Bayesian Perspectives on Probability and Statistics

Leila Peitsch | Erin Lipman

Table of Contents

Brief Overview: Bayes' Rule

$$\text{posterior} = \frac{\text{prior} \cdot \text{likelihood}}{\text{normalizing constant}}$$

Bayesian statistics is built on the method of using **prior** knowledge along with new incoming **data** in order to develop a **posterior** understanding.

It is helpful as we can update our beliefs and it is often more *interpretable* than frequentist statistics.

Bayes' Rule in STAT 311

Most will recognize **Bayes' Rule** in terms of events of A and B.

In introductory classes, Bayesian statistics is often used in examples that involve medical testing (such as asking: given my positive test result, what's the chance that I actually have the disease?)

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Brief Overview: Bayes' Rule

In some cases when we use certain priors and likelihoods called **conjugate priors**, we are able to calculate the normalizing constant easily.

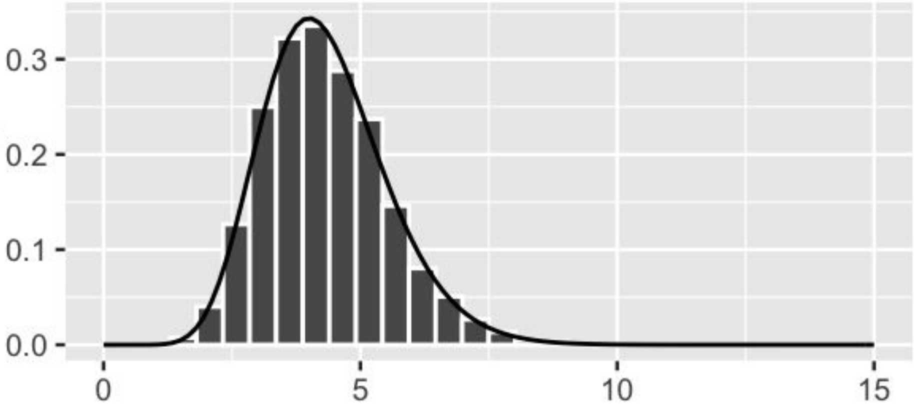
However, calculating the normalizing constant can involve large integrals as we have to consider *all* possible outcomes. Due to this, the normalizing constant is sometimes omitted:

$$\text{posterior} \propto \text{prior} \cdot \text{likelihood}$$

Approximating the Posterior

While conjugate priors are effective within simple models, they aren't as useful if the posterior is seemingly impossible to identify. Such as when the prior and likelihood are comprised of complicated *multivariate* functions.

In this case, instead of specifying the posterior, we instead **approximate** the posterior via simulation!



Markov chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) is a class of algorithms used to sample from the posterior probability distribution.

MCMC simulation produces a chain of dependent samples - these samples are *not* drawn of the posterior pdf. But as long as we have enough samples (*N* is large enough), the simulation will reflect the posterior model.

It is helpful as it is easier to scale up for more complicated Bayesian models, unlike other approximation methods such as grid approximation.

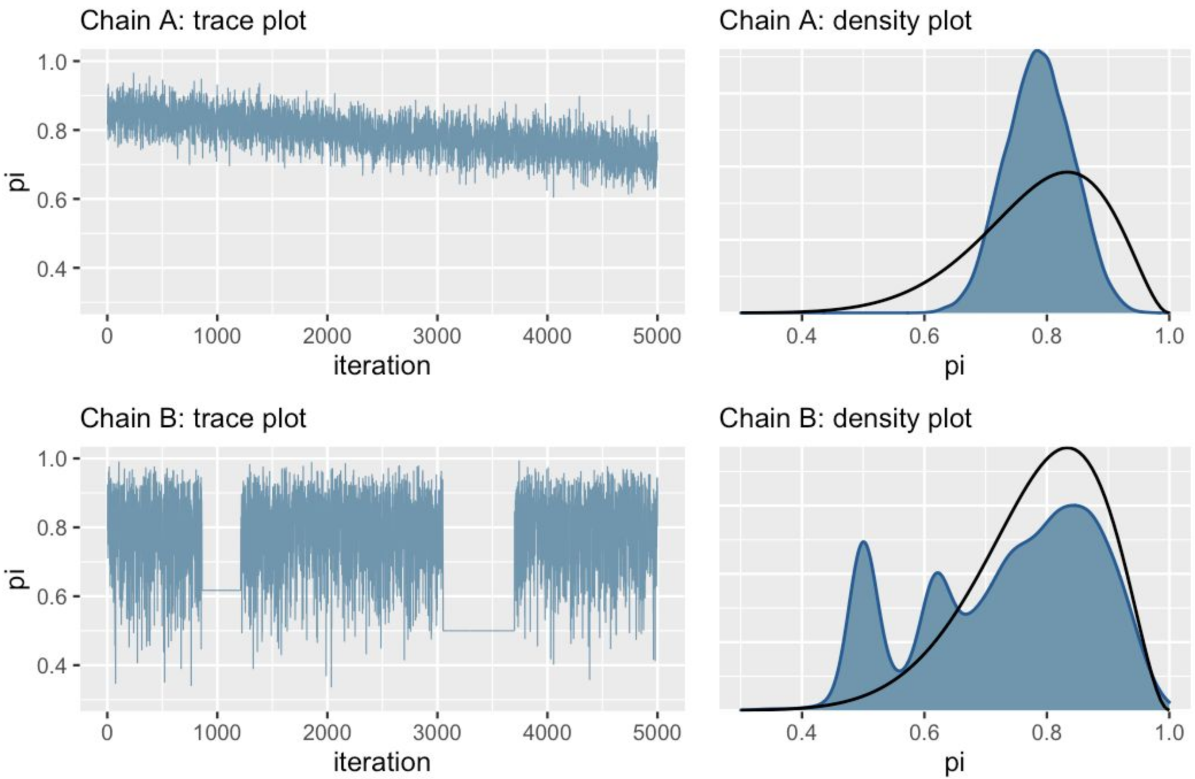
Evaluating MCMC Simulations

Simulations cannot all be perfect - in order to tell if a Markov chain is "good", there are a few checks we put in place.

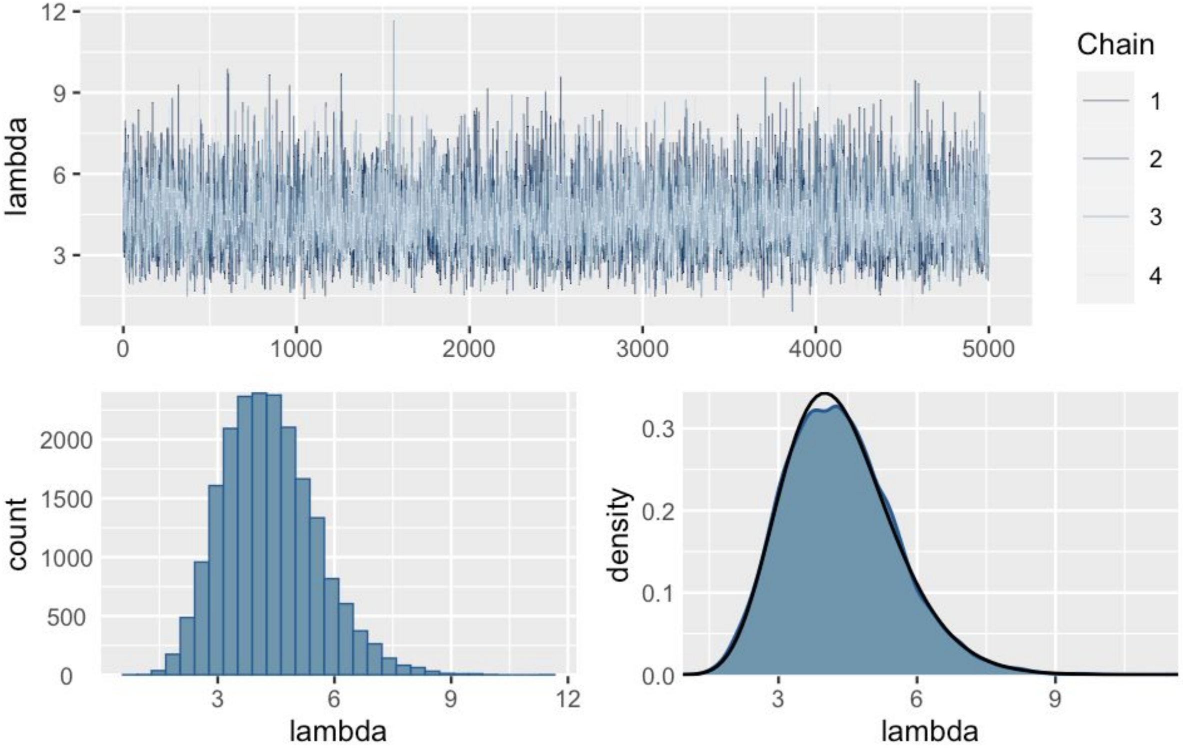
We can utilize trace plots and parallel chain to observe any inconsistencies visually. This can show *fast-mixing* vs *slow-mixing* - fast mixing is good as it exhibits behavior similar to that of an independent sample compared to slow mixing which does not explore the range of possible posterior values.

On a numeric level, we can check effect sample size, autocorrelation, and R-hat.

Examples of “Bad” Trace Plots



Examples of a “Good” Trace Plot



Difference between MCMC Algorithms

- Metropolis-Hastings
  - An algorithm that goes through an iterative step process in order to propose and accept step locations
- Hamiltonian Monte Carlo
  - Employed in the *rstan* package, uses derivatives and allows for an 100% acceptance prob.
- Gibbs sampling
  - Employed in the *rjags* package, uses conditional probs. of different parameters and also allows for 100% acceptance prob.

Metropolis-Hastings Algorithm

Step 1: Propose a new location

Conditioned on the current location μ, draw a location μ’ from a proposal model with pdf q(μ’|μ)

Step 2: Decide whether or not to go there

Calculate the *acceptance probability*

$$A(\mu, \mu') = \min\left(1, \frac{f(\mu')P(y|\mu')}{f(\mu)P(y|\mu)} \frac{q(\mu)}{q(\mu')} \right)$$

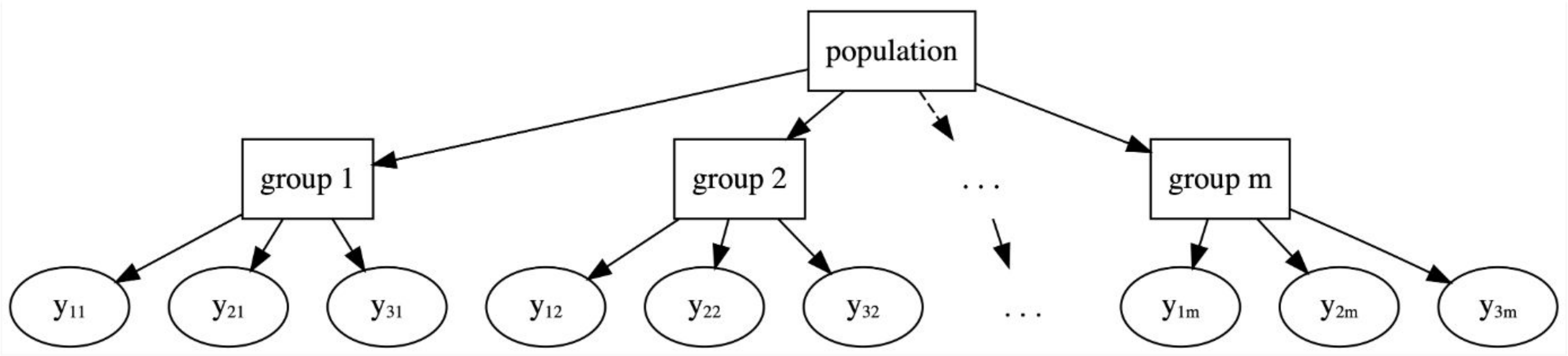
Hierarchical Models

Hierarchical models (or multilevel models) are a way to model complex systems such as cases when individuals are organized at more than one level

There are limitations in models that are *fully pooled* or *not pooled at all*

- Fully pooled models sacrifice independence between samples and can give us misleading conclusions
- Models with no pooling doesn’t allow us to observe within-group variability or between-group variability

Hierarchical Models



A **partial pooled model** (as we use in our applied analysis) allows us to see what we can be inferred with distinct subgroups as well as how different subgroups *inform one another* - a big advantage of Bayesian inference!

Applied Analysis: Mental Health Medication in the US

We looked at a dataset from Kaggle that reports on mental health following the COVID-19 pandemic within the US.

There were 4 indicators - we chose to look at the percentage of individuals who reported taking medication for mental health in the past 4 weeks.

	subgroup	time_period	time_period_label	value	low_ci	high_ci	sd
1	Alabama	13	Aug 19 – Aug 31	22.3	19.6	25.1	1.4030870
2	Alaska	13	Aug 19 – Aug 31	16.7	14.2	19.5	1.3520657
3	Arizona	13	Aug 19 – Aug 31	18.3	16.1	20.7	1.1734910
4	Arkansas	13	Aug 19 – Aug 31	24.1	20.8	27.7	1.7602364

$$y_{it} \sim N(\mu_i, S_{it}^2)$$

is the likelihood

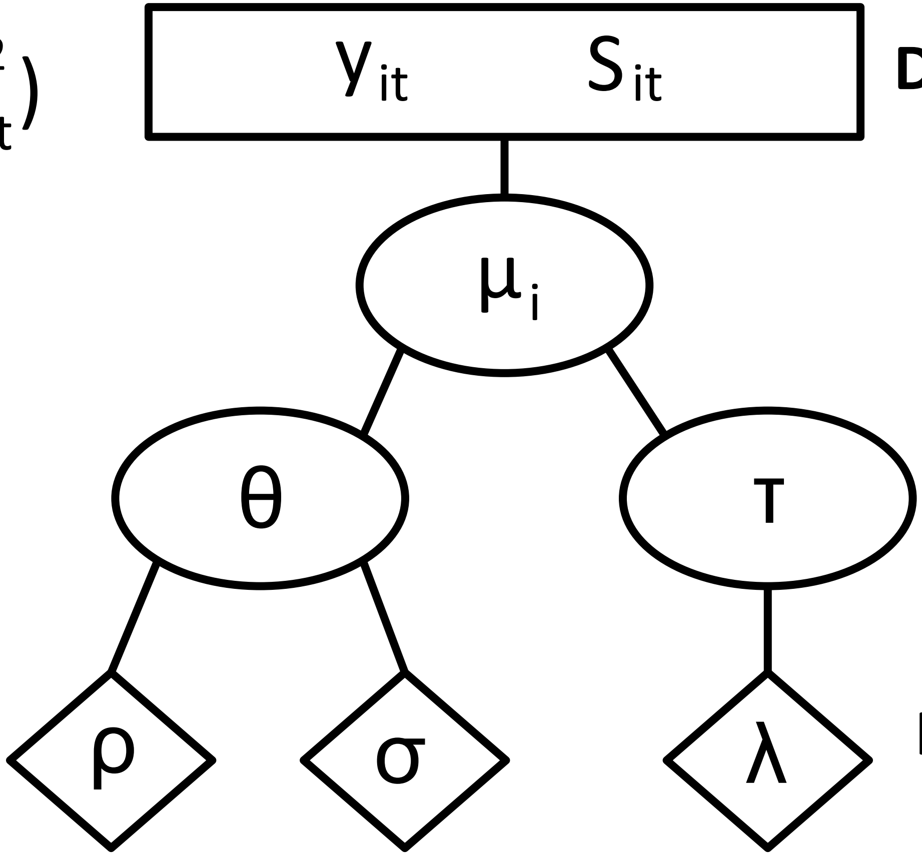
$$\mu_i \sim N(\theta, \tau^2)$$

is the prior

$$\theta \sim N(\rho, \sigma^2)$$

$$\tau \sim \exp(\lambda)$$

are hyperpriors



Data (from sample)

Parameters
(population)

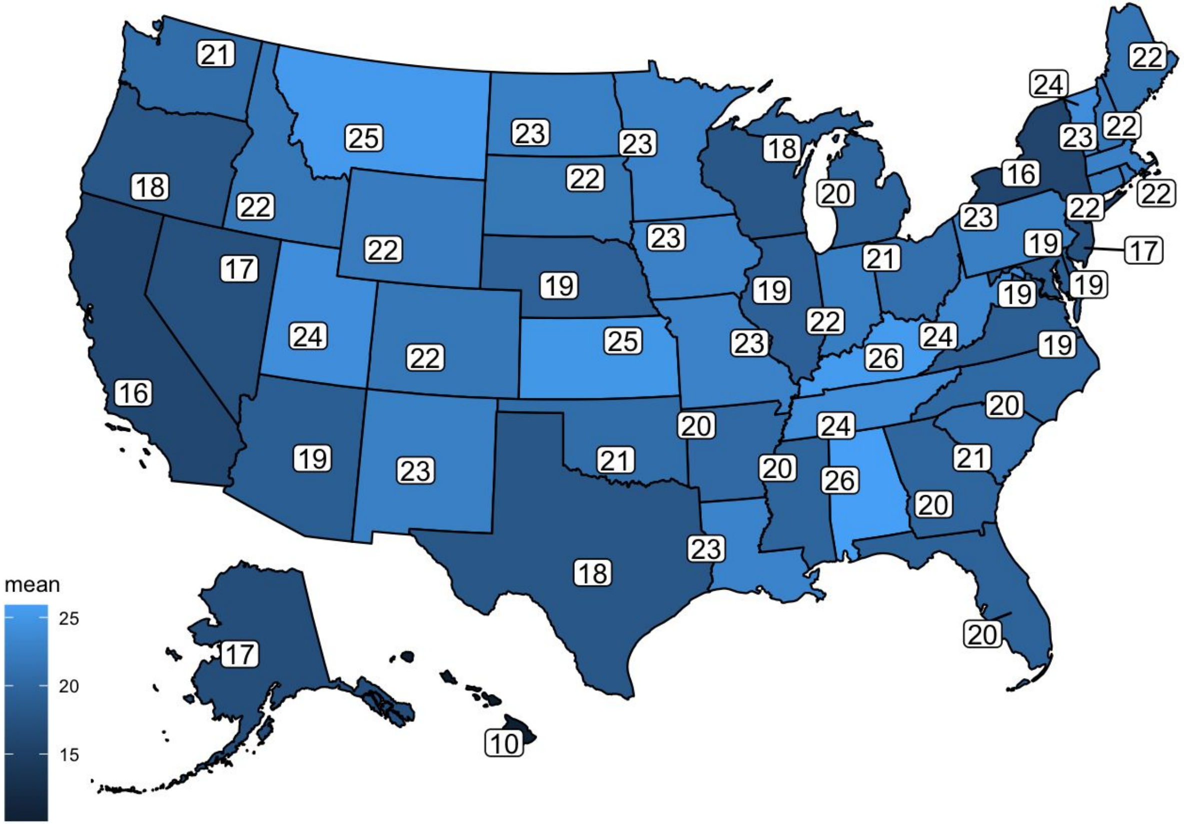
Hyperparameters

Our Model in R

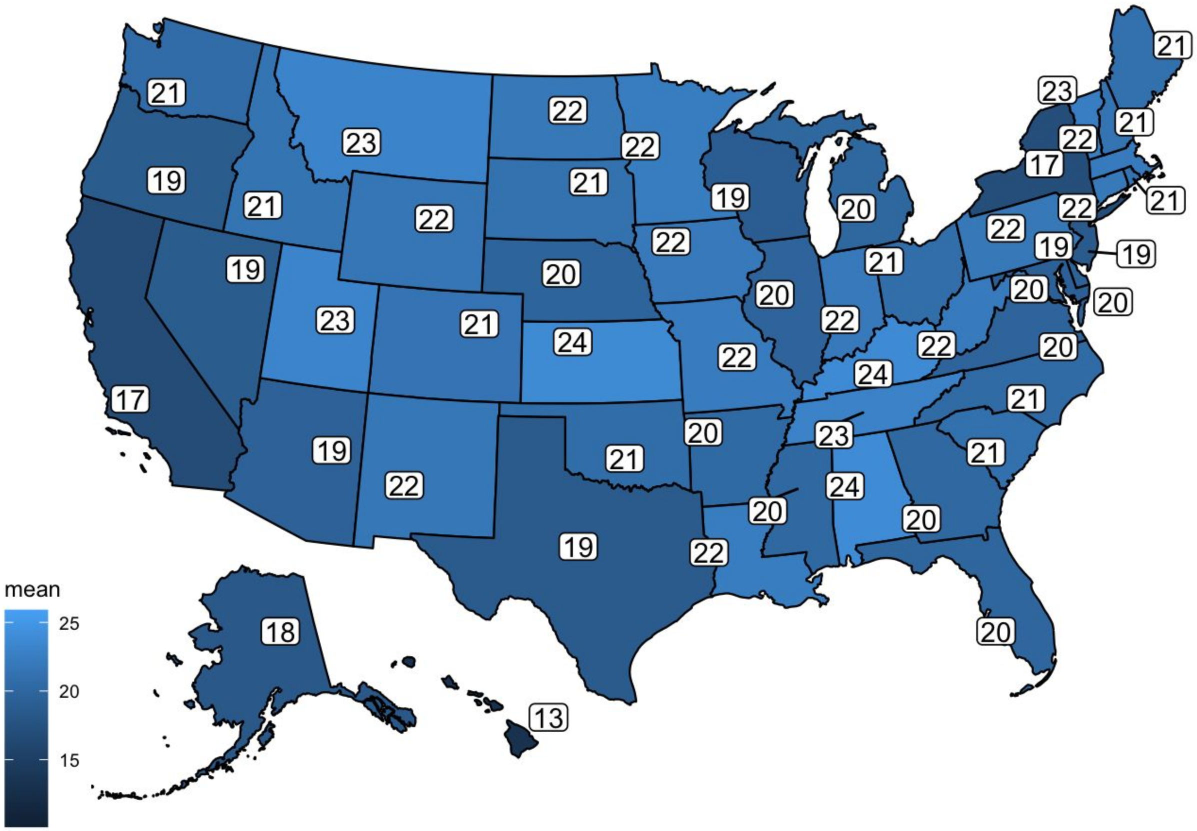
We implemented our model using the *rstan* package - this stan file encompasses the partial pooled hierarchical model pictured previously.

```
data {
  int<lower=0> N;              # Number of states
  int<lower=0> M;              # Number of time periods
  matrix[N, M] y;              # Outcome
  matrix[N, M] s;              # Outcome sd
  real prior_mean;             # prior mean for mu
  real<lower=0> prior_sd;      # prior sd for mu
}

parameters {
  vector[N] mu_state;
  real mu_overall;
  real<lower=0> sigma_overall;
}

model {
  for (i in 1:N){ # likelihood
    for (t in 1:M){
      y[i,t] ~ normal(mu_state[i], s[i,t]);
    }
    mu_state[i] ~ normal(mu_overall,sigma_overall); # prior
  }

  mu_overall~normal(prior_mean, prior_sd);
  sigma_overall~exponential(prior_sd);
}
```

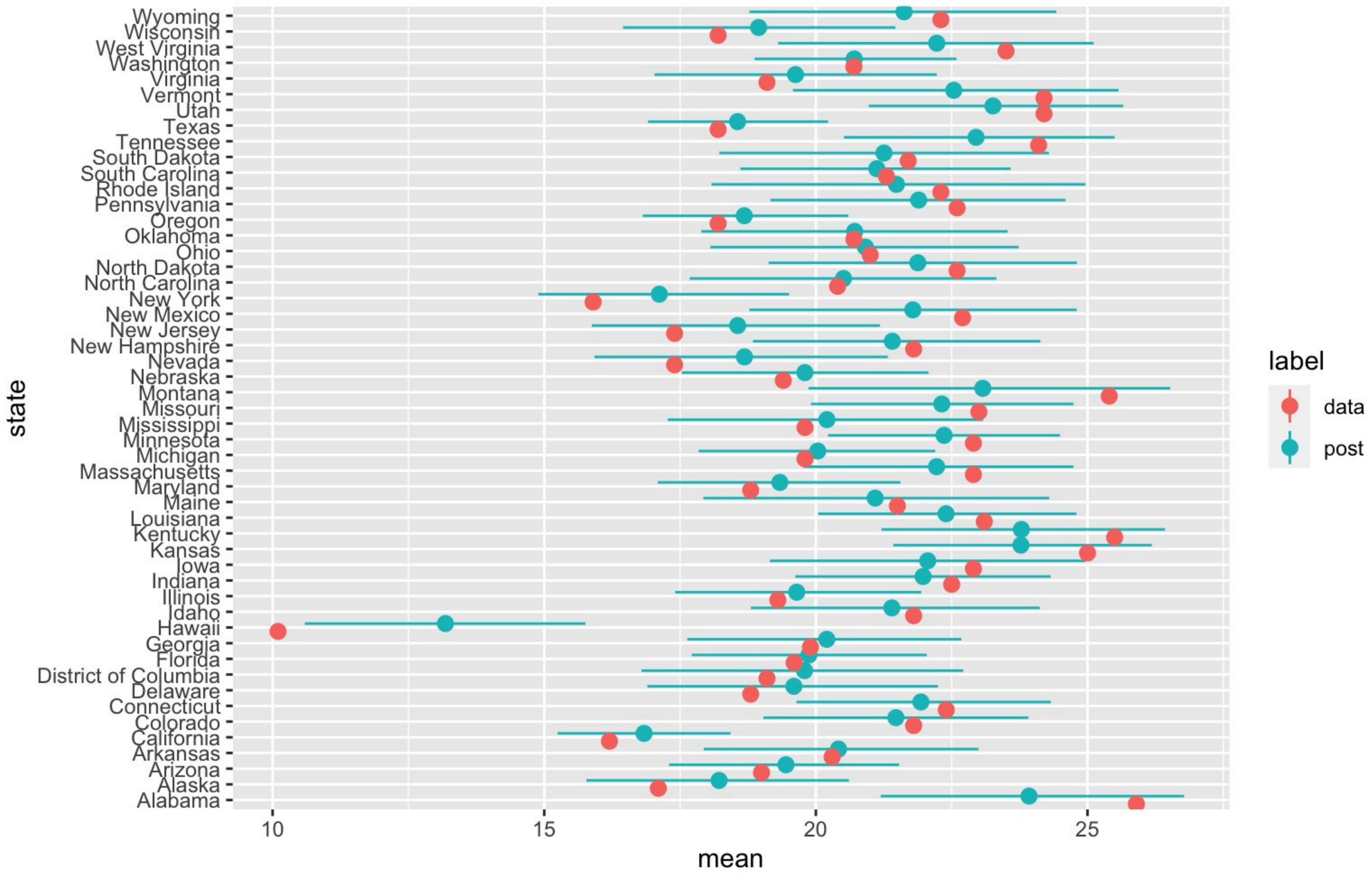
Unadjusted State Means (Sep 16 - Sep 28, 2020)



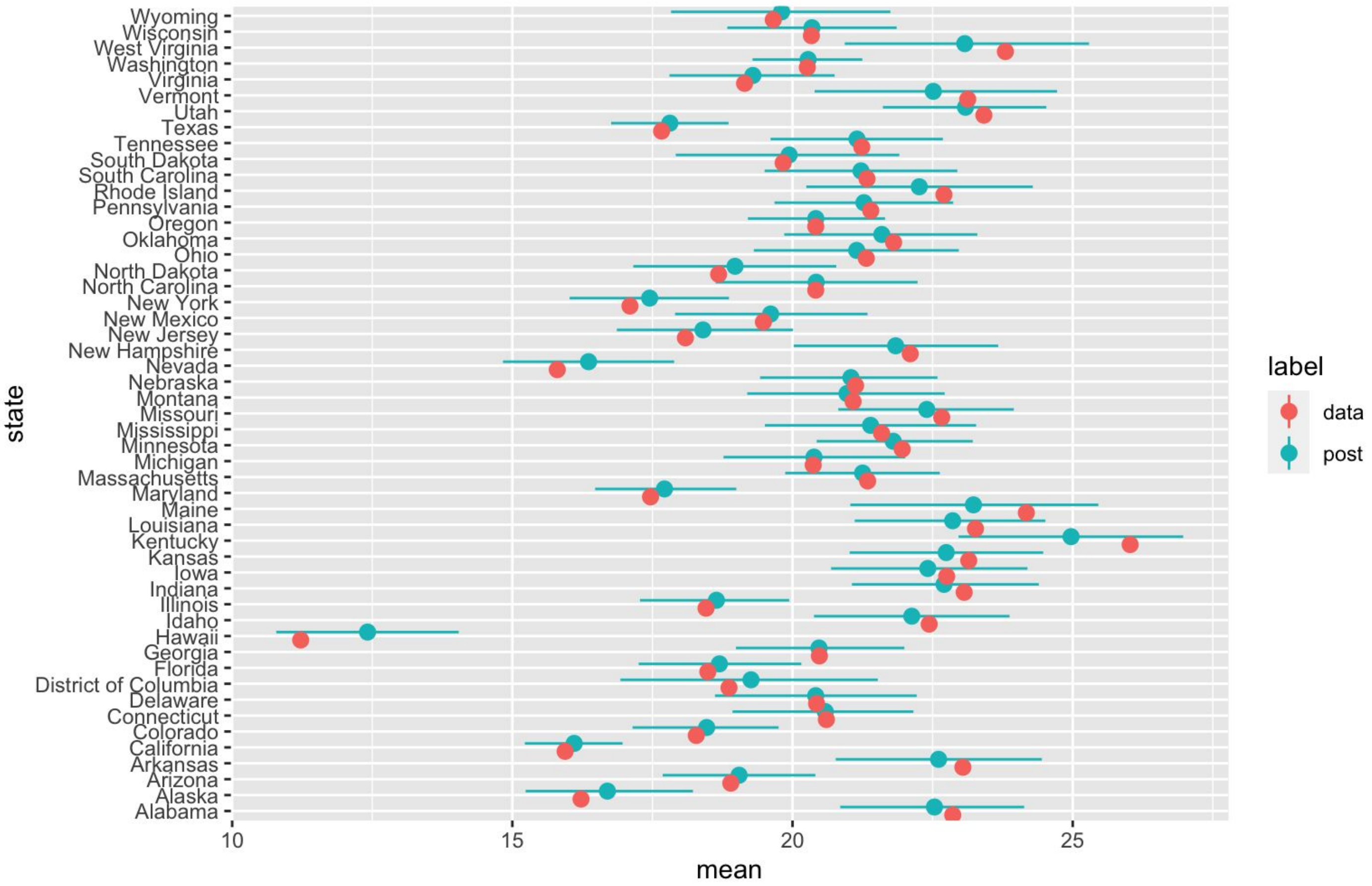
Adjusted State Means (Sep 16 - Sep 28, 2020)



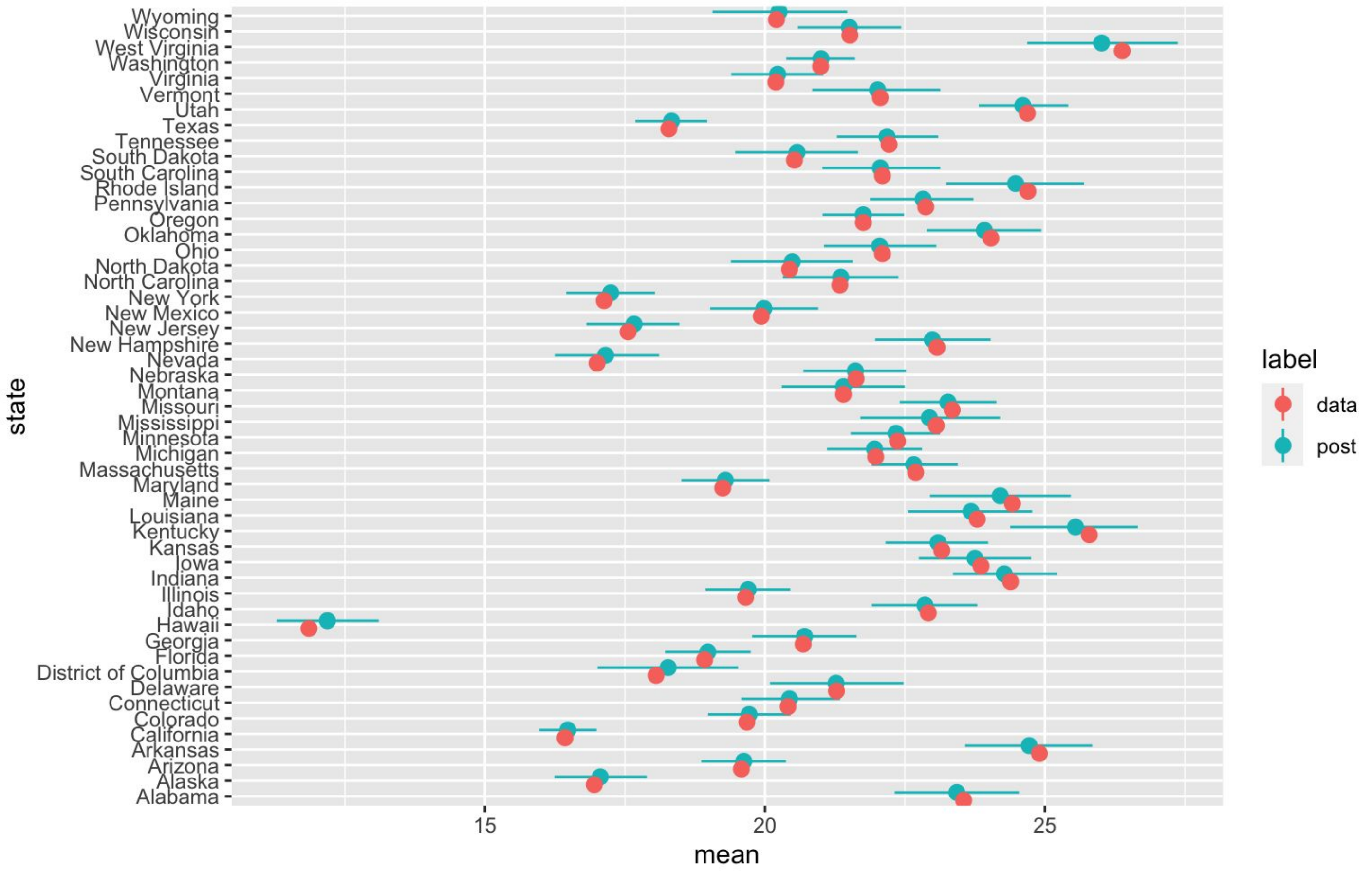
Unadjusted & Adjusted State Means (Sep 16 - Sep 28, 2020)



Unadjusted & Adjusted State Means (Aug 19 - Sep 28, 2020)



Unadjusted & Adjusted State Means (Aug 19, 2020 - Mar 1, 2021)



Thank you for listening!

Any questions or comments?

Credits

- Dataset from Kaggle → U.S. Household Mental Health and COVID-19
  - Data comes from *US Open Data Portal*
- Johnson, Alicia A., et al. *Bayes Rules! An Introduction to Applied Bayesian Modeling* (2021)