Lab 1

Name: __Udayraj Suthapalli_____ (Also, click on the header and enter your full name)

Note: The word document of this assignment can be downloaded from CANVAS course page.

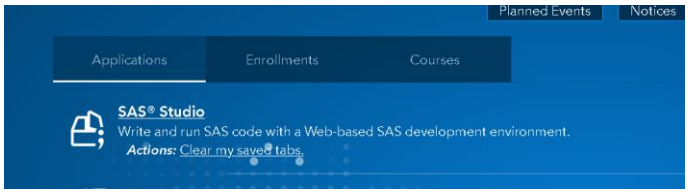
Please first read instructions and/or examples below and then answer questions in the highlighted box. Manual calculation is required for each question although you can verify your calculations using the SAS output. The SAS simulation results may not perfectly match your calculation but they should be very close.

In this lab, we are going to start using SAS Studio. If it is your first time using SAS Studio, do not panic! The instructions should be enough to complete the tasks. It is assumed that you already created a free profile on SAS® OnDemand for Academics platform.

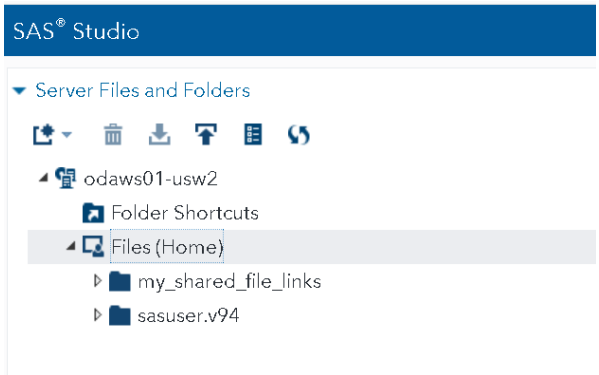
Info about the data: *"we will use PEW September 2020 Survey data. Telephone interviews conducted Sept. 22-28, 2020, among a national sample of 1,007 adults, 18 years of age or older, living in the United States (301 respondents were interviewed on a landline telephone, and 706 were interviewed on a mobile phone, including 487 who had no landline telephone). A combination of landline and mobile phone random-digit-dial samples were used. Interviews were conducted in English (972) and Spanish (35). The combined landline and mobile phone sample is weighted to provide nationally representative estimates of the adult population 18 years of age and older."*

TASK: Login to SAS Studio and Create a new folder 'Lab1_your initials' under Home folder and directly import the data from our course folder in SAS Studio. Alternatively, you can download data to your local folder and upload to SAS Studio. You can perform various analysis things in SAS Studio with point-and-click. Programming is also available for SAS Studio.

- Start SAS Studio (https://welcome.oda.sas.com/)
- Click on SAS Studio icon

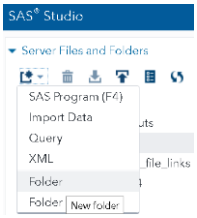


Now, you should see something as below:

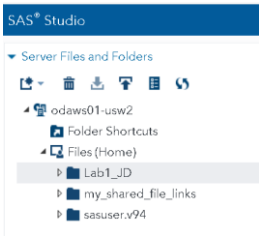


my_shared_file_links under Home folder is our class folder. Your instructor will add data over there; you will add it to your own lab folder and work on it. As an example in this lab, we will create a new folder '**Lab1_your initials' (e.g., Lab1_JD for John Doe and I will use JD when naming is needed)** under home folder and then copy the PEW data **PEW2020.sav** that is under class folder (**my_shared_file_links**) to the new Lab1_JD folder. The step-by-step instruction is as below.

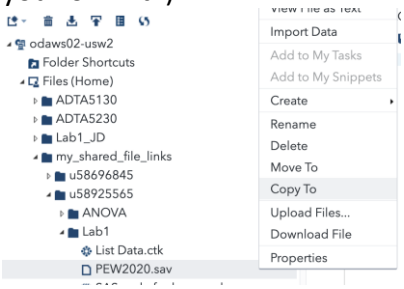
- Click on Server Files and Folders. When Files (Home) folder is selected, click on New icon  to create a folder.

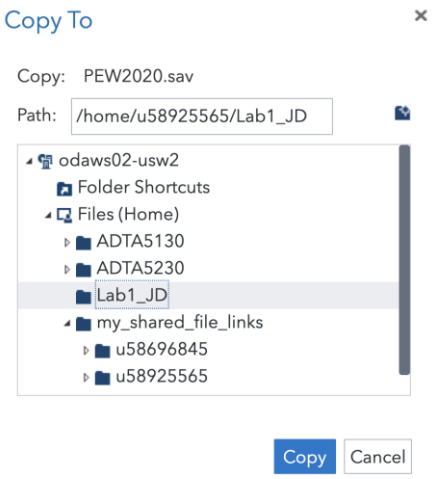


Name your folder: and then click Save. Our new folder is showing up under Files(Home).

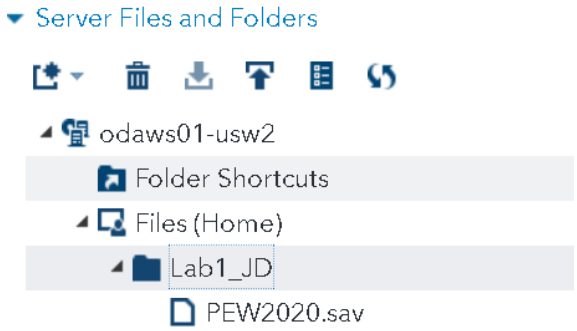


- Locate the PEW data under my_shared_file_links/u58925565/Lab1 folder, **right click** on **PEW2020.sav**, select **copy** and choose newly created **Lab1_JD** folder as destination. Click **copy.** (Please note that u58925565 is the instructor's id. you also have a folder named in your own id.)





- Now, your data is under Lab1_JD folder.

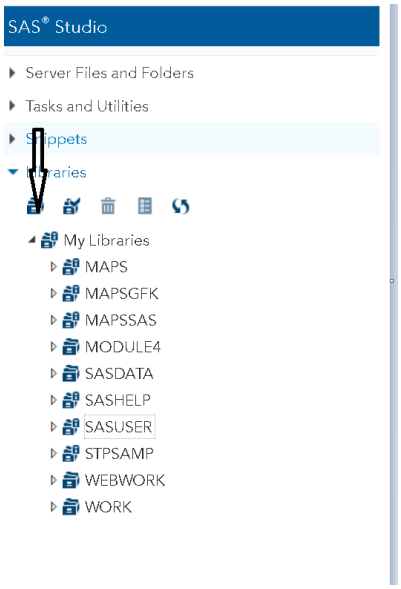


NOTE: If you have the PEW data saved in your local computer, you could just import it directly by selecting Lab1_JD folder and clicking on upload button . Meaning, you can bring your own data and use it in here for your analysis.

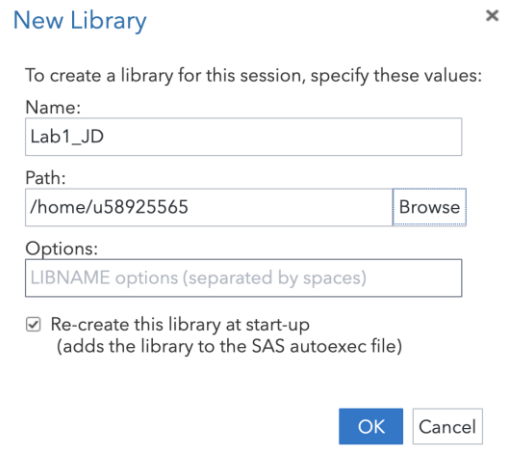
- Next, we will create a SAS data file from this PEW file and save it in a library to perform analysis. We will create a library first. We can talk about library later but the reason for creating a new library is to have long term access to the data we worked on. Otherwise, SAS will put all your work in temporary WORK library and delete them later.

Create a new library under Libraries to prepare and conduct analysis on the PEW data in SAS format

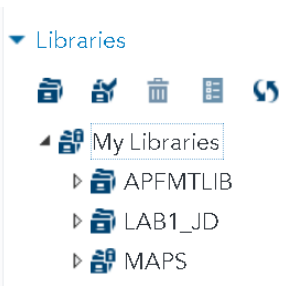
- On SAS Studio, under **Libraries/My Libraries**, click on new library icon



- Name you library in the same format of 'Lab1_yourinitials'. For path, select Files(home) folder. Click **Re-create this library at start-up.**

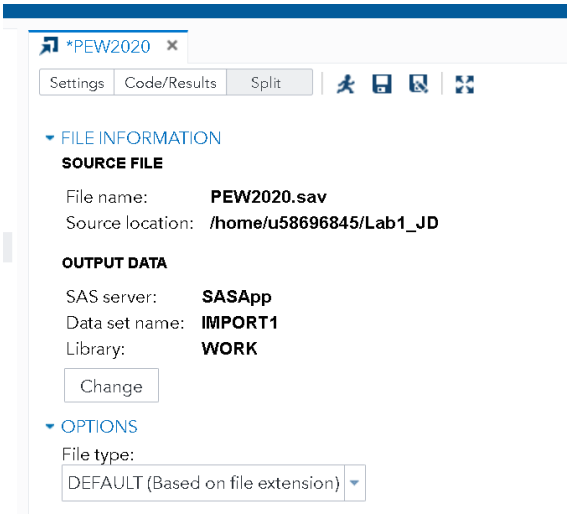


Our library is ready! Next step is to call the PEW data in this library.

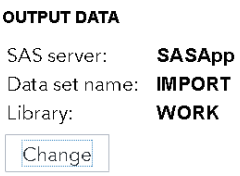


Convert PEW2020.sav data into SAS7bdat format and open in newly created library to work on it

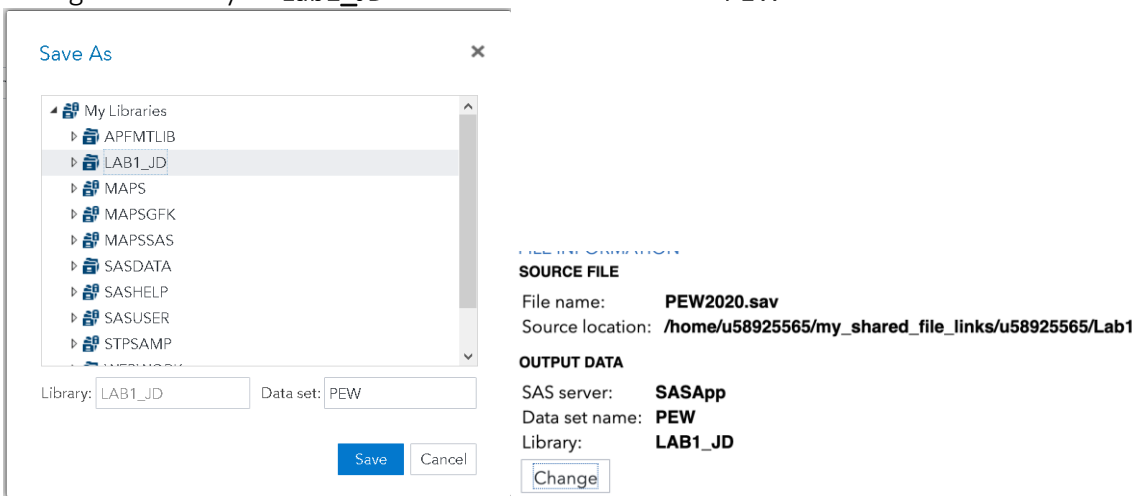
- Double click on **PEW202.sav** file under Lab1_JD folder to open it, import utility will open. If the import utility was previously opened to import a different worksheet, a new tab will not open. Do not panic. You can import a new worksheet just by changing the source and output file of the same tab.



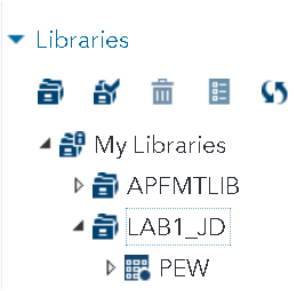
- Under **OUTPUT DATA**, click **Change**.



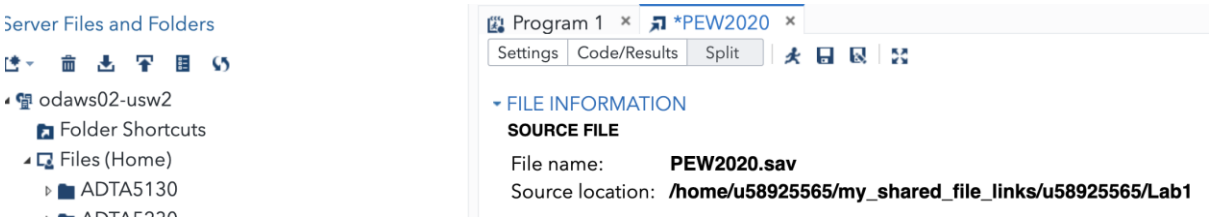
- Change the library to **Lab1_JD** and the data set name to **PEW**. Click on Save.



- Click on Run  icon.
- If you see your data under **Lab1_JD** library, you are on the right track.



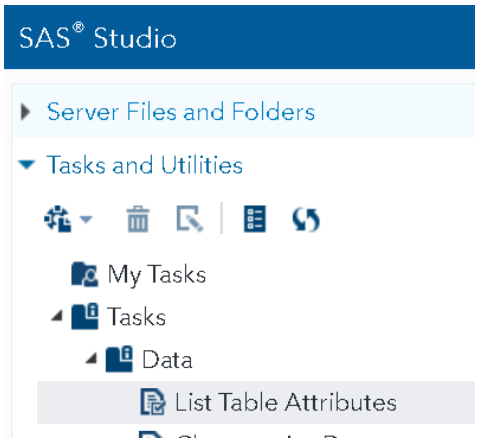
Now, we can close the PEW2020 import utility tab from the window.



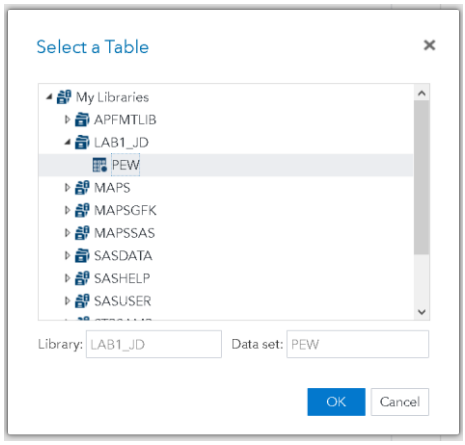
When it prompts you with “Do you want to save "PEW2020"?”, choose Don’t Save.

Now, our PEW Data is stored in Lab1_JD library in SAS data format forever. Next we will look at the attributes of this dataset.

1. Navigate to **Task and Utilities** ⇨ **Tasks** ⇨ **Data**.
2. Double-click the **List Table Attributes** task to open it.

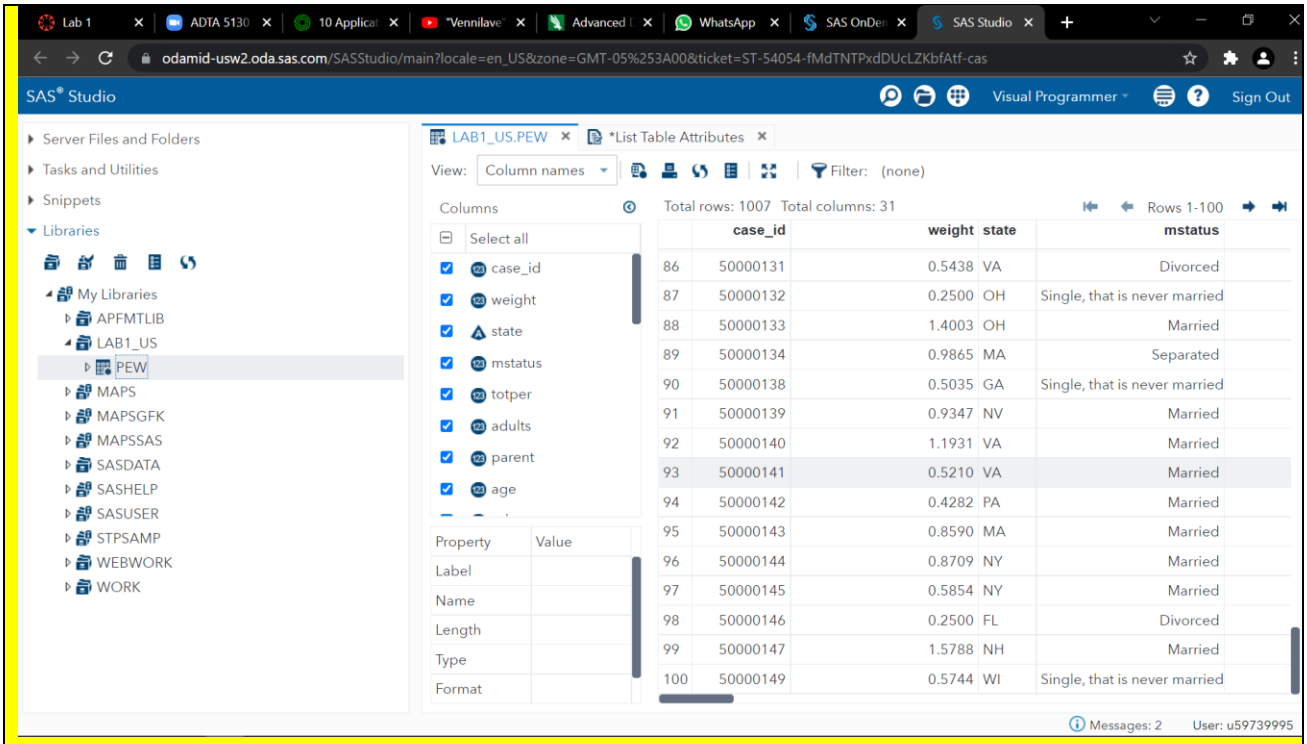


Under DATA, click and select **Lab1_JD.PEW**. Click on **OK.**



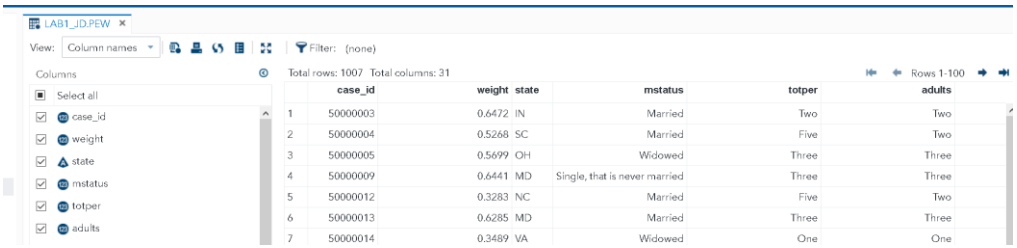
Click on Run icon or press (F3) to answer the questions in Box 1.

BOX 1: Complete this box based on the Results tab in the last step
1. Take a screenshot of your SAS data table that shows your file under the newly created library

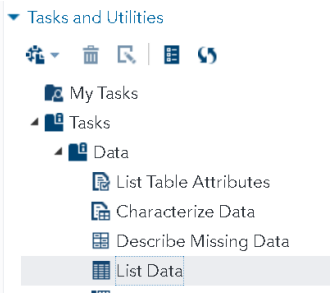


2. Based on the RESULTS table when we run **List Table Attributes** task, there are __1007_____observations and _31___ variables in our dataset.

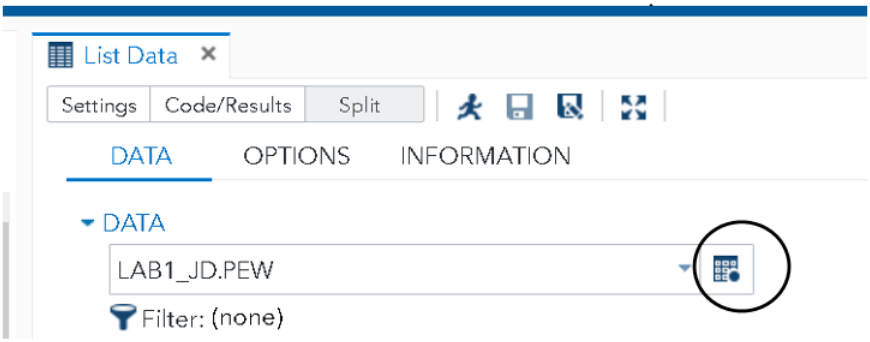
LISTING THE SAS DATA SET: An easy way to see a listing of the PEW data set is to select the Libraries tab in the navigation pane and select My Libraries. Expand the Lab1_JD library and double-click PEW. (I know, you just closed it ☺) You should be seeing something like below:



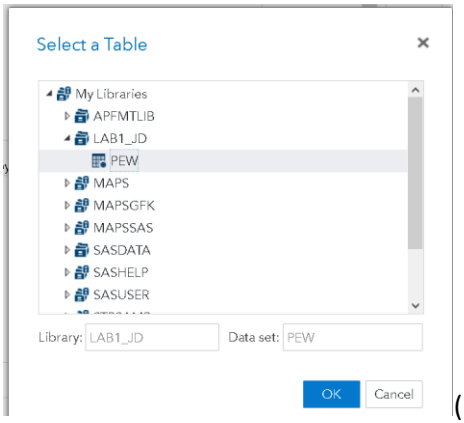
You can use scroll to the right or to the down to see the rest of the table. But we want to create a better looking report, so click the Tasks and Utilities tab of the navigation pane and select Data followed by List Data, as shown below.



Double-click List Data.

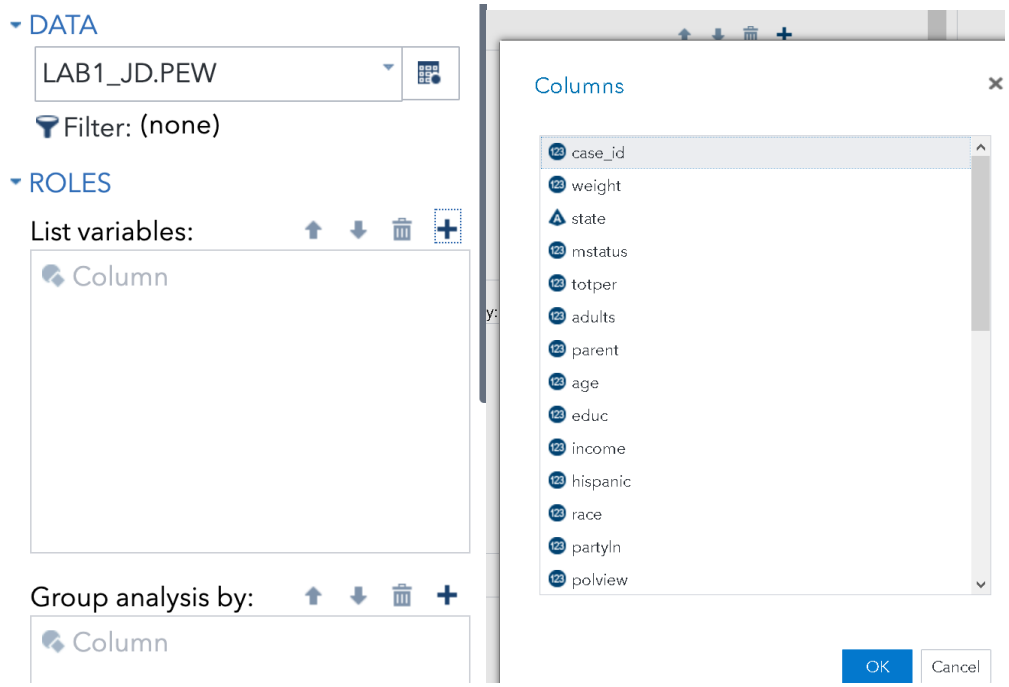


Click the icon at the far right side of the box labelled DATA (circled in the figure below) to select PEW data set. You can also type the library.SAS data set into the box to open a certain data in a certain library.

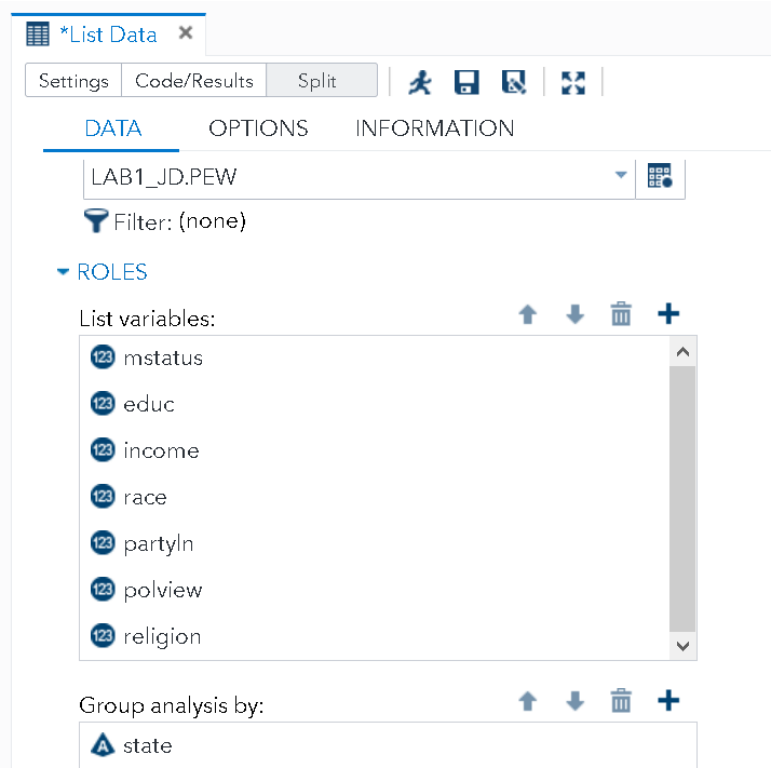


(In fact, it was already selected; I am showing just in case.)

Click OK. Your next task is to select variables to be in the list. Under ROLES, you see four boxes. At this time, you are only interested in the first two--List variables and Group analysis by. Click the plus sign next to List Variables. A list of variables from the PEW data set will show up as below:



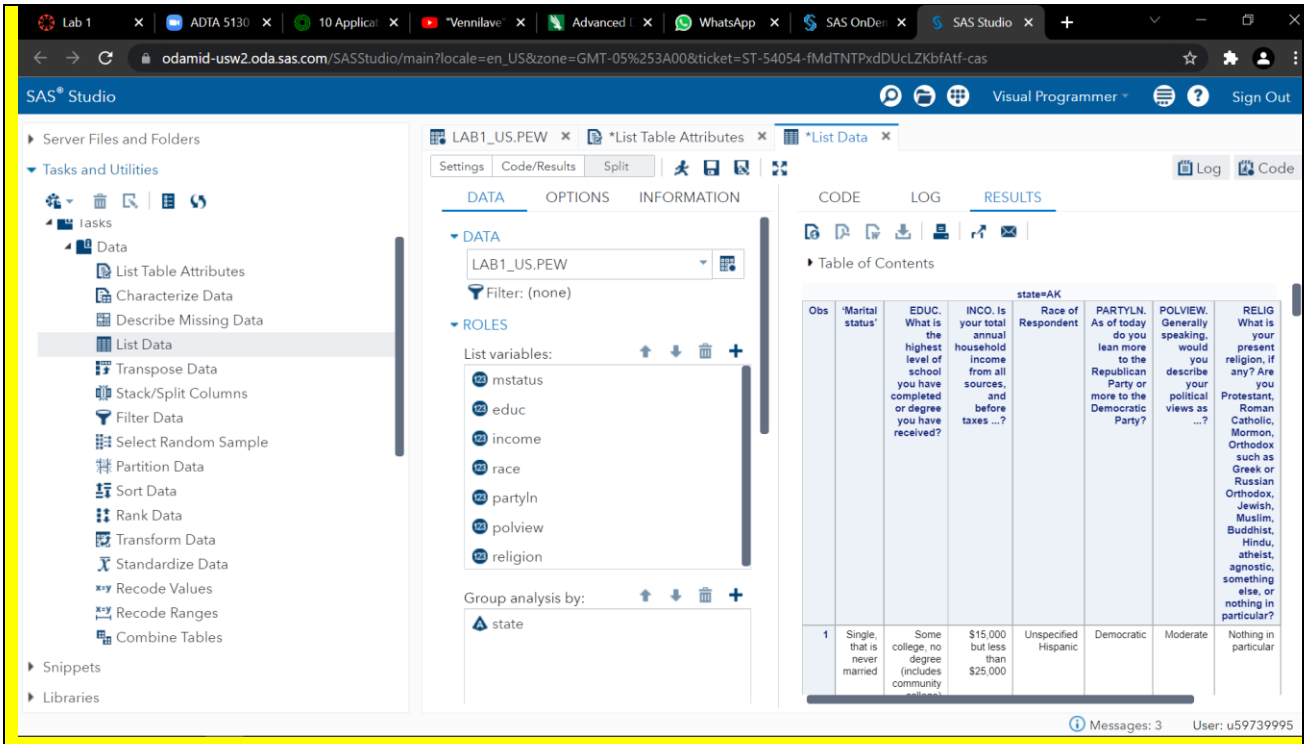
You can decide which variable to list and which variable to group by. For illustration purpose, make the selection of mstatus, educ, income, race, partyln, polview, religion grouped by state as below.



Click on Run  icon or press (F3) to answer the questions in Box 2.

BOX 2
Take a screenshot of all observations for variables mstatus, educ, income, race, partyln, polview, religion from Arkansas (state='AK') and insert below:

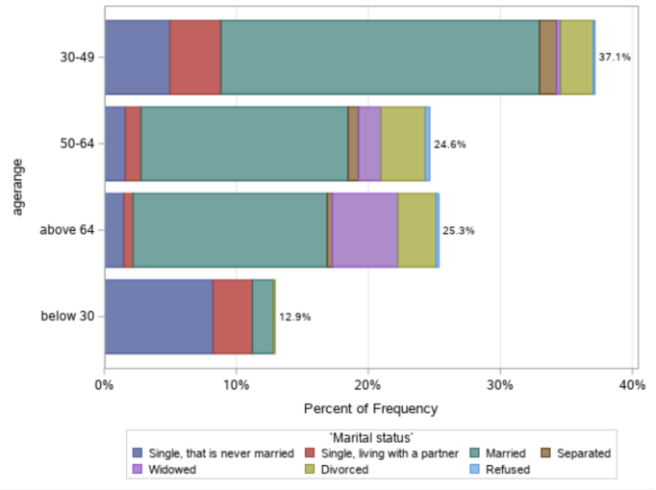
[Udayraj Suthapalli], lab1



For each variable below, indicate their data type as numerical (continuous or discrete) or categorical (nominal or ordinal)

Variable	Data Type
mstatus	Categorical nominal
educ	Categorical ordinal
income	Categorical ordinal
race	Categorical nominal
partyln	Categorical nominal
polview	Categorical ordinal
religion	Categorical nominal

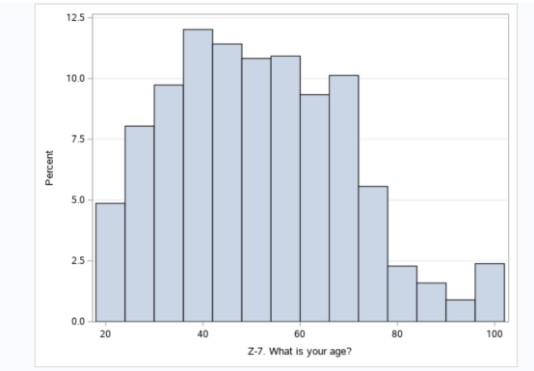
Now, we will replicate the following graph showing the marital status for different age groups.

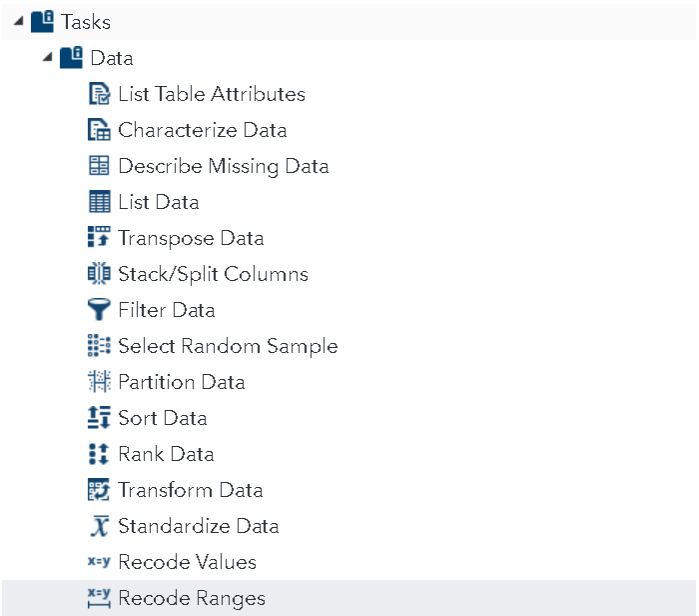


For this task, we will focus on the following two variables.

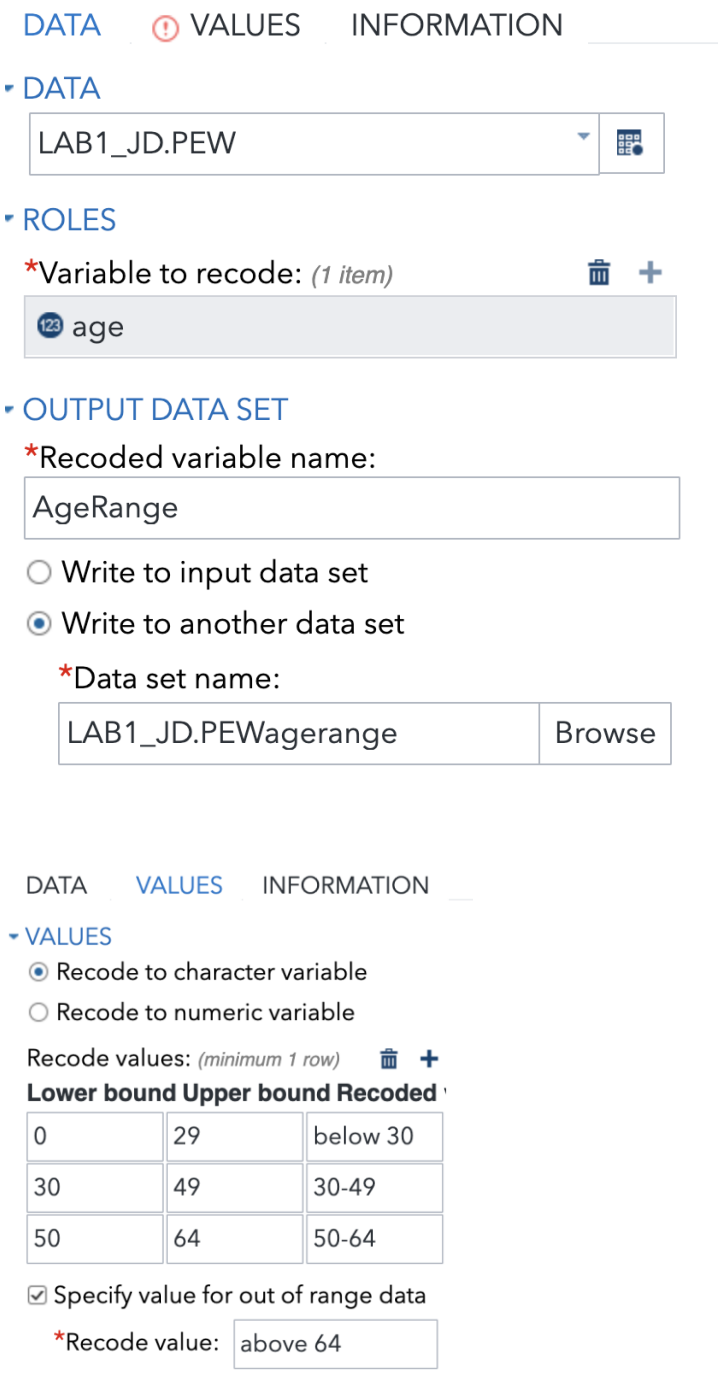
- **Variable 1:** response to the question 'What is your age?' (age)
- **Variable 2:** Marital status (mstatus)

The age variable is not a range variable originally. Therefore, we need to create a new range variable from Age using Recode Ranges under DATA Task.

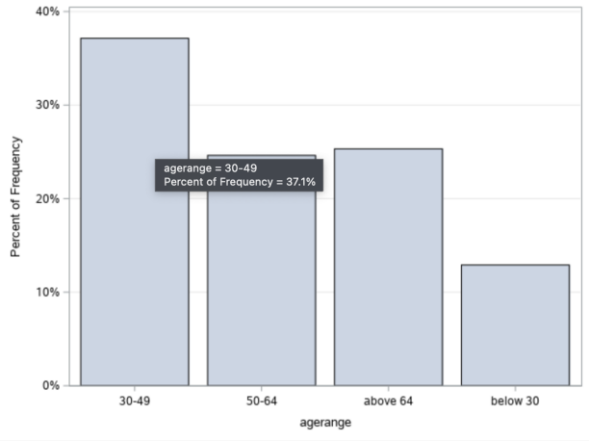




Use the following specifications to recode Age by the ranges specified and call it AgeRange.



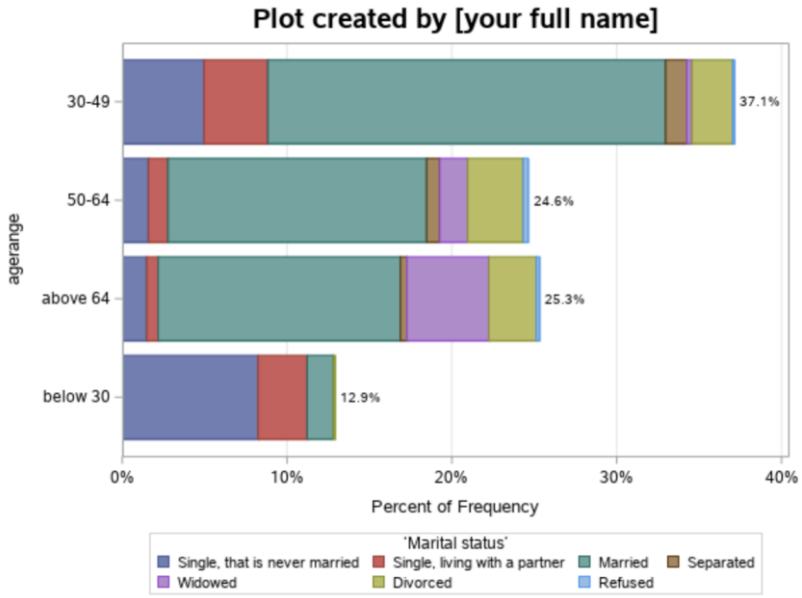
So far we created a new variable called AgeRange the distribution of which can be generated from bar chart using the measure of frequency percent.



mstatus is a categorical variable and possible values with frequency distribution can be shown using One-Way Frequencies under Statistics Tasks.

'Marital status'				
mstatus	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Single, that is never married	164	16.29	164	16.29
Single, living with a partner	88	8.74	252	25.02
Married	565	56.11	817	81.13
Separated	25	2.48	842	83.61
Widowed	70	6.95	912	90.57
Divorced	89	8.84	1001	99.40
Refused	6	0.60	1007	100.00

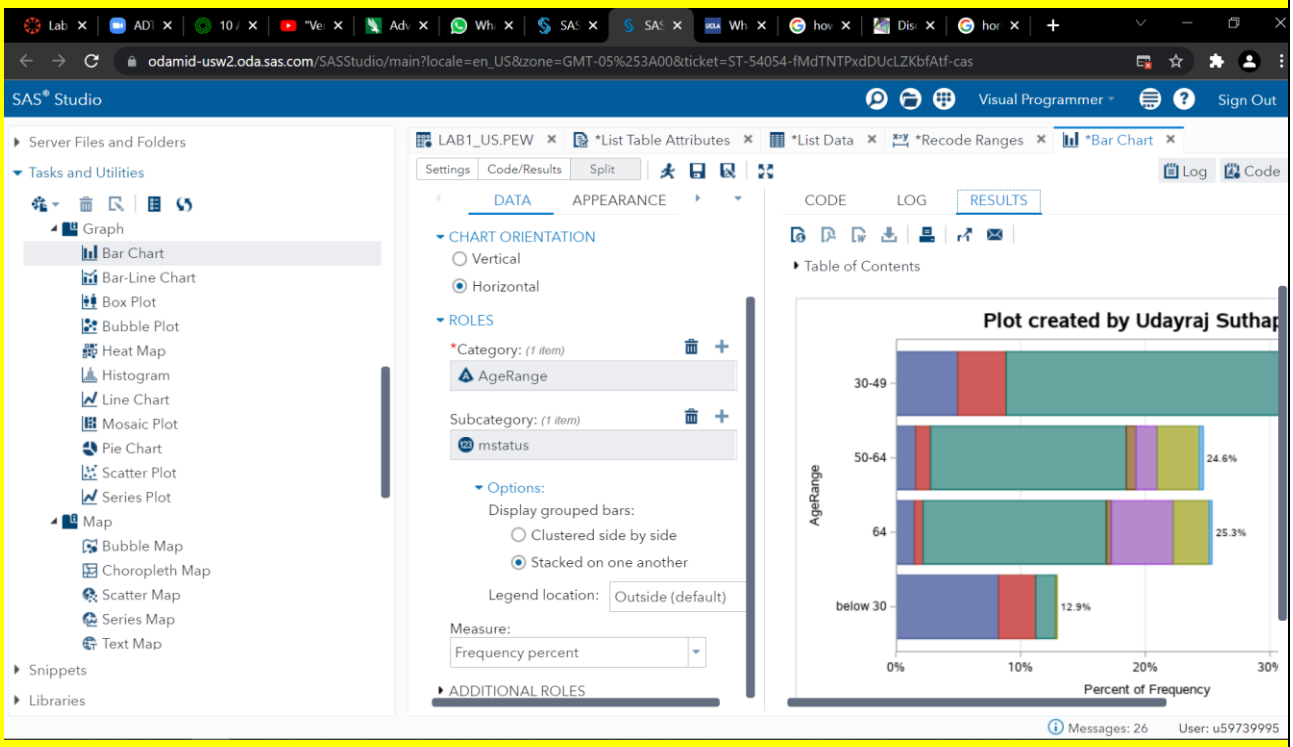
Please consider how to reproduce the following chart and insert your answer in box 3 below.



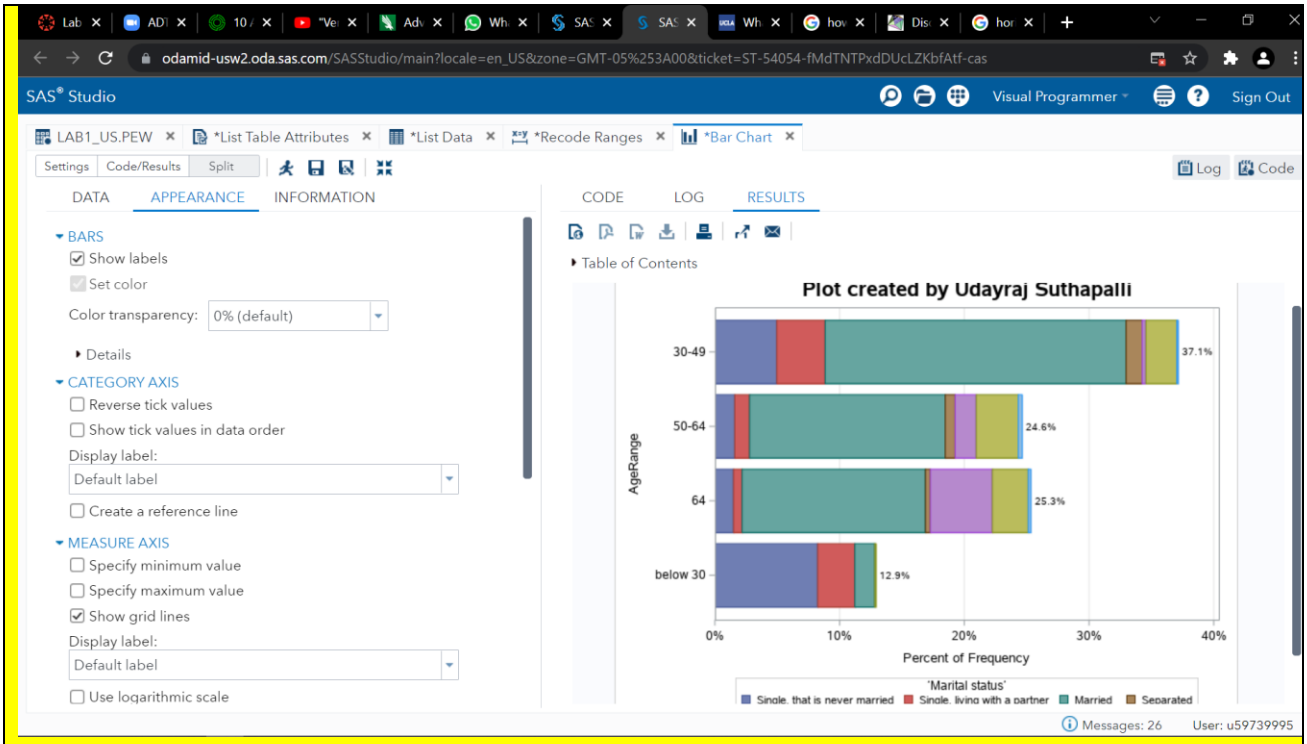
BOX 3

Insert your plot with your full name inside the bracket in the title.

Take a screenshot of the options you selected to obtain this plot.



[Udayraj Suthapalli], lab1



Survey question (not for credit): Overall, I spent __1__ hours in Lab 1.