Individual HW04

Your Name Here

```
library(bayesrules) # R package for our textbook
library(tidyverse) # Collection of packages for tidying and plotting data
library(janitor) # Helper functions like tidy and tabyl
library(rstan) # for MCMC
library(bayesplot) # for plotting
```

1 Q1

Explain the difference between N_{eff} and the actual number of MCMC samples

2 Q2: `mcmc_rank_hist` and `mcmc_rank_overlay`

Trace plots are a natural way to view a chain, but they become hard to read when we have many samples and/or many chains. An alternative way to view the chains is called a **rank histogram** or **trace rank plot**. What this means is to (a) take all of the samples for a parameter and rank them, (b) draw a histogram of the ranks for each individual chain

The code chunk below loads two pre-fit stan models.

```
# TBD ONCE WEBSITE UPDATED
```

    (a) Create traceplots, rank histograms (`mcmc_rank_hist`), and trace rank plots (`mcmc_rank_overlay`) for both MCMC samples.
    (b) Which sample is "healthy"? Which is "unhealthy"? Explain how you can tell from the new plots
    (c) Which type of plot do **you** find more useful?

3 Revisiting BR Exercise 5.11

Prof. Abebe and Prof. Morales both recently finished their PhDs and are teaching their first statistics classes at Bayesian University. Their colleagues told them that the average final exam score across all students, μ, varies Normally from year to year with a mean of 80 points and a standard deviation of 4. Further, individual students' scores Y vary Normally around μ with an *unknown* standard deviation σ.

(a) Suggest three possible prior distributions for σ (at least two should be different named distributions). Include density plots for each prior. (You can find available probability distributions in the Stan documentation: https://mc-stan.org/docs/functions-reference/positive_continuous_distributions.html)

(b) Below is starter code for running a Stan model in this scenario. Delete the #| eval: false line and fill in the blanks

```
fn_model2 <- "
  data {
    int<lower=0> N;
    vector[N] y;
  }

  parameters {
    real mu;
    real<lower=0> sigma;
  }

  model {
    y ~ normal(_____, ____);
    mu ~ normal(_____, _____);
    sigma ~ _____;
  }
"
```

(c) Run your model on the data below from Prof. Abebe and Prof. Morales' combined scores.

```
scores<- c(90.3, 81.3,  85.2, 85.6, 86.4, 83.8, 86.7, 90.1, 82.3, 88.4, 89.2, 90.6,81.2, 85.9, 8
```

(d) Run thorough diagnostics on your MCMC sample. This should include: traceplots of all relevant parameters, density plots for each chain for all relevant parameters, R-hat values, and N-eff. What do you conclude about the validity of your sample?

(e) What do you conclude about the posterior mean $\mu|\vec{y}$? How does your answer differ from HW3's conjugate analysis?

(f) Rewrite the stan model code to take a flexible normal prior for μ. That is, we should be able to specify the prior mean θ and prior sd τ, instead of hard-coding them as 80 and 4.