# Name:

# Classwork 11: **Analyzing Yeager, Hirschi, & Josephs’ Hypothesis**

**Issues from Quiz 2**

1. How much do you agree with these statements (poll everywhere):

**To do well on the quizzes, you needed to memorize a lot of stuff.**

**To do well on the quizzes, you needed to know a lot of R code.**

**The quiz felt really long.**

1. Which of these things helps: online textbook, online practice quizzes, classwork, lab, study groups?
2. What did you spend time on during the quiz? What parts took you a long time?
3. Some students did well… what did they do?
4. “Most difficult questions”: 5, 6, 17
5. What do we need to do differently next time?

**Clean Up RStudio.cloud**

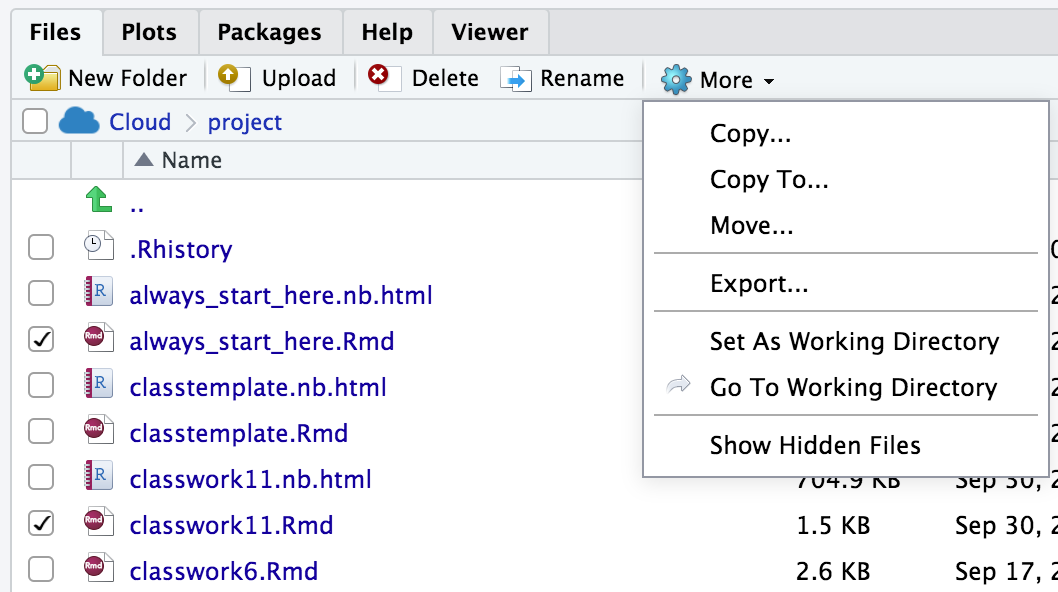
Think of a “project” on RStudio.cloud is like a “computer.” You probably just want one computer to keep all your files in. You probably don’t want each file on a bunch of different computers. If you have a bunch of different computers, it will be tough to find your files! If you only have one project, you can mostly ignore this section.

1. Just sign in to RStudio.cloud.
2. Figure out which of your “projects” you will use as your main computer. Make sure it has a name that is meaningful to you.
3. Now we need to delete the other projects. Are there any files on the other projects that you need? If so, you want to download them first. Downloading is taking them “down” from “the cloud.”
   1. There are two kinds of files that often have the same name. Circle the one you need.

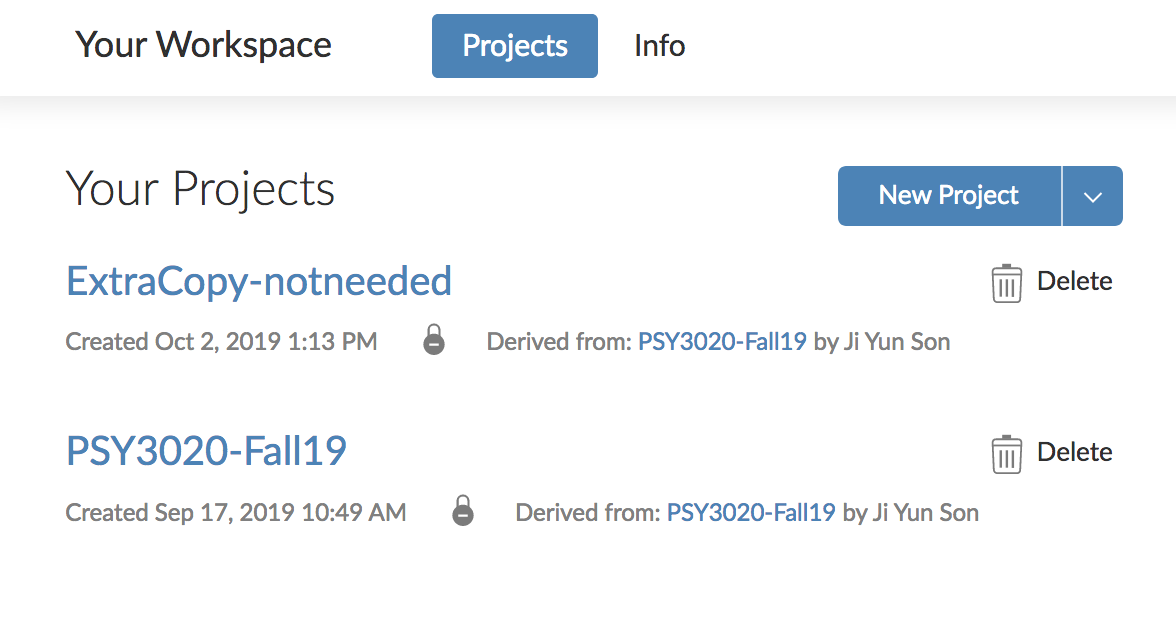
**.Rmd**

**.html**

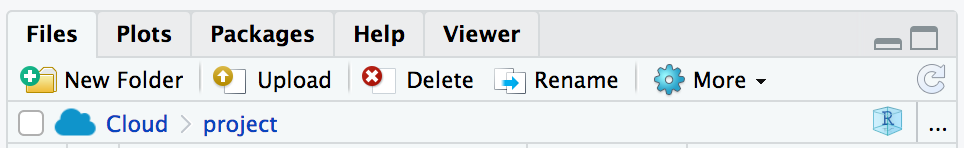
* 1. Choose the ones you want to keep. Then click on “More” and “Export…”



* 1. Did you download all the stuff you need? If so, delete those extra projects by going to your “workspace” (this is the place with all your projects) and using the trash “Delete” button.



1. Now you can upload all your files into your main project. Uploading is putting files “up” on “the cloud.” Remember where this is… we will need to upload stuff again soon!

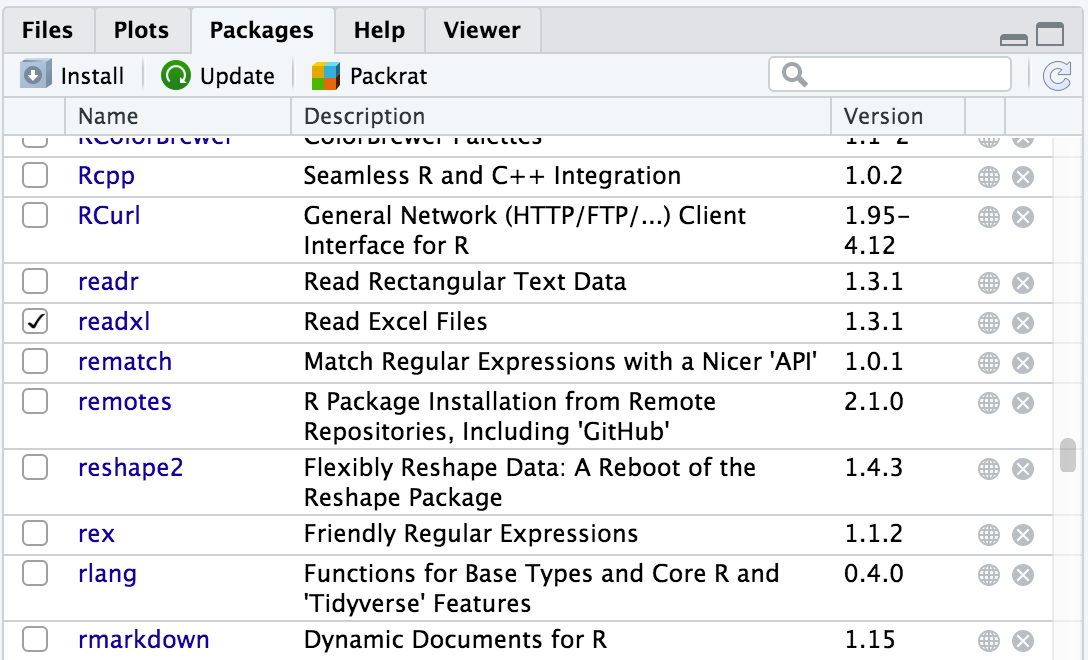


**Importing Excel Files into RStudio.cloud**

1. Let’s practice making the data available in RStudio Cloud! I’m going to teach you something new – how to put an excel file into RStudio. Run this code:

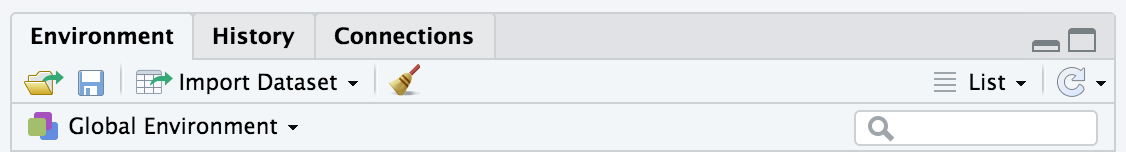
install.packages(pkgs = c("readxl"))

Check and make sure that this special package readxl appears in your list of Packages:

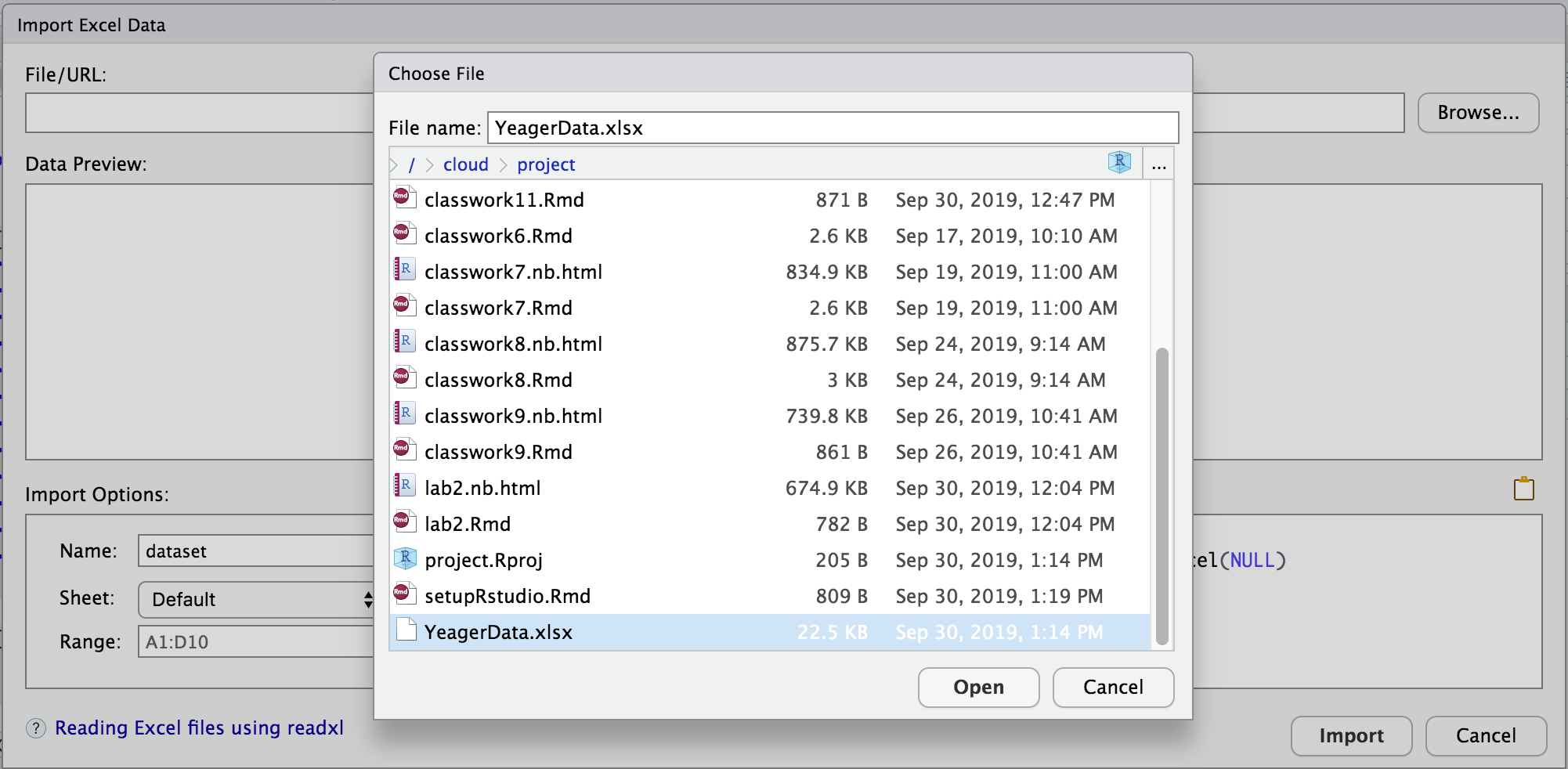


Remember, R needs to load all the libraries every time you start cooking. You may want to go to your **always\_start\_here.Rmd** file and include this library (readxl) in the list of the libraries you’ll need.

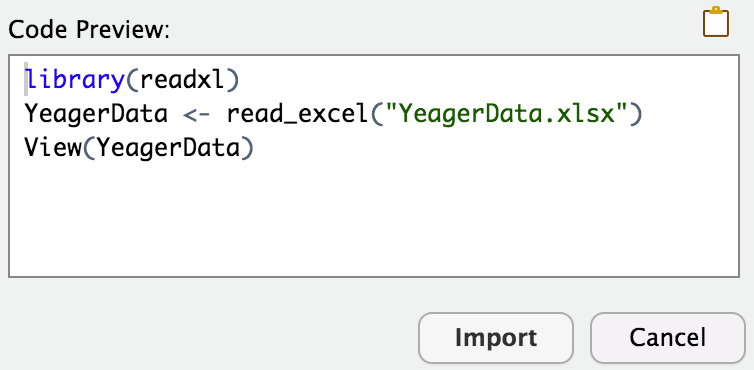
1. I have emailed a file called **YeagerData.xlsx** (a regular old excel file) to your calstatela email account. Download that. That will place this file on your computer.
2. Upload this file by clicking on the “Upload” button in R Studio. (Remember #4?)
3. Now let’s import this file as a data frame. Click the “Import Dataset” button.



Then “Browse” for this file and “Open” it.



Before you “Import” it, check the code preview. Do you want to give this data frame a different name? If so, now is the time to change it.



**Let’s return to psychology! Respect? Does it work?**

1. We are going to explore a hypothesis that Yeager, Hirschi, & Josephs’ had: *Respectful instructions will create better medical adherence (especially in adolescents)*. What do you think about this hypothesis? Also, write this idea as a word equation.
2. Now let me tell you a bit more about their experiment. Write notes here.
3. Here are all the variables in this data frame. Circle the ones that are most relevant to their hypothesis. Which is the explanatory variable? Which is the outcome variable?

Data Information

The data in each row corresponds to a single individual in the study.

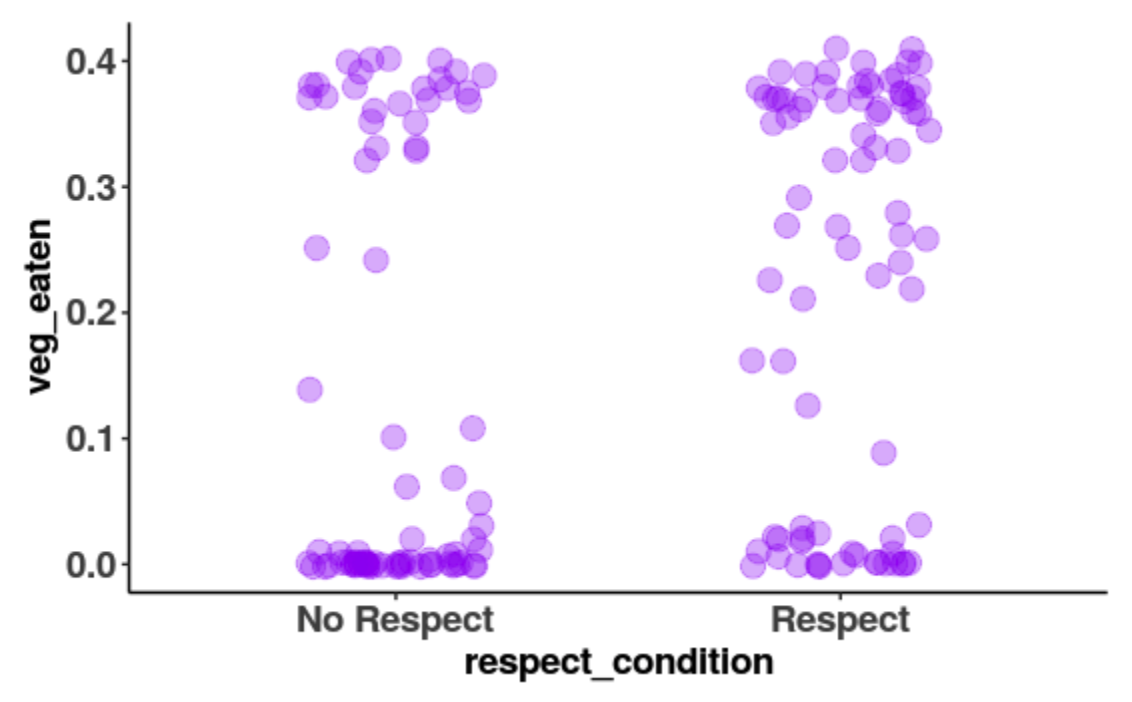
Column descriptions

|  |  |
| --- | --- |
| **subject** | ID for the participant. |
| **respect\_condition** | Whether the participant was spoken to respectfully (1) or not 0) |
| **testosterone** | The measure of the testosterone via the saliva sample |
| **spoon1\_before** | The mass (grams) of the vegemite in the 1st spoon before it was given. |
| **spoon1\_after** | The mass (grams) of the vegemite in the 1st spoon after it was given. |
| **spoon2\_before** | The mass (grams) of the vegemite in the 2nd spoon before it was given. |
| **spoon2\_after** | The mass (grams) of the vegemite in the 2nd spoon after it was given. |
| **ravens\_correct** | Number of items answered correctly in a Standardized Progressive Matrix |
| **openness** | Score from the Big Five personality test (OCEAN), 1 (low) - 7 (high) |
| **conscientious** | Score from the Big Five personality test (OCEAN), 1 (low) - 7 (high) |
| **extraversion** | Score from the Big Five personality test (OCEAN), 1 (low) - 7 (high) |
| **agreeable** | Score from the Big Five personality test (OCEAN), 1 (low) - 7 (high) |
| **narcissism** | Score from the Big Five personality test (OCEAN), 1 (low) - 7 (high) |
| **emotion\_stability** | Average score from questions regarding emotional stability, 1 (low) - 7 (high) |
| **reactance** | Average score from questions regarding reactance, 1 (low) - 7 (high) |
| **subjective\_power** | Average score from questions regarding subjective power, 1 (low) - 7 (high) |
| **aggressive\_right\_now** | “How aggressive are you feeling right now?”, 1 (low) - 7 (high) |
| **sex\_drive\_right\_now** | “How high is your sex drive right now?”, 1 (low) - 7 (high) |
| **campus\_greek** | Are you in a fraternity or sorority? |
| **status** | Average score from questions regarding personal status, 1 (low) - 7 (high) |
| **competent** | Average score from questions regarding self-competence, 1 (low) - 7 (high) |
| **autonomous** | Average score from questions regarding autonomy, 1 (low) - 7 (high) |
| **respectful** | Measure of how respectful they felt the researcher was, 1 (low) - 7 (high) |

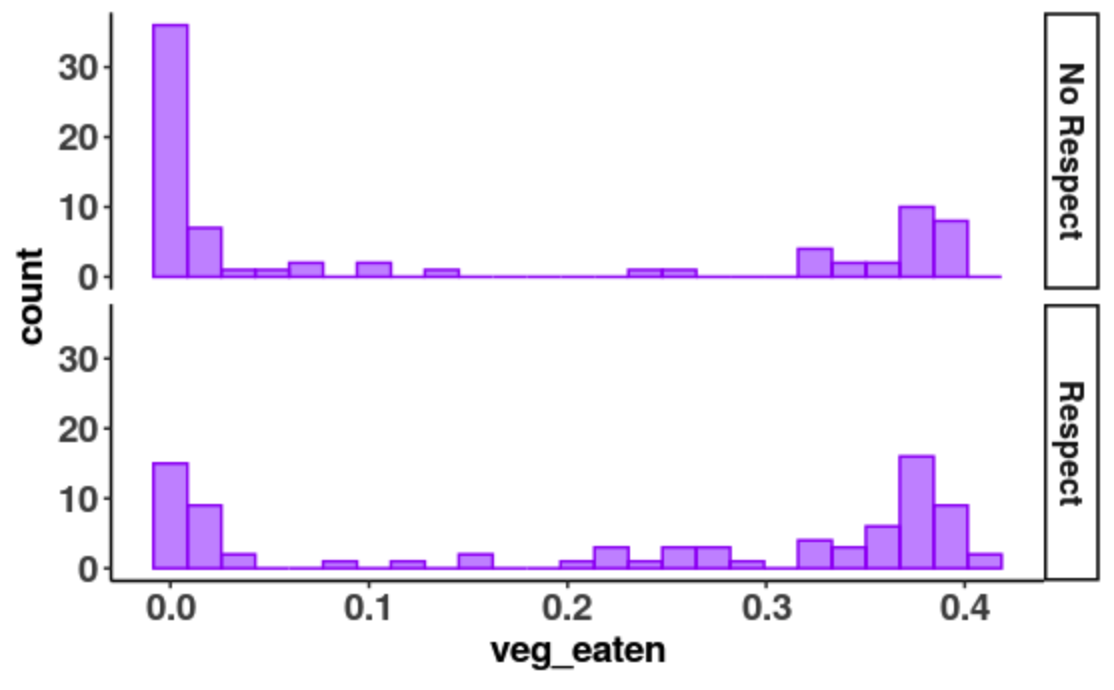
1. **Explore variation.** Explore the variation in the outcome variable using a plot or graph.
   1. What would a 0 mean on this outcome variable?
   2. Did respect make a difference? Argue each side with a partner.
      1. Made a difference:
      2. Did not make a difference:
      3. Could this pattern of data be the result of randomness?
2. **Model variation**. Let’s create a best fitting empty model.
   1. If we used the mean as our empty model to predict how much vegemite someone in this study would eat, what would we have predicted? How many people would we have predicted correctly?
   2. Why is the mean a terrible model for this data?
   3. If we used the mode as our empty model to predict how much vegemite someone in this study would eat, what would we predict? How many people would we have predicted correctly?
   4. Even though we would have predicted more people correctly using the mode, why is the mean *still* a useful model for this data?

**Mean vs. Mode**

1. Let’s write these two models, (1) mean as the empty model and (2) mode as the empty model, in GLM notation. And then let’s also put them in our jitter plot (use two different colors). Draw a few residuals off of each model.



1. Try putting in these two models into a faceted histogram as well.



1. If we calculated the sum of squared residuals off the mean versus off the mode, which number will be lower?
2. Let’s calculate the sum of squares off the *mean* as the model. You can do it the fast and easy way (e.g., using anova()) but also create a column of residuals, square them, and sum them up as well to make sure you get the same answer.
3. Let’s calculate the sum of squares off the *mode* as the model. There is no easy function but you can create a column of residuals, square them, and sum them up. You can take notes about the R code here.
4. Compare the SS from 9 and 10. What makes them different? What should we notice about these?
5. In summary, why is the mean such a “great” model (even when it is clearly not such a great prediction)?