Bayesian Statistics and Modeling
Part II

Time series modeling with `pymc`

Based on this example

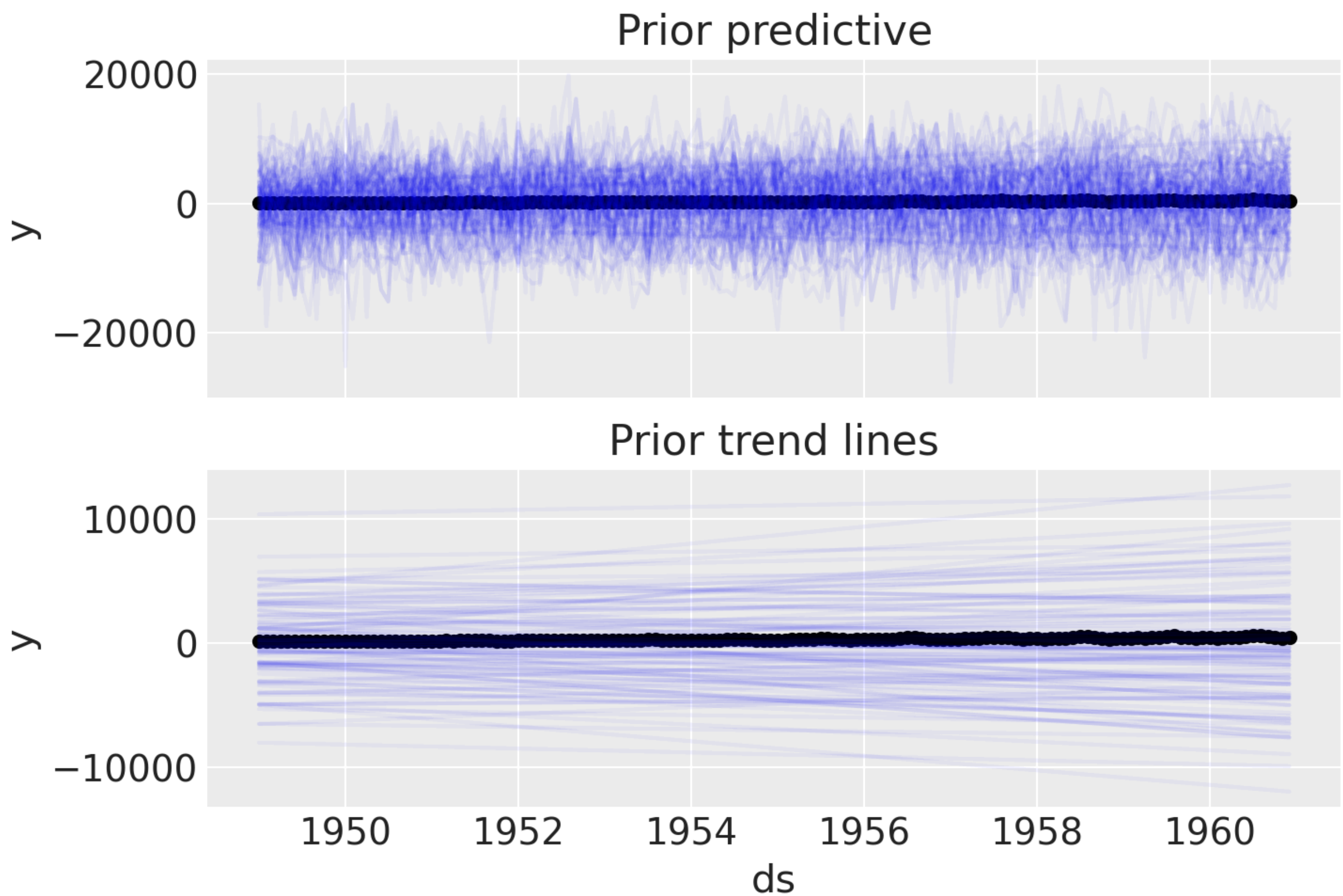
Modeling airline trends

$$Passengers \sim \alpha + \beta \cdot t$$

Just a simple linear trend for now. α is an intercept term, and β is our slope term

Let's go to the code here to start building our model

Prior predictions (WHAT??)



Prior predictions (WHAT??)

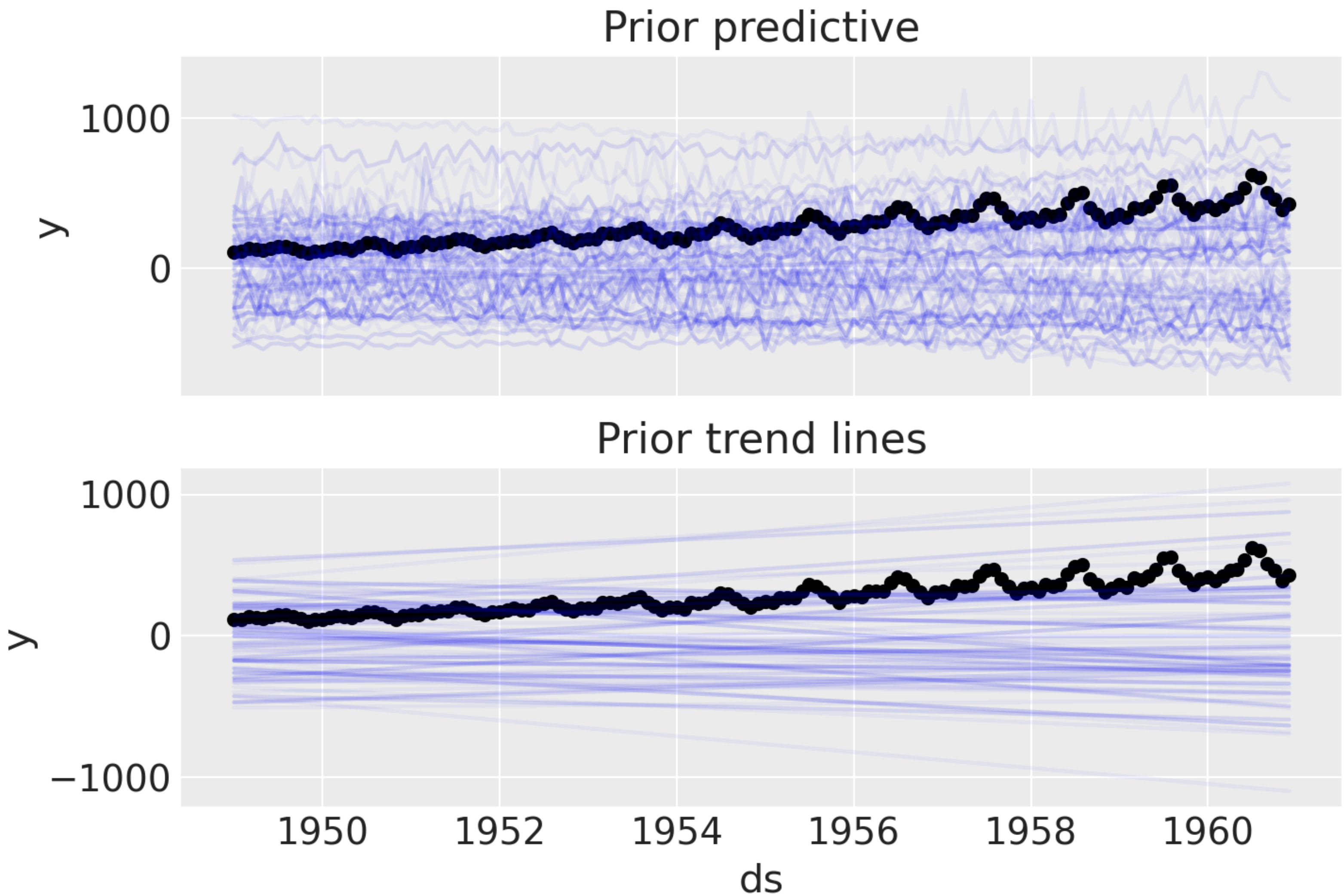
- Look at a large array of possible "reasonable" outcomes given our assumptions about the data
- Gives us an idea of whether our priors make sense
- In this case, we want to make some corrections

Note on Prior Selection

Why choose better priors?

Part of the challenge with a Bayesian model is compuational. Our prior is intended to help us search the parameter space effectively in order to find truth. If we cast too wide a net, we slow things down (they're alread not fast), and we make it harder to find truth.

Updated priors



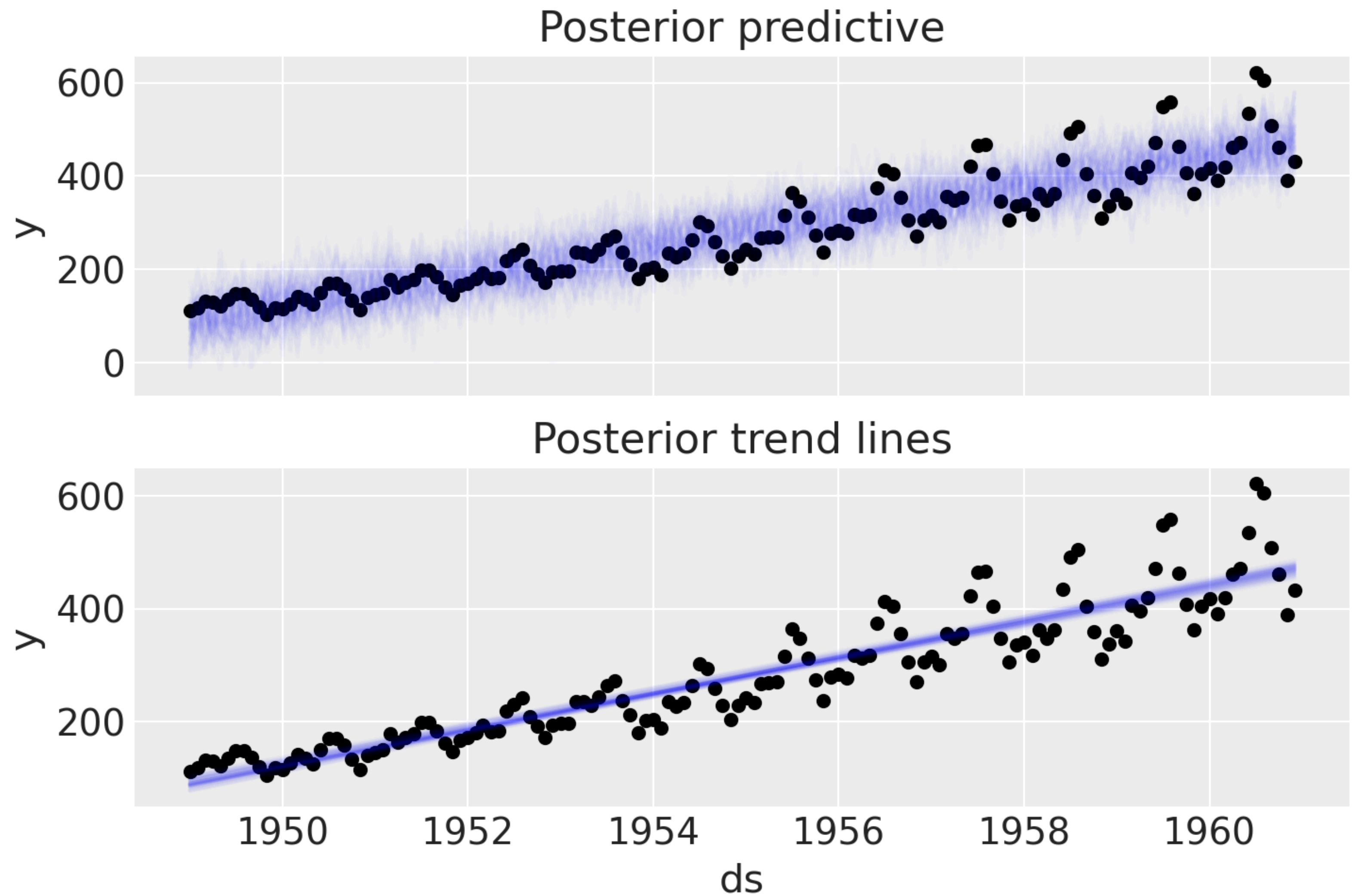
One question, though...

Why do we have negative slopes and intercepts in our prior space?

Posterior predictions

- Incorporate our actual data and then compare our model to observed outcomes
- Decide if we think that our model can make reasonable predictions

Posterior predictions



Adding seasonality

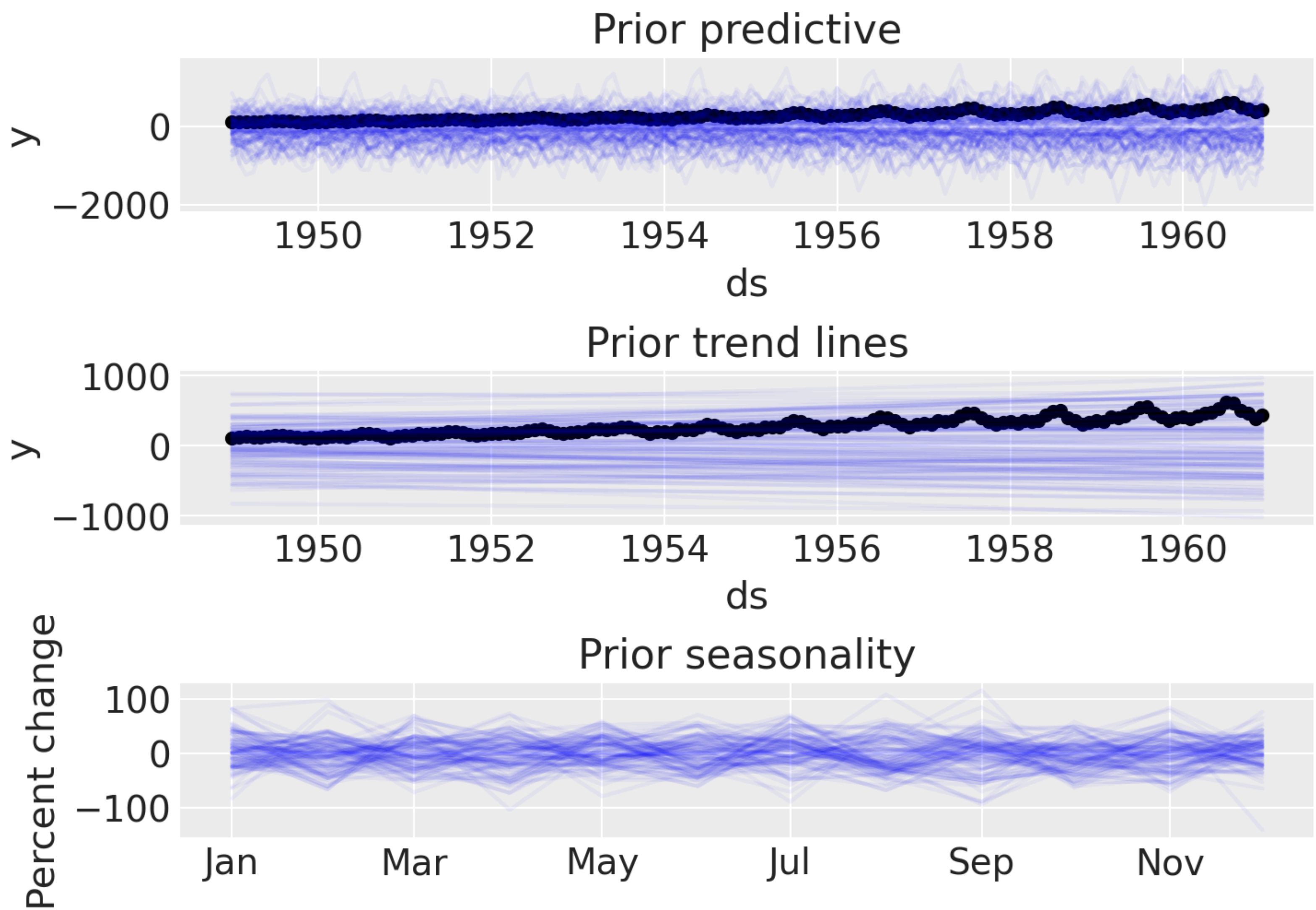
We add a group of periodic functions (fourier features) to function as our "seasonality splines" (if we think of our model as a GAM). They will get stretched or weighted based on observations.

Seasonality (multiplicative)

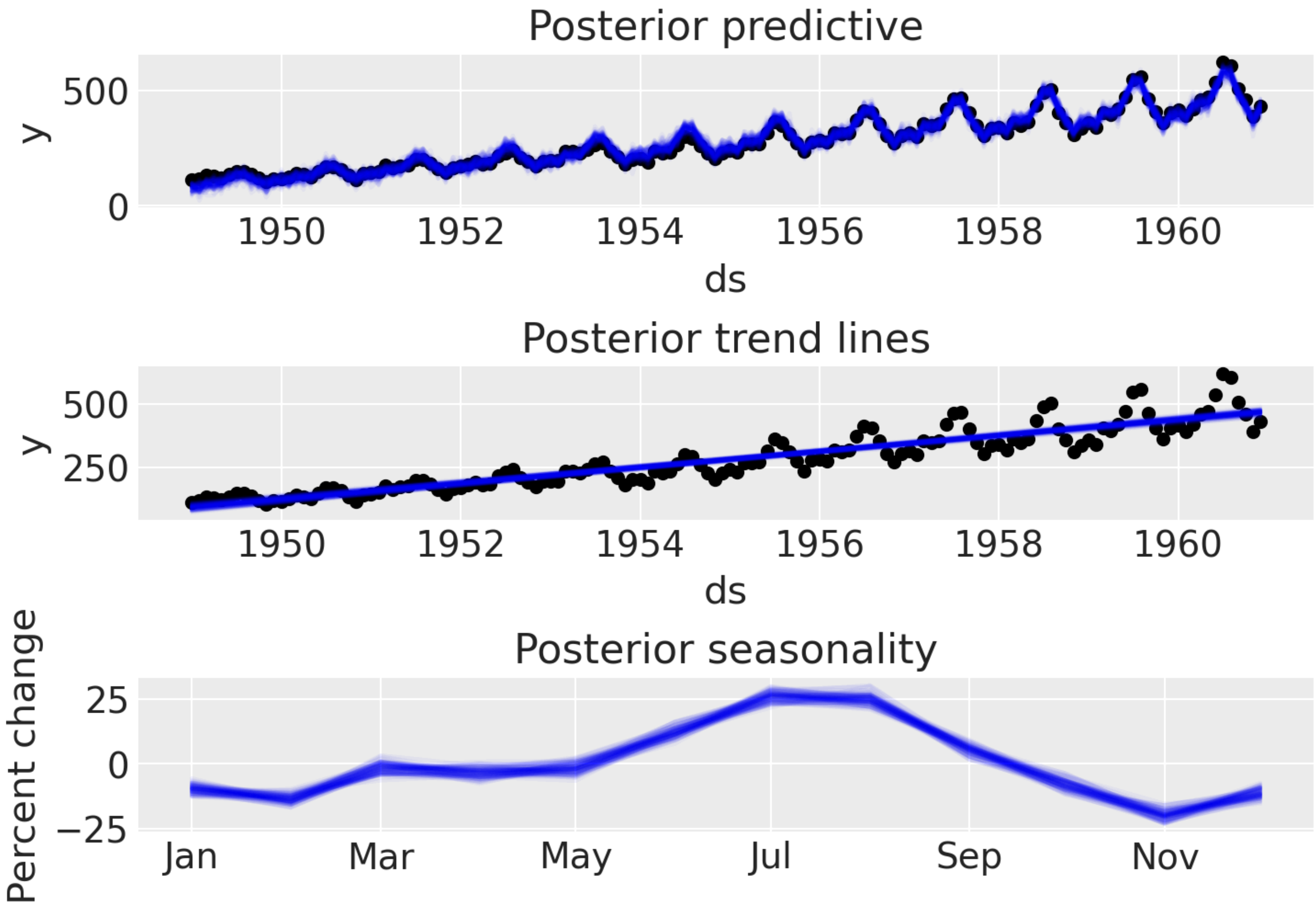
$$Passengers \sim (\alpha + \beta \cdot t) \cdot (1 + seasonality)$$

Our seasonal terms interact with each term in our original model to increase/decrease the expected number of passengers

Seasonal priors



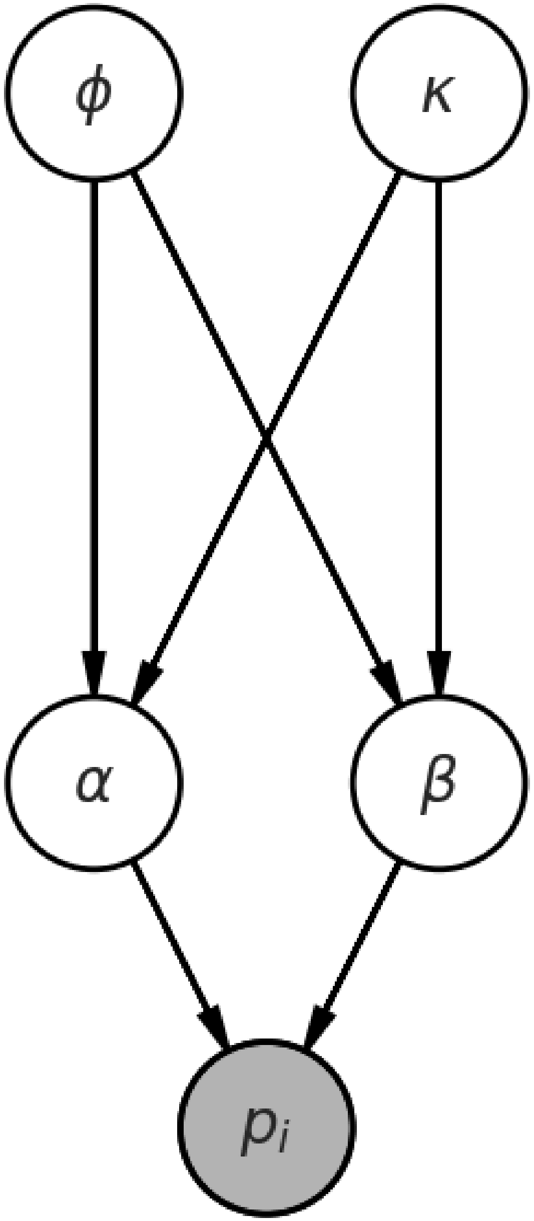
Seasonal posteriors



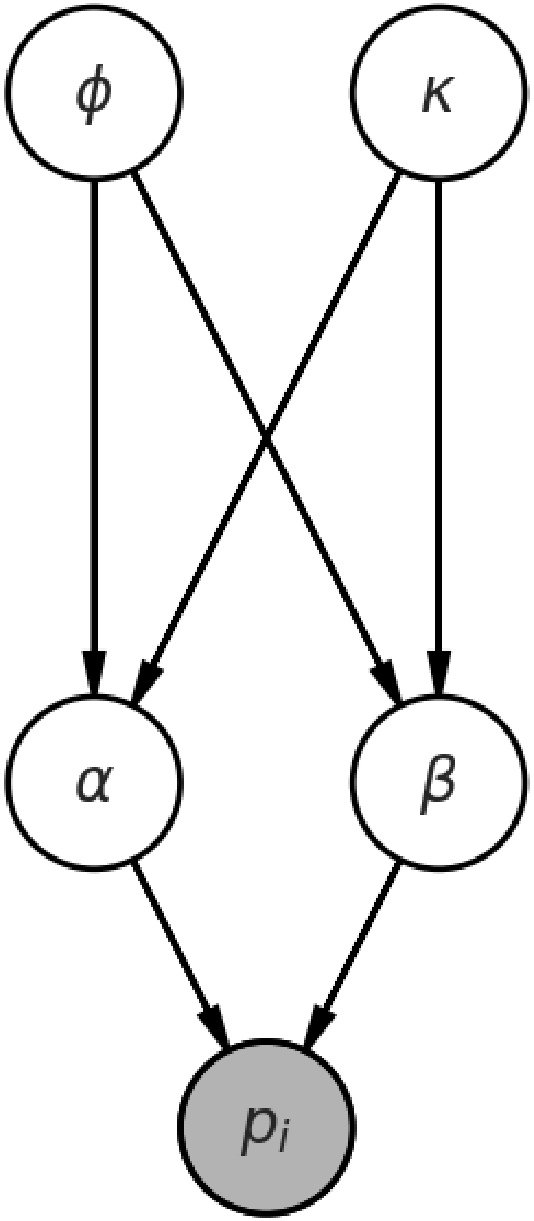
Modeling baseball outcomes

A revised/updated version of this tutorial

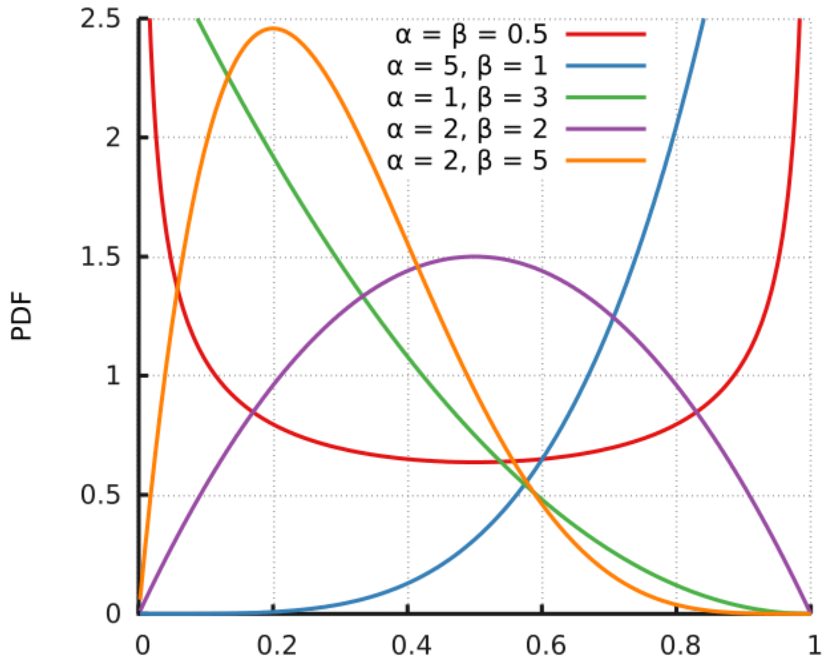
Follow along with the tweaked code here



- ϕ (phi) - Our population-level expectation of batting average
- κ (kappa) - Population variance in batting average
- α, β - Parameters of our beta distribution
- p_i - Individual batting average



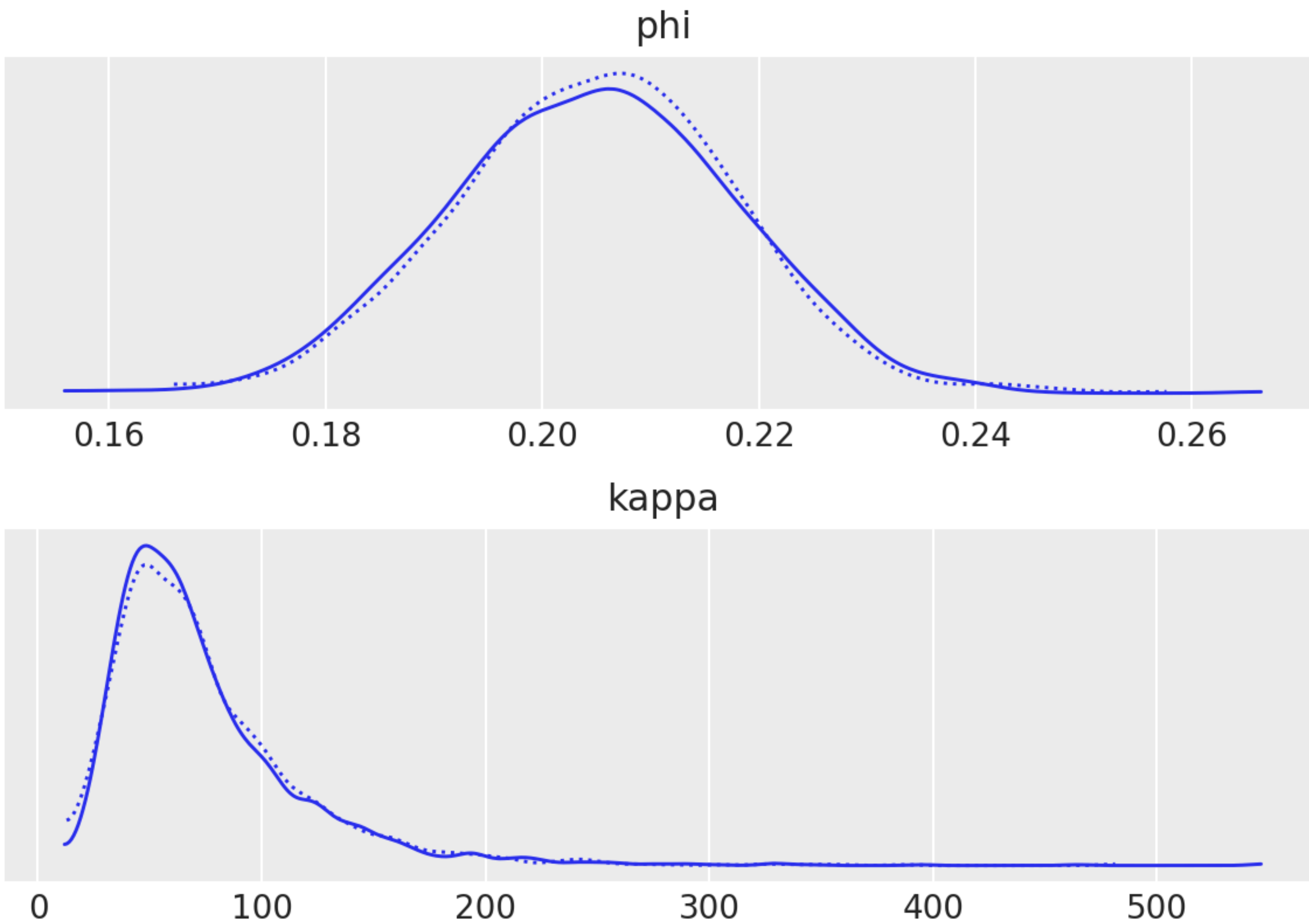
$$\alpha = \phi \cdot \kappa$$
$$\beta = (1 - \phi) \cdot \kappa$$



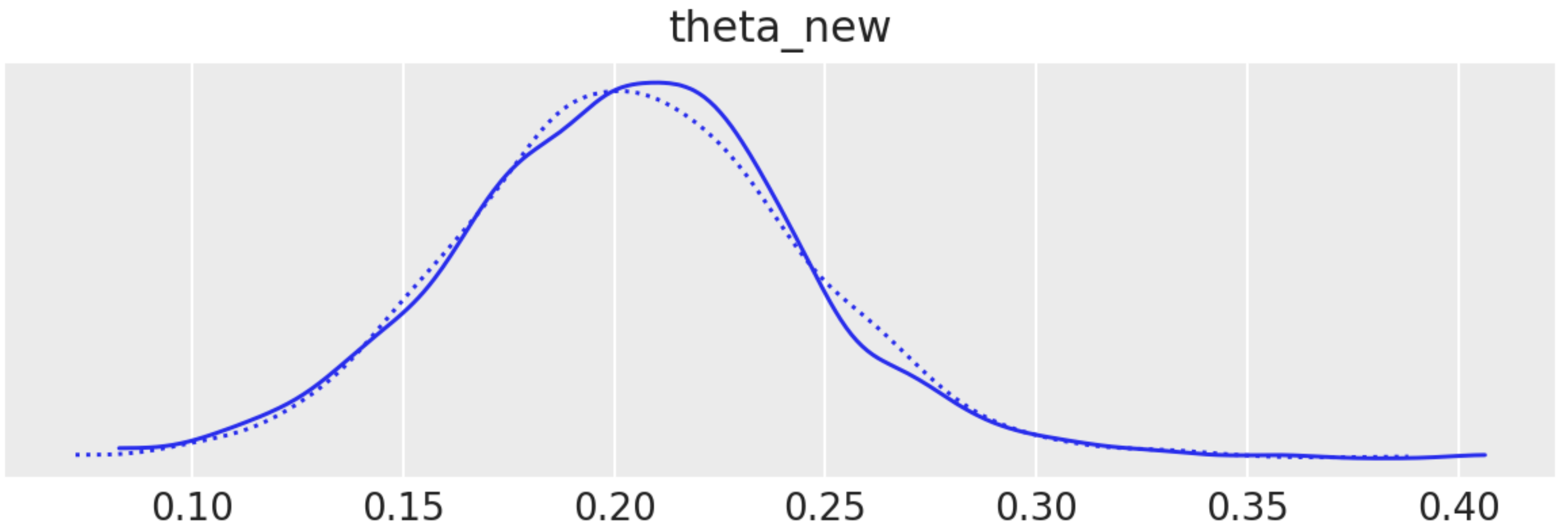
Beta distribution

- Used where there are binary outcomes (hit or no hit)
- Tilts toward 1 or 0 based on observed outcomes and concentration of those outcomes

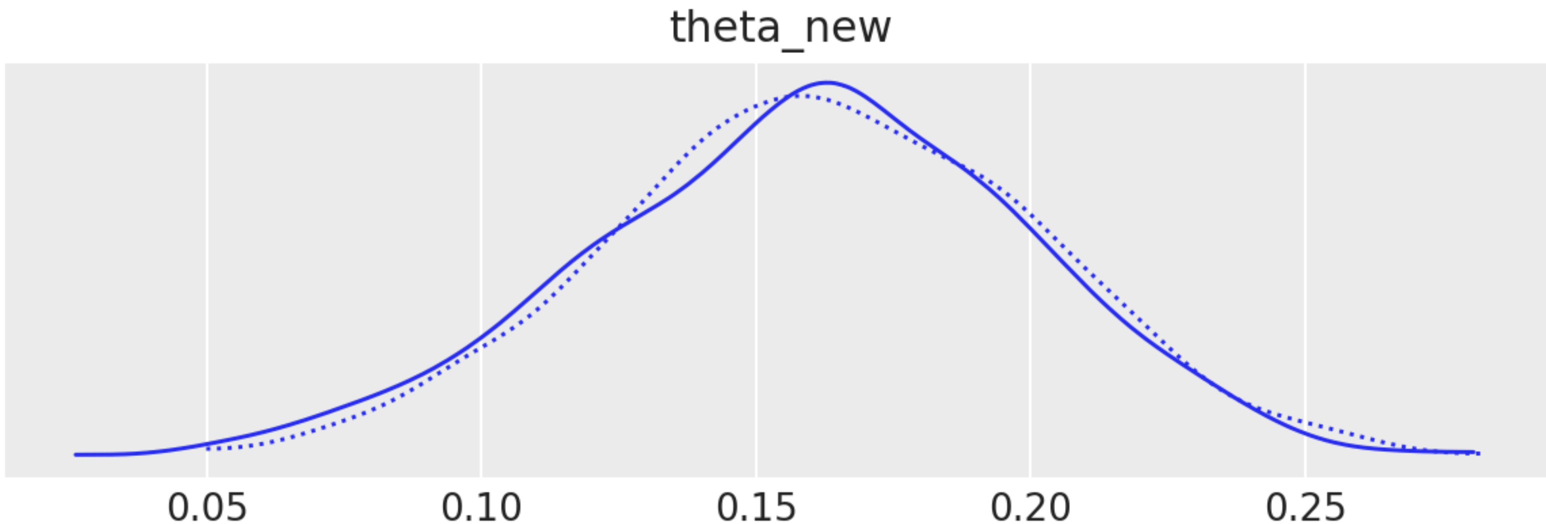
Population values



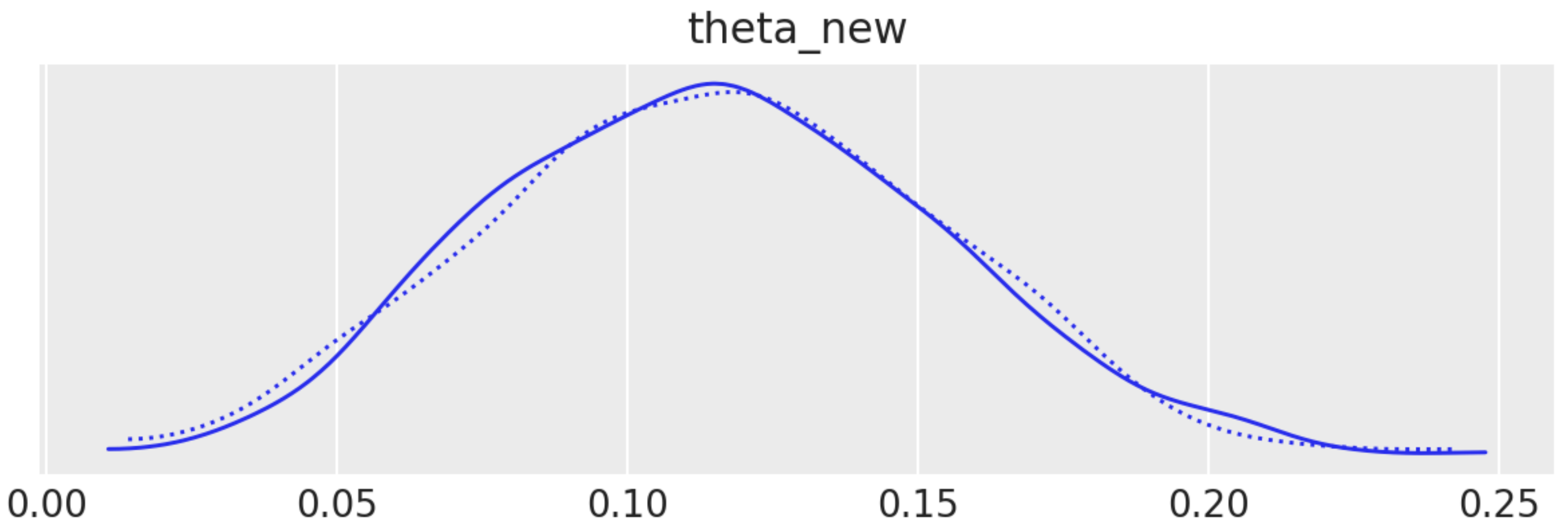
Player with 4 at-bats, no hits



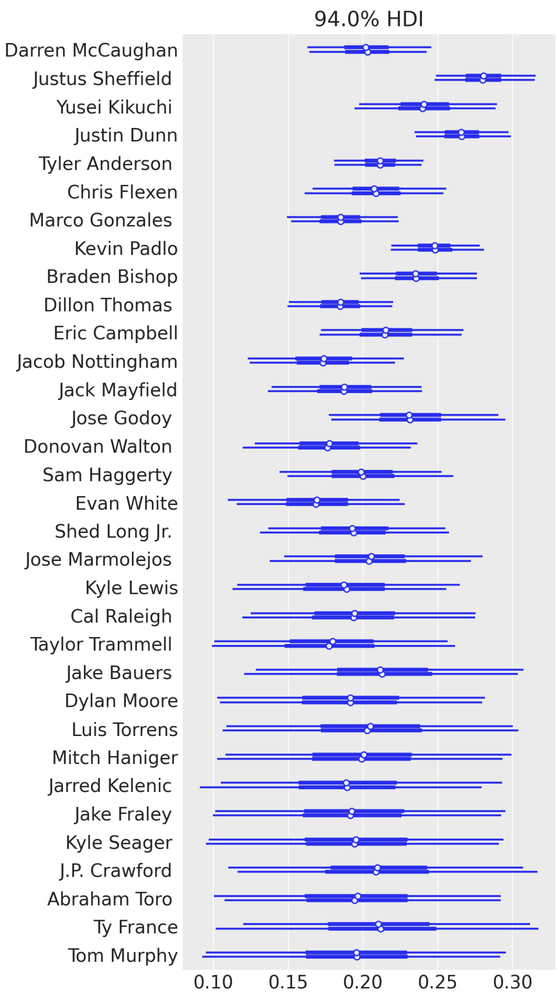
Player with 25 at-bats, no hits



Player with 50 at-bats, no hits



Mariners 2021



Data Storytelling

> Probabilistic programming will unlock narrative explanations of data, one of the holy grails of business analytics and the unsung hero of scientific persuasion. People think in terms of stories - thus the unreasonable power of the anecdote to drive decision-making, well-founded or not. But existing analytics largely fails to provide this kind of story; instead, numbers seemingly appear out of thin air, with little of the causal context that humans prefer when weighing their options.

-- B. Cronin (full article)