

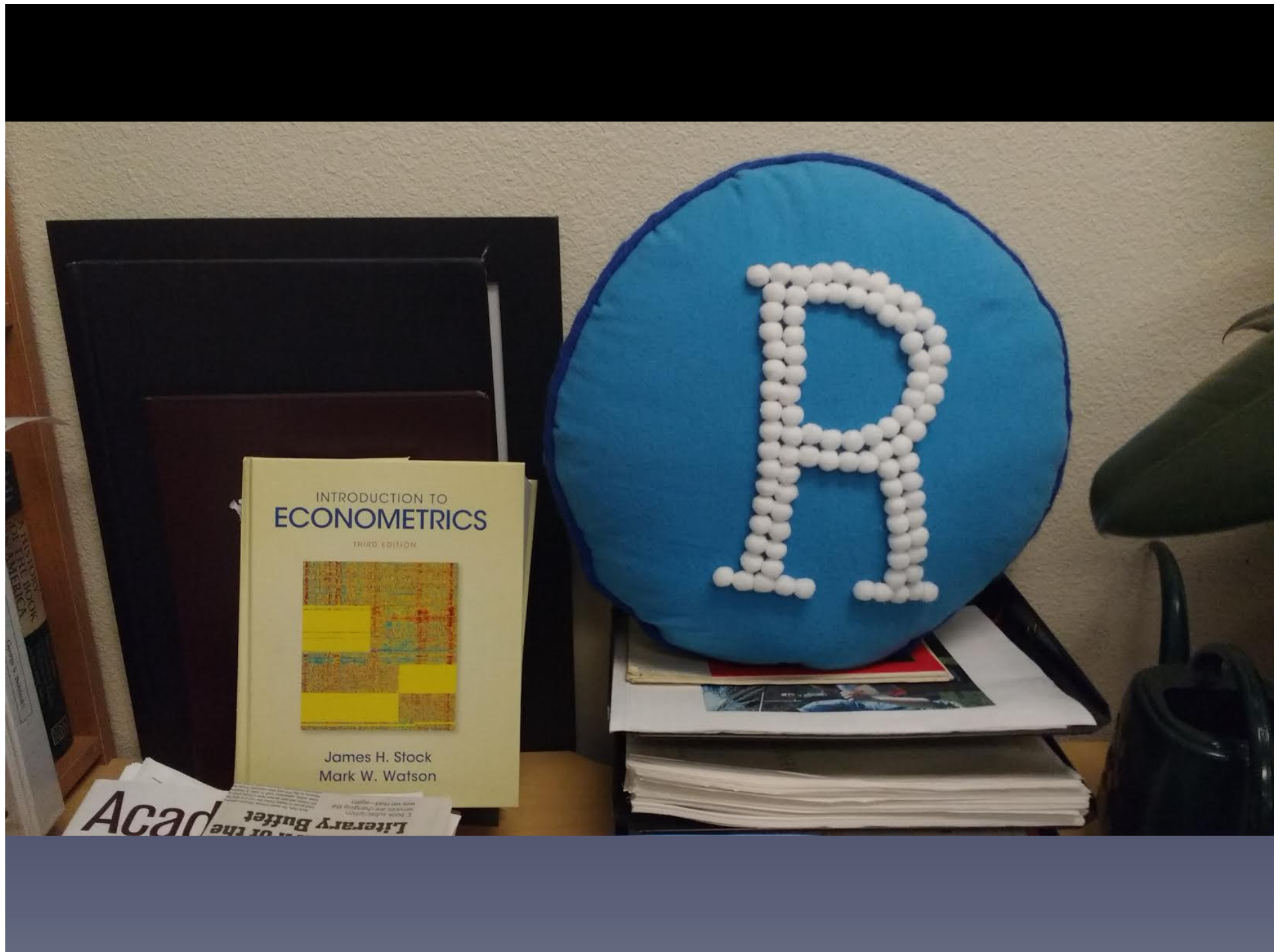
# Teaching Undergrad Econometrics with R

Bill Sundstrom

Dept. of Economics, Santa Clara U.

[wsundstrom@scu.edu](mailto:wsundstrom@scu.edu)

BARUG February 9, 2016



# Context

- Large undergrad economics major
- Intro stats no longer sufficient
  - Modern economics is highly empirical
  - Data analysis skills are highly valued

# What are we trying to do?

- New required econometrics course:
  - Regression and causal explanation
  - Hands-on data analysis using R
- Why R?

# Challenges

- Many students find stats hard
  - They have taken just one intro stats course
- Thinking about causal identification is hard
- Digital natives are also digital *naifs*

Hi Professor,

Something terrible happened just now. I normally dont save often in Rstudio because it seems to have no problems staying open, so I hadnt saved the code for my final essay and now it suddenly crashed. Theres just that rainbow wheel and it says application not responding. Do you know of anything I should/could do? I'm totally fucked

Thanks!

# Making it work

- Standard 10-week lecture course (two meetings per week)
- Accompanying weekly R lab course

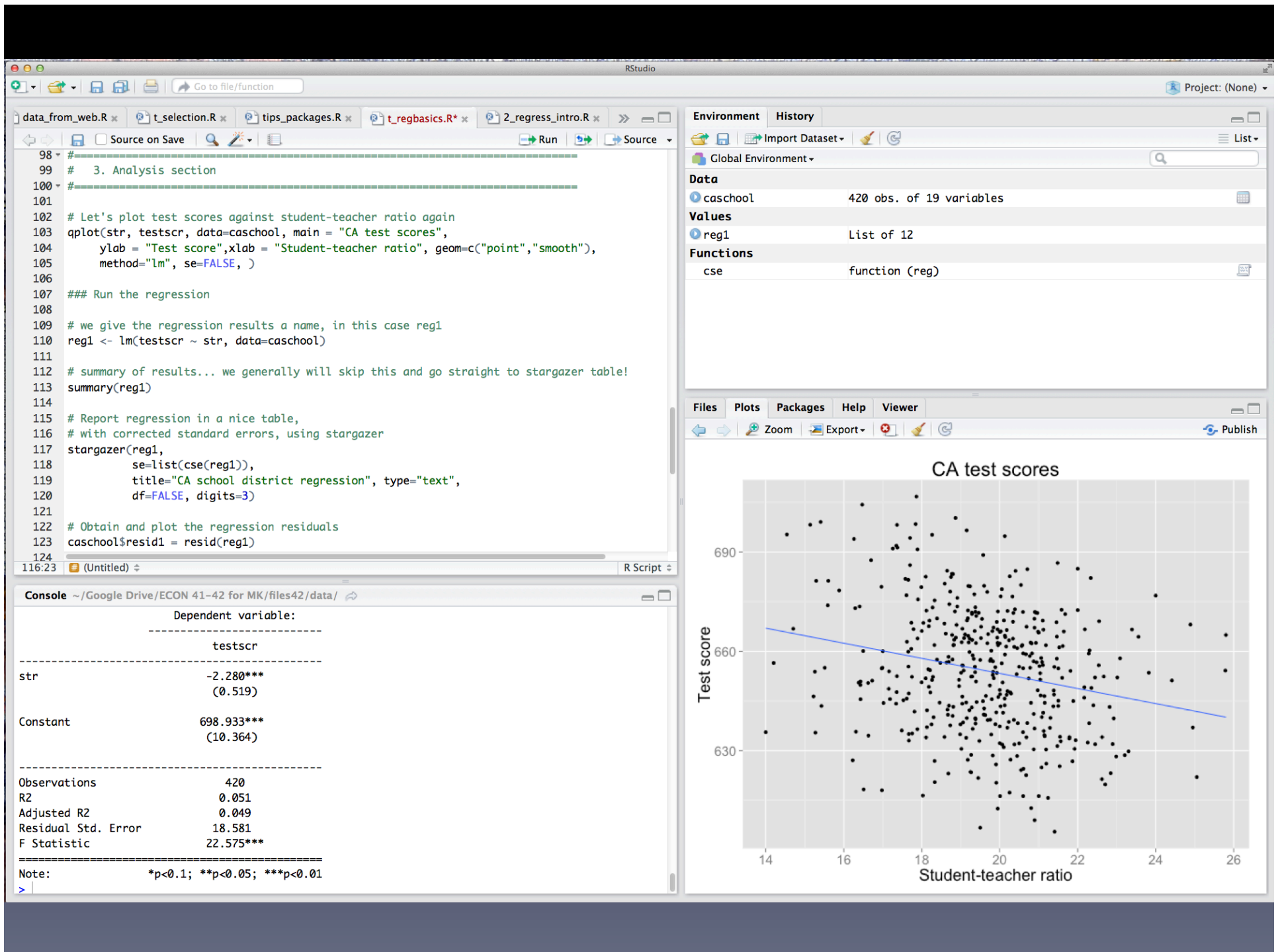
# In the lab

- We *do not* try to teach coding from scratch
- We *do* insist that students run R from scripts, from soup to nuts: replicability
- We provide sample scripts and tutorials
- Trouble-shooting for the first two weeks, then running regressions and interpreting results
- Culminates in data analysis project

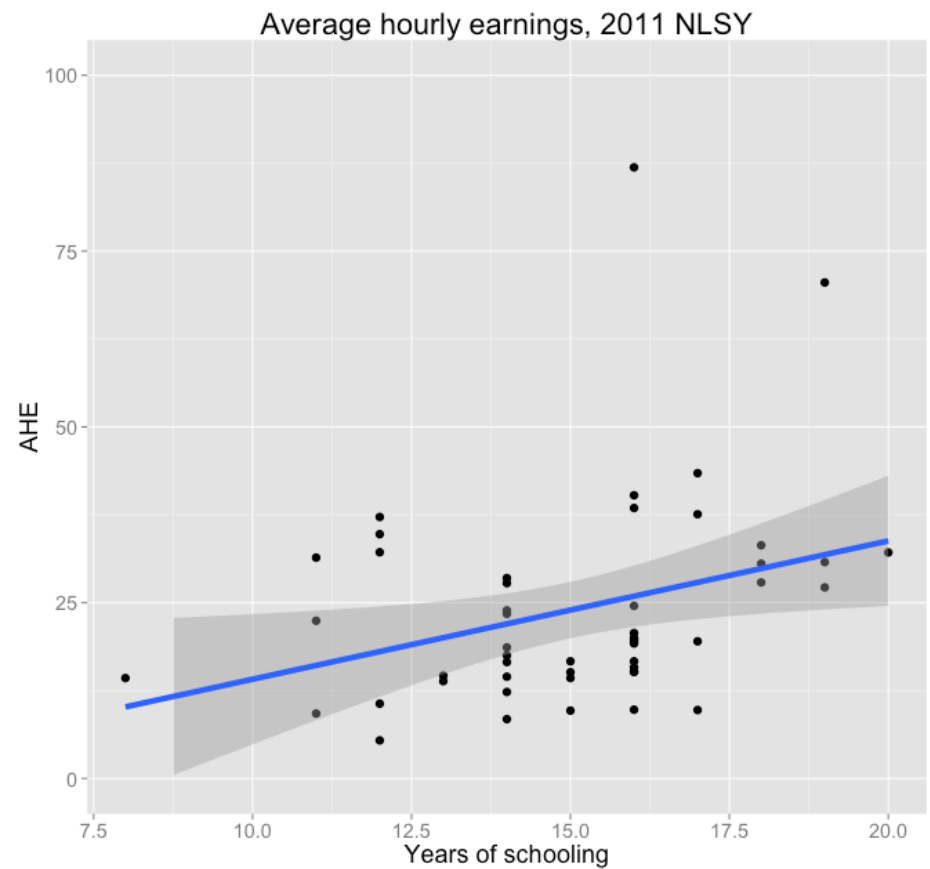
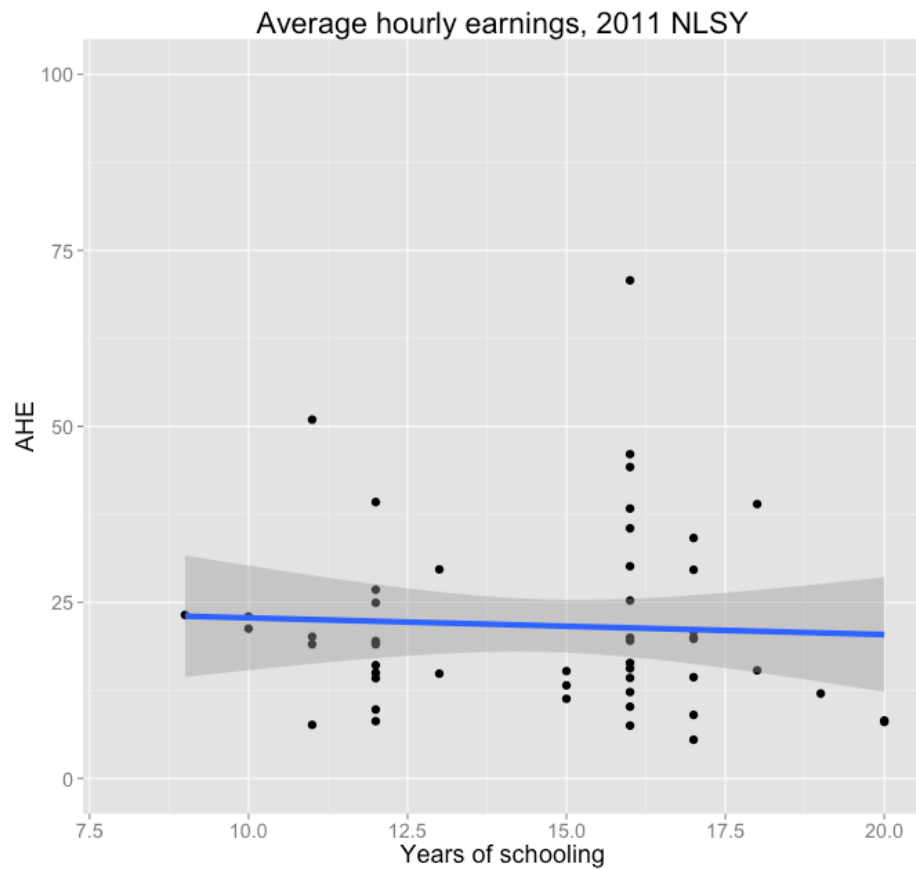


# A tutorial script in action

- Effect of class size on student performance
- Plot the data
- Run the regression
- Nice regression tables with `stargazer`



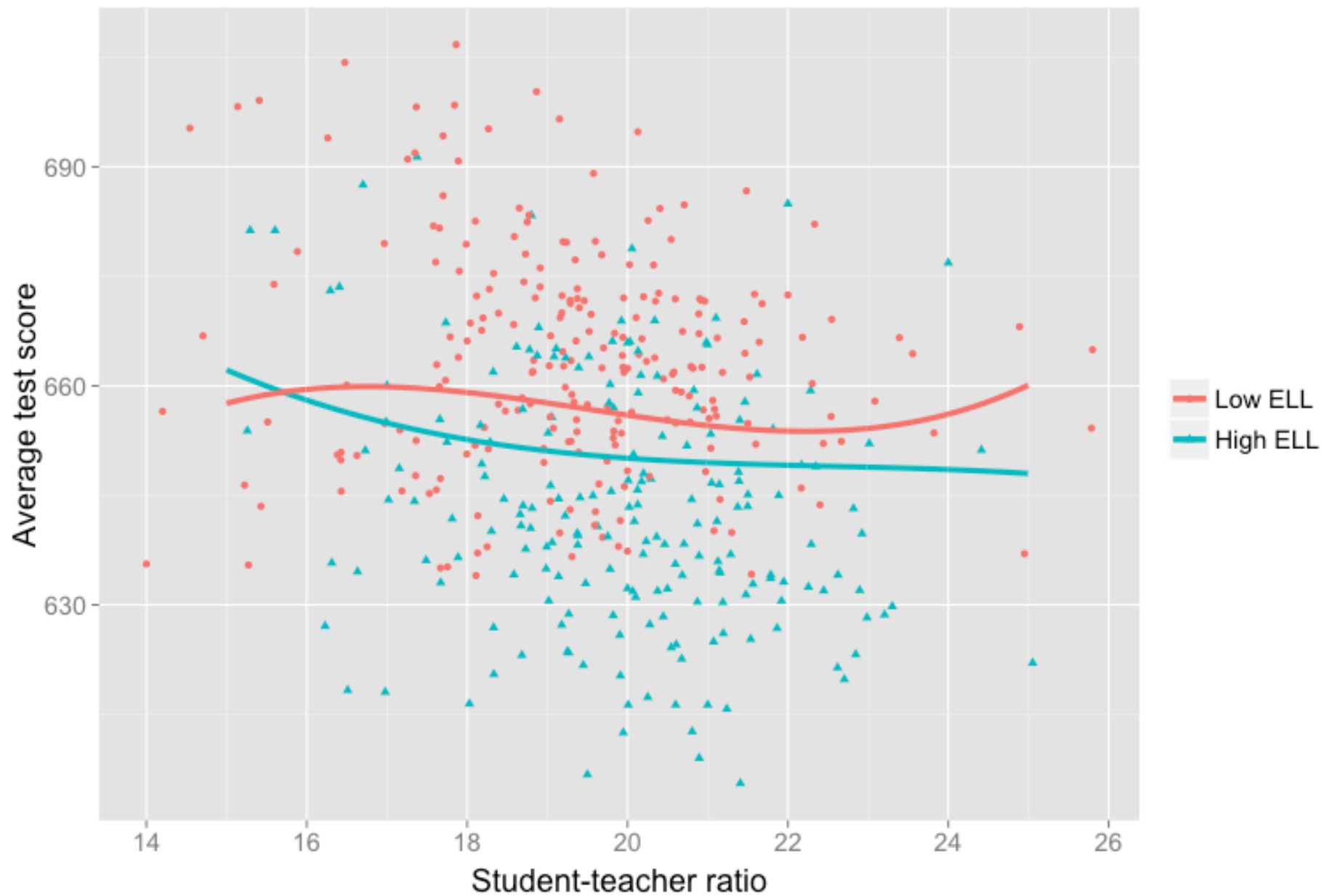
# $\beta^{\wedge}$ is a random variable



# Multiple regression

- Control variables as partial solution to omitted variable bias
- Specification
  - Using interactions to facilitate comparisons
  - Nonlinear specifications
- Thinking about effect size, not just statistical significance

## Test scores and class size



# Data analysis project

- Replicate and reevaluate a classic economics article using up-to-date data
  - Barry Chiswick, “The Effect of Americanization on the Earnings of Foreign-Born Men” (JPE, 1978)

TABLE 2

REGRESSION ANALYSIS OF EARNINGS FOR NATIVE- AND  
FOREIGN-BORN ADULT WHITE MEN, 1970

	NATIVE BORN	NATIVE AND FOREIGN BORN			FOREIGN BORN
	(1)	(2)	(3)	(4)	(5)
EDUC .....	.07154 (53.78)	.07058 (55.68)	.07004 (55.18)	.07164 (54.11)	.05740 (12.93)
<i>T</i> .....	.03167 (22.99)	.03050 (22.86)	.03071 (22.99)	.03097 (23.10)	.02028 (3.47)
<i>T</i> <sup>2</sup> .....	-.00052 (-20.77)	-.00049 (-20.45)	-.00050 (-20.78)	-.00051 (-20.93)	-.00031 (-3.18)
LN WW .....	1.10335 (81.75)	1.10326 (84.78)	1.10169 (84.70)	1.10111 (84.67)	1.07151 (21.97)
RURALEQ1 ...	-.17222 (-20.28)	-.16970 (-20.25)	-.17080 (-20.39)	-.16915 (-20.18)	-.05821 (-1.13)
SOUTHEQ1 ...	-.12090 (-14.17)	-.12620 (-15.01)	-.12530 (-14.91)	-.12389 (-14.74)	-.21587 (-4.38)
NOTMSP .....	-.30647 (-27.76)	-.31078 (-28.97)	-.30947 (-28.86)	-.30874 (-28.79)	-.34498 (-7.66)
FOR .....	*	.02951 (1.75)	-.16359 (-1.32)	.00990 (0.18)	*
(FOR) (YSM)	*	*	.01461 (3.98)	.01555 (4.23)	.01500 (3.87)
(FOR) (YSM2)	*	*	-.00016 (-2.47)	-.00018 (-2.79)	-.00019 (-2.82)
(FOR) (EDUC)	*	*		-.01619 (-4.23)	*
CONSTANT ...	-1.03646	-1.01537	-1.00016	-1.02156	-.78891
Observations ( <i>N</i> ) .....	34,321	36,245	36,245	36,245	1,924
<i>R</i> .....	.55423	.55455	.55533	.55564	.58194
<i>R</i> <sup>2</sup> .....	.30717	.30753	.30839	.30873	.33866
Standard error	.70900	.71008	.70966	.70949	.71676

SOURCE.—U.S. Bureau of the Census 1972.

NOTE.—*t*-ratios in parentheses; dependent variable: natural logarithm of earnings in hundreds of dollars.

\* Variable not entered.

Chiswick Table 2

	Dependent variable:				
	US-born (1)	All (2)	lnearn All (3)	All (4)	For-born (5)
educ_years	0.1216*** (0.0032)	0.1130*** (0.0027)	0.1128*** (0.0028)	0.1222*** (0.0032)	0.0895*** (0.0057)
exper	0.0476*** (0.0030)	0.0433*** (0.0027)	0.0434*** (0.0027)	0.0440*** (0.0027)	0.0153* (0.0079)
I(exper2)	-0.0008*** (0.0001)	-0.0007*** (0.0001)	-0.0007*** (0.0001)	-0.0007*** (0.0001)	-0.0002 (0.0001)
ftfy	1.2034*** (0.0249)	1.1942*** (0.0233)	1.1936*** (0.0233)	1.1903*** (0.0233)	1.0651*** (0.0649)
south	-0.0364*** (0.0141)	-0.0412*** (0.0134)	-0.0409*** (0.0134)	-0.0416*** (0.0133)	-0.0820** (0.0403)
notmsp	-0.3124*** (0.0154)	-0.3064*** (0.0145)	-0.3063*** (0.0145)	-0.3036*** (0.0145)	-0.2168*** (0.0442)
forborn		-0.0339 (0.0219)	-0.0891 (0.0638)	0.3923*** (0.1020)	
forysm			0.0031 (0.0052)	-0.0004 (0.0052)	0.0043 (0.0054)
I(forysm2)			-0.00002 (0.0001)	0.00005 (0.0001)	-0.00002 (0.0001)
educ_years : forborn				-0.0352*** (0.0058)	
Constant	7.5147*** (0.0629)	7.6839*** (0.0554)	7.6877*** (0.0556)	7.5521*** (0.0612)	8.2743*** (0.1460)



# A work in progress

- Many students don't run the tutorials
- Need more emphasis on data visualization
- Teaching with simulations?
- We give them data that is *way too tidy*

# Questions for you...

- What are the most important things for an informed econ/ business major to learn about data analysis?
- We could use good examples/ projects from business, data analytics
- How can good students interested in data analytics start networking, build a reputation?
- Can you use our students? Internships, jobs

# Thank you

- Co-conspiratoR: Michael Kevane, Economics, Santa Clara University
- My email: [wsundstrom@scu.edu](mailto:wsundstrom@scu.edu)
- Guide to R online version (under revision):  
<http://rpubs.com/wsundstrom/home>