

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.



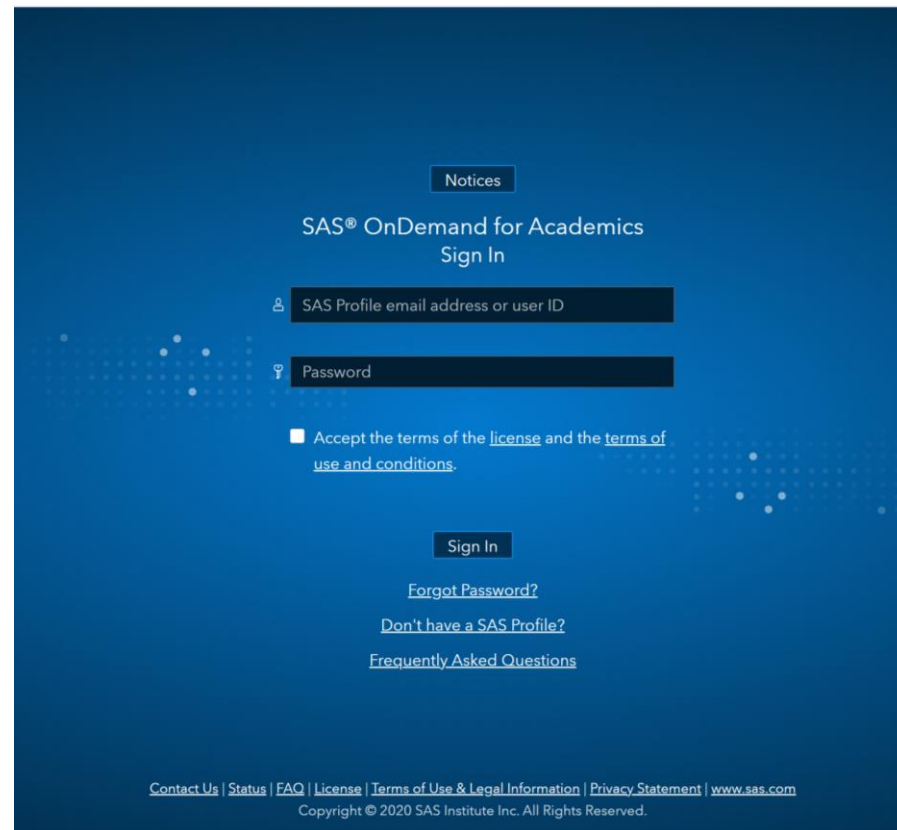
Dataset

- 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>



The screenshot shows the SAS OnDemand for Academics Sign In page. The background is a dark blue gradient with a subtle pattern of white dots. At the top, there is a "Notices" button. Below it, the text "SAS® OnDemand for Academics" and "Sign In" are displayed. The sign-in form includes two input fields: "SAS Profile email address or user ID" and "Password". Below the password field, there is a checkbox labeled "Accept the terms of the [license](#) and the [terms of use and conditions](#)." A "Sign In" button is positioned below the checkbox. At the bottom of the form, there are three links: "Forgot Password?", "Don't have a SAS Profile?", and "Frequently Asked Questions". The footer contains a row of links: "Contact Us", "Status", "FAQ", "License", "Terms of Use & Legal Information", "Privacy Statement", and "www.sas.com", followed by the copyright notice "Copyright © 2020 SAS Institute Inc. All Rights Reserved."



SAS OnDemand for Academics (SODA) Control Center

The screenshot displays the SAS OnDemand for Academics (SODA) Control Center dashboard. At the top, the SAS logo is on the left, and the user's location (United States) and name (Stefanie Reay) are on the right. The main heading is "SAS® OnDemand for Academics Dashboard". Below this, there are tabs for "Planned Events" and "Notices". A navigation bar includes "Applications", "Enrollments", and "Courses". The "Applications" tab is active, showing a list of SAS products:

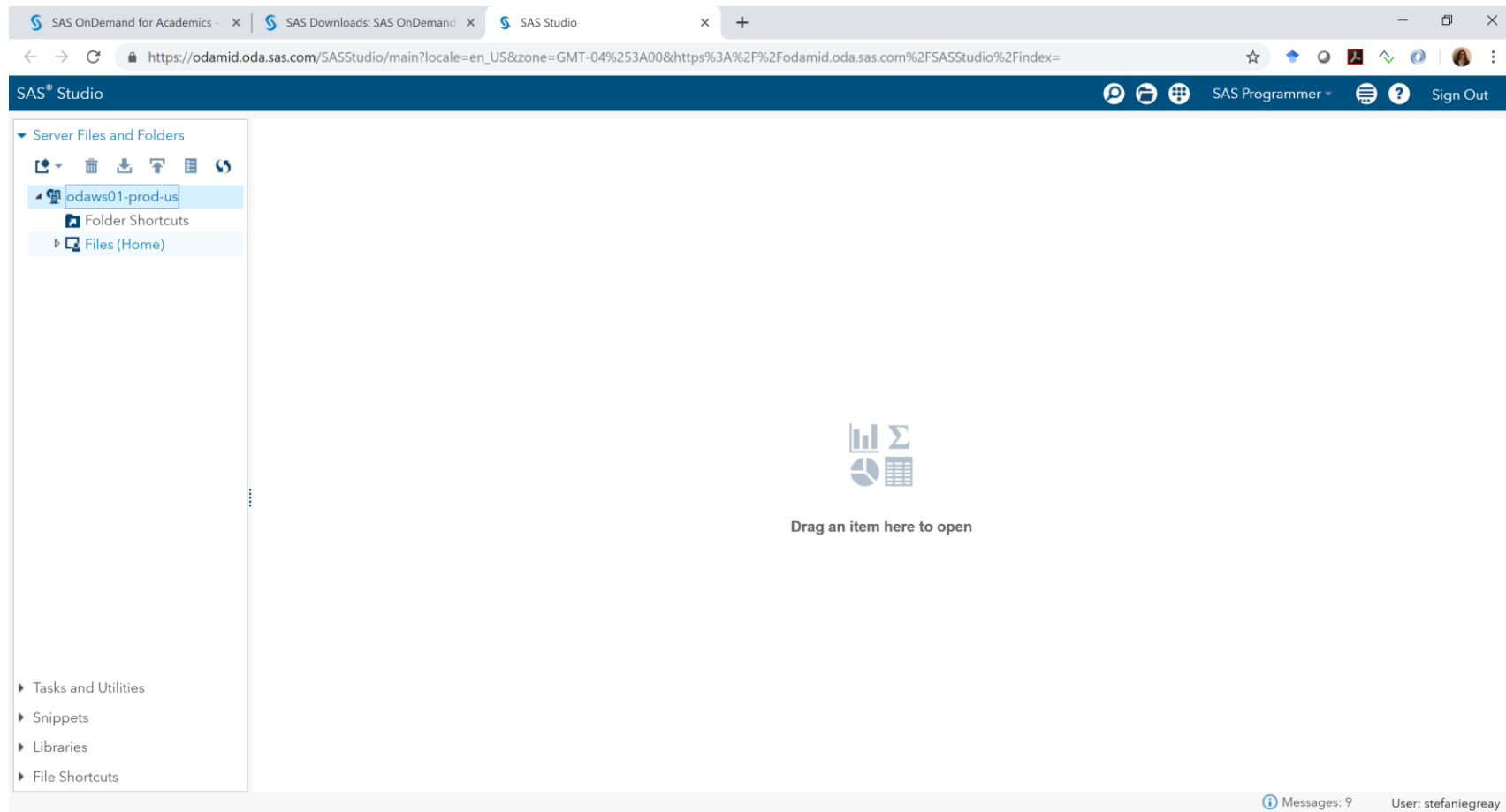
- SAS® Studio**: Write and run SAS code with a Web-based SAS development environment. *Actions: [Clear my saved tabs.](#)*
- SAS® Enterprise Guide®**: Deliver the power of SAS from an easy-to-use, point-and-click interface. ([Download Required](#))
- SAS® Enterprise Miner™**: Reveal valuable insights with powerful data mining software. ([Configuration Steps Required](#)) *Actions: [Clear my project locks.](#)*
- SAS® Forecast Studio**: Generate large numbers of high-quality forecasts automatically. ([Configuration Steps Required](#)) *Actions: [Manage your personal environment.](#)*
- JMP® Software access to SAS® hosted servers**: Statistical discovery software. Users must have a copy of JMP® software. ([Configuration Steps Required](#))

On the right side, there is a "Reference" section with links to the [Support Site](#), [Step-by-Step Reference Guides](#), and [Frequently Asked Questions](#). Below this is a "Quotas (learn more)" section showing progress bars for "Home Directory (46.5MB/5120MB)" at 1% and "Course Directory (207.0MB/3072MB)" at 7%.

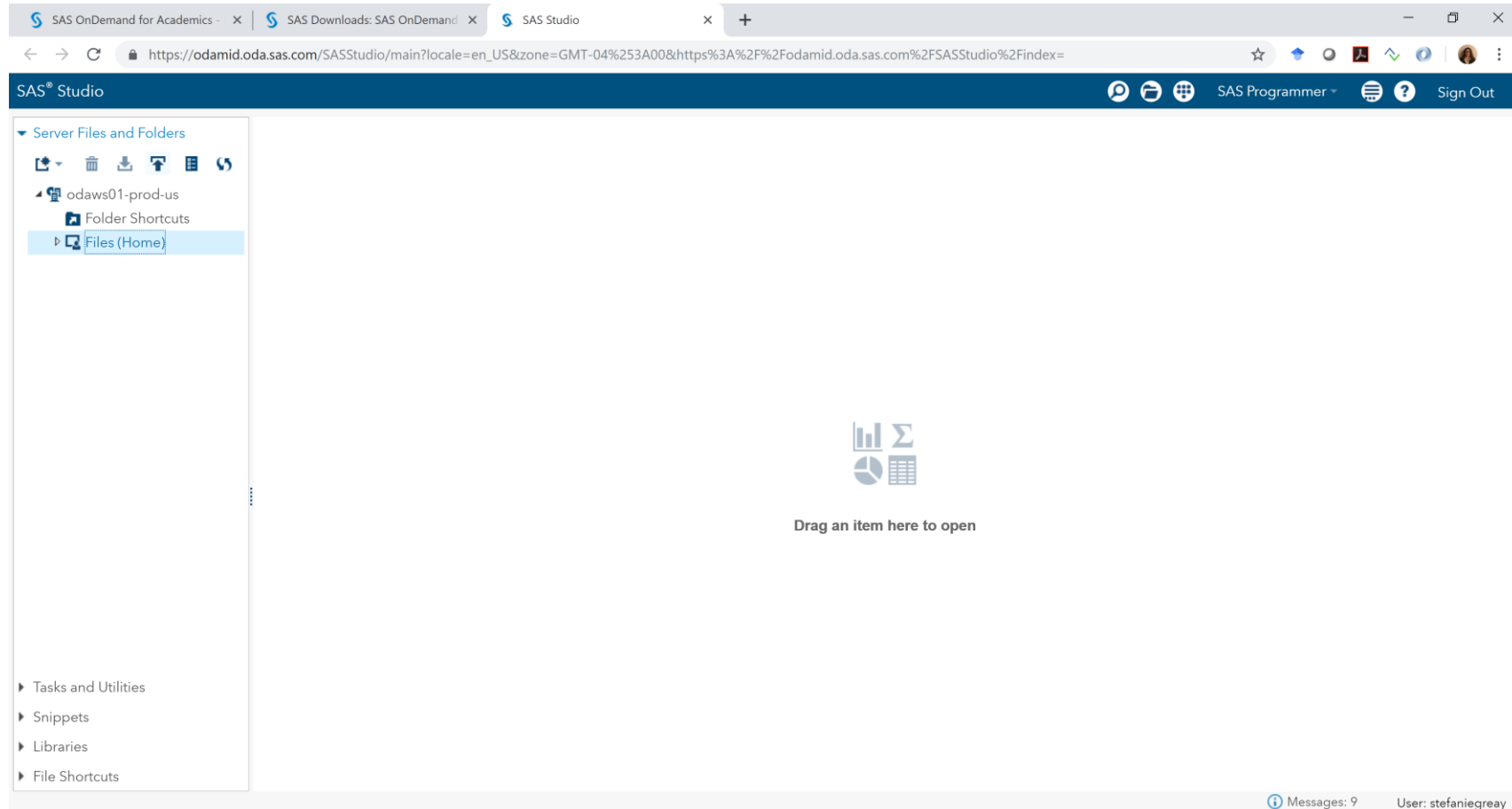
At the bottom, there is a link for "Other Ways to Access SAS® OnDemand for Academics Resources".



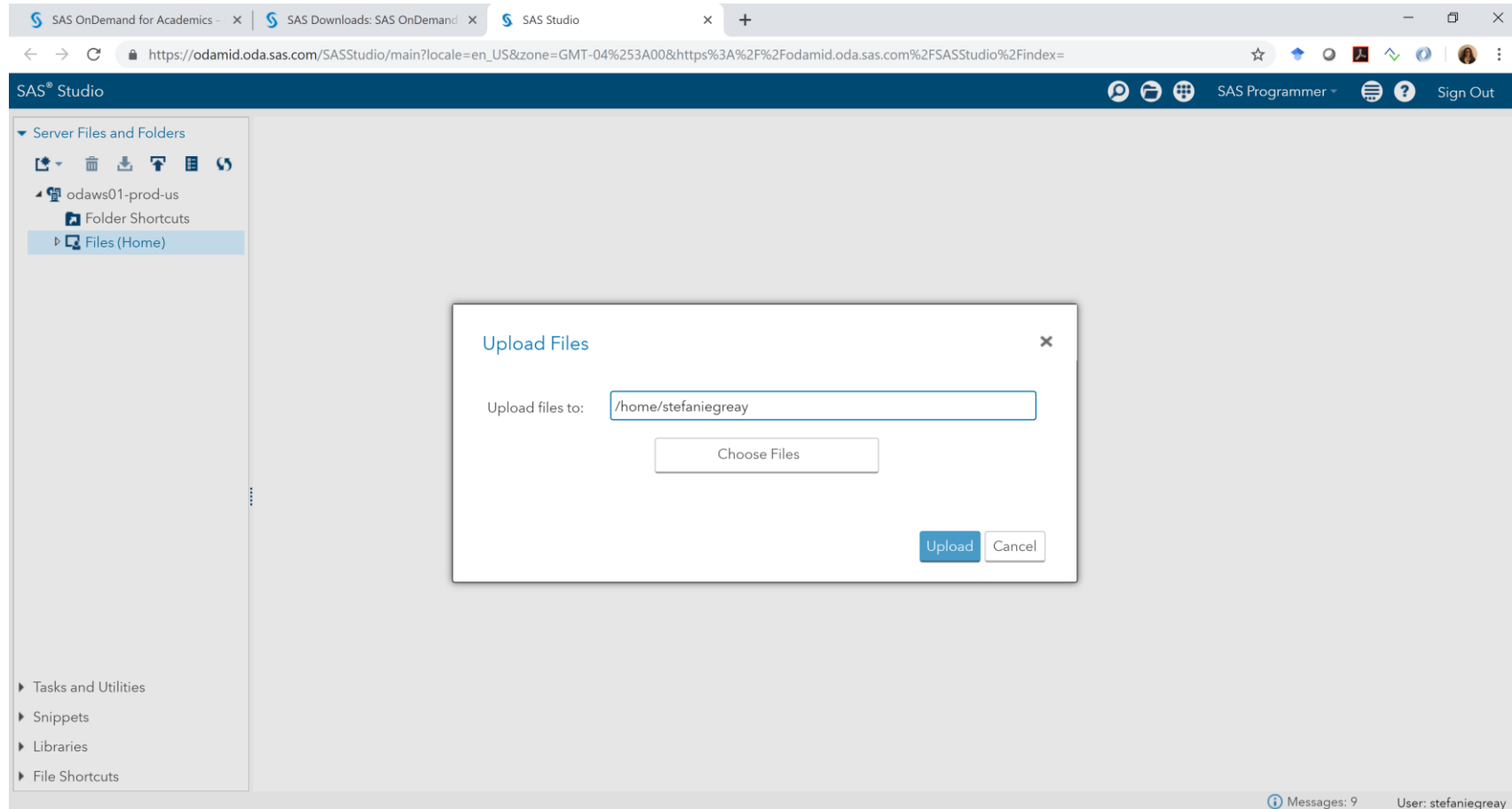
Click on Files(Home)



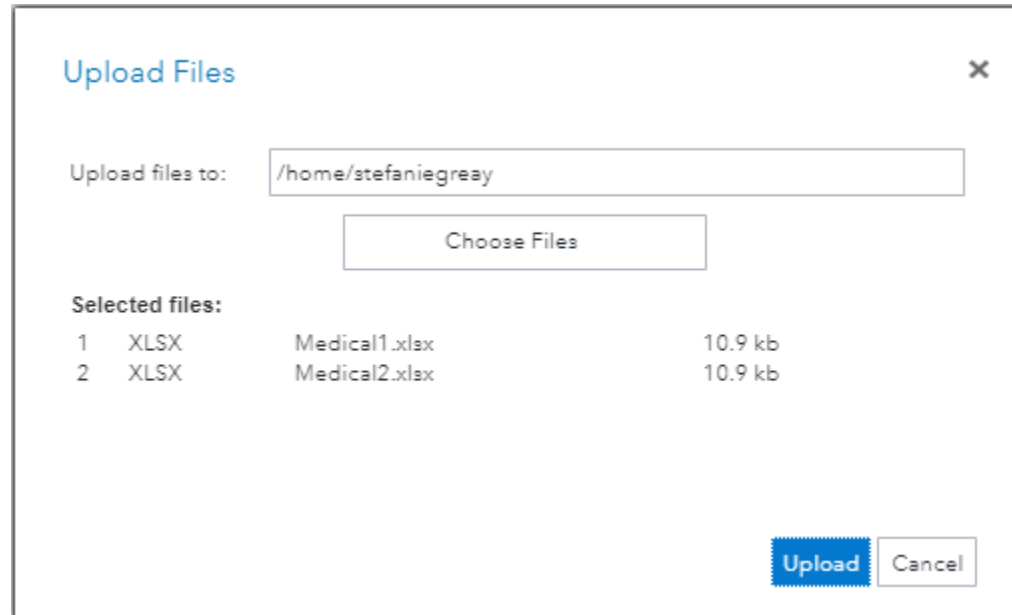
The Upload button will display in dark blue



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.”

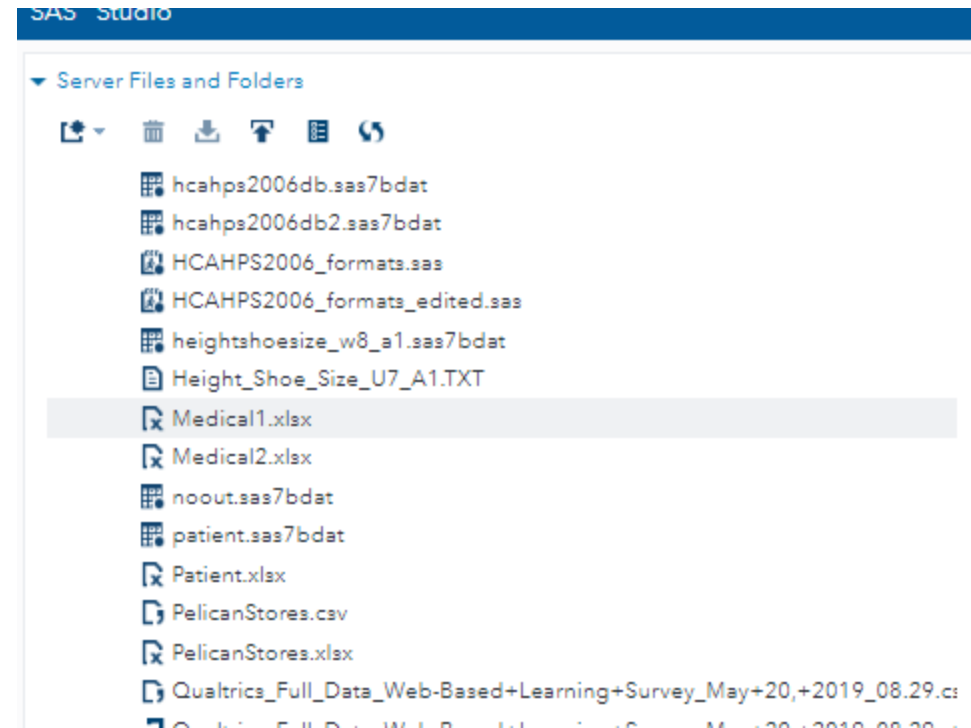


The screenshot shows a web-based 'Upload Files' dialog box. At the top, the title 'Upload Files' is in blue, followed by a close button (X). Below the title, there is a text input field labeled 'Upload files to:' containing the path '/home/stefaniegreay'. Underneath this field is a button labeled 'Choose Files'. Below the 'Choose Files' button, there is a section titled 'Selected files:' which contains a table of two files. The first file is 'Medical1.xlsx' (10.9 kb) and the second is 'Medical2.xlsx' (10.9 kb). At the bottom right of the dialog box, there are two buttons: 'Upload' (highlighted in blue) and 'Cancel'.

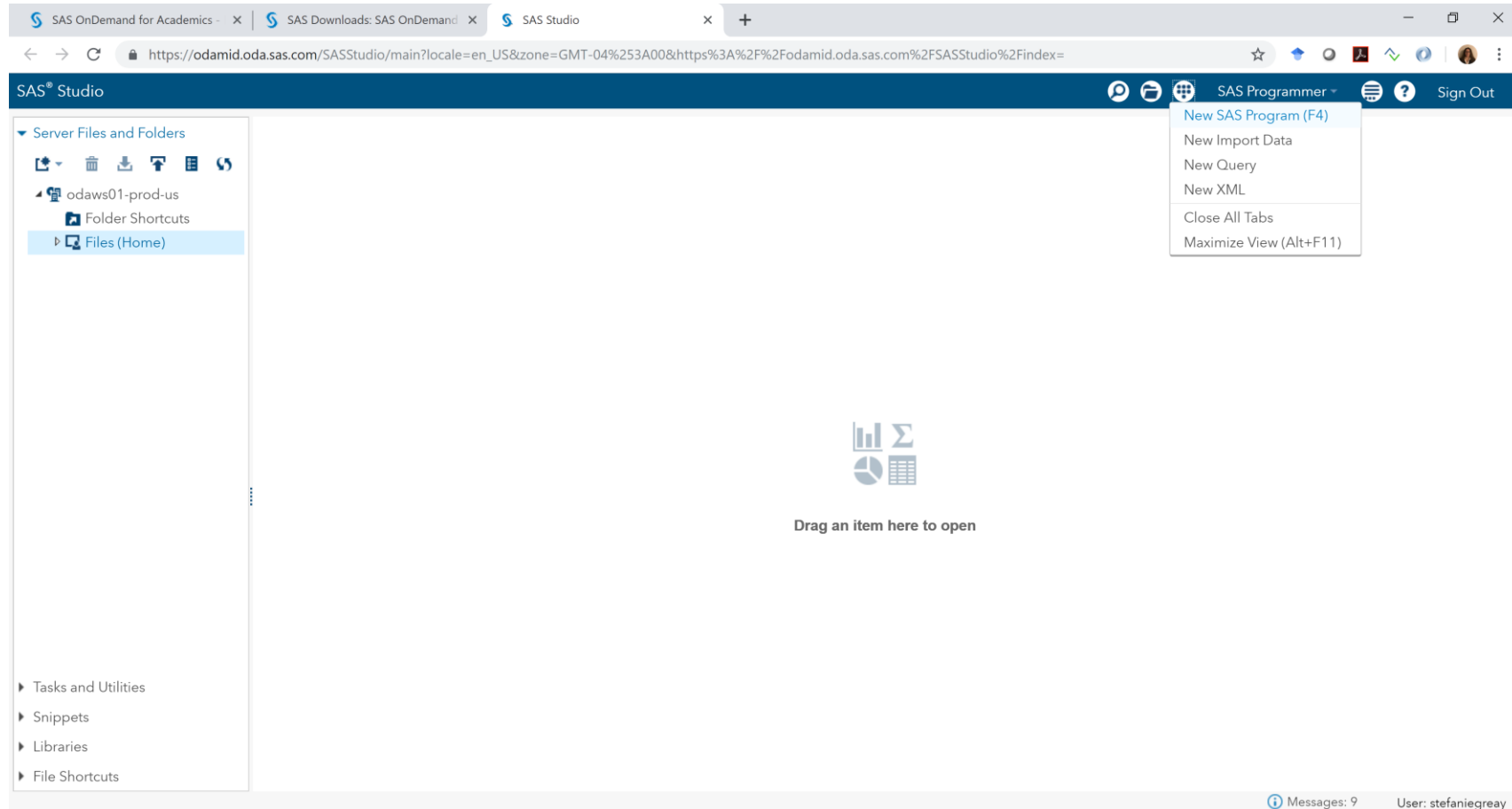
Selected files:		
1	XLSX	Medical1.xlsx 10.9 kb
2	XLSX	Medical2.xlsx 10.9 kb



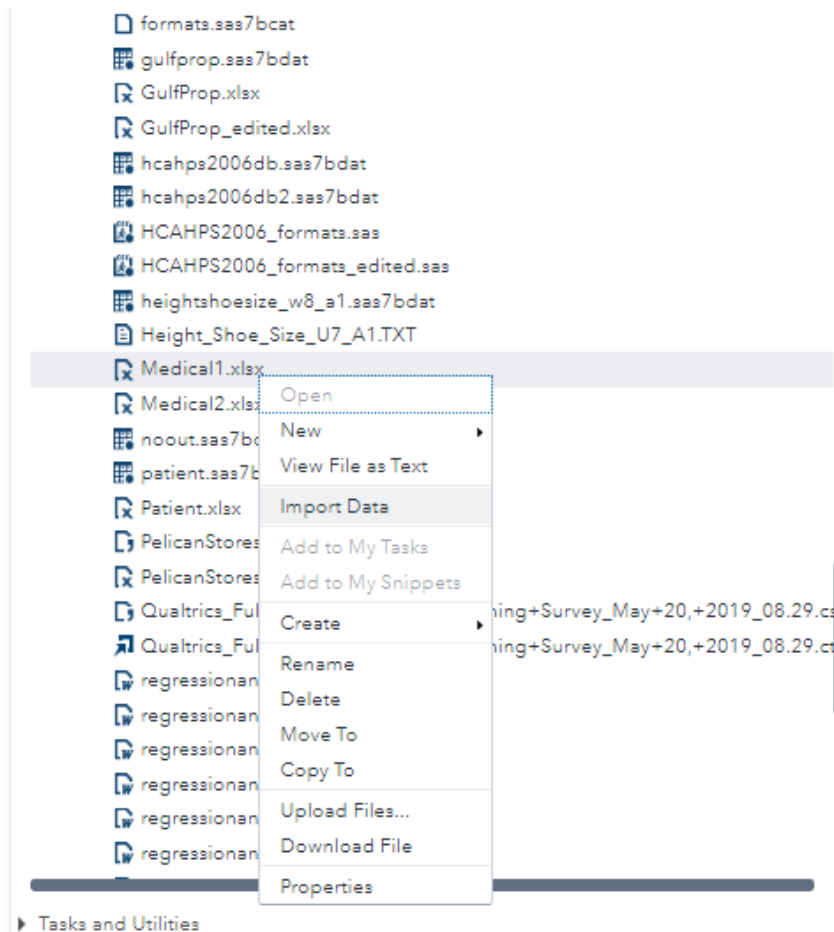
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



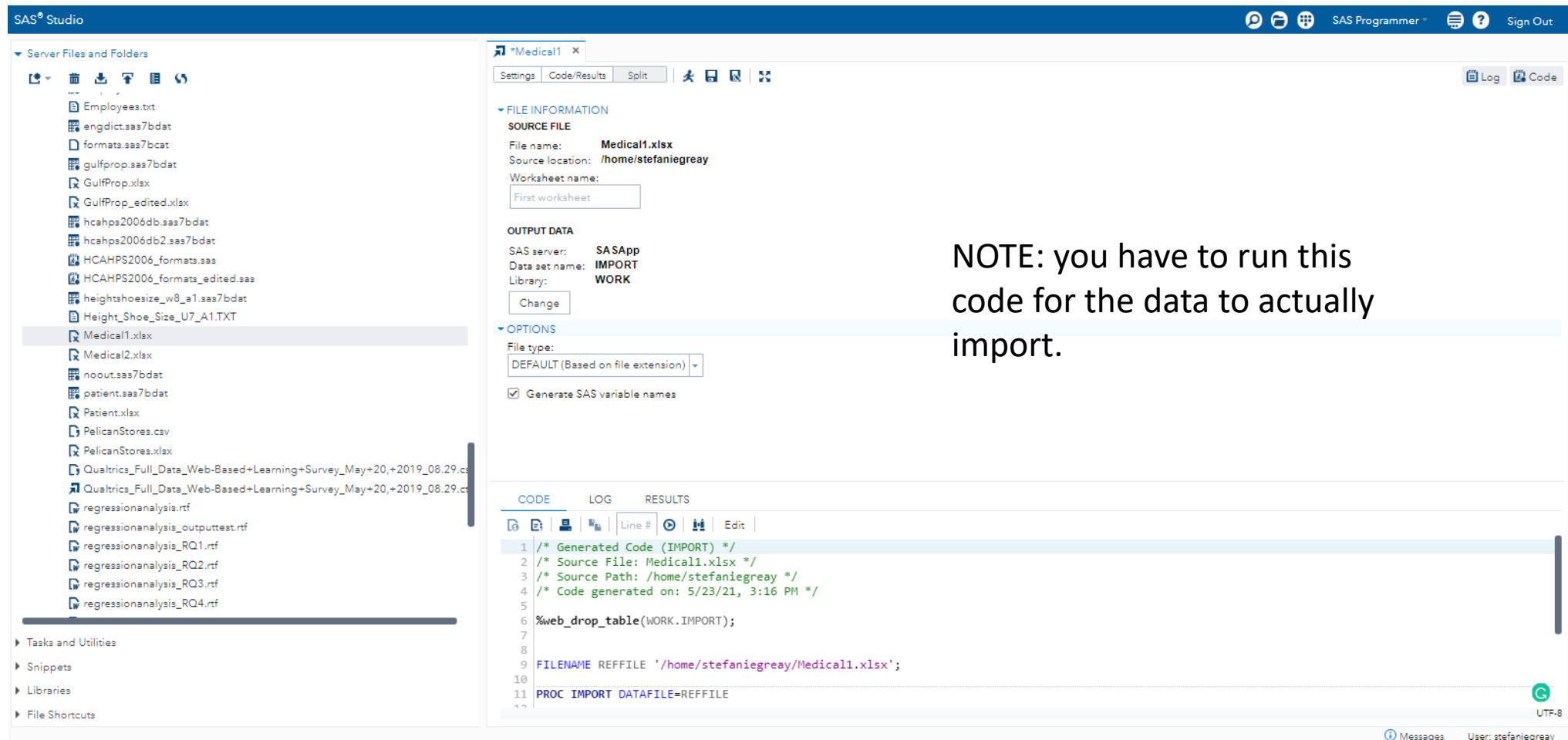
To get started with the SAS portion of the assignment, start a new SAS program.



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!)



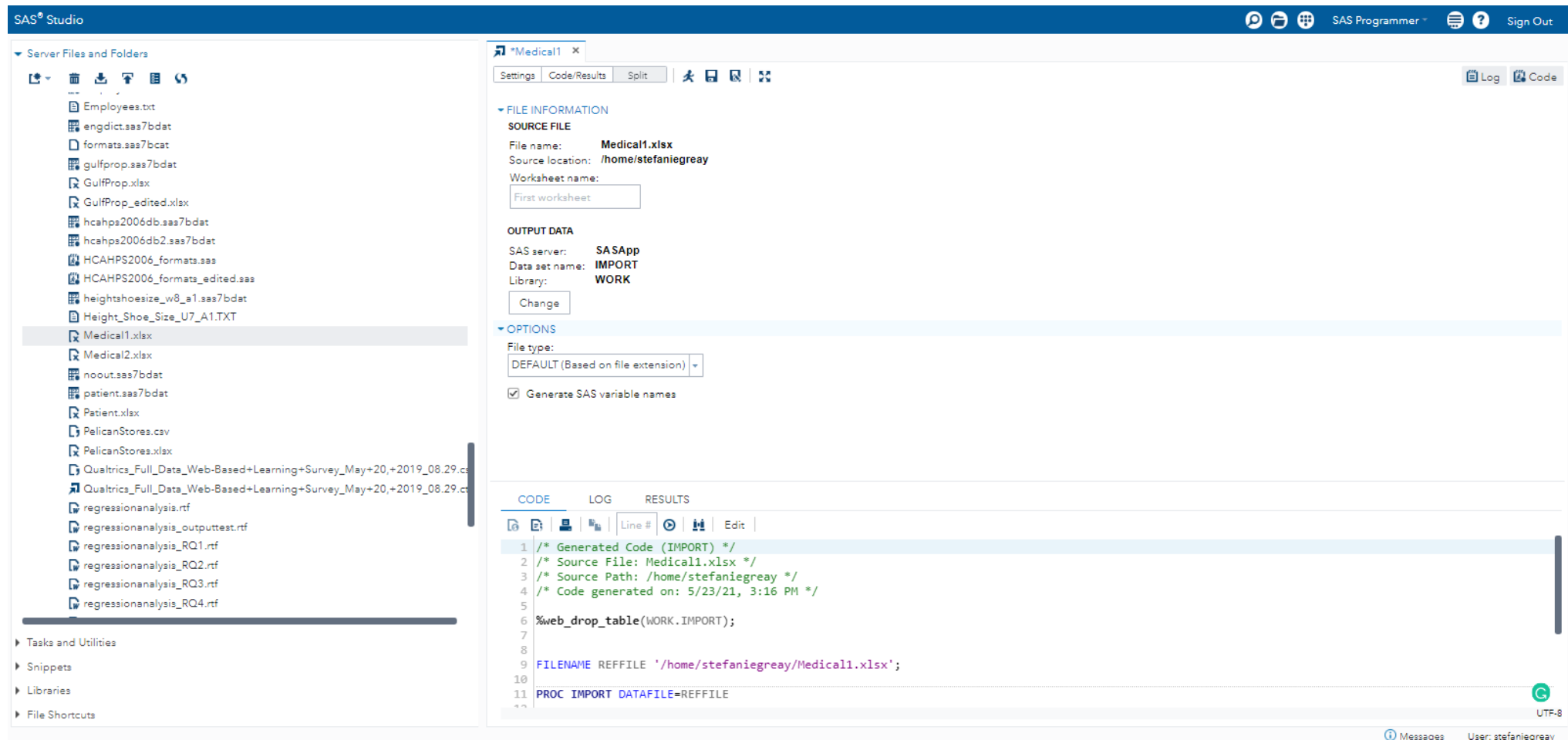
The screenshot displays the SAS Studio interface. On the left, the 'Server Files and Folders' pane lists various files, with 'Medical1.xlsx' selected. The main window is titled '*Medical1' and shows the 'Proc Import' configuration. The 'SOURCE FILE' section indicates the file name is 'Medical1.xlsx', the source location is '/home/stefaniegreay', and the worksheet name is 'First worksheet'. The 'OUTPUT DATA' section shows the SAS server is 'SASApp', the data set name is 'IMPORT', and the library is 'WORK'. The 'OPTIONS' section shows the file type is 'DEFAULT (Based on file extension)' and the checkbox 'Generate SAS variable names' is checked. The 'CODE' pane at the bottom displays the generated SAS code for importing the file.

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

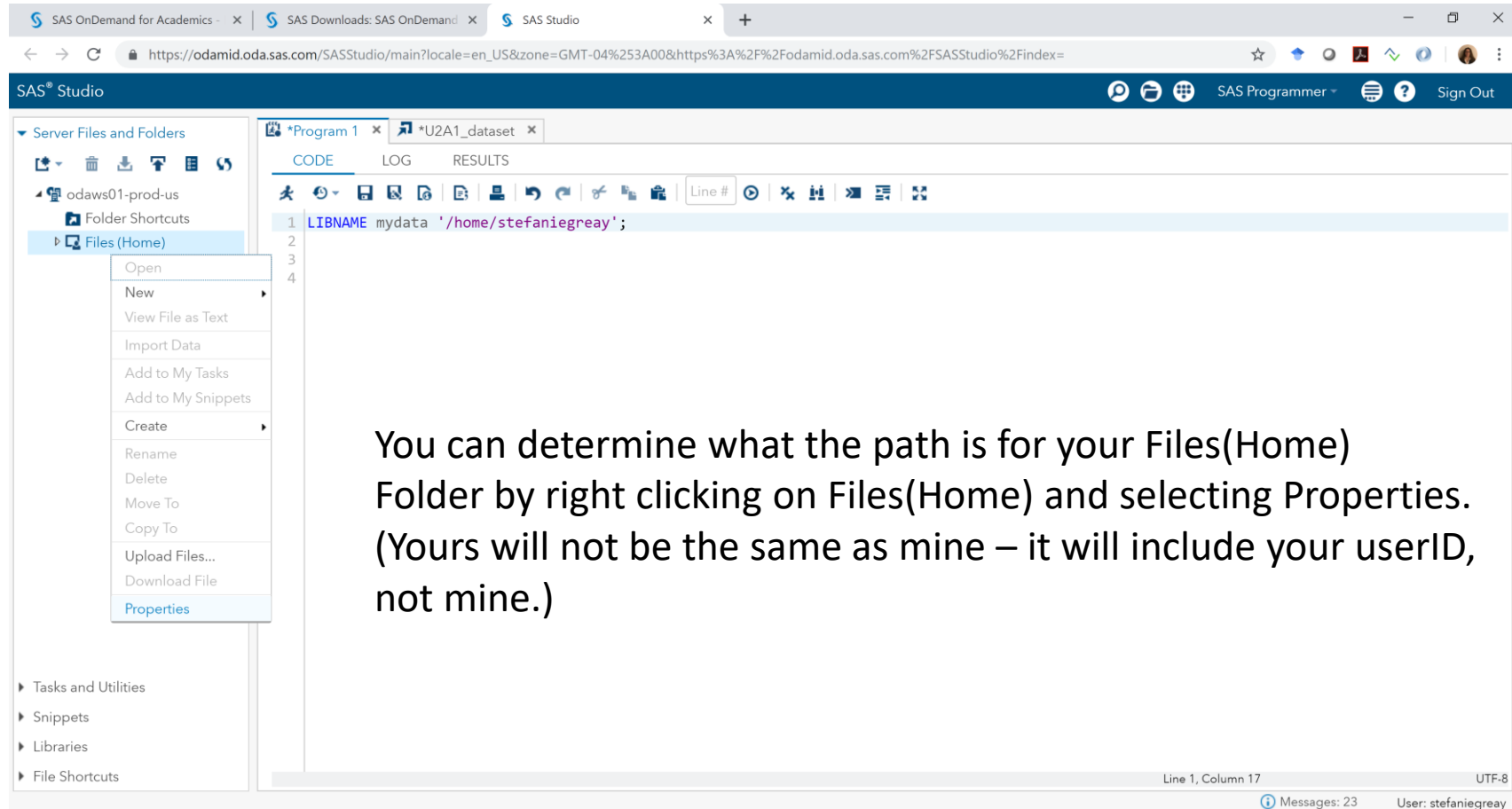
```
1 /* Generated Code (IMPORT) */
2 /* Source File: Medical1.xlsx */
3 /* Source Path: /home/stefaniegreay */
4 /* Code generated on: 5/23/21, 3:16 PM */
5
6 %web_drop_table(WORK.IMPORT);
7
8
9 FILENAME REFFILE '/home/stefaniegreay/Medical1.xlsx';
10
11 PROC IMPORT DATAFILE=REFFILE
```



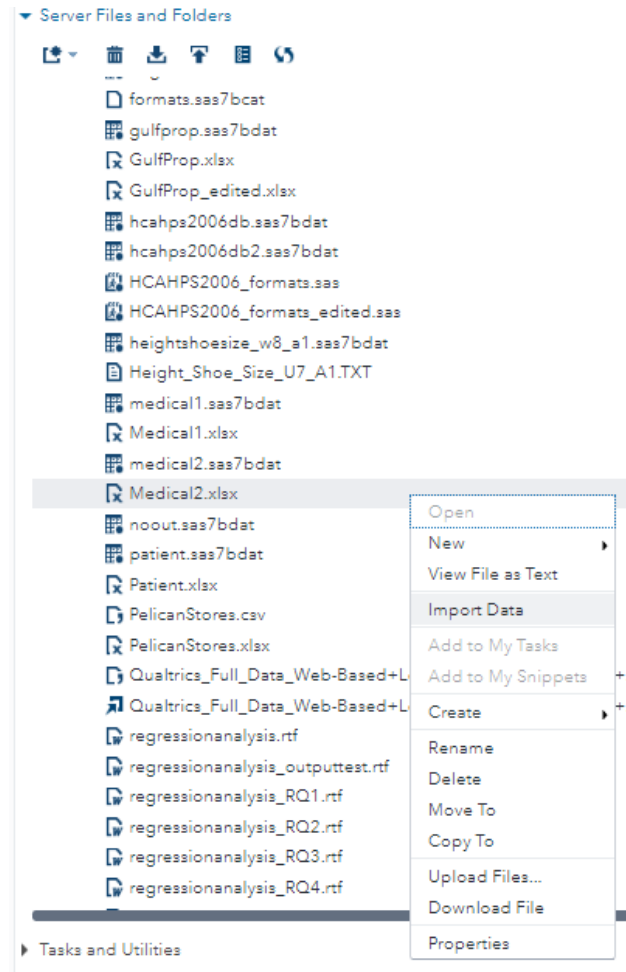
To run the code, click the icon that looks like a guy running.



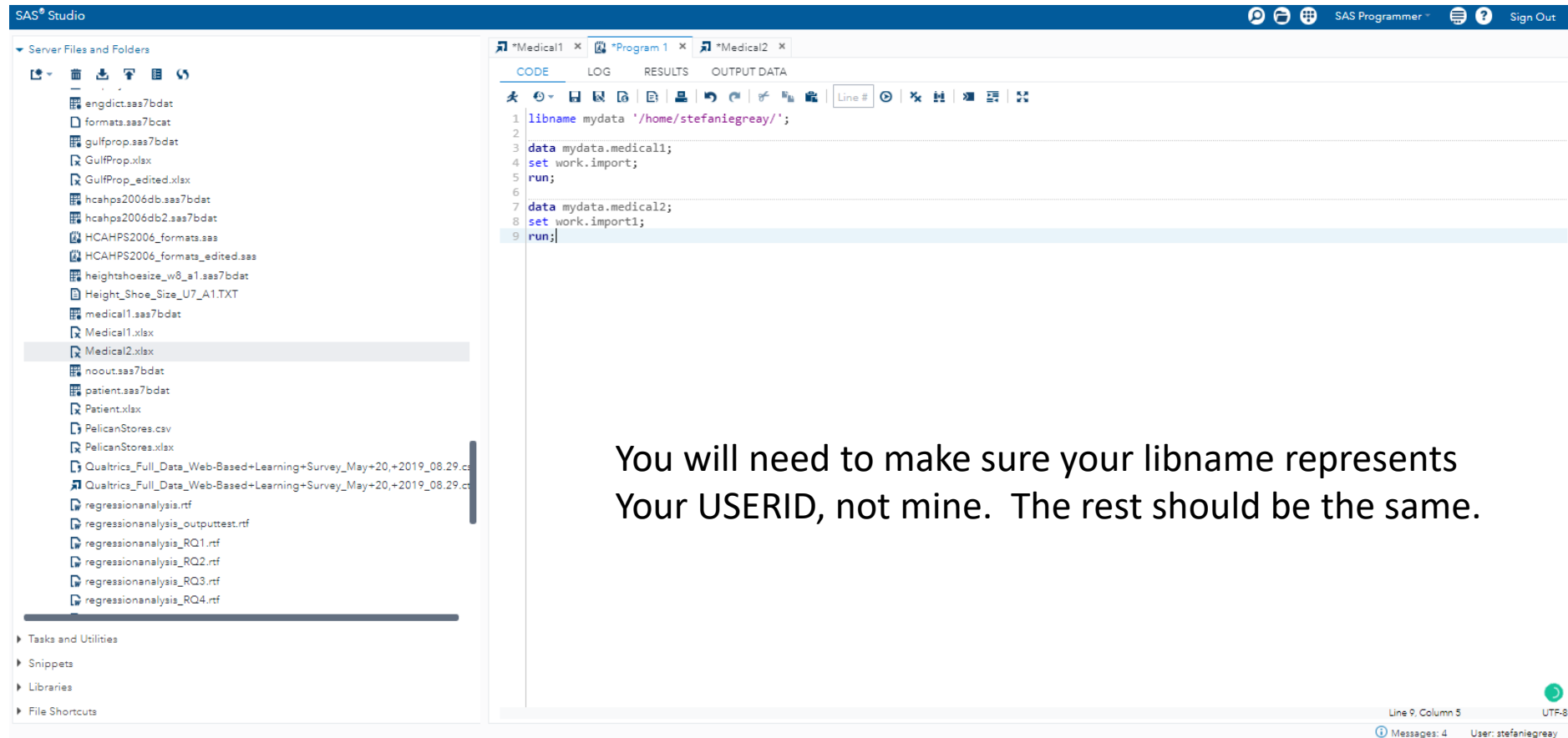
To create a SAS Library for your Files(Home) folder, you need to use a libname statement



Repeat the process for the second dataset.



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



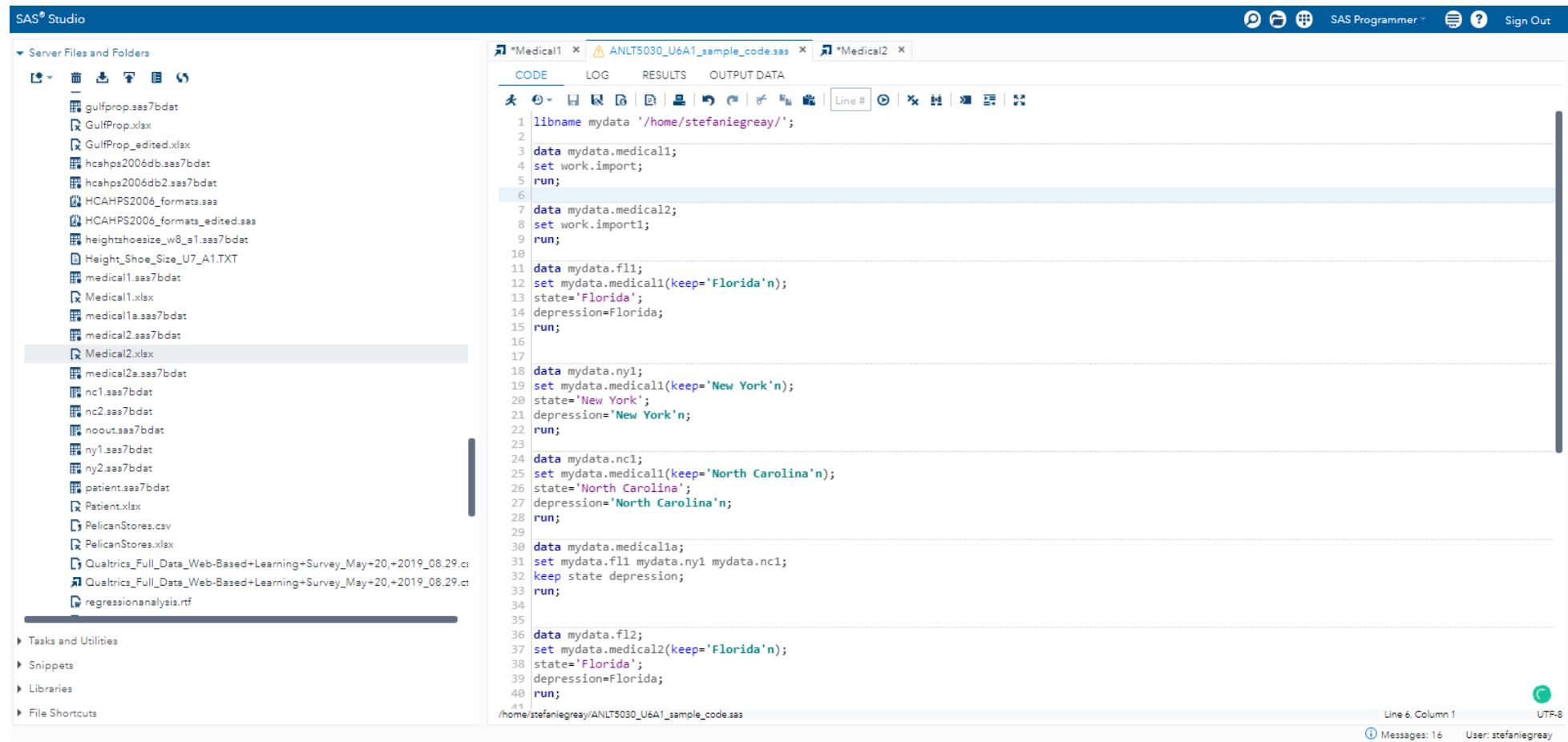
When you run the code, you will see the dataset in the output data window and can verify its success.

The screenshot displays the SAS Studio interface. On the left, the 'Server Files and Folders' pane shows a list of files, with 'Medical2.xlsx' highlighted. The main window is divided into several tabs: 'CODE', 'LOG', 'RESULTS', and 'OUTPUT DATA'. The 'OUTPUT DATA' tab is active, showing a table with 20 rows and 3 columns: 'Florida', 'New York', and 'North Carolina'. The table contains numerical data for each row. Below the table, there is a 'Property Value' section with fields for Label, Name, Length, Type, Format, and Informat. The bottom status bar indicates 'Messages: 6' and 'User: stefaniegreay'.

	Florida	New York	North Carolina
1	3	8	10
2	7	11	7
3	7	9	3
4	3	7	5
5	8	8	11
6	8	7	8
7	8	8	4
8	5	4	3
9	5	13	7
10	2	10	8
11	6	6	8
12	2	8	7
13	6	12	3
14	6	8	9
15	9	6	8
16	7	8	12
17	5	5	6
18	4	7	3
19	7	7	8
20	3	8	11



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 H_0
If $p > \alpha$	Fail to Reject H_0



Step 5: Conclusion

Result	Evidence in favor of H_a /claim?	Statistical Significance
Reject H_0	Sufficient Evidence (to suggest claim shown in H_a may be valid)	Statistically Significant
Fail to Reject H_0	Insufficient Evidence (or not sufficient evidence) (to suggest claim shown in H_a 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 H_0 for the ANOVA 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.

