Mostly Dates…
and a few other useful STATA commands

Jen Cocohoba, Pharm.D., MAS
Associate Clinical Professor
UCSF School of Pharmacy

In this portion of the lecture

- How to manipulate dates in STATA
- Performing loops
- Basic “how to” merge datasets

- “Follow along”
  - No lab exercises
  - Sample Excel spreadsheet “practicedates” on syllabus

Dealing with Dates in STATA

- Dates in your research

- STATA can help you manipulate dates
  - Add, subtract, calculate time between dates
  - Comparing dates (e.g. before 1999, after 1999)
  - Extracts components of dates (year, day of week)

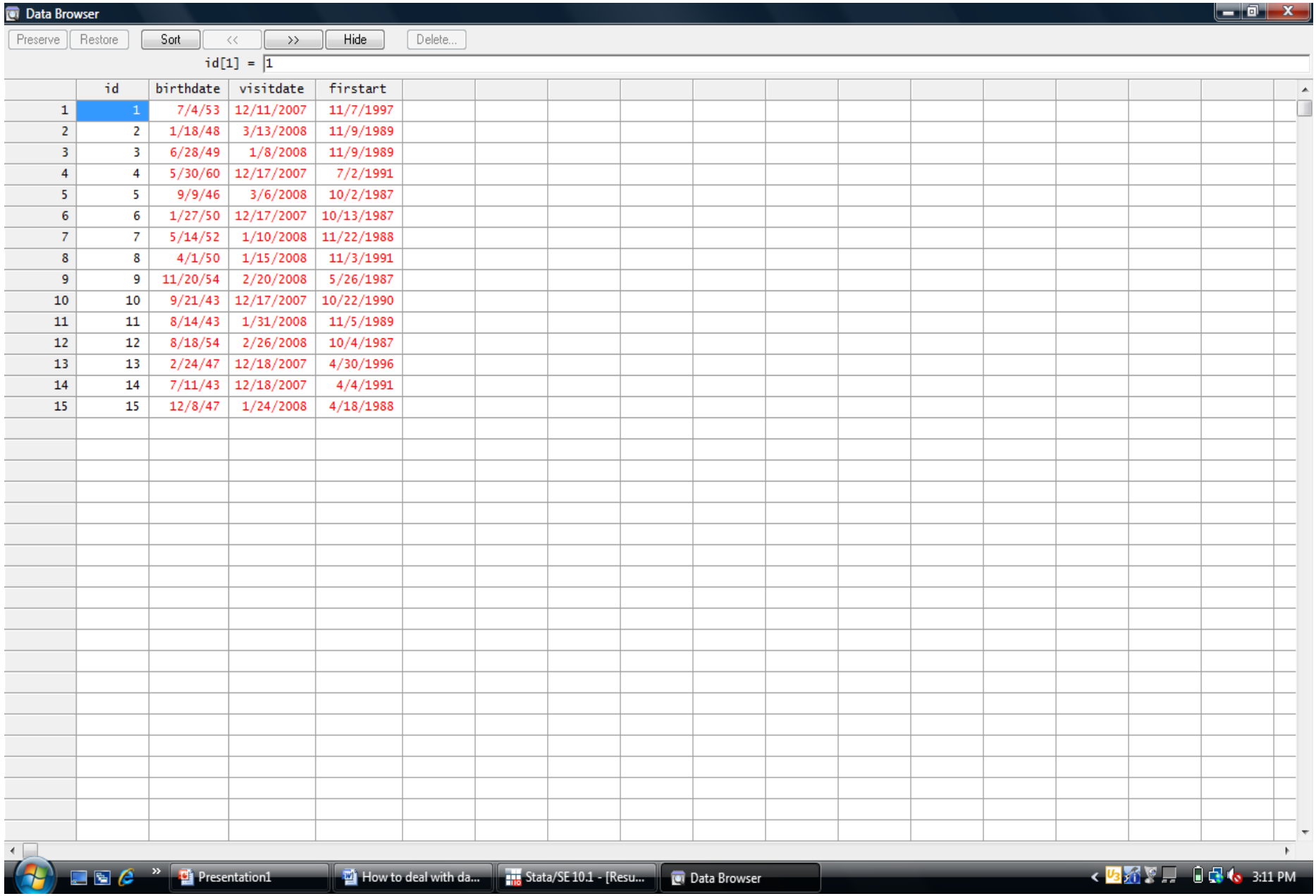
How STATA thinks about dates

- Sees them as a number
- "Counts" date as the # of days from a specific reference
  - January 1, 1960 = 0
  - January 2, 1960 = 1
  - January 3, 1960 = 2
  - December 31, 1960 = 364
- This makes it "easy" for STATA to manipulate them mathematically
- We will come back to this in formatting dates

Cleaning strings to STATA dates

- Most often imported from a program (Excel)

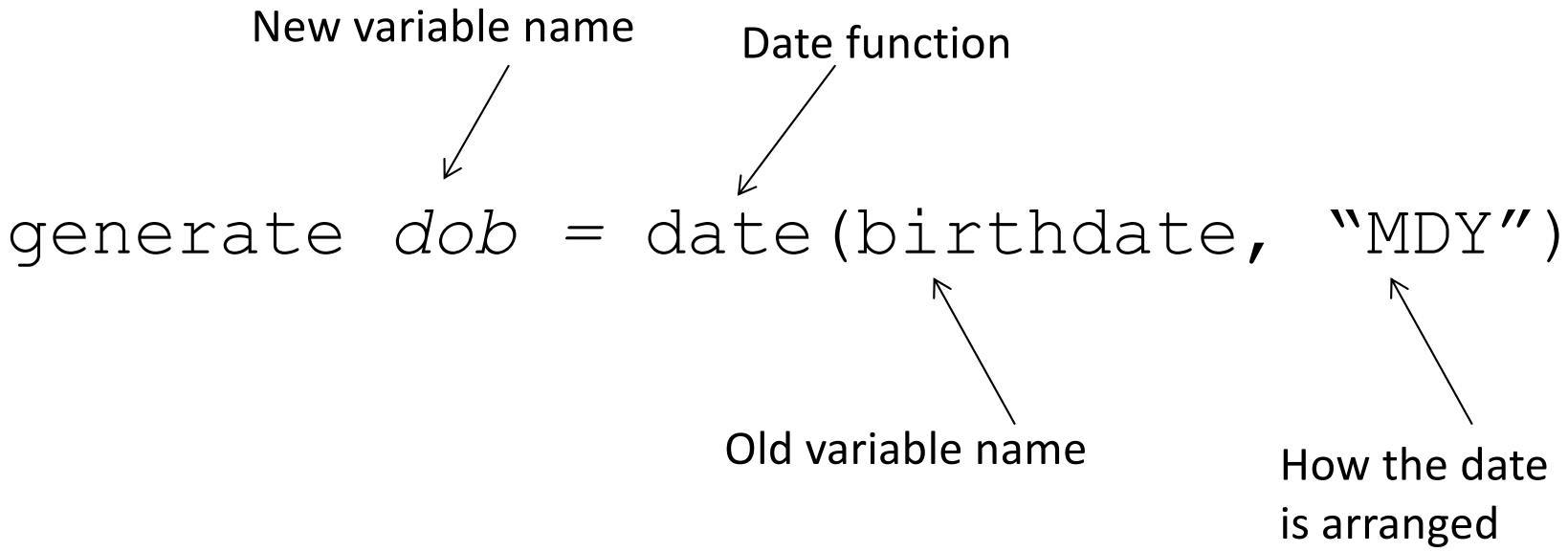
- You can follow along
  - Practicedates.xls
  - Open file
  - Copy dates
  - Paste into STATA data editor
  - Look at dates …they are red!


Data Browser
Preserve
Restore
Sort
<<
>>
Hide
Delete...
id[1] = 1
id birthdate visitdate firstart
1 1 7/4/53 12/11/2007 11/7/1997
2 2 1/18/48 3/13/2008 11/9/1989
3 3 6/28/49 1/8/2008 11/9/1989
4 4 5/30/60 12/17/2007 7/2/1991
5 5 9/9/46 3/6/2008 10/2/1987
6 6 1/27/50 12/17/2007 10/13/1987
7 7 5/14/52 1/10/2008 11/22/1988
8 8 4/1/50 1/15/2008 11/3/1991
9 9 11/20/54 2/20/2008 5/26/1987
10 10 9/21/43 12/17/2007 10/22/1990
11 11 8/14/43 1/31/2008 11/5/1989
12 12 8/18/54 2/26/2008 10/4/1987
13 13 2/24/47 12/18/2007 4/30/1996
14 14 7/11/43 12/18/2007 4/4/1991
15 15 12/8/47 1/24/2008 4/18/1988
Presentation1
How to deal with da...
Stata/SE 10.1 - [Resu...
Data Browser
3:11 PM

Cleaning STATA dates

- STATA sees dates from Excel as text, not #
- Even if you type dates directly into data editor, still seen as text

- Date conversion
  - Generate a new date variable using the date function
  - Tell it which old variable contains the date you want to convert
  - Give it a format (most common is month, day, year)
  - Try it and look at results


New variable name
Date function
generate dob = date(birthdate, "MDY")
Old variable name
How the date is arranged

Number nonsense

dob
-2372
-4366
-3839
150
-4862
-3626
-2788
-3562
-1868
-5946
-5984
-1962
-4694
-6018
-4407
0

- Can format (mask) the numerical date so that it is easier for you to understand

Command:

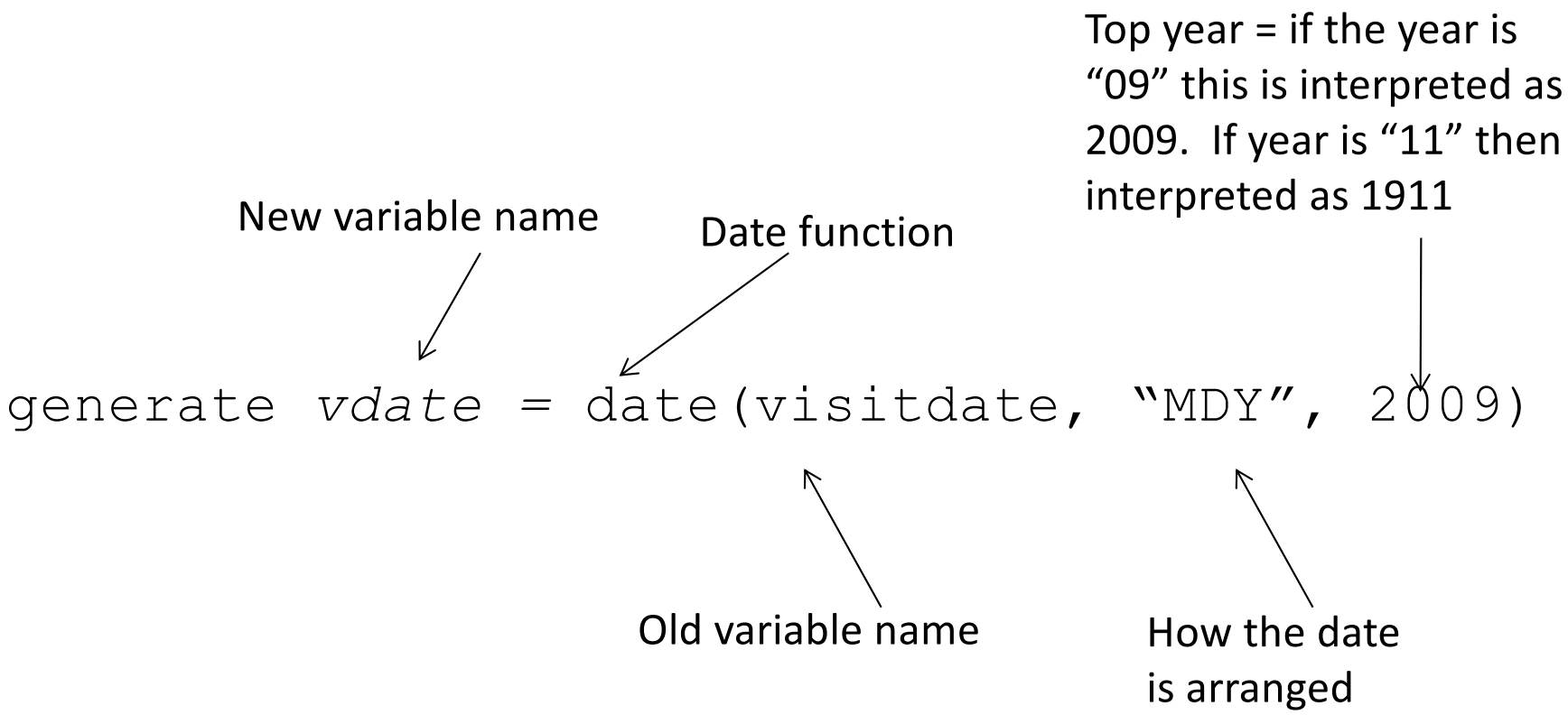
`format dob %td`

A series of two commands

- Most will be like this 2-command example
  - `generate dob = date(birthdate, "MDY")`
  - `format dob %td`

- Change the first ART date to a STATA interpretable date

- STATA has issues with dates with 2 digit years
  - Try converting visitdate (2 digit years)
  - Should get a "missing values" generated
  - Need to add a "topyear" which is the cutoff value. STATA will interpret years up to this year.


Top year = if the year is "09" this is interpreted as 2009. If year is "11" then interpreted as 1911
New variable name
Date function
generate vdate = date(visitdate, "MDY", 2009)
Old variable name
How the date is arranged

Date formatted – what can you do with it?

- Extract components of the date into new variables (columns)
  - gen *nameofdayvariable* = day(*datevariable*)
  - gen *weekdayvariable* = dow(*datevariable*)
    - Lists as 0(Sunday) - 6(Saturday)
  - gen *monthvariable* = month(*datevariable*)
  - gen *yearvariable* = year(*datevariable*)

What can you do with dates

- Find time between dates

- Suppose you wanted to find participants' age at the date of their study visit.
  - Generate new variable called ageatvisit

```
gen ageatvisit = vdate - dob
```

  - Note this gives you their age in number of DAYS
  - Can do this more efficiently by

```
gen ageatvisit =(vdate - dob)/365.25
*gen agevisityears = int(ageatvisit)*
```

Comparing dates

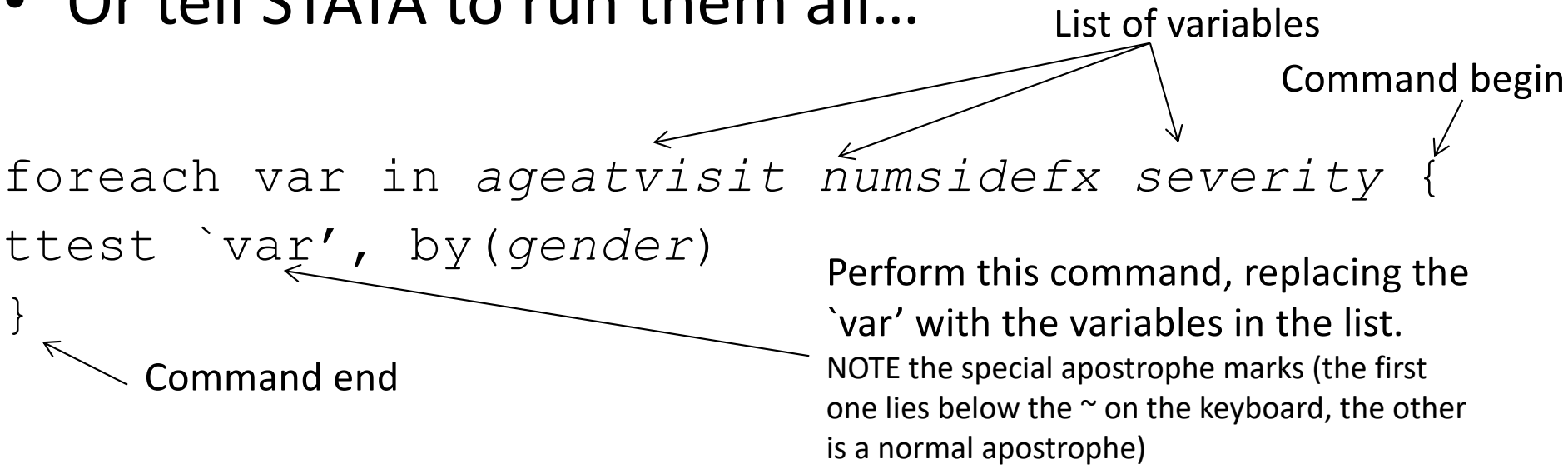
- Suppose you wanted to categorize patients by their visit dates
  - Those who had a visit before 12/31/07 = earlyvisit
- Using literal dates
  - Formatted as day month year (01jan1960)
  - Must be denoted by parenthesis and letter d
  - Example: d(01jan1960)
- Example
  - `gen earlyvisit = 0`
  - `replace earlyvisit = 1 if vdate <= d(31dec2007)`
  - `replace earlyvisit =. if vdate==.`

Programming loops

- Same command to a bunch of variables
- Example
  - Test whether age at visit, number of side effects, and average severity of side effects differ by gender (sex)
- Could do this…
  - `ttest` *`ageatvisit`*`, by(`*`sex`*`)`
  - `ttest` *`numsidefx`*`, by(`*`sex`*`)`
  - `ttest` *`severity`*`, by(`*`sex`*`)`

Loop Syntax

- Or tell STATA to run them all…




```
foreach var in ageatvisit numsidefx severity {
summarize`var', detail
}
```

Merging datasets, simplest example

- Merge versus append
  - Merge = add new variables from 2nd dataset to existing observations
  - Append = add new observations to existing variables

- Merging requires datasets to have a common variable (ID)

- Nomenclature for two datasets
  - One dataset is defined as the "master" (in memory) dataset
  - The other dataset is called the "using" dataset

- Many merge types
  - One to one: master file w/demographics, using data has labs (`merge 1:1`)
  - One to many: master file w/demographics, using file with multiple visits (`merge 1:m`)
  - Many to one: Master file with multiple visits, using with demographics (`merge m:1`)
  - Many to many: master with multiple visits + using file with multiple visits (`merge m:m`)

How to merge

- Need to make sure they are sorted AND saved
  - STATA 11 may do this automatically for you!
  - `sort idvariable`

- Steps
  - Load the master dataset into memory
  - Sort (just to be safe)
  - Command

`merge type commonvariable using "name of 2nd dataset.dta"`

Example: `merge 1:1 wihsid using "socdem.dta"`

See appearance of a "merge" variable which tells you where the observations came from (dataset 1, dataset 2, etc.)

The wonders of STATA on the Web

- Many things STATA can help you do

- To figure out how…
  - STATA help is one place to start
  - I've had luck with Google searches
  - UCLA has a helpful STATA site
  - Other discussion strings

- Good luck with your final projects