# activate environment

conda activate qiime2-amplicon-2024.2

# will summarize the information created in the demux file by telling qiime to do it

# create demux visualization

qiime demux summarize \

--i-data demux.qza \

--o-visualization demux.qzv

#this code is telling qiime to do a quality control in paired-end reads

#trim on the left side of forward read at 0 and truncate on the right side of forward reads at 190. The next two codes are doing the same at different locations for reverse reads.

#the next three codes will tell qiime to create 3 files with req seqs, a table and statistics from the data found in demux.qza

qiime dada2 denoise-paired \

--i-demultiplexed-seqs demux.qza \

--p-trim-left-f 0 \

--p-trunc-len-f 190 \

--p-trim-left-r 0 \

--p-trunc-len-r 160 \

--o-representative-sequences rep-seqs.qza \

--o-table table.qza \

--o-denoising-stats stats.qza

#this command is telling qiime to summarize and format the metadata file associated with our samples

#this is creating a visualization based off the file stats.qza generated in the code above

qiime metadata tabulate \

--m-input-file stats.qza \

--o-visualization stats.qzv

#this command is telling qiime to summarize all the data in feature table using the medata.txt found in our folder

#using this information, it will create visualizations for the table and sequences, with the files also generated in the code above

qiime feature-table summarize \

--i-table table.qza \

--o-visualization table.qzv \

--m-sample-metadata-file metadata.txt

qiime feature-table tabulate-seqs \

--i-data rep-seqs.qza \

--o-visualization rep-seqs.qzv

#this is telling the terminal to download a specific file that we will use for our taxonomy base

wget \

-O "gg-13-8-99-515-806-nb-classifier.qza" \

"https://data.qiime2.org/2024.2/common/gg-13-8-99-515-806-nb-classifier.qza"

#this will take the data created in the rep-seqs file and classify the taxonomy based on the species found on the species found in the subjects

qiime feature-classifier classify-sklearn \

--i-classifier gg-13-8-99-515-806-nb-classifier.qza \

--i-reads rep-seqs.qza \

--o-classification taxonomy.qza

#this code will create a table where we view the taxonomic classification based on the file created before

qiime metadata tabulate \

--m-input-file taxonomy.qza \

--o-visualization taxonomy.qzv

#this code excludes mitochondria and chloroplast species off our taxonomy classification

#it will do this by using the data created in the table.qza file

qiime taxa filter-table \

--i-table table.qza \

--i-taxonomy taxonomy.qza \

--p-exclude mitochondria,chloroplast \

--o-filtered-table table.qza

#now excluding mitochondria and chloroplast

#this command is telling qiime to summarize all the data in feature table using the medata.txt found in our folder

#using this information, it will create visualizations for the table and sequences, with the files also generated in the code above

qiime feature-table summarize \

--i-table table.qza \

--o-visualization table.qzv \

--m-sample-metadata-file metadata.txt

#this code will create a table where we view the taxonomic classification based on the file created before now excluding mitochondria

qiime feature-classifier classify-sklearn \

--i-classifier gg-13-8-99-515-806-nb-classifier.qza \

--i-reads rep-seqs.qza \

--o-classification taxonomy.qza

qiime metadata tabulate \

--m-input-file taxonomy.qza \

--o-visualization taxonomy.qzv

#this code creates taxonomy barplot using the table.qza file, metadata file and taxonomy file now excluding mitochondria

qiime taxa barplot \

--i-table table.qza \

--i-taxonomy taxonomy.qza \

--m-metadata-file metadata.txt \

--o-visualization taxa-bar-plots.qzv

#the code will remove positions that are highly variable

#this code will generate multiple sequence alignments

# this code will use the rep seqs file to generate a phylogenetic an unrooted and rooted tree

qiime phylogeny align-to-tree-mafft-fasttree \

--i-sequences rep-seqs.qza \

--o-alignment aligned-rep-seqs.qza \

--o-masked-alignment masked-aligned-rep-seqs.qza \

--o-tree unrooted-tree.qza \

--o-rooted-tree rooted-tree.qza

# generate alpha and beta diversity numbers using a sample depth of 8545 - can find this value from table.qzv generated before this

#this code will be using the rooted tree file created in the code above and the table where the sample depth was chosen from

#this code is also using the metadata file to create all the outputs in an folder called core metrics results

qiime diversity core-metrics-phylogenetic \

--i-phylogeny rooted-tree.qza \

--i-table table.qza \

--p-sampling-depth 8545 \

--m-metadata-file metadata.txt \

--output-dir core-metrics-results

#alpha diversity significance

#this code will use the file for Shannon and Observed features in the core metrics results folder to create visualizations in order to see any significance for the alpha diversity

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/shannon\_vector.qza \

--m-metadata-file metadata.txt \

--o-visualization core-metrics-results/shannon\_vector.qzv

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/observed\_features\_vector.qza \

--m-metadata-file metadata.txt \

--o-visualization core-metrics-results/observed\_features\_vector.qzv

#this code will create beta diversity significance visualization for sex, population and flock using Bray Curtis dissimilarity file in the core-metrics folder and using the data from the metadata file

#the codes needed to be separated for each different subject

qiime diversity beta-group-significance \

--i-distance-matrix core-metrics-results/bray\_curtis\_distance\_matrix.qza \

--m-metadata-file metadata.txt \

--m-metadata-column sex \

--o-visualization core-metrics-results/bray-curtis-sex-significance.qzv \

--p-pairwise

qiime diversity beta-group-significance \

--i-distance-matrix core-metrics-results/bray\_curtis\_distance\_matrix.qza \

--m-metadata-file metadata.txt \

--m-metadata-column population \

--o-visualization core-metrics-results/bray-curtis-population-significance.qzv\

--p-pairwise

qiime diversity beta-group-significance \

--i-distance-matrix core-metrics-results/bray\_curtis\_distance\_matrix.qza \

--m-metadata-file metadata.txt \

--m-metadata-column flock \

--o-visualization core-metrics-results/bray-curtis-flock-significance.qzv \

--p-pairwise