

Welcome to the course!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R



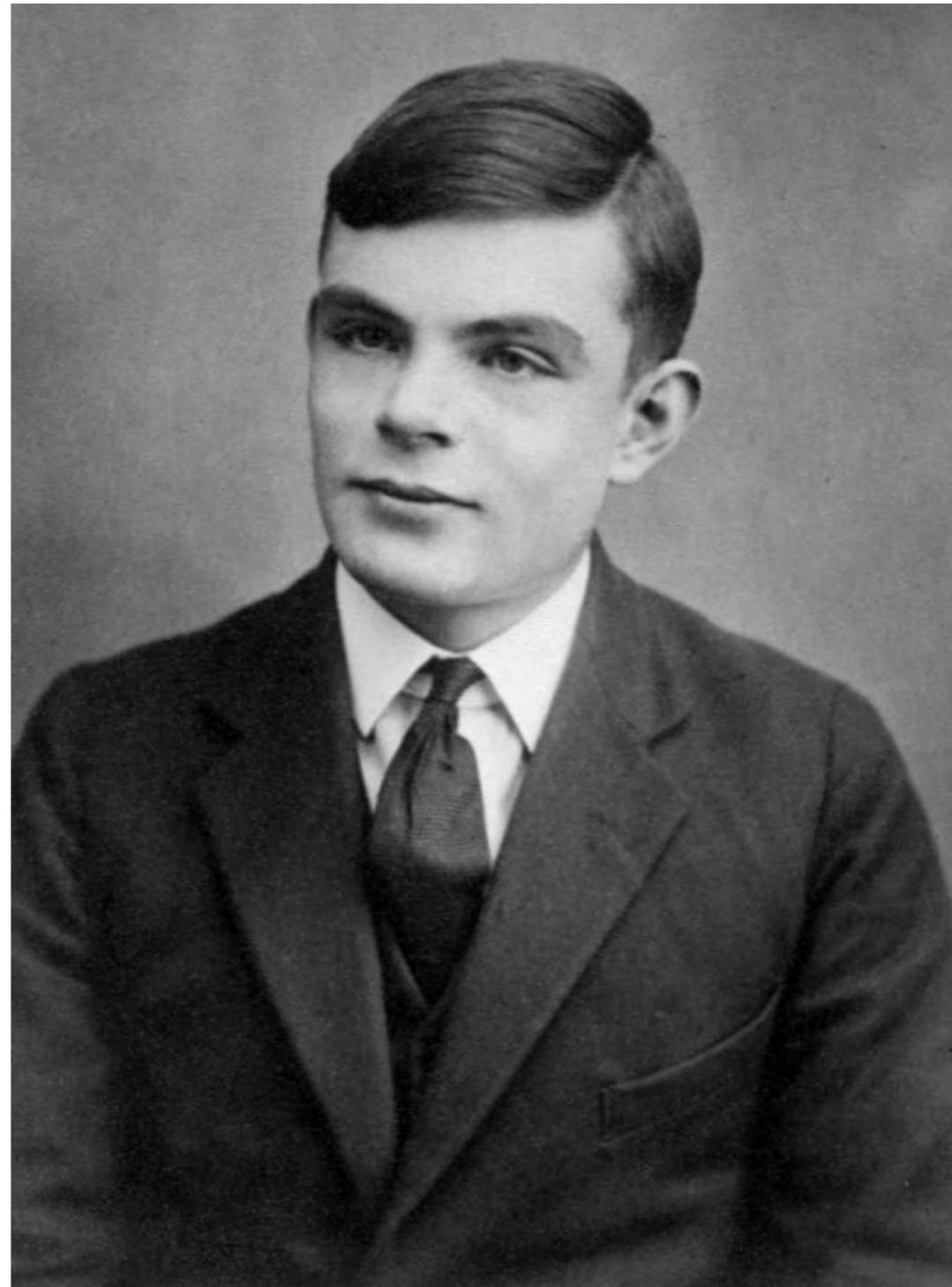
Rasmus Bååth
Data Scientist

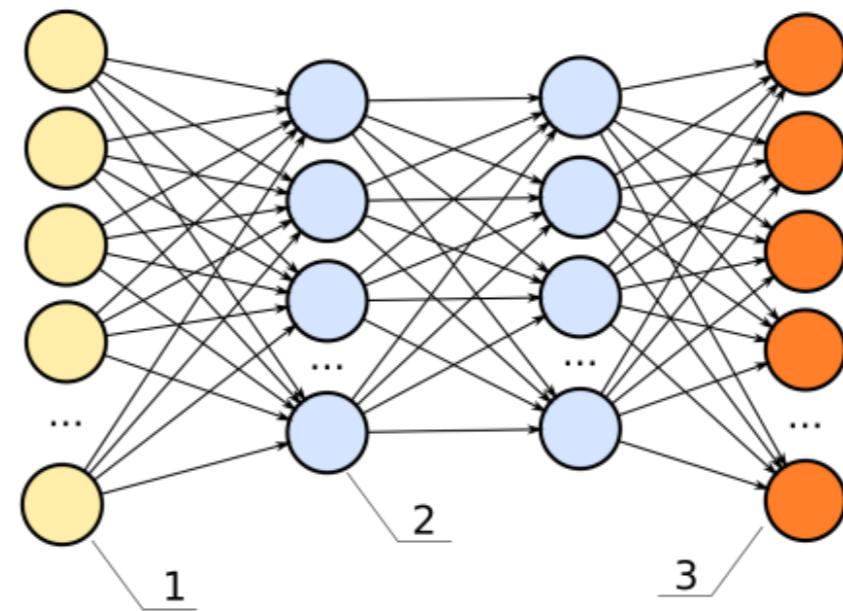
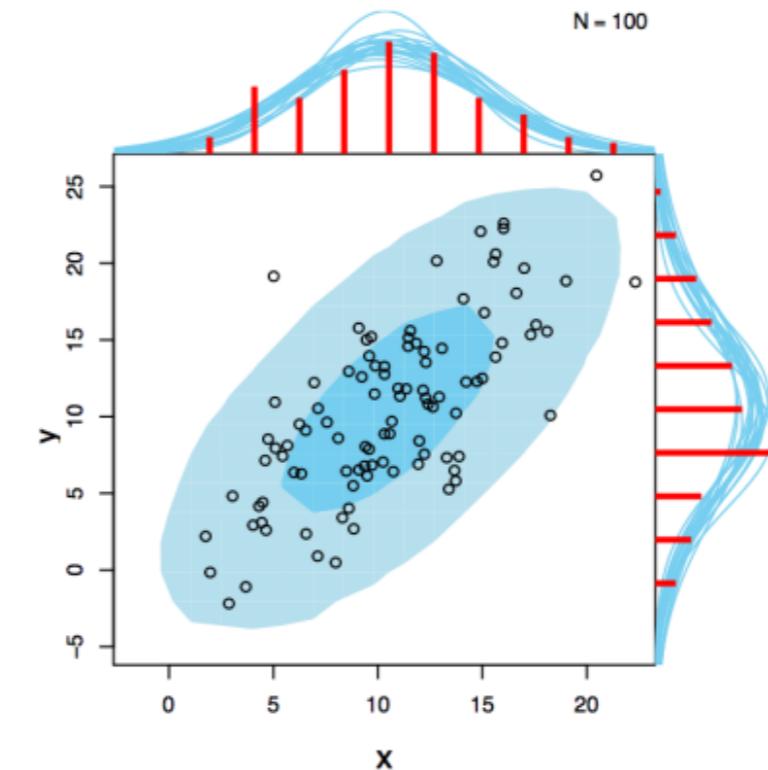
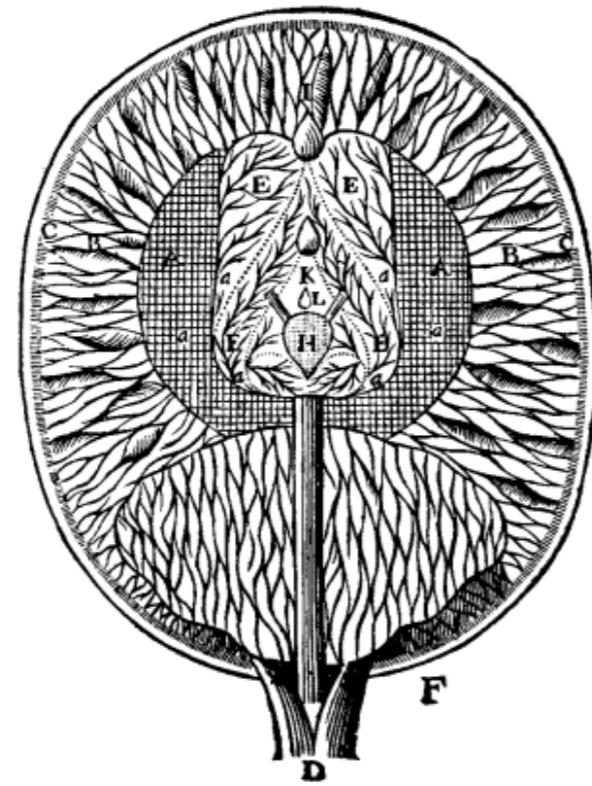






¹ https://commons.wikimedia.org/wiki/File:Enigma_08.jpg



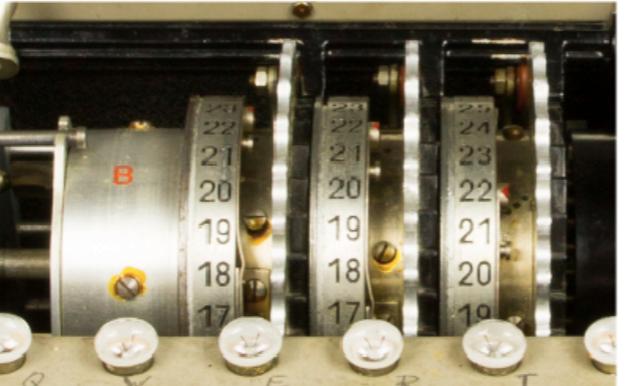


Bayesian inference in a nutshell

A method for figuring out unobservable quantities given known facts that uses probability to describe the uncertainty over what the values of the unknown quantities could be.



Wheel settings

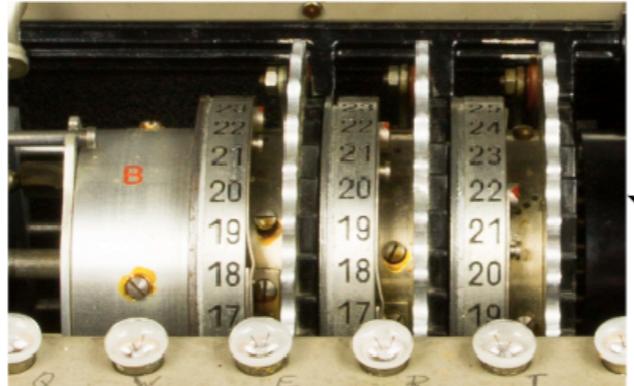


Enigma
model



JAZSFOXRQERSPXEIYUA
PARHCWSMYXCJIMFGVOAH
SJPQJYYKEOABSAUZYNQL

Wheel settings



Enigma
model

JAZSFOXRQERSPXEIYUA
PARHCWSMYXCJIMFGVOAH
SJPQJYYKEOABSAUZYNQL

Bayesian Inference

Bayesian data analysis

- The use of Bayesian inference to learn from data.
- Can be used for hypothesis testing, linear regression, etc.
- Is flexible and allows you to construct problem-specific models.

Course overview

- **Chapter 1:** A small Bayesian analysis.
- **Chapter 2:** How Bayesian inference works.
- **Chapter 3:** Why you would want to use Bayesian data analysis?
- **Chapter 4:** Bayesian inference with Bayes theorem.
- **Chapter 5:** Wrapping up + a practical tool for Bayesian Data Analysis in R.

Bayesian Data
analysis: a tool to
make sense of your
data.

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

A little bit of background

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R



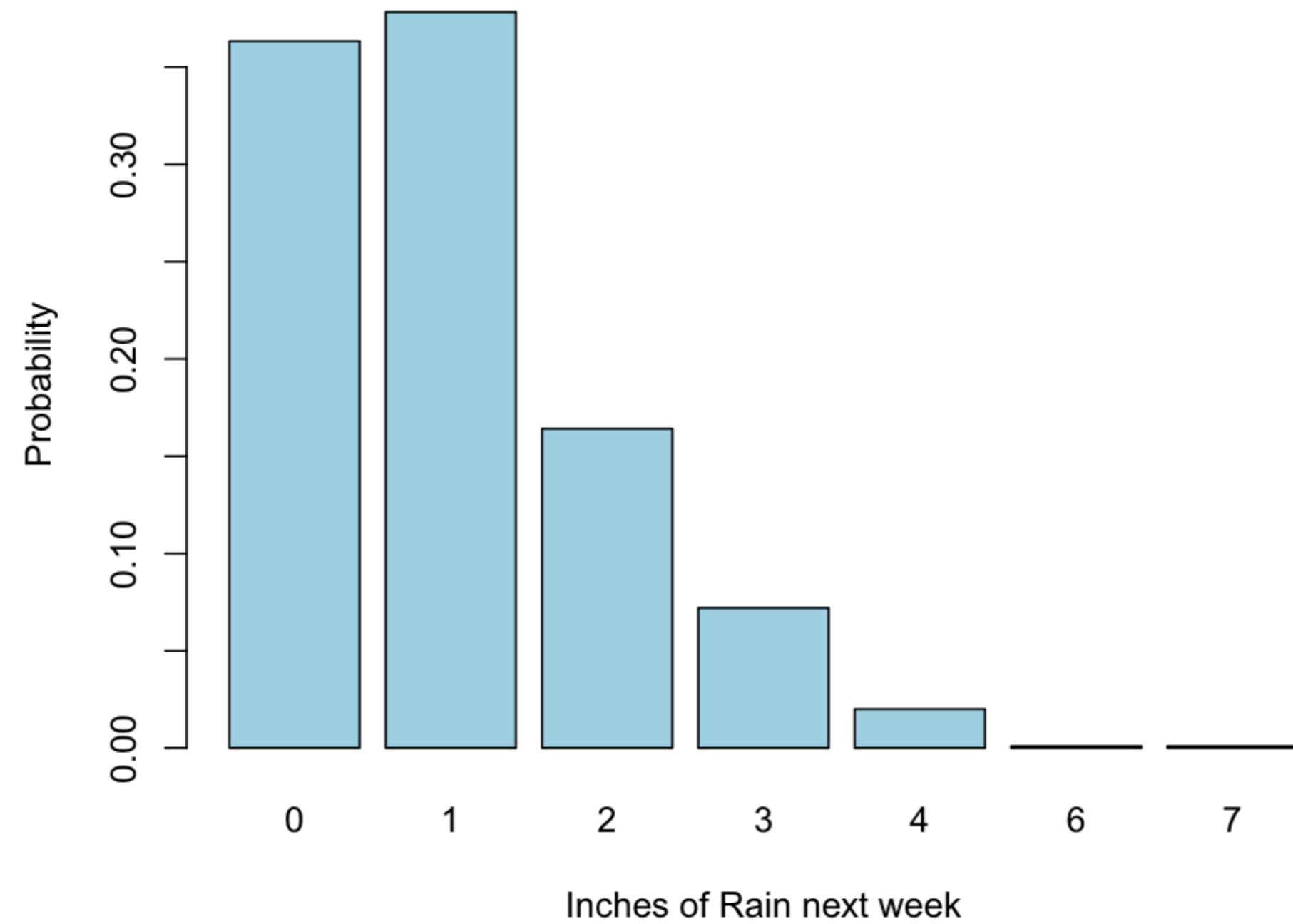
Rasmus Bååth
Data Scientist



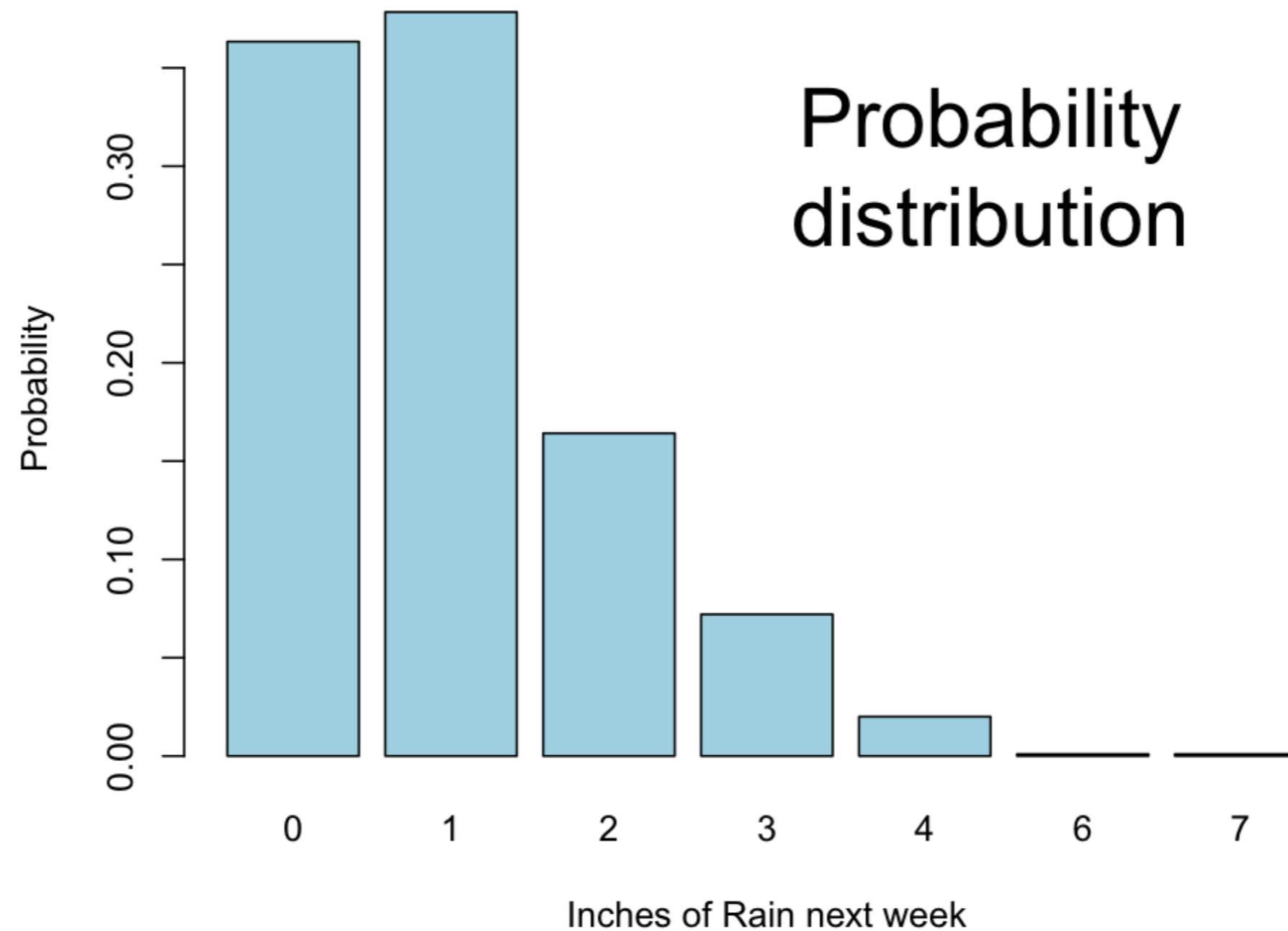
Thomas Bayes (1702-1761)

Probability

- A number between 0 and 1.
- A statement about certainty / uncertainty.
- 1 is complete certainty something is the case.
- 0 is complete certainty something is *not* the case.
- Not only about yes/no events.



Probability distribution



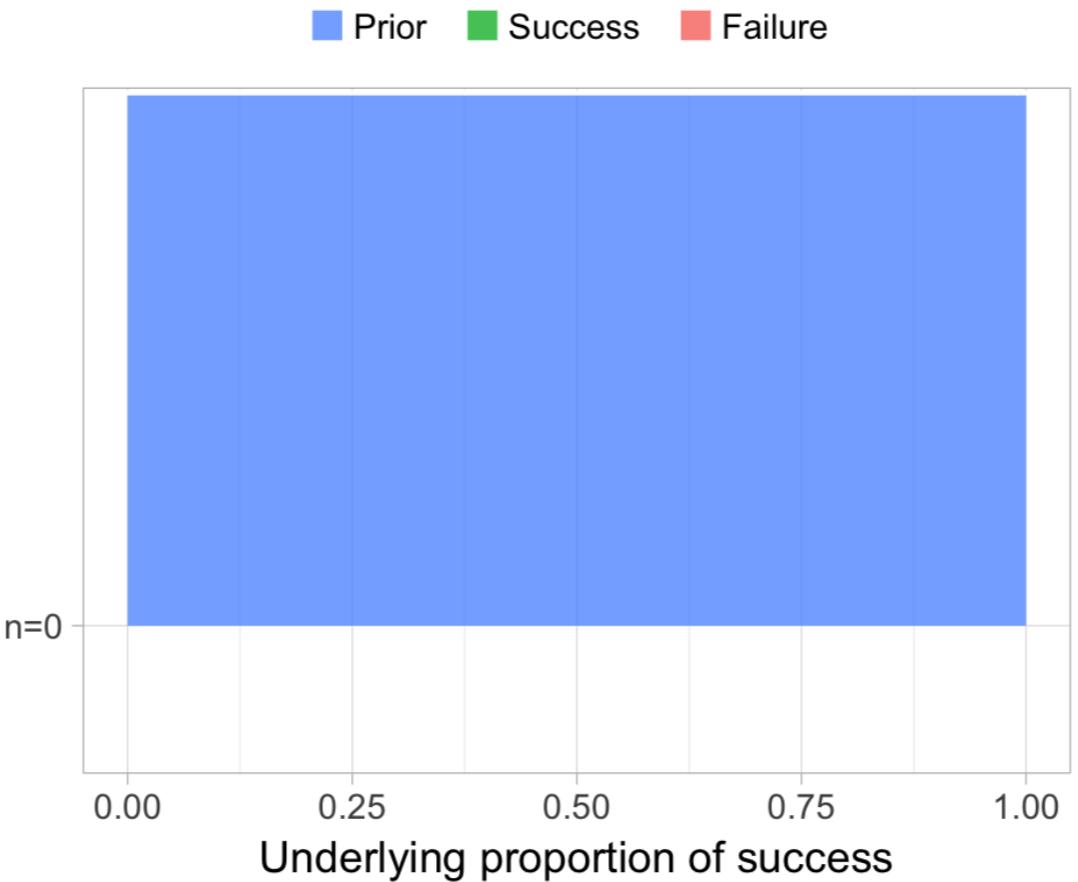
The role of probability distributions in Bayesian data analysis is to represent uncertainty, and the role of Bayesian inference is to update probability distributions to reflect what has been learned from data.

A Bayesian model for the proportion of success

- `prop_model(data)`
- The `data` is a vector of successes and failures represented by `1` s and `0` s.
- There is an unknown underlying *proportion of success*.
- If data point is a success is only affected by this proportion.
- Prior to seeing any data, any underlying proportion of success is equally likely.
- The result is a probability distribution that represents what the model knows about the underlying proportion of success.

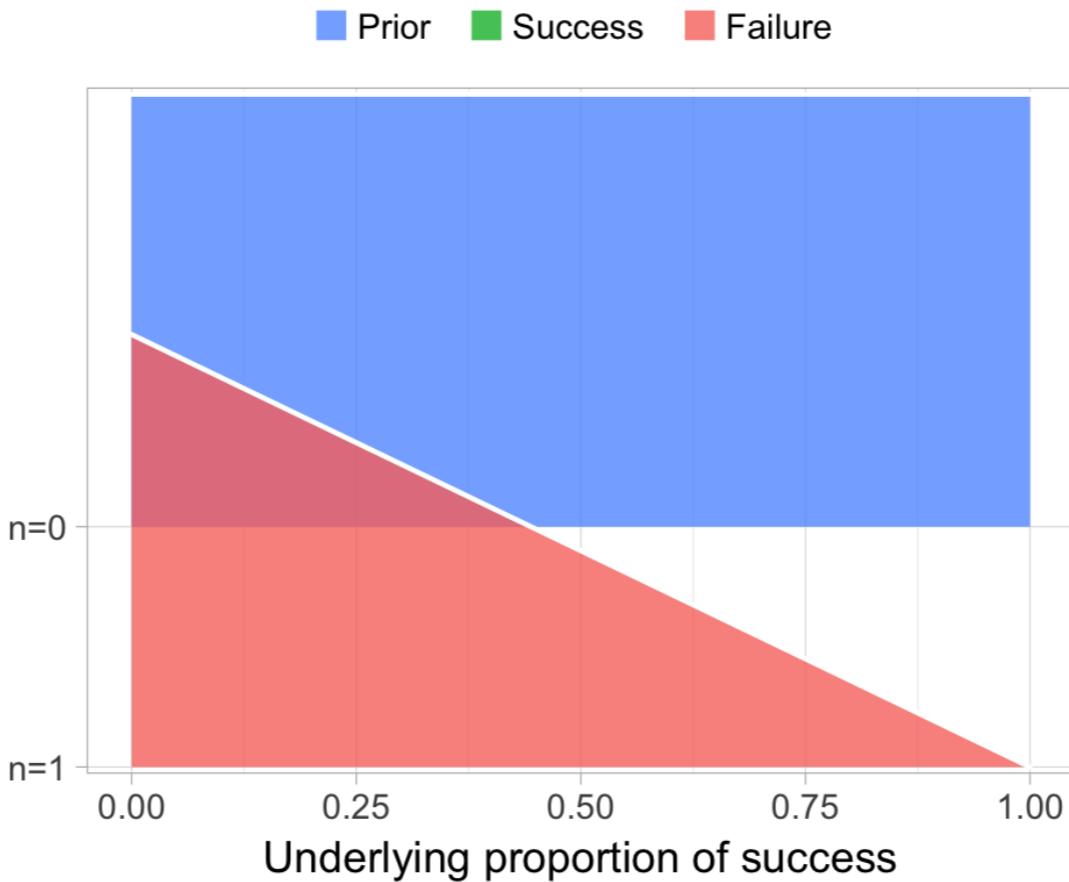
Trying out prop_model

```
data <- c()  
prop_model(data)
```



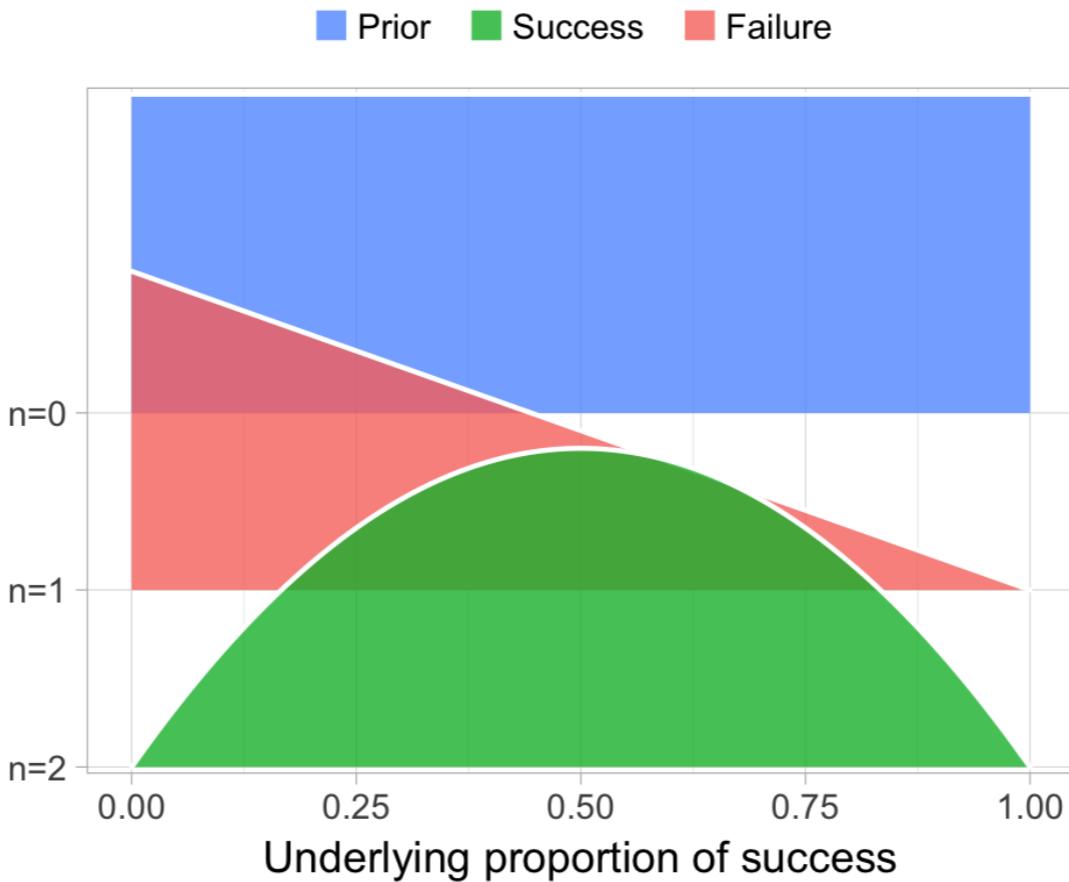
Trying out prop_model

```
data <- c(0)  
prop_model(data)
```



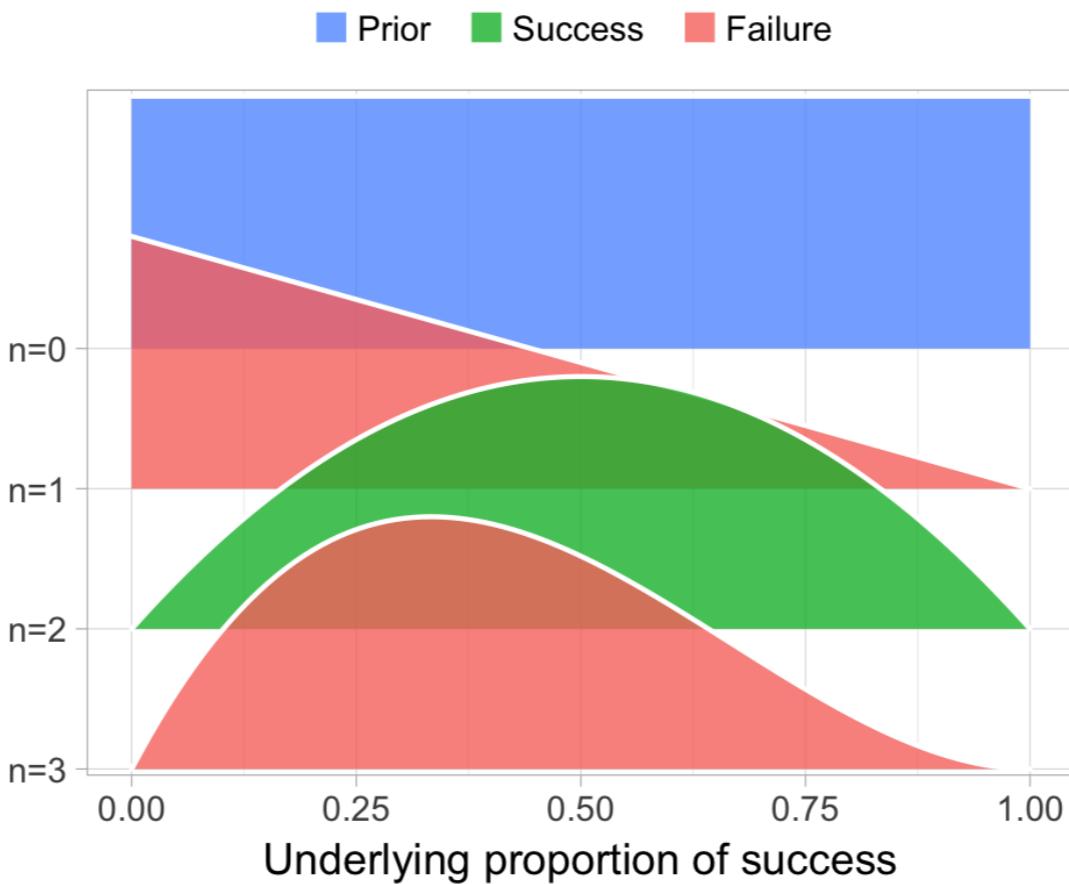
Trying out prop_model

```
data <- c(0, 1)  
prop_model(data)
```



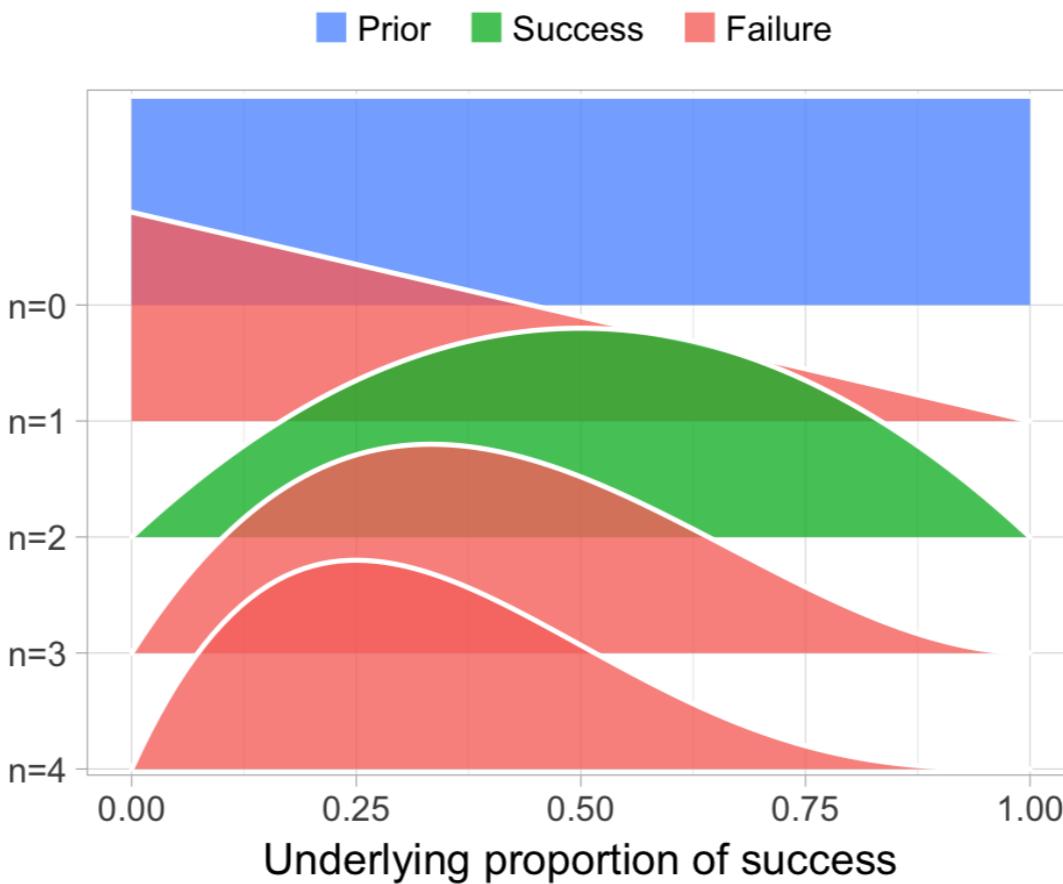
Trying out prop_model

```
data <- c(0, 1, 0)  
prop_model(data)
```



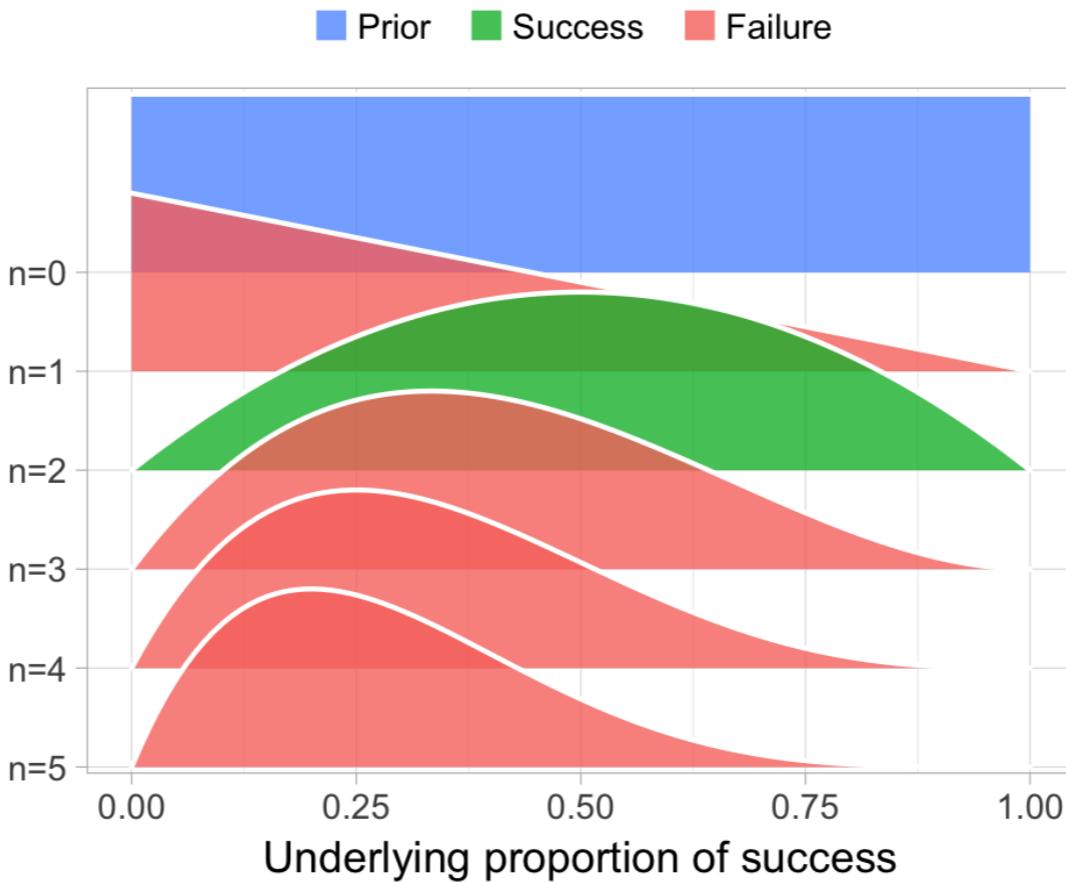
Trying out prop_model

```
data <- c(0, 1, 0, 0)  
prop_model(data)
```



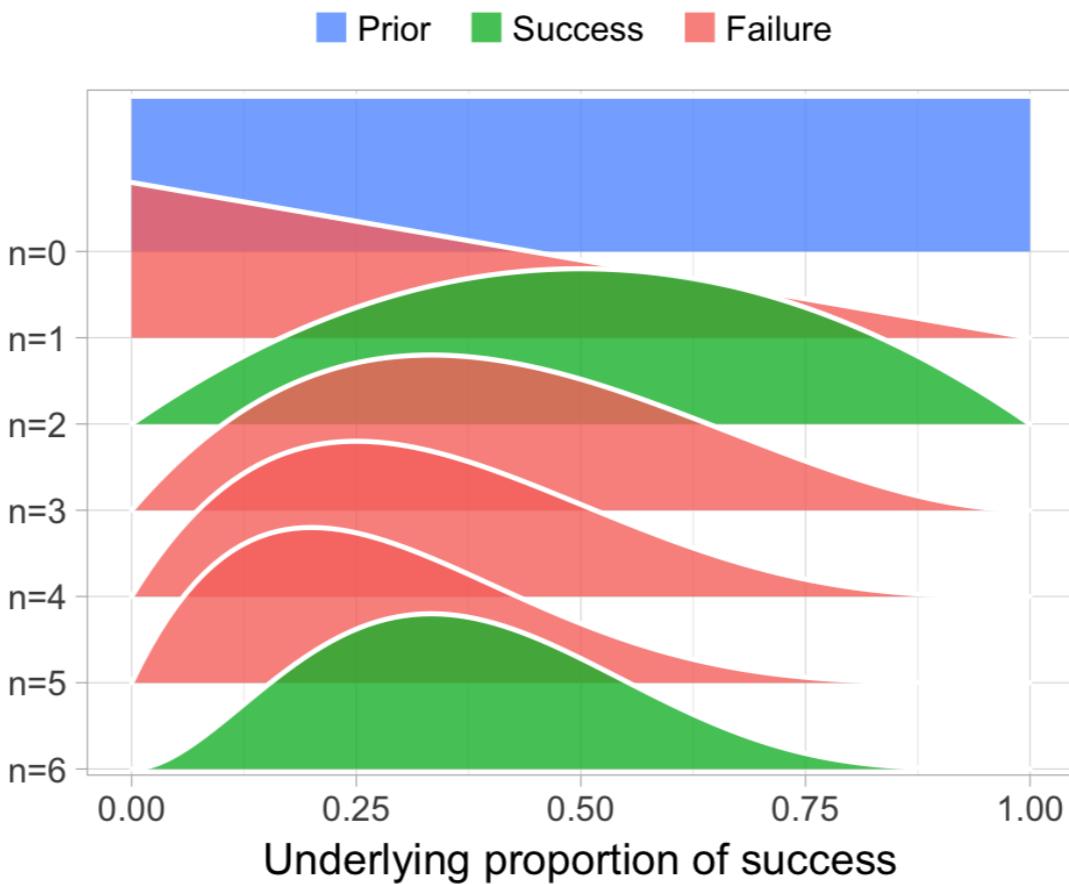
Trying out prop_model

```
data <- c(0, 1, 0, 0, 0)  
prop_model(data)
```



Trying out prop_model

```
data <- c(0, 1, 0, 0, 0, 1)  
prop_model(data)
```



**Now, you try out
prop_model!**

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

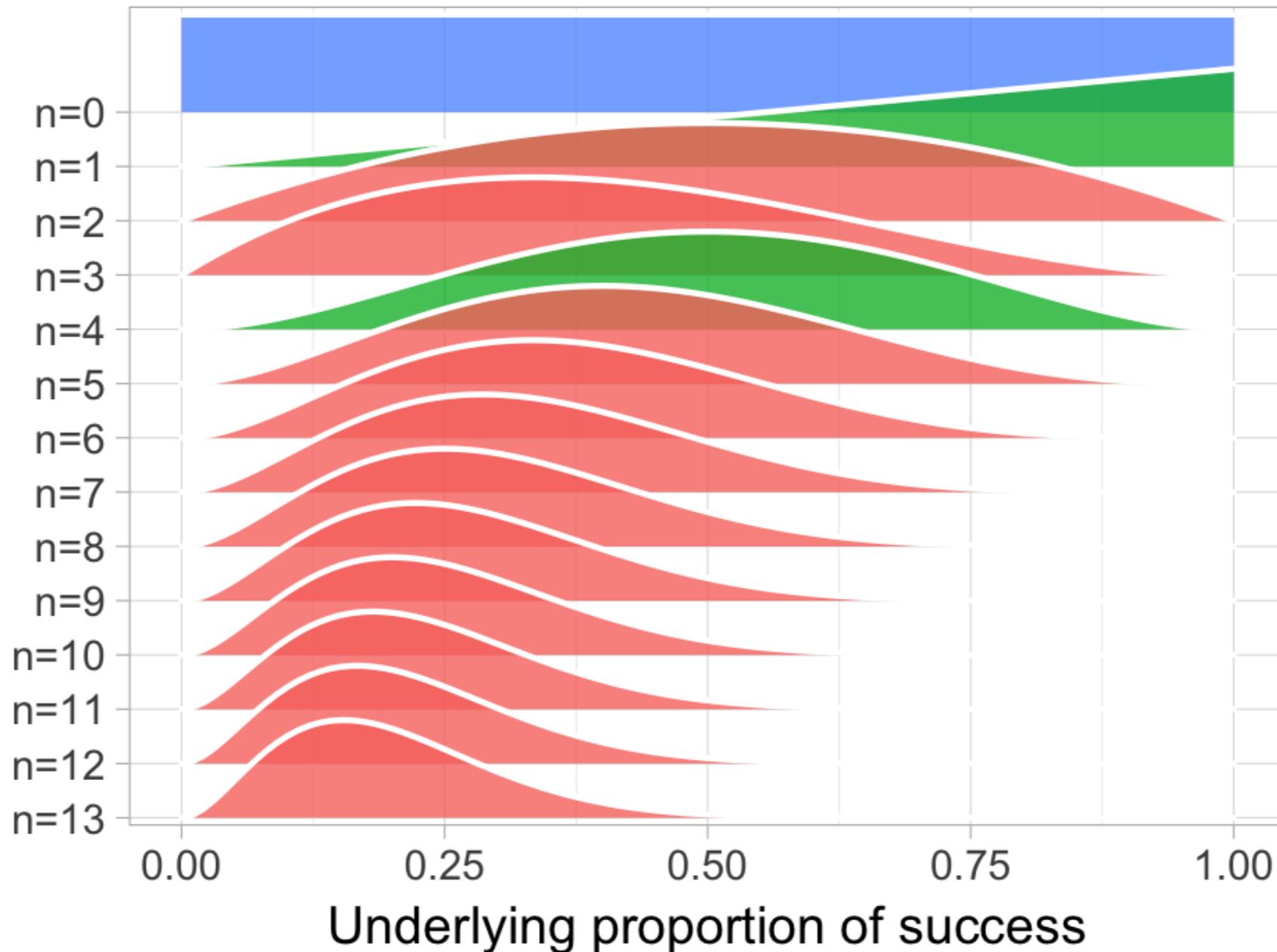
You just did some
Bayesian data
analysis!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

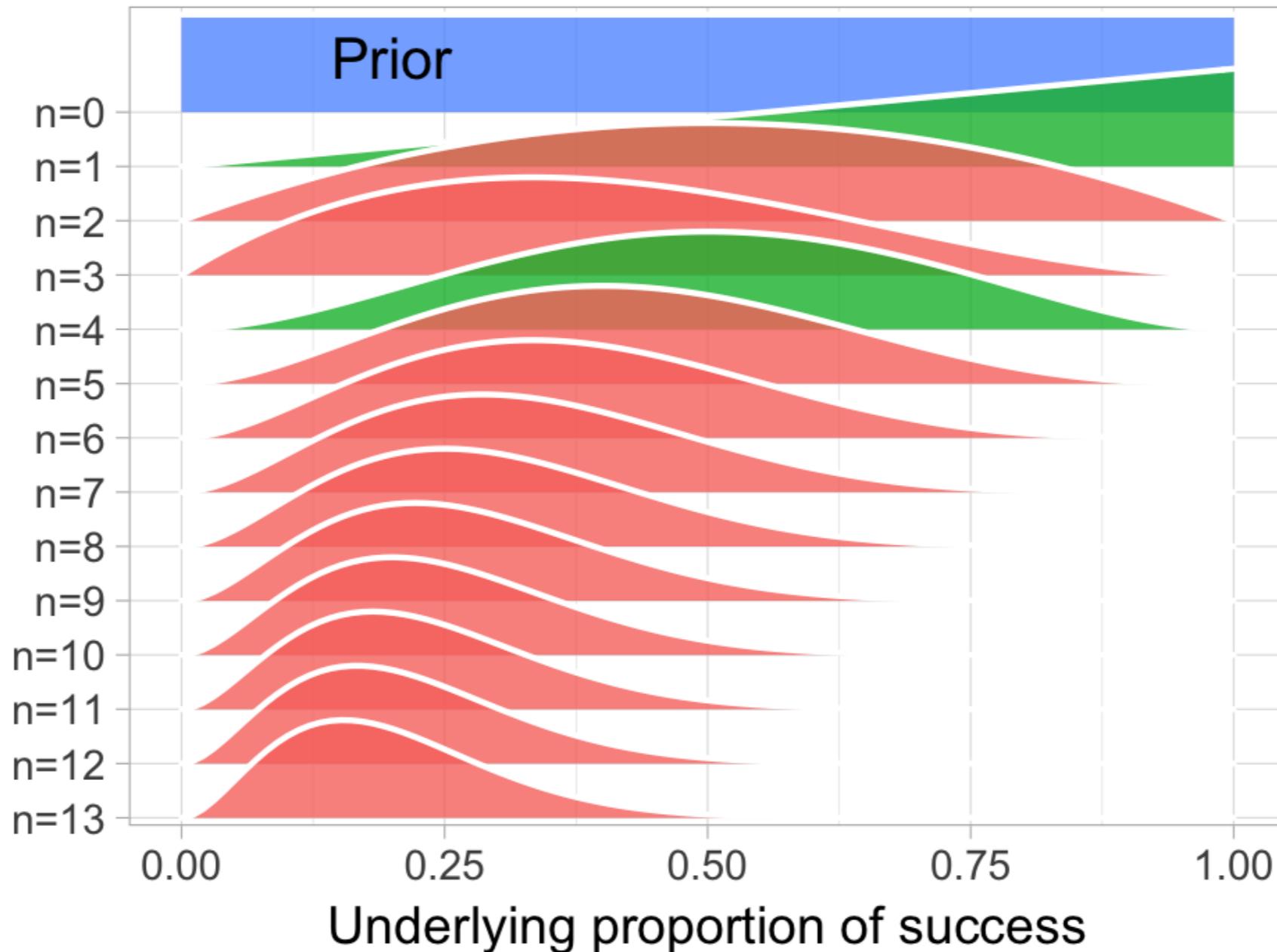


Rasmus Bååth
Data Scientist

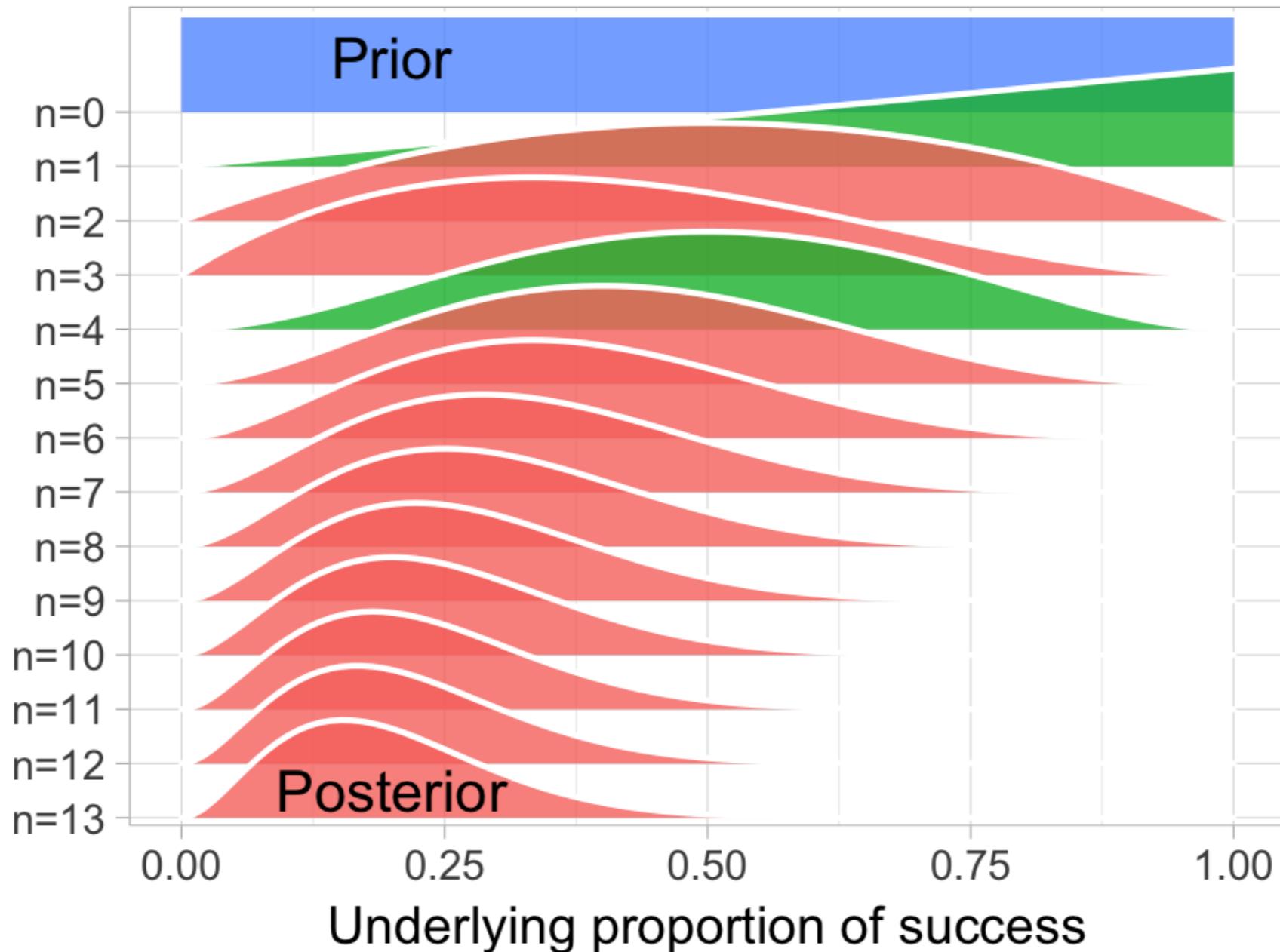
Prior Success Failure



Prior Success Failure



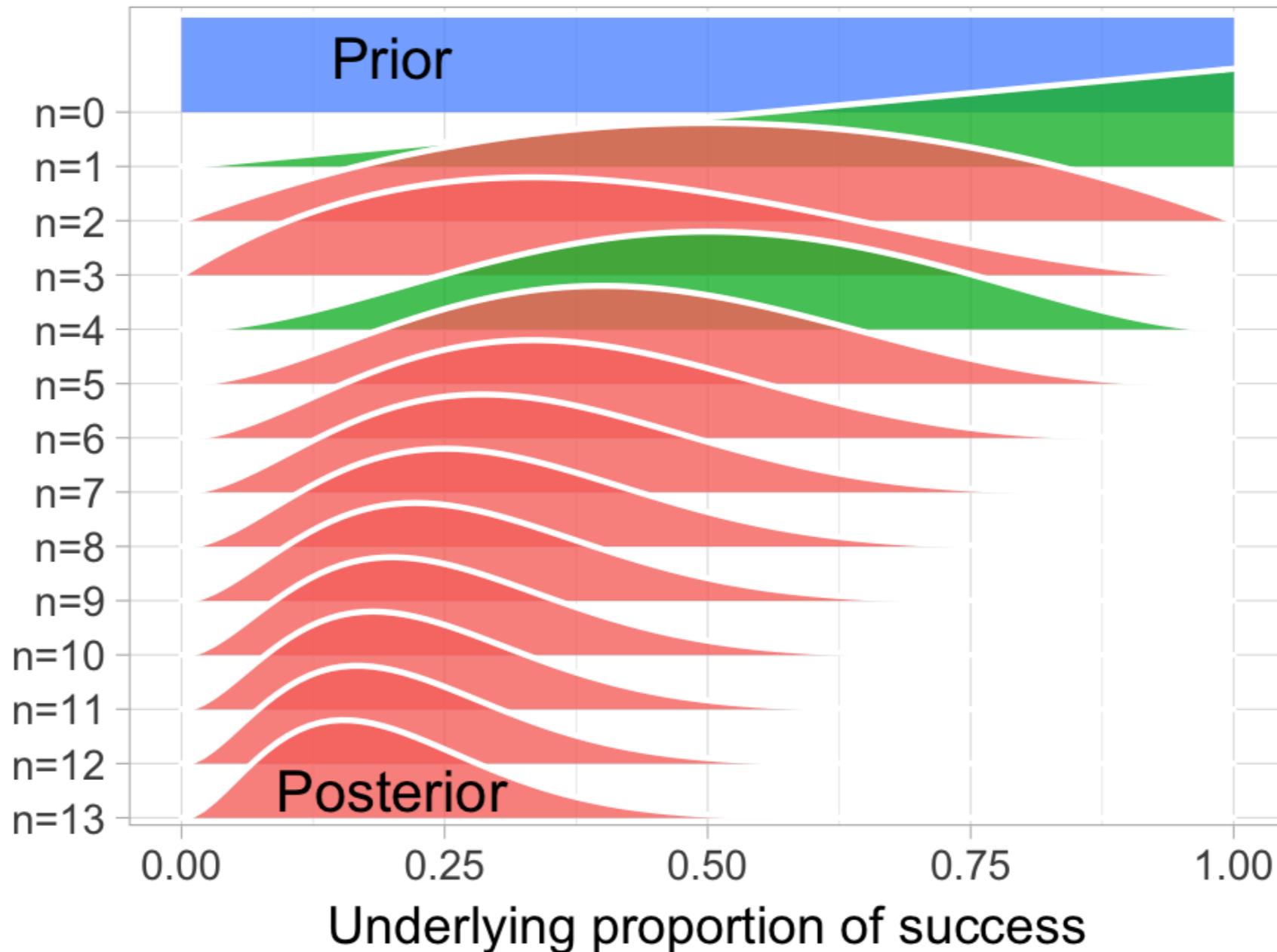
Prior Success Failure



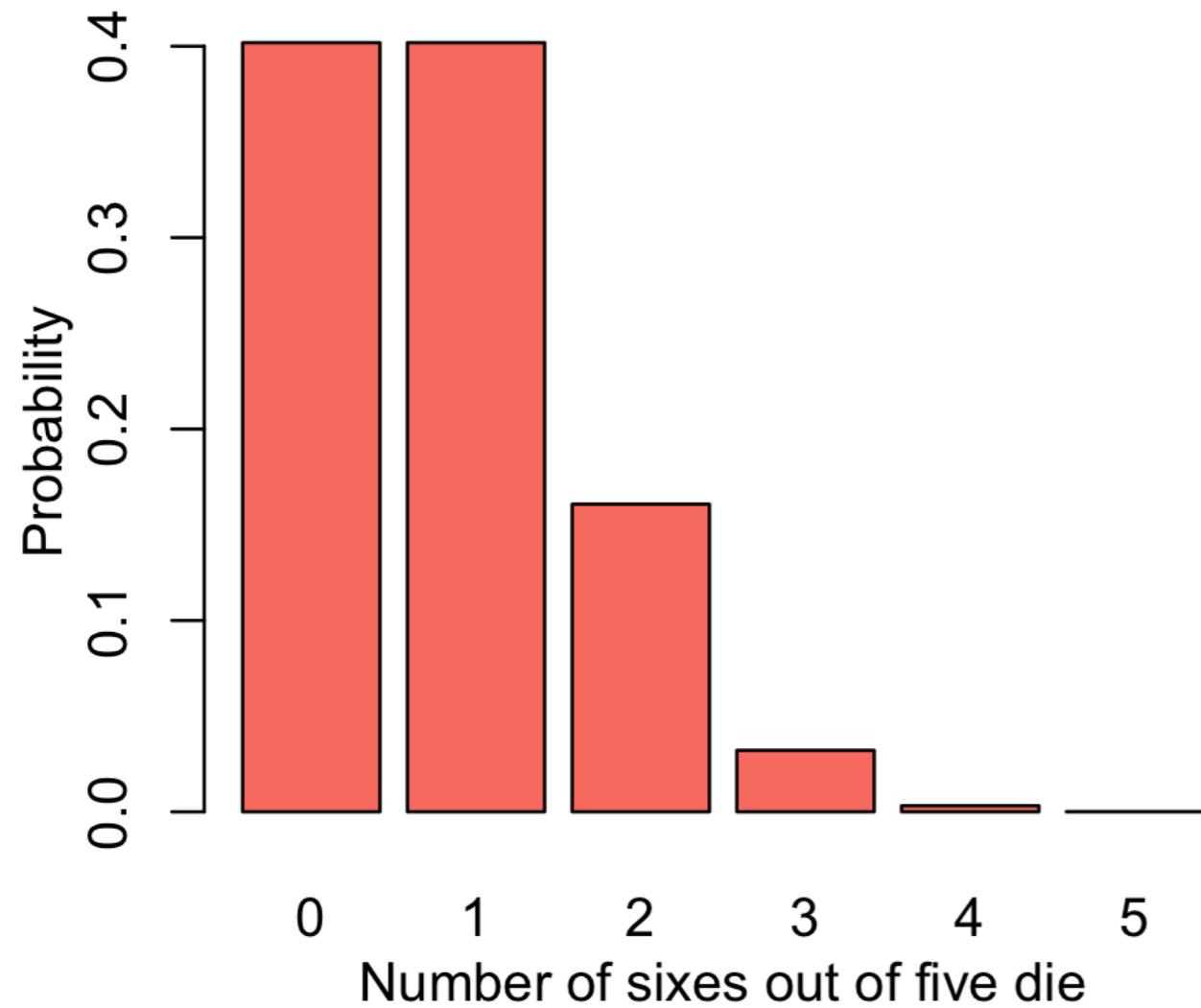
Priors & Posteriors

- A **prior** is a probability distribution that represents what the model knows before seeing the data.
- A **posterior** is a probability distribution that represents what the model knows after having seen the data.

Prior Success Failure



The probability distribution over the number of 6's when rolling 5 dice



$$p(x) = \binom{5}{x} \left(\frac{1}{6}\right)^x \left(1 - \left(\frac{1}{6}\right)\right)^{5-x}$$

```
number_of_sixes
```

```
1 1 0 0 1 0 1 2 0 1 0 0 1 0 0 0 0 0 1 1 1 0 1 1 0 0 2 0  
0 1 0 0 1 0 0 1 0 1 2 0 1 0 0 0 1 2 1 2 0 0 1 1 3 3 0 0  
1 1 1 1 1 0 0 1 2 0 1 3 1 1 1 0 1 0 1 2 0 1 1 0 1 1 1 0  
2 1 0 4 0 1 2 1 1 1 2 0 1 0 1 1 0 0 2 0 0 0 0 0 1 1 0 1  
0 0 0 0 2 0 0 0 0 0 1 1 0 0 2 1 1 1 0 2 1 1 1 0 0 1 1 1  
...
```

```
mean(number_of_sixes)
```

```
0.83
```

```
posterior <- prop_model(data)
print(posterior)
```

```
0.23 0.36 0.20 0.21 0.12
0.10 0.03 0.16 0.09 0.14
0.23 0.05 0.15 0.26 0.22
...
```

Finish off the Zombie drug analysis!

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R

Wrapping up the zombie analysis

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R



Rasmus Bååth
Data Scientist

The result of the zombie analysis

```
data = c(1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)  
posterior <- prop_model(data)  
median(posterior)
```

```
0.19
```

```
quantile(posterior, c(0.05, 0.95))
```

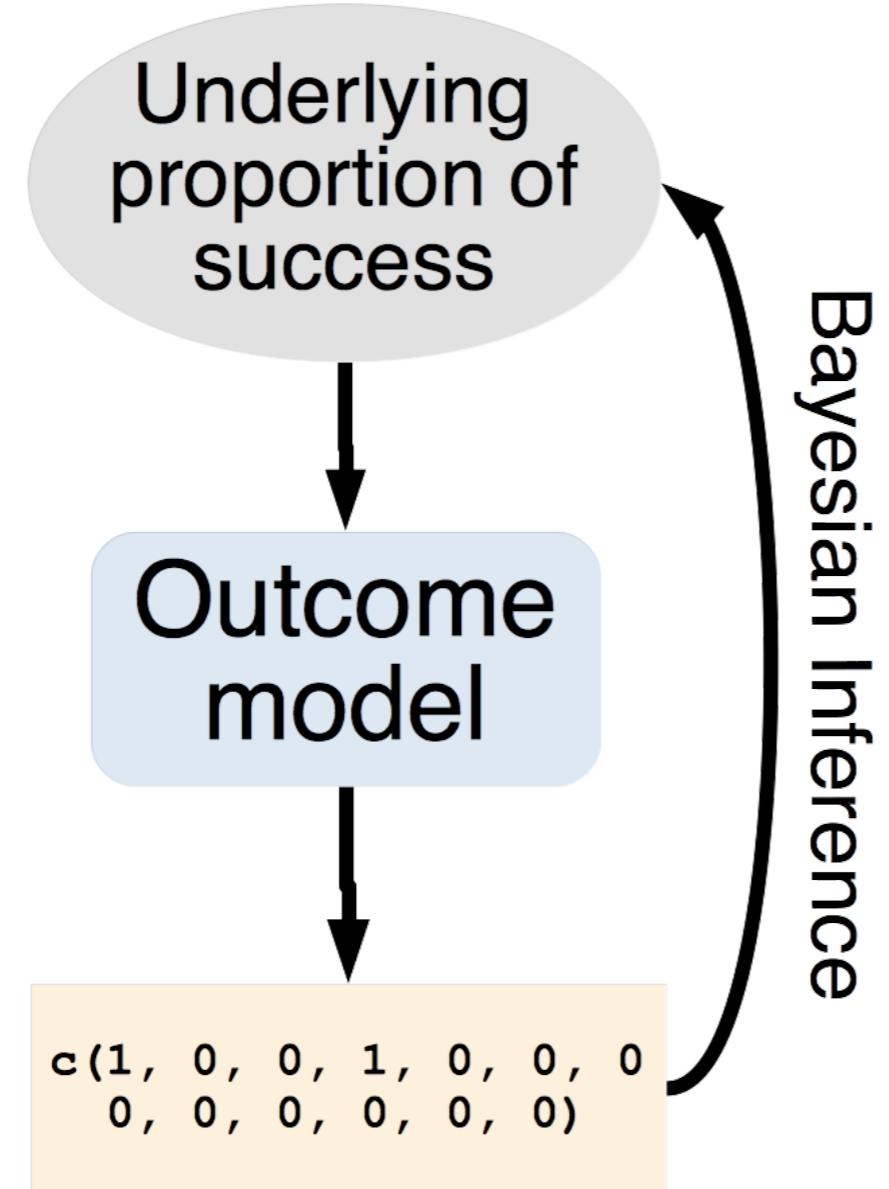
```
5%      95%  
0.06   0.39
```

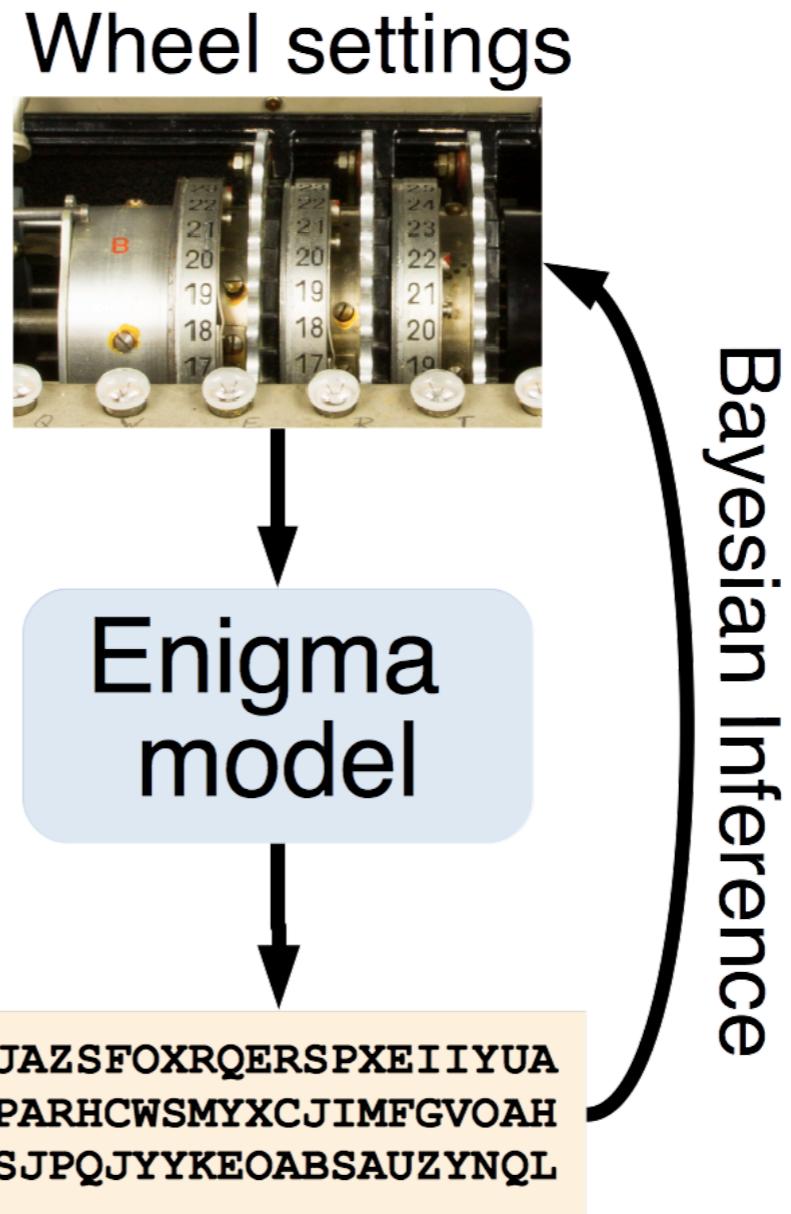
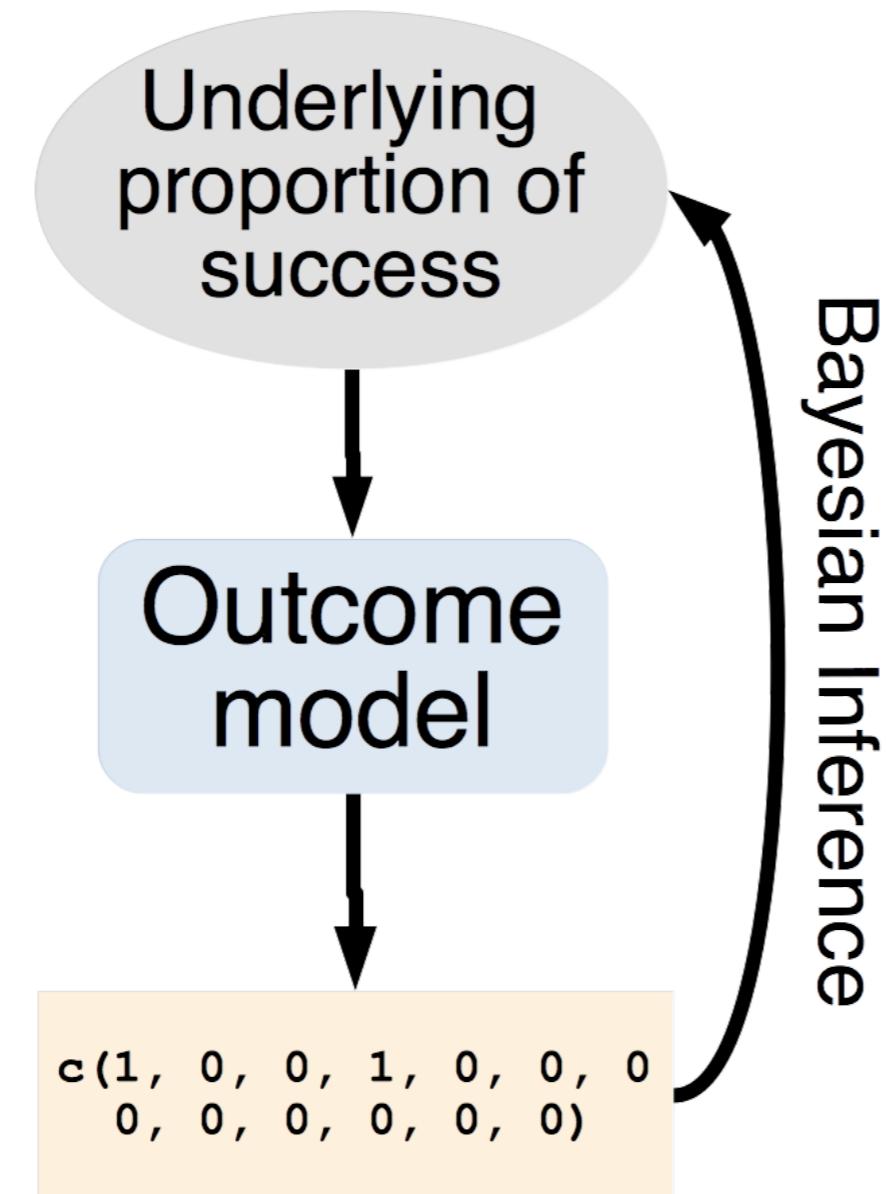
```
sum(posterior > 0.07) / length(posterior)
```

```
0.93
```

The result in a journal

Given the data of two cured and 11 relapsed zombies, and using the Bayesian model described before, there is a 90% probability that our drug cures between 6% and 39% of treated zombies. Further, there is 93% probability that our drug cures zombies at a higher rate than the current state of the art drug.





Next up: How does Bayes work?

FUNDAMENTALS OF BAYESIAN DATA ANALYSIS IN R