Reproducible Research at the Turing







- Research fellow at the Alan Turing Institute for Data Science
- Research associate in the Brain Mapping Unit, Department of Psychiatry, University of Cambridge
- 2016/17 Mozilla Fellow for Science



Reproducible vs Replicable





Same	Different



Different

S@kirstie_j

Same	Different
Reproducible	

Different

python



Same	Different
Reproducible	Replicable

Different

python (



Same	Different
Reproducible	Replicable
Robust	



Different

g@kirstie_j

Same Different Reproducible Replicable Generalisable Robust



Different

y @kirstie_j ucibleResearch

Barriers to reproducible research





Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings



Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings



Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings



Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings



Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings



Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings



Requires additional skills

Plead the 5th

Support additional users

Barriers to Takes time

reproducible research



Held to higher standards than others

Publication bias towards novel findings

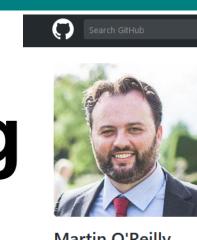


Reproducible research at the Turing





The Alan Turing Institute



Martin O'Reilly martintoreilly

Research Software Engineer at the UK's national data science institute.

Follow

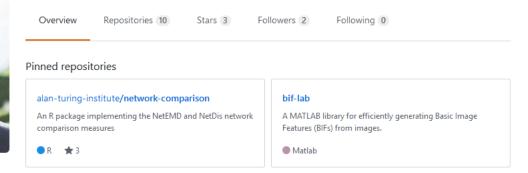
Block or report user

A The Alan Turing Institute

O London

Organizations

The Alan Turing



Pull requests Issues Marketplace Explore

815 contributions in the last year

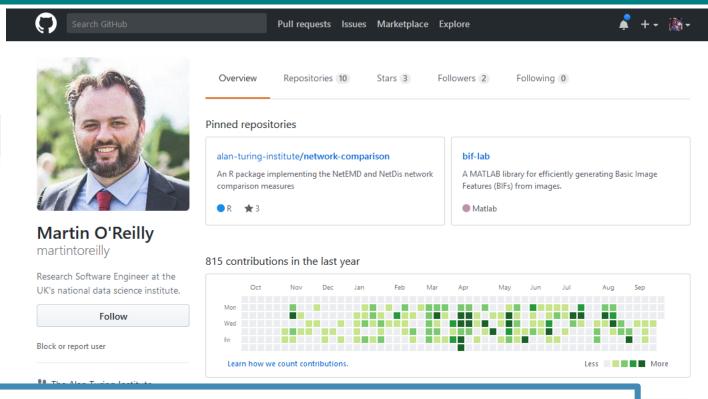








The Alan Turing Institute



My colleagues in the research software engineering team and I **seek to position The Alan Turing Institute as a world leader for reproducible research** by embedding a culture of
"reproducible by default" at the institute and providing training and tools to make reproducible
research "too easy not to do".





Reproducible research at the Turing

- Training & support for incoming students
- Software development support
 - From the beginning if possible
 - Posthoc for now
- Turing Reproducibility Champions
- Institutional buy in



Reproducible research at the Turing

- Training & support for incoming students
- Software development support
 - From the beginning if possible
 - Posthoc for now
- Turing Reproducibility Champions
- Institutional buy in
- "The Turing Way"





Attr: edhiggins, CC BY 3.0, via Wikimedia Commons

Start small

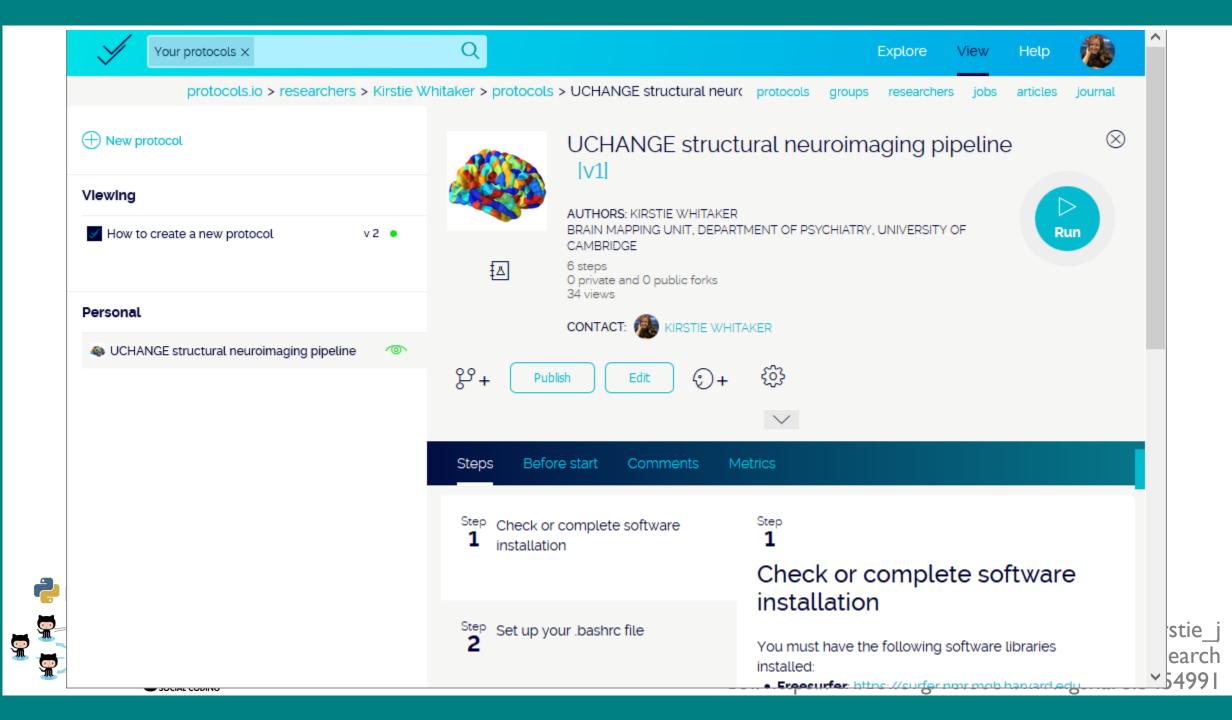




Protocols.io







My protocols







save in journal

the following text:

UCHANGE structural neuroimaging pipeline

13 steps

KIRSTIE WHITAKER, BRAIN MAPPING UNIT, DEPARTMENT OF PSYCHIATRY, UNIVERSITY OF CAMBRIDGE

CONTACT: KIRSTIE WHITAKER



Check or complete software installation You must have the following software libraries installed: Freesurfer: https://s... read more

Set up your bashrc file Open gedit and make sure your bashrc file contains

Set up directory structure You must have the appropriate input files in the

Put the fsaverageSubP folder in the SUB_DATA directory The fsaverageSubP

displayed directory structure: Inside SUB_... read more

directory contains the 308 parcellation and nee... read more

Check or complete software installation

You must have the following software libraries installed:

- Freesurfer: https://surfer.nmr.mgh.harvard.edu/fswiki/DownloadAndInstall
- FSL: http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation
- Anaconda: https://www.continuum.io/downloads

From within Anaconda install the following python packages following the command below:

- VTK
- Mayavi
- Nibabel
- pysurfer

COMMAND (LINUX)



conda install vtk

Coding your analyses



Comments are your friend!





```
1 #!/bin/bash
    # Created by Kirstie Whitaker on 13th April 2016
    # DESCRIPTION:
         This code conducts a brain and head extraction of the PDw image to which
           the quantitative multiparametric mapping (MPM) images have been aligned.
           It then uses the head mask to set all voxels outside of the head to
9 #
10 #
           zero for the quantitative MPM images and uses the brain mask to create
11 #
           brain extracted versions of the MPM images (where all voxels outside of
12
           the brain have been set to zero.
13 #
   # USAGE:
15
         NSPN mpm bet mask.sh <pdw file>
16
    # INPUTS:
17
         pdw file: Proton density weighted file to which the MPM
19
                      quantitative maps are aligned.
20
21
    # EXPECTS:
         The following files should be in the same directory as the
22
23
           input file:
24
                              MT.nii.gz
25
    #
             R1.nii.gz
26
             R1s.nii.gz
                              A.nii.gz
27 #
    # OUTPUTS:
28
         All output are in the same directory as the input file.
         A sub-directory called PDw bet is created and contains all the
31 #
           files created by FSL's bet command
32 #
33 #
            R1 head.nii.gz
                               R1 brain.nii.gz
```

```
# Erode the brain mask by 3mm
    if [[ ! -f ${mpm dir}/PDw brain ero3.nii.gz ]]; then
       fslmaths ${bet dir}/PDw brain.nii.gz -ero ${bet dir}/PDw brain ero3.nii.gz
94
    fi
95
96
    #-----
    # Now make the brain and head files for each of the
    # calculated MPM files
    #-----
    echo -n " Applying masks"
101
    for f name in PDw ${calc filename list[@]}; do
102
103
       # Don't run if it's already complete!
104
       if [[ ! -f ${mpm dir}/${f name} head.nii.gz ]]; then
105
          echo -n " - ${f name}"
106
107
          fslmaths ${bet dir}/PDw brain ero3.nii.gz \
108
                    -bin \
                    -mul ${mpm dir}/${f name}.nii.gz \
109
                    ${mpm dir}/${f name} brain.nii.gz
110
111
112
          fslmaths ${bet dir}/PDw brain outskin mask.nii.gz \
113
                    -bin \
114
                    -mul ${mpm dir}/${f name}.nii.gz \
115
                    ${mpm dir}/${f name} head.nii.gz
116
    done # Close the mpm calculated file loop
117
118
    echo ""
119
    #-----
    # All done!
    #-----
```



Aim for 40% comments in your code





Share your comments with the original author





(which is almost always going to be YOU!)







** It is embarrassing to me how often I forget not just details of experiments, but entire experiments. For example, for the manuscript I am working on now, I forgot that we had done an experiment to test for vertical transmission of the parasite. Fortunately, the undergrad who has been working on the project remembered and had it in his writeup!

https://dynamicecology.wordpress.com/2015/02/18/the-biggest-benefit-of-my-shift-to-r-reproducibility

python

https://twitter.com/kcranstn/status/370914072511791104



Asking for help: StackOverflow & MWEs

http://stackoverflow.com









https://mrcolley.com/2014/07/01/talk-to-the-duck-debugging-and-resilience



What if I can't share my code until l'm published?

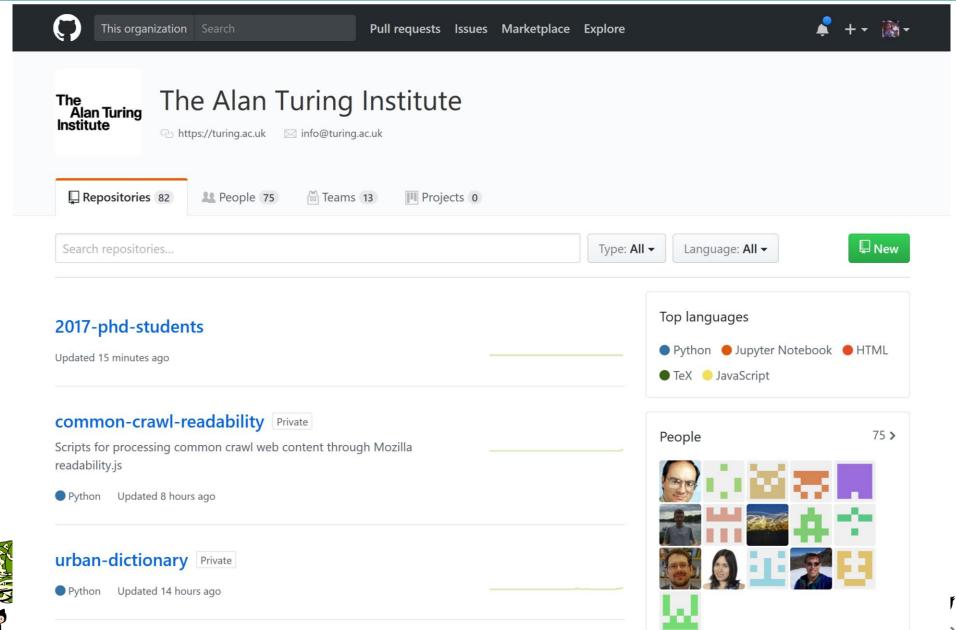








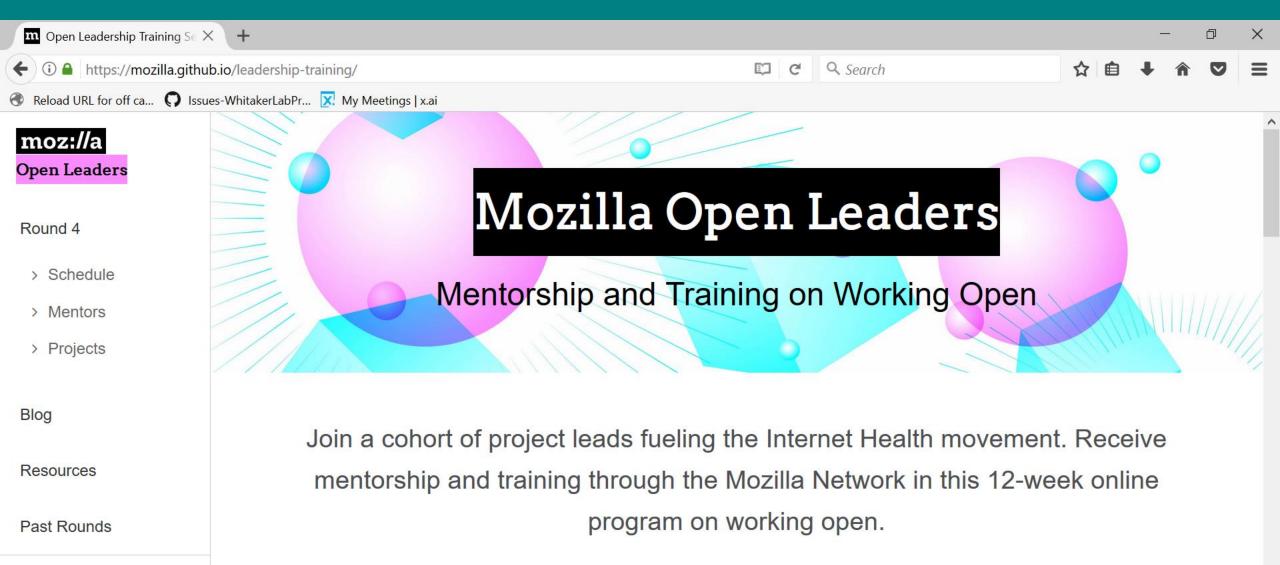




/@kirstie_j pleResearch

Reproducible <> Open (and that's fine)





About the Program

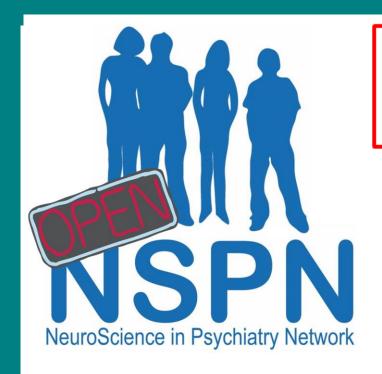
See the Code on

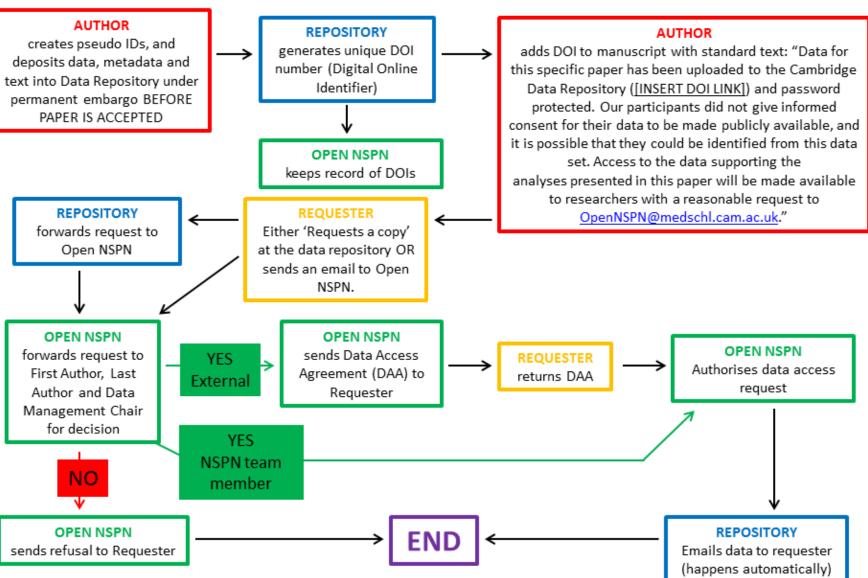
GitHub

Our fourth cohort of Mozilla Open Leaders starts in September 2017! Wa're looking for Open Project Leaders to fuel the Internet

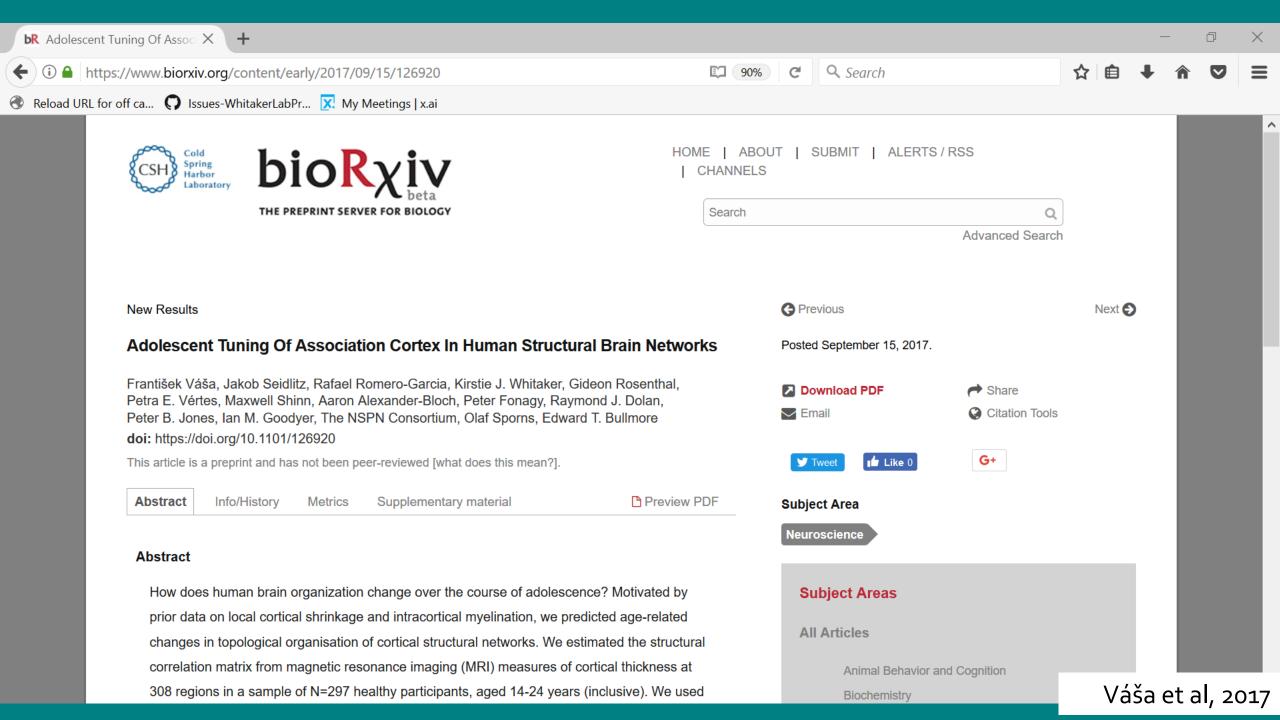
What if I can't (ever) share my data?

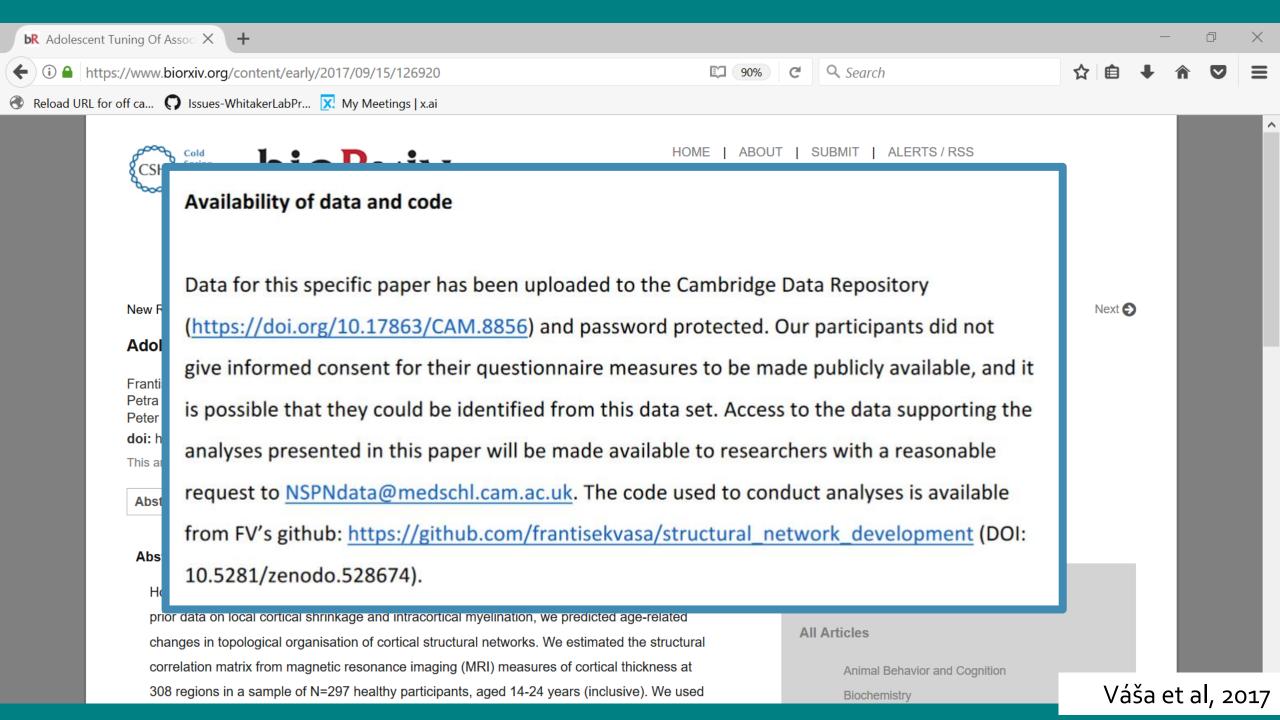


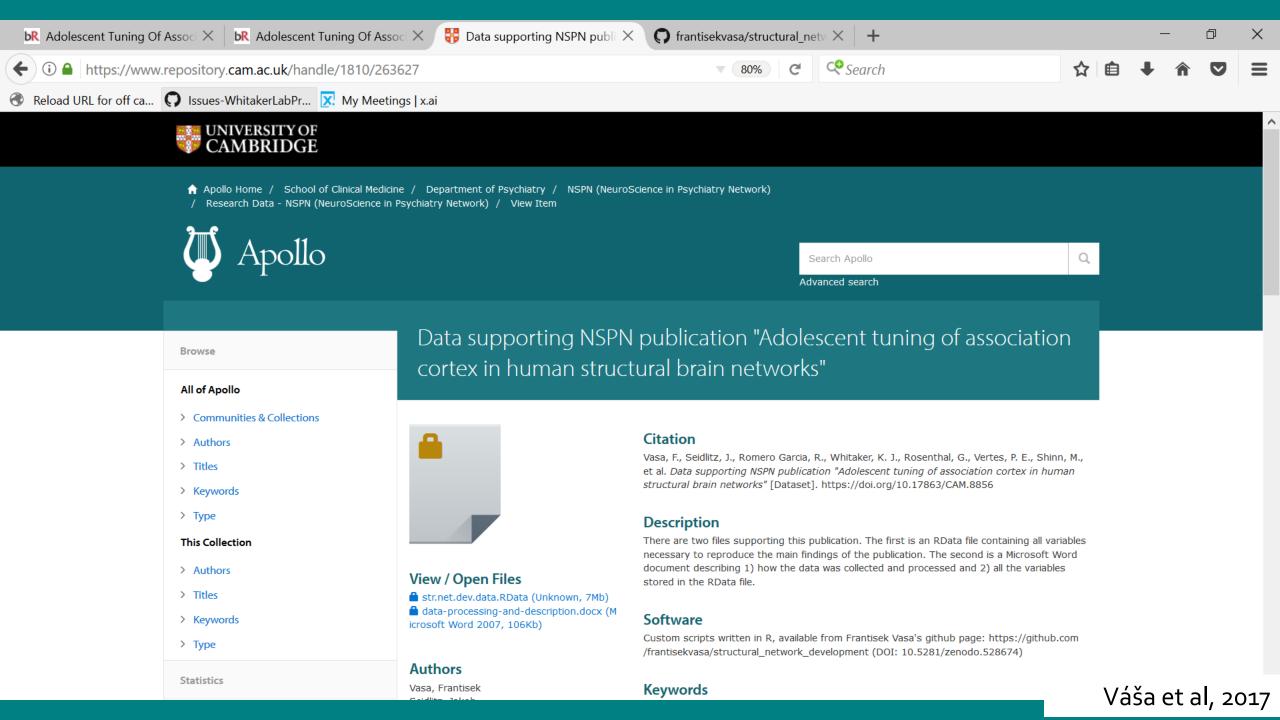


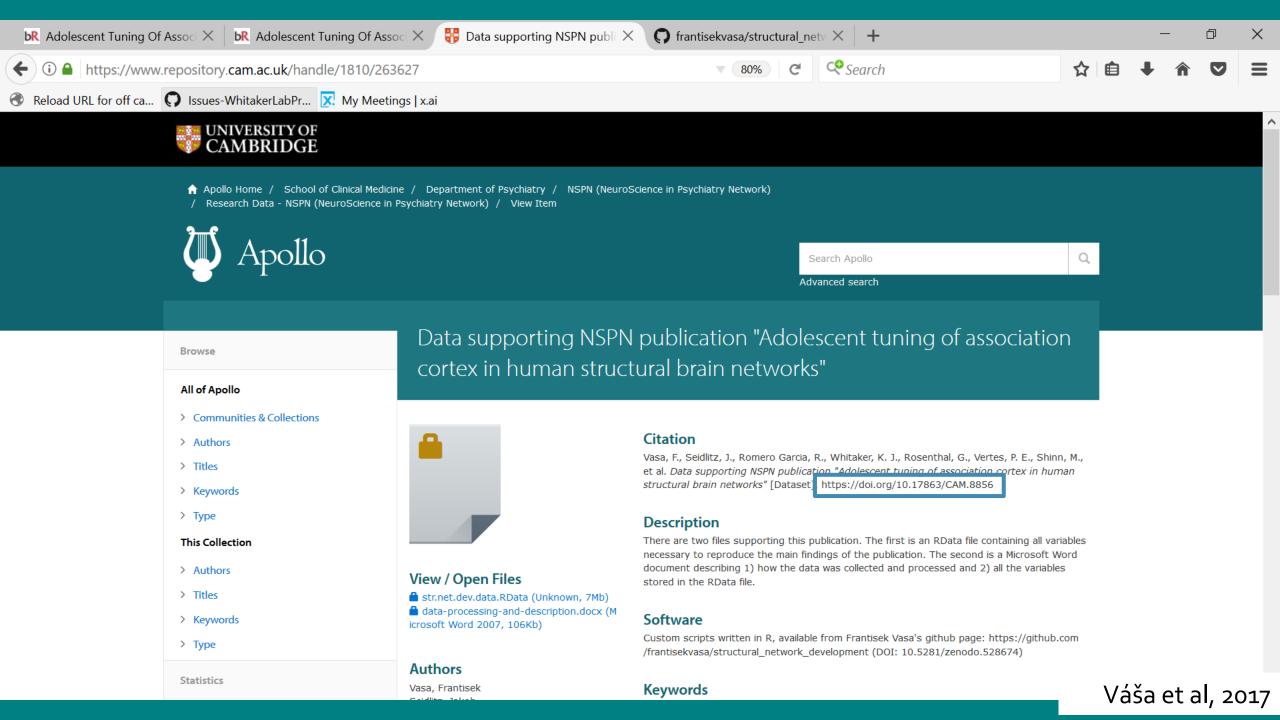


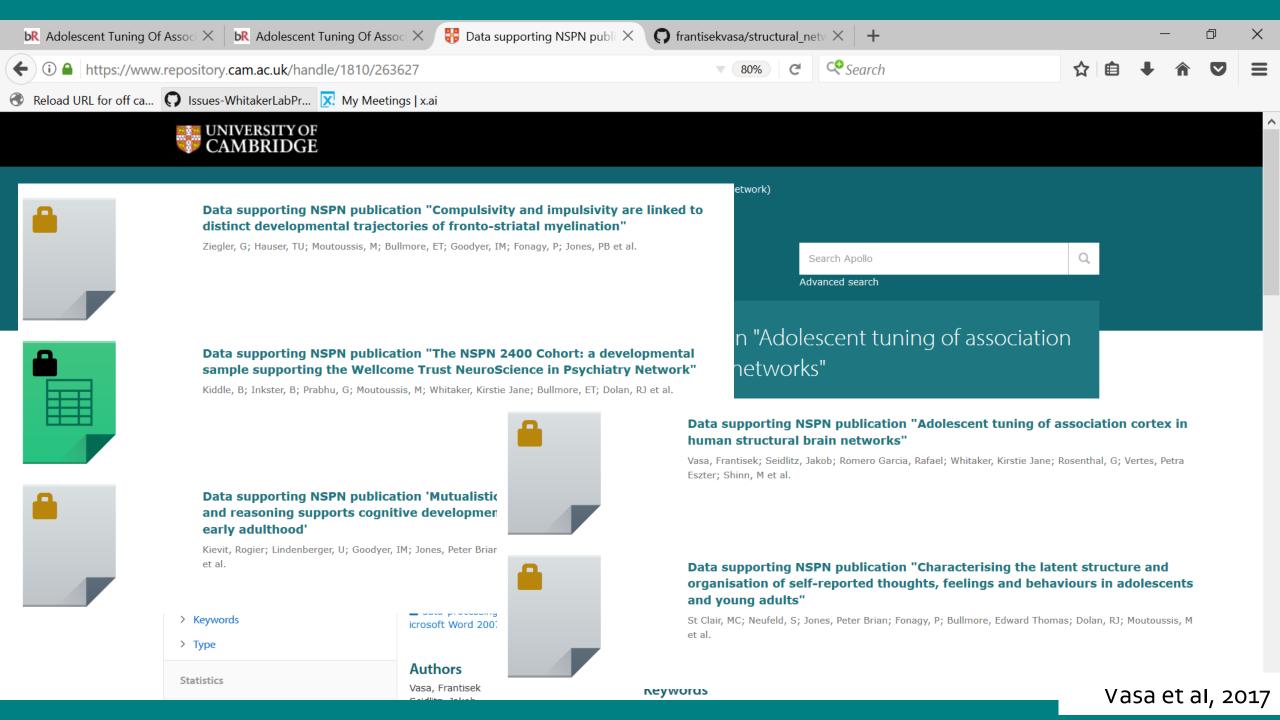


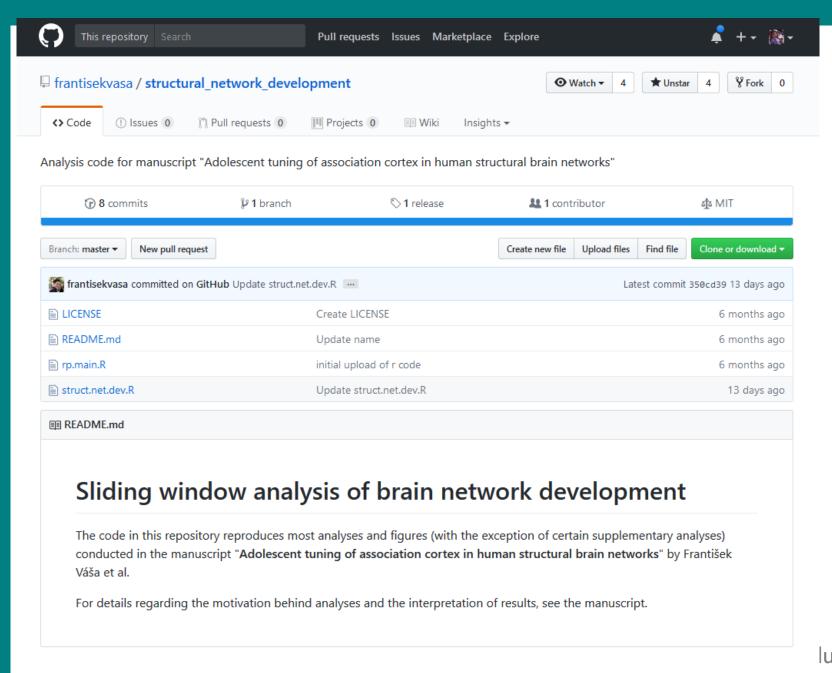




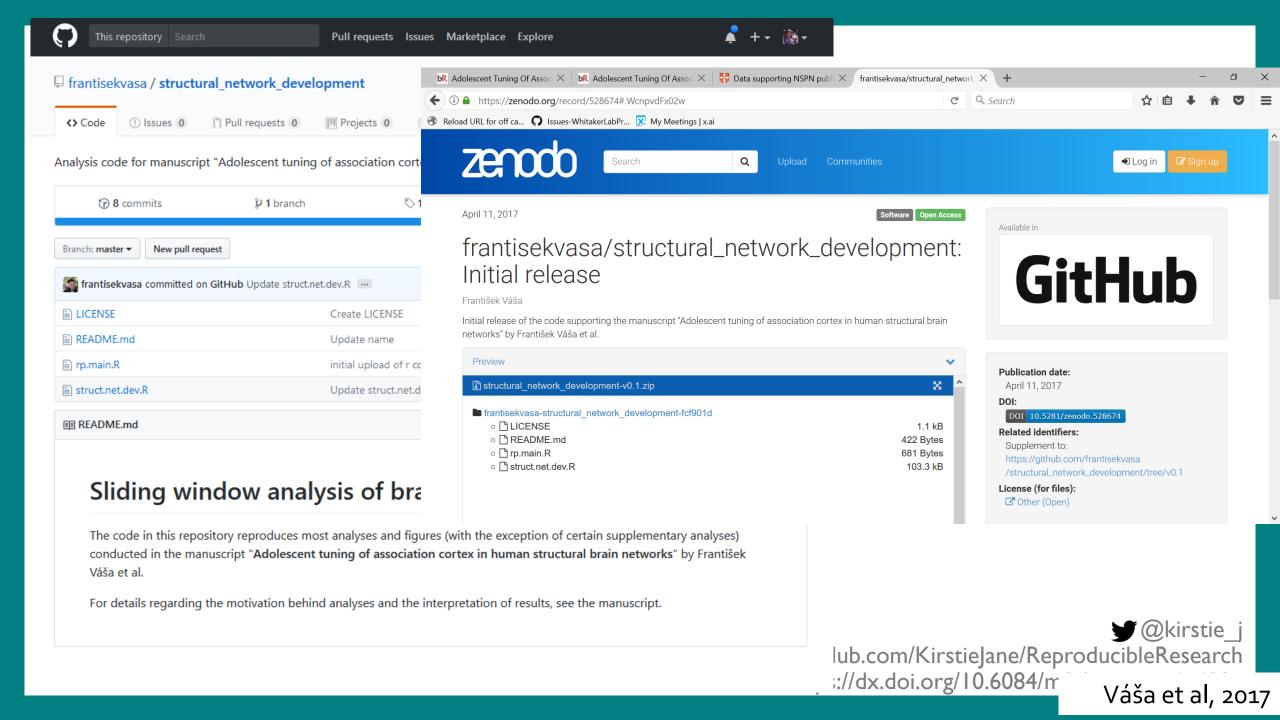






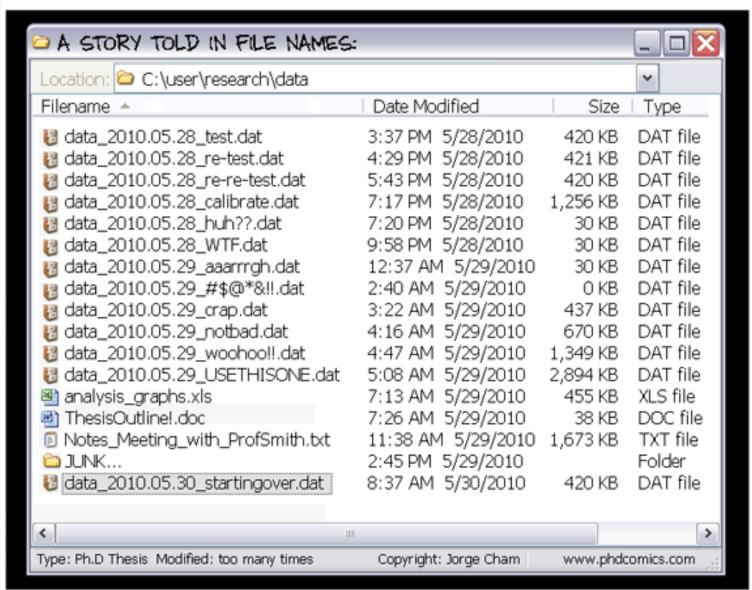


| which is a second of the contract of the con

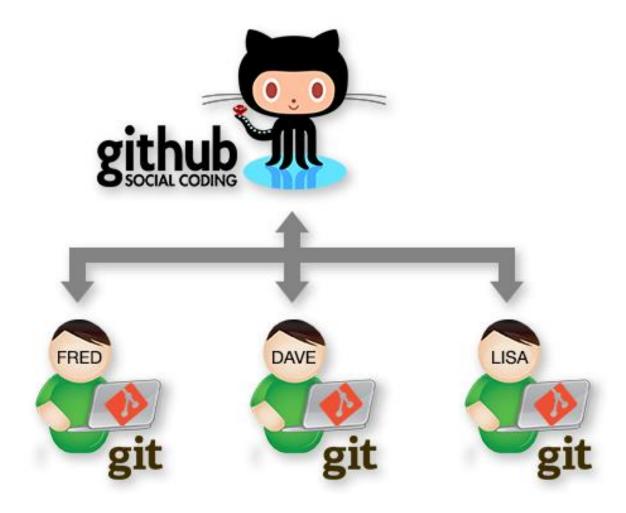


Version Control











Repository

Clone

Branch

Pull request

Version control

Some jargon busting

Merge

Issues

Release

Commit

Fork



Markdown



Submit your first pull request!

Inspired by: https://yourfirstpr.github.io







DATA CARPENTRY UIIII



Community

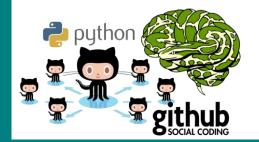
https://www.mozillascience.org

http://software-carpentry.org

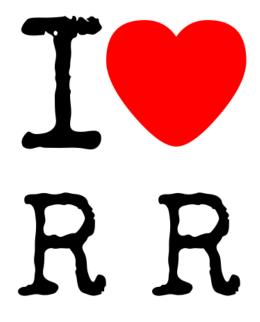
http://data-carpentry.org

https://www.coursera.org/specializations/jhu-data-science

https://www.coursera.org/learn/python/home/info







© 2016 iLoveHeartStudio.com

Thank you!



alan-turing-institute/2017-phd-students



@kirstie_j



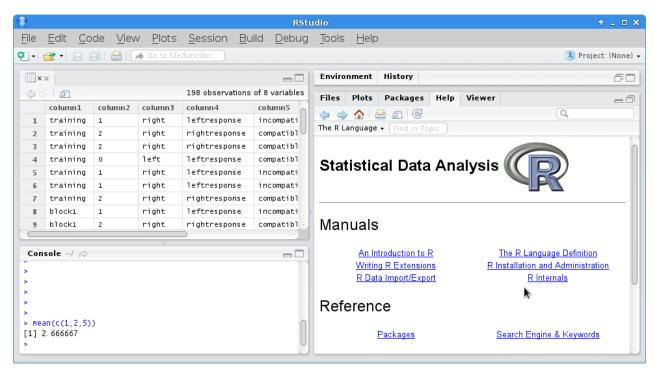
doi: 10.6084/m9.figshare.5454991



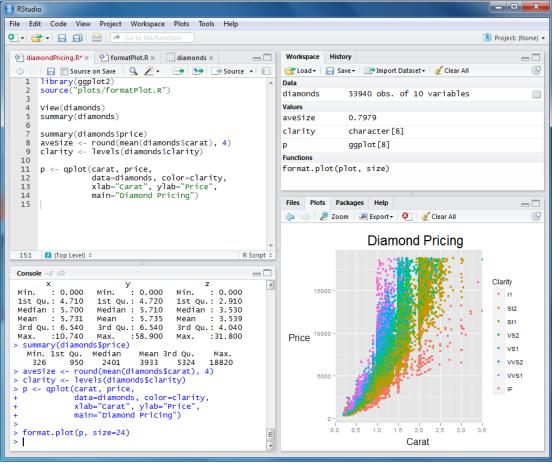
y @kirstie j

Python vs R (vs Matlab vs STATA etc...)









g @kirstie_j

OVERVIEW

TUTORIAL

ARTICLES

GALLERY

REFERENCE

DEPLOY

HELP

Gallery

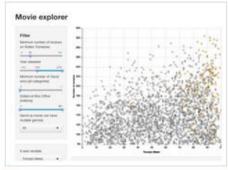
This gallery contains useful examples to learn from. Visit the Shiny User Showcase to see an inspiring set of sophisticated apps.

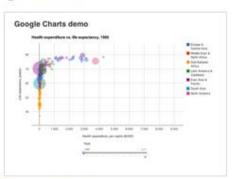
Interactive visualizations

Shiny is designed for fully interactive visualization, using JavaScript libraries like d3, Leaflet, and Google Charts.









SuperZip example

Bus dashboard

Movie explorer

Google Charts



Shiny gallery



PYTHON THE FASTEST GROWING OPEN DATA SCIENCE PLATFORM



Leading Open Data Science Platform
Powered by Python

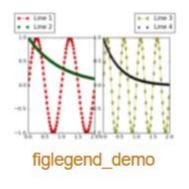
DOWNLOAD FOR FREE

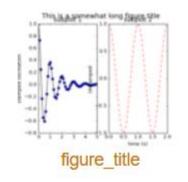


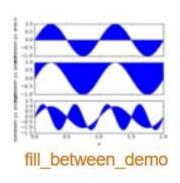


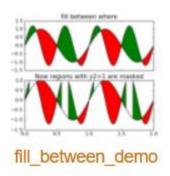
Matplotlib gallery

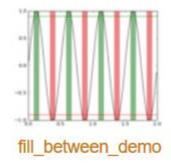
Click each example to see source code

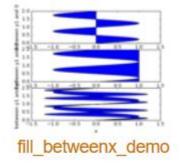


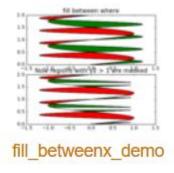


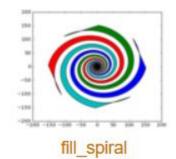






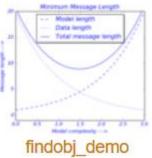


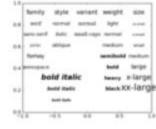












emo fonts_demo





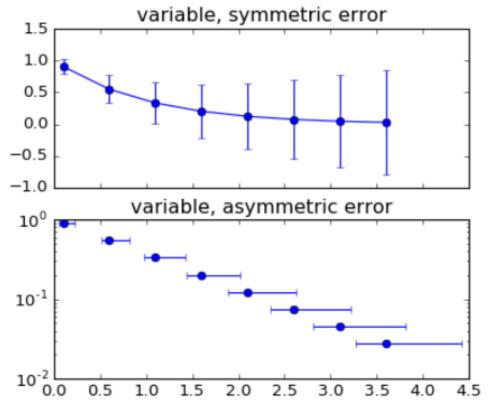






statistics example code: errorbar_demo_features.py

(Source code, png, hires.png, pdf)



```
Demo of errorbar function with different ways of specifying error bars.
Errors can be specified as a constant value (as shown in `errorbar demo.py`),
or as demonstrated in this example, they can be specified by an N x 1 or 2 x N,
where N is the number of data points.
N x 1:
    Error varies for each point, but the error values are symmetric (i.e. the
    Lower and upper values are equal).
2 x N:
    Error varies for each point, and the Lower and upper limits (in that order)
    are different (asymmetric case)
In addition, this example demonstrates how to use Log scale with errorbar.
import numpy as np
import matplotlib.pyplot as plt
# example data
x = np.arange(0.1, 4, 0.5)
# example error bar values that vary with x-position
error = 0.1 + 0.2 * x
# error bar values w/ different -/+ errors
lower_error = 0.4 * error
upper error = error
asymmetric error = [lower_error, upper_error]
fig, (ax0, ax1) = plt.subplots(nrows=2, sharex=True)
ax0.errorbar(x, y, yerr=error, fmt='-o')
ax0.set_title('variable, symmetric error')
ax1.errorbar(x, y, xerr=asymmetric_error, fmt='o')
ax1.set_title('variable, asymmetric error')
ax1.set yscale('log')
plt.show()
```



http://matplotlib.org/examples/statistics/errorbar_demo_features.html



Jupyter Notebook







SIGNAL PROCESSING WITH GW150914 OPEN DATA

Welcome! This ipython notebook (or associated python script GW150914_tutorial.py) will go through some typical signal processing tasks on strain time-series data associated with the LIGO GW150914 data release from the LIGO Open Science Center (LOSC):

- https://losc.ligo.org/events/GW150914/
- View the tutorial as a web page https://losc.ligo.org/s/events/GW150914/(In [6]:
- Download the tutorial as a python script https://losc.ligo.org/s/events/GW1
- Download the tutorial as iPython Notebook https://losc.ligo.org/s/events/G

To begin, download the ipython notebook, readligo.py, and the data files listed becan run the python script GW150914_tutorial.py. You will need the python package.

On Windows, or if you prefer, you can use a python development environment s /why-anaconda) or Enthought Canopy (https://www.enthought.com/products/cai

Questions, comments, suggestions, corrections, etc: email losc@ligo.org

v20160208b

https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.html



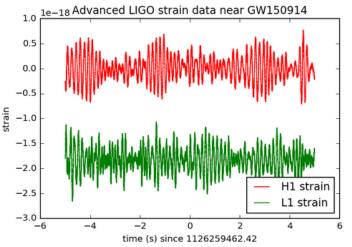
Observation of Gravitational Waves from a Binary Black Hole Merger

B. P. Abbott *et al.**
(LIGO Scientific Collaboration and Virgo Collaboration)
(Received 21 January 2016; published 11 February 2016)



```
# plot +- 5 seconds around the event:
tevent = 1126259462.422  # Mon Sep 14 09:50:45 GMT 2015
deltat = 5.  # seconds around the event
# index into the strain time series for this time interval:
indxt = np.where((time_H1 >= tevent-deltat) & (time_H1 < tevent+deltat))

plt.figure()
plt.plot(time_H1[indxt]-tevent, strain_H1[indxt], 'r', label='H1 strain')
plt.plot(time_L1[indxt]-tevent, strain_L1[indxt], 'g', label='L1 strain')
plt.xlabel('time (s) since '+str(tevent))
plt.ylabel('strain')
plt.legend(loc='lower right')
plt.title('Advanced LIGO strain data near GW150914')
plt.savefig('GW150914 strain.png')</pre>
```



01. 11cp3.// 47.401.01 6/ 10.000 1/11/1.116311a1 0.3 13 1//