Assignment 4

EDH7916

Benjamin Skinner

Using the `hsls_small.csv` data set and the online codebook, answer the following questions. You **do not** need to save the final output as a data file: just having the final result print to the console is fine. For each question, I would like you to try to pipe all the commands together. Throughout, you **should** account for missing values by dropping them.

For each question, show your data work and, if necessary, answer the question in a short (1-2 sentence(s)) comment.

Questions

1. Compute the average test score by region and join back into the full data frame. Next, compute the difference between each student's test score and that of the region. Finally, return the mean of these differences by region.
2. Compute the average test score by region and family income level. Join back to the full data frame. **HINT** You can join on more than one key.
3. Select the following variables from the full data set:

   - `stu_id`
   - `x1stuedexpct`
   - `x1paredexpct`
   - `x4evratndclg`

   From this reduced data frame, reshape the data frame so that it is long in educational expectations, meaning that each observation should have two rows, one for each educational expectation type.

   e.g. (your column names and values may be different)

stu_id	expect_type	expectation	x4evratndclg
0001	x1stuedexpct	6	1
0001	x1paredexpct	7	1
0002	x1stuedexpct	5	1
0002	x1paredexpct	5	1

Submission details

- Save your script (`<lastname>_assignment_4.R`) in your `scripts` directory.
- Push changes to your repo (the new script and new folder) to GitHub prior to the next class session.