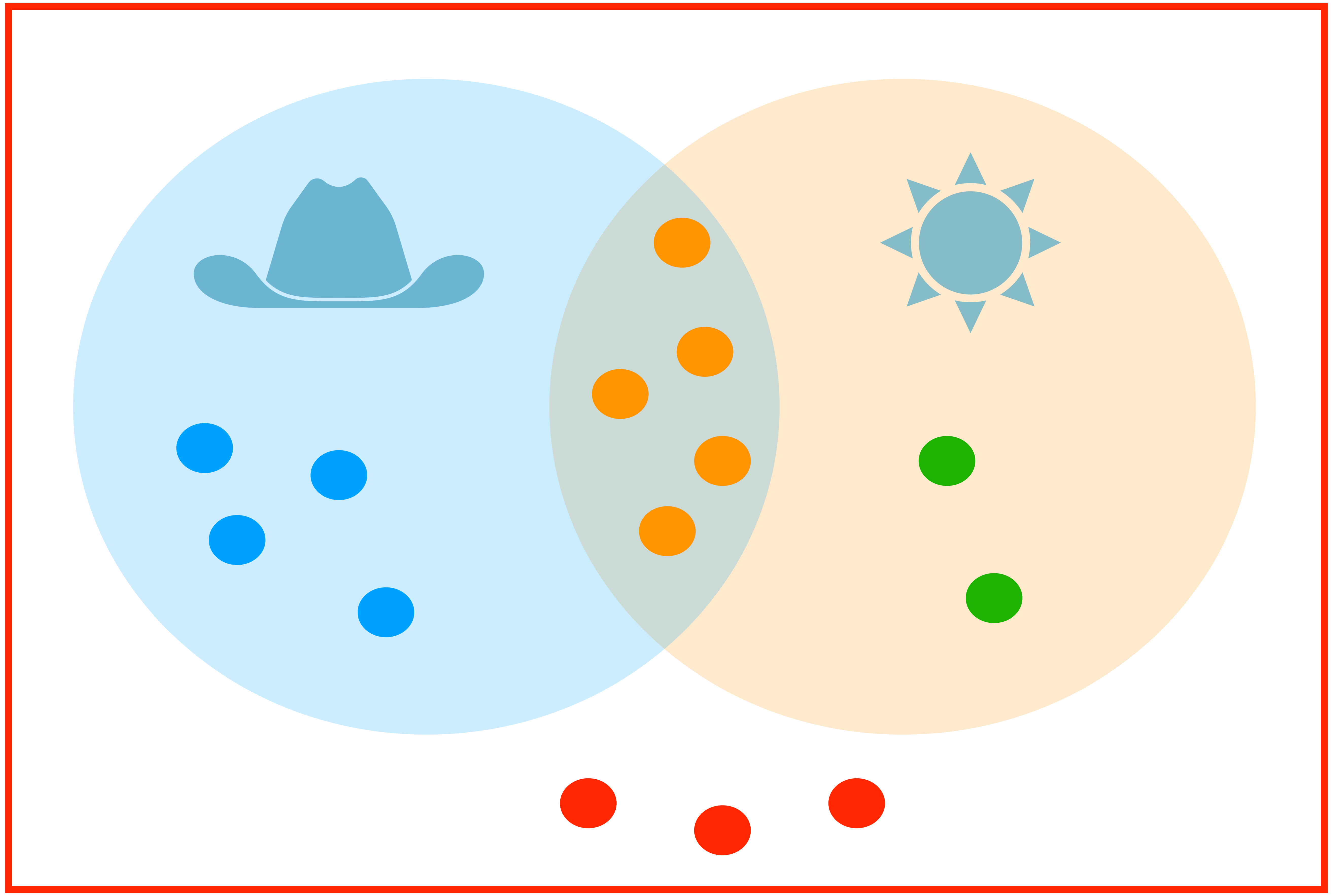
$p(\text{parameter} \mid \text{data})$

Thinking about probability

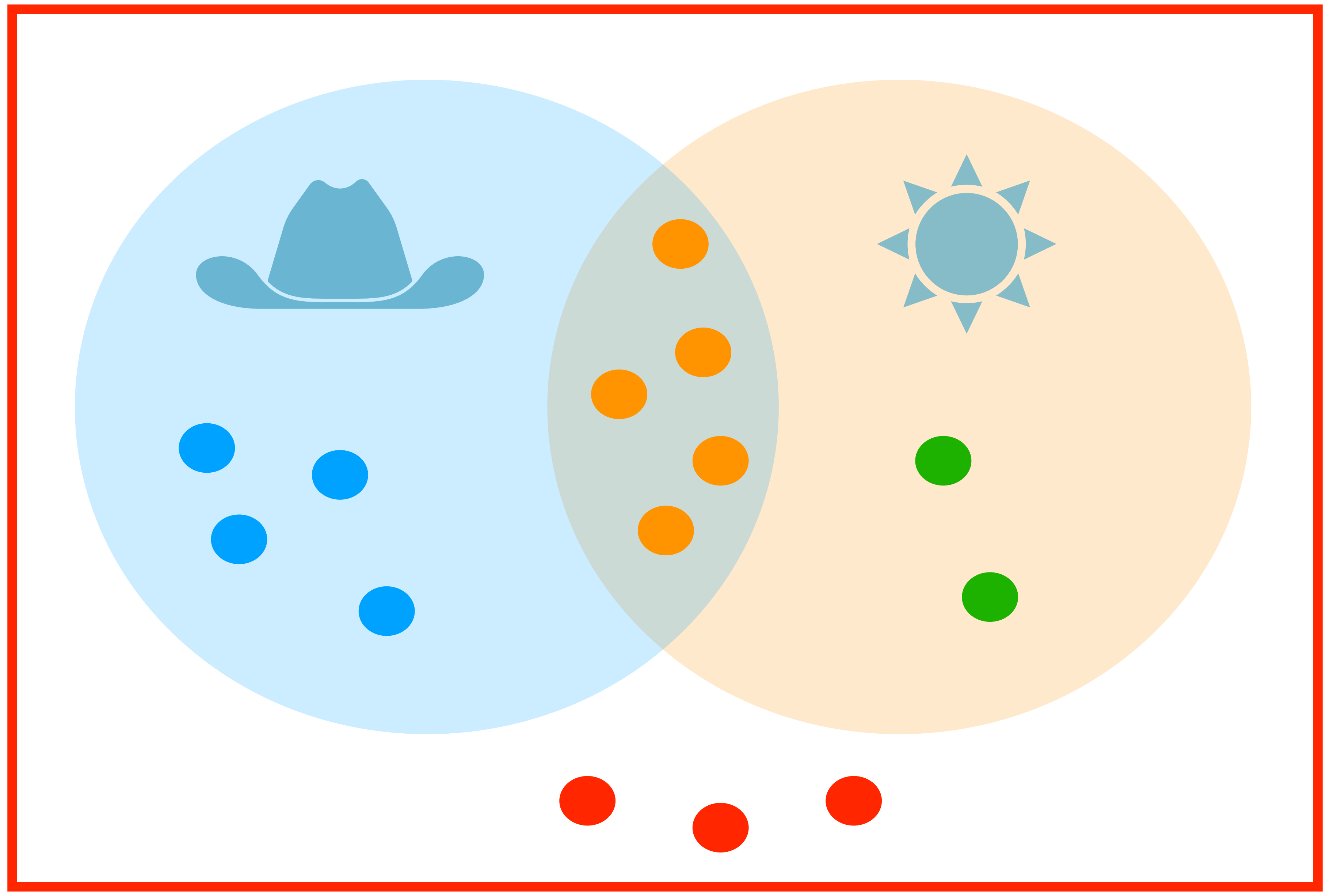


Check out: https://www.youtube.com/watch?v=9wCnvr7Xw4E&ab_channel=StatQuestwithJoshStarmer

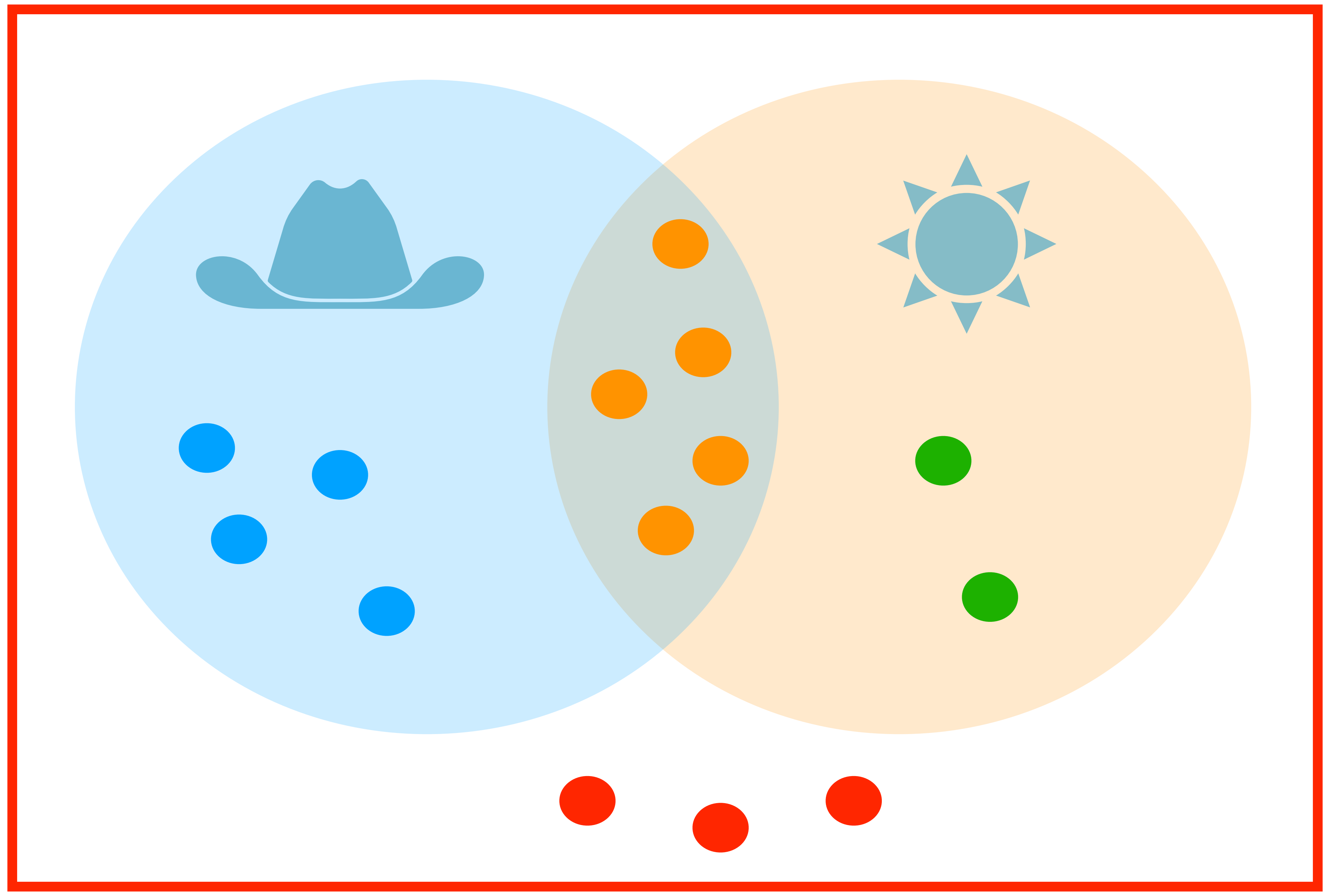
Thinking about probability



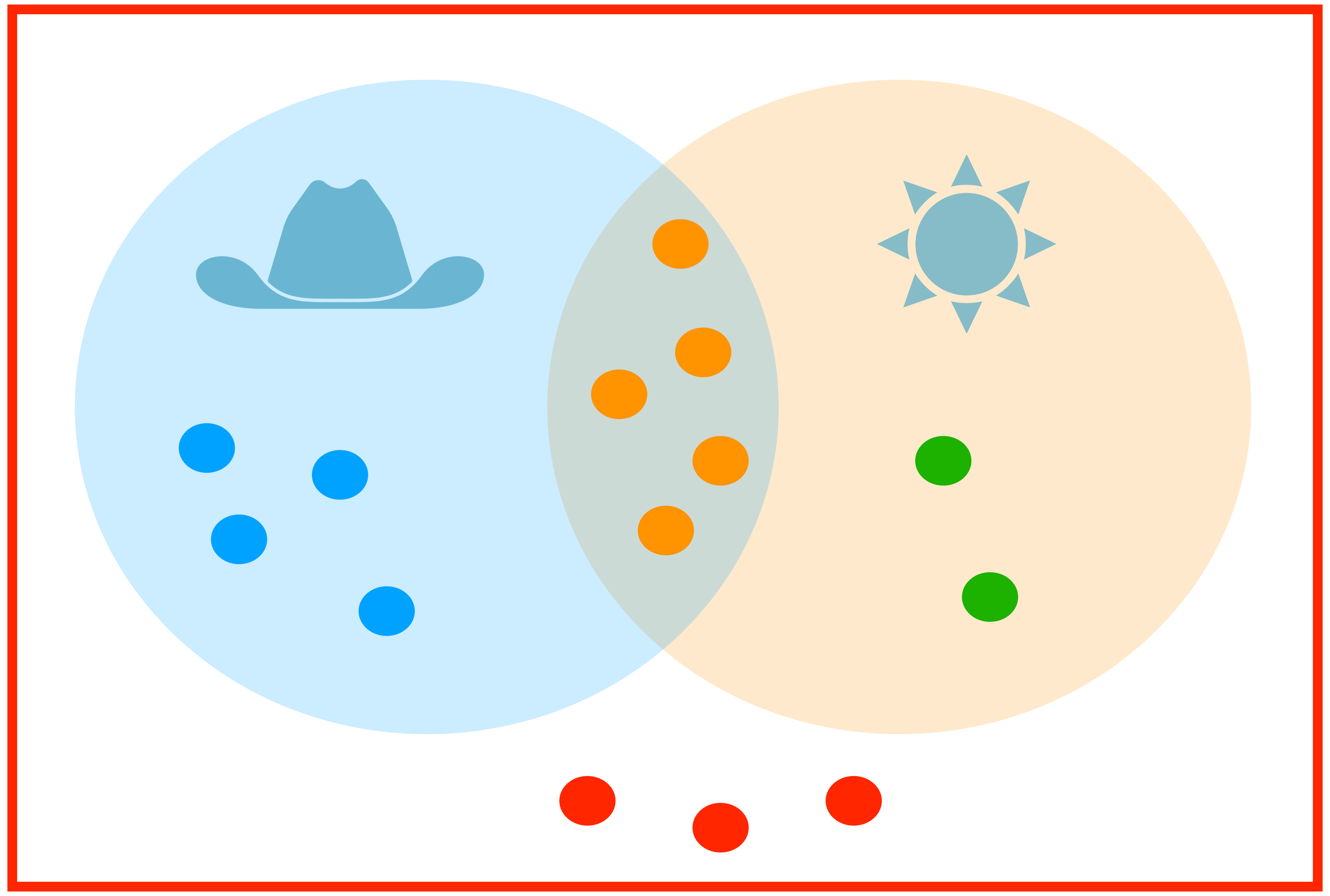
Thinking about probability



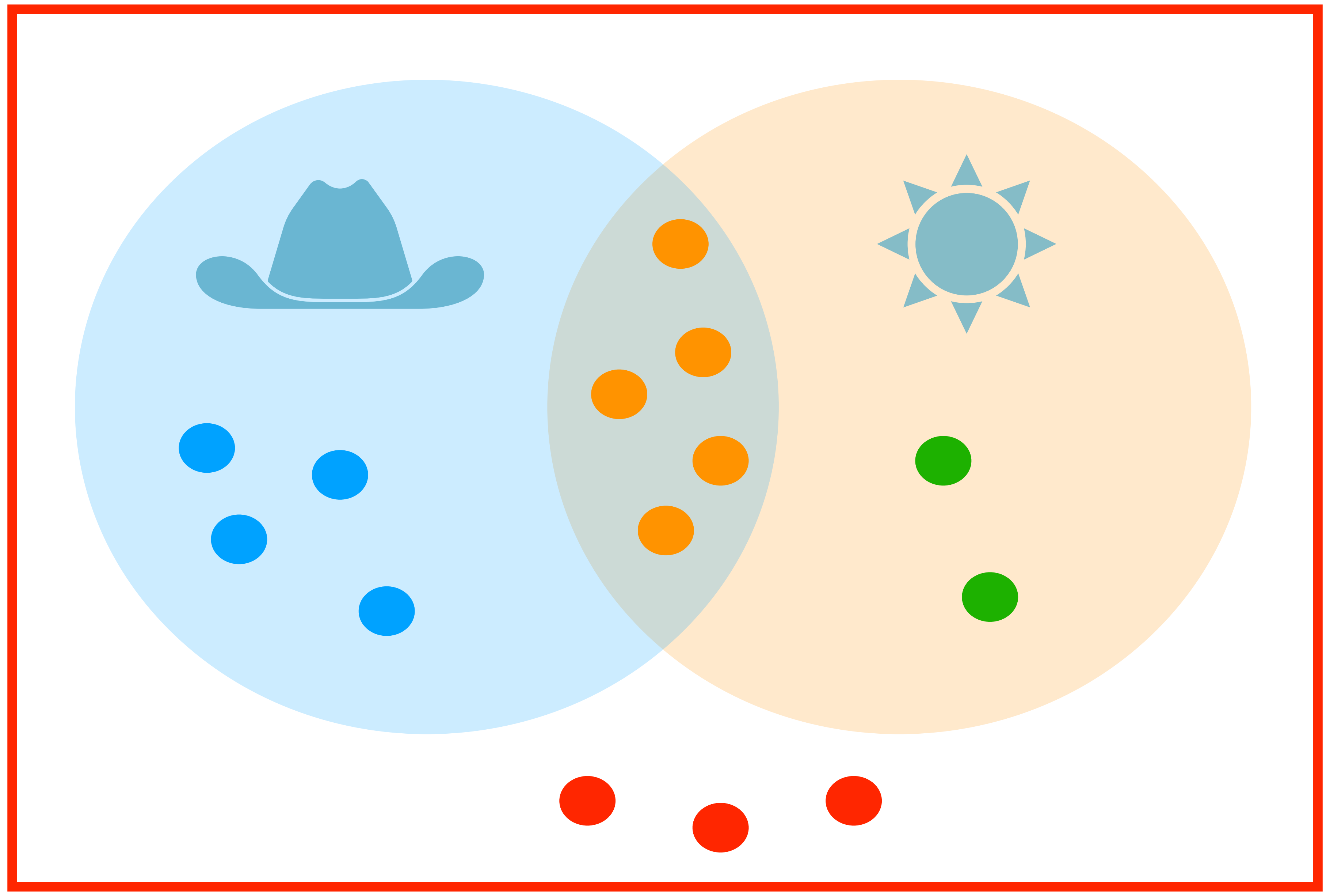
	No sun	Sun	Total
Wore hat	4/14		
Did not wear hat			
Total			



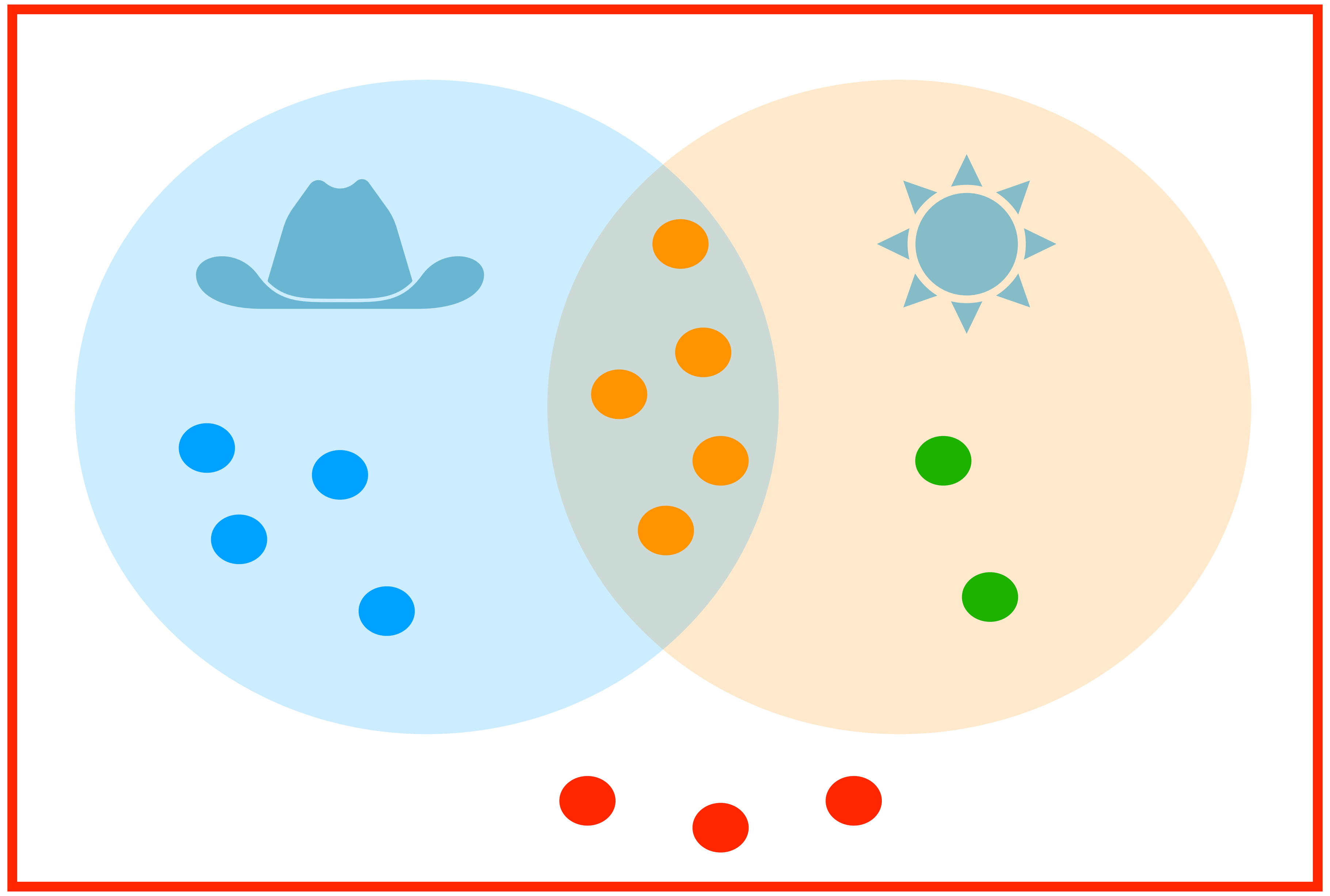
	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat			
Total			



	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14		
Total			

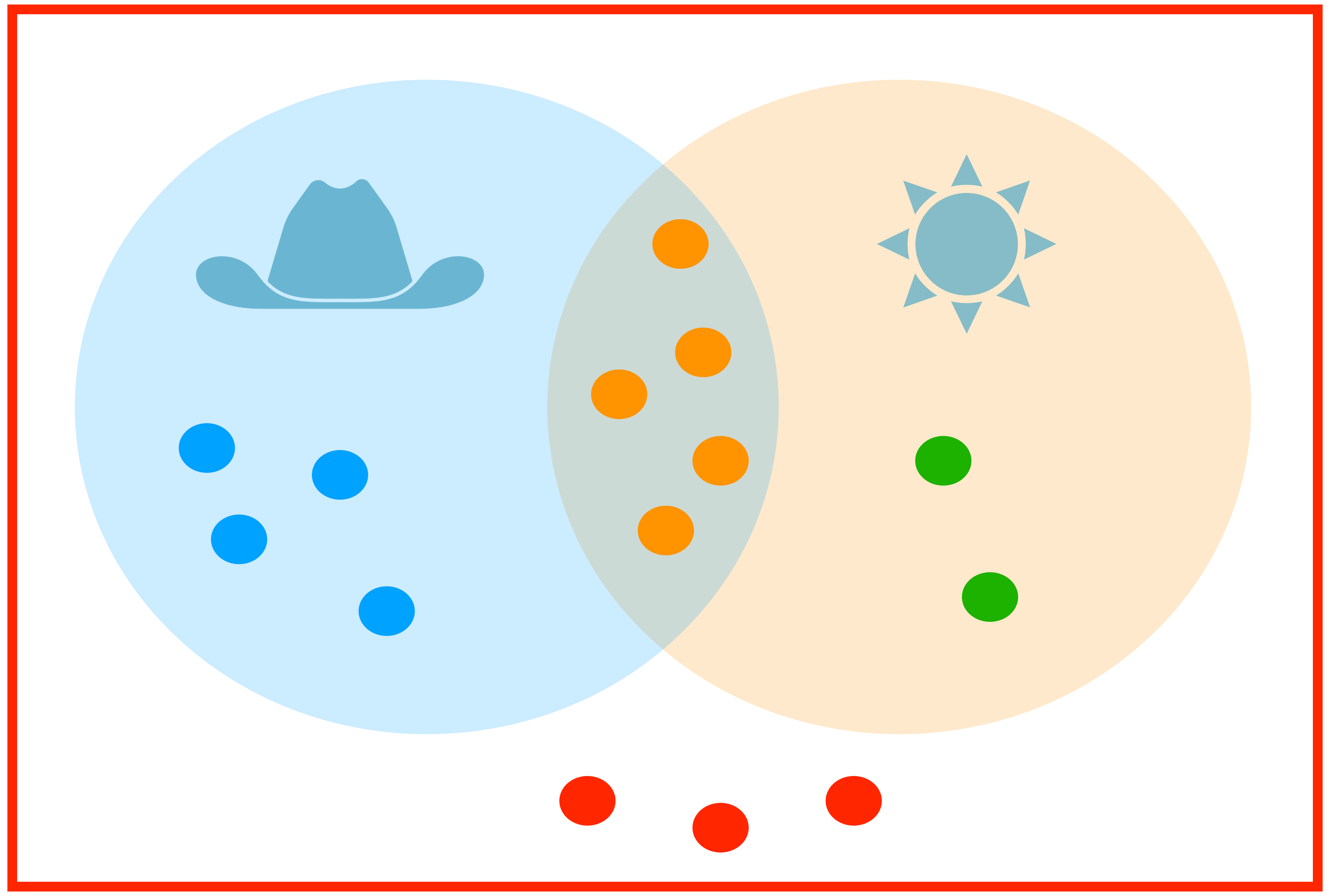


	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	



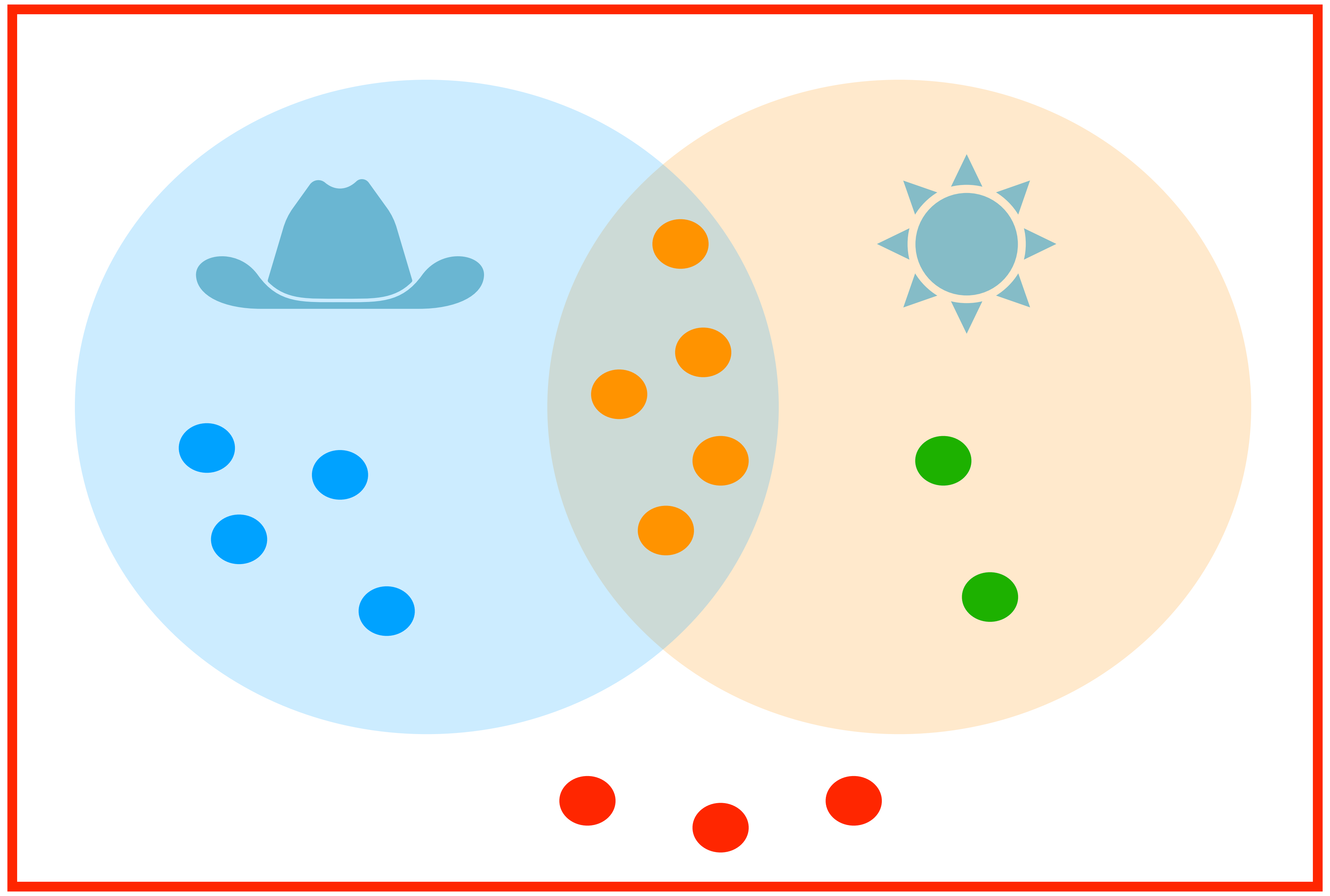
	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun} \;\&\; \text{hat} \mid \text{hat}) =$$



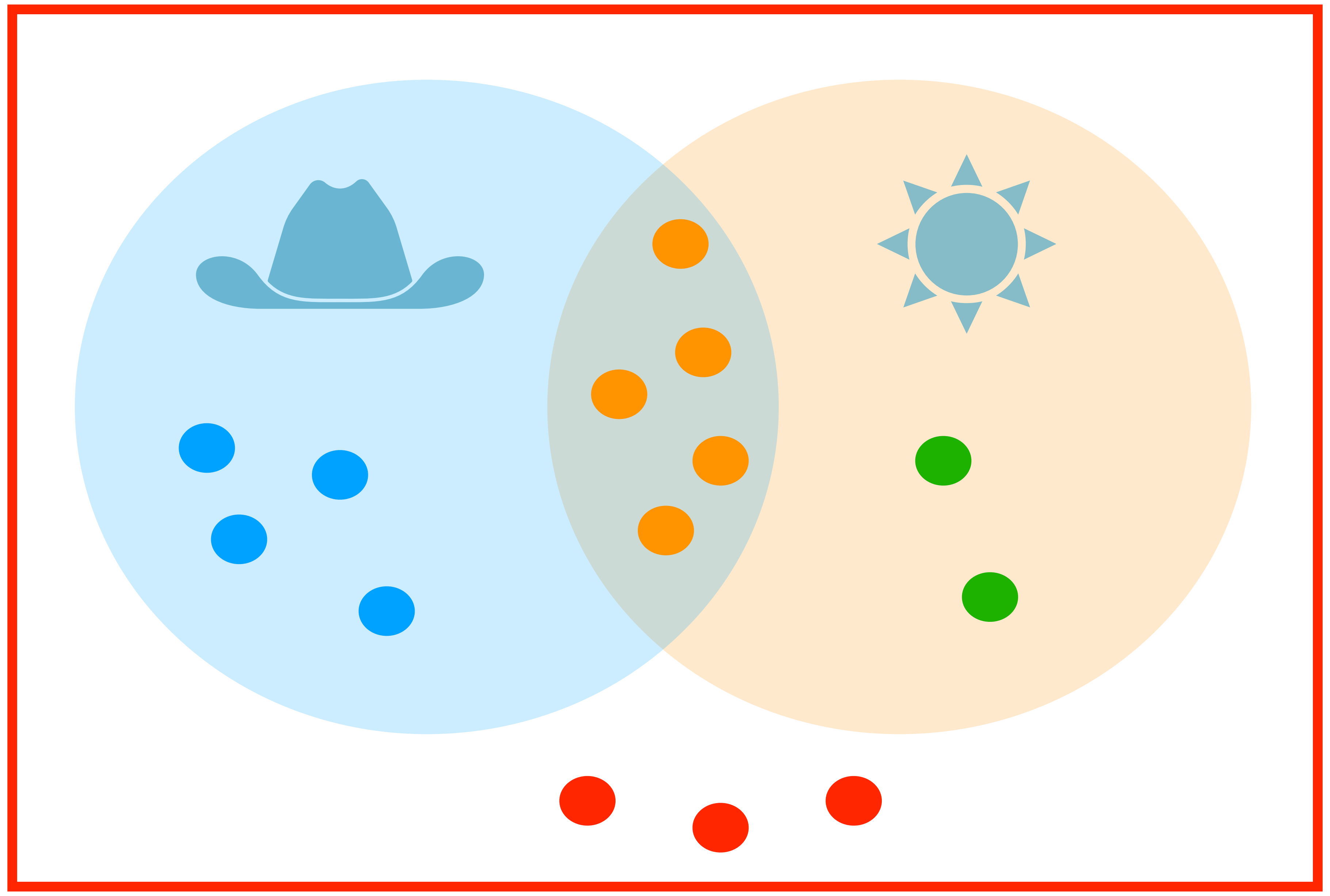
	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun \& hat} \mid \text{hat}) = \frac{}{p(\text{hat})}$$



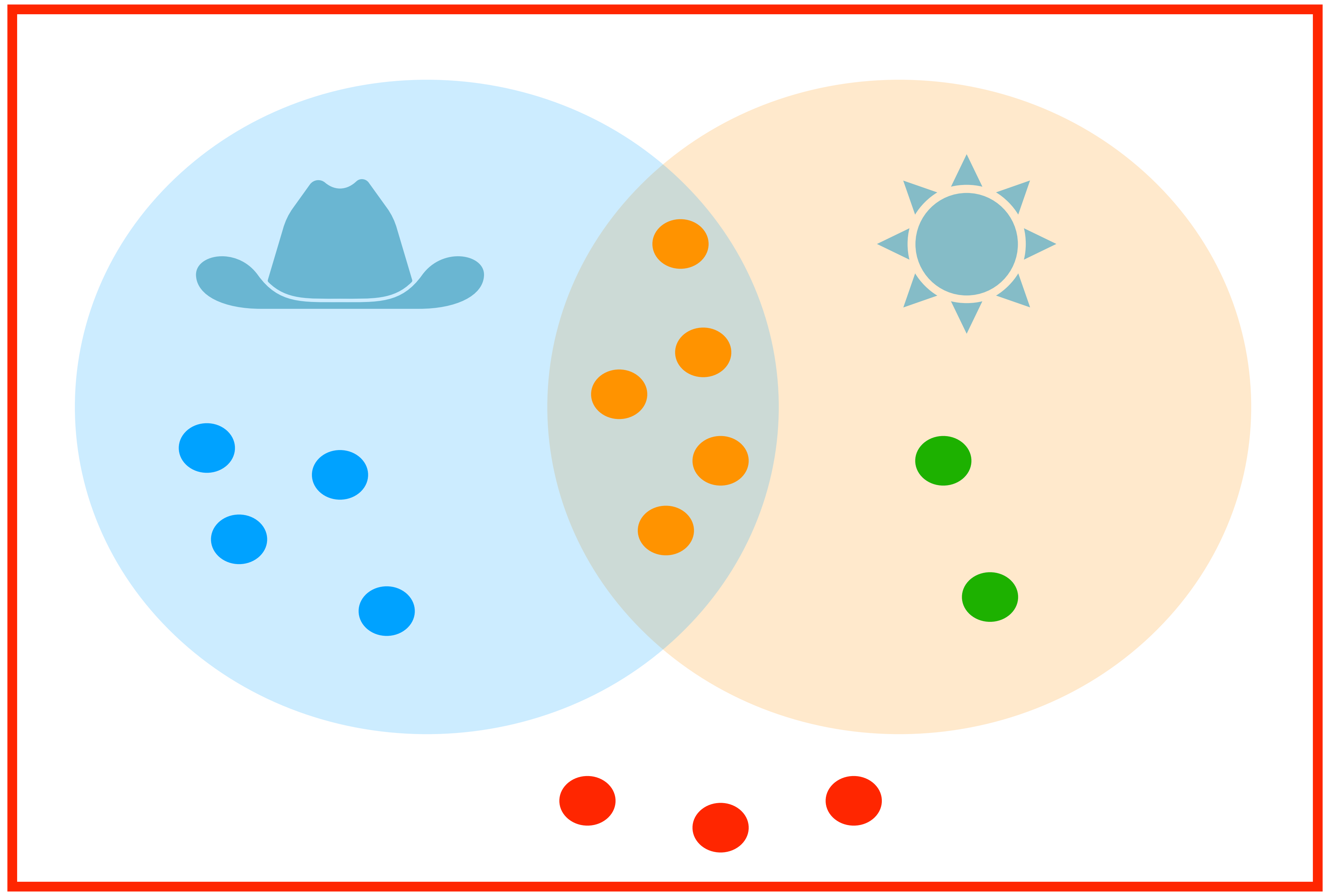
	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun \& hat} \mid \text{hat}) = \frac{p(\text{sun \& hat})}{p(\text{hat})}$$



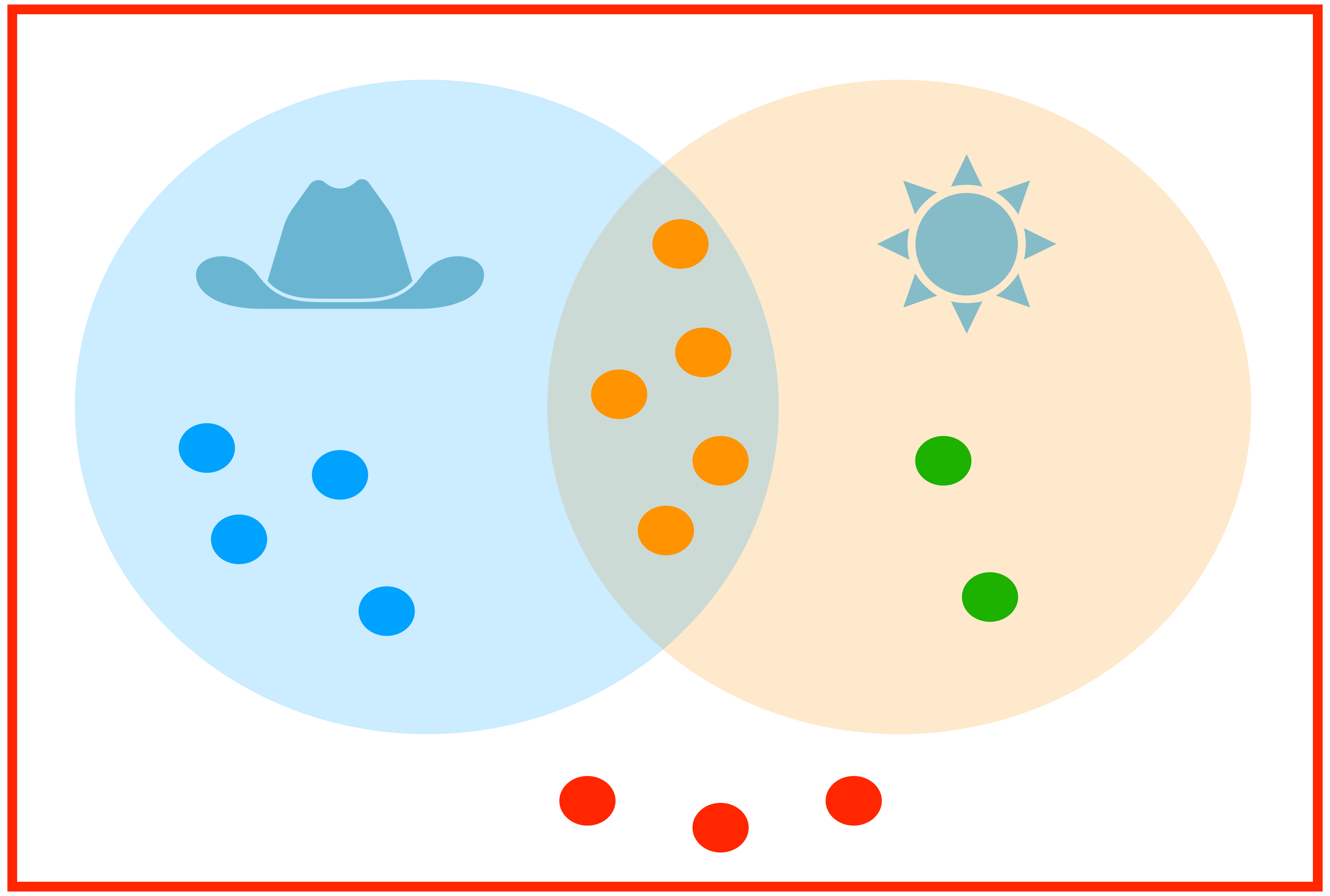
	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun \& hat} \mid \text{hat}) = \frac{5/14}{9/14} = .55$$



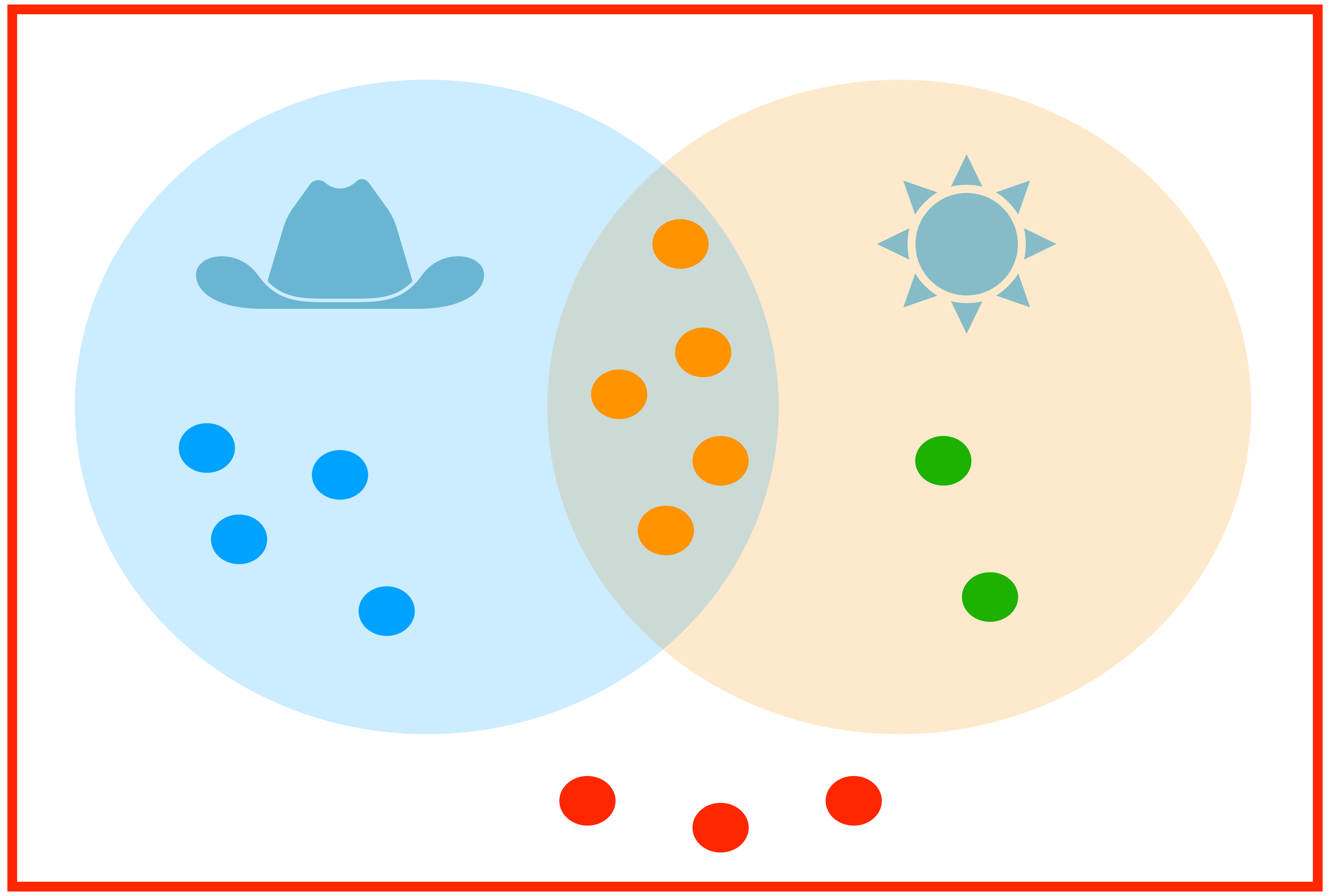
	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) = \frac{}{p(\text{sun})}$$



	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun \& hat} \mid \text{sun}) = \frac{p(\text{sun \& hat})}{p(\text{sun})}$$



	No sun	Sun	Total
Wore hat	4/14	5/14	9/14
Did not wear hat	3/14	2/14	5/14
Total	7/14	7/14	

$$p(\text{sun \& hat} \mid \text{sun}) = \frac{5/14}{7/14} = .71$$

$$p(\text{sun \& hat} \mid \text{hat}) = \frac{p(\text{sun \& hat})}{p(\text{hat})}$$

$$p(\text{sun \& hat} \mid \text{sun}) = \frac{p(\text{sun \& hat})}{p(\text{sun})}$$

$$p(\text{sun \& hat} \mid \text{hat}) = \frac{p(\text{sun \& hat})}{p(\text{hat})}$$

$$p(\text{sun \& hat} \mid \text{sun}) = \frac{p(\text{sun \& hat})}{p(\text{sun})}$$

We don’t usually know the probabilities of both events,
so we ask, is it possible to estimate the conditional p, without that data?

$$p(\text{sun} \,\&\, \text{hat} \mid \text{hat}) * p(\text{hat}) = \quad p(\text{sun} \,\&\, \text{hat})$$

$$p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) * p(\text{sun}) = \quad p(\text{sun} \,\&\, \text{hat})$$

Thinking about probability

$$p(\text{sun} \,\&\, \text{hat} \mid \text{hat}) * p(\text{hat}) = p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) * p(\text{sun})$$

$$p(\text{sun} \,\&\, \text{hat} \mid \text{hat}) * p(\text{hat}) = p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) * p(\text{sun})$$

$$p(\text{sun} \,\&\, \text{hat} \mid \text{hat}) = \frac{p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) * p(\text{sun})}{p(\text{hat})}$$

Thinking about probability

$$p(\text{sun} \,\&\, \text{hat} \mid \text{hat}) * p(\text{hat}) = p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) * p(\text{sun})$$

🥳

$$p(\text{sun} \,\&\, \text{hat} \mid \text{hat}) = \frac{p(\text{sun} \,\&\, \text{hat} \mid \text{sun}) * p(\text{sun})}{p(\text{hat})}$$

$$p(\text{sun \& hat} \mid \text{hat}) = \frac{p(\text{sun \& hat} \mid \text{sun}) * p(\text{sun})}{p(\text{hat})}$$

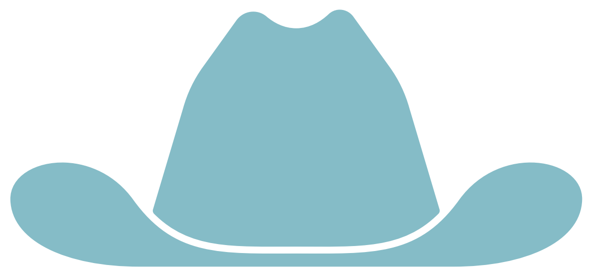
== Bayes Theorem! 🥳🥳🥳

Thinking about probability

$$p(\text{sun} \;\&\; \cancel{\text{hat}} \mid \text{hat}) = \frac{p(\cancel{\text{sun}} \;\&\; \text{hat} \mid \text{sun}) * p(\text{sun})}{p(\text{hat})}$$

$$p(\text{sun} \mid \text{hat}) = \frac{p(\text{hat} \mid \text{sun}) * p(\text{sun})}{p(\text{hat})}$$

Thinking about probability



evidence/
data

hypothesis/
parameter

$$p(\text{parameter} \mid \text{data}) = \frac{p(\text{data} \mid \text{parameter}) * p(\text{parameter})}{p(\text{data})}$$

$$p(\text{parameter} \mid \text{data}) = \frac{p(\text{data} \mid \text{parameter}) * p(\text{parameter})}{p(\text{data})}$$

IV. Bayesian analysis

$$p(\text{parameter} \mid \text{data}) = \frac{p(\text{data} \mid \text{parameter}) * p(\text{parameter})}{p(\text{data})}$$

Likelihood

$$p(\text{parameter} \mid \text{data}) = \frac{p(\text{data} \mid \text{parameter}) * p(\text{parameter})}{p(\text{data})}$$

What is the probability of observing this data, given this parameter value?

Posterior **Likelihood**

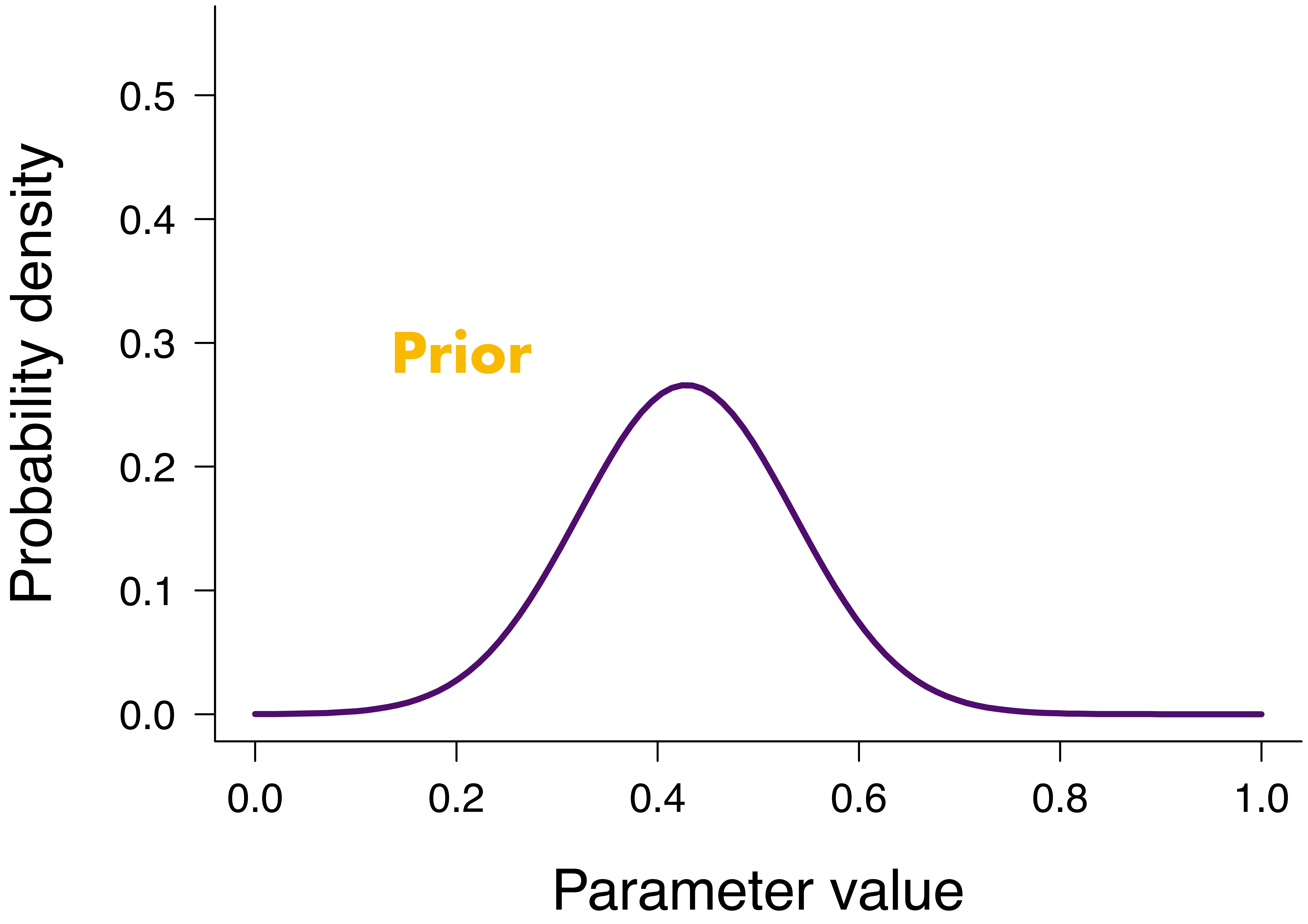
$$\underbrace{p(\text{parameter} \mid \text{data})}_{\text{Posterior}} = \frac{\underbrace{p(\text{data} \mid \text{parameter})}_{\text{Likelihood}} * p(\text{parameter})}{p(\text{data})}$$

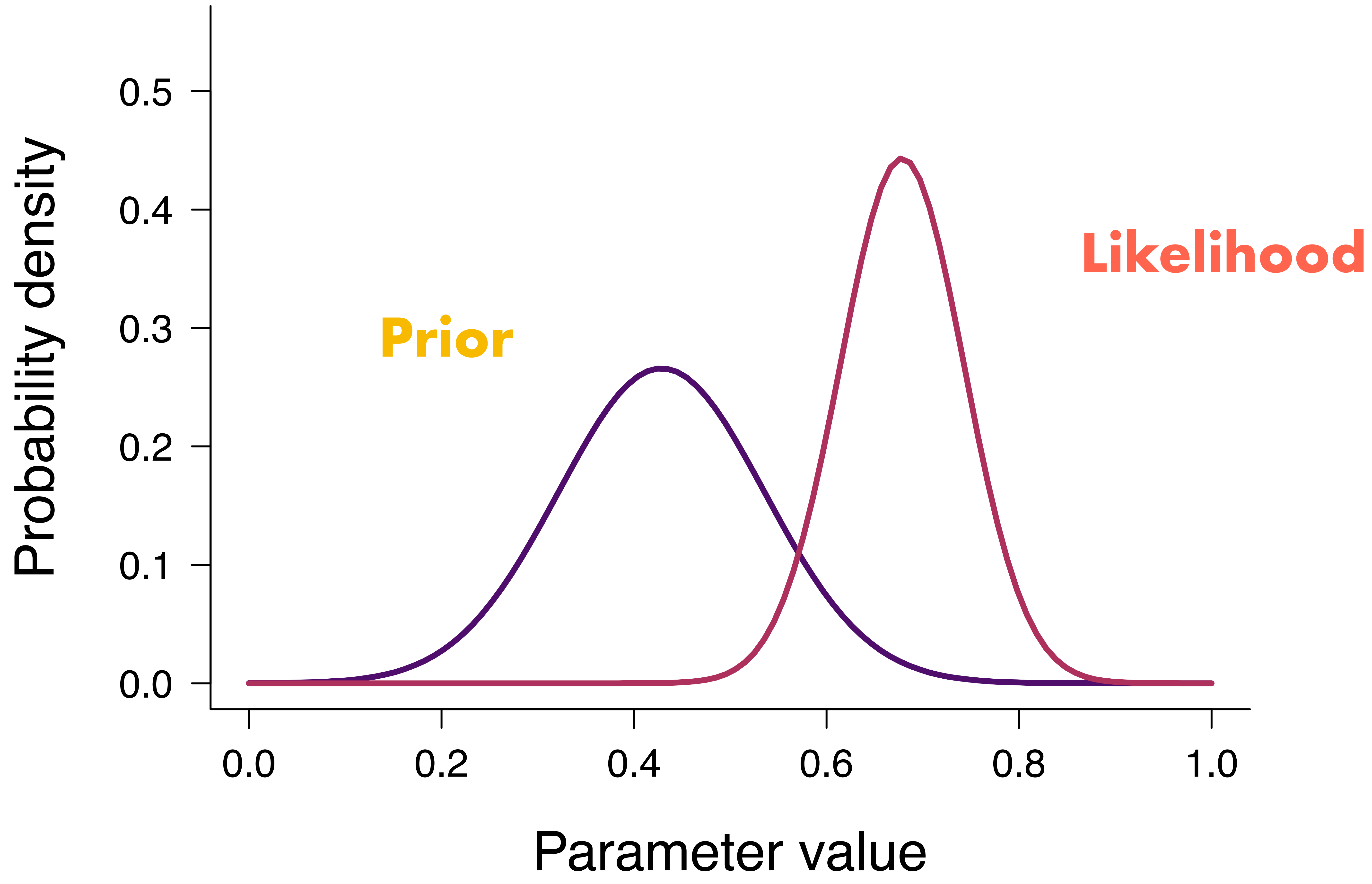
What is the probability of a given parameter value, given the data we have observed?

$$\underbrace{p(\text{parameter} \mid \text{data})}_{\text{Posterior}} = \frac{\underbrace{p(\text{data} \mid \text{parameter})}_{\text{Likelihood}} * \underbrace{p(\text{parameter})}_{\text{Prior}}}{p(\text{data})}$$

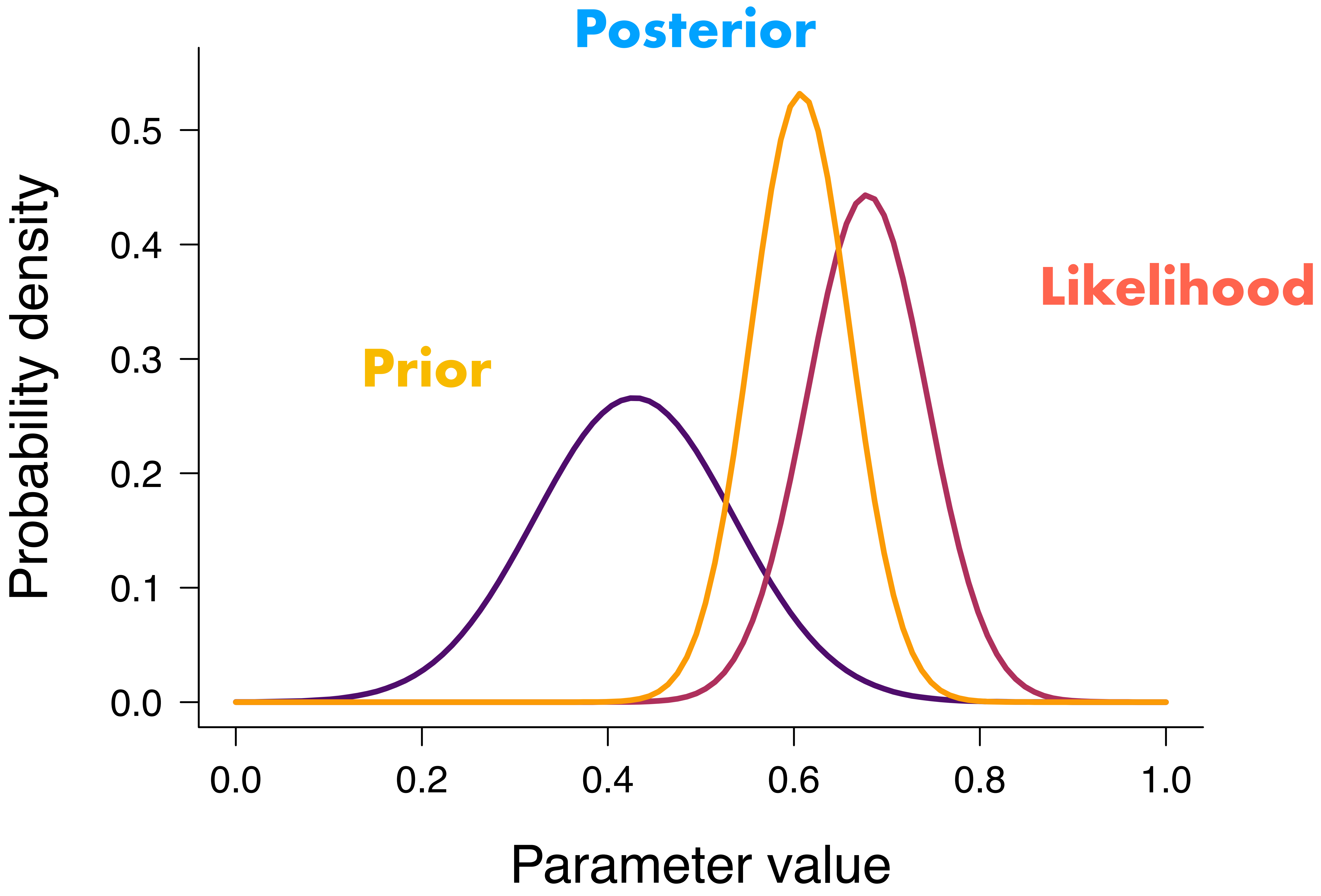
What is our expectation for the probability of the parameter?

We have three quantities of interest




Probability density
0.5
0.4
0.3
0.2
0.1
0.0
Prior
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
Parameter value

Bayesian analysis



Use prior knowledge to define *prior distributions*

Observe data and calculate the *likelihood*

Apply Bayes' theorem to estimate *posterior distributions*

Exercise 3: Explore impact of likelihood and prior on posterior probability with online tool

Defining a prior

Defining a prior

- vague/minimally informative

Defining a prior

- vague/minimally informative
- subjective/expert opinion

Defining a prior

- vague/minimally informative
- subjective/expert opinion
- estimate from previous data

Defining a prior

- vague/minimally informative
- subjective/expert opinion
- estimate from previous data

DON’T define your prior based on examining your data

Exercise 4: Defining a prior

Instead of a single value estimate for a parameter, we have an entire probability distribution across all unknowns.

$$y = a + bx + \epsilon$$

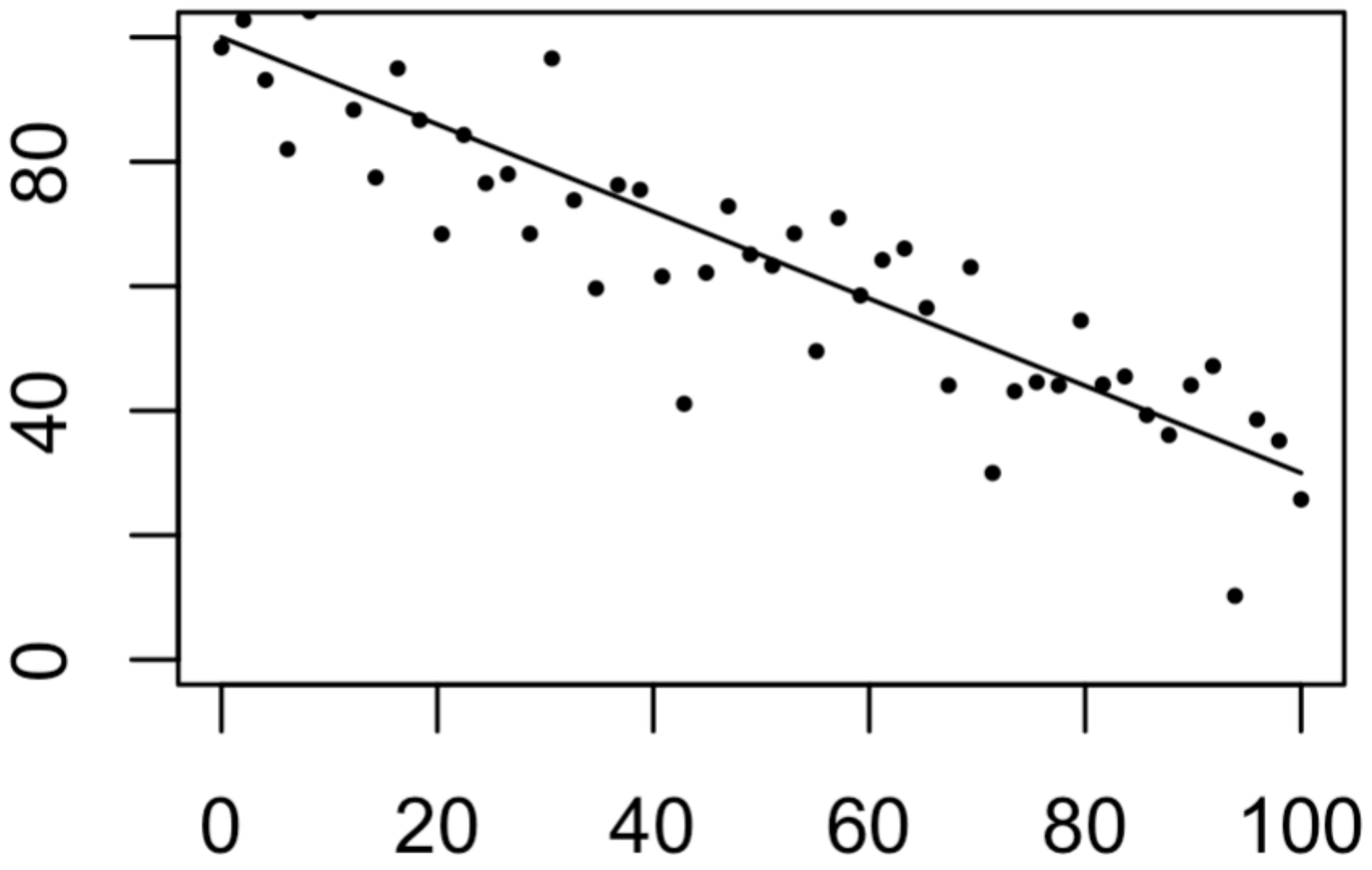
$$\epsilon \sim N(0, \sigma^2)$$

MLE

$$b = -0.67$$

Posterior mean

$$b = -0.67$$



Bayesian inference is intuitive

Frequentist inference is confusing

Bayesian inference is intuitive

Frequentist inference is confusing

Exercise 5: Confidence intervals

Frequentist confidence interval

"Were this procedure to be repeated on numerous samples, the fraction of calculated confidence intervals (which would differ for each sample) that encompass the true population parameter would tend toward 90%."

Frequentist confidence interval

"Were this procedure to be repeated on numerous samples, the fraction of calculated confidence intervals (which would differ for each sample) that encompass the true population parameter would tend toward 90%."

Bayesian credible interval

"An interval within which an unobserved parameter value falls with a particular probability."

Frequentist approach:

- No place for “prior beliefs”
- Inference should only depend on the data (likelihood)
- Probability is the same as frequency
- Point estimate

Frequentist approach:

- No place for “prior beliefs”
- Inference should only depend on the data (likelihood)
- Probability is the same as frequency
- Point estimate

Bayesian approach:

- Inference depends on prior knowledge and available data
- Probability is subjective; it is a degree of belief
- It is more intuitive! “I am 95% certain that…”

V. Introduction to brms

$$\overbrace{p(\text{parameter} \mid \text{data})}^{\text{Posterior}} = \frac{\overbrace{p(\text{data} \mid \text{parameter})}^{\text{Likelihood}} * \overbrace{p(\text{parameter})}^{\text{Prior}}}{p(\text{data})}$$

$$\overbrace{p(\text{parameter} \mid \text{data})}^{\text{Posterior}} = \frac{\overbrace{p(\text{data} \mid \text{parameter})}^{\text{Likelihood}} * \overbrace{p(\text{parameter})}^{\text{Prior}}}{\text{😒}}$$

$$\overbrace{p(\text{parameter} \mid \text{data})}^{\textbf{Posterior}} \propto \overbrace{p(\text{data} \mid \text{parameter})}^{\textbf{Likelihood}} * \overbrace{p(\text{parameter})}^{\textbf{Prior}}$$

We can approximate the posterior by drawing a large random sample from the distribution using Markov chain Monte Carlo (MCMC)

‘Stan’ is a software package that comes with a programming language to implement MCMC.

‘Stan’ is a software package that comes with a programming language to implement MCMC.

Stan uses Hamiltonian Monte Carlo and No-U-Turn Sampler (NUTS) algorithms to implement the MCMC sampling.

'brms()' is a software package that leverages lme4-like syntax to make implementation of Stan functionality more accessible.

‘brms()’ is a software package that leverages lme4-like syntax to make implementation of Stan functionality more accessible.

R practical: introduction to brms()

R/2-linear_models.R

Chains:

Iterations:

Warmup:

Thin:

Draws:

Chains: Number of Markov chains

Iterations:

Warmup:

Thin:

Draws:

Chains: Number of Markov chains

Iterations: Number of steps per chain

Warmup:

Thin:

Draws:

Chains: Number of Markov chains

Iterations: Number of steps per chain

Warmup: First walks around parameter space that you throw away as the chain searches for the right area.

Thin:

Draws:

Chains:	Number of Markov chains
Iterations:	Number of steps per chain
Warmup:	First walks around parameter space that you throw away as the chain searches for the right area.
Thin:	Prevents correlation between steps by removing steps at this rate.
Draws:	

Chains: Number of Markov chains

Iterations: Number of steps per chain

Warmup: First walks around parameter space that you throw away as the chain searches for the right area.

Thin: Prevents correlation between steps by removing steps at this rate.

Draws: Iterations x chains

VI. Model checking

Prior predictive checks: generate data according to the priors to assess whether they are appropriate.

Prior predictive checks: generate data according to the priors to assess whether they are appropriate.

R practical: prior predictive checks in brms

Convergence: Do the samples in the chains converge in to the same maxima of the posterior distribution

1. Whether each chain converges on an estimate

2. Whether all chains converge on the same estimate

Convergence: Do the samples in the chains converge in to the same maxima of the posterior distribution

- Traceplot
- Rhat
- Effective sample size

Convergence: Do the samples in the chains converge in to the same maxima of the posterior distribution

- Traceplot
- Rhat
- Effective sample size

R practical: diagnosing convergence

Other checks

Posterior predictive check = simulate your fitted model
Check residuals

Addressing your question

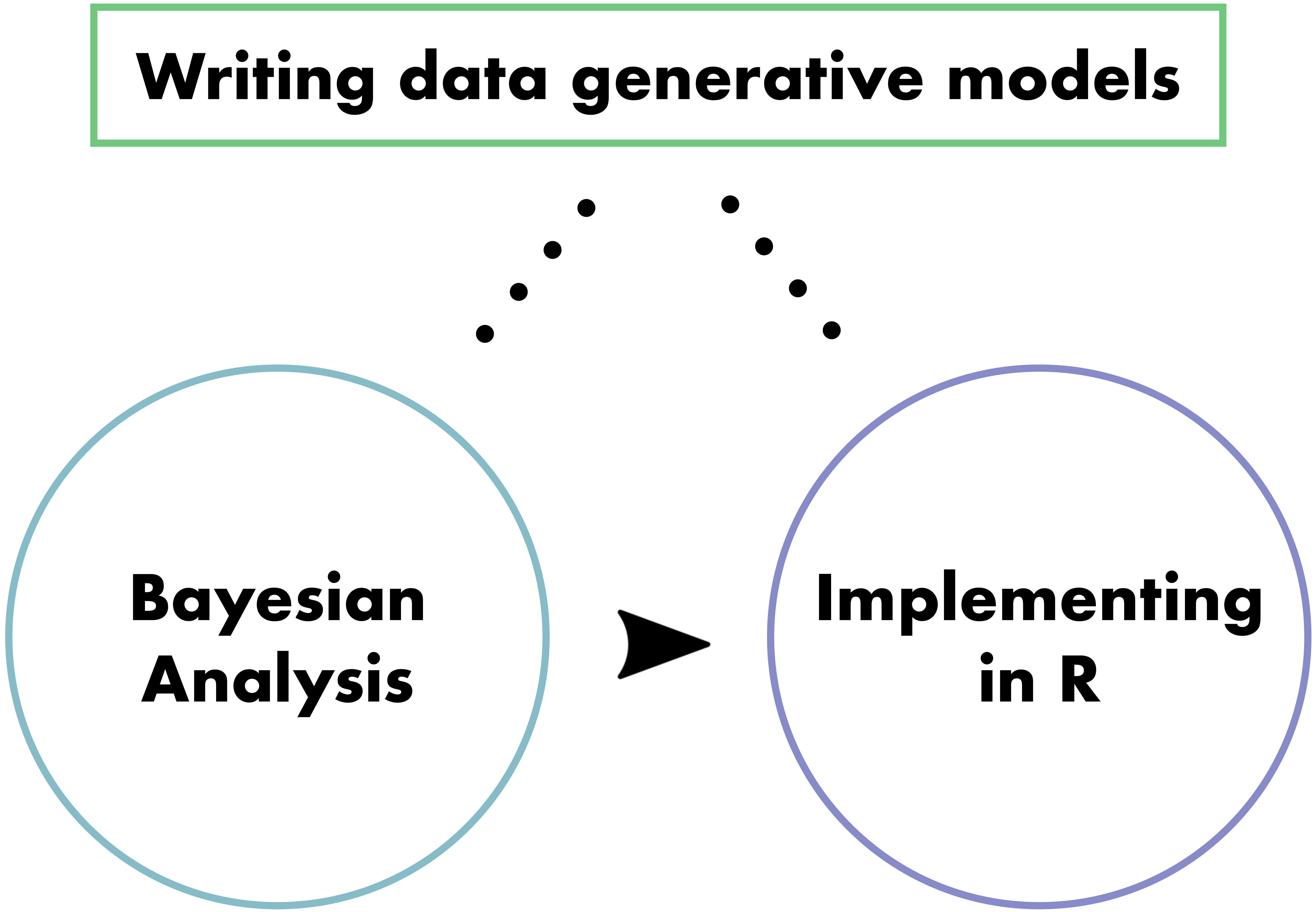
Estimating value of parameter:
Forest plot, summarise, mean and sd of posterior of parameter.

Hypothesis testing:
Bayesian p-value. "what is the posterior probability that X has an effect on Y."

Prediction:
calculate, effect/response plot to new data. Cross-validation - testing on new data.

Review

Workshop aims



Posterior **Likelihood** **Prior**

$$\underbrace{p(\text{parameter} \mid \text{data})}_{\text{Posterior}} \propto \underbrace{p(\text{data} \mid \text{parameter})}_{\text{Likelihood}} * \underbrace{p(\text{parameter})}_{\text{Prior}}$$

We can approximate the posterior by drawing a large random sample from the distribution using Markov chain Monte Carlo (MCMC)

Defining a prior

- estimate from previous data
- subjective/expert opinion
- vague/minimally informative

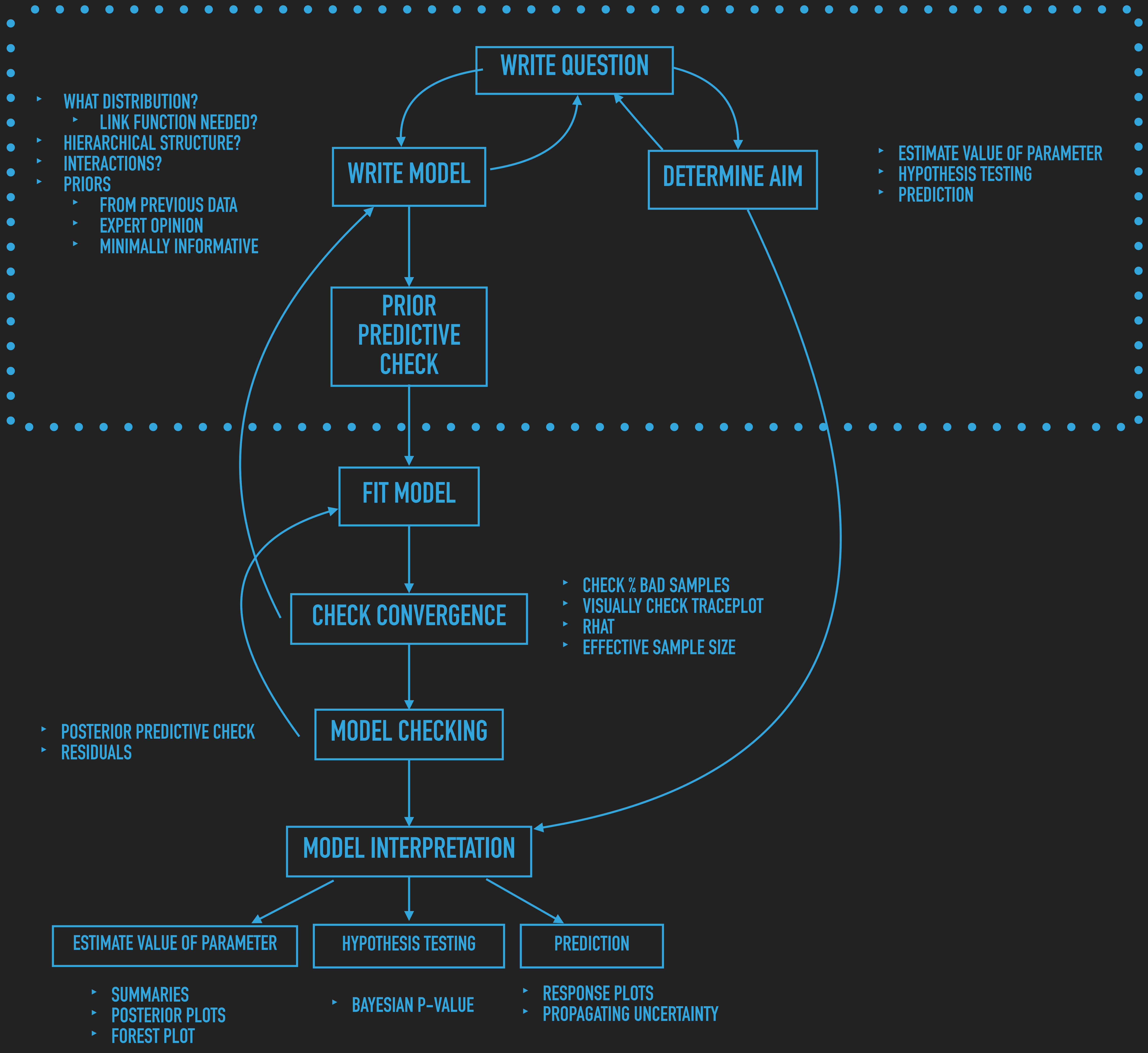
DON'T define your prior based on examining your data

Frequentist approach:

- No place for “prior beliefs”
- Inference should only depend on the data (likelihood)
- Probability is the same as frequency
- Point estimate

Bayesian approach:

- Inference depends on prior knowledge and available data
- Probability is subjective; it is a degree of belief
- It is more intuitive! “I am 95% certain that…”



WRITE QUESTION
WHAT DISTRIBUTION?
LINK FUNCTION NEEDED?
HIERARCHICAL STRUCTURE?
INTERACTIONS?
PRIORS
FROM PREVIOUS DATA
EXPERT OPINION
MINIMALLY INFORMATIVE
WRITE MODEL
DETERMINE AIM
ESTIMATE VALUE OF PARAMETER
HYPOTHESIS TESTING
PREDICTION
PRIOR PREDICTIVE CHECK
FIT MODEL
CHECK CONVERGENCE
CHECK % BAD SAMPLES
VISUALLY CHECK TRACEPLOT
RHAT
EFFECTIVE SAMPLE SIZE
MODEL CHECKING
POSTERIOR PREDICTIVE CHECK
RESIDUALS
MODEL INTERPRETATION
ESTIMATE VALUE OF PARAMETER
HYPOTHESIS TESTING
PREDICTION
SUMMARIES
POSTERIOR PLOTS
FOREST PLOT
BAYESIAN P-VALUE
RESPONSE PLOTS
PROPAGATING UNCERTAINTY

Special thanks to

Nick Golding
Marc Kery
Gerry Ryan
Richard McElreath