Bayesian Statistics

Overview of Bayesian Stats

- Why bother?
- Principles of Probability
- Setting Appropriate Priors
- The Bayes Factor
- BRMs & Stan – The R Code
- Reporting Your Results

Why Bother?

The classic example:
Do you believe in ESP?

Why Bother?

The classic example:
Do you believe in ESP?

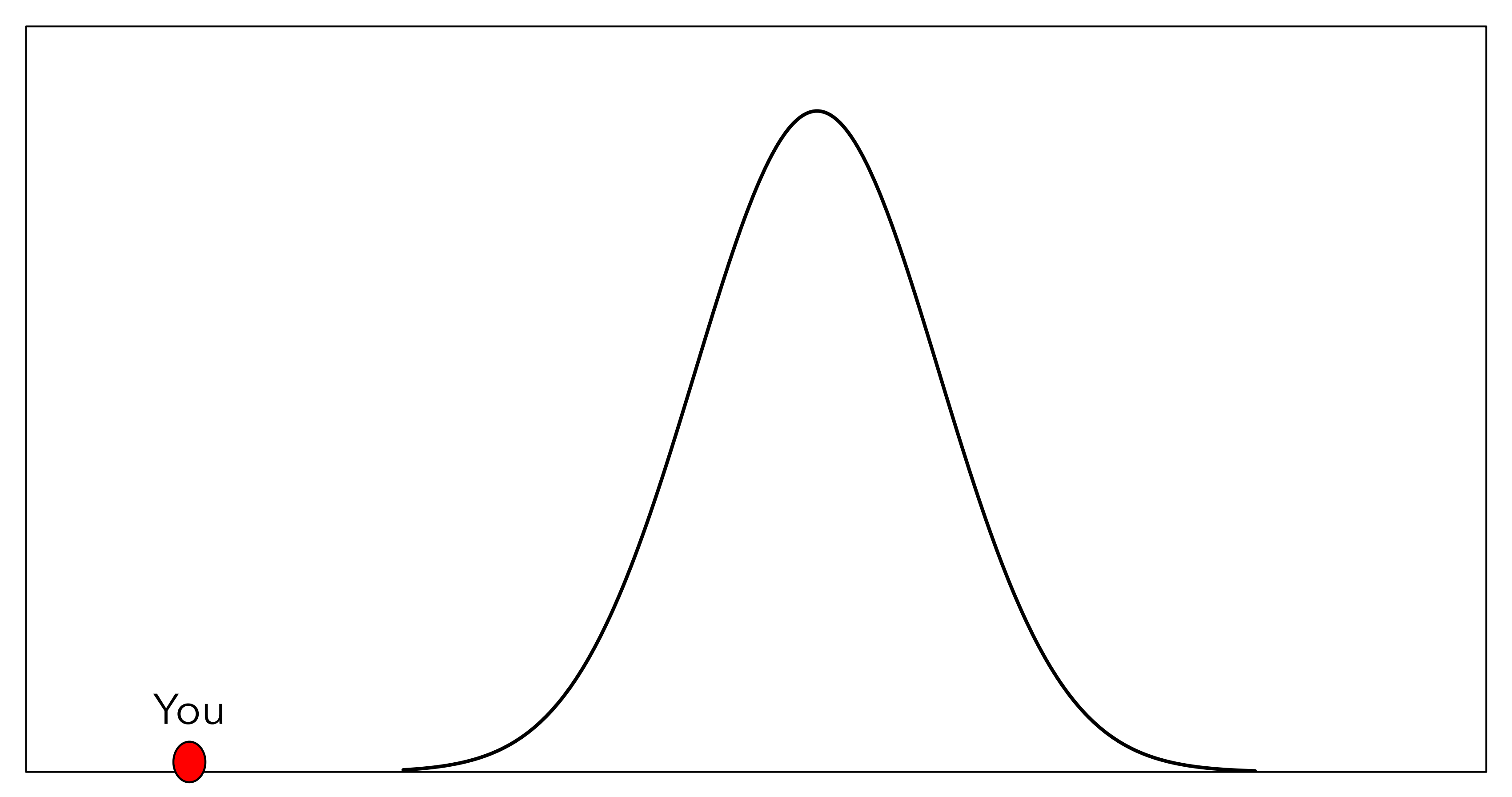
Real Slate Headline:

Daryl Bem Proved ESP Is Real

Which means science is broken.

Why Bother?

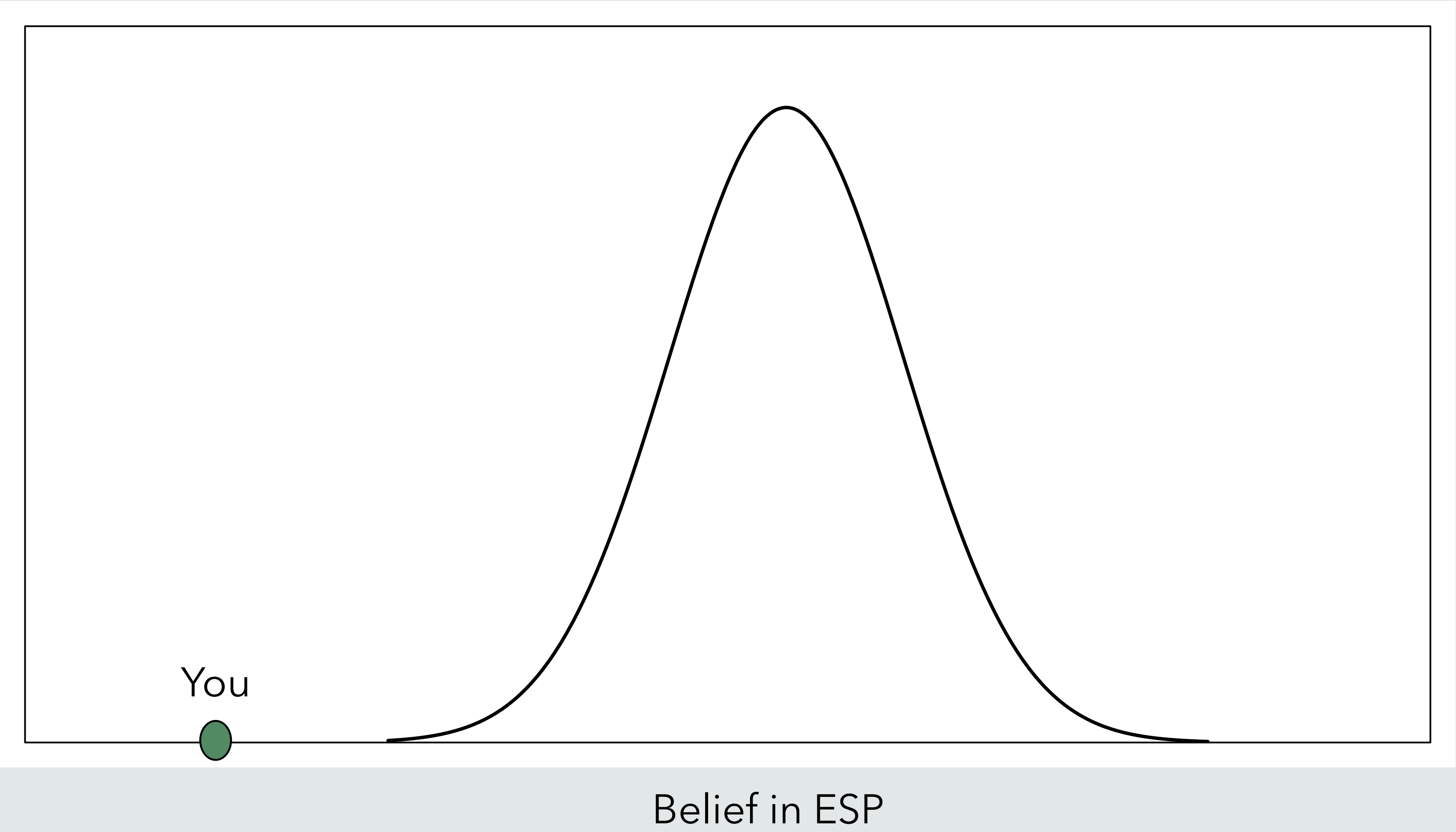
The classic example: Do you believe in ESP



Belief in ESP

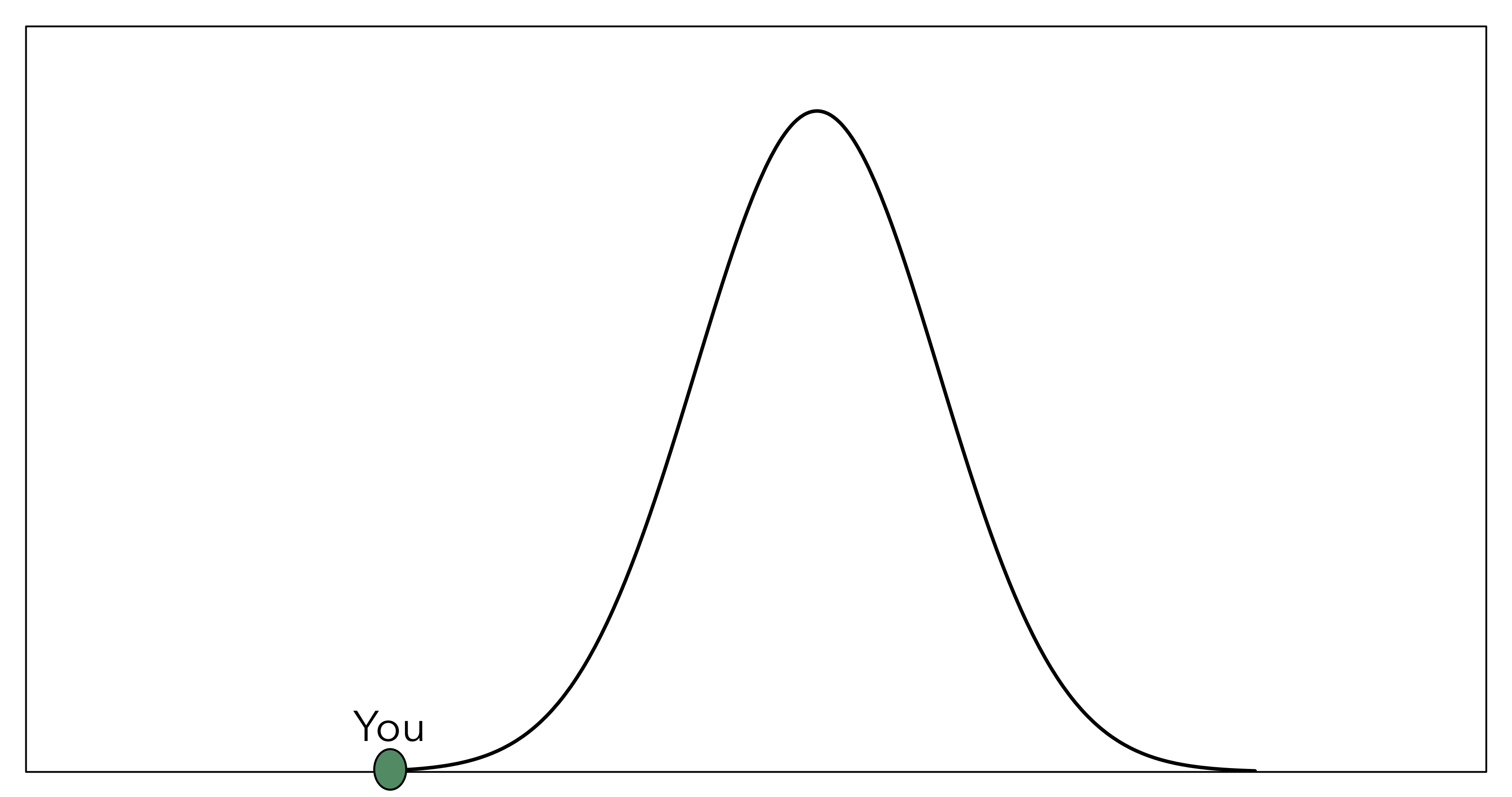
Why Bother?

The classic example: 1 article



Why Bother?

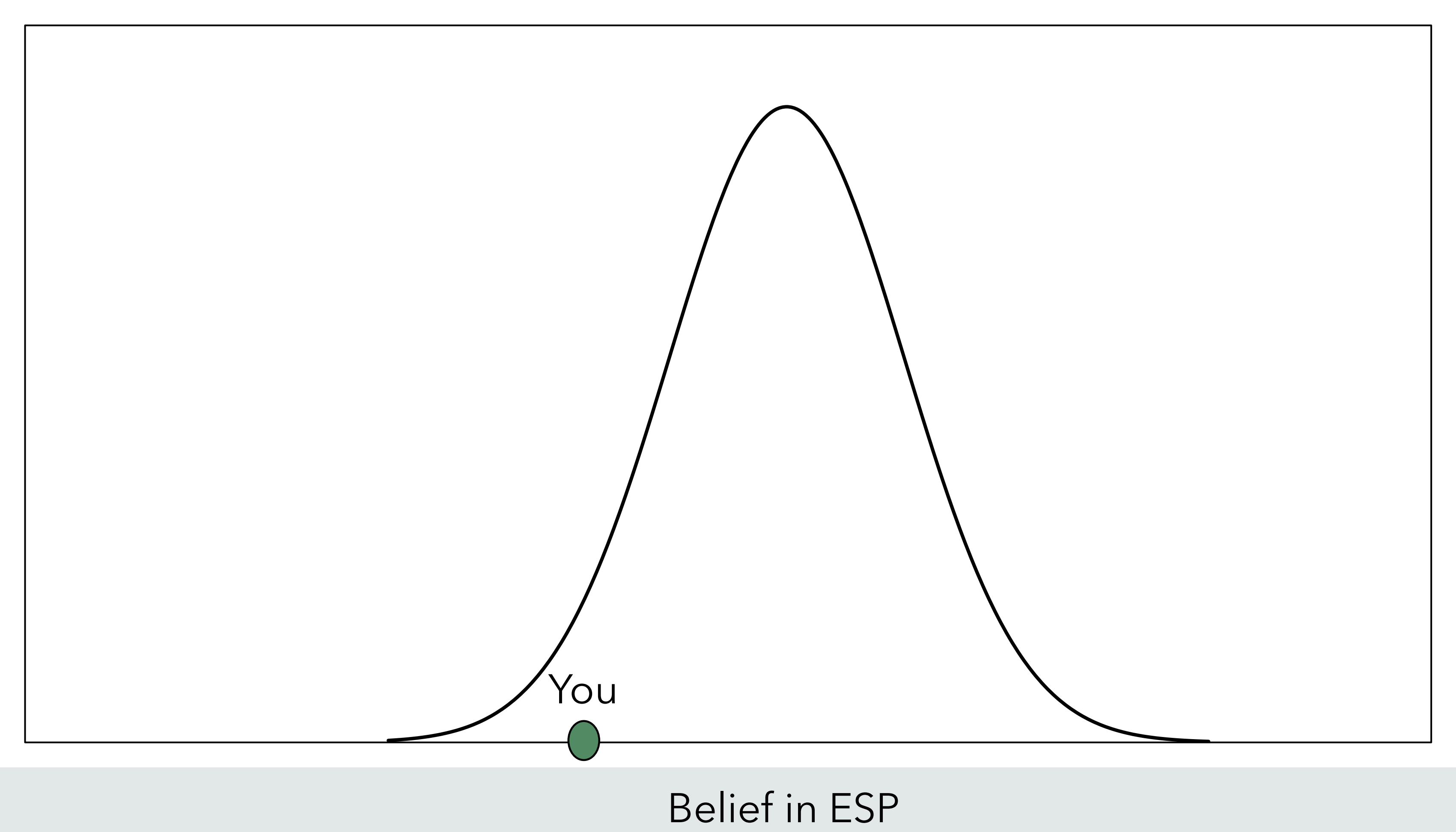
The classic example:
100 article



Belief in ESP

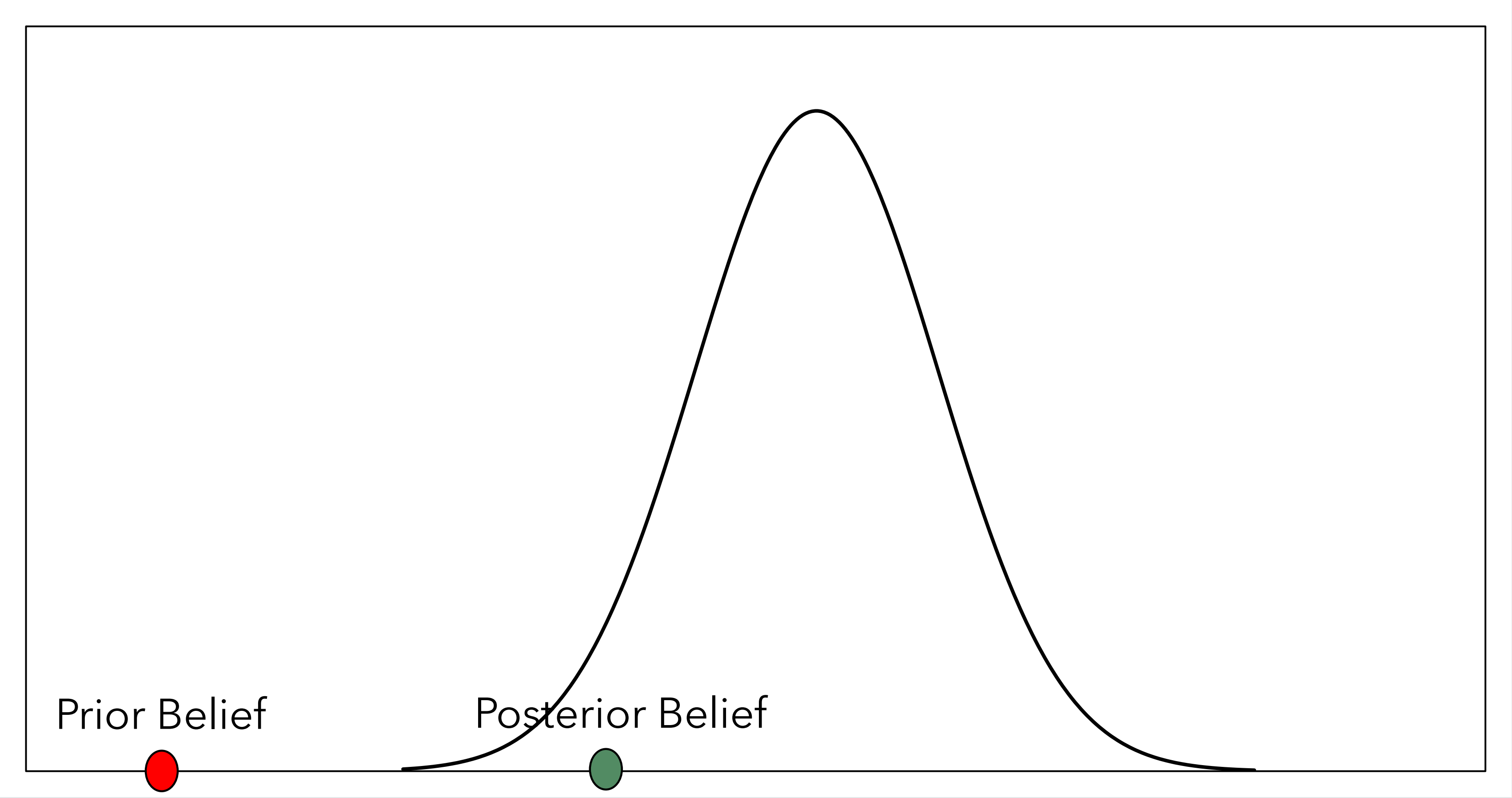
Why Bother?

The classic example: Neil deGrasse Tyson



Why Bother?

The classic example: Neil deGrasse Tyson



Belief in ESP

Why Bother?

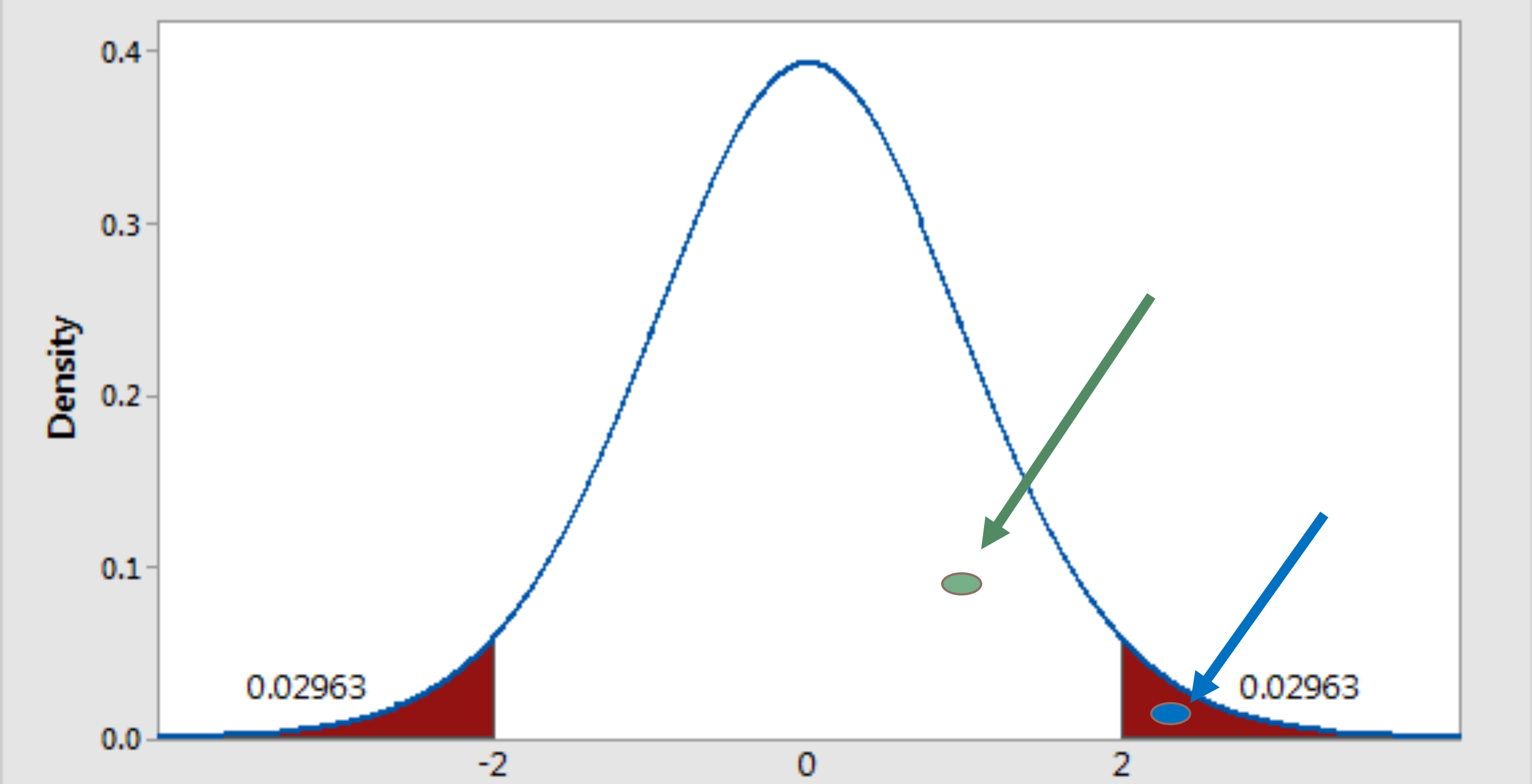
- Allows us to account for a wealth of prior literature
- Allows for proving the null

    What is a p-value?

    - The probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis correct.

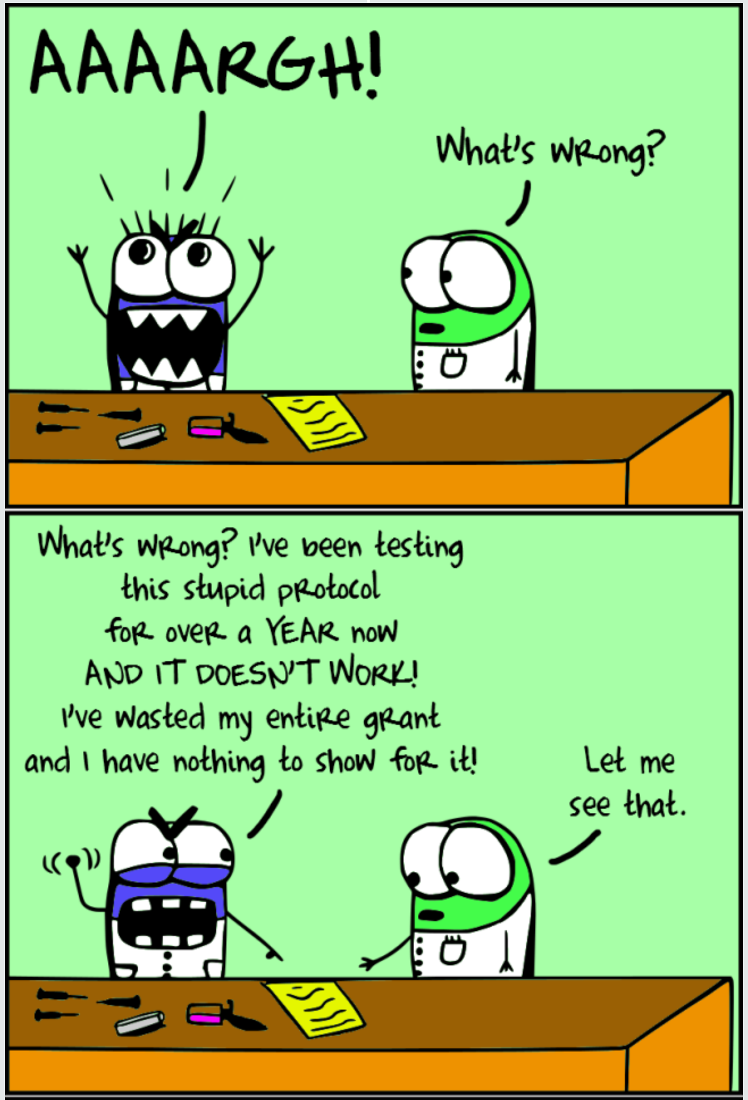
Why Bother?

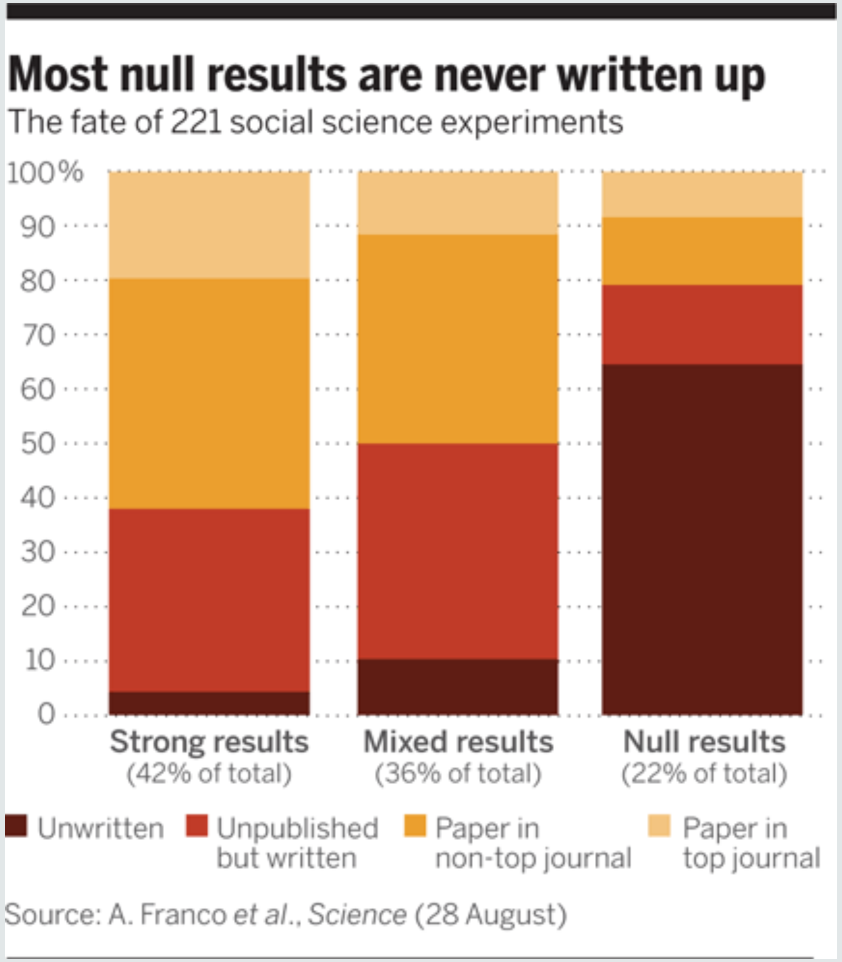
p-values



Why Bother?

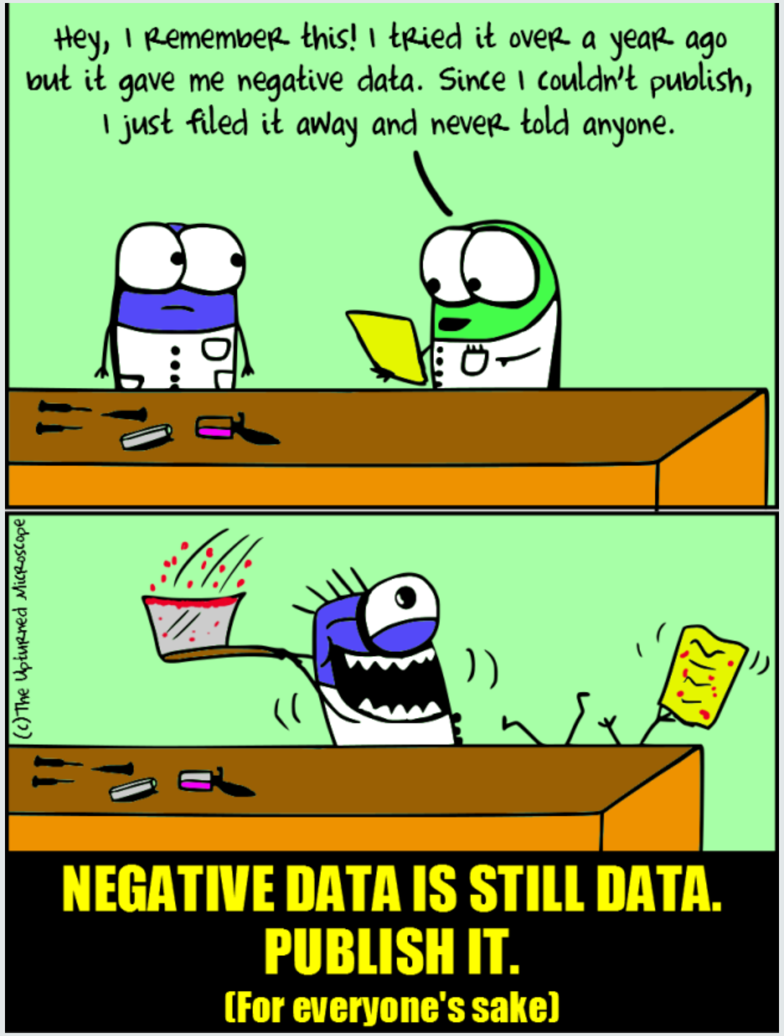
p-values

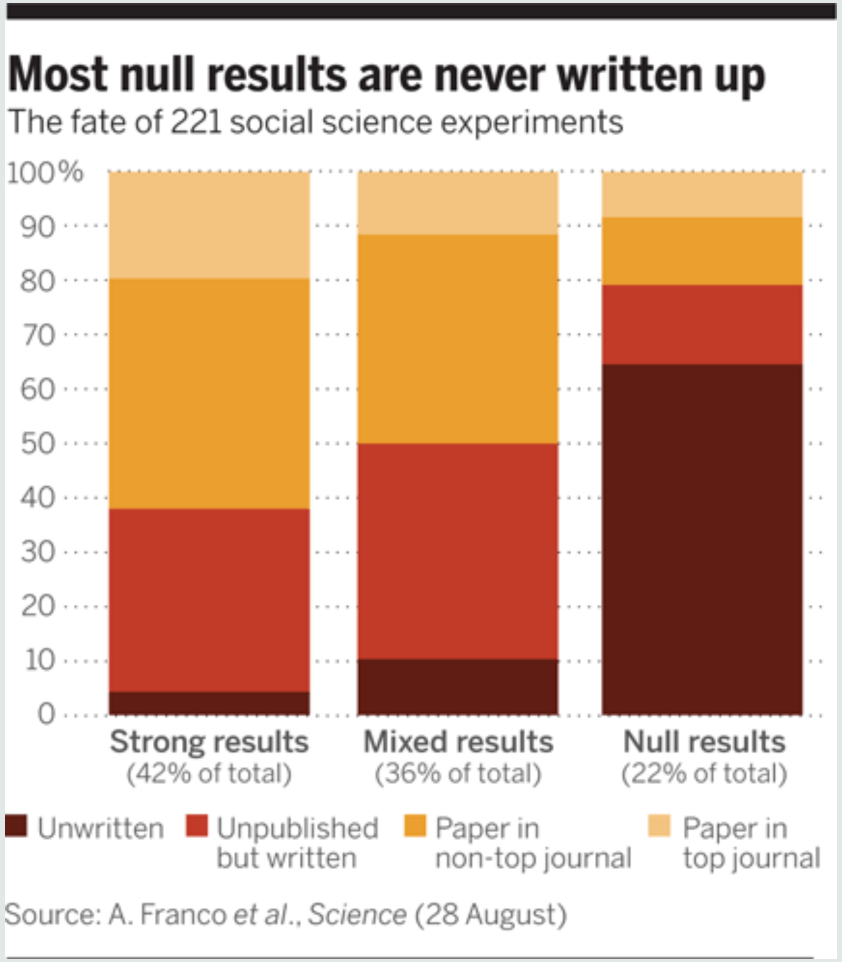




Why Bother?

p-values





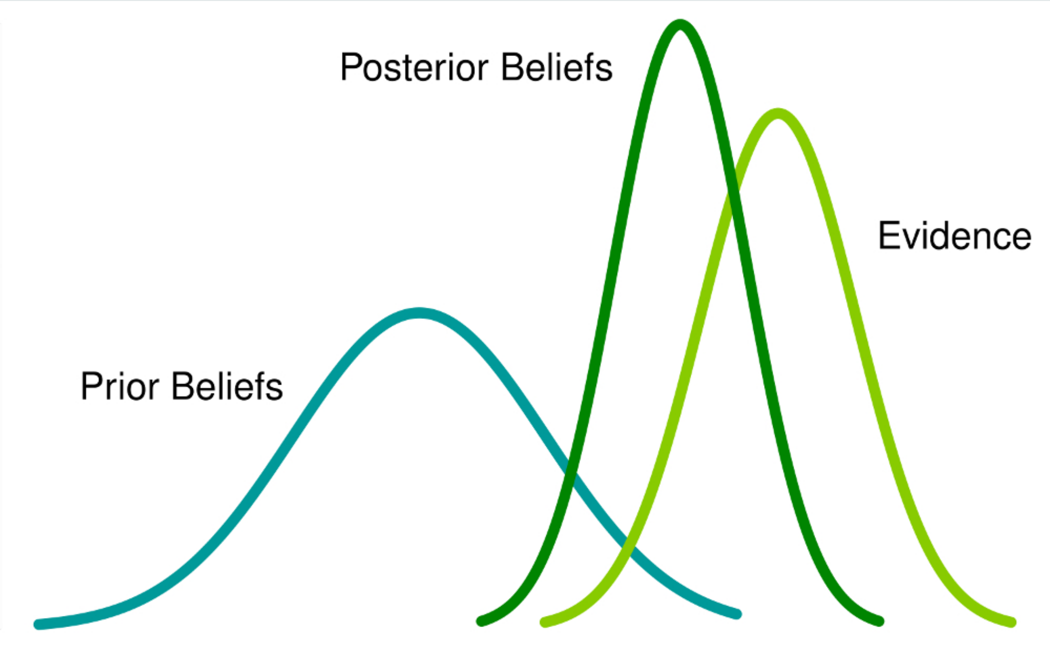
Why Bother?

Proving the null

- Given the null hypothesis, how likely is my data
- Given the data, which hypothesis is more likely (the null or the alternative)

$$y = b0$$

$$y = b0 + b1$$



Why Bother?

- Allows us to account for a wealth of prior literature
- Allows for proving the null

    What is a p-value?

    - The probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis correct.
- More reliable results with less data

Why Bother?

Few Observations

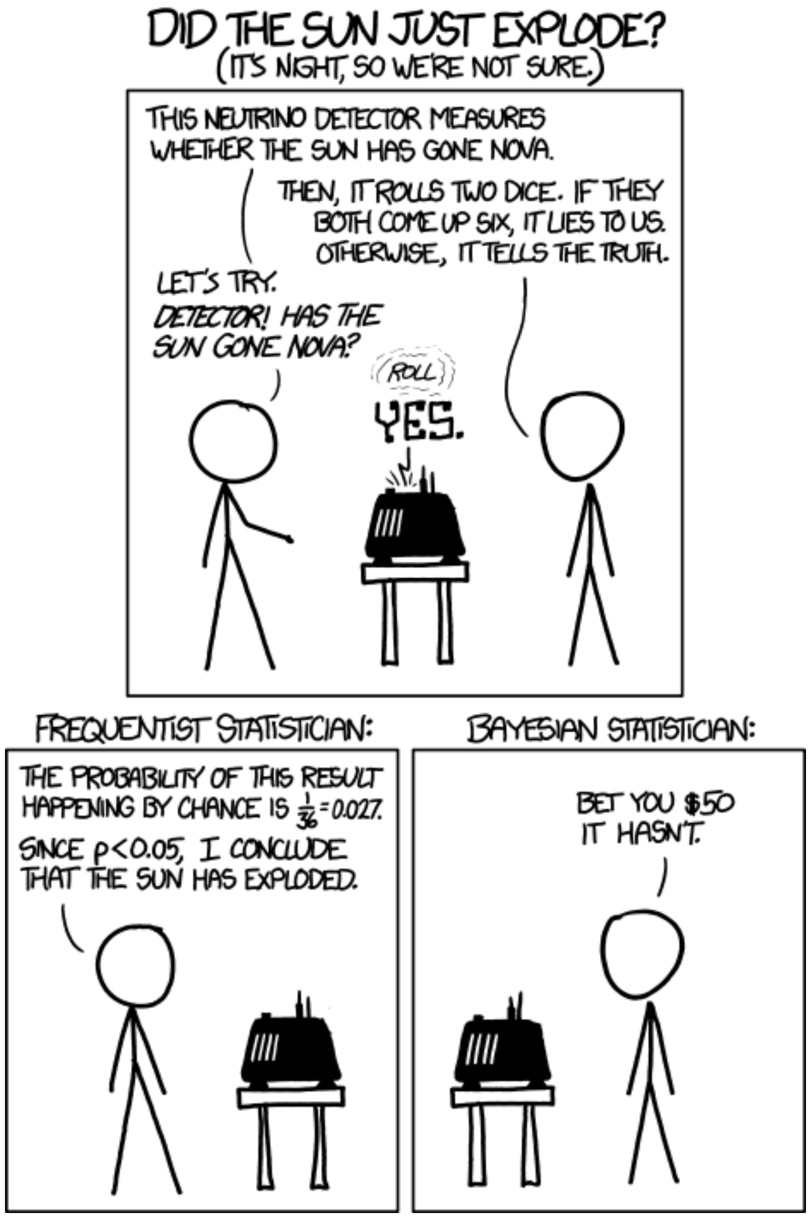
# of Coin Tosses	# of heads	% of heads	Heads/Tails are not equal
2	0	0%	undetermined
10	2	20%	p = .05
20	13	65%	p = .19
100	46	46%	p = .43

Why Bother?

Few Observations

# of Coin Tosses	# of heads	% of heads	Heads/Tails are not equal
2	0	0%	undetermined
10	2	20%	p = .05
20	13	65%	p = .19
100	46	46%	p = .43

# of Coin Tosses	# of heads	% of heads
1002	500	49.9%
1010	502	49.7%
1020	513	50.29%
1100	546	49.64%


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.

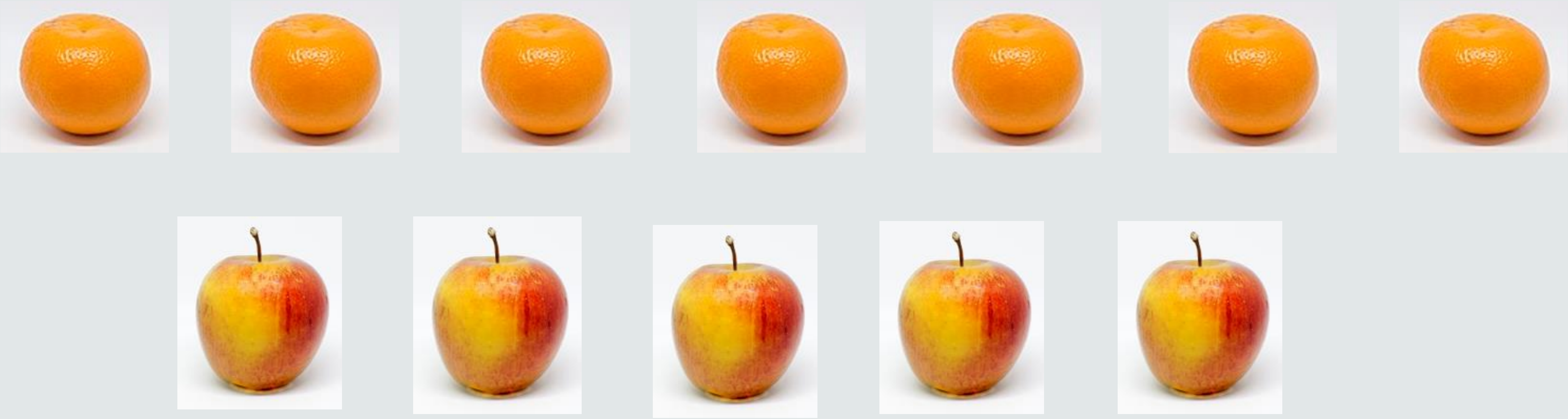
Why Bother?

Overview of Bayesian Stats

- ~~Why bother?~~
- Principles of Probability
- Setting Appropriate Priors
- The Bayes Factor
- BRMs & Stan – The R Code
- Reporting Your Results

Principles of Probability

- What is the probability of randomly selecting an orange?



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

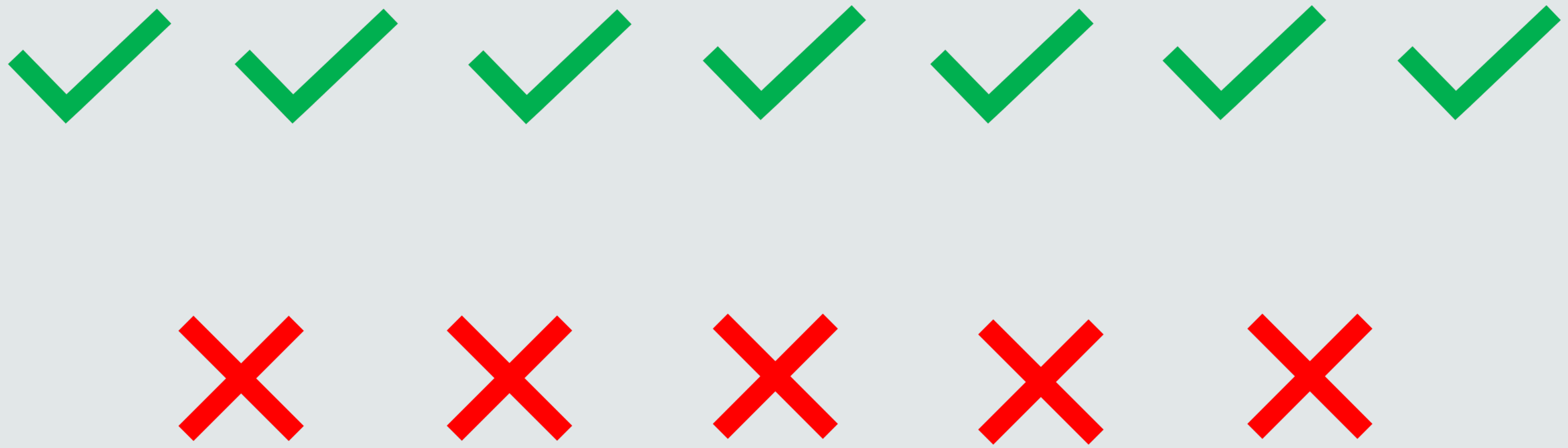
- What is the probability of randomly selecting an orange?

$$\frac{7}{12}$$

$$p(\text{orange}) = \frac{7}{12}$$

Principles of Probability

- What is the probability of randomly getting an event?



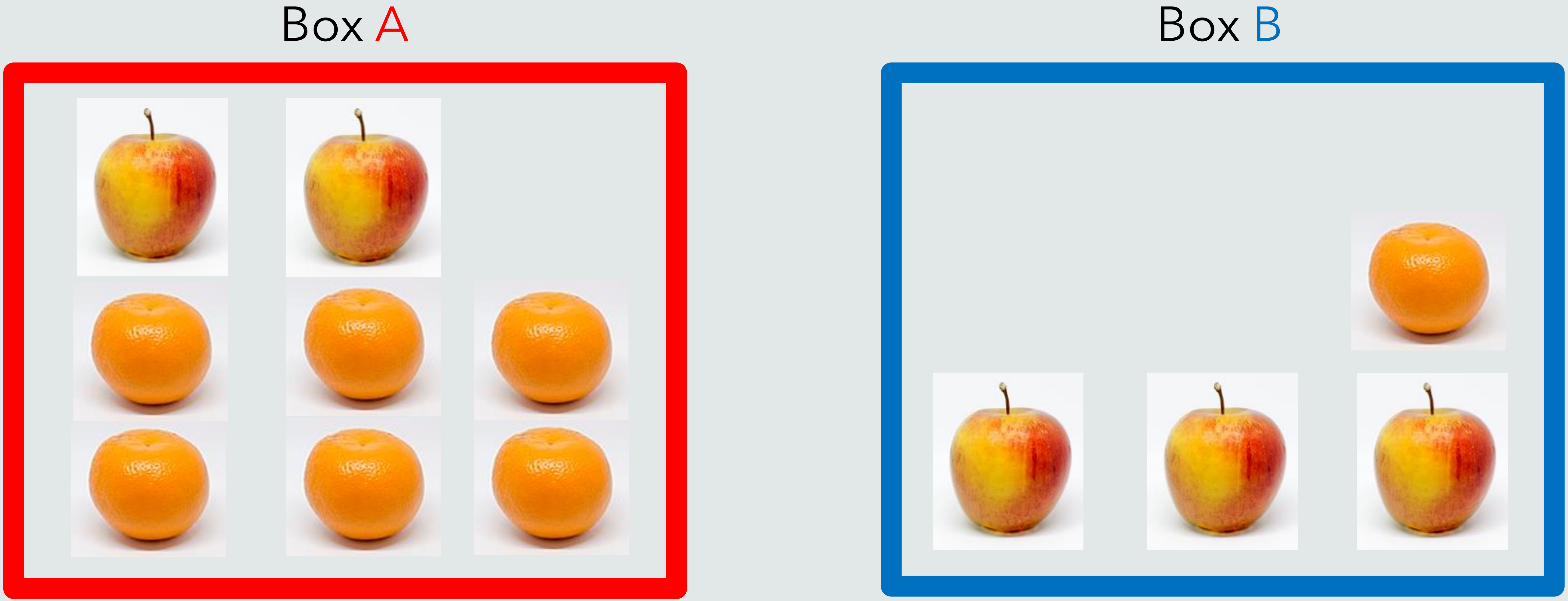
$$\frac{7}{12}$$

$$p(\text{event}) = \frac{7}{12}$$

Principles of Probability

$$p(\text{Box A}) = \frac{8}{12}$$

- What is the probability of randomly selecting fruit from Box A?



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

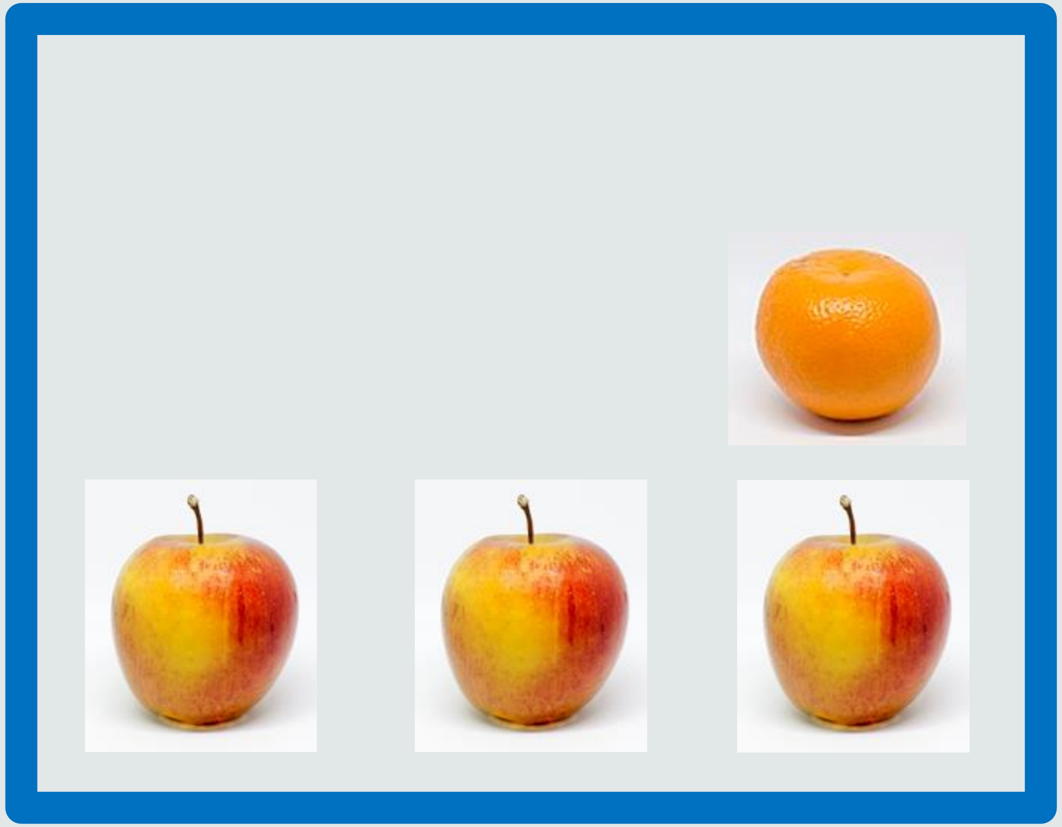
$$p(\text{Box A}) = \frac{8}{12}$$

- What is the probability of randomly selecting a piece of fruit from a Box A?

Marginal Probability

Box A

Box B



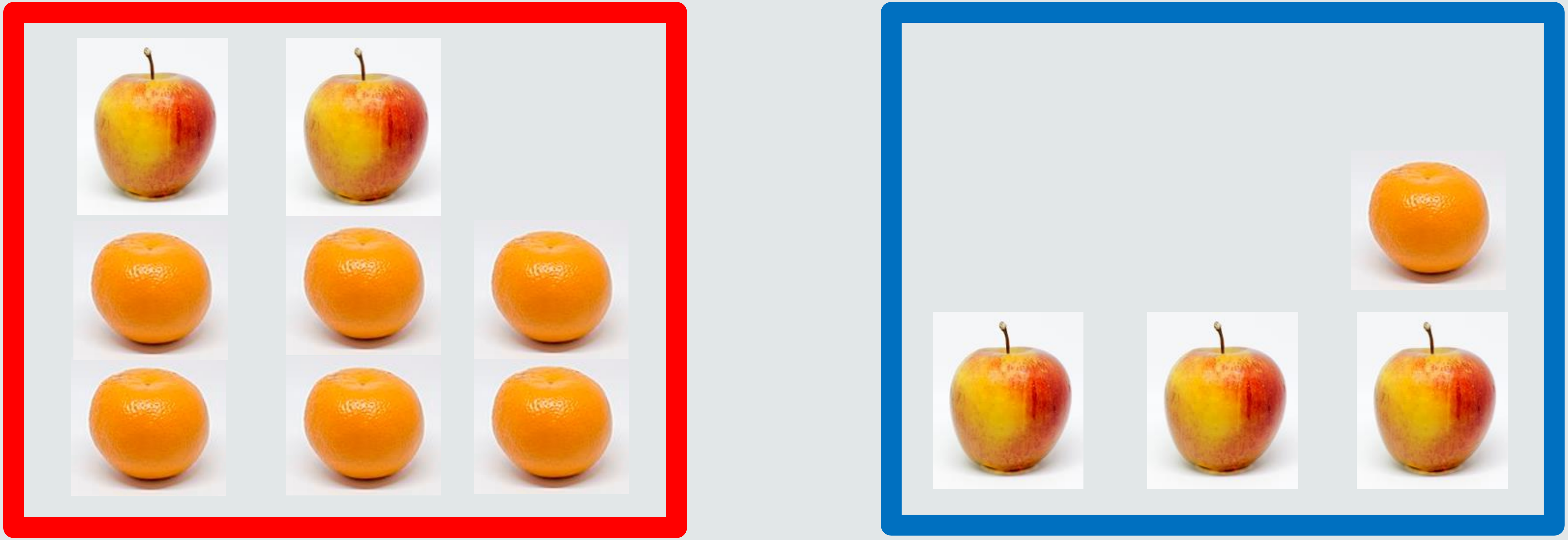
Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

$$p(\text{Orange}|\text{Box A}) = \frac{6}{8} = \frac{3}{4}$$

- What is the probability of randomly selecting an orange given Box A?

Box A   Box B



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

$$p(\text{Orange}|\text{Box A}) = \frac{6}{8}$$

- What is the probability of randomly selecting an orange given Box A?

Box A

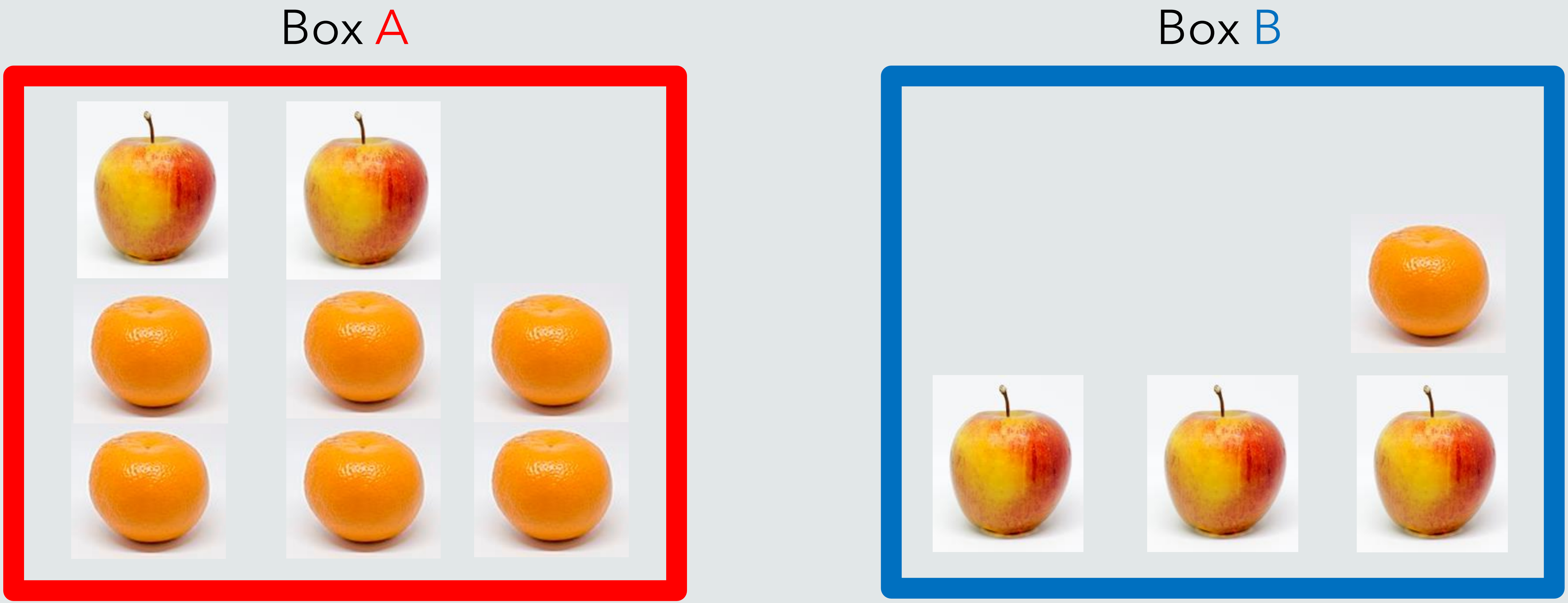
Conditional Probability

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

$$p(\text{Orange}|\text{Box B}) = \frac{1}{4}$$

- What is the probability of randomly selecting an orange given Box B?



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

- What is the conditional probability of each combination?

Fruit	A	B
Apple	2/8	3/4
Orange	6/8	1/4

Box

Principles of Probability

- What is the **joint probability** of each combination?

Fruit		A	B
	Apple	p(Apple & Box A) 2/	p(Apple & Box B) 3/
	Orange	p(Orange & Box A) 6/	p(Orange & Box B) 1/

Box

Principles of Probability

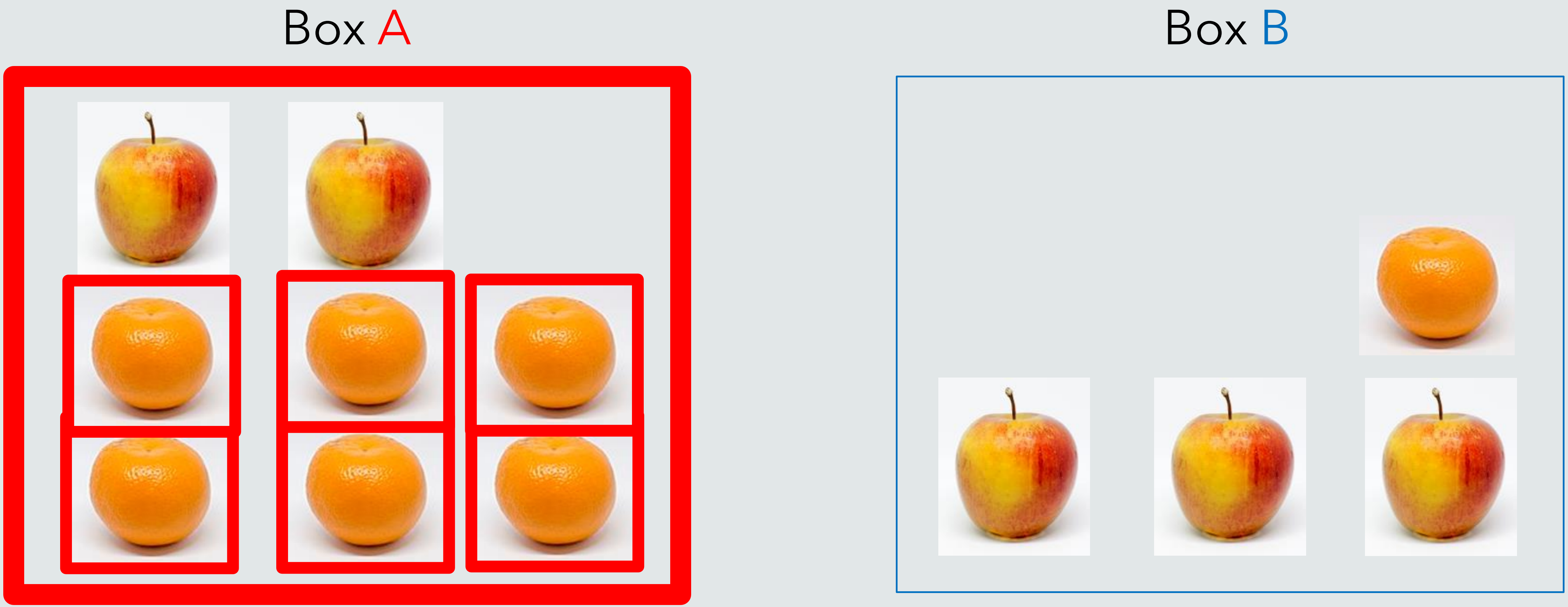
- What is the **joint probability** of each combination?

Fruit	A	B
Apple	p(Apple & Box A) 2/12	p(Apple & Box B) 3/12
Orange	p(Orange & Box A) 6/12	p(Orange & Box B) 1/12

Box

Principles of Probability

- **Joint Probability: Box A and Orange p(Box A, Orange)** $\frac{6}{12}$



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

- What is the joint probability of each combination?

Fruit		A	B
	Apple	p(Apple , Box A) 2/12	p(Apple , Box B) 3/12
	Orange	p(Orange , Box A) 6/12	p(Orange , Box B) 1/12

Box

Principles of Probability

- Sum Rule of Probability

$$\frac{2}{12} + \frac{6}{12} = \frac{8}{12} \qquad \frac{1}{12} + \frac{3}{12} = \frac{4}{12}$$

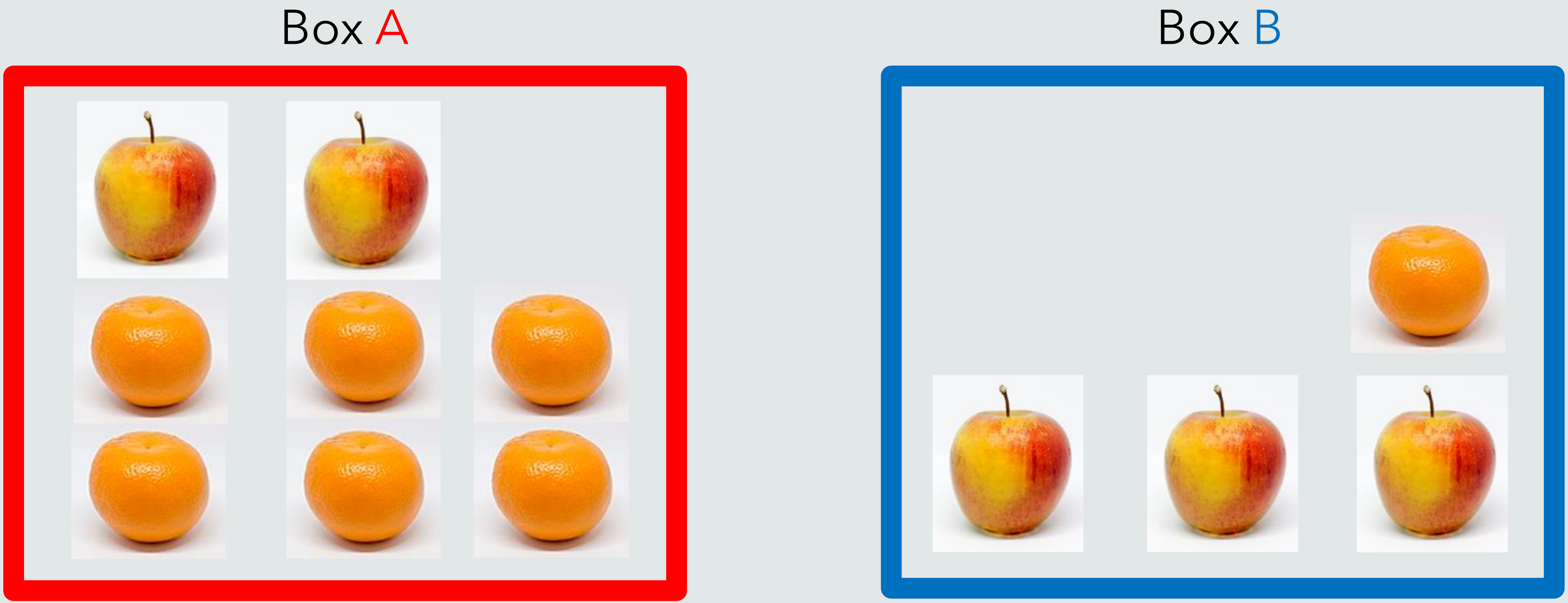
Fruit		A	B
	Apple	2/12	3/12
	Orange	6/12	1/12

Box

Principles of Probability

$$p(\text{Orange}\ ,\ \text{Box A}) = \frac{6}{12}$$

- What is the probability of randomly selecting an Orange & Box A?



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

Box A

$$p(\text{Orange , Box A}) = \frac{6}{12} * \frac{8}{8}$$

How many pieces of fruit in Box A?

Principles of Probability

Box A

$$p(\text{Orange}\ ,\ \text{Box A}) = \frac{6}{12} * \frac{8}{8} = \frac{6}{8} * \frac{8}{12}$$

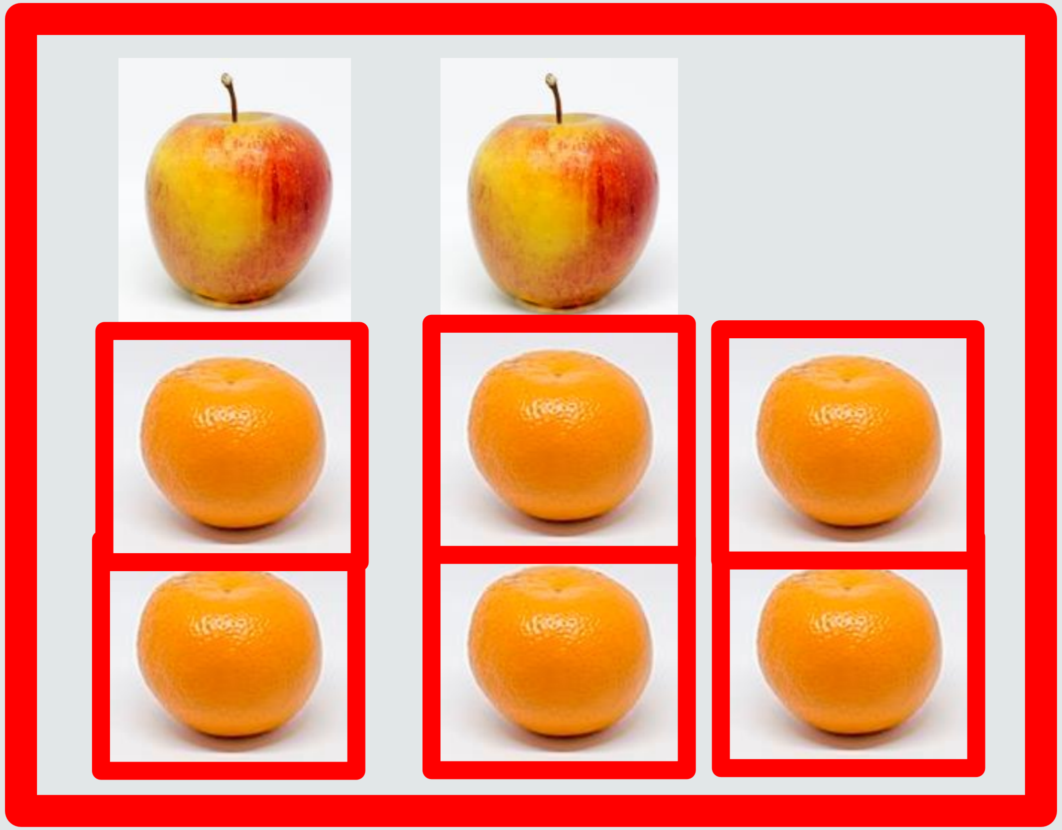
How many oranges in Box A?

How much fruit in Box A?

Principles of Probability

- **Conditional Probability: Orange given Box A p(Orange|Box A)** $\frac{6}{8}$

Box A



Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

- **Marginal Probability: p(Box A)** $= \frac{8}{12}$

Box A

Box B

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

Product Rule:
Joint probability is the product of the conditional and marginal probabilities:

$$p(\text{Orange}, \text{Box A}) = \frac{6}{12} * \frac{8}{8} = \boxed{\frac{6}{8}} * \boxed{\frac{8}{12}}$$

$$\underbrace{p(\text{Orange}, \text{Box A})}_{\textbf{Joint Probability}} = \underbrace{\boxed{p(\text{Orange} \mid \text{Box A})}}_{\textbf{Conditional Probability}} \times \underbrace{\boxed{p(\text{Box A})}}_{\textbf{Marginal Probability}}$$

Principles of Probability

Sum Rule

p(Orange) =p(Box A, Orange) + p(Box B , Orange)

Principles of Probability

Sum Rule

$p(\text{Orange}) = p(\text{Box A, Orange}) + p(\text{Box B , Orange})$

Product Rule

$p(\text{Orange , Box A}) = p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})$

$p(\text{Orange , Box B}) = p(\text{Orange} \mid \text{Box B}) \times p(\text{Box B})$

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

Sum Rule

p(Orange) =p(Box A, Orange) + p(Box B , Orange)

Product Rule

p(Orange , Box A) = p(Orange | Box A) x p(Box A)

p(Orange , Box B) = p(Orange | Box B) x p(Box B)

Combined

***p*(Orange) = *p*(Orange | Box A) x *p*(Box A) + *p*(Orange | Box B) x *p*(Box B)**

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

Combine

p(Orange) = p(Orange | Box A) x p(Box A) + p(Orange | Box B) x p(Box B)

$$p(\text{Orange}) = \frac{6}{8}\ x\ \frac{8}{12} + \frac{1}{4}\ x\ \frac{4}{12} = \frac{6}{12} + \frac{1}{12} = \frac{7}{12}$$

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Principles of Probability

Joint Probabilities

p(Box A, Orange) = p(Orange | Box A) x p(Box A)

p(Orange, Box A) = p(Box A | Orange) x p(Orange)

p(Orange, Box A) = p(Box A, Orange)

***p*(Orange | Box A) x *p*(Box A) = *p*(Box A | Orange) x *p*(Orange)**

Principles of Probability

Joint Probability

$p(\text{Box A, Orange}) = p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})$

$p(\text{Orange, Box A}) = p(\text{Box A} \mid \text{Orange}) \times p(\text{Orange})$

$p(\text{Orange, Box A}) = p(\text{Box A, Orange})$

$p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A}) = \boldsymbol{p(\textbf{Box A} \mid \textbf{Orange})} \times p(\text{Orange})$

$$\boldsymbol{p(\textbf{Box A} \mid \textbf{Orange})} = \frac{p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})}{p(\text{Orange})}$$

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Bayes' Theorem

$$\boldsymbol{p}(\textbf{Box A} \mid \textbf{Orange}) = \frac{p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})}{p(\text{Orange})}$$

Principles of Probability

Bayes' Theorem

$$p(\text{Box A} \mid \text{Orange}) = \frac{p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})}{p(\text{Orange})}$$

$$\boldsymbol{p(\text{prior} \mid \text{data}) = \frac{p(\text{data} \mid \text{prior}) \times p(\text{prior})}{p(\text{data})}}$$

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Bayes' Theorem

$$p(\text{Box A} \mid \text{Orange}) = \frac{p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})}{p(\text{Orange})}$$

$$p(\text{prior} \mid \text{data}) = \frac{p(\text{data} \mid \text{prior}) \times p(\text{prior})}{p(\text{data})}$$

$$\boldsymbol{p(\textbf{prior} \mid \textbf{evidence}) = \frac{\textbf{likelihood} \times p(\textbf{prior})}{p(\textbf{evidence})}}$$

Principles of Probability

Bayes' Theorem

$$p(\text{Box A} \mid \text{Orange}) = \frac{p(\text{Orange} \mid \text{Box A}) \times p(\text{Box A})}{p(\text{Orange})}$$

$$p(\text{prior} \mid \text{data}) = \frac{p(\text{data} \mid \text{prior}) \times p(\text{prior})}{p(\text{data})}$$

$$p(\text{prior} \mid \text{evidence}) = \frac{\text{likelihood} \times p(\text{prior})}{p(\text{evidence})}$$

***Posterior* ∝ likelihood x *p*(prior)**

Credit Joseph Yurko, Machine Learning, Information Sciences 2595

Overview of Bayesian Stats

- ~~Why bother?~~
- ~~Principles of Probability~~
- Setting Appropriate Priors
- The Bayes Factor
- BRMs & Stan – The R Code
- Reporting Your Results

Setting Appropriate Priors

- Informative: Strong Beliefs
- Weakly informative: Weak Beliefs
- Uninformative/Diffuse/Flat: No Beliefs (every possibility is equally probable)
- Improper: The prior doesn’t integrate to 1
- Over the reals: Over the real numbers (not infinity)

Setting Appropriate Priors

- y = vocabulary knowledge (a summed total across a variety of measures)
- x = age
- H_1 = Vocabulary knowledge goes up as age goes up

  One data set rejecting the null with 40 observations

  One data set rejecting the null with 400 observations
- H_0 = Vocabulary knowledge remains the same over the lifespan (unlikely)

  One data set failing to reject the null with 40 observations

  One data set failing to reject the null with 400 observations

Setting Appropriate Priors

lm() output

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 198.0078    41.9681   4.718 3.19e-05 ***
age           0.2625     2.1876   0.120    0.905
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.66 on 38 degrees of freedom
Multiple R-squared:  0.0003787, Adjusted R-squared:  -0.02593
F-statistic: 0.0144 on 1 and 38 DF,  p-value: 0.9051
```

brms() output with flat priors

Bayesian Regression Models using Stan (brms)

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   198.48     41.52   116.21   281.39 1.00     3435     2751
age           0.25      2.18    -4.05     4.60 1.00     3491     2731
```

Setting Appropriate Priors

brms() output with flat priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   198.48     41.52   116.21   281.39 1.00     3435     2751
age           0.25      2.18    -4.05     4.60 1.00     3491     2731
```

brms() prior output

```
                  prior     class coef group resp dpar nlpar bound       source
                 (flat)         b                                       default
                 (flat)         b  age                             (vectorized)
 student_t(3, 201.1, 44.9) Intercept                                    default
     student_t(3, 0, 44.9)     sigma                                    default
```

Setting Appropriate Priors

diffuse <- c(set_prior("normal(5,100)", class = "b", coef = "age"),

set_prior("student_t(3,200,2.5)", class="Intercept"))

weak <- set_prior("normal(5,10)", class = "b", coef = "age")

informative <- set_prior("normal(5,1)", class = "b", coef = "age")

hyperinformative <- set_prior("normal(5,0.1)", class = "b", coef = "age")

```
                prior     class coef group resp dpar nlpar bound  source
               (flat)         b                                  default
          normal(5,10)        b  age                                user
student_t(3, 201.1, 44.9) Intercept                              default
    student_t(3, 0, 44.9)     sigma                              default
```

Setting Appropriate Priors

brms() output with flat priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   198.48     41.52   116.21   281.39 1.00     3435     2751
age           0.25      2.18    -4.05     4.60 1.00     3491     2731
```

brms() output with diffuse priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   194.82     42.73   109.94   279.18 1.00     3194     2713
age           0.31      2.25    -4.14     4.77 1.00     3195     2693
```

Setting Appropriate Priors

brms() output with diffuse priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   194.82     42.73   109.94   279.18 1.00     3194     2713
age           0.31      2.25    -4.14     4.77 1.00     3195     2693
```

brms() output with weak priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   194.00     42.30   109.03   274.08 1.00     3272     2568
age           0.47      2.19    -3.72     4.89 1.00     3288     2635
```

Setting Appropriate Priors

brms() output with weak priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   194.00     42.30   109.03   274.08 1.00     3272     2568
age           0.47      2.19    -3.72     4.89 1.00     3288     2635
```

brms() output with informative priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   123.46     18.58    86.99   160.70 1.00     3028     2629
age           4.20      0.91     2.39     5.96 1.00     3120     2645
```

Setting Appropriate Priors

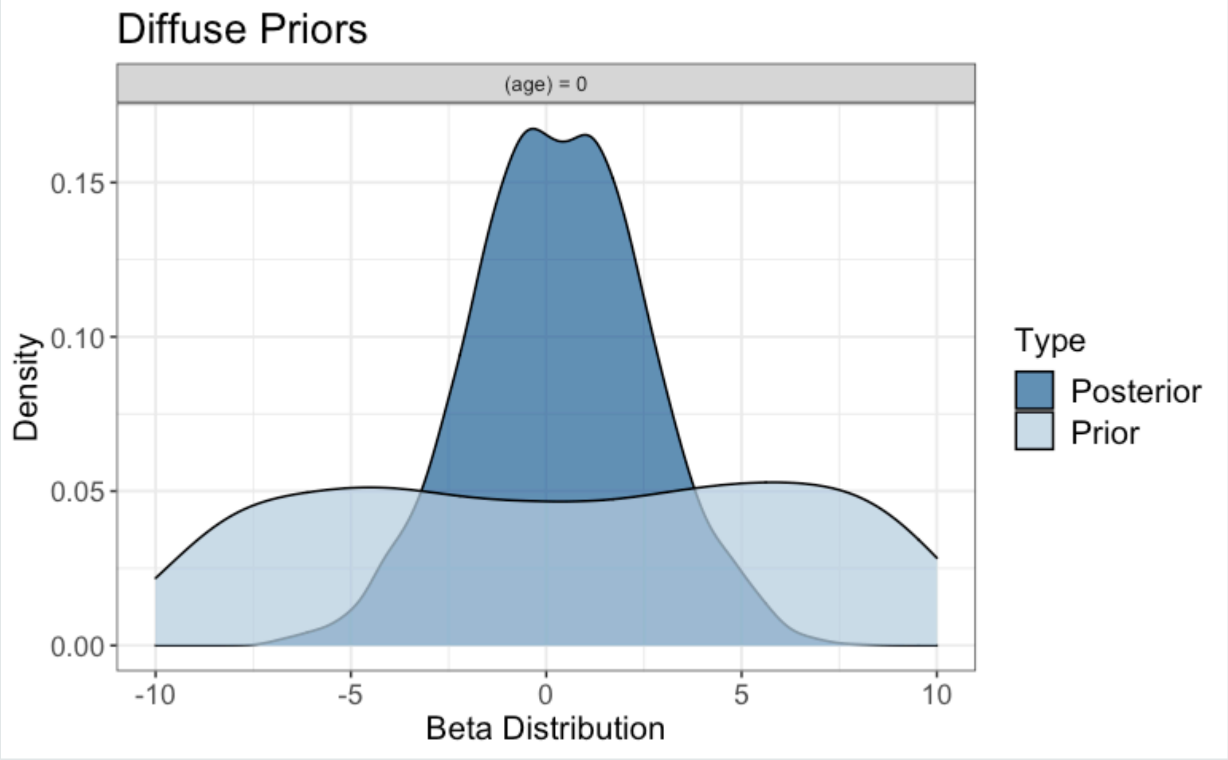
brms() output with informative priors

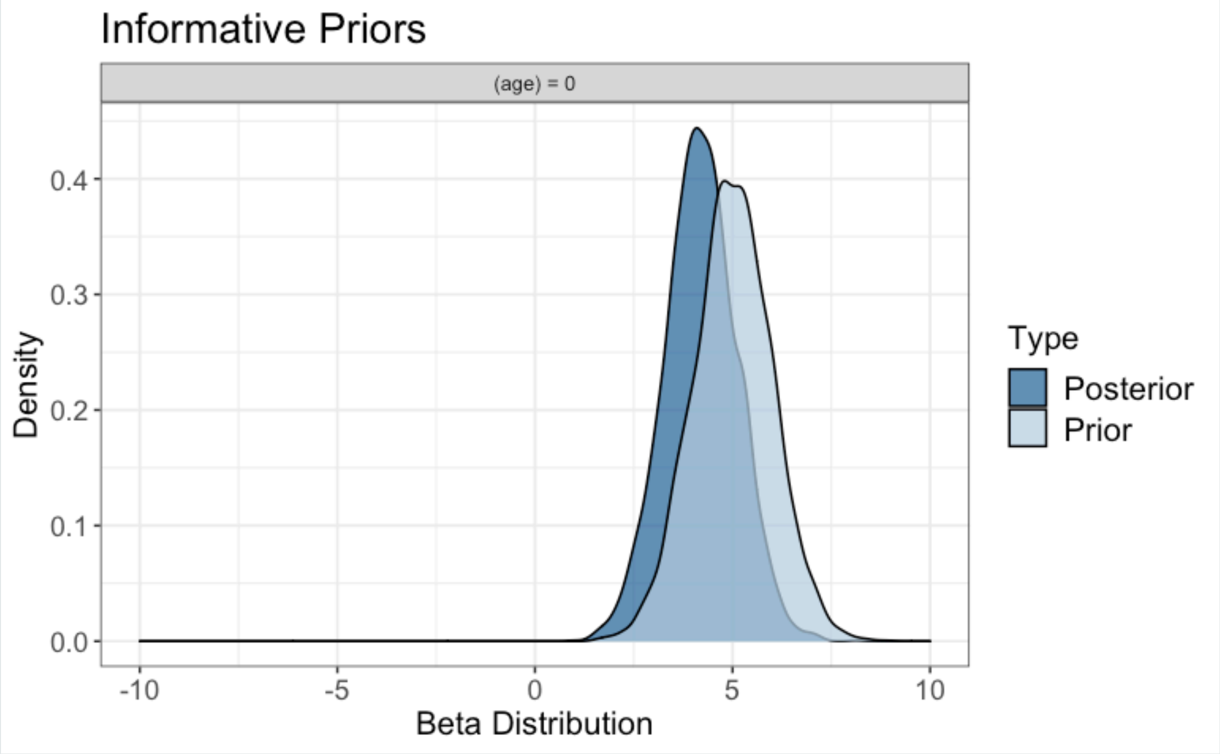
```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   123.46     18.58    86.99   160.70 1.00     3028     2629
age           4.20      0.91     2.39     5.96 1.00     3120     2645
```

brms() output with hyper informative priors

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   108.27      7.13    94.04   122.56 1.00     3632     2858
age           4.99      0.10     4.79     5.17 1.00     4330     2773
```

Setting Appropriate Priors





Setting Appropriate Priors

brms() output with informative priors – 40 observations

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	123.46	18.58	86.99	160.70	1.00	3028	2629
age	4.20	0.91	2.39	5.96	1.00	3120	2645

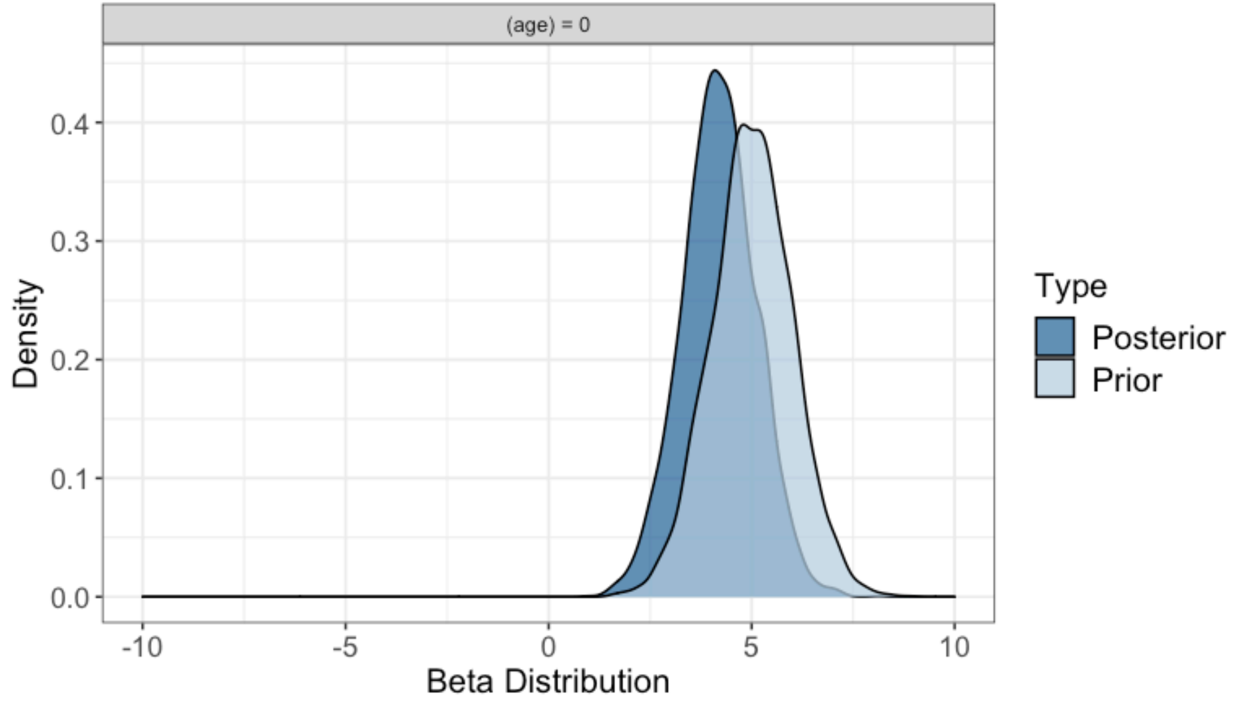
brms() output with informative priors – 400 observations

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	176.07	9.47	157.03	194.22	1.00	4074	3173
age	1.51	0.51	0.53	2.53	1.00	4194	2982

Setting Appropriate Priors

Informative Priors 40 obs

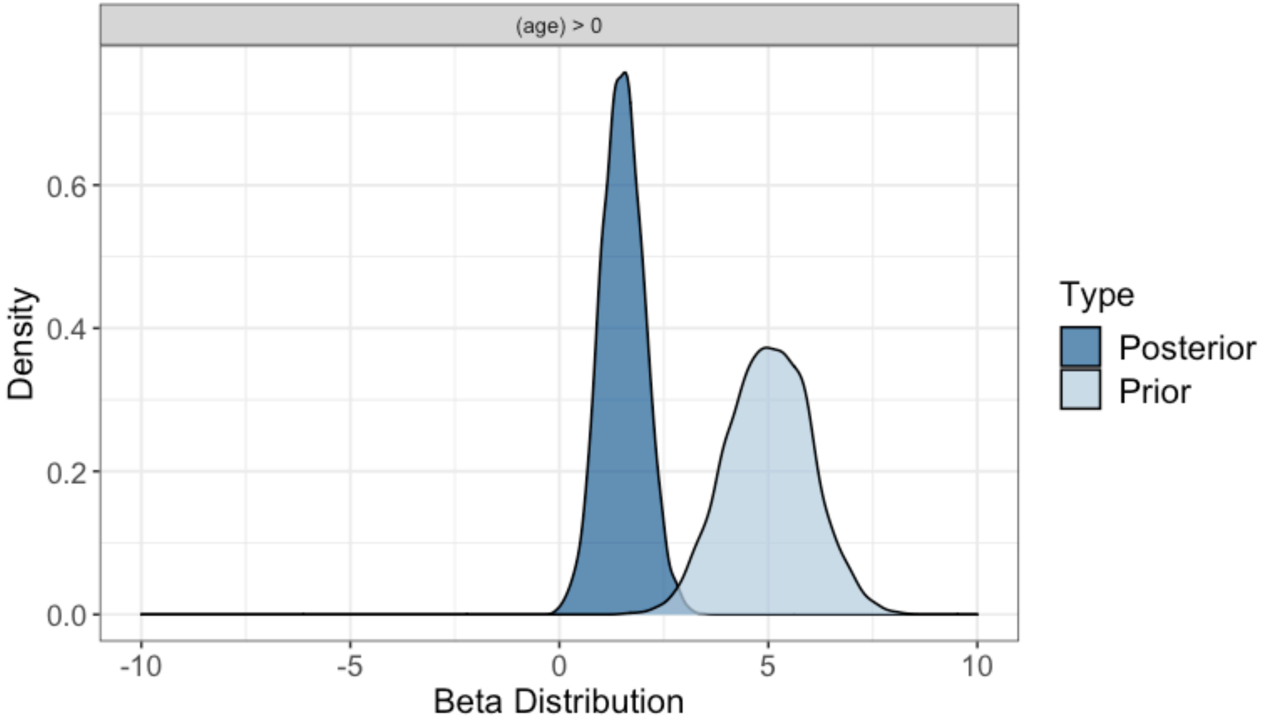


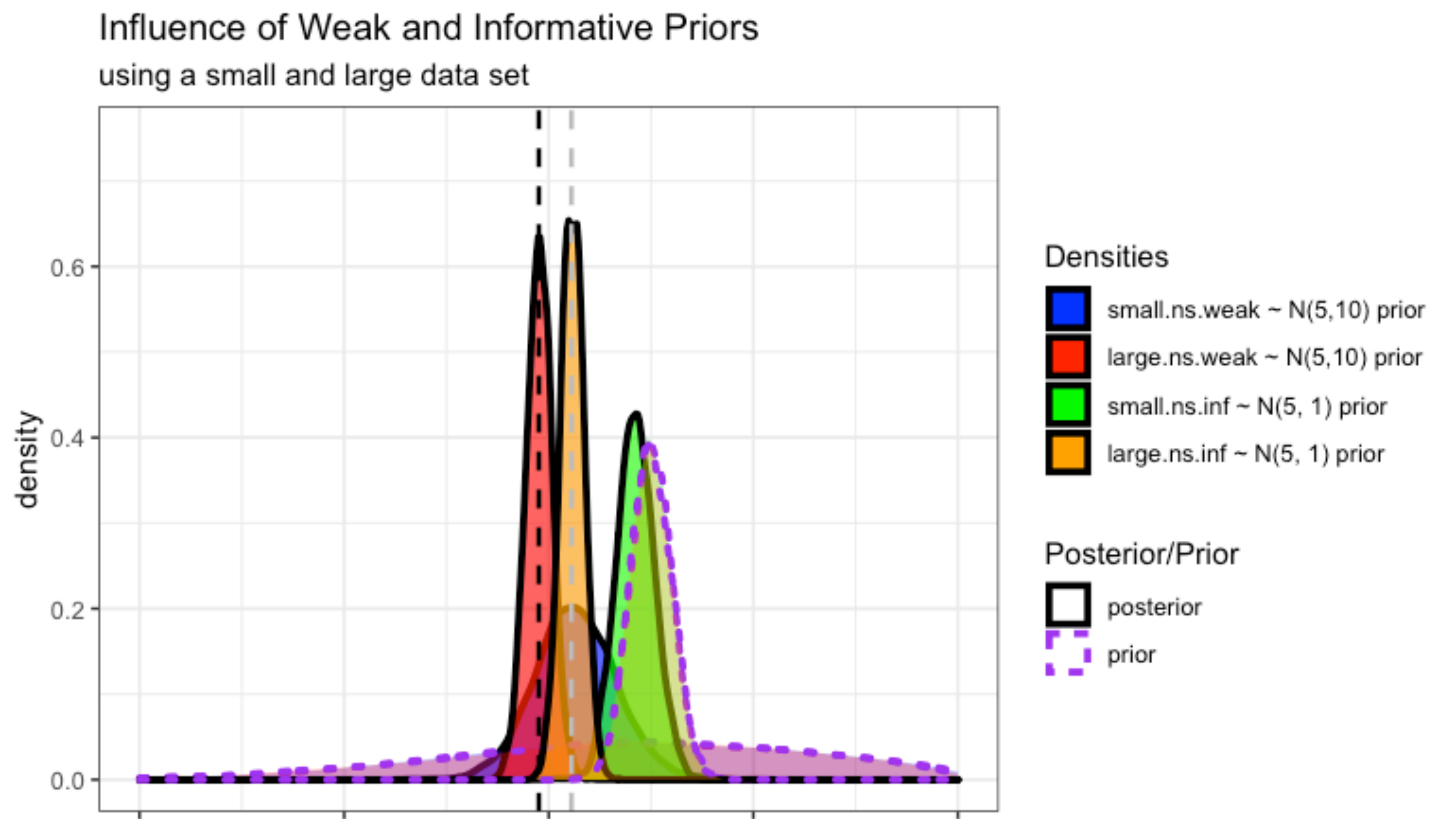
riors – 40 observations

```
Rhat Bulk_ESS Tail_ESS
1.00     3028     2629
```

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI
Intercept   176.07      9.47   157.03   194.22
age           1.51      0.51     0.53     2.53
```

Informative Priors 400 obs





Influence of Weak and Informative Priors
using a small and large data set
density
0.6
0.4
0.2
0.0
Densities
small.ns.weak ~ N(5,10) prior
large.ns.weak ~ N(5,10) prior
small.ns.inf ~ N(5, 1) prior
large.ns.inf ~ N(5, 1) prior
Posterior/Prior
posterior
prior

Setting Appropriate Priors

lm() output – sig data

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1155     0.1060   1.090    0.283
age           0.5284     0.1033   5.115 9.27e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6691 on 38 degrees of freedom
Multiple R-squared:  0.4077,	Adjusted R-squared:  0.3922
F-statistic: 26.16 on 1 and 38 DF,  p-value: 9.271e-06
```

brms() output with flat priors – sig data

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.12      0.11    -0.10     0.33 1.00     3914     2464
age           0.53      0.11     0.31     0.73 1.00     3887     2845
```

Setting Appropriate Priors

brms() output with flat priors – sig data

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.12      0.11    -0.10     0.33 1.00     3914     2464
age           0.53      0.11     0.31     0.73 1.00     3887     2845
```

brms() output with informative priors – sig data

```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.12      0.11    -0.10     0.33 1.00     3496     2878
age           0.58      0.11     0.36     0.80 1.00     3310     2746
```

Setting Appropriate Priors

brms() output with informative priors – sig data

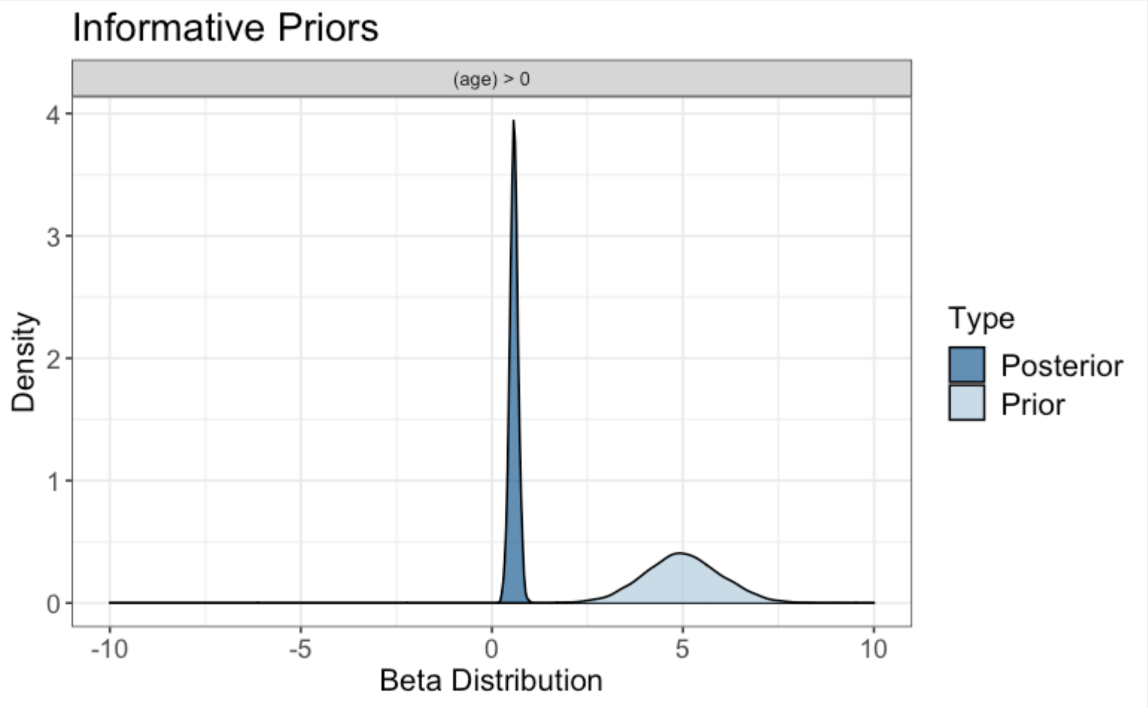
```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.12      0.11    -0.10     0.33 1.00     3496     2878
age           0.58      0.11     0.36     0.80 1.00     3310     2746
```

brms() output with hyper informative priors – sig data

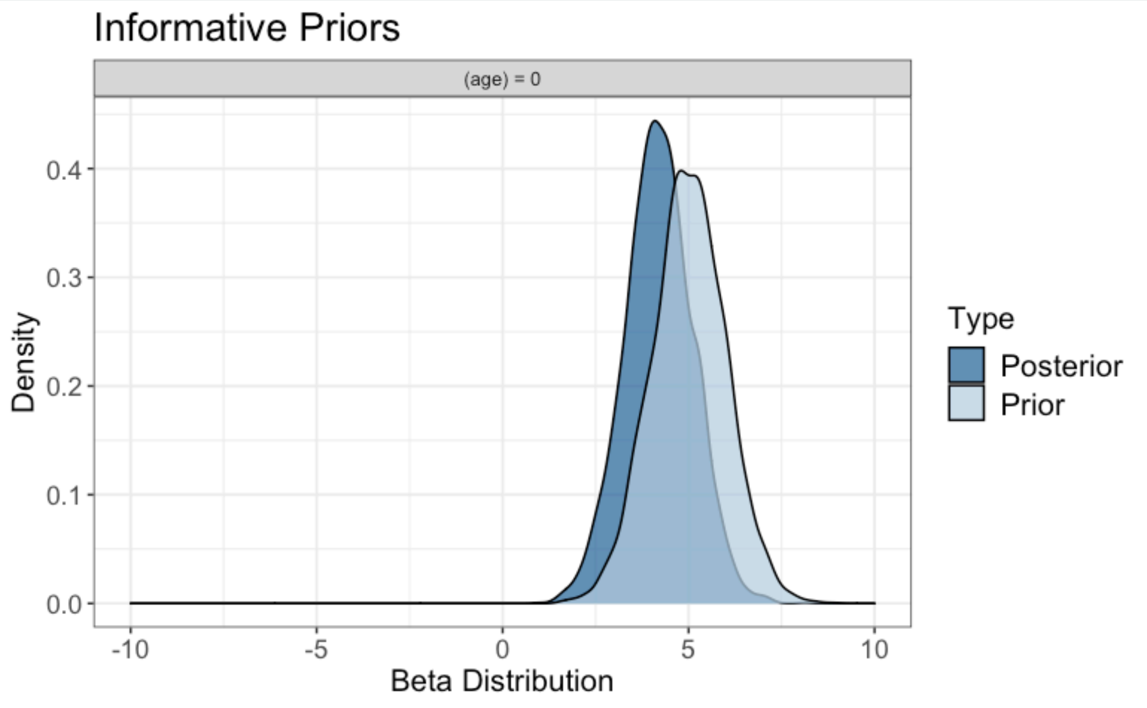
```
Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     0.36      0.69    -1.03     1.74 1.00     3556     2667
age           4.91      0.10     4.71     5.11 1.00     3412     3027
```

Setting Appropriate Priors

Sig Data

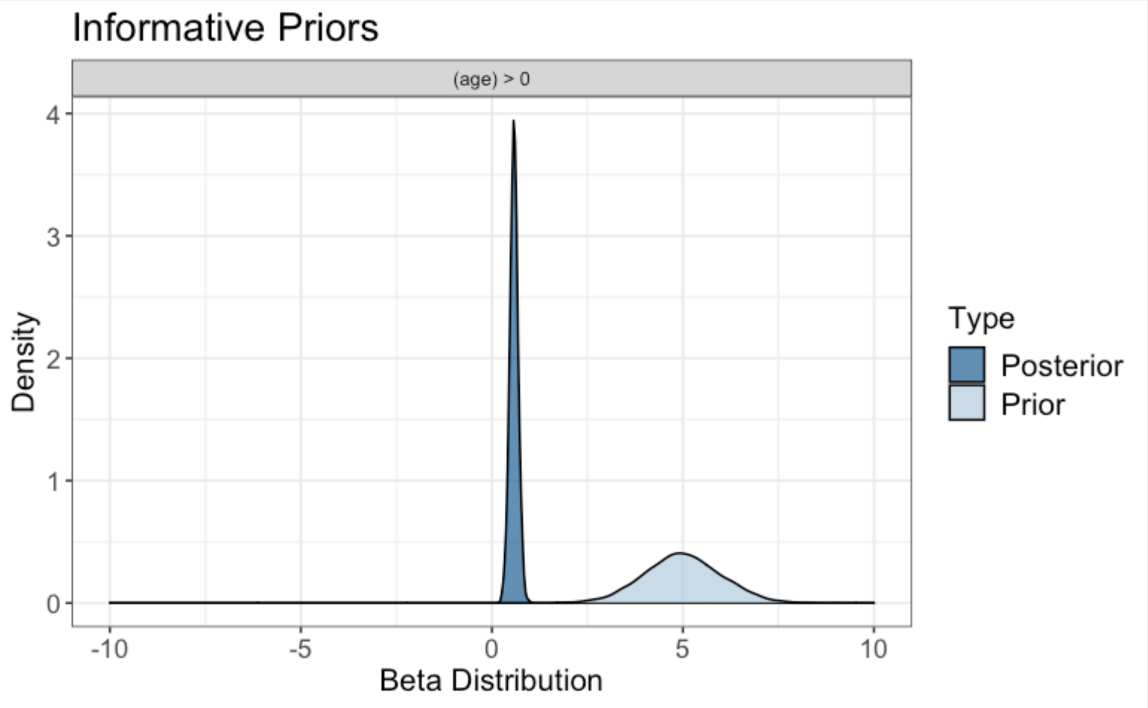


n.s. Data

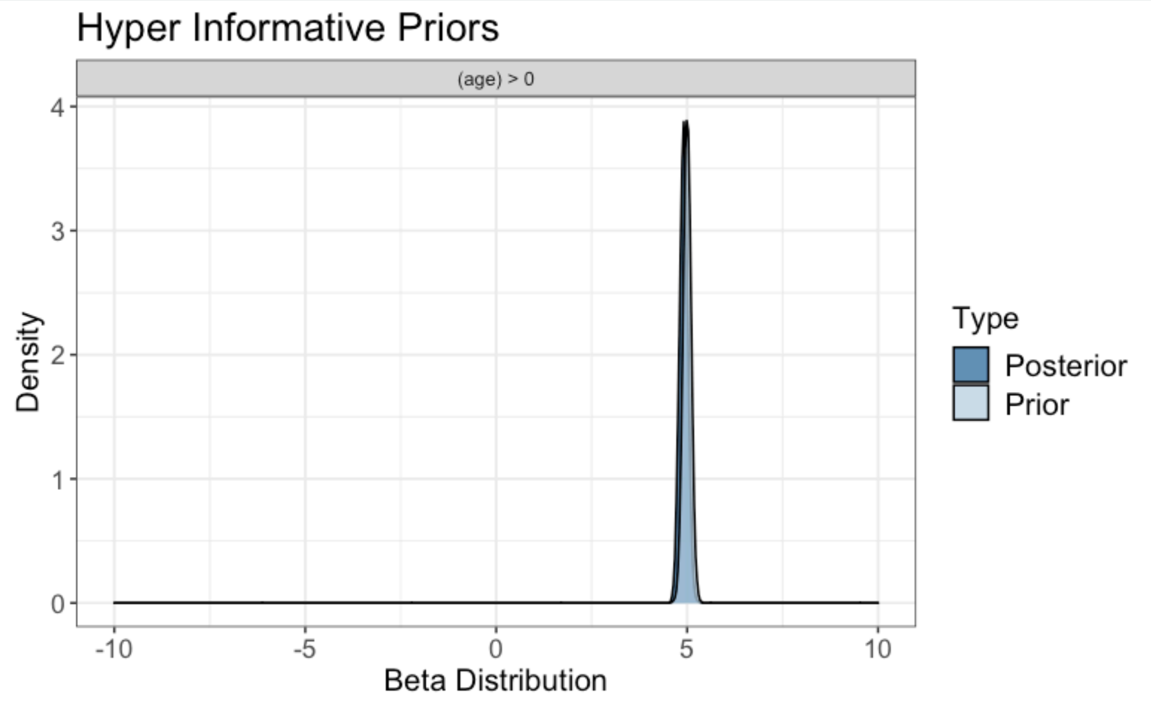


Setting Appropriate Priors

Sig Data



Sig Data



Overview of Bayesian Stats

- ~~Why bother?~~
- ~~Principles of Probability~~
- ~~Setting Appropriate Priors~~
- The Bayes Factor
- BRMs & Stan – The R Code
- Reporting Your Results

The Bayes Factor

Controversy Warning

Bayes Factor

- H_1 = Vocabulary knowledge goes up as age goes up
- H_0 = Vocabulary knowledge remains the same over the lifespan (unlikely)

$$p(\text{prior} \mid \text{data}) = \frac{\boldsymbol{p}\textbf{(data | prior)} \times p(\text{prior})}{p(\text{data})}$$

***p*(data | prior) = likelihood**

$$\frac{likelihood\ of\ the\ data\ given\ H_1}{likelihood\ of\ data\ given\ H_0} = \frac{p(D|\ H_1)}{p(D|\ H_0)}$$

The Bayes Factor

$$\frac{likelihood\ of\ the\ data\ given\ H_1}{likelihood\ of\ data\ given\ H_0} = \frac{p(D|\ H_1)}{p(D|\ H_0)}$$

$$\frac{y = \beta_0 + \beta_{age}x}{y = \beta_0}$$

The Bayes Factor

$$\frac{likelihood\ of\ the\ data\ given\ H_1}{likelihood\ of\ data\ given\ H_0} = \frac{p(D|\ H_1)}{p(D|\ H_0)}$$

$$\frac{y=\beta_0+\beta_{age}x}{y=\beta_0} \text{ --> } \frac{y=\beta_0}{y=\beta_0+\beta_{age}x}$$

The Bayes Factor

$$\frac{likelihood\ of\ the\ data\ given\ H_1}{likelihood\ of\ data\ given\ H_0} = \frac{p(D|\ H_1)}{p(D|\ H_0)}$$

$$\frac{y=\beta_0+\beta_{age}x}{y=\beta_0} \text{ --> } \frac{y=\beta_0}{y=\beta_0+\beta_{age}x}$$

$$\frac{y = \beta_0 + \beta_{age}x * \beta_{SES}x}{y = \beta_0 + \beta_{age}x + \beta_{SES}x}$$

The Bayes Factor

Data Size	Significant?	Prior	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Small	No	Flat	3.23	1/3.23 = .31

The likelihood of the Null is 3.23 times greater than the likelihood of the alternative.

The Bayes Factor

Jeffrey's Levels

Bayes factor	Evidence category
> 100	Extreme evidence for
30 - 100	Very strong evidence for
10 - 30	Strong evidence for
3 - 10	Moderate evidence for
1 - 3	Anecdotal evidence for
1	No evidence

Bayes Factor BF_{10} **Favors the alternative**
1/3.23 = .31

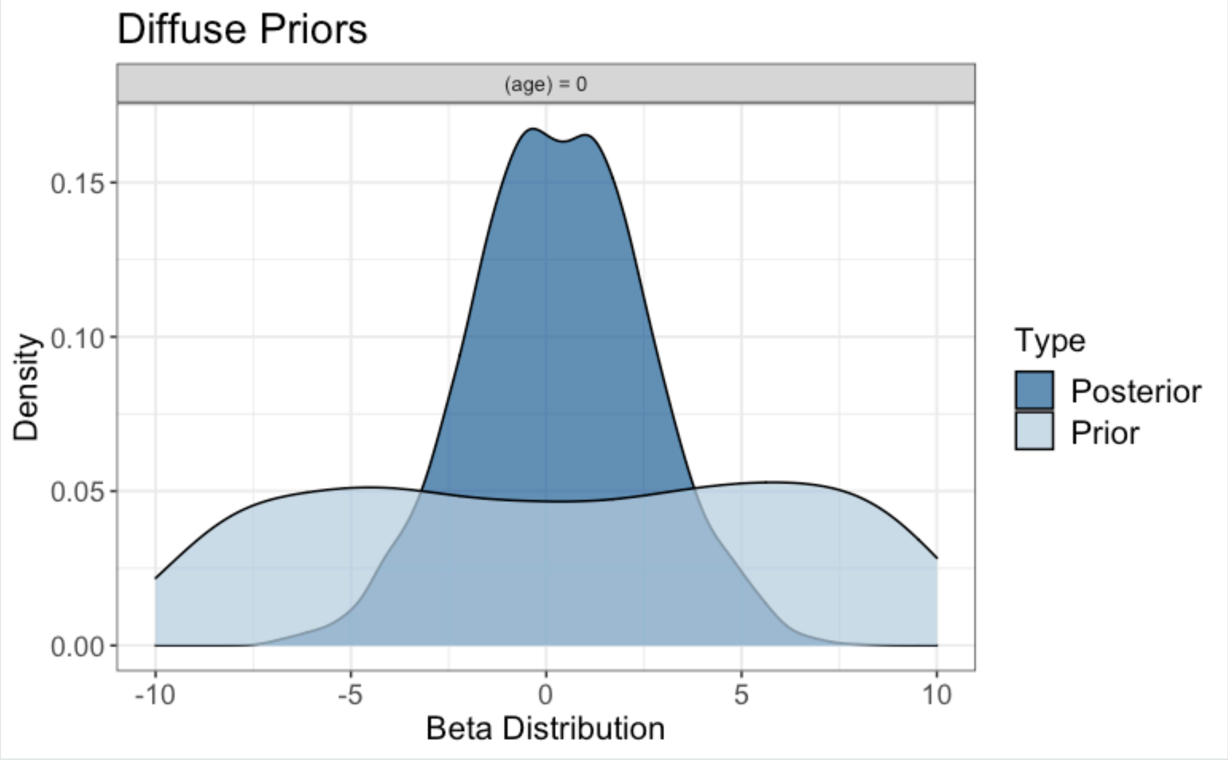
The likelihood of the Null is 3.23 times greater than the likelihood of the alternative.

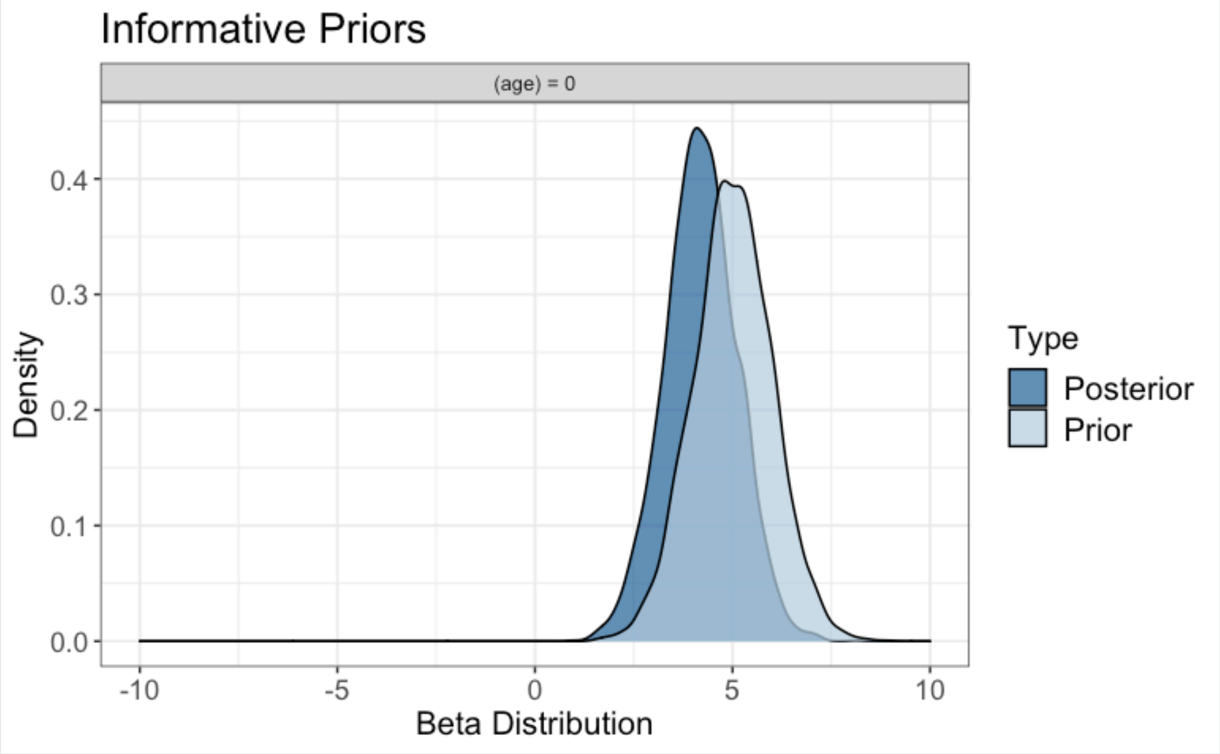
The Bayes Factor

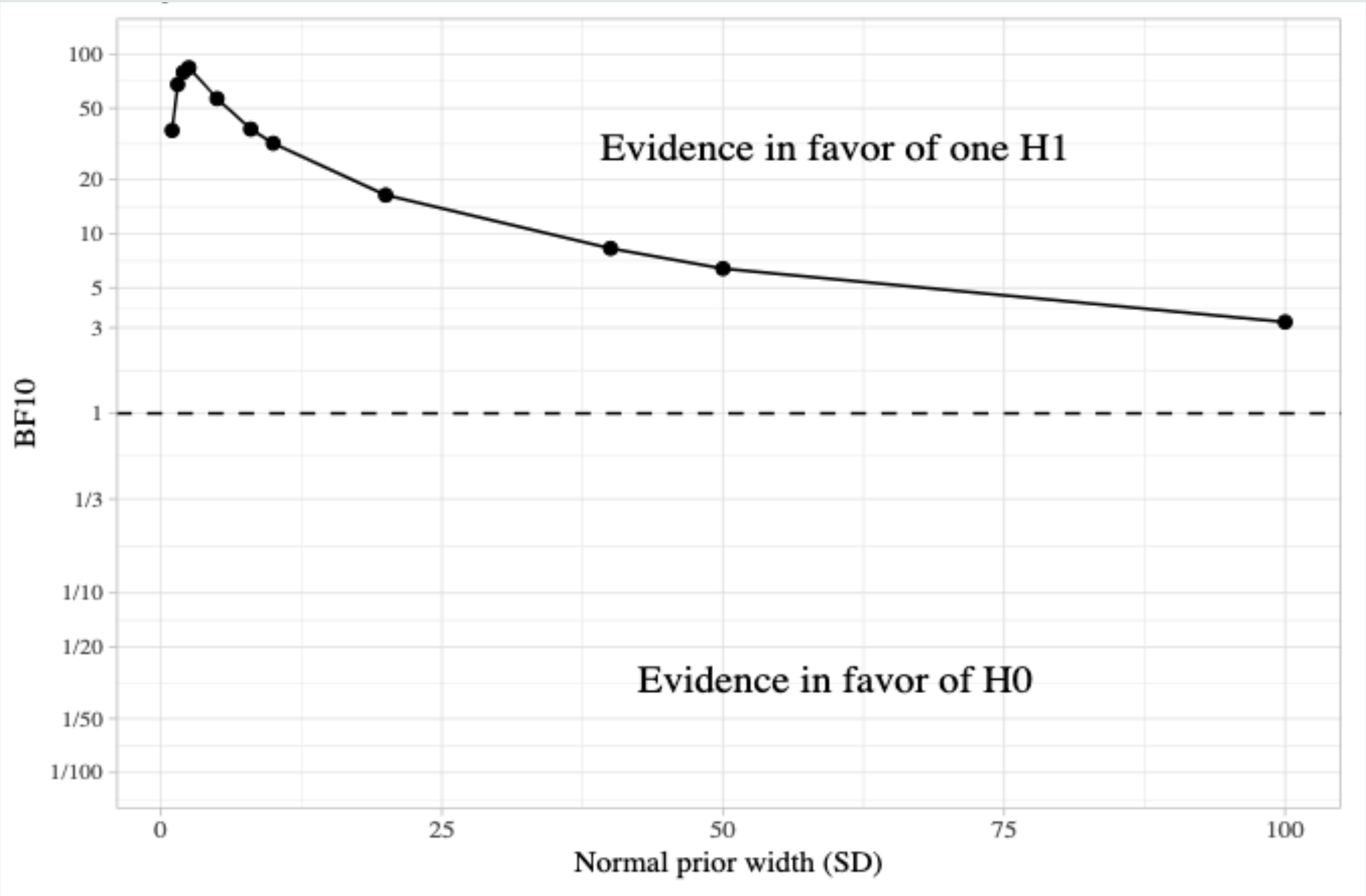
Data Size	Significant?	Prior	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Small	No	Flat	3.23	1/3.23 = .31
Small	**No**	**Diffuse**	**45.6**	**.02**
Small	No	Weak	5.17	.19
Small	No	Informative	7.25	.14
Small	No	Hyper-Informative	9.54	.10

The Bayes Factor

When we include effect sizes that are not part of our data in our priors, our evidence for the null grows.

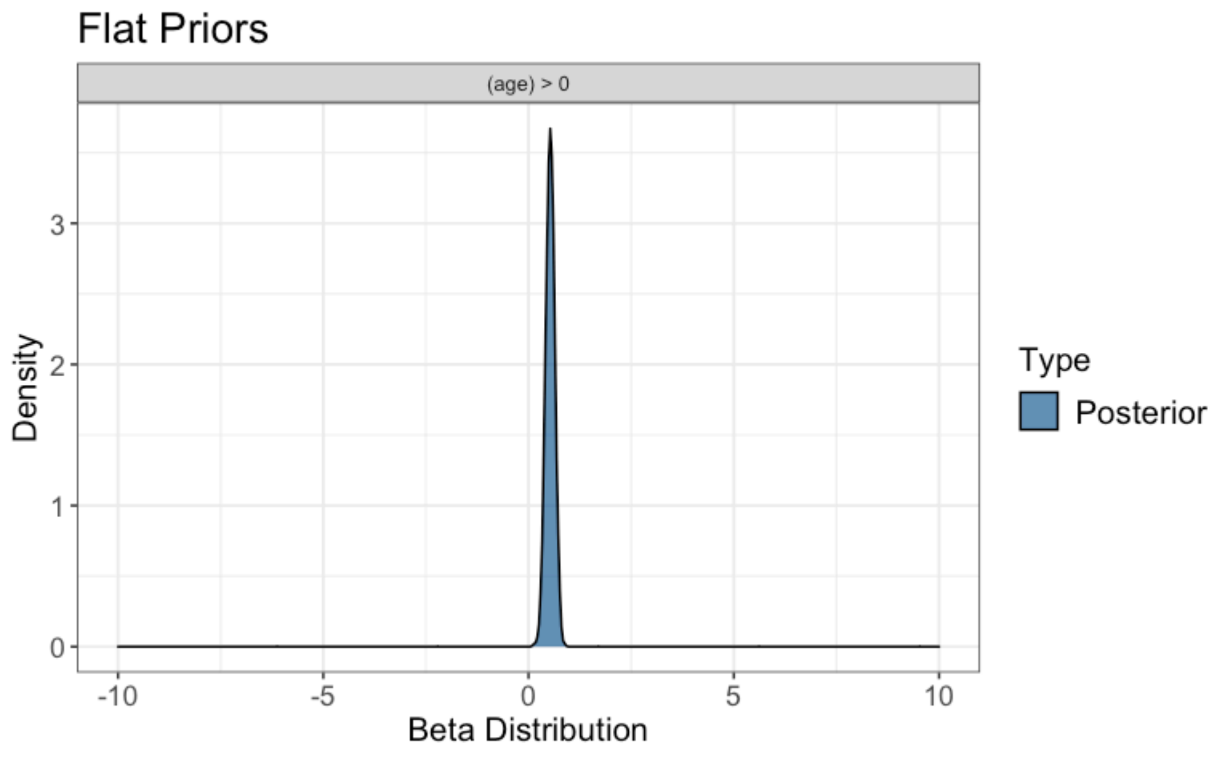




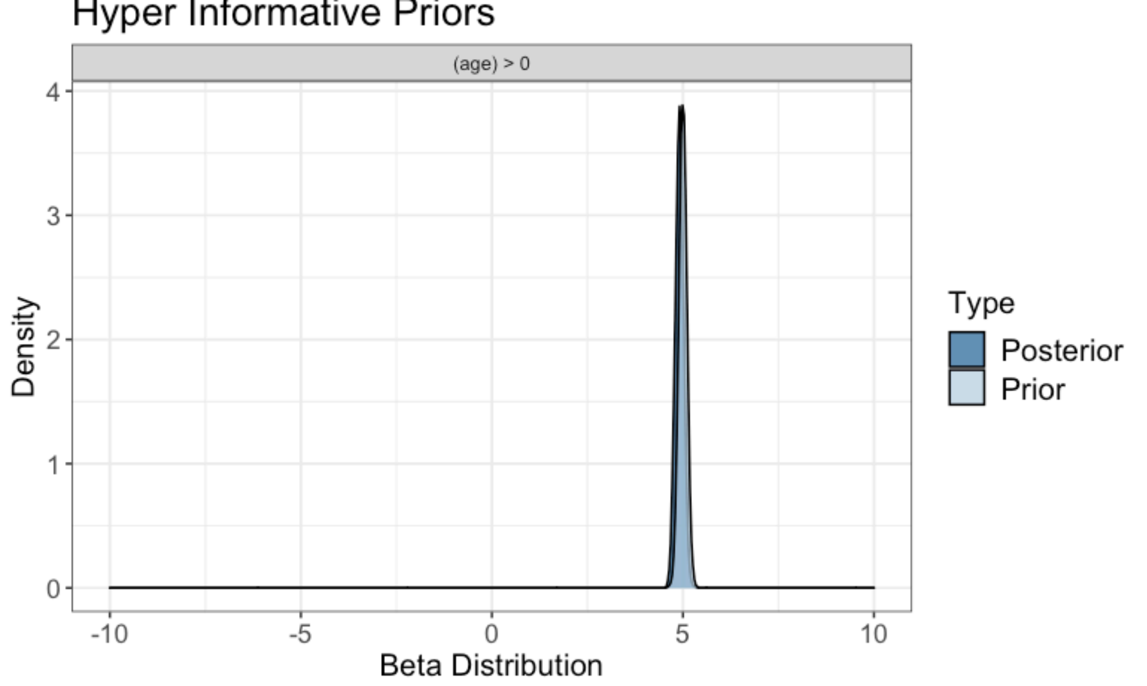


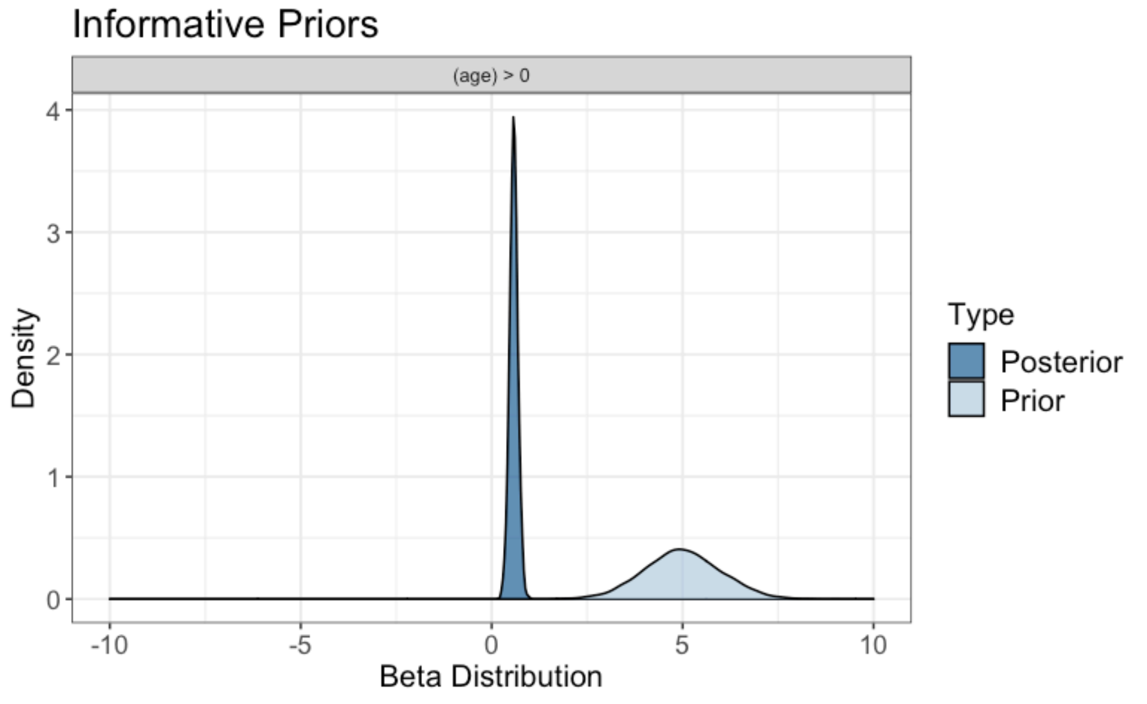
https://vasishth.github.io/bayescogsci/book/sec-N400BF.html

The Bayes Factor



Significant data (40 obs)





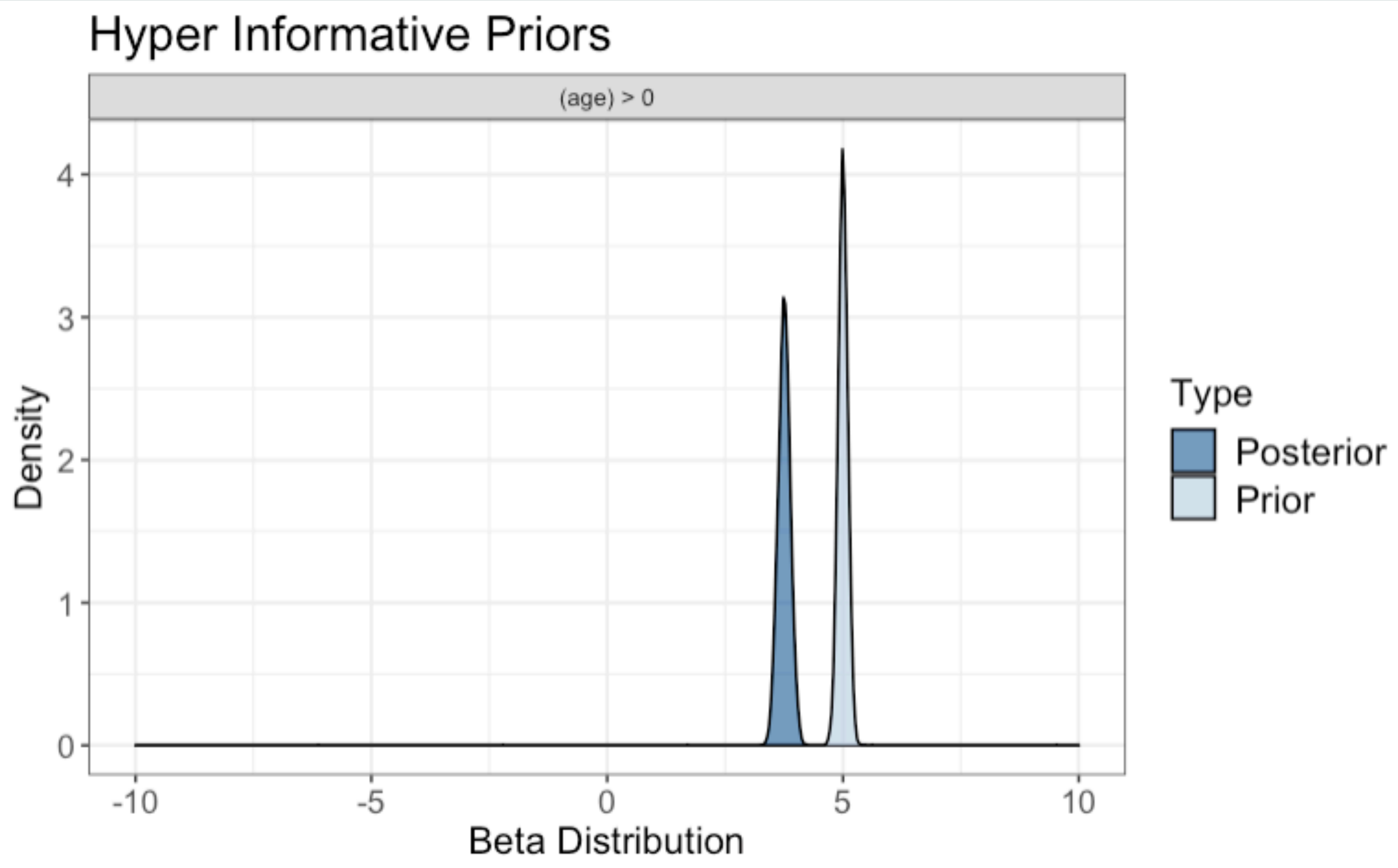
The Bayes Factor

Data Size	Significant?	Prior supports H_A	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Small	No	Flat	3.23	1/3.23 = .31
Small	No	Diffuse	45.6	.02
Small	No	Weak	5.17	.19
Small	No	Informative	7.25	.14
Small	No	Hyper-Informative	9.54	.10
Small	Yes	Flat	.0006	1719.15
Small	Yes	Diffuse	**.04**	25
Small	Yes	Weak	.005	200
Small	Yes	Informative	**8.67**	.12
Small	Yes	Hyper-Informative	**>1000**	<.0001

The Bayes Factor

Data Size	Significant?	Prior supports H_A	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Large	No	Flat	8.33	.12
Large	No	Diffuse	150.58	.007
Large	No	Weak	16.75	.06
Large	No	Informative	6392.59	.0002
Large	No	Hyper-Informative	>10,000	< .0001
Large	Yes	Flat	.17	5.85
Large	Yes	Diffuse	<.0001	>10,000
Large	Yes	Weak	<.0001	>10,000
Large	Yes	Informative	<.0001	>10,000
Large	Yes	Hyper-Informative	**>10,000**	< .0001

The Bayes Factor



Setting Appropriate Priors

diffuse <- c(set_prior("normal(**0**,100)", class = "b", coef = "age"),

set_prior("student_t(3,200,2.5)", class="Intercept"))

weak <- set_prior("normal(**0**,10)", class = "b", coef = "age")

informative <- set_prior("normal(**0**,1)", class = "b", coef = "age")

hyperinformative <- set_prior("normal(**0**,0.1)", class = "b", coef = "age")

Prior belief = no effect

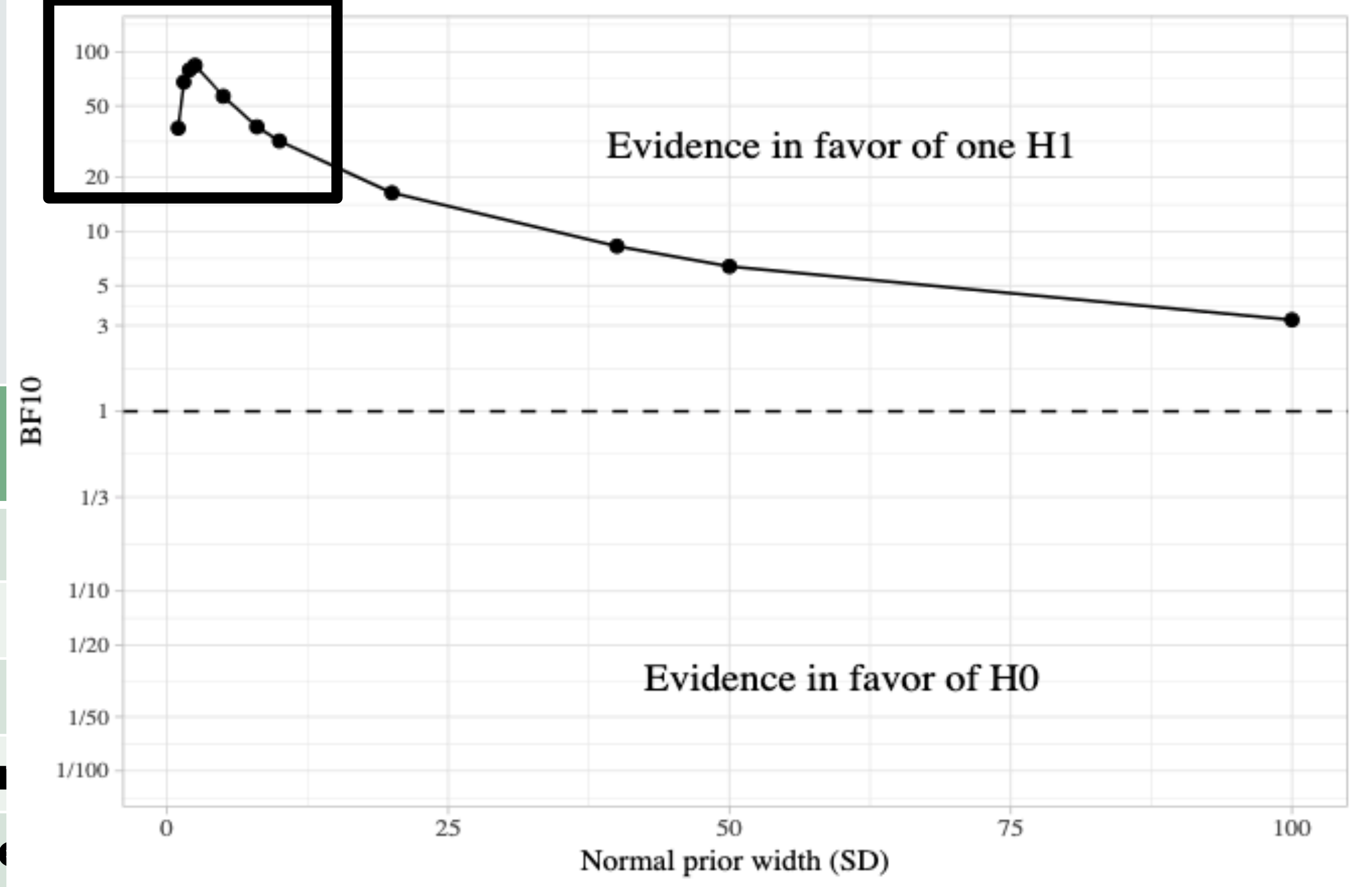
The Bayes Factor

Data Size	Significant?	Prior supports H_A	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Large	No	Flat	8.33	.12
Large	No	Diffuse	150.58	.007
Large	No	Weak	16.75	.06
Large	Yes	Flat	.17	5.85
Large	Yes	Diffuse	<.0001	>10,000
Large	Yes	Weak	<.0001	>10,000
Large	Yes	Informative	<.0001	>10,000
Large	Yes	Hyper-Informative	<.0001	>10,000

The Bayes Factor

Data Size	Significant?	Prior supports H_A	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Large	No	Flat	8.33	.12
Large	No	Diffuse	150.58	.007
Large	No	Weak	16.75	.06
Large	**No**	**Informative**	**1.77**	**.56**
Large	**No**	**Hyper-Informative**	**1.01**	**.99**
Large	Yes	Flat	.17	5.85
Large	Yes	Diffuse	<.0001	>10,000
Large	Yes	Weak	<.0001	>10,000
Large	Yes	Informative	<.0001	>10,000
Large	Yes	Hyper-Informative	<.0001	>10,000

The Bayes Factor



Data Size	Significant?			
Large	No			
Large	No			
Large	No			
Large	**No**	**I**		
Large	**No**	**Hyp**		
Large	Yes	Flat	.17	5.85
Large	Yes	Diffuse	<.0001	>10,000
Large	Yes	Weak	<.0001	>10,000
Large	Yes	Informative	<.0001	>10,000
Large	**Yes**	**Hyper-Informative**	**<.0001**	**>10,000**

The Bayes Factor

Data Size	Significant?	Prior supports H_A	Bayes Factor BF_{01} Favors the Null	Bayes Factor BF_{10} Favors the alternative
Large	No	Flat	8.33	.12
Large	No	Diffuse	150.58	.007
Large	No	Weak	16.75	.06
Large	No	Informative	1.77	.56
Large	No	Hyper-Informative	1.01	.99
Large	**Yes**	**Flat**	**.17**	**5.85**
Large	Yes	Diffuse	<.0001	>10,000
Large	Yes	Weak	<.0001	>10,000
Large	Yes	Informative	<.0001	>10,000
Large	Yes	Hyper-Informative	<.0001	>10,000

Overview of Bayesian Stats

- ~~Why bother?~~
- ~~Principles of Probability~~
- ~~Setting Appropriate Priors~~
- ~~The Bayes Factor~~
- BRMs & Stan – The R Code
- Reporting Your Results

The R Code

Packages:

- **brms**
- **BayesFactor**
- **rstanarm**

The R Code

Flat <- brm(**vk~age**, **data=df.ns.small**, family=gaussian(),

chains=4,iter=2000,cores=2, sample_prior = TRUE,save_pars =

save_pars(all=TRUE))

The R Code

Flat <- brm(vk~age, data=df.ns.small, **family=gaussian(),** chains=4, iter=2000, cores=2, sample_prior = TRUE,save_pars = save_pars(all=TRUE))

The R Code

Flat <- brm(vk~age, data=df.ns.small, family=gaussian(), **chains=4, iter=2000, cores=2**, **sample_prior = TRUE**, save_pars = save_pars(all=TRUE))

The R Code

```
Flat <- brm(vk~age, data=df.ns.small, family=gaussian(), chains=4,
iter=2000, cores=2, sample_prior = TRUE, save_pars =
save_pars(all=TRUE))
```

The R Code

Flat <- brm(vk~age, data=df.ns.small, family=gaussian(), chains=4, iter=2000, cores=2, sample_prior = TRUE, save_pars = save_pars(all=TRUE))

Inf <- brm(vk~age, data=df.ns.small, family=gaussian(), chains=4, cores=2**, prior=informative**, sample_prior = TRUE, save_pars=save_pars(all=TRUE))

The R Code

```
prior_summary(Inf)

informative <- c(set_prior("normal(0,100)", class = "b", coef = "age"),
          set_prior("student_t(3,200,2.5)", class="Intercept"))
```

The R Code

With BayesFactor Package

```
Bf <- lmBF(vk~age, df)
```

The R Code

With BayesFactor Package

```
Bf <- lmBF(vk~age, df)
```

With brms

Set a null model

```
null <- update(inf, formula = ~ . -age)
```

The R Code

With BayesFactor Package

```
Bf <- lmBF(vk~age, df)
```

With brms

Set a null model

```
null <- update(inf, formula = ~ . -age)
```

Compare the models

```
bayes_factor(null, inf)
```

Overview of Bayesian Stats

- ~~Why bother?~~
- ~~Principles of Probability~~
- ~~Setting Appropriate Priors~~
- ~~The Bayes Factor~~
- ~~BRMs & Stan – The R Code~~
- Reporting Your Results

Write-Up

- Effect size from Bayes Analysis should be reported with credible intervals
- Explain what bayes factors are

  This includes establishing notation and levels for interpretation.

  BF_{01} *or* BF_{10}

- Explain how BF were calculated (iterations, chains, priors, sensitivity)
- Next to a p-value, write the notation for the BF_{01} or BF_{10} and give the value ($BF_{01} = 25$)
- Explain the strength of the evidence using Jeffrey’s levels (or other)

Write-Up

… p = .65, BF_{01} = 25. Here, the Bayes Factor provides strong evidence for the null hypothesis that cleft and connective presence neither increase nor decrease reading times for the non-focused target word.

Overview of Bayesian Stats

- ~~Why bother?~~
- ~~Principles of Probability~~
- ~~Setting Appropriate Priors~~
- ~~The Bayes Factor~~
- ~~BRMs & Stan – The R Code~~
- ~~Reporting Your Results~~