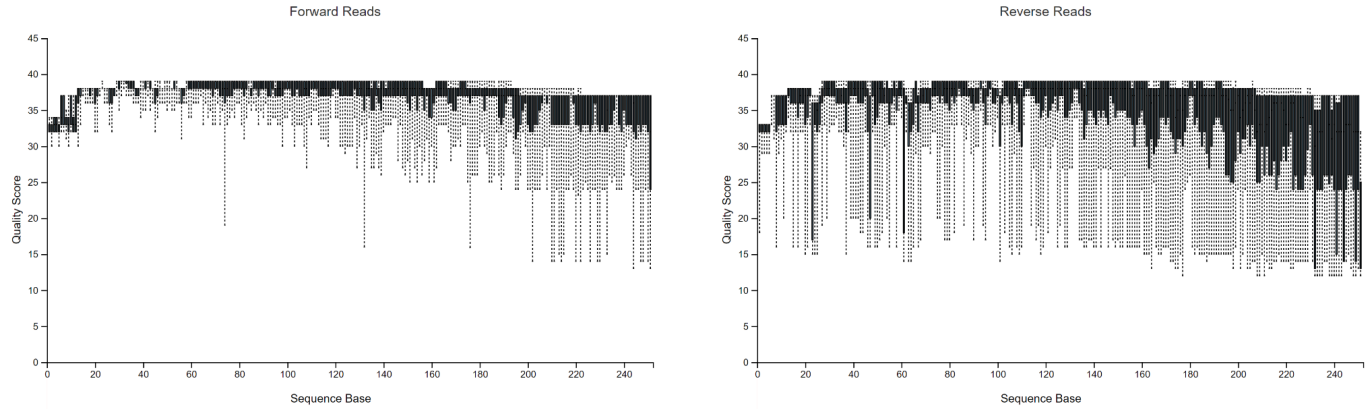
https://docs.qiime2.org/2024.2/tutorials/moving-pictures/

1) Include a screenshot of your interactive quality plot. Based on this plot, what values would you choose for `--p-trunc-len` and `--p-trim-left` for both the forward and reverse reads? Why have you chosen those numbers?



Based on this plot, I would choose –p-trunc-len to be 13 to 219 and --p-trim-left to be 13 to 205. I cut my reverse reads a little shorter because they were lower quality than my forward reads.

For questions 2 and 3: Because these are paired-end reads, you will have to modify the dada2 code in order to perform the quality trimming on both the forward and reverse reads. You will not do the deblur. You will need to adjust this code to account for `--p-trunc-len` and `--p-trim-left` for both the forward and reverse reads. The basics of the code you need to change are here.

```
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs demux.qza \
  --p-trim-left-f  \
  --p-trunc-len-f  \
  --p-trim-left-r  \
  --p-trunc-len-r  \
  --o-representative-sequences rep-seqs-dada2.qza \
  --o-table table-dada2.qza \
  --o-denoising-stats stats-dada2.qza`
```

2) How would you modify the code above to truncate and trim in your desired way?

```
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs demux.qza \
  --p-trim-left-f 0 \
  --p-trunc-len-f 249 \
  --p-trim-left-r 0 \
  --p-trunc-len-r 230 \
  --o-representative-sequences rep-seqs-dada2.qza \
  --o-table table-dada2.qza \
  --o-denoising-stats stats-dada2.qza
```

3) In the tutorial, you had to `mv` the files to rename them to just `rep-seqs.qza`, `table.qza`, and `stats.qza`. How could you modify the above code to skip that step? How do you need to modify `qiime metadata tabulate` in order to account for the renamed files being generated?

You can name your output files whatever you want:

```
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs demux.qza \
  --p-trim-left-f 0 \
  --p-trunc-len-f 249 \
  --p-trim-left-r 0 \
  --p-trunc-len-r 230 \
  --o-representative-sequences rep-seqs.qza \
  --o-table table.qza \
  --o-denoising-stats stats.qza
```

Here is how you would need to modify `qiime metadata tabulate` in order to account for the renamed files being generated (remove the -dada2 in the file names, same as above):

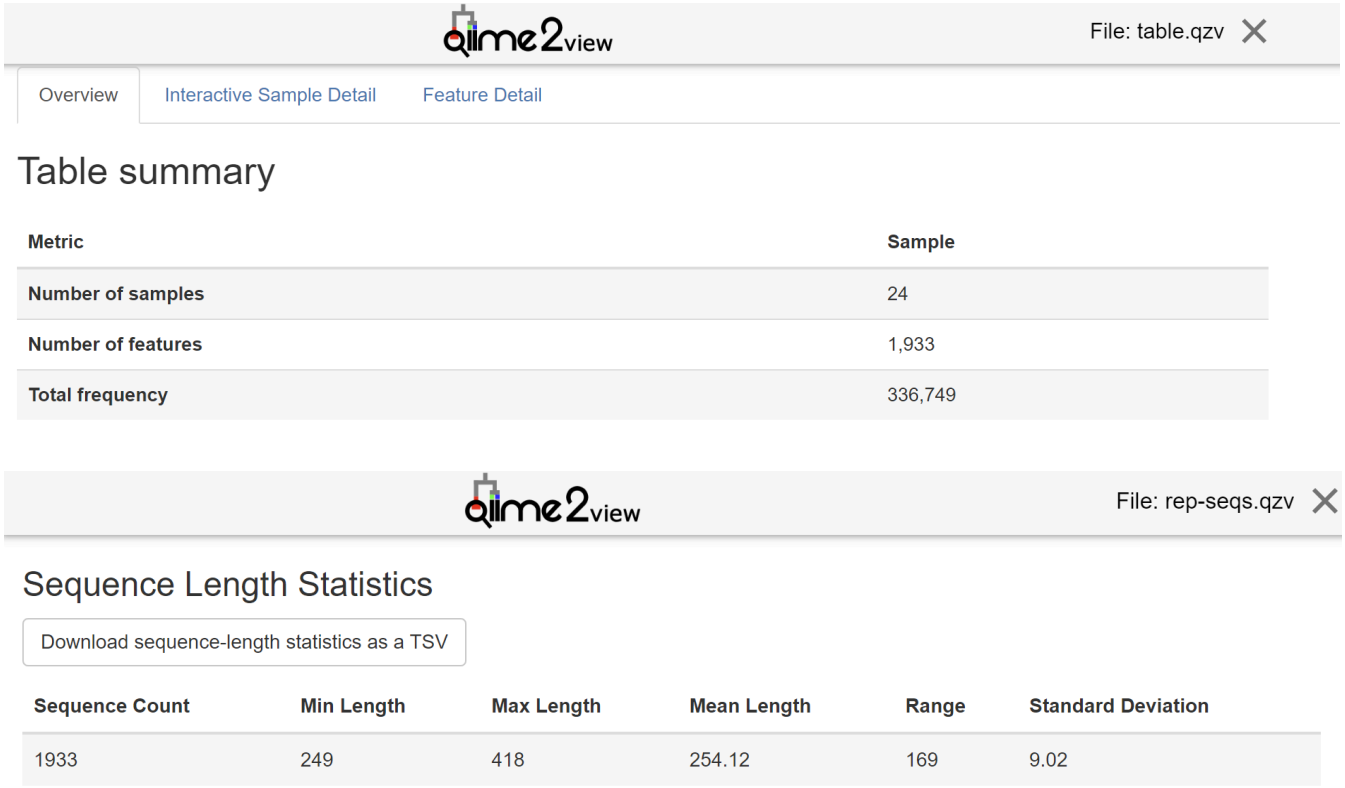
```
qiime metadata tabulate \
  --m-input-file stats.qza \
  --o-visualization stats.qzv
```

4) Your metadata file has a different name than that in the tutorial. How do you adjust your code in order to use the metadata file you have been given?

```
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
```

5) Include a screenshot of the table summary from visualizing your table and a screenshot of the sequence length statistics from the rep-seqs file.



6) Jump down to taxonomy. Once you have generated your taxonomy visualization, sort it by confidence. What are your top hits?

My top hits are sequences of mitochondrial origin.

qiime2view File: taxonomy.qzv Visualization Details Provenance

Feature ID (#q2:types)	Taxon (categorical)	Confidence (categorical)
366740d587f49e93540cfd17bf7a8589	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	1.0
7a999f4aa0ccd929825ed68bb9bbd5c3	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__MIZ46; f__; g__; s__	1.0
6804f86f07345226b90a3bef182e8e32	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999991
0d71c12227f08234d9185ff62fbae7cd	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999984
39970fbf18bfab4571bd0f32cac0482d	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999968
fc53a5208a8c6cb0f47461141ca5060e	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999956
72beafb48d2312340ebb3f1638801682	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999936
8b36c645b65b016e7c3539c78b17d8e2	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales	0.9999999999999494
5b9a6dbfc952e746732fc2d8d850487d	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae	0.9999999999999324
de6d76b596ef11e4329a247e295298cd	k__Bacteria; p__Chloroflexi; c__Thermomicrobia; o__Ellin6537; f__; g__; s__	0.9999999999999147
541dc80624fa12591e881959d1fd3e01	k__Bacteria; p__Armatimonadetes; c__Armatimonadia; o__FW68; f__; g__; s__	0.9999999999998721
55a36dc45106d8f105b9d0cf231b2321	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Legionellales; f__Coxiellaceae; g__Rickettsiella; s__	0.999999999999801

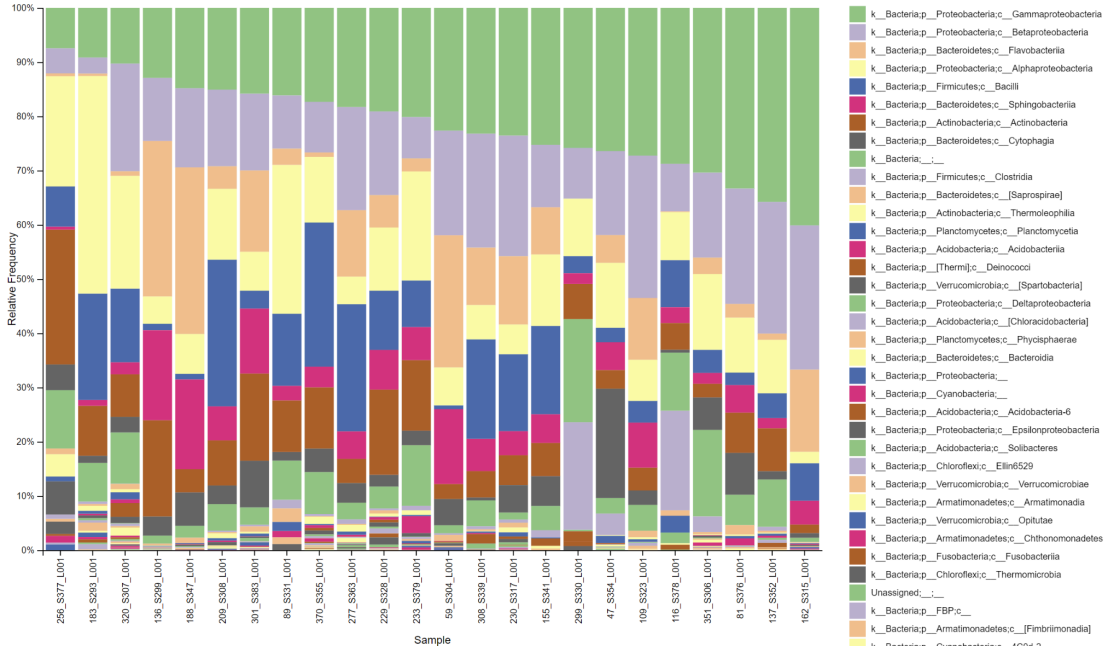
7) What do you think this code is doing? Why do you think this is a necessary or important step?
For question 7: Run this code

```
qiime taxa filter-table \
  --i-table table.qza \
  --i-taxonomy taxonomy.qza \
  --p-exclude mitochondria,chloroplast \
  --o-filtered-table table.qza
```

I think this code is going through my table of organisms and deleting anything that came from a mitochondria or chloroplast. This is important because we amplify bacterial ribosomes in PCR, we are also amplifying mt and chloroplast ribosomes. That will also amplify mt and chloroplasts seqs because their ribosomes are prokaryotic.

8) Re-do your table visualization and re-do your taxonomy commands. Do you have any differences now in the hits with the highest confidence? Why or why not? Really think about what the code is doing.
My top hits are the exact same: lots of mitochondrial sequences. This makes sense given our code. We are telling the computer to edit table.qza and exclude sequences that are assigned to mitochondria or chloroplasts. It successfully does this because (after visualizing the table) I went from having 1,911 features to 1,848. This means that 63 of my reads were from mitochondria or chloroplasts. When I re-run my taxonomy commands, nothing changes because my taxonomy commands do not reference my table file at any point. I use rep-seqs.qza to generate my taxonomy file. So I don't expect anything to change.

9) Looking at taxa bar plots, what are your top 2 phyla? Include a screenshot. What are the top 5 most abundant classes? Include a screenshot.
Top 2 phyla are Proteobacteria and Bacteroidetes.
Top 2 classes are Gammaproteobacteria and Betaproteobacteria.
From taxa-bar-plots.qzv



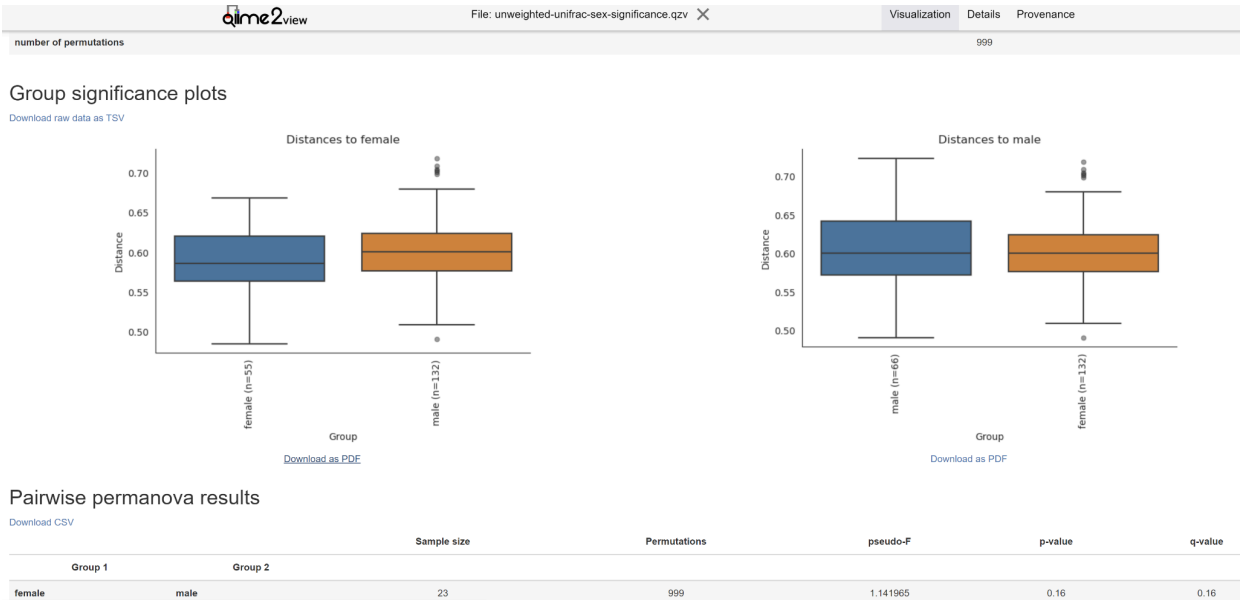
10) What is the difference between alpha and beta diversity? You will have to read outside resources to answer this question. Your response should be in your own words.
Alpha diversity measures what species are present in a community. It can be used to see what spp. two groups have in common and whether or not they are represented evenly. Beta diversity asks if two communities are the same, what species are there, and in what ratios. How different are these two communities. 0 = they're identical. 1 = no overlap in spp. at all.

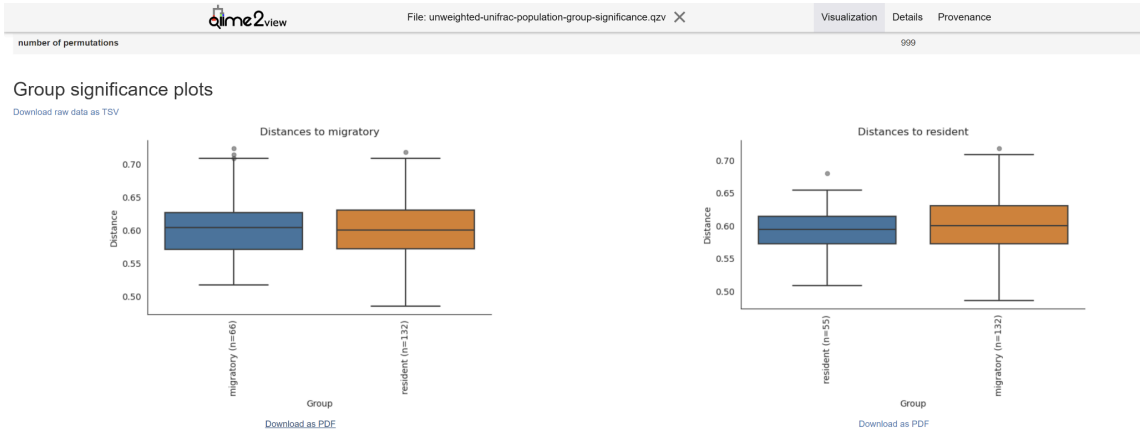
11) Before you calculate your diversity metrics, you have to choose a sampling depth. What file previously generated will you use to help you determine what to choose? Defend your choice of sampling depth. How many samples do you retain and how many do you lose?
I used table.qzv to view all my samples using "interactive sample detail." I chose a sampling depth of 7,850. The sample just under that one was less than half of that, with 3,806 reads. I thought that jump was large enough to warrant a cutoff, since there wasn't another jump that big where I would lose a reasonable number of samples. I lost 6 samples but retained 18.

12) For alpha diversity, you need to create visualizations for Shannon diversity and Observed features. This will require you to modify the `alpha-group-significance` code. For which metadata values were graphs generated? Were any of those comparisons significant? How do you know whether they were or were not significant? Briefly describe what Shannon diversity and Observed features are measuring (less than 1 paragraph).
Graphs were generated for the metadata values of population, sex, and flock.
A p-value of ≤0.05 is significant and >0.05 is not significant.
The Shannon diversity didn't show any significant comparisons.
The comparisons between all ♂ and all ♀ were significant, as well as between resident ♂ and resident ♀.

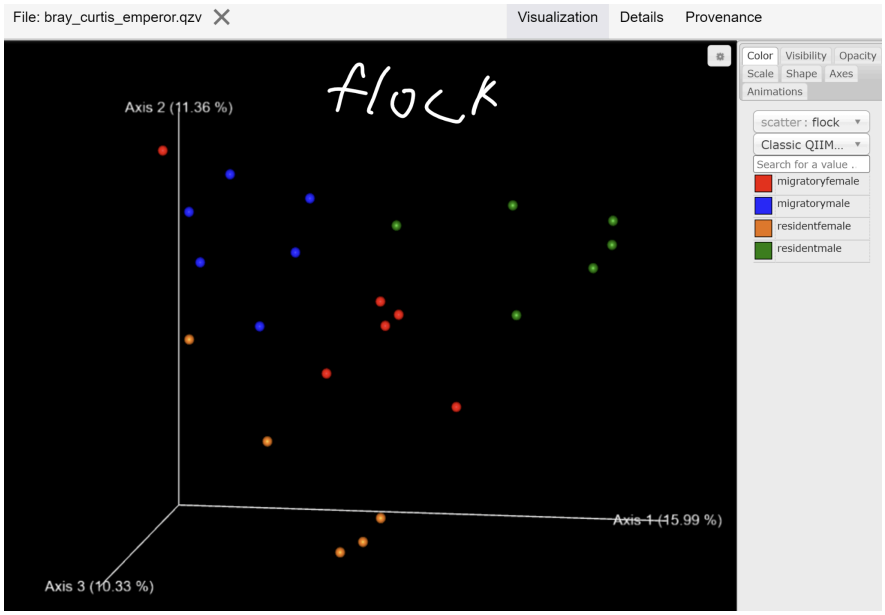
Shannon diversity is measuring community richness and evenness in terms of quantity. It is calculated using what percentage of the community is made up by each species. It doesn't consider phylogenetics.
Observed features are measuring community richness in terms of quality. It measure the # of unique features between 2 samples.

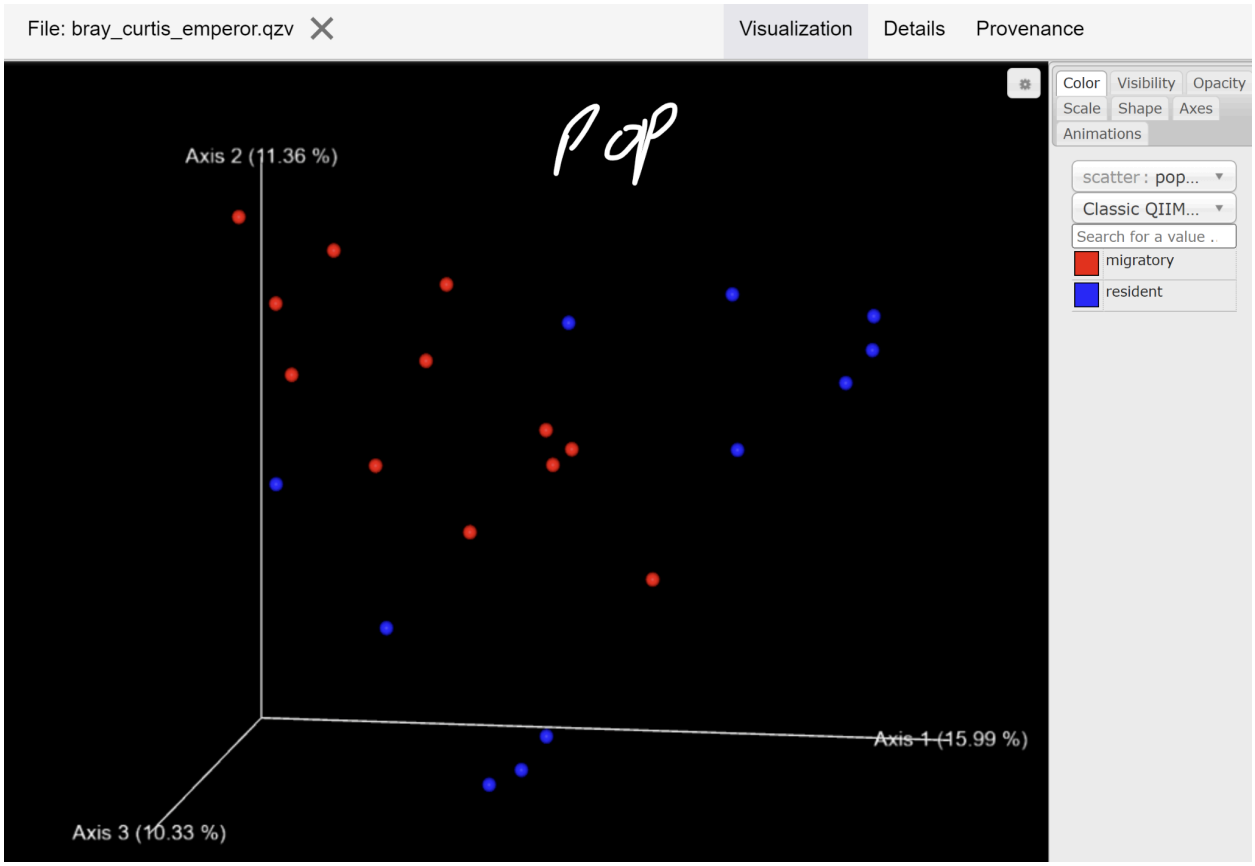
13) For beta diversity, you will need to create visualizations for Bray Curtis dissimilarity. This will require you to modify the `beta-group-significance` code. You should have one visualization for sex, one for population, and one for flock. Include a screenshot of each visualization. Is there any significance? Regardless of significance, how can you interpret these results (hint: what is beta diversity looking at?)
We can interpret these results as answering this question: how similar are these two populations?
We see that there are significant differences /no overlap between microbiomes of resident males and females because the p-value is less than 0.05.

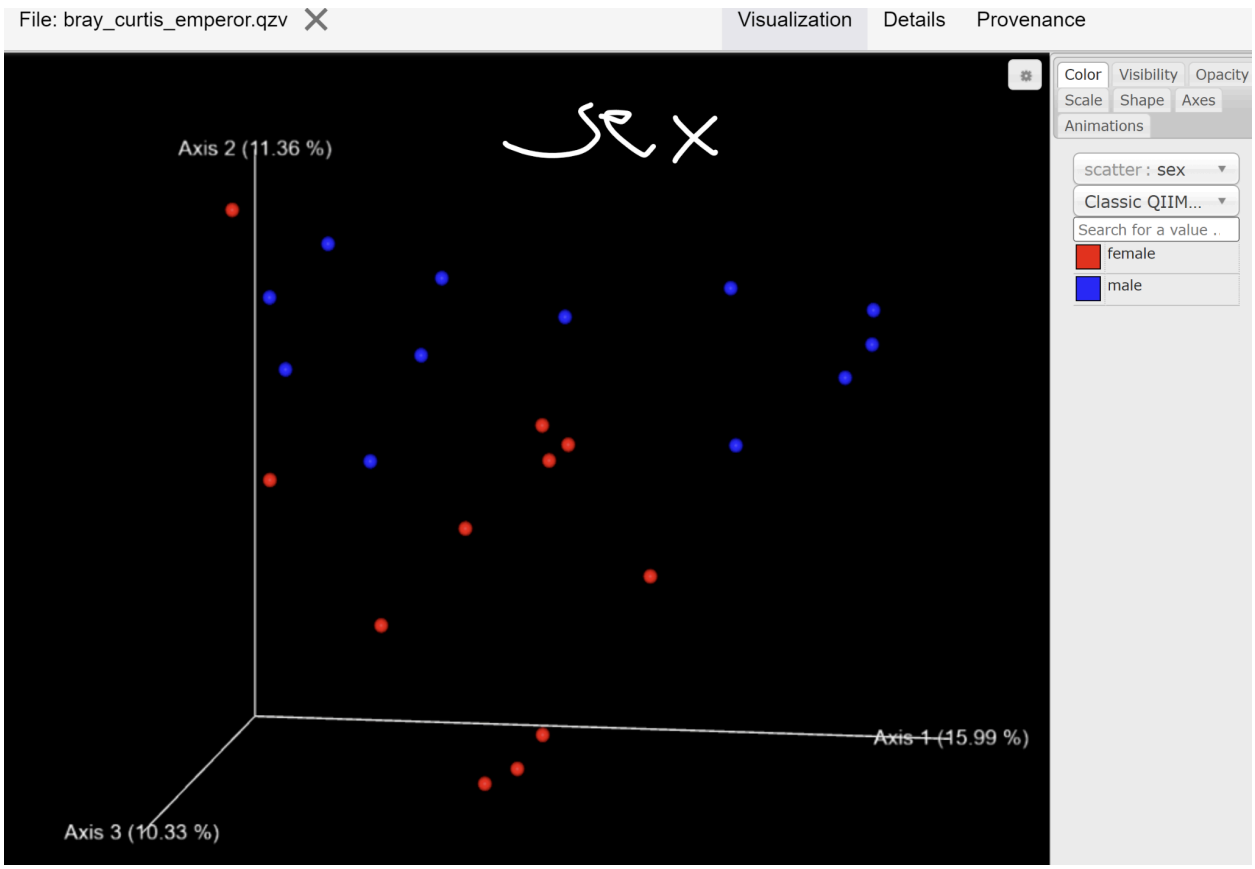




14) The `core-metrics-phylogeny` command generates a file called `bray-curtis-emporer.qzv.` Include 3 screenshots total (1 where the points are colored based on sex, one on population, one on flock). How do these results help you make sense of the results you got from question 13?




File: bray_curtis_emperor.qzv
Visualization
Details
Provenance
pop
Axis 2 (11.36 %)
Axis 1 (15.99 %)
Axis 3 (10.33 %)
Color
Visibility
Opacity
Scale
Shape
Axes
Animations
scatter : pop...
Classic QIIM...
Search for a value .
migratory
resident


File: bray_curtis_emperor.qzv
Visualization
Details
Provenance
sex
Axis 2 (11.36 %)
Axis 1 (15.99 %)
Axis 3 (10.33 %)
Color
Visibility
Opacity
Scale
Shape
Axes
Animations
scatter : sex
Classic QIIM...
Search for a value .
female
male