# PHAE Part 5: Plot Data Exploration

## Overview

In this week’s lab, we will start our analysis of the PHAE data, comparing woody plant species richness and diversity across our different plots. We’ll also examine whether there is a relationship between the degree of human disturbance in a plot and the woody plant diversity in that site.

## Data Notes

I have compiled and edited the submitted data. We’ve collected a lot of data!

Some of the variables you have measured are more useful in making comparisons across broad scales, and less useful for comparing our class plots to each other. For instance, measures like annual precipitation and average temperature are great for comparing sites to each other across the country, but not as informative when most of our plots are in one area.

In this lab, we will focus on our class data for our analysis. That means we won’t be analyzing every variable you measured. However, those variables will still be relevant to our analysis using the NEON data in a future lab, and will also be valuable in future studies that build on the work we started this semester.

Here’s a quick list of the data processing that created the spreadsheet you’ll be working with for this analysis:

* I hid any variables that you don’t need to use for this data analysis lab (e.g., variables that are only useful for national-scale analyses, or variables that turned out to be pretty similar across all plots)
* I edited some entered data to correct errors (e.g., in land cover classes, spelling of species names, etc.)
* I added the Shannon and Simpson diversity measures for each plot (the measures that you calculated in a previous lab).
* I added a broader land cover categorization (any of the NLCD “developed” types were lumped into a “Developed” category, and any of the “forest” types were lumped into a “Forest” category).
* I calculated the average tree diameter, as well as the “coefficient of variation” (CV) of tree diameter. Coefficient of variation is a measure of variability in a dataset – a small CV indicates all the trees were a similar size, and a large one indicates there was a lot of variability – both large and small trees.

## Summary Tables and Graphs

First, let’s summarize some variables from our plots by land cover type. If we generalize the NLCD land cover classes from our first PHAE lab, we can say that all of our plots are either in a type of forest or in a type of developed area. This is represented in the LandCoverGeneral column in the datasheet.

Using what you learned about Pivot Tables from previous labs, make a summary table that shows the following averages for all forest plots and for all developed plots (fill in the values below). *This should be a quick thing to calculate, using the Pivot Tables.*

|  |  |  |
| --- | --- | --- |
| **Variable** | **Forest Plots** | **Developed Plots** |
| **Average** number of trees |  |  |
| **Average** species richness |  |  |
| **Average** tree diameter |  |  |
| **Average** variation in tree diameter |  |  |
| **Standard deviation (SD)**: number of trees |  |  |
| **Standard deviation (SD)**: species richness |  |  |
| **Standard deviation (SD)**: tree diameter |  |  |
| **Standard deviation (SD)**: variation in tree diameter |  |  |
| **Number of plots (n)** |  |  |

### Bar Graphs

We will make bar graphs for each of these variables, to compare the forest and the developed plots. We want to put *error bars* on our bar graphs.

When we calculate the average of values in a sample (like the average number of trees in forested plots), we are really trying to make a more general scientific statement – our average is an *estimate* of the number of trees we expect to see in forested plots like this. We want to be able to represent how *certain* we are that the average we calculated is a good estimate of the number of trees we’d tend to see if we keep looking at more plots.

We can use *error bars* to indicate this in a bar graph. We will use *standard error bars*.

The *standard error* is calculated based on the standard deviation and the sample size:

$$SE= \frac{SD}{\sqrt{n}}$$

Use the numbers above to calculate the standard error for each variable.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Forest Plots** | **Developed Plots** |
| **Standard error (SE):** number of trees |  |  |
| **Standard error (SE):** species richness |  |  |
| **Standard error (SE):** tree diameter |  |  |
| **Standard error (SE):** variation in tree diameter |  |  |

Next, make four bar graphs – one for each of these four variables – comparing the forest plots to the developed plots. *Watch the demonstration video for help on how to make these bar graphs.*

The bar graphs should:

* Have two bars (one for forest, and one for developed)
* Have the x-axis labelled with the types of plots (forest and developed)
* Have a y-axis label indicating the variable
* NOT have a title (in scientific reports, we use captions instead of figure titles, so you can delete the title that Excel creates for you)
* NOT have gridlines (for a simple bar graph with two bars, they are not needed)
* Have error bars that indicate the standard errors you created above (Excel’s automatic error bars are incorrect – see the demo video for how to add your own error bars).

**Paste your bar graphs into this document here.**

### Scatterplots

We have two core measure of human impact that we measured in our plots – the “viewscape” impact (Human View Impact, which was VH0 through VH350 in your datasheets), and % Impervious (which you estimated in quadrants in your datasheets).

Please make *scatterplots* *with trendlines* for the following variables.

|  |  |  |
| --- | --- | --- |
| **Figure** | **x-axis** | **y-axis** |
| 1 | Human View Impact\* | Shannon Diversity |
| 2 | Human View Impact | Simpson Diversity |
| 3 | % Impervious | Shannon Diversity |
| 4 | % Impervious | Simpson Diversity |

The scatterplots should:

* Have the x and y axis labeled, correctly showing the variables listed above.
* Have a trendline showing the relationship between the two variables.
* NOT have a title (in scientific reports, we use captions instead of figure titles, so you can delete the title that Excel creates for you)
* NOT have gridlines (for a simple bar graph with two bars, they are not needed)

**Paste your scatterplots into this document here.**