Showing your working:
a how to guide to reproducible research



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201



- 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

mozilla Science Lab



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Reproducible vs Replicable



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

		Data	
		Same	Different
Code	Same		
	Different		



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

		Data	
		Same	Different
Code	Same	Reproducible	
	Different		



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different		



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Barriers to reproducible research



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Is not considered for promotion

Requires additional skills

Plead the 5th

Support additional users

Barriers to reproducible research

Takes time

Held to higher standards than others

Publication bias towards novel findings



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Start small



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Protocols.io



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Your protocols ×

Explore View Help

protocols.io > researchers > Kirstie Whitaker > protocols > UCHANGE structural neuro

protocols groups researchers jobs articles journal

New protocol

Viewing

How to create a new protocol v 2

Personal

UCHANGE structural neuroimaging pipeline



UCHANGE structural neuroimaging pipeline [v1]

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

6 steps
0 private and 0 public forks
34 views

CONTACT: KIRSTIE WHITAKER

Run

Publish Edit

Steps Before start Comments Metrics

Step 1 Check or complete software installation

Step 2 Set up your .bashrc file

Step 1

Check or complete software installation

You must have the following software libraries installed:

- **Freesurfer** https://surfer.nmr.mgh.harvard.edu



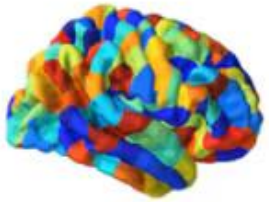


stie_j
earch
43201

search protocols, groups and more

My protocols Journal

save in journal



UCHANGE structural neuroimaging pipeline

13 steps

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

CONTACT: KIRSTIE WHITAKER

0/13 STEPS COMPLETED

Step 1 Check or complete software installation You must have the following software libraries installed: Freesurfer: https://s... read more

Step 2 Set up your .bashrc file Open gedit and make sure your .bashrc file contains the following text:

Step 3 Set up directory structure You must have the appropriate input files in the displayed directory structure: Inside SUB_... read more

Step 4 Put the fsaverageSubP folder in the SUB_DATA directory The fsaverageSubP directory contains the 308 parcellation and nee... read more

Step 1

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
```

Help

Coding your analyses



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Comments are your friend!



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

```
#!/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
#      zero for the quantitative MPM images and uses the brain mask to create
#      brain extracted versions of the MPM images (where all voxels outside of
#      the brain have been set to zero.
#
# USAGE:
#    NSPN_mpm_bet_mask.sh <pdw_file>
#
# INPUTS:
#    pdw_file : Proton density weighted file to which the MPM
#                  quantitative maps are aligned.
#
# EXPECTS:
#    The following files should be in the same directory as the
#      input file:
#
#        R1.nii.gz         MT.nii.gz
#        R1s.nii.gz        A.nii.gz
#
# OUTPUTS:
#    All output are in the same directory as the input file.
#    A sub-directory called PDw_bet is created and contains all the
#      files created by FSL's bet command
#
#       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
fi

#==============================================================
# Now make the brain and head files for each of the
# calculated MPM files
#==============================================================
echo -n "  Applying masks"
for f_name in PDw ${calc_filename_list[@]}; do

    # Don't run if it's already complete!
    if [[ ! -f ${mpm_dir}/${f_name}_head.nii.gz ]]; then
        echo -n " - ${f_name}"
        fslmaths ${bet_dir}/PDw_brain_ero3.nii.gz \
                    -bin \
                    -mul ${mpm_dir}/${f_name}.nii.gz \
                    ${mpm_dir}/${f_name}_brain.nii.gz

        fslmaths ${bet_dir}/PDw_brain_outskin_mask.nii.gz \
                    -bin \
                    -mul ${mpm_dir}/${f_name}.nii.gz \
                    ${mpm_dir}/${f_name}_head.nii.gz
    fi
done # Close the mpm calculated file loop
echo ""

#==============================================================
# All done!
#==============================================================
```



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Aim for 40% comments in your code



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Share your comments with the original author

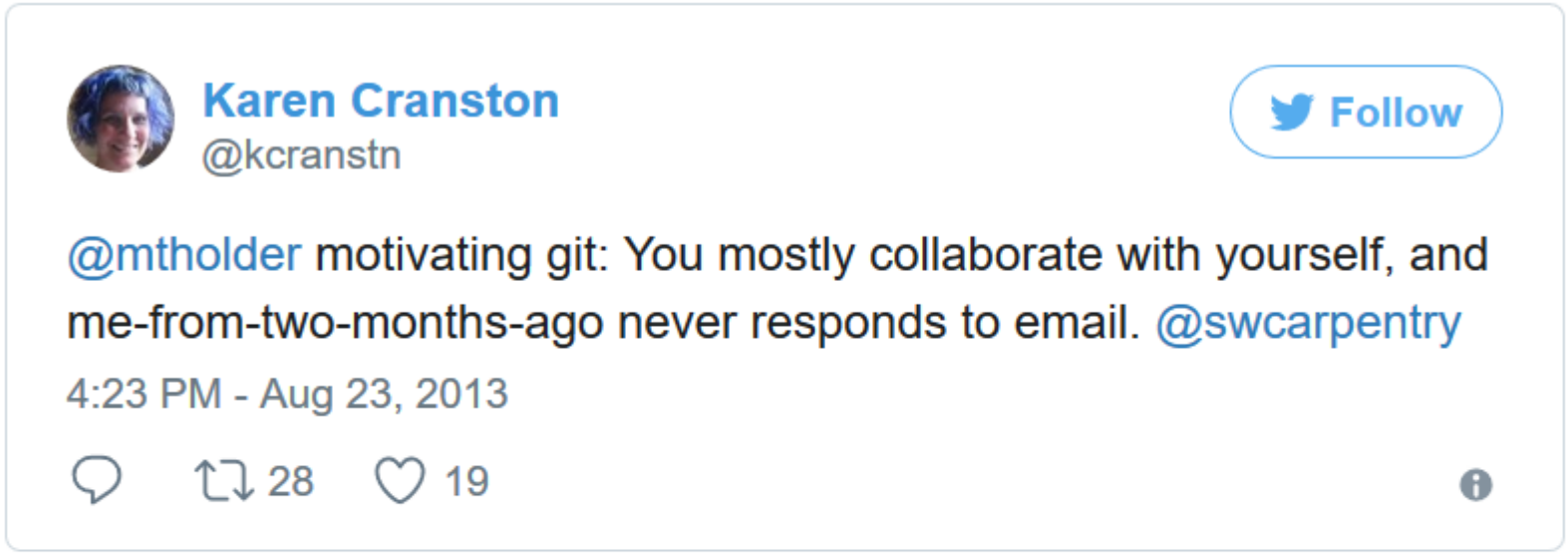


@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

(which is almost always going to be YOU!)



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201



** 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

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



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

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



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201



https://education.github.com



@kirstie_j

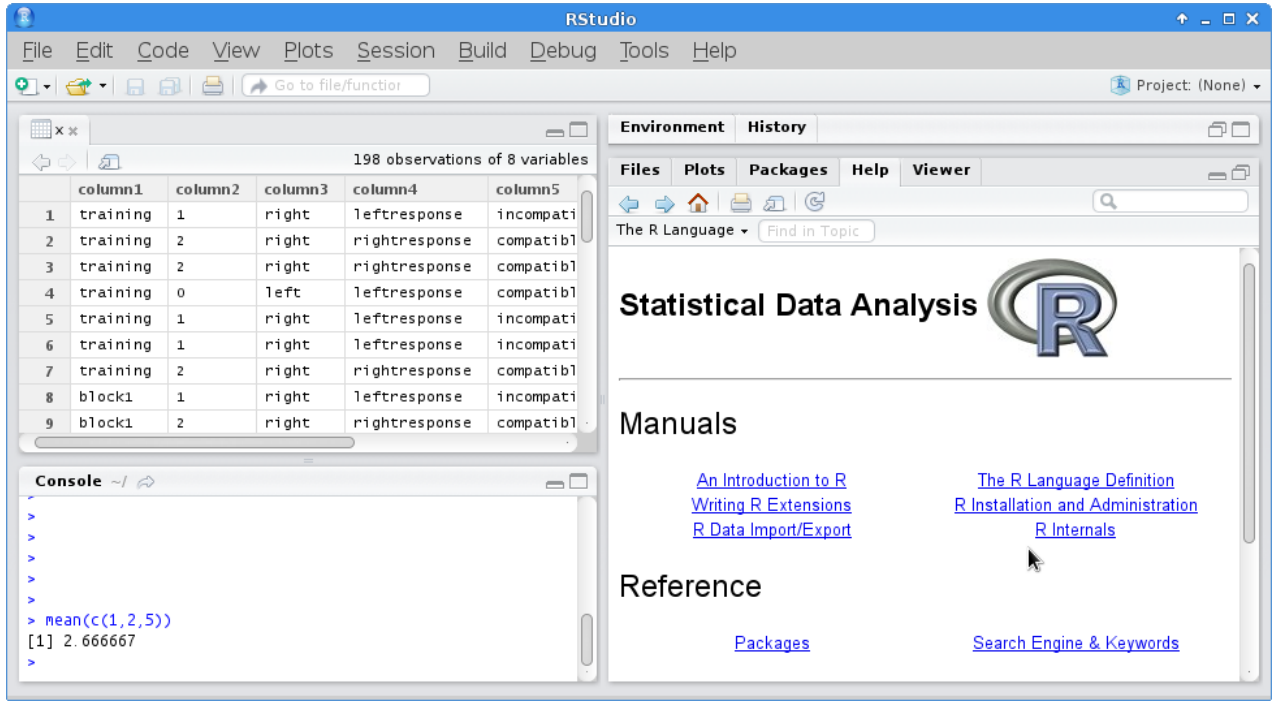
https://GitHub.com/KirstieJane/ReproducibleResearch

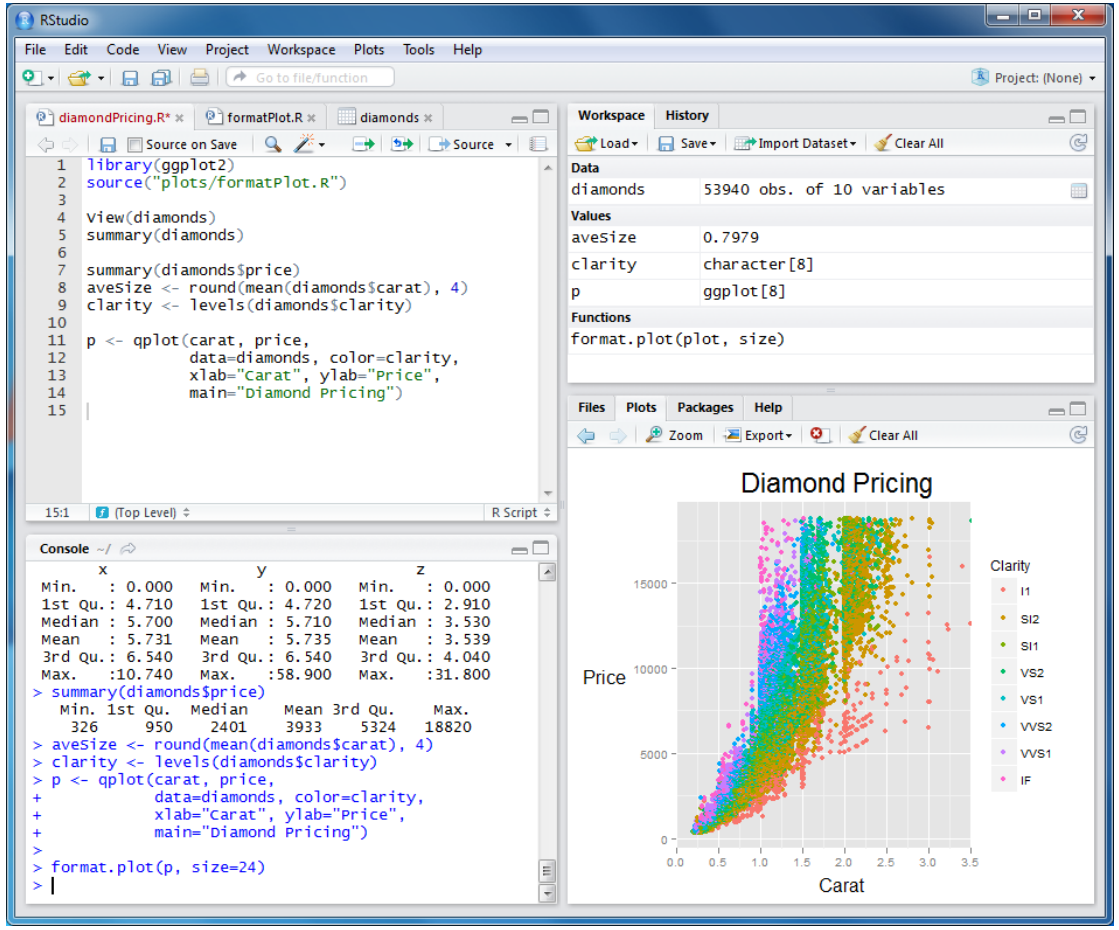
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Python vs R
(vs Matlab vs STATA etc…)



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201







@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Shiny by RStudio

Search

OVERVIEW

TUTORIAL

ARTICLES

GALLERY

REFERENCE

DEPLOY

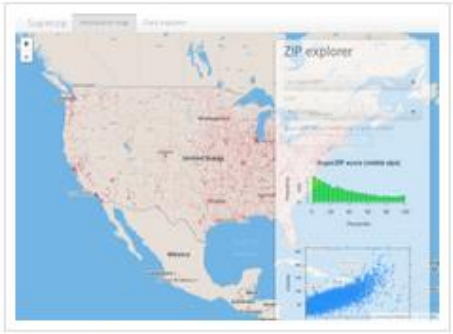
HELP

Gallery

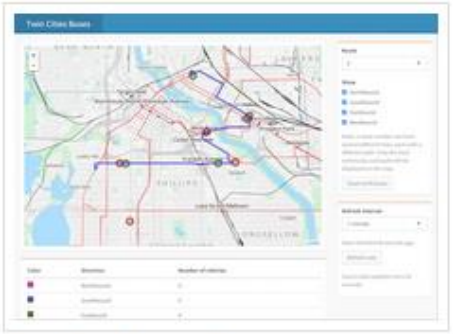
This gallery contains useful examples to learn from. Visit the Shiny User Showcase to see an inspiring set of sophisticated apps.

Interactive visualizations

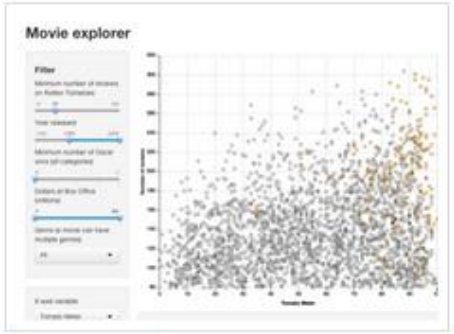
Shiny is designed for fully interactive visualization, using JavaScript libraries like d3, Leaflet, and Google Charts.



SuperZip example



Bus dashboard



Movie explorer



Google Charts



Shiny gallery

@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

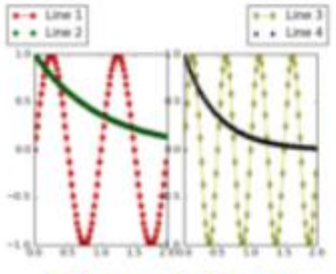


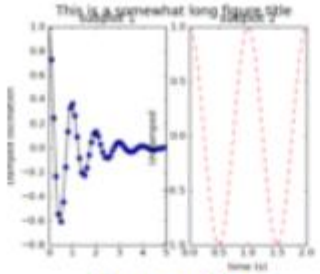


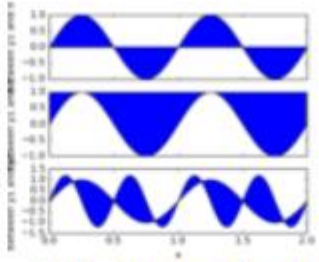
@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

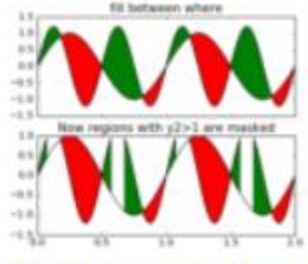
Matplotlib gallery

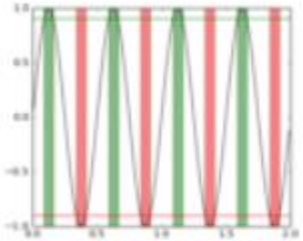
Click each example to see source code

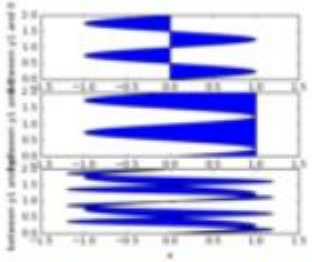

figlegend_demo

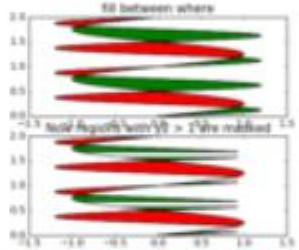

figure_title

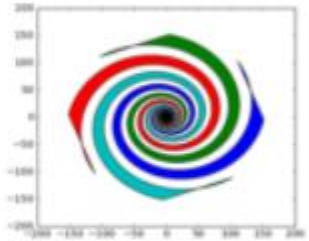

fill_between_demo


fill_between_demo

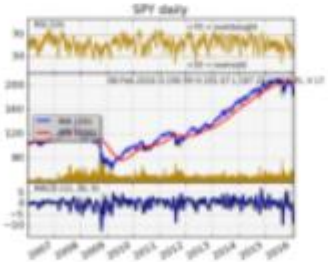

fill_between_demo

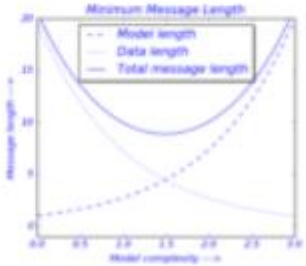

fill_betweenx_demo

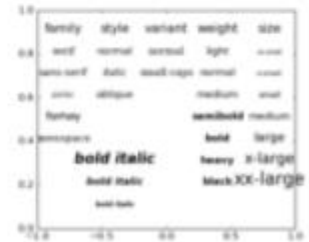

fill_betweenx_demo


fill_spiral


finance_demo


finance_work2


findobj_demo


fonts_demo





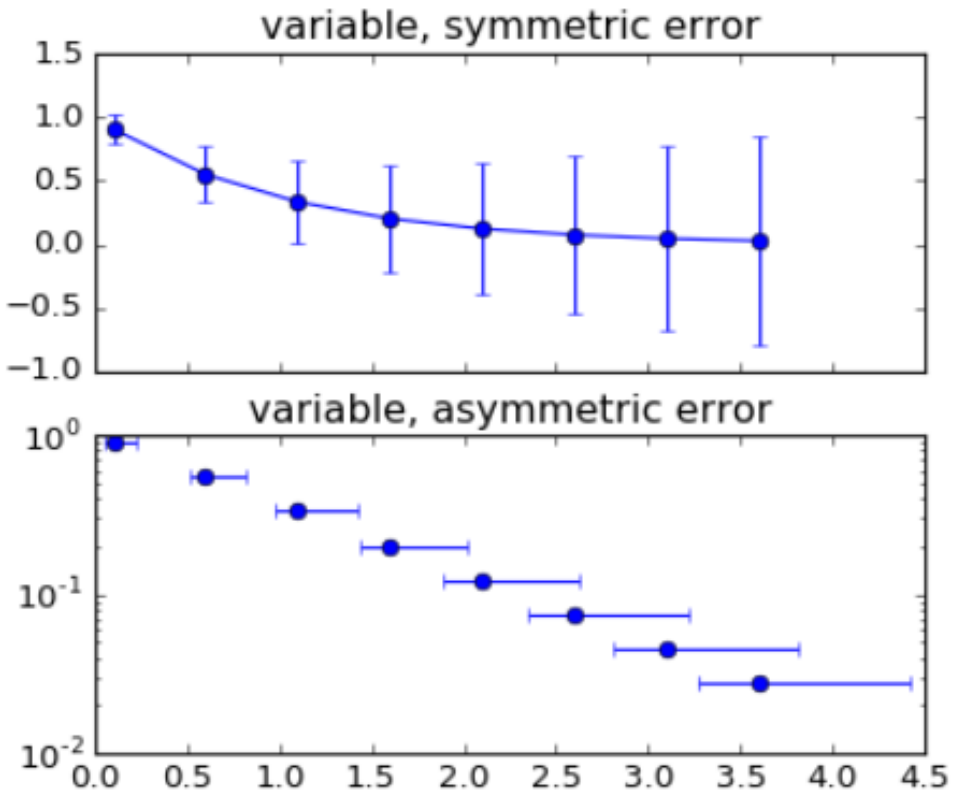






statistics example code: errorbar_demo_features.py

(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)
y = np.exp(-x)
# 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

@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Jupyter Notebook



http://jupyter.org



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

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/
- 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 b
can run the python script GW150914_tutorial.py. You will need the python packa

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

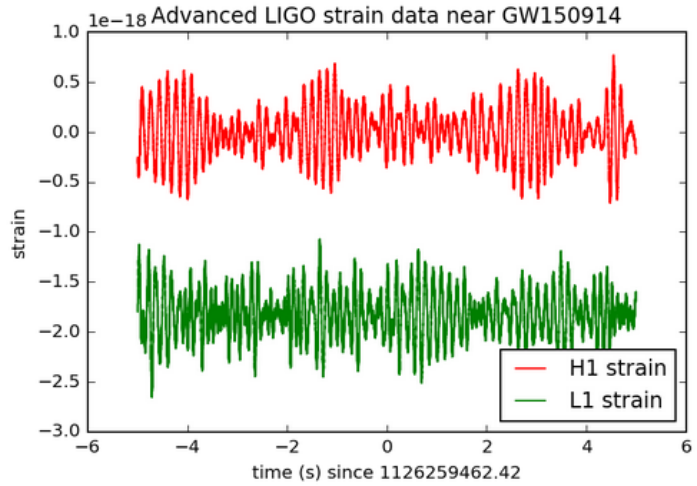
Questions, comments, suggestions, corrections, etc: email losc@ligo.org

v20160208b

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

In [6]:
```
# 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')
```



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)



Asking for help: StackOverflow & MWEs

http://stackoverflow.com



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201



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

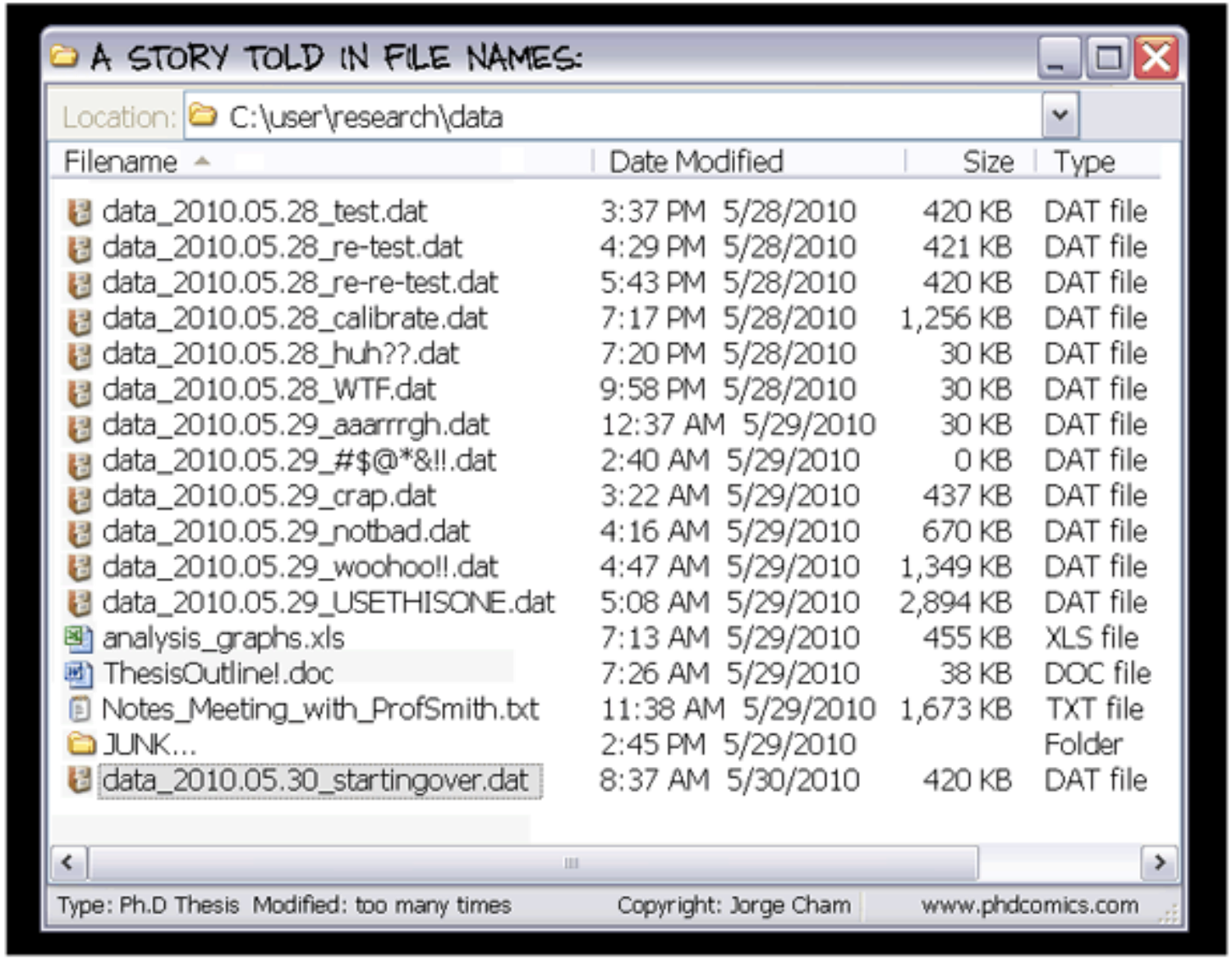


@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Version Control



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

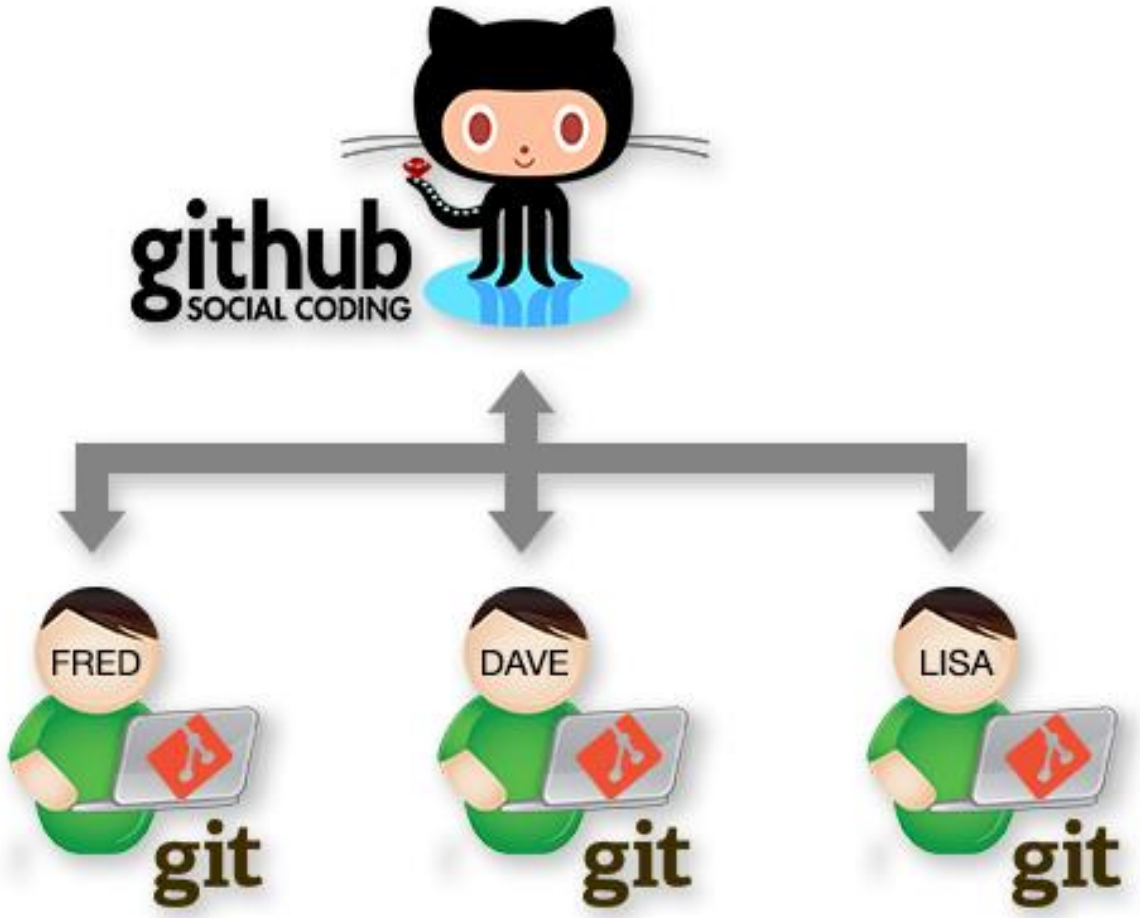




@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201





@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

Repository

Clone

Pull request

Branch

Version control

Some jargon busting

Merge

Issues

Release

Commit

Fork

Markdown



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201

Submit your first pull request!

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



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201

ReproducibleResearch

This repository supports Kirstie's presentation on tips and tricks for making your research reproducible.

The goal is to build a directory of useful links, and a jargon busting glossary.

Guide for contributors

See our guidelines for how to contribute to the project.

Code of conduct

Everyone is welcome to join this project, particularly people who have not used GitHub before and are feeling unsure of how to begin!

Please follow our code of conduct in all your on and offline interactions.



© 2016 iLoveHeartStudio.com



@kirstie_j
https://GitHub.com/KirstieJane/ReproducibleResearch
doi: https://dx.doi.org/10.6084/m9.figshare.443201



DATA CARPENTRY

software carpentry

Community

https://www.mozillascience.org

http://software-carpentry.org

http://data-carpentry.org

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

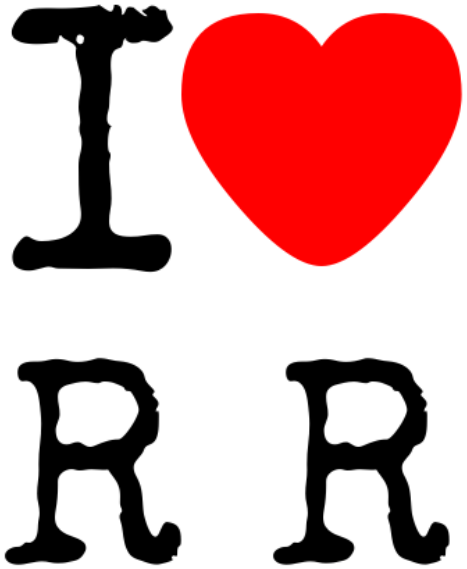
https://www.coursera.org/learn/python/home/info



@kirstie_j

https://GitHub.com/KirstieJane/ReproducibleResearch

doi: https://dx.doi.org/10.6084/m9.figshare.443201



Thank you!

KirstieJane/ReproducibleResearch

@kirstie_j

doi: 10.6084/m9.figshare.4244996

