Free Assignment 2: Using the Tidyverse and graphs with ggplot

This assignment will build on a lot of the work in Interactive Assignments 3 and 4. We will use the same data that we used in Interactive Assignment 4 where we examined how older and younger adults rated the facial appearance of older and younger faces.

In this assignment, we will need the `tidyverse` and `ggplot` libraries. Fortunately, ggplot is part of the `tidyverse` so all we have to load is the `tidyverse` package.

In this assignment, you will produce an R-Notebook answering all the questions here. Make sure you include code and annotations telling me what each code chunk is doing.

Step 1: Create a new folder for this assignment. Download the IA4graph.csv data that we used for the last assignment in that folder and create new R notebook. Save the notebook in that folder.

Step 2: Now we should load the data that was used in Interactive Assignment 4. Go ahead and load the file named IA4graph.csv as the data.frame `face`, just like you did for Interactive Assignment 4. Create a code chunk which loads the `tidyverse` library.

As a reminder from Interactive Assignment 4, these data are a series of face ratings and were used in several studies looking at how older adult (older than 65 years) and younger adult (18-30) raters perceive older and younger adult faces. The variables are as follows:

- FaceSex - the sex of each face rated (F for female, M for male)

- FaceAgeGroup - the age group of each face rated (OA for older adult faces, YA for younger adult faces)

- FaceAgeNumber - the numeric age of the individual rated

- ag.old and ag.young - ratings of aggressiveness by older adult and younger adult raters respectively

- at.old and at.young - ratings of attractiveness by older adult and younger adult raters respectively

- ba.old and ba.young - ratings of babyfaceness by older adult and younger adult raters respectively

- co.old and co.young - ratings of competence by older adult and younger adult raters respectively

- he.old and he.young - ratings of health by older adult and younger adult raters respectively

- un.old and un.young - ratings of untrustworthiness by older adult and younger adult raters respectively

Filtering Data and Histograms

In this assignment, we are going to look at differences between older faces and younger faces. The first step is to create two new data frames. One data frame will be for older faces and one data frame will be for younger faces. As we discussed in Interactive Assignment 3, you can use the `filter()` command to create a data frame that contains only rows that fit a certain condition.

Step 3: Use the `filter()` command to create a new data frame called "face.of" which only contains older faces, or those faces where the variable FaceAgeGroup is equal to "OA".

Step 4: Use the `filter()` command to create a new data frame called "face.ya" which only contains younger faces, or faces where the variable FaceAgeGroup is equal to "YA".

Step 5: Answer the following questions as annotations in your notebook:

1. How would you use the `filter()` command to select only faces that are younger adult and males? How would you select older adult females? For this question, you do not need to type the code into R, but just write down the answer below or type the answer in your R notebook

2. How would you use the `filter()` command to select only those faces that are above the median in younger adult attractiveness ratings and above the median in older adult attractiveness ratings. Indicate which command you would type in your R notebook.

Step 6 Using ggplot, create a histogram for the aggressiveness rating variables for older faces using the variables ag.old and ag.young for the dataframe face.of. Make your histogram have 10 bins, have bars that have black lines and dark red fill. In your annotation, what pattern do you see with the two histograms? Do the ratings appear to be normally distributed? Are there any differences between the ag.old and ag.young variables?

Your histogram for ag.old should look like this:
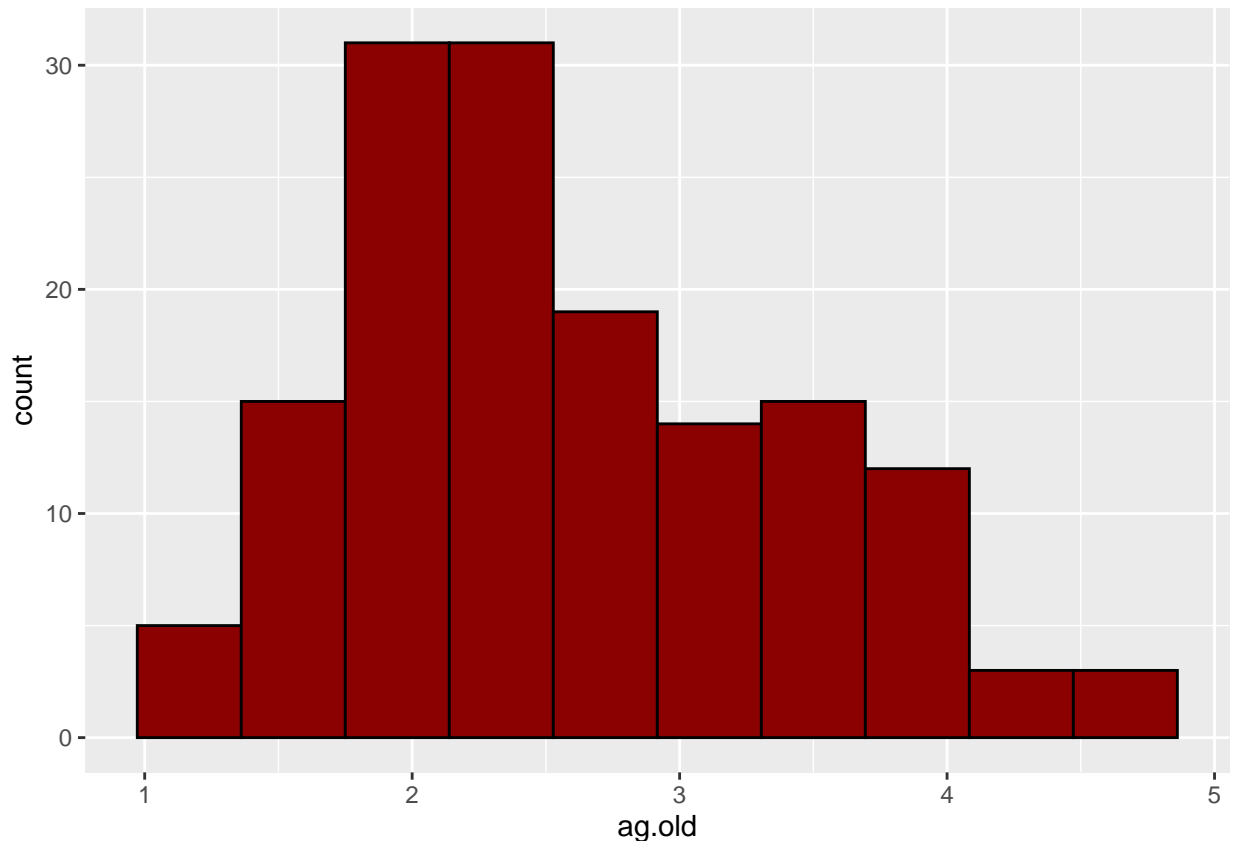
Step 7 Now create two histograms for the aggressiveness variables for the younger faces. The only difference is that instead of using the data frame "face.of", you should use the data frame "face.yf". In your annotation answer this question: What pattern do you see with the two histograms of younger adult faces? Do the ratings appear to be similar to the ratings for older adult faces?

Scatterplots

Now we are going to create a scatterplot, looking at the correlation between aggressiveness ratings by older raters and aggressiveness ratings by younger raters. We want to see whether the correlation is stronger for older faces or is it stronger for younger faces. That is, do older and younger adults agree more when rating how aggressive older faces are or do they agree more when rating how aggressive younger faces are?

Step 8 Create a scatterplot looking at the relationship between ag.old and ag.young for older faces. When making your scatterplot, make sure that you have the following features:

- Your plot title should say "Older Faces" so that we know that this plot is for older faces only
- Your x axis label and y axis label should say "Older Adult Raters" and "Younger Adult Raters" on the correct axis, depending on which one you plot on the x axis and which one you plot on the y axis.

You can change the other elements such as the type of points and colors to anything you want. Your plot should look like the plot below:
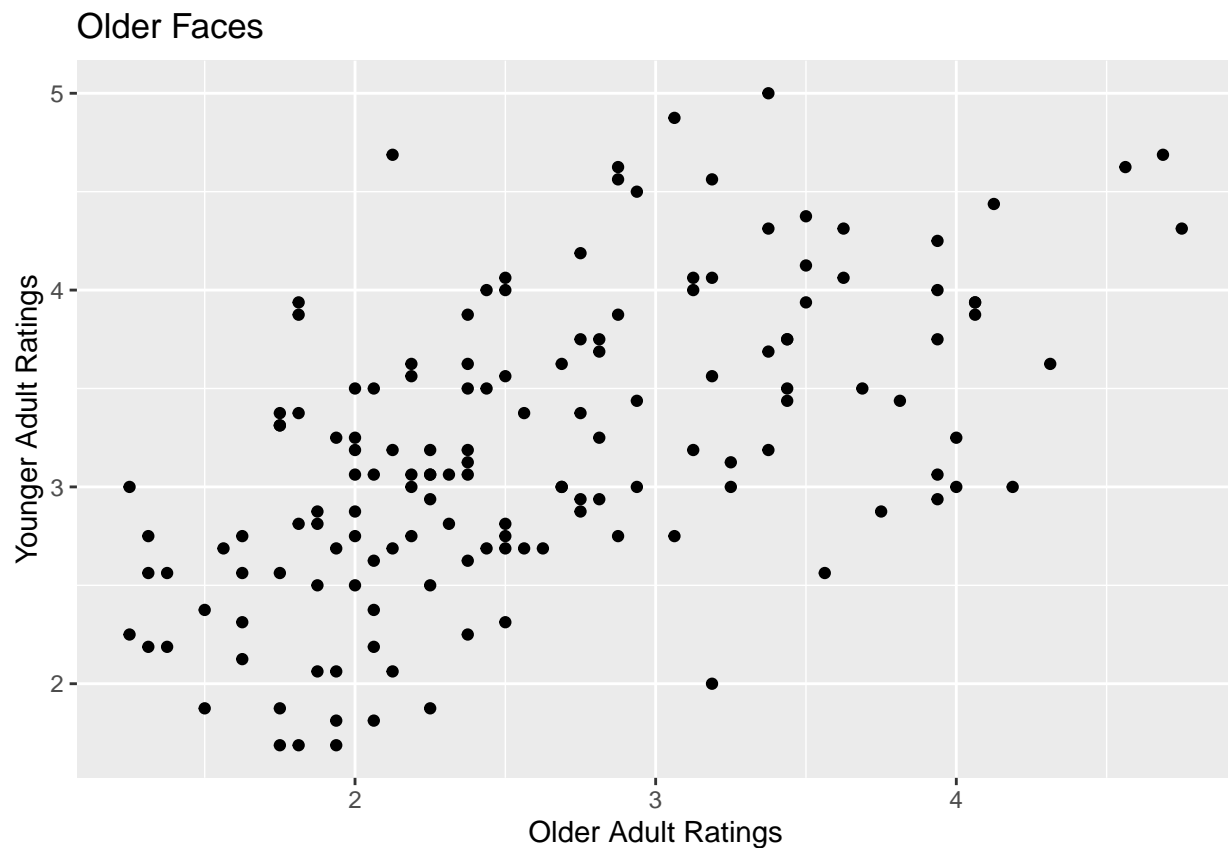
Step 9 Create the same plot you did in Step 6 only instead using younger faces. Make sure you change the title, but everything else can be the same. In your annotation, compare the plot for older faces and younger faces. Does it look like the correlation is stronger for older faces or younger faces? Does it look like it is the same?

One concept we will talk about later in this class is the idea of a **moderator**. A moderator is a third variable that affects or *moderates* the relationship between two variables. When there is a moderator, the correlation between two variables is different, depending on the moderating third variable, which indicates what group the variables are in.

In this case, the age of the face may be a moderator. The age of the face indicates what group that the variable is in (older face or younger face). The two variables being correlated are ag.old and ag.young. In this case, ag.old and ag.young is looking at whether ratings of the faces by older people is correlated with ratings of the face by younger people. That is, do older and younger people agree on how aggressive a face is.

The moderator is the age of the face. If the correlations are different for older and y ounger faces, it may be the case that older and younger raters agree on how aggressive faces are when they are one age but don't agree for another age group. We will discuss this in more detail in the next section!

Step 10 Use the `cor.test()` function to see whether the correlation between ag.old and ag.young is different for the older faces versus the younger faces. What are the correlations? Do the results support your answer for the graphs above? Is there more agreement for one age versus the other, or are they roughly equal?

Step 11: Repeat steps 5 - 8 for the competence ratings (the variables co.old and co.young), producing a histogram, a scatterplot, and the correlations for the competence ratings. What patterns do you see and what do you conclude? Are the results similar or different from the results for aggressiveness?

Step 12: Knit the notebook and submit it!