Get these slides!

github.com/jmlammert/HBN_methods_lunch

Research Roundtable: Healthy Brain Network Dataset

Jessica Lammert

CBS Methods Lunch

Part 1:
Introduction to the Healthy Brain Network (HBN)

Part 2:
Applications of Machine Learning with the HBN

Child Mind Institute
Healthy Brain Network

INDI
International Neuroimaging Data-Sharing Initiative

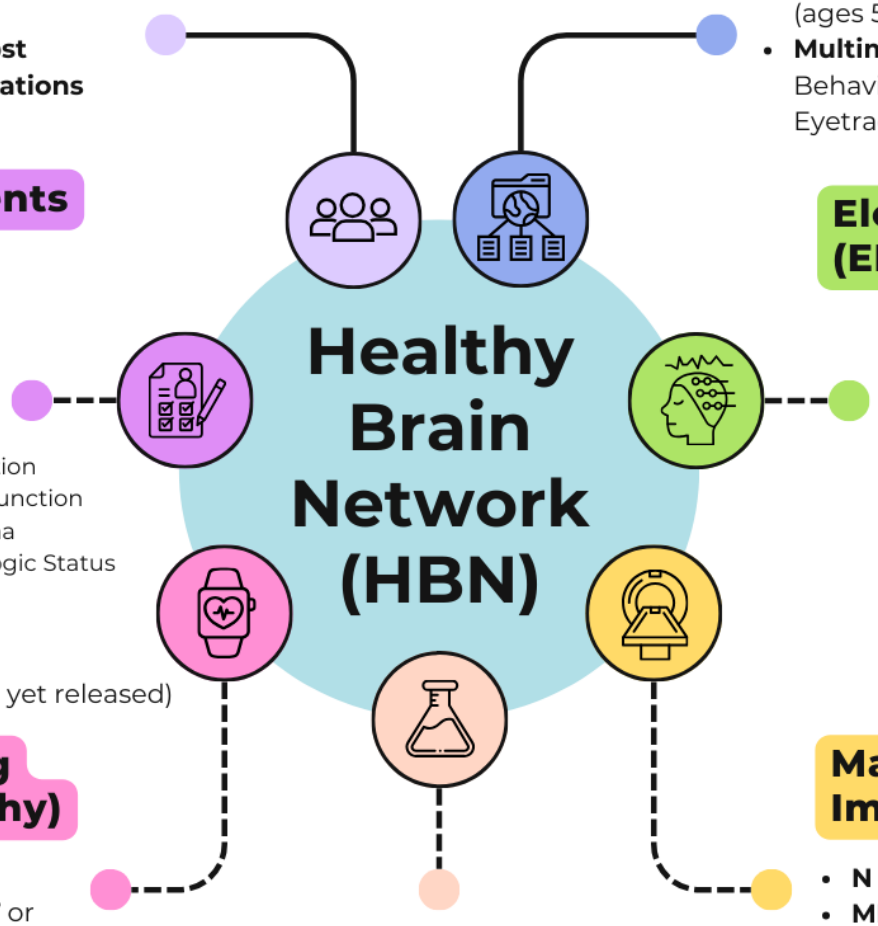
Last updated: May 2024
Data Release 11

Community-Referred Recruitment Model

- Inclusion of **diverse communities** across **New York City**
- Recruitment based on **perceived clinical concerns**
- Participants provided with **no-cost** comprehensive **diagnostic evaluations**

Large-scale Transdiagnostic Open Science Database

- **Open data** sharing initiative (**N > 4000**)
- Broad range of **mental health and learning disorders** in the **developing human brain** (ages 5-21)
- **Multimodal approach**: Phenotypic, Behavioral Monitoring, MRI, EEG, Eyetracking, Biological Samples



Phenotypic Assessments

- **N = ~5000**
- **Assessments > 140**
- **Domains**:
  - Demographics
  - Cognition
  - Language
  - Emotional and Psychological Function
  - Social, Emotional, and Behavioral Function
  - Family Structure, Stress, and Trauma
  - Physical Skills, Fitness, and Physiologic Status
  - Substance Use and Addiction
  - Medical Status
  - Diagnosis
- **Audio and video recordings** (not yet released)

Electroencephalography (EEG) & Eyetracking

- **N = ~4000**
- **128-channel high-density** EEG (SR: 500 Hz, BP: 0.1 - 100 Hz)
- **Infrared video-based** eyetracker (SR: 120 Hz, res: 0.1°)
- **Resting state + tasks**:
  - Sequence Learning
  - WISC Symbol Search
  - Inhibition/Excitation
  - Naturalistic video watching

Behavioral Monitoring Technology (Actigraphy)

- **N = ~2000** (not yet released)
- Wrist-worn **ActiGraph wGT3X-BT** or **GENEActiv**
- Day and night for **1 month**
- **Recordings**:
  - Physical Activity
  - Sleep
  - Ambient Light
  - Temperature

Biological Samples

- **Saliva samples** (not yet released)
- **Blood samples** (not yet released)
- **Teeth samples** (not yet released)
- **Hair samples** (not yet released)

Magnetic Resonance Imaging (MRI)

- **N = ~4000**
- **MRI Sequences**: T1, T2, FLAIR, DWI, EPI
- **Resting state + tasks**:
  - Eye-movement task (PEER)
  - Naturalistic video watching

Healthy Brain Network Dataset

- Clinically referred large-scale study

Community Website

Project Website

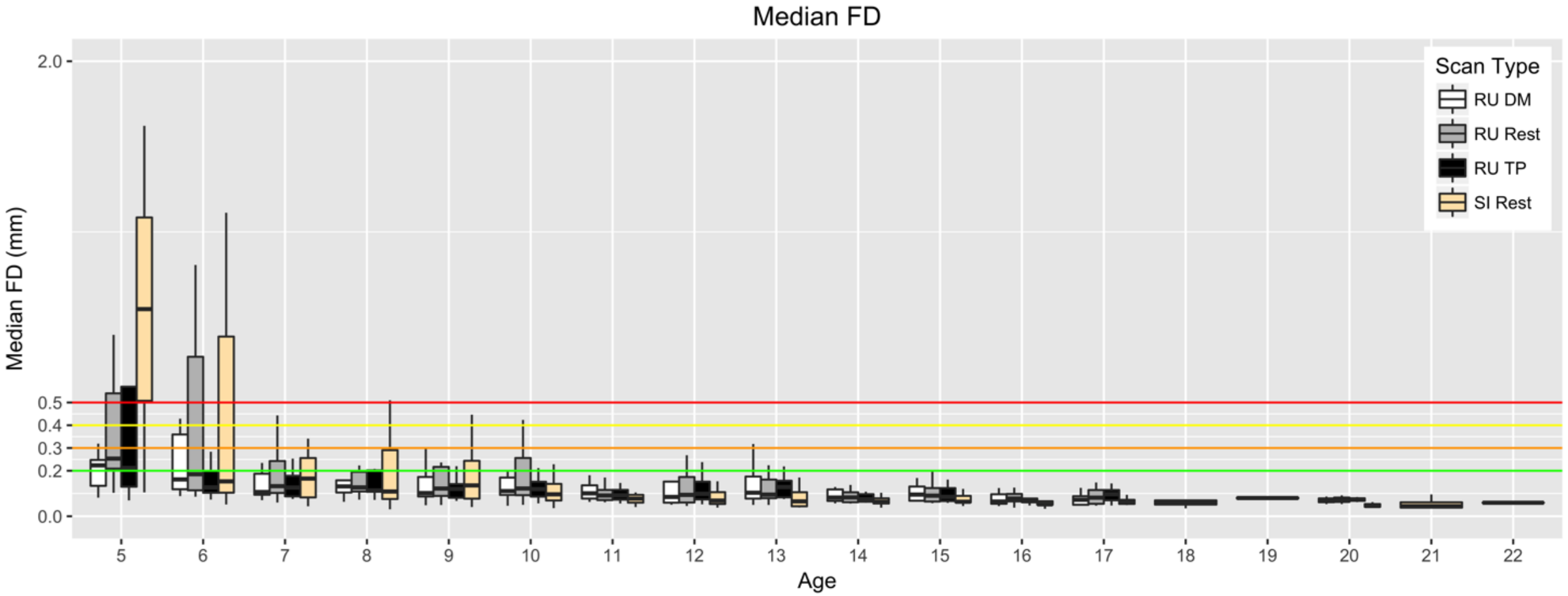
Alexander et al., 2017

CHILD MIND® INSTITUTE
healthy brain network

- MRI (*N* ~4,000)
  - T1w, FLAIR, T2w, DWI, resting-state, *Despicable Me*, *The Present*, Peer
- EEG (*N* ~4,000)
  - Resting-state, WISC symbol search, sequence learning, inhibition/excitation, visual perception/decision-making, naturalistic viewing, *The Present*

**datasets are not guaranteed to be complete (neuro/pheno) or "good" quality*

Motion


Median FD
Median FD (mm)
2.0
0.5
0.4
0.3
0.2
0.0
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Age
Scan Type
RU DM
RU Rest
RU TP
SI Rest

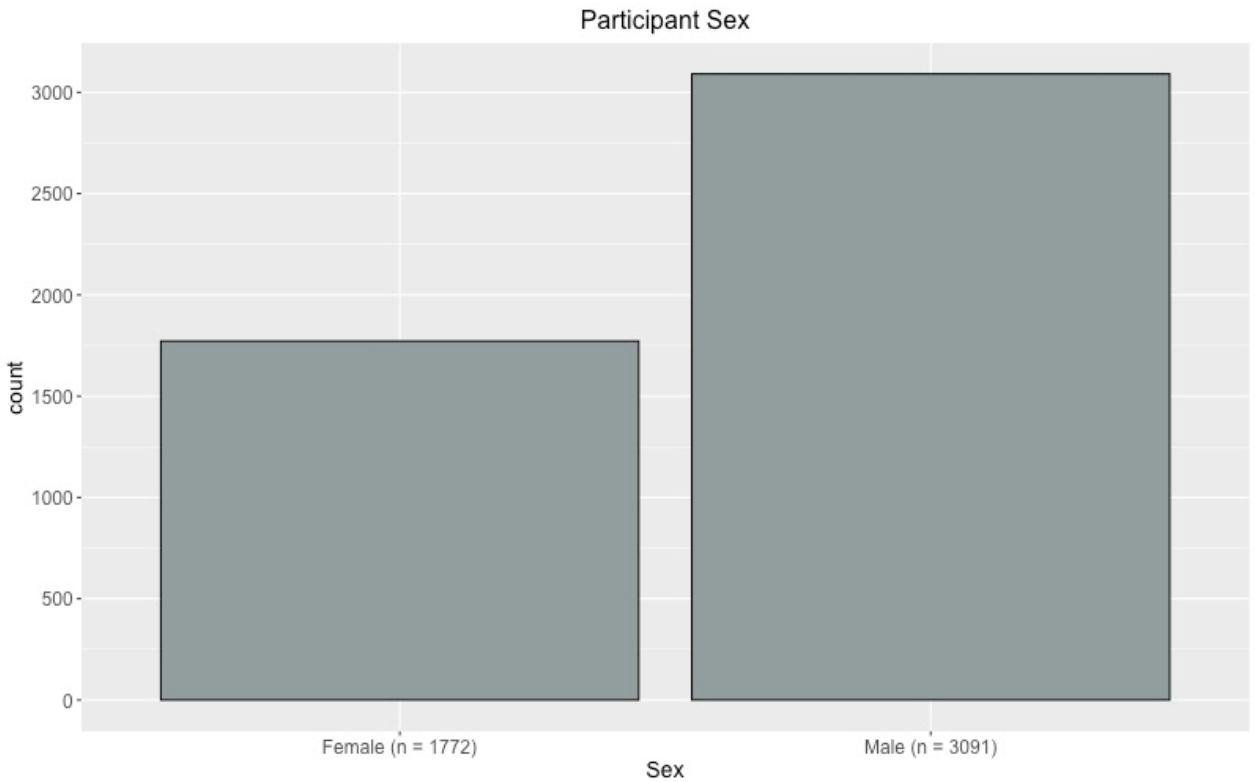
Phenotypic Assessments

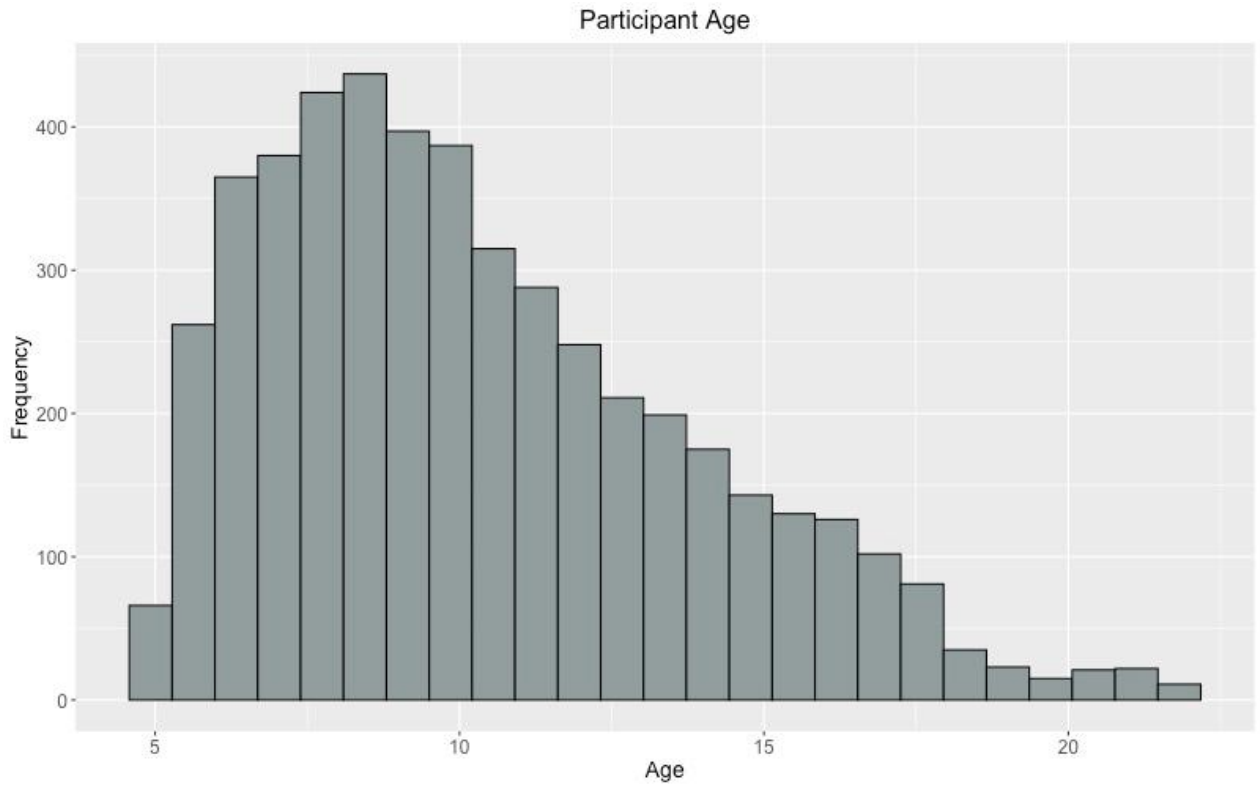
Phenotypic Assessment Protocol

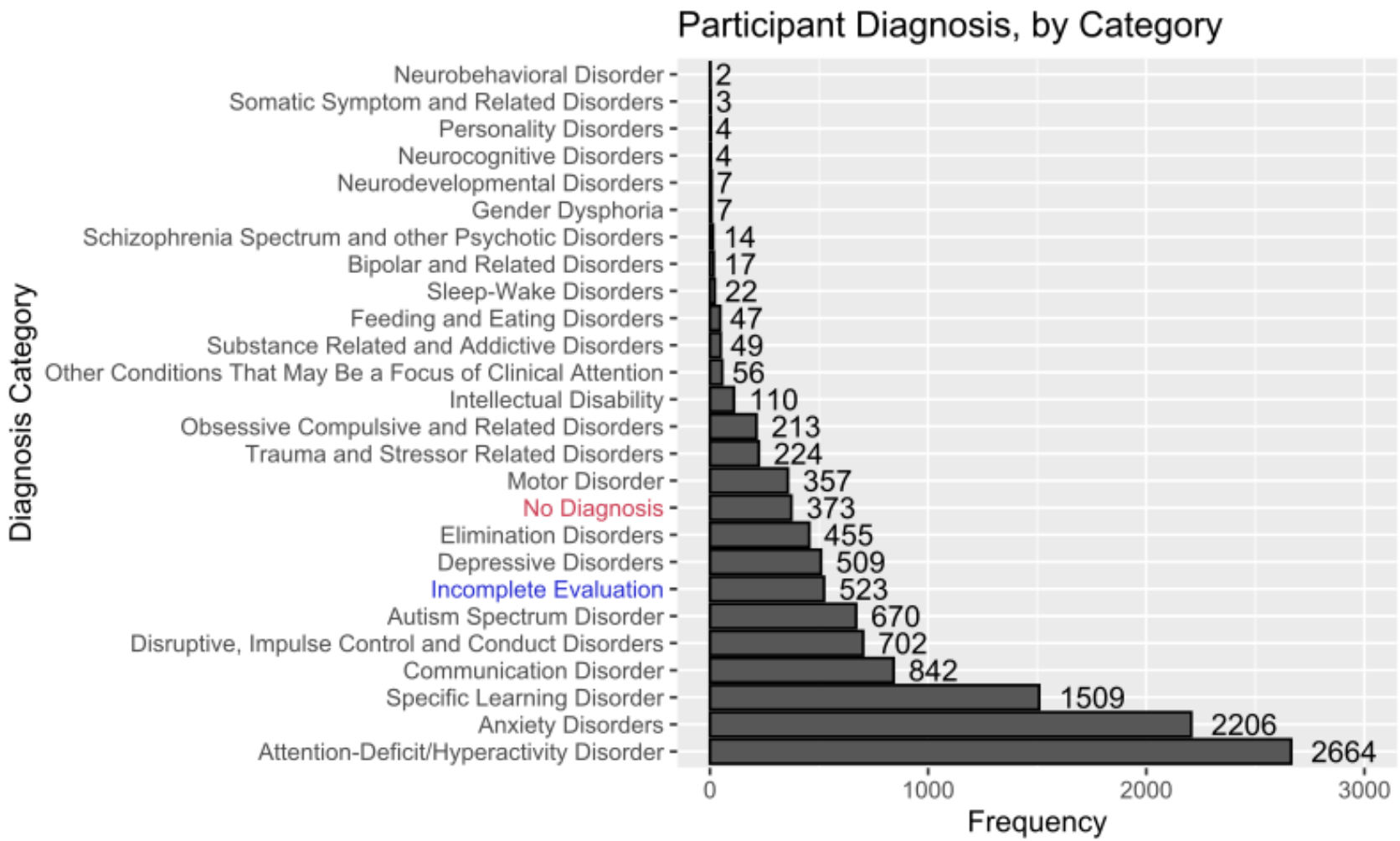
General Information	Behavioral Measures
Demographics CMI Symptom Checker Edinburgh Handedness Inventory Intake Interview Physical Activity Questionnaire for Older Children (PAQ-C) (8–14) Physical Activity Questionnaire for Adolescents (PAQ-A) (14–19) Barratt Simplified Measure of Social Status Financial Support Questionnaire Medical History Questionnaire—Family Pregnancy and Birth Questionnaire	Child Behavior Checklist (CBCL) (5–17) Youth Self Report (YSR) (11–18) Adult Self Report (ASR) (18+) Screen for Child Anxiety Related Disorders (SCARED)—Parent Report & Self Report (8–18) State Trait Anxiety Inventory (STAI) (18+)—Self Report Mood & Feelings Questionnaire (MFQ)—Parent Report & Self Report (8+) Affective Reactivity Index—(ARI-S) Self Report Columbia Suicide Severity Rating Scale (C-SSRS)—Self Report (7+) Extended Strengths and Weaknesses Assessment of Normal Behavior (E-SWAN) (5–17) Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN) (6+) Conners ADHD Rating Scales Self Report Short Form (Conners) (8+) Repetitive Behavior Scale (RBS) (5–21) Autism Spectrum Screening Questionnaire (ASSQ) (5+) Social Communication Questionnaire (SCQ) (5+) Social Responsiveness Scale-2 (SRS-2) (5+) Strengths and Difficulties Questionnaire (5+) The Columbia Impairment Scale (CIS) Parent ad self report (5+) Social Aptitudes Scale (SAS) (5+) WHO Disability Assessment Schedule (WHODAS) Parent and Self-Report (5+) Food Frequency Questionnaire (FFQ) (5–17) Inventory of Callous-Unemotional Traits—Parent Report (5+)
Physical Measures	**Family Structure, Stress and Trauma**
FITNESSGRAM (Pushups, Curl-ups, Trunk-Lift, Sit and Reach, Grip Strength) Cardiovascular Fitness Test Vitals (Heart Rate, Blood pressure) Measurements (Height/weight, Waist circumference, Bio-impedance) Blood Draw (Endocrine, Immunologic, and Metabolic profiling; Genetics) Buccal Swabs (Genetics) Urine Sample (Toxicology screen, Pregnancy test: 11+) Ishihara Color Vision Test Electroencephalography (EEG)/Eye Tracking Magnetic Resonance Imaging (MRI) Peterson Puberty Scale (6–17) Sleep Disturbance Scale for Children (SDSC) (6–15)	Family History-Research Diagnostic Criteria (FH-RDC) Parental Stress Index IV (PSI-IV) Alabama Parenting Questionnaire—Self Report (APQ) (6–18) Alabama Parenting Questionnaire—Parent Report (APQ) (6–18) Children's Perception of Interparental Conflict (CPIC) (8–18) Distress Tolerance Index—Parental Self Report Children's Coping Strategies Checklist—Revised (CCSC) (8–18) UCLA Trauma Reactivity Sale for DSM-V (UCLA) (5–18) Negative Life Events Scale (NLES)—Self Report (8–18) Negative Life Events Scale Parent Report (8–18) Adverse Childhood Experiences Scale (ACES) (18+)

Cognition and Language Tasks	Substance Use and Addiction Measures
NIH Toolbox Tasks: Flanker, Card Sort and Processing Speed Temporal Discounting Task Quotient ADHD System Rapid Automatic Naming & Rapid Alternating Stimulus Test (RAN/RAS) (5) Wechsler Intelligence Scale for Children-V (WISC-V) (6–17) Wechsler Adult Intelligence Scale-IV (WAIS-IV): (17+) Wechsler Abbreviated Scale of Intelligence-II (WASI): (17+) Wechsler Individual Achievement Test—III (WIAT) Differential Ability Scales—II (DAS) (5 or IQ below 70) Clinical Evaluation of Language Fundamentals—5th Edition (CELF-5) Goldman Fristoe Test of Articulation—II (GFTA) Comprehensive Test of Phonological Processing—II (CTOPP) Test of Word Reading Efficiency (TOWRE) (6+) Expressive Vocabulary Test (EVT) (when indicated) Peabody Picture Vocabulary Test (PPVT) (when indicated)	Fagerstrom Test for Nicotine Dependence (FTND) (18+) Alcohol Use Disorders Identification Test (AUDIT) (11+) Modified Fagerstrom Tolerance Questionnaire- Adolescents (FTQA) (13–17) European School Survey Project on Alcohol & Other Drugs (ESPAD) (10+) Internet Addiction Test (IAT) Parent-Child Internet Addiction Test (PCIAT) Yale Food Addiction Scale (YFAS) and YFAS-Child
Diagnostic Assessments	**Longitudinal Follow Up Measures**
Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) Child and Adolescent Psychiatric Assessment Schedule (Cha-PAS) (when indicated) Vineland Adaptive Behavior Scale—Parent/Caregiver Rating Form (when indicated) Yale Global Tic Severity Scale (YGTSS) (when indicated, 6+) Yale-Brown Obsessive Compulsive Scale (Y-BOCS) (when indicated, 18+) Children's Yale-Brown Obsessive Compulsive Scale (when indicated, 6–18)	Youth Services Survey (YSS) & Services Assessment for Children and Adolescents (SACA) Follow Up: CBCL Follow Up: Columbia Impairment Scale Parent snd Self Report Follow Up: WHODAS Parent and Self Report

Age and Sex distribution (*N* = 4,863)







Levels of Certainty

Confirmed

Presumptive

Requires Confirmation

Rule-out

+

By History

*NOTE

- Principal Investigators must sign a Data Usage Agreement (DUA) with the CMI to access phenotypic data *only* (see example)

- Phenotypic data is accessed through LORIS: https://data.healthybrainnetwork.org/main.php

- Assessment list and data dictionaries *are available* on CBS server

HBN on the CBS server

- Raw and pre-processed MRI and EEG data are available

- Users can **store** processed data, scripts, etc. **or call** to the drive from other directories
  - **25 TB**
  - Store data outputs and code (e.g., .csv files) on personal drive where possible

- **Shared drive:** keep original files intact, tag folders with your initials if needed

Request access to the **open_data_hbn** project on Schulich HelpDesk:

https://westernu.atlassian.net/servicedesk/customer/portal/18

Network Services > Network Folder Access Request

Pre-processed Data

- MRI
  - fMRIprep 23.0.1 run on DRAC Graham cluster
  - skull-stripping, brain tissue segmentation, alignment, etc.
  - Output space: MNI152 non-linear asymmetrical, 2mm resolution

  - Out of ~1,500 selected datasets, ~1,000 were successfully pre-processed
  - ~500 met manual QC criteria (1/3) = **TP-tasks dataset**
  - Denoising regresses out confounds (CSF, WM, motion)

- EEG and remaining MRI pre-processing ETA Summer 2025
  - MRI: *N* ~2,500 remaining, 1/3 acceptable quality = ~1,300 participants

Links (Part 1)

HBN Community website
https://healthybrainnetwork.org/

HBN Project website
https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/

HBN data publication
https://www.nature.com/articles/sdata2017181

HBN phenotypic protocol
https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Pheno_Protocol.html

DUA example
https://github.com/jmlammert/HBN_methods_lunch/blob/c2e43f7dc08c780d399faeb2971d2540f2aecece/DUA_CMI_template.docx

LORIS
https://data.healthybrainnetwork.org/main.php

Schulich HelpDesk
https://westernu.atlassian.net/servicedesk/customer/portal/18

Post-processing package
https://github.com/tkkuehn/hbn_postprocessing

Denoising application
https://github.com/akhanf/denoise-fmri.git

Contact:

E-mail: jlammert@uwo.ca

Slack: westernbrainandmind.slack.com > #healthy-brain-network

Part 2:

Analysis & Feedback

“Big data” and Predictive Modeling

- Want to model the relationship between multiple variables (often biological and behavioural)
  - E.g., What is the relationship between IQ and activity in PFC?
- Supervised and unsupervised learning
- Categorization, classification, regression, etc.

Study Overview

- **Background:** Narrative (story) comprehension is a multimodal process, important for learning and communication in childhood

- **Research Question:** How do socio-cognitive skills support narrative comprehension in children?

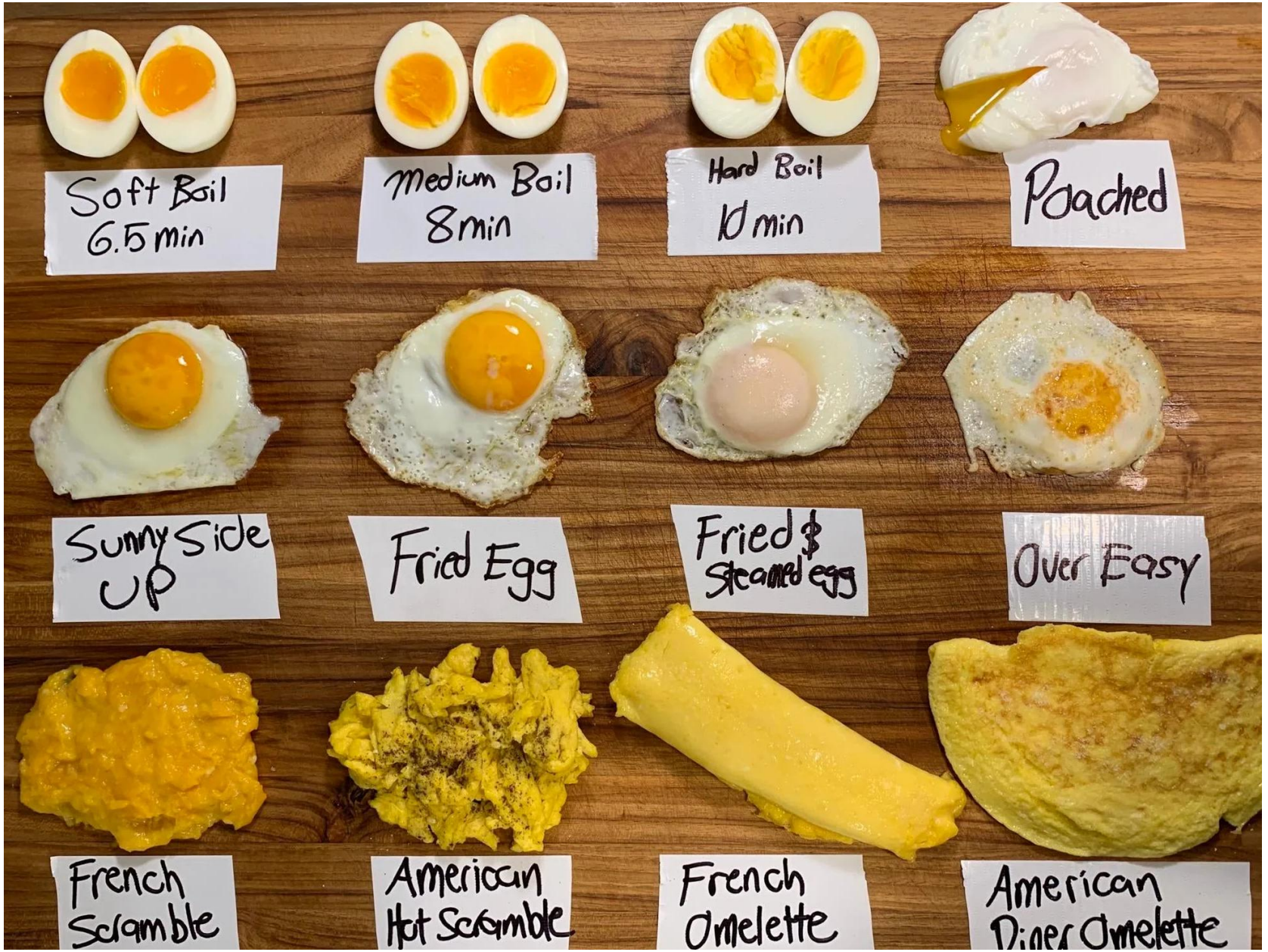
- **Approach:**
  - Multiple socio-cognitive tests
  - fMRI during movie-watching
  - Cluster participants into “narrative comprehension profiles”
  - Predict (classify) profile using socio-cognitive test scores

Question:
Are there more ways to cook an egg or analyze neuroimaging data?

A. Cook an egg

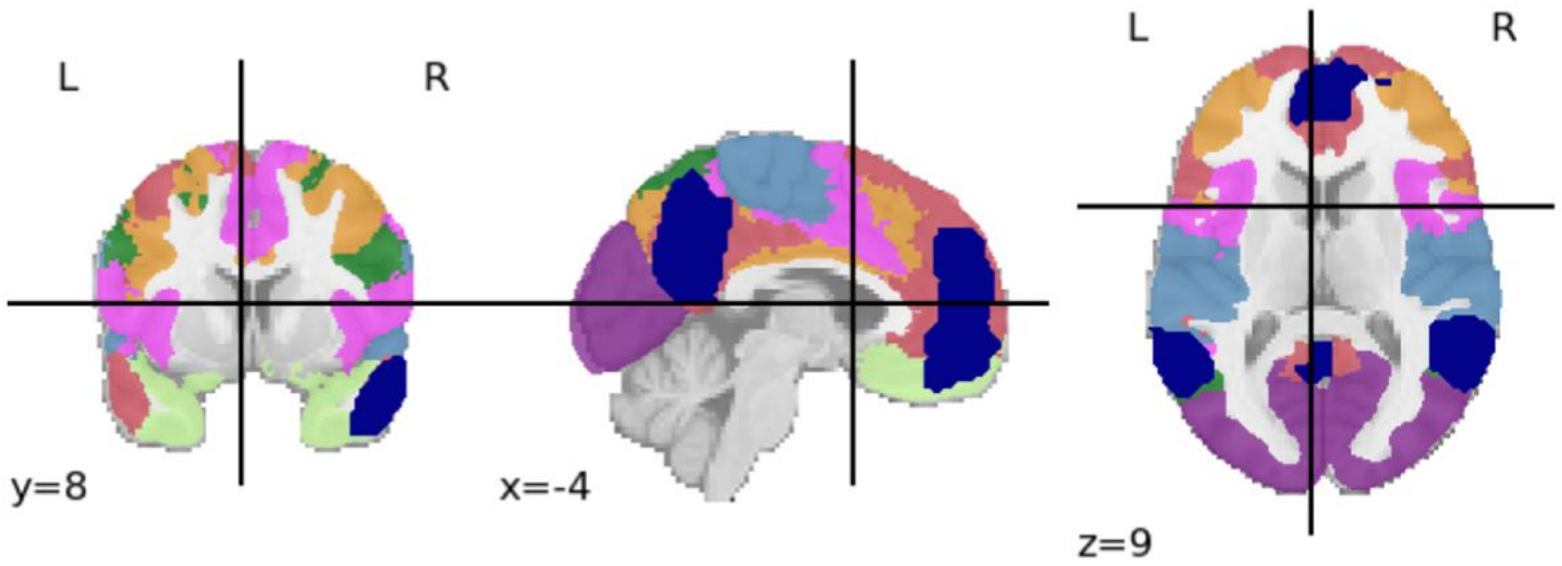
B. Analyze data

C. The limit does not exist

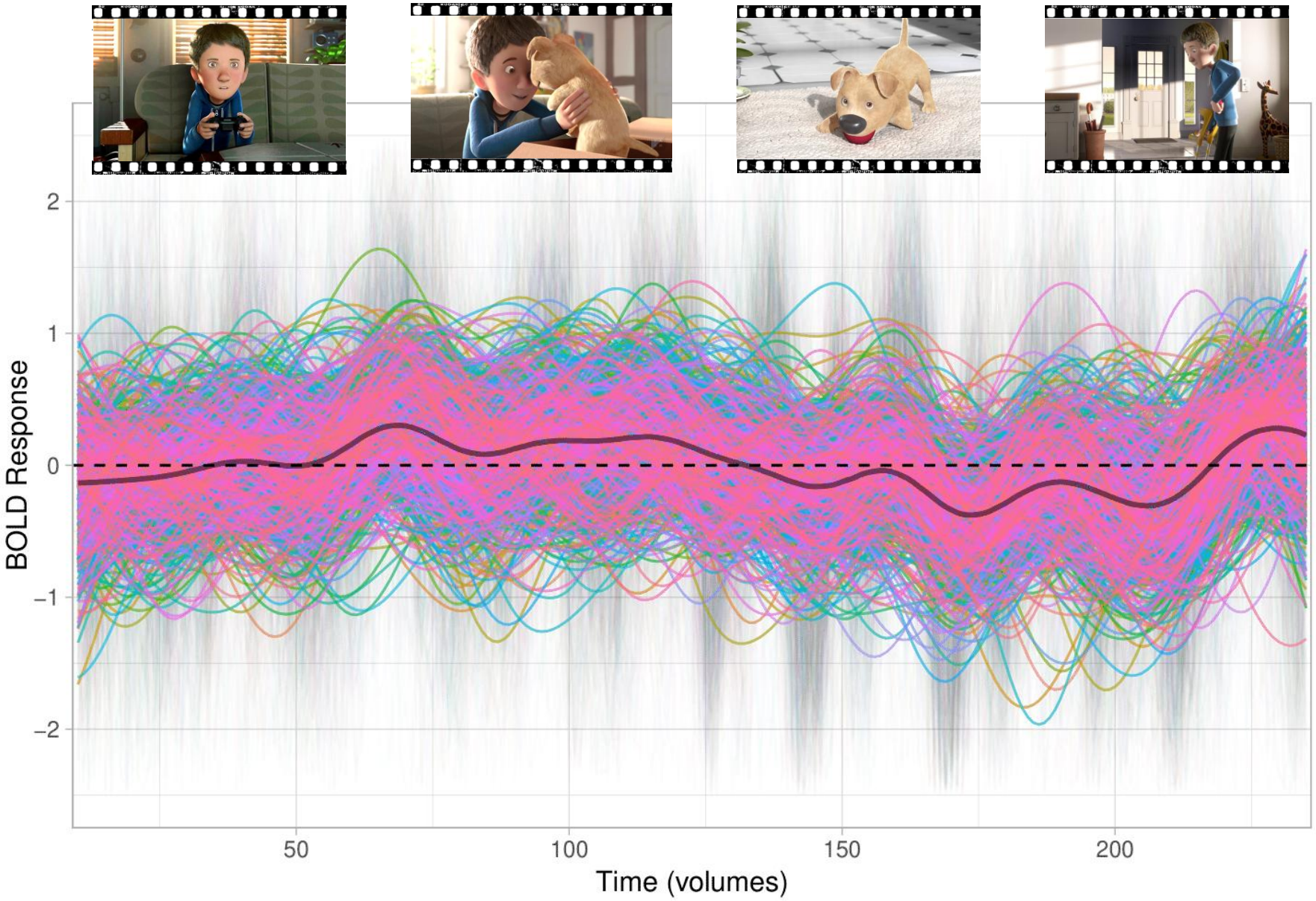


Brain Networks

Index	Network Name
1	Visual
2	Somatomotor
3	Dorsal Attention
4	Ventral Attention
5	Limbic
6	Frontoparietal
7	Default-mode
8	Theory of Mind



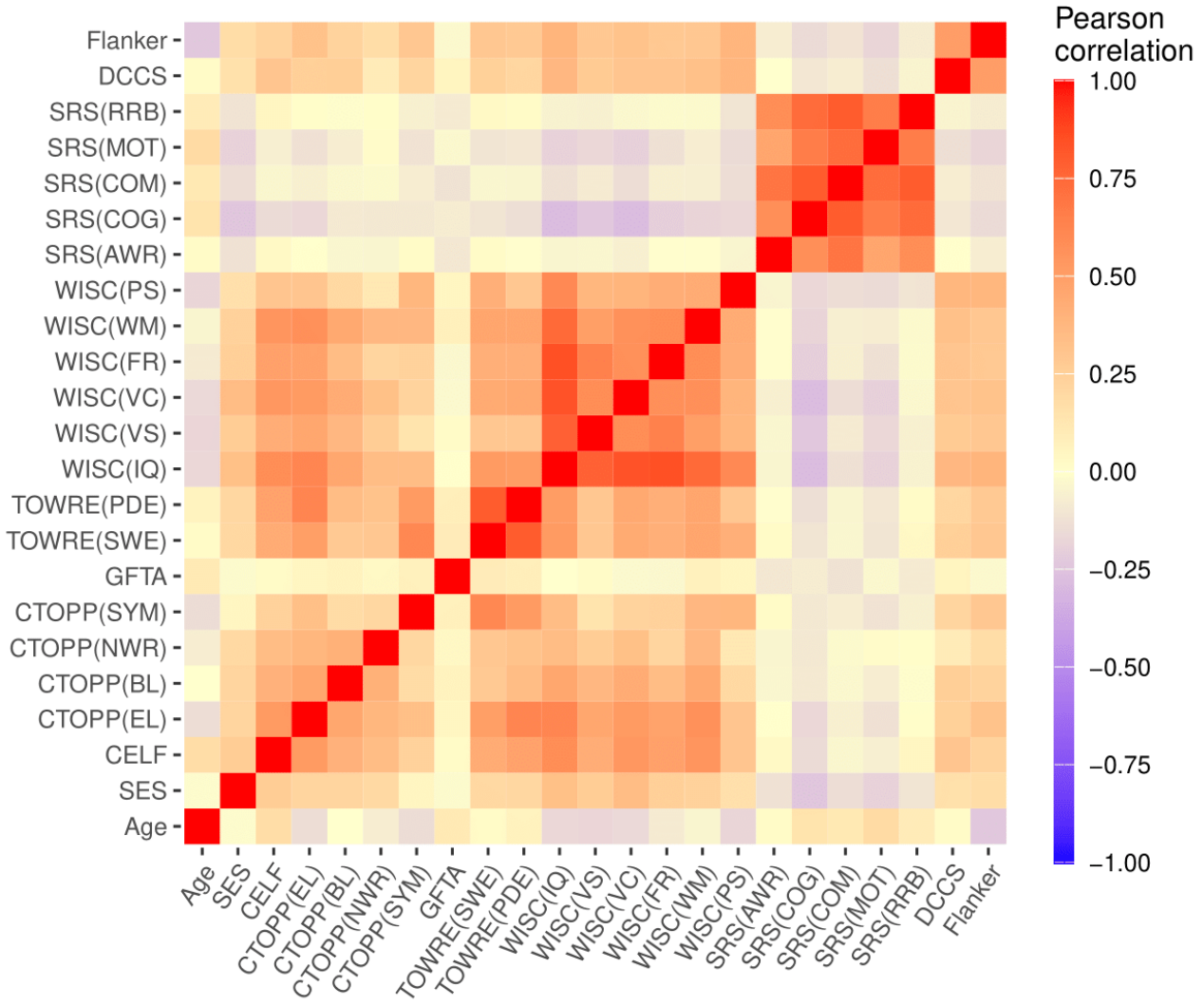
BOLD activation to *The Present* in the Dorsal Attention network



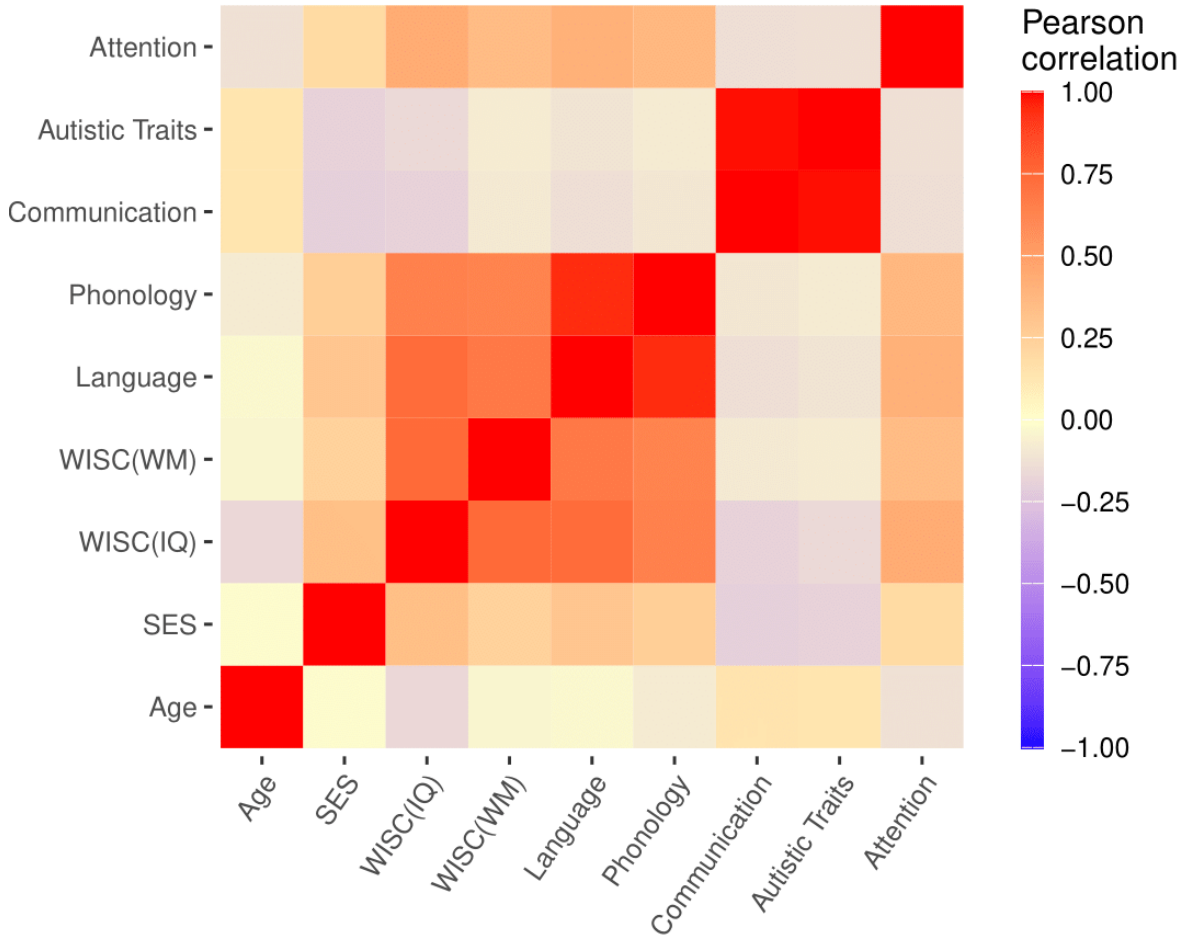
N = 526

Age 7-17

Phenotypic Measures



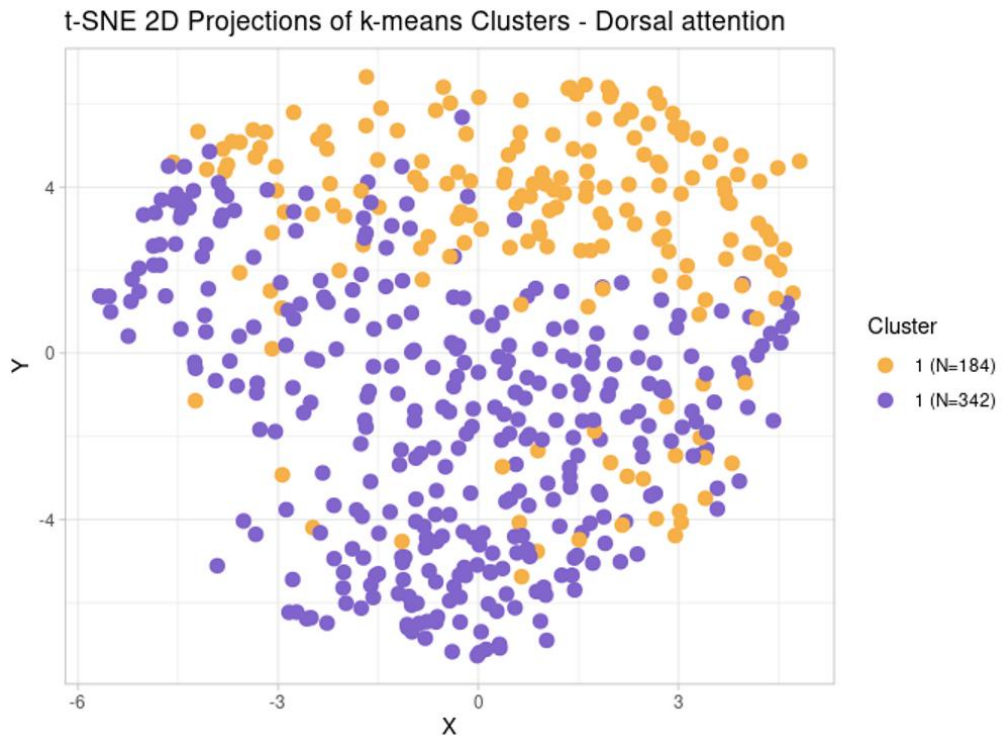
Composite Variables

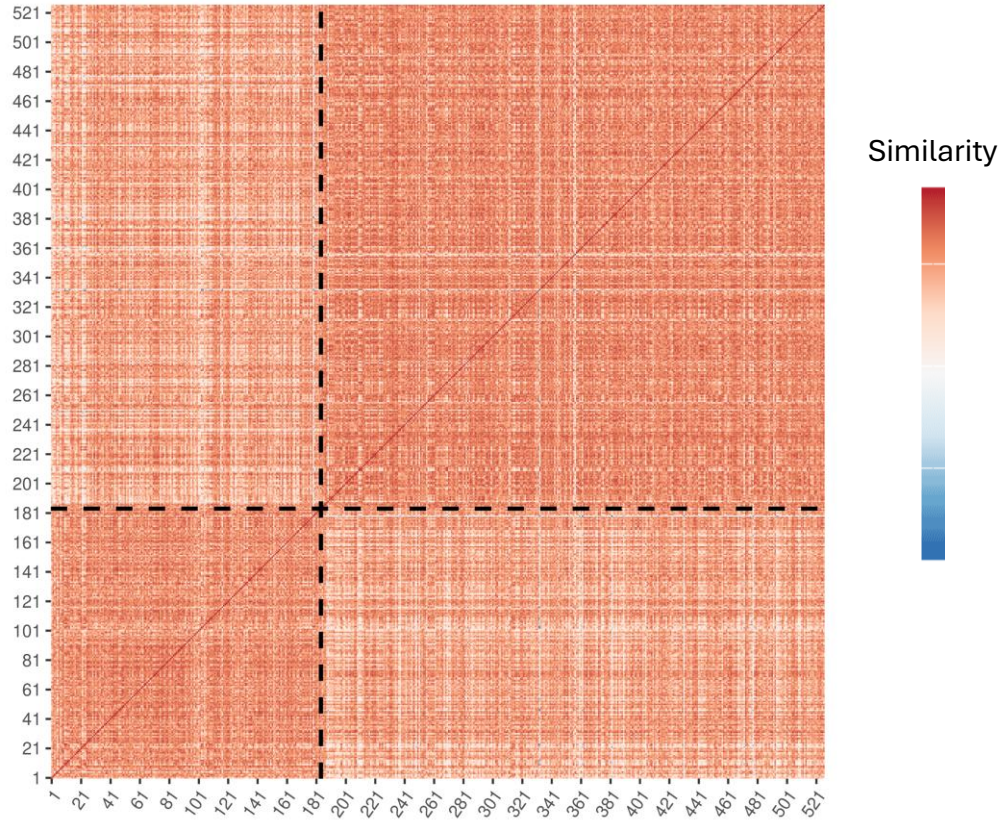


Dimensionality Reduction

24 phenotypic measures → 526 continuous time courses?

24 phenotypic measures → 2-4 “narrative comprehension profiles”

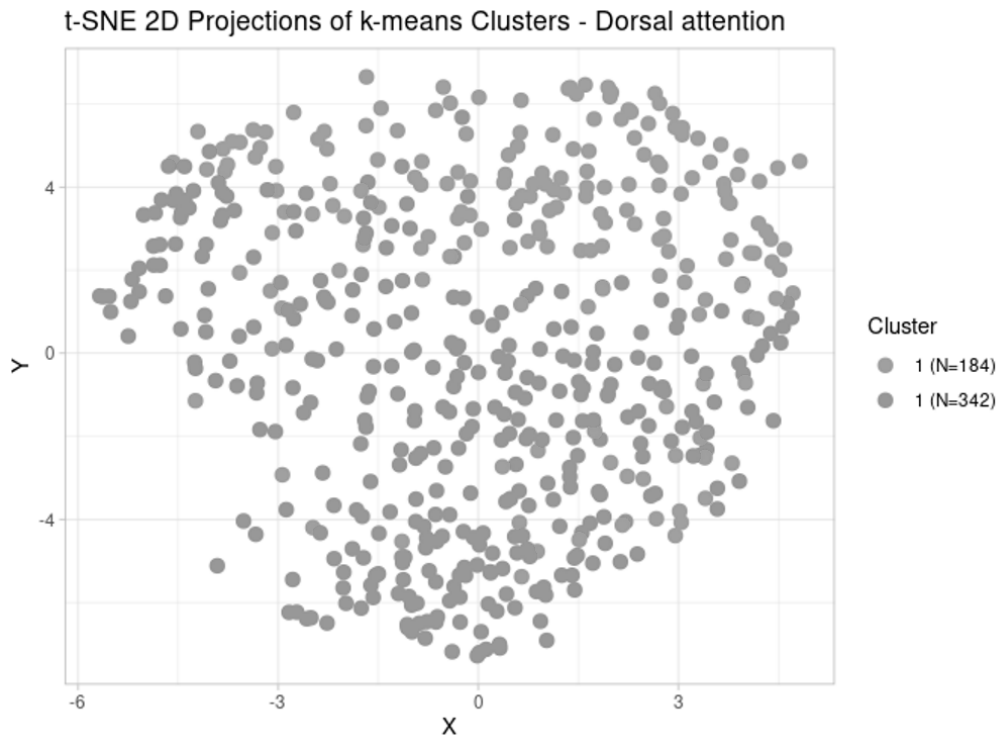


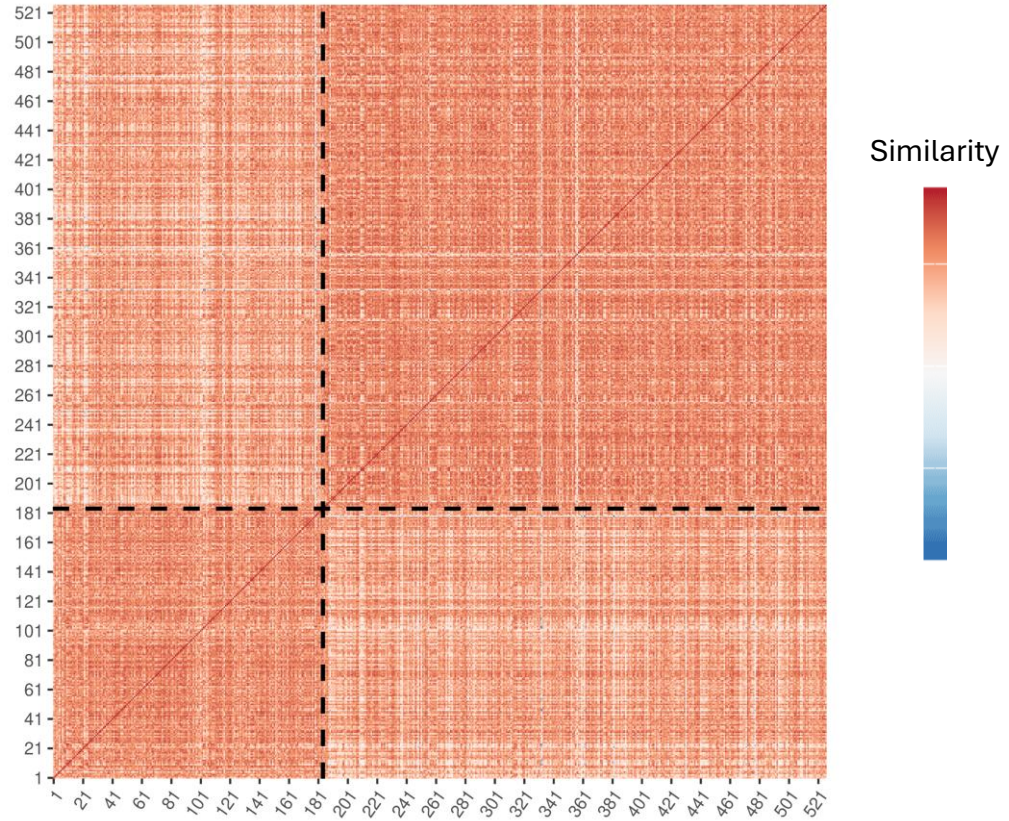


Dimensionality Reduction

24 phenotypic measures → 526 continuous time courses?

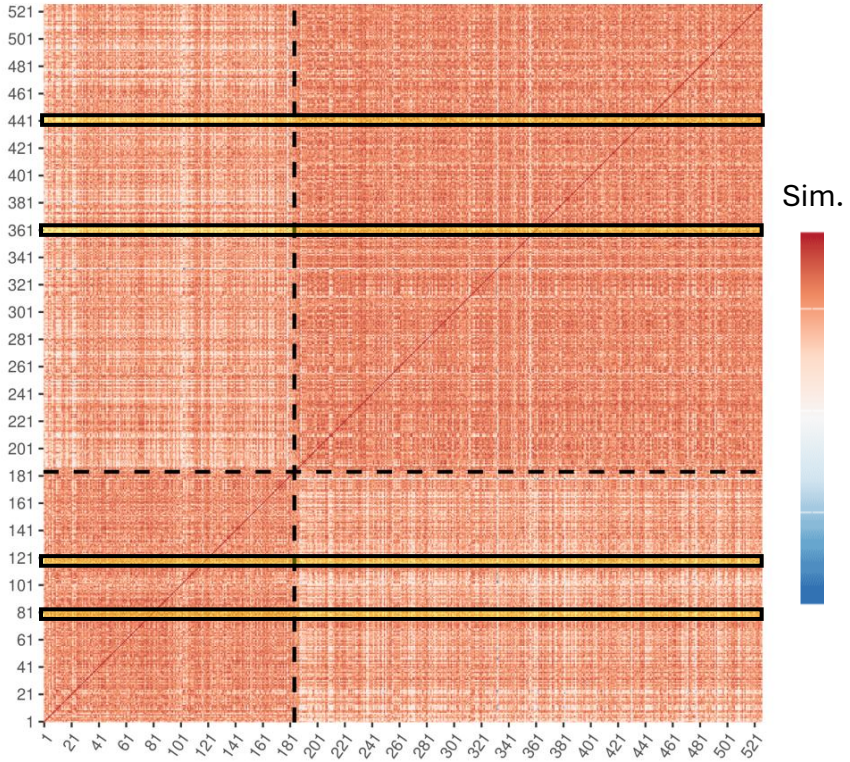
24 phenotypic measures → 2-4 “narrative comprehension profiles”

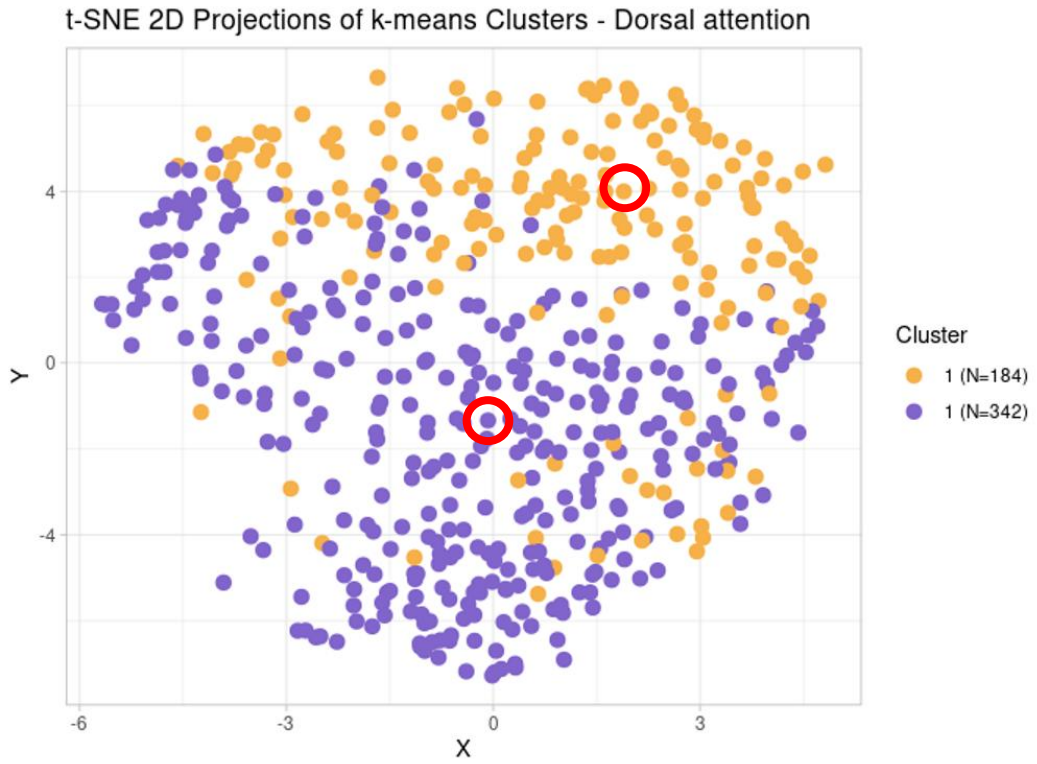




k-medoids Clustering

- medoid = median (actual data point) serves as cluster centroids
- centroid = center of the cluster; the point from which distance to other data is measured
- *k* = number of clusters

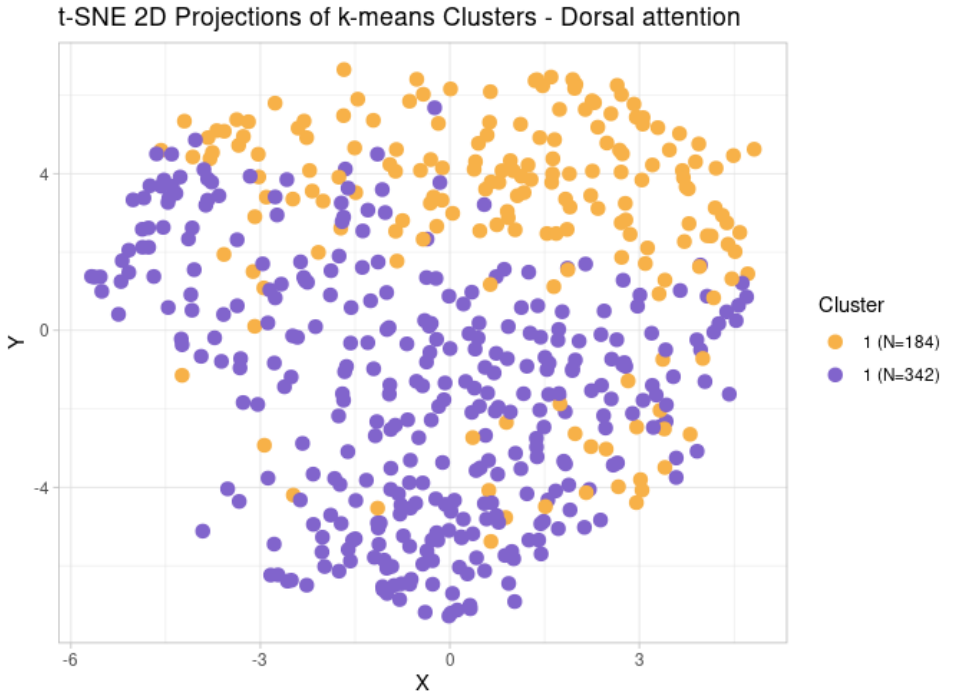


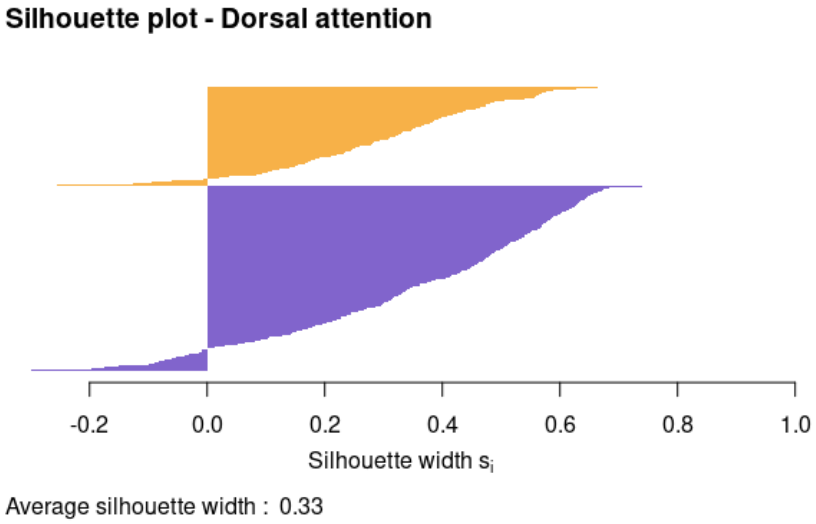


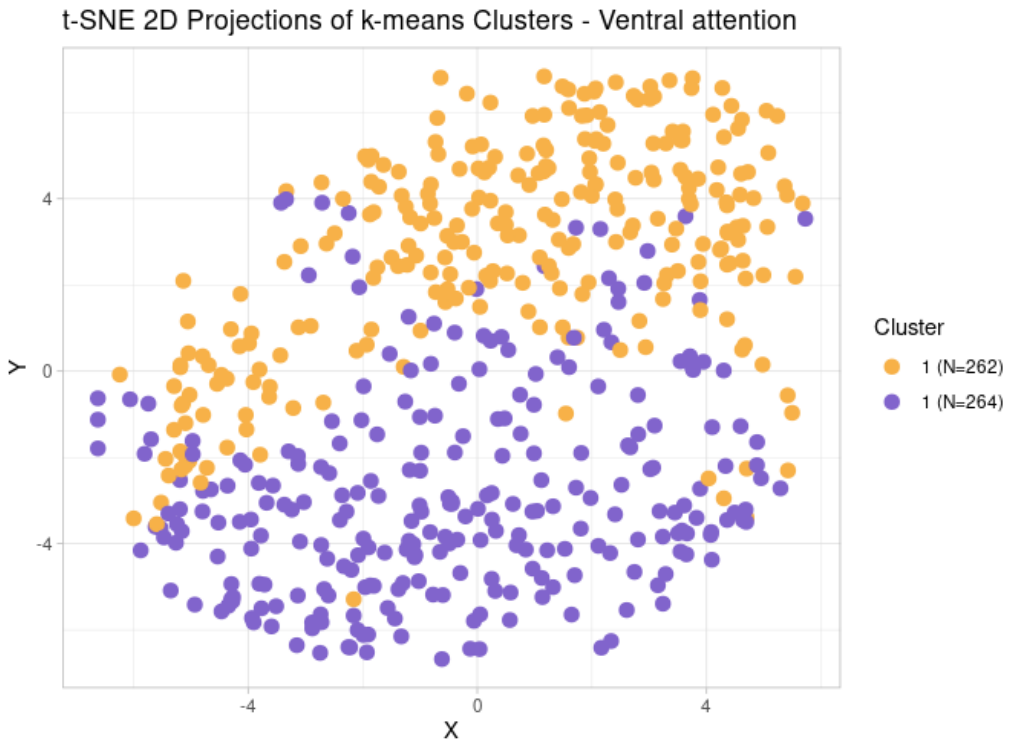
Clustering Results

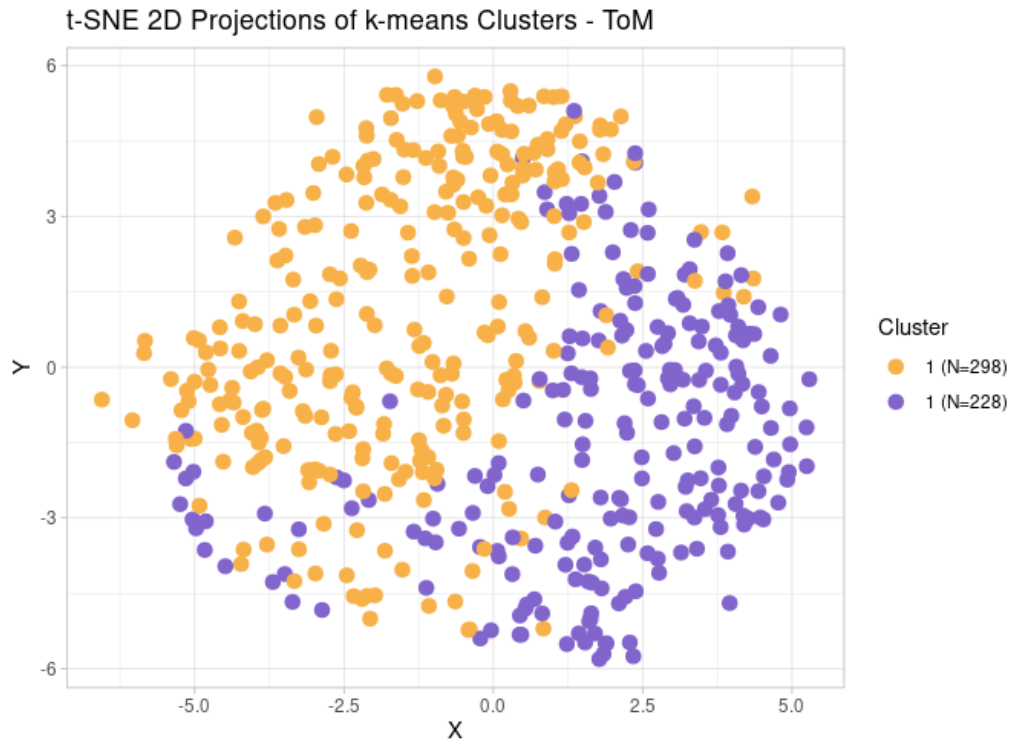
Network	Optimal # Clusters	Average Sil. Width	Cluster Sizes
Dorsal attention	2	0.33	184:342
Ventral attention	2	0.34	262:264
Frontoparietal	4	0.28	91:131:186:249
DMN	3	0.32	149:152:225
ToM	2	0.38	298:228

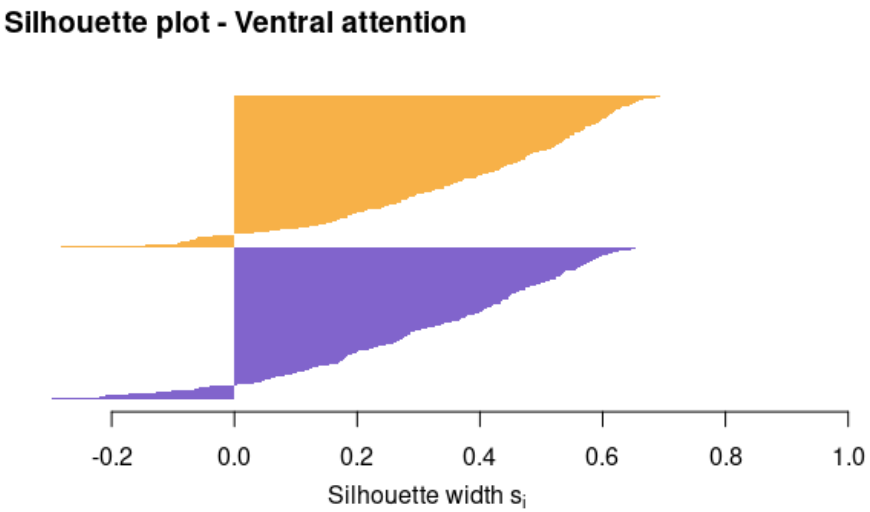
Sil. > 0.7 = “strong”
Sil. = 0.5-0.7 = “reasonable”
Sil. < 0.25 = “weak” **



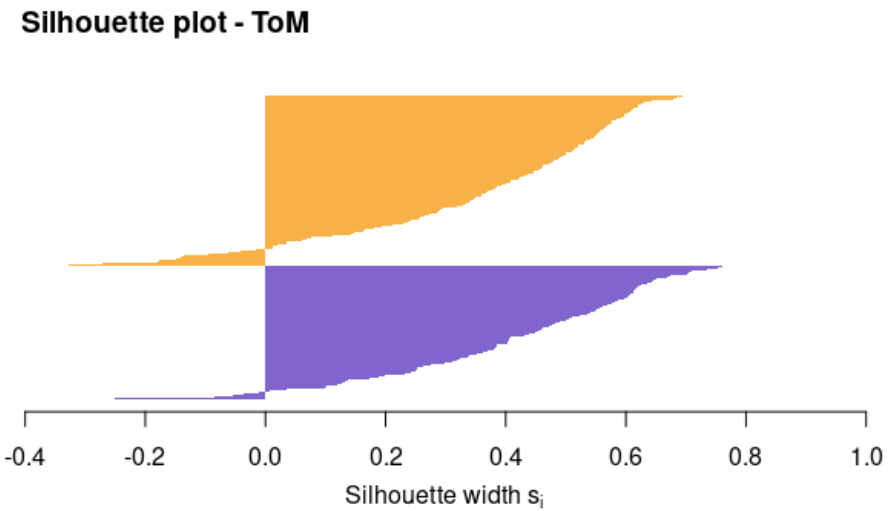




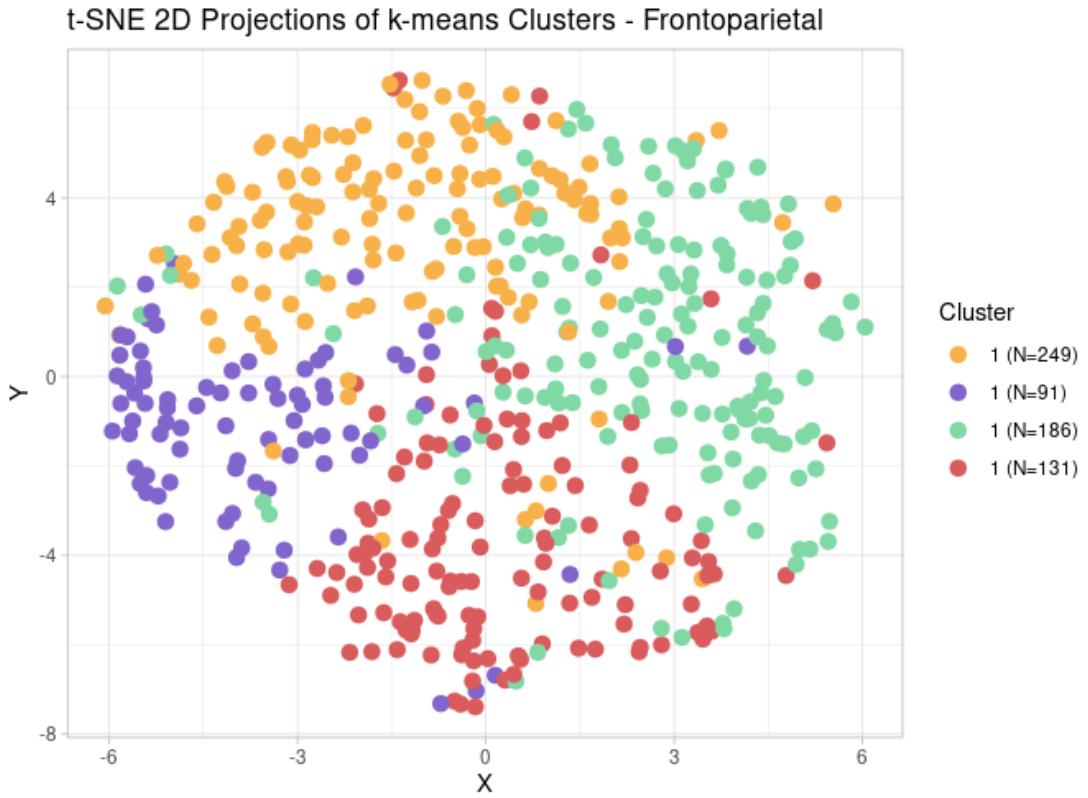


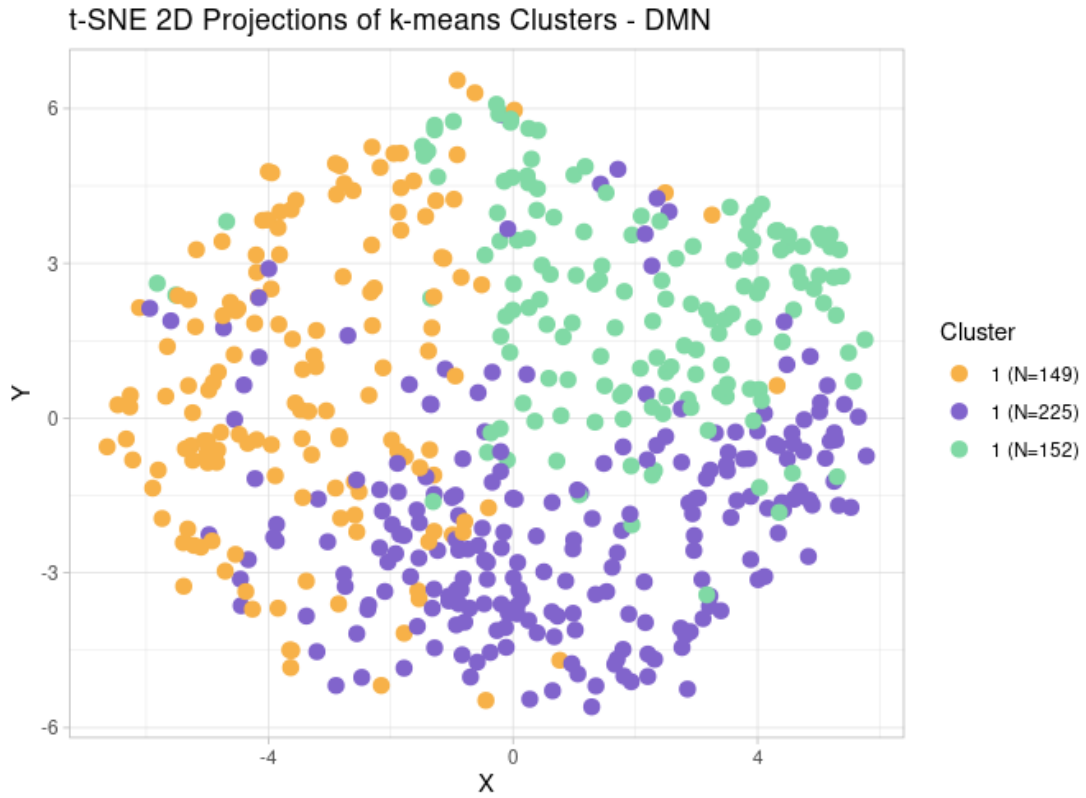


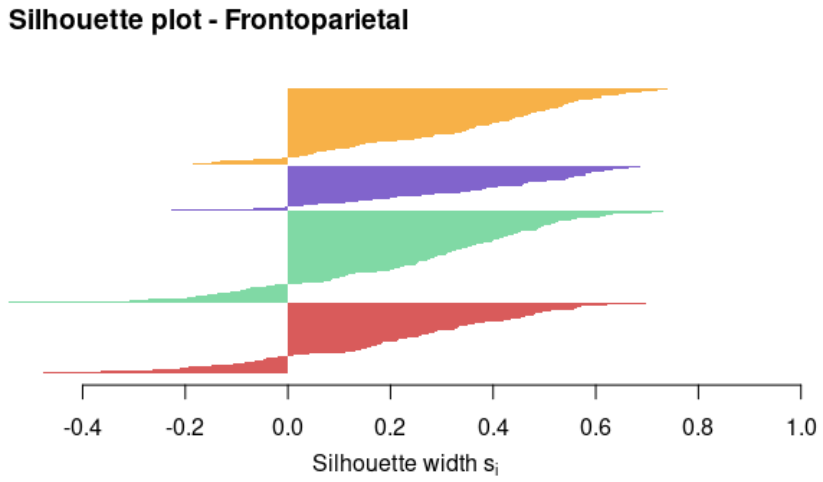
Average silhouette width : 0.34



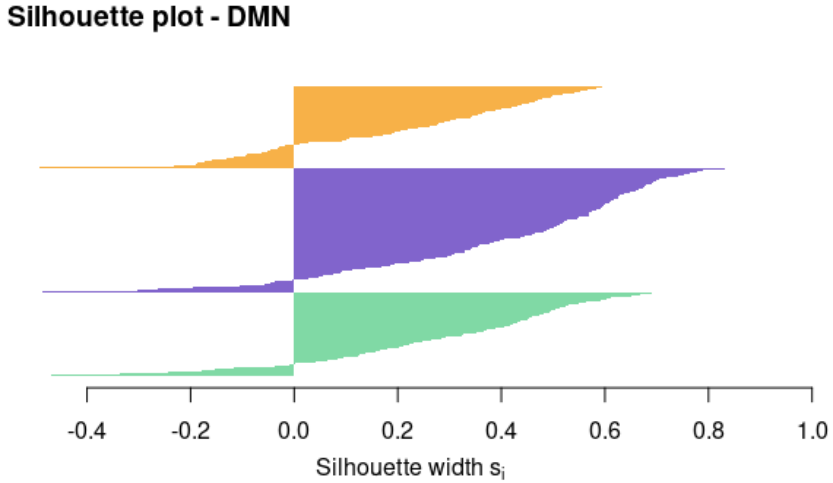
Average silhouette width : 0.38







Average silhouette width : 0.28



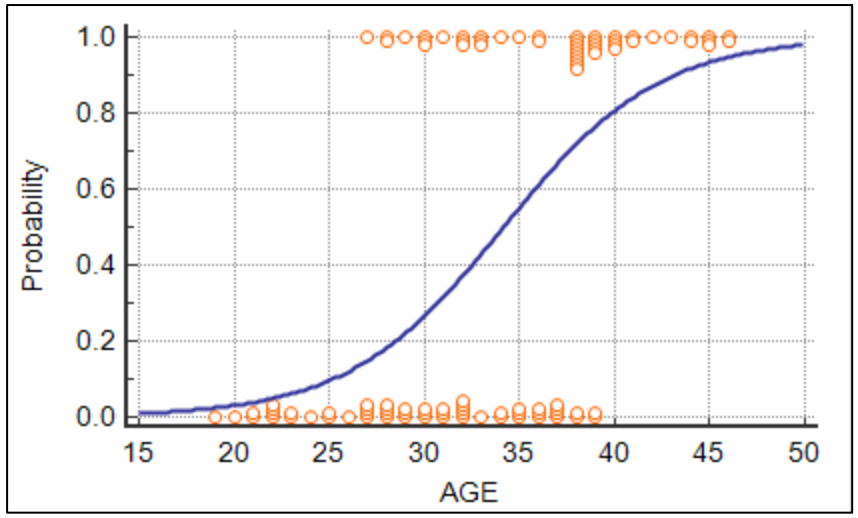
Average silhouette width : 0.32

Predictive Modeling

24 phenotypic measures → 2-4 “narrative comprehension profiles”

Logistic Regression

- Ridge regularization: shrinks all parameters
- LASSO regularization: shrinks some variables to 0 (feature selection)



Model Construction

Input: phenotypic measures

Output: narrative comprehension profile

- **5 models * 8 networks**
  1. LASSO – all measures
  2. Ridge – all measures
  3. LASSO – composite measures
  4. Ridge – composite measures
  5. Ridge – LASSO selected measures

- **80/20 train/test split**
  - Stratified by cluster
  - Optimized hyperparameter lambda: 10-fold cross-validation using training data

- **Result:** classification model with minimized loss/error, coefficient values for variables

Model Performance

all = all measures
comp = composite measures
sel = lasso selected measures

2 Clusters

Network	Model	Sensitivity	Specificity	Precision	F1 score	Balanced Accuracy	Accuracy
Dorsal attn.	LASSO (all)	0.57	0.39	0.33	0.42	0.48	0.45
Ventral attn.	Ridge (comp)	0.58	0.51	0.54	0.56	0.55	0.55
ToM	Ridge (all)	1	0	0.43	0.61	0.5	0.43

F1 > 0.9 = “excellent”
F1 = 0.5-0.8 = “average”
F1 < 0.5 = “poor”

Model Performance

all = all measures
comp = composite measures
sel = lasso selected measures

3 Clusters

Network	Model	Cluster	Sensitivity	Specificity	Precision	F1 score	Balanced Accuracy	Accuracy
DMN	Ridge (sel)	1	0.07	0.83	0.13	0.09	0.45	0.34
		2	0.64	0.26	0.39	0.49	0.45	
		3	0.16	0.84	0.29	0.21	0.50	

F1 > 0.9 = “excellent”
F1 = 0.5-0.8 = “average”
F1 < 0.5 = “poor”

Model Performance

all = all measures
comp = composite measures
sel = lasso selected measures

4 Clusters

Network	Model	Cluster	Sensitivity	Specificity	Precision	F1 score	Balanced Accuracy	Accuracy
Frontoparietal	Ridge (sel)	1	0.98	0.09	0.42	0.59	0.54	0.42
		2	0.045	0.99	0.5	0.08	0.52	
		3	0.09	0.96	0.4	0.14	0.53	
		4	0	1	NA	NA	0.5	

F1 > 0.9 = “excellent”
F1 = 0.5-0.8 = “average”
F1 < 0.5 = “poor”

Future Directions

- Improve model performance
  - Improve clustering?
  - Reduce overfitting?
  - Try different algorithms? (e.g., SVM)

- Obtain confidence intervals for model accuracy
  - Nested cross-validation

- Explore different permutations of data
  - E.g., within-subject functional connectivity, leave-one-out ISC

Links (Part 2)

R code: pre-processing (phenotypic)
https://github.com/jmlammert/HBN_methods_lunch/blob/c5b9ed9cf31e33e879bed14e97efdd4d4ed5ccb9/R%20code/preprocess_MRI-behav_v5.R

R code: pre-processing (time series)
https://github.com/jmlammert/HBN_methods_lunch/blob/c5b9ed9cf31e33e879bed14e97efdd4d4ed5ccb9/R%20code/preprocess_MRI-timecourse_v3.R

R code: descriptive statistics
https://github.com/jmlammert/HBN_methods_lunch/blob/c5b9ed9cf31e33e879bed14e97efdd4d4ed5ccb9/R%20code/descriptives_MRI-behav_v3.R

R code: clustering
https://github.com/jmlammert/HBN_methods_lunch/blob/c5b9ed9cf31e33e879bed14e97efdd4d4ed5ccb9/R%20code/analyses_exp1_v9.R

R code: logistic regressions
https://github.com/jmlammert/HBN_methods_lunch/blob/c5b9ed9cf31e33e879bed14e97efdd4d4ed5ccb9/R%20code/regressions_exp1_v5.R

Questions & Discussion