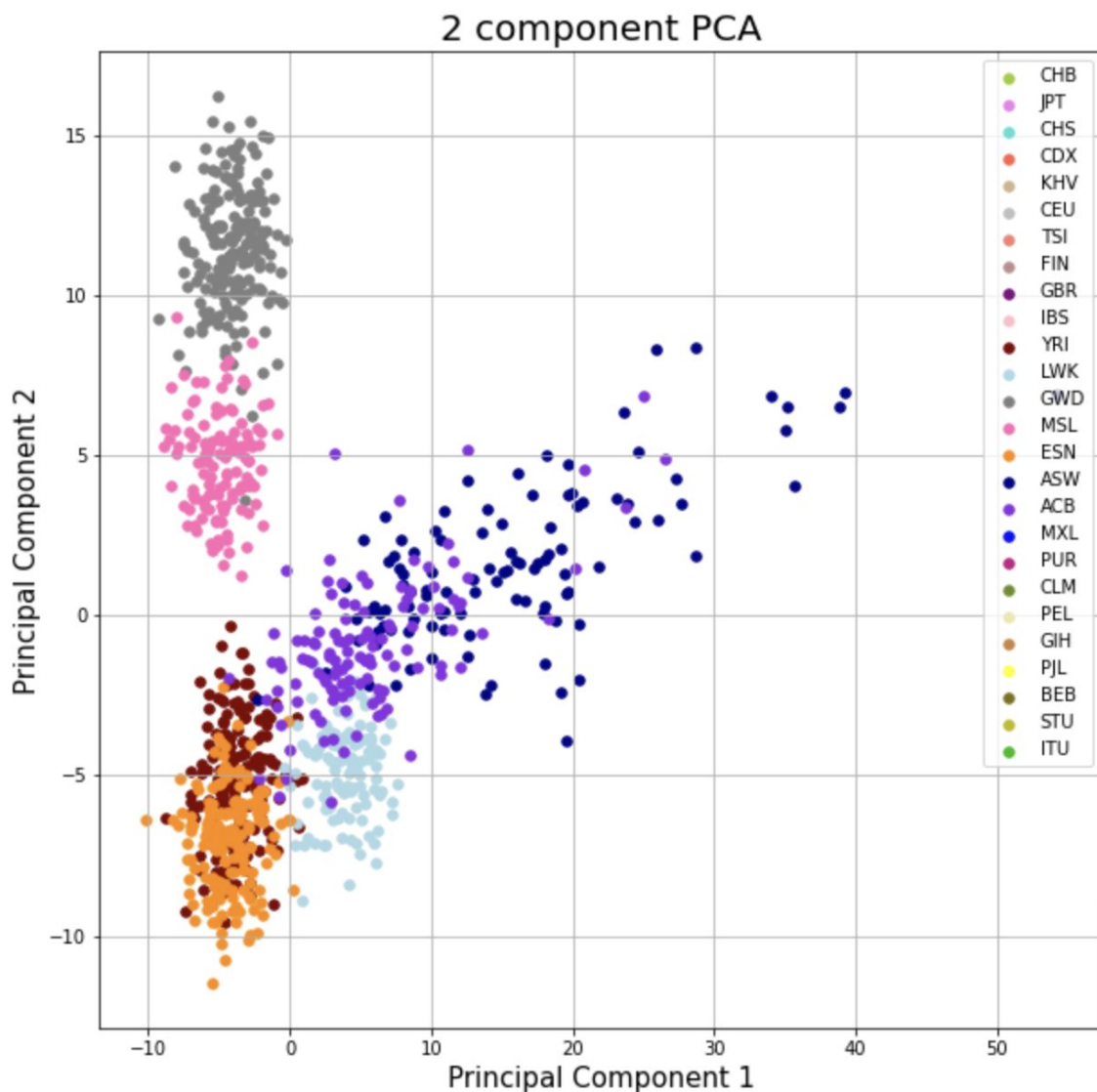Mini-Project #1
Anna Fischer
2/24/21
CSE 40625

a.) If we ran PCA on the matrix, we would get vectors with dimension 995 for each principal component, because there are 995 individuals.

b.)

c.) The plot in part b helps interpret the data by looking at the various population groupings. Principal components 1 and 2 look at the top 2 correlated attributes according to each population grouping. pc1 and pc2 show populations ESN and YRI are similarly correlated, which makes sense because they are both Nigerian. On the other hand, a different population GWD (is

positively correlated for pc2 but similarly correlated for pc1 as ESN and YRI) because it is also in Africa geographically, but varies in pc2.

d.)
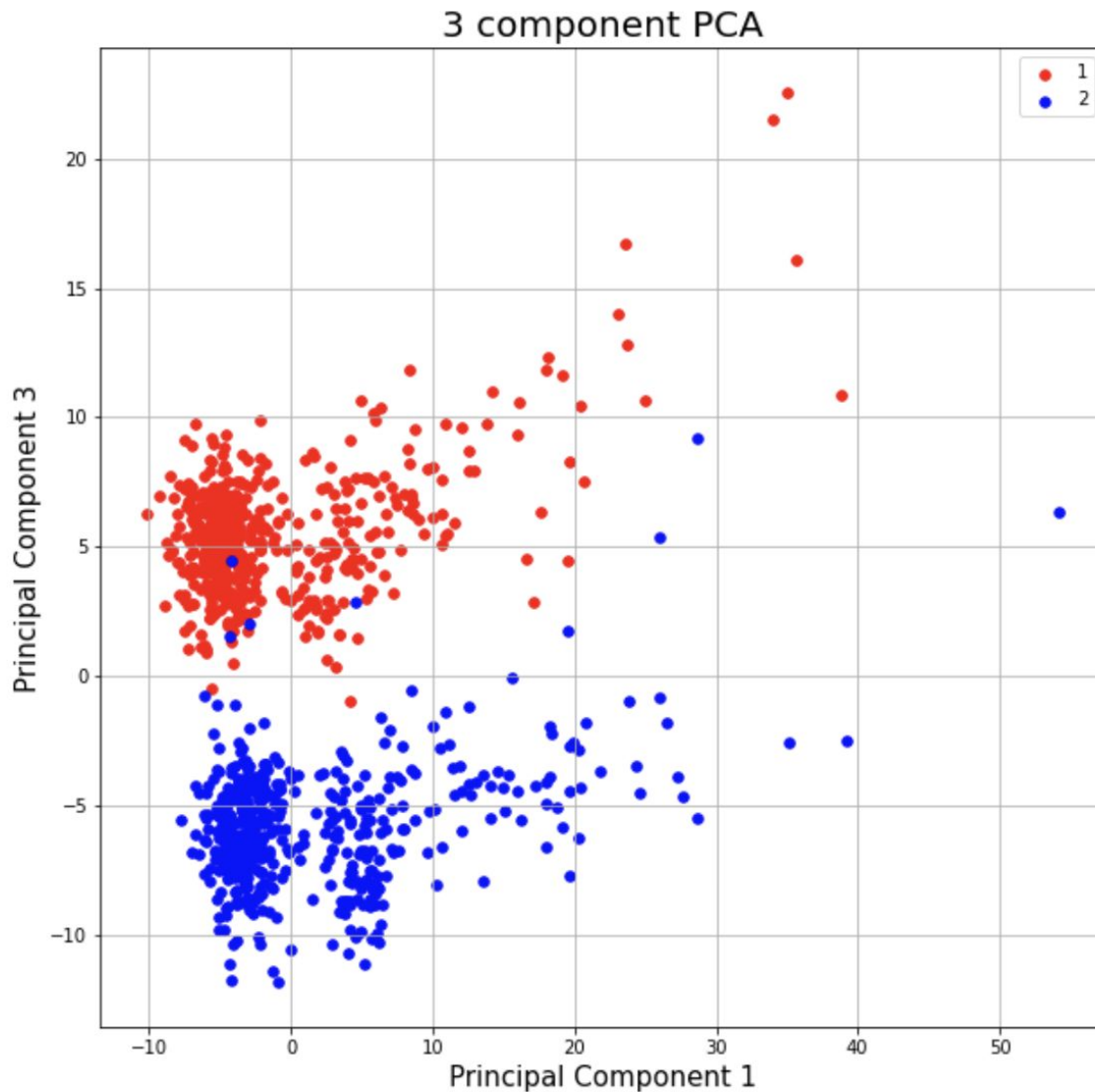
e.) This plot shows positive correlation between pc3 and males, and negative for females and pc3. Thus, it is obvious that the 3rd principal component correlates gender for each individual.

f.) Since pc3 correlates to gender, the plot of nucleobase and absolute value of pc3 shows similar correlation for most of the nucleotides except for some on the end, which is because some individuals are female and some are male, meaning that they would have different

nucleobases for XX chromosome vs XY chromosomes.