STATISTICAL RETHINKING 2026
HOMEWORK B01

These problems use data from a 2019 publication that reports date and time of submission of papers to the medical journal *The BMJ* (doi:10.1136/bmj.l6460). I have provided these data on the course github and on the Discord (homework channel) as `BMJSubmissions.csv`. There are 6 columns and 49,464 rows. Each row is an article submission. The columns are:

`T`: The local time of the submission
`Y`: The year of the submission
`country_name`: Country of corresponding author
`L`: ID value for country (an index variable from 1 to 38)
`W`: 0-1 indicator variable for weekend submission
`H`: 0-1 indicator variable for holiday submission

1. For each country in the sample, estimate the probability that a corresponding author submits on a weekend. In this problem, use a no-pooling (no shrinkage) model. This means the model does not have hierarchical priors. Which countries have the highest probability of submitting on the weekend? Can you visualize these estimates in a way that also shows their uncertainties?

2. Now use a partial-pooling model. This means that the model has a hierarchical prior for each country's estimate. Hint: Think of each country like a tadpole tank. How are the partial-pooling estimates different from the no-pooling estimates? Can you figure out why they differ?