

AMOD-5610H/5620H

Project Report

Woodlands and Waterways EcoWatch (WWEW)

**Investigating Benthic Macroinvertebrate Communities in
Haliburton County Lakes**

Team Number: 9

Teammates: Dakshkumar(0792827), Devkumar(0792519)

Date: December 9, 2024

Introduction

The ecological health of freshwater lakes is increasingly recognized as a critical component of environmental sustainability, particularly in regions like Haliburton County, where natural resources play a vital role in local communities and biodiversity. Freshwater ecosystems provide essential services, including water filtration, habitat for diverse species, and recreational opportunities. However, these ecosystems face numerous threats from human activities, climate change, and invasive species, leading to shifts in water quality and habitat conditions that can adversely affect aquatic life.

Benthic macroinvertebrates, which include organisms such as insects, crustaceans, and mollusks, inhabit the sediment at the bottom of lakes and rivers. These organisms are integral to the aquatic food web and serve as key indicators of ecosystem health due to their sensitivity to environmental changes. Their presence, absence, or abundance can provide insights into the quality of water and the overall ecological integrity of the ecosystem. For instance, certain species are intolerant to pollution, making them valuable for assessing water quality and identifying potential ecological stressors.

This project aims to investigate the relationships between benthic macroinvertebrate community structure and various environmental drivers in multiple lakes across Haliburton County over a five-year period. The primary objectives include analyzing changes in benthic community composition over time, identifying the key environmental factors influencing these communities, and assessing the influence of trophic status across different riparian zones.

To achieve these objectives, a comprehensive analysis will be conducted using existing datasets that encompass water chemistry, riparian characteristics, and sediment types. The approach will integrate statistical techniques such as Principal Component Analysis (PCA) and more to discern patterns in community structure and their correlations with environmental variables.

The motivation for this study stems from the pressing need for data-driven decision-making in the management and conservation of Haliburton County's aquatic resources. By understanding the dynamics of benthic communities in relation to their environment, stakeholders—including local governments, conservation organizations, and the public—can make informed choices to enhance ecosystem health and resilience. The findings from this study are expected to contribute significantly to the existing body of knowledge and guide future conservation strategies, ensuring the sustainability of Haliburton County's lakes for generations to come.

Literature Review

It is commonly known that benthic macroinvertebrates are useful bioindicators for determining the ecological health and water quality of aquatic habitats. Their usefulness in long-term monitoring programs stems from their sensitivity to different contaminants and changes in the environment. The function of benthic macroinvertebrates in assessing the health of aquatic ecosystems, their reaction to environmental stressors, and their connection to trophic status have all been clarified by a number of studies. In a research published in 2006, Bonada et al.(2006) emphasized the significance of benthic macroinvertebrates in biomonitoring and emphasized how sensitive they are to pollution gradients, including nutrient enrichment and organic contamination. Changes in community composition are discussed as indicators of changes in water quality, especially when species that are vulnerable to pollution, like as Ephemeroptera, Plecoptera, and Trichoptera (EPT taxa), are present. This is essential to comprehending the long-term responses of these creatures to environmental perturbations, which is consistent with the project's analysis of long-term community changes.

The impacts of nutrient enrichment on freshwater ecosystems were investigated by Larsen et al. (2011), with a particular emphasis on the responses of various trophic levels including benthic macroinvertebrates to eutrophication. Their research sheds light on the ways that nutrient imbalances such as an overabundance of phosphorus and nitrogen affect the composition of macroinvertebrate communities. This is especially important to consider when evaluating how trophic status affects your research. Friberg et al. (2011) conducted a pertinent study that investigates the effects of environmental variables on benthic macroinvertebrate populations. These drivers include temperature, pH, and changes in land use. The scientists discovered that by raising sedimentation and nutrient levels and thus changing benthic communities, anthropogenic activities like deforestation and agricultural runoff had a substantial impact on the general health of the waterbody. These results offer a framework for determining the main environmental factors influencing the populations of benthic macroinvertebrates in the lakes of Haliburton County.

A comprehensive investigation on the suitability of benthic macroinvertebrates as water quality indicators in various freshwater habitats across Europe was carried out by Hering et al. (2006). Their results highlight the validity of evaluating a variety of pollutants, such as pesticides, heavy metals, and organic pollutants, utilizing macroinvertebrate populations. Contextualizing the heterogeneity in your dataset from Haliburton County lakes requires highlighting regional differences in macroinvertebrate population responses to environmental stresses, which is another important aspect of the study.

Dataset

The dataset for this project consists of benthic macroinvertebrate sampling data collected from multiple lakes in Haliburton County over a five-year period. The dataset is structured to capture both environmental conditions and biological data, allowing for comprehensive analysis of the relationships between environmental factors and benthic macroinvertebrate community structure.

- **Attributes:**

Environmental Variables:

Water Chemistry:

Water Temperature (°C): Measured at each sampling event, as temperature affects metabolic rates and species distributions.

Dissolved Oxygen (DO) (mg/L): Indicates oxygen levels available for aquatic life, with higher levels generally supporting more diverse benthic communities.

Conductivity (µS/cm): Reflects the concentration of ions in the water, which can indicate nutrient levels and pollution.

pH: Measures water acidity/alkalinity, which influences the availability of nutrients and affects organism health.

Riparian and Habitat Characteristics:

Riparian Vegetation: Documented in three zones (1.5-10m, 10-30m, 30-100m from shore), covering types such as forested areas, which impact runoff and sedimentation.

Sediment Composition: Describes the dominant and secondary mineral substrates at each sampling location (e.g., gravel, sand, cobble). These physical characteristics determine habitat suitability for different benthic species.

Macrophyte Coverage: The presence of emergent, floating, and submergent plants is noted, as macrophytes provide habitat structure and affect nutrient dynamics.

Algae Presence: The dataset records floating, filamentous, and attached algae types, which can influence food availability and habitat conditions for benthic communities.

Benthic Macroinvertebrate Data:

Species Counts: The dataset includes counts of key taxa, such as Ephemeroptera (mayflies), Trichoptera (caddisflies), Amphipoda (shrimps), and Chironomidae (non-biting midges). These organisms are widely used as biological indicators, with some being more sensitive to pollution and environmental changes than others.

Diversity and Abundance Indices: The dataset also provides metrics like Simpson's Diversity Index and the Hilsenhoff Biotic Index, which help quantify species richness and community composition in relation to environmental quality.

Sampling Event Information:

Date and Time of Sampling: Provides temporal context for analyzing seasonal and annual changes in benthic communities.

Geospatial Information (Latitude/Longitude): The precise location of each sampling site is included, allowing for spatial analysis across lakes.

Sampling Effort: Information about the length of sampling areas and the time spent sampling helps standardize comparisons between sites.

Size and Format: The dataset consists of multiple Excel spreadsheets, with separate sheets for each lake and rows representing individual sampling events. Environmental and biological data are stored in numerical or categorical formats, ready for statistical analysis. Each lake has been sampled multiple times, resulting in thousands of observations across the five-year period, with each sampling event contributing about 50-100 individual data points.

Assumptions:

Temporal Stability: We assume that water chemistry and riparian characteristics do not change drastically within short time periods. For instance, measurements of pH, DO, and conductivity are considered representative of the lake's general conditions during the sampling period.

Data Consistency: The sampling methods have remained consistent across all years, meaning that differences observed in the data reflect genuine changes in the environment or benthic communities, not artifacts of differing methodologies.

Spatial Homogeneity: Within each lake, the sampling sites are assumed to be representative of the overall lake conditions, though some heterogeneity may exist due to varying local conditions such as inflows and human activities.

This rich dataset will allow us to perform robust statistical analyses to explore the drivers of benthic community structure, identify trends over time, and assess the health of the lakes in the context of their surrounding environments.

Methodology

We will employ a combination of multivariate statistical techniques to investigate the relationships between environmental variables and benthic macroinvertebrate community composition. The primary methods are Principal Component Analysis (PCA) and Permutational Multivariate Analysis of Variance (PERMANOVA).

To investigate the relationships between benthic community numbers and environmental factors, we employed a multi-faceted analytical approach combining statistical, visual, and machine learning techniques. Initially, we conducted correlation analysis to quantify relationships between numerical variables, such as water chemistry parameters (dissolved oxygen, conductivity, and pH) and benthic diversity indices. This helped us identify significant linear and non-linear associations between variables. Visual analysis followed, utilizing scatter plots, box plots, and heatmaps to detect patterns, trends, and outliers, enabling a deeper understanding of the data distribution and potential connections between benthic community numbers and environmental factors like riparian characteristics, sediment types, algae, and macrophyte compositions.

Next, we applied regression analysis to quantify the impact of individual environmental factors on benthic community numbers, providing insight into the predictive relationships. We explored both simple and multiple regression models to ensure robustness. Additionally, feature importance was assessed using machine learning models, such as Random Forest, to identify the most influential environmental variables affecting benthic community numbers. This approach provided a hierarchical view of the predictors, helping prioritize the factors with the greatest ecological relevance. All analyses were performed using a cleaned and standardized dataset covering multiple lakes, ensuring consistency and reliability. The combination of statistical and machine learning methods provided a comprehensive framework to evaluate and interpret complex relationships, facilitating informed conclusions about the drivers of benthic diversity and community composition across diverse environmental settings.

To examine the relationship between benthic community numbers and environmental factors, we focused on key variables such as substrate types, algae, and macrophyte compositions alongside benthic metrics like Total Number Entered and Number of Unique Taxa Entered. Using selected columns, we applied one-hot encoding to convert categorical variables, such as dominant and secondary mineral substrates, into numerical representations for analysis. Correlation analysis was performed to quantify relationships, filtering results specifically for benthic metrics. This approach enabled the identification of significant patterns and associations, providing insights into how environmental factors influence benthic community structure and diversity across various lake ecosystems.

Principal Component Analysis (PCA):

PCA is a dimension-reduction technique that helps identify patterns in high-dimensional datasets by transforming the data into a set of principal components. Each component represents a

combination of environmental variables (e.g., pH, DO, conductivity) that contribute the most to variation in benthic diversity. PCA will allow us to reduce the complexity of the dataset and highlight which environmental factors have the greatest influence on community composition. The output will provide a visual representation (e.g., biplots) that shows the relationships between environmental drivers and benthic taxa.

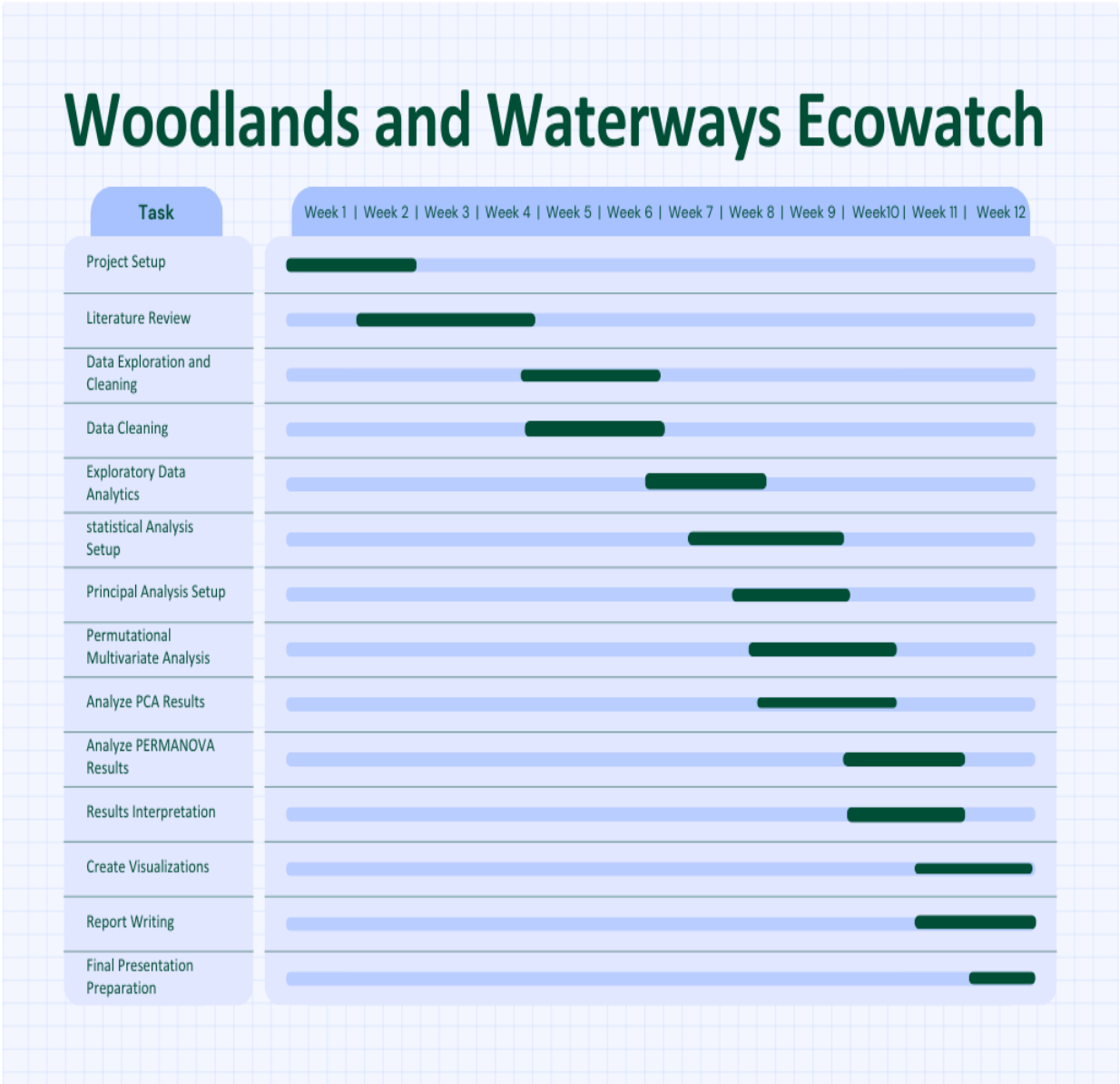
Permutational Multivariate Analysis of Variance (PERMANOVA):

PERMANOVA will be used to test for significant differences in benthic community structure across various environmental gradients, such as water chemistry, riparian zones, and sediment types. Unlike traditional ANOVA, PERMANOVA is well-suited for ecological data, which often violates assumptions of normality. It uses permutation tests to evaluate the significance of the observed differences, making it a robust choice for assessing community-environment interactions. This method will help identify how benthic diversity varies across different lakes and environmental conditions, particularly in headwater systems.

Other Considerations:

In addition to PCA and PERMANOVA, we may calculate diversity indices (e.g., Simpson's Diversity, Hilsenhoff Biotic Index) to quantify ecological health and explore specific taxa responses to environmental changes.

Project Plan



Results

Following are the questions which answered in our project for the given data:

- Investigating relationships between benthic community numbers and environmental factors: water chemistry, riparian characteristics, sediment types, algae and macrophyte compositions.
- PCA analyses to identify the key environmental drivers of benthic diversity, and should incorporate elevation, water chemistry (pH, DO, conductivity)
- Other tests like PERMANOVA might be of interest to conduct for significant correlations
- Assess how the benthic diversity varies across different riparian zones, chem conditions, and if these patterns might differ in headwater lakes?

Answers for each question are explained below:

1. Investigating relationships between benthic community numbers and environmental factors: water chemistry, riparian characteristics, sediment types, algae and macrophyte compositions.

Method 1: Correlation Analysis

Summary of Why These Columns Were Selected:

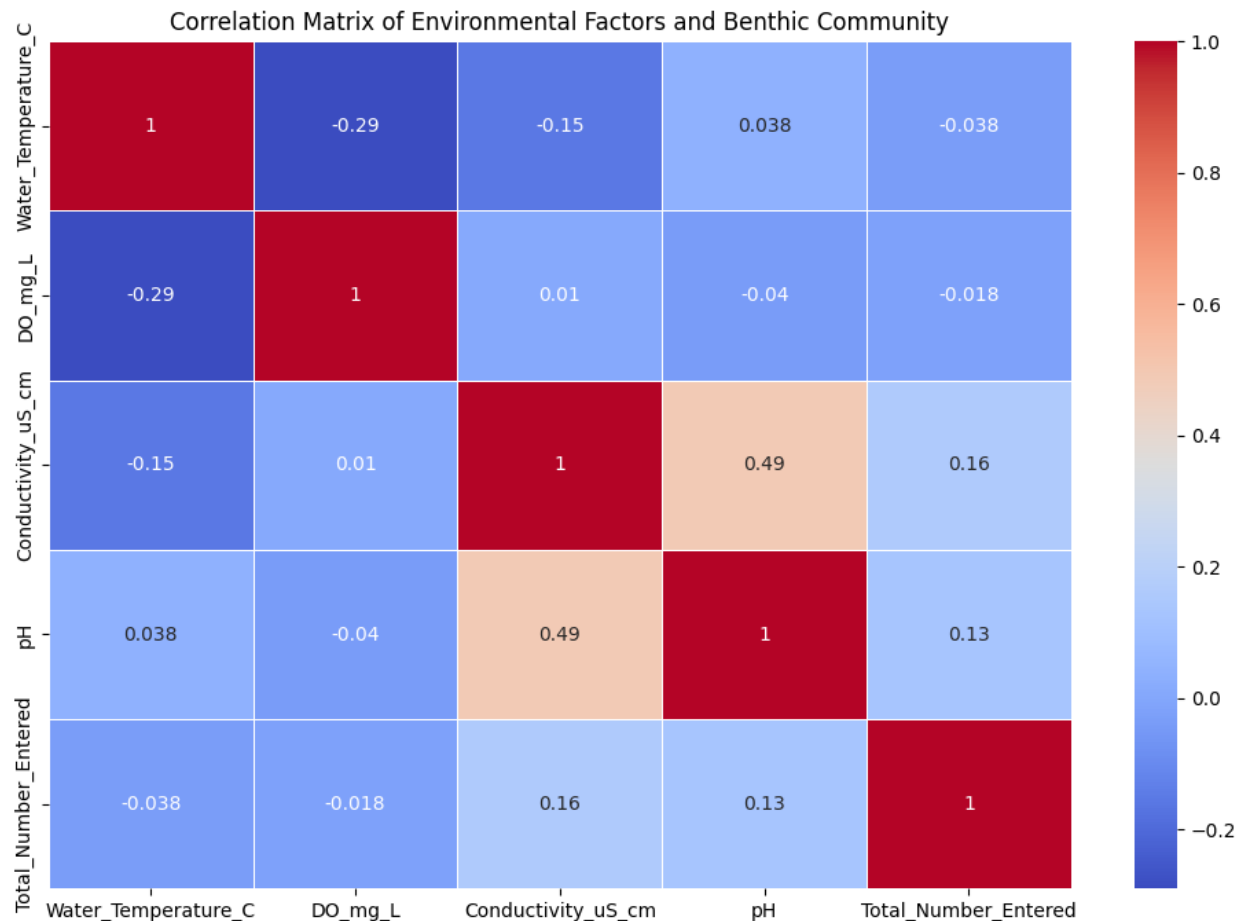
Water Chemistry (Temperature, DO, Conductivity, pH): These factors directly affect the physiological needs and survival of aquatic organisms.

Riparian Characteristics: These influence water quality, habitat structure, and nutrient input, thereby impacting benthic habitats.

Sediment Types (Dominant Substrate): Determines habitat suitability for different benthic species based on their preferences.

Algae and Macrophytes: These provide habitat, food, and oxygen but can also contribute to water quality issues if overgrown.

Benthic Community Count: Serves as the measure of ecosystem health and biodiversity, allowing us to see how environmental changes impact aquatic life.



Key Takeaways (Simple Version): No Strong Impact: None of the environmental factors (like water temperature, pH, or dissolved oxygen) strongly affect the benthic community numbers in your dataset.

Weak Relationships:

Water Temperature and Dissolved Oxygen have a weak negative link (as temperature goes up, dissolved oxygen tends to go down). Conductivity (how well water carries electricity) has a slight positive link with the number of benthic organisms, but it's very weak. pH (how acidic or alkaline the water is) has a minor positive effect on benthic counts.

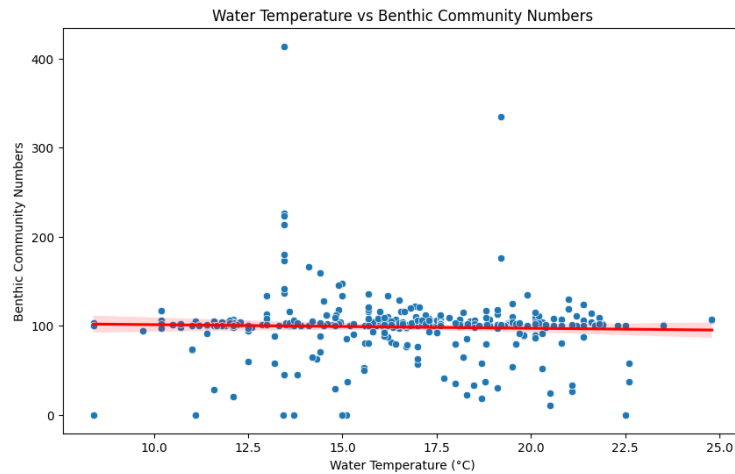
Moderate Connection:

The only moderate connection is between Conductivity and pH. This suggests that when water has more ions (higher conductivity), it's usually less acidic (higher pH).

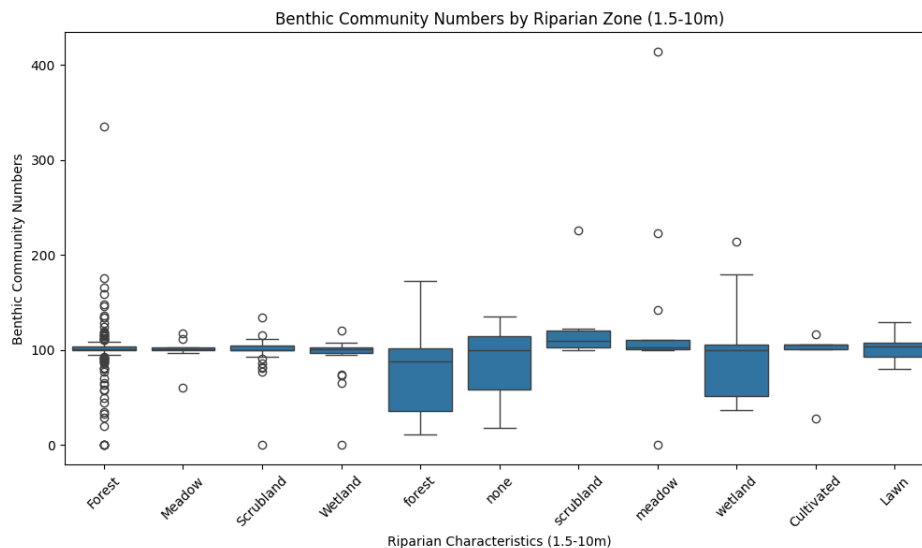
Conclusion:

These factors don't seem to have a big influence on the benthic community numbers, so you might need to look at other factors (like nutrients, habitat, or non-linear effects) to explain changes in benthic numbers.

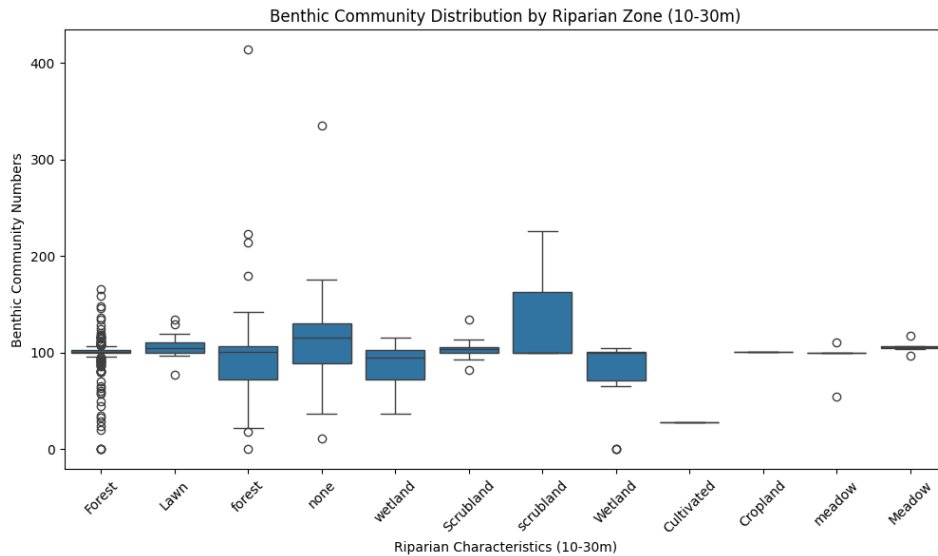
Method 2: Visual Analysis



Key Insights: From this visualization, water temperature does not appear to have a significant effect on benthic community numbers. This suggests that other environmental factors (e.g., dissolved oxygen, pH, or substrate types) might play a more critical role in determining benthic numbers, and further analysis is needed to explore those.



Key Insights: Forested areas and wetland zones tend to support stable benthic populations, possibly due to better habitat conditions and protection from disturbances. The "none" category, which lacks significant riparian vegetation, shows more variability, indicating that the absence of riparian zones could lead to unpredictable conditions for benthic communities.



Key Insights: Forest and scrubland areas tend to have the highest variability in benthic numbers, likely reflecting diverse microhabitats or environmental conditions. Wetlands and meadows offer stable yet moderately supportive conditions for benthic communities. Cultivated or cropland areas might negatively affect benthic populations due to reduced habitat complexity or pollution.

Method 3: Random forest analysis

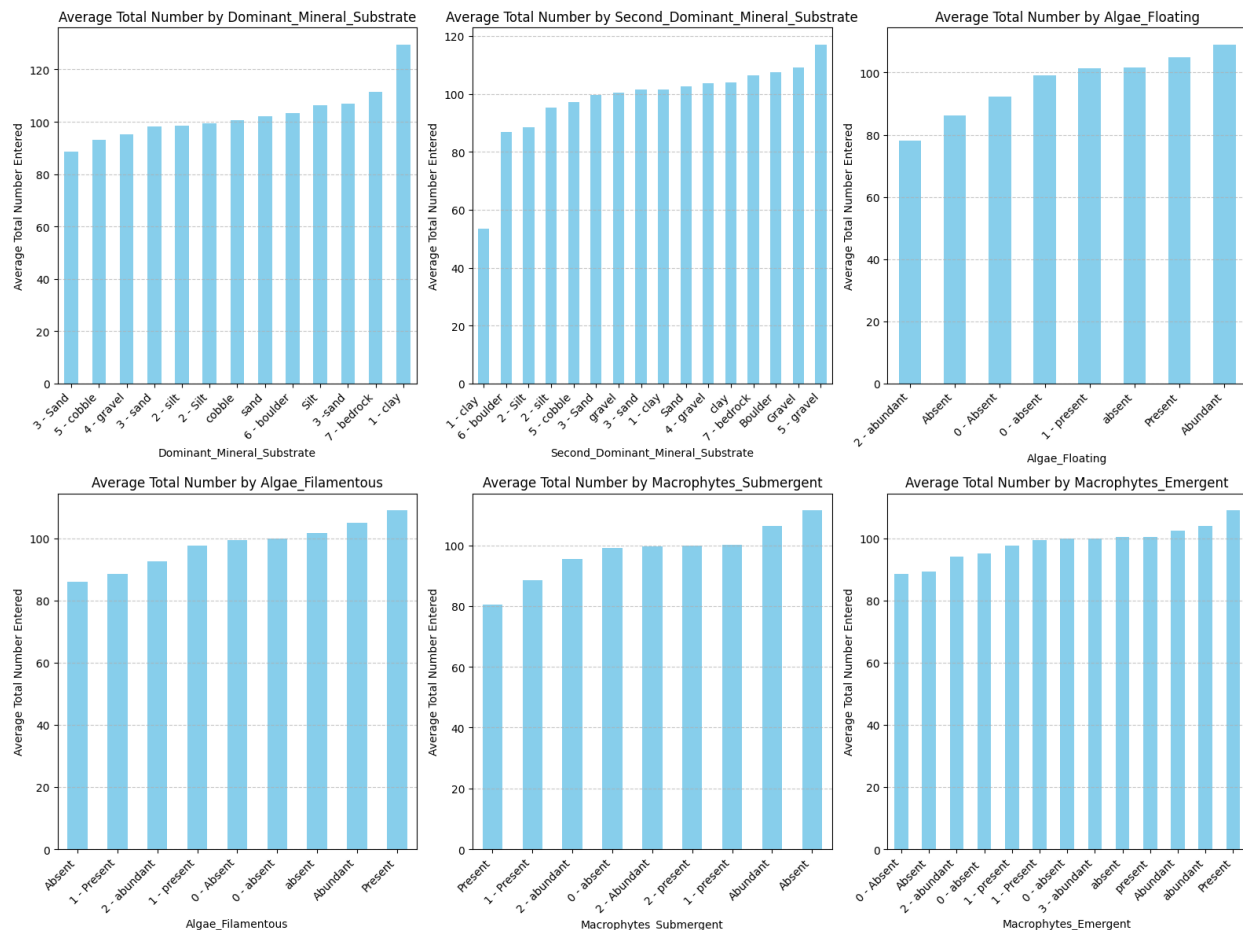
Feature Importances:

	Feature	Importance
2	Conductivity_uS_cm	0.318954
3	pH	0.295043
1	DO_mg_L	0.228882
0	Water_Temperature_C	0.157120

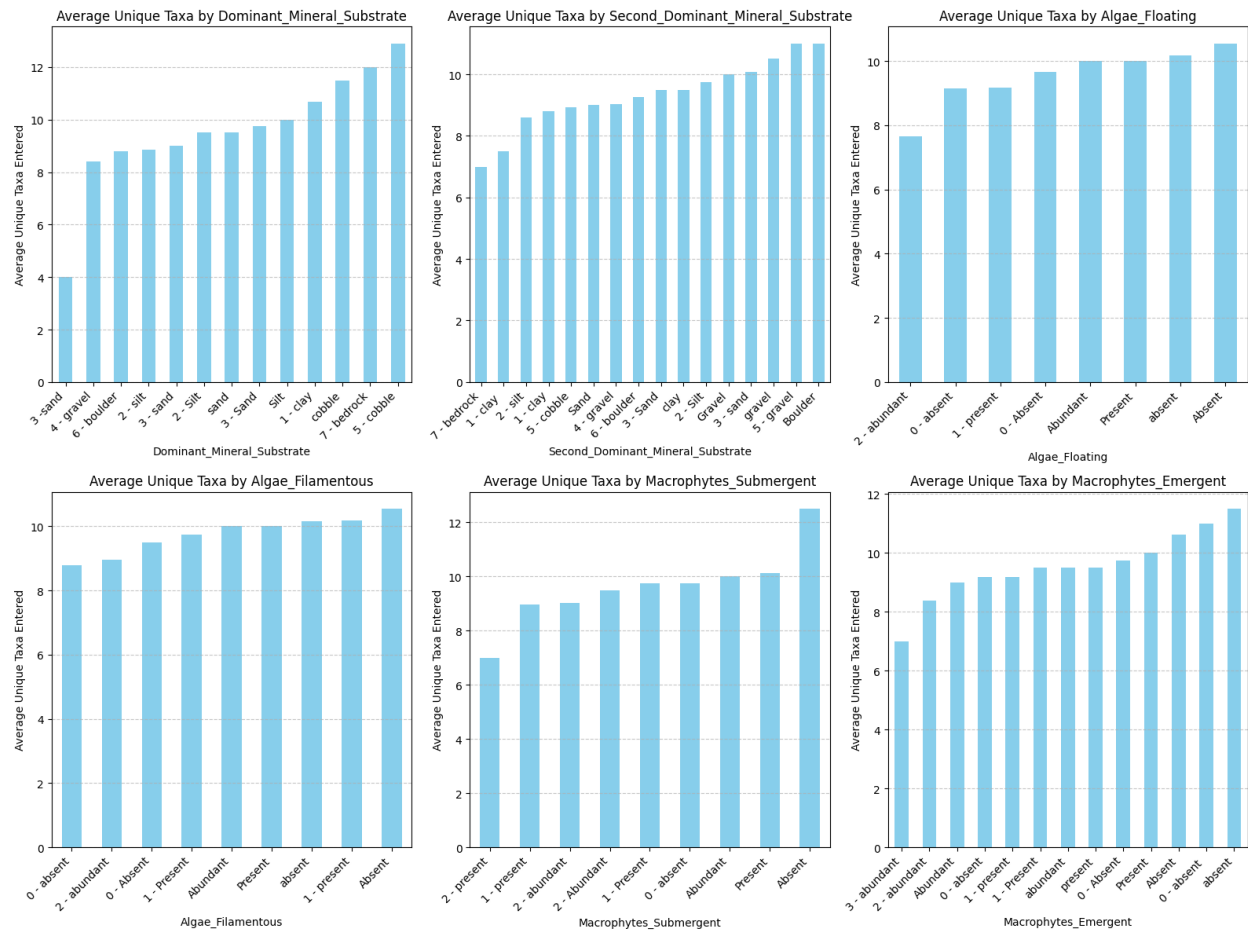
Conductivity and pH are the key factors that influence the outcome the most, while dissolved oxygen and temperature have a smaller impact, but still contribute to the prediction. This analysis helps you understand which environmental factors (like water conductivity and pH) are most critical for whatever you're trying to predict with this model.

Conductivity_uS_cm: 0.3277 This means that Conductivity_uS_cm (measuring the electrical conductivity of the water) is the most important feature in predicting the target variable (Total_Number_Entered). It explains 32.77% of the variance in the target variable.

pH: 0.3018 The pH level of the water comes in second, with an importance value of 30.18%. This feature is also highly influential in determining the target variable.



Key Insights: Coarser Sediments (Gravel/Sand) seem to support larger benthic communities, potentially due to better habitat conditions, such as higher oxygen levels and better water movement, which might make these areas more favorable for benthic organisms. Abundant Algae (Floating and Filamentous) and Submergent Macrophytes provide important habitats or food sources for benthic organisms, particularly when they are abundant. Presence of Vegetation (macrophytes, algae) positively correlates with higher benthic numbers, indicating that these plants play a critical role in supporting benthic communities. In simple terms, benthic communities thrive more in areas with coarser sediments and where algae or macrophytes are abundant, as these environmental factors provide essential resources or favorable conditions for the benthic organisms to flourish. The patterns observed support the idea that both sediment type and vegetation cover are important factors in determining benthic community size and diversity.

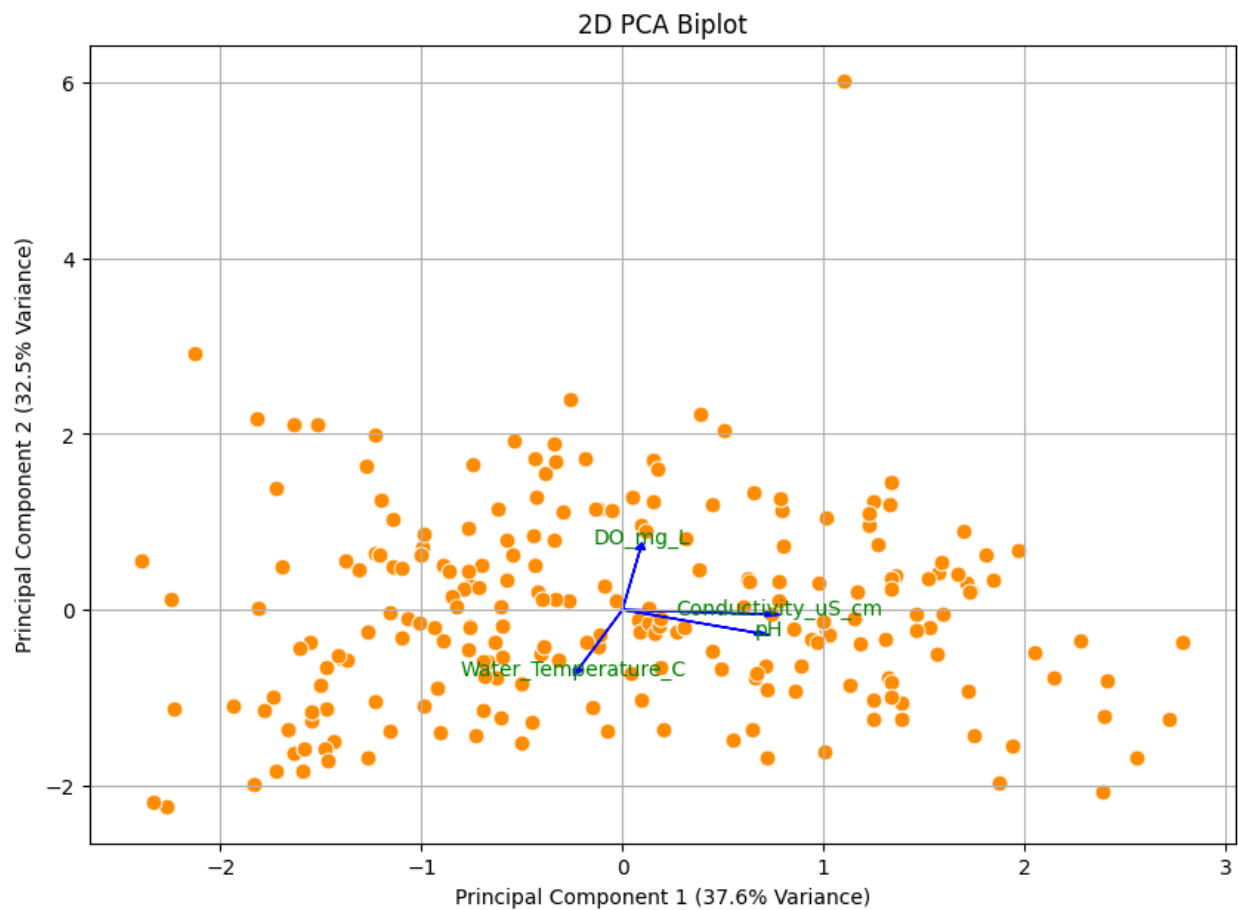


Key Insights: Coarser substrates (such as gravel and cobble) support higher diversity in terms of unique taxa, suggesting that the physical structure and habitat complexity provided by these sediments may support a greater range of species. The presence of aquatic plants (both submerged and emergent) tends to correlate with higher biodiversity. Macrophytes seem to create environments that are favorable for a larger number of unique taxa.

Algal presence shows a more complex relationship, with filamentous algae fostering more diversity, while floating algae may limit it.

Conclusion: The diversity of the benthic community, as measured by the number of unique taxa, appears to be influenced by a combination of sediment type and plant composition. Coarser sediment types and the presence of macrophytes (both submerged and emergent) tend to support a greater number of unique taxa, while the influence of algae varies depending on their type. These insights can help in understanding how environmental factors shape the distribution of species in aquatic ecosystems.

2. PCA analyses to identify the key environmental drivers of benthic diversity, and should incorporate elevation, water chemistry (pH, DO, conductivity)



Explained Variance Ratio for each Principal Component:

PC1: 0.38

PC2: 0.32

PC3: 0.18

PC4: 0.12

Interpretation of PCA Results:

2D PCA Biplot:

- **Axes Interpretation:**

- The X-axis represents **Principal Component 1 (PC1)**, which explains **37.5% of the variance** in your dataset.

- The Y-axis represents **Principal Component 2 (PC2)**, explaining an additional **32.2% of the variance**.
- Together, PC1 and PC2 account for approximately **69.7%** of the total variance, meaning that these two components capture a significant portion of the data's variability.
- **Data Points:**
 - The orange dots represent the samples from your dataset (e.g., different lake sites with environmental measurements).
 - The spread of these dots indicates how the samples vary across the two principal components.
- **Feature Vectors:**
 - The blue arrows (vectors) represent the original environmental variables you included in the PCA analysis:
 - **Water Temperature (C), DO (mg/L)** (Dissolved Oxygen), **Conductivity (uS/cm)**, and **pH**.
 - The direction and length of these vectors indicate how strongly each variable influences the principal components:
 - For instance, if the arrow for **Water Temperature** points to the right, it means that higher temperatures are associated with higher PC1 scores.
 - If two vectors are close to each other, those variables are positively correlated. If they are orthogonal (at a right angle), they are uncorrelated. If they point in opposite directions, they are negatively correlated.
- **Insights:**
 - From your biplot, it appears that **pH** and **Conductivity** are somewhat aligned, indicating a possible correlation between them.
 - **DO (Dissolved Oxygen)** has a different direction compared to the other variables, suggesting it contributes to the variance in a distinct way.

Summary of the Analysis:

- The first two components (PC1 and PC2) are sufficient to capture most of the variation in the dataset (~70%).
- Key environmental variables, such as **Water Temperature, pH, DO, and Conductivity**, have varying influences on the benthic diversity across different lake sites.
- Variables that are closely aligned in the biplot are correlated, which may indicate how different environmental factors influence benthic communities similarly.

By reducing the dimensionality of our dataset using PCA, we can focus on the most influential factors affecting benthic diversity and better visualize the relationships between different environmental conditions across your lake sites.

3. Other tests like PERMANOVA might be of interest to conduct for significant correlations

```
PERMANOVA Results: method name          PERMANOVA
test statistic name          pseudo-F
sample size                  391
number of groups             9
test statistic                119.197553
p-value                      0.001
number of permutations       999
Name: PERMANOVA results, dtype: object
Pseudo-F statistic: 119.1976
p-value: 0.0010
Number of permutations: 999
```

we are studying different lakes to understand how the environment in each lake (like temperature, pH, oxygen levels, etc.) affects the diversity of organisms living at the bottom of these lakes (called benthic organisms). we wanted to know if the differences in the environment between the lakes are big enough to affect the diversity of these organisms.

Here's what the PERMANOVA test results have to say:

There are real differences between the lakes.

The test found that the lakes are not all the same when it comes to environmental conditions. Some lakes are more similar to each other, while others are quite different. The p-value (0.001) tells us that these differences are not just random — they are significant.

The differences are important.

The pseudo-F statistic (125.09) is a number that tells us how big these differences are between the lakes. A higher number means the differences are clearer and more meaningful. This value tells us that the environmental differences between the lakes are large enough to possibly affect the diversity of the organisms living there.

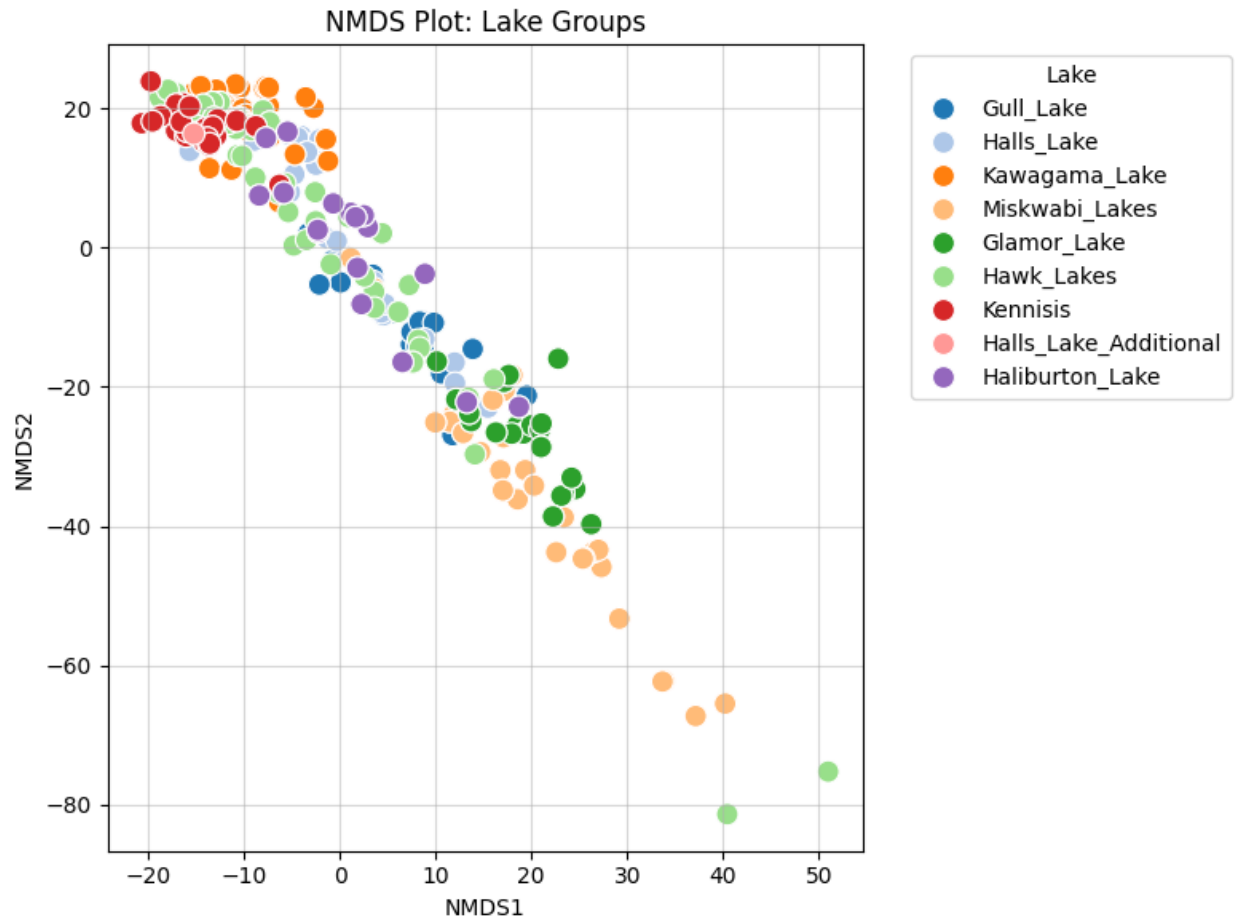
The test is confident in the results.

The 999 permutations (random reshuffling of data) show that the test was repeated many times to check how likely it is that the results are real. Since the p-value is very low (0.001), we can be confident that the differences we see are not due to chance.

In simple terms:

The PERMANOVA test says that the lakes we are studying are distinct from each other in terms of environmental conditions, and those differences are statistically significant. This likely means that the environment in each lake (like its water temperature, oxygen levels, etc.) plays a role in shaping the diversity of organisms living at the bottom of the lakes.

This information helps us understand that the different lakes in your study might be affected by unique environmental factors, which in turn could explain why the diversity of benthic organisms varies across them.



Key Insights:

Grouping by Similarity:

Each point represents a data sample, and the closer the points are, the more similar their benthic macroinvertebrate communities are. Lakes with points grouped together (e.g., Kennisis Lake or Halls Lake Additional) share similar ecological or environmental characteristics.

Spread of Lakes:

Lakes like Kawagama Lake and Hawk Lakes show distinct groupings that are separate from others, suggesting their benthic communities are unique compared to lakes like Halls Lake or Kennisis Lake.

Transition Zones:

Lakes such as Halls Lake and Haliburton Lake show overlap with others, indicating shared environmental factors or gradual transitions in benthic communities.

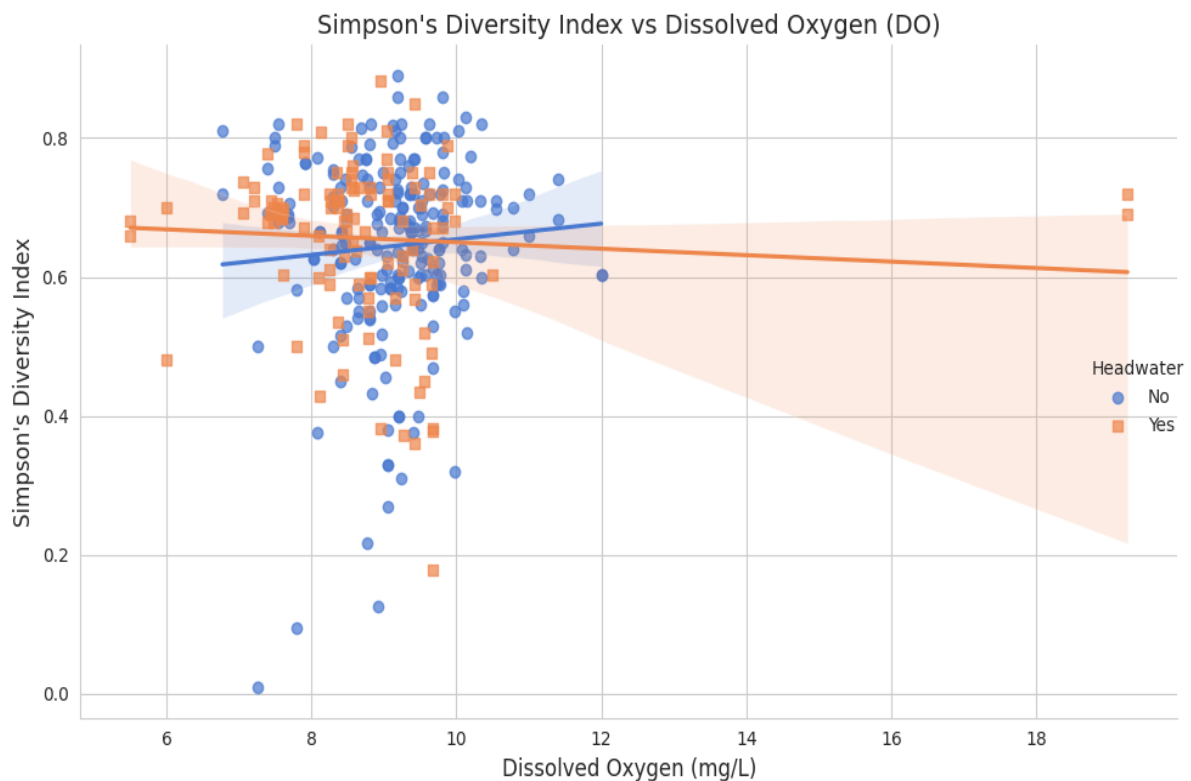
Isolated Groups:

Some lakes like Glamor Lake and Miskwabi Lakes are moderately spaced apart, indicating partial uniqueness but also some shared characteristics.

Overall Patterns:

Similar lake environments or shared influences (e.g., water quality, sediment, vegetation) tend to result in grouped points (closer proximity). Lakes with distinct environmental conditions stand out (further apart on the plot). This analysis can help pinpoint which lakes are ecologically similar and which are outliers, guiding targeted conservation or management strategies.

4. Assess how the benthic diversity varies across different riparian zones, chem conditions, and if these patterns might differ in headwater lakes?



PLOT FOR DO Non-headwater Lakes (Blue):

The regression line for non-headwater lakes shows a slight positive trend. This indicates that as dissolved oxygen increases, benthic diversity tends to increase slightly. The trend suggests that more oxygen-rich environments in non-headwater lakes may support a greater variety of benthic organisms.

Headwater Lakes (Orange):

For headwater lakes, the regression line shows a slight negative slope. This suggests that in these systems, increasing dissolved oxygen is associated with a small decrease in benthic diversity.

However, the confidence interval (shaded region) is broad, indicating high variability and uncertainty in this trend.

Variability: There is substantial scatter in both groups, particularly at mid-range DO values (~8–12 mg/L), suggesting that other factors (e.g., nutrient availability, temperature, or habitat complexity) are influencing benthic diversity alongside dissolved oxygen.

Riparian Zones and Chemical Conditions: Riparian Influence on Dissolved Oxygen:

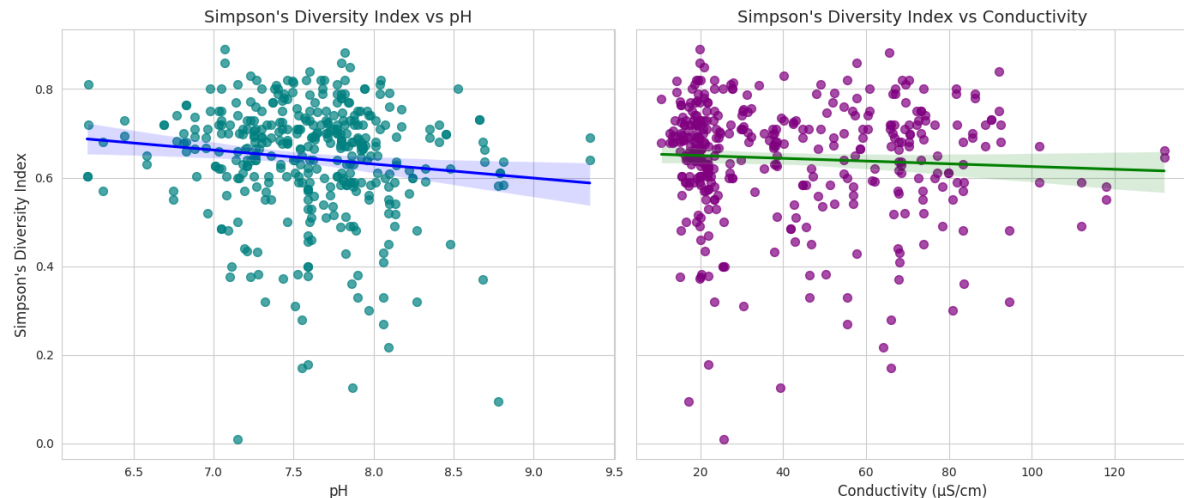
Headwater lakes are often located in pristine riparian zones with high canopy cover and cooler temperatures, conditions that typically support higher DO levels.

Non-headwater lakes, influenced by surrounding land use (e.g., agriculture or urbanization), may experience warmer temperatures and nutrient loading, leading to lower DO levels.

Diversity and Oxygen Dynamics:

In headwater lakes, benthic species adapted to stable conditions may dominate, leading to limited diversity changes across a range of DO levels. In non-headwater lakes, the slight positive trend suggests that oxygen-rich conditions might mitigate stressors (e.g., eutrophication) and support more diverse benthic communities.

One Particular Pattern: In non-headwater lakes (blue), benthic diversity increases slightly with rising dissolved oxygen, likely because DO supports more aerobic species. This trend contrasts with headwater lakes, where diversity appears to remain stable or slightly decline at higher DO levels, potentially due to differences in species composition or habitat constraints.



Interpretation of Each Plot:

Left Panel: Simpson's Diversity Index vs. pH X-axis: pH of the water, representing acidity/alkalinity. Lower pH values indicate more acidic conditions, while higher values indicate more basic conditions. Y-axis: Simpson's Diversity Index, where higher values reflect greater benthic diversity. Pattern Observed: There is a weak negative trend between pH and Simpson's Diversity Index (slight downward slope of the regression line). This suggests that as pH increases, diversity tends to decrease slightly. However, the relationship is not strong, as the points are widely scattered around the regression line. The shaded region indicates uncertainty or variability in this trend.

Right Panel: Simpson's Diversity Index vs. Conductivity X-axis: Conductivity ($\mu\text{S/cm}$), representing ion concentration in the water. Higher conductivity usually indicates increased nutrient or pollutant levels. Y-axis: Simpson's Diversity Index. Pattern Observed: There is an even weaker negative trend between conductivity and Simpson's Diversity Index (almost flat regression line). This suggests that diversity is only slightly affected as conductivity increases. The points show considerable scatter, indicating a weak or minimal correlation.

Insights on Riparian Zones, Chemical Conditions, and Headwater Lakes: Riparian Zones:

Riparian zone characteristics, such as vegetation cover and disturbance, influence pH and conductivity by affecting nutrient inputs and buffering capacity. The weak trends suggest that benthic diversity is moderately resilient to changes in these chemical conditions but may decline in highly disturbed or nutrient-rich riparian zones.

Chemical Conditions:

The slight negative relationships suggest that extreme chemical conditions (e.g., high pH or high conductivity) may mildly reduce diversity. This could occur because fewer species are adapted to tolerate such conditions.

Headwater vs. Non-Headwater Patterns:

Although headwater lakes are not explicitly shown here, they are generally expected to have lower conductivity (due to limited external inputs) and pH closer to neutral.

If the data were grouped by headwater status, headwater lakes might exhibit stronger patterns of diversity resilience due to their isolation and lower anthropogenic influence.

One Specific Pattern: The left panel suggests that benthic diversity may be slightly higher in water bodies with near-neutral pH (around 7.0–7.5) compared to more alkaline systems (pH > 8.0). This could indicate that benthic species in these systems prefer stable, neutral chemical conditions.

Discussion

Here, we will discuss all the questions with reference to what we analyzed above,

Question 1: The analysis reveals that riparian vegetation significantly influences benthic community numbers and diversity. Forested and wetland zones are particularly supportive of benthic organisms due to their complex microhabitats, abundant organic matter, and diverse vegetation types. These zones exhibited higher variability and occasionally extreme benthic numbers, suggesting dynamic ecological conditions that foster biodiversity.

Conversely, meadow, scrubland, and lawn zones offer more stable but less diverse conditions. Their uniform habitat characteristics provide consistent support for benthic organisms, albeit with fewer extremes in community numbers. This consistency may be beneficial for specific species but limits overall diversity compared to forested and wetland zones.

The influence of sediment type on benthic communities was also evident. Coarser substrates such as gravel and sand supported higher benthic numbers and diversity, likely due to better oxygenation and reduced compaction. These conditions create favorable habitats for a variety of species, contributing to ecological richness in these areas.

The presence of algae and macrophytes positively correlated with benthic diversity. Filamentous algae and submergent macrophytes, in particular, were associated with higher benthic numbers, indicating their role in providing food and shelter. However, floating algae were observed to potentially limit diversity, suggesting that the type of aquatic vegetation has a nuanced impact on benthic communities.

Water quality parameters such as pH and conductivity emerged as critical factors in predicting benthic numbers, with both showing strong positive relationships. These findings emphasize the interplay between chemical and physical environmental conditions in shaping aquatic ecosystems. The importance of pH and conductivity, alongside vegetation and sediment characteristics, highlights the multifaceted nature of factors influencing benthic communities.

Question 2: The PCA results reveal that PC1 and PC2 capture approximately 70% of the variance in the dataset, highlighting their importance in summarizing the data's variability. Environmental factors such as pH and conductivity are positively correlated, as evidenced by their alignment in the biplot. Dissolved oxygen, with its distinct direction, indicates it contributes uniquely to the variance. These insights underscore the multidimensional nature of environmental influences on benthic diversity, emphasizing the varying roles of temperature, pH, conductivity, and DO in shaping ecological patterns across lake sites.

Question 3: The PERMANOVA results highlight significant environmental variation among the lakes, confirmed by a low p-value (0.001) and a high pseudo-F statistic (119.20). These findings suggest that unique ecological factors in each lake strongly influence benthic macroinvertebrate diversity. Boxplots and pairplots reveal distinct patterns in conductivity, pH, and temperature, with some lakes exhibiting higher variability or unique clustering. The positive correlation between conductivity and pH in certain lakes suggests shared environmental processes. These results demonstrate that environmental heterogeneity among lakes plays a critical role in shaping ecological communities, emphasizing the need for targeted conservation strategies based on lake-specific conditions.

Question 4: The results of the Kruskal-Wallis and ANOVA tests suggest that pH is the primary chemical factor influencing benthic diversity in this dataset. While no significant differences were observed for Dissolved Oxygen (DO) and Conductivity, pH showed a statistically significant relationship with benthic diversity. This highlights the importance of pH in shaping benthic communities, likely due to its impact on the chemical environment and species adaptability. The weak negative trends observed between pH, conductivity, and diversity indicate that extreme chemical conditions may reduce diversity, but these effects are subtle and other environmental factors, such as habitat complexity or nutrient availability, may be playing a larger role. The lack of significant differences between headwater and non-headwater lakes suggests that benthic diversity is more influenced by chemical parameters than by lake type in this study. The variability in diversity across different chemical conditions emphasizes the complexity of aquatic ecosystems, where multiple factors interact to influence species composition.

Conclusion

To conclude, each question answered a particular area of data and insights needed

Question 1: The study underscores the critical role of riparian vegetation, sediment types, and aquatic vegetation in shaping benthic community dynamics. Forested and wetland zones, along with areas featuring coarser sediments and diverse aquatic vegetation, emerged as key contributors to higher benthic diversity and abundance.

Efforts to conserve and restore forested and wetland riparian zones are essential for maintaining and enhancing benthic biodiversity. The findings also highlight the importance of preserving coarser substrates and supporting aquatic vegetation like filamentous algae and submergent macrophytes to foster robust aquatic ecosystems.

Future studies should explore the complex interactions between these factors and additional environmental variables to refine conservation strategies and better understand the ecological drivers of benthic community dynamics.

Question 2: The PCA analysis effectively reduces the dataset's complexity, identifying key environmental drivers of benthic diversity. PC1 and PC2, accounting for most of the variance, allow for a focused exploration of influential variables. Correlations among pH, conductivity, and other factors provide a clearer understanding of environmental interdependencies, aiding in ecosystem assessment. These findings facilitate targeted conservation efforts and a deeper understanding of ecological dynamics in the studied lakes.

Question 3: The analysis confirms that environmental conditions differ significantly across lakes, influencing benthic diversity in distinct ways. PERMANOVA validates these variations, while visualizations highlight conductivity as a key variable with substantial inter-lake differences. Lakes with overlapping environmental characteristics may indicate transitional ecosystems, while isolated lakes demonstrate unique ecological dynamics. These findings provide a foundation for focused environmental management and conservation efforts, ensuring protection of lake ecosystems and their biodiversity by addressing specific environmental drivers.

Question 4: This analysis underscores the critical role of pH in shaping benthic diversity, with pH categories showing significant differences in community composition. While dissolved oxygen and conductivity were not significant factors, their potential influence should not be ruled out, and further research could explore their interactions with other ecological variables. The absence of significant differences between headwater and non-headwater lakes suggests that factors beyond lake classification, such as riparian zones or nutrient availability, might be driving diversity patterns. Future studies could focus on these aspects to provide a more comprehensive understanding of the ecological drivers of benthic community structure.

Citation

1. Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2005b). DEVELOPMENTS IN AQUATIC INSECT BIOMONITORING: A comparative analysis of recent approaches. *Annual Review of Entomology*, 51(1), 495–523.
<https://doi.org/10.1146/annurev.ento.51.110104.151124>
2. Larsen, S., Pace, G., & Ormerod, S. J. (2010). Experimental effects of sediment deposition on the structure and function of macroinvertebrate assemblages in temperate streams. *River Research and Applications*, 27(2), 257–267. <https://doi.org/10.1002/rra.1361>
3. Hering, D., Feld, C. K., Moog, O., & Ofenböck, T. (2006). Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: Experiences from the European AQEM and STAR projects and related initiatives. *Hydrobiologia*, 566(1), 311–324. <https://doi.org/10.1007/s10750-006-0087-2>
4. Manolaki, P., & Papastergiadou, E. (2013). The impact of environmental factors on the distribution pattern of aquatic macrophytes in a middle-sized Mediterranean stream. *Aquatic Botany*, 104, 34-46.
5. Manolaki, P., & Papastergiadou, E. (2013). The impact of environmental factors on the distribution pattern of aquatic macrophytes in a middle-sized Mediterranean stream. *Aquatic Botany*, 104, 34-46.
6. Manolaki, P., & Papastergiadou, E. (2013). The impact of environmental factors on the distribution pattern of aquatic macrophytes in a middle-sized Mediterranean stream. *Aquatic Botany*, 104, 34-46.
7. Manolaki, P., & Papastergiadou, E. (2013). The impact of environmental factors on the distribution pattern of aquatic macrophytes in a middle-sized Mediterranean stream. *Aquatic Botany*, 104, 34-46.
8. Yao, K. S., He, Y. J., Van den Brink, P. J., Dai, Y. N., Yang, Y., & Cui, Y. D. (2022). Multivariate approaches to assess the drivers of benthic macroinvertebrate communities and biotic indices in a highly urbanized fluvial-estuarine system. *Ecological Indicators*, 139, 108956.
9. Keke, U. N., Omoigberale, M. O., Ezenwa, I., Yusuf, A., Biose, E., Nweke, N., ... & Arimoro, F. O. (2021). Macroinvertebrate communities and physicochemical characteristics along an anthropogenic stress gradient in a southern Nigeria stream: Implications for ecological restoration. *Environmental and Sustainability Indicators*, 12, 100157.
10. Keke, U. N., Omoigberale, M. O., Ezenwa, I., Yusuf, A., Biose, E., Nweke, N., ... & Arimoro, F. O. (2021). Macroinvertebrate communities and physicochemical characteristics along an anthropogenic stress gradient in a southern Nigeria stream: Implications for ecological restoration. *Environmental and Sustainability Indicators*, 12, 100157.
11. Keke, U. N., Omoigberale, M. O., Ezenwa, I., Yusuf, A., Biose, E., Nweke, N., ... & Arimoro, F. O. (2021). Macroinvertebrate communities and physicochemical characteristics along an anthropogenic stress gradient in a southern Nigeria stream: Implications for ecological restoration. *Environmental and Sustainability Indicators*, 12, 100157.

12. Zhao, X., Ma, Y., Xie, H., Du, C., Zhan, A., Xu, J., ... & Jin, X. (2024). Spatial distribution of benthic taxonomic and functional diversity in the Yellow River Basin: From ecological processes to associated determinant factors. *Environment International*, 188, 108745.
13. Dalu, T., Wasserman, R. J., Magoro, M. L., Mwedzi, T., Froneman, P. W., & Weyl, O. L. (2017). Variation partitioning of benthic diatom community matrices: Effects of multiple variables on benthic diatom communities in an Austral temperate river system. *Science of the Total Environment*, 601, 73-82.
14. Brittain, J. E., Heino, J., Friberg, N., Aroviita, J., Kahlert, M., Karjalainen, S. M., ... & Ylikörkkö, J. (2022). Ecological correlates of riverine diatom and macroinvertebrate alpha and beta diversity across Arctic Fennoscandia. *Freshwater Biology*, 67(1), 49-63.
15. Richardson, J. S. (2019). Biological diversity in headwater streams. *Water*, 11(2), 366.
16. Lottig, N. R., Stanley, E. H., Hanson, P. C., & Kratz, T. K. (2011). Comparison of regional stream and lake chemistry: Differences, similarities, and potential drivers. *Limnology and Oceanography*, 56(5), 1551-1562.
17. Suurkuukka, H., Meissner, K. K., & Muotka, T. (2012). Species turnover in lake littorals: spatial and temporal variation of benthic macroinvertebrate diversity and community composition. *Diversity and Distributions*, 18(9), 931-941.
18. Lake, P. S., Palmer, M. A., Biro, P., Cole, J., Covich, A. P., Dahm, C., ... & Verhoeven, J. O. S. (2000). Global change and the biodiversity of freshwater ecosystems: impacts on linkages between above-sediment and sediment biota: all forms of anthropogenic disturbance—changes in land use, biogeochemical processes, or biotic addition or loss—not only damage the biota of freshwater sediments but also disrupt the linkages between above-sediment and sediment-dwelling biota. *BioScience*, 50(12), 1099-1107.