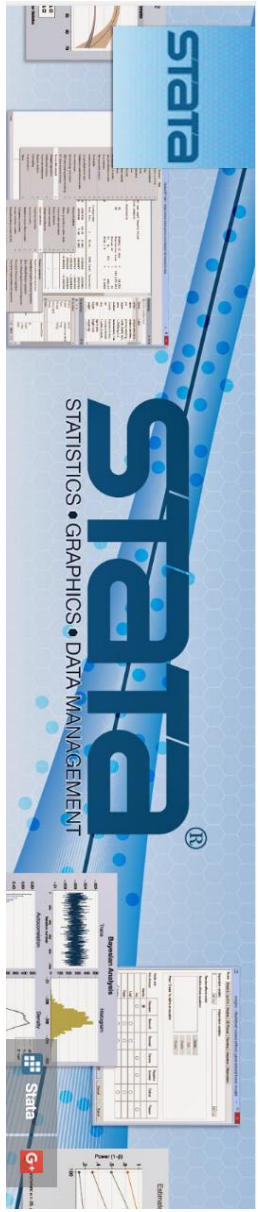
# GET DATA

Into Stata

# Stata YouTube Channel

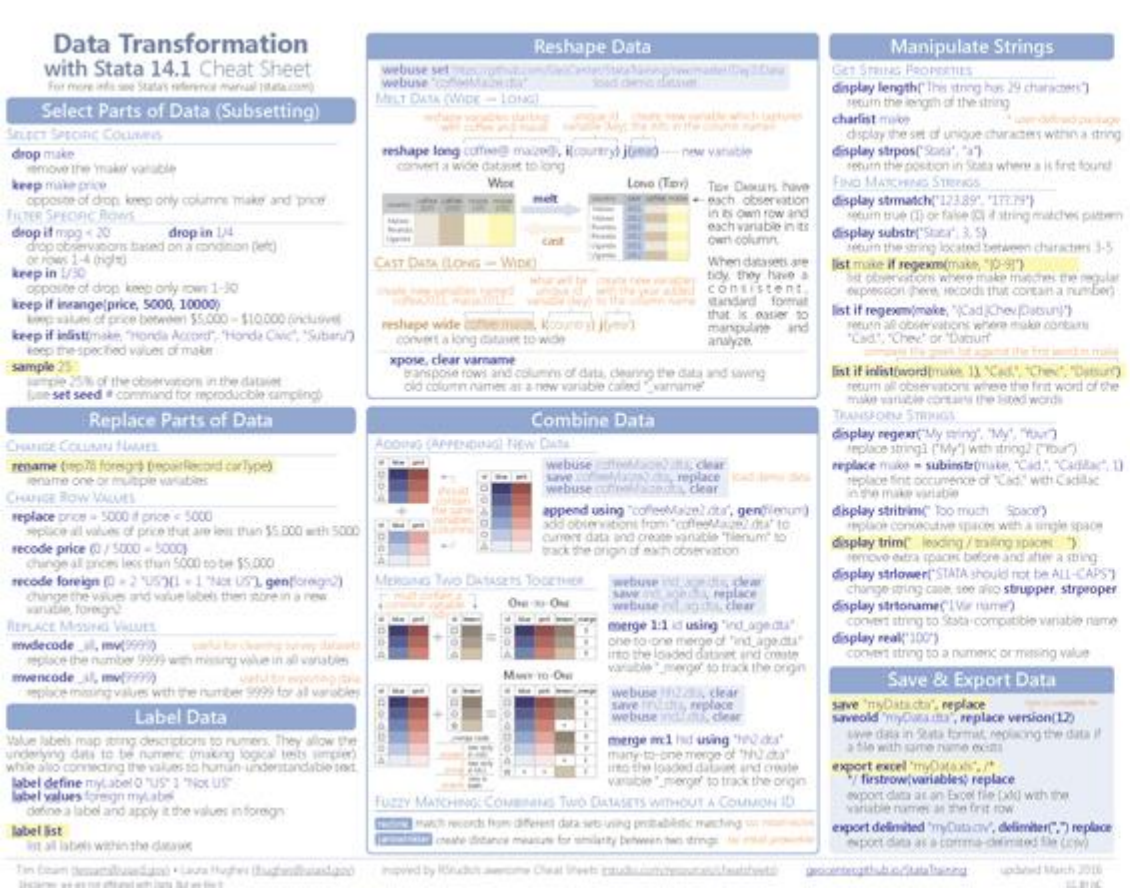
Project 1

- Tour of Stata Interface
- Copy/paste data from Excel into Stata
- Import Excel data into Stata
- Importing delimited data in Stata



# Cheat Sheets

Link




# Lab 2 & Project 1

Intro Stata

Learning Goals

- Types of Variables & Control Structures
- Import data
- Save dataset
- Create new variables
- Use basic operators

# The 5 Basic Concepts

Found in all Programming Languages, Link.

1. Variables
2. Control Structures
3. Data Structures
4. Syntax
5. Tools

# Variables

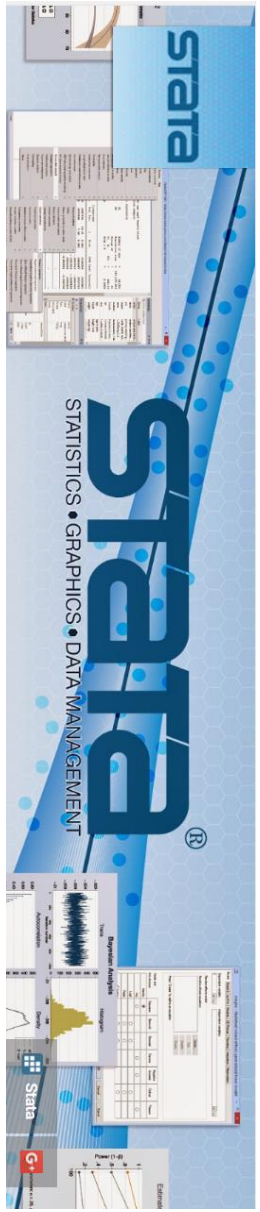
*"Any way to store & retrieve information at will"*

Stata **Datatypes** ( small -> large ), see Link

- Numeric: byte, int, long, float, double
- String: str1, str2, … , str2045, strL

Does Var Size Matter?

- Increasing dataset size may slow Stata down
- Converting types incorrectly will truncate data



# Control Structures

**Code Flow**: Left-to-Right, Top-to-Bottom

**Decisions**: Repeat, Skip, or Jump To
Will often be conditional on Variable values

In Stata…

- If { … } Else { … } Statements, Link
- Foreach *x* in '*varlist*' { … } Loops, Link
- While *exp* { … } Loops, Link
- Function(…) Calls, Link

# The 5 Basic Concepts

Found in all Programming Languages, Link.

1. Variables
2. Control Structures
3. Data Structures
4. Syntax
5. Tools

# help summarize

**Title**

[R] summarize — Summary statistics

**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 and base and empty cells |

# Stata Syntax

Link

**summarize** [varlist] [if *exp*] [in r*ange*] [weight] [, *options*]

Example—`sysuse auto`

- `summarize mpg price`
- `summarize in 1/10`
- `summarize if foreign == 1`
- `summarize if foreign == 1 & mpg > 30`
- `summarize , detail`

# Stata Shortcuts

Link

**Prefix Commands:** Repeat Stata command on subsets of the data

[*by varlist*:] summarize [varlist] [if *exp*] [in r*ange*] [weight] [, *options*]

Example—`sysuse auto`

- `by foreign:  summarize rep78`
- `sort foreign make`
- `by foreign (make): summarize rep78`
- `by foreign rep78, sort: summarize price`

| cc | year | life |
|---|---|---|
| Afghanistan | 1800 | 28.21 |
| Afghanistan | 1801 | 28.2 |
| Afghanistan | 1802 | 28.19 |
| Afghanistan | 1803 | 28.18 |
| Afghanistan | 1804 | 28.17 |
| Afghanistan | 1805 | 28.16 |
| Afghanistan | 1806 | 28.15 |
| Afghanistan | 1807 | 28.14 |
| Afghanistan | 1808 | 28.13 |
| Afghanistan | 1809 | 28.12 |
| Afghanistan | 1810 | 28.11 |
| Afghanistan | 1811 | 28.1 |
| Afghanistan | 1812 | 28.09 |
| Afghanistan | 1813 | 28.08 |
| Afghanistan | 1814 | 28.07 |
| Afghanistan | 1815 | 28.06 |
| Afghanistan | 1816 | 28.05 |
| Afghanistan | 1817 | 28.04 |
| Afghanistan | 1818 | 28.03 |

Excel files

# import excel

Import & export Excel files

- **import excel** [using] **filename** [**,** *import_excel_options*]

| **Options** | **Descriptions** |
|---|---|
| sheet("Sheetname") | worksheet to load |
| cellrange([start][:end]) | cell range to load |
| firstrow | treat first row of Excel data as variable names |
| allstring | import all Excel data as strings |
| clear | replace data in memory |

**Example**

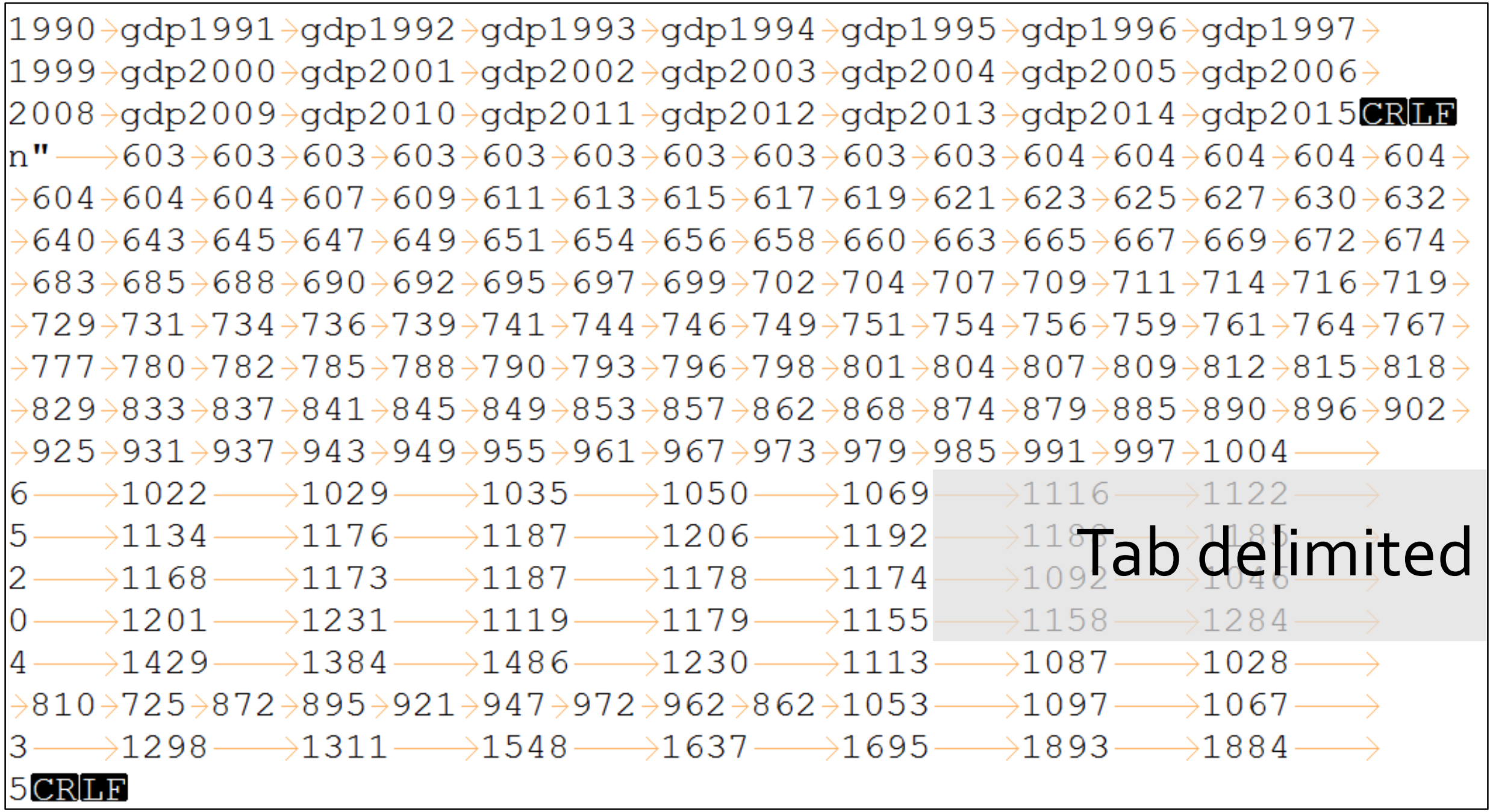
UCLA

**References**

Princeton UNC Stata

p2003,tp2004,tp2005,tp2006,tp2007,tp2008,tp2009,tp2010,tp2011,tp2012,
14,tp2015CRLF
"Afghanistan",3280000,3280000,3323519,3448982,3625022,3810047,3973968
19695,4710171,5021241,5351413,5813814,6394908,7034081,7752118,7839426
38312,8150037,8270024,8398309,8534913,8679848,8833127,8994793,9164945
31555,9728645,9935358,10148841,10368600,10599790,10849510,11121097,11
6896,12022514,12315553,12582954,12831361,13056499,13222547,13283279,1
96923,12667001,12279095,11912510,11630498,11438949,11337932,11375768,
067570,12789374,13745630,14824371,15869967,16772522,17481800,18034130
9038420,19701940,20531160,21487079,22507368,23499850,24399948,25183611
26528741,27207291,27962207,28809167,29726803,30682500,31627506,325265
"Albania",410445,423591,438671,457234,478227,506889,552800,610036,672
819950,901122,963956,1015991,1123210,1263171,1287499,1316086,1348097,
9969,1459089,1500152,1543224,1588478,1636054,1685001,1735646,1786530,
6125,1947786,1998695,2049147,2099657,2150602,2202040,2253842,2305999,
1229,2464338,2517869,2571845,2626290,2681245,2735329,2788315,2842620,
6799,3041003,3121336,3197064,3253659,3281453,3275438,3240613,3189623,
6727,3092034,3092471,3102898,3114851,3121965,3124093,3123112,3117045,
2172,3050741,3010849,2968026,2929886,2901883,2886010,2880667,2883281,
6679CRLF

Comma delimited




Tab delimited

# Insheet

Read data created by a spreadsheet

- **insheet** [varlist] **using filename** [**, options**]

| **Options** | **Descriptions** |
|---|---|
| tab | tab-delimited data, .txt |
| comma | comma-delimited data, .csv |
| delimiter("char") | use char as delimiter |
| clear | replace data in memory |
| [*no*]names | variable names included on first line of the file |

**Examples**

UCLA UCLA

**References**

Princeton Stata

```
"Benin"                          "Sub-Saharan Africa"
"BJ"                    "Africa"
"Botswana"                       "Sub-Saharan Africa"
"BW"                    "Africa"
"Burkina Faso"                   "Sub-Saharan Africa"
"BF"                    "Africa"
"Burundi"                        "Sub-Saharan Africa"
"BI"                    "Africa"
"Cameroon"                       "Sub-Saharan Africa"
"CM"                    "Africa"
"Cape Verde"                     "Sub-Saharan Africa"
"CV"                    "Africa"
"Central African Rep."           "Sub-Saharan Africa"
"CF"                    "Africa"
"Chad"                           "Sub-Saharan Africa"
"TD"                    "Africa"
"Comoros"                        "Sub-Saharan Africa"
"KM"                    "Africa"
"Congo, Dem. Rep."               "Sub-Saharan Africa"
"CD"                    "Africa"
"Congo, Rep."                    "Sub-Saharan Africa"
"CG"                    "Africa"
"Cote d'Ivoire"                  "Sub-Saharan Africa"
"CI"                    "Africa"
"Equatorial Guinea"              "Sub-Saharan Africa"
"GQ"                    "Africa"
"Eritrea"                        "Sub-Saharan Africa"
"ER"                    "Africa"
"Eritrea and Ethiopia"           "Sub-Saharan Africa"
"Eritrea_Ethiopia"      "Africa"
"Ethiopia"                       "Sub-Saharan Africa"
"ET"                    "Africa"
"Gabon"                          "Sub-Saharan Africa"
"GA"                    "Africa"
"Gambia"                         "Sub-Saharan Africa"
"GM"                    "Africa"
```

# Infile

Read unformatted text data

- **infile varlist** [*varlist* [_*skip*[(#)] ...]]] **using filename** [if] [in] [**,** *options*]

| Options | Descriptions |
|---|---|
| clear | replace data in memory |
| automatic | create value labels from nonnumeric data |

**Example**

UCLA

**References**

Princeton Stata

# The 5 Basic Concepts

Found in all Programming Languages, Link.

1. Variables
2. Control Structures
3. Data Structures
4. Syntax
5. Tools

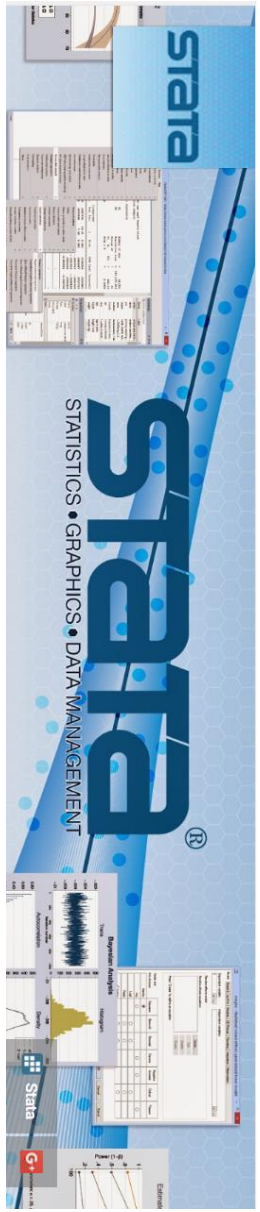
# Tools

*Anything that helps gets your job done faster!*

IDE—Limited options when using Stata, Link.

- Integrated Development Environment
- Check syntax, organize & view files, Code completion & navigation

User-written procedure files—see findit carryforward



# Findit

Uncertain source—**findit** estout

- Search Stata docs for user package, select, install.
- Not always trivial to find correct package ☹
- See Link or Link.

Known source—ssc install estout

List of UCLA developed packages, Link.

# Carryfoward

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

| Options | Descriptions |
|---|---|
| by sort : | Used with extended macro function<br>Limits flow of values to within by groups |

| x | y |
|---|---|
| 12 | 12 |
| 4 | 4 |
| . | 4 |
| . | 4 |
| . | 4 |
| 3 | 3 |
| . | 3 |
| 7 | 7 |
| . | 7 |
| . | 7 |

**Example**

UCLA

**References**

Statalist Replace [ _n-1]

# Stata Blog

Project 1

- 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.