

# Raw Data

Yuxiang

```
#packages
```

```
library(reshape2)
library(readxl)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(ggplot2)
library(ggpubr)
library(ggpmisc)
```

```
## Loading required package: ggpp
```

```
##
## Attaching package: 'ggpp'

## The following object is masked from 'package:ggplot2':
##
##   annotate
```

```
library(reshape2)
library(readr)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2
## --
```

```
## v tibble  3.1.8    v dplyr   1.1.0
## v tidyr   1.3.0    v stringr 1.5.0
## v purrr   1.0.1    v forcats 1.0.0
## -- Conflicts ----- tidyverse_conflicts() --
## x ggpp::annotate()      masks ggplot2::annotate()
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()      masks base::date()
## x dplyr::filter()        masks stats::filter()
```

```
## x lubridate::intersect() masks base::intersect()
## x dplyr::lag() masks stats::lag()
## x lubridate::setdiff() masks base::setdiff()
## x lubridate::union() masks base::union()
```

```
library(here)
```

```
## here() starts at /Users/min/Desktop/TSA/FinalProject/FinalProject_ENV790
```

```
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:ggpubr':
##
##   get_legend
##
## The following object is masked from 'package:lubridate':
##
##   stamp
```

```
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo
```

```
#install.packages("vegan")
#install.packages("permute")
#install.packages("lattice")
library(permute)
library(lattice)
library(mgcv)
```

```
## Loading required package: nlme
##
## Attaching package: 'nlme'
##
## The following object is masked from 'package:dplyr':
##
##   collapse
##
## This is mgcv 1.8-42. For overview type 'help("mgcv-package")'.
```

```
library(vegan)
```

```
## This is vegan 2.6-4
```

```
#install.packages("relaimpo")  
library(relaimpo)
```

```
## Loading required package: MASS  
##  
## Attaching package: 'MASS'  
##  
## The following object is masked from 'package:dplyr':  
##  
##     select  
##  
## Loading required package: boot  
##  
## Attaching package: 'boot'  
##  
## The following object is masked from 'package:lattice':  
##  
##     melanoma  
##  
## Loading required package: survey  
## Loading required package: grid  
## Loading required package: Matrix  
##  
## Attaching package: 'Matrix'  
##  
## The following objects are masked from 'package:tidyr':  
##  
##     expand, pack, unpack  
##  
## Loading required package: survival  
##  
## Attaching package: 'survival'  
##  
## The following object is masked from 'package:boot':  
##  
##     aml  
##  
## Attaching package: 'survey'  
##  
## The following object is masked from 'package:vegan':  
##  
##     calibrate  
##  
## The following object is masked from 'package:graphics':  
##  
##     dotchart  
##  
## Loading required package: mitools  
## This is the global version of package relaimpo.  
##  
## If you are a non-US user, a version with the interesting additional metric pmvd is available  
##
```

```
## from Ulrike Groempings web site at prof.beuth-hochschule.de/groemping.
```

```
library(zoo)
```

```
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
#install.packages("magrittr")
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##   set_names
##
## The following object is masked from 'package:tidyr':
##
##   extract
```

```
library(knitr)
library(dplyr)
#install.packages("kableExtra")
#install.packages("mgcv")
```

```
#phyto
```

```
phyto <- read.csv( file = "./Data/Phyto_raw.csv")
phyto$sampldate <- ymd(phyto$sampldate)
```

## infomation

```
###site_observation
```

```
## # A tibble: 4 x 2
##   lakeid Observation_Count
##   <chr>             <int>
## 1 ME                 402
## 2 MO                 355
## 3 WI                 23
## 4 FI                 1
```

Table 1: Site Information

Site	Observation Date	Count
ME		402
MO		355
WI		23
FI		1

###Division level Total Biomass

Table 2: Division level Total Biomass (*mg/L*)

Division	Count	Total Biomass	Max Biomass	Min Biomass	Mean Biomass
Cyanophyta	8581	1824.2521267	76.0000000	0.0000000	0.2125920
Bacillariophyta	1914	378.0188786	13.4027978	0.0000319	0.1975020
Chlorophyta	4368	244.9520459	84.6924000	0.0000000	0.0560788
Cryptophyta	1876	76.1100881	2.9630000	0.0000363	0.0405704
Pyrrhophyta	415	29.1295004	5.3194000	0.0000000	0.0701916
Chrysophyta	871	11.3487339	0.5268000	0.0000000	0.0130295
Haptophyta	190	7.4487959	0.7395927	0.0000256	0.0392042
Miscellaneous	387	4.4740805	0.1359000	0.0000000	0.0115609
Euglenophyta	44	1.1695524	0.3695402	0.0002000	0.0265807
Xanthophyta	7	0.0690431	0.0517453	0.0005859	0.0098633
Rhodophyta	1	0.0309404	0.0309404	0.0309404	0.0309404

###Genus level Total Biomass

Table 3: Genus level Total Biomass (*mg/L*)

Division	Genus	Count	Total Biomass	Max Biomass	Min Biomass	Mean Biomass
Cyanophyta	Aphanizomenon	812	923.88321	76.000000	0.0004000	1.1377872
Cyanophyta	Microcystis	1065	436.73355	23.370568	0.0000284	0.4100785
Bacillariophyta	Stephanodiscus	501	259.59263	13.402798	0.0001000	0.5181490
Chlorophyta	Coelastrum	42	85.44968	84.692400	0.0004569	2.0345162
Chlorophyta	Actinastrum	53	84.45011	84.070000	0.0001421	1.5933983
Cyanophyta	Anabaena	525	79.18206	5.336000	0.0003000	0.1508230
Cyanophyta	Planktothrix	96	70.65535	8.519000	0.0004000	0.7359932
Cyanophyta	Lyngbya	247	50.47694	9.042087	0.0000570	0.2043601
Cryptophyta	Cryptomonas	880	47.31605	2.963000	0.0000363	0.0537682
Bacillariophyta	Aulacoseira	180	44.14470	4.202500	0.0005140	0.2452483

## Data

```
biomass_D <- rawdata[,c(1,3,5,6,15,13)]%>%
  group_by(lakeid, sampledate,depth_range, division) %>% summarize(total_biomass = sum(biomass_conc))
```

## 'summarise()' has grouped output by 'lakeid', 'sampledate', 'depth\_range'. You  
## can override using the '.groups' argument.

```
biomass_G <- rawdata[,c(1,3,5,6,15,13)] %>% filter( division == "Cyanophyta") %>%
  group_by(lakeid, sampleddate, depth_range, genus) %>% summarize(total_biomass = sum(biomass_conc))
```

## 'summarise()' has grouped output by 'lakeid', 'sampledate', 'depth\_range'. You  
## can override using the '.groups' argument.

```
biomass_cyan <- biomass_D %>% filter(division == "Cyanophyta")
#ME lake
cyan_ME <- filter(biomass_cyan, lakeid == "ME")
unique(cyan_ME$depth_range)
```

```
## [1] "0-8m" "0-2m"
```

```
common_dates <- cyan_ME[, c(2,3)] %>%
  filter(depth_range == "0-8m") %>%
  inner_join(cyan_ME[, c(2,3)] %>% filter(depth_range == "0-2m"), by = "sampledate")

common_dates
```

```
## # A tibble: 0 x 3
## # Groups:   sampledate [0]
## # ... with 3 variables: sampledate <date>, depth_range.x <chr>,
## #   depth_range.y <chr>
```

*## 2 depth range in different day*

#Physical

```
physical <- read.csv( file = "../Data/Physical_raw.csv")
physical$sampledate <- ymd(physical$sampledate)
```

###site\_observation

```
## # A tibble: 11 x 2
##   lakeid Observation_Count
##   <chr>                <int>
## 1 CR                    725
## 2 SP                    698
## 3 TR                    698
## 4 BM                    695
## 5 AL                    692
## 6 TB                    691
## 7 CB                    685
## 8 ME                    427
## 9 MO                    388
## 10 WI                   375
## 11 FI                   357
```

Table 4: Site Information

Site	Observation Date Count
CR	725
SP	698
TR	698
BM	695
AL	692
TB	691
CB	685
ME	427
MO	388
WI	375
FI	357

```
##Data
```

```
physical_ME_08 <- rawdata[,c(1,4,5,7,8,9,10,11)]%>%
  filter(depth %in% c(0:8)) %>%
  filter(lakeid == "ME") %>% group_by(sampledate) %>% summarize(Temperature = mean(wtemp), o2 = mean(o2))

physical_ME_02 <- rawdata[,c(1,4,5,7,8,9,10,11)]%>%
  filter(depth %in% c(0:2)) %>%
  filter(lakeid == "ME") %>% group_by(sampledate) %>% summarize(Temperature = mean(wtemp), o2 = mean(o2))

physical_ME_combined <- rbind(physical_ME_08, physical_ME_02)

cyan_physical_ME <- cyan_ME %>%
  left_join(physical_ME_combined, by = c("sampledate", "depth_range"))
```

```
#Chemical
```

```
chemical <- read.csv( file = "./Data/Chemical_raw.csv")
chemical$sampledate <- ymd(chemical$sampledate)
```

```
##Data wangling
```

```
#08
rawdata <- chemical
chemical_ME_08_totpuf_sloh <- rawdata%>%
  filter(depth %in% c(0:8)) %>%
  filter(lakeid == "ME") %>%
  filter(!is.na(totpuf_sloh)) %>% #
  group_by(sampledate) %>%
  summarize(totpuf_sloh = mean(totpuf_sloh , na.rm = TRUE)) %>% mutate(depth_range = "0-8m")

chemical_ME_08_no3no2_sloh <- rawdata%>%
  filter(depth %in% c(0:8)) %>%
  filter(lakeid == "ME") %>%
  filter(!is.na(no3no2_sloh )) %>% #
  group_by(sampledate) %>%
```

```

    summarize(no3no2_sloh = mean(no3no2_sloh, na.rm = TRUE)) %>% mutate(depth_range = "0-8m")

chemical_ME_08_nh4_sloh <- rawdata%>%
  filter(depth %in% c(0:8)) %>%
  filter(lakeid == "ME") %>%
  filter(!is.na(nh4_sloh)) %>% #
  group_by(sampledate) %>%
  summarize(nh4_sloh = mean(nh4_sloh, na.rm = TRUE)) %>% mutate(depth_range = "0-8m")

#02
chemical_ME_02_totpuf_sloh <- rawdata%>%
  filter(depth %in% c(0:2)) %>%
  filter(lakeid == "ME") %>%
  filter(!is.na(totpuf_sloh)) %>% #
  group_by(sampledate) %>%
  summarize(totpuf_sloh = mean(totpuf_sloh, na.rm = TRUE)) %>% mutate(depth_range = "0-2m")

chemical_ME_02_no3no2_sloh <- rawdata%>%
  filter(depth %in% c(0:2)) %>%
  filter(lakeid == "ME") %>%
  filter(!is.na(no3no2_sloh)) %>% #
  group_by(sampledate) %>%
  summarize(no3no2_sloh = mean(no3no2_sloh, na.rm = TRUE)) %>% mutate(depth_range = "0-2m")

chemical_ME_02_nh4_sloh <- rawdata%>%
  filter(depth %in% c(0:2)) %>%
  filter(lakeid == "ME") %>%
  filter(!is.na(nh4_sloh)) %>% #
  group_by(sampledate) %>%
  summarize(nh4_sloh = mean(nh4_sloh, na.rm = TRUE)) %>% mutate(depth_range = "0-2m")

#totpuf_sloh
chemical_ME_combined_totpuf_sloh <- rbind(chemical_ME_02_totpuf_sloh, chemical_ME_08_totpuf_sloh)
ME_raw1 <- cyan_physical_ME %>%
  left_join(chemical_ME_combined_totpuf_sloh, by = c("sampledate", "depth_range"))

#totnuf_sloh
chemical_ME_combined_nh4_sloh <- rbind(chemical_ME_02_nh4_sloh, chemical_ME_08_nh4_sloh)
ME_raw2 <- ME_raw1 %>%
  left_join(chemical_ME_combined_nh4_sloh, by = c("sampledate", "depth_range"))

chemical_ME_combined_no3no2_sloh <- rbind(chemical_ME_02_no3no2_sloh, chemical_ME_08_no3no2_sloh)
ME_raw3 <- ME_raw2 %>%
  left_join(chemical_ME_combined_no3no2_sloh, by = c("sampledate", "depth_range"))

#final raw data

ME_rawME_dayI <- ME_raw3 %>%
  ungroup() %>%
  arrange(sampledate) %>%

```



```
mutate(date_diff = as.numeric(difftime(sampledate, lag(sampledate), units = "days"))) %>% mutate(TN =
Table_ME_rawME_dayI_10 <- ME_rawME_dayI[c(1:10),c(1,2,5,6,12,13,14)]
kable(Table_ME_rawME_dayI_10,
caption = "rawdata")
```

Table 5: rawdata

lakeid	sampledate	total_biomass	Temperature	date_diff	TN	TP
ME	1995-01-24	0.0128965	NA	NA	NA	NA
ME	1995-03-28	0.0013183	NA	63	NA	NA
ME	1995-04-11	0.0017578	NA	14	NA	NA
ME	1995-04-24	0.0019157	NA	13	NA	NA
ME	1995-05-23	0.8052959	13.64444	29	0.7305	0.0895000
ME	1995-06-06	0.0738443	16.84444	14	0.7195	0.0756667
ME	1995-06-20	0.9552335	NA	14	NA	NA
ME	1995-07-06	9.4816674	21.48889	16	0.1940	0.0423333
ME	1995-07-18	2.9985433	25.05556	12	0.0075	0.0295000
ME	1995-08-02	1.6772605	24.80000	15	NA	0.0310000

#save final raw data

```
save(ME_rawME_dayI, file="ME_rawME_dayI.RDATA")
```

```
rawdata_all <- ME_rawME_dayI %>%
  mutate( date = ymd(sampledate)) %>%
  mutate( Year = year(date),
          Month = month(date),
          Day = day(date)) %>%
  group_by( Year, Month ) %>%
  summarise('Biomass' = mean(total_biomass), 'TN' = mean(TN), 'TP' = mean(TP), 'Temperature' = mean(Temperature))
```

## 'summarise()' has grouped output by 'Year'. You can override using the  
## '.groups' argument.

```
start_date <- ym("1995-01")
end_date <- ym("2020-12")

monthly_dates <- seq(from = start_date, to = end_date, by = "month") %>%
  as.data.frame()
colnames(monthly_dates) <- c("date")

rawdata_all <- right_join(rawdata_all, monthly_dates, by = 'date' )

#fill the NA
# due to no T, TN, TP before 1995-5-1, this all data also start at 1995-05-01
rawdata_all_fill <- rawdata_all[,c(7,6,4,5,3)] %>% arrange(date) %>% filter(!is.na(date)) %>%
  filter(!date <= "1995-04-01" ) %>% mutate('Biomass'=na.approx(Biomass),rule=2)%>%
```

```

mutate(Temperature =na.approx(Temperature),rule=2)%>%
mutate(TN=na.approx(TN),rule=2)%>%
mutate(TP=na.approx(TP),rule=2)

rawdata_all_fill_10 <- rawdata_all_fill[c(1:10),c(1:5)]
kable(rawdata_all_fill_10,
      caption = "rawdata_1")

```

Table 6: rawdata\_1

date	Temperature	TN	TP	Biomass
1995-05-01	13.644444	0.7305000	0.0895000	0.8052959
1995-06-01	18.458333	0.4156250	0.0627083	0.5145389
1995-07-01	23.272222	0.1007500	0.0359167	6.2401053
1995-08-01	25.318518	0.1928333	0.0252778	4.6301317
1995-09-01	20.220370	0.2849167	0.0681389	0.7740042
1995-10-01	15.122222	0.3770000	0.1110000	2.1857864
1995-11-01	8.711111	0.5245000	0.1170000	0.6125286
1995-12-01	2.300000	0.6720000	0.1230000	0.7866917
1996-01-01	3.702222	0.7277500	0.1163333	0.4872730
1996-02-01	5.104444	0.7835000	0.1096667	0.1878542