Activity: Does interacting with dogs help relieve exam stress?

Group members:

Instructions: Work with a neighbor on the following activity. I will collect the handout at the end of class, and it will be part of your class participation grade. You will be graded only on effort – it is ok if you don't finish all the questions, or get them all correct.

Background

With the goal of helping students manage stress, many colleges and universities provide the opportunity for students to interact with animals (dogs, goats, rabbits, etc), particularly during final exam season. While these sessions are often popular, however, we may wonder whether they do really help reduce feelings of stress.

To investigate this question, a 2022 paper by Binfet *et al.* analyzed the results of an experiment in which students interacted with dogs, and reported their feelings of stress before and after the interaction. Students were randomly assigned to one of three groups with different levels of dog interaction, and the researchers hypothesized that students who interacted directly with a dog would benefit more.

The data

To download the data, run the following chunk of code in R:

```
library(tidyverse)
raw_data <- read_csv("https://raw.githubusercontent.com/ciaran-evans/dog-data-analysis/main/
```

You can then inspect the data, for example by running `View(raw_data)` or `glimpse(raw_data)`.

You will notice that there are many columns (210, in fact!). The first few columns give some information about the student in the experiment (their age, their year in school, etc.).

Group assignment

An important column here is `GroupAssignment`, which takes values Control (the student did not interact with a dog), Indirect (a dog was present but the student didn't pet them), and Direct (a dog was present and the student could interact directly with the dog).

Variables measuring well-being and ill-being

The rest of the variables record student responses to a series of questions about their stress levels and well-being. The variable prefix describes what the question is designed to measure ("HA" for Happiness, "SC" for social connectedness, "L" for loneliness, etc.), while the suffix is the question number for that section.

For example, student Happiness is measured with four different questions. Student responses *before* the dog interaction are recorded in the `HA1_1`, `HA1_2`, `HA1_3`, and `HA1_4` columns, while their responses *after* the interaction are recorded in the `HA2_1`, `HA2_2`, `HA2_3`, and `HA2_4` columns. For each question, students are asked to respond with a number between 1 and 5. For example, the first question (`HA1_1` and `HA2_1`) is this:

> `Please circle the point on the scale that you feel is most appropriate in describing you`
>
> `In general, I consider myself:`
>
> `1 2 3 4 5`
>
> `where 1 is not a very happy person, and 5 is a very happy person`

A student's **overall happiness** is measured by averaging their responses to these four questions. (Likewise, social connectedness is measured by averaging the `SC...` answers, loneliness is measured by averaging the `L...` answers, etc.).

Questions

1. **Data:** What does a row in the data represent?

2. **Wrangling:** We will need to do some data wrangling with the raw data before we can use it to answer the research question. Think about what you want the *final* data to look like, to make it easy to answer the research question (do students who interact more with dogs see a greater reduction in stress and a greater improvement in well-being?). Describe what variables you want – are these variables available in the raw data, or will you need to create them?

3. **Wrangling:** Brainstorm some of the steps that need to be done to turn the raw data into the data you *want* from question 2.

4. **Wrangling:** Are you familiar with any tools (e.g. R functions) that would allow you to carry out the data processing steps from question 3?

5. **Ethics:** The original data from the Binfet *et al.* paper actually includes a *lot* more demographic information on the students (their ethnicity, gender identity, etc.). Why do you think I removed these variables before sharing the data with you?