# *SOC 5050: Lab 11*

*Christopher Prener, Ph.D.*

*October 31st, 2016*

## *Directions*

Please complete all steps below. This lab requires that you use the 2011 CPS dataset. Your final work by do-file, log-file, and markdown file with answers should be uploaded to your GitHub assignment repository by 4:20pm on Monday, October 31st, 2016.

## *Deep Cleaning Data*

1. Create a new race variable (from the current variable `PTDTRACE`) that (a) properly handles missing data and (b) reduces categories to "White only", "Black only", "American Indian/Alaskan Native only", "Asian only", "Hawaiian/Pacific Islander only", and "multi-racial" (2 or more races). Be sure that your new variable is properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
2. Construct logic checks using the `assert` command to ensure that your recodes worked correctly.
3. Create a new variable for Latino identity (from the current variable `PRDTHSP`) where "0" is used to represent data currently coded as missing and "1" is used to represent any Latino origin. The category "OTHER SPANISH" should be coded as "0" since this typically refers to individuals who identify as Spanish or another country of origin outside of Latin America.[1] Be sure that your new variable is properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
4. Construct logic checks using the `assert` command to ensure that your recodes worked correctly.
5. Create a new variable that combines your recoded race and ethnicity variables. The new variable should have the following categories: "White, non-Latino only", "Black, non-Latino only", "Latino", "American Indian/Alaskan Native only", "Asian only", "Hawaiian/Pacific Islander only", and "multi-racial" (2 or more races).[2] Be sure that your new variable is properly labeled with

[1] The social construction of race and ethnicity!

[2] *Hint*: Combine the `replace` command with `if` qualifiers, relational operators, and logical operators to produce this variable.

well-formatted value labels. Using variable notes, document the recoding process.

6. Construct logic checks using the `assert` command to ensure that your recodes worked correctly.
7. Create dummy variables of your latest variable (that combines race and ethnicity). Be sure that your new variables are properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
8. Use a loop to create new versions of the following variables: `PEHRUSL1`, `PEHRUSL2`, and `PEHRUSLT`. These new variables should be binary variables where "0" is equal to values less than or equal to the parent variable's mean and "1" is equal to value greater than the parent variable's mean. This loop should *not* rely on manually entered values; instead, you should use local macros. Be sure that your new variables are properly labeled with well-formatted value labels. Using variable notes, document the recoding process.
9. Drop all variables other than those that you have created in this lab.
10. Add a dataset label and dataset-level notes to your dataset.
11. Create data documentation files (a variable index and codebook).
12. In your analysis file, output the descriptive statistics for each of the new variables you created.

## *Document Details*

Document produced by Christopher Prener, Ph.D. for the Saint Louis University course SOC 5050 - QUANTITATIVE ANALYSIS: APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository `README.md` file for additional details. Data are drawn from the ULCA Institute for Digital Research and Education.