**Objectives:** Before you begin coding, it's important to have a plan for data wrangling. In this activity, you will be asked to obtain a specific data frame fitting one or more requirements. You will need to determine which dplyr verbs can be applied to obtain which variables in which order to obtain the desired data. You only need to worry about the **verbs**, the **variables**, and the **order**, not the particular syntax in R.

Assignment Organize yourself into groups of 2 or 3. Work together to answer the following questions. Write your answers in your notes.

## The Data

In Spring 2022, students in an intro statistics class were asked to complete a survey containing a series of serious and not-soserious questions. A key for the variables in the table is printed on the reverse side of this page, and a .csv file of the data itself is available on today's section of the schedule page of the course website.

## The Tasks

For this task, you do not need to create the actual data frames or actually compute any values. Just described the verbs, variables, and order needed to obtain the desired data frame.

- 1. Sort students by division, and within division, by year.
- 2. Determine the proportion of students who list each of the different locations as their preferred study place.
- 3. Find the number of students in each year whose primary mode of transportation is walking.
- 4. Calculate the median number of college applications submitted by students who play Wordle.
- 5. Create a data set consisting of just the categorical variables in the original data set, along with the variable Timestamp. Responses should be ordered by time submitted, from earliest to latest.
- 6. Identify students whose social views are 1 or more points higher than their economic views.
- 7. Create a data set consisting of two columns: height in centimeters and height in inches.
- 8. Count how many students think both that dogs should wear pants on their back legs and that hot dogs are sandwiches.
- 9. Among students who drink coffee and whose hometown is at least 100 miles from the college, create a data set that could be used make a scatterplot of bedtime vs. weekly hours spent studying.
- 10. Create your own data wrangling task based on this survey data frame.

## Variable Key

Explanation
Time when survey was submitted (Month / Day / Year Hour: Minute)
Height in centimeters
Number of colleges applied to
How far (in miles) the student's hometown is from the college.
Year in college
Division of the college that major belongs to
Typical location students studies in
Personal political views on social issues (1 is very conservative, 5 is very liberal)
Personal political views on economic issues (1 is very conservative, 5 is very liberal)
Number of roommates
Number of books read in the past 6 months
Primary mode of transportation
How a dog would wear pants
Are hotdogs sandwhiches?
What is your preferred caffeinated beverage?
What is your typical bedtime?
Do you have any dietary restrictions?
Do you play Wordle?

## dplyr Verbs

- select: Extracts the specified columns from the data frame.
- filter: Extracts rows meeting certain conditions from the data frame.
- mutate: Add columns to the data frame according to a formula.
- arrange: Sort the data frame according to the values of one or more variables
- group\_by: Prepare the data frame to be summarized within values of a certain variable.
- summarize: Compute specified summary statistics for specified columns of the data frame.