

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

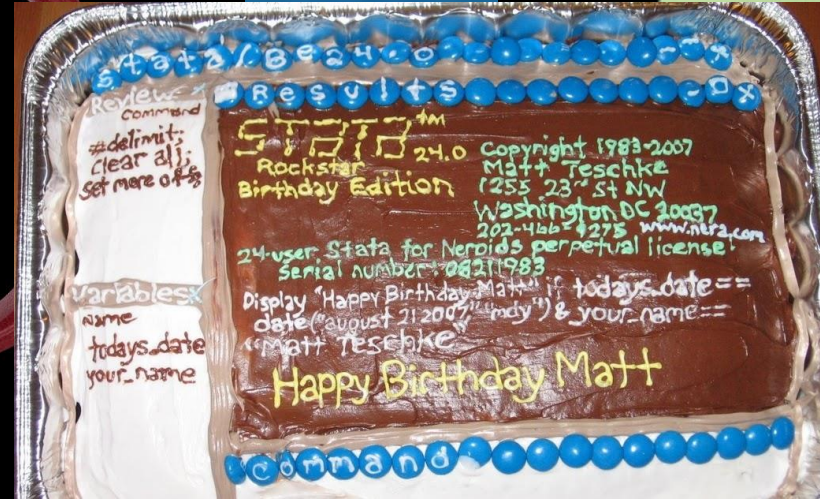
Support

A beautiful solution.



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

What can Stata do for you?

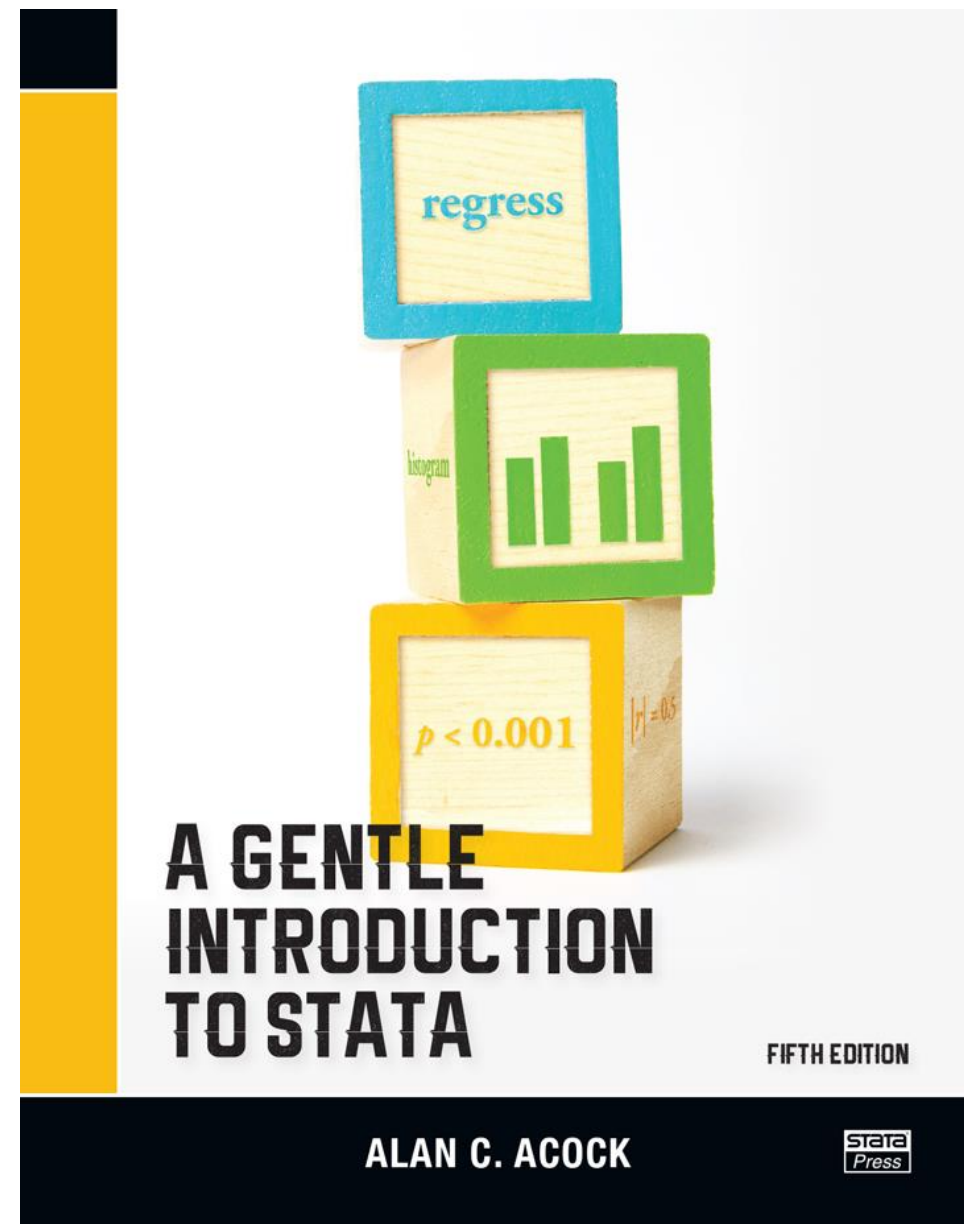
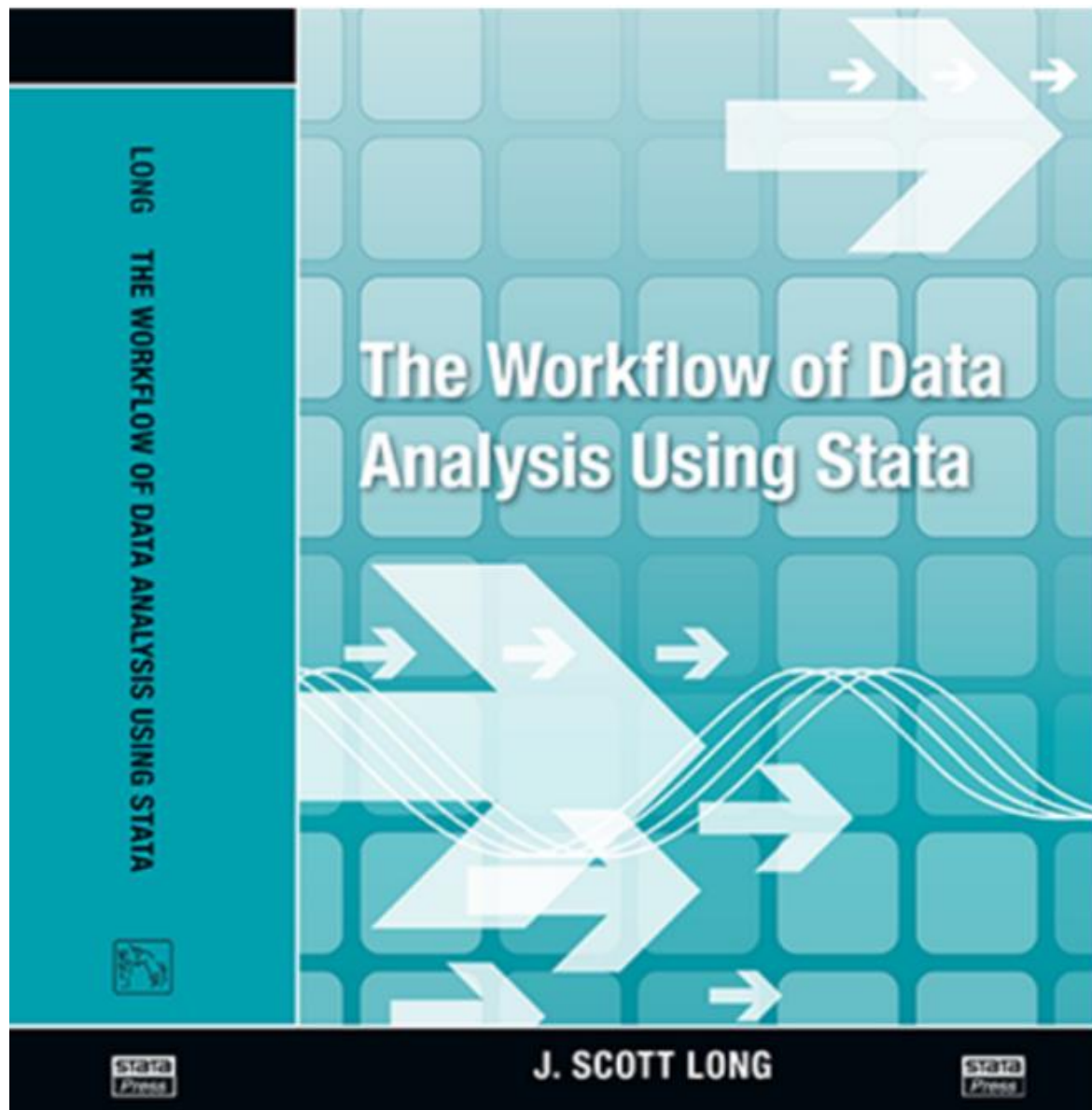


Why Stata?

- 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.

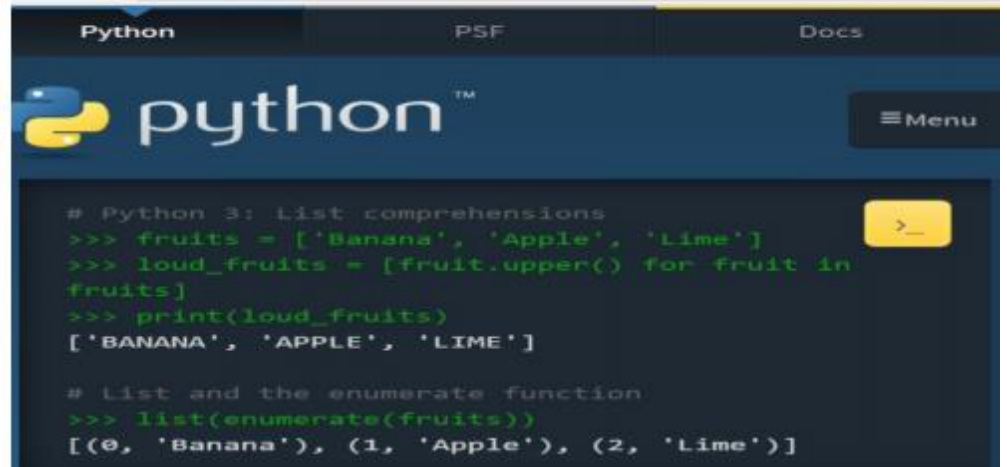
If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

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 (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.

Properties

Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

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

Command

\\CANTUS\User43\u\1474300\Documents

CAP NUM OVR

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

CAPNUMOVR09:3908/07/2019

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)

Special Edition

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.

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
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

Line: 1, Col: 1 CAP NUM OVR

\\CANTUS\User43\u1474300\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

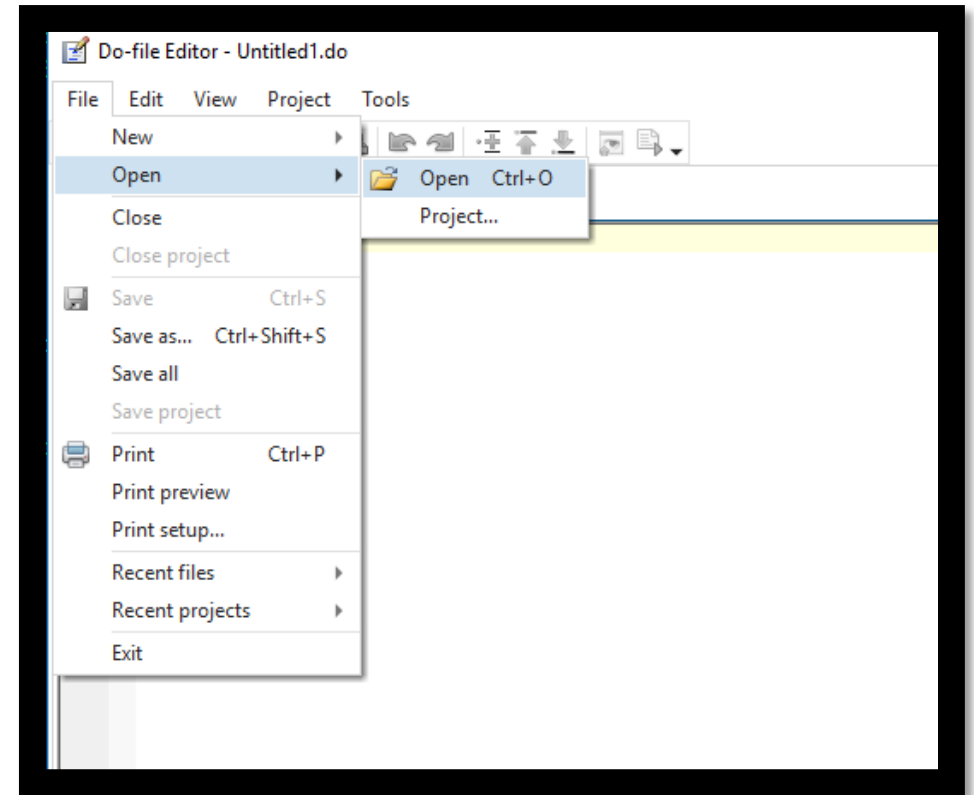
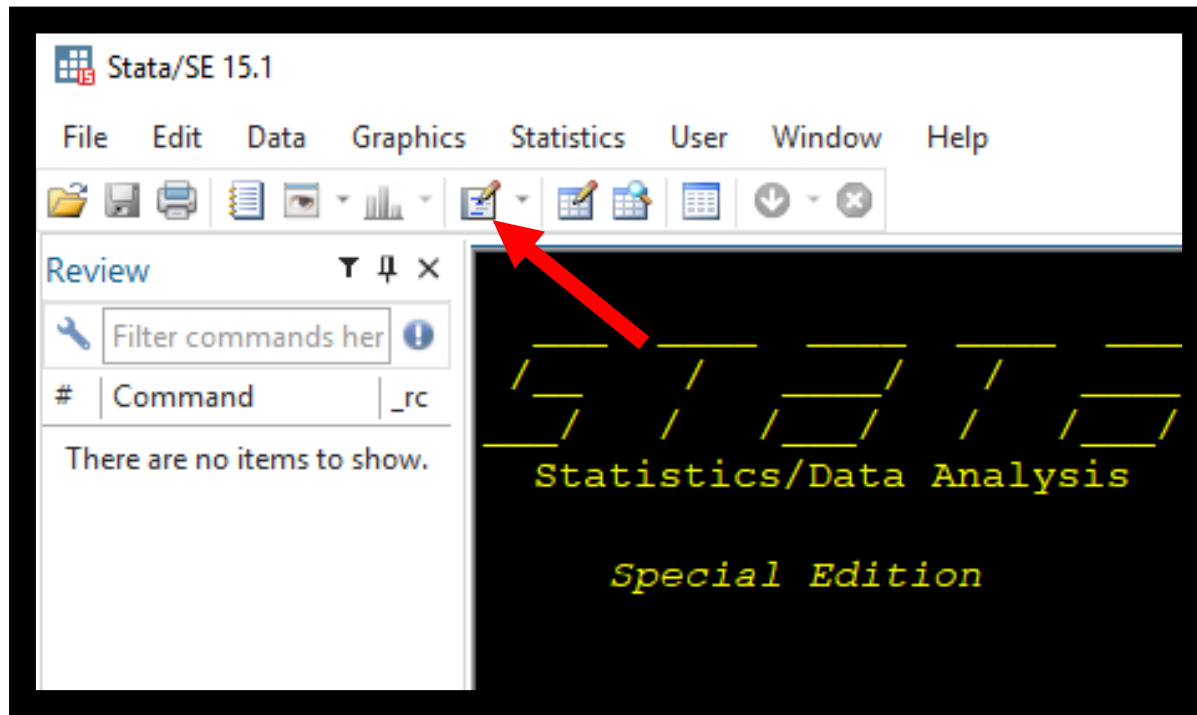
Materials can be downloaded here:

link

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

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

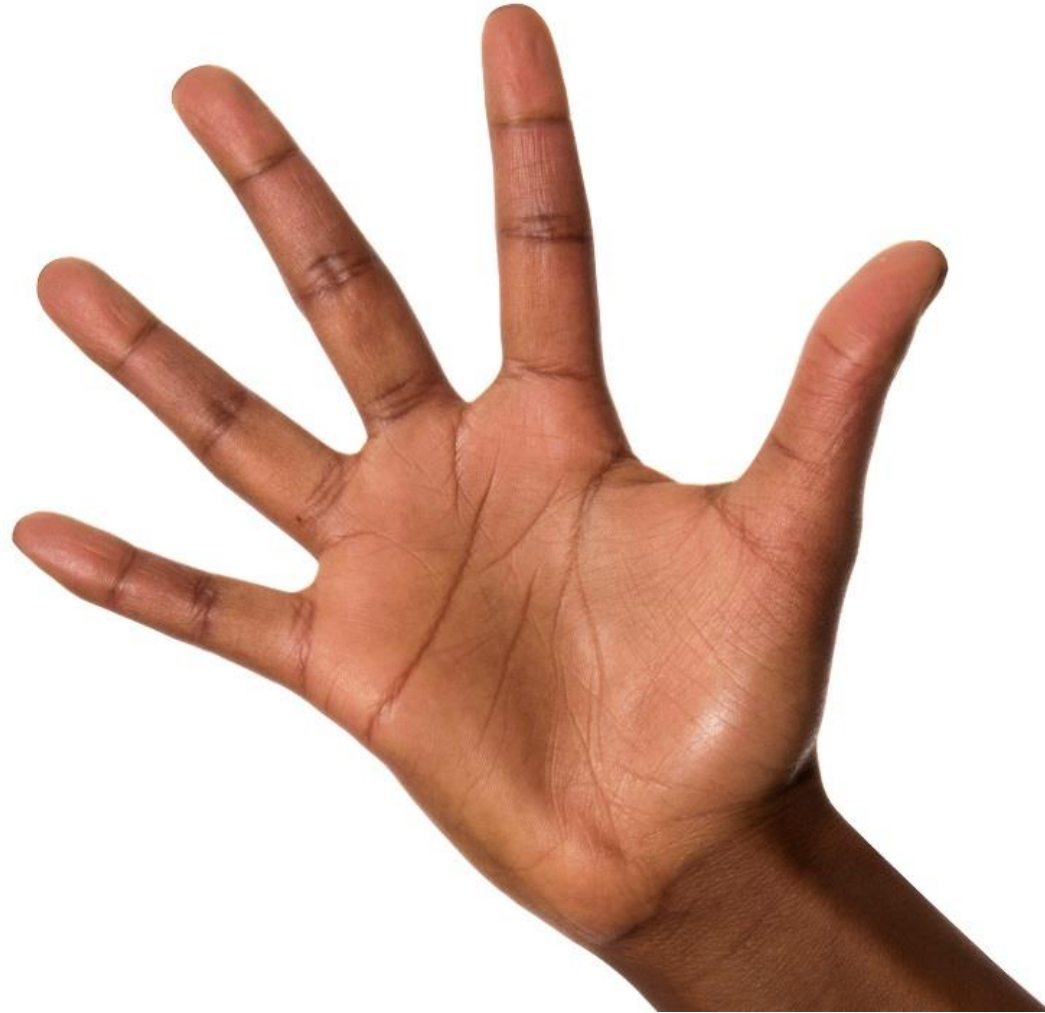
Have you ever lost a file?



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

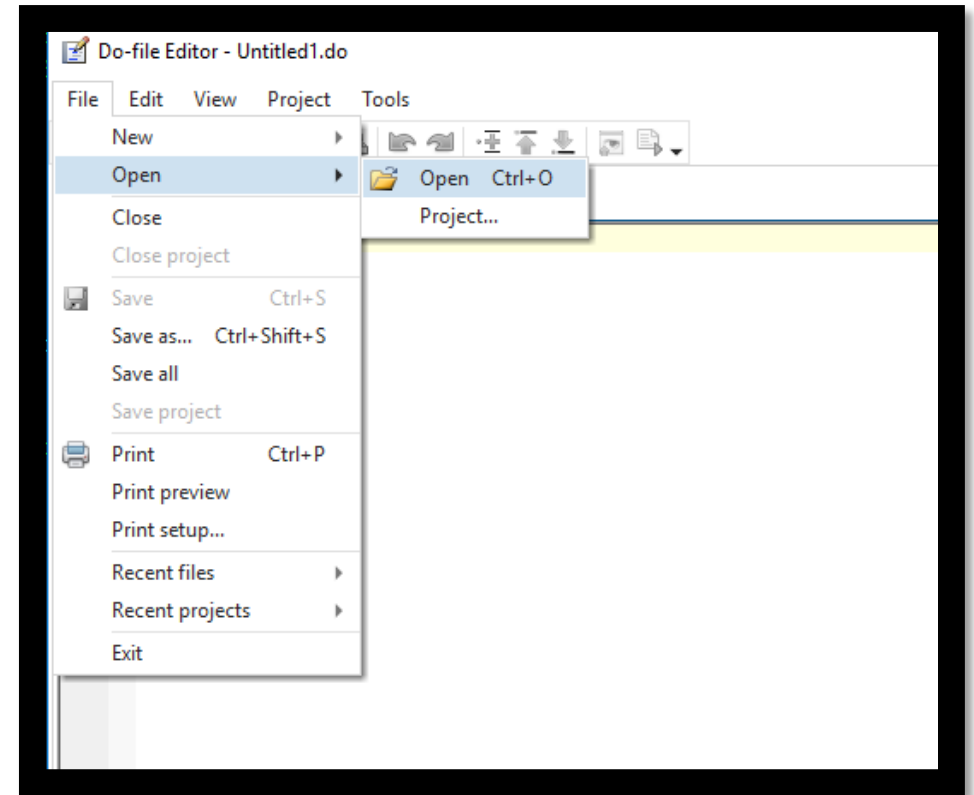
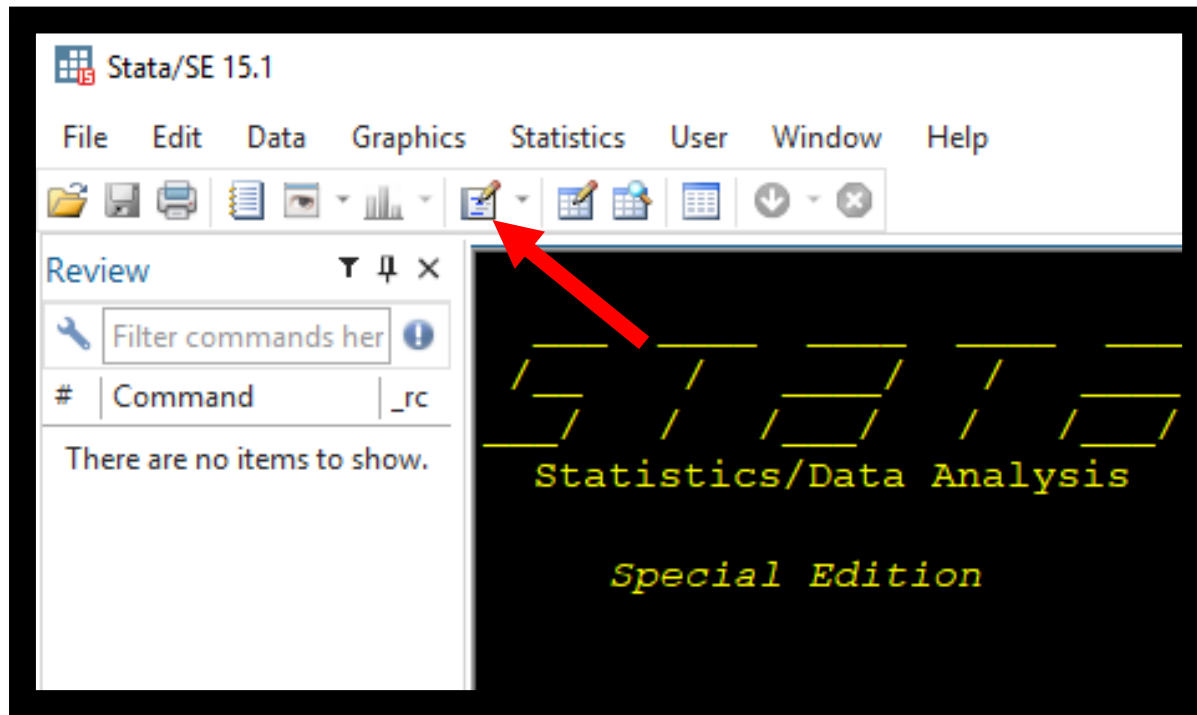


The Connelly-Ralston Constant



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

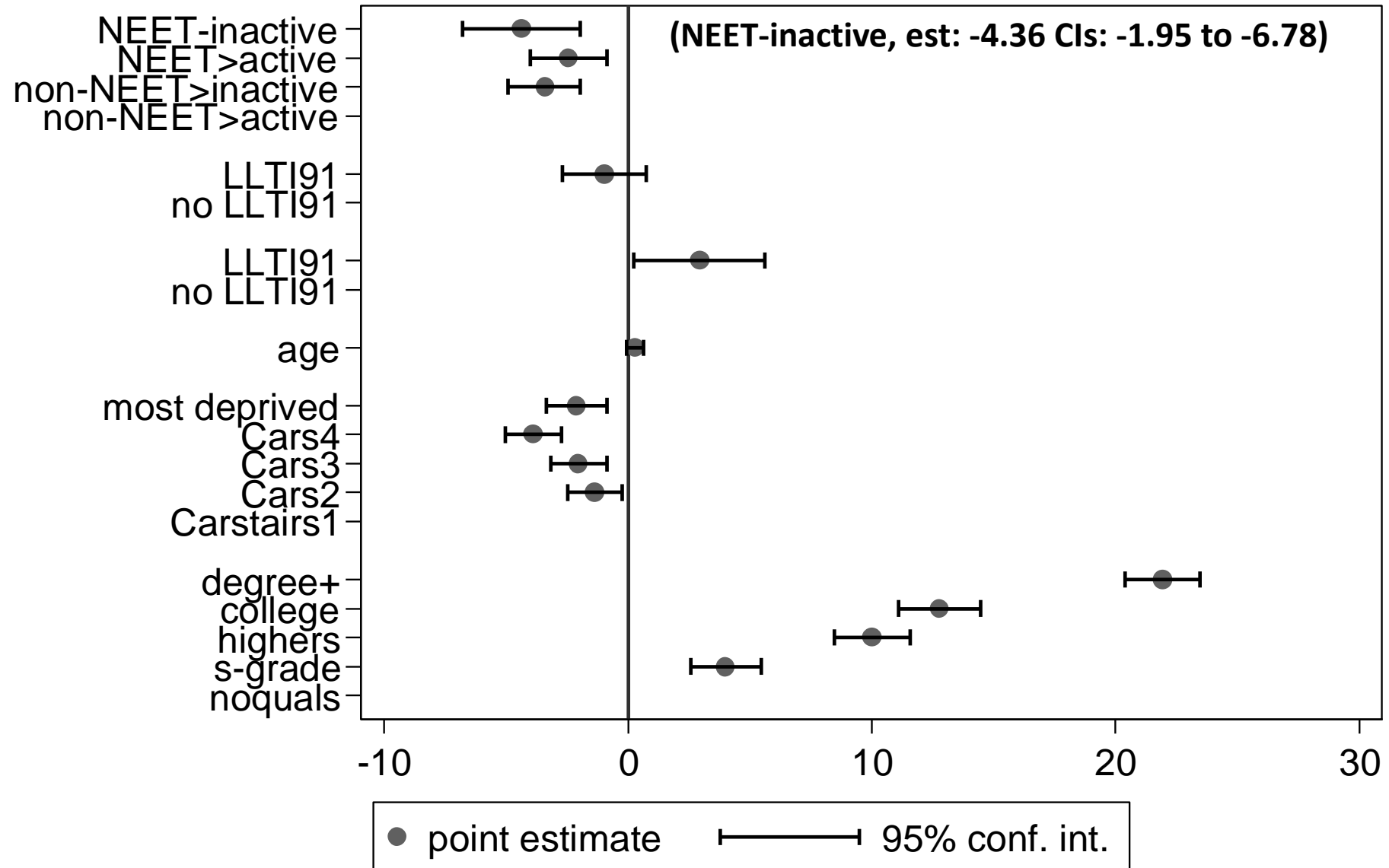
Results:

Is there occupational scarring related to NEET status?

OLS Regressions

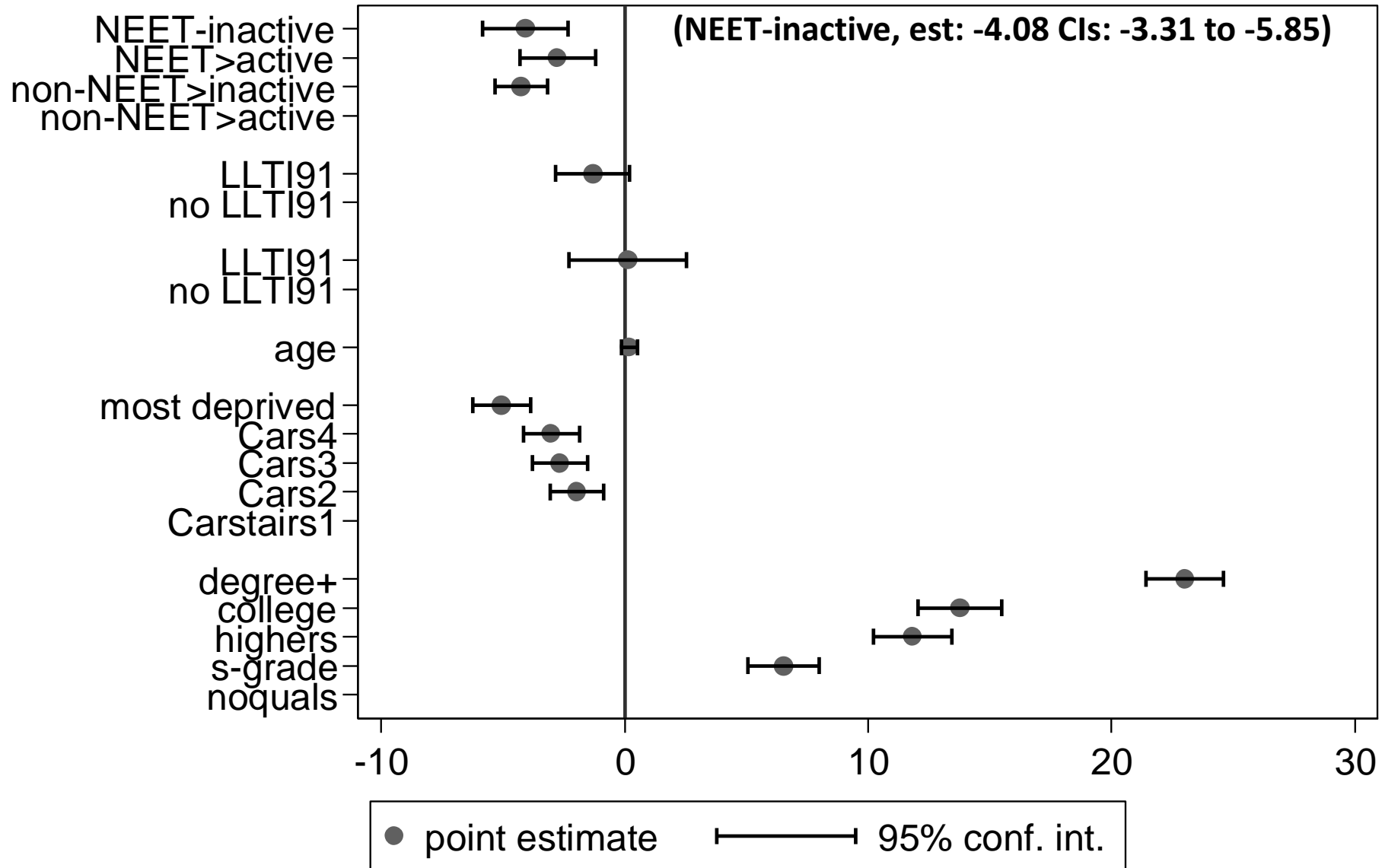
$$Y_i = b_0 + b_1 x_{i1} + e_i$$

CAMSIS score of men



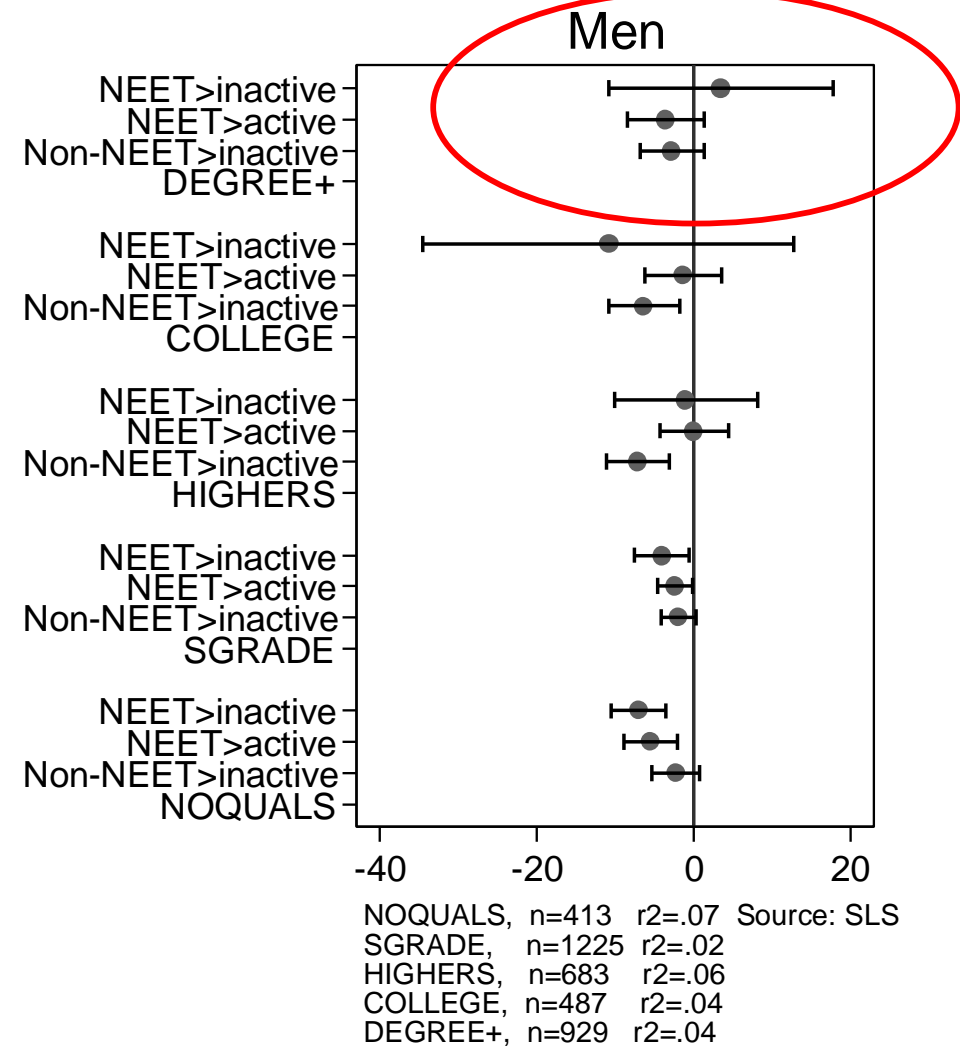
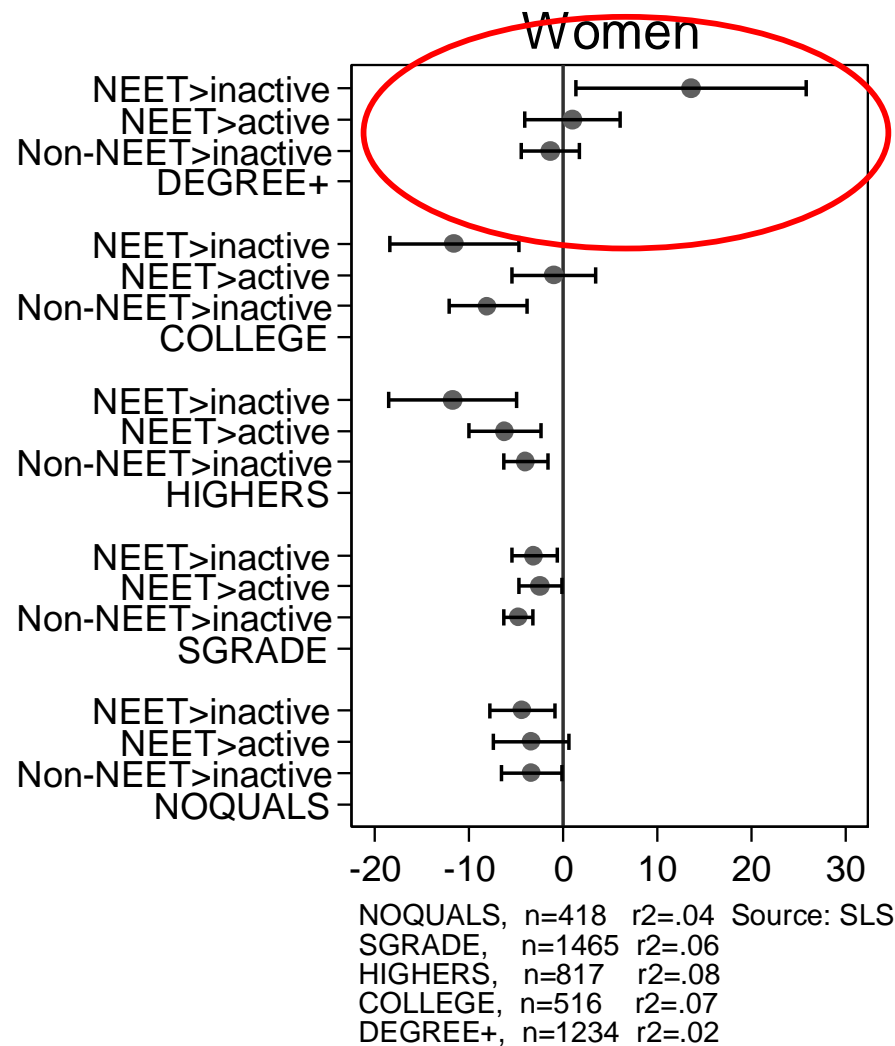
r²=0.35, n=3737, Source: SLS

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

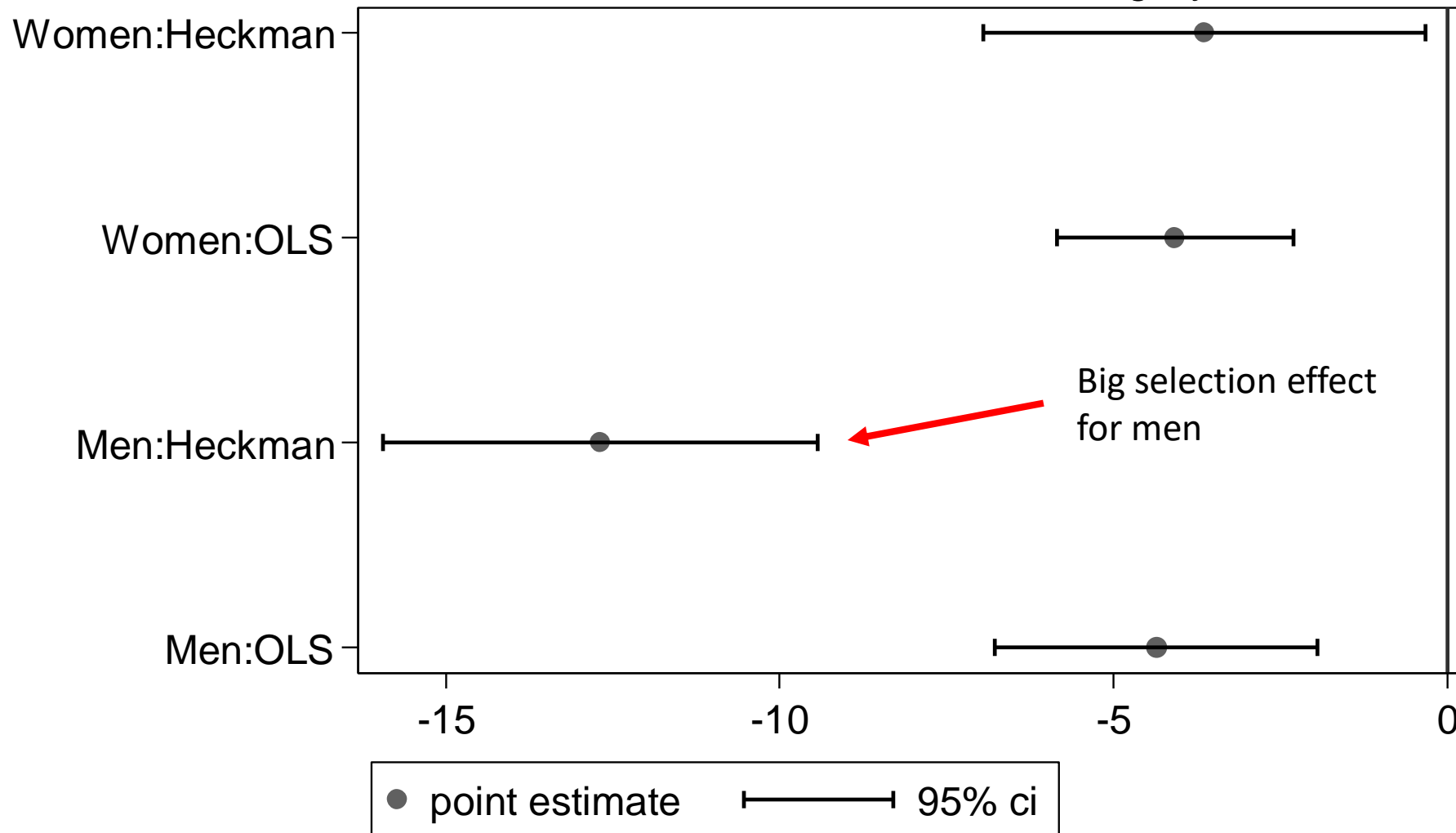
- Gregg and Tominey (2001) used selection to examine the relationship between youth unemployment and adult unemployment, NCDS.
- They also use an instrumental variable approach:
 - *'the unemployment rate prevalent locally for individuals aged 16 is used to instrument youth unemployment in the wage equation 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



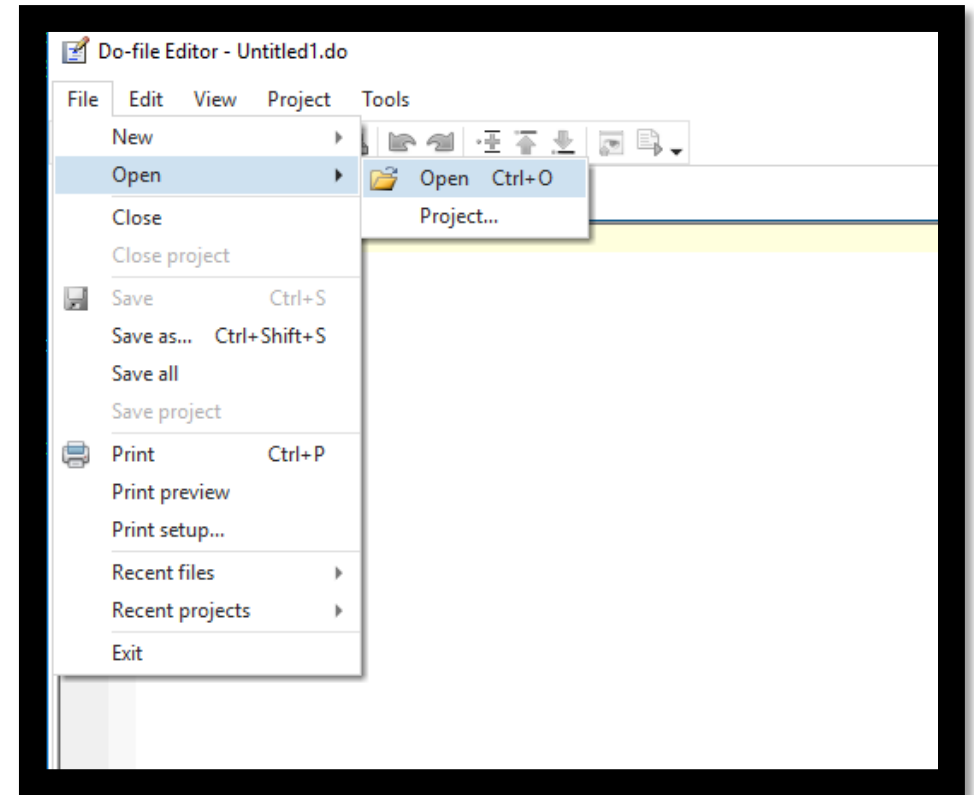
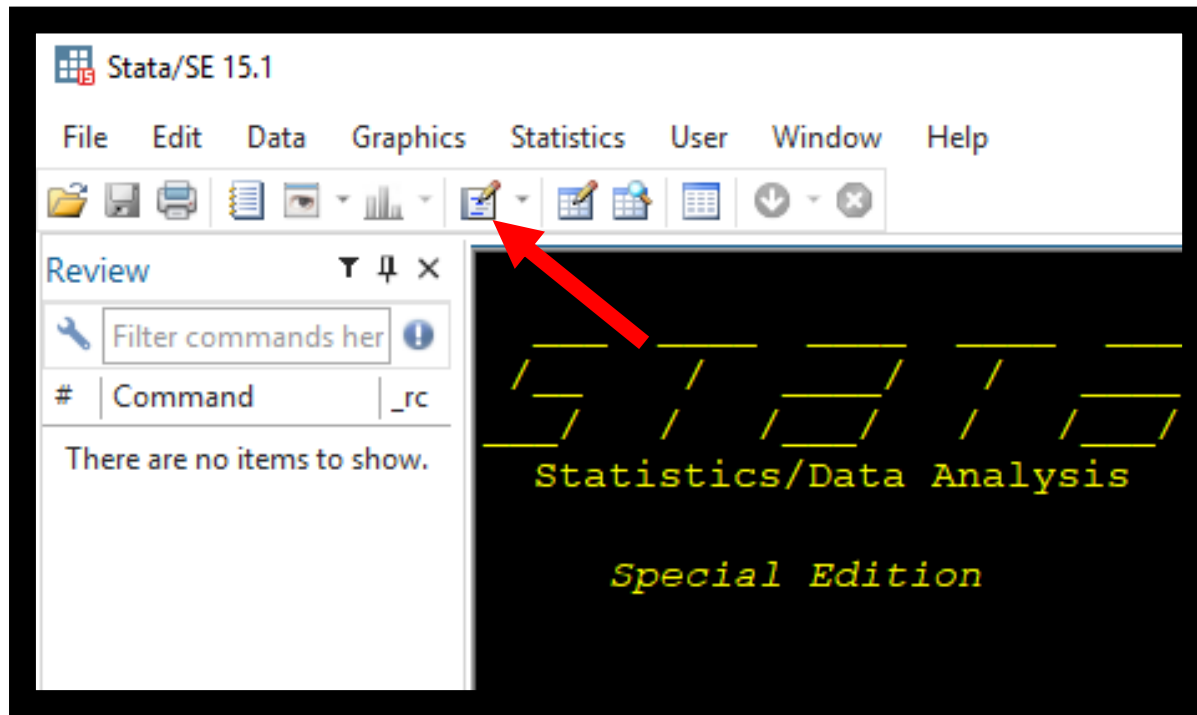
Source: SLS

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

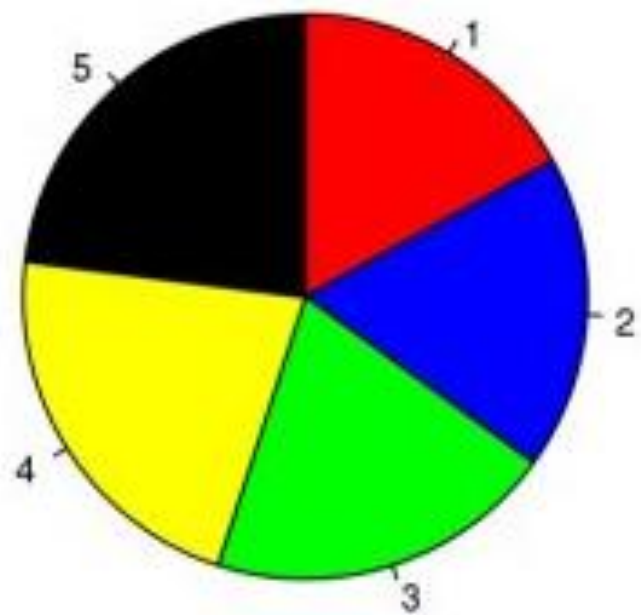


Publication Ready Tables

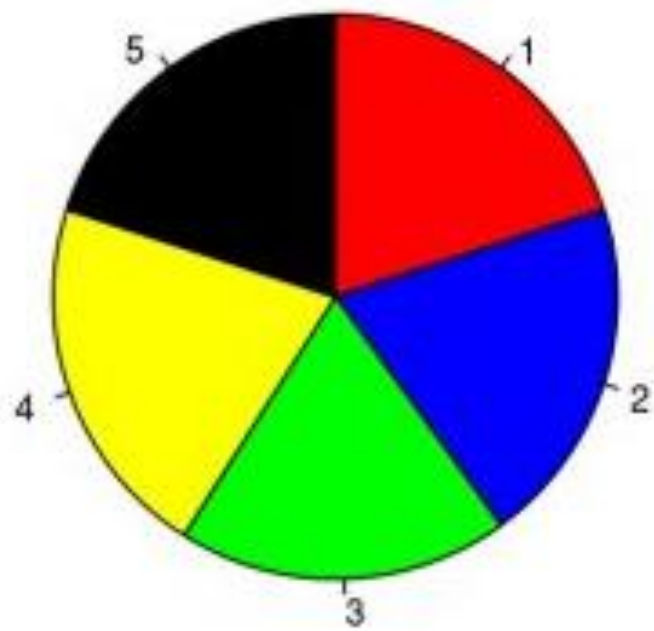
Publication Reading Tables

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

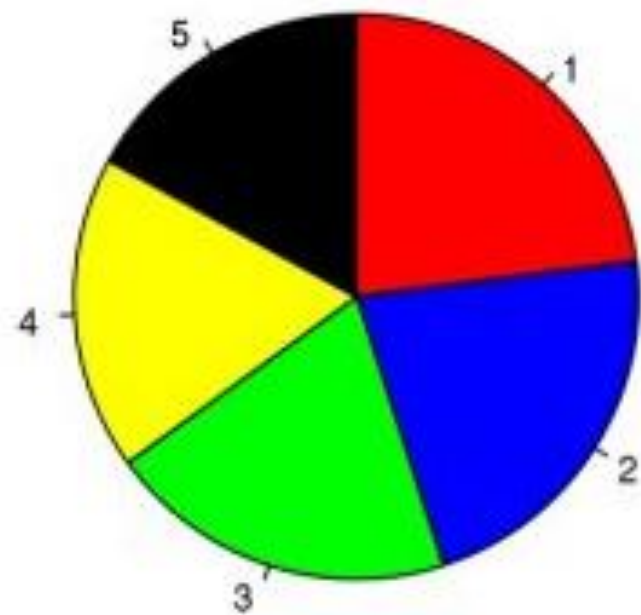
A

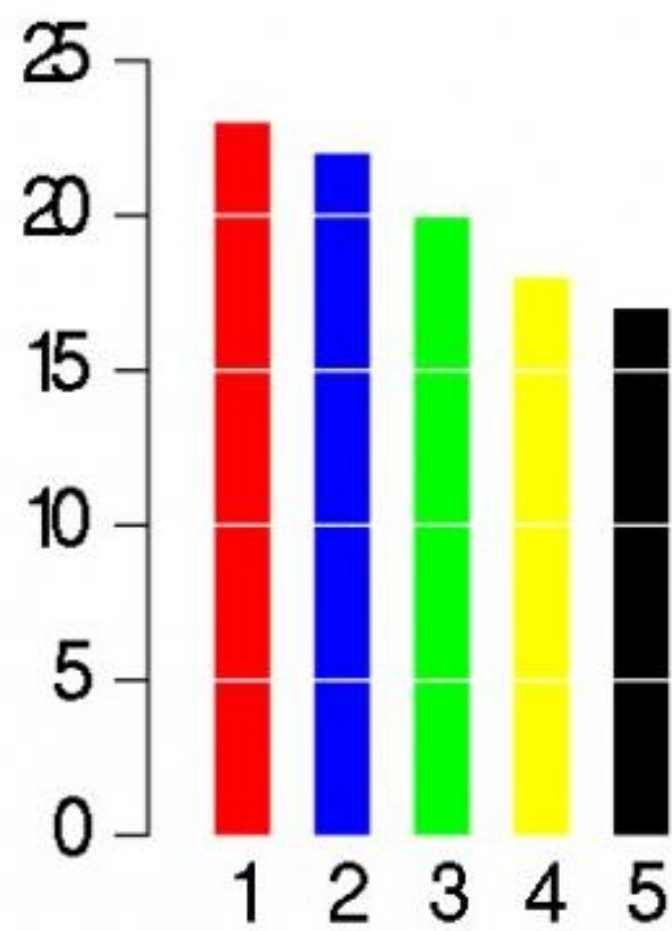
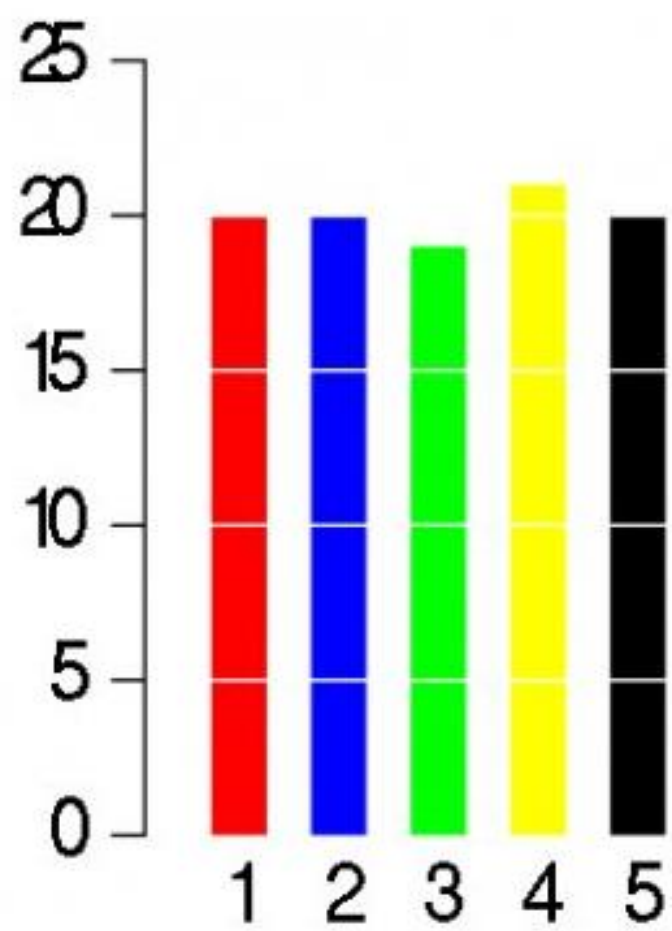
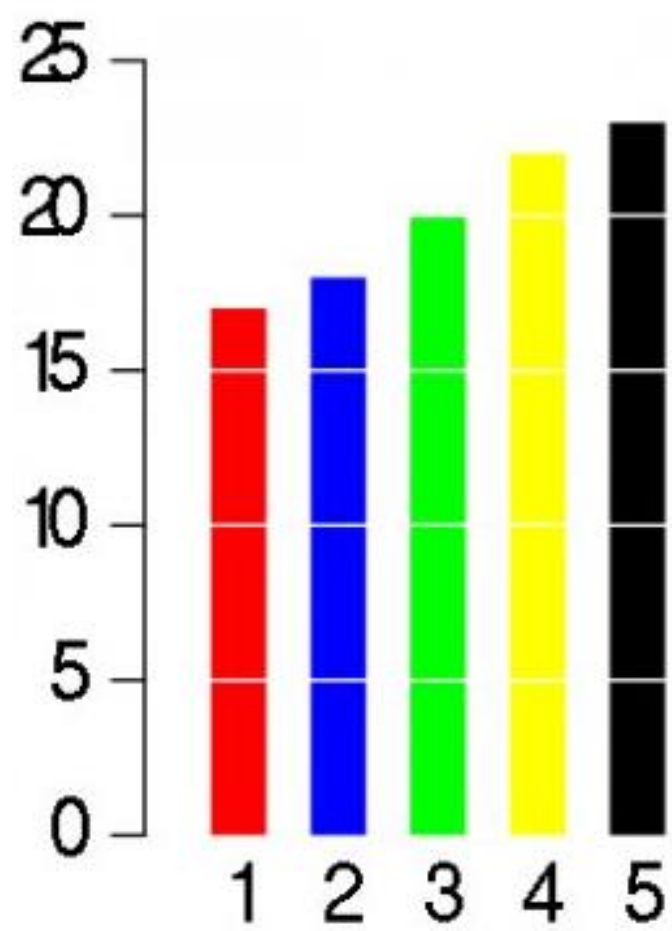


B



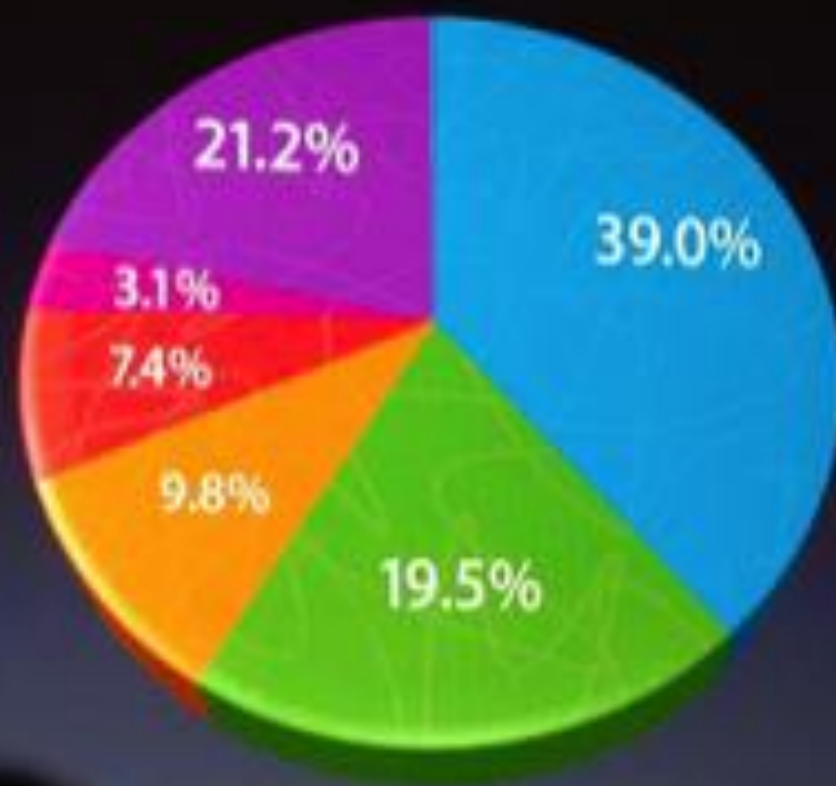
C





U.S. SmartPhone Marketshare

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





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



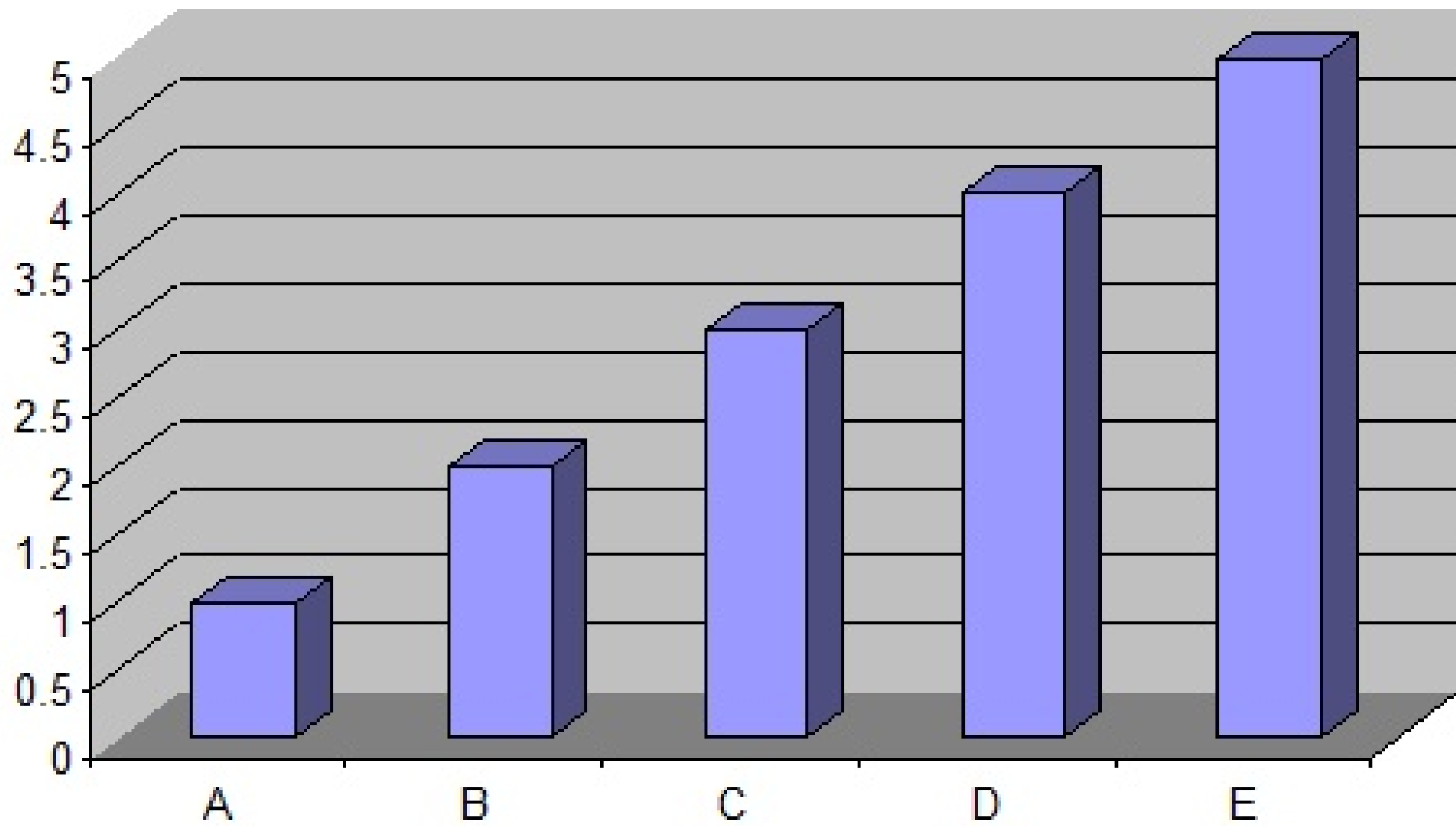
10

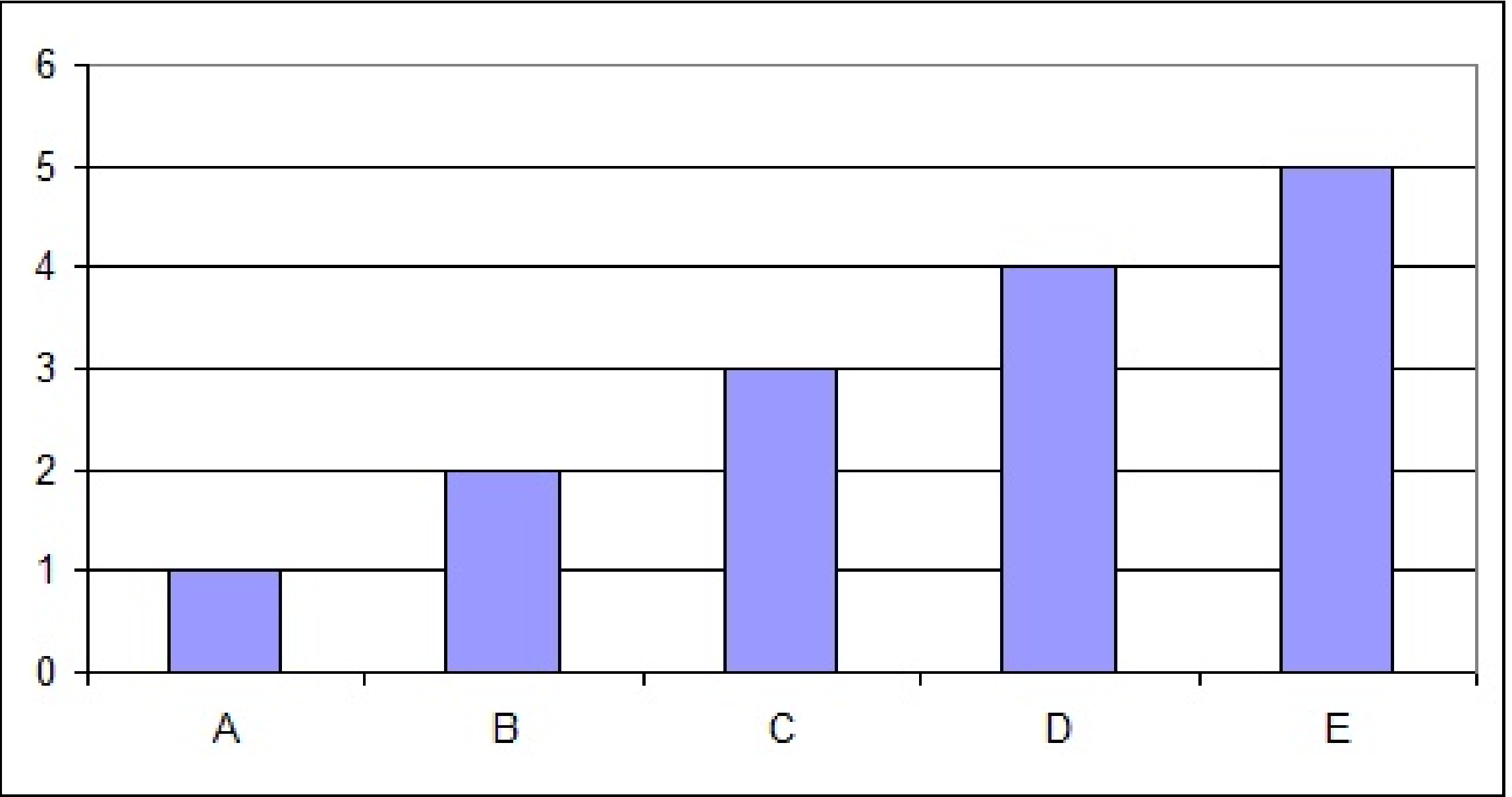


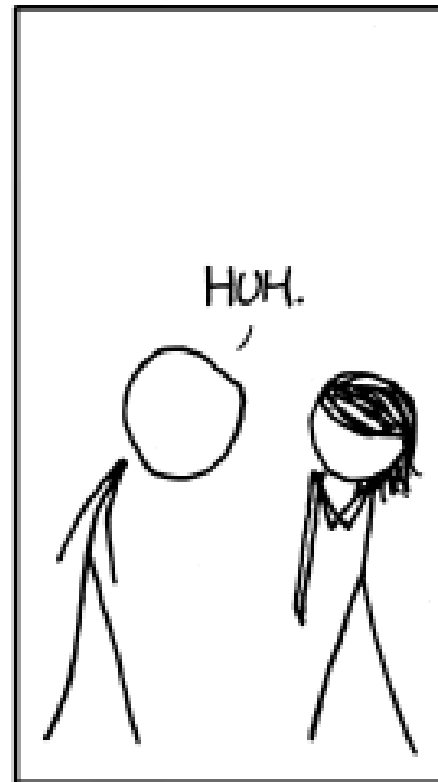
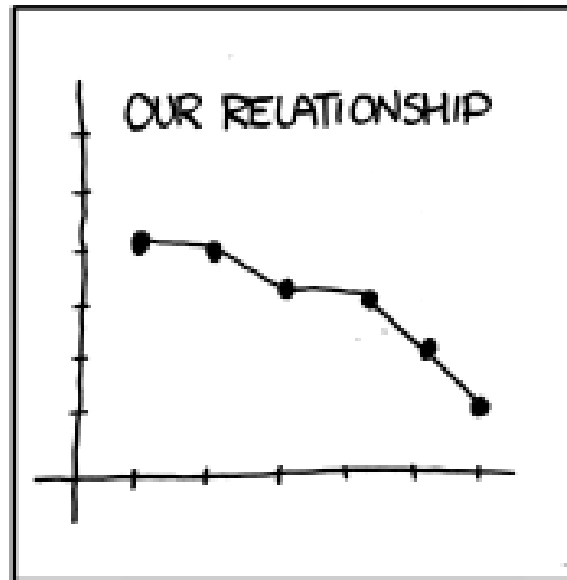
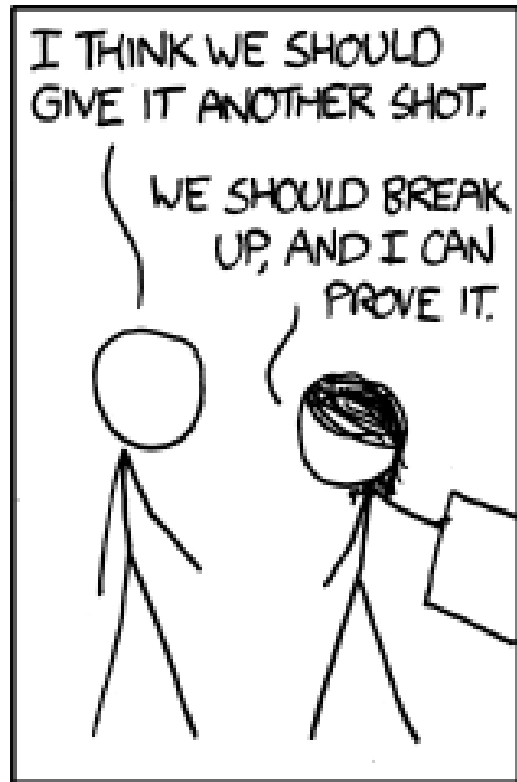
133



67







Effective Graphs

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

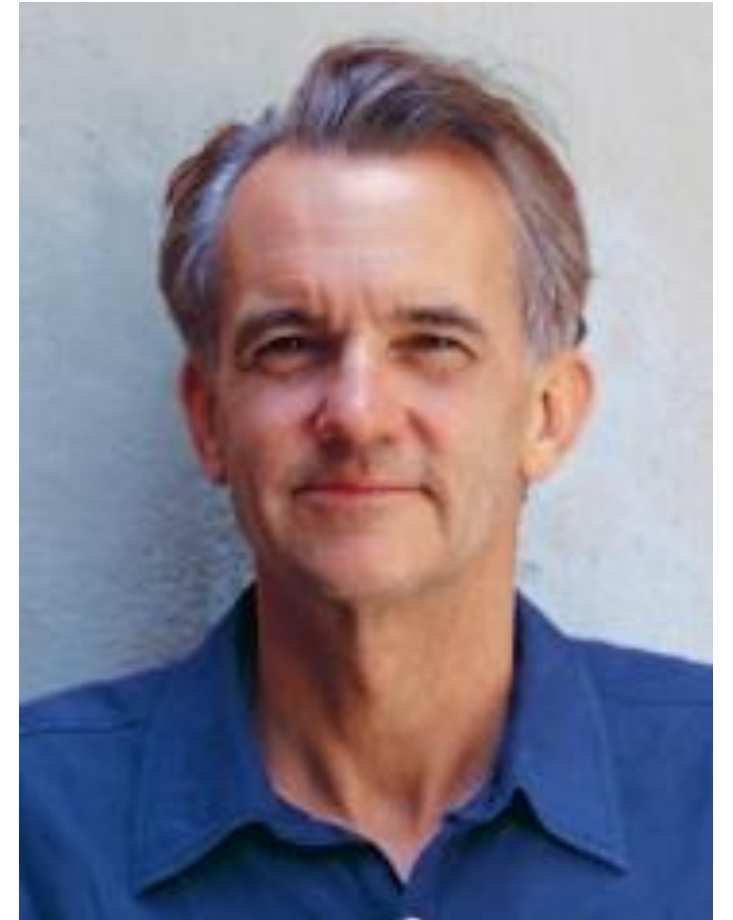
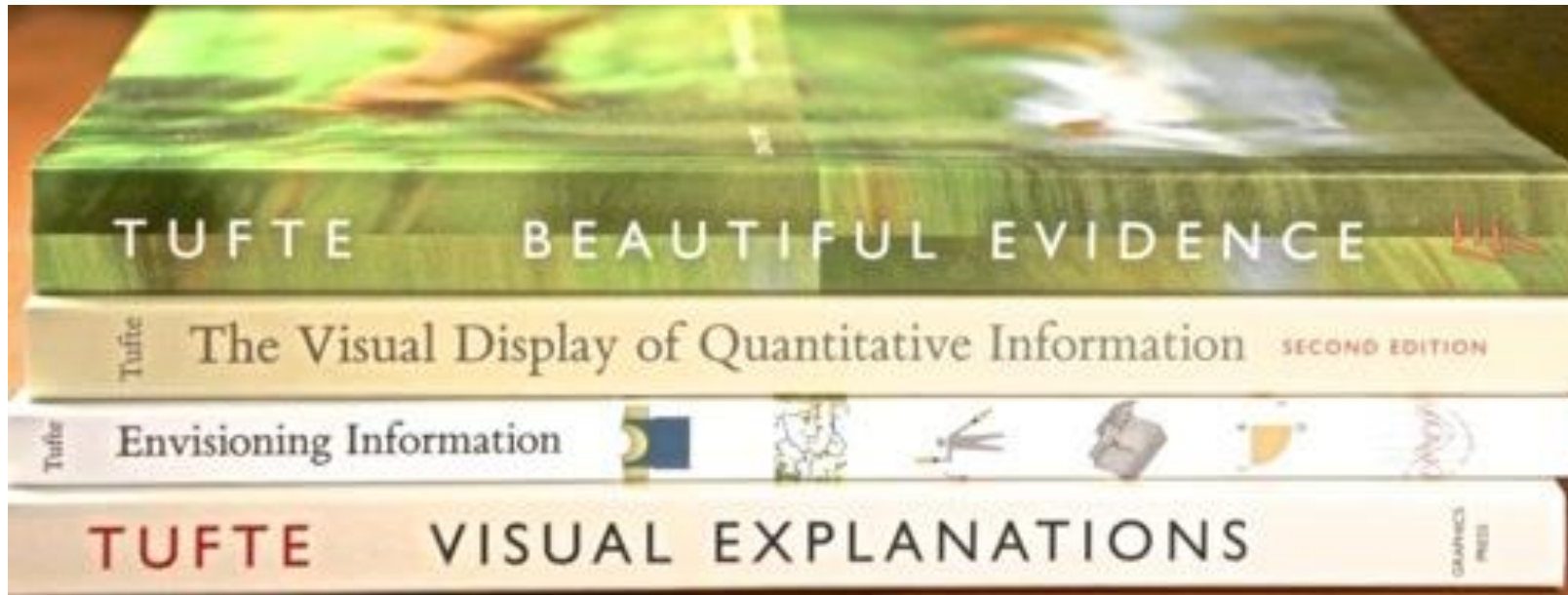


Table 1: OLS Regression model of maths test scores.

	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	

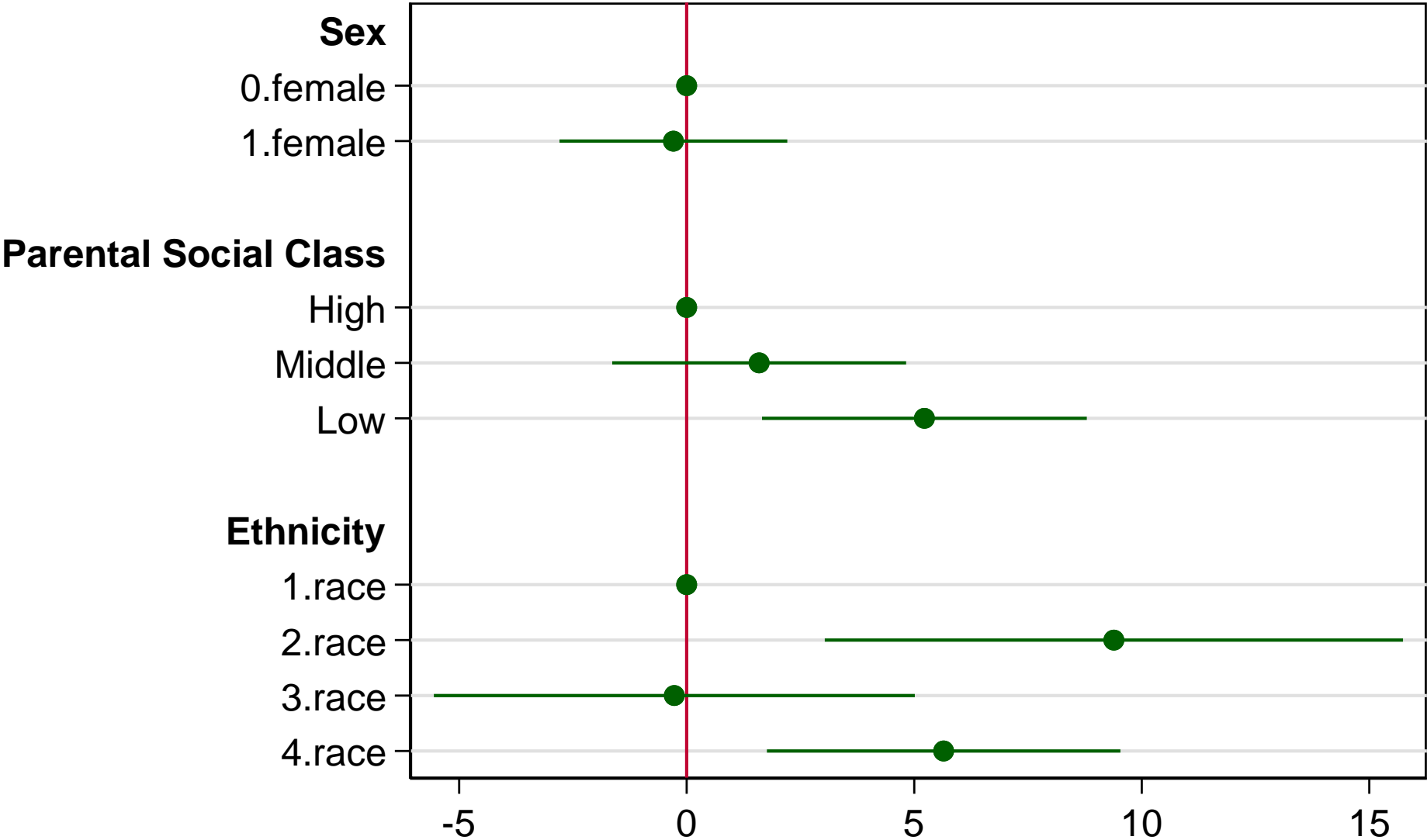
Data: High School and Beyond Survey

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



OLS Regression Model of Maths Test Scores

Coefficients and 95% Standard Errors

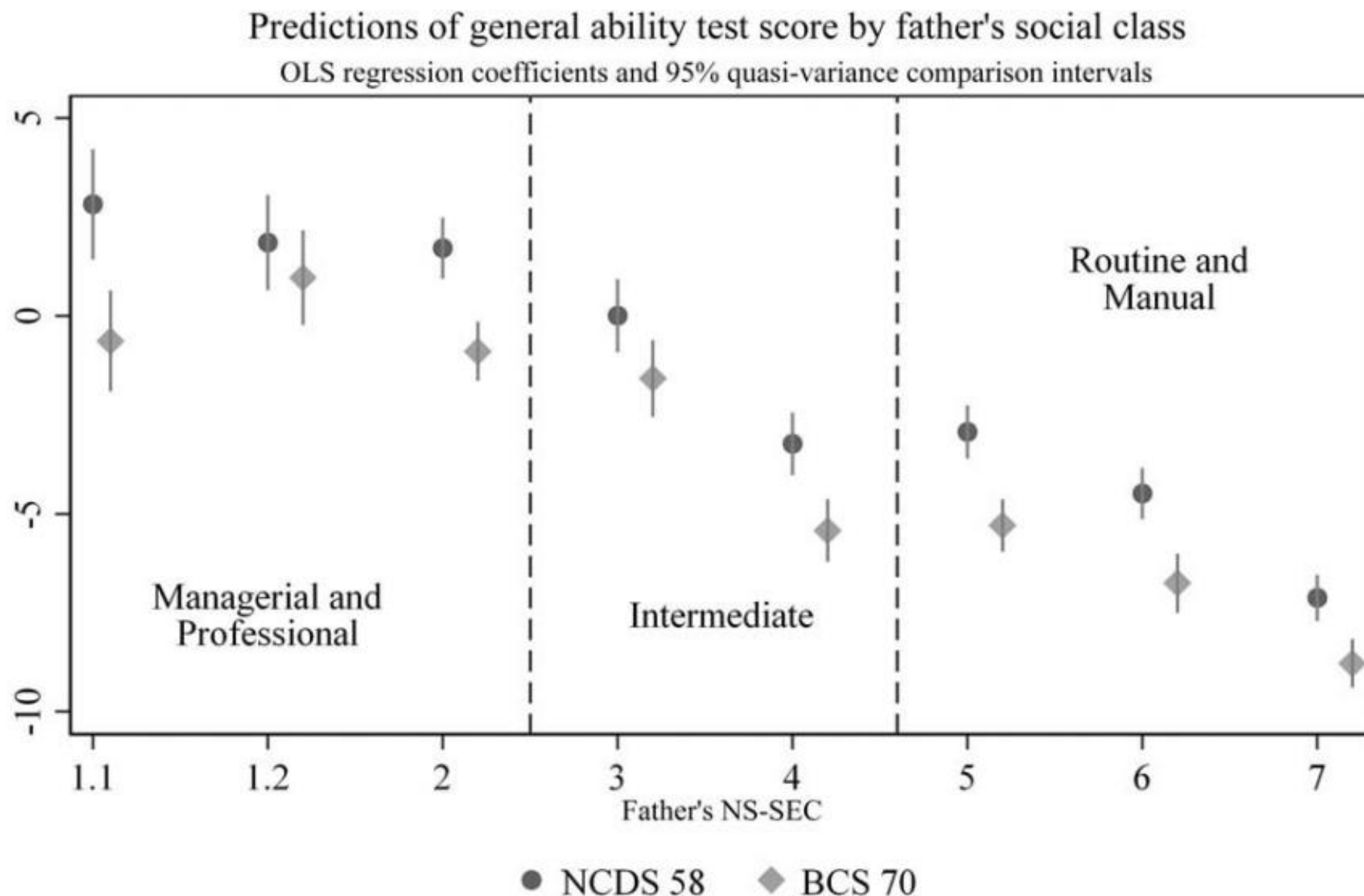


Data: High School and Beyond Data, n = 200, Adjusted R2 = 0.12

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. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

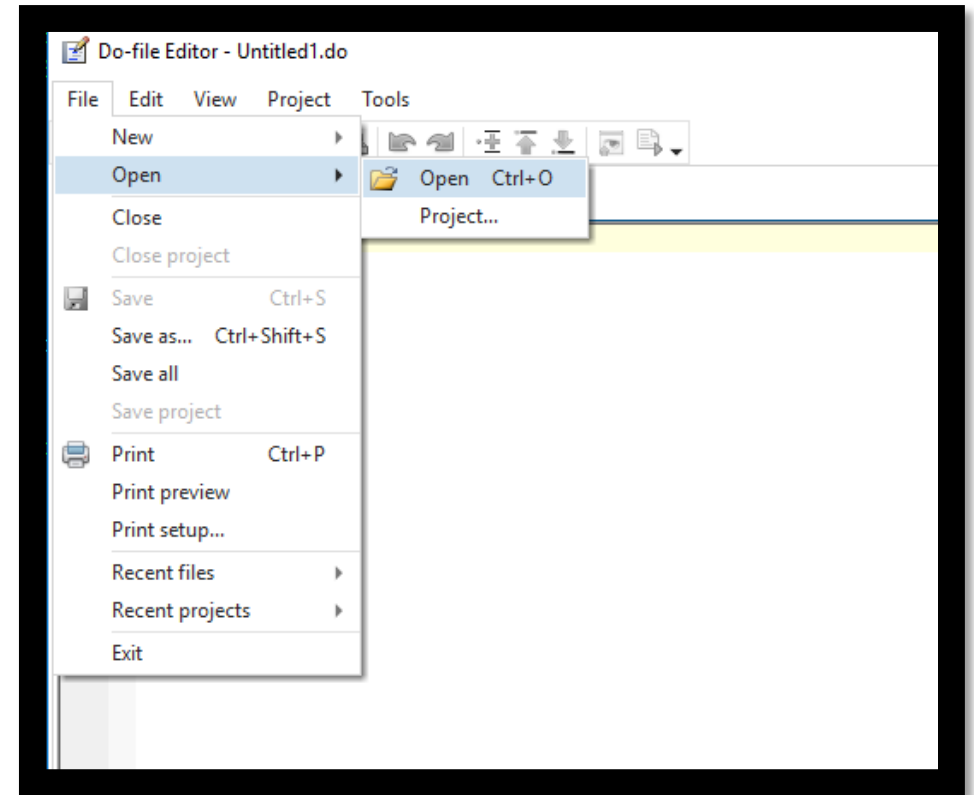
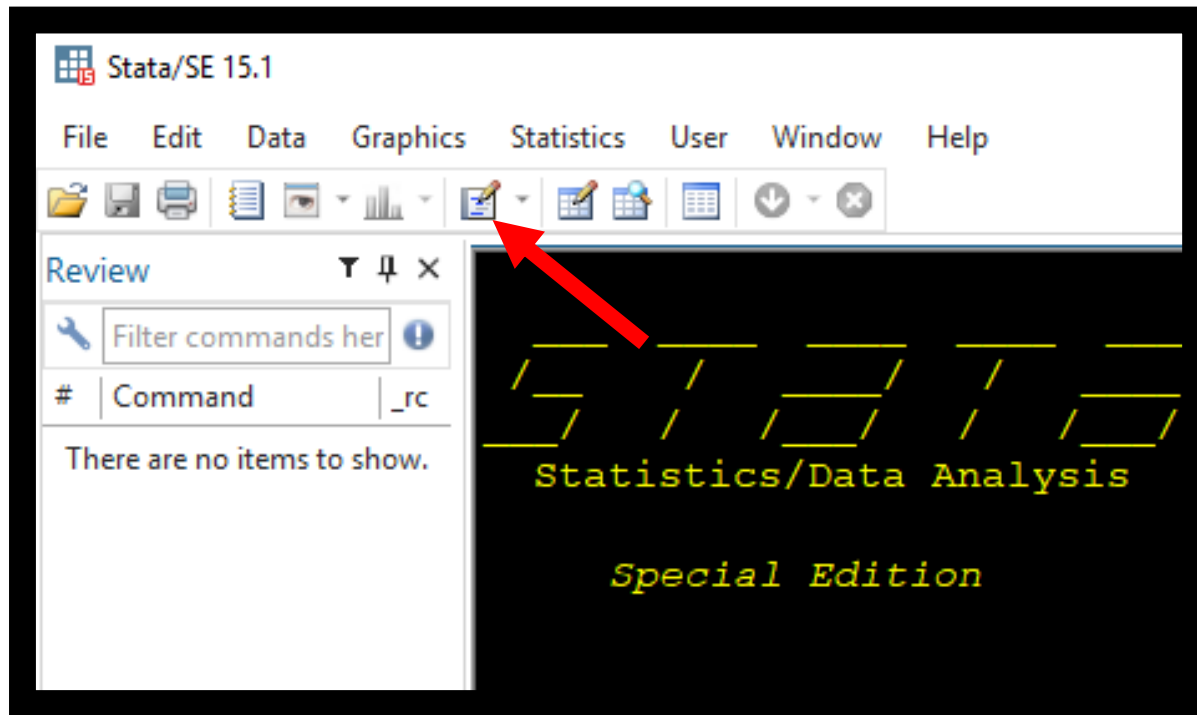


Note: Estimates are taken from Table IV, model 2. Model also contains gender and parent's highest education.

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

