

Introduction to Stata

Roxanne Connelly
University of York

Kevin Ralston
York St John University

9th July, 2019
University of York



UNIVERSITY
of York



Hello!



Outline

1000 – 1015	Welcome and Introduction to the Stata Environment
1015 – 1100	Hands On Workshop
1100 – 1115	The Workflow and Data Management
1115 – 1200	Hands on Workshop
1200 – 1300	Lunch
1300 – 1315	Regression
1315 – 1415	Hands on Workshop
1415 – 1430	Break
1430 – 1445	Publication Ready Tables
1445 – 1545	Hands on Workshop
1545 – 1600	Questions and Close

Introduction to the Stata Environment

Products

Purchase

Training

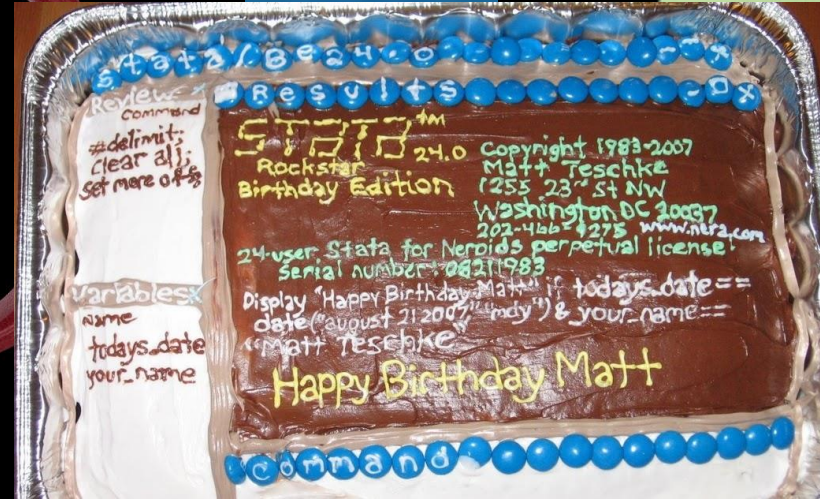
Support

A beautiful solution.



No matter what your field, put Stata to work for you.

What can Stata do for you?

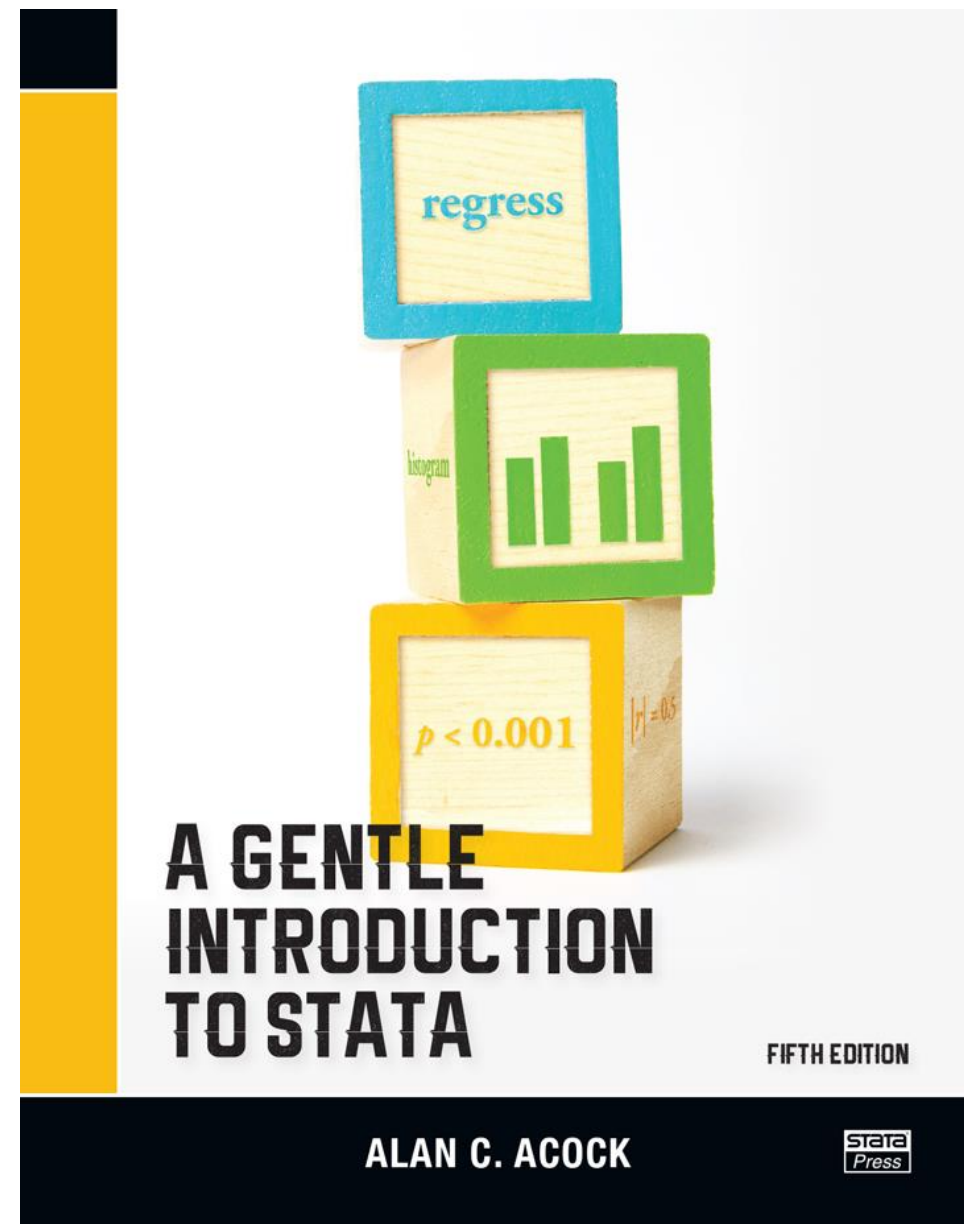
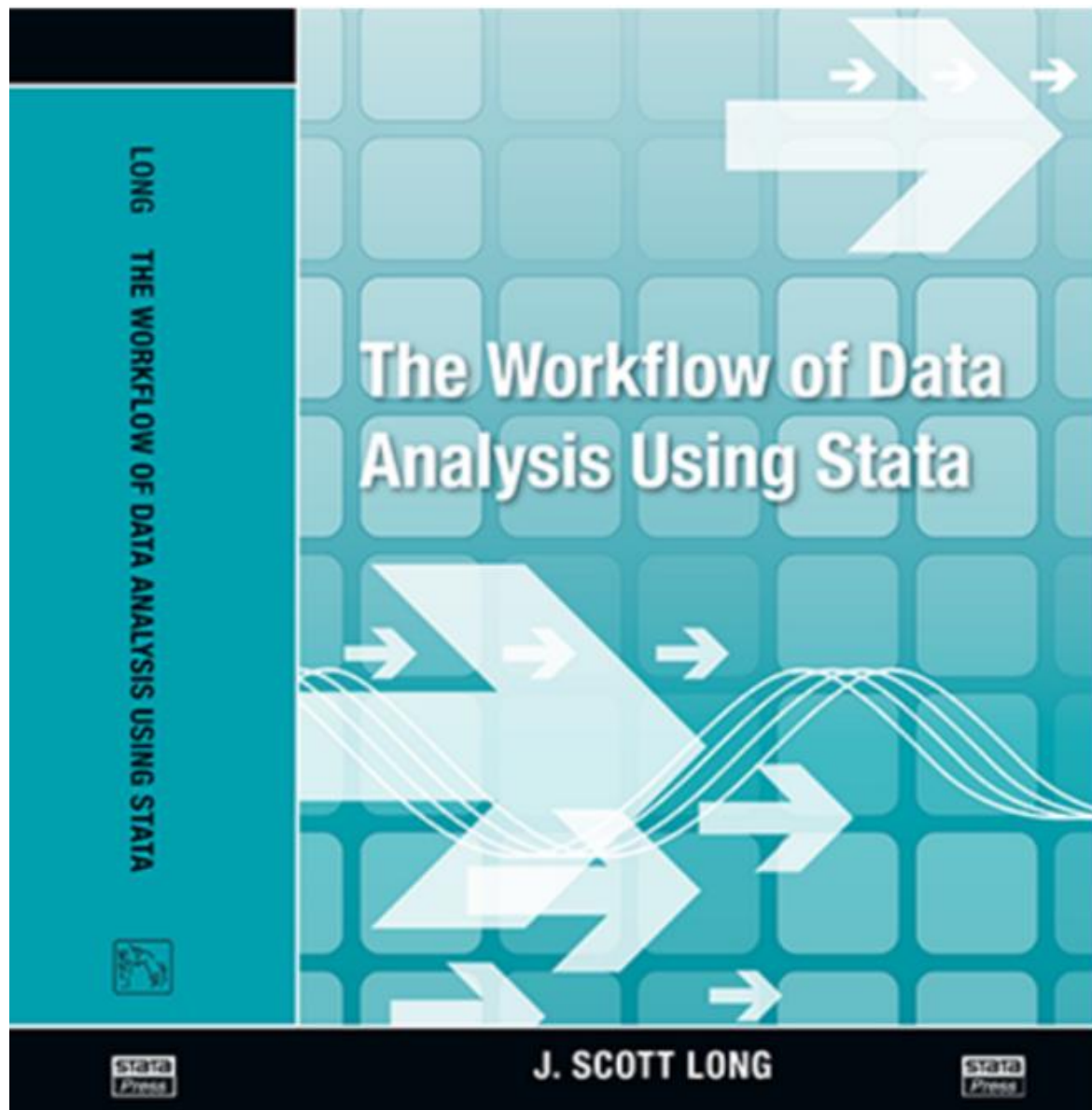


Why Stata?

- Excellent documentation and help files.
- A helpful forum (www.statalist.org).
- Many good quality and accessible training resources:
 - Stata (<http://www.stata.com/training/>)
 - UCLA (<http://stats.idre.ucla.edu/stata/>)
 - Timberlake (<http://www.timberlake.co.uk/training>)
 - Princeton (<http://data.princeton.edu/stata/>)

Why Stata?

- Data Enabling (i.e. preparing complex survey data for analyses)
- Exploratory Data Analysis
- Comprehensive data analysis
 - Models in the regression family (glm)
 - Selection models
 - Multilevel models
 - Event history models
 - Longitudinal models (e.g. panel models)





IBM Analytics

IBM SPSS Software

Predict with confidence what will happen next so you can make smarter decisions for your organization



Featured products



Free SPSS trials

Secure | https://www.sas.com/en_gb/home.html



Sign In | United Kingdom | Cx

[Products & Solutions](#) [Industries](#) [Support](#) [Learn SAS](#) [Customer Success](#) [Partners](#)

Analytics software & solutions



[Home]

Download

[CRAN](#)

R Project

[About R](#)
[Logo](#)
[Contributors](#)
[What's New?](#)
[Reporting Bugs](#)
[Development Site](#)
[Conferences](#)
[Search](#)

R Foundation

[Foundation](#)
[Board](#)
[Members](#)
[Donors](#)
[Donate](#)

Help With R

[Getting Help](#)

Documentation

[Manuals](#)
[FAQs](#)
[The R Journal](#)
[Books](#)
[Certification](#)
[Other](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred CRAN mirror.

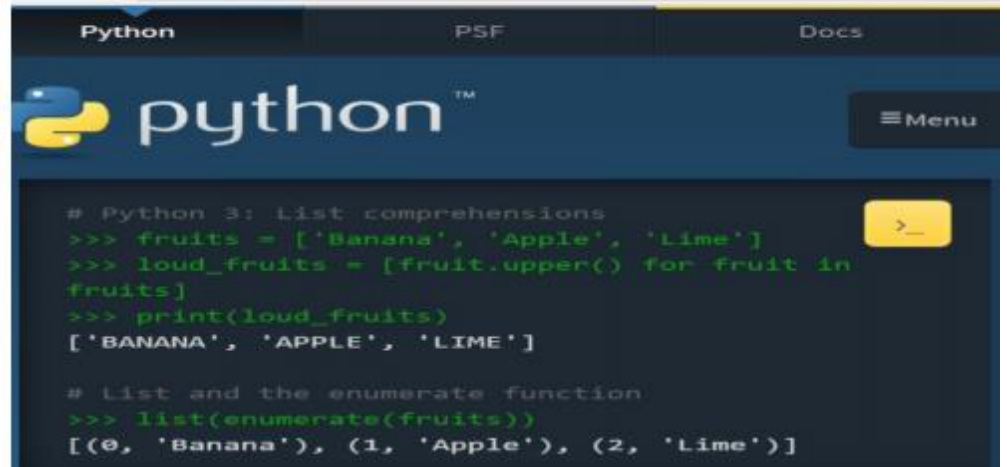
If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

R version 3.3.3 (Another Canoe) has been released on Monday 2017-03-06.

- **useR! 2017** (July 4 - 7 in Brussels) has opened registration and more at <http://user2017.brussels/>
- Tomas Kalibera has joined the R core team.
- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- **The R Journal Volume 8/1** is available.
- The **useR! 2017** conference will take place in Brussels, July 4 - 7, 2017.
- **R version 3.2.6 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, have taken place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.

Python Software Foundation [US] | <https://www.python.org>





IBM Analytics

IBM SPSS Sof

Predict with confidence what will happen
can make smarter decisions for your



Featured products



Free

Secure | https://www.sas.com/en_gb/home.html



Products & Solutions Industries Support Learn SAS

Analytics software & solutions



[Home]

Download

CRAN

R Project

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

VS

tion 3.3.3 (Another Canoe) has been released on Monday 2017-03-06.

useR! 2017 (July 4 - 7 in Brussels) has opened registration and more at <http://user2017.brussels/>

has Kalibera has joined the R core team.

R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, has Kalibera, and Balasubramanian Narasimhan.

R Journal Volume 8/1 is available.

useR! 2017 conference will take place in Brussels, July 4 - 7, 2017.

version 3.2.6 (Very, Very Secure Dishes) has been released on 2016-04-14. This is a rebadging the quick-fix release 3.2.4-revised.

ice XQuartz users (Mac OS X) A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.

R Logo is available for download in high-resolution PNG or SVG formats.

useR! 2016, have taken place at Stanford University, CA, USA, June 27 - June 30, 2016.

R Journal Volume 7/2 is available.

version 3.2.3 (Wooden Christmas-Tree) has been released on 2015-12-10.

version 3.1.3 (Smooth Sidewalk) has been released on 2015-03-09.

on Software Foundation [US] | <https://www.python.org>

PSF

Docs

thon™

Menu

```
# List comprehensions
fruits = ['Banana', 'Apple', 'Lime']
loud_fruits = [fruit.upper() for fruit in
```

```
fruits]
>>> print(loud_fruits)
['BANANA', 'APPLE', 'LIME']
```

```
# List and the enumerate function
>>> list(enumerate(fruits))
[(0, 'Banana'), (1, 'Apple'), (2, 'Lime')]
```




Stata/SE 14.2

File Edit Data Graphics Statistics User Window Help

Review

Filter commands here

Command _rc

There are no items to show.

Statistics/Data Analysis

Special Edition

14.2

Copyright 1985-2015 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Unlimited-user Stata network license expires 30 Apr 2018:
Serial number: 401409002392
Licensed to: IT Services
University of Warwick

Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set_maxvar.

running C:\Program Files (x86)\Stata14\sysprofile.do ...

Variables

Filter variables here

Name Label

There are no items to show.

Command

Properties

Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

\\CANTUS\User43\u\1474300\Documents

111

12:38

05/10/2017

Stata/SE 15.1

FileEditDataGraphicsStatisticsUserWindowHelp

Filter commands here

Command

_rc

There are no items to show.

Statistics/Data Analysis

15.1

Copyright 1985-2017 StataCorp LLC

StataCorp

4905 Lakeway Drive

College Station, Texas 77845 USA

800-STATA-PC

979-696-4600

979-696-4601 (fax)

<http://www.stata.com>

stata@stata.com

Special Edition

100-user Stata network perpetual license:

Serial number: 401506202441

Licensed to: IT Services

University of York

Notes:

1. Unicode is supported; see [help unicode_advice](#).

2. Maximum number of variables is set to 32767; see [help set_maxvar](#).

running C:\Program Files (x86)\Stata15\sysprofile.do ...

.

Command

Variables

Filter variables here

Name	Label
There are no items to show.	

Properties

Variables

Name	
Label	
Type	
Format	

\\userfs\rc1531\w2k

CAPNUMOVR

09:39

08/07/2019

Stata/SE 14.2

File Edit Data Graphics Statistics User Window Help

Review

Filter commands here

Command _rc

There are no items to show.

Statistics/Data Analysis (R) 14.2 Copyright 1985-2015 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Special Edition

Unlimited-user Stata network license expires 30 Apr 2018:
Serial number: 401409002392
Licensed to: IT Services
University of Warwick

Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set_maxvar.

running C:\Program Files (x86)\Stata14\sysprofile.do ...

Variables

Filter variables here

Name Label

There are no items to show.

Do-file Editor - QS904Session1_20170922_rc_v2

File Edit View Project Tools

QS904Session1_20170922_rc_v2 X Untitled.do X

```
1 STOP
2
3 /*
4 */
5
6 QS904: RESEARCH DESIGN, WORKFLOW AND DATA MANAGEMENT
7
8 Lab 1: Getting Started in Stata
9
10
11
12 INTRODUCTION:
13
14 This file is sequential and must to run line by line. You can run this whole
15 file through in a matter of minutes, but you will not learn much by doing that.
16
17 I suggest that you save copy of this file, work through it slowly, and
18 annotate it to aid your learning.
19
20 Throughout the file there are tasks for you to do. I also recommend that you
21 try to repeat the operations in this file using different variables to
22 practice the commands.
23
24
25
26
27 LEARNING AIMS:
28
29 By the end of lab session 1 you should be able to:
30
31 1. Understand the benefits of keeping a well curated .do file.
32 2. Develop an organised filing structure.
33 3. Become familiar with Stata.
34 4. Open data from different sources and in different formats.
35 5. Explore data using basic commands (this may be revision for some).
36 6. Know where to find help.
37
38
39
40 */
41
42
43
44
45
46
47
```

Line: 1, Col: 1 CAP NUM OVR

Properties

Variables

Name	Label	Type	Format	Value label	Notes

Data

Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

\\CANTUS\User43\ud1474300\Documents

12:39 05/10/2017

No Drop Down Menus?

GUIs will leave you in a sticky mess!



Workshop Sessions

The best way to learn Stata is to use Stata!

There are 4 hands on workshop sessions.

- 20 Minutes Work
- 5 Minute Break
- 20 Minutes Work

Please ask for help!

You all have different backgrounds and different levels of expertise. You are not expected to finish all the lab materials. Please work at your own pace, and take time to practice the commands you are shown.

Workshop Sessions

Materials can be downloaded here:

[**https://tinyurl.com/y2jbm9tt**](https://tinyurl.com/y2jbm9tt)

Please save these materials on your H: drive or memory stick.

https://tinyurl.com/y2jbm9tt

Unstar 1 Fork 0

<> Code ! Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

One Data Introduction to Stata Workshop

Edit

18 commits

1 branch

0 releases

1 contributor



Branch: master

New pull request

Create new file

Upload files

Find File

Clone or download

RoxanneConnelly Delete Slides.pdf

Data

Add files via upload

Session1_Introduction

Delete Lab1

Session2_DataManagement

Delete Lab2

Session3_Regression

Delete Lab3

Session4_Tables&Graphs

Delete Lab4

README.md

Update README.md

2 minutes ago

2 months ago

Clone with HTTPS

Use SSH

Use Git or checkout with SVN using the web URL.

https://github.com/RoxanneConnelly/Intro

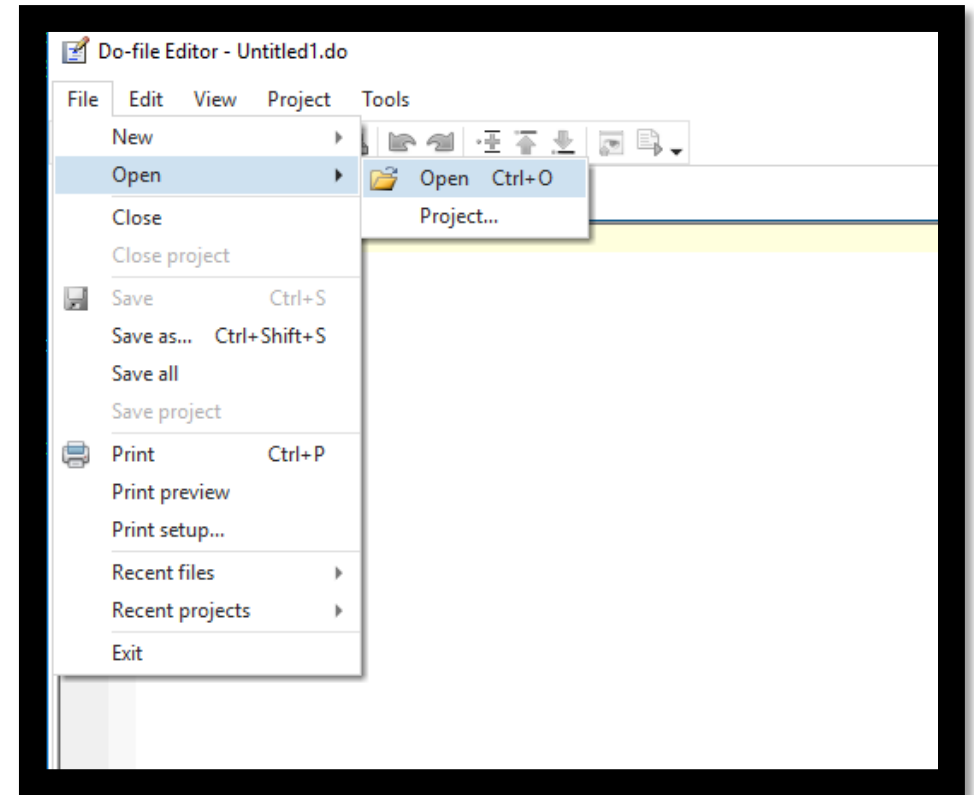
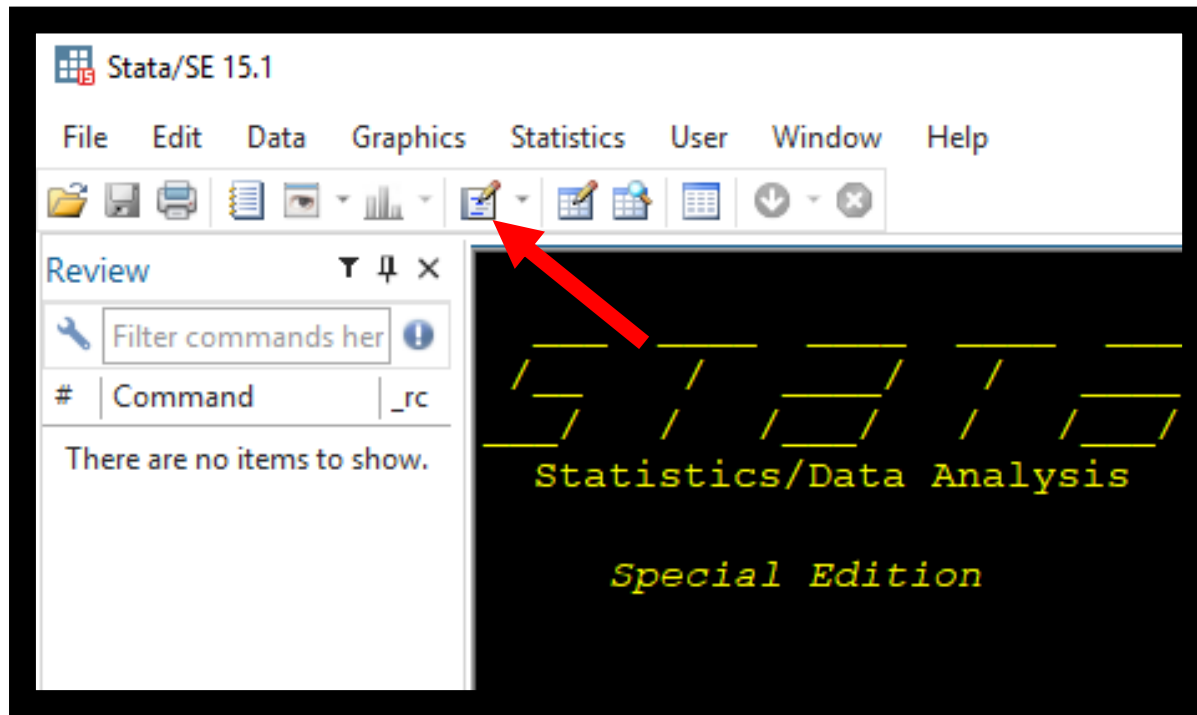


Open in Desktop

Download ZIP

Hands On Session 1

- Open Stata
- Open the Lab1 .do file



The Workflow and Data Management

The Workflow and Data Management

Data management is “the entire process of data analysis including planning and documenting your work, cleaning data and creating variables, producing and replicating statistical analyses, presenting findings, and archiving your work.”

(Long 2009: 1)

- Tasks associated with preparing and enhancing data for statistical analysis, such as recoding measures, constructing new variables and linking datasets.

Why the workflow is essential...

- Reproducibility
 - Results that others can reproduce are essential for good science.
 - The workflow is crucial for reproducibility.
- Effective
 - You want to make sure your answers are correct
 - With open science others will find your mistakes
- Efficient
 - You have limited time and resources
 - You also need to work carefully

Organisation

Have you ever lost a file?

Organisation

Have you ever lost a file?



Organisation

Have you ever lost a file?



Documentation

- It is always easier to document today than it is tomorrow!
- Corollary 1:
 - Nobody likes to write documentation
- Corollary 2:
 - Nobody ever regrets having written documentation

The Gayle-Lambert Constant

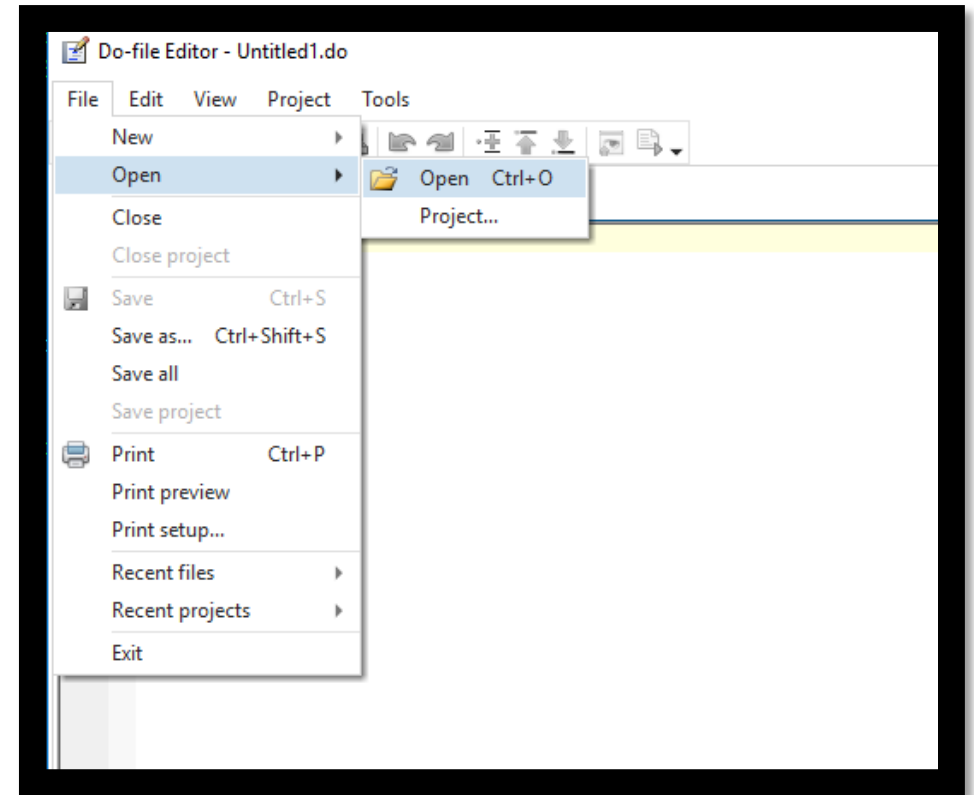
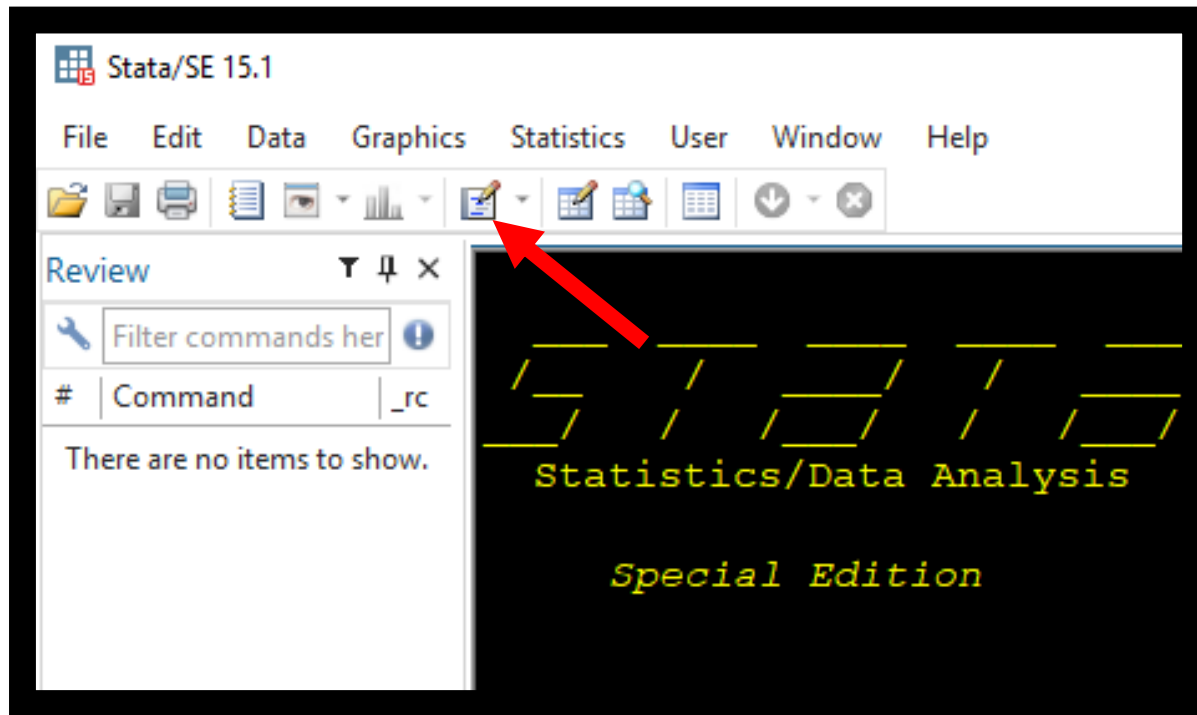


The Connelly-Ralston Constant



Hands On Session 2

- Open Stata
- Open the Lab2 .do file



Regression

Regression in Stata

a powerful **multivariate** technique

Type of outcome variable	Type of regression	Stata code
Scale	OLS regression	reg
Categories (2 only)	Logit, Probit, loglinear	logit probit
Nominal categories (>2)	M-logit	mlogit
Ordinal categories (>2)	O-Logit	ologit
Count variable	Poisson	poisson

- Clustered/hierarchical data
 - `meglm`
- Panel data
 - `xt` suite of commands
 - `xtset, i(pid) t(wave)`
- Event history data
 - `st` suite of commands
 - `stset t1, id(id) failure(died)`
- Structural equation models
 - `gsem`
 - `gsem (m1 m2 <- L1) (L2 -> m3 m4) (L3 -> m5 m6 m7) (L3 <- L1 L2)`
- Latent Class Models
 - `gsem (v1 v2 v3 v4 v5 <-), logit lclass(C 3)`

Regression

$$Y_i = \beta_0 + \beta_1 X_1 \dots \beta_k X_k + e_i$$

A research example using regression

Also testing for selection bias using a Heckman model

Do young people not in education employment and training (NEET) experience long term occupational scarring?

A longitudinal analysis over 20 years of follow up.

Ralston, K¹. Dawn Everington². Dibben, C³. Lambert, P⁴. and Feng, Z².
kev.ralston@ed.ac.uk

1. National Centre for Research Methods, University of Edinburgh 2.
University of St Andrews 3. University of Edinburgh 4. University of Stirling

Background

- Categorising those ***Not in Education Employment and Training*** (NEET)
- NEET, constructed from *Economic Activity* indicator of 16-19 year olds (at 1991 in our data):

NEET/Economically inactive	Non-NEET/Economically Active
unemployed, permanently sick, retired, looking after home/family and other inactive	students, those on training schemes and waiting to start a job, the employed

Data and methods

- Scottish Longitudinal Study (SLS)
 - 5.3% semi-random sample of the population of Scotland, based on 20 birth dates
 - 1991, 2001 and 2011 Censuses
 - Sample, 16-19 at 1991 Census, n=14567
 - NEET \approx 13.5%
 - Main analytic sample =8187 (56.2%)
 - Those recorded and traced at all 3 censuses
 - Examined the missing and they are more likely to be NEET
 - If anything we therefore think it likely any bias will underestimate the effect of NEET

Table 1, Examples of occupations by CAMSIS score, based on soc2010

Men		Women	
CAMSIS score	Soc2010 occupation	CAMSIS score	Soc2010 occupation
8.83	Street cleaners	8.76	Industrial cleaning process occupations
20.46	Kitchen and catering assistants	21.37	Cleaners and domestics
30	Quarry workers and related operatives	30.39	Launderers, dry cleaners and pressers
40.1	Aircraft maintenance and related trades	40.16	Customer service occupations
50.03	Leisure and sports managers	50.18	Library clerks and assistants
60.01	Senior officers in fire, ambulance, prison and related services	60.07	School secretaries
70.16	Biological scientists and biochemists	70.81	Company secretaries
80.02	Higher education teaching professionals	80.29	Authors, writers and translators
92.04	Social and humanities scientists	92.63	Medical practitioners

Data and methods: variables

Table 2, Descriptive statistics by men and women

	% Men	% Women
No qualifications (by 2001)	9	8
Standard Grade 1	33	34
Highers 2	19	19
HNC/HND 3	13	12
Degree+ 4	25	28
Carstairs 1 – least deprived	24	21
Carstairs 2	20	21
Carstairs 3	19	20
Carstairs 4	19	19
Carstairs 5 – most deprived	17	19
Age 16	24	22
17	26	24
18	25	26
19	26	28
LLTI 01- no	94	93
LLTI 01- yes	6	7
LLTI 91- no	98	98
LLTI 91- yes	2	2
Non-NEET/Active01	83	73
Non-NEET/Inactive01	8	15
NEET/Active01	6	6
NEET/non-active01	3	5
n=	3737	4450

The n and percentages are given with missing removed.

1,These are high school graduate level qualification in Scotland

2,These are high school qualifications usually used to gain university entrance

3,Further educational level qualifications

4, Degree and higher degrees

Source: SLS

Data and methods: methods

- Model **CAMSIS** using OLS regression, controlling for:
 - Educational attainment, deprivation, age, LLTI91 and 01, NEET-economic activity interaction
- CAMSIS stratified by level of education:
 - No-qualifications, high-school graduate, university entrance level, FE-college, degree+
- Selection model:
 - selecting on whether the individual is in work or not and using the unemployment rate at 1991 as an instrument

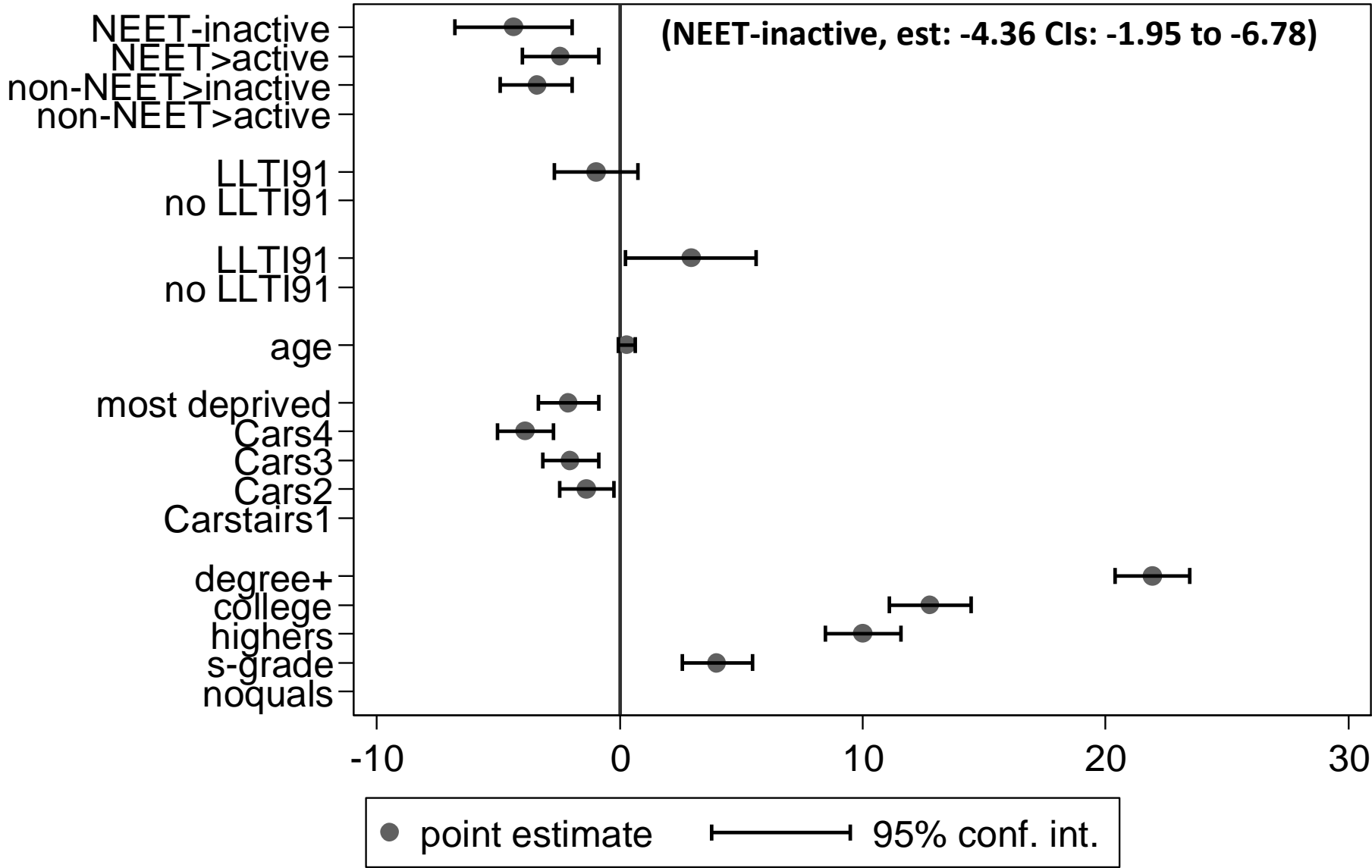
Results:

Is there occupational scarring related to NEET status?

OLS Regressions

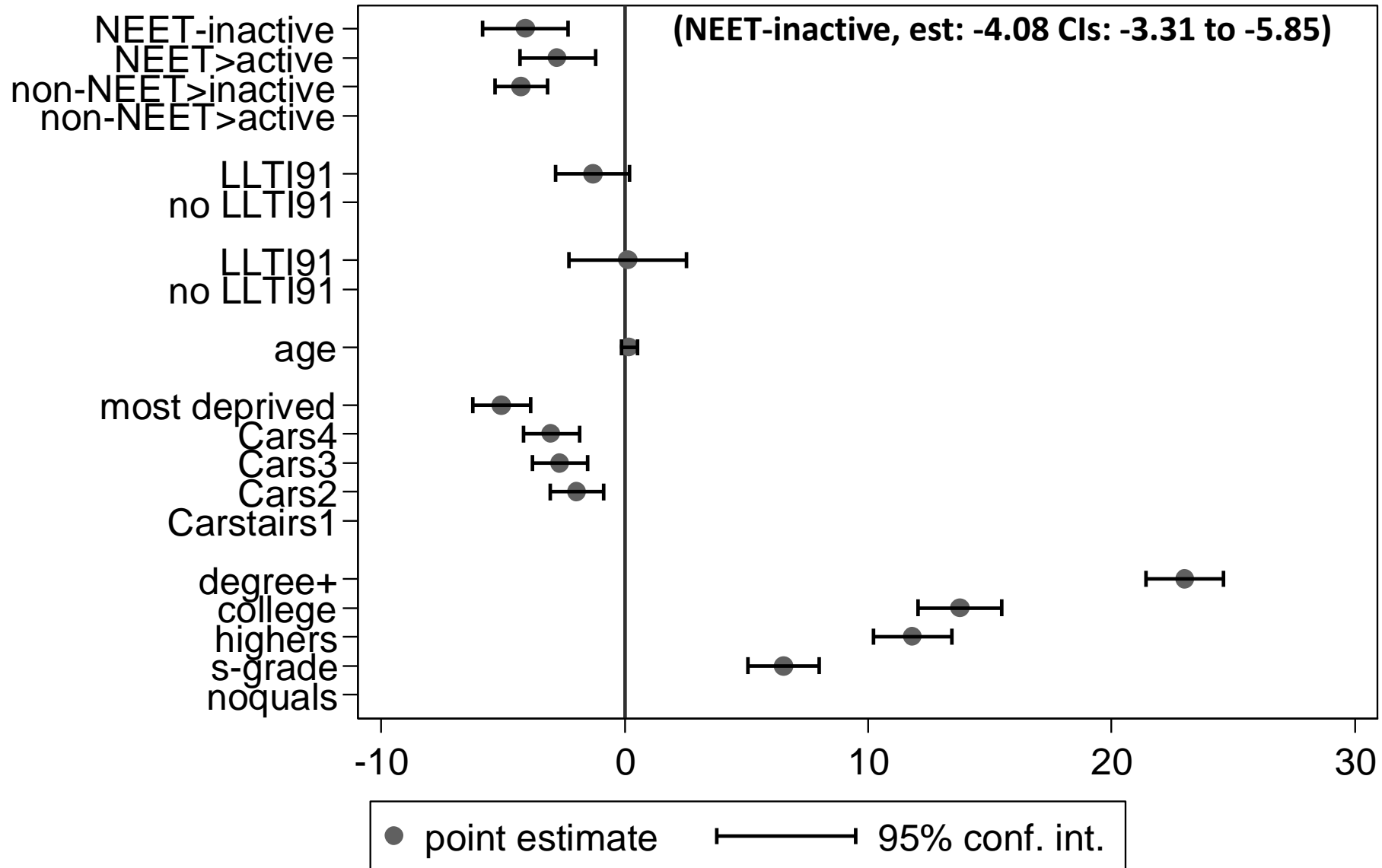
$$Y_i = b_0 + b_1 x_{i1} + e_i$$

CAMSIS score of men



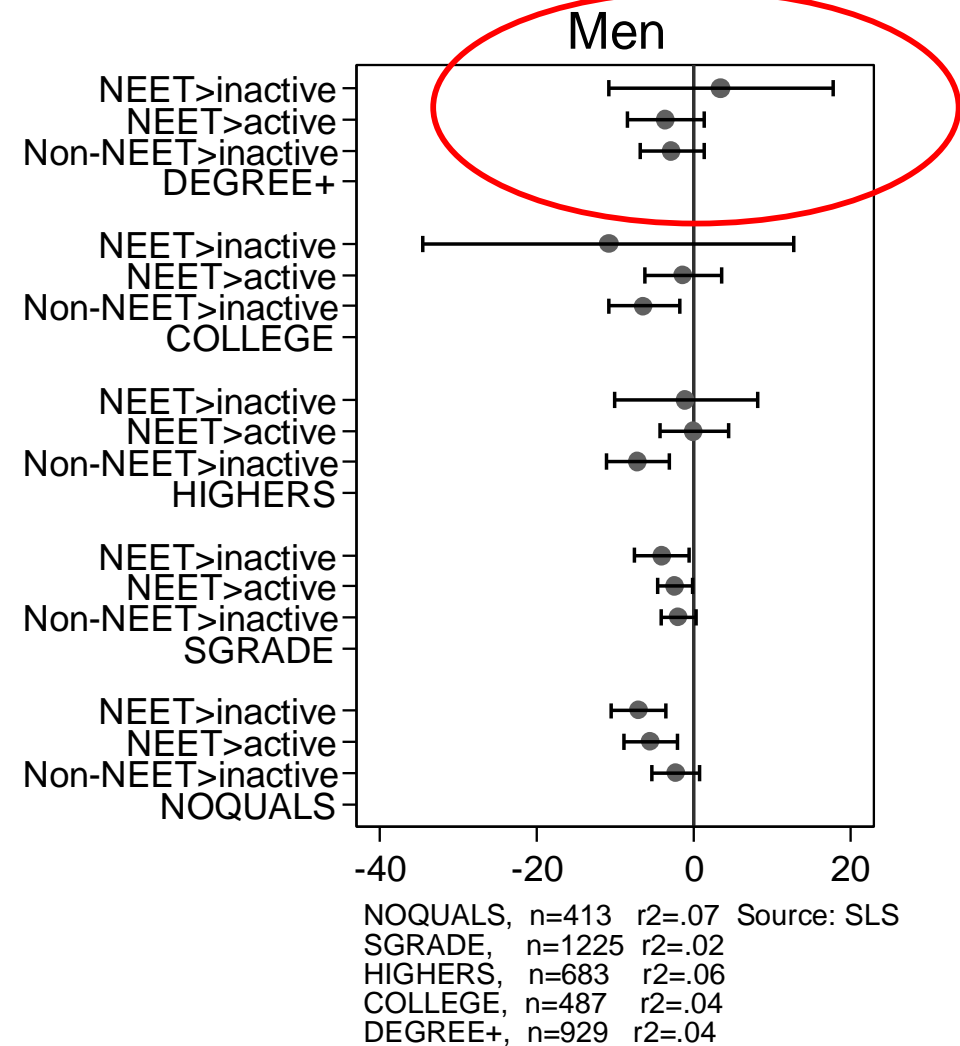
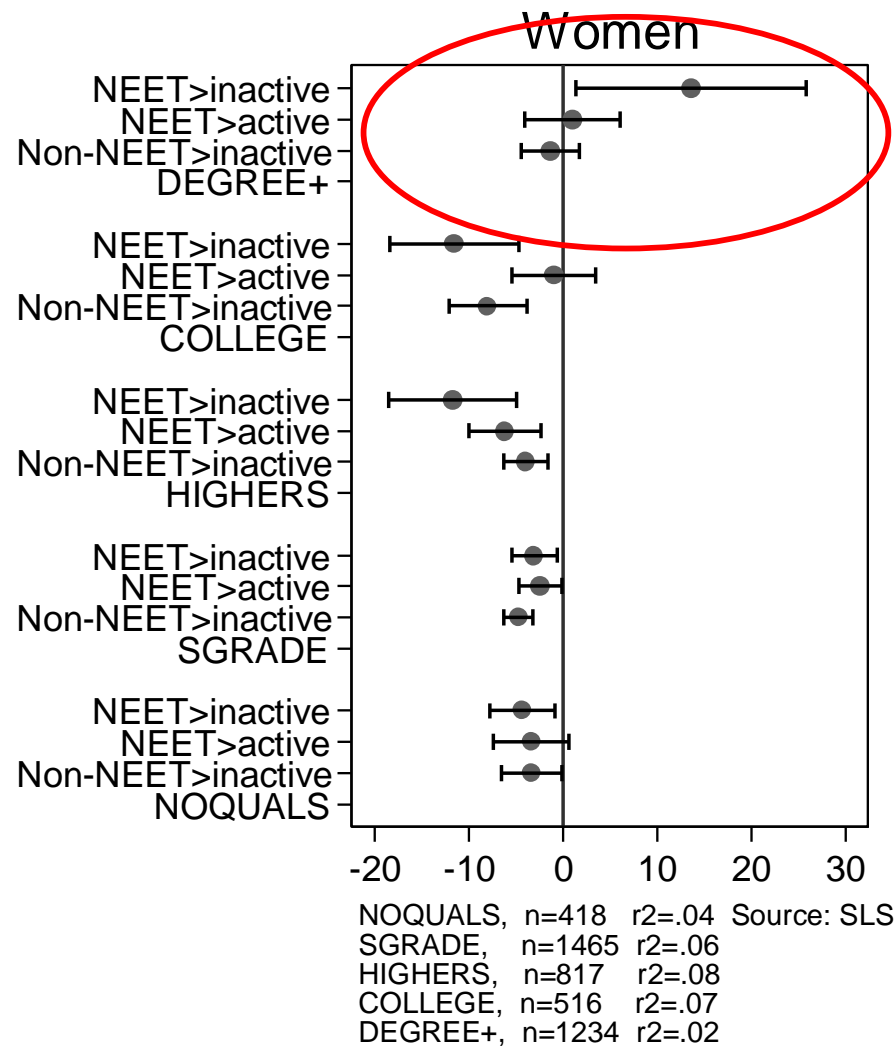
r2=0.35, n=3737, Source: SLS

CAMSIS score of women



$r^2=0.36$, $n=4450$, Source: SLS

CAMSIS score of models stratified by level of education showing only the interaction term



Area deprivation, age and LLTI 91 and 01 controlled in the models, results not reported

Selection model

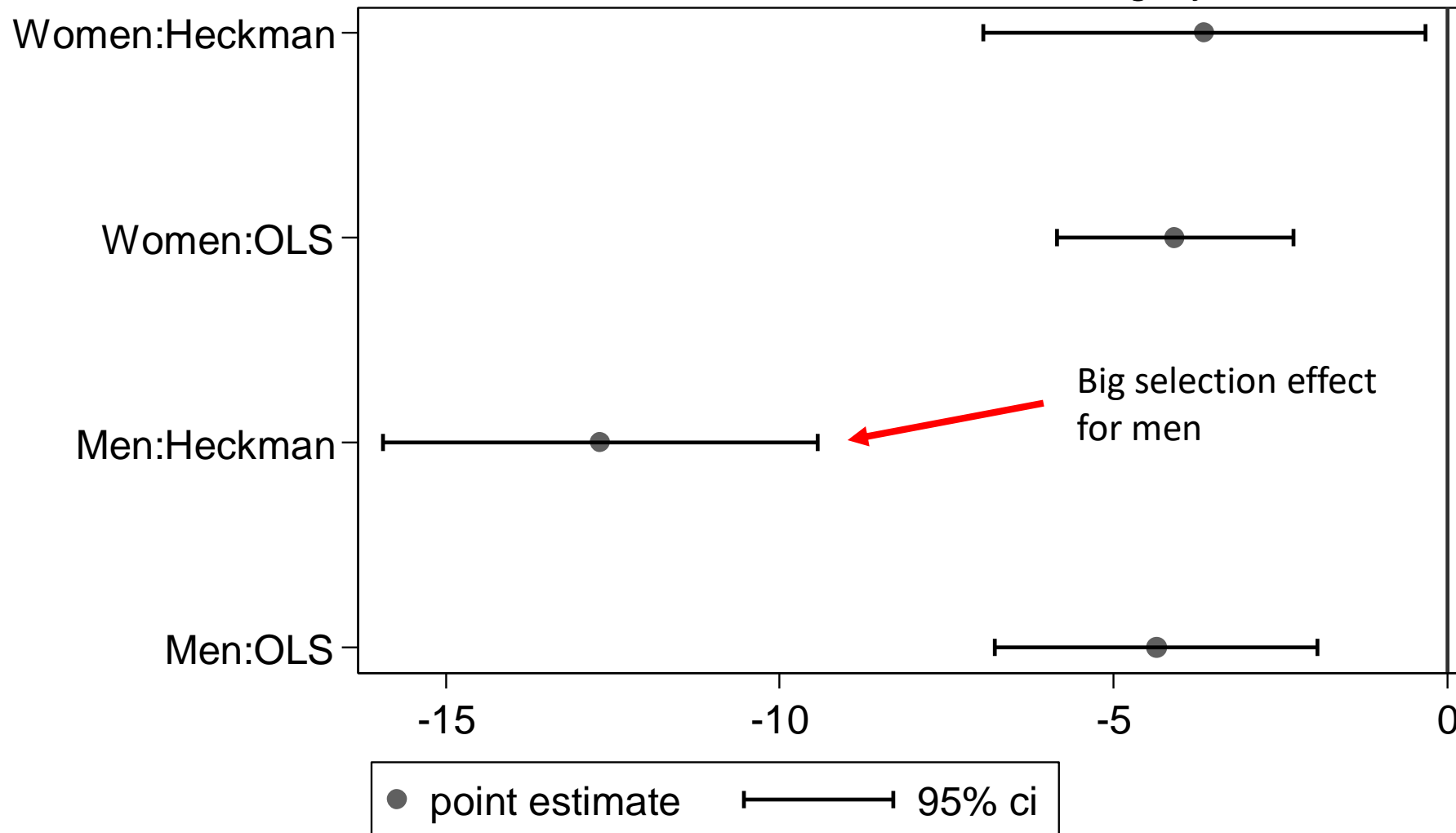
- Gregg and Tominey (2001) used selection to examine the relationship between youth unemployment and adult unemployment, NCDS.
- They also use an instrumental variable approach:
 - *'the unemployment rate prevalent locally for individuals aged 16 is used to instrument youth unemployment in the wage equation for individuals aged 33'*
 - They argue:
 - *instrument drives the unemployment experience, the endogenous factor, but which is exogenous to the individual themselves.*
 - *The intuition is that at such a young age, the individuals have little autonomy over their area of residence, thus the personal characteristics of the individuals are removed from the equation.*

Selection Model

- Account for selection effects in CAMSIS score, selecting on those in work or not, when aged 36-39 in the 2011 Census
- Instrument:
 - We included the unemployment rate at output area at 1991, when young people aged 16-19
 - At this point we have applied it in the selection equation
- Not perfect
 - Personal characteristics may not relate to area of residence at 16?
 - We have 16 to 19 year olds
 - In any case it's a difficult assumption to sustain, it assumes children's characteristics don't relate to parents
 - Gregg and Tominey (2001) therefore control for background effects
 - Location is influenced by various factors,
 - control for deprivation Carstairs quintiles, account for parental characteristics and go some way to mean these are not picked up in the instrument

CAMSIS score:

Comparison of OLS and Heckman results
NEET and non-active category



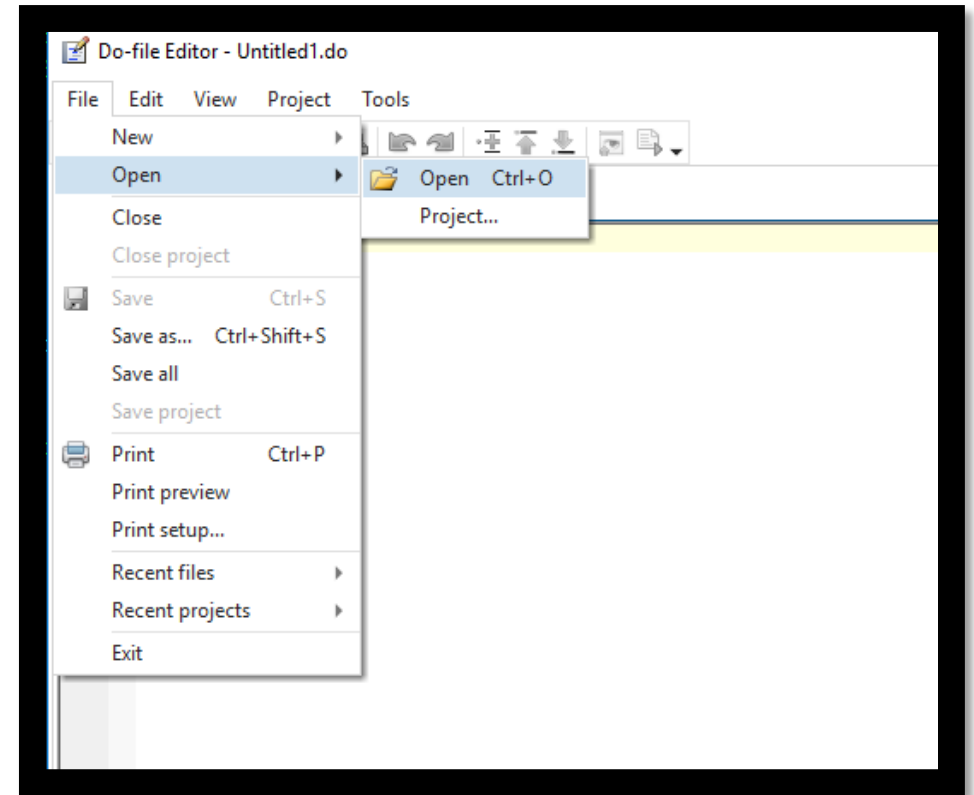
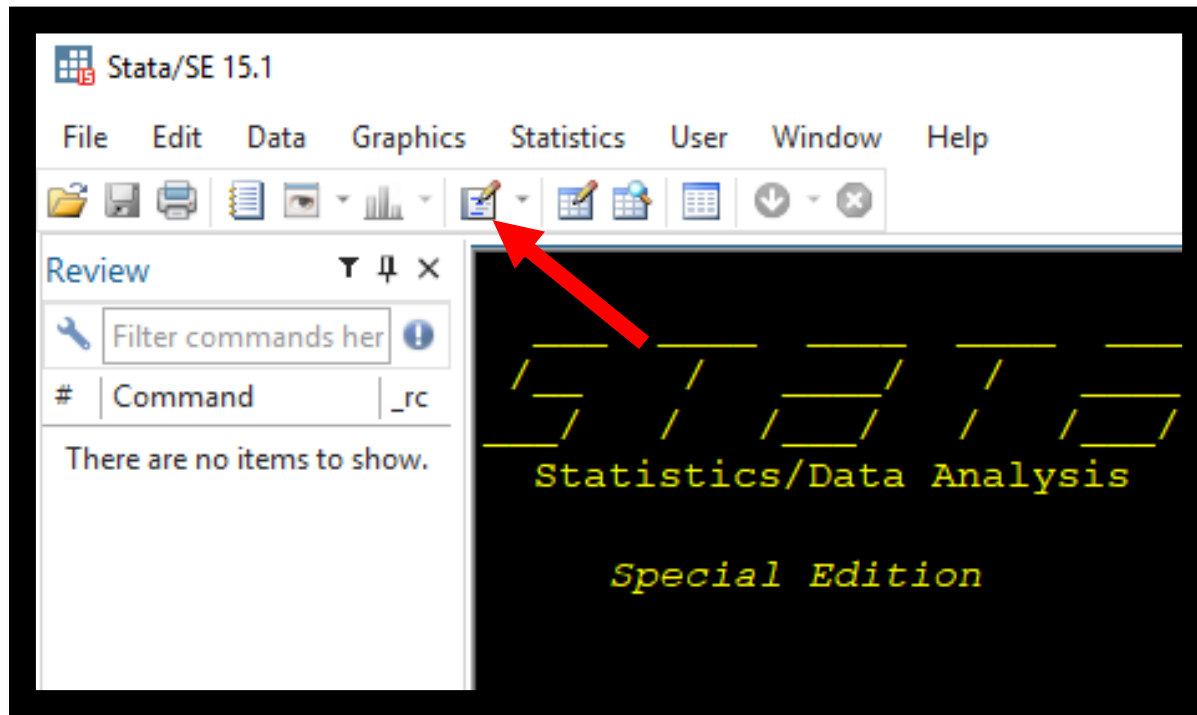
Source: SLS

Results

1. The results confirm occupational scarring
2. Results show those moving into work from NEET status also seem to still do worse
3. CAMSIS score stratified by education provides some evidence for different types of NEET
 1. Where NEET is generally a disadvantage (although there is a small sample size and not all categories are significant)
 2. Women with degrees who were NEET>inactive are not disadvantaged by NEET, no significant effects for either men or women
4. Differences between men and women, big selection effect for men

Hands On Session 3

- Open Stata
- Open the Lab3 .do file



Publication Ready Tables

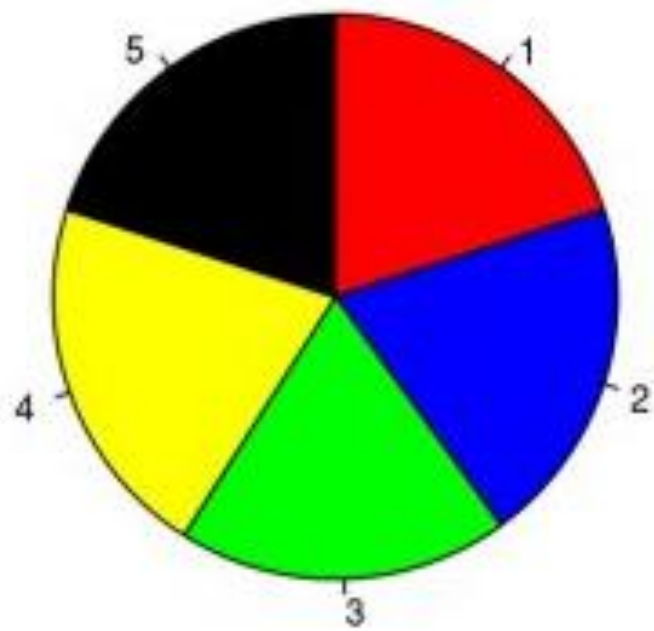
Publication Reading Tables

- Do not copy and paste
 - You will make mistakes
 - This is inefficient
 - Stata will prepare tables for you (in word)!!!

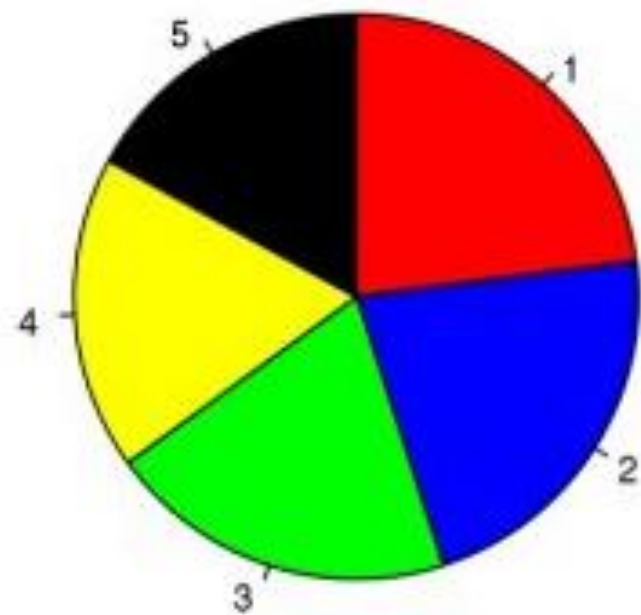
A

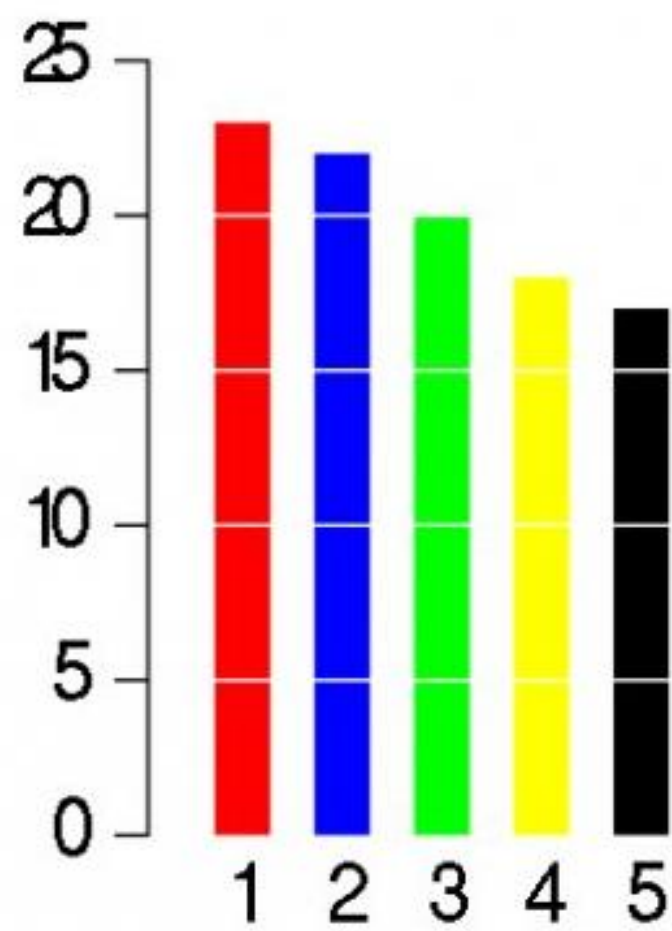
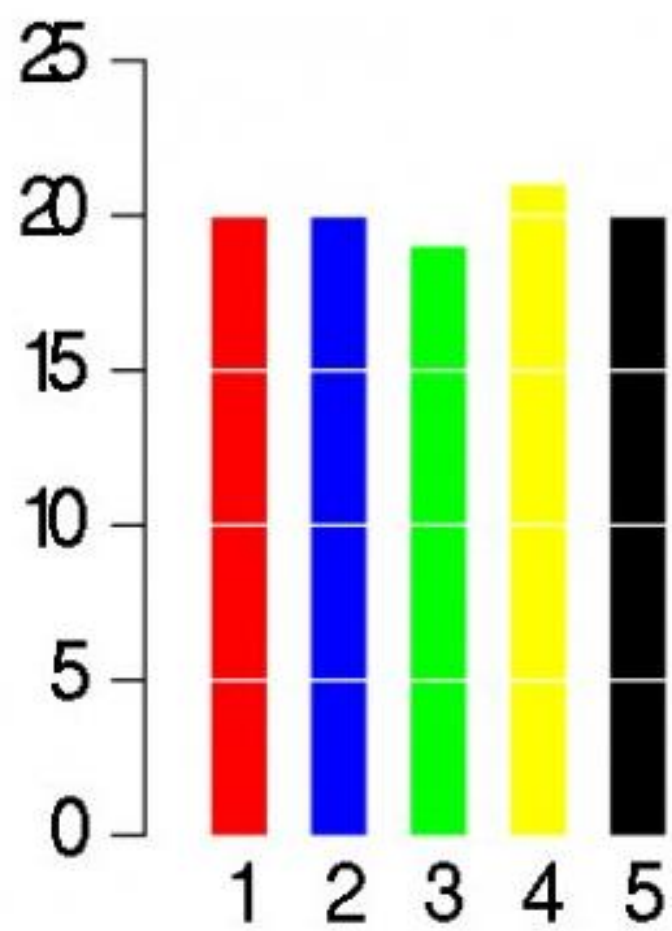
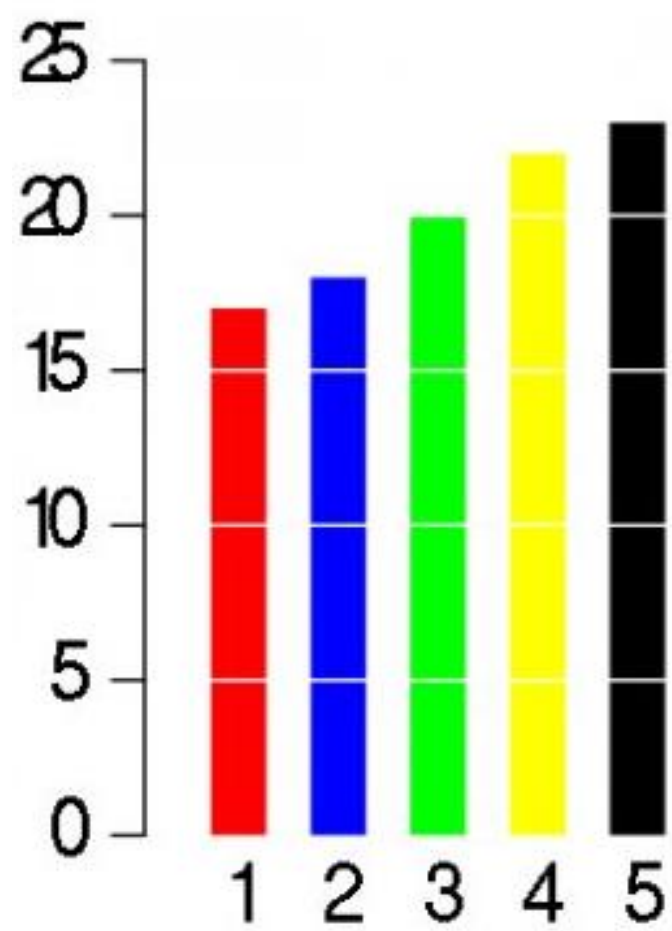


B



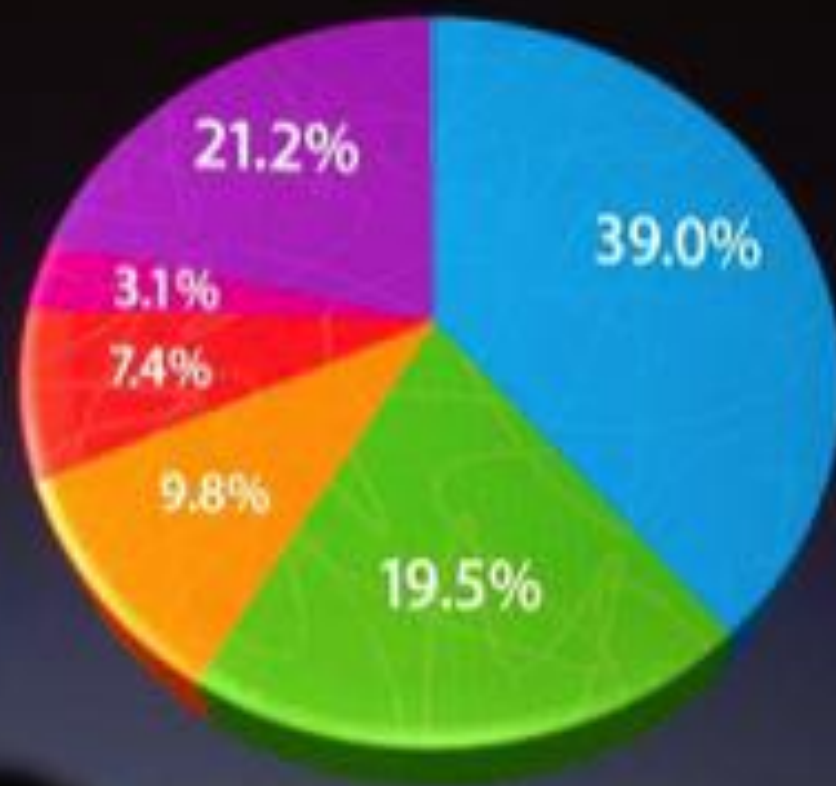
C





U.S. SmartPhone Marketshare

- RIM
- Apple
- Palm
- Motorola
- Nokia
- Other



Engage Gartner for



Edward Tufte

@EdwardTufte

Following



Pie chart users deserve same suspicion+skepticism as those who mix up its/it's, there/their. To compare, use little table, sentence, not pies.

RETWEETS

133

LIKES

67



7:13 PM - 9 Jan 2013



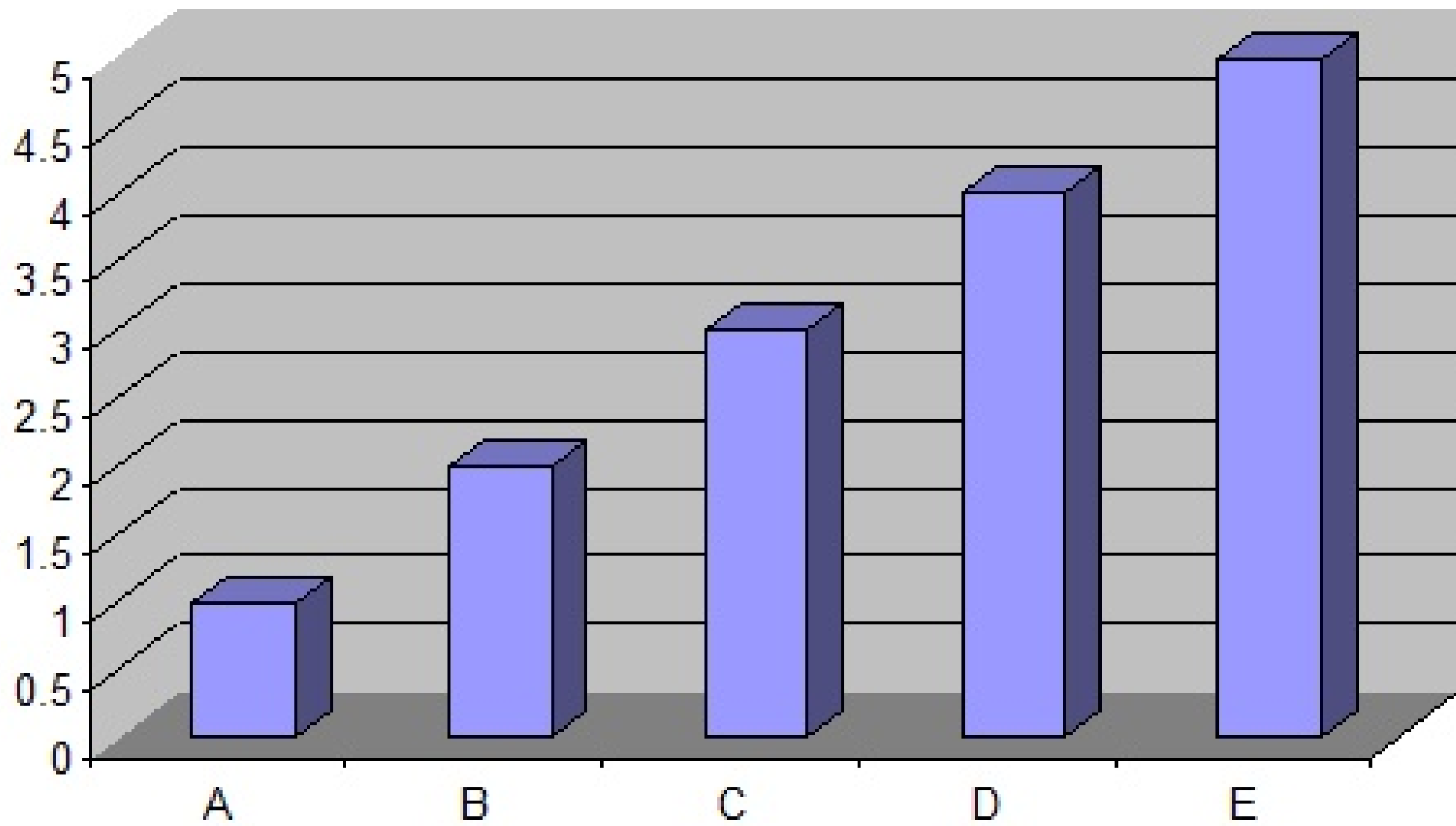
10

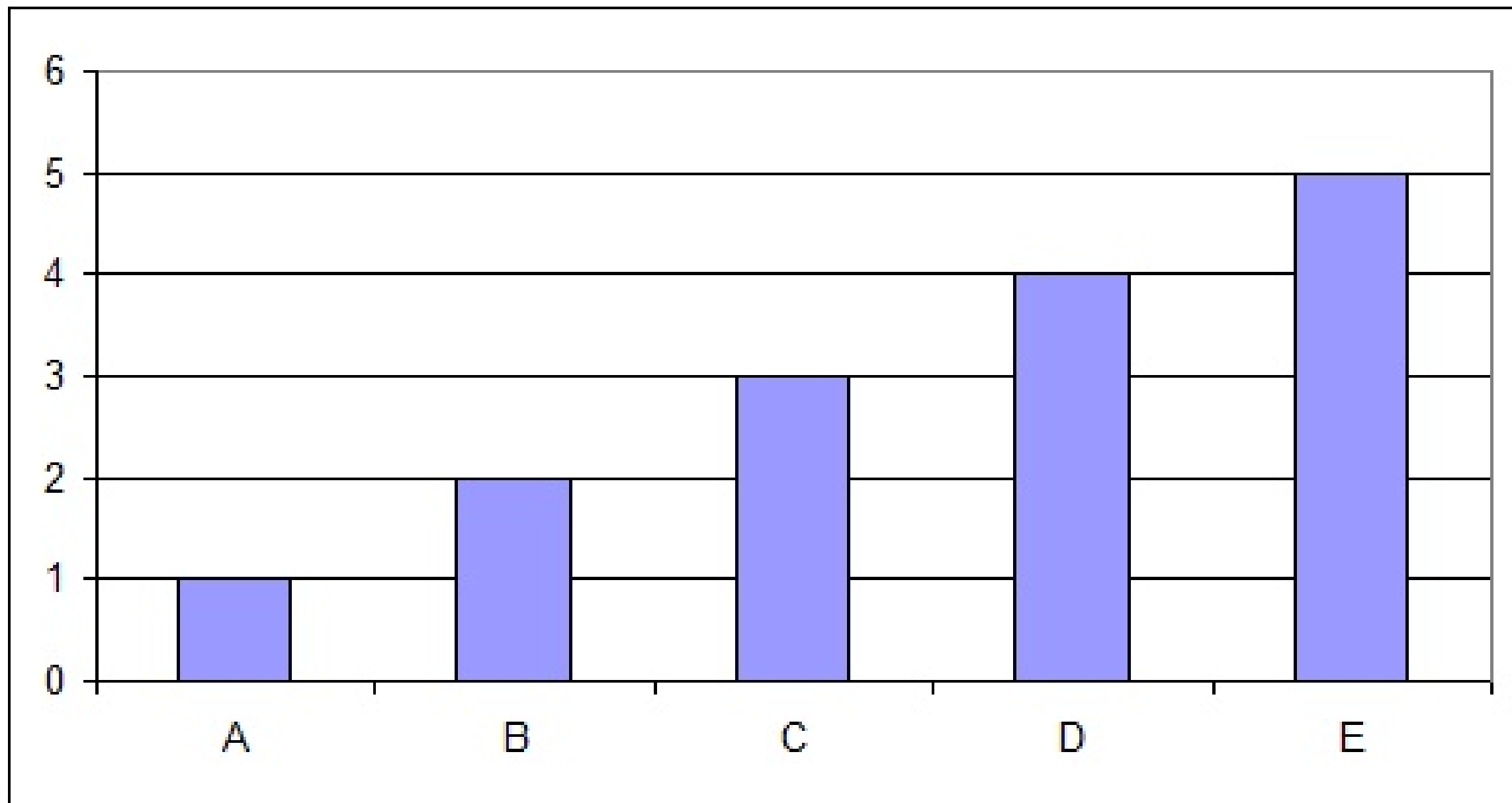


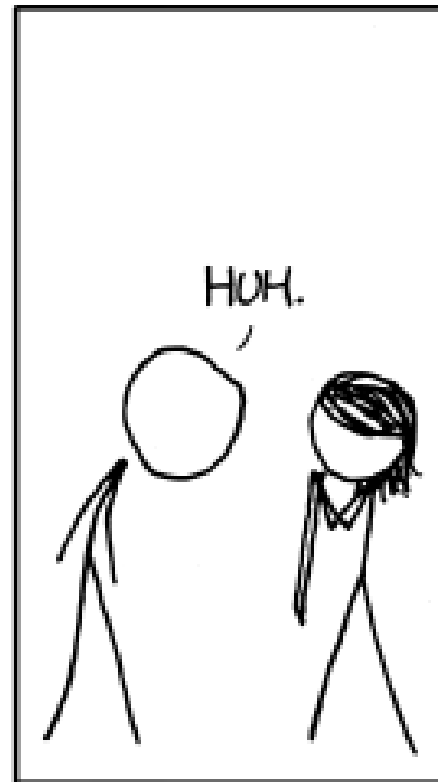
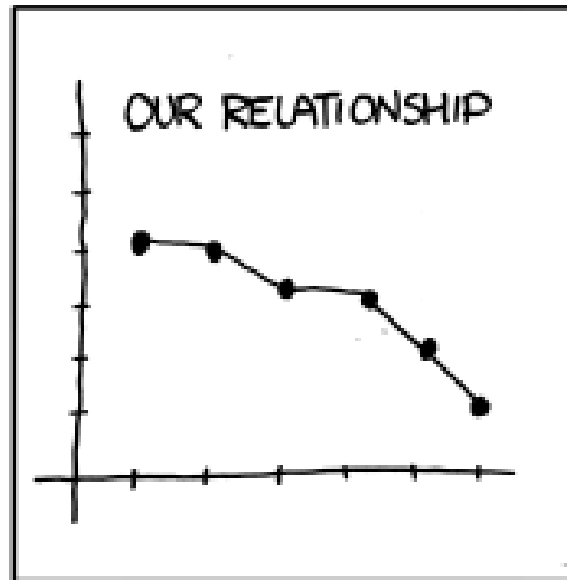
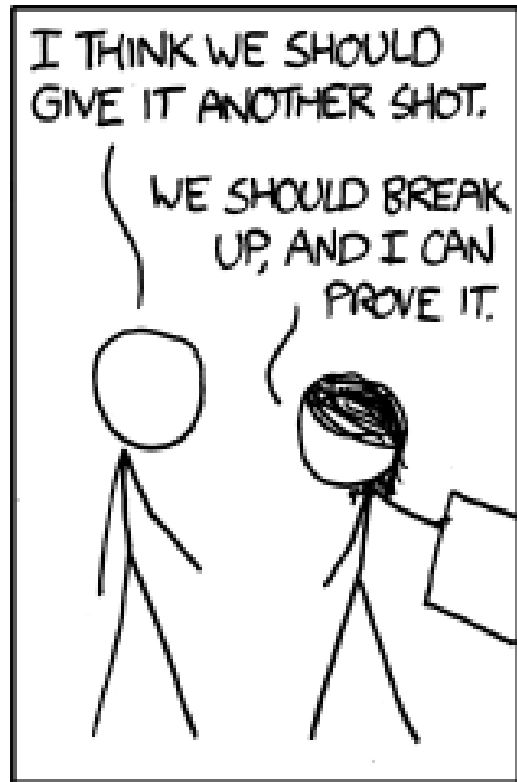
133



67







Effective Graphs

- Edward Tufte
 - *'The daVinci of data'* New York Times

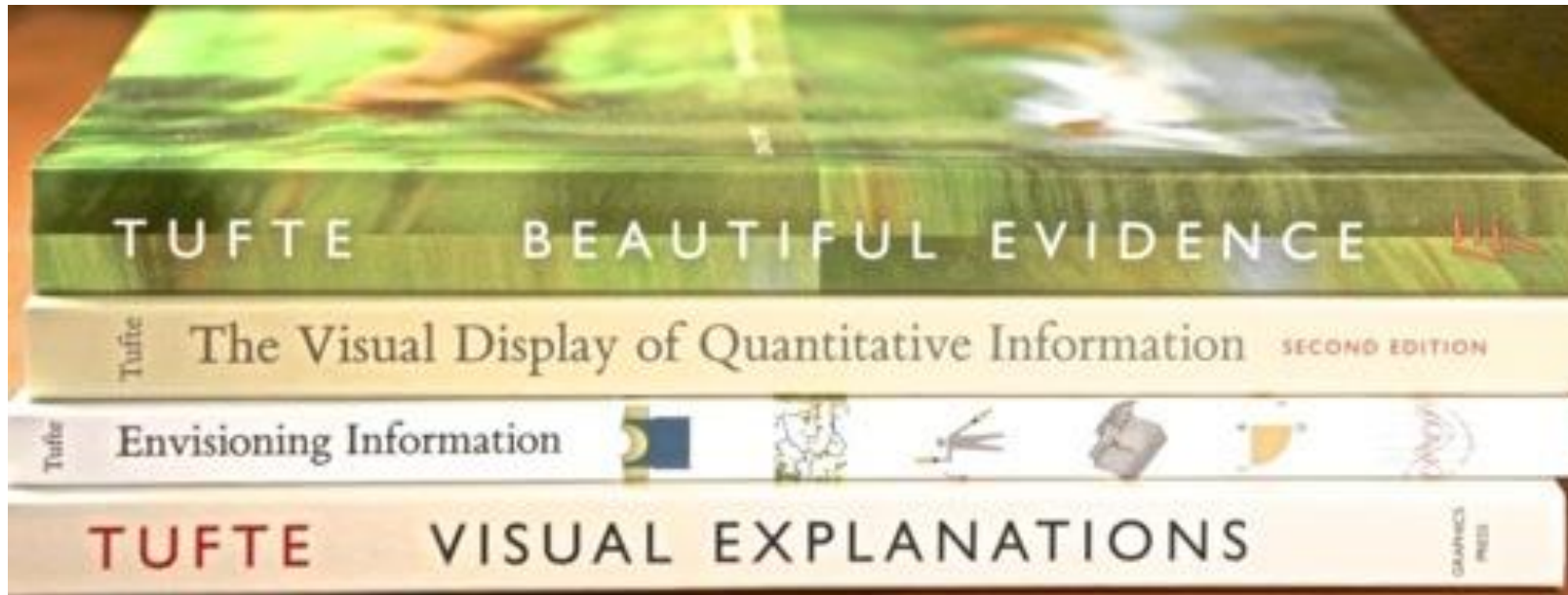


Table 1: OLS Regression model of maths test scores.

	Coef.		SE
Sex			
Male	Ref.		
Female	-0.292		(-0.23)
Parental Social Class			
Low	Ref.		
Medium	1.592		(0.97)
High	5.224	**	(2.89)
Ethnicity			
Hispanic	Ref.		
Asian	9.387	**	(2.91)
African American	-0.272		(-0.10)
White	5.646	**	(2.87)
Constant	45.95	***	(21.30)
n	200		
Adjusted R ²	0.12		

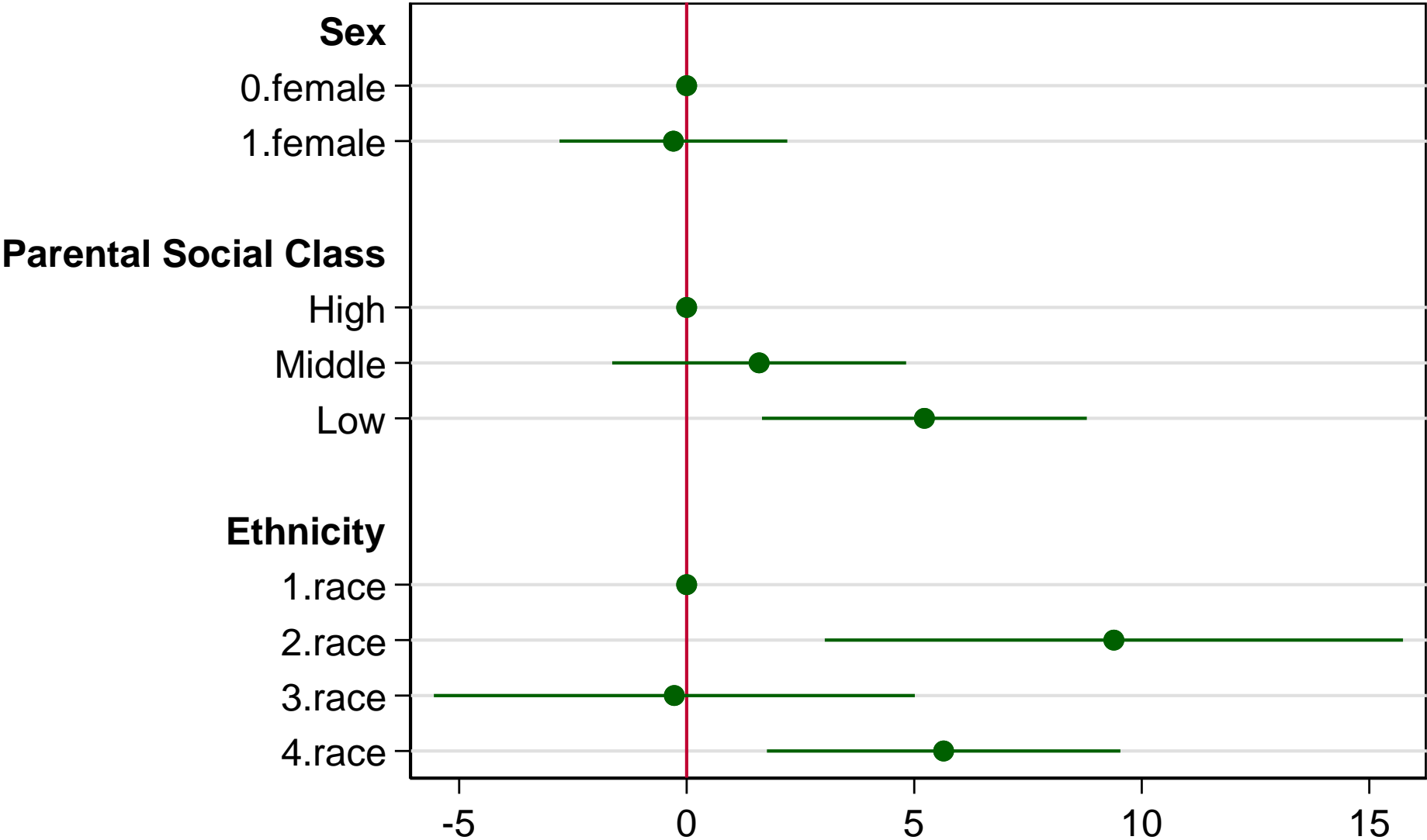
Data: High School and Beyond Survey

Note: * p < 0.05, ** p < 0.01, *** p < 0.001



OLS Regression Model of Maths Test Scores

Coefficients and 95% Standard Errors

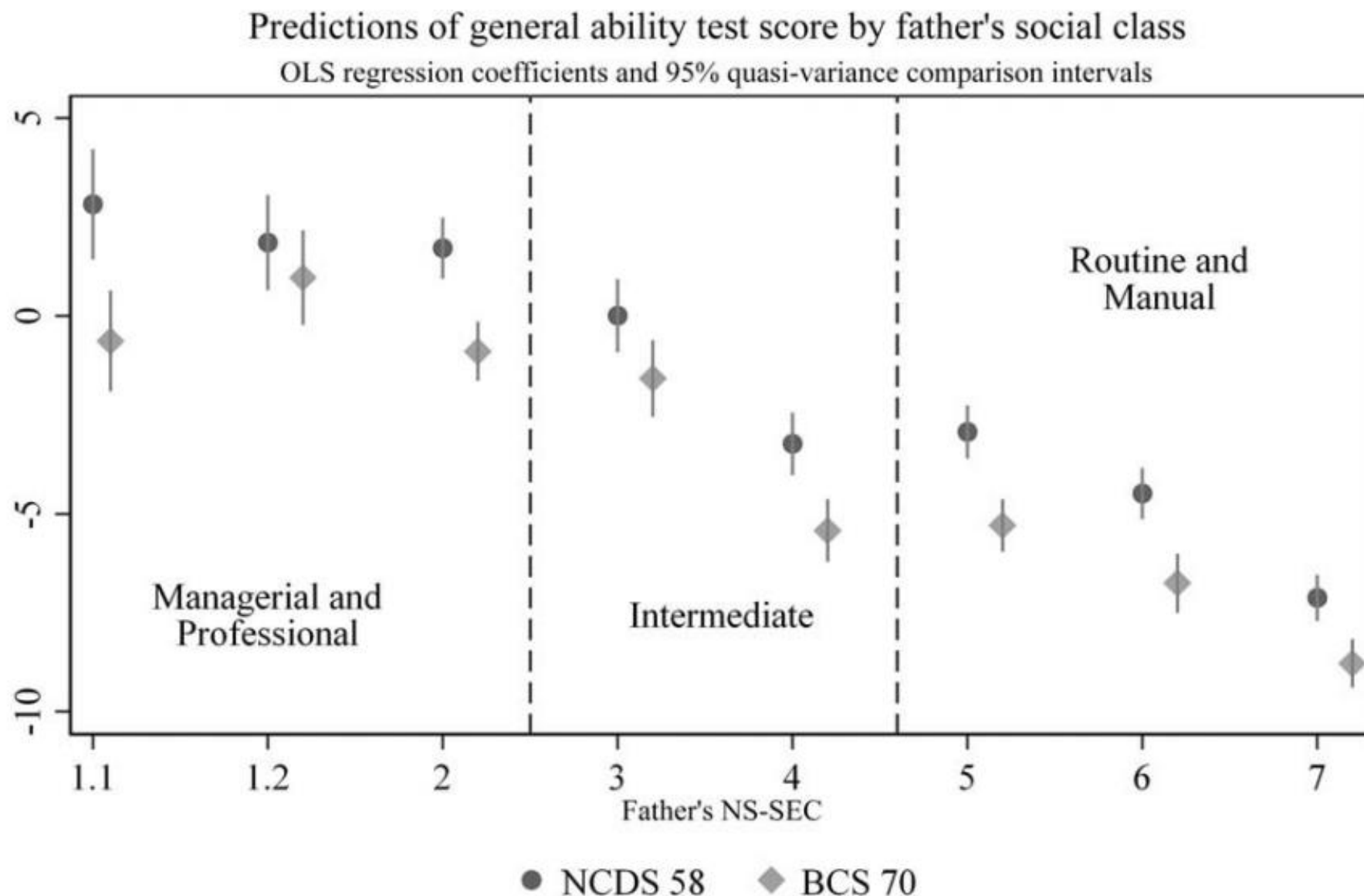


Data: High School and Beyond Data, n = 200, Adjusted R2 = 0.12

Table IV: Regression analysis (OLS) of general ability test scores pooled NCDS and BCS data, with adjustments for missing data

	Model 1 (MI+IPW)		Model 2 (MI+IPW)	
	Coef.	SE	Coef.	SE
Gender				
Female	Ref.		Ref.	
Male	-0.55**	(0.18)	-0.56**	(0.18)
Parent's highest education				
Compulsory school only	Ref.		Ref.	
Compulsory school + 1 to 3 years	5.87***	(0.24)	5.87***	(0.24)
Compulsory school + 4 to 5 years	8.30***	(0.54)	8.31***	(0.53)
Compulsory school + 6 or more years	10.63***	(0.46)	10.65***	(0.46)
Father's NS-SEC				
NS-SEC1.1	1.79**	(0.58)		
NS-SEC1.2	2.28***	(0.59)		
NS-SEC2	1.19**	(0.43)		
NS-SEC3	Ref.			
NS-SEC4	-3.53***	(0.43)		
NS-SEC5	-3.31***	(0.41)		
NS-SEC6	-4.80***	(0.43)		
NS-SEC7	-7.17***	(0.41)		
Cohort				
NCDS (1958)	Ref.			
BCS (1970)	-2.09***	(0.18)		
Father's NS-SEC \times Cohort interaction				
NS-SEC1.1 \times NCDS			2.82**	(0.86)
NS-SEC1.1 \times BCS			-0.64	(0.78)
NS-SEC1.2 \times NCDS			1.85*	(0.77)
NS-SEC1.2 \times BCS			0.96	(0.77)
NS-SEC2 \times NCDS			1.71**	(0.62)
NS-SEC2 \times BCS			-0.90	(0.59)
NS-SEC3 \times NCDS			Ref.	
NS-SEC3 \times BCS			-1.59*	(0.68)
NS-SEC4 \times NCDS			-3.24***	(0.63)
NS-SEC4 \times BCS			-5.42***	(0.62)
NS-SEC5 \times NCDS			-2.94***	(0.59)
NS-SEC5 \times BCS			-5.29***	(0.57)
NS-SEC6 \times NCDS			-4.48***	(0.60)
NS-SEC6 \times BCS			-6.75***	(0.59)
NS-SEC7 \times NCDS			-7.12***	(0.56)
NS-SEC7 \times BCS			-8.78***	(0.56)
Constant	104.06***	(0.43)	101.74***	(0.49)
<i>n</i>	28,331		28,331	
Adjusted <i>R</i> ²	0.14		0.14	

Note: IPW = Inverse Probability Weights; MI = Multiple Imputation. AIC and log likelihood cannot be calculated for models using Multiple Imputation. The sample size for models 1 and 2 is all non-deceased cohort members who were present in the third survey. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

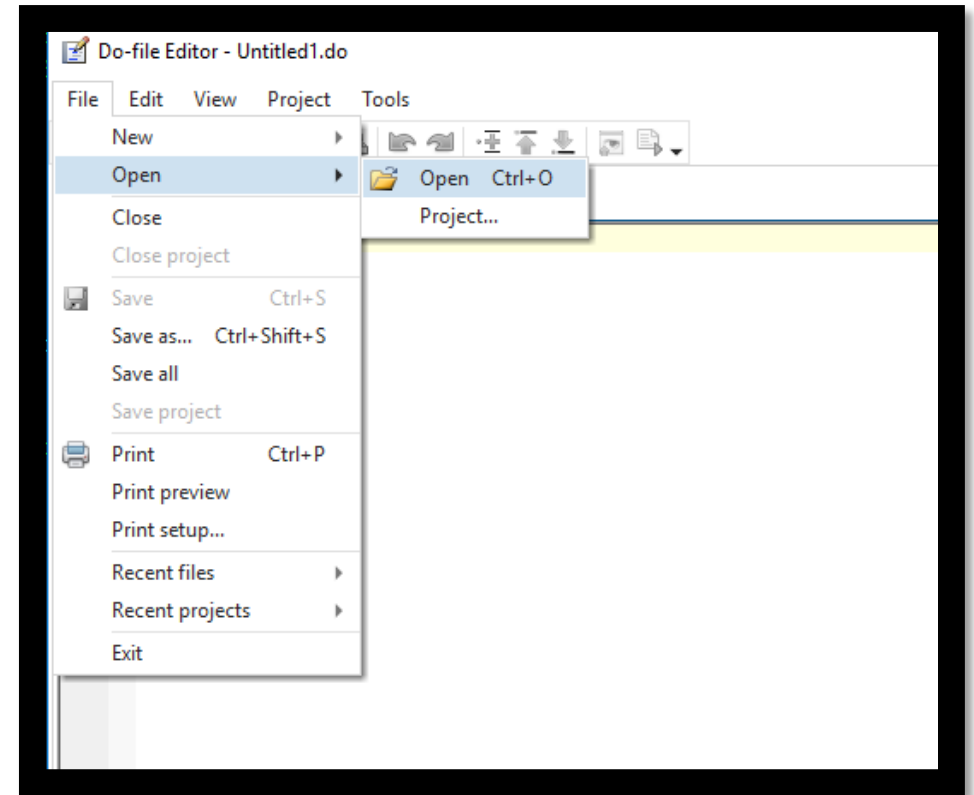
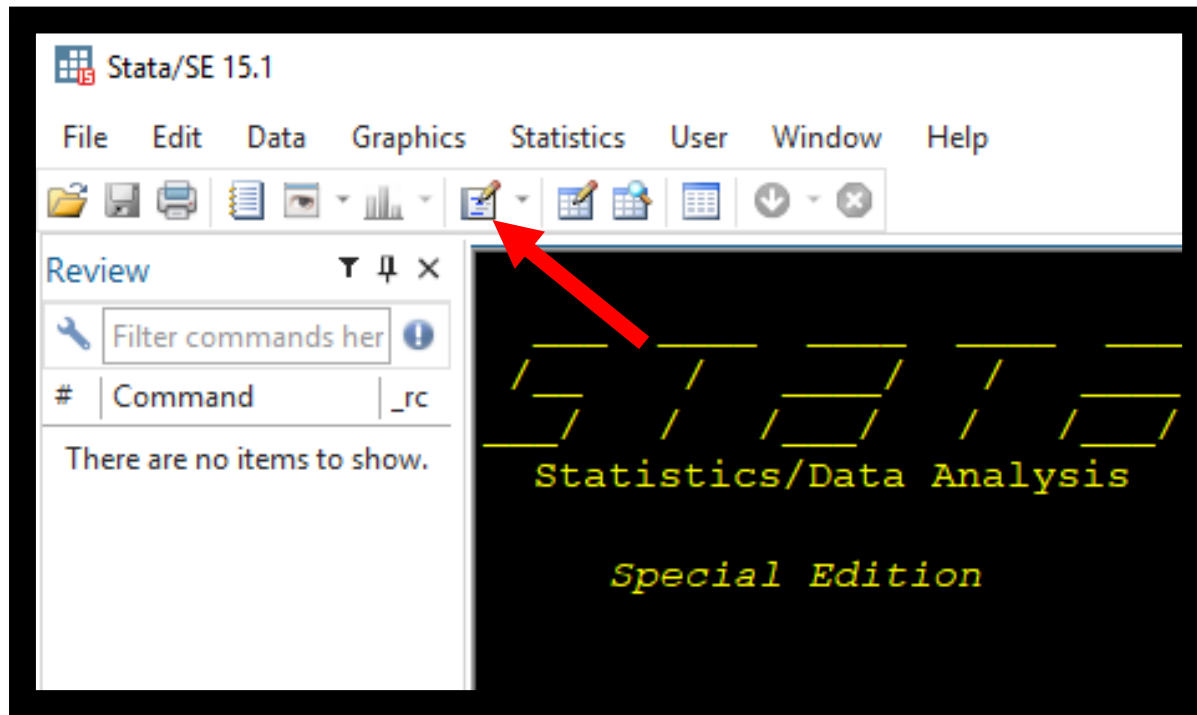


Note: Estimates are taken from Table IV, model 2. Model also contains gender and parent's highest education.

Data source: 1958 National Child Development Study and 1970 British Cohort Study.

Hands On Session 4

- Open Stata
- Open the Lab4 .do file



Questions?

Roxanne Connelly
roxanne.connelly@york.ac.uk

Kev Ralston
k.ralston@yorks.ac.uk

