

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2006, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Some Useful Stata Information

John McGready

The Johns Hopkins University

Lecture Topics

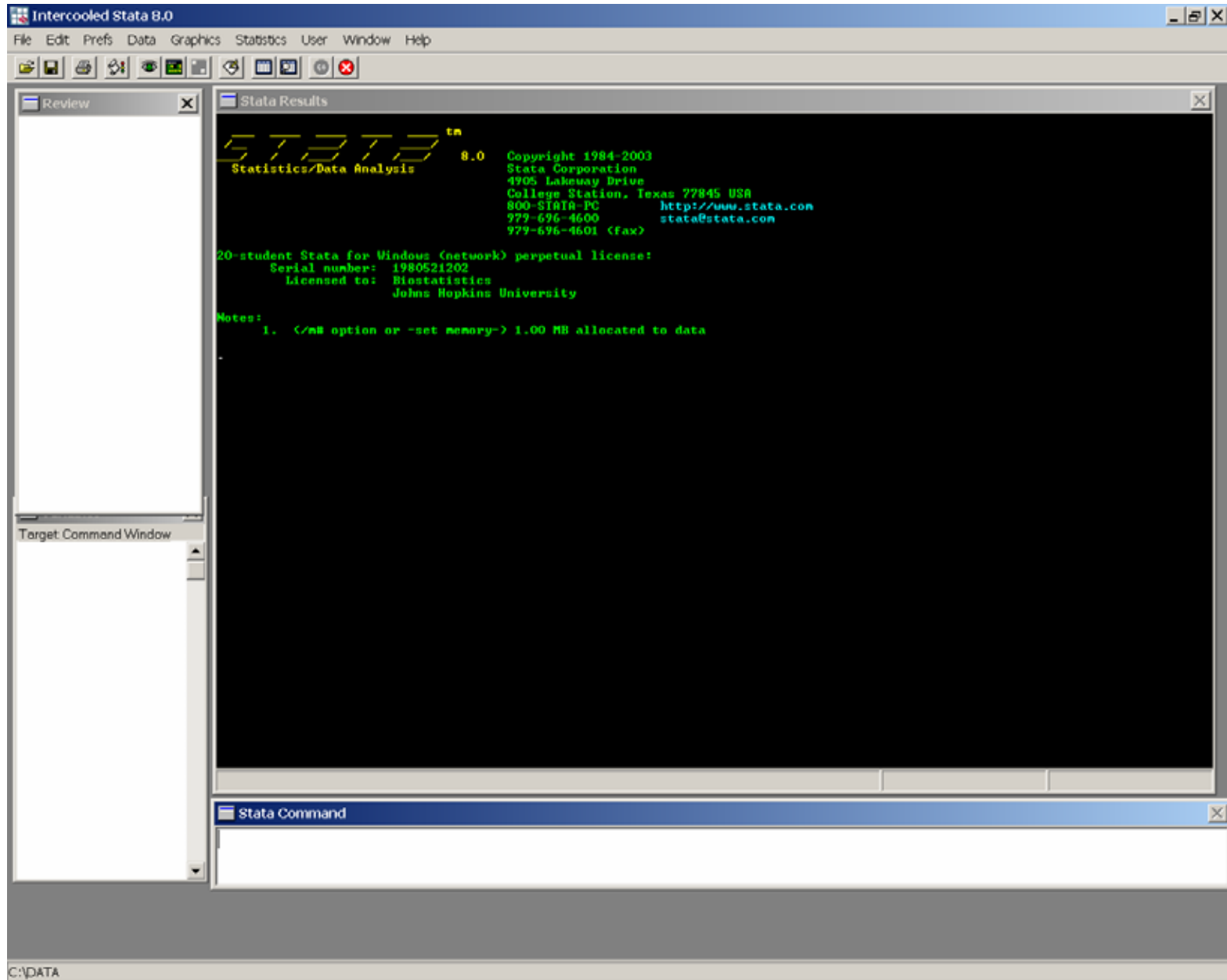
- ◆ Stata's appearance
- ◆ Entering data and holding on to it
- ◆ Re-visiting and updating Stata data sets
- ◆ Syntax diagrams and "immediate commands"

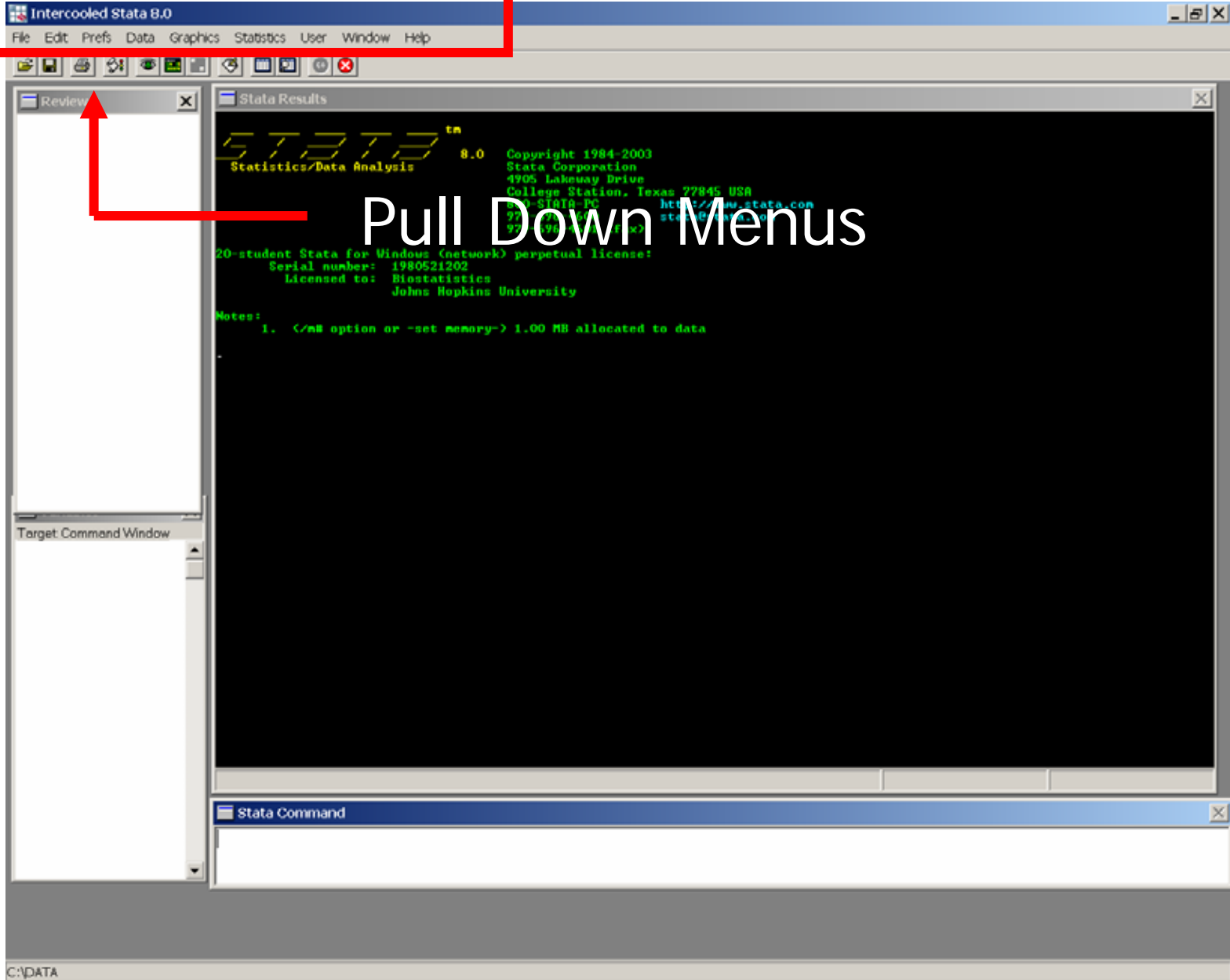


JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

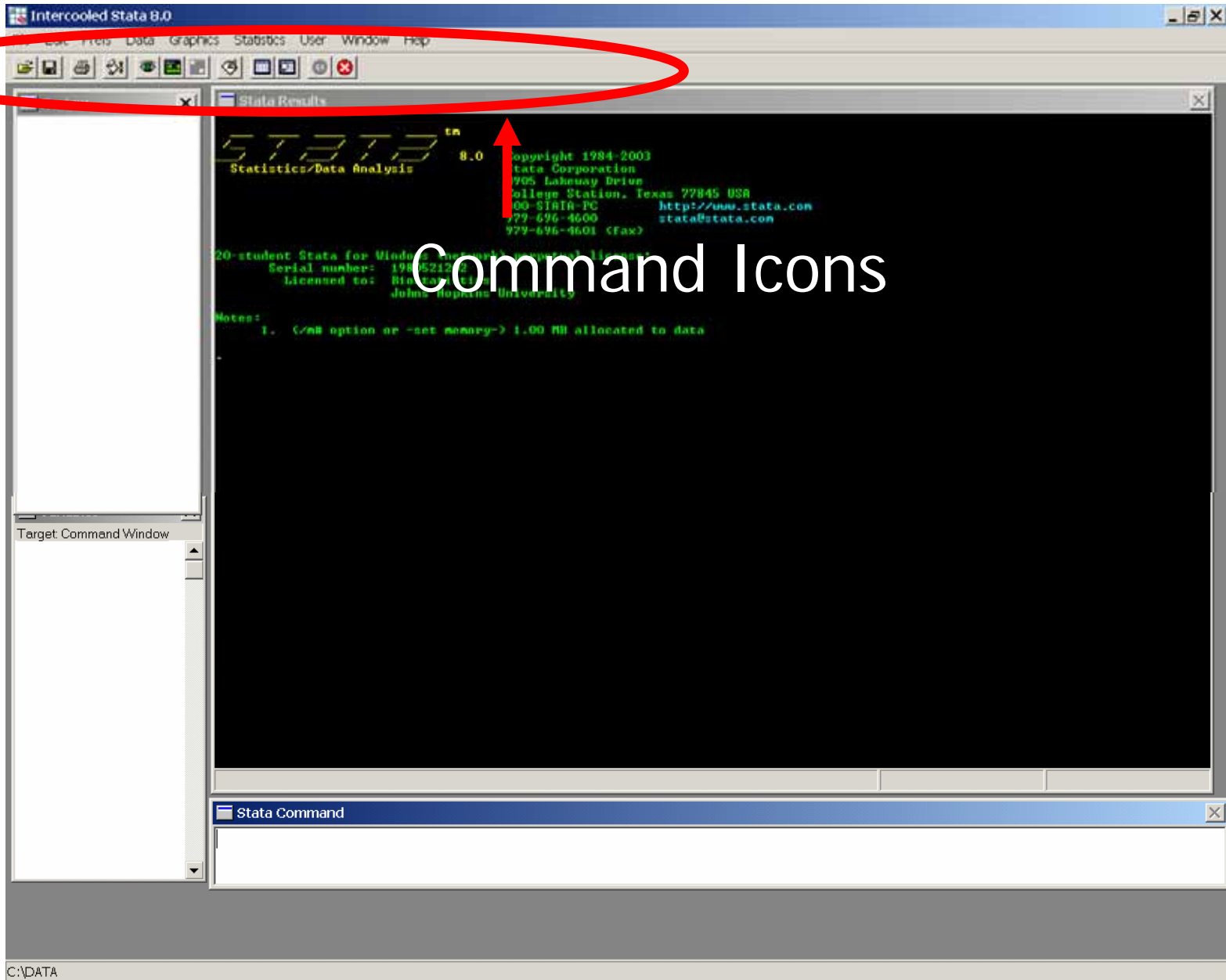
Section A

What Does Stata Look Like?





Pull Down Menus

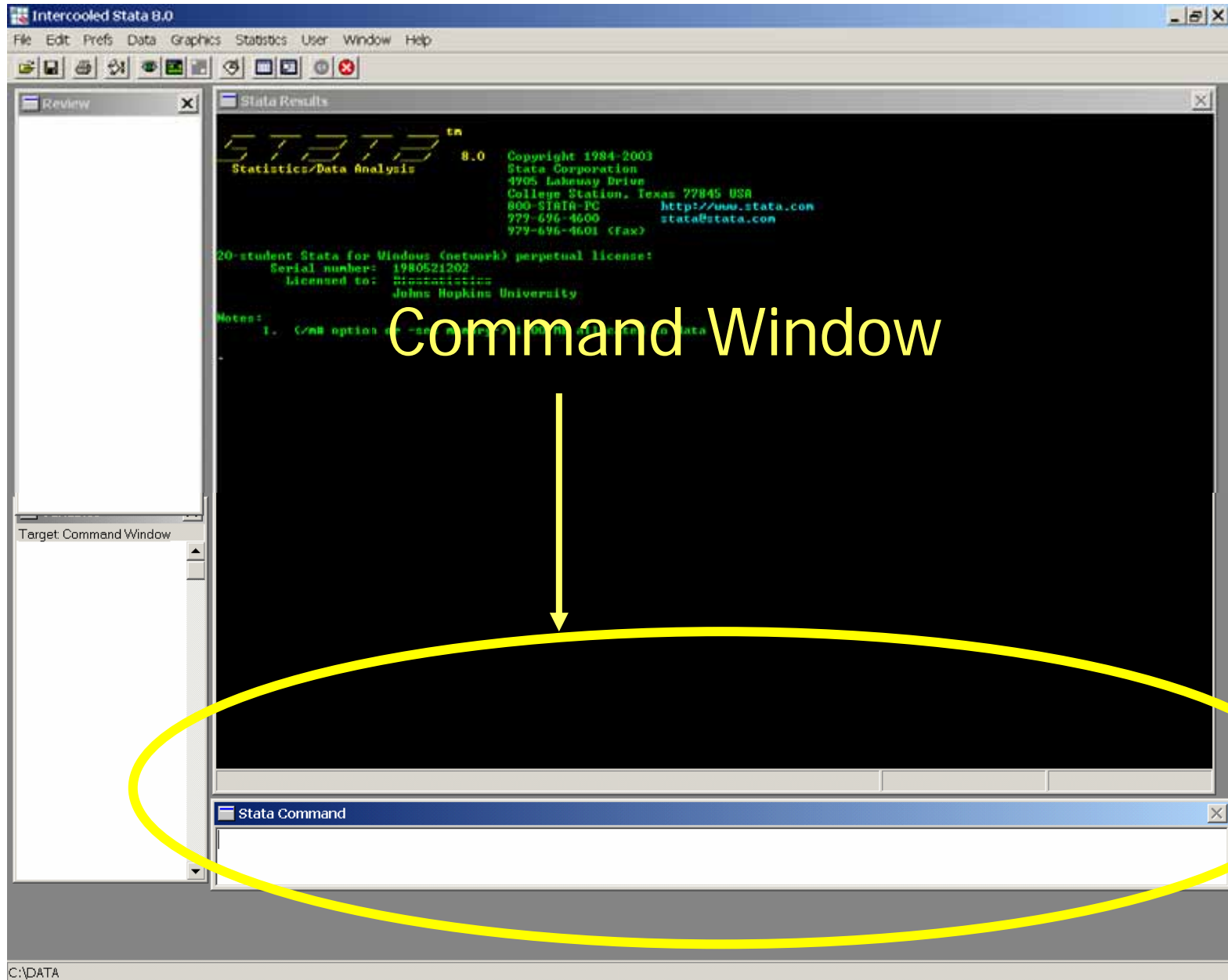


Command Icons

The Windows

1. Command

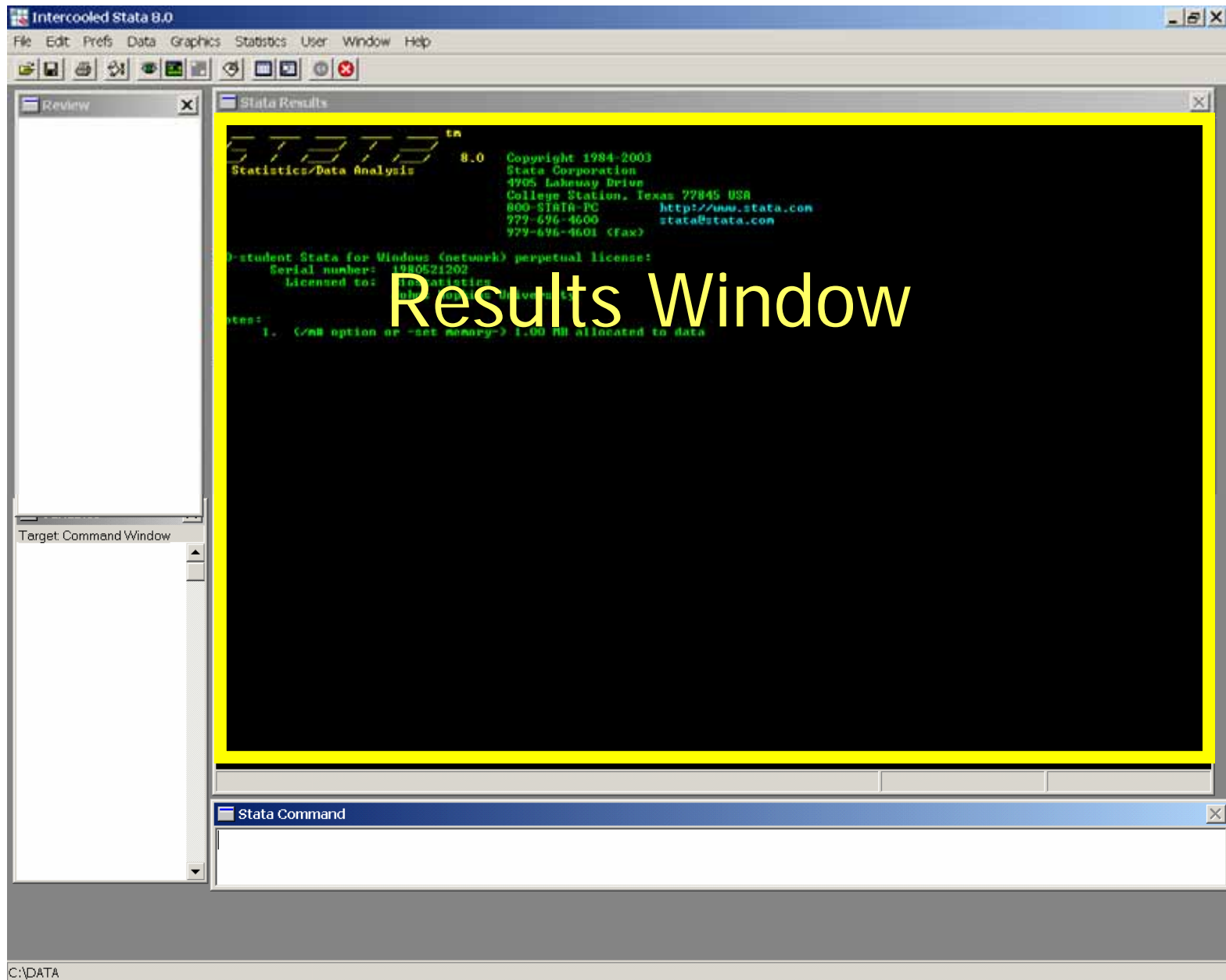
- Window that allows user to input commands
- Stata is a “commands driven” program—commands are entered individually as opposed to entered as a group (batch submission)



The Windows

2. Results

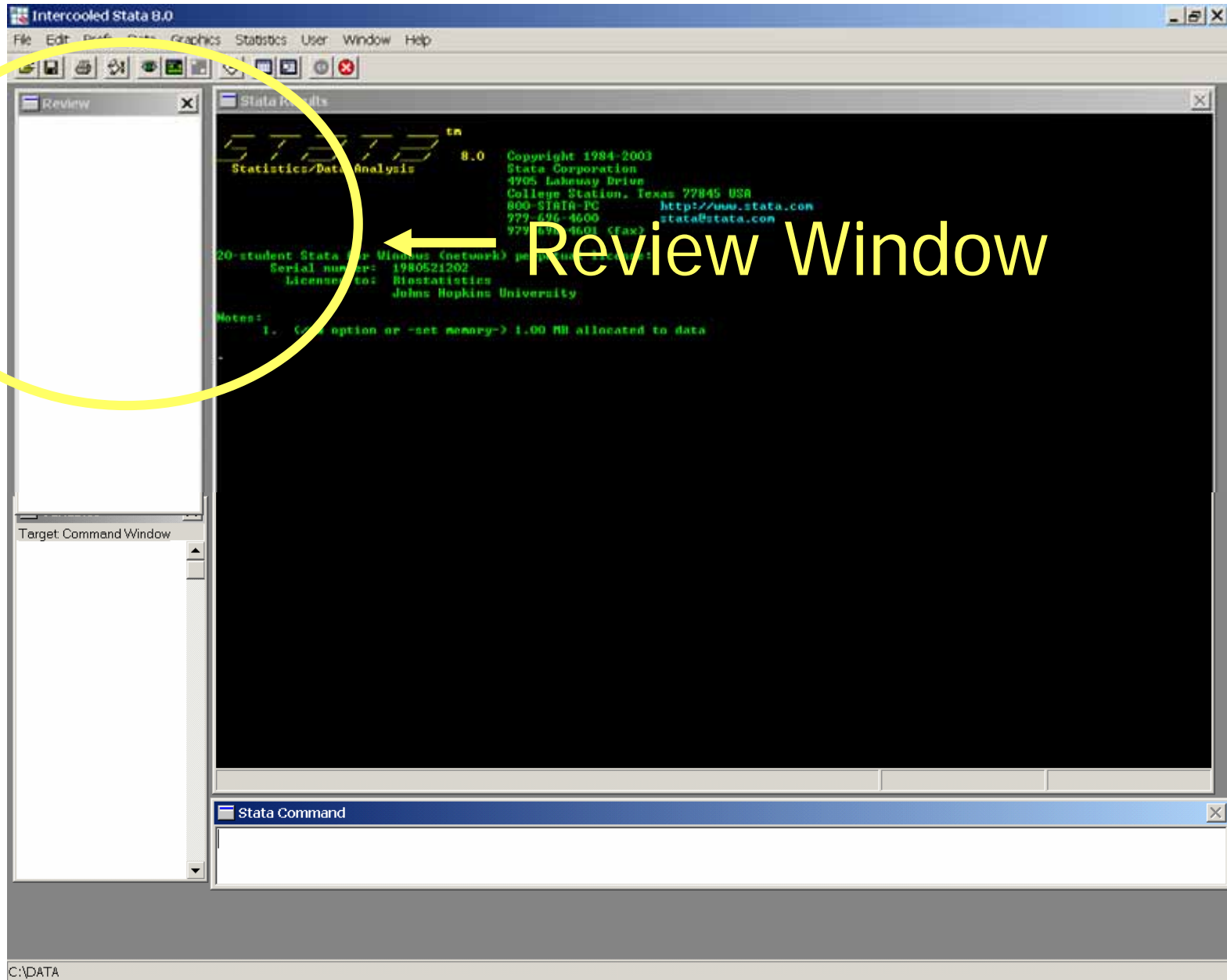
- Window that displays all Stata output
- When the users enter a command that produces output, this output will appear in the “results” window



The Windows

3. Review

- This window keeps a “running tab” of all commands used in current Stata session

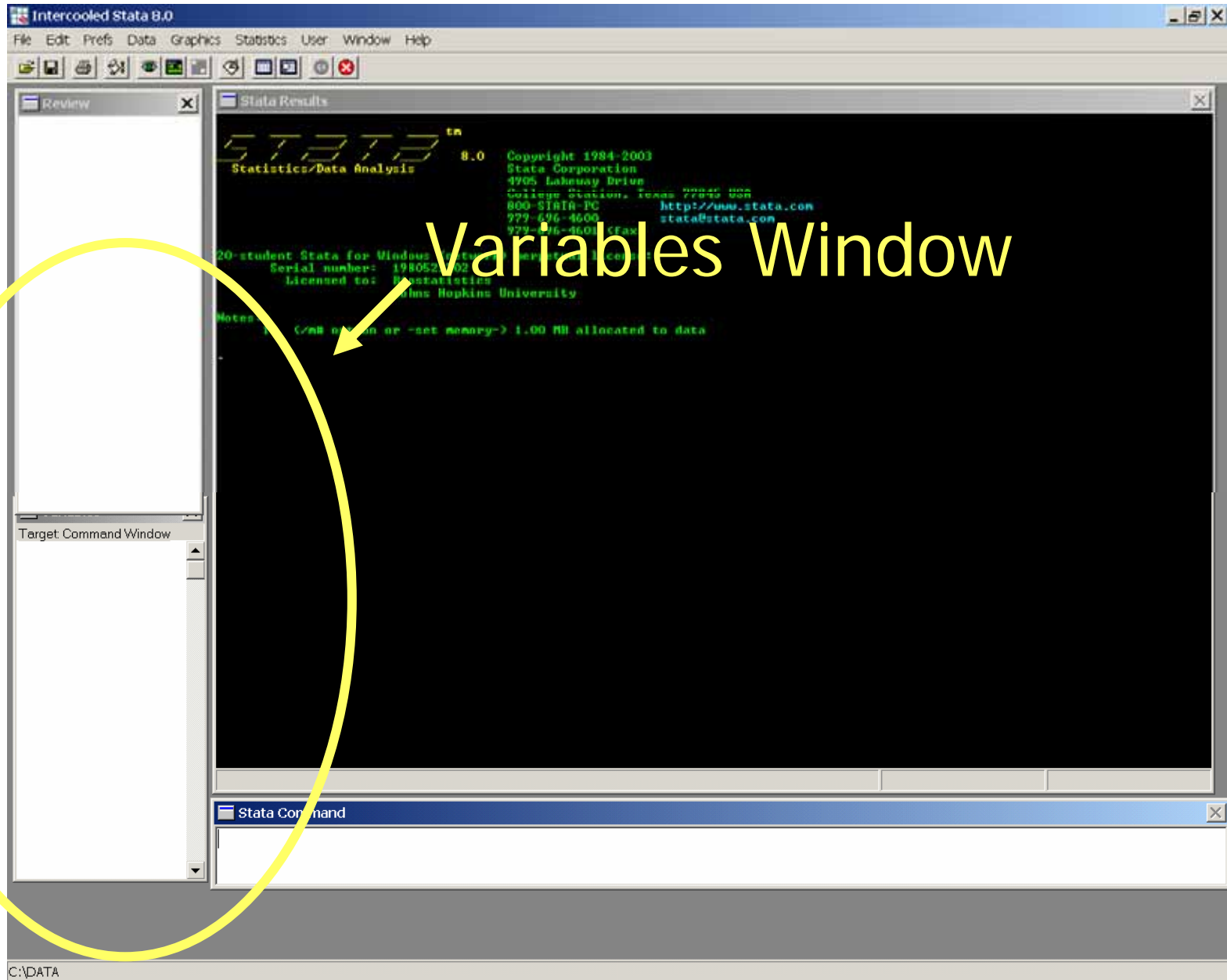


Review Window

The Windows

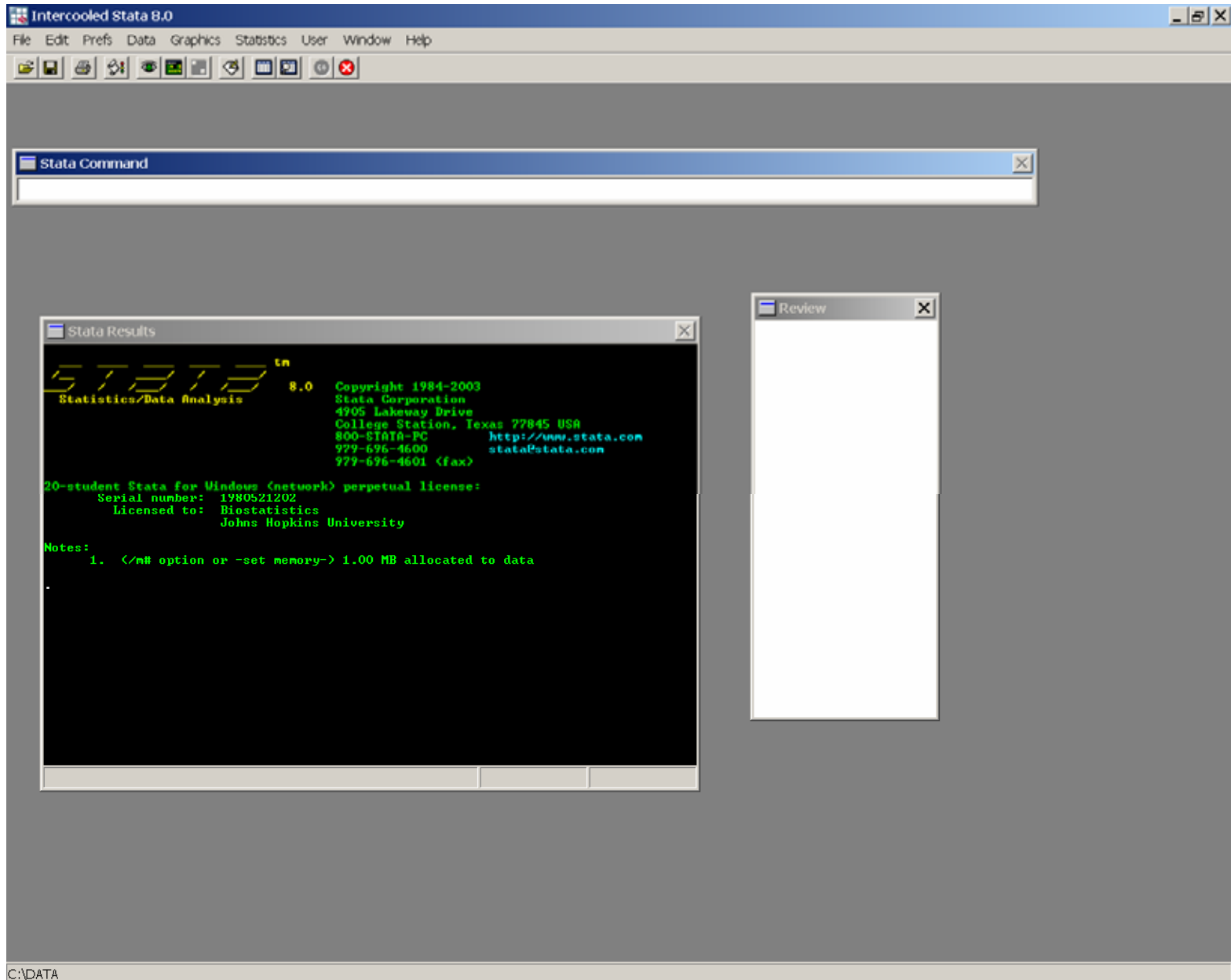
4. Variables

- Displays all variables in data set used in the Stata session.

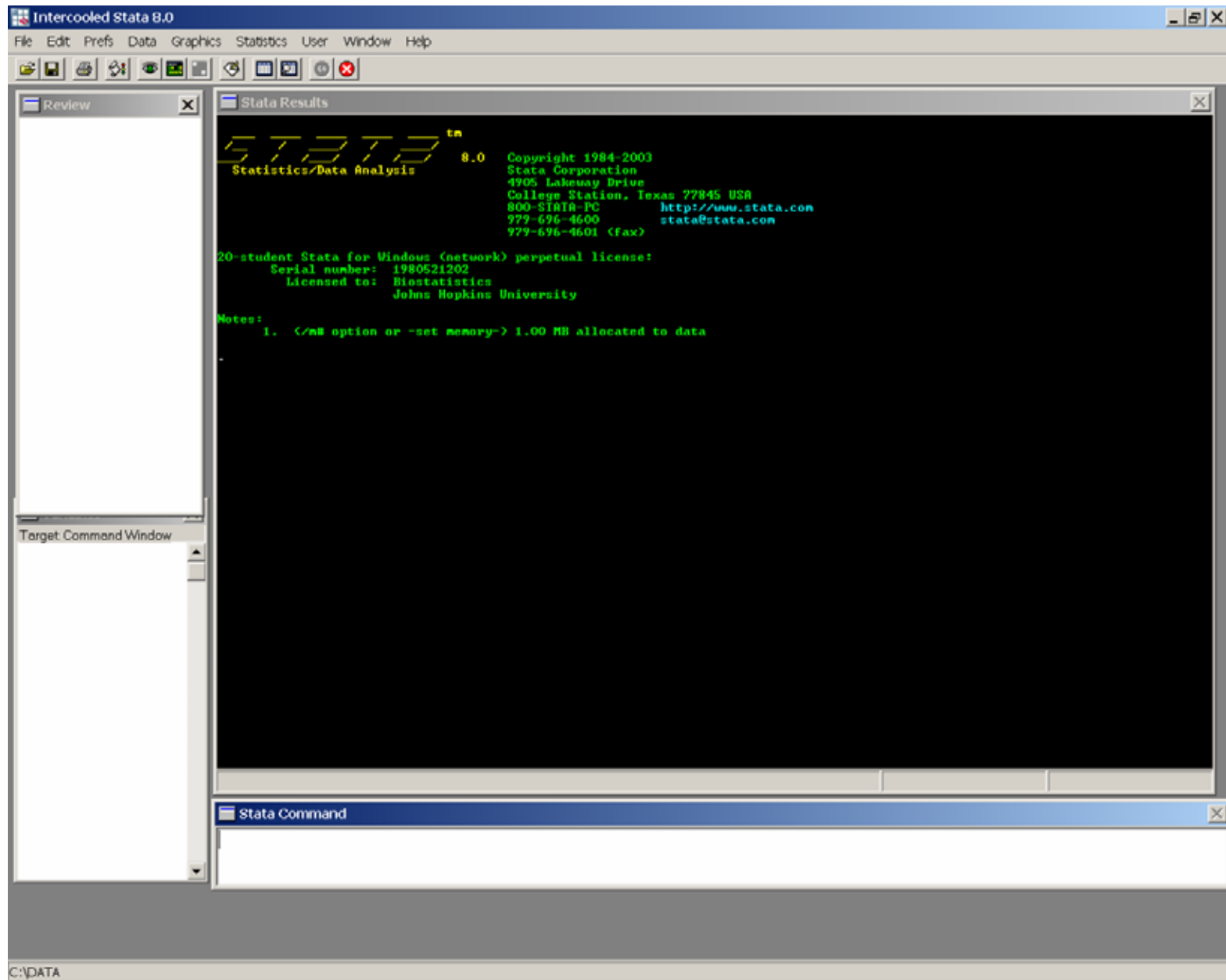


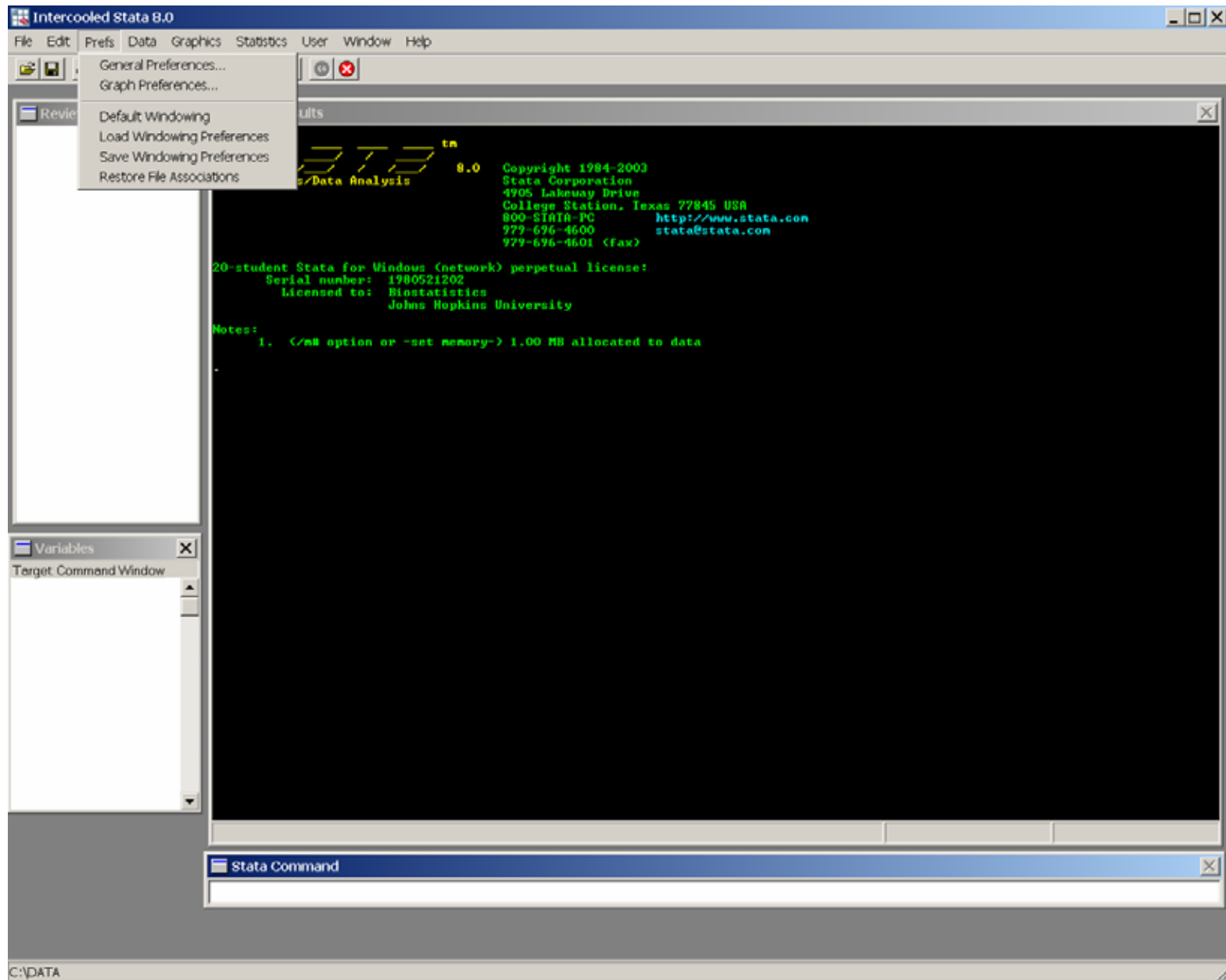
The Screen

- ◆ You, the user, can customize the screen to your liking
- ◆ You can designate which of the four windows you want to have open on the screen, and size up each window to suit your preferences



C:\DATA







JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section B

Entering and Saving Data

Data Set

- ◆ How can we input this into Stata?

Age (yrs)	Weight (lbs)		Age (yrs)	Weight (lbs)
12	123		43	150
65	215		8	58
31	187		76	126
15	89		50	312

Data Set

- ◆ You would like to use Stata to analyze this data set collected from eight subjects

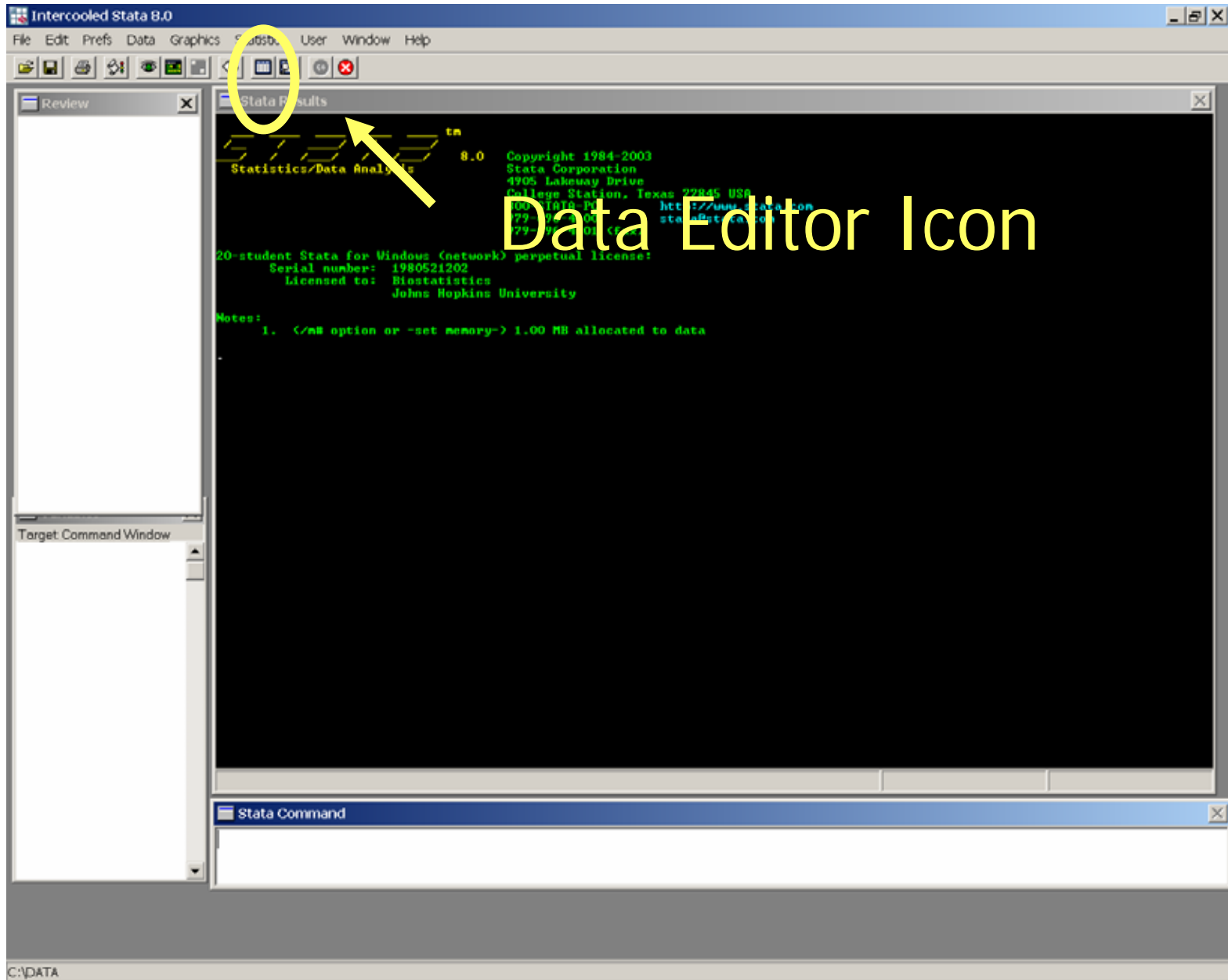
Age (yrs)	Weight (lbs)		Age (yrs)	Weight (lbs)
12	123		43	150
65	215		8	58
31	187		76	126
15	89		50	312

Data Editor

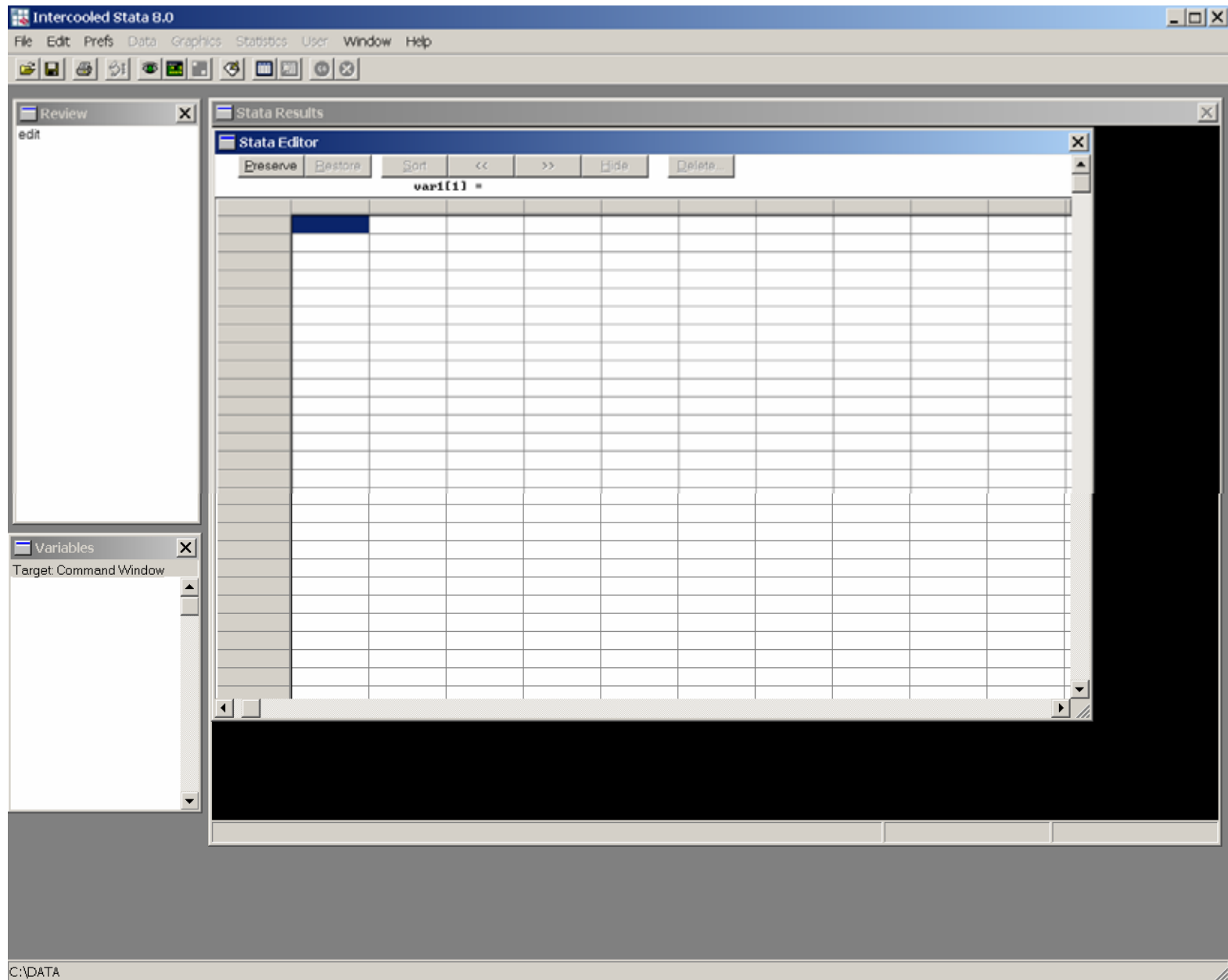
- ◆ Stata has a very nice and easy to use data editor

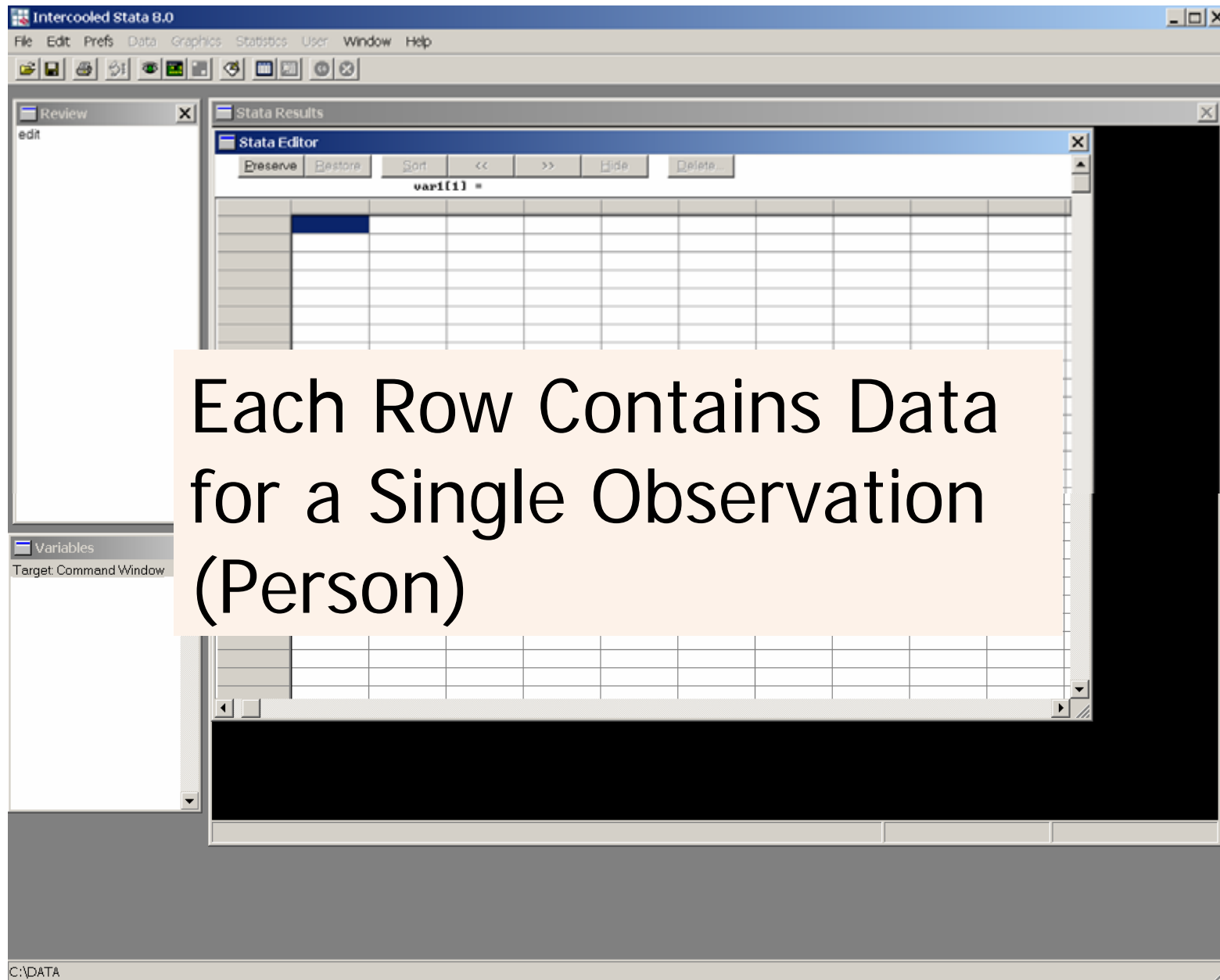
Data Editor

- ◆ To access the data editor, click on the editor icon in the Stata toolbar

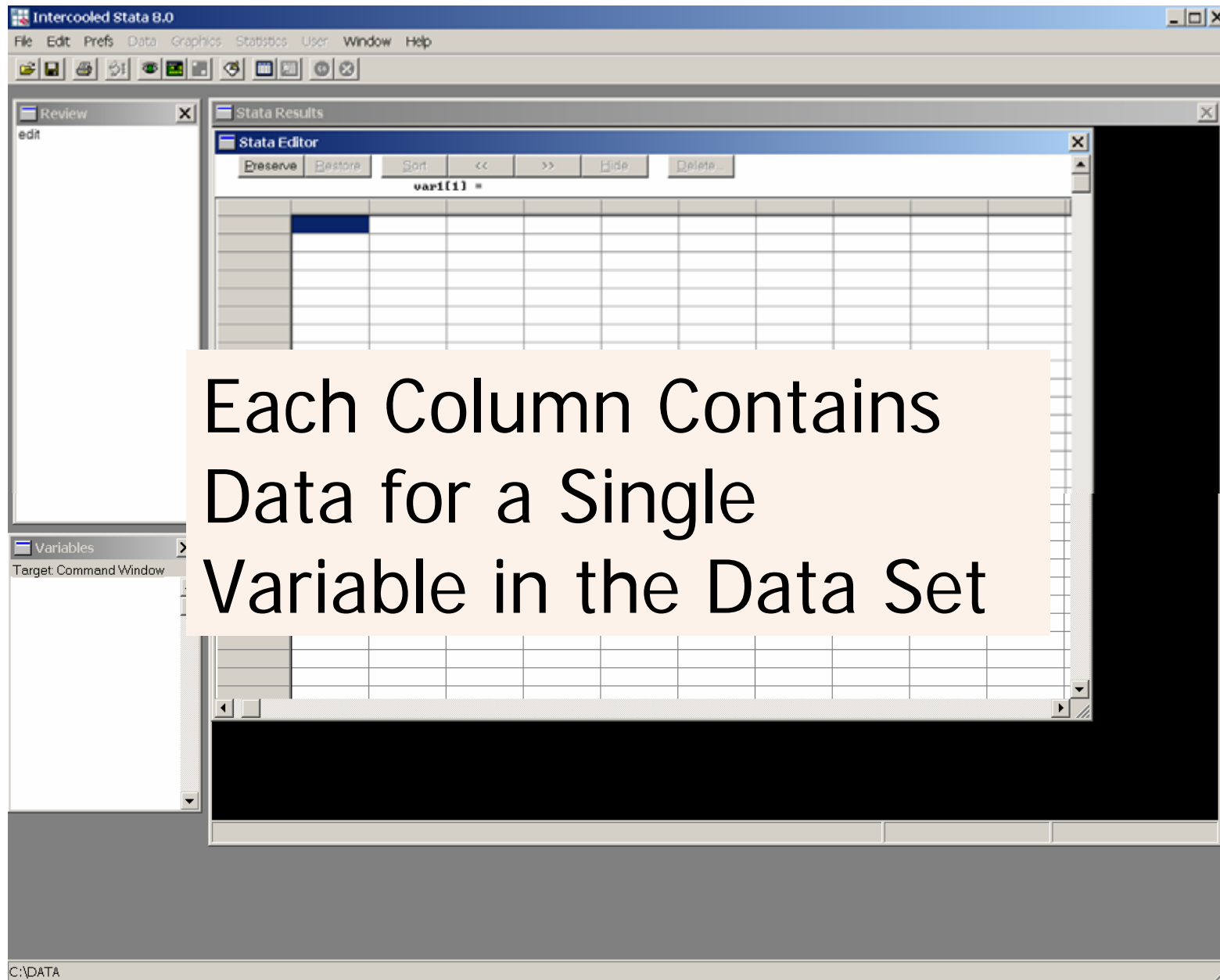


Data Editor Icon





Each Row Contains Data
for a Single Observation
(Person)



Each Column Contains
Data for a Single
Variable in the Data Set

Data Set

- ◆ In this data set, each of the eight subjects is an individual observation

Age (yrs)	Weight (lbs)		Age (yrs)	Weight (lbs)
12	123		43	150
65	215		8	58
31	187		76	126
15	89		50	312

Data Set

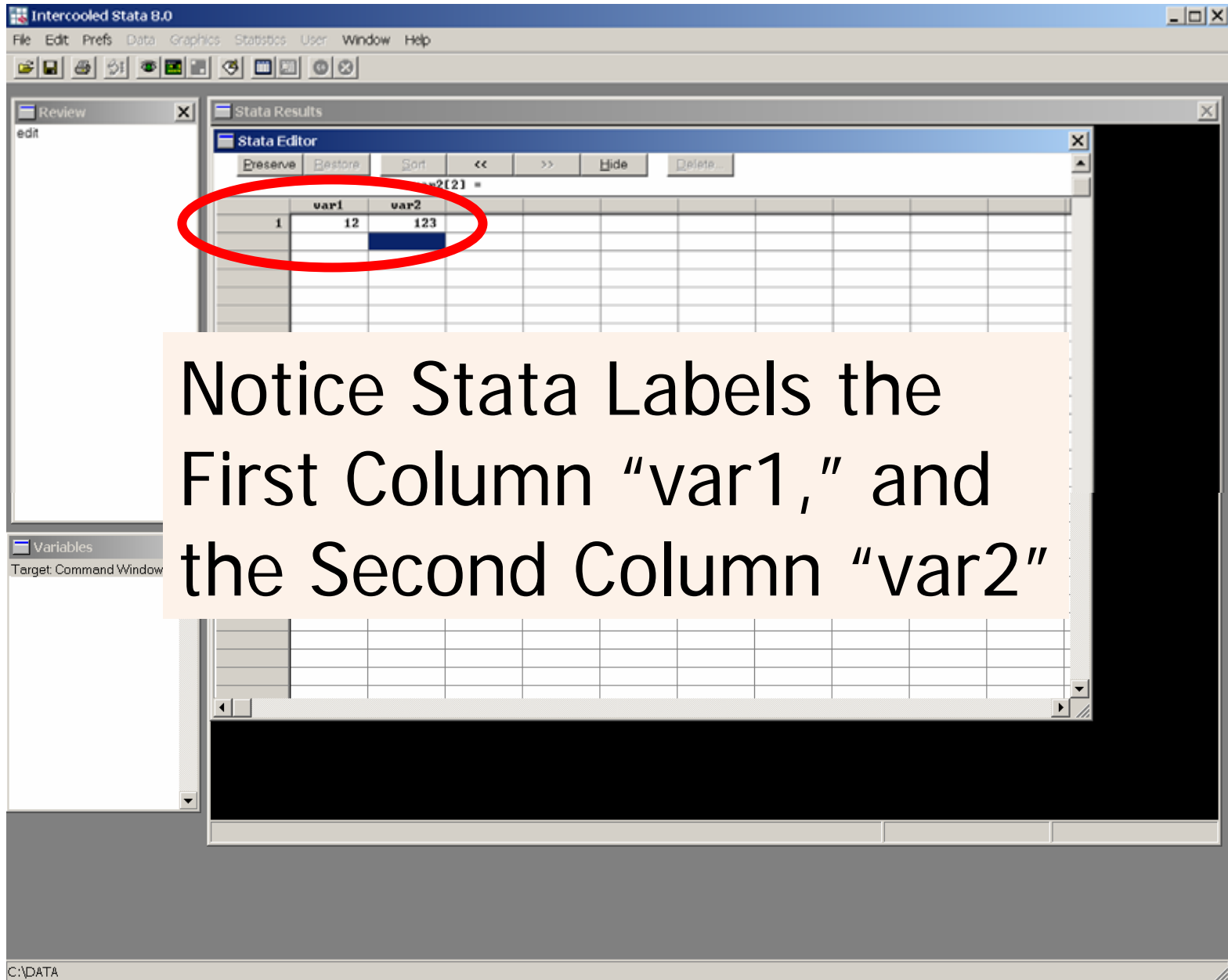
- ◆ Each observation consists of two variables, **age** and **weight**

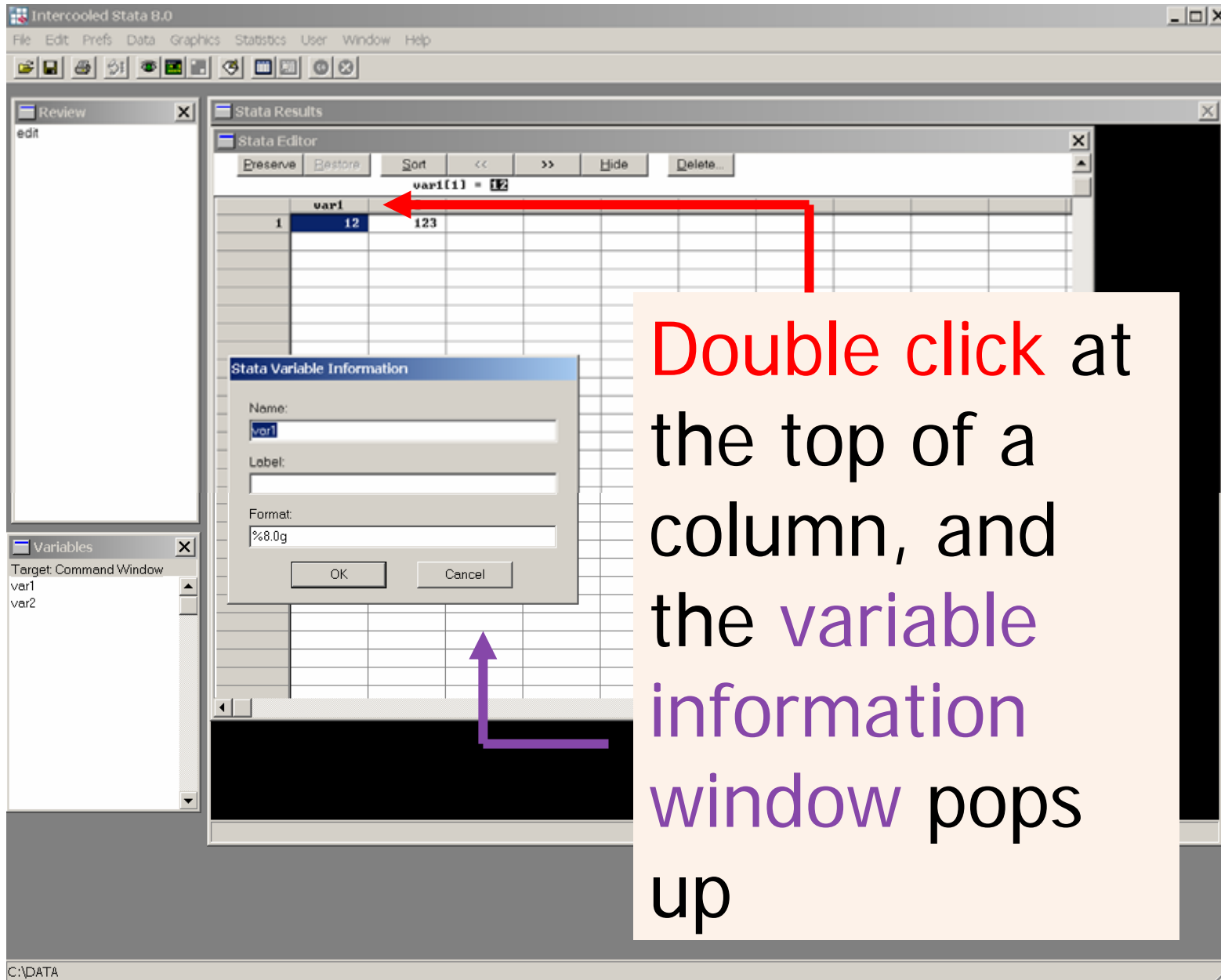
Age (yrs)	Weight (lbs)		Age (yrs)	Weight (lbs)
12	123		43	15
65	215		8	58
31	187		76	126
15	89		50	312

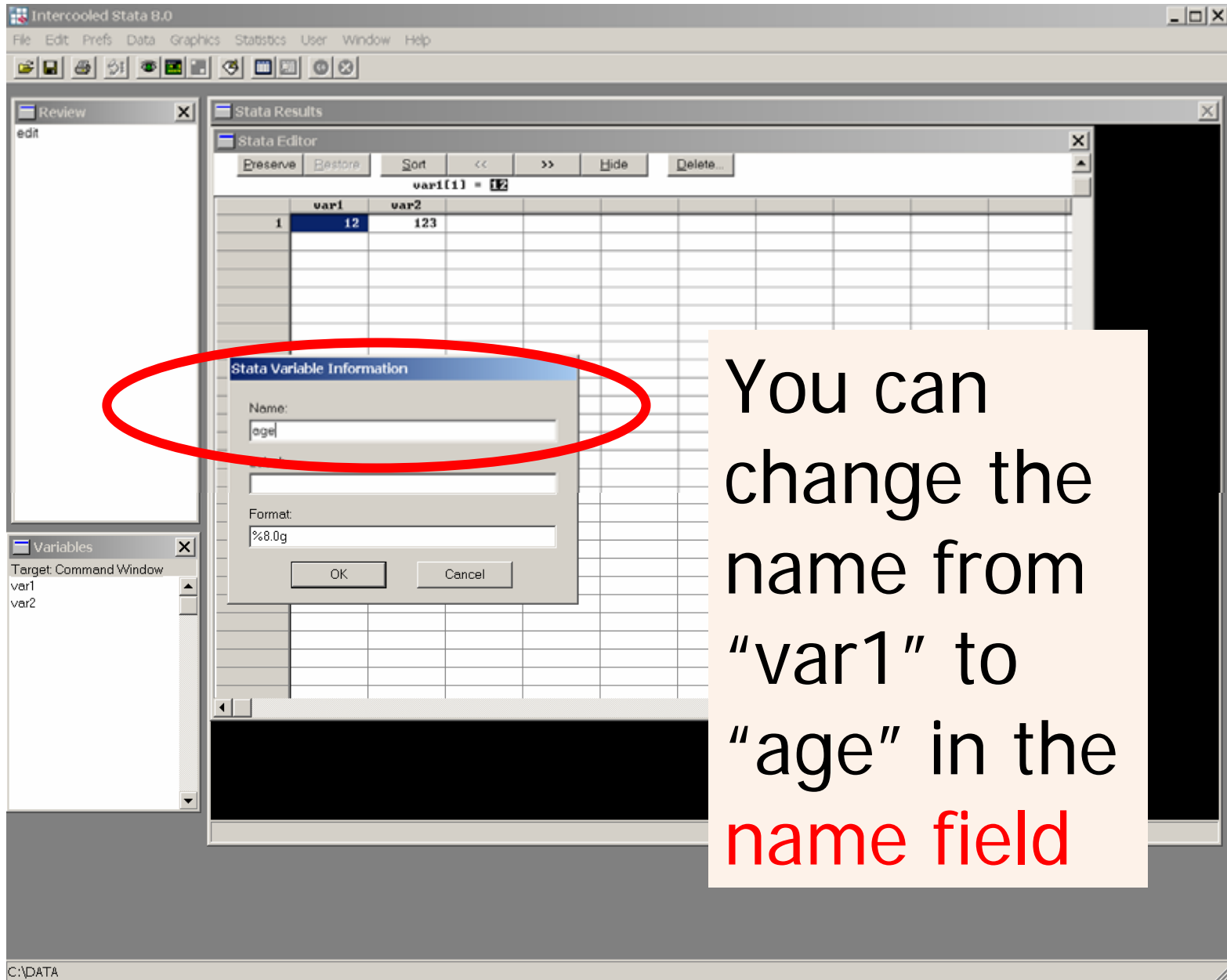
The screenshot displays the Intercooled Stata 8.0 interface. The main window is the Stata Editor, which shows a data editor grid. The first row of data is highlighted, showing the values 12 and 123 for variables var1 and var2, respectively. Two black arrows point from the text box to these values. The interface includes a menu bar (File, Edit, Prefs, Data, Graphics, Statistics, User, Window, Help), a toolbar, and several docked windows: Review, Stata Results, Variables, and Target Command Window. The status bar at the bottom left indicates the current directory is C:\DATA.

	var1	var2
1	12	123

Data (age, weight) on the first subject in the study has been entered in the data editor







You can change the name from "var1" to "age" in the name field

Naming Variables

- ◆ Stata is case sensitive—so far as it is concerned, “age” and “Age” are two unique variable names

Intercooled Stata 8.0

File Edit Prefs Data Graphics Statistics User Window Help

Review

```
edit
clear
use "C:\JMCGRADY\bio611_
edit
```

Stata Results

Stata Editor

age[1] = 12

	age	weight
1	12	123
2	65	215
3	31	187
4	15	89
5	43	150
6	8	58
7	76	126
8	50	312

Variables

Target: Command Window

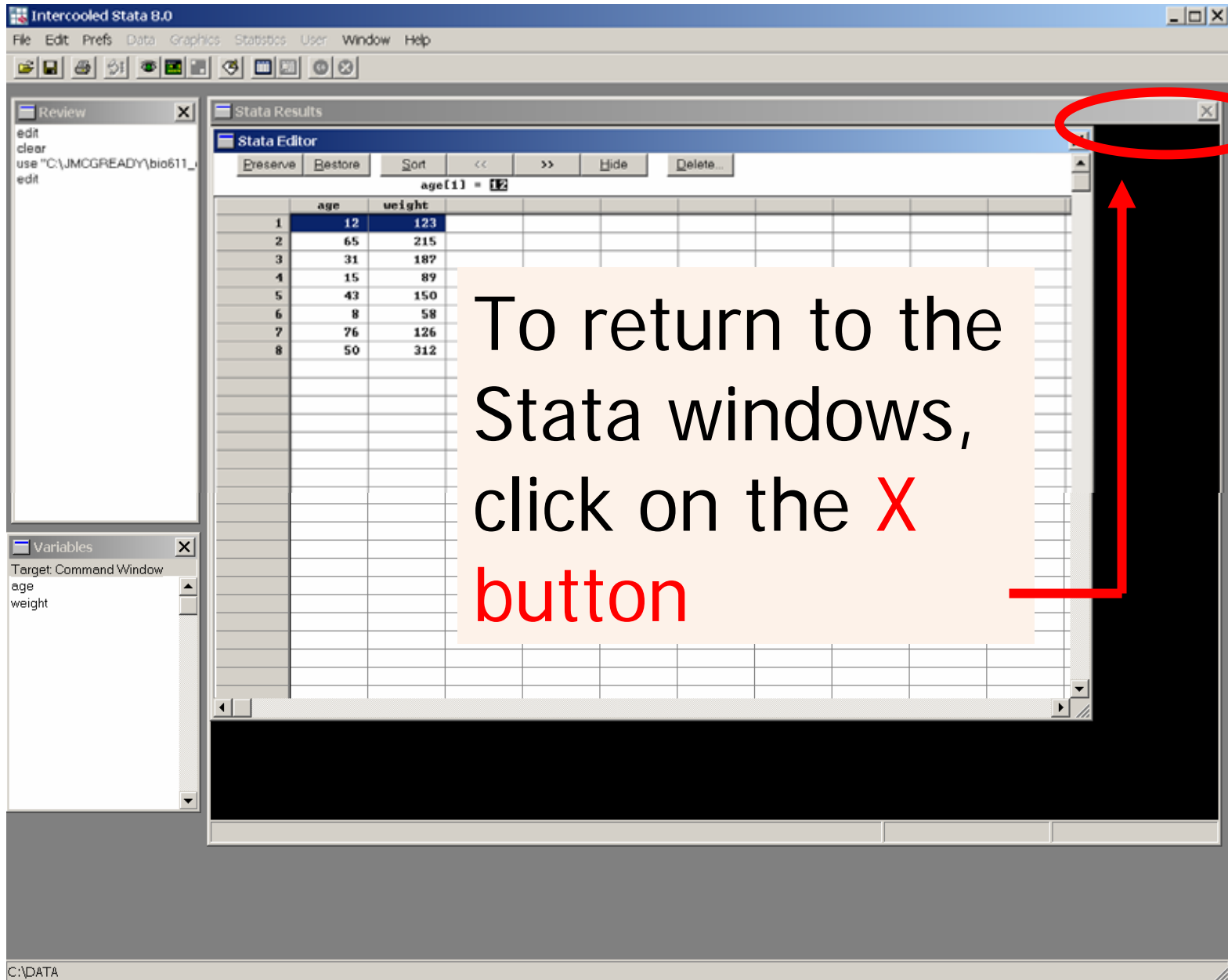
age
weight

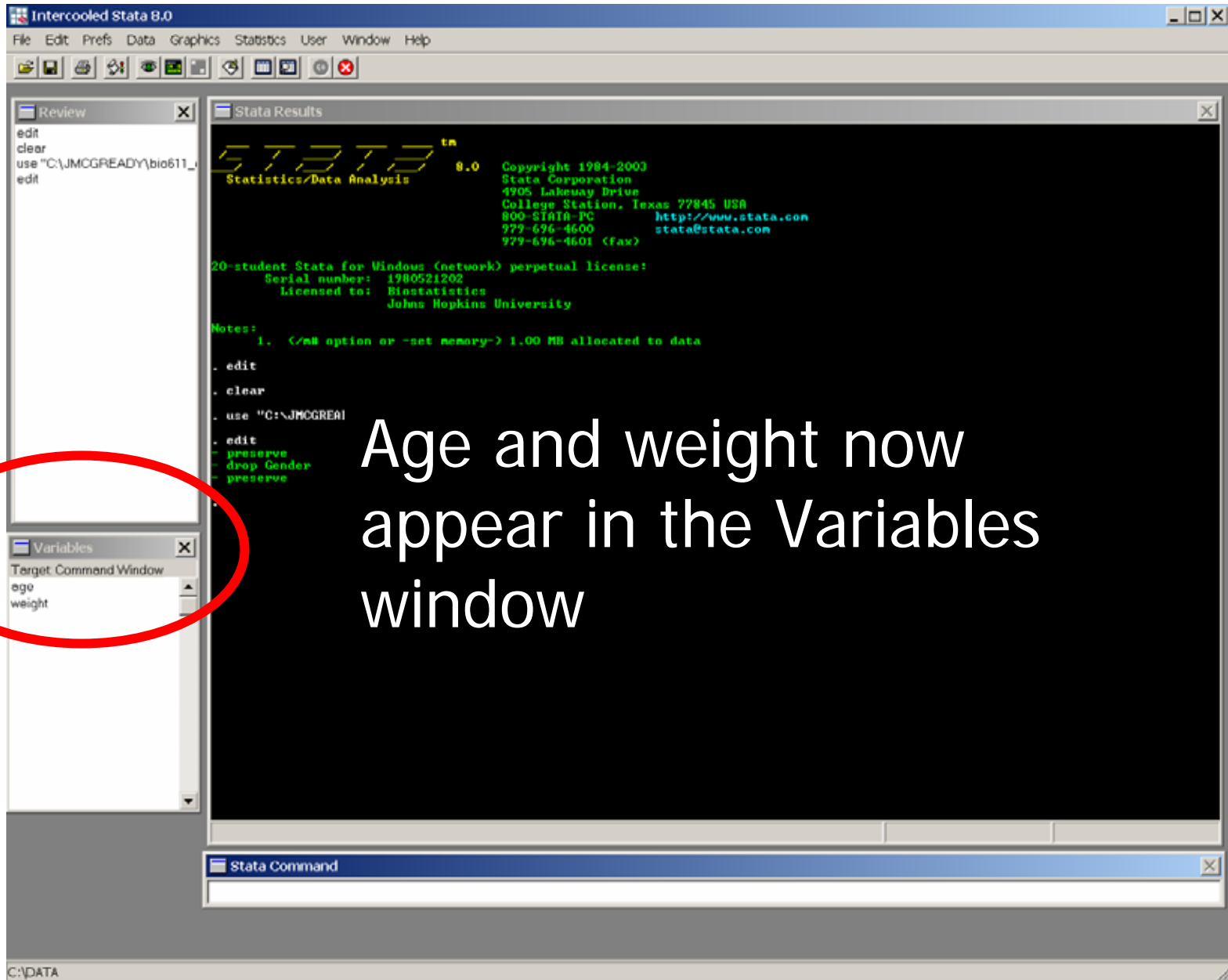
All the data is now entered, and both variables (columns) are renamed

The screenshot shows the Stata 8.0 interface. The main window is the Stata Editor, displaying a data table with columns 'age' and 'weight'. The 'Preserve' button is circled in red, and a red arrow points from it to the data table. A text box on the right explains that to formally put data into Stata's memory, the 'Preserve' button should be clicked.

	age	weight
1	12	123
2	65	215
3	31	187
4	15	89
5	43	150
6	8	58
7	76	126
8	50	312

To formally put the data into Stata's memory, click the **preserve** button





Age and weight now
appear in the Variables
window

The List Command

- ◆ The `list` command allows the user to view the entire data set
- ◆ Individual variables can also be listed—for example, `list age`

The screenshot shows the Stata 8.0 interface. The main window displays the Stata logo and version information. The command window at the bottom contains the command `list`, which is circled in red. The results window shows the output of the `list` command, which is a table of age and weight data for 8 observations.

	age	weight
1.	12	123
2.	65	215
3.	31	187
4.	15	89
5.	43	150
6.	8	58
7.	26	126
8.	50	312

Stata 8.0 Copyright 1984-2003
Stata Corporation
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
779-696-4600 stata@stata.com
779-696-4601 (fax)

20-student Stata for Windows (network) perpetual license:
Serial number: 1980521202
Licensed to: Biostatistics
Johns Hopkins University

Notes:
1. `</n>` option or `-set memory-` 1.00 MB allocated to data

```
. use "C:\JMCGREADY\bio611_online 2003\ppoint\lectures\Stata review\age_weight.dta", clear  
. drop Gender  
. list
```

Type the command
in the command
window and press
"enter"

The screenshot shows the Stata 8.0 software interface. The main window, titled 'Stata Results', displays the following text:

```
-----+-----
Stata 8.0 Copyright 1984-2003
Statistics/Data Analysis Stata Corporation
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
779-696-4600 stata@stata.com
779-696-4601 (fax)

20-student Stata for Windows (network) perpetual license:
Serial number: 1980521202
Licensed to: Biostatistics
Johns Hopkins University

Notes:
1. /n# option or -set memory- 1.00 MB -----+-----

. use "C:\JMCGREADY\bio611_online 2003\ppoint\
. drop Gender
. list
```

The 'list' command output is displayed in a table format:

	age	weight
1.	12	123
2.	65	215
3.	31	187
4.	15	89
5.	43	150
6.	8	58
7.	26	126
8.	50	312

The 'Stata Command' window at the bottom shows the command 'list' entered. The 'Review' window on the left shows the commands: 'use "C:\JMCGREADY\bio611_', 'drop Gender', and 'list'. The 'Variables' window shows the variables 'age' and 'weight'. The status bar at the bottom indicates 'C:\DATA', 'Slide 22 of 41', 'Default Design', and 'English (U.S.)'.

Results displayed in "Results" window

The List Command

- ◆ A close-up of what appears in the review window after using the **list** command

```
. list
```

	age	weight
1.	12	123
2.	65	215
3.	31	187
4.	15	89
5.	43	150
6.	8	58
7.	76	126
8.	50	312

Intercooled Stata 10.0

File Edit Prefs Data Graphics Statistics User Window Help

Review

```
use "C:\JMCGREADY\bio611_
drop Gender
list
```

Stata Results

```
*****
                |
                |      Statistics
                |_____
                |
                |      20-student Sta
                |      Serial
                |      Licen
                |
                |      Notes:
                |      1.  C:\nd
                |
                |      . use "C:\JMCG
                |      drop Gender
                |      list
                |
                |      age
                |_____
                |
                |      1.      12
                |      2.      65
                |      3.      31
                |      4.      15
                |      5.      43
                |
                |      6.      8      58
                |      7.      26     126
                |      8.      50     312
```

Variables

Target: Command Window

age
weight
Gender

Stata Command

C:\DATA

Slide 25 of 41 Default Design English (U.S.)

Notice, the **list** command now appears in the Review window

Describe Command

- ◆ The `describe` command returns output giving the user a general overview of the data set
- ◆ The `describe` command can also be used to profile individual variable (for example, `describe age`)

Describe Command

- ◆ A close up of the review window after using the **describe** command

```
. describe
Contains data from C:\JMCGREADY\bio611_online 2003\ppoint\lectures\Stata review\age_weight.dta
  obs:           8
  vars:           2              7 Aug 2000 12:32
  size:          56 <99.9% of memory free>
-----
variable name    storage  display  value  variable label
                type    format
age              byte    %8.0g
weight           int     %8.0g
-----
Sorted by:
Note: dataset has changed since last saved
```

Describe Command

- ◆ A close up of the review window after using the describe command

```
. describe
Contains data from C:\MCGREADY\bio611_online 2003\ppoint\lectures\Stata review\age_weight.dta
obs:      8
vars:     2                               7 Aug 2000 12:32
size:     56 <99.9% of memory free>
-----
variable name  storage  display  value  variable label
                type   format   label
age            byte   %8.0g
weight         int    %8.0g
Sorted by:
Note: dataset has changed since last saved
```

Information on the size of the data set.

Describe Command

- ◆ A close up of the review window after using the describe command

```
. describe  
Contains data from C:\JMCGREADY\bio611_online 2003\ppoint\lectures\Stata review\age_weight.dta  
obs:      8  
vars:      2          7 Aug 2000 12:32  
size:      56 <99.9% of memory free>  
-----  
variable name  storage  display  value  variable label  
              type   format   label  
-----  
age           byte    %8.0g  
weight        int     %8.0g  
-----  
Sorted by:  
Note: dataset has changed since last saved
```

Information on all variables in the data set.

Saving Data

- ◆ Click on the **File** option in the upper left hand corner of your screen

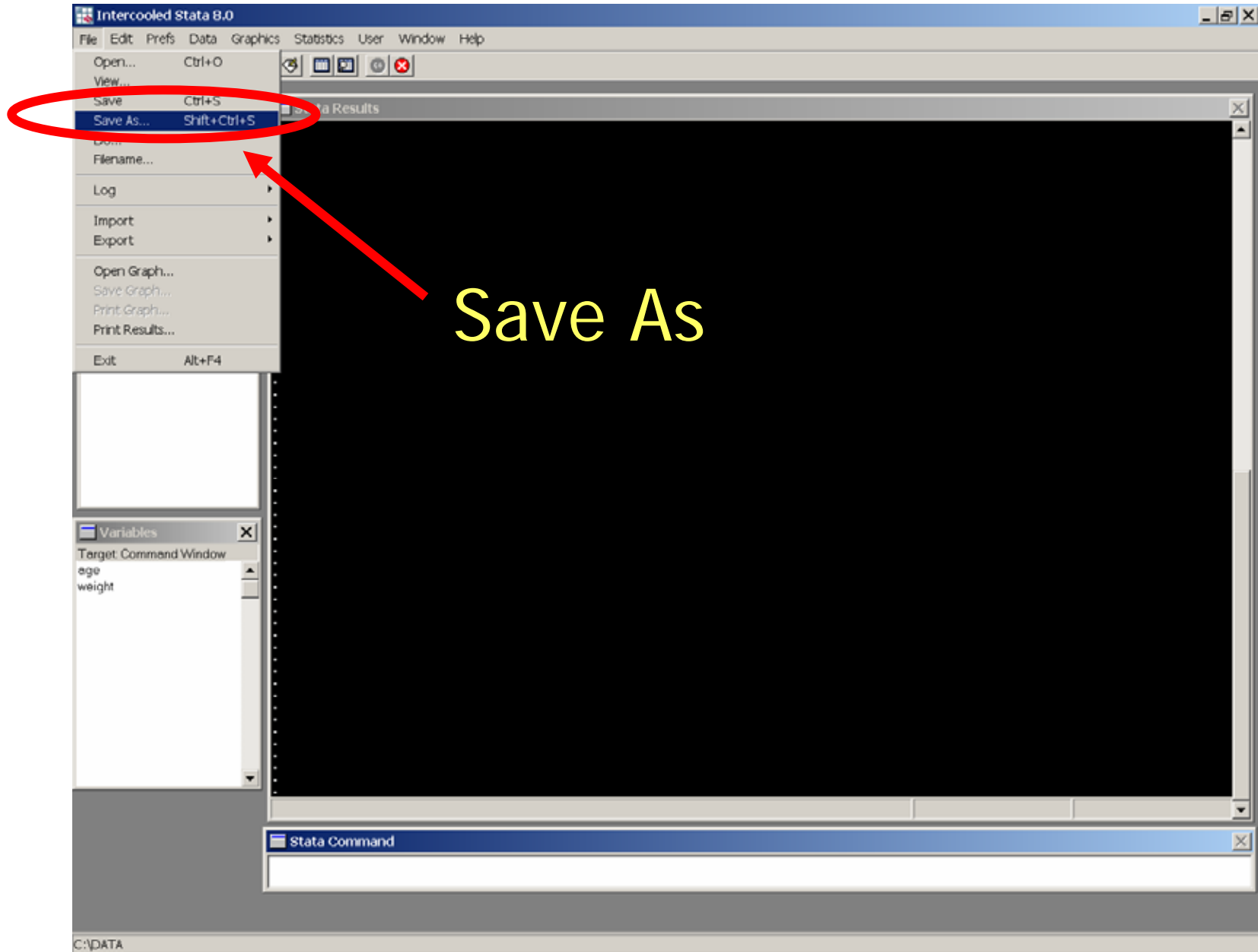
File option

```
use "C:\JMCGRADY\bio611_...", clear
drop Gender
list
```

	age	weight
1.	12	123
2.	65	215
3.	31	187
4.	15	89
5.	43	150
6.	8	58
7.	76	126
8.	50	312

Saving Data

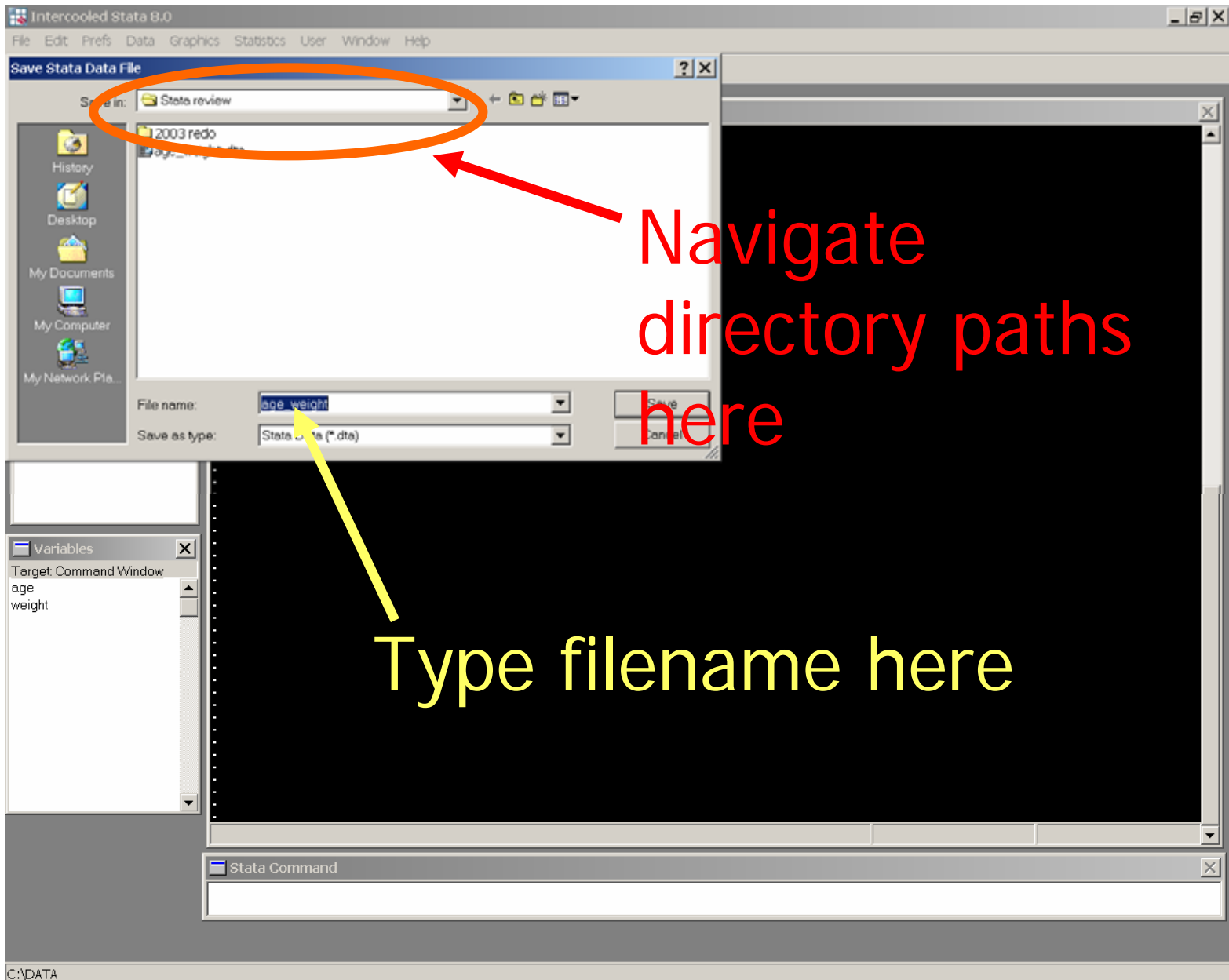
- ◆ You will see a menu appear, choose the “save as” option



Save As

Saving Data

- ◆ You will get a file chooser window
- ◆ Here, you can select the proper directory path (i.e., put the file where you want it) and name the file

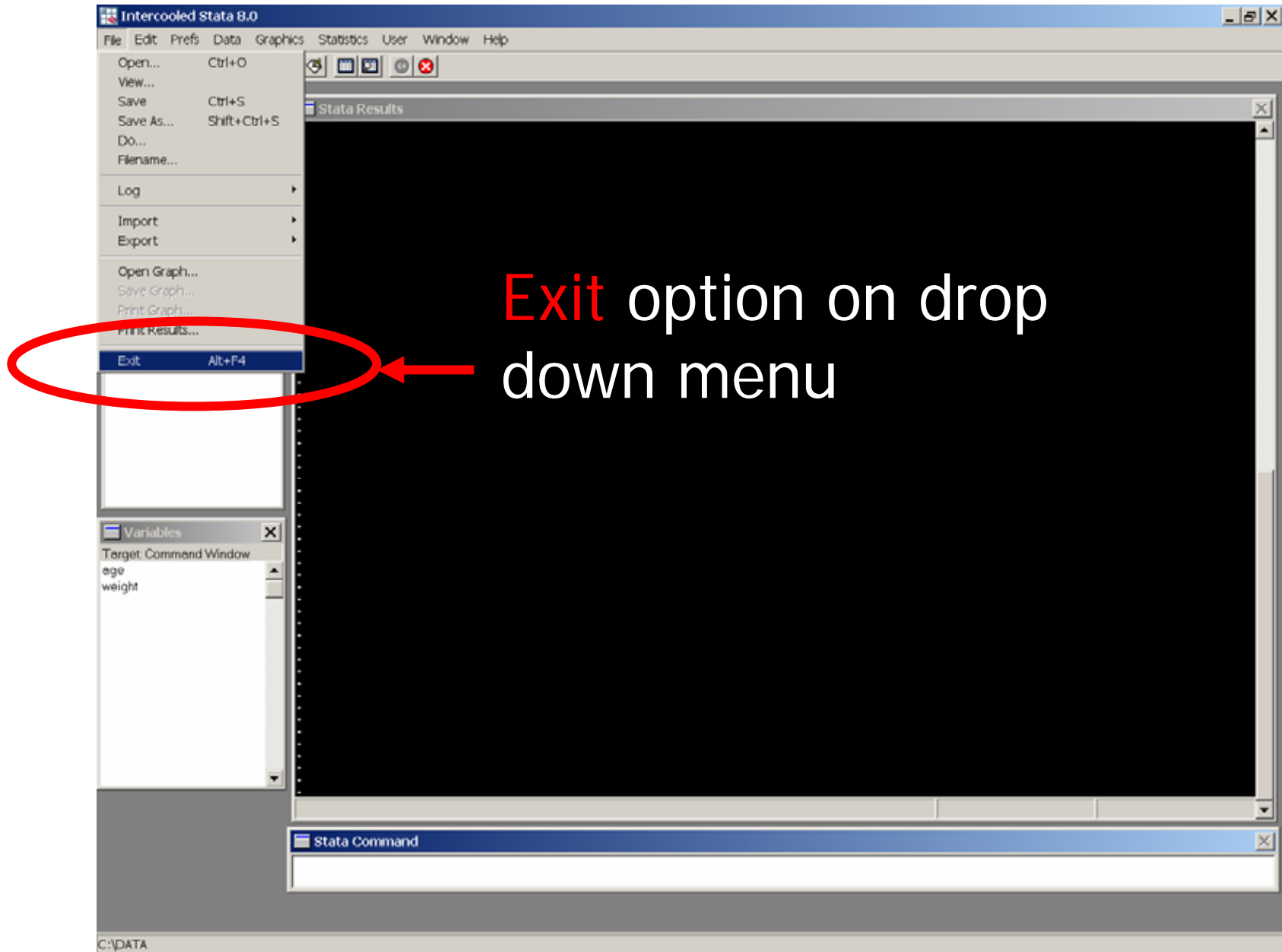


Naming Data Files

- ◆ With either approach, your file will be saved with extension “.dta” (the extension for Stata data sets)
- ◆ Example—if you call your file mydata, it will be saved as mydata.dta

Leaving Stata

- ◆ When you are done for the day, and have saved your data, you may want to leave Stata
- ◆ Click on the file option in the upper left hand corner and choose exit on the drop down menu



Exit option on drop
down menu



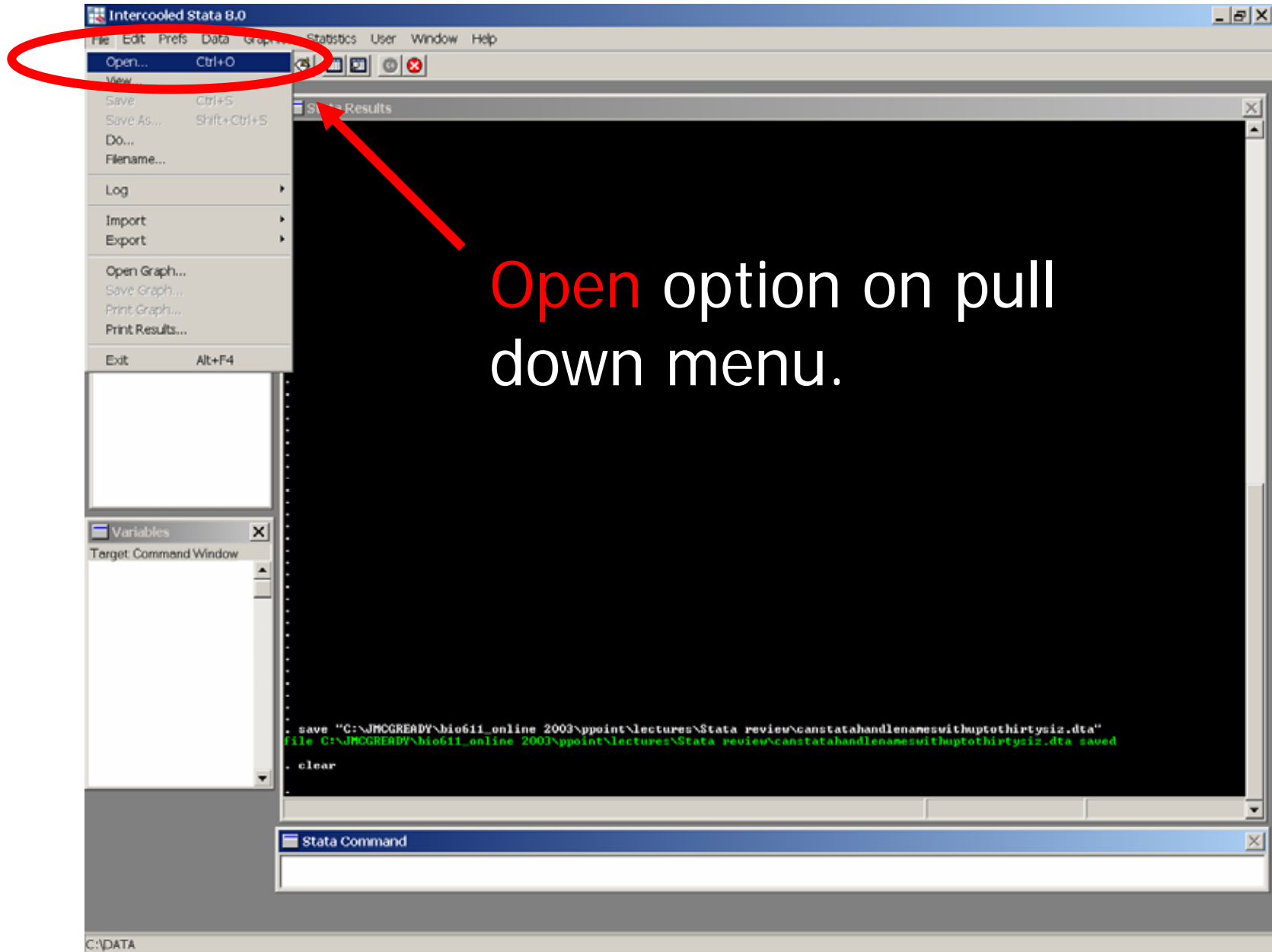
JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section C

*Reusing, Updating, and Labeling
Stata Data*

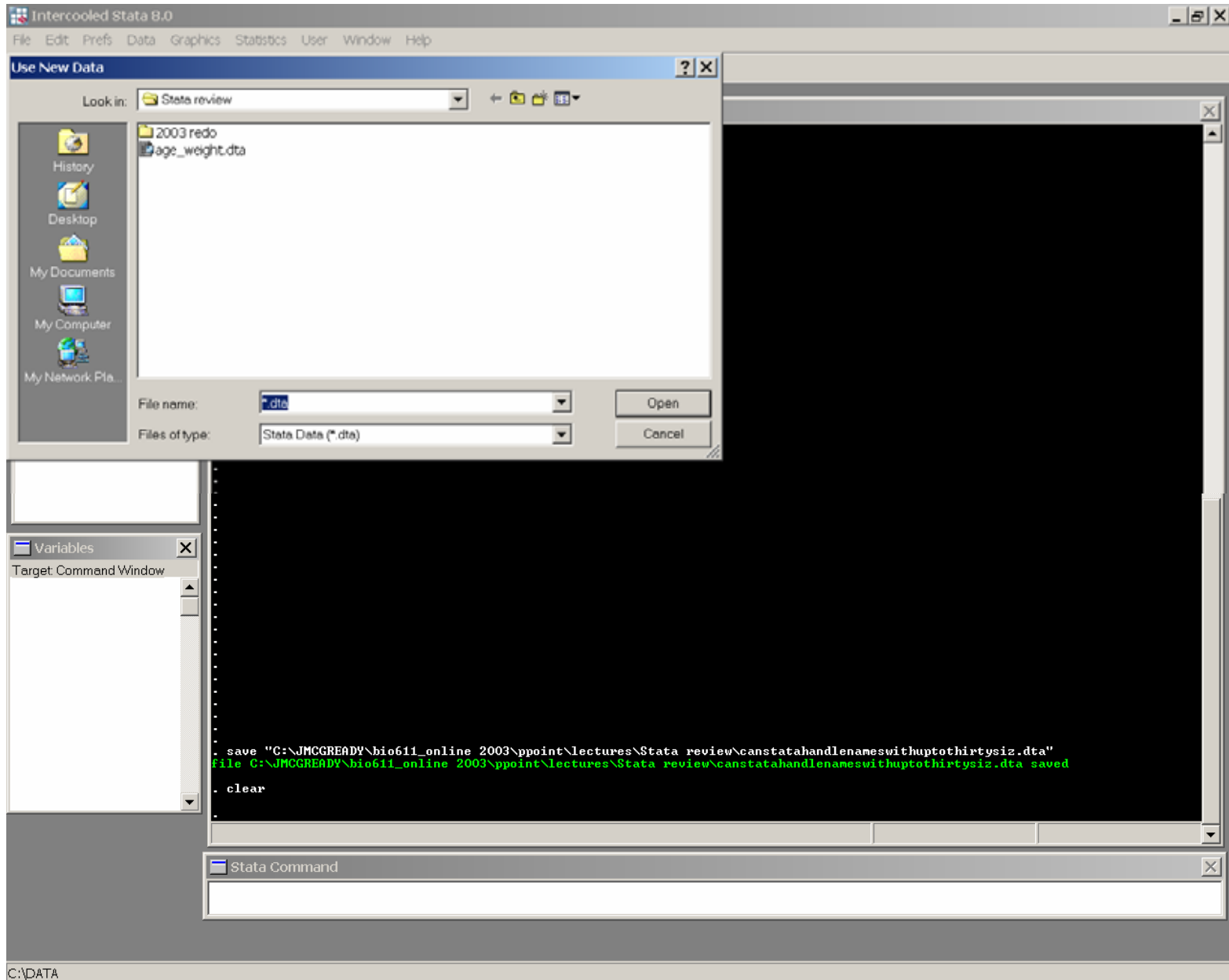
Retrieving Saved Data

- ◆ Now that you have created your data set, and saved it, you can also revisit it at some point in the future
- ◆ The file option allows you to open a saved Stata data set



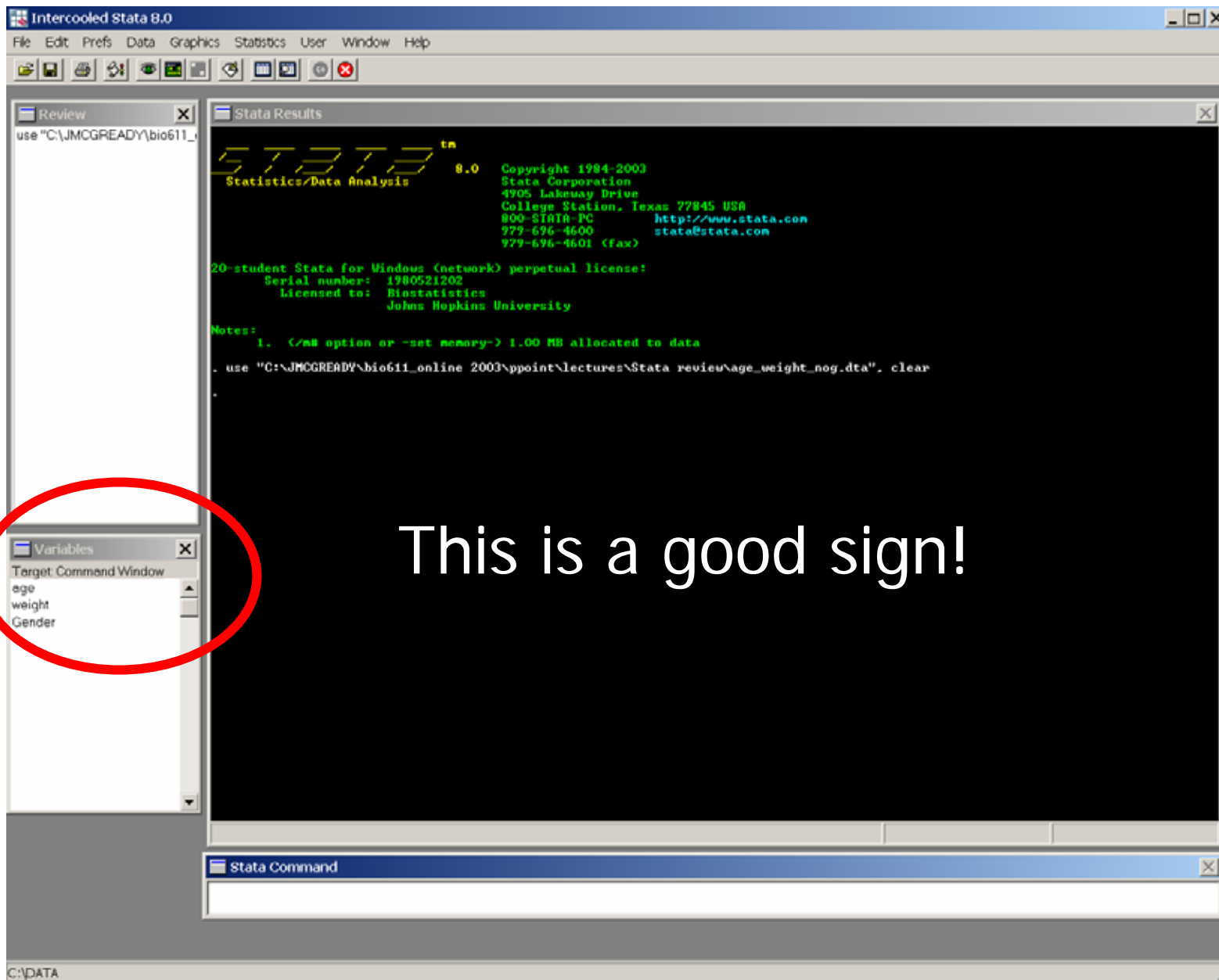
Retrieving Saved Data

- ◆ As when saving, the file chooser box will appear, allowing you to designate the directory and filename of the data set you wish to retrieve



Retrieving Saved Data

- ◆ After opening the file, the names of the variables in the data set should appear in the variable window



This is a good sign!

Labeling Variables

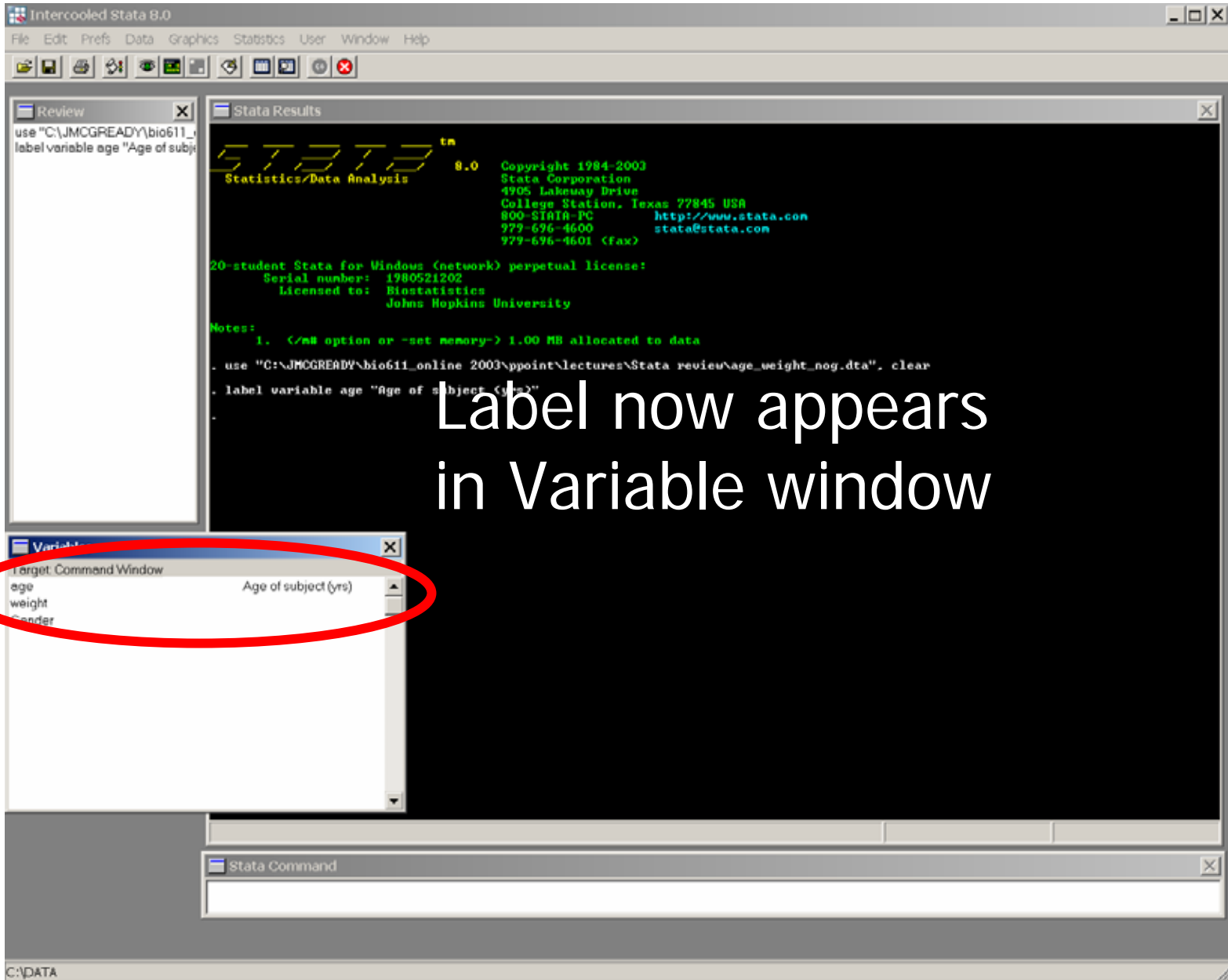
- ◆ Sure, we have given intuitive names to the two variables in this data set; however, with large data sets containing multiple observations on the same data point (example: weight measured at birth, one year, and five years) it becomes increasingly more difficult to assign variable names that tell the “whole story”

Labeling Variables

- ◆ Luckily, Stata allows us the option of labeling variables—inputting a short descriptive phrase, which will appear in the variable window alongside the variable name

Labeling Variables

- ◆ Syntax
 - label variable `varname` "label"
- ◆ Example
 - label variable `age` "Age of subject (yrs)"



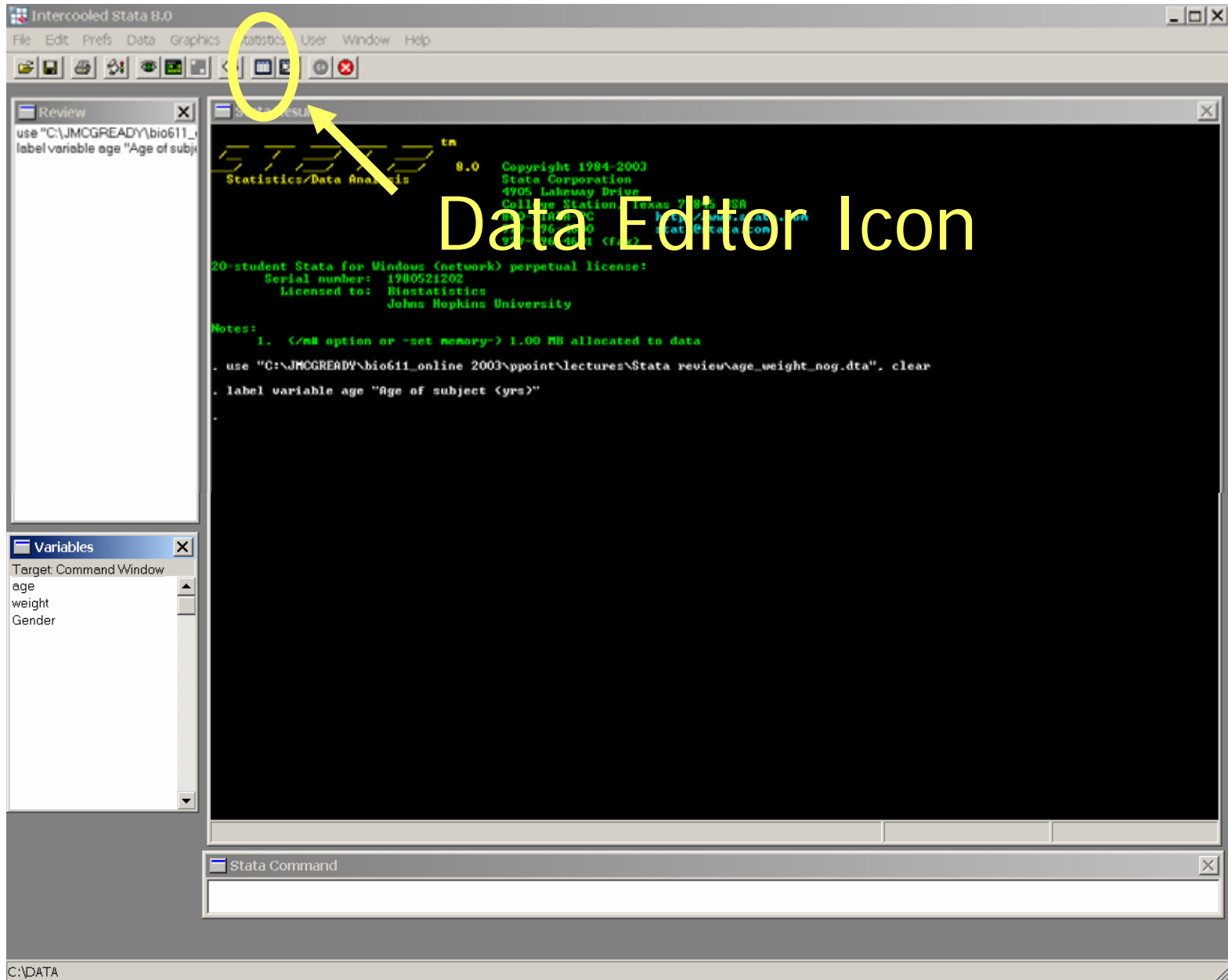
Label now appears
in Variable window

Labeling Variables

- ◆ Labeling variables is optional
 - Can do it for none, some, or all of the variables in a data set

Adding Variable(s)

- ◆ Stata data sets can be updated—more variables can be added, as can more observations
- ◆ To add more data, return to the data editor screen



Data Editor Icon

Adding Variable(s)

- ◆ Suppose we obtained the following information about the gender of each subject

Age (yrs)	Weight (lbs)	Gender		Age (yrs)	Weight (lbs)	Gender
12	123	F		43	115	F
65	215	M		8	58	F
31	187	M		76	126	M
15	89	F		50	312	M

Adding Variable(s)

- ◆ Instead of entering "Male" or "Female," let's enter "0" for males, and "1" for females

Age (yrs)	Weight (lbs)	Gender		Age (yrs)	Weight (lbs)	Gender
12	123	F		43	115	F
65	215	M		8	58	F
31	187	M		76	126	M
15	89	F		50	312	M

Adding Variable(s)

- ◆ Open the data editor and add gender data

Age (yrs)	Weight (lbs)	Gender		Age (yrs)	Weight (lbs)	Gender
12	123	F		43	115	F
65	215	M		8	58	F
31	187	M		76	126	M
15	89	F		50	312	M

Intercooled Stata 8.0

File Edit Prefs Data Graphics Statistics User Window Help

Review

```
use "C:\JMCGREADY\bio611_...
label variable age "Age of subj
drop GEnde
edit
drop Gender
edit
```

Stata Results

Stata Editor

Preserve Restore Sort << >> Hide Delete...

var3[5] = 1

	age	weight	var3
1	12	123	1
2	65	215	0
3	31	187	0
4	15	89	1
5	43	150	.
6	8	58	.
7	76	126	.
8	50	312	.

Gender in Column 3

Variables

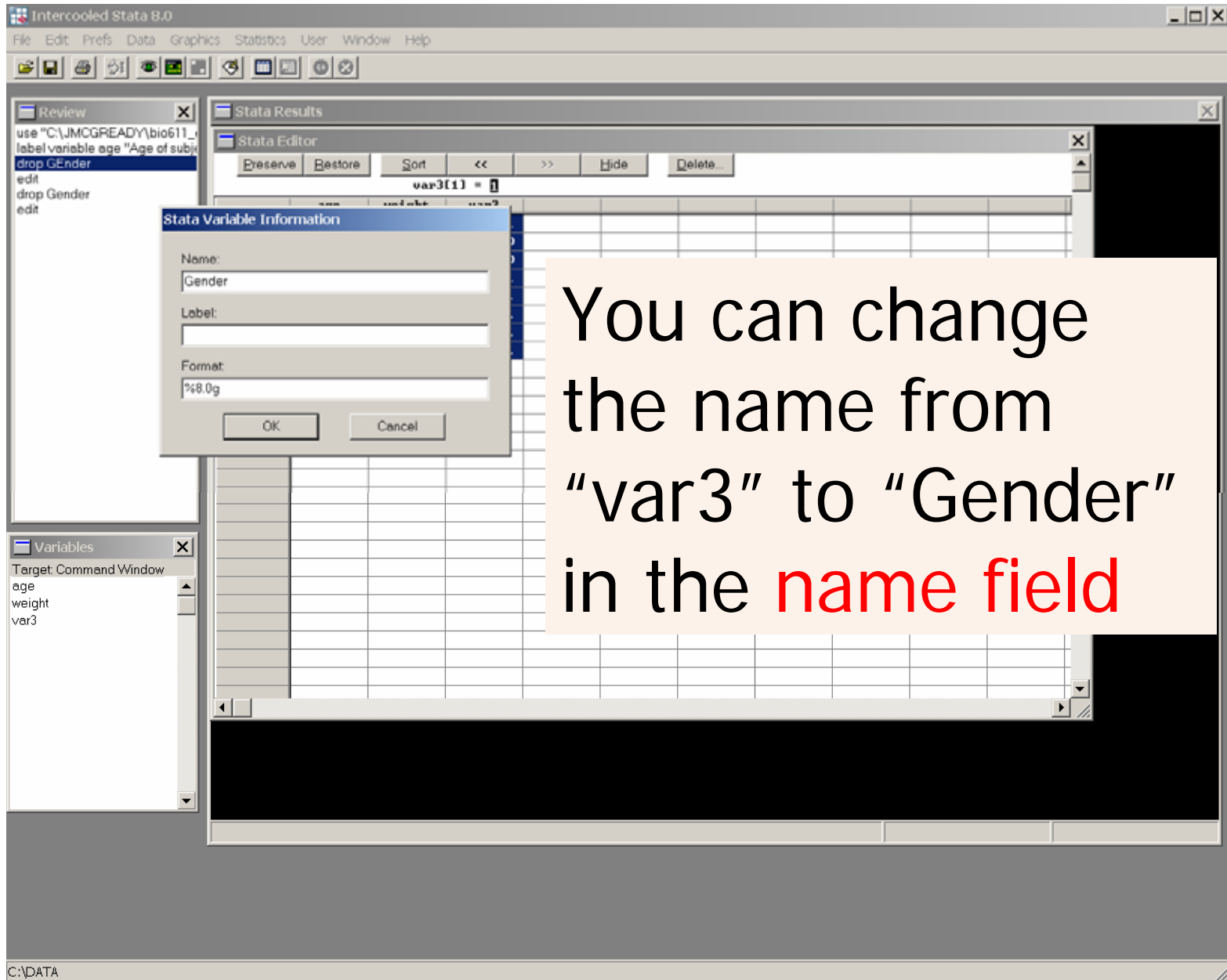
Target: Command Window

```
age
weight
```

C:\DATA

Adding Variable(s)

- ◆ When finished entering the data, change the name of the new variable from "var3" to "Gender"



Adding Variable(s)

- ◆ Click on the “Preserve” button to save changes and then close the data editor
- ◆ Check your update via the “List” command

Intercooled Stata 8.0

File Edit Prefs Data Graphics Statistics User Window Help

Review

```
use "C:\JMCGREADY\bio611_
label variable age "Age of subj
drop GEnde
edit
drop Gender
edit
clear
use "C:\JMCGREADY\bio611_
```

Stata Results

```
STATA
Statistics/Data Analysis 8.0 Copyright 1984-2003
Stata Corporation
4905 Lakesway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

20-student Stata for Windows (network) perpetual license:
Serial number: 1980521202
Licensed to: Biostatistics
Johns Hopkins University

Notes:
1. /MB option or -set memory- 1.00 MB allocated to data

. use "C:\JMCGREADY\bio611_online 2003\ppoint\lectures\Stata review\age_weight_nog.dta", clear
. label variable age "Age of subject <yrs>"
. drop GEnde
variable GEnde not found
r(111);
. edit
- preserve
. drop Gender
edit
- preserve
- byte var3 = 1 in 1
- replace var3 = 0 in 2
- replace var3 = 0 in 3
- replace var3 = 1 in 4
. clear
. use "C:\JMCGREADY\bio611_online 2003\ppoint\lectures\Stata review\age_weight_nog.dta", clear
```

Variables

Target: Command Window

```
age
weight
Gender
```

Stata Command

C:\DATA

Results from List Command

```
. list, nolab
```

	age	weight	Gender
1.	12	123	1
2.	65	215	0
3.	31	187	0
4.	15	89	1
5.	43	150	1
6.	8	58	1
7.	76	126	0
8.	50	312	0

Defining a Label

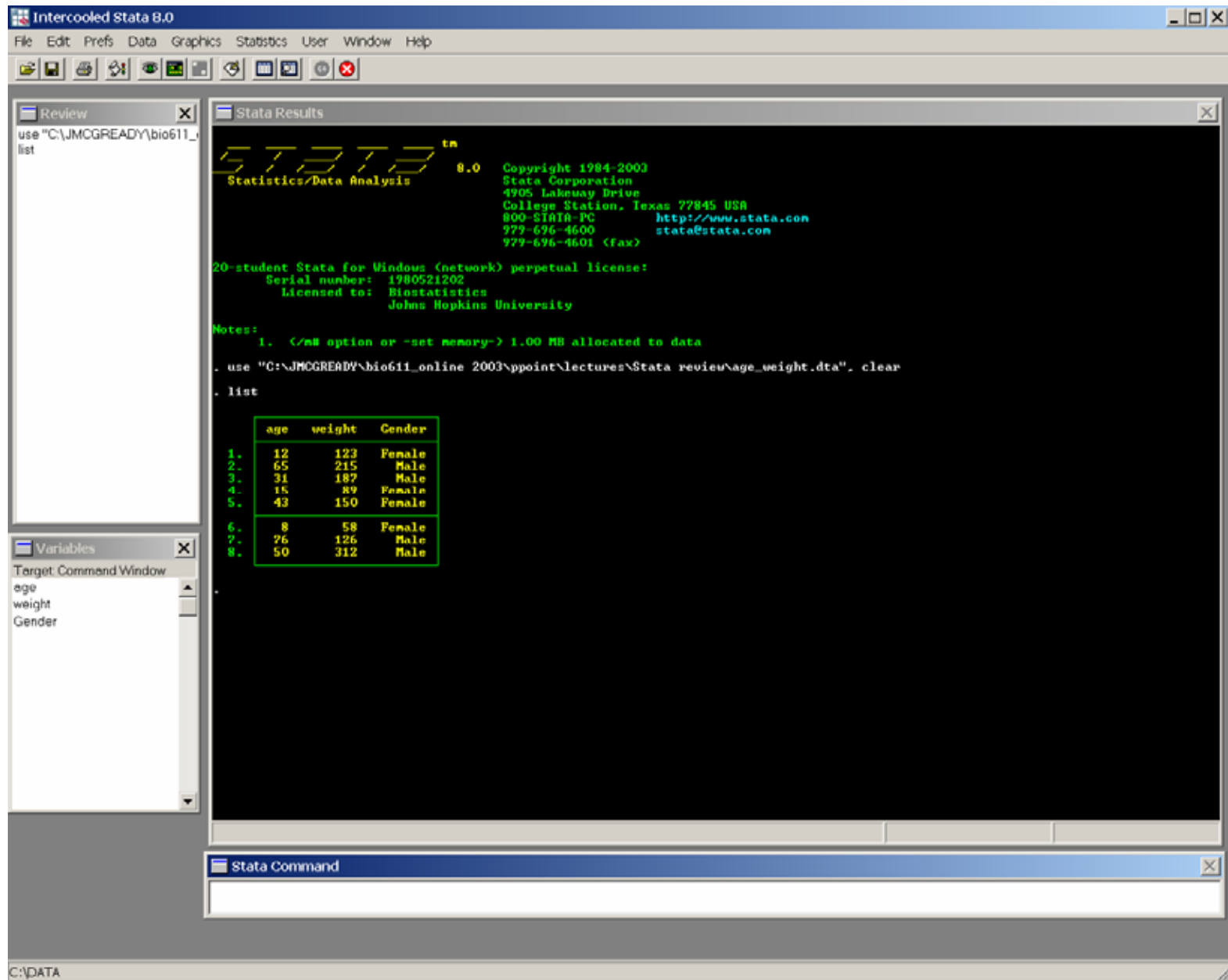
- ◆ First, we need to create a label to apply to the variable
- ◆ Syntax (in command window)
 - label define *labelname label*
- ◆ Example
 - label define **gend** 1 "female" 0 "male"

Assigning a Label

- ◆ Once we have defined a label, we need to assign to the variable(s) for which it was created
- ◆ Syntax (in command window)
 - label values *varname labelname*
- ◆ Example
 - label values **Gender** gend

Assigning a Label

- ◆ Now, the variable "Gender" is still stored as "1" or "0," but will appear as "Female" or "Male"



C:\DATA

Results from List Command

```
. list
```

	age	weight	Gender
1.	12	123	Female
2.	65	215	Male
3.	31	187	Male
4.	15	89	Female
5.	43	150	Female
6.	8	58	Female
7.	76	126	Male
8.	50	312	Male

Why All this Effort?

- ◆ Why not just type in “male” or “female” into the data editor in the first place?
 - More typing
 - Data would then be a text variable instead of numeric—less flexible for using in Stata functions

General Rule

- ◆ When categorical variable is discrete (gender, race, age group, etc.)—usually better to code numerically, then label
- ◆ When it can take on many values—(favorite food, name, etc.)—enter as text



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section D

*A Quick Look at Syntax Diagrams and
Running Commands with Menus*

Stata Syntax Diagrams: The Help Menu

- ◆ Stata has a useful help menu, which allows the user to reference information about specific commands and their syntax
- ◆ The command for getting the help menu is to type `help commandname`, where `commandname` is the name of command of interest

Stata Syntax Diagrams: The Help Menu

- ◆ For example, if you were interested in more information about the list command, you would type **help list** in the command window

Stata Syntax Diagrams: The Help Menu

- ◆ Selected results from `help list`

```
[by varlist:] list [varlist] [if exp] [in range]  
[, [no]display nolabel noobs ]
```

Stata Syntax Diagrams: The Help Menu

- ◆ This is a syntax diagram

```
[by varlist:] list [varlist] [if exp] [in range]  
[, [no]display no label noobs ]
```

Stata Syntax Diagrams: The Help Menu

- ◆ Let's try to demystify!

```
[by varlist:] list [varlist] [if exp] [in range]  
[, [no]display no label no obs ]
```

Stata Syntax Diagrams: The Help Menu

- ◆ First big rule—anything in square brackets is optional

```
[by varlist:] list [varlist] [if exp] [in range]  
[, [no]display no label noobs ]
```

Stata Syntax Diagrams: The Help Menu

- ◆ If we take out everything in brackets, all we have left is “list”



list

- ◆ This is the minimum number of “words” we need to execute the list command

Stata Syntax Diagrams: The Help Menu

- ◆ “varlist”—refers to a list of variable(s)

```
[by varlist:] list [varlist] [if exp] [in range]  
[, [no]display nolabel noobs ]
```

- ◆ Because “varlist” is in brackets, it means listing variables is optional

Stata Syntax Diagrams: The Help Menu

- ◆ Recall from before . . .
 - Typing list at the command line will list the entire dataset
 - Typing list age will list the age variable only for every observation in the dataset

Stata Syntax Diagrams: The Help Menu

- ◆ [by varlist]
- ◆ Optional (of course, it's in brackets)
- ◆ Allows user to repeat the same command over various subsets of data set

by gender: list age

Stata Syntax Diagrams: The Help Menu

- ◆ [if exp]
- ◆ (exp is abbreviation for expression)
- ◆ Optional (of course, it's in brackets)
- ◆ Allows user to run command for observations meeting certain criteria

```
list age if age < 35
```

Stata Syntax Diagrams: The Help Menu

- ◆ [in range]
- ◆ Optional (of course, it's in brackets)
- ◆ Allows user to run command for specific observations in dataset

`list age in 5/25`

Stata Syntax Diagrams: The Help Menu

- ◆ [, [no]display nolabel noobs]
- ◆ Optional (of course, it's in brackets)
- ◆ These are all options specific to the **list** command—more detail in help entry
- ◆ Generally, in syntax diagram, listings after comma are command specific options

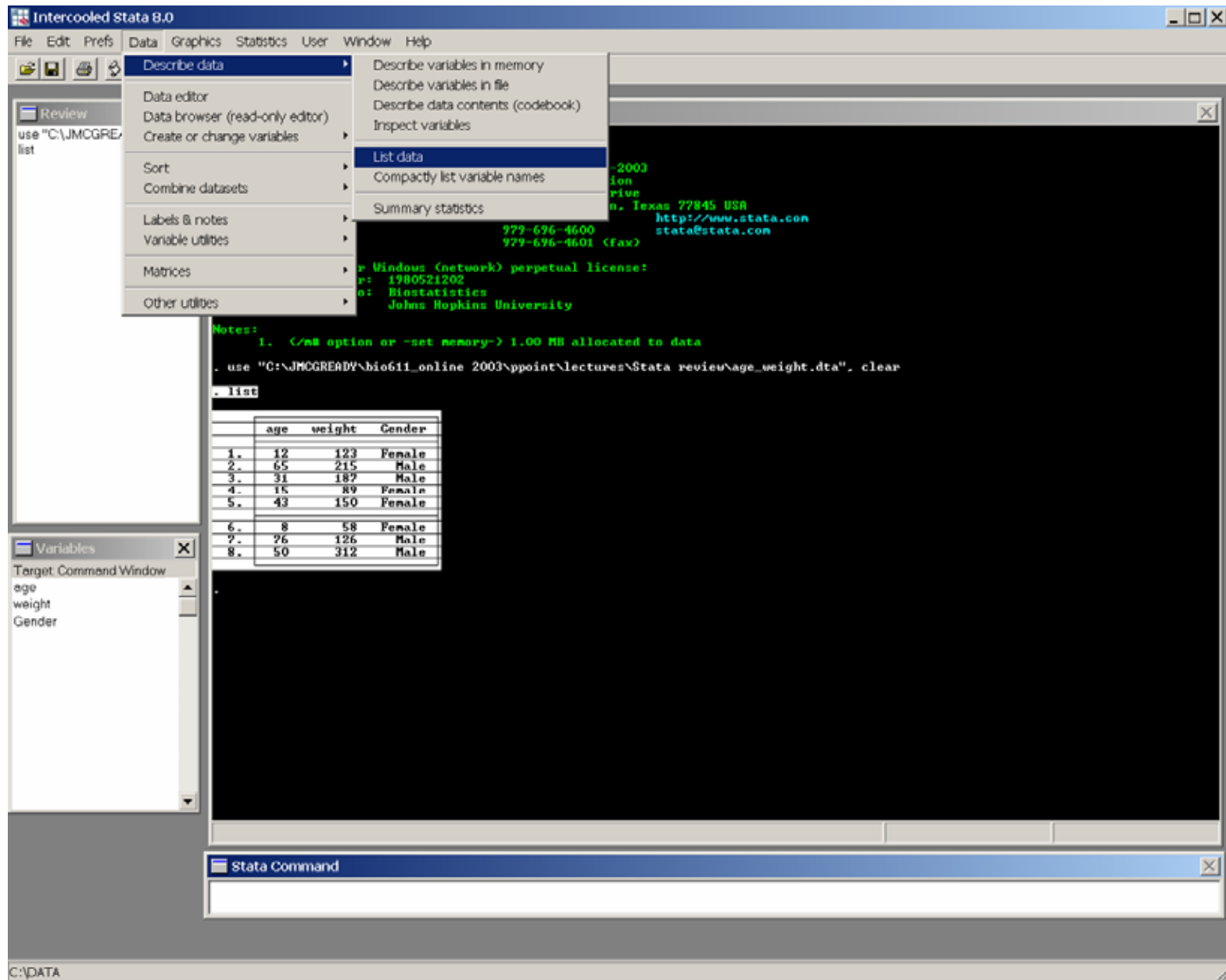
Stata Syntax Diagrams: The Help Menu

The Take-Home Message

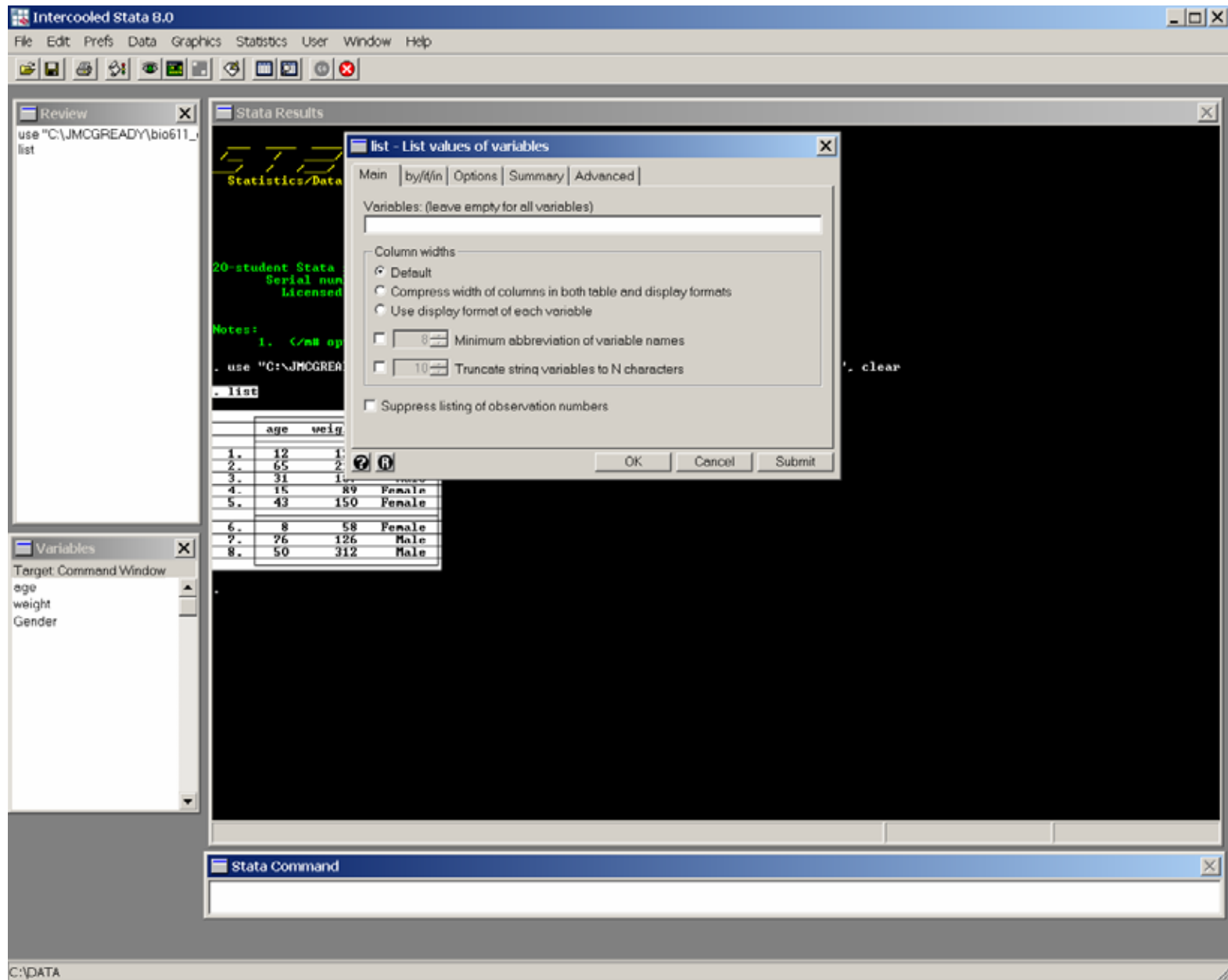
- ◆ “help” useful for getting detailed information on a command
- ◆ Syntax diagrams cumbersome—brackets mean optional!

Using the Menus to Run Commands

- ◆ Stata 8 is the first version of Stata to include a “built-in” menus system for running data related commands



C:\DATA



Intercooled Stata 8.0

File Edit Prefs Data Graphics Statistics User Window Help

Review

```
use "C:\JMOGREADY\bio611_
list
```

Stata Results

```

20-student Stata
Serial num
Licensed

Notes:
1. C:\ml op
. use "C:\JMOGREADY\
. list
. clear
```

list - List values of variables

Main by//in Options Summary Advanced

Repeat command for groups defined by:

Restrict to observations:

if age < 35 Create...

Obs. in range: 1 to 8

OK Cancel Submit

	age	weig	
1.	12	1	
2.	65	2	
3.	31	1	
4.	15	89	Female
5.	43	150	Female
6.	8	58	Female
7.	76	126	Male
8.	50	312	Male

Variables

Target Create...

age
weight
Gender

Stata Command

C:\DATA

Intercooled Stata 8.0

File Edit Prefs Data Graphics Statistics User Window Help

Review

```
use "C:\JMOGREADY\bio611_
list
```

Stata Results

```

20-student Stata
Serial num
Licensed

Notes:
1. C:\ml op
. use "C:\JMOGREADY\
. list

```

list - List values of variables

Main by//in Options Summary Advanced

Repeat command for groups defined by:

Restrict to observations:

if Create...

Obs. in range: to:

OK Cancel Submit

	age	weig	
1.	12	1	
2.	65	2	
3.	31	1	
4.	15	89	Female
5.	43	150	Female
6.	8	58	Female
7.	76	126	Male
8.	50	312	Male

Variables

Target Create...

age
weight
Gender

Stata Command

. clear

C:\DATA