Stata Tutorial: Session I

Basics

- Interface
- Workflow
- Understanding command structure
- Frequently used commands
- Data types

Stata Interface

- Starting Stata: 4 ways
  - Start Menu
  - Icon
  - .do file
  - .dta file

Stata/IC 15.0 - C:\Users\ckilby\Documents\Doc\Teach\Econometrics\Stata tutorials\DA1_B.dta

File Edit Data Graphics Statistics User Window Help

lnUSaid	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnpop	.6326675	.0760239	8.32	0.000	.482183	.7831521
lnGDPpc	-.7523999	.1594348	-4.72	0.000	-1.067991	-.4368085
_cons	-.1132217	1.870742	-0.06	0.952	-3.81624	3.589797

```
. reg USA_TOFG_real pop GDPpc if year==2014
```

Source	SS	df	MS		
				Number of obs	= 163
				F(2, 160)	= 5.02
Model	522583.264	2	261291.632	Prob > F	= 0.0077
Residual	8323060.52	160	52019.1282	R-squared	= 0.0591
				Adj R-squared	= 0.0473
Total	8845643.78	162	54602.7394	Root MSE	= 228.08

USA_TOFG_r~l	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pop	1.36e-07	1.20e-07	1.14	0.257	-1.00e-07	3.73e-07
GDPpc	-.0026992	.0009339	-2.89	0.004	-.0045435	-.0008549
_cons	144.9632	22.89221	6.33	0.000	99.75337	190.1731

```
.
```

Review

Filter commands here

#	Command	_rc
42	su GDP_PPP	
43	gen GDPpc=GDP...	
44	gen lnGDPpc=lo...	
45	reg lnUSaidpc ln...	
46	reg lnUSaidpc ln...	
47	reg lnUSaidpc ln...	
48	reg lnUSaidpc ln...	
49	duplicates exam...	
50	reg lnUSaidpc ln...	
51	reg lnUSaid lnpo...	
52	reg lnUSaid lnpo...	
53	gen lnUSaid=log...	
54	reg lnUSaid lnpo...	
55	reg lnUSaid lnpo...	
56	reg USA_TOFG_r...	
57	reg USA_TOFG_r...	
58	reg USA_TOFG_r...	
59	reg USA_TOFG_r...	
60	reg USA_TOFG_r...	
61	reg USA_TOFG_r...	
62	reg lnUSAaid lnp...	111
63	reg lnUSaid lnpo...	
64	reg lnUSaid lnpo...	
65	reg lnUSaid lnpo...	
66	reg lnUSaid lnpo...	
67	reg lnUSaid lnpo...	
68	reg USA_TOFG_r...	

Command

```
reg lnUSaid lnpop lnGDPpc if year==2014
```

Variables

Filter variables here

Name	Label
country	WDI cou
countrycode	Country
cid	WDI cou
region	Region C
year	calendar
deflator	GDP defl
pop	Populatic
GDP_nominal	GDP (cur
GDP_real	GDP (cor
GDP_PPP	GDP, PPP
CAN_ODA_c...	Canada

Properties

Variables

Name	GDP_PPP
Label	GDP, PPP (c
Type	double
Format	%8.0g
Value label	
Notes	

Data

Filename	DA1_B.dta
Label	Data Analys
Notes	
Variables	271
Observations	7,265
Size	7.91M

C:\Users\ckilby\Documents\Doc\Teach\Econometrics\Stata tutorials

CAP NUM OVR

11:56 PM
9/9/2017

Workflow

Economic analysis is a scientific research: others must be able to replicate your results

➔ Need to document exactly what you do

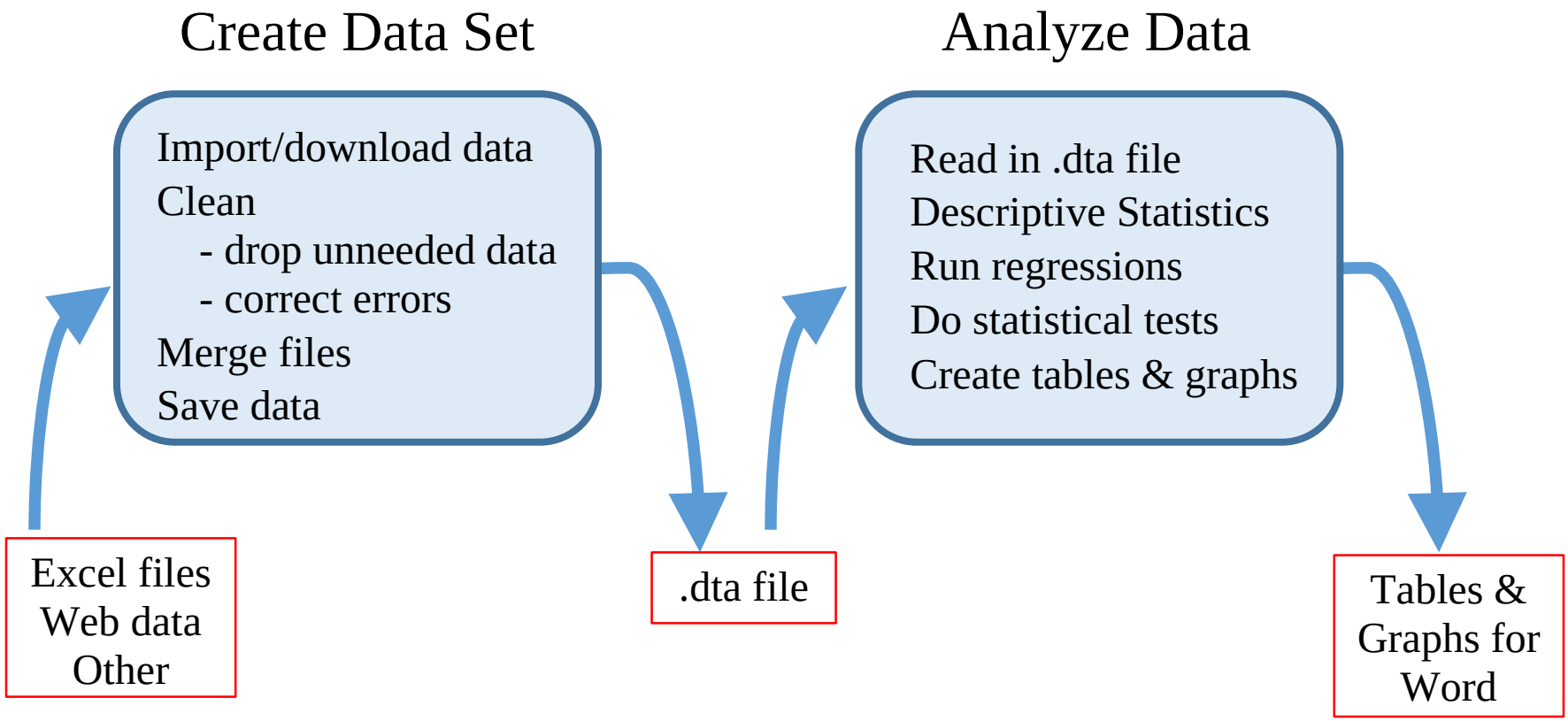
- when creating data set
- when analyzing data set

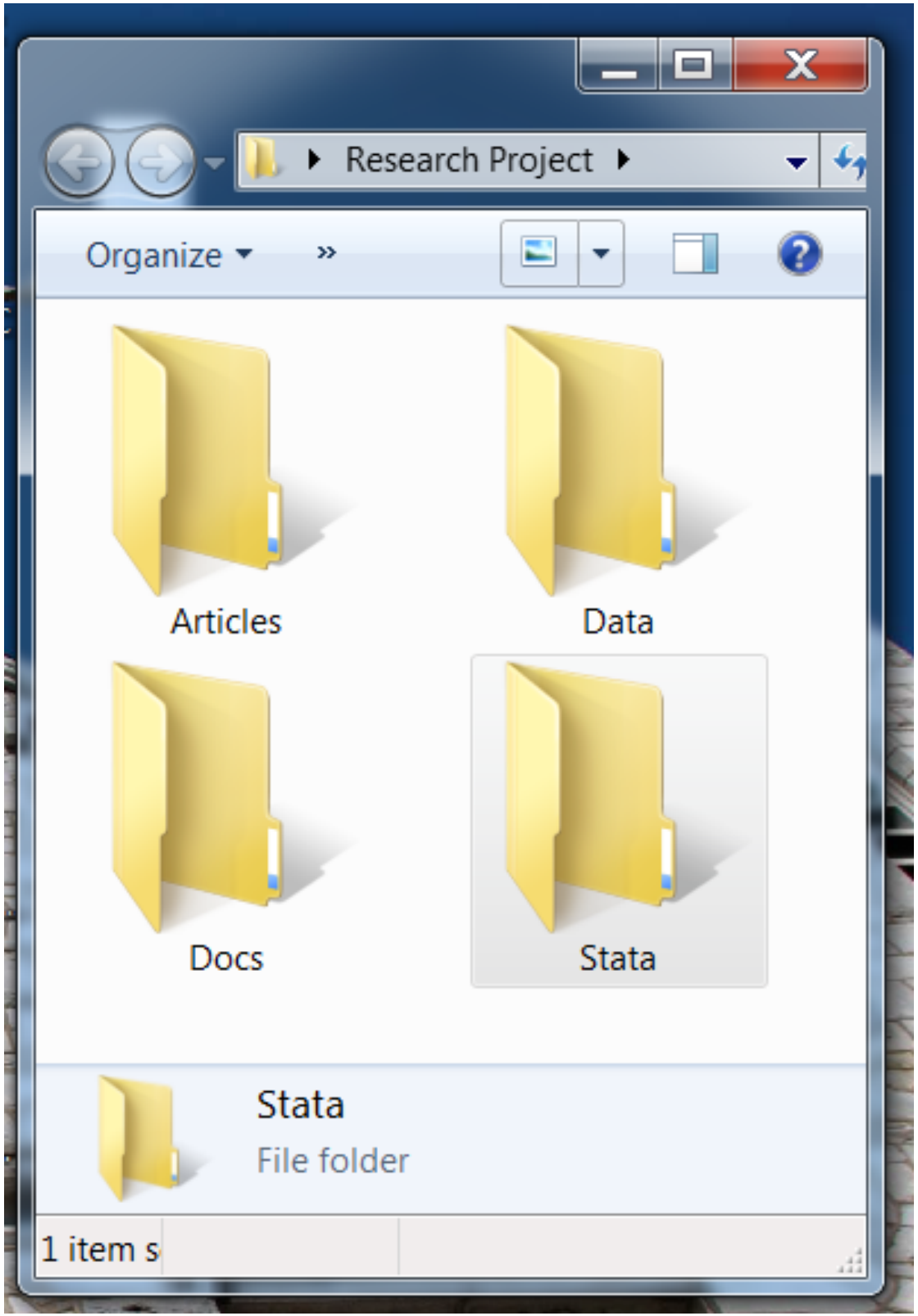
Goal is to create “program” (also called batch file—a Stata .do file) that creates data set & another that carries out analysis.

Stata has 2 modes—interactive & batch files

- Experiment in interactive mode
  - Using menu & dialog boxes
  - Typing commands directly
- Copy commands into .do file
  - Edit .do file
  - Save & Run!

Typical Workflow for Research Projects




Research Project
Organize
Articles
Data
Docs
Stata
Stata
File folder
1 item s

Try it yourself!

- Create folder named “data”
- Go to my website and click on date
  http://www.homepage.villanova.edu/christopher.kilby/
- Download `DA1_B.dta` to your new “data” folder

Double click on `DA1_B.dta`

Understanding Stata's command structure

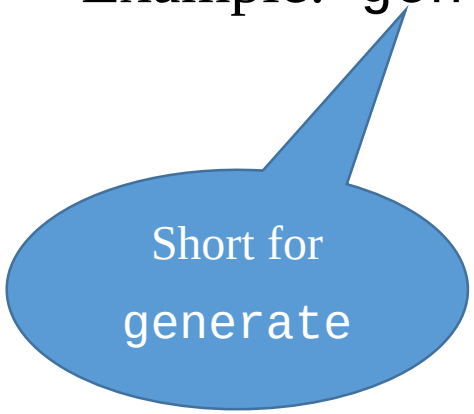
Most Stata commands manipulate or analyze data

- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row, etc.
  - Example: `gen Z = X + Y`

Understanding Stata's command structure

Most Stata commands manipulate or analyze data

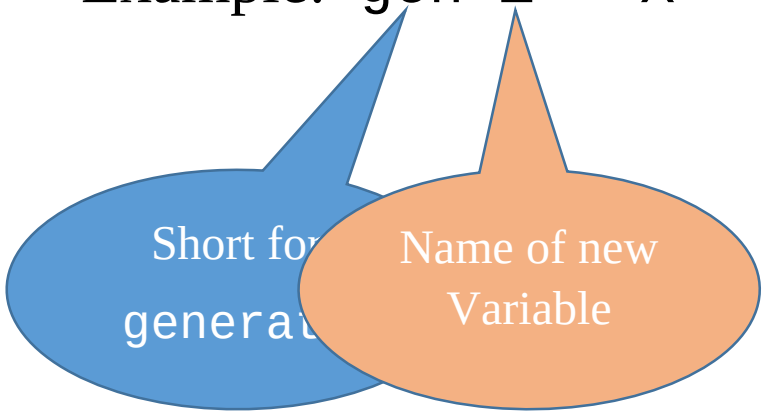
- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row, etc.
  - Example: `gen Z = X + Y`



Understanding Stata's command structure

Most Stata commands manipulate or analyze data

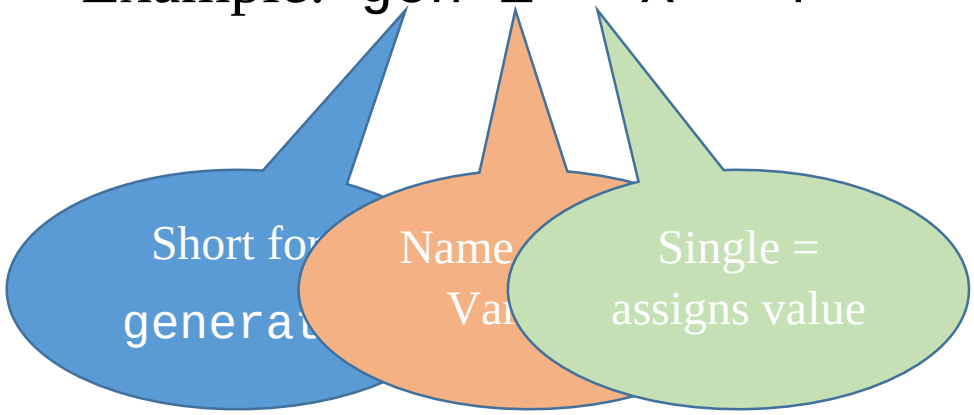
- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row, etc.
  - Example: `gen Z = X + Y`



Understanding Stata's command structure

Most Stata commands manipulate or analyze data

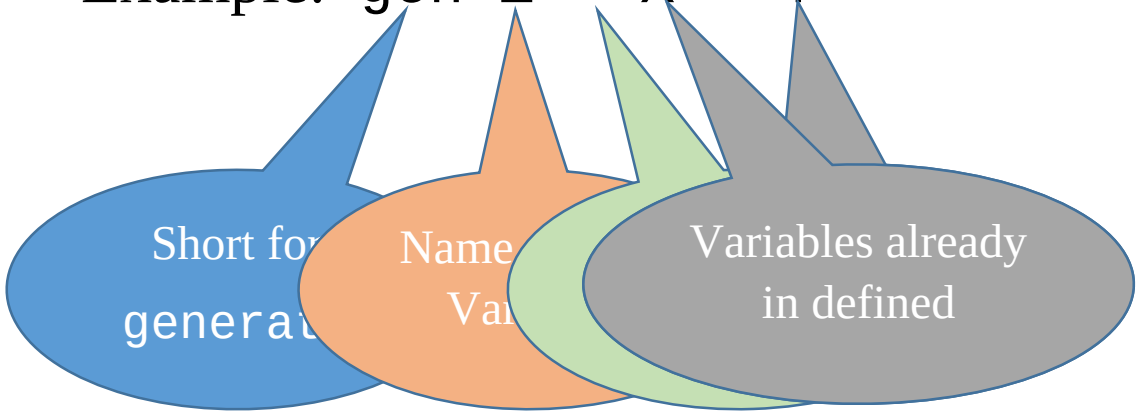
- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row, etc.
  - Example: `gen Z = X + Y`



Understanding Stata's command structure

Most Stata commands manipulate or analyze data

- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row, etc.
  - Example: `gen Z = X + Y`



Understanding Stata's command structure

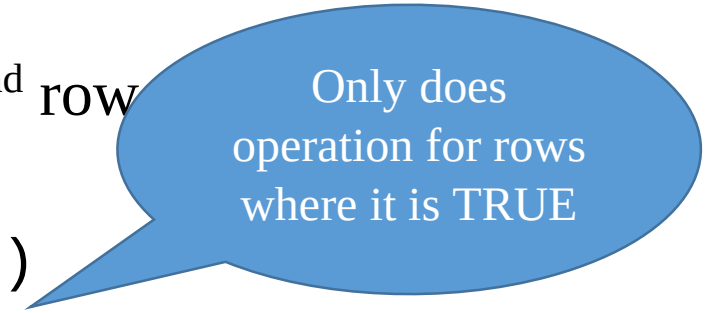
Most Stata commands manipulate or analyze data

- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row, etc.
  - Example: `gen Z = X + Y`
  - Example: `gen lnX = log(X)`
  - Example:
```
gen absY = Y if Y>=0
replace absY = -1*Y if Y<0
```

Understanding Stata's command structure

Most Stata commands manipulate or analyze data

- Data organized like Spreadsheet
  - Columns=variables
  - Rows=observations (data points)
- Commands process 1st row, then 2nd row
  - Example: `gen Z = X + Y`
  - Example: `gen lnX = log(X)`
  - Example:
```
gen absY = Y if Y>=0
replace absY = -1*Y if Y<0
```



Understanding Stata's command structure

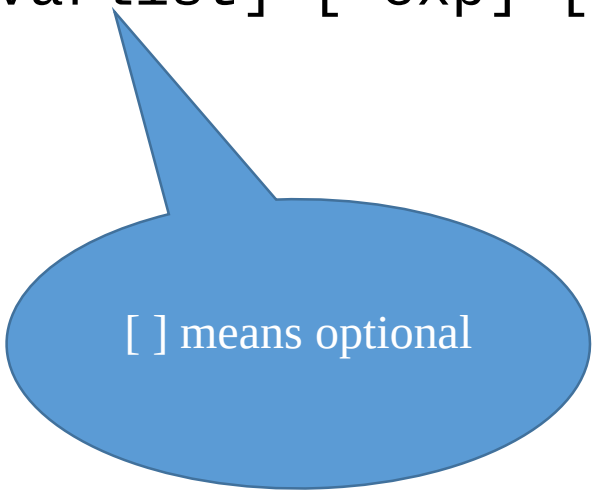
Basic language syntax

```
command [varlist] [=exp] [if] [, options]
```

Understanding Stata's command structure

Basic language syntax

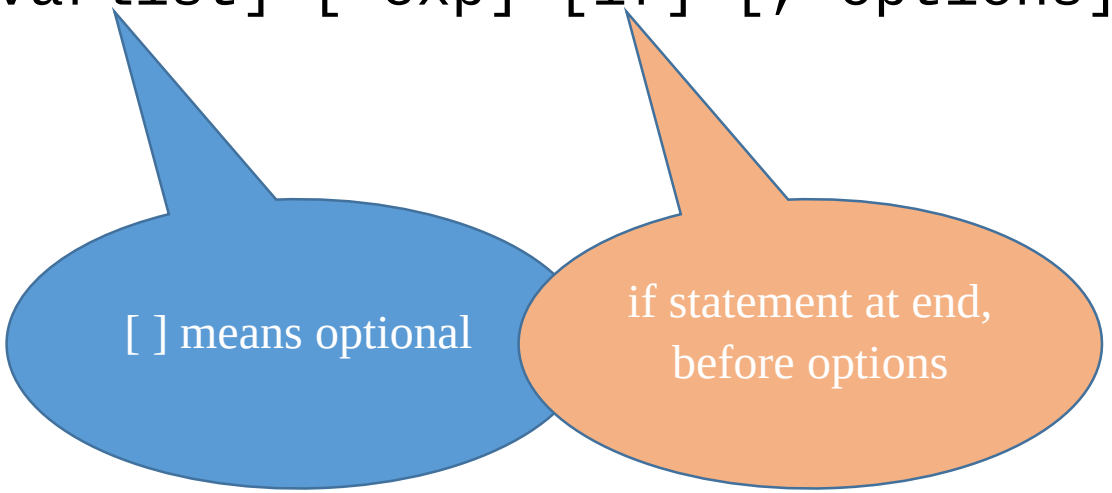
```
command [varlist] [=exp] [if] [, options]
```



Understanding Stata's command structure

Basic language syntax

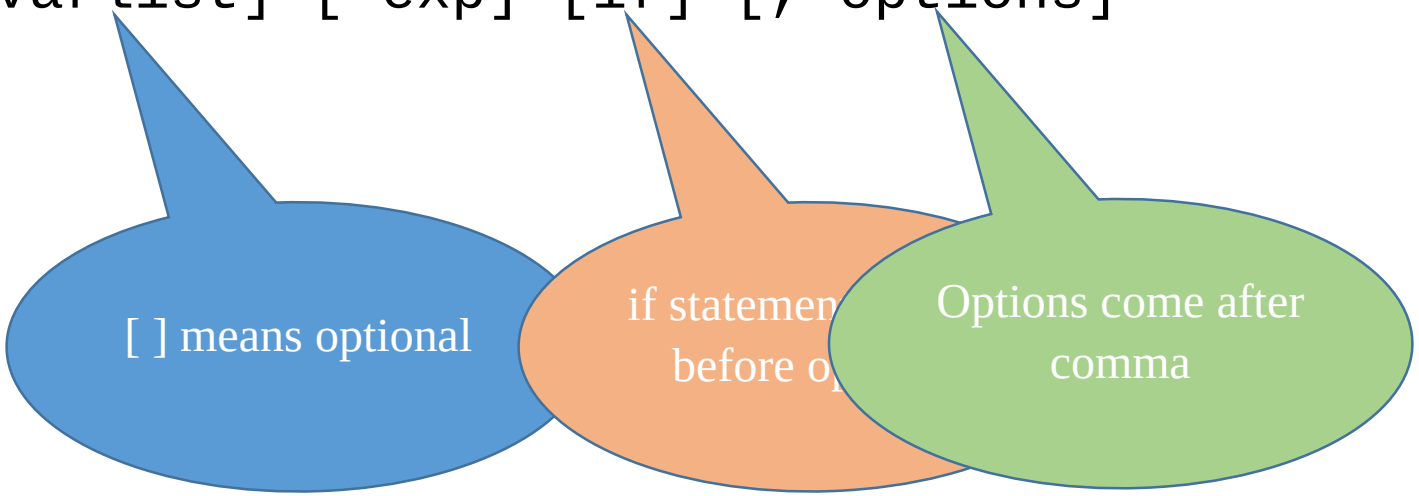
```
command [varlist] [=exp] [if] [, options]
```



Understanding Stata's command structure

Basic language syntax

```
command [varlist] [=exp] [if] [, options]
```



Understanding Stata's command structure

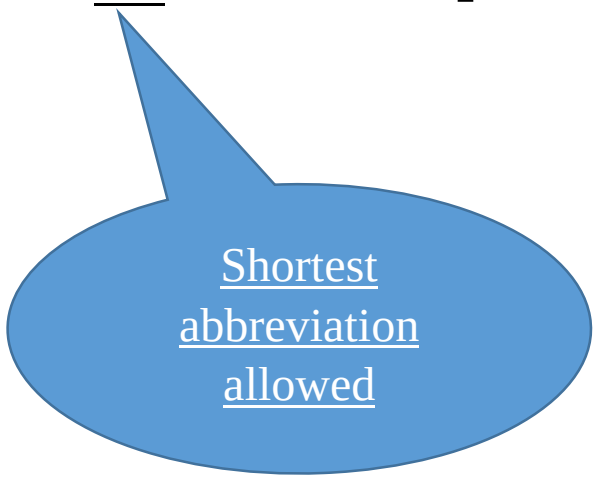
Examples

```
summarize [varlist] [if] [, options]
```

Understanding Stata's command structure

Examples

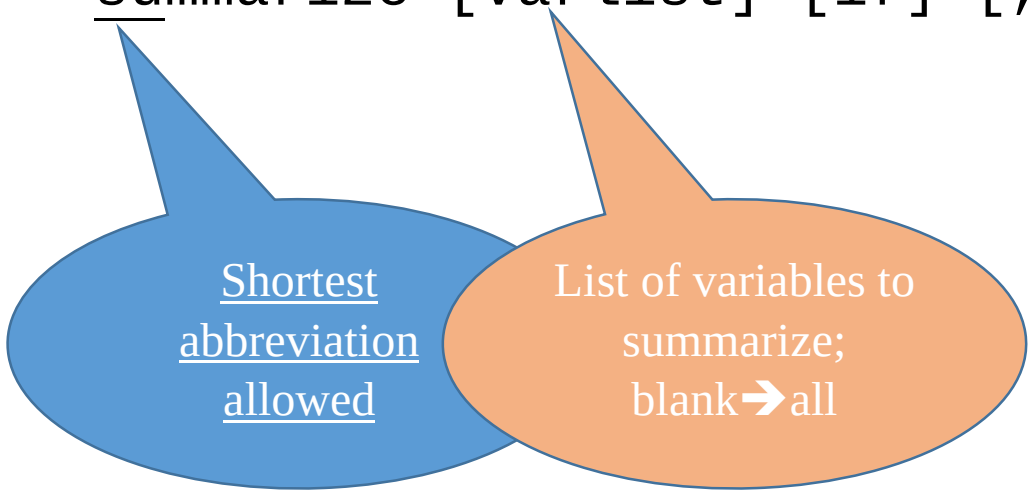
```
summarize [varlist] [if] [, options]
```



Understanding Stata's command structure

Examples

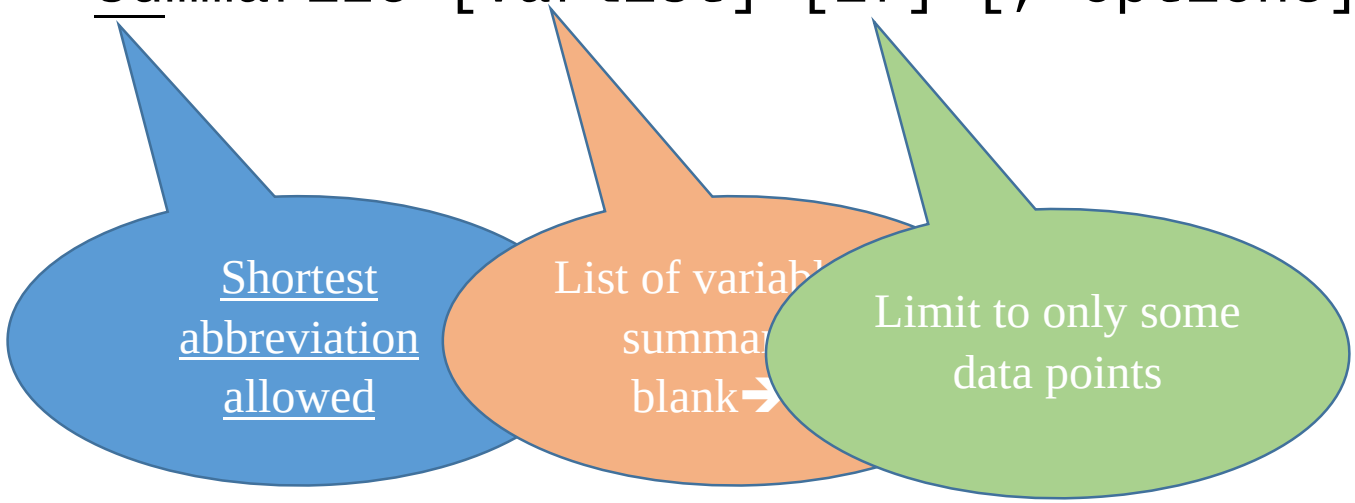
```
summarize [varlist] [if] [, options]
```



Understanding Stata's command structure

Examples

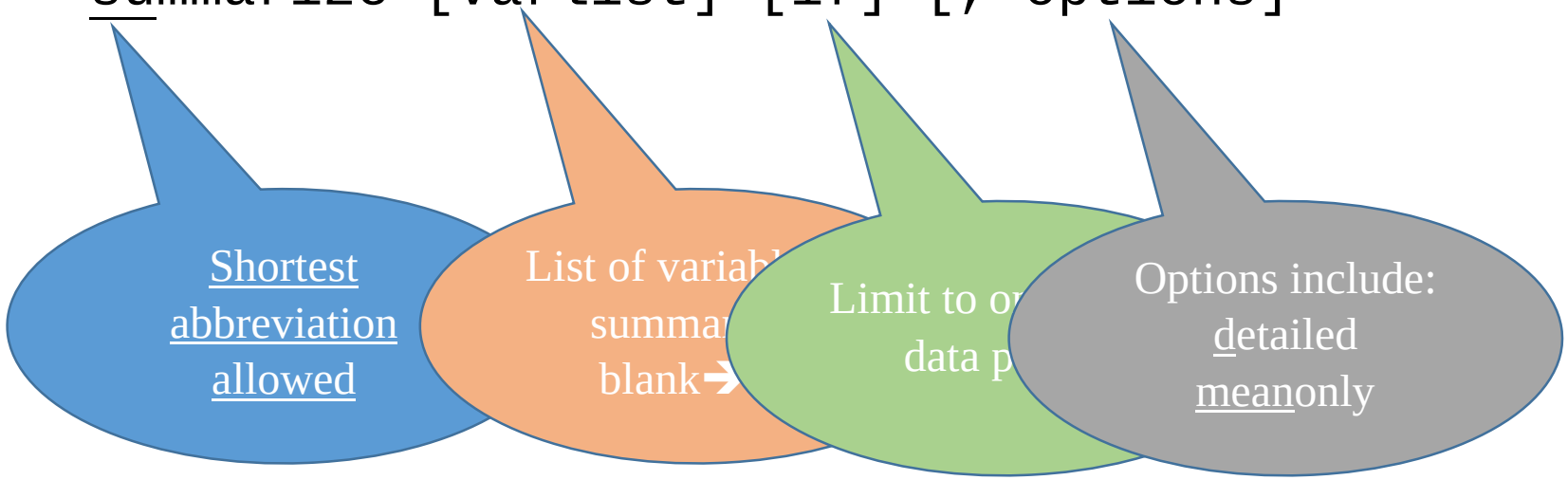
```
summarize [varlist] [if] [, options]
```



Understanding Stata's command structure

Examples

```
summarize [varlist] [if] [, options]
```

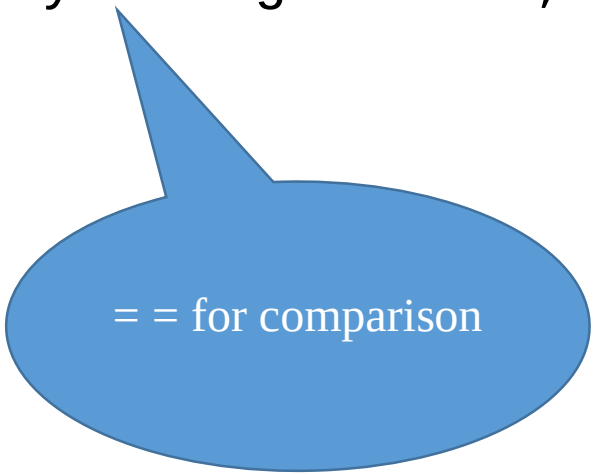


Understanding Stata's command structure

Examples

```
summarize [varlist] [if] [, options]

su GDP_real if country=="Argentina", detail
```



Understanding Stata's command structure

Examples

```
summarize [varlist] [if] [, options]
```

```
su GDP_real if country==”Argentina”, detail
```

Try it yourself!

- Summarize `year`
- Summarize `pop` for China
- Find median population in 2014

Understanding Stata's command structure

Digression on comparison

=	equal (assignment)
==	equal (comparison)
!=	not equal
~=	not equal
>	greater than
>=	greater than or equal to
<	less than
<=	less than or equal to
|	or
&	and

True (1) or False (0)

Understanding Stata's command structure

Digression on comparison: Examples

```
su GDP_real if year>=1990 & year<=1999
su GDP_real if country=="China" | country=="India"
gen coldwar=0
replace coldwar=1 if year<=1991
gen ColdWar=(year<=1991)
su coldwar ColdWar
```

Understanding Stata's command structure

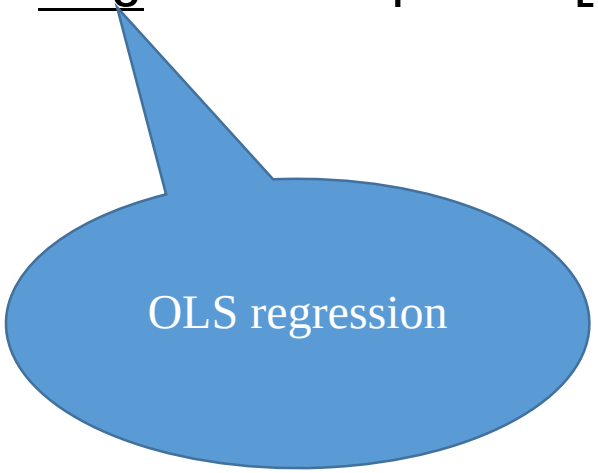
Examples

```
regress depvar [indepvars] [if] [, options]
```

Understanding Stata's command structure

Examples

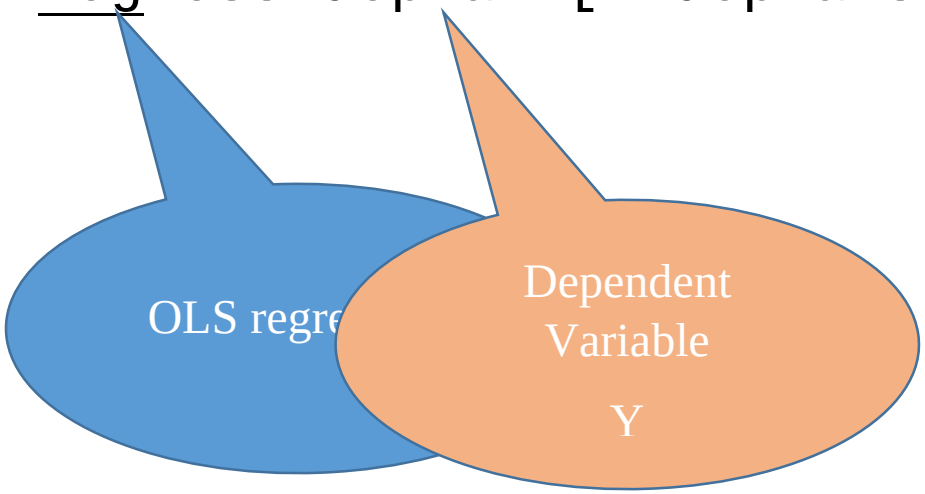
```
regress depvar [indepvars] [if] [, options]
```



Understanding Stata's command structure

Examples

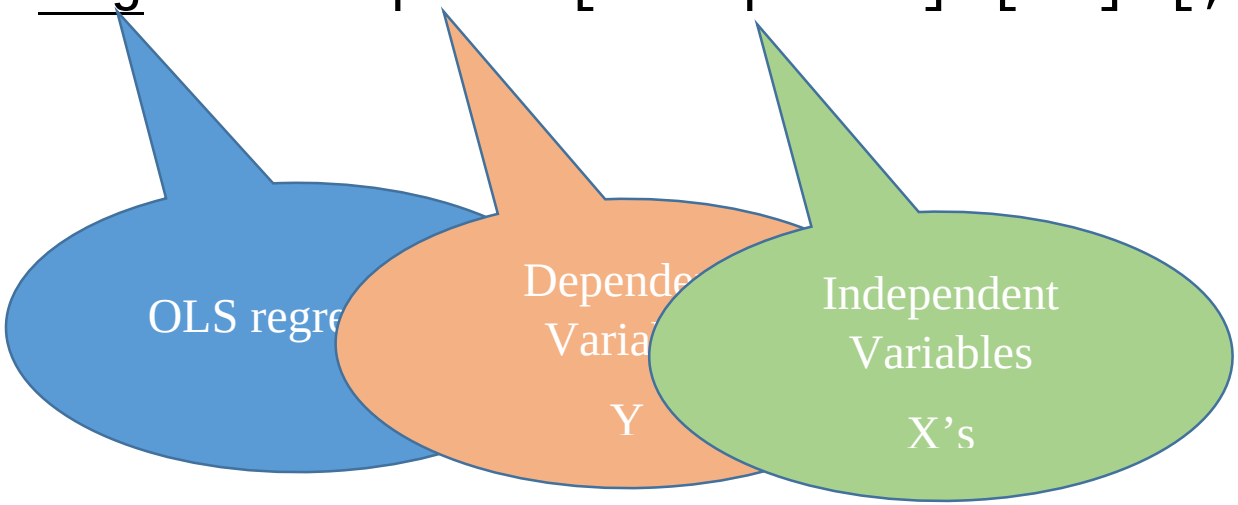
```
regress depvar [indepvars] [if] [, options]
```



Understanding Stata's command structure

Examples

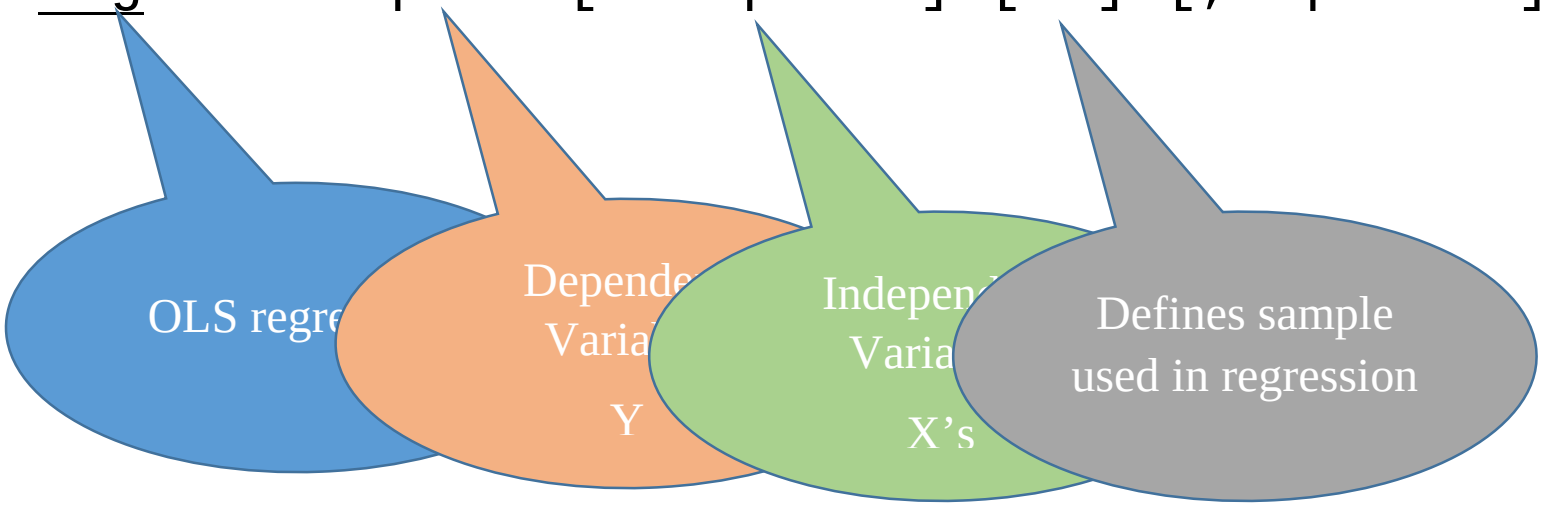
```
regress depvar [indepvars] [if] [, options]
```



Understanding Stata's command structure

Examples

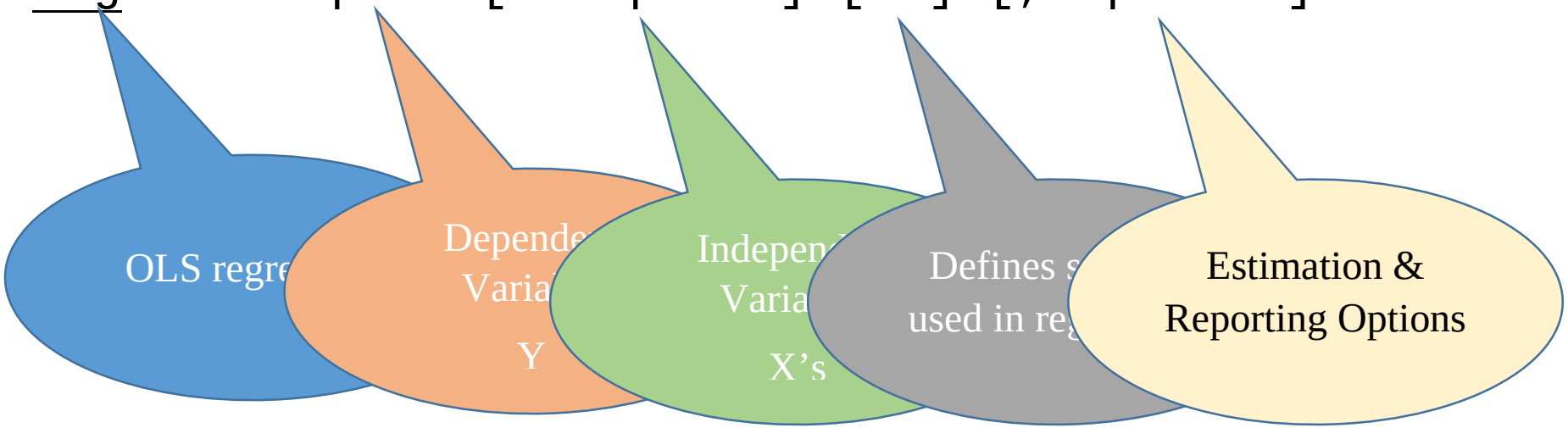
```
regress depvar [indepvars] [if] [, options]
```



Understanding Stata's command structure

Examples

```
regress depvar [indepvars] [if] [, options]
```



Understanding Stata's command structure

Examples

```
regress depvar [indepvars] [if] [, options]
```

```
gen lnUSaid=log(USA_TOFG_real)
gen lnPop=log(pop)
gen lnGDP=log(GDP_PPP)
reg lnUSaid lnPop lnGDP if year>2000, cluster(cid)
```

Frequently Used Stata Commands

`br`	browse data
`corr`	correlation between variables
`drop`	drop observations or variables
`gen`	generate new variable
`hist`	histogram
`keep`	keep obs or variables (drop all others)
`label var`	label variable
`missing`	identifies missing observations
`reg`	standard (OLS) regression
`rename`	rename variable
`replace`	replace values in existing variable
`sort`	sorts data
`su`	summarize variables
`twoway`	combines two-way graphs

Exercise

Create a graph that shows how US 2014 aid per person varied with country size.

- measure US aid as USA_TOFG_real (note: this is overall; you have to make it per capita by generating a new variable)
- put both measures in logs (log of US aid per person and log of population)
- put population on X-axis, US aid on Y-axis
- graph should include both scatter plot & best fit line
- figure out which three countries are at top left of graph
- save graph into Word file

  Hint: select “Graphics>Twoway graph” from menu or type “help twoway”

Stata Data types

Numbers:

- Generally can ignore what "type" of number (handled automatically)
- Integer & Real
- Different levels of precision
- Represented as digits
  - Examples: `234`
    `-14.5`
    `1,204` ✘ Comma not used in #s

Strings (Text):

  - Examples: `"JEAN"`
    `"Sally"`
    `Christopher` ✘ Must be in quotes

Stata Data types

Each Data type has different color when viewing data:

- Black: numbers
- Red: strings
- Blue: label for number

Missing Data:

- Numbers: .
  - Examples:

```
replace Age=. if Age<0
replace ReportedIncome=0 if Income>=.
replace ReportedIncome=0 if missing(Income)
```

- Strings: ""

```
gen NoName=(Name=="")
```

Stata Data types

Dates:

- Really numbers but formatted as dates
- Measure in time since Jan. 1, 1960
  - + numbers = after Jan. 1, 1960
  - − numbers = before Jan. 1, 1960
- Different units are possible
  - Years, Quarters, Months, Weeks, Days
  - Hours, Minutes, Seconds
- Formatting may carry over from data source (e.g., Excel) or you can set format:

```
format %td ApprovalDate
```

- Can tell Stata that a variable measures time in data:

```
tsset ApprovalDate
```

More Practice

Use a regression to see how U.S. foreign aid depends on recipient country need as measured by GDP per capita and Population.

- compare 1991 with 2014
- look at variables all in levels and then all in logs
- try including one or two other relevant variables