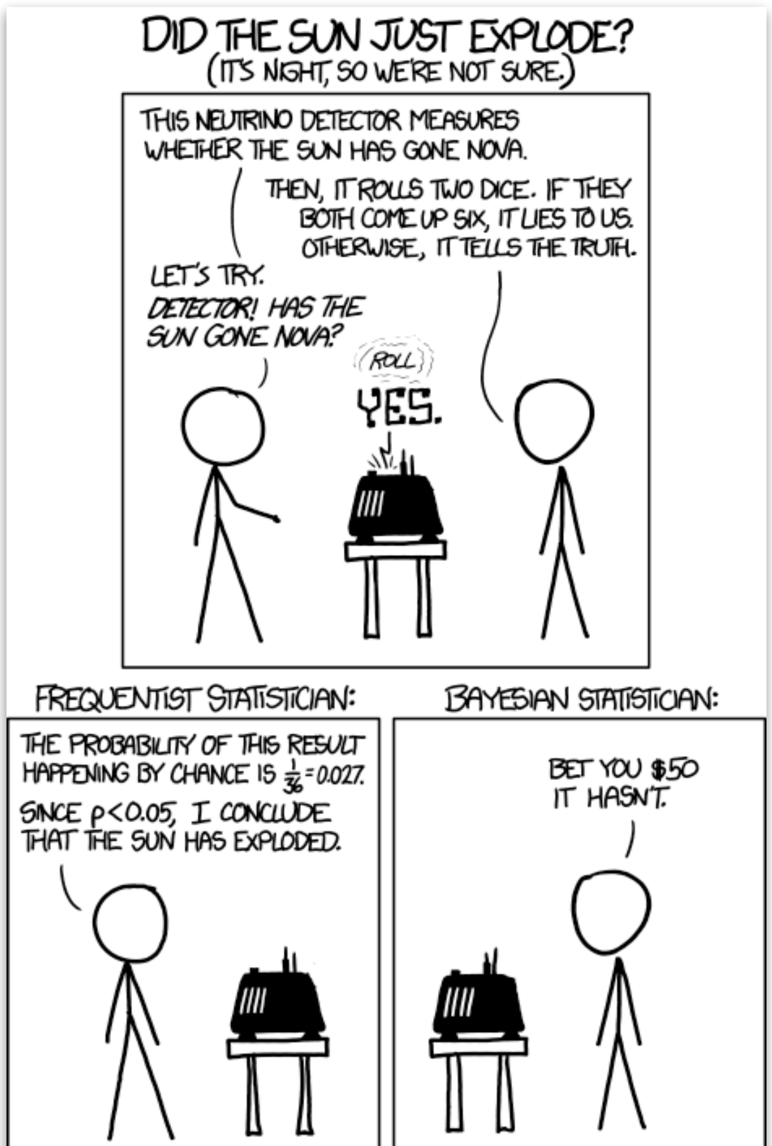
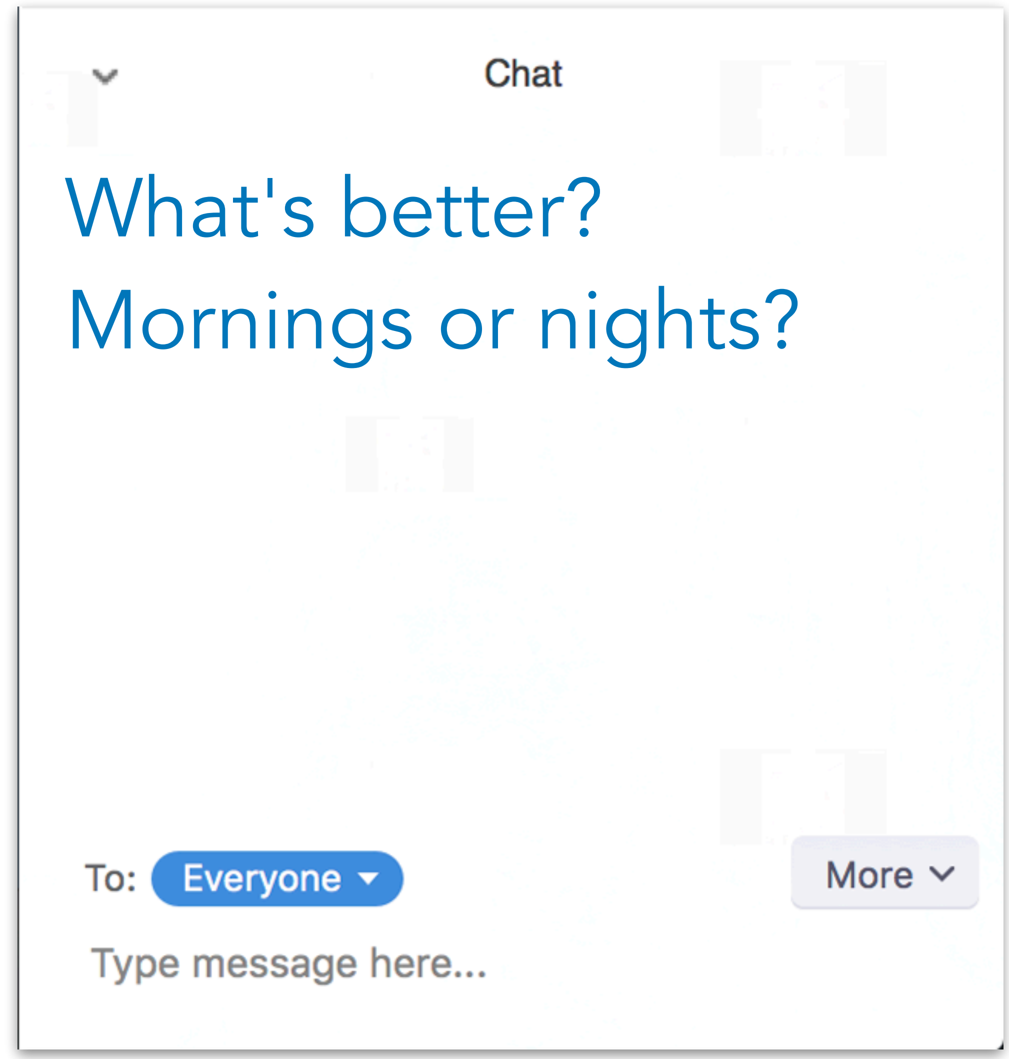
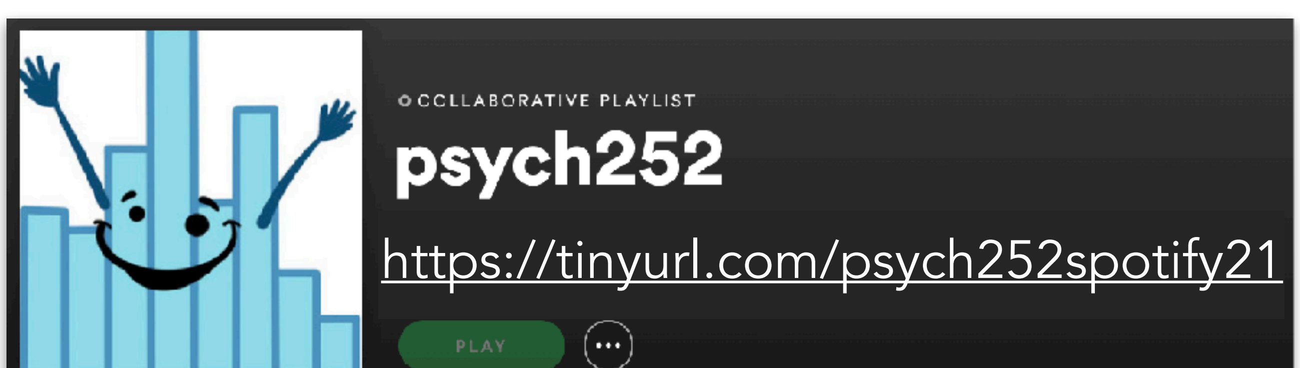
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

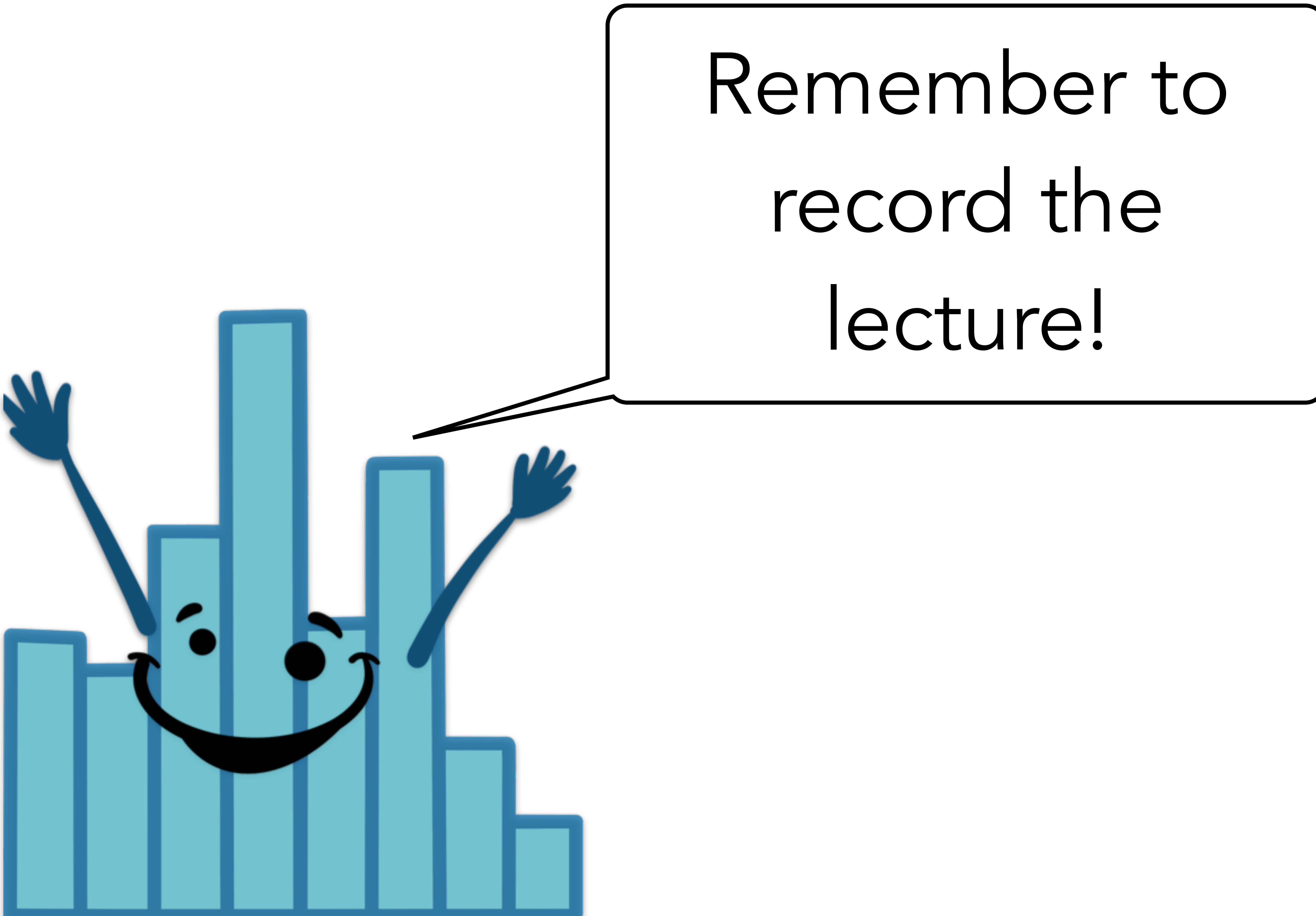
Bayesian data analysis 1





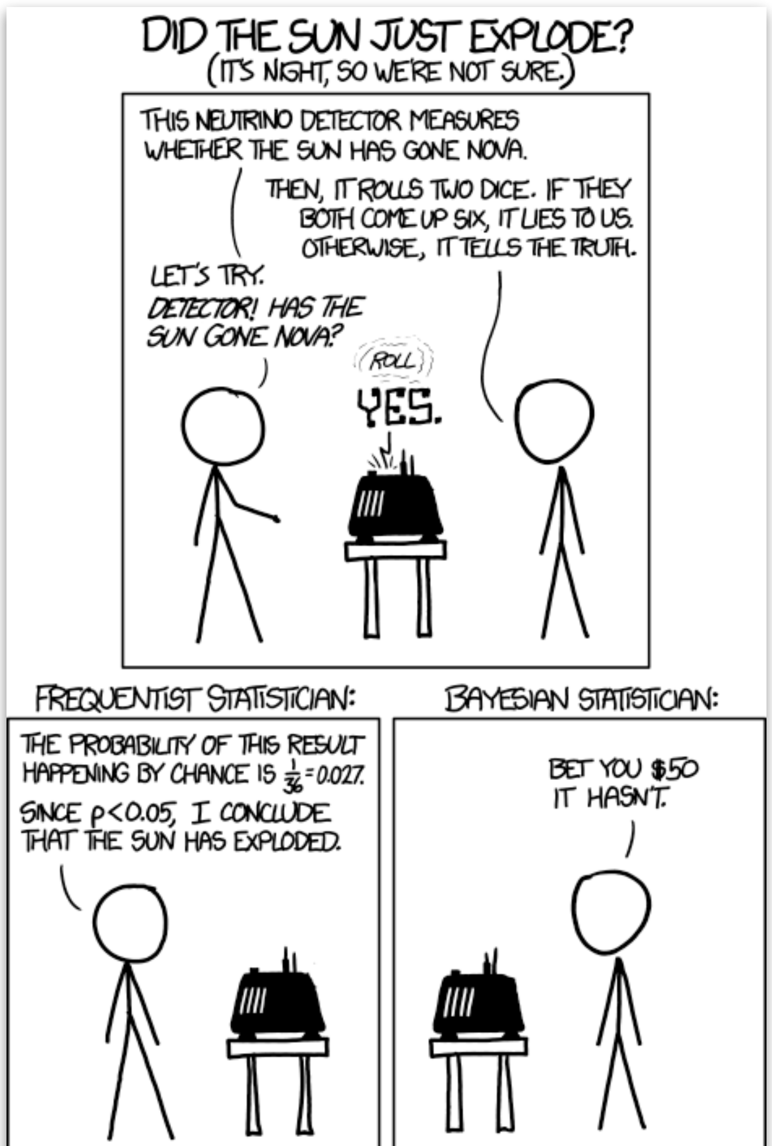


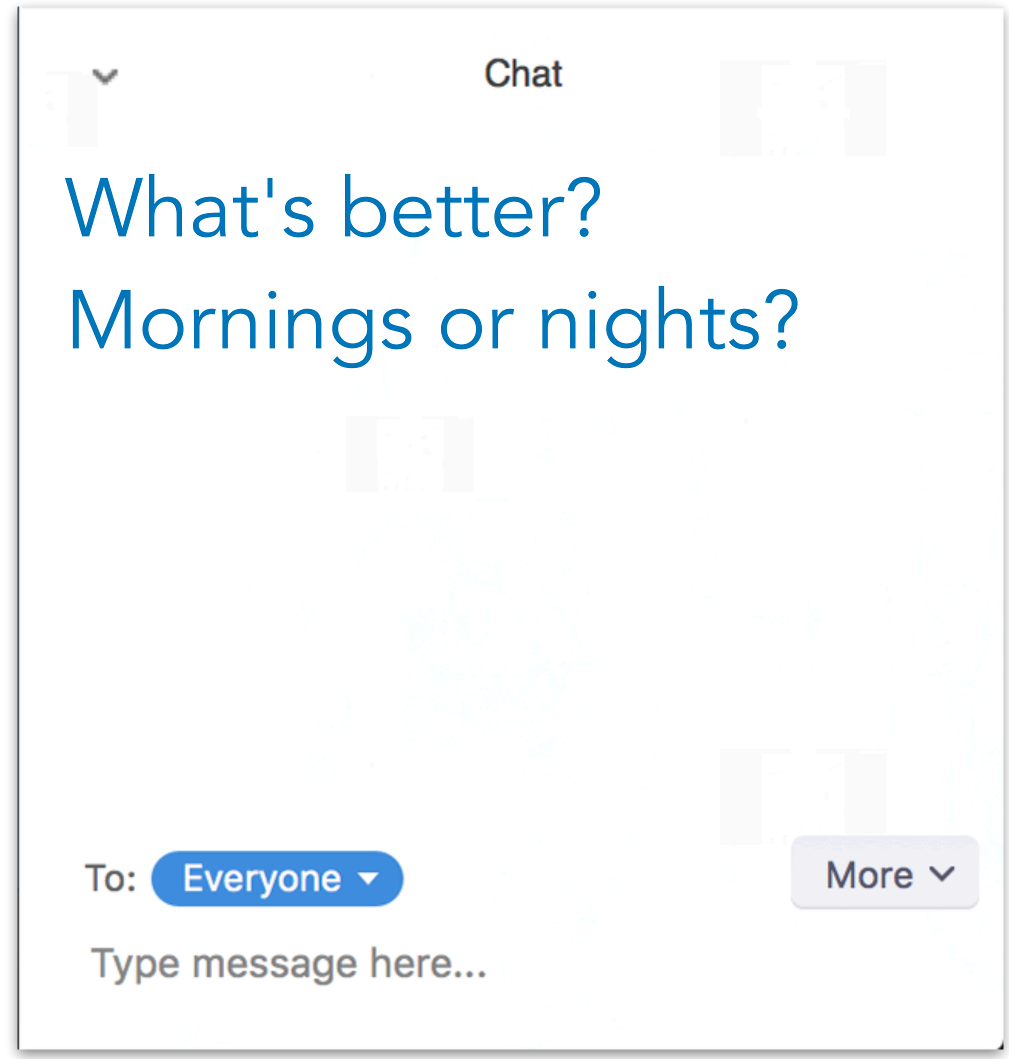
03/05/2021

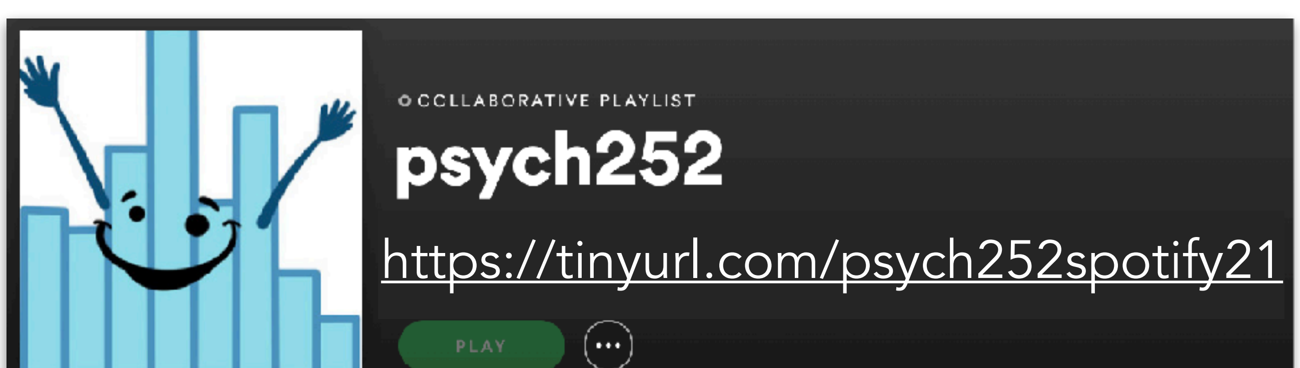

Remember to record the lecture!

PSYCH 252: Statistical Methods for Behavioral and Social Sciences

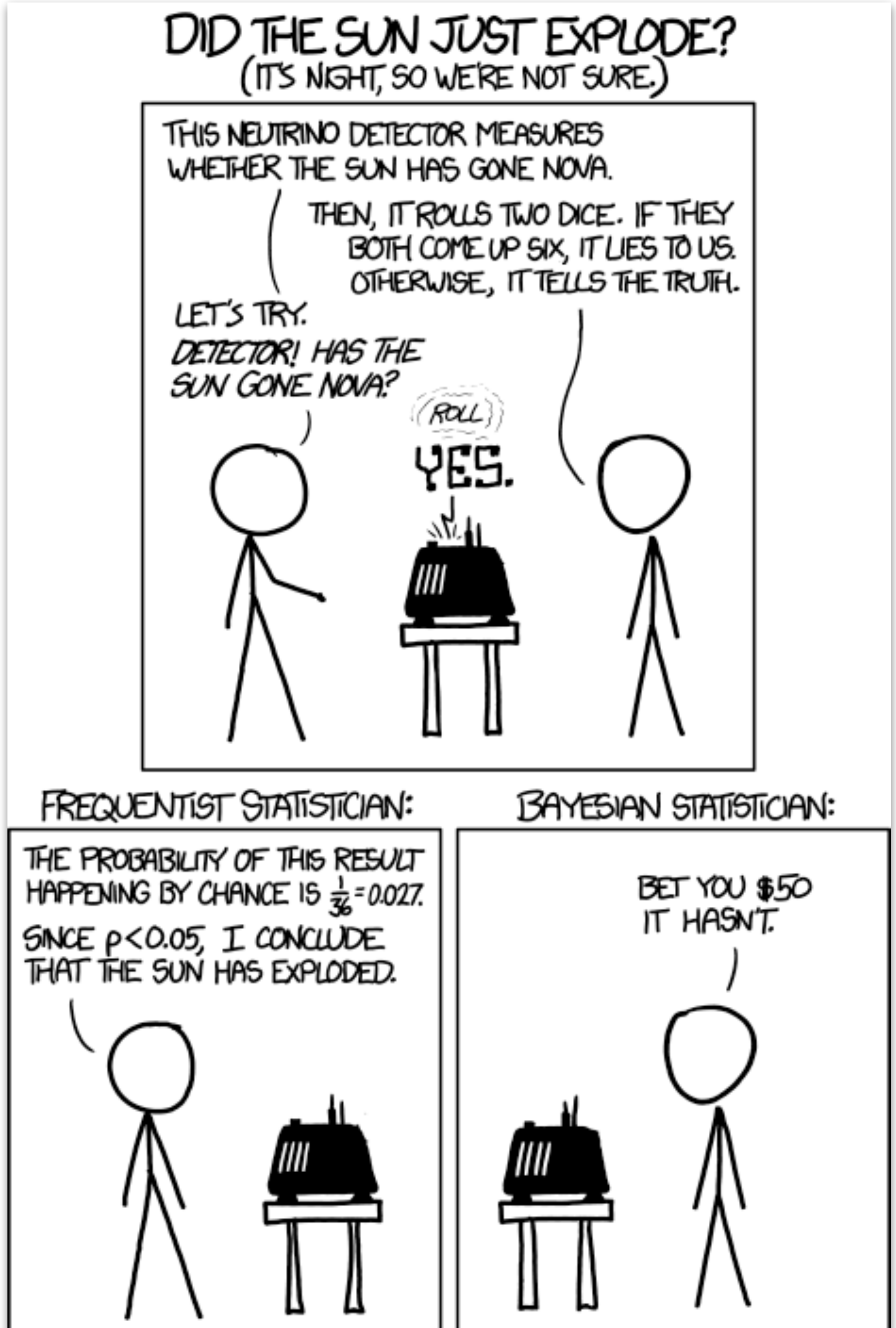
Bayesian data analysis 1







03/05/2021


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36}$ = 0.027.
SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU \$50 IT HASN'T.

Things that came up

Rohan Alexander
@RohanAlexander

When grad students want to learn something they take a course on it. When professors want to learn something they teach a course on it.

6:44 AM · Mar 3, 2021 · Twitter for iPad

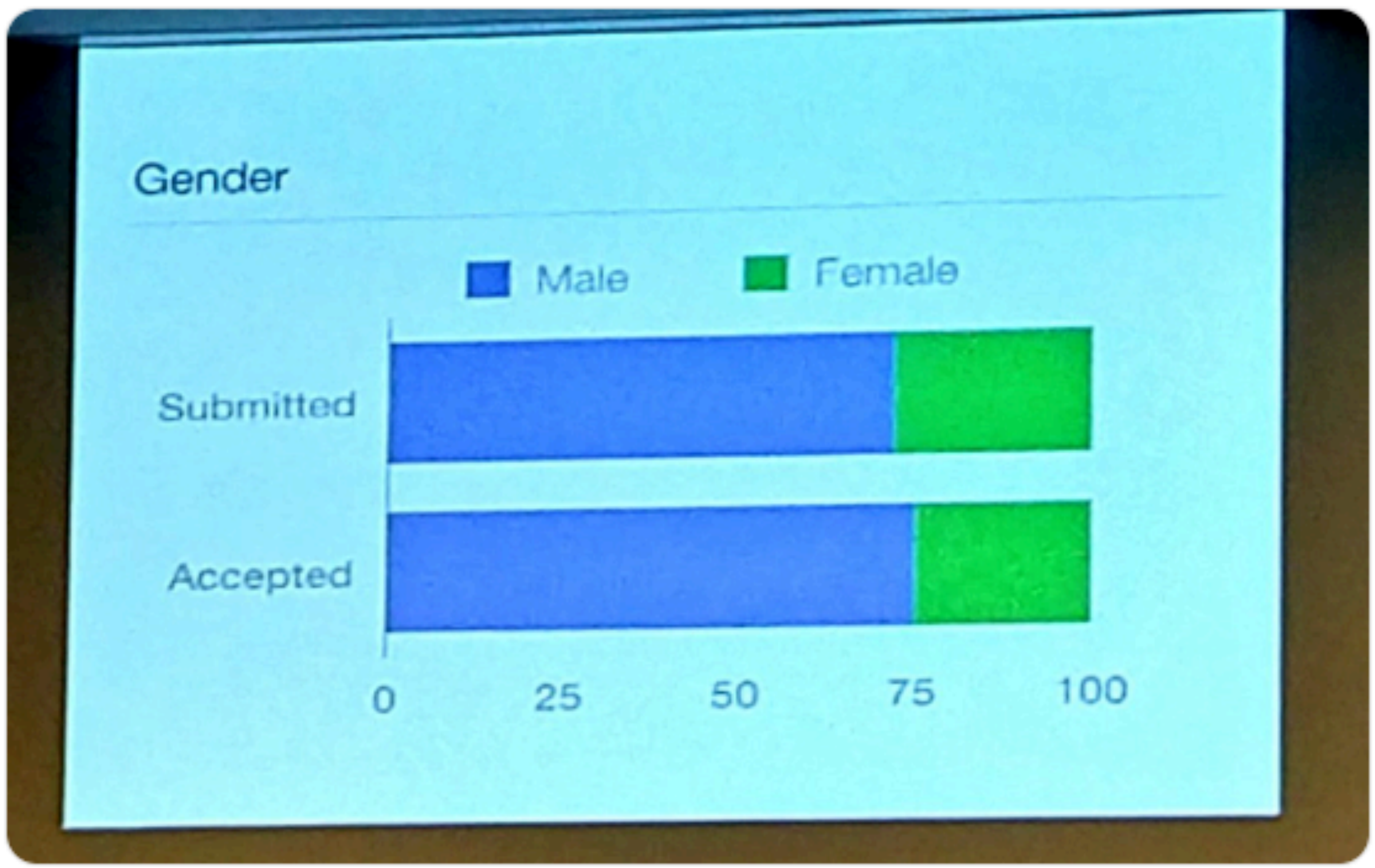
34 Retweets **16** Quote Tweets **447** Likes

Bias in conference admission?

Adam J Calhoun
@neuroecology

Follow

Accepted and submitted abstracts by gender roughly the same at #cosyne19 cc @neimarkgeffen @TrackingActions



10:29 AM - 28 Feb 2019

Bias in conference admission?

Adam J Calhoun @neuroecology · Feb 28

Accepted and submitted abstracts by gender roughly the same at #cosyne19 cc @neimarkgeffen @TrackingActions

4 10

Yael Niv
@yael_niv

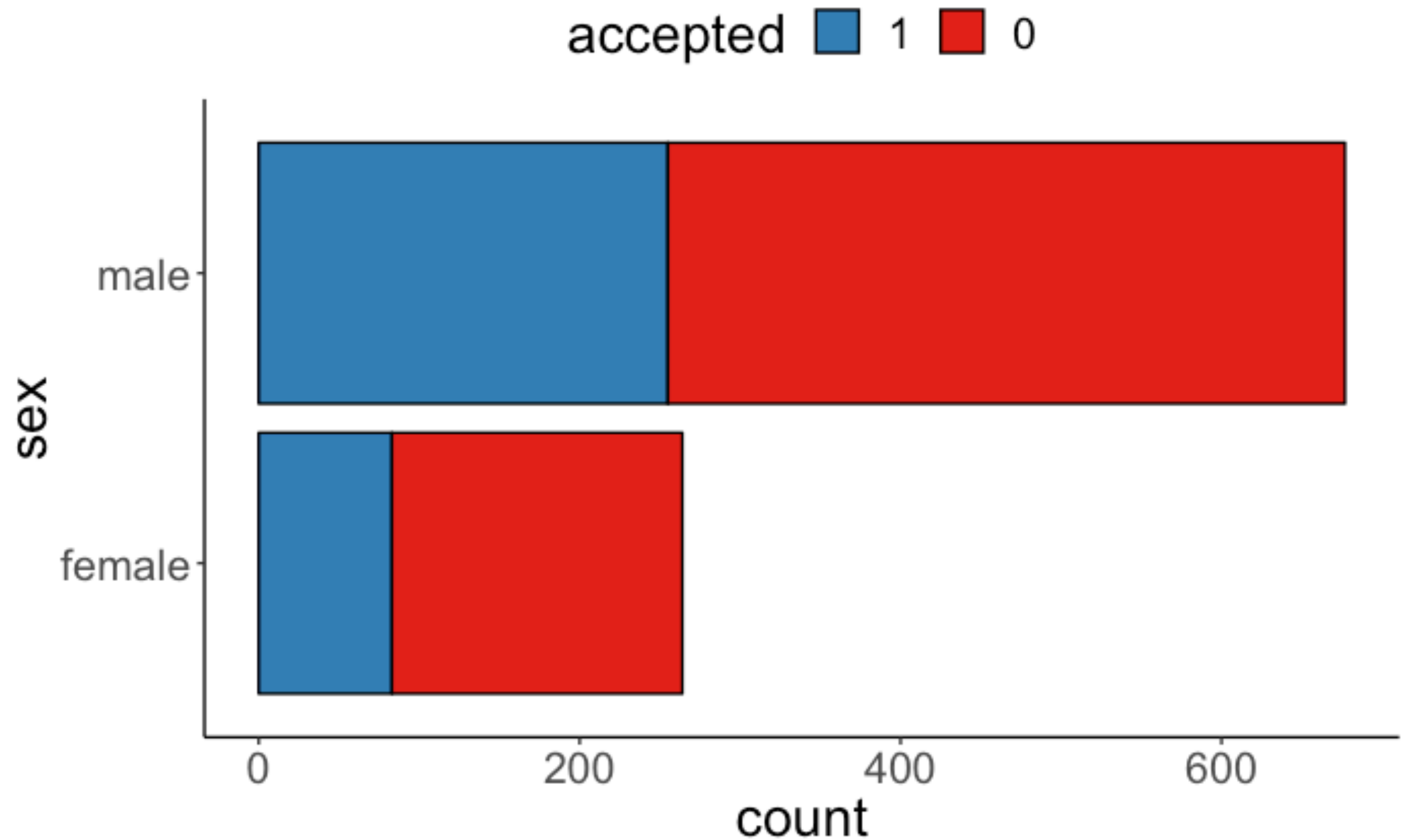
Follow

Replying to @neuroecology @neimarkgeffen @TrackingActions

Doesn't look the same to me...

10:55 AM - 2 Mar 2019

Bias in conference admission?



different representation of the data

Bias in conference admission?

```r
# logistic regression
fit.glm = glm(formula = accepted ~ 1 + sex,
              family = "binomial",
              data = df.conference)

# model summary
fit.glm %>%
  summary()
```

```
Call:
glm(formula = accepted ~ 1 + sex, family = "binomial", data = df.conference)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9723  -0.9723  -0.8689   1.3974   1.5213

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7797     0.1326  -5.881 4.07e-09 ***
sexmale       0.2759     0.1545   1.786   0.0741 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1228.9  on 940  degrees of freedom
Residual deviance: 1225.6  on 939  degrees of freedom
AIC: 1229.6

Number of Fisher Scoring iterations: 4
```

Bias in conference admission?

Megan Carey @meganinlisbon · Mar 2

I presented the math for this at the #cosyne19 diversity lunch today.

Success rates for first authors with known gender:
Female: 83/264 accepted = 31.4%
Male: 255/677 accepted = 37.7%

37.7/31.4 = a 20% higher success rate for men

Adam J Calhoun @neuroecology

Accepted and submitted abstracts by gender roughly the same at #cosyne19 cc @neimarkgeffen @TrackingActions

Show this thread

9 37 83

Mehrdad Jazayeri
@mjaztwit

Following

Replying to @meganinlisbon

That's a really large difference. It's seems like this year we really messed up as a community. What's the distribution of difference under the null (if you do the same analysis but shuffle the gender labels)?

8:06 AM - 2 Mar 2019

8 Likes

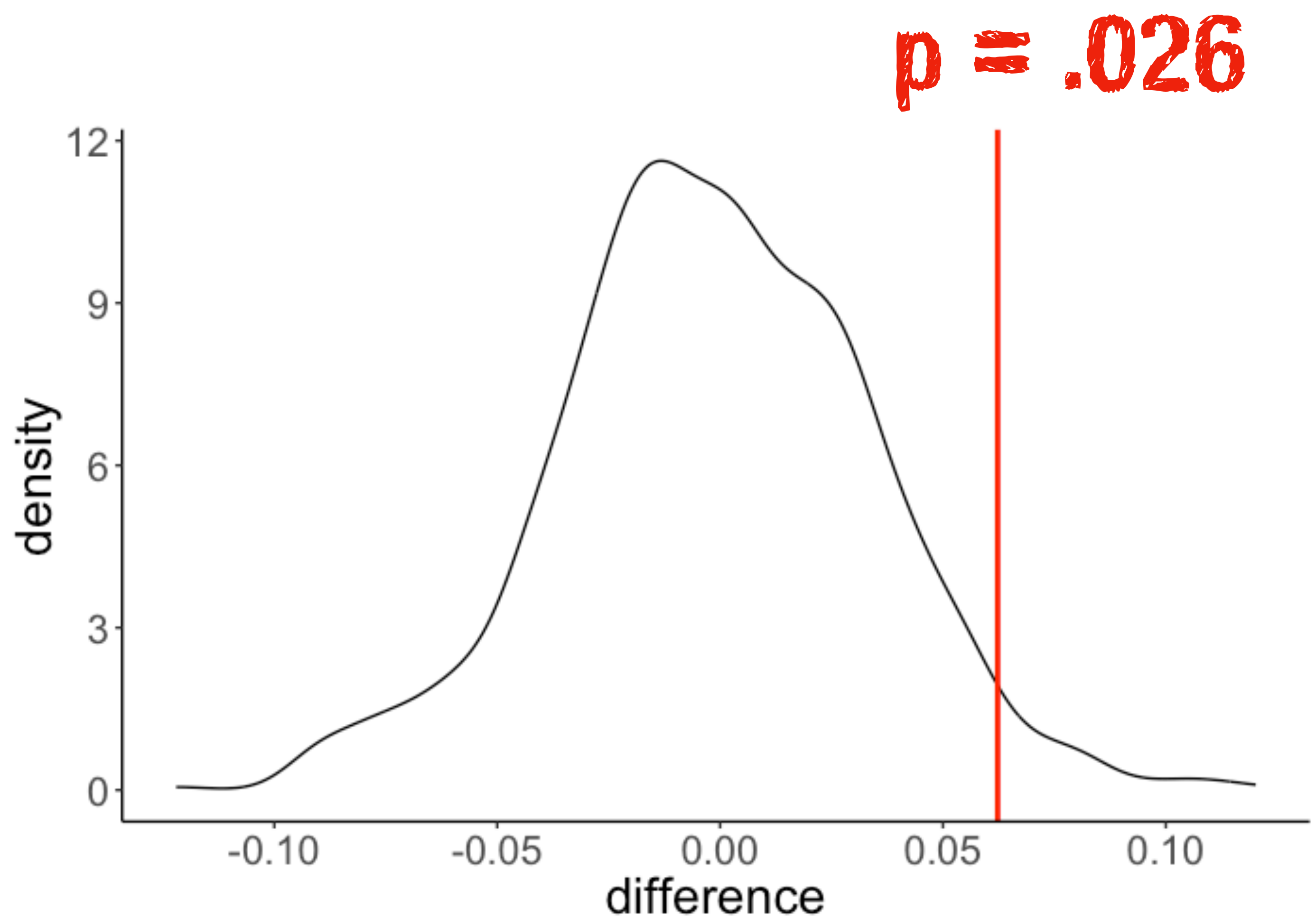
permutation test!

Bias in conference admission?

```r
# difference in proportion
fun.difference = function(df){
  df %>%
    as_tibble() %>%
    count(sex, accepted) %>%
    group_by(sex) %>%
    mutate(proportion = n / sum(n)) %>%
    filter(accepted == 1) %>%
    select(sex, proportion) %>%
    pivot_wider(names_from = sex,
                values_from = proportion) %>%
    mutate(difference = male - female) %>%
    pull(difference)
}

# actual difference
difference = df.conference %>%
  fun.difference()

# permutation test
df.permutation = df.conference %>%
  permute(n = 1000, sex) %>%
  mutate(difference = map_dbl(perm, ~ fun.difference(.)))
```



significant association between sex and acceptance rate

Feedback

Your feedback

On a more general note (not this class specific), I really like Tobi's teaching style, and have found his explanations and lessons to make this one of the most useful stats classes I've ever taken. **I was wondering if at some point this quarter, Tobi could take a few minutes to highlight additional stats courses that would either build off the material we've been introduced to, or would be good follow-up courses for students interested in more stats classes?** I'm not sure if Tobi himself will teach other classes, or has other course recommendations. Thanks!

thanks! and yes, I will do that

Logistics

Homework 6

- Topic: **Linear mixed effects models**
- will be released later today
- will be due **Thursday, March 11th, at 8pm**

My name goes here

The names of the people I have worked with go here

2021-03-04 17:21:49

Instructions

This homework is due by **Thursday, March 11th, 8:00pm**.

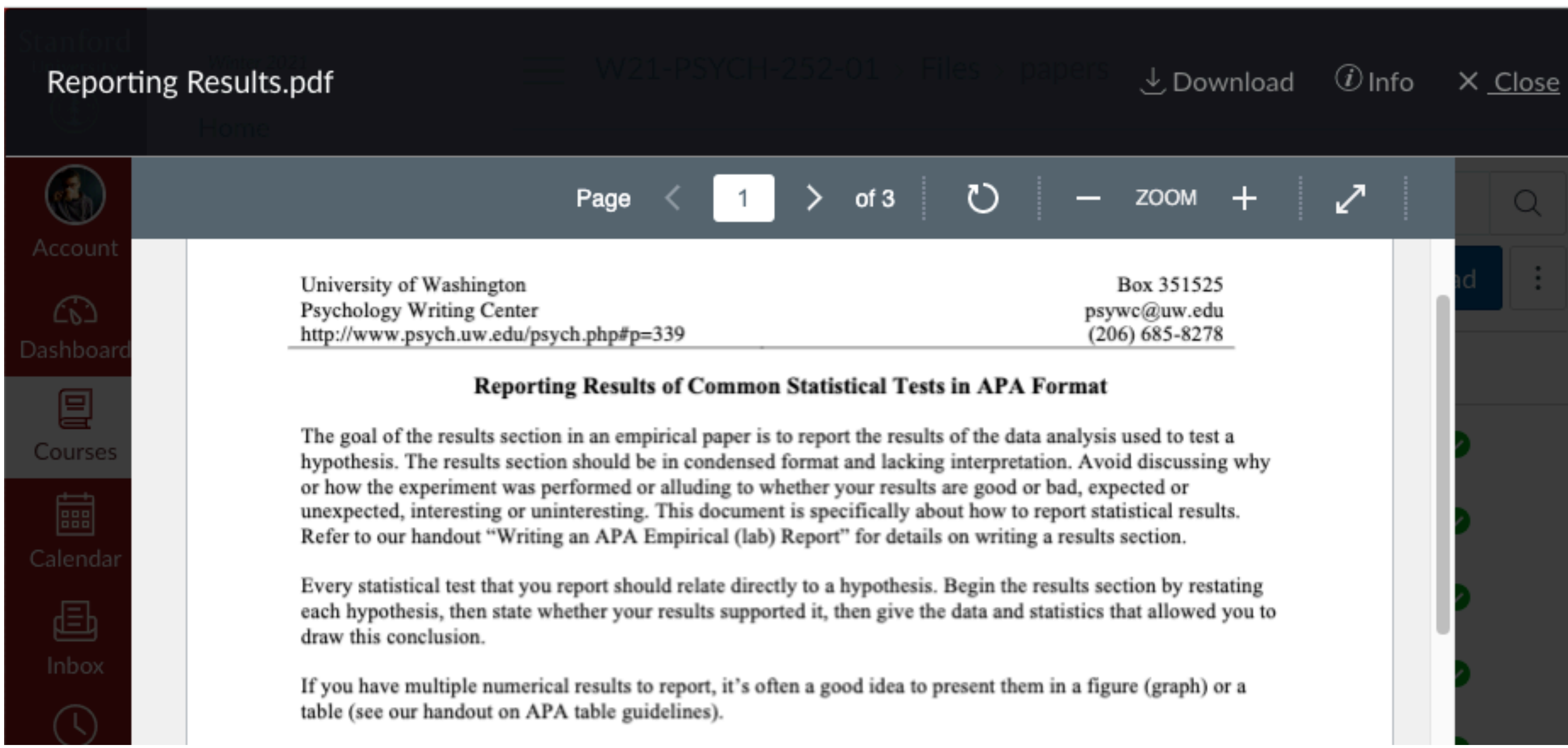
As per usual, please upload your rendered pdf on Canvas, and note that late submissions will receive 0 points.

Note:

- When asked to report results, please do so like you would in a scientific article (see examples from lectures, as well as in 'Reporting Results.pdf' on Canvas under Files > papers).
- Some code chunks contain some skeleton code. The code chunk option for these chunks is set to `eval=F` so that knitting the RMarkdown document doesn't throw any errors. Make sure to set these chunks to `eval=T` when you knit your homework, so that your calculations are shown in the pdf.
- Make sure to show the results of your calculations in the knitted pdf, for example, by using the `print()` function at the end of a code chunk.
- Some questions ask for a short written response as indicated by this prompt: **Your answer:**

Reporting results

- When we ask you to report results in the homework, try and do this like you would do in a scientific paper (reporting the statistical test, test-statistic, p-value, effect size)



University of Washington
Psychology Writing Center
http://www.psych.uw.edu/psych.php#p=339

Box 351525
psywc@uw.edu
(206) 685-8278

Reporting Results of Common Statistical Tests in APA Format

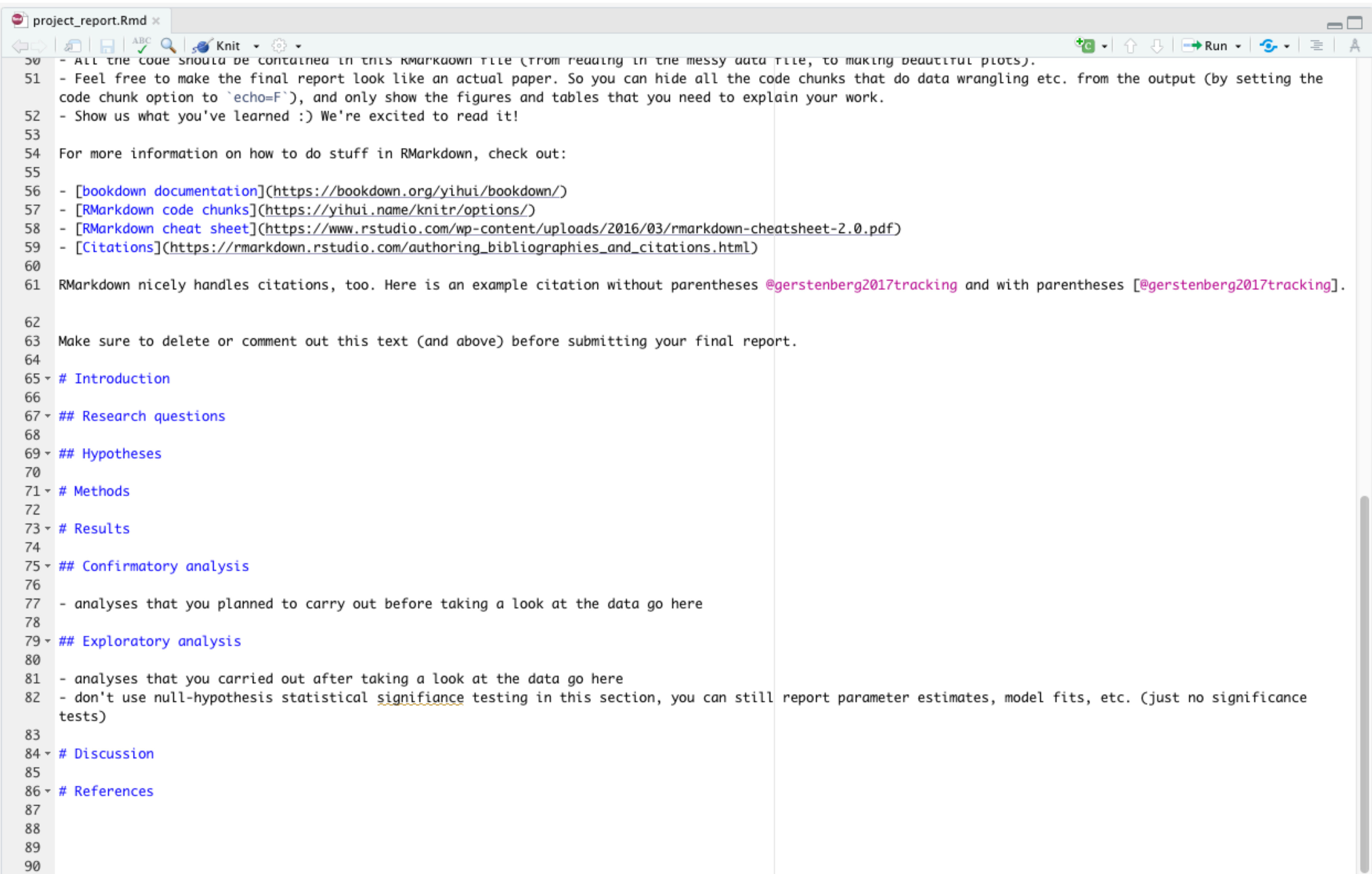
The goal of the results section in an empirical paper is to report the results of the data analysis used to test a hypothesis. The results section should be in condensed format and lacking interpretation. Avoid discussing why or how the experiment was performed or alluding to whether your results are good or bad, expected or unexpected, interesting or uninteresting. This document is specifically about how to report statistical results. Refer to our handout "Writing an APA Empirical (lab) Report" for details on writing a results section.

Every statistical test that you report should relate directly to a hypothesis. Begin the results section by restating each hypothesis, then state whether your results supported it, then give the data and statistics that allowed you to draw this conclusion.

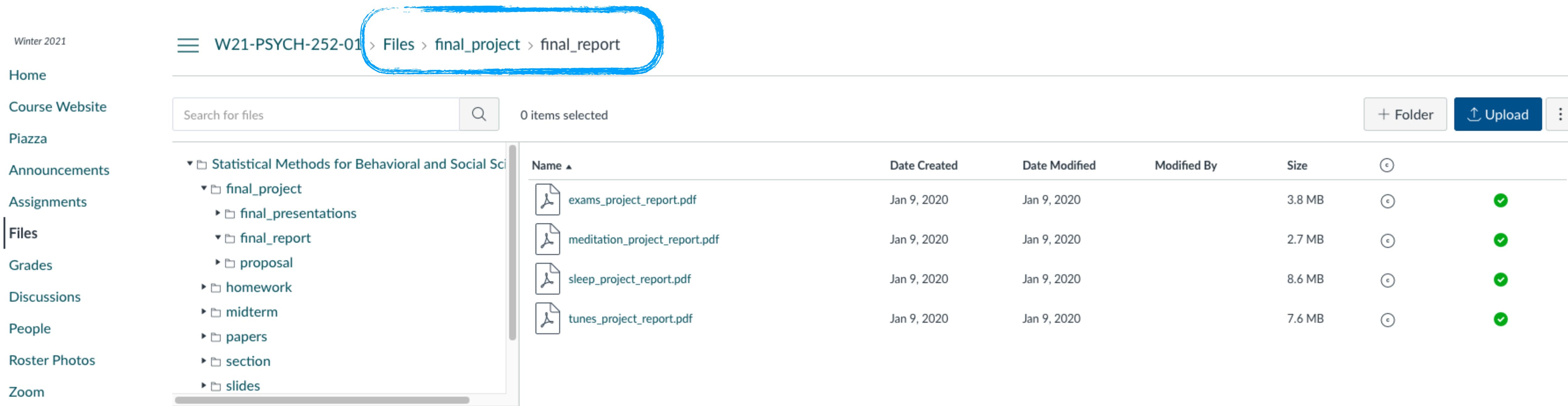
If you have multiple numerical results to report, it's often a good idea to present them in a figure (graph) or a table (see our handout on APA table guidelines).

Files > papers > Reporting results.pdf

Final project



Final project



some examples are on Canvas

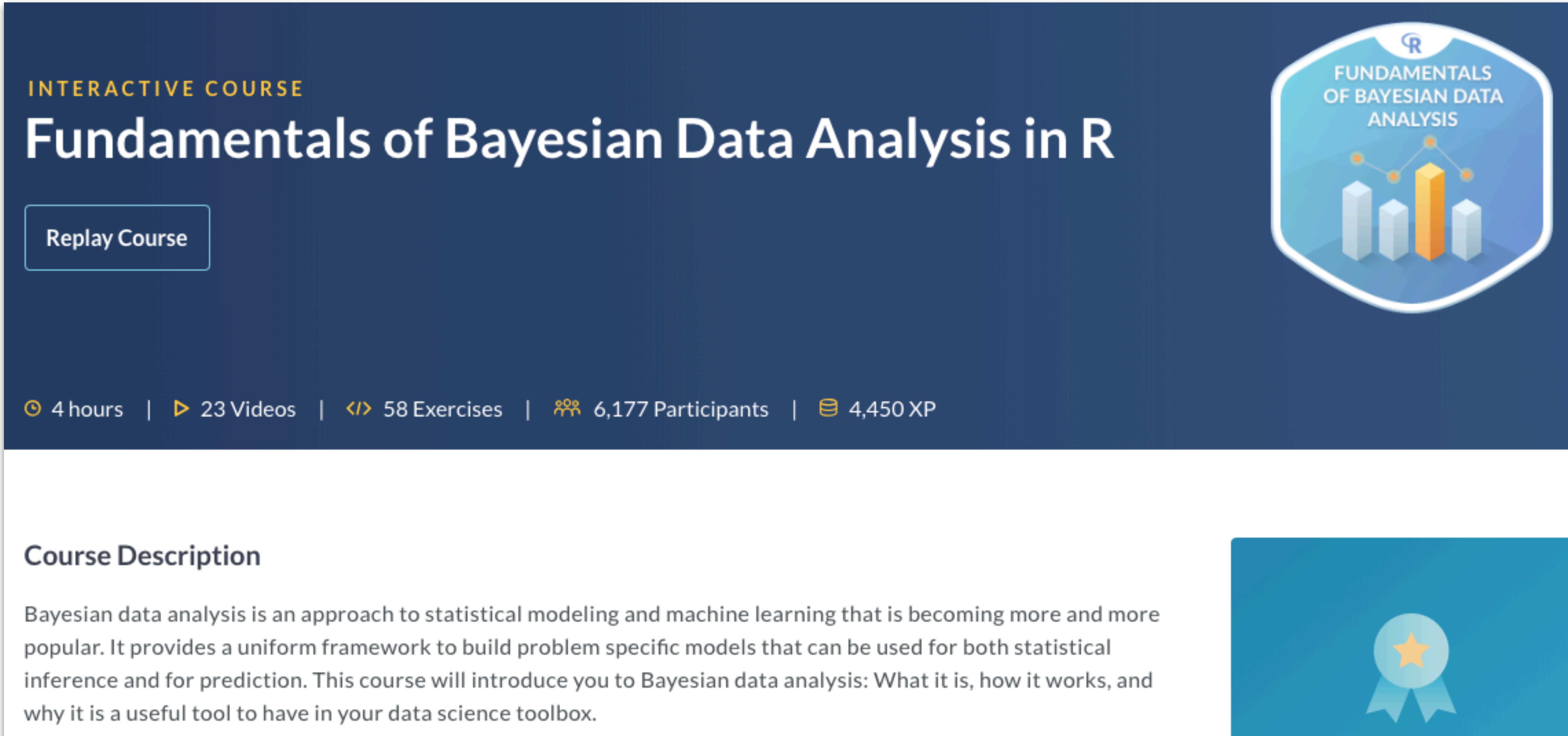
The final project is due on **Friday, March 20th at 8pm**.

Here are some guidelines:

- The length of the final report should be around 2000 words per person in the group.
- All the code should be contained in this RMarkdown file (from reading in the messy data file, to making beautiful plots).
- Feel free to make the final report look like an actual paper. So you can hide all the code chunks that do data wrangling etc. from the output (by setting the code chunk option to `echo=F`), and only show the figures and tables that you need to explain your work.
- Show us what you've learned :) We're excited to read it!

Datacamp course

recommended!!



https://www.datacamp.com/courses/fundamentals-of-bayesian-data-analysis-in-r

Great online book

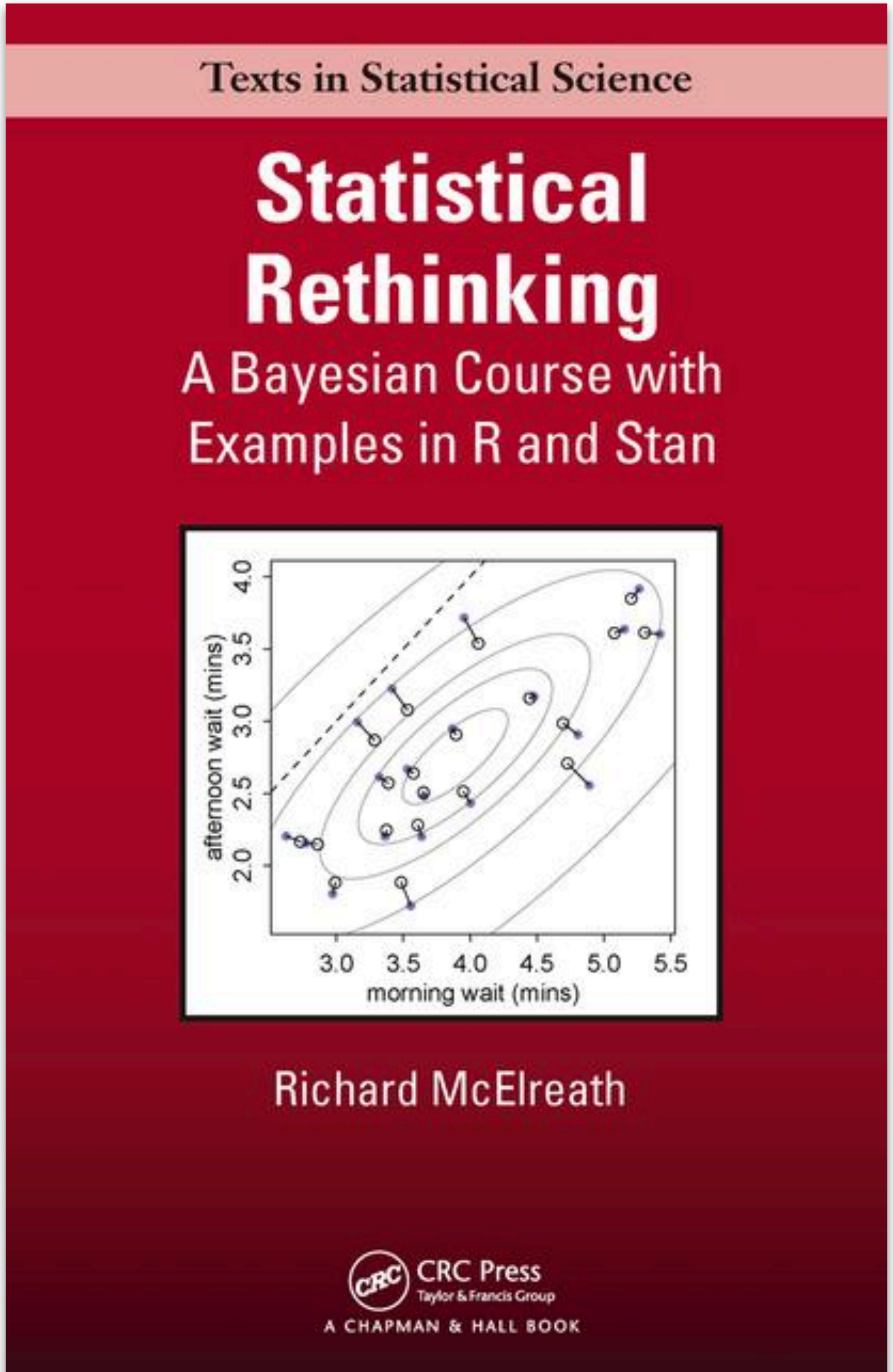
An Introduction to Data Analysis

Michael Franke

last rendered at: 2021-02-23 12:07:27

https://michael-franke.github.io/intro-data-analysis/index.html

Great book on Bayesian data analysis



- nice hands-on book (which uses R throughout)
- rewrite of all the code with `"tidyverse"` and `"BRMS"` is here: https://bookdown.org/content/4857/
- video lectures are available here: https://www.youtube.com/watch?v=4WVelCswXo4&list=PLDcUM9US4XdNM4Edgs7weiyIguLSToZRI&ab_channel=RichardMcElreath

Plan for today

- Bayesian data analysis
  - Comparison between frequentist and Bayesian data analysis
  - Quick flash from the past
  - Flipping coins
  - What affects the posterior?
  - Ingredients: likelihood, prior, inference
  - Doing Bayesian data analysis

What comes to mind when you hear Bayesian Data Analysis?

Bayes' Theorem

Conditional probabilities and an interesting formula

A preacher screaming at files of data, proselytizing to masses of data

In econ

People on Twitter like it

Challenge to standard approach - disruptive

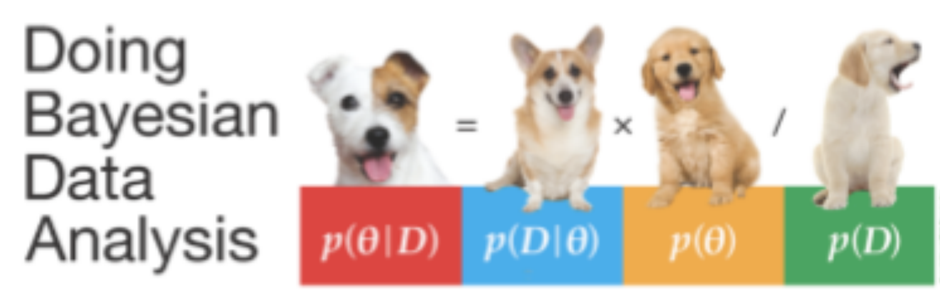
Conditionality

Mike Frank

Cool kids' game

Constraining space of possibility

Prior, posterior, inference,



BRMS

Handle the maximal effects models that lmer chokes on +1

Comparison between frequentist and Bayesian data analysis

Goal of data analysis: Inference about the world

Frequentist statistics

- generate a sampling distribution of the test statistic assuming H_0
- compare observed value of the test statistic with the sampling distribution
- reject the H_0 if probability of observed value (or more extreme values) is less than α

Bayesian statistics

- directly test hypotheses of interest
- define prior over hypotheses **p(H)**
- compute likelihood of the data for each hypothesis **p(D|H)**
- use Bayes' rule to infer the posterior over hypotheses given the data **p(H|D)**

Objections to frequentist NHST

null hypothesis significance testing

- p-value is not a measure of evidential support
  - becomes smaller as N increases
- results are often misinterpreted (both p-values and confidence intervals are not particularly intuitive)
- what we want to know: $p(\text{Hypothesis} \mid \text{Data})$
- what we calculate: $p(\text{Data} \mid \text{Null Hypothesis})$

Wasserstein, Schirm, & Lazar (2019) Moving to a World Beyond p < 0.05. *The American Statistician*

Frequentists vs. Bayesian

- both want to evaluate the evidence for a hypothesis using a sample of data **p(H|D)**
- it's often easier to calculate the inverse: the probability of the data given a hypothesis **p(D|H)**
- frequentists use a rule of thumb (p-value) to make a decision
- Bayesians use Bayes' rule

Why don't more people use Bayesian Statistics?

- supposedly more difficult
  - relies on the logic of probability theory
- reliance on a *prior*
- reliance on computing and simulation
  - we can't just use SPSS
  - but we can use JASP (Just Another Statistics Program)

and we've already learned how to simulate and visualize data in this class!

What are (some of) the benefits of Bayesian data analysis?

- intuitive model testing and comparison
  - compare simulated data with the real data
- straightforward interpretation of results
  - Bayesian credible intervals vs. Confidence intervals
- more model flexibility
  - adequately express assumptions about the data-generating process
- better predictions!

Flash from the past

Clue guide to probability

we derived this using the definition of conditional probability

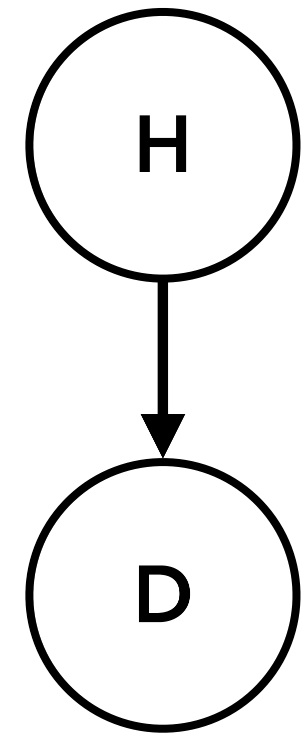
$$p(B\,|\,A) = \frac{p(A\,|\,B)\cdot p(B)}{p(A)}$$

$$\underbrace{p(H\,|\,D)}_{\text{posterior}} = \frac{\overbrace{p(D\,|\,H)}^{\text{likelihood}}\cdot \overbrace{p(H)}^{\text{prior}}}{p(D)}$$

subjective probability interpretation

H = Hypothesis

D = Data



formal framework for learning from data

updating our prior belief $p(H)$, to a posterior belief $p(H\,|\,D)$ given some data

Clue guide to probability

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is “95% accurate”; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that **$P(T|D) = 0.95$** and **$P(\neg T|\neg D) = 0.95$**. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

Clue guide to probability

what we know

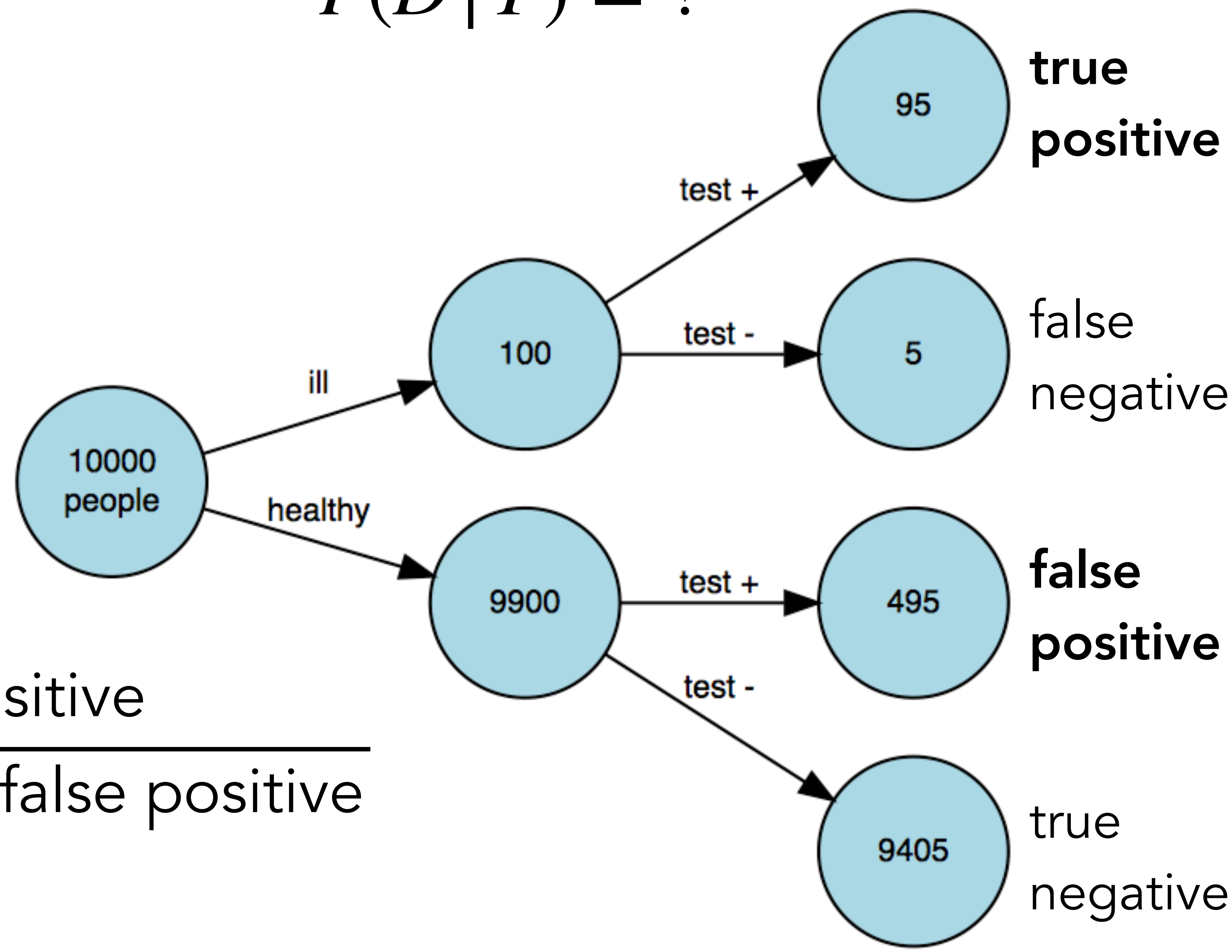
$$P(D) = 0.01$$

$$P(T \mid D) = 0.95$$

$$P(T \mid \neg D) = 0.05$$

what we want to know

$$P(D \mid T) = ?$$



$$P(D \mid T) = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$= \frac{95}{95 + 495}$$

$$\approx 0.16$$

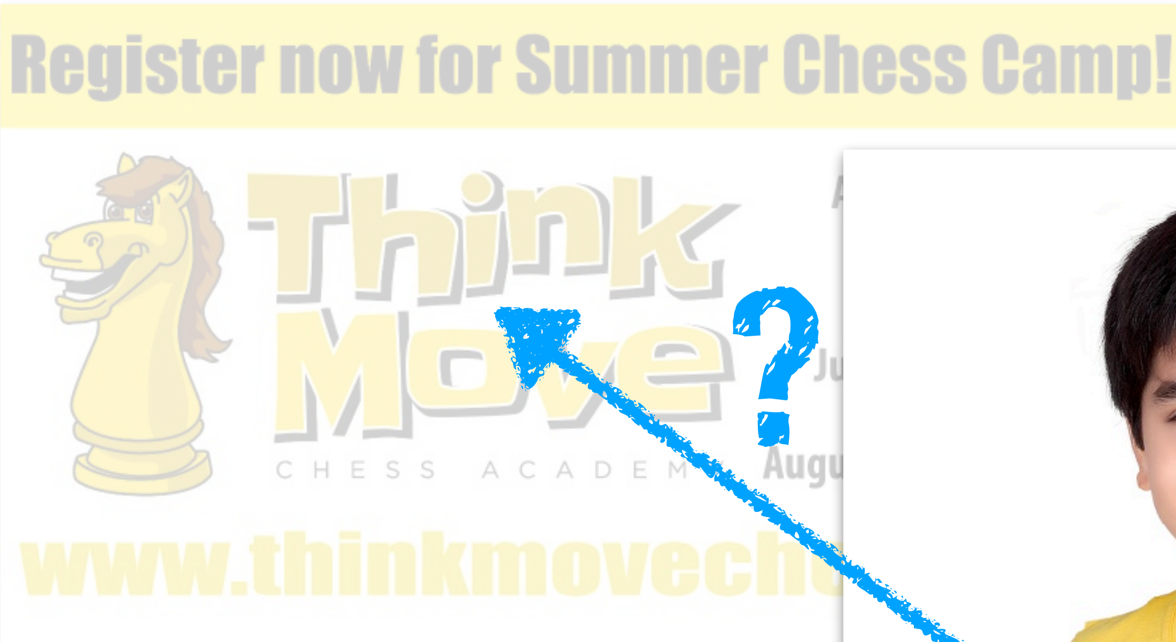
Summer camp

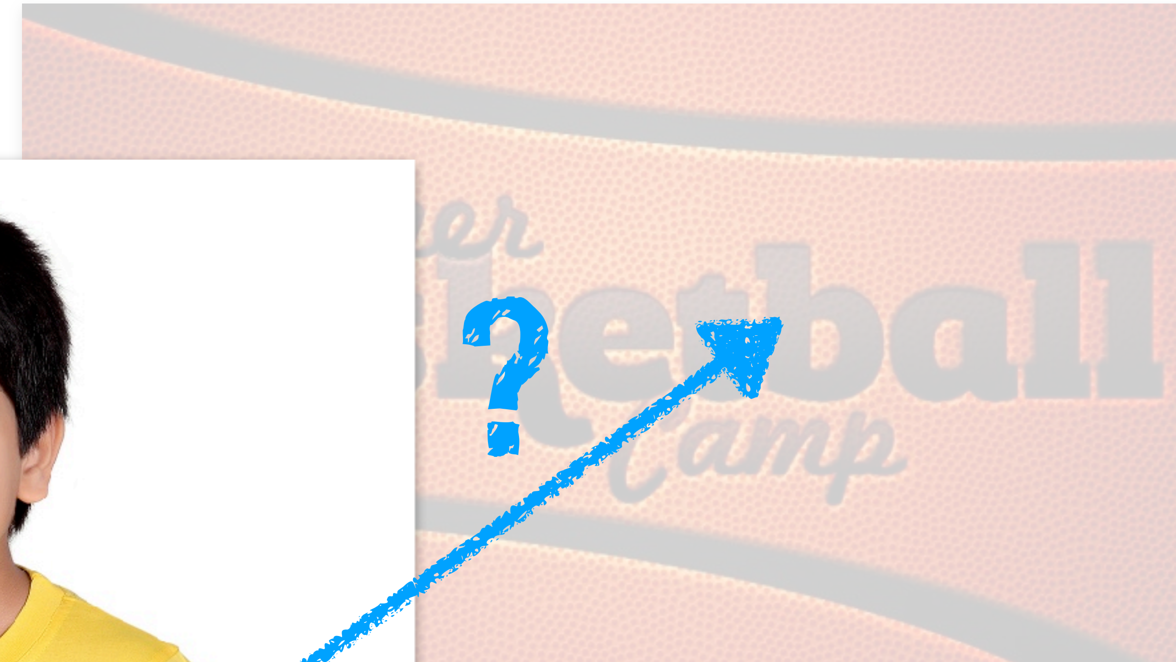
twice as many kids go to the basketball camp

$X \sim$ Normal$(\mu = 170,\ \sigma = 8)$

$X \sim$ Normal$(\mu = 180,\ \sigma = 10)$

Summer camp





twice as ma … etball camp

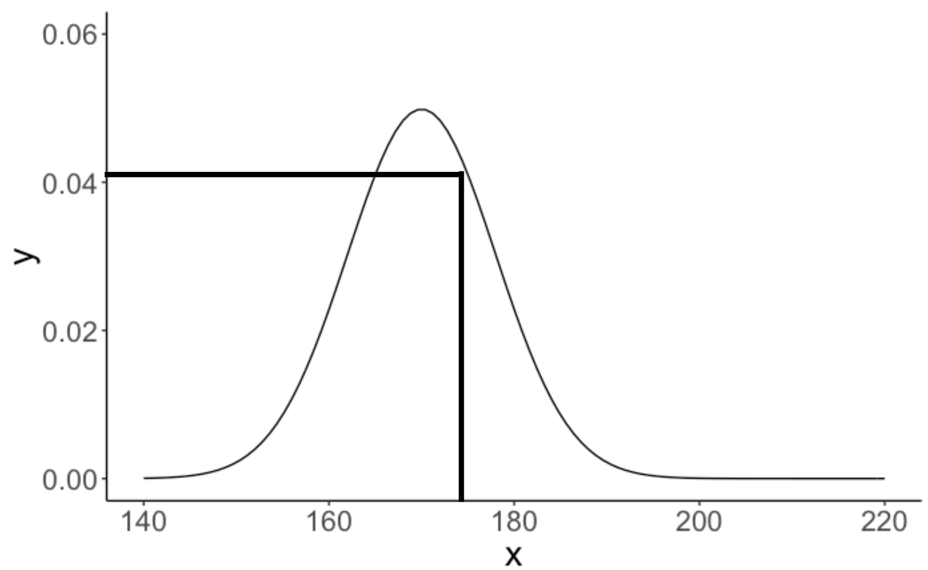
$X \sim$ Normal($\mu = 170$ … al($\mu = 180$, $\sigma = 10$)

Summer camp

prior

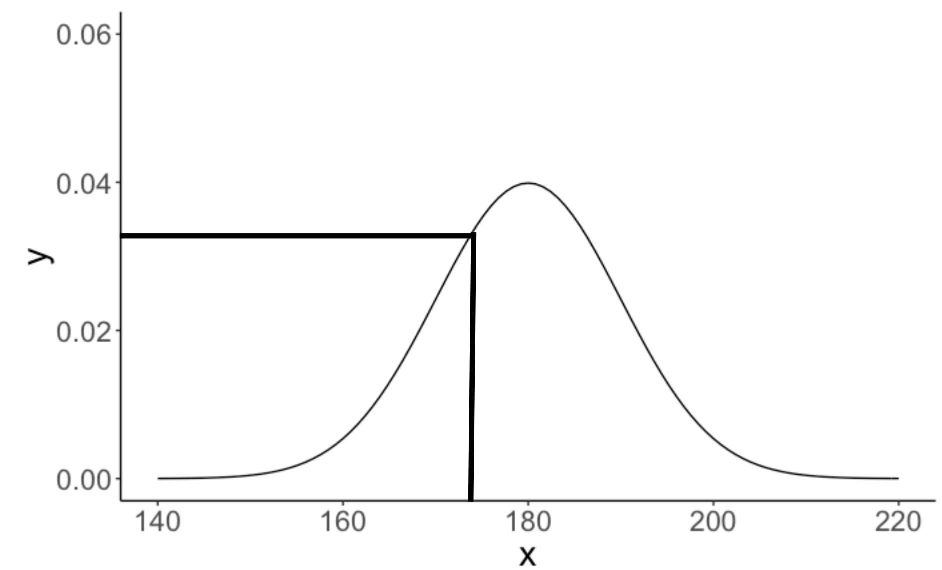
$$p(\text{chess}) = \frac{1}{3} \qquad p(\text{basketball}) = \frac{2}{3}$$

likelihood



`dnorm(175, mean = 170, sd = 8)`

= 0.041



`dnorm(175, mean = 180, sd = 10)`

= 0.035

posterior

$$p(\text{basketball} \mid 175) = \frac{p(175 \mid \text{basketball}) \cdot p(\text{basketball})}{p(175 \mid \text{basketball}) \cdot p(\text{basketball}) + p(175 \mid \text{chess}) \cdot p(\text{chess})}$$

$$p(\text{basketball} \mid 175) = \frac{0.035 \cdot 2/3}{0.035 \cdot 2/3 + 0.041 \cdot 1/3} \approx 0.63$$

send the kid to the basketball gym!

Flipping coins

Flipping coins

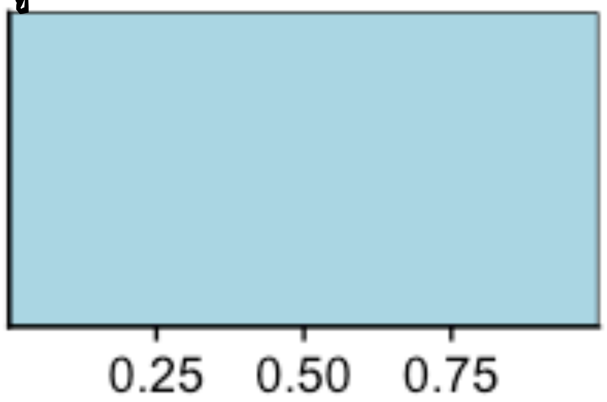
What's the probability of my coin coming up heads?

$$\theta = ?$$

Learning from data

How does/should our belief change as evidence comes in?

prior



Today's posterior is tomorrow's prior.

$$p(\theta \,|\, \mathsf{n}_{\mathsf{success}} = 6,\, \mathsf{n}_{\mathsf{trials}} = 8)$$

Coin flip example

Which coin did I flip?

Hypotheses

$\theta = 0.3$

$\theta = 0.5$

$\theta = 0.9$

Data

#8 tails, #2 heads

Bayesian Recipe

- Hypotheses
- Prior over hypotheses
- Data
- Likelihood of the data given each hypothesis
- Posterior over hypotheses given the data

+ a healthy dose of Bayes' rule

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}$$

Coin flip example

```r
# data
data = rep(0:1, c(8, 2))

# parameters
theta = c(0.1, 0.5, 0.9)

# prior
prior = c(0.25, 0.5, 0.25)

# likelihood
likelihood = dbinom(sum(data == 1), size = length(data), prob = theta)

# posterior
posterior = likelihood * prior / sum(likelihood * prior)
```

8 tails and 2 heads

hypotheses

prior over the hypotheses

binomial distribution

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

law of total probability:

$$p(D) = \sum_{i=1}^{n} p(D|H_i) \cdot p(H_i)$$

Coin flip example

```r
# data
data = rep(0:1, c(8, 2))

# parameters
theta = c(0.1, 0.5, 0.9)

# prior
prior = c(0.25, 0.5, 0.25)

# likelihood
likelihood = dbinom(sum(data == 1), size = length(data), prob = theta)

# posterior
posterior = likelihood * prior / sum(likelihood * prior)
```

8 tails and 2 heads

hypotheses

prior over the hypotheses

binomial distribution

multiply

re-normalize

theta	prior	likelihood	prior_x_like	posterior
0.1	0.25	0.19	0.0475	0.69
0.5	0.50	0.04	0.02	0.31
0.9	0.25	0.00	0.00	0.00

Binomial distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

out of *n* trials *k* successes occur with probability p^k, and *n*-*k* failures occur with probability $(1\text{-}p)^{n\text{-}k}$

all the possible ways to get *k* successes out of *n* trials

What's the probability of getting 2 heads, when flipping a fair coin 10 times?

```
dbinom(x = 2, size = 10, prob = 0.5)
```

$n = 10$

$k = 2$

$p = 0.5$

$$P(X = 2) = \binom{10}{2} 0.5^2 (1-0.5)^{10-2} = 0.044$$

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?

what the model knows **before** having seen the data

learning by **conditioning** on the data

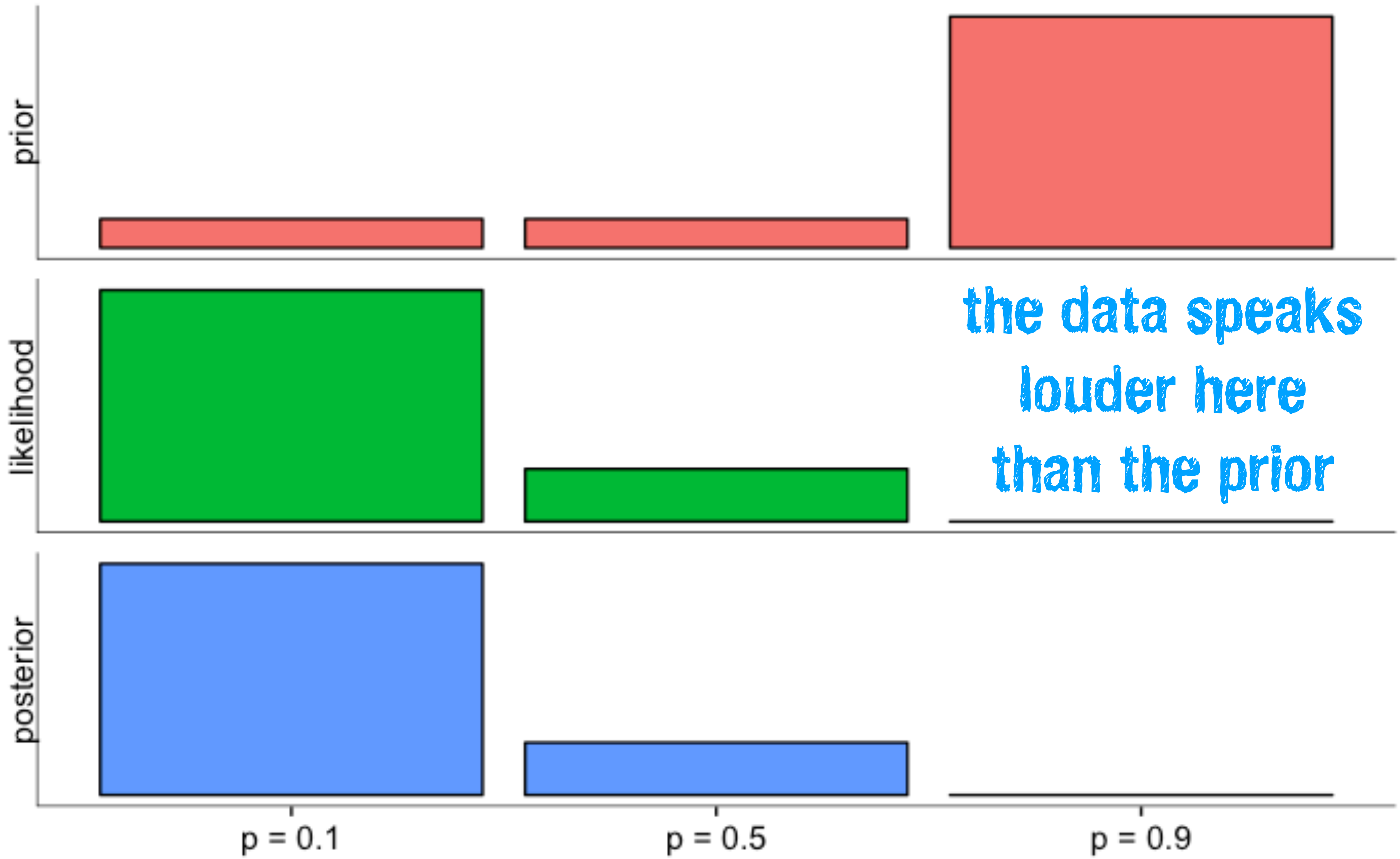
what the model knows **after** having seen the data

posterior = multiplicative weighting of prior and likelihood

Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?

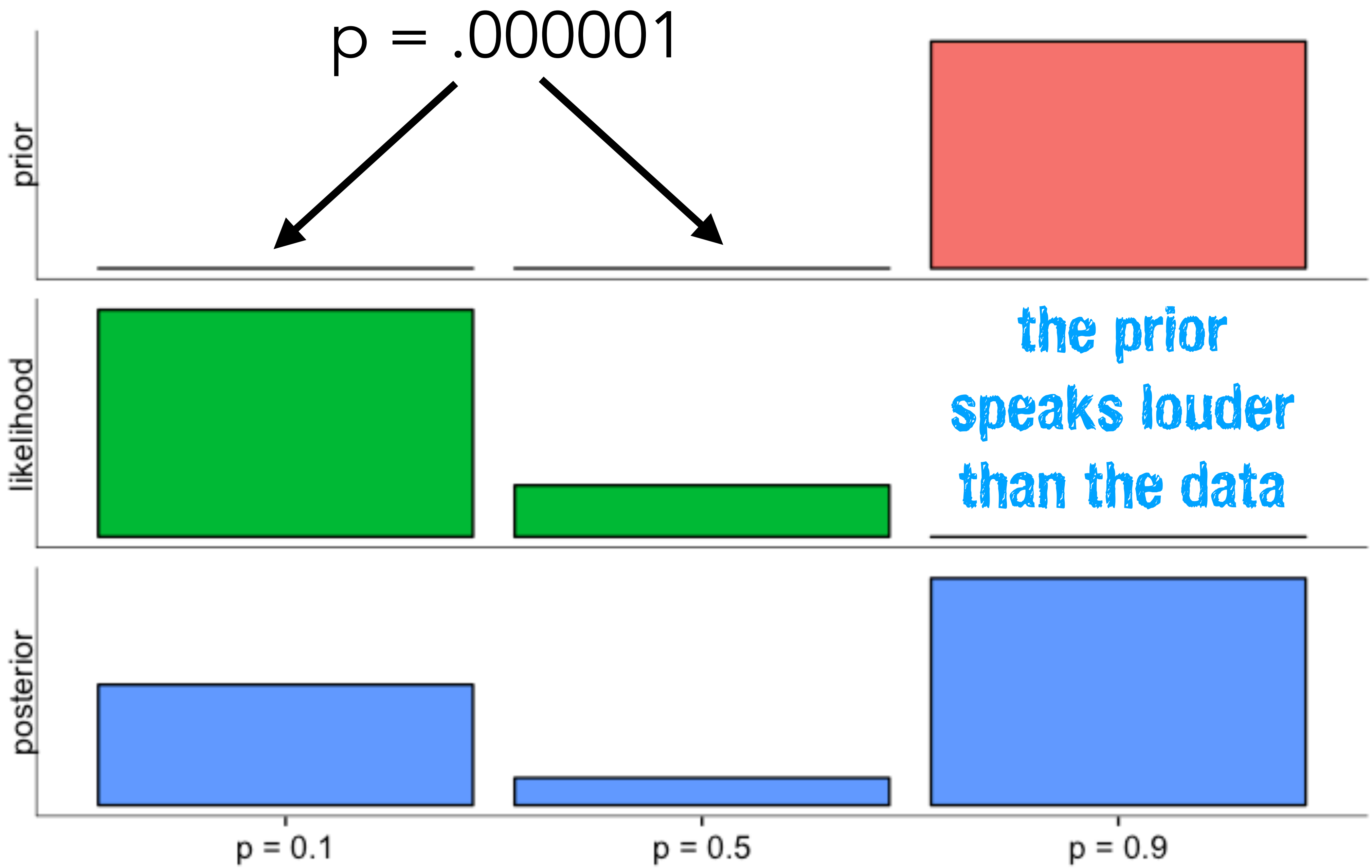


posterior = multiplicative weighting of prior and likelihood

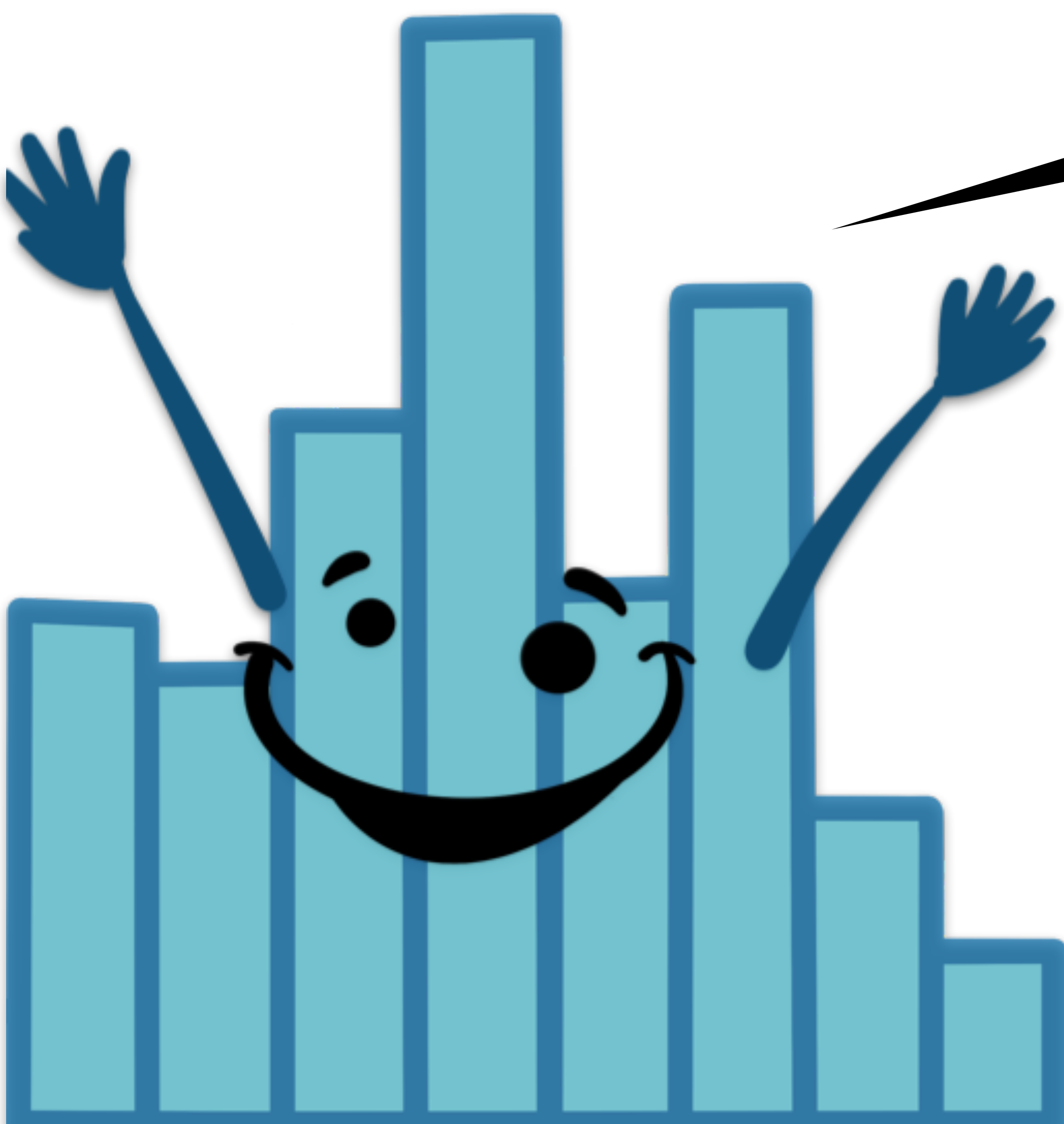
Coin flip example

data: #8 tails, #2 heads

Which coin was flipped?



posterior = multiplicative weighting of prior and likelihood



01:00

stretch break!

What affects the posterior?

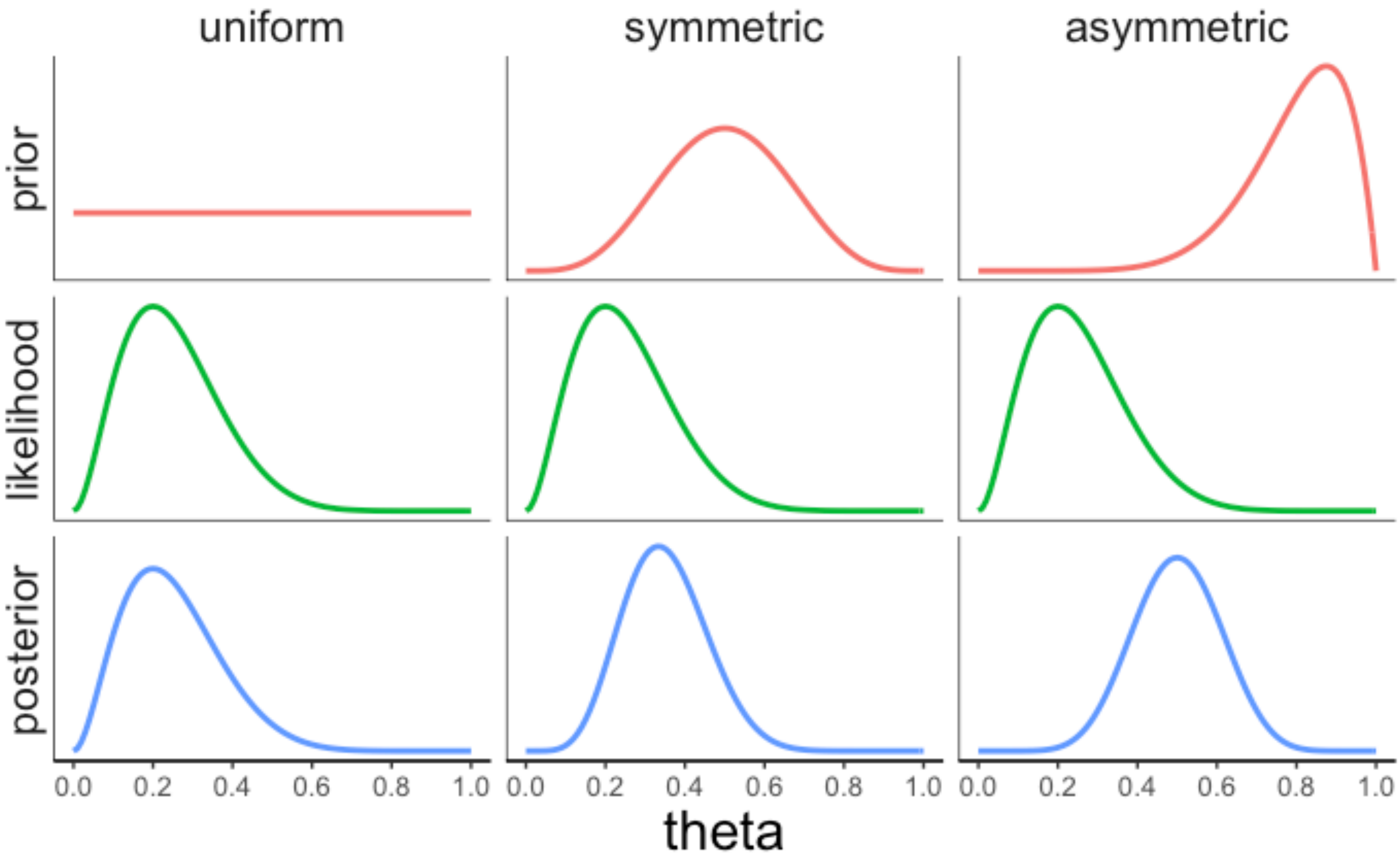
What affects the posterior?

1. the prior over hypotheses
2. the likelihood of the data given each hypothesis

$$\underbrace{p(H \mid D)}_{\text{Posterior}} = \frac{\overbrace{p(D \mid H)}^{\text{Likelihood}} \cdot \overbrace{p(H)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Normalizing constant}}}$$

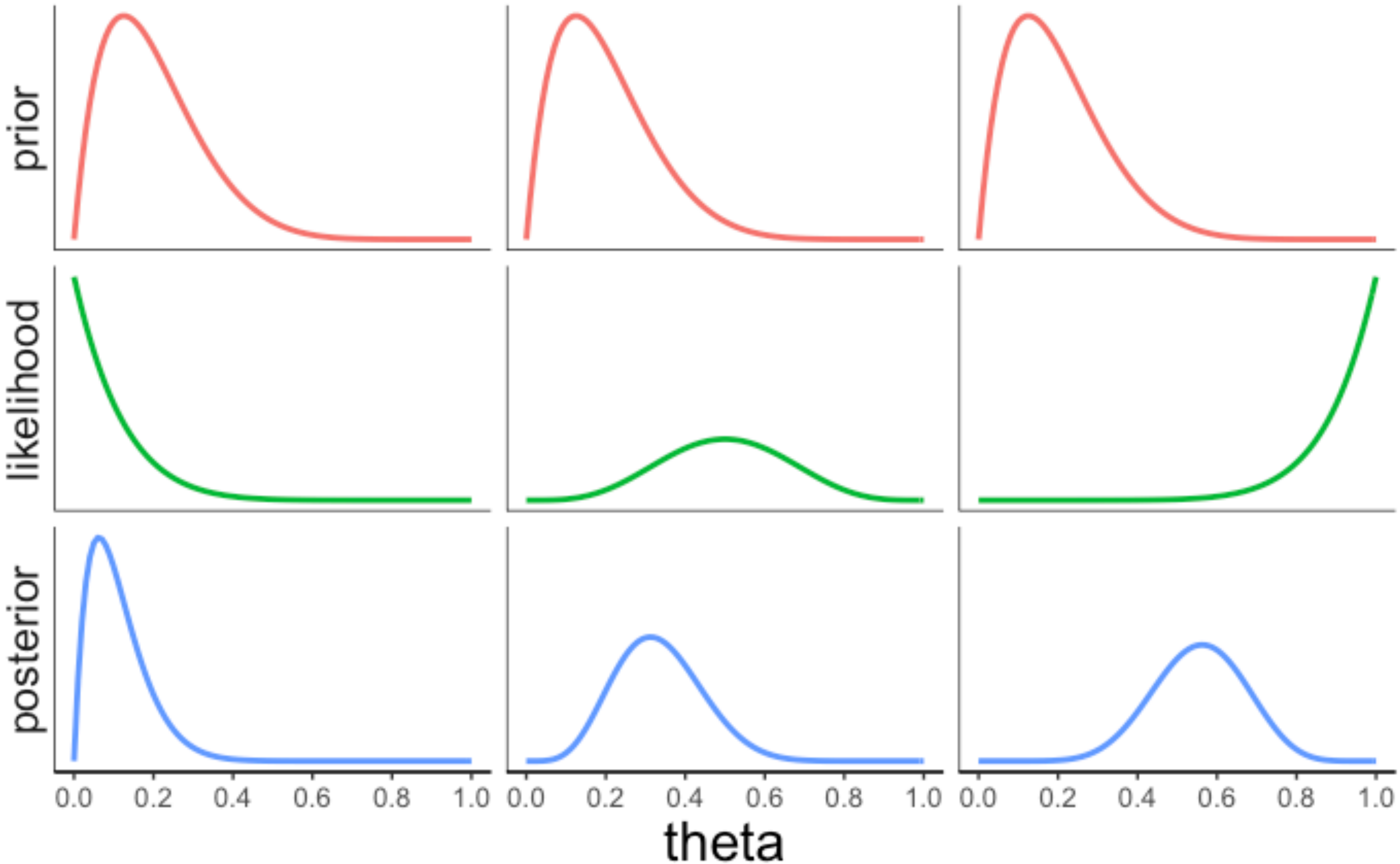
The effect of the **prior**

same data, different priors

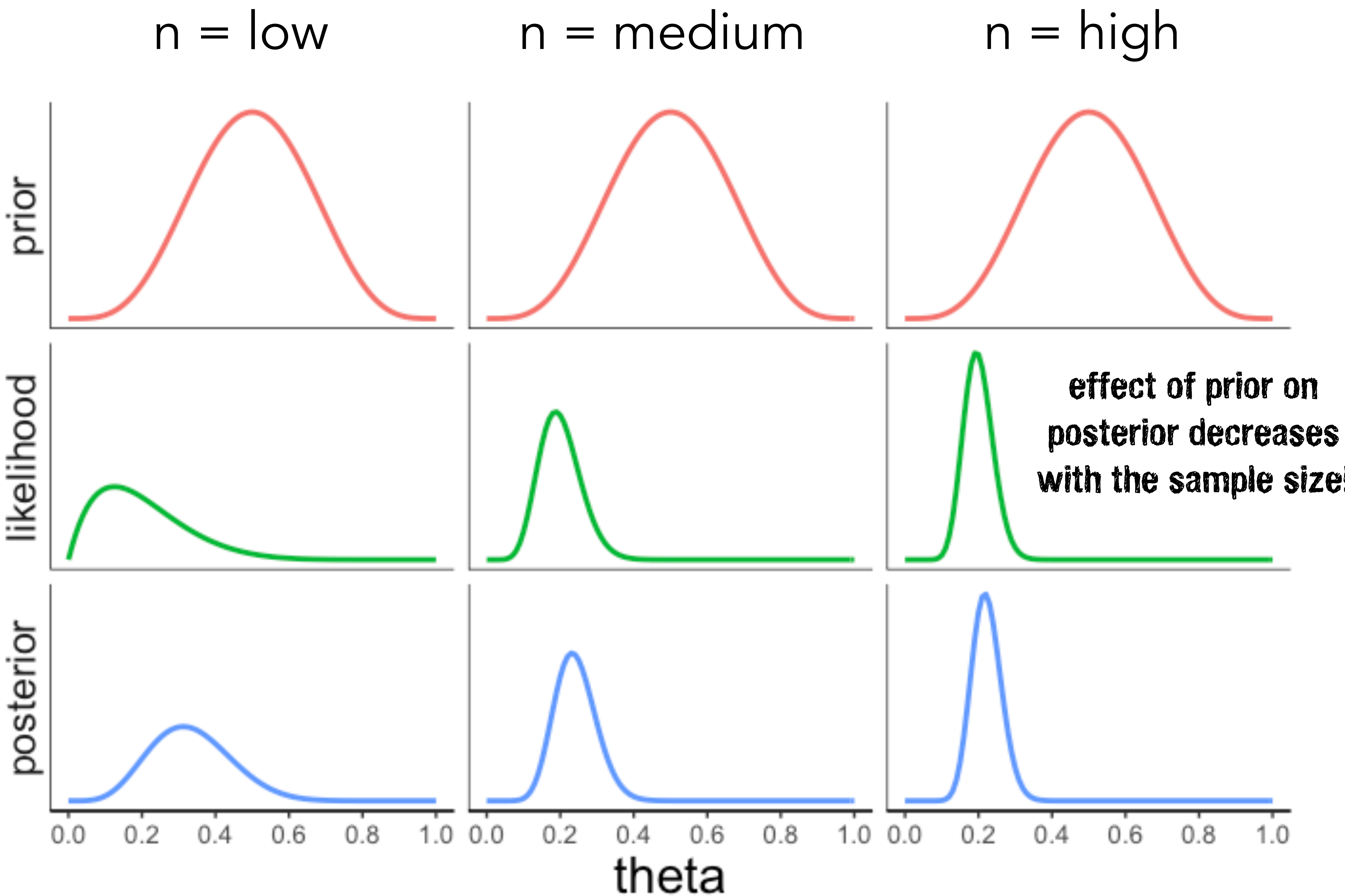


The effect of the **likelihood**

same prior, different data



The effect of **sample size**



Ingredients: likelihood, prior, inference

Ingredients

Likelihood Prior

Posterior

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

Normalizing constant

$$\underbrace{p(H \mid D)}_{\text{Posterior}} = \frac{\overbrace{p(D \mid H)}^{\text{Likelihood}} \cdot \overbrace{p(H)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Normalizing constant}}}$$

Likelihood

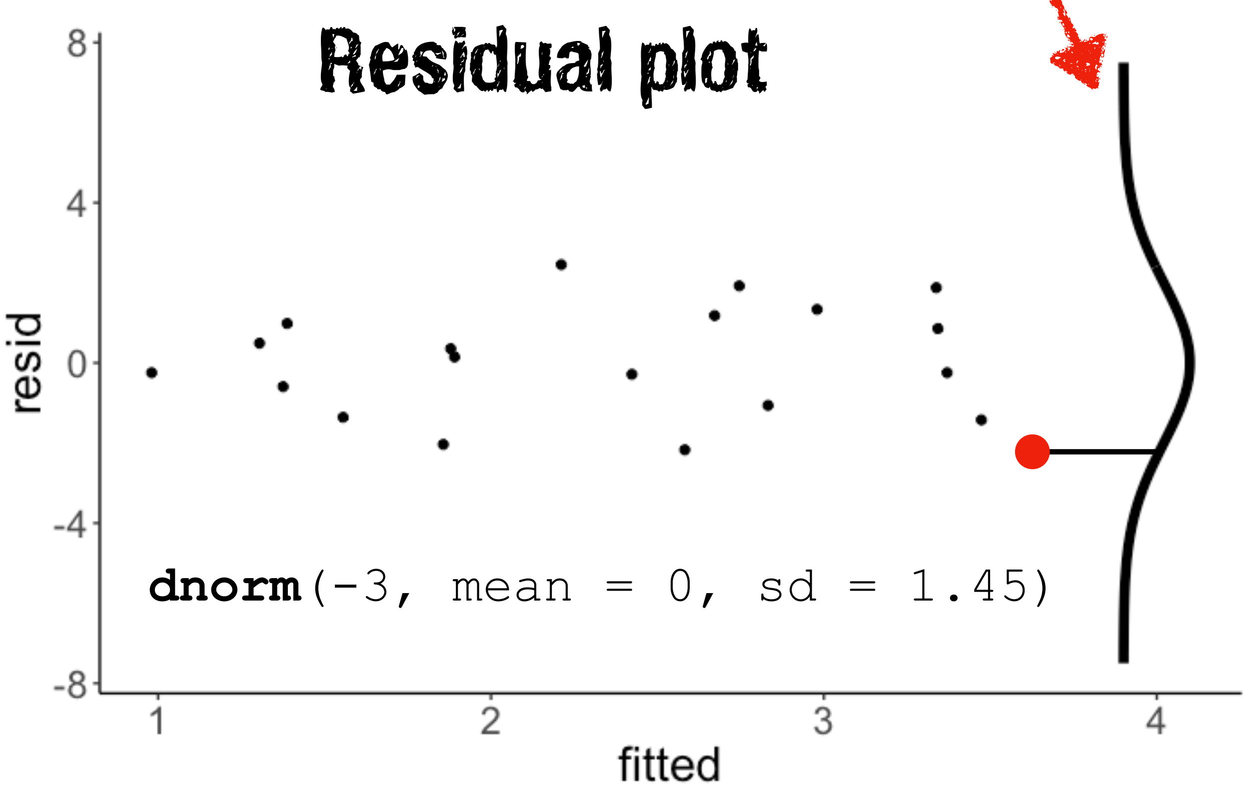
- **What probabilistic model describes best how the data were generated?**
  - What assumptions can you make about the data?
    - What's the nature of your dependent variable (e.g. binary, ordered, continuous)?
  - Does the model re-create the behavior of interest?

Likelihood

Gaussian distribution

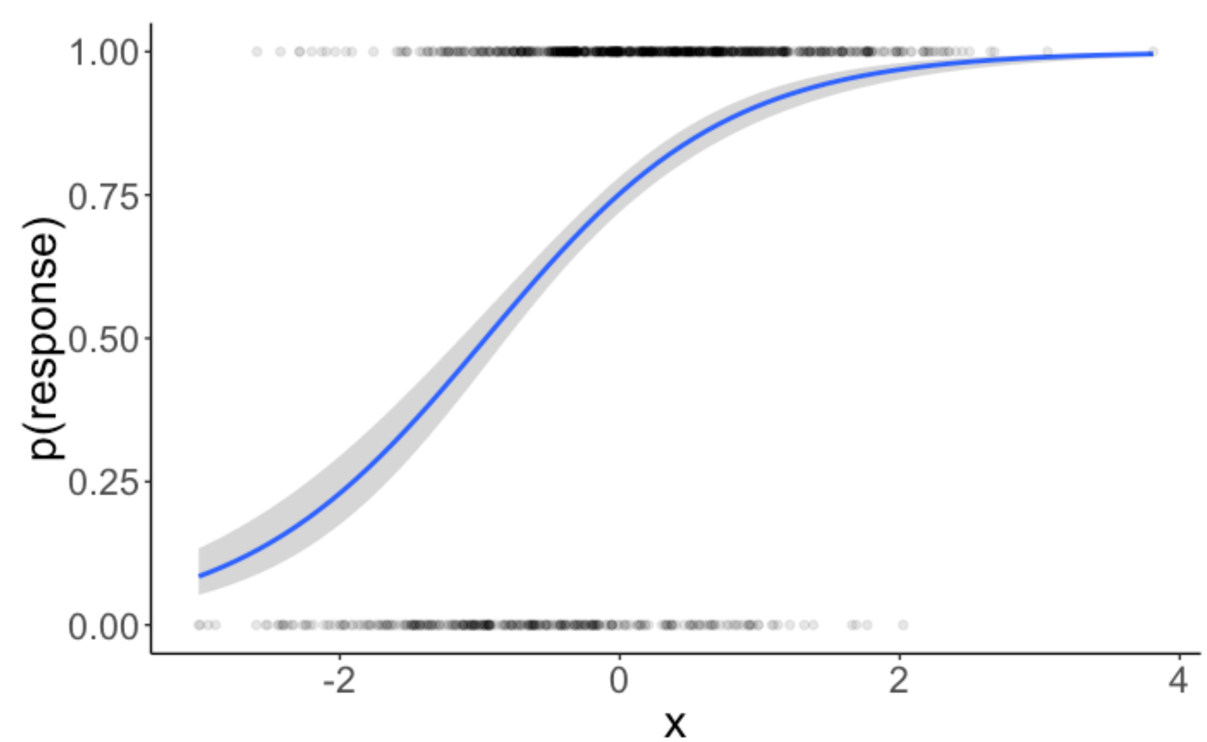
$$Y_i = b_0 + b_1 \cdot x_i + e_i$$

$$e_i \sim \mathcal{N}(0, \sigma)$$

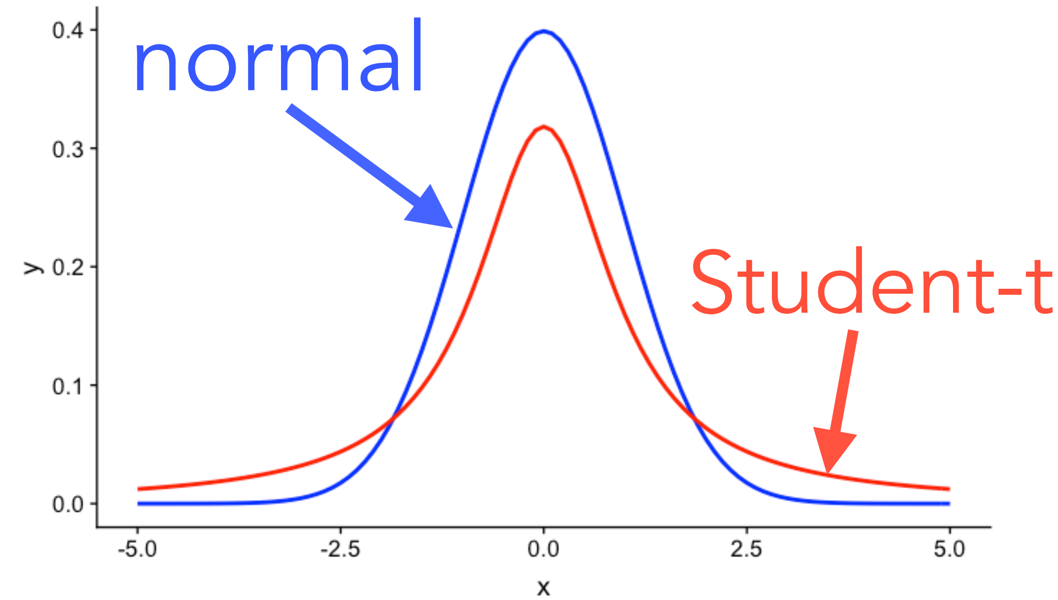


Binomial distribution

```
fit.glm = glm(formula = survived ~ 1 + fare,
              family = "binomial",
              data = df.titanic)
```

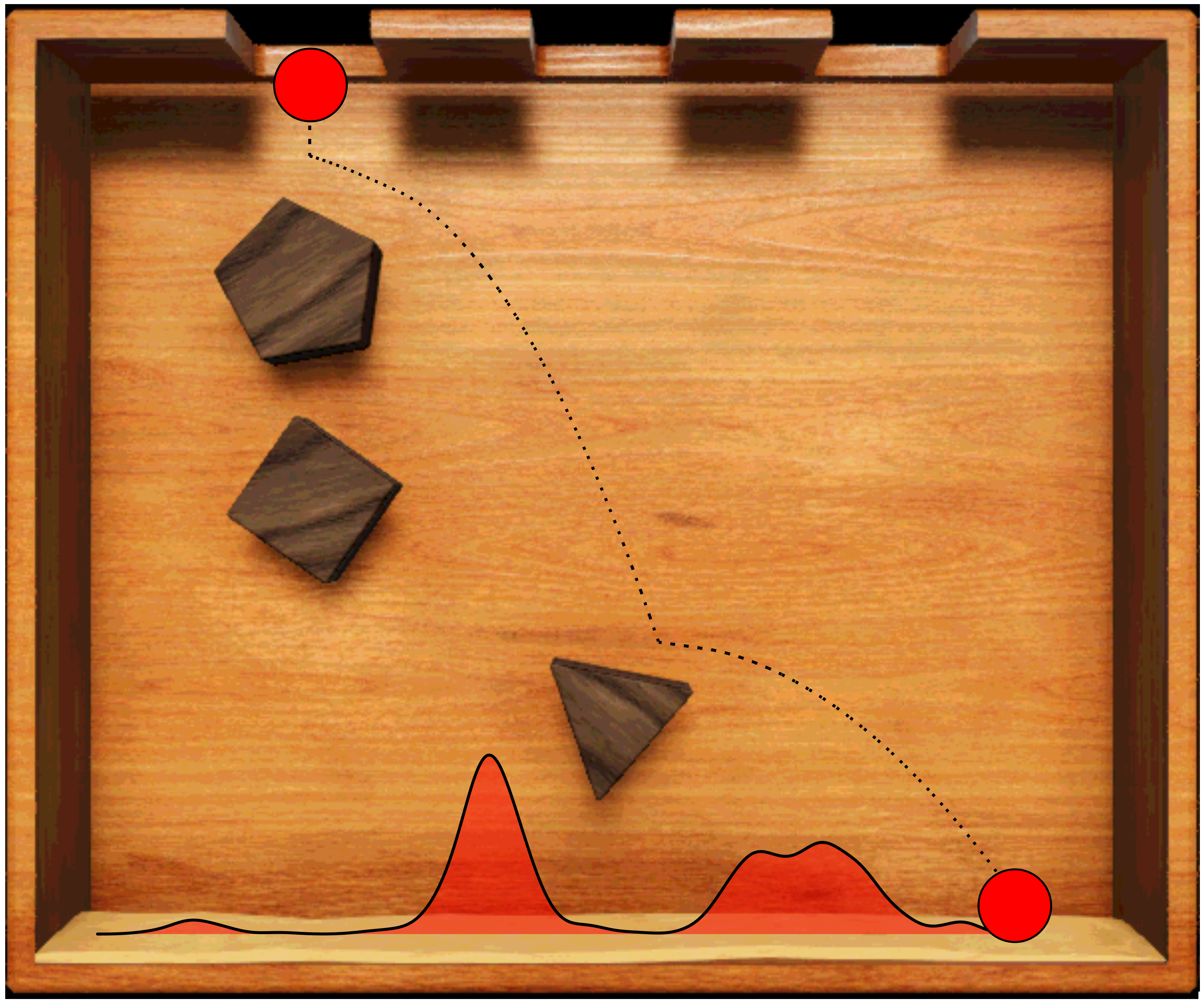


Likelihood



- **Bernoulli**:
  - binary data
  - a single trial
- **Poisson**: count of discrete events
- **Beta-binomial**: like binomial but probability of success may change across trials
- **Student-*t***:
  - same as Normal
  - handles greater variability in the data (distribution has **fat tails**)
- ...

Prediction: Where will the ball land?

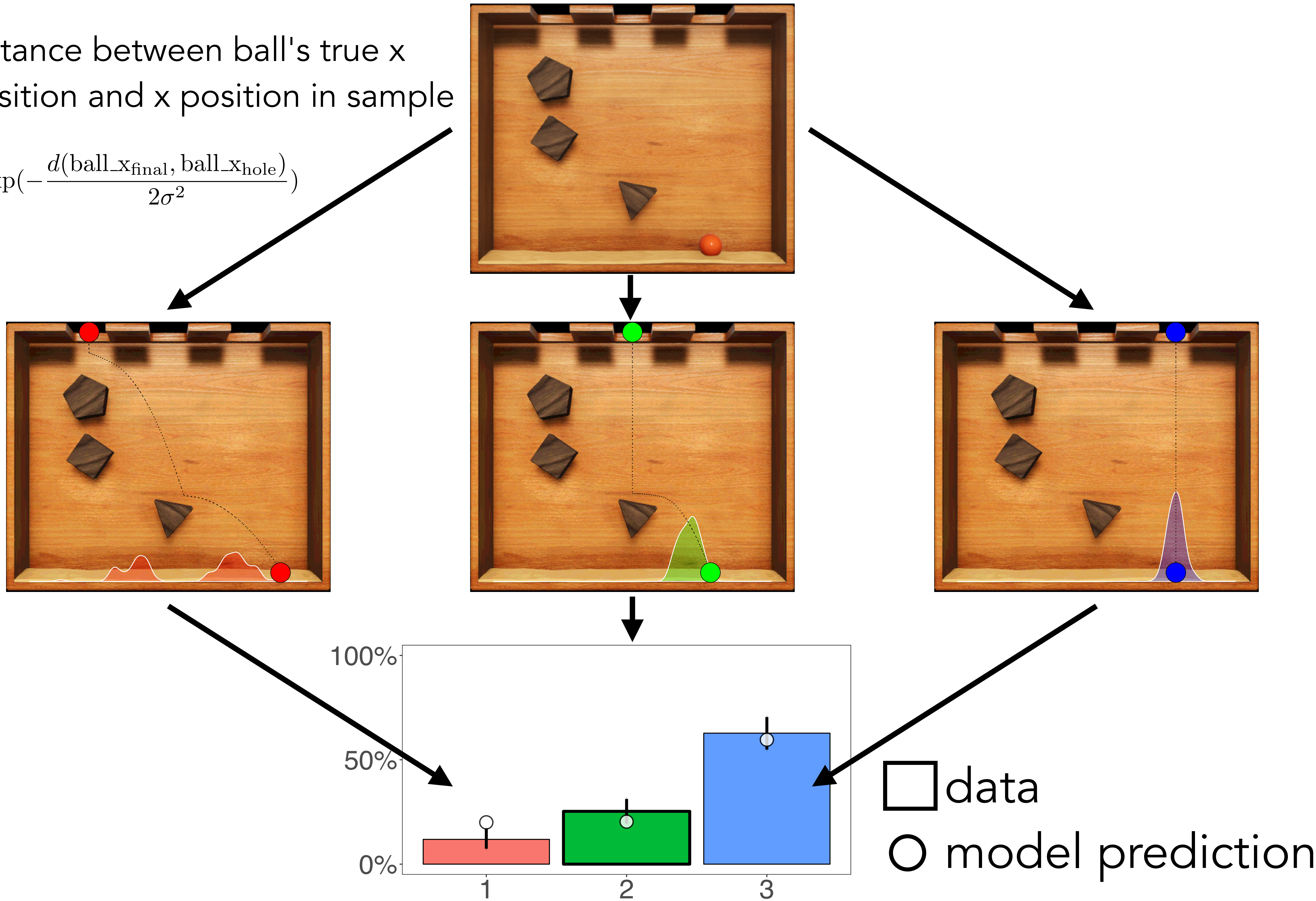


Aggregated responses

Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

$$\exp\left(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2}\right)$$



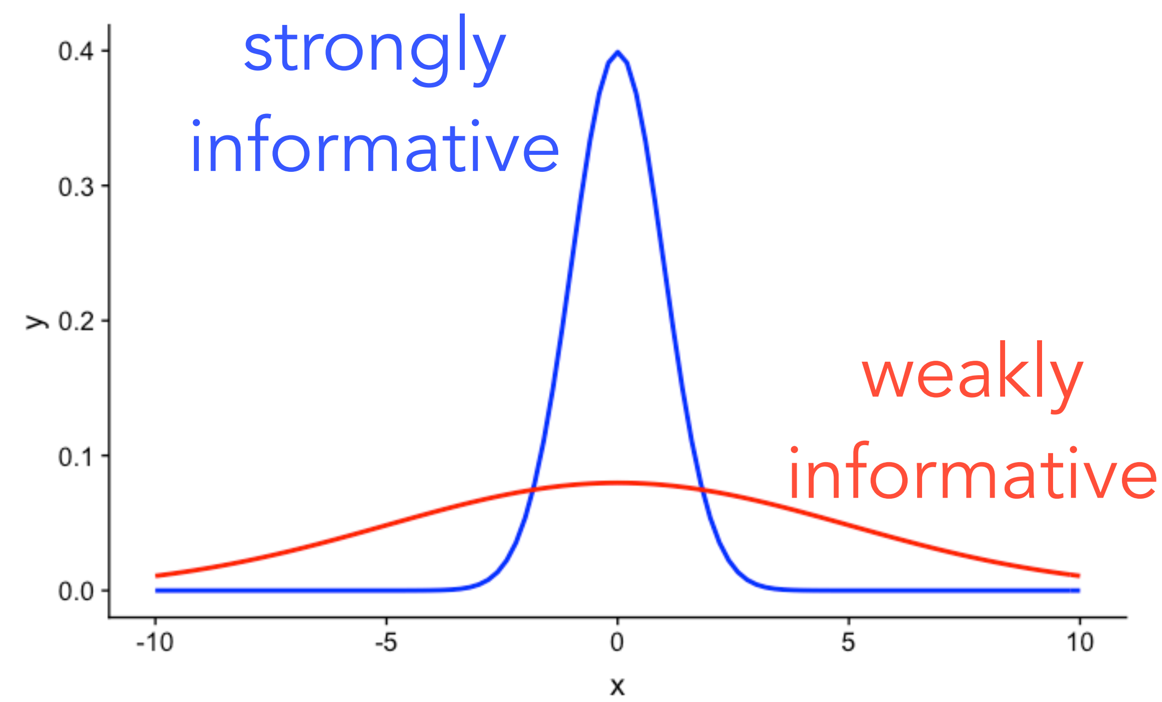
Prior

$$\underbrace{p(H \mid D)}_{\text{Posterior}} = \frac{\overbrace{p(D \mid H)}^{\text{Likelihood}} \cdot \overbrace{p(H)}^{\text{Prior}}}{\underbrace{p(D)}_{\text{Normalizing constant}}}$$

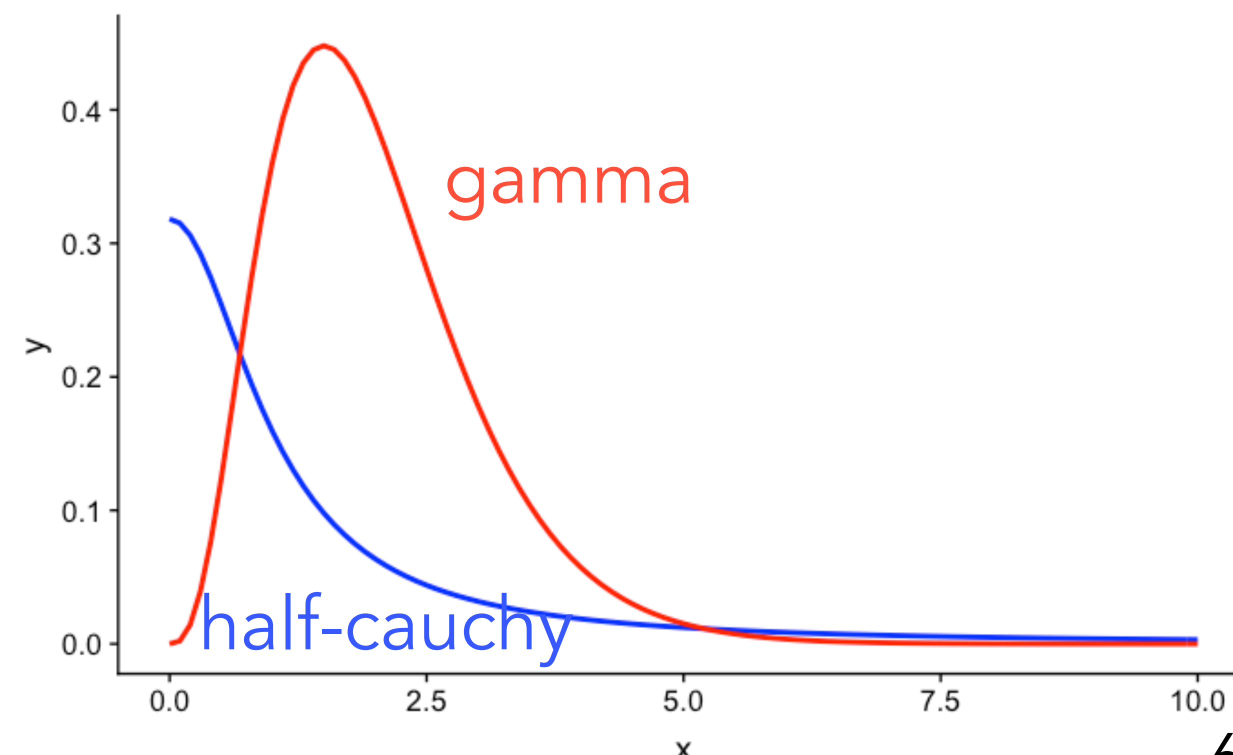
Prior

- **uniform**:
  - continuous or discrete
  - bounded between minimum and maximum
- **gaussian:**
  - sd determines how informative the prior is
- **gamma, half-cauchy**:
  - for parameters we know are positive

for **beta coefficients** in a regression



for **standard deviation** of the Gaussian



Inference

$$\overset{\textbf{Posterior}}{p(H \mid D)} = \frac{\overset{\textbf{Likelihood}}{p(D \mid H)} \cdot \overset{\textbf{Prior}}{p(H)}}{\underset{\textbf{Normalizing constant}}{p(D)}}$$

the devil is in the denominator ...

Doing Bayesian inference

Discrete hypothesis space

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{\sum_{i=1}^{n} p(D \mid H_i) \cdot p(H_i)}$$

sum over all possibilities

Continuous hypothesis space

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{\int_{-\infty}^{\infty} p(D \mid H_i) \cdot p(H_i) dH_i}$$

integral over all possibilities

Discretizing the parameters

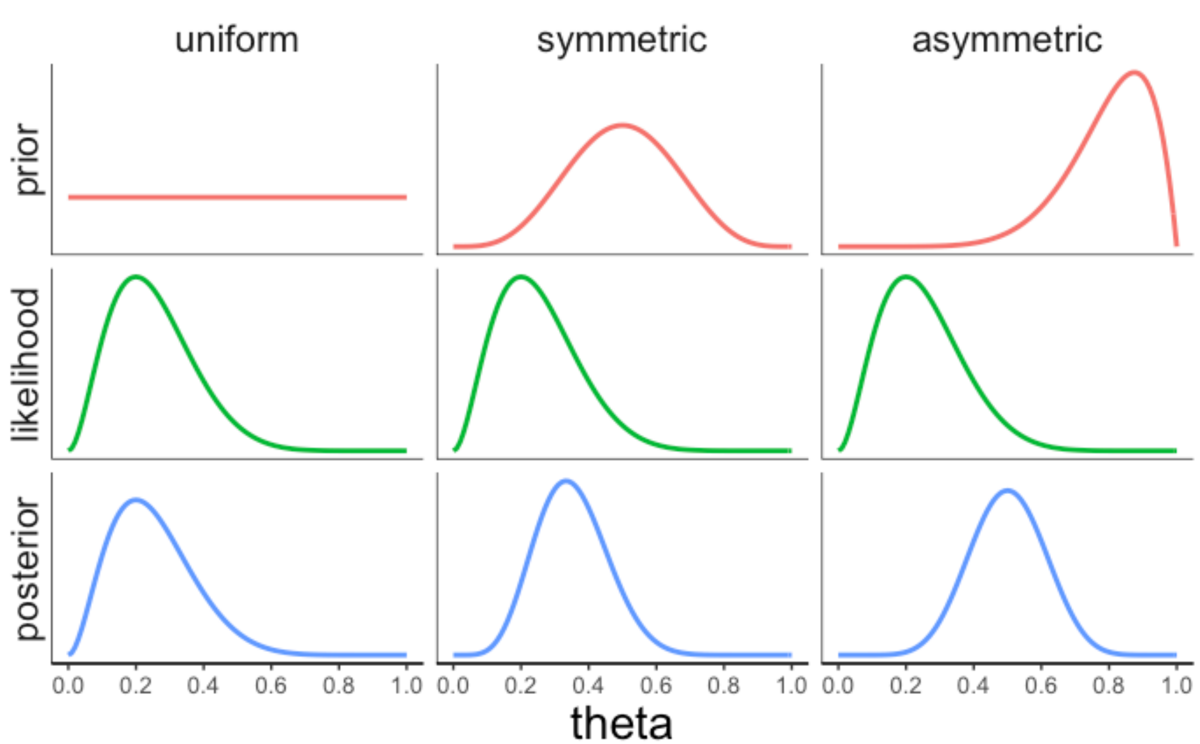
```r
# grid
theta = seq(0, 1, 0.01)

# data
data = rep(0:1, c(8, 2))

# calculate posterior
df.prior = tibble(theta = theta,
                  prior_uniform = dbeta(grid, shape1 = 1, shape2 = 1),
                  prior_normal = dbeta(grid, shape1 = 5, shape2 = 5),
                  prior_biased = dbeta(grid, shape1 = 8, shape2 = 2)) %>%
  pivot_longer(cols = -theta,
               names_to = "prior_index",
               values_to = "prior") %>%
  mutate(likelihood = dbinom(sum(data == 1),
                             size = length(data),
                             prob = theta)) %>%
  group_by(prior_index) %>%
  mutate(posterior = likelihood * prior / sum(likelihood * prior)) %>%
  ungroup() %>%
  pivot_longer(cols = -c(theta, prior_index),
               names_to = "index",
               values_to = "value")
```

100 discrete values

for 3 variables, we would already need 1 Mio combinations



The CURSE of dimensionality

Inference via sampling

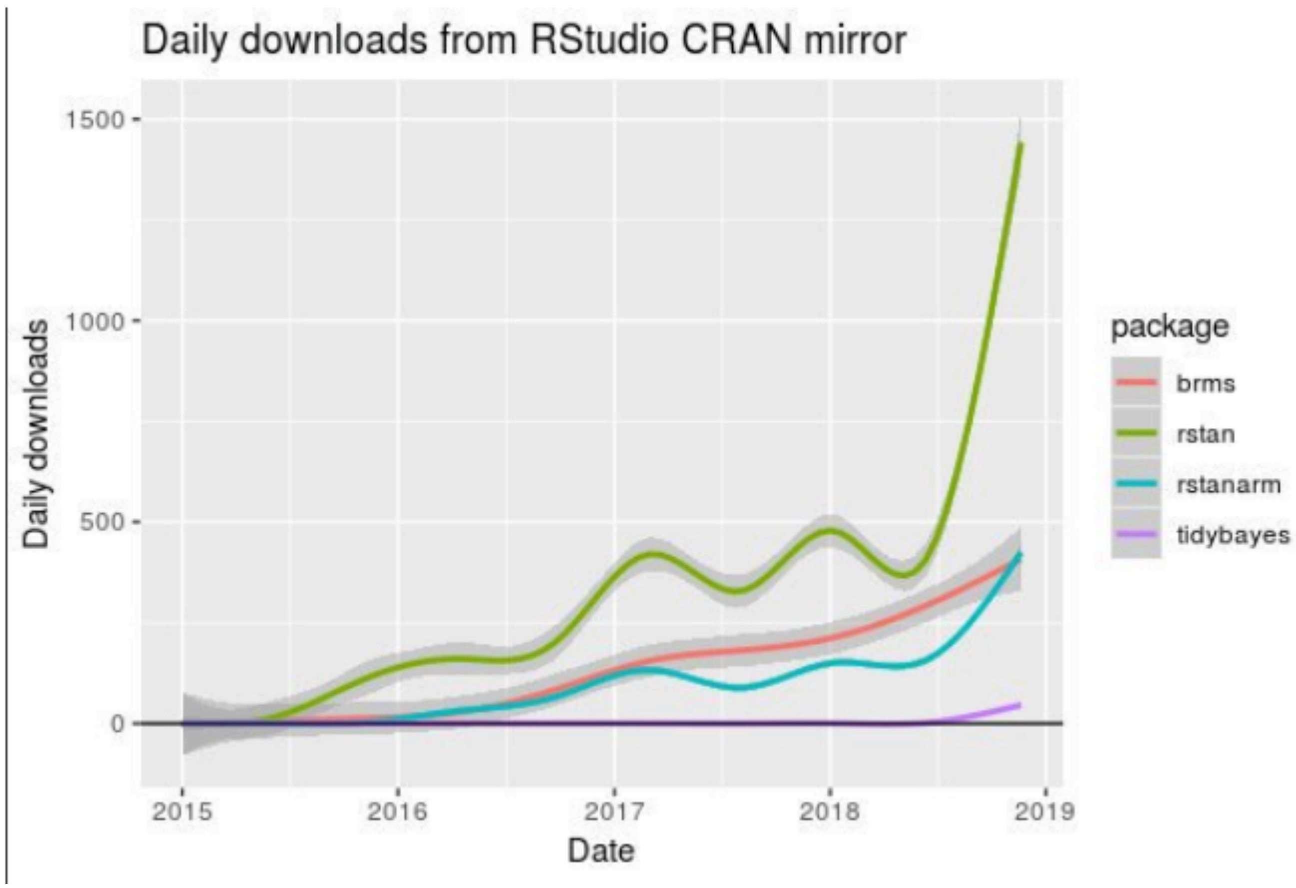
- we cannot directly calculate the probability of the posterior (because it might have a pretty weird shape)
- **but**: we can draw random samples from the posterior
- we can then use our data wrangling and visualization skills to make inferences based on these samples

It's as if ...

we don't have `pnorm()` but we do have `rnorm()`

Inference via sampling

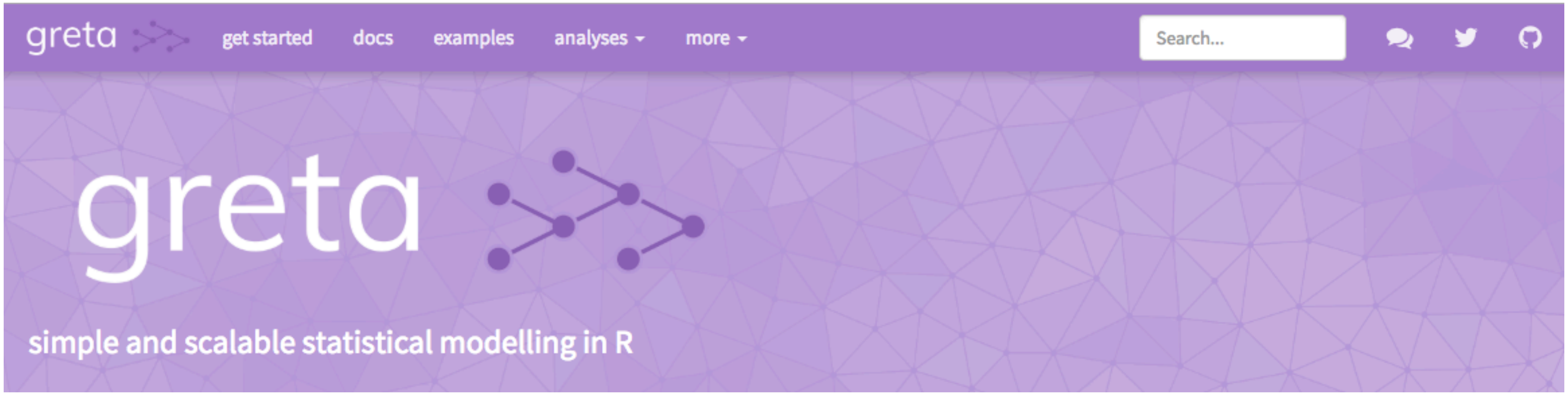
- Bayesian data analysis is becoming more popular because:
  - computers are getting more powerful
  - inference techniques are getting better
  - software packages become easier to use



Doing Bayesian data analysis

Software packages

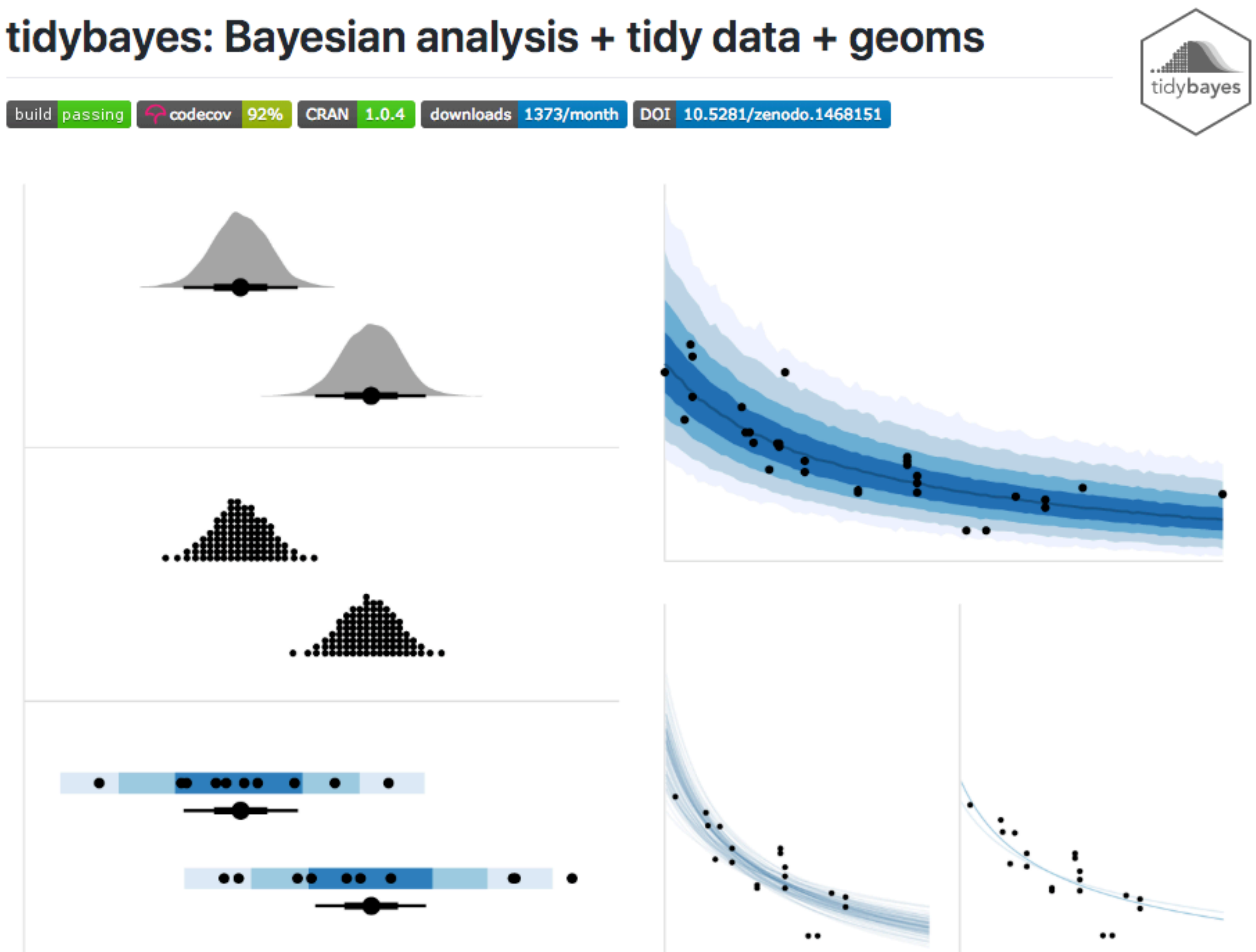
`library("greta")`



- let's us write Bayesian models directly in R with a simple syntax
- uses Tensorflow to implement Hamiltonian Monte Carlo sampling (a fast inference algorithm ...)

Software packages

`library("tidybayes")`

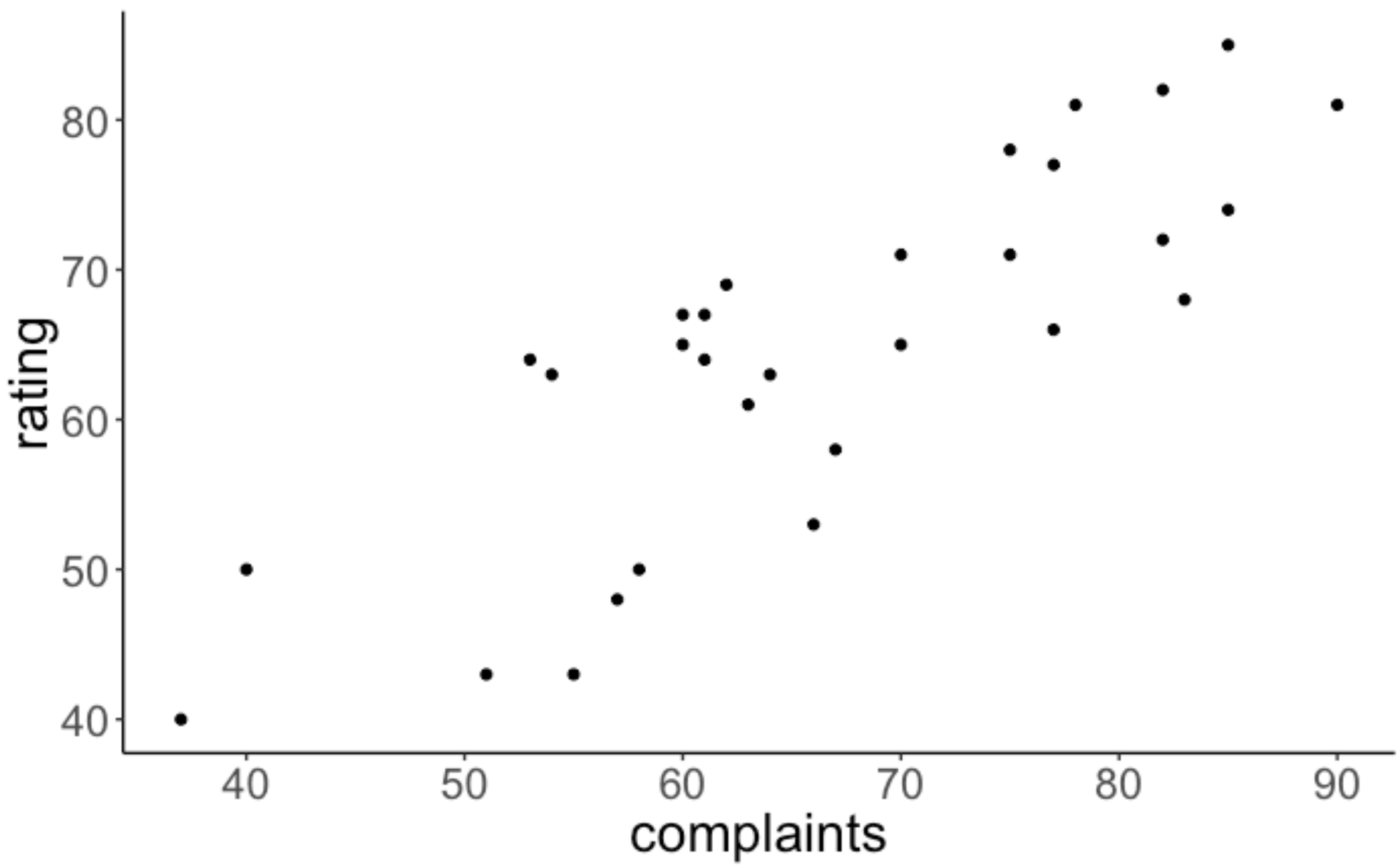


Matthew Kay

great tool for wrangling and visualizing the results of Bayesian data analysis

Attitude data set

What's the relationship between how well an employee handles complaints and their overall rating?



Frequentist analysis

Frequentist analysis

```r
# fit model
fit = lm(formula = rating ~ 1 + complaints,
         data = df.attitude)

# print summary
fit %>% summary()
```

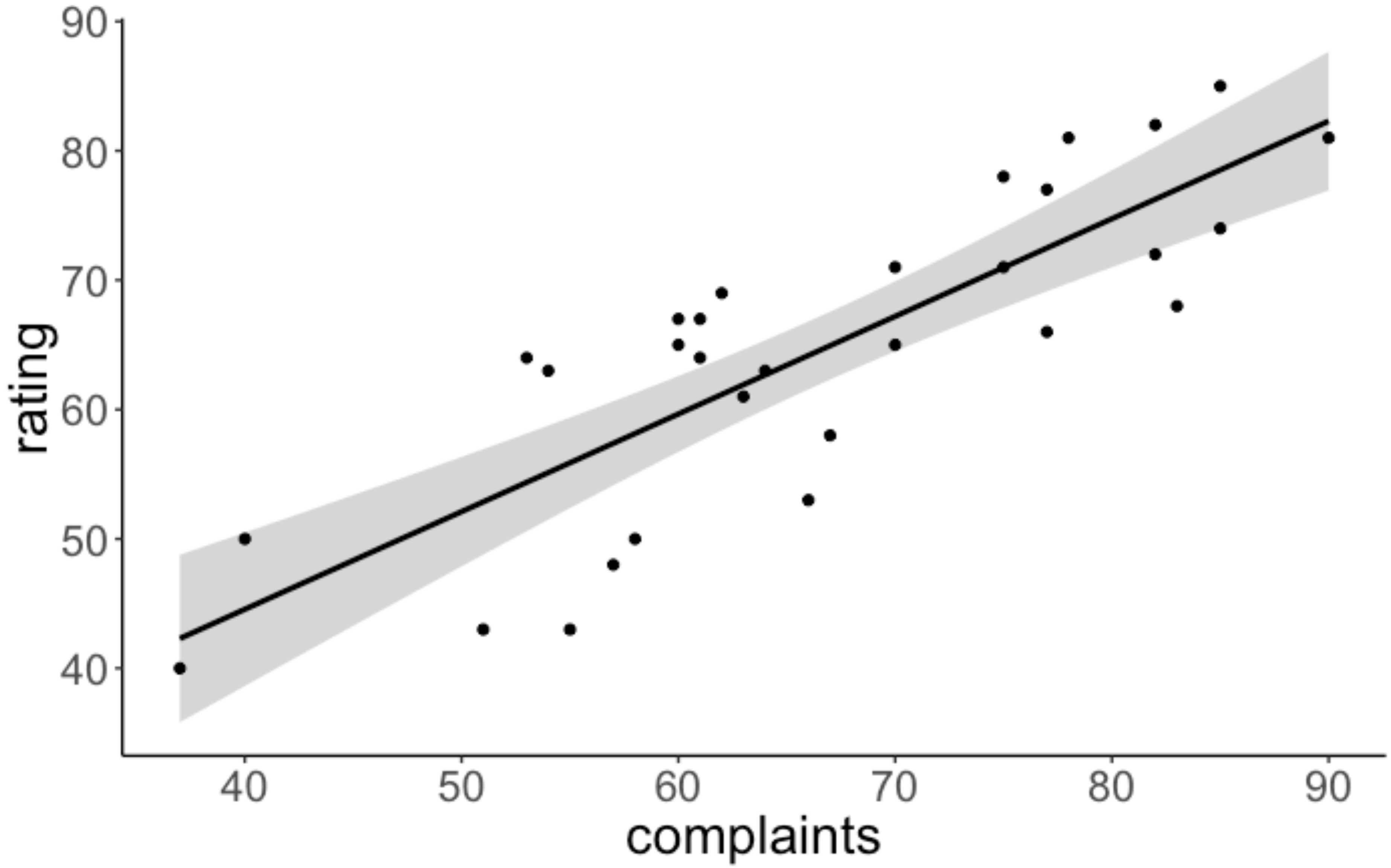
```
Call:
lm(formula = rating ~ 1 + complaints, data = df.attitude)

Residuals:
     Min       1Q   Median       3Q      Max
-12.8799  -5.9905   0.1783   6.2978   9.6294

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.37632    6.61999   2.172   0.0385 *
complaints   0.75461    0.09753   7.737 1.99e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.993 on 28 degrees of freedom
Multiple R-squared:  0.6813,	Adjusted R-squared:  0.6699
F-statistic: 59.86 on 1 and 28 DF,  p-value: 1.988e-08
```

Visualize model predictions



Best-fitting regression line with confidence interval

Bayesian analysis

Model specification

```r
library("greta")
library("tidybayes")

# variables & priors
b0 = normal(0, 10)
b1 = normal(0, 10)
sd = cauchy(0, 3, truncation = c(0, Inf))

# linear predictor
mu = b0 + b1 * attitude$complaints

# observation model (likelihood)
distribution(attitude$rating) = normal(mu, sd)

# define the model
m = model(b0, b1, sd)
```

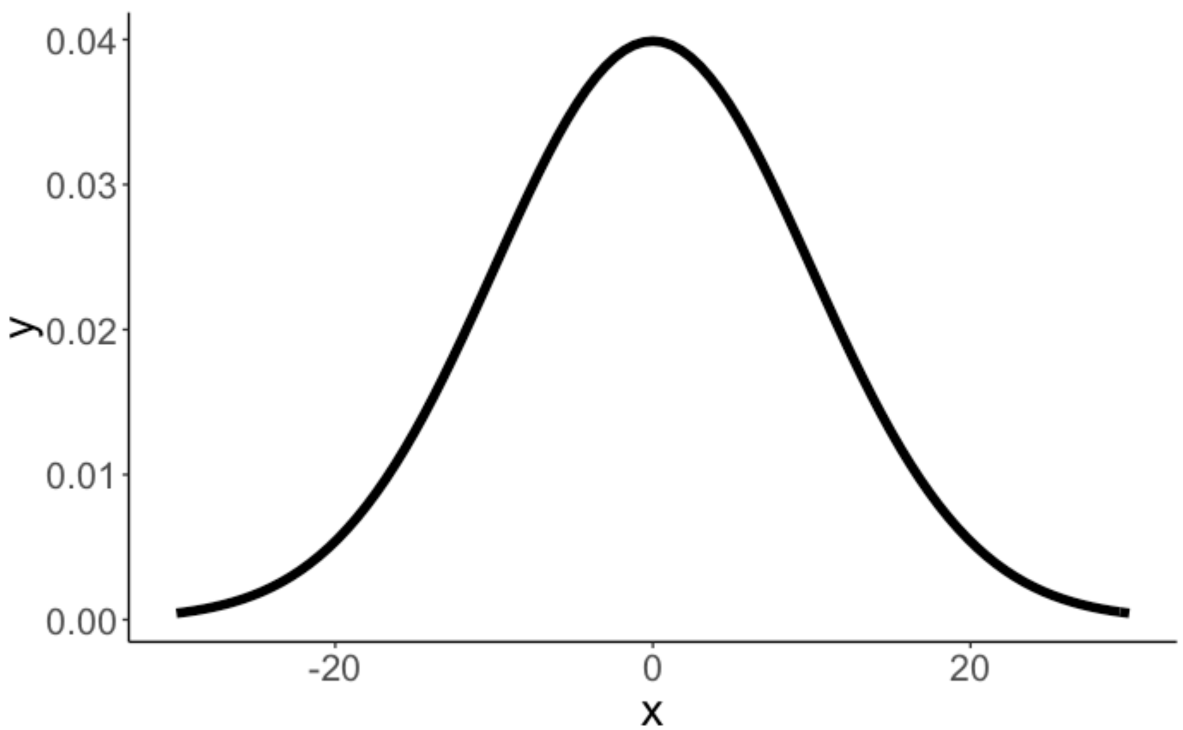
priors

linear combination

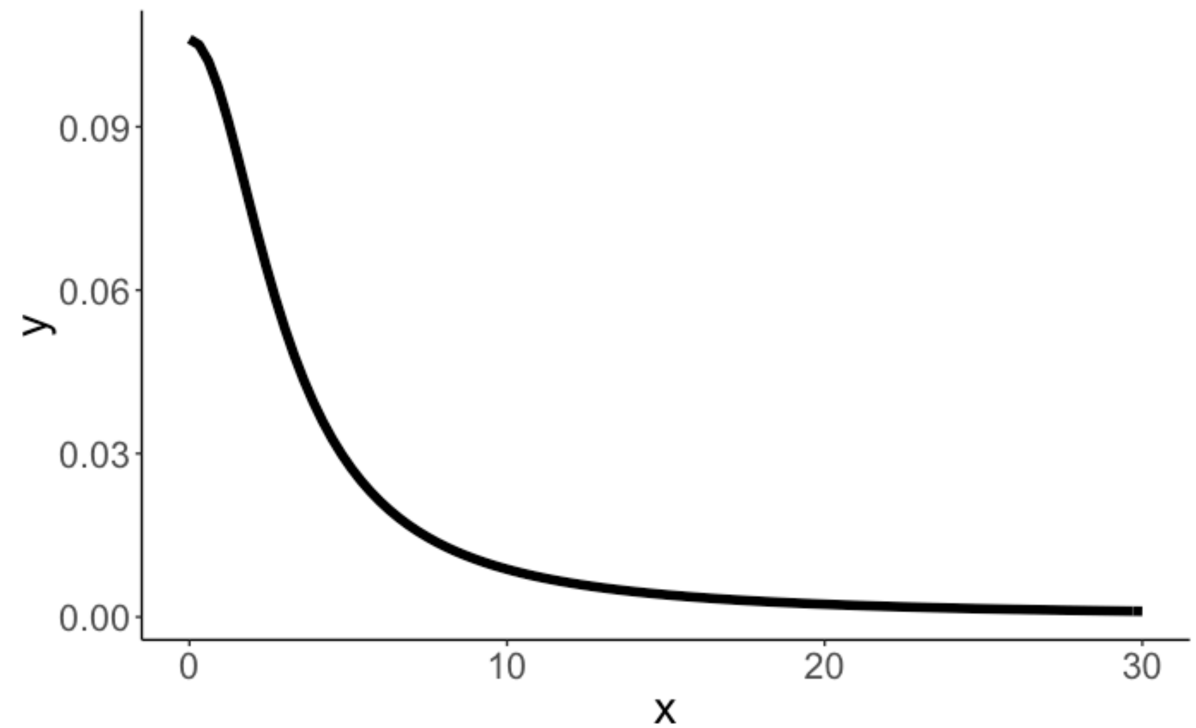
Gaussian likelihood

build the model

Priors



Gaussian prior on intercept and coefficient

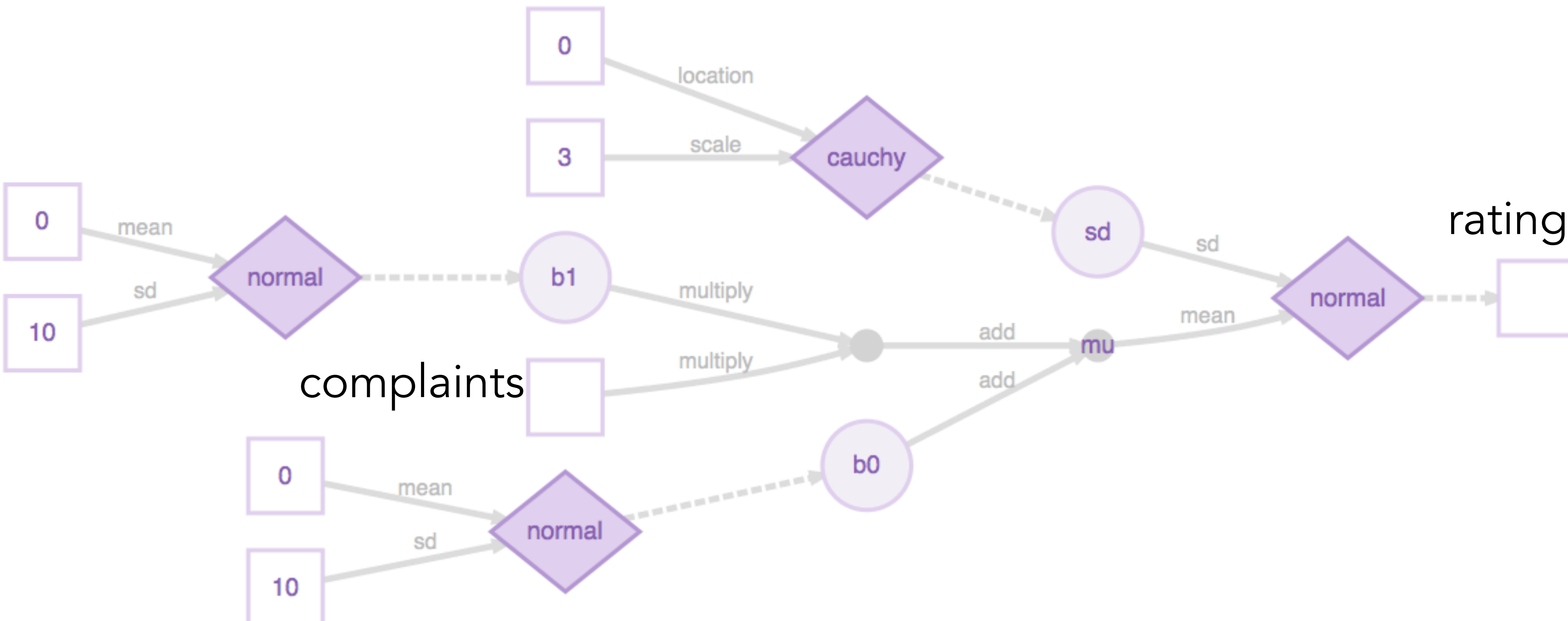


Truncated Cauchy prior on the standard deviation

weakly informative priors (allow for a wide range of possible values)

Graphical representation of the model

```
1 # plotting
2 plot(m)
```



Inference via sampling

Markov Chain
Monte Carlo
inference

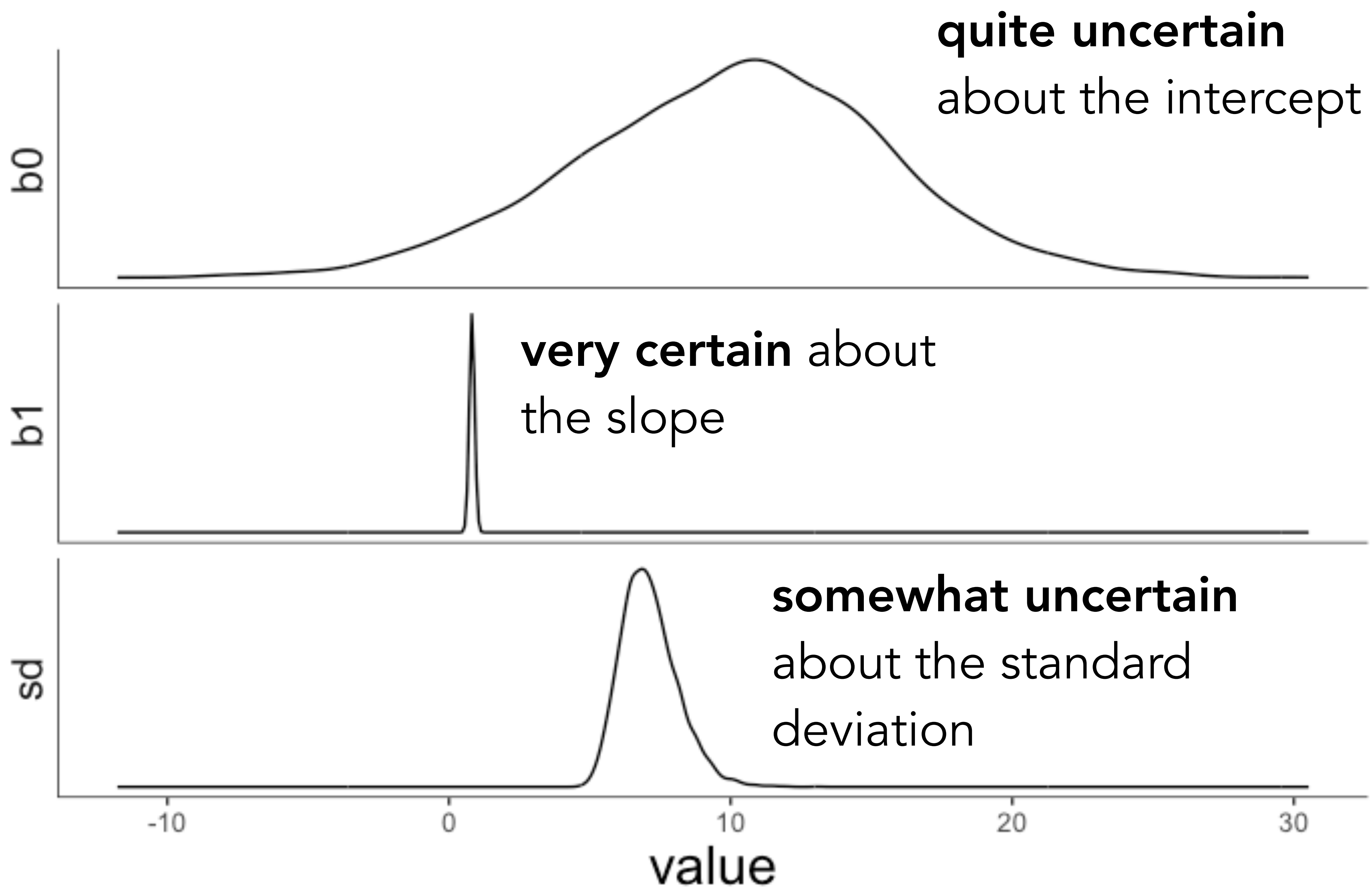
```r
# sampling
draws = mcmc(m, n_samples = 1000)

# tidy up the draws
df.draws = tidy_draws(draws) %>%
  clean_names()
```

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

each of these is a solution for explaining the data

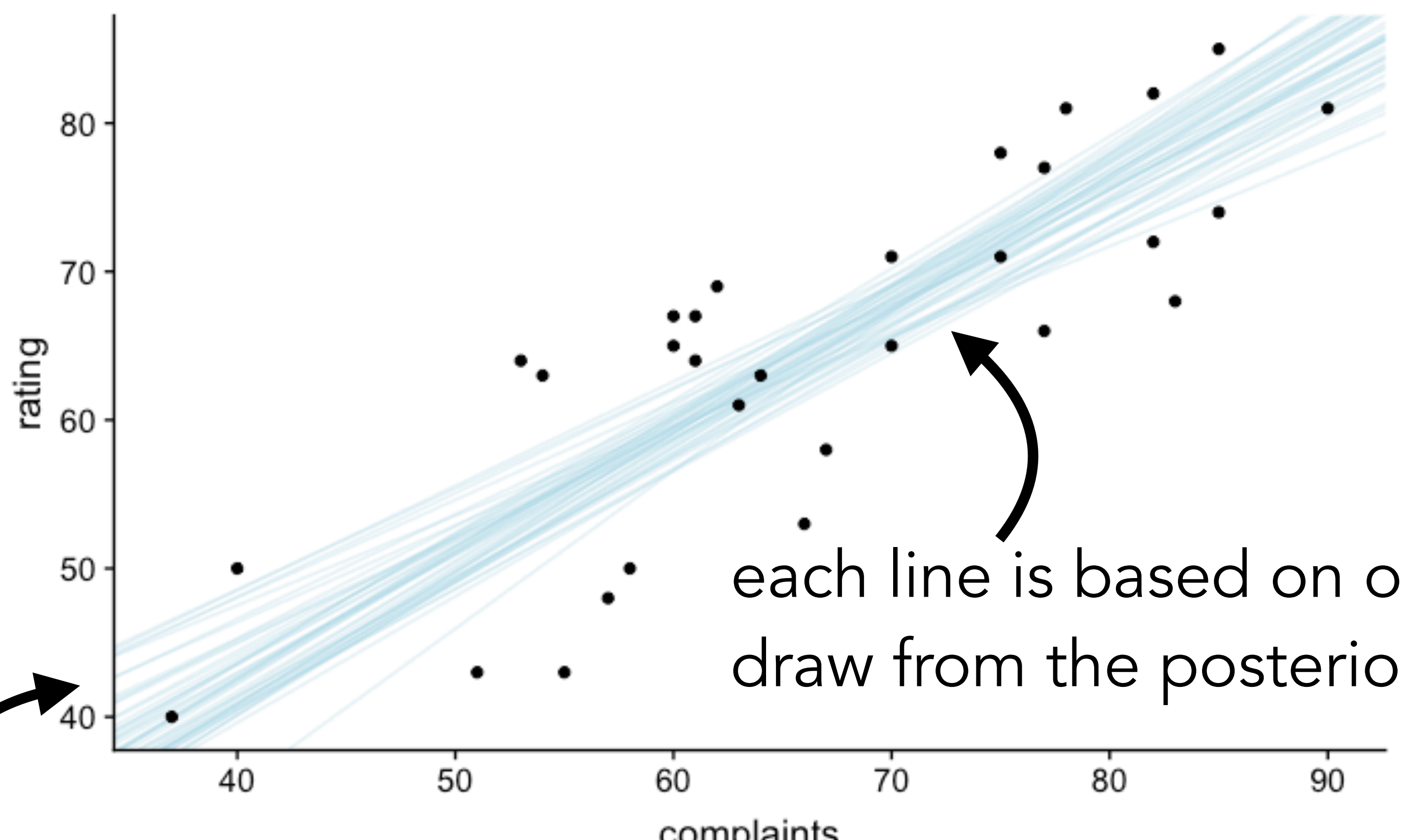
Visualize the posterior



Posterior distribution over the three parameters in the model

Visualize the model predictions

```r
ggplot(data = df.attitude,
       mapping = aes(x = complaints,
                     y = rating)) +
  geom_abline(data = df.draws %>%
                sample_n(size = 50),
              aes(intercept = b0,
                  slope = b1),
              alpha = 0.3,
              color = "lightblue") +
  geom_point()
```



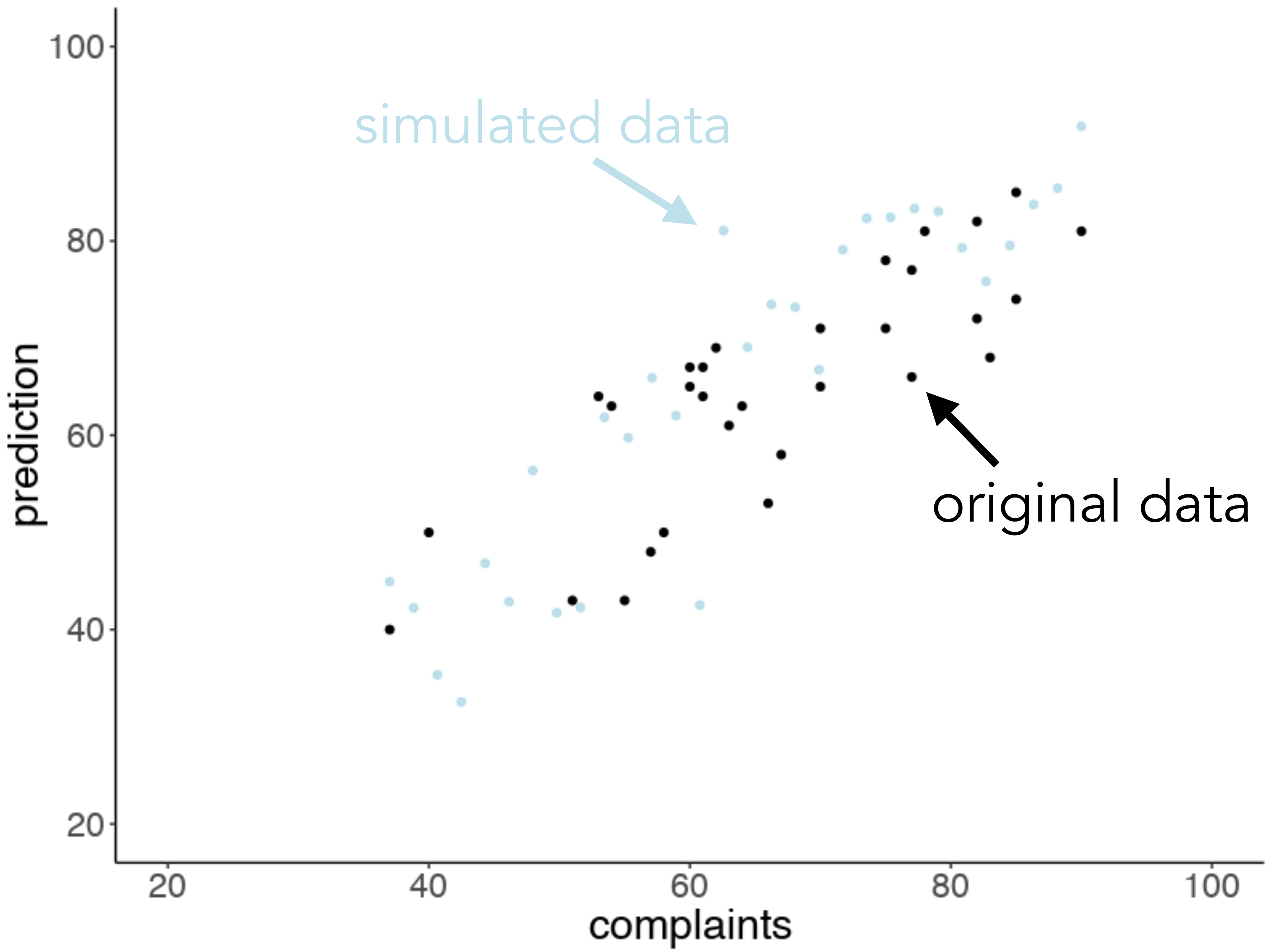
each line is based on one draw from the posterior

explains the "high" uncertainty about the intercept

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

Posterior predictive check

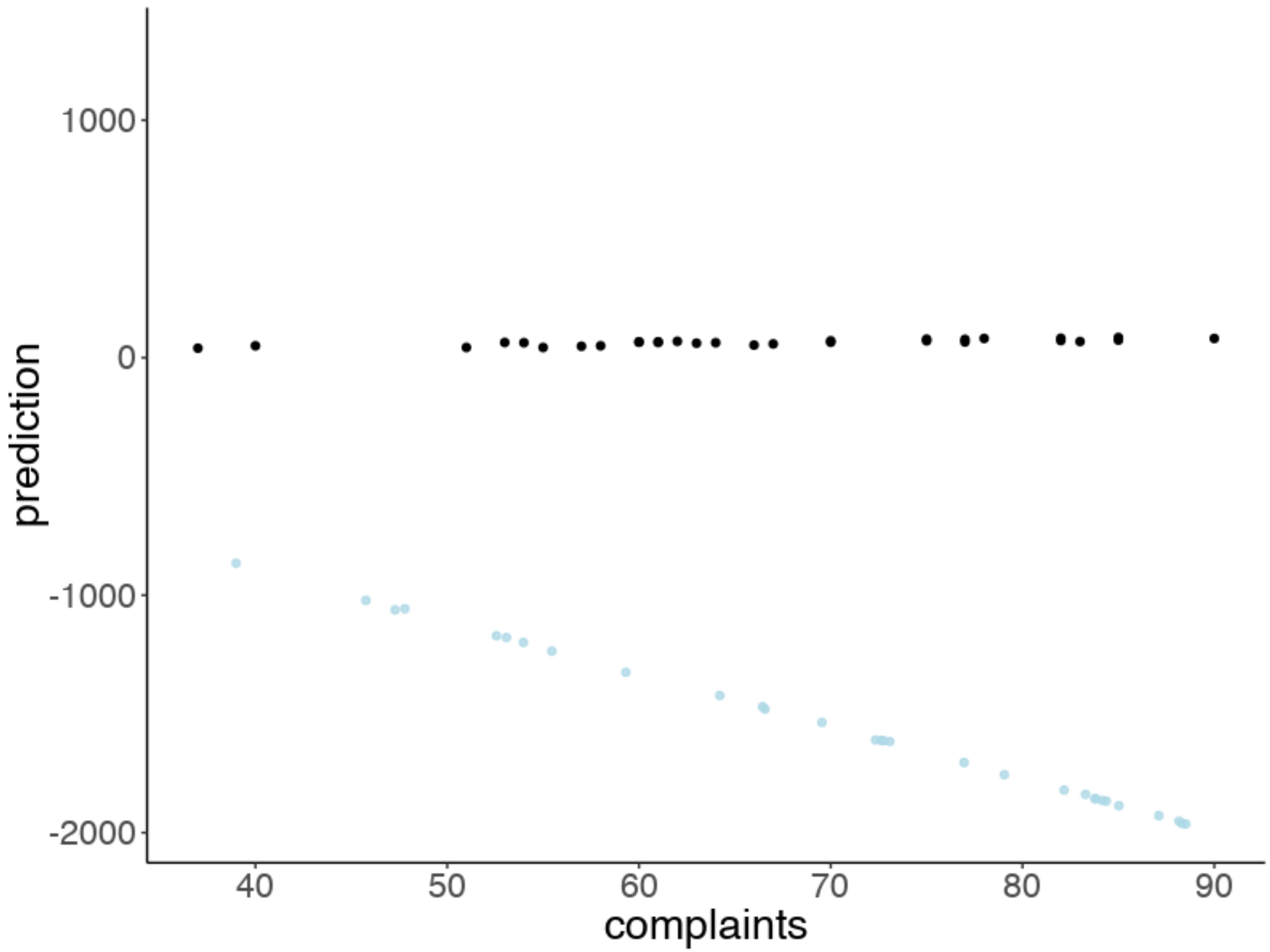
1. sample parameters from the **posterior distribution**
2. generate data using these parameters (using the likelihood function)



Hypothetical outcome plot

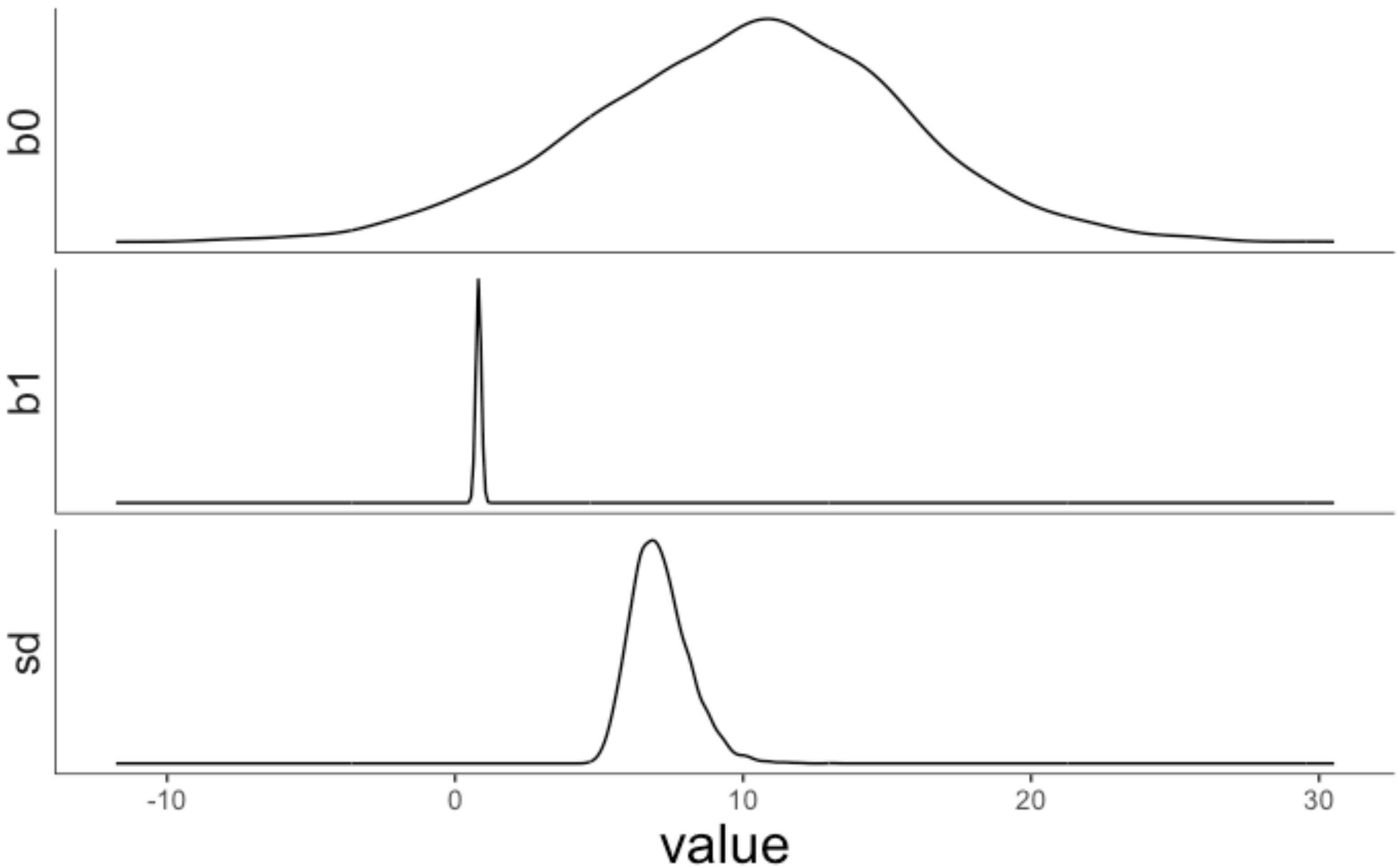
Prior predictive check

1. sample parameters from the **prior distribution**
2. generate data using these parameters (using the likelihood function)



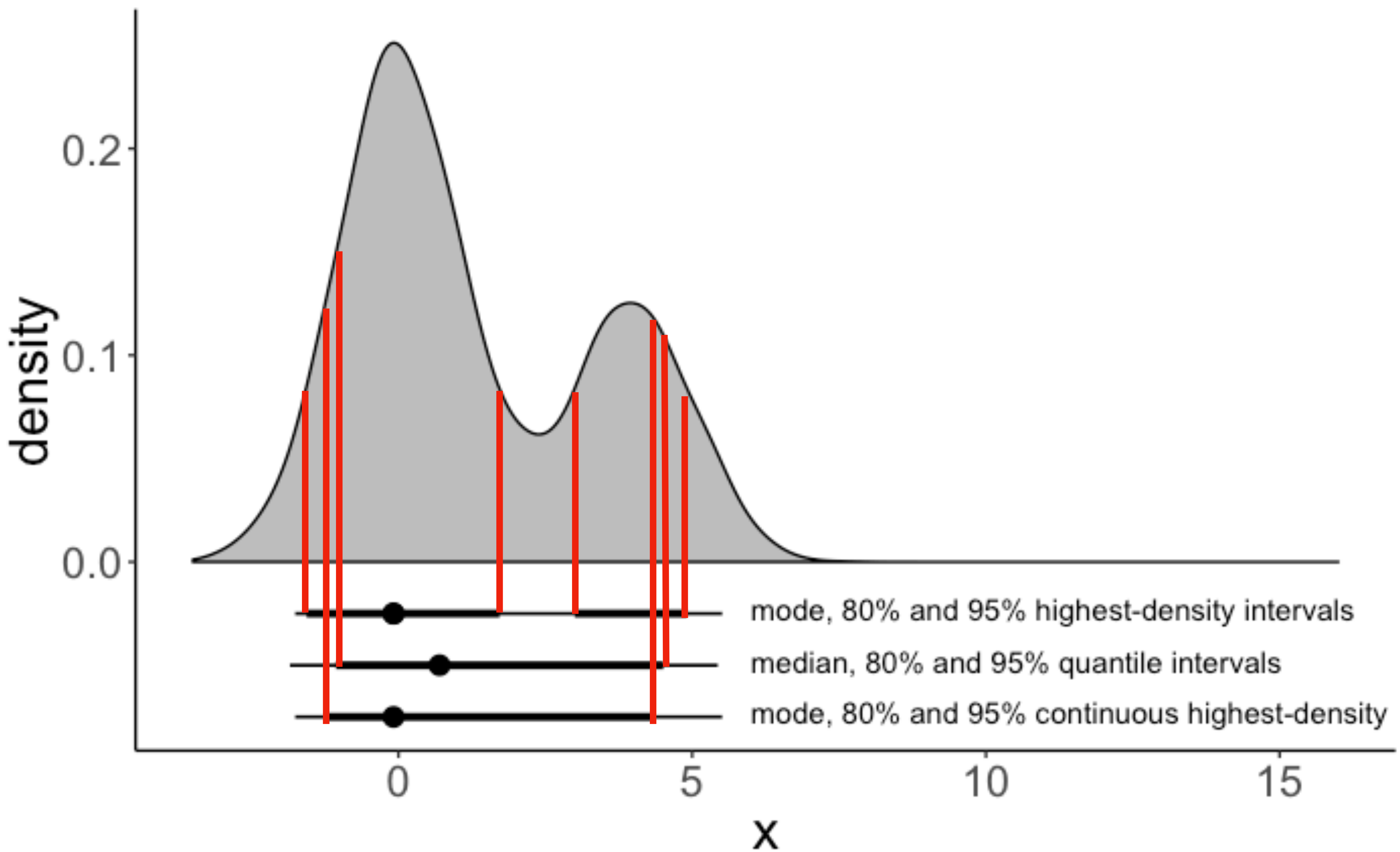
Hypothetical outcome plot

Summarizing results



- Posterior over each parameter is the result of the Bayesian data analysis.
- no p-values
- no confidence intervals

Different kinds of credible intervals



Summary

- Bayesian data analysis
  - Comparison between frequentist and Bayesian data analysis
  - Quick flash from the past
  - Flipping coins
  - What affects the posterior?
  - Ingredients: likelihood, prior, inference
  - Doing Bayesian data analysis

Feedback

What did you like about today's class? What could be improved next time?

Thank you!