Group Work 09

```r
library(bayesrules) # R package for our textbook
library(tidyverse) # Collection of packages for tidying and plotting data
library(janitor) # Helper functions like tabyl
library(rstan) # for fitting models
library(rstanarm) # for fitting standard regression models
library(broom.mixed) # for tidy() function
library(bayesplot) # helpful plotting functions
library(tidybayes) # helpful for wrangling Bayesian model output
library(patchwork)
```

1 Exercise 15.3 (Hierarchical data)

Interested in the impact of sleep deprivation on reaction time, Belenky et al. (2003) enlisted 18 subjects in a study. The subjects got a regular night's sleep on "day 0" of the study, and were then restricted to 3 hours of sleep per night for the next 9 days. Each day, researchers recorded the subjects' reaction times (in ms) on a series of tests. The results are provided in the `sleepstudy` dataset in the {lme4} package.

This data is hierarchical. Draw a diagram in the spirit of Figure 15.8 that captures the hierarchical framework. Think: What are the "groups?"
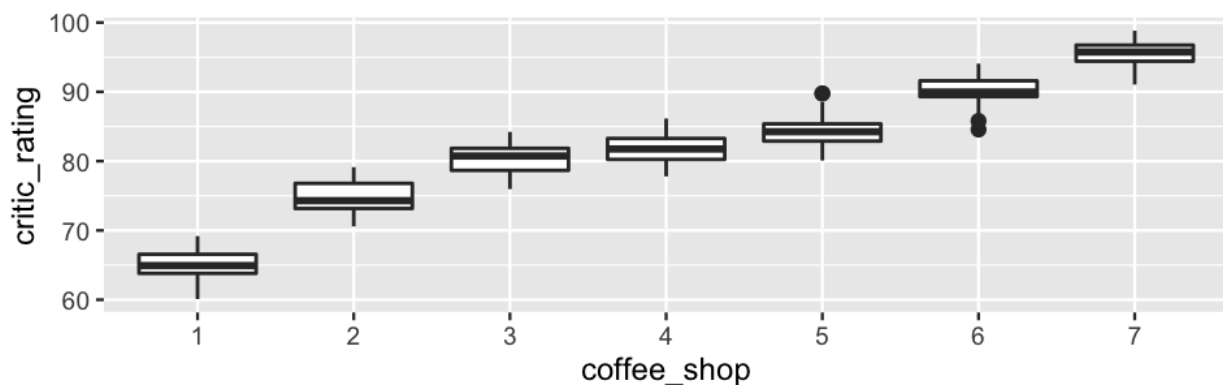
```r
lme4::sleepstudy |> head(10)
```

```
  Reaction Days Subject
1 249.5600    0     308
2 258.7047    1     308
3 250.8006    2     308
4 321.4398    3     308
5 356.8519    4     308
6 414.6901    5     308
7 382.2038    6     308
```

```
8  290.1486    7     308
9  430.5853    8     308
10 466.3535    9     308
```

2 Exercise 16.1 (Shrinkage)

The plot below illustrates the distribution of critic ratings for 7 coffee shops. Suppose we were to model coffee shop ratings using three different approaches: complete pooled, no pooled, and hierarchical. For each model, sketch what the posterior mean ratings for the 7 coffee shops might look like on the plot below.

3 Exercise 16.2 (Grouping variable or predictor?)

(a) The `climbers_sub` data in the {bayesrules} package contains outcomes for 2076 climbers that have sought to summit a Himalayan mountain peak. In a model of climber `success`, is `expedition_id` a potential predictor or a grouping variable? Explain.
(b) In a model of climber `success`, is `season` a potential predictor or a grouping variable? Explain.
(c) The `coffee_ratings` data in the {bayesrules} package contains ratings for 1339 different coffee batches. In a model of coffee ratings (`total_cup_points`), is `processing_method` a potential predictor or a grouping variable? Explain.
(d) In a model of coffee ratings (`total_cup_points`), is `farm_name` a potential predictor or a grouping variable? Explain.

4 Exercise 16.3 (Speed typing: interpret the coefficients)

Alicia loves typing. To share the appreciation, she invites four friends to each take 20 speed-typing tests. Let Y_{ij} be the time it takes friend j to complete test i.

(a) In modeling Y_{ij}, explain why it's important to account for the grouping structure introduced by observing each friend multiple times.

(b) Suppose we were to model the outcomes Y_{ij} by (16.5). Interpret the meaning of all model coefficients in terms of what they might illuminate about typing speeds:

- μ_j:
- μ:
- σ_y:
- σ_μ:

5 BR 17.1

To test the relationship between reaction times and sleep deprivation, researchers enlisted 3 people in a 10-day study. Let Y_{ij} denote the reaction time (in ms) to a given stimulus and X_{ij} be the number of days of sleep deprivation for the ith observation on subject j. For each set of assumptions below, use mathematical notation to represent an appropriate Bayesian hierarchical model of Y_{ij} vs X_{ij}.

(a) Not only do some people tend to react more quickly than others, sleep deprivation might impact some people's reaction times more than others.

(b) Though some people tend to react more quickly than others, the impact of sleep deprivation on reaction time is the same for all.

(c) Nobody has inherently faster reaction times, though sleep deprivation might impact some people's reaction times more than others.

6 BR 17.6

> **i** Note:
>
> In the group exercise, you'll write out the model and do an EDA. You will fit the models on the individual homework this week

You will explore this relationship in more depth here. To this end, suppose researchers tell us that on a typical day, the average person should have a reaction time of roughly 250ms to the stimulus used in the sleep study. Beyond this baseline, we'll balance weakly informative priors with the `sleepstudy` data from the {lme4} package to better understand reaction times. Specifically, consider two possible models as expressed by `stan_glmer()` syntax:

1. `Reaction ~ Days + (1 | Subject)`
2. `Reaction ~ Days + (Days | Subject)`

(a) What's the grouping variable in the `sleepstudy` data and why is it important to incorporate this grouping structure into our analysis?

(b) Use formal notation to define the hierarchical regression structure of models 1 and 2.
(c) Summarize the key differences between models 1 and 2. Root this discussion in the sleep study.
(d) Using the `sleepstudy` data, construct and discuss a plot that helps you explore which model is more appropriate: 1 or 2.

```
library(lme4)
data(sleepstudy)
```