Introduction

This notebook contains all of the code from the corresponding post on the One Codex Blog. These snippets are exactly what are in the blog post, and let you perfectly reproduce those figures.

This is meant to be a starting off point for you to get started analyzing your own samples. You can copy this notebook straight into your account using the button in the header. To "run" or execute a cell, just hit `Shift + Enter`. A few other resources you may find useful include: notes on getting started with our One Codex library; the full documentation on our API (more technical); a cheat sheet on getting started with Pandas, a Python library for data manipulation; and reading a few of our blog posts (where we plan to have nice demos with these notebooks). As always, also feel free to send us questions or suggestions by clicking the chat icon in the bottom right!

Now we're going to dive right in and start crunching some numbers!

Fetching data

To get started, we create an instance of our API, grab the DIABIMMUNE project, and download 500 samples from the cohort.

classification_id	gender	host_age	geo_loc_name	totalige	eggs	vegetables	milk	wheat	rice
001b3ea2093b426d	Male	1093	Finland:Espoo	62.90	True	True	False	True	True
850ba22531cd4cde	Female	686	Russia:Petrozavodsk	91.50	True	True	True	False	True
db177a540a1c43b0	Female	673	Estonia:Tartu	112.00	True	True	True	True	True
d281a52b08f54b6a	Female	173	Russia:Petrozavodsk	NaN	False	False	False	False	False
a0cc5e58e2074fab	Male	493	Russia:Petrozavodsk	42.30	False	True	False	False	False
4b3aa0d6eabb48dc	Male	229	Estonia:Tartu	36.50	False	False	False	False	False
d0238007374a4dab	Male	502	Estonia:Tartu	7.39	True	True	True	True	True
8abf53a2cf4341fe	Male	390	Estonia:Tartu	698.00	True	True	True	True	True
8599190018b045c8	Male	427	Estonia:Tartu	88.10	True	True	True	True	True
11ded36641bf4be2	Female	587	Estonia:Tartu	25.20	False	True	True	True	True
ffb265bc656c4afc	Female	598	Estonia:Tartu	131.00	True	True	True	True	True
3bf2f958380647b5	Female	594	Estonia:Tartu	30.80	True	True	True	True	True
77cad5fb325f45fa	Male	410	Estonia:Tartu	7.39	True	True	True	True	True

classification_id	gender	host_age	geo_loc_name	totalige	eggs	vegetables	milk	wheat	rice
b84abe31b6bb4caf	Male	500	Estonia:Tartu	86.60	True	True	True	True	True
081424d940bd4cd9	Female	406	Estonia:Tartu	13.70	True	True	True	True	True
3aa4f451de8149a2	Female	278	Russia:Petrozavodsk	12.10	True	True	True	False	True
f3cd80e4ec4f4169	Male	483	Estonia:Tartu	698.00	True	True	True	True	True
4661a4510b124ee5	Female	500	Estonia:Tartu	23.90	True	True	True	False	False
61cd95290ed84876	Male	675	Estonia:Tartu	25.00	True	True	True	True	True
efafd7a62a6e434f	Male	479	Estonia:Tartu	88.10	True	True	True	True	True
12327f55ae0d4fb8	Male	400	Finland:Espoo	36.00	True	True	True	True	True
d4517cfd2a98419c	Male	649	Finland:Espoo	36.00	True	True	True	True	True
e81430008e1347e4	Male	588	Finland:Espoo	36.00	True	True	True	True	True
ddddd16149da436c	Male	212	Finland:Espoo	41.50	False	True	True	False	True
aa5ed3675cb54eb3	Male	304	Finland:Espoo	41.50	False	True	True	True	True
4ac0bfdeefc04893	Female	431	Finland:Espoo	4.16	True	True	True	True	True
7afcf62f9ea74a80	Male	703	Russia:Petrozavodsk	NaN	True	True	True	True	True
d9ce42afc60240c3	Male	1075	Finland:Espoo	36.00	True	True	True	True	True
f7283fc3621a4c8e	Female	593	Estonia:Tartu	193.00	True	True	True	True	True
344a8bcb2c73426b	Female	217	Estonia:Tartu	24.50	False	True	False	False	False
54986602b6fe4e0b	Female	495	Estonia:Tartu	5.43	True	True	True	True	True
df91a248bca34f8d	Male	210	Estonia:Tartu	24.00	True	True	True	True	True
2b8464bc86b448fa	Male	392	Finland:Espoo	127.00	True	True	True	True	True
8f0aaff9b67944e1	Female	669	Finland:Espoo	15.30	True	True	False	True	True
ce65d5efd36d4c14	Female	676	Finland:Espoo	92.20	True	True	True	True	True
ee361483f00549a7	Male	670	Finland:Espoo	24.30	True	True	True	True	True
c000446b505e4d1d	Male	556	Finland:Espoo	24.30	True	True	True	True	True
139c880885a544da	Male	539	Russia:Petrozavodsk	NaN	True	True	True	True	True
2d287a0836964fef	Female	541	Russia:Petrozavodsk	NaN	True	True	True	True	True
8dbdbdd3e7e0438e	Male	743	Russia:Petrozavodsk	58.00	False	False	False	False	False
c19c28ff39c54cbc	Male	535	Russia:Petrozavodsk	19.40	True	True	True	True	True
1693d10e542d4d21	Male	217	Russia:Petrozavodsk	13.00	True	True	False	False	True
f47fd5aa29a5434f	Female	434	Russia:Petrozavodsk	10.30	False	False	False	False	False
8aa0a16b36cd4f3f	Male	400	Russia:Petrozavodsk	30.20	False	False	False	False	False

classification_id	gender	host_age	geo_loc_name	totalige	eggs	vegetables	milk	wheat	rice
c053a1cc63fe4752	Male	498	Russia:Petrozavodsk	30.20	False	False	False	False	False
b1e800f58204406b	Female	286	Russia:Petrozavodsk	2.00	False	True	True	False	True
20b6217e78d643a4	Female	367	Russia:Petrozavodsk	10.30	False	False	False	False	False
0e6afb347fa14281	Female	248	Russia:Petrozavodsk	12.10	False	True	True	False	True
1e8467e5f1784fcf	Male	213	Russia:Petrozavodsk	10.60	False	False	True	False	True
0466bf5d5a4145c5	Female	519	Russia:Petrozavodsk	8.91	False	False	False	False	False

Question #1: How does alpha diversity vary by sample group?

Here, we display observed taxa, Simpson's Index, and Shannon Entropy side-by-side, grouped by the region of birth. Each group includes samples taken across the entire three-year longitudinal study.

```
2024-08-20 17:15:52,253 WARNING: observed_otus is deprecated as of 0.6.0.
```

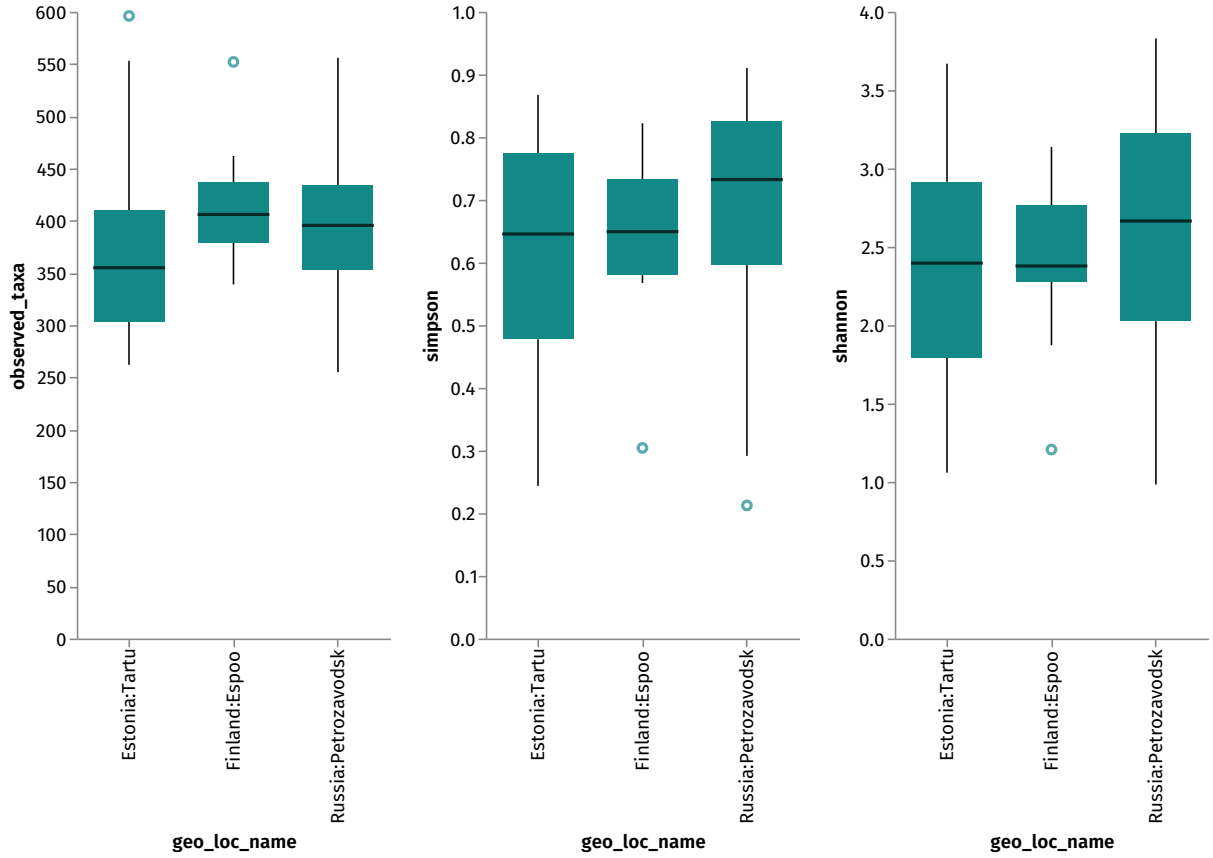
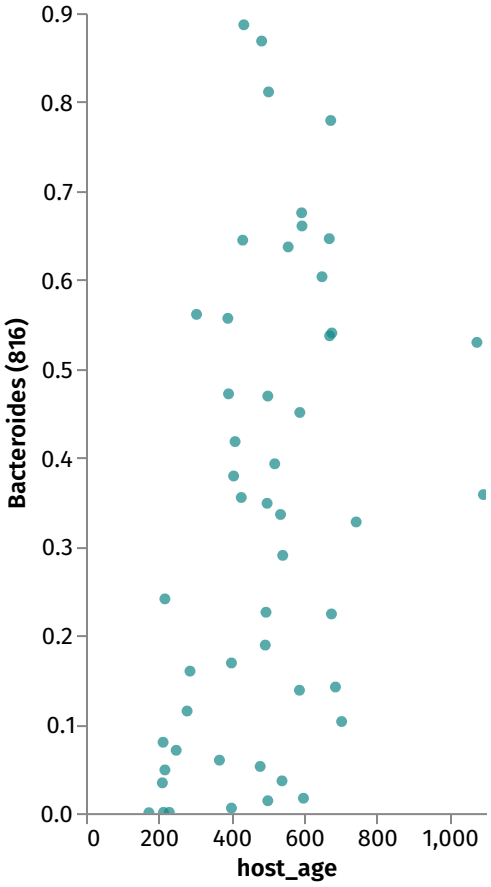


Figure 1. *Alpha diversity by location of birth*[1]

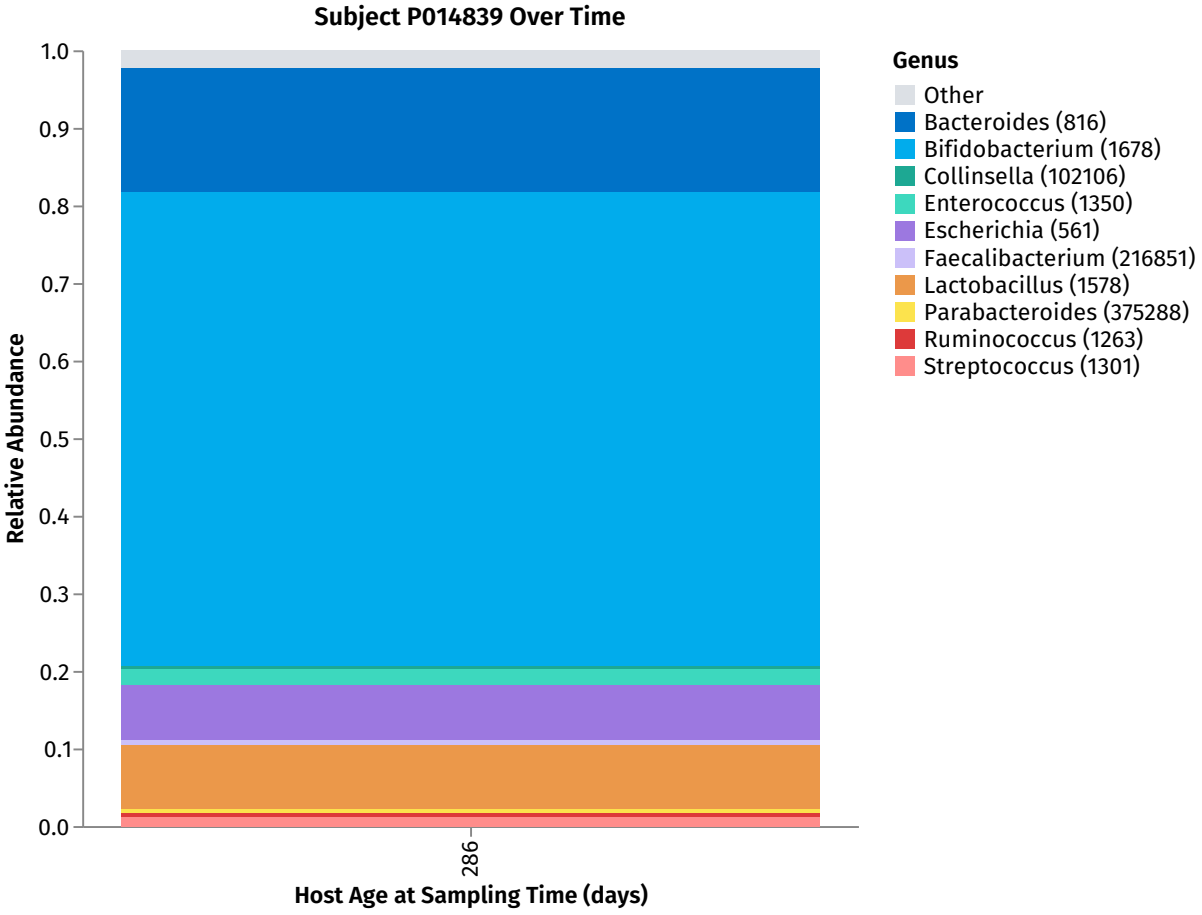
Question #2: How does the microbiome change over time?

The `plot_metadata` function can search through all taxa in your samples and pull out read counts or relative abundances.



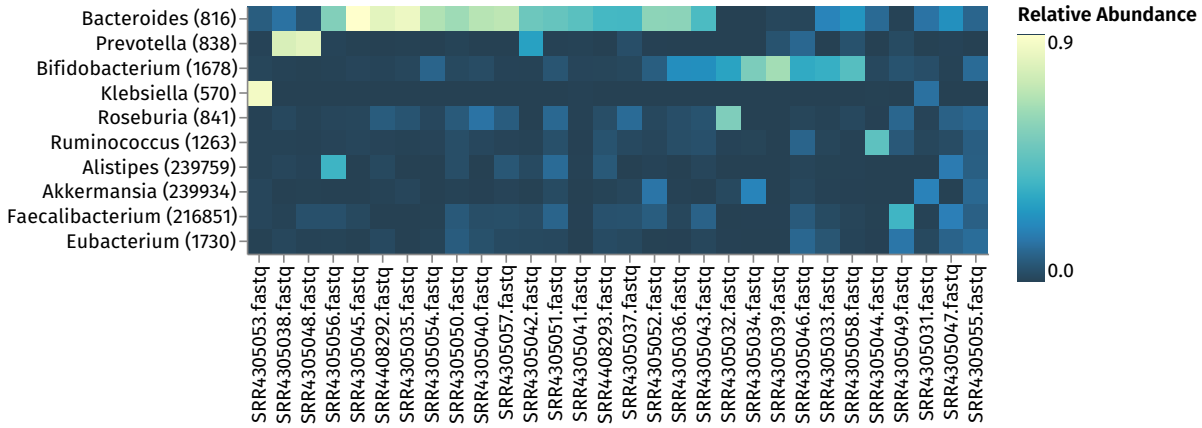
Question #3: How does an individual subject's gut change over time?

Here, we're going to drop into a dataframe, slice it to fetch all the data points from a single subject of the study, and generate a stacked bar plot. It's nice to see the expected high abundance of Bifidobacterium early in life, giving way to Bacteroides near age three!



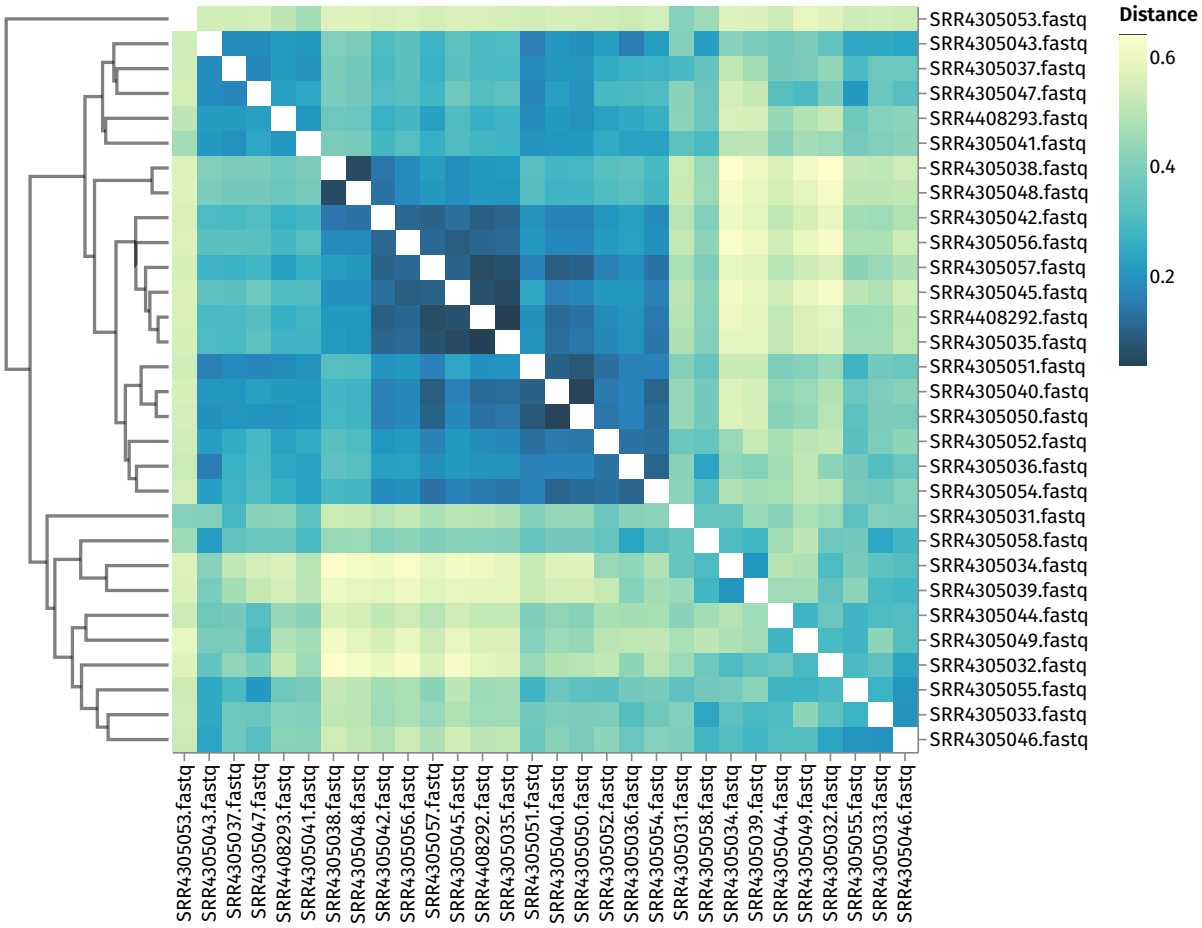
Question #4: Heatmaps?!

```
WARN Scale bindings are currently only supported for scales with unbinned,
continuous domains.
WARN Scale bindings are currently only supported for scales with unbinned,
continuous domains.
```

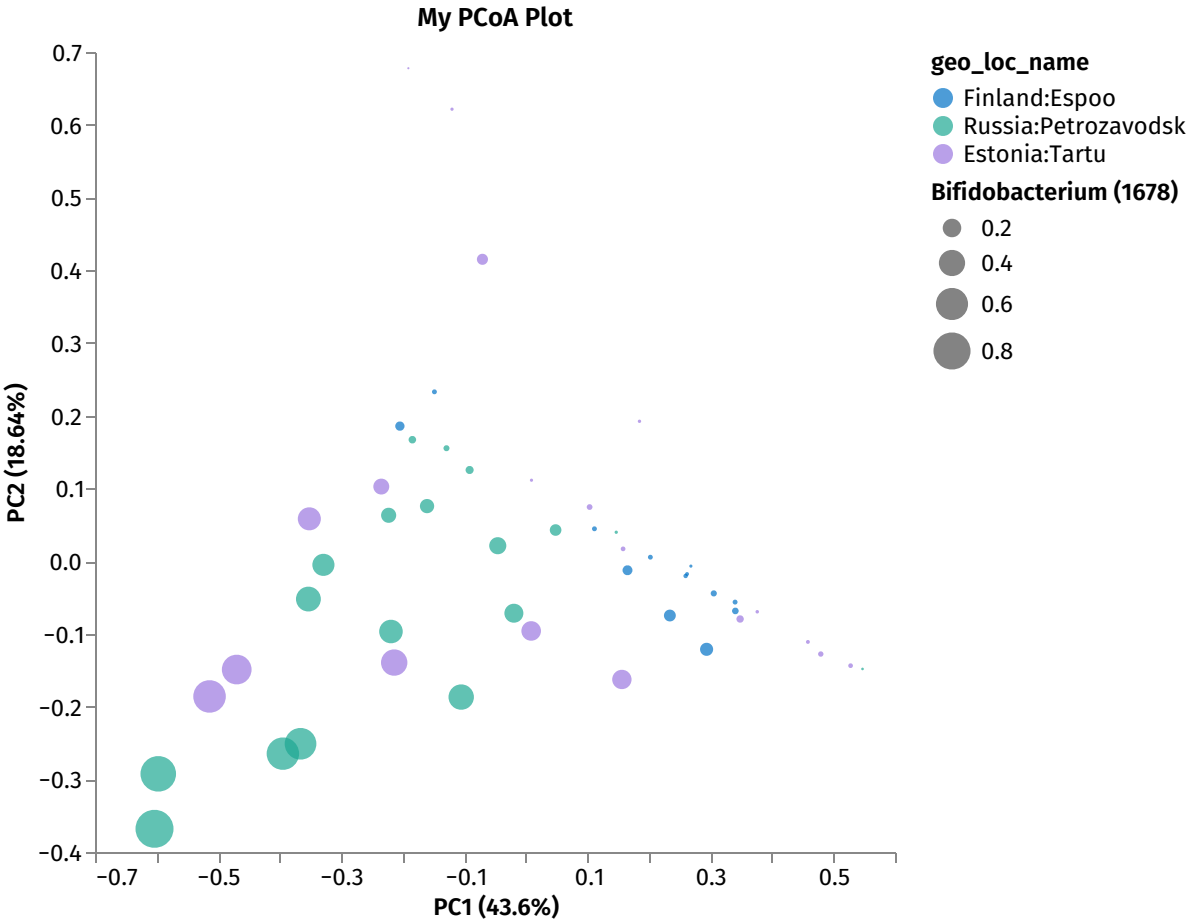


Question #5: How do samples cluster?

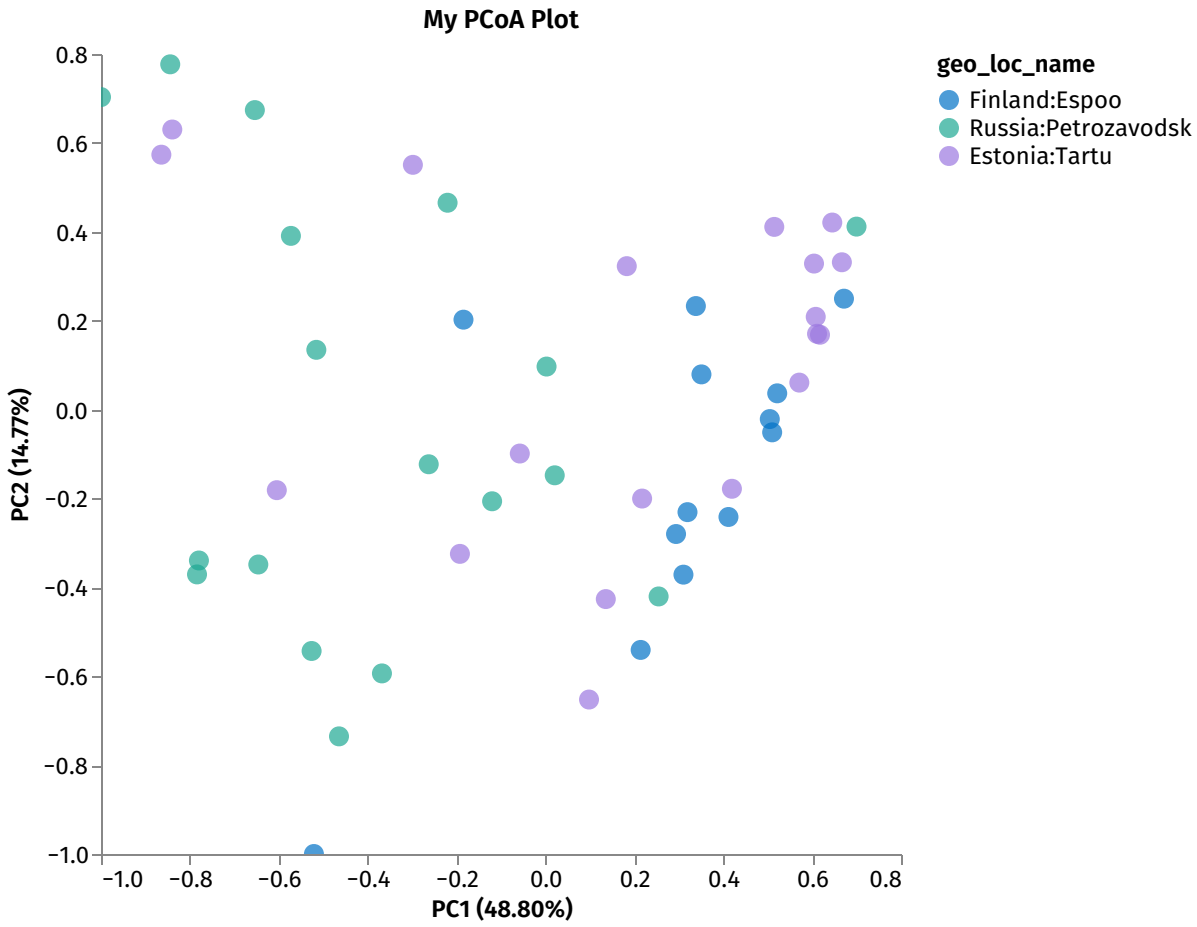
First up, we'll plot a heatmap of weighted UniFrac distance between the first 30 samples in the dataset. This requires unnormalized read counts, so we'll generate a new, unnormalized dataframe.



Question #6: Can I do PCA?



Question #6: Can I do something *better* than PCA?


My PCoA Plot
geo_loc_name
Finland:Espoo
Russia:Petrozavodsk
Estonia:Tartu
PC2 (14.77%)
PC1 (48.80%)
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8

References

1. Roo, et al. "How to Python." Nature, 2019.