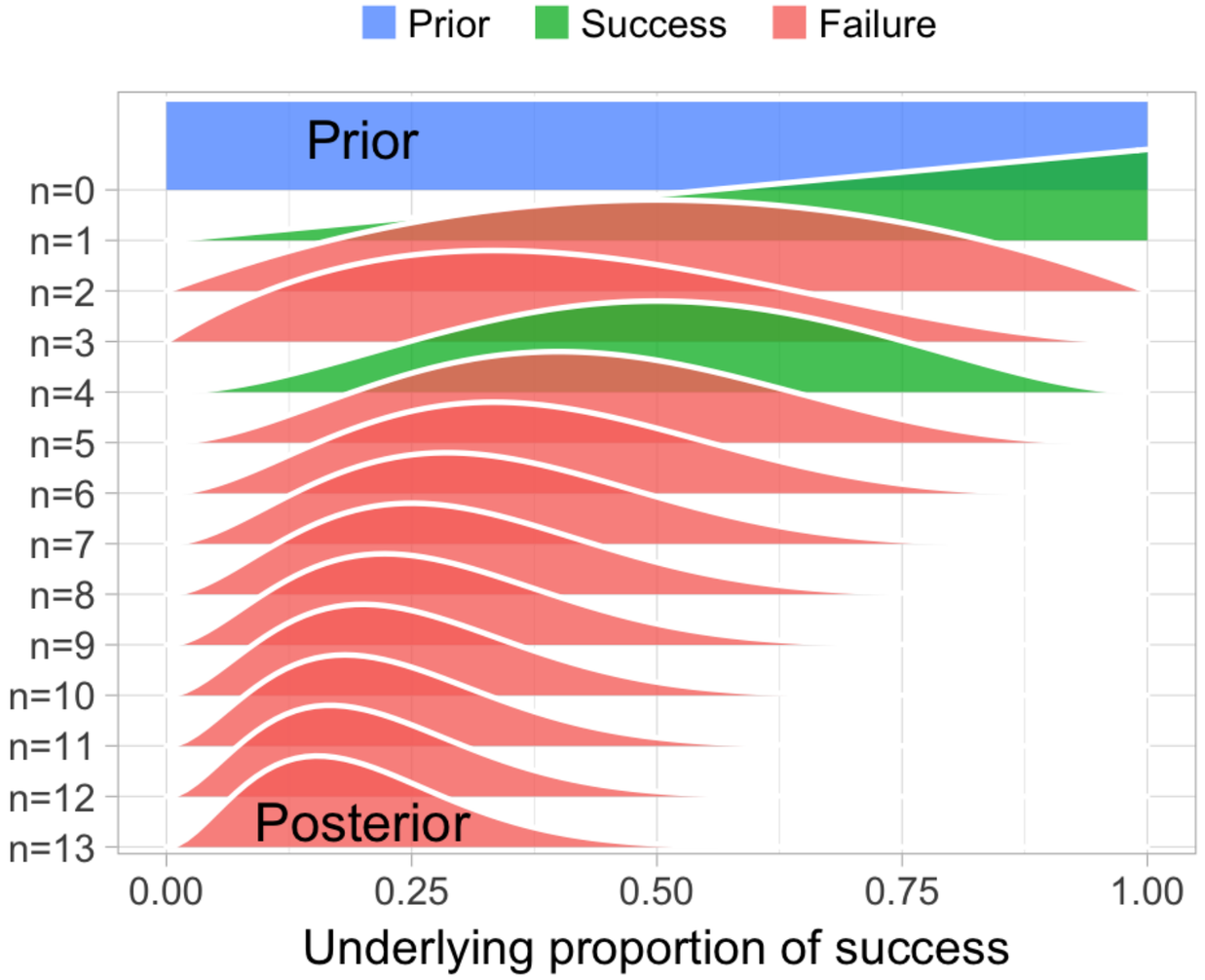
The parts needed for Bayesian inference

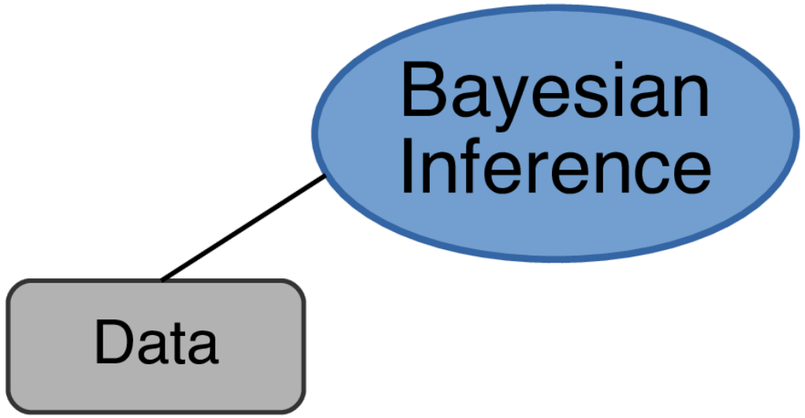
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

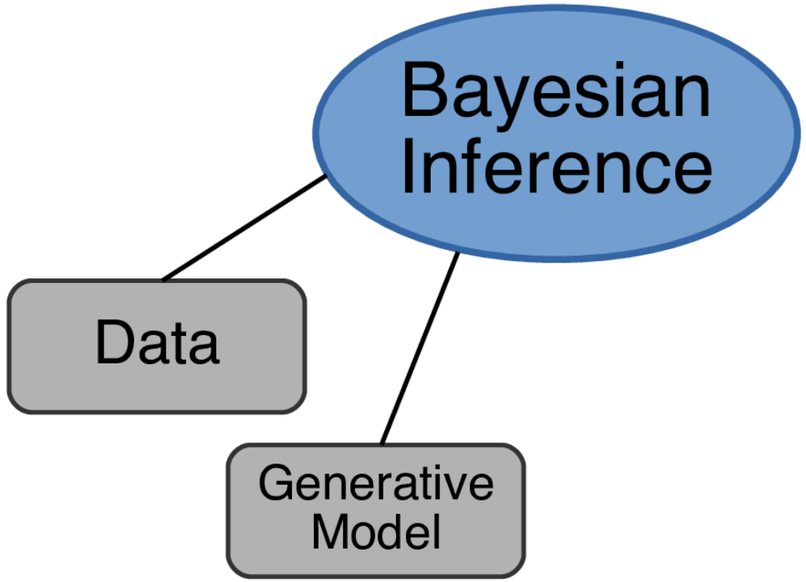
Rasmus Bååth
Data Scientist

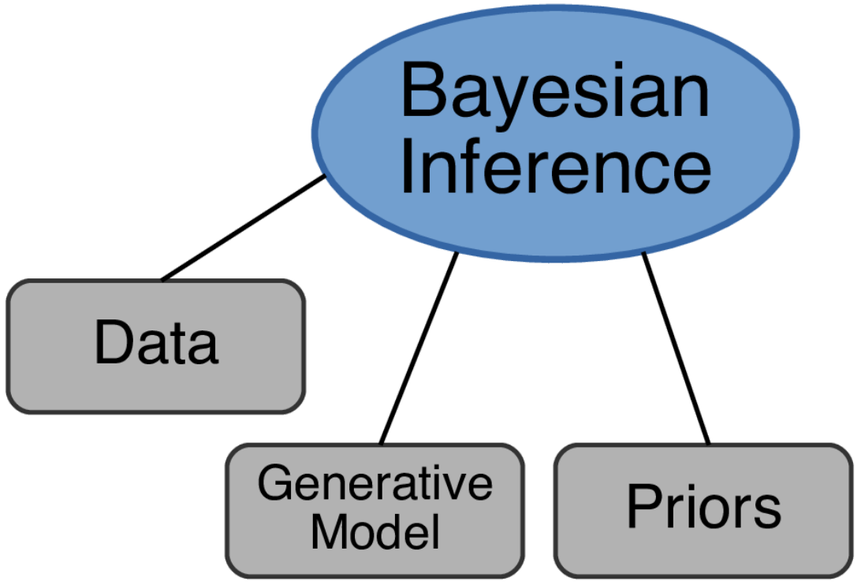
datacamp


Prior
Success
Failure
Prior
n=0
n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
n=11
n=12
n=13
Posterior
0.00
0.25
0.50
0.75
1.00
Underlying proportion of success

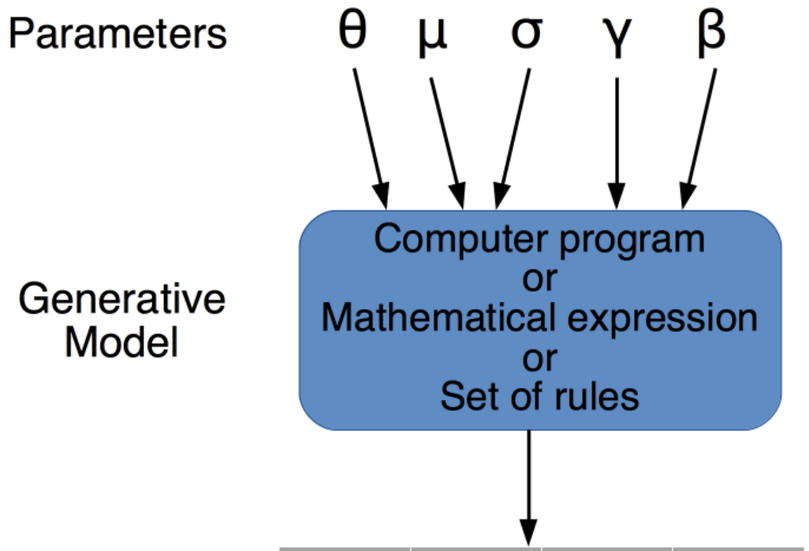

Bayesian
Inference


Bayesian
Inference
Data


Bayesian Inference
Data
Generative Model


Bayesian Inference
Data
Generative Model
Priors

What is a generative model?



x	y	animal	cuteness
4	0.34	Dog	13
1	0.63	Dog	12
-7	0.22	Cat	7
3	0.62	Dog	11
...	...	...	...

Generative zombie drug model

Generative zombie drug model

```
# Parameters
prop_success <- ???
n_zombies <- ???
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- ???
}
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success}
```

```
data
```

```
FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
data <- as.numeric(data)
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
data <- as.numeric(data)
data
```

```
0 0 0 1 0 0 0 0 1 0 1 0 0
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
data <- as.numeric(data)
data
```

```
0 0 1 0 0 0 0 0 0 0 0 0 0
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
data <- as.numeric(data)
data
```

```
0 1 0 1 1 0 0 1 0 1 0 0 0
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
data <- as.numeric(data)
data
```

```
0 0 0 0 0 0 0 1 0 0 0 0 0
```

Generative zombie drug model

```
# Parameters
prop_success <- 0.15
n_zombies <- 13
# Simulating data
data <- c()
for(zombie in 1:n_zombies) {
  data[zombie] <- runif(1, min = 0, max = 1) < prop_success
}
data <- as.numeric(data)
data
```

```
0 0 0 0 1 0 0 0 0 1 0 1 0
```

Take this model for a spin!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Using a generative model

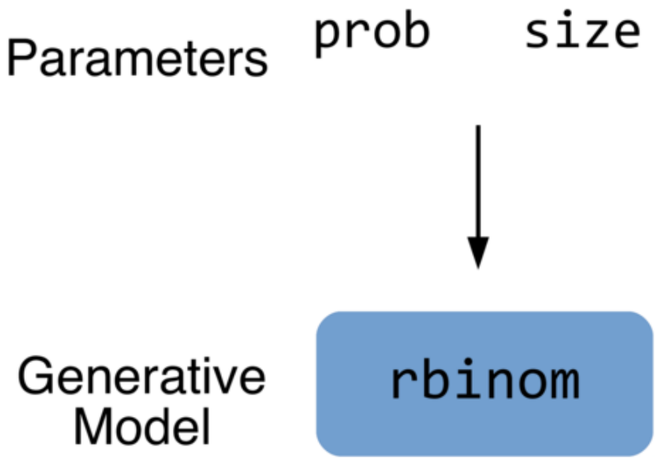
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

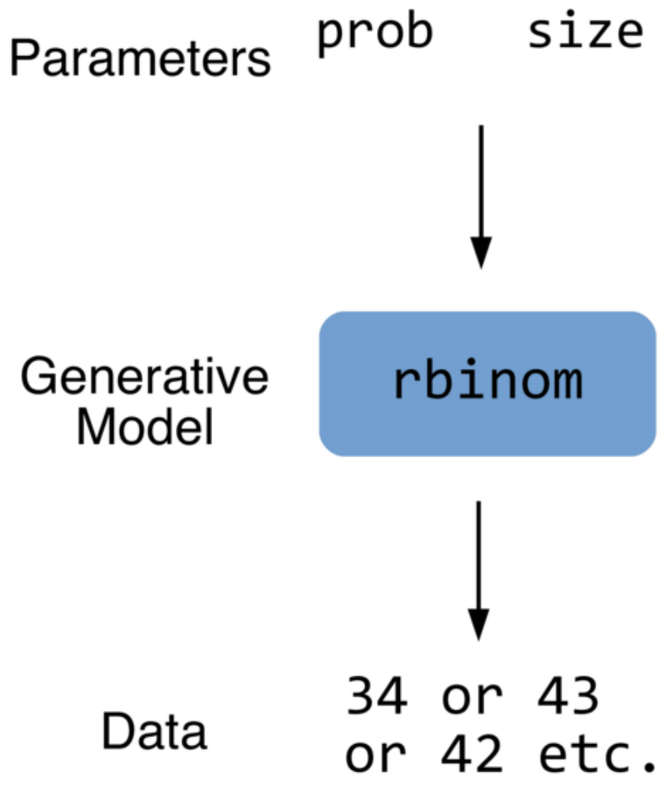
Rasmus Bååth
Data Scientist

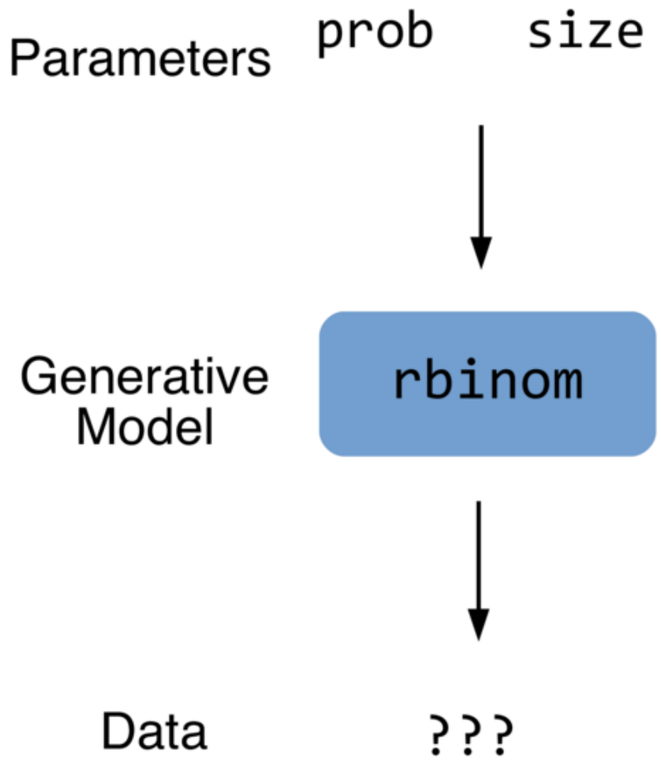
datacamp

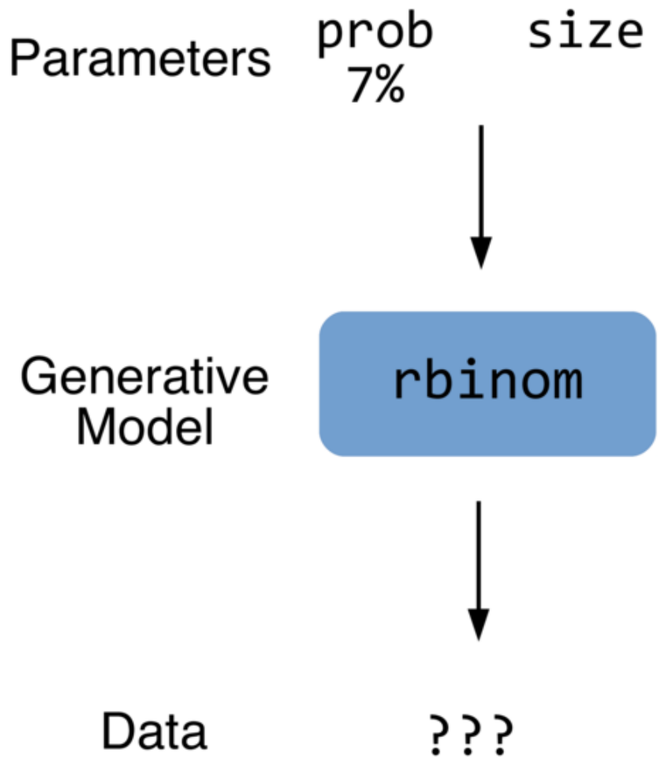
```
rbinom(n, size, prob)
```

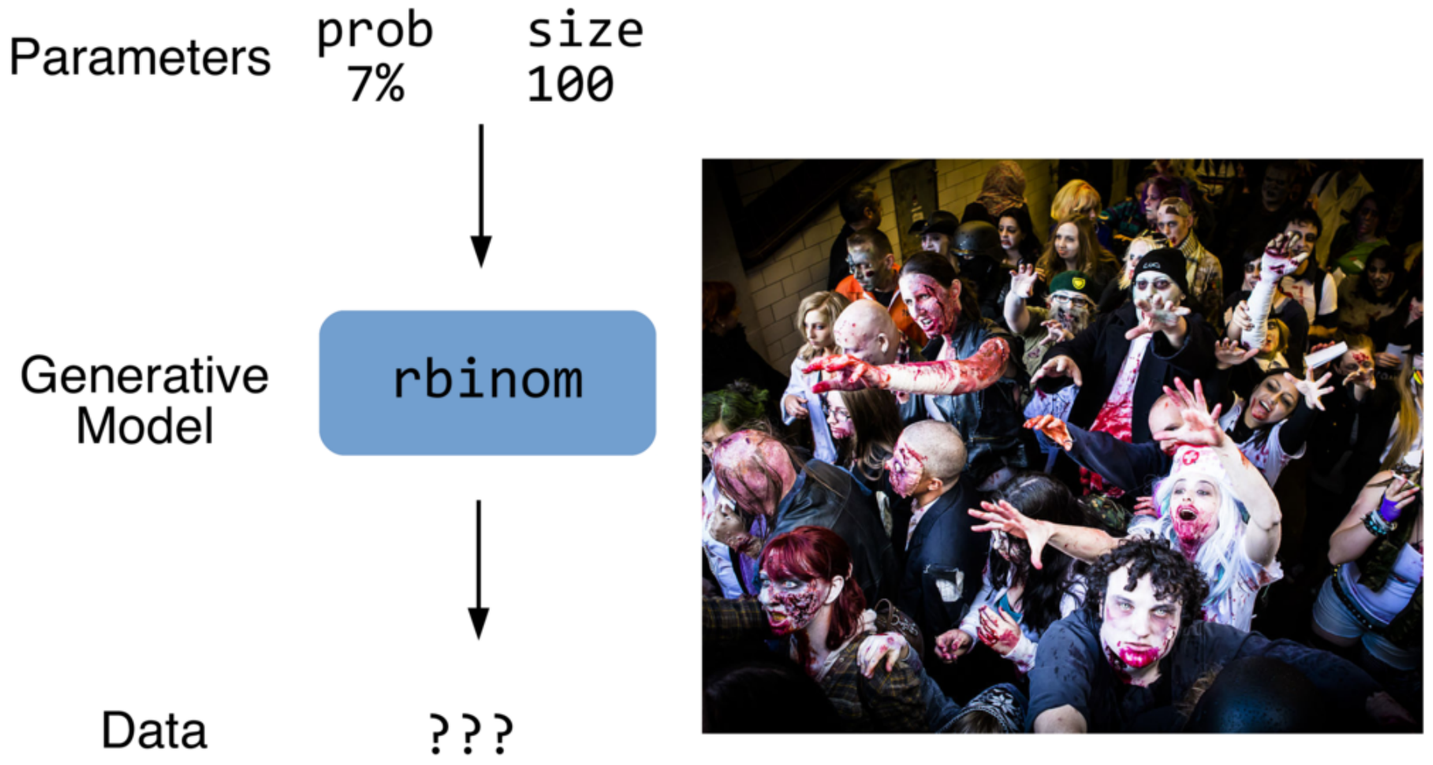

Generative
Model
rbinom


Parameters
prob
size
Generative Model
rbinom


Parameters
prob
size
Generative Model
rbinom
Data
34 or 43
or 42 etc.


Parameters
prob
size
Generative Model
rbinom
Data
???


Parameters
prob
7%
size
Generative
Model
rbinom
Data
???



[1] https://commons.wikimedia.org/wiki/File:The_Hoard_III_-_Flickr_-_SoulStealer.co.uk.jpg

```
cured_zombies <- rbinom(n = 100000, size = 100, prob = 0.07)
```

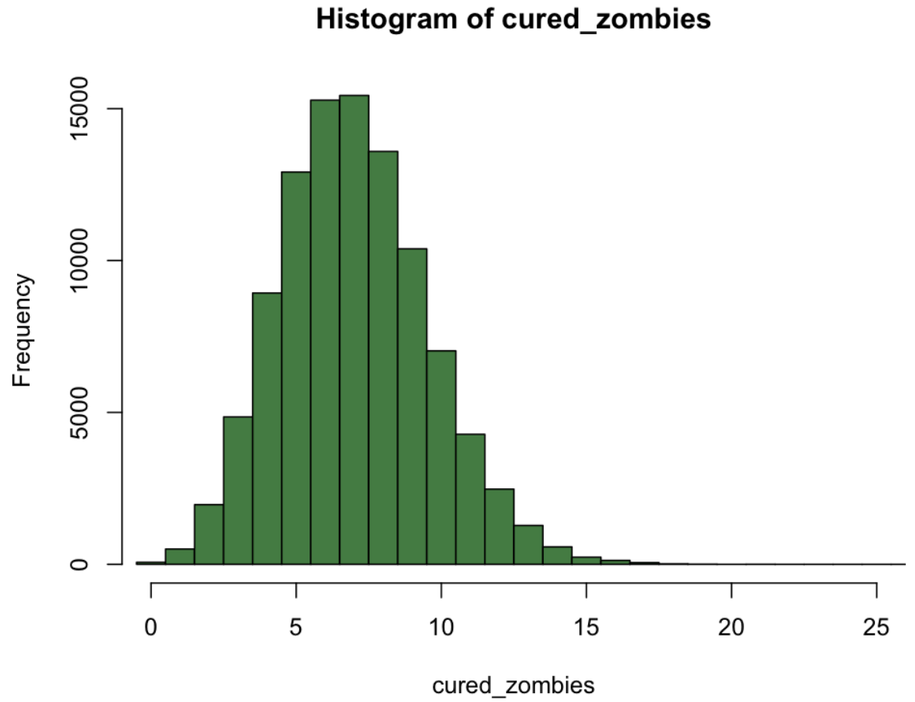
```
cured_zombies <- rbinom(n = 100000, size = 100, prob = 0.07)
```

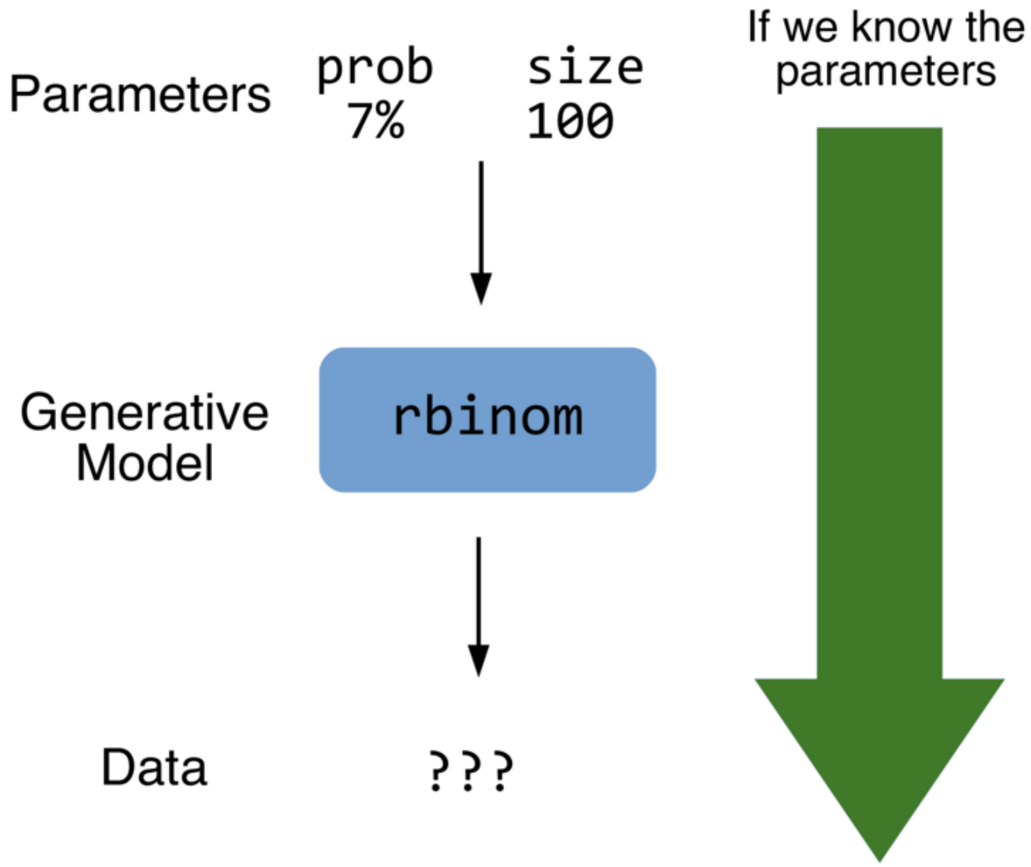
```
cured_zombies
```

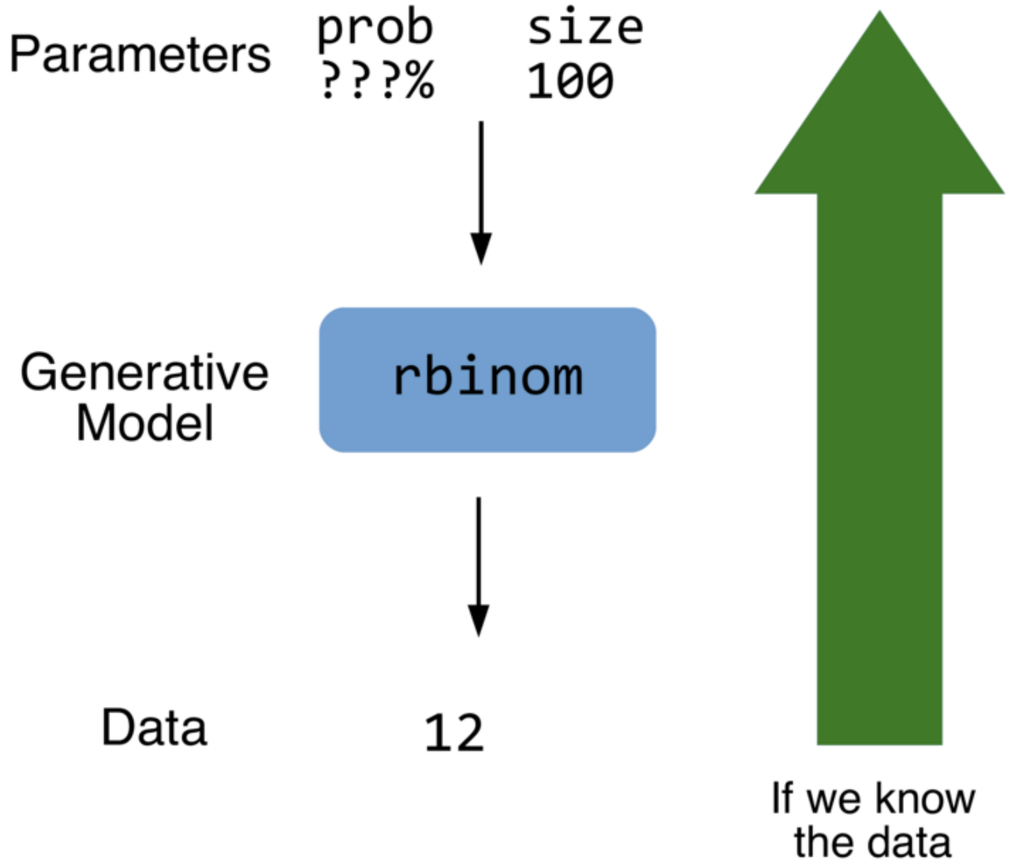
```
 8  6  5  8 11  5  7  7  5  8  7 11  6  8  8
 9  9  4  7  4  8  7  5  5  6  9 14  4  4  8
 ...
```

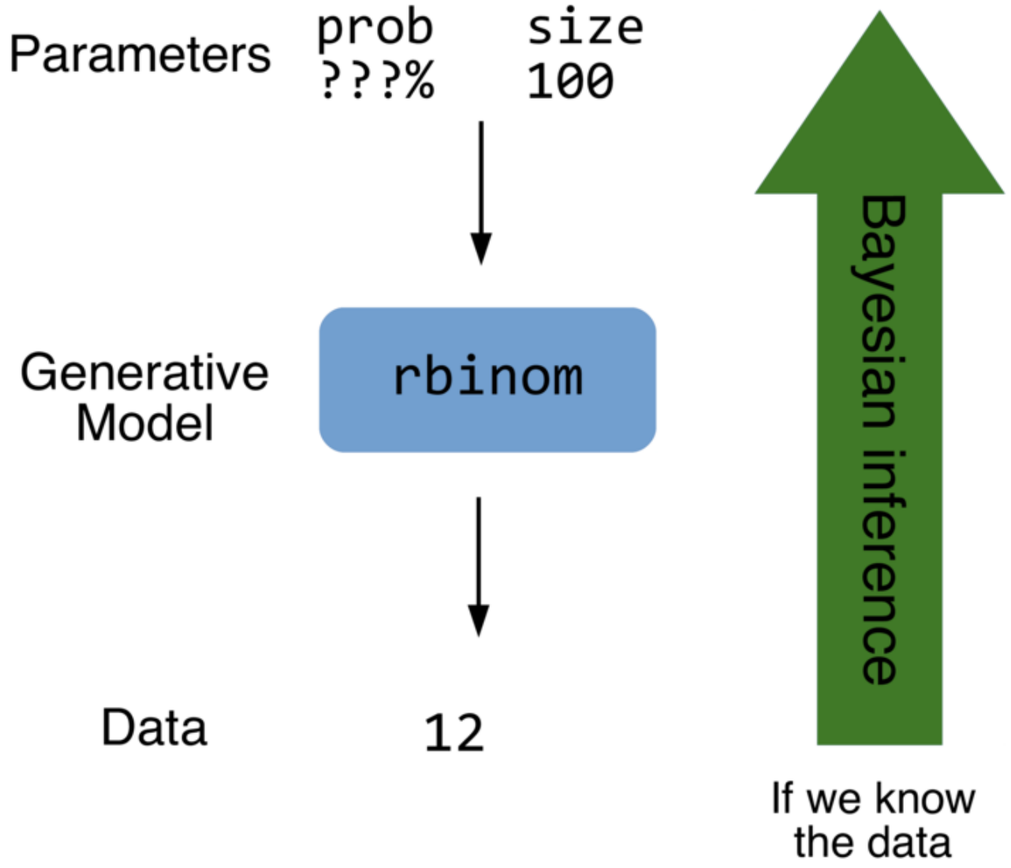
```
cured_zombies <- rbinom(n = 100000, size = 100, prob = 0.07)
```

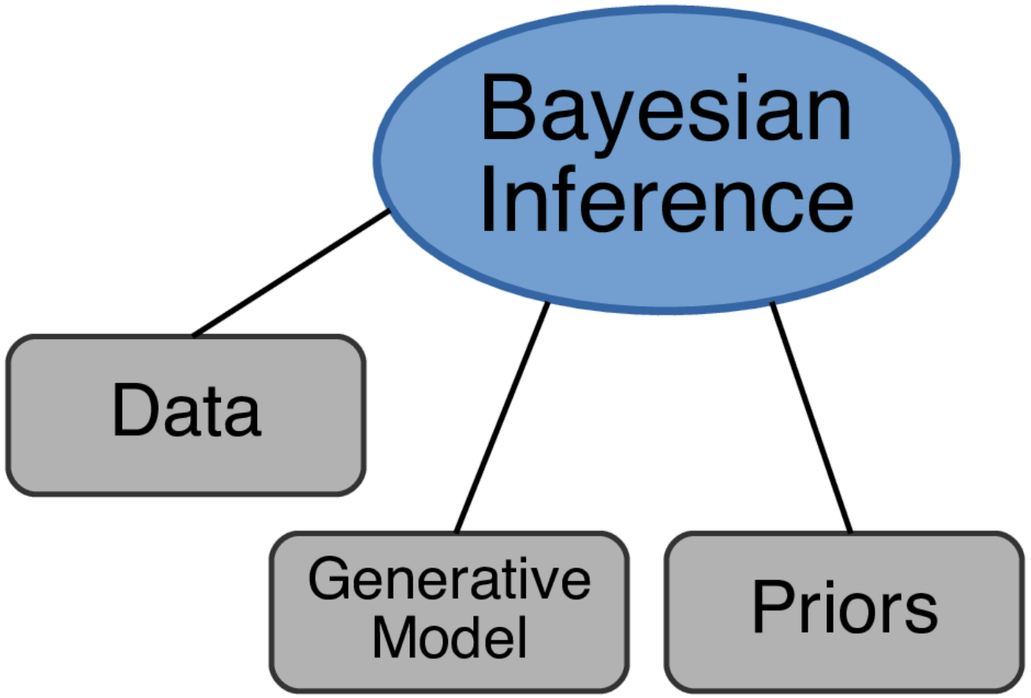
```
hist(cured_zombies)
```

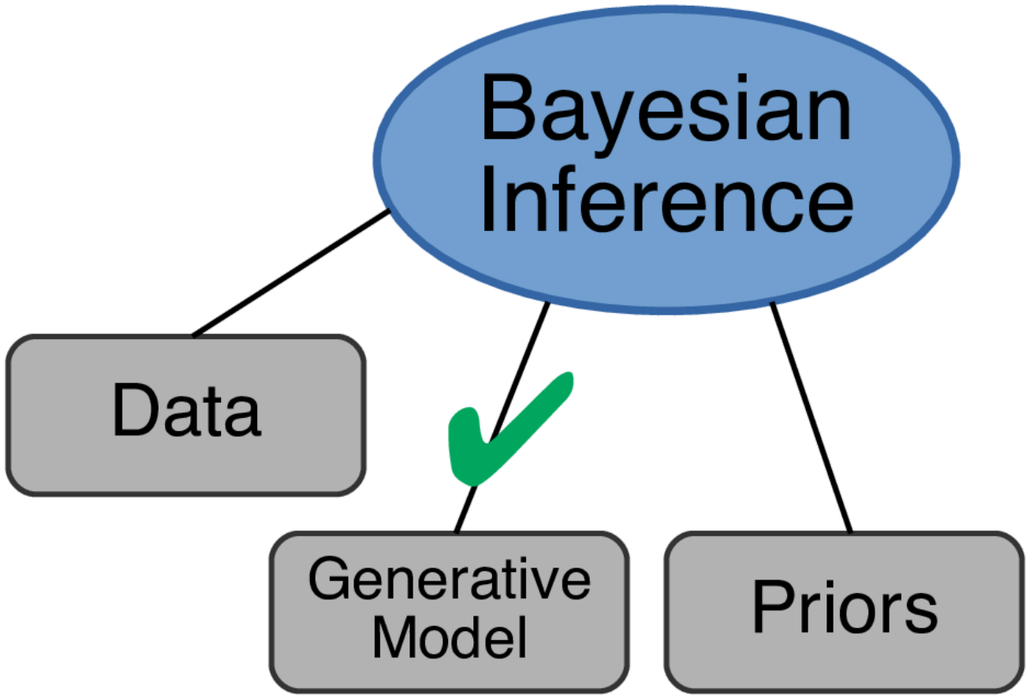



Parameters
prob
7%
size
100
If we know the parameters
Generative Model
rbinom
Data
???


Parameters
prob
???%
size
100
Generative Model
rbinom
Data
12
If we know the data


Parameters
prob
???%
size
100
Generative Model
rbinom
Data
12
Bayesian inference
If we know the data


Bayesian Inference
Data
Generative Model
Priors


Bayesian Inference
Data
Generative Model
Priors

[1] https://commons.wikimedia.org/wiki/File:The_Hoard_III_-_Flickr_-_SoulStealer.co.uk.jpg


dmoz
Google
YAHOO!
Ask Jeeves
AOL
S
E
O

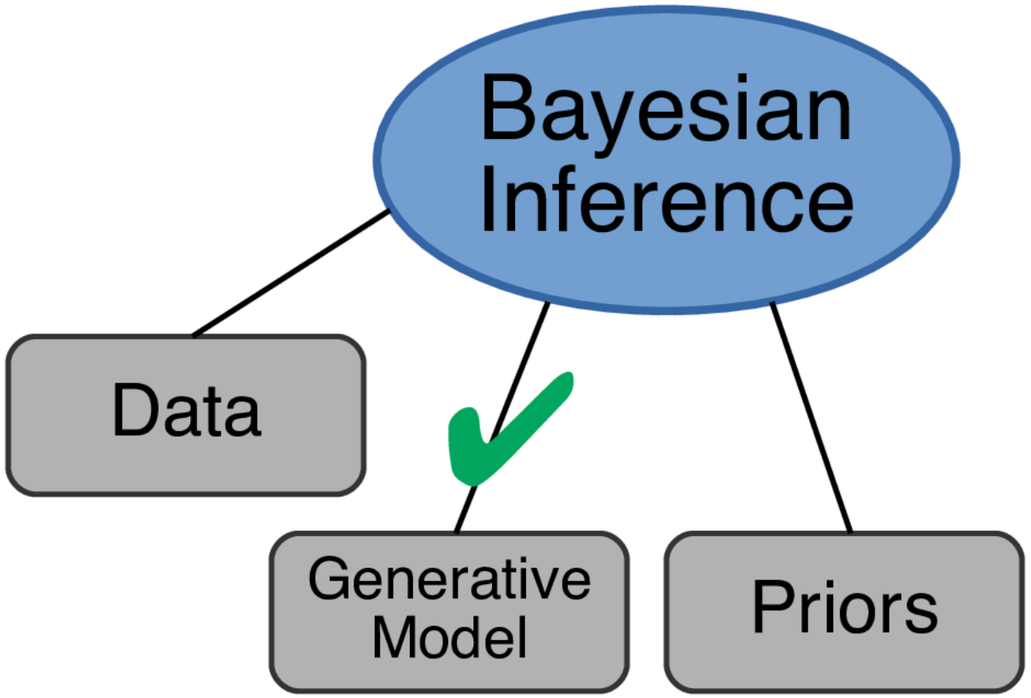
How many visitors?

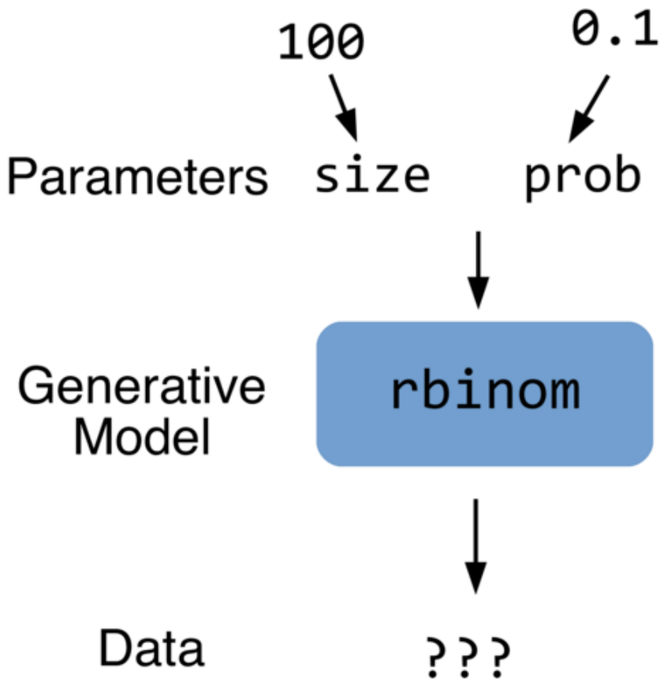
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

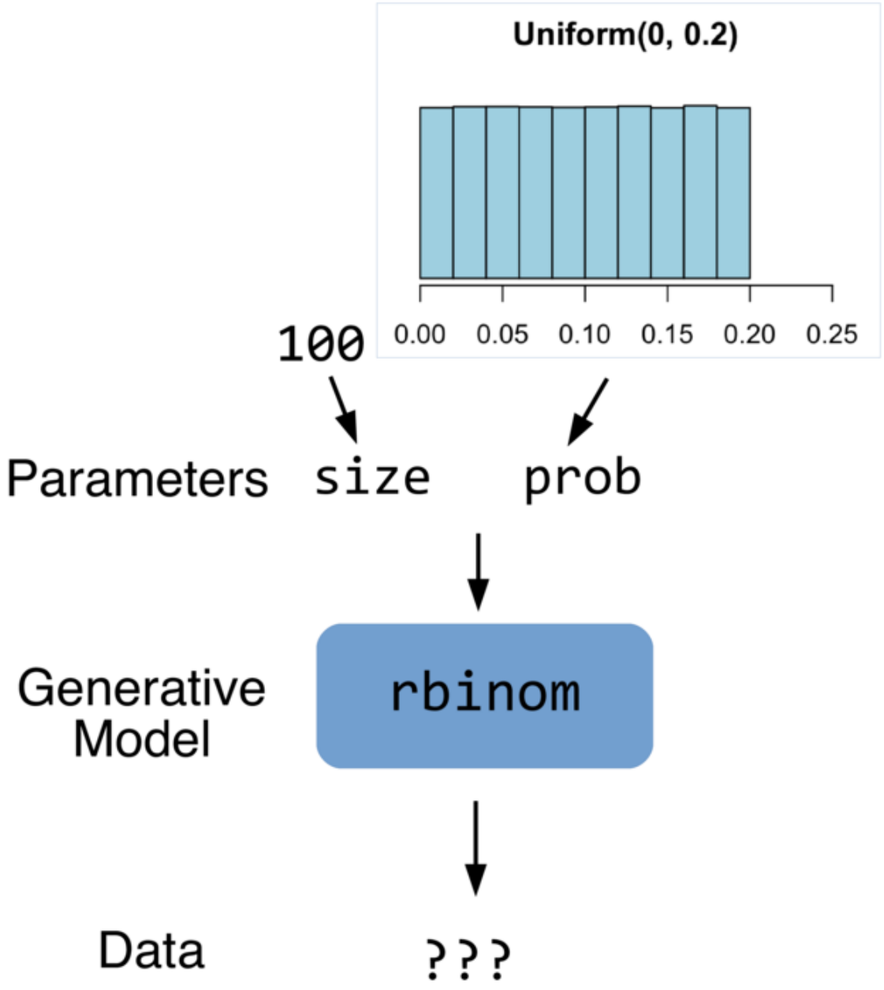
One more part

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Rasmus Bååth
Data Scientist


Bayesian Inference
Data
Generative Model
Priors


100
0.1
Parameters
size
prob
Generative Model
rbinom
Data
???


Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
???

rbinom Bayesian model 2

```
n_samples <- 100000
n_ads_shown <- 100
proportion_clicks <- 0.1
n_visitors <- rbinom(n_samples, size = n_ads_shown,
                     prob = proportion_clicks)
```

runif - Random Uniform samples

runif - Random Uniform samples

```
proportion_clicks <- runif(n = 6, min = 0.0, max = 1.0)
```

```
proportion_clicks
```

```
0.05  0.58  0.21  0.61  0.69  0.39
```

runif - Random Uniform samples

```
proportion_clicks <- runif(n = 6, min = 0.0, max = 1.0)
n_clicks <- rbinom(n = 6, size = 100, proportion_clicks)
```

```
proportion_clicks
```

```
0.05  0.58  0.21  0.61  0.69  0.39
```

runif - Random Uniform samples

```
proportion_clicks <- runif(n = 6, min = 0.0, max = 1.0)
n_clicks <- rbinom(n = 6, size = 100, proportion_clicks)
```

```
proportion_clicks
```

```
0.05  0.58  0.21  0.61  0.69  0.39
```

```
n_clicks
```

```
7     59    13    63    67    29
```

Try this in practice!

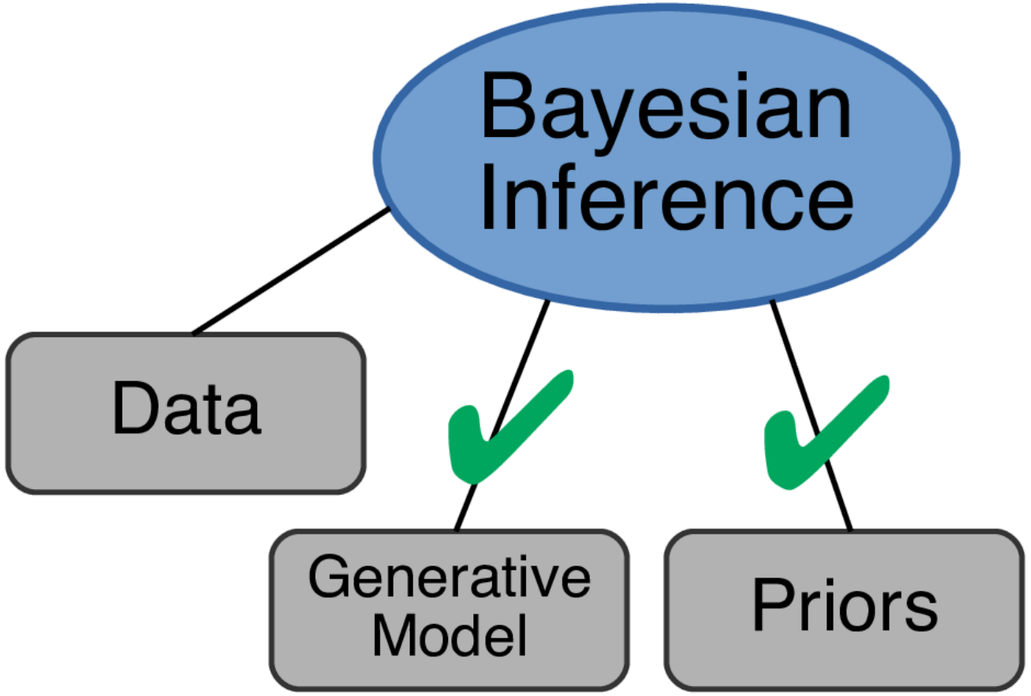
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

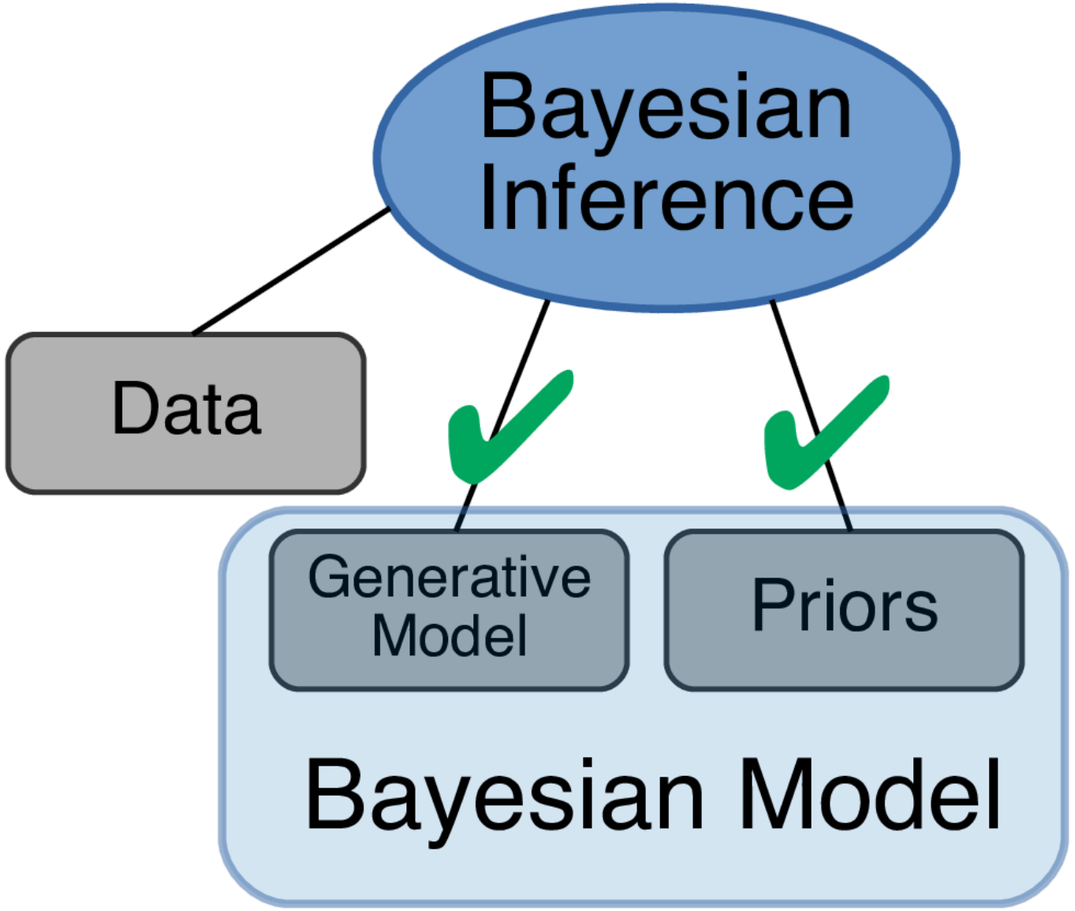
Bayesian models and conditioning

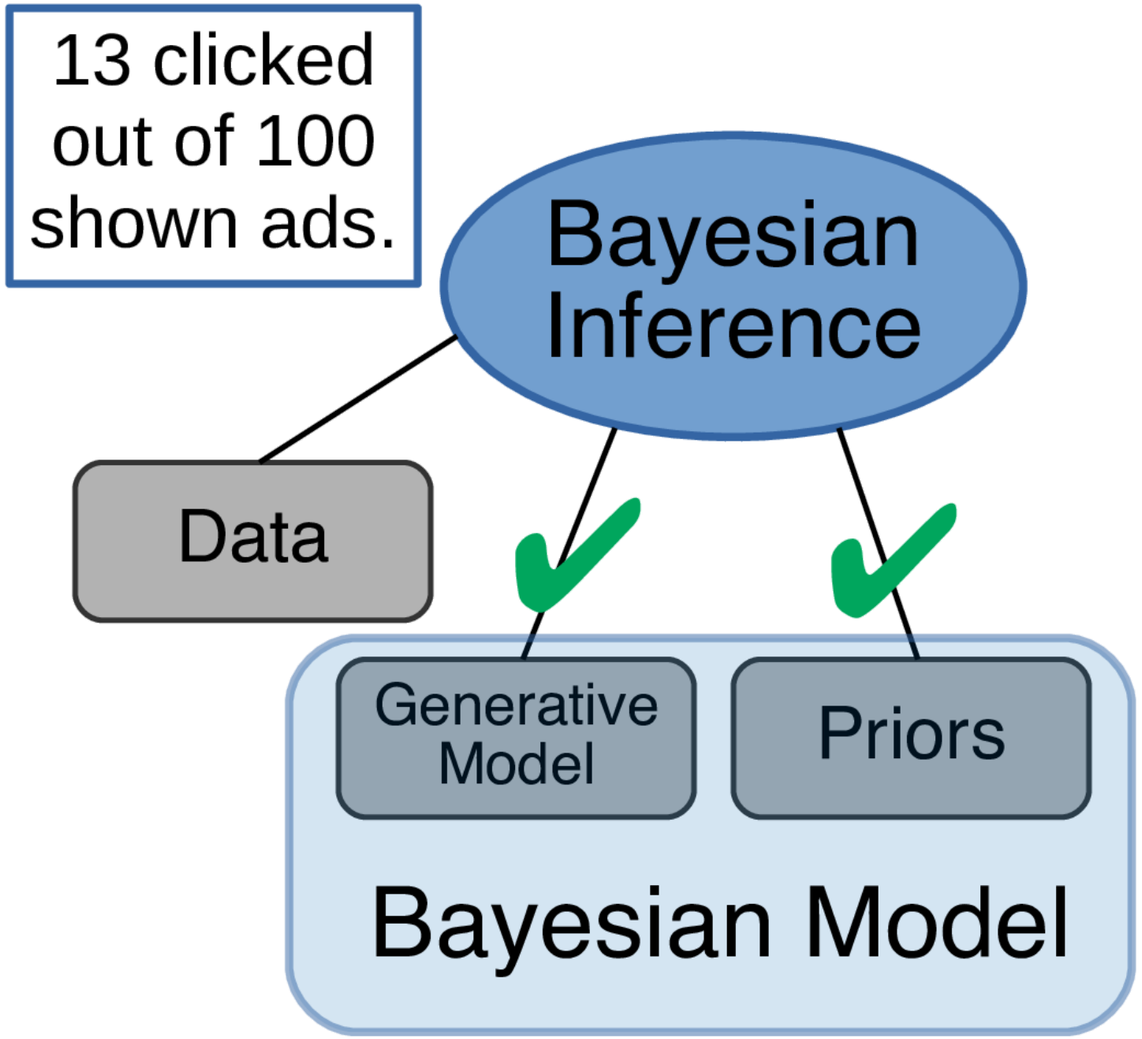
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

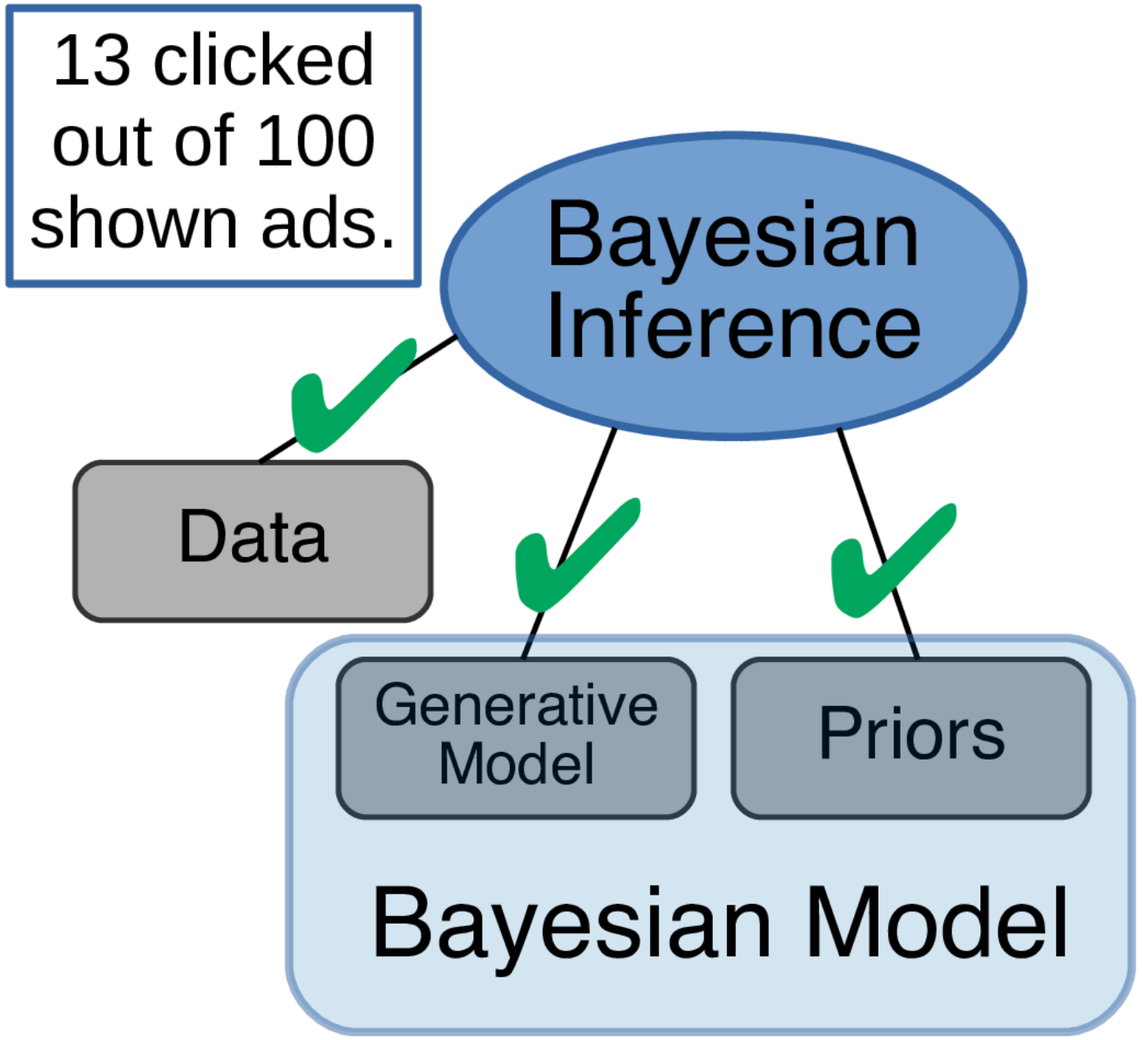
Rasmus Bååth
Data Scientist

datacamp


Bayesian Inference
Data
Generative Model
Priors


Bayesian Inference
Data
Generative Model
Priors
Bayesian Model


13 clicked out of 100 shown ads.
Bayesian Inference
Data
Generative Model
Priors
Bayesian Model

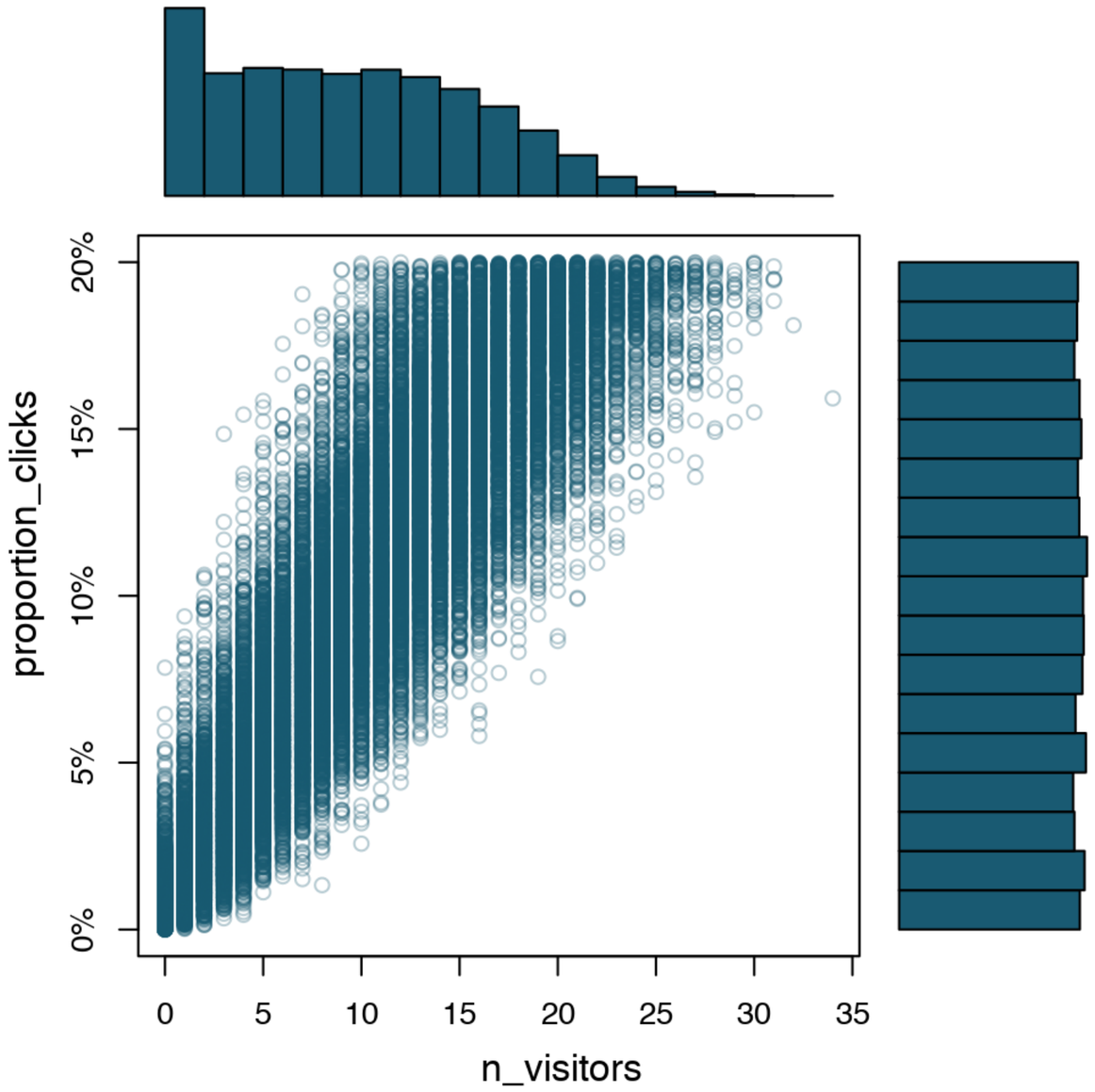

13 clicked out of 100 shown ads.
Bayesian Inference
Data
Generative Model
Priors
Bayesian Model

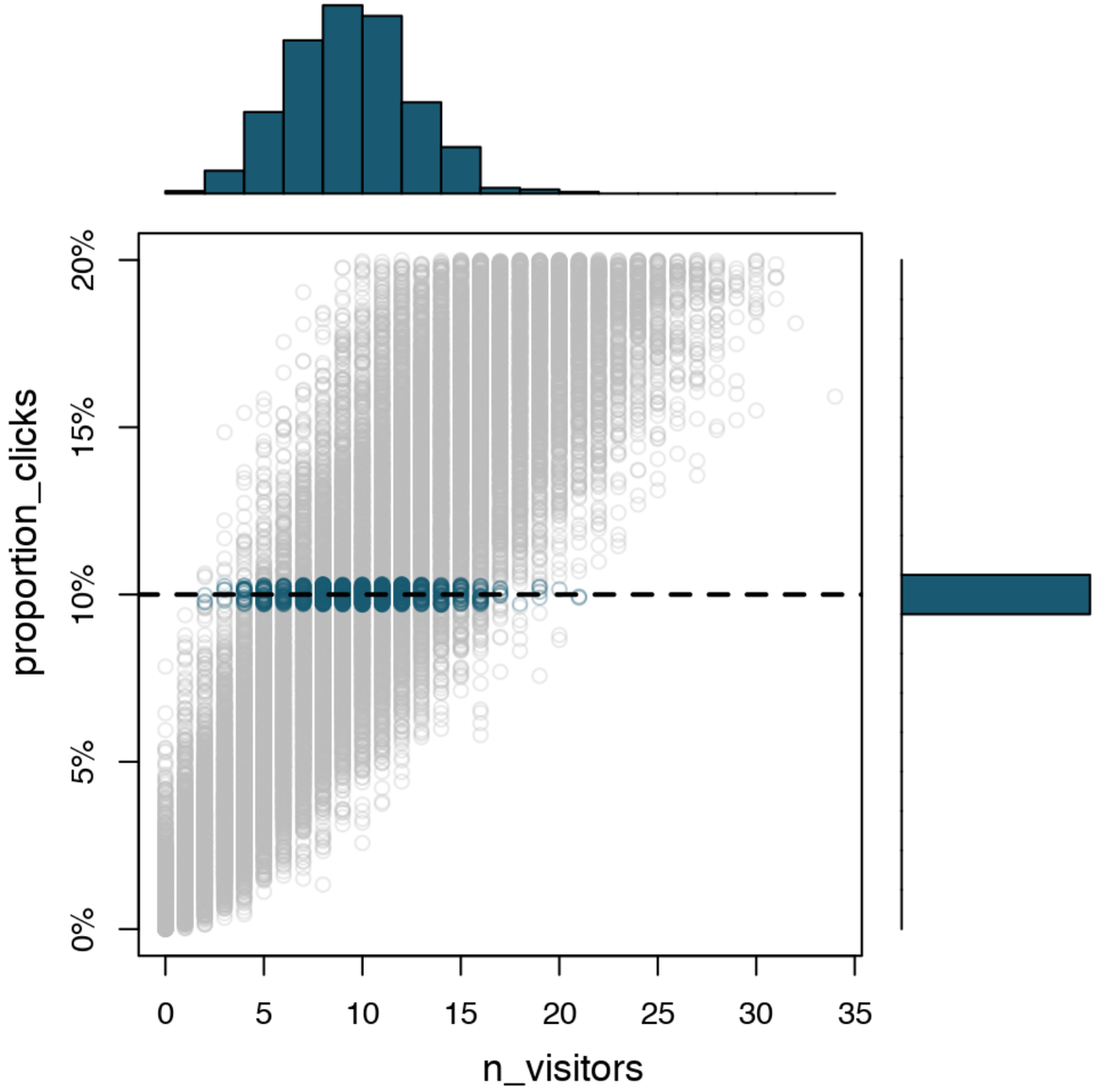
```
n_samples <- 100000
n_ads_shown <- 100
proportion_clicks <- runif(n_samples, min = 0.0, max = 0.2)
n_visitors <- rbinom(n = n_samples, size = n_ads_shown,
                     prob = proportion_clicks)
```

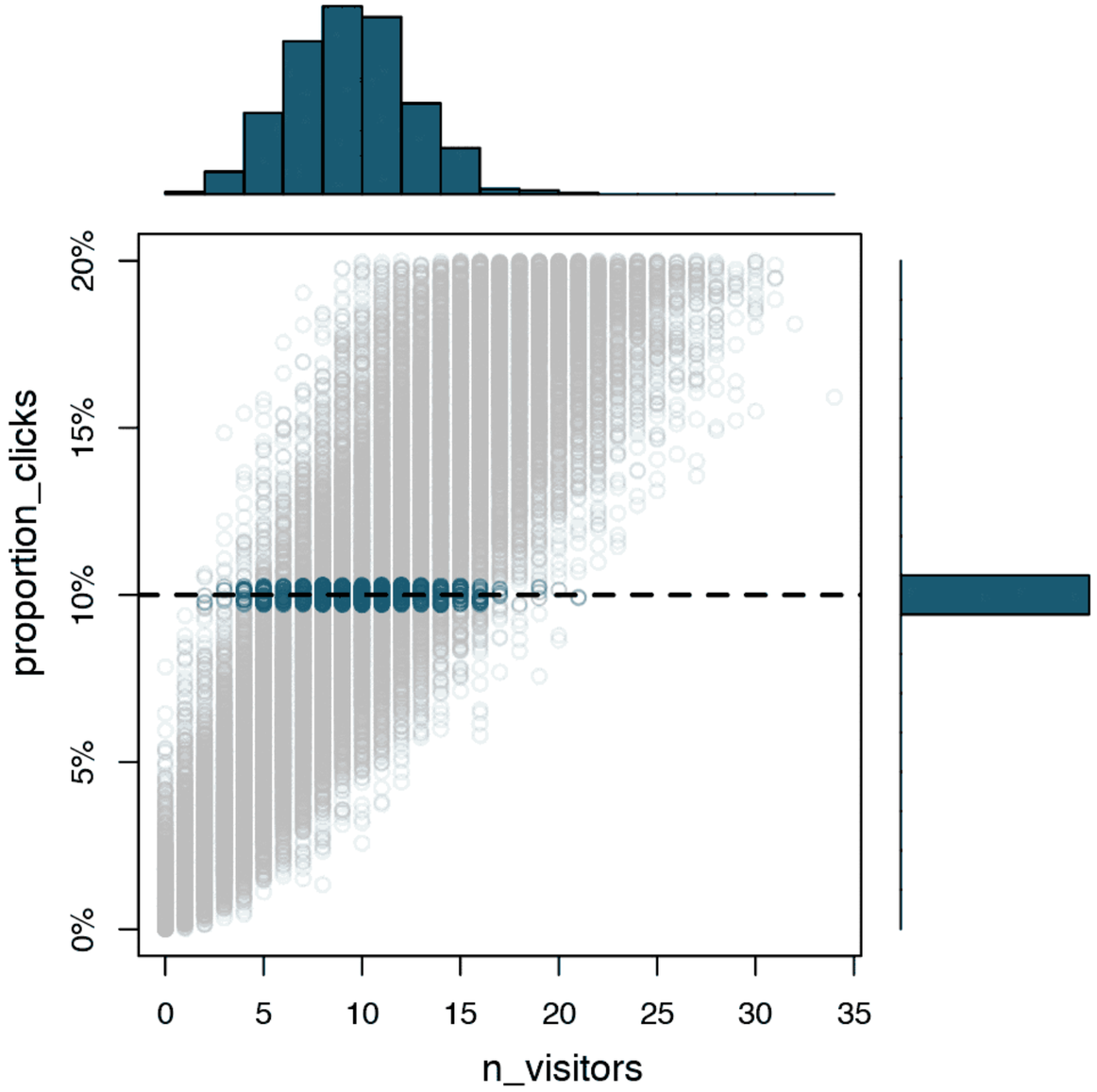
```
n_samples <- 100000
n_ads_shown <- 100
proportion_clicks <- runif(n_samples, min = 0.0, max = 0.2)
n_visitors <- rbinom(n = n_samples, size = n_ads_shown,
                     prob = proportion_clicks)
prior <- data.frame(proportion_clicks, n_visitors)
```

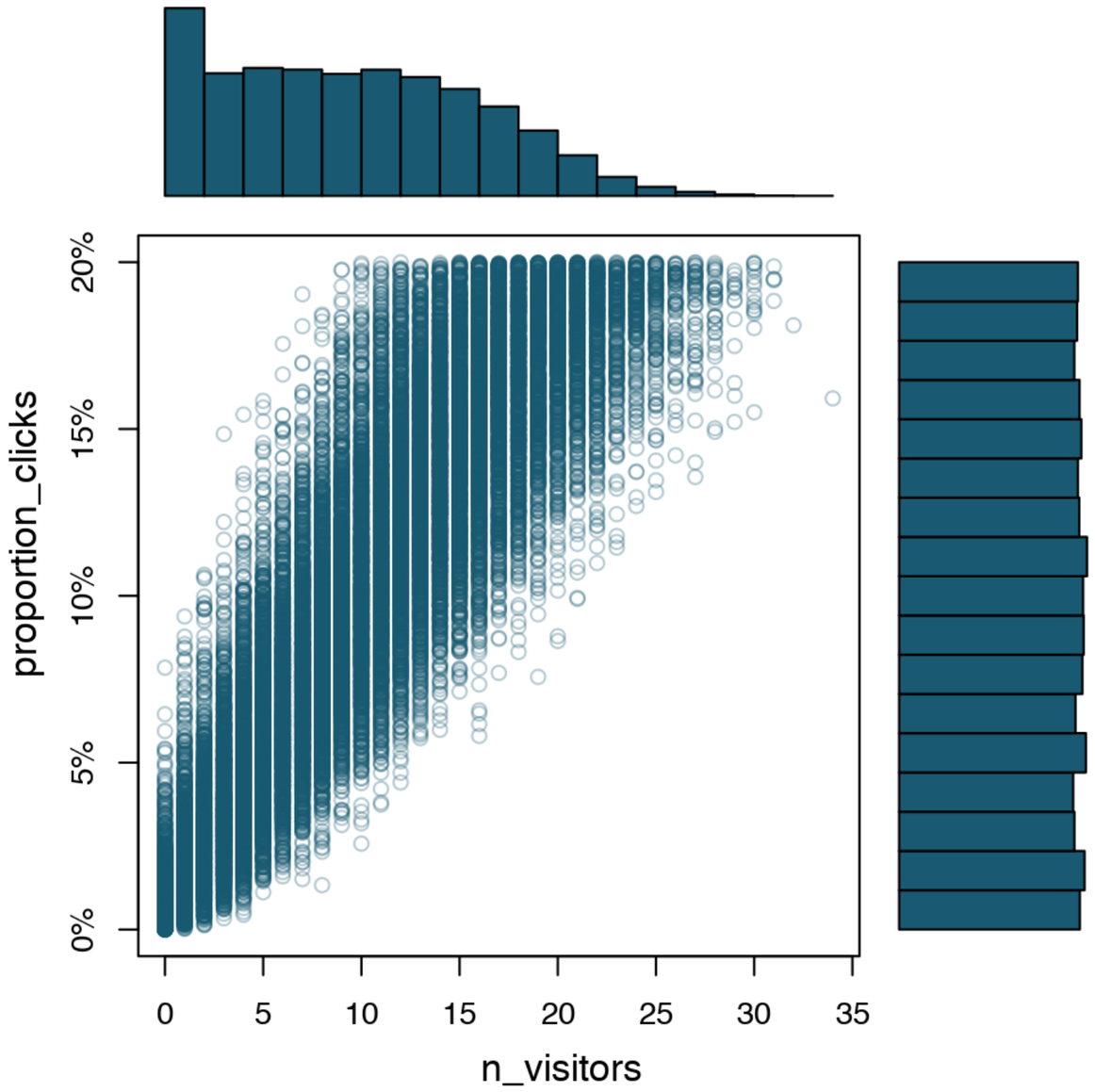
```
prior
```

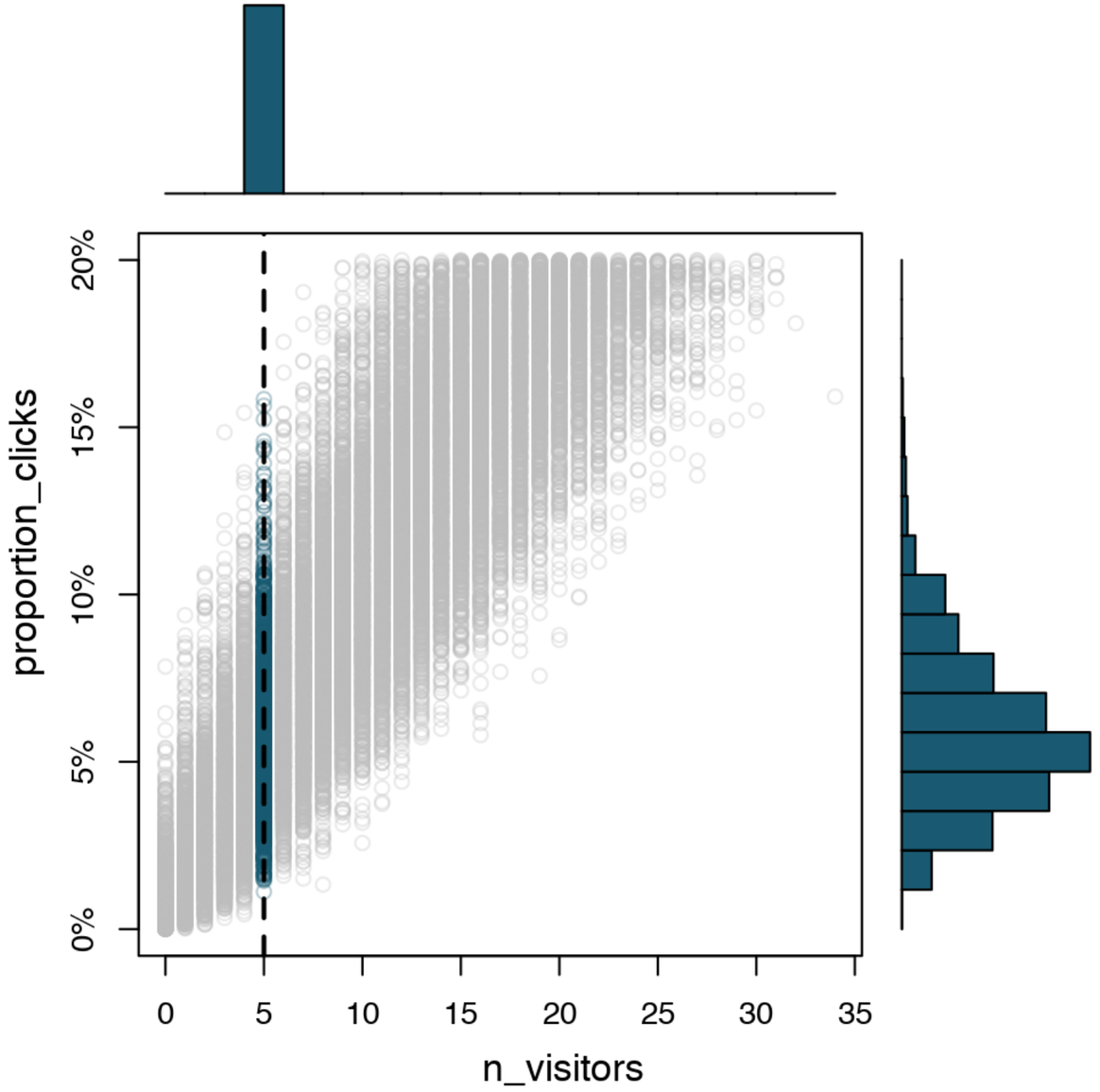
```
   proportion_clicks n_visitors
1               0.12         10
2               0.04          3
3               0.11         14
4               0.15         14
5               0.15         12
6               0.16         13
7               0.04          6
8               0.04          3
9               0.09         10
10              0.04          3
11              0.08          8
12              0.13         12
13              0.02          3
14              0.18         19
15              0.04          5
16              0.10         10
...              ...        ...
```

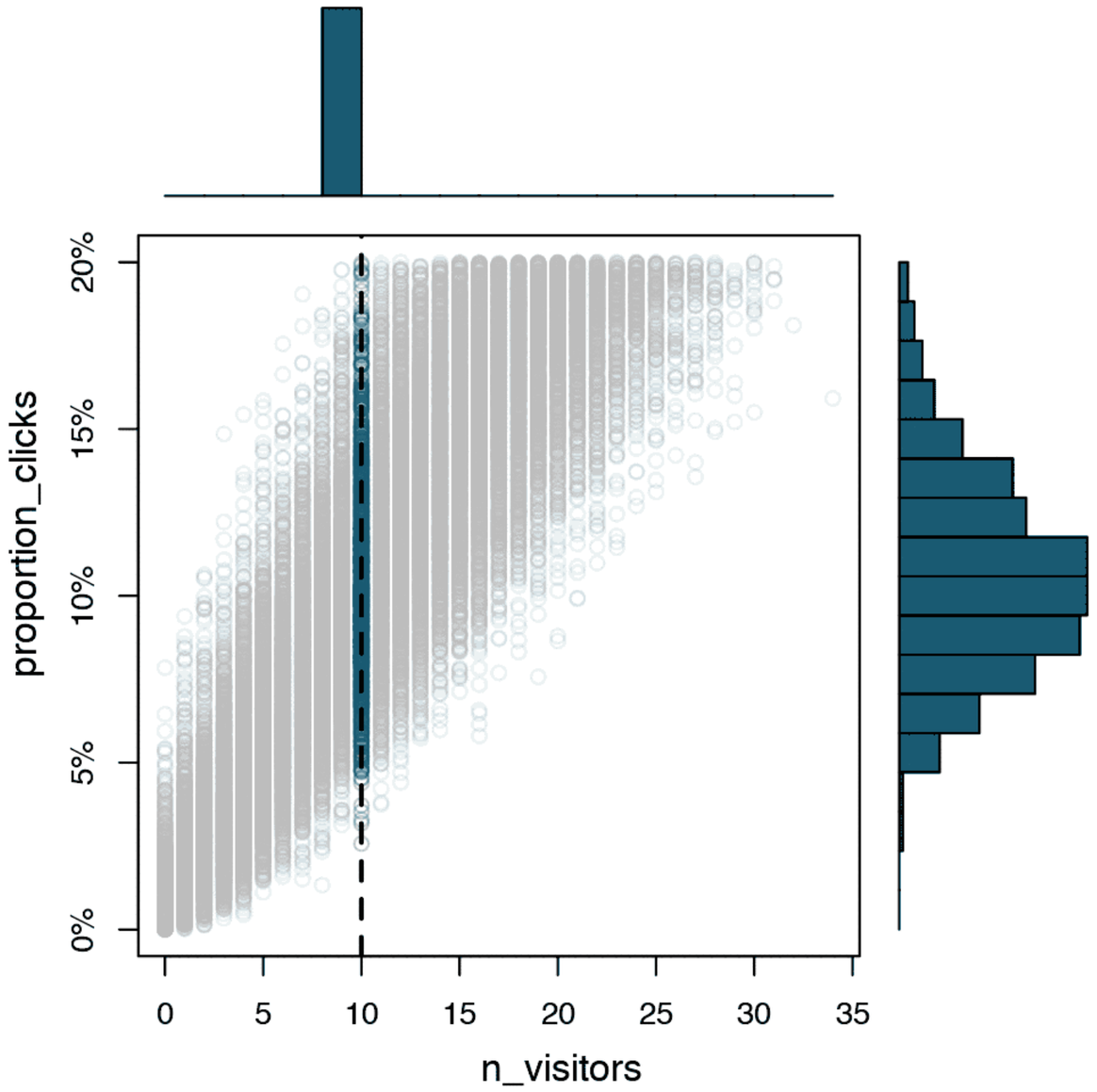

proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors

The essence of Bayesian inference

> Bayesian inference is conditioning on data, in order to learn about parameter values.

Try some Bayesian inference yourself!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

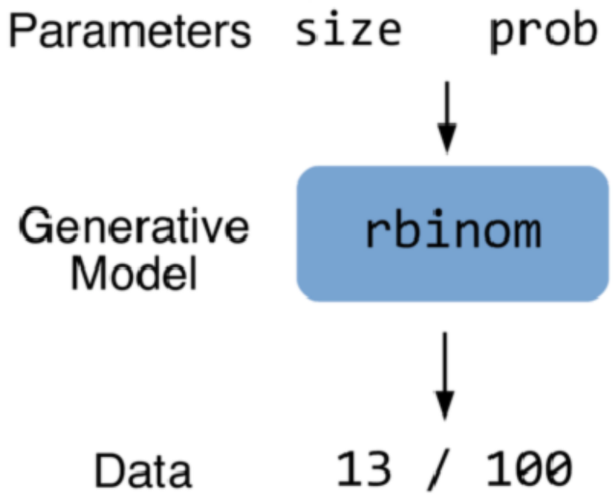
Bayesian inference, again!

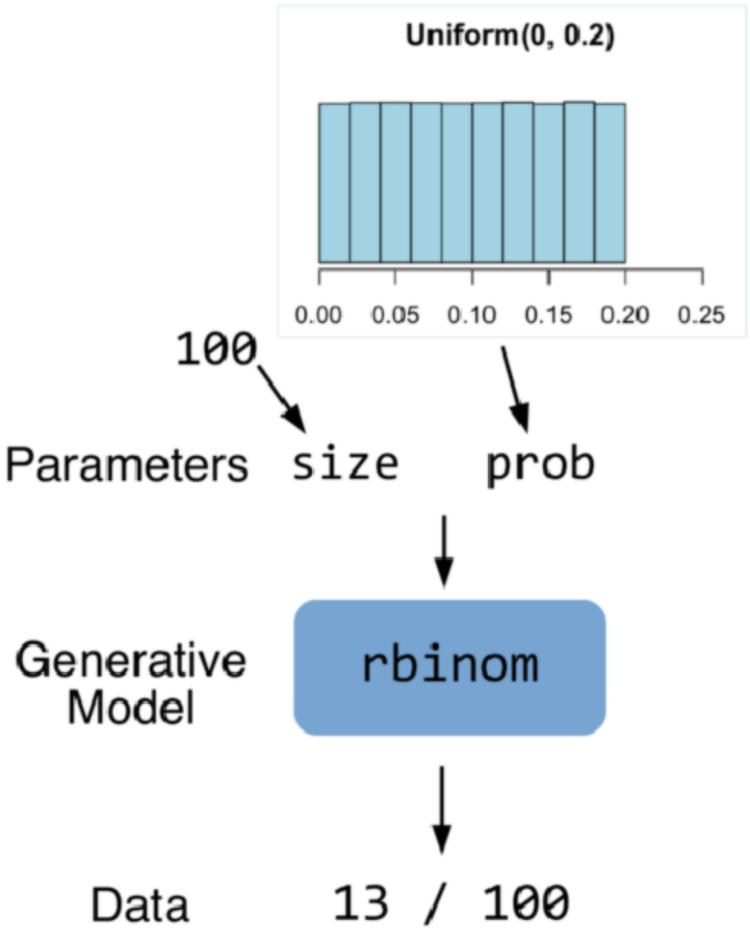
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

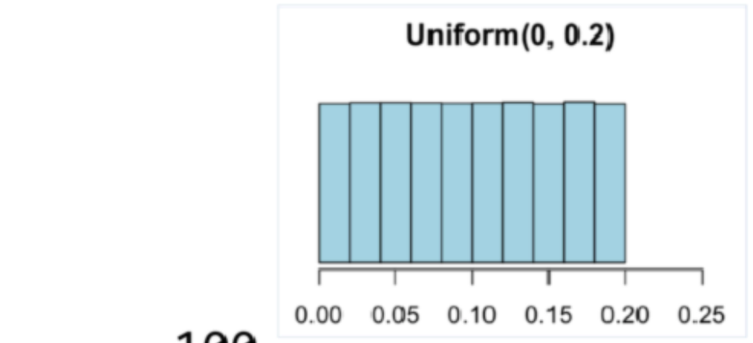
Rasmus Bååth
Data Scientist

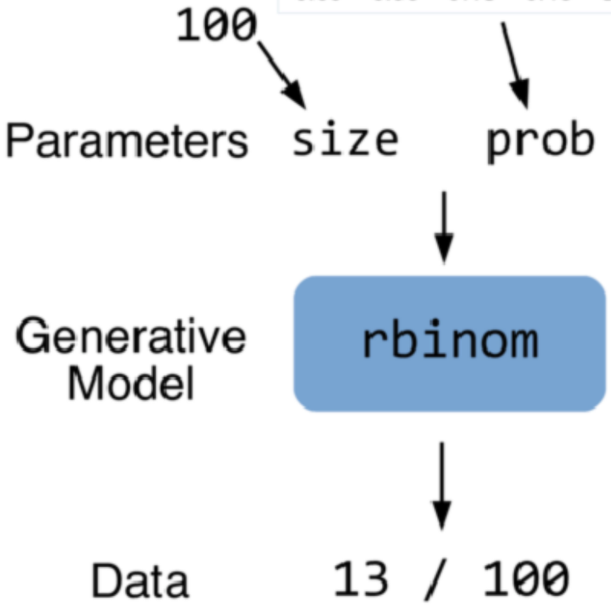
datacamp

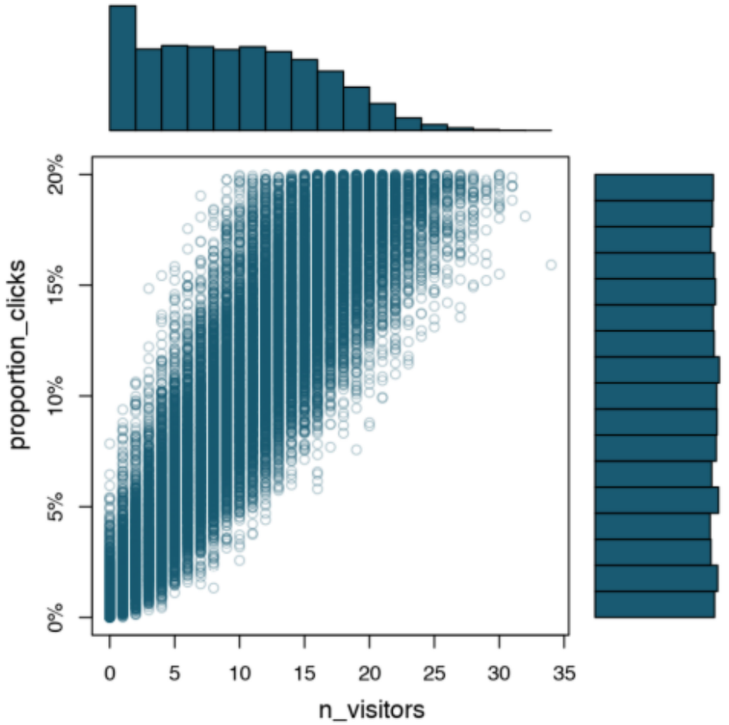
```
prop_success <- 0.1
size <- 100
# Simulating data
data <- c()
for(i in 1:size) {
  data[i] <-
    runif(1, min = 0, max = 1) <
      prop_success
}
data <- as.numeric(data)
```

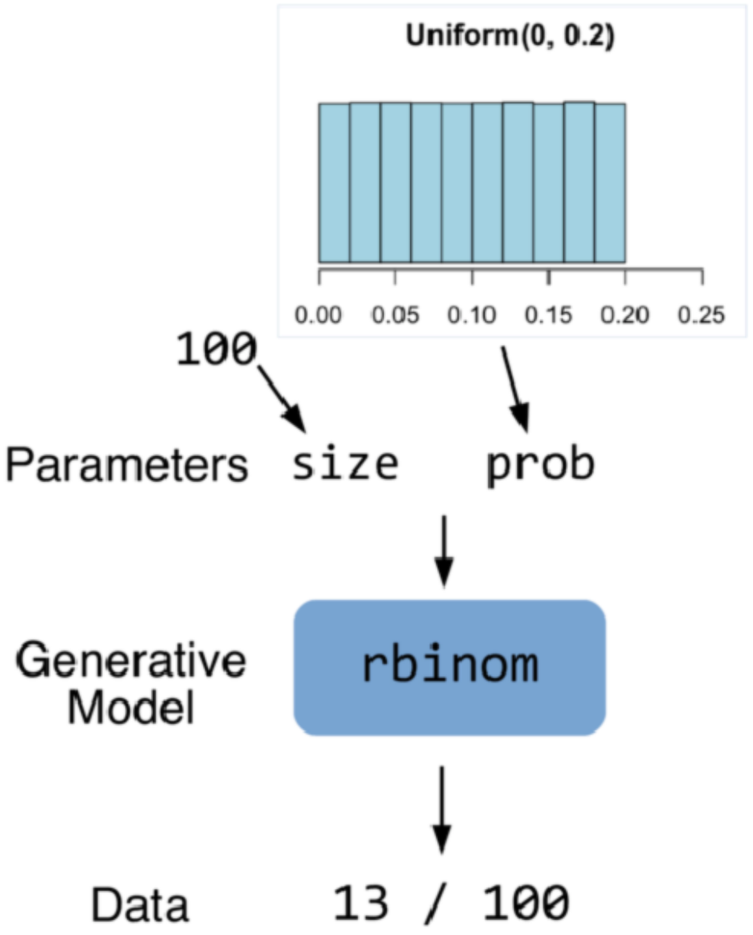

Parameters
size
prob
Generative Model
rbinom
Data
13 / 100

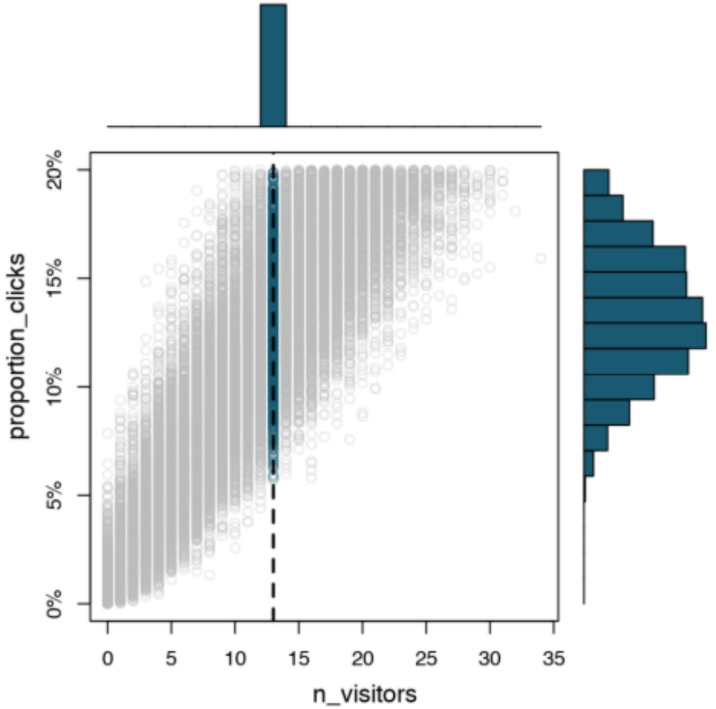

Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100

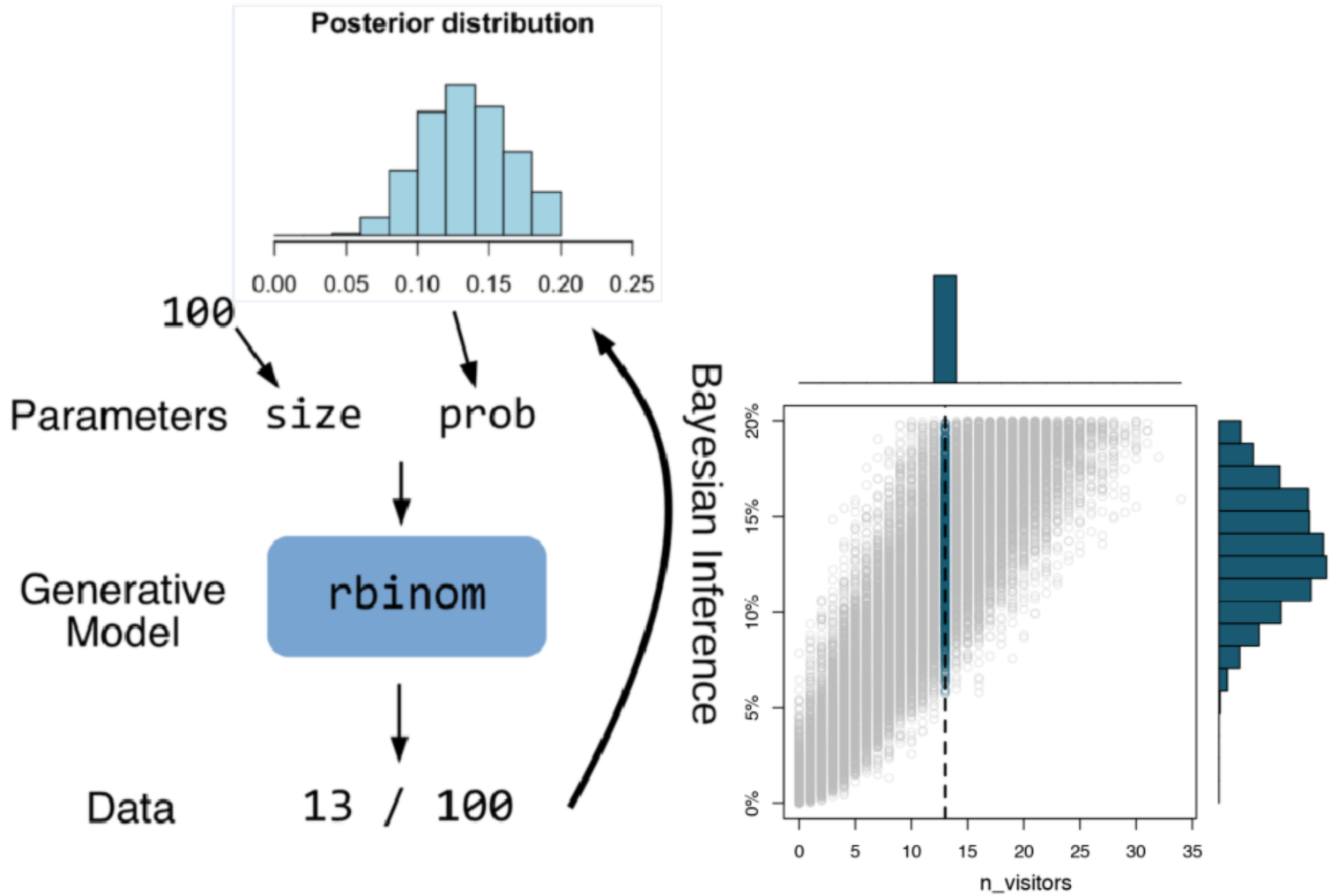

Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25

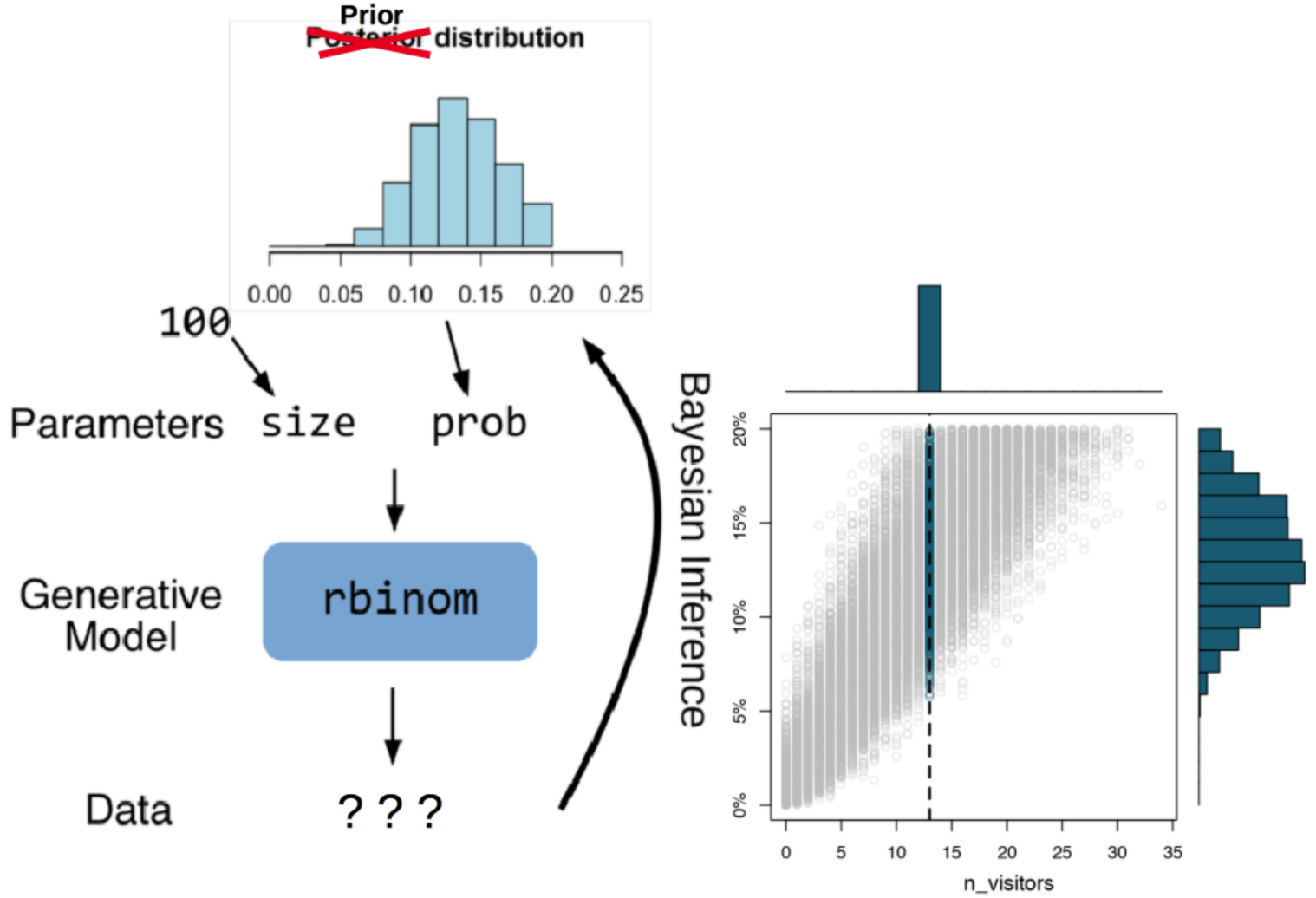

100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100

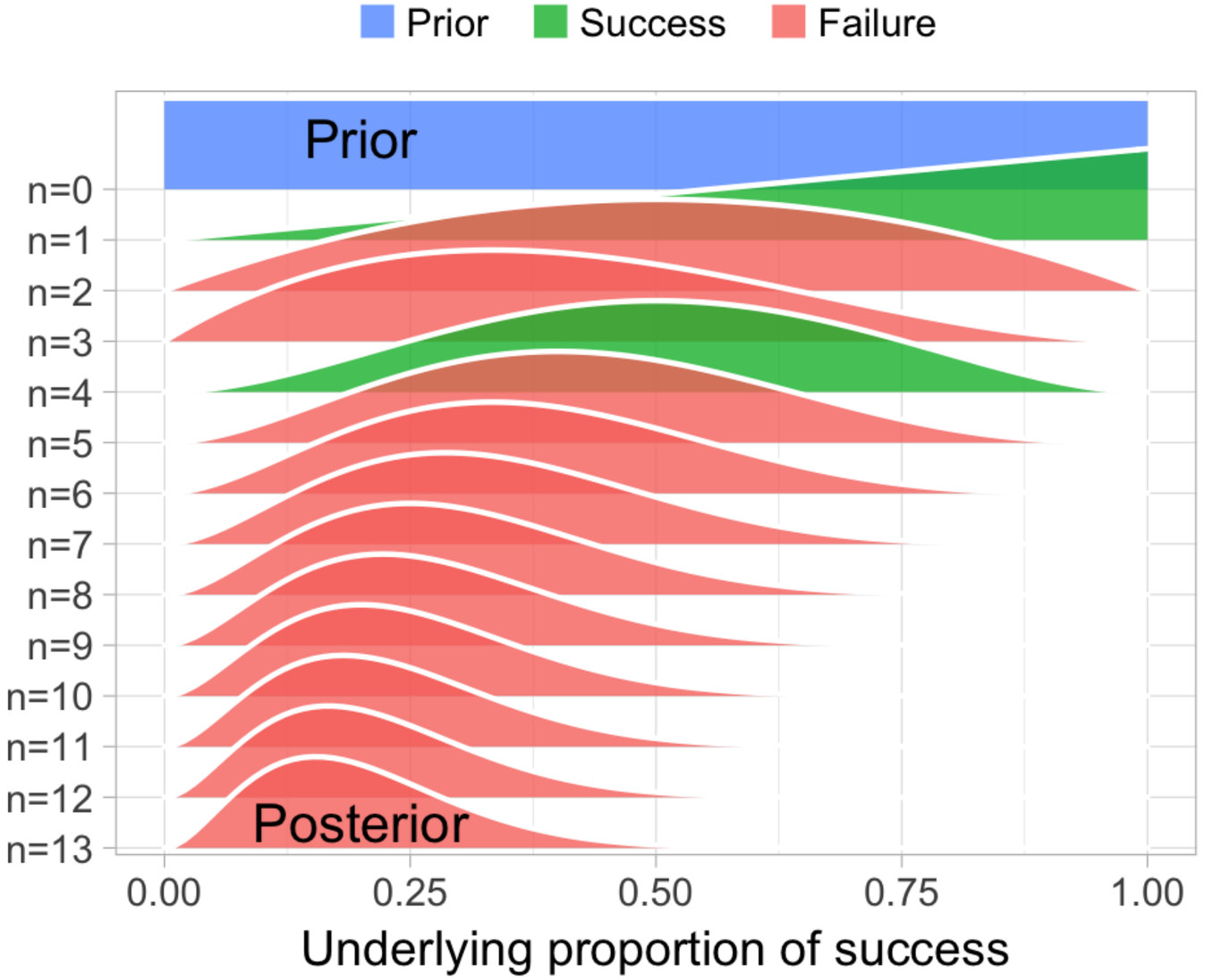

proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors

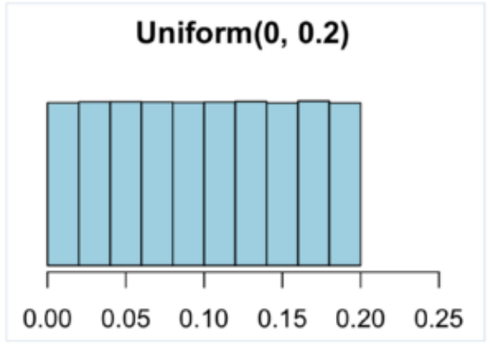

Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100


proportion_clicks
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors


Posterior distribution
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100
Bayesian Inference
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors

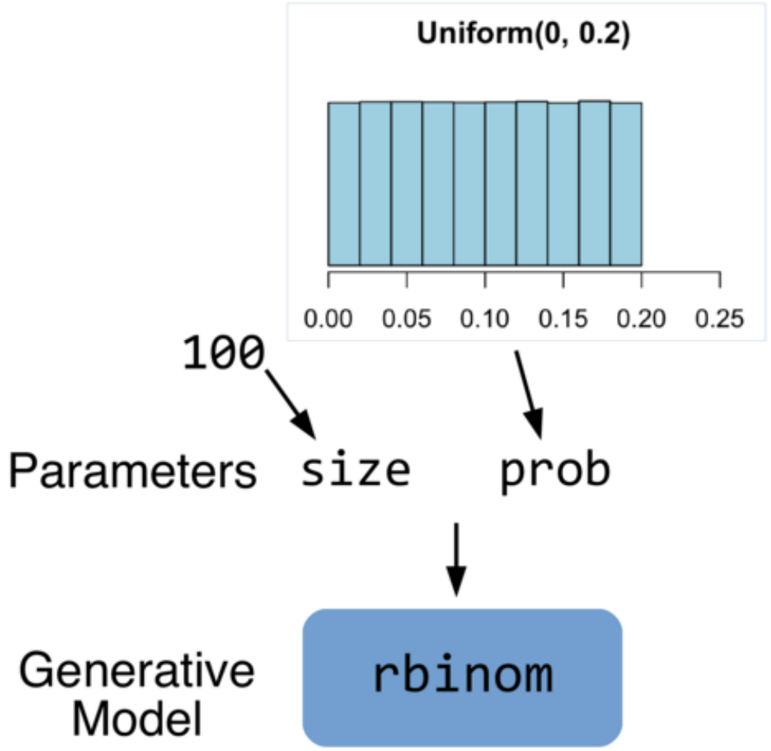

Prior
distribution
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
? ? ?
Bayesian Inference
0%
5%
10%
15%
20%
0
5
10
15
20
25
30
35
n_visitors

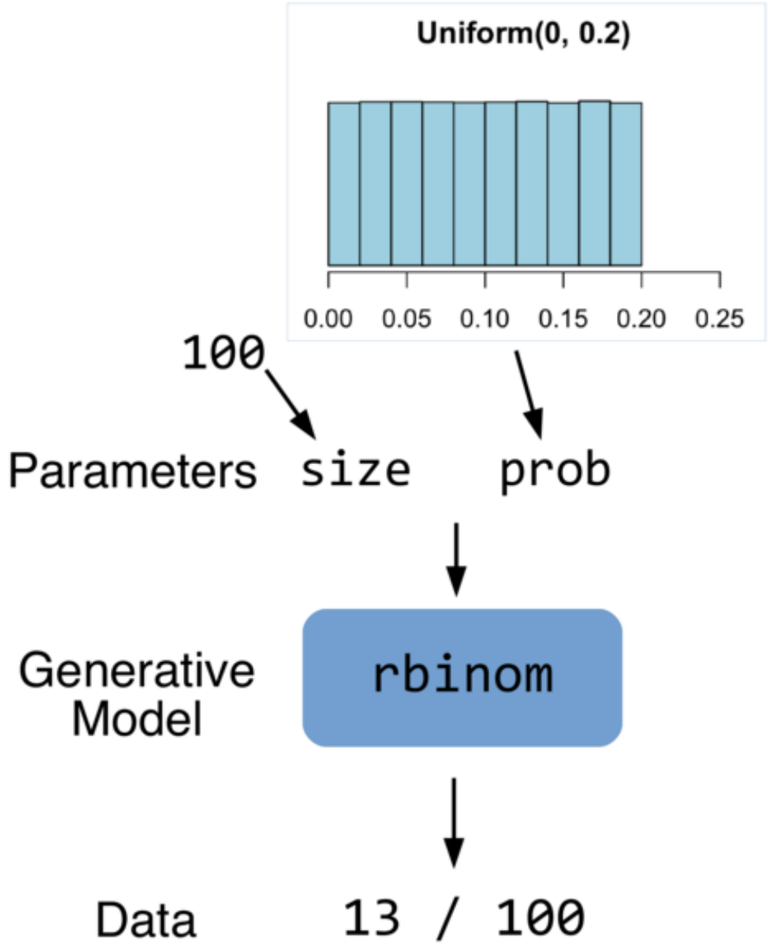

Prior
Success
Failure
Prior
n=0
n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
n=11
n=12
n=13
Posterior
0.00
0.25
0.50
0.75
1.00
Underlying proportion of success

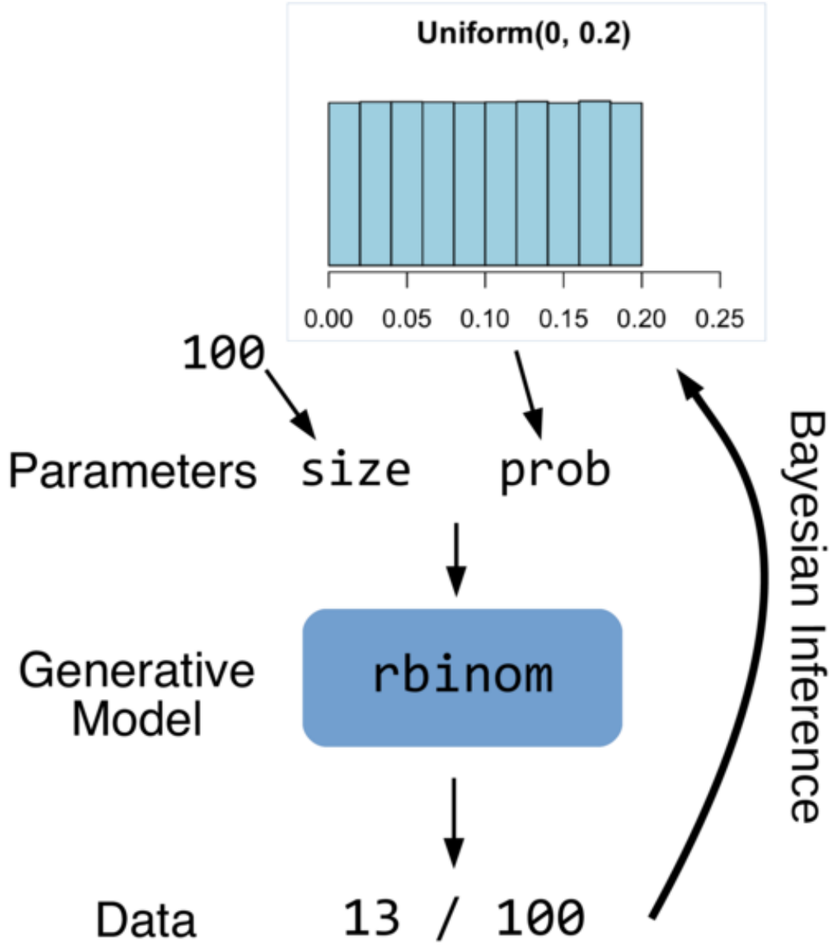


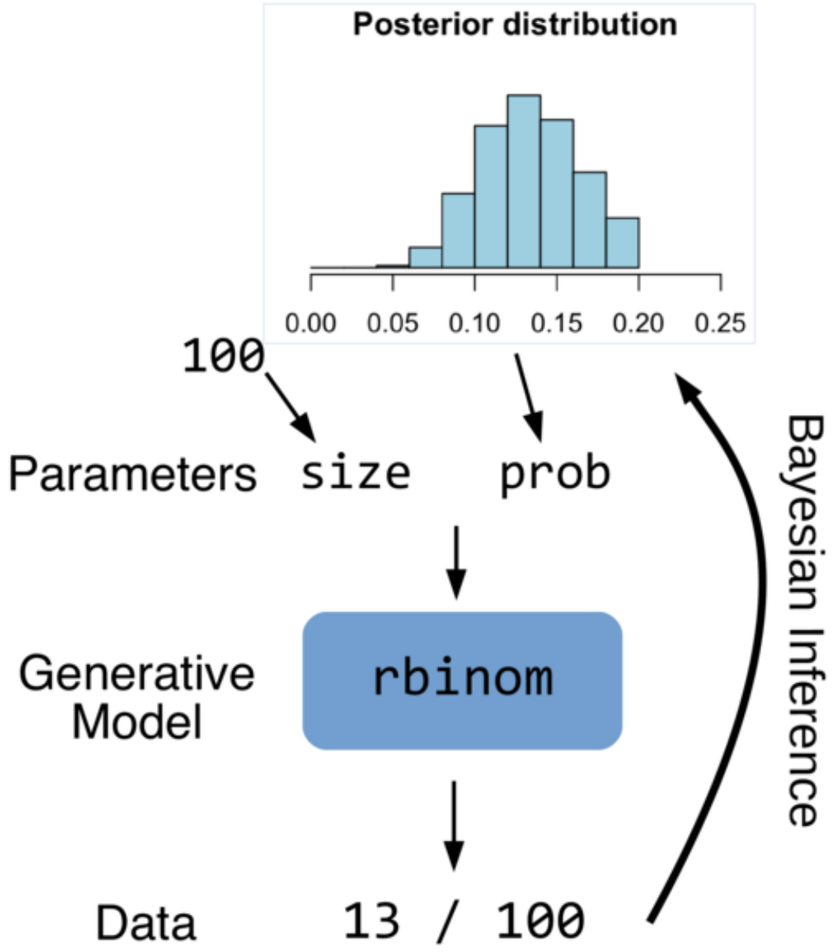

```
100
```

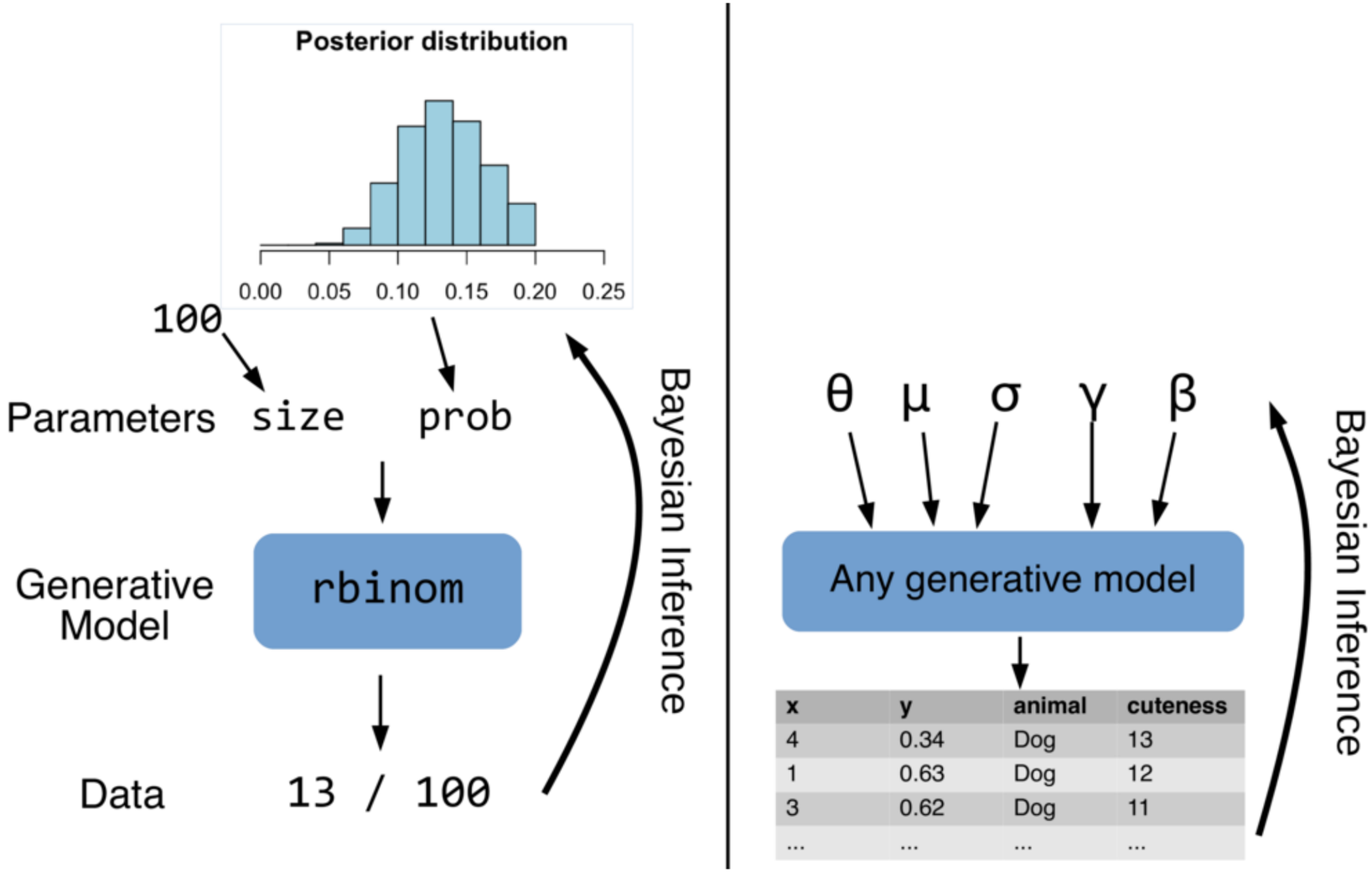
Parameters `size` `prob`

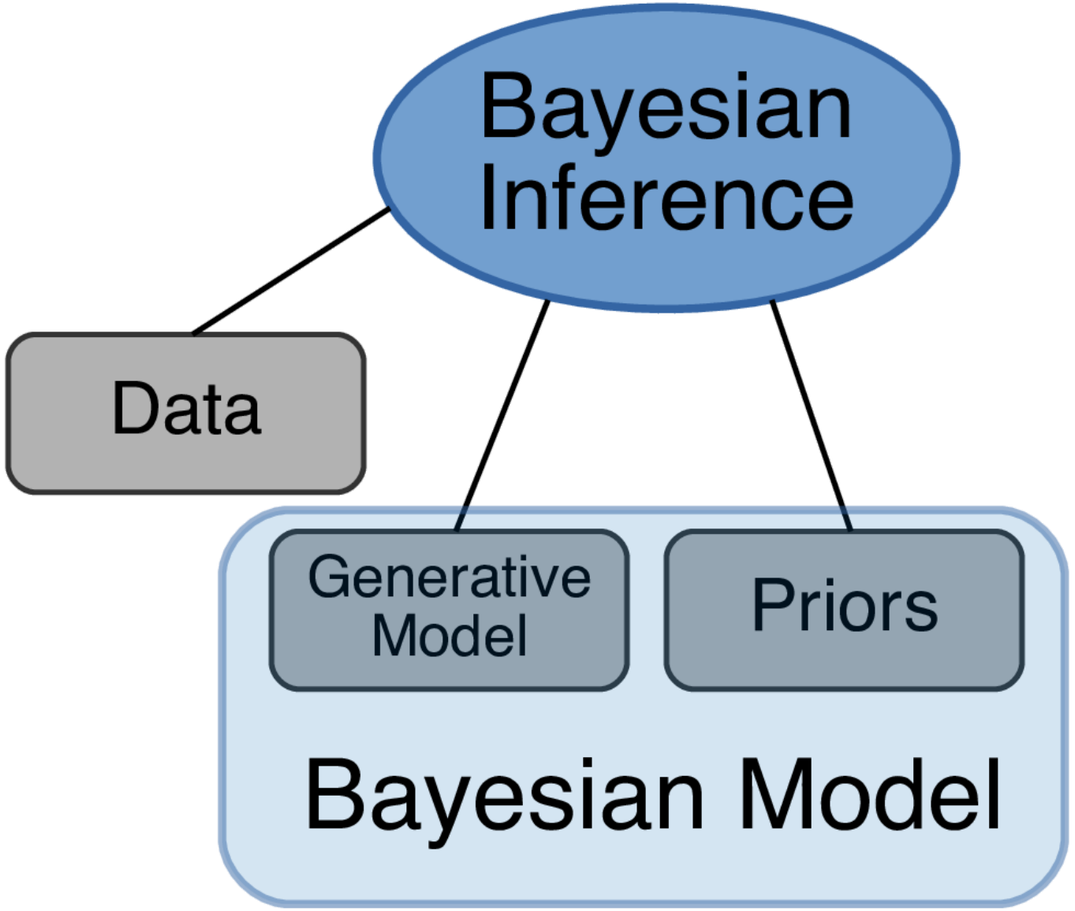

Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom

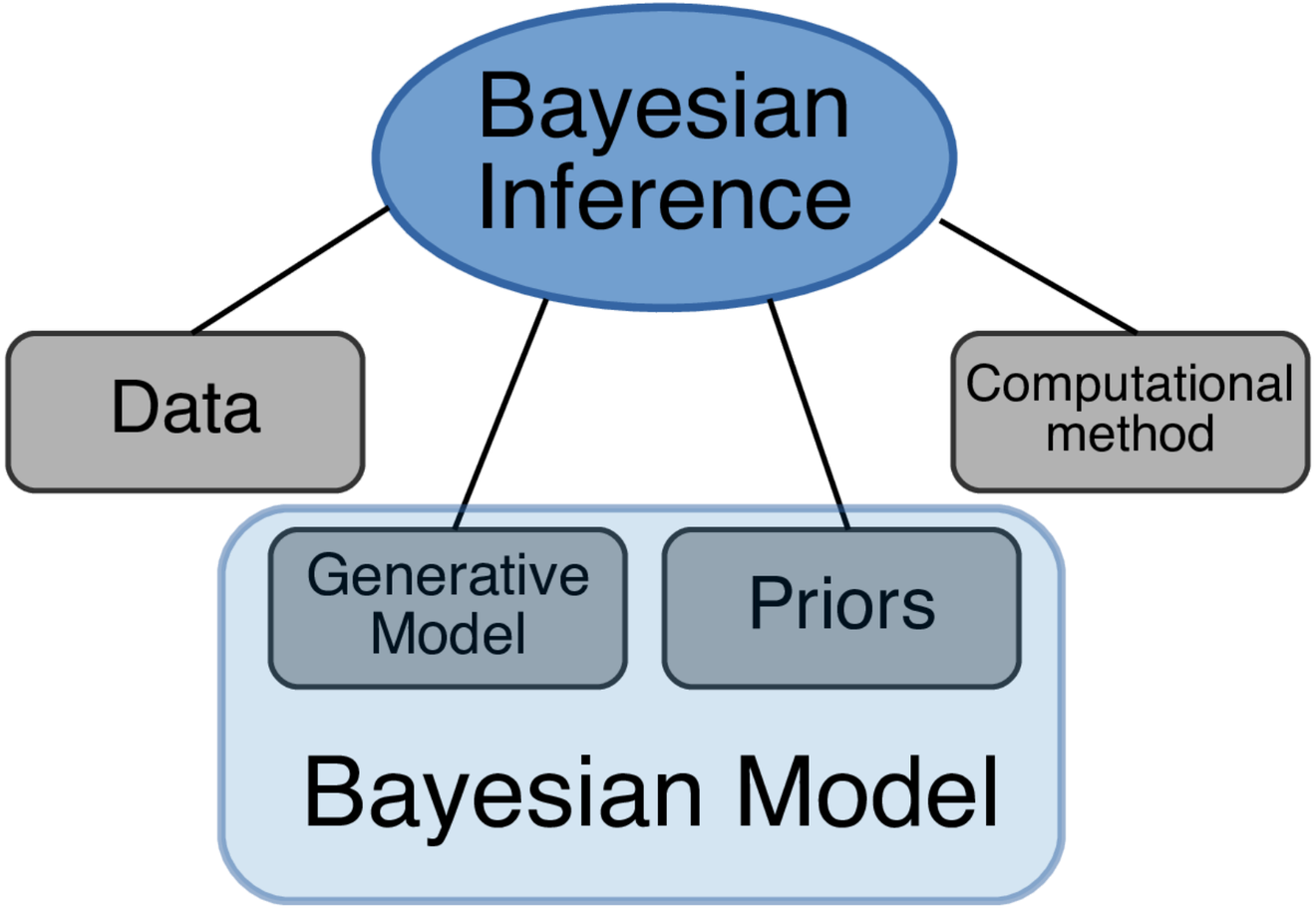

Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100


Uniform(0, 0.2)
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100
Bayesian Inference


Posterior distribution
0.00
0.05
0.10
0.15
0.20
0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100
Bayesian Inference


Posterior distribution
0.00 0.05 0.10 0.15 0.20 0.25
100
Parameters
size
prob
Generative Model
rbinom
Data
13 / 100
Bayesian Inference
θ μ σ γ β
Any generative model
x y animal cuteness
4 0.34 Dog 13
1 0.63 Dog 12
3 0.62 Dog 11
... ...
Bayesian Inference


Bayesian Inference
Data
Generative Model
Priors
Bayesian Model


Bayesian Inference
Data
Computational method
Generative Model
Priors
Bayesian Model

Next up: Why use Bayes?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R