

Introduction to STATA

Spring 2021 - SAM Lab Workshop



What is STATA?

Statistical/ Data software tool commonly used in social sciences.

Relatively easy to use and allows for completion of a wide range of data analyses.

How can you access STATA?

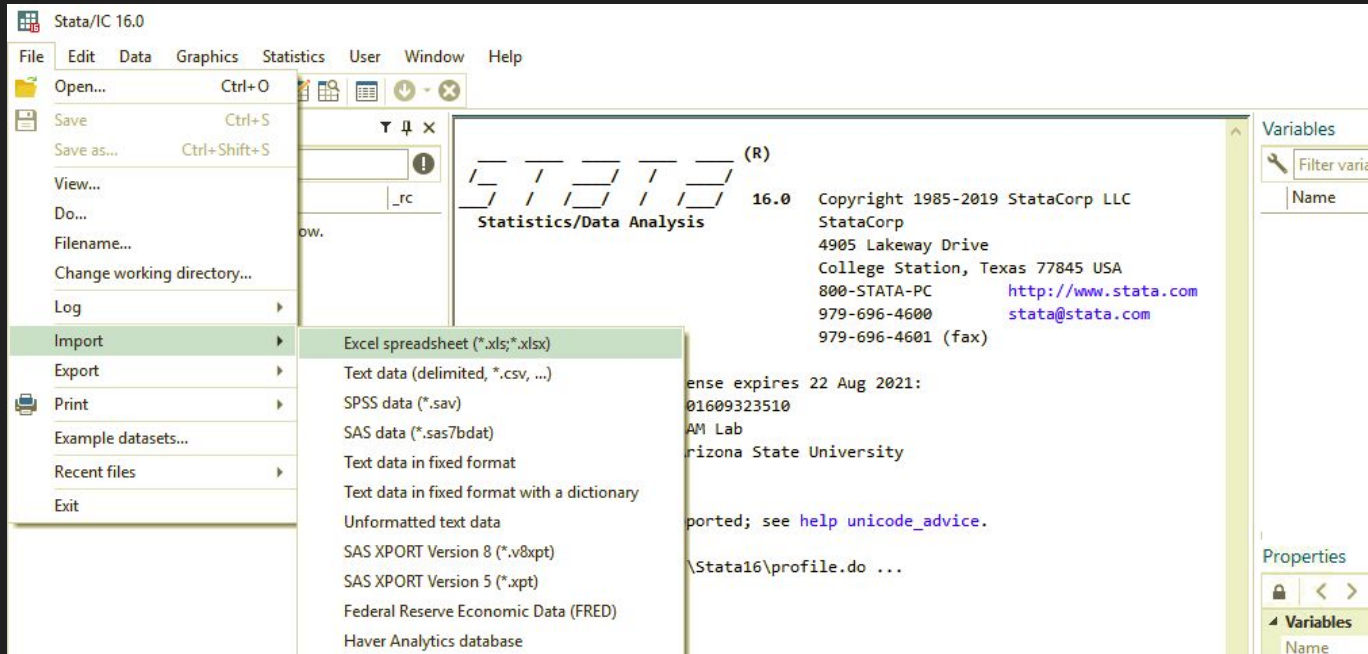
- Purchase a license through STATA corporation to download the software onto your personal computer
- As a SBS student at ASU, you can also use SAM Lab computers which already have STATA on them

Learning Goals:

1. Learn how to import data and create/edit data
2. Learn the different ways to navigate through STATA (Menu & Dialog System, Command Window, Do - File)
3. Learn how to perform basic descriptive statistics and correlations
4. Gain some familiarity with output display and interpretation

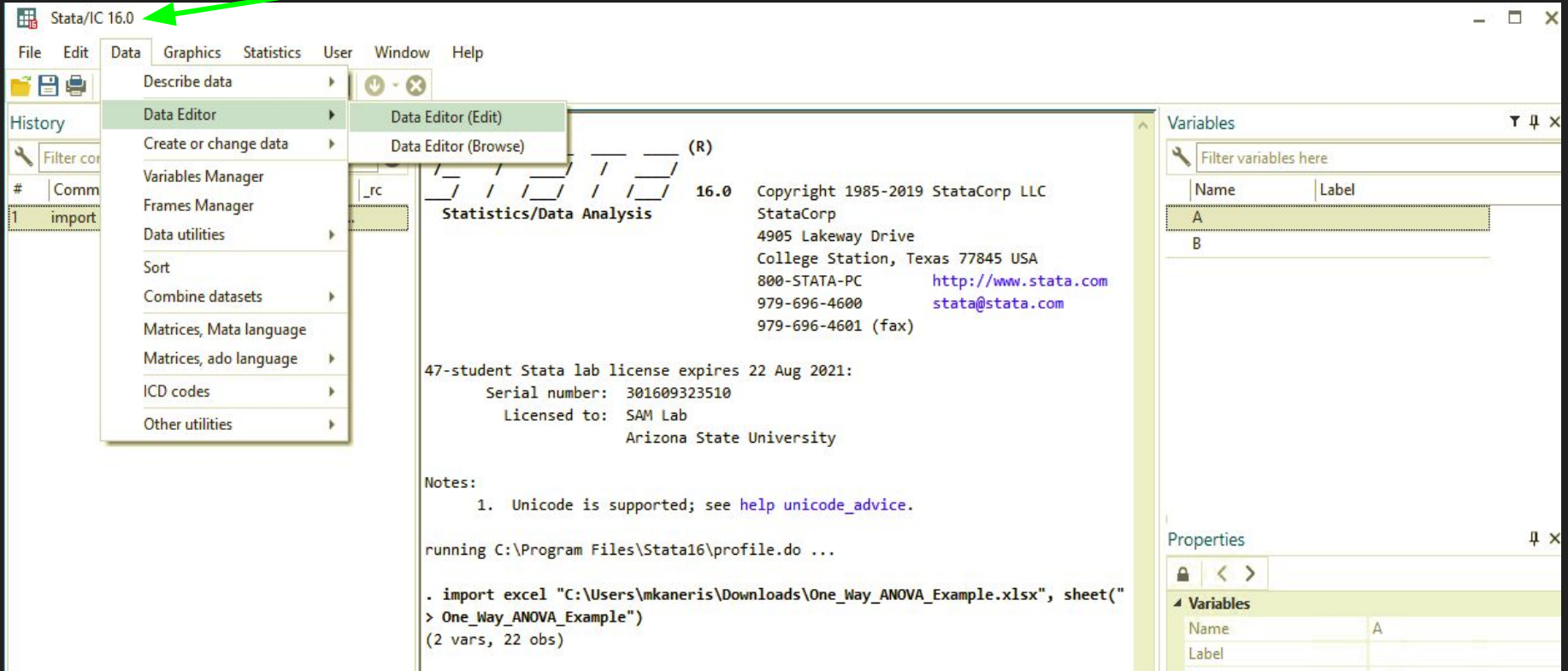
Importing Data

- Stata stores data in a special format that cannot be read by other programs. Stata data files have extension **.dta** . However, you can also import other formats (e.g., .xlsx, .csv, .sav, etc... files) into Stata
- To import data, go to **File -> Import -> select your desired dataset format to import**



Viewing, editing or creating raw data

Select "data" in the menu bar and the appropriate option based on what you are looking to do.



The screenshot shows the Stata/IC 16.0 software interface. The 'Data' menu is open, highlighting 'Data Editor (Edit)'. A green arrow points to the 'Data' menu bar. The main window displays the Stata startup screen with the following text:

```

(R)
Statistics/Data Analysis 16.0 Copyright 1985-2019 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

47-student Stata lab license expires 22 Aug 2021:
Serial number: 301609323510
Licensed to: SAM Lab
Arizona State University

Notes:
1. Unicode is supported; see help unicode_advice.

running C:\Program Files\Stata16\profile.do ...

. import excel "C:\Users\mkaneris\Downloads\One_Way_ANOVA_Example.xlsx", sheet("
> One_Way_ANOVA_Example")
(2 vars, 22 obs)

```

The interface also shows a 'History' panel on the left with 'import' selected, and a 'Variables' panel on the right with a table:

Name	Label
A	
B	

Below the Variables panel is a 'Properties' panel with a table:

Name	Label
A	
Label	

Data Editor (Edit) - [Untitled]

File Edit View Data Tools

Animal[1]

Raw data values visible on the left side.

	Animal	Gender	Relationship-s	Happiness
1	1	1	4	2
2	1	1	1	3
3	1	1	2	2
4	1	0	1	2
5	1	0	1	2
6	1	0	1	3
7	1	0	3	1
8	2	1	2	6
9	2	1	2	7
10	2	1	1	6
11	2	1	2	7
12	2	0	2	4
13	2	0	1	4
14	2	0	1	1
15	3	1	4	4
16	3	1	2	5
17	3	1	1	4
18	3	0	3	1
19	3	0	2	3
20	3	0	3	2
21	3	0	3	3

Information about variables on the right.

Variables

Filter variables here

<input checked="" type="checkbox"/>	Name	Label	Type	Format	Value la
<input checked="" type="checkbox"/>	Animal	Animal	byte	%10.0g	
<input checked="" type="checkbox"/>	Gender	Gender	byte	%10.0g	
<input checked="" type="checkbox"/>	Relationship_Sta...	Relationship_Status	byte	%10.0g	
<input checked="" type="checkbox"/>	Happiness	Happiness	byte	%10.0g	

Variables | Snapshots

Properties

Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

Frame	default
Filename	
Label	
Notes	
Variables	4
Observations	21
Size	84
Memory	64M

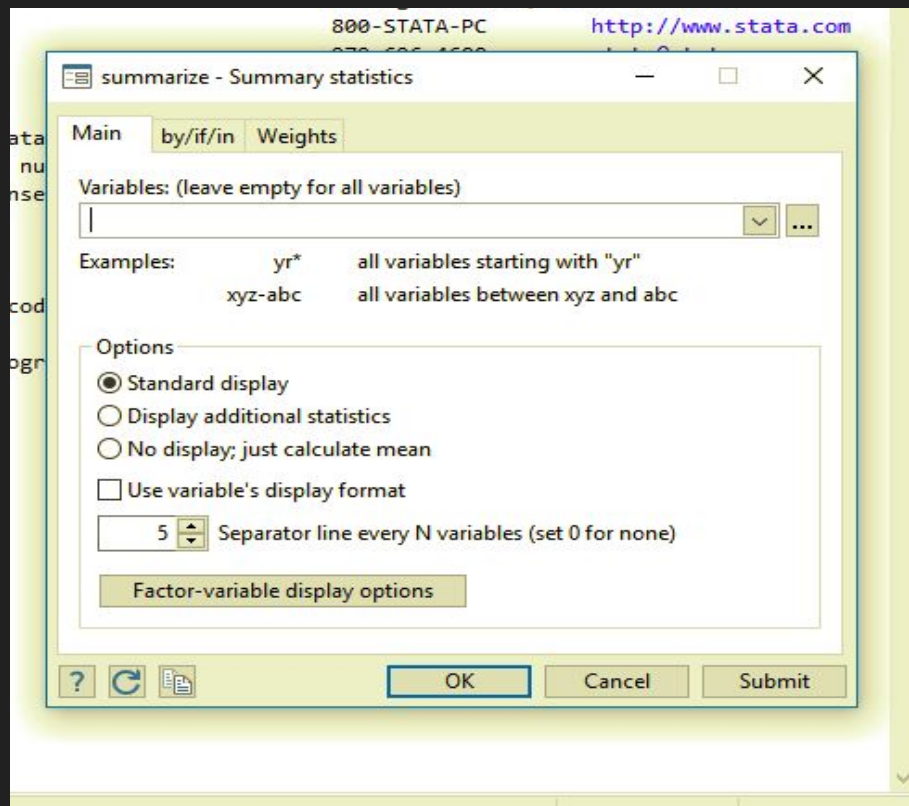
STATA Interface

Methods of navigation:

1. Menu system & dialog boxes
2. Command window
3. Do-file

1. Menu system and dialog boxes

Dialog Box



2. Using the Command Window

Once performed, results of your commands will be displayed in this window

A list of your recent commands will be displayed here

The screenshot displays the Stata/IC 16.0 software interface. The main window is the Command Window, which is highlighted with a green border. It contains the following text:

```
STATA (R) 16.0 Copyright 1985-2019 StataCorp LLC
Statistics/Data Analysis
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

17-student Stata lab license expires 22 Aug 2021:
Serial number: 301609323510
Licensed to: SAM Lab
Arizona State University

Notes:
1. Unicode is supported; see help unicode_advice.

Running C:\Program Files\Stata16\profile.do ...
```

Below the Command Window is a blue bar labeled "Command" with a cursor, indicating where to type commands.

On the left side, the History panel is visible, showing a search filter and a table with columns for command number and command text. It currently displays "There are no items to show."

On the right side, the Variables panel is visible, showing a search filter and a table with columns for Name and Label. It also displays "There are no items to show."

At the bottom right, the Properties panel is visible, showing a search filter and a table with columns for Name and Value. It is currently expanded to show the "Variables" section, which includes fields for Name, Label, Type, Format, Value label, and Notes. Below this is the "Data" section, which includes fields for Frame, Filename, Label, Notes, Variables (0), Observations (0), Size (0), Memory (64M), and Sorted by.

Here is where you type your commands (i.e., tasks/ analysis you want to perform)

2. Using the Command Window

You tell Stata what to do by typing commands in the Command Window. The general form of all Stata commands is:

command **variables**, **options**

command tells STATA which command you want to execute.

variables (each variable name separated by a space) are the list of variables used to perform the command.

options tells Stata how you want to execute the command.

Example:

Say we want to list the number of observations we have, specifically for the variable “Animal”

We could type the following in the command window:

list **Animal**



History



#	Command	_rc
1	import excel "C:\Users\mkaneris\Downloads...	
3	list Animal	

. list Animal

	Animal
1.	1
2.	1
3.	1
4.	1
5.	1
6.	1
7.	1
8.	2
9.	2
10.	2
11.	2
12.	2
13.	2
14.	2
15.	3
16.	3
17.	3
18.	3
19.	3
20.	3
21.	3

Command

list Animal

Variables



Name	Label
Animal	Animal
Gender	Gender
Relationship_Sta...	Relationship_Status
Happiness	Happiness

Properties



Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	4
Observations	21
Size	84
Memory	64M
Sorted by	

2. Using the Command Window

Now, say we want to run descriptive statistics (e.g., # of observations, mean, standard deviation) for this “Animal” variable. We would type in the following command:

```
summarize Animal
```

If we want to know more descriptive stats (e.g., skew or kurtosis), we could add a “detail” option to the above command.

```
summarize Animal, detail
```



History



Filter commands here



#	Command	_rc
1	import excel "C:\Users\mkaneris\Downloads...	
3	list Animal	
4	summarize Animal	
5	summarize Animal, detail	

```
16.      3
17.      3
18.      3
19.      3
20.      3
21.      3
```

```
. summarize Animal
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Animal	21	2	.83666	1	3

```
. summarize Animal, detail
```

Animal					
Percentiles		Smallest			
1%	1	1			
5%	1	1			
10%	1	1	Obs		21
25%	1	1	Sum of Wgt.		21
50%	2		Mean		2
75%	3	Largest	Std. Dev.		.83666
90%	3	3	Variance		.7
95%	3	3	Skewness		0
99%	3	3	Kurtosis		1.5

Command

Variables



Filter variables here

Name	Label
Animal	Animal
Gender	Gender
Relationship_Sta...	Relationship_Status
Happiness	Happiness

Properties



Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

Frame	default
Filename	
Label	
Notes	
Variables	4
Observations	21
Size	84
Memory	64M
Sorted by	

List of common commands for descriptive information and statistics:

describe provides basic info about your STATA data file (e.g. # of observations, # of variables, etc...)

codebook provides an overview of the different variables in your data file (e.g., name of variables, type of variable, mean, standard deviation etc...)

inspect provides another type of overview of different variables in your data file. Includes plots of observations and unique values.

list provides a list of observations of all variables in your data file (unless you specify only a specific variable)

tabulate (or tab) provides a frequency table of observations for specified variable(s)

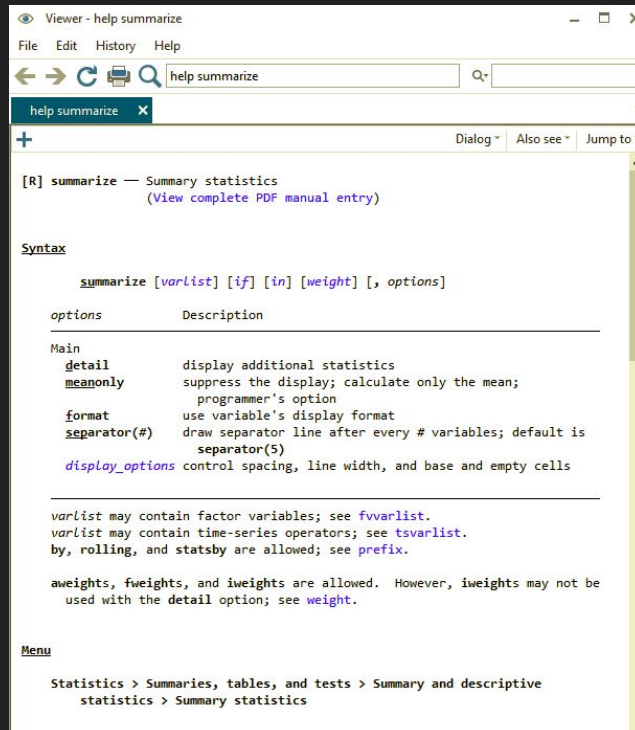
column option allows you to generate column percentages

summarize provides descriptive statistics about a variable(s) (e.g., mean, standard deviation)

detail option provides additional descriptive stats (e.g., skew, kurtosis)

To learn more about what a specific command does

Help-> Stata Command... -> type in the command for detailed information about it



The screenshot shows a window titled "Viewer - help summarize" with a menu bar (File, Edit, History, Help) and a search bar containing "help summarize". The main content area displays the help text for the [R] summarize command, including its syntax and a list of options with their descriptions.

```
[R] summarize — Summary statistics
      (View complete PDF manual entry)

Syntax

      summarize [varlist] [if] [in] [weight] [, options]

options      Description
-----
Main
  detail      display additional statistics
  meanonly    suppress the display; calculate only the mean;
              programmer's option
  format      use variable's display format
  separator( # ) draw separator line after every # variables; default is
              separator(5)
  display_options control spacing, line width, and base and empty cells

varlist may contain factor variables; see fvvarlist.
varlist may contain time-series operators; see tsvarlist.
by, rolling, and statsby are allowed; see prefix.

aweight, fweight, and iweight are allowed. However, iweight may not be
used with the detail option; see weight.

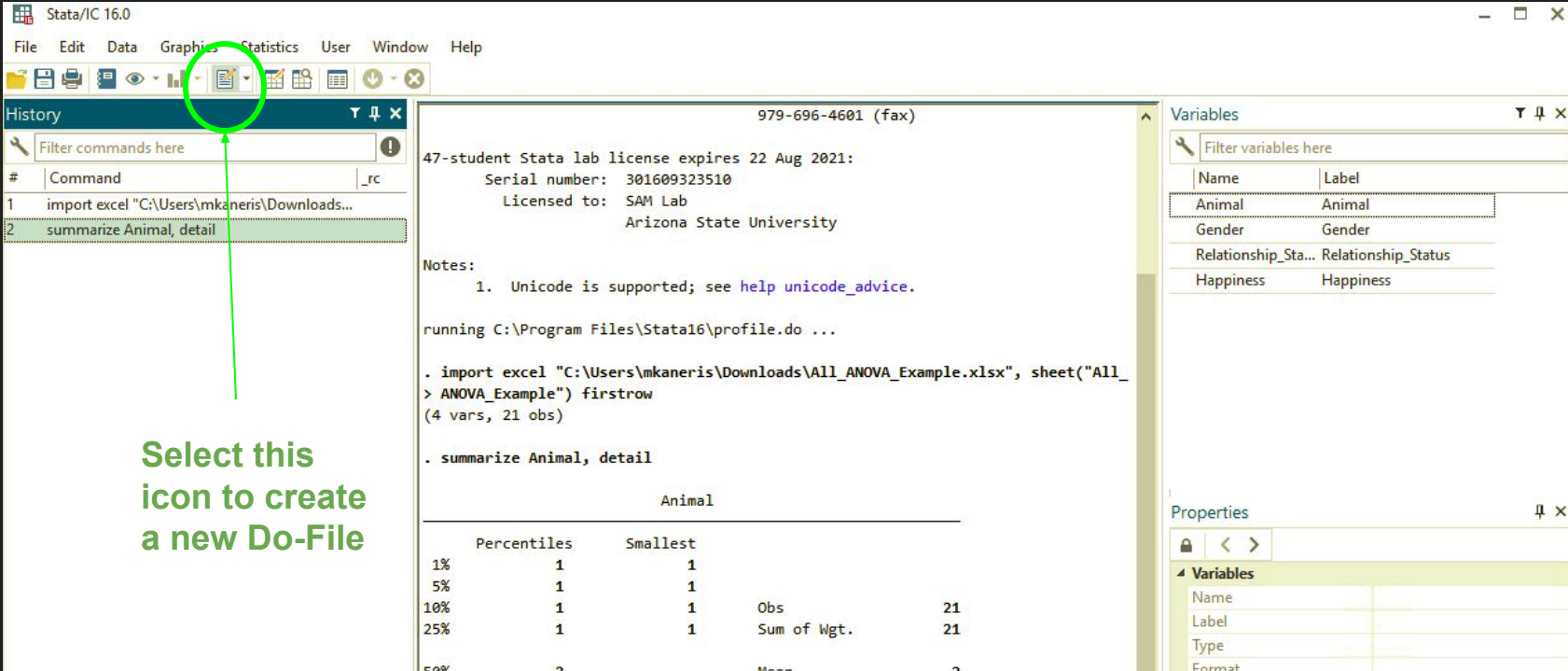
Menu

Statistics > Summaries, tables, and tests > Summary and descriptive
statistics > Summary statistics
```

3. Creating and using a Do-File

To create a new Do-File : **Window -> Do-File Editor -> New Do-File Editor**

OR



The screenshot shows the Stata/IC 16.0 interface. The History window on the left contains the following commands:

```
# Command _rc
1 import excel "C:\Users\mkaneris\Downloads\All_ANOVA_Example.xlsx", sheet("All_
2 summarize Animal, detail
```

The main window displays the following output:

```
979-696-4601 (fax)
47-student Stata lab license expires 22 Aug 2021:
Serial number: 301609323510
Licensed to: SAM Lab
Arizona State University

Notes:
1. Unicode is supported; see help unicode_advice.

running C:\Program Files\Stata16\profile.do ...

. import excel "C:\Users\mkaneris\Downloads\All_ANOVA_Example.xlsx", sheet("All_
> ANOVA_Example") firstrow
(4 vars, 21 obs)

. summarize Animal, detail

              Animal
-----
Percentiles   Smallest
1%             1             1
5%             1             1
10%            1             1   Obs           21
25%            1             1   Sum of Wgt.   21
50%            2
              Mean
```

The Variables window on the right shows the following variables:

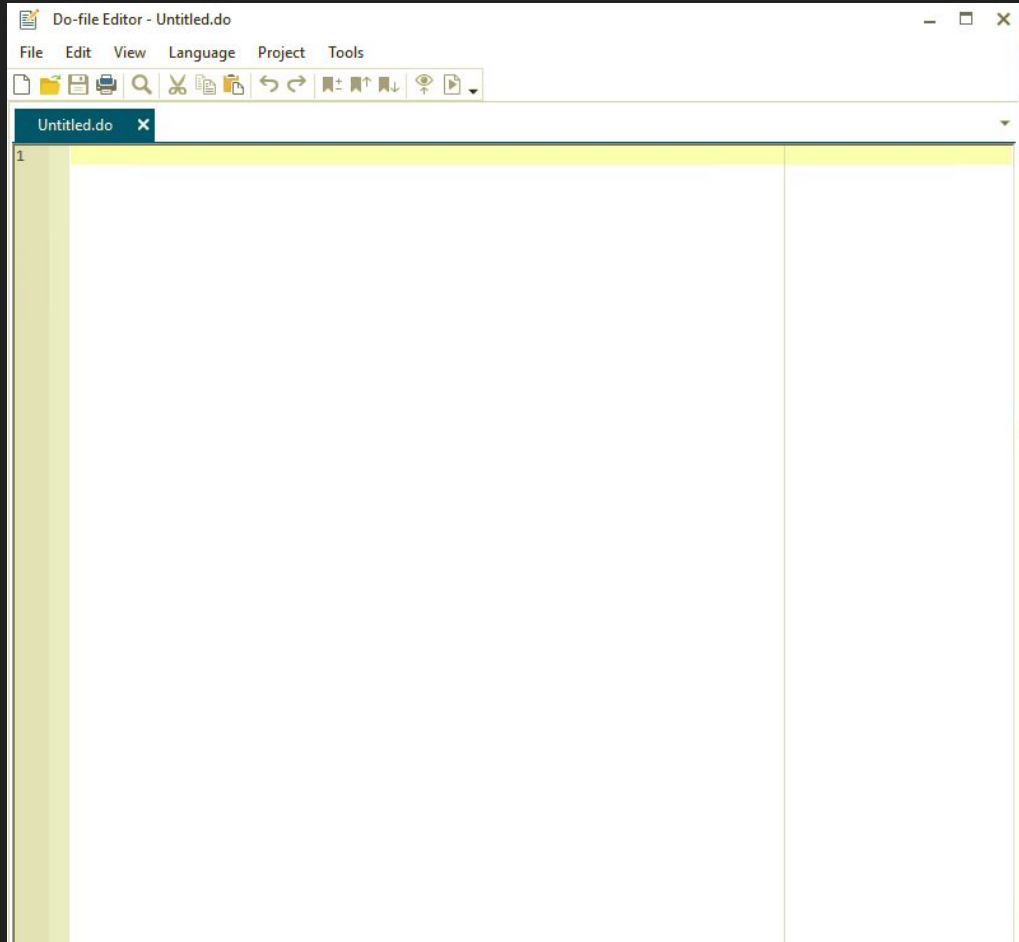
Name	Label
Animal	Animal
Gender	Gender
Relationship_Sta...	Relationship_Status
Happiness	Happiness

The Properties window at the bottom right shows the following variables:

Name	Label	Type	Format
Animal	Animal		
Gender	Gender		
Relationship_Sta...	Relationship_Status		
Happiness	Happiness		

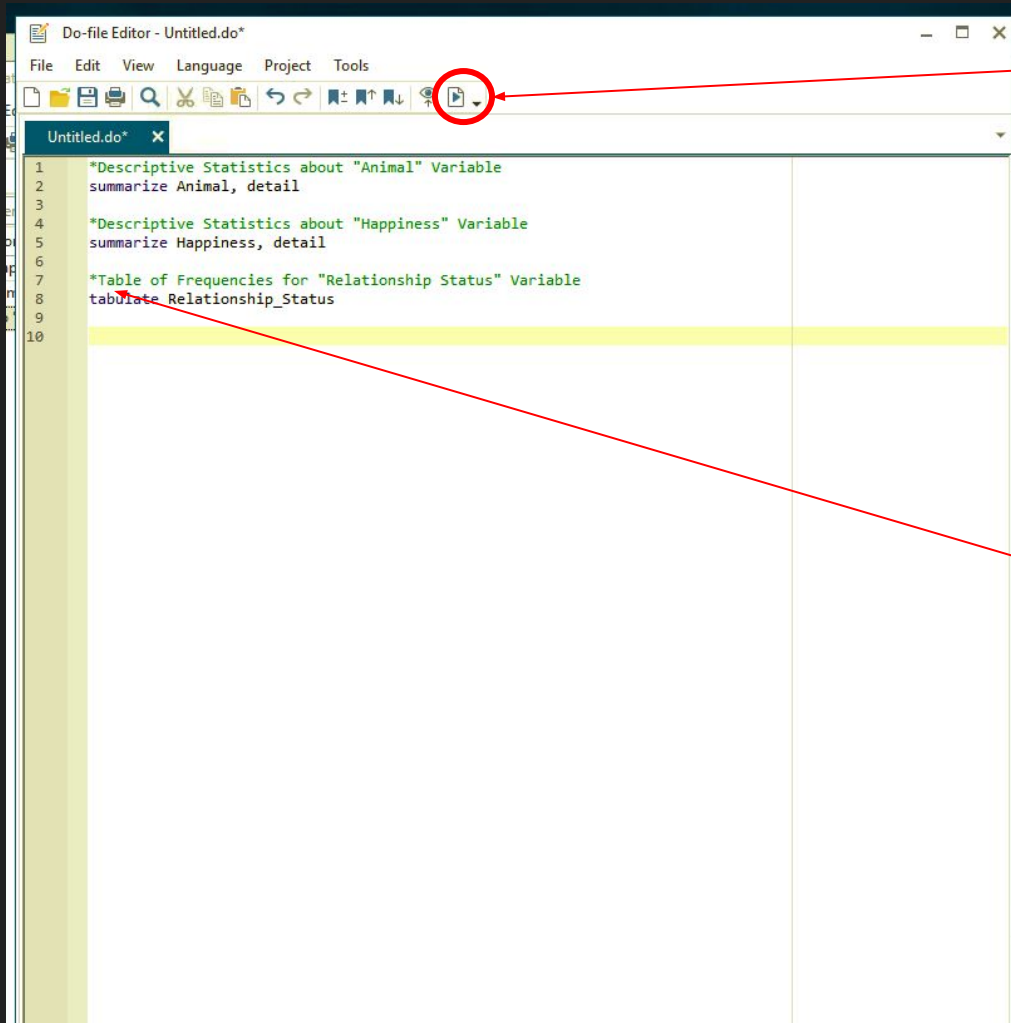
Select this icon to create a new Do-File

3. Creating and using a Do-File



This blank text document is your new Do-File.

You can write the commands you wish to execute in this file, same as you would in the command window.



Select this icon to execute your commands. If you only want to execute a specific command, highlight that specific command in the text and then select this icon to execute it.

Alternatively, you can execute the commands using the menu system. Go to **Tools -> Execute (do)**

To make your Do-File easier to read, you may want to make comments or leave spaces.

To add a comment, put an asterisk (`*`) before typing your comment. The comment will then automatically turn green colored. Note that comments and blank lines will not be executed in Stata.

How to save your Do-File in the Do-File Editor : **File -> Save As**

Simple Practice Exercises:

Q: Does the number of pets someone has *correlate* with their happiness?

Command:

```
correlate Animal Happiness
```

Or Using Menu & Dialog Box:

Statistics-> Summaries, tables, and tests -> Summary and descriptive statistics -> Correlations and covariances

```
. correlate Animal Happiness
(obs=25)
```

	Animal Happiness	
Animal Happiness	1.0000	
	0.8063	1.0000

Simple Practice Exercises:

Graph the correlation between number of pets and happiness

Command:

scatter Y-axis variable X-axis variable

scatter Happiness Animal

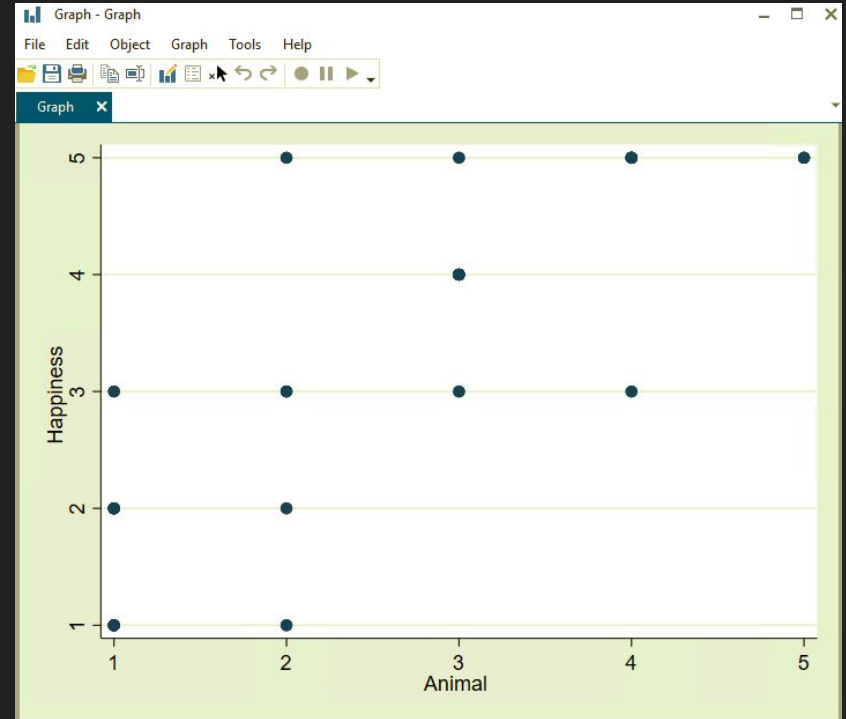
Or using Menu & Dialog Box:

Graphics -> Twoway graph (scatter, line ,etc...)

Once graph is created, to edit it select:

File -> Start Graph Editor

And make the changes you like.



Simple Practice Exercises:

Q. Does the relationship between number of pets one has vary based on gender?

Command for independent samples t-test:

```
ttest Animal, by(Gender)
```

Or using Menu & Dialog Box:

Statistics-> Summaries, tables, and tests -> Classical tests of hypotheses -> t test (mean comparison test)

```
. ttest Animal, by(Gender)
```

```
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	12	2.5	.398862	1.381699	1.622111	3.377889
1	13	2.692308	.3648782	1.315587	1.897306	3.487309
combined	25	2.6	.2645751	1.322876	2.053944	3.146056
diff		-.1923077	.5394759		-1.308299	.9236832

```
diff = mean(0) - mean(1)                                t = -0.3565  
Ho: diff = 0                                           degrees of freedom = 23
```

```
Ha: diff < 0  
Pr(T < t) = 0.3624
```

```
Ha: diff != 0  
Pr(|T| > |t|) = 0.7247
```

```
Ha: diff > 0  
Pr(T > t) = 0.6376
```


Further Resources:

The STATA website has links to numerous types of resources for learning STATA and performing specific analyses, as well as cheatsheets of common commands:

<https://www.stata.com/links/resources-for-learning-stata/#official>

STATA Corp LLC YouTube Channel:

<https://www.youtube.com/channel/UCV4k4G4nEtBS4tLOyHqustDA>

UCLA Institute for Digital Research and Education Statistical Consulting:

<https://stats.idre.ucla.edu/stata/modules/>