Q1. How many species were identified by the study, considering both NWAs? 70 species.

Q2. Plot a Species Accumulation Curve with the three vegetation types in each of the NWAs.

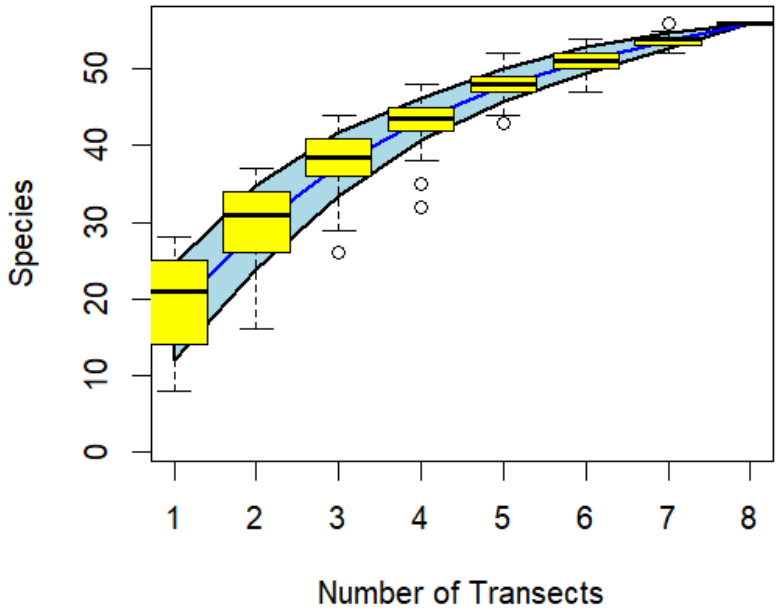

Figure 1. Species accumulation curve in Long Point National Wildlife Area (LPNWA) among 8 50 m long transects where 808 points at 50 cm intervals with no *Phragmites australis* invasion (Reference) were sampled, which observed a total of 37 species (S = 37).

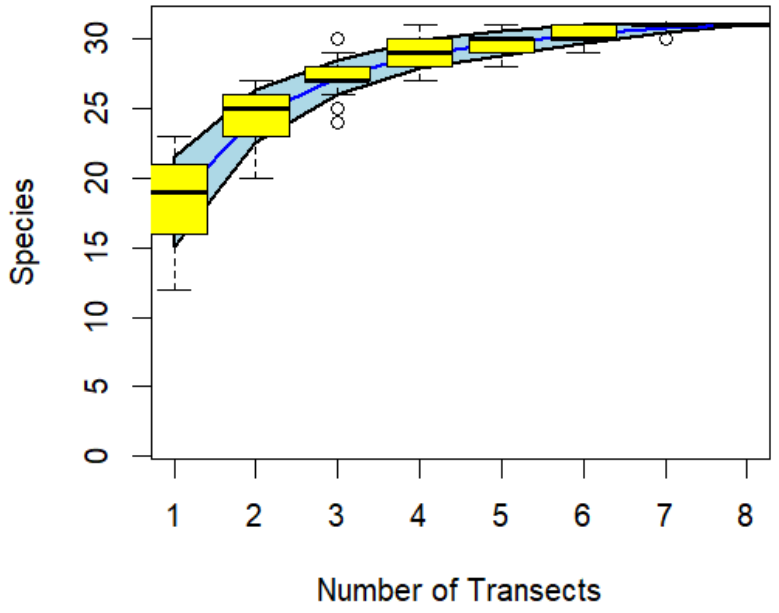

Figure 2. Species accumulation curve in Long Point National Wildlife Area (LPNWA) among 8 50 m long transects where 808 points at 50 cm intervals formerly invaded by *Phragmites australis* but then treated with herbicide (Treated) were sampled, which observed a total of 19 species (S = 19).

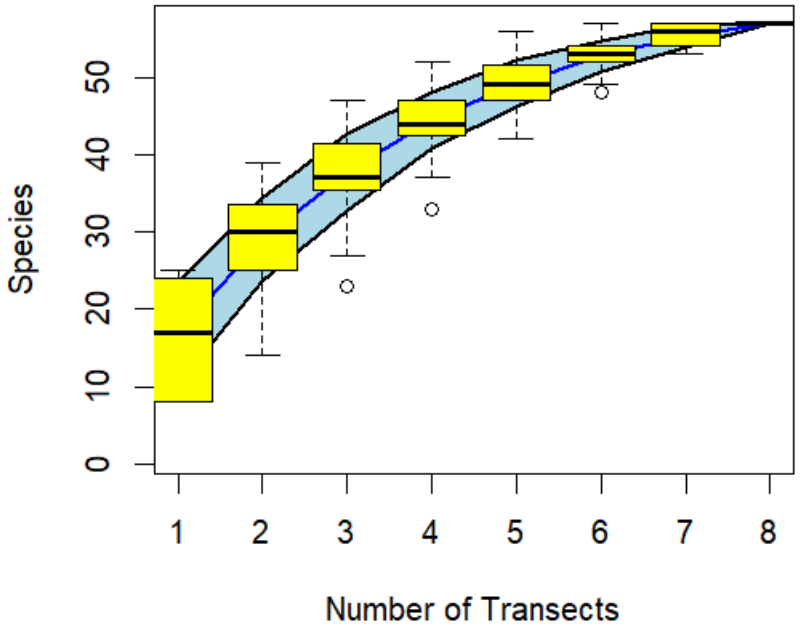

Figure 3. Species accumulation curve in Long Point National Wildlife Area (LPNWA) among 8 50 m long transects where 808 points at 50 cm intervals invaded by *Phragmites australis* (Phragmites) were sampled, which observed a total of 39 species (S = 39).

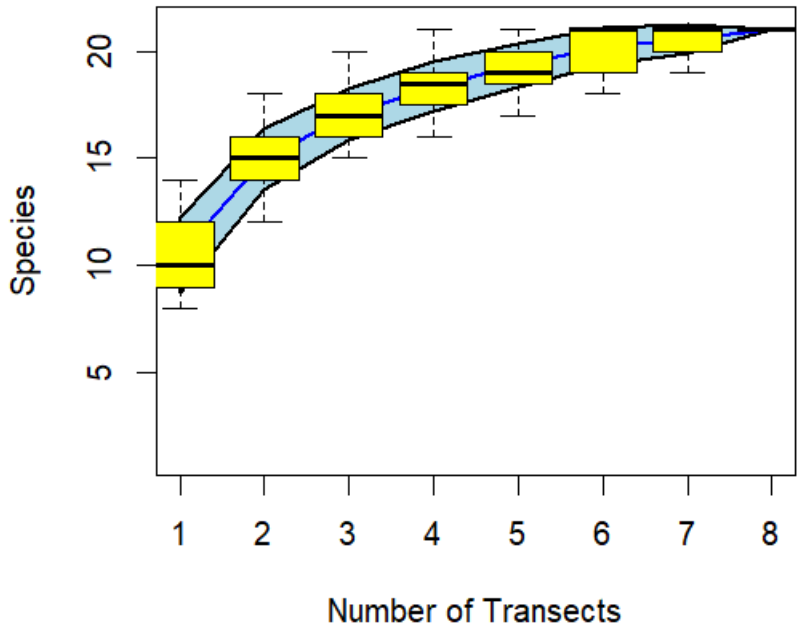

Figure 4. Species accumulation curve in Big Creek Point National Wildlife Area (BCNWA) among 8 50 m long transects where 808 points at 50 cm intervals with no *Phragmites australis* invasion (Reference) were sampled, which observed a total of 13 species (S = 13).

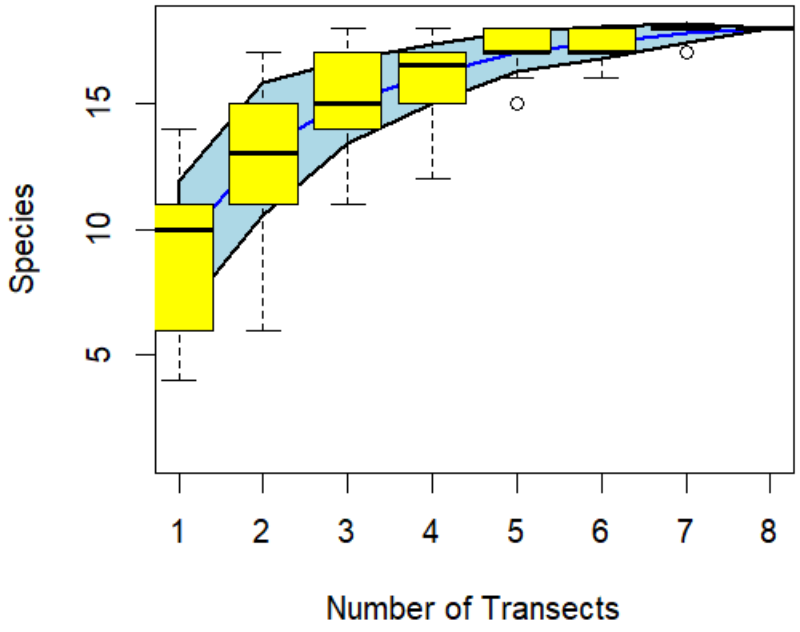

Figure 5. Species accumulation curve in Big Creek National Wildlife Area (BCNWA) among 8 50 m long transects where 808 points at 50 cm intervals formerly invaded by *Phragmites australis* but then treated with herbicide (Treated) were sampled, which observed a total of 12 species (S = 12).

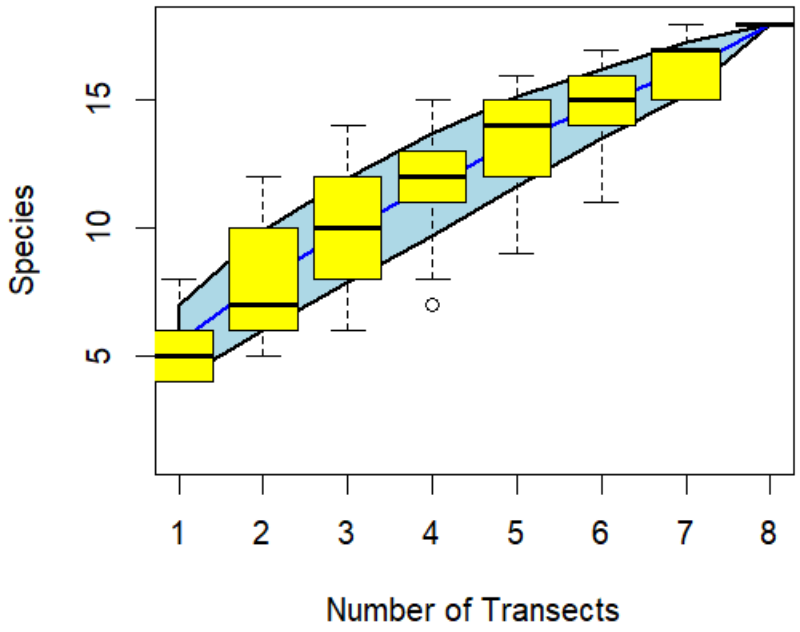

Figure 6. Species accumulation curve in Big Creek National Wildlife Area (BCNWA) among 8 50 m long transects where 808 points at 50 cm intervals invaded by *Phragmites australis* (Phragmites) were sampled, which observed a total of 12 species (S = 12).

Q3. In a table, indicate whether you think the number of transects was adequate to accurately represent the diversity of plants in three habitat types across each of our two NWAs, based on the species accumulation curves. Was one more thoroughly sampled than the other? Explain.

Table 1. Summary assessment of sampling adequacy of number of transects for plant diversity in three habitat types (Reference, Treated, Phragmites) across Long Point National Wildlife Area (LPNWA) and Big Creek National Wildlife Area (BCNWA) based on species accumulation curves.

Sites	Sampled Enough?
LP Reference	Yes
LP Treated	Yes
LP Phragmites	No
BC Reference	No
BC Treated	Yes
BC Phragmites	No

Based on the species accumulation curves, there was adequate sampling for the following sites: LP Reference, LP Treated, and BC Treated, while more sampling would have been beneficial for LP Phragmites, BC Reference, and BC Phragmites. Regarding the first three, more sampling would not have been necessary since as the number of transects increased, the variance in the number of species observed continuously decreases with the 7th transect having very minimal to no variance at all, indicating that it is unlikely that species have been missed. These three curves also reached plateaus, suggesting that adding more transects would not observe and sample significantly more species.

On the other hand, the latter three would benefit greatly from more sampling since there remains relatively high variance in the number of species observed by the 7th transect and have not yet reached lengthy plateaus. Ultimately, there was stronger sampling completed for LPNWA compared to BCNWA, and sampling of the Phragmites habitat type could be improved throughout both locations.

Q4. Include two figures, one comparing singleton data by Location and another comparing singleton data by HabitatType. What does this information tell you about these vegetation communities? What does the number of uniques tell us about the two NWAs?

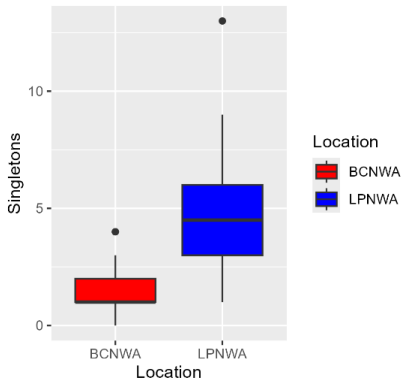

Figure 7. Boxplot showing the distribution of singleton species (y-axis) across two sampling locations: Long Point National Wildlife Area (LPNWA) and Big Creek National Wildlife Area (BCNWA) (x-axis). Each box represents the interquartile ranges (IQR) with the median shown as a horizontal line. Outliers are displayed as individual points. The plot highlights variability in singleton occurrences across locations.

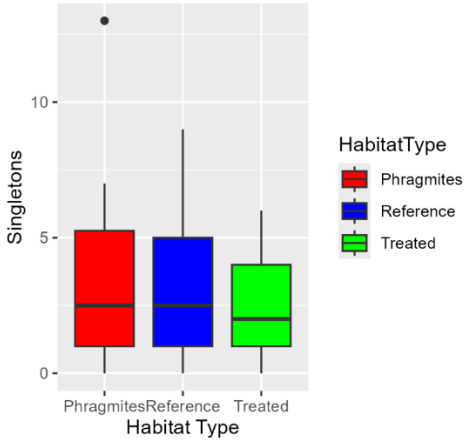

Figure 8. Boxplot showing the distribution of singleton species (y-axis) across three habitat types: Reference, Treated, and Phragmites within LPNWA and BCNWA (x-axis). Each box represents the interquartile ranges (IQR) with the median shown as a horizontal line. Outliers are displayed as individual points. The plot highlights variability in singleton occurrences across locations.

Singletons refer to species that are represented by a single individual in a transect or sample, which is often used to identify rare species and evaluate the completeness of sampling efforts. When comparing singletons by location, evidently there is a larger interquartile range (IQR) in the LPNWA, suggesting more variation and thus richer, more diverse species in this location. However, when comparing singletons by habitat type, there is no significant difference between the three IQRs, but it is worth noting that the Phragmites sites have the largest IQRs while Treated has the lowest. These observations suggest that habitats invaded by *Phragmites australis* are more species diverse than those treated with herbicide.

Regarding uniques, this refers to species that are found in only one sampling site, which may highlight species restricted to specific locations. The sum of uniques found in the LPNWA and BCNWA are 7 and 0, respectively. Similarly to the discussion on singletons, LPNWA has evidently more compared to BCNWA, such that 7 species are exclusively found in LPNWA but not the BCNWA. This may suggest that the LPNWA has higher habitat specialization to provide specialized environments for species, greater biodiversity contribution, and may also reflect the better sampling of the LPNWA as described in Q3.

Q5. Explain how a species could be a unique, but not a singleton. Based on the definitions of these terms, what is the foundation for this difference?

Although a species is a unique due to being found in one exclusive sampling location, it may occur multiple times within that location, which strips the species of its singleton status. Thus, the foundation for this difference lies in spatial occurrence versus frequency of observation. For example, uniques are spatially restricted to a single location, but may occur frequently in that location, while singletons are “rare” species due to its very low frequency by appearing only once across all locations.

Q6. Include all six figures that you generate with the observed species richness and four different estimators. Include captions for each plot.

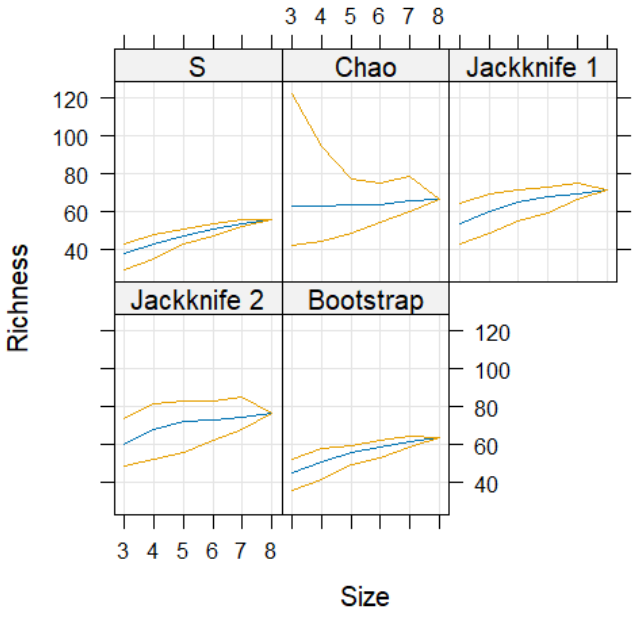

Figure 9. Observed species richness (S) and estimations using four different incidence-based estimators: Chao, First order jackknife, Second order jackknife, and Bootstrap based on data from Long Point National Wildlife Area (LPNWA) among 8 50 m long transects where 808 points at 50 cm intervals with no *Phragmites australis* invasion (Reference) were sampled.

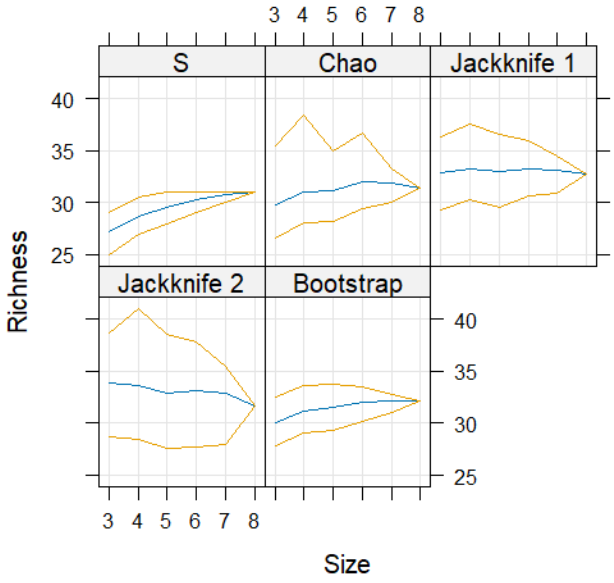

Figure 10. Observed species richness (S) and estimations using four different incidence-based estimators: Chao, First order jackknife, Second order jackknife, and Bootstrap based on data from Long Point National Wildlife Area (LPNWA) among 8 50 m long transects where 808 points at 50 cm intervals formerly invaded by *Phragmites australis* but then treated with herbicide (Treated) were sampled.

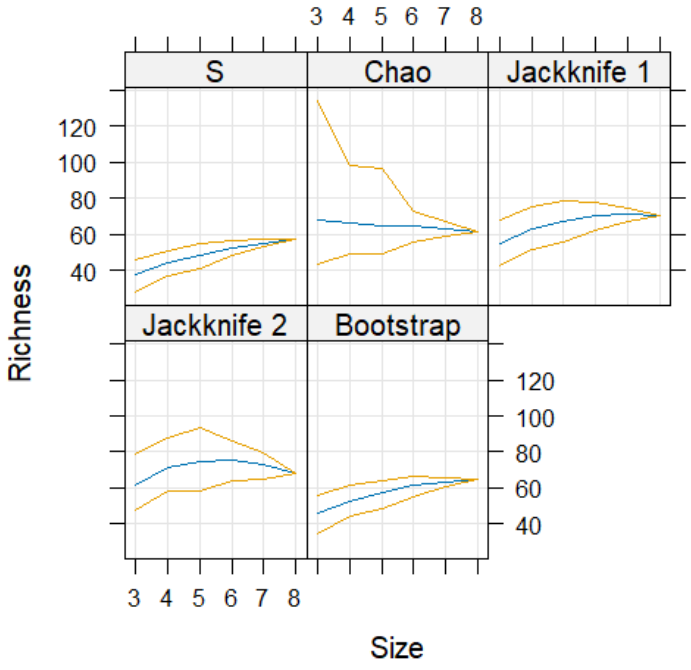

Figure 11. Observed species richness (S) and estimations using four different incidence-based estimators: Chao, First order jackknife, Second order jackknife, and Bootstrap based on data from Long Point National Wildlife Area (LPNWA) among 8 50 m long transects where 808 points at 50 cm intervals invaded by *Phragmites australis* (Phragmites) were sampled.

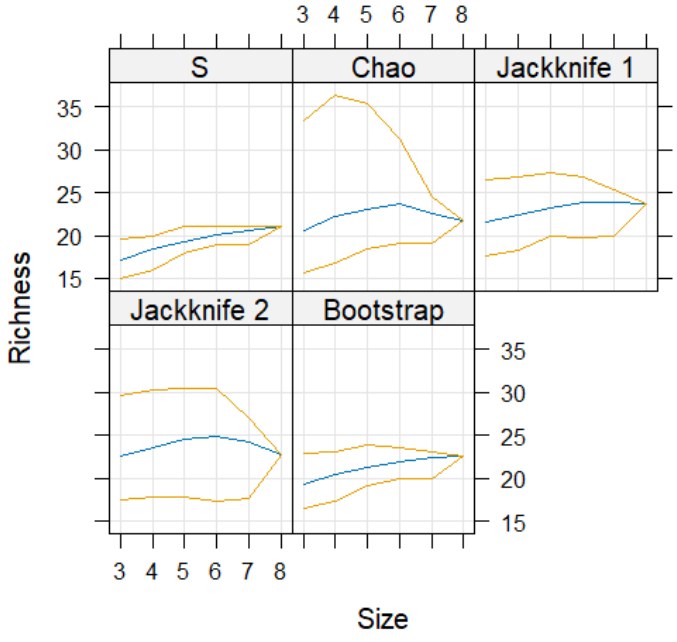

Figure 12. Observed species richness (S) and estimations using four different incidence-based estimators: Chao, First order jackknife, Second order jackknife, and Bootstrap based on data from Big Creek National Wildlife Area (BCNWA) among 8 50 m long transects where 808 points at 50 cm intervals with no *Phragmites australis* invasion (Reference) were sampled.

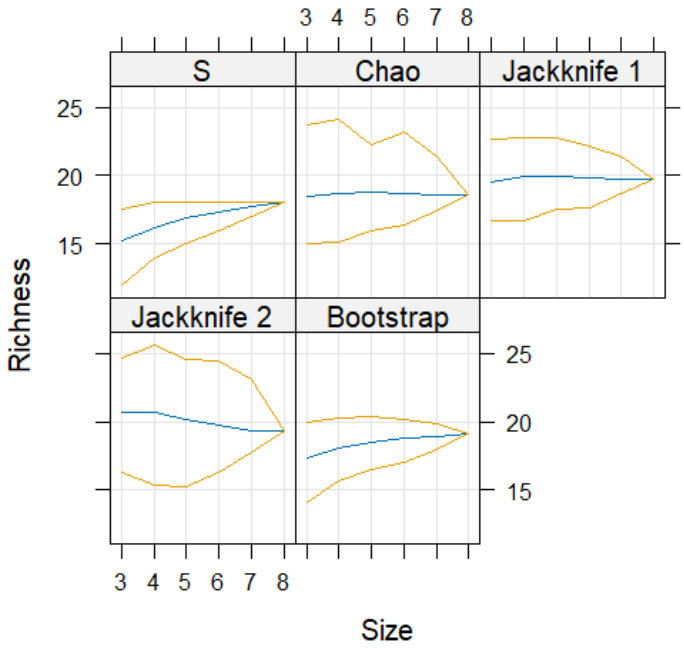

Figure 13. Observed species richness (S) and estimations using four different incidence-based estimators: Chao, First order jackknife, Second order jackknife, and Bootstrap based on data from Big Creek National Wildlife Area (BCNWA) among 8 50 m long transects where 808 points at 50 cm intervals formerly invaded by *Phragmites australis* but then treated with herbicide (Treated) were sampled.

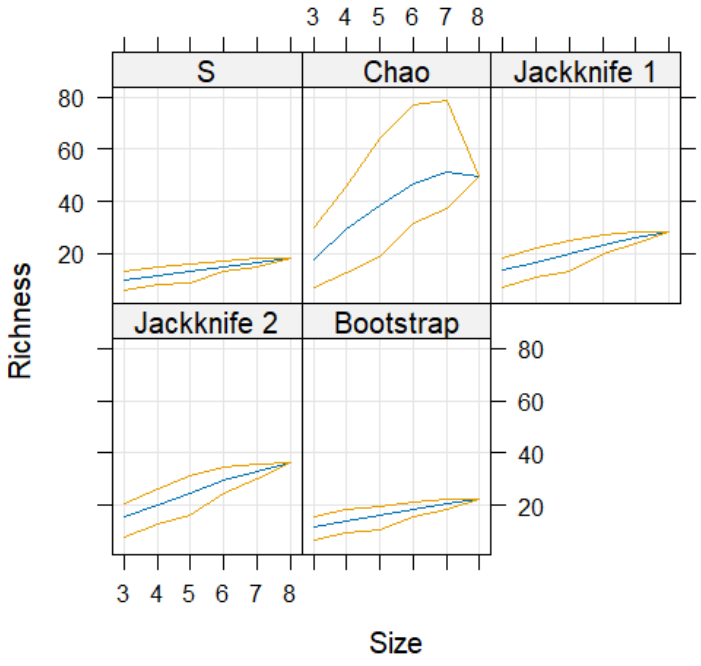

Figure 14. Observed species richness (S) and estimations using four different incidence-based estimators: Chao, First order jackknife, Second order jackknife, and Bootstrap based on data from Big Creek National Wildlife Area (BCNWA) among 8 50 m long transects where 808 points at 50 cm intervals invaded by *Phragmites australis* (Phragmites) were sampled.

Q7. In a table, show the total richness observed at each of the NWAs across the three habitat types. In additional columns, show the maximum estimated richness and what estimated yielded it and the minimum estimated richness and its estimator from among the 4 estimators. Based on the descriptions given and your analysis, do you think that some estimators are consistently over- or underestimating "true" richness, relative to the others? Which is most conservative? Which type of estimator(s) would you choose to use if you were doing a thesis project looking at the diversity of plants in wetlands?

Table 2. Observed and estimated species richness at Long Point National Wildlife Area (LPNWA) and Big Creek National Wildlife Area (BCNWA) across three habitat types (Reference, Treated, Phragmites) with corresponding maximum and minimum incidence-based estimators.

Site	Total Observed Richness (S)	Maximum Estimated Richness	Maximum Estimated Richness Estimator	Minimum Estimated Richness	Minimum Estimated Richness Estimator
LP Reference	56.00	76.89286	Second Order Jackknife	63.77774	Bootstrap
LP Treated	31.00	32.75000	First Order Jackknife	31.43750	Chao
LP Phragmites	57.00	70.12500	First Order Jackknife	61.68750	Chao
BC Reference	21.00	23.62500	First Order Jackknife	21.78750	Chao
BC Treated	18.00	19.75000	First Order Jackknife	18.58333	Chao
BC Phragmites	18.00	49.50000	Chao	22.32392	Bootstrap

Yes, there are estimators that consistently overestimate the "true" richness, relative to others, such as the Jackknife estimators. In nearly all six combinations of NWAs and habitat types, the Jackknife estimators do not resemble a horizontal line that meets the species accumulation curve where it asymptotes and instead continues to climb and even decline as more transects are added. A few exceptions include Figure 10 and 13 where the Jackknife estimators are relatively flat, but still end above the species accumulation curve asymptote, which is supported by the many maximum estimated richness values coming from Jackknife estimators as illustrated in Table 2.

On the other hand, the Chao estimator is the most conservative. Across nearly all six combinations of NWAs and habitat types, the Chao estimator remained the most flat and horizontal and met the species accumulation curve asymptote. Its conservative nature is also supported by its many minimum estimated richness values as illustrated in Table 2. Since the Chao estimator most closely followed the characteristics of an ideal estimator, it would be the best type of estimator to use for a thesis project looking at the diversity of plants in wetlands. As a secondary opinion, the Bootstrap estimator would be a decent choice since it also did not overly over- or underestimate a majority of the six scenarios.

Q8. Based on the table you created for Q7, which habitat type and location do you think was the most adequately sampled? Compare your conclusions from this table with your interpretation of the species accumulation curves in Q2.

Based on the results in Table 2, the Treated habitat type across both the LPNWA and BCNWA locations were the most adequately sampled. For both habitat type and location combinations, the range between the minimum and maximum estimated richness were narrow and close to the total observed richness, indicating the convergence of multiple approaches reaching a similar estimate. In comparison to the conclusions drawn from the species accumulation curves, these conclusions continue to hold true.

Q9. What are the three diversity indices that the function *diversity* can calculate? What is the formula for the *invsimpson* index? How does this differ from *Simpson*? How does a high *Simpson* value differ from a high *invsimpson*?

The three diversity indices calculated by the function *diversity* include Shannon, Simpson, and InvSimpson. For the InvSimpson index, it is mainly based on $D = \sum {p_i}^2$ and returns 1/D, which differs from the 1-D of Simpson. Although both Simpson and InvSimpson are related measures, the two differ due to interpretations and scales. For example, Simpson measures the probability that two randomly selected individuals from a community belong to the same species, so a high Simpson value means that the community is dominated by one or a few species and reflects low diversity. Alternatively, InvSimpson represents diversity, so a high InvSimpson value indicates high diversity and even distribution of species with no domination of a single species.

Q10. Before you look at the results, consider which HabitatType you would expect to have the smaller Simpson's Inv Mean value: Phragmites or Reference? Record your prediction and justify it. Now, compare the Simpson's Inv Mean values that you calculated in R for the six difference HabitatType by Location combinations and record them in a table.

It is expected that the Phragmites habitat type would have the smaller Simpson's Inv Mean value, thus reflecting low diversity and unequal distribution of species due to the domination of the *Phragmites australis* species. This expectation is based on the invasive nature of the *Phragmites australis* plants that outcompete native wetland plants, which would reduce biodiversity.

Table 3. Simpson's Inverse Mean values for six combinations of Reference, Treated, and Phragmites habitat types at Long Point National Wildlife Area (LPNWA) and Big Creek National Wildlife Area (BCNWA).

Site	Simpson's Inv Mean
LP Reference	6.723871
LP Treated	7.547399
LP Phragmites	6.268833
BC Reference	4.562585
BC Treated	3.127329
BC Phragmites	2.133628

As expected, the Phragmites habitat type across both the LPNWA and BCNWA locations had the lowest Simpson's Inv Mean value, suggesting that these areas that had been invaded by *Phragmites australis* lost diversity due to the invasive species outcompeting native plant species. In contrast, the Reference habitat type had higher Simpson's Inv Mean values since there had been no negative influence of the presence of an invasive plant, enabling native species to flourish. For the Treated habitat type, the BCNWA location followed an expected trend such that the use of herbicides likely killed off the Phragmites plant and native plants were able to move in and recover diversity, but not entirely. However, the Treated habitat type of the LPNWA location shows an interesting result with the highest Simpson's Inv Mean, indicating a high diversity greater than the Reference habitat type. Although surprising, plausible explanations may include competitive release, colonization of new species, etc.

Q11. Looking at the *merged_df*, are the Sorenson, Jaccard, and Bray-Curtis indices of similarity or dissimilarity? How can you tell by looking at the data? What does it mean if two sites have an index value of 0?

The Sorenson, Jaccard, and Bray-Curtis indices are of dissimilarity. This is deduced from the data by comparing two of the exact same sites, which will result in all indices returning a value of 0. For dissimilarity indices, a value of 0 implies the sites are completely similar. To further support this, when comparing different sites, the indices are all greater than 0 and capped at 1, which refers to sites that are not similar and thus very different from each other.

Q12. Graph the relationship between the Sorenson and Jaccard indices, present this figure with axis titles and a caption. Which pair vegetation types are the most similar? Which pair of vegetation types are the least similar? Be sure to include the actual index values in text to justify your reasoning.

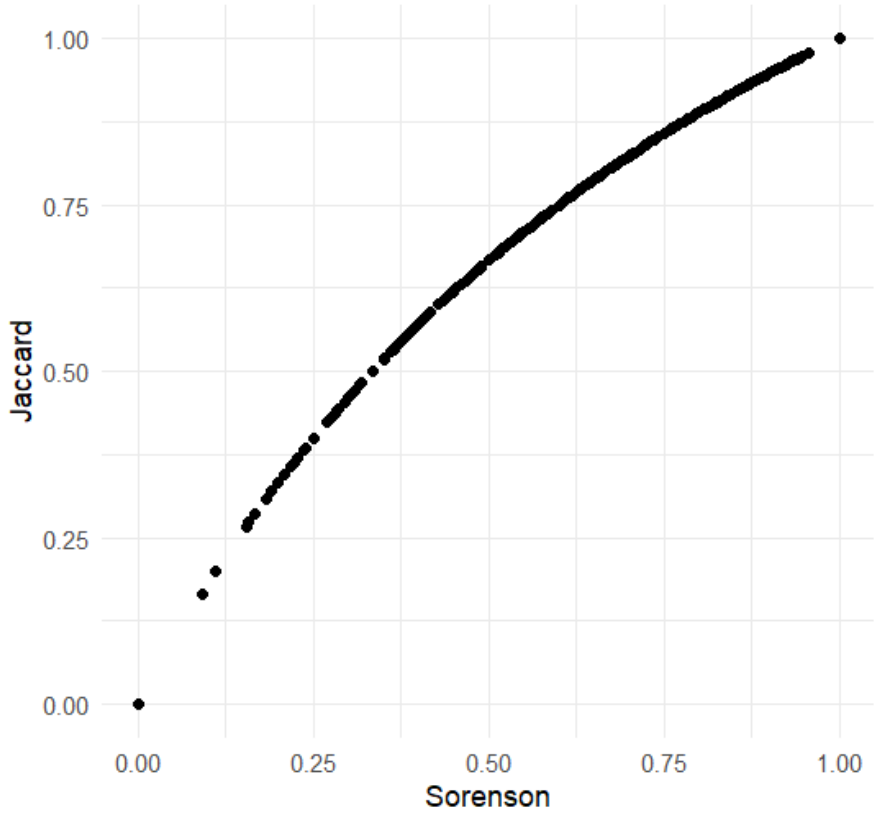

Figure 15. Mathematical relationship between the Sorenson index and the Jaccard index across varying community compositions to illustrate their correlation in quantifying community similarity.

The Treated-Treated pair is the most similar as the means of the indices are the lowest and closest to 0 (i.e., 0.541 Bray-Curtis, 0.520 Sorenson, and 0.650 Jaccard), which for dissimilarity indices, represents sites that are more similar. Alternatively, the Phragmites-Reference pair is the least similar as the means of the indices are the highest and closest to 1 (i.e., 0.876 Bray-Curtis, 0.747 Sorenson, 0.846 Jaccard), which for dissimilarity indices, represents sites that are more dissimilar.

Q13. As you did in question 12, plot the relationship between the Sorenson and the Bray-Curtis indices. Present this figure with axis titles and a caption. Describe the relationship between the two in one sentence. What is the reason for this in terms of the input data used to calculate the two indices? How could the two indices yield different values for the same pair of sites?

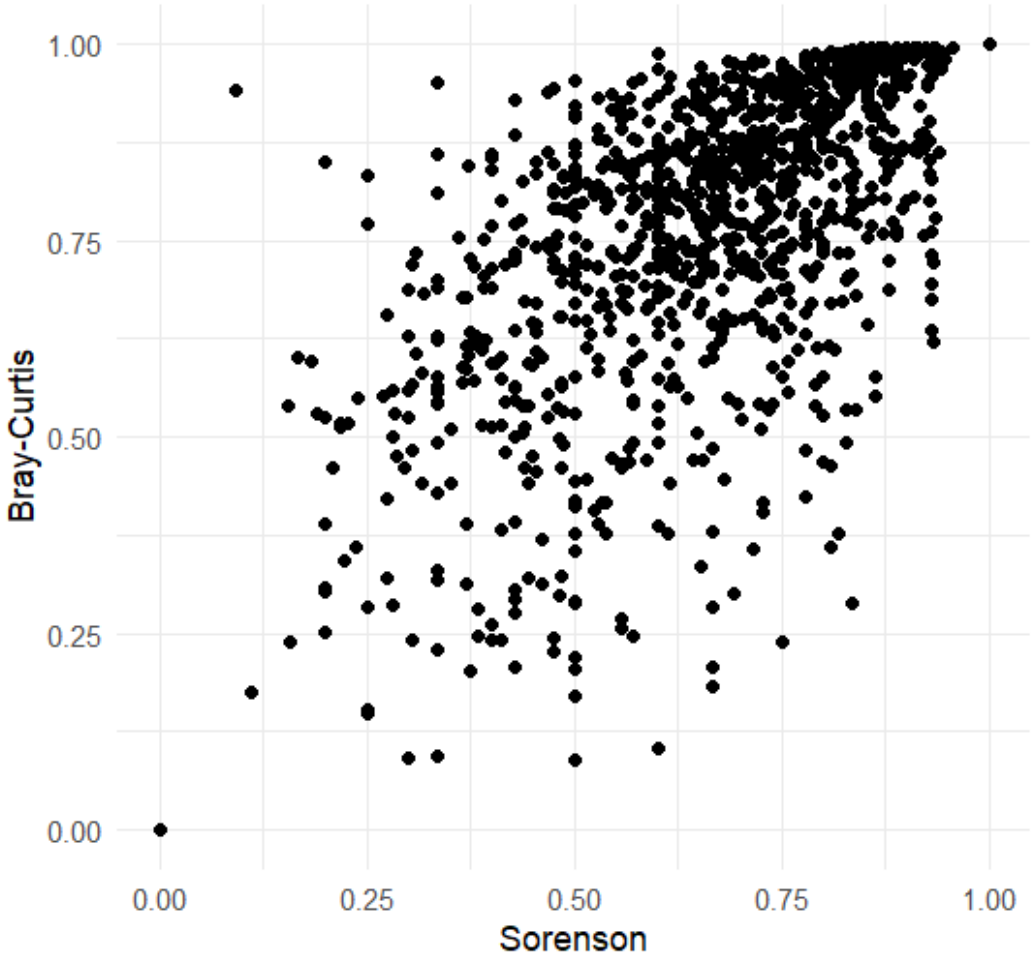

Figure 16. Mathematical relationship between the Sorenson index and the Bray-Curtis index across varying community compositions to illustrate their correlation in quantifying community similarity.

The Sorenson and Bray-Curtis indices are generally positively correlated, but there are evident variations such that one index may exhibit slightly or significantly lower or higher values, particularly towards the origin where species are expected to be completely similar. This variation and non-linear relationship stem from the binary abundance data input. Sorenson typically uses binary presence-absence data so the data does not undergo manipulation prior to calculating the index, while Bray-Curtis uses the actual abundance of each species which requires a presence-absence standardization prior to its calculation. Due to this, sites with similar species compositions but highly unequal abundances will yield different results between the indices such that Sorenson will treat all species equally, while Bray-Curtis will account for abundance differences despite sites having shared species.

Q14) Based on your Sorensen and Bray-Curtis index values, has removal of *Phragmites australis* in the treated transects made the Treated transects more similar to the Reference transects in the two NWAs, or are they still more like the Phragmites transects? Would you consider restoration of this site to have been successful in terms of the vegetation community composition and diversity? Why or why not?

Yes, the removal of *Phragmites australis* in the treated transects has made the Treated transects more similar to the Reference transects in the two NWAs, and more dissimilar to the Phragmites transects. Previously, the Bray-Curtis and Sorenson indices for the Treated-Phragmites pair was 0.821 and 0.693, respectively, and are now 0.746 and 0.628, respectively, for the Treated-Reference pair. Evidently from these values, the indices are lower for the Treated-Reference pair, indicating more similarity.

Generally, yes, these values would imply that restoration of this site has been successful in terms of the vegetation community composition and diversity. With transects becoming more similar to the Reference transects, it suggests that the herbicide was able to halt the damage of *Phragmites australis*, thus enabling native plants to move in and recover native plant diversity. However, it is worth noting that the difference in indices between the Treated-Phragmites and Treated-Reference pairs is not major. Yes, there has been improvements, but the Treated transects are still not completely similar to the Reference transects (i.e., the indices are not 0), so further work will be needed.

Q15) If you were working for the Nature Conservancy of Canada and you were able to purchase the land adjacent to one of these NWAs to protect in perpetuity, which NWA would you select to maximize the conservation of biodiversity and why?

Purchasing the land adjacent to the LPNWA would maximize the conservation of biodiversity. After comparing the Simpson's Inv Mean values in Table 3, it is evident that LPNWA had higher values, which reflects high diversity and more equal distribution of species. Compared to the BCNWA values, LPNWA values were 2-3x greater, which implies that the LPNWA is in a better biodiversity position than BCNWA. Similarly, Table 2 illustrates the greater species richness of LPNWA compared to BCNWA. Additionally, after the herbicide treatment, the Treated site of the LPNWA had a greater Simpson's Inv Mean value than its Reference site, while BCNWA's Treated site remained short of the Reference. LPNWA's Phragmites site was also relatively similar to the LPNWA Reference site, indicating that although a *Phragmites australis* invasion occurred, it had not significantly reduced biodiversity as it had with BCNWA where its Reference to Phragmites site value halved. Thus, with LPNWA's current biodiversity state, it provides a strong, promising foundation to protect and focus conservation efforts towards.