Assignment 2: Physical Properties of Lakes

Jackie Van Der Hout

OVERVIEW

This exercise accompanies the lessons in Water Data Analytics on the physical properties of 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 PDF file.
5. After completing your assignment, fill out the assignment completion survey in Sakai.

Having trouble? See the assignment's answer key if you need a hint. Please try to complete the assignment without the key as much as possible - this is where the learning happens!

Target due date: 2022-01-25

Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, lubridate, and rLakeAnalyzer packages
3. Import the NTL-LTER physical lake dataset and set the date column to the date format.
4. Using the `mutate` function, add a column called Month. Remove temperature NAs.
5. Set your ggplot theme (can be theme_classic or something else)

```
getwd()
```

```
## [1] "C:/Users/Jackie/Box/Classes Spring 2022/Water Data Analytics/Water_Data_Analytics_2022/Assignmen
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

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

```
library(rLakeAnalyzer)
NTLdata <- read.csv("~/../Box/Classes Spring 2022/Water Data Analytics/Water_Data_Analytics_2022/Data/Ra
NTLdata$sampledate <- as.Date(NTLdata$sampledate, format = "%m/%d/%y")
class(NTLdata$sampledate) #checking it is read as a date
```

```
## [1] "Date"
```

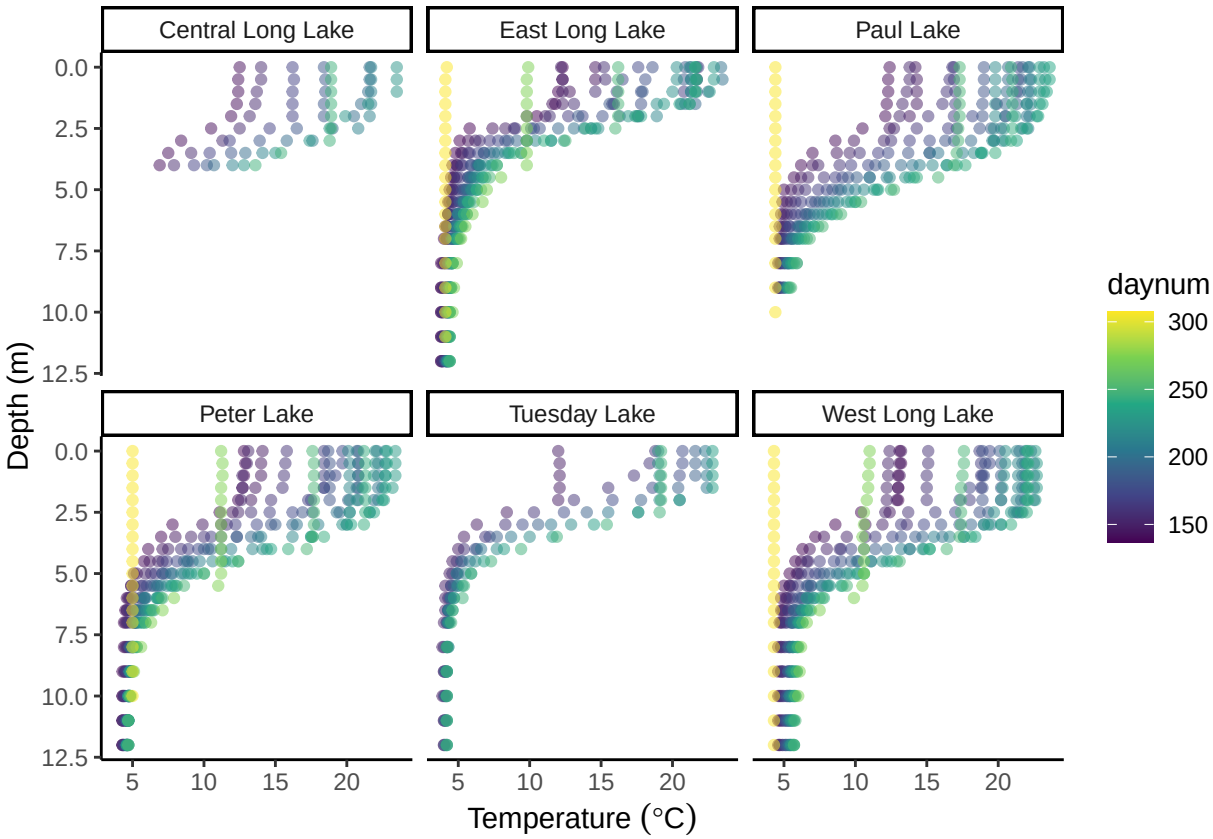
```
theme_set(theme_classic())
```

Creating and analyzing lake temperature profiles

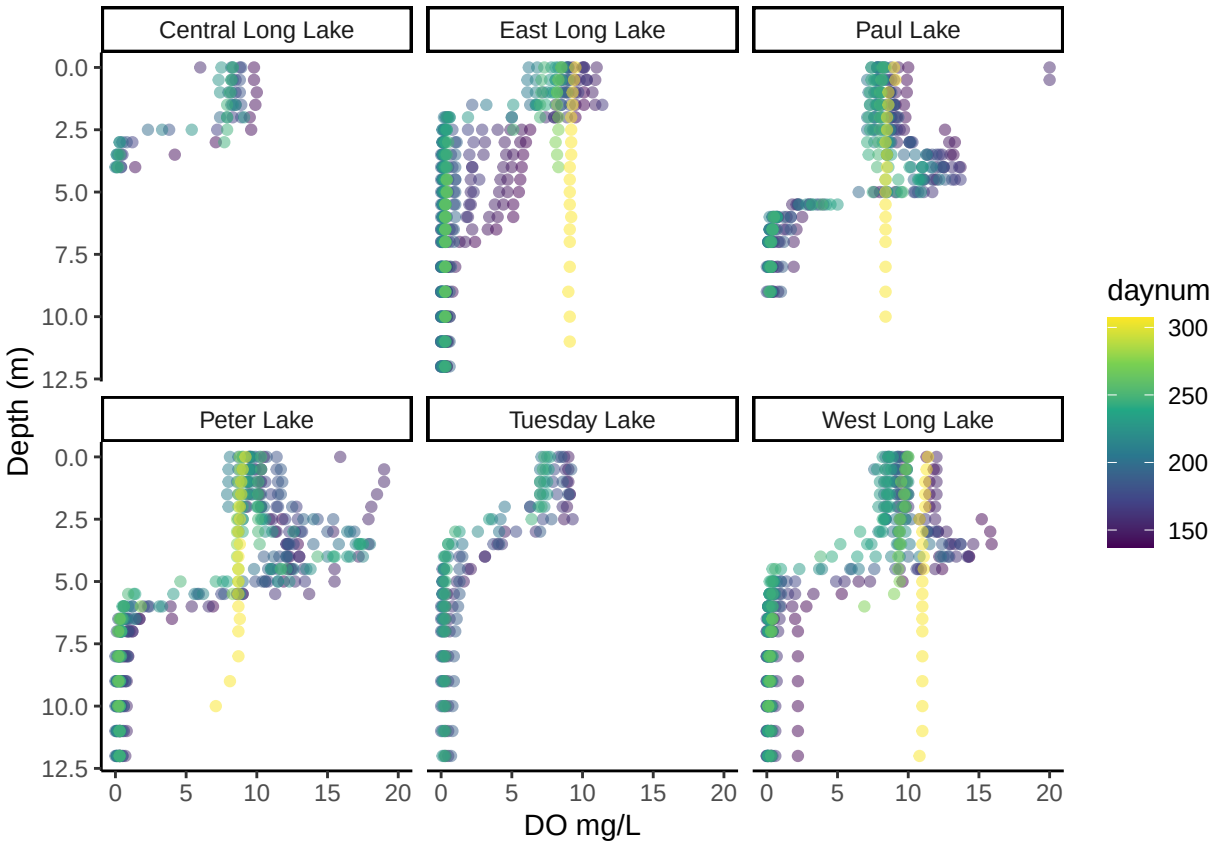
5. For the year 1993, plot temperature and dissolved oxygen profiles for all six lakes in the dataset (as two separate ggplots). Use the `facet_wrap` function to plot each lake as a separate panel in the plot. Plot day of year as your color aesthetic and use a reverse y scale to represent depth.

What seasonal trends do you observe, and do these manifest differently in each lake?

```
NTLdata.no.na.temp <- NTLdata %>% #getting rid of NAs
  drop_na(temperature_C)

ggplot(subset (NTLdata.no.na.temp, year4 == 1993), aes(x = temperature_C, y= depth, color = daynum))+
  geom_point(alpha = 0.5)+
  facet_wrap(~lakename)+
  scale_color_viridis_c()+
  scale_y_reverse()+
  labs(x = expression("Temperature " (degree*C)), y = "Depth (m)" )
```




```
#Temperature Seasonal Trends: Early in the year, temperature is broadly less stratified across the six

NTLdata.no.na.DO <- NTLdata %>% #getting rid of NAs in DO
  drop_na(dissolvedOxygen)

ggplot(subset (NTLdata.no.na.DO, year4 == 1993), aes(x = dissolvedOxygen, y= depth, color = daynum))+
  geom_point(alpha = 0.5)+
  facet_wrap(~lakename)+
  scale_color_viridis_c()+
  scale_y_reverse()+
  labs(x = "DO mg/L", y = "Depth (m)")
```




```
#Dissolved Oxygen seasonal trends that can be seen in this visualization are slightly more complex to d
```

6. Create a new dataset that calculates thermocline depths for all lakes on all dates (hint: you will need group by lake, year, month, DOY, and sample date).

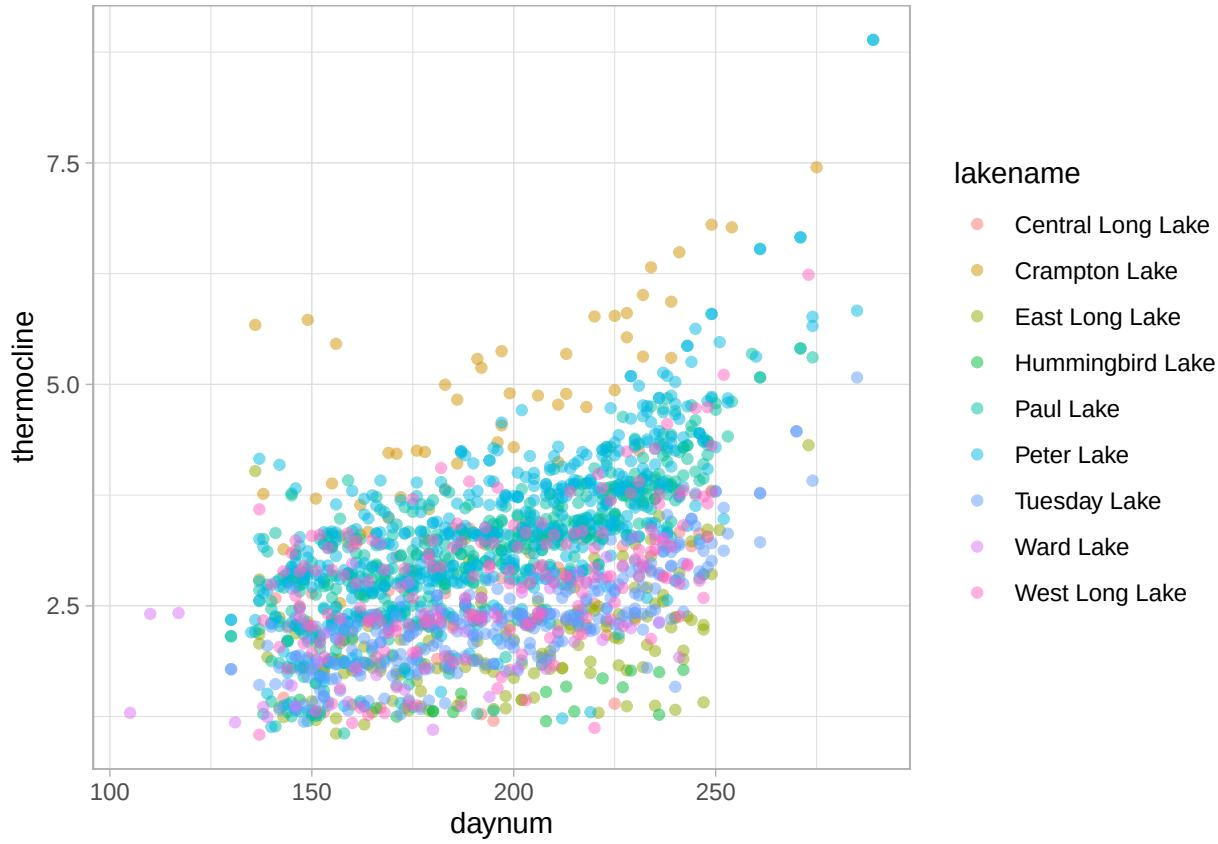
```
NTLdata.no.na.temp$month <- strftime(NTLdata.no.na.temp$sampledate,"%m") #make new column with month
View(NTLdata) #check that new column was created

#needed to remove NA values from temp in order for pipe to work with thermocline

thermocline.data <- NTLdata.no.na.temp %>%
  group_by(lakename, year4, month, daynum, sampledate) %>%
  summarise(thermocline = thermo.depth(wtr= temperature_C,   depths = depth, seasonal = FALSE)) %>%
  filter(thermocline > 1) #ran without thermocline filter and got more lakes
```

```
## `summarise()` has grouped output by 'lakename', 'year4', 'month', 'daynum'. You can override using tl
```

```
#had to use answer sheet for this part
```

7. Plot thermocline depth by day of year for your newly made dataset. Color each point by lake name, make the points 50% transparent, and choose a color palette other than the ggplot default.

```
ggplot(thermocline.data, aes(x = daynum, y = thermocline, color = lakename))+
  geom_point(alpha = 0.5)+
  theme_light()
```

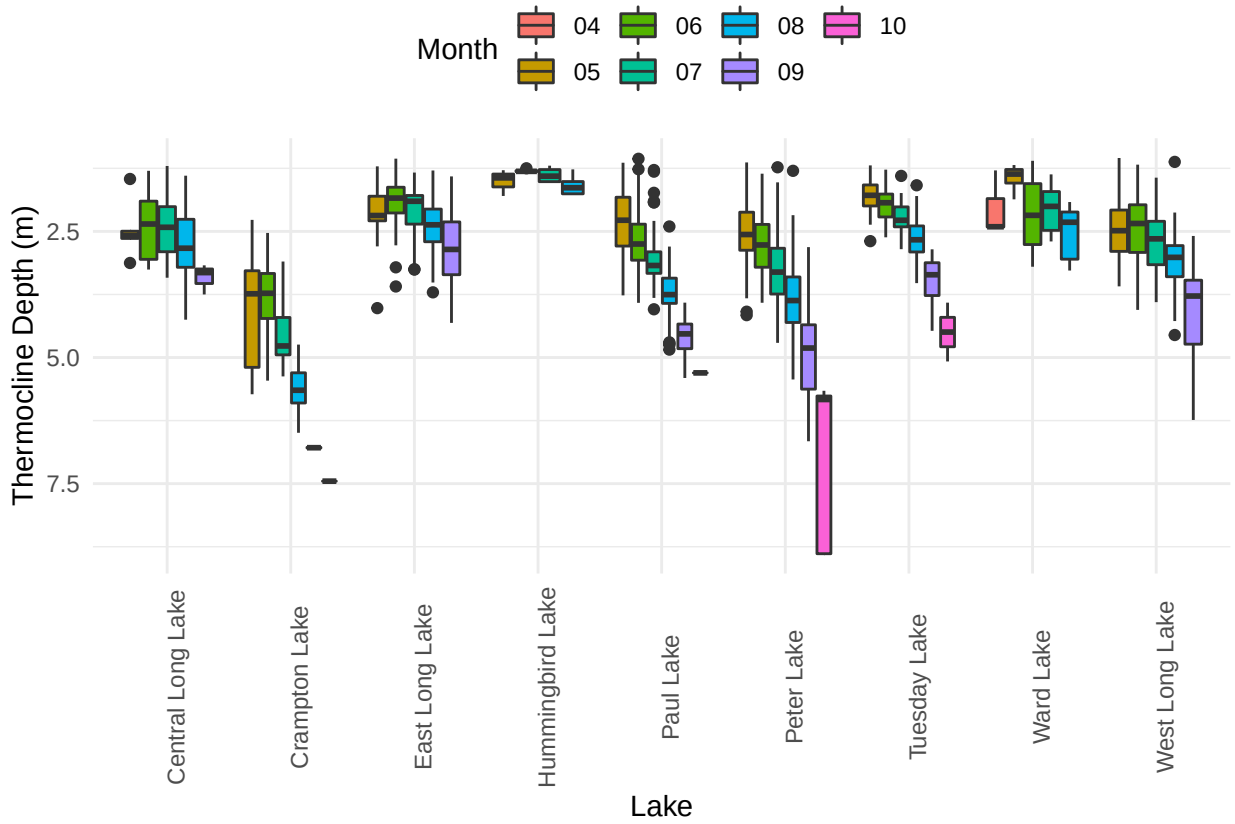


8. Create a boxplot of thermocline depth distributions split up by lake name on the x axis and by month as the fill color (hint: you will need to set Month as a factor). Choose a color palette other than the ggplot default, relabel axes and legend, and place the legend on the top of the graph.

```
class(thermocline.data$month)
```

```
## [1] "character"
```

```
ggplot(thermocline.data, aes(x = lakename, y= thermocline, fill = as.factor(month)))+
  geom_boxplot()+
  theme_minimal()+
  labs(x = "Lake", y ="Thermocline Depth (m)", fill = "Month")+
  theme(legend.position = "top")+
  theme(axis.text.x = element_text(angle = 90))+
  scale_y_reverse()
```



Do all the lakes have a similar seasonal progression of thermocline deepening? Which lakes have the deepest thermoclines, and how does this relate to their maximum depth?

The lakes seem to for the most part have a similar progression in their deepening of the thermocline, but the deeper lakes seem to be grouped into one progression and the shallower lakes into another. Peter lake is by far the deepest lake and has the deepest progression of it's thermocline in October, along with Tuesday Lake. It appears that the deeper the lakes are, the later on in the year the thermocline will persist and the slower temperatures will shift.