DATA MANIPULATION

There are now four rules that apply to all projects:

a) Follow instructions *precisely*. If I do not tell you what to write on a particular line, leave it blank.
b) Do not use any functions or approaches to problems that we have not yet learned in this course.
c) All code must be *scalable by sample size* unless specifically noted otherwise. This means your code should work equally well on a dataset with N=10 as N=1000.
d) Any code using *magrittr* should contain a max of one verb per line. Also, you are not required to use all lines specified (e.g., code on just line 4 would satisfy "line 4-6").

You're continuing working on datasets related to the project in Week 4. This time, you have four datafiles from two different labs: two participants data files and two files containing experimenter notes! The parameters are similar to last week:

1. Each participant completed the study 4 times with different versions of the stimulus.
2. The software used to collect data differed by lab, resulting in different file formats.
3. There are 9 focal study variables. You also have experimenter notes in separate files, sorted by participant number. An experimenter note indicates there was a problem with that participant.
4. The 9 focal variables and their names should ultimately be: case number (*casenum*), participant number (*parnum*), stimulus version (*stimver*), date and time of data collection (*datadate*), and *q1 – q5*. One file additionally has *q6-q10.* Case and participant numbers begin with 1 within each lab.

Part 1 – Set up a new R Studio Project with one R script called week5.R

Part 2 – Data Import and Cleaning

1. **Lines 1-3:** Write a comment that says: **R Studio API Code**, and set the wd as usual.
2. **Line 5:** Write a comment that says: **Data Import**
3. **Lines 6-10:** Using four tidy import functions, convert the four datafiles into: *Adata_tbl*, *Anotes_tbl*, *Bdata_tbl*, and *Bnotes_tbl* with appropriate column names (including *qs* in *Adata_tbl*).
4. **Line 12:** Write a comment that says: **Data Cleaning**
5. **Lines 13-17:** In *Adata_tbl*, using a single series of pipes, split **qs** into the 5 correct variable names and convert them all to numeric. Also convert the *datadate* column to its correct type. Extra credit if your numeric conversion is scalable to any number of variables; you will need a new *dplyr* verb.
6. **Lines 18-23:** Using one series of pipes per data file, aggregate across conditions such that participant number and mean scores of q1-q5 across all 4 conditions are saved in two new tbls: *Aaggr_tbl* and *Baggr_tbl*. In other words, each aggregated table should have 1 row per participant, not 4. Extra credit if you do this scalable by any number of variables.
7. **Lines 24-25:** Using a join, add participant notes to each of your newly aggregated tbls as a new column called *notes*.
8. **Lines 26-29:** Using a single series of pipes, combine cases from *Aaggr_tbl* and *Baggr_tbl*, drop rows with research notes, and report the final Ns split by datafile source as a tibble displayed to the console.

Part 4 – Submission