Tidy data and Bayesian analysis make uncertainty visualization fun

Matthew Kay, Assistant Professor
School of Information & Department of Computer Science and Engineering
University of Michigan

What happens when we ignore uncertainty?

A mixed-design ANOVA with sex of face (male, female) as a within-subjects factor and self-rated attractiveness (low, average, high) and oral contraceptive use (true, false) as between-subjects factors revealed a main effect of sex of face, $F(1, 1276) = 1372$, $p < .001$, $\eta_p^2 = .52$. This was qualified by interactions between sex of face and SRA, $F(2, 1276) = 6.90$, $p = .001$, $\eta_p^2 = .011$, and between sex of face and oral contraceptive use, $F(1, 1276) = 5.02$, $p = .025$, $\eta_p^2 = .004$. The predicted interaction among sex of face, SRA and oral contraceptive use was not significant, $F(2, 1276) = 0.06$, $p = .94$, $\eta_p^2 < .001$. All other main effects and interactions were non-significant and irrelevant to our hypotheses, all $F \leq 0.94$, $p \geq .39$, $\eta_p^2 \leq .001$.

A mixed-design ANOVA with sex of face (male, female) as a within-subjects factor and self-rated attractiveness (low, average, high) and oral contraceptive use (true, false) as between-subjects factors revealed a main effect of sex of face, $F(1, 1276) = 1372$, $p < .001$, $\eta_p^2 = .52$. This was qualified by interactions between sex of face and SRA, $F(2, 1276) = 6.90$, $p = .001$, $\eta_p^2 = .011$, and between sex of face and oral contraceptive use, $F(1, 1276) = 5.02$, $p = .025$, $\eta_p^2 = .0$[illegible]. The predicted interaction among sex of face, SRA and oral contraceptive use was not significant, $F(2, 1276) = 0.06$, $p = .94$, $\eta_p^2 < .001$. All other main effects and interactions were not significant and irrelevant to our hypotheses, all $F \leq 0.9$[illegible], $p \geq .39$, $\eta_p^2 \leq .001$.

Alternatives…

Table 7
Stevens et al. 2006, table 2: Determinants of authoritarian aggression

Variable	Coefficient (Standard Error)
Constant	.41 (.93)
Countries	
Argentina	1.31 (.33)**[B,M]
Chile	.93 (.32)**[B,M]
Colombia	1.46 (.32)**[B,M]
Mexico	.07 (.32)[A,CH,CO,V]
Venezuela	.96 (.37)**[B,M]
Threat	
Retrospective egocentric economic perceptions	.20 (.13)
Prospective egocentric economic perceptions	.22 (.12)#
Retrospective sociotropic economic perceptions	−.21 (.12)#
Prospective sociotropic economic perceptions	−.32 (.12)*
Ideological distance from president	−.27 (.07)**
Ideology	
Ideology	.23 (.07)**
Individual Differences	
Age	.00 (.01)
Female	−.03 (.21)
Education	.13 (.14)
Academic Sector	.15 (.29)
Business Sector	.31 (.25)
Government Sector	−.10 (.27)
R^2	.15
Adjusted R^2	.12
N	500

**$p < .01$, *$p < .05$, #$p < .10$ (twotailed)

[A]Coefficient is significantly different from Argentina's at $p < .05$;

[B]Coefficient is significantly different from Brazil's at $p < .05$;

[CH]Coefficient is significantly different from Chile's at $p < .05$;

[CO]Coefficient is significantly different from Colombia's at $p < .05$;

[M]Coefficient is significantly different from Mexico's at $p < .05$;

[V]Coefficient is significantly different from Venezuela's at $p < .05$.

Alternatives…

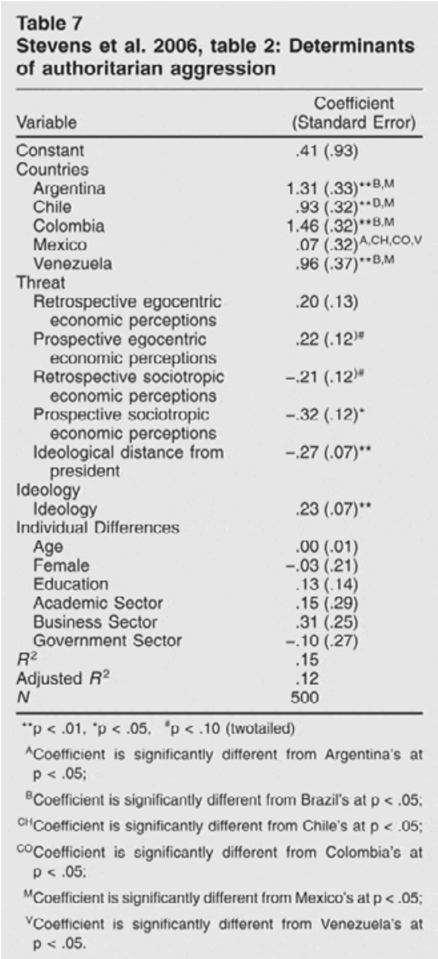
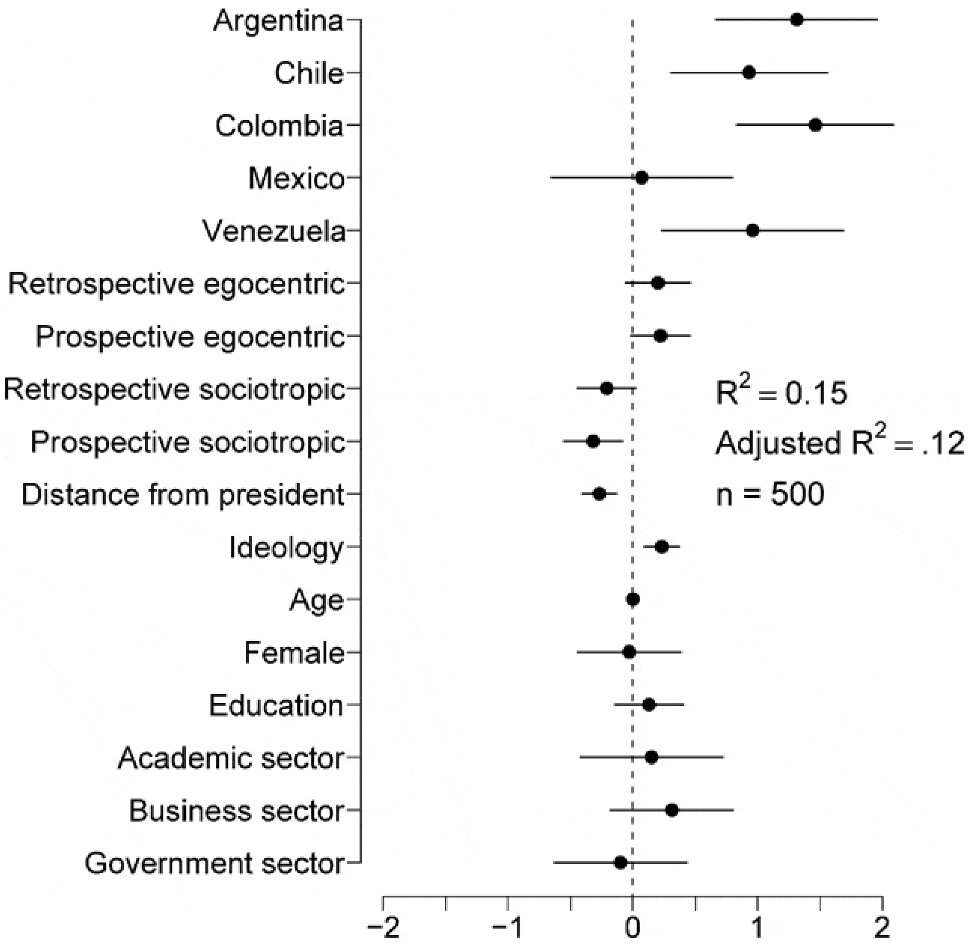


Table 7
Stevens et al. 2006, table 2: Determinants of authoritarian aggression

Variable	Coefficient (Standard Error)
Constant	.41 (.93)
Countries	
Argentina	1.31 (.33)**B,M
Chile	.93 (.32)**B,M
Colombia	1.46 (.32)**B,M
Mexico	.07 (.32)A,CH,CO,V
Venezuela	.96 (.37)**B,M
Threat	
Retrospective egocentric economic perceptions	.20 (.13)
Prospective egocentric economic perceptions	.22 (.12)#
Retrospective sociotropic economic perceptions	−.21 (.12)#
Prospective sociotropic economic perceptions	−.32 (.12)*
Ideological distance from president	−.27 (.07)**
Ideology	
Ideology	.23 (.07)**
Individual Differences	
Age	.00 (.01)
Female	−.03 (.21)
Education	.13 (.14)
Academic Sector	.15 (.29)
Business Sector	.31 (.25)
Government Sector	−.10 (.27)
R^2	.15
Adjusted R^2	.12
N	500

**p < .01, *p < .05, #p < .10 (twotailed)
ACoefficient is significantly different from Argentina's at p < .05;
BCoefficient is significantly different from Brazil's at p < .05;
CHCoefficient is significantly different from Chile's at p < .05;
COCoefficient is significantly different from Colombia's at p < .05;
MCoefficient is significantly different from Mexico's at p < .05;
VCoefficient is significantly different from Venezuela's at p < .05.



[Jonathan P Kastellec and Eduardo L Leoni. 2007. Using Graphs Instead of Tables in Political Science. Perspectives on politics 5, 4: 755–771]

Alternatives...

Table 7
Stevens et al. 2006, table 2: Determinants of authoritarian aggression

Variable	Coefficient (Standard Error)
Constant	.41 (.93)
Countries	
Argentina	1.31 (.33)**B,M
Chile	.93 (.32)**B,M
Colombia	1.46 (.32)**B,M
Mexico	.07 (.32)A,CH,CO,V
Venezuela	.96 (.37)**B,M
Threat	
Retrospective egocentric economic perceptions	.20 (.13)
Prospective egocentric economic perceptions	.22 (.12)#
Retrospective sociotropic economic perceptions	−.21 (.12)#
Prospective sociotropic economic perceptions	−.32 (.12)*
Ideological distance from president	−.27 (.07)**
Ideology	
Ideology	.23 (.07)**
Individual Differences	
Age	.00 (.01)
Female	−.03 (.21)
Education	.13 (.14)
Academic Sector	.15 (.29)
Business Sector	.31 (.25)
Government Sector	−.10 (.27)
R^2	.15
Adjusted R^2	.12
N	500

**$p < .01$, *$p < .05$, #$p < .10$ (twotailed)

[A]Coefficient is significantly different from Argentina's at $p < .05$;

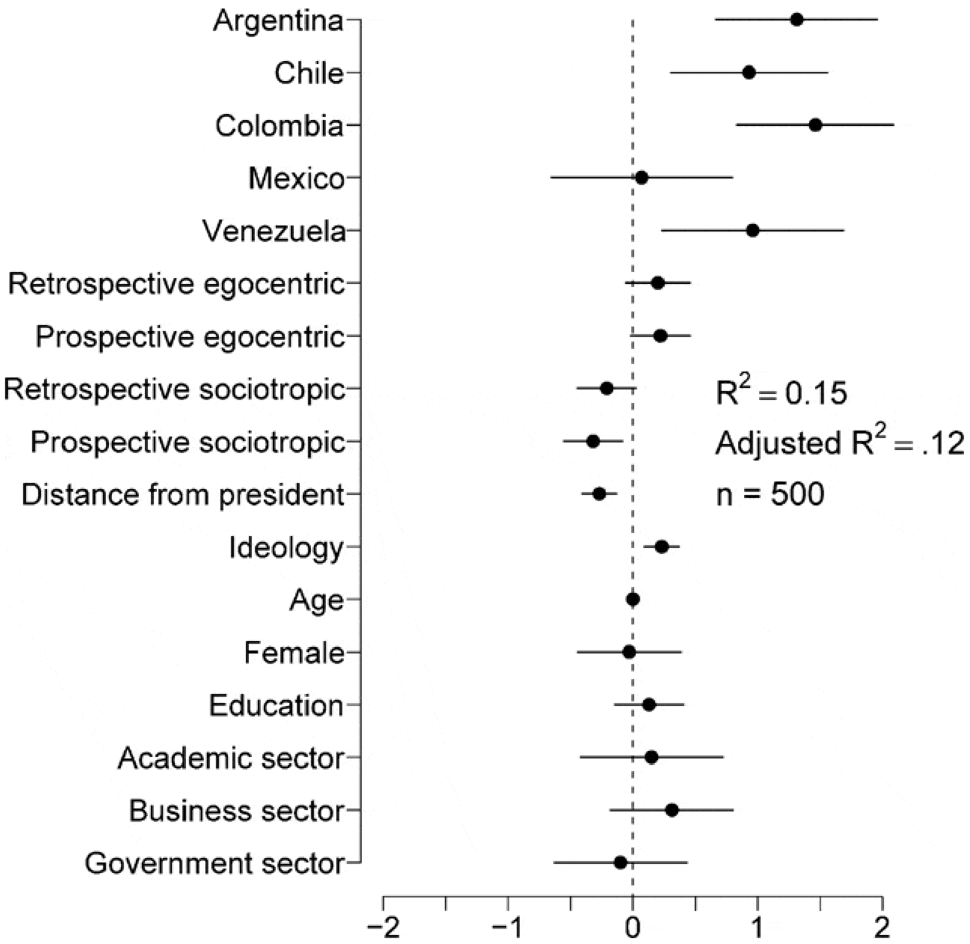
[B]Coefficient is significantly different from Brazil's at $p < .05$;

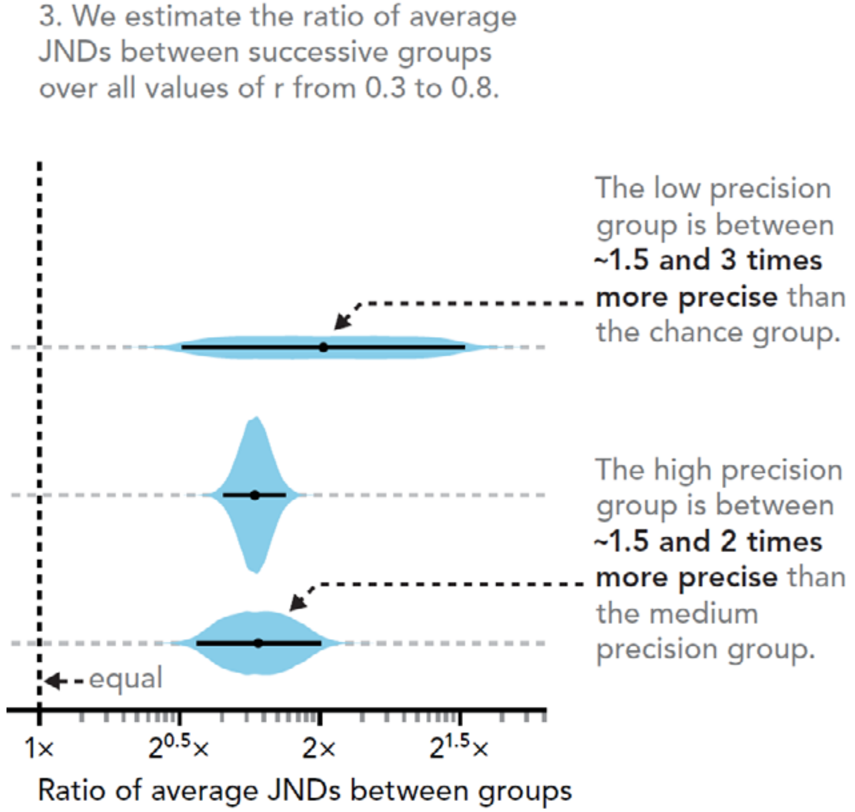
[CH]Coefficient is significantly different from Chile's at $p < .05$;

[CO]Coefficient is significantly different from Colombia's at $p < .05$;

[M]Coefficient is significantly different from Mexico's at $p < .05$;

[V]Coefficient is significantly different from Venezuela's at $p < .05$.

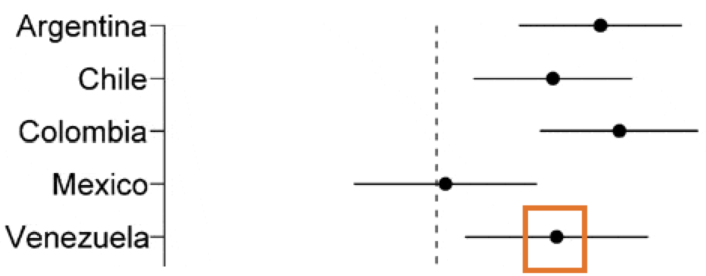


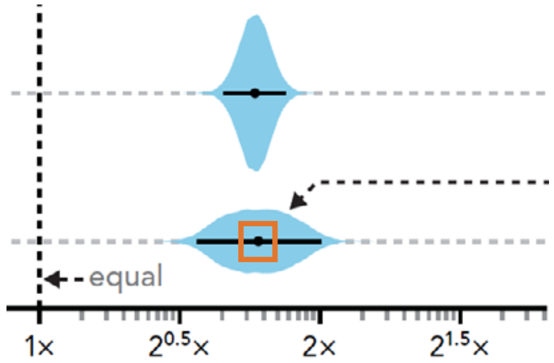


[Jonathan P Kastellec and Eduardo L Leoni. 2007. Using Graphs Instead of Tables in Political Science. Perspectives on politics 5, 4: 755–771]

How easy is it to ignore the uncertainty?

Variable	Coefficient (Standard Error)
Constant	.41 (.93)
Countries	
Argentina	1.31 (.33)**B,M
Chile	.93 (.32)**B,M
Colombia	1.46 (.32)**B,M
Mexico	.07 (.32)A,CH,CO,V
Venezuela	.96 (.37)**B,M





This contributes to dichotomania

Dichotomania...

Predictions from last US presidential election

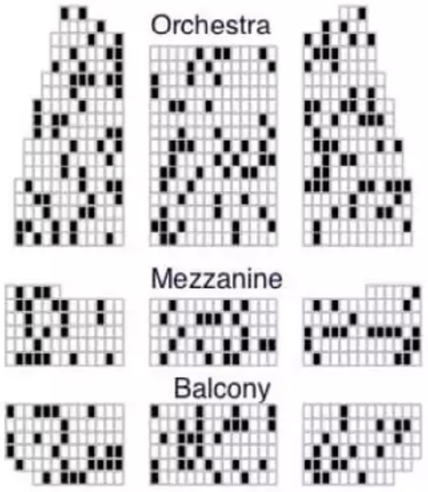
[http://wapo.st/2fCYvDW]

FiveThirtyEight: Trump's Chances	NYT Upshot: Trump's Chances	HuffPo Pollster: Trump's Chances
28%	15%	2%

Predictions from last presidential election

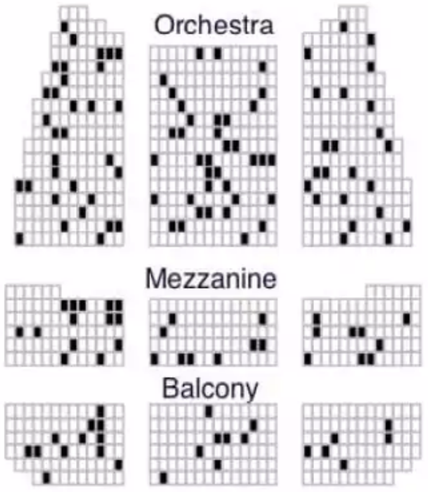
[http://wapo.st/2fCYvDW]

FiveThirtyEight: Trump's Chances



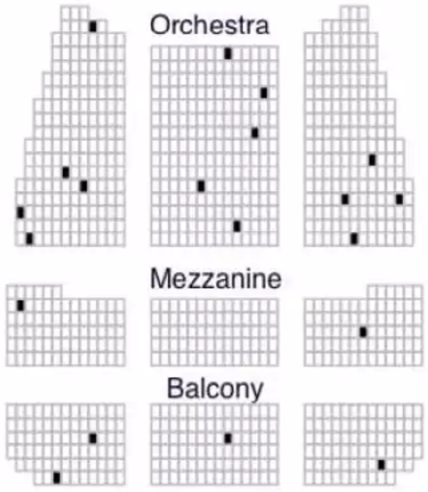
286 cases in 1,000

NYT Upshot: Trump's Chances



150 cases in 1,000

HuffPo Pollster: Trump's Chances



20 cases in 1,000

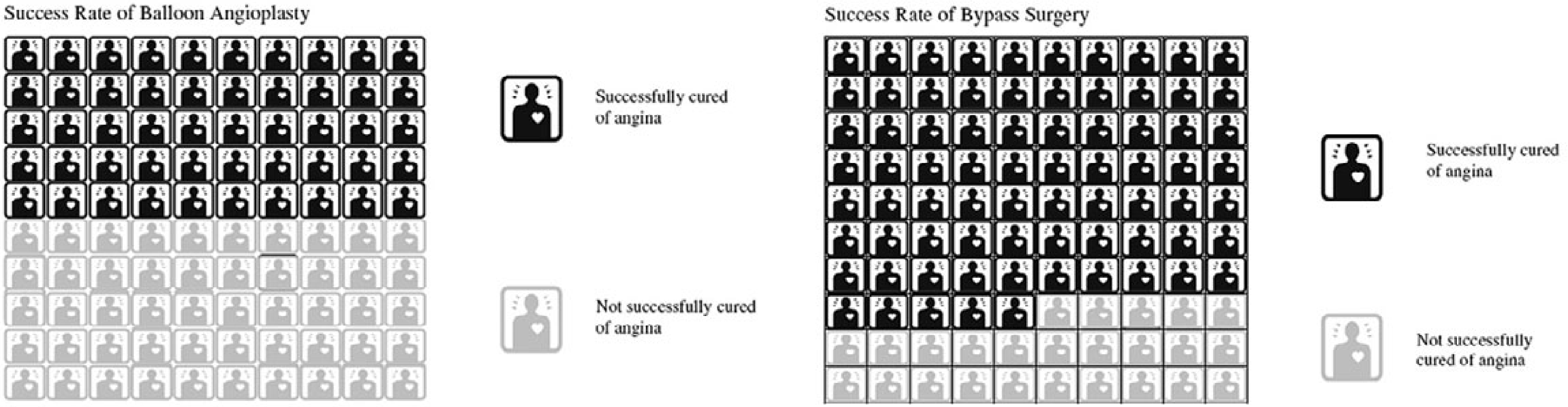
People are very good at ignoring uncertainty...

People are very good at ignoring uncertainty…

Especially when we provide bad uncertainty representations

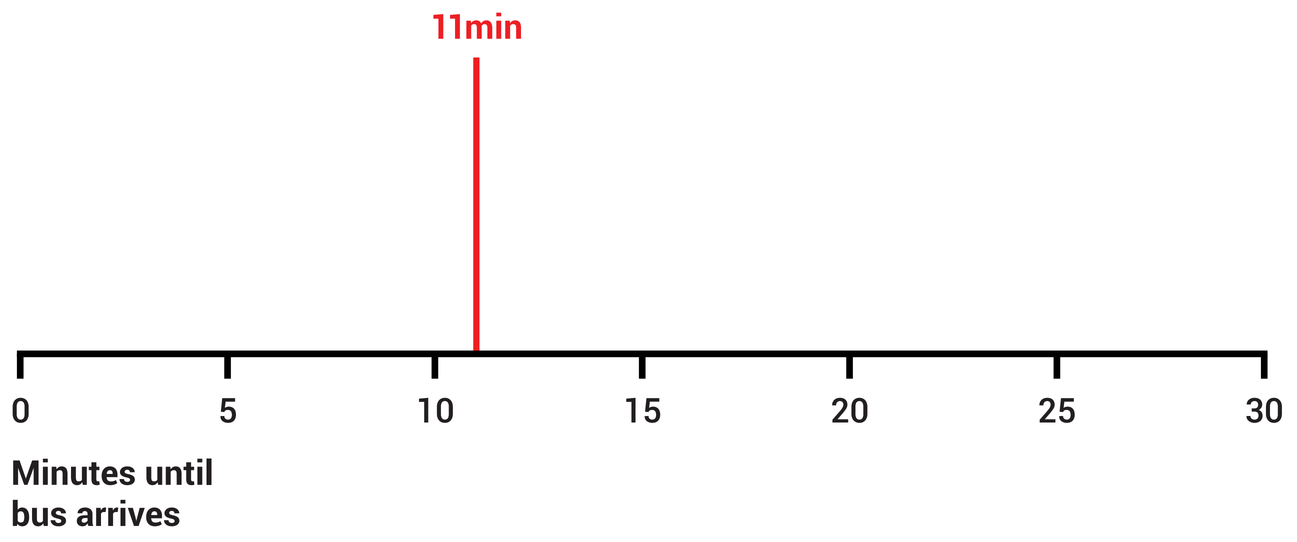
Icon arrays in medical risk communication

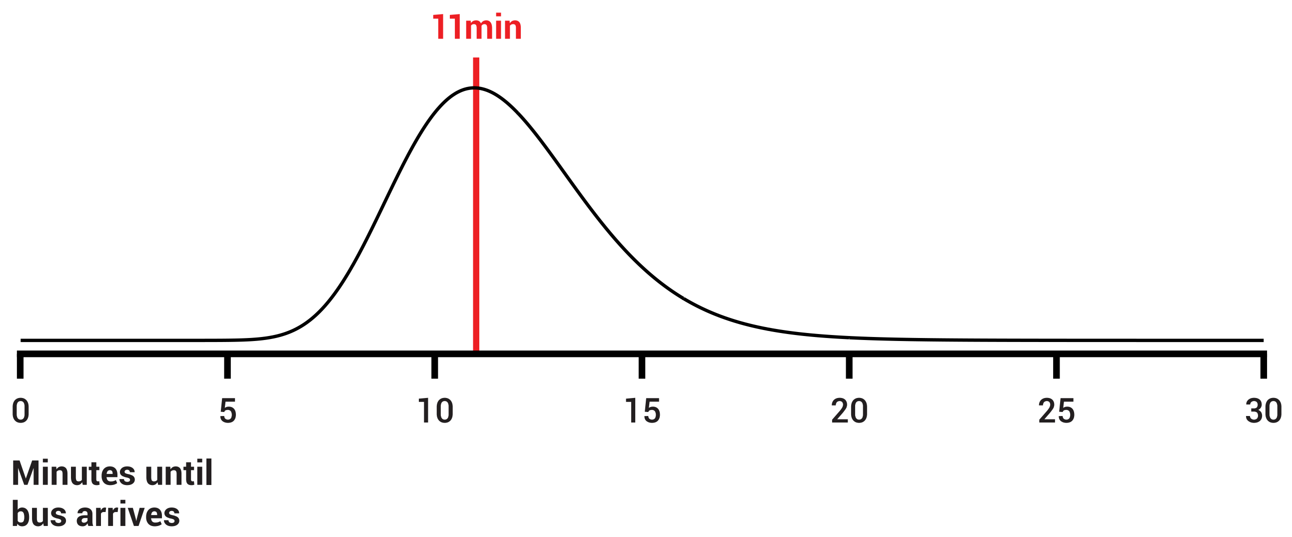
[Figure from Fagerlin et al, 2005]

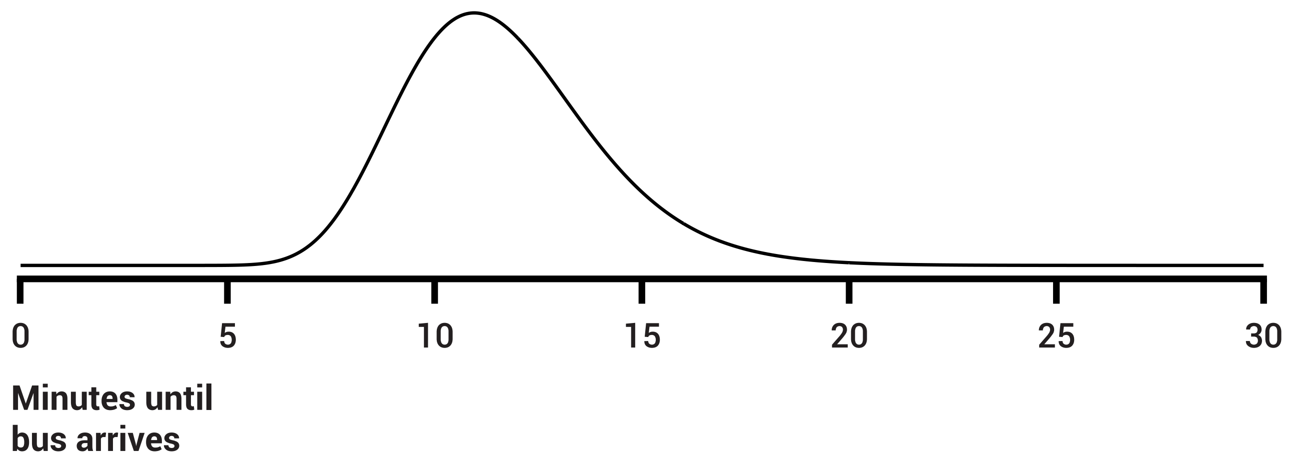


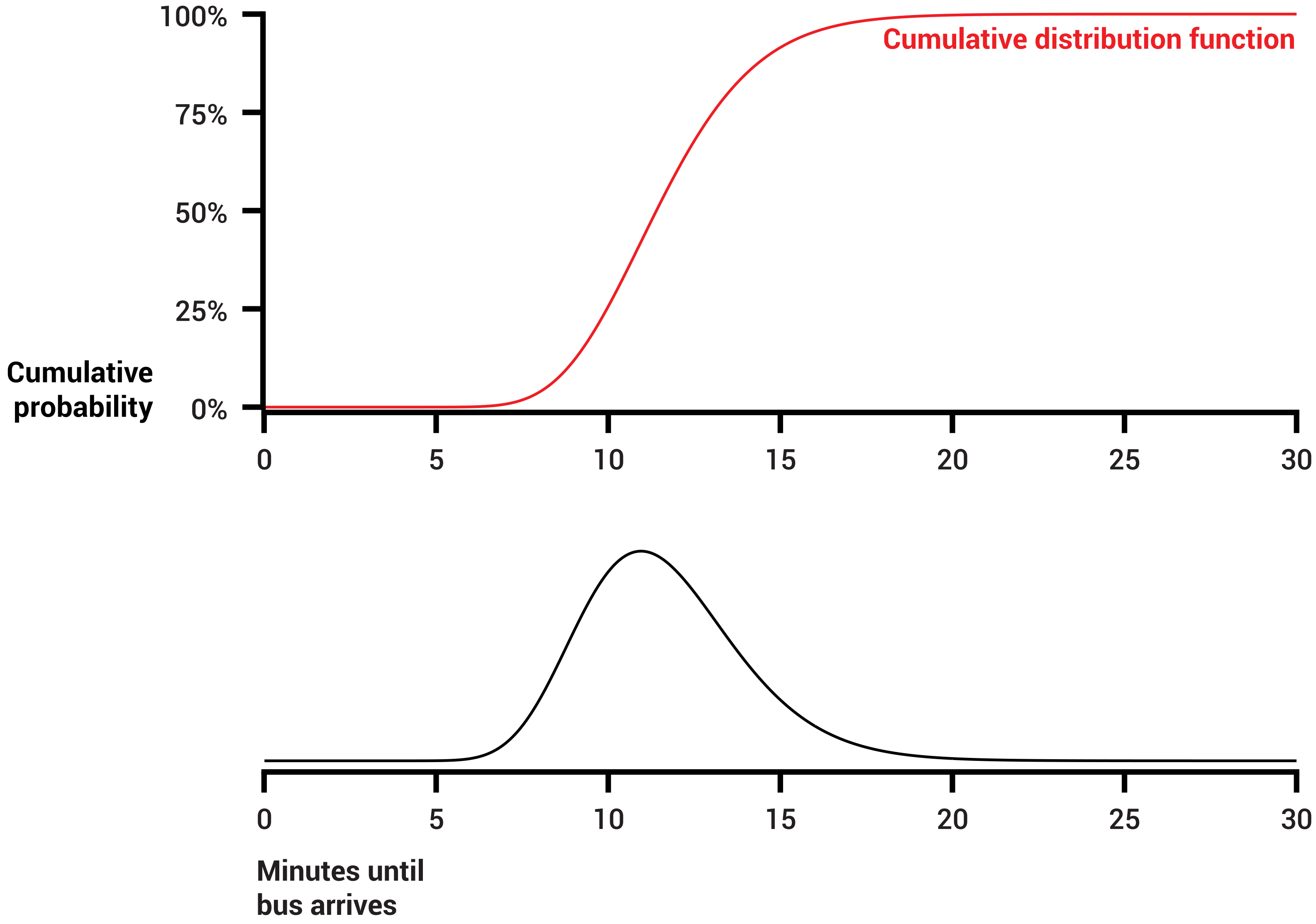
Frequency framing or discrete outcome visualization

What is an icon array for a continuous distribution?

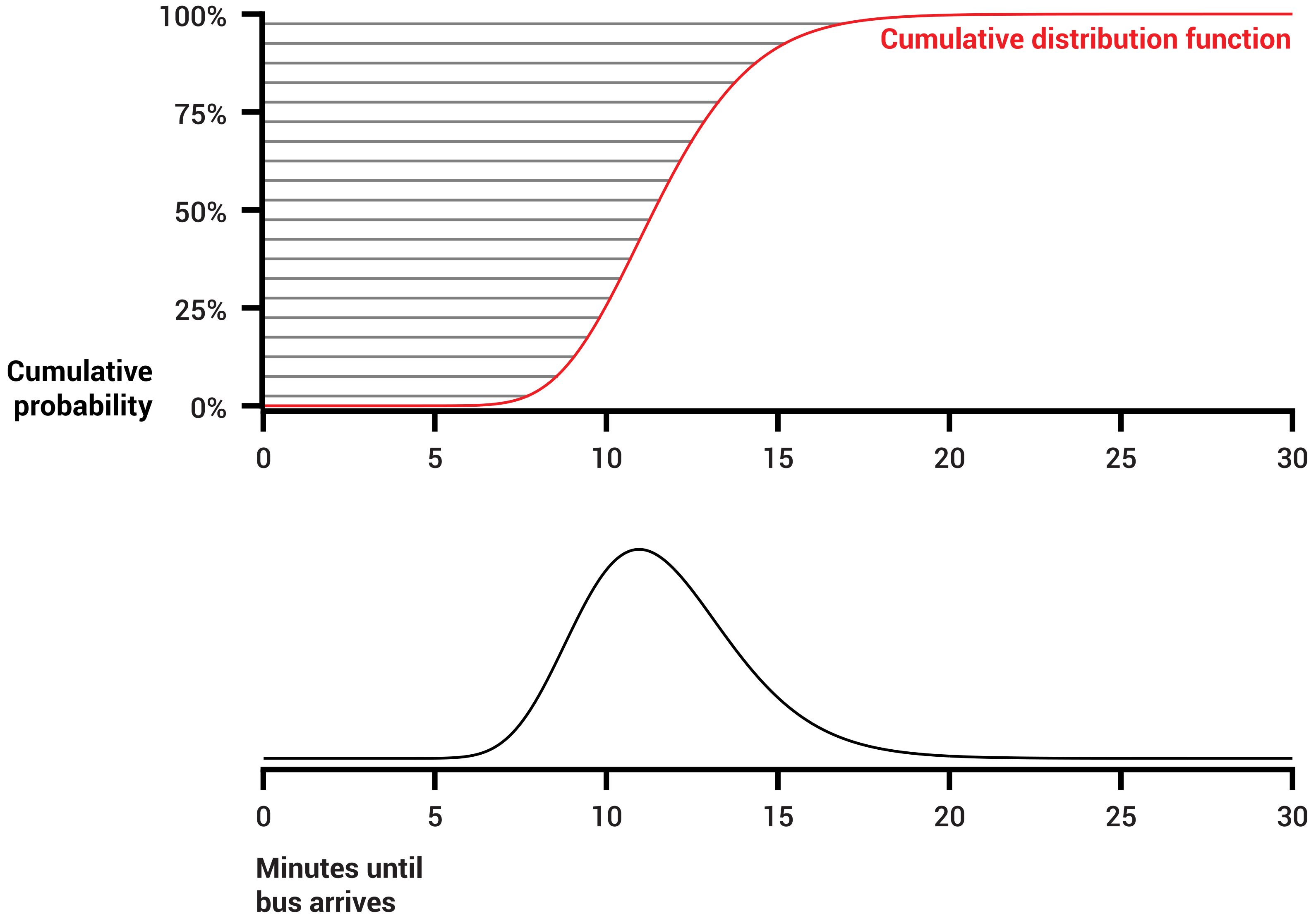

11min
0
5
10
15
20
25
30
Minutes until
bus arrives

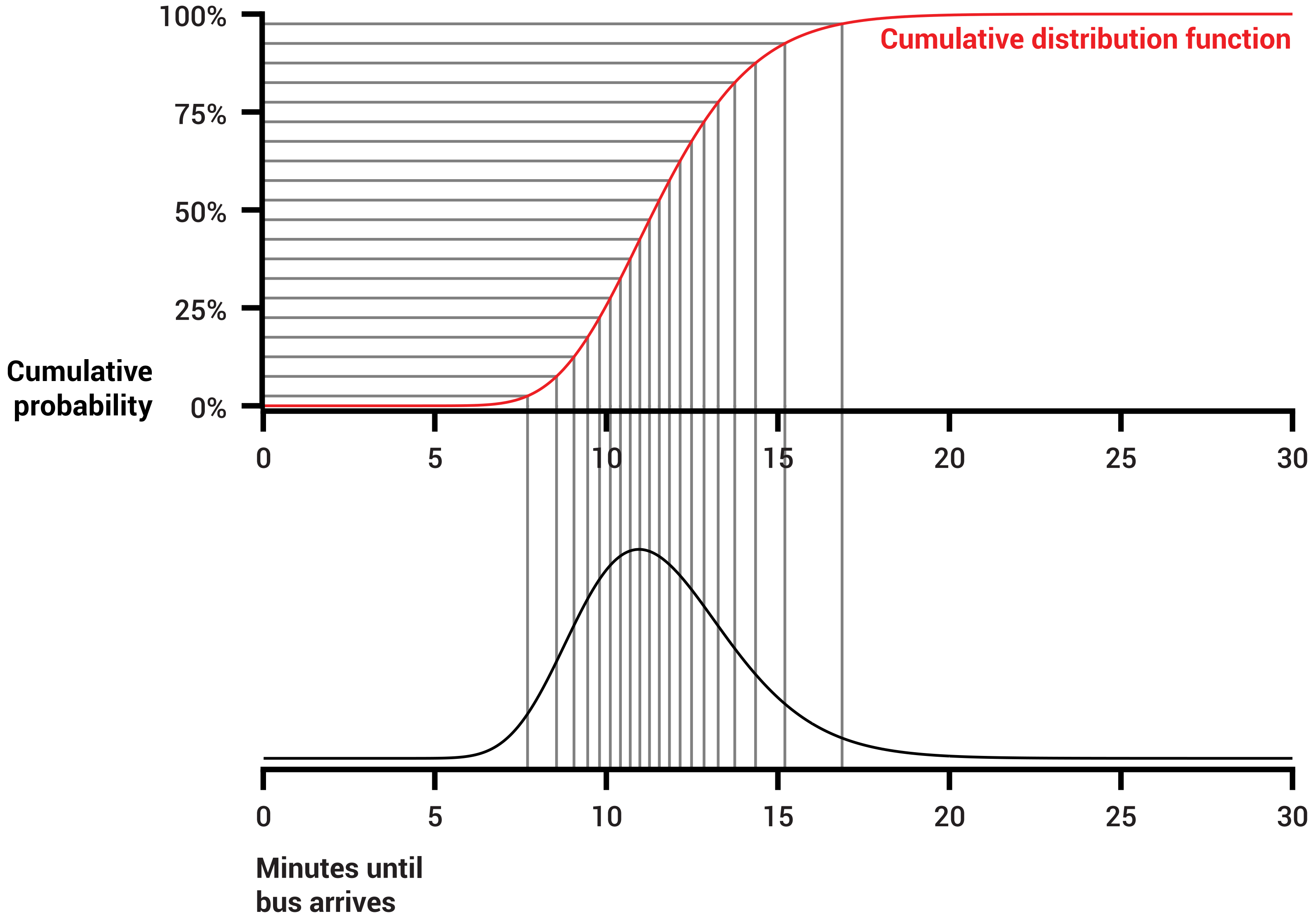

11min
0
5
10
15
20
25
30
Minutes until
bus arrives

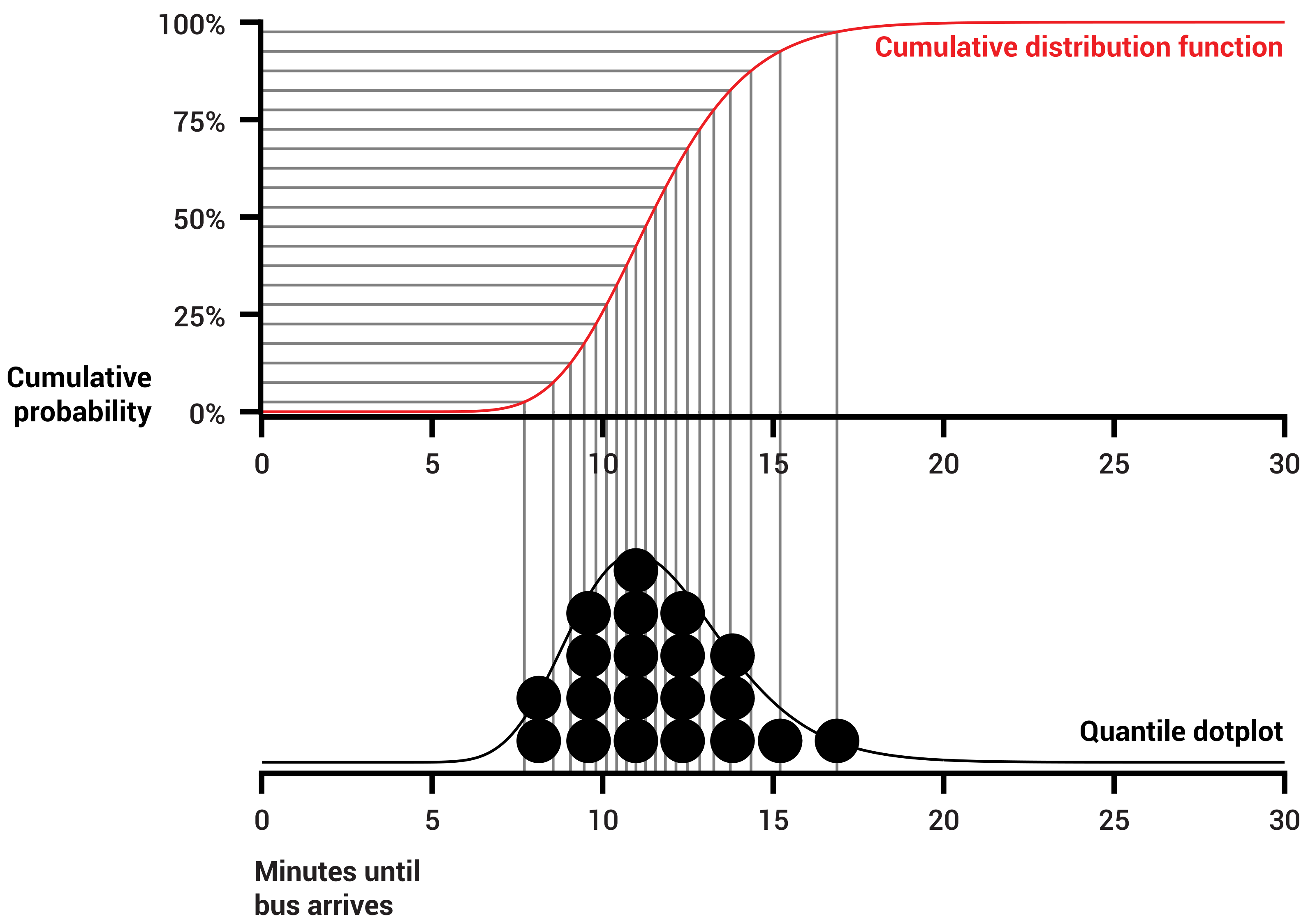

0
5
10
15
20
25
30
Minutes until
bus arrives



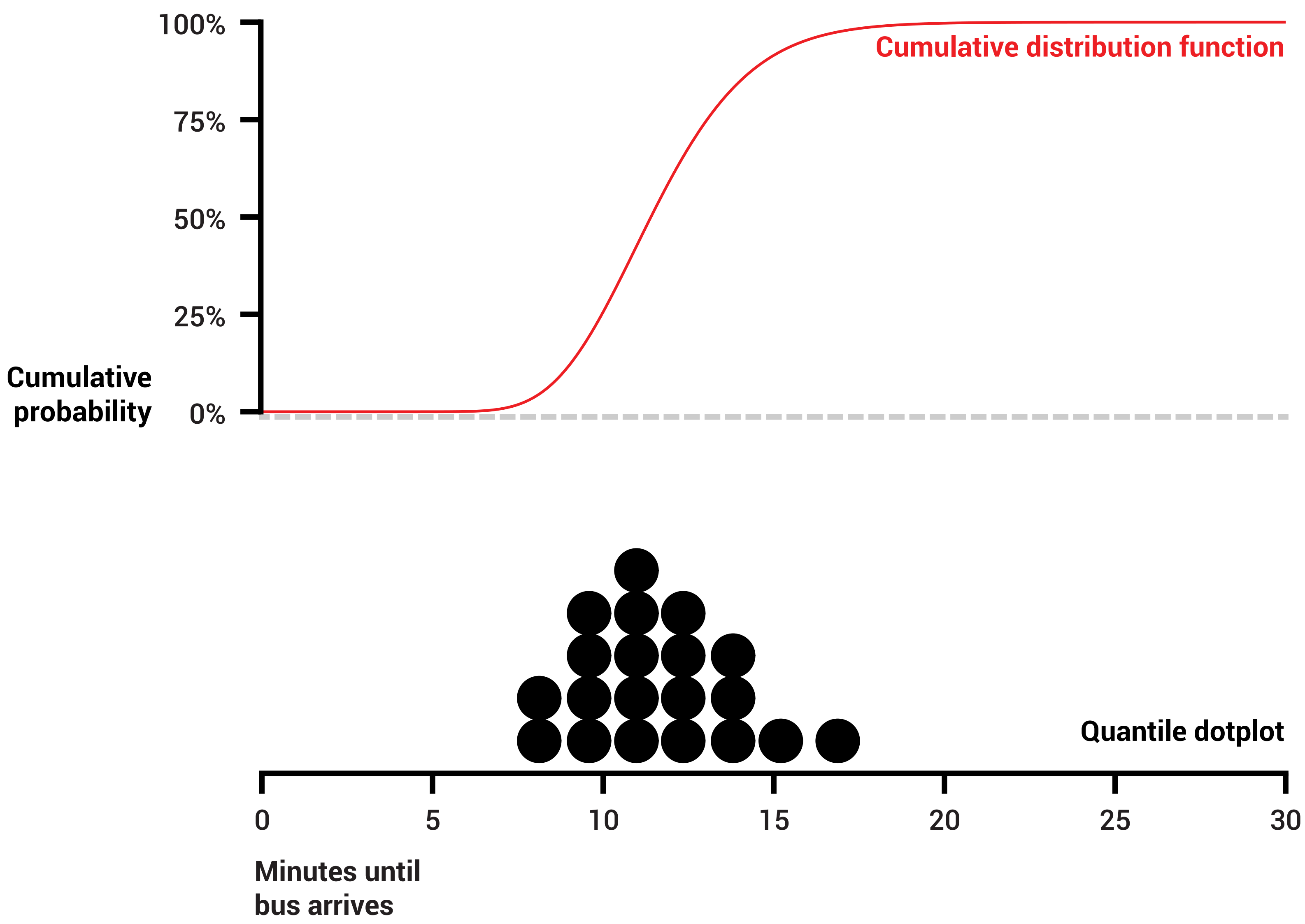
Cumulative distribution function
100%
75%
50%
25%
0%
Cumulative probability
0
5
10
15
20
25
30
0
5
10
15
20
25
30
Minutes until bus arrives

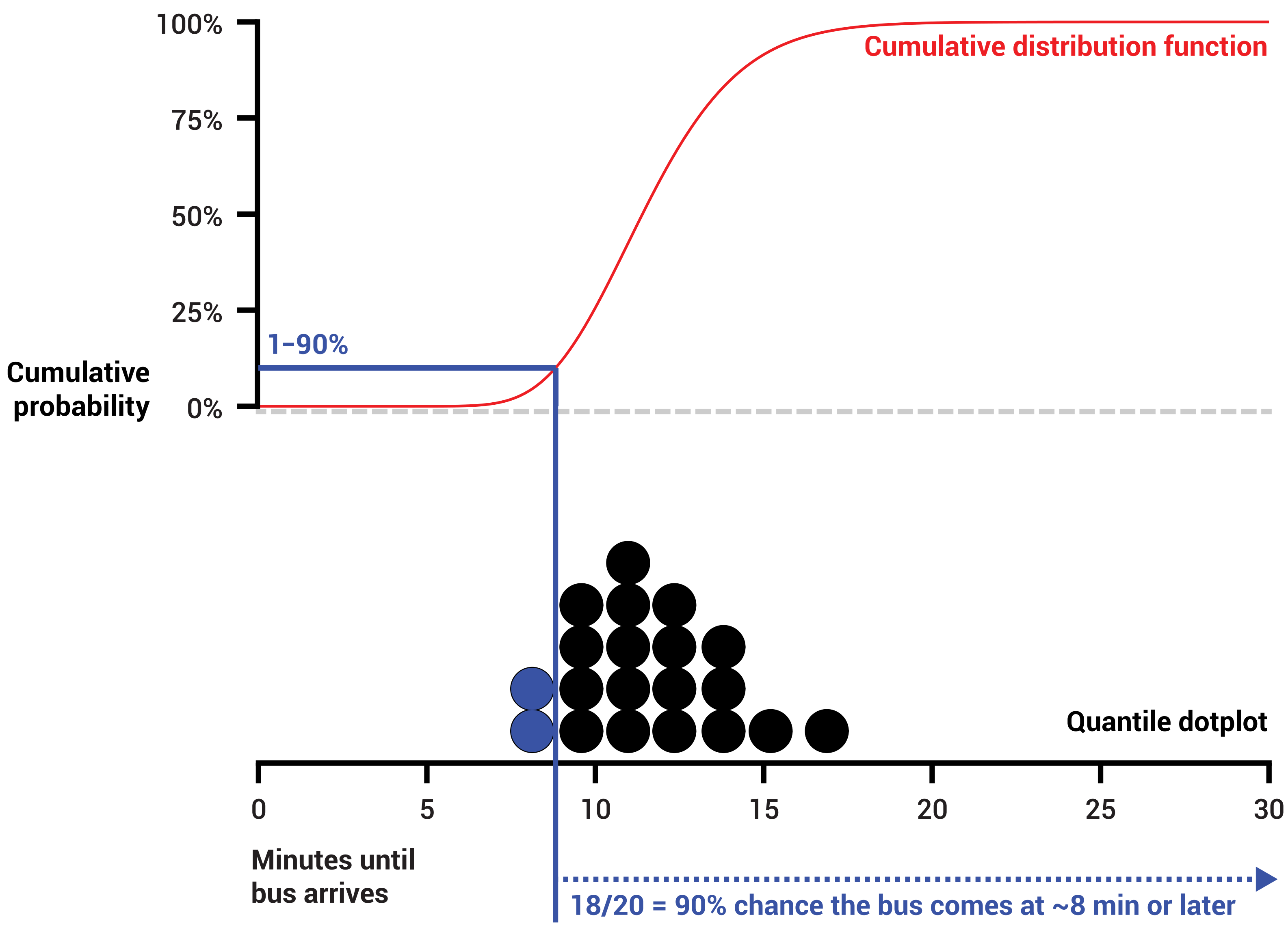

100%
75%
50%
25%
0%
Cumulative probability
Cumulative distribution function
0
5
10
15
20
25
30
0
5
10
15
20
25
30
Minutes until bus arrives


100%
75%
50%
25%
0%
Cumulative probability
Cumulative distribution function
0
5
10
15
20
25
30
0
5
10
15
20
25
30
Minutes until bus arrives



100%
75%
50%
25%
0%
Cumulative probability
Cumulative distribution function
0
5
10
15
20
25
30
Quantile dotplot
0
5
10
15
20
25
30
Minutes until bus arrives


100%
75%
50%
25%
0%
Cumulative probability
Cumulative distribution function
Quantile dotplot
0
5
10
15
20
25
30
Minutes until bus arrives



100%
75%
50%
25%
0%
Cumulative probability
1–90%
Cumulative distribution function
Quantile dotplot
0
5
10
15
20
25
30
Minutes until bus arrives
18/20 = 90% chance the bus comes at ~8 min or later

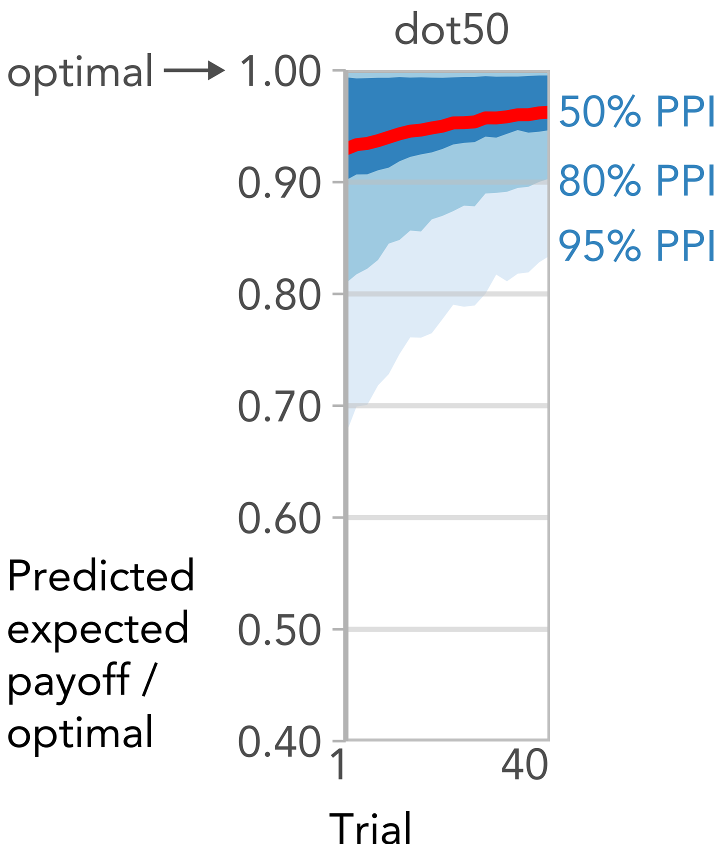
Quantile dotplots

[Kay et al 2016, Fernandes et al 2018]

Better estimates, decisions with time

Variance decreases:
Even worst performers improve

Good uncertainty displays are possible!



Okay, sure, so we should visualize uncertainty.

Okay, sure, so we should visualize uncertainty.

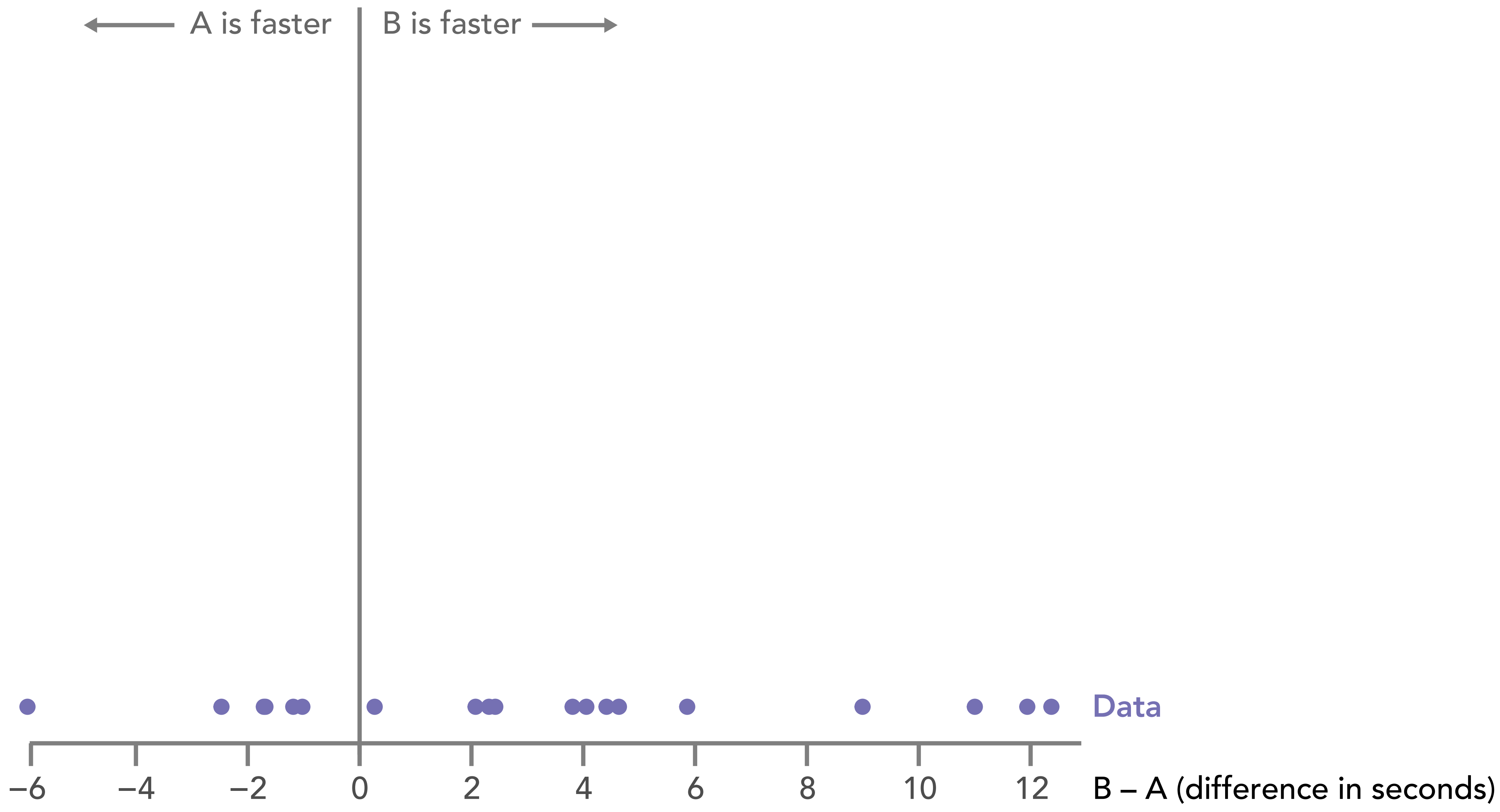
But it's such a pain...

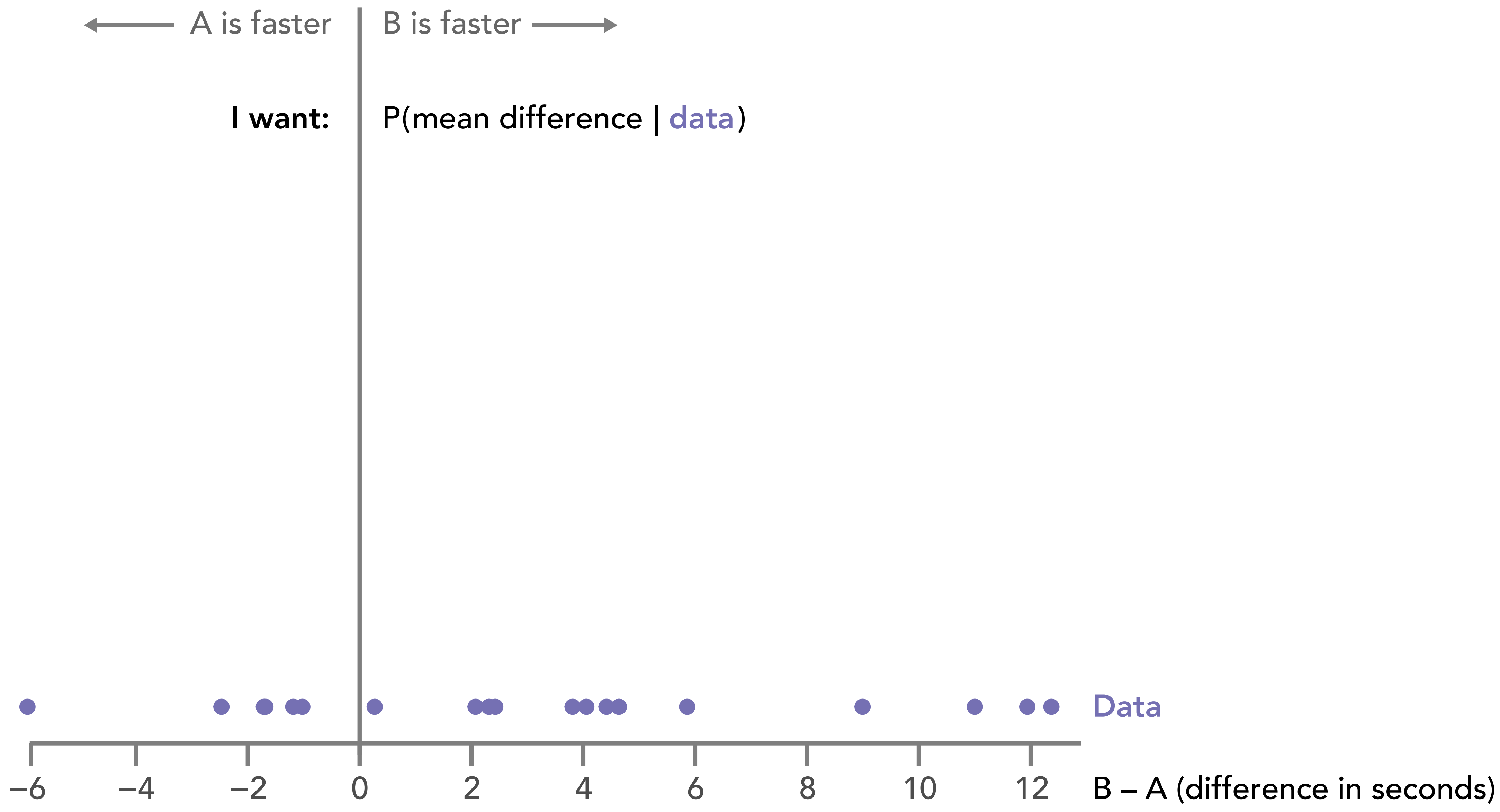
Building uncertainty displays the fun way

Use:

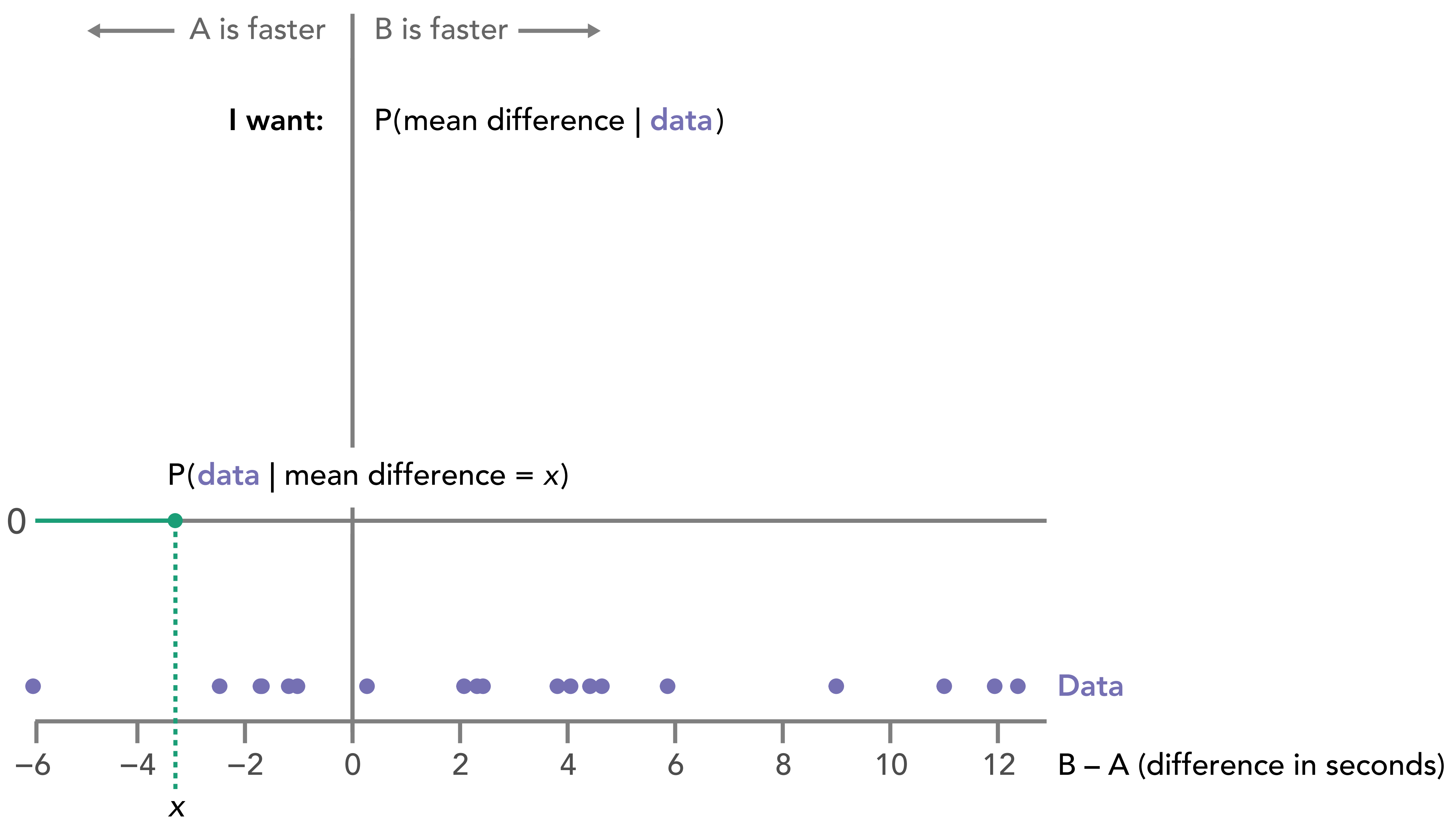
1. Bayesian analysis to get samples from distributions
2. Tidy data to organize those samples
3. Grammar of graphics to visualize samples easily

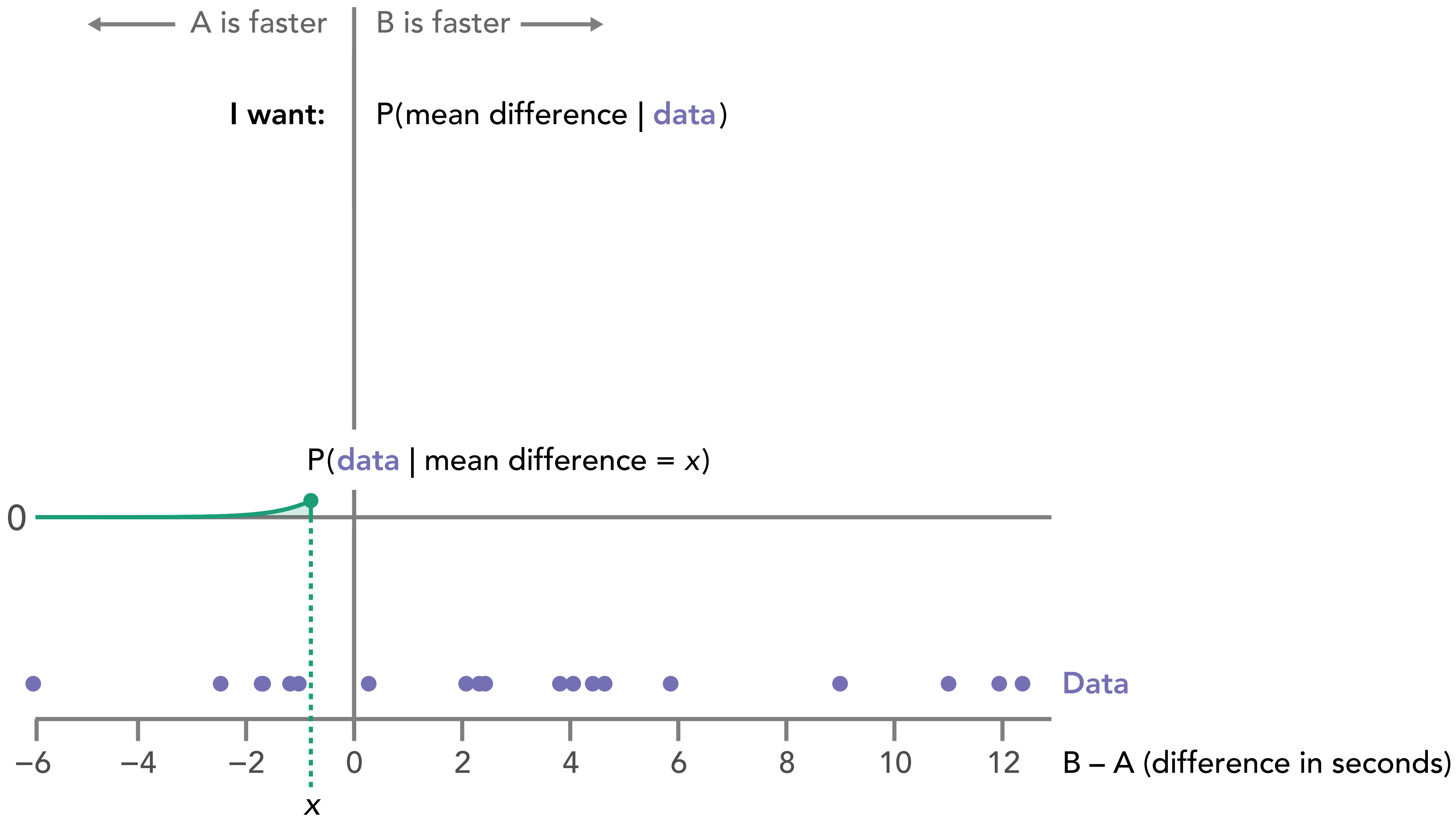
Step 1. Bayesian analysis


A is faster
B is faster
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

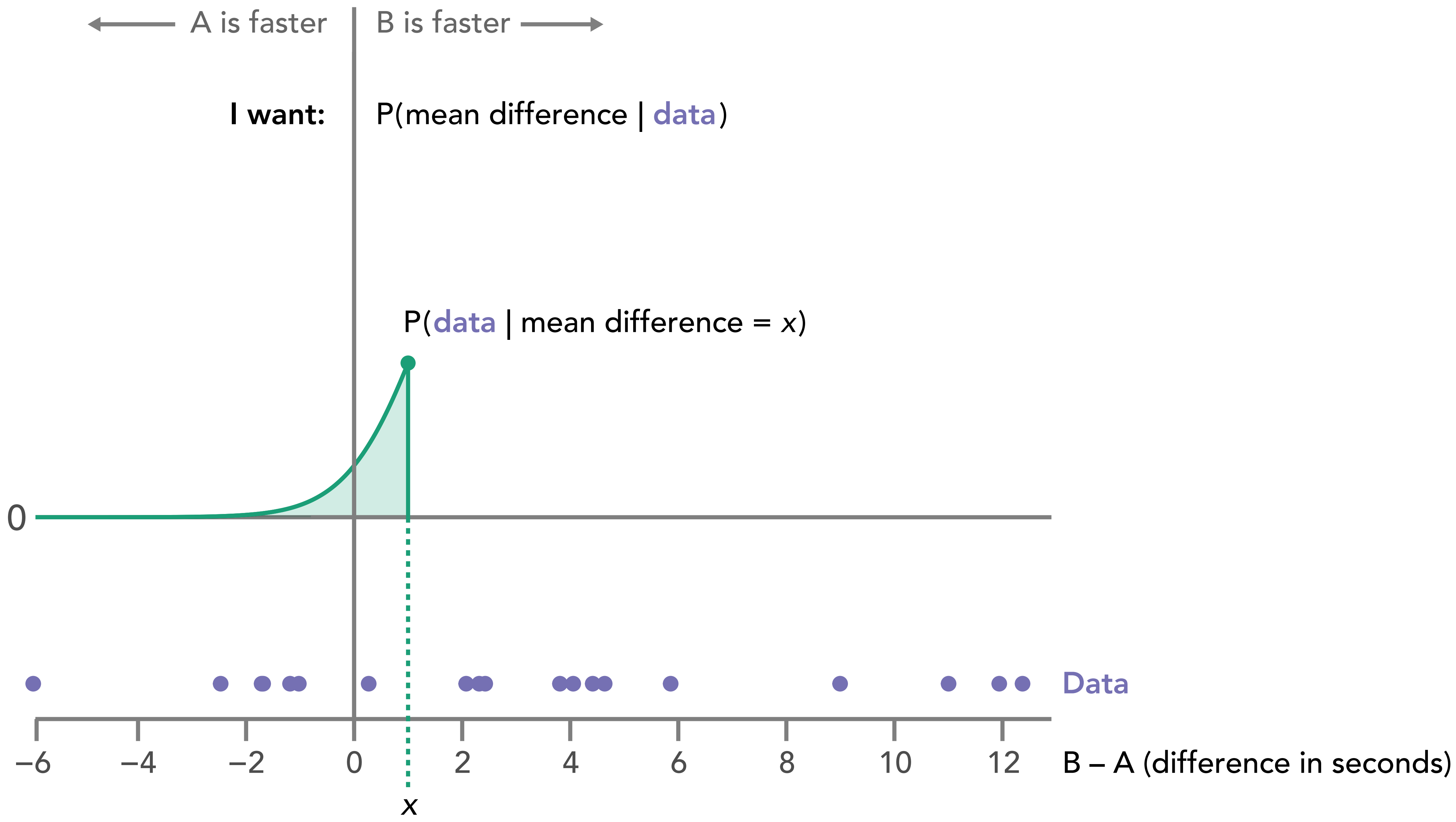


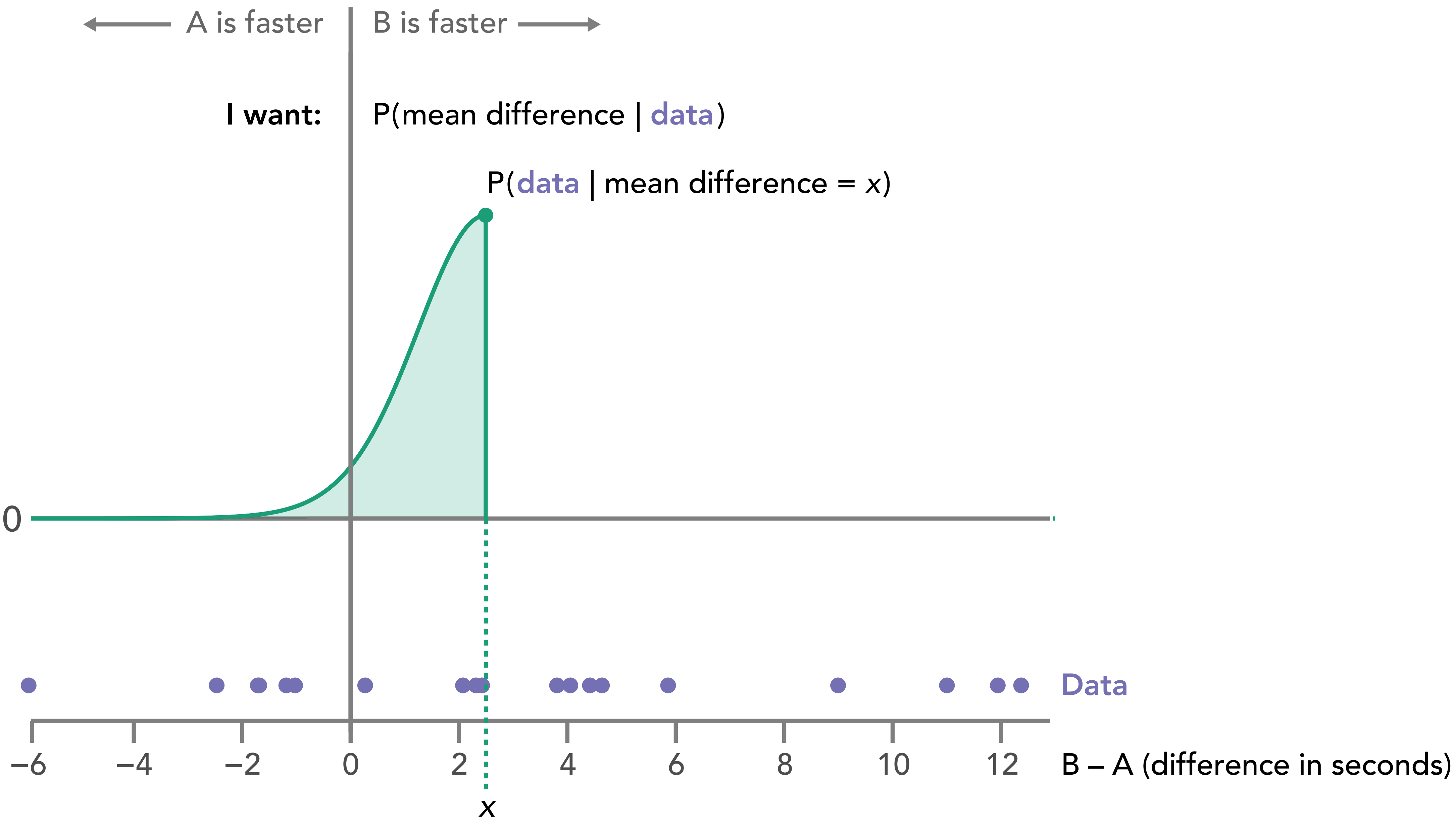
A is faster
B is faster
I want:
P(mean difference | data)
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

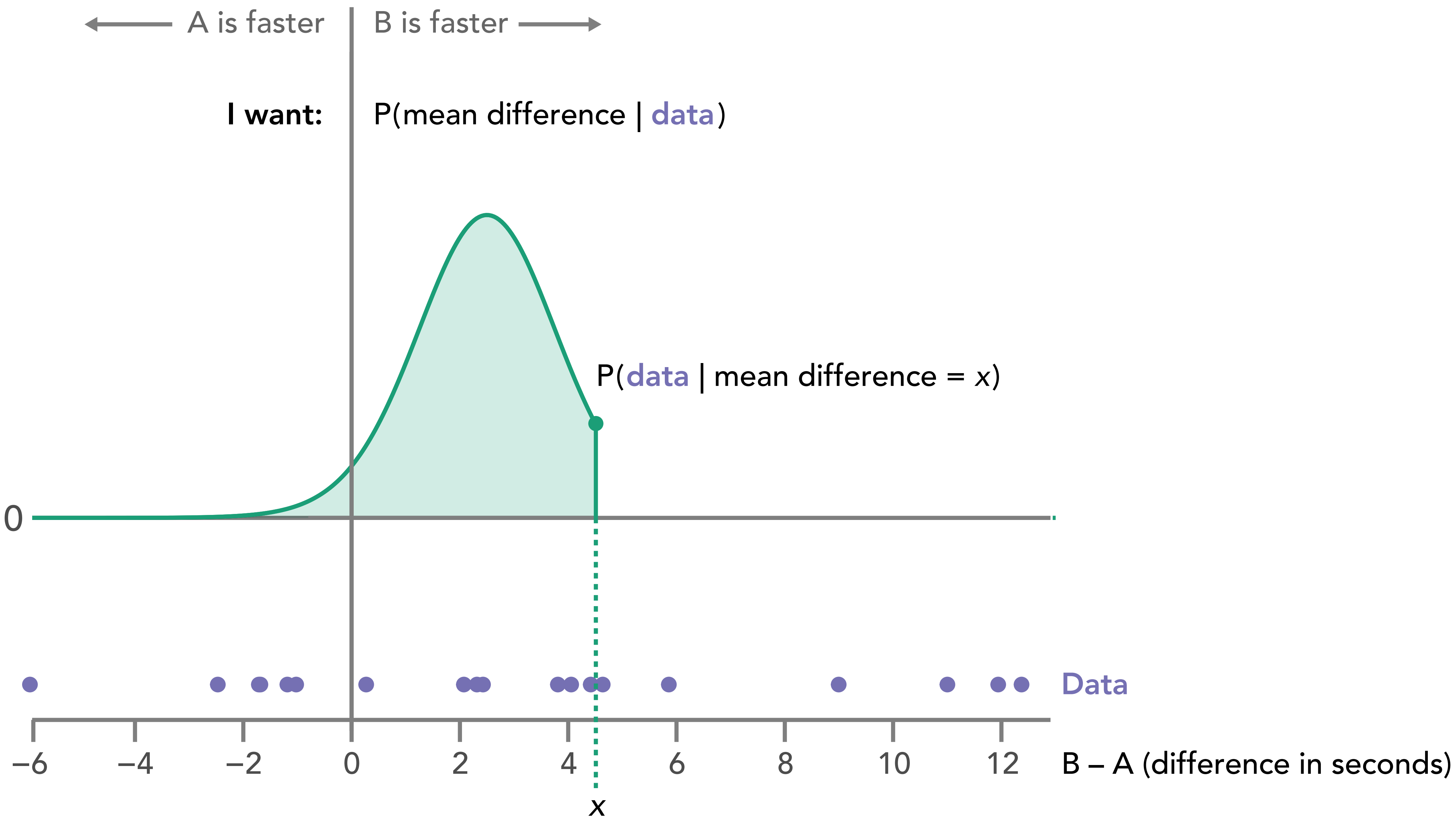

A is faster
B is faster
I want:
P(mean difference | data)
P(data | mean difference = x)
0
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)
x

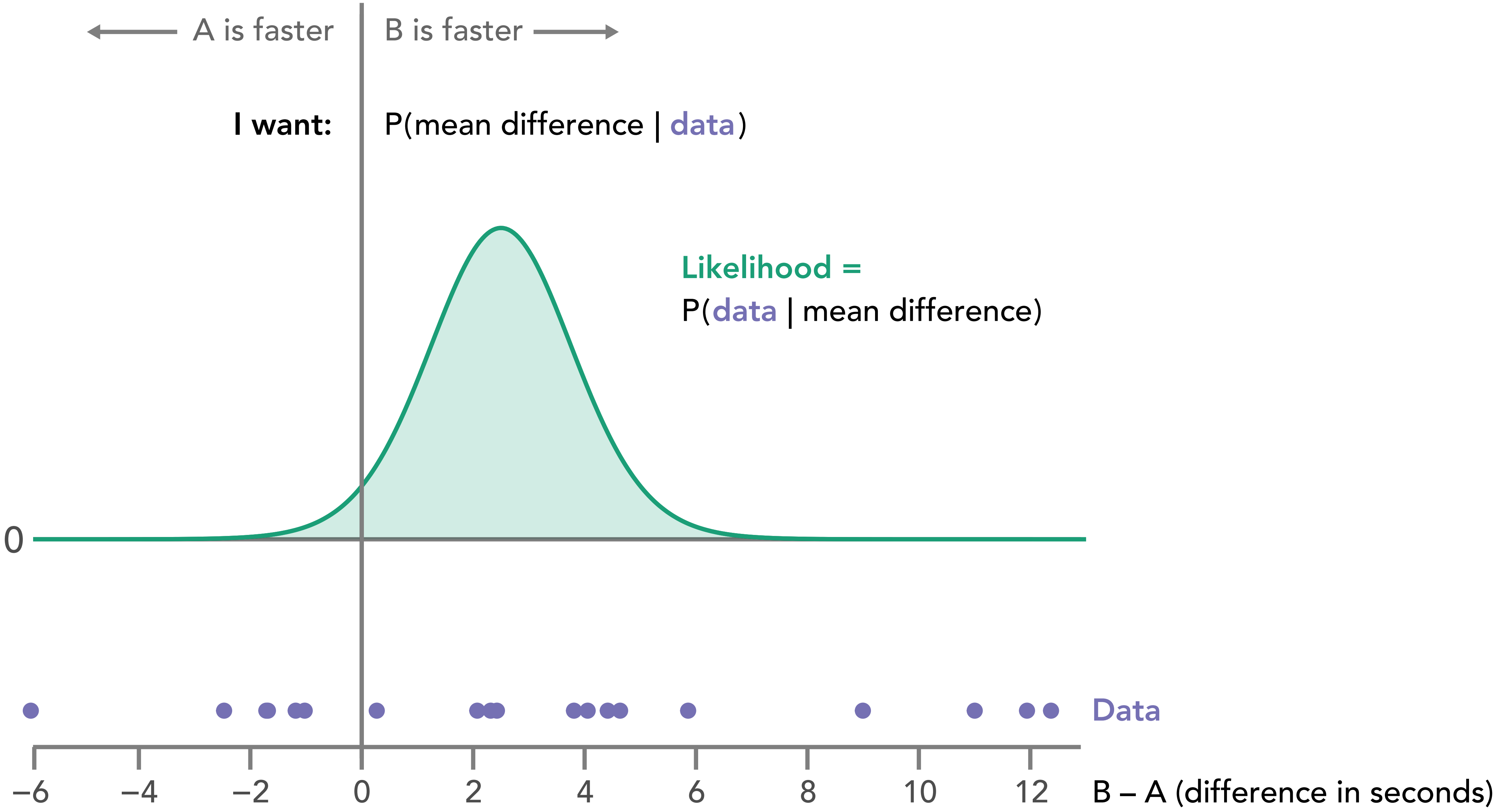


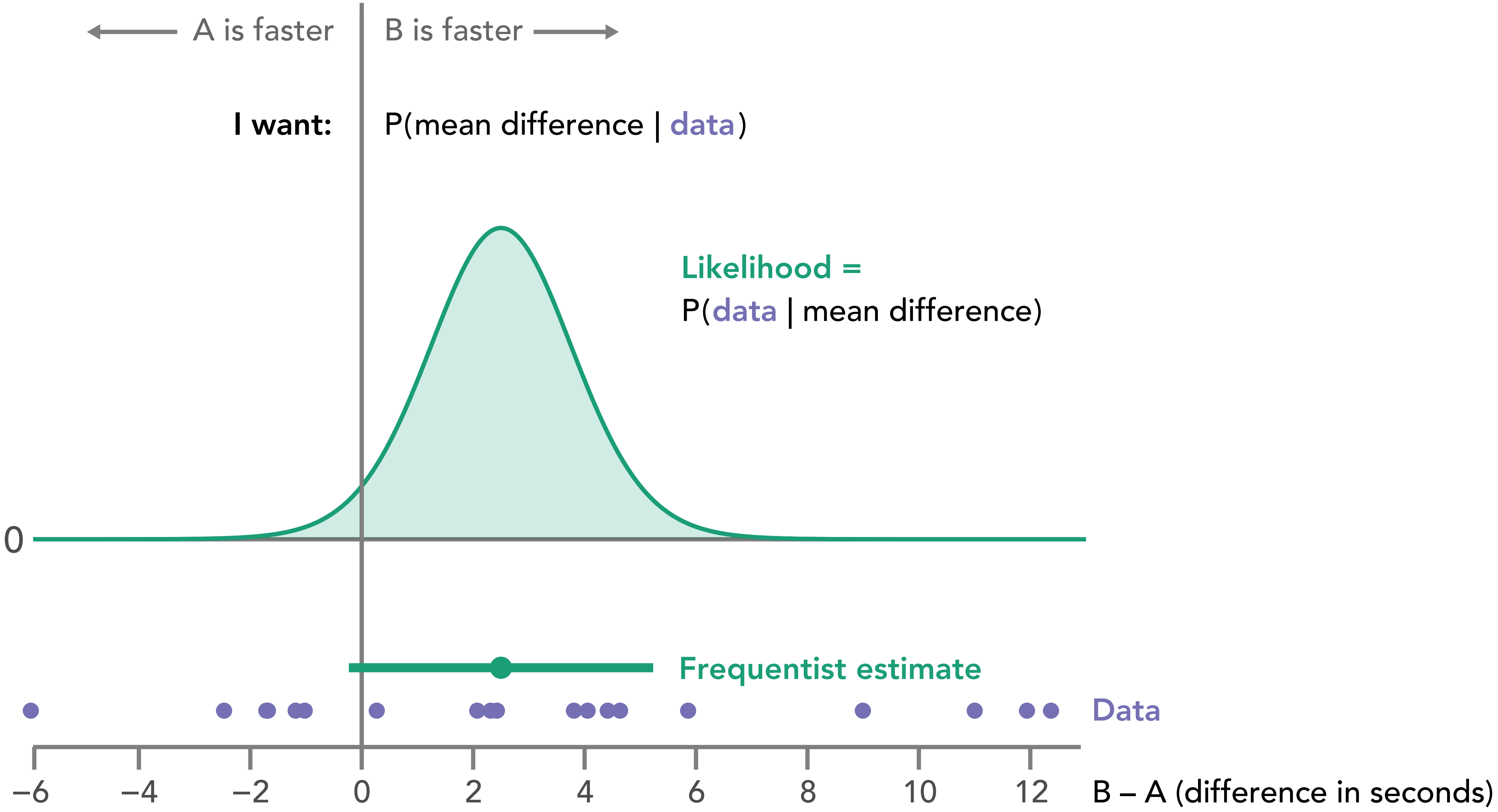
A is faster
B is faster
I want:
P(mean difference | data)
P(data | mean difference = x)
0
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)
x


A is faster
B is faster
I want:
P(mean difference | data)
P(data | mean difference = x)
0
Data
−6
−4
−2
0
2
4
6
8
10
12
x
B – A (difference in seconds)

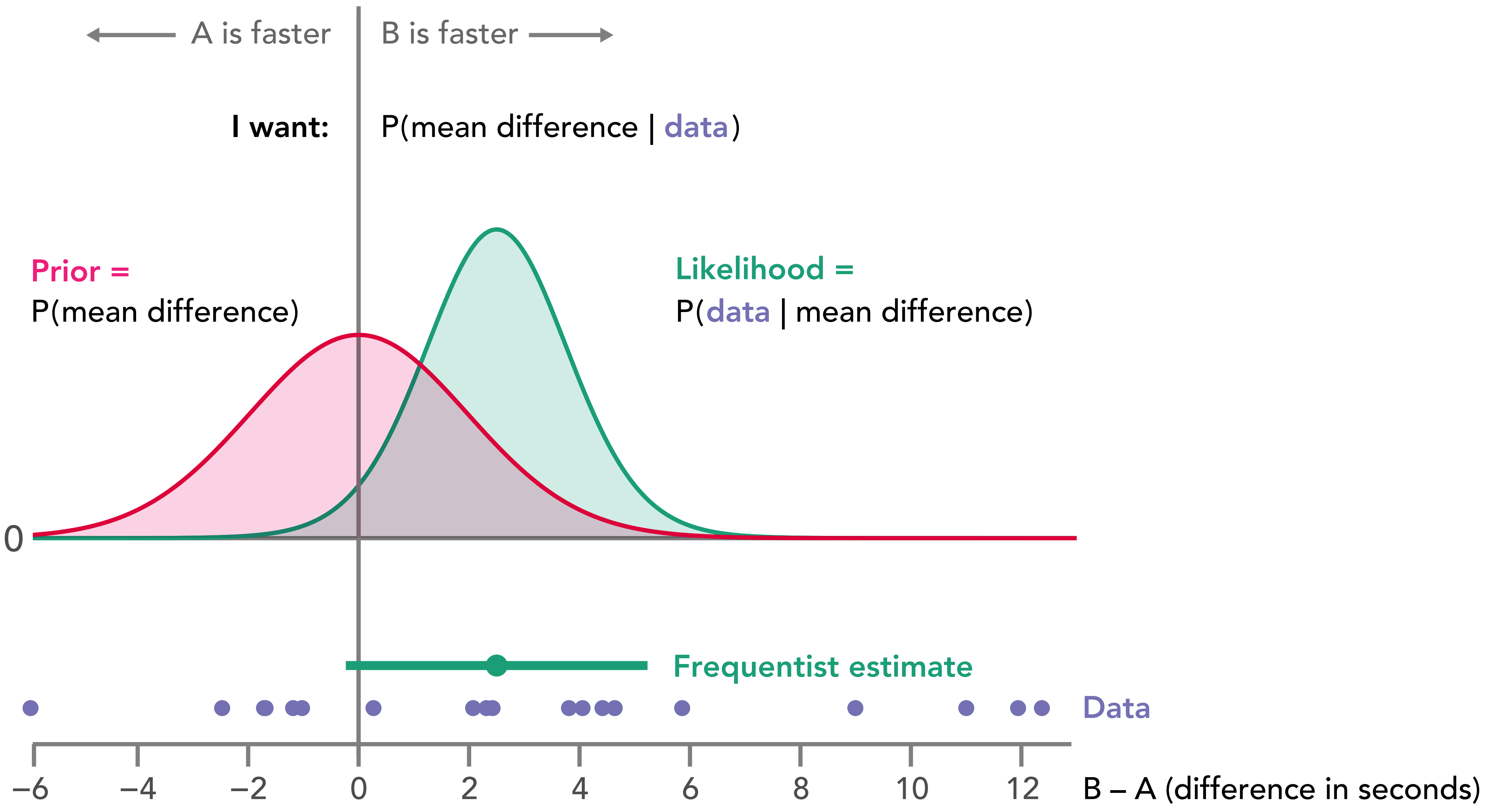

A is faster
B is faster
I want:
P(mean difference | data)
P(data | mean difference = x)
0
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)
x

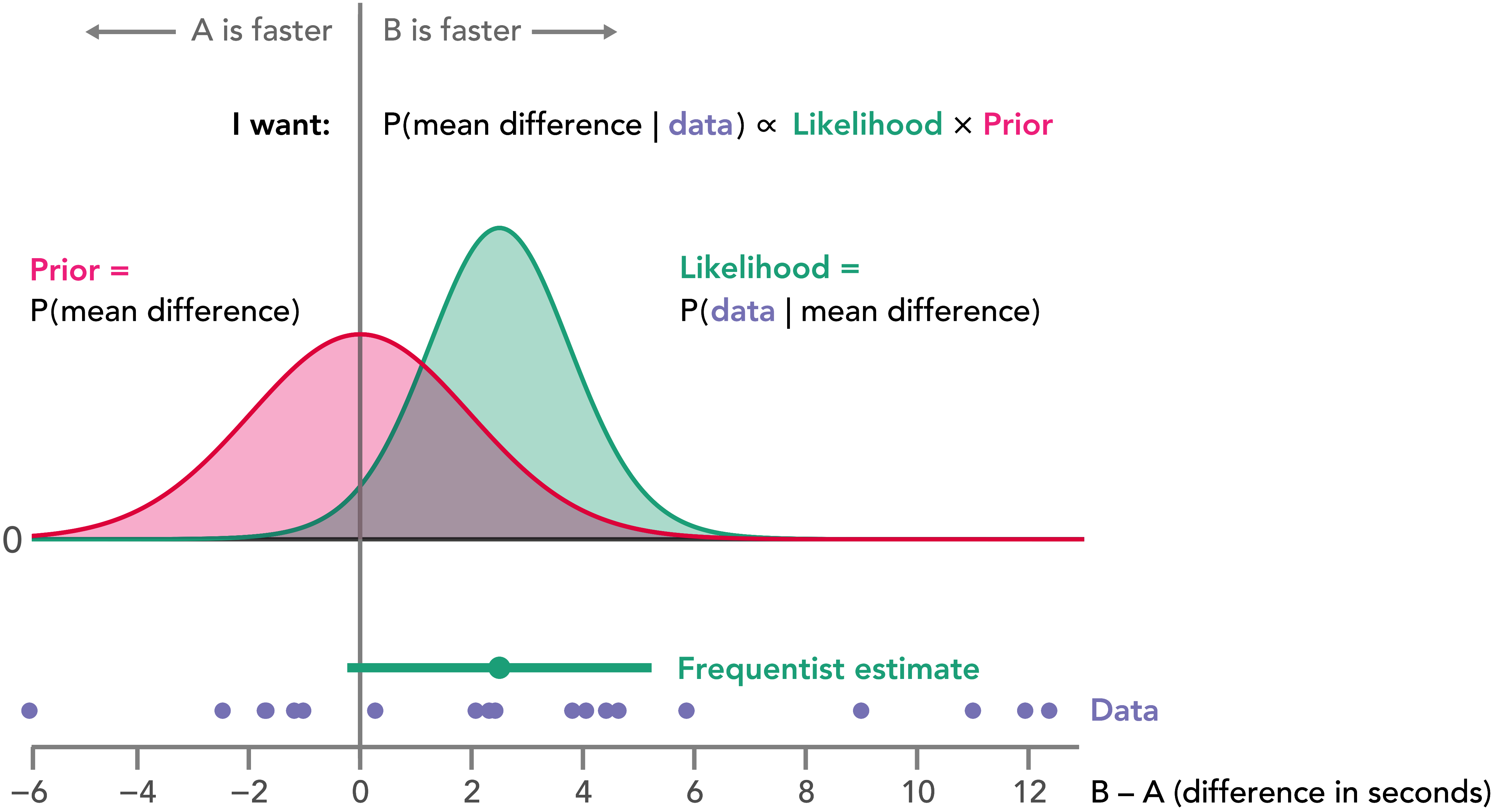

A is faster
B is faster
I want:
P(mean difference | data)
P(data | mean difference = x)
0
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)
x

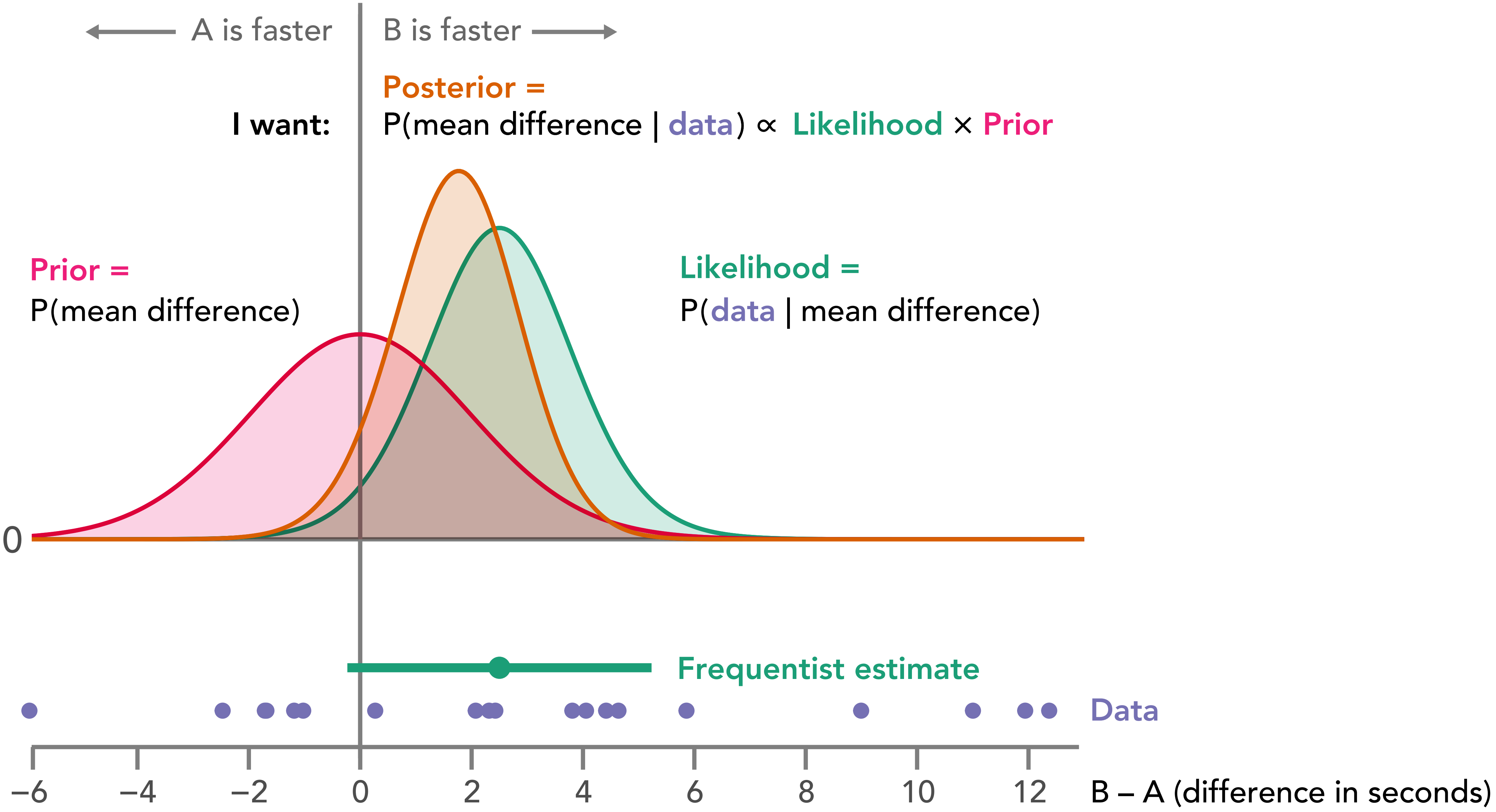

A is faster
B is faster
I want:
P(mean difference | data)
Likelihood =
P(data | mean difference)
0
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)



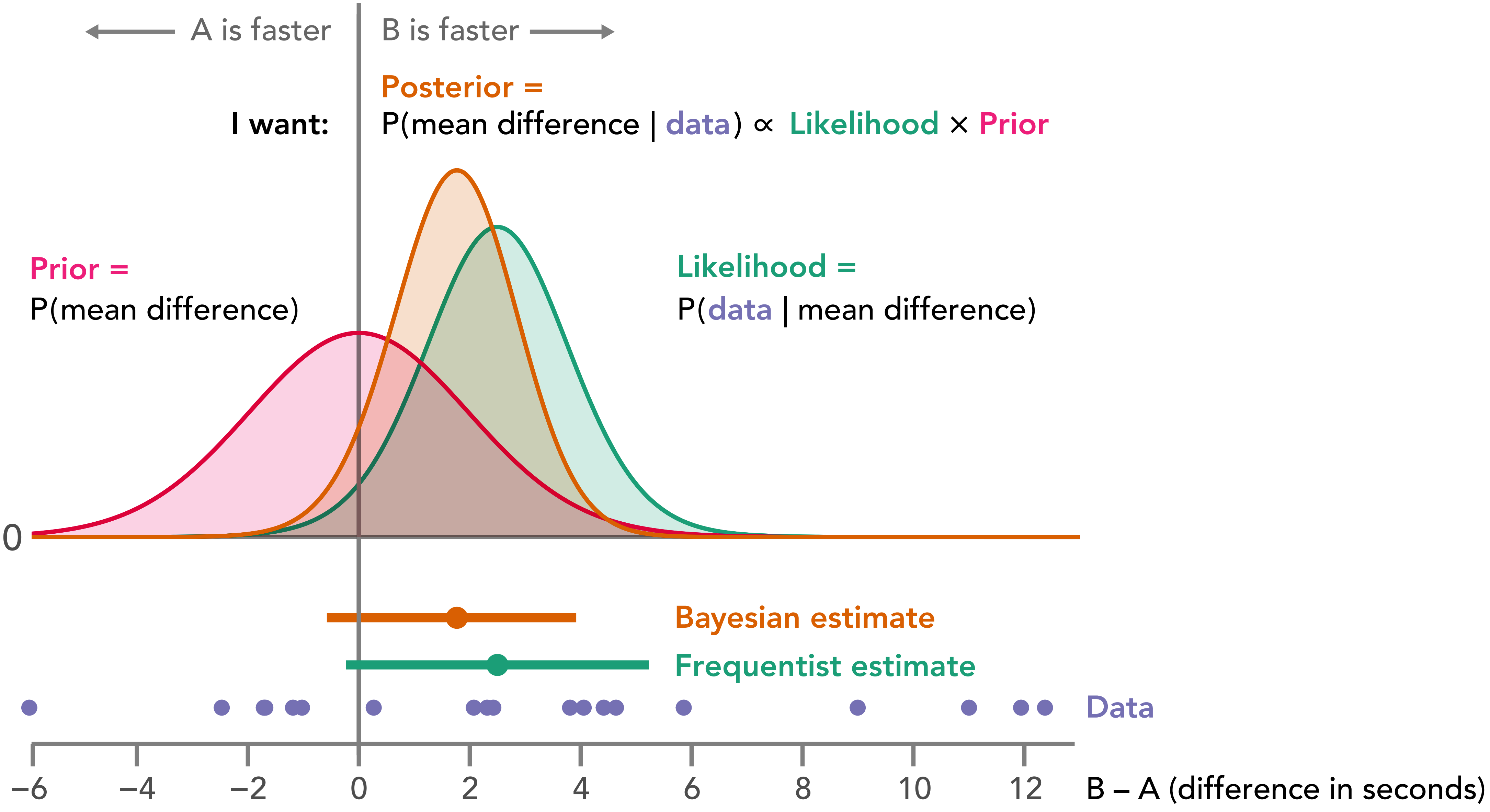
A is faster
B is faster
I want:
P(mean difference | data)
Likelihood =
P(data | mean difference)
0
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

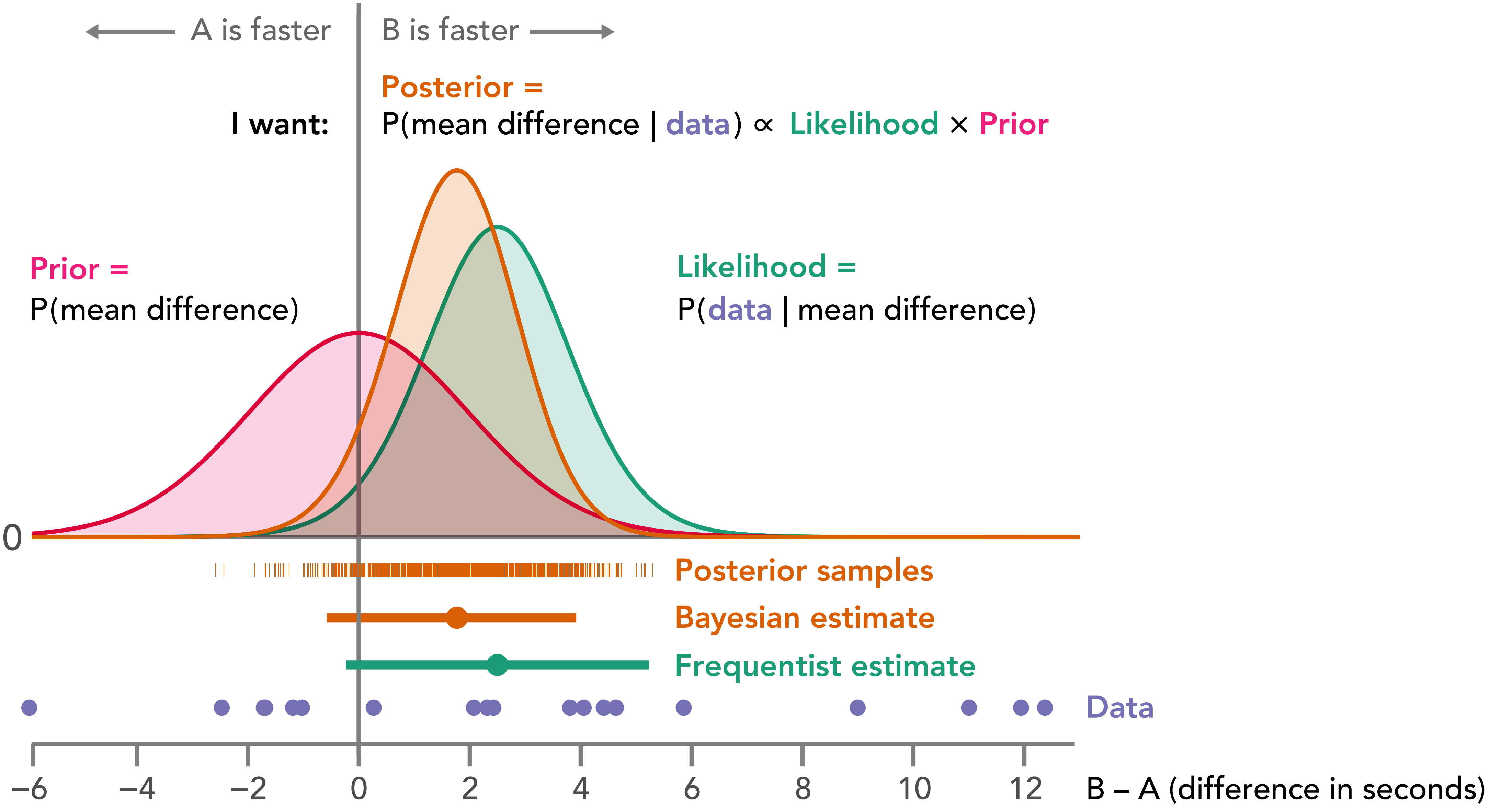

A is faster
B is faster
I want:
P(mean difference | data)
Prior =
P(mean difference)
Likelihood =
P(data | mean difference)
0
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

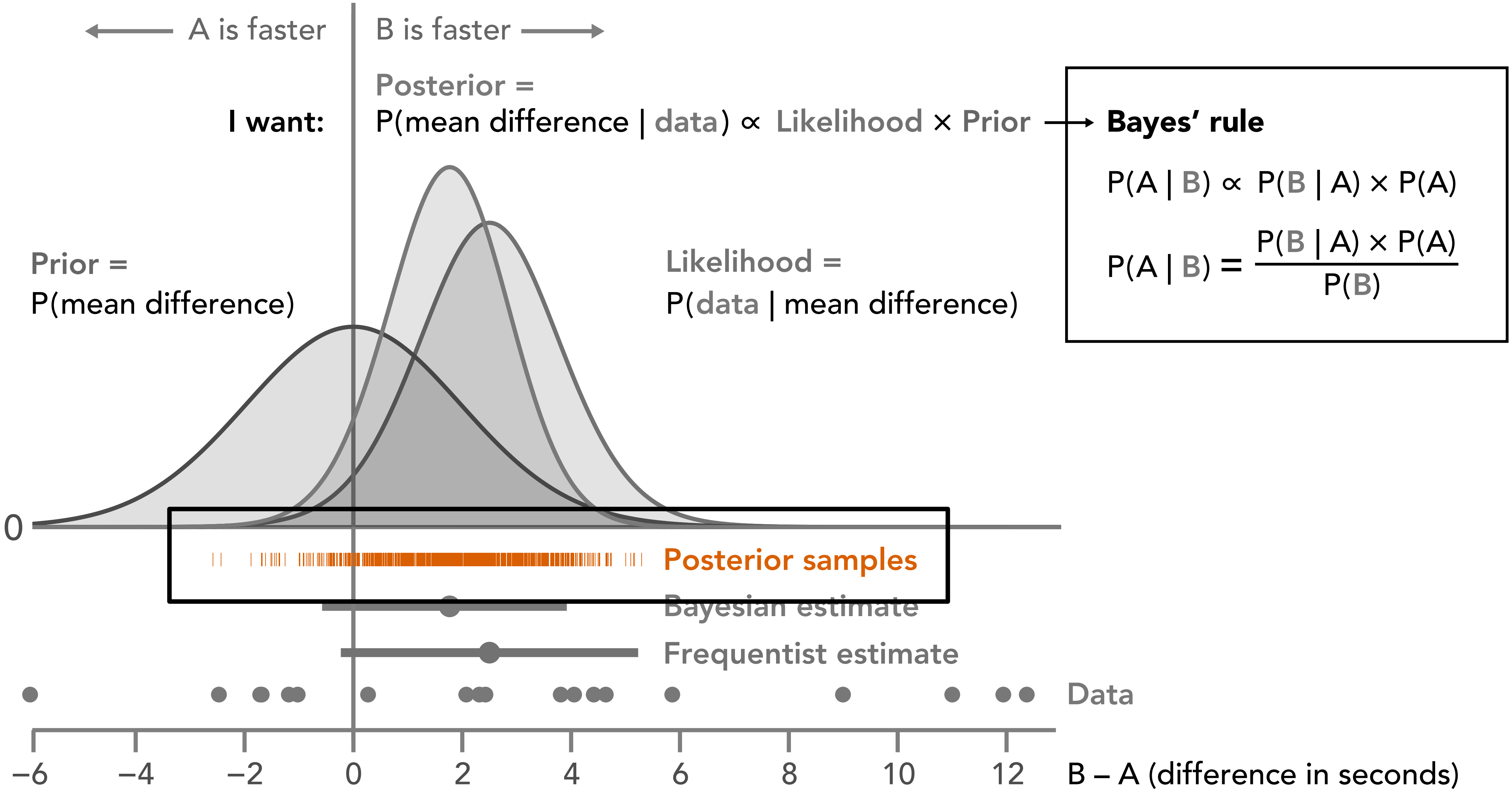

A is faster
B is faster
I want:
P(mean difference | data) ∝ Likelihood × Prior
Prior =
P(mean difference)
Likelihood =
P(data | mean difference)
0
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)



A is faster
B is faster
Posterior =
I want:
P(mean difference | data) ∝ Likelihood × Prior
Prior =
P(mean difference)
Likelihood =
P(data | mean difference)
0
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

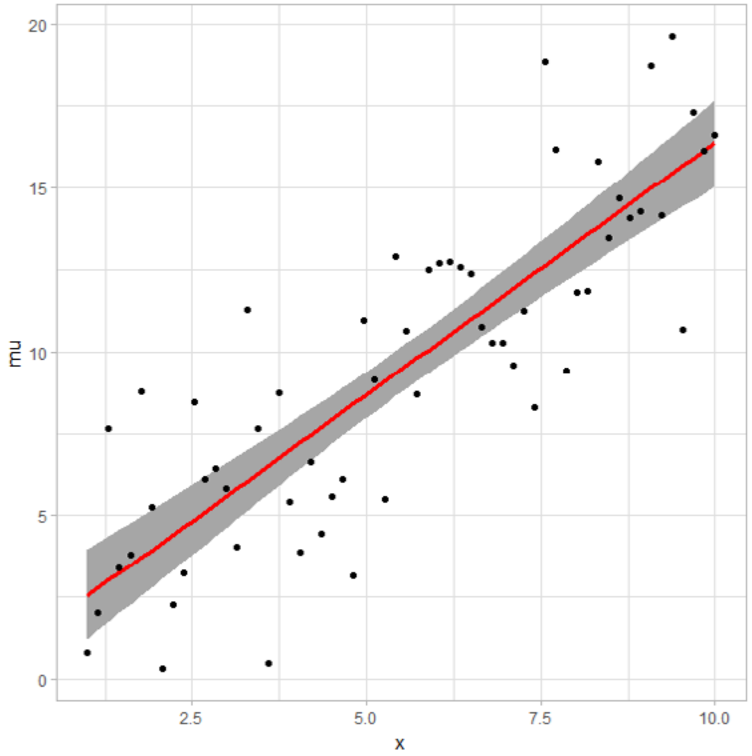

A is faster
B is faster
Posterior =
I want:
P(mean difference | data) ∝ Likelihood × Prior
Prior =
P(mean difference)
Likelihood =
P(data | mean difference)
0
Bayesian estimate
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)


A is faster
B is faster
I want:
Posterior =
P(mean difference | data) ∝ Likelihood × Prior
Prior =
P(mean difference)
Likelihood =
P(data | mean difference)
0
Posterior samples
Bayesian estimate
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

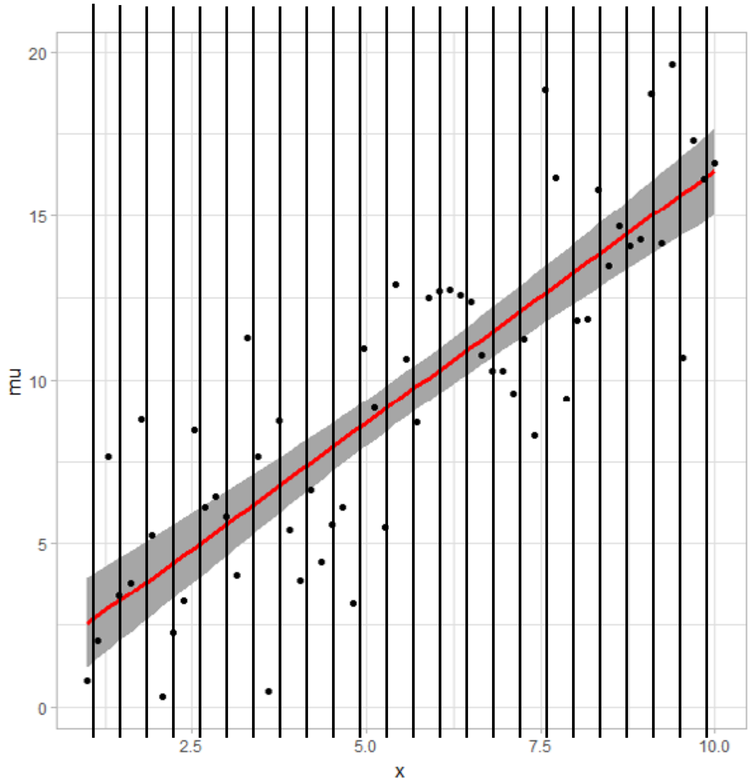

A is faster
B is faster
I want:
Posterior =
P(mean difference | data) ∝ Likelihood × Prior
Bayes' rule
$P(A \mid B) \propto P(B \mid A) \times P(A)$
$P(A \mid B) = \dfrac{P(B \mid A) \times P(A)}{P(B)}$
Prior =
P(mean difference)
Likelihood =
P(data | mean difference)
0
Posterior samples
Bayesian estimate
Frequentist estimate
Data
−6
−4
−2
0
2
4
6
8
10
12
B – A (difference in seconds)

Step 2. Tidy data, tidy samples

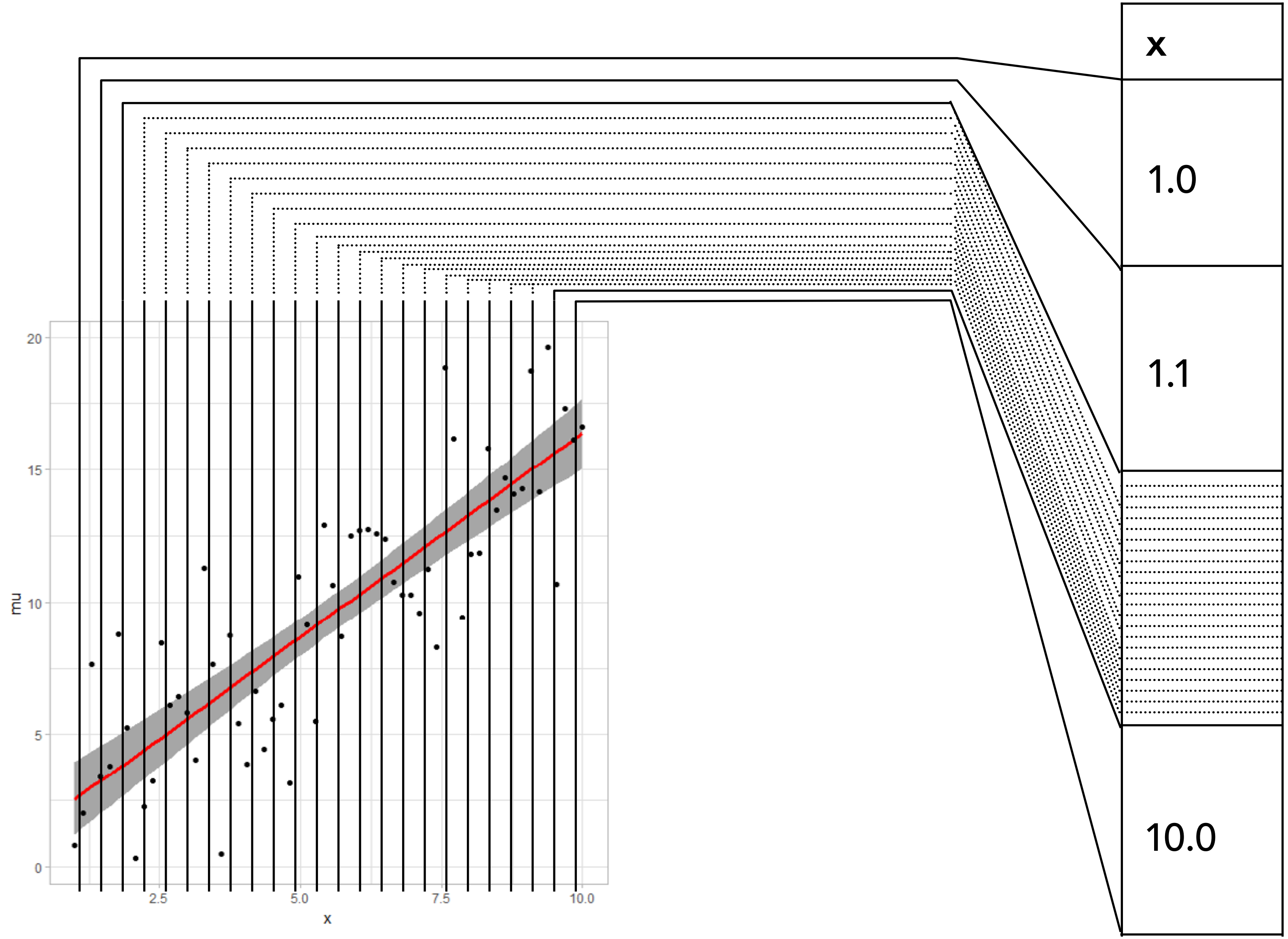
How to build this chart



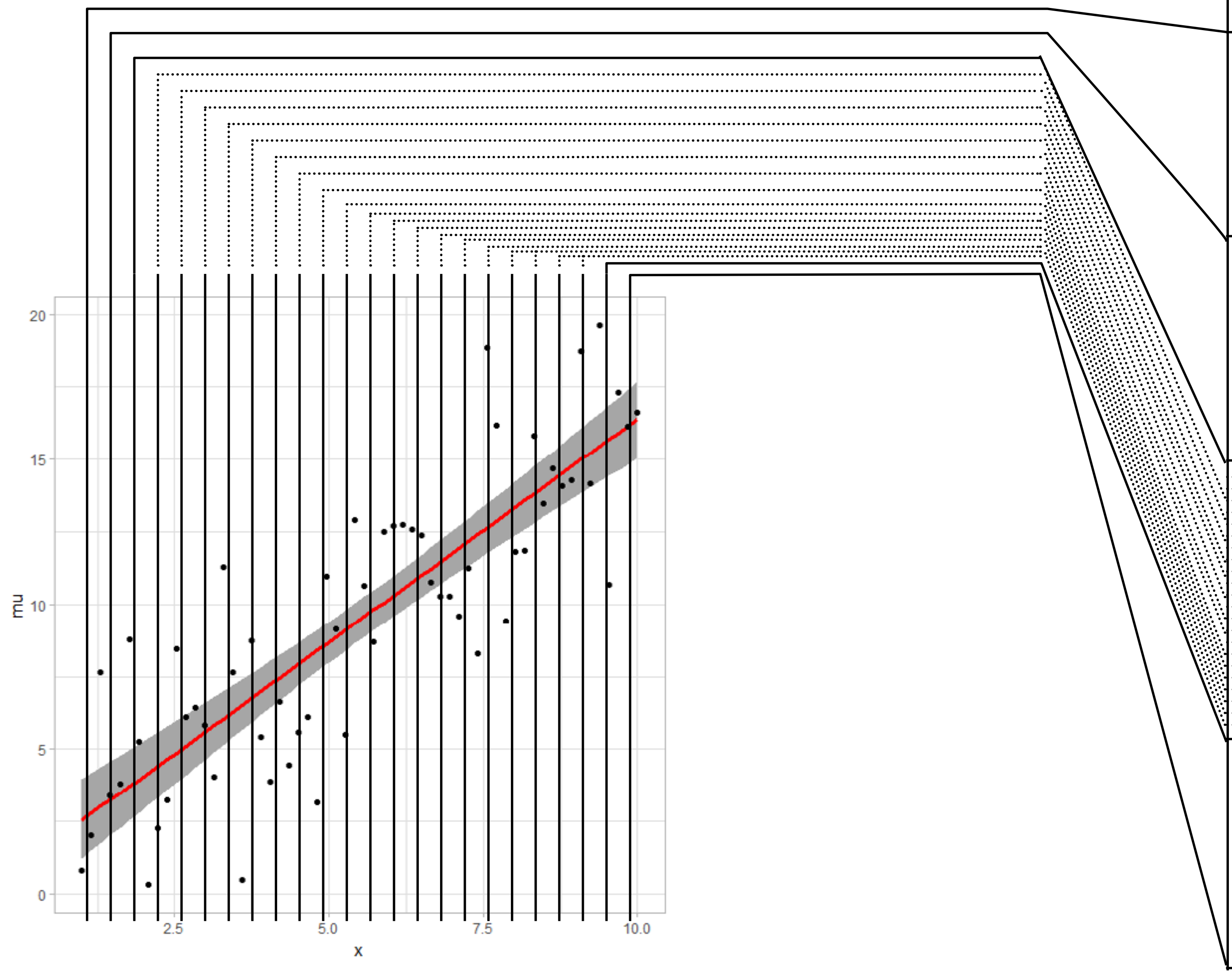
How to build this chart



How to build this chart



How to build this chart

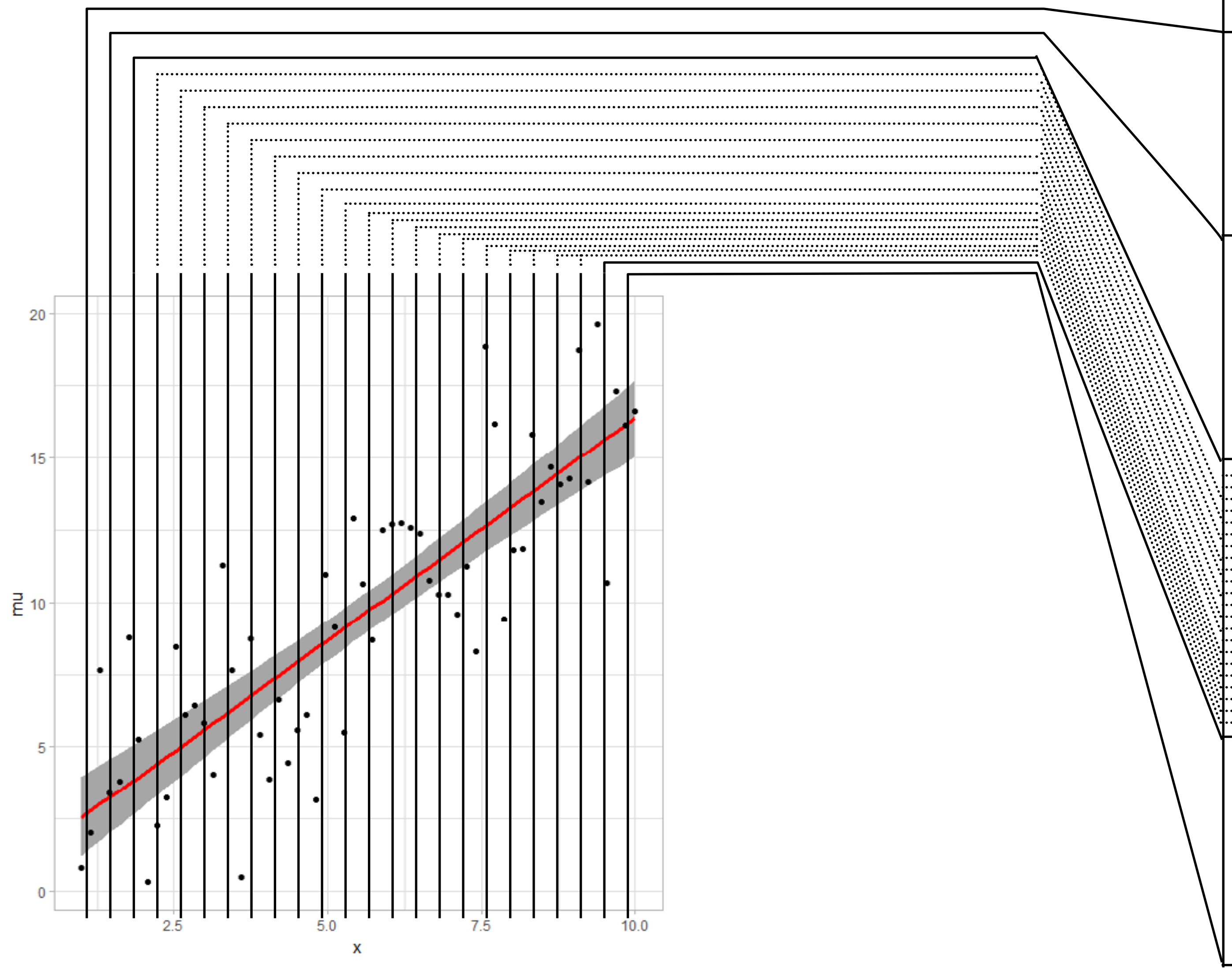


Predicted from the model

x	Mean at x
1.0	2.6
1.1	2.7
10.0	17.2

How to build this chart

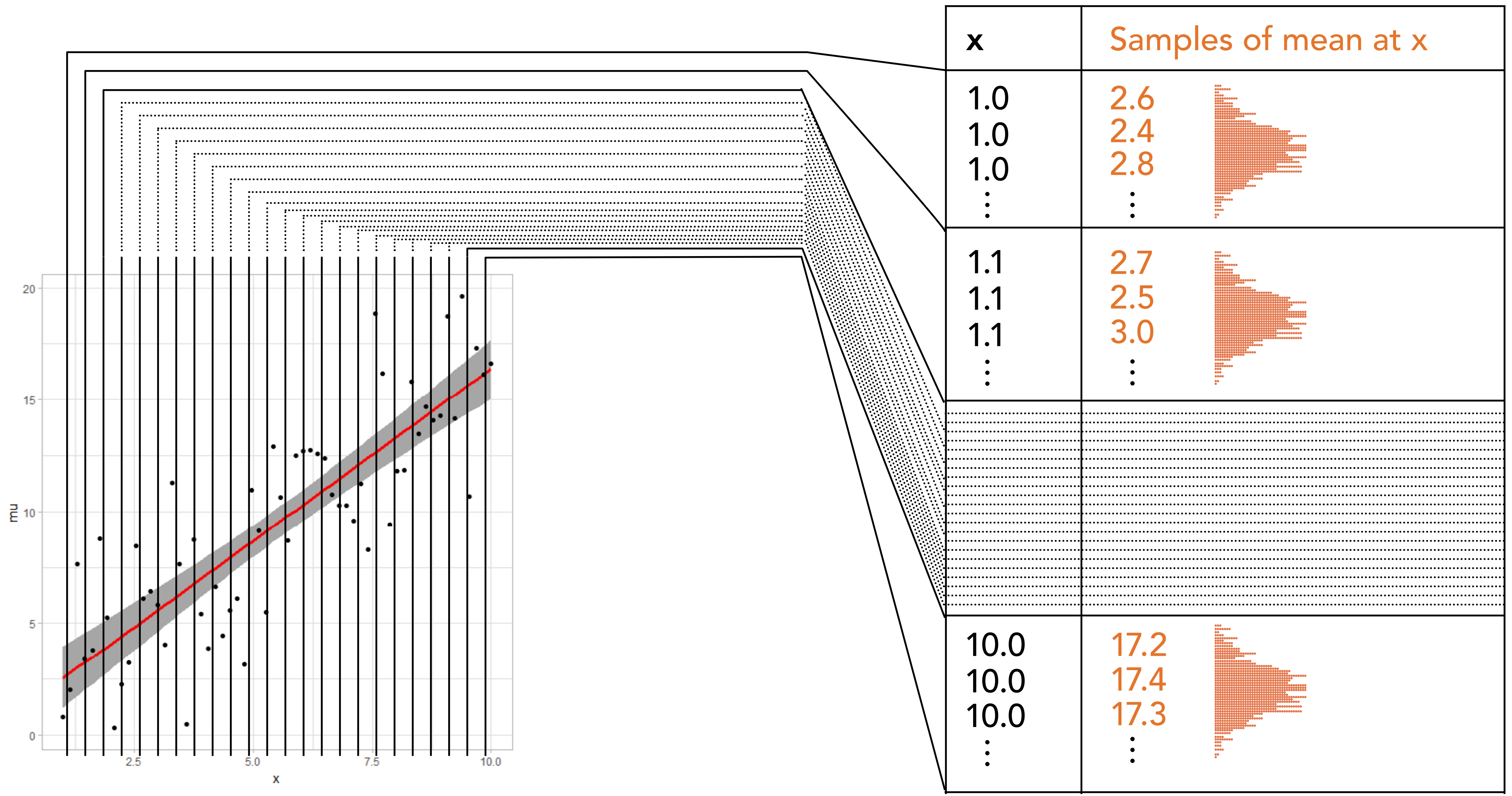
Predicted from the model



x	Samples of mean at x
1.0 1.0 1.0 ⋮	2.6 2.4 2.8 ⋮
1.1 1.1 1.1 ⋮	2.7 2.5 3.0 ⋮
10.0 10.0 10.0 ⋮	17.2 17.4 17.3 ⋮

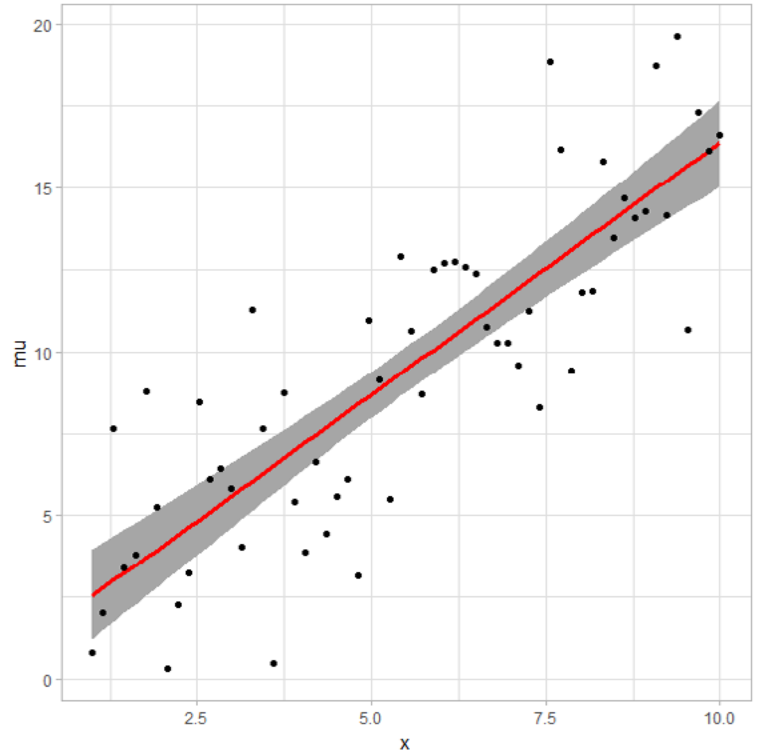
How to build this chart

Predictors Samples from fits or predictions



How to build this chart

tidybayes (R package) outputs tables like this given model + table of predictors



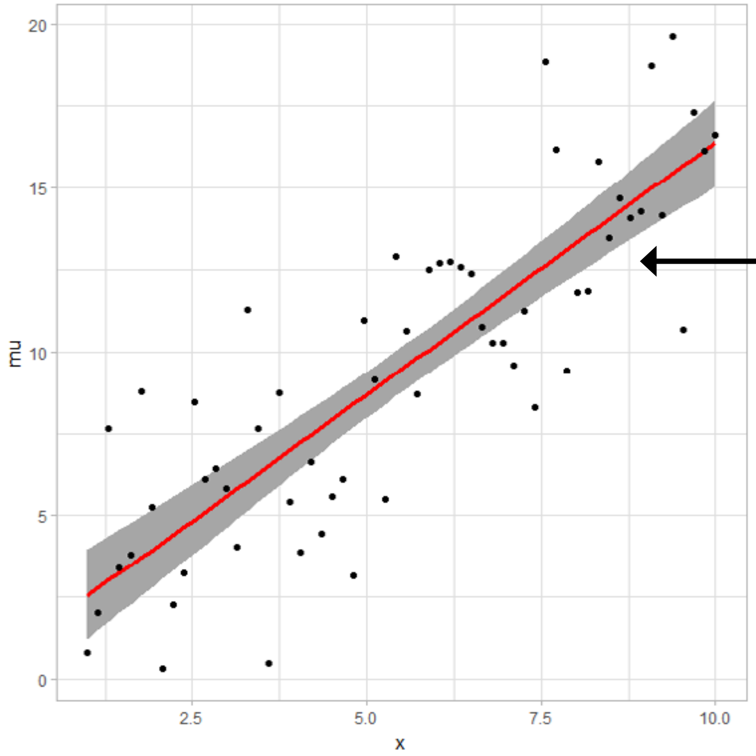
Predictors | Samples from fits or predictions

x	Samples of mean at x
1.0 1.0 1.0 ⋮	2.6 2.4 2.8 ⋮
1.1 1.1 1.1 ⋮	2.7 2.5 3.0 ⋮
⋯	⋯
10.0 10.0 10.0 ⋮	17.2 17.4 17.3 ⋮

How to build this chart

What aggregation do I do in each group?

How do I map onto channels/marks?



?????

Predictors | Samples from fits or predictions

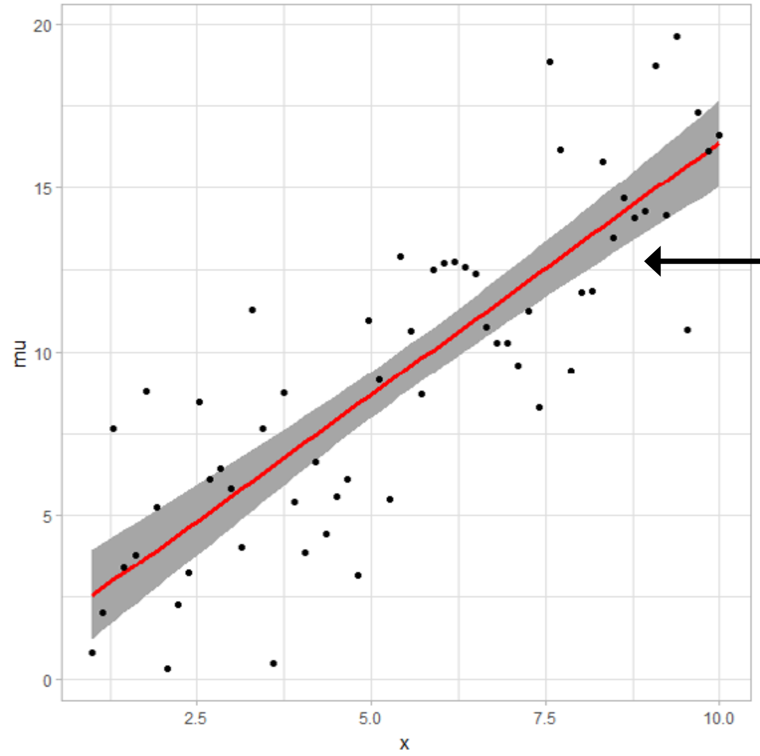
x	Samples of mean at x
1.0	2.6
1.0	2.4
1.0	2.8
⋮	⋮
1.1	2.7
1.1	2.5
1.1	3.0
⋮	⋮
10.0	17.2
10.0	17.4
10.0	17.3
⋮	⋮

Step 3. Grammar of graphics

How to build this chart

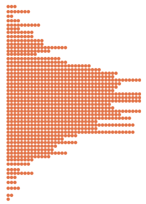
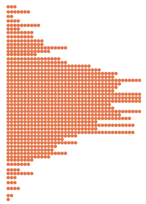
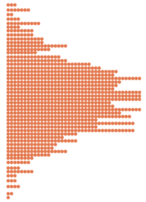
What aggregation do I do in each group?

How do I map onto channels/marks?



?????

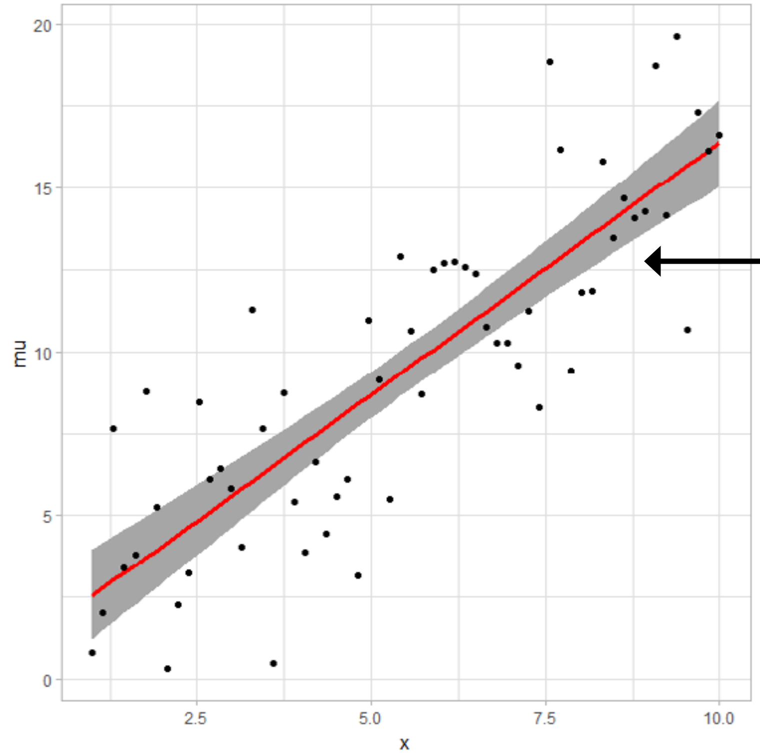
Predictors | Samples from fits or predictions

x	Samples of mean at x	
1.0	2.6	
1.0	2.4	
1.0	2.8	
⋮	⋮	
1.1	2.7	
1.1	2.5	
1.1	3.0	
⋮	⋮	
10.0	17.2	
10.0	17.4	
10.0	17.3	
⋮	⋮	

How to build this chart

What aggregation do I do in each group?

How do I map onto channels/marks?



Take mean,
95% interval

Geom: line +
band

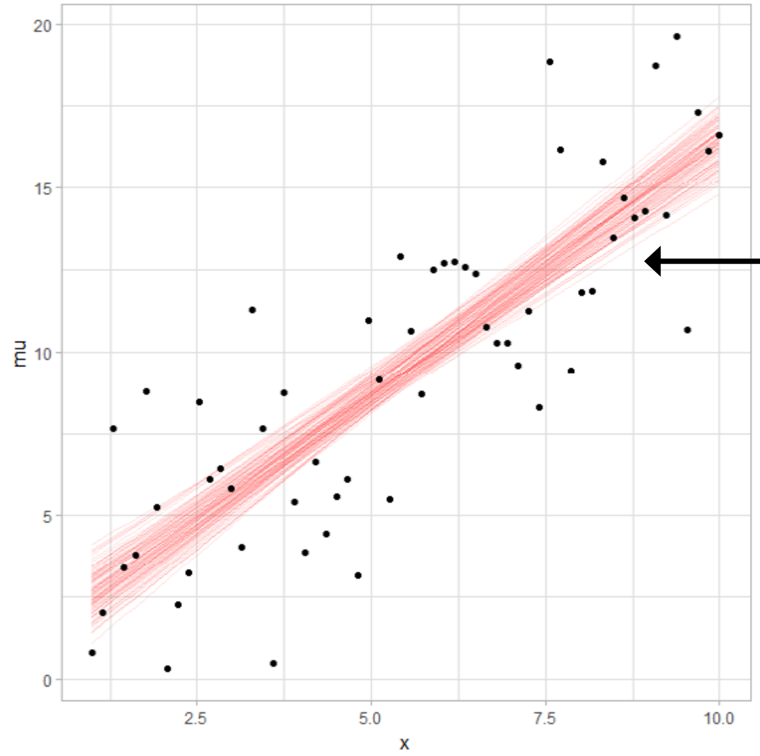
Predictors Samples from fits or predictions

x	Samples of mean at x
1.0 1.0 1.0 ⋮	2.6 2.4 2.8 ⋮
1.1 1.1 1.1 ⋮	2.7 2.5 3.0 ⋮
10.0 10.0 10.0 ⋮	17.2 17.4 17.3 ⋮

How to build this chart

What aggregation do I do in each group?

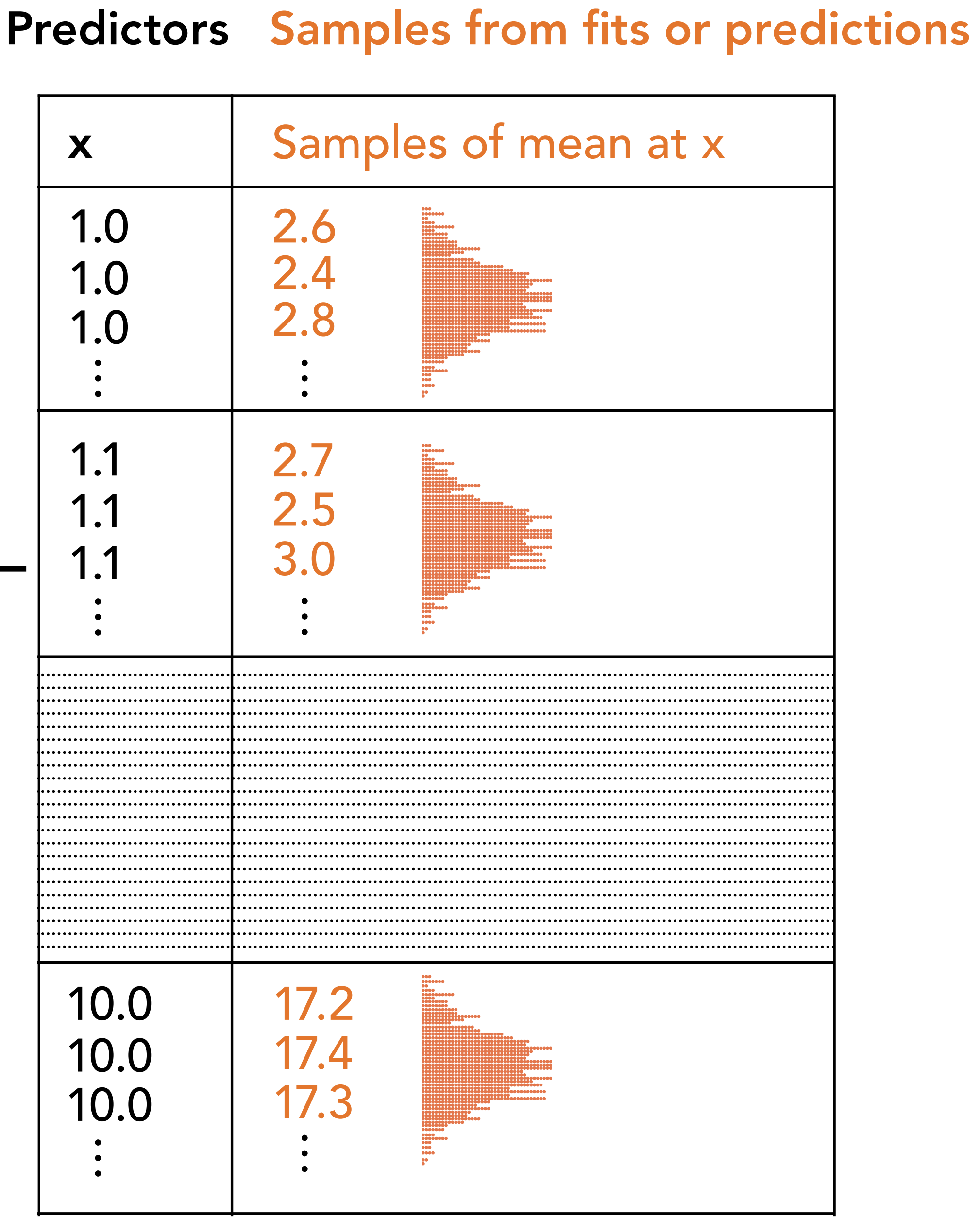
How do I map onto channels/marks?



Take first 100 samples

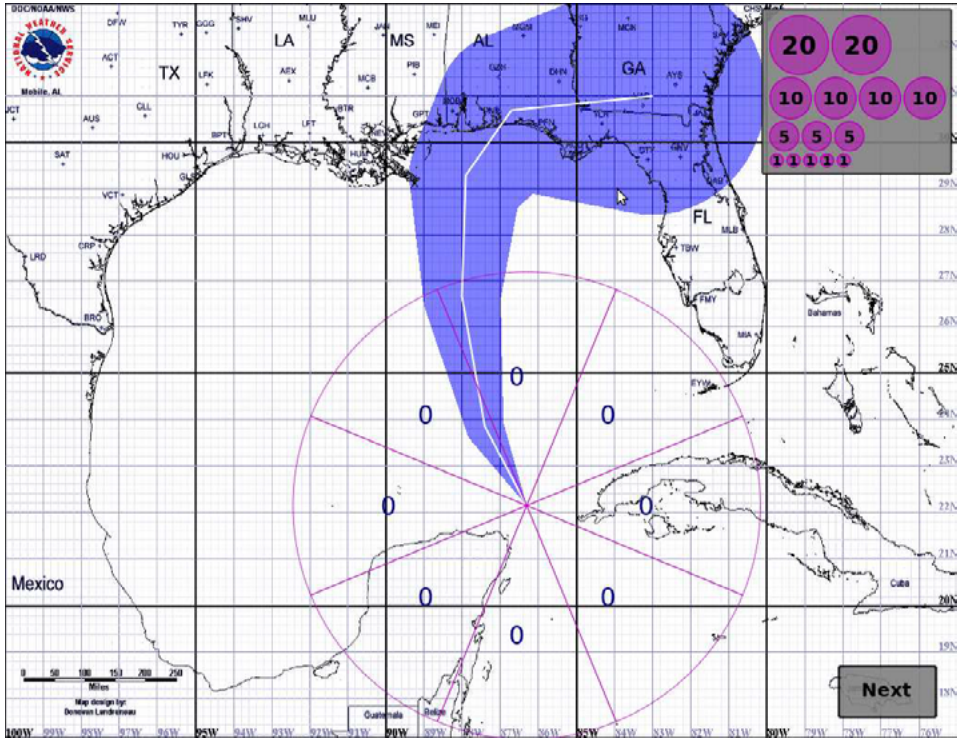
Geom: line

Group by sample



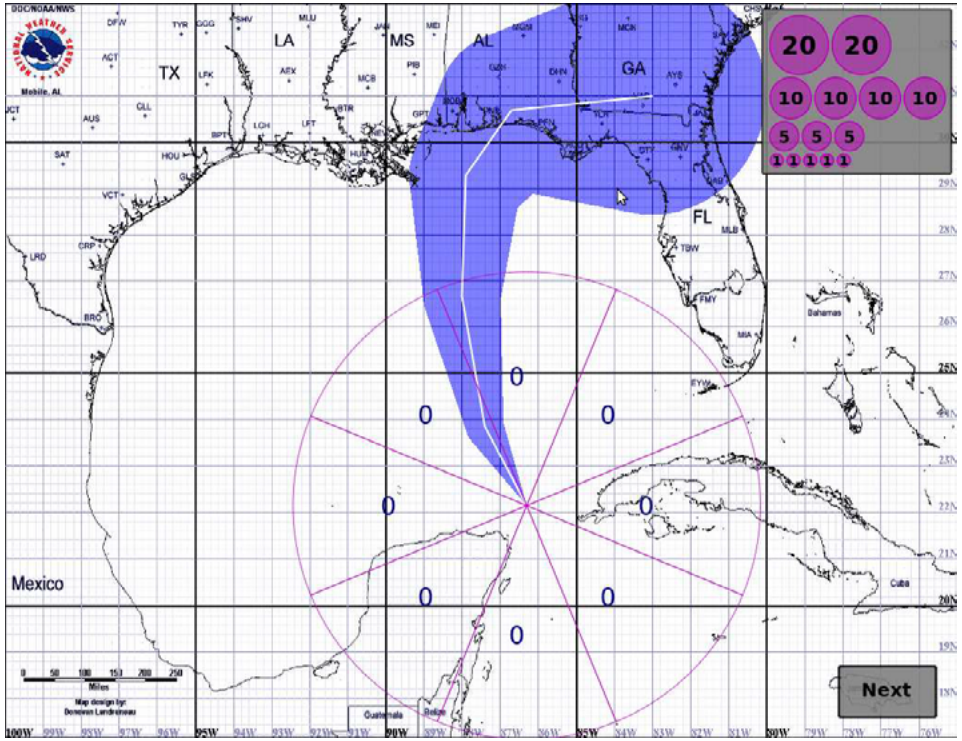
(Hurricane error cones)

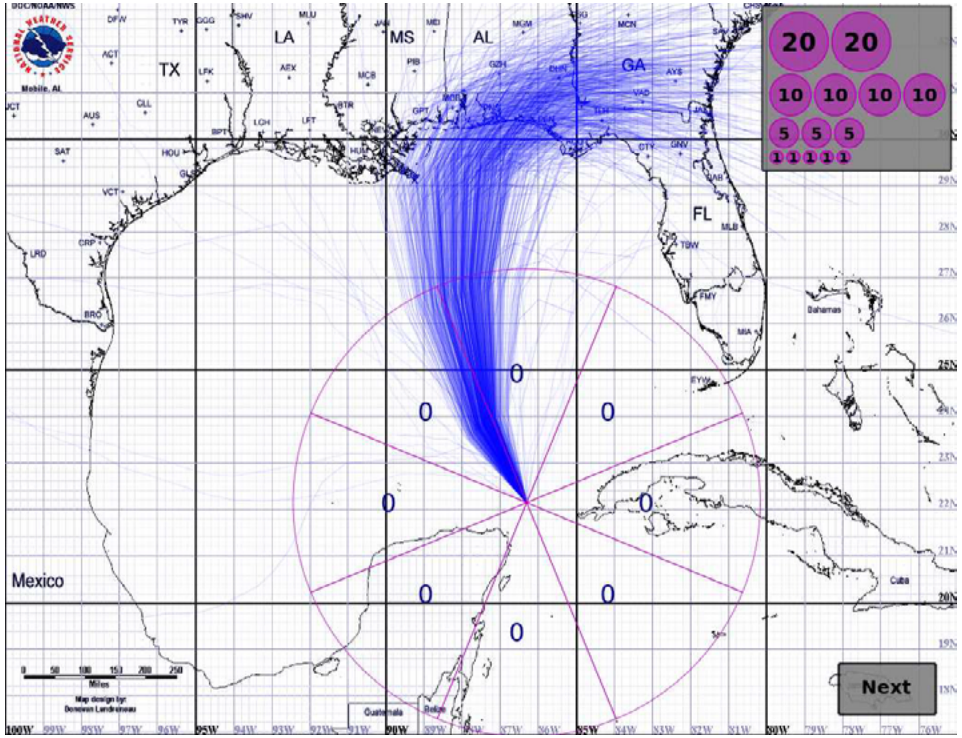
[Cox et al, Visualizing Uncertainty in Predicted Hurricane Tracks, 2013]



(Hurricane error cones)

[Cox et al, Visualizing Uncertainty in Predicted Hurricane Tracks, 2013]

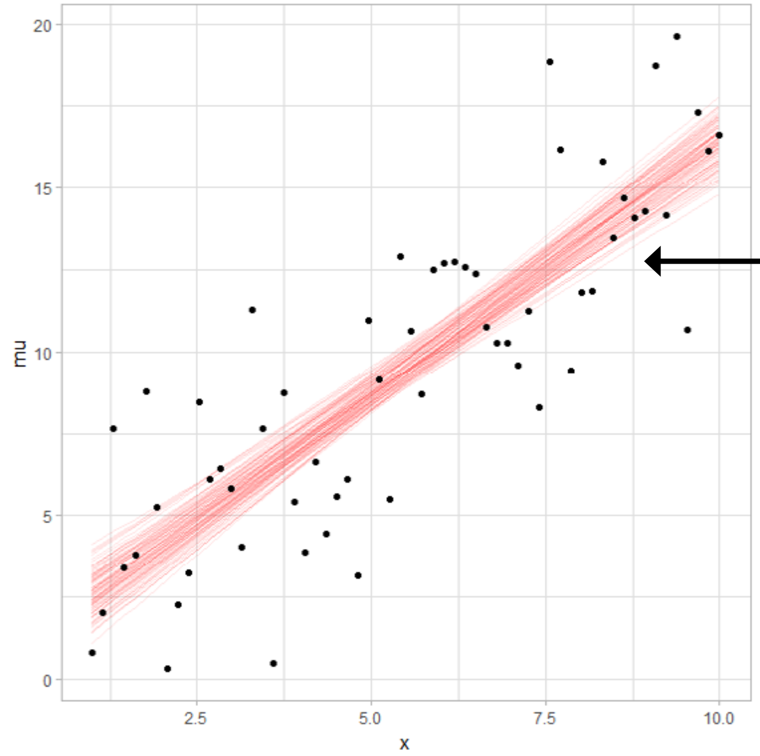




How to build this chart

What aggregation do I do in each group?

How do I map onto channels/marks?



Take first 100 samples

Geom: line

Group by sample

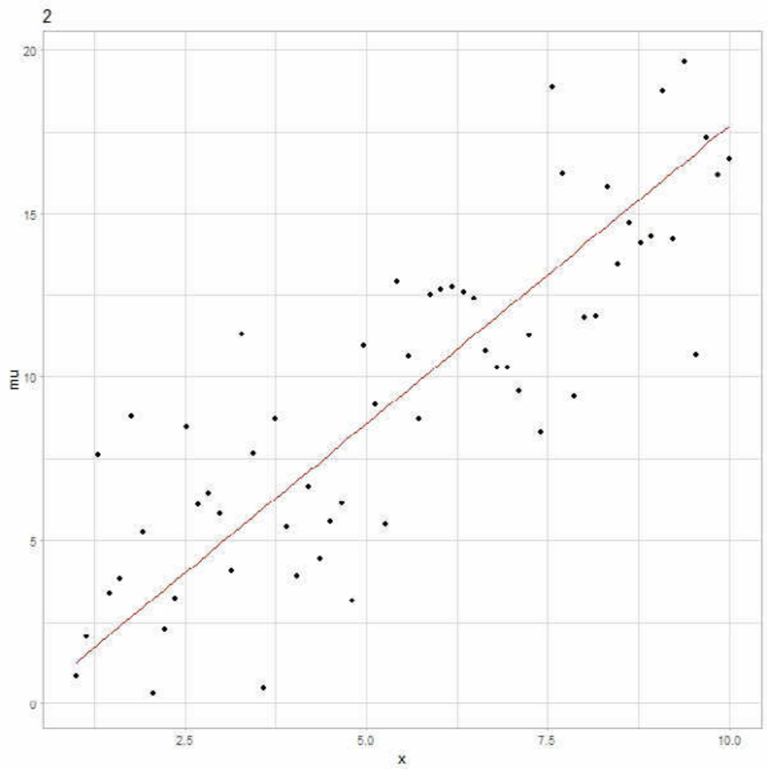
Predictors | Samples from fits or predictions

x	Samples of mean at x
1.0 1.0 1.0 ⋮	2.6 2.4 2.8 ⋮
1.1 1.1 1.1 ⋮	2.7 2.5 3.0 ⋮
⋮	⋮
10.0 10.0 10.0 ⋮	17.2 17.4 17.3 ⋮

How to build this chart

What aggregation do I do in each group?

How do I map onto channels/marks?



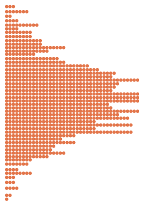
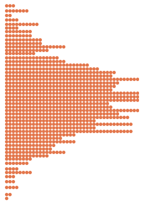
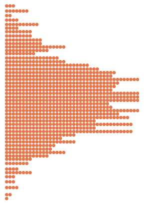
Geom: line

Map sample -> frame

(gganimate)

[Hullman et al, HOPs, 2015]

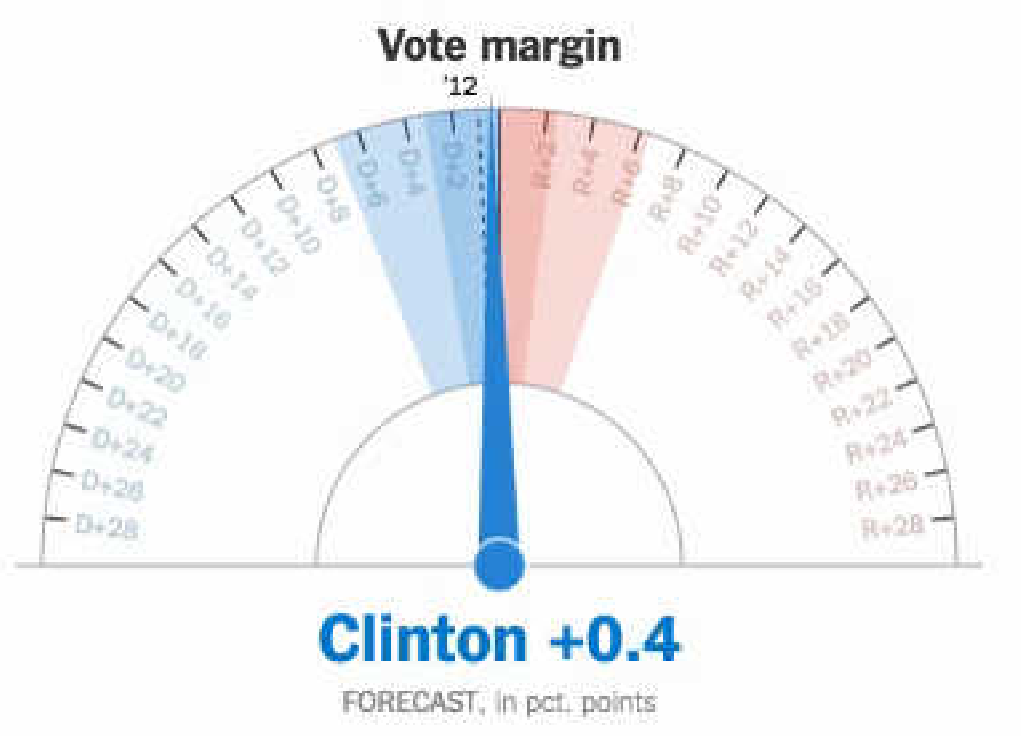
Predictors | Samples from fits or predictions

x	Samples of mean at x	
1.0 1.0 1.0 ⋮	2.6 2.4 2.8 ⋮	
1.1 1.1 1.1 ⋮	2.7 2.5 3.0 ⋮	
10.0 10.0 10.0 ⋮	17.2 17.4 17.3 ⋮	

Okay, but on the subject of HOPs

New York Times Election Needle

[https://www.nytimes.com/interactive/2016/11/08/us/elections/trump-clinton-election-night-live.html]



The Fake Twitchy Hell Dials of the New York *Times*’ Forecast Only Made Last Night Worse

By **Jake Swearingen**

Photo: rhyselsmore/Twitter

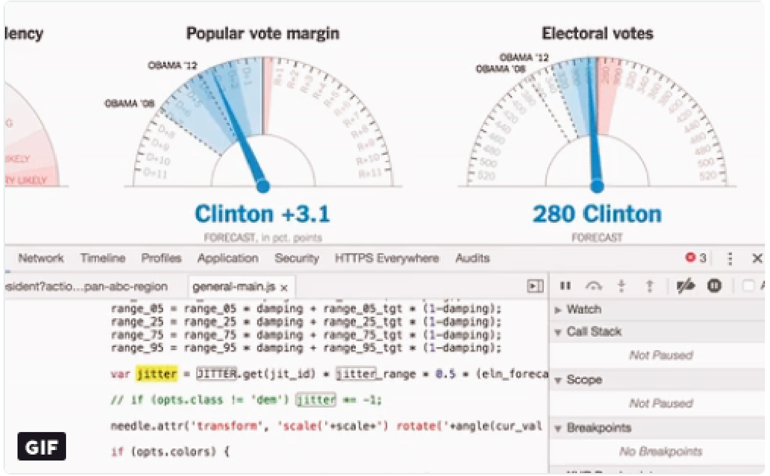
Around 9:30 last night, this tweet popped up on my timeline:

> stop tweeting the fucking hell dial
>
> — **erictoral vote (@ericlimer)** November 9, 2016

Alp Toker @atoker

Follow

Looking for trends in @nytimes's presidential forecast needle? Don't look too hard - the bounce is random jitter from your PC, not live data



Richard Porczak @tsiro

Follow

straight up: the NYT needle jitter is irresponsible design at best and unethical design at worst and you should stop looking at it

9:58 PM - 8 Nov 2016

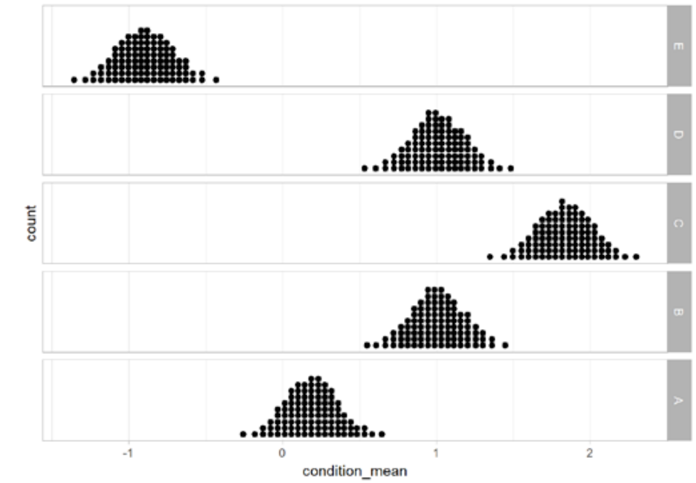
509 Retweets **882** Likes

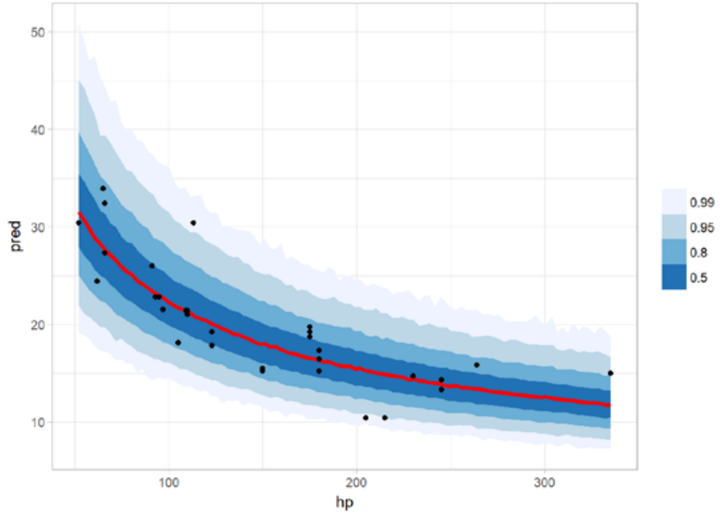
17 509 882

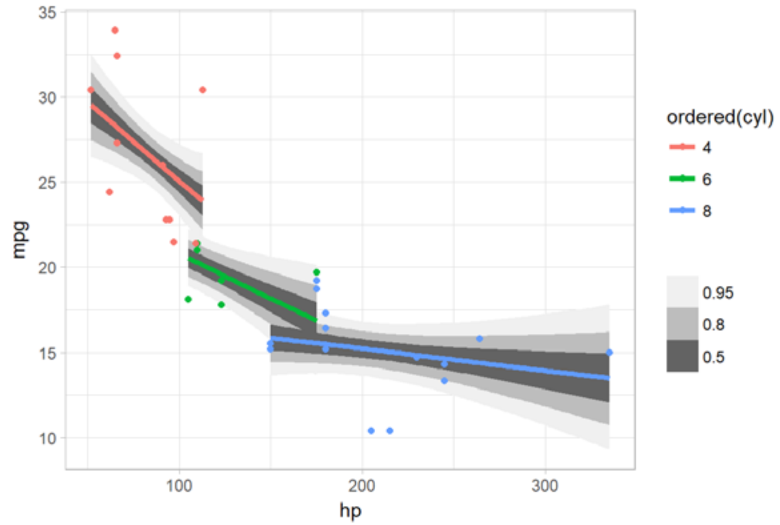
But shouldn't anxiety be proportional to uncertainty?

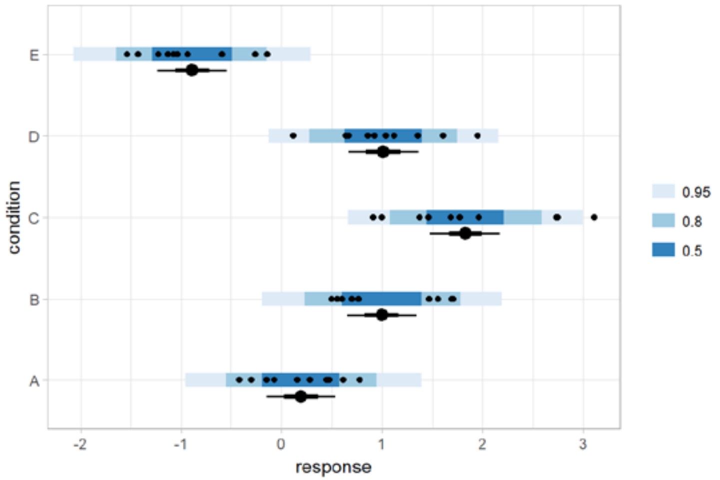
Tidy tables of samples are powerful and generic

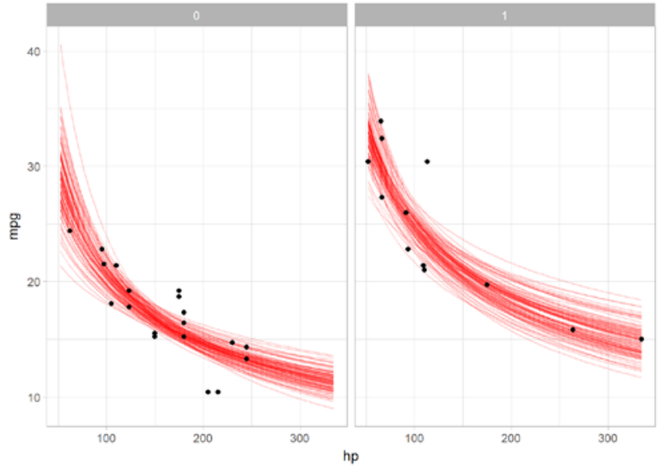
More examples

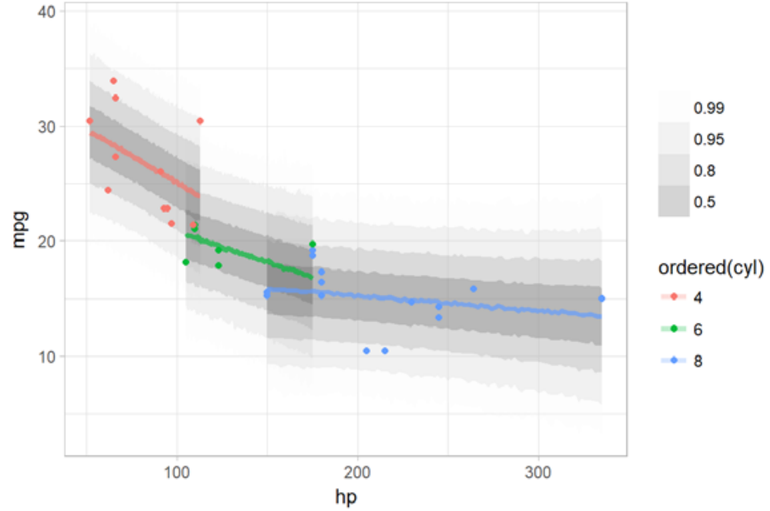












Uncertainty visualization can be fun!

And Bayesian analysis + tidy data + grammar of graphics makes it an easier-to-explore design space.