ENV 710 Final Project

Emma Kaufman, Cara Kuuskvere, Jenn McNeill

2024-04-10

Contents

Introduction

In this report we examine the relationship of total phosphorus concentration on dissolved oxygen levels in Wisconsin lakes. Adequate levels of dissolved oxygen in lakes are crucial for the aquatic species that reside there. DO levels are at risk for becoming dangerously low when excess organic materials, such as algae, decompose, a process called eutrophication (US EPA, 2013). As a result, it is important to limit algae growth in lakes. When there is too much phosphorus in lakes conditions become suitable for algae blooms resulting in eutrophication. In addition, toxins found in algae blooms are harmful to human and animal health (US EPA, 2013).

Total phosphorus in lakes is influenced by rainfall, stream flow, overland runoff, groundwater discharge, decomposition of organic matter, and soil erosion. Humans contribute to phosphorus levels in lakes due to fertilization of agricultural fields and lawns, leaking waste water, and increased urbanization (Lee, 1973). Most phosphorus in lakes takes the form of dissolved or particulate phosphorus, and total phosphorus is a measure of all forms of phosphorus potentially available to algae (Total Nitrogen/Total Phosphorus and Chlorophyll a, n.d.).

Our research questions and associated hypotheses are three-fold:

- Is there a difference in the impact of total phosphorus on dissolved oxygen levels in the spring versus the summer?
  - H_0: There is no difference in the impact of phosphorus on DO in the spring versus the summer.
  - H_a: There is a difference in the impact of phosphorus on DO in spring versus summer.
- Is there a difference in the impact of total phosphorus on dissolved oxygen levels in shallow versus subsurface lake water?
  - H_0: There is no difference in the impact of phosphorus on DO in surface and subsurface lake water.
  - H_a: There is a difference in the impact of phosphorus on DO in surface and subsurface lake water.
- Do West Long, East Long, Peter, and Paul Lake all exhibit the same relationship of how total phosphorus impacts dissolved oxygen levels?
  - H_0: All of these lakes display the same relationship of how total phosphorus impacts dissolved oxygen levels.
  - H_a: The impact of phosphorus on DO is different across these lakes.

Methods

The data for this project are from lakes in the North Temperate Lakes District of Wisconsin. They were collected at the Long Term Ecological Research station located there. The physical and chemical variables were measured at a central station in the deepest point of each lake. Most measurements were made between 8 and 9 am. The nutrient measurements were made along vertical profiles at varied depth intervals. The dissolved oxygen chemistry data from 1991-1999 were obtained from a Lachat auto-analyser. The methods of data collection from 1991-1997 are described in more detail by Carpenter at al. (2001). More information on this data can be found here: https://lter.limnology.wisc.edu/core-datasets/.

These datasets contain daily nutrient and lake chemistry data at 26 different lakes within the North Temperate Lakes District in Wisconsin over 25 years. The variables between the two datasets include the lake depth, total nitrogen, total phosphorus, nh34, no23, and po4 concentrations, as well as lake temperature, dissolved oxygen, and irradiance. For the purposes of this project we focus on using lake depth, total phosphorus, and dissolved oxygen data across 4 lakes: East Long Lake, Paul Lake, Peter Lake, and West Long Lake from 1994 to 1999. A table of these variables is found below:

Variable Analyzed	Unit of Measure	Type of Variable
Lake Name	Name Factor	Character/ Factor
Depth	Meters below surface	Numeric ; binary indicator
Season	Spring (May)/ Summer (August)	Binary indicator
Dissolved Oxygen (DO)	mg/ Liter	Numeric
Phosphorus	μg	Numeric

Our dependent variable is dissolved oxygen (mg/L) and our independent variables are total phosphorus (μg), season (May or August, coded as spring or summer), and depth (<6 ft as surface, deeper than 6 ft as subsurface).

In this report we use linear models to estimate dissolved oxygen across these lakes based on total phosphorus content sampled within the lakes' water columns. We examine the difference in the relationship of DO and total phosphorus in spring and summer, as well as the difference in the relationship of DO and total phosphorus at the water surface versus deeper in the water column. Finally, we use a hierarchical linear model to examine the DO and total phosphorus relationship across East Long, West Long, Peter, and Paul lakes.

In order to check the assumptions of these models, we plot the residuals against a normal distribution (qq plot) as well as the residuals vs fitted values to check for constant variance. We also fit training models with 80% of the data and checked our mean absolute percentage error (MAPE) values to see the accuracy of our model fit versus our observed data.

Results

MODEL FIT ONE: Dissolved Oxygen (DO) by Total Phosphorus (TP) with a Seasonal Binary Indicator Variable and a Seasonal Interaction Term

Variable	Value	Interpretation
β_0 (intercept)	10.15 mg/L	Expected value of DO in spring with 0 TP
β_1	-2.36 mg/L	Difference in DO between spring and summer with 0 TP
β_2	-0.067 (mg/L)/μg	Expected change in D0 for each unit increase in TP in the spring
β_3	0.02 (mg/L)/μg	Adjustment to slope for interaction for each unit increase in TP during the summer

Results Model One

The results of this model are that the expected value of dissolved oxygen across all of the lakes with no phosphorus is 10.15 mg/L in the spring. In the summer, dissolved oxygen levels when there is no phosphorus decrease by 2.36 mg/L.

The expected change in DO for each unit increase in total phosphorus during the spring is -0.067 (mg/L)/μg . In the summer this change increases by 0.02.

These model results are displayed with the observed data in Figure 1. This linear model assumes that the residuals are normally distributed and have constant variance. The plots that examine these assumptions are found in the appendix plot 1. They show that these assumptions were not met, as the residuals do not follow a normal distribution (as seen in the qq plot), and the variance is not constant (in the residuals v fitted plot).

We tried to improve our model by normalizing the data with a log transformation of total phosphorus. The results of this are in the appendix (Appendix plot log-transform). We still did not see normal distribution of residuals or constant variance of residuals. Additionally, we tried to use a subset of the DO data and fit our model to those data (Appendix plot subset DO), but still did not have promising results for our model assumptions. We chose to go forward with the original data because it is the most interpretable.

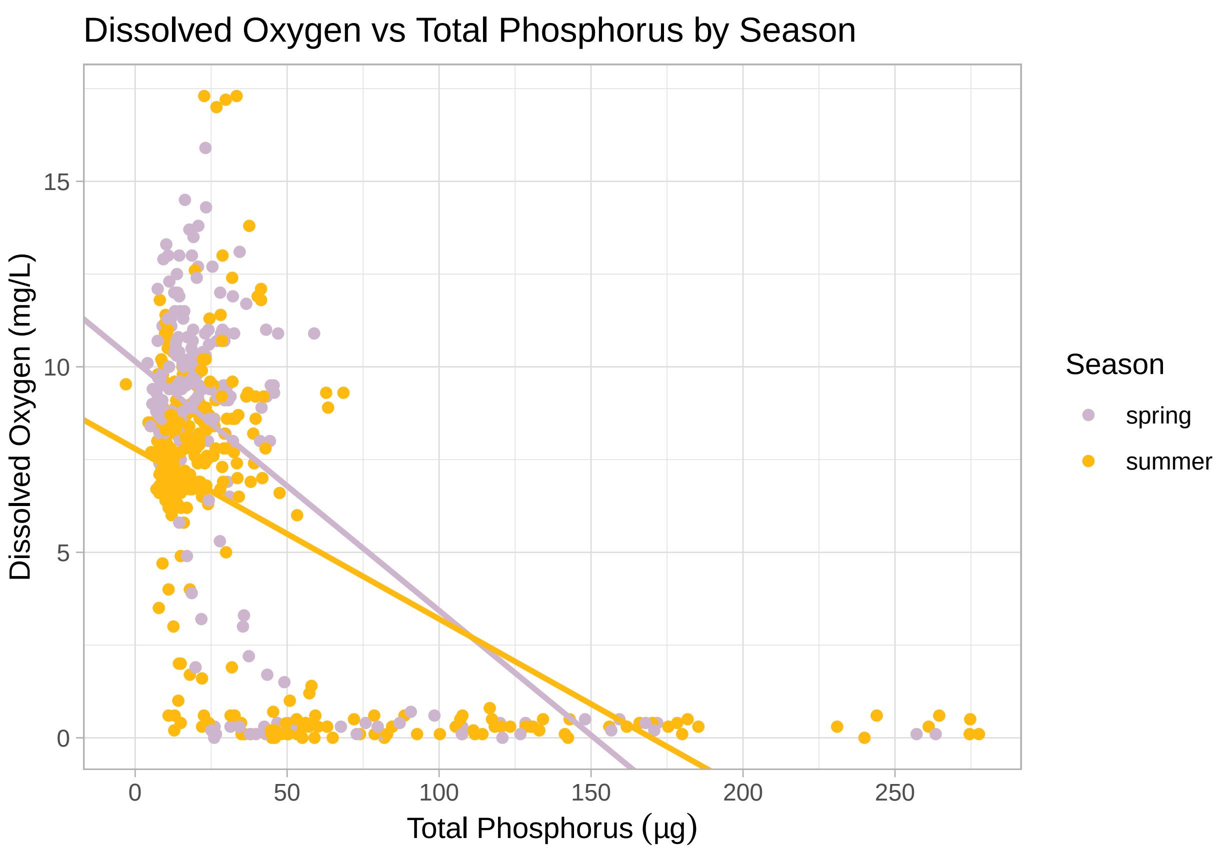


Figure 1: Figure 1

These results show that with increasing concentrations of phosphorus, there is a decrease in dissolved oxygen, and that the impact of phosphorus on DO levels is slightly stronger in the spring. These results are significant (all of the p-values are less than 0.05 for each coefficient), but that doesn't mean they are the best model of these data. The MAPE score of 2.99 shows that there is high error when you compare the fitted and observed data.

MODEL FIT TWO: Dissolved Oxygen (DO) by Total Phosphorus (TP) with a Water Depth Binary Indicator Variable and a Water Depth Interaction Term

Variable	Value	Interpretation
β_0 (Intercept)	8.52 mg/L	Expected value of DO in surface water with 0 tp
β_1	-7.67 mg/L	Difference in DO b/w sub- and surface water with 0 tp
β_2	-.0002 mg/L	Expected change in D0 for each unit increase in tp in surface water
β_3	-0.003 mg/L	Adjustment to slope for interaction term between tp in subsurface water

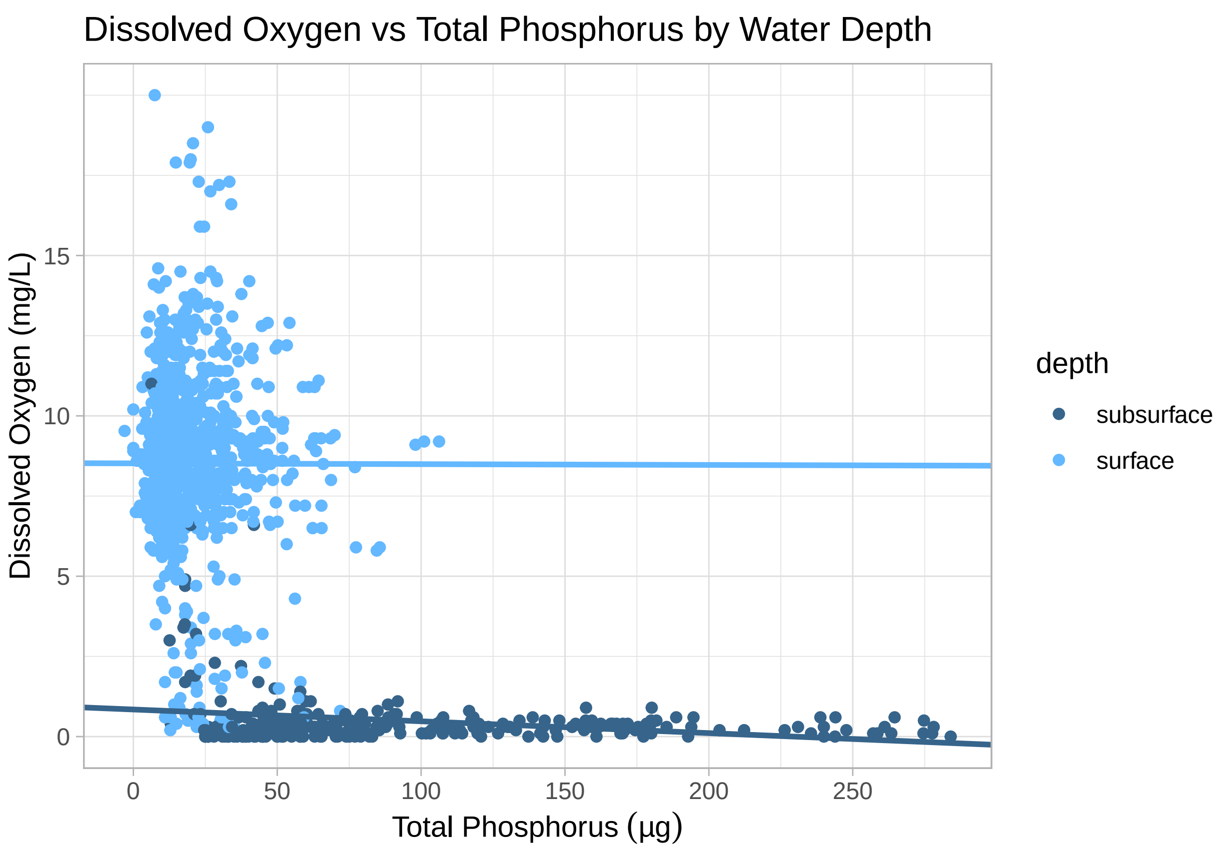


Figure 2: Figure 2

Results Model Two

The results from this model tell us that the expected value for dissolved oxygen in surface waters with zero total phosphorus is 8.52 mg/L. In subsurface waters with zero total phosphorus, the dissolved oxygen level decreases by 7.67 mg/L. The rate of change of the dissolved oxygen concentration for each unit increase of micrograms of total phosphorus in the water is -0.0002 in surface waters and -0.003 for subsurface waters.

Figure 2 shows a graphical representation of this linear model. These data are consistent with our assumptions given what we know about the relationship between dissolved oxygen and total phosphorus. As total phosphorus levels increase, the phosphorus in the water stimulates the growth of algae, a process known as eutrophication. This algae consumes the oxygen in the water as it decomposes, and the level of dissolved oxygen decreases. Simultaneously, the algae growth in the water halts the production of oxygen because the water gets murkier and plants are not able to photosynthesize as efficiently. Knowing that this process occurs, it is logical that the levels of dissolved oxygen are lower in deeper waters because the sun penetration is weaker and plants are not photosynthesizing enough to overcome the depletion of oxygen caused by eutrophication. It follows that the downward slope of the line is also logical because the dissolved oxygen is decreasing as total phosphorus increases.

Although the results of this linear model are consistent with our predictions, the model still has a high degree of error. The QQ plot shown in Appendix Plot 2 does not form a straight line, which means that the residuals from this model do not follow a normal distribution. The clouds of data points in Figure 2 are consistent with this observation; the points are not evenly distributed about the best fit line. Additionally, the MAPE score of 2.01 is an improvement as compared to the first model we fit but still shows a high degree of error.

MODEL FIT THREE: Dissolved Oxygen by Total Phosphorus with a Hierarchical Linear Mixed Effects Lake Variable

a. β_0 (Intercept): 8.68 mg/L

This represents the expected value of dissolved oxygen with zero total phosphorus.

b. β_1 (tp_ug): -0.055 mg/L

This represents the change in dissolved oxygen for each unit increase in total phosphorus in surface water.

Variable	Value	Interpretation
β_0 (Intercept)	8.68 mg/L	Expected value of D0 with zero phosphorus
β_1 (tp_ug)	-0.55 (mg/L)/μg	Change in dissolved oxygen for each unit increase in TP

Model Three Results

This model evaluates dissolved oxygen (dependent) based upon total phosphorus (independent) in each lake in our subset. The hierarchical model evaluates these two variables, with consideration that each lake may have introduce differences in the relationship between the two, whose impacts should not be reflected in the pure randomness between the two variables but can be generalized between all lakes. Like the previous two models, linearity is assumed in the data, along with independence due to the nature of the data collection process. Normality of residuals and homoscedasticity are assumptions made of successful models, however our diagnostic plots demonstrate that this does not

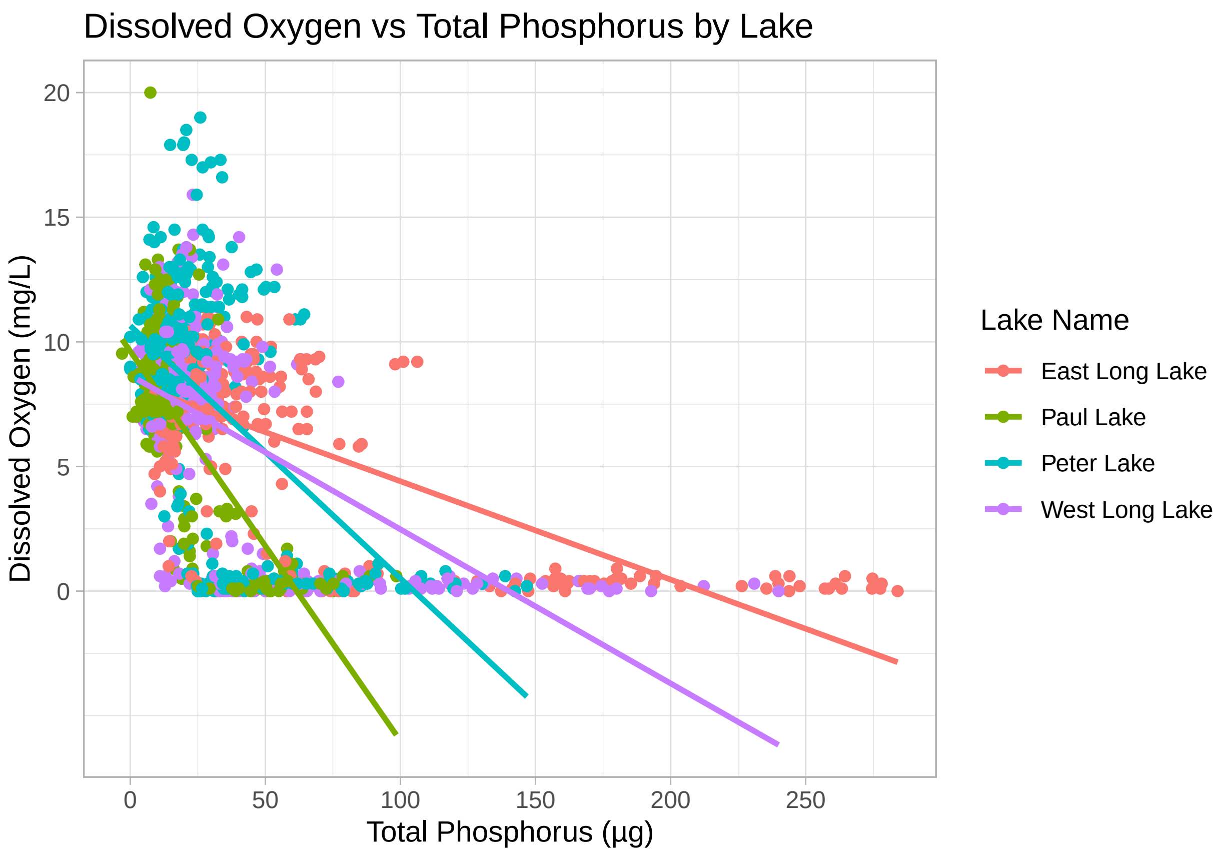


Figure 3: Figure 3

Based on the full dataset, the fixed estimate for the total phosphorus in micrograms is approximately -0.056, and for the training dataset it is approximately -0.054. Given the fixed estimate based on the full dataset, this suggests that for every one-unit increase in total phosphorus, dissolved oxygen decreases by approximately 0.056 units.

Analysis and Conclusion

Model	MAPE
Depth Model	2.01
Seasonal Model	2.93
Hierarchical Model by Lakename	3.07

While each of our fits demonstrate slight relationships between the variables investigated, the residual analysis of each indicates that error is not randomly distributed in every model. Further, the plots show two main “clouds” of data, with clusters of points at both high dissolved oxygen and low phosphorus, and consistently low (near zero) dissolved oxygen at the full range of phosphorus levels. This, along with other efforts seen in the appendix aimed at transforming or separating data based on different indicators, failed to separate these two clouds of data points beyond manually sub-setting the data based on dissolved oxygen and phosphorus values.

The depth model has the lowest MAPE value, indicating that this fit is the best of our three we evaluate. This aligns with expectations, as depth should have a larger effect on the quantity of dissolved oxygen due to gas exchange at the surface and mixing. While the depth model is best of the three we evaluate, there are also distinct difference between spring (may) and summer (august) data as well as differecnes between each lake in the relationship between phosphorus and dissolved oxygen.

Scientific evidence demonstrates increases in phosphorus lead to decreases in dissolved oxygen due to eutrophication events in bodies of water. However, this data, along with our research, does not indicate an exact relationship. The two clusters of data may suggest that beyond a certain quantity of phosphorus, or perhaps after a threshold required for a eutrophication event, dissolved oxygen drops to near zero. Thus, while we were unable to uncover quality models to predict the outcome of the quantity of dissolved oxygen based upon our variables, we were able to see this slight negative relationship and critical quantity at which dissolved oxygen largely does not exist beyond near zero values at high quantities of phosophorus.

Bibliography

Carpenter, S. R., Cole, J. J., Hodgson, J. R., Kitchell, J. F., Pace, M. L., Bade, D., … & Schindler, D. E. (2001). Trophic cascades, nutrients, and lake productivity: whole-lake experiments. Ecological monographs, 71(2), 163-186.

Lee, G. F. (1973). Role of phosphorus in eutrophication and diffuse source control. In Phosphorus in Fresh Water and the Marine Environment (pp. 111-128). Pergamon.

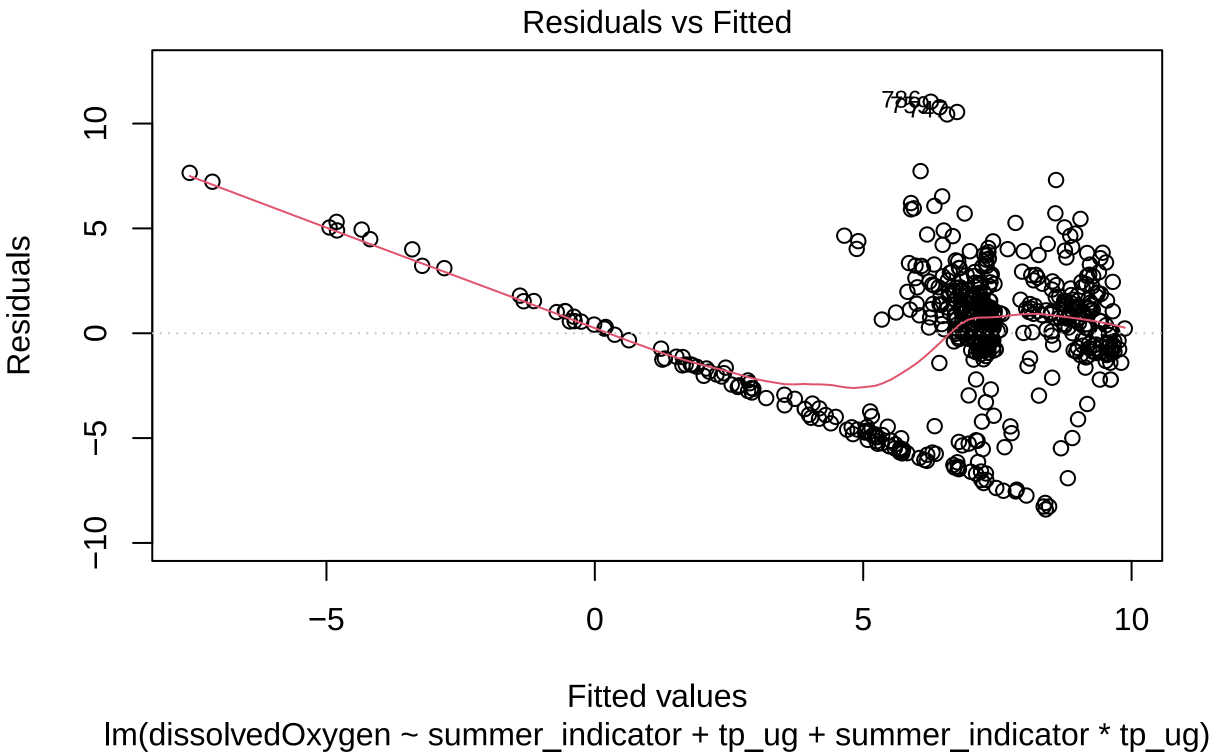
Total nitrogen/total phosphorus and chlorophyll a: Volunteer data entry and information: indiana clean lakes program: indiana university. (n.d.). Indiana Clean Lakes Program. Retrieved April 15, 2024, from https://clp.indiana.edu/volunteer-data/phosphorus-chlorophyll.html

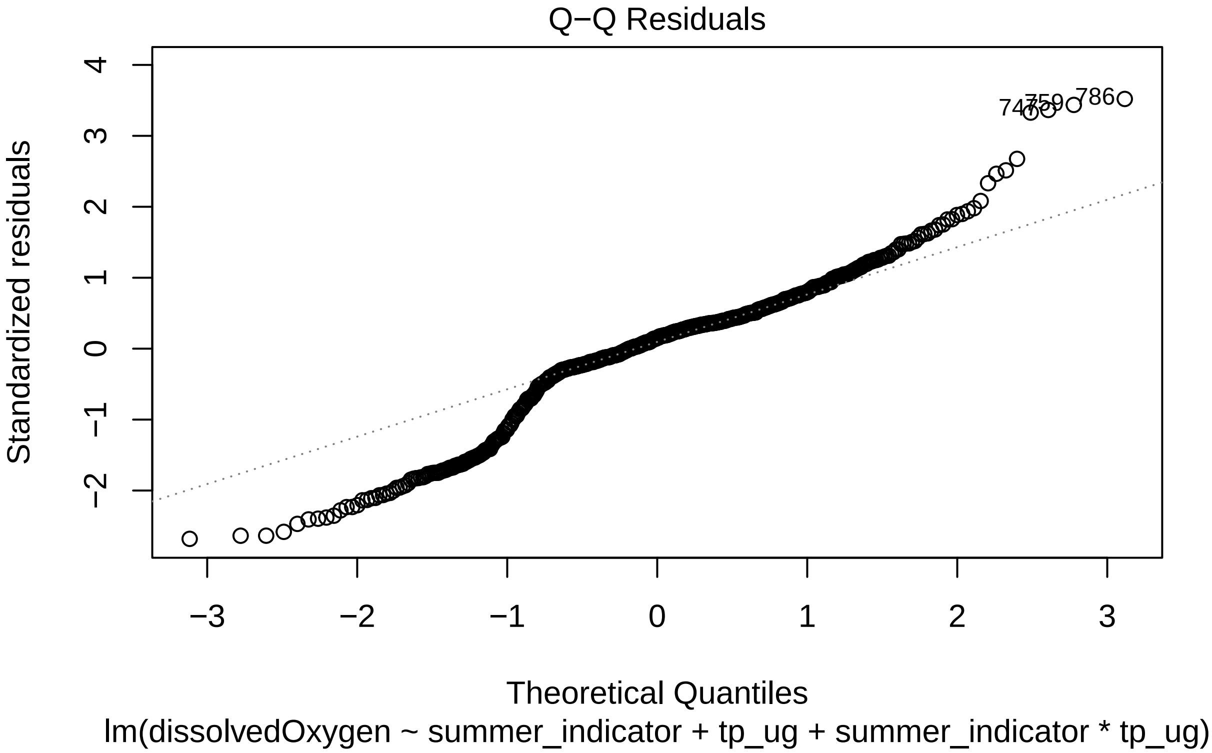
US EPA, O. (2013, November 20). Indicators: Dissolved oxygen [Overviews and Factsheets]. https://www.epa.gov/national-aquatic-resource-surveys/indicators-dissolved-oxygen

US EPA, O. (2013, November 27). Indicators: Phosphorus [Overviews and Factsheets]. https://www.epa.gov/national-aquatic-resource-surveys/indicators-phosphorus

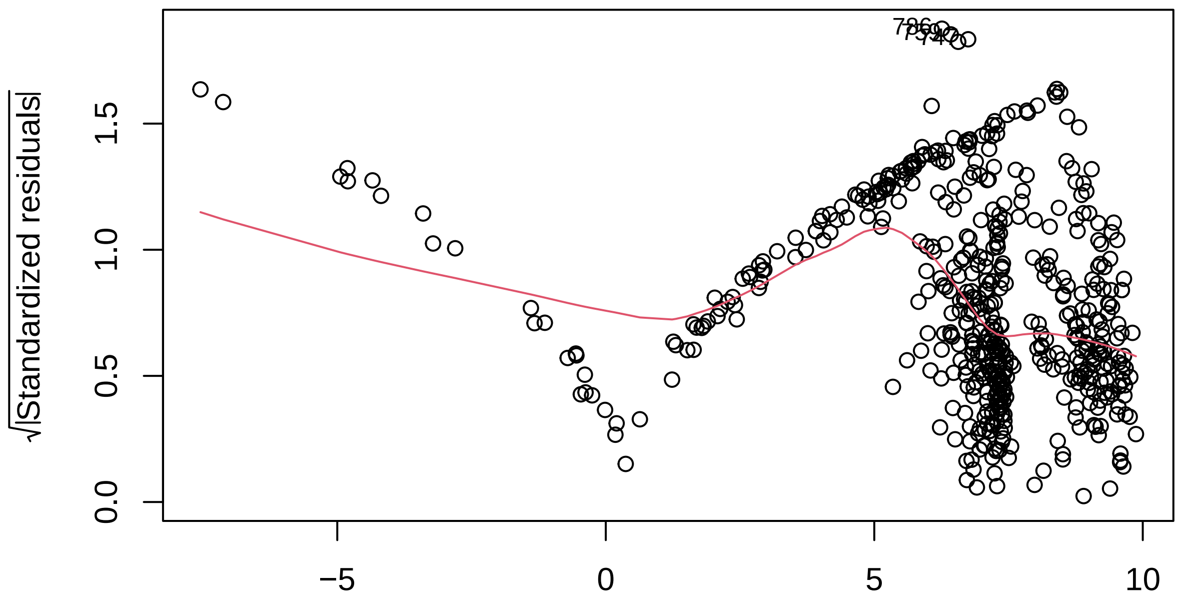
Appendix

Model One Residuals



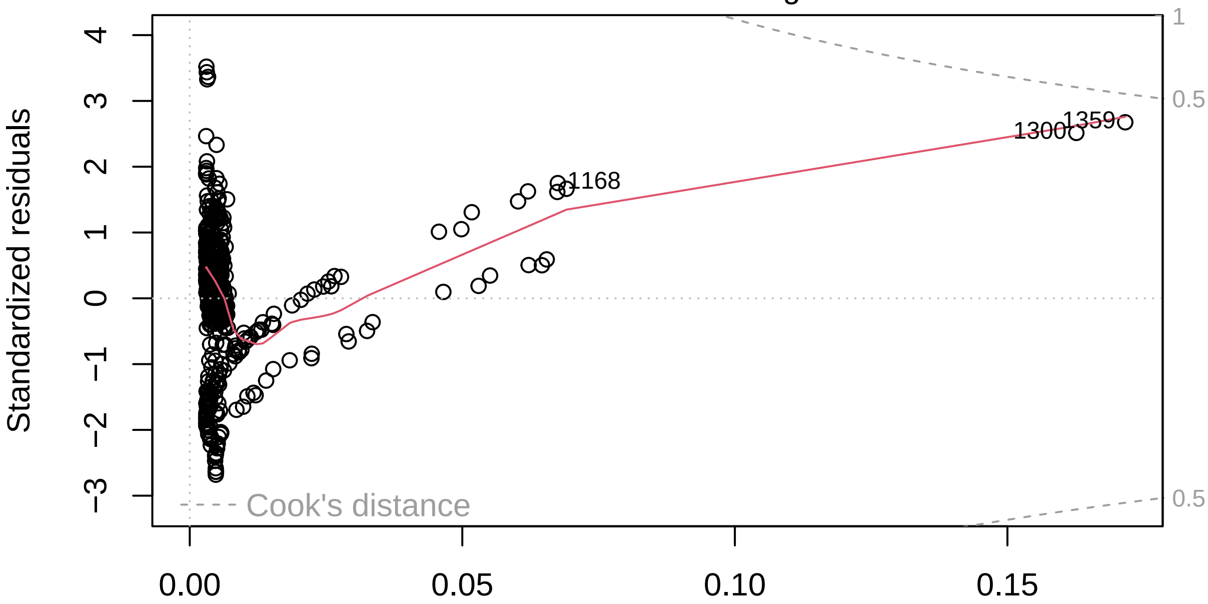


Scale−Location



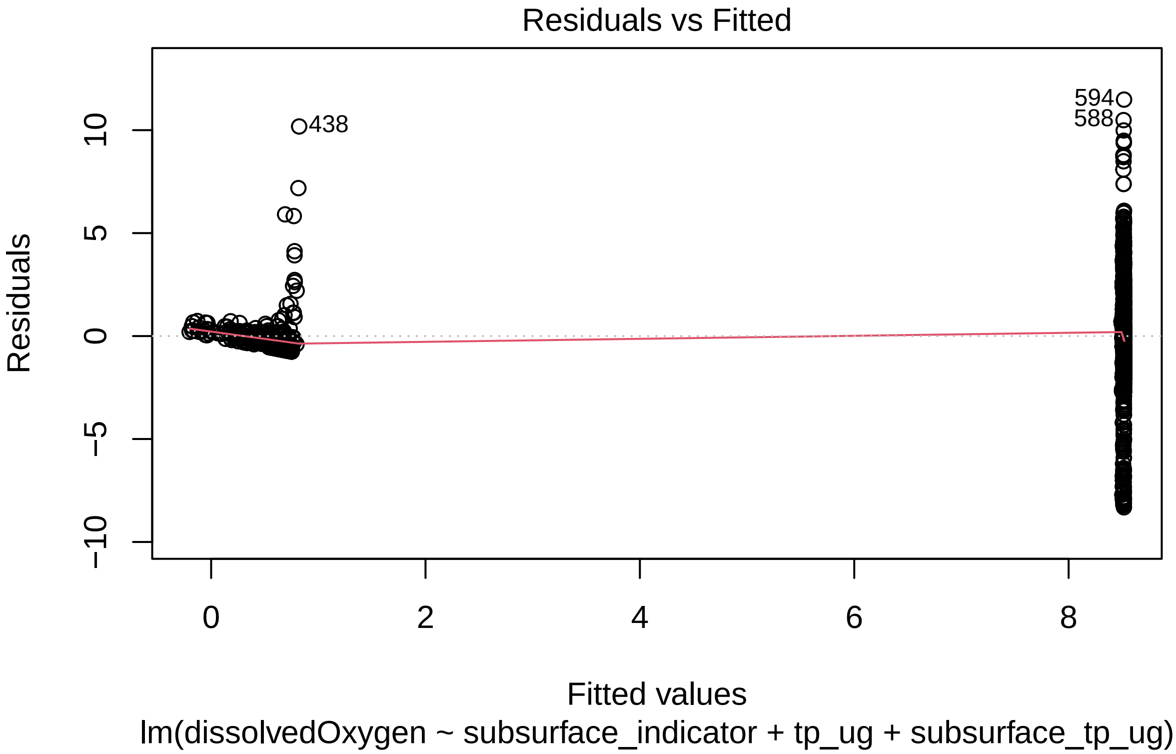
Fitted values
lm(dissolvedOxygen ~ summer_indicator + tp_ug + summer_indicator * tp_ug)

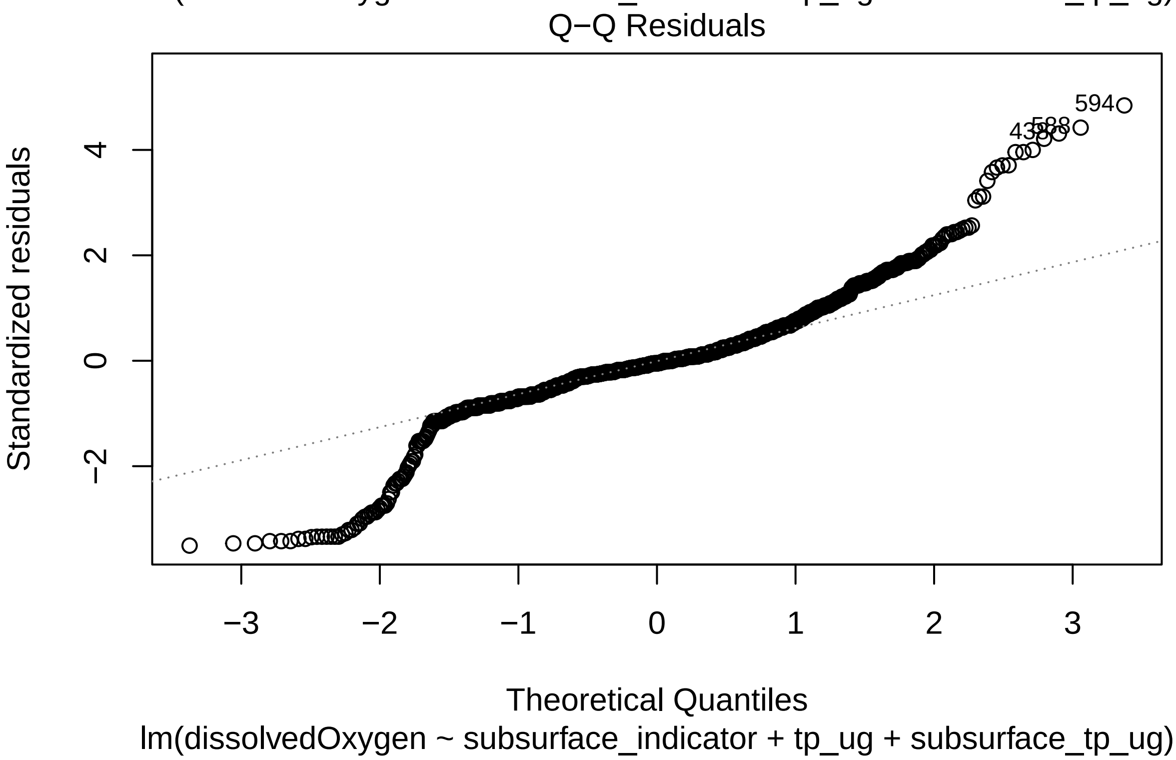
Residuals vs Leverage



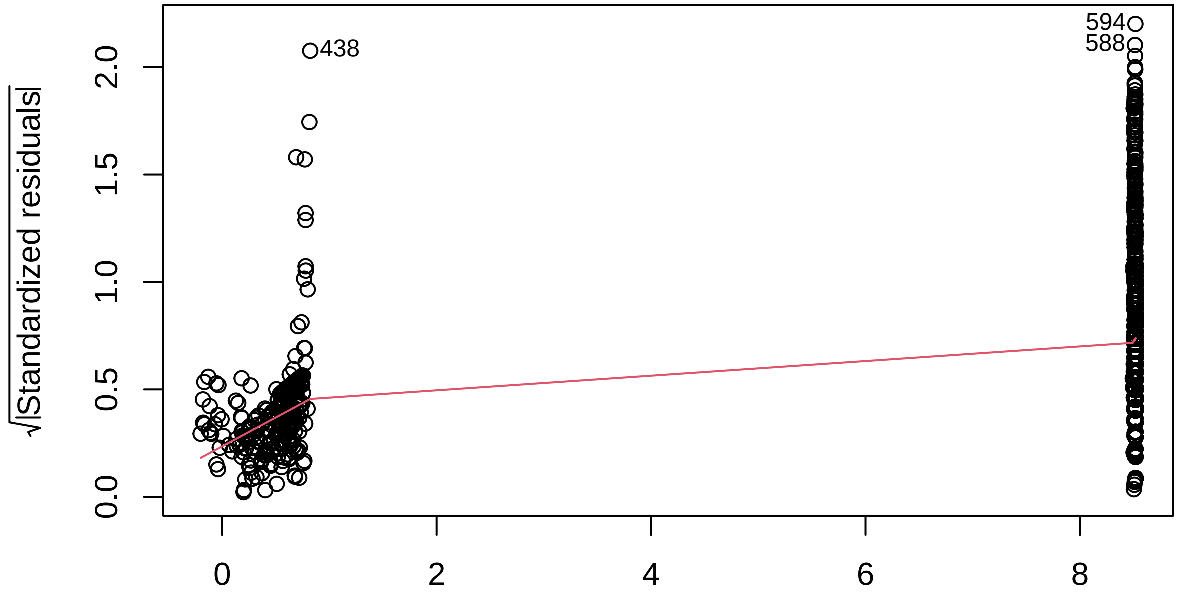
Leverage
lm(dissolvedOxygen ~ summer_indicator + tp_ug + summer_indicator * tp_ug)

Model Two Residuals



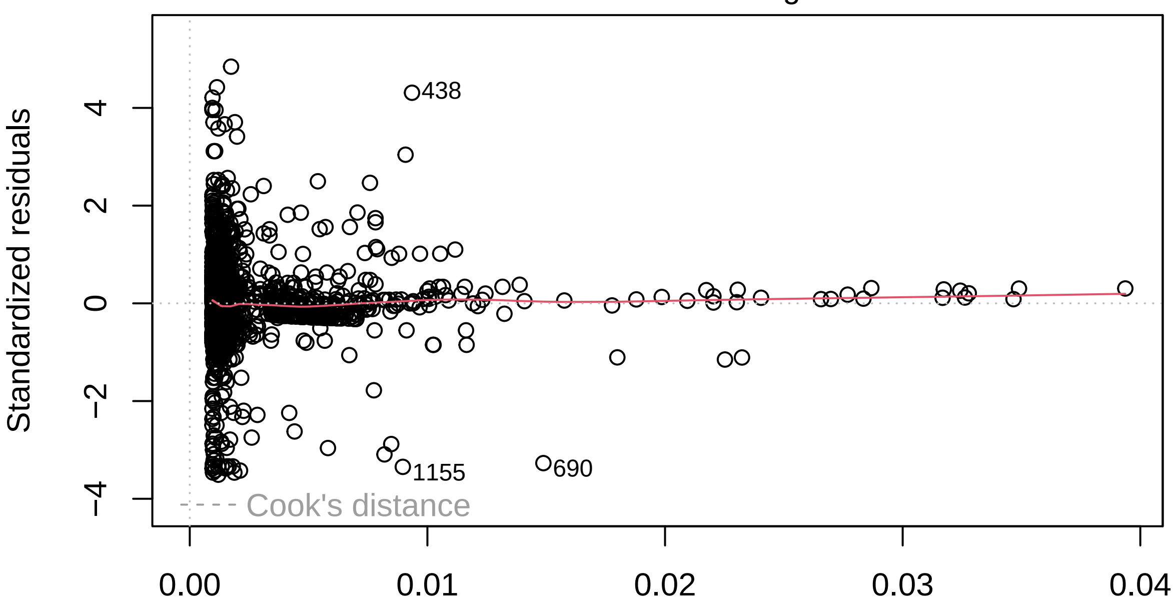


Scale−Location



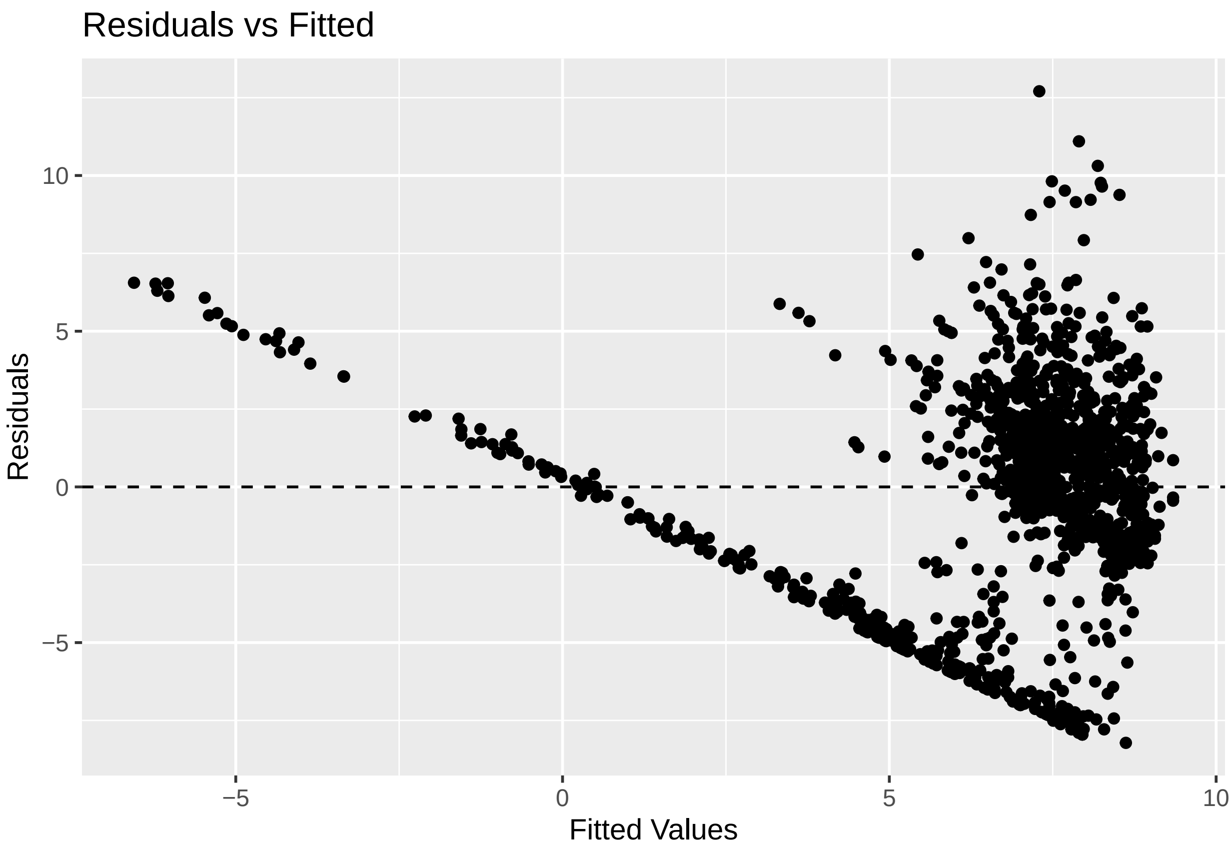
Fitted values
lm(dissolvedOxygen ~ subsurface_indicator + tp_ug + subsurface_tp_ug)

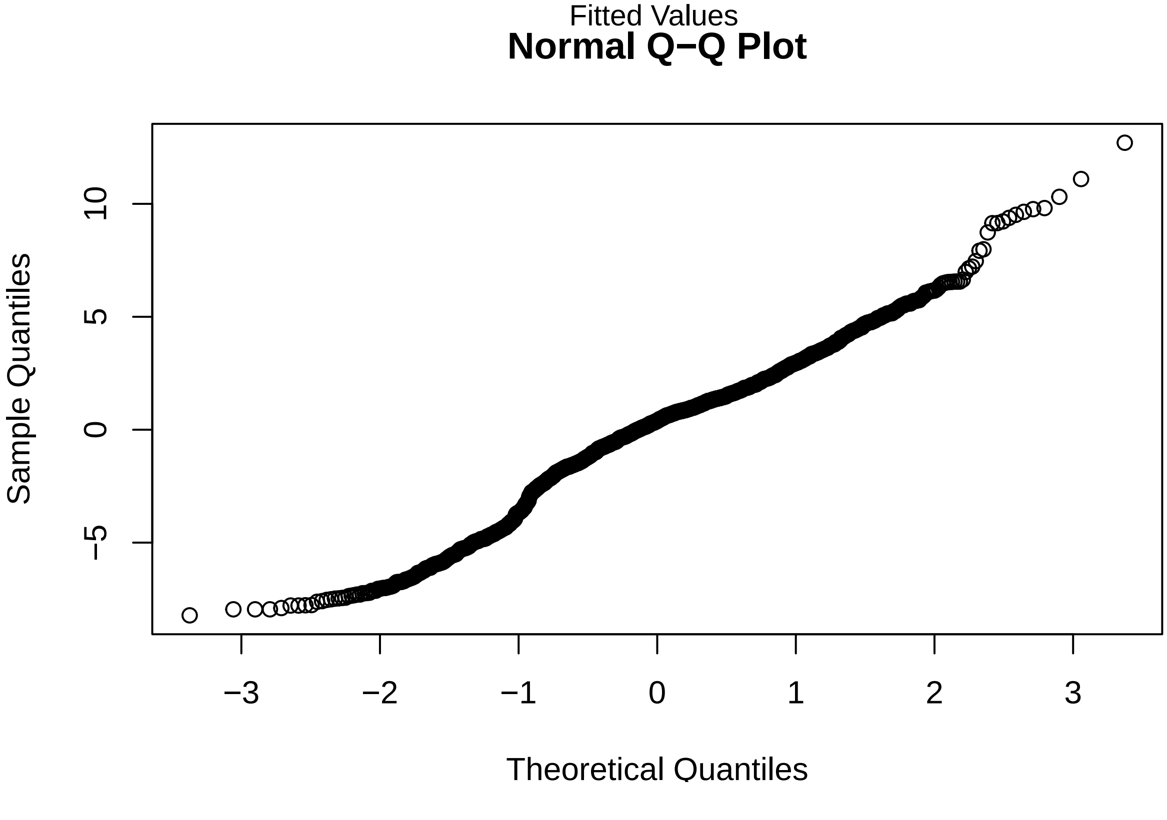
Residuals vs Leverage

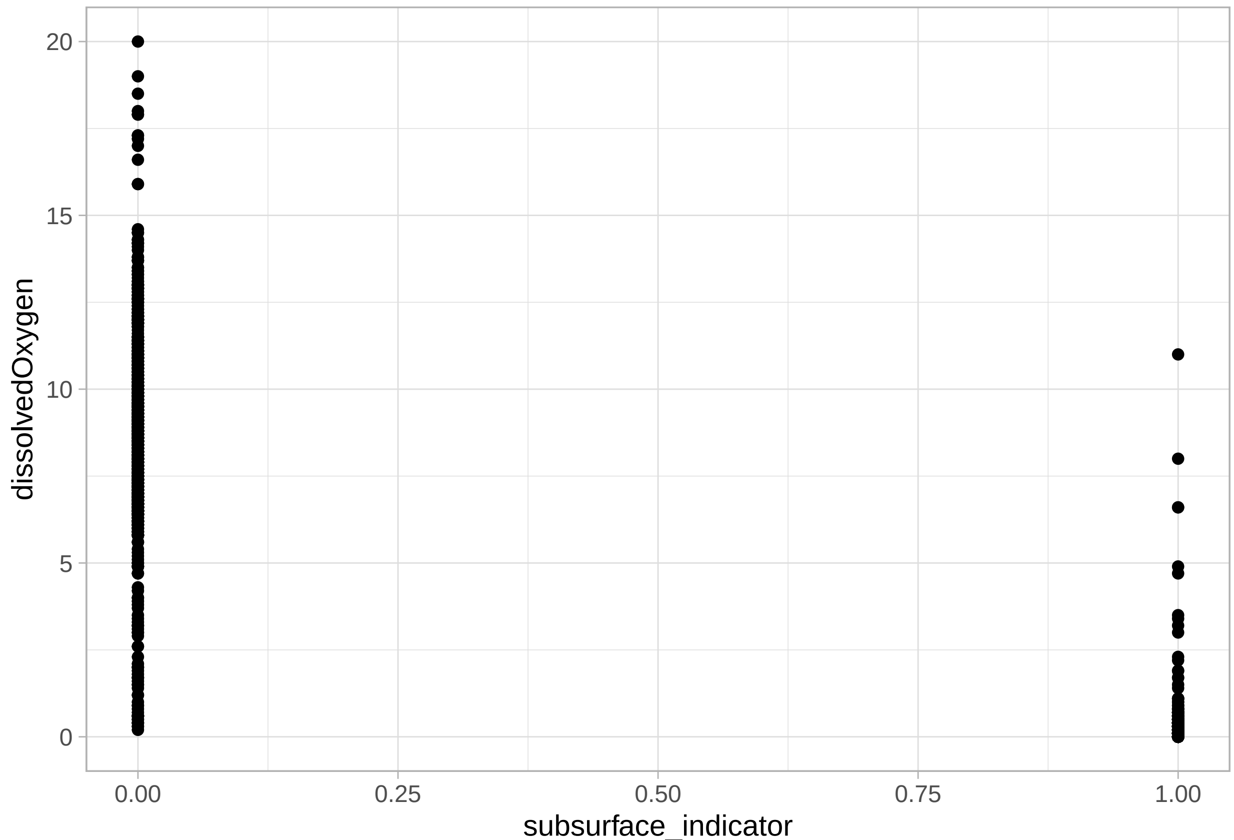


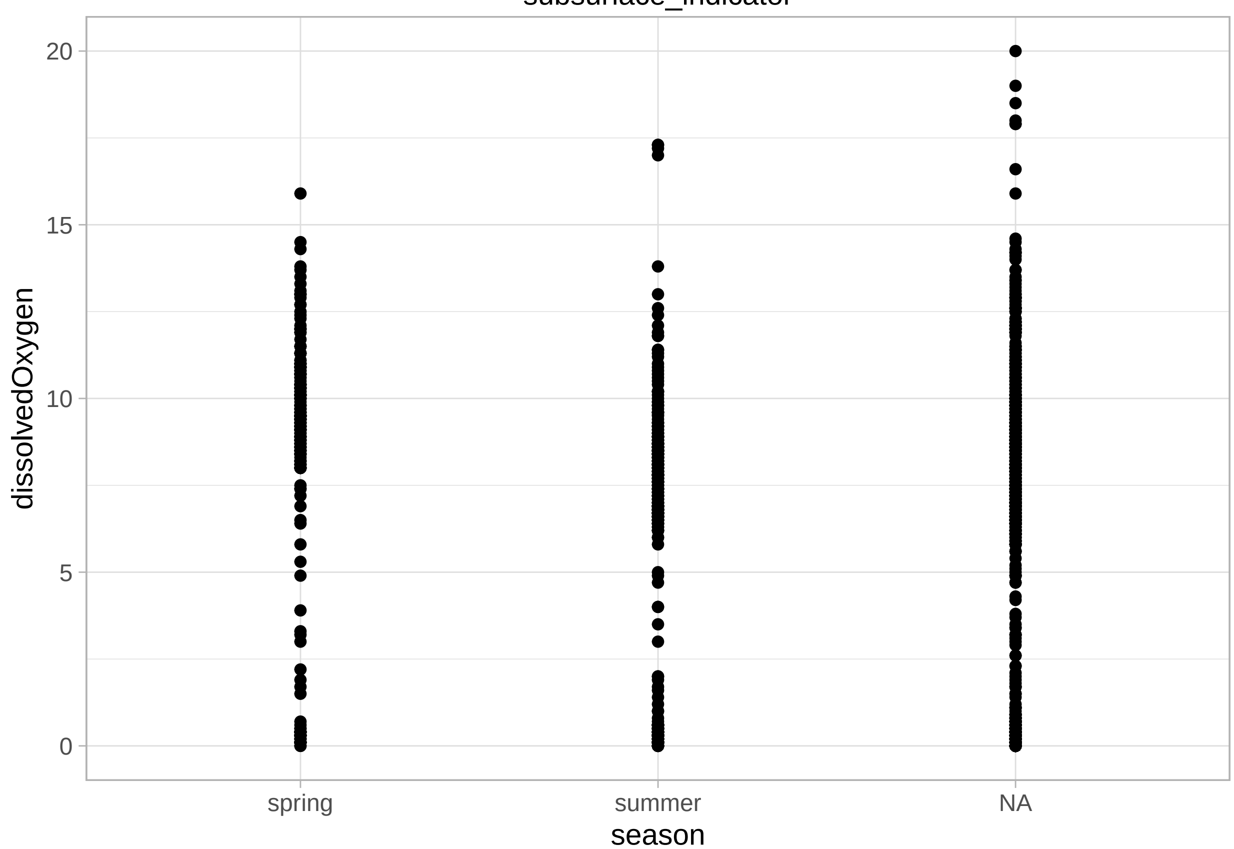
Leverage
lm(dissolvedOxygen ~ subsurface_indicator + tp_ug + subsurface_tp_ug)

Model Three Residuals

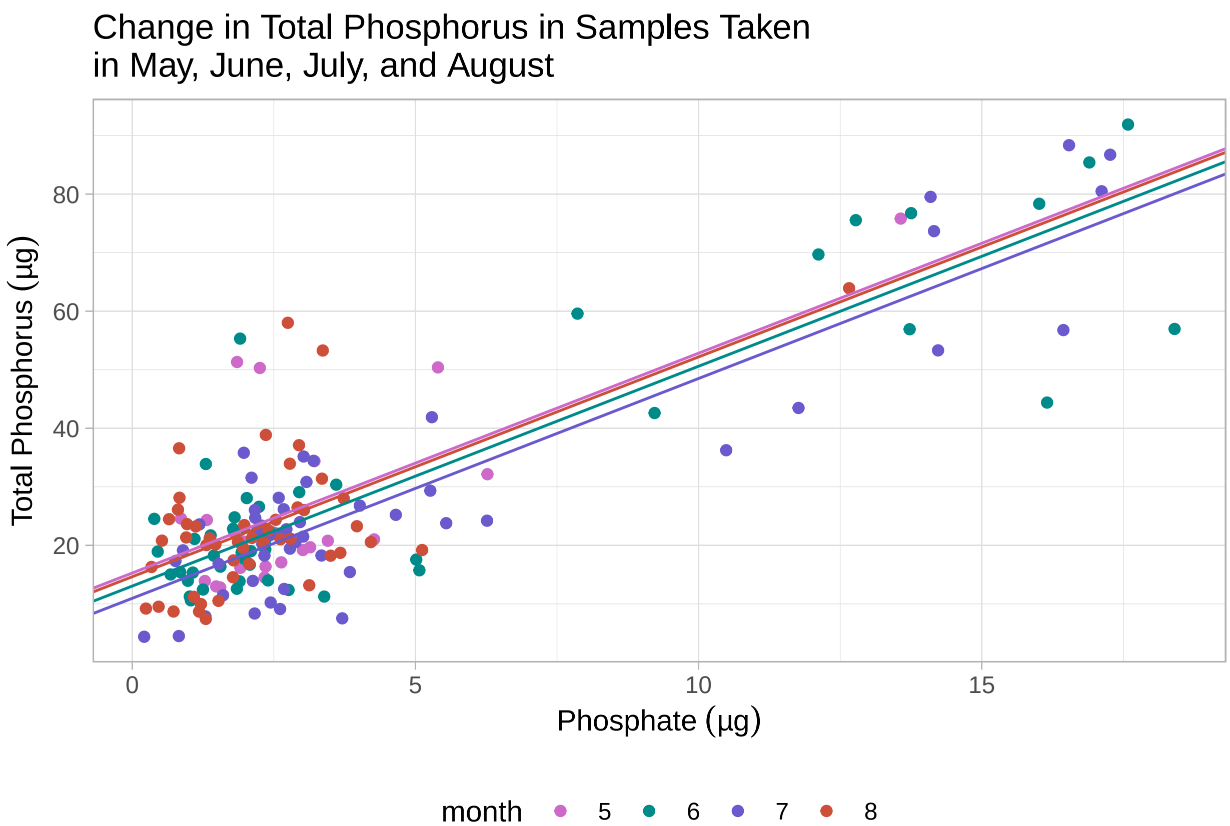





20
15
10
5
0
dissolvedOxygen
0.00
0.25
0.50
0.75
1.00
subsurface_indicator

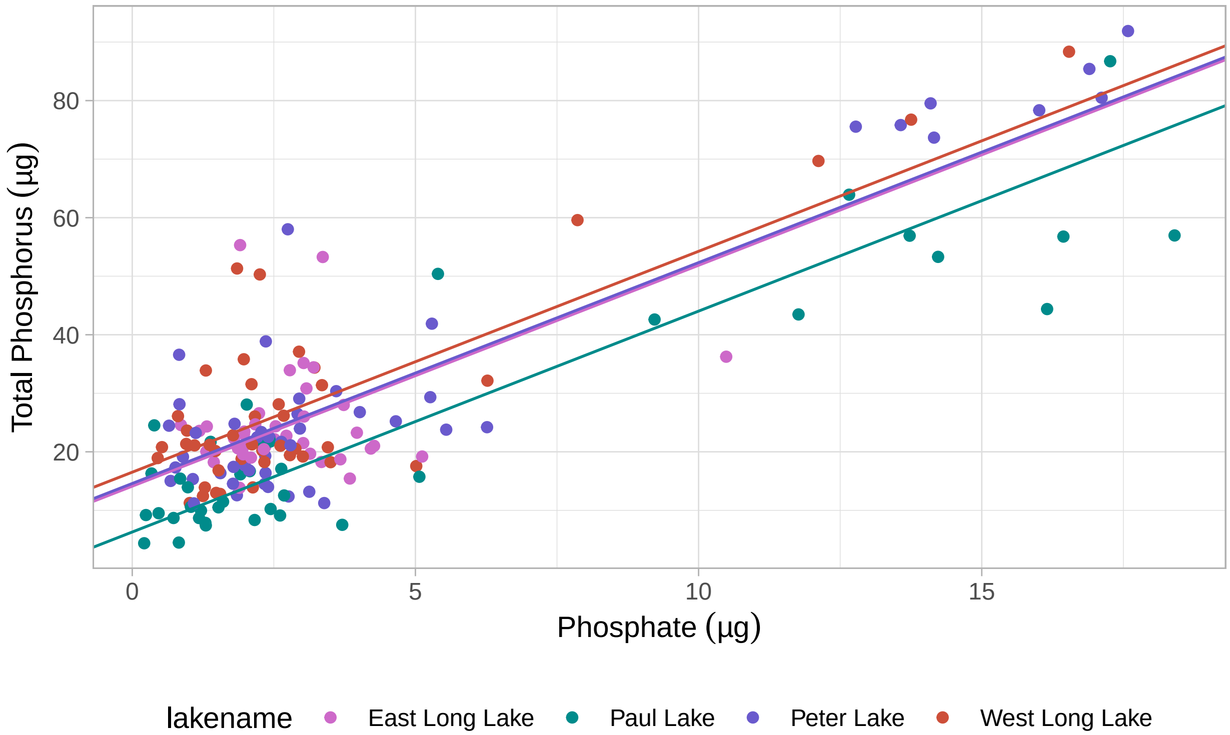

20
15
10
5
0
dissolvedOxygen
spring
summer
NA
season

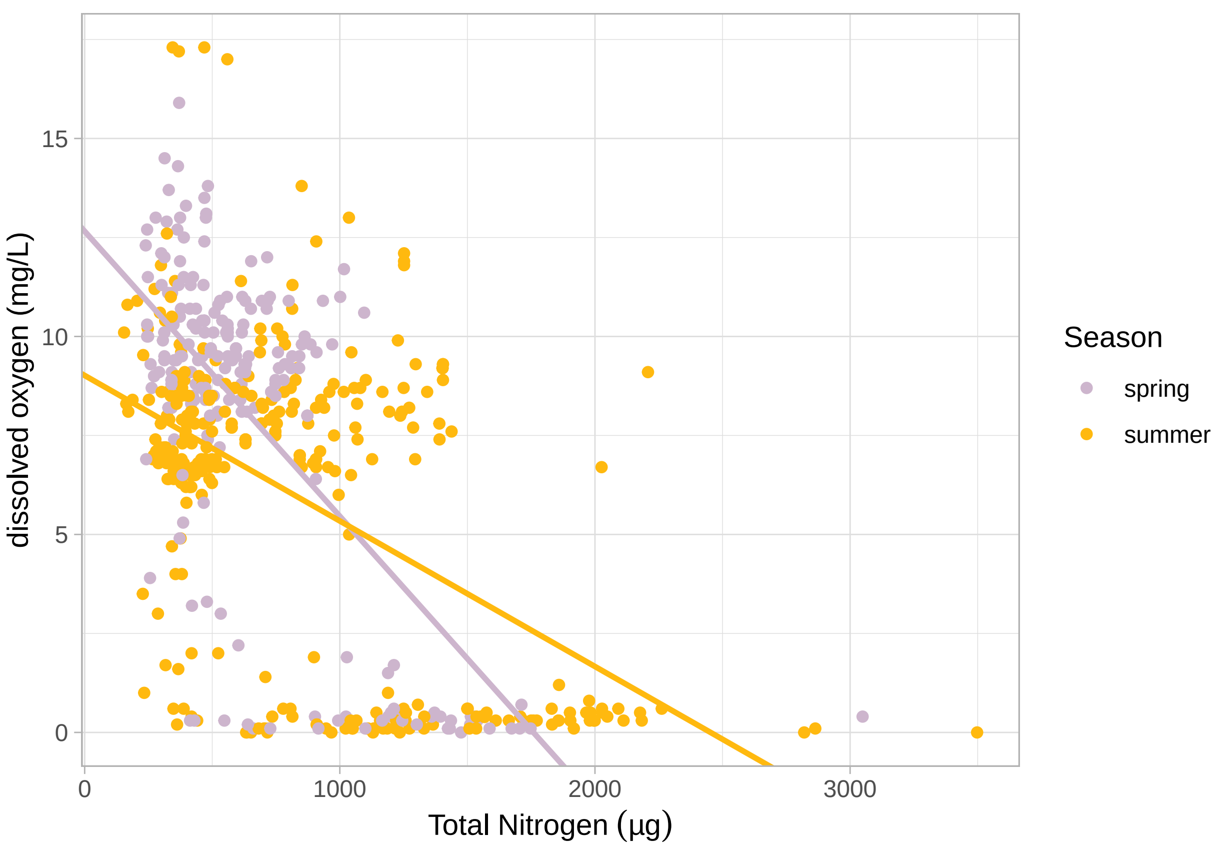
```
##
## Call:
## lm(formula = tp_ug ~ month + po4, data = lake_1994)
##
## Coefficients:
## (Intercept)       month6       month7       month8          po4
##     15.2690      -2.2260      -4.3199      -0.6307       3.7553
```




```
##
## Call:
## lm(formula = tp_ug ~ lakename + po4, data = lake_1994)
##
## Coefficients:
##           (Intercept)      lakenamePaul Lake      lakenamePeter Lake
##               14.1293                -7.8305                  0.4614
## lakenameWest Long Lake                    po4
##                2.3909                 3.7737
```

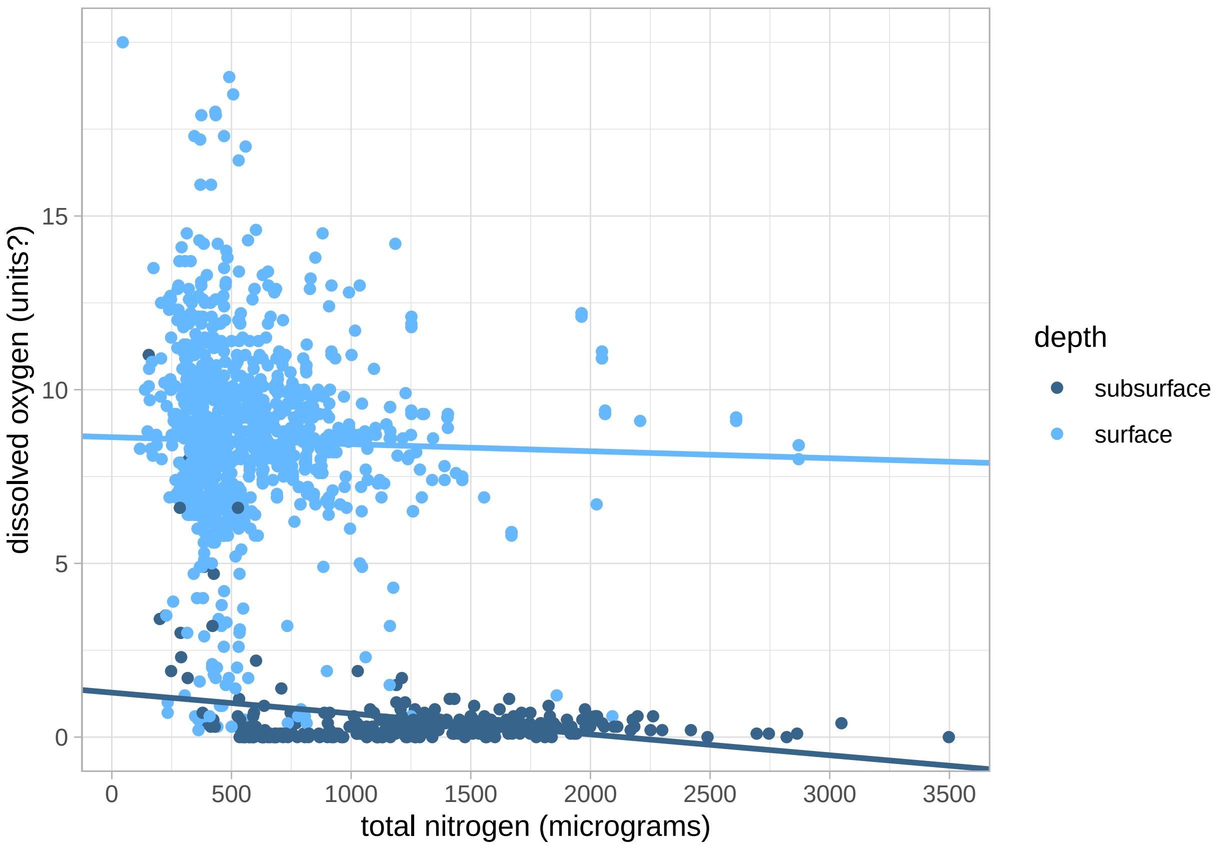
Change in Total Phosphorus in Samples Taken
at East Long Lake, Paul Lake, Peter Lake, and West Long Lake





##

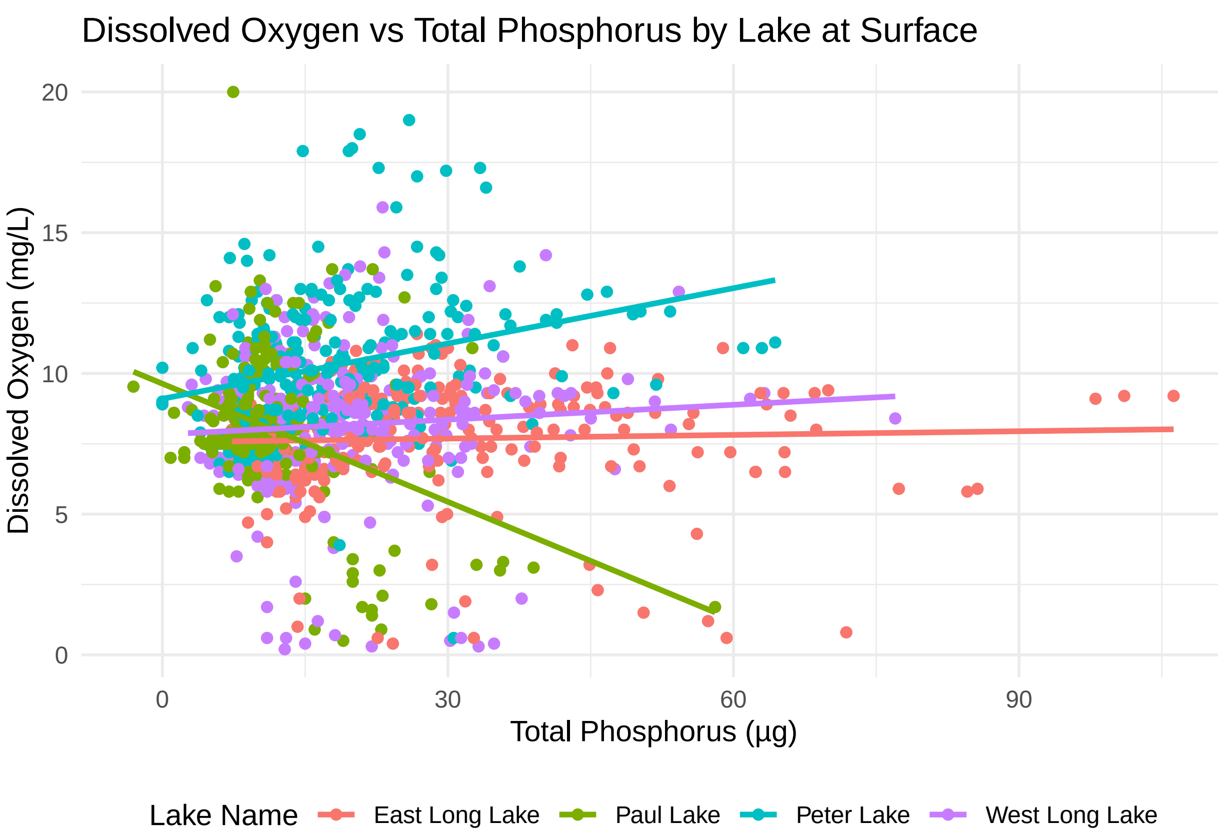
```
## Call:
## lm(formula = dissolvedOxygen ~ subsurface_indicator + tn_ug +
##     subsurface_tn_ug, data = jenn)
##
## Coefficients:
##          (Intercept)  subsurface_indicator                 tn_ug
##            8.6334713            -7.3533011            -0.0002020
##     subsurface_tn_ug
##           -0.0003988
```




```
## Linear mixed model fit by REML ['lmerMod']
## Formula: dissolvedOxygen ~ tp_ug + (1 | lakename)
##    Data: lakes_processed_surface
##
## REML criterion at convergence: 4851.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0558 -0.5070  0.0015  0.5606  5.0058
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  lakename (Intercept) 1.441    1.200
##  Residual             5.819    2.412
## Number of obs: 1050, groups:  lakename, 4
##
```

```
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 8.330860   0.616156  13.521
## tp_ug       0.009972   0.006026   1.655
##
## Correlation of Fixed Effects:
##       (Intr)
## tp_ug -0.191
```

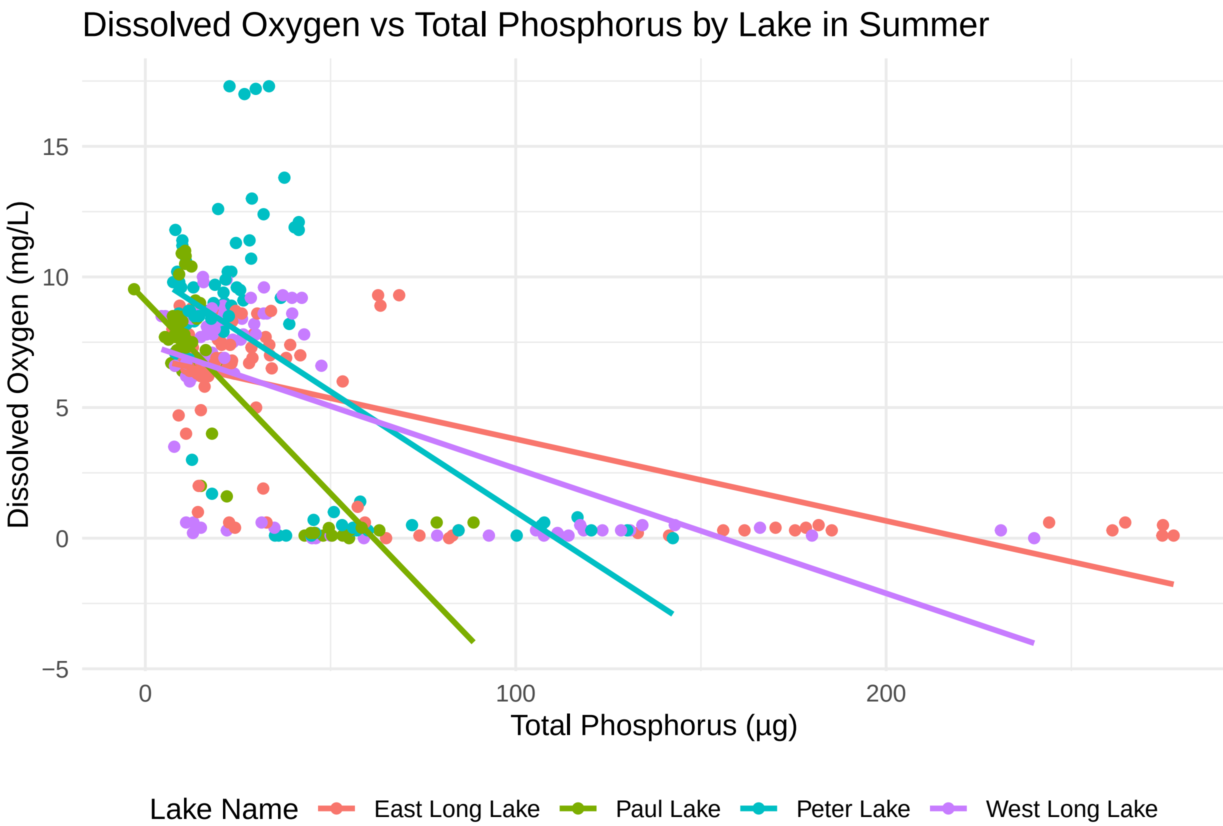
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: dissolvedOxygen ~ tp_ug + (1 | lakename)
##    Data: lakes_processed_surface
## REML criterion at convergence: 4851.247
## Random effects:
##  Groups   Name        Std.Dev.
##  lakename (Intercept) 1.200
##  Residual             2.412
## Number of obs: 1050, groups:  lakename, 4
## Fixed Effects:
## (Intercept)        tp_ug
##    8.330860     0.009972
```




```
## Linear mixed model fit by REML ['lmerMod']
## Formula: dissolvedOxygen ~ tp_ug + (1 | lakename)
##    Data: lakes_processed_summer
##
## REML criterion at convergence: 1703.4
```

```
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2227 -0.3910  0.1498  0.5243  3.3339
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  lakename (Intercept) 0.4257   0.6524
##  Residual             9.4989   3.0820
## Number of obs: 332, groups:  lakename, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  7.750824   0.391567   19.79
## tp_ug       -0.045821   0.003436  -13.34
##
## Correlation of Fixed Effects:
##       (Intr)
## tp_ug -0.342
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: dissolvedOxygen ~ tp_ug + (1 | lakename)
##    Data: lakes_processed_summer
## REML criterion at convergence: 1703.426
## Random effects:
##  Groups   Name        Std.Dev.
##  lakename (Intercept) 0.6524
##  Residual             3.0820
## Number of obs: 332, groups:  lakename, 4
## Fixed Effects:
## (Intercept)        tp_ug
##     7.75082     -0.04582
```



Dissolved Oxygen vs Total Phosphorus by Lake in Summer
Dissolved Oxygen (mg/L)
15
10
5
0
−5
0
100
200
Total Phosphorus (µg)
Lake Name
East Long Lake
Paul Lake
Peter Lake
West Long Lake

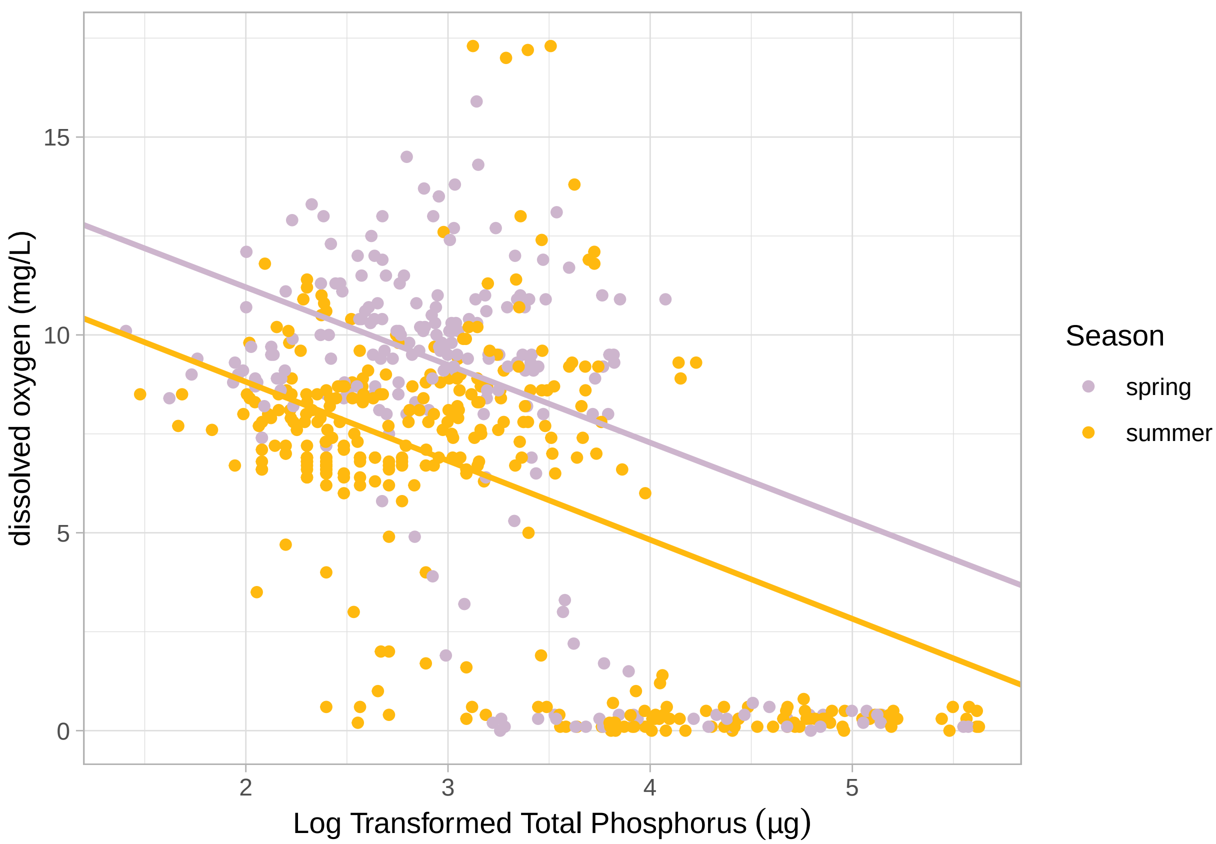


Figure 4: Appendix plot log-transform

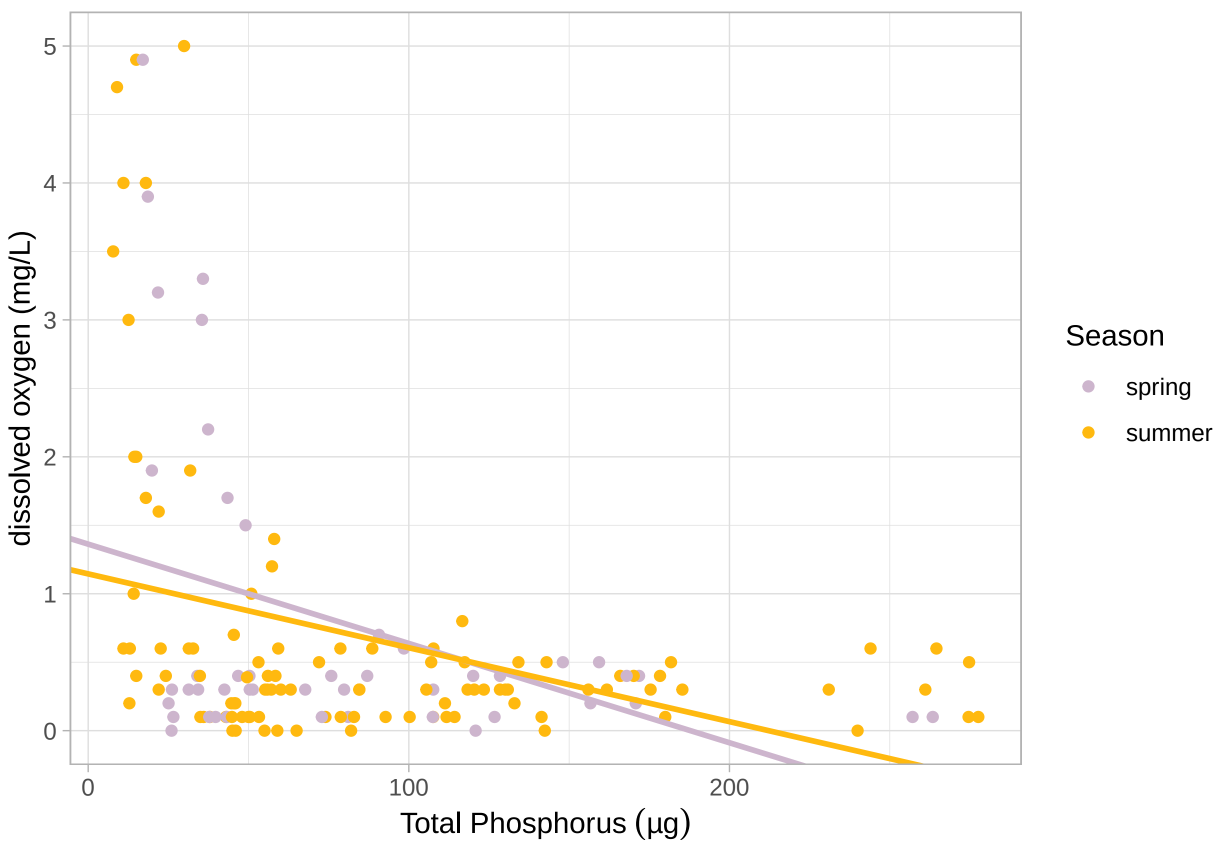


Figure 5: Appendix plot subset DO