

Getting it together

Unit 2 - Lab 8

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

Putting data together

- In this lab, we'll continue our investigation into whether the personality color traits are meaningful.
 - We'll investigate whether your personality color says anything about how stressed or chill you are during the day.
 - To do this, we'll need to merge two datasets together.
- Load your personality color data and name it **colors**.
- Then *Download, upload, load* your stress/chill data set and call it **stress**.

Looking at stress/chill

- Go ahead and **View** the stress/chill data.
- Based on the variables in the stress/chill data:
 - **Which of the provided variables do you think might tend to raise or lower people's stress/chill level?**
 - **Do you think people's personalities might effect their general level of stress?**
 - **Using just the stress/chill data, could you answer the question above?**

Deciding how to merge

- Before we merge data, we need to decide *how* we plan on merging it.
 - We can *stack* our data sets, that is take one data set's rows and add them to the bottom of the other.
 - We can also *join* data sets horizontally. Where we take one data set's columns and take them on at 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 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 we want, it makes sense for us to save it for later use.
- 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'll analyze our newly merged data. But for now, try creating a plot to answer the following question:
 - Did some (predominant) personality colors spent more time outside than others? If so, which? Write down why you think that might be the case.