Teaching frequentist and Bayesian side by side

John K. Kruschke

http://www.indiana.edu/~kruschke/

USCOTS 2019

Evaluating Evidence

Why Teach Bayesian?

- Bayesian is valuable in real applications; sought by researchers
- Bayesian is easy to teach; easier than frequentist
- Bayesian clarifies frequentist ideas

→ Bayesian should be in curriculum

Bayesian is Valuable in Real Applications

Examples from personal experience:

- ✓ in quadratic logistic regression, need credible interval on the position of the nadir of parabola
- ✓ need simultaneous estimation of regression coefficients and latent scale values of items (with credible intervals) for end-censored data
- ✓ need models for unbalanced designs, with empty cells, heterogeneous variances, outliers, in customized hierarchical structure
- ✓ need variety of customized, specialized trend models, w. credible intervals on all parameters

All are straight forward in Bayesian.

Bayesian is Sought by Researchers

45 workshops with audiences of working professionals and graduate students from

- academia (econ, educ, climate sci, bio, cognitive sci, etc.)
- business and industry
  (retail sales, charitable giving, food production, etc.)
- government
  (human factors @FAA, survival analysis @FDA, etc.)

Bayesian is Easy to Teach
(and easier than frequentist)

- I was initially driven to Bayesian by aversion to *teaching* frequentist (and experienced extraordinary research usefulness later)
- I've taught frequentist and Bayesian courses separately for years…

A Tale of Two Courses

Frequentist	Bayesian

A Tale of Two Courses

Frequentist	Bayesian
Fundamental concepts: data and models	

A Tale of Two Courses

Frequentist	**Bayesian**
Fundamental concepts: data and models	
Generalized Linear Model: dependent variable types metric, dichotomous, nominal, ordinal, count.	

A Tale of Two Courses

Frequentist	**Bayesian**
Fundamental concepts: data and models	
sampling distrib's, *p* values, confidence intervals, test & stop intentions	Bayesian re-allocation of credibility across parameter values, MCMC representation
Generalized Linear Model: dependent variable types metric, dichotomous, nominal, ordinal, count.	

Bayesian Clarifies Frequentist Ideas

"The *p* value is not the probability of the null hypothesis!"

"What is?"

Answer from Bayesian…

Aha: The *p* value is about *imaginary data*.

"The 95% confidence interval is not the range of most probable values!"

"What is?"

Answer from Bayesian…

Aha: The CI is about *not-rejected* values.

Why Teach Bayesian?

- Bayesian is valuable in real applications; sought by researchers
- Bayesian is easy to teach; easier than frequentist
- Bayesian clarifies frequentist ideas

→ Bayesian should be in curriculum

How to Include Bayesian?

- Replace stand-alone frequentist course?

  No: frequentist methods are entrenched, and do address the issue of error rates.
- Add optional stand-alone Bayesian course?

  No: students won’t take it and instructors won’t prep it.
- Add required stand-alone Bayesian course?

  No: won’t be required any time soon.
- Any *separate* courses?

  No: Juxtaposition can clarify both approaches.

→ Inject Bayesian+frequentist into existing courses

How to Inject Bayesian+Frequentist into Existing Courses

Need: A *module* that

- is self-contained (minimizes teacher prep)
- has a complete tutorial explaining Bayesian and frequentist analyses
- has interactive software (browser-based, no installation needed)
- has interactive exercises
- has clear learning objectives and assessment

Accomplished by new Shiny App

Overview of the Shiny App

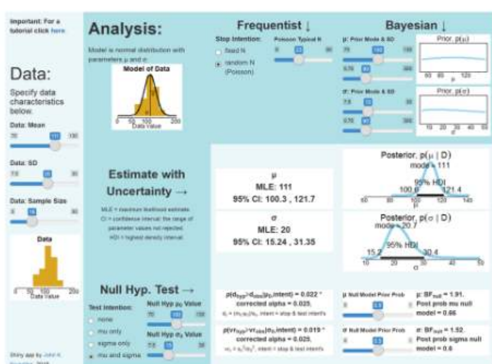
http://www.indiana.edu/~kruschke/

John K. Kruschke | Home | App | Book | Pub's | Teach | More

The Shiny App: Bayesian and Frequentist Side by Side

This is an interactive web app for doing Bayesian and frequentist analysis side by side. An extensive tutorial guides you through all the interactive capabilities. You get to see the different information delivered by Bayesian and frequentist analyses. You get to see the different information delivered by parameter estimation with uncertainty versus hypothesis testing. You get to interactively experience the dependencies of the analyses on different assumptions.

The Shiny App: Frequentist and Bayesian Side by Side (external site)



Screen Layout of the App

Data
specification

Frequentist **Bayesian**

Screen Layout of the App

Data specification		**Frequentist**	**Bayesian**
	Estimate with Uncertainty		
	Null Hypothesis Test		

Why teach hypothesis testing and estimation with uncertainty

- They have different goals:
  - Hypothesis testing seeks decision regarding a specific hypothesis.
  - Estimation with uncertainty seeks description of data with precision.
- Hypothesis testing is traditional and ubiquitous
  - but has issues (e.g., cognitive trap of "black and white thinking").
- Estimation with uncertainty is encouraged by best practices (e.g., ASA *Statement*, Am. Stat. *Beyond p<.05*)

- Juxtaposition clarifies both.
- Estimation w. uncertainty is more intuitive and easier to teach.
  - App's default view has no hypothesis tests!

Simultaneous info for all analyses

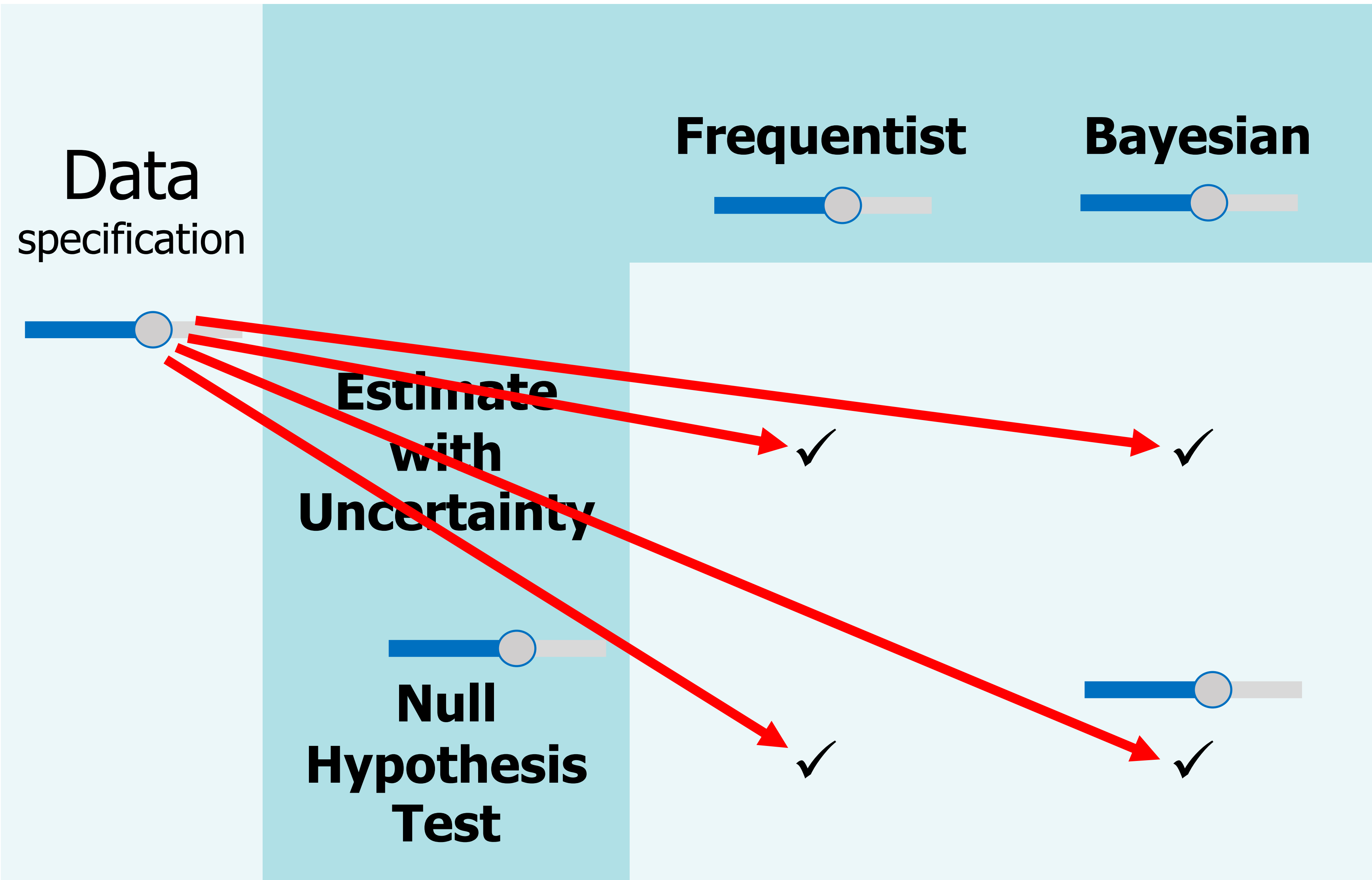
Data specification		Frequentist	Bayesian
	Estimate with Uncertainty	✓	✓
	Null Hypothesis Test	✓	✓

Interactive controls for assumptions

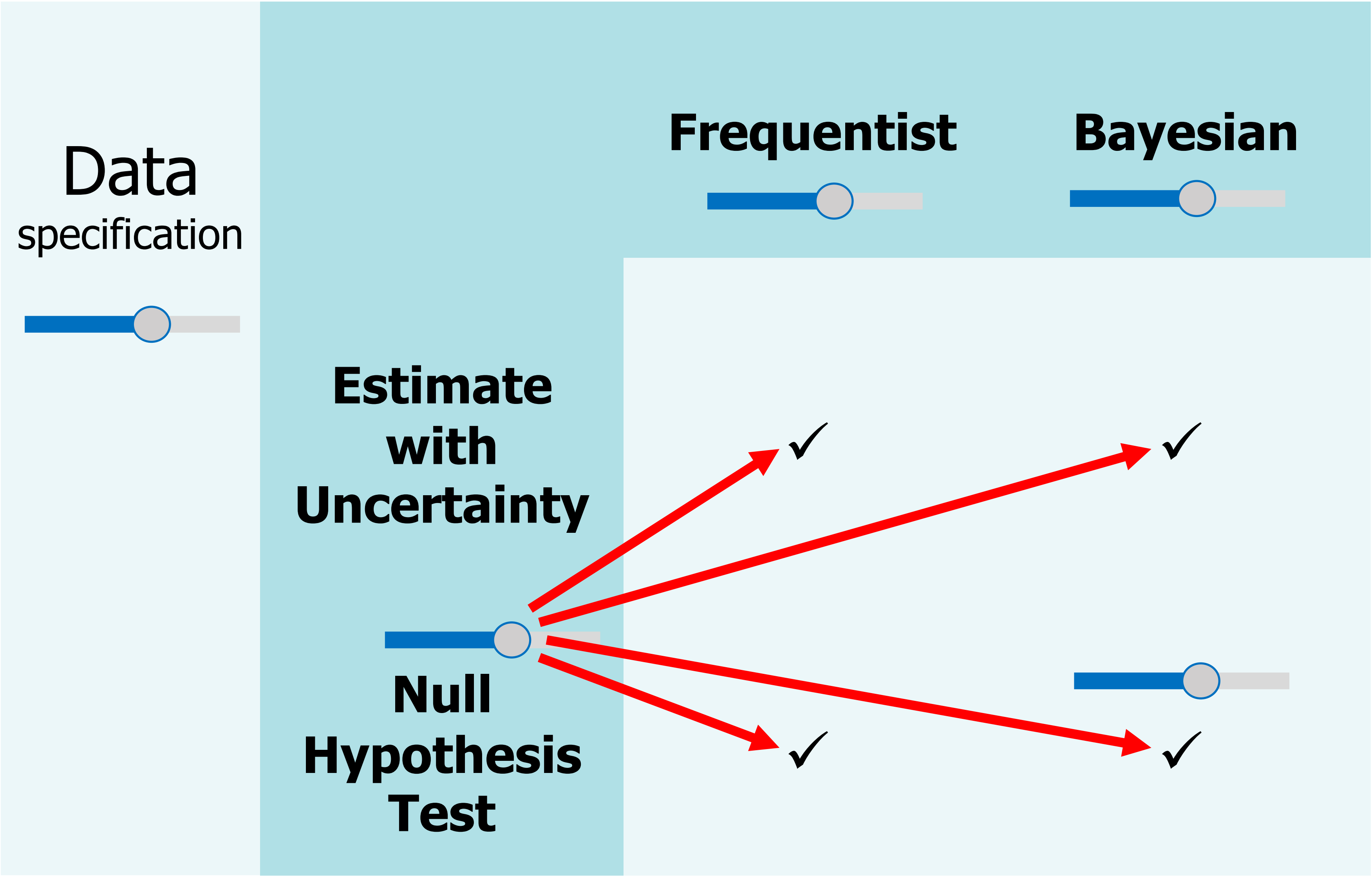
Data specification

	Frequentist	Bayesian
Estimate with Uncertainty	✓	✓
Null Hypothesis Test	✓	✓

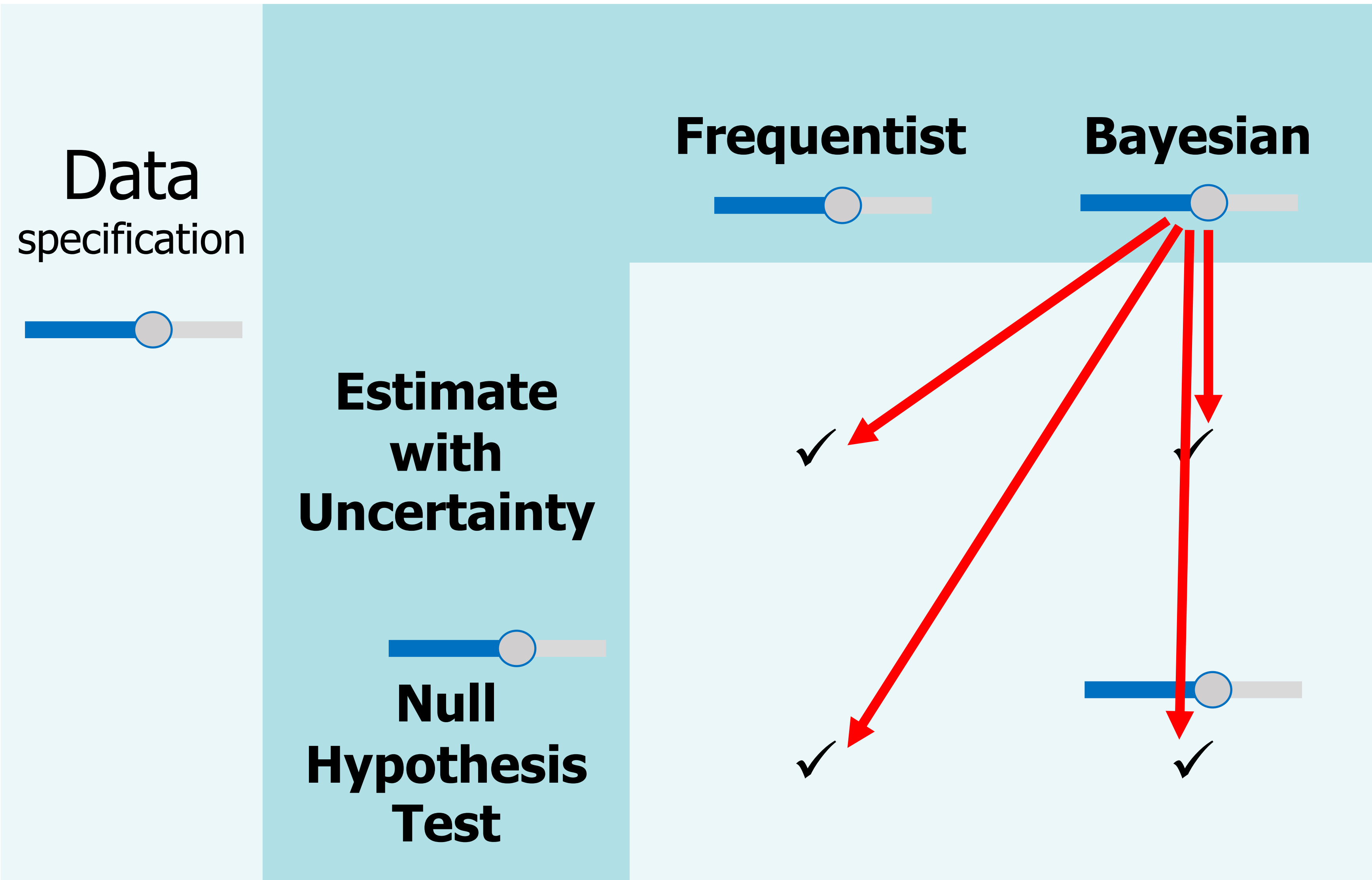
Observe all influences simultaneously



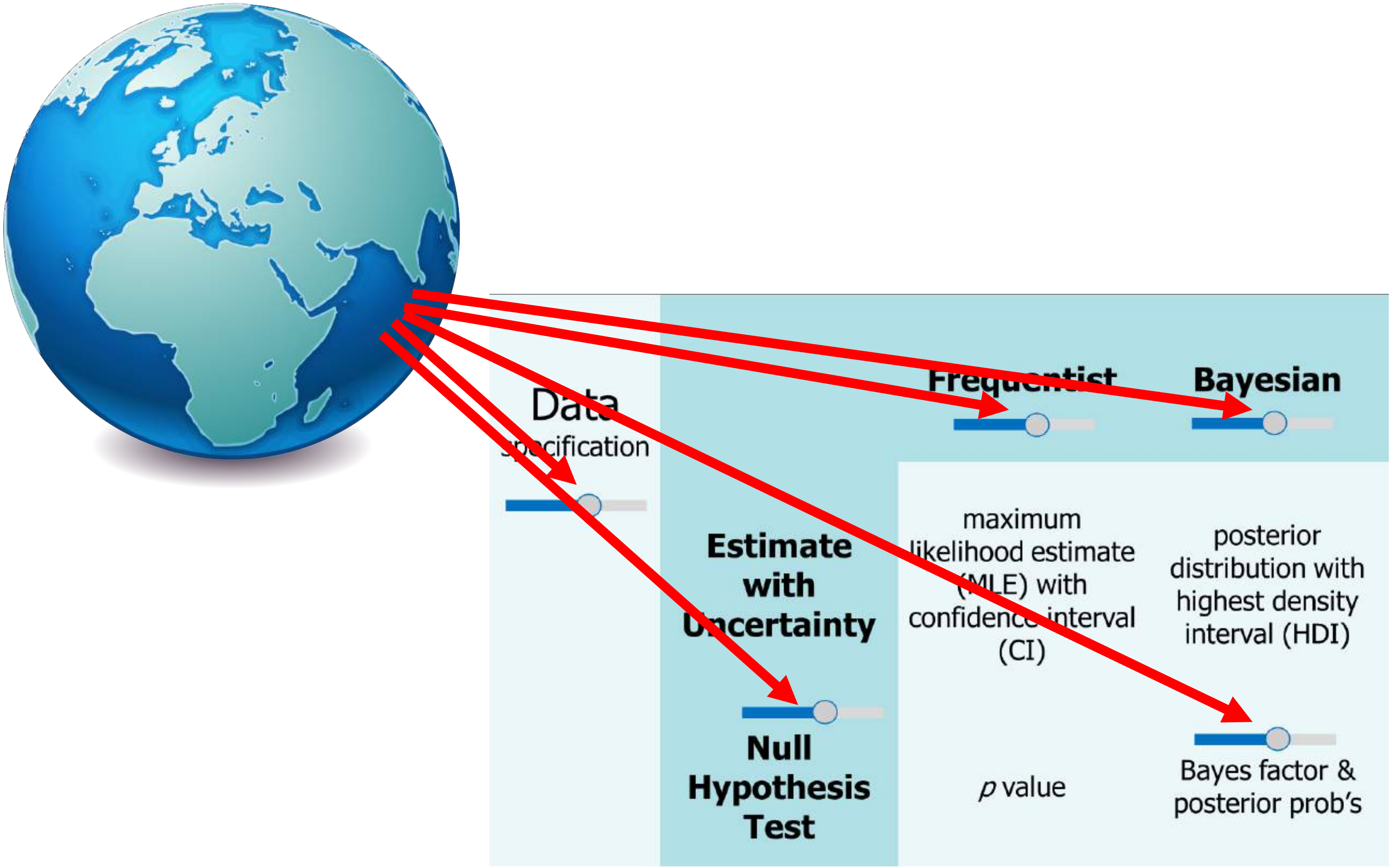
Observe all influences simultaneously



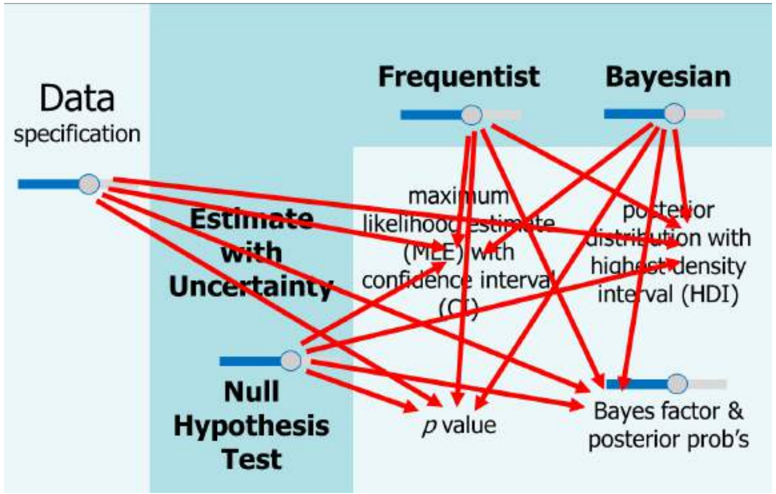
Observe all influences simultaneously

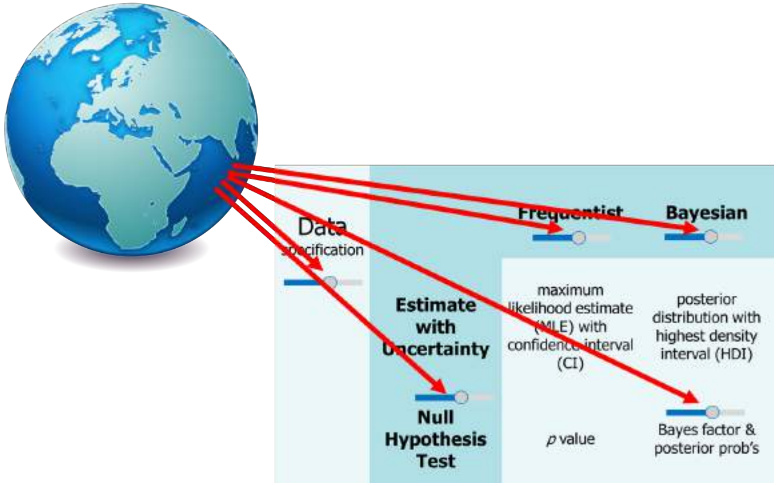


Application: Translate situation to settings



Intended Learning Outcomes

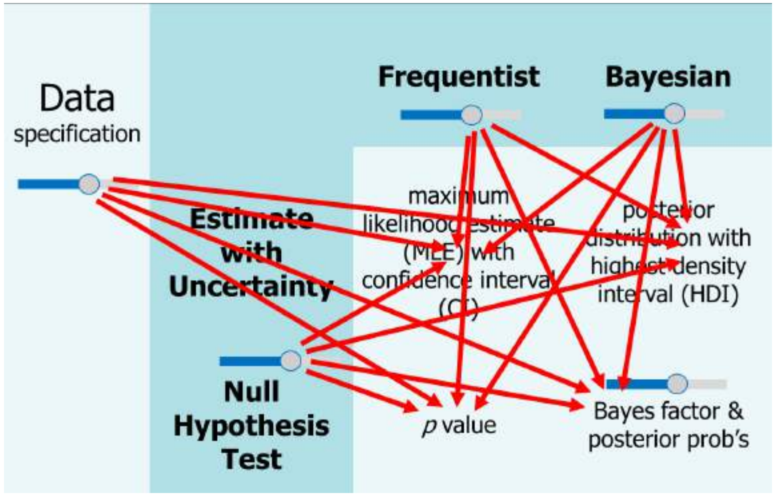


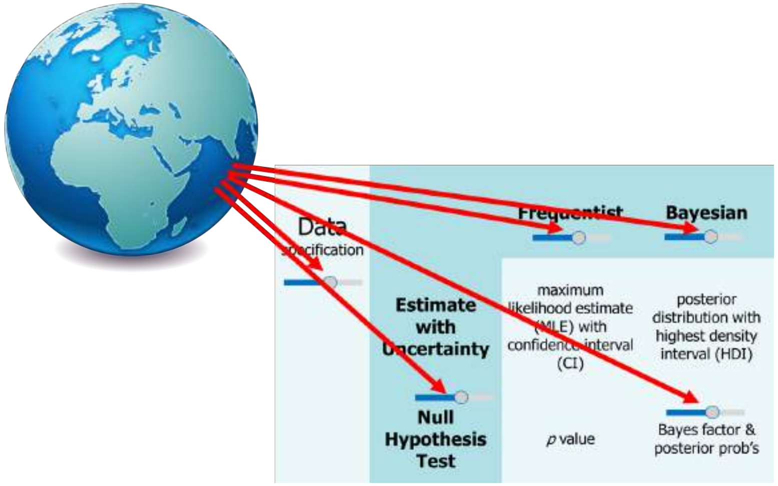


In-depth understanding of frequentist and Bayesian analyses and their inter-relation.
Achieved interactively by manipulating the sliders, watching what happens, and figuring out why.

In-depth understanding of how to apply to real situations.
Achieved interactively by translating situation to settings of the sliders. The app makes explicit what info needs to be found.

Assessing Learning Outcomes





Be able to predict the qualitative effect of every slider and button on the results in every cell of the table, and explain why.

Be able to set the sliders appropriately to reflect real-world scenarios, and explain why.

Extensive Online Tutorial

Getting oriented

This tutorial guides you through a Shiny app that puts frequentist and Bayesian analysis side by side.

This tutorial is best viewed in a wide window so the dynamic table of contents (TOC) appears on the left of the text. With the TOC visible, you can click in it to navigate to any section you like. In a narrow window, however, the TOC appears at the top of the screen and disappears when you scroll down.

Core structure of the app

The app is organized as a 2 $\times$ 2 table: There is one column for **frequentist** analysis and a second column for **Bayesian** analysis; there is one row for **estimation with uncertainty** and a second row for **null hypothesis tests**. The cells of the 2 $\times$ 2 table indicate the typical information provided by each type of analysis, as noted in the figure below:

Data specification		**Frequentist**	**Bayesian**
	Estimate with Uncertainty	maximum likelihood estimate (MLE) with confidence interval (CI)	posterior distribution with highest density interval (HDI)
	Null Hypothesis Test	*p* value	Bayes factor & posterior probabilities

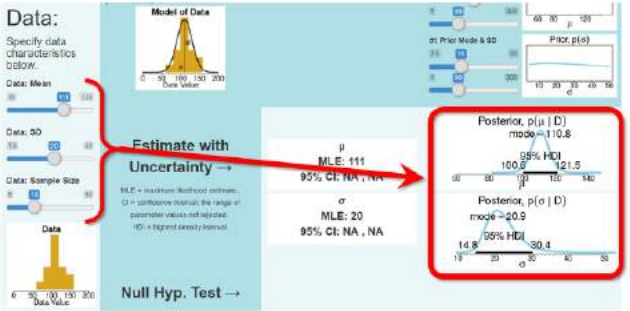
The app's 2 x 2 table of analyses.

Interactive “Try It!” Exercises

Try It!

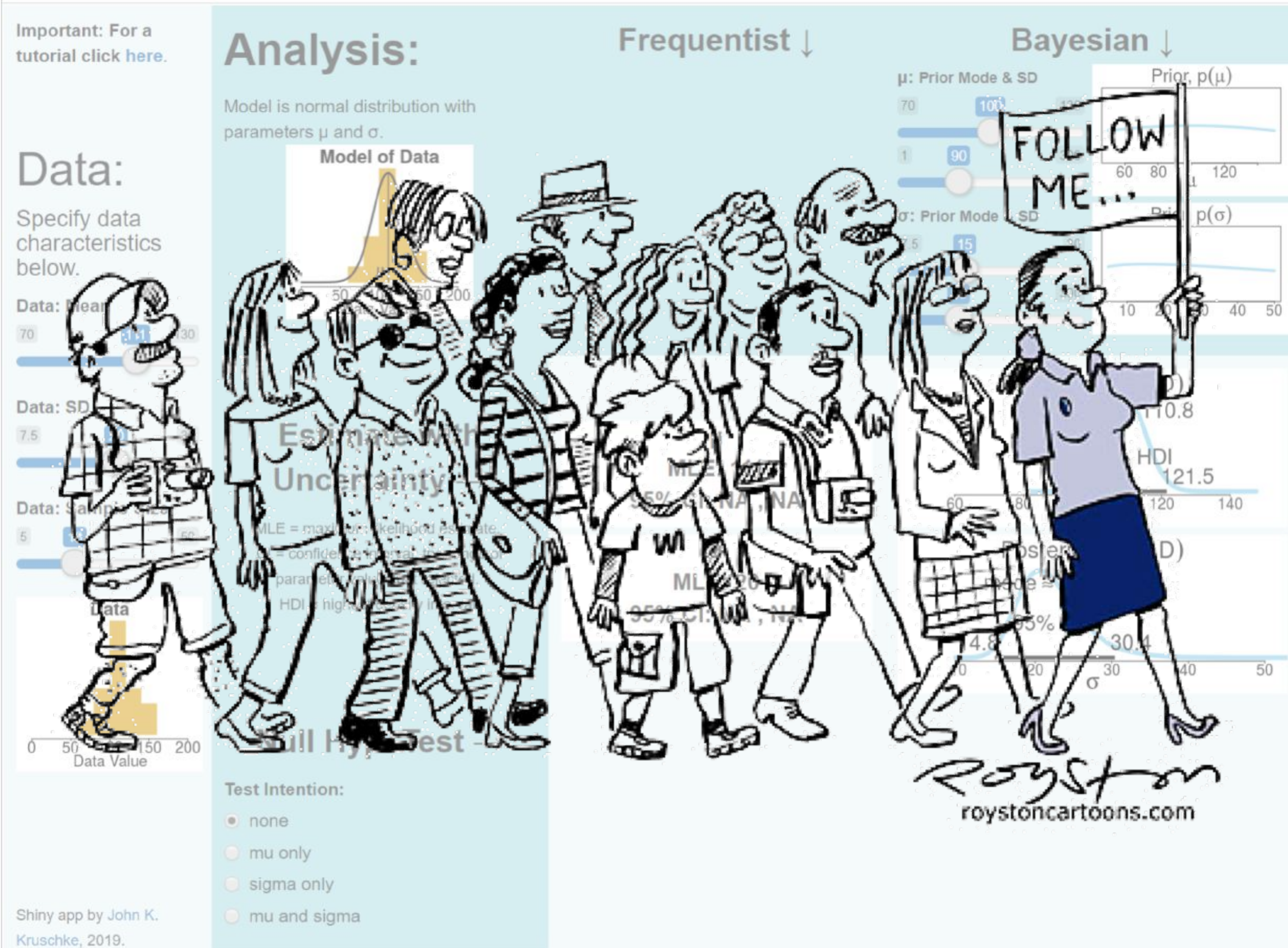
For these exercises, leave the prior distributions broad (as they are by default when the app is invoked).

Try It!
Manipulate the data sliders and watch the effect on the posterior distribution.

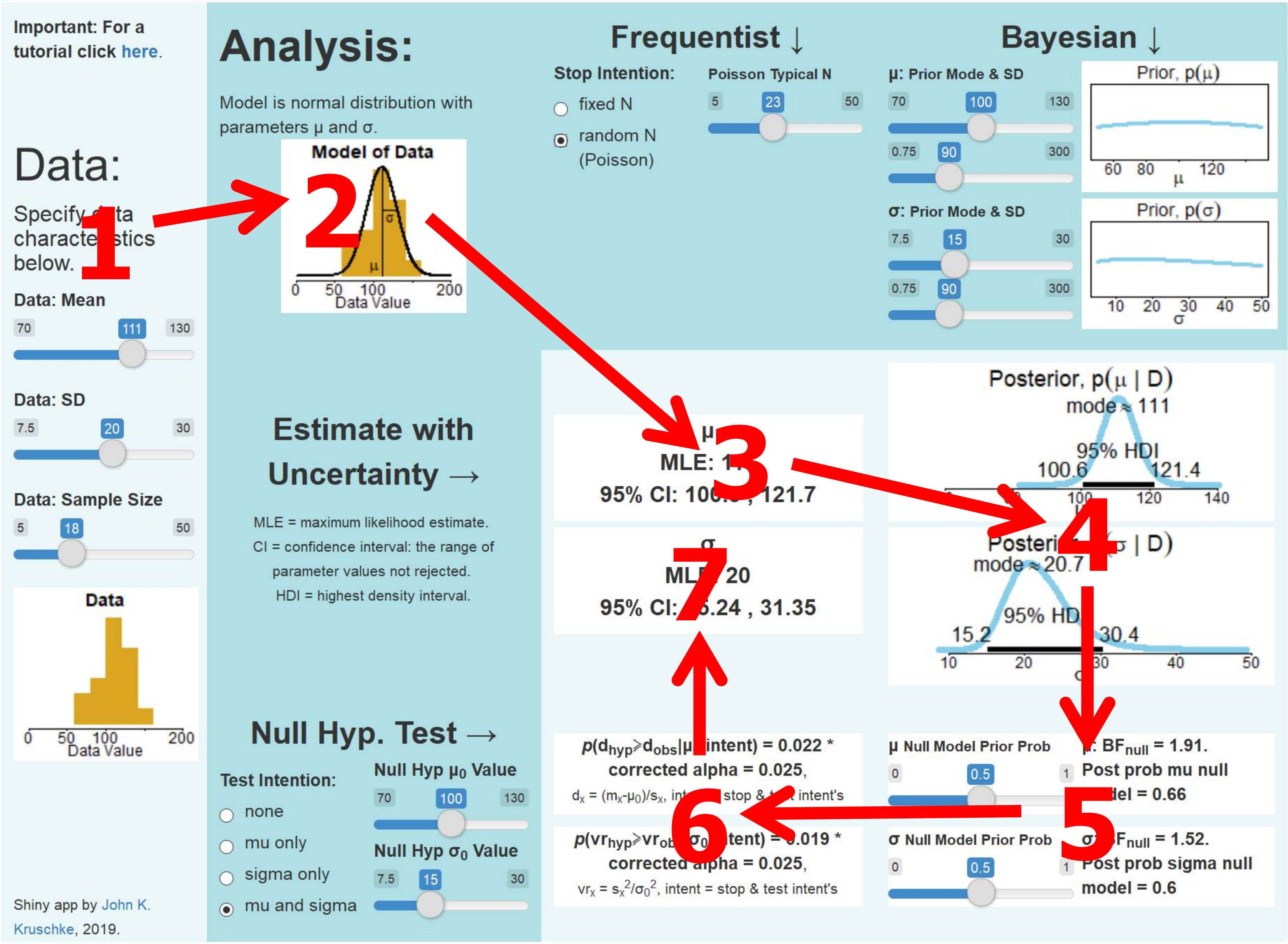


- *Slide the data mean to a different value and watch the posterior mode of μ.* The posterior mode of μ should be close to the data mean. This makes sense because credible values of μ should be near the data mean when there is not strong prior information to suggest otherwise.
- *Slide the data standard deviation to a different value and watch*

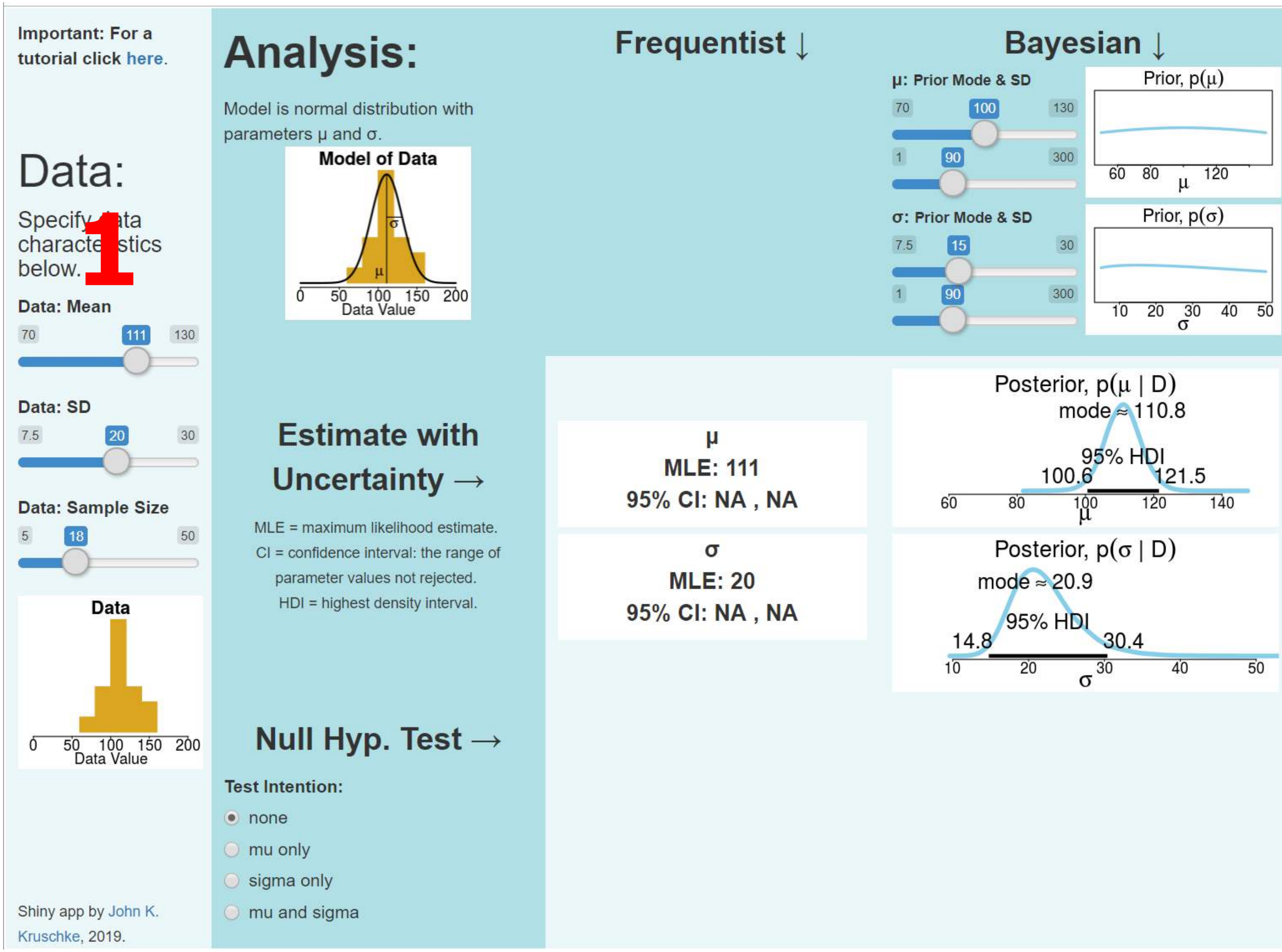
A Tour of the Shiny App



Ordering of topics



1. The Data

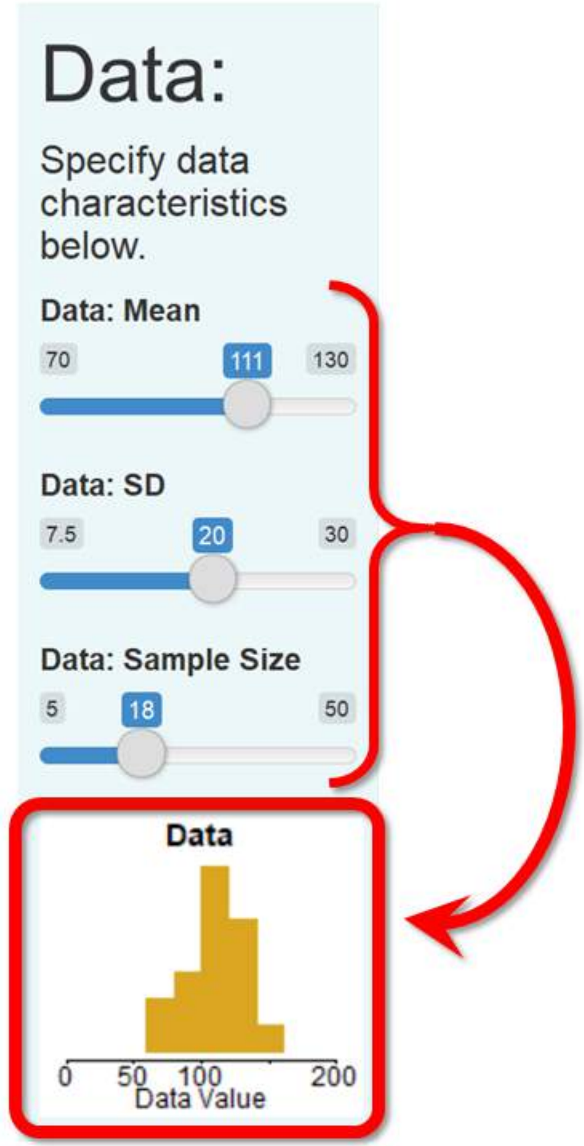


Important: For a tutorial click here.
Analysis:
Frequentist ↓
Bayesian ↓
Model is normal distribution with parameters μ and σ.
Model of Data
Data Value
Data:
Specify data characteristics below.
1
Data: Mean
70
111
130
Data: SD
7.5
20
30
Data: Sample Size
5
18
50
Data
Data Value
μ: Prior Mode & SD
70
100
130
1
90
300
Prior, p(μ)
σ: Prior Mode & SD
7.5
15
30
1
90
300
Prior, p(σ)
Estimate with Uncertainty →
μ
MLE: 111
95% CI: NA , NA
σ
MLE: 20
95% CI: NA , NA
Posterior, p(μ | D)
mode ≈ 110.8
95% HDI
100.6
121.5
Posterior, p(σ | D)
mode ≈ 20.9
95% HDI
14.8
30.4
MLE = maximum likelihood estimate.
CI = confidence interval: the range of parameter values not rejected.
HDI = highest density interval.
Null Hyp. Test →
Test Intention:
none
mu only
sigma only
mu and sigma
Shiny app by John K. Kruschke, 2019.

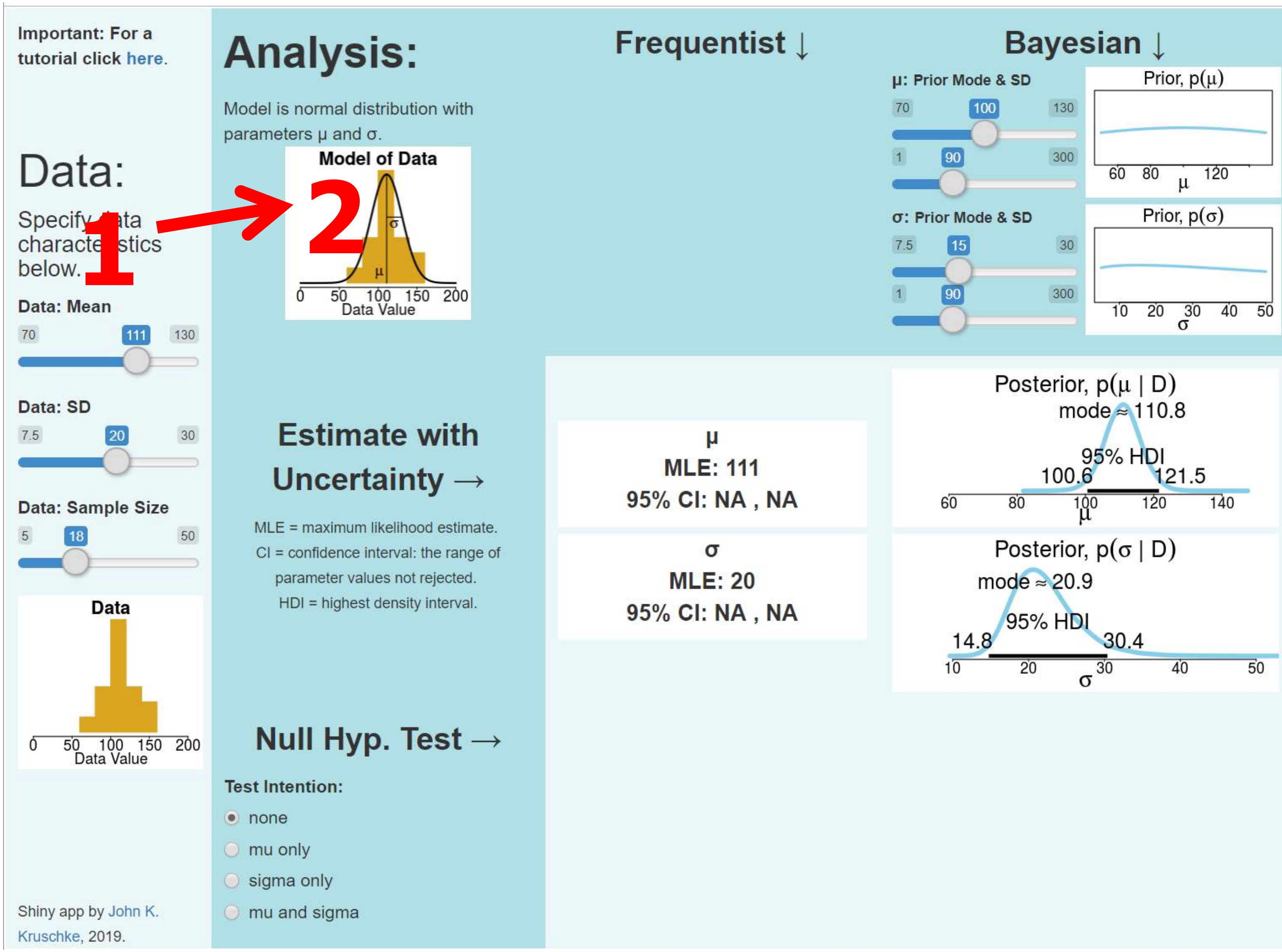
1. The Data

Try It!

Watch the effect of the data sliders on the data histogram.

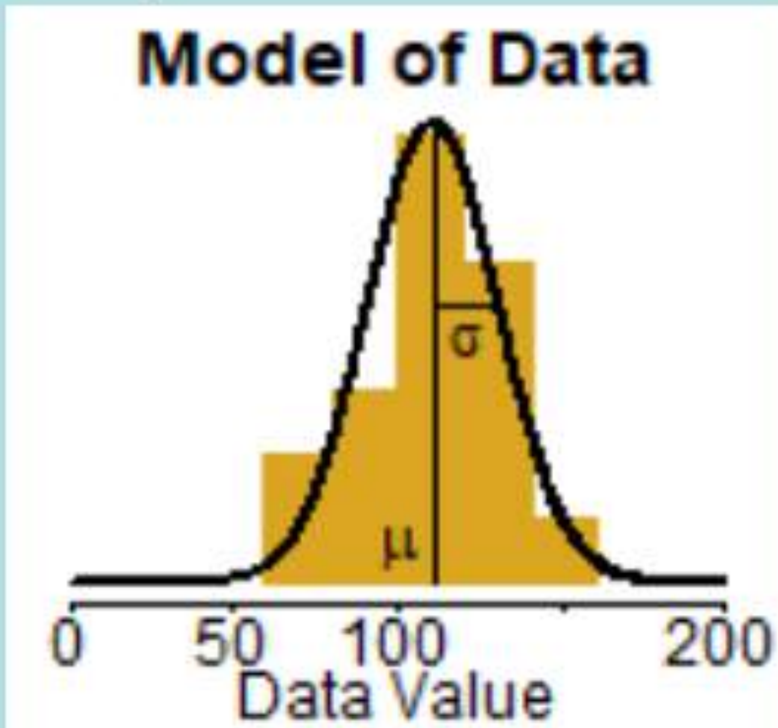


2. The Model


Important: For a tutorial click here.
Data:
Specify data characteristics below.
Data: Mean
70 111 130
Data: SD
7.5 20 30
Data: Sample Size
5 18 50
Data
0 50 100 150 200
Data Value
Shiny app by John K. Kruschke, 2019.
Analysis:
Model is normal distribution with parameters μ and σ.
Model of Data
0 50 100 150 200
Data Value
1
2
Estimate with Uncertainty →
MLE = maximum likelihood estimate.
CI = confidence interval: the range of parameter values not rejected.
HDI = highest density interval.
Null Hyp. Test →
Test Intention:
none
mu only
sigma only
mu and sigma
Frequentist ↓
μ
MLE: 111
95% CI: NA , NA
σ
MLE: 20
95% CI: NA , NA
Bayesian ↓
μ: Prior Mode & SD
70 100 130
1 90 300
Prior, p(μ)
60 80 120
μ
σ: Prior Mode & SD
7.5 15 30
1 90 300
Prior, p(σ)
10 20 30 40 50
σ
Posterior, p(μ | D)
mode ≈ 110.8
95% HDI
100.6 121.5
60 80 100 120 140
μ
Posterior, p(σ | D)
mode ≈ 20.9
95% HDI
14.8 30.4
10 20 30 40 50
σ

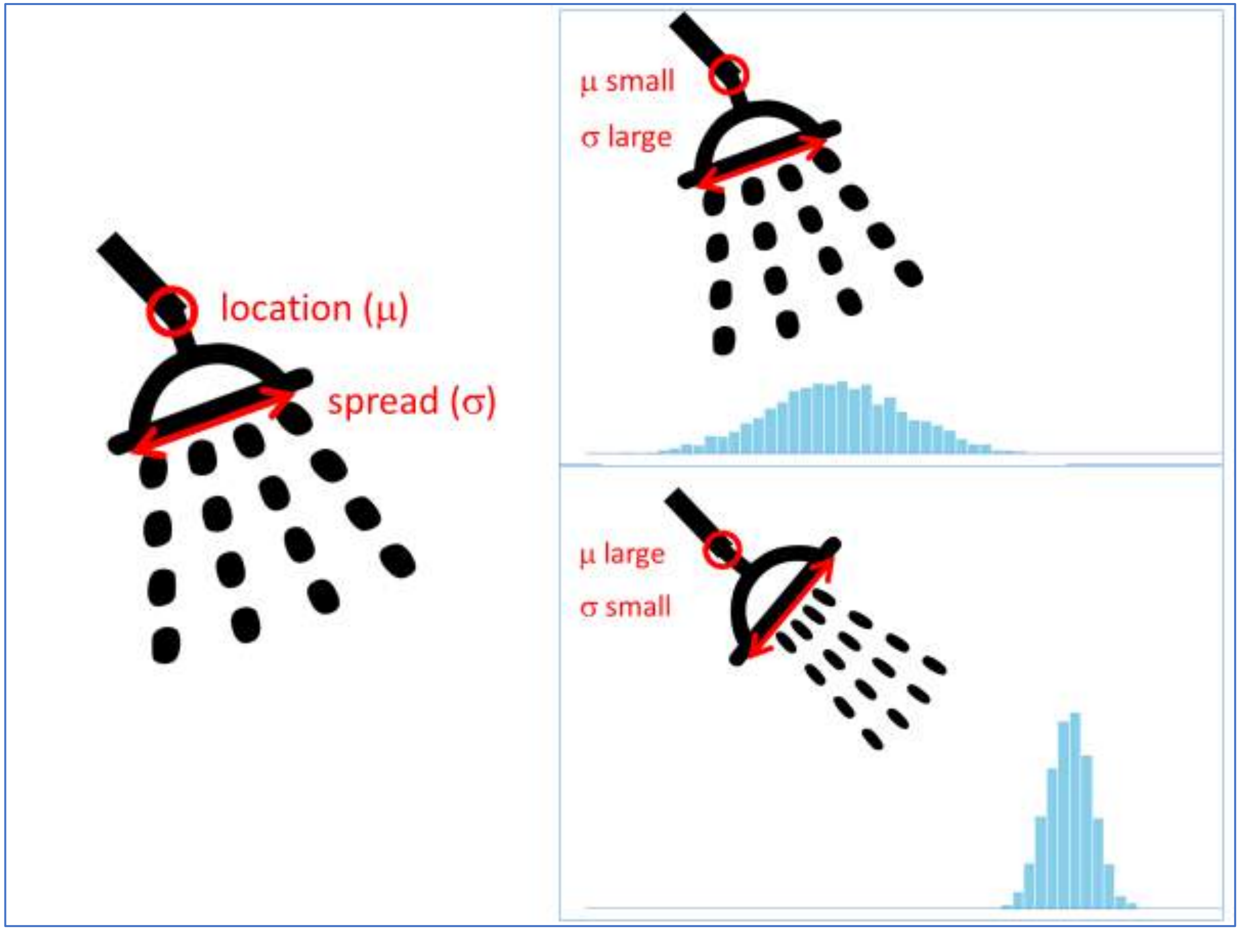
2. The Model

Model is normal distribution with parameters μ and σ.

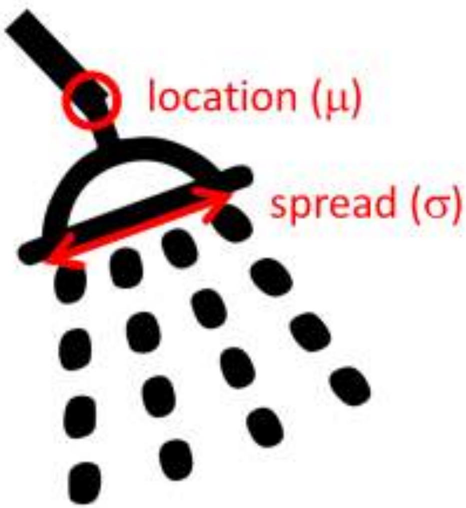


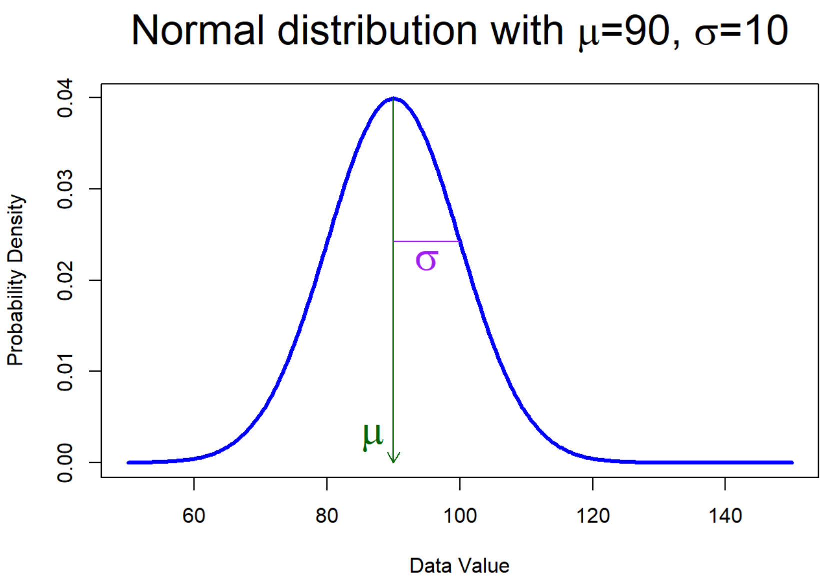
2. The Model

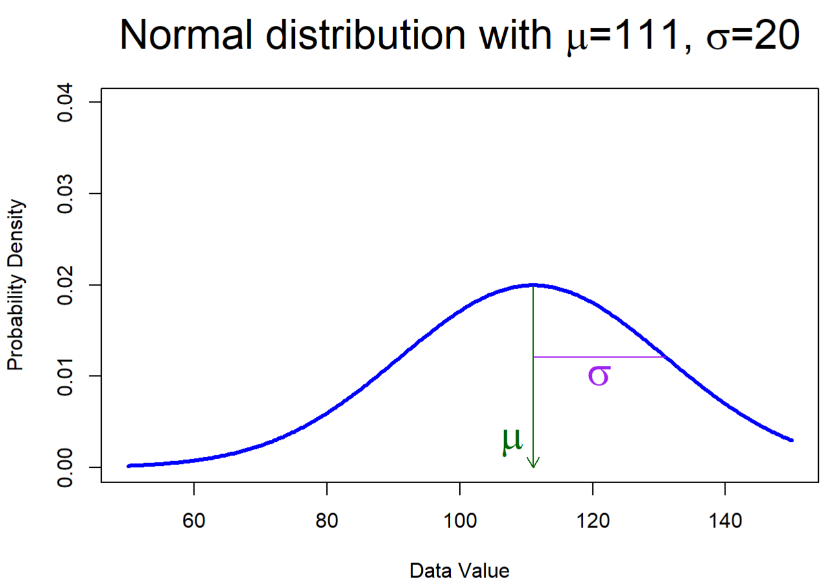
A *model* is a data-generating machine with control knobs called parameters.



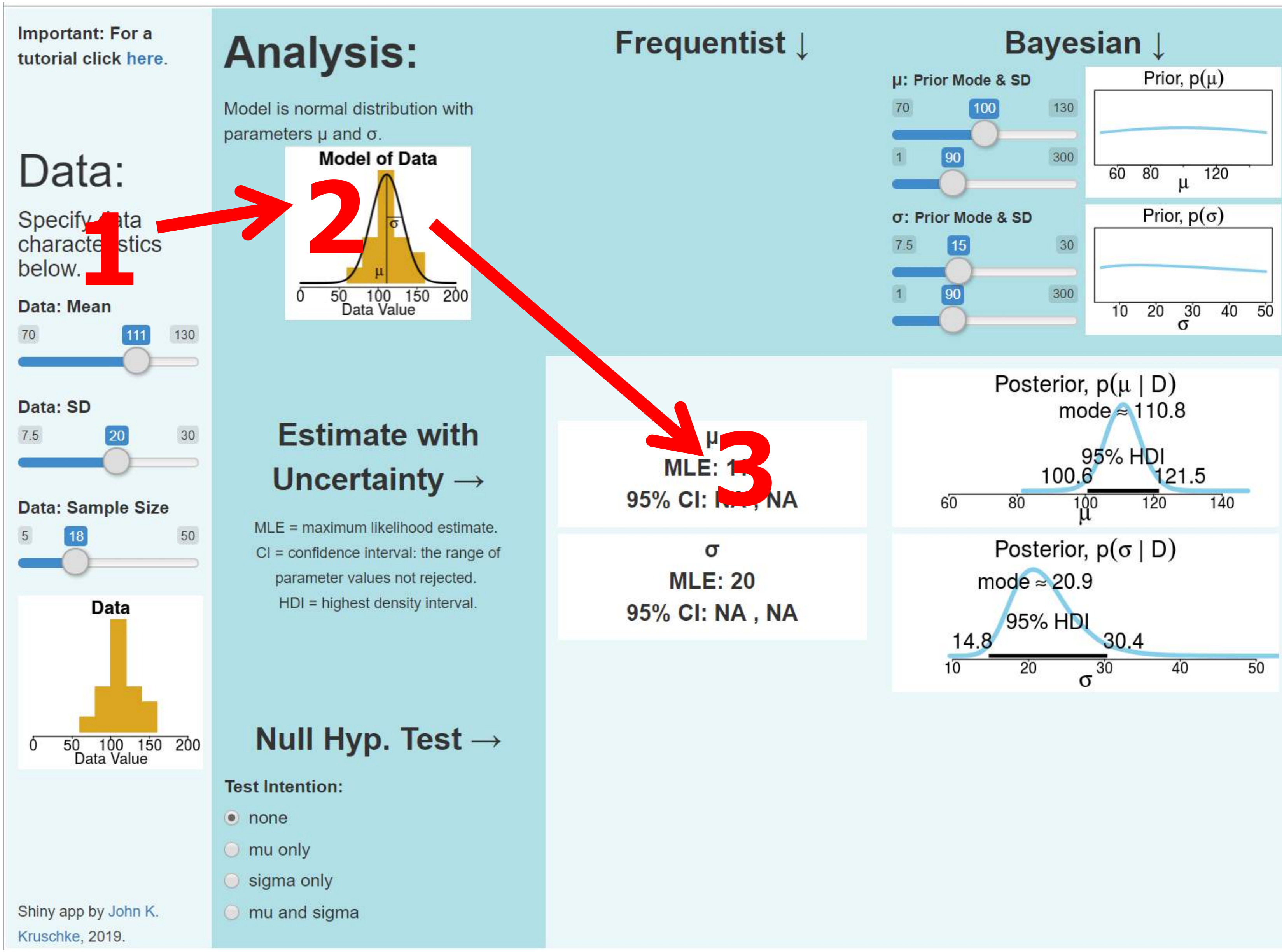
2. The Model


location (μ)
spread (σ)

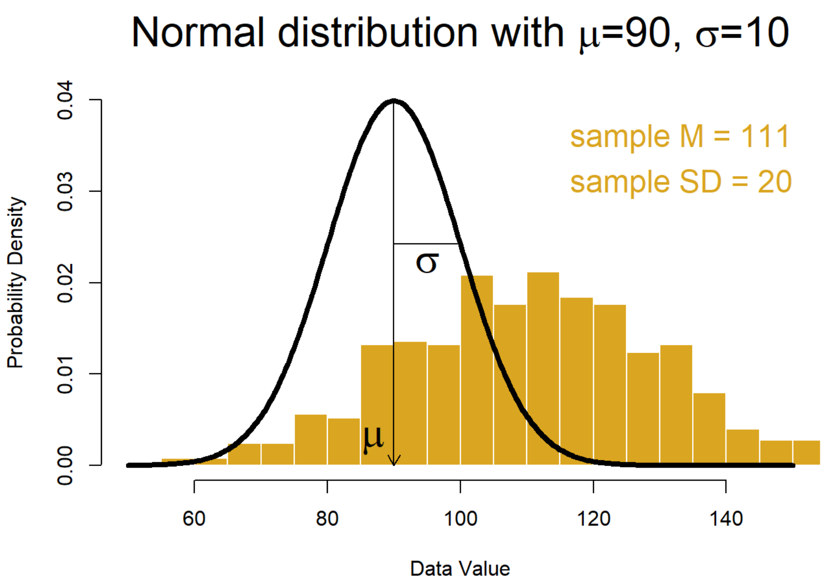

Normal distribution with μ=90, σ=10
Probability Density
0.00
0.01
0.02
0.03
0.04
σ
μ
60
80
100
120
140
Data Value

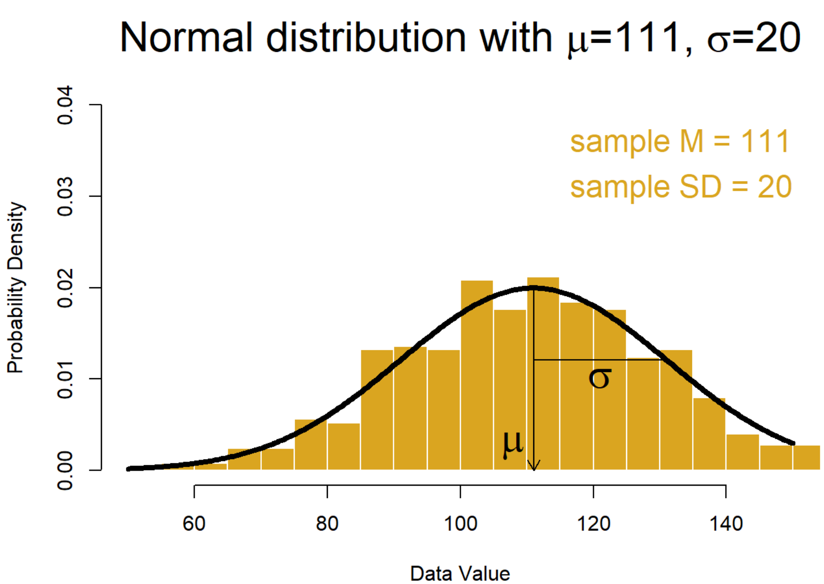

Normal distribution with μ=111, σ=20
Probability Density
0.00
0.01
0.02
0.03
0.04
σ
μ
60
80
100
120
140
Data Value

3. Frequentist (point) Estimation



3. Frequentist (point) Estimation





μ

MLE: 111

95% CI: NA , NA

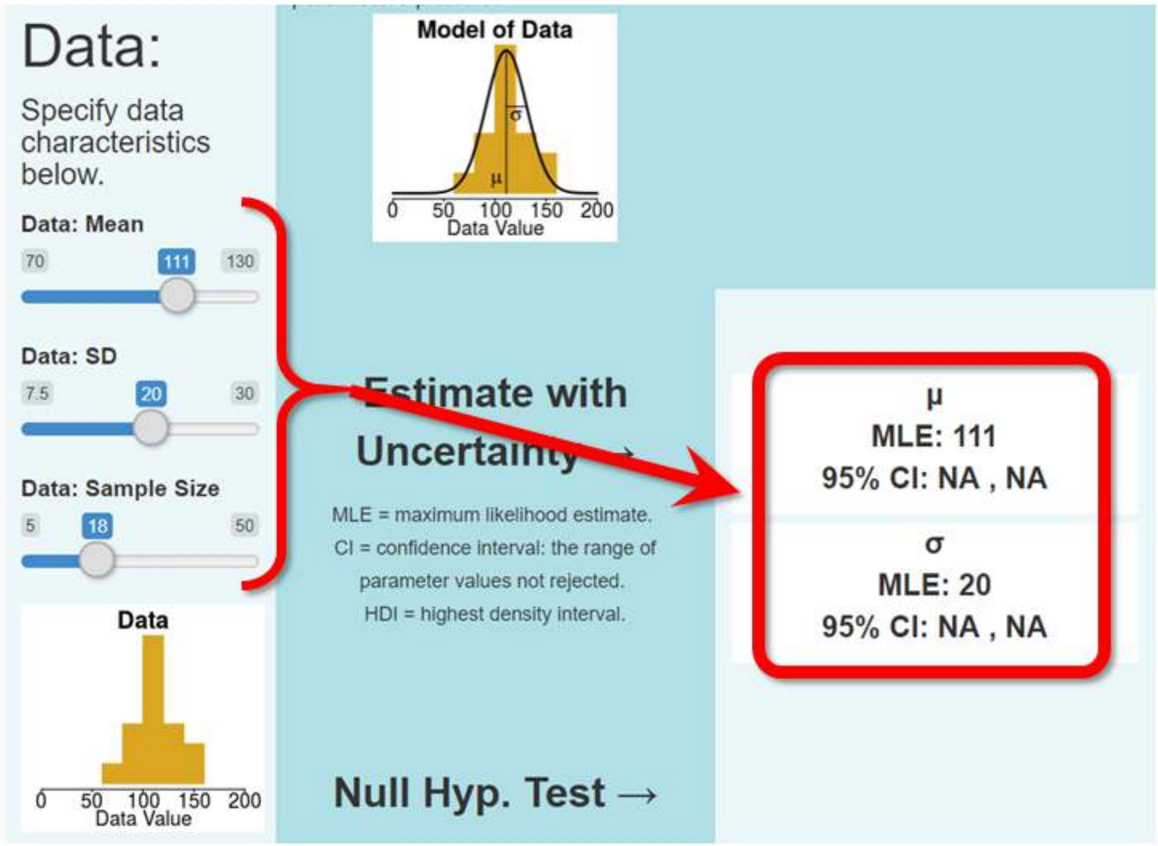
σ

MLE: 20

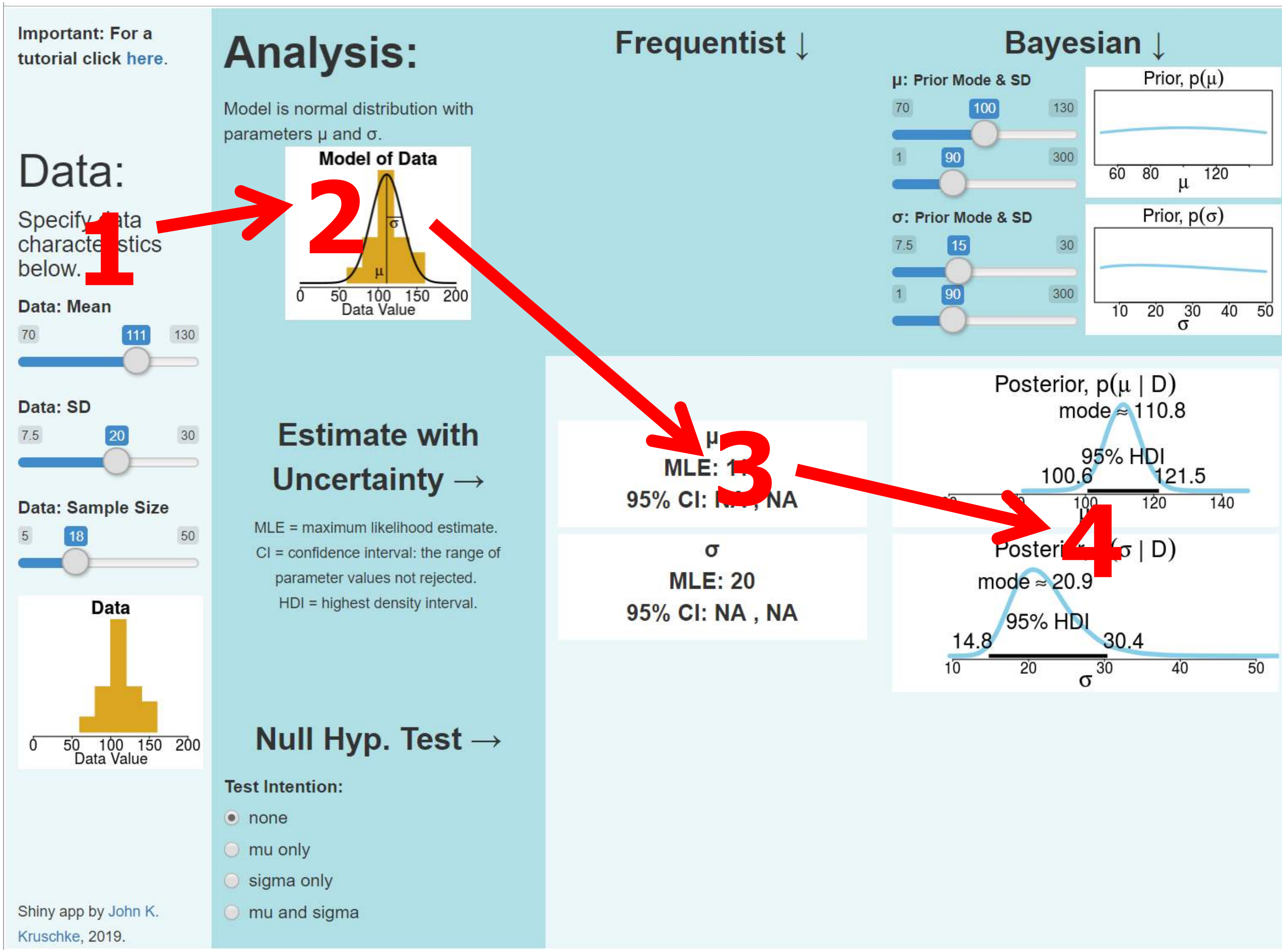
95% CI: NA , NA

3. Frequentist (point) Estimation

Try It!
Manipulate the data sliders and watch the MLE's.



4. Bayesian Estimation



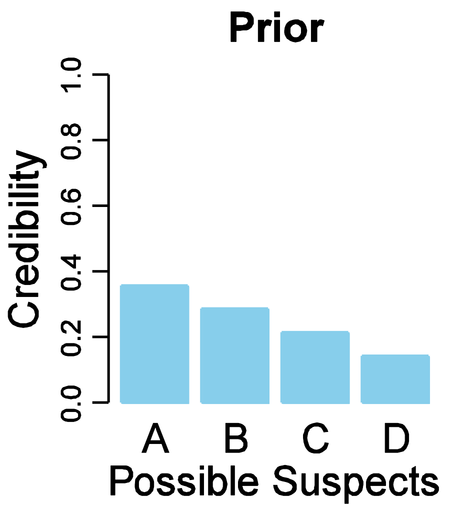
Two Foundational Ideas of Bayesian Reasoning

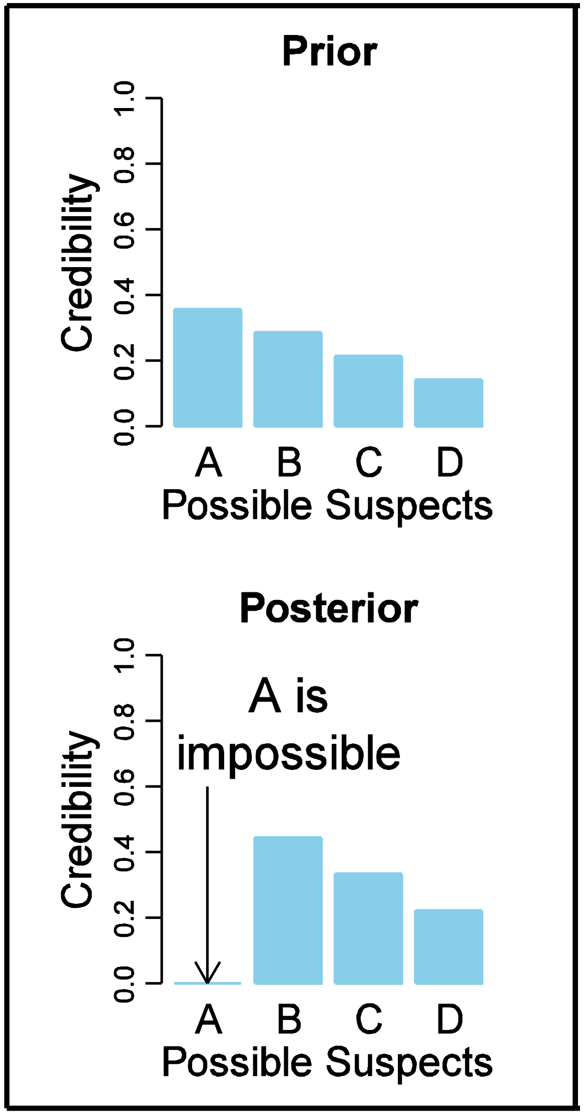
1. Bayesian reasoning is re-allocation of credibility across possibilities.
2. The possibilities are parameter values in a mathematical model of data.

Bayesian reasoning is re-allocation of credibility across possibilities.

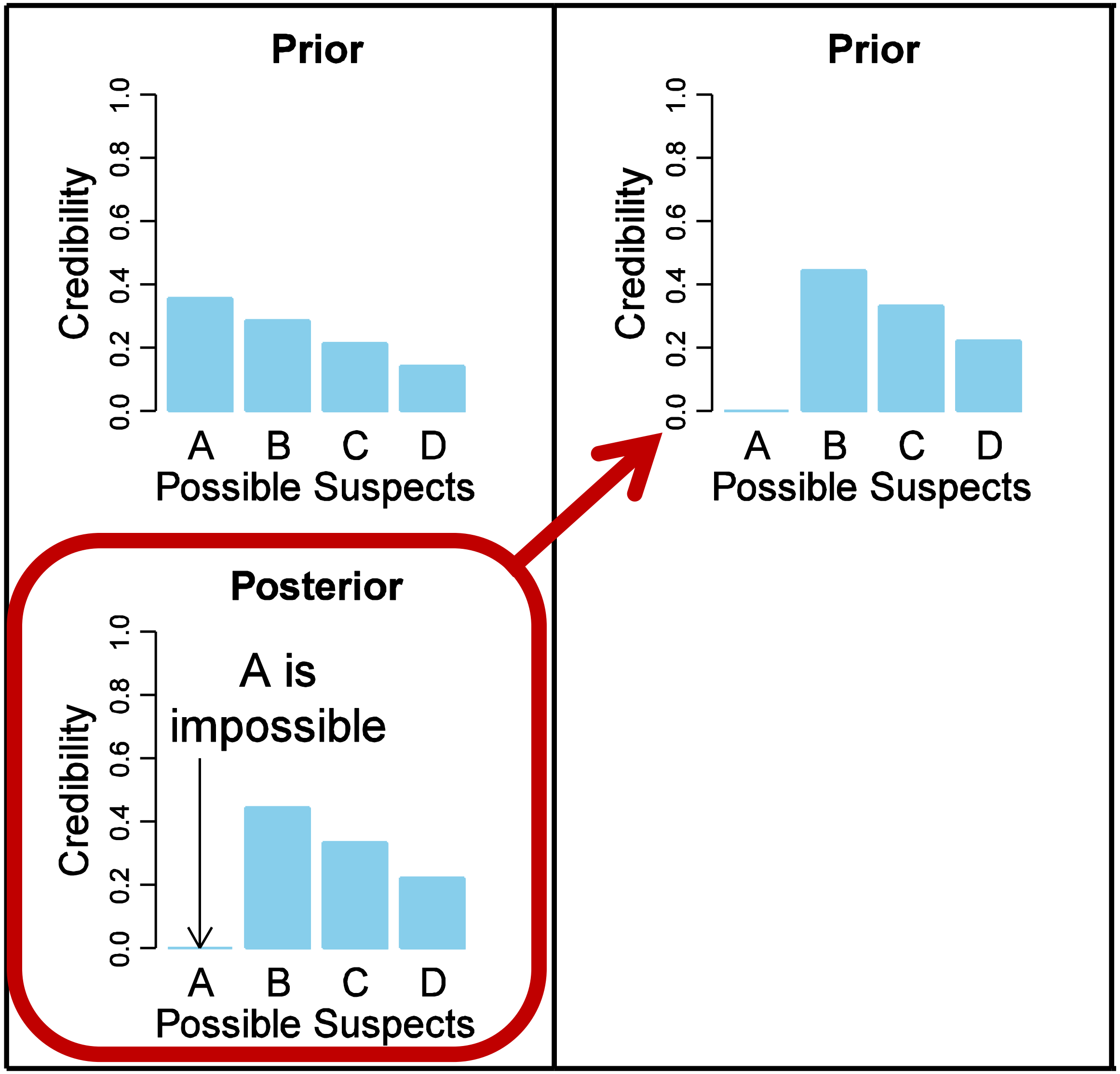
Sherlock Holmes: “How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?”

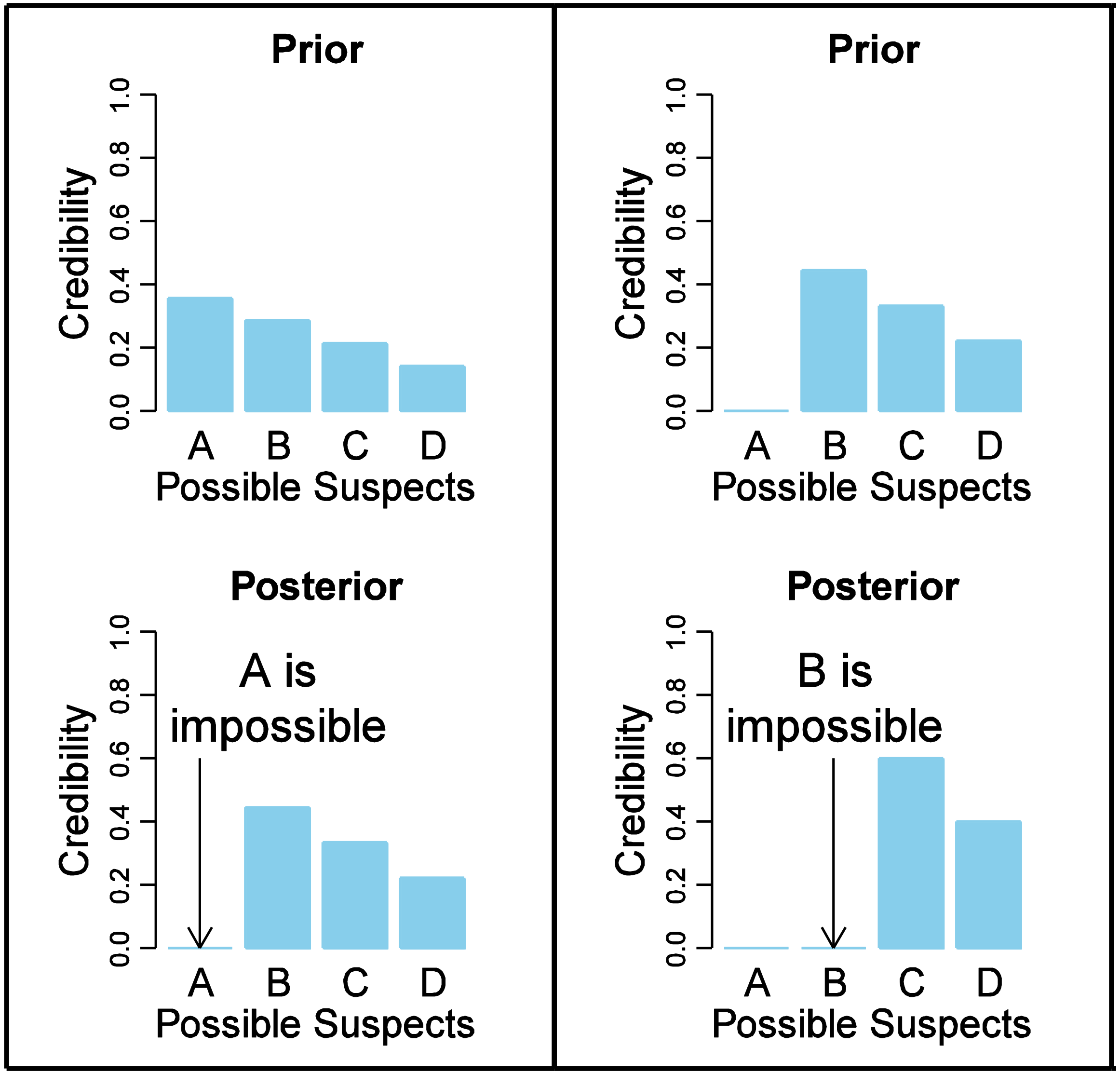
(Doyle, 1890)


Prior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A
B
C
D
Possible Suspects

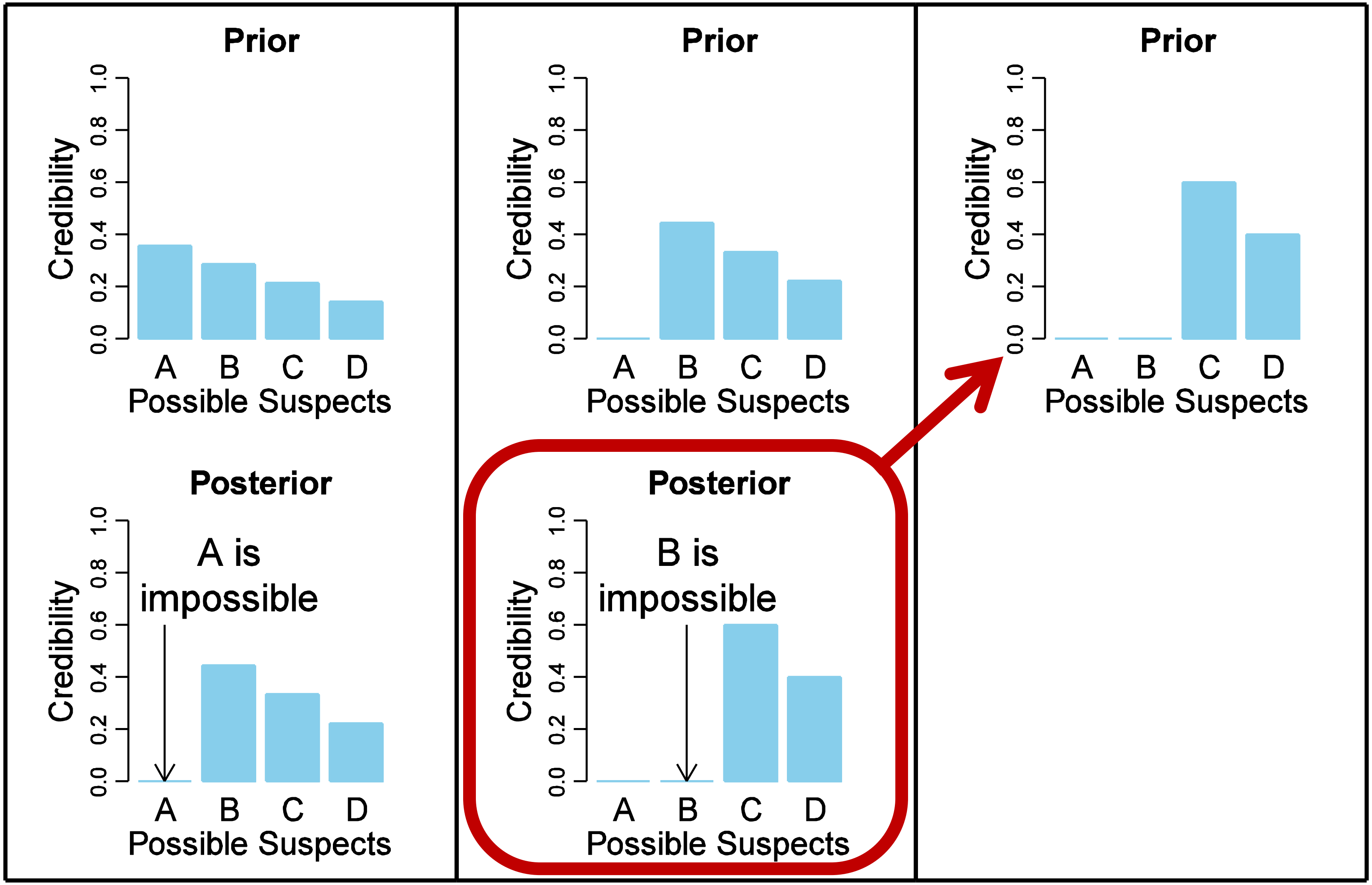


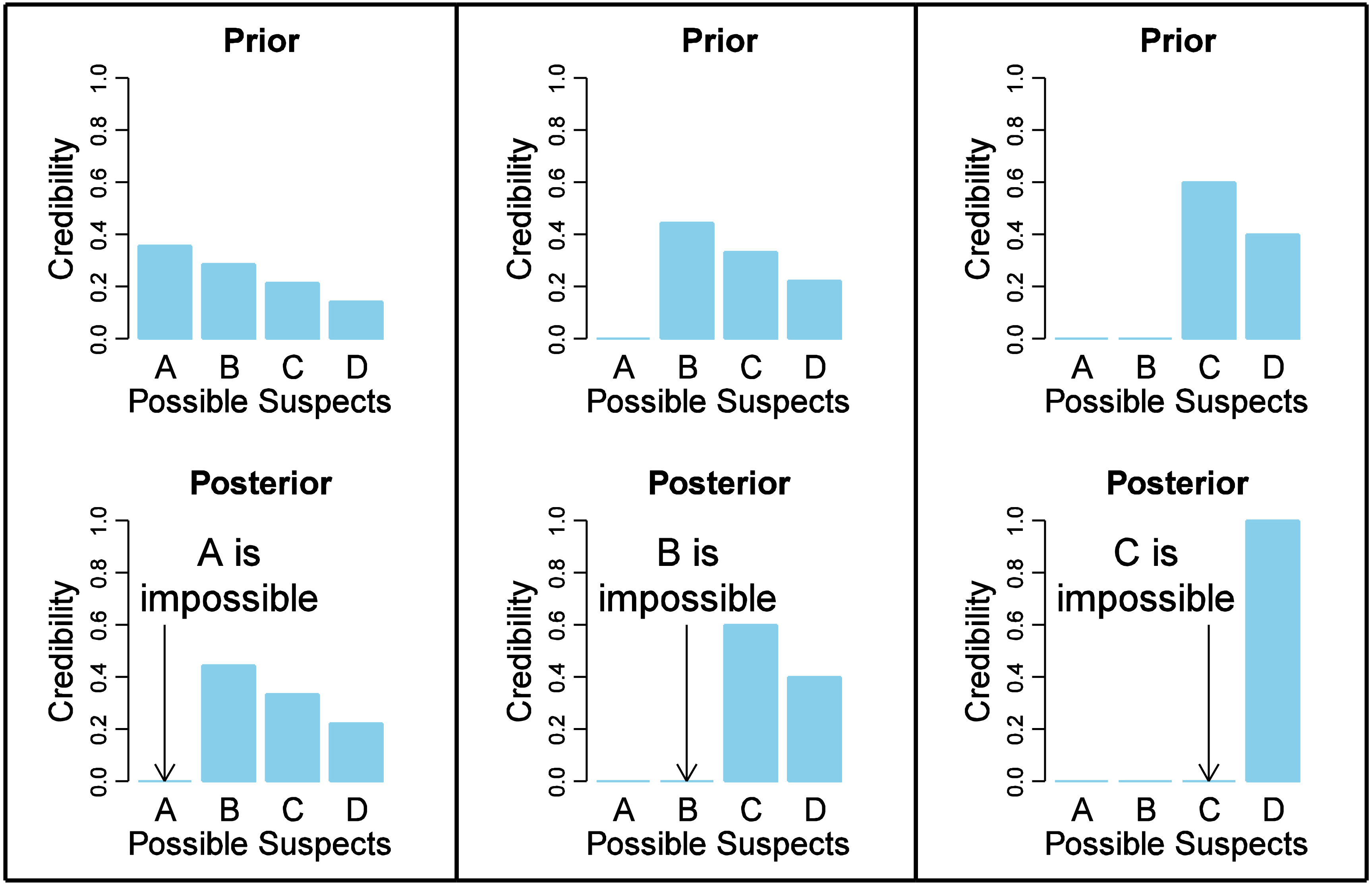
Prior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A B C D
Possible Suspects
Posterior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A is impossible
A B C D
Possible Suspects


Prior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A
B
C
D
Possible Suspects
Posterior
A is impossible
Prior
Possible Suspects



Prior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A
B
C
D
Possible Suspects
Prior
Credibility
Possible Suspects
Posterior
A is
impossible
Credibility
Possible Suspects
Posterior
B is
impossible
Credibility
Possible Suspects


Prior
Credibility
0.0 0.2 0.4 0.6 0.8 1.0
A B C D
Possible Suspects
Posterior
A is impossible
B is impossible


Prior
Credibility
0.0
0.2
0.4
0.6
0.8
1.0
A
B
C
D
Possible Suspects
Posterior
A is impossible
B is impossible
C is impossible

Two Foundational Ideas of Bayesian Reasoning

1. Bayesian reasoning is re-allocation of credibility across possibilities.
2. The possibilities are parameter values in a mathematical model of data.

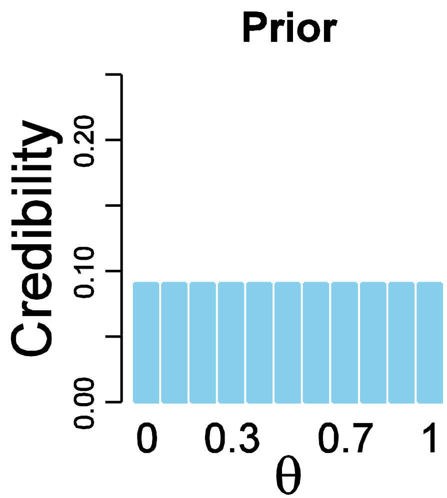
The tendency of a coin to come up heads

Data values:

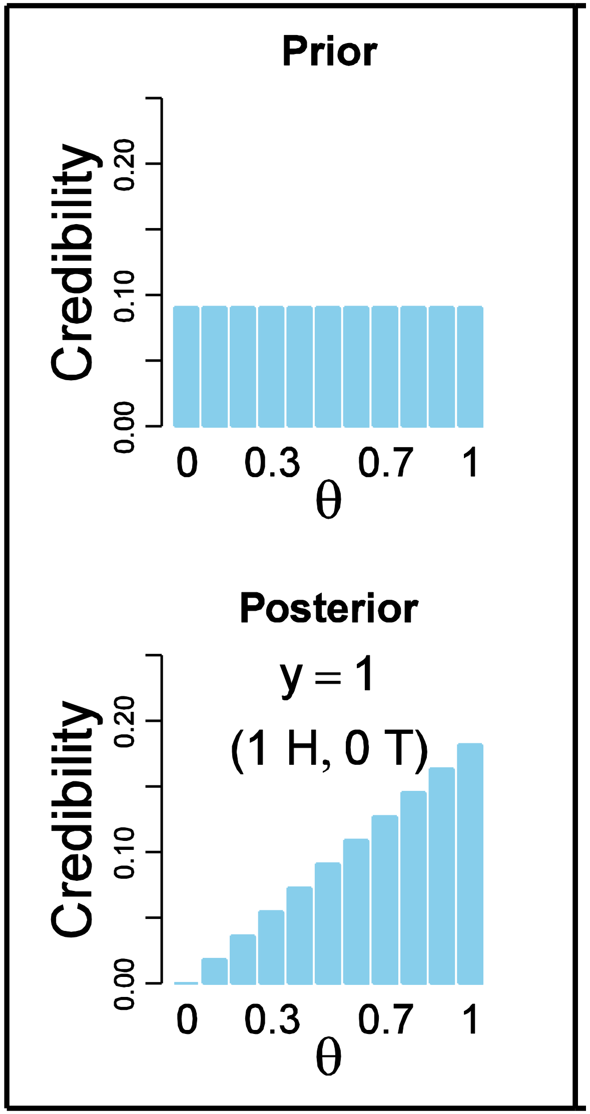
y=1 for “heads” and y=0 for “tails”

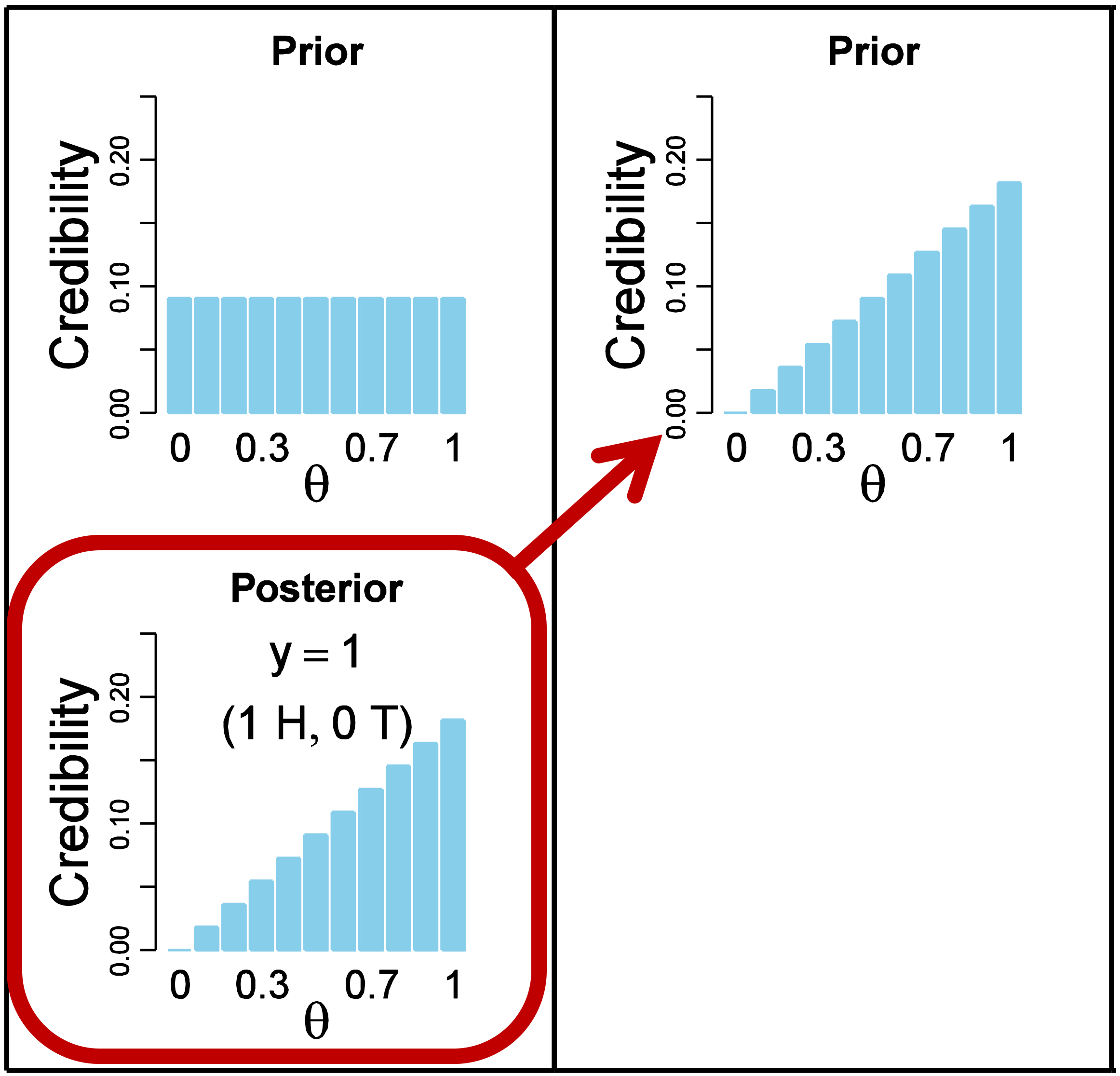
The tendency for heads is the value of the parameter θ:

$$p(y=1|\theta) = \theta$$

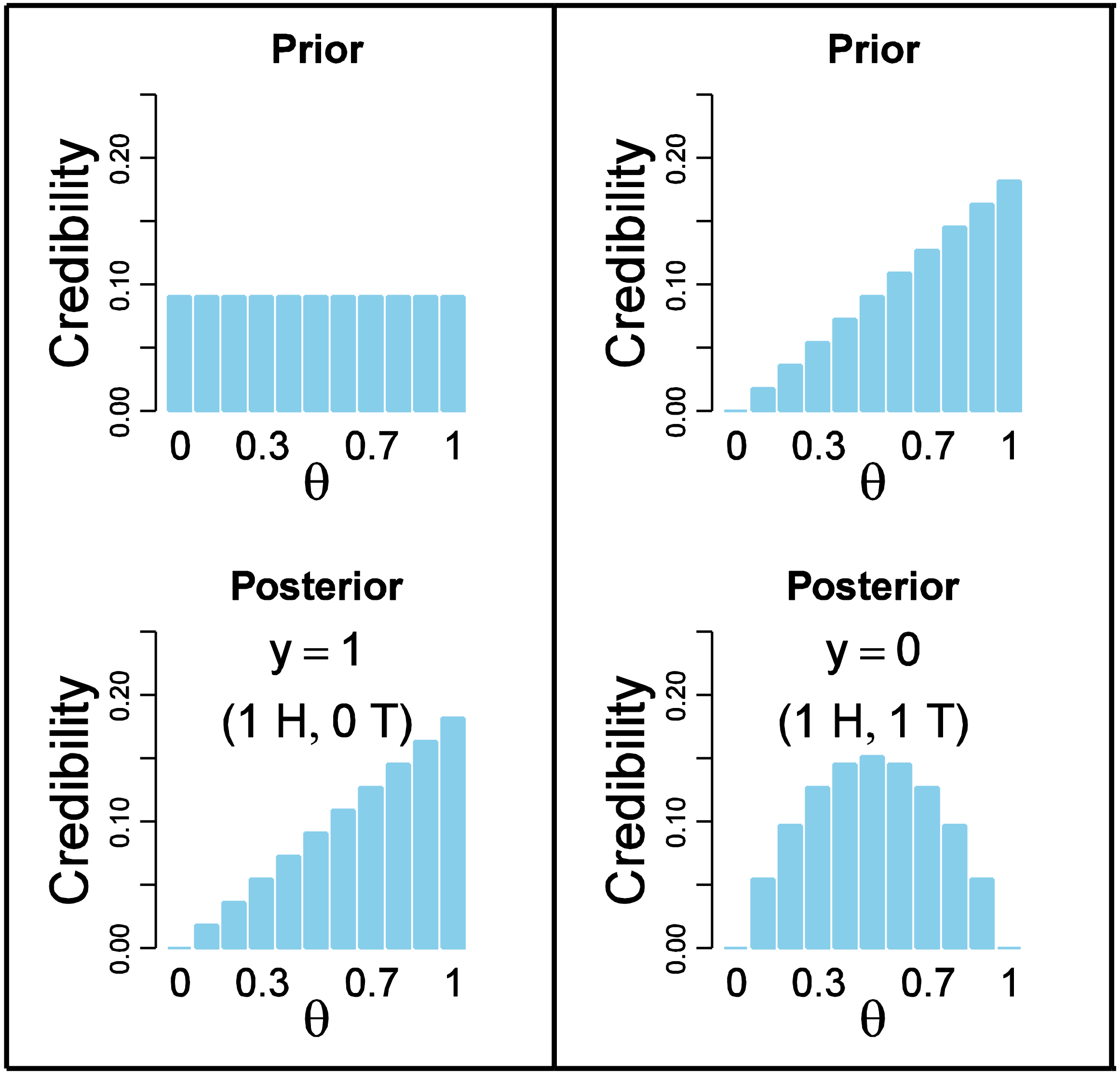


11 discrete candidate values for θ

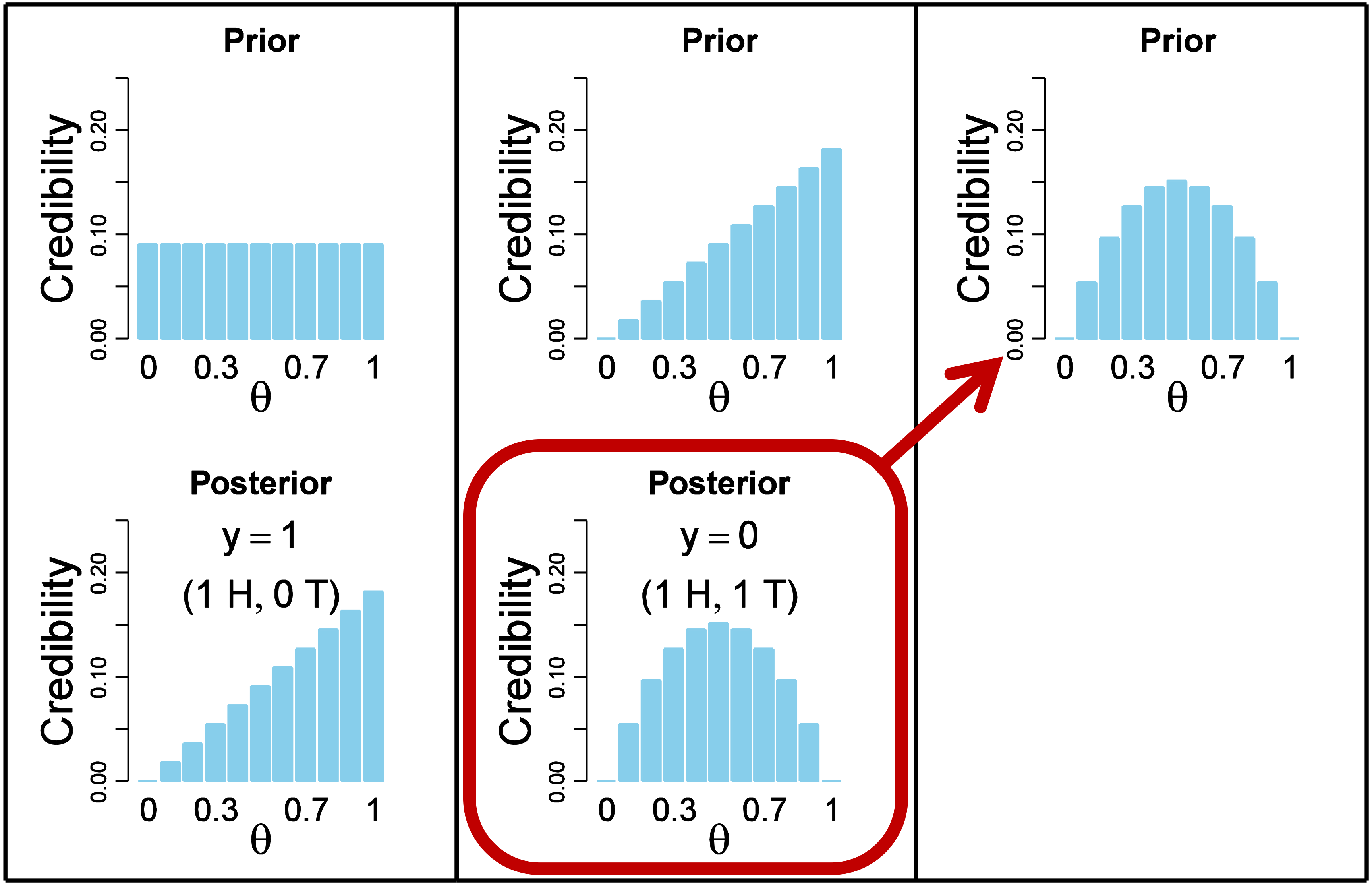

Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 1
(1 H, 0 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ

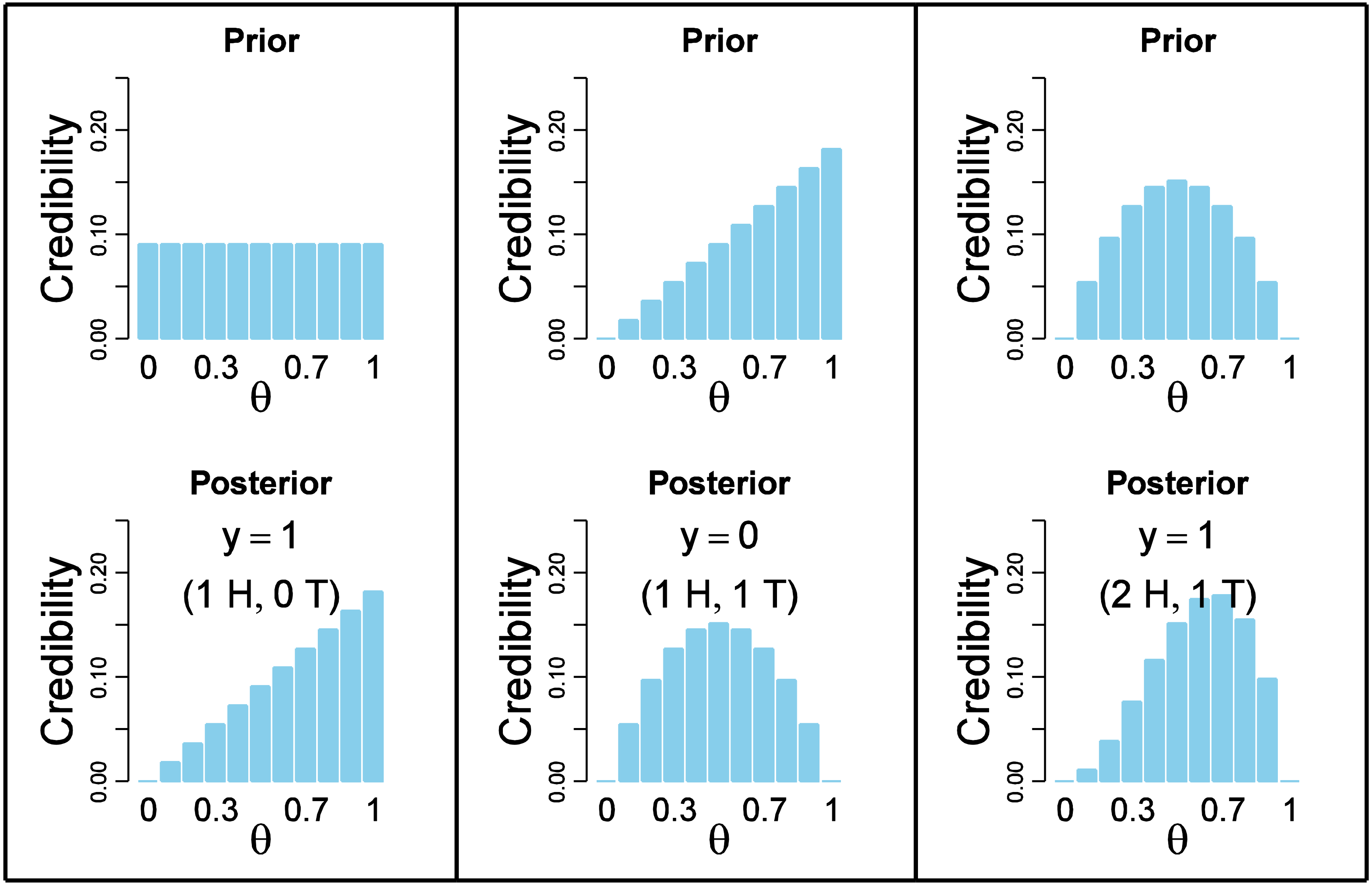


Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 1
(1 H, 0 T)



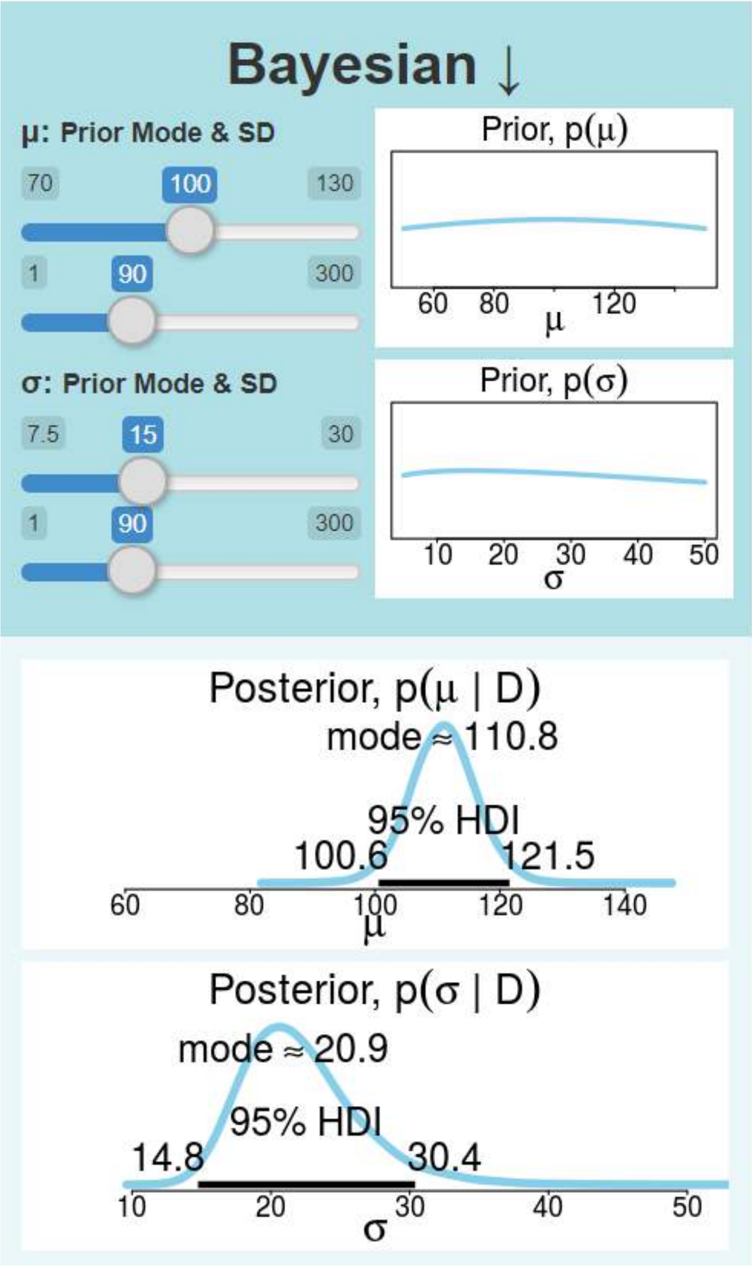
Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 1
(1 H, 0 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 0
(1 H, 1 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ


Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 1
(1 H, 0 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 0
(1 H, 1 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ

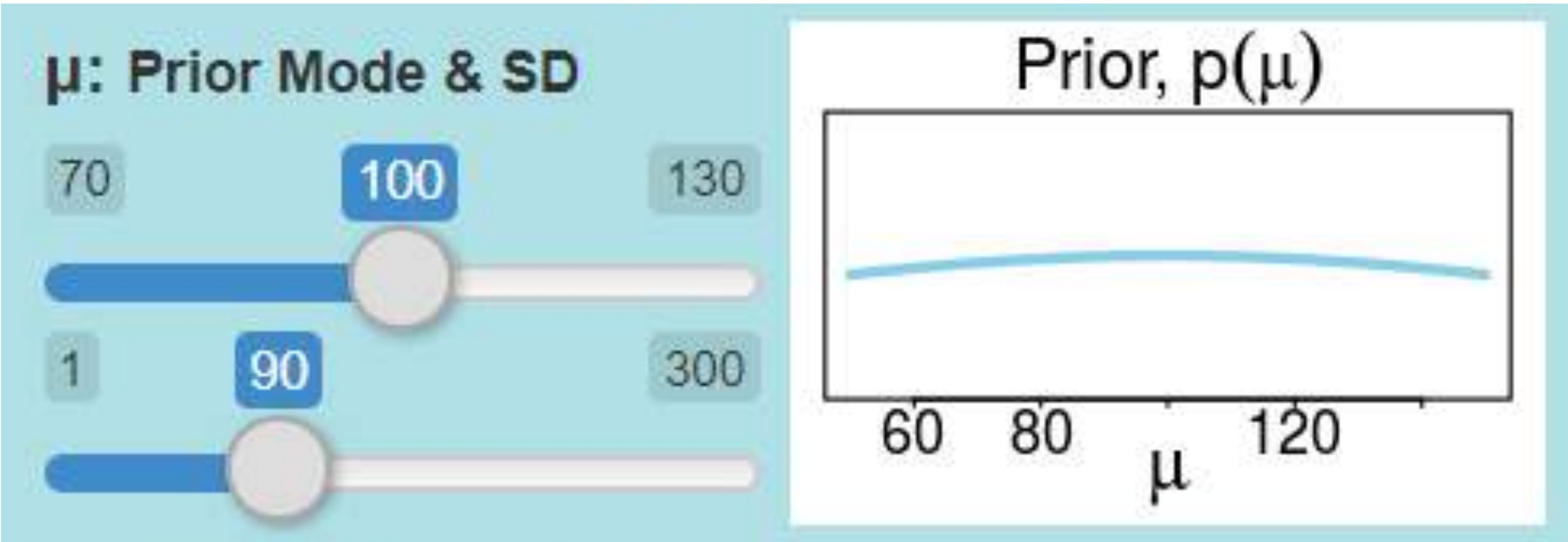


Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Prior
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 1
(1 H, 0 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 0
(1 H, 1 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ
Posterior
y = 1
(2 H, 1 T)
Credibility
0.00
0.10
0.20
0
0.3
0.7
1
θ

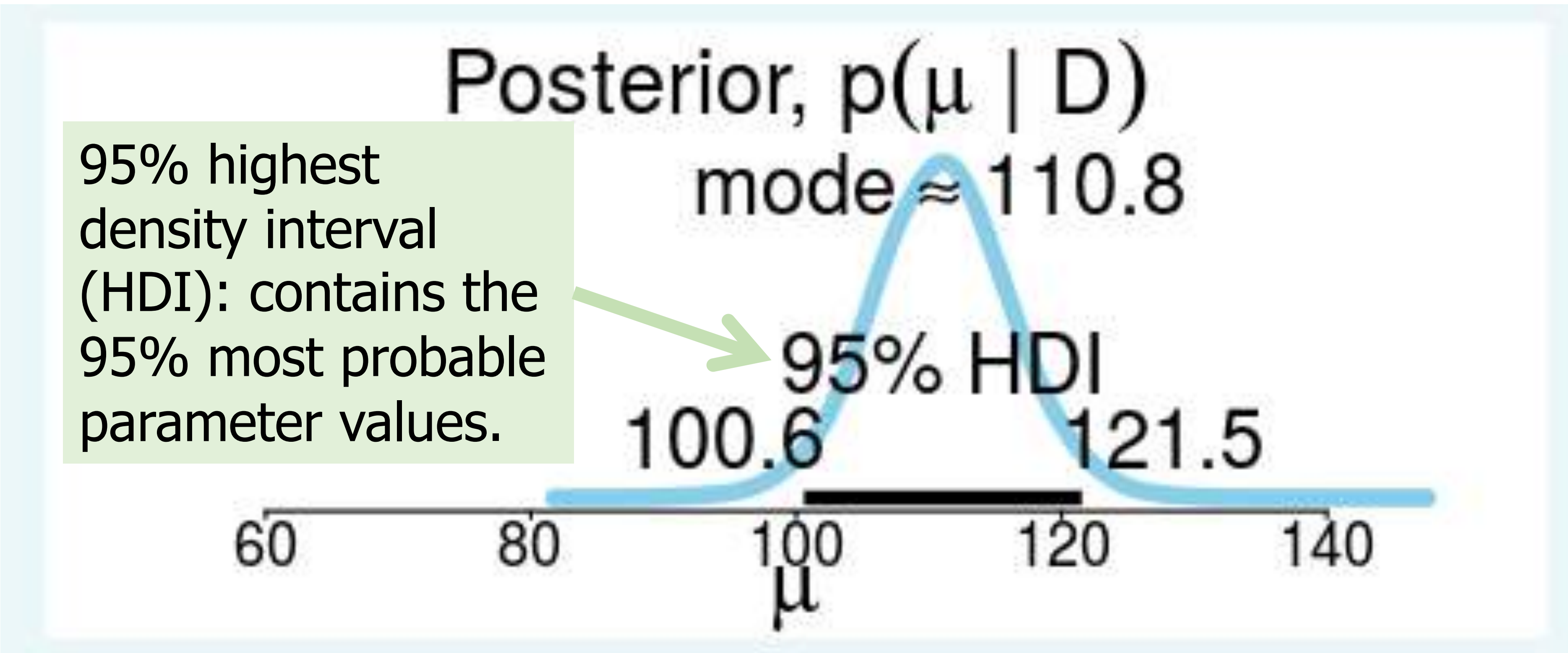
4. Bayesian Estimation



4. Bayesian Estimation



4. Bayesian Estimation

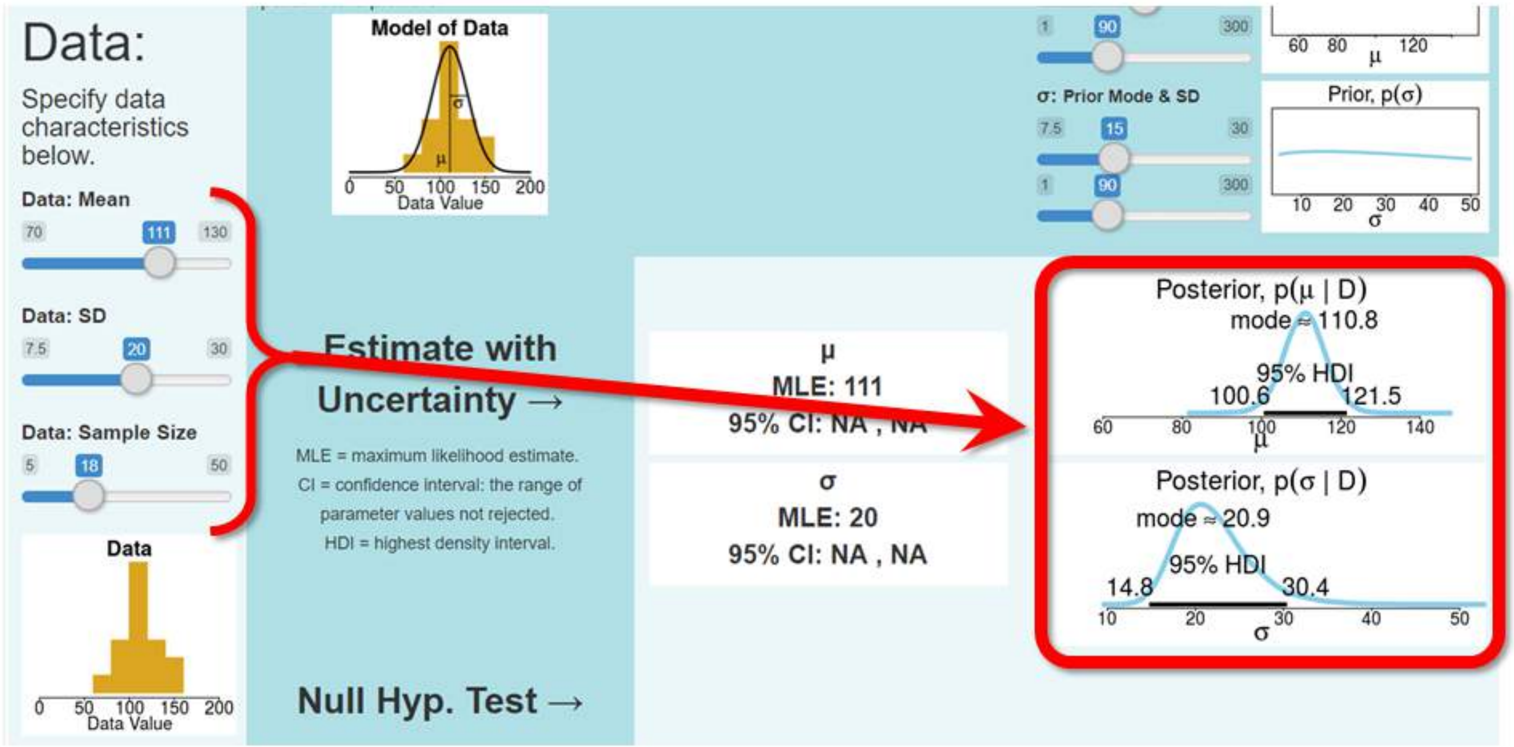


95% highest density interval (HDI): contains the 95% most probable parameter values.

4. Bayesian Estimation

Try It!

Manipulate the data sliders and watch the effect on the posterior distribution.

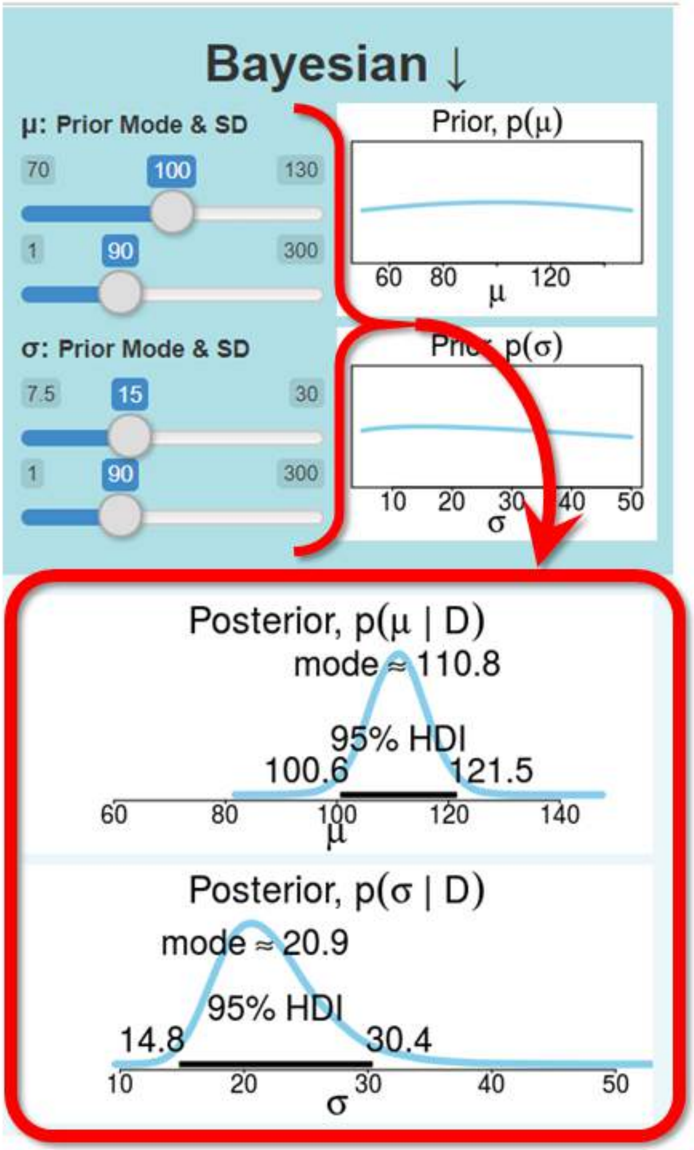


Notice: Posterior modes track data. HDI is narrower for smaller SD, & for larger N.

4. Bayesian Estimation

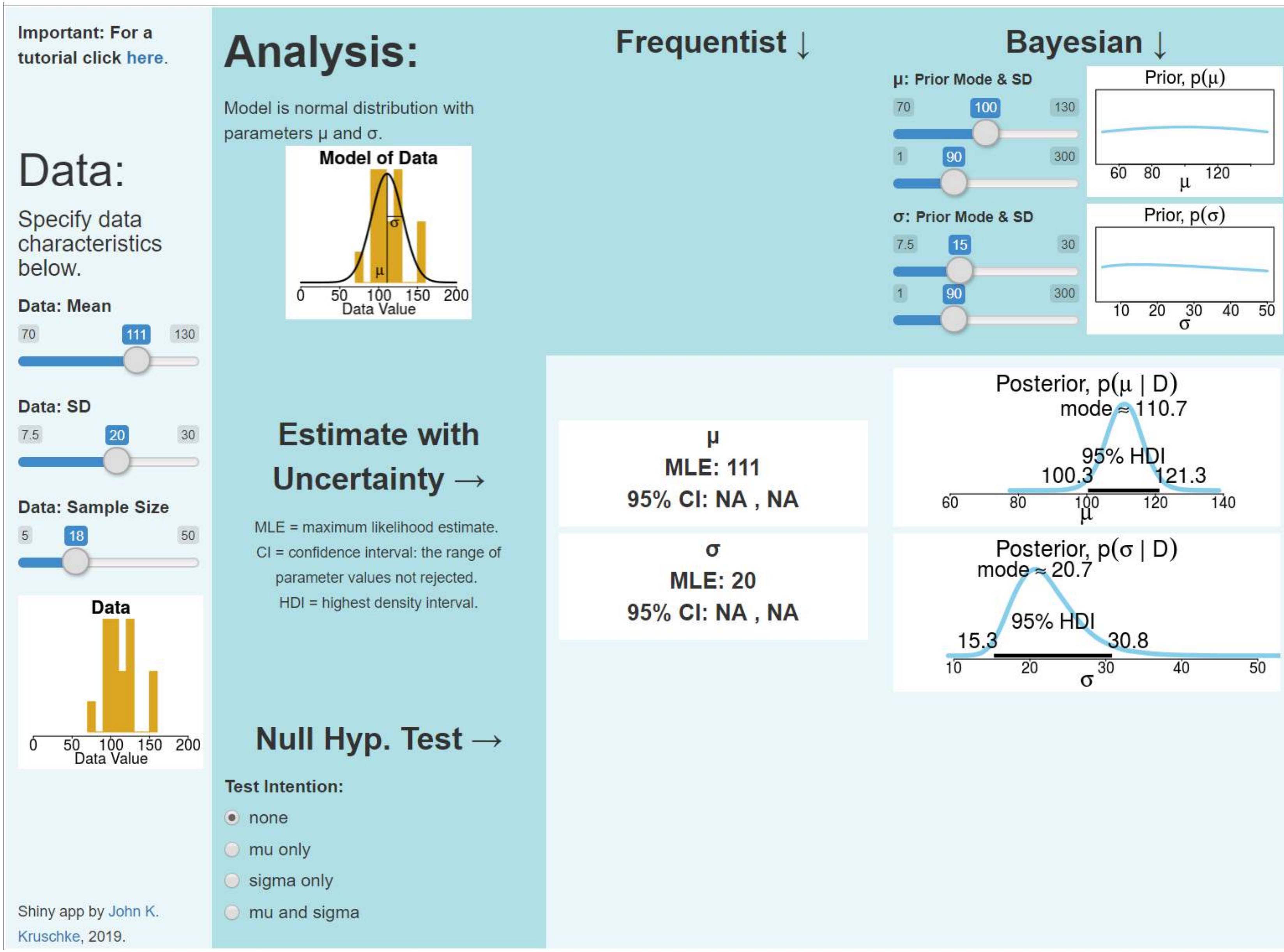
Try It!

Manipulate the prior distribution sliders and watch the effect on the posterior distribution.

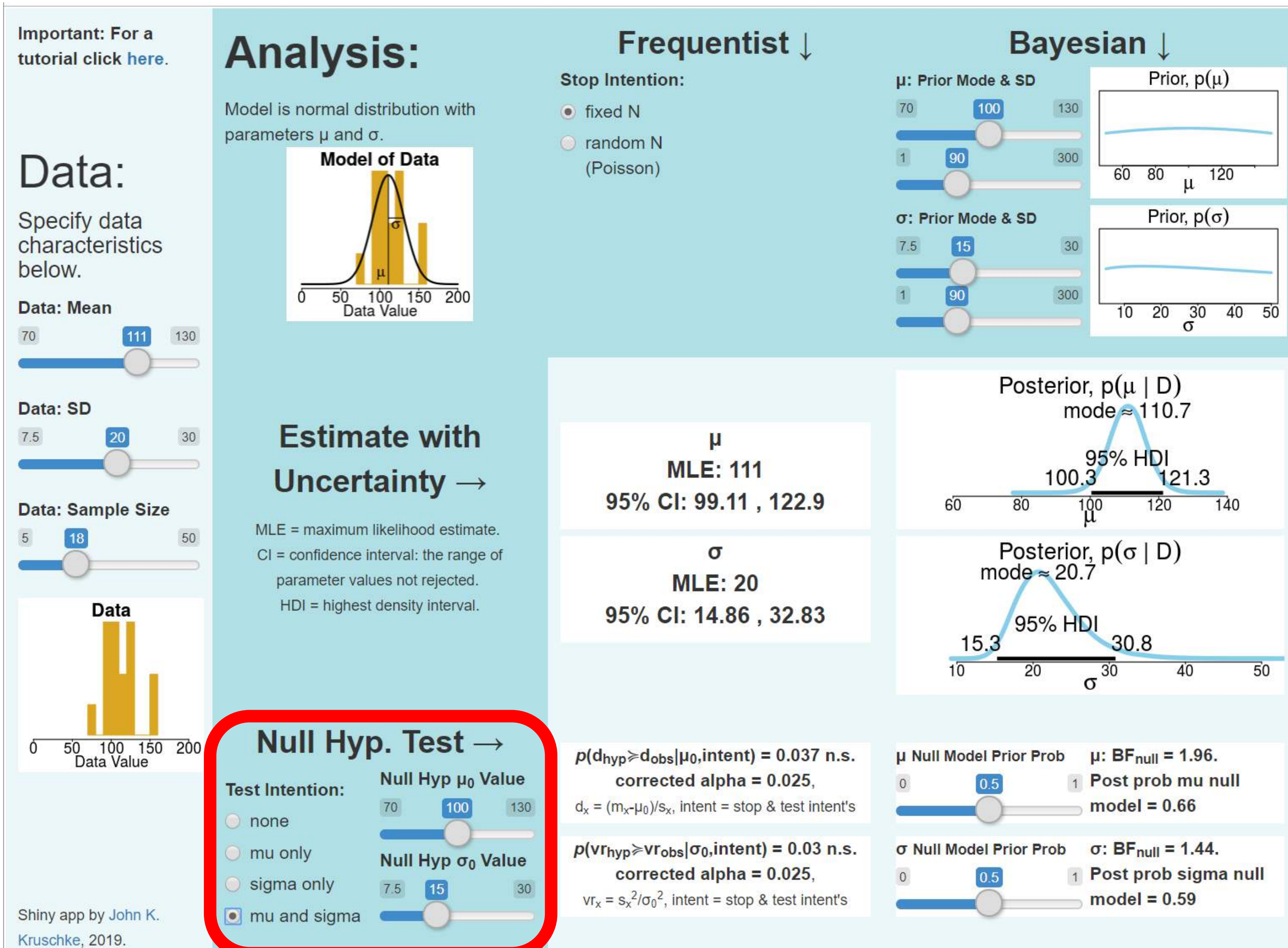


Notice: Any broad prior has minimal influence on posterior.

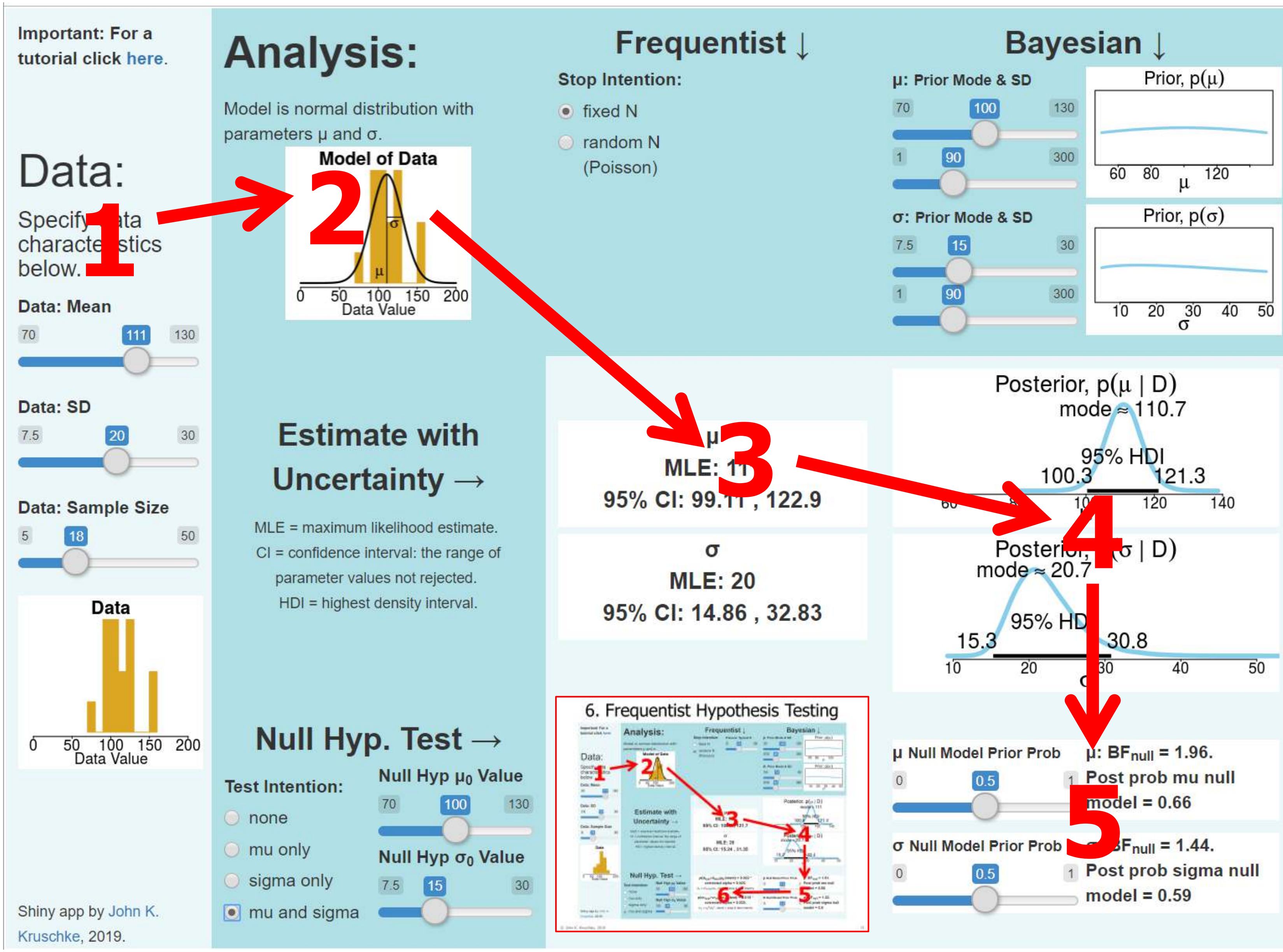
What If There Were No Significance Tests?



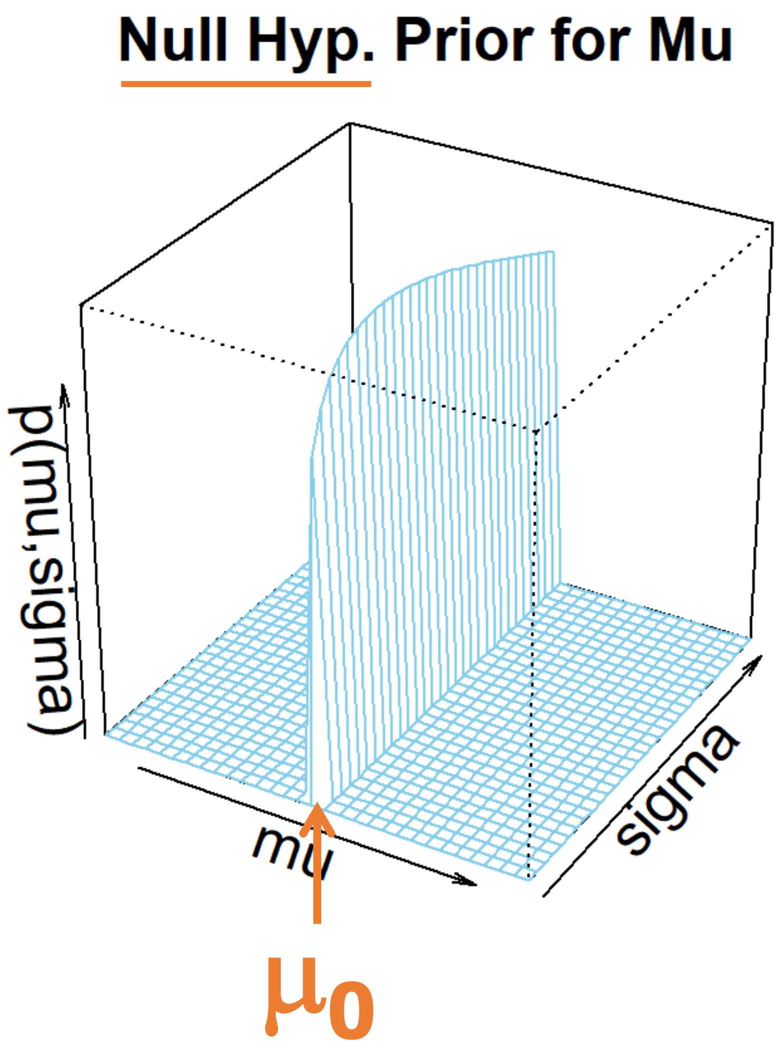
Hypothesis Testing



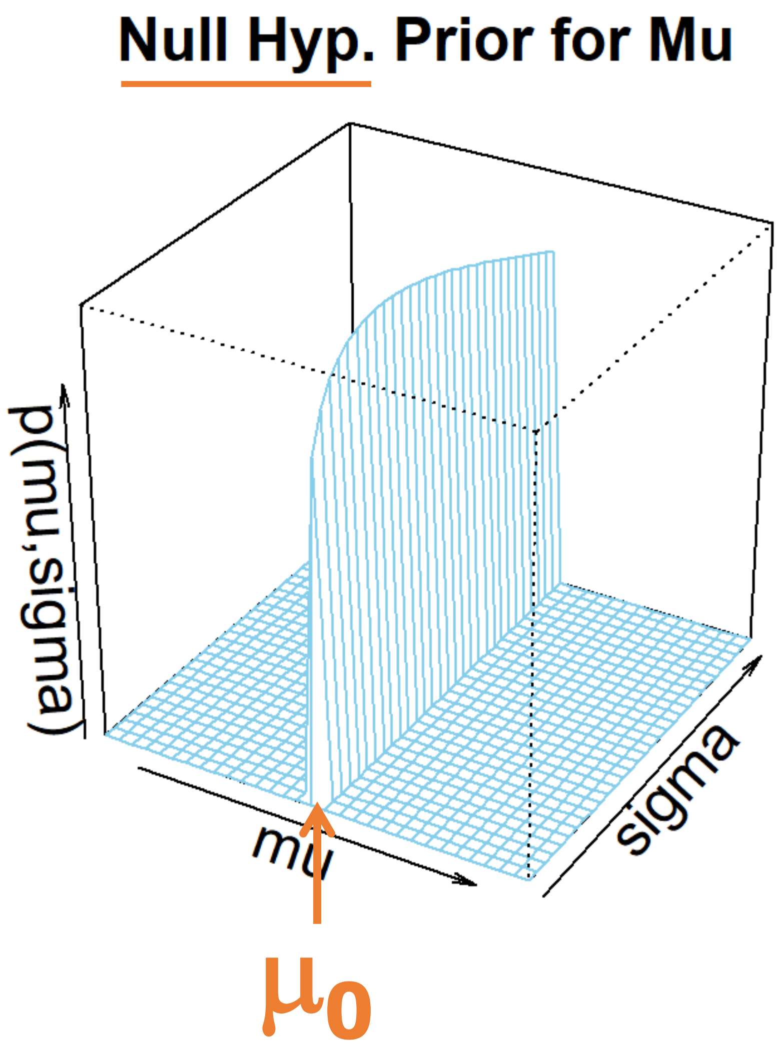
5. Bayesian Hypothesis Testing

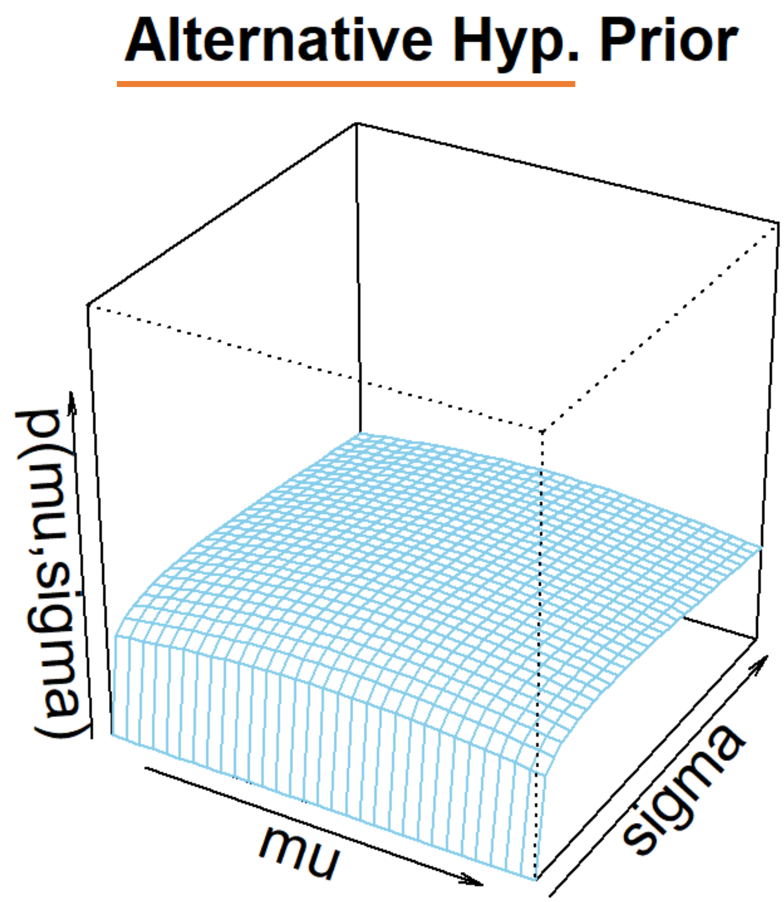


5. Bayesian Hypothesis Testing



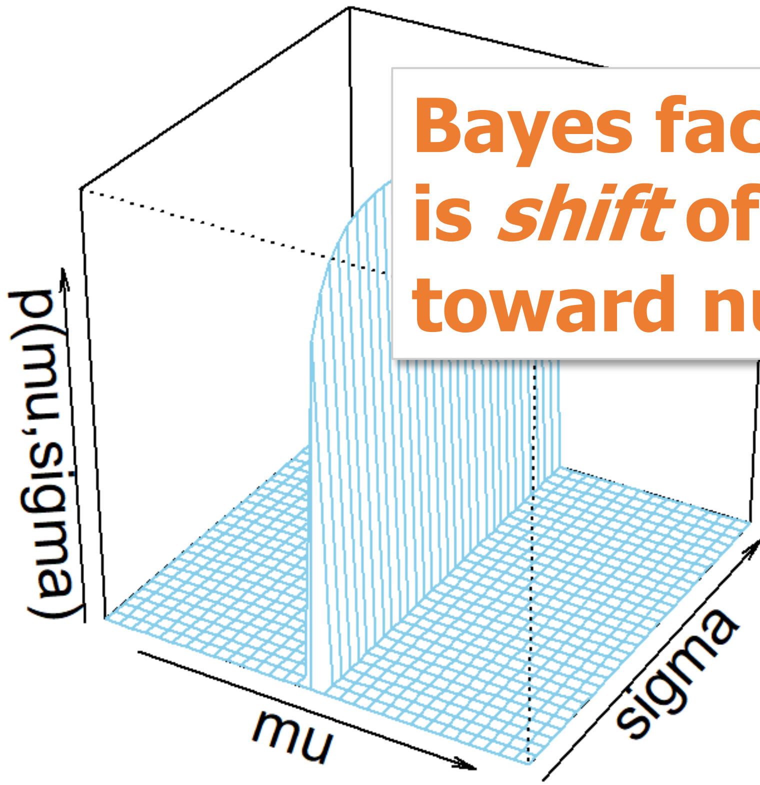
5. Bayesian Hypothesis Testing



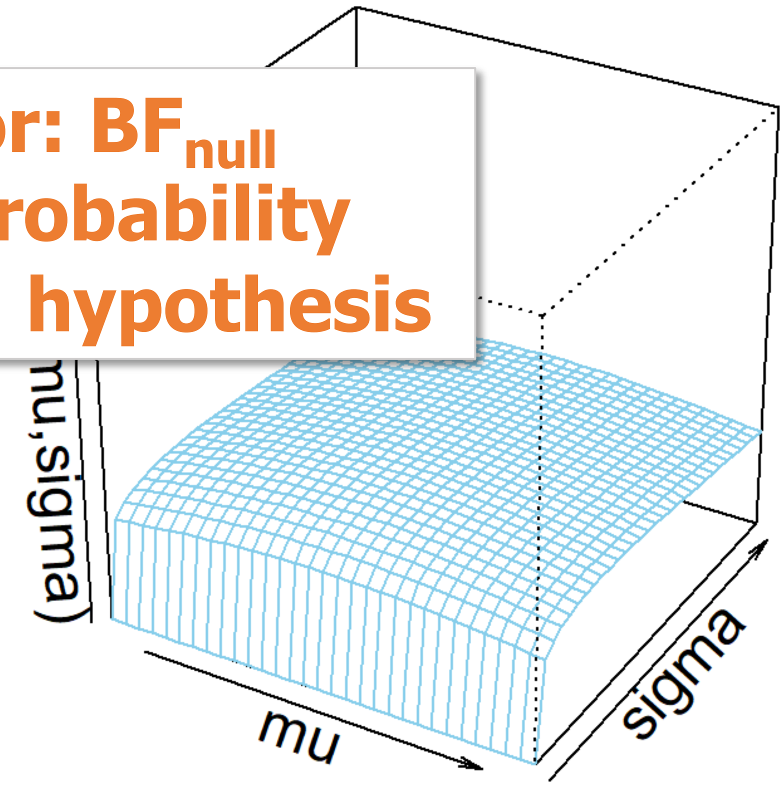


5. Bayesian Hypothesis Testing

Null Hyp. Prior for Mu



Alternative Hyp. Prior

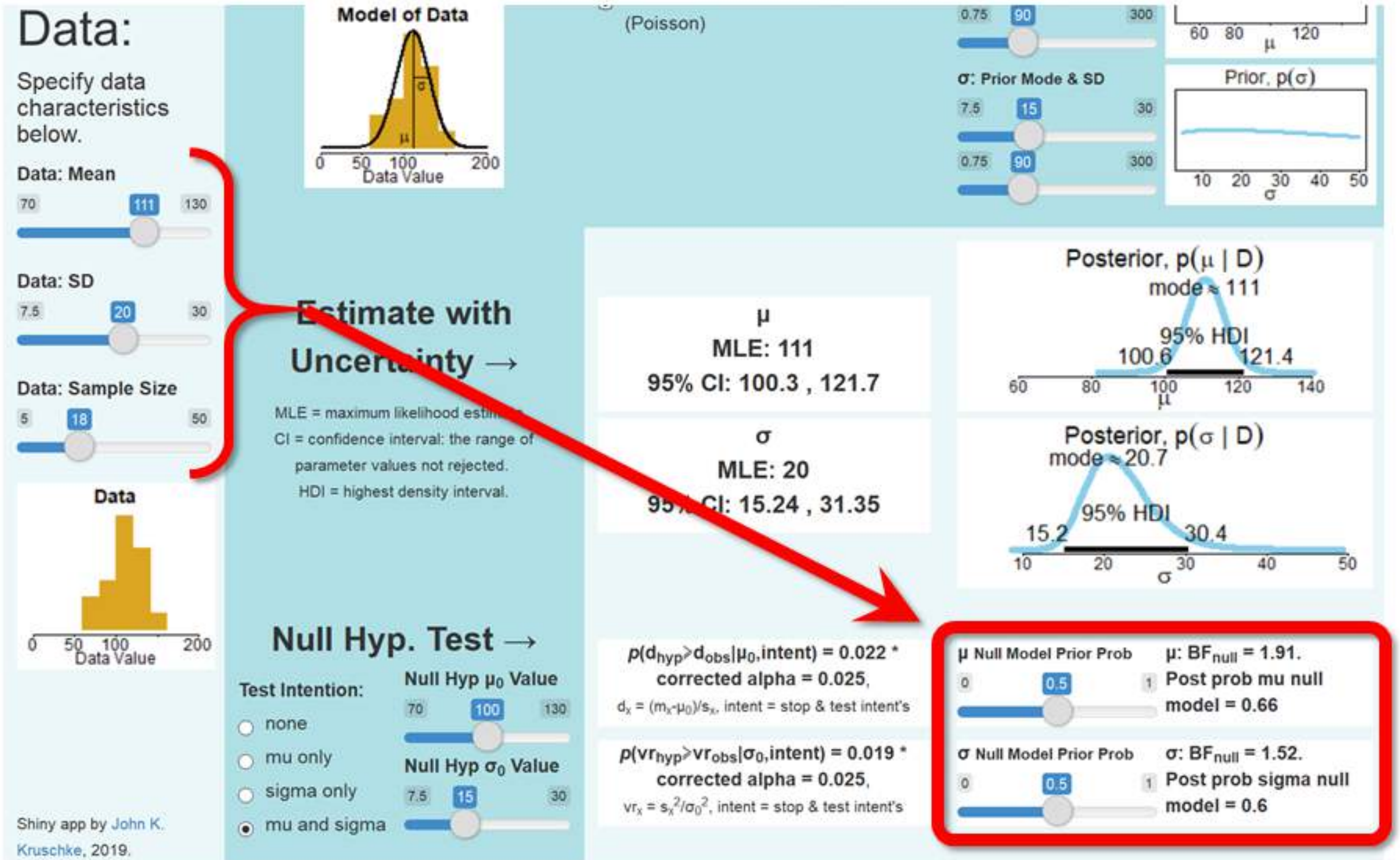


Bayes factor: BF_{null} is *shift* of probability toward null hypothesis

5. Bayesian Hypothesis Testing

Try It!

Manipulate the **Data** sliders, watch the effect on the BF_{null}'s and posterior prob's of the models.

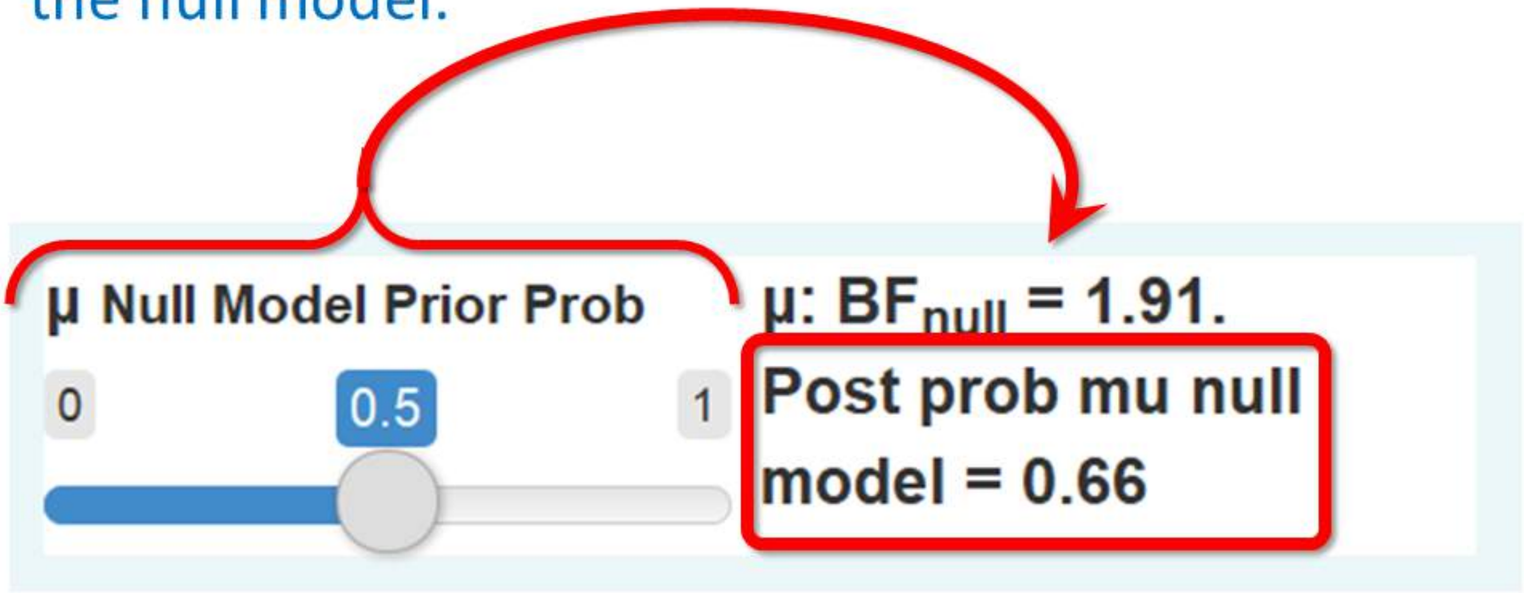


Notice: When data mean is close to μ_0 then BF_{null} is greater than 1.0; *etc.*

5. Bayesian Hypothesis Testing

Try It!

Manipulate the prior probability of the null model and watch the effect on the posterior probability of the null model.

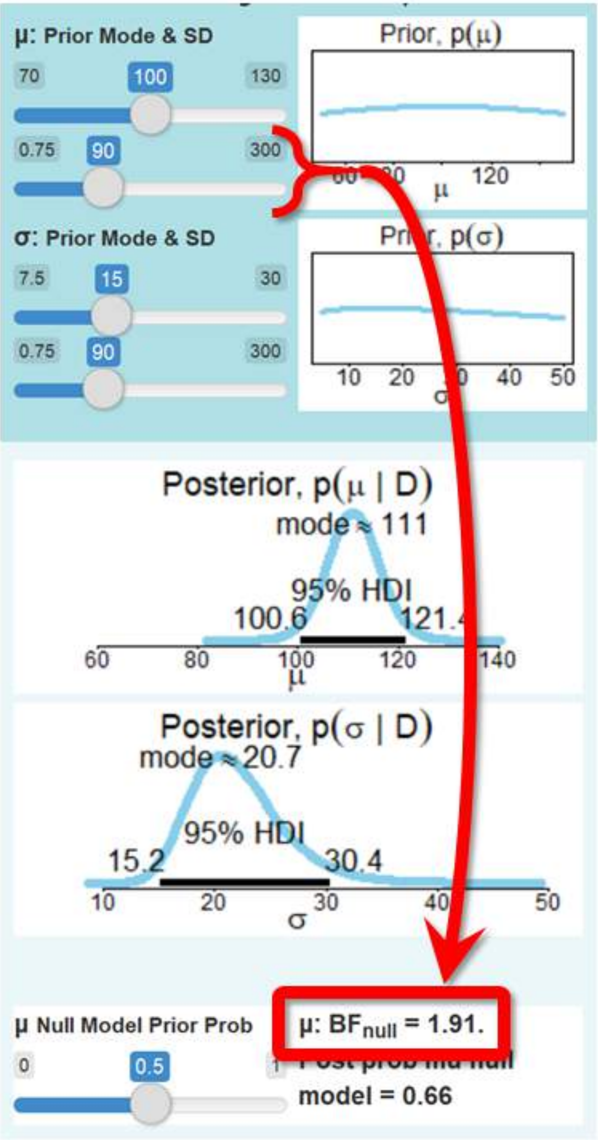


Notice: BF_{null} is *shift* of prob, not the posterior prob of null.

5. Bayesian Hypothesis Testing

Try It!

Manipulate the
μ Prior SD
and watch the effect
on the Bayes factor
(and on the posterior
HDI).

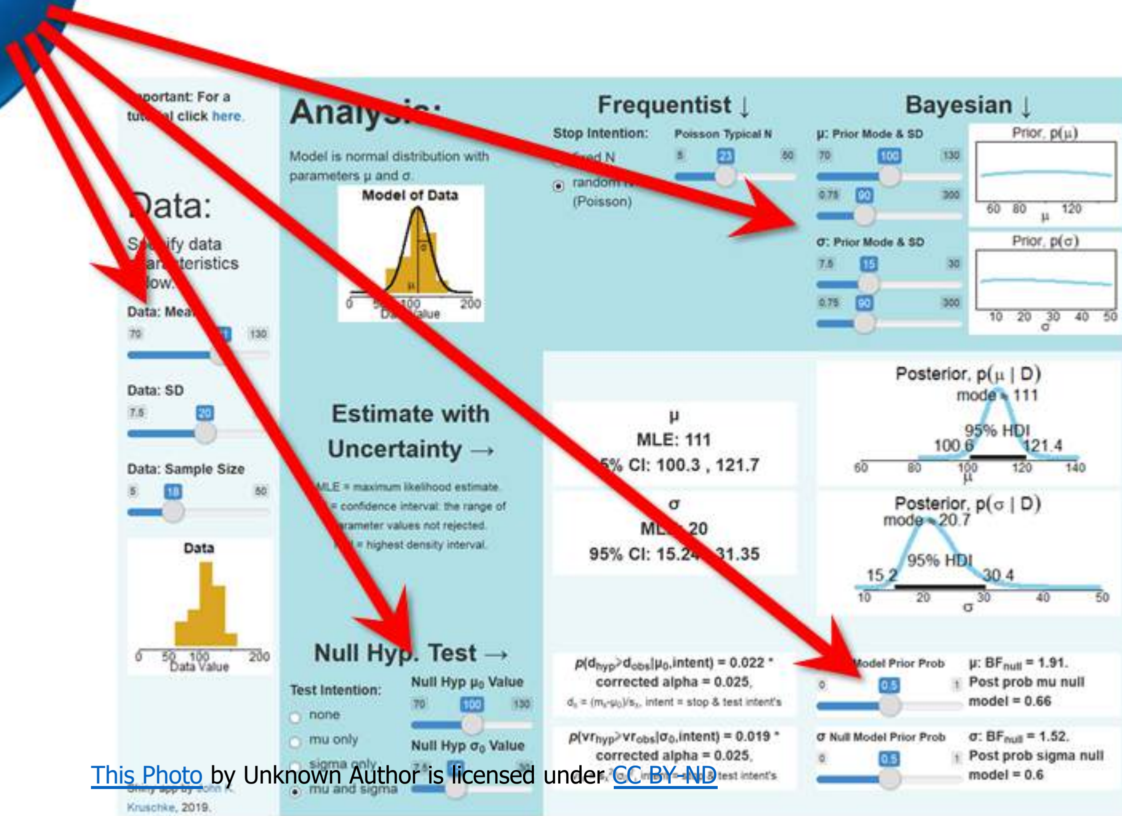


Notice: BF_{null} is strongly affected by SD of prior (but HDI is not).

5. Bayesian Hypothesis Testing

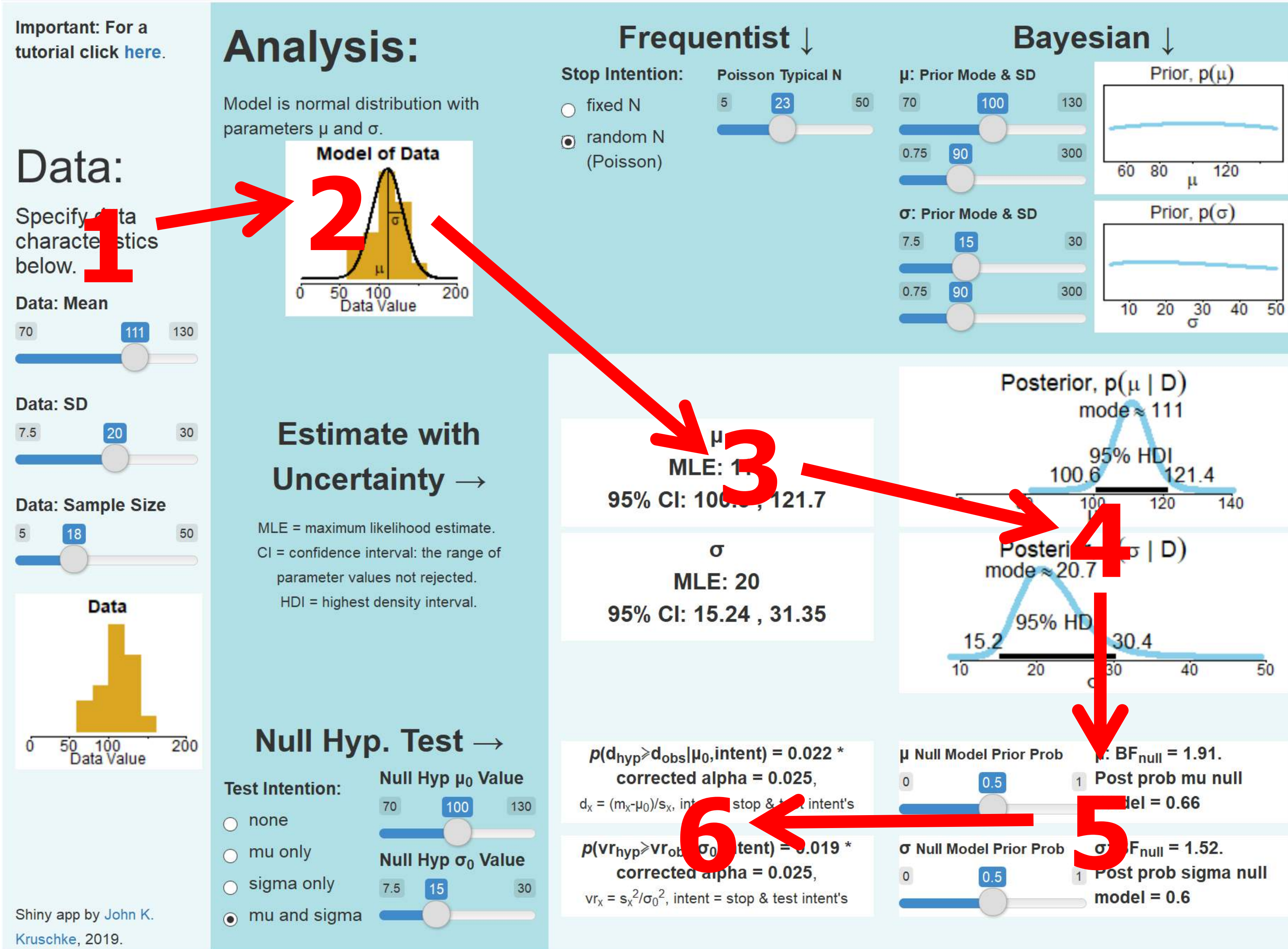
Try It!

Set the sliders to represent situational information.

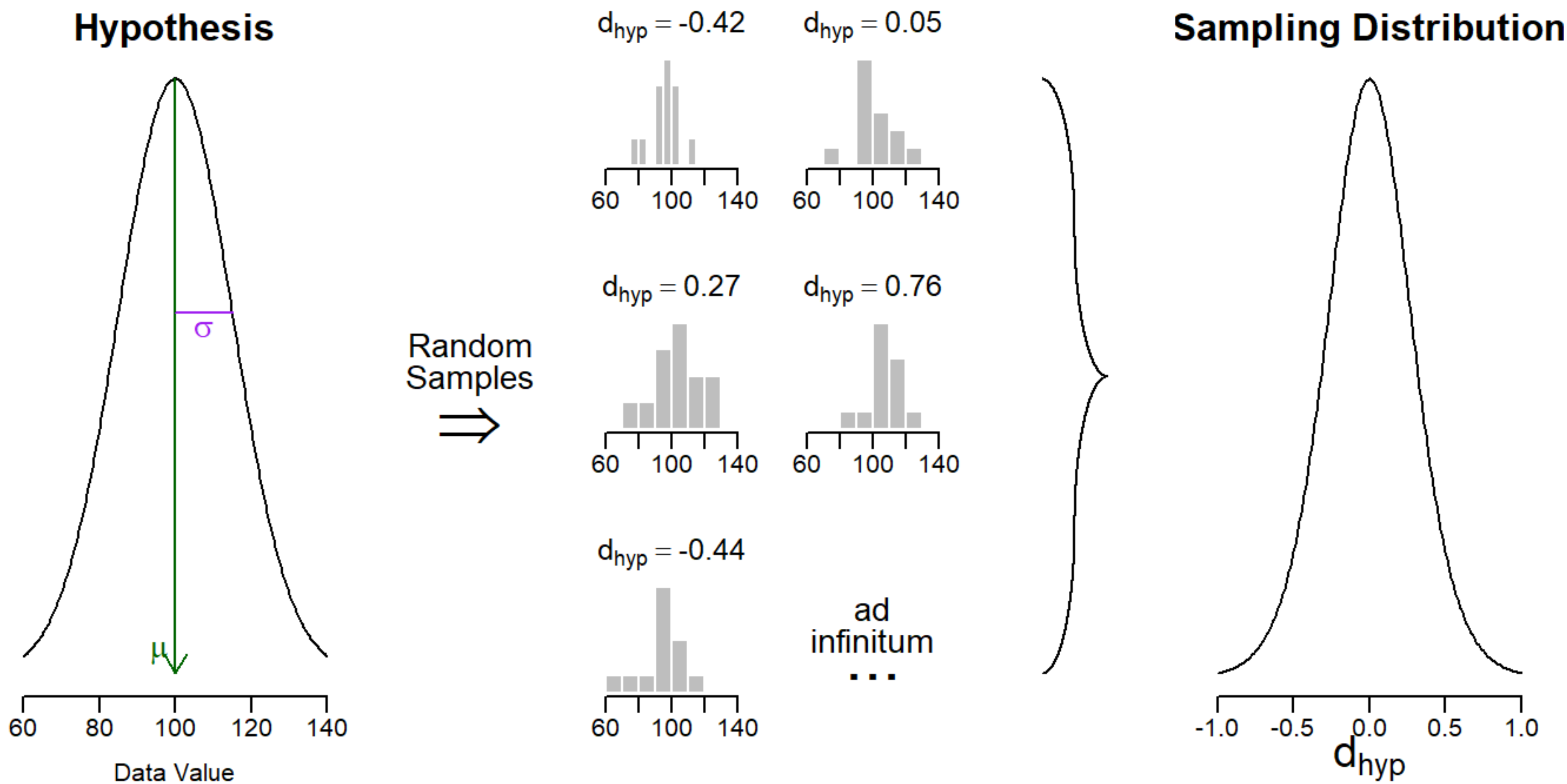


Notice: Prior on parameter is not same as prior on null model.

6. Frequentist Hypothesis Testing



6. Frequentist Hypothesis Testing



6. Frequentist Hypothesis Testing

$p(d_{hyp} \geqslant d_{obs} | \mu_0, \text{intent}) = 0.022$ *

corrected alpha = 0.025,

$d_x = (m_x - \mu_0)/s_x$, intent = stop & test intent's

$p(vr_{hyp} \geqslant vr_{obs} | \sigma_0, \text{intent}) = 0.019$ *

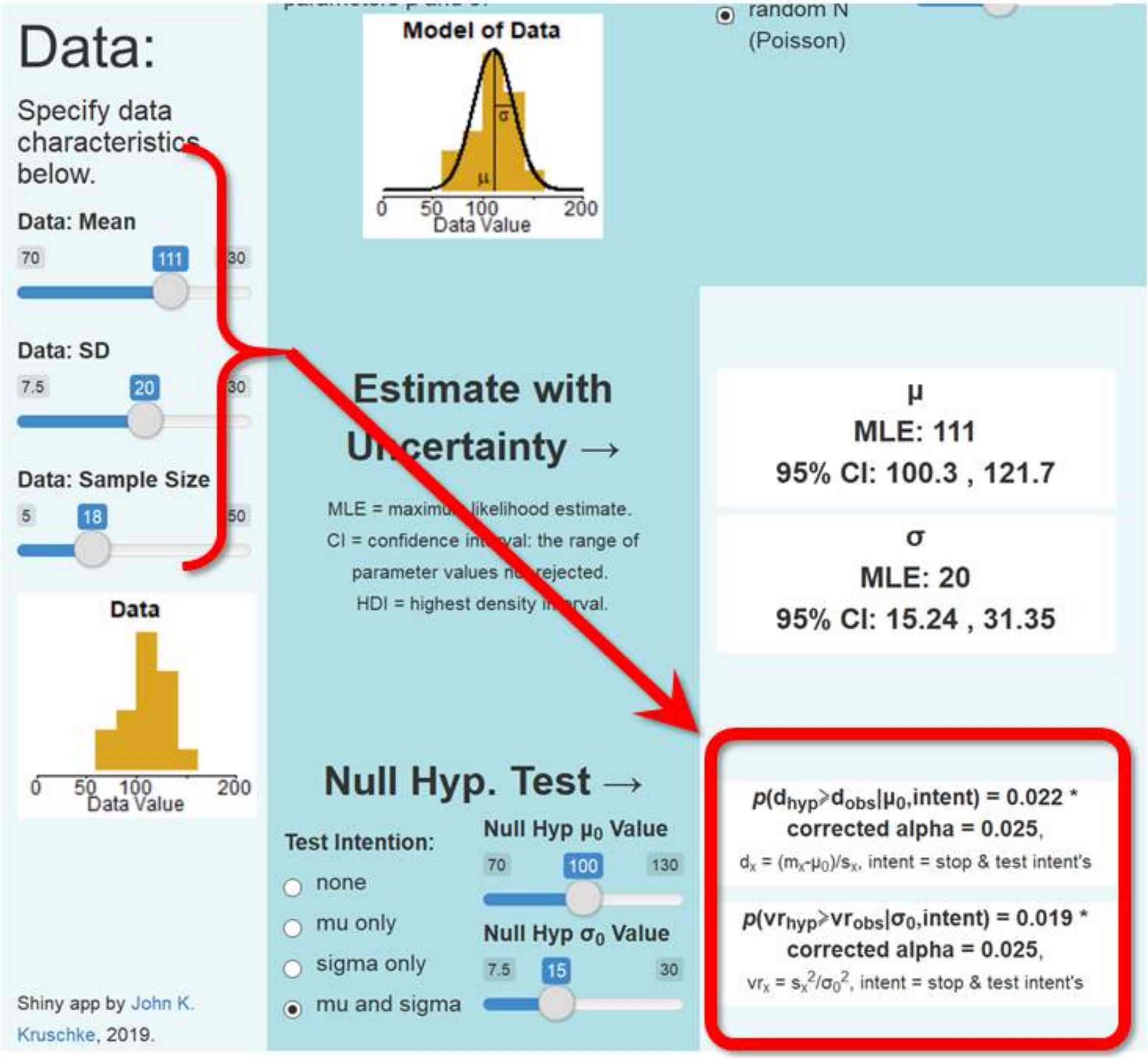
corrected alpha = 0.025,

$vr_x = s_x^2/\sigma_0^2$, intent = stop & test intent's

6. Frequentist Hypothesis Testing

Try It!

Manipulate the data sliders and watch the effect on the *p* values.

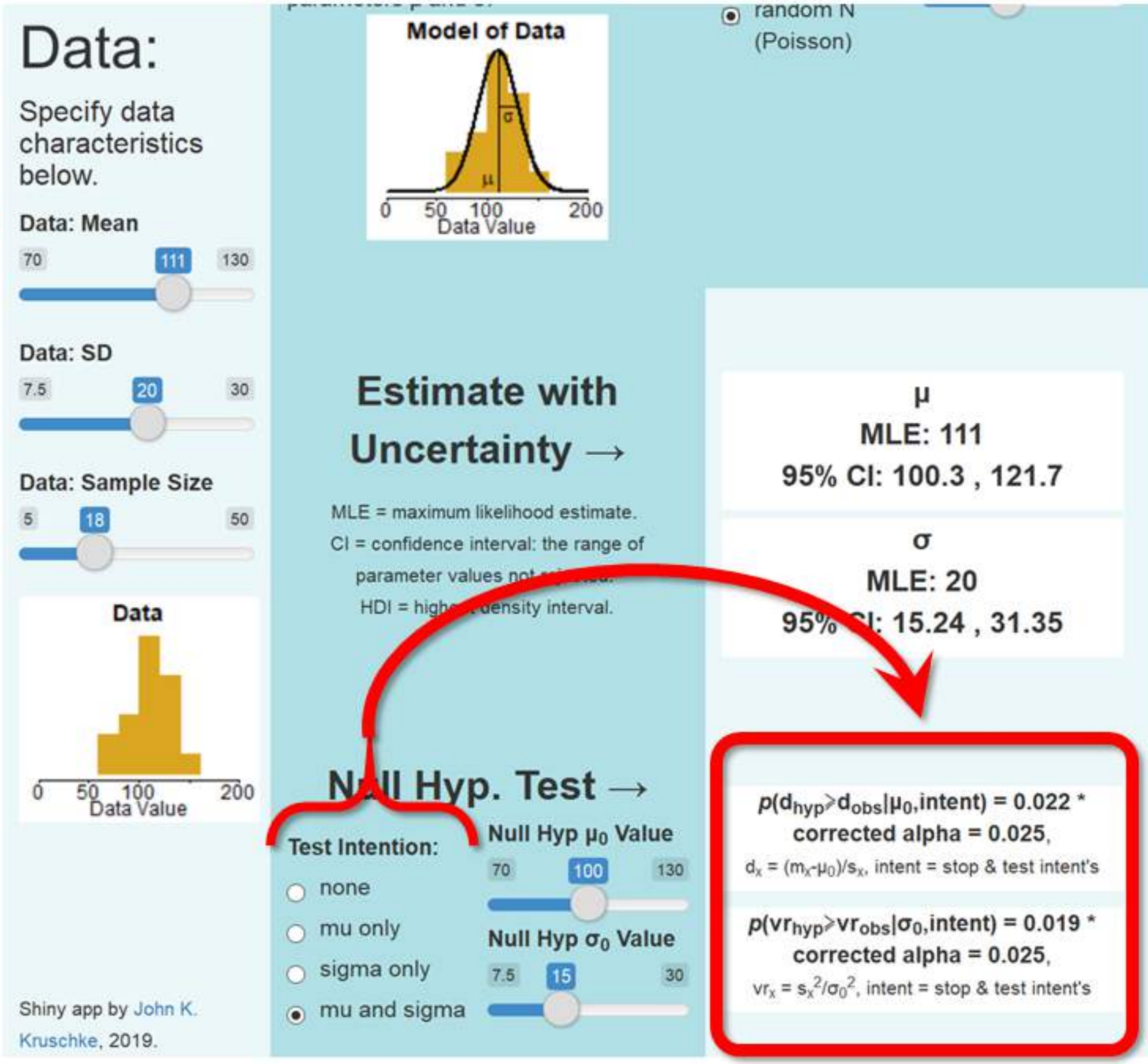


Notice: When data mean is close to μ_0 then *p* is large; *etc.*

6. Frequentist Hypothesis Testing

Try It!

Manipulate the **Test Intention** button and watch the effect on the alpha value and * vs **n.s.** marking.

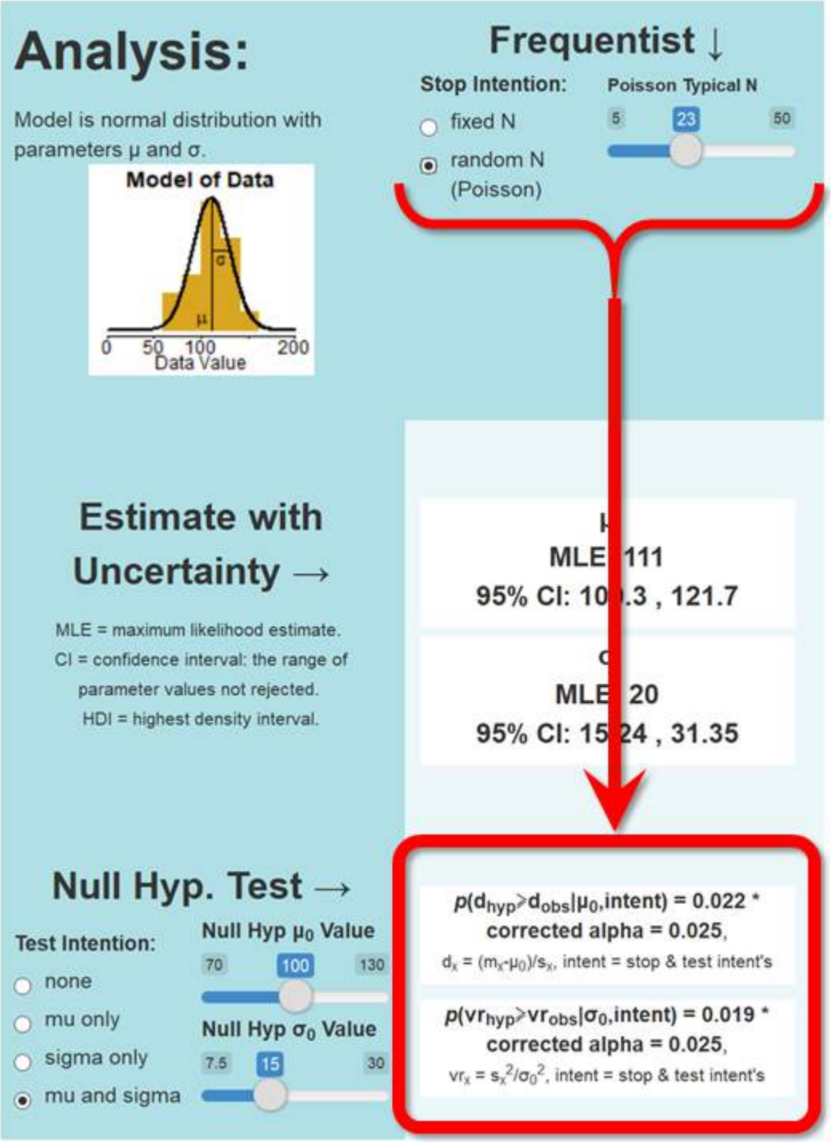


Notice: Multiple tests imply more stringent alpha.

6. Frequentist Hypothesis Testing

Try It!

Manipulate the Stop Intention button and Poisson Typical N slider and watch the effect on the *p* values (and on the CI's).

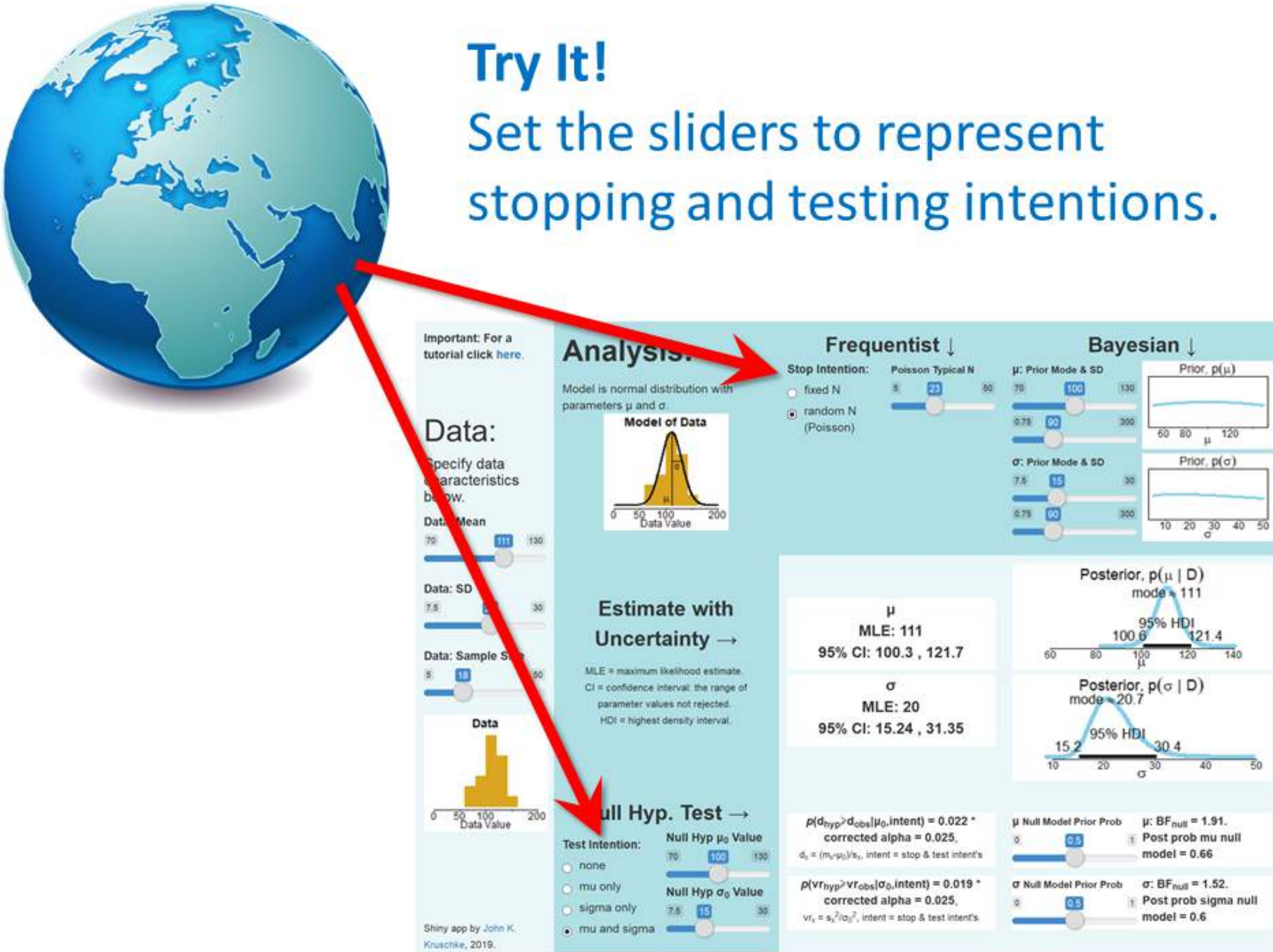


Notice: Larger typical N implies smaller *p* values (and narrower CI's).

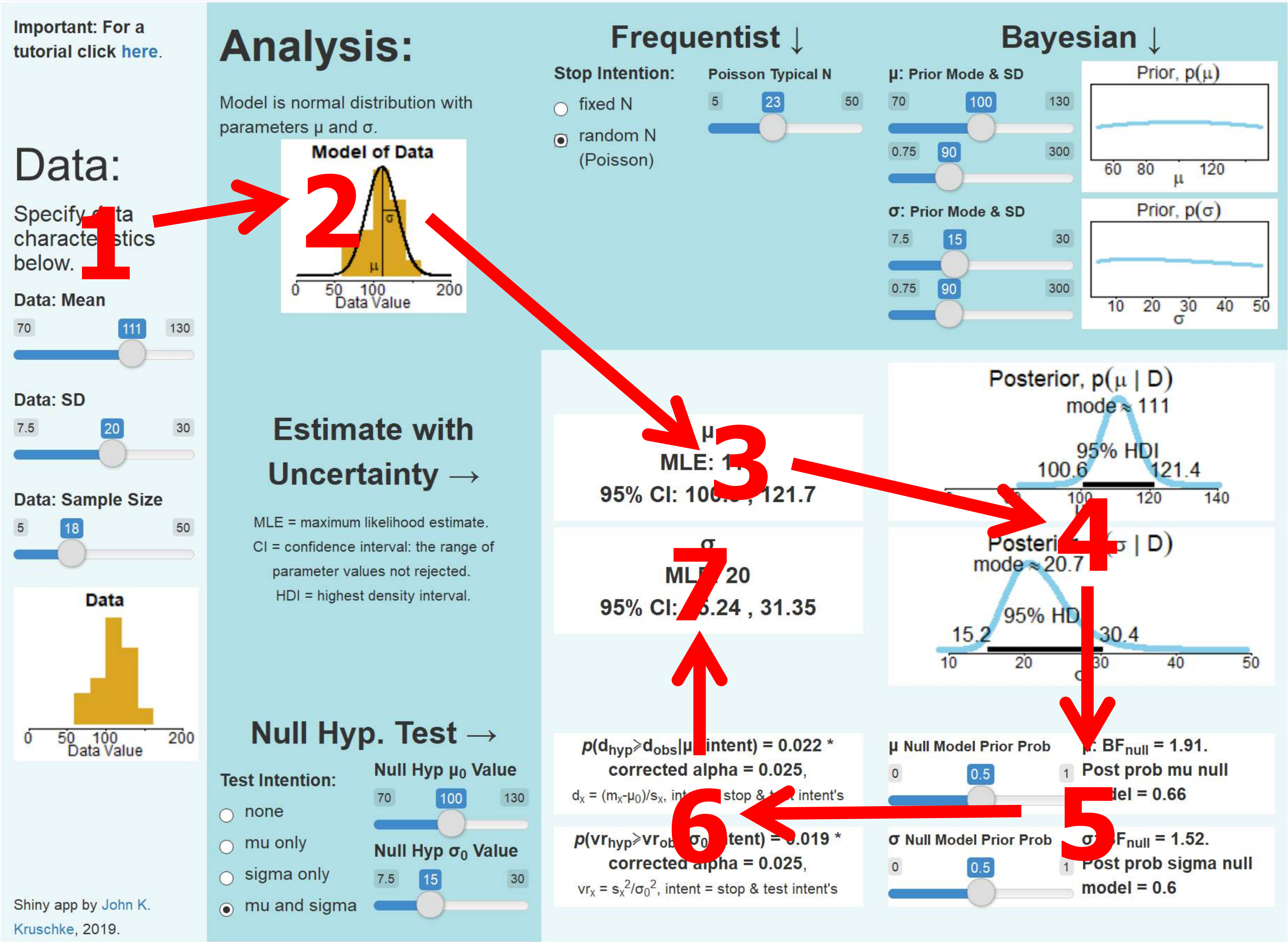
6. Frequentist Hypothesis Testing

Try It!

Set the sliders to represent stopping and testing intentions.



7. Frequentist Uncertainty



7. Frequentist Uncertainty

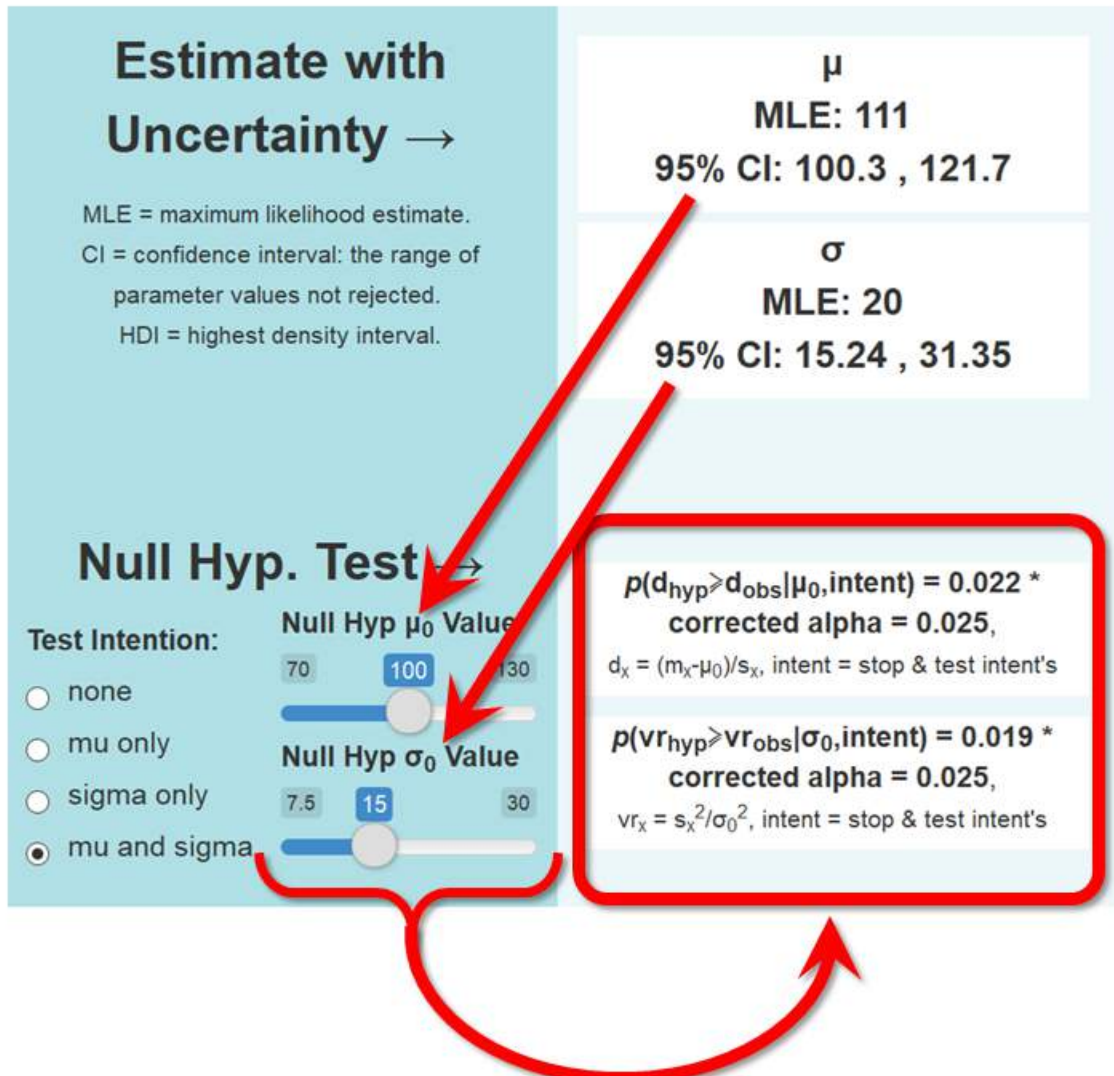
The 95% Confidence Interval
is
the parameter values
not rejected at $p < .05$

i.e.,
the parameter values
with $p \geq .05$

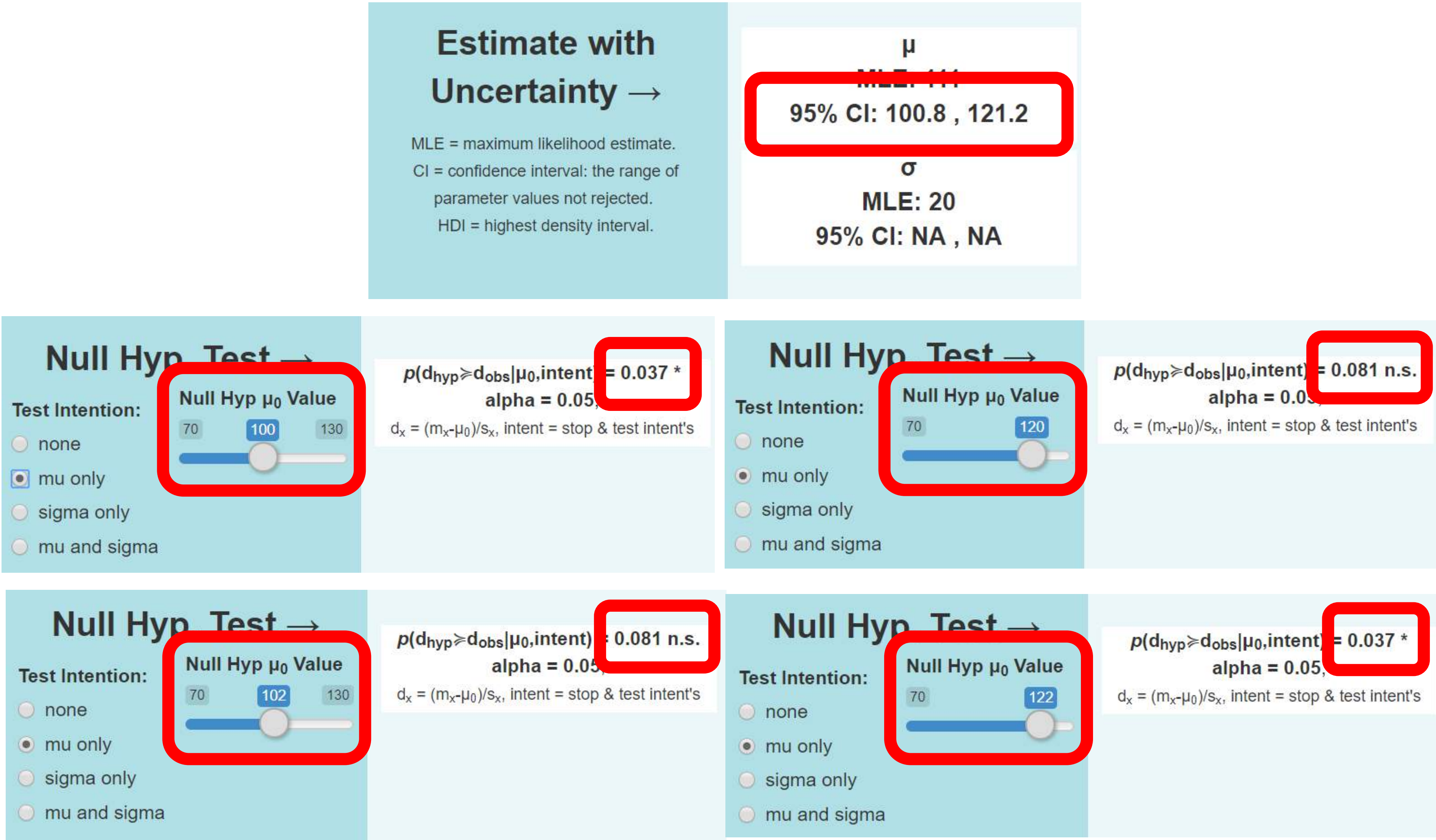
7. Frequentist Uncertainty

Try It!

Manipulate **Null Hyp** sliders to values inside and outside of CI limits and watch the *p* values.



7. Frequentist Uncertainty



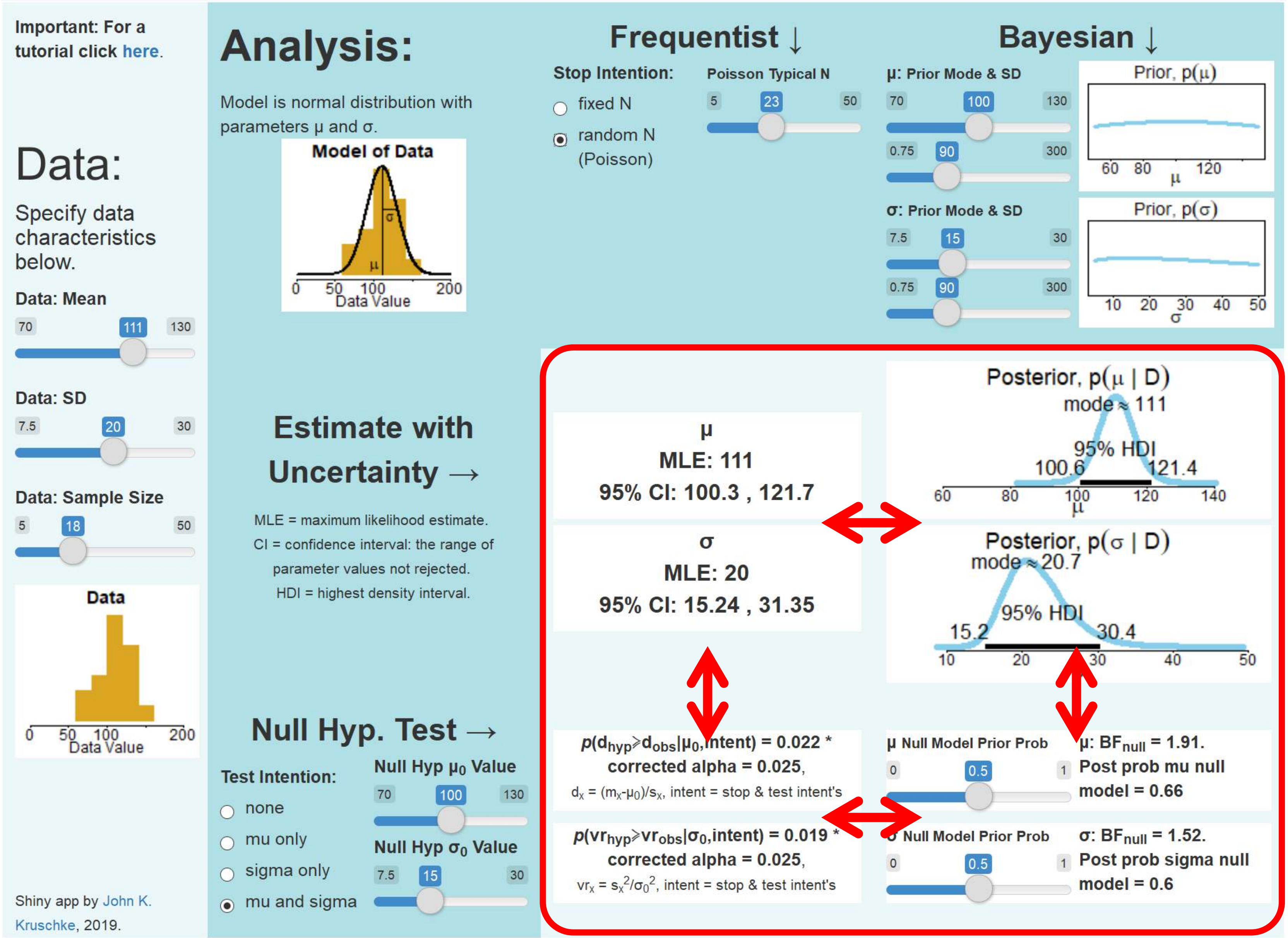
7. Frequentist Uncertainty

Try It!

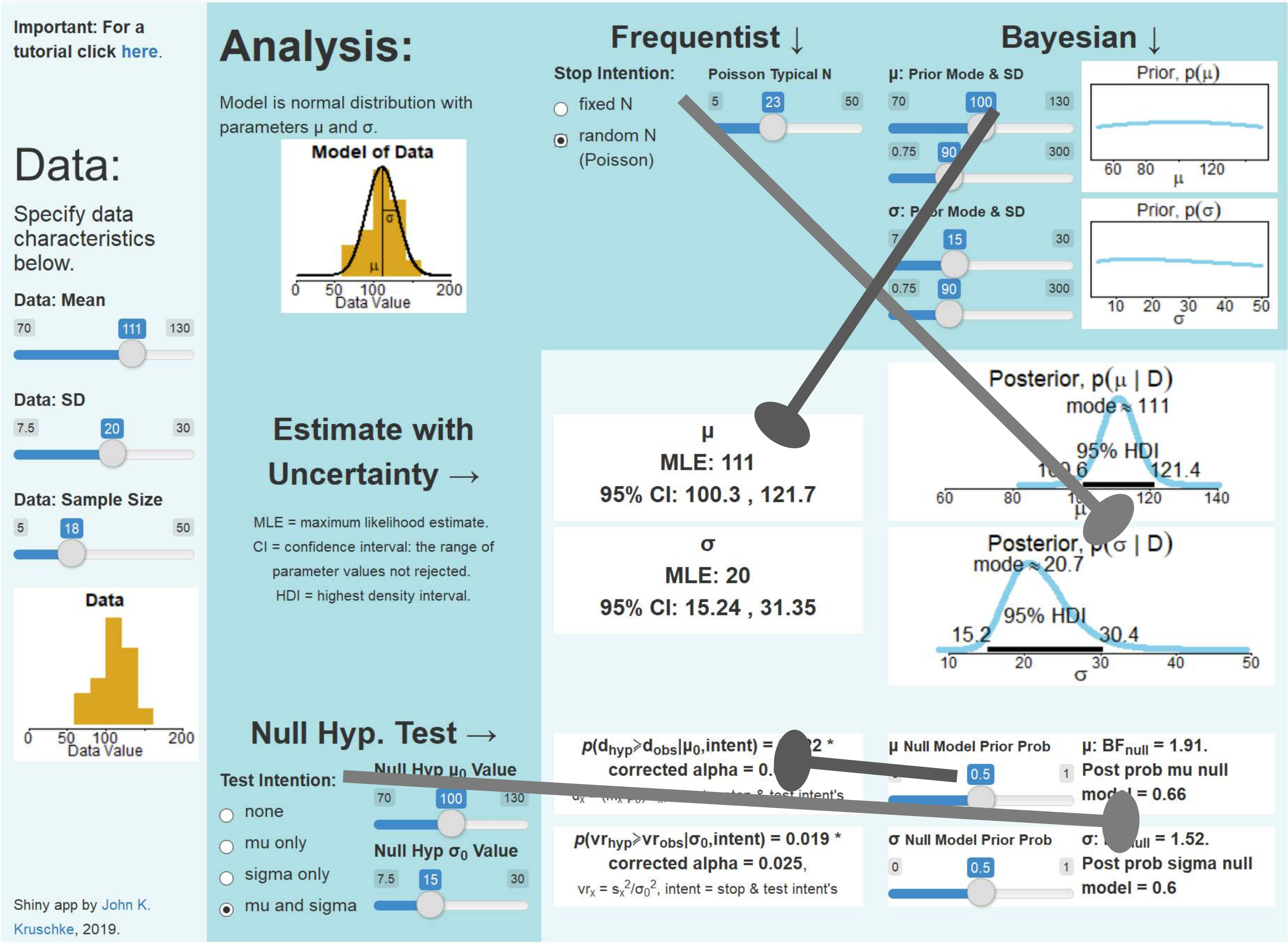
• *Change the testing intentions, and watch the CI's change*. Switch from "mu only" to "mu and sigma", watch the change in CI on μ. Switch from "sigma only" to "mu and sigma", watch the change in CI on σ. Do the CI's get wider or narrower when more tests are intended?

• *Change the stopping intention, and watch the CI's change*. Switch from "fixed N" to "random N (Poisson)". Use different settings of **Poisson Typical N**. The CI's will change accordingly. Do the CI's get wider or narrower when the Typical N is increased?

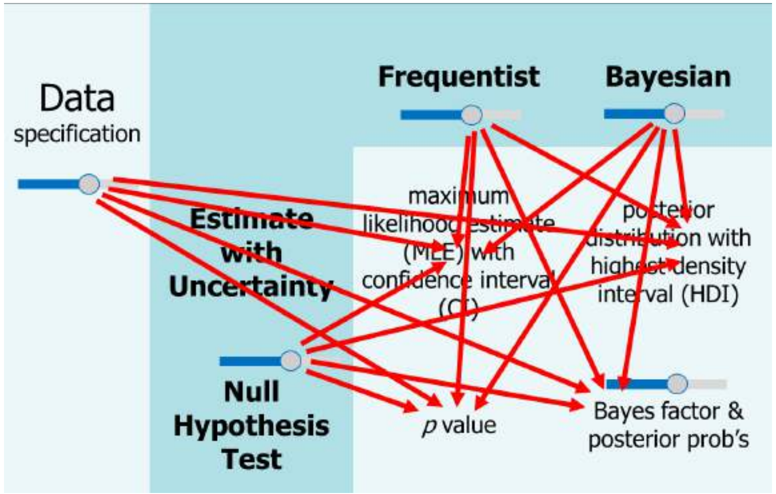
Review: Compare info side by side

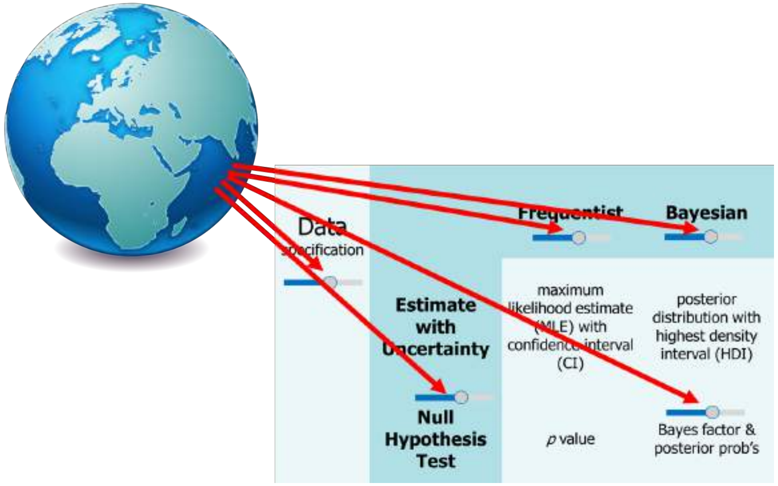


Review: What sliders do *not* affect



Reprise: Learning Outcomes





Be able to predict the qualitative effect of every slider and button on the results in every cell of the table, and explain why.

Be able to set the sliders appropriately to reflect real-world scenarios, and explain why.

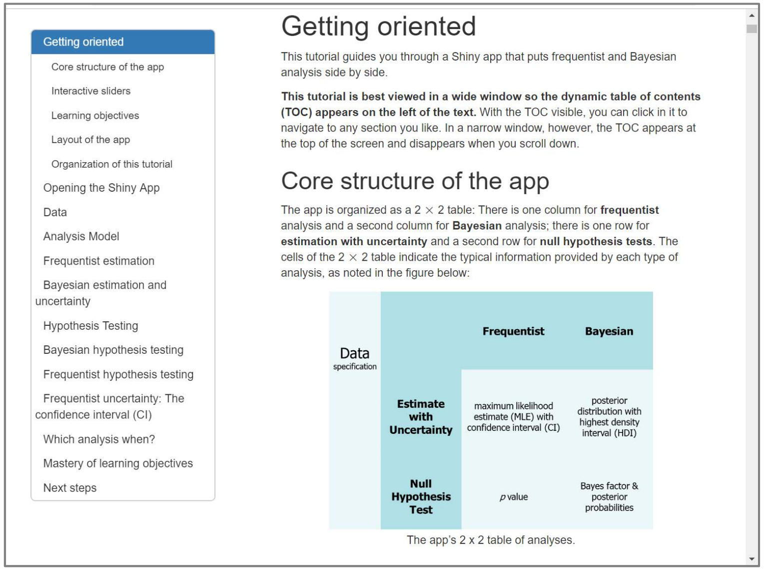
How to Improve the App?

- Guidelines for where and how to insert into existing classes?
- Sub-modules? With videos?
- Specific exercise sets and quiz banks?
- Simpler, one-parameter version?
  (could not show corrections for multiple tests)
- Software that allows inserting user data?
  (would be challenging in Shiny)
- **Discuss at Breakout Session!**

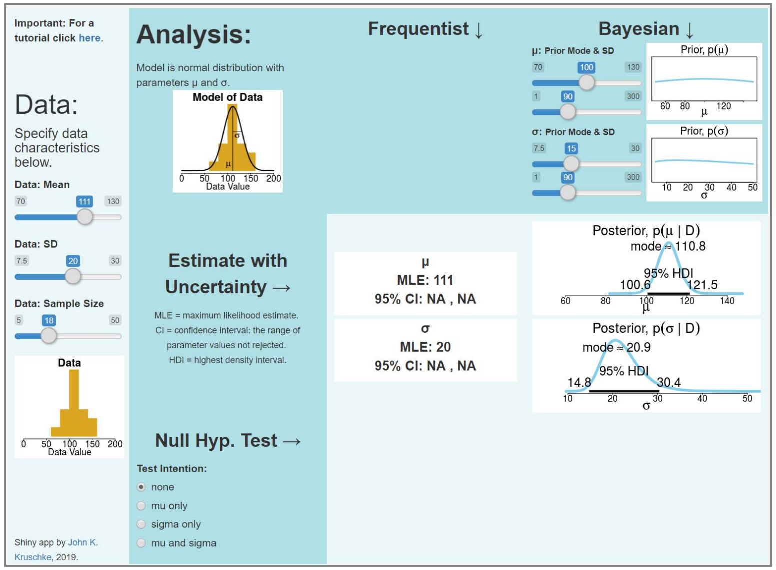
How to Get Teachers to Adopt the App?

- More arguments in favor of Bayesian?
- Demo's of how intuitive and easy Bayesian is?
- Examples of how juxtaposition clarifies both?
- More arguments against frequentist? (people react badly to this)
- Evidence of efficaciousness?
- Endorsement by agencies, societies, leading instructors?
- **Discuss at Breakout Session!**

The Tutorial



The App



http://www.indiana.edu/~kruschke/

Thank you!