Question 1

A difference in Phosphate levels was found. To begin with, our data was not normal as shown by the **Shapiro-Wilk normality test,** then the bestNormalise() function was used from the bestNormalize library to transform the data and make it normal. Another **Shapiro-Wilk normality test** was performed to test if we were able to transform the data into normality or not and it did pass the test. Then a histogram was drawn for a general overview of the data. Then a t-test was performed with the normalized phosphate levels as normality is one of the assumptions for a t-test. The t-test did find a significant difference in phosphate levels between the two Ecozones with a p level of < 0.001 which means there is a difference in phosphate levels between the two different ecosystems. We chose the t-test because there were only 2 types of Ecozones, and they were independent.

Question 2

Here we are talking about two different variables and their phosphate levels so we have to do a multi-way ANOVA test whose two assumptions are homo-variance and normality. **Shapiro-Wilk normality test** revealed that the phosphate levels were normal. **Levene's Test** further revealed that our data was normal. We would have to further conduct an ANOVA test for each fishery and Ecozone and their influences which needs contrast code which was not in the scope of the class hence we cannot know about the interactions.

However just to know about phosphate levels for fishery we can still conduct a parametric one-way ANOVA test. The result from that ANOVA test shows that there is a significant difference in phosphate levels of different fisheries with a $p<0.001$.
With that being said, our ANOVA does not provide details about which fishery group has the different phosphate levels for that we would need a post hoc test. We chose TukeyHSD for post hoc as we are performing a parametric ANOVA test.
The test revealed that the difference was originating from the groups- (Walleye-Lake Trout, Walleye) and (Walleye-Brook Trout, Lake Trout). Graphs were drawn to visualize the data.

Question 3

First, we made a scatterplot in which all the data seemed to accumulate on the left side. An attempt to normalize species was made which was unsuccessful. A correlation test was done and a statistically significant correlation was not found.
Then linear model function was applied which resulted in a statistically significant intercept of 11.968 which means if phosphate levels are 0 our species richness would be 11.97 but the slope between these two variables was found to be statistically insignificant.

To calculate species richness for 18.2 phosphate levels we will use $0*18.2+11.968 = 11.968$ as our slope is insignificant.

Our prediction is likely to be very unreliable as –

- **Normality-** Species are not normal so we used non-normal phosphate levels as well, so both of our variables are not normal.

- **Correlation**- was found to be statistically not significant.
- The slope was found to be insignificant.
- **Regression graph** did not show strong visual evidence of correlation and all the points were somewhat clustered on the left. The data did not look distributed over a straight line.

-Residual Normality
Residuals are largely normal except residuals from Walleye and Lake Trout according to the **Shapiro-Wilk normality test.**

PART 2

Table 1-Key descriptive data relating to the spread and normality of the data.

Variable	Mean	Median	Min	Max	Skew	Variance	P value for normality
Phosphate levels	10.3	8.5	0	101.5	4.83	52.84	<0.001
pH	7.18	7.16	5.48	8.64	0.029	0.23	0.059
Calc	9.69	6.12	1.02	77.40	2.35	104.37	<0.001
Species	12.74	11	2	80	2.39	68.97	<0.001
Residual(Phosphate-micrograms per litre)	3.487326e-16	-1.58	-11.06	66.43	2.34	68.66	<0.001

1)
Method
A Welch two-sample t-test was used to compare the Phosphate levels for two different Ecozones, Mixedwood Plains and Ontario Shield. A **Shapiro-Wilk normality test** indicated that the data does not come from a normal distribution. **OrderNorm (ORQ)** transformation was performed to make the data normal. **Shapiro-Wilk normality test** on the transformed data showed that it comes from a normal distribution (Mixedwood Plains W=0.98, p=0.68; Ontario Shield W=0.99, p=0.99). Both the Ecozones are unrelated, as such independent t-test was chosen. Alpha for this analysis was 0.05. The statistical software used was R version 4.3.2.

Result

The Welch Two Sample t-test revealed a significant difference in Phosphate contents among the two ecozones (t=4.6558; df=58.189; p<0.001)(Figures 1 and 2). Henceforth we reject the null hypothesis of no difference between phosphate levels in Mixedwood Plains (mean=0.62) and Ontario Shield (-0.049).

Table 2 Results from Welch Two Sample t-test comparing Phosphate(micrograms per litre) levels in two different Ecozones- Mixedwood Plains and Ontario Shield.

Parameter	Value
t	4.65
df	58.189
p-value	<0.001
Mean in Mixedwood Plains	0.62
Mean in Ontario Shield	-0.049

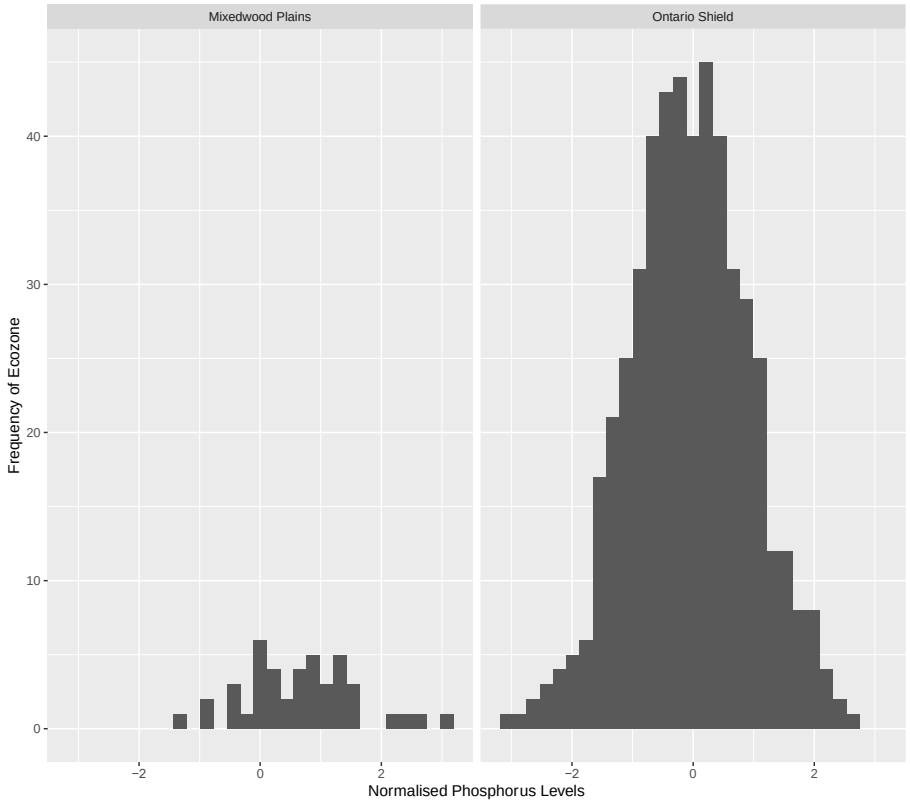

Figure 1 Histogram visualizing frequency of different Ecozones for different levels of Phosphorus.

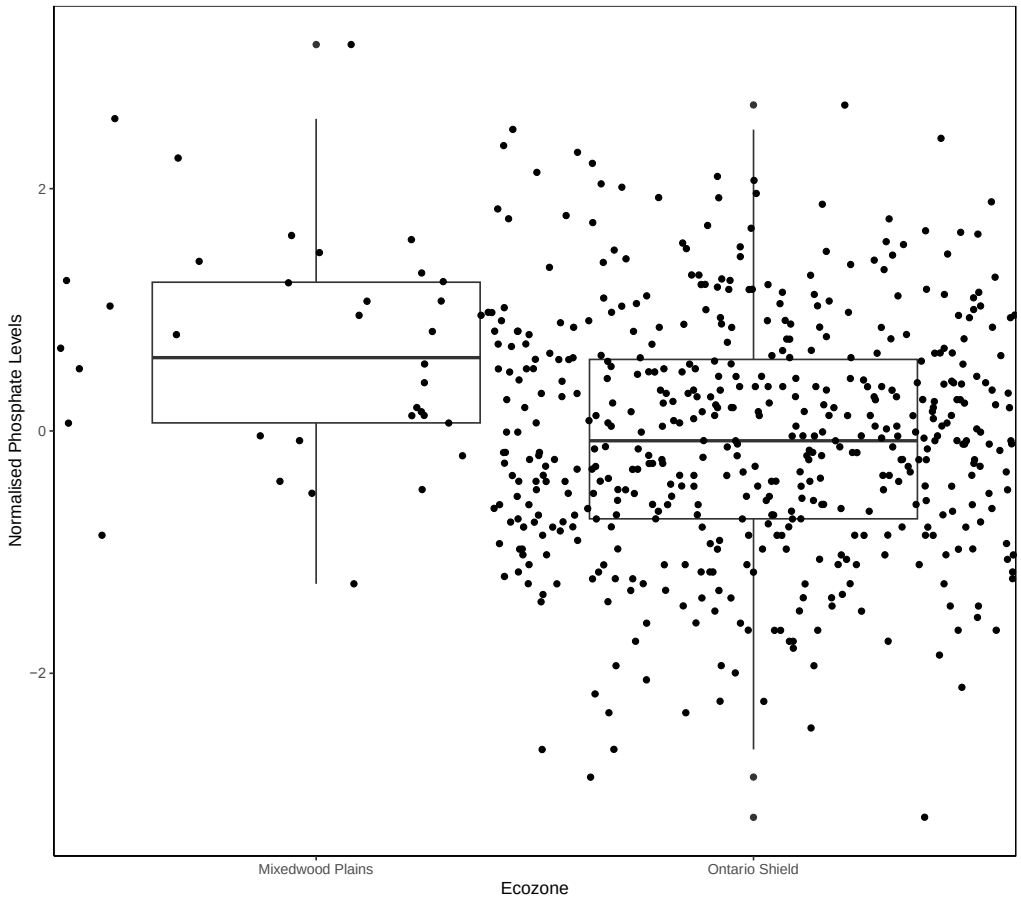

Figure 2 Boxplot visualizing the distribution of Phosphate levels across different Ecozones and their variances.

2)

Method

A one-way ANOVA was used to compare the Phosphate levels across among different fisheries Shapiro-Wilk's test indicated that the data does not come from a normal distribution. **OrderNorm (ORQ)** transformation was performed to make the data normal. **Shapiro-Wilk normality test** on the transformed data showed that it comes from a normal distribution (Brook Trout W=0.99, p=0.82; Brook Trout, Lake Trout W=0.89, p=0.36; Lake trout W=0.99, p=0.67; Lake Trout, Walleye W=0.95, p=0.47; Walleye W=0.89, p=0.89). Levene's test revealed that variances were equal among the groups (F=0.90, d.f=6,573, p=0.49). As such, a parametric ANOVA test was chosen. Alpha for the analysis was 0.05. The statistical software used was R version 4.3.2.

Result

The ANOVA revealed a significant difference in Phosphate levels among different fisheries (F=56.84; d.f= 4,575; p<0.001) (Figures 3 and 4). A Tukey test showed the source of the difference was the mean Phosphate content of lakes with fisheries Walleye (mean= 0.42, SE=0.047) and Brook Trout, Lake Trout (mean= -0.85, SE=0.28); and Walleye (mean= 0.42,

SE=0.047) and Lake Trout, Walleye (mean= -0.42, SE=0.17). The other groups were not different.

Table 3 Results from one-way ANOVA test comparing Phosphate levels (micrograms per litre) favoured by each fishery.

Effect	Degrees of Freedom	Sum of squares	Mean Squares	F-Statistic	Probability
Fishery	4	165.6	41.39	56.84	<0.001
Error	575	418.7	0.73		

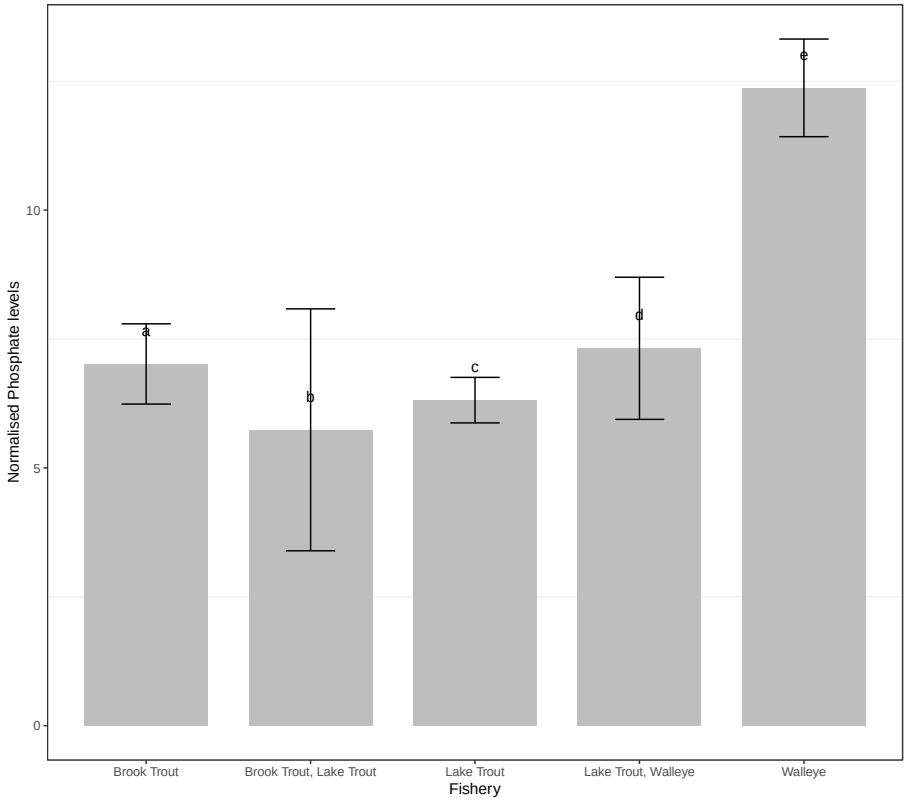

Figure 3 Mean (+-95% confidence intervals) Phosphorus by each Fishery in the dataset. Letters above the bars denote statistically similar groups.

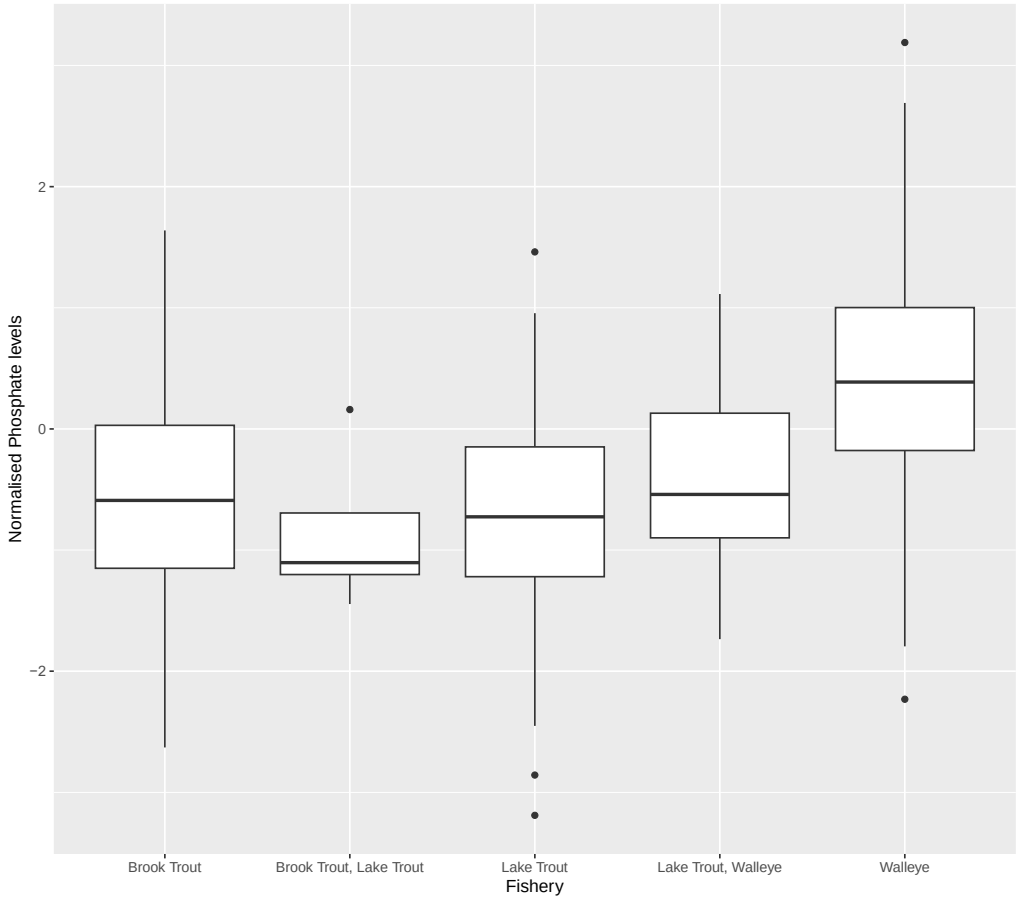

Figure 4 Variations in Normalised Phosphorus levels across different Fishery types.

3)

Method

A Shapiro Wilk normality test was conducted which showed that Species considered significantly from the normal distribution and attempts to transform the data failed. Henceforth non-transformed data was used for Phosphate data and Species. Pearson's product-moment correlation test was used to assess the correlation and the linear model was used to assess residuals and the significance of the intercept and slope of the two variables. A scatterplot was used to visually assess the data.

Result

Pearson's product-moment correlation test indicated that there was no statistically significant correlation between Phosphate levels in lakes and the number of species. The linear model showed a non-significant slope but a significant intercept of 11.97 ($p<0.001$). Residuals from only Brook Trout ($W=0.96$, $p=0.06$) , Brook trout, Lake trout($W=0.94$, $p=0.70$) and Lake trout, Walleye ($W=0.94$, $p=0.32$) showed normal distribution. The scatterplot

(Figure 5)showed no obvious correlation, and the points were not along the line of best fit but clustered in the bottom left.

Table 4 Key parameters from model predicting species richness and Phosphate levels (micrograms per Litre) in lakes in Ontario.

Parameter	Value	Probability
Median residual	-1.580	
First quartile of residuals	-5.31	
Third quartile of residuals	3.054	
Minimum residual	-11.061	
Maximum residual	66.436	
Y intercept	11.97	
Slope	0.077	0.090
Adjusted R^2	0.0029	
F-statistic	2.88	0.090

*** p < .001 ** p < .01 * p < .05

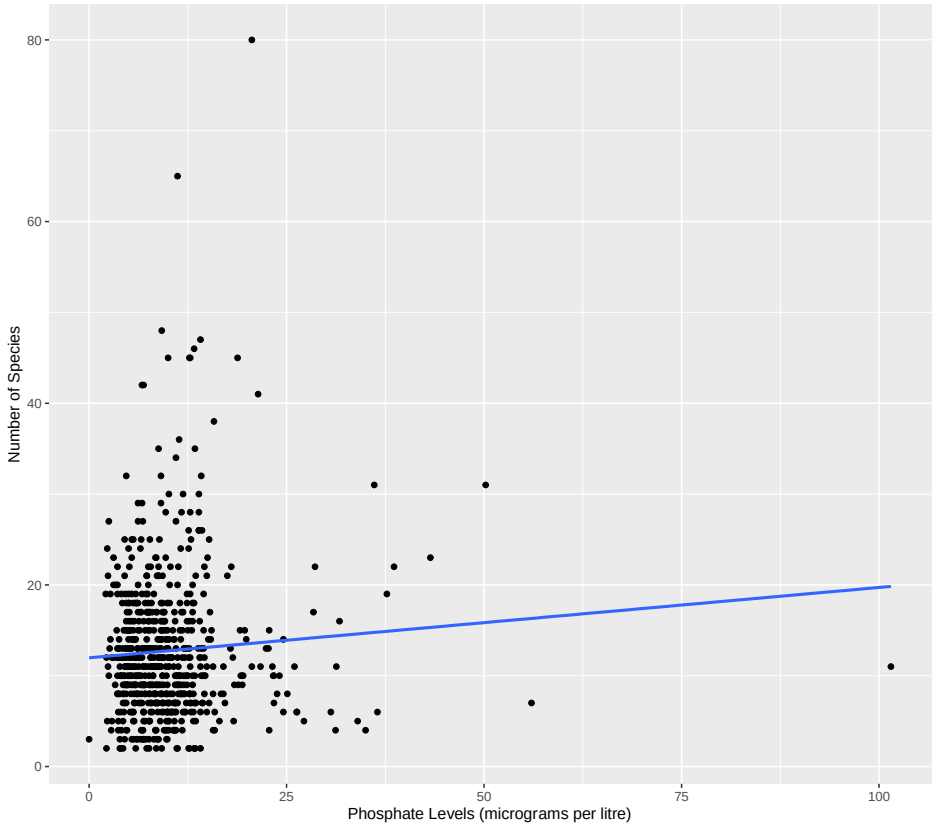

Figure 5 Scatterplot visualizing the relationship between species richness and Phosphate levels.

Conclusion-

The Findings of significant differences in phosphate levels across ecozones and fisheries suggest that spatially determined chemical factors significantly influence aquatic biodiversity. The appropriateness of statistical methods post normalizations lends credibility to the t-test and ANOVA results. However, the limitations are noteworthy, the normalization of data could potentially obscure ecological nuances and the lack of correlation between phosphate levels and species richness indicates other ecological variables may be influencing biodiversity.