The temperature in a Normal lake

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist

The model we've used so far

$$n_{\text{ads}} = 100$$

$$p_{\text{clicks}} \sim \text{Uniform}(0.0, 0.2)$$

$$n_{\text{visitors}} \sim \text{Binomial}(n_{\text{ads}}, p_{\text{clicks}})$$

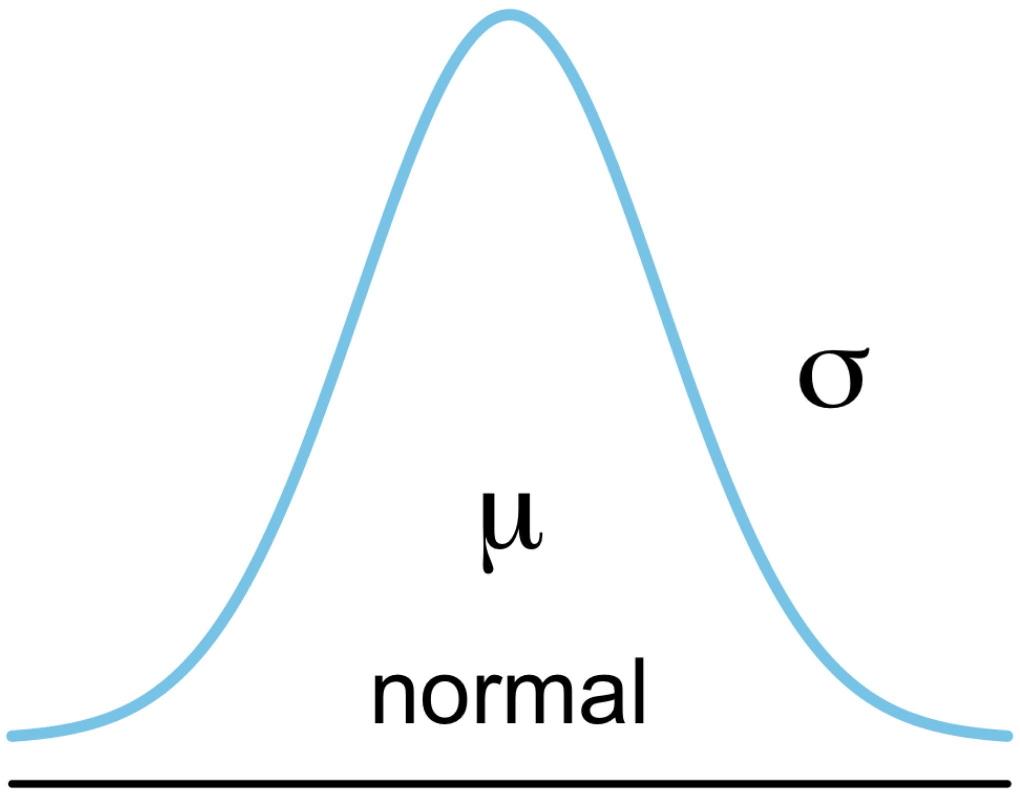
Some temperature data

```
temp   <- c(19, 23, 20, 17, 23)
```

```
temp_f <- c(66, 73, 68, 63, 73)
```

The Normal distribution

$$\text{Normal}(\mu, \sigma)$$



The Normal distribution in R

```
rnorm(n = , mean = , sd = )
```

The Normal distribution in R

```r
rnorm(n = 5, mean = 20, sd = 2)
```

```
20.3 24.1 22.4 24.7 21.6
```

```r
rnorm(n = 5, mean = 20, sd = 2)
```

```
16.3 22.1 23.1 18.9 16.3
```

```r
rnorm(n = 5, mean = 20, sd = 2)
```

```
20.3 20.9 18.0 16.8 22.6
```

```r
temp   <- c(19, 23, 20, 17, 23)
```

The Normal distribution in R

```r
temp <- c(19, 23, 20, 17, 23)
like <- dnorm(x = temp, mean = 20, sd = 2)
like
```

```
0.176 0.065 0.199 0.065 0.065
```

```r
prod(like)
```

```
9.536075e-06
```

```r
log(like)
```

```
-1.737086 -2.737086 -1.612086 -2.737086 -2.737086
```

Try out using rnorm and dnorm!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

A Bayesian model of water temperature

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist

Let's define the model

$$\text{temp} = 19, 23, 20, 17, 23$$

Let's define the model

$$\text{temp}_i \sim \text{Normal}(\mu, \sigma)$$
$$\text{temp} = 19, 23, 20, 17, 23$$

Let's define the model

$$\sigma \sim \text{Uniform}(\text{min}: 0, \text{max}: 10)$$
$$\text{temp}_i \sim \text{Normal}(\mu, \sigma)$$
$$\text{temp} = 19, 23, 20, 17, 23$$

Let's define the model

$$\mu \sim \text{Normal}(\text{mean: } 18, \text{sd: } 5)$$
$$\sigma \sim \text{Uniform}(\text{min: } 0, \text{max: } 10)$$
$$\text{temp}_i \sim \text{Normal}(\mu, \sigma)$$
$$\text{temp} = 19, 23, 20, 17, 23$$

Let's fit the model

```
n_ads_shown <- 100
n_visitors <- 13
proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)

proportion_clicks <- seq(0, 1, by = 0.01)
pars <- expand.grid(proportion_clicks = proportion_clicks)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <-
sigma <-
pars <- expand.grid(proportion_clicks = proportion_clicks)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

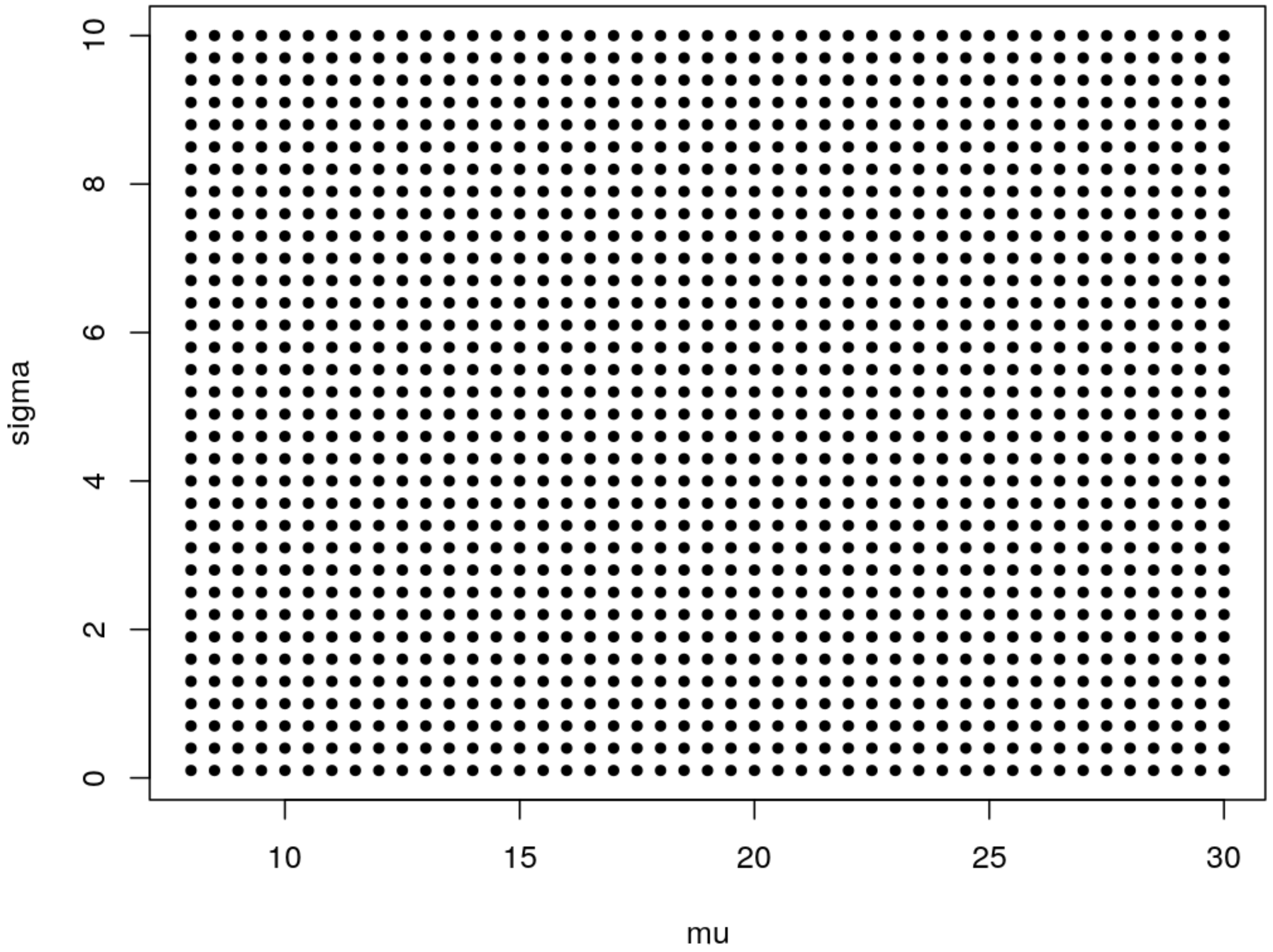
```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(proportion_clicks = proportion_clicks)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

The parameter space

```
plot(pars, pch=19)
```



Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$mu_prior <- dnorm(pars$mu, mean = 18, sd = 5)

pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$mu_prior <- dnorm(pars$mu, mean = 18, sd = 5)
pars$sigma_prior <- dunif(pars$sigma, min = 0, max = 10)
pars$prior <- dunif(pars$proportion_clicks, min = 0, max = 0.2)
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$mu_prior <- dnorm(pars$mu, mean = 18, sd = 5)
pars$sigma_prior <- dunif(pars$sigma, min = 0, max = 10)
pars$prior <- pars$mu_prior * pars$sigma_prior
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

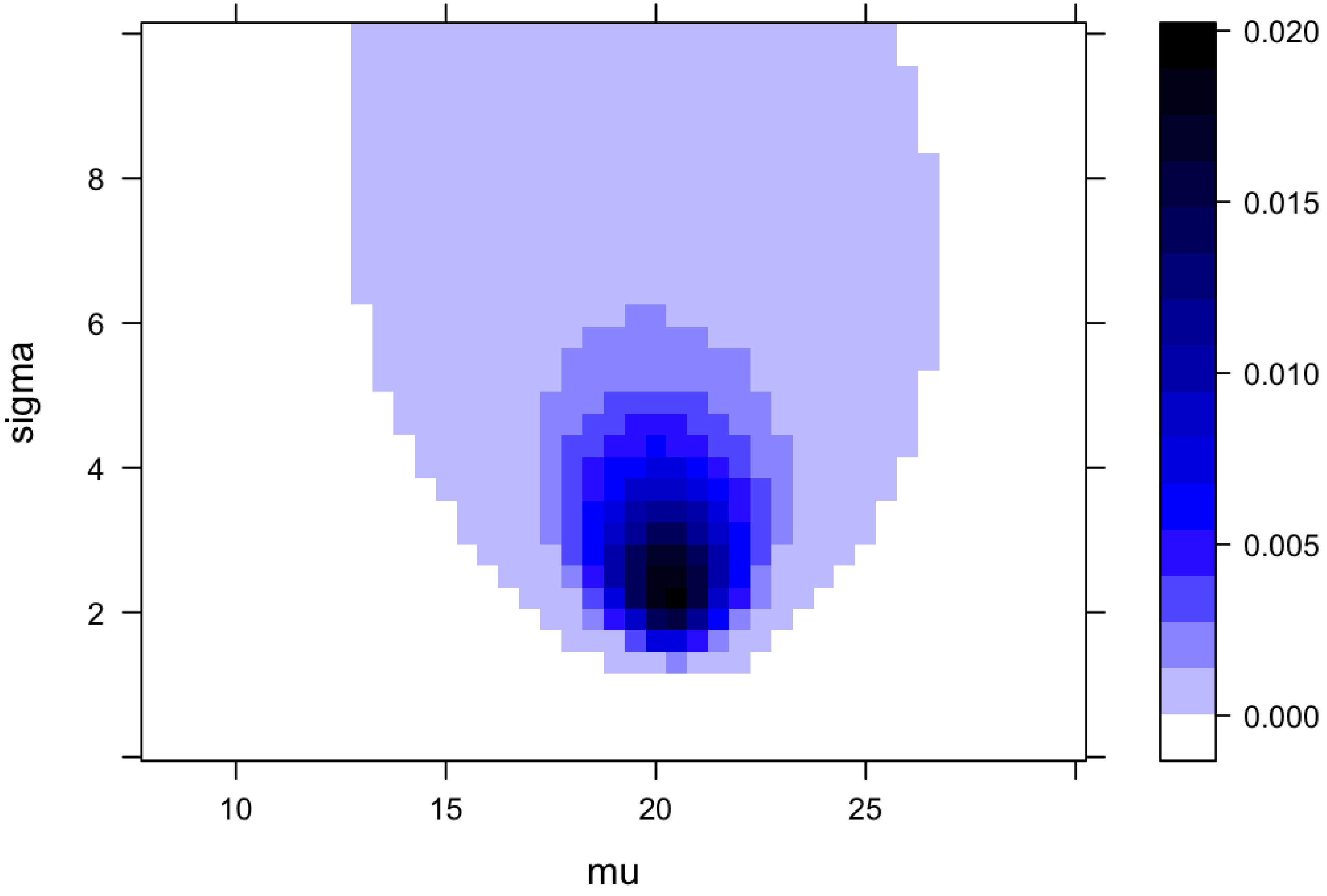
```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$mu_prior <- dnorm(pars$mu, mean = 18, sd = 5)
pars$sigma_prior <- dunif(pars$sigma, min = 0, max = 10)
pars$prior <- pars$mu_prior * pars$sigma_prior
for(i in 1:nrow(pars)) {

pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$mu_prior <- dnorm(pars$mu, mean = 18, sd = 5)
pars$sigma_prior <- dunif(pars$sigma, min = 0, max = 10)
pars$prior <- pars$mu_prior * pars$sigma_prior
for(i in 1:nrow(pars)) {
    likelihoods <- dnorm(temp, pars$mu[i], pars$sigma[i])
pars$likelihood <- dbinom(n_visitors,
    size = n_ads_shown, prob = pars$proportion_clicks)
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```

Let's fit the model

```
temp <- c(19, 23, 20, 17, 23)
mu <- seq(8, 30, by = 0.5)
sigma <- seq(0.1, 10, by = 0.3)
pars <- expand.grid(mu = mu, sigma = sigma)
pars$mu_prior <- dnorm(pars$mu, mean = 18, sd = 5)
pars$sigma_prior <- dunif(pars$sigma, min = 0, max = 10)
pars$prior <- pars$mu_prior * pars$sigma_prior
for(i in 1:nrow(pars)) {
    likelihoods <- dnorm(temp, pars$mu[i], pars$sigma[i])
    pars$likelihood[i] <- prod(likelihoods)
}
pars$probability <- pars$likelihood * pars$prior
pars$probability <- pars$probability / sum(pars$probability)
```


sigma
mu
10
15
20
25
2
4
6
8
0.020
0.015
0.010
0.005
0.000

Replicate this analysis using zombie data!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Answering the question: Should I have a beach party?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist

The questions

- What's likely the average water temperature on 20th of Julys?
- What's the probability that the water temperature is going to be 18 or more on the *next* 20th?

The posterior distribution

```
pars
```

```
  mu sigma probability
17.5   1.9      0.0001
18.0   1.9      0.0003
18.5   1.9      0.0014
19.0   1.9      0.0043
19.5   1.9      0.0094
20.0   1.9      0.0142
20.5   1.9      0.0151
21.0   1.9      0.0112
21.5   1.9      0.0058
22.0   1.9      0.0021
...    ...       ...
```

```
sample_indices <- sample(1:nrow(pars), size = 10000,
                         replace = TRUE, prob = pars$probability)
```

```r
sample_indices <- sample(1:nrow(pars), size = 10000,
                         replace = TRUE, prob = pars$probability)
```

```r
head(sample_indices)
```

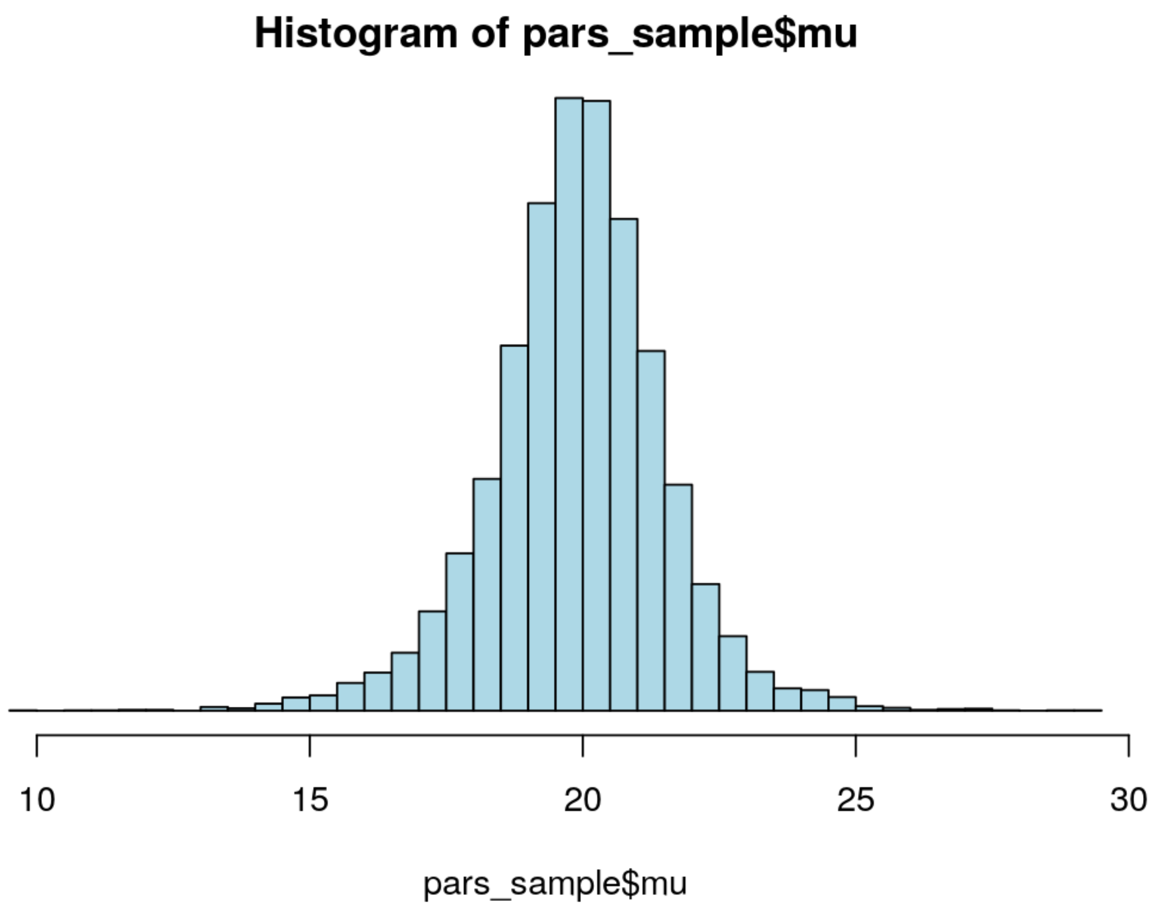
```
430  428 1010  383  343  385
```

```r
pars_sample <- pars[sample_indices, c("mu", "sigma")]
head(pars_sample)
```

```
     mu sigma
1  20.0   2.8
2  19.0   2.8
3  17.5   6.7
4  19.0   2.5
5  21.5   2.2
6  20.0   2.5
7  20.0   2.8
8  20.5   1.6
9  19.0   2.5
10 17.0   4.0
```

The probability distribution over the mean temperature

```
hist(pars_sample$mu, 30)
```



The probability distribution over the mean temperature

```
quantile(pars_sample$mu, c(0.05, 0.95))
```

```
  5%  95%
17.5 22.5
```

Is the temperature 18 or above on the 20th?

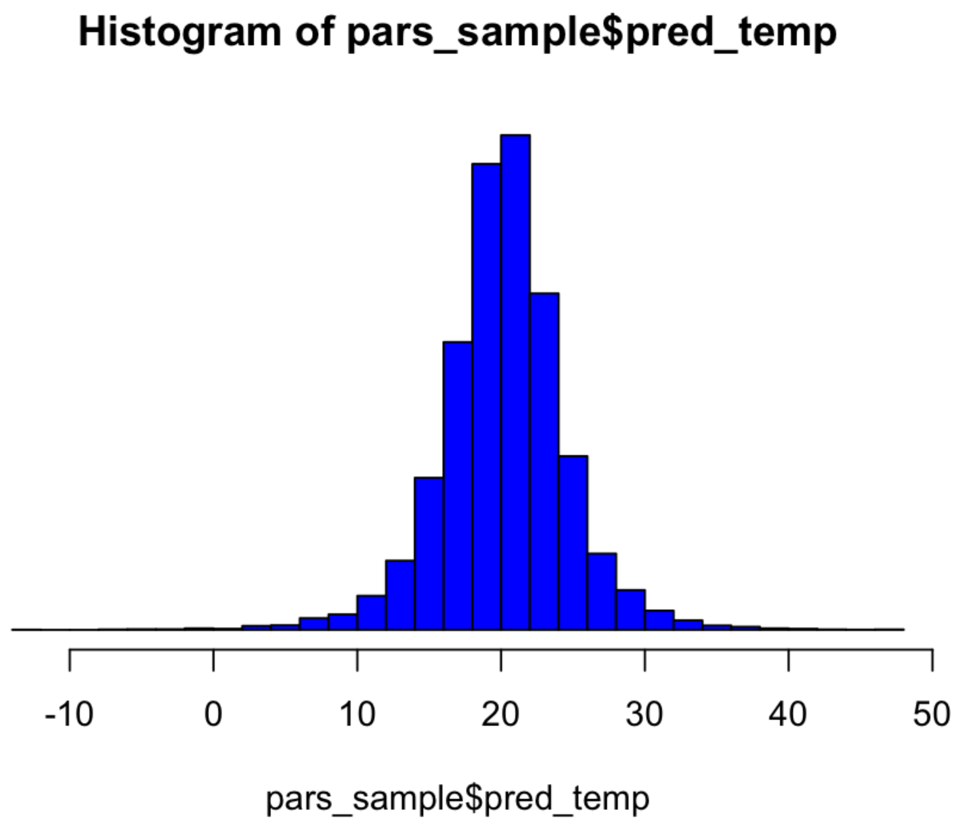
```
pred_temp <- rnorm(10000, mean = , sd = )
```

Is the temperature 18 or above on the 20th?

```
pred_temp <- rnorm(10000, mean = pars_sample$mu, sd = pars_sample$sigma)
```

Is the temperature 18 or above on the 20th?

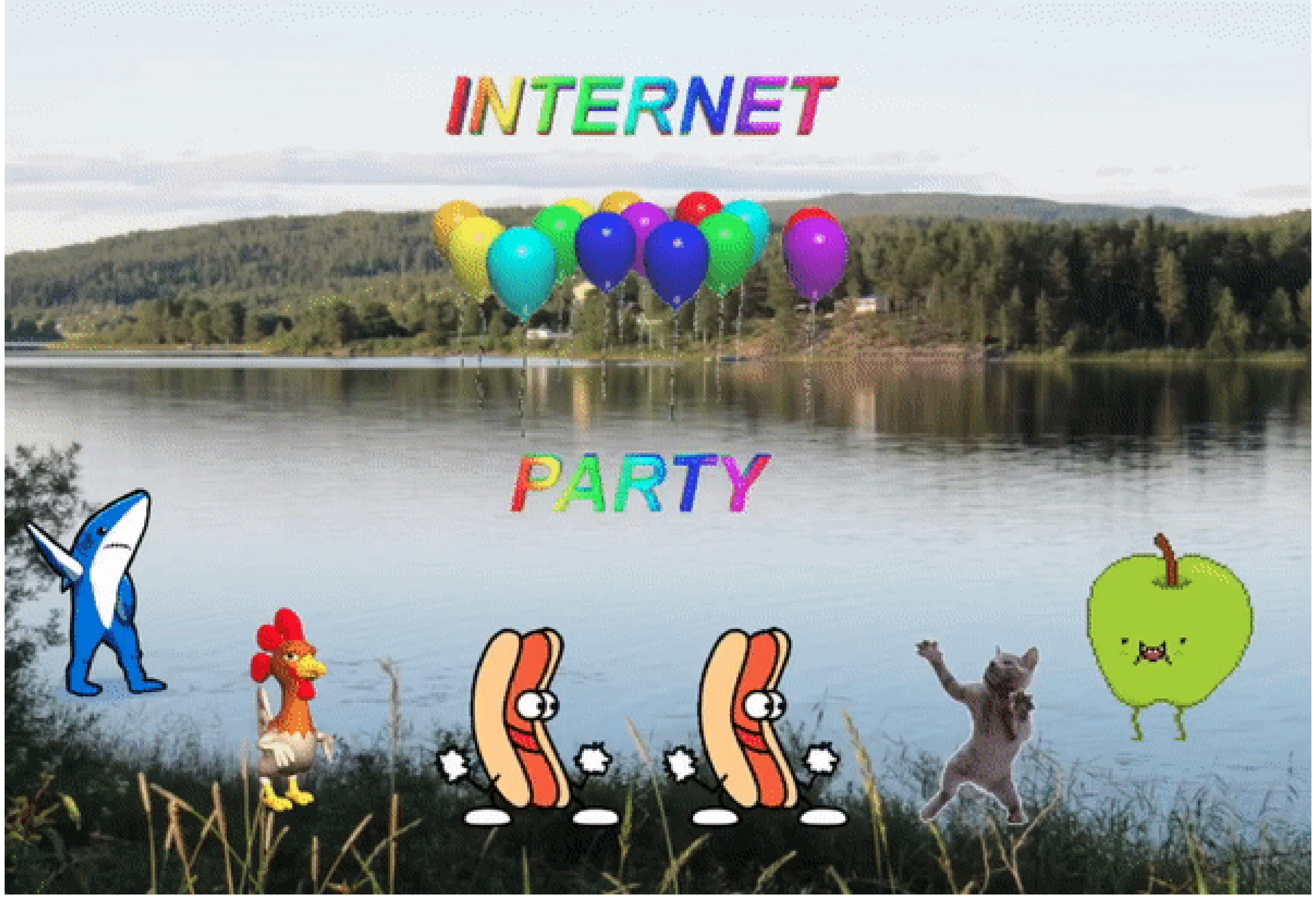
```
pred_temp <- rnorm(10000, mean = pars_sample$mu, sd = pars_sample$sigma)
hist(pred_temp, 30)
```



Is the temperature 18 or above on the 20th?

```
pred_temp <- rnorm(10000, mean = pars_sample$mu, sd = pars_sample$sigma)
hist(pred_temp, 30)
sum(pred_temp >= 18) / length(pred_temp )
```

```
0.73
```


INTERNET
PARTY

What about the IQ of zombies?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

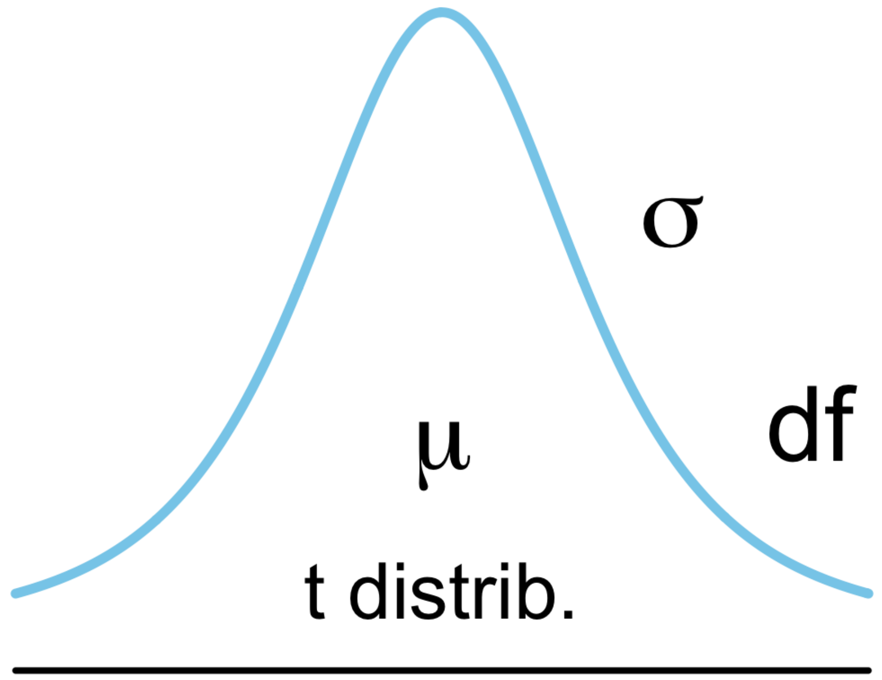
You've fitted a Bayesian Normal model!

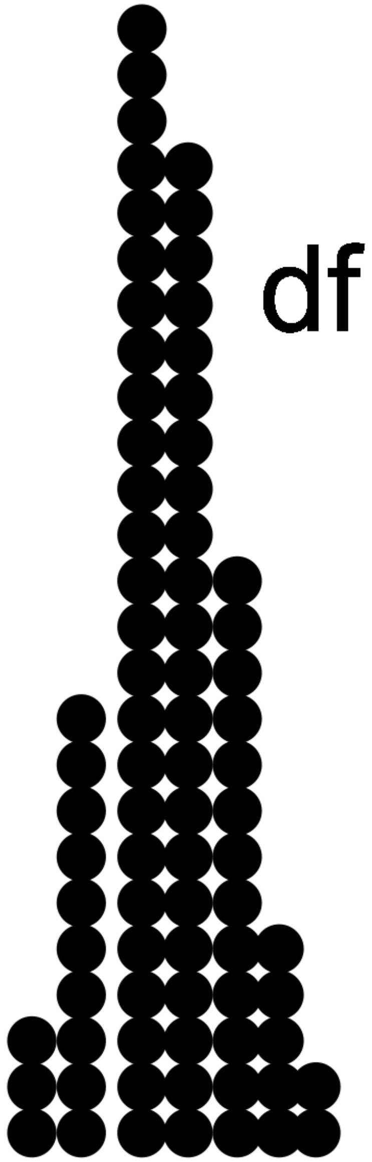
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

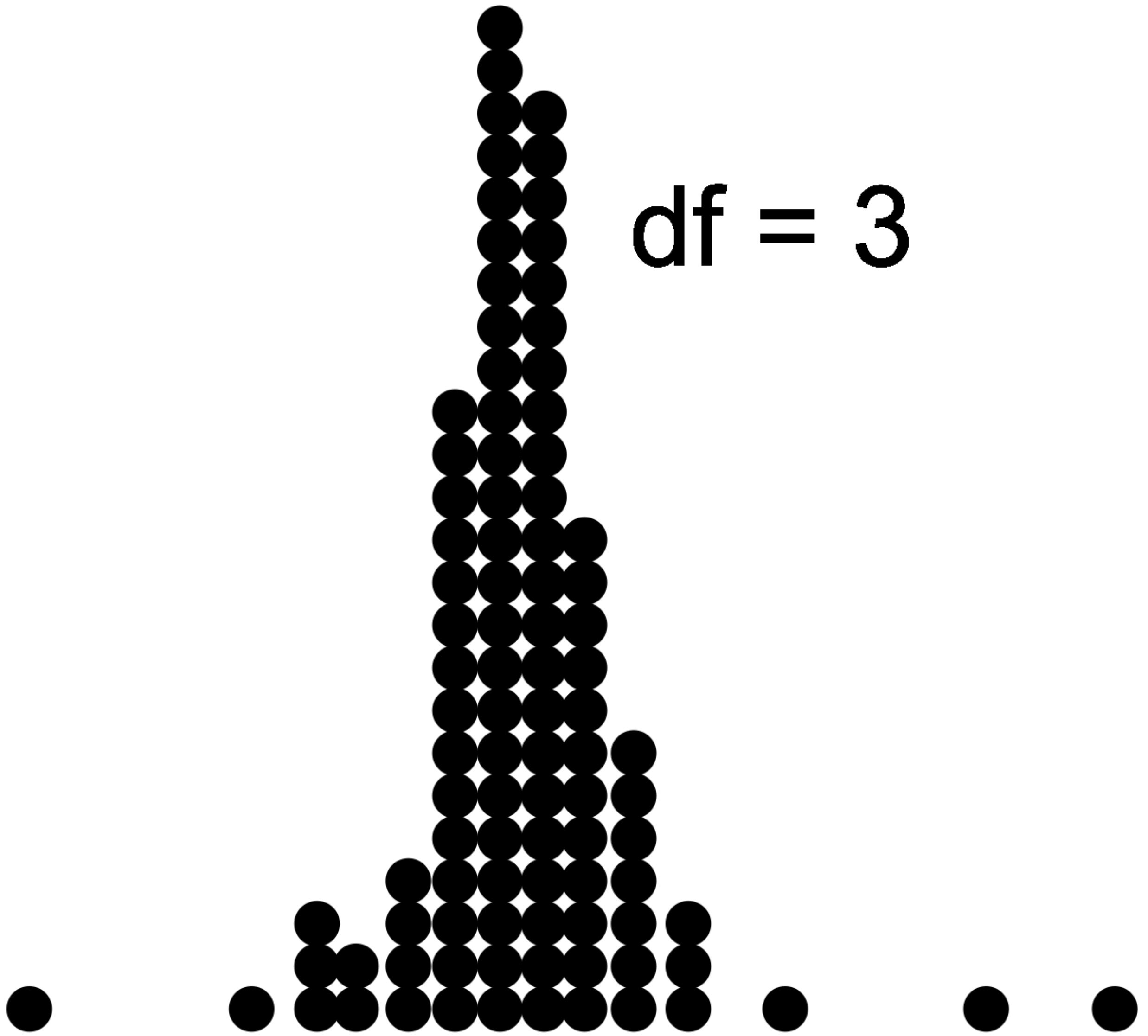
Rasmus Bååth
Data Scientist

BEST

- A Bayesian model developed by John Kruschke.
- Assumes the data comes from a t-distribution.




df > 30


df = 3

BEST

- A Bayesian model developed by John Kruschke.
- Assumes the data comes from a t-distribution.
- Estimates the mean, standard deviation and degrees-of-freedom parameter.
- `library(BEST)`
- Uses Markov chain Monte Carlo (MCMC).

Let's use BEST!

```
library(BEST)
iq <- c(55, 44, 34, 18, 51, 40, 40, 49, 48, 46)
```

Let's use BEST!

```
library(BEST)
iq <- c(55, 44, 34, 18, 51, 40, 40, 49, 48, 46)
fit <- BESTmcmc(iq)
```

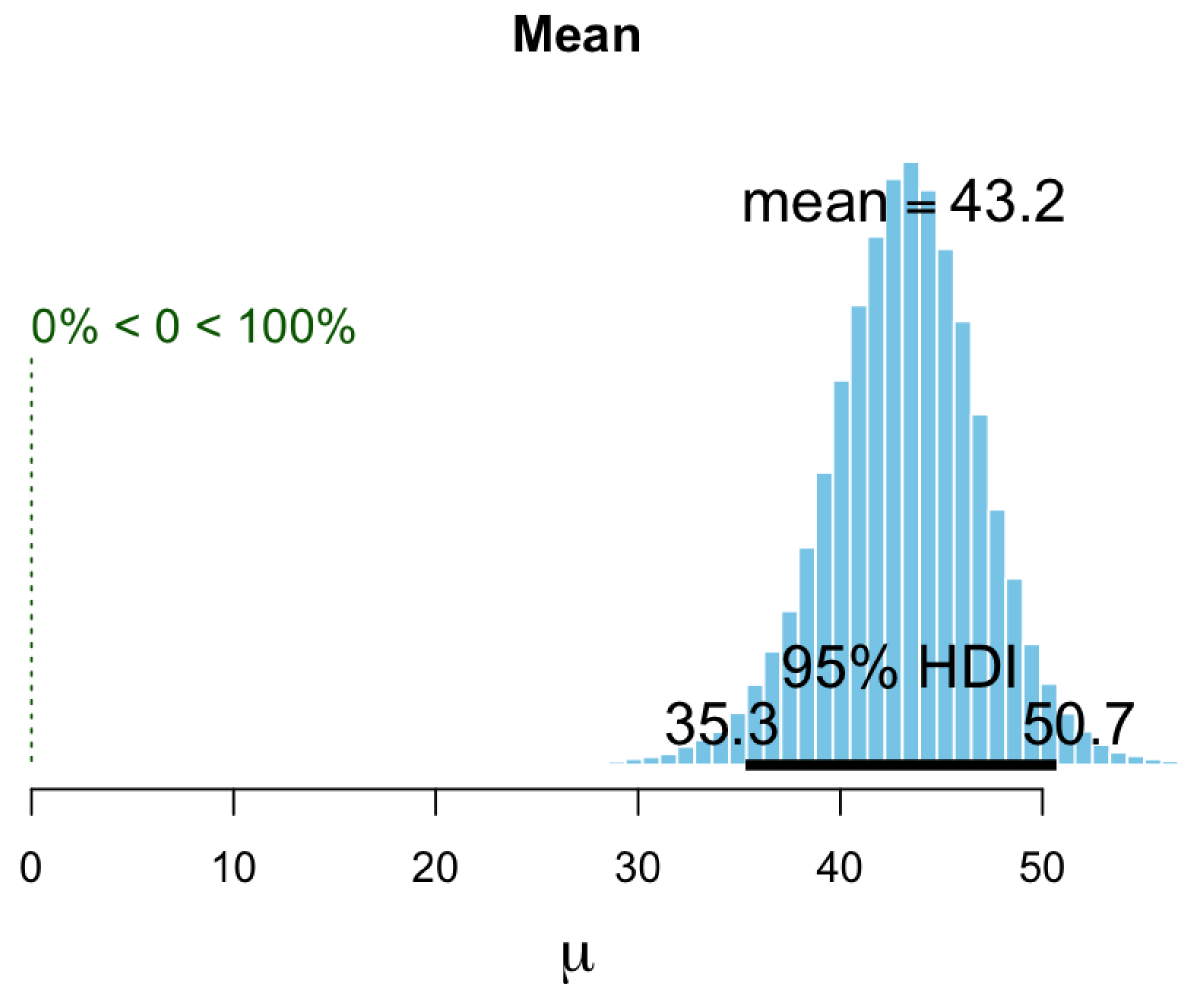
Let's use BEST!

```
library(BEST)
iq <- c(55, 44, 34, 18, 51, 40, 40, 49, 48, 46)
fit <- BESTmcmc(iq)
fit
```

```
MCMC fit results for BEST analysis:


       mean     sd median  HDIlo HDIup
mu    43.15  3.810  43.28 35.367 50.49
nu    27.42 26.647  18.91  1.001 81.59
sigma 11.00  3.754  10.44  4.857 18.38
```

Let's use BEST!

```
library(BEST)
iq <- c(55, 44, 34, 18, 51, 40, 40, 49, 48, 46)
fit <- BESTmcmc(iq)
plot(fit)
```



Try out BEST yourself!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

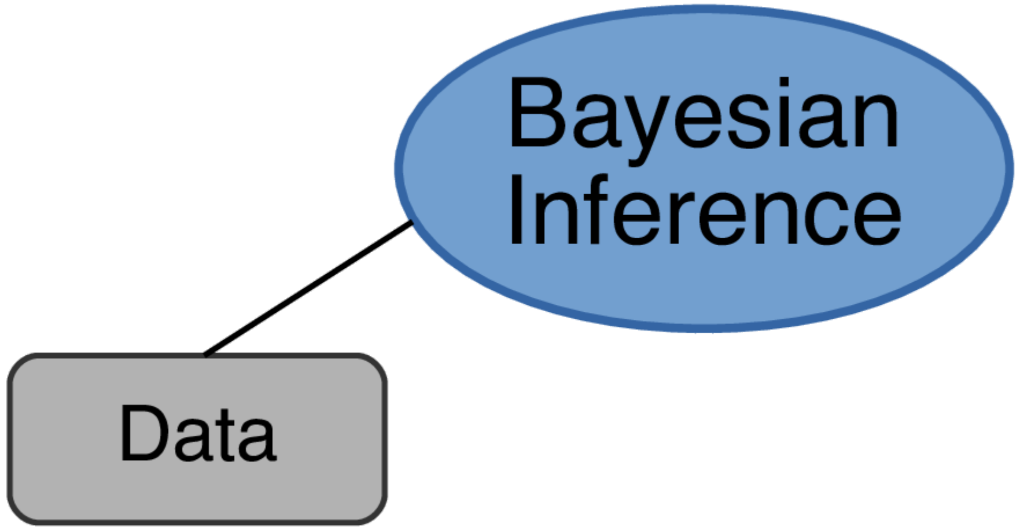
What have you learned? What did we miss?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

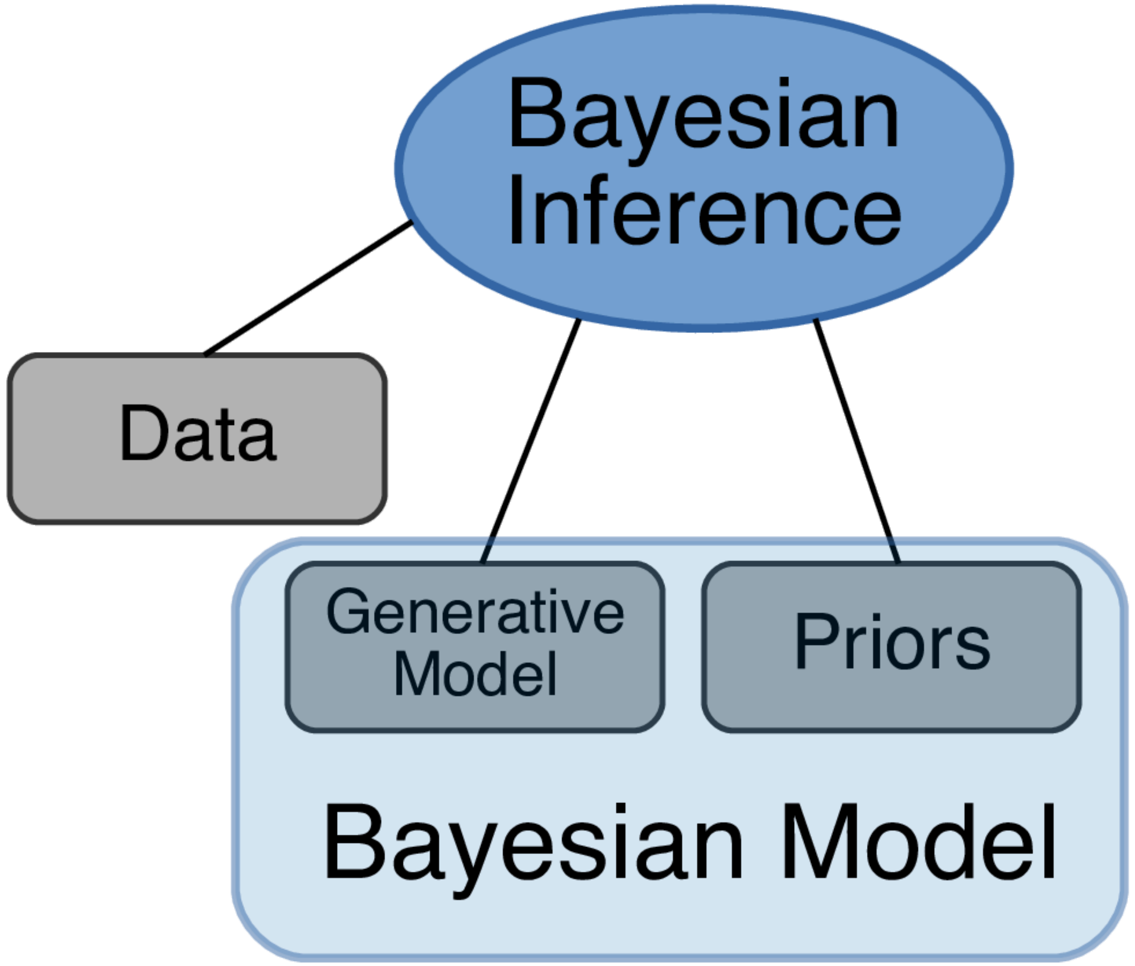
Rasmus Bååth
Data Scientist

We have covered

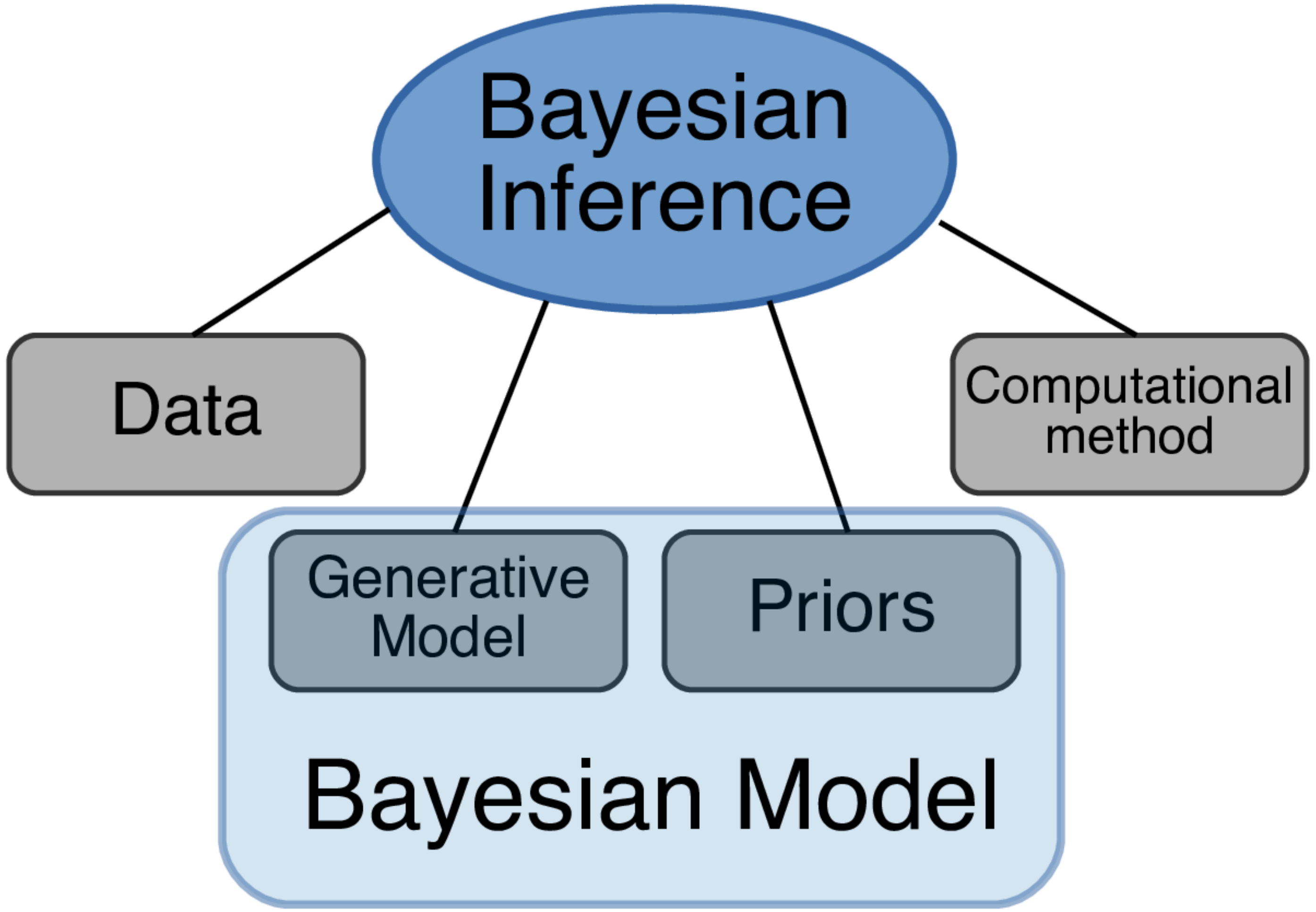
We have covered



We have covered



We have covered

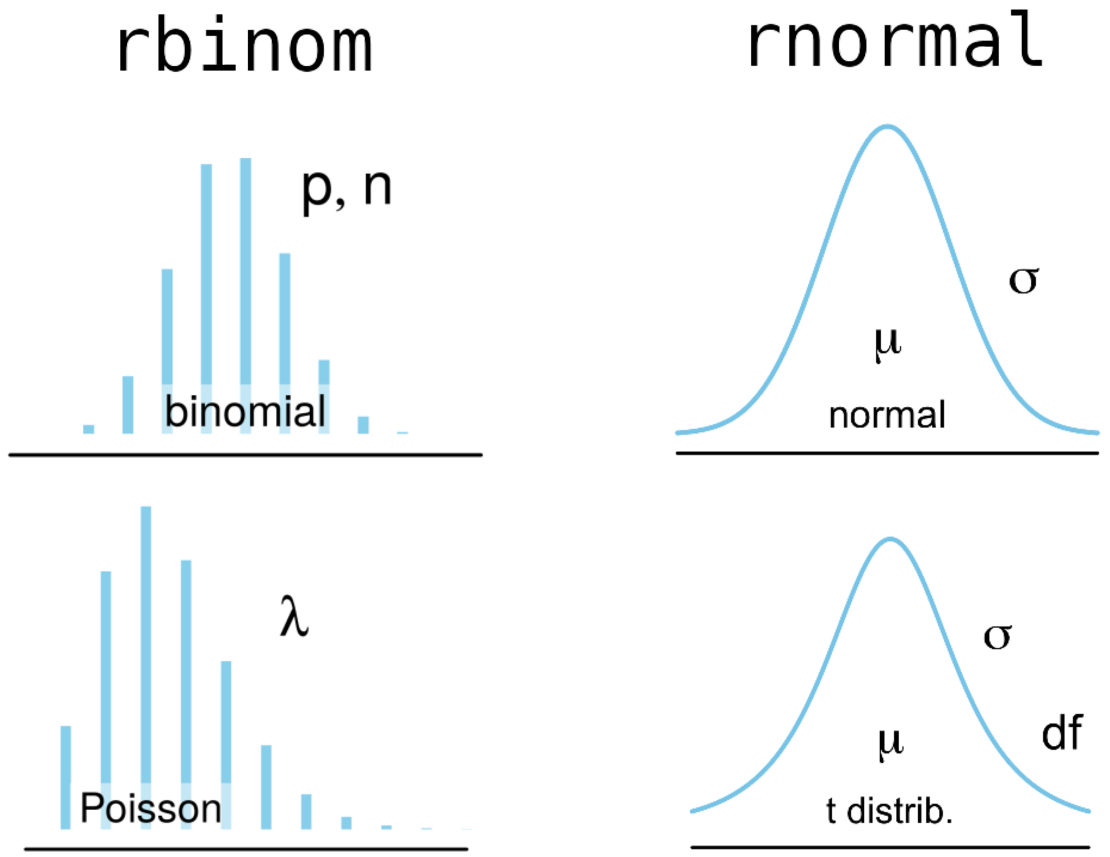


We have covered

- Computational methods
  - Rejection sampling
  - Grid approximation
  - Markov chain Monte Carlo (MCMC)

We have covered

- Generative models:



- Working with samples representing probability distributions:

```
> head(sample)
```

```
mu    sigma
39.39 10.18
39.39 21.77
40.90 20.26
45.45 13.20
34.84 12.70
40.90 12.70
```

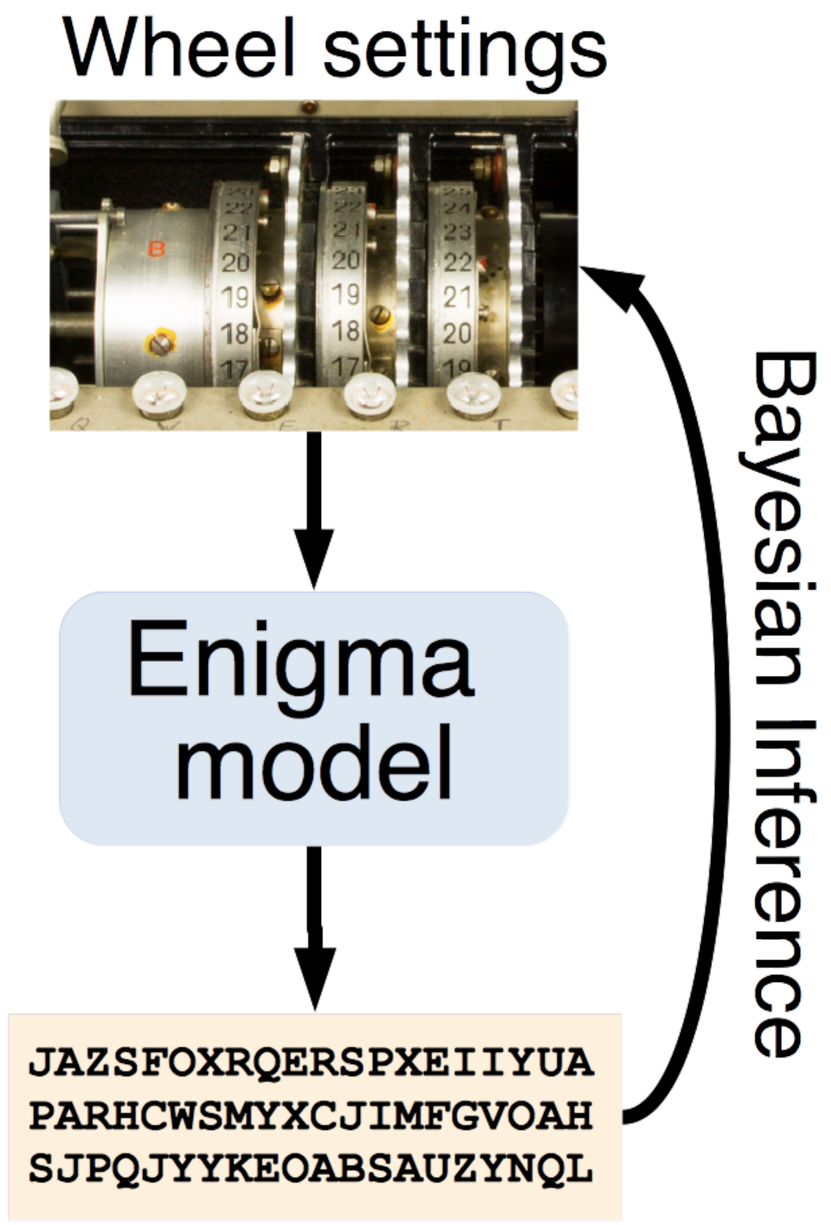
```
pred_iq <- rnorm(10000, mean = sample$mu, sd = sample$sigma)
sum(pred_iq >= 60) / length(pred_iq)
```

```
0.0901
```

Things we didn't cover

- That a Bayesian approach can be used for much more than simple models.
- How to decide what priors and models to use.
- How Bayesian statistics relate to classical statistics.
- More advanced computational methods.
- More advanced computational tools.

Things we didn't cover



Go explore Bayes!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Bye and thanks!

Let's practice!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R