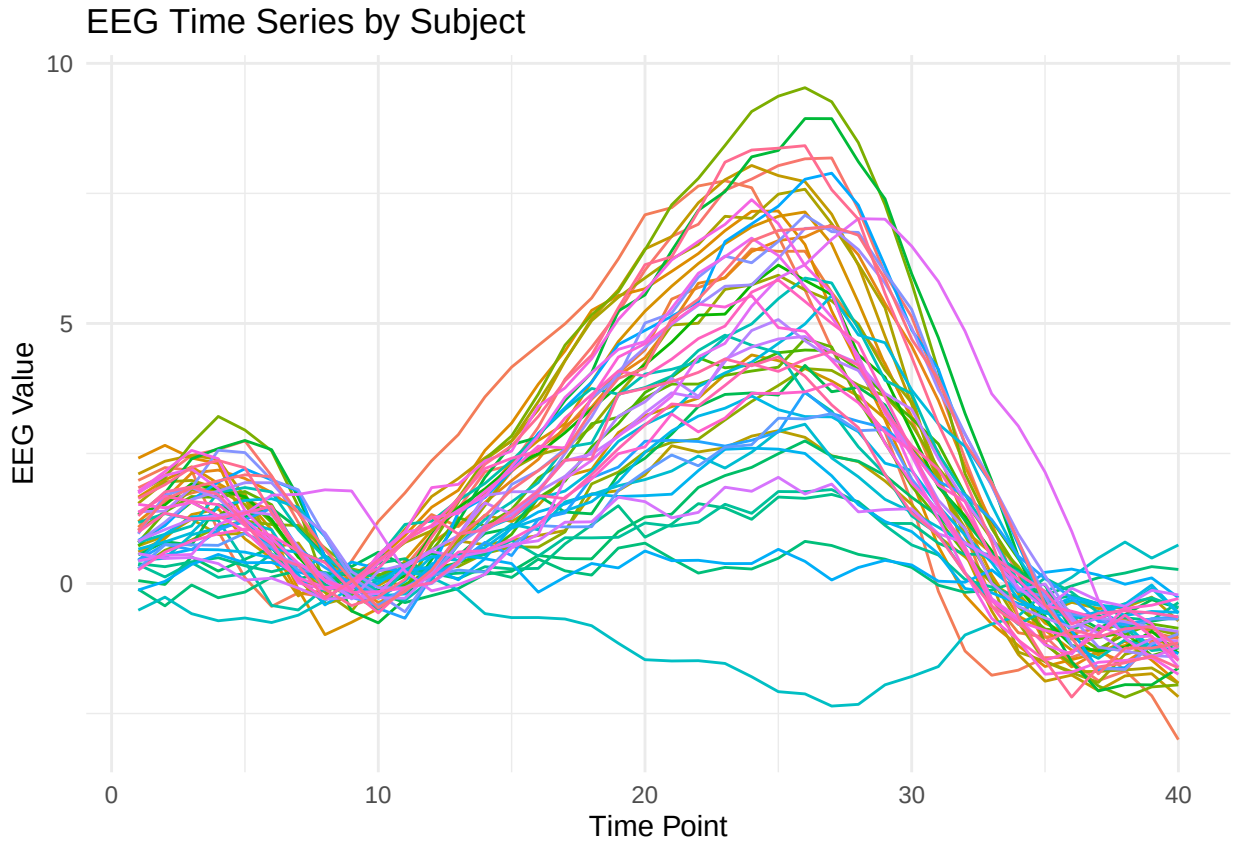
Midterm

Jacob Thielemier

2025-03-16

Question 1

- The plot below shows pretty consistent behavior between different time points. Most of the subjects follow a similar rise and then fall pattern.
- Most subjects peak at a time period of 25. Several subjects follow each other by rising and falling at the same time periods. But a few outliers exists that show sudden deviations near 10 and 30.

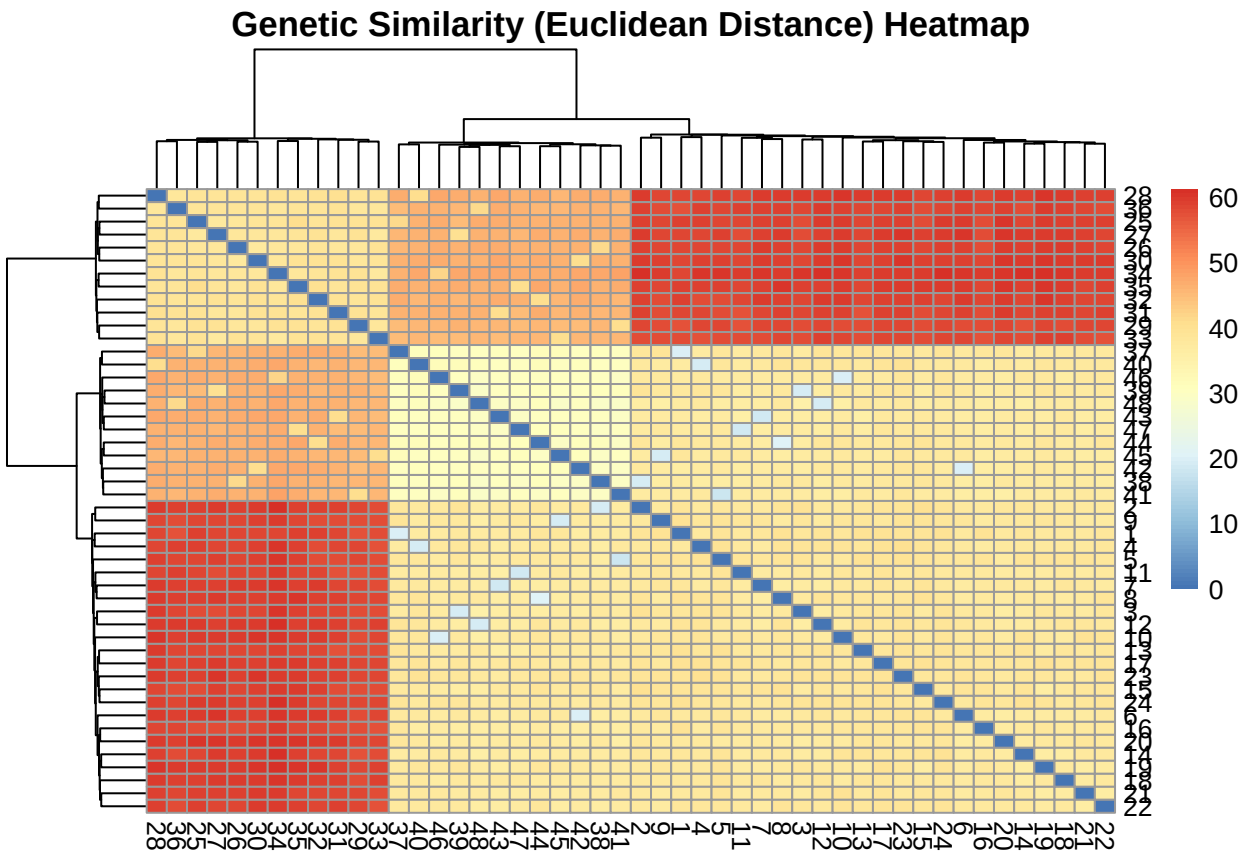


Question 2

- The heatmap below shows that there is clustering among the subjects.
  - One cluster in the top-left where the subjects are relatively genetically similar. The lighter yellow shades mean smaller distances.

  - Another cluster in the bottom-right is darker red. Meaning those are more similar withing their own group, but also more distant from the other cluster.

- The overall distance between the two clusters is far and that suggests there is a high genetic dissimilarity between the subjects in the two clusters.



Question 3

- We can see in the table below that the Mantel test shows significant genetic-EEG association (RV = 0.459, $p < 0.0001$).

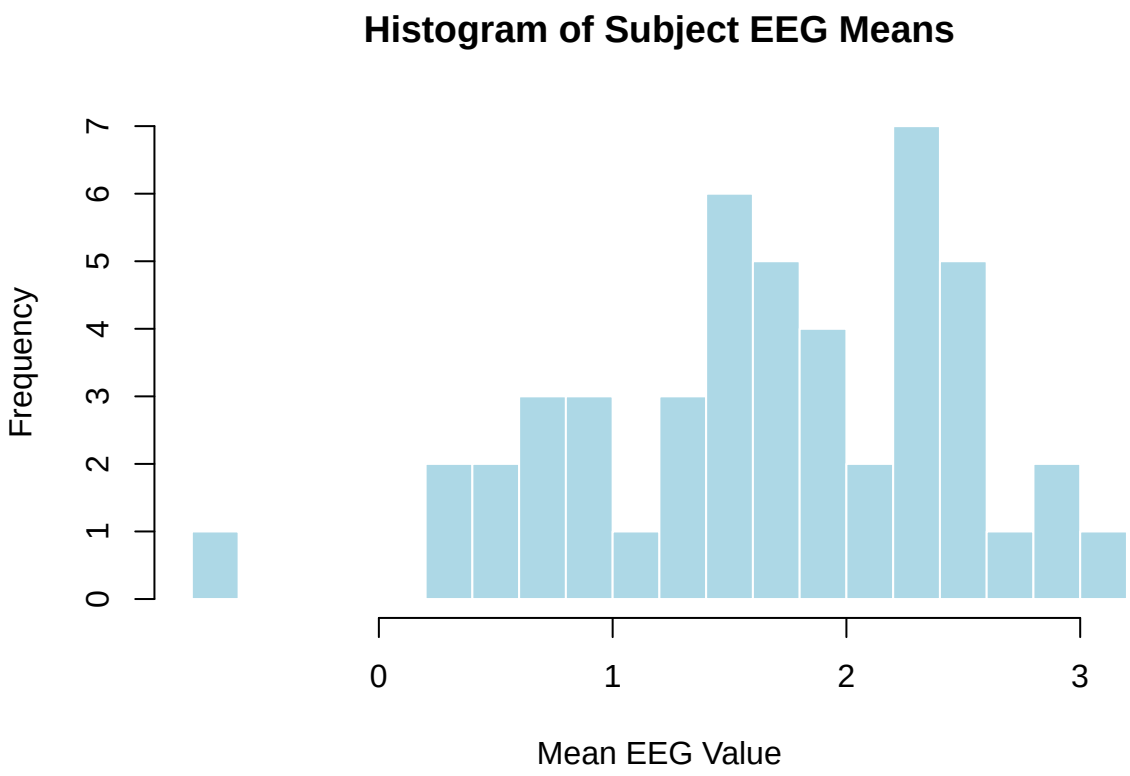
- Genetic differences among the subjects significantly influence EEG responses.

Table 1: Summary of RV Coefficient Results Between EEG and SNP Data

Metric	Value
RV	0.459
RV Std	14.51
Mean	0.067
Variance	7.3e-04
Skewness	2.354
p-value	8.17e-07

Question 4

- In the histogram below, EEG data for each subject is summarized by the mean EEG value across all time points. This provides a single EEG measure per subject.



Question 5

```
Linear mixed model fit by REML ['lmerMod']
Formula: EEG_Mean ~ (1 | Subject)
   Data: merged_data

REML criterion at convergence: 75.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.13977 -0.26561 -0.07704  0.25669  1.12586

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 0.5819   0.7628
 Residual             0.0191   0.1382
Number of obs: 48, groups:  Subject, 48

Fixed effects:
            Estimate Std. Error t value
```

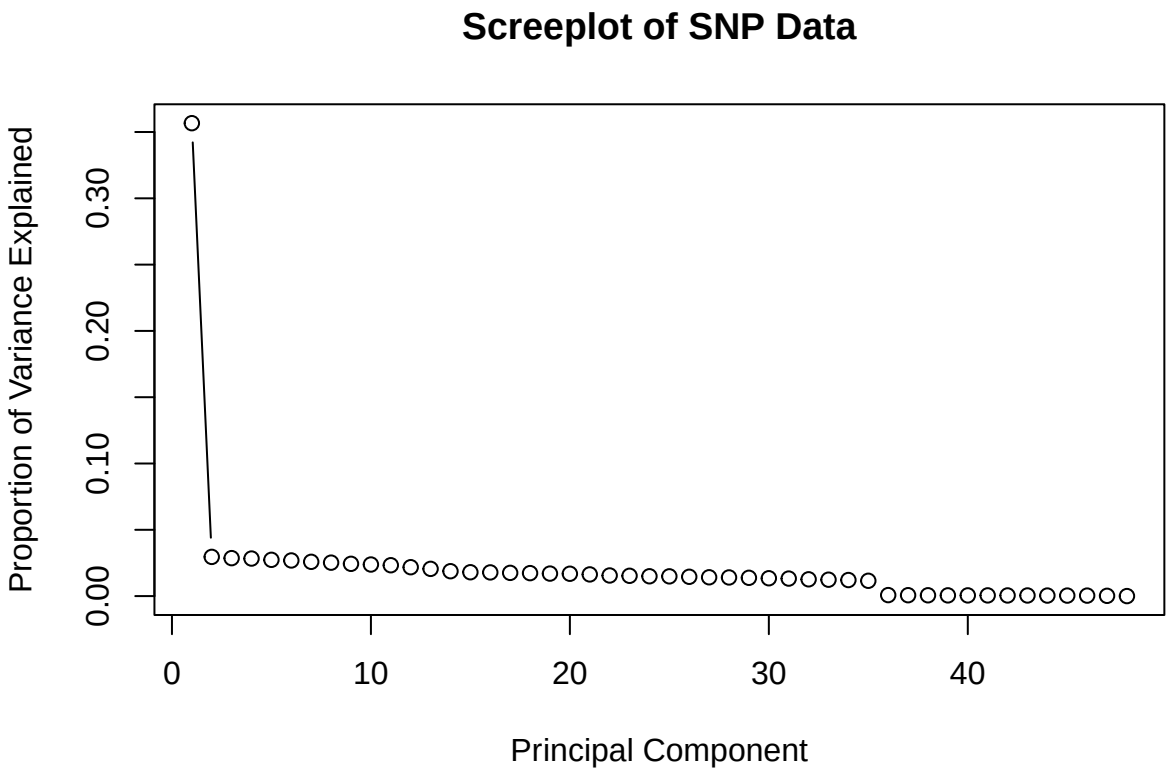
```
(Intercept)  1.67070    0.01995   83.75

Proportion of variance explained by SNPs: 0.968

Approximate 95% CI for proportion explained: 0.948 - 0.977
```

Question 6

- Based on the Scree plot below we see the elbow takes shape at the second PCA. We should only use two components.



Question 7

- There is strong evidence ($p < 0.0001$) that the EEG means are associated with PC1. However, PC2 is not significantly associated with EEG means ($p = 0.729$). This tell me that at least one genetic principal component (PC1) has a clear relationship with the EEG phenotype.
- The adjusted R-squared is 0.396, meaning that 39.6% of the total variance in EEG means is explained by these genetic principal components.

Table 2: Linear Regression of EEG Means on Genetic Principal Components

	Predictor	Estimate	Std. Error	t-value	p-value
(Intercept)	(Intercept)	1.671	0.091	18.260	1.96e-22
PC1	PC1	-0.028	0.005	-5.741	7.55e-07
PC2	PC2	-0.006	0.017	-0.349	0.729

Question 8

- PCA-based regression shows that genetic variation significantly influences EEG means:
  - PC1 had a strong association ($p < 0.0001$); PC2 did not ($p = 0.729$).
  - Genetic principal components explained 39.6% (adjusted R^2) of EEG variability.
  - Scree plot suggested retaining two principal components, with a clear elbow at PC2.
- The variance components analysis provided an overall measure of genetic influence:
  - Genetic relatedness explained a substantial proportion of EEG variance (96.8%, 95% CI: 94.8–97.7%, from Question 5).
  - Mantel/RV coefficient analysis also indicated significant genotype-EEG association (RV = 0.459, $p < 0.0001$, from Question 3).
- Together, these methods confirm a clear genetic contribution to EEG phenotypes, capturing both specific genetic components and overall genetic relatedness.