Introduction to Stata and data analysis

September 8, 2023

Learning Objective

- Ability to explore/use Stata Menus and Toolbar
- Know how to get help from Stata
  - Understand help files to a certain degree
- Understand Stata data files
  - Know variable naming convention, variable types, labels, and missing values
  - Able to explore/convert between numeric variables and string variables
- Able to create a clear and reproducible do-file:
  - Produce summary statistics
  - Perform statistical analysis
  - Execute the specific command line(s)/whole do-file, read error messages if there are any
- Create/edit graphs by menus/graph editor
- Remember some frequently used Stata commands
  - Execute commands in the Command window or in a do-file

Outline of topics

- Part 1. Stata Windows & Menus (Windows OS)
- Part 2. Get Help
- Part 3. Stata Files
- Part 4. Stata Graphs
- Part 5. Data Analysis by Menus and/or Commands
- Part 6. Working with A Do-file

Outline of practices

- Practice 1 - Launch Stata
- Practice 2 - Commands Executed in the Command Window
- Practice 3 - Work with the Variables Manager Window
- Practice 4 - Interact with Stata by Commands & Menus
- Practice 5 - Get help from Stata
- Practice 6 - Examples of Help Files
- Practice 7 - Stata commands
- Practice 8 - Work with the Windows
- Practice 9 - From the Variables Manager window

Outline of practices (Cont')

- Practice 10 - Convert between numeric & string variables
- Practice 11 - Working with data
- Practice 12 - Graphs by Stata Menu
- Practice 13 - Editing Graphs by the Graph Editor
- Practice 14 - Execute & Edit the do-file
- Practice 15 - Data Analysis
- Practice 16 - Further Edit: Add Comments & Empty Lines
- Practice 17 - continue editing the do-file if needed
- Practice 18 - Perfect your Do-file

Why Stata ?

Far from the following:

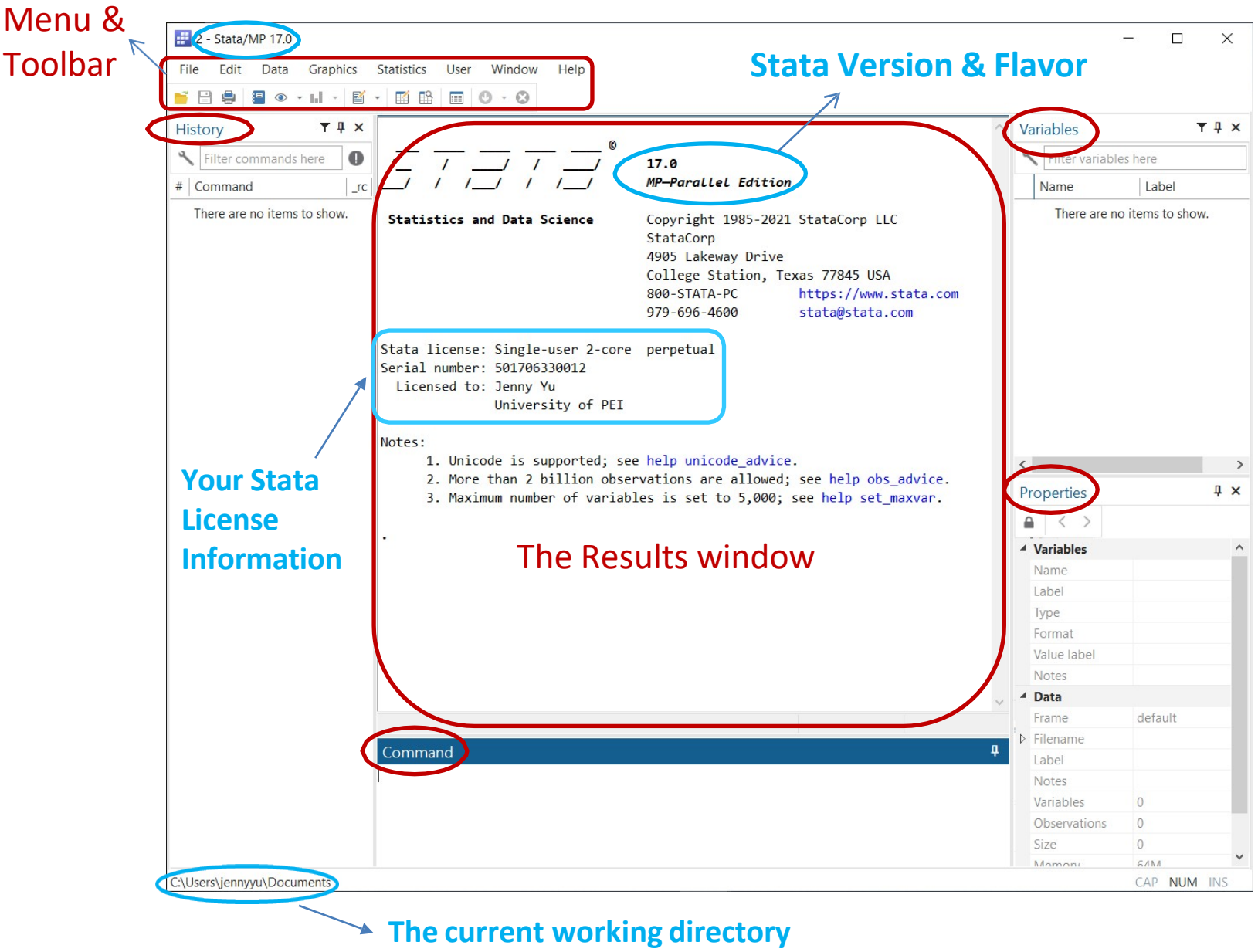
- A complete, integrated software that provides
  - Data management
  - Visualization
  - Statistics
  - Automated reporting
- Easy to use, grow with, automate, extend
  - Pointing and clicking
  - Intuitive and easy to learn, consistent
  - Stata's do-file automation
  - User written program and many more …
- Comprehensive resources
  - Video tutorial, Help files with worked examples, Stata Blog, Free webinars, Stata technical support...

Practice 1 - Launch Stata

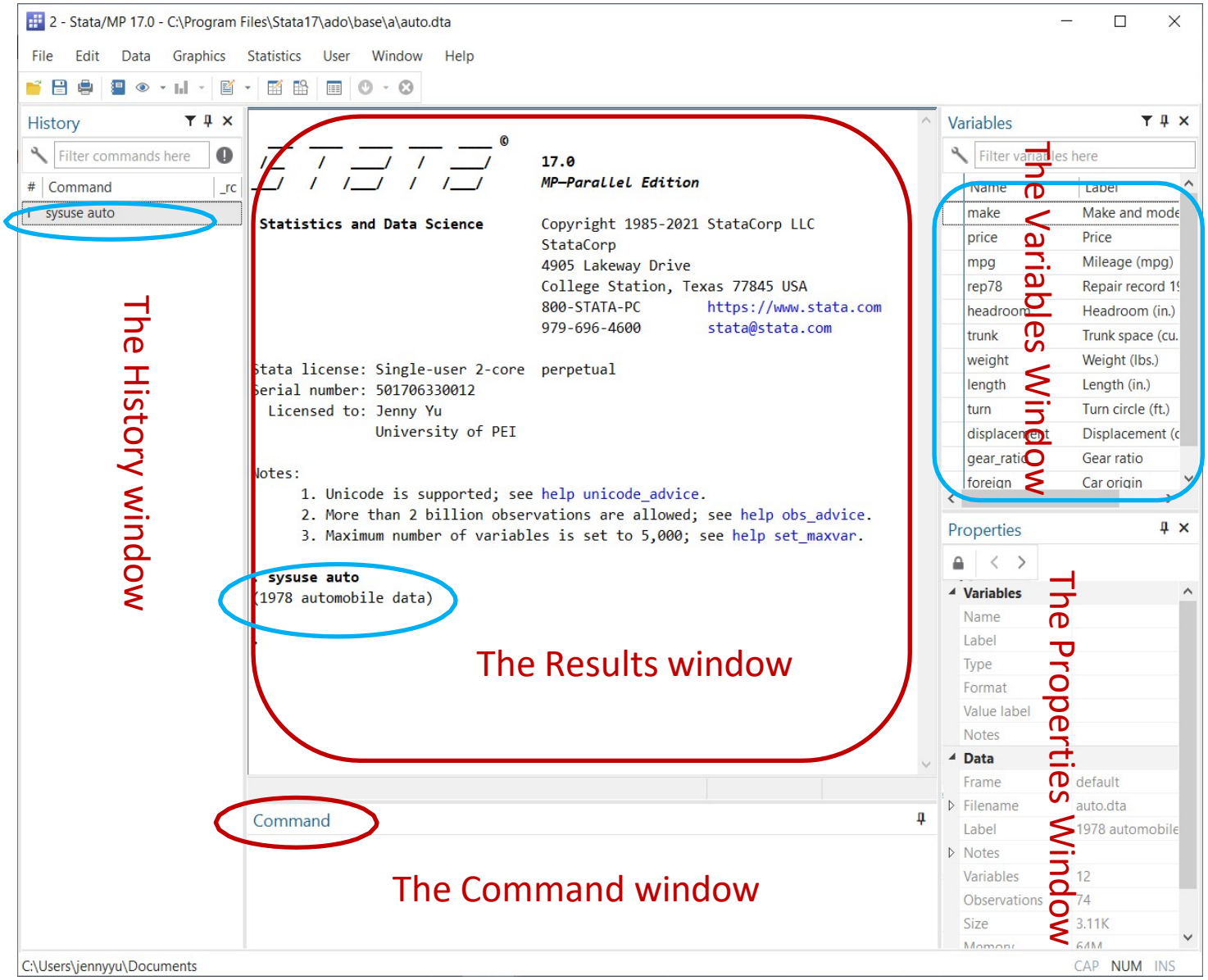
- Create a working folder (e.g.: named introStata) on your computer desktop (or anywhere else),
  - Download all files for the session you received into this folder
  - We will save all files created from this session into this folder
- Install Stata (Already purchased Stata (17), installed on your computer)
- Create a Stata shortcut on your computer's desktop and/or taskbar
- Click the Stata icon to launch your Stata

Part 1 - Stata Windows & Menus (Windows OS)

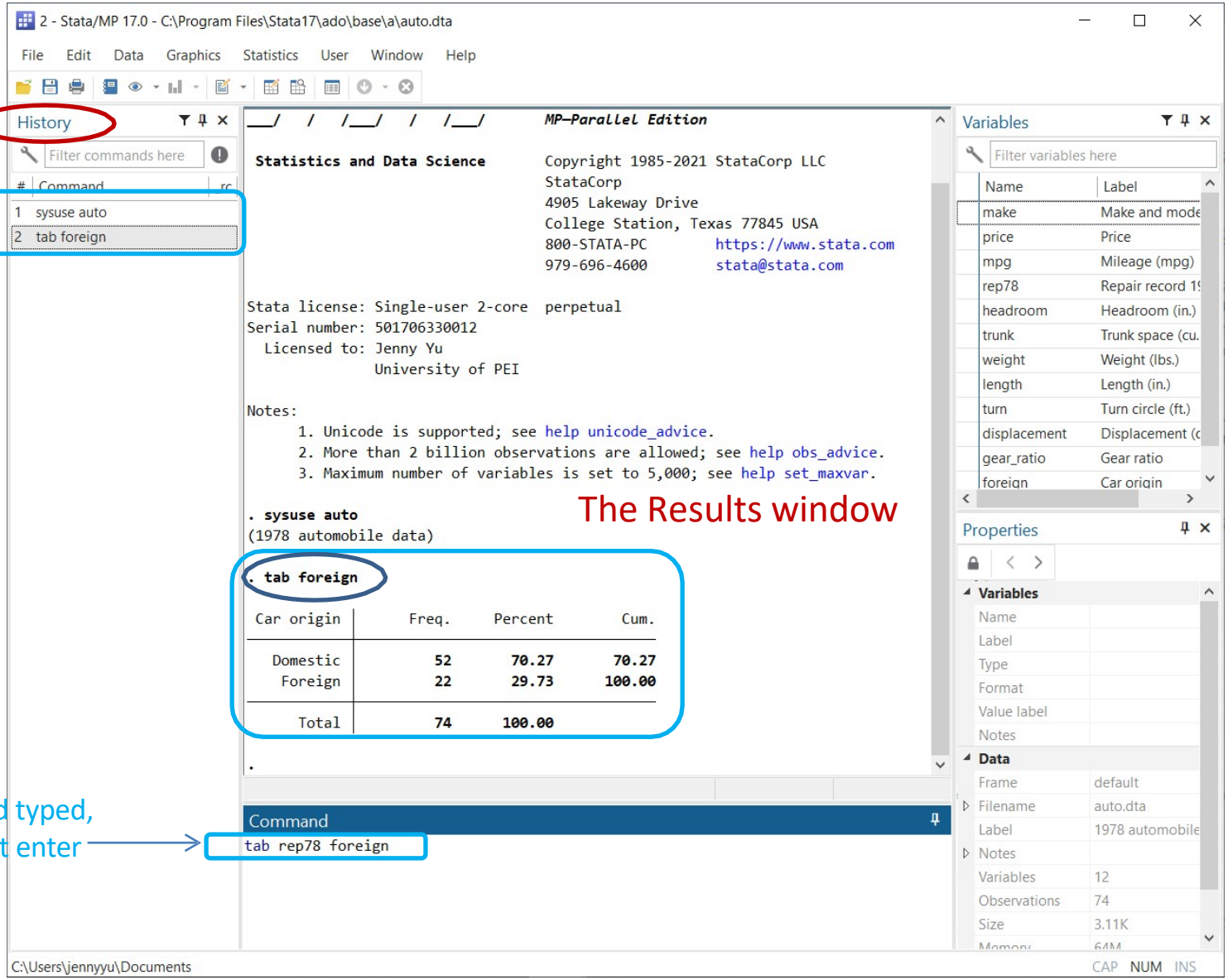
Stata's default interface



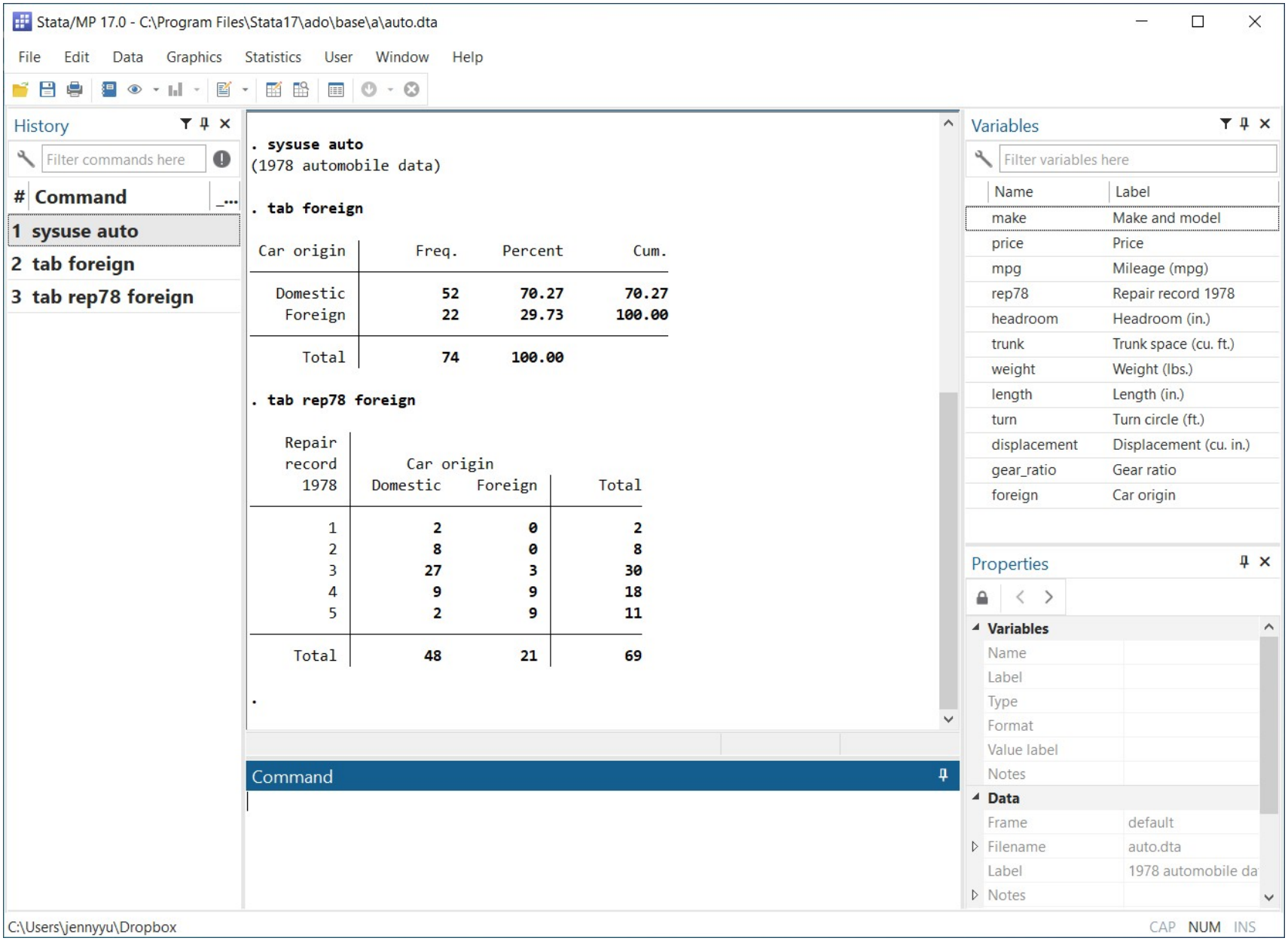
Stata's interface with data loaded



Implement commands



Practice 2 - Commands Executed in the Command Window



Stata's Windows, Menu & Toolbar

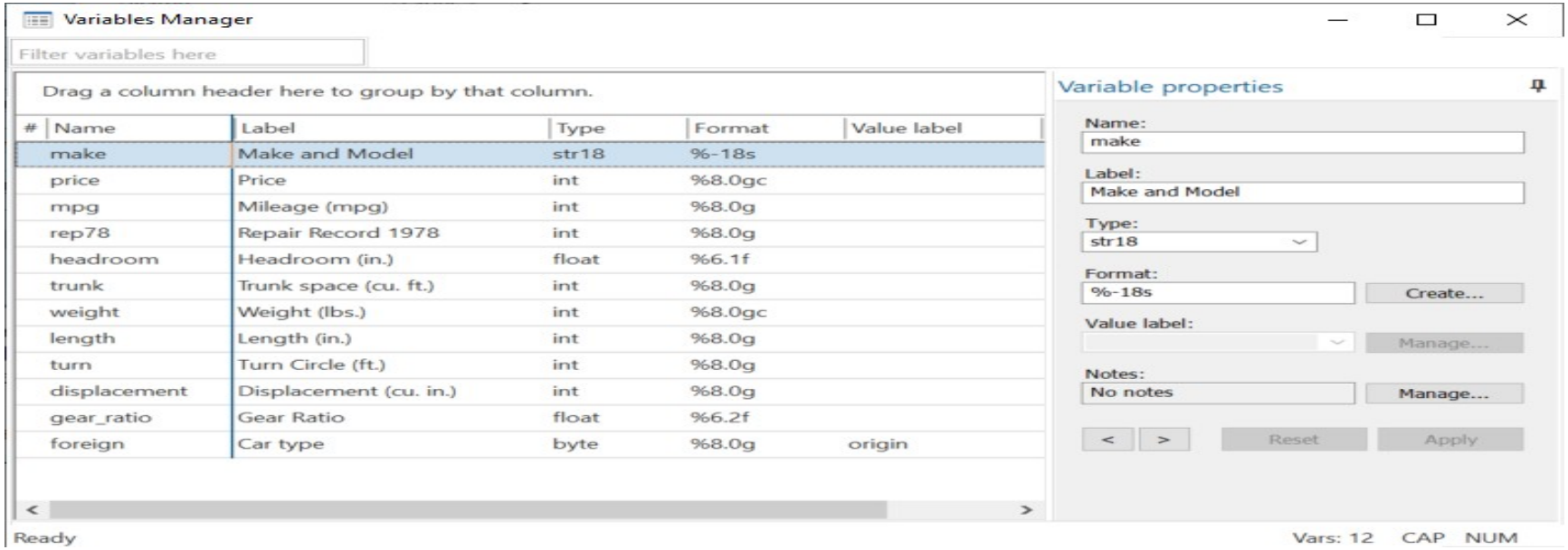
- **Windows:**
  - The five main windows are typically in use the whole time Stata is open; each with a particular task
    - Windows position & size can be rearranged* - practice later
    - All windows allowed to disappear except the Results window; (if it disappears)
      - ❖ Go to the menu: Window, then click the window name,
      - ❖ Or: Edit>Preferences>Load preference set>Factory window settings
  - More specialized windows:
    - Variables Manager
    - Data Editor (Browser & Edit)
    - **Do-file Editor**
    - Graph & Graph Editor
    - Viewer
- **Toolbar & Menu:**
  - The toolbar provides quick access to Stata's more commonly used features - quick access to the more specialized windows
  - Menu Can access all of Stata's features

Stata Menus

- **File menu:** File management
  - Load/save Stata datasets;
  - Import/export external datasets: Excel files, .csv files…
- **Data menu:** Data manipulation
  - Describe data, create/edit labels, manage datasets…
  - Create or change variables needed for the analyses…
- **Graph menu:** Data visualization
  - Create different types of graphs;
- **Statistics menu:** Statistical analyses
  - Descriptive Statistics…
  - Statistical modelling & model postestimation…
- **Window menu:** get the Windows back if lost
- **Help menu:** get help from Stata

Practice 3 - work with the Variables Manager Window

- Click the **Variables Manager icon on the Toolbar (Or from Data menu)**



- **Rename** the variable make to car_make
  - Change make to car_make in the box under Names:
- Edit(/Create) the **variable (**car_make **) label:**
  - Edit the variable car_make label to “Make & model” in the box under Label:
- **Keep** or **drop** selected variable(s)
  - Select the variables trunk & turn, then right click them, click “drop selected
- Pay attention to the commands in the Results/History window
- Use the 2 commands -**rename** & **label variable**- created by Stata again

Interact with Stata: Three Ways

– **Commands**

•Type a command in the Command window & press the enter after each line

– **Toolbar & Menus**

- For the toolbar, hold the mouse pointer over the button for a moment, a tooltip will appear
- Click Stata drop-down menu, navigate to the different commands and select the options you need in the corresponding dialog box;

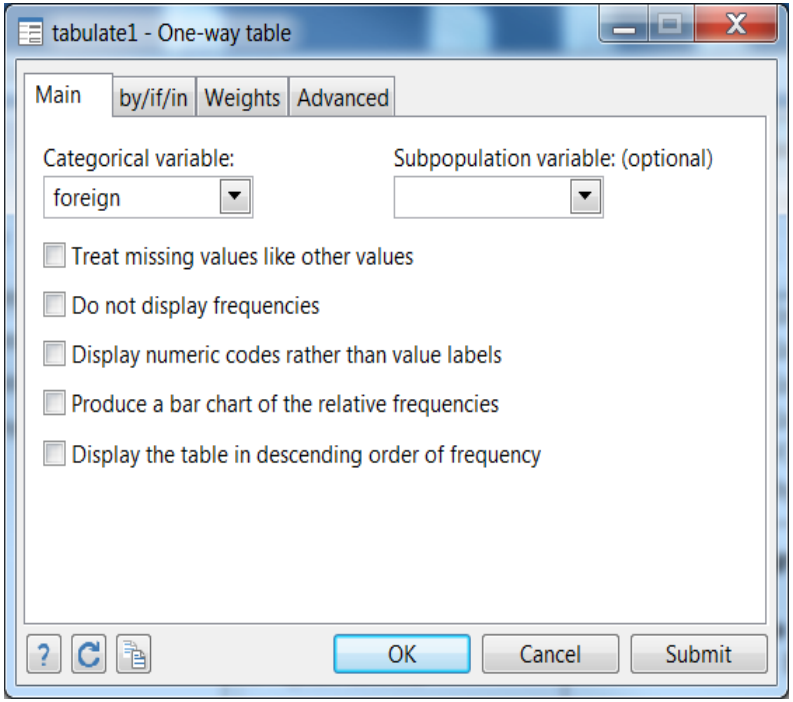
– **Do-files**

- Type commands in a do-file editor, execute the commands
  - ➢ Clicking the button of “New Do-file Editor” on the toolbar

Practice 4* - Interact with Stata by Commands & Menus

➢ Type in the Command window

- . use auto //?
- . use auto, clear
- . tab foreign



➢ Do the above from the Menus

- **Open auto.dta dataset**
  - ✓ Click File > Example datasets
  - ✓ Click '*Example datasets installed with Stata*'
  - ✓ Click *use* after *auto.dta*
- **Issue -tab- command for foreign variable**
  - ✓ Click 'Statistics > Summaries, tables, and tests > Frequency tables > One-way table';
  - ✓ Find & click variable 'foreign' under the box of Categorical variable, then click **Submit** or OK button;

❖ Compare the commands & results created by Stata in the Results window with the typed in the Command window. Questions?

Interact with Stata - Which way to choose?

- **Command window**
  - Need **to remember** the commands (get help files);
  - Typically, much **faster** for often used commands;
- **Menus & Toolbar**
  - Easier for beginners;
  - Can do practically **almost everything** as typing the commands;
  - Faster when building up complex commands, such as those that create graphs;
  - **Learn & remember the commands** that Stata generated for you when using Stata menus
    - ❖ Displayed on both the Results window and the History window.

❖ Combination of using menus and commands

- **Do-files**
  - Reproducible

Part 2 - Get Help

Practice 5* - Gethelp from Stata

- **Click the ? button** on the bottom left corner in the dialogue box:
  - Continue with the previous dialogue box: 'tabulate 1 - One-way table'
  - ✓ **tabulate** *varname [if] [in] [weight] [, tabulate1_options]*
    - *varname is required;*
    - *If a part of a command word is underlined,* ***the underlined part is the minimum abbreviation****. Any abbreviation at least this long is acceptable;*
- **Type help & the command** in the Command window**.** If know the commands, **typing is easier & faster.**
  - **. help tabulate**
  - **. h ta**
- **Get help from the Help Menu:**
  - Help > Stata command: tab
  - Help > PDF documentation or/ Advice or/Contents…

Practice 6* - Examples of Help Files

- In the Command window, type:

. help language

The basic language syntax is (with few exceptions)

[prefix :] ***command*** *[varlist] [=exp] [if] [in] [weight] [***using** *filename] [, options]*

- ❖ Square brackets distinguish optional from required options;
- ❖ Underlining is used to indicate the shortest abbreviations allowed;
- ❖ Options, denoted as *options* in the above, are specified at the end of the command. A comma must precede the first option.

. help data_types

- ❖ Read the help file. Try one of the worked examples at the bottom of the file. Any questions?

. help format

- ❖ Read the help file. Try one of the worked examples at the bottom of the file. Any questions?

Practice 7 - Stata commands

- Try the following commands(// indicating comment):
  - . sysuse auto [, clear]
  - . count if rep78 > 4 //?16
  - . tab rep78 //?, check the data
  - . tab rep78, mis //11 + 5
  - . count if rep78 > 4 & rep78 != .
  - . by foreign: count if rep78>4
  - . by foreign: tab rep78
  - . tab foreign rep78, miss
  - . table foreign rep78, nototal miss

Get help with Stata

- Stata website:
  - https://www.stata.com/training/webinar/
  - https://www.stata.com/links/video-tutorials/
  - https://www.stata.com/learn/

- Statalist - a web Forum: http://www.statalist.org

- Stata support: https://www.stata.com/support/

- Many more …

Practice 8 - work with the Windows

- **The Command window**
  - Click the Command window to activate it, then press
    - ❖ **Page Up** or **Page Down** key: steps backward or forward through the commands history to bring the command to the Command window;
- **The History window**
  - Shows the history of commands that have been entered;
  - Displays **successful commands** in **black;**
  - Displays unsuccessful commands, along with their error codes, in **red**;
    - ❖ **Click once** on a command in the History window to put it back in the Command window (usually for editing) or **double-click** on it to execute it again as-is.
- **The Variables window**
  - Click once on a variable to select it; Right-click on a variable/variables to explore;
  - **Double-click** on a variable or **one click** on the one-click paste column puts the selected variable at the insertion point in the Command window.

Part 3 - Stata Files: Data Files & Related

Stata Files

A glimpse at Stata files:

Stata File Name	File Extension	Notes
Data files	.dta	Datasets saved in Stata format
Do-files	.do	Program files: Stata Commands saved for future reference
Log files	.log .smcl	Record your session in the Results window • Stata commands and • The results from the commands
Graph files	.gph	Graphs saved in Stata format
Ado files*	.ado	• Automatically loaded do-file • User-written Stata programs that needs extra steps to install in your computer
Help files	.pdf	**Get help:** • Know how to get help, • Read and understand the help files

Stata Data Files

– **A data file consists of:**

- **Observations (rows)**
- **Variables (columns):**
  - Variable names - refer to slide: Rules for naming variables
  - Variable types - refer to slide: Variable storage types
  - Variable labels - refer to slide: Variable labels & Value labels

- **Values (cells): displayed in red, black, blue**
  - Value labels
    - **Value labels** (**texts**) displayed in **blue** in data browser/editor;
    - **Internally, Stata store the values as numeric, not the value labels**
      - ✓ . h labels
  - Missing values:
    - A missing value for a numeric variable displayed as a black dot
    - A missing value for a string variable displayed as an empty/blank cell
      - ✓ . h missing

Missing Values*

- The basic missing value for numeric variables is represented by a dot . , which is **larger than any nonmissing;** (. count if rep78 > 4 //16)
- *There are 26 others you can use for missing values: .a through .z, called extended missing values. Stata treats them all the same, but you can assign meanings to them. For example, if you were working with a survey you might decide to code "the question did not apply" as .a and "the respondent refused to answer" as .b. The ordering of missing values is:
  - **all nonmissing values < . < .a < .b < ... < .z**
- Missing values for string variables are denoted by a **blank cell** in the data browser/editor, or the **empty string “”** if referred in the command line

.

Variable storage types* - . help data types

- **Numeric variables or String variables:**
  - Each variable is said to be either type numeric or type string according to their storage type;
- Numeric variables stored as byte, int, long, float, or double
  - Default being **float**;
  - **byte, int, & long** are integer type - can **hold only integers**;
  - The values of numeric variables are **black** in data browser/editor;
- String variables are stored as str1 … str2045 or as strL.
  - **The values of string variables** are **red** in data browser/editor;
  - **The string values** must put inside **quotation marks if need to refer**
- The value 1 and the character "1" are completely different things.
  - For example: 1+1 is 2, but **"1" + "1" is "11"**

Rules for naming variables:

- Stata is **case sensitive**, Make, make, and MAKE are all different names to Stata
  - . h Naming conventions

- A variable name is a sequence of **1-32 characters**
  - The characters can be letters (A-Z, a-z, and any Unicode letter), digits (0-9), underscores(_)

- **Spaces** or other characters **are not allowed**

- The first character of a name must be a letter or an underscore. (However, recommend you not begin your variable names with an underscore.)

Making data readable*: Stata labels

- Labeling dataset, variables, and values
  - **Labeling variables with descriptive names** clarifies their meanings;
    - Variable labels are how we would refer to the variable in normal conversation
  - **Labeling values of numerical categorical variables** ensures that the real-world meanings of the encodings are not forgotten;
  - Proper labeling of the dataset produces much more readable results;
  - Are crucial when sharing data with others, and your future self.

Variable labels & Value labels - . help labels

- **Variables** can be labeled/edited in several ways:
  - By the Variables Manager - Practice 3
  - Select the variable in the Variables window and editing the Label field in the Properties window (unlock it first);
  - Remember the commands
- Variable/data **values** can be labeled/edited:
  - **Value labels** (**texts**) displayed in **blue** in datasets; **internally, Stata store the values as numeric, not the labels (texts).**
  - By Menu (or remember the commands)
    - *Data > Data utilities > Label utilities > Manage value label…Create label*
    - *Data > Data utilities > Label utilities > Assign value label to variables*
  - By the Variables Manager (or the Properties) window
    - Click a numeric variable you want to label
    - Click Manage… button, then click Create label or Edit label or…

Practice 9* - From the Variables Manager window

- **Rename** variable(s) or Create/edit **variable labels:**
  - Try to rename make to: **car make**; **car's_make**; or **make_&_model**
  - Try to label make to: **car make; car's_make;** or **make_&_model**
- Edit **data types -** click the variables' name at the left; click Create button beside the Format/edit the format in the box under Format:
  - Change the format of make to %18s; price to %8.0g; headroom to %6.2f
- Create/edit **value labels (Variables Manager)**
  - . gen byte origin= foreign //browse data
  - Click origin at the left, click Manage…button beside Value label:, click Create label
  - Type originlabel in the box under Label name:
  - Type 0 in the box under Value:, type domestic under the box of Label:, click Add;
  - Type 1 in the box under Value:, type foreign under the box of Label:, click Add;
  - Click OK, click Close;
  - Make sure variable origin is selected, click the dropdown arrow under **Variable label**:, select originlabel, click Apply button;
    - label the variable origin: type car origin in the box under Label:, click Apply button
- **Browse the data; learn and save the commands created by Menu**

Continuous, categorical, and indicator variables*

- A continuous variable: **a numerical variable**, measures something, such as person's age, height, or weight…
- A categorical variable identifies a group to which the thing belongs.
  - Could categorize persons according to their gender, race or ethnicity
  - **Sometimes,** categorical variables are stored as strings

Continuous, categorical, and indicator variables*

- An indicator variable denotes whether something is true;
  - Indicator variables are a special case of categorical variables;
  - Any categorical variable that divides the data into two groups is a categorical, is also an indicator variable;
  - All indicator variables are categorical variables, but the opposite is not true. A categorical variable might divide the data into more than two groups;
  - For clarity, let's reserve the terms
    - ❖ **A categorical variable divides the data into more than two groups**;
    - ❖ **An indicator variable divides the data into exactly two groups.**

- Stata **can convert** continuous variables to categorical or indicator variables; convert categorical variables to indicator variables

Practice 10* - Convert between numeric & string variables

- Menu:
  - Data > Create or change data > Other variable-transformation commands >

➢ **. destring**

❖ > Convert variables from string to numeric

➢ **. tostring**

❖ > Convert variables from numeric to string

➢ **. encode**

❖ > Encode value labels from string variable

➢ **. decode**

❖ > Decode strings from labeled numeric variable

Stata Operators

- **Syntax**

Arithmetic	Logical	(numeric and string)
+ addition	& and	> greater than
- subtraction	| or	< less than
* multiplication	! not	>= > or equal
/ division	~ not	<= < or equal
^ power		**== equal**
- negation		!= not equal
+ string concatenation		~= not equal

Stata Operators

- The order of evaluation (from first to last) of all operators is ! (or ~), ^, - (negation), /, *, -(subtraction), +, != (or ~=), >, <, <=, >=, ==, &, and |.

- When **performing equality tests** for variables with value labels, need to refer to the numbers in commands, not the labels. Eg: if sub-setting data, the if condition would be:
  - **if foreign==1**
  - **if foreign=="Foreign" is not correct.**

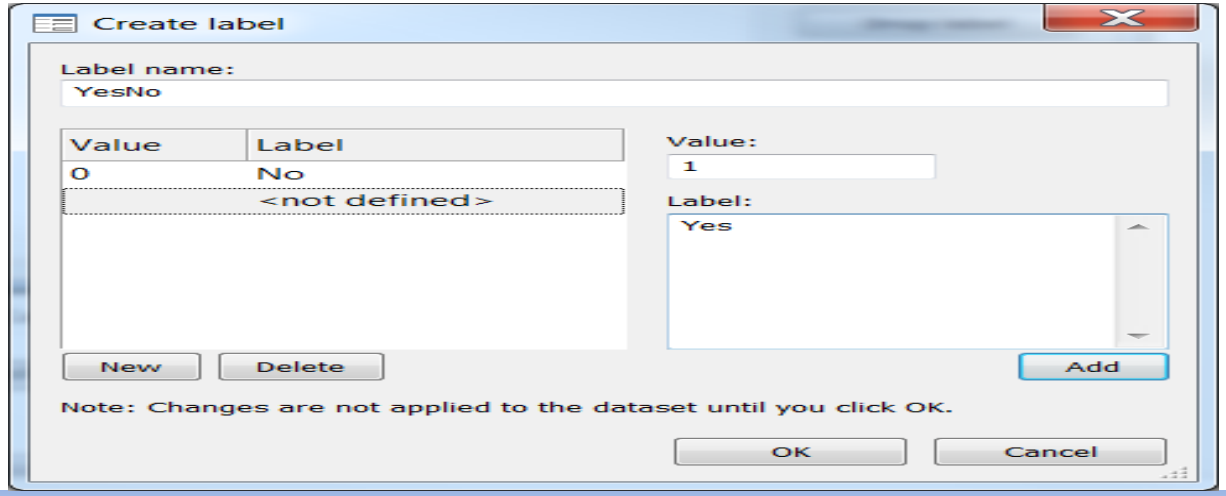
Practice 11* - Working with data

- Change working directory

. **cd** "C:\Users\...\Desktop\2023vhm8110IntroStata"

- . use autoMod.dta, clear
- . count if (rep78 > 4 & rep78 != .) & weight < 3000
- . Browse rep78 weight if (rep78 ==5) & weight < 3000

- Click the Variable Manager:
  - Label variable foreignNum "isForeignCar"
  - Assign value labels: give the values of 0 & 1 of the variable foreignNum a text description, 0 - 'No', 1 - 'yes'
    - Create a value label called 'YesNo'
    - Assign the value label 'YesNo' to the variable foreignNum



Part 4 -Stata Graphs

Practice 12*: Graphs by Stata Menu

- **Create a linear prediction plot overlayed on a scatter plot (sysuse auto)**
  - Click the menu: Graphics > Twoway graph (scatter, line, etc.) …
  - Click Create… button under the tab of Plots
    - ✓ Select “Basic plots” on the left, “Scatter” on the right,
    - ✓ Select “mpg” under Y variables:, “weight” under X variable,
    - ✓ Click Accept; Plot 1 of scatter plot of mpg VS weight was created;
  - Click Create… button again under the tab of Plots
    - ✓ Select “Fit plots” on the left, “Linear prediction” on the right,
    - ✓ Select “mpg” under Y variables:, “weight” under X variable,
    - ✓ Click Accept; Plot 2 of linear prediction line plot of mpg VS weight was created;
  - Click Submit to see the graph; then back to the dialogue box of Twoway graphs:
    - ✓ Click Y axis tab, type Mileage (mpg);
    - ✓ Click Legend tab, select Hide legend;
    - ✓ Click Overall tab, type mpgVSwgt for Name of graph, check box of Replace
    - ✓ Click OK button
    - ✓ Review and save the command created by Stata Menu

Practice 12*: Graphs by Stata Menu

- **Create a linear prediction plot overlayed on a scatter plot (sysuse auto)**
  - Click the menu: Graphics > Twoway graph (scatter, line, etc.) …
  - Click Create… button under the tab of Plots
    - ✓ Select “Basic plots” on the left, “Scatter” on the right,
    - ✓ Select “mpg” under Y variables:, “weight” under X variable,
    - ✓ Click Accept; Plot 1 of scatter plot of mpg VS weight was created;
  - Click Create… button again under the tab of Plots
    - ✓ Select “Fit plots” on the left, “Linear prediction” on the right,
    - ✓ Select “mpg” under Y variables:, “weight” under X variable,
    - ✓ Click Accept; Plot 2 of linear prediction line plot of mpg VS weight was created;
  - Click Submit to see the graph; then back to the dialogue box of Twoway graphs:
    - ✓ Click Y axis tab, type Mileage (mpg);
    - ✓ Click Legend tab, select Hide legend;
    - ✓ Click Overall tab, type mpgVSwgt for Name of graph, check box of Replace
    - ✓ Click OK button
    - ✓ Review and save the command created by Stata Menu

Practice 13*: Editing Graphs by the Graph Editor

- **Click "Start Graph Editor" icon on the graph "mpgVSwgt.gph"**
  - ❑ Click red dot button of "Start recording" to record the changes made by the Graph Editor:
    - ➢ Edit Y-axis: Click somewhere along the y-axis
      - ✓ Change Label angel to Horizontal,
      - ✓ Change Label size to small
    - ➢ Change the Label size to small for X-axis
    - ➢ Edit marker size and symbol: click any marker symbol
      - ✓ Select small for Size; Hollow triangle for Symbol
    - ➢ Click the icon of "Stop Graph Editor" to end recording, click Save button,
    - ➢ Give it a name: myscatter for the recording file & save it in your current working folder
    - ➢ To apply the above recording/changes made from the Graph Editor, add an option play(myscatter) in the Stata graphing command, such as:
      - ✓ . twoway (scatter mpg weight) (lfit mpg weight), ytitle(Mileage (mpg)) legend(off) name(mpgVSwgt, replace) play(myscatter)

Practice 13*: Editing Graphs by the Graph Editor

- **Click "Start Graph Editor" icon on the graph "mpgVSwgt.gph"**
  - Click red dot button of "Start recording" to record the changes made by the Graph Editor:
    - Edit Y-axis: Click somewhere along the y-axis
      - ✓ Change Label angel to Horizontal,
      - ✓ Change Label size to small
    - Change the Label size to small for X-axis
    - Edit marker size and symbol: click any marker symbol
      - ✓ Select small for Size; Hollow triangle for Symbol
    - Click the icon of "Stop Graph Editor" to end recording, click Save button,
    - Give it a name: myscatter for the recording file & save it in your current working folder
    - To apply the above recording/changes made from the Graph Editor, add an option play(myscatter) in the Stata graphing command, such as:
      - ✓ . twoway (scatter mpg weight) (lfit mpg weight), ytitle(Mileage (mpg)) legend(off) name(mpgVSwgt, replace) play(myscatter)

Any questions?

Part 5 - Data analysis by Menus and/or Commands

Data analysis: Plan

- Prepare your data: Part 1 - 4
- Task: statistical analysis and save the results:
  - Tell Stata where my working folder is; start a log file; load a dataset
  - Browse the dataset; describe and understand the relevant variables in your dataset numerically and graphically; prepare your data for the further statistical analysis
    - ❖ Display summary statistics for relevant variables;
    - ❖ Visualize your data;
    - ❖ Generate variable(s) if needed;
  - Save the dataset with a different name
    - ❖ **not overwrite the original dataset**
  - perform statistical modelling for your interest variable;
  - *Model Evaluation
  - *Interpretation and presentation of the results
  - Save the analysis results, close the log file
  - Create a reproducible do-file: save all useful commands typed or generated by the menus to a do-file, **save the do-file**

Data analysis - Step 1

- Preparation:
  - -**cd-** change directory: Tell Stata what my working folder is
    - . cd “C:\Users\...\Desktop\2022vhm8110IntroStata”
    - Or File>Change working directory…
  - -**log using-** Start a log file, put it into my working folder
    - . log using "introStata.smcl"
    - Or File>Log>Begin…
  - **-use-** Load a Stata dataset into memory
    - . use autoMod.dta
    - Or Click folder open icon from the toolbar
- Browse the data, pause, think, and investigate

Data analysis - Step 2

- Descriptive statistics: Data Menu & Statistics Menu
  - Detailed summary statistics for variable(s)
    - . des
    - . sum price mpg
  - One-way table of frequencies:
    - . tab foreign, sum(price)
- Visualization: Graphs Menu
  - Visualize distribution of the variable **price**: Create a histogram for variable **price,** add normal-density plot
    - Graphics > Histogram
      - ✓ Select price
      - ✓ Click Density plot tab, select "Add normal-density plot"
      - ✓ Click OK
    - Right skewed?

Data analysis - Step 3

- Linear regression analysis
- Create necessary variable(s): **log of price**
  - . gen ln_price = ln(price)
- Visualize the association between DV ln_price with mpg
  - . twoway (scatter ln_price mpg) (lfit ln_price mpg)
- Interested in the other variable(s): eg: foreign

*About factor-variable operators in Stata:

Factor variables are extensions of **varlists** of existing variables. When a command allows factor variables, in addition to typing variable names from your data, you can type factor variables using factor-variable operators.

There are five factor-variable operators:

Operator	Description
i.	**unary operator to specify indicators**
c.	unary operator to treat as continuous
o.	unary operator to omit a variable or indicator
#	binary operator to specify interactions
##	binary operator to specify factorial interactions

Data analysis - Step 4

- Statistical modelling*:
  - Statistics > Linear models and related > Linear regression Regression analysis for ln_price

```
. regress ln_price mpg i.foreign
```

Source	SS	df	MS	Number of obs	=	74
				F(2, 71)	=	17.80
Model	3.74819416	2	1.87409708	Prob > F	=	0.0000
Residual	7.47533892	71	.105286464	R-squared	=	0.3340
				Adj R-squared	=	0.3152
Total	11.2235331	73	.153747029	Root MSE	=	.32448

ln_price	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
mpg	-.0421151	.0071399	-5.90	0.000	-.0563517	-.0278785
foreign						
Foreign	.2824445	.0897634	3.15	0.002	.1034612	.4614277
_cons	9.4536	.1485422	63.64	0.000	9.157415	9.749785

- * Model postestimation
- *Present the results and interpret the findings

Data analysis - Step 5

- Save the new dataset: give it a different name
  - **. save autoMod2.dta, replace**
- Close the log file
  - **. log close**
- Save the commands to a do-file:
  - Copy the useful commands in the History window;
    - ❖ Hold Ctrl/Shift key, left click the commands in the History window that you want to save;
    - ❖ Right click the selected commands, click either 'Copy' (or 'Send selected to do-file editor');
  - Click on the "New Do-file Editor" icon on the toolbar to open a new do-file editor;
  - Paste the copied commands to the do-file;
  - Add comments &/or empty lines to the do-file
  - Edit(simplify)/add commands if necessary
  - Save the do-file to your working folder, give it a name.

Part 6 - Working with A Do-file

Practice 14 - Execute & Edit the do-file

- Run the specific command(s) in the do-file:
  - Hightlight the specific commands to re-display statistics or graph(s):
    - **. sum price mpg**
    - **. twoway (scatter ln_price mpg) (lfit ln_price mpg)**
    - **. regress ln_price mpg i.foreign**
- Run the whole do-file: not select/highlight any specific command line, click Execute(do) button on the Toolbar
  - Error with **log using**…; Fix it:
    - . h log
    - . log using…, **replace** ;
  - Error with . use autoMod.dta //fix it: add the option "**, clear**"
  - Error with . save autoMod2.dta //add the option: "**, replace**"
- Run the whole do-file again

Practice 15 - Data Analysis

- Perform data analysis from Step 1 to 5;
  - If you have finished the previous 5 steps, you'll have created a do-file (you) named 'introStata.do' in your working folder

    - cd "C:\Users\...\2021vhm811introStata" //… computer independent
    - log using "introStata.smcl"
    - use "autoMod.dta"

    - des
    - sum price mpg
    - tab foreign, sum(price)

    - hist price, normal
    - generate ln_price = ln(price)
    - twoway (scatter ln_price mpg) (lfit ln_price mpg)

    - regress ln_price mpg i.foreign

    - save "autoMod2.dta", replace

    - log close

Interact with Stata: Why Do-files?

- All commands typed & submitted from the Command window or generated by using menus are **lost** once Stata is closed;

- Need a program file: a **do-file**, to save our Stata commands, which can be used and edited later;

- Might also need a file: **a log file**, to save our work - both the Stata commands and the results from those commands.

Practice 16 - Further Edit: Add Comments & Empty Lines

- . help comments
- comments:
  - Used in do-files, **not used in the Command window;**
  - Ignored by Stata
  - Are for you and your collaborators; comments
    - ❖ Begin the line with *;
    - ❖ Placed inside /* and */ delimiters;
    - ❖ Placed after two forward slashes, that is, //; everything after the // to the end of the current line is considered a comment;

- /// is used to make long lines more readable
- Add commands? If so, type them directly on the do-file;
- Add blank lines, alignment or indentation;
- If it runs, save and close the do-file; exit Stata; then launch Stata again, open the do-file:
  - Execute the entire do-file: introStata.do, by clicking the icon "Execute (do)" in your Do-file Editor Toolbar

Practice 17 - continue editing the do-file if needed

- Run the specific commands in the introStata.do:
  - Hightlight the specific commands to re-display statistics or graphs;
  - Edit the following lines:
- Run the whole do-file again
- Add comments, blank lines, alignment and/or indentation to your do-file to make your do-file meaningful & more clear
- Want to add any commands?
  - . version 17
- Save your updated do-file

Practice 18 - Perfect your Do-file

- **Continue to edit your do-file 'introStata.do', adding empty lines and meaningful comments:**

```
/* vhm811: Introduction to Stata and data analysis - Sep 14, 2022*/
*1. change working directory, open a log file
        . cd "C:\Users\...\2022vhm8110introStata“ //this line is computer dependent
        . log using "introStata.smcl", replace
        . version 17

*2. load dataset
        . use "autoMod.dta", clear

*3. describe data by summary statistics and graphs
        . des
        . sum price mpg, d
        . tab foreign, sum(price)

        . hist price, normal
    *generate a new variable from the variable price
        . generate ln_price = ln( price)
        . twoway (scatter ln_price mpg) (lfit ln_price mpg)

*4. linear regression:
        . regress ln_price mpg i.foreign

* 5. save the new dataset, close log file
        . save "autoMod2.dta", replace
        . log close
```

Questions?

Summary

Commands you should know

- **Operating system interface**
  - . pwd
  - . cd
- **Using & saving data**
  - . save
  - . use
  - . compress
- **Inputting data into Stata**
  - . import
- **Graphing data**
  - . graph
- **Keeping track of your work**
  - . log
  - . notes

- **Basic data reporting**
  - . describe
  - . codebook
  - . list
  - . browse
  - . count
  - . inspect
  - . table
  - . tabulate
  - . summarize
- **Convenience**
  - . display

- **Data manipulation**
  - . append
  - . merge
  - . generate
  - . egen
  - . rename
  - . clear
  - . drop
  - . keep
  - . Sort
  - . tostring, destring
  - . encode, decode
  - . order
  - . by
  - . reshape
  - . frames

Other Stata Commands & Data Management Concepts

- Working with dates and times:
  - . h datetime
- List of Stata variable names:
  - . h varlist
- Stata internal variable and looping: _n and _N
  - . h _variables
- Extensions to generate command:
  - . h egen
- Stata functions: used in expressions, abbreviated exp in syntax diagrams
  - . h functions

Leaning objectives restated

- Ability to explore/use Stata Menus and Toolbar
- Know how to get help from Stata
  - Understand help files to a certain degree
- Understand Stata data files
  - Know variable naming convention; variable types; Labels; Missing values;
  - Able to explore/convert between numeric variables and string variables;
- Able to create a clear and reproducible do-file:
  - Produce summary statistics, *Perform statistical analysis;
  - Execute the specific command line(s)/whole do-file, read error messages if there are any;
- Create/edit graphs by menus/graph editor
- Remember some frequently used Stata commands
  - Execute commands in the Command window or in a do-file

Appendix

- ***Install user-written Stata programs:**
  - Hundreds of user-written programs are available for use with Stata. With Stata's Internet features, obtaining these programs is easy.
    - ❖ Type: findit keyword/s, in the Command window;
    - ❖ Click the link that matches to your search to begin installation;
  - Once the program is installed, you will see: installation completed

- Stata is an invented word, not an acronym.
  - ❖ The correct spelling is **"Stata", not "STATA".**

Acknowledgement

- This module session was originally developed by Dr. Jenny Yu, our CVER member now based in Ontario Veterinary College.

Suggested tutorial materials

- Stata Tutorial: Introduction to Stata
- Stata for beginners course: Stats basics, creating variables, data entry, descriptive stats