# Week of 9/18 Deliverables

EEG

# General Update: What we are actually doing

Our plan is to create a visualization and analysis tool to facilitate the exploratory data analysis of HBNB and other high dimensional and phenotypically rich neuroimaging datasets. Particularly, a tool that can visualize both biological and behavioral metrics in a meaningful way, and can aid researchers who are looking to develop data driven nosologies and dimensional theories of mental health.

# Deliverables Summary (mainly background reading):

- Background on statistical methods for handling high dimensional multi-modal data.
  - MGC, CCA, kCCA
- Background on analyzing large phenotypic datasets
  - p-Factor paper, Multivariate Distance Based Analytic Framework, Beyond DSM-5 paper
- Background on MRI, MRI analysis
  - Greg's Thesis, Mindboggle
- Background on EEG analysis
  - Age-Sex and naturalistic stimuli

# Deliverables: Background on Statistical Methods for handling High Dimensional Multimodal Data (Ryan, Nitin)

- MGC paper
- CCA / kCCA note

# Progress

- MGC paper. ✔
  - DoD: notes https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/mgc.md
- kCCA note. ✔
  - DoD: notes https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/nkumar14/papers/kcca.md

# MGC Paper ✔

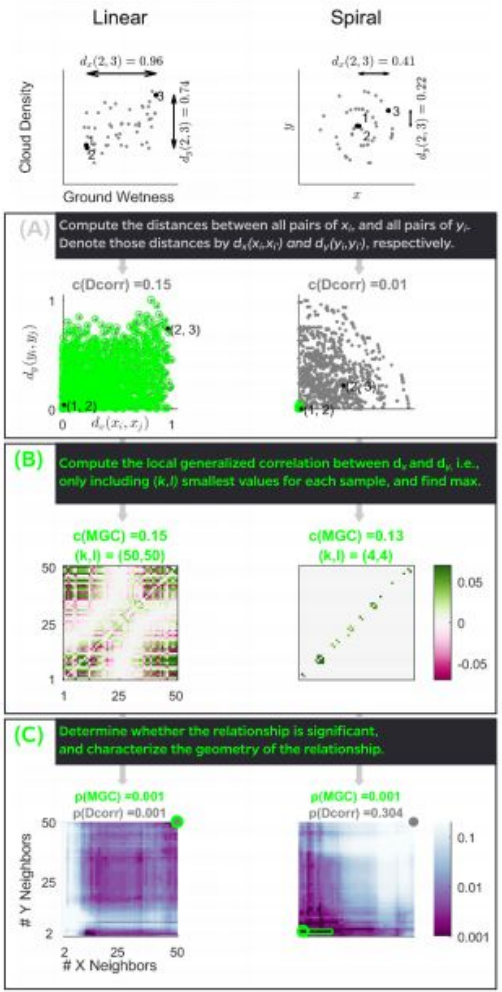
New method for testing an independence hypothesis when low sample size data is high dimensional and highly nonlinear. Method theoretically dominates global tests for dependence, and gives insight into the problem rather than just testing for a relationship.

- null: f(x, y) = fxfy
- alt: f(x, y) =/= fxfy

## Significance to Us

- MGC seems like a promising way to link behavioral and biological measures. We can use MGC to test whether a relationship exists between high dimensional neural activity and behavioral scores.

DoD: https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/mgc.md



# Algorithm Note on kCCA ✔

- Example case: 2 copies of a text, one in German, one in English
- We want to find projections of maximal correlation to find topics that are the same in 2 different languages
- Try maximizing correlation of projected German and English vectors
- Translates into an eigenvalue equation
- Kernel equation uses Legrange multipliers and regularization
- Presents a mathematical representation of problem, need other references for simpler solution

## Significance to Us

- It's a relevant method, but we need to look elsewhere to simplify it
  Potential open source solution Pyrcca

DoD: link here

# Deliverables: Background on analyzing large phenotypic datasets (Ryan, Nitin, Yuka)

- p-Factor paper
- Multivariate Distance Analytic Framework
- Beyond DSM-5 paper

# Progress

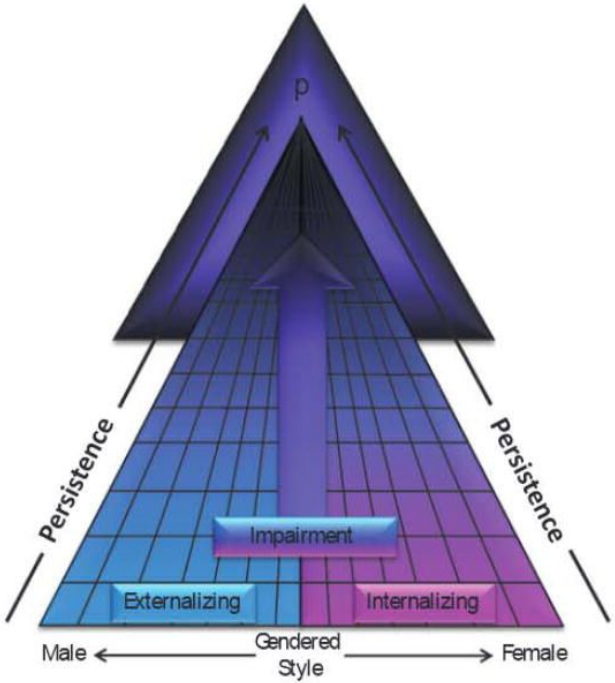
- p-Factor paper. ✔
  - DoD: notes https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/pfactor.md
- Multivariate Distance Analytic Framework ✔
  - DoD: notes
  - https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/nkumar14/papers/multi-var-dist-based.md
- Beyond DSM-5. ✔
  - DoD: notes
  - https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/nkumar14/papers/data-driven-categorization.md
  - https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/beyonddsm.md

# p-Factor Paper ✔

Although useful for clinical purposes, DSM 5 not a great model (doesn't explain comorbidity, sequential diagnoses, multiple diagnoses)

Dimensional models are better, but often dimensions are highly correlated with one another (Internalizing and Externalizing highly correlated)

When accounting for a p-Factor -- one number summary of the propensity for someone to develop a mental illness of any kind, these dimensions correlate less and can be used to discriminate among illnesses.




## Significance to Us

- In visualizing dimensional models of phenotypic data, accounting for a p-Factor may let us see differences between diagnostic clusters more. (We can also develop a better p-Factor!)

DoD:
https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/pfactor.md

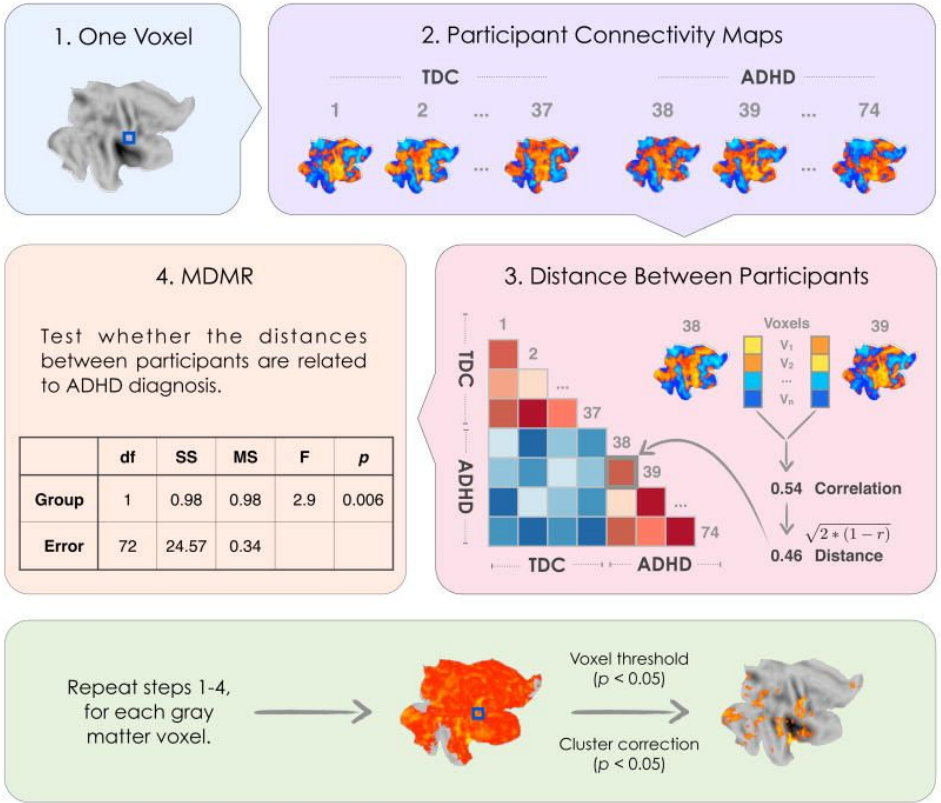
# Multivariate Distance-Based Analytic Framework

- Advantages of framework:
  - Publicly available code
  - Don't need to know number of dimensions
  - Don't need to lower resolution
  - No seed selection necessary
  - Don't need to set any parameters
- Use MDMR (multivariate distance matrix regression) yields pseudo-F statistic

**Significance to Us**

- Defined parameter-less analysis algo robustly tested on many results
- Already shown as extendable with the clustering in the phenotypic paper
- Can be extended by using more metrics from fMRI (partial correlation, directional info, etc)

DoD: [Link to notes]



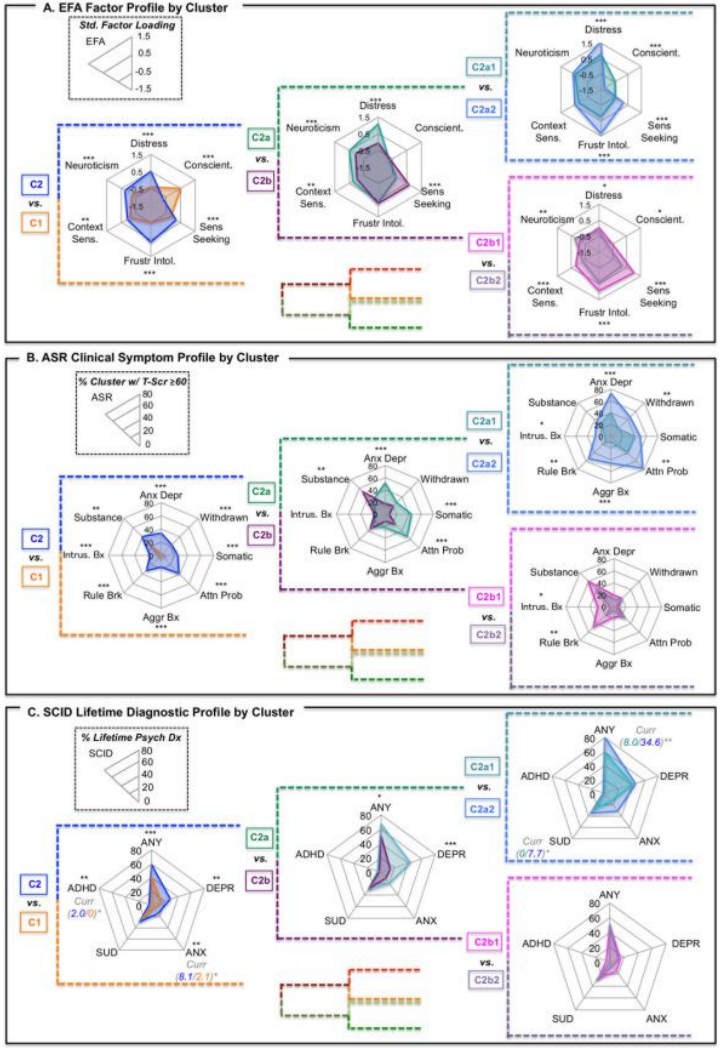
# Beyond DSM 5 ✔

- Weaknesses of categorical definitions of illnesses are increasingly apparent
- Data-driven approaches may help identify more behaviorally refined biological phenotypes to address the profound heterogeneity

**Significance to Us**

- Number of fMRI and phenotypic methods that we can use
- Clustering methods across different dimensions and combining with fMRI data
- Shown results from such measures on another large dataset

DoD: [link to notes]



# Deliverables: Background reading on MRI, fMRI, and MRI analysis (Ryan, Ronak, Vidur, Yuka)

- Greg Kair's thesis, ndmg
- Mindboggle summary paper
- Survey of computational methods used on fMRI data, especially those useful to the project.
- Neural Correlates of Reading Fluency Deficits using fMRI
- A database of age-appropriate average MRI templates

# Progress

- Greg's thesis ✔
  - DoD: link
- Mindboggle ✔
  - DoD: link
- Survey of preprocessing, analysis, and computational methods associated with MRI/fMRI studies. ✔
  - DoD: MRI Analysis Write-up
- Neural Correlates of Reading Fluency Deficits using fMRI ✔
  - DoD: https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/vidurkailash/papers/Langer-fmri.md

## Greg's Thesis ✔

- Provides a one-click pipeline to take DTI images to brain graphs
- Useful tool to pre-process, register, and generate useful mathematical objects from brain images
- Has all of the niceties that would make it easy to use to add to a visualization tool (dockerized, amazonized, pythonized)

DoD:
https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/gregs.md

## Mindboggle ✔

- Tool to take preprocessed MRI data to features representing the shape of the brain
    - Area, volume, curvature, thickness, depth, ect.
- Many studies show these features can be used as biomarkers for different illnesses
- Fast, dockerized, usable
- Comes with a dataset with expert annotated cortical information

DoD:
https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/rmarren1/papers/mindboggle.md

## Survey on computational methods used for fMRI ✔

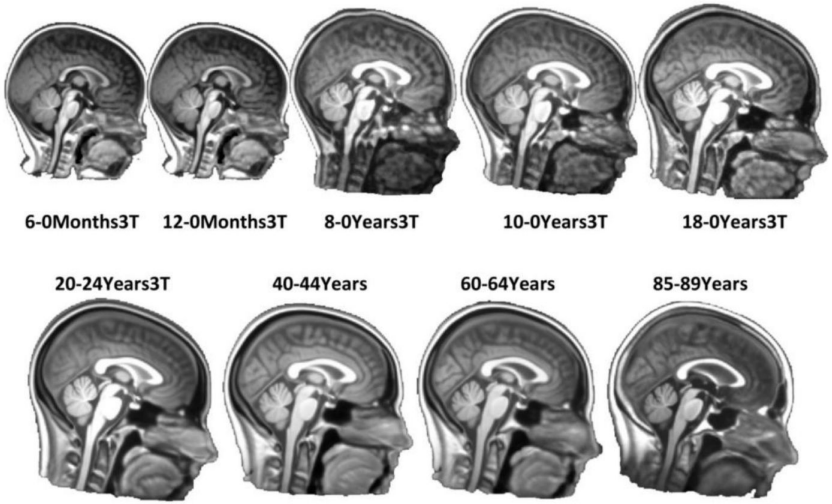
Overview of Content:

- Background and difference between structural and functional MRI.
- General Important Points.
- Data Structure.
- Common Preprocessing Steps.
- Statistical/Computational Methods of Interest.
- Useful resources.

DoD: MRI Analysis Write-up

# Age-appropriate MRI template ✔

- Problems exist in normalizing:
  - pediatric populations to adult reference (solved by Fonov et al. (2011) using NIHPD)
  - young adults to older adults*
- Creation of fine-grained, age-appropriate average MRI reference templates could propose a possible solution (http://jerlab.psych.sc.edu/NeurodevelopmentalMRIDatabase/)




## Significance to Us

- Age-appropriate templates are great for studies of developments in brain structure and functions
- Be more precise in analysis of neurological and mental diseases that are sensitive to age.
- **DoD:**

## Neural Correlates of Reading Fluency Deficits using fMRI ✔

- Investigated “natural reading” at varying and individual-based reading speeds in RD (Reading Deficient) and TYP (Typical Developing Children) children in order to investigate how the pattern of activation in core reading regions changes as the opportunity to read fluently is manipulated.

### Importance to Us:

- Uses fMRI to help classify differences in activation patterns between RD and TYP children.
- This is a specific analysis of a certain disorder. We recreate this analysis and make it more general.

# Deliverables: Background reading on EEG analysis

**(Nitin, Vidur)**

- Age-sex modulating responses to naturalistic stimuli
- Assessing information processing using EEG and eye tracking.

# Progress

- Age-Sex and Naturalistic Stimuli ✔
  - DoD: link to notes

- EEG and Eye tracking ✔
  - DoD: https://github.com/NeuroDataDesign/eeg-panda-s17f18/blob/master/docs/vidurkailash/papers/Langer-EEG.md

# Age-Sex and Naturalistic Stimuli ✔

- Investigated influence of age and sex on responses to naturalistic video stimuli
  - First EEG study to report a measure of across-subject neural similarity with clear age and sex effects

**Importance to Us**

- ISC: potential "discriminability" like metric with strong support
- Analyses done here are very exploratory; can identify phenotype groups (or known combinations of groups) and compare stimuli on these groups
- CCA: relevant fMRI metric to help determine locations of high correlations

DoD: link to notes

## EEG and Eye Tracking ✔

- Presents a dataset combining electrophysiology and eye tracking intended as a resource for the investigation of information processing in the developing brain.
    a. Many links between cognitive deficits and mental health disorders. With more and more datasets, the hope is to have a much deeper/clearer understanding of what these relationships mean.

## Importance to Us:

- Explained how all behavioral tests were conducted from which we can develop our own visualization tool because current visualizations aren't very good.
- There are performance measures on the cognitive tests from a behavioral view and biological metric view. Would be interesting to work with this data set.

# Goals for Next Week

- Project proposal -- check in wednesday
- Begin work on data we have available (EEG, MRI), particularly getting data in a workable form
- Test software explored this week for usability