

Geographic Data Science - Lecture II

Modern Computational Environments

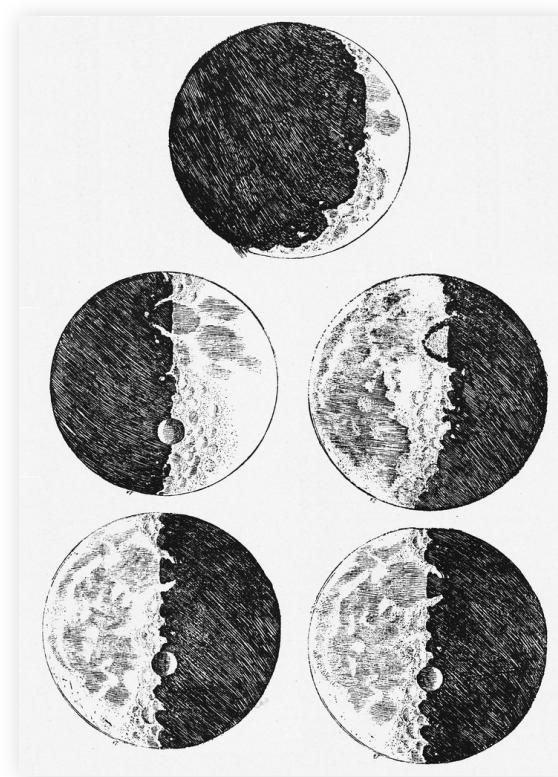
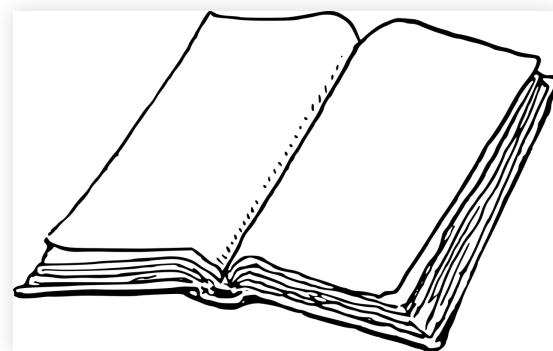
Dani Arribas-Bel

How does Science “get done”?

- Reproducibility in Science
- Modern scientific tools
- JupyterLab demo

Reproducibility in Science

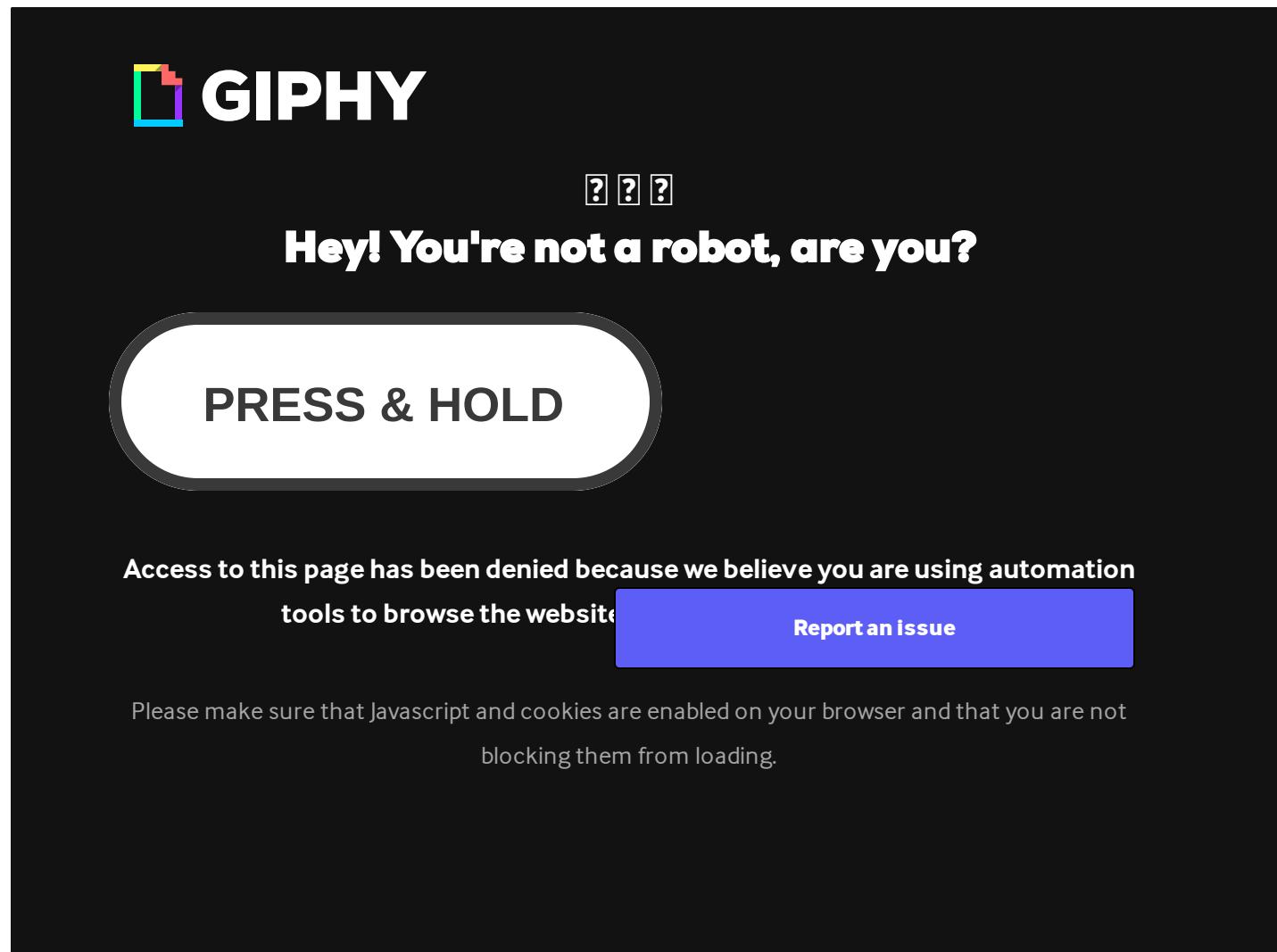
In the old days...



Reproducibility

- Ability to reproduce scientific procedures
(e.g. experiments, results)
- Key to the scientific endeavour
- Embedded in early work

But...



SHARE

Duke University is at the center of a whistleblower lawsuit concerning potential research misconduct. USCHOOLS
UNIVERSITY IMAGES/ISTOCKPHOTO

url

Whistleblower sues Duke, claims doctored data helped win \$200 million in grants

By Alison McCook, Retraction Watch | Sep. 1, 2016, 2:00 PM

On a Friday in March 2013, a researcher working in the lab of a prominent pulmonary scientist at Duke University in Durham, North Carolina, was arrested on charges of embezzlement. The researcher, biologist Erin Potts-Kant, later pled guilty to siphoning more than \$25,000 from the Duke University Health System, buying merchandise from Amazon, Walmart, and Target—even faking receipts to legitimize her purchases. A state judge ultimately levied a fine, and sentenced her to probation and community service.

Then Potts-Kant's troubles got worse. Duke officials took a closer look at her work and didn't like what they saw. Fifteen of her papers, mostly dealing with pulmonary biology, have now been retracted, with many notices citing "unreliable" data. Several others have been modified with either partial retractions, expressions of concern, or corrections. And last month, a U.S. district court unsealed a whistleblower lawsuit filed by a former colleague of Potts-Kant. It accuses the researcher, her former supervisor, and the university of including fraudulent data in applications and reports involving more than 60 grants worth some \$200 million. If successful, the suit—brought under the federal False Claims Act (FCA)—could force Duke to return to the government up to three times the amount of any ill-gotten funds, and produce a multimillion-dollar payout to the whistleblower.

The Duke case "should scare all [academic] institutions around the country," says attorney Joel

■ April 18, 2013, 11:31 AM GMT+1

FAQ: Reinhart, Rogoff, and the Excel Error That Changed History

By Peter Coy

url



PHOTOGRAPH BY GREGOR SCHUSTER

Harvard University economists Carmen Reinhart and Kenneth Rogoff have acknowledged making a spreadsheet calculation mistake in a 2010 research paper, “[Growth in a Time of Debt](#)” (PDF), which has been widely cited to justify budget-cutting. But the authors stand by their conclusion that higher government debt is associated with slower economic growth. Here’s what you need to know:

Over half of psychology studies fail reproducibility test

Largest replication study to date casts doubt on many published positive results.

Monya Baker

27 August 2015



Rights & Permissions

Don't trust everything you read in the psychology literature. In fact, two thirds of it should probably be distrusted.

In the biggest project of its kind, Brian Nosek, a social psychologist and head of the Center for Open Science in Charlottesville, Virginia, and 269 co-authors repeated work reported in 98 original papers from three psychology journals, to see if they independently came up with the same results.

The studies they took on ranged from whether expressing insecurities perpetuates them to differences in how children and adults respond to fear stimuli, to effective ways to teach arithmetic.



Brian Nosek's team set out to replicate scores of studies.

url

Modern Scientific Tools

Reproducible Science

- **Transparent** (computational) processes
- Enough **detail** to reproduce the entire analysis
- Efficient model of **reusability** (D.R.Y.)

Building Blocks

- Computational Notebooks
- Open-Source Packages
- Open Platforms

Computational Notebooks

**omic analysis of elongated skulls
extensive female-biased immigration
Early Medieval Bavaria**

Krishna R. Veeramah^a, Andreas Rott^{b,1}, Melanie Groß^{c,1}, Lucy van Dorp^d, Saioa López^e, Karola Kirsanow^c, Christian Sell^c, Jens Blöcher^f, Daniel Wegmann^{f,g}, Vivian Link^{f,g}, Zuzana Hofmanová^{f,g}, Joris Peters^{b,h}, Bernd Trautmann^b, Anja Gairhosⁱ, Jochen Haberstroh^j, Bernd Päffgen^k, Garrett Hellenthal^k, Brigitte Haas-Gebhard^k, Michaela Harbeck^{b,2,3}, and Joachim Burger^{c,2,3}

^aDepartment of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245; ^bState Collection for Anthropology and Palaeoanatomy, Bavarian Natural History Collections, 80333 Munich, Germany; ^cPalaeogenetics Group, Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz, 55099 Mainz, Germany; ^dUCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, WC1E 6BT London, United Kingdom; ^eCancer Institute, University College London, WC1E 6DD London, United Kingdom; ^fDepartment of Biology, University of Fribourg, 1700 Fribourg, Switzerland; ^gSwiss Institute of Bioinformatics, 1700 Fribourg, Switzerland; ^hArchaeoBioCenter and Institute for Palaeoanatomy, Domesticate Research and the History of Veterinary Medicine, Ludwig Maximilian University, 80539 Munich, Germany; ⁱBavarian State Archaeological Collection, 80538 Munich, Germany; ^jBavarian State Department of Monuments and Sites, 80539 Munich, Germany; and ^kInstitute of Prehistoric and Protohistoric Archaeology, Ludwig Maximilian University, 80799 Munich, Germany

Edited by Eske Willerslev, University of Copenhagen, Copenhagen, Denmark, and approved January 30, 2018 (received for review November 21, 2017)

Modern European genetic structure demonstrates strong correlations with geography, while genetic analysis of prehistoric humans has indicated at least two major waves of migration into Europe. During the period of cultural change, little is known on the demographic processes occurring during the intervening periods have been absent. Therefore, we generated genomic data from 41 individuals dating mostly to the late 5th/early 6th century AD from present-day Bavaria in southern Germany, including 11 whole genomes (mean depth 5.56x). In addition we developed a capture array to sequence neutral regions spanning a total of 5 Mb and 486 functional polymorphic sites to high depth (mean 72x) in all individuals. Our data indicate that while men generally had to form in the 5th century AD, and that it emanated from a combination of the romanized local population of the border province of the former Roman Empire and immigrants from north of the Alps. The latter were less well known than some other groups, such as the Franks. An interesting archaeological feature in Bavaria from this period is the presence of skeletons with artificially deformed or elongated skulls (Fig. 1A).

Artificial cranial deformation (ACD), which is only possible during early childhood, is a deliberate and permanent shaping of the head performed with great effort. In some societies reshaping the human skull has been seen as an ideal of beauty, while it

Significance

PNAS / Richard Goerg / Getty / The Atlantic

url

The Scientific Paper Is Obsolete

Here's what's next.

JAMES SOMERS | APR 5, 2018 | SCIENCE

Share

Tweet

...

TEXT SIZE

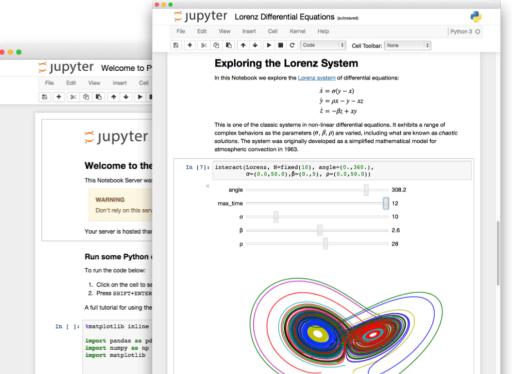
- +

Like *The Atlantic*? Subscribe to [The Atlantic Daily](#), our free weekday email newsletter.

Email

SIGN UP

THE SCIENTIFIC paper—the actual form of it—was one of the enabling inventions of modernity. Before it was developed in the 1600s, results were communicated privately in letters, ephemerally in lectures, or all at once in books. There was no public forum for *incremental* advances. By making room for reports of single experiments or minor technical advances, journals made the chaos of science accretive. Scientists from that point forward became like the social insects: They made their progress steadily, as a buzzing mass.



The Jupyter Notebook

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

[Try it in your browser](#) | [Install the Notebook](#)



Language of choice

Jupyter supports over 40 programming languages, including Python, R, Julia, and Scala.



Share notebooks

Notebooks can be shared with others using email, Dropbox, GitHub and the [Jupyter Notebook Viewer](#).



Interactive output

Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.



Big data integration

Leverage big data tools, such as Apache Spark, from Python, R and Scala. Explore that same data with pandas, scikit-learn, ggplot2, TensorFlow.

RStudio Blog

Home About Categories Tags Archives

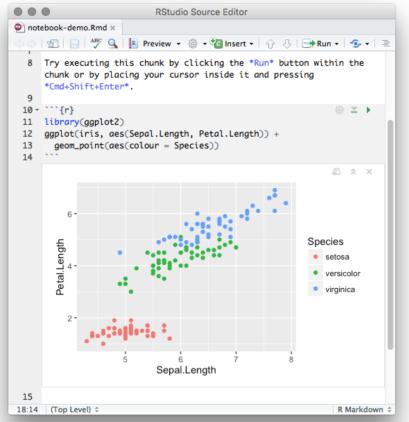
R Notebooks

Jonathan McPherson

2016-10-05

Categories: [Featured](#) [R Markdown](#) [RStudio IDE](#)

Today we're excited to announce [R Notebooks](#), which add a powerful notebook authoring engine to [R Markdown](#). Notebook interfaces for data analysis have compelling advantages including the close association of code and output and the ability to intersperse narrative with computation. Notebooks are also an excellent tool for teaching and a convenient way to share analyses.



Computational Notebooks

The screenshot shows a computational notebook interface with a title bar "lab_03". The main area contains a section titled "Loading up spatial data". It includes narrative text explaining how to load spatial data into a GeoDataFrame using the geopandas library. Below the text are three code cells:

- In [3]:**

```
lsoas_link = lcp_dir + 'shapefiles/Liverpool_lsoa11.shp'  
lsoas = gpd.read_file(lsoas_link)
```
- In [4]:**

```
lsoas.head()
```

Out[4]:

Lsoa11CD	geometry
0	POLYGON ((336103.358 389628.58, 336103.416 389...
1	POLYGON ((335173.781 389691.538, 335169.798 38...
2	POLYGON ((335495.676 389697.267, 335495.444 38...
3	POLYGON ((334953.001 389029, 334951 389035, 33...
4	POLYGON ((335354.015 388601.947, 335354 388602...
- In [5]:**

```
lsoas.plot()
```

Out[5]:

At the bottom, a note states: "This might not be the most aesthetically pleasant visual representation of the LSOAs geography, but it is hard to argue it is not quick to produce. We will work on styling and customizing spatial plots later on."

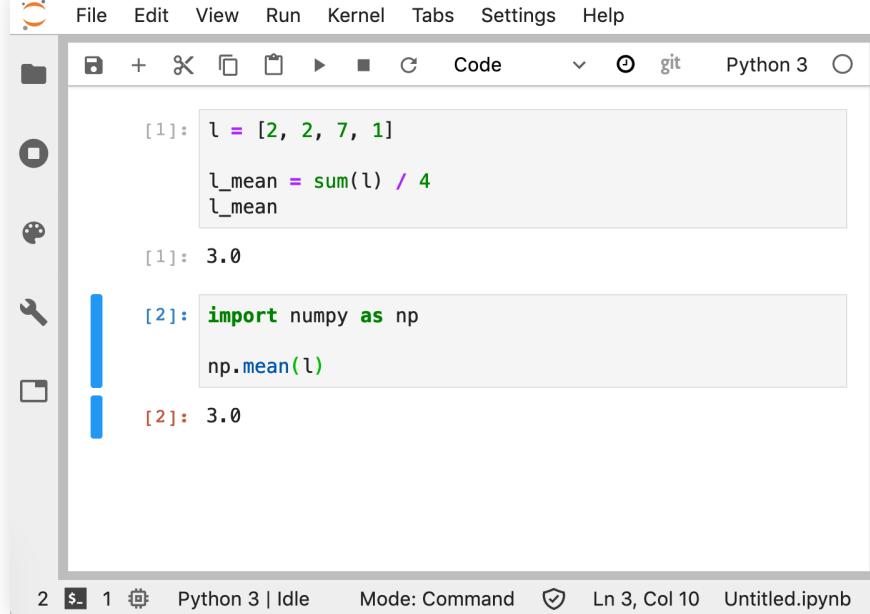
One-file documents:

- (Executable) code
- Output
- Narrative text

Open-Source Packages

D.R.Y. (Don't Repeat Yourself)

- Encapsulate **reusable** functionality
- Easy **access** + more **reliable** (if package is good!)
- Code available (**free** as in beer... but also as in speech!)



```
[1]: l = [2, 2, 7, 1]
l_mean = sum(l) / 4
l_mean
[1]: 3.0

[2]: import numpy as np
np.mean(l)
[2]: 3.0
```

Platforms

Hardware and low-level software (OS) that supports computations

Change of models:

- Desktop Vs cloud
- Integrated Vs distributed
- Native installation Vs virtualisation/containerisation

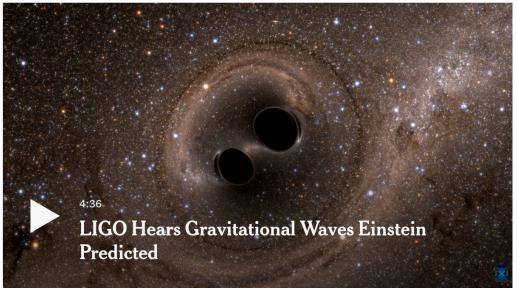
Examples

Ligo gravitational waves

The New York Times

SCIENCE

2017 Nobel Prize in Physics Awarded to LIGO Black Hole Researchers



4:36 LIGO Hears Gravitational Waves Einstein Predicted

About a hundred years ago, Einstein predicted the existence of gravitational waves, but until now, they were undetectable. Artist's rendering/Simulating eXtreme Spacetimes

By Dennis Overbye

Oct. 3, 2017

f t m r b

GW Open Science Center https://www.gw-openscience.org/tutorials/ 150% LOG IN

Gravitational Wave Open Science Center

Data Software Online Status About GWOSC

Tutorials

Each tutorial will lead you step-by-step through some common data analysis tasks. While GWOSC data can be analyzed using libraries in many software languages (C, C++, Matlab, etc.), most of these tutorials use Python. See also the [software page](#) for more examples.

See the [tutorial setup page](#) for help installing software to run these tutorials.

Tutorials shown here are not used to produce published results. For gravitational-wave software analysis packages that are used to produce LSC and Virgo Collaboration publications, see [software page](#).

Gravitational Wave Open Data Workshop Web Course (2019)



Lecture videos and tutorials from 2019 workshop

Course Material

2019



SPECTATOR INDEX **The Spectator Index** @spectatorindex

IMAGE: Reaction of Katie Bouman, who led the creation of an algorithm to produce first image of black hole.

A photograph of Katie Bouman sitting at a desk in an office. She is looking at a laptop screen displaying the first image of a black hole, which appears as a bright orange ring against a black background. She is smiling and has her hands near her face in a gesture of excitement or disbelief.

1:15 AM · Apr 11, 2019 · [TweetDeck](#)

2.5K Retweets 8.5K Likes

NATIONAL GEOGRAPHIC

A photograph of the first image of a black hole, which appears as a bright orange ring against a black background. This is the image that was captured by the Event Horizon Telescope.

The Event Horizon Telescope—a planet-scale array of ground-based radio telescopes—has obtained the first image of a supermassive black hole and its shadow. The image reveals the central black hole of Messier 87, a massive galaxy in the Virgo cluster.

PHOTOGRAPH BY EVENT HORIZON TELESCOPE COLLABORATION

SCIENCE & INNOVATION | STARSTRUCK

First-ever picture of a black hole unveiled

Using a telescope the size of the planet, astronomers have captured the first image of this space oddity. Here's why that matters.

BY NADIA DRAKE

9 MINUTE READ



PUBLISHED APRIL 10, 2019

JupyterLab (live) demo



Geographic Data Science'19 by Dani Arribas-Bel is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.