ANLT5020 – Unit 8 Assignment 1 Tutorial

SAS Studio

Instructions

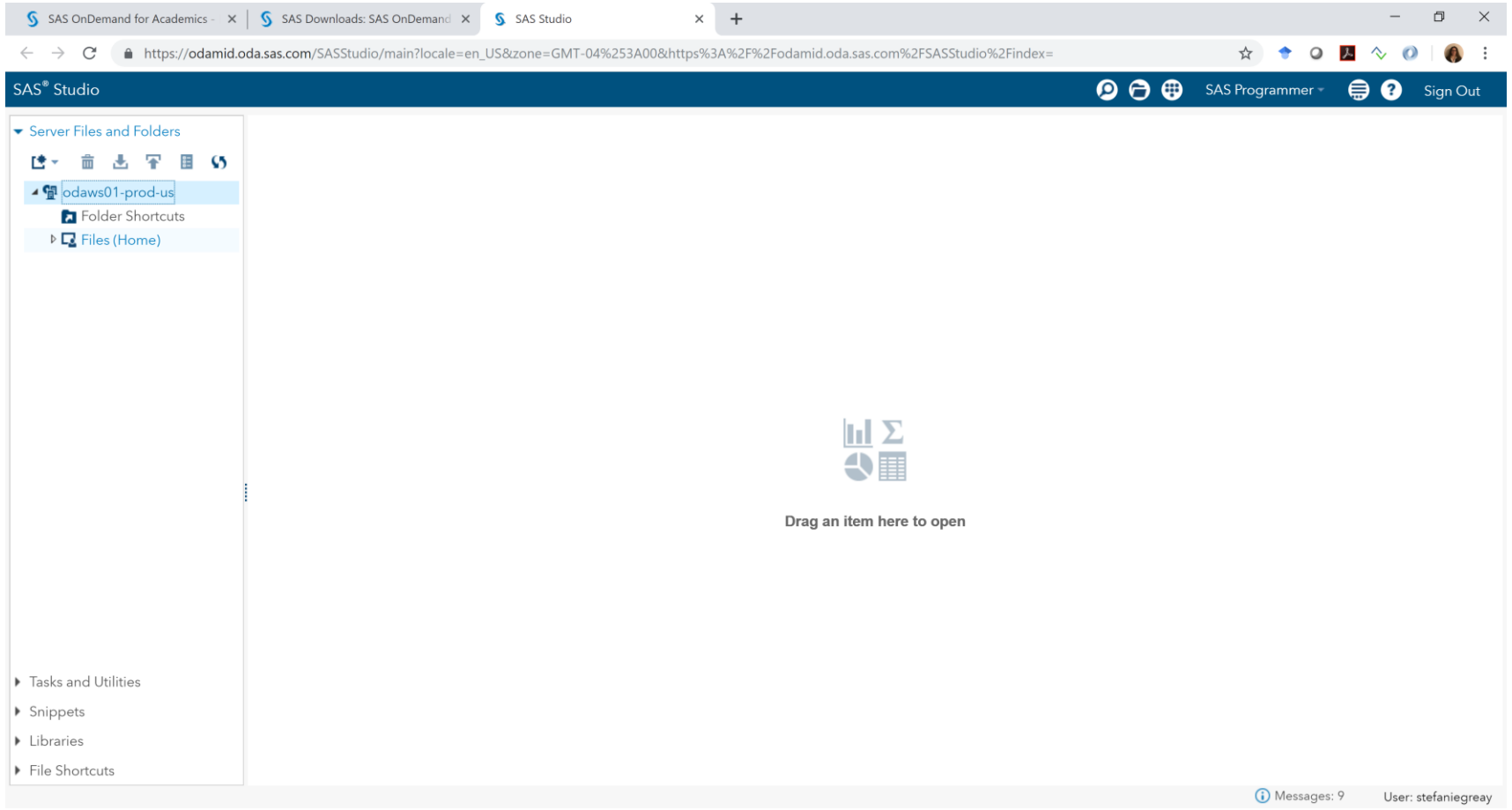
For this assignment, complete the following:

- Write SAS code to merge the datasets Employees.txt and Departments.txt. You first need to sort each dataset by the common variable to each dataset using PROC SORT. Write SAS code to sort each dataset by that common variable.
- Write SAS code using the merged dataset to determine the average salary, minimum salary, and maximum salary by Department_Name.
- Describe how you can merge all three datasets. Notice that all three datasets do not share a common variable.
- Write SAS code to merge all three datasets. Remember you first need to use PROC SORT.
- Using the merged dataset, write SAS code to determine the average salary, minimum salary, and maximum salary by Supervisor.

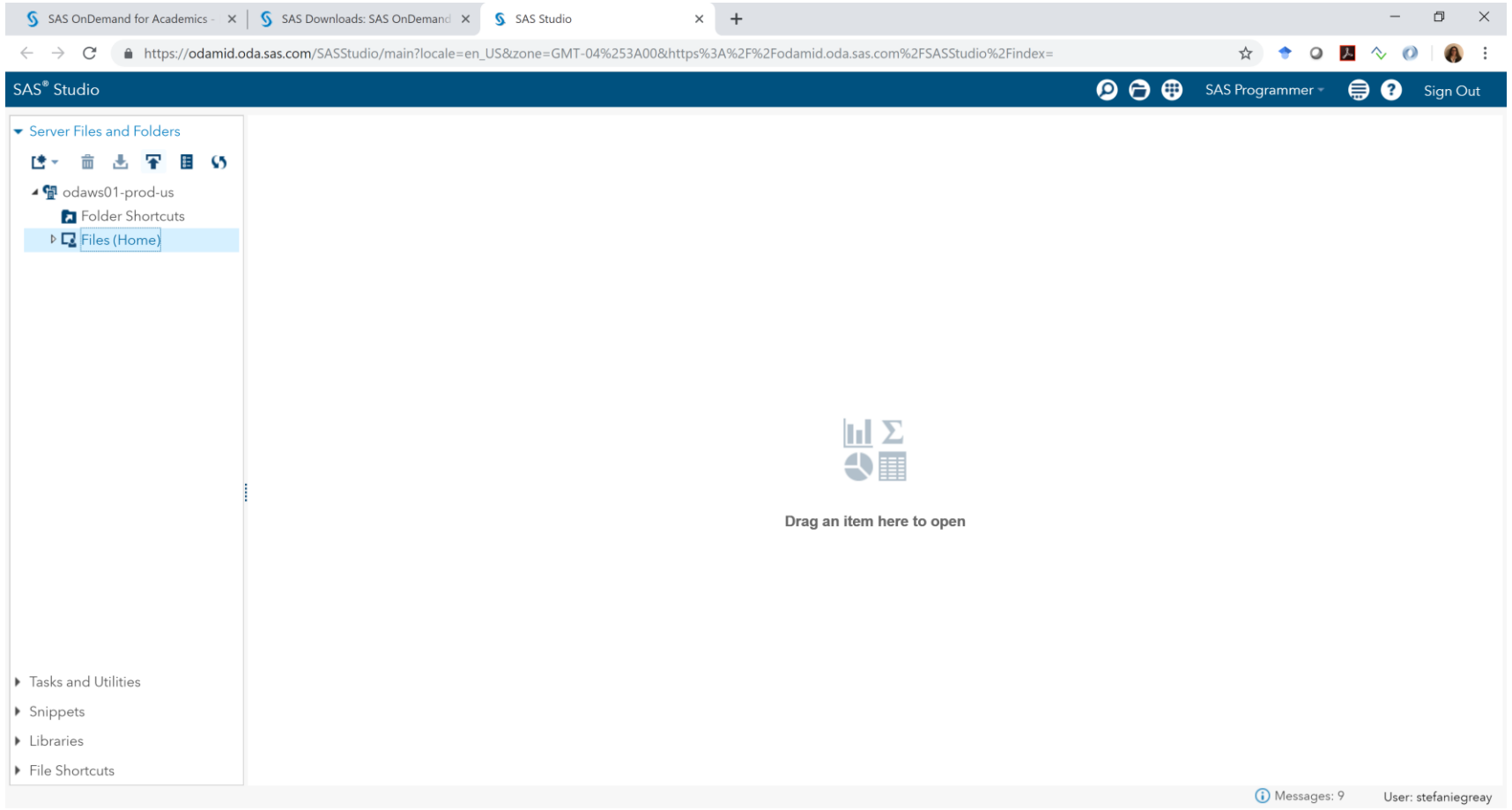
Dataset

- Download the Employees.txt, Departments.txt, and Supervisor.txt files from the course datasets zip file or from the Unit 8 Welcome announcement in the course announcements.

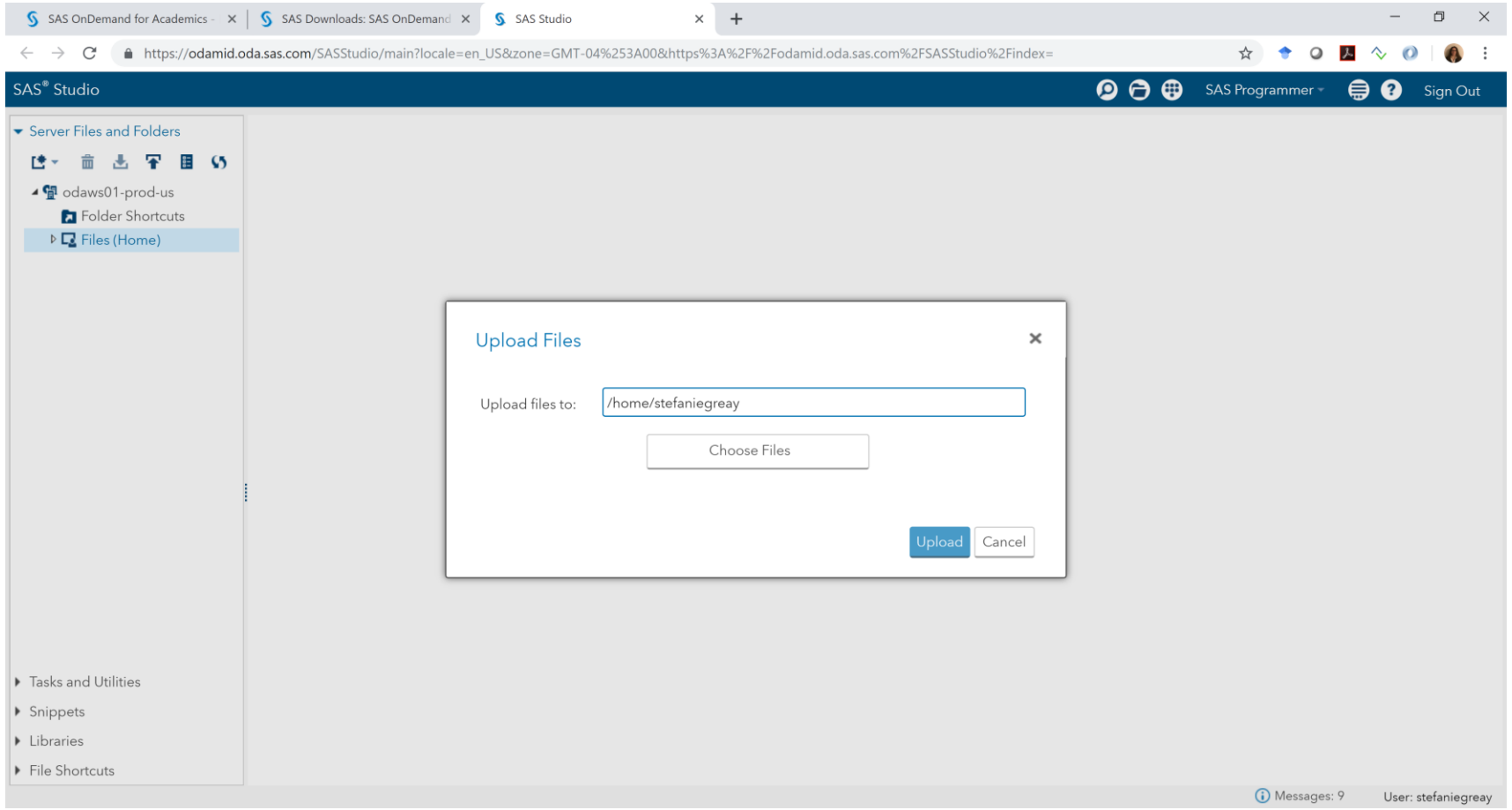
Click on Files(Home)



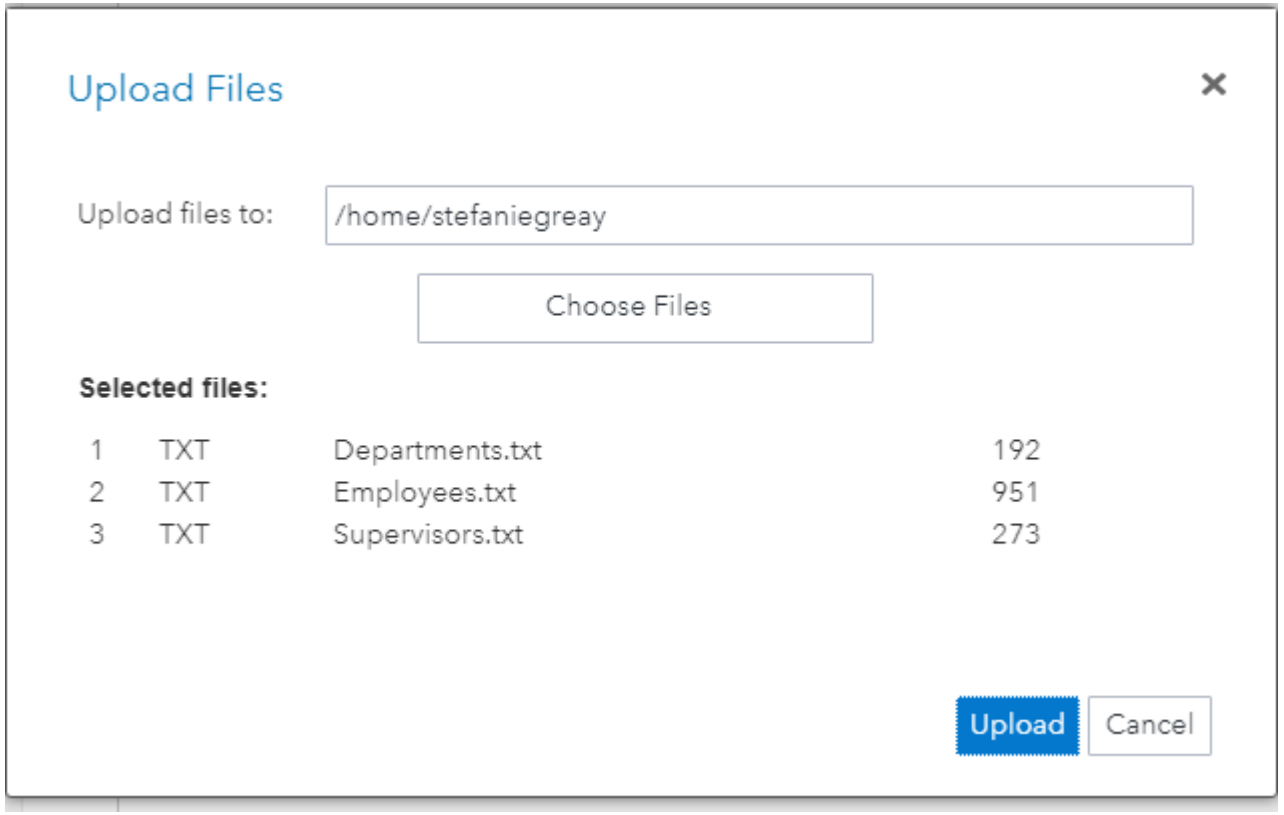
The Upload button will display in dark blue



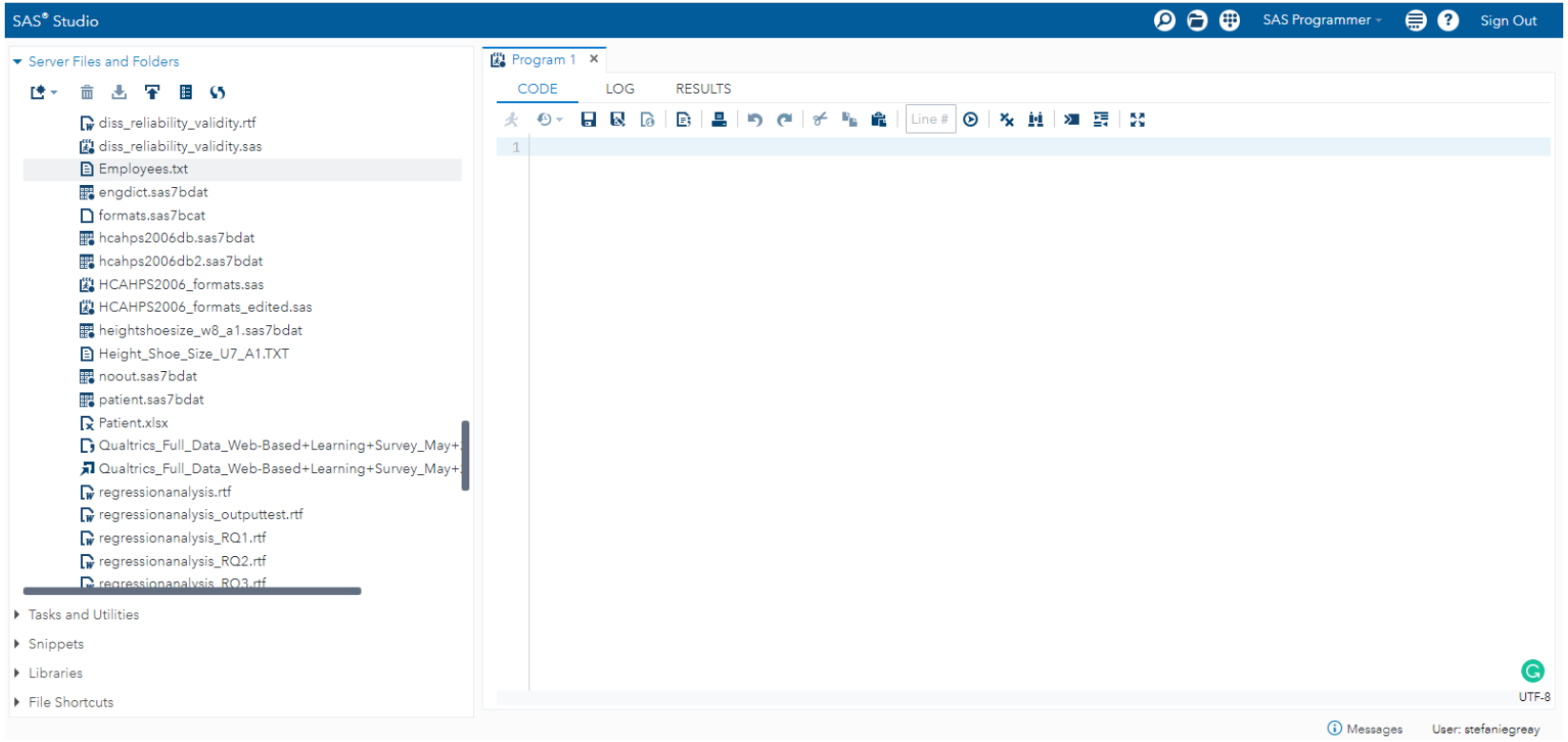
You can create a folder at this point, if you wish, or simply upload to your home directory.



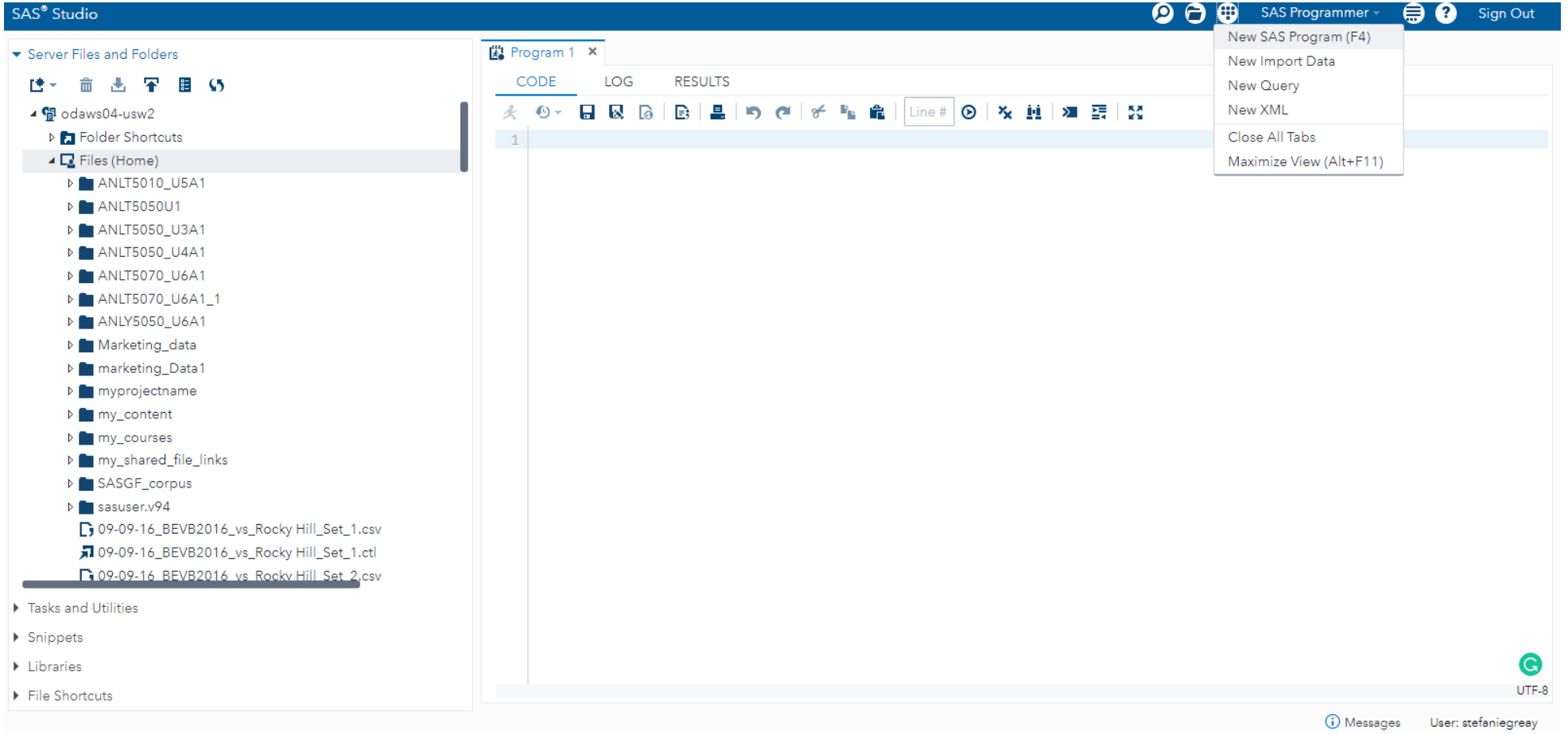
Select “Choose Files” to browse your computer for the dataset you want to upload. Once the dataset has been selected, click “Upload.”



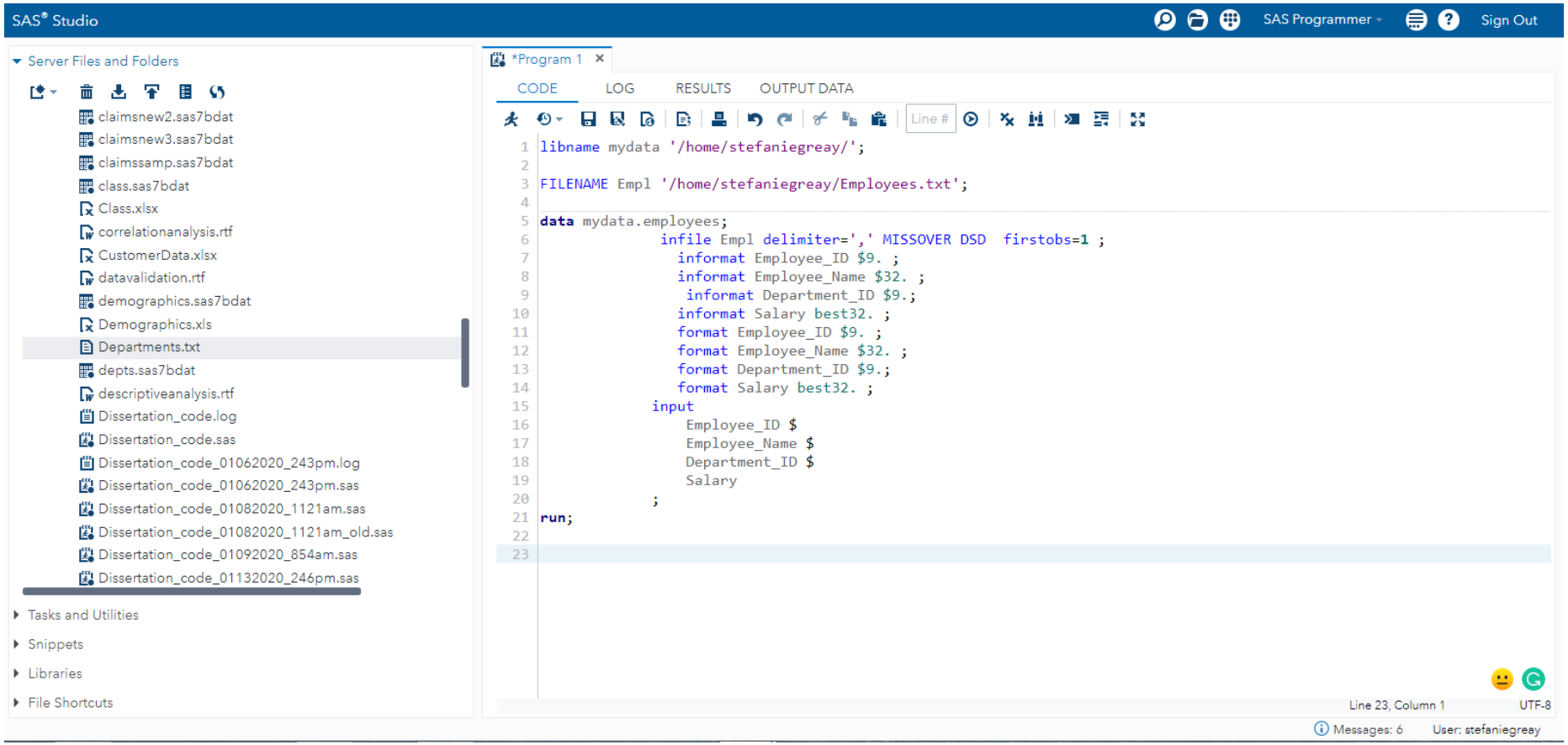
You will be able to view your files by clicking on “Files(Home)” to verify that your file successfully uploaded.



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



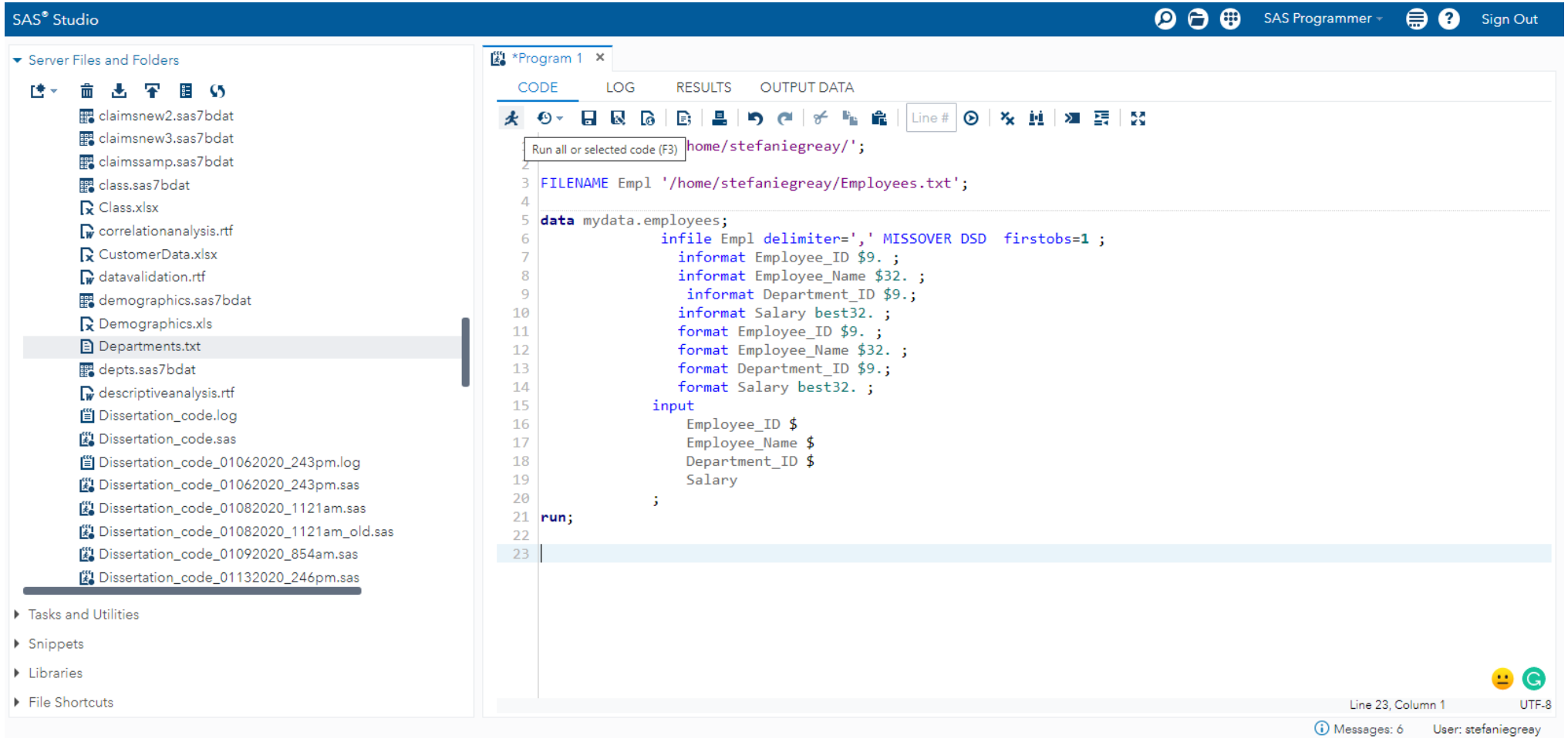
Import the dataset into a SAS dataset format using a SAS data step, as requested in the assignment. (You will need to change the location of your folder and the file you uploaded.)



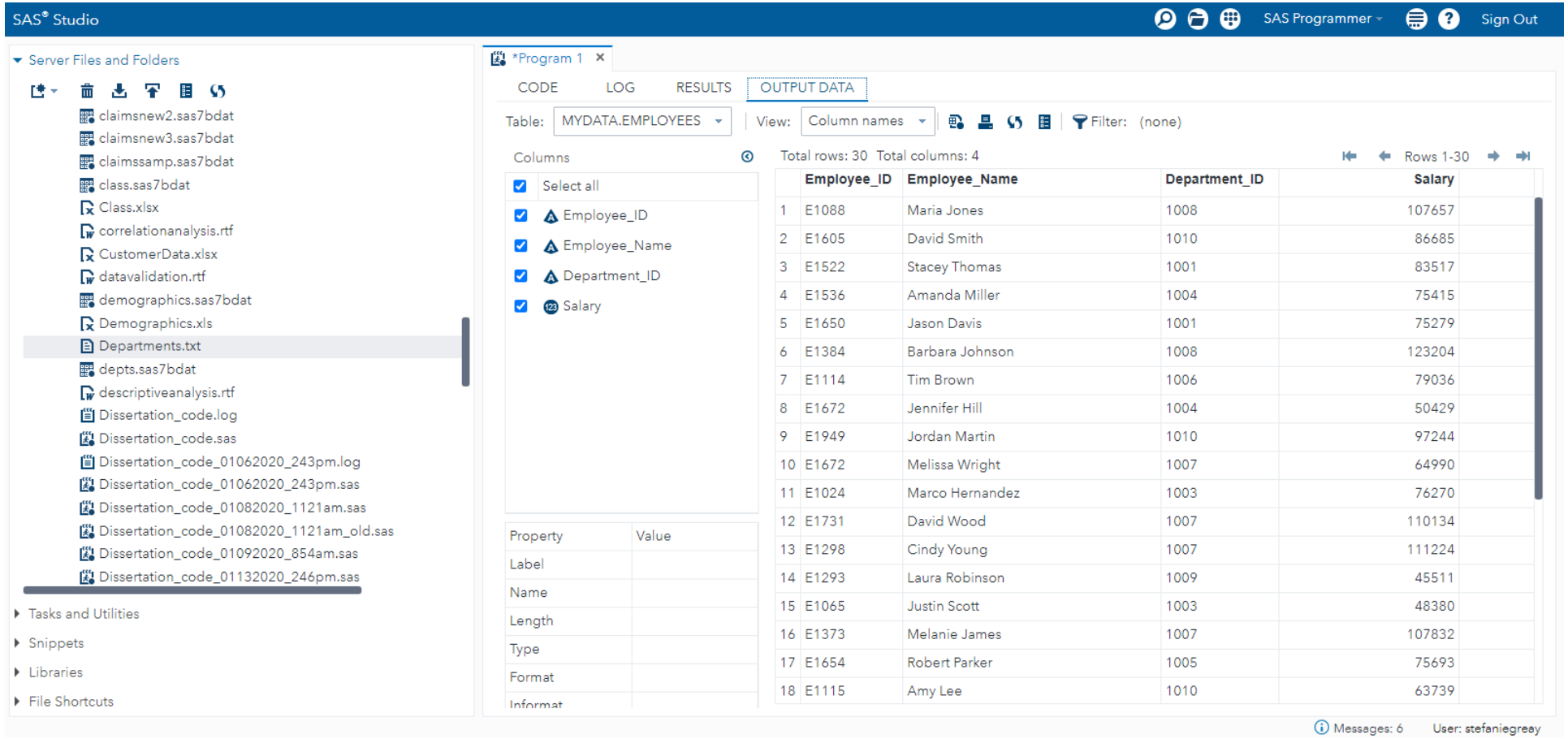
Code to import data in comma-delimited text file using data step. (This is for the employee.txt file only.)

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

FILENAME Empl '/home/stefaniegreay/Employees.txt';

data mydata.employees;
              infile Empl delimiter=',' MISSOVER DSD  firstobs=1 ;
                informat Employee_ID $9. ;
                informat Employee_Name $32. ;
                 informat Department_ID $9.;
                informat Salary best32. ;
                format Employee_ID $9. ;
                format Employee_Name $32. ;
                format Department_ID $9.;
                format Salary best32. ;
             input
                 Employee_ID $
                 Employee_Name $
                 Department_ID $
                 Salary
             ;
run;
```

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



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



Repeat the import process for each of the 3 datasets

The import code will need to be written and run for each of the 3 datasets to import the text files into SAS datasets.

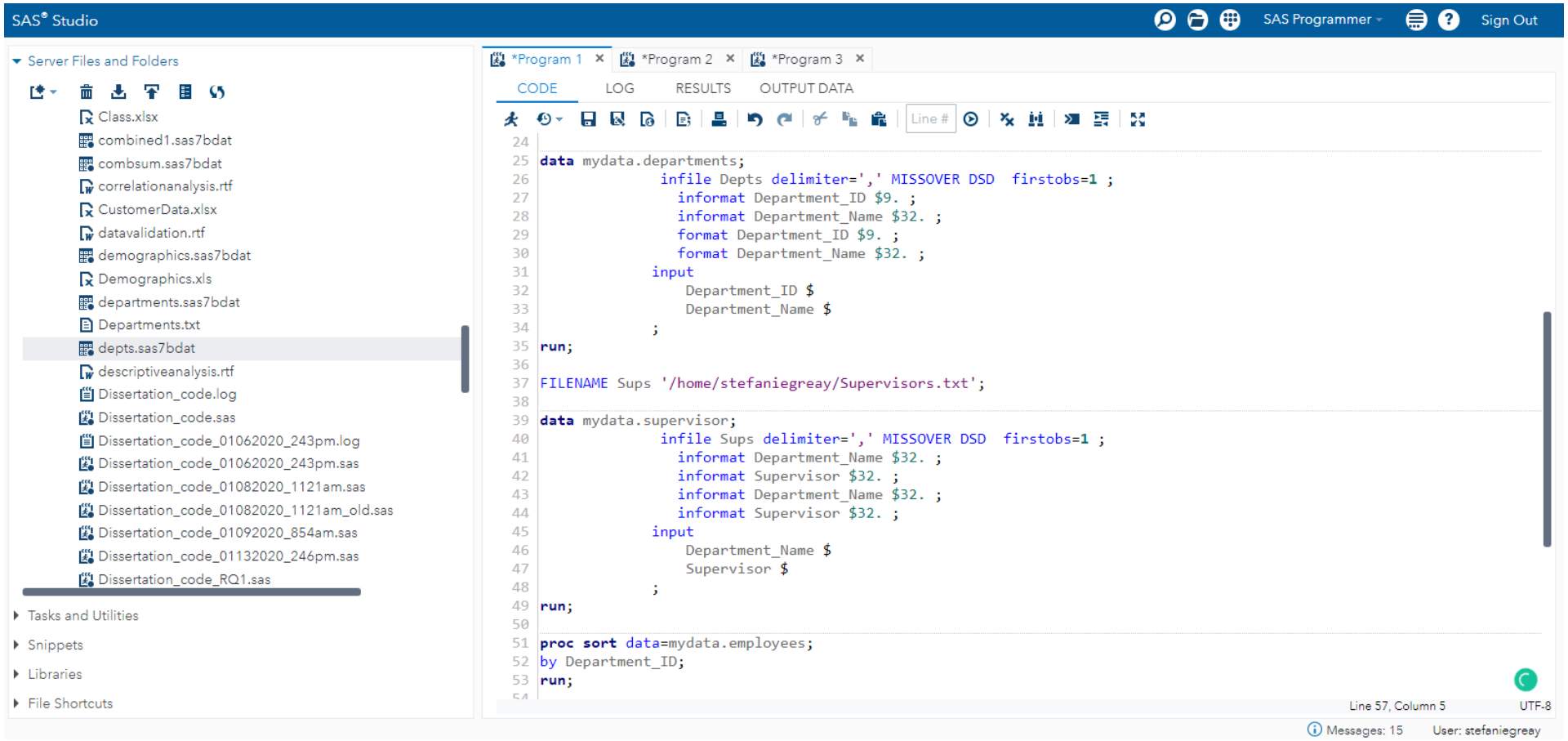
Code to import data in comma-delimited text file using data step. (This is for the departments.txt file only.)

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

FILENAME Depts '/home/stefaniegreay/Departments.txt';

data mydata.departments;
              infile Depts delimiter=',' MISSOVER DSD  firstobs=1 ;
                informat Department_ID $9. ;
                informat Department_Name $32. ;
                format Department_ID $9. ;
                format Department_Name $32. ;
             input
                 Department_ID $
                 Department_Name $
             ;
run;;
run;
```

Code to import data in comma-delimited text file using data step. (This is for the supervisors.txt file only.)

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

FILENAME Sups '/home/stefaniegreay/Supervisors.txt';

data mydata.supervisor;
              infile Sups delimiter=',' MISSOVER DSD  firstobs=1 ;
                informat Department_Name $32. ;
                informat Supervisor $32. ;
                informat Department_Name $32. ;
                informat Supervisor $32. ;
             input
                 Department_Name $
                 Supervisor $
             ;
run;
```

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



Sample Code for combining the employees.txt and departments.txt datasets

```
proc sort data=mydata.employees;
by Department_ID;
run;

proc sort data=mydata.departments;
by Department_ID;
run;

data mydata.combined1;
merge mydata.employees mydata.departments;
by Department_ID;
run;
```

Sample Code to determine the average salary, minimum salary, and maximum salary by Department_Name.

```
proc sort data=mydata.combined1;
by Department_Name;
run;

proc means data=mydata.combined1;
by Department_Name;
output out=mydata.combsum min=minsal max=maxsal mean=meansal;
run;
```

Sample Code to merge the supervisor data with the already merged other 2 datasets

```
proc sort data=mydata.supervisor;
by Department_Name;
run;

data mydata.combined2;
merge mydata.combined1 mydata.supervisor;
by Department_Name;
run;
```

Sample code to determine the average salary, minimum salary, and maximum salary by Supervisor.

```
proc sort data=mydata.combined2;
by supervisor;
run;

proc means data=mydata.combined2;
by Supervisor;
output out=mydata.combsum2 min=minsal max=maxsal mean=meansal;
run;
```