Horkan_Final_Project_668

Erin Horkan

2024-05-04

Importing Data

I imported the ag_map_with_alpha dataset. I went through the dataset and selected some variables of interest for analysis as well as all of the columns containing diversity metrics to create a smaller dataframe that will be easier to work with and named it ag.ready.

After looking through the different variables I decided to test what has a greater influence on gut microbiome diversity, cats, dogs, and/or household plants.

I made a smaller dataframe of just the variables I was interested in along with the diversity metrics named home.conditions.
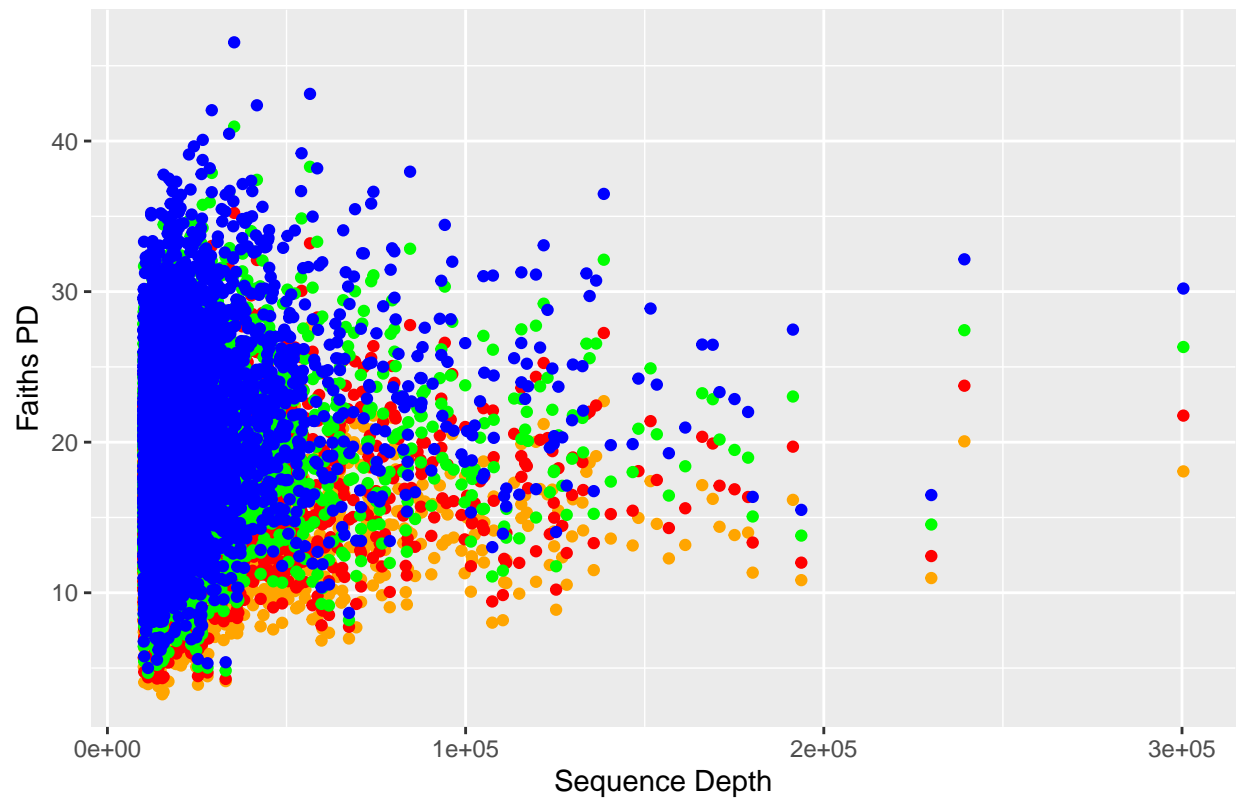
Rarefaction Depth

Next I needed to decide which rarefaction depth to use so I plotted all of the rarefaction depths for the 3 diversity indices, Shannon, Faith's Phylogenetic Diversity, and the Observed OTUs.

Shannon Diversity by Rarefaction Depth
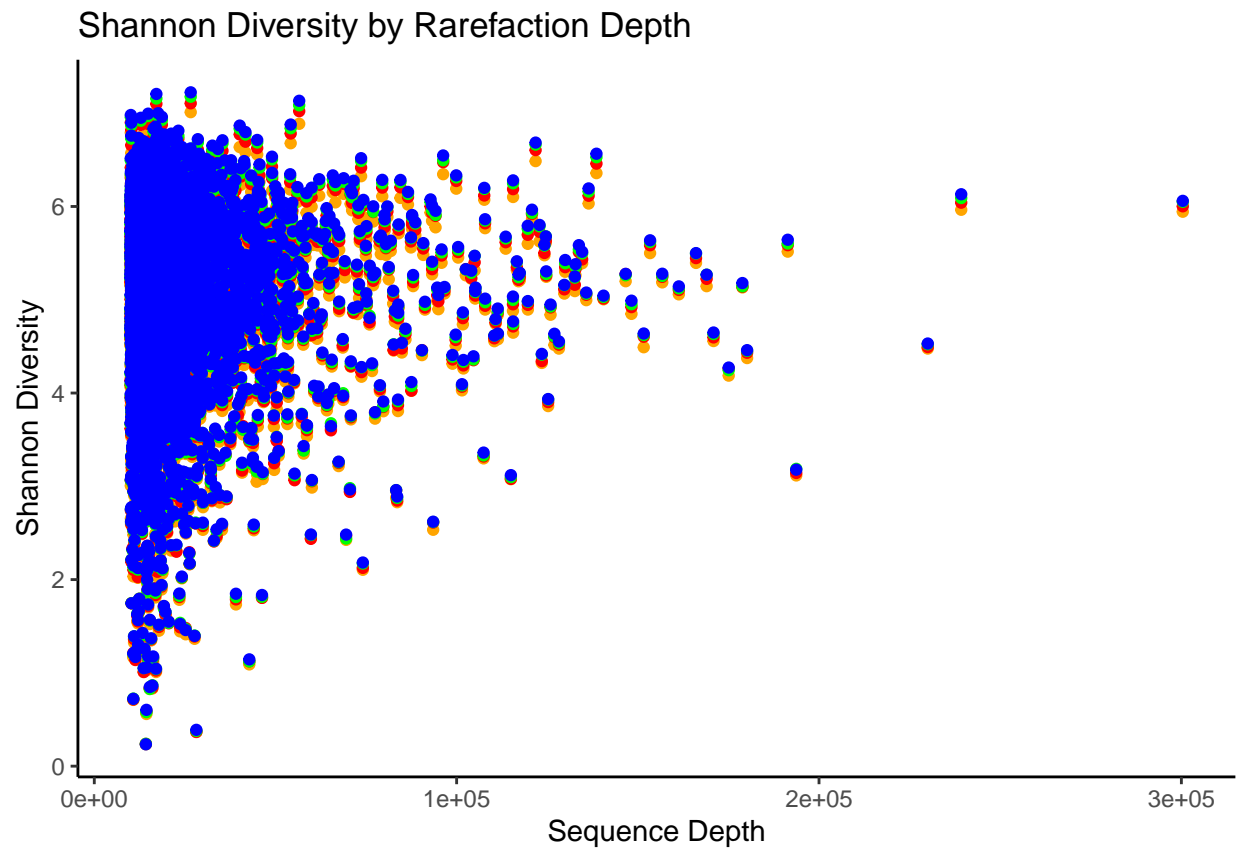
Shannon Diversity plots show little difference between rarefaction depths, I will use 1250 for the rarefaction depth for Shannon Diversity.

Faiths Phylogenetic Diversity by Rarefaction depth

Faith's Phylogenetic Diversity shows that the larger samples have higher phylogenetic diversity for a larger rarefaction depth. I will use 10,000 for the rarefaction depth for PD.

Observed Taxonomic Units by Rarefaction depth
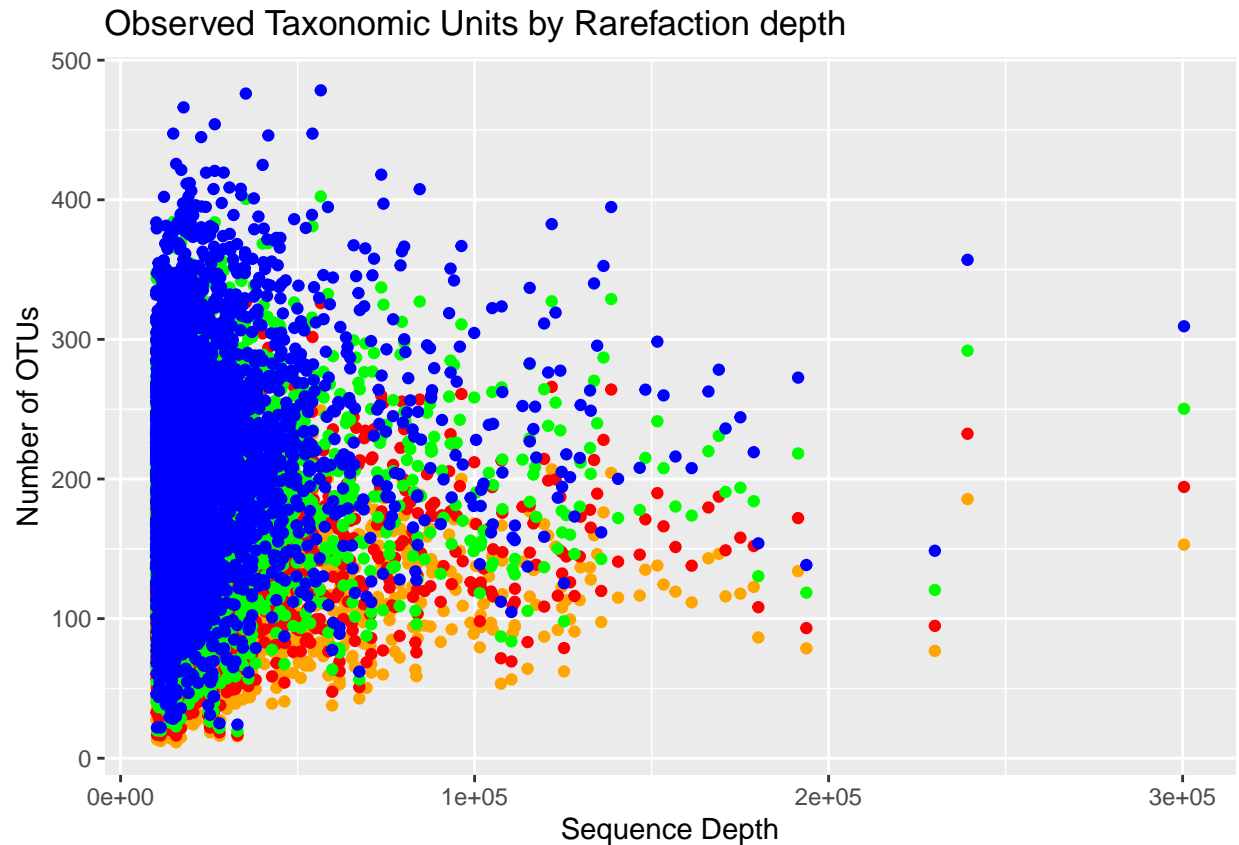
Observed OTUs are higher for larger rarefaction depth. I will use 10,000 for the rarefaction depth for Observed OTUs.

## Further Preparing the dataset

To finally get the data ready for analysis I decided to remove all of the samples that did not include a response for my variables of interest, cat, dog, and types of plants. This cut the number of samples from 5094 to 2920. I also reclassified the plant types as none, low, medium, or high.
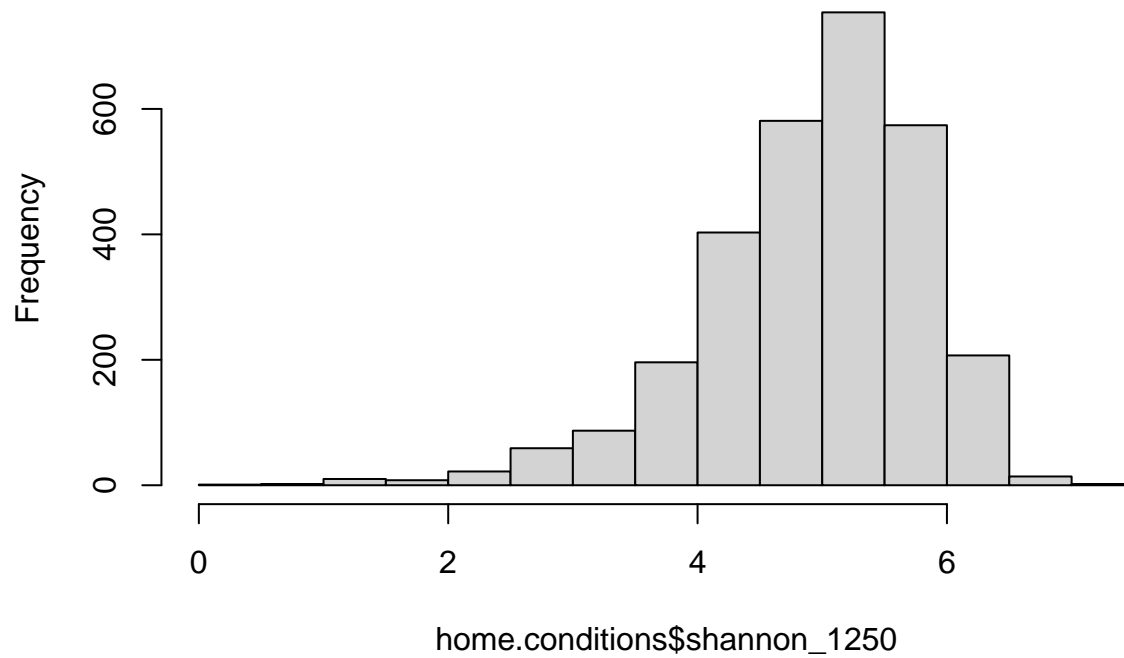
Check For Normal Distribution

Now that the data has been prepared I needed to check for normal distribution
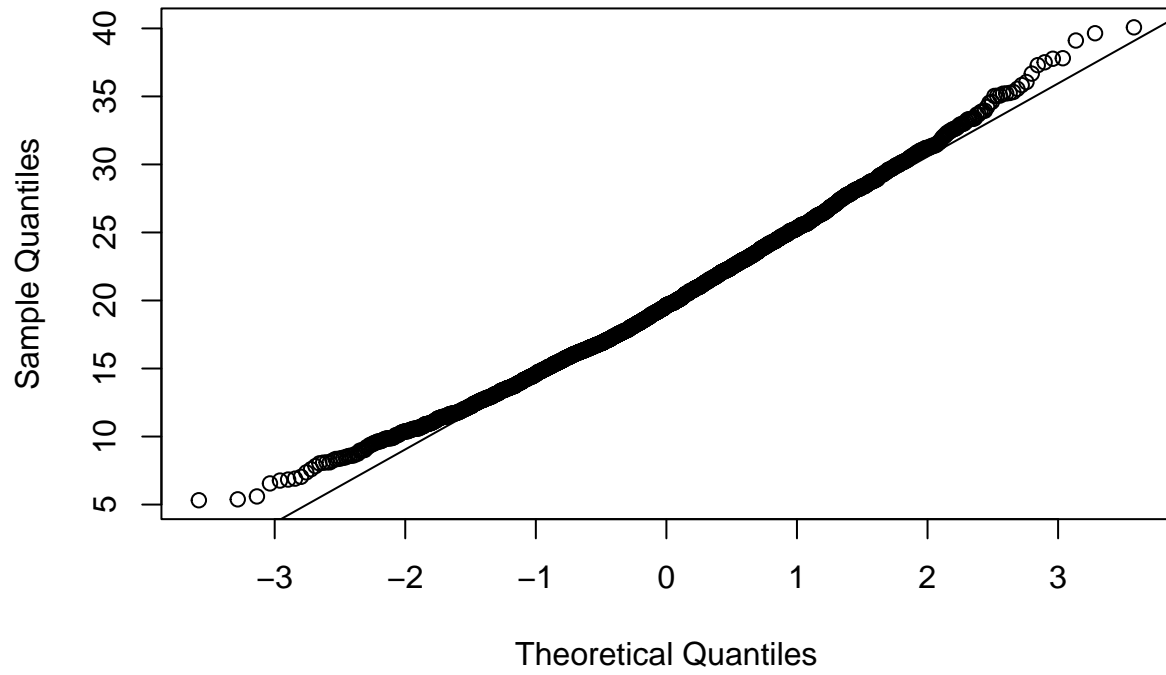
```
## [1] 0.770191
```

Normal Q−Q Plot

Theoretical Quantiles

Sample Quantiles

Histogram of home.conditions$shannon_1250
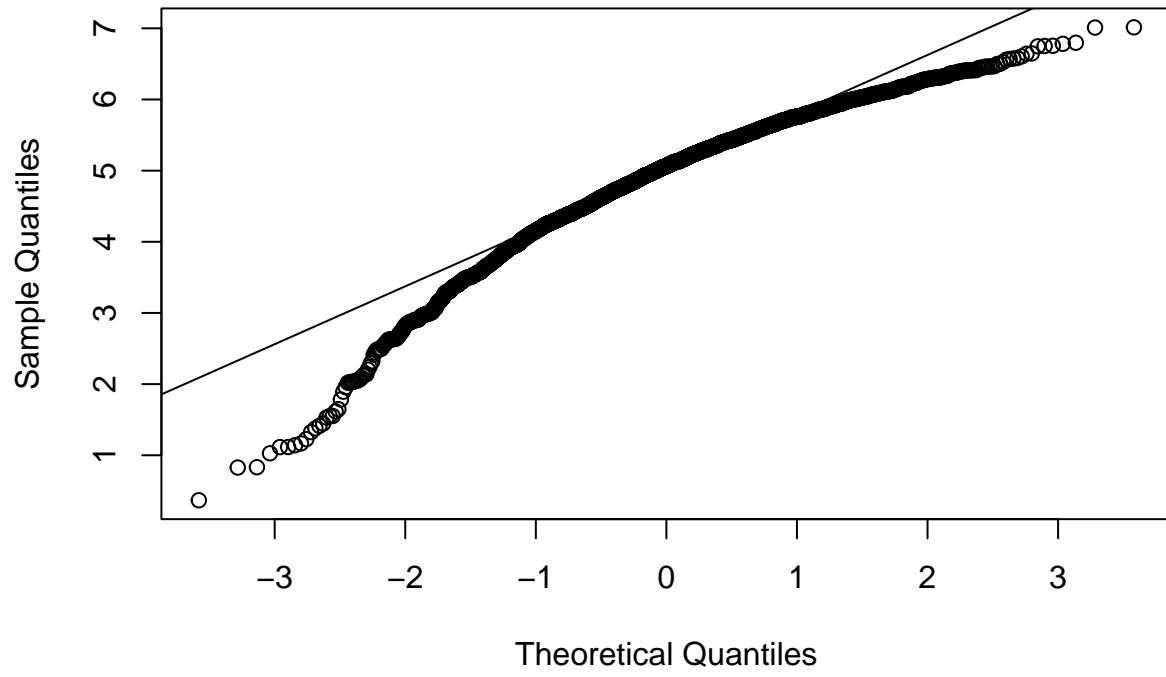
home.conditions$shannon_1250

```
##
##  Shapiro-Wilk normality test
##
## data:  home.conditions$shannon_1250
## W = 0.94946, p-value < 2.2e-16

## [1] 28.5816
```

Normal Q−Q Plot

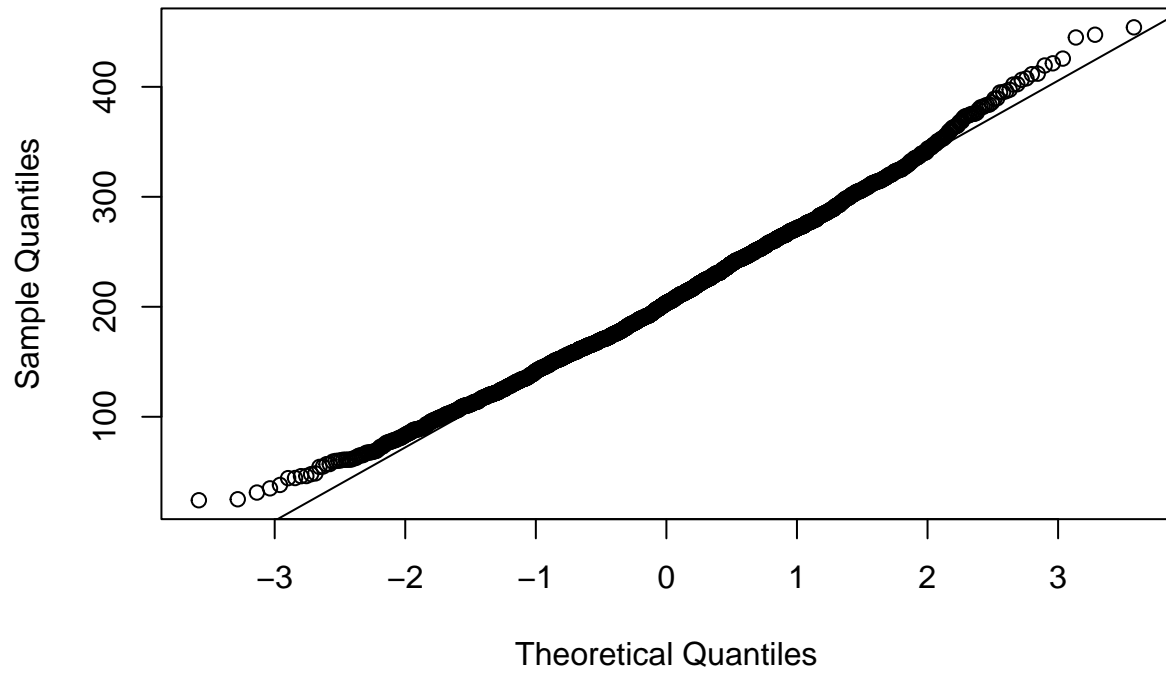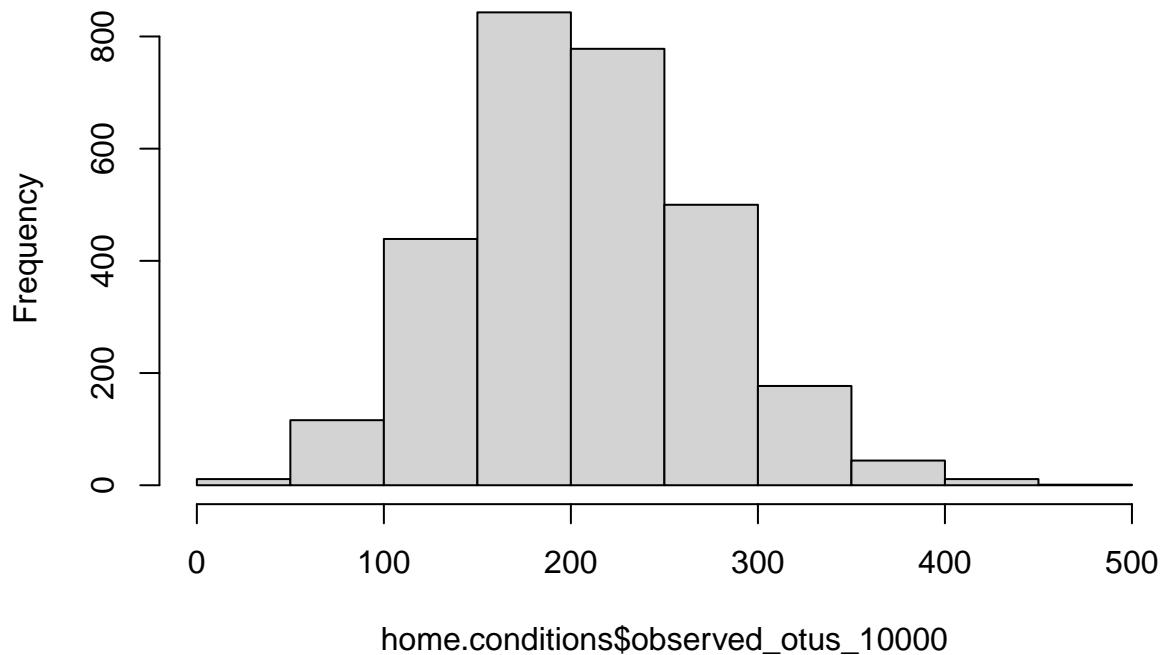# Histogram of home.conditions$faiths_pd_10000
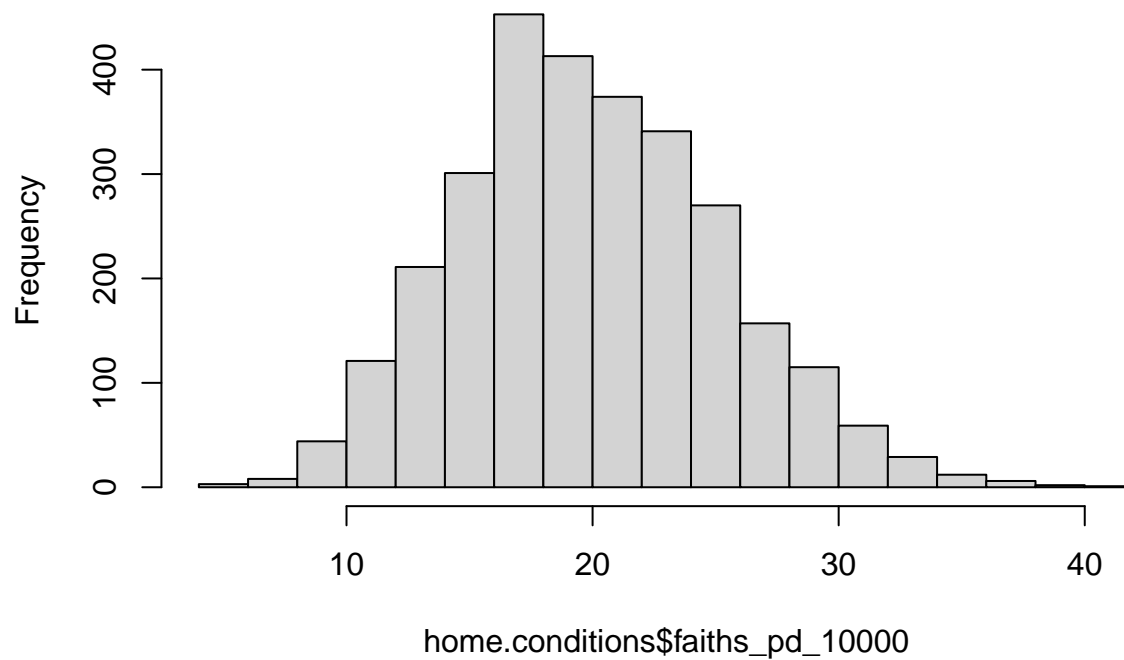

```
##
##  Shapiro-Wilk normality test
##
## data:  home.conditions$faiths_pd_10000
## W = 0.99326, p-value = 2.131e-10
```

```
## [1] 4283.666
```

Normal Q−Q Plot

# Histogram of home.conditions$observed_otus_10000


```
## 
##  Shapiro-Wilk normality test
## 
## data:  home.conditions$observed_otus_10000
## W = 0.99554, p-value = 1.108e-07
```

Based on the QQplots and the histogram plots of the distribution, the observed OTUs are the most normally distributed while the shannon diversity is skewed right, and the phylogenetic diversity is skewed left. The Shapiro-Wilkes test results W values for all of the diversity metrics are all near 1, Observed OTUs W value of 0.99554 was closest to 1, while the Shannon Diversity W value of 0.94946 was furthest from 1. All of the p-values for the Shapiro-Wilkes test were well below 0.05 indicating the variation in the dataset does not vary significantly from normal so I will not transform the data before further analysis.

Visualize the Data

Shannon Diversity for Cats

PB for Cats
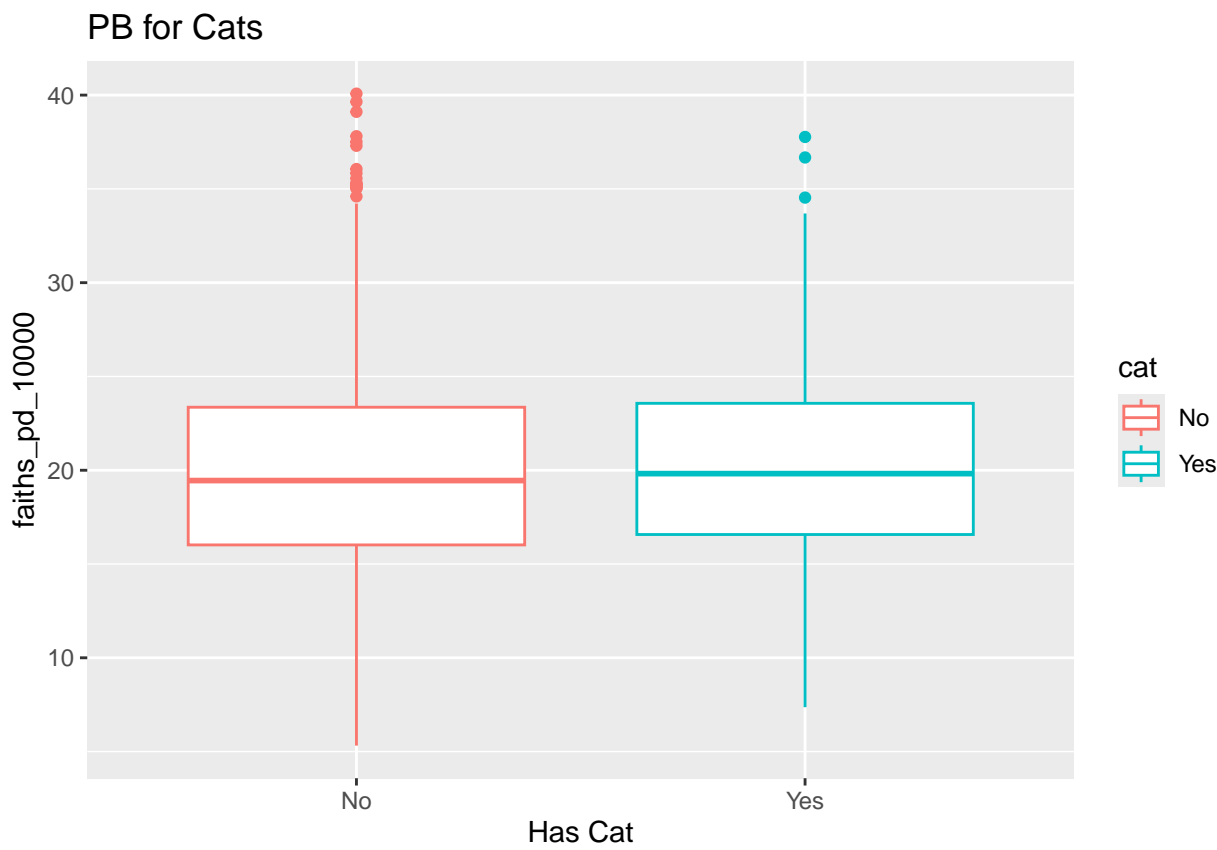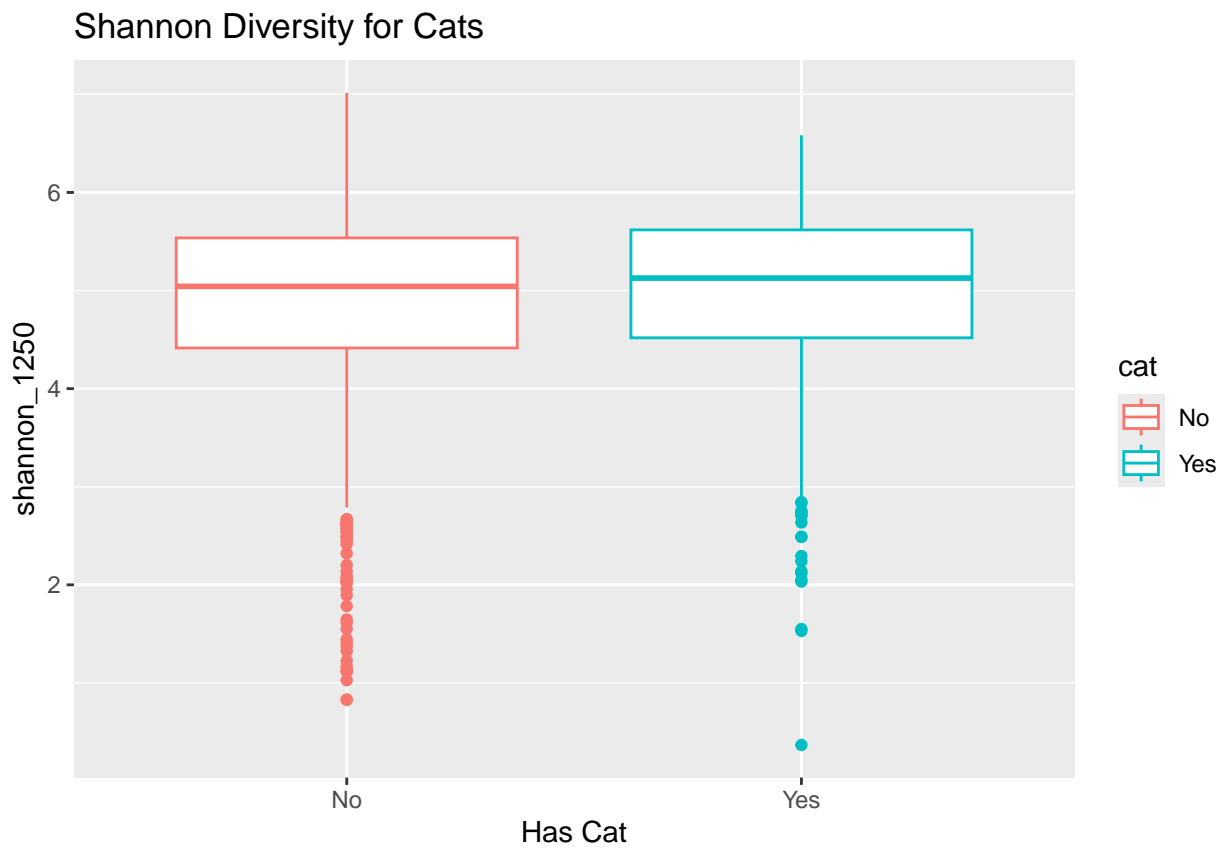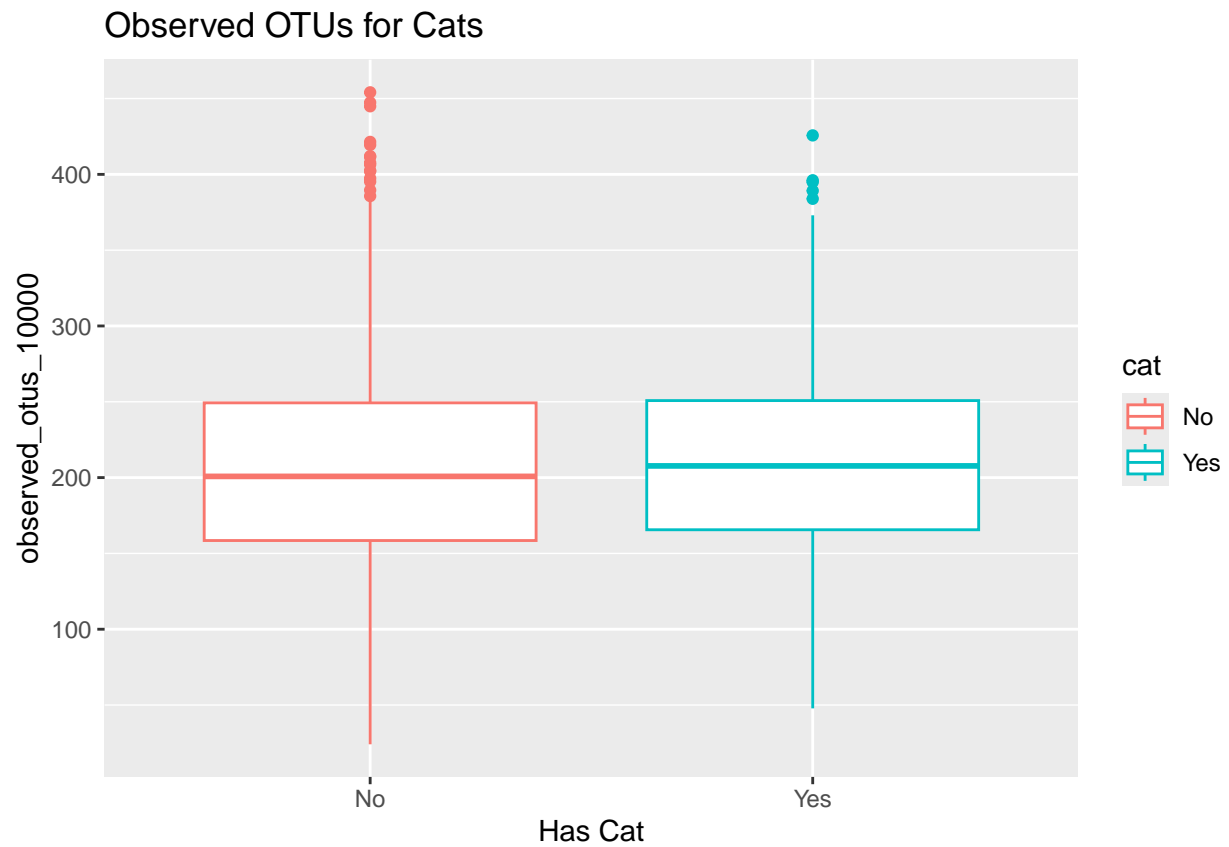
Diversity Metrics Appear to show no difference between samples from people who have cats versus those who do not.

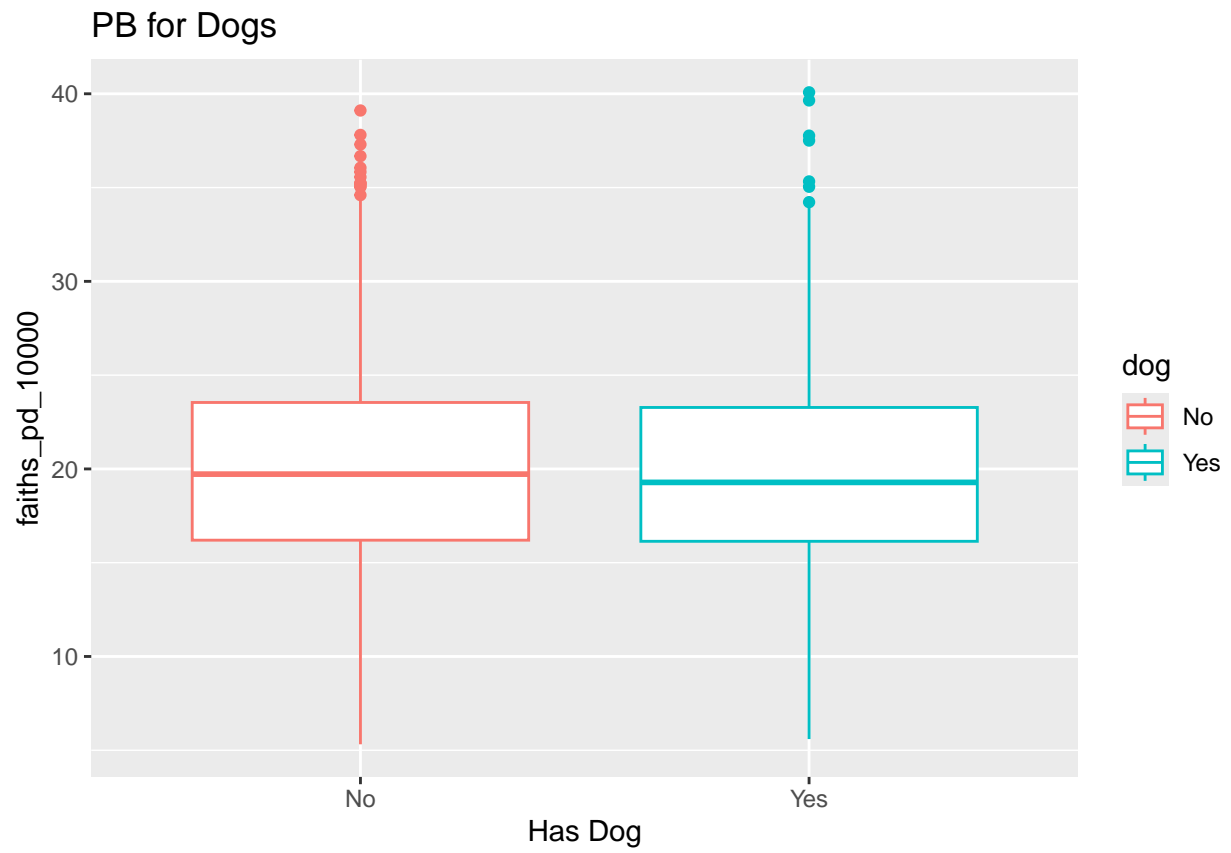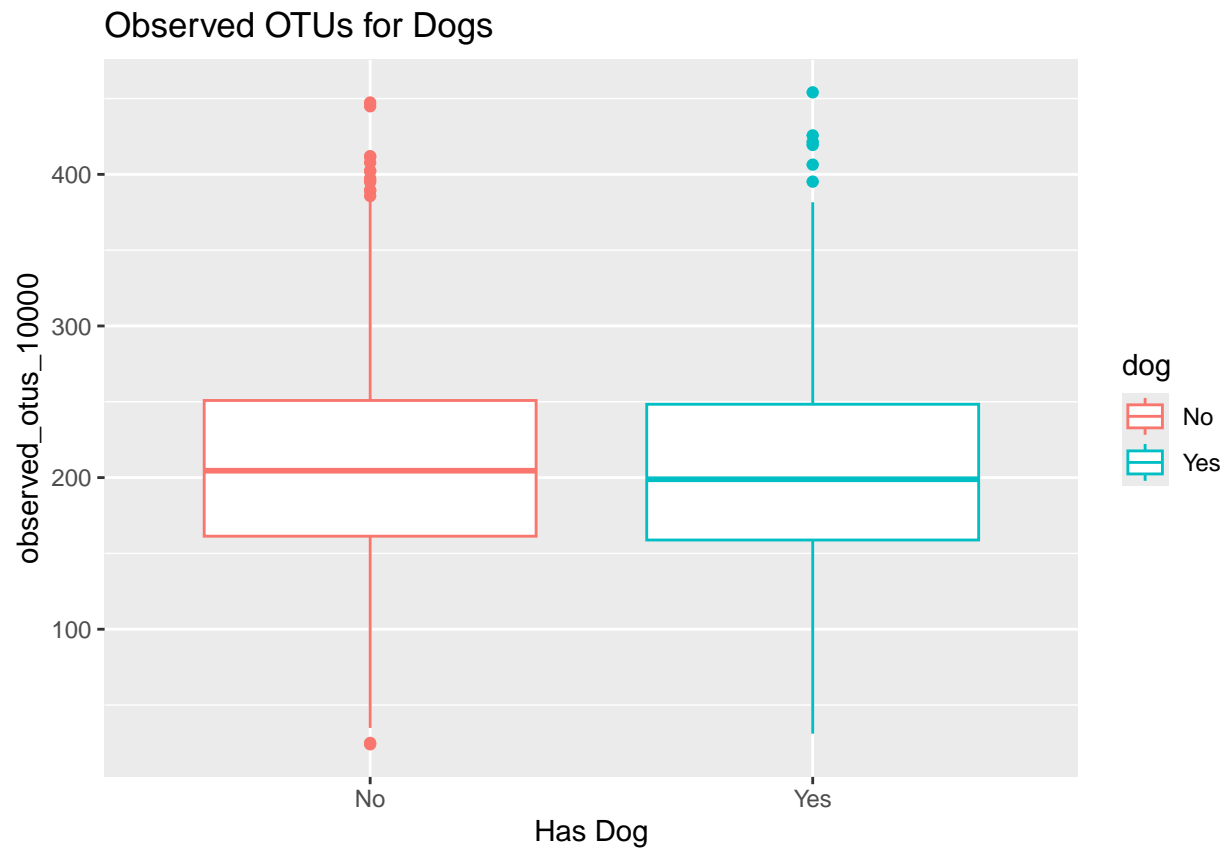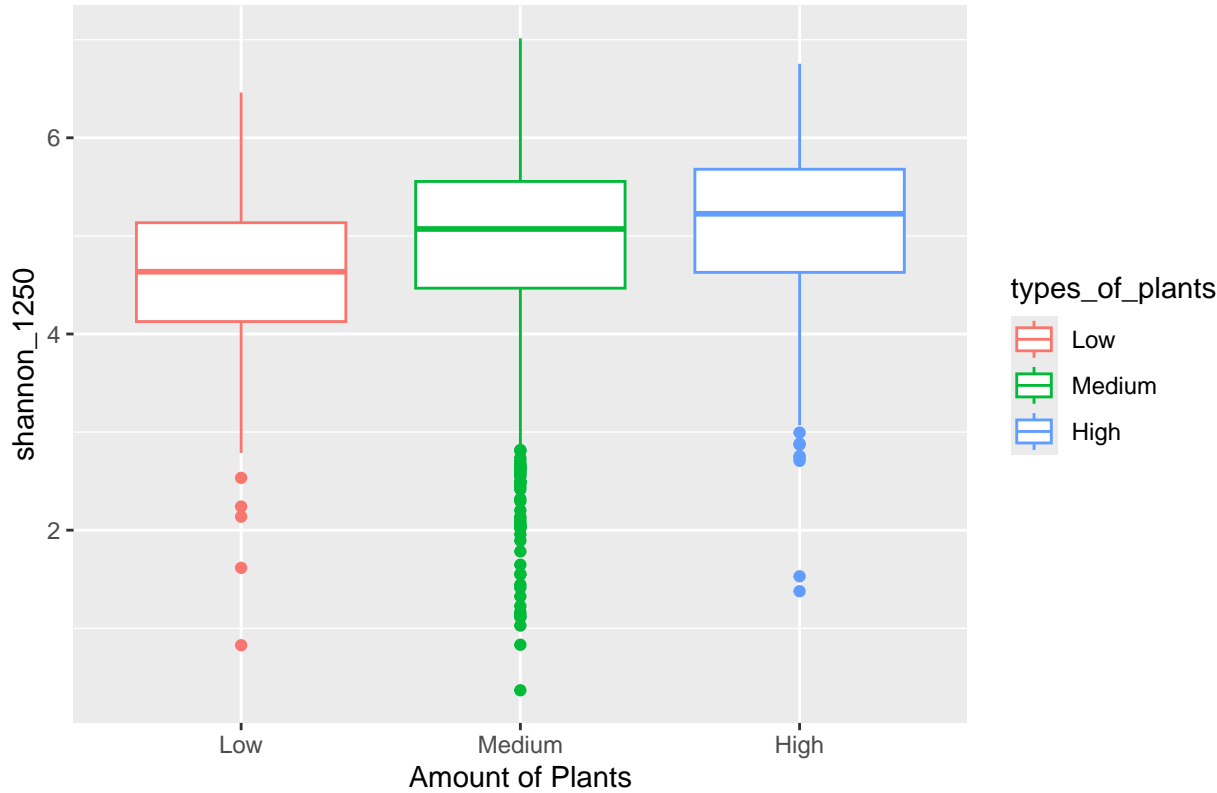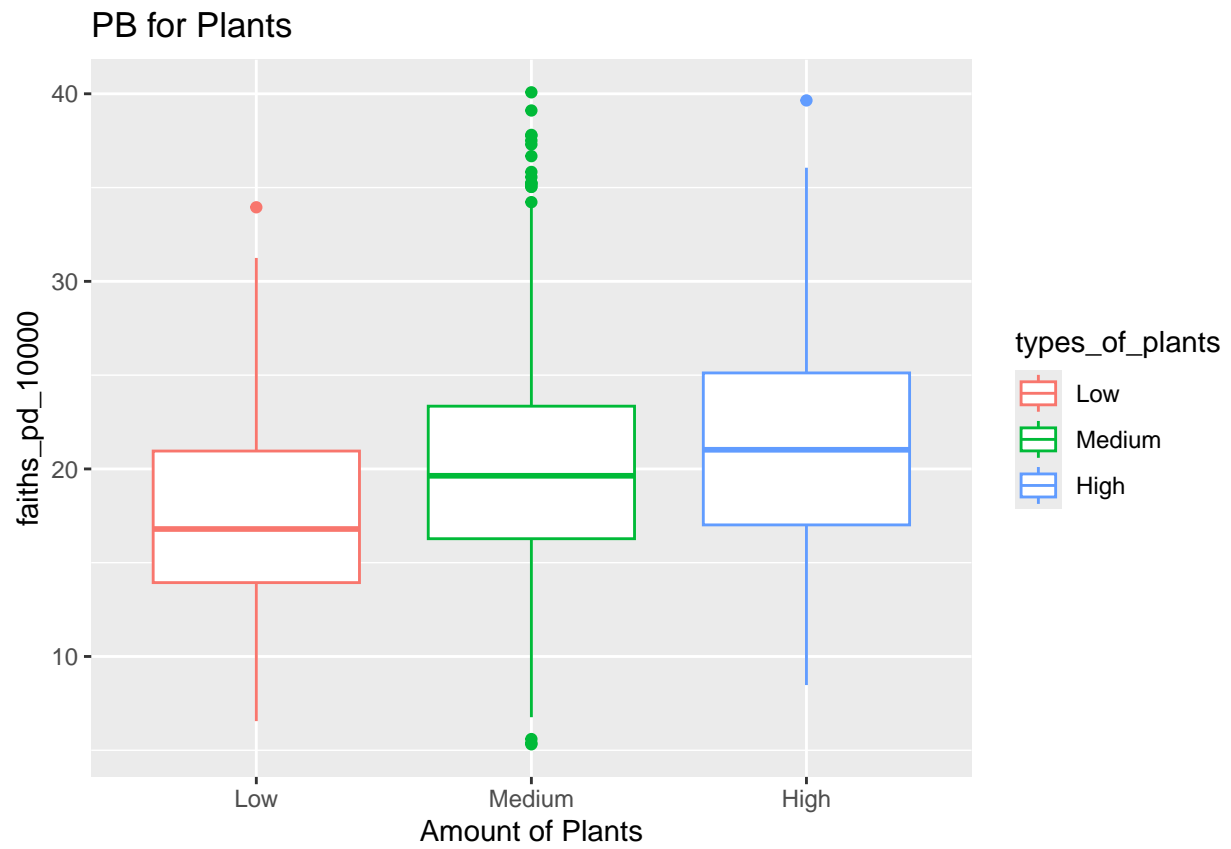Shannon Diversity for Dogs

PB for Dogs
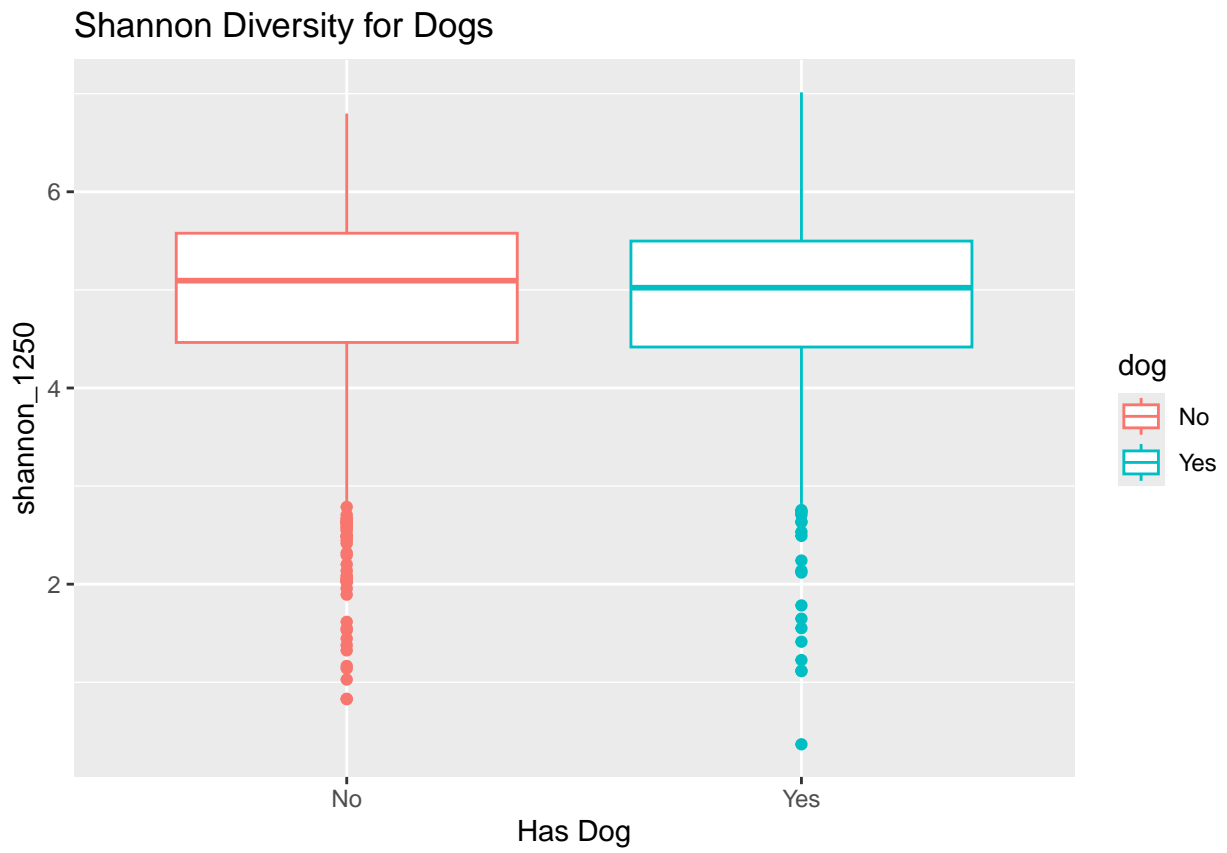
Diversity Metrics Appear to show little difference between samples from people that have dogs versus those

Shannon Diversity for Plants

who do not.

PB for Plants

Obsered OTUs for Plants
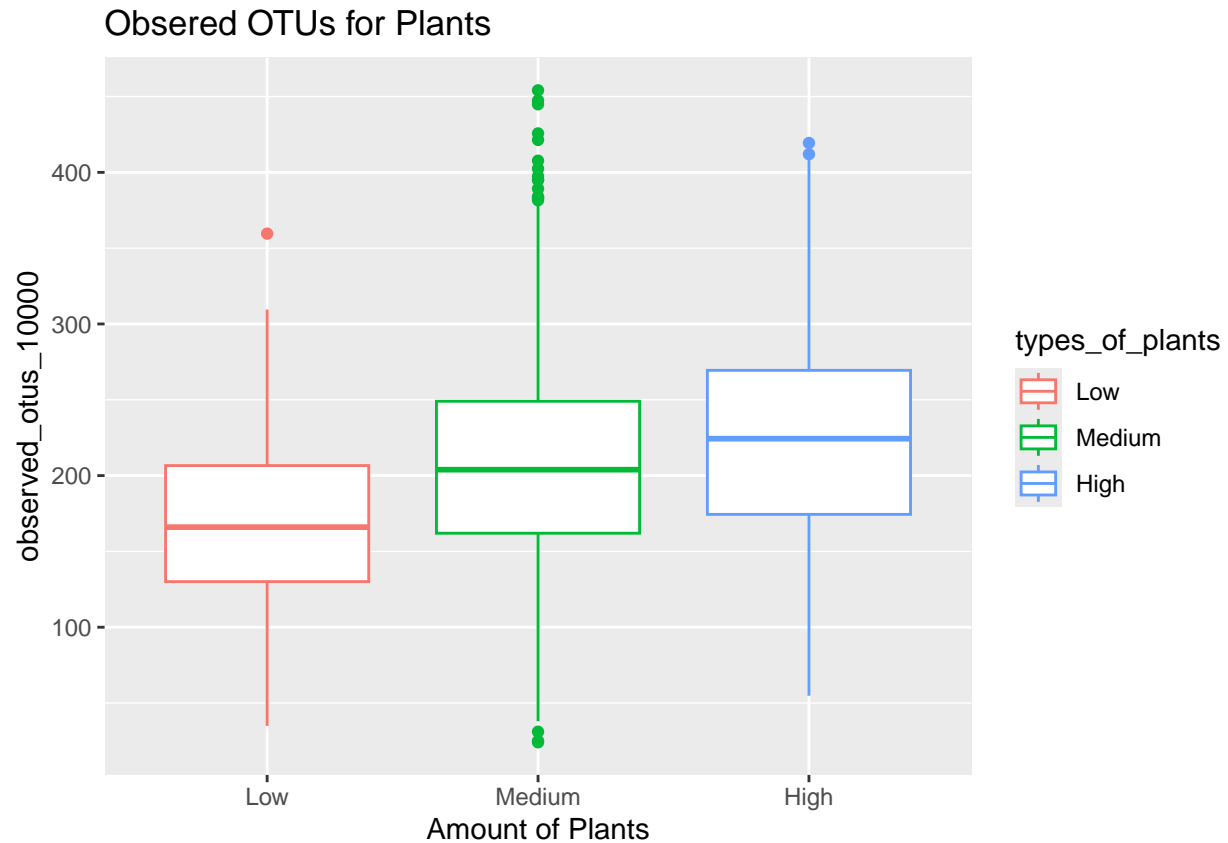
Diversity metrics for types of plants show a trend of higher diversity metrics for a higher number of types of plants.

Significance Testing

Now that we have prepared the data and visualized the data it is time to see if any variables have an effect on the gut diversity that is significant.

Anova

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## dog              1    1.1   1.068   1.388 0.2388
## cat              1    3.8   3.770   4.903 0.0269 *
## Residuals     2917 2243.3   0.769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## cat                1    3.8   3.769   4.975   0.0258 *
## types_of_plants    2   35.0  17.508  23.107 1.11e-10 ***
## Residuals       2916 2209.4   0.758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## dog               1    1.1   1.068    1.408     0.236
## types_of_plants   2   35.8  17.888   23.589 6.88e-11 ***
## Residuals       2916 2211.3   0.758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##               Df    Sum Sq Mean Sq F value Pr(>F)
## dog            1      2730    2730   0.638  0.425
## cat            1     11389   11389   2.660  0.103
## Residuals   2917 12489903    4282


##               Df    Sum Sq Mean Sq F value Pr(>F)
## cat            1     11386   11386   2.742 0.0979 .
## types_of_plants 2    382993  191496  46.112 <2e-16 ***
## Residuals   2916 12109643    4153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##               Df   Sum Sq Mean Sq F value Pr(>F)
## dog            1     2730    2730   0.657  0.418
## types_of_plants 2   387147  193573  46.595 <2e-16 ***
## Residuals   2916 12114145    4154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##               Df    Sum Sq Mean Sq F value Pr(>F)
## dog            1      2730    2730   0.638  0.425
## cat            1     11389   11389   2.660  0.103
## Residuals   2917 12489903    4282


##               Df    Sum Sq Mean Sq F value Pr(>F)
## cat            1     11386   11386   2.742 0.0979 .
## types_of_plants 2    382993  191496  46.112 <2e-16 ***
## Residuals   2916 12109643    4153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


##               Df    Sum Sq Mean Sq F value Pr(>F)
## dog            1      2730    2730   0.657  0.418
## types_of_plants 2    387147  193573  46.595 <2e-16 ***
## Residuals   2916 12114145    4154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The 2-way ANOVA tests show that number of types of plants has a significant effect on the gut diversity. This was true for all of the diversity metrics. Whether or not the person sampled had a dog does not have a significant effect, while whether the person has a cat was not significant for all but the shannon diversity by types of plants. Types of plants was significant for all tests for all diversity metrics.

Plants Additional Significance Tests

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = observed_otus_10000 ~ types_of_plants, data = home.conditions)
##
## $types_of_plants
##                  diff        lwr       upr    p adj
## Medium-Low   34.72384  24.064760  45.38292  0.0e+00
## High-Low     52.93011  40.050332  65.80990  0.0e+00
## High-Medium  18.20627   9.729142  26.68340  1.5e-06
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  home.conditions$observed_otus_10000 and home.conditions$types_of_plants
##
##        Low     Medium
## Medium 4.4e-14 -
## High   < 2e-16 5.0e-07
##
## P value adjustment method: BH
```

The Tukey test shows that the difference between groups increases in a stepwise manner with the largest difference between high and low, compared to the difference between low to medium categories, and the medium to high categories.

The pairwise T-test show that there is a significant difference between low, medium, and high when it comes to types of plants influence on the OTUs found in a person's gut microbiome sample.

Conclusions and Summary

The number of types of plants a person has in their home has a more significant influence on the diversity of a persons gut microbiome than the kinds of pets they have in their home. Whether the person had a dog or a cat had no significant influence on the diversity of a persons gut microbiome diversity. I found these results surprising as I had read before that having pets can lead to kids in the same home having more diverse microbiomes. It seems like having a different animal in the home would have a larger influence than plants, though the high category for plants is over 30 plants which depending on the size of a persons home could meant they have an indoor jungle. It is also possible since I did not read the metadata that types of plants is indicative of the number of different kinds of plants a person eats not the types of plants in their house which makes a little more sense.