

- (5%) Coding Budget: You are allowed at most one DATA steps and thirteen PROC steps to complete this assignment.

Specifications:

- (5%) Setting up paths: Programmatically change your working directory as needed to set up FILENAME and LIBNAME statements. Do not use absolute paths anywhere in your code other than in these X commands at the top of your program!
- (5%) Ensure no output is sent to the listing window. The only results should appear in the PDF and PNG files.
- (15%) The raw file L:\st555\ipums 2005 values.txt contains information from the 2005 US Census on over 1.15 million households. It includes an identifier, Serial, and three pairs of variables. In each pair the second value is an amount and the first value is a description for that value. For example, in the first record the household has no mortgage payment, makes 12 thousand dollars a year, and lives in a home (apparently) worth \$9,999,999. (If that makes you suspicious that something is wrong - you're exactly correct!)
 - Read this data into a data set named Amounts.
 - Use a macro variable to specify the number of pairs of data on each household. (This allows us to use this for other years of the Census where we may have recorded only 2 of these pairs, for example.)
 - Ensure the data set ends up having a vertical structure appropriate for BY-processing. For comparison, my data set is available in L:\ST446\MP#2.
 - In some cases, such as the very first record, the respondent is not the home-owner but instead is a renter. Since the renter should not be associated with the value of the house, it should be missing. However, the value entered was actually \$9,999,999 and that value is consistently used in place of a missing value. Change all of these values back to missing.
- (10%) Demographic data is also provided by the SAS data set Demographics in the L:\ST555 folder. Use this in conjunction with the Amounts data set you created to produce a joint data set named AllData. As above, my data set is available for comparison.
- (40%) Now that you've created the data, it's time to produce some reports and graphics.
 - Produce a report like the one I've provided: State-Level Report.pdf. A few items to pay close attention to:
 - * The states included in the report must be controlled by a macro variable. I strongly suggest testing your final program using a couple of different combinations of states to make sure this works.
 - * The title is static, but the footnote is dynamic. The footnote uses 8pt font.
 - * The report is clearly dynamic - it should only include the states that appear in your macro variable.
 - * The report uses the Sapphire style in two columns with a landscape layout. It includes page numbers but not dates.
 - Produce a graph like the one I've provided: HistoBoxes.png. A few items to pay close attention to here:
 - * The main title is dynamic. You'll need to create macro variables *for whatever descriptions appear in the data*. This means if I run this on the 2010 Census and instead of calling it Home Value they called it Property Value then your code should still work.
 - * The variables that get plotted are also dynamic. For the same reason as the title, if the variable changes then the graphs must change.
 - * The labels are also dynamic. If the year changes, we would expect the median value to change and we don't want to have to manually set all these with %LET statements so you'll need to dynamically figure out the medians.
 - * The grid uses 80% of the space for the first row and the columns are evenly divided.
- (20%) Programmatically demonstrate that the following are equivalent for your results and my results. In all validation, numeric values must differ in absolute value by no more than 1E-6. All validation difference data sets should include only records related to a difference between the base and comparison data sets, should be based on using my data as the base data set, and should include copies of my record, your record, and the difference for any records where a difference does exist.
 - The Amounts data (content and descriptor portions). Save the differences as DIFF1A and DIFF2A, respectively.
 - The AllData data (content and descriptor portions). Save the differences as DIFF1B and DIFF2B, respectively.