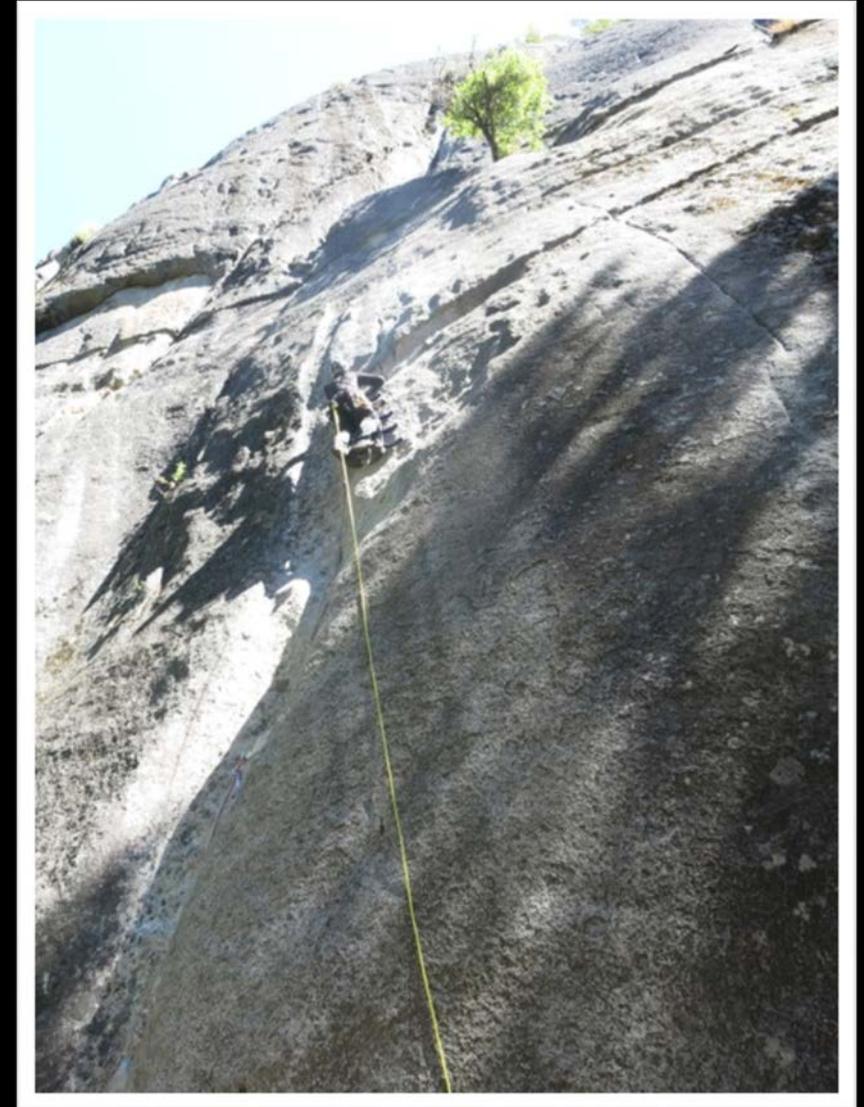


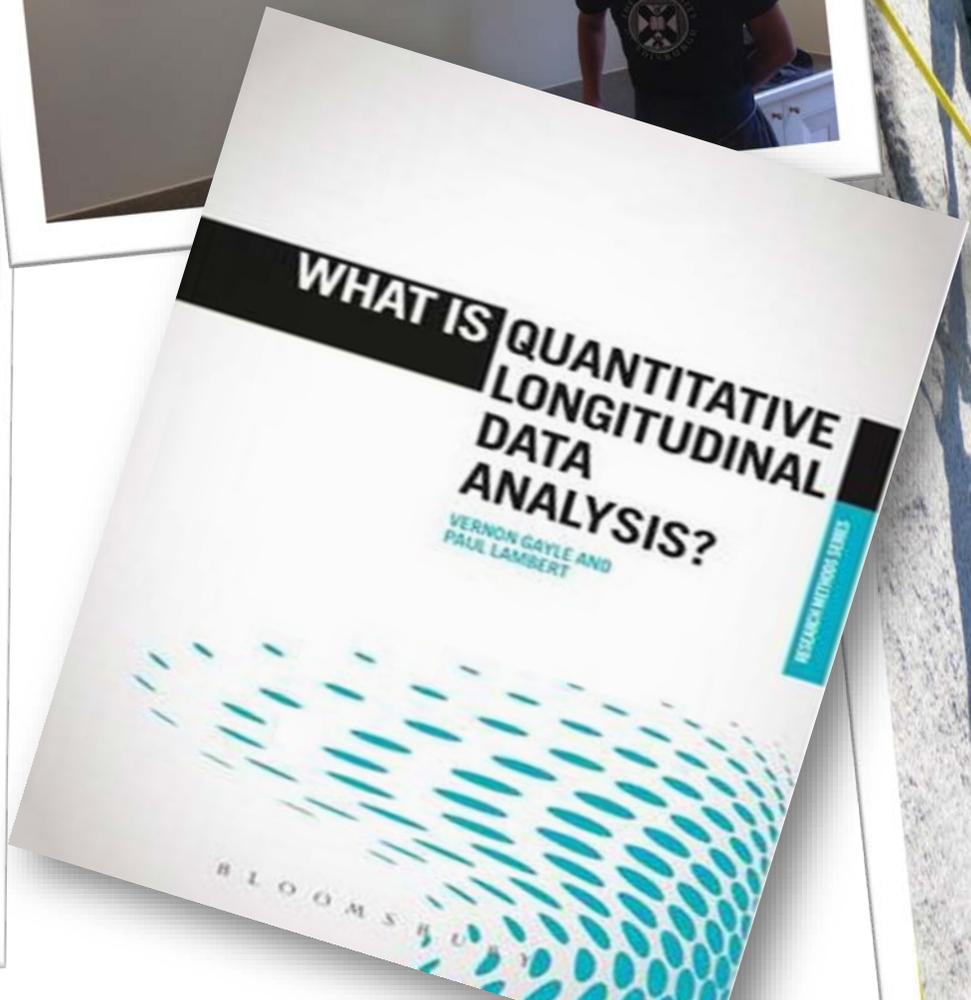
Why the Paper is Just a Palimpsest?

An appeal for reproducible
sociological research

Q – Step Edinburgh, December 2016

Vernon Gayle
University of Edinburgh
@profbigvern





A Thought Experiment

You are examining a PhD thesis or refereeing a paper

You turn the page and read the following output...

A Thought Experiment

A logistic regression model of some data from

The Youth Cohort Study of England and Wales (1997)

A Thought Experiment

The outcome is passing 5+ GCSEs at grades A*- C

The explanatory variables are

Gender, Ethnicity and Parental Social Class

n = 19,216

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	-0.039	0.392	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.80			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	-0.039	0.392	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.80			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	-0.039	0.392	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	-0.039	0.392	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	-0.392	0.039	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	-0.392	0.039	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	0.392	0.039	10.066	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
Indian	1.028	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

**The final set of results looked highly plausible.
Alarmingly the results were produced on data
that are completely fabricated!**

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	0.392	0.039	10.006	0.000
Boys	0.000	.	.	.
Chinese	1.719	0.287	5.988	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Black	0.000	.	.	.
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000	.	.	.
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*- C)				
Girls	0.405	0.039	10.305	0.000
Boys	0.000	.	.	.
Chinese	2.002	0.377	5.306	0.000
Indian	1.066	0.208	5.117	0.000
White	0.643	0.171	3.757	0.000
Bangladeshi	0.766	0.345	2.222	0.026
Pakistani	0.531	0.245	2.169	0.030
Black	0.000	.	.	.
Professional/Non-Man	2.192	0.109	20.179	0.000
Other Non-Manual	1.773	0.108	16.423	0.000
Skilled Manual	0.932	0.104	8.954	0.000
Semi-Skilled Manual	0.576	0.113	5.112	0.000
Unskilled	0.000	.	.	.
Constant	-2.208	0.198	-11.152	0.000
n	12789			

Produced in Stata using `svy`; Connolly (2006) used SPSS with data weighted incorrectly!

Appendix: details of binary logistic regression models derived from the three cohorts (see Table 3)

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A* - C)				
Girls	0.405	0.039	10.305	0.000
Boys	0.000	.	.	.
Chinese	2.002	0.377	5.306	0.000
Indian	1.066	0.208	5.117	0.000
White	0.643	0.171	3.757	0.000
Bangladeshi	0.766	0.345	2.222	0.026
Pakistani	0.531	0.245	2.169	0.030
Black	0.000	.	.	.
Professional/Non-Manual	2.192	0.109	20.179	0.000
Other Non-Manual	1.773	0.108	16.423	0.000
Skilled Manual	0.932	0.104	8.954	0.000
Semi-Skilled Manual	0.576	0.113	5.112	0.000
Unskilled	0.000	.	.	.
Constant	-2.208	0.198	-11.152	0.000
n	12789			

Produced in Stata using [svy](#); Connolly (2006) used SPSS with data weighted incorrectly!

Table 5. Binary logistic regression on whether school leavers in England and Wales in 1997 gained five or more GCSE grades A*-C or not¹

	B	S.E.	Wald	df	Sig.	Exp(B)
<i>Gender</i> ²						
Girls	0.405	0.038	114.340	1	<.001	1.499
<i>Ethnicity</i> ³						
Chinese	2.002	0.341	34.436	1	<.001	7.406
Indian	1.066	0.193	30.389	1	<.001	2.903
White	0.643	0.159	16.372	1	<.001	1.902
Bangladeshi	0.766	0.332	5.330	1	.021	2.151
Pakistani	0.531	0.230	5.338	1	.021	1.701
<i>Social Class</i> ⁴						
Professional/Managerial	2.192	0.110	396.863	1	<.001	8.954
Other Non-Manual	1.773	0.110	261.000	1	<.001	5.886
Skilled Manual	0.932	0.107	76.255	1	<.001	2.540
Semi-Skilled Manual	0.576	0.115	24.965	1	<.001	1.779
Constant	-2.208	0.189	136.885	1	<.001	0.110

¹Source of data: Secondary analysis of data derived from first sweep of Cohort 9 of the Youth Cohort Study of England and Wales.

²Reference category: boys.

³Reference category: Black pupils.

⁴Reference category: Unskilled Manual Occupations.

Connolly, Paul. "The effects of social class and ethnicity on gender differences in GCSE attainment: a secondary analysis of the Youth Cohort Study of England and Wales 1997–2001." *British Educational Research Journal* 32.1 (2006): 3-21.

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*-C)				
Girls	0.392	0.039	10.066	0.000
Boys	0.000
Chinese	1.719	0.287	5.988	0.000
Indian	0.928	0.210	4.899	0.000
White	0.997	0.160	6.241	0.000
Bangladeshi	1.351	0.317	4.267	0.000
Pakistani	0.677	0.209	3.241	0.001
Black	0.000
Prof/Managerial	1.945	0.078	25.093	0.000
Non-Manual(other)	1.335	0.085	15.687	0.000
Skilled Manual	0.779	0.079	9.908	0.000
Semi-skilled	0.310	0.086	3.611	0.000
Unskilled	0.000
Constant	-0.958	0.175	-5.475	0.000
Pseudo R ²	0.08			
n	19216			

Are we really examining or refereeing what we think we are?



- HOME
- ABOUT US
- CREATE & MANAGE DATA
- DEPOSIT DATA
- HOW WE CURATE DATA
- FIND DATA
- NEWS & EVENTS

CELEBRATING OUR HISTORY

As the University of Essex celebrates 50 years, take a look back at the history of the Archive

READ ON



- What's new
- Data lifecycle
- Find data
- Who are we?

DEPOSITING YOUR DATA

Depositing your data with the Archive ensures that they will be professionally curated and accessible...

DEPOSIT DATA

FINDING DATA TO USE

We can help you find data for research and teaching with our catalogue of over 5,000 data collections

OUR CATALOGUE

HOW WE CURATE DATA

We follow best practices in preparing and curating our data to ensure usability

DATA CURATION



OUR SERVICES

UK Data Service data promoting evidence-based research

VIEW SITE



FIRST TIME HERE?

HELPFUL INFORMATION

A QUICK GUIDE TO THE ARCHIVE

3 of 10: Each of our data collections has a unique persistent identifier (DOI) that makes it easier to find and cite

WHO GIVES US DATA?

Find out what kind of data are available to you

LATEST NEWS & EVENTS

A+ Futuretrack: a new longitudinal study of applicants to higher education
The UK Data Service is pleased to announce the ...

ESSEX SUMMER SCHOOL Social Science Summer School
The UK Data Service has been teaching students ...



Edinburgh Research Explorer

University Homepage > Research > Explorer home

- Explorer home
- Staff
- Research projects
- Research outputs
- Research activities
- Colleges & Schools
- Research press coverage
- Datasets



Search the Research Explorer

Media enquiries

Commercial enquiries

Staff

Meet our international academics whose visions are shaping tomorrow's world

Staff

Projects

Explore the pioneering research projects that make us a world-leading centre of excellence.

Projects

Activities

Search our research activities and events, and see the awards conferred on our researchers.

Activities

Research outputs

Discover our influential reputation through our research outputs and publications.

Outputs

Research press coverage

ECCEI to advise on carbon capture methods in Tees Valley
12/08/16:
Stuart Haszeldine

300 volunteers to take part in whisky study lead by Dr Adam Moore
11/08/16:
Adam Moore

Cancer drug for mums-to-be may curb baby girls' future fertility, finds Prof Norah Spears
11/08/16:
Norah Spears

Dr Claudi Pagliari highlights hidden privacy breaches in OTC genetic testing kits
11/08/16:
Claudia Pagliari

Li-Fi, invented by Prof Harald Haas, could help connectivity in India
11/08/16:

Access to Data



THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES

[HOME](#) [ABOUT US](#) [CREATE & MANAGE DATA](#) [DEPOSIT DATA](#) [HOW WE CURATE DATA](#) [FIND DATA](#) [NEWS & EVENTS](#)

CELEBRATING OUR HISTORY

As the University of Essex celebrates 50 years, take a look back at the history of the Archive

[READ ON](#)

[What's new](#) [Data lifecycle](#) [Find data](#) [Who are we?](#)

DEPOSITING YOUR DATA

Depositing your data with the Archive ensures that they will be professionally curated and accessible

[DEPOSIT DATA](#)

HOW WE CURATE DATA

We follow best practices in preparing and curating our data to ensure usability

[DATA CURATION](#)

A BLACK BOX

Lack of access to the research code that produced the research output

Command files
Syntax files
.do Stata files
R scripts
.sps SPSS files
Jupyter notebooks



Edinburgh Research Explorer

[Research press coverage](#) [Datasets](#)

Search the Research Explorer

[Media](#)

[Inquiries](#)

[Commercial](#)

[Inquiries](#)

Research press coverage

DCI to advise on carbon capture methods in Tees Valley
08/16
[Geart Haszeldine](#)

10 volunteers to take part in whisky study lead by Dr Adam Moore
08/16
[Adam Moore](#)

Cancer drug for mums-to-be may curb baby girls' future fertility, finds Prof Sarah Spears
08/16
[Sarah Spears](#)

Prof Claudi Pagliari highlights hidden privacy breaches in OTC genetic testing kits
08/16
[Claudia Pagliari](#)

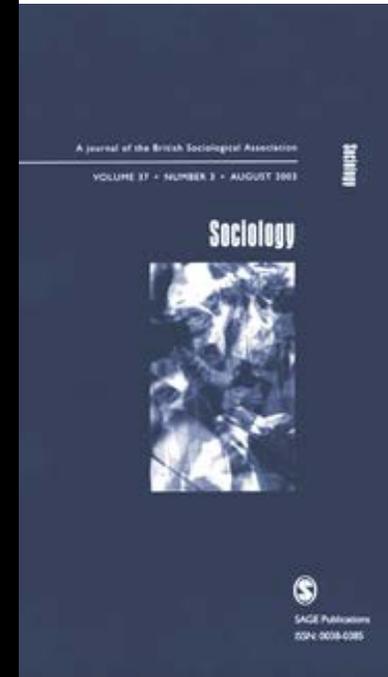
FI, invented by Prof Harald Haas, could help connectivity in India
08/16
[Harald Haas](#)



A BLACK BOX

???

[How was NVivo used]





THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES

- HOME
- ABOUT US
- CREATE & MANAGE DATA
- DEPOSIT DATA
- HOW WE CURATE DATA
- FIND DATA
- NEWS & EVENTS

CELEBRATING OUR HISTORY

As the University of Essex celebrates 50 years, take a look back at the history of the Archive

[READ ON](#)



DEPOSITING YOUR DATA

Depositing your data with the Archive ensures that they will be professionally curated and accessible

[DEPOSIT DATA](#)

FINDING DATA TO USE

We can help you find data for research and teaching with our catalogue of over 6,000 data collections

[OUR CATALOGUE](#)

HOW WE CURATE DATA

We follow best practices in preparing and curating our data to ensure usability

[DATA CURATION](#)

OUR SERVICES

[UK Data Service](#) data promoting evidence-based research

[VIEW SITE](#)

A BLACK BOX

Historically lack of space in paper journals



A parchment or other writing surface on which the original text has been effaced or partially erased, and then overwritten by another; OED

palimpsest, *n.* and *adj.*

Pronunciation: Brit. /'palm(p)sɛst/ , U.S. /'pæləm(p),sɛst/

Young people and school General Certificate of Secondary Education attainment: looking for the ‘missing middle’

Vernon Gayle^{a*}, Susan Murray^b and Roxanne Connelly^c

^a*School of Social and Political Science, University of Edinburgh, Edinburgh, UK;*

^b*School of Applied Social Science, University of Stirling, Stirling, UK;* ^c*Centre for Longitudinal Studies, Institute of Education, University of London, London, UK*

(Received 6 July 2013; final version received 12 June 2014)

In Britain, educational qualifications gained at school continue to play an important and central role in young people’s educational and employment pathways. Recently there has been growing interest in documenting the lives of ‘ordinary’ young people. In this paper we analyse the Youth Cohort Study of England and Wales in order to better document the experiences of those with ‘middle’ levels of school General Certificate of Secondary Education (GCSE) attainment. The overall pattern of school GCSE attainment is one of increasing levels of performance. GCSE attainment is still highly stratified. Girls performed better than boys, and there were some marked differences in attainment for pupils from the main minority ethnic groups. Most notably, parental socio-economic positions are the most important factor. The analyses fail to persuade us that there are clear boundaries that demark a ‘middle’ category of school GCSE attainment. We conclude that sociologists should study ‘ordinary’ young people; however, school GCSE attainment is best understood as a continuum, and measures such as the number of GCSEs or point scores are preferable.

Keywords: youth transitions; sociology of youth; educational attainment; General Certificate of Secondary Education; missing middle; Youth Cohort Study of England and Wales

Introduction

The idea of a ‘marginalised mainstream’ or ‘missing middle’ is currently

```
capture drop t03cat
gen t03cat = .
replace t03cat = 0 if (t0examac2==0)
replace t03cat = 1 if (t0examac2>=1)&(t0examac2<=4)
replace t03cat = 2 if (t0examac2>=5)&(t0examac2<=14)
tab t03cat, missing

label define t03catll 0 "none" 1 "1-4" 2 "5+"
label values t03cat t03catll

* first model mlogit

svyset [pw=t1weight]

svy:mlogit t03cat ///
    cohort93 cohort95 cohort97 cohort99 cohort01 boys ///
    black indian pakistani bangladeshi oasian other ///
    rented ohouse ///
    mumonly dadonly ohh ///
    gradpar ///
    nssec11 nssec12 nssec2 nssec4 nssec5 nssec6 nssec7 ///
    ,baseoutcome(1)

estimates store mlogit1

*****
```

Young people and school General Certificate of Secondary Education attainment: looking for the ‘missing middle’

Vernon Gayle^{a*}, Susan Murray^b and Roxanne Connelly^c

^aSchool of Social and Political Science, University of Edinburgh, Edinburgh, UK; ^bSchool of Applied Social Science, University of Stirling, Stirling, UK; ^cCentre for Longitudinal Studies, Institute of Education, University of London, London, UK

(Received 6 July 2013; final version received 12 June 2014)

replace t03cat = 2 if (t0examac2 >= 5) & (t0examac2 <= 14)

tab t03cat, missing

label define t03cat 0 "none" 1 "1-4" 2 "5+"

label values t03cat t03cat

* first model mlogit

svyset [pw=t1weight]

svy:mlogit t03cat //

cohort93 cohort99 cohort07 cohort09 cohort11 boys //

black indian pakistani bangladeshi asian other //

rented_ohouse //

numonly hadonly ohh //

gradpar //

nssec11 nssec12 nssec2 nssec4 nssec5 nssec6 nssec7 //

basenutcmesh

In Britain, educational qualifications gained at school continue to play an important and central role in young people’s educational and employment pathways. Recently there has been growing interest in documenting the lives of ‘ordinary’ young people. In this paper we analyse the Youth Cohort Study of England and Wales in order to better document the experiences of those with ‘middle’ levels of school General Certificate of Secondary Education (GCSE) attainment. The overall pattern of school GCSE attainment is one of increasing levels of performance. GCSE attainment is still highly stratified. Girls performed better than boys, and there were some marked differences in attainment for pupils from the main minority ethnic groups. Most notably, parental socio-economic positions are the most important factor. The analyses fail to persuade us that there are clear boundaries that demark a ‘middle’ category of school GCSE attainment. We conclude that sociologists should study ‘ordinary’ young people; however, school GCSE attainment is best understood as a continuum, and measures such as the number of GCSEs or point scores are preferable.

Keywords: youth transitions; sociology of youth; educational attainment; General Certificate of Secondary Education; missing middle; Youth Cohort Study of England and Wales

estimates store mlogit1

Introduction

The idea of a ‘marginalised mainstream’ or ‘missing middle’ is currently



Reproducing Results

'today it's impossible to verify most of the computational results that scientists present at conferences and in papers'

The Yale Law School Roundtable on Data and Code Sharing (2010: 8)

Sharing Research Code

Sufficient information to enable other researchers to

understand

evaluate

build upon the work

Sharing Research Code

Sufficient information to enable other researchers to understand, evaluate and build upon the work

Enough information for a third party to reproduce (i.e. duplicate) results without needing to get additional information from the authors

Sharing Research Code

Sufficient information to enable other researchers to understand, evaluate and build upon the work

Enough information for a third party to reproduce results without needing to get additional information from the authors

Required by *Science*, *Am. Econ. Rev.*, *Econometrica*, *Rev. Econ. Studies*

500 journals signed up to the Transparency and Openness Promotion (TOP) Guidelines

Some Existing Examples of Code Sharing in Social Stratification Research

Harry Ganzeboom

<http://www.harryganzeboom.nl/index.htm>

Don Treiman

<https://ccpr.ucla.edu/dtreiman/>

Paul Lambert

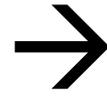
<http://www.camsis.stir.ac.uk/>

Kim Weedon

<http://www.kimweeden.com/research/occupations/>

“Show me your working out”

Show me

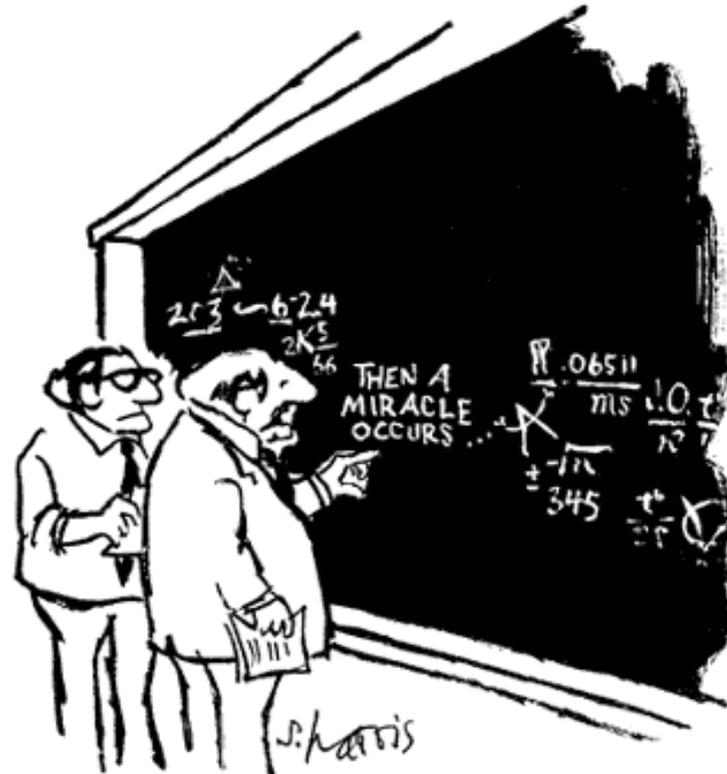


Trust me



“Show me your working out”

The Royal Society's motto '*Nullius in verba*' (take nobody's word for it)



"I think you should be more explicit here in step two."

Duplication

Step 1

Work can be duplicated if sufficient information is made available which ensures consistent results can be produced using the same data and the same analytical techniques

Replication

Step 2

A replication study can 'duplicate' the original work....

But can also further test the robustness of the original work by employing new or additional data (or measures) and alternative data analysis techniques

Duplication

100 lines of code (Stata syntax) to enable the data

I know the YCS very well and know Stata

Barrier to duplication

Connolly, Paul. "The effects of social class and ethnicity on gender differences in GCSE attainment: a secondary analysis of the Youth Cohort Study of England and Wales 1997–2001."
" *British Educational Research Journal* 32.1 (2006): 3-21.

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A*- C)				
Girls	0.405	0.039	10.305	0.000
Boys	0.000	.	.	.
Chinese	2.002	0.377	5.306	0.000
Indian	1.066	0.208	5.117	0.000
White	0.643	0.171	3.757	0.000
Bangladeshi	0.766	0.345	2.222	0.026
Pakistani	0.531	0.245	2.169	0.030
Black	0.000	.	.	.
Professional/Non-Manual	2.192	0.109	20.179	0.000
Other Non-Manual	1.773	0.108	16.423	0.000
Skilled Manual	0.932	0.104	8.954	0.000
Semi-Skilled Manual	0.576	0.113	5.112	0.000
Unskilled	0.000	.	.	.
Constant	-2.208	0.198	-11.152	0.000
n	12789			

Produced in Stata using `svy`. Connolly (2006) used SPSS with data weighted incorrectly!

Replications

Logistic Regression 5+ GCSEs (A*-C) YCS Cohort 9

	b	se	t	p
5+ GCSEs (A* - C)				
Boys	0.000	.	.	.
Girls	0.405	0.039	10.305	0.000
White	0.000	.	.	.
Chinese	1.359	0.337	4.031	0.000
Indian	0.423	0.122	3.469	0.001
Bangladeshi	0.123	0.301	0.409	0.682
Pakistani	-0.112	0.177	-0.630	0.529
Black	-0.643	0.171	-3.757	0.000
Professional/Non-Manual	0.000	.	.	.
Other Non-Manual	-0.420	0.060	-6.962	0.000
Skilled Manual	-1.260	0.053	-23.558	0.000
Semi-Skilled Manual	-1.616	0.069	-23.543	0.000
Unskilled	-2.192	0.109	-20.179	0.000
Constant	0.627	0.047	13.209	0.000
n	12789			

Produced in Stata using [svy](#); Connolly (2006) used SPSS with data weighted incorrectly!

Logistic Regression 5+ GCSEs (A*- C) YCS Cohort 9

	Original b		NS-SEC b	
5+ GCSEs (A*- C)				
Boys	0.000		0.000	
Girls	0.405	***	0.434	***
White	0.000		0.000	
Chinese	1.359	***	1.491	***
Indian	0.423	***	0.598	***
Bangladeshi	0.123		0.320	
Pakistani	-0.112		0.208	
Black	-0.643	***	-0.715	***
Professional/Non-Manual	0.000			
Other Non-Manual	-0.420	***		
Skilled Manual	-1.260	***		
Semi-Skilled Manual	-1.616	***		
Unskilled	-2.192	***		
Large employers and higher managerial			0.000	
Higher professional occupations			0.371	***
Lower managerial & professional			-0.417	***
Intermediate occupations			-0.848	***
Small employers and own account workers			-1.437	***
Lower supervisory and technical occupations			-1.749	***
Semi-routine occupations			-1.944	***
Routine occupations			-2.361	***
Constant	0.627	***	0.716	***
n	12789		12788	

Produced in Stata using ~~svy~~; Connolly (2006) used SPSS with data weighted incorrectly!

Why bother?

- Improves transparency - don't just trust me – I will show you
- Allows others to understand, evaluate, and build upon the work
- Checks on accuracy
- Facilitates incremental development (and comparative work)

Are we serious about what we do?

(Edinburgh University's mission is the creation, dissemination and curation of knowledge)

The Horror of a Retraction...

The authors of a March 2015 *Journal of Health and Social Behavior* (JHSB) study, "In Sickness and in Health? Physical Illness as a Risk Factor for Marital Dissolution in Later Life" (2015, 56(1):59-73), have retracted the article.

There was a major error in the coding in their dependent variable of marital status. The conclusions of that study should be considered invalid.

A corrected version of the article will appear in the September 2015 issue of JHSB.

2011 Three Valleys, bottled March 2013
This, our only multi-vineyard wine, comes from three of the finest appellations in Sonoma County. As with our single-vineyard wines, the grapes are harvested by hand, fermented on native yeasts and aged in small cooperage. Cool summer weather delayed ripening; we began harvesting in mid-September, finishing in late October. We chose the most accessible lots from eight ranches for the 2011. This elegant wine has generous fruit, and will be most enjoyable over the next four to five years. JO (12/12)

Ingredients: Sustainably grown grapes; indigenous yeasts; naturally occurring malolactic bacteria; tartaric acid; egg whites; SO₂. PD (12/12)

750 mL
408.867.3233

© REGISTERED TRADEMARK
www.ridgevine.com

RIDGE PULL NO
PUNCHES WITH
THEIR LABELLING,
AND ARE ONE OF
THE FEW WINERIES
TO LIST ALL THE
"INGREDIENTS" ON
THE BACK LABEL



Private Reproducibility to Public Reproducibility

A small step from private reproducibility
could lead to a giant leap for public reproducibility

NEIL ARMSTRONG'S HISTORIC FIRST WORDS ON MOON: 'HOLY LIVING F * CK'

THE MOON—Just fucking
great.

The distant, lonely, mysterious
satellite that has fascinated
mankind since the dawn of time
is distant and lonely no more.

At 8:17 p.m. EST yesterday,
astronaut Neil Armstrong and
Edwin E. Aldrin Jr. touched
down on the Sea of Tranquility
in the lunar module Eagle and
returned back to Earth the mor-

ning in the moon's low-gravity
atmosphere, Aldrin joined his
colleague on the lunar surface.

"Hell, yeah. Hell, yeah," an
ecstatic Aldrin said, pumping
his space-gloved fist into the
thin air.

"It's like I told you on the way
up here," replied Armstrong to
his shipmate. "Remember? I told
you this was going to be fucking
amazing."

Levels of Reproducibility

1. Unreproducible

2. Reproducible with Effort

3. Completely Reproducible

Levels of Reproducibility

1. Unreproducible

Gayle, Vernon. "The determinants of student loan take-up in the United Kingdom: another gaze." *Applied Economics Letters* 3.1 (1996): 25-27.

Levels of Reproducibility

2. Reproducible with Effort

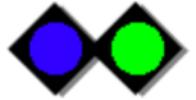
Connolly (2006) reproducible once effort is put in

Levels of Reproducibility

3. Completely Reproducible

Gayle, V. (1996) 'Modelling Tabular Data with an Ordered Outcome', *Sociological Research Online*, vol. 1, no. 3, <<http://www.socresonline.org.uk/1/3/4.html>>

© [Copyright Sociological Research Online, 1996](#)



Sociological Research Online

Modelling Tabular Data with an Ordered Outcome

by [Vernon Gayle](#)

Department of Applied Social Science, University of Stirling

Received: 19/01/96 Accepted: 23/9/96 Published: 2/10/96

Gayle, V. and Lambert, P.S., 2007. Using quasi-variance to communicate sociological results from statistical models. *Sociology*, 41(6), pp.1191-1208.

Full companion website <http://www.restore.ac.uk/Longitudinal/qv/>
can be used to reproduce the work in both Stata and SPSS

Ottoboni, K., Boring, A. and Stark, P., 2016. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.



<https://github.com/BuzzFeedNews/2016-01-tennis-betting-analysis>
<http://www.buzzfeed.com/heidiblake/the-tennis-racket>

1 commit 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

jtemplon Initial commit Latest commit 1f08da2 on Jan 17

data	Initial commit	7 months ago
notebooks	Initial commit	7 months ago
.gitignore	Initial commit	7 months ago
README.md	Initial commit	7 months ago
requirements.txt	Initial commit	7 months ago

README.md

Methodology and Code: Detecting Match-Fixing Patterns In Tennis

A closer look at the data analysis behind BuzzFeed News' investigation into corruption in tennis.

General Notes

In "The Tennis Racket," a yearlong investigation into match-fixing in professional tennis, BuzzFeed News published findings from an original data analysis we performed. That analysis revealed many examples of one particularly suspicious pattern: heavy betting against a player, followed by that player's loss.

Betting patterns alone aren't proof of fixing. Players can underperform for all sorts of reasons — injury, fatigue, bad luck — and sometimes that underperformance will just happen to coincide with heavy betting against them. But it's extremely unlikely for a player to underperform repeatedly in matches on which people just happen to be betting massive sums against him.

Edinburgh Research Explorer



Search the Research Explorer

- Media enquiries
▶
- Commercial enquiries
▶

- Research press coverage**
-
- ECGI to advise on carbon capture methods in Tees Valley**
 12/08/16
[Stuart Haszeldine](#)
- 300 volunteers to take part in whisky study lead by Dr Adam Moore**
 11/08/16
[Adam Moore](#)
- Cancer drug for mums-to-be may curb baby girls' future fertility, finds Prof Norah Spears**
 11/08/16
[Norah Spears](#)
- Dr Claudi Pagliari highlights hidden privacy breaches in OTC genetic testing kits**
 11/08/16
[Claudia Pagliari](#)
- Li-Fi, invented by Prof Harald Haas, could help connectivity in India**
 11/08/16
[Harald Haas](#)

Staff

Meet our international academics whose visions are shaping tomorrow's world.

Staff

Projects

Explore the pioneering research projects that make us a world-leading centre of excellence.

Projects

Activities

Search our research activities and events, and see the awards conferred on our researchers.

Activities

Research outputs

Discover our influential reputation through our research outputs and publications.

Outputs

STORRE: Stirling Online Research Repository



STORRE holds a growing collection of the research output of University of Stirling authors.

Learn More

STORRE

Welcome to STORRE: Stirling Online Research Repository

STORRE holds a continually growing collection of the full text of the research outputs of University of Stirling authors. It includes published journal articles, conference papers, book chapters, working papers, etc. As a result of the University [policy](#) encouraging open access, the repository will continue to develop as an important source of free full text access to Stirling's research.

STORRE also holds the full text of all University of Stirling research theses from [September 2006](#) onwards; covering PhDs, Masters of Philosophy and Masters by Research plus Professional Higher Degrees by Research. A small collection of our older theses is also included, and is continually added to, due to our involvement in the British Library's [ETHOS](#) service.

Recent Submissions REF ID REF ID RSS

Bracciali A, Cataldo E, Felicioi C, Marangoni R & Stano P From "cells as computation" to "cells as apps" (Forthcoming), Gadducci F, Tavocanis M (ed.) Volume of IFIP Advances in Information and Communication Technology, Third International Conference on the History and Philosophy of Computing (HaPoC 2015), Pisa, Italy, 8.10.2015 - 11.10.2015, Cham, Switzerland: Springer.

See

◀ We reflect on the computational aspects that are embedded in life at the molecular and cellular level, where life machinery can be understood as a massively distributed system whose macroscopic behaviour is an emerging property of the interaction of its components. Such a relatively new perspective, clearly pursued by systems biology, is contributing to the view that biology is, in several respects, a quantitative science. The recent developments in biotechnology and synthetic biology, notice... ▶

● ○ ○ ○ ○

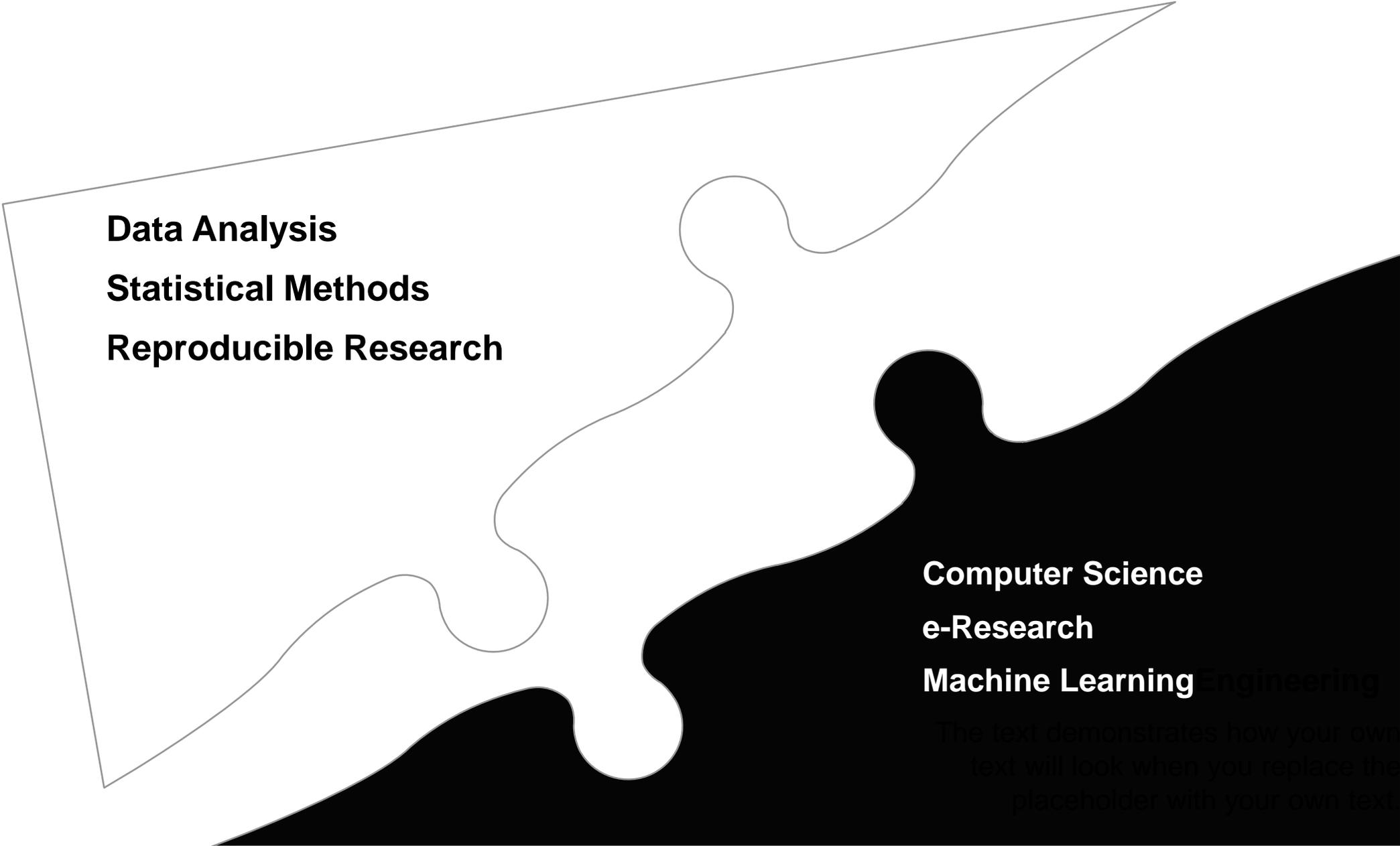
Deposit in STORRE

Depositing published research
 Stirling academic staff: [find out how to deposit](#).

eTheses Service

Is this all of Stirling's theses?
 No. This is only a very small subset. Browse a [list of all our theses titles](#).

All research theses must be deposited
 Stirling research postgraduates: [find out how to deposit](#).



Data Analysis

Statistical Methods

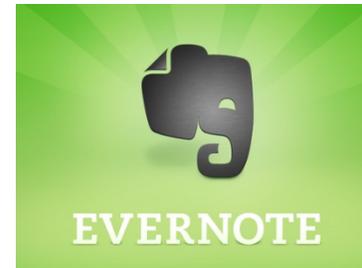
Reproducible Research

Computer Science

e-Research

Machine Learning Engineering

The text demonstrates how your own
text will look when you replace the
placeholder with your own text



GitLab



PAUL WELLER LIVE ST... vernongayle/SLS_Paper_2 x

GitHub, Inc. [US] https://github.com/vernongayle/SLS_Paper_2

This repository Search Pull requests Issues Gist

vernongayle / SLS_Paper_2 PRIVATE Unwatch 4 Star 0 Fork 0

Code Issues 0 Pull requests 0 Wiki Pulse Graphs Settings

Pattern of Subject Specific Performance in School Attainment in Scotland — Edit

28 commits 1 branch 0 releases 2 contributors

Branch: master New pull request New file Click for language details md file HTTPS https://github.com/vernongayle/SLS_Paper_2 Download ZIP

ChrisPlayford Added files via upload Latest commit 8127831 18 days ago

LLCS_abstract_20160303_text_v1.txt	Added files via upload	2 months ago
README.md	Create README.md	3 months ago
e-mail_to_Playford_20160307_vg_v1.docx	Added files via upload	2 months ago
e-mail_to_Playford_20160308_vg_v1.d...	Added files via upload	2 months ago
model1_model4_20160313_vg_v1.do	Exploring models 1-4	2 months ago
model1_qv_20160304_vg_v1.do	Added files via upload	2 months ago
playford_sls_additional_models_201602...	Added files via upload	3 months ago
playford_sls_additional_models_201602...	Added files via upload	3 months ago
playford_sls_additional_models_progre...	Added files via upload	2 months ago
playford_sls_par_occ_analysis_freqs.xlsx	Added NS-SEC 8 Class by Parental Education Freqs	2 months ago
playford_sls_par_occ_analysis_model0...	Playford SLS Parental Occupational Analysis	2 months ago
playford_sls_par_occ_analysis_model0...	Playford SLS Parental Occupational Analysis	2 months ago
playford_sls_par_occ_analysis_model1...	Playford SLS Parental Occupational Analysis	2 months ago
quasi_variance_20140810_vg_v3.do	Added files via upload	3 months ago
qv_sls_education_20160220_vg_v3.do	Update qv_sls_education_20160220_vg_v3.do	3 months ago
...

Windows taskbar: 17:26 13/05/2016



Things we should do immediately

- Deposit annotated scripts that a THIRD PARTY can use to completely duplicate ALL the results included in the published work
 - Clearly state the data source (and release)
 - Clearly state software used including versions, libraries and dependencies (even seeds)
 - Include all the script needed for data enabling
 - Deposit well annotated code books (detailing variables)
 - Check that a THIRD PARTY can genuinely duplicate the work

Things we could do

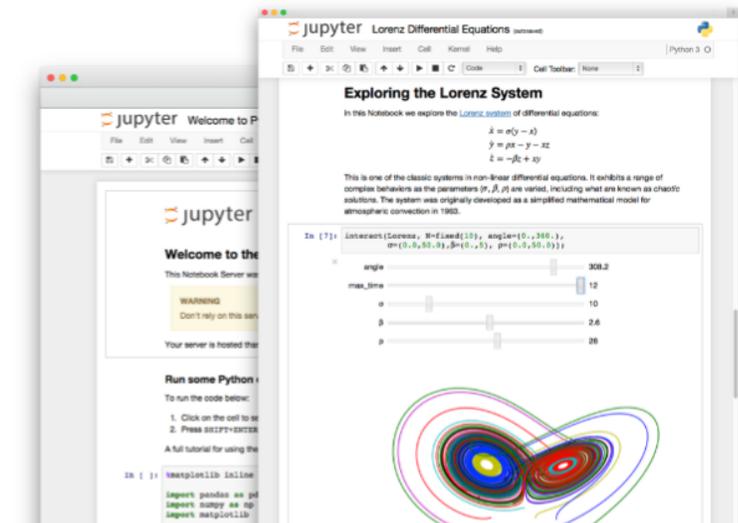
- Be clear about 'all' the work you did rather than just the work that has been reported
- Provide a justification of the micro-actions undertaken
 - minor recodes
 - variable choices (using wQFACHI not wQFEDHI)
- Peer programming (one player on the ball one player off the ball) ??
- More internal checking within research teams



Open source, interactive data science and scientific computing across over 40 programming languages.

Jupyter Notebook

The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.



<http://jupyter.org/>

<http://jupyter.org/>



THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SOCIAL SCIENCES AND HUMANITIES

- HOME
- ABOUT US
- CREATE & MANAGE DATA
- DEPOSIT DATA
- HOW WE CURATE DATA
- FIND DATA
- NEWS & EVENTS

CELEBRATING OUR HISTORY

As the University of Essex celebrates 50 years, take a look back at the history of the Archive

READ ON



- What's new
- Data lifecycle
- Find data
- Who are we?

DEPOSITING YOUR DATA

Depositing your data with the Archive ensures that they will be professionally curated and accessible

DEPOSIT DATA

FINDING DATA TO USE

We can help you find data for research and teaching with our catalogue of over 5,000 data collections

OUR CATALOGUE

HOW WE CURATE DATA

We follow best practices in preparing and curating our data to ensure usability

DATA CURATION

OUR SERVICES

UK Data Service data promoting evidence-based research

VIEW SITE

SEARCH OUR SITE

FIRST TIME HERE?

HELPFUL INFORMATION

A QUICK GUIDE TO THE ARCHIVE

3 of each of our data collections has a unique persistent identifier (DOI) that makes it easy to find and cite

WHO USES US DATA?

Find out what kind of data are available to you

LATEST NEWS & EVENTS

Futuretrack: a new longitudinal study of applicants to higher education

The UK Data Service is pleased to announce

Social Science Summer School The UK Data Service has been teaching students

Open the BLACK BOX

Routinely sharing research code

Edinburgh Research Explorer

University Homepage Research Explorer home

- Explorer home
- Staff
- Research projects
- Research outputs
- Research activities
- Colleges & Schools
- Research press coverage
- Datasets



Search the Research Explorer

Search

Media enquiries

Commercial enquiries

Staff

Meet our international academics whose visions are shaping tomorrow's world.

Staff

Projects

Explore the pioneering research projects that make us a world leading centre of excellence.

Projects

Activities

Search our research activities and events and see the awards conferred on our researchers.

Activities

Research outputs

Discover our influential reputation through our research outputs and publications.

Outputs

Research press coverage

ECGI to advise on carbon capture methods in Tees Valley
12/08/16
Stuart Haszeldine

300 volunteers to take part in whisky study lead by Dr Adam Moore
11/08/16
Adam Moore

Cancer drug for mums-to-be may curb baby girls' future fertility, finds Prof Norah Spears
11/08/16
Norah Spears

Dr Claudi Pagliari highlights hidden privacy breaches in OTC genetic testing kits
11/08/16
Claudia Pagliari

Li-Fi, invented by Prof Harald Haas, could help connectivity in India
11/08/16
Harald Haas

Research Talks

This page contains files related to research talks that I have delivered

Gayle, V. (2016) 'Why the Paper is Just a Palimpsest – An appeal for reproducible sociological research', *Q-Step Centre, Research Seminar, Edinburgh.*

Files supporting reproducibility

Stata syntax file (research code) [cambridge_20160901_vg_v4.do](#)

Stata 14 data file [fake_data_20160828_vg_v1.dta](#)

Stata 13 data file [fake_data_20160828_stata13_vg_v1.dta](#)

Excel versions of the data file [fake_data_20160828_vg_v1.xlsx](#)

Excel versions of the data file (csv) [fake_data_20160828_vg_v1.csv](#)

Vernon Gayle

Professor of Sociology & Social Statistics

University of Edinburgh

vernon.gayle@ed.ac.uk

@profbigvern

