

---

# **VIDEO: INTRODUCTION TO STATA**

EMBA Data Analysis

Professor Timothy Simcoe

Boston University School of Management

# Agenda

---

Loading data

Stata's windows

Creating a variable

Advanced stuff

Learning resources

# Loading Data

---

Download “Brookline Home Prices” data set

- <http://people.bu.edu/tsimcoe/teaching.html>

Two ways to import data

- Excel Import Tool: “File > Import > Excel...”
- As Stata (.dta) data set: “File > Open...”
- Type “clear” in Command box before loading new data

Stata always stores data as single “rectangular” worksheet

- Row = Observation, Column = Variable

# Stata's Windows

Menu Bar: Single location to access all features of program

Results Window: Where Stata displays output after you tell it to do something

The screenshot shows the Stata 12.0 interface with several windows open. The menu bar at the top includes File, Edit, View, Data, Graphics, Statistics, User, Window, and Help. The Review window on the left lists commands: 1 set mi... and 2 import... The Results window in the center displays the Stata logo and version information (12.0), copyright details (1985-2011 StataCorp LP), contact information, a single-user perpetual license for Timothy Simcoe at Boston University, and a list of notes including a memory setting change and an Excel import command. The Variables window on the right shows a table of variables with their names and labels. The Command window at the bottom is empty.

| Name        | Label         |
|-------------|---------------|
| City        | City          |
| State       | State         |
| StoreID     | Store ID      |
| FiscalYear  | Fiscal Year   |
| Sales       | Sales         |
| CostofSales | Cost of Sales |
| Region      | Region        |

| Variables   |  |
|-------------|--|
| Name        |  |
| Label       |  |
| Type        |  |
| Format      |  |
| Value Label |  |
| Notes       |  |

| Data         |       |
|--------------|-------|
| Filename     |       |
| Label        |       |
| Notes        |       |
| Variables    | 7     |
| Observations | 99    |
| Size         | 3.67K |
| Memory       | 96M   |

Variables window: Useful info about the variables in current data set

Review window: List of previous Commands, click for a "do over"

Command window: Where you type in commands (unless you prefer to use the drop down menus)

# Stata also has a “spreadsheet” view

Data Editor allows editing (not recommended)

Data Browser only permits viewing data

Click here to view data in “spreadsheet” format

Note that **string** variables are displayed in **RED**, & **numeric** variables are displayed in **BLACK**

The screenshot shows the Stata 12.0 interface. The main window is titled "Data Editor (Browse)" and displays a spreadsheet view of data. The spreadsheet has columns for City, State, StoreID, FiscalYear, Sales, CostofSales, and Region. The City column is highlighted in yellow. The data is displayed in a grid format with string variables in red and numeric variables in black. The spreadsheet shows 19 rows of data for various cities in Massachusetts and Connecticut. The Stata command window on the left shows the following commands:

```
1 set mi...
2 import...

Single-user
Serial
Lic...

Notes:
. set min_mer
(set min_mer...

. import exc...
> sheet("She...

Command
```

The status bar at the bottom indicates "Vars: 7" and "Obs: 99".

|    | City      | State | StoreID | FiscalYear | Sales     | CostofSales | Region |
|----|-----------|-------|---------|------------|-----------|-------------|--------|
| 1  | Boston    | MA    | 1       | 2001       | 250000    | 226624.75   | NE     |
| 2  | Boston    | MA    | 1       | 2002       | 249534.87 | 209588.05   | NE     |
| 3  | Boston    | MA    | 1       | 2003       | 260841.68 | 231461.58   | NE     |
| 4  | Boston    | MA    | 2       | 2001       | 300000    | 275673.04   | NE     |
| 5  | Boston    | MA    | 2       | 2002       | 315063.55 | 266654.66   | NE     |
| 6  | Boston    | MA    | 2       | 2003       | 355224.74 | 285674.59   | NE     |
| 7  | Boston    | MA    | 3       | 2001       | 190000    | 157543.44   | NE     |
| 8  | Boston    | MA    | 3       | 2002       | 186397.18 | 162927.55   | NE     |
| 9  | Boston    | MA    | 3       | 2003       | 204596.22 | 181135.98   | NE     |
| 10 | Worcester | MA    | 1       | 2001       | 100000    | 82490.636   | NE     |
| 11 | Worcester | MA    | 1       | 2002       | 109960.48 | 92311.208   | NE     |
| 12 | Worcester | MA    | 1       | 2003       | 109217.88 | 97099.96    | NE     |
| 13 | Hartford  | CT    | 1       | 2001       | 180000    | 168250.27   | NE     |
| 14 | Hartford  | CT    | 1       | 2002       | 207275.08 | 158811.19   | NE     |
| 15 | Hartford  | CT    | 1       | 2003       | 226873.53 | 174659.56   | NE     |
| 16 | Hartford  | CT    | 2       | 2001       | 150000    | 137601.41   | NE     |
| 17 | Hartford  | CT    | 2       | 2002       | 162738.83 | 133662.67   | NE     |
| 18 | Hartford  | CT    | 2       | 2003       | 187883.1  | 175823.13   | NE     |
| 19 | Hartford  | CT    | 3       | 2001       | 90000     | 72010.62    | NE     |

# Creating a new variable

---

**Goal: Make new variable that measures assessed value per square foot**

## Using Menu Bar

- Select “Data > Create or change data > Create new variable”
- Using dialog box, provide a name (“**dpsf**”)... and expression (“**value/area**”)
- *No spaces allowed in variable names*
- Can remove variables by typing “drop [variable name]” on the Command line

## Using Stata’s “generate” command

- Type “generate dpsf = value / area” on Command line

## Three ways to reduce typing (and typos)

- Click next to variable name to paste in Command line
- Click in Review Window to paste previous commands
- Renaming variables, e.g. “rename dpsf warm-and-fuzzy”

# Advanced stuff

---

Creating tabular summaries of the data

- “table nbhd, contents(p25 value p75 value)”

Visualizing the data

- “histogram dpsf”

Modeling relationships in the data

- “regress value rooms”
- “twoway (scatter value rooms) (lfit value rooms)”

# Additional Resources

---

## The Instructor

- I am available to answer Stata questions any time
- I prefer to take these questions by email

## Help Manuals

- Type “help” or “help [command name]” in Stata to see the manuals

## Internet

- Google “Stata Help” for a wealth of information

## SMG Tools

- I will post a Tutorial, Cheat Sheet and Tips to a “Stata Help” folder

## Reference Books

- “A Gentle Introduction to Stata”