Soc212 Stata Tutorial

Philosophical overview: using statistical software is not difficult, but you have to invest the time in learning the basics. It is the nuts and bolts of doing demography. It should be fun (or rather, you know you are a demographer when it is fun).

I have written a basic tutorial to help everyone grasp the fundamentals of using Stata.

This tutorial uses a simple data set to illustrate basic Stata fundamentals:

```
list
```

	id	inc	sex	ed	_merge
1.	1	10	1	6	3
2.	2	20	1	12	3
3.	3	30	1	18	3
4.	4	40	1	12	3
5.	5	10	2	6	3
6.	6	20	2	12	3
7.	7	30	2	18	3
8.	8	40	2	16	3

It's a small data set of income, gender, and education.

The tutorial addresses the following issues:
1. Basics such as are you in the right directory etc.
2. Merging data
3. Getting a basic overview of the data
4. Making basic changes to the data
5. Doing a basic descriptive analysis. (Descriptive tables etc.)
6. Some useful egen commands
7. Making a dataset of aggregate statistics (i.e., the collapse command).
8. Dealing with log files and pasting results.
9. Making and running do-file.
10. Questions.

What I want you to do:
1) Run the do-file.
2) Read the do-file carefully.
I want you to annotate the do-file to write explanations of commands you don't understand.
You must understand each of these basic Stata commands. If you already understand the command, place a check mark beside the command on the printed version of the do-file. If you don't understand the command type "help commandname" at the Stata prompt **and** look at the effect of the command on the data. Write an explanation of the command.
3) On many of the commands, I have given you multiple examples of the same command, each slightly different. Make sure you understand each of them. It might help

to load the tut-temp.dta data in (it's the data in memory at the end of the do-file) and try these out interactively.
After you have changed the data (say by accidentally deleting it, you can get back to the tut-temp data by typing:
clear
use tut-temp
4) Answer the questions at the end of the do-file (they are reproduced below). These questions only use commands illustrated in this do-file, but they challenge your ability to go from English to the logical language of a computer statistical program. This, obviously, is an important skill for a demographer, no matter how frustrating it may seem at first.
Turn the annotated do-file and the answers to the questions in next Monday
5) The goal of this tutorial is for you to make sure you know the basics. Anything we do in lab will build on this concepts and commands. I will teach you any additional commands you will need as we get to them. If you are getting stuck on this tutorial, then I want to know where you are having problems. If it is a minor issue, we can deal with it in the next lab. If it is serious, please come by my office sometime on Thursday or Friday of this week (between 9-2:30 on Thursday and 11-5 on Friday).

Questions. Give me the Stata command needed to answer these questions with this data. You do not need to give the answer, just the Stata command.

```
(1) What are the names of the variables
(2) What percent of the men in the data are in each of the education
    categories?
(3) What percent of individuals in each of the education categories are
    men?
(4) What percent of women in the data are in each of the income
    categories?
(5) What is the mean level of income for individuals in each of the
    education categories?
(6) What is the mean level of income by education and gender?
 (7) I want to change values of education that are 18 to 19
(8) I want to replace the education of all men and set it equal to 0
(9) I want to replace the income of the first case and set it to 100
(10) I want to make a new variable that is the sum of education and
    income
(11) I want to make a data set of aggregate data of the mean education
    and income by gender.
(12) I want to save the data as "ugh"
(13) I want to make a (1-way) table of men's education.
(14) I want to sort the data in descending order of income.
(15) I want to get rid of all the men in the data.
(16) I want to keep only the cases in the data with 20 years of
    education.
(17) I want to get rid of everything in memory.
```