**Pandas and Scipy Lab Exercise**

Goal: Make a **python script** or (ideally) a **jupyter notebook** that uses **pandas** and **scipy stats** to manipulate a large data frame and generate statistical analysis.

Scipy Stats Link: <https://docs.scipy.org/doc/scipy/reference/stats.html>

Filename: **YOURLASTNAME**pandas.ipynb or **YOURLASTNAME**pandas.py

Exercise: Use the example python script Get\_Map.txt to do the following.

1. Read in ag\_map\_with\_alpha.txt
2. Test the relationship of three categorical variables – Antibiotic use, Autoimmune and Alcohol consumption each with two diversity indices (observed otus 1250 and Faith’s PD 1250). Make sure to remove any Unspecified data first. Use the f\_oneway test for ANOVA.
3. Use linear regression to correlate BMI with otus and faith’s PD. (remove missing data first).
4. Export a data file that includes all the columns "#SampleID","barcode","pcr\_primers","sex","age\_corrected", plus Anbiotic, Autoimmune, Alcohol consumption and all the diversity measures. See example in Get\_Map.txt of how to export.

VARIABLES: alcohol\_frequency

antibiotic\_history

autoimmune

bmi

observed\_otus\_1250

faiths\_pd\_1250