Showing your working: a how to guide to reproducible research







- 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

% @kirstie_j

Same	Different
Reproducible	



python

y @kirstie_j

Same	Different
Reproducible	Replicable



python



Same	Different
Reproducible	Replicable
Robust	



g@kirstie_j

Same Different Reproducible Replicable Generalisable Robust



Different

¥ @kirstie_j

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



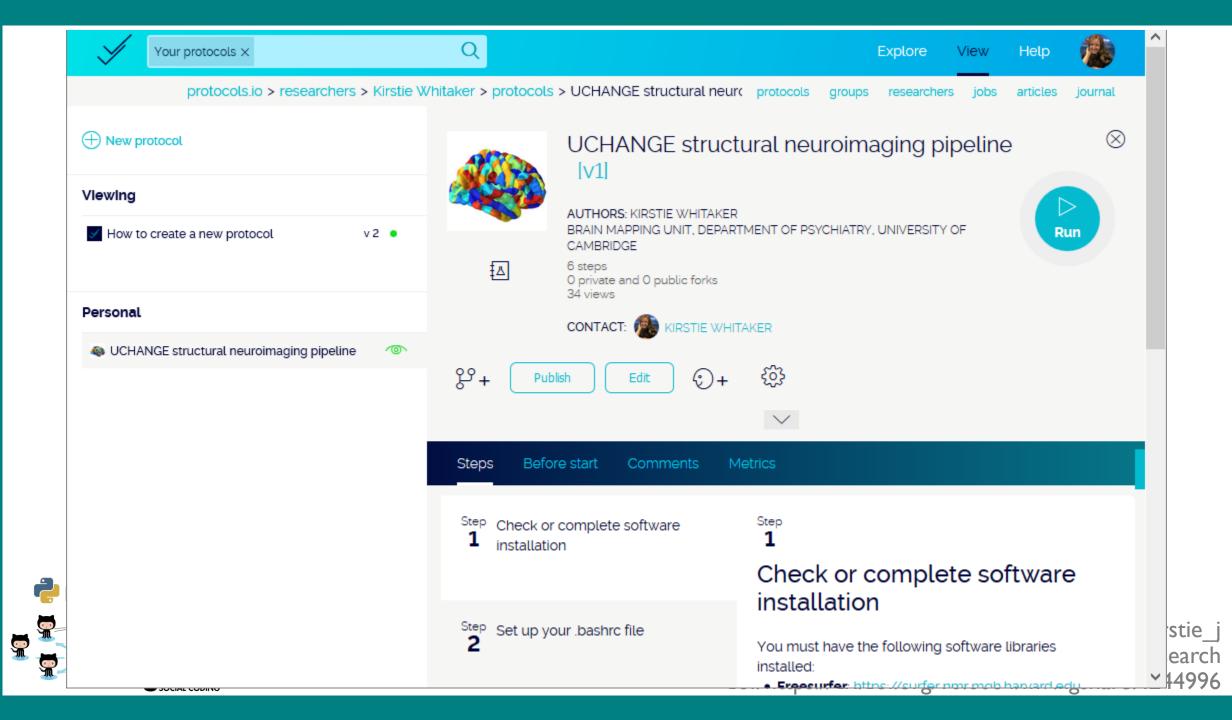
Start small

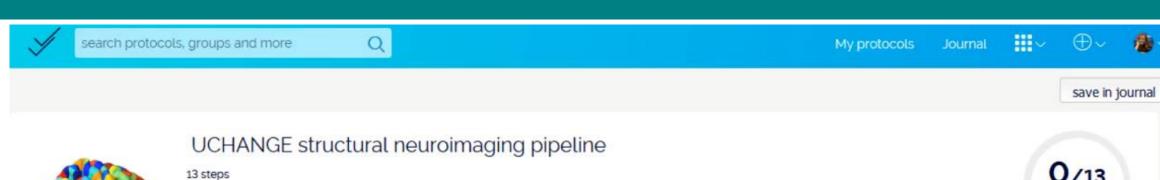


Protocols.io









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 Check or complete software installation Set up your bashrc file Open gedit and make sure your bashrc file contains You must have the following software libraries installed: the following text: • 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 Set up directory structure You must have the appropriate input files in the displayed directory structure: Inside SUB_... read more below: VTK Mayavi Nibabel pysurfer

Put the fsaverageSubP folder in the SUB_DATA directory The fsaverageSubP

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





(LINUX)

COMMAND

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!)



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

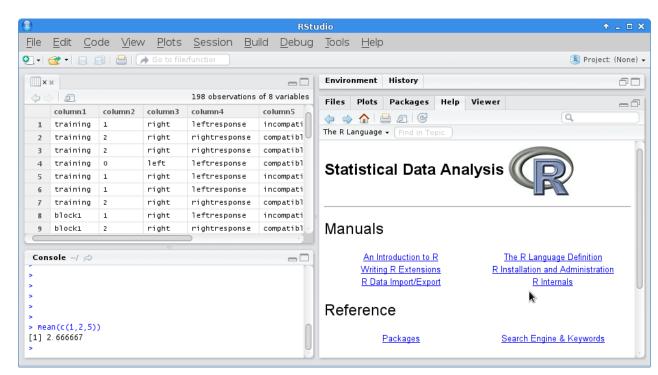


R Studio

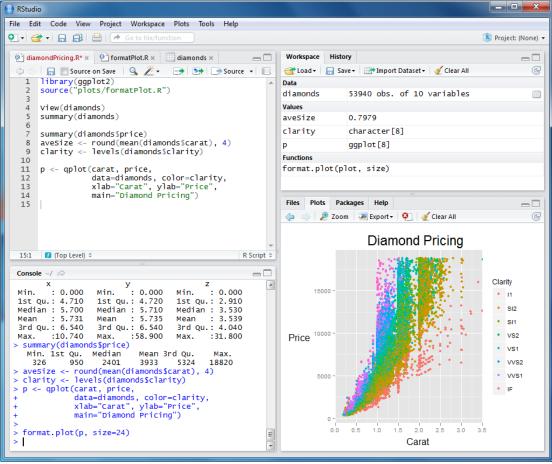
https://www.rstudio.com











g @kirstie_j

OVERVIEW

TUTORIAL

ARTICLES

GALLERY

REFERENCE

DEPLOY

HELP

Gallery

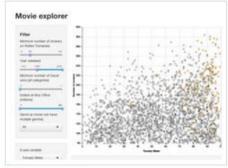
This gallery contains useful examples to learn from. Visit the Shiny User Showcase to see an inspiring set of sophisticated apps.

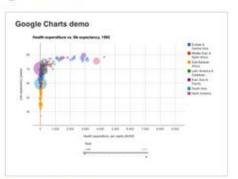
Interactive visualizations

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 & Seaborn

http://matplotlib.org

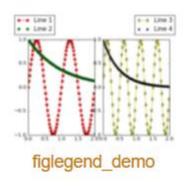
http://stanford.edu/~mwaskom/software/seaborn

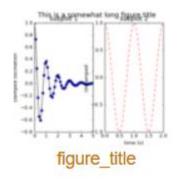


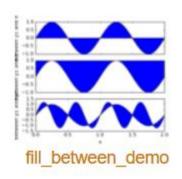


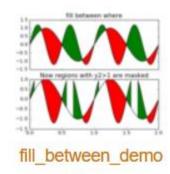
Matplotlib gallery

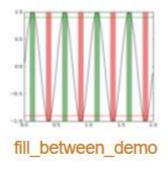
Click each example to see source code

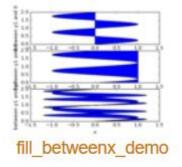


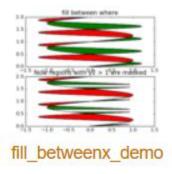


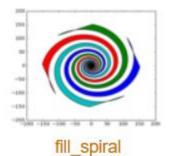






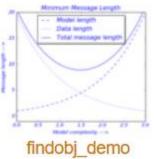


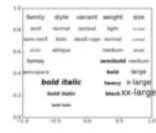












no fonts demo





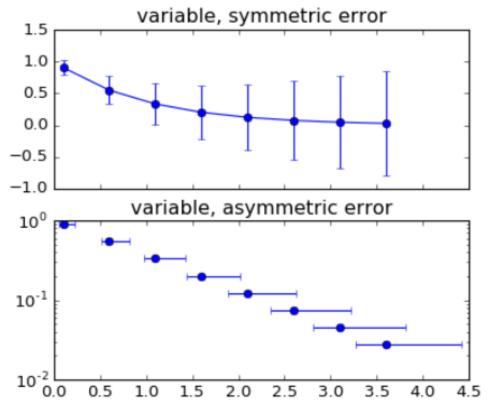






statistics example code: errorbar_demo_features.py

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

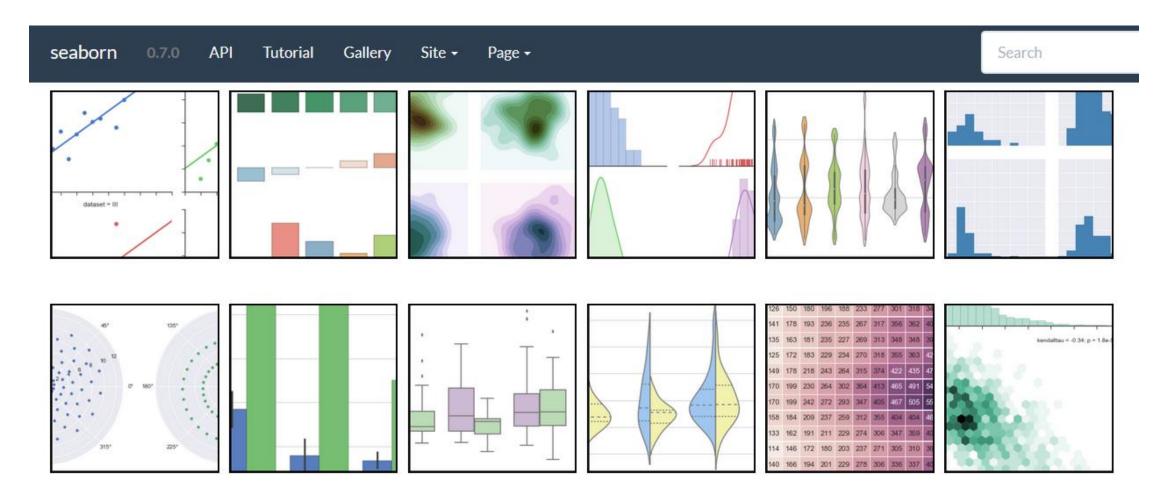


```
Demo of errorbar function with different ways of specifying error bars.
Errors can be specified as a constant value (as shown in `errorbar demo.py`),
or as demonstrated in this example, they can be specified by an N x 1 or 2 x N,
where N is the number of data points.
N x 1:
    Error varies for each 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







Seaborn gallery

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

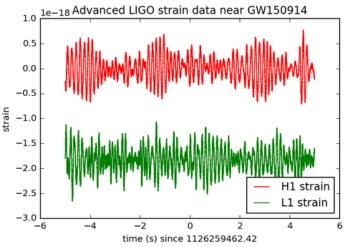
https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.html



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)



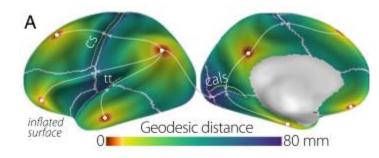


101. 11cp3// 47.401.01 8/ 10.000 1/11/.118311a1 c. 12 1 1//

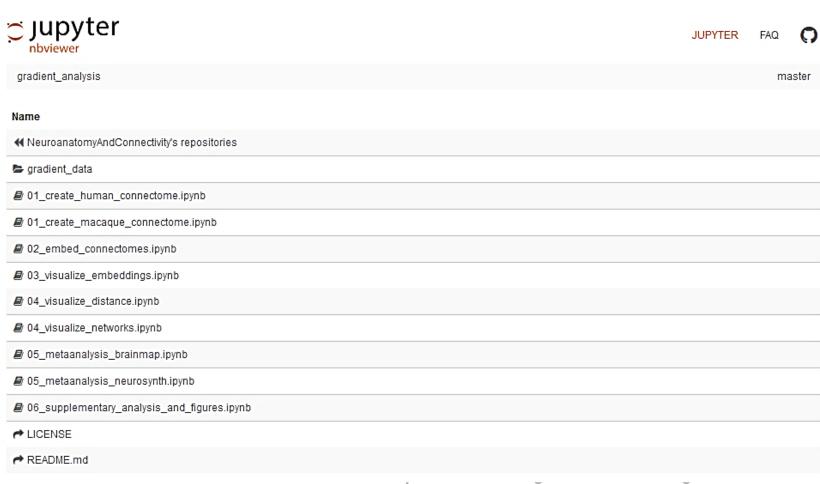
Situating the default-mode network along a principal gradient of macroscale cortical organization

Daniel S. Margulies^{a,1}, Satrajit S. Ghosh^{b,c}, Alexandros Goulas^d, Marcel Falkiewicz^a, Julia M. Huntenburg^{a,e}, Georg Langs^{f,g}, Gleb Bezgin^h, Simon B. Eickhoff^{i,j}, F. Xavier Castellanos^{k,l}, Michael Petrides^m, Elizabeth Jefferies^{n,o}, and Jonathan Smallwood^{n,o}

DOI: 10.1073/pnas.1608282113











```
gradient_analysis / 04_visualize_distance.ipynb
```

jupyter

```
In [1]: % matplotlib inline
        import nibabel as nib
        import nibabel.gifti
        import nilearn
        from nilearn import datasets, plotting
        import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib as mpl
        import matplotlib.pyplot as plt
        from matplotlib.colors import ListedColormap
        mpl.rcParams['svg.fonttype'] = 'none'
        import sys
        sys.path.append('gradient data/src/')
        from viz import *
        import surfdist as sd
        from surfdist import viz, load, utils, surfdist
        /Applications/miniconda3/envs/topography/lib/python2.7/site-packages/matplotlib/font manager.py:273: UserWarning: Matplo
          warnings.warn('Matplotlib is building the font cache using fc-list. This may take a moment.')
        /Applications/miniconda3/envs/topography/lib/python2.7/site-packages/matplotlib/ init .py:872: UserWarning: axes.color
          warnings.warn(self.msg depr % (key, alt key))
```

Asking for help: StackOverflow & MWEs

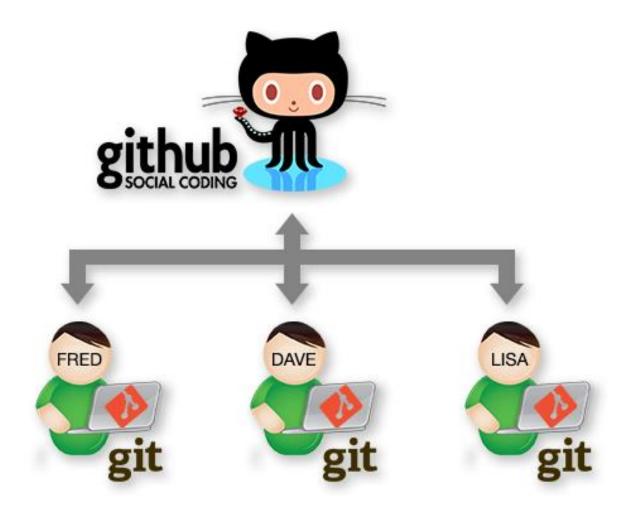
http://stackoverflow.com





Git vs GitHub

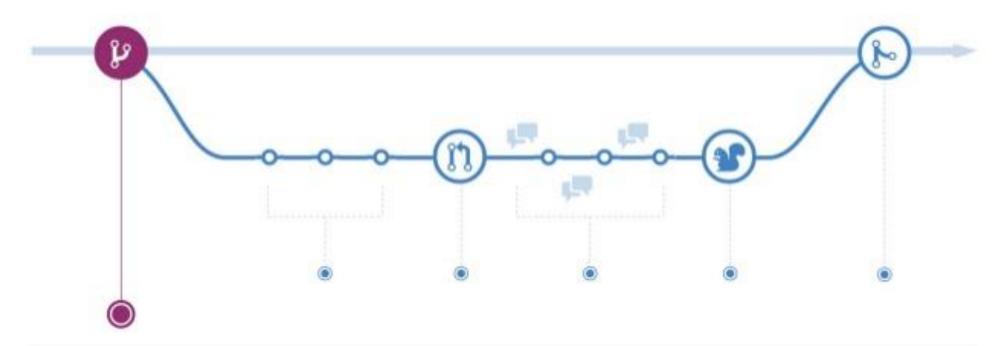






GitHub Flow - Create a Branch





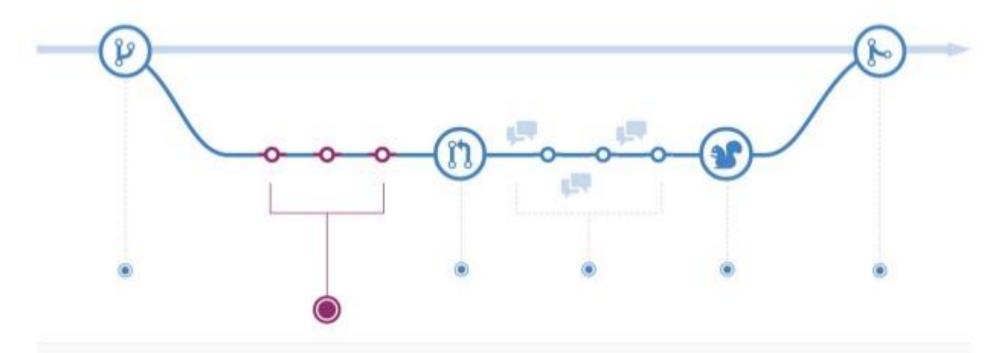




J@kirstie_j

GitHub Flow - Add Commits

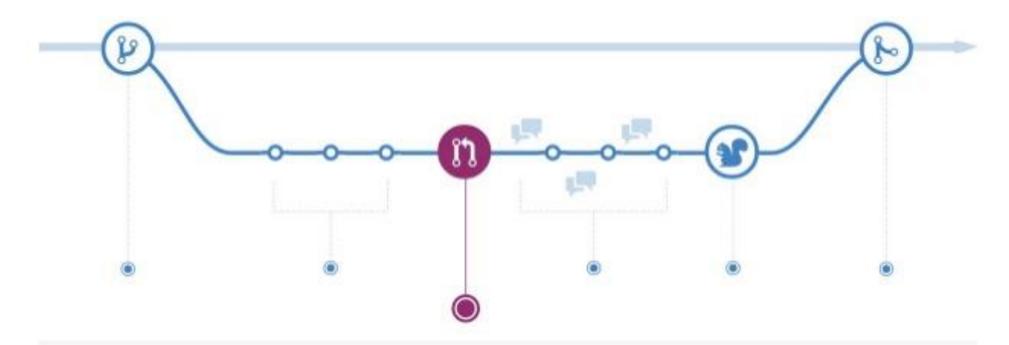






GitHub Flow - Open a pull request



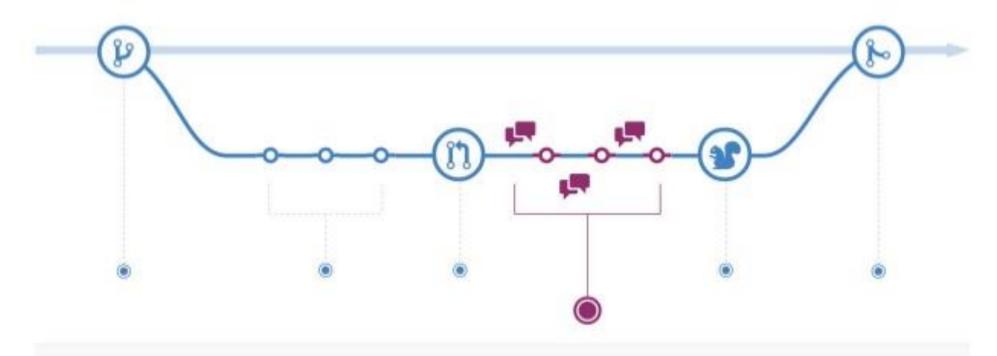






GitHub Flow - Discuss and Review your code









GitHub Flow - Deploy / Test



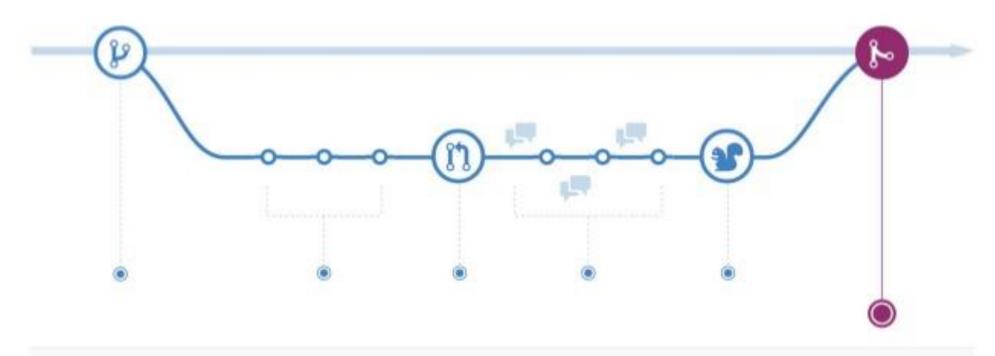






GitHub Flow - Merge









www.GitHub.com/ KirstieJane/ ReproducibleResearch





Repository

Clone

Branch

Pull request

Version control

Some jargon busting

Merge

Issues

Release

Commit

Fork



Markdown



https://GitHub.com/KirstieJane/ReproducibleResearch doi: https://dx.doi.org/10.6084/m9.figshare.4244996

Submit your first pull request!

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





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.



@ 2016 iLoveHeartStudio.com

Code of conduct

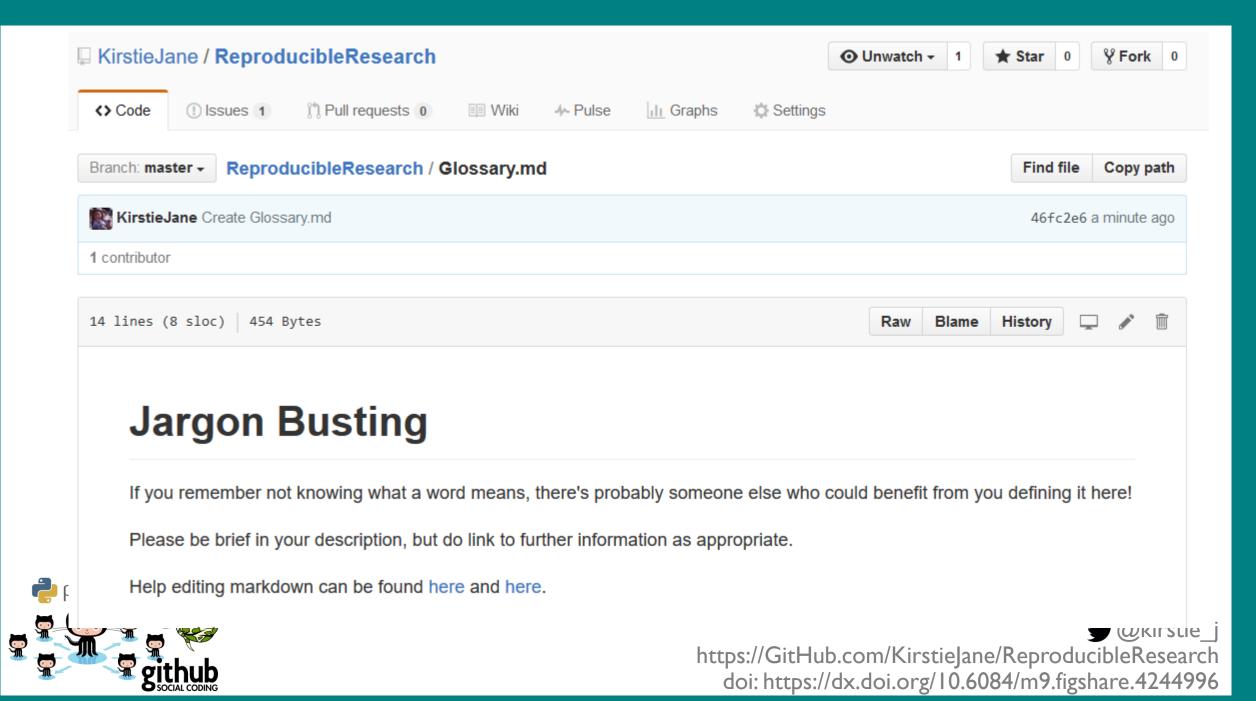
Everyone is welcome to join this project, particularly people who have not used GitHub before and are feeling unsure of how to begin!





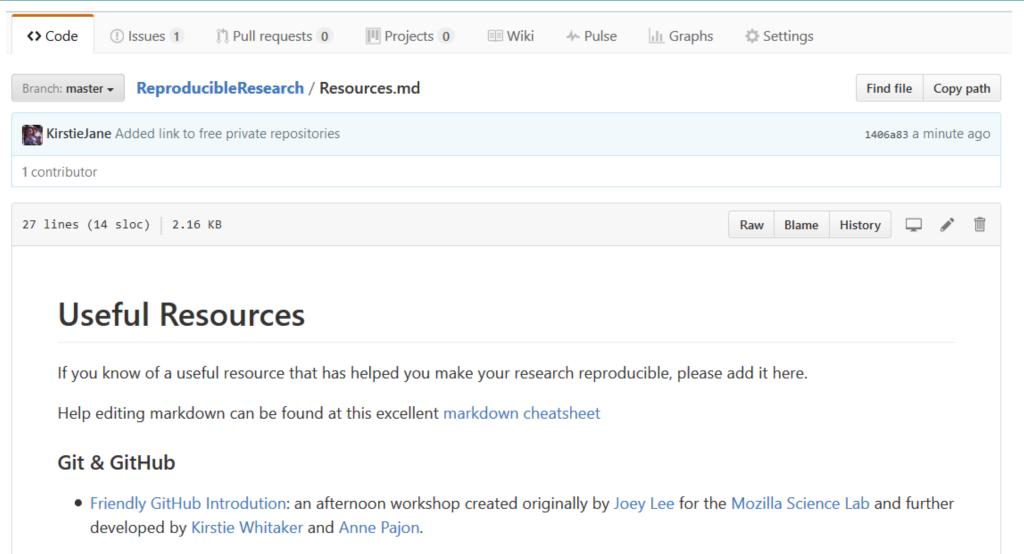
Create a glossary





Share resources







• Get private GitHub repositories: a paid GitHub account allows you to have private repositories that only people you choose to make collaborators can see. If you're a member of an academic institution you can request a discount (to \$0) and therefore have free private repositories. The discount lasts for 2 years but you can request another at the end of that period no problem .

@kirstie_j eResearch e.4244996

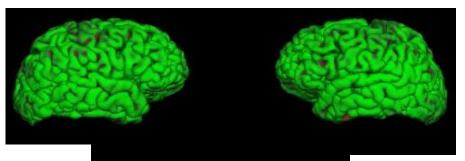


Nipype: **Neuroimaging in Python Pipelines and Interfaces**











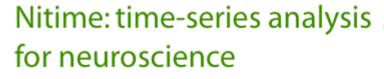
MEG + EEG ANALYSIS & VISUALIZATION

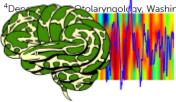
Reconstruction Tracking Deterministic Probabilistic Registration Streamline-based Image-based Streamline-based Segmentation Image-based

Automatic analysis (aa): efficient **FreeSurfer** neuroimaging workflows and parallel processing using Matlab and XML

> Rhodri Cusack^{1*}, Alejandro Vicente-Grabovetsky², Daniel J. Mitchell³, Conor J. Wild¹, Tibor Auer³, Annika C. Linke¹ and Jonathan E. Peelle⁴

claryngology, Washington University in St. Louis, St. Louis, MO, USA





¹Brain and Mind Institute, Western University, London, ON, Canada

²Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

³MRC Cognition and Brain Sciences Unit, Cambridge, UK



DATA CARPENTRY UIIII



Community

https://www.mozillascience.org

http://software-carpentry.org

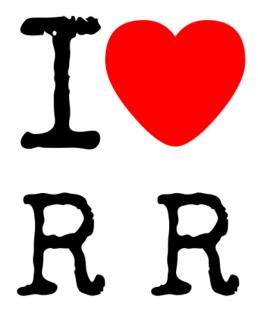
http://data-carpentry.org

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

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







Thank you!

© 2016 iLoveHeartStudio.com



KirstieJane/ReproducibleResearch



@kirstie_j



doi: 10.6084/m9.figshare.4244996



y @kirstie j

https://GitHub.com/KirstieJane/ReproducibleResearch doi: https://dx.doi.org/10.6084/m9.figshare.4244996