

Data Manipulation in Stata

November 9, 2020



01

*Append, Merge,
Collapse*

02

Egen

03

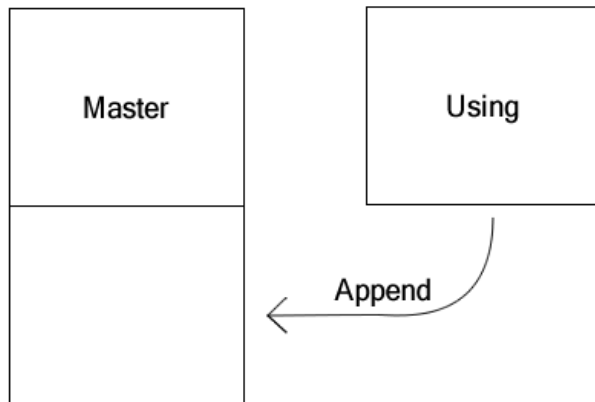
Graphing

04

Export Data

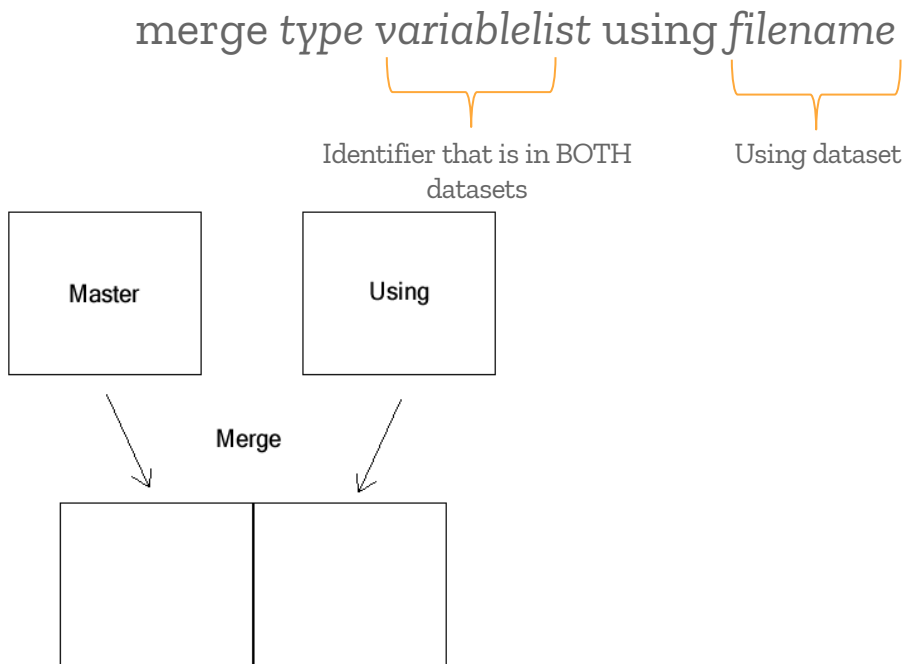
- Working with more than one dataset
- Add observations/cases to an existing data set with the same or similar variables

append using dataset.dta



Appending Datasets

- Working with more than one dataset
- Adding variables
- 4 different cases



Merging Datasets

1. One-to-one merge

`merge 1:1 variablelist using filename`

2. One-to-many merge

`merge 1:m variablelist using filename`

3. Many-to-one merge

`merge m:1 variablelist using filename`

4. Many-to-many merge

`merge m:m variablelist using filename`

Merging Datasets

The command will also create an additional variable “_merge” that identifies if an observation was matched in the merge.

It provides three indicators:

- 1 = observation found only in the master dataset
- 2 = observation found only in the using dataset
- 3 = observation found in both master and using dataset (complete match)

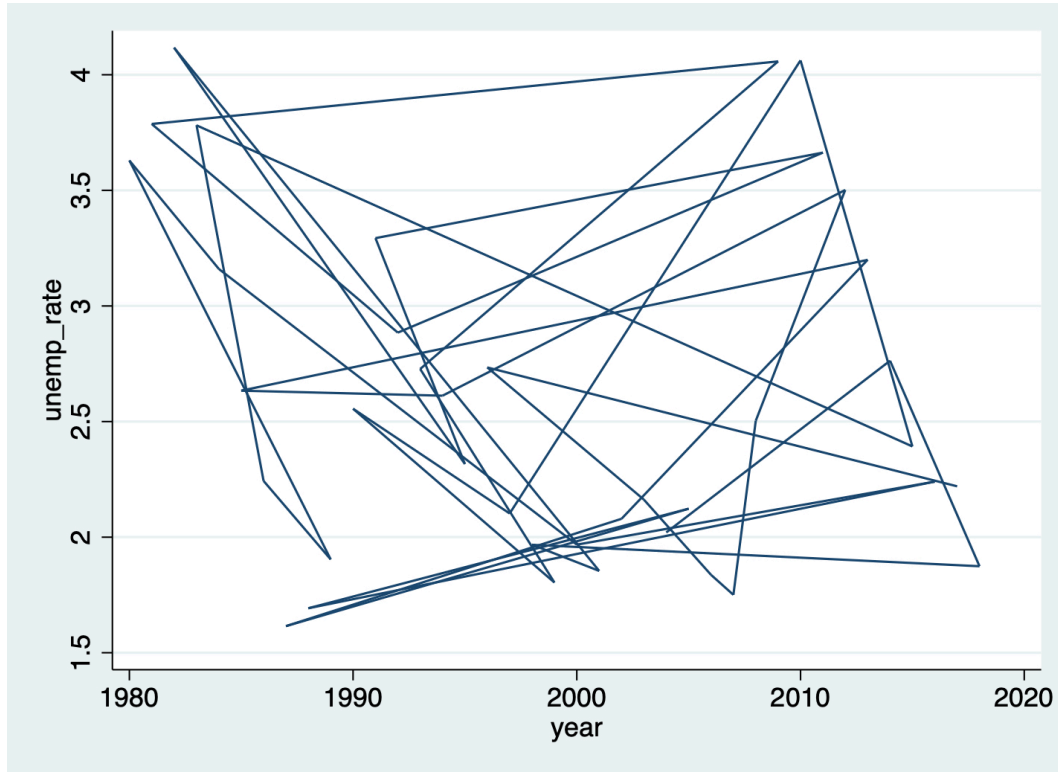
Merging Datasets

- This command takes an open (or master) dataset and creates a new dataset by summarizing statistics on a selected variable
- IOW, this creates a dataset of summary statistics

`collapse (statistic) variablelist, by (what variable to collapse on)`

Collapsing Datasets

```
graph twoway line unemp_rate year if state == "DE"
```



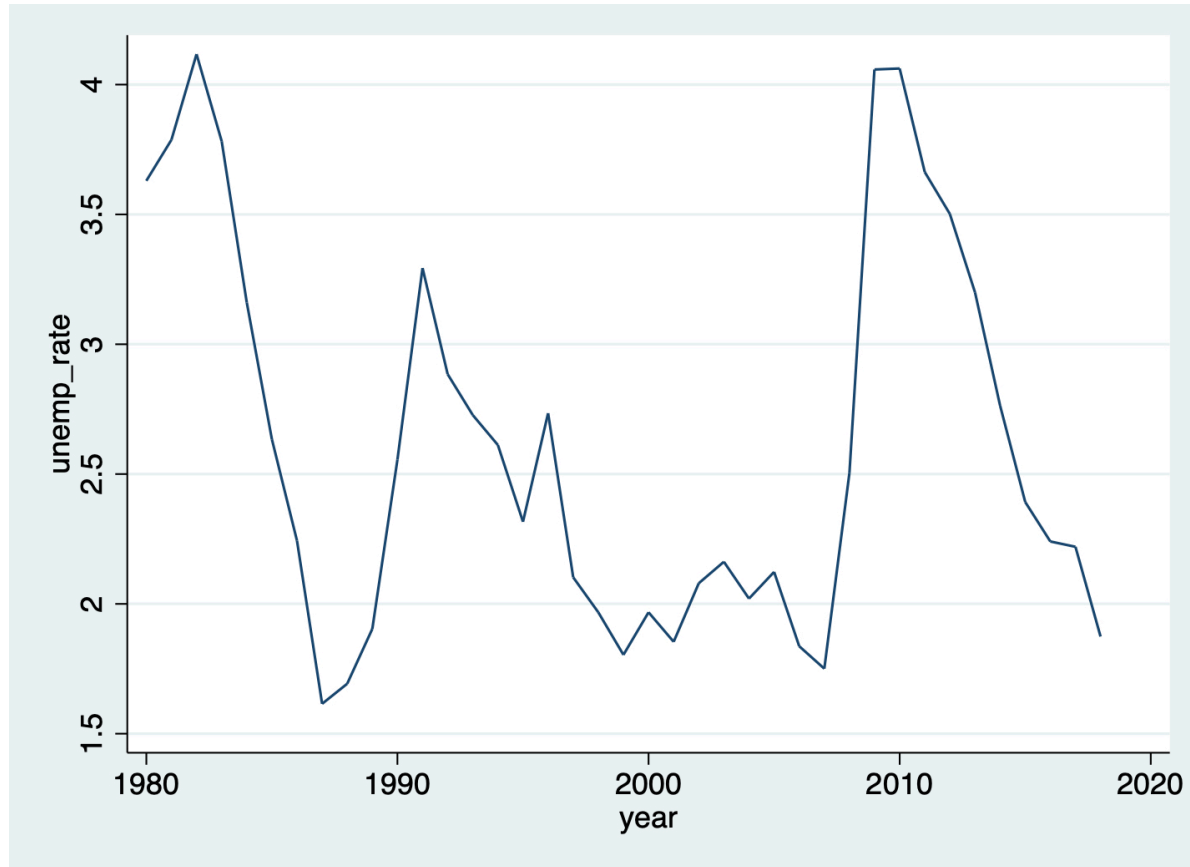
Line Graph

8

This graph looks strange. Let's sort our data by year.

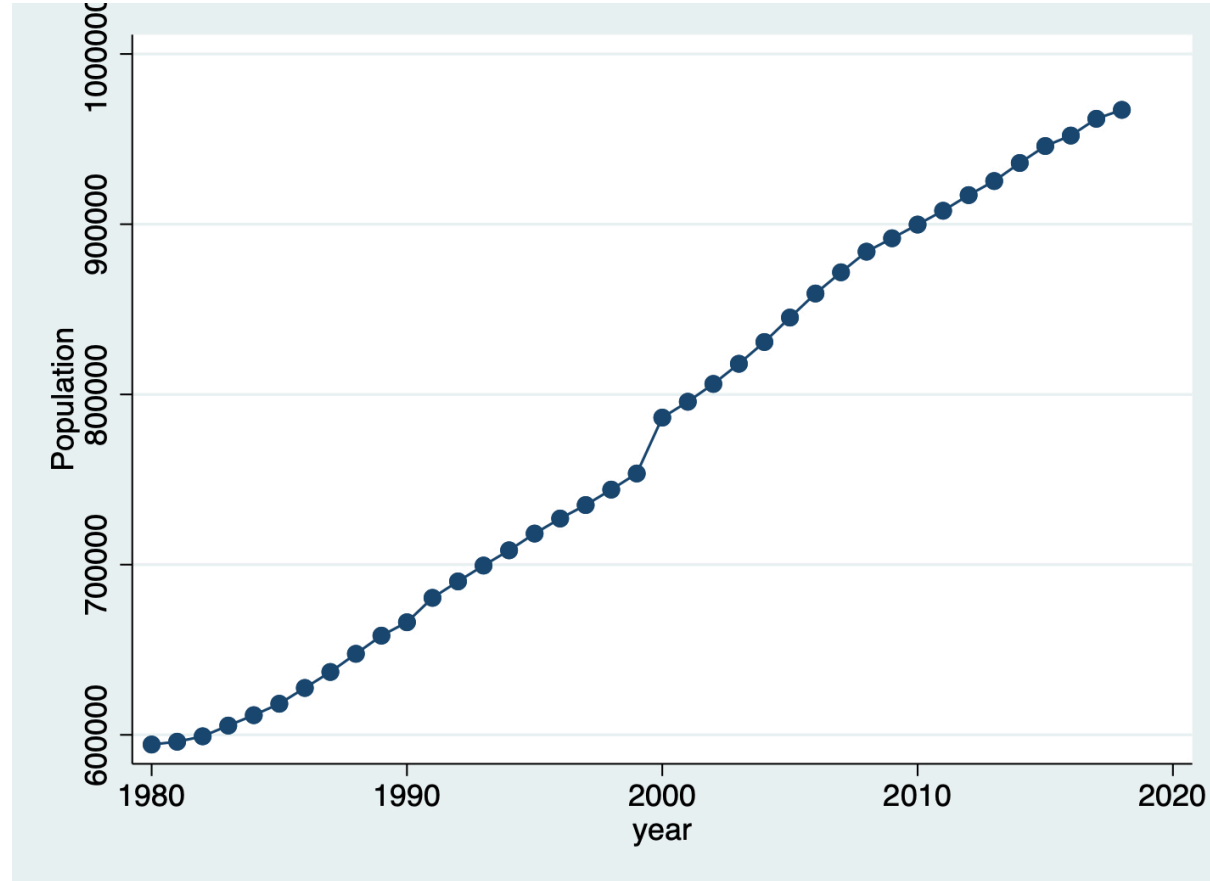

```
sort year
```

```
graph twoway line unemp_rate year if state == "DE"
```



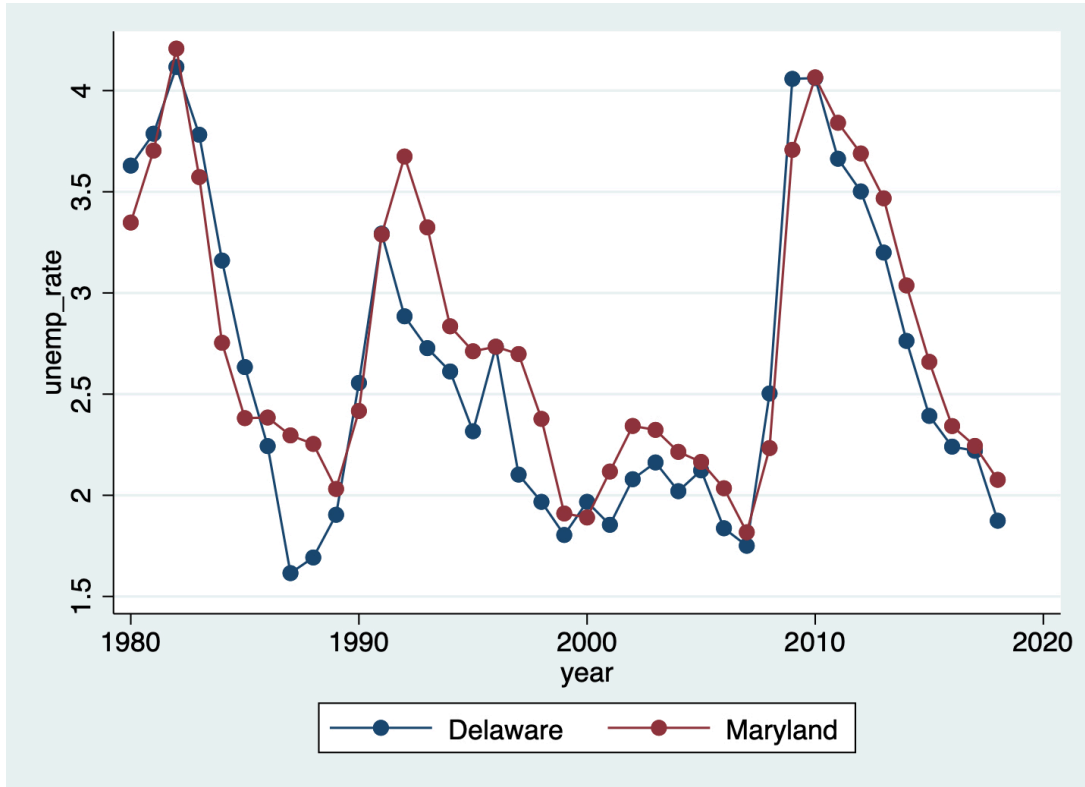
Line Graph

```
graph twoway connected pop year if state == "DE"
```



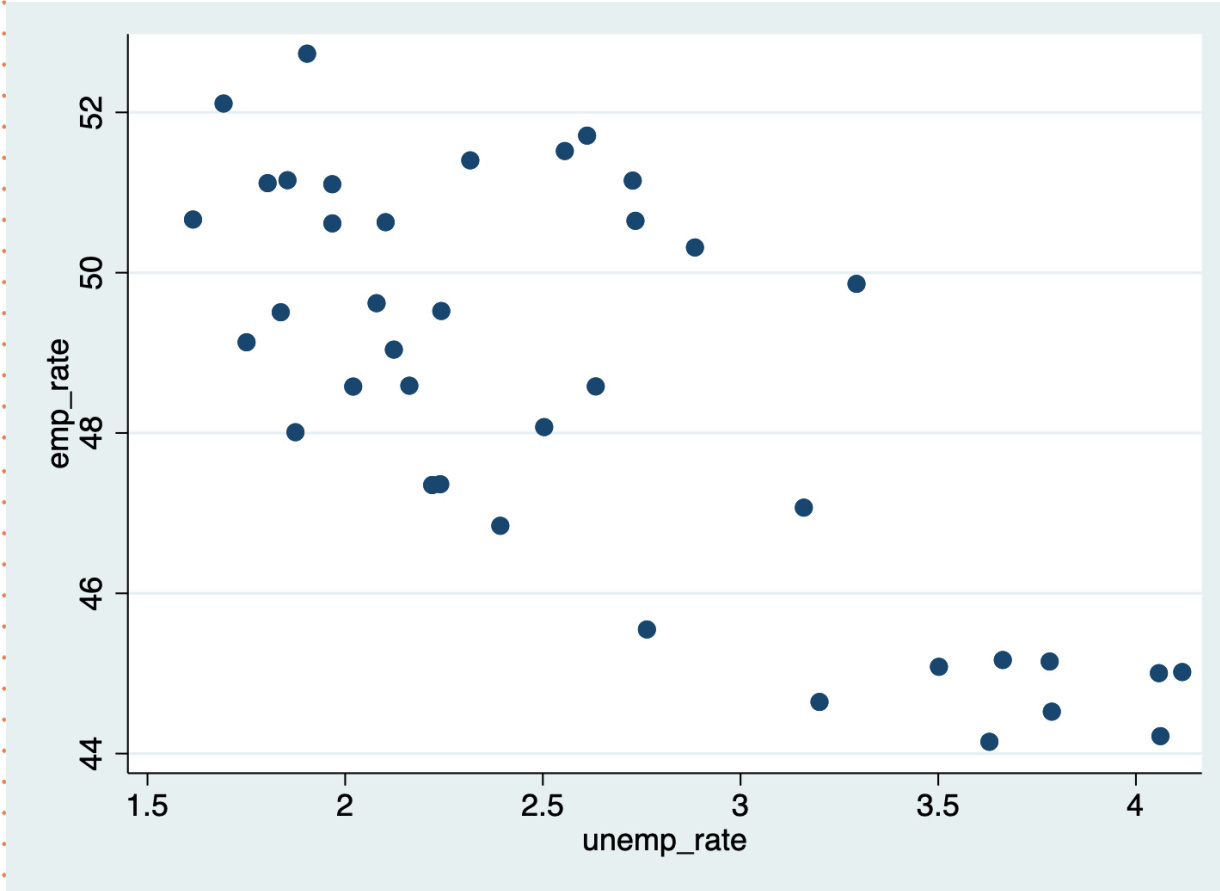
Line Graph

```
graph twoway (connected unemp_rate year if state == "DE") (connected unemp_rate year if state == "MD"), legend(order(1 "Delaware" 2 "Maryland"))
```



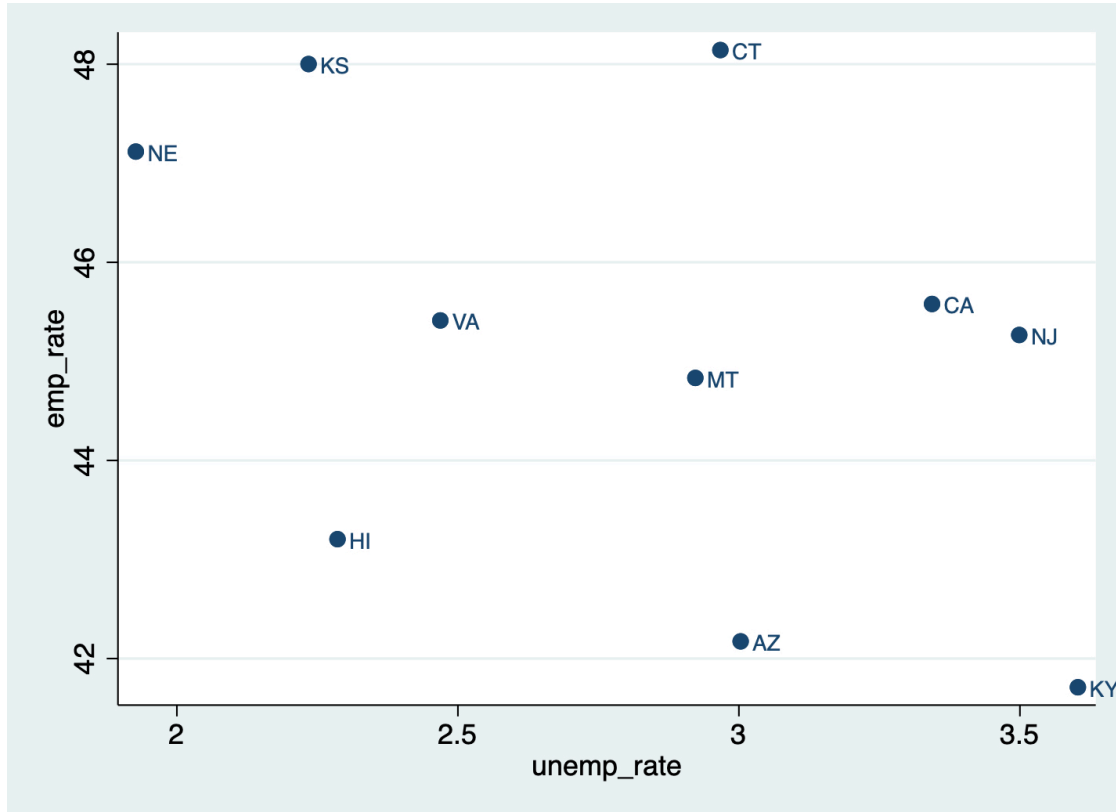
Line Graph

```
scatter emp_rate unemp_rate if state == "DE"
```



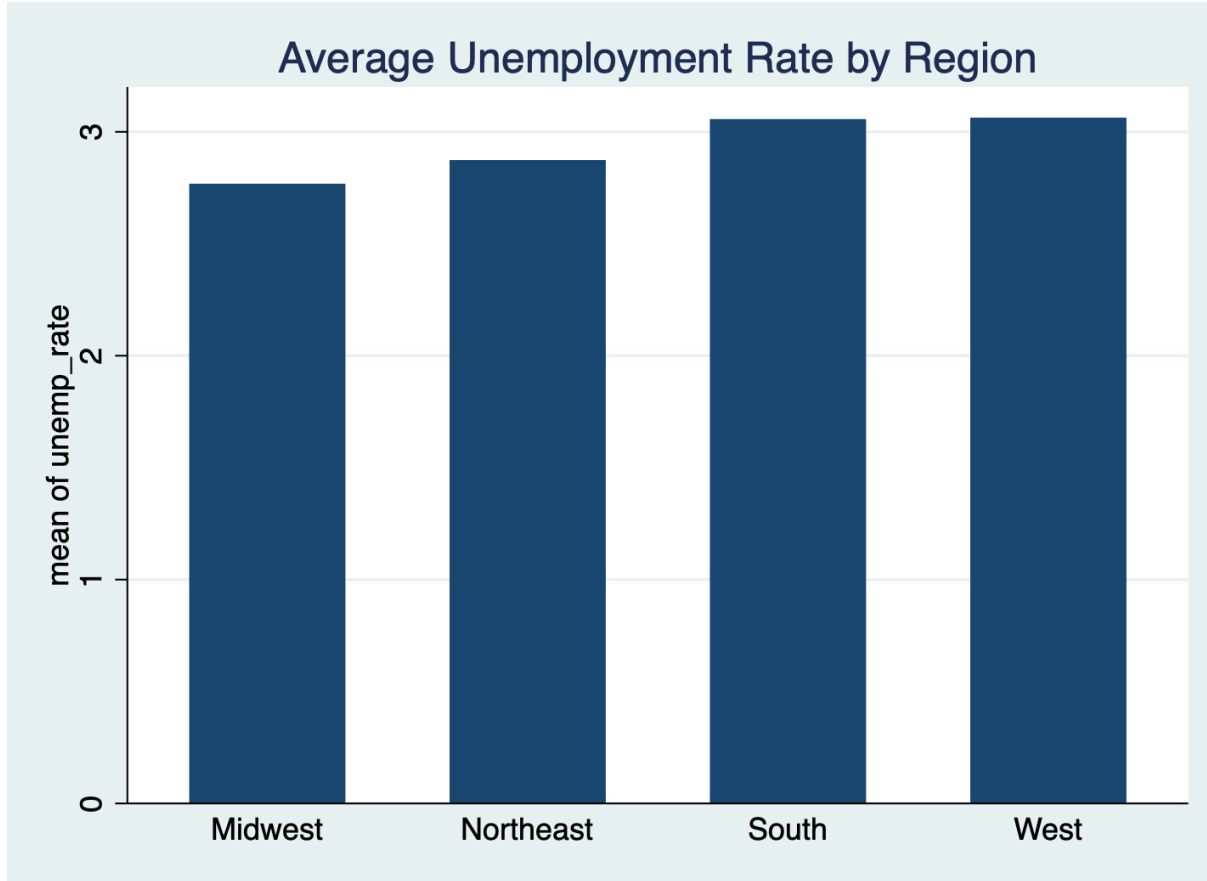
Scatter Plots

```
scatter emp_rate unemp_rate if _n<=10, mlabel(state)
```



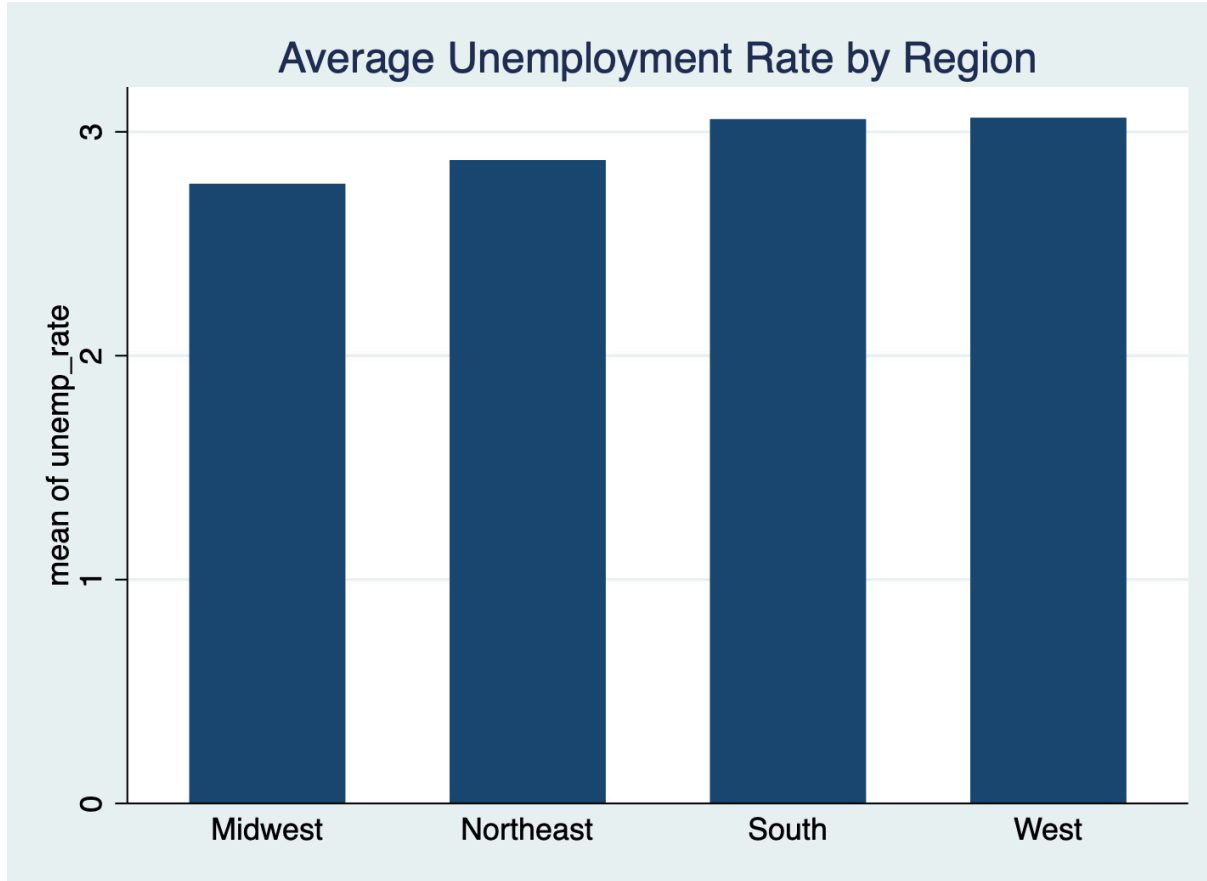
Scatter Plots

```
graph bar unemp_rate, over(region) title("Average Unemployment Rate by  
Region")
```



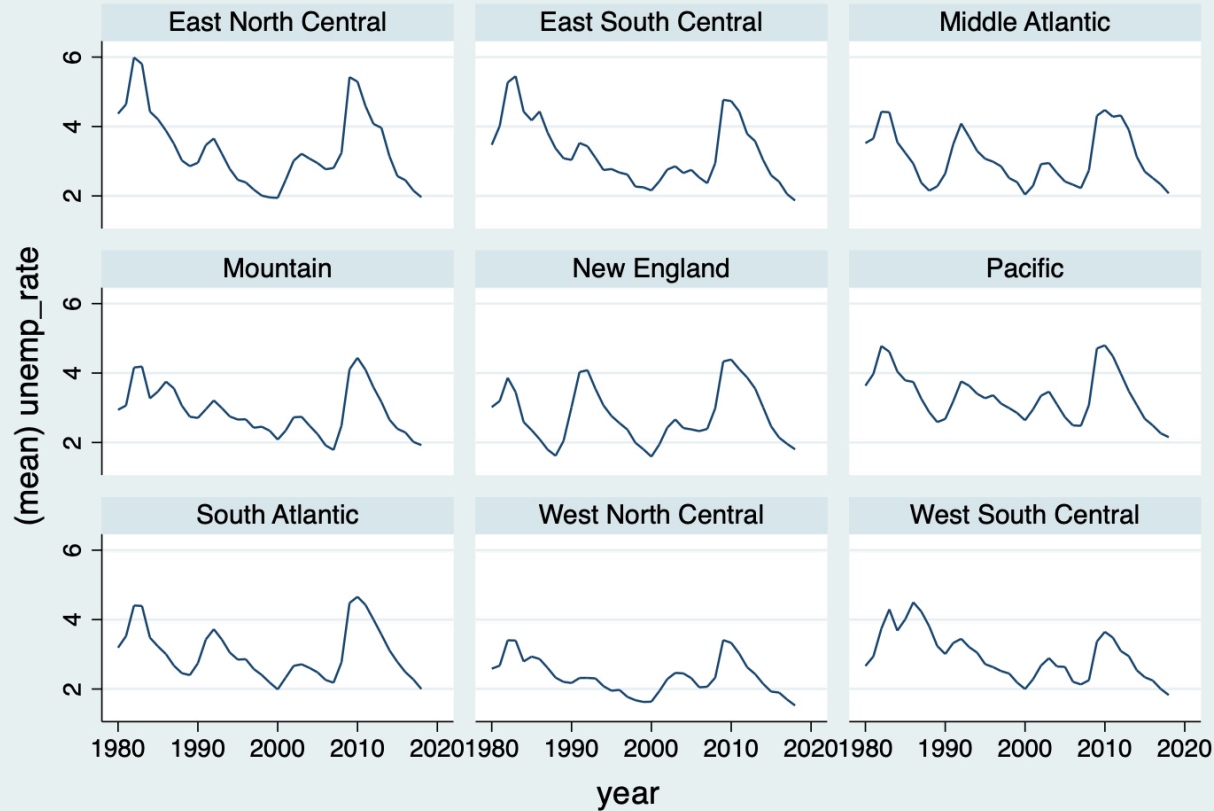
Bar Graphs

```
graph bar unemp_rate, over(region) title("Average Unemployment Rate by  
Region")
```



Bar Graphs

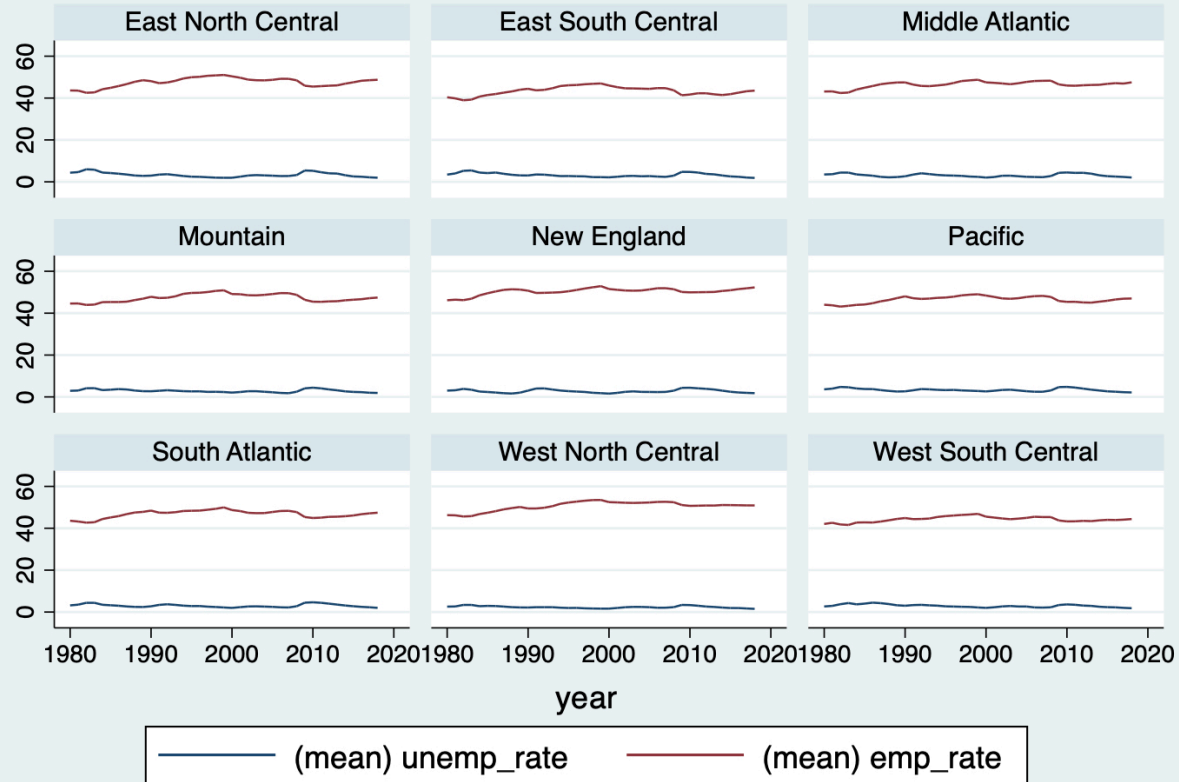
line unemp_rate year, by(division)



Graphs by Division

More Line
Graphs

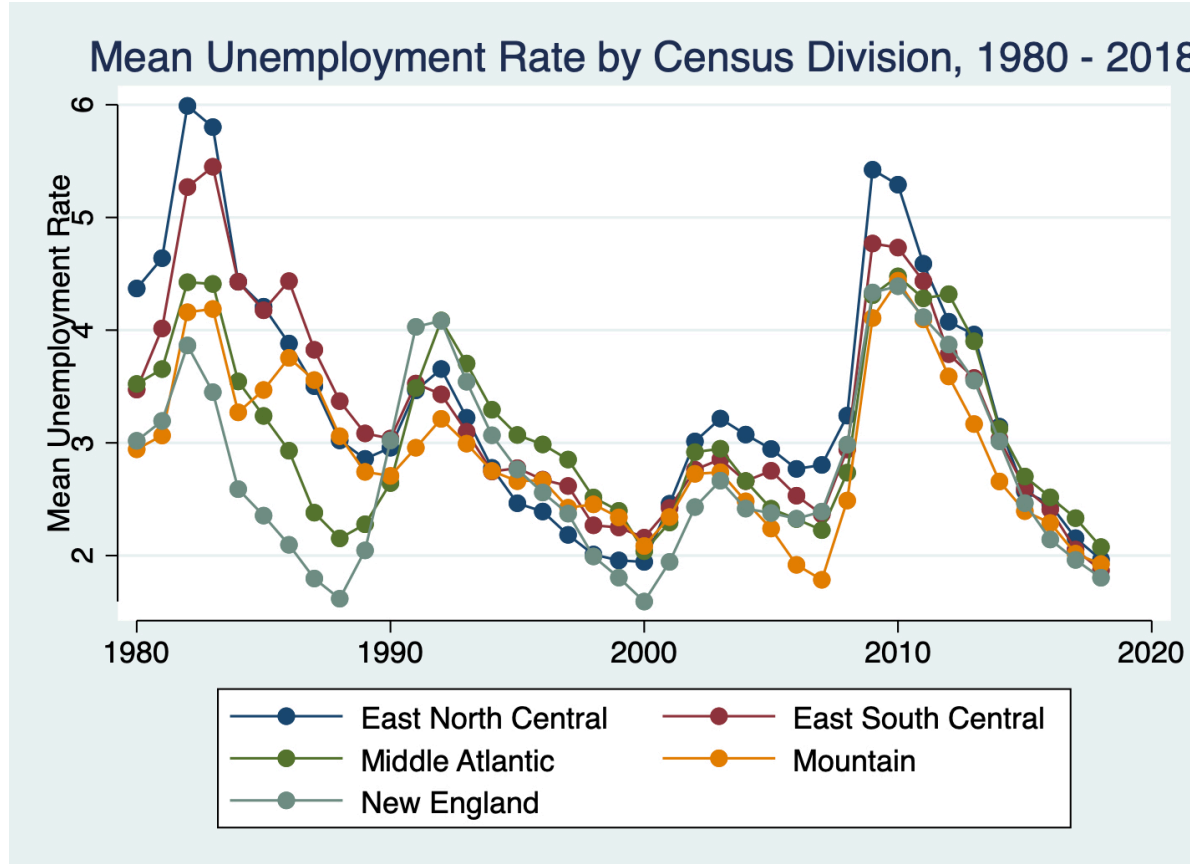
line unemp_rate emp_rate year, by(division)



Graphs by Division

More Line
Graphs

line unemp_rate emp_rate year, by(division)



More Line
Graphs