Samuel L. Peoples
BBIO 393: Computational Biology
Project 1

Exploratory Analysis of Beta Diversity of Human Gut Microbiome Between Lean and Obese Twins

1. Introduction

This project was completed based on Jeff Gordon's *A Core Gut Microbiome in Obese and Lean Twins* which was used to visualize any noticeable dissimilarity when comparing the Weighted-Unifrac Beta Diversity between Lean-vs-Lean, Obese-vs-Obese, and Lean-vs-Obese Twins. This is completed using the *beta-diversity.py* script from within a *Qiime2* virtual machine. Visualizations were initially created using a Bray-Curtis-Faith metric, but a more accurate result is achieved by introducing the *97_otus.tree* and using a Weighted Unifrac metric.

2. Set-up and use

Parameters:
     filepath = path to *study_77_011618-113533* directory

Dependencies (Qiime2 scripts):
     beta_diversity.py
     single_rarefaction.py
     principal_coordinates.py
     make_emperor.py

Sample console output can be found in /docs/*subsequent_output.txt* and /docs/*firstRun_output.txt*
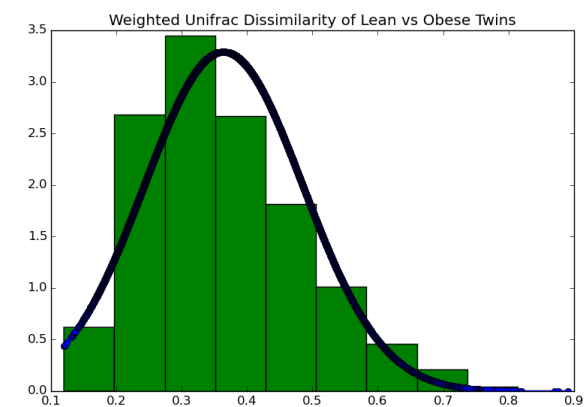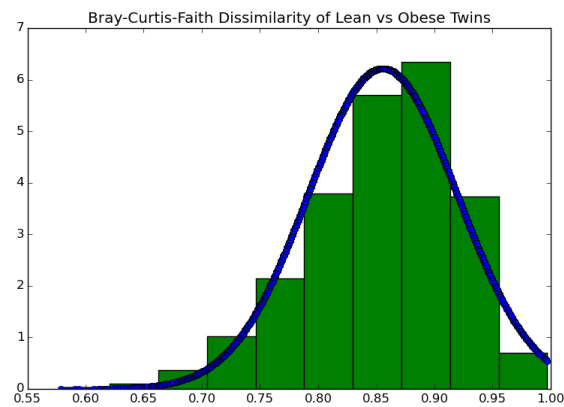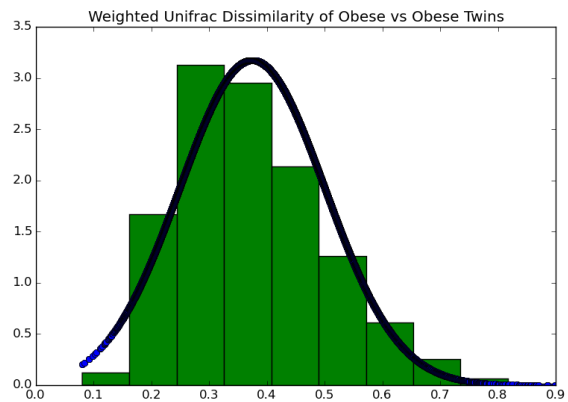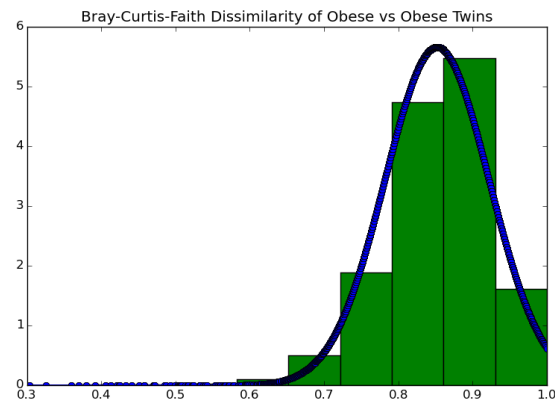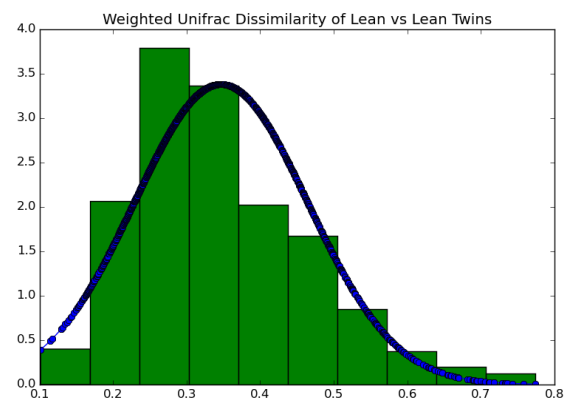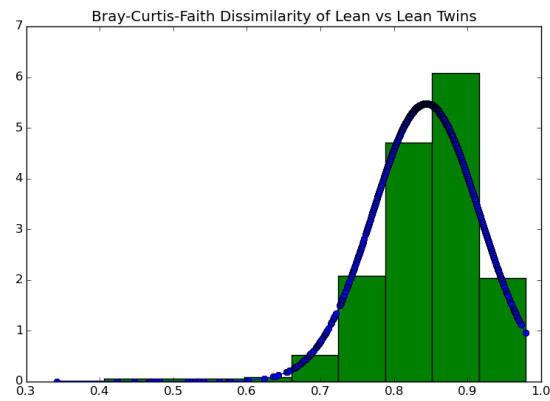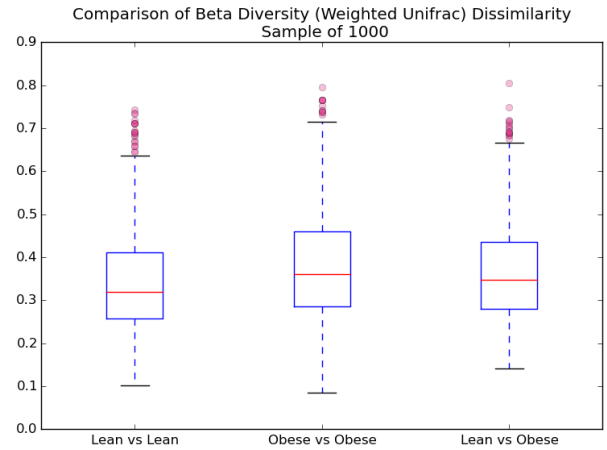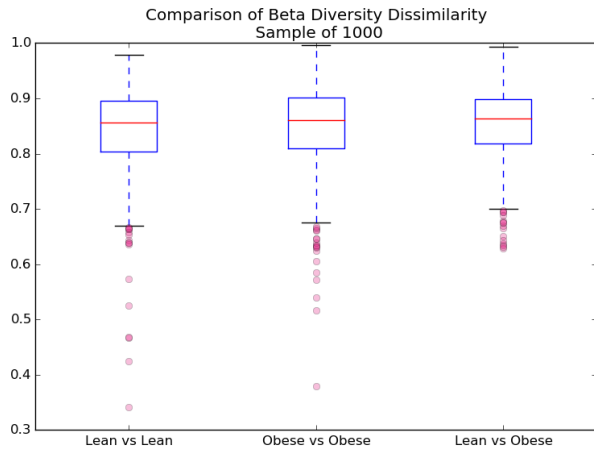
3. Conclusion

Viewing the figures side-by-side lets us see just how significant the changes are when using the *weighted-unifrac* metric. In project 2, I plan to increase the scope of the script, where I will make more of the options parameterized, because just the *filepath* to the study folder is parameterized at this time. I also plan to explore the significance of the Unifrac distances, and implement a monte-carlo simulation to better sample the data.

The differences that can be seen between the distributions shows the significant changes that occur from the *Weighted-Unifrac* metric, which is much more accurate than the *Bray-Curtis-Faith* metric. In the *Unifrac* boxplot, there are more differences that can be seen other than outliers, which can provide exciting opportunities for the second project. In particular, comparing the beta diversity of Lean-vs-Lean patients shows a mean of .348, while the mean of Obese-vs-Obese and Lean-vs-Obese are .375 and .365 respectively. The middle 50% of the data is aggregated lower for Lean-vs-Lean than the other two categories, and the maximum (excluding outliers) is highest when comparing Obese-vs-Obese patients.
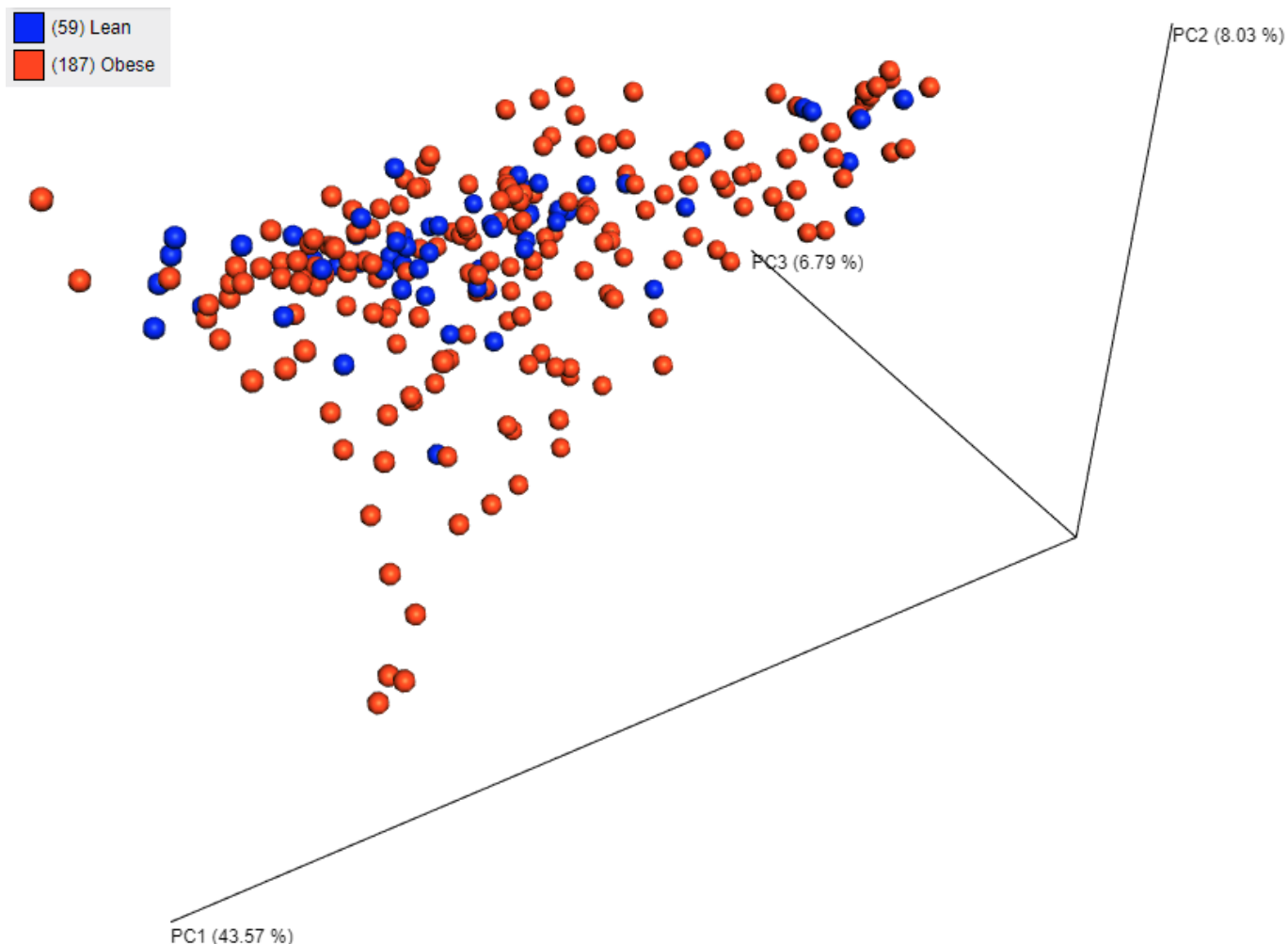
In the PCoA plots, samples labeled obese seem to deviate from the band of blue lean samples. This suggests that the obese samples have more variance in their diversity because the script is developed from the output from the *weighted-unifrac* beta diversity measure. These findings are incredibly useful, and leave a lot to be explored in the subsequent projects.

These differences suggest that there does exist a measurable difference in the beta diversity of the human gut microbiome based on obesity, and the statistical significance of these differences as well as a more detailed analysis can be completed as a part of the second project. Visualizing these differences can be a great way to begin exploratory analysis on the data, and can direct our questioning while simply expressing any trends that may exist.

4. Figures

Comparison of Beta Diversity Dissimilarity
Sample of 1000

Comparison of Beta Diversity (Weighted Unifrac) Dissimilarity
Sample of 1000

Bray-Curtis-Faith Dissimilarity of Lean vs Lean Twins

Weighted Unifrac Dissimilarity of Lean vs Lean Twins

Bray-Curtis-Faith Dissimilarity of Obese vs Obese Twins

Weighted Unifrac Dissimilarity of Obese vs Obese Twins

Bray-Curtis-Faith Dissimilarity of Lean vs Obese Twins

Weighted Unifrac Dissimilarity of Lean vs Obese Twins

PCoA Samples:

5. References

Qiime: http://qiime.org/
Qiita: https://qiita.ucsd.edu/
Bray-Curtis-Faith: http://readiab.org/book/0.1.3/3/1#4.1.1
Gut Microbiome Dataset: https://qiita.ucsd.edu/study/description/77
Biom-Format: http://biom-format.org/documentation/biom_format.html
Lozupone/ Knight's UniFrac: http://aem.asm.org/content/71/12/8228.full
Qualitative/ Quantitative: https://www.ncbi.nlm.nih.gov/pubmed/17220268
Weighted UniFrac: https://liorpachter.wordpress.com/2013/09/18/unifrac-revealed/