

# Why use Bayesian data analysis?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R



**Rasmus Bååth**  
Data Scientist

# Bayes is flexible

1. You can include information sources in addition to the data.
2. You can make any comparisons between groups or data sets.
3. You can use the result of a Bayesian analysis to do Decision Analysis.
4. You can change the underlying statistical model.

# Including information in addition to data

- Background information
- Expert opinion
- Common knowledge



**You**

So what are really the range of proportion of clicks you see for ads?



**You**

So what are really the range of proportion of clicks you see for ads?



**Social media company person**

Hi You! Most ads gets clicked on 5% of the time, but for some ads it is as low as 2% and for others as high as 8%.



**You**

So what are really the range of proportion of clicks you see for ads?



**Social media company person**

Hi You! Most ads gets clicked on 5% of the time, but for some ads it is as low as 2% and for others as high as 8%.



**You**

Ah, but you've written 10% on your webpage!? 🤔



**You**

So what are really the range of proportion of clicks you see for ads?



**Social media company person**

Hi You! Most ads gets clicked on 5% of the time, but for some ads it is as low as 2% and for others as high as 8%.



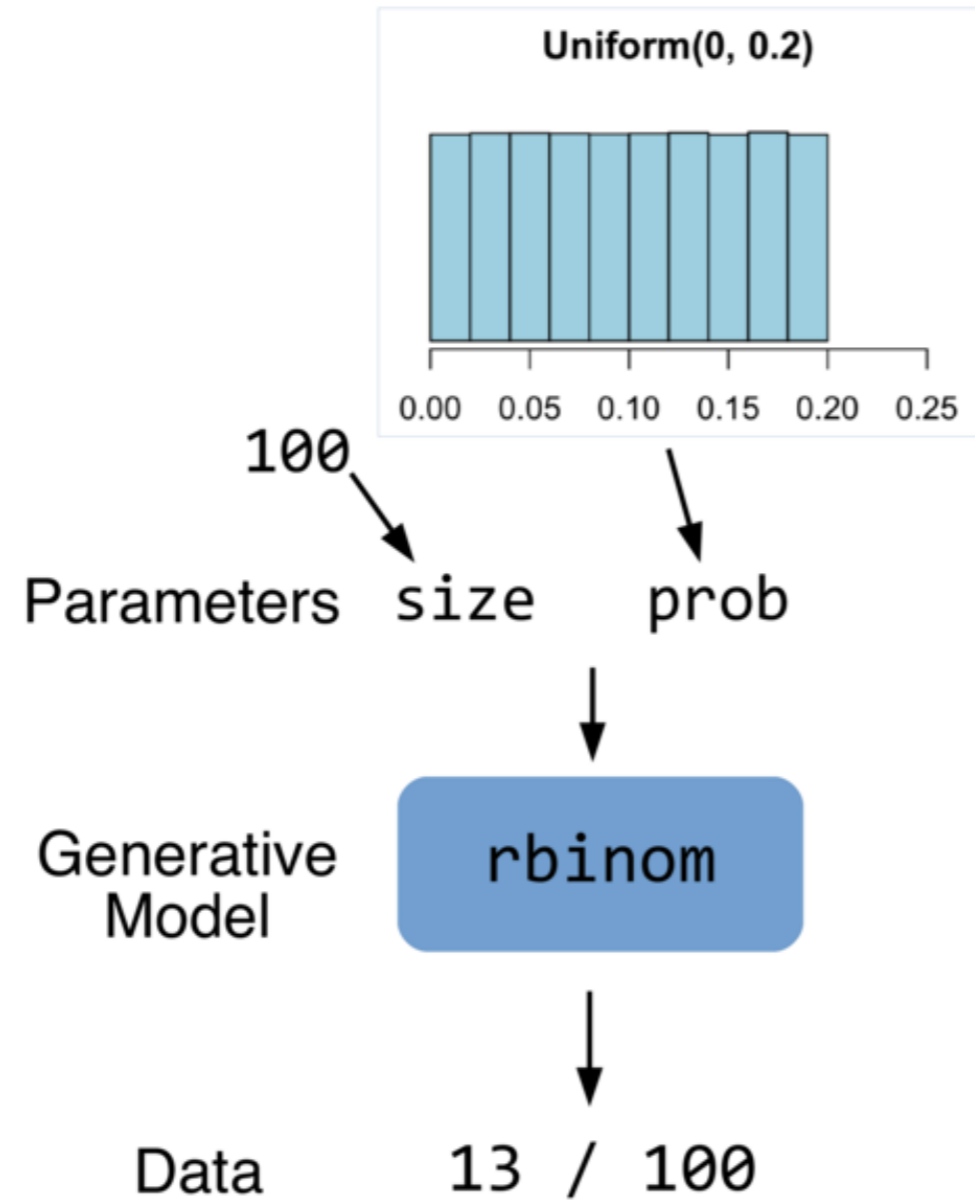
**You**

Ah, but you've written 10% on your webpage!? 🤔



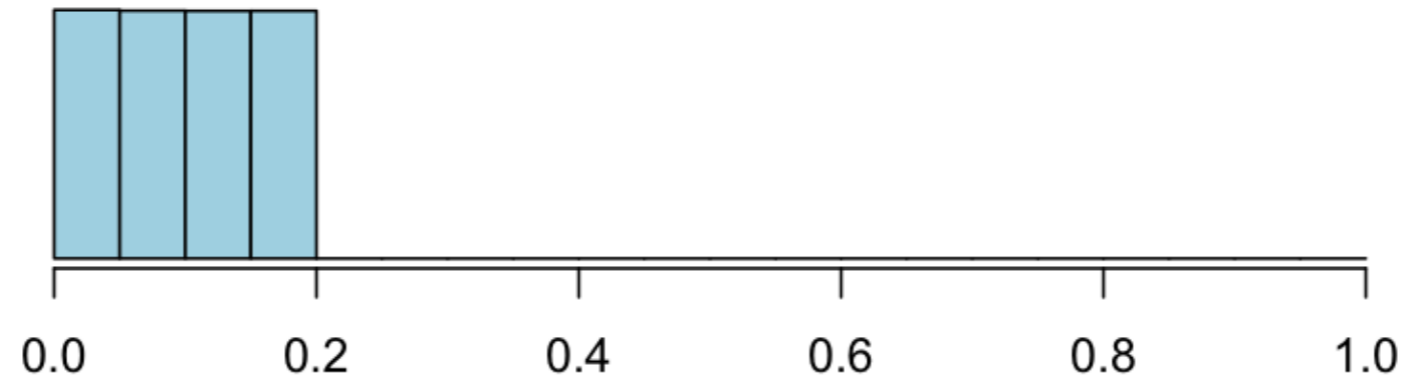
**Social media company person**

That's marketing, don't listen to them! 😜

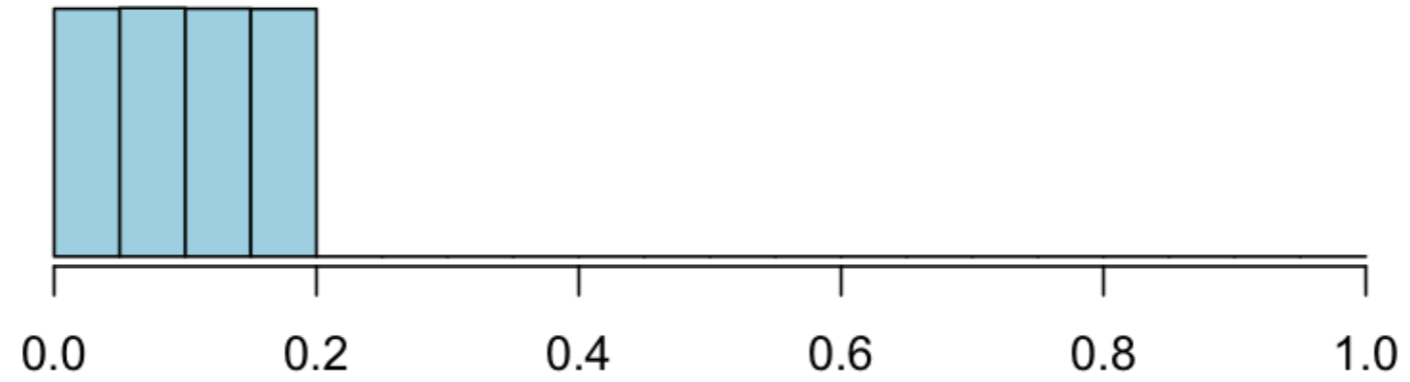




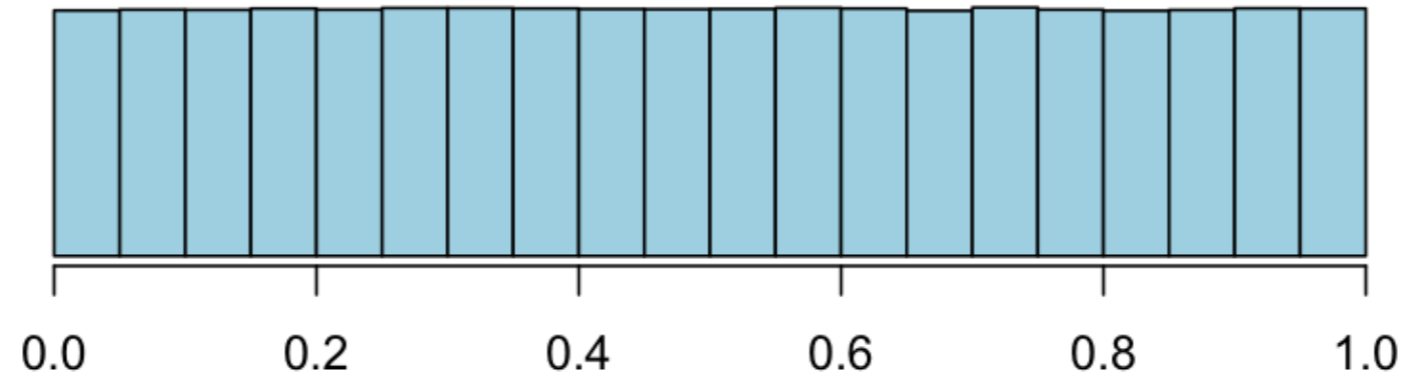
**Uniform(0, 0.2) prior**



**Uniform(0, 0.2) prior**



**Uniform(0, 1.0) 'Uninformative' prior**





**You**

So what are really the range of proportion of clicks you see for ads?



**Social media company person**

Hi You! Most ads gets clicked on 5% of the time, but for some ads it is as low as 2% and for others as high as 8%.



**You**

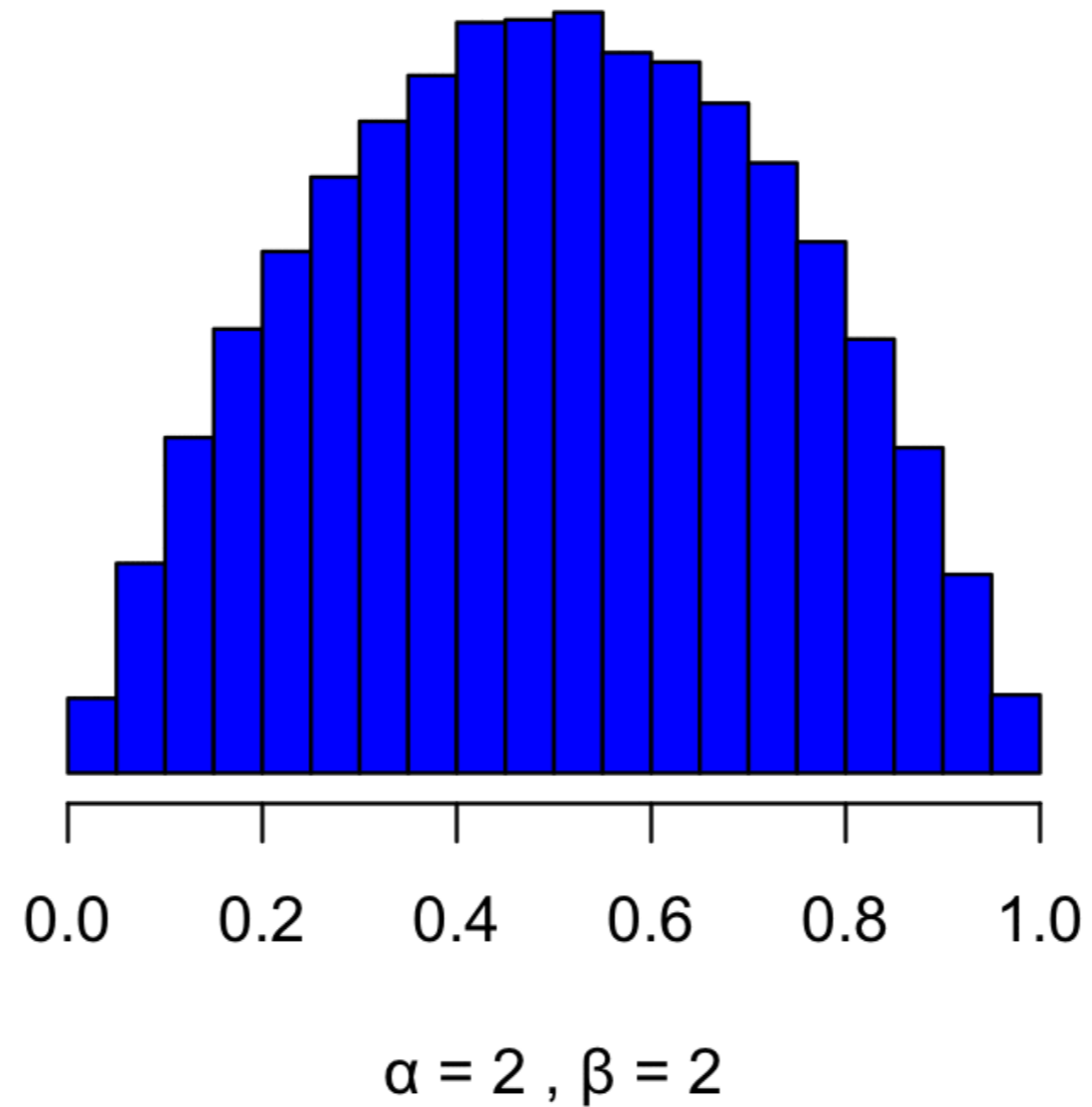
Ah, but you've written 10% on your webpage! 🤔



**Social media company person**

That's marketing, don't listen to them! 🤔

## Some shapes of the beta distribution



# Define an informed prior!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

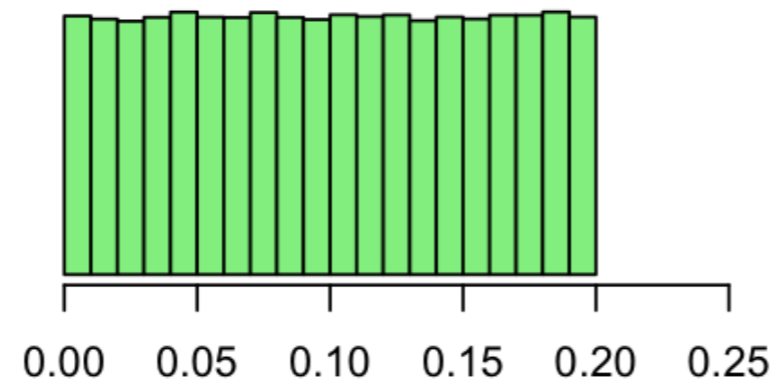
# You've changed the prior!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

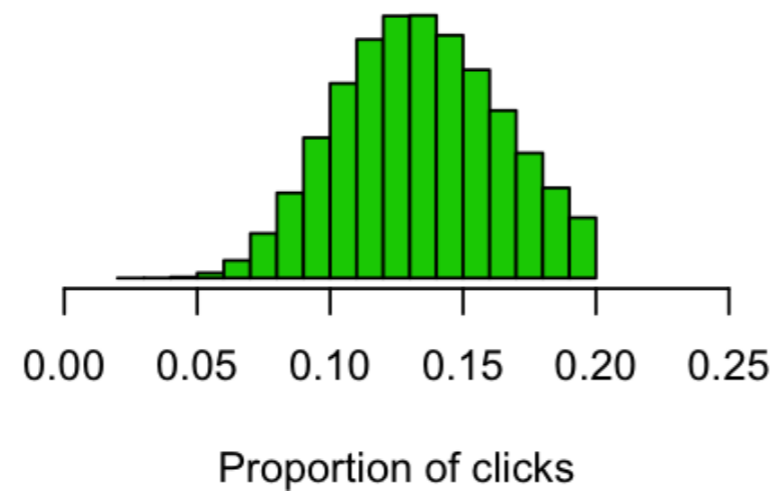


**Rasmus Bååth**  
Data Scientist

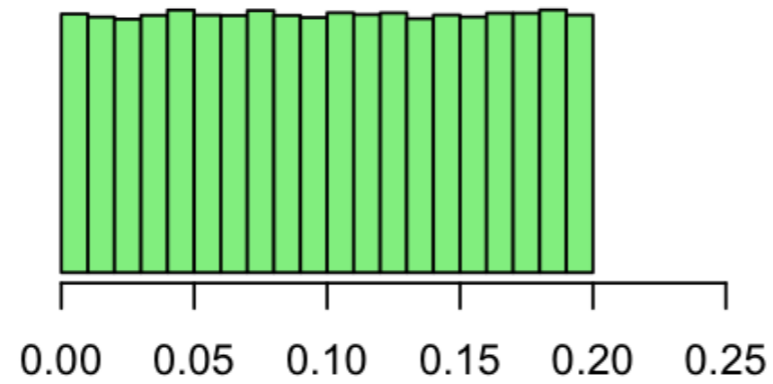
**Old prior**



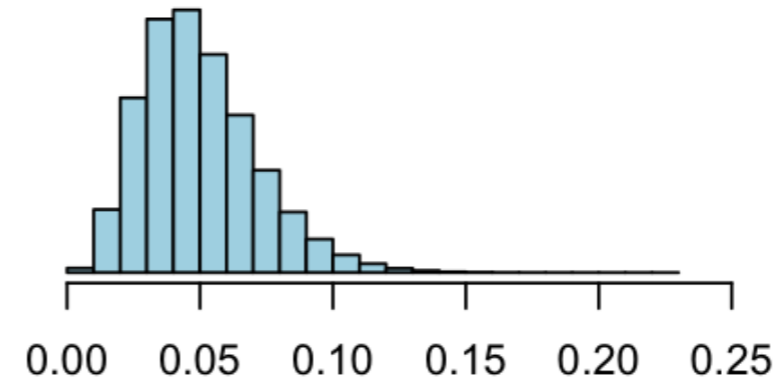
**Old posterior**



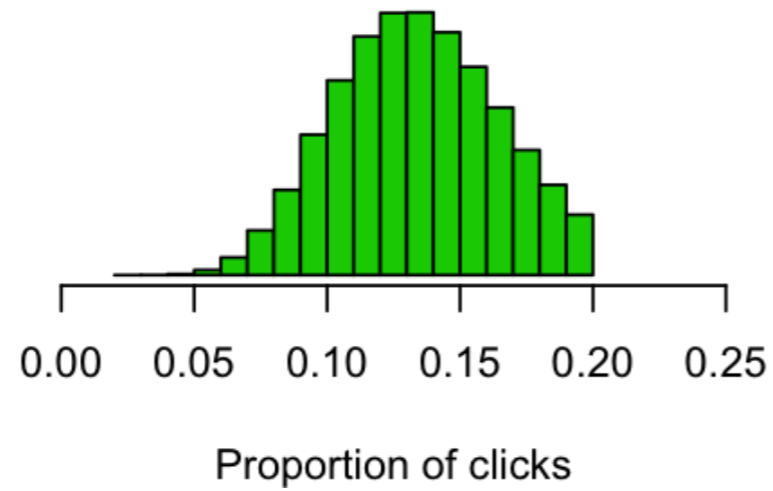
**Old prior**



**Informed prior**

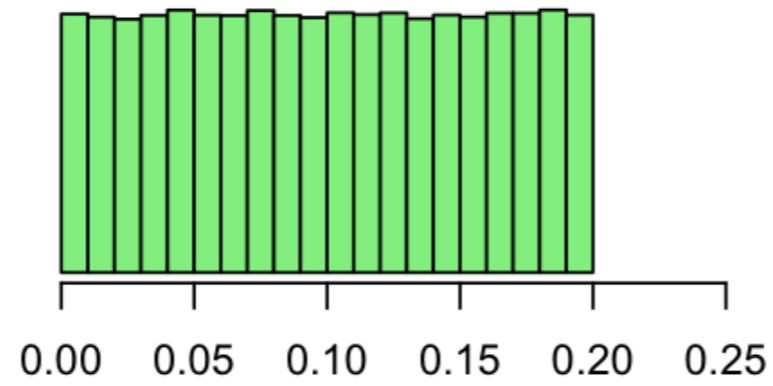


**Old posterior**



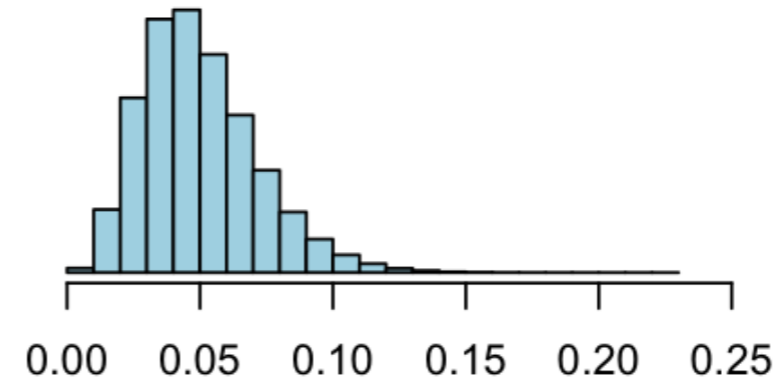


**Old prior**



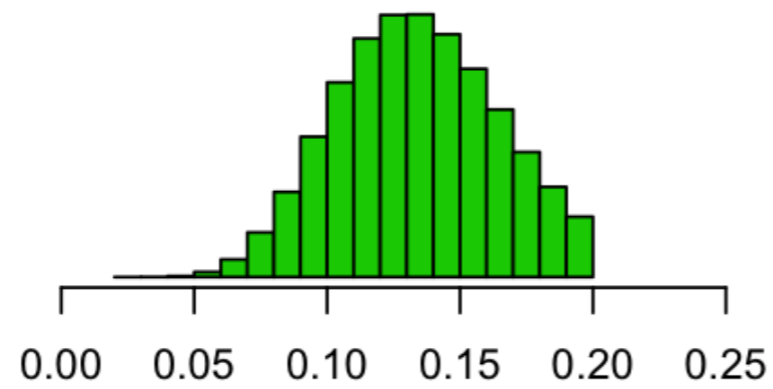
Proportion of clicks

**Informed prior**



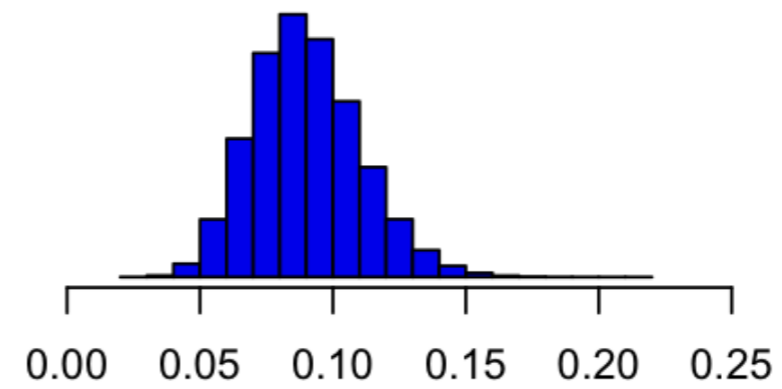
Proportion of clicks

**Old posterior**



Proportion of clicks

**Informed posterior**



Proportion of clicks

# Next up on reasons to use Bayesian data analysis

1. You can include information sources in addition to the data.
2. **You can make any comparisons between groups or datasets.**
3. You can use the result of a Bayesian analysis to do Decision Analysis.
4. You can change the underlying statistical model.

# Video vs Text



# Video vs Text



[datacamp.com](https://datacamp.com) | Learn data science intuitively

Ad [www.datacamp.com/](https://www.datacamp.com/)

Complete short online exercises and watch brief videos.

Tailored For Your Needs · Free And Premium Courses · Learn A  
Courses: Intro to R, Python for Data Science, Intro to SQL, Git for

# Video vs Text



[datacamp.com](https://datacamp.com) | Learn data science intuitively

Ad [www.datacamp.com/](https://www.datacamp.com/)

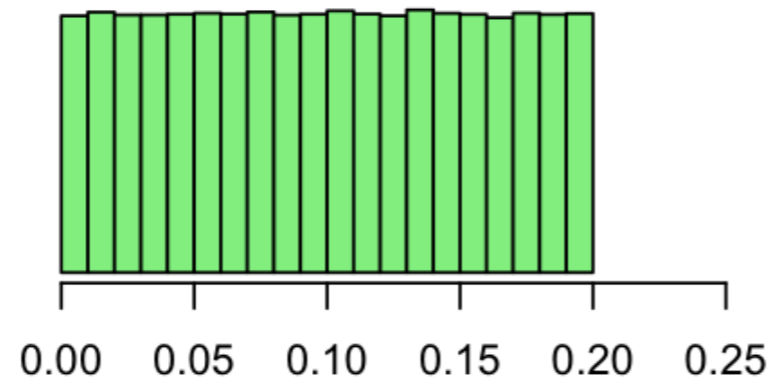
Complete short online exercises and watch brief videos.

Tailored For Your Needs · Free And Premium Courses · Learn A  
Courses: Intro to R, Python for Data Science, Intro to SQL, Git f

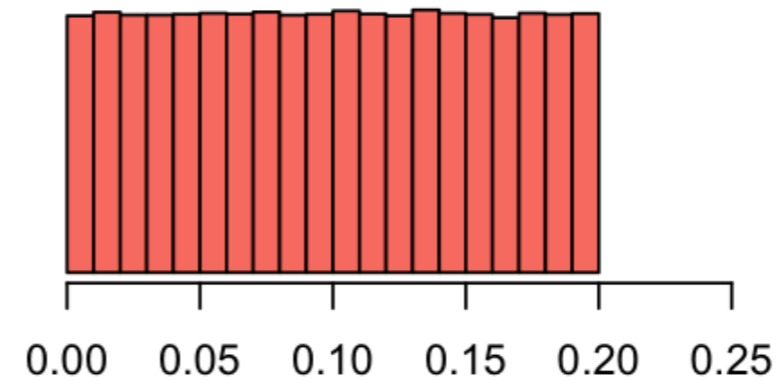
13 / 100

6 / 100

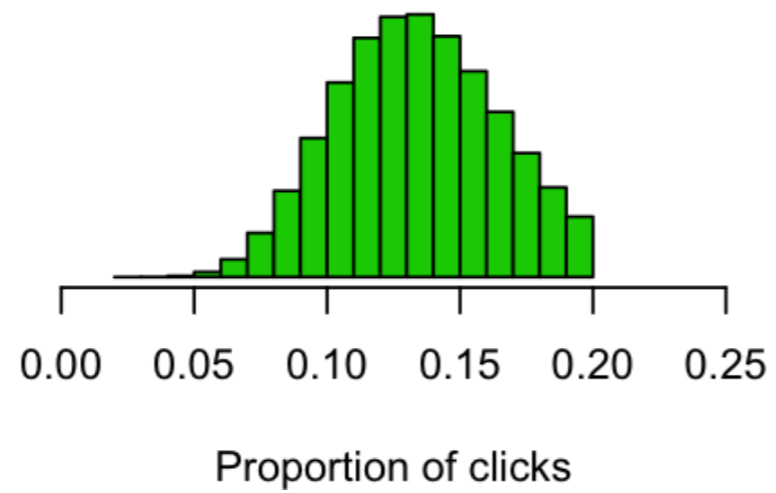
**Video prior**



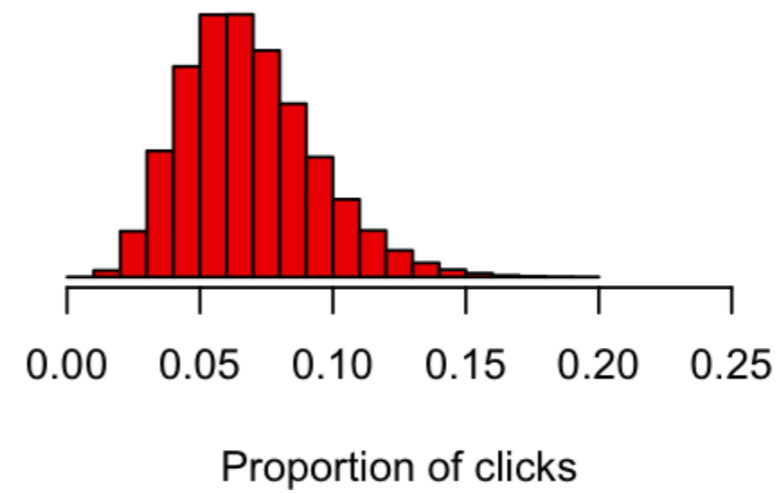
**Text prior**



**Video posterior (13 / 100)**



**Text posterior (6 / 100)**



# Comparing Video and Text ads

```
posterior
```

```
  video_prop text_prop
1      0.08      0.10
2      0.10      0.07
3      0.16      0.05
4      0.09      0.05
5      0.18      0.03
6      0.13      0.05
7      0.12      0.10
8      0.10      0.04
9      0.11      0.09
10     0.18      0.05
11     0.12      0.04
12     0.13      0.07
13     0.10      0.13
14     0.15      0.03
15     0.07      0.05
16     0.14      0.09
...     ...     ...
```

# Comparing Video and Text ads

```
posterior$prop_diff <- posterior$video_prop - posterior$text_prop  
posterior
```

```
  video_prop text_prop prop_diff  
1      0.08      0.10     -0.02  
2      0.10      0.07      0.03  
3      0.16      0.05      0.11  
4      0.09      0.05      0.04  
5      0.18      0.03      0.15  
6      0.13      0.05      0.08  
7      0.12      0.10      0.02  
8      0.10      0.04      0.06  
9      0.11      0.09      0.02  
10     0.18      0.05      0.13  
11     0.12      0.04      0.08  
12     0.13      0.07      0.06  
13     0.10      0.13     -0.03  
14     0.15      0.03      0.11  
15     0.07      0.05      0.01  
...     ...     ...     ...
```



# How does the `prop_diff()` distribution look?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

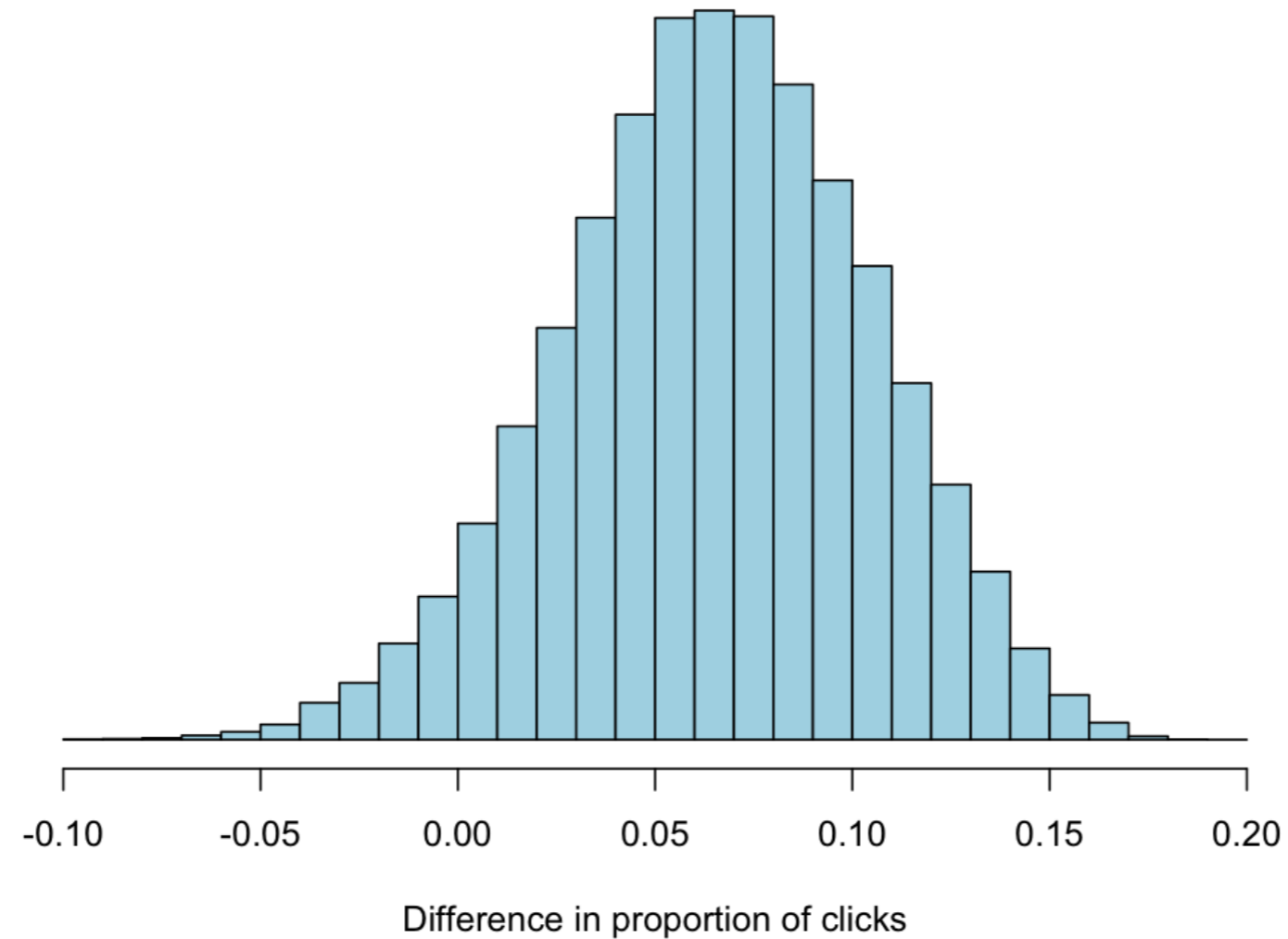
# It's easy to compare and contrast!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

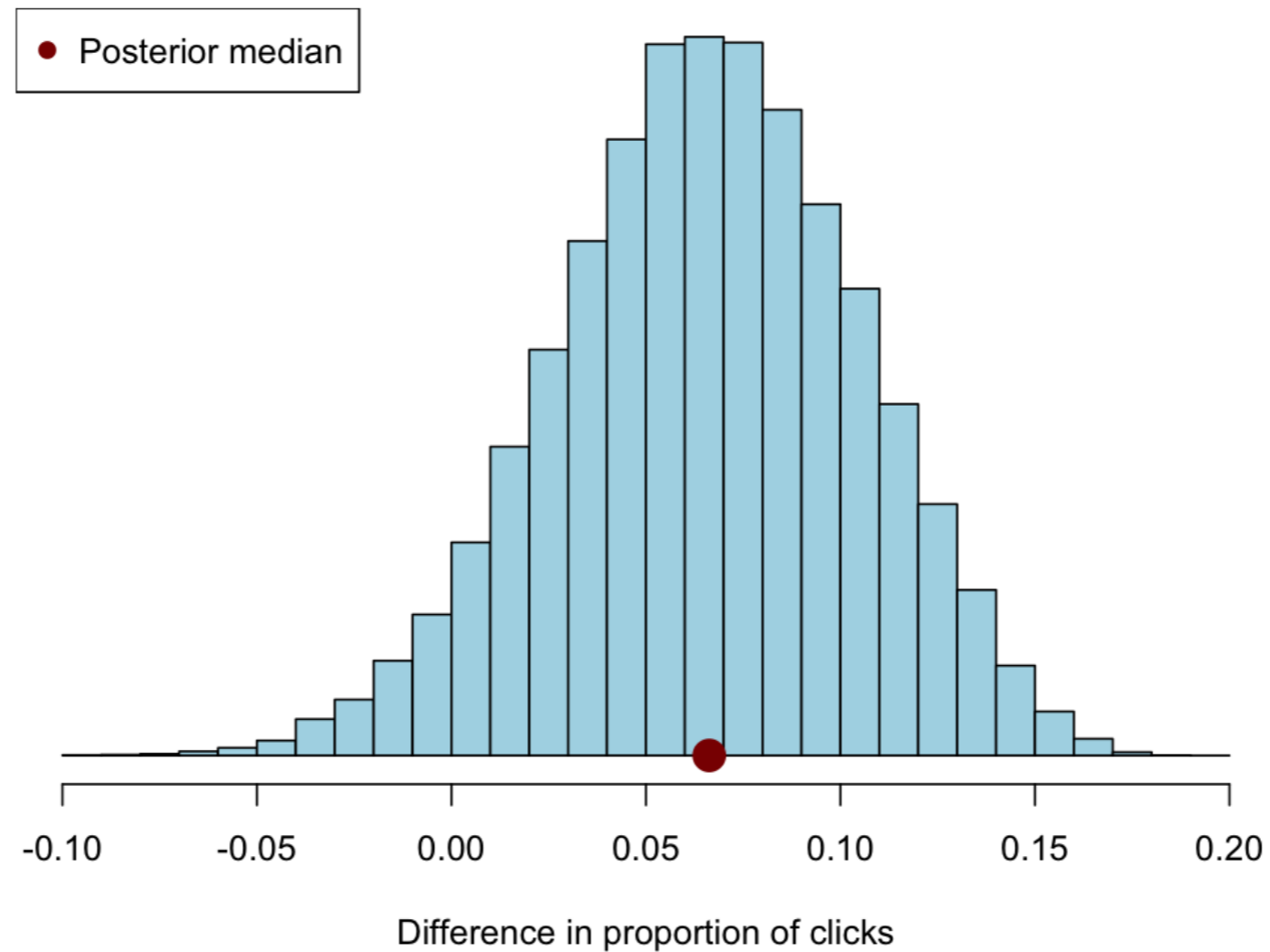


**Rasmus Bååth**  
Data Scientist

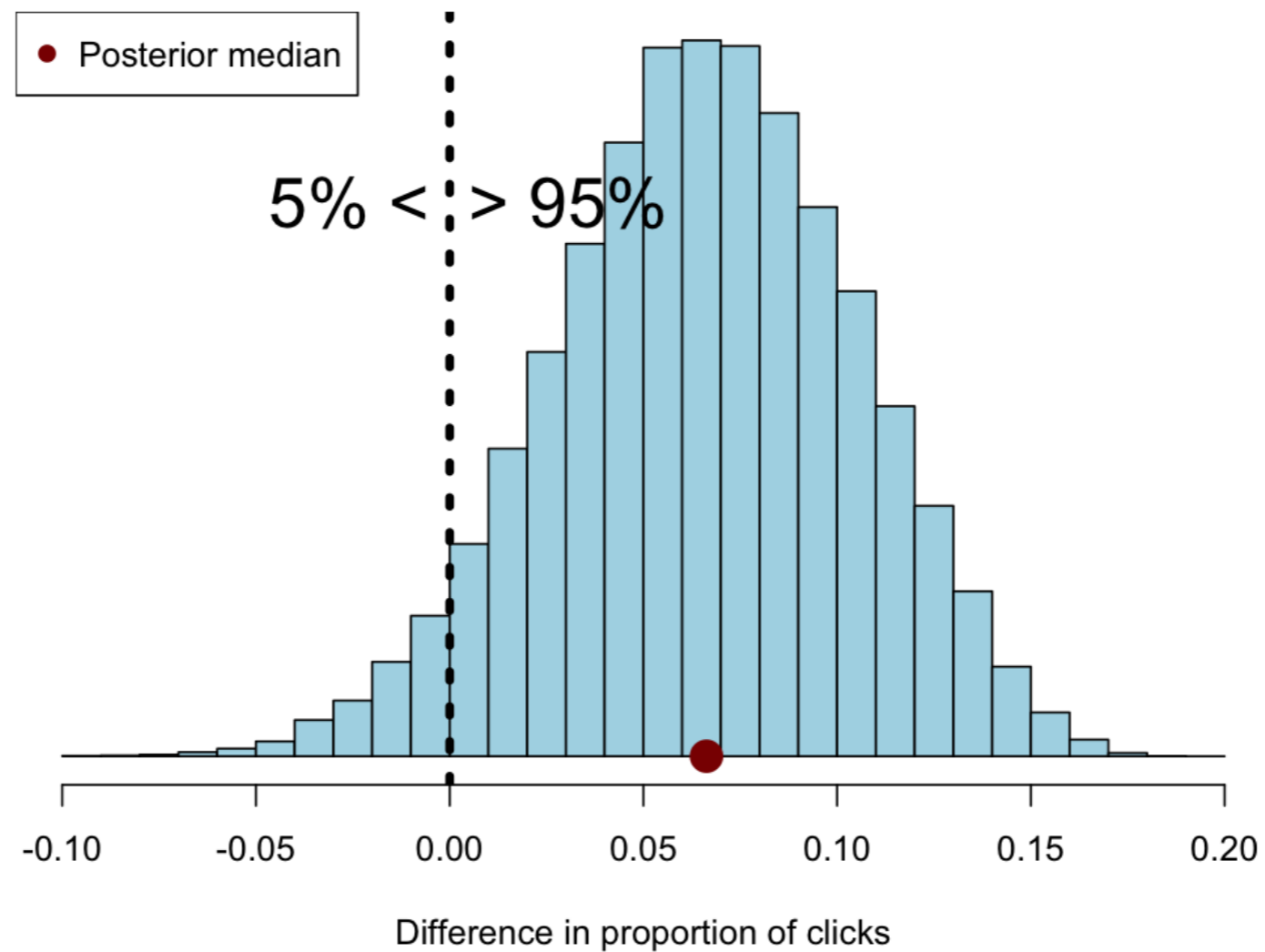
# A probability distribution over an interesting parameter



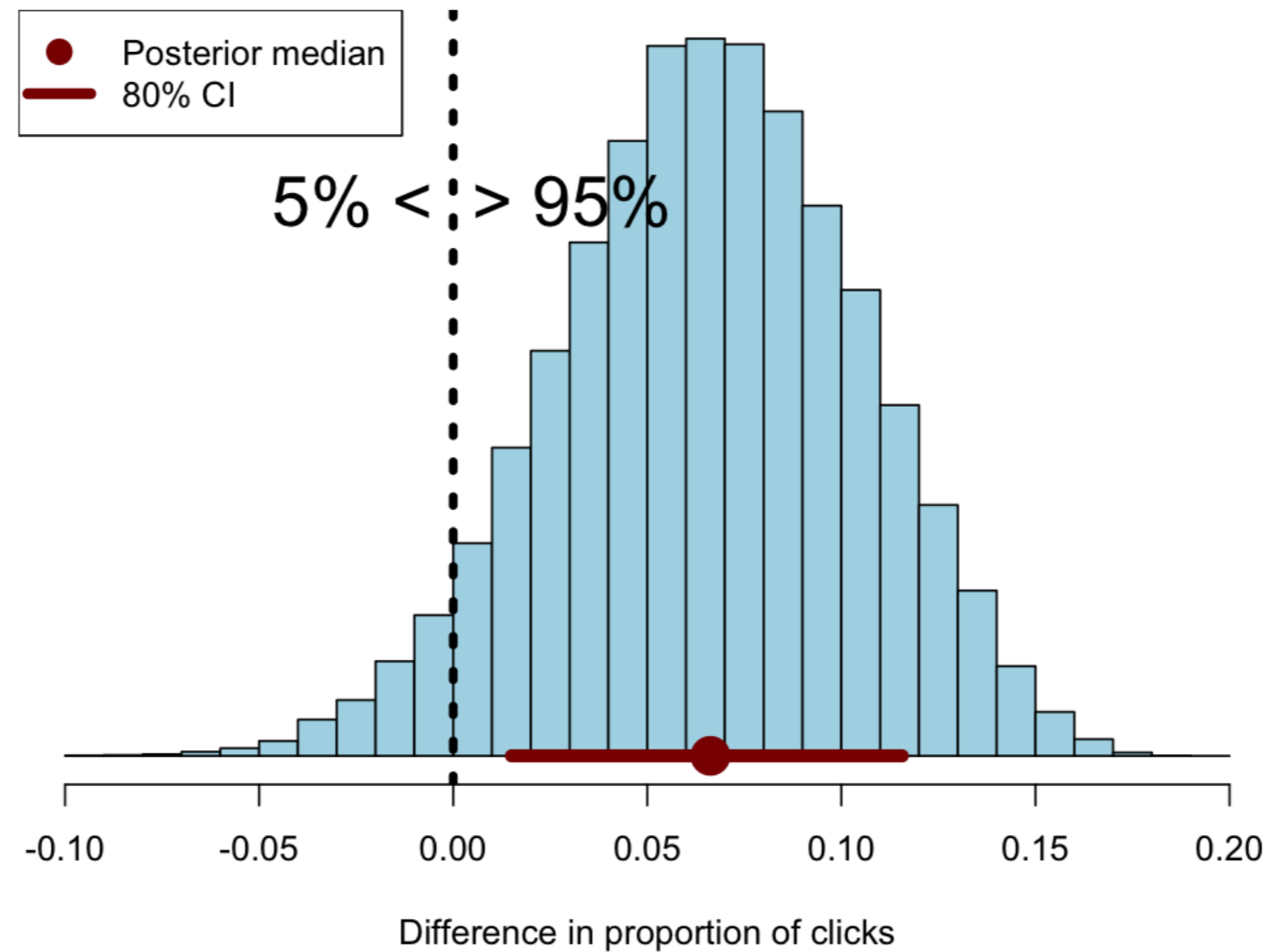
# A probability distribution over an interesting parameter



# A probability distribution over an interesting parameter



# A probability distribution over an interesting parameter



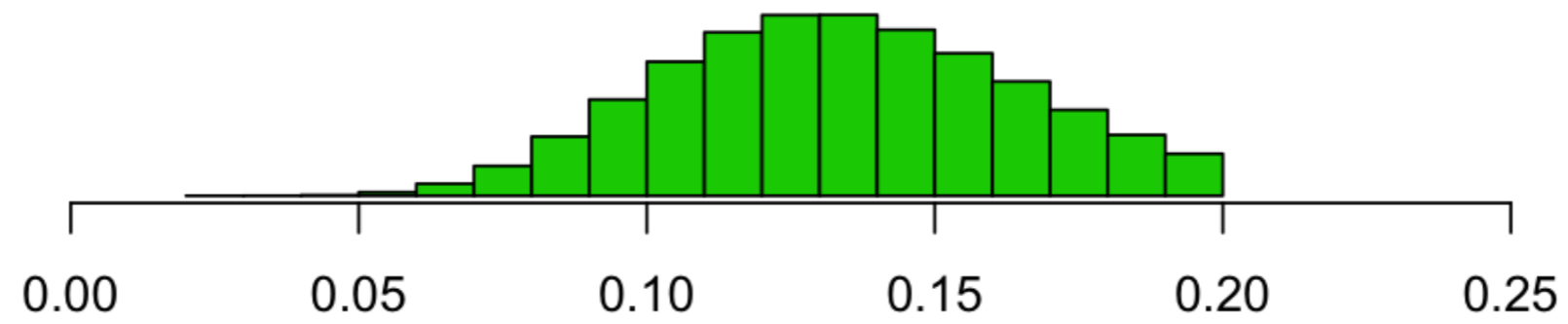
# Next up on reasons to use Bayesian data analysis

1. You can include information sources in addition to the data.
2. You can make any comparisons between groups or data sets.
3. **You can use the result of a Bayesian analysis to do Decision Analysis.**
4. You can change the underlying statistical model.



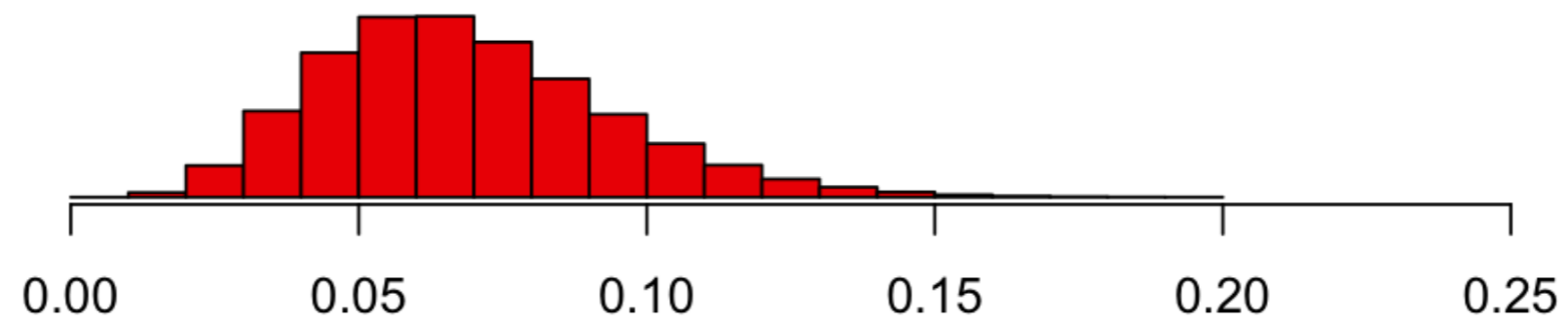


### Video posterior



Proportion of clicks

### Text posterior



Proportion of clicks

# A small decision analysis

```
video_cost <- 0.25  
text_cost <- 0.05  
visitor_spend <- 2.53
```

```
video_cost <- 0.25
text_cost <- 0.05
visitor_spend <- 2.53
posterior
```

```
  video_prop text_prop
1      0.08      0.10
2      0.10      0.07
3      0.16      0.05
4      0.09      0.05
5      0.18      0.03
6      0.13      0.05
7      0.12      0.10
8      0.10      0.04
9      0.11      0.09
10     0.18      0.05
11     0.12      0.04
12     0.13      0.07
```

```
video_cost <- 0.25
text_cost <- 0.05
visitor_spend <- 2.53
posterior$video_profit <- posterior$video_prop * visitor_spend - video_cost
posterior
```

```
  video_prop text_prop video_profit
1      0.08      0.10      -0.04
2      0.10      0.07       0.00
3      0.16      0.05       0.15
4      0.09      0.05      -0.02
5      0.18      0.03       0.21
6      0.13      0.05       0.08
7      0.12      0.10       0.06
8      0.10      0.04       0.01
9      0.11      0.09       0.02
10     0.18      0.05       0.21
11     0.12      0.04       0.06
12     0.13      0.07       0.08
```

```
video_cost <- 0.25
text_cost <- 0.05
visitor_spend <- 2.53
posterior$video_profit <- posterior$video_prop * visitor_spend - video_cost
posterior$text_profit <- posterior$text_prop * visitor_spend - text_cost
posterior
```

```
  video_prop text_prop video_profit text_profit
1      0.08      0.10      -0.04      0.21
2      0.10      0.07       0.00      0.12
3      0.16      0.05       0.15      0.09
4      0.09      0.05      -0.02      0.08
5      0.18      0.03       0.21      0.02
6      0.13      0.05       0.08      0.09
7      0.12      0.10       0.06      0.20
8      0.10      0.04       0.01      0.05
9      0.11      0.09       0.02      0.17
10     0.18      0.05       0.21      0.09
11     0.12      0.04       0.06      0.05
12     0.13      0.07       0.08      0.12
```

```
video_cost <- 0.25
text_cost <- 0.05
visitor_spend <- 2.53
posterior$video_profit <- posterior$video_prop * visitor_spend - video_cost
posterior$text_profit <- posterior$text_prop * visitor_spend - text_cost
posterior$profit_diff <- posterior$video_profit - posterior$text_profit
posterior
```

	video_prop	text_prop	video_profit	text_profit	profit_diff
1	0.08	0.10	-0.04	0.21	-0.26
2	0.10	0.07	0.00	0.12	-0.12
3	0.16	0.05	0.15	0.09	0.07
4	0.09	0.05	-0.02	0.08	-0.10
5	0.18	0.03	0.21	0.02	0.18
6	0.13	0.05	0.08	0.09	0.00
7	0.12	0.10	0.06	0.20	-0.14
8	0.10	0.04	0.01	0.05	-0.04
9	0.11	0.09	0.02	0.17	-0.15
10	0.18	0.05	0.21	0.09	0.12
11	0.12	0.04	0.06	0.05	0.00
12	0.13	0.07	0.08	0.12	-0.04

# Make a data informed decision!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

# Change anything and everything

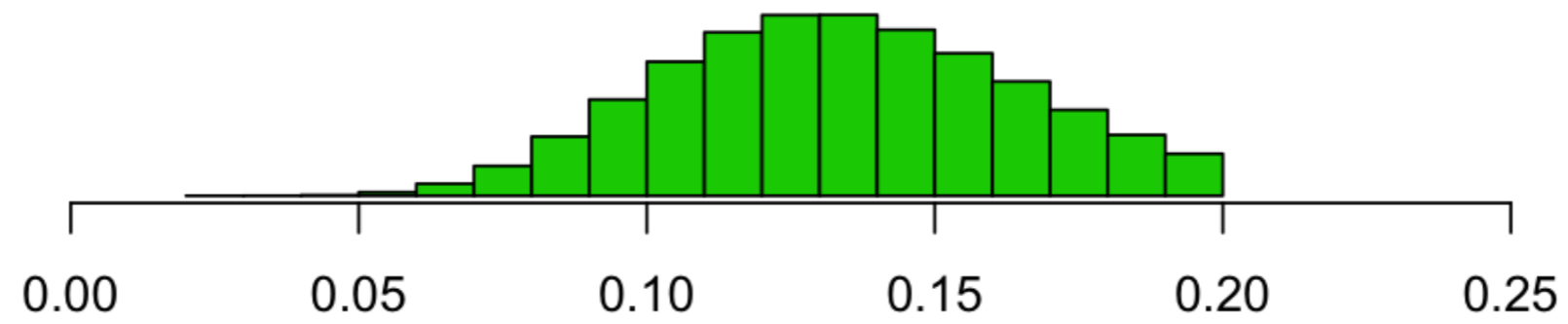
FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R



**Rasmus Bååth**  
Data Scientist

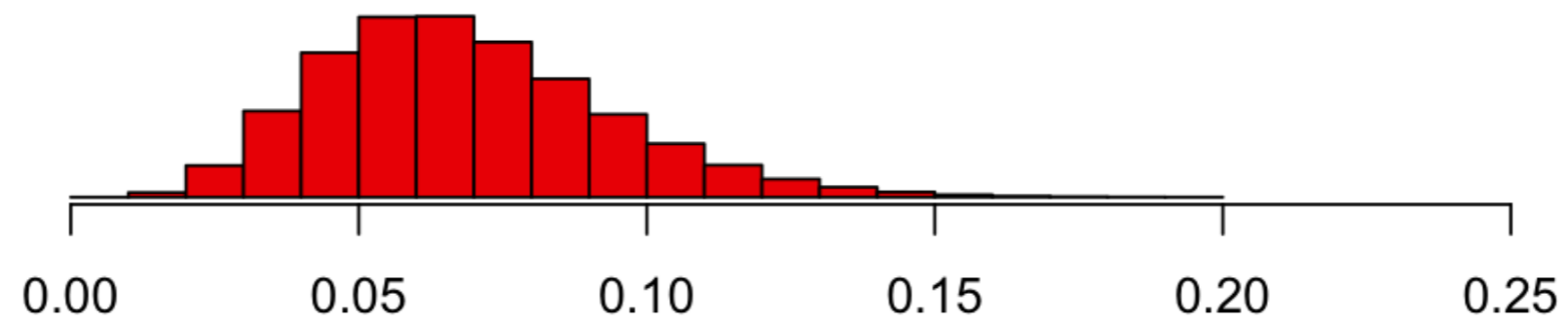


### Video posterior



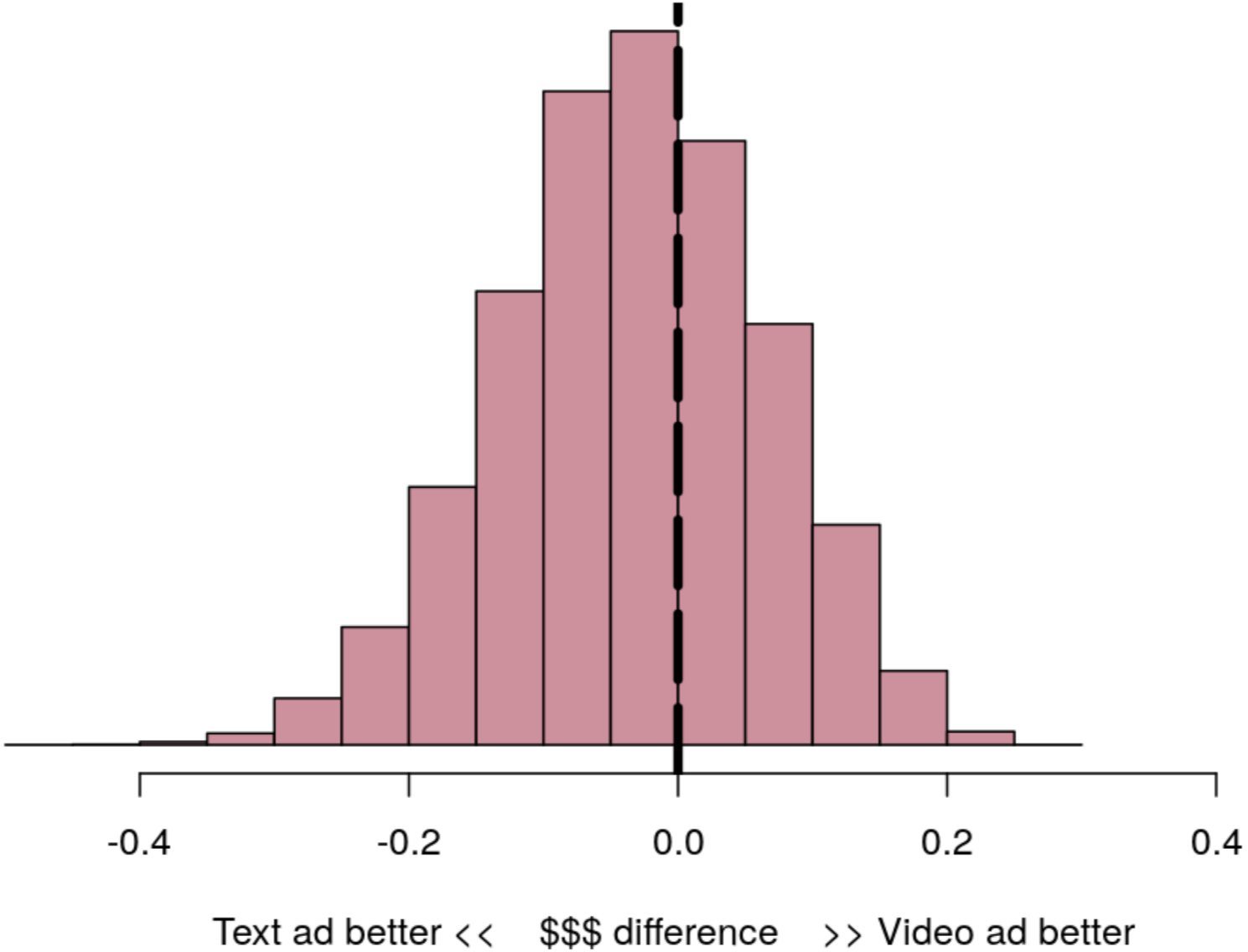
Proportion of clicks

### Text posterior



Proportion of clicks

### Posterior difference in profit



# Next up on reasons to use Bayesian data analysis

1. You can include information sources in addition to the data.
2. You can make any comparisons between groups or data sets.
3. You can use the result of a Bayesian analysis to do Decision Analysis.
4. **You can change the underlying statistical model.**

# Completely switch out the binomial model

- Why? Well, you have some new data...
- A *banner ad* for your site.
- You **don't** pay per view, you pay **per day**.
- A trial resulted in 19 clicks in a day
- *How many daily site visits, should we expect, on average, if we pay for this banner?*

# A model for counts per day

- Split the day into 1440 minutes.
- What proportion of minutes results in a click on the ad?

# A model for counts per day

- ~~Split the day into 1440 minutes.~~
- ~~What proportion of minutes results in a click on the ad?~~
- Split the day into 86400 seconds.
- What proportion of seconds results in a click on the ad?

# A model for counts per day

- ~~Split the day into 1440 minutes.~~
- ~~What proportion of minutes results in a click on the ad?~~
- ~~Split the day into 86400 seconds.~~
- ~~What proportion of seconds results in a click on the ad?~~
- Split the day into 86400000 milliseconds.
- What proportion of milliseconds results in a click on the ad?

# A model for counts per day

- ~~Split the day into 1440 minutes.~~
- ~~What proportion of minutes results in a click on the ad?~~
- ~~Split the day into 86400 seconds.~~
- ~~What proportion of seconds results in a click on the ad?~~
- ~~Split the day into 86400000 milliseconds.~~
- ~~What proportion of milliseconds results in a click on the ad?~~
- Split the day into infinite parts...
- ???

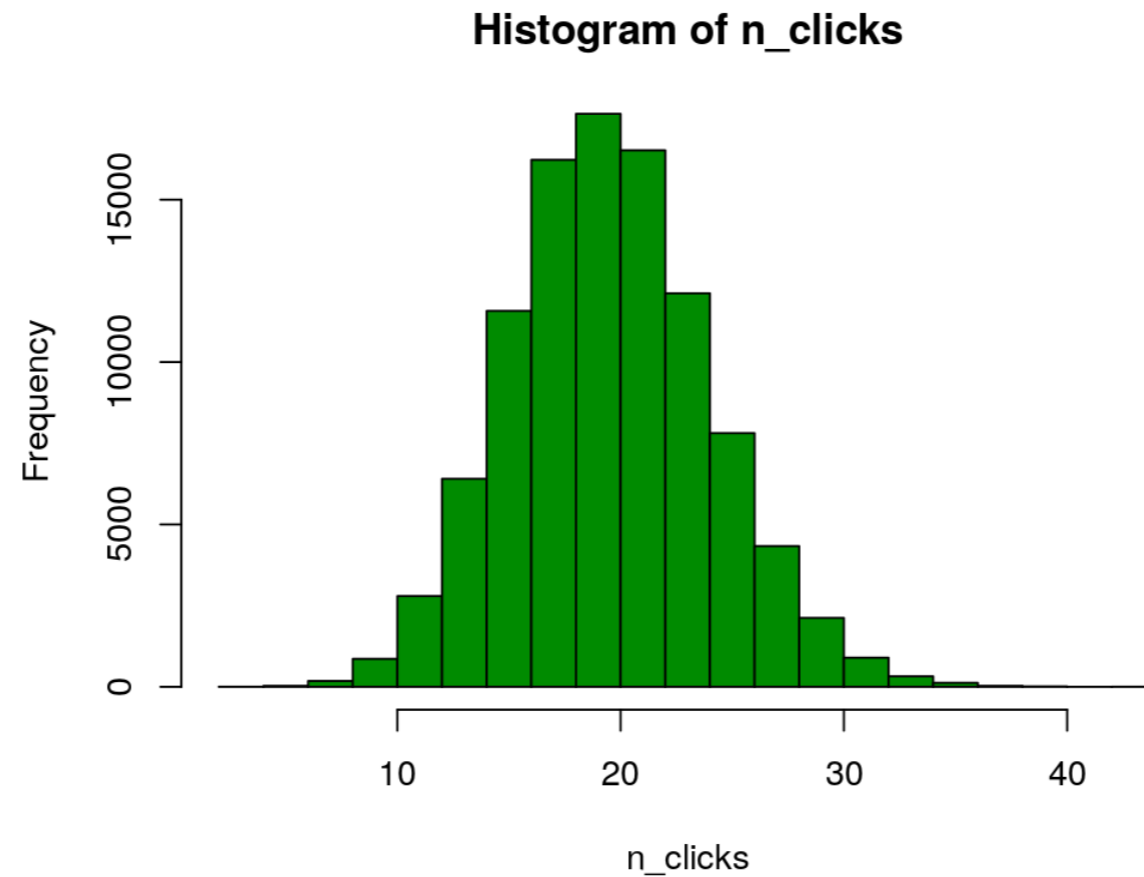


# The Poisson distribution

- One parameter: The mean number of events per time unit.
- `rpois` samples from the Poisson distribution.

# The Poisson distribution

```
n_clicks <- rpois(n = 100000, lambda = 20)  
hist(n_clicks)
```



**Let's find out in the  
exercises!**

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

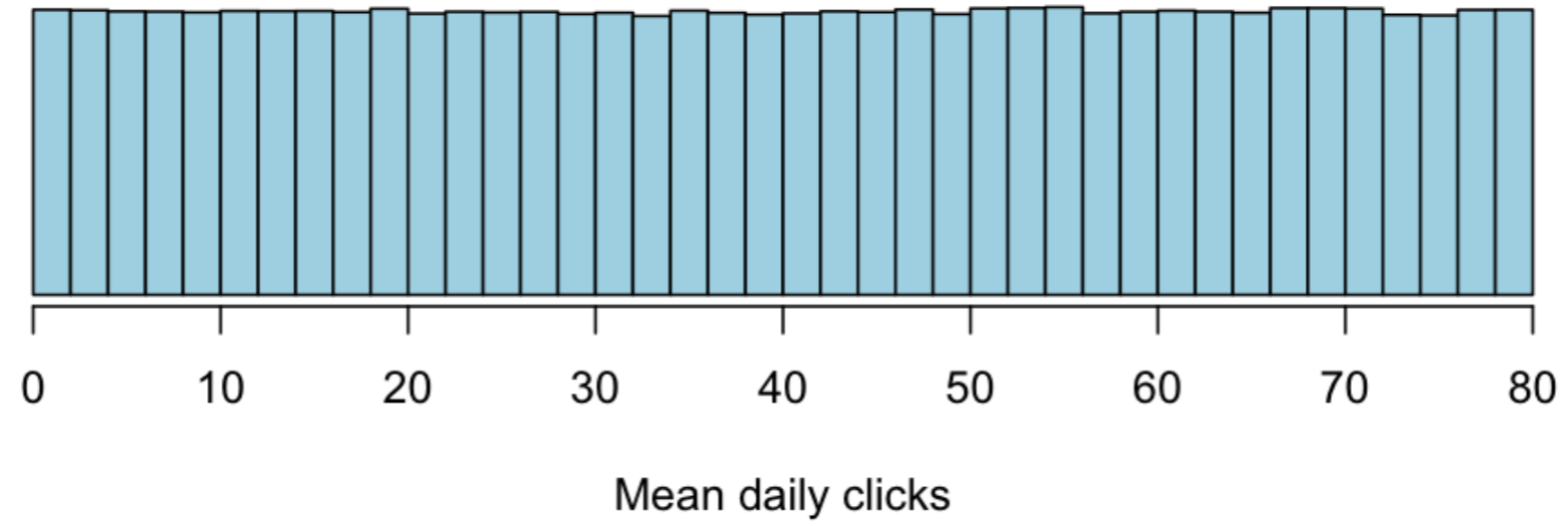
# You just replaced the whole model!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

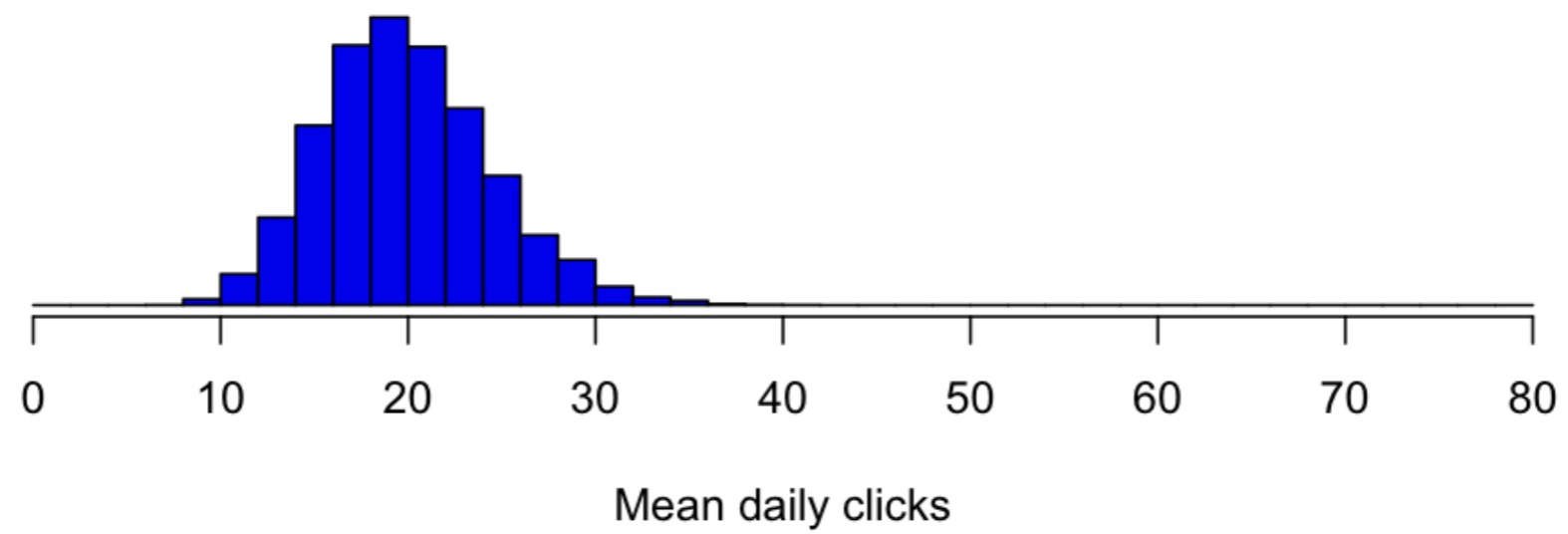


**Rasmus Bååth**  
Data Scientist

## Prior

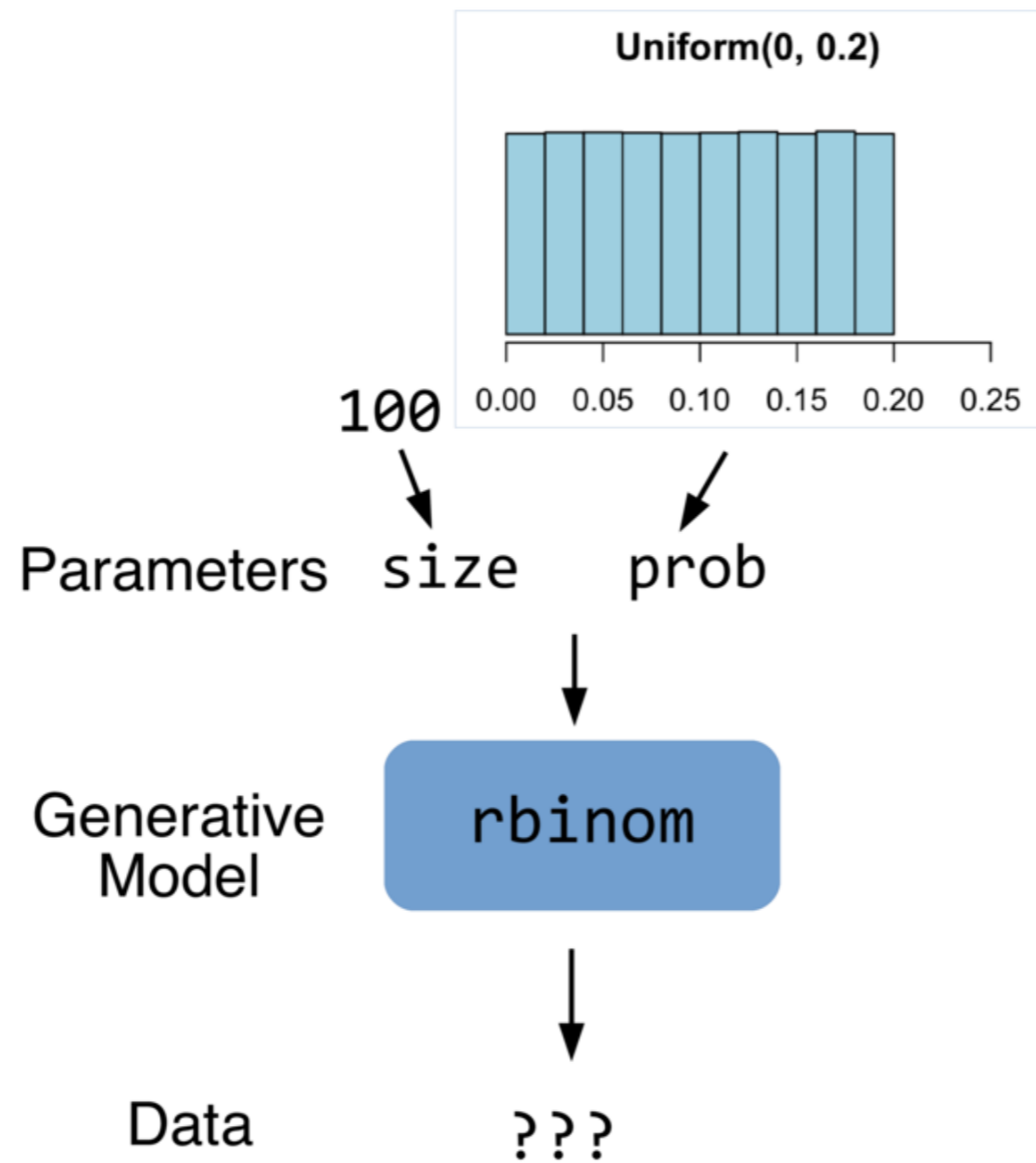


## Posterior



# Some ways Bayesian data analysis can be useful

1. You can include information sources in addition to the data.
2. You can make any comparisons between groups or data sets.
3. You can use the result of a Bayesian analysis to do Decision Analysis.
4. You can change the underlying statistical model.
5. **Bayesian inference is optimal, kind of.**



# Nice properties of Bayes

- *Bayes is optimal, in the small world of the model.*
- *In Bayesian data analysis there is a separation between model and computation.*



# Next up: How to fit Bayesian models more efficiently!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R