Getting It Together

Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

Putting data together

- Previously, we have looked at data sets by themselves.
- But sometimes, we can learn more from a data set by adding in information from a separate data set.
- In this lab, we will investigate whether your personality color says anything about how stressed or chill you are during the day.
  - To do this, we will need to merge two data sets together.
- *Download, upload, load* your *Personality Color* data set and name it `colors`.
- Then, *download, upload, load* your *Stress/Chill* data set and name it `stress`.

Looking at Stress/Chill

- Go ahead and `View` the *Stress/Chill* data.
- Based on the variables in the *Stress/Chill* data set:
  - **Which of the provided variables do you think might tend to raise or lower a person's stress/chill level?**
  - **Do you think a person's personality might affect his/her general level of stress?**
  - **Using just the *Stress/Chill* data, would you be able to answer the question above? Why or why not?**

Deciding how to merge

- Before we merge data, we need to decide *how* we plan to merge it.
  - We can *stack* our data sets - that is, take one data set's rows and add them to the bottom of the other data set.
  - We can also *join* our data sets horizontally. This is where we take one data set's columns and add them to the end of the other data set's columns.
- We have data about people's stress levels in our `stress` data and data about people's personalities in our `colors` data.
  - **Write down the steps you would take if you had to merge these data sets together manually.**

Finding variables in common:

- Look at the `names` of the variables in each data set.
- To merge different data sets together, we need to find variables they have in common.
  - **Which variables do the data sets have in common?**
  - **Which variable would make sense to merge the data sets together with? Why not one of the others?**

Putting them together

- Before merging the variables together:
  - **Write down the dimensions of each data set.**
  - **Write down the number of rows and columns you think the *merged* data set should have.**
- To merge the data sets together, run the following command:

```
merge(stress, colors, by = "user.id")
```

- Save this `merged` data set as `stress_colors`.

Looking at our data

- `View` your merged data. Does it look *right*?
- **What are the dimensions of this new `stress_colors` data set?**
  - **Why did it make sense to combine the data by joining the columns?**
  - **Why didn't we stack the rows of data instead?**
- **What happens if you swap the order of the data sets in the code above? Do the dimensions change?**

Saving our file to use later

- Now that we've created the data set of interest, it makes sense for us to save it so we can use it again later.
- Run the following code to save your `stress_colors` data:

```
save(stress_colors, file = "~/stress_colors.rda")
```

- Look in your home folder in the *Files* tab to make sure your data was saved.

A preview...

- In the next lab, we will analyze our newly merged data. But for now, try creating a plot to answer the following question:
  - **Did some (predominant) personality colors spend more time outside than others? If so, which? Write down why you think that might be the case.**