

**Introduction to Stata
2017-18**

**02.
Getting Data into Stata**

1. Enter Data: Create a New Data Set in Stata	2
2. Enter Data: How to Import an Excel Data Set	4
3. Import a Stata Data Set Directly from the Internet	9
4. Importing Excel Dates into Stata	10

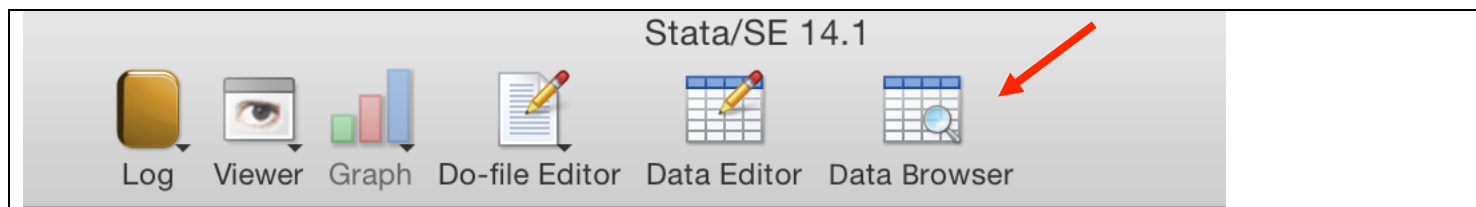
Please note I do a lot of comments!

You will see many of my commands that begin with an **asterisk (*)**. I've put some of these in green (but not all) so that they are easier to see. Commands in STATA that begin with an asterisk (*) are comments. While recommended, you don't actually have to type these comments.

1. Enter Data – Create a New Data Set in Stata

Most of the time, you will import data into Stata for analysis from an Excel file or from a Stata data set available on the internet. Once in a while, however, you may want to create a new data set by launching **Data Editor** in Stata.

Tip – The **Data Editor** icon is located on the top, horizontal, navigation bar



YOUR TURN – Create a Stata Data Set Containing the Following Data

Follow the commands below to create a data set containing the n=4 observations on the variables **id**, **dob**, **gender**, and **weight** that are shown below.

id type: numeric	dob type: date	gender type: string/character	weight type: numeric
1	3/26/1926	male	161.3
2	6/9/1956	female	120.1
3	4/1/1954	male	223.2
4	11/4/1951	female	124.0

Tip – The type of variable matters!!

In this illustration, you will be creating three different *types* of variables; **numeric**, **string**, and **date**.

Launch Stata

Then type the following two commands into the command window

```
clear
set more off
```

Then follow along, issuing the following commands, to create your Stata data set

```
. * STEP 1: Define your variables (lower case recommended), set type, and initialize
. generate id=.
. generate str8 dob_string=""
. generate str8 gender=""
. generate weight=.

. * STEP 2: Click on DATA EDITOR icon. This will bring you to an empty spreadsheet
. * ---- Enter the data on the previous page. Then close the data editor window ----*

. * STEP 3: Create dob (date of birth) that is a Stata date variable
. * For date variables with year in 4 digits, use function date with option "MDY"
. generate dob=date(dob_string, "MDY")
. format dob %tdNN/DD/CCYY
. drop dob_string
. list

. * STEP 4: Create 0/1 indicator of female gender
. generate female=(gender=="female")
. list

. * STEP 5: Attach labels to variable names
. label variable id "Subject id"
. label variable weight "weight (lbs)"
. label variable dob "Date of birth"
. label variable female "0/1 female"

. * STEP 6: Define dictionary of discrete variable values
. label define femalef 0 "male" 1 "female"

. * STEP 7: Attach coding labels to discrete variable values
. label values female femalef
. list

. * To see data with numeric labels provided
. numlabel, add
. list

. * To drop display of numeric labels
. numlabel, remove
. list

. * STEP 8 - Save data set using FILE > SAVE AS
```

2. Enter Data – How to Import an Excel Data Set

Beware - Take care that the data types are correct, especially for date variables!

There are multiple methods for importing an excel data set: (1) Copy and paste; (2) Importing the excel spreadsheet; and (3) using StatTransfer (The simplest but requires purchase of StatTransfer).

METHOD 1 – Copy (from Excel) and Paste (into Stata)

Step 1 – Launch Excel. Open the file stata_lab1.xls *You should see:*

A	B	C	D
id	dob	gender	weight
1	3/26/26	male	161.3
2	6/9/56	female	120.1
3	4/1/54	male	223.2
4	11/4/51	female	124.0

Step 2 – In Excel, use FORMAT > CELLS to format each column of data (numeric, text, custom, etc)

Tips – (1) For each variable, position cursor over the letter of the column (eg column A for formatting the variable ID) (2) If you format a column in Excel as a date variable, it will NOT import correctly into Stata. You must format it as type = custom.

Variable

Format Cells as

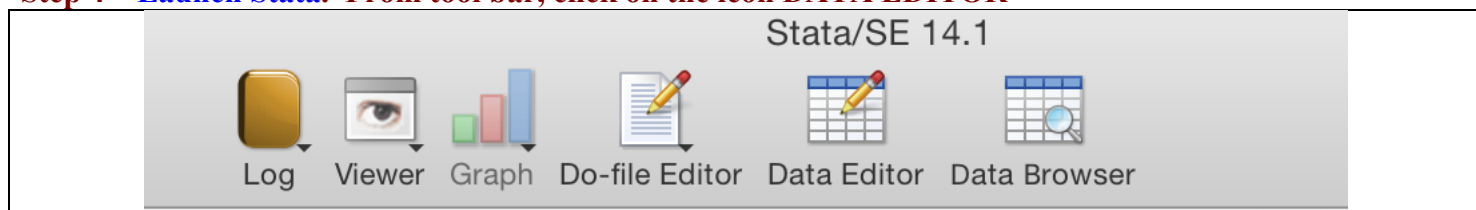
id
dob
gender
weight

numeric – then choose 0 places after the decimal point
custom – then choose the type: m/d/yy
text
numeric - then choose 2 places after the decimal point

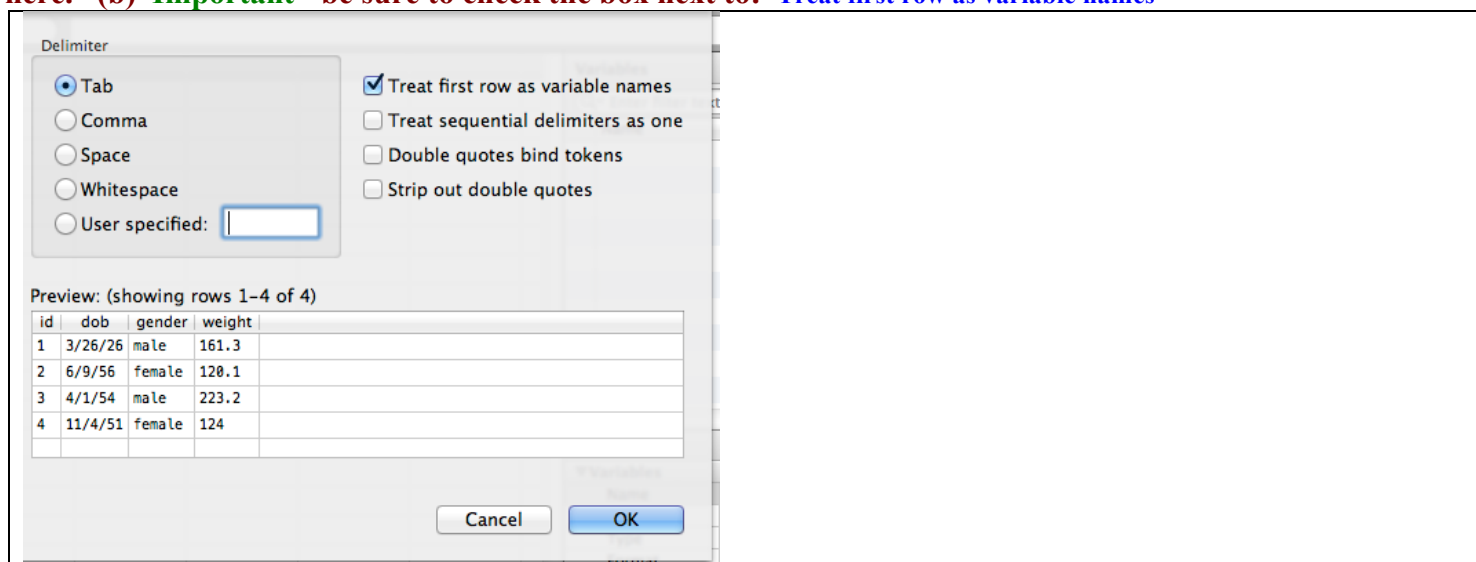
Step 3 – In Excel, select the data to be copied, including row headings with variable names.

Use EDIT > COPY to complete selection.

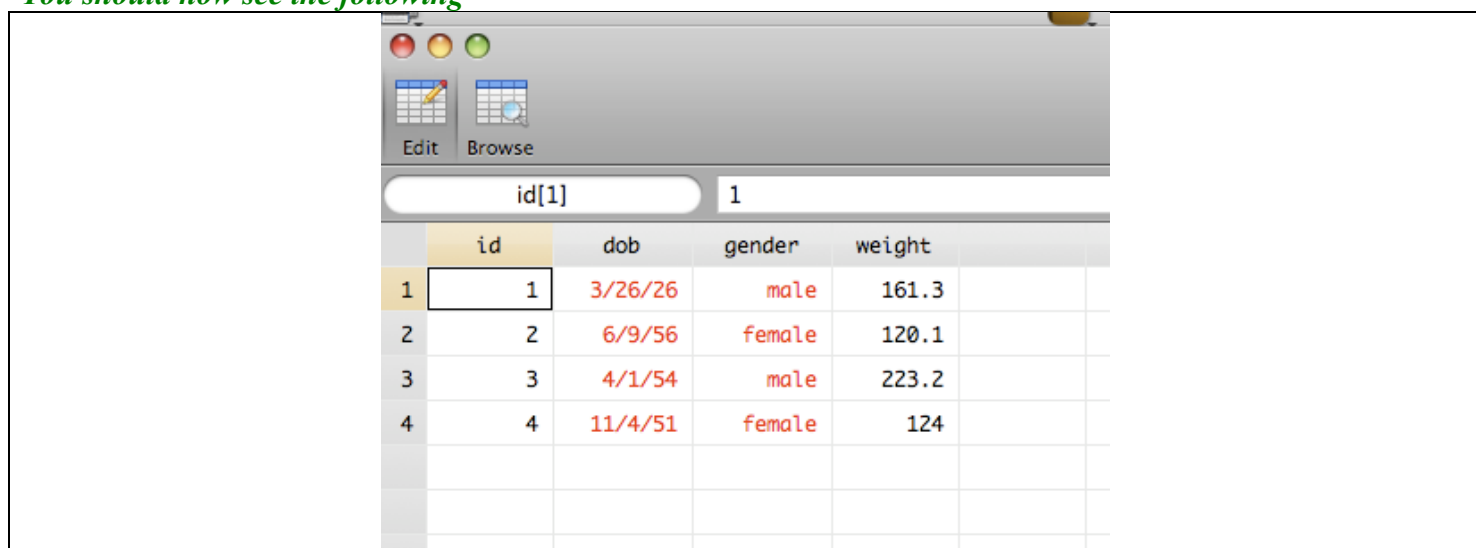
A	B	C	D
id	dob	gender	weight
1	3/26/26	male	161.30
2	6/9/56	female	120.10
3	4/1/54	male	223.20
4	11/4/51	female	124.00

Step 4 – Launch Stata. From tool bar, click on the icon **DATA EDITOR**

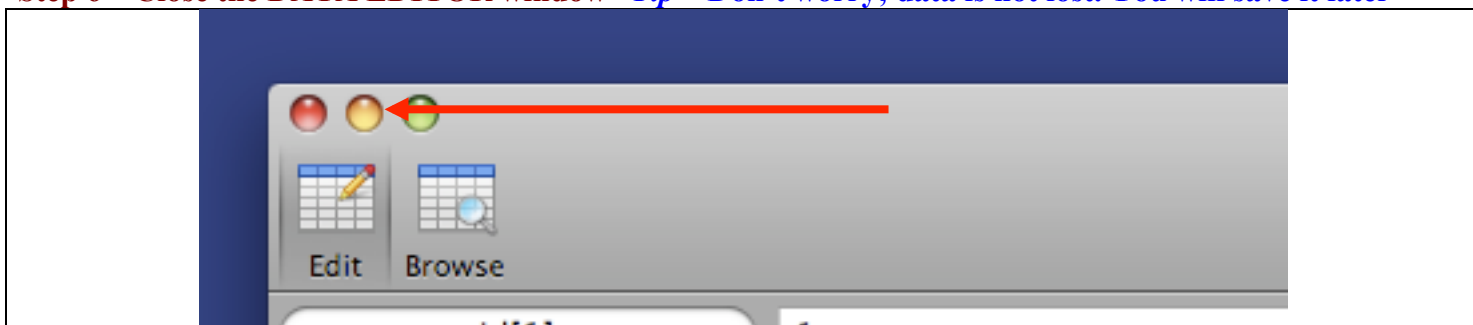
Step 5 – (a) Position cursor in cell Var1[1]. From the menu bar, use **EDIT > PASTE SPECIAL** to paste data here. **(b) Important** - be sure to check the box next to: **Treat first row as variable names**



You should now see the following



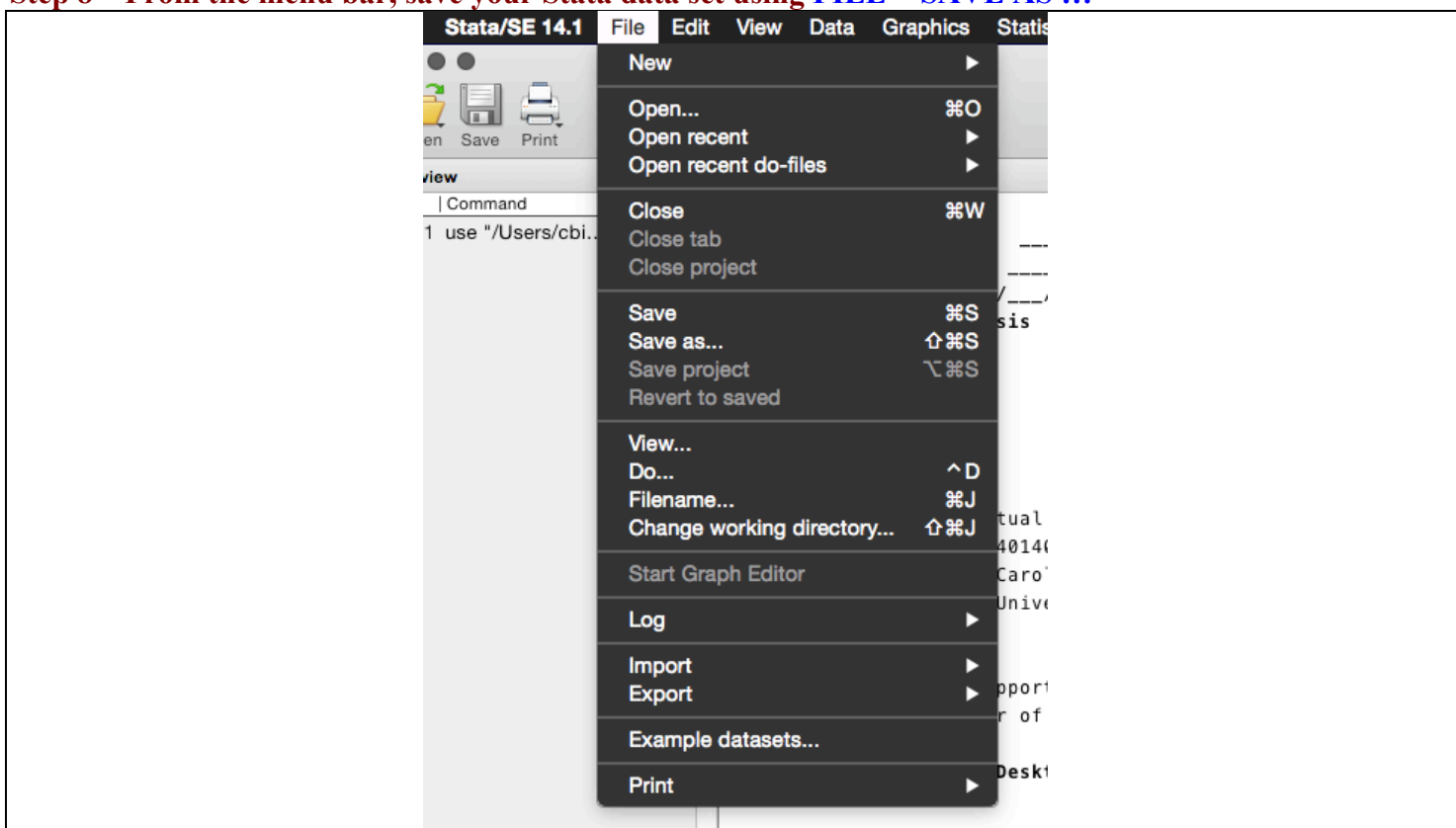
Step 6 – Close the DATA EDITOR window *Tip – Don't worry; data is not lost. You will save it later*



Step 7 – In the command window, issue the following commands to convert the Excel date variable (that fails to import as a date variable) into a Stata date variable that is a bona fide date variable. Eg; -

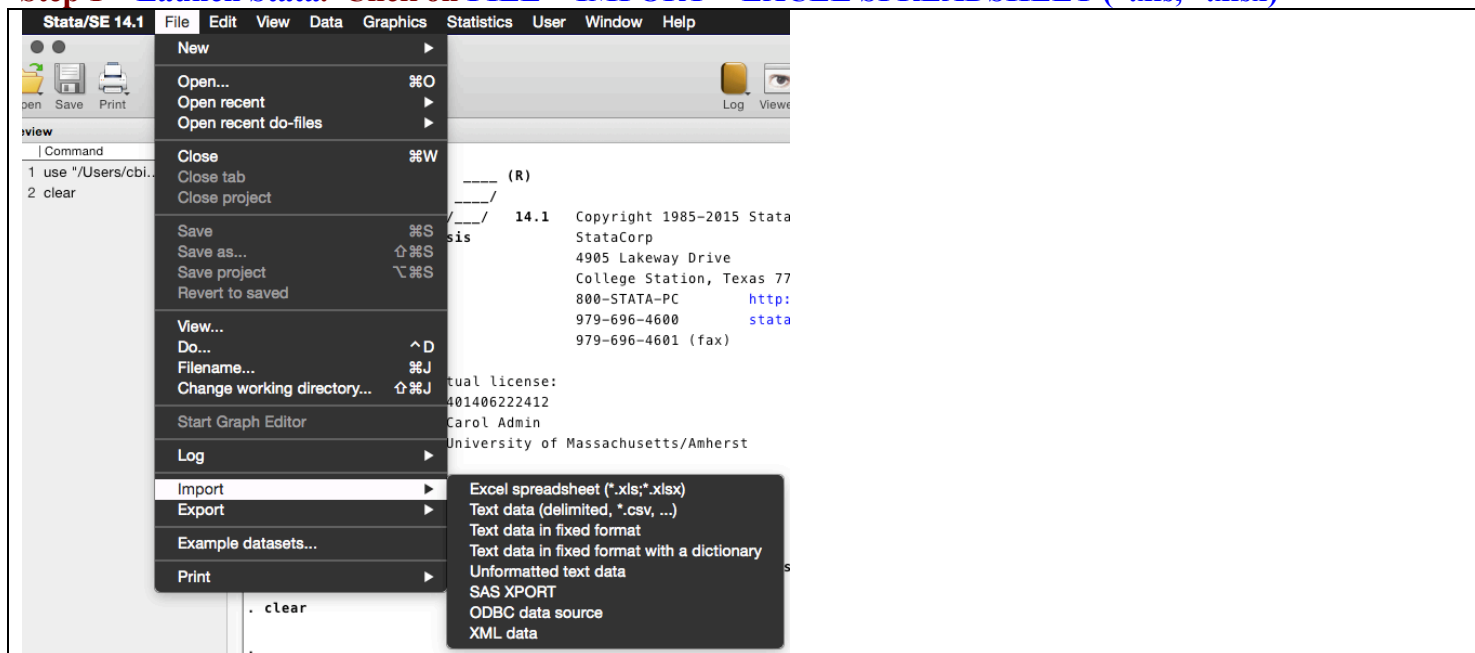
```
* For dates with year recorded in 2 digits, all in the 1900's:
* Use the function date(stringvariable, "MD19Y") with option "MD19Y" in quotes.
generate dob2 = date(dob, "MD19Y")
format dob2 %tdNN/DD/CCYY
drop dob
rename dob2 dob
describe
```

Step 8 – From the menu bar, save your Stata data set using FILE > SAVE AS ...

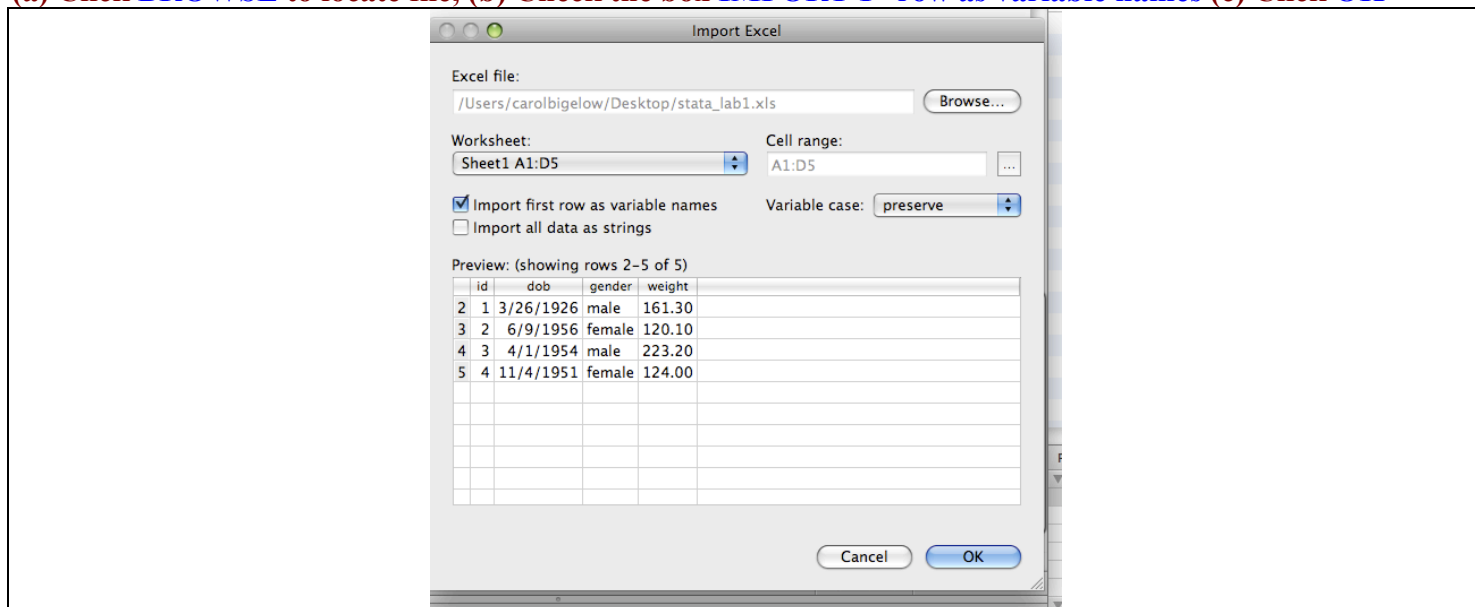


METHOD 2 – Importing the Excel spreadsheet.

Tip – Before you do this, make sure that you have previously formatted (and saved) your data types in Excel! See METHOD 1.

Step 1 – Launch Stata. Click on FILE > IMPORT > EXCEL SPREADSHEET (*.xls, *.xlsx)**Step 2 –**

(a) Click BROWSE to locate file, (b) Check the box IMPORT 1st row as variable names (c) Click OK



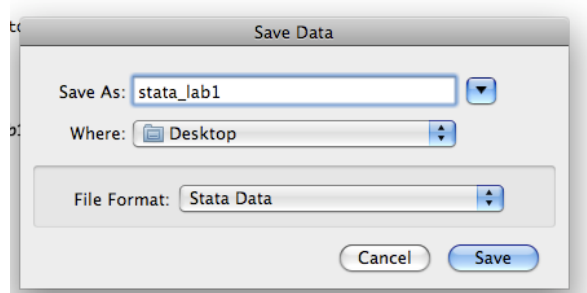
You should see the following (but with your path and name, not mine) in your results window

```
. import excel "/Users/carolbigelow/Desktop/stata_lab1.xls", sheet("Sheet1") firstrow
```

Step 3 – In the command window, issue the following commands to convert the Excel date variable (that fails to import as a date variable) into a Stata date variable that is a bona fide date variable. Eg; -

```
* For dates with year recorded in 2 digits, all in the 1900's:
* Use the function date(stringvariable, "MD19Y") with option "MD19Y" in quotes.
generate dob2 = date(dob, "MD19Y")
format dob2 %tdNN/DD/CCYY
drop dob
rename dob2 dob
describe
```

Step 4 – From the menu bar, save your Stata data set using FILE > SAVE AS ...



YOUR TURN – Create an Excel Data set. Bring it into Stata by method 1 or 2. Save.

- __ 1. Launch Excel.
- __ 2. Create an Excel data set called stata_lab1.xls
- __ 3. Create a Stata data set stata_lab1.dta using your excel data.
- __ 4. Save your Stata data set.

Here is the excel data for you. Note that the dob variable has year in 2 digits, not 4 digits.

id type: numeric	dob type: date	gender type: string/character	weight type: numeric
1	3/26/26	male	161.3
2	6/9/56	female	120.1
3	4/1/54	male	223.2
4	11/4/51	female	124.0

3. Import a Stata Data Set Directly from the Internet

When to Import a Stata Data Set

You will do this often for class and perhaps not so often in your work. **Tips** – (1) Be sure to enclose the url in quotes, (2) Be sure to use the option **clear** after the comma; and (3) Be sure to save the data onto your computer so that you have it for your use later.

Use the command **use** to import a stata data set. Be sure to enclose the full url path in quotes. The basic command is of the following form and is issued in the command window

use “http://fullurlpath”, clear

To save the data onto your computer, from the top menu bar issue:

FILE > SAVE AS ..

Tip!! Be sure to include the extension “.dta” in the name of the data set; see examples below.

Examples -

use “http://www.pauldickman.com/survival/ivf.dta”, clear
use “http://people.umass.edu/biep640w/datasets/week02.dta”, clear
use “http://people.umass.edu/biep640w/datasets/larvae.dta”, clear

YOUR TURN – Import ivf.dta from the internet

- __ 1. Launch Stata (if you have not already done so)
- __ 2. In the command window, type: **set more off**
- __ 3. In the command window, type: **use “http://www.pauldickman.com/survival/ivf.dta”, clear**
- __ 4. Use FILE > SAVE AS... to save it to your computer

4. Importing Excel Dates into Stata

Suppose you have the following messy excel file that you'd like to import into Stata.

Problem! Stata is not so kind in how it handles incoming dates data.

studentid	dob	gender	age	d_intake
1	11/4/25	female	90.10	6/1/15
2	3/28/26	male	89.80	10/24/99
3	4/1/54	male	61.50	8/4/15
4	5/1/01	female	14.60	11/6/15

There are 5 variables

- studentid is *integer*
- dob is a *date* variable
- gender is a *text* variable
- age is a *numeric* variable
- d_intake is another *date* variable

Step 1 – Excel

- ___ a. For each column of data separately, use FORMAT > CELLS to format the data in that column.
Use the selections shown below

Variable	FORMAT > CELLS. Then, under CATEGORY: choose format cells as
studentid	numeric – then choose 0 places after the decimal point
dob	custom – then choose the type: m/d/yy
gender	text
age	numeric - then choose 2 places after the decimal point
d_intake	custom – then choose the type: m/d/yy

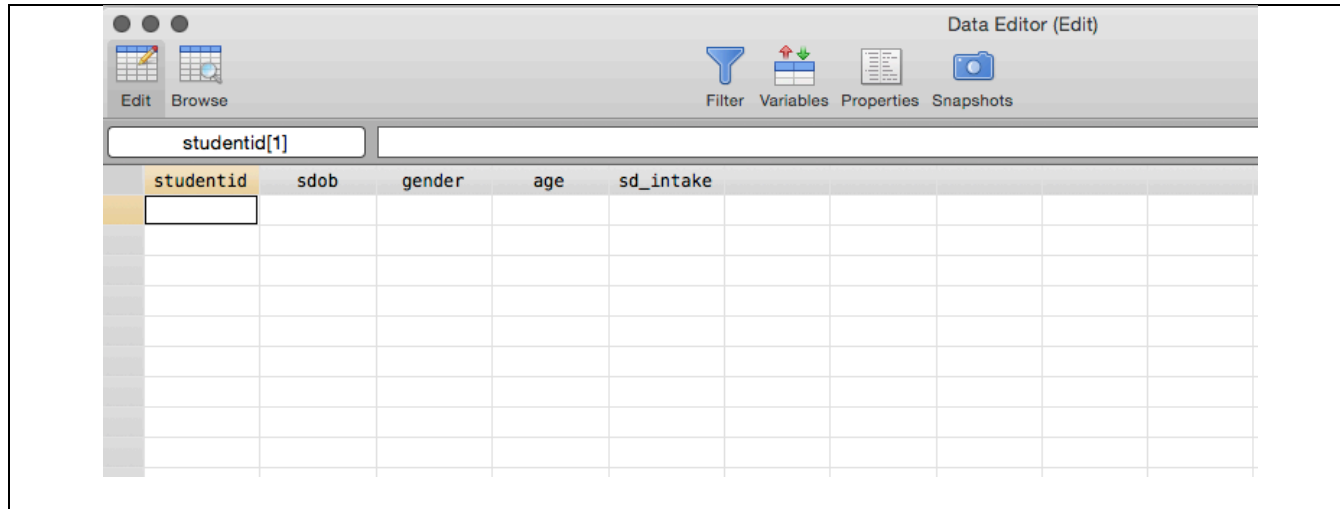
- ___ b. SAVE your excel file, but leave it open. Minimize.

Step 2 – Stata

- ___ a. Initialize all your variables, taking care to initialize them correctly according to type.
IMPORTANT – Despite your saving dates in excel as bona fide dates, Stata will input these as string.
Tip – Thus, in my initial import, I give these date variables a slightly different name. See below.

```
. generate studentid=.  
. generate sdob="."  
. generate gender="."  
. generate age=.  
. generate sd_intake="."
```

- ___ b. In Stata, click on the DATA EDITOR icon located on the top, horizontal, menu bar.
This will bring you to an empty spread sheet with the correct column headings (except that we've got those pesky string variables for dates – we'll fix these later)

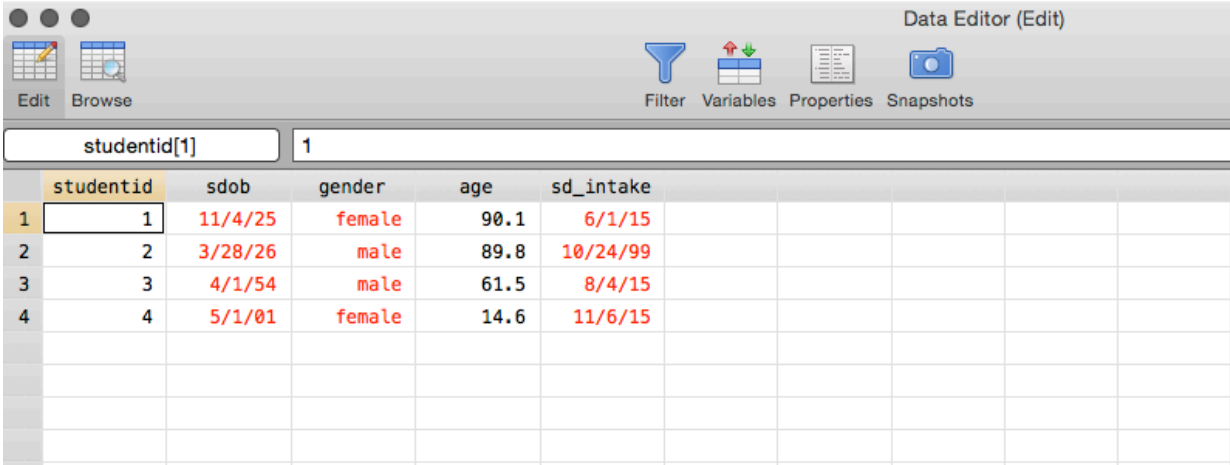


__c. Copy and Paste step:

In EXCEL: Select all your data using EDIT > COPY.

In STATA (be sure your cursor is in row 1 column 1): EDIT > PASTE

You should now see a “populated” spreadsheet. **Note** - The red columns are text data in Stata



	studentid	sdob	gender	age	sd_intake					
1	1	11/4/25	female	90.1	6/1/15					
2	2	3/28/26	male	89.8	10/24/99					
3	3	4/1/54	male	61.5	8/4/15					
4	4	5/1/01	female	14.6	11/6/15					

__d. Close the DATA EDITOR window. No worries. Your data is not lost. It's in active memory.

Step 3 – Create bona fide Stata date variables from the strings predecessors. Format. Check.

```
. ** --- Initially, give all dates century=19
. generate dob=date(sdob,"MD19Y")
. generate d_intake=date(sd_intake,"MD19Y")

.
. ** --- Format for readability
. format dob %tdNN/DD/CCYY
. format d_intake %tdNN/DD/CCYY

.
. ** --- First look at dates to illustrate that all is century=19
. list studentid dob d_intake
```

	studentid	dob	d_intake
1.	1	11/04/1925	06/01/1915
2.	2	03/28/1926	10/24/1999
3.	3	04/01/1954	08/04/1915
4.	4	05/01/1901	11/06/1915

```
. *--- DOB: Dates of birth 1900-1915 should really be 2000-2015
. generate check1=year(dob)
. replace dob=date(sdob,"MD20Y") if check1<1916
(1 real change made)

. *--- D_INTAKE: Dates of intake 900-1915 should really be 2000-2015
. generate check2=year(d_intake)
. replace d_intake=date(sd_intake,"MD20Y") if check2<1916
(3 real changes made)

. ** --- Final look to be sure all is well.
. list studentid dob d_intake
```

	studentid	dob	d_intake
1.	1	11/04/1925	06/01/2015
2.	2	03/28/1926	10/24/1999
3.	3	04/01/1954	08/04/2015
4.	4	05/01/2001	11/06/2015

Step 4 – Last but not least, drop the no longer needed variables and save your Stata data set.

```
. ** All is well. Drop the variables not needed anymore
. drop sdob sd_intake checkyear check2
. list
```

	studentid	gender	age	dob	d_intake
1.	1	female	90.1	11/04/1925	06/01/2015
2.	2	male	89.8	03/28/1926	10/24/1999
3.	3	male	61.5	04/01/1954	08/04/2015
4.	4	female	14.6	05/01/2001	11/06/2015

```
. describe
```

```
Contains data
  obs:      4
 vars:      5
 size:     88
```

variable name	storage type	display format	value label	variable label
studentid	float	%9.0g		
gender	str6	%9s		
age	float	%9.0g		
dob	float	%td..		
d_intake	float	%td..		

```
Sorted by:
```

```
Note: Dataset has changed since last saved.
```

```
. ** Don't forget to save your data
. save "/Users/cbigelow/Desktop/excel_to_stata dates.dta"
file /Users/cbigelow/Desktop/excel_to_stata dates.dta saved
```