



# Jupyter Notebooks

University of Luxembourg  
October 2018

Vernon Gayle  
University of Edinburgh

Roxanne Connelly  
University of York

**RUSSELL  
GROUP**





# Research Serendipity



Some ultramarathons I've run:

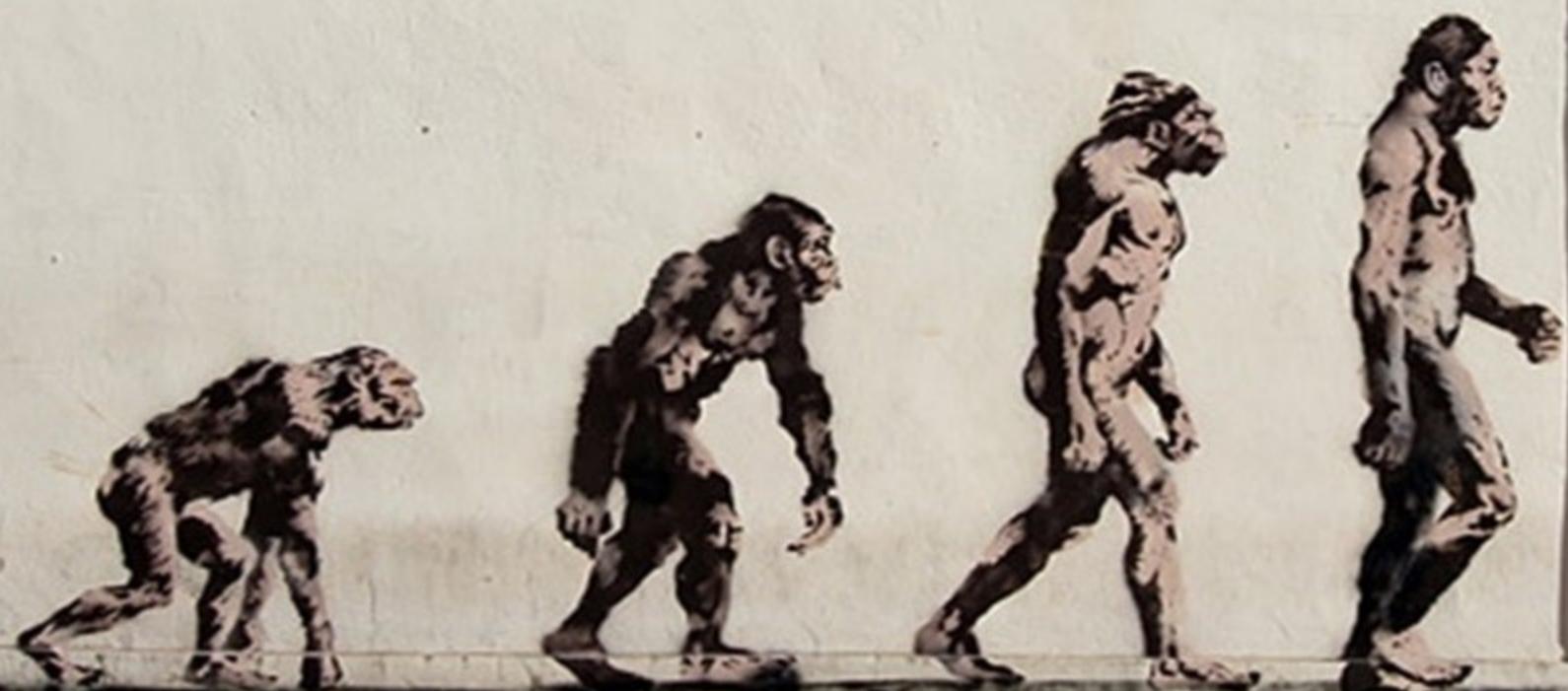
Event	Distance	Year	State	Climb
Angel Island 50k	50 km	2005	CA	4,200'
Backward Western States	100+ mi	2004	CA	21,970'
Backyard Hundred	100 mi	2006	CA	~20,000'
The Bear	100.33 mi	2004	ID	21,061'
		8	CA	~15,000'
		5	WY	18,308'
		5	CA	3,890'
		WA	20,470'	
		CA	28,102'	
		CA	7,800'	
		CA	~10,000'	
		CA	8,200'	
		CA	18,552'	
		CA	15,600'	
		Z	~6000'	
		X		
				3,710'
				4,750'
				11,300'
				11,300'



CARPENTERS PLACE

12000

200



Jupyter Notebooks

Playford, C.J., Gayle, V., Connelly, R. and Gray, A.J., 2016.  
Administrative social science data: The challenge of  
reproducible research. *Big Data & Society*, 3(2).

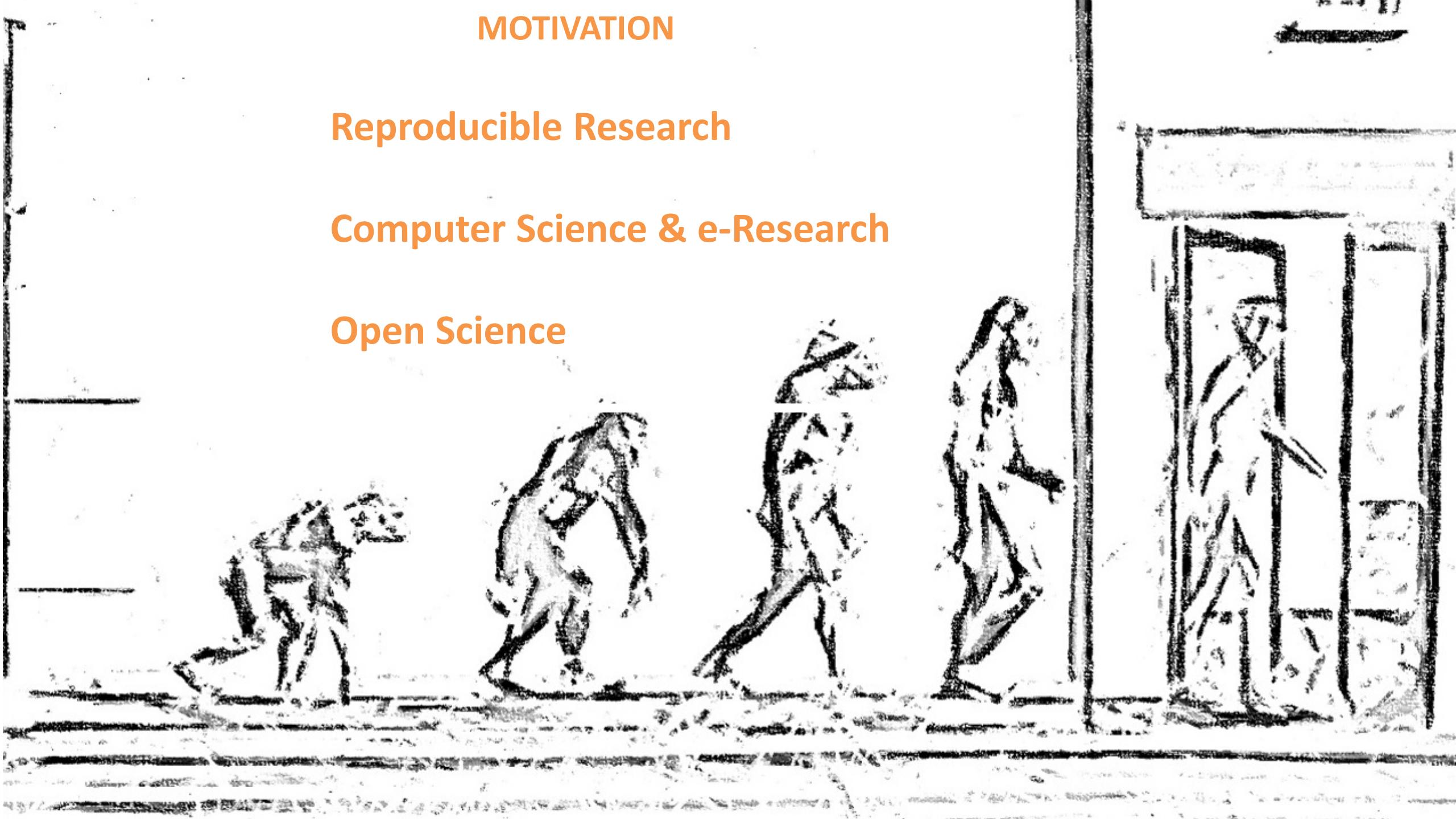


# MOTIVATION

Reproducible Research

Computer Science & e-Research

Open Science



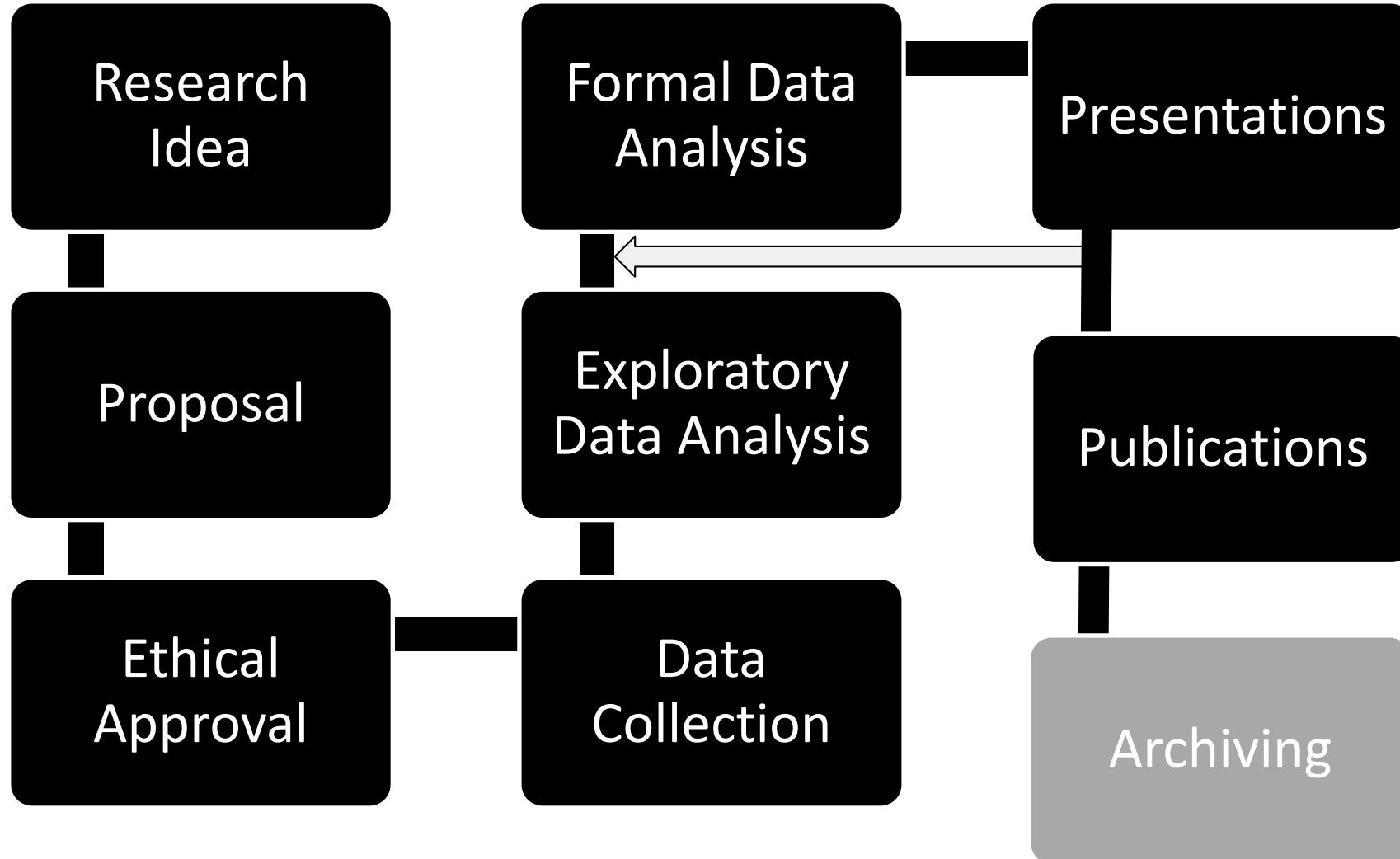
# Structure of this Session

- Workflow
- Literate computing
- Notebooks in research
- Jupyter in general
- Jupyter demonstration
- Concluding remarks



*A tigger warning – there will be live software demonstrated and things might get bouncy and could potentially go wrong!*

# The Workflow



Workflow home

What's new?

Additions by chapter

Downloading Stata files

My hardware & software

Reader's comments

Reader's stories

Quotes

Getting help

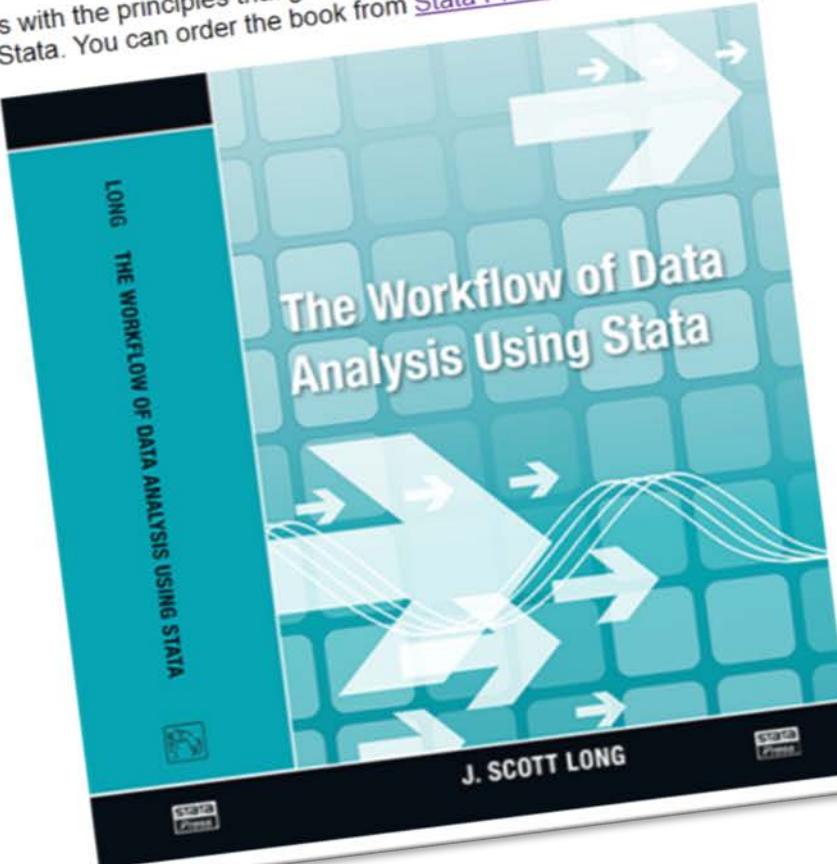
Disclaimer

Home

## The Workflow of Data Analysis Using Stata

Principles and practice for effective data management and analysis.

This project deals with the principles that guide data analysis and how to implement those principles using Stata. You can order the book from [Stata Press](#).



# The Workflow: A Practical Guide to Producing Accurate, Efficient, Transparent and Reproducible Social Survey Data Analysis

Vernon Gayle, Paul Lambert



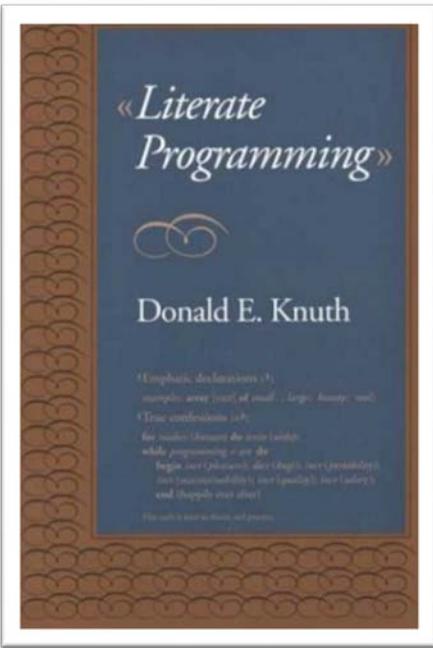
<http://eprints.ncrm.ac.uk/4000/>

# Literate Computing

Fernando Perez says

Literate Computing is the weaving of a narrative directly into a live computation, interleaving text with code and results to construct a complete piece that relies equally on the textual explanations and the computational components, for the goals of communicating results in scientific computing and data analysis.

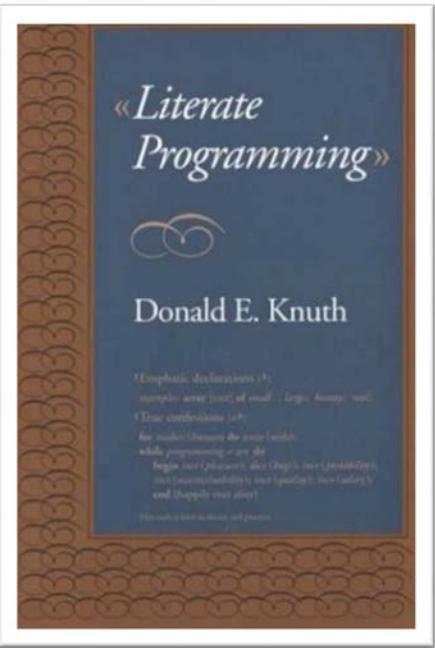
<http://blog.fperez.org/>



Knuth says

Treat your program as literature

People publish scores of symphonies they don't just listen to them



Knuth says

Treat your program as literature

People publish scores of symphonies they don't just listen to them

Both people and computers should be able to read your program

If others can read my program may I will understand my own program better



As with many scientists, Linus Pauling utilized bound notebooks to keep track of the details of his research as it unfolded. A testament to the remarkable length and diversity of Dr. Pauling's career, the Pauling Papers holdings include forty-six research notebooks spanning the years of 1922 to 1994 and covering any number of the scientific fields in which Dr. Pauling involved himself. In this regard, the notebooks contain many of Pauling's laboratory calculations and experimental data, as well as scientific conclusions, ideas for further research and numerous autobiographical musings.

[Research Notebook 01](#)

1922

[Research Notebook 02](#)

1922-1923, 1932, 1934, 1936, 1973,  
1985

[Research Notebook 03](#)

1923-1925

[Research Notebook 04](#)

1923-1924, 1928-1930

[Research Notebook 05](#)

[Research Notebook 13](#)

1935-1936, 1938-1939

[Research Notebook 14](#)

1936-1939, 1949, 1952

[Research Notebook 15](#)

1935, 1937, 1968

[Research Notebook 16](#)

1935-1956

[Research Notebook 17](#)

1939-1941, 1971, 1988

[Research Notebook 24](#)

1953, 1956, 1962, 1963, 1967, 1968,

[Research Notebook 25](#)

1969, 1970, 1973

[Research Notebook 26](#)

1958, 1964-1966

[Research Notebook 27](#)

1955, 1964-1969, 1974-1976, 1980-

[Research Notebook 28](#)

1982, 1987, 1990-1991

[Research Notebook 35b](#)

1938-1939, 1946, 1955, 1968, 1986-

[1988](#)

[Research Notebook 36](#)

1980-1981, 1986-1987

[Research Notebook 37](#)

1971, 1983

[Research Notebook 38](#)

1980-1981, 1983, 1985, 1989

[Research Notebook 39](#)

---

But on January 10th the stars appeared in the following position with regard to Jupiter; there were two only, and both on the east side

Ori.



Occ.

of Jupiter, the third, as I thought, being hidden by the planet.

---

# Jupiter



Type: planet  
Magnitude: -1.80 (extincted to: -0.80)  
Absolute Magnitude: 25.86  
RA/Dec (J2000.0): 15h19m45.40s/-17°34'18.7"  
RA/Dec (J2018.8): 15h20m49.15s/-17°38'19.4"  
Hour angle/DE: 3h40m34.18s/-17°32'09.1" (apparent)  
Az/Alt: +232°02'41.0"/+7°06'06.7" (apparent)  
Ecliptic longitude/latitude (J2000.0): +232°08'52.8"/+0°45'32.7"  
Ecliptic longitude/latitude (J2018.8): +232°24'35.7"/+0°45'25.8"  
Galactic Longitude/Latitude: -14°01'36.3"/+32°37'54.3"  
Obliquity (of date, for Earth): +23°26'12.7"  
Distance: 6.064AU (907.221 Mio km)  
Apparent diameter: +0°00'32.5"  
Sidereal period: 4331.87 days (11.860 a)  
Sidereal day: 9h55m29.7s  
Mean solar day: 9h55m33.1s  
Phase Angle: +7°17'08"  
Elongation: +42°55'32"  
Phase: 1.00  
Illuminated: 99.6%





# Juila, Python and R almost spell JuPyteR

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

<https://jupyter.org/>

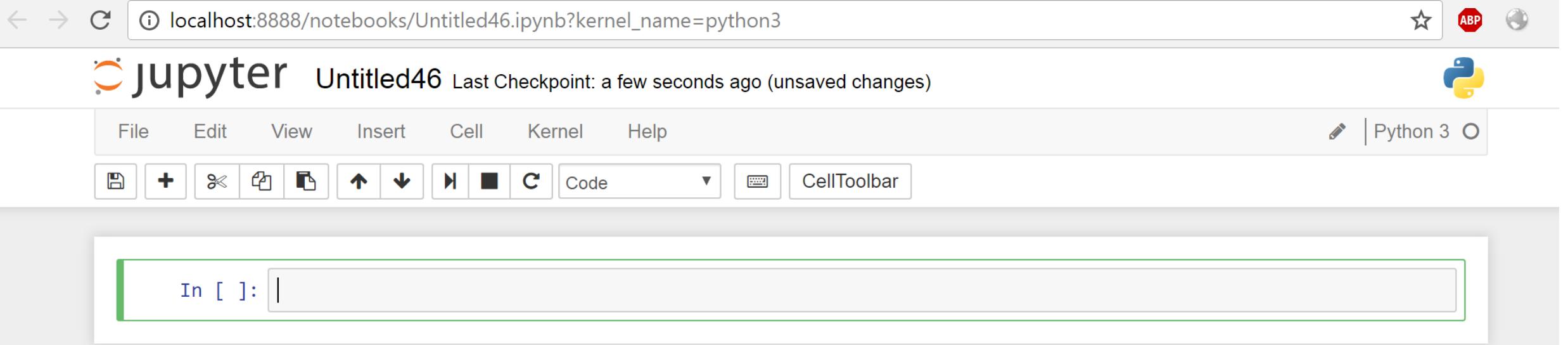




<https://www.youtube.com/watch?v=BmHPoBpZoJ4>



- Easy documentation alongside research code
- ‘Language agnostic’ 40+ languages
- Rich visual outputs
- Big data tools e.g. python
- Teaching and training
- Collaborative work
- Portability (publication) easy to share



Open source web application

Creates documents which include live code, output and explanatory text

Single platform for the complete workflow

# Code

```
In [4]: summarize
```

# Output

In [4]: summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
case	1,580	517.7411	284.8605	1	1003
femp	1,580	.6455696	.4784918	0	1
mune	1,580	.0740506	.2619362	0	1
time	1,580	7.2	3.981019	0	13
und1	1,580	.0746835	.2629633	0	1
-----+-----					
und5	1,580	.2974684	.4572891	0	1
age	1,580	36.01013	9.114841	18	60

# Text (Markdown)

In [4]: summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
case	1,580	517.7411	284.8605	1	1003
femp	1,580	.6455696	.4784918	0	1
mune	1,580	.0740506	.2619362	0	1
time	1,580	7.2	3.981019	0	13
und1	1,580	.0746835	.2629633	0	1
und5	1,580	.2974684	.4572891	0	1
age	1,580	36.01013	9.114841	18	60

The data mirror a real example of data analysed in Davies et al. (1992).

The dataset is a panel of 155 married women.

Davies, Richard B., Peter Elias, and Roger Penn. "The relationship between a husband's unemployment and his wife's participation in the labour force." *Oxford Bulletin of Economics and Statistics* 54.2 (1992): 145-171.

# Markdown

- *Markdown* is an easy way to write documents
- It is written in what computer geeks like to call 'plaintext'
- Plaintext is just the regular alphabet plus a few other familiar symbols (for example the asterisk \* )
- Unlike cumbersome word processing applications, text written in Markdown can be easily shared between computers

# Markdown

- It's quickly becoming the writing standard in some academic areas and in science
- Websites like GitHub and reddit use Markdown to style their comments
- Here is a summary of *Markdown* codes <https://en.wikipedia.org/wiki/Markdown#Example>
- If you have half an hour you can learn *Markdown* here <http://www.markdowntutorial.com/> (try a different browser)

# Images within the notebook cell...

localhost:8888/notebooks/adrcs\_20160922\_vg\_v8.ipynb

jupyter adrcs\_20160922\_vg\_v8

nunc 1 of 1

File Edit View Insert Cell Kernel Help Python 3

CellToolbar

A man and a bear...

A photograph of a man with a beard standing next to a large yellow teddy bear mascot. The bear is wearing a white shirt with red stripes and the word "PRUDENTIAL" printed on it. They are outdoors on a grassy field at dusk or night, with stadium lights visible in the background.

# LaTeX

«Lah-tech» rhymes with «Bertolt Brecht»

*to render cell contents as LaTeX*

In [8]:

```
%%latex
\begin{align}
a = \frac{1}{2} \quad & b = \frac{1}{2} \quad c = \frac{1}{4} \\
\end{align}
```

$$a = \frac{1}{2} \quad b = \frac{1}{2} \quad c = \frac{1}{4}$$

In [9]:

```
%%latex
$e^{i\pi} + 1 = 0
$
```

$$e^{i\pi} + 1 = 0$$

# The Swivel Chair – Language Agnostic Work



```
In [11]: logit femp mune und5
```

```
Iteration 0:  log likelihood = -1027.2309
Iteration 1:  log likelihood = -879.88806
Iteration 2:  log likelihood = -878.68101
Iteration 3:  log likelihood = -878.67998
Iteration 4:  log likelihood = -878.67998
```

```
Logistic regression                               Number of obs     =      1,580
                                                LR chi2(2)      =     297.10
                                                Prob > chi2    =     0.0000
Log likelihood = -878.67998                      Pseudo R2       =     0.1446
```

---

	femp	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
+	mune	-1.703308	.2358489	-7.22	0.000	-2.165563 -1.241053
	und5	-1.733521	.1221909	-14.19	0.000	-1.973011 -1.494031
	_cons	1.306829	.0744154	17.56	0.000	1.160978 1.452681

---

```
In [3]: mylogit <- glm(femp ~ mune + und5, data = mydata, family = "binomial")
summary(mylogit)
```

Call:  
glm(formula = empf ~ mune + und5, family = "binomial", data = mydata)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7586	-1.0024	0.6922	0.6922	2.1177

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.30683	0.07442	17.561	< 2e-16 ***
mune	-1.70331	0.23585	-7.222	5.12e-13 ***
und5	-1.73352	0.12219	-14.187	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2054.5 on 1579 degrees of freedom  
Residual deviance: 1757.4 on 1577 degrees of freedom  
AIC: 1763.4

```
In [6]: independentVar = ['mune', 'und5', 'Int']
logReg = sm.Logit(df['femp'] , df[independentVar])
answer = logReg.fit()
```

```
Optimization terminated successfully.
    Current function value: 0.556127
    Iterations 5
```

the results are in the object "answer"

```
In [9]: answer.summary()
```

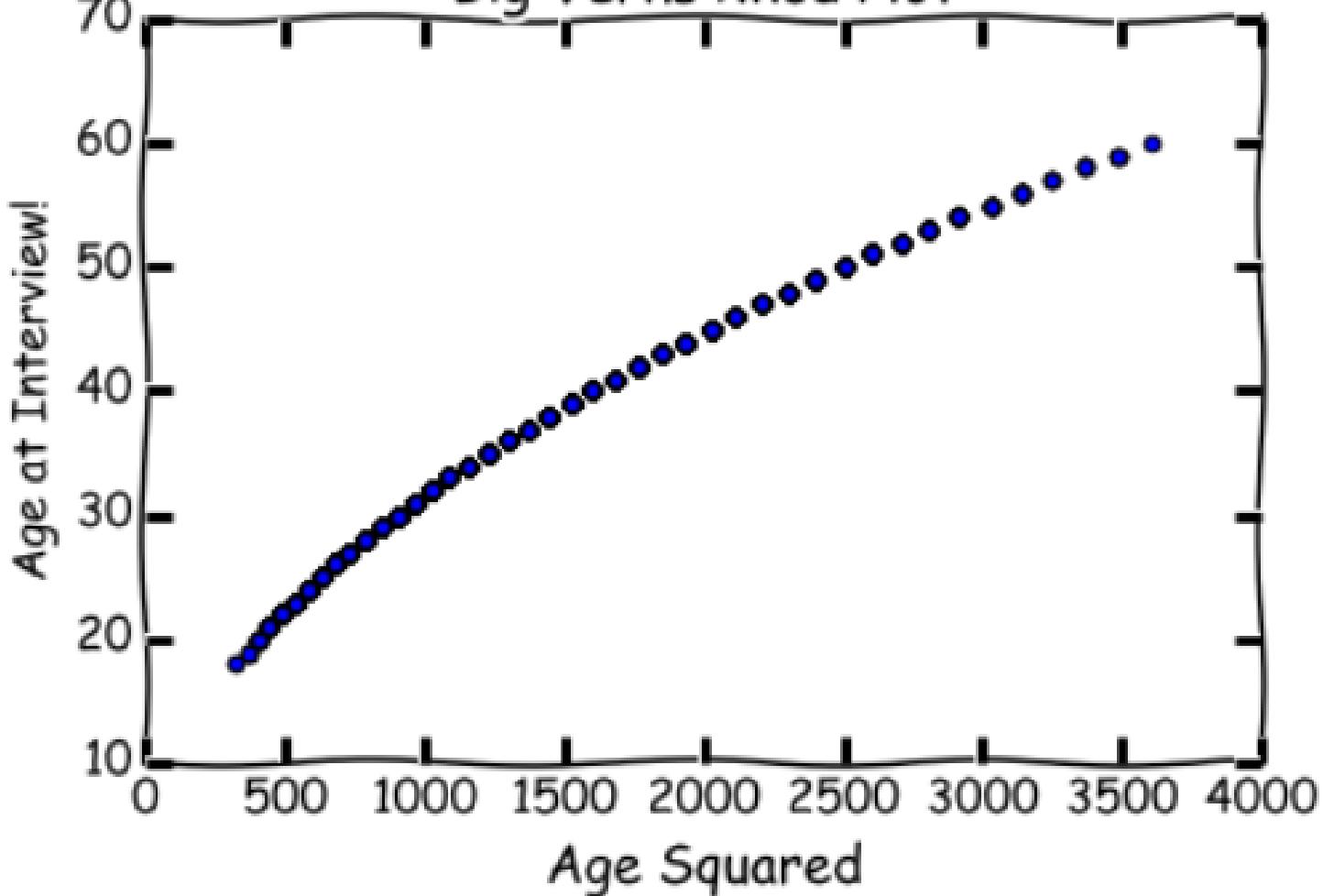
Out[9]: Logit Regression Results

Dep. Variable:	femp	No. Observations:	1580
Model:	Logit	Df Residuals:	1577
Method:	MLE	Df Model:	2
Date:	Fri, 14 Oct 2016	Pseudo R-squ.:	0.1446
Time:	10:13:23	Log-Likelihood:	-878.68
converged:	True	LL-Null:	-1027.2
		LLR p-value:	3.056e-65

	coef	std err	z	P> z	[95.0% Conf. Int.]
mune	-1.7033	0.236	-7.222	0.000	-2.166 -1.241
und5	-1.7335	0.122	-14.187	0.000	-1.973 -1.494
Int	1.3068	0.074	17.561	0.000	1.161 1.453

# Rich Visual Outputs

# Big Verns xkcd Plot

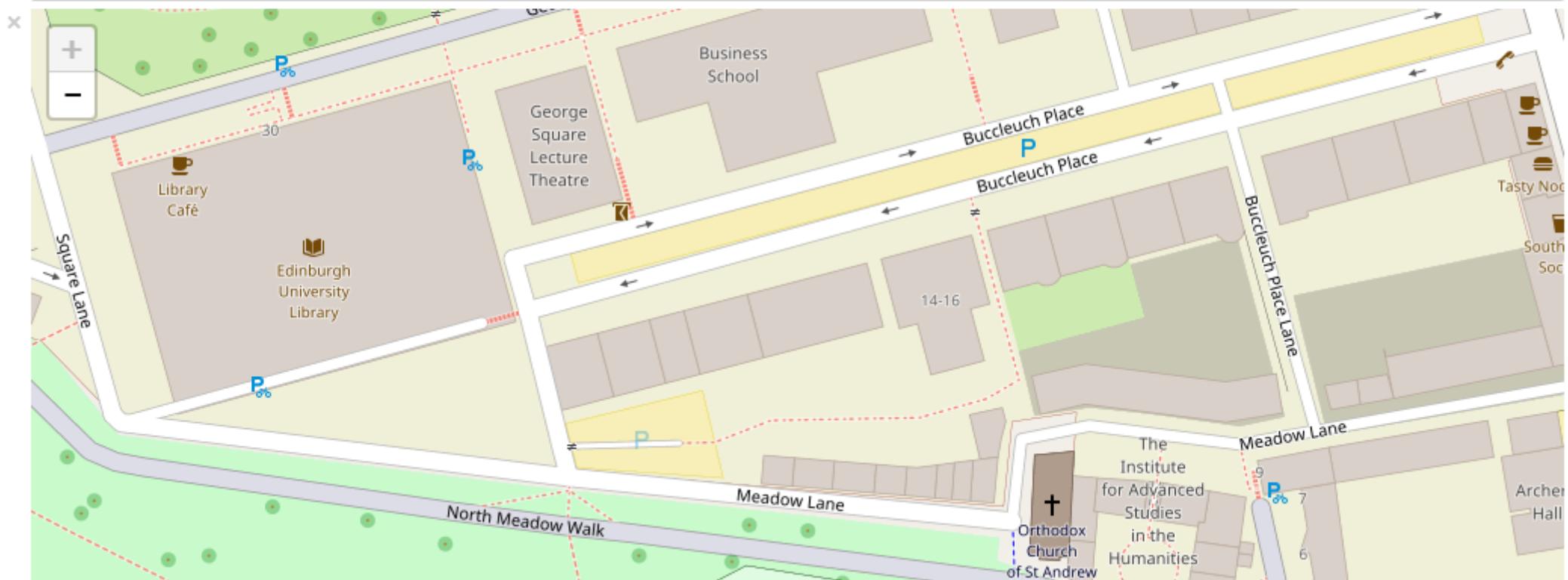


Another inventive use of the wemp dataset

## Using an open street map

I've recently moved to a more commodious office in Buccleuch Place. Here is an example of an open source map on my new hood.

```
In [5]: from ipyleaflet import Map  
Map(center=[55.942535, -3.187269], zoom=20)
```





# Teaching with Jupyter Notebooks





Computational Fluid Dynamics  
Algorithms **Fluid Mechanics**  
HIGH-PERFORMANCE COMPUTING  
**CFD** *Immersed Boundary Methods*  
**Biomolecular Physics**  
**GPU Computing**

## PUBLICATIONS



 RT @NumFOCUS: Tis the season of giving back! Support NumFOCUS & our projects by donating to our End-of-Year Fundraising Drive: <https://t.co...>  
[View](#) // [Reply](#) // [Retweet](#) // [Favorite](#)

 Donoho does not vouch for & will not cite the computational work of his own students who...refuse to work reproducibly <https://t.co/N0IQZ0hTKC>  
[View](#) // [Reply](#) // [Retweet](#) // [Favorite](#)

## CODE



Prof. Barba awarded a 2016 Leamer-Rosenthal Prize for Open Social Science



The 2016 Leamer-Rosenthal Prizes were announced on 15 December 2016, at the

# nbgrader

**nbgrader** is a tool that facilitates creating and grading assignments in the Jupyter notebook

It allows instructors to easily create notebook-based assignments that include both coding exercises and written free-responses

**nbgrader** then also provides a streamlined interface for quickly grading completed assignments

<https://nbgrader.readthedocs.io/en/stable/>

# nbconvert

**nbconvert** converts notebooks to familiar formats e.g. PDF HTML LaTeX

Presentation, publishing, sharing and collaboration

Original Article |  Free Access |

## An investigation of social class inequalities in general cognitive ability in two British birth cohorts\*

Roxanne Connelly , Vernon Gayle

First published: 19 December 2017 | <https://doi.org/10.1111/1468-4446.12343>

[Read the full text >](#)

 PDF  TOOLS  SHARE



### Early View

Online Version of Record before inclusion in an issue

 Figures  References  Related  Information

### Metrics

 score 39

Connelly, R. and Gayle, V., 2017. An investigation of social class inequalities in general cognitive ability in two British birth cohorts. *The British journal of sociology*.  
<https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-4446.12343>



Search or jump to...

Pull requests Issues Marketplace Explore

Watch 1 Star 1 Fork 2

## RoxanneConnelly / Social-Class-Inequalities-in-General-Cognitive-Ability-in-Two-British-Birth-Cohorts

Watch 1 Star 1 Fork 2

Code

Issues 0

Pull requests 0

Projects 0

Wiki

Insights

No description, website, or topics provided.

5 commits

1 branch

0 releases

0 contributors

Branch: master ▾

New pull request

Create new file

Upload files

Find file

Clone or download ▾

Roxanne Connelly Add files via upload

Latest commit 2615e6c on Nov 22, 2017

JupyterNotebook\_20171122.ipynb

Add files via upload

10 months ago

README.md

Update README.md

10 months ago

README.md

## An investigation of Social Class Inequalities in General Cognitive Ability in Two British Birth Cohorts

### British Journal of Sociology

Roxanne Connelly ([R.Connelly@warwick.ac.uk](mailto:R.Connelly@warwick.ac.uk))

Vernon Gayle ([vernon.gayle@ed.ac.uk](mailto:vernon.gayle@ed.ac.uk))

This repository hosts a Jupyter Notebook which accompanies the paper above. Details of how to access the required data are provided in the notebook.

# Observation of Gravitational Waves from a Binary Black Hole Merger

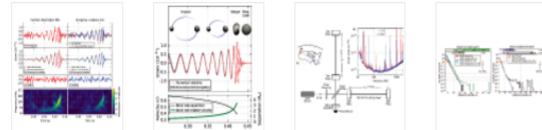
B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration)  
 Phys. Rev. Lett. **116**, 061102 – Published 11 February 2016

**Physics** See Viewpoint: [The First Sounds of Merging Black Holes](#)

Article References Citing Articles (185) PDF HTML Export Citation

## ABSTRACT

On September 14, 2015 at 09:50:45 UTC the two detectors of the Laser Interferometer Gravitational-Wave Observatory simultaneously observed a transient gravitational-wave signal. The signal sweeps upwards in frequency from 35 to 250 Hz with a peak gravitational-wave strain of  $1.0 \times 10^{-21}$ . It matches the waveform predicted by general relativity for the inspiral and merger of a pair of black holes and the ringdown of the resulting single black hole. The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203 000 years, equivalent to a significance greater than  $5.1\sigma$ . The source lies at a luminosity distance of  $410^{+160}_{-180}$  Mpc corresponding to a redshift  $z = 0.09^{+0.03}_{-0.04}$ . In the source frame, the initial black hole masses are  $36^{+5}_{-4} M_{\odot}$  and  $29^{+4}_{-4} M_{\odot}$ , and the final black hole mass is  $62^{+4}_{-4} M_{\odot}$ , with  $3.0^{+0.5}_{-0.5} M_{\odot}c^2$  radiated in gravitational waves. All uncertainties define 90% credible intervals. These observations demonstrate the existence of binary stellar-mass black hole systems. This is the first direct detection of gravitational waves and the first observation of a binary black hole merger.



Received 21 January 2016

# PREDICTING CORONAL MASS EJECTIONS USING MACHINE LEARNING METHODS

M. G. Bobra and S. Ilinidis

Published 2016 April 21 • © 2016. The American Astronomical Society. All rights reserved.  
*The Astrophysical Journal*, Volume 821, Number 2

GitHub, Inc. [US] | <https://github.com/RoxanneConnelly/Social-Class-Inequalities-in-General-Cognitive-Ability-in-Two-British-Birth-Cohorts>

RoxanneConnelly / Social-Class-Inequalities-in-General-Cognitive-Ability-in-Two-British-Birth-Cohorts

Code Issues Pull requests Marketplace Explore

No description, website, or topics provided.

5 commits 1 branch 0 releases 0 contributors

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

Roxanne Connelly Add files via upload Latest commit 2615e6c on Nov 22, 2017

JupyterNotebook 20171122.ipynb Add files via upload 10 months ago

README.md Update README.md 10 months ago

README.md

An investigation of Social Class Inequalities in General Cognitive Ability in Two British Birth Cohorts

British Journal of Sociology

Roxanne Connelly ([R.Connelly@warwick.ac.uk](mailto:R.Connelly@warwick.ac.uk))

Vernon Gayle ([vernon.gayle@ed.ac.uk](mailto:vernon.gayle@ed.ac.uk))

This repository hosts a Jupyter Notebook which accompanies the paper above. Details of how to access the required data are provided in the notebook.

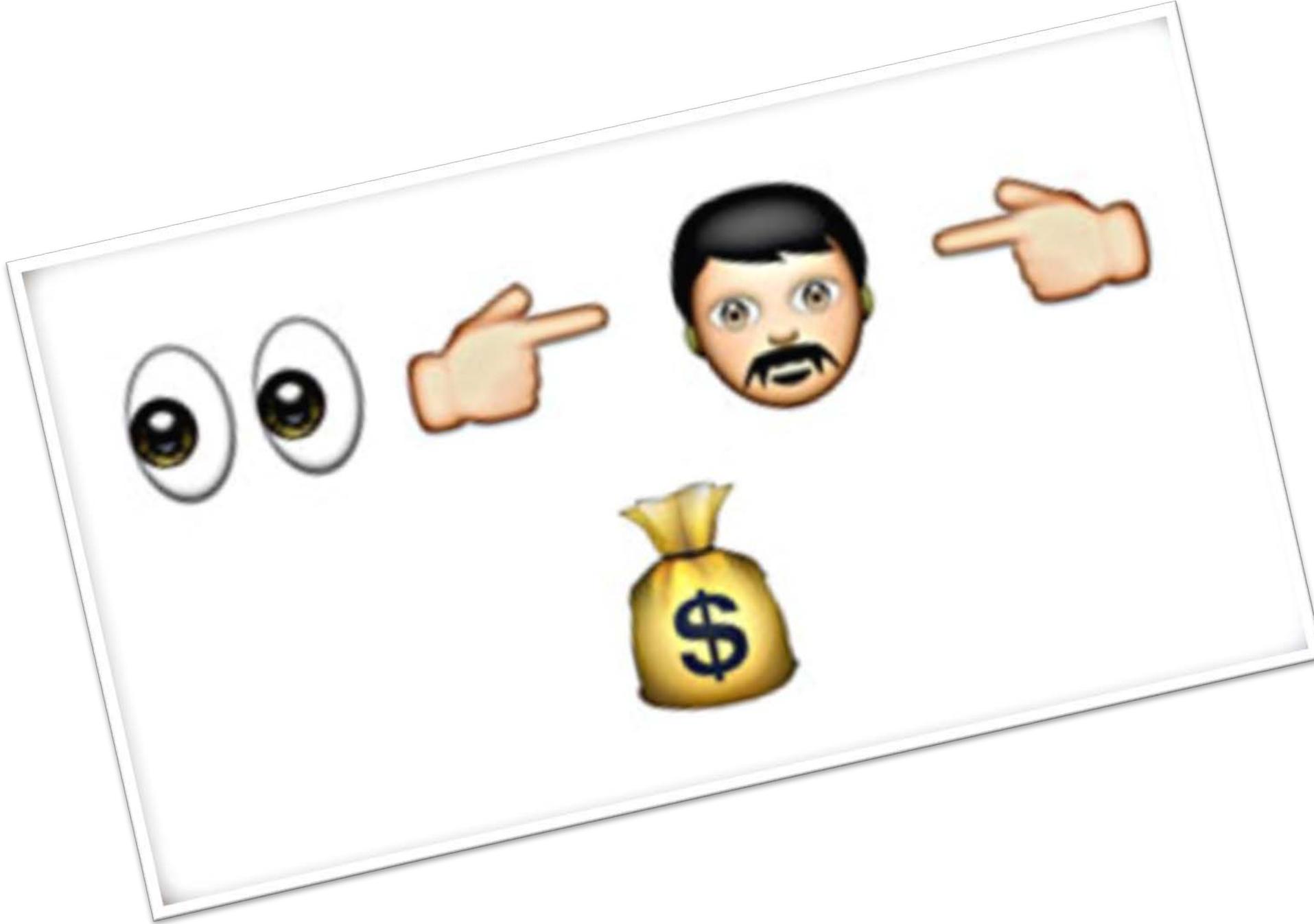


- Easy documentation alongside research code
- ‘Language agnostic’ 40+ languages
- Rich visual outputs
- Big data tools e.g. python
- Teaching and training
- Collaborative work
- Portability (publication) easy to share



# Some Points of Caution

- Easy to install but dependencies can be complex
- Windows 10, university systems etc. conspire against you
- Open source = less help
- Stack Overflow, blogs etc. assume low-level programming skills





# Jupyter Notebooks

University of Luxembourg  
October 2018

Vernon Gayle  
University of Edinburgh

Roxanne Connelly  
University of York

**RUSSELL  
GROUP**

