

# Lecture 1 - Introuduction

## ECON 4651: Principles of Econometrics for Business and Analytics

---

Jason Cook  
Fall 2020

# Prologue

# Who am I?

## Jason Cook

- Applied microeconomist and econometrician
- I study the economics of education and public economics.
  - Charter schools
  - Racial segregation
  - Food Assistance: SNAP/WIC

Where can you find me?

- **Office hours:** T/Th 3:30-4:30pm or by appointment. Schedule with **Calendly**
  - **Zoom Meeting ID:** 331 921 1844
  - **Password:** 4651
- **Email:** **Don't email (use Slack Instead!)**

# Syllabus

# Syllabus

## Grading

- Attendance **(1 pt per class up to 25 pts - 5%)**
  - Watch entire lecture, either by Zoom or Canvas, or combo
  - Lab days, submit Stata do-file for credit
  - Must be completed before Sunday at midnight following the lecture

# Syllabus

## Grading continued...

- 6 Problem Sets **(25 pts each - 25%)**
  - Mix of theory and empirical tasks
  - Optional groups of up to 4 **(turn in own responses)**
  - Will drop the lowest-scored problem set
- 4 Group Assignments **(25 pts each - 20%)**
  - Assigned a group of 5<sup>†</sup>
  - Open-ended empirical assignments
  - Submit report and video presentation
  - Give feedback on 3 other presentations
- Midterm **(100 pts - 20%)**
- Comprehensive Final **(150 pts - 30%)**

[†]: Contact TA if group member isn't helping and we will contact entire group and may dock points to that member

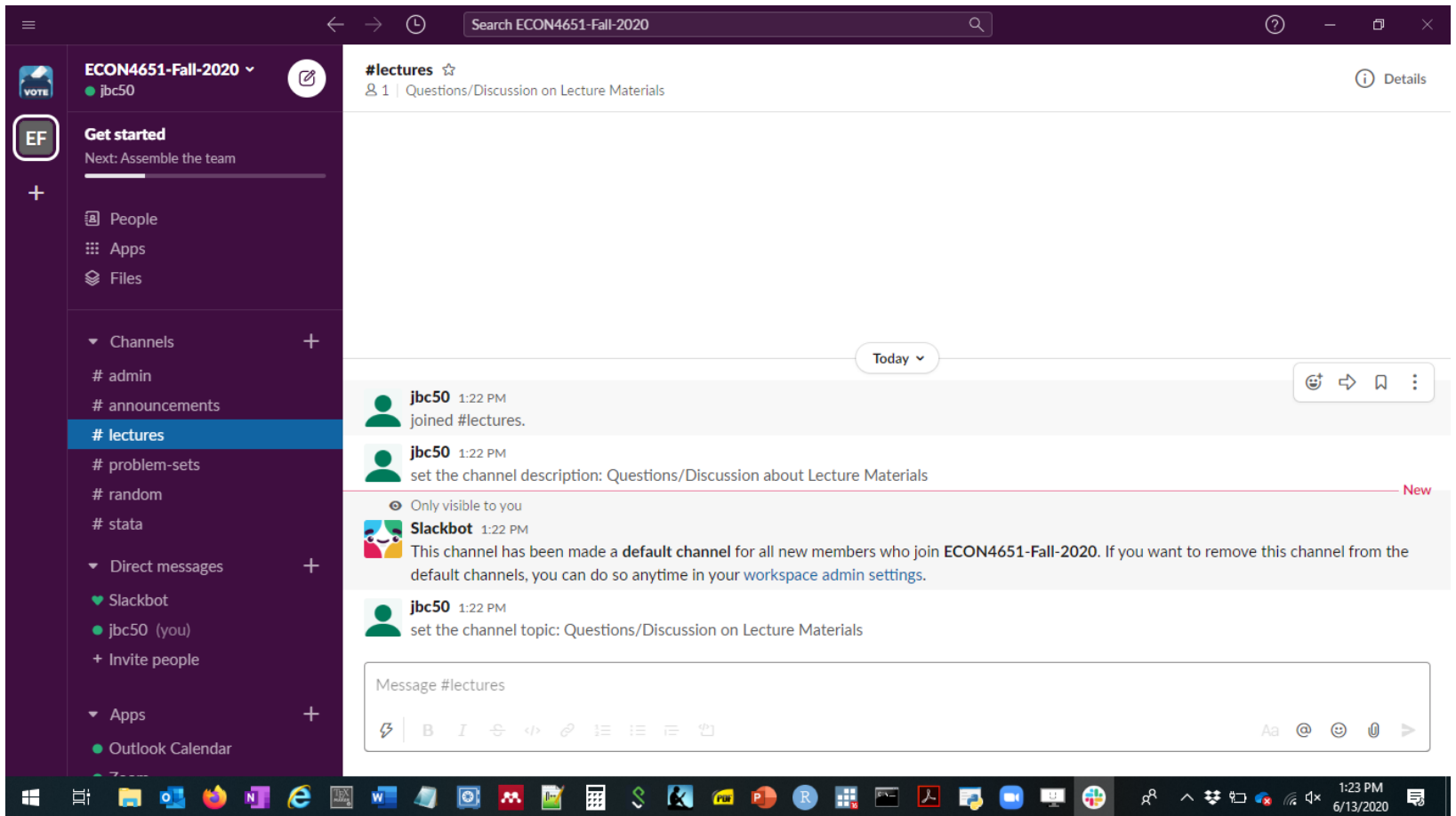
# Slack

# Slack

- I will utilize Slack for this course rather than email (response guaranteed)
- Please do not email me (you may not get a response)
- Create a Slack Account and join our Workspace [Econ 4651: Fall-2020](#)
- [Getting Started with Slack](#)



# Slack




- Channels (#) along the left panel organize topics for discussion
- Send Direct Messages in lower left panel

## In Class

- We will periodically use Slack in class for polling and other activities, so have it ready

## Ettiquete

- Respond to comments and questions in **threads**
  - Hover over message and click thread  icon
- Be kind and professional
- Slack is **judgment-free zone**, no such thing as "stupid questions"
- TAs and I will monitor Slack to answer questions as well
- To ask a specific question to me or the TAs, be sure to tag us in the post using @
- Please do answer other students questions and contribute to discussion
- In fact...

## Bonus

To help facilitate activity on Slack. At the end of the semester, I will subjectively award up to **25 points** (equivalent to one full problem-set) based on each student's activity on Slack

Activity includes both asking and answering questions, though I'll put more weight on answering

## Channels:

- #admin
- #announcements
- #lectures
- #problem-sets
- #in-class
- #random
- #stata

## Direct Messages:

- Send direct messages jointly to myself **and both TAs** for anything personal questions or concerns
  - TA will usually respond, but I'll step in as needed

# Zoom

## Ettiquette

- keep video on
- raise hand
- use chat feature (TAs will be monitoring chat)
- mic off unless called on

# Poll Everywhere



# Poll Everywhere

- I'll use poll everywhere for in-class polling
- These polls are to gauge understanding and will not be for points
- No need to enroll in Poll Everywhere, just click links in Slack channel

# Motivation

## Why study econometrics?

1. Develop **skills that employers value**.
2. Cultivate **healthy skepticism**.
3. Learn about the world using **data**.

# Motivation

## Why study econometrics?

Provide answers to important questions

- Do minimum wage policies **reduce poverty**?
- Does the death penalty **deter violent crime**?
- Does recreational marijuana **cross state lines**?
- Are recessions **good for your health**?
- How will global warming **affect the economy**?
- Will Donald Trump **win again**?
- Do mandatory mask policies **reduce the spread of COVID-19**?

# Econometrics

Most econometric inquiry concerns one of two distinct goals:

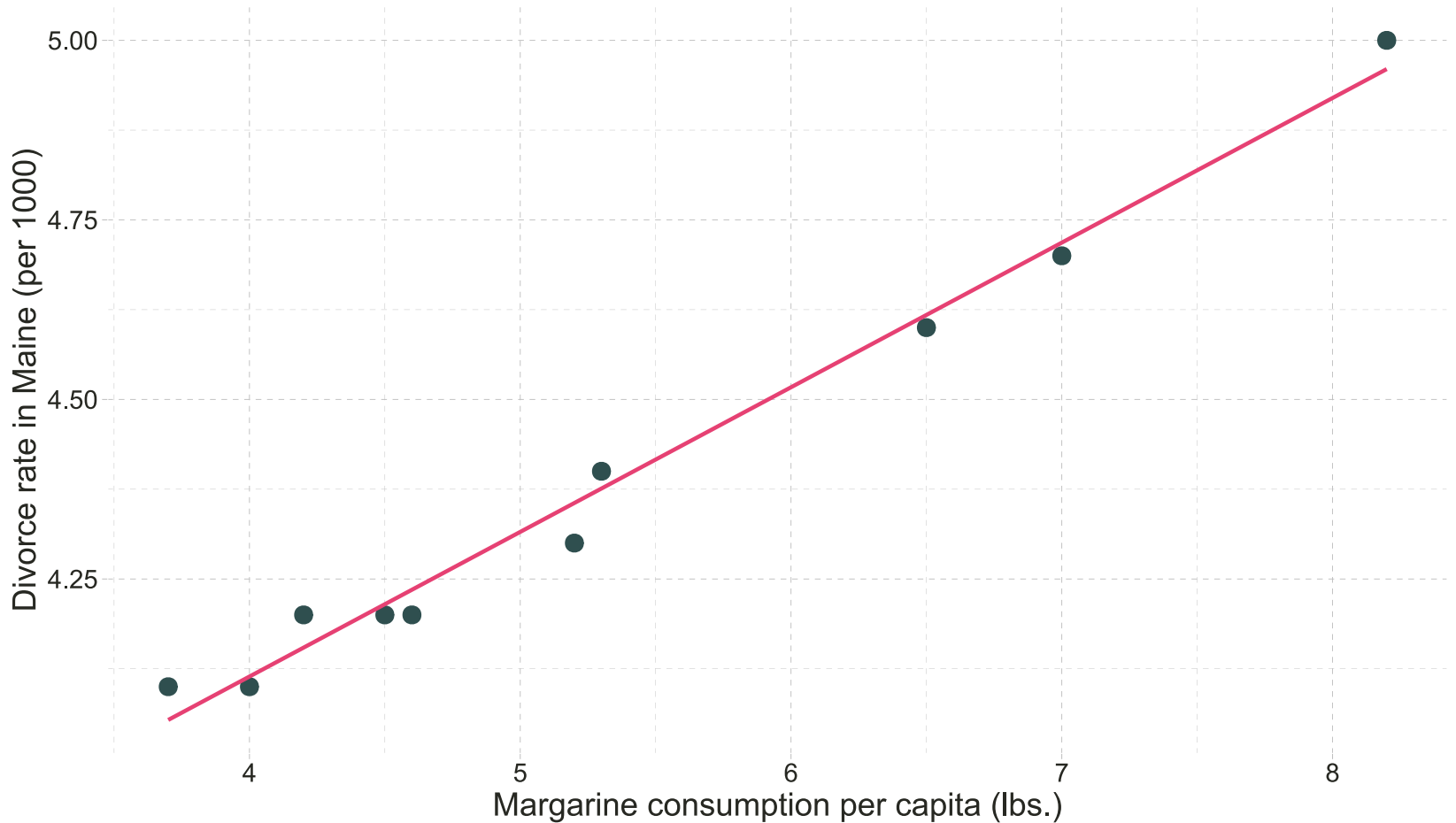
1. **Prediction:** Accurately predict or forecast an outcome given a set of predictors. Given what we know about  $x$ , what values do we expect  $y$  to take?
2. **Causal identification:** Estimate the effect of an intervention on an outcome. How does  $y$  change when we change  $x$ ?

The main focus of ECON 4651 and QAMO 3030 is causal identification.

- But...both rely on a common set of statistical techniques.

# Econometrics

Not all relationships are causal



# Econometrics

## Correlation vs. Causation

Common refrain: *"Correlation doesn't necessarily imply causation!"*

- **Q:** Why might correlation fail to describe a causal relationship?
- **A:** Omitted-variables bias, selection bias, simultaneity, reverse causality.

Correlation can imply causation.

- Requires strong assumptions.
- **Real life often violates these assumptions!**
- **Solutions:** Conduct an experiment or find a natural experiment.

# Example: *Blue Paradox*

Recent study by economist Grant McDermott and coauthors.

**Question:** Do commercial fishers preempt fishing bans by increasing their fishing effort before the bans go into effect?

## Motivation

- Recent conservation efforts seek to preserve aquatic habitat and increase fish stocks.
- Policy lever: Restrict fishing activity in marine protected areas.
- Concern: Preemptive behavior could *decrease* fish stocks.

## Data

- Vessel-level data on fishing effort/intensity.

# Example: *Blue Paradox*

## **Natural Experiment**

Phoenix Islands Protected Area (PIPA)

- First mentioned on 1 September 2014; implemented 1 January 2015.
- *Treatment group*: PIPA.
- *Control group*: Outlying Kiribati islands.



# Example: *Blue Paradox*

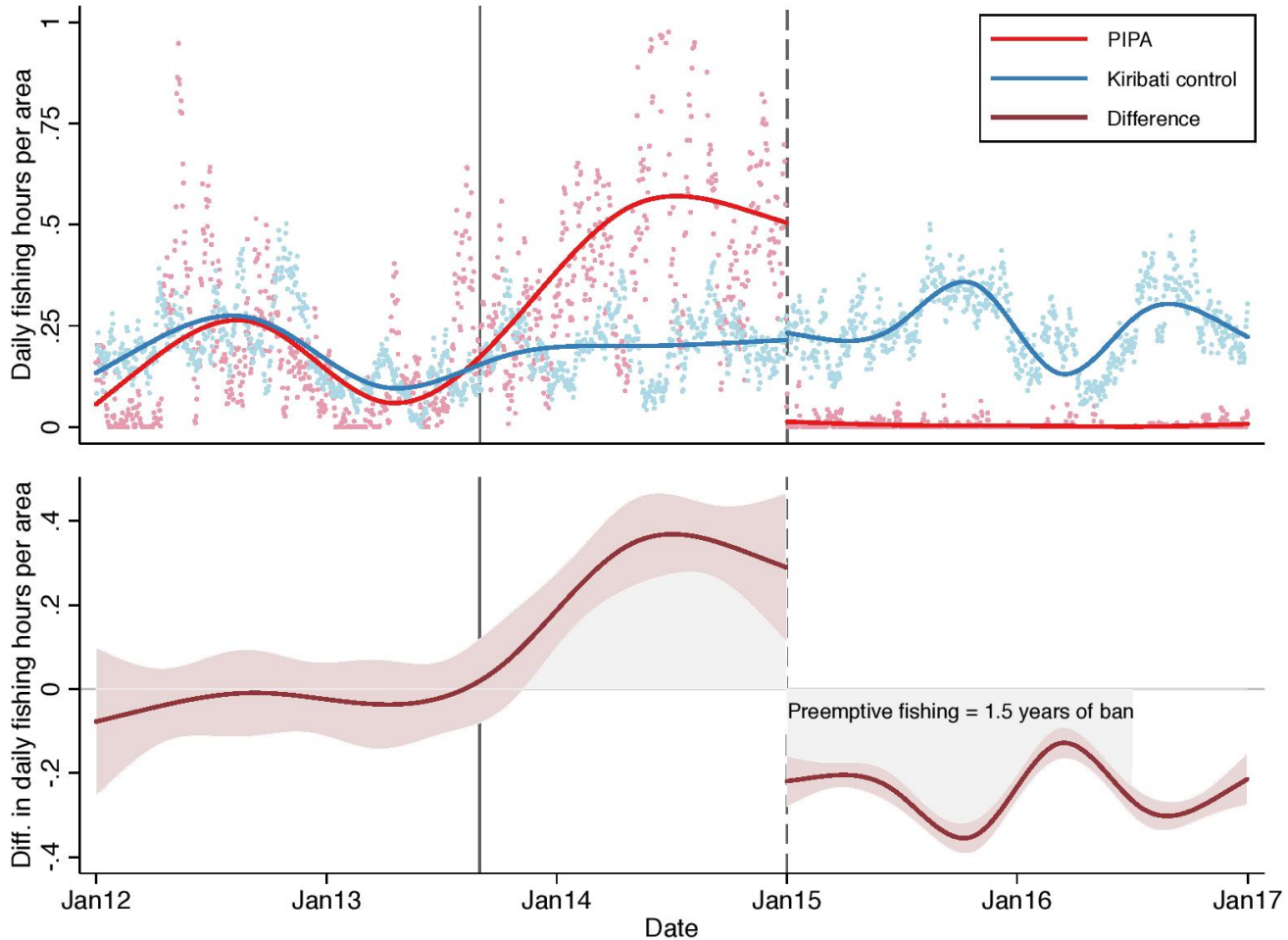
## Natural Experiment

Measure the causal effect of the fishing ban by comparing fishing effort in treatment and control regions, before-and-after PIPA.

- A *difference-in-differences* comparison.
- **Assumption:** *Parallel trends*. If we believe this assumption, then the observed change supports a causal interpretation. If not, then the change could reflect other factors and thus fail to isolate the causal effect of the ban.

# Example: *Blue Paradox*

## Results



# Example: *Blue Paradox*

## Discussion

Results provide causal evidence that commercial fishers engage in preemptive behavior in response to conservation policy changes.

Results are *consistent* with economic theory, but *cannot prove* that the theory is correct.

- **Science cannot prove anything.**
- Science can **falsify or reject** existing hypotheses or **corroborate** existing evidence.

Also...the causal statement rests on a critical assumption.

- Cannot prove that the assumption is true, but can falsify it.
- Failure to falsify  $\neq$  assumption is true.

# Econometrics

An applied econometrician<sup>†</sup> needs a solid grasp on (at least) three areas:

1. The **theory** underlying econometrics (assumptions, results, strengths, weaknesses).
2. How to **apply theoretical methods** to actual data.
3. Efficient methods for **working with data**—cleaning, aggregating, joining, visualizing.

**This course** aims to deepen your knowledge in each of these three areas.

- 1: As before.
- 2–3: **Stata**

[†]: *Applied econometrician* = Practitioner of econometrics, e.g., analyst, consultant, data scientist.

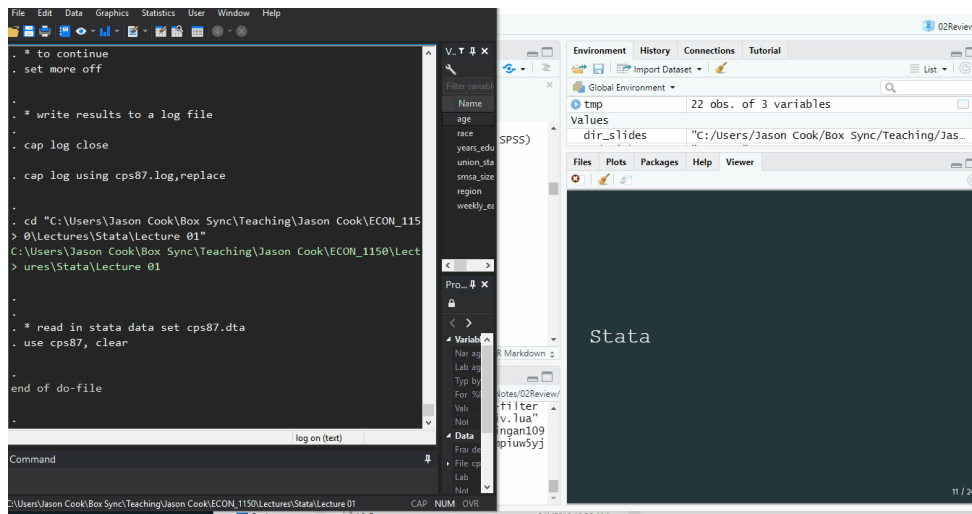


# Stata

- There are several statistical packages useful for econometrics, we will use Stata
- Concepts learned easily translate to other packages (e.g., R, SAS, SPSS)
- **Useful Resources:**
  - [Stata FAQ](#)
  - [The Stata listserv](#)
  - [UCLA's resources for learning Stata](#)
- **In Stata:** CPS data
  - Navigating Stata
  - Help files
  - Do-files

# Stata - Do File

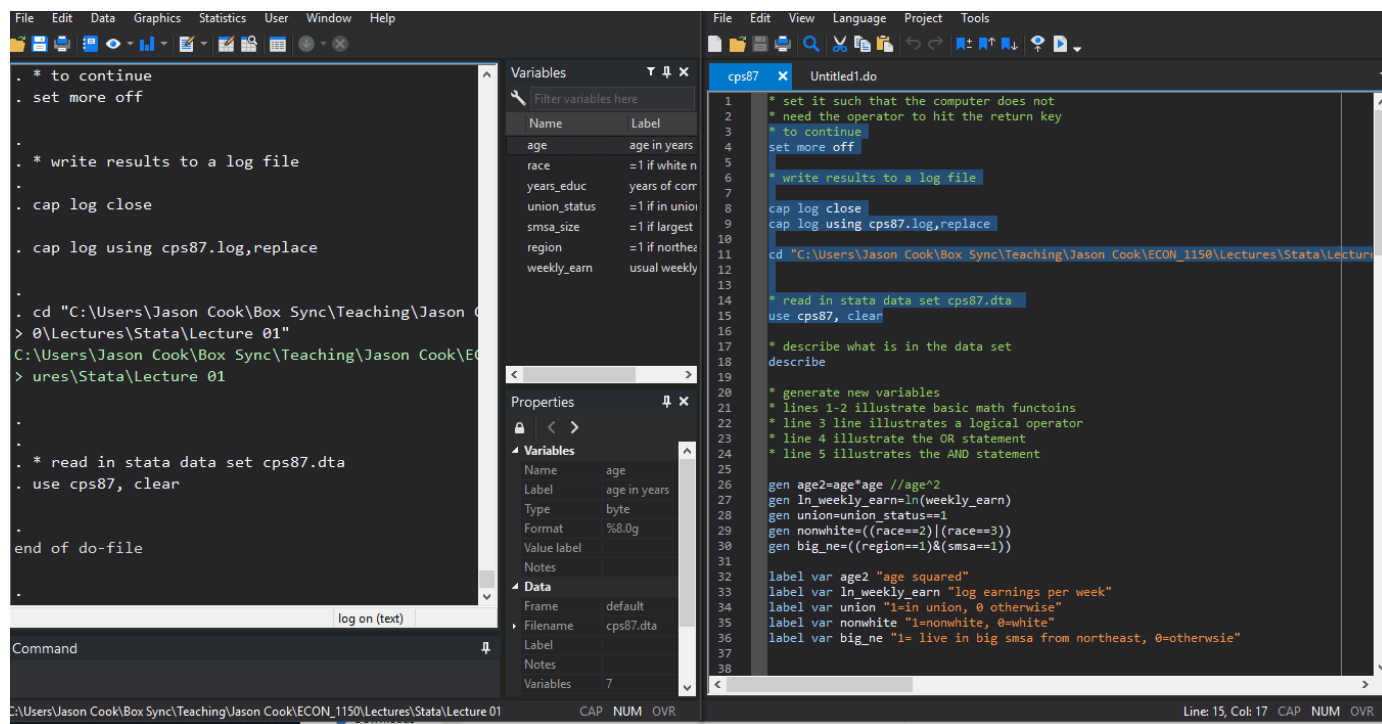
- Everything you **do** in Stata should be recorded in a **do file**
- The do file should:
  - Begin by opening raw data (source data unedited by user)
  - Perform data cleaning
  - Run analyses and save output
- **Never save over raw data**
- Comment by starting lines with \* to organize code



Opening a do file

# Stata - help

- Use the help feature liberally
  - Explains what a command does
  - At the end gives examples of using the syntax
  - Sometimes even has video explanations



**Looking up help file for summarize command**



# Stata - Basic Operations

- **Generate new variables:** `gen age2=age*age`
  - Create a variable (gen) called age2 and assign (=) to it the value of age times age (aka  $\text{age}^2$ )
- **Natural Log:** `gen ln_weekly_earn=ln(weekly_earn)`
- **Binary Variable:** `gen union=union_status=1*`
  - Create a variable called union that equals 1 if union\_status equals 1, otherwise assign union\_status to be 0

[\*] = "assign"; = "compare"

# Stata - Basic Operations

- **Logical Conditions (OR):** `gen nonwhite=race==2 | race==3`
  - Create a variable called `nonwhite` that equals 1 if `race` equals 2 *OR* (`|`) 3, otherwise assign `nonwhite` to be 0
- **Logical Conditions (AND):** `gen big\_ne=region==1 & smsa==1`
  - Create a variable called `big_ne` that equals 1 if `region` equals 1 *AND* (`&`) `smsa` equals 1, otherwise assign `big_ne` to be zero

# Stata - Summary Statistics

## `summarize, detail`

- Provides summary statistics of given variable(s), (here, `age`)
- `, detail` option that tells Stata to provide additional info like quantiles

	<b>Age</b>	<b>Education</b>
Mean	37.97	13.16
Median	36	12
Variance	124.4	7.81

# Stata - Summary Stats by Categories

```
bysort race: sum weekly_earn
```

- Runs `sum` command separately for each value of `race`
- Here `race=1` is white, `race=2` is black, and `race=3` is Hispanic

Race	Mean Earnings
White	506.5
Black	383.1
Hispanic	368.6

# Stata - if Statements

- Add `if` conditions after command to run on subset of data

The screenshot displays the Stata software interface with three main windows:

- Do-file Editor (Untitled1.do):** Contains Stata commands for data analysis, including sorting by race, summarizing weekly earnings, tabulating frequencies for race and smsa, and creating a histogram of age.
- Command Window:** Shows the command `log on (text)`.
- Results Window:** Displays the output of the `sum` command, showing summary statistics for three groups defined by the `if` statement.

**Results Window Output:**

	1	2	3	Total	
	1	2,806	1,349	842	4,997
		56.15	27.00	16.85	100.00
		38.46	18.89	15.39	25.10
	2	1,501	1,742	1,592	4,835
		31.04	36.03	32.93	100.00
		20.58	24.40	29.10	24.29
	3	1,501	2,542	1,904	5,947
		25.24	42.74	32.02	100.00

**Do-file Editor Content:**

```
52 sum weekly_earn age, detail
53
54 * to get means across different subgroups
55 * sample, first sort the data, then generate
56 * summary statistics by subgroup
57
58 sort race
59 by race: sum weekly_earn
60
61 * get weekly earnings for only those with
62 * high school education
63 sum weekly_earn if years_educ>=12
64
65 * get frequencies of discrete variables
66 tabulate race
67
68 * get two-way table of frequencies
69 tabulate region smsa, row column
70
71 *Histogram of age and plot line at average
72 sum age,d
73 histogram age, normal xtitle(Age) xline(`
74 graph export age_hist.png,replace as(png)
75
76 *****
77
78
79
80
81 * one-sample ttest
```

# Stata - tabulate Command

- Provides counts of observations in each unique value of variable

The screenshot displays the Stata software interface. The main window shows the results of the 'tabulate' command for the variable 'weekly\_earn'. The results are presented in three tables, one for each race (2 and 3). The first table shows the overall distribution of 'weekly\_earn' for race 2. The second table shows the distribution for race 3. The third table shows the distribution for race 4. The command window shows the command 'tabulate race' being executed. The command window also shows the command 'sum weekly\_earn if years\_educ>=12' being executed. The command window also shows the command 'sum age,d' being executed. The command window also shows the command 'histogram age, normal xtitle(Age) xline(' being executed. The command window also shows the command 'graph export age\_hist.png,replace as(png)' being executed. The command window also shows the command '\*\*\*\*\*' being executed. The command window also shows the command '\* one-sample ttest' being executed.

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	17,103	506.4874	237.2567	60	999

-> race = 2

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,642	383.095	196.2224	90	999

-> race = 3

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,161	368.5512	200.6758	66	999

Command: log on (text)

```
sum weekly_earn age, detail
* to get means across different subgroups
* sample, first sort the data, then gener
* summary statistics by subgroup
sort race
by race: sum weekly_earn
* get weekly earnings for only those with
* high school education
sum weekly_earn if years_educ>=12
* get frequencies of discrete variables
tabulate race
* get two-way table of frequencies
tabulate region smsa, row column
*Histogram of age and plot line at averag
sum age,d
histogram age, normal xtitle(Age) xline(
graph export age_hist.png,replace as(png)
*****
* one-sample ttest
```

# Stata - Figures

- Easiest way to make figures in Stata is using GUI in menu under Graphics → Twoway Graph

The screenshot displays the Stata software interface with three main panels: a command window on the left, a variables list in the center, and a command editor on the right.

**Command Window (Left):**

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,642	383.095	196.2224	90	999

-> race = 3

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,161	368.5512	200.6758	66	999

. sum weekly\_earn if race==3

Variable	Obs	Mean	Std. Dev.	Min	Max
weekly_earn	1,161	368.5512	200.6758	66	999

log on (text)

Command

C:\Users\Jason Cook\Box Sync\Teaching\Jason Cook\ECON\_1150\Lectures\Stata\Lecture 01

**Variables List (Center):**

Name	Label
age	age
race	=1
years_educ	years_educ
union_status	=1
smsa_size	=1
region	=1
weekly_earn	ust

**Command Editor (Right):**

```
52 sum weekly_earn age, detail
53
54 * to get means across different subgroups
55 * sample, first sort the data, then gener
56 * summary statistics by subgroup
57
58 sort race
59 by race: sum weekly_earn
60
61 * get weekly earnings for only those with
62 * high school education
63 sum weekly_earn if years_educ>=12
64
65 * get frequencies of discrete variables
66 tabulate race
67
68 * get two-way table of frequencies
69 tabulate region smsa, row column
70
71 *Histogram of age and plot line at averag
72 sum age,d
73 histogram age, normal xtitle(Age) xline(`
74 graph export age_hist.png,replace as(png)
75
76 *****
77
78
79
80
81 * one-sample ttest
```

Line: 74, Col: 15 CAP NUM OVR

# Stata - Documenting code

- Organize code into sections
- Use `*` to comment entire lines of code or `\\` to comment out everything following it on the same line
- This can be used to write notes about what the chunk of code is doing

## Example

```
*Create new variables
```

```
gen age2 = age^2 \\other comment here
```

In the example above, only the `gen age2=age^2` part of the code is executed, Stata will skip over the rest of the text



# Questions?