Canonical Correlation Analysis on Functional Connectivity of Human Connectome Project

Coleman Ferrell and Malavika Mampally
STOR 674 Final Project

1 Introduction

The study of brain connectomes, which encompasses both structural connectivity (SC) and functional connectivity (FC), has gained significant attention due to its potential to uncover relationships between the human brain and various traits. Structural connectivity examines the physical architecture of neural pathways, while functional connectivity focuses on the temporal correlations of brain activity across different regions. Together, these perspectives offer a comprehensive understanding of the brain's interconnected networks.

There is a growing demand to investigate how SC and FC are linked to human traits such as cognitive abilities, psychiatric assessments, and other behavioral dimensions. Advances in imaging techniques such as diffusion Magnetic Resonance Imaging (dMRI) and functional Magnetic Resonance Imaging (fMRI) have enabled the extraction of rich connectivity data, facilitating these explorations.

Building on prior research, including work by Kessler and Levina [2023], which utilized Canonical Correlation Analysis (CCA) to study the relationships between FC and behavioral performance, we aim to replicate and extend their findings. The focus is to test the connections between FC data and cognitive, psychological, and other behavioral traits in a more comprehensive manner.

For this study, high-quality SC and FC data were sourced from the Human Connectomes Project (HCP), leveraging a large dataset of approximately 1206 subjects. Dimensionality was reduced using the Desikan-Killiany Atlas, which divides the brain into 68 parcels. This approach ensures that the analysis remains both computationally efficient and biologically meaningful.

2 Methodology

2.1 Datasets

The study involves the use of three main datasets:

Traits: This contains 175 traits of about 1065 subjects that measure a person's cognition, substance use, psychiatric and life function, sensory, emotion, health and family history.

Functional Connectomes: This contains the connectomes for 68 brain regions based on the Desikan-Killiany Atlas. For each subject, a 68 X 68 matrix with functional connectivity (FC) for each pair. The extraction of FC was based on Zhang et al. [2019] Since this matrix is naturally symmetric. The upper (or lower) triangle data contains all the information needed for each subject.

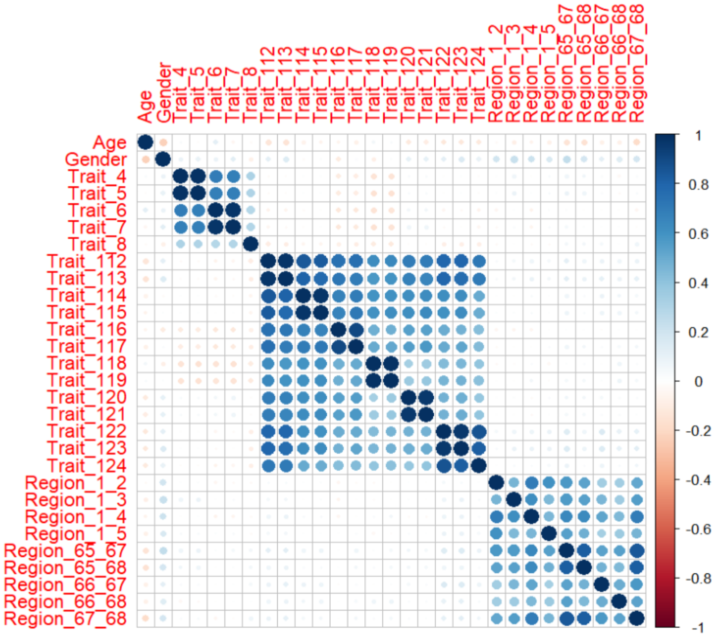


Figure 1: Correlation Heat map

Confounding Varibles: Variables that can potentially add bias or unwanted effect on the response variable. This include age and gender of the subjects.

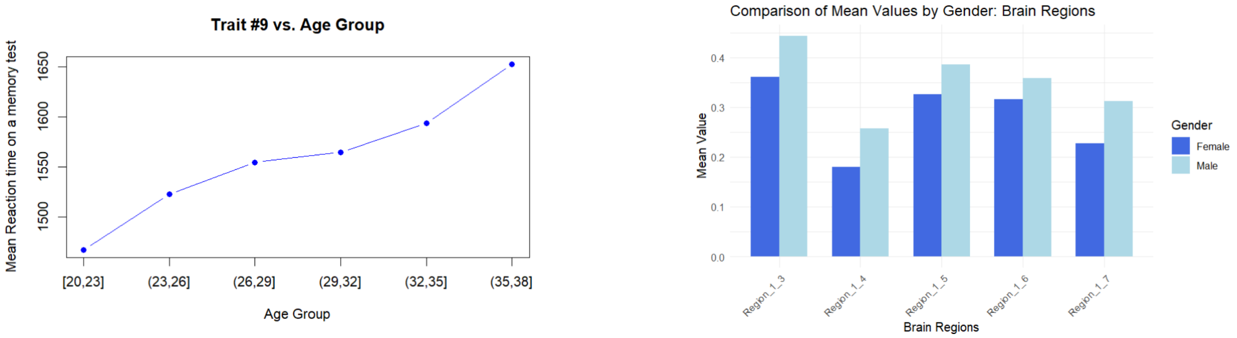
2.2. Exploratory Data Analysis

To gain an initial understanding of the relationships within the dataset, we conducted a correlation analysis. This involved generating a heatmap to visualize the correlations among the traits and brain regional correlation. Due to high dimensionality of the data, we select a set of traits and brain regions to visualize easily. Figure 1 shows the corresponding heat map.

Traits exhibited strong intra-group correlations, indicating potential redundancies or shared underlying factors. Brain regions also demonstrated high correlations within their group, suggesting interdependencies in their structural or functional roles. This analysis provided a foundation for selecting features and guided subsequent steps in the modeling process.

Age and gender were identified as potential confounding variables, as they can influence both brain development and cognitive traits. To quantify their effects, a formal correlation analysis and ANOVA were performed to assess the significance of age and gender across traits and brain regions. The level of significance for these tests was set to 0.05 initially and can be modified based on the results desired.

There tests revealed brain regions to have 1,603 significant correlations with age and 1,966 with gender. It also revealed traits to have 49 significant correlations with age and 67 with gender. These findings underscored the necessity of controlling for these variables to ensure that the results focused solely on meaningful relationships between traits and brain regions. These effects can also be visualized in the following figures.



(a)

(b)

Figure 2: (a) The plot illustrates the mean reaction time on a memory test across different age groups.(b) This bar plot shows the mean values of FC Correlation for different brain regions grouped by gender

In Figure 2a the mean reaction time increases progressively with age (almost linearly) suggesting a potential correlation between age and memory-related reaction time. Similarly in Figure 2b, the FC correlation between the region pairs (1,3), (1,4), (1,5), (1,6) and (1,7) display a clear distinction between the two genders, with male always being higher, suggesting potential gender-based variations in brain region characteristics.

By systematically addressing confounding variables and ensuring a clean dataset, this methodology established a solid foundation for the subsequent multivariate analysis. In this paper, we chose to carry out Canonical Correlation Analysis (CCA) due to its ability to capture the bidirectional relationships between two sets of variables. This approach is particularly useful when either set of variables can serve as both the response and the explanatory variable. For instance, a researcher might aim to identify behavioral traits that predict brain structure, while simultaneously exploring how brain structures can predict behavioral traits. Because of these academic questions, performing an analysis with an emphasis on correlation is desired. This naturally leads to CCA, which is a common practice when handling complex neuroimaging data.

2.3. Canonical Correlation Analysis

The framework of CCA is closely related to the more classical regression problem. Hence, it can aid one's intuition and understanding of CCA. Thus, we first recall the classical regression problem and formulate the corresponding optimization problem, which is found in Equation 1.

Let $Y \in \mathbb{R}^N, X \in \mathbb{R}^{N \times p}, \epsilon \in \mathbb{N}$. The common regression model is $Y = X\beta + \epsilon$. Classical regression aims to find the $\hat{\beta}$ that solves the following optimization problem:

$$\hat{\beta} = \arg\min_{\beta} \|Y - X\beta\|_2^2 \tag{1}$$

This technique is sufficient when one aims to construct a simple and interpretable model for prediction, and when multicollinearity is not an issue. However, one is often more interested in identifying correlations between two sets of variables. This is a frequent question in the field of neuroimaging since numerous researchers focus on identifying shared structures between subject traits and brain function or structures. Mathematically, instead of solely finding one set of coefficients $\hat{\beta}$, CCA aims to find two sets of coefficients: $\hat{\gamma}$ and $\hat{\beta}$. CCA simultaneously identifies linear combinations of both $\hat{\gamma}$ and $\hat{\beta}$ that maximize the correlations between γY and βX. This is akin to Equation 1 and can be formulated into an optimization problem, but the target of CCA is correlations, whereas classical regression has the objective of minimizing the error in explaining the variation. Therefore, we adjust the classical regression problem to align with the objective of CCA. This problem is shown in Equation 2. The constraint is included to make the solution unique up to a sign flip. In this paper, we will take X to be FC measures and Y to be behavioral traits and characteristics of participating subjects.

$$\hat{\beta}, \hat{\gamma} = \arg\max_{\beta,\gamma} \ \mathrm{Corr}(X\beta, Y\gamma) \quad \text{s.t. } \mathrm{Var}(X\beta) = \mathrm{Var}(Y\gamma) = 1. \tag{2}$$

Prior to conducting CCA, it is necessary to preprocess the data to ensure accurate and meaningful results. As mentioned in the methodology, we possess each subject's functional connectivity (FC) as a symmetric 68 x 68 matrix, where the rows and columns corresponded to different brain regions. Since the matrix is symmetric, only the unique pairwise correlations between the brain regions were considered. These pairwise correlations (off-diagonal elements) were extracted and arranged into a single row for each subject, effectively flattening the unique elements of the matrix.

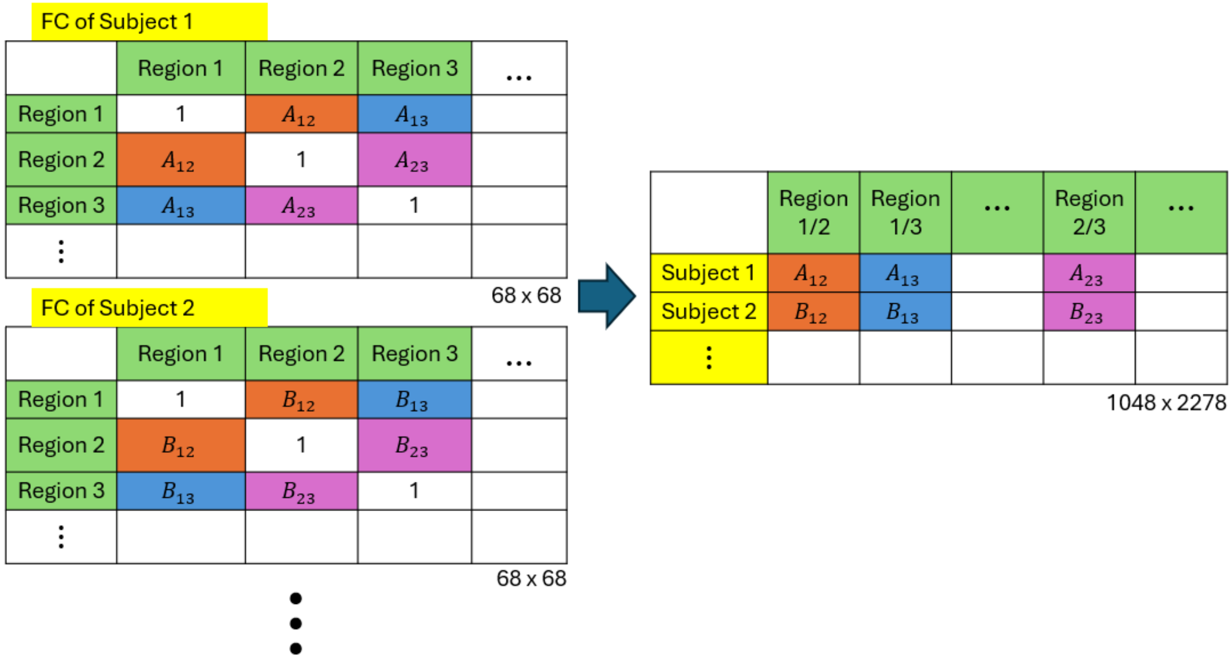
This process resulted in a dataset where each subject was represented by a single row, containing all pairwise correlations. The total number of pairwise correlations is

$$\binom{68}{2} = 2278,$$

corresponding to the unique region pairs. The final consolidated dataset had dimensions of 1065 x 2278 where 1065 is the number of subjects, and 2278 is the number of unique pairwise correlations. This format allowed all subjects to be contained in a single, structured dataset for subsequent analysis. This process is illustrated in Figure 3.
Handling missing values in the functional connectivity (FC) data reduced the dataset to 1,058 subjects. Additionally, addressing missing information in the traits dataset required a multi-step process. First, we identified and removed subjects with the highest number of missing trait observations to maximize the retention of traits. Subsequently, traits with any remaining missing values were eliminated. As a result, the dataset dimensions were reduced to 1,048 subjects and 123 traits, down from the original 175 traits.

Next, we addressed redundancy in the dataset by identifying and removing likely twins. The Human Connectome Project (HCP) mentions the presence of twin pairs in their dataset, which could introduce similarities and reduce the diversity of the data due to genetic and environmental factors. To account for this, we utilized features such as Family ID and age to identify potential twin pairs—specifically, subjects who shared the same Family ID and were within one year of age. To prevent redundancy, one twin from each identified pair was randomly removed, ensuring a more representative and independent dataset for further analysis.

With the bulk of the preprocessing complete, the methodology for the analysis is outlined in detail. As mentioned in 1 Introduction, we intend to build upon previous research, specifically, we employ the R package developed by Kessler and Levina [2023]. The purpose of this study was to enhance methods for performing inference in CCA. While the current literature predominantly emphasizes analysis, there is comparatively little focus on inference. This gap is addressed in the work by utilizing bootstrapping techniques, and a R package, `combootcca`, is created to implement this procedure. An empirical example using neuroimaging data, specifically FC metrics, is examined in the article to demonstrate the effectiveness of the methods. Accordingly, we adopt their methodological framework, applying it to our data by formatting it appropriately and implementing the CCA using the `combootcca` package. A summary of such methodology is laid out in Figure 4.



FC of Subject 1

	Region 1	Region 2	Region 3	…
Region 1	1	A_{12}	A_{13}	
Region 2	A_{12}	1	A_{23}	
Region 3	A_{13}	A_{23}	1	
⋮				

68 x 68

FC of Subject 2

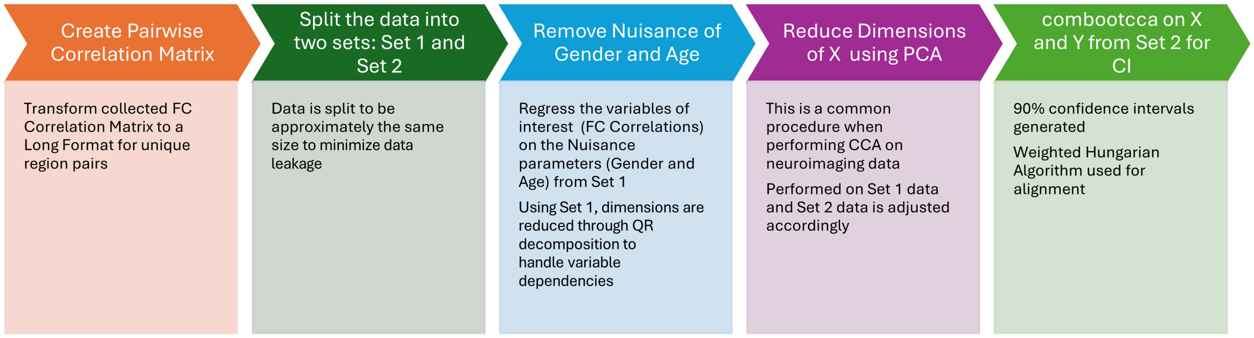
	Region 1	Region 2	Region 3	…
Region 1	1	B_{12}	B_{13}	
Region 2	B_{12}	1	B_{23}	
Region 3	B_{13}	B_{23}	1	
⋮				

68 x 68

	Region 1/2	Region 1/3	…	Region 2/3	…
Subject 1	A_{12}	A_{13}		A_{23}	
Subject 2	B_{12}	B_{13}		B_{23}	
⋮					

1048 x 2278

Figure 3: Overview of the formation of Pairwise correlation Matrix



Create Pairwise Correlation Matrix

Transform collected FC Correlation Matrix to a Long Format for unique region pairs

Split the data into two sets: Set 1 and Set 2

Data is split to be approximately the same size to minimize data leakage

Remove Nuisance of Gender and Age

Regress the variables of interest (FC Correlations) on the Nuisance parameters (Gender and Age) from Set 1

Using Set 1, dimensions are reduced through QR decomposition to handle variable dependencies

Reduce Dimensions of X using PCA

This is a common procedure when performing CCA on neuroimaging data

Performed on Set 1 data and Set 2 data is adjusted accordingly

combootcca on X and Y from Set 2 for CI

90% confidence intervals generated

Weighted Hungarian Algorithm used for alignment

Figure 4: Detailed procedure for data processing following Kessler and Levina [2023]

We begin by creating the pairwise correlation matrix. This was discussed earlier because this step was necessary to perform the EDA; consequently, we omitted the details here. Note that X is the filtered pairwise FC matrix, and Y is the filtered trait matrix. The next step was dividing the data into sets: Set 1 and Set 2. The total number of samples from preprocessing n was found, and approximately 50% of the samples were allocated to a training set, with the indices selected randomly without replacement. The remaining samples were assigned to the testing set. The pairwise FC metrics were divided into training X_1 and testing X_2 matrices. Similarly, the subject-level traits were split into training Y_1 and testing Y_2 matrices, and the confounding variables were divided into corresponding training W_1 and testing W_2 matrices. This stratification ensures that all datasets are appropriately aligned across training and testing subsets for further analysis. But more importantly, this combats data leakage and increases the validity of results.

To account for the potential influence of confounding variables, the variables of interest were regressed onto the confounding variables. Specifically, the coefficients for the nuisance covariates were calculated using the following formulas:

$$\hat{A}_X = (W_1^\top W_1)^{-1} W_1^\top X_1, \quad \hat{A}_Y = (W_1^\top W_1)^{-1} W_1^\top Y_1. \tag{3}$$

The contributions of the nuisance covariates were then removed from the variables of interest using:

$$\begin{aligned} X_1^{\text{adjusted}} &= X_1 - W_1 \hat{A}_X, \quad X_2^{\text{adjusted}} = X_2 - W_2 \hat{A}_X, \\ Y_1^{\text{adjusted}} &= Y_1 - W_1 \hat{A}_Y, \quad Y_2^{\text{adjusted}} = Y_2 - W_2 \hat{A}_Y. \end{aligned}$$

One requirement for well-defined canonical correlations in CCA is that both input matrices are full rank. This ensures results are meaningful and interpretable, and prevents issues arising in the `combootcca` package. After adjusting for the confounding variables, we discover there is dependence between certain pairwise regions and between traits, meaning there is multicollinearity, and the X and Y matrices are not full ranks. The linearly dependent variables were identified in X_1^{adjusted} and Y_1^{adjusted} and removed using the `findLinearCombos` function from the `caret` package. This function uses QR decomposition to detect linear combinations within X_1^{adjusted} and Y_1^{adjusted} and groups the dependencies, and outputs recommended columns to remove so both matrices are full rank. The columns identified using X_1^{adjusted} and Y_1^{adjusted} are removed from both X_2^{adjusted} and Y_2^{adjusted}.

Despite being full rank, the X_2^{adjusted} matrix still retained minor dependency issues that led to challenges when performing later analyses. Thus to reduce the dimensionality further, Principal Component Analysis (PCA) via SVD was applied to the adjusted training dataset (X_1^{adjusted}). This is a standard procedure when performing CCA on neuroimaging data (Kessler and Levina [2023]). To ensure consistency, the adjusted testing dataset X_2^{adjusted} was projected onto the principal component basis derived from the training data, using the same centering and scaling parameters. The projected testing dataset X_2^{PCA} was then truncated to the first 150 principal components, capturing the most significant variance while reducing dimensionality.

We now have two matrices, X_2^{PCA} and Y_2^{adjusted}, that have independent columns. The final step before performing CCA is to standardize our data. We standardize each column of both matricies to have mean 0 and standard deviation 1. We define these new matrices as $\tilde{X}_2$ and $\tilde{Y}_2$. With this complete, we can start initiating the arguments for `combootcca`.

In Kessler and Levina [2023], 90% confidence intervals are constructed from bootstrapping, so we also generated 90% intervals to follow this. After examining the data, it appears to be approximately symmetric. Thus, we decided to do 1000 bootstraps using the basic type. Also in the reference, there is notable discussion about the alignment of the canonical direction. This is vital because Equation 2 only has a unique solution up to a sign flip, hence a reference solution is required to provide the reference direction. We begin by creating a reference solution with `cc()` function from the `CCA` package. The recommended alignment from Kessler and Levina [2023] is Assignment via Weighted Hungarian Algorithm because it balances the risks of underestimating and overestimating the variance. Therefore, we set this method as the alignment procedure in `combootcca`. The current setup is now ready for the execution of CCA using `combootcca`.

3 Methods for Reproducibility

An emphasis was placed on reproducibility practices in this project to ensure transparency and validate our work. The code relevant to this work is available on GitHub: `https://github.com/colemanferrell2/STOR674-Final-Project.git`. We faced several challenges regarding reproducibility due to the sensitive nature of the data. Consequently, we took deliberate steps to comply with the terms of the restricted dataset provided by the HCP. More importantly, we carefully designed the analysis to protect the privacy of HCP subjects.

Robust data transformations and filtering were performed during preprocessing, and these methods are detailed in Section 2. We specifically must avoid providing data that could lead to identifiability on the individual level. To ensure anonymity, we implemented various redaction and randomization techniques. For the trait data, sensitive behavioral traits, such as substance use and psychiatric well-being, are reported at the individual level. We used the `synthpop` package in R to create synthetic data based on the distributions of raw behavioral traits. Additionally, we removed the trait names, reordered them randomly, and assigned unidentifiable numbers. The specific arguments and methods used in `synthpop` are not disclosed to prevent traceability.

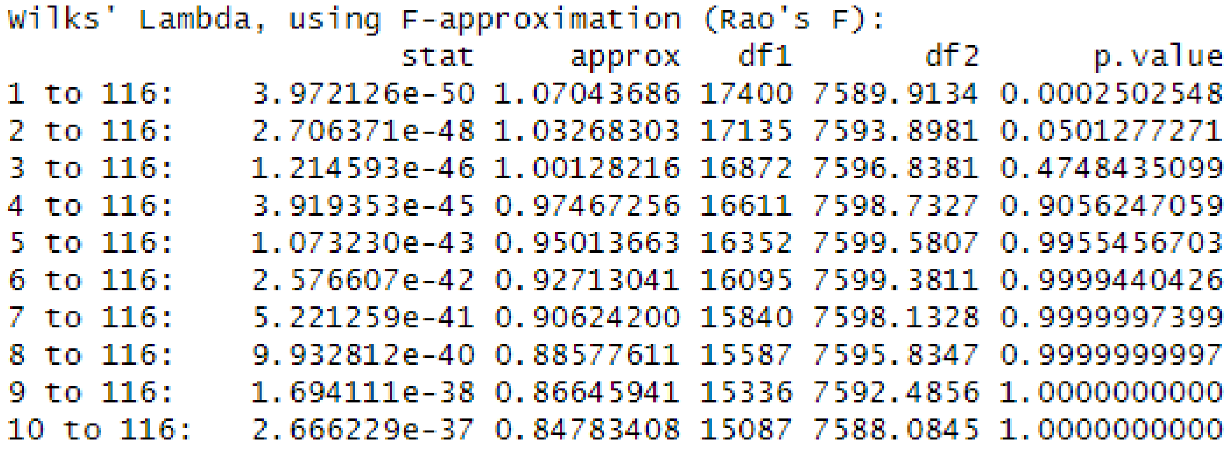
For the brain region data, identifiability is already reduced through PCA, requiring minimal additional anonymization. To further mitigate the risk of reverse engineering, we did not report specific loadings for each subject, rather we again created synthetic data. For both sample datasets, subject identifiers were removed, and the rows were arranged in a disclosed random order. The sample data is provided in an `.Rdata` file on GitHub, located in `data\Sample_Data.Rdata`.

With the privacy concerns addressed, we next focus on the reproducibility practices we implement to ensure others can replicate our work. First, we structure the folders in our project clearly and logically, separating the repository into `\data`, `\scripts`, and texttt\results folders. We documented the project thoroughly with a detailed README file in the home directory and within each subfolder. These files contain step-by-step instructions for setting up the environment, running the analysis, and interpreting the results.

All analysis code was annotated with inline comments in the `scripts\analysis.R`, and functions for the analysis are included in the `scripts\functions.R` with function purpose. Additionally, we used the `renv` package to manage R package dependencies, ensuring that all collaborators and future users can replicate the computational environment accurately. Finally, we make use of `Git` for version control, ensuring that all changes to the code are well-tacked and documented. As mentioned at the beginning of this section, the project was hosted on a GitHub repository, enabling easy sharing and access for team members and public,

4 Results

We aim to uncover correlations between distinct brain regions and subject traits. Using `combootcca`, we achieve this by determining the variables that significantly contribute to the most important canonical directions. Consequently, we first learn which canonical directions are significant. From



```
Wilks' Lambda, using F-approximation (Rao's F):
                  stat      approx   df1       df2      p.value
1 to 116:    3.972126e-50 1.07043686 17400 7589.9134 0.0002502548
2 to 116:    2.706371e-48 1.03268303 17135 7593.8981 0.0501277271
3 to 116:    1.214593e-46 1.00128216 16872 7596.8381 0.4748435099
4 to 116:    3.919353e-45 0.97467256 16611 7598.7327 0.9056247059
5 to 116:    1.073230e-43 0.95013663 16352 7599.5807 0.9955456703
6 to 116:    2.576607e-42 0.92713041 16095 7599.3811 0.9999440426
7 to 116:    5.221259e-41 0.90624200 15840 7598.1328 0.9999997399
8 to 116:    9.932812e-40 0.88577611 15587 7595.8347 0.9999999997
9 to 116:    1.694111e-38 0.86645941 15336 7592.4856 1.0000000000
10 to 116:   2.666229e-37 0.84783408 15087 7588.0845 1.0000000000
```

Figure 5: Wilks' Lambda Test on Canonical Directions

Table 1: Variables with significant canonical weights in first canonical direction

Variables in $\tilde{X}$	Variables in $\tilde{Y}$	Trait Category
PC 1	Trait 42	Cognition
PC 4	Trait 47	Motor
PC 5	Trait 49	Motor
PC 10	Trait 51	Motor
PC 59	Trait 53	Substance Use
	Trait 54	Substance Use
	Trait 104	Psychiatric and Life Function

the reference solution, we perform Wilks' Lambda Test, Wilks [1932], on the CCA correlations as seen in Figure 5. At a significance level of $\alpha = 0.05$, only the first correlation is significant. However, the second correlation is near the significance threshold, so it will also be considered in subsequent analysis. The remaining correlations are far from the significance threshold and, therefore, were not analyzed further.

Remaining aligned with our goal, we utilize `combootcca` to construct confidence intervals for each variable in $\tilde{X}$ and $\tilde{Y}$ on the first two canonical directions. These confidence intervals are for the canonical weights of each variable. The confidence intervals for the PCA brain regions and traits are in Figure 6 and 7, respectively. All four graphs contain intervals that do not contain 0, thereby indicating the variables make a statistically significant contribution to the corresponding canonical direction. We extract these variables and investigate their practical interpretation. The variables with significant weights in each direction are shown in Tables 1 and 2. The category of trait is also included for the variables associated with $\tilde{Y}$ to aid in visualization.

The principal components in $\tilde{X}$ are meaningless without direct ties in the context of brain regions. To facilitate interpretation, we project these components back into the original brain region feature space. Since PCA was performed using SVD, we can invert the components back into the brain region feature space by reversing the transformation. This allows us to retrieve the corresponding loadings of the original brain region features associated with each principal component. Since multiple principal components were found to be significant, we compute the total contribution of each brain region to these components by summing the squared loadings of that region across all significant components. To interpret the relative importance of each region,

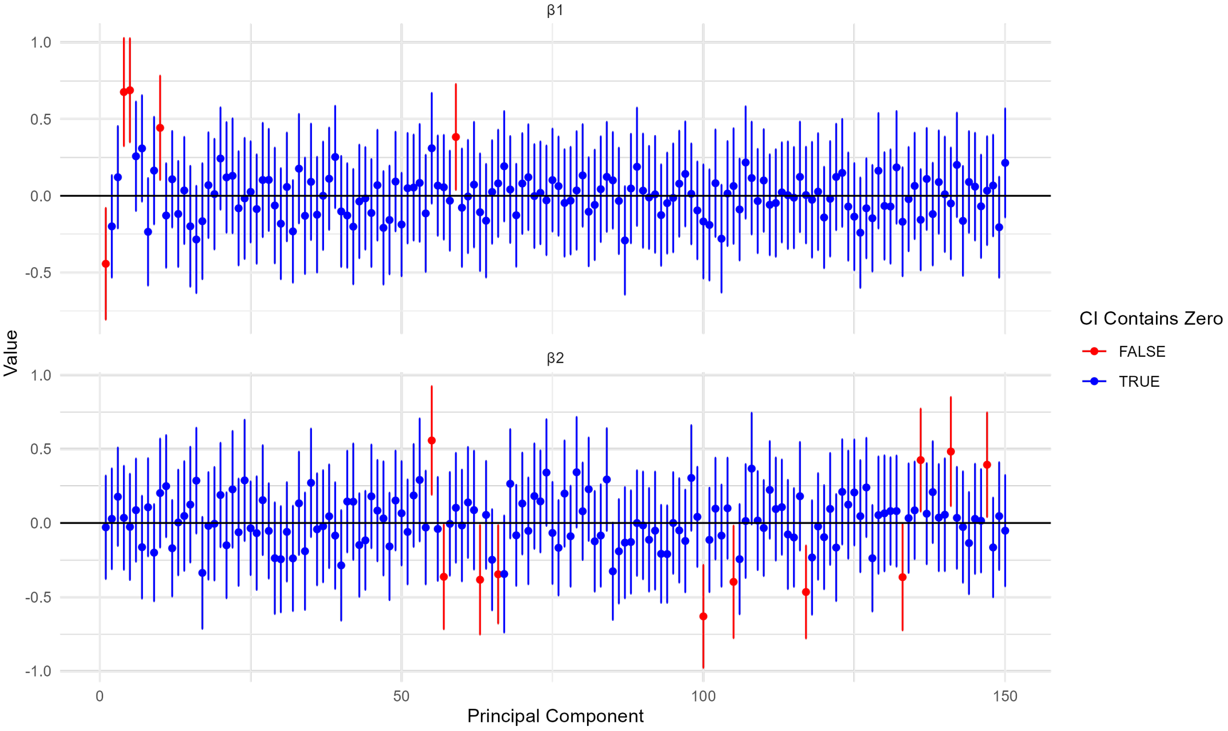


Figure 6: Weights confidence interval for first two canonical directions for β (Brain Regions)

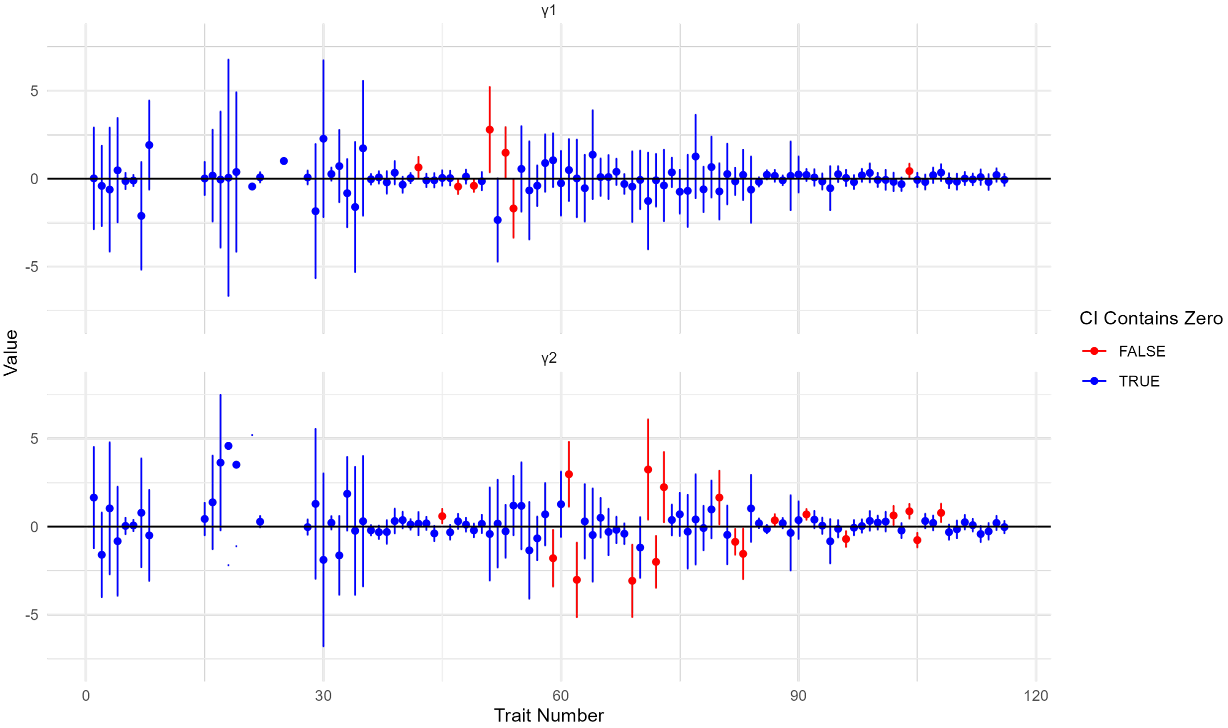


Figure 7: Weights confidence interval for first two canonical directions for γ (Traits)

Table 2: Variables with significant canonical weights in second canonical direction

Variables in $\tilde{X}$	Variables in $\tilde{Y}$	Trait Category
PC 55	Trait 45	Cognition
PC 57	Trait 59	Substance Use
PC 63	Trait 61	Substance Use
PC 66	Trait 62	Substance Use
PC 100	Trait 69	Substance Use
PC 105	Trait 71	Substance Use
PC 117	Trait 72	Substance Use
PC 133	Trait 73	Substance Use
PC 136	Trait 80	Substance Use
PC 141	Trait 82	Substance Use
PC 147	Trait 83	Substance Use
	Trait 87	Substance Use
	Trait 91	Psychiatric and Life Function
	Trait 96	Psychiatric and Life Function
	Trait 102	Psychiatric and Life Function
	Trait 104	Psychiatric and Life Function
	Trait 105	Psychiatric and Life Function
	Trait 108	Psychiatric and Life Function

we calculate the percent distribution of these contributions by normalizing the total contribution of each region against the sum of all regional contributions across the components. We can express this mathematically as follows:

Let $L \in \mathbb{R}^{p \times k}$ be the matrix of loadings, where L_{ij} represents the loading of brain region pair i on principal component j, $j \in [1, k]$. p is the number of pairwise brain regions and k is the number of significant principal components. The squared loadings, L_{ij}^2, reflect the contribution of region i to the variance explained by component j. The total contribution of region i across all significant principal components is given by:

$$C_i = \sum_{j=1}^{k} L_{ij}^2$$

The percent distribution of contributions for region i is calculated as:

$$P_i = \frac{C_i}{\sum_{i=1}^{p} C_i} \times 100$$

Where C_i is the total contribution of brain region i across the significant components, and $\sum_{i=1}^{p} C_i$ is the total contributions of all brain regions.

We make the transformation and percent contribution calculations for both the first and second canonical directions. Truncatated tables of these findings are in Tables 3 and 4. These tables can be utilized to identify correlations with the traits from Tables 1 and 2

The first major observation is the prominent contribution of Region 3 in the first canonical direction. Of the top eight brain region pair contributors, Region 3 was included in seven of the variables. Region 3 corresponds to the caudal portion of the middle frontal gyrus in the left hemisphere of the brain, located in the frontal lobe. From Table 1, we can see a majority of the traits with significant canonical weights are in either the Cognition or Motor category. This

Table 3: First direction brain region pairs loading contribution

Brain Region Pair	Percentage Contribution
Region_3_9	0.026908
Region_3_13	0.018428
Region_3_10	0.017397
Region_3_18	0.016895
Region_3_24	0.016404
Region_3_59	0.015455
Region_5_44	0.015235
Region_3_12	0.014927
Region_2_44	0.014681
Region_2_56	0.014538

Table 4: Second direction brain region pairs loading contribution

Brain Region Pair	Percentage Contribution
Region_3_9	0.038561
Region_1_51	0.034711
Region_1_9	0.033603
Region_4_21	0.033562
Region_2_20	0.031879
Region_2_22	0.029619
Region_3_11	0.029141
Region_1_56	0.028542
Region_4_33	0.026332
Region_1_41	0.025772
Region_2_32	0.023865
Region_5_27	0.023859
Region_3_51	0.023592
Region_5_47	0.023198
Region_2_66	0.022375
Region_1_29	0.022143
Region_3_17	0.021930
Region_3_34	0.021910
Region_2_16	0.021902
Region_2_34	0.021859

suggests a correlation between an individual's cognition and motor skills and Region 3. This is consistent with the results provided in Ridderinkhof et al. [2004]. This article provides indirect evidence that the caudal middle frontal gyrus may indeed be linked to both cognitive control and motor functions, as it likely works in concert with the medial frontal cortex for monitoring and regulating behaviors. This strengthens our argument that Region 3 is relevant for these domains in our analysis.

Looking at the second canonical direction, it is less apparent that there are any correlations. This may be expected, as this direction was technically found to be insignificant in the Wilks' Lambda Test. Table 4 has varying brain region pairs and no pattern was found on initial inspection. However, the traits in Table 2 are dominated by Substance Use. This observation is particularly intriguing given that Zhang et al. [2019] highlights associations between structural connectivity and substance use. Although it does not address functional connectivity, future research exploring potential correlations between functional connectivity and substance use could yield valuable insights.

References

Daniel Kessler and Elizaveta Levina. Computational inference for directions in canonical correlation analysis, 2023. URL `https://arxiv.org/abs/2308.11218`.

K. Richard Ridderinkhof, Markus Ullsperger, Eveline A. Crone, and Sander Nieuwenhuis. The role of the medial frontal cortex in cognitive control. *Science*, 306(5695):443–447, 2004.

S. S. Wilks. Certain generalizations in the analysis of variance. *Biometrika*, 24(3/4):471–494, 1932.

Zhengwu Zhang, Genevera I. Allen, Hongtu Zhu, and David Dunson. Tensor network factorizations: Relationships between brain structural connectomes and traits. *NeuroImage*, 197:330–343, 2019. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2019.04.027. URL `https://www.sciencedirect.com/science/article/pii/S1053811919303131`.