# Association Between Mice Connectome and Behavioral Indices

Youngsoo Baek, Yunran Chen, and Brian Kundinger

# Background

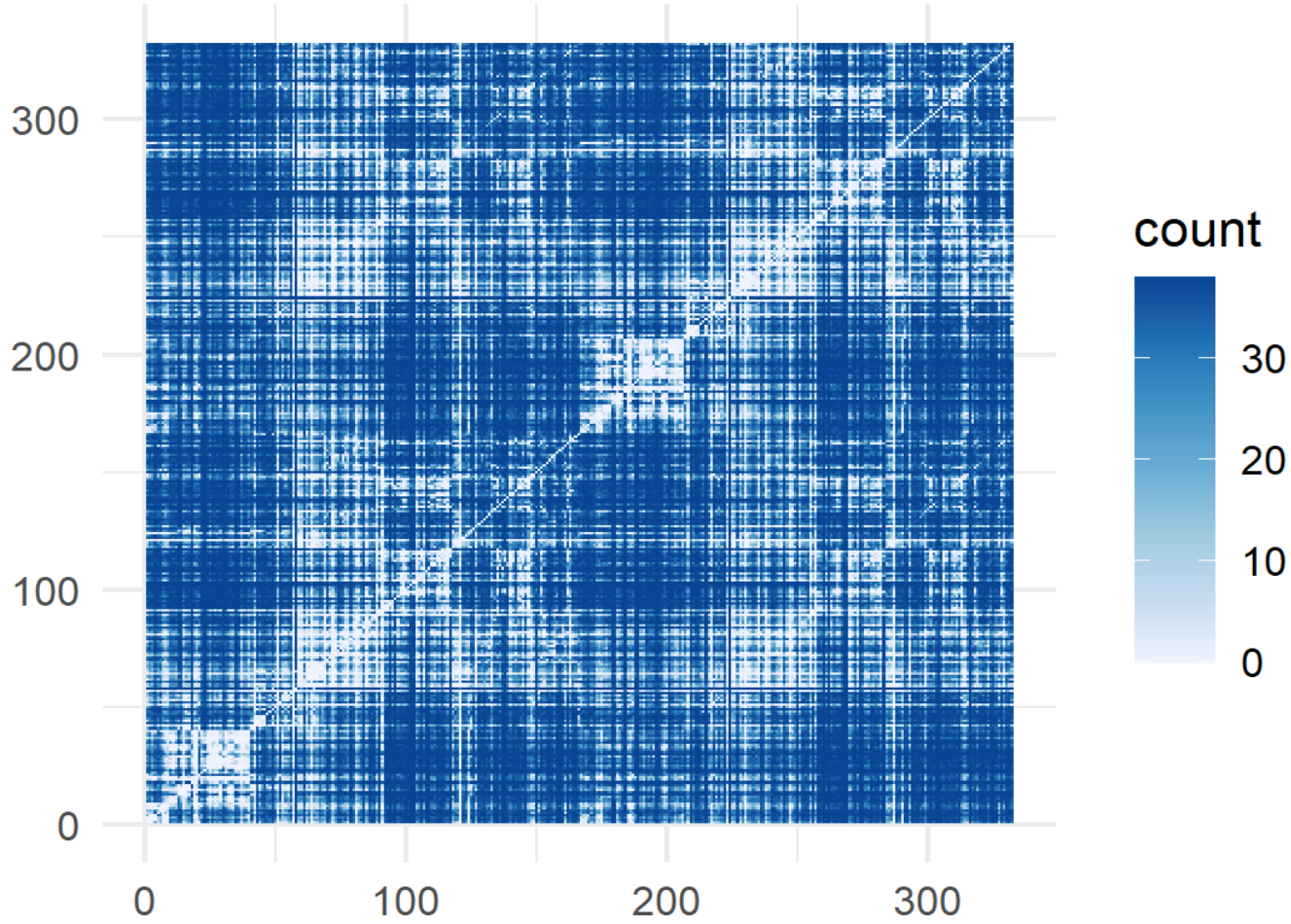
- Behavioral variables of interest mainly include `NormSWTime` (normalized time to reach target) and `RI_T2`, `RI_T3` (recognition indices)
- **Connectomes** are a collection of white matter fiber tracts that connect different regions of the brain.
- Can be thought of as a weighted undirected graph with 332 vertices
- Goal: Study the possible relationship between behavioral variables and mice connectomes

# Data Processing

- Need to match mice across different datasets (Connectomes, NOR, MWM)
- MWM has several mice with multiple "runno" identifiers (eg N54716/N54915), but most of these can be matched with unique connectome
- One mouse "N54891/N54900LRspecific" and another with runno "NA" are unidentifiable, and removed from analysis.
- Once Connectomes are matched with entries in MWM, we use Animal ID to match with NOR
- After cleaning, we conduct analysis with 38 mice

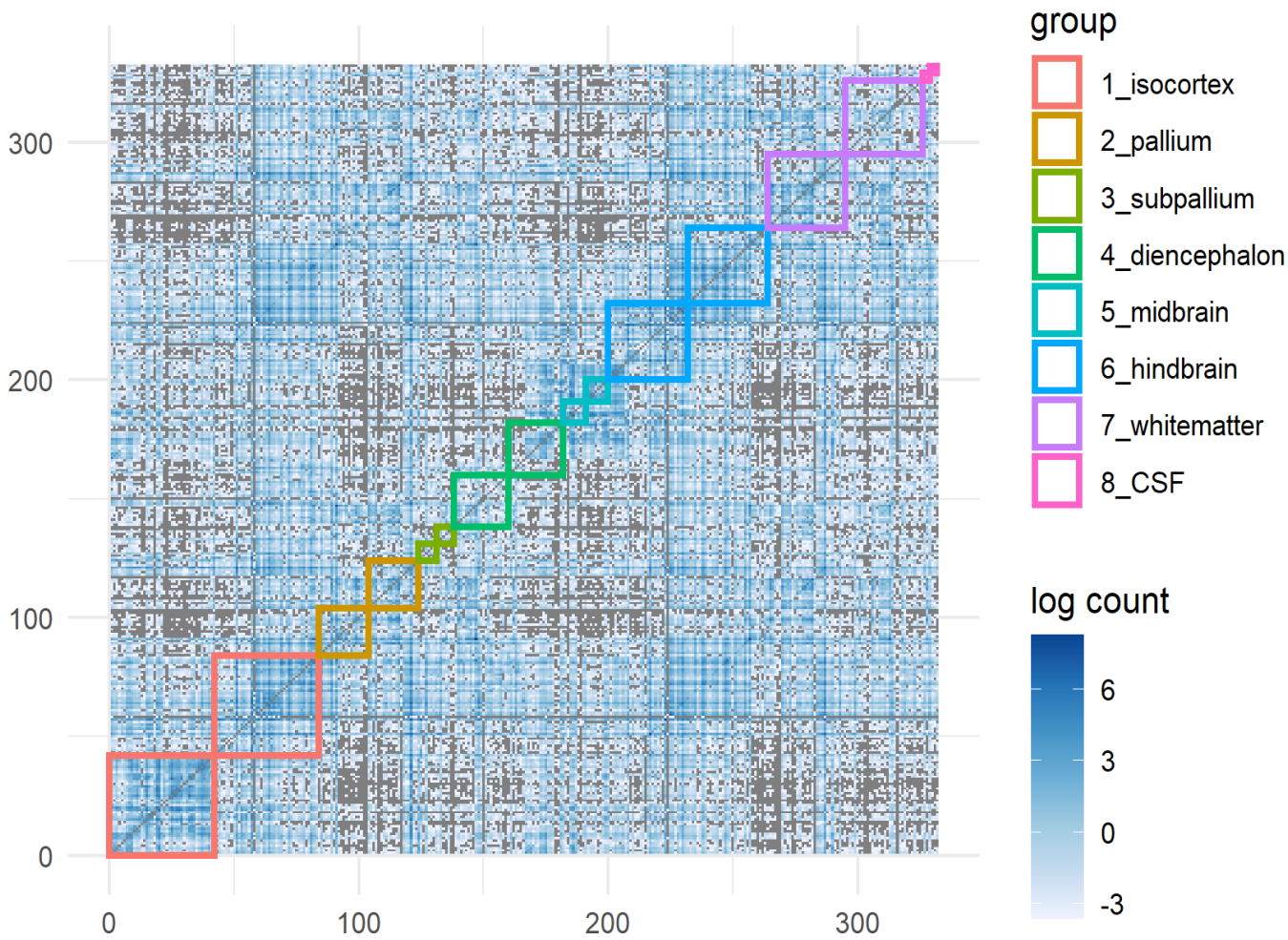
# Visualizing Connectomes

- Sparsity of connectomes + small between-mice variation regarding sparse entries



# Visualizing Connectomes

- Apparent clusterings of graph nodes do not coincide with known compartments



# Approaches to Analysis

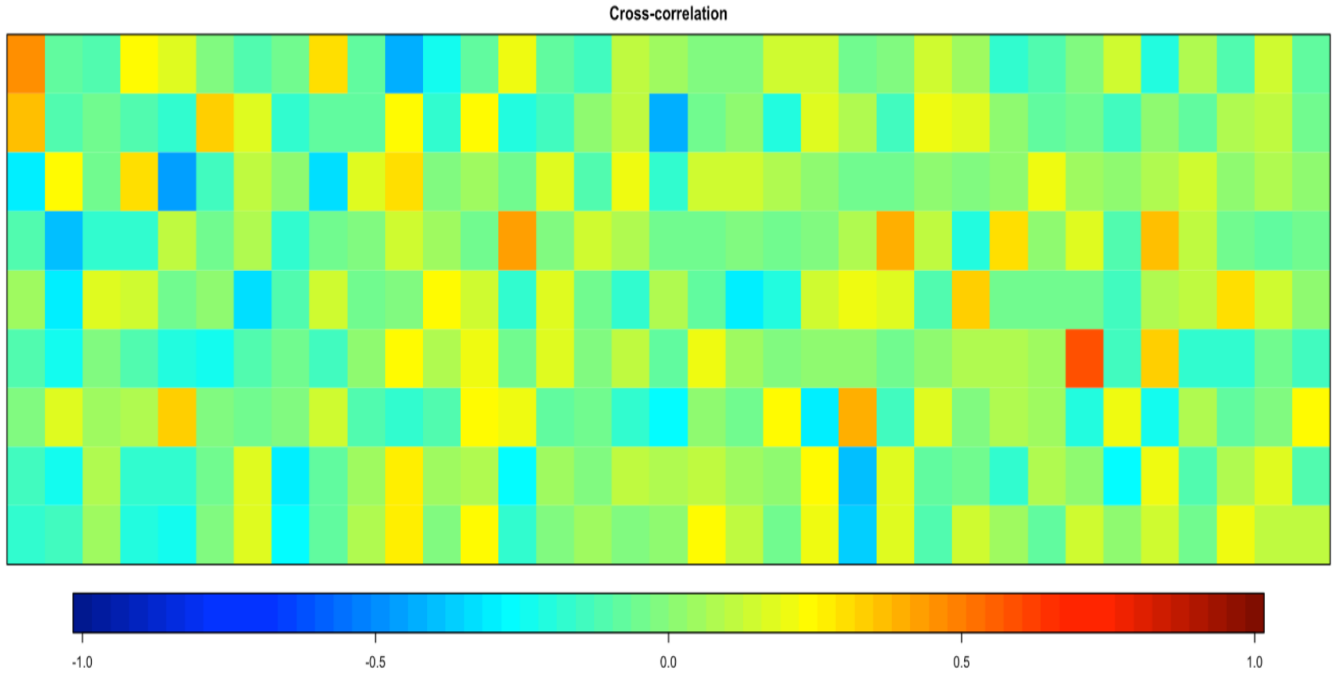
- Exploratory techniques to explore connectomes matrices in a lower dimension $\Longrightarrow$ *PCA* and *CCA* (canonical correlation analysis)
- Clustering, both between graphs and of nodes within graphs (*community detection*) are important ways to summarise connectomes
- *Beta regression* to study association, both within behavioral variables and between behavioral variables and summaries of connectomes

# Canonical Correlation Analysis (CCA)

- Analogue to PCA in high-dimensional settings
- Goal: For $X_1, \ldots, X_n \in \mathbb{R}^p$ and $Y_1, \ldots, Y_n \in \mathbb{R}^q$, want to estimate the covariance matrix $\Sigma_{X,Y}$
- Through SVD, find vectors $a \in \mathbb{R}^q$ and $b \in \mathbb{R}^p$ maximizing $\mathrm{Corr}[a^T X, b^T Y]$ subject to orthogonality constraints

# PCA and CCA

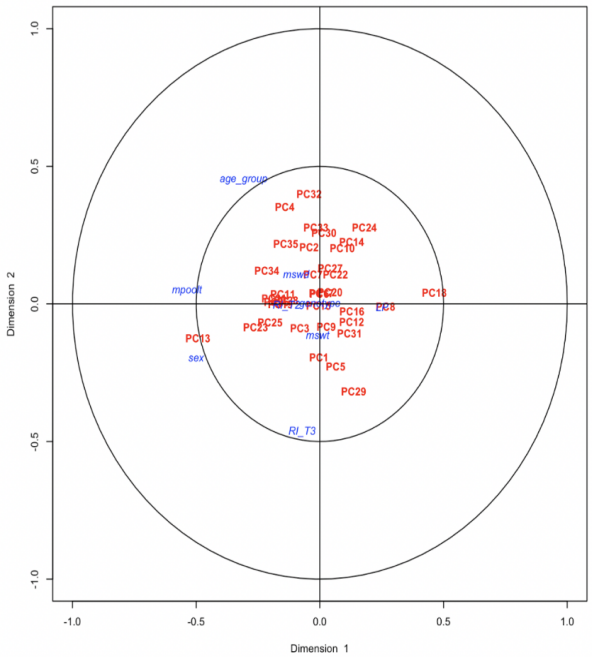
- Canonical Correlation Analysis (CCA): Explore how brain connectomes and behaviorial traits co-vary in a similar way
- Preprocessing using PCA: keep the within subdivision connectomes ($38 \times 2107$) and apply PCA to reduce dimension to ($38 \times 35$)
- There may exist correlation between PCs and traits.

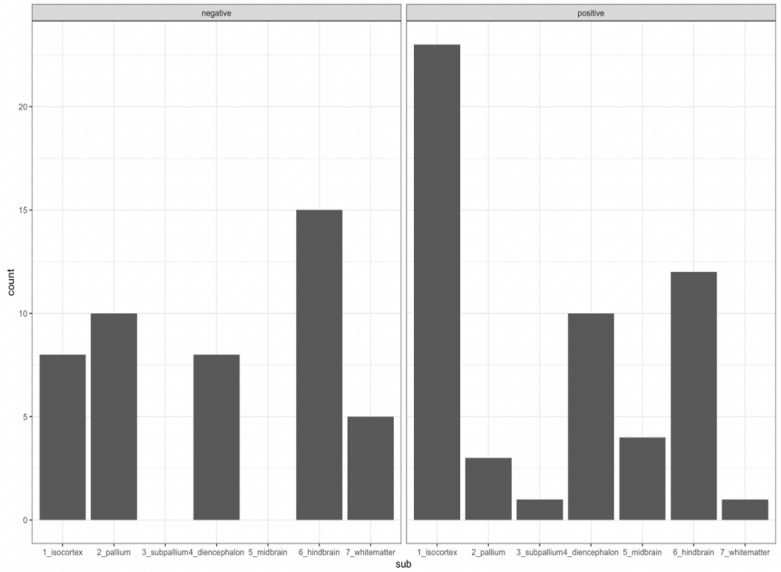


Youngsoo Baek, Yunran Chen, and Brian Kundinger

# CCA Results

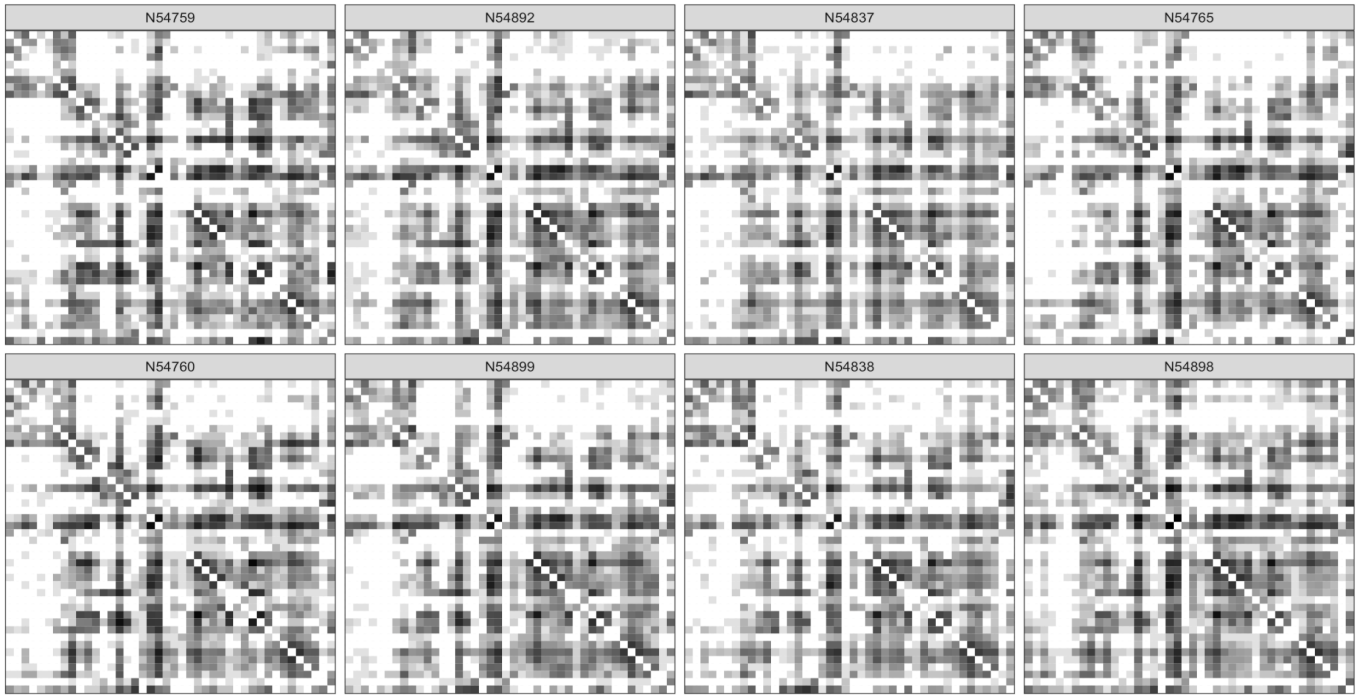
- (Left) Correlation between traits/PCs and first/second CCA mode
- (Right) The subdivision of top 100 ROIs loading in PC13
- `pooltime` may co-vary with PC13
- `pooltime` may mostly relate to the isocoritex region positively, relate to the hindbrain region negatively



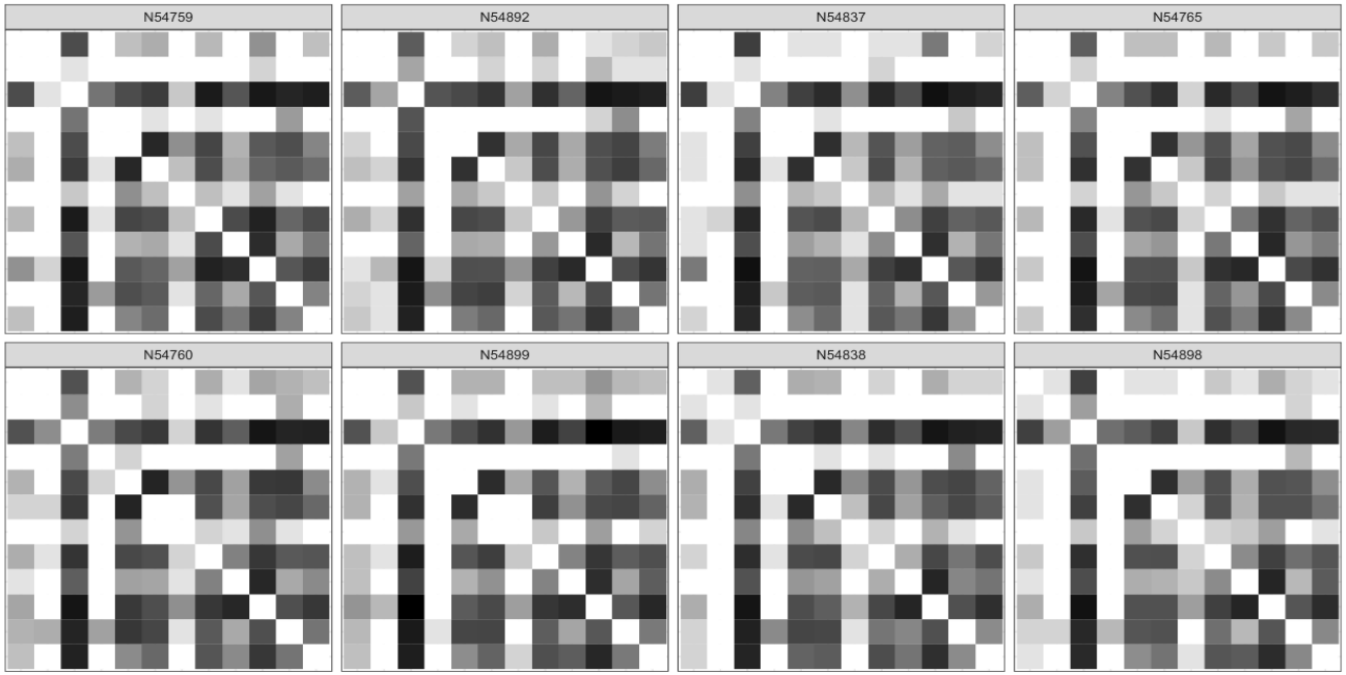
# Pool time and Connectomes in isocortex

- Female, old, genotype(0), arranged based on values of small pool time (from small to large)
- More active connections between ROIs may relate to larger pool time.



# Pool time and Connectomes in hindbrain

- Female, old, genotype(0), arranged based on values of small pool time (from small to large)
- Less active connections between ROIs may relate to larger pool time.



# Method: Beta Regression

- Suppose $Y_i \sim Beta(a, b)$. Then

$$f(y|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1}$$

- We reparameterize for mean $\mu = \frac{a}{a+b}$, and precision $\phi = a + b$

$$f(y|\mu,\phi) = \frac{\Gamma(\mu)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}$$

- Then we model $g(\mu_i) = x_i^T\beta \implies \mu_i = g^{-1}(x_i^T\beta)$ for some link function $g : (0,1) \to \mathbb{R}$.

# Method: Beta Regression

- Beta regression also allows us to model heteroskedastic variance

$$g_2(\phi) = z_i^T \gamma$$

for covariates $z_i$ and regression coefficient vector $\gamma$

- Convert discrete "Day" variable into continuous "time" variable to capture change in variance over the course of the experiment

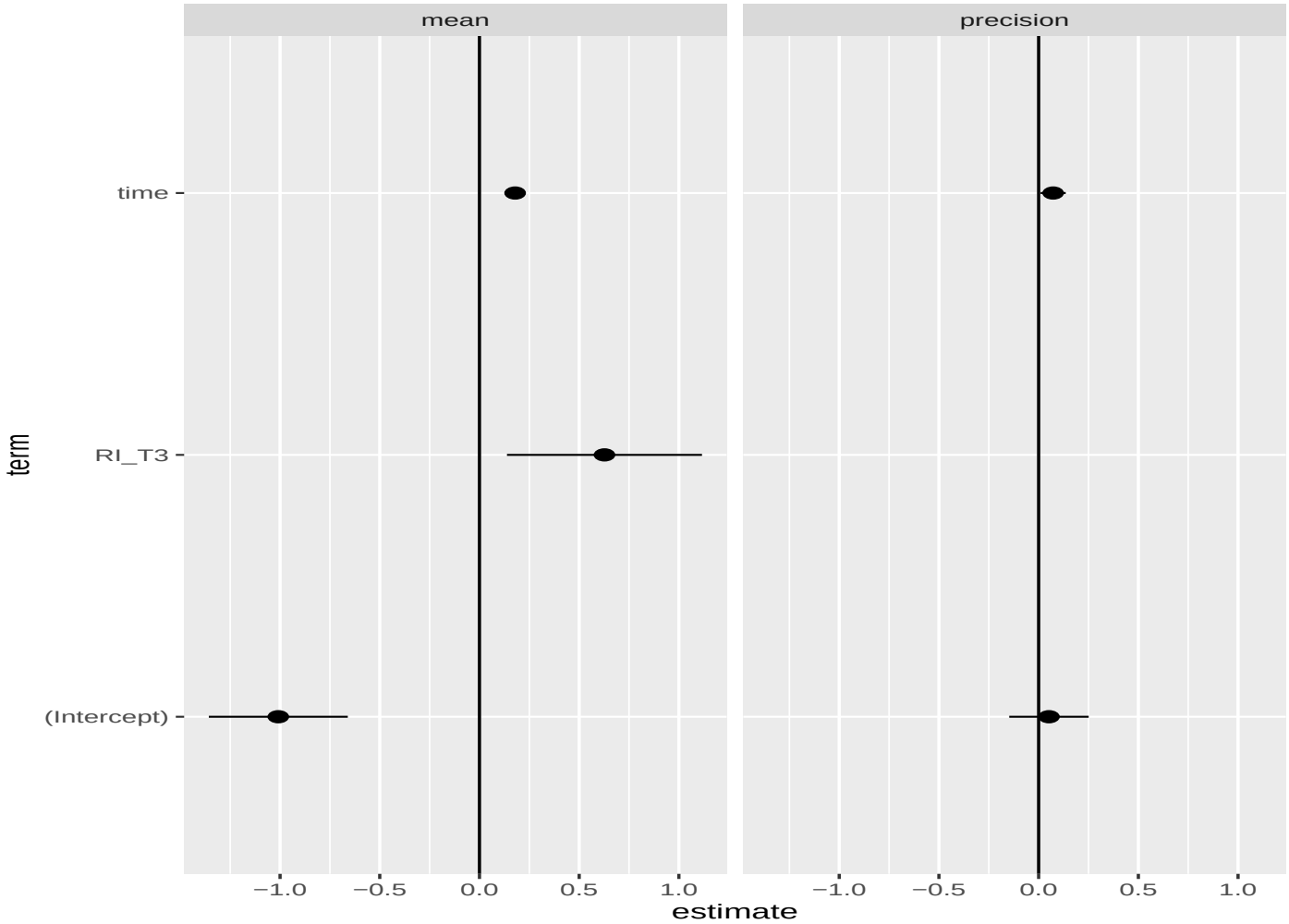
# Method: Beta Regression (NormSWTime)

- When modeling NormSWTime, many values take on 0 or 1, which causes problems (think of logit function)
- No continuous mapping from $[0, 1]$ (Closed Set) $\rightarrow \mathbb{R}$ (Open Set)
- Remove observations where Pool Time $< 60$ but NormSWTime $= 0$
- Remove probe trials, all have Pool Time $= 60$ (not useful)
- Remove mouse with zero RI on trial 3
- Apply the following transformation for the remainders (Smithson and Verkuilen, 2006):

$$y^* = \frac{y(n-1) + .5}{n}$$

# Beta Regression: Results

$$\text{Formula : NormSWTime} \sim \text{time} + \text{RI}_{\text{T3}}|\text{time}$$



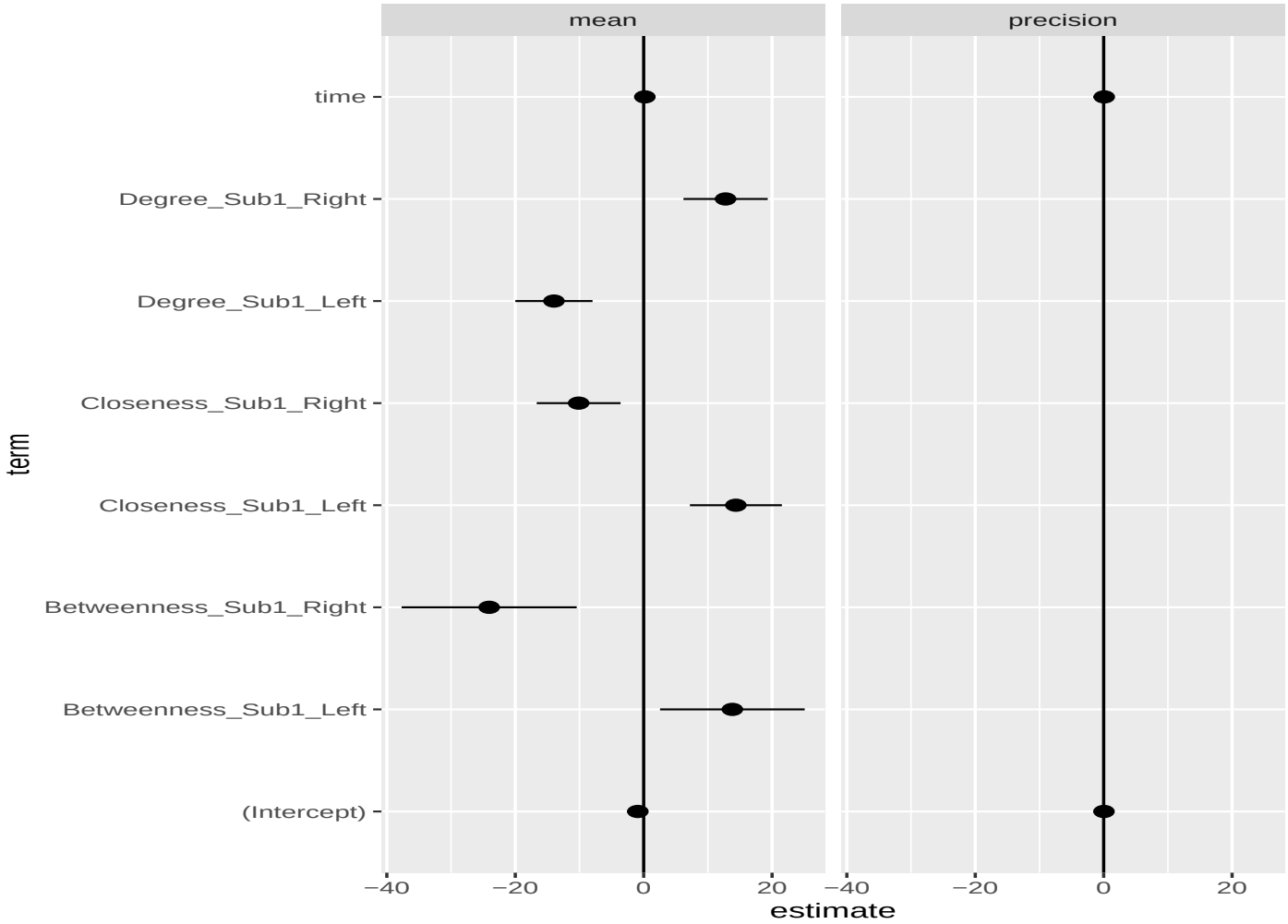
# Beta Regression: Results

| component | term | estimate | p.value |
|---|---|---|---|
| precision | (Intercept) | 0.0517465 | 0.6033946 |
| precision | time | 0.0728208 | 0.0208468 |

Youngsoo Baek, Yunran Chen, and Brian Kundinger

# Beta Regression: Results

NormSWTime ~ time + Isocortex Summary Statistics |time



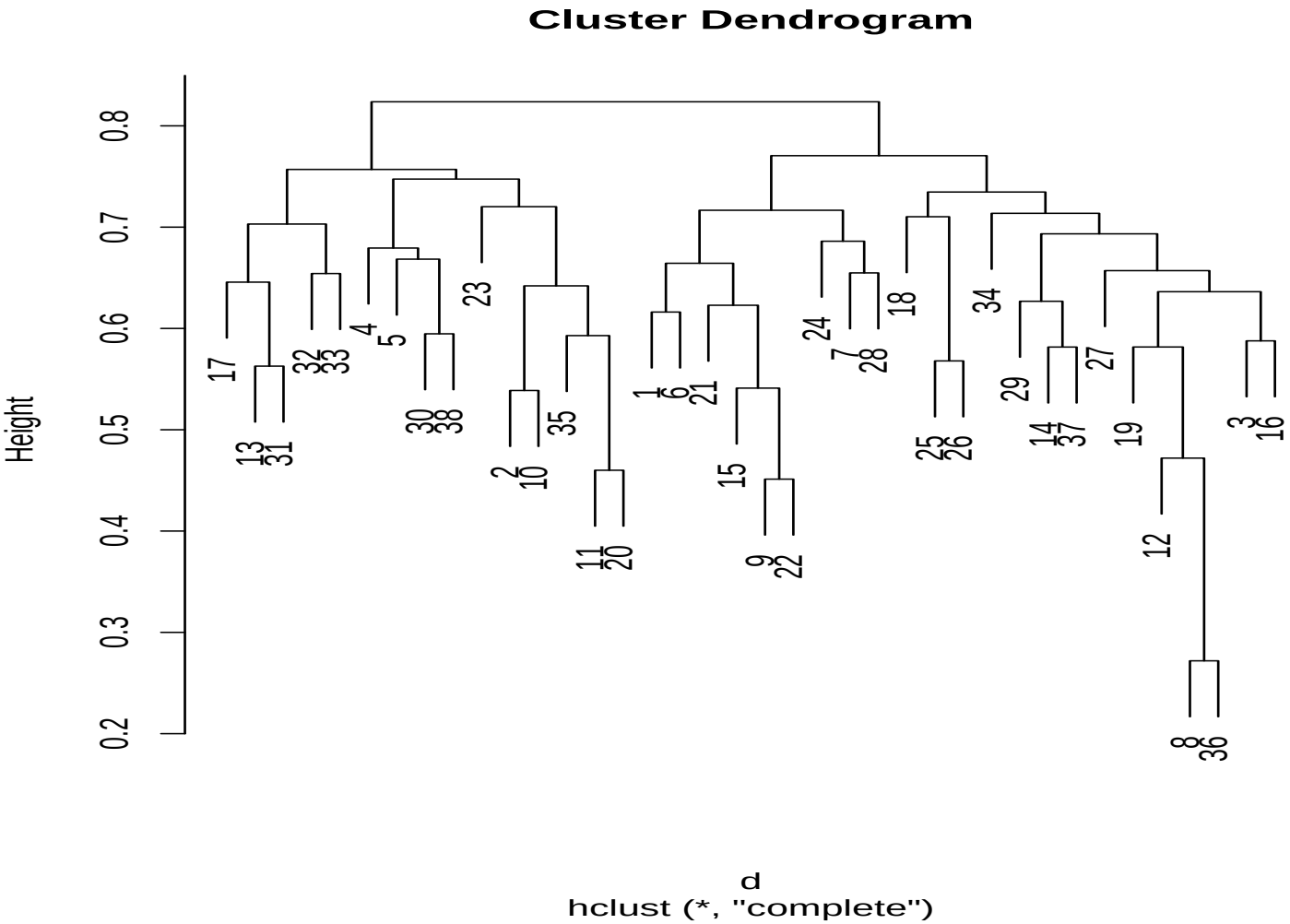
# Beta Regression: Remarks

- Need more accurate clustering tool and knowledge of network summary statistics for this method to provide meaningful results
- Increase in response variable precision over time indicates all have more similar NormSWTime at the end than at the beginning
- Perhaps Connectome effects might not increase with time, but rather decrease with time

# Mice Grouping

- Explored methods to cluster the different mice connectomes (graph clustering), and to cluster the vertices within a graph (community detection)
- Rather than identify clusters within the networks, we attempt to cluster the mice by characteristics in the connectome data
- Identify clusters within each connectome through Louvain community detection
- Calculate pairwise Normalized Mutual Informormation (NMI, measure of similarity between two network community structures), compile into "distance matrix"
- Use hierarchical clustering to divide mice into 4 groups. Use group membership as regressors.

# Mice Grouping




- Method not successful, NMI scores are too similar across the mice.

# On Community Detection

- Standard search algorithms (e.g., `igraph`'s greedy search based on modularity) often yields unsatisfying results
- Cannot merge information from different mice into clustering

# On Community Detection

- A simple mixture membership model (Newman and Leicht, 2007) can detect latent structures that do not necessarily align with the known compartment
- EM algorithm can be used to estimate the posterior modes / MLEs for membership probabilities ($\pi_r$), and connection probabilities between each community $r$ and node $i$ ($\theta_{ri}$)
- **Problem**: EM implementation became numerically unstable and too sensitive to initialization

# Beta Regression of Behavioral Responses on Connectomes

- For now, simply sum over the fiber counts belonging to the same compartments (332-dim. $\rightarrow$ 8-dim.)
- Stack the diagonal / lower triangular entries of connectomes matrices, which become our covariates ($p = 8(8+1)/2 = 36$)
- Taken apart separately, no individual compartment entries seem to have significant association with behavioral variables
- **Problem**: How to systematically induce strong shrinkage?

# Discussion

- Exploring ways to better summarise connectomes will be interesting with more time
- Connectomes matrices may serve as better covariates for modeling second order structure of different response variables
- More structure that induces strong shrinkage is needed in tensor regression model

# Reference


- "Mixture models and exploratory analysis in networks," Newman, M. E. J. and Leight, E. A. *PNAS*, 104(23), 2007.
- Smith S M, Nichols T E, Vidaurre D, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior[J]. Nature neuroscience, 2015, 18(11): 1565.
- "A better lemon squeezer? Maximum-likelihood regression with Beta-distributed dependent variables," Smithson, M. and Verkuilen, J. *Psychological Methods*, 11, 2006.