OVERVIEW OF THE WORKSHOP AND LOGISTICS

Instructor: Professor Kendrick Kay, PhD (http://cvnlab.net)
Teaching assistant: Emily Allen, PhD

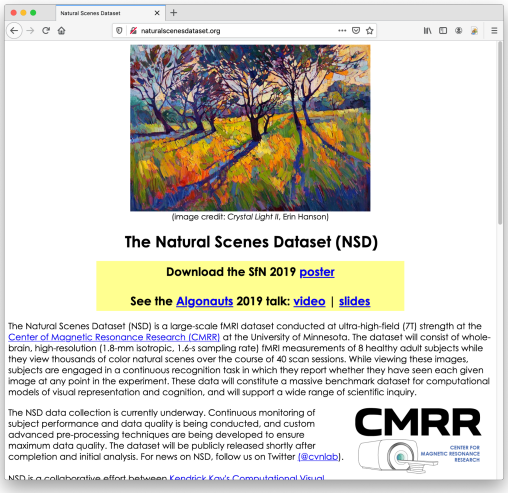
Sponsor: Professor Bas Rokers, PhD
Teaching assistant: Nate Miller

1

Natural Scenes Dataset

A very large visual fMRI dataset collected at 7T on 8 subjects
http://naturalscenesdataset.org

This workshop is an **intensive introduction to cognitive neuroimaging from a data science perspective**.



2

Overview of this workshop

- Heavy on programming and learning how to 'do it yourself'. (As opposed to relying on GUIs.)
- Knowing how things work at many levels:
  - Conceptual level
  - Algorithmic/statistical level
  - Coding level
- KISS (Keep It Simple Stupid) principle
- Multiplicity principle
  - Often, there is not 'One Way' to carry out neuroimaging/data analysis/coding, but it depends on what you are trying to do.
- Will use cvnlab's tools... this is out of necessity, but hopefully you will learn general principles

3

Goal is to learn:

- Broad sampling of styles/frameworks/analysis approaches
- How to deal with very large structured data
- Good coding style, automated and large-scale analysis
- Knowing how to mix-and-match different tools, learning how to put them together

4

What this workshop is NOT:

- The IT stuff has been done already. Dealing with IT, software management, etc. is not the focus.
- Not about preprocessing of neuroimaging data, as that has already been done.
- This is not intended to be learning any specific tool/software/package/analysis (aside from perhaps ITK-SNAP)
- Not really *specifically* about programming, but it will come along for the ride...

…Though, if interested, we can certainly have in-depth side discussions on these topics.

5

Overview of example scripts

- Designed to be as short as possible (with function abstraction) while still being informative and instructive. Prioritize clarity and understanding.
- Commented extensively
- Examples of good coding (modular, flexible, general when it makes sense to be)

6

Overview of example scripts

http://github.com/kendrickkay/nsdexamples

From http://github.com/kendrickkay/nsdexamples/

Example 1: Basic exploration of the NSD data files

Example 2: Loading data into MATLAB

Example 3: Understand mapping between spaces

Example 4: Automated surface visualization

Example 5: Inspect data at scale

Example 6: Basic loading and inspection of NSD betas

Example 7: A simple contrast-based analysis of the NSD betas

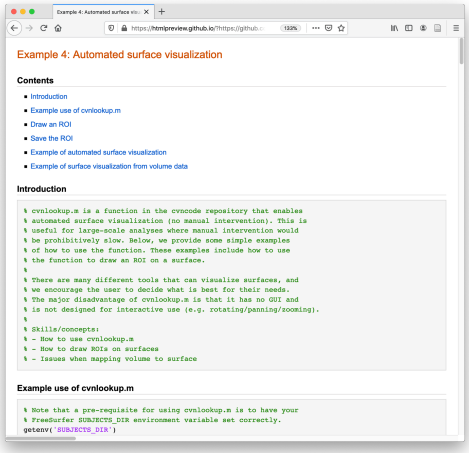
Example 8: A simple example of MVPA on the NSD betas

Example 9: Some example analyses of the behavioral data

Example 10: Building encoding models

Example 11: Representational similarity analysis

Example 12: Resting-state functional connectivity



7

Software/computational setup

- The following materials have been prepared:
  - **Data:** the NSD data (already pre-processed)
  - **Software:** a variety of neuroimaging, programming, and software tools
  - **Code:** Toolboxes and a series of 12 example scripts that illustrate a variety of analyses of the NSD data

8

Software/computational setup

- **Data:**
  - Data is stored on shared file server
  - Can mount using **smb://rcsfileshare.abudhabi.nyu.edu/Vision**
  - Hopefully will be fast enough to enable rapid data access
  - Directory on local Mac is /Volumes/Vision/nsd/

9

Software/computational setup

- **Software:**

OS:
- Mac OS 10.14.6 (Mojave)
- XQuartz (2.7.11)
- switch to tcsh, set up .tcshrc

Neuroimaging:
- FreeSurfer [freesurfer-Darwin-OSX-stable-pub-v6.0.0-2beb96c]
- ITK-SNAP [3.8.0]
- SPM [spm12]
- FSL [6.0.1, in /usr/local/fsl]
- AFNI [AFNI_19.2.19 (Claudius), Aug 26 2019] [commented out apsearch in .bashrc] [afni from Terminal]
- MRIcroGL [1.9.0]

Programming:
- Matlab [R2019a] [also setup alias in .tcshrc so we can run in terminal] [setup a startup.m that sets up path]
- Psychtoolbox
- Python

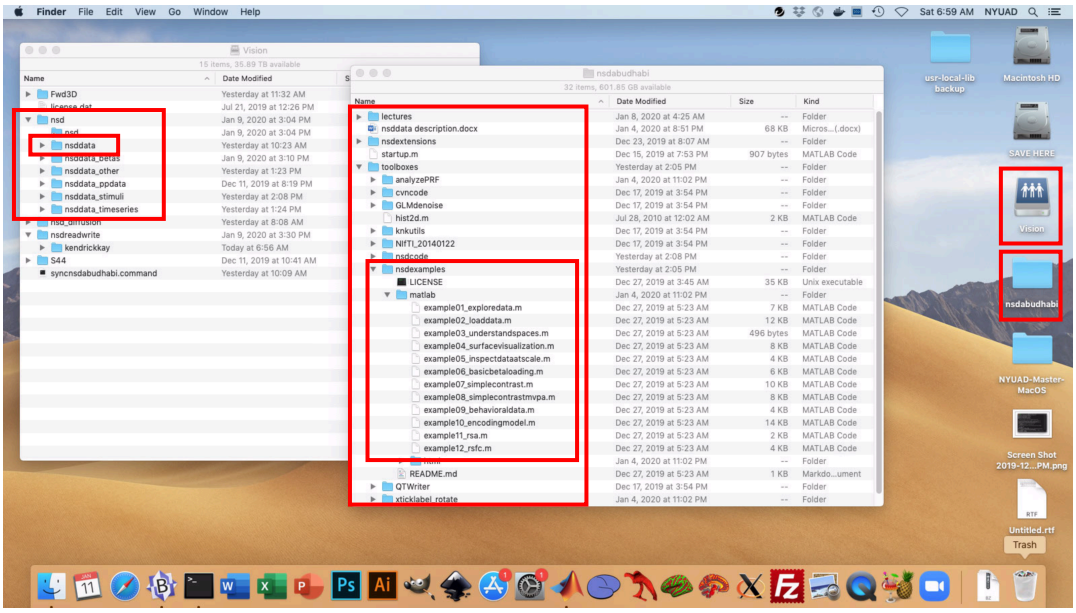
Utilities:
- FileZilla [3.45.1]
- Zoom
- Safari
- Terminal

General-purpose:
- BBEdit [12.6.7]
- BBEdit Matlab Language module
- Photoshop
- Illustrator
- GIMP
- Inkscape
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

Image/video-related:
- HandBrake
- ffmpeg (brew install ffmpeg)
- ImageMagick (brew install imagemagick)
- Ghostscript (brew install ghostscript)

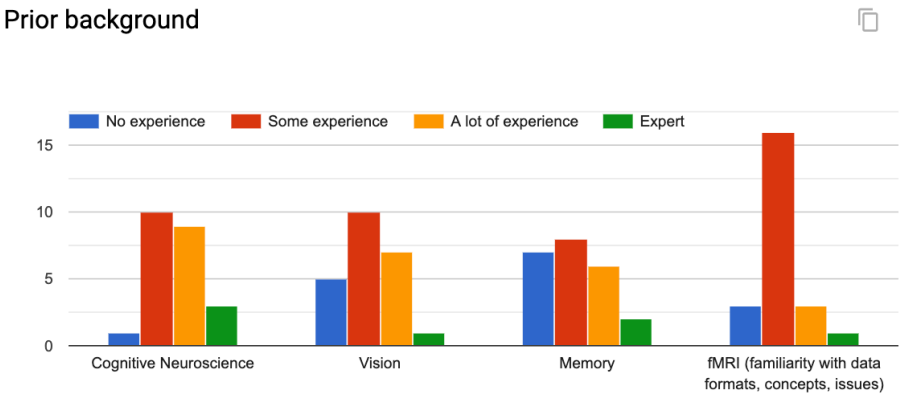
10

File organization



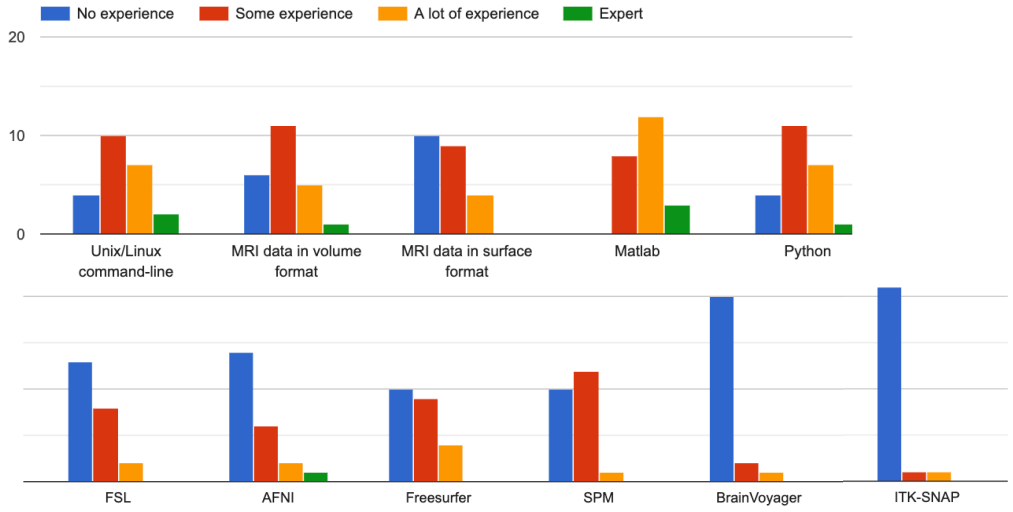
11

Background of workshop attendees



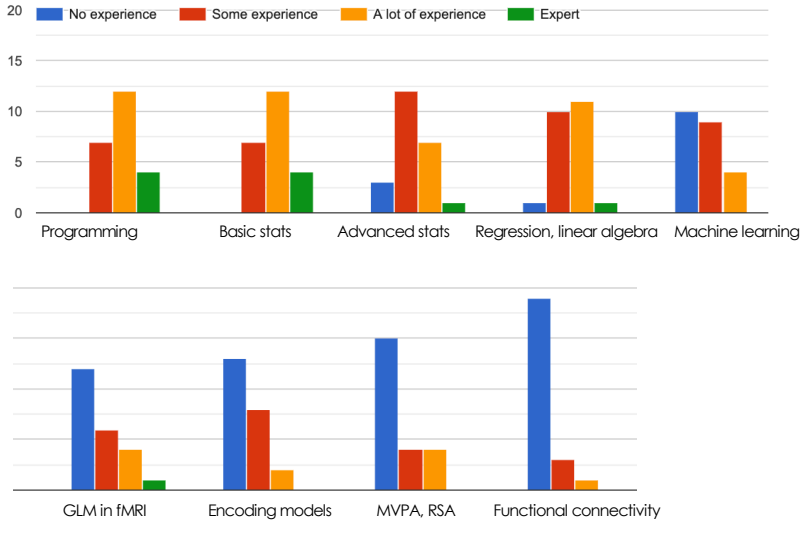
12

Background of workshop attendees



13

Background of workshop attendees



14

Should you use a tool (method) without understanding it?

Attitude 1: "Don't reinvent the wheel" / "It's probably right" / "Who cares?" / "Yeah, I used it, but don't really know the details."

Attitude 2: "You don't get to use the tool if you don't know how it works." Danger.

- When should you take **Attitude 1** vs. **Attitude 2**? When is it okay to be ignorant?
- Proposal: It's more okay to take **Attitude 1** if the following conditions are met: (i) the **black box** has only one way to use it, (ii) the **black box** has only one type of input and only one type of output, (iii) we can look at the results of the **black box** and can confirm that it worked correctly.
  - Example in neuroimaging: FreeSurfer
  - Example in real life: microwave, car for travel, light switch, Google Maps, hammer.
- But when the black box has multiple ways of usage, many inputs, many outputs, and it's hard to confirm it did what we wanted....
  - Example: fMRI analysis in general??

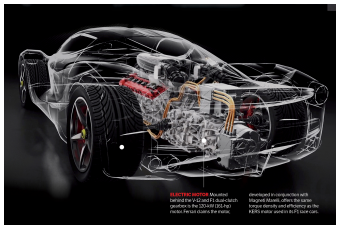
15

Different metaphors on methods

- Black box vs. the details

- User vs. developer

- Learn how to drive vs. learn how to build a car



- Prepared food vs. food from scratch

16

The journey of a data point

- What is the full set of transformations that lead to that data point?
- Do you know what happened to your data?
- Can you communicate that to the reader?
- Can the reader hold these steps in their head (so as to appreciate the validity of the results)?

17

Complexity and largeness pose a special problem

- Provenance tends to be lost with big data or shared data.
- Try to avoid the game of telephone.
- Try to avoid complex, convoluted analyses.
- With obfuscated analyses, we cannot even assign a truth value to a figure. (Analogy with prose: highly convoluted text has no meaning.)
- The challenge of reproducibility:
  - Sharing code is helpful
  - But can you also achieve clear conceptual description? (Can you reproduce an analysis from the Methods description alone?)

18