

# Project organisation

Complex Human Data Summer  
School

# Why workflow matters

- A good workflow
  - saves time in the long run
  - helps you avoid errors
  - makes it easy for others to reproduce your work
- Folder organization is critical for all of these things

mcmcsamples\_expt2\_subj19\_scenario1.mat  
mcmcsamples\_expt2\_subj19\_scenario2.mat  
mcmcsamples\_expt2\_subj19\_scenario3.mat  
mcmcsamples\_expt2\_subj20\_scenario1.mat  
mcmcsamples\_expt2\_subj20\_scenario2.mat  
mcmcsamples\_expt2\_subj20\_scenario3.mat  
modelfits1.mat  
modelfits1 copy.mat  
modelfits2.mat  
modelfits2 copy.mat  
parameterdescriptives.m  
participant1scenario3expt1.eps  
participant4scenario3expt1.eps  
participant10scenario3expt1.eps  
participant13scenario1expt1.eps  
participant15scenario1expt1.eps  
participant19scenario3expt1.eps  
recalculateVAF.m  
recalculateVAF.m~  
showtestcases.m  
stupidplots.m  
stupidplots.m~  
stupidplots2.m  
stupidplots2.m~  
vafplots.m  
.DS Store

# One of Danielle Navarro's folders from 2009

mcmcsamples\_expt2\_subj19\_scenario1.mat  
mcmcsamples\_expt2\_subj19\_scenario2.mat  
mcmcsamples\_expt2\_subj19\_scenario3.mat  
mcmcsamples\_expt2\_subj20\_scenario1.mat  
mcmcsamples\_expt2\_subj20\_scenario2.mat  
mcmcsamples\_expt2\_subj20\_scenario3.mat  
modelfits1.mat  
modelfits1 copy.mat  
modelfits2.mat  
modelfits2 copy.mat  
parameterdescriptives.m  
participant1scenario3expt1.eps  
participant4scenario3expt1.eps  
participant10scenario3expt1.eps  
participant13scenario1expt1.eps  
participant15scenario1expt1.eps  
participant19scenario3expt1.eps  
recalculateVAF.m  
recalculateVAF.m~  
showtestcases.m  
stupidplots.m  
stupidplots.m~  
stupidplots2.m  
stupidplots2.m~  
vafplots.m  
.DS\_Store

Multiple copies of  
files

mcmcsamples\_expt2\_subj19\_scenario1.mat  
mcmcsamples\_expt2\_subj19\_scenario2.mat  
mcmcsamples\_expt2\_subj19\_scenario3.mat  
mcmcsamples\_expt2\_subj20\_scenario1.mat  
mcmcsamples\_expt2\_subj20\_scenario2.mat  
mcmcsamples\_expt2\_subj20\_scenario3.mat  
modelfits1.mat  
modelfits1 copy.mat  
modelfits2.mat  
modelfits2 copy.mat  
parameterdescriptives.m  
participant1scenario3expt1.eps  
participant4scenario3expt1.eps  
participant10scenario3expt1.eps  
participant13scenario1expt1.eps  
participant15scenario1expt1.eps  
participant19scenario3expt1.eps  
recalculateVAF.m  
recalculateVAF.m~  
showtestcases.m  
stupidplots.m  
stupidplots.m~  
stupidplots2.m  
stupidplots2.m~  
vafplots.m  
.DS\_Store

White space in  
filenames is a bad idea

- mcmcsamples\_expt2\_subj19\_scenario1.mat
- mcmcsamples\_expt2\_subj19\_scenario2.mat
- mcmcsamples\_expt2\_subj19\_scenario3.mat
- mcmcsamples\_expt2\_subj20\_scenario1.mat
- mcmcsamples\_expt2\_subj20\_scenario2.mat
- mcmcsamples\_expt2\_subj20\_scenario3.mat
- modelfits1.mat
- modelfits1 copy.mat
- modelfits2.mat
- modelfits2 copy.mat
- parameterdescriptives.m
- participant1scenario3expt1.eps
- participant4scenario3expt1.eps
- participant10scenario3expt1.eps
- participant13scenario1expt1.eps
- participant15scenario1expt1.eps
- participant19scenario3expt1.eps
- recalculateVAF.m
- recalculateVAF.m~
- showtestcases.m
- stupidplots.m
- stupidplots.m~
- stupidplots2.m
- stupidplots2.m~
- vafplots.m
- .DS\_Store

Slightly embarrassing  
filenames

mcmcsamples\_expt2\_subj19\_scenario1.mat  
mcmcsamples\_expt2\_subj19\_scenario2.mat  
mcmcsamples\_expt2\_subj19\_scenario3.mat  
mcmcsamples\_expt2\_subj20\_scenario1.mat  
mcmcsamples\_expt2\_subj20\_scenario2.mat  
mcmcsamples\_expt2\_subj20\_scenario3.mat  
modelfits1.mat  
modelfits1 copy.mat  
modelfits2.mat  
modelfits2 copy.mat  
parameterdescriptives.m  
participant1scenario3expt1.eps  
participant4scenario3expt1.eps  
participant10scenario3expt1.eps  
participant13scenario1expt1.eps  
participant15scenario1expt1.eps  
participant19scenario3expt1.eps  
recalculateVAF.m  
recalculateVAF.m~  
showtestcases.m  
stupidplots.m  
stupidplots.m~  
stupidplots2.m  
stupidplots2.m~  
vafplots.m  
.DS\_Store

Inconsistent separator  
characters

# Inconsistent naming scheme

mcmcsamples\_expt2\_subj19\_scenario1.mat  
mcmcsamples\_expt2\_subj19\_scenario2.mat  
mcmcsamples\_expt2\_subj19\_scenario3.mat  
mcmcsamples\_expt2\_subj20\_scenario1.mat  
mcmcsamples\_expt2\_subj20\_scenario2.mat  
mcmcsamples\_expt2\_subj20\_scenario3.mat  
modelfits1.mat  
modelfits1 copy.mat  
modelfits2.mat  
modelfits2 copy.mat  
parameterdescriptives.m  
participant1\_scenario3expt1.eps  
participant4scenario3expt1.eps  
participant10scenario3expt1.eps  
participant13scenario1expt1.eps  
participant15scenario1expt1.eps  
participant19scenario3expt1.eps  
recalculateVAF.m  
recalculateVAF.m~  
showtestcases.m  
stupidplots.m  
stupidplots.m~  
stupidplots2.m  
stupidplots2.m~  
vafplots.m  
.DS\_Store



mcmcsamples\_expt2\_subj19\_scenario1.mat  
mcmcsamples\_expt2\_subj19\_scenario2.mat  
mcmcsamples\_expt2\_subj19\_scenario3.mat  
mcmcsamples\_expt2\_subj20\_scenario1.mat  
mcmcsamples\_expt2\_subj20\_scenario2.mat  
mcmcsamples\_expt2\_subj20\_scenario3.mat  
modelfits1.mat  
modelfits1 copy.mat  
**modelfits2.mat**  
modelfits2 copy.mat  
parameterdescriptives.m  
participant1scenario3expt1.eps  
participant4scenario3expt1.eps  
participant10scenario3expt1.eps  
participant13scenario1expt1.eps  
participant15scenario1expt1.eps  
participant19scenario3expt1.eps  
recalculateVAF.m  
recalculateVAF.m~  
showtestcases.m  
stupidplots.m  
stupidplots.m~  
stupidplots2.m  
stupidplots2.m~  
vafplots.m  
.DS\_Store

Filename that aren't  
very helpful

# Not just Dani!

README  
allconceptslow.m  
allconceptveryslow.m  
atomiceval.m  
atomicevalmex.mexmaci  
bool.mexmaci  
booleval.m  
boolevalm.m  
boolevalmBROKEN.m  
boolevalmex.mexmaci  
boolvs.m  
choosestim3.m  
cnf/  
cnf.m  
cnf\_prep.m  
cnfold.m  
combineconj.m  
combineconj2.m  
completiongroups.mat  
computeextensions.m  
computeextensions4.m  
computeextensions4PAR.m  
computeextensions5.m  
computeextensions5PAR.m  
computeextensions5expt.m  
computeextensionsOLD.m  
computeextensionscompletion.m  
computeextensionsspar.m  
computeextensionstweak.m  
computeprior.m  
conceptnum3.m  
conjsizes3.m  
conjsizes4.m  
conjunctions.m  
conjunctions3.m  
conjunctions4.m  
conjunctionsOLD.m  
conjunctionslow.m  
consistent.m  
dnf.m  
dnf\_prep.m  
enumprop.m  
enumquant.m  
expprod.m

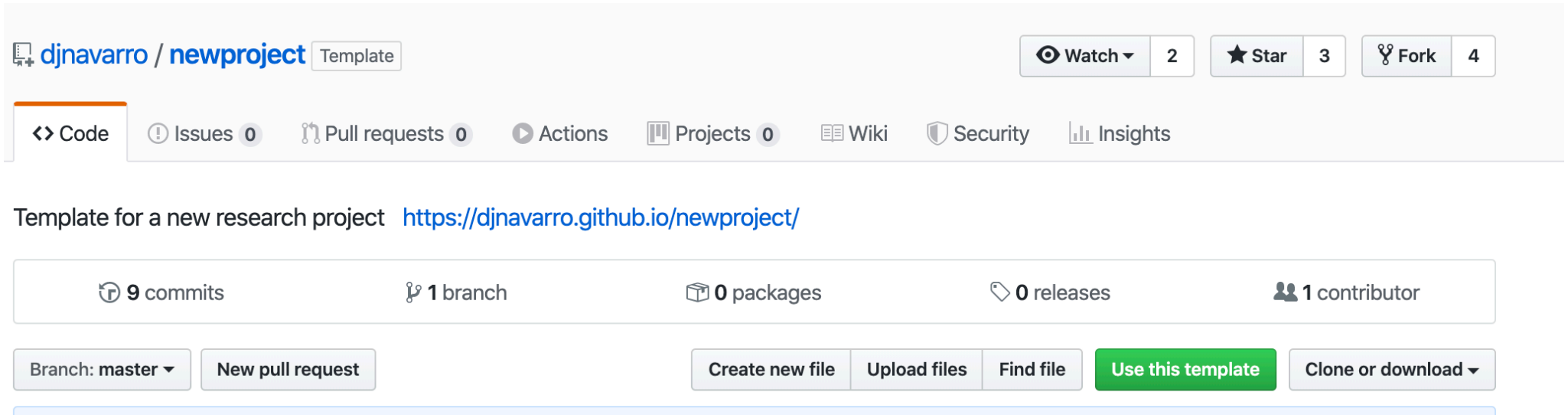
expt/  
exptoneshot.m  
exptquartet.m  
exptquartetconj.m  
expttreadsOLD.m  
expttriads.m  
expttriads3.m  
expttriads3conj.m  
feat3lev3.mat  
feat3lev3\_2ac.mat  
feat3lev3\_3ac.mat  
feat3lev3\_selectedtriple.mat  
feat3lev3\_selectedtripleOLD.mat  
feat3lev3\_selectedtripleTEST.mat  
fparse.m  
fsign3v3.mat  
fsign3v3prior.mat  
fsign3v5.mat  
fsign3v5prior.mat  
fsign5v5.mat  
implicationremove.m  
loopbool.m  
makefeatprops.m  
makemodels5.m  
makemodelscompletion.m  
makeprops.m  
makeprops2.m  
makeprops3.m  
makepropsTEST.m  
makequartetsactual.m  
maketriads.m  
maketriadsOLD.m  
maketriadsort.m  
maketriadsortactual.m  
mergeconjuncts.m  
mergedisjuncts.m  
mex/  
modelgroups.m  
modexpt2.mat  
modexpt2miss.m  
mygrid.m  
negationin.m  
newpreds.txt  
notes.txt

oldpreds.txt  
oneshotadd.m  
oneshotmodelpreds.mat  
parallel/  
parteval.m  
proporder1.txt  
proporder2.txt  
proporder3.txt  
propositionalize.m  
propositionalizeBUG.m  
propositionalizeOLD.m  
propstring.m  
propstringTEST.m  
propstringglobal.m  
propstringnew.m  
propstringnewtest.m  
propsub.m  
run4combine.m  
saveCONJ.m  
savePAR.m  
schema4top.mat  
schema4topconj.mat  
schemasizes4.m  
schemasizes4conj.m  
testcnf.m  
testdnf.m  
testmex.m  
testout.m  
testrule.m  
testruleOLD.m  
testtriad.m  
testtriadOLD.m  
testtriadconj.m  
threebythreeexpt.txt  
triadmodelpreds.mat  
triadmodelpredsconj.mat  
urows.m  
urows4.m  
urowsOLD.m  
urowsconj3.m  
urowsconj4.m  
varconstrain.m  
varconstraindnf.m  
variablize.m

File organization and naming are  
powerful weapons against chaos

– Jenny Bryan

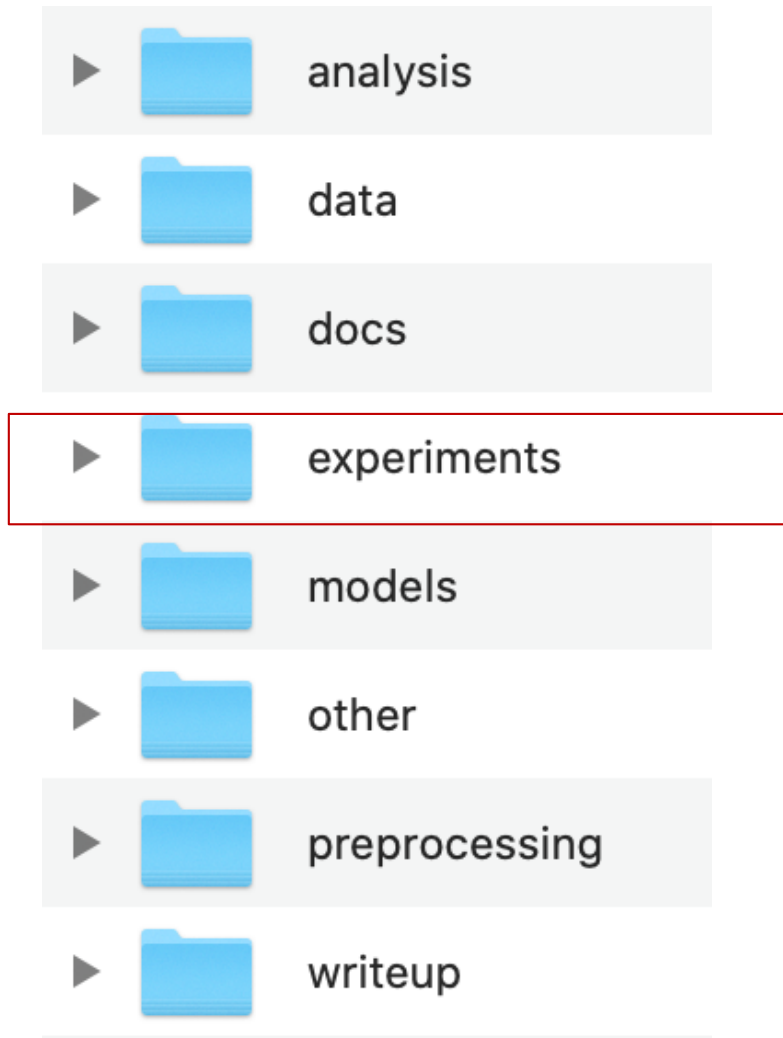
# newproject template



The screenshot shows the GitHub interface for the repository 'djnavarro / newproject'. At the top, the repository name is displayed with a 'Template' badge. To the right are buttons for 'Watch' (2), 'Star' (3), and 'Fork' (4). Below this is a navigation bar with links for 'Code', 'Issues' (0), 'Pull requests' (0), 'Actions', 'Projects' (0), 'Wiki', 'Security', and 'Insights'. The main content area states 'Template for a new research project' followed by the URL <https://djnavarro.github.io/newproject/>. Below this is a summary bar showing '9 commits', '1 branch', '0 packages', '0 releases', and '1 contributor'. At the bottom, there is a 'Branch: master' dropdown, a 'New pull request' button, and a row of buttons: 'Create new file', 'Upload files', 'Find file', 'Use this template' (highlighted in green), and 'Clone or download'.

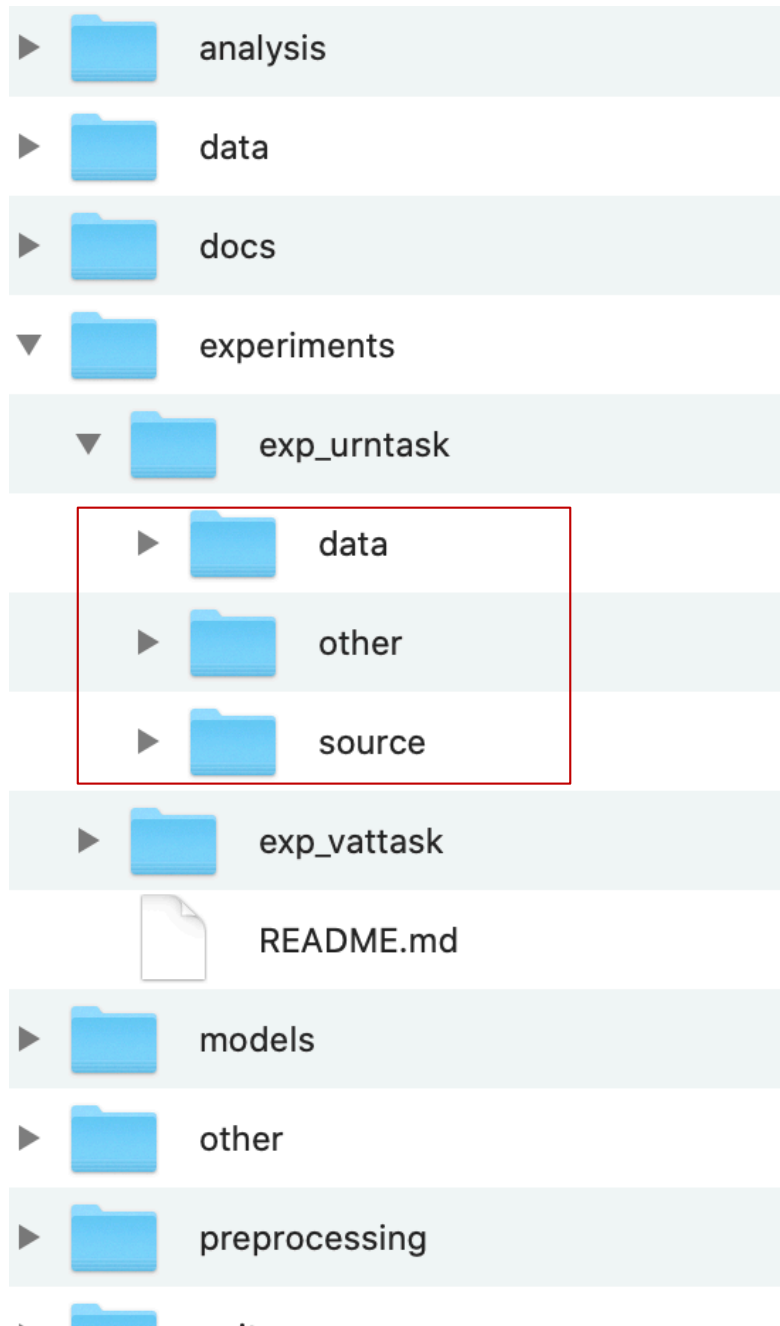
- Exercise:
  - Use this template to create a new repository called pitchmemory
  - Clone the repository inside your CHDS folder

# Folder structure



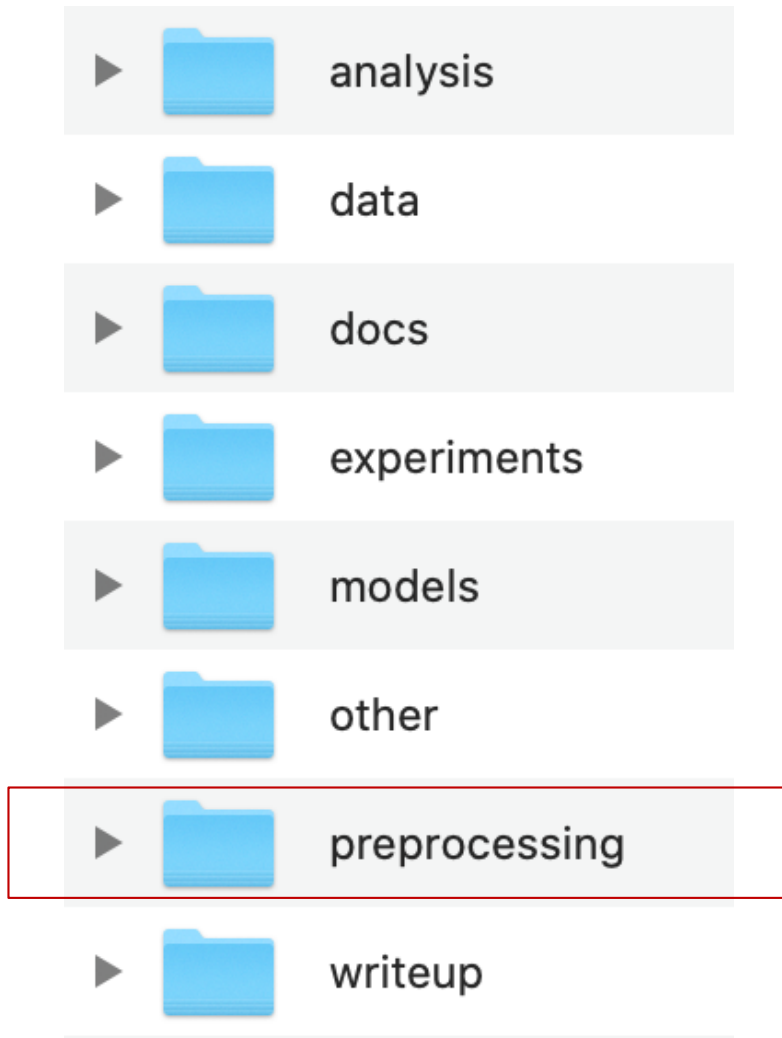
Contains a  
subfolder for each  
experiment

# Folder structure



Each experiment folder has subfolders for data and source code

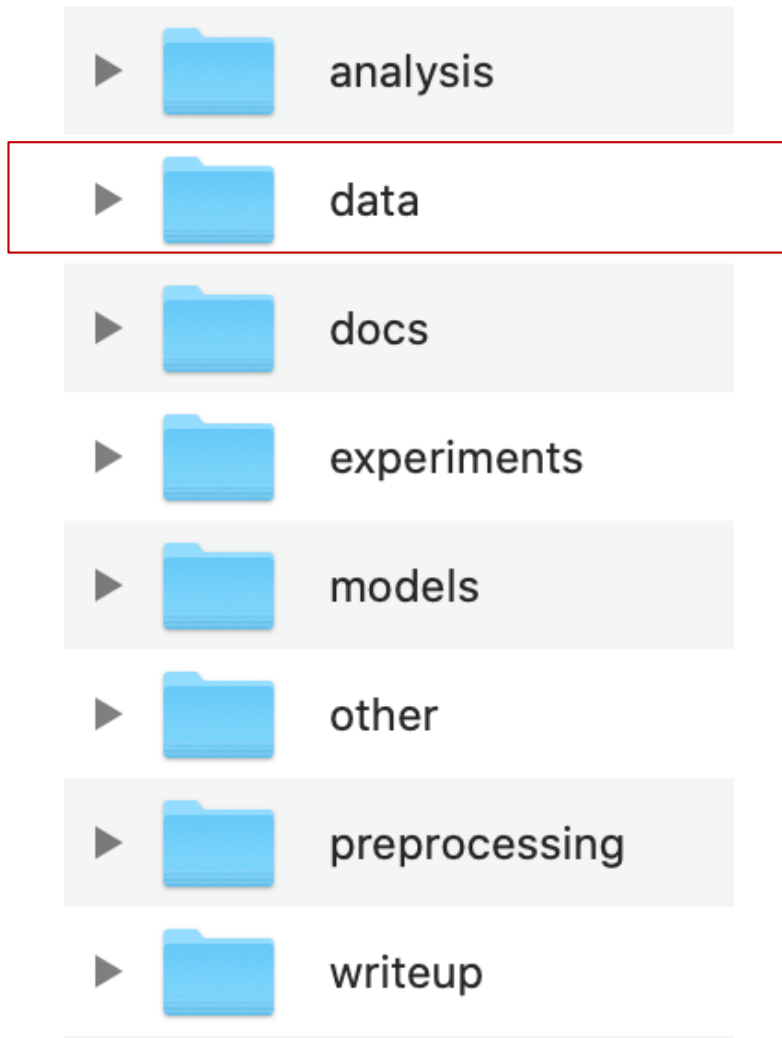
# Folder structure



Scripts for cleaning  
the data

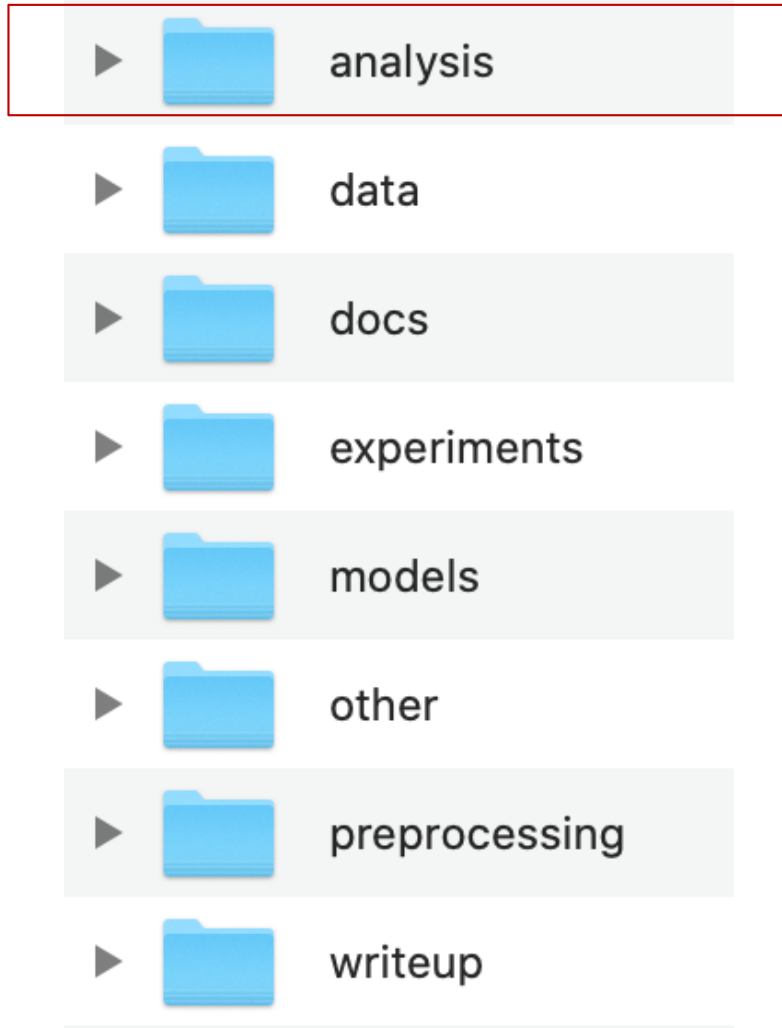


# Folder structure



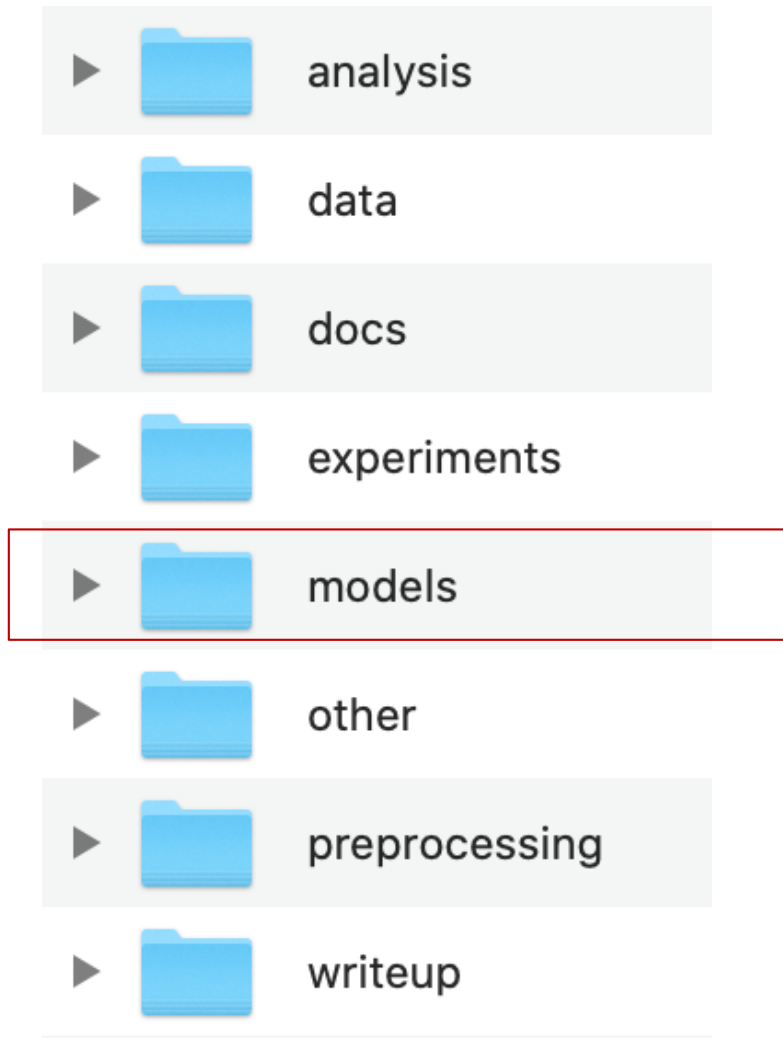
Cleaned data files

# Folder structure



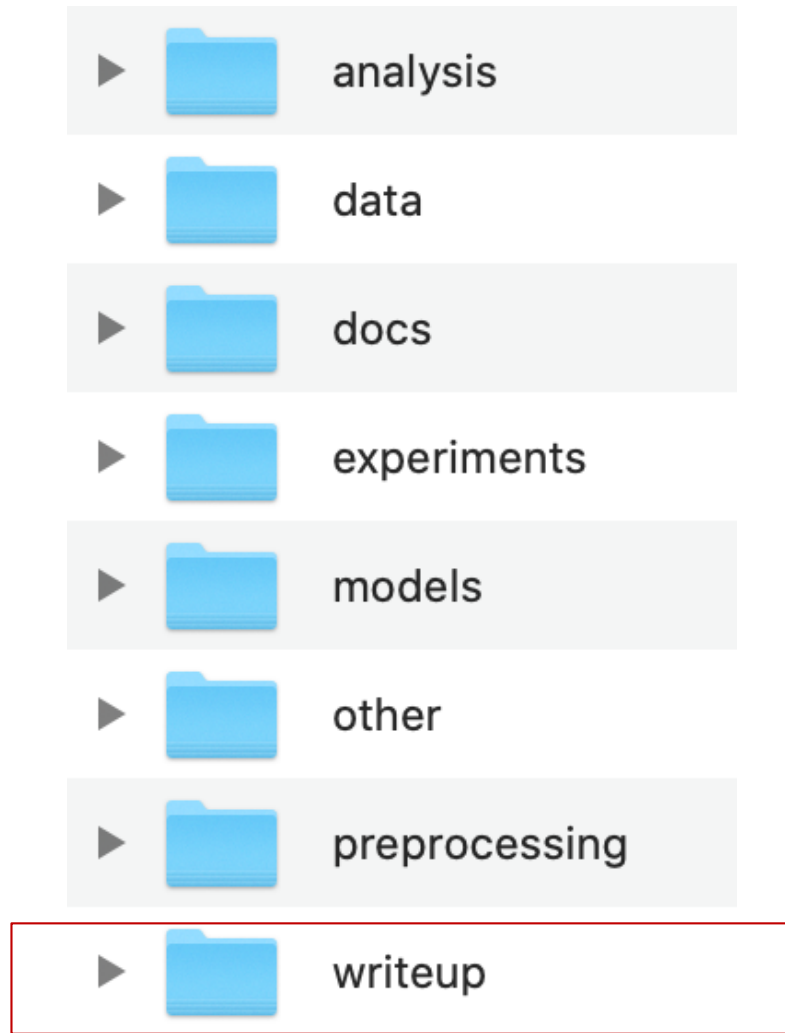
Scripts for exploratory  
analyses and statistical  
tests

# Folder structure



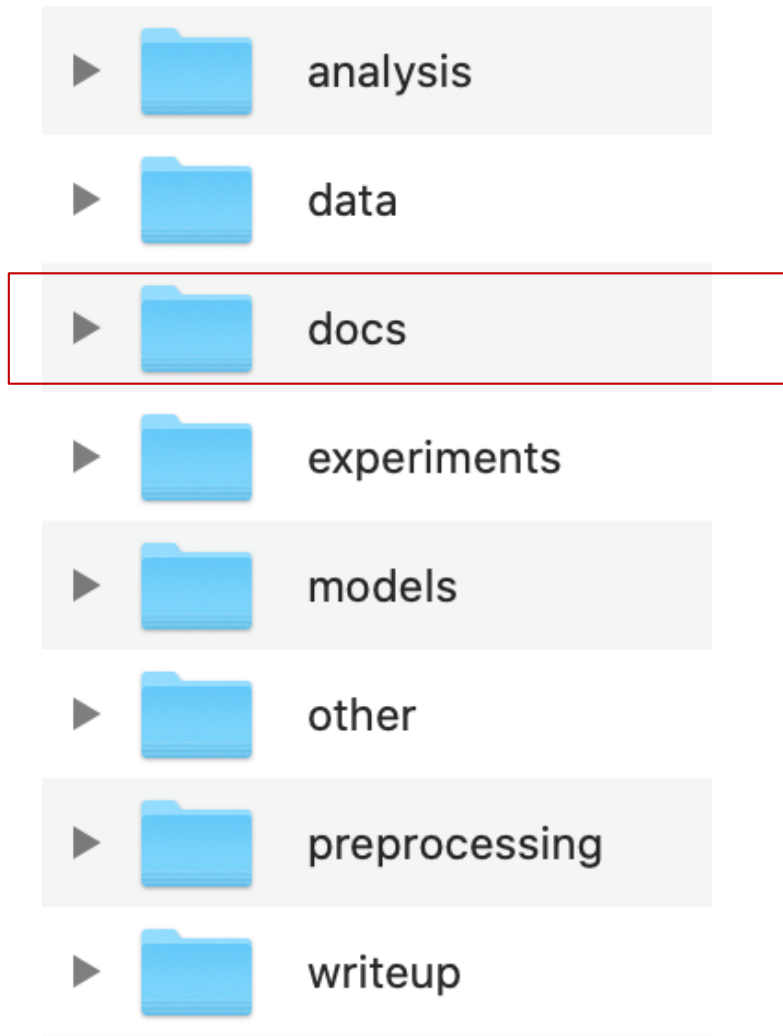
Models (if this is a modelling project)

# Folder structure



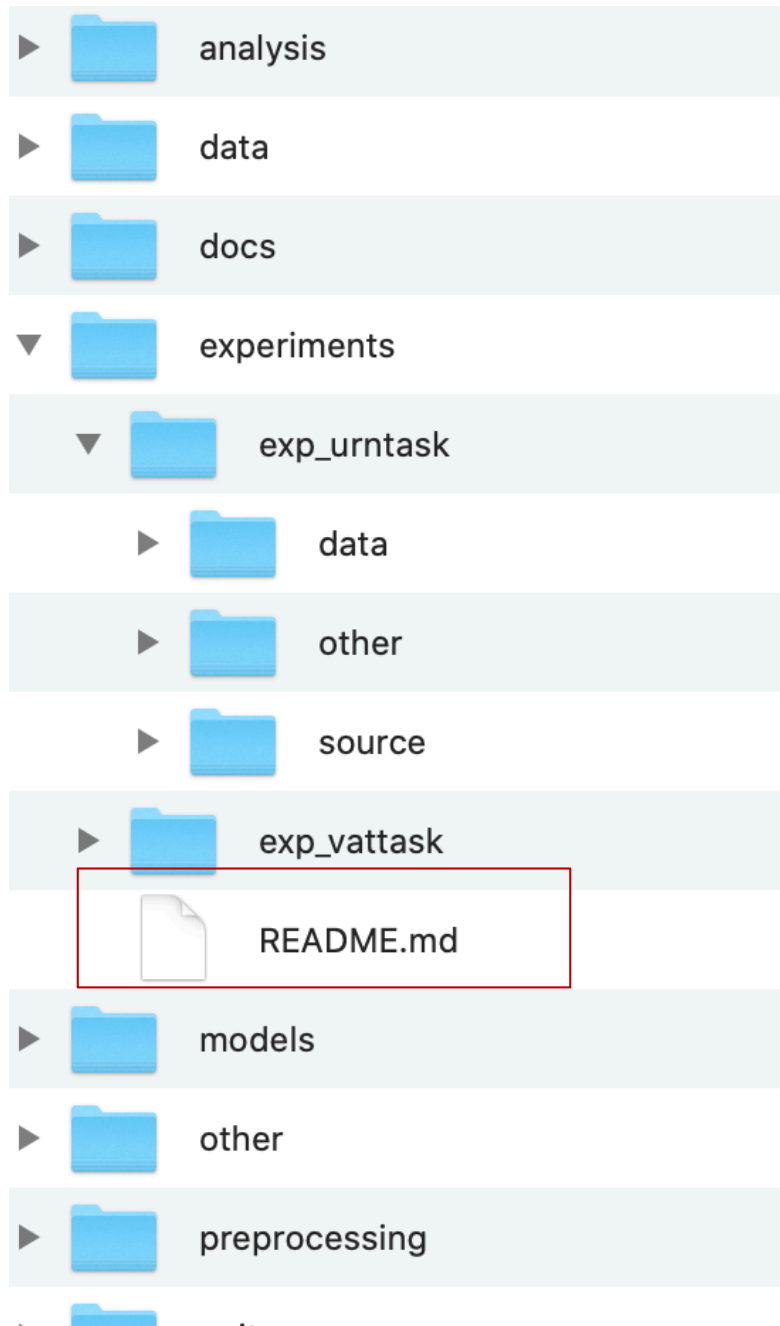
Manuscripts, talks, posters

# Folder structure



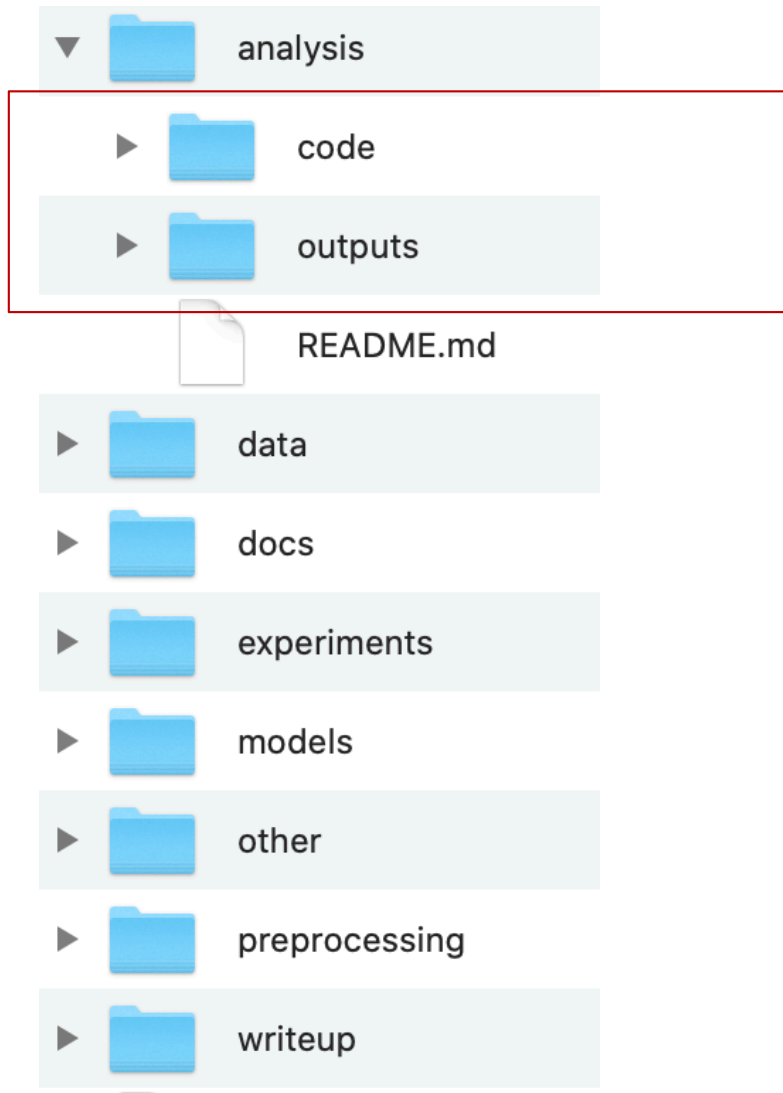
Used if you want to turn your project into a GitHub pages website

# Folder structure























A README in *every* folder

# Other ideas



Separate outputs (e.g. figures) from the code that generated them

# Filenames

- ▶  analysis
- ▶  data
- ▼  experiments
  - ▶  exp\_dice\_all
  - ▶  exp\_dice\_intention
  - ▶  exp\_dice\_story
  - ▶  exp\_story\_hospital
  - ▶  exp\_story\_long
  - ▶  exp\_story\_short
  - ▶  exp\_vignette\_generation
  - ▶  exp\_vignette\_intention
  - ▶  exp\_vignette\_probabilities
-  README.md
- ▶  other
- ▶  preprocessing
- ▶  .git
- ▶  .Rproj.user
-  conjunction.Rproj
-  README.md
-  README.Rmd

Files and folders should use a consistent naming scheme that is easy for both humans and machines to read



# RStudio projects

- Projects let you separate and easily switch between different things that you might be working on.

[DEMO]

- Exercise: create a project for your summerschool repository

# Benefits of Projects

- If you open a project by clicking on the .Rproj file, you'll automatically be in the right directory
- You can easily switch between projects using the menu in RStudio.
- Projects help you specify paths using the `here` package

*If the first line of your R script is*

```
setwd("C:\\Users\\jenny\\path\\that\\only\\I\\have")
```

*I will come into your office and SET YOUR COMPUTER ON FIRE 🔥.*

– Jenny Bryan

## A better approach

```
> library(here)
```

```
here() starts at /Users/ckemp/u/mygithub/chdss2019_content/samplingframes
```

```
> location <- here("data", "data_samplesize.csv")
```

```
> print(location)
```

```
[1] "/Users/ckemp/u/mygithub/chdss2019_content/samplingframes/data/data_samplesize.csv"
```

# R Markdown

- R Markdown lets you make documents that combine text and code.
- Demo/Exercise:
  - Switch to your summerschool project and create a new R Markdown document
  - Knit the document to create an .html file
  - Figure out the buttons at the top-right of a chunk

```
```{r pressure, echo=FALSE}  
plot(pressure)  
```
```



- Set `eval=FALSE` at the top of a chunk and see how that changes the .html produced by Knit

# What's R Markdown useful for?

- Preprocessing and analyzing data
- Making runnable research notebooks
- Making slides
- Writing journal papers (using `papa`)
- ...