Assignment 5: Water Quality in Lakes

Keith Bollt

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis on water quality in lakes

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single HTML file.
5. After Knitting, submit the completed exercise (HTML file) to the dropbox in Sakai. Add your last name into the file name (e.g., "A05_Salk.html") prior to submission.

The completed exercise is due on 2 October 2019 at 9:00 am.

Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, lubridate, and LAGOSNE packages.
3. Set your ggplot theme (can be theme_classic or something else)
4. Load the LAGOSdata database and the trophic state index csv file we created on 2019/09/27.

```
getwd()
```

```
## [1] "Z:/Hydrologic_Data_Analysis2/Assignments"
```

```
library(tidyverse)
library(lubridate)
library(LAGOSNE)

theme_set(theme_classic())
options(scipen = 100)

#lagosne_get(dest_folder = LAGOSNE:::lagos_path(), overwrite = TRUE)
LAGOSdata <- lagosne_load()
```

Trophic State Index

5. Similar to the trophic.class column we created in class (determined from TSI.chl values), create two additional columns in the data frame that determine trophic class from TSI.secchi and TSI.tp (call these trophic.class.secchi and trophic.class.tp).

```
LAGOSlocus <- LAGOSdata$locus
LAGOSstate <- LAGOSdata$state
LAGOSnutrient <- LAGOSdata$epi_nutr

LAGOSlocus$lagoslakeid <- as.factor(LAGOSlocus$lagoslakeid)
LAGOSnutrient$lagoslakeid <- as.factor(LAGOSnutrient$lagoslakeid)

LAGOSlocations <- left_join(LAGOSlocus, LAGOSstate, by = "state_zoneid")
```

```
LAGOSlocations <-
  within(LAGOSlocations,
         state <- factor(state, levels = names(sort(table(state), decreasing=TRUE))))
LAGOStrophic <-
  left_join(LAGOSnutrient, LAGOSlocations, by = "lagoslakeid") %>%
  select(lagoslakeid, sampledate, chla, tp, secchi,
         gnis_name, lake_area_ha, state, state_name) %>%
  mutate(sampleyear = year(sampledate),
         samplemonth = month(sampledate),
         season = as.factor(quarter(sampledate, fiscal_start = 12))) %>%
  drop_na(chla:secchi)

levels(LAGOStrophic$season) <- c("Winter", "Spring", "Summer", "Fall")

LAGOStrophic <-
  mutate(LAGOStrophic,
         TSI.chl = round(10*(6 - (2.04 - 0.68*log(chla)/log(2)))),
         TSI.secchi = round(10*(6 - (log(secchi)/log(2)))),   #In R, log is the natural log.
         TSI.tp = round(10*(6 - (log(48/tp)/log(2)))),
         trophic.class =
            ifelse(TSI.chl < 40, "Oligotrophic",
                   ifelse(TSI.chl < 50, "Mesotrophic",
                          ifelse(TSI.chl < 70, "Eutrophic", "Hypereutrophic"))),
         trophic.class.secchi =
            ifelse(TSI.secchi < 40, "Oligotrophic",
                   ifelse(TSI.secchi < 50, "Mesotrophic",
                          ifelse(TSI.secchi < 70, "Eutrophic", "Hypereutrophic"))),
         trophic.class.tp =
            ifelse(TSI.tp < 40, "Oligotrophic",
                   ifelse(TSI.tp < 50, "Mesotrophic",
                          ifelse(TSI.tp < 70, "Eutrophic", "Hypereutrophic"))))
```

6. How many observations fall into the four trophic state categories for the three metrics (trophic.class, trophic.class.secchi, trophic.class.tp)? Hint: `count` function.

```
chlcount <-
  LAGOStrophic %>%
  count(trophic.class)
print(chlcount)
```

```
## # A tibble: 4 x 2
##   trophic.class      n
##   <chr>          <int>
## 1 Eutrophic      41861
## 2 Hypereutrophic 14379
## 3 Mesotrophic    15413
## 4 Oligotrophic    3298
```

```
secchicount <-
  LAGOStrophic %>%
  count(trophic.class.secchi)
print(secchicount)
```

```
## # A tibble: 4 x 2
##   trophic.class.secchi     n
```

```
##   <chr>                 <int>
## 1 Eutrophic             28659
## 2 Hypereutrophic         5099
## 3 Mesotrophic           25083
## 4 Oligotrophic          16110
```

```r
tpcount <-
  LAGOStrophic %>%
  count(trophic.class.tp)
print(tpcount)
```

```
## # A tibble: 4 x 2
##   trophic.class.tp     n
##   <chr>            <int>
## 1 Eutrophic        24839
## 2 Hypereutrophic    7228
## 3 Mesotrophic      23023
## 4 Oligotrophic     19861
```

7. What proportion of total observations are considered eutrohic or hypereutrophic according to the three different metrics (trophic.class, trophic.class.secchi, trophic.class.tp)?

```r
chl.proportion <-
  sum(LAGOStrophic$trophic.class == 'Eutrophic' | LAGOStrophic$trophic.class == 'Hypereutrophic') /
  sum(LAGOStrophic$trophic.class == 'Eutrophic' | LAGOStrophic$trophic.class == 'Hypereutrophic' |
      LAGOStrophic$trophic.class == 'Mesotrophic' | LAGOStrophic$trophic.class == 'Oligotrophic')
print(chl.proportion)
```

```
## [1] 0.7503569
```

```r
secchi.proportion <-
  sum(LAGOStrophic$trophic.class.secchi == 'Eutrophic' | LAGOStrophic$trophic.class.secchi == 'Hypereut
  sum(LAGOStrophic$trophic.class.secchi == 'Eutrophic' | LAGOStrophic$trophic.class.secchi == 'Hypereut
      LAGOStrophic$trophic.class.secchi == 'Mesotrophic' | LAGOStrophic$trophic.class.secchi == 'Oligot
print(secchi.proportion)
```

```
## [1] 0.4504009
```

```r
tp.proportion <-
  sum(LAGOStrophic$trophic.class.tp == 'Eutrophic' | LAGOStrophic$trophic.class.tp == 'Hypereutrophic')
  sum(LAGOStrophic$trophic.class.tp == 'Eutrophic' | LAGOStrophic$trophic.class.tp == 'Hypereutrophic'
      LAGOStrophic$trophic.class.tp == 'Mesotrophic' | LAGOStrophic$trophic.class.tp == 'Oligotrophic')
print(tp.proportion)
```

```
## [1] 0.4278395
```

```r
#Note to self for 10.1.19: repeat for the tp and secchi dataframes
```

Which of these metrics is most conservative in its designation of eutrophic conditions? Why might this be?

> Total phosphorus. Using this metric as a proxy for eutrophic conditions assumes that phosphorus is the limiting nutrient. While this is true in summer, this dataset contains yearround observations. For this reason, we would expect total phosphorus to underestimate the true amount of biomass in our lakes and therefore its true trophic state.

Note: To take this further, a researcher might determine which trophic classes are susceptible to being differently categorized by the different metrics and whether certain metrics are prone to categorizing trophic class as more or less eutrophic. This would entail more complex code.

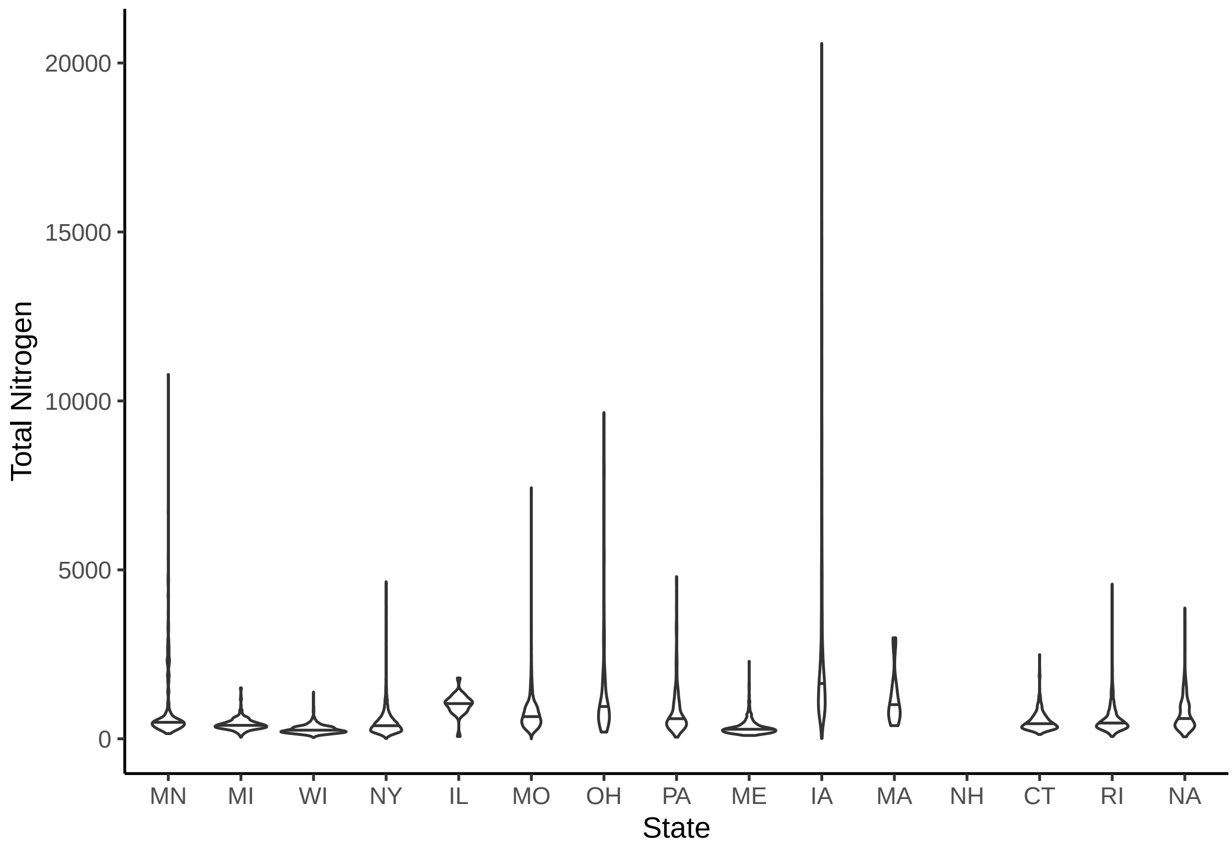
Nutrient Concentrations

8. Create a data frame that includes the columns lagoslakeid, sampledate, tn, tp, state, and state_name. Mutate this data frame to include sampleyear and samplemonth columns as well. Call this data frame LAGOSNandP.

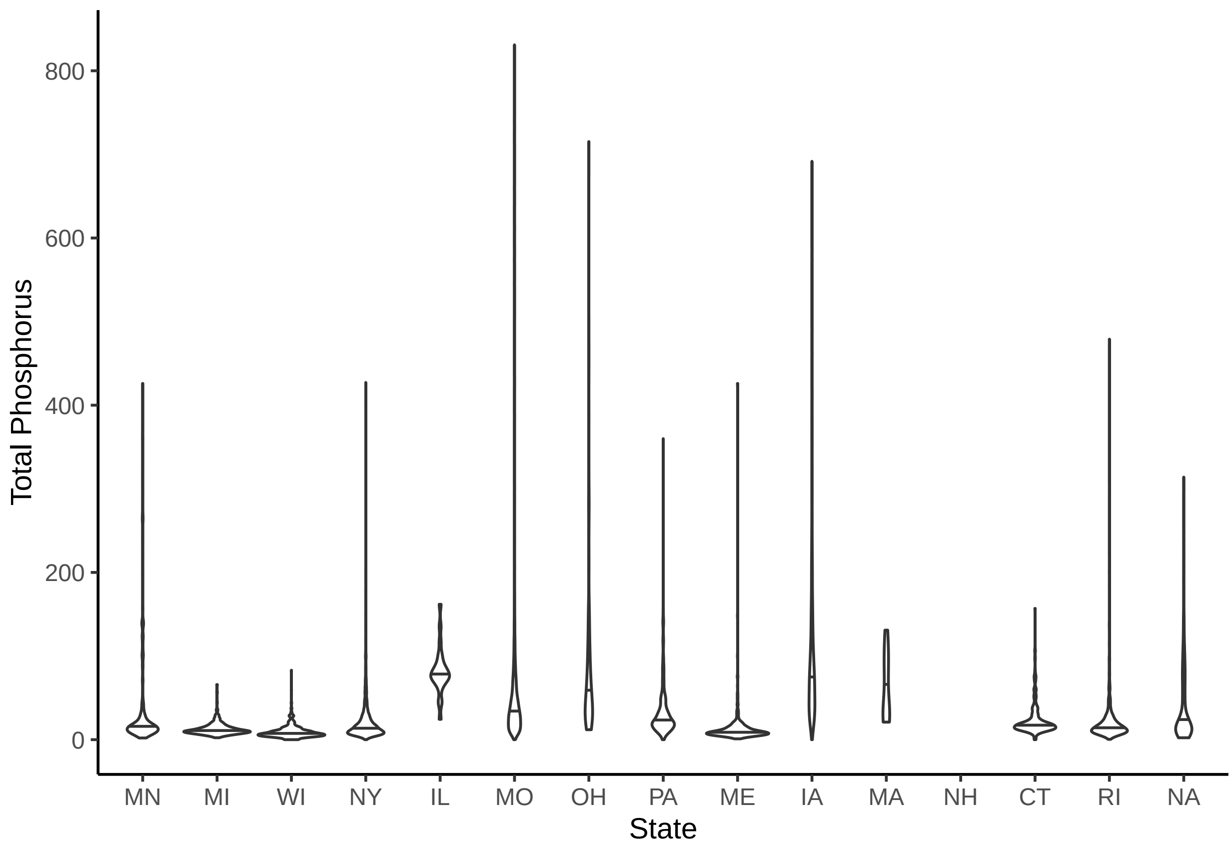
```
LAGOSNandP <-
  left_join(LAGOSnutrient, LAGOSlocations, by = "lagoslakeid") %>%
  select(lagoslakeid, sampledate, chla, tn, tp, secchi,
         gnis_name, lake_area_ha, state, state_name) %>%
  mutate(sampleyear = year(sampledate),
         samplemonth = month(sampledate)) %>%
  drop_na(chla:secchi)
```

9. Create two violin plots comparing TN and TP concentrations across states. Include a 50th percentile line inside the violins.

```
TN_violin <- ggplot(LAGOSNandP, aes(x = state, y = tn)) +
  geom_violin(draw_quantiles = 0.50)+
  labs(x = "State", y = "Total Nitrogen")
print(TN_violin)
```




```
TP_violin <- ggplot(LAGOSNandP, aes(x = state, y = tp)) +
  geom_violin(draw_quantiles = 0.50)+
  labs(x = "State", y = "Total Phosphorus")
print(TP_violin)
```



Which states have the highest and lowest median concentrations?

TN:

TP:

Which states have the highest and lowest concentration ranges?

TN:

TP:

10. Create two jitter plots comparing TN and TP concentrations across states, with samplemonth as the color. Choose a color palette other than the ggplot default.

Which states have the most samples? How might this have impacted total ranges from #9?

TN:

TP:

Which months are sampled most extensively? Does this differ among states?

TN:

TP:

11. Create two jitter plots comparing TN and TP concentrations across states, with sampleyear as the color. Choose a color palette other than the ggplot default.

Which years are sampled most extensively? Does this differ among states?

TN:

TP:

Reflection

12. What are 2-3 conclusions or summary points about lake water quality you learned through your analysis?

13. What data, visualizations, and/or models supported your conclusions from 12?

14. Did hands-on data analysis impact your learning about water quality relative to a theory-based lesson? If so, how?

15. How did the real-world data compare with your expectations from theory?