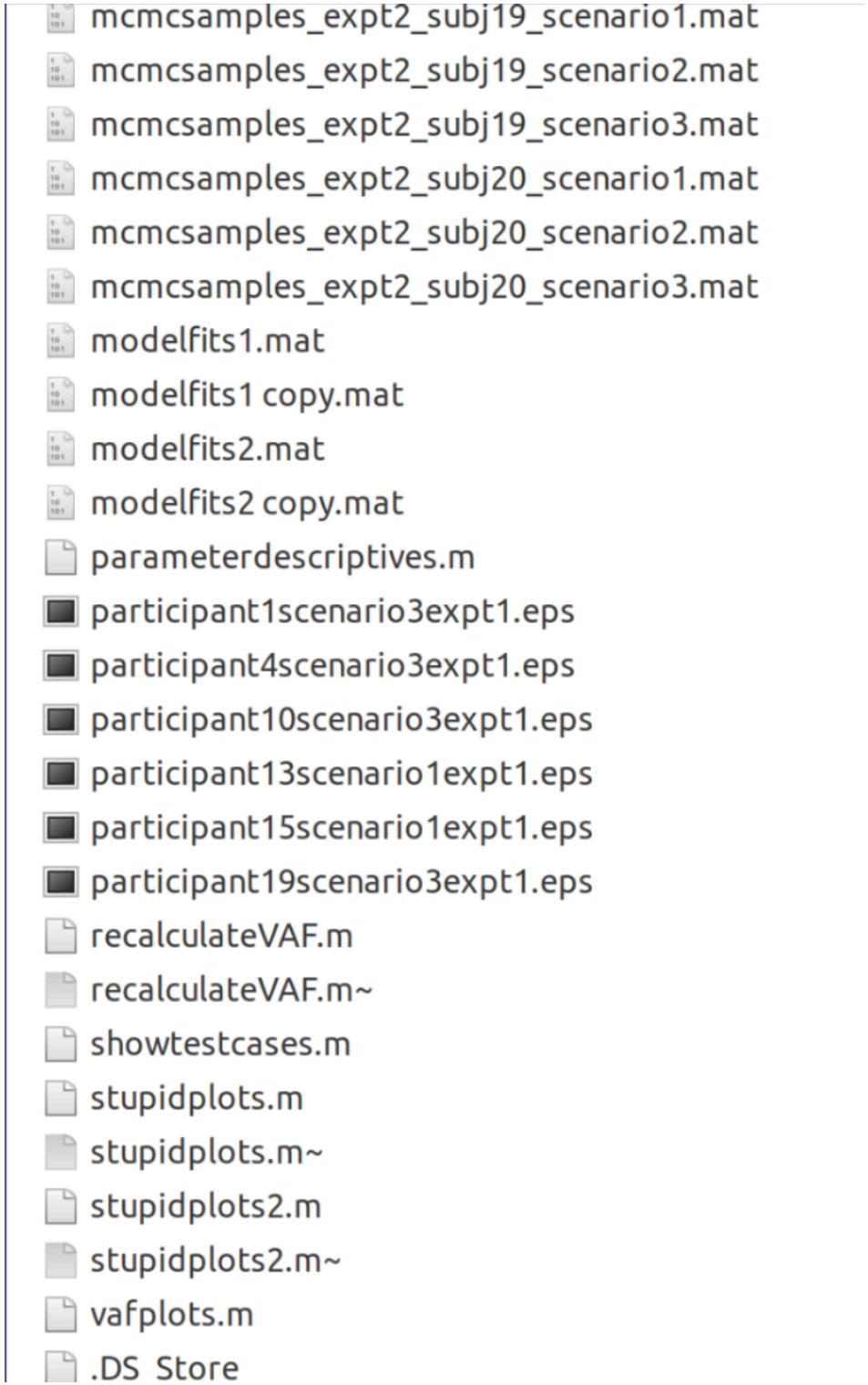
Project organisation

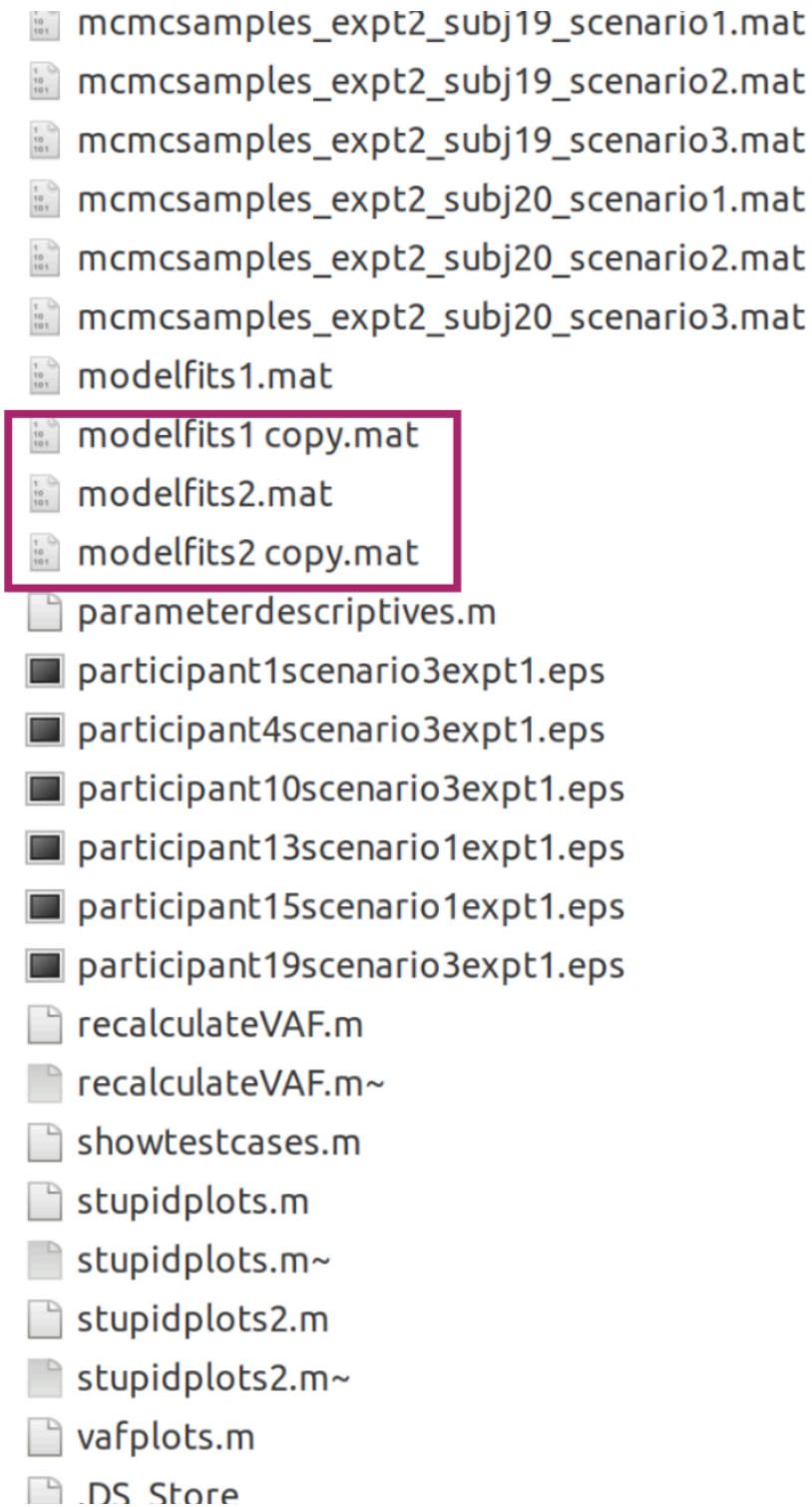
Complex Human Data Summer School

Why workflow matters

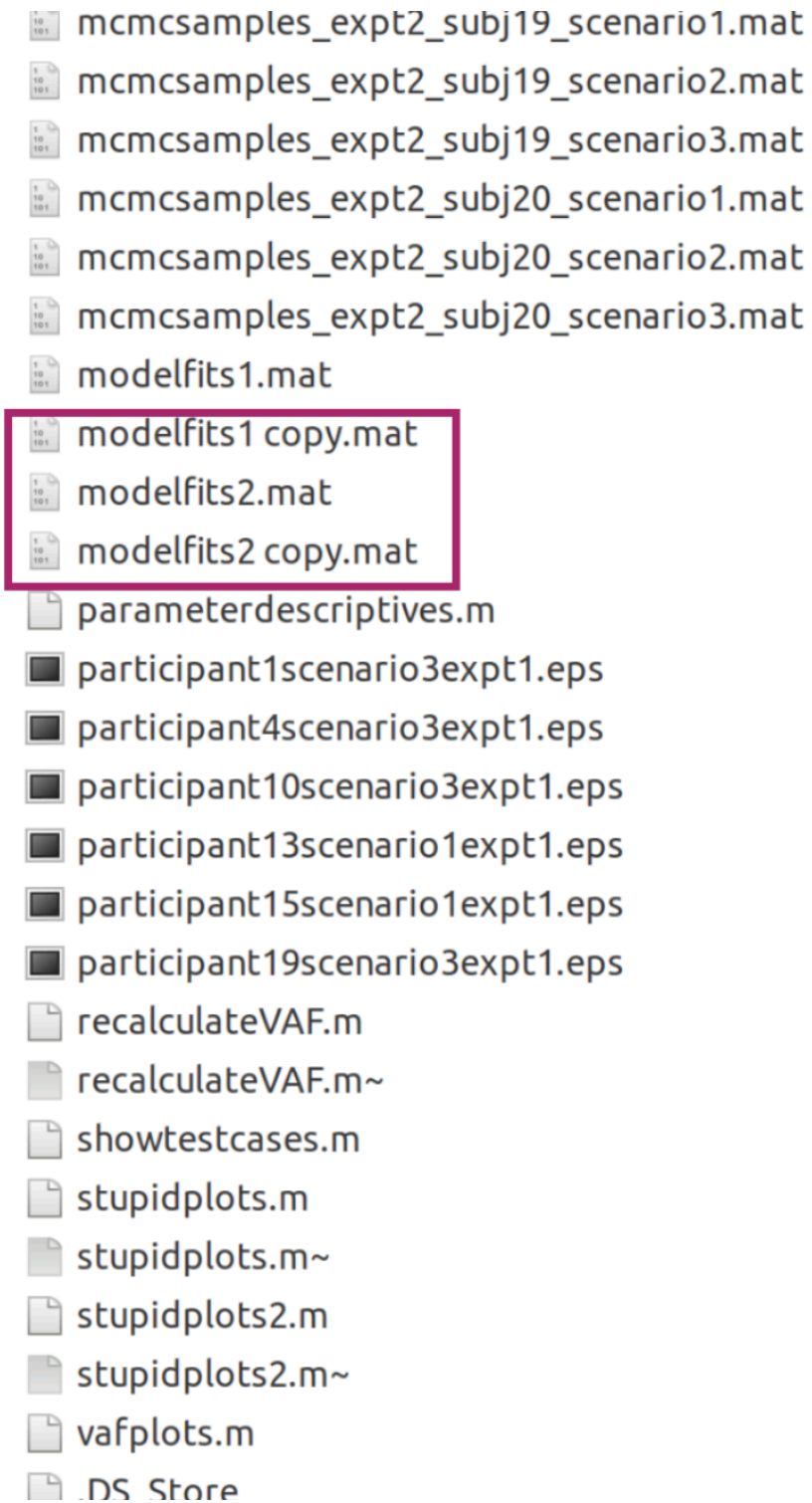
- A good workflow
  - saves time in the long run
  - helps you avoid errors
  - makes it easy for others to reproduce your work
- Folder organization is critical for all of these things



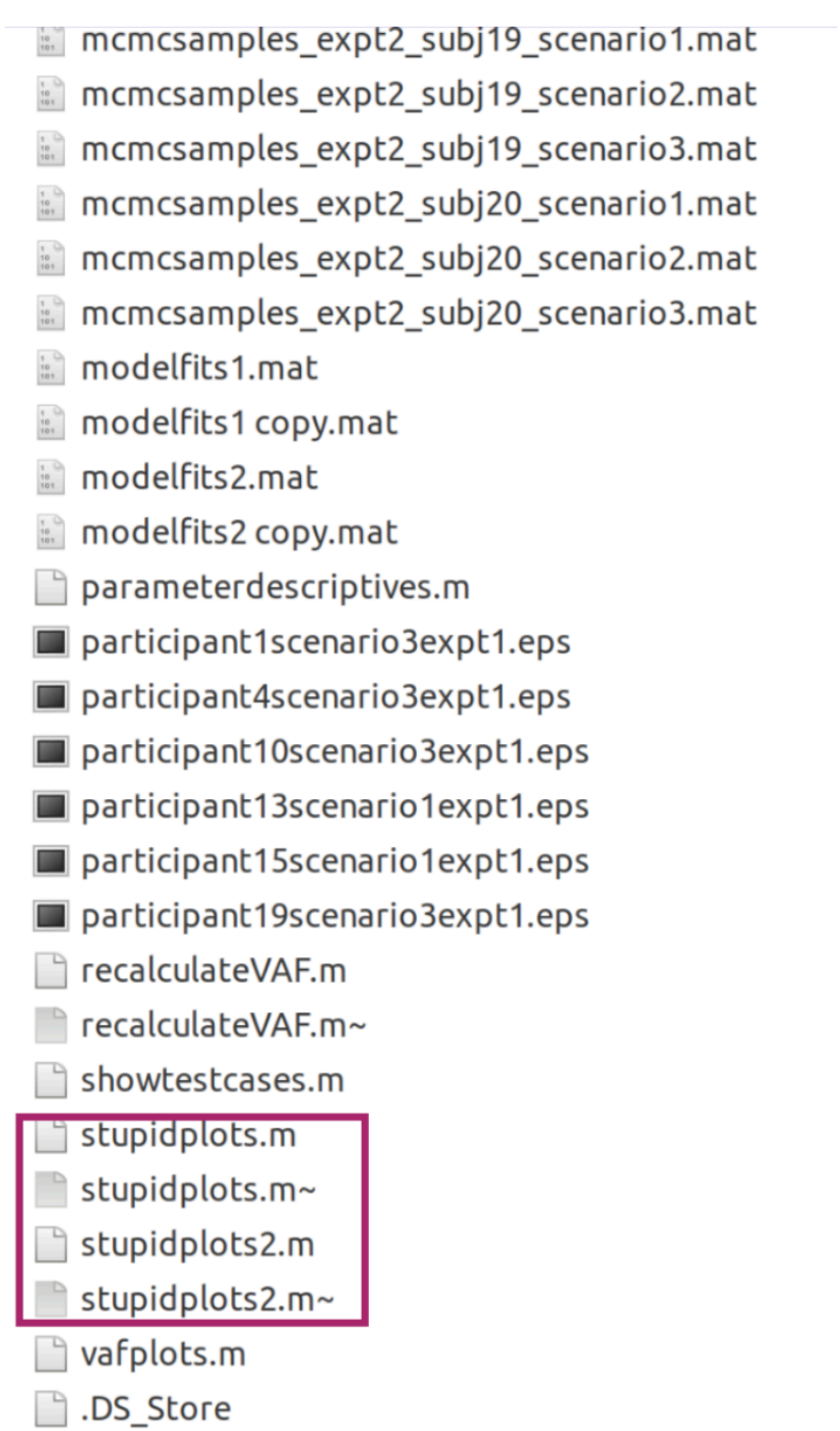
One of Danielle Navarro's folders from 2009



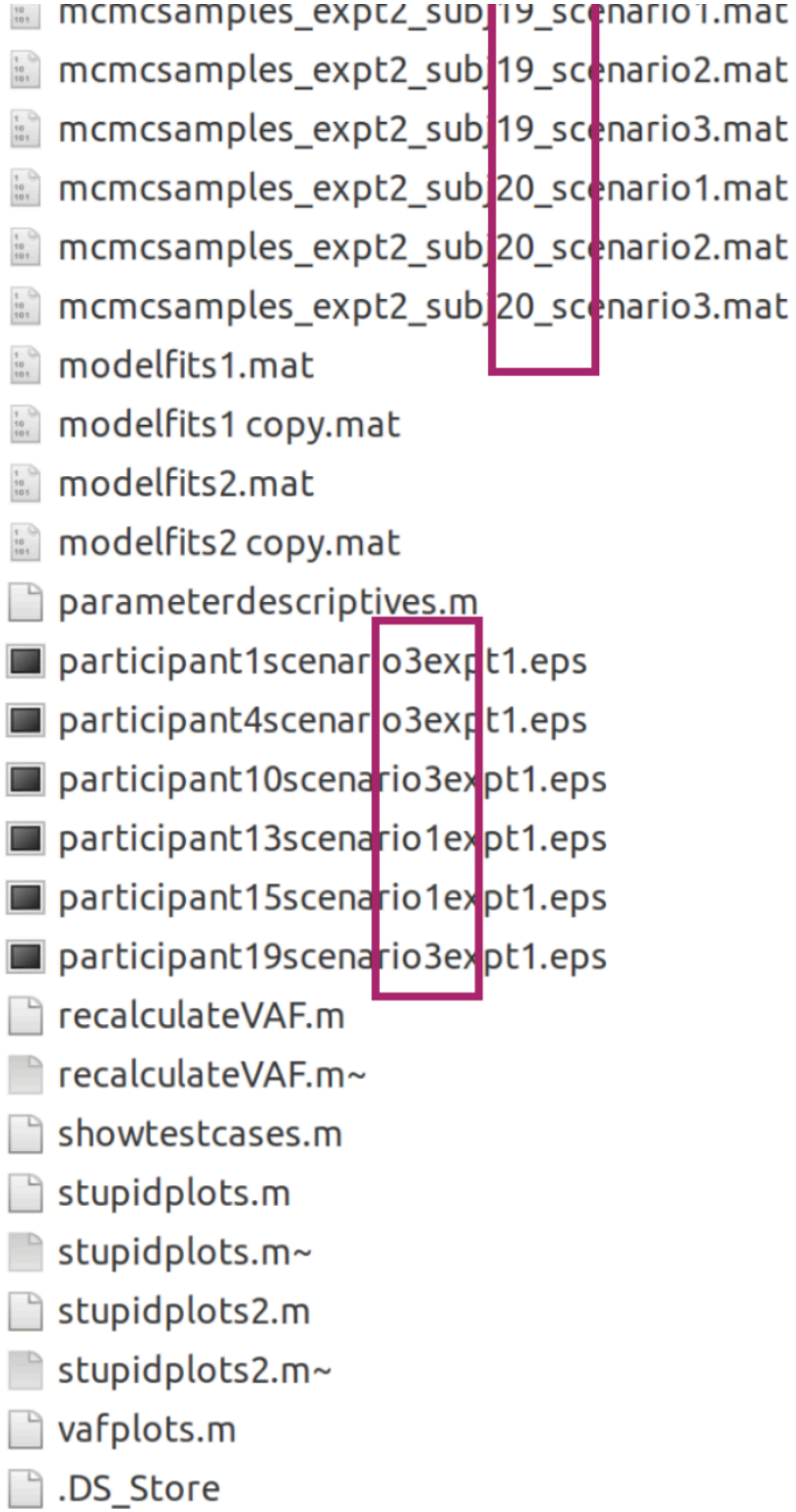
Multiple copies of files



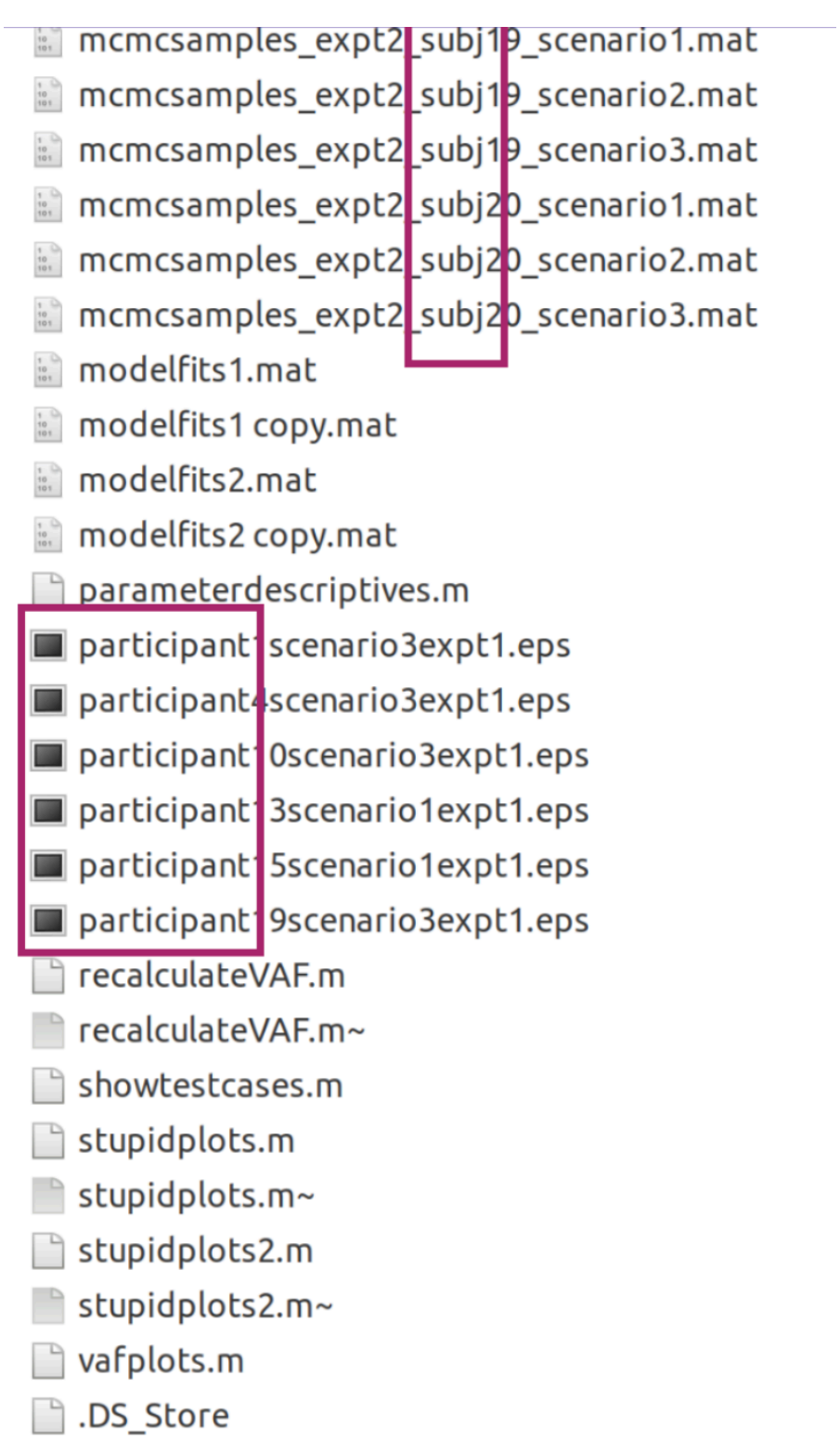
White space in filenames is a bad idea



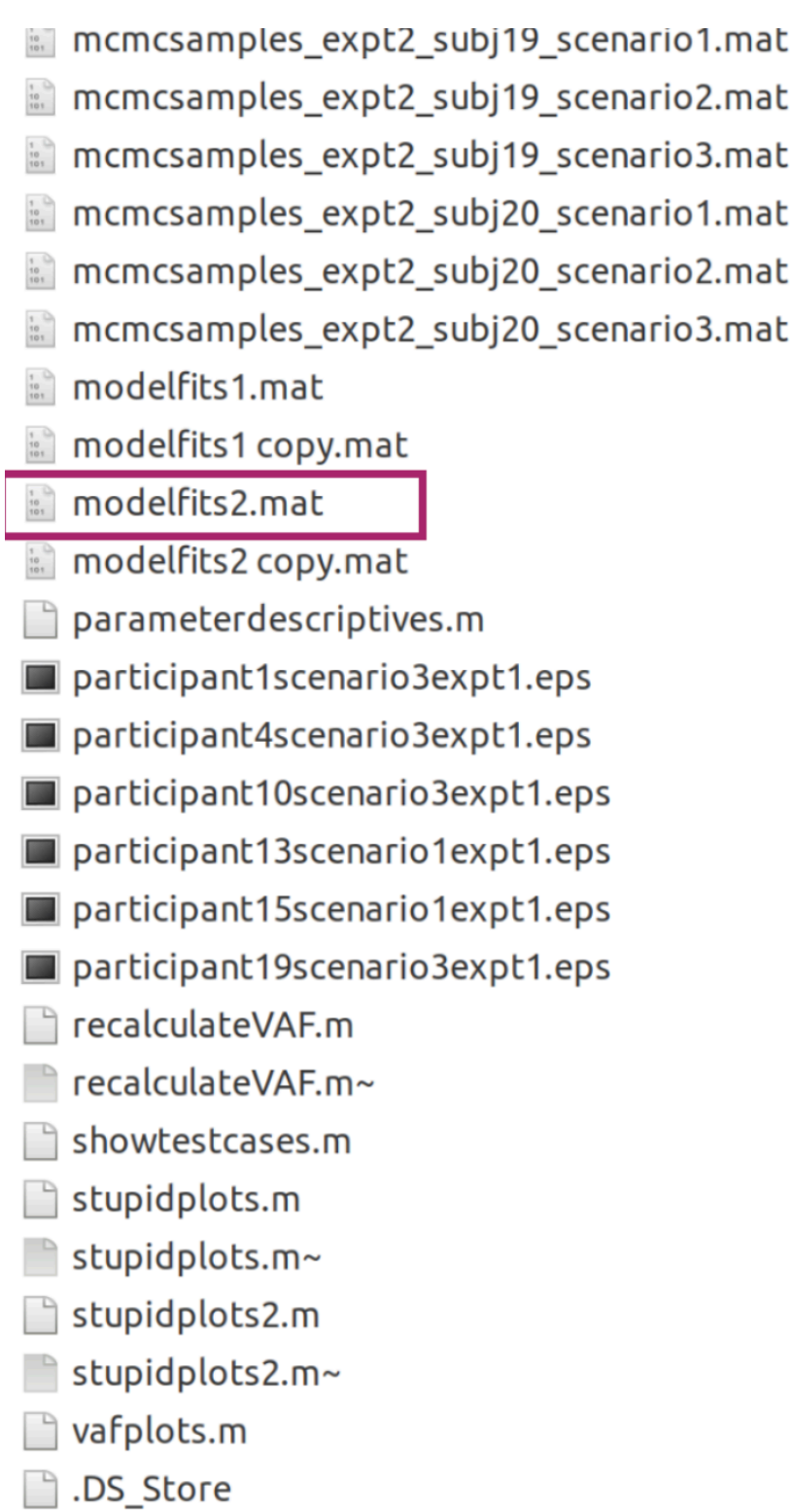
Slightly embarrassing filenames



Inconsistent separator characters



Inconsistent naming scheme



Filenames that aren't very helpful

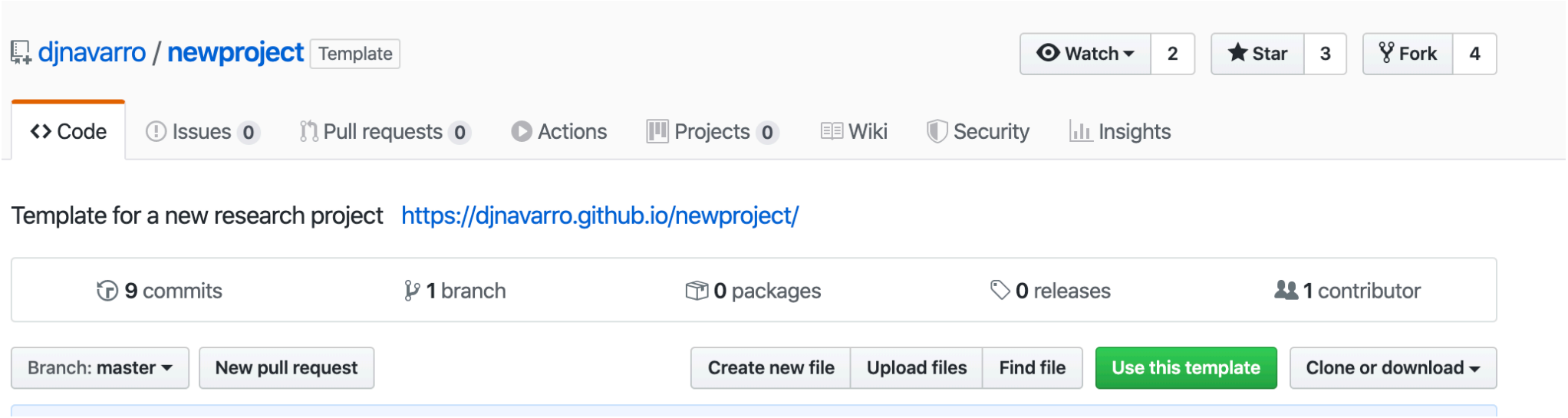
Not just Dani!

```
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
computeextensionspar.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

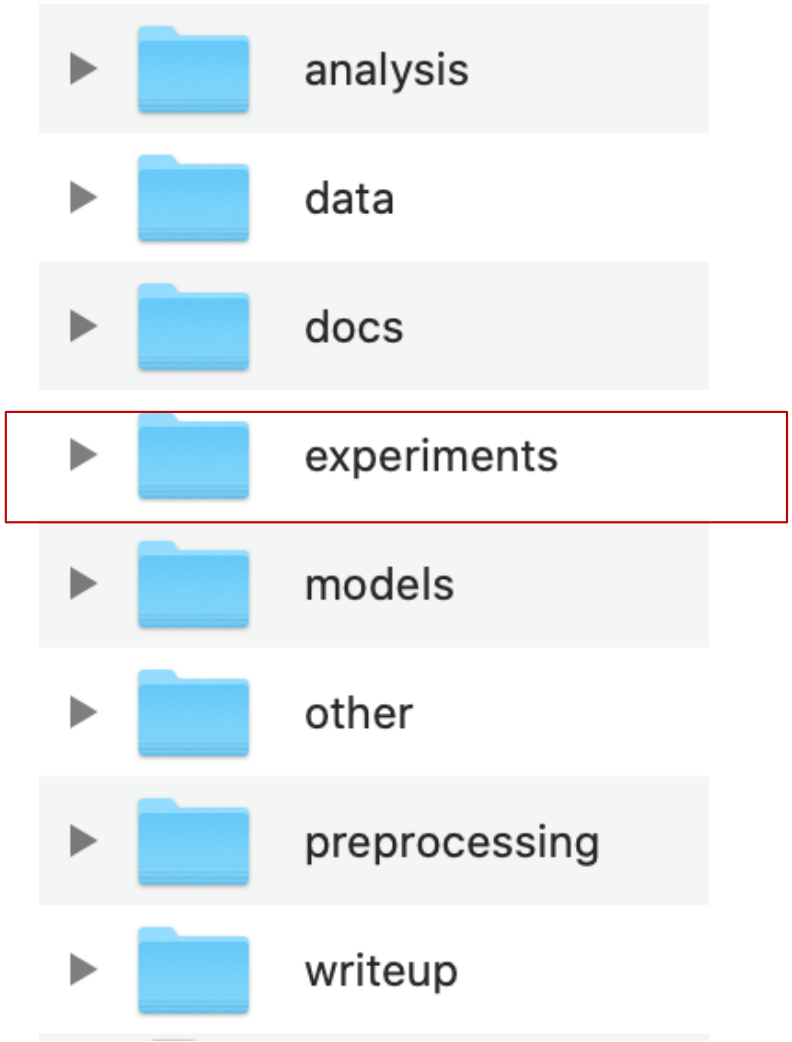
– Jenny Bryan

newproject template



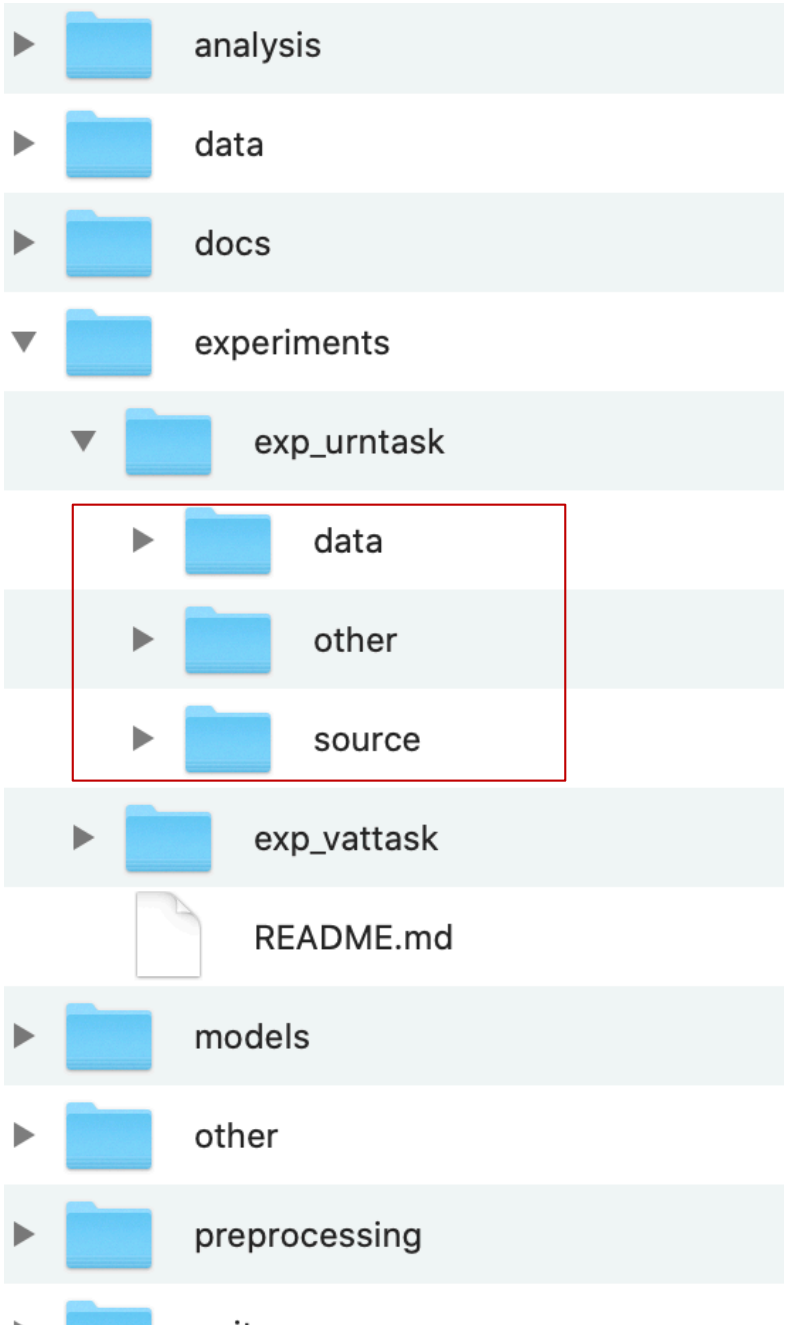
- Exercise:
  - Use this template to create a new repository called pitchmemory
  - Clone the repository inside your CHDS folder

Folder structure



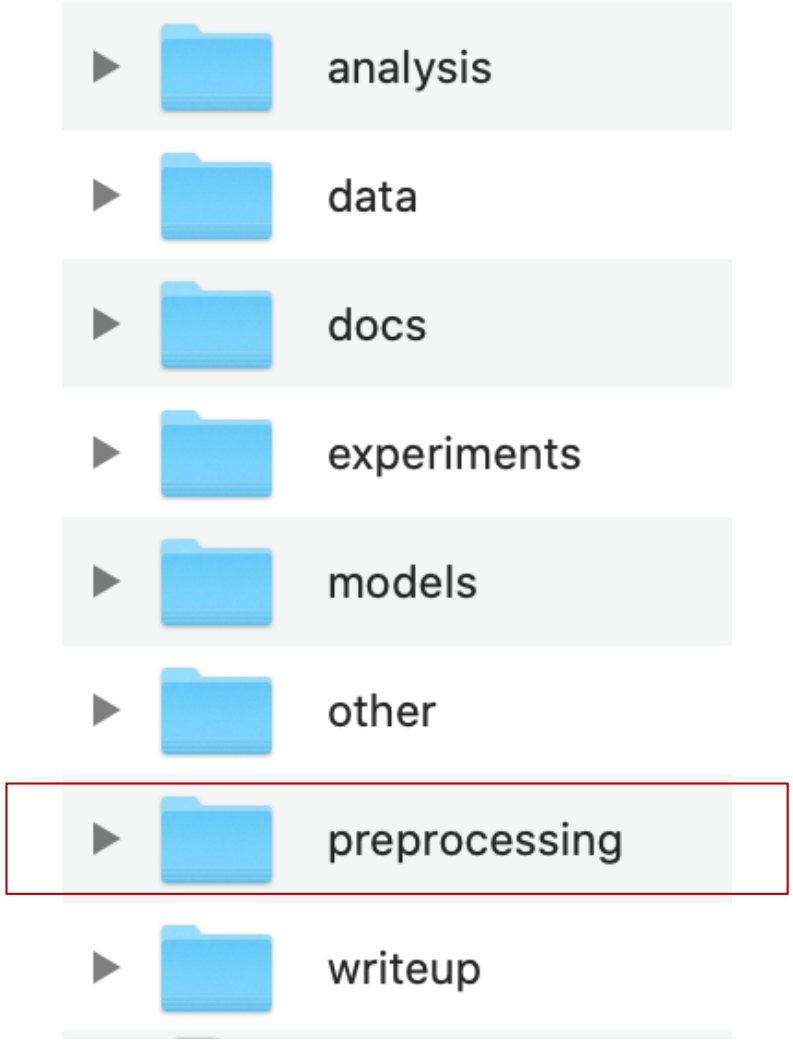
Contains a subfolder for each experiment

Folder structure



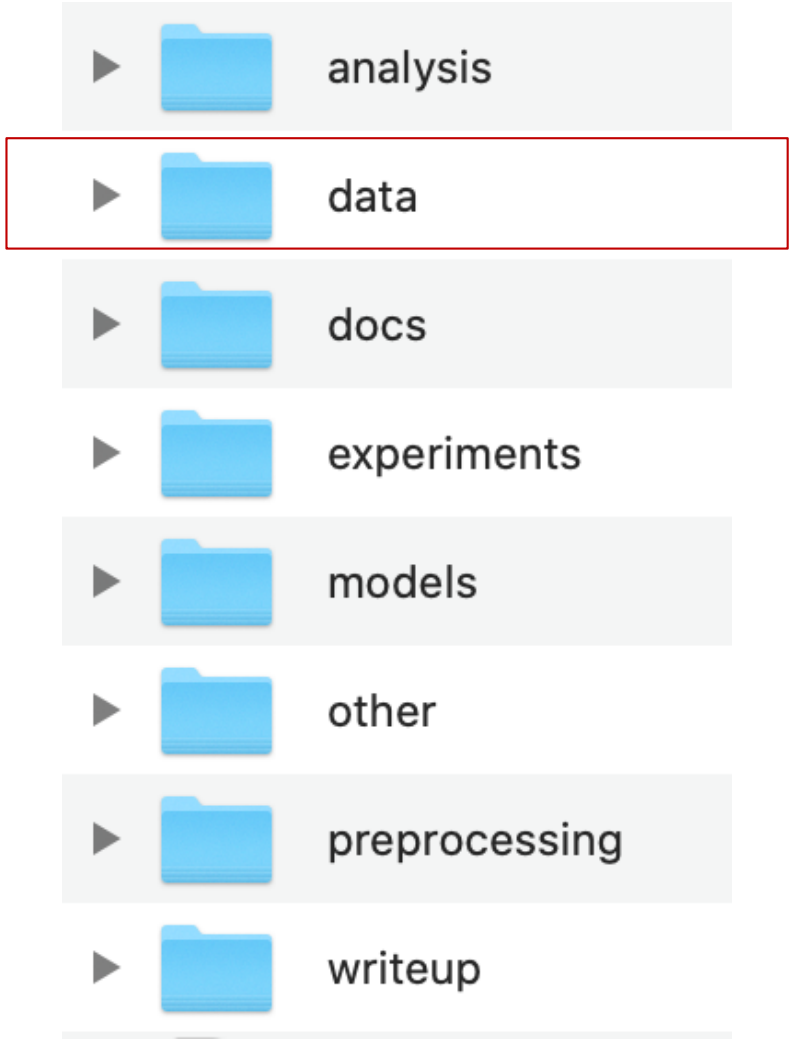
Each experiment folder has subfolders for data and source code

Folder structure



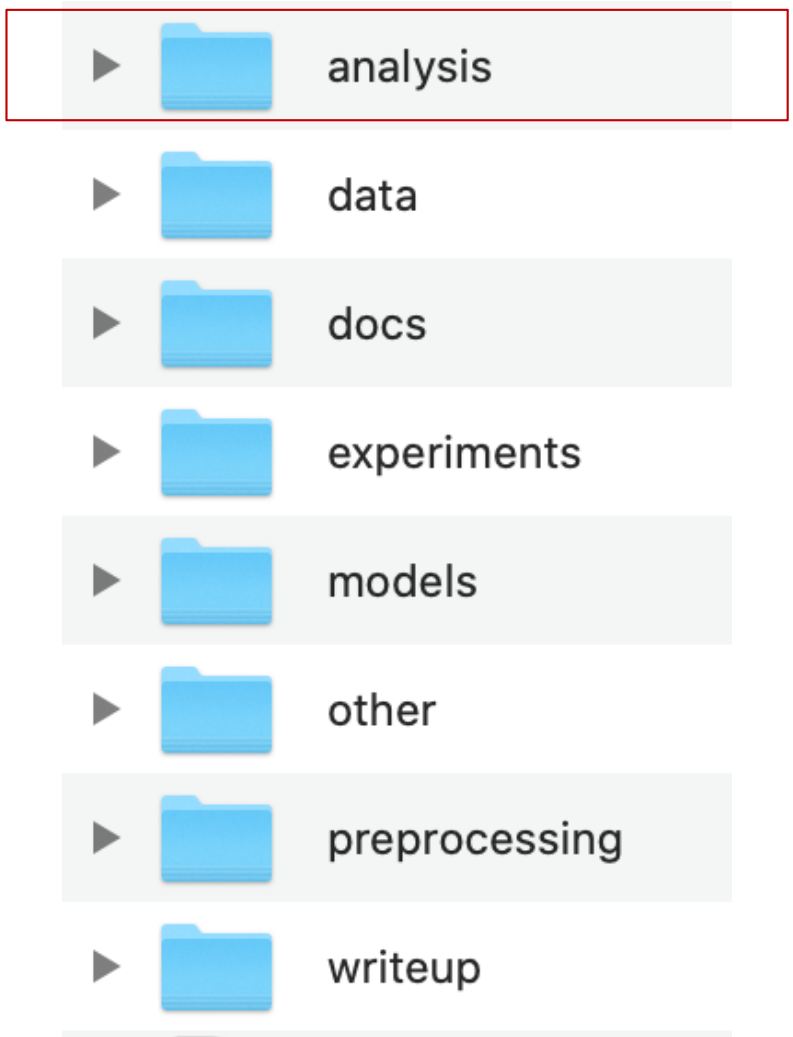
Scripts for cleaning the data

Folder structure



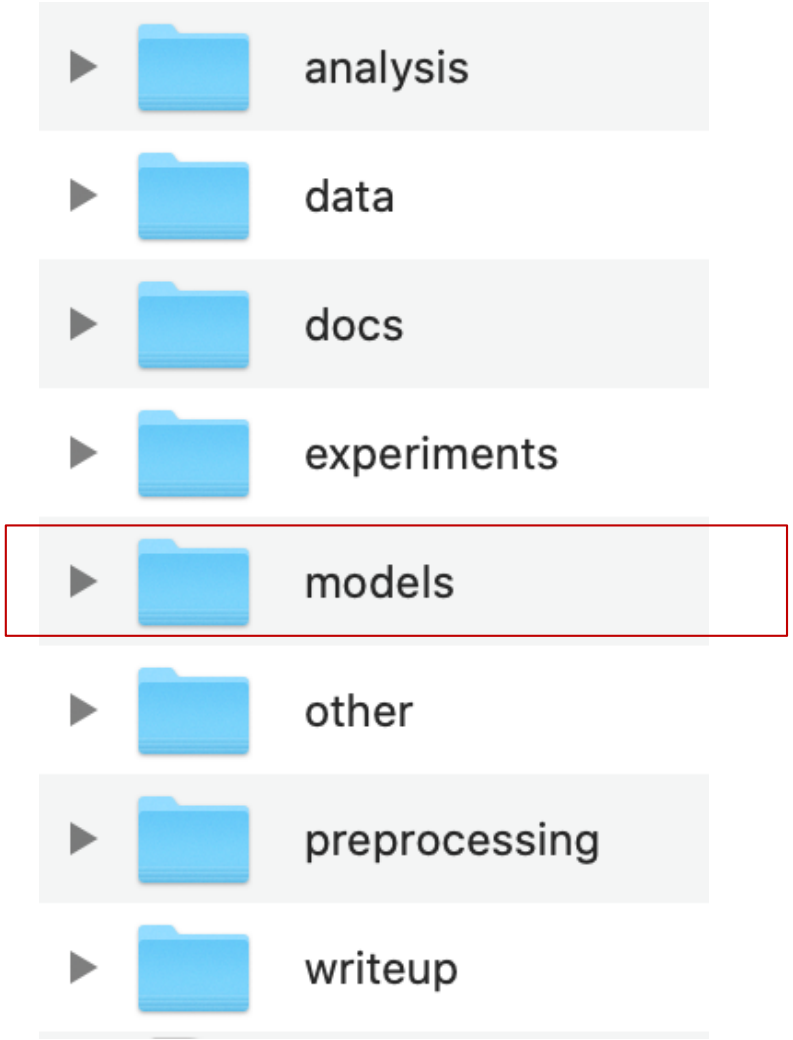
Cleaned data files

Folder structure



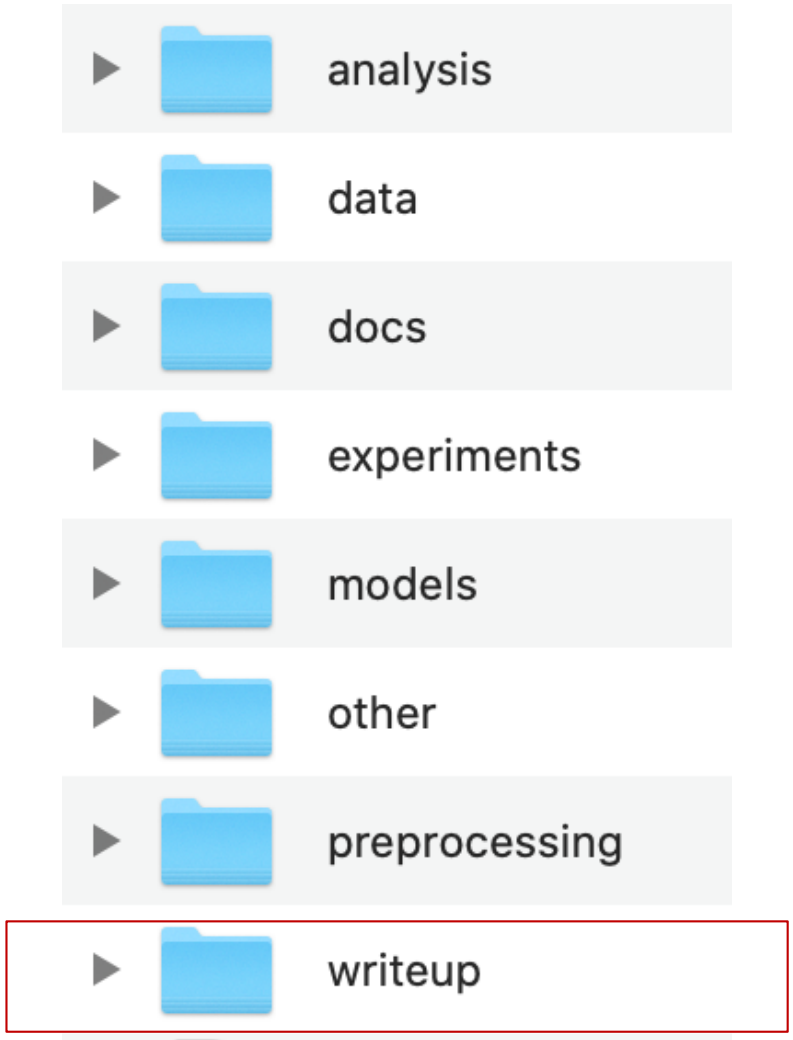
Scripts for exploratory analyses and statistical tests

Folder structure



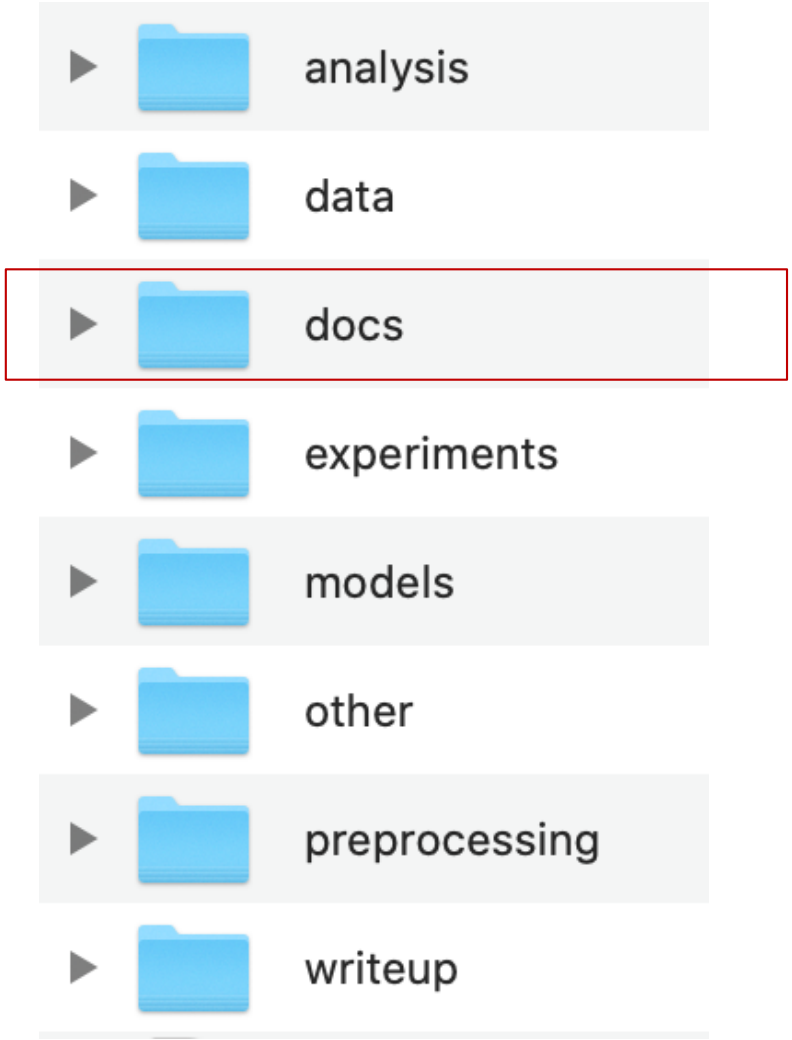
Models (if this is a modelling project)

Folder structure



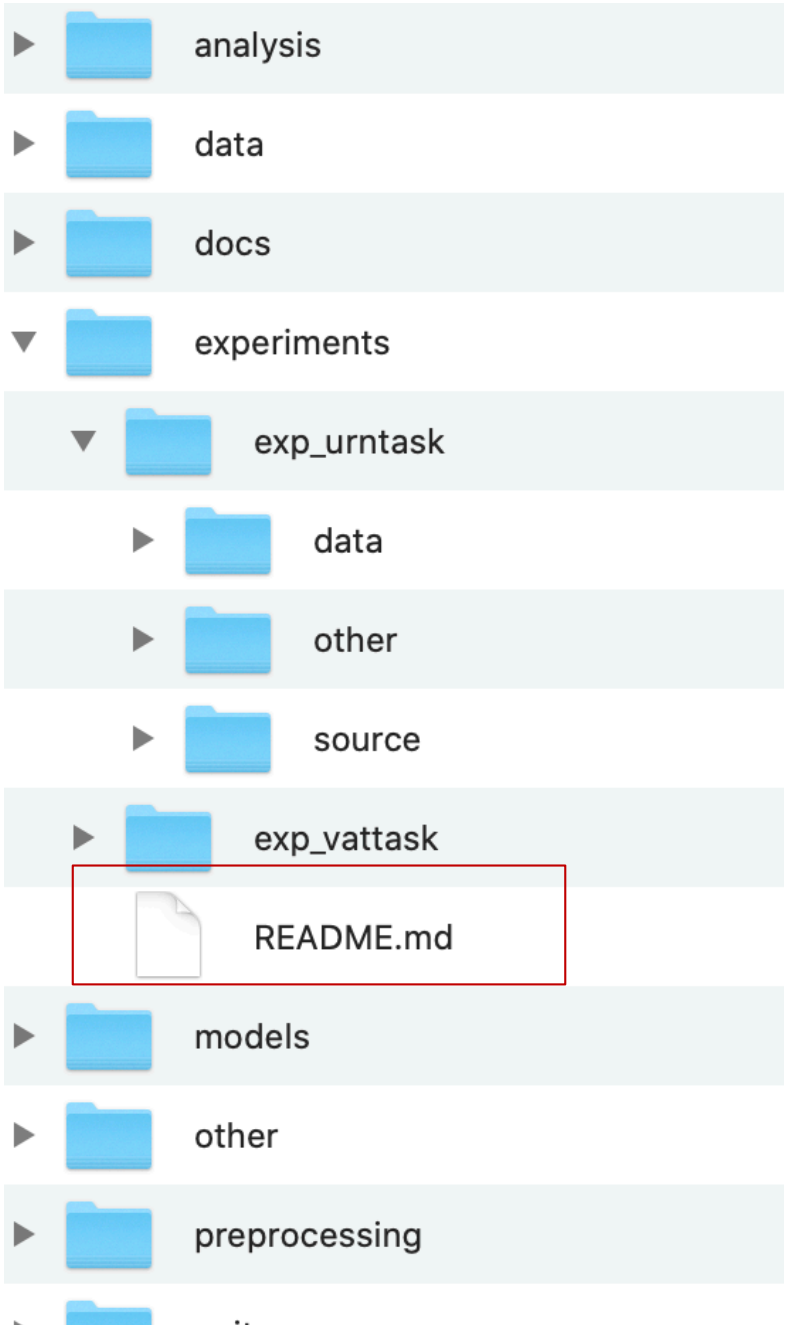
Manuscripts, talks, posters

Folder structure



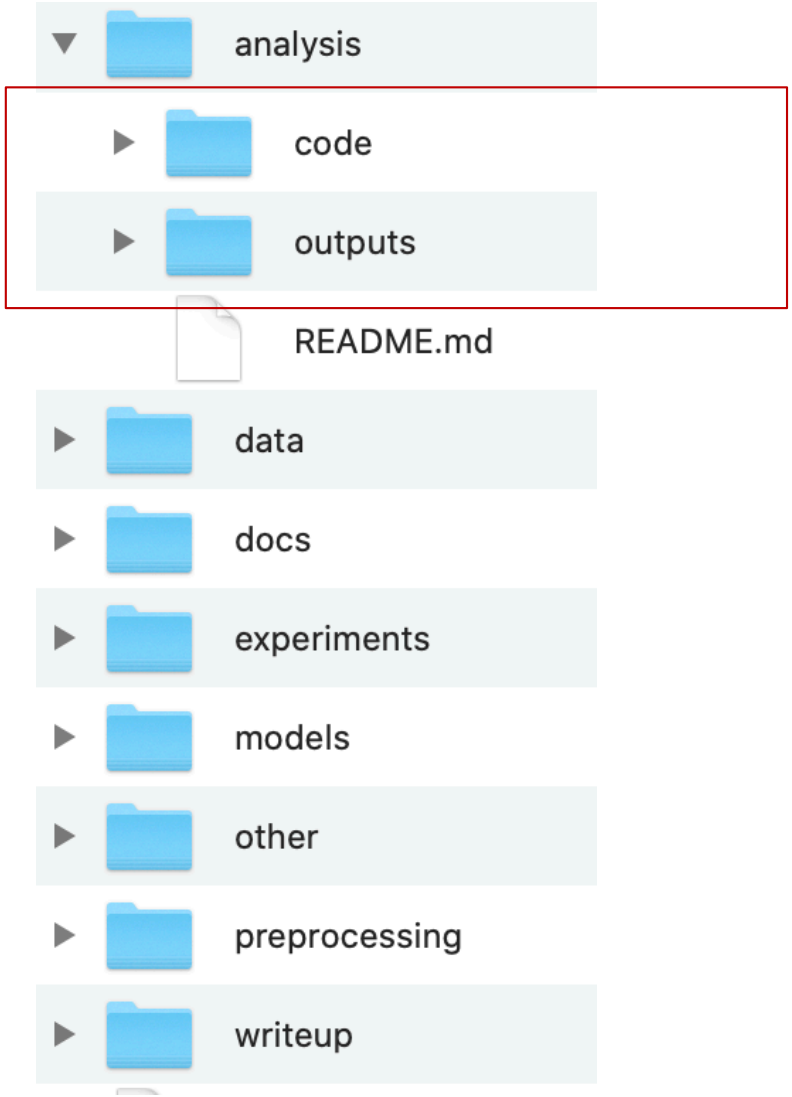
Used if you want to turn your project into a GitHub pages website

Folder structure



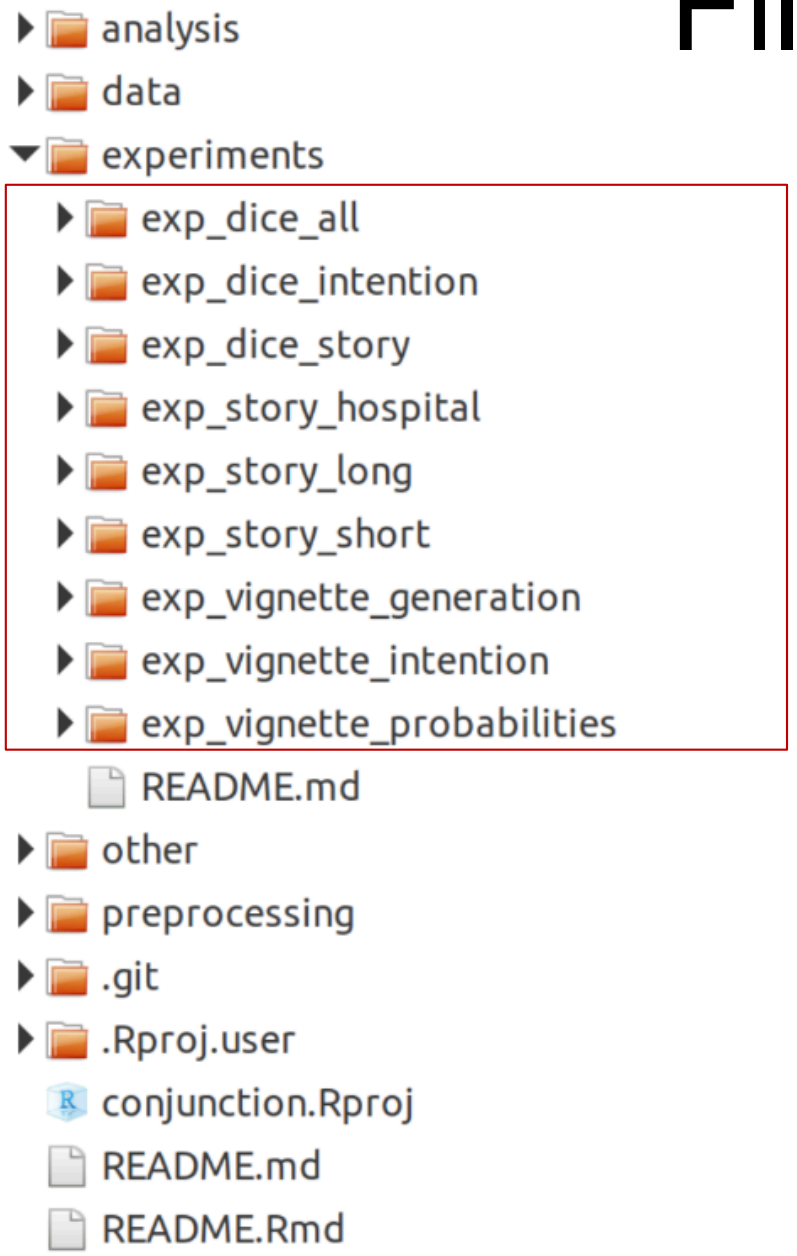
A README in *every* folder

Other ideas



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

Filenames



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

  - 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 `papaja`)
- …