Nicole Kayambu
CSCI 4341-02

HW4 Pairwise Associations

Problem 1.

Contingency Table:

Genome2	0	1	All
Genome1			
0	128	21	149
1	50	0	50
All	178	21	199

Mutual Information: 0.03257317770864432

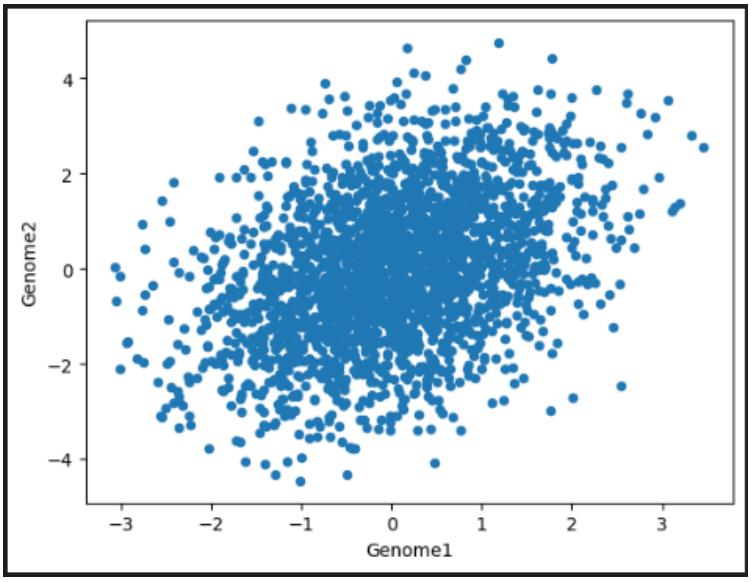
Jaccard Index: 0.0

Chi-Squared: 7.87836513083478 (p value = 0.09613683666331894)

Problem 2.

Part a.

Plot of data:



Pearson Correlation: 0.3808750357837303

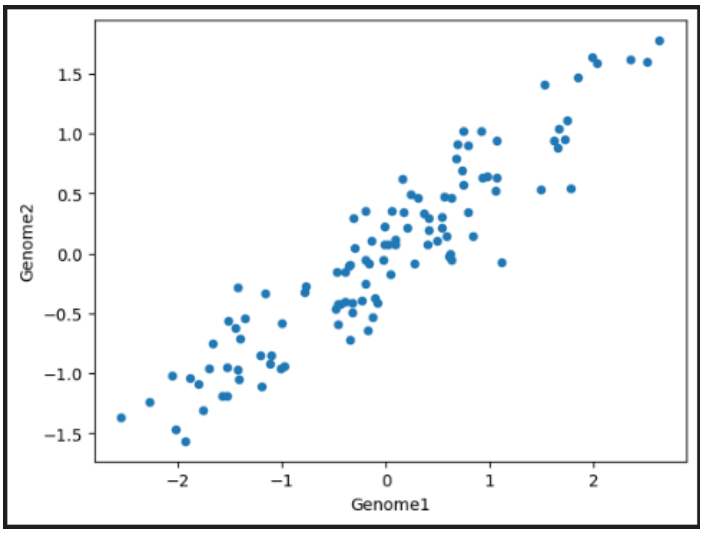
P-value: 1.0409455130017367e-83

Assuming α is .05, we reject the null hypothesis as the p value is less than α. The Pearson correlation of the genomes is .38, this means that there is some correlation between the two genomes in the positive direction.

Nicole Kayambu
CSCI 4341-02

Part b.

Plot of data:



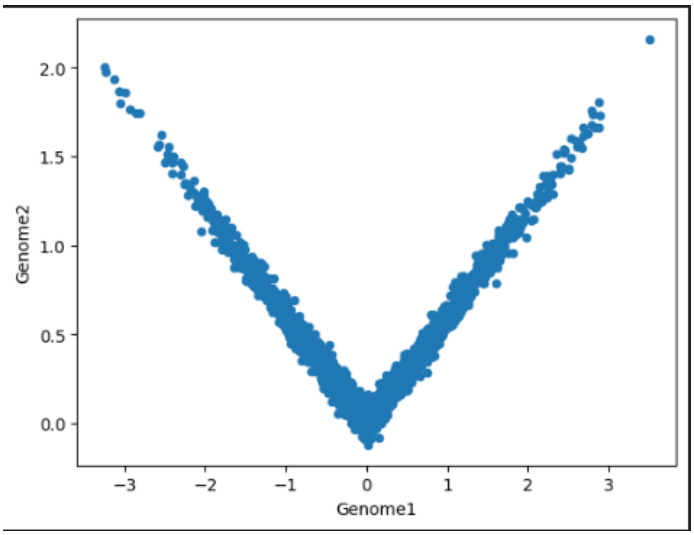
Pearson Correlation: 0.9312196333264214

P-value: 3.737321008438923e-49

Assuming α is .05, we reject the null hypothesis as the p value is less than α. The Pearson correlation between these two genomes is .93, this means there is a very high correlation between the two genomes in the positive direction.

Part c.

Plot of data:



Pearson Correlation: 0.04117899777683178

P-value: 0.059195916605452974

Nicole Kayambu
CSCI 4341-02

Assuming α is .05, we fail to reject the null hypothesis as the p value is greater than α. The Pearson correlation between these two genomes is .04 which is very low, meaning that there is not much correlation between the two genomes.

Part e.

The associations between the two in terms of correlation and p value is the strongest with 2b's data. With the scatter plots it looks like 2c's data also might be heavily correlated however, the shape of the scatterplot being a V will make the numbers calculated and what is seen have some discrepancy. The scatterplot of 2b's data shows a loose line in the positive direction which is what the calculations give us as well, since it is a strong linear relationship it reflects that. The scatterplot of 2c's data is not linear and that could be where the discrepancy lies. The scatterplot of 2a's is a cluster with a somewhat linear shape and a positive direction which reflects in the calculations.