Assessing phytoplankton abundance in North Temperate Lakes

https://github.com/Sylvia-C/872_Final.git

Siying Chen

Abstract

Phytoplankton growth depends on the availability of carbon dioxide, sunlight, and nutrients. This project uses data from studies on several lakes in the North Temperate Lakes District in Wisconsin, USA as part of the Long Term Ecological Research station established by the National Science Foundation. The purpose of this project is to find out the relationship between phytoplankton abundance (biovolume) and nutrients (nitrogen and phosphorus) using regression analysis. There are 2 datasets used in this project, which are the nutrients dataset and phytoplankton dataset of the lakes.

Contents

List of Tables

List of Figures

r Figure 1, echo =F, fig.cap="x" r Figure 2, echo =F, fig.cap="x" r Figure 3, echo =F, fig.cap="x" r Figure 4, echo =F, fig.cap="x" r Figure 5, echo =F, fig.cap="x" r Figure 6, echo =F, fig.cap="x"

<Note: set up autoreferencing for figures and tables in your document>

1 Research Question and Rationale

Phytoplankton growth depends on the availability of carbon dioxide, sunlight, and nutrients. Like other plants, phytoplanktons require nutrients like nitrogen and phosphorus to survive. In this project, I would like to look at total nitrogen and total phosphorus concentration and see how they interact with phytoplankton abundance, which was measured in biovolume. This project can help peole to understand the relationship between phytoplankton growth and the concentration of nutrients like nitrogen and phosphorus at North Temperate Lakes.

Due to data availability, we will only be looking at East Long Lake, Paul Lake, Peter Lake, Tuesday Lake, and West Long Lake. For this project, I'll be focusing on the division level of phytoplankton abundance, which are measured in biovolume in millimeter cubed per liter. I'll sum up the total biovolume of different phytoplankton division at each lake on each data. For nutrients data, since the nutrient concentration was collected at different depth, I'll use the median concentration for each lake on each date to avoid outliers. The research question for this project are: 1. How total nitrogen concentration affect the abundance of phytoplankton in these 5 lakes? 2. How total phosphorus concentration affect the abundance of phytoplankton in these 5 lakes?

2 Dataset Information

Nutrient data were collected from 1991-2016. They are usually collected in the morning and collected at different depth. Chemical measurements are collected in several ways: pooled mixed layer sample (PML), epilimnion, metalimnion, and hypolimnion, or vertical profiles. On each sampling date, there are up to seven samples due to Van Dorn collections across a depth interval according to percent irradiance. Nutrient samples were sent to the Cary Institute of Ecosystem Studies for analysis beginning in 2000. The Kjeldahl method for measuring nitrogen is not used at IES, and so measurements reported from 2000 onwards are Total Nitrogen.

Data on epilimnetic phytoplankton were collected from 1984-2015, determined by light microscopy from pooled Van Dorn samples at 100 percent, 50 percent, and 25 percent of surface irradiance. Samples collected after 1995 were counted by Phycotech Inc. Sampling were conducted at 5 lakes and the frequency varies. Data include taxonomic information and split into distinct columns of genus, species and description for archival as best possible.

Dataset	Data Column	Unit
Nutrients	Total nitrogen concentration, total phosphorus concentration	Micrograms Per Liter
Phytoplankton	Total biovolume	Millimeter Cubed Per Liter

3 Exploratory Data Analysis and Wrangling

```
# Read in datasets as dataframe
nutrients <- read.csv("./Raw/NTL-LTER_Lake_Nutrients_Raw.csv")
phytoplankton <- read.csv("./Raw/NTL-LTER_Lake_Phytoplankton_Raw.csv")

# Format sampledate as date
nutrients$sampledate <-
  as.Date(nutrients$sampledate, format = "%m/%d/%y")
phytoplankton$sampledate <-
  as.Date(phytoplankton$sampledate, format = "%Y-%m-%d")

# Select and filter dataframe to slim the data
nutrients_slim <- nutrients %>%
  select(lakename, sampledate, depth, tn_ug, tp_ug) %>%
  na.omit() %>%
  group_by(lakename, sampledate) %>%
  summarize(medianTN = median(tn_ug),
            medianTP = median(tp_ug))

phytoplankton_slim <- phytoplankton %>%
  select(lakename, sampledate, division, total_biovol) %>%
  group_by(lakename, sampledate, division) %>%
  summarize(biomass = sum(total_biovol))

# Join datasets
biomass_nutrients <- phytoplankton_slim %>%
  summarize(tot_biomass = sum(biomass)) %>%
  left_join(nutrients_slim, by = c("lakename" = "lakename", "sampledate" = "sampledate")
  na.omit()
```

```
## Warning: Column `lakename` joining factors with different levels, coercing
## to character vector
```

```
#colnames(biomass_nutrients) <- c("Lake Name", "Sample Date", "Biomass", "Median TN",

# Summary statistics
head(phytoplankton_slim)
```

```
## # A tibble: 6 x 4
## # Groups:   lakename, sampledate [1]
##   lakename       sampledate division        biomass
##   <fct>          <date>     <fct>             <dbl>
## 1 East Long Lake 1989-05-26 Bacillariophyta   0.012
## 2 East Long Lake 1989-05-26 Chlorophyta       0.012
```

```
## 3 East Long Lake 1989-05-26 Chrysophyta       0.304
## 4 East Long Lake 1989-05-26 Cryptophyta       0.081
## 5 East Long Lake 1989-05-26 Miscellaneous     0.073
## 6 East Long Lake 1989-05-26 Pyrrhophyta       0.137
```

```
colnames(phytoplankton_slim)
```

```
## [1] "lakename"   "sampledate" "division"   "biomass"
```

```
dim(phytoplankton_slim)
```

```
## [1] 3911    4
```

```
summary(phytoplankton_slim)
```

```
##            lakename      sampledate                  division
##  East Long Lake: 498   Min.   :1984-05-27   Chlorophyta  :678
##  Paul Lake     :1147   1st Qu.:1987-07-28   Cryptophyta  :672
##  Peter Lake    :1109   Median :1991-05-22   Chrysophyta  :645
##  Tuesday Lake  : 658   Mean   :1990-07-05   Cyanophyta   :563
##  West Long Lake: 499   3rd Qu.:1993-06-28   Miscellaneous:520
##                        Max.   :1995-09-05   Pyrrhophyta  :497
##                        NA's   :28           (Other)      :336
##     biomass
##  Min.   : 0.0000
##  1st Qu.: 0.0120
##  Median : 0.0340
##  Mean   : 0.1969
##  3rd Qu.: 0.1020
##  Max.   :23.4510
##
```

```
head(biomass_nutrients)
```

```
## # A tibble: 6 x 5
## # Groups:   lakename [1]
##   lakename       sampledate tot_biomass medianTN medianTP
##   <chr>          <date>           <dbl>    <dbl>    <dbl>
## 1 East Long Lake 1991-05-22       0.144     426.       13
## 2 East Long Lake 1991-05-29       0.307     557        13
## 3 East Long Lake 1991-06-05       0.251     540        11
## 4 East Long Lake 1991-06-12       0.277     490.       17
## 5 East Long Lake 1991-06-19       0.425     480        18
## 6 East Long Lake 1991-06-26       0.558     496        16
```

```
colnames(biomass_nutrients)
```

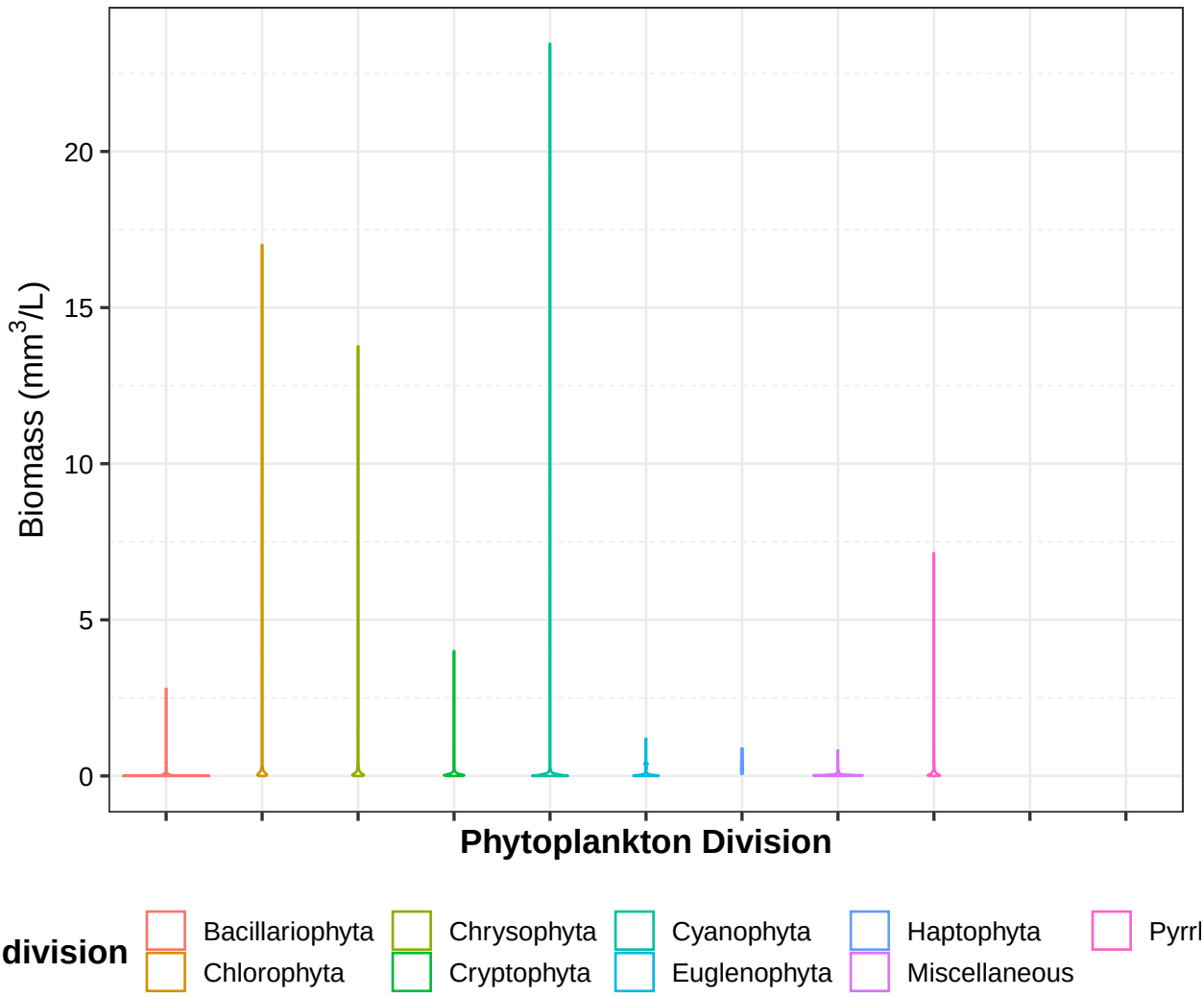
```
## [1] "lakename"    "sampledate"  "tot_biomass" "medianTN"    "medianTP"
```

```r
dim(biomass_nutrients)
```

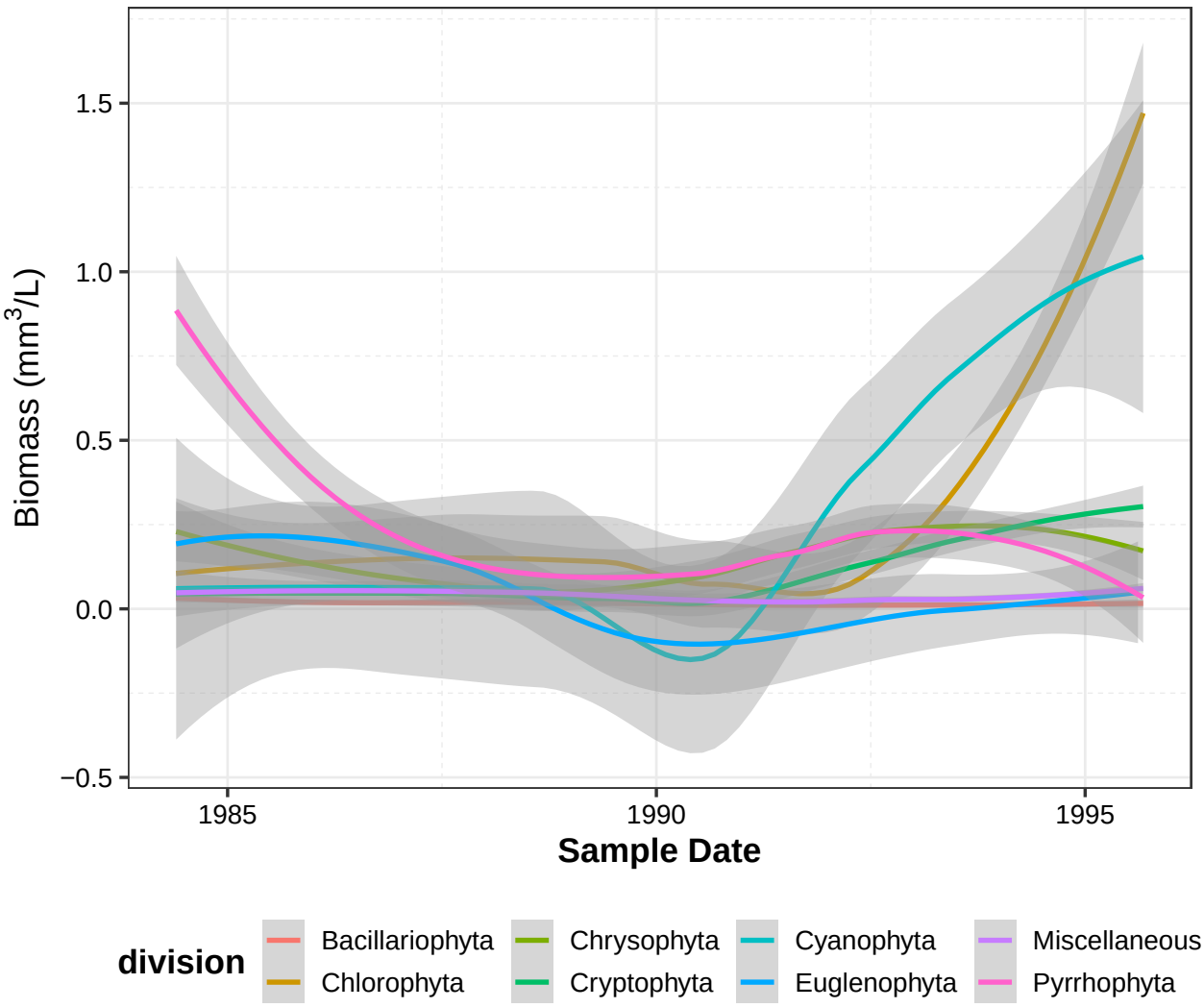
```
## [1] 302   5
```

```r
summary(biomass_nutrients)
```

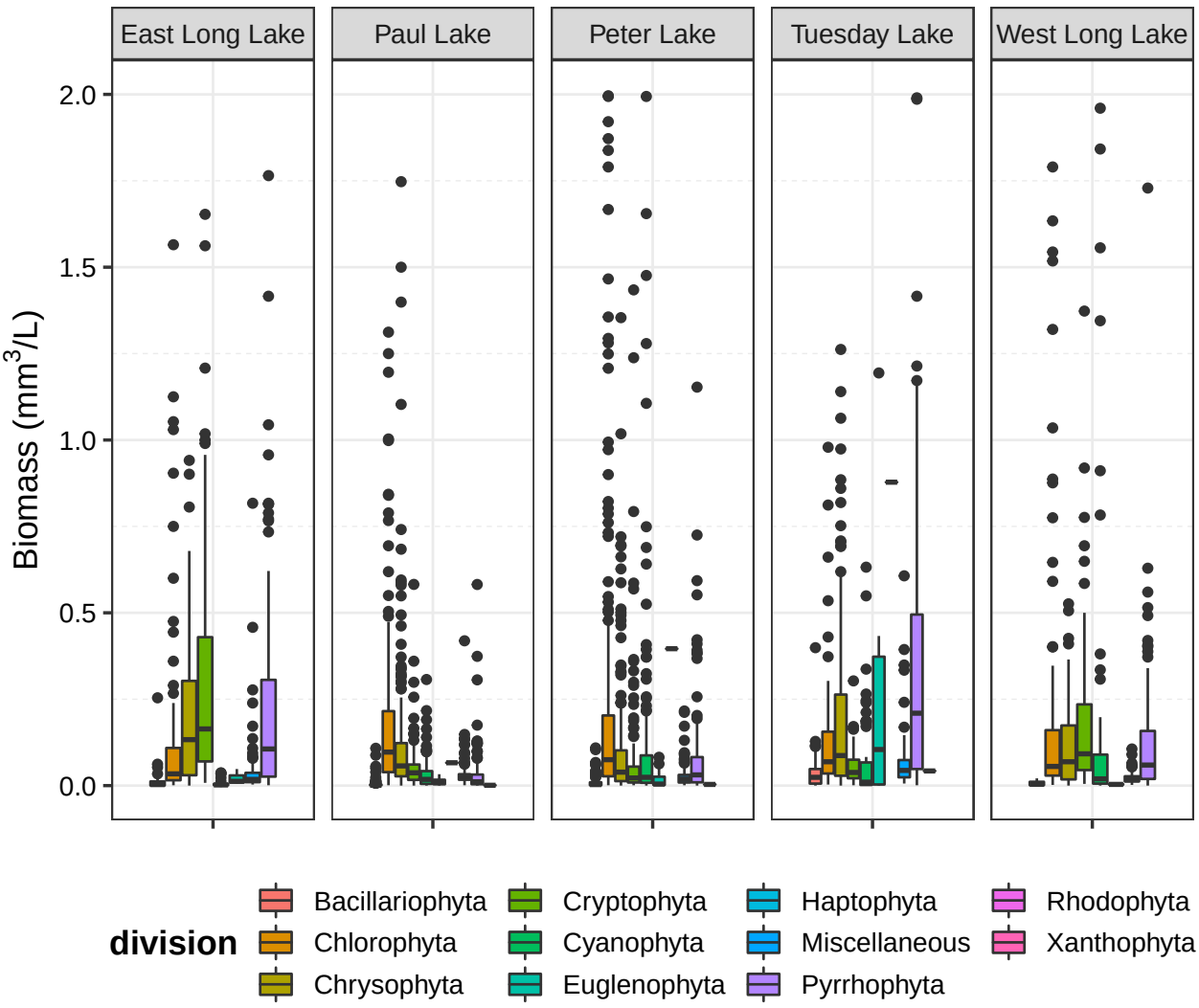
```
##    lakename           sampledate          tot_biomass
##  Length:302         Min.   :1991-05-20   Min.   : 0.0690
##  Class :character   1st Qu.:1992-06-11   1st Qu.: 0.3603
##  Mode  :character   Median :1993-07-17   Median : 0.7075
##                     Mean   :1993-07-13   Mean   : 1.6355
##                     3rd Qu.:1994-08-09   3rd Qu.: 1.6247
##                     Max.   :1995-08-29   Max.   :24.6680
##     medianTN         medianTP
##  Min.   : 163.1   Min.   : 4.501
##  1st Qu.: 411.3   1st Qu.:11.000
##  Median : 494.5   Median :16.000
##  Mean   : 618.8   Mean   :20.143
##  3rd Qu.: 779.5   3rd Qu.:25.920
##  Max.   :2091.7   Max.   :89.501
```




```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 28 rows containing non-finite values (stat_smooth).
```




```
## Warning: Removed 65 rows containing non-finite values (stat_boxplot).
```

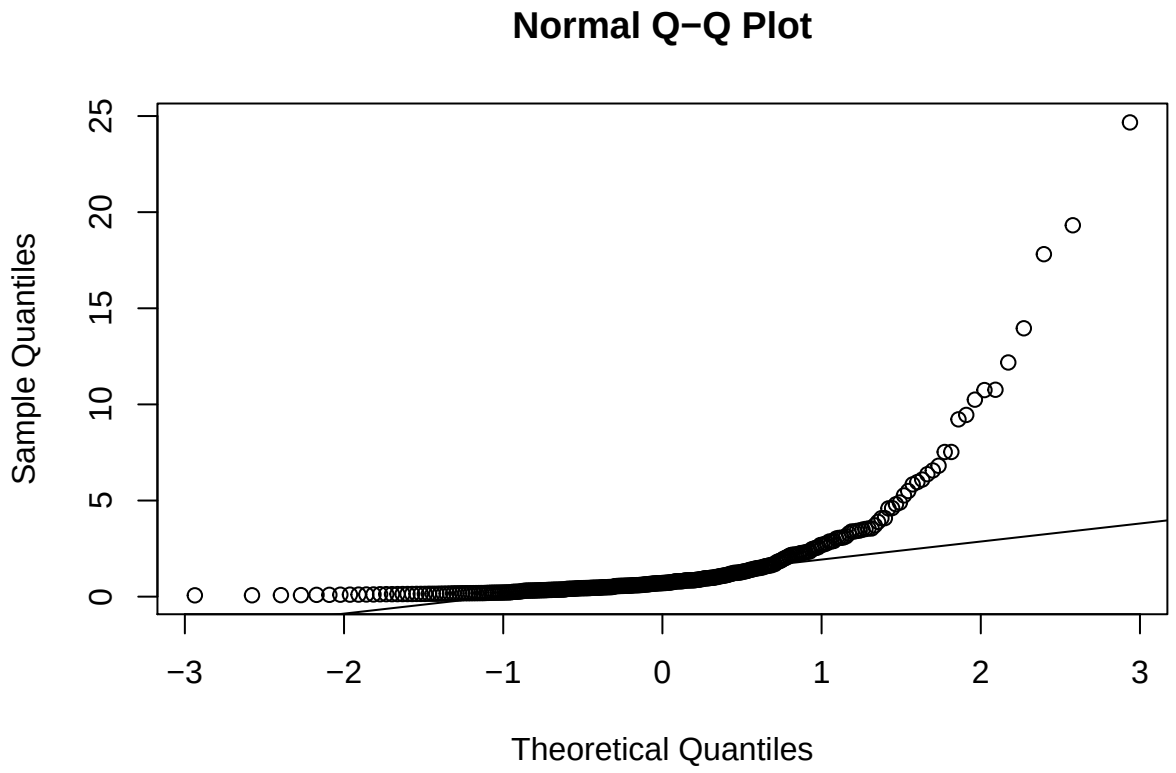

East Long Lake
Paul Lake
Peter Lake
Tuesday Lake
West Long Lake
Biomass (mm3/L)
2.0
1.5
1.0
0.5
0.0
division
Bacillariophyta
Chlorophyta
Chrysophyta
Cryptophyta
Cyanophyta
Euglenophyta
Haptophyta
Miscellaneous
Pyrrhophyta
Rhodophyta
Xanthophyta

4 Analysis

```
# Evaluate assumption of normal distribution
shapiro.test(biomass_nutrients$tot_biomass) # less than 0.05, not normally distributed
```

```
##
##  Shapiro-Wilk normality test
##
## data:  biomass_nutrients$tot_biomass
## W = 0.51532, p-value < 2.2e-16
```

```
qqnorm(biomass_nutrients$tot_biomass); qqline(biomass_nutrients$tot_biomass)
```




```
#
```

5 Summary and Conclusions

Overall,