Poli 5D Social Science Data Analytics

Functions in Excel (2); Intro to Stata

Shane Xinyang Xuan
ShaneXuan.com

January 26, 2017

Contact Information

Shane Xinyang Xuan

xxuan@ucsd.edu

The teaching staff is a team!

Professor Roberts	M	1600-1800 (SSB 299)
Jason Bigenho	Th	1000-1200 (Econ 116)
Shane Xuan	Th	1200-1400 (SSB 332)

Supplemental Materials

UCLA STATA starter kit

`http://www.ats.ucla.edu/stat/stata/sk/`

Princeton data analysis

`http://dss.princeton.edu/training/`

Road map

Here is what we are going to cover in today's section:

- Logical functions in Excel
- Nested IF functions in Excel
- Macros in Excel
- Introduction to Stata

Road map

Here is what we are going to cover in today's section:

- Logical functions in Excel
- Nested IF functions in Excel
- Macros in Excel
- Introduction to Stata

Let's take a poll really quick – in which order do you want me to cover these topics in today's section?

Road map

Here is what we are going to cover in today's section:

- Logical functions in Excel
- Nested IF functions in Excel
- Macros in Excel
- Introduction to Stata

Let's take a poll really quick – in which order do you want me to cover these topics in today's section?

- Should I slow down in the section?

Road map

Here is what we are going to cover in today's section:

- Logical functions in Excel
- Nested IF functions in Excel
- Macros in Excel
- Introduction to Stata

Let's take a poll really quick – in which order do you want me to cover these topics in today's section?

- Should I slow down in the section?
- Any questions that I can answer (w.r.t lectures, sections, homework, ...)?

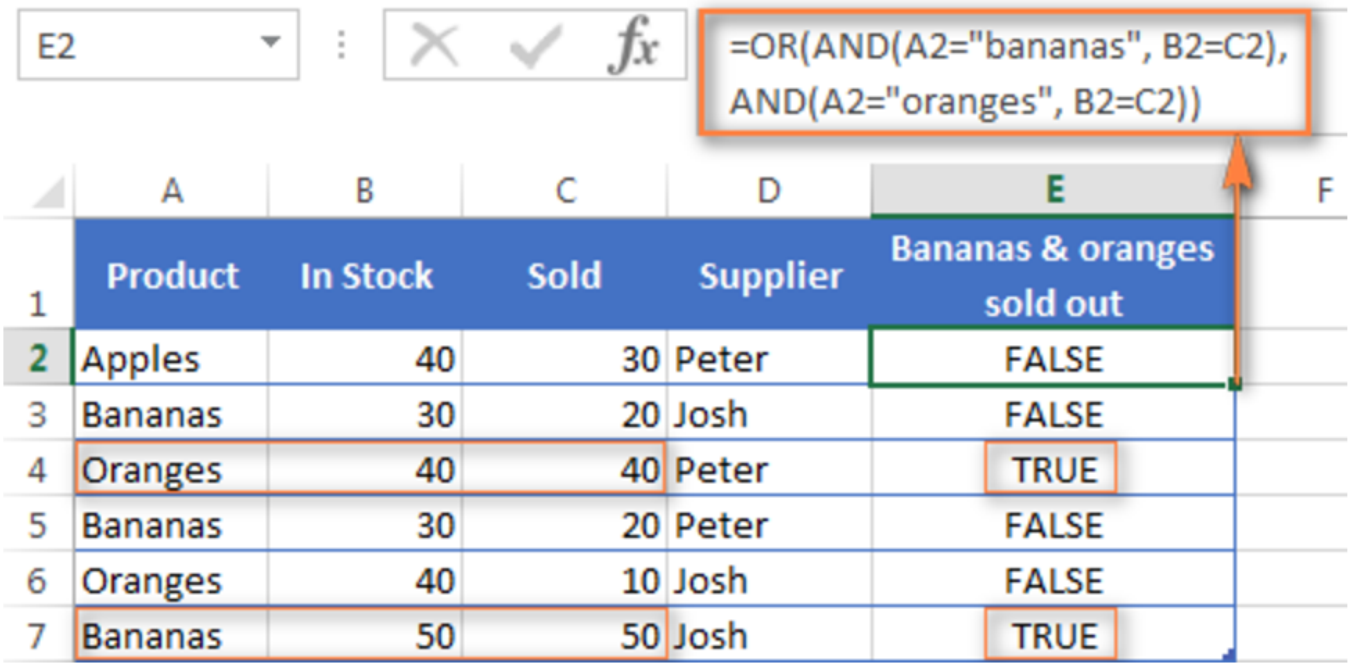
Logical functions

We are going to go through four examples in the next few slides to wrap up logical functions.

Logical functions

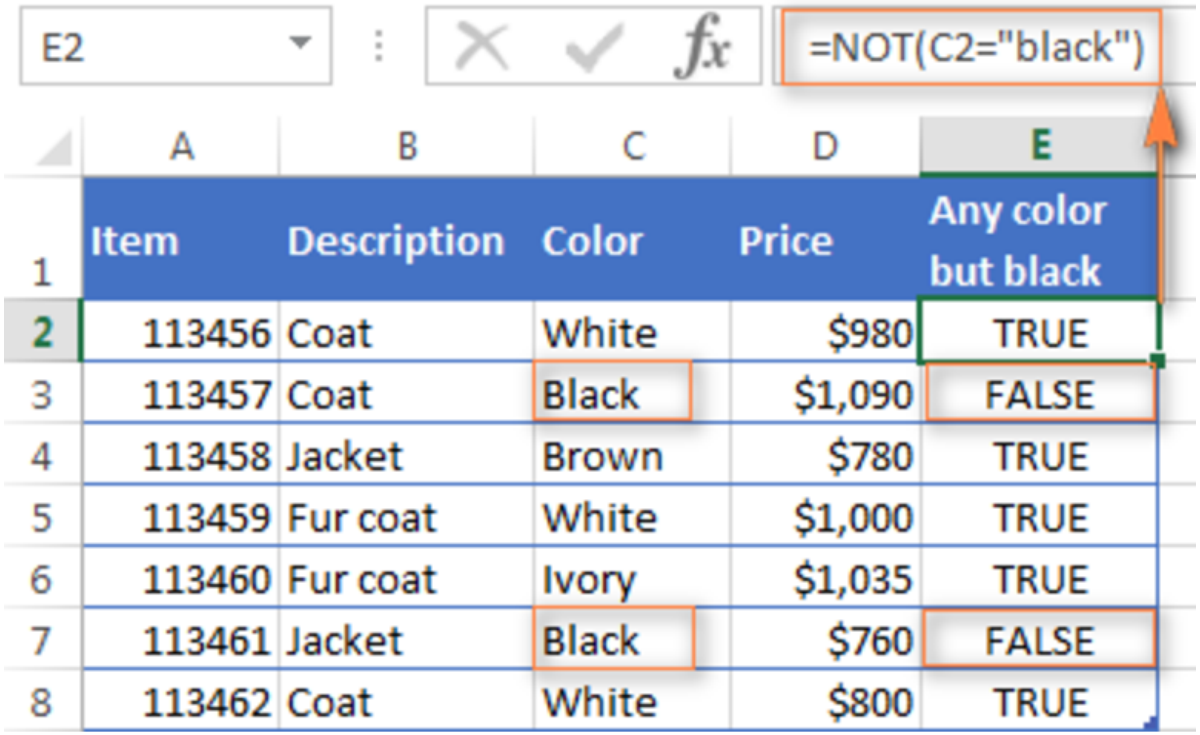
	A	B	C	D	E	F
1	Product	In Stock	Sold	Formula 1	Formula 2	Formula 3
2				=OR(A2="Bananas", A2="Oranges")	=OR(B2>=40, C2>=20)	=OR(B2="", C2="")
3	Bananas	30	10	TRUE	FALSE	FALSE
4	Oranges		20	TRUE	TRUE	TRUE
5	Cherries	20		FALSE	FALSE	TRUE
6	Oranges	30	10	TRUE	FALSE	FALSE
7	Cherries			FALSE	FALSE	TRUE

Logical functions



	A	B	C	D	E	F
1	Product	In Stock	Sold	Supplier	Bananas & oranges sold out	
2	Apples	40	30	Peter	FALSE	
3	Bananas	30	20	Josh	FALSE	
4	Oranges	40	40	Peter	TRUE	
5	Bananas	30	20	Peter	FALSE	
6	Oranges	40	10	Josh	FALSE	
7	Bananas	50	50	Josh	TRUE	

Logical functions



	A	B	C	D	E
1	Item	Description	Color	Price	Any color but black
2	113456	Coat	White	$980	TRUE
3	113457	Coat	Black	$1,090	FALSE
4	113458	Jacket	Brown	$780	TRUE
5	113459	Fur coat	White	$1,000	TRUE
6	113460	Fur coat	Ivory	$1,035	TRUE
7	113461	Jacket	Black	$760	FALSE
8	113462	Coat	White	$800	TRUE

Logical functions

	A	B	C	D
1	Value 1	Value 2	Value 3	Is Value 1 between Value 2 & Value 3?
2				=IF(A2=MEDIAN(A2:C2),"Yes","No")
3	5	1	10	Yes
4	10	6	8	No
5	4	2	5	Yes
6	12	15	3	Yes
7	7-Oct	5-Oct	27-Oct	Yes
8	24-Nov	26-Dec	21-Oct	Yes
9	13-Oct	13-Oct	17-Oct	Yes
10	Bananas	Apples	Cherries	#NUM!
11	Apples	Apricot	Bananas	#NUM!

Nested IF

Scenario: You want to test for multiple outcomes.
Solution: A nested IF function is when one IF function is placed inside another IF function to test an additional condition.

Purpose: To determine the fee for a driver's license

Logic Scenario: Driver's license fee varies by age

Below 16	"Too Young"
16–45	$30
46–60	$25
61 and older	$20

Formula: Nested IF functions

```
=IF(B1<16,"Too Young",IF(B1<=45,30,IF(B1<=60,25,20)))
```

Data: cell B1 stores the driver's age

Example:

Data	Condition1	Condition2	Condition3	Results
Cell B1	B1<16	B1<=45	B1<=60	(Fee)
15	True	Not evaluated	Not evaluated	Too Young
25	False	True	Not evaluated	30
55	False	False	True	25
65	False	False	False	20

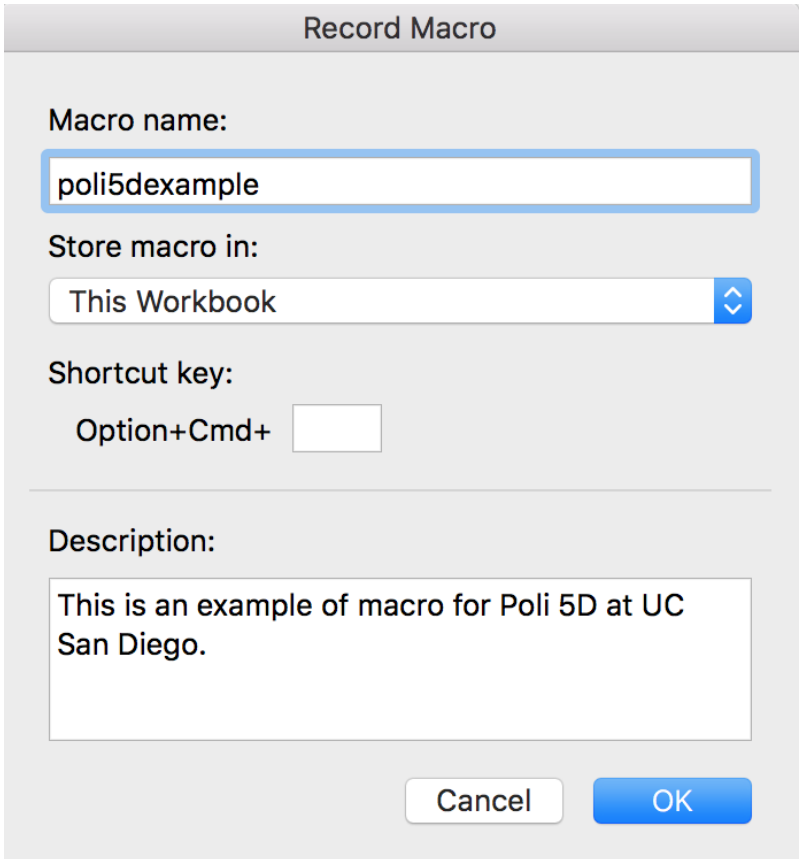
Macros in Excel

Scenario: We want to perform repetitive tasks consistently
Solution: Use macros!

- You record a macro
- You then apply the macro to other cells

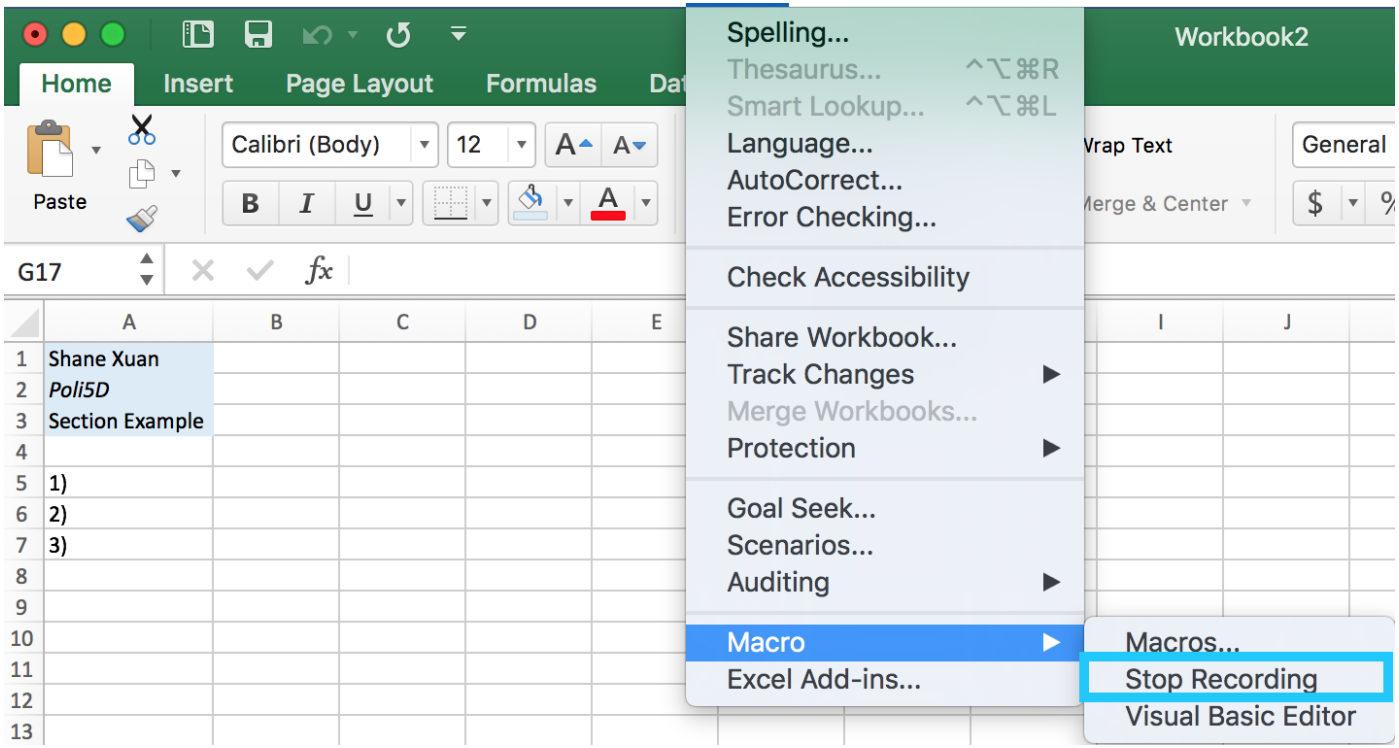
Macros in Excel

Record Macro



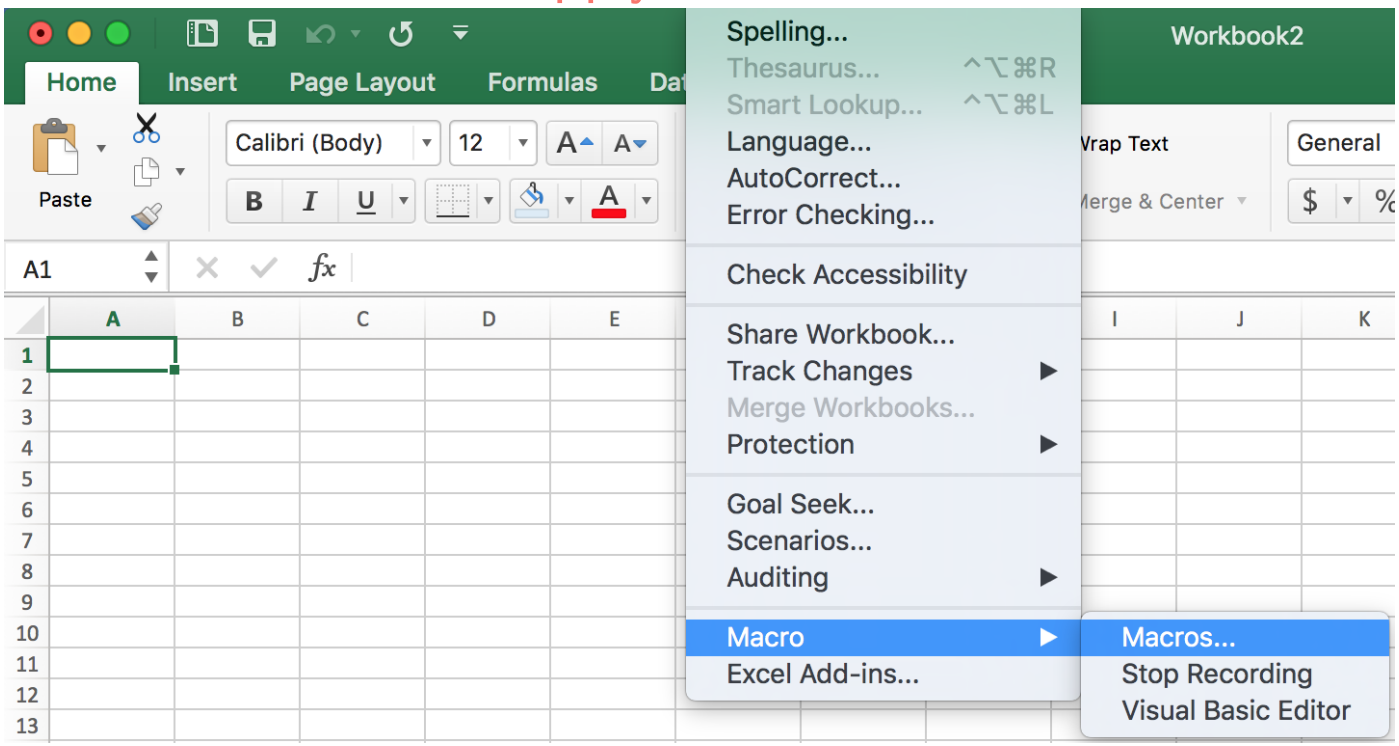
Macros in Excel

Stop Recording



Macros in Excel

In a new worksheet, apply Macros



Macros in Excel

What's under the hood?

Visual Basic Editor

```
Sub poli5dexample()
'
' poli5dexample Macro
' This is an example of macro for Poli 5D at UC San Diego.
'

'
    ActiveCell.FormulaR1C1 = "Shane Xuan"
    Range("A2").Select
    ActiveCell.FormulaR1C1 = "Poli5D"
    Range("A3").Select
    ActiveCell.FormulaR1C1 = "Section Example"
    Range("A5").Select
    ActiveCell.FormulaR1C1 = "1)"
    Range("A6").Select
    ActiveCell.FormulaR1C1 = "2)"
    Range("A7").Select
    ActiveCell.FormulaR1C1 = "3)"
    Range("A1:A3").Select
    ExecuteExcel4Macro "PATTERNS(1,0,5,TRUE,2,4,0,0.799981688894314)"
    Columns("A:A").EntireColumn.AutoFit
    Range("A2").Select
    Selection.Font.Italic = True
End Sub
```

We are done with Excel!

We are done with Excel! What to do next?

- Check out VBA if you want know more about Excel
- We are going to move on to Stata and R for the rest of the quarter

We are done with Excel!

We are done with Excel! What to do next?

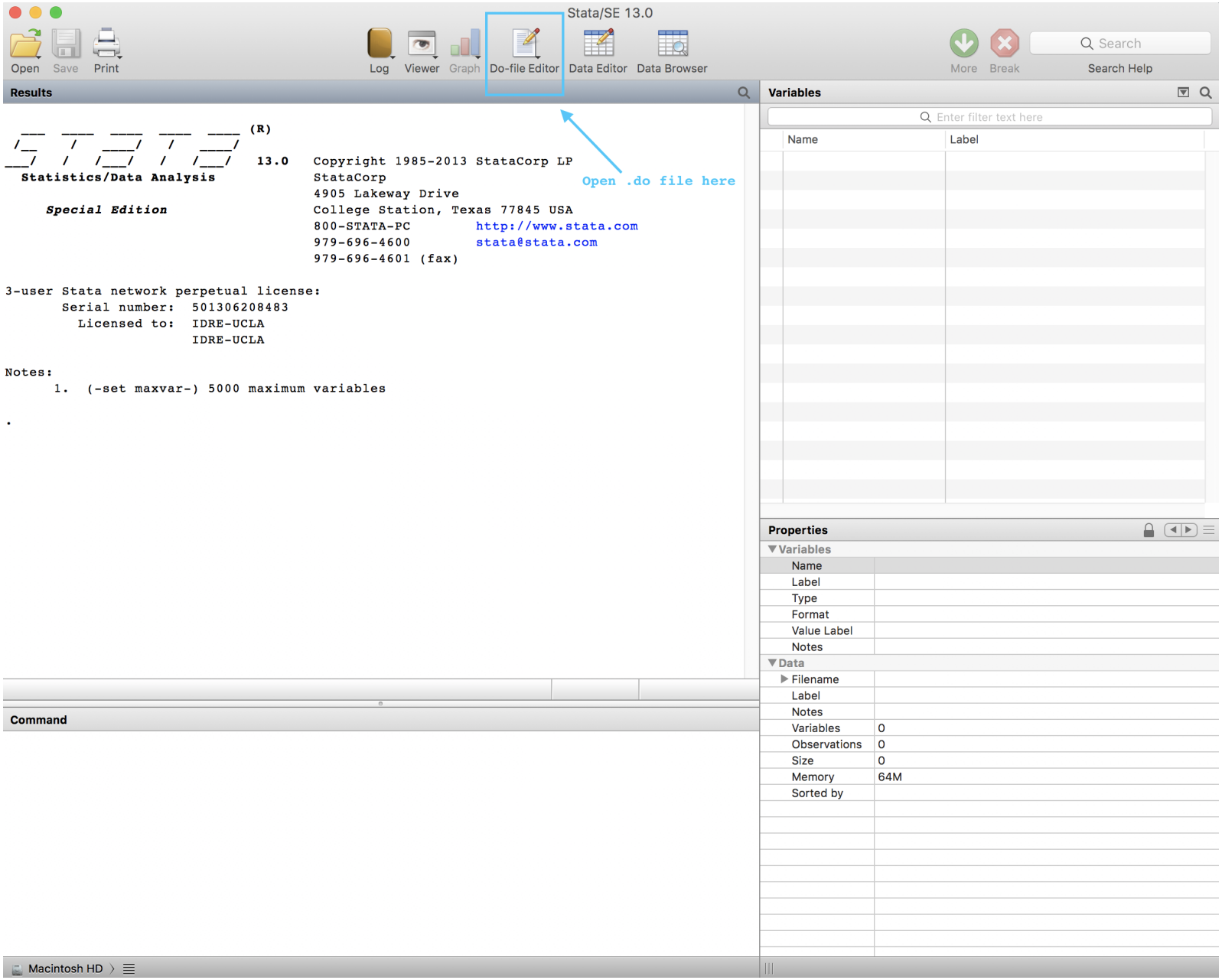
- Check out VBA if you want know more about Excel
- We are going to move on to Stata and R for the rest of the quarter

Before then, let's take attendance.

Introduction to Stata

- Change CD (cd "/Users/Shane/Dropbox/Poli5D/psets/")
- Import data
- Summarize variable
- IF statement in Stata
- Lookup function in Stata
- Help command
- Sort command
- Create .log files

Import data in Stata



Import data in Stata

Example:

```
import excel "h1_fams_data.xlsx", sheet("Families") firstrow clear
```

Import data in Stata

Example:

```
import excel "h1_fams_data.xlsx", sheet("Families") firstrow clear
```

- import data

Import data in Stata

Example:

import excel "h1_fams_data.xlsx", sheet("Families") firstrow clear

- specify sheet name

Import data in Stata

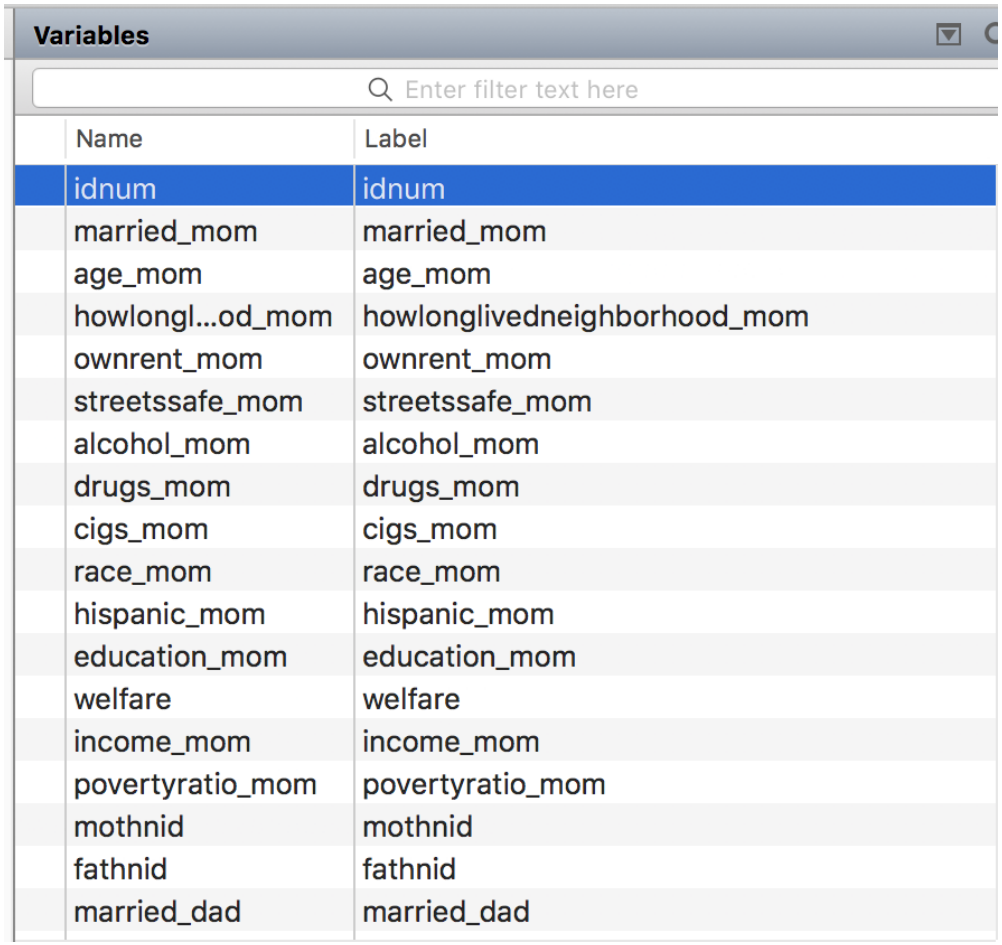
Example:

import excel "h1_fams_data.xlsx", sheet("Families") firstrow clear

- Make the first row to be variable names

Import data in Stata

Your data look like:



Summarize variables in Stata

Example:

```
. summarize age_mom
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age_mom	3895	25.2439	6.009458	15	43

Summarize variables in Stata

Example:

```
. tab married_mom
```

married_mom	Freq.	Percent	Cum.
FUnKnown	9	0.23	0.23
No	2,913	74.96	75.19
Yes	964	24.81	100.00
Total	3,886	100.00	

Summarize variables in Stata

Example:

```
. tab married_mom race_mom
```

married_mom	race_mom					
	AmInd	Asian	Blck	Hispanic	Other	Total
FUnKnown	0	0	8	0	0	9
No	130	42	1,636	3	356	2,856
Yes	37	63	251	0	99	949
Total	167	105	1,895	3	455	3,814

IF statements in Stata

Example:

```
. count if mom_older == 1
  623

. count if mom_older != 1
 3276
```

Syntax:

- Count if (variable == value)
- Other operations: != > < <= >=

Note: In Stata, the single equal (=) is used as a set equal operator; the double equals (==) is used to test for equality; exclamation mark (!) means "NOT"

IF statements in Stata

Example:

```
. count if (mom_older == 1) & (race_mom=="Asian")
   19

. count if  (mom_older == 1) | (race_mom=="Asian")
  710
```

Note: In Stata, & means "AND", and | means "OR"

IF statements in Stata

Example:

```
. summarize howlonglivedneighborhood_dad if mom_older == 1
```

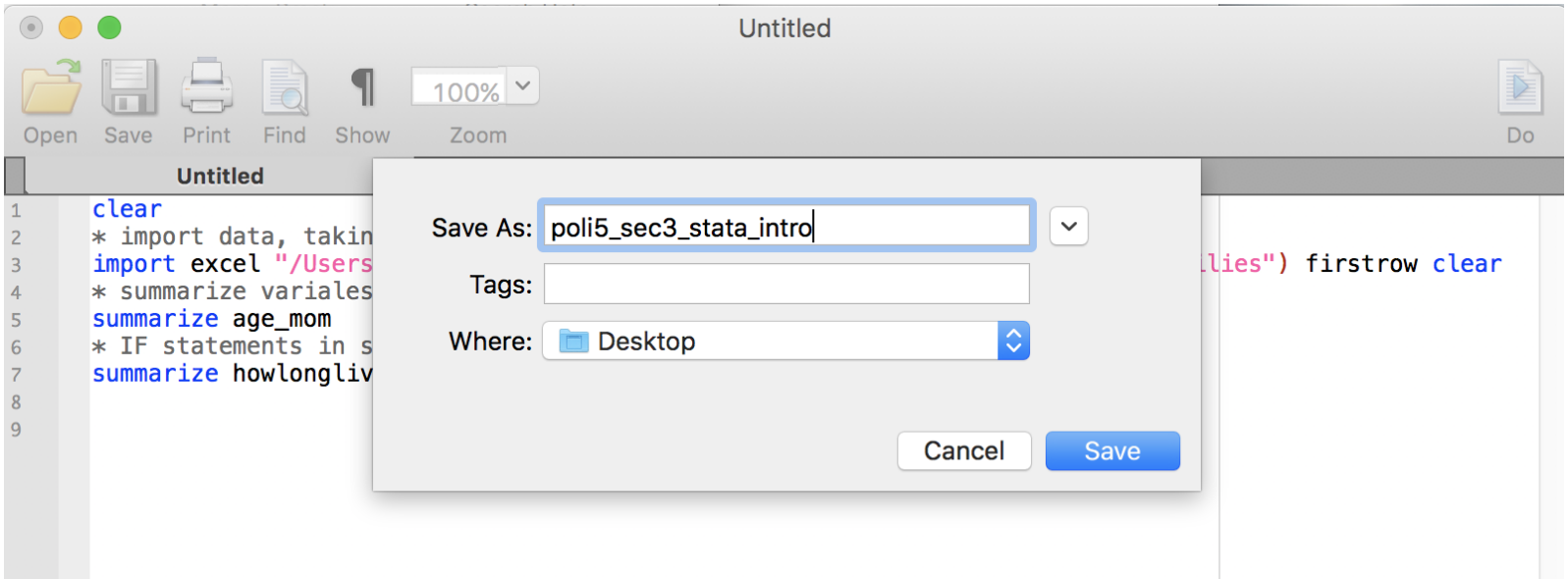
Variable	Obs	Mean	Std. Dev.	Min	Max
howlongliv~d	582	2.051546	3.045963	0	12

Note: In Stata, the single equal (=) is used as a set equal operator; the double equals (==) is used to test for equality

Others

- Lookup function:
  list if (variable) == (value) & (variable) != .
- Help command: help tabulate
- Sort command: sort (variable)
- .log file:
  - log using "log_example", text
  - clear more
    ⋮ ←(your code here)
  - log close

Save your .do file!



Concluding Remarks

We covered many things (but definitely not everything) that are related to your Problem Set 2. Make sure that you

1. Do the readings (!!)
2. Read lecture slides and section slides
3. Learn through trial and error

Be ready to put in hard work while doing the second problem set. Start early, because I hold office hours the day after the due date.