

CLEAN DATA

In Stata

Stata YouTube Channel



Project 1

Stata 14 Video Tutorials

[Link](#)

100 Best Stata Videos

[Link](#)



Stata Blog

Project 1

- NEW—Merges on Multiple Keys & Merges Gone Bad, [Link](#).
- Using import excel with real world data—Import Raw *GapMinder* Data
- Example, [Link](#).
- Data; [Link1](#), [Link2](#).

Advanced Topic—Extended Macro Function

- See **help extended_fcn**
- More Info, [Link](#).



Lab 3 & Project 1

Intro **Stata**

Learning Goals

- Gen new variables & use control structures
- Clean data sets
- Merge data sets
- Save as one dataset

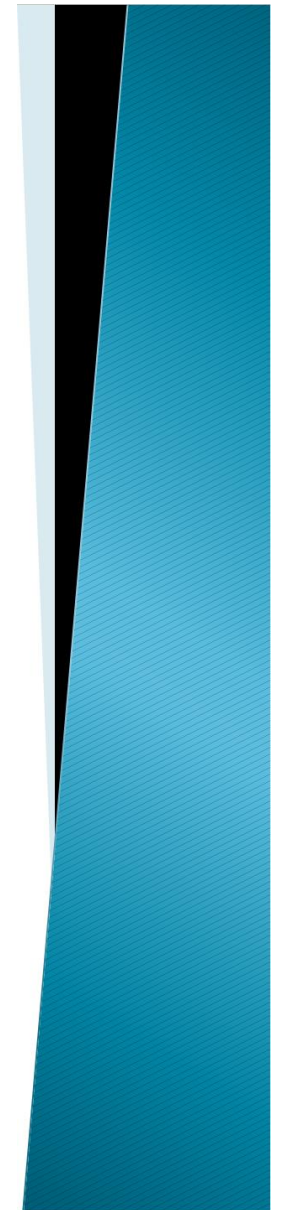
Shape Data

Command	Variables (Cols), Observations (Rows)
Collapse	Combine; Obs
Reshape	Combine & Separate; Current Vars & Obs
Merge	Combine; Datasets, Add New Vars & Obs
Append	Combine; Datasets, Add New Vars & Obs

Example

See Slides

References



Carryforward

- `carryforward` varlist [*if exp*] [*in range*], {gen(*newvarlist*) | replace}

Options

by sort :

Descriptions

Limits copying of values to within by groups
See extended macro functions

x	y
12	12
4	4
.	4
.	4
.	4
3	3
.	3
7	7
.	7
.	7

Example

[UCLA](#)

References

[Statalist](#)

[Replace](#) [_n-1]

Macros—Global

- `global gblname [=exp | :extended_fcn | [`]"[string]"[']]`

- Idea: Variable to store **global** text
- Accessible: Entire Stata session & all do-files, as ***\$gblname***
- Common usage: Program parameters

Example

References

[SSC 1](#)

[SSC 2](#)

[Princeton](#)

Macros—Local

- `local localname [=exp | :extended_fcn | [']"[string]"[']]`

- Idea: Variable to store **local** text
- Accessible: Only within current do-file, as '*localname*'
- Common usage: Loops—For & While

Example

References

[SSC 1](#)

[SSC 2](#)

[Princeton](#)

[Other](#)

While Loops

- **while** **exp** {
 ... *stata_commands* ...
}

exp

examples

while `i' < 40 {

... local i = `i' + 1 ...

while abs(`x_{new}' - `x_{old}') > 0.001 & `count' < 100 {

... *stata_commands* ...

}

Example

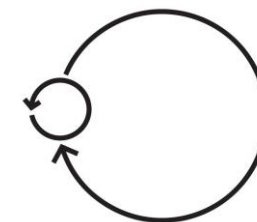
[SSC 2](#)

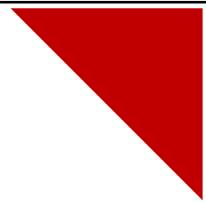
References

[SSC 1](#)

[Princeton](#)

[Other](#)





For Loops

- **foreach** *lname* in *list_type* any_list {
 ... *commands referring to 'lname'* ...
}

- **forvalues** *lname* = *range* {
 ... *commands referring to 'lname'* ...
}

list_types

local/global	Variable to store text list
varlist	list of variable names
numlist	list of numbers

range from steps

#1(#d)#2	#1 to #2	#d
#1/#2		1

Example

[SSC 2](#)

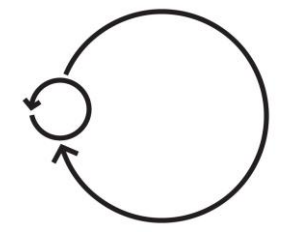
References

[Stata](#)

[SSC 1](#)

[UCLA](#)

[Princeton](#)



_Variables

- Built-in system variables that are created and updated by Stata

Common	Contains
_n	Number of the current observation
_N	Total number of observations.
_rc	Return code, most recent capture command

Example

[UCLA](#)

[Other](#)

References

[Stata _n](#)

[Stata _N](#)

_N, _n

- Additional Usage

Example	Source
Creating group Identifiers	Stata 1
Find first/last occurrence	Stata 2
Find duplicates in data	Stata 3
Lag/Difference data	Stata Daily

Example

[UCLA](#)

References

[Stata_n](#)

[Stata_N](#)

log file

- **log using filename** [, append replace [text|smcl] name(logname)]

Options

Descriptions

append | replace

Add or overwrite results to existing log file

text | smcl

Format in which the log is to be recorded

name(logname)

Optional name to refer to log file while open

Example

[UCLA](#)

References

[Stata](#)

do files

- Text files ending with **.do** extension, containing only Stata commands
- Stata has a built-in do file editor
- Create a new .do file by typing `doedit` into the command window
- Run .do file by typing `do myDoFile`, pressing `Ctrl-d`, or “play” button

Example

References

[SSC1](#)

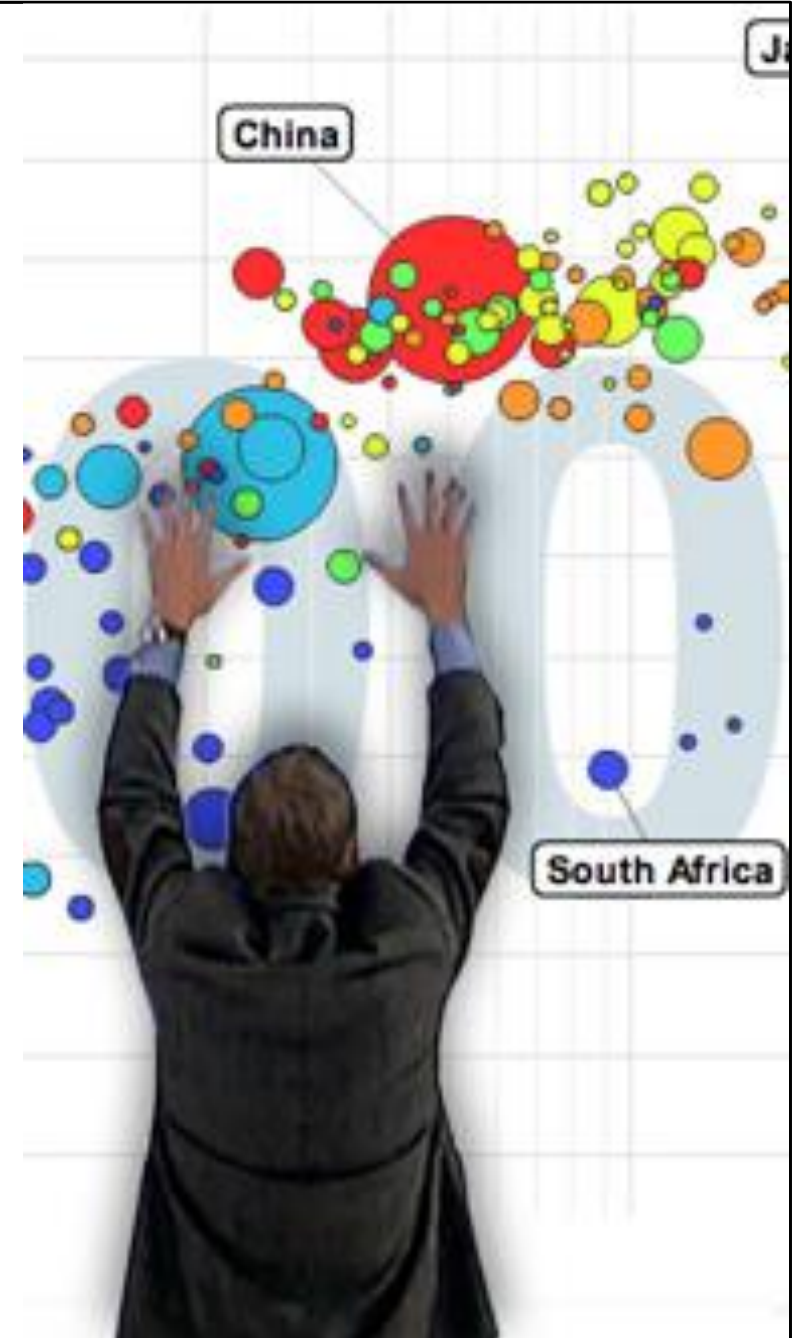
[SSC2](#)

Clean

Project 1

Do-file

- Use Imported Data
- Describe & Summarize
- Reshape long
- Carryforward, fill-in missing
- Truncate lifespan (<24)
- Ln gdp
- Save

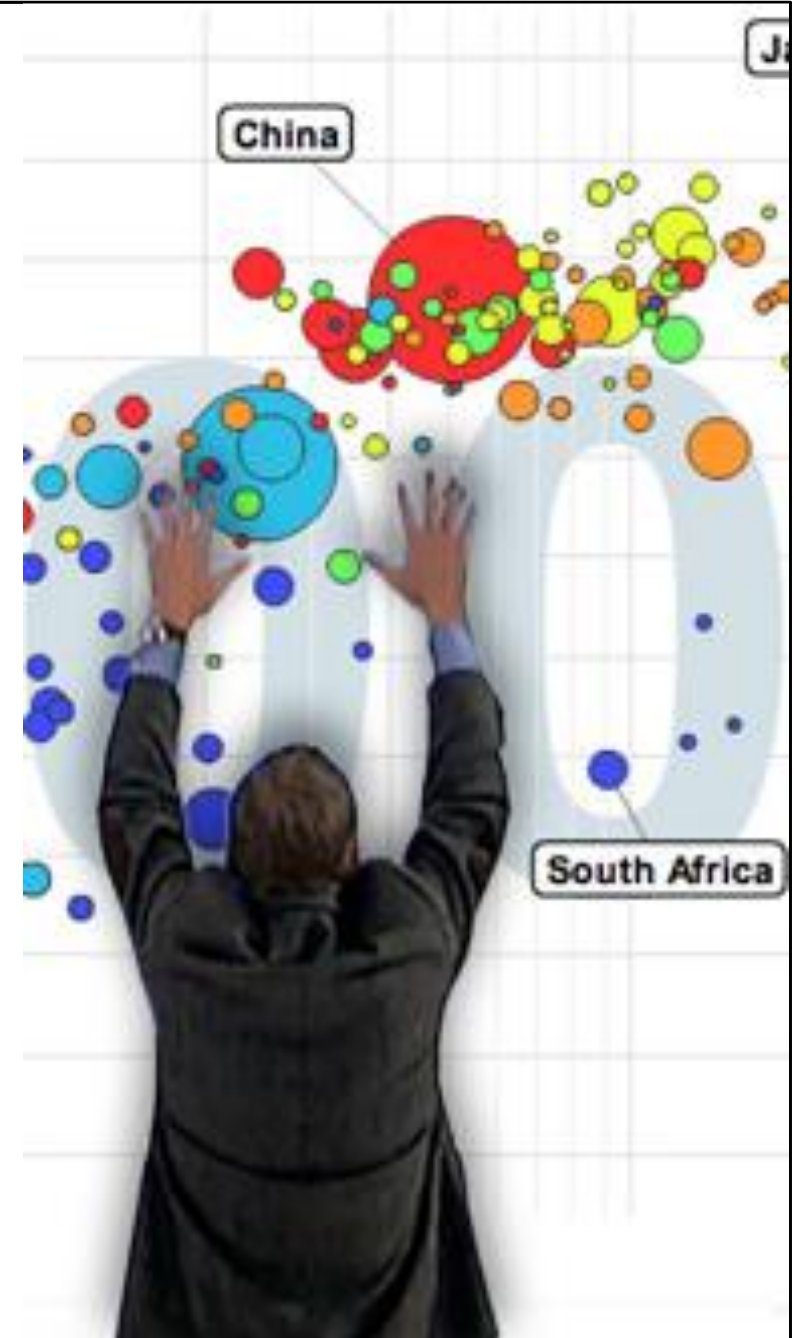


Merge

Project 1

Do-file

- Merge 1:1, m:1, 1:m
- Drop _merge
- Sort by cc & year
- Save Merged Data



Exercise I

- Complete your first .do file—use Project1 data—modify Header file

```
/*  
 * Project: GapMinder  
 * Author: Zdinak, Michael  
 * Email: zdinakmg@wustl.edu  
 * Date: 05 Jan 2017  
 * File: Main.do  
 * Depends:  
*/  
version 12  
clear all  
macro drop _all
```

Exercise II

```
cd "..."  
insheet using "...", ... Clear  
drop if missing( ... )  
reshape ... ..., i( ... ) j( ... )  
replace ...  
bysort ... : carryforward ...  
rename ...  
sort ...  
cd ../../input  
save ..., ...  
merge ... :1 ... ... using ...  
drop _merge
```

//
//
//
// log, truncate data
// fill-in missing

// save each
// merge saved