

Periodontitis

Seunghoon Kim

Jaewoong Lee

Semin Lee

2020-11-26

Contents

1	Introduction	4
1.1	Microbiome	4
1.2	Ribosomal RNA	4
1.3	16S rRNA Gene Sequencing	4
1.4	Periodontitis	4
2	Materials	4
2.1	16S rRNA Gene Sequencing	4
3	Methods	4
3.1	QIIME2 Workflow	4
3.1.1	Denoising techniques	4
3.1.2	Taxonomy Classification	4
3.1.3	Rarefaction	4
3.1.4	Alpha-diversity	4
3.1.5	Beta-diversity	7
3.1.6	ANCOM	7
3.2	Python Packages	7
3.2.1	Pandas	7
3.2.2	Scikit-learn	7
3.2.3	Matplotlib	7
3.2.4	Seaborn	7
4	Results	7
4.1	Quality Filter	7
4.2	Rarefaction	7
4.3	Alpha-diversity	7
4.4	Beta-diversity	7
4.5	ANCOM	7
5	Discussion	7
6	References	7

List of Tables

List of Figures

1	Concept of a Core Human Microbiome (Turnbaugh et al., 2007)	5
2	A Theoretic Overview of QIIME2 Workflow (Bolyen et al., 2019, 2018)	5
3	Denoising Techniques which provided by QIIME2	6
4	Taxonomy Classification which provided by QIIME2	6
5	Example ANCOM Volcano Plot which Provided by QIIME2 (Bolyen et al., 2019, 2018)	8
6	Sequence Quality Plot	8
7	Frequency per Sample by DADA2	8
8	Frequency per Sample by DADA2	9
9	Evenness Index from DADA2	9
10	Faith PD Index from DADA2	9
11	Observed Features Index from DADA2	10
12	Shannon's Diversity Index from DADA2	10
13	Evenness Index from Deblur	10
14	Faith PD Index from Deblur	11
15	Observed Features Index from Deblur	11
16	Shannon's Diversity Index from Deblur	11
17	Bray-Curtis Distance Index with DADA2	12
18	Jaccard Distance Index with DADA2	12
19	Unweighted Unifrac Distance Index with DADA2	13
20	Weighted Unifrac Distance Index with DADA2	13

21	Bray-Curtis Distance Index with Deblur	14
22	Jaccard Distance Index with Deblur	14
23	Unweighted Unifrac Distance Index with Deblur	15
24	Weighted Unifrac Distance Index with Deblur	15
25	ANCOM Volcano Plot with DADA2 and Greengenes	16
26	ANCOM Volcano Plot with DADA2 and SILVA	16
27	ANCOM Volcano Plot with Deblur and Greengenes	16
28	ANCOM Volcano Plot with Deblur and SILVA	18

1 Introduction

1.1 Microbiome

1.2 Ribosomal RNA

Ribosomal RNA (rRNA)

1.3 16S rRNA Gene Sequencing

1.4 Periodontitis

2 Materials

2.1 16S rRNA Gene Sequencing

- 100 Healthy samples
- 50 Chronic Early Periodontitis Sample
- 50 Chronic Moderate Periodontitis Sample
- 50 Chronic Severe Periodontitis Sample

3 Methods

3.1 QIIME2 Workflow

QIIME2 is a capable, expandable and distributed microbiome analysis package with transparent analysis (Bolyen et al., 2019, 2018). A theoretic overview of QIIME2 workflow is shown as figure 2.

3.1.1 Denoising techniques

There are two denoising techniques provided by QIIME2: DADA2 (Callahan et al., 2016) and Deblur (Amir et al., 2017). Major difference between DADA2 and Deblur is a strategy, the strategy used to divide as different species.

3.1.2 Taxonomy Classification

There are two taxonomy classification databases which provided by QIIME2: Greengenes (GG) (DeSantis et al., 2006) and SILVA (Pruesse et al., 2007).

3.1.3 Rarefaction

Rarefaction is a statistical method of estimating the number of species expected in a random sample which taken from a collection (James & Rathbun, 1981). Moreover, rarefaction allows comparisons of the species richness among communities. Thus, rarefaction is a good choice for normalization (Weiss et al., 2017).

3.1.4 Alpha-diversity

Alpha-diversity is a metric which shows the richness of taxa at a single community. There are four alpha-diversity indices which provided from QIIME2:

- Shannon's diversity index.
- Observed features.
- Faith's phylogenetic diversity.
- Evenness index.



Figure 1: Concept of a Core Human Microbiome (Turnbaugh et al., 2007)



Figure 2: A Theoretic Overview of QIIME2 Workflow (Bolyen et al., 2019, 2018)

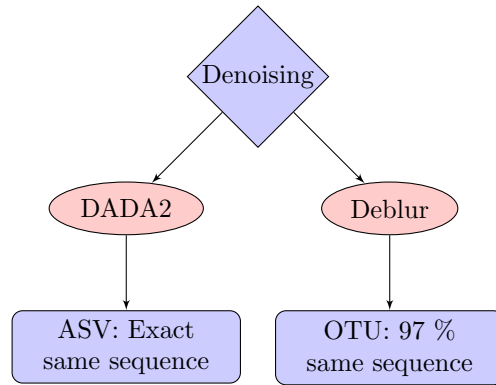


Figure 3: Denoising Techniques which provided by QIIME2

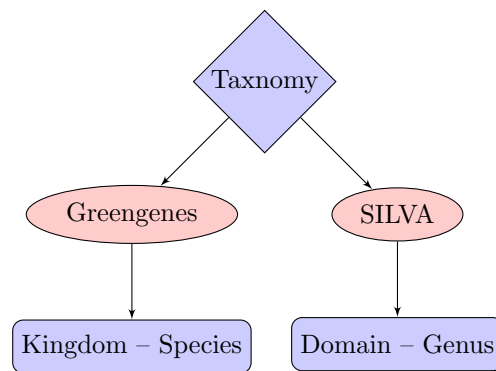


Figure 4: Taxonomy Classification which provided by QIIME2

3.1.5 Beta-diversity

Beta-diversity is a metric which indicates the taxonomic differentiation between multiple communities. There are four beta-diversity indices which provided from QIIME2:

- Jaccard distance.
- Bray-Curtis distance.
- Unweighted UniFrac distance.
- Weighted UniFrac distance.

3.1.6 ANCOM

ANCOM (Analysis of composition of microbiomes) can be used for analyzing the composition of microbiome in multiple populations (Mandal et al., 2015). Example ANCOM volcano plot is shows as figure 5.

3.2 Python Packages

3.2.1 Pandas

Pandas is a Python package of rich data structures and tools for analyzing with structured data sets (McKinney et al., 2011).

3.2.2 Scikit-learn

Scikit-learn grants state-of-the-art implementation of many machine learning algorithms, while controlling an easy-to-use interface tightly integrated the Python code (Pedregosa et al., 2011).

3.2.3 Matplotlib

Matplotlib is a Python graphics package which used for application development, interactive scripting and publication quality image generation (Barrett, Hunter, Miller, Hsu, & Greenfield, 2005). Matplotlib, also, is designed to create simple plots with a few commands (Hunter, 2007).

3.2.4 Seaborn

Seaborn is a Python data visualization package which based on matplotlib, allows a high-level interface for displaying engaging and descriptive statistical graphics (Waskom & the seaborn development team, 2020).

4 Results

4.1 Quality Filter

4.2 Rarefaction

4.3 Alpha-diversity

4.4 Beta-diversity

4.5 ANCOM

5 Discussion

6 References

- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Xu, Z. Z., ... others (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, 2(2).
- Barrett, P., Hunter, J., Miller, J. T., Hsu, J.-C., & Greenfield, P. (2005). matplotlib—a portable python plotting package. In *Astronomical data analysis software and systems xiv* (Vol. 347, p. 91).
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., ... others (2018). *Qiime 2: Reproducible, interactive, scalable, and extensible microbiome data science* (Tech. Rep.). PeerJ Preprints.

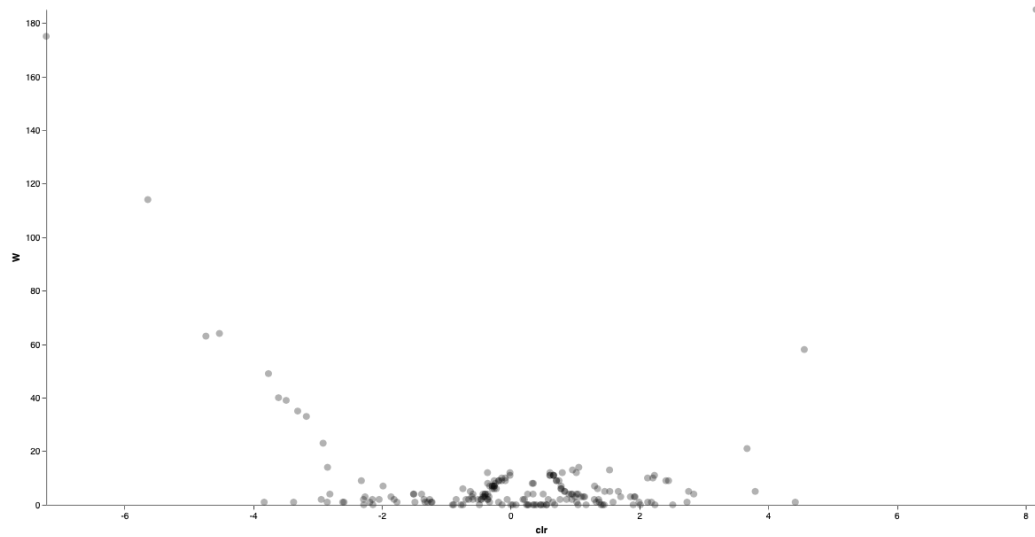


Figure 5: Example ANCOM Volcano Plot which Provided by QIIME2 (Bolyen et al., 2019, 2018)

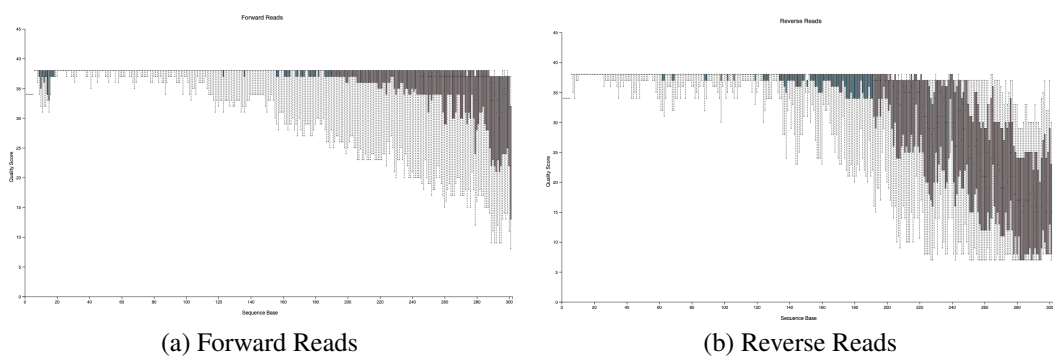


Figure 6: Sequence Quality Plot

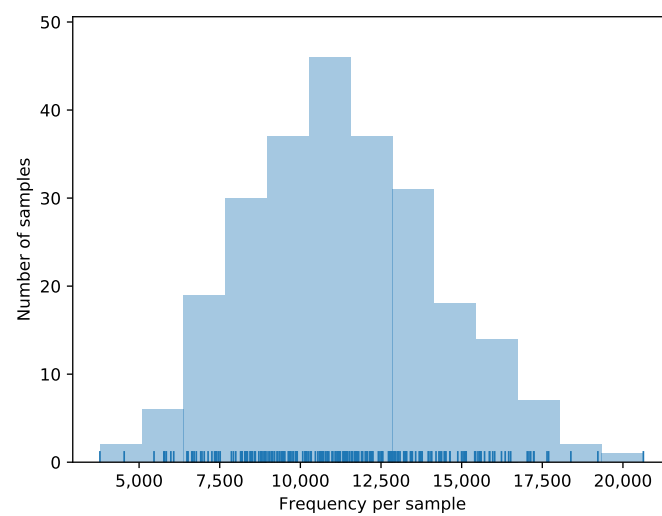


Figure 7: Frequency per Sample by DADA2

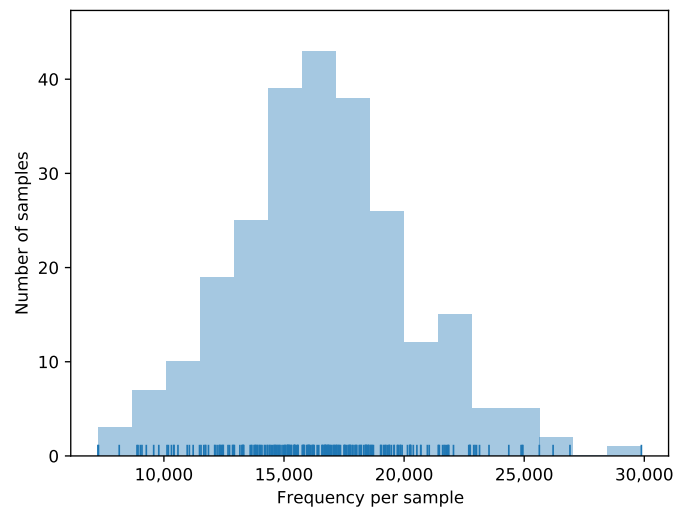


Figure 8: Frequency per Sample by DADA2

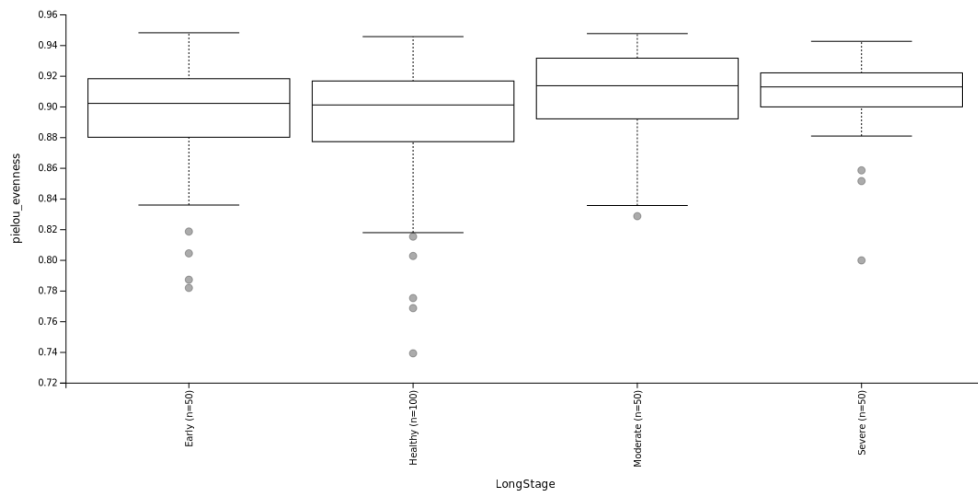


Figure 9: Evenness Index from DADA2

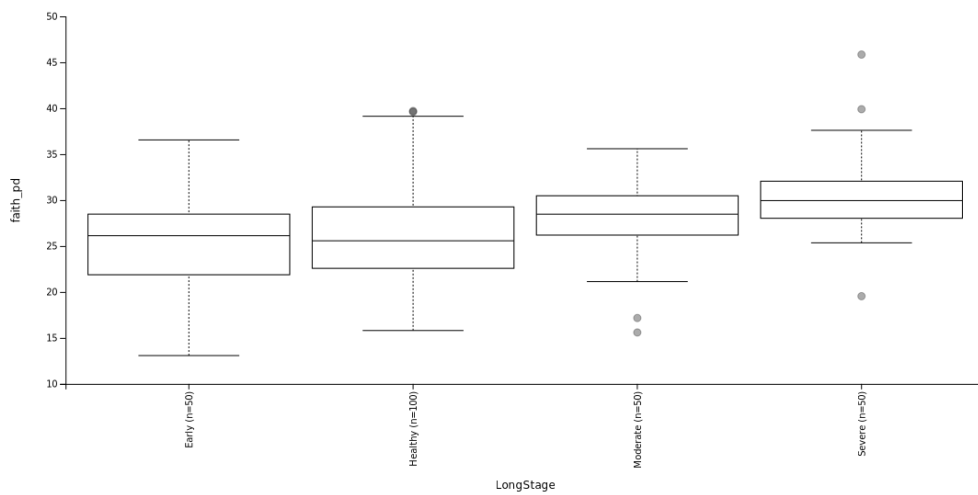


Figure 10: Faith PD Index from DADA2

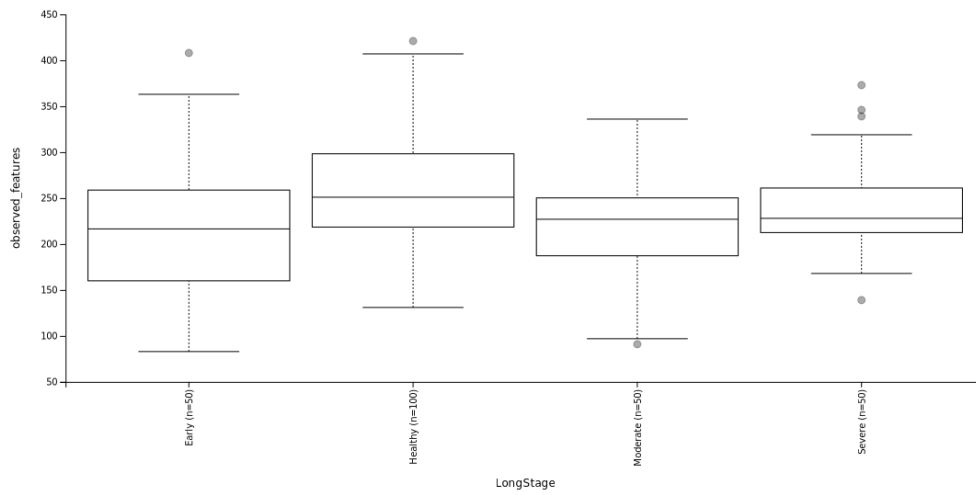


Figure 11: Observed Features Index from DADA2

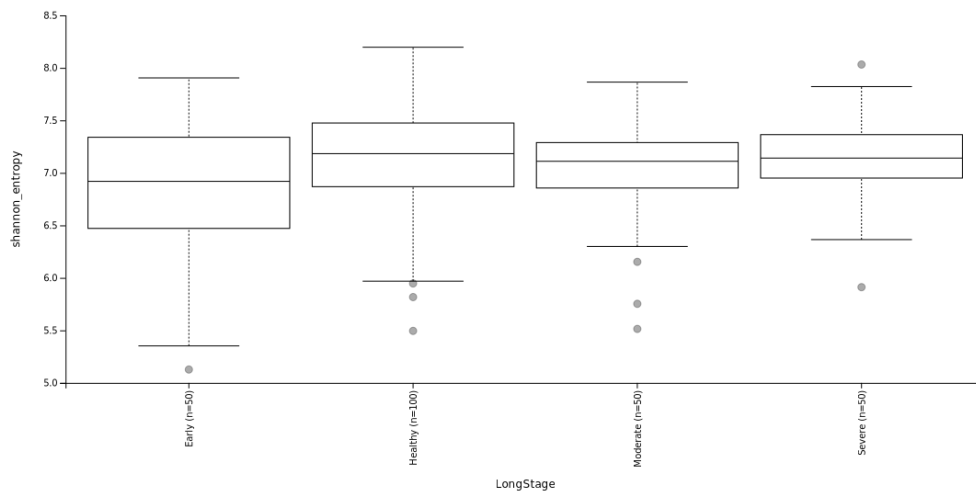


Figure 12: Shannon's Diversity Index from DADA2

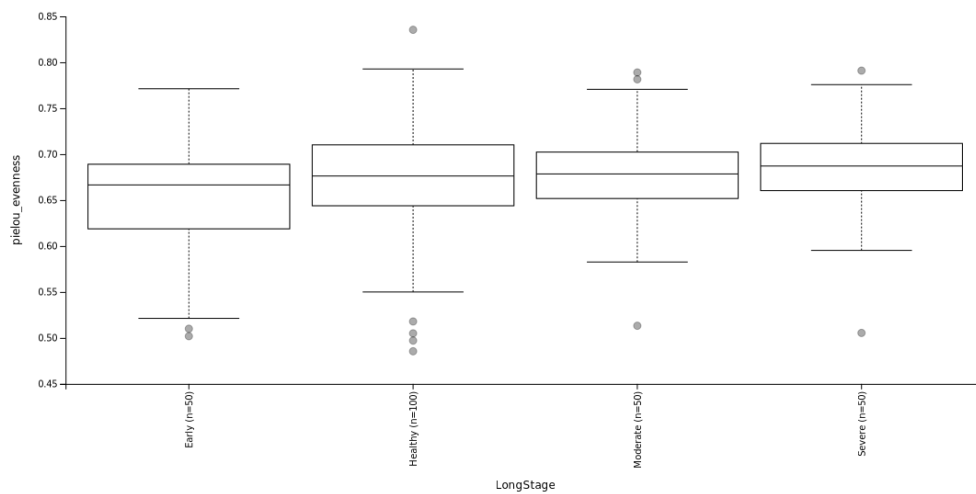


Figure 13: Evenness Index from Deblur

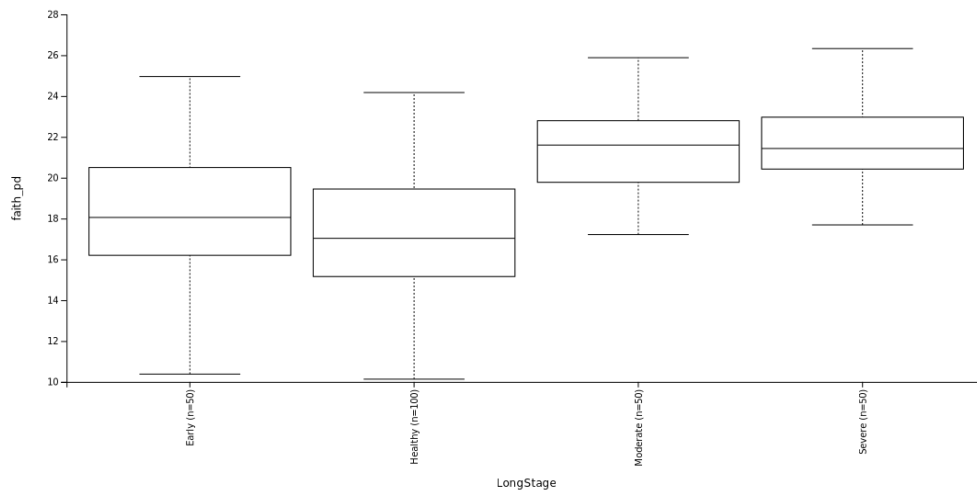


Figure 14: Faith PD Index from Deblur

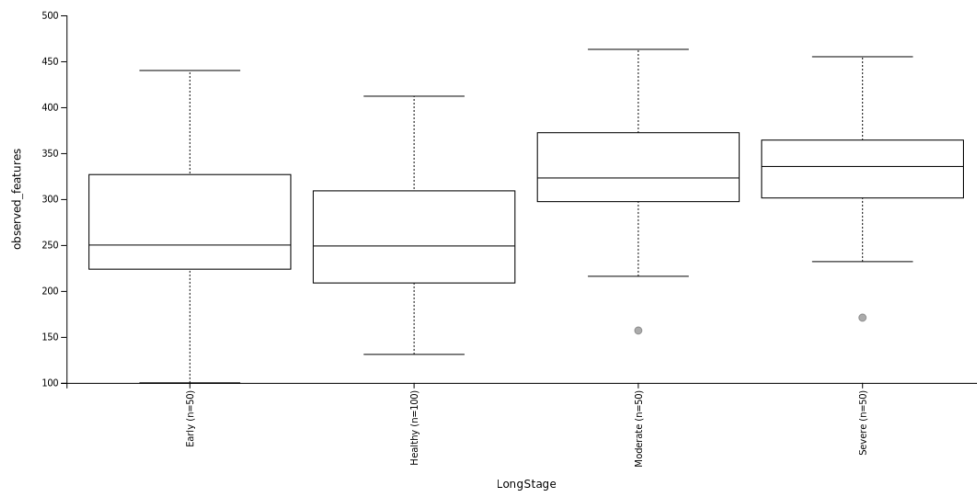


Figure 15: Observed Features Index from Deblur

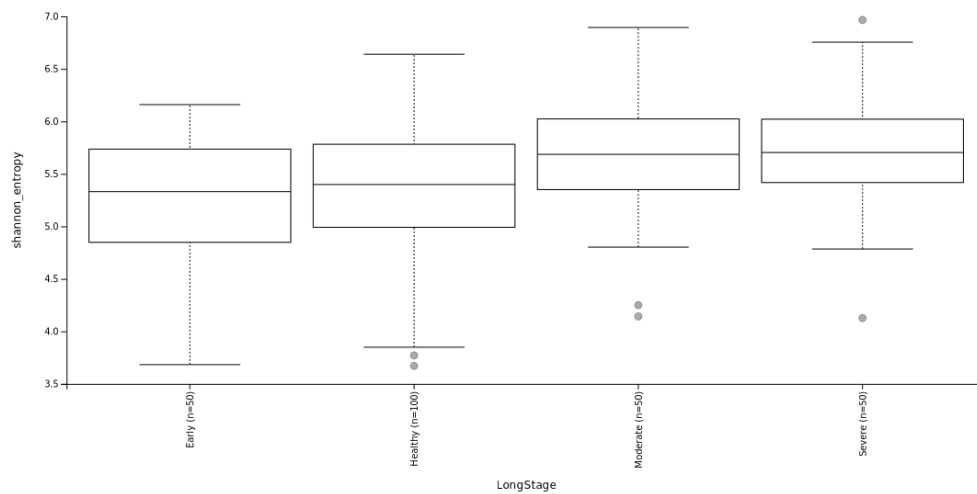


Figure 16: Shannon's Diversity Index from Deblur

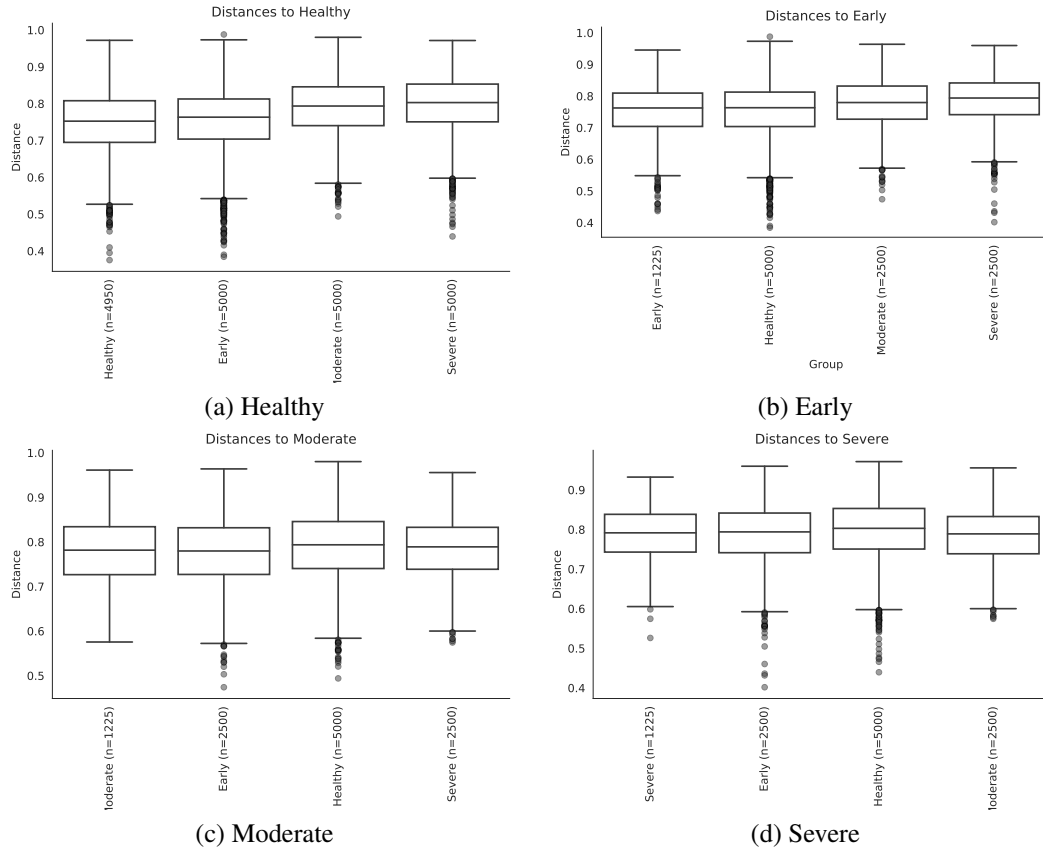


Figure 17: Bray-Curtis Distance Index with DADA2

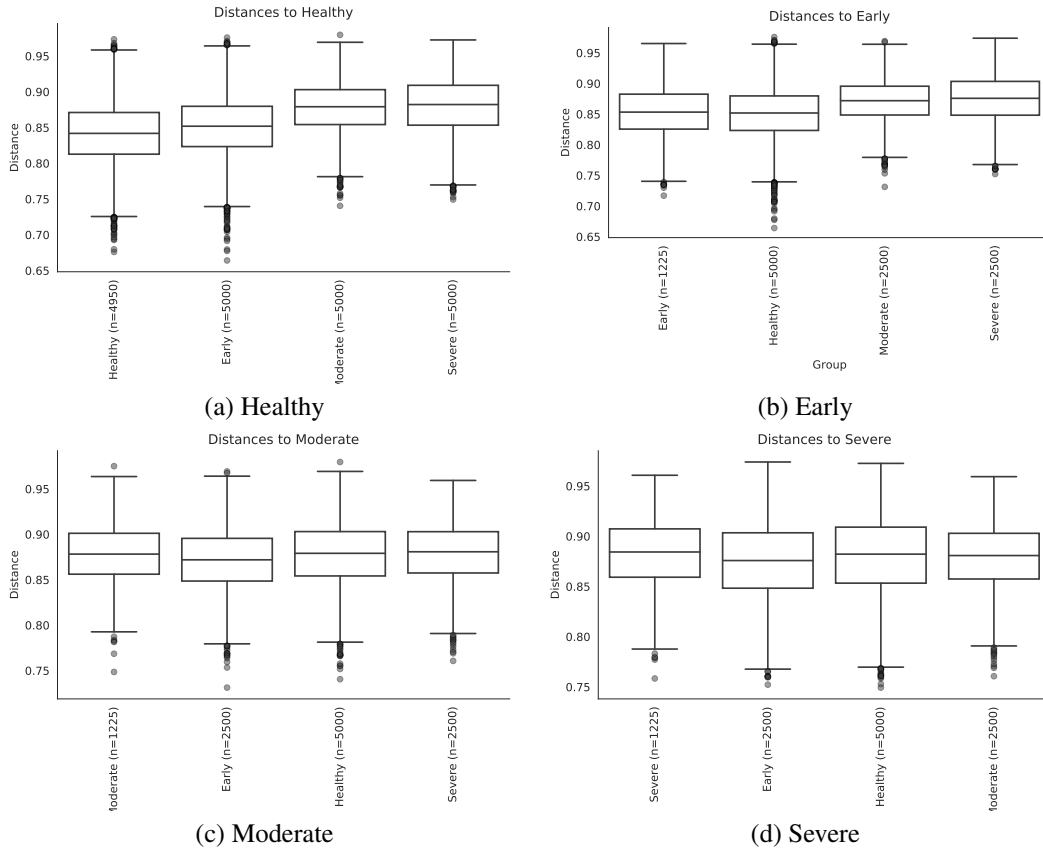


Figure 18: Jaccard Distance Index with DADA2

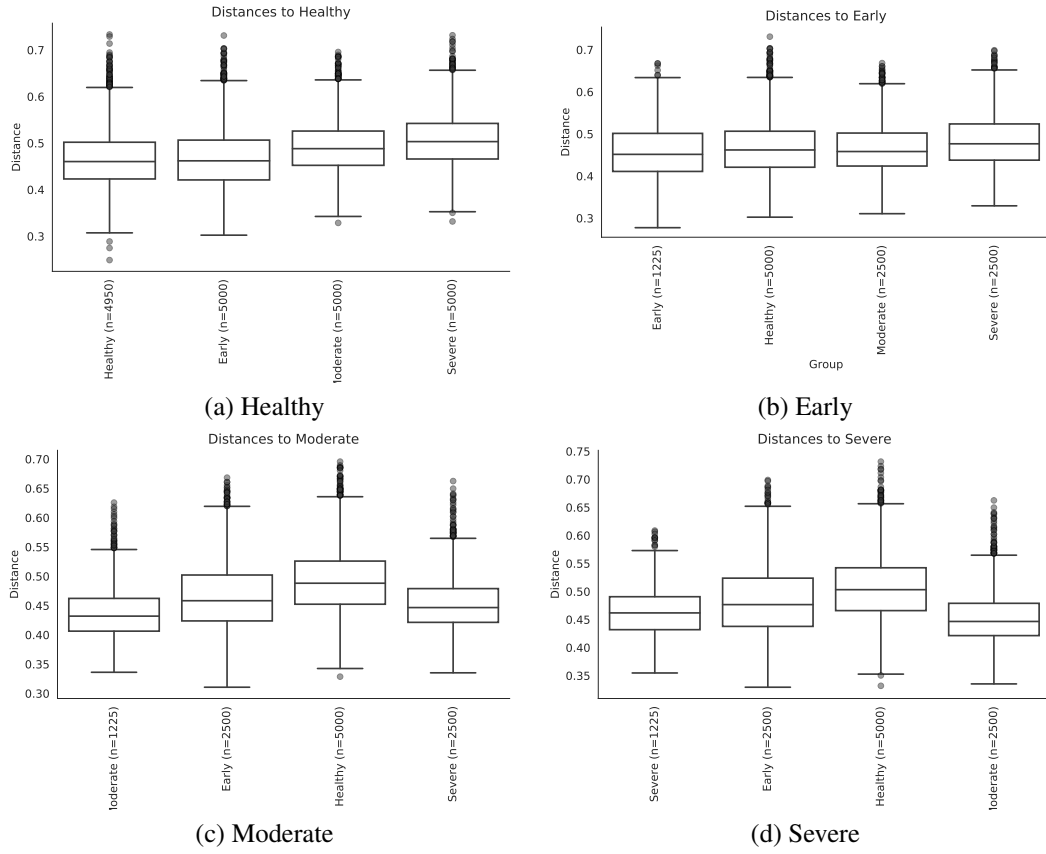


Figure 19: Unweighted Unifrac Distance Index with DADA2

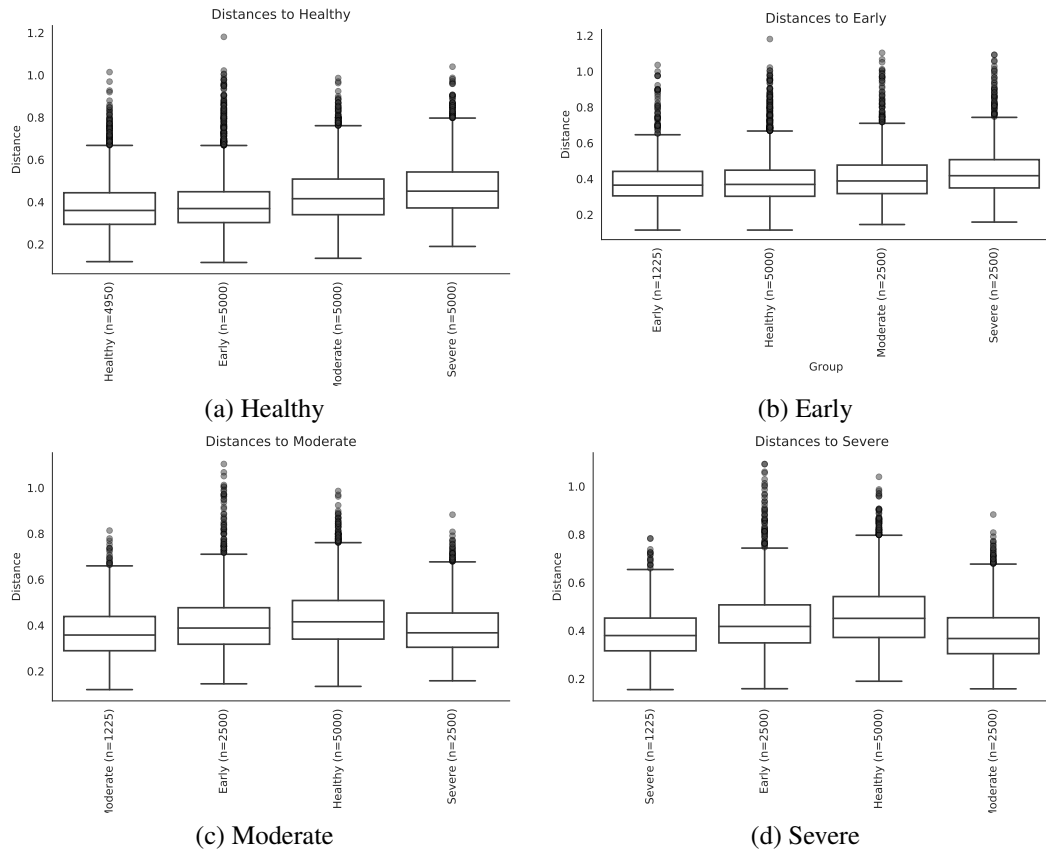


Figure 20: Weighted Unifrac Distance Index with DADA2

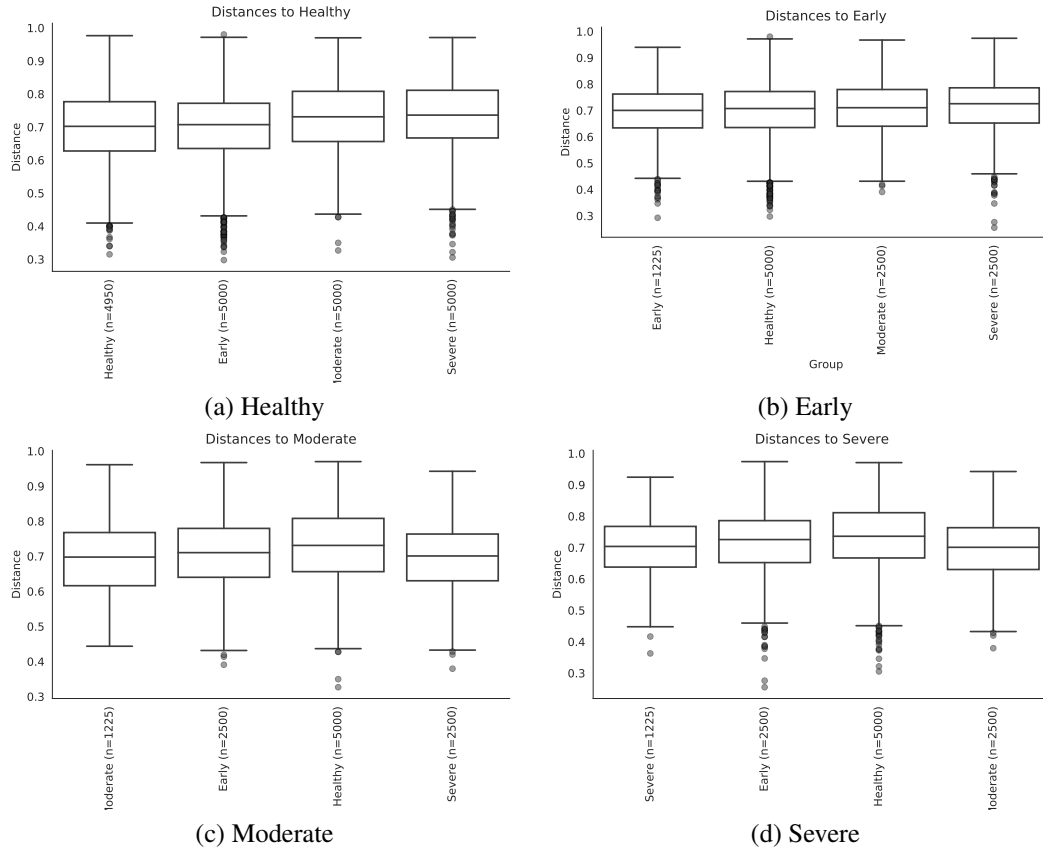


Figure 21: Bray-Curtis Distance Index with Deblur

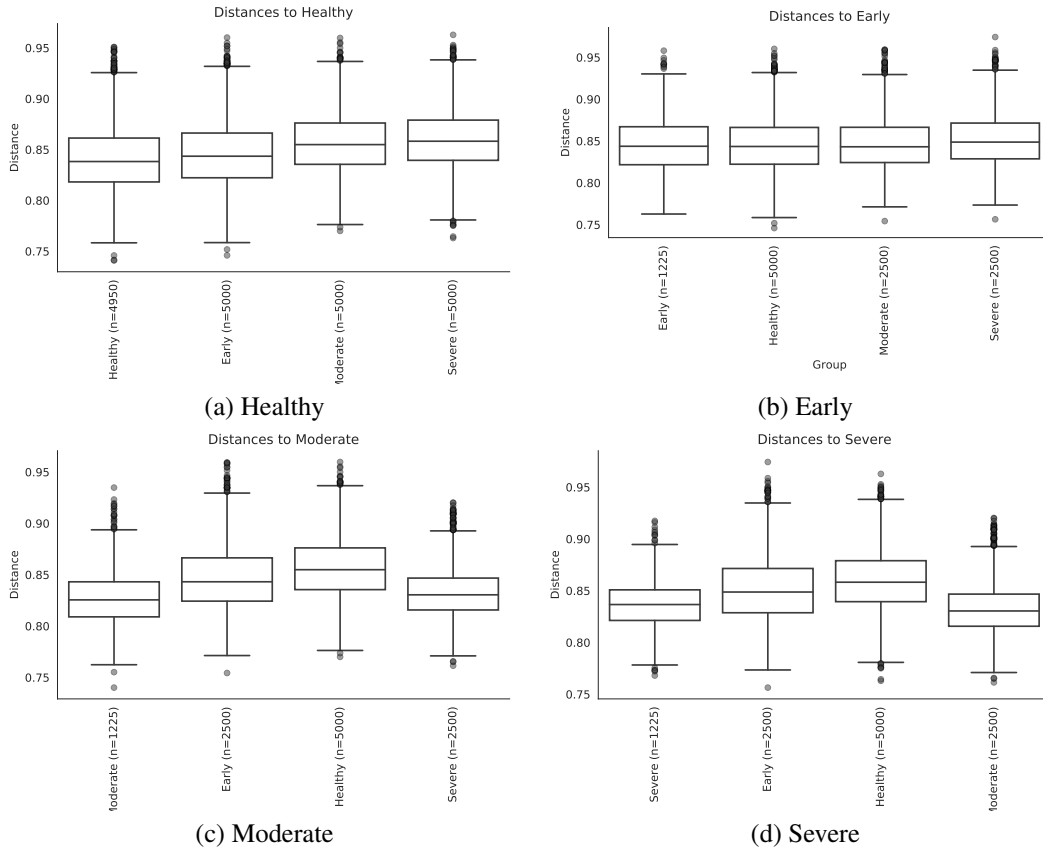


Figure 22: Jaccard Distance Index with Deblur

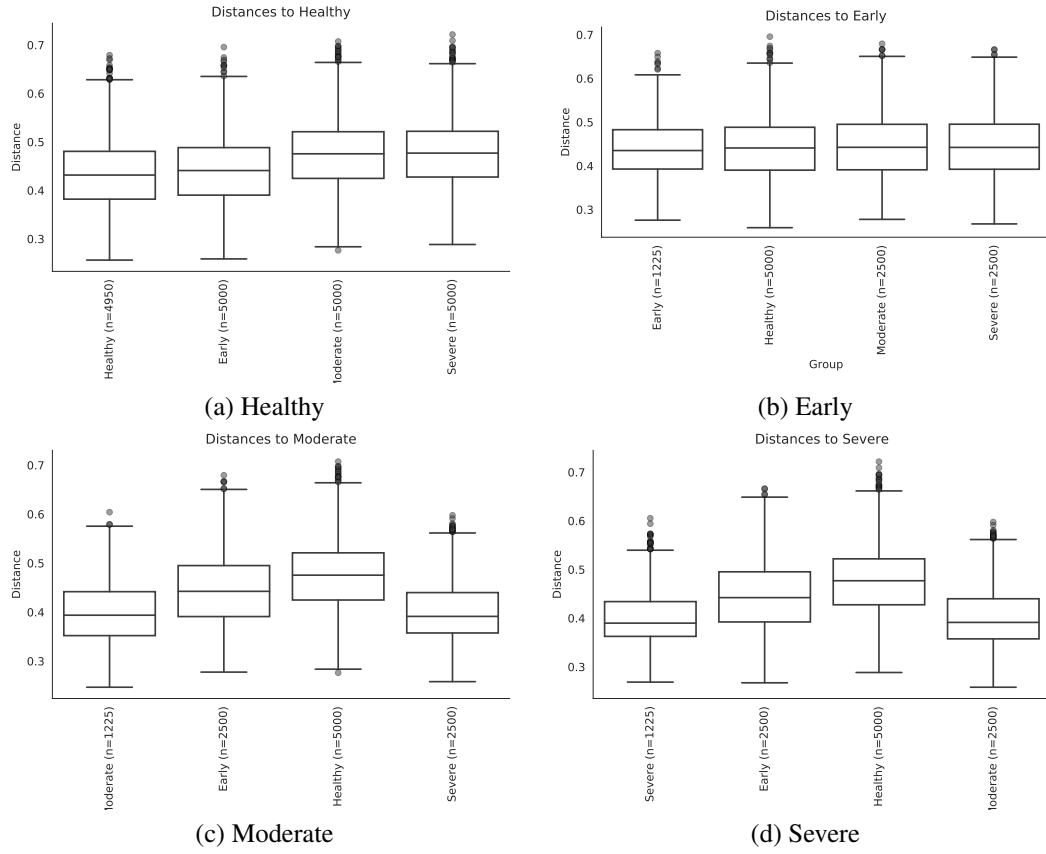


Figure 23: Unweighted Unifrac Distance Index with Deblur

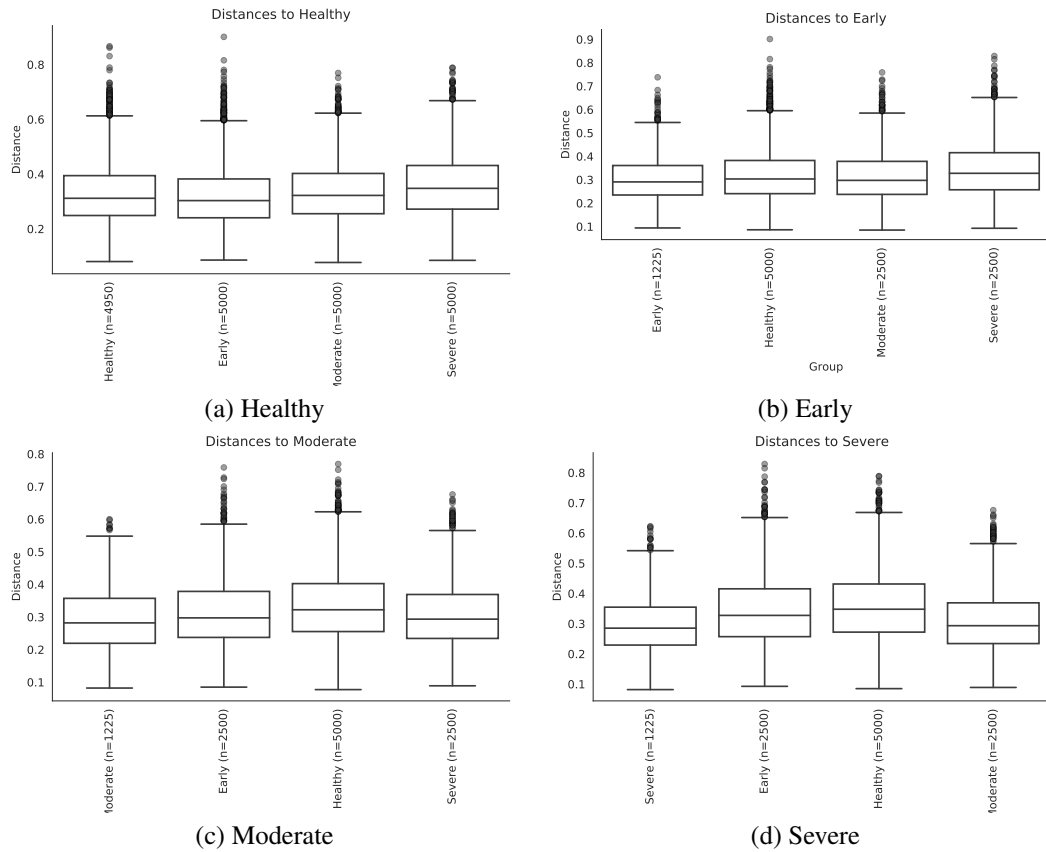


Figure 24: Weighted Unifrac Distance Index with Deblur

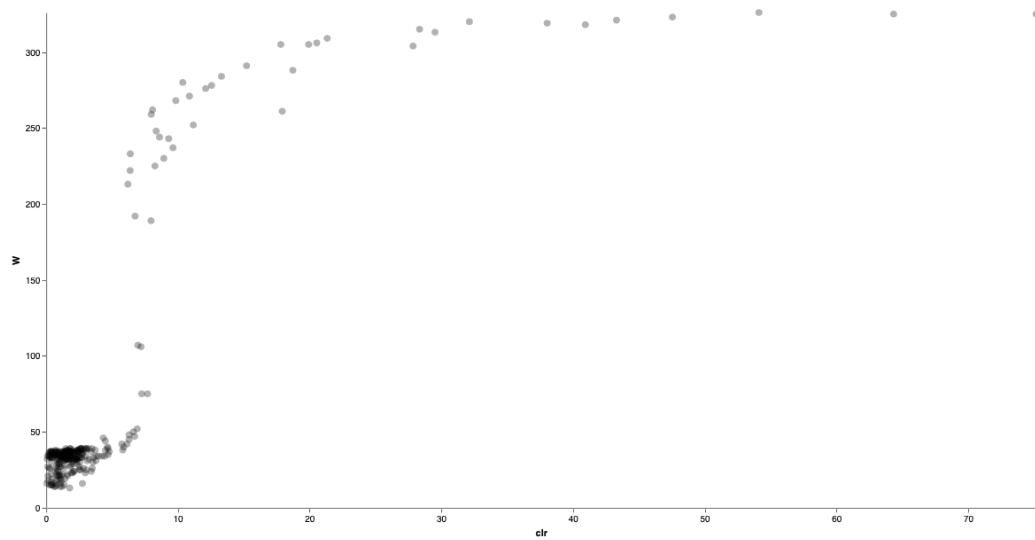


Figure 25: ANCOM Volcano Plot with DADA2 and Greengenes

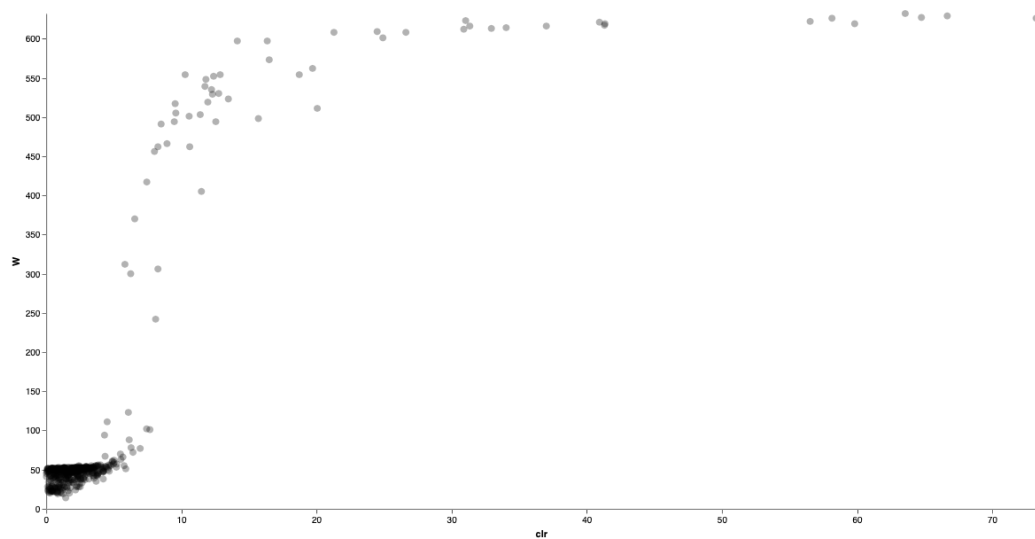


Figure 26: ANCOM Volcano Plot with DADA2 and SILVA

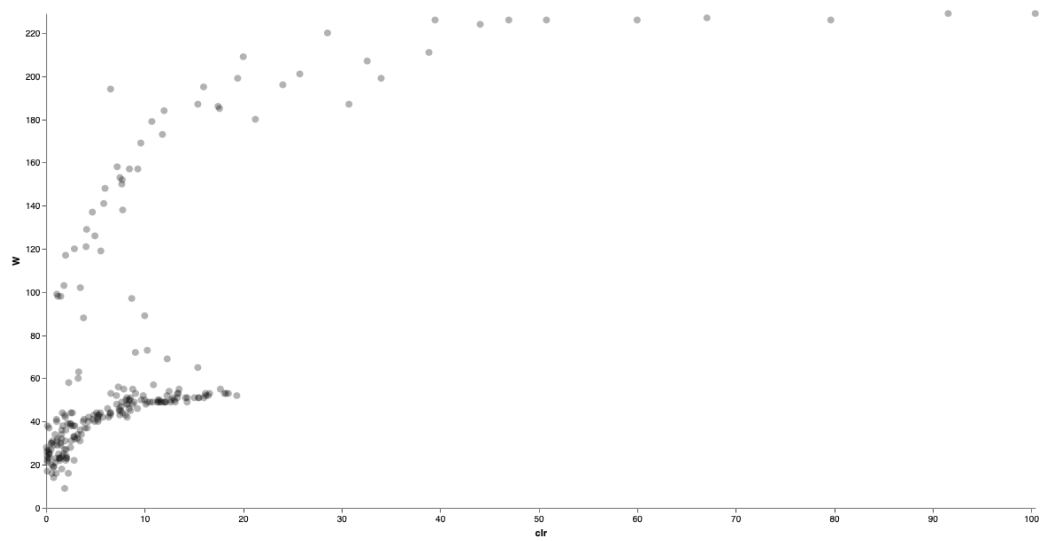


Figure 27: ANCOM Volcano Plot with Deblur and Greengenes

- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... others (2019). Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature biotechnology*, 37(8), 852–857.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). Dada2: high-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7), 581–583.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... Andersen, G. L. (2006). Greengenes, a chimera-checked 16s rna gene database and workbench compatible with arb. *Applied and environmental microbiology*, 72(7), 5069–5072.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3), 90–95.
- James, F. C., & Rathbun, S. (1981). Rarefaction, relative abundance, and diversity of avian communities. *The Auk*, 98(4), 785–800.
- Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease*, 26(1), 27663.
- McKinney, W., et al. (2011). pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12, 2825–2830.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). Silva: a comprehensive online resource for quality checked and aligned ribosomal rna sequence data compatible with arb. *Nucleic acids research*, 35(21), 7188–7196.
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, 449(7164), 804–810.
- Waskom, M., & the seaborn development team. (2020, September). *mwaskom/seaborn*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.592845> doi: 10.5281/zenodo.592845
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., ... others (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1), 27.

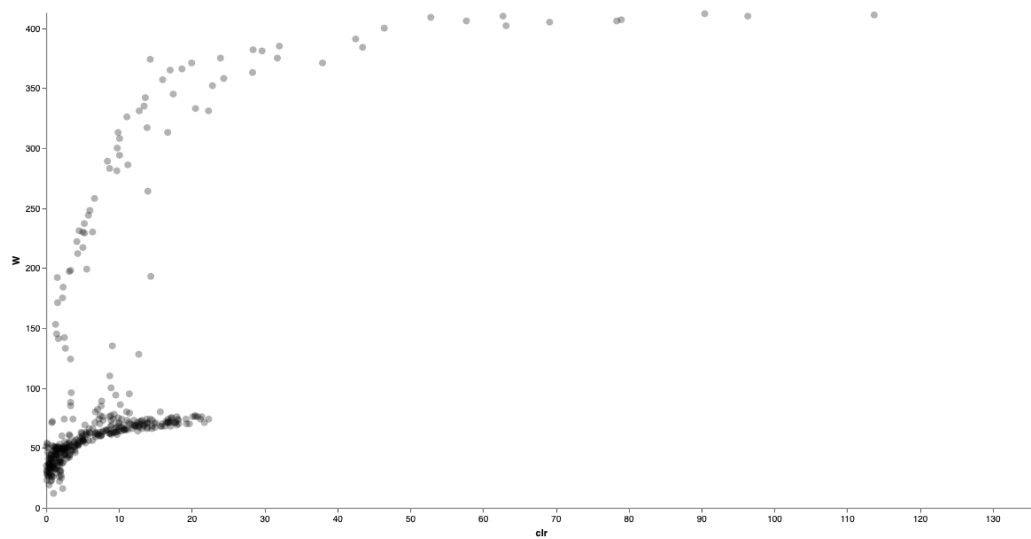


Figure 28: ANCOM Volcano Plot with Deblur and SILVA