Sta-230 Takehome (S23)

Instructions:

- You are expected to complete this assignment *entirely on your own*
  - You *may* ask Prof. Miller *clarification questions*, but not coding-related questions
  - You *may* use any course materials or textbooks
  - You *may* view other online resources that pre-date the assignment (ie: you can browse existing Stack Overflow questions, but cannot post your own question).
  - You *may not* discuss the assessment with your classmates until after the due-date has passed
  - You *may not* consult with class mentors, DASIL mentors, or anyone else regarding content related to the assignment
- You have until *midnight on Sunday 2/19* to complete the assignment
  - I suggest budgeting approximately 3-4 hours of working time
  - There will be no class on Friday 2/17
  - You should submit (via P-web) an R Script or R Markdown file that contains all code needed to recreate the visualizations shown later in this document
  - See the project description for additional details related to grading, expectations, etc.

Data Description:

The data needed for this project are available from the following links:

```
Ready = read.csv("https://remiller1450.github.io/data/Ready.csv")
Blood = read.csv("https://remiller1450.github.io/data/Blood.csv")
```

These data are from an actual driving simulator study exploring the impact of cannabis use on driving performance at various time-points following acute use. The study had each participant engage in 4 simulated drives. The first (Time 1) was a baseline drive, where the participant was completely sober. Baseline blood measurements were also collected prior to this drive. After the baseline drive, each participant smoked cannabis for 15-minutes, then had a second set of blood measurements collected. Afterwards, they engaged in three additional drives at approximately 30-minutes, 90-minutes, and 180-minutes after the end of cannabis dosing. Prior to each drive, participants were asked a yes/no question if felt ready to drive (recorded as "RTD").

Your task will be to create visualizations that relate the readiness to drive (RTD) responses with a participant's blood measurements.

Below you will find a more detailed description of the contents of each data file:

- "Ready.csv" contains participant IDs for everyone who completed the study, along with four columns ("RT1", "RT2", "RT3", "RT4") documenting their self-reported readiness prior each drive (ie: "RT1" is the baseline drive, "RT2" the 30-minutes post-dose drive, etc.)
  - There are three additional columns: "Days" records the participant's number of days in the past 30 where they had driven within 2-hours of cannabis use, "Ever" indicates if they've ever driven within 2-hours of cannabis use, and "Better" indicates if they believe they are a safer driver after using cannabis.
- "Blood.csv" contains participants IDs and blood measures for everyone with at least one blood draw. Some of the recruited participants may have had one or more blood measurements recorded but did not complete the study. Your visualizations should exclude anyone who does not have a corresponding record in "Ready.csv".
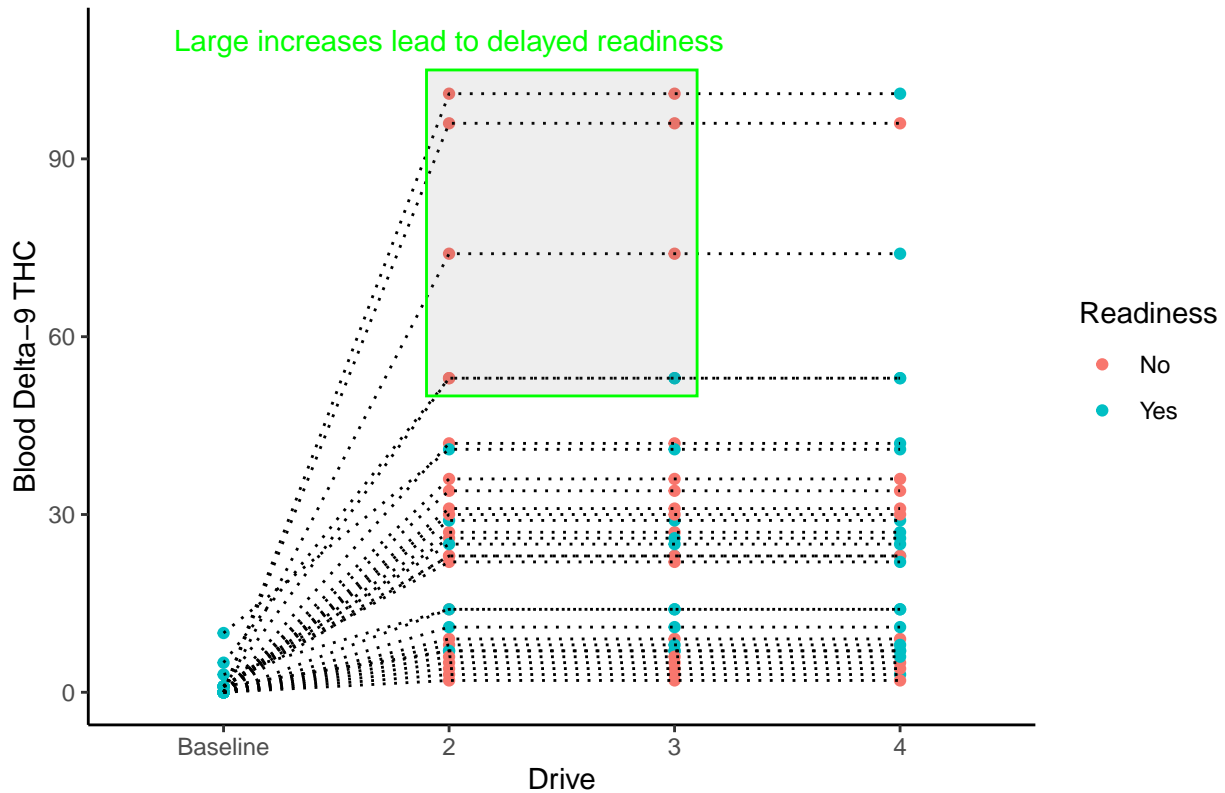
– For your reference, "delta_9_thc" is considered the "active" THC molecule within the body, while "THC_COOH" is a by-product resulting from the breakdown of "delta_9_thc" that reflects recent cannabis.

Visualizations:

Visualization #1 – Changes in Delta−9 by Drive and Readiness
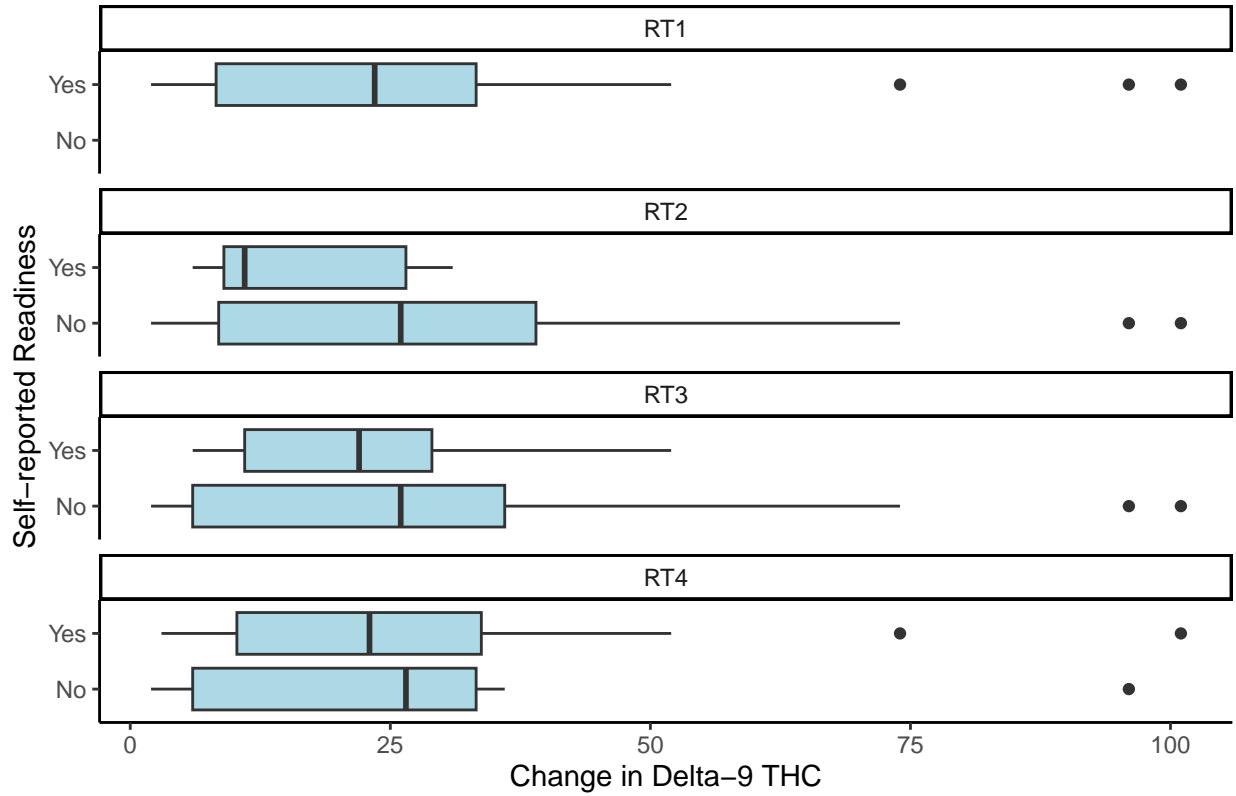
Visualization #2 – Individual trends in Delta−9 and Readiness

Hints:

For Visualization #1:

- The x-axis represents the *change in Delta-9 THC* calculated by the *post-dose measurement minus the baseline measurement*
- The boxes are filled with the color named "lightblue"

For Visualization #2:

- The y-axis heights at Baseline are the pre-dose blood THC concentrations, while the y-axis heights at Drives 2-4 are post-dose concentrations.
- You might consider *local specification of aesthetics* (described in the preamble of Lab 3, and used in Question #3 of HW #2) to allow the lines and points to be colored differently.
- You might consider setting up the variable used on the graph's x-axis using the `factor()` function (as described towards the end of Lab 2)
- The shaded rectangle has x-coordinates at 1.9 and 3.1, and y-coordinates at 50 and 105. It also uses `color = "green", alpha = 0.1`.

Partial Credit Guidelines:

These data visualizations require the proper application of several different data manipulation steps. If you cannot determine these steps, you may still receive partial credit for constructing visualizations that adequately demonstrate your knowledge of `ggplot`.

- For Visualization #1, partial credit will be given to any visualization that displays side-by-side boxplots with faceting. Greater amounts of partial credit will be given for visuals that incorporate additional elements of the one shown above (ie: theme changes, the blue coloration, etc.)

- For Visualization #2, partial credit will be given to any visualization that displays a scatter plot containing 4 different points per subject. Greater amounts of partial credit will be given for visuals that incorporate additional elements of the one shown above (ie: lines connecting subjects, coloring by readiness, theme changes, annotations, etc.)