Stata 12 Merging Guide

Nathan Favero
Texas A&M University
October 19, 2012

Contents

Best Practices

- Backup everything
  - Save a separate copy of the original files somewhere before you start modifying/merging
- Always use a do-file to make changes
  - This makes it much, much easier to come back later and fix mistakes or update data
- Never merge by school/district name (use IDs)

Merge or Append?

Merge	Append
• Adding more variables	• Adding more observations (individuals and/or years)
• The same observations can be found in both files.	• The same variables can be found in both files

Merge

- Adding Variables

	year	campus	elementary	middle	high
1	2009	1902001	0	0	1
2	2009	1902041	0	1	0
3	2009	1902103	1	0	0
4	2009	1903001	0	0	1
5	2009	1903041	0	1	0
6	2009	1903101	1	0	0
7	2009	1904001	0	0	1
8	2009	1904041	0	1	0

	year	campus	teachersal~y	avgexper
1	2009	1902001	43244	11.90697
2	2009	1902041	42505	11.42857
3	2009	1902103	42965	15.95833
4	2009	1903001	43151	13.8
5	2009	1903041	41443	13.77419
6	2009	1903101	40684	14.8913
7	2009	1904001	40417	12.6241
8	2009	1904041	40157	11.36735

Append

- Adding Observations (Years)

	year	campus	elementary	middle	high
1	2009	1902001	0	0	1
2	2009	1902041	0	1	0
3	2009	1902103	1	0	0
4	2009	1903001	0	0	1
5	2009	1903041	0	1	0
6	2009	1903101	1	0	0
7	2009	1904001	0	0	1
8	2009	1904041	0	1	0

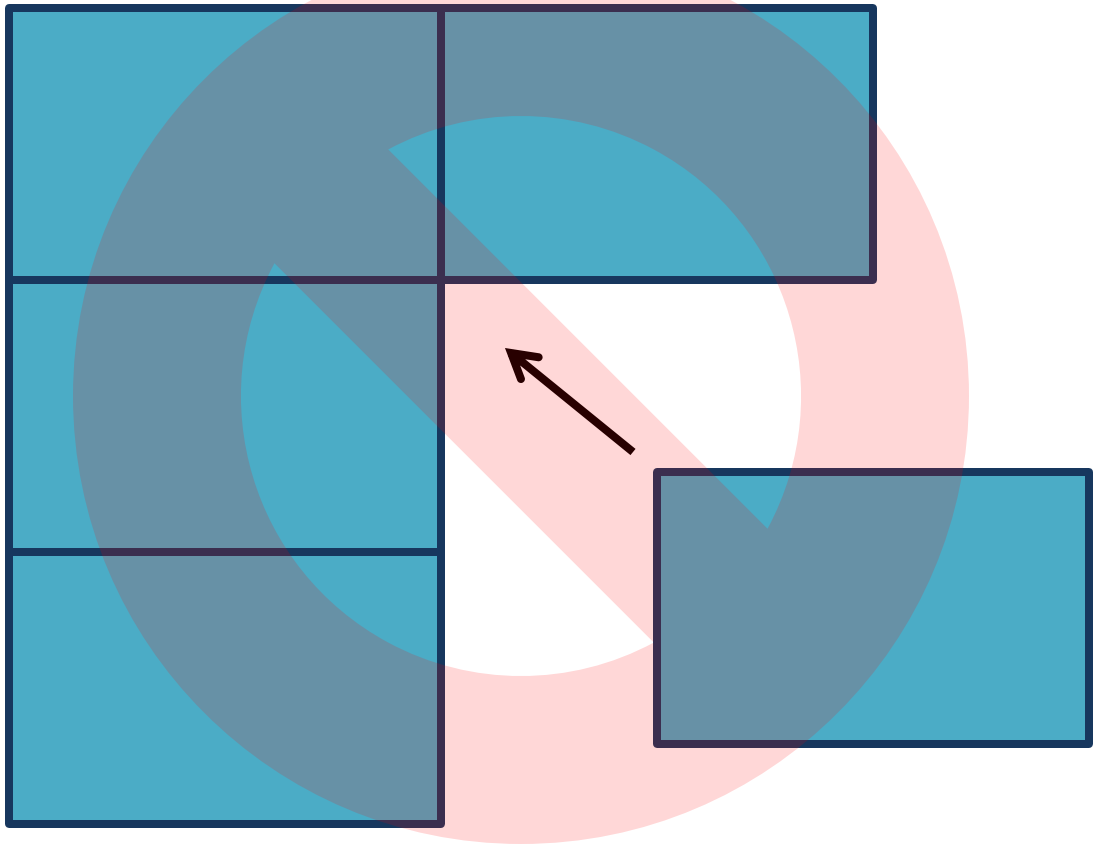
	year	campus	elementary	middle	high
1	2008	1902001	0	0	1
2	2008	1902041	0	1	0
3	2008	1902103	1	1	1
4	2008	1903001	0	0	1
5	2008	1903041	0	1	0
6	2008	1903101	1	0	0
7	2008	1904001	0	0	1
8	2008	1904041	0	1	0

Mixing Merge & Append

- You can only bind 1 direction (horizontally or vertically) at once.
- If you're combining both directions, you have to plan the order in which you perform your steps so that you never have to bind in 2 directions at once.

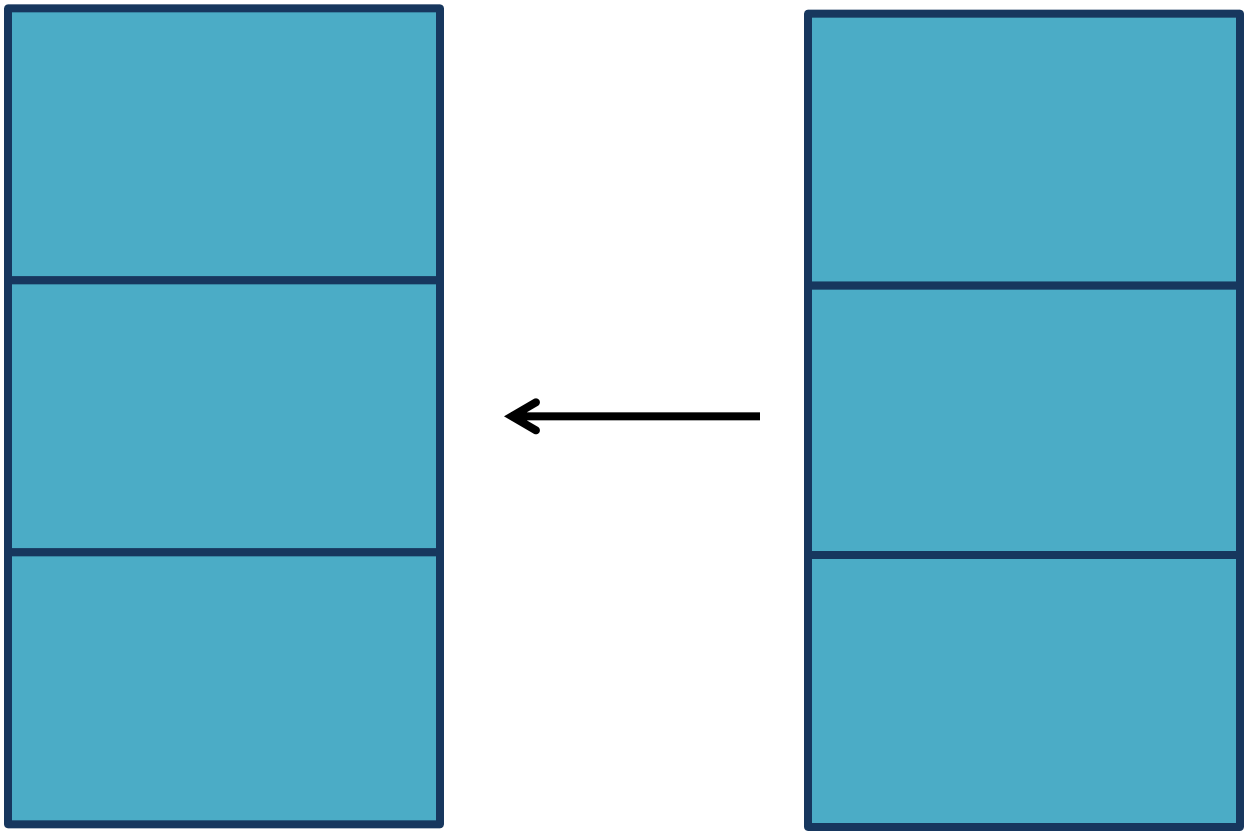
Mixing Merge & Append

- For example, this won't work.



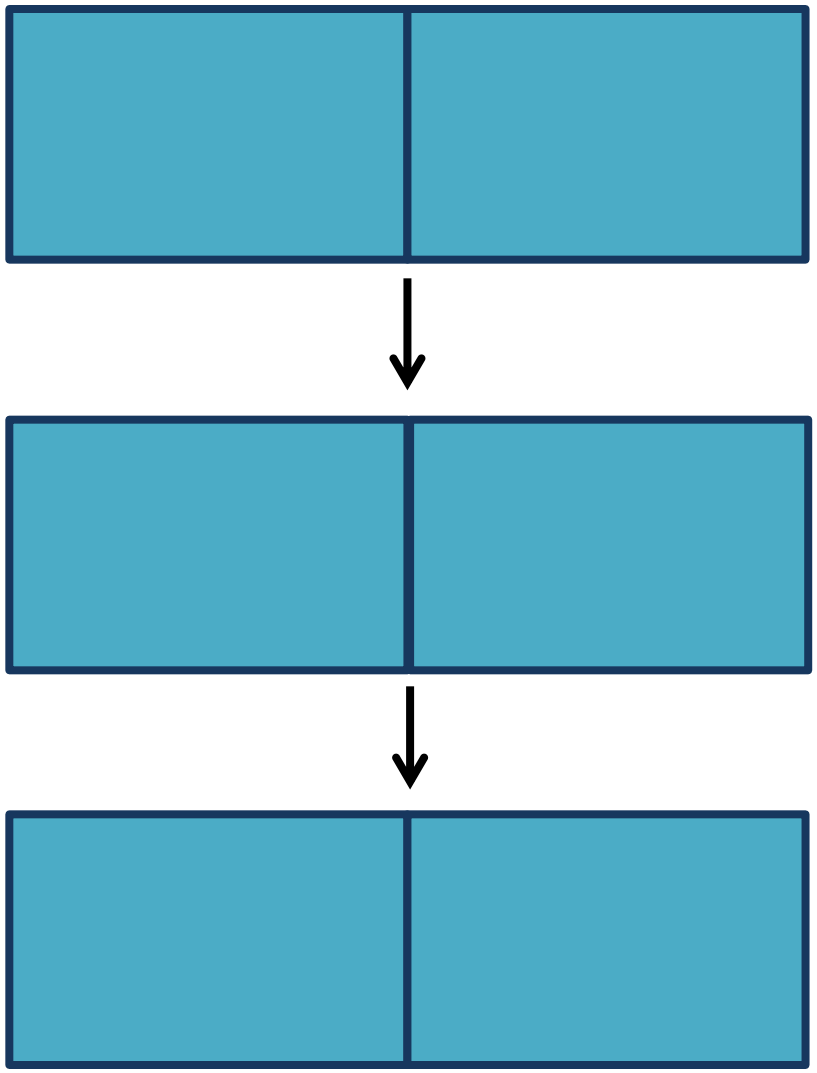
Mixing Merge & Append

- This will work.



Mixing Merge & Append

- Or this will work.



Using Merge

```
merge [n]:[n] [varlist] using [filename]
```

- merge 1:1
  - Try using this if you're unsure.
  - Merging two data files with the same unit of observation
  - Note: If using panel data, varlist must uniquely identify both individual and year
- merge m:m
  - Rarely used

Using Merge

`merge [n]:[n] [varlist] using [filename]`

- merge m:1 or 1:m
  - Merging smaller unit of analysis (e.g., school) with larger unit of analysis (e.g., district)
  - Merging panel data (school-year) with cross-sectional (school) or time-series data (year)
  - m corresponds to the more specific data; 1 corresponds to the more general data (if you get this switched, you'll simply get an error)
  - The [varlist] should uniquely identify the more general data (e.g, district, cross-section, or time-series)

Troubleshooting with Merge

- Does my variable list uniquely identify my observations? (Remember, you must specify both ID and Year variables in 1:1 for panel data.)
- Have I correctly specified 1:1, 1:m, or m:1?
- Should I be using append?
- Do I already have a variable named "_merge"?
- Are both files saved as ".dta"?
- Are the variable names exactly the same in both files?
- Are some of my variables strings in one file and numeric in the other?

Using Append

```
append using [filename]
```

- append is a much simpler command than merge
- Just make sure that:
  - The variable names are exactly the same in both files.
  - The variable types (string or numeric) are the same in both files.
  - Both files are saved as Stata files (“.dta”)

Other Useful Commands

```
destring [varlist], [replace or gen([var
  name])] ignore([characters to ignore])
```

- Convert a string variable to a numeric variable

```
tab [var name] if regexm([var name], "[^0-9
  .]")
```

- Show all of the values of a variable that are non-numeric

Other Useful Commands

`tostring [varlist], [replace or gen([var name])] [force]`

- Convert a numeric variable to a string variable

`gen [new var name] = string([numeric var name], "%12.0f")`

- Use this command instead of “tostring” if you have trouble with Stata giving you scientific notation

Other Useful Commands

```
gen [new var name] = substr([string var
  name],[starting position],[number of
  characters])
```

- Create a new variable with a fixed number of characters from another string variable (e.g., first 3 characters)

```
order [varlist], after([var name])
order [varlist], first
```

- Change the order in which your variables appear

Other Useful Commands

`duplicates report [varlist]`

- Learn about the number of duplicates

`duplicates drop`

- Drop any exact duplicates

`duplicates tag [varlist], gen([var name])`

- Create a variable that tags duplicate observations

Other Useful Commands

```
foreach var of varlist [varlist] {
rename `var' [prefix]_`var'
}
```

- Adds a prefix to the beginning of each variable name

```
reshape wide [vars that contain varying
  data], i([identifying var(s)]) j([var that
  distinguishes observations that have the
  same identifying var(s)])
```

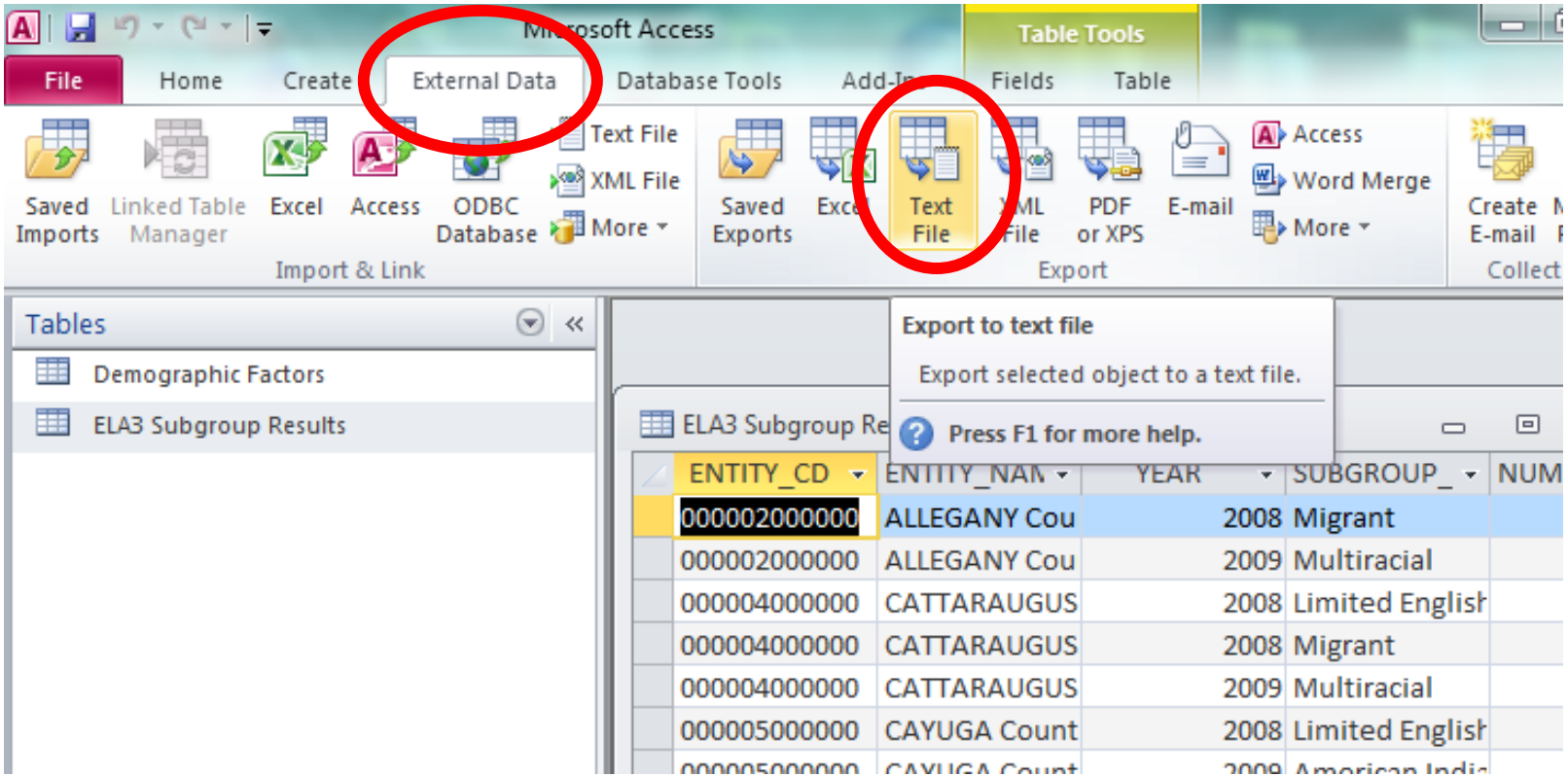
- Consolidates multiple observations into a single observation by adding variables

Importing Data into Stata: Delimited Files

- How to get Stata to read in the first row as variable names
  - Logic: At least one of the variables must be a completely numeric variable.
  1. Open your file in Excel
  2. Create a new column, and give it a variable name in the first cell. (You can put numbers in the column's other cells, or just leave them blank.)
  3. Save the file, and reload it into Stata.

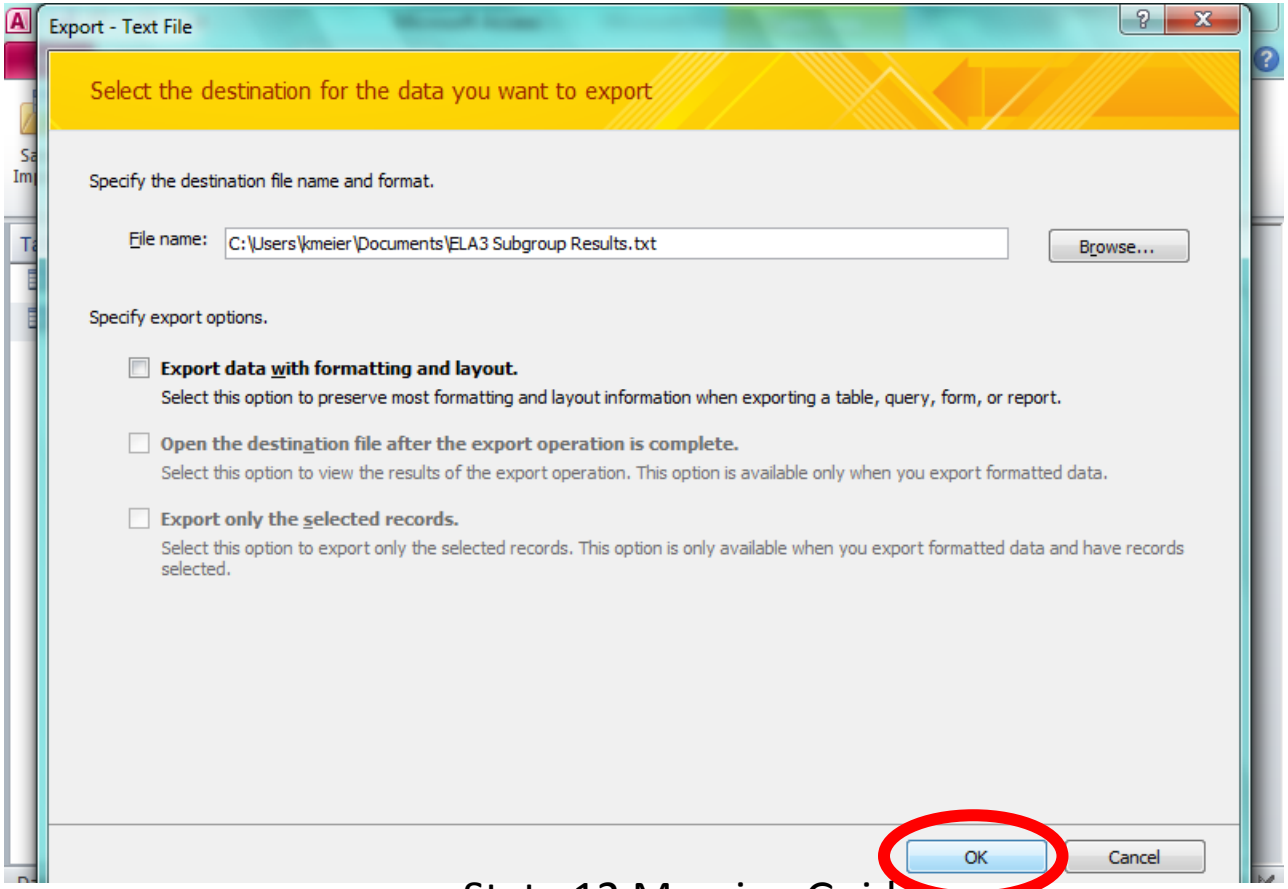
Importing Data into Stata: Microsoft Access 2010

- External Data → Text File (Export)



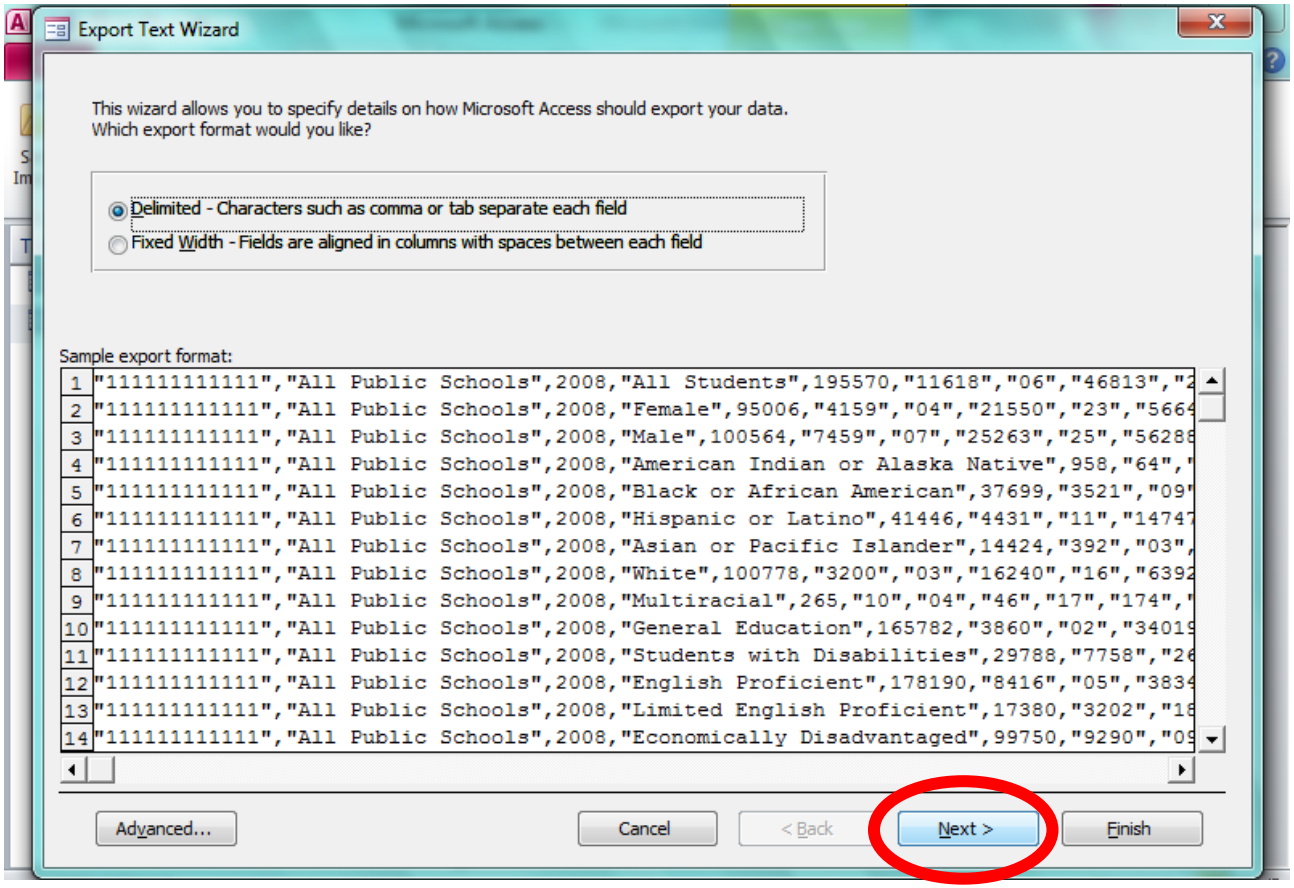
Importing Data into Stata: Microsoft Access 2010

- Select a file location and name. Click OK.



Importing Data into Stata: Microsoft Access 2010

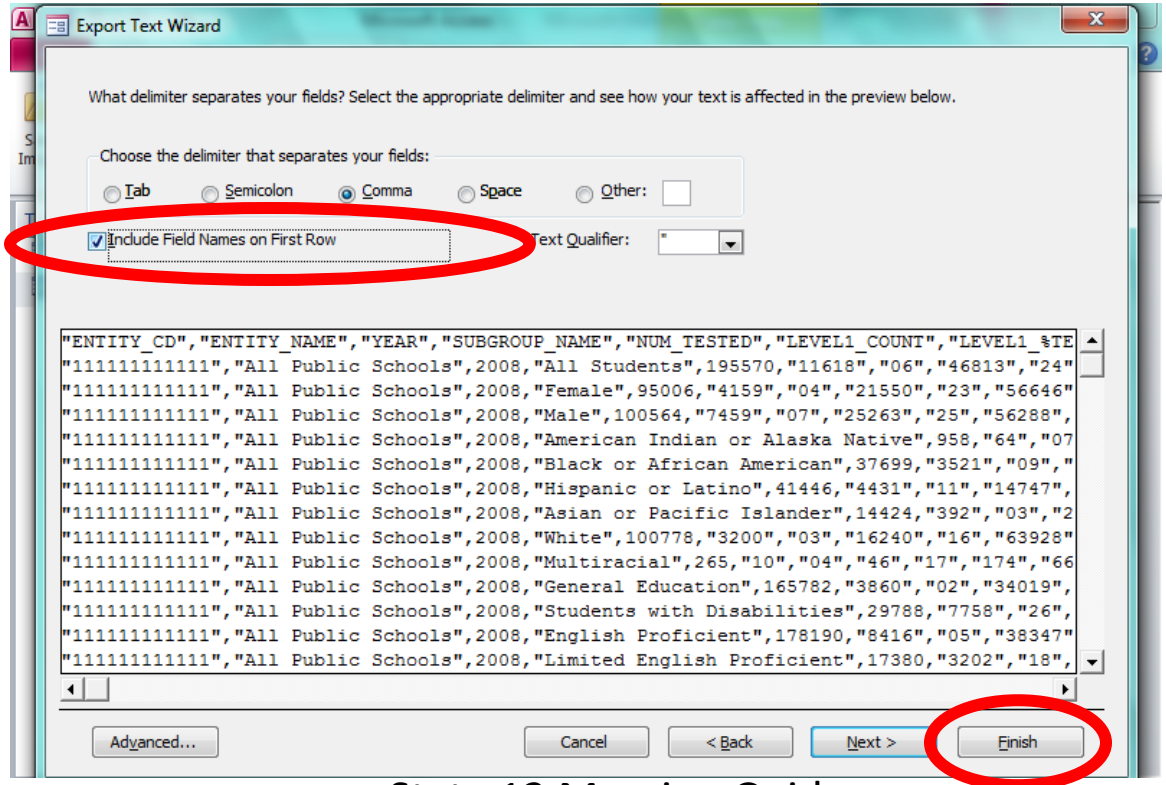
- Click Next.



Stata 12 Merging Guide

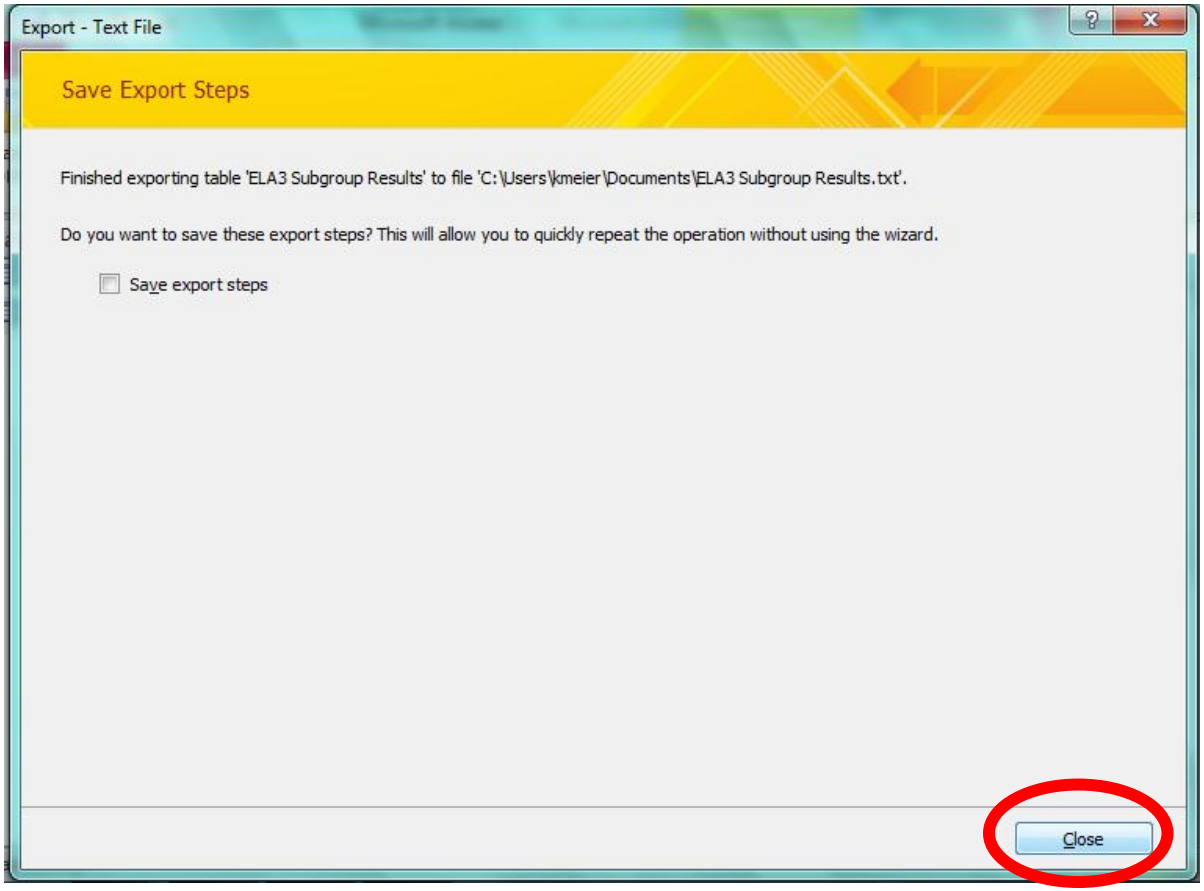
Importing Data into Stata: Microsoft Access 2010

- Select “Include Field Names on First Row.”
- Click Finish.



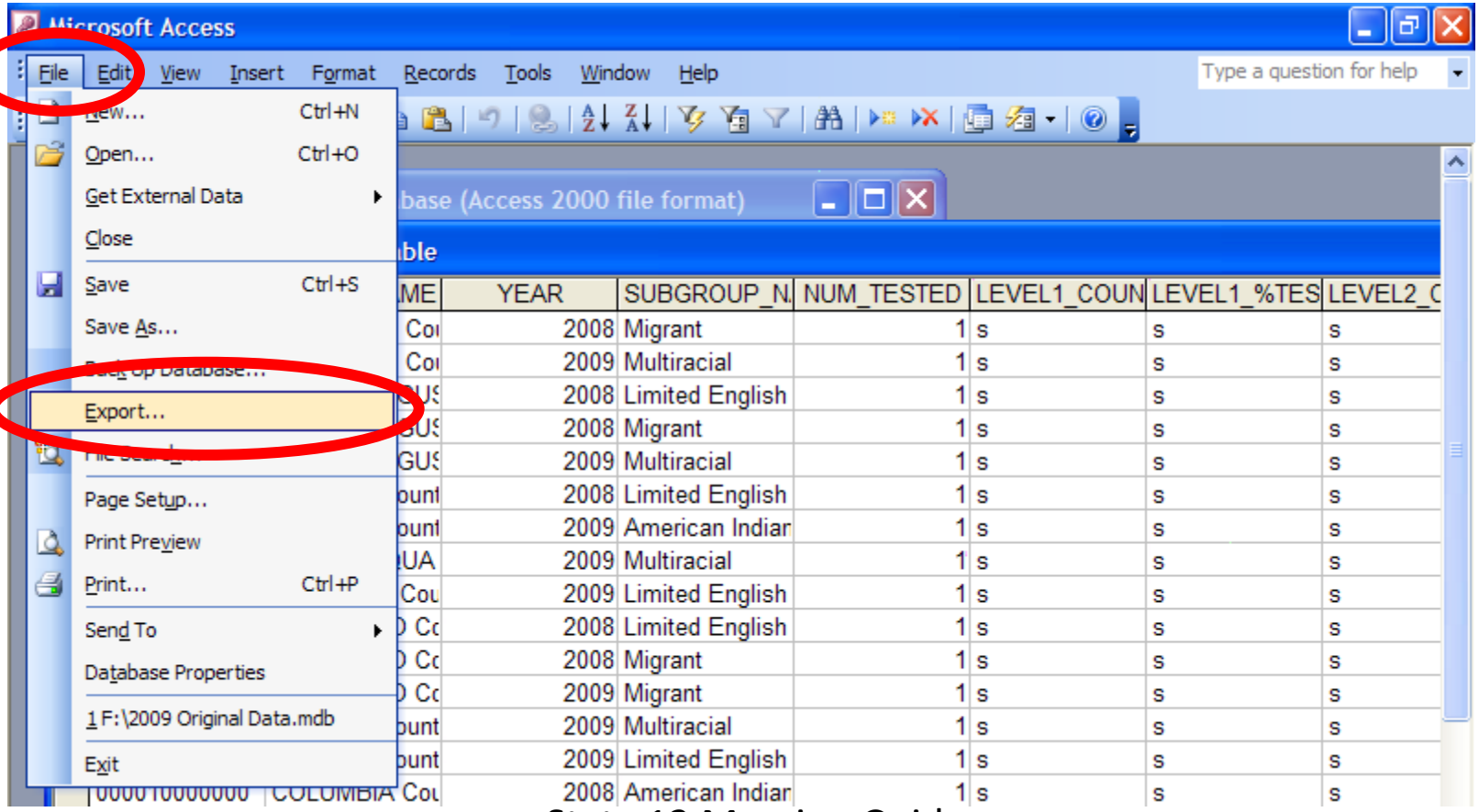
Importing Data into Stata: Microsoft Access 2010

- Click Close.



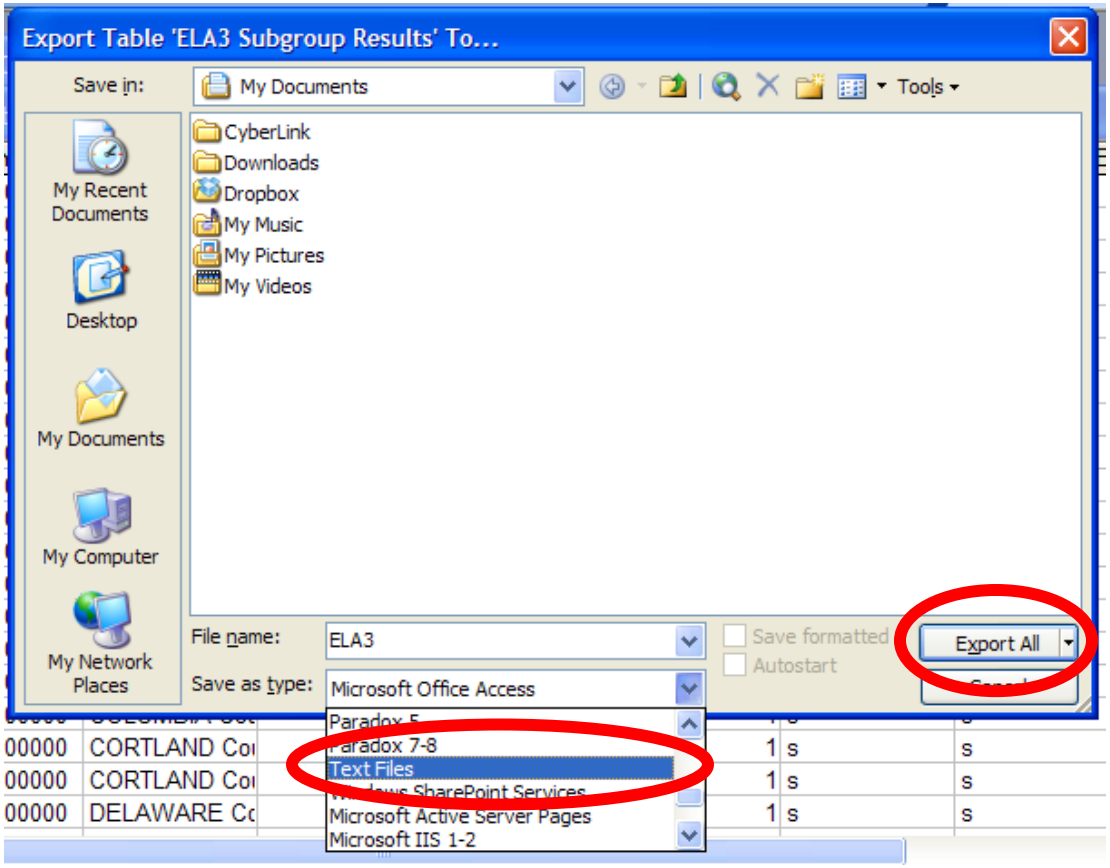
Importing Data into Stata: Microsoft Access 2003

- File → Export



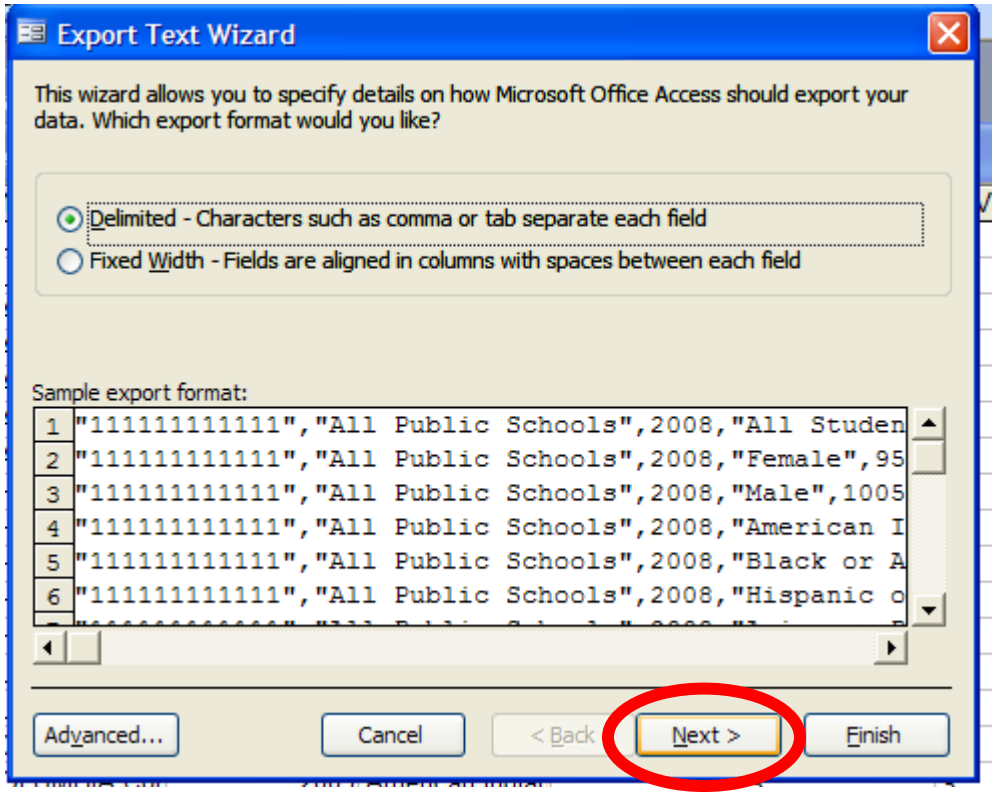
Importing Data into Stata: Microsoft Access 2003

- Choose a location/name.
- Under “Save as type,” select “Text Files.”
- Click Export All



Importing Data into Stata: Microsoft Access 2003

- Click Next.



Importing Data into Stata: Microsoft Access 2003

- Select “Include Field Names on First Row.”
- Click Finish.