ANLT5030 – Unit 6 Assignment 1 Tutorial

SAS Studio

Instructions

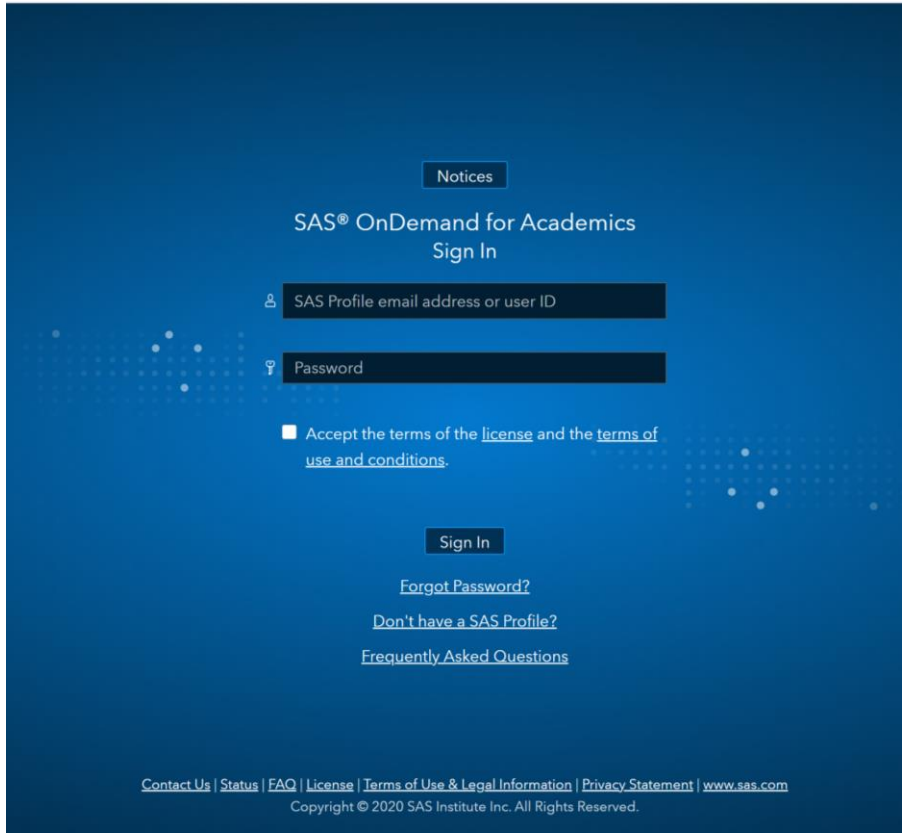
- For this assignment, you will be assessed on your use and interpretation of ANOVA through a case in your textbook. Read Case Problem 1, Wentworth Medical Center, on page 616 of your text, and download the accompanying data sets from CengageBrain to create the following in your report:
- Use descriptive statistics to summarize the data from the two studies.
- Provide your preliminary observations about the depression scores for both studies.
- Using analysis of variance (ANOVA) on both data sets, state the hypothesis being tested in each case.
- Provide your conclusions based on your ANOVA calculations.
- Use inferences about individual treatment means where appropriate and state your conclusions.
- Review the Analysis of Variance Scoring Guide for the detailed grading expectations of this assignment.

Dataset

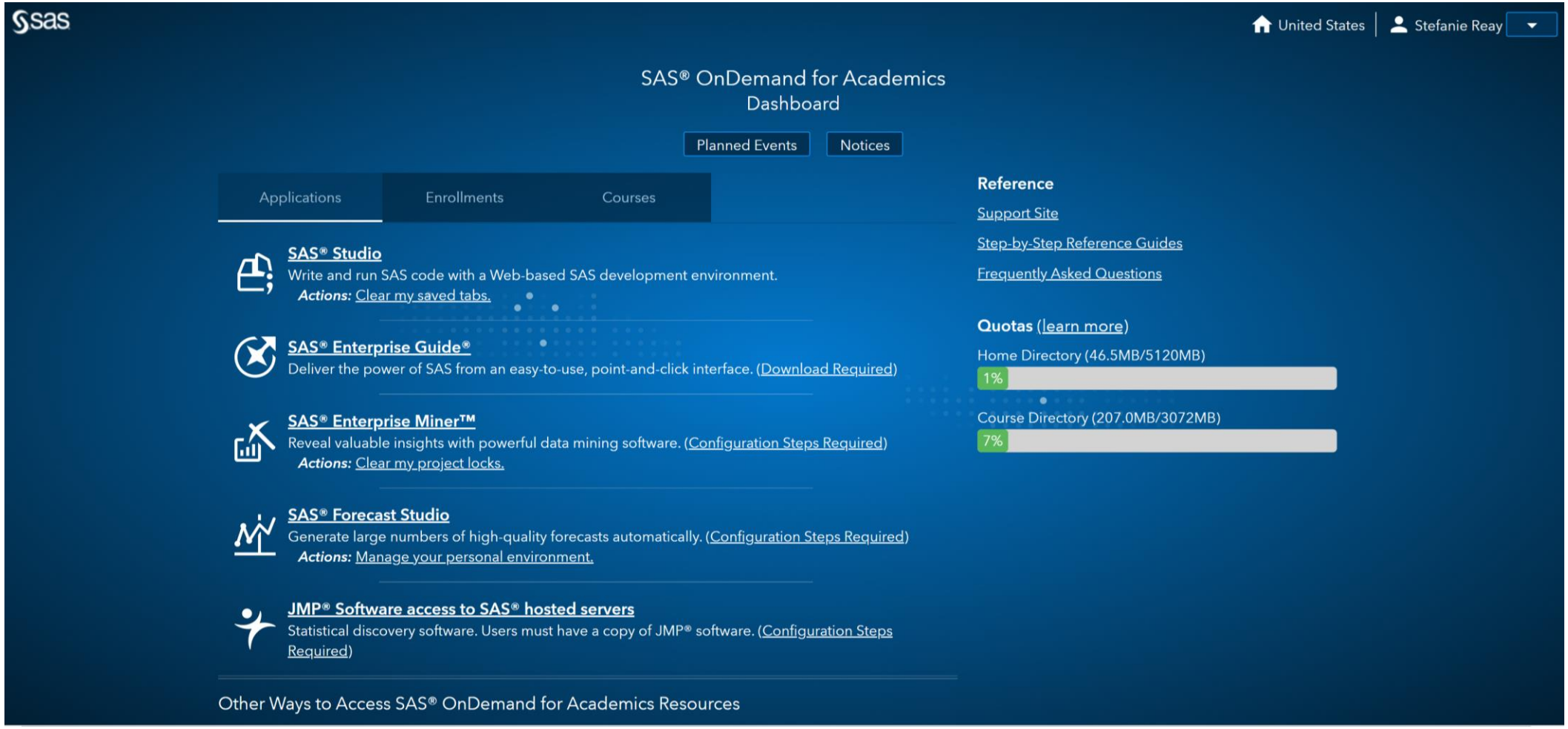
- Download the Medical1.xlsx and Medical2.xlsx files from the course datasets or from the Unit 6 Welcome announcement in the course announcements.

Access the SAS OnDemand for Academics Control Center

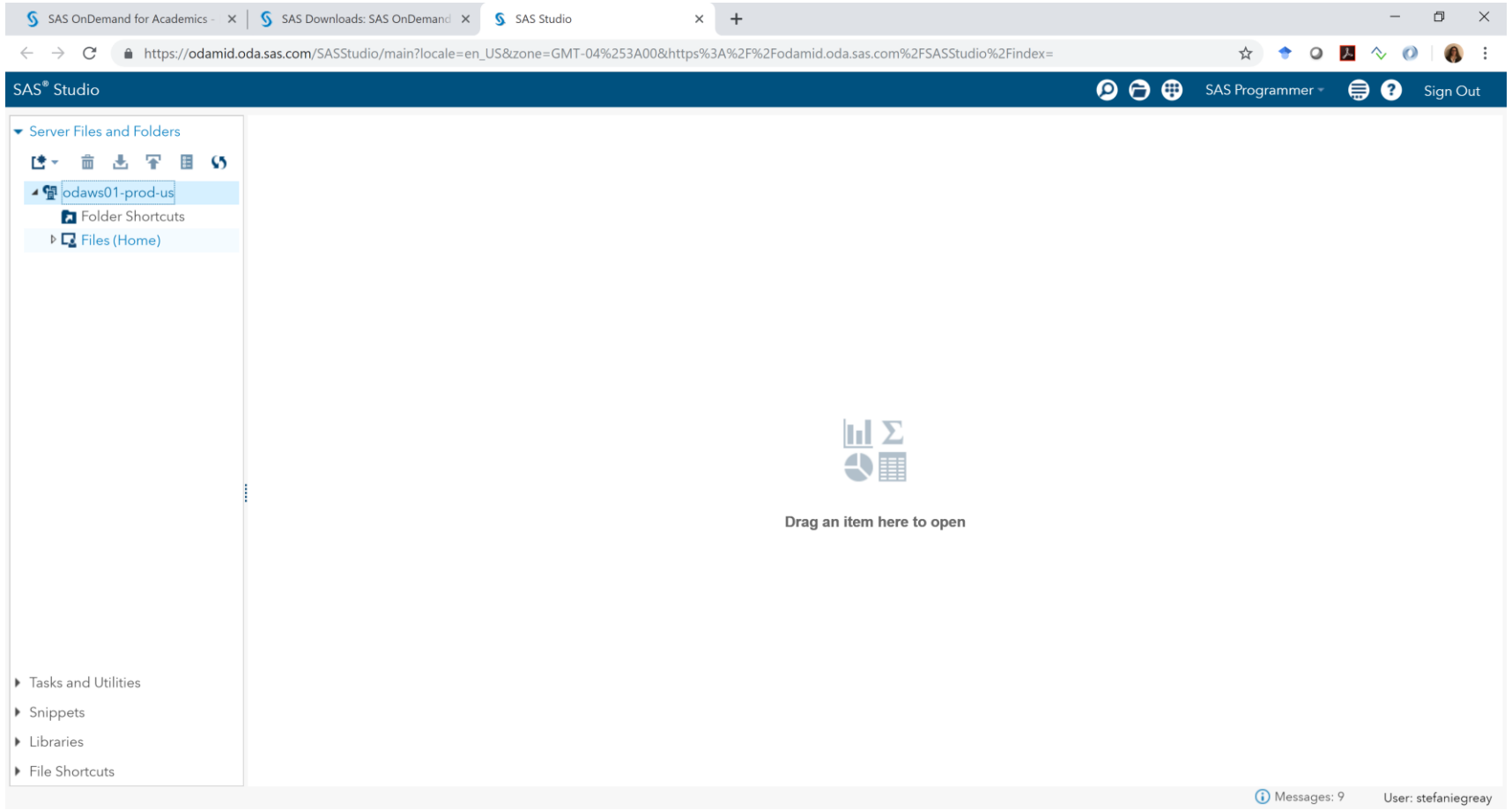
https://odamid.oda.sas.com/SASODAControlCenter



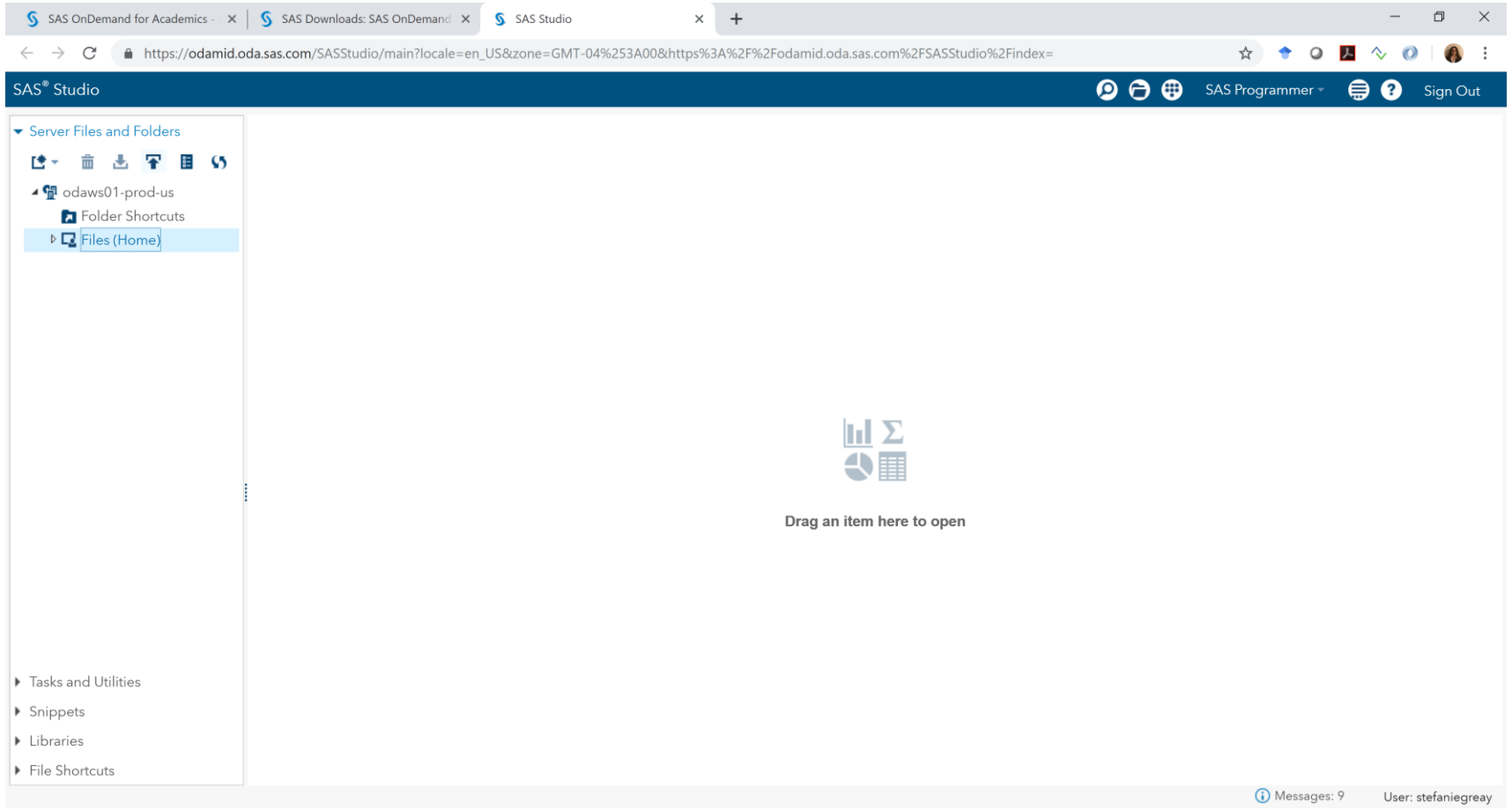
SAS OnDemand for Academics (SODA) Control Center



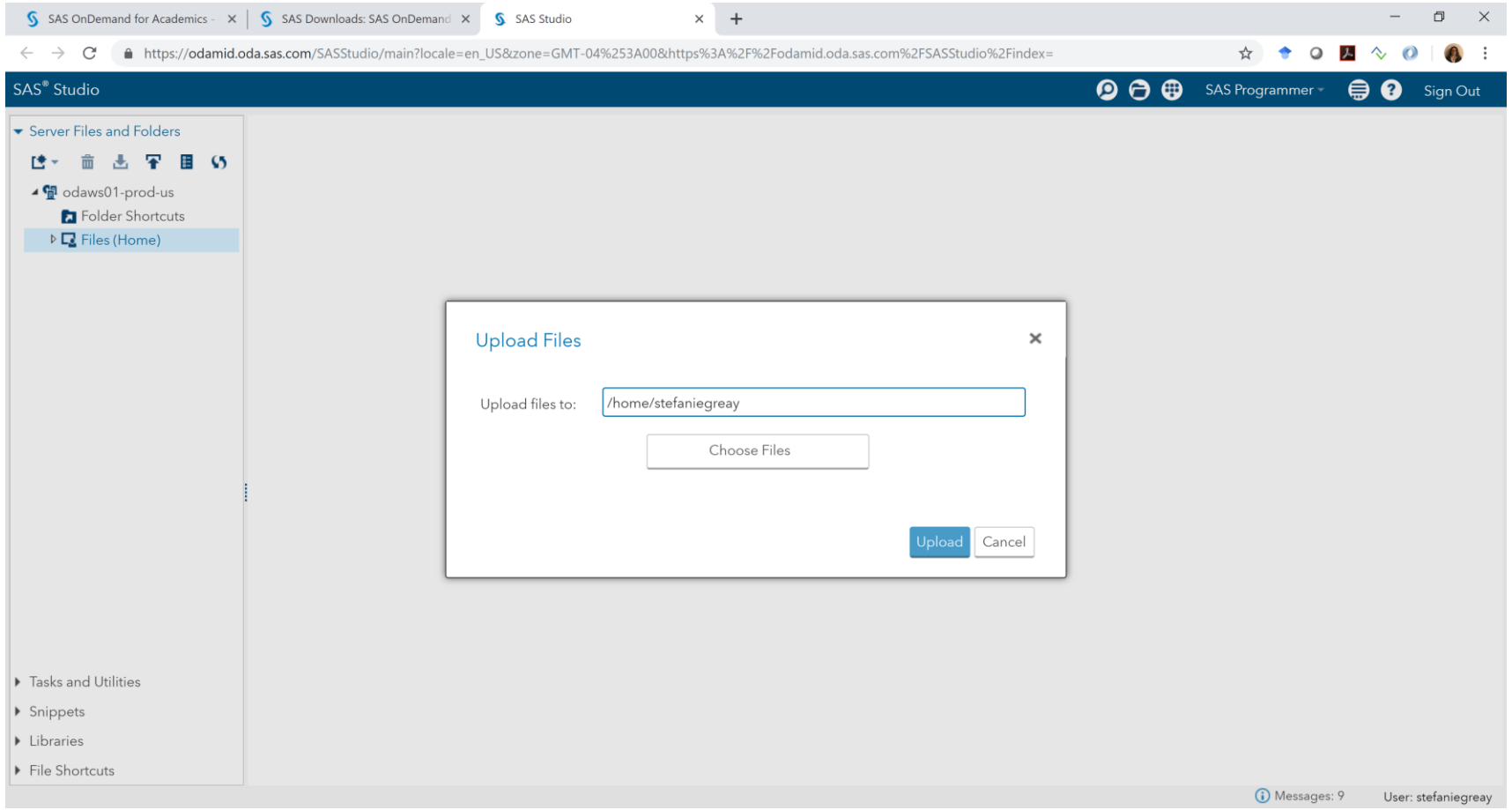
Click on Files(Home)



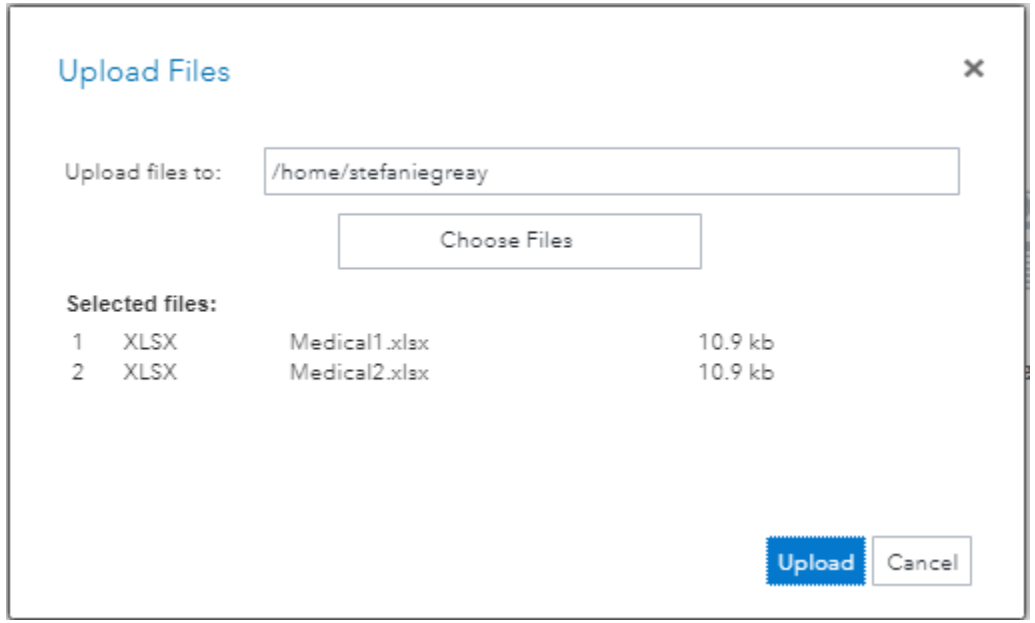
The Upload button will display in dark blue



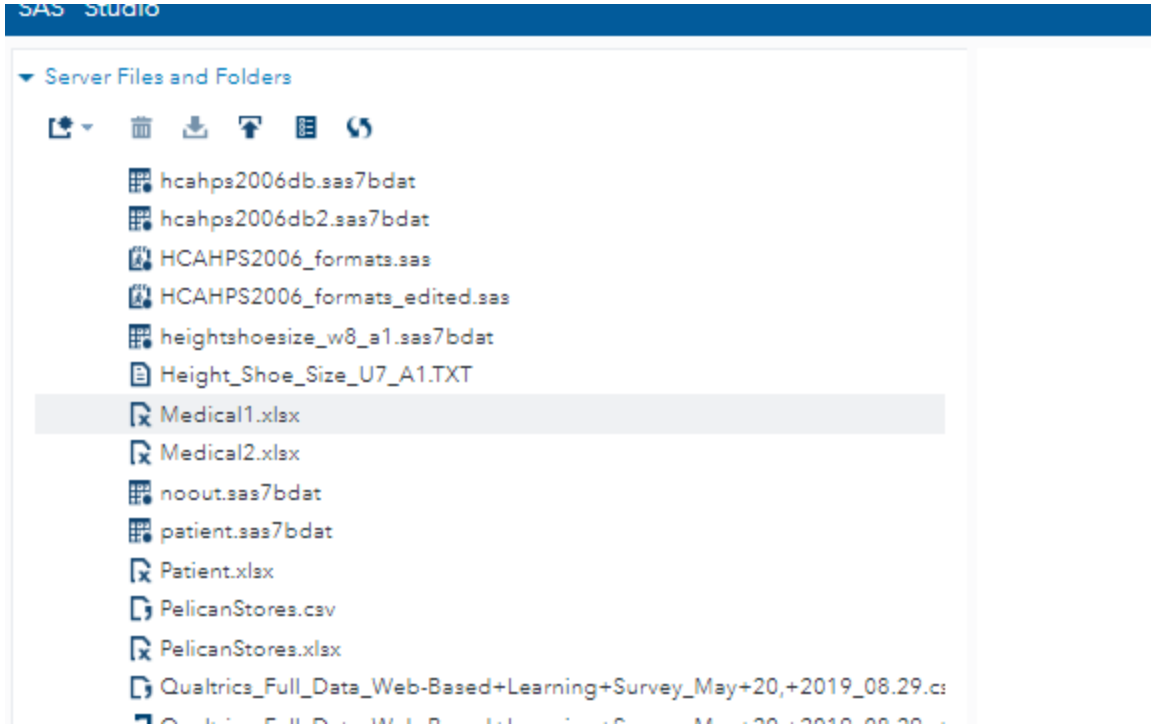
You can create a folder at this point, if you wish, or simply upload to your home directory.



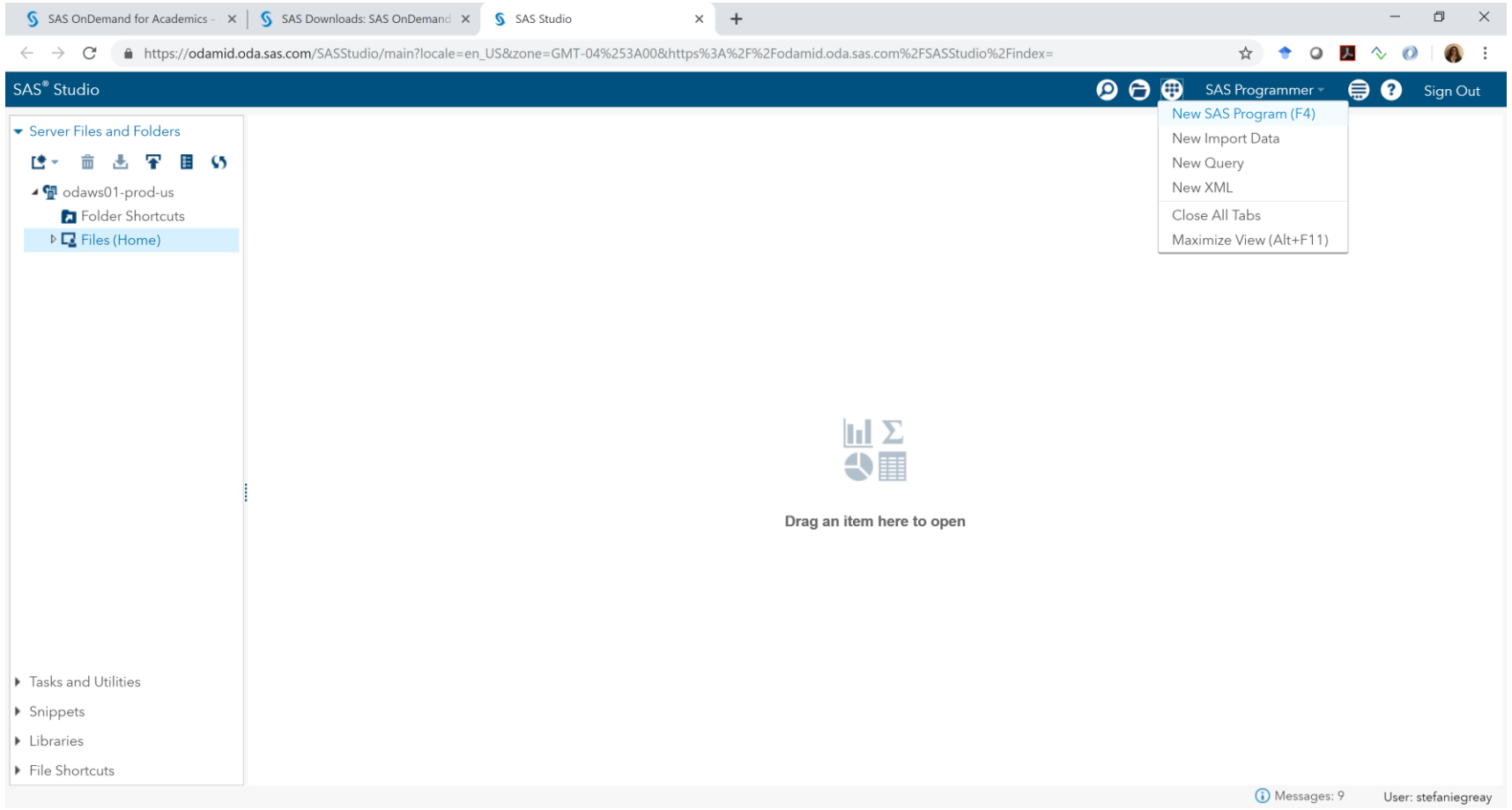
Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”



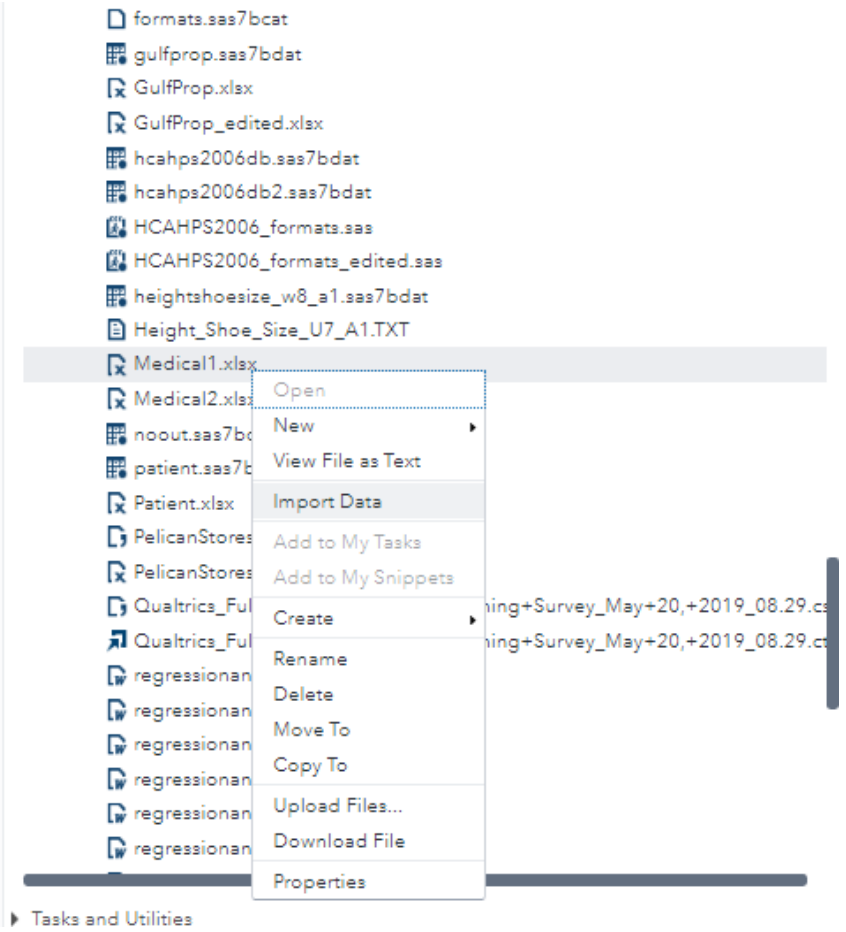
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



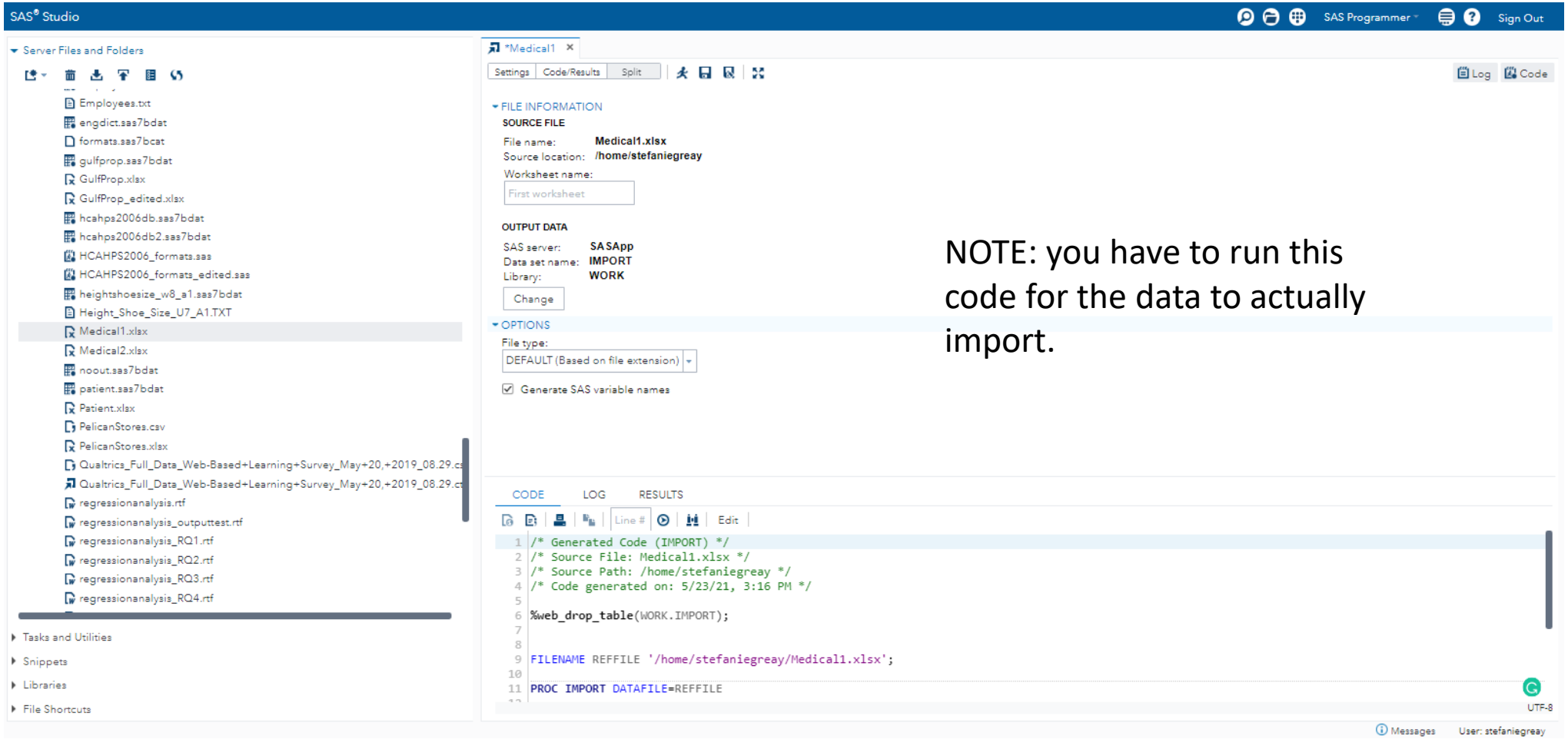
To get started with the SAS portion of the assignment, start a new SAS program.



Import the dataset into a SAS dataset format (from the current xlsx format)

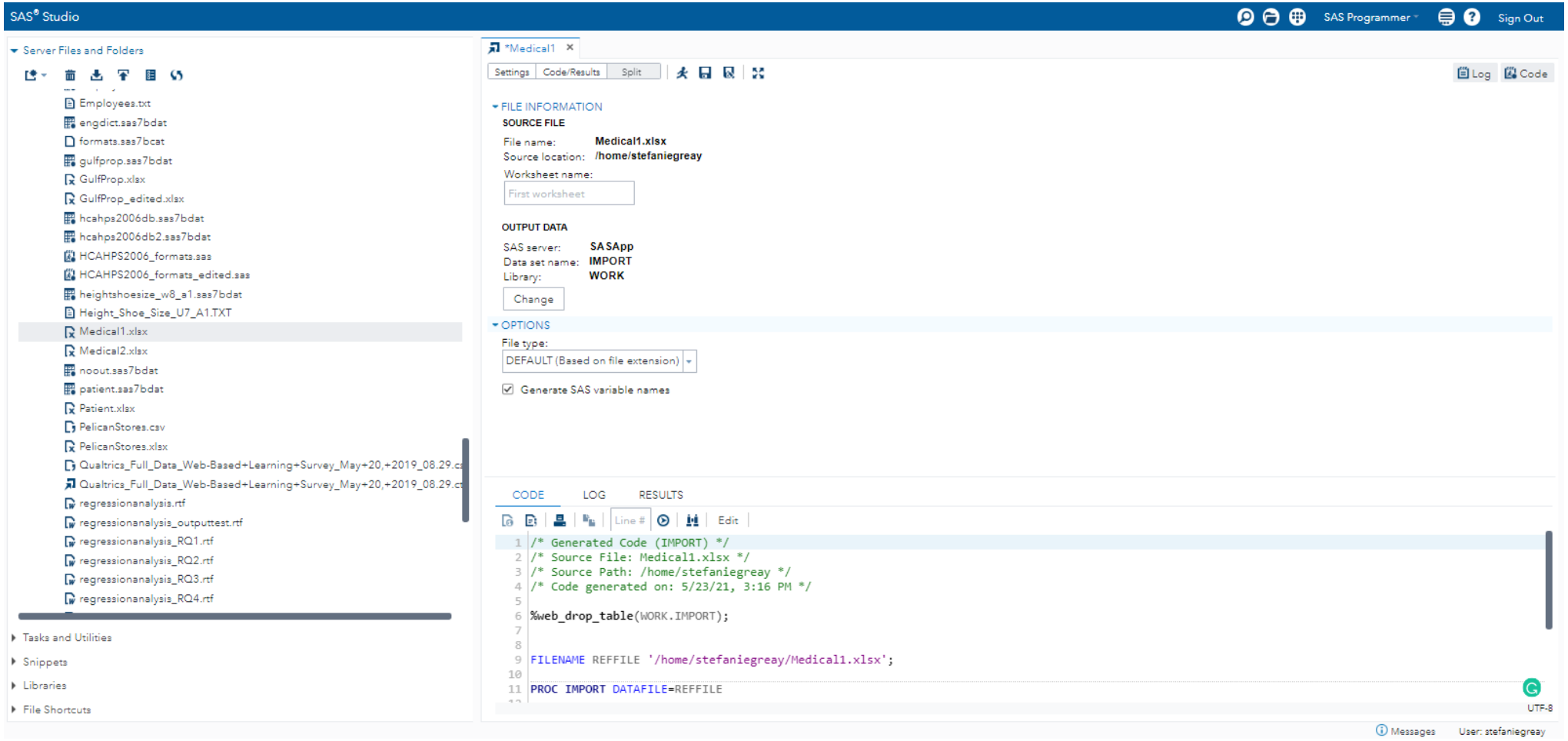


The Proc Import code will be written for you (save this as a template to use for future imports!)

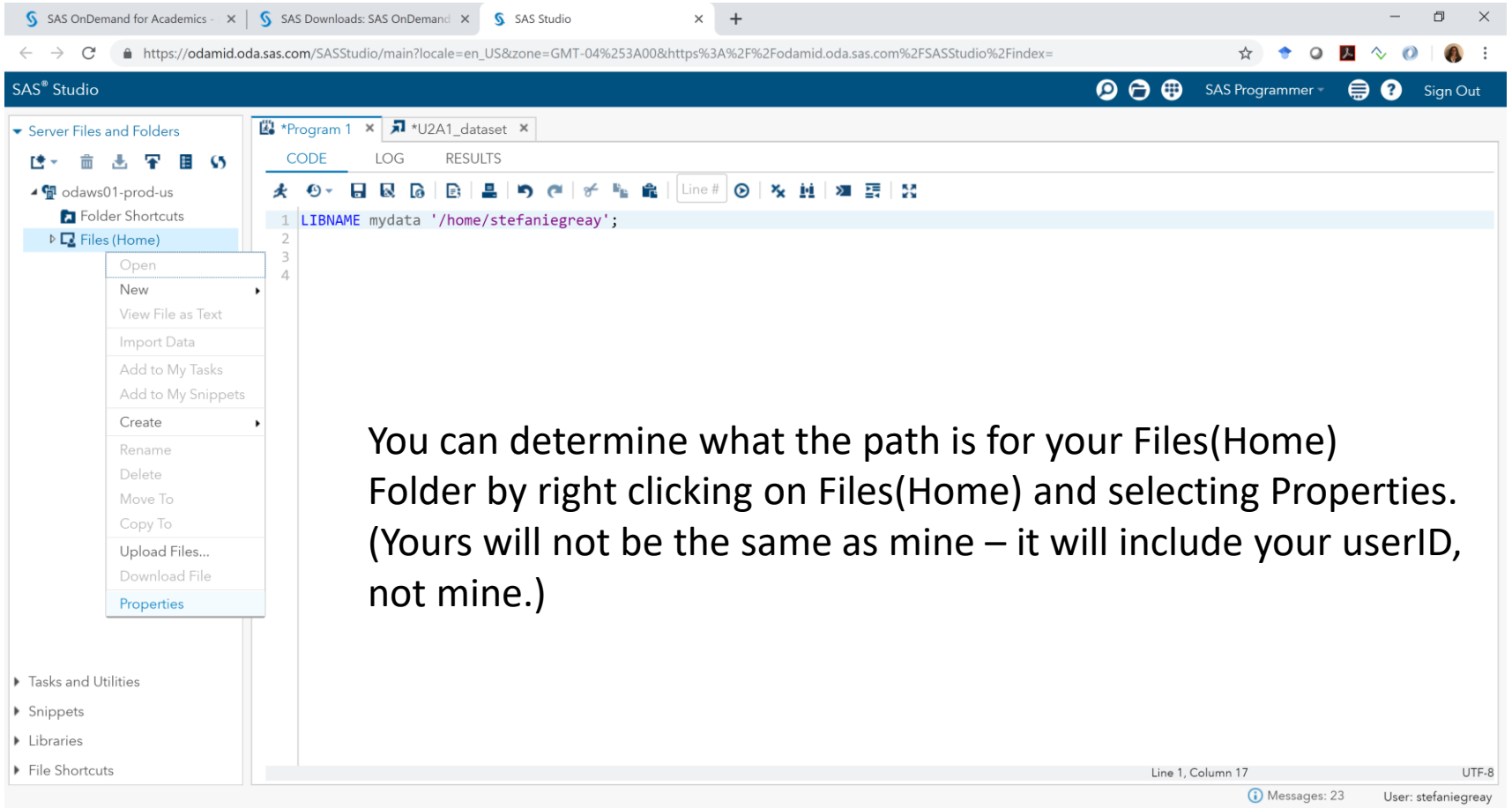


NOTE: you have to run this code for the data to actually import.

To run the code, click the icon that looks like a guy running.

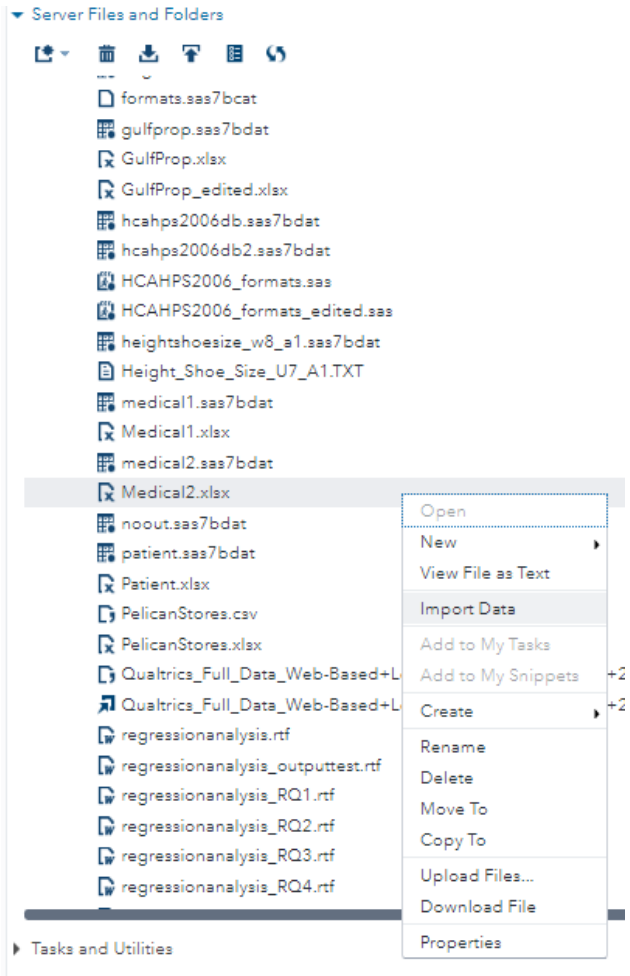


To create a SAS Library for your Files(Home) folder, you need to use a libname statement

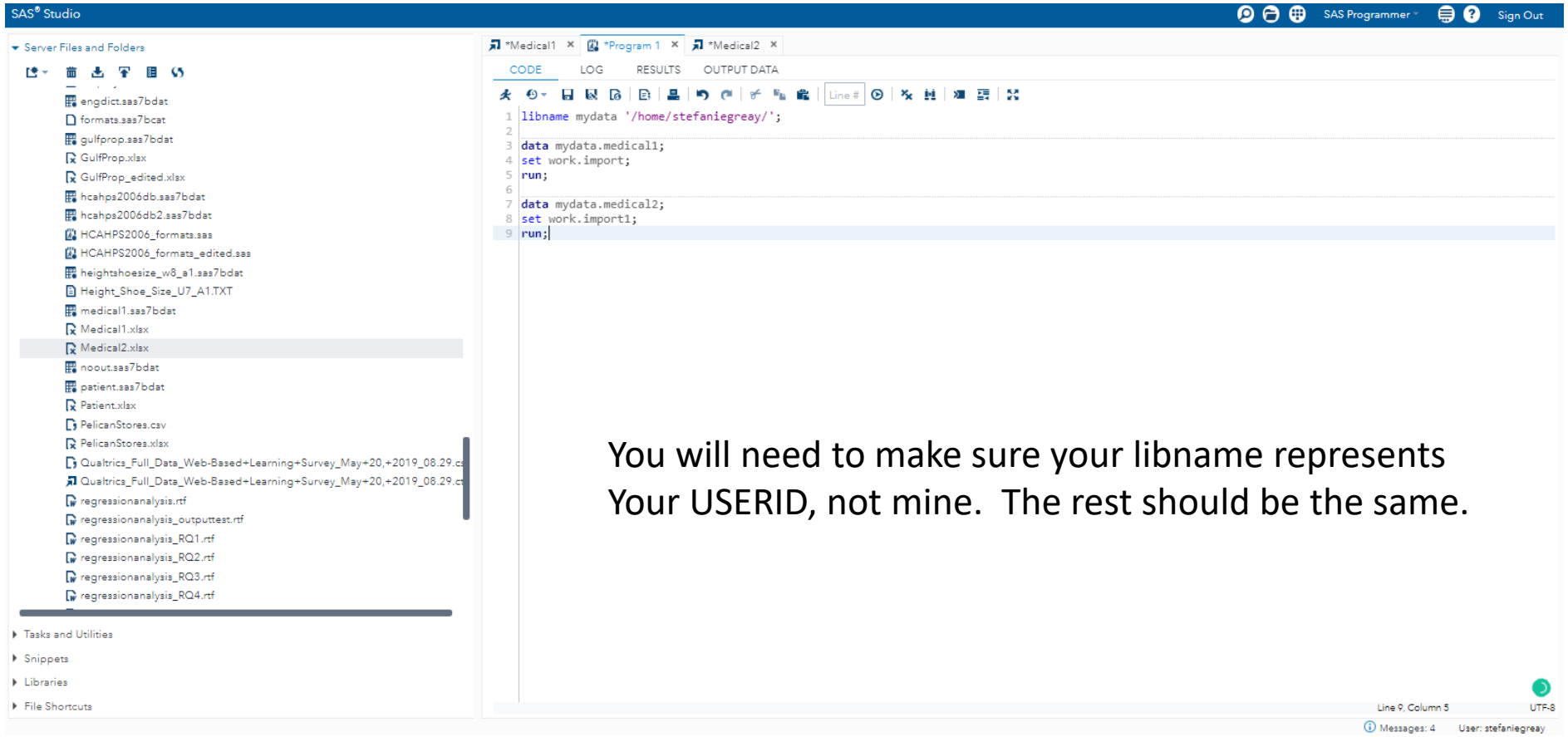


You can determine what the path is for your Files(Home) Folder by right clicking on Files(Home) and selecting Properties. (Yours will not be the same as mine – it will include your userID, not mine.)

Repeat the process for the second dataset.

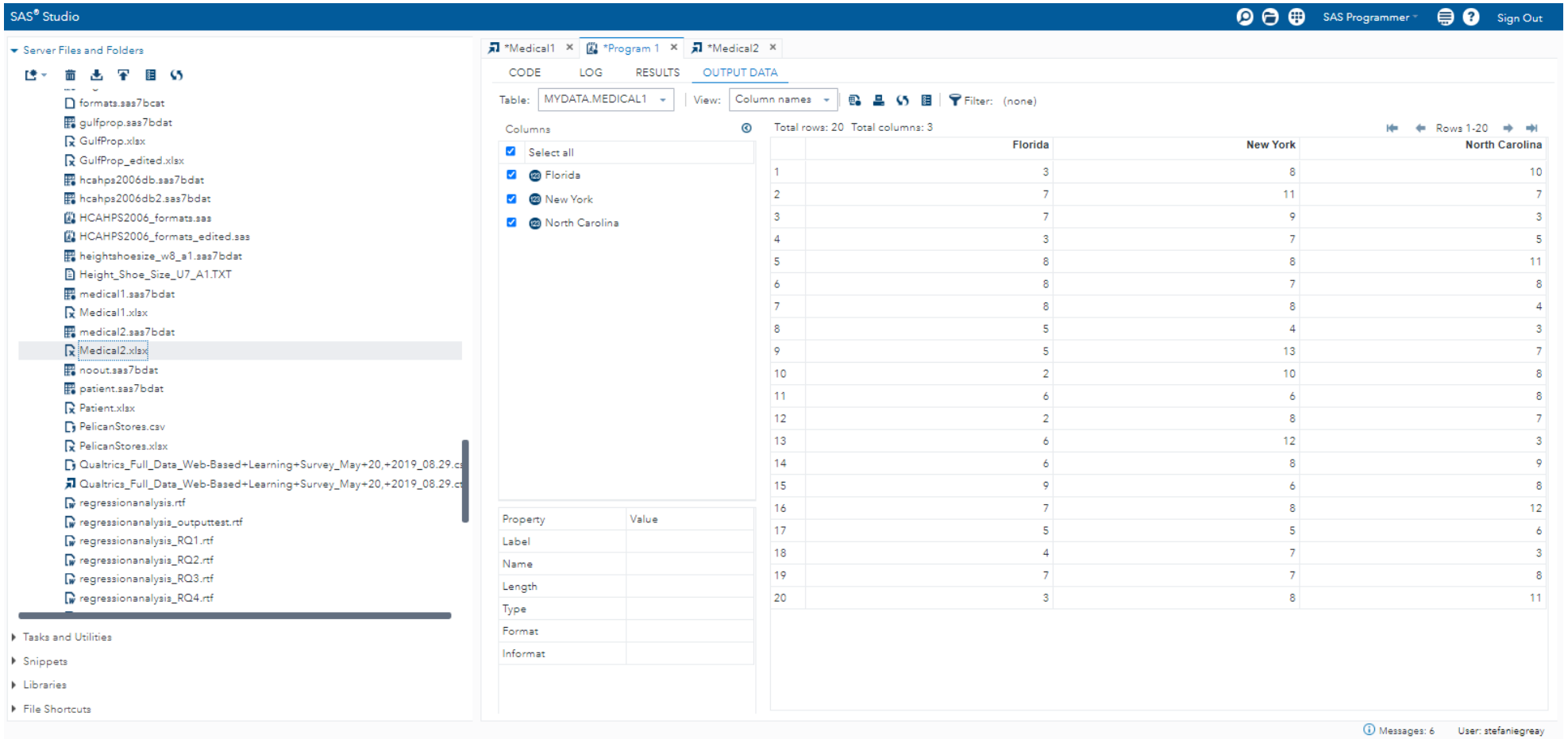


Save the temporary SAS datasets created by the import to your library using the following sample code.

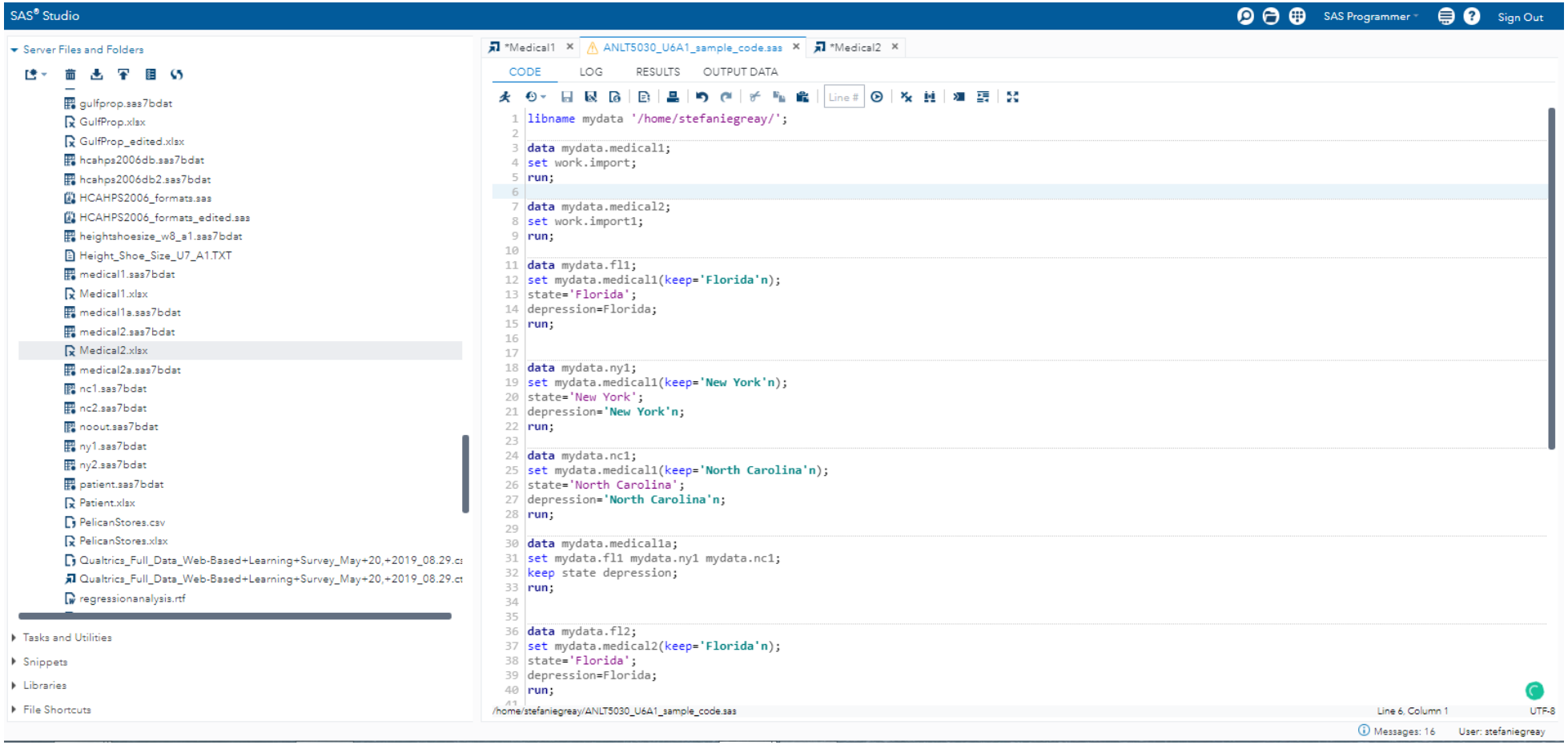


You will need to make sure your libname represents Your USERID, not mine. The rest should be the same.

When you run the code, you will see the dataset in the ouput data window and can verify its success.



You can now run any procedures against that dataset via the code window.



Sample Code for the required data manipulation portion of the assignment

```
libname mydata '/home/stefaniegreay/';

data mydata.medical1;
set work.import;
run;

data mydata.medical2;
set work.import1;
run;
```

```
data mydata.fl1;
set mydata.medical1(keep='Florida'n);
state='Florida';
depression=Florida;
run;


data mydata.ny1;
set mydata.medical1(keep='New York'n);
state='New York';
depression='New York'n;
run;

data mydata.nc1;
set mydata.medical1(keep='North Carolina'n);
state='North Carolina';
depression='North Carolina'n;
run;

data mydata.medical1a;
set mydata.fl1 mydata.ny1 mydata.nc1;
keep state depression;
run;
```

```
data mydata.fl2;
set mydata.medical2(keep='Florida'n);
state='Florida';
depression=Florida;
run;


data mydata.ny2;
set mydata.medical2(keep='New York'n);
state='New York';
depression='New York'n;
run;

data mydata.nc2;
set mydata.medical2(keep='North Carolina'n);
state='North Carolina';
depression='North Carolina'n;
run;


data mydata.medical2a;
set mydata.fl2 mydata.ny2 mydata.nc2;
keep state depression;
run;
```

Sample Code for the descriptive portion of this assignment

```
proc sort data=mydata.medical1a;
by state;
run;

proc univariate data=mydata.medical1a;
by state;
var depression;
title 'Descriptive Summary of Depression by State Medical1.xlsx';
run;


proc sort data=mydata.medical2a;
by state;
run;

proc univariate data=mydata.medical2a;
by state;
var depression;
title 'Descriptive Summary of Depression by State Medical2.xlsx';
run;
```

Sample Code for the ANOVA portion of this assignment

```
proc anova data=mydata.medical1a;
class state;
model depression = state;
means state/tukey;
title 'ANOVA of Depression by State Medical1.xlsx';
run;

proc anova data=mydata.medical2a;
class state;
model depression = state;
means state/tukey;
title 'ANOVA of Depression by State Medical2.xlsx';
run;
```

Additional Resources for ANOVA and SAS's Proc ANOVA

- SAS's Proc ANOVA documentation:
  https://support.sas.com/documentation/onlinedoc/stat/131/anova.pdf

- Laerd Statistics One-Way ANOVA Explanation:
  https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php

Assumptions for ANOVA Hypothesis tests

Assumptions
1) The responses for each category (here each state) have a normal population distribution.
2) These distributions have the same variance.
3) The observations are independent.

5 Steps of a Hypothesis Test

- Hypotheses
  - (null and alternative hypothesis)
- Test Statistic
  - (from software output)
- P-value (or Rejection Region)
  - (from software output)
- Result
  - (reject or fail to reject the null hypothesis)
- Conclusion
  - (result written in terms of claim)

Step 1: ANOVA Hypotheses

- H_o: $\mu_{FL} = \mu_{NY} = \mu_{NC}$
- H_a: at least one of the means ($\mu_{FL}, \mu_{NY}, \mu_{NC}$) does not equal the others.

In Words:

- H_o: all of the means are equal
- H_a: at least one of the means does not equal the others

Step 2: Test Statistic and 3: P-value

- Step 2: Test Statistic
  - This is the “F-Value” shown in the ANOVA output from the software
- Step 3: P-value
  - This is the “Pr>F” value shown in the ANOVA output from the software

Step 4: Result

p-value comparison	Result
If $p \leq \alpha$	Reject Ho
If $p > \alpha$	Fail to Reject Ho

Step 5: Conclusion

Result	Evidence in favor of Ha/claim?	Statistical Significance
Reject Ho	Sufficient Evidence (to suggest claim shown in Ha may be valid)	Statistically Significant
Fail to Reject Ho	Insufficient Evidence (or not sufficient evidence) (to suggest claim shown in Ha may be valid)	Statistically Insignificant (or not statistically significant)

Post-Hoc pairwise comparison tests (i.e. Tukey Tests)

If we end up with a significant difference (i.e. we reject Ho for the ANOFA F-test), this tells us that at least one of the means is significantly different than the others, but it does not tell us which of them is or are significantly different.

At this point, the next step is to conduct pairwise comparisons to see which of the means differ from the others. This is essentially conducting a two-sample t-test for the difference in population means, but doing it for each pair of means/populations. A common approach here is the Tukey test (which compares each pair and the SAS Proc ANOVA output represents it in a simple, easy to interpret graphical output). These would need to be discussed and interpreted as separate hypothesis tests, just like the ANOVA F-test itself.