Assignment 2: Physical Properties of Lakes

Lydie Costes

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)

```
# 1. Get working directory
getwd()
```

```
## [1] "/Users/lydiecostes/Documents/Duke/WaterDataAnalytics/Water_Data_Analytics_2022/Assignments"
```

```
# 2. Load packages
library(tidyverse)
```

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     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'
```

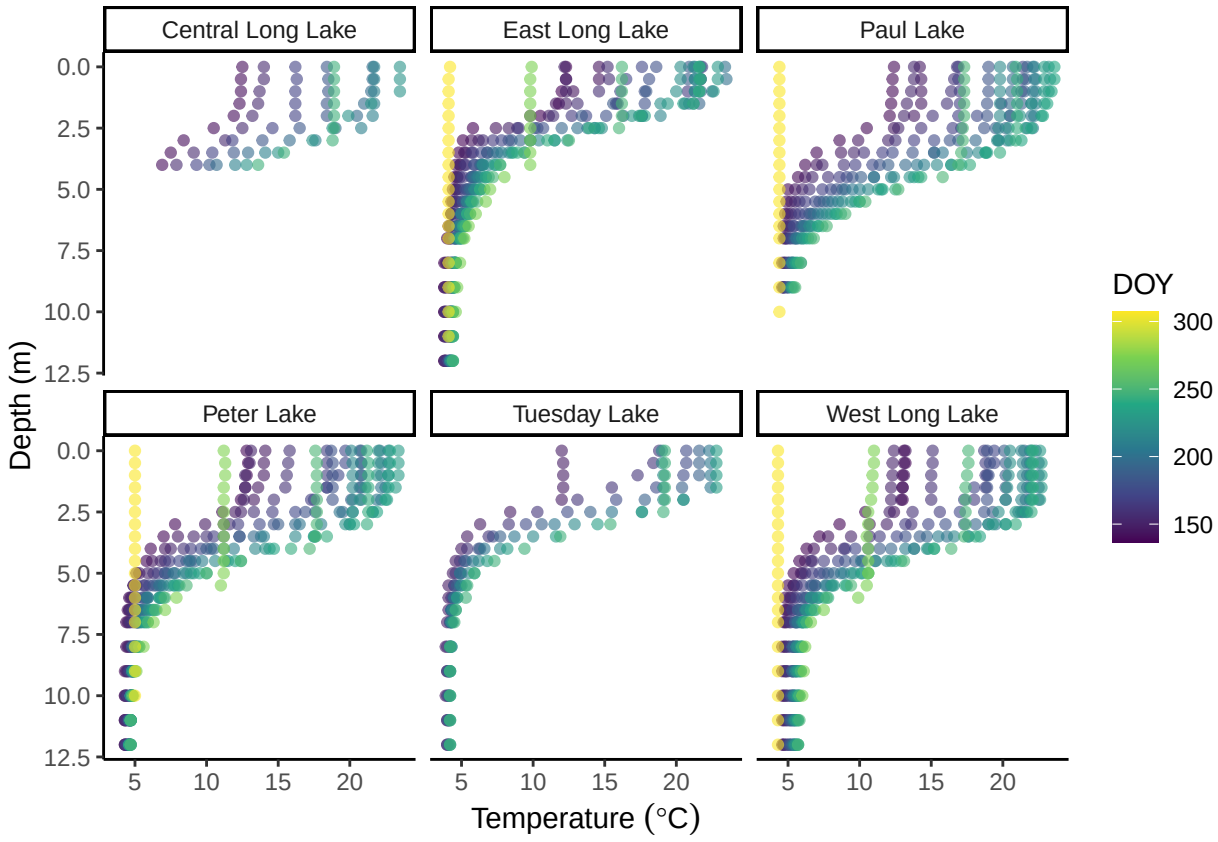
```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(rLakeAnalyzer)

# 3. Load dataset
plake <- read.csv("../Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")
# Set date column as date format
plake$sampledate <- as.Date(plake$sampledate, format="%m/%d/%y")

# 4. Add month column and omit temperature NAs; save changes
plake <- plake %>%
  mutate(Month = format(sampledate, "%m")) %>%
  drop_na(temperature_C)

# 5. Set theme to classic
theme_set(theme_classic())
```

Creating and analyzing lake temperature profiles

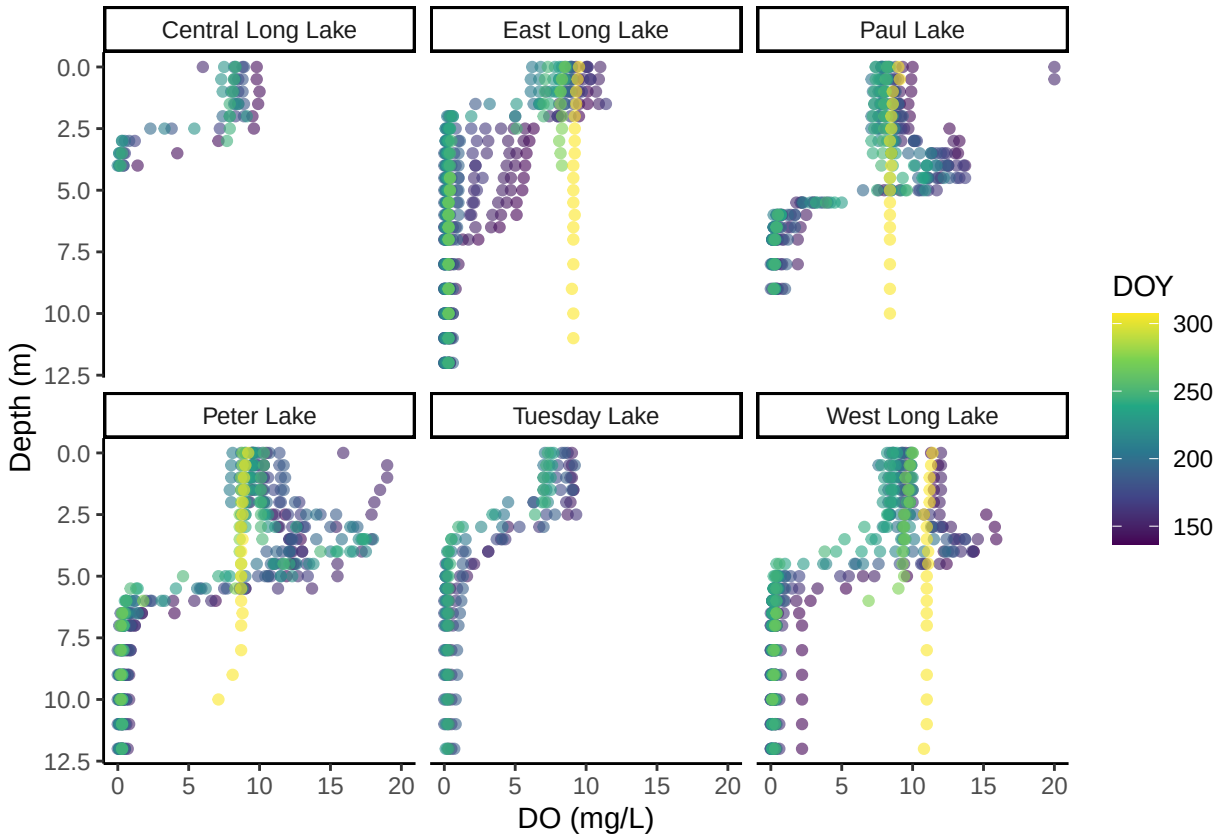
5. For the year 1993, plot temperature and dissolved oxygen profiles for all six lakes in the dataset (as two separate ggplots). Use the `facet_wrap` function to plot each lake as a separate panel in the plot. Plot day of year as your color aesthetic and use a reverse y scale to represent depth.

```
# 5. Plot temperature for 1993 by lake
ggplot(subset(plake, year4 == 1993),
       aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point(alpha = 0.6) +
  scale_y_reverse() +
  scale_color_viridis_c() +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)", color = "DOY") +
  facet_wrap(~lakename)
```




```
# Plot dissolved oxygen for 1993 by lake
ggplot(subset(plake, year4 == 1993),
       aes(x = dissolvedOxygen, y = depth, color = daynum)) +
  geom_point(alpha = 0.6) +
  scale_y_reverse() +
  scale_color_viridis_c() +
  labs(x = expression("DO (mg/L)"), y = "Depth (m)", color = "DOY") +
  facet_wrap(~lakename)
```

```
## Warning: Removed 8 rows containing missing values (geom_point).
```



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

Temperature shows the widest range in summer, which makes sense because that is a typical time for stratification. Unsurprisingly, temperatures are warmest in the summer. The range of temperatures is much broader in the summer/early fall. The limited winter data suggests that the epilimnion and hypolimnion have mixed and that the water temperature is stable throughout the water column. All the lakes show similar temperature patterns, except that Central Long Lake is clearly more shallow (and Paul Lake too, to some extent) and therefore does not show the same range of lower temperatures.

The dissolved oxygen patterns are more diverse across the different lakes. In general, dissolved oxygen levels are highest in the spring and decline over the year, increasing again in the winter. All of the lakes have a cluster of data at around 10 mg/L - for some lakes like East Long Lake, this is highest recorded dissolved oxygen level in the spring, whereas lakes like Peter and West Long have a chunk of higher oxygen levels at a depth of 2.5-5m in the spring. Probably these increased levels correspond to the presence of algae. All of the lakes show a drop off of oxygen levels moving into deeper waters, but that dropoff occurs at different levels - for example, East Long Lake has a drop in oxygen at about 2.5 m, whereas the dropoff in Paul Lake and Peter Lake is below 5 m. Perhaps physical properties like turbidity and light penetration could influence the depth at which dissolved oxygen levels drop precipitously (with algae presence as a mediator).

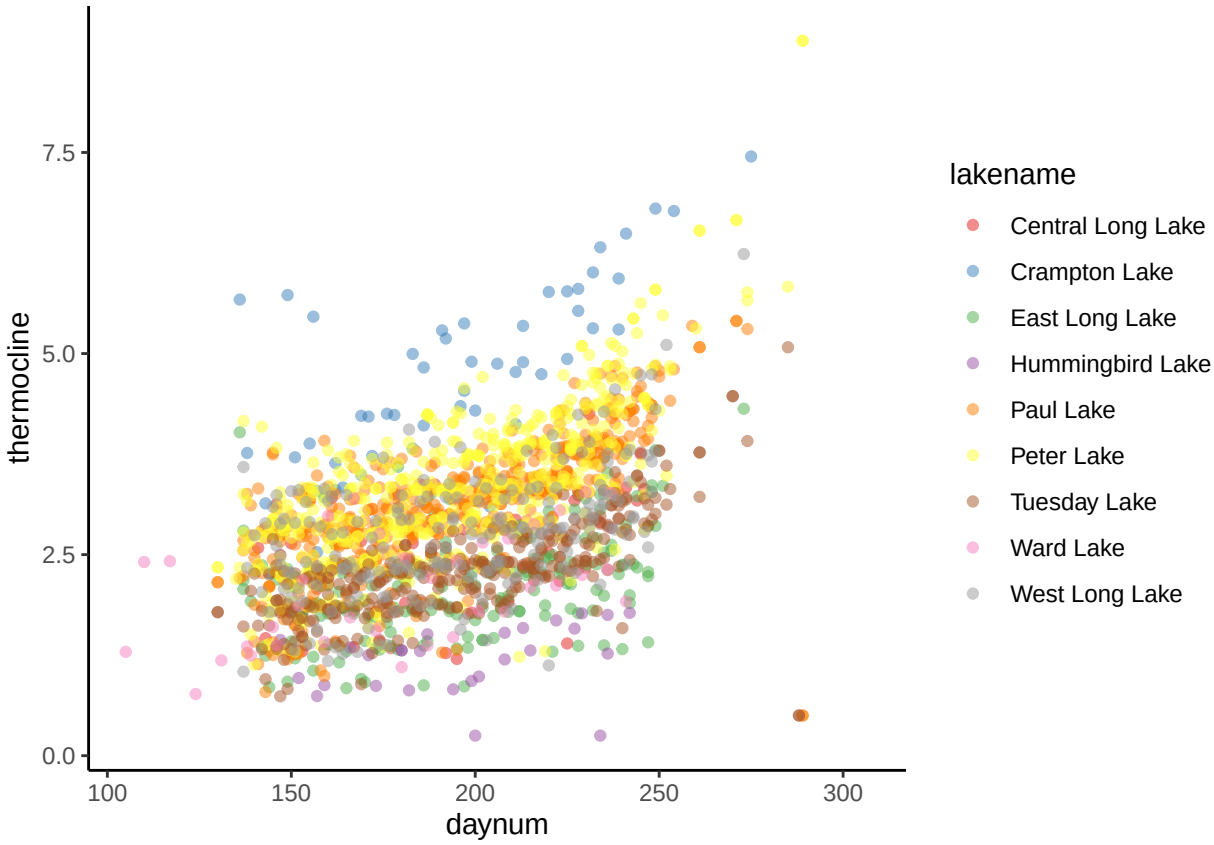
6. Create a new dataset that calculates thermocline depths for all lakes on all dates (hint: you will need group by lake, year, month, DOY, and sample date).
7. Plot thermocline depth by day of year for your newly made dataset. Color each point by lake name, make the points 50% transparent, and choose a color palette other than the ggplot default.
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.

```r
# 6. Calculate thermocline depths
thermo <- plake %>%
  group_by(lakename, year4, Month, daynum, sampledate) %>%
  summarise(thermocline = thermo.depth(wtr = temperature_C, depths = depth, seasonal = FALSE))
```

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

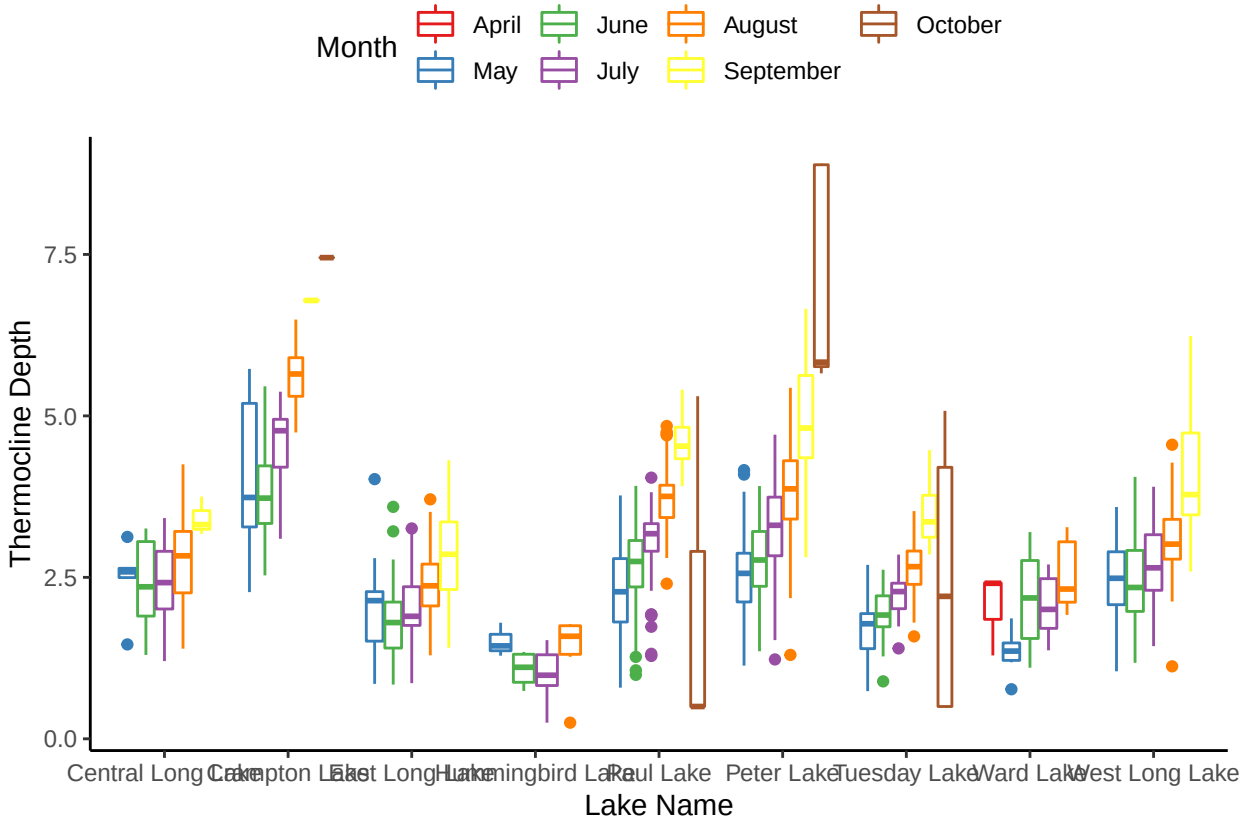
```r
# 7. Plot thermocline by day of year, colored by lake
ggplot(thermo, aes(x = daynum, y = thermocline, color = lakename)) +
  geom_point(alpha = 0.5) +
  scale_colour_brewer(palette = "Set1")
```

```
## Warning: Removed 8 rows containing missing values (geom_point).
```




```r
# Set Month as factor
thermo$Month <- as.factor(thermo$Month)

# 8. Boxplot of thermocline depth distributions by lake
ggplot(thermo, aes(x = lakename, y = thermocline, color = Month)) +
  geom_boxplot() +
  labs(x = "Lake Name", y = "Thermocline Depth") +
  theme(legend.position = "top") +
  scale_colour_brewer(palette = "Set1", labels = c("April", "May", "June", "July",
                                                   "August", "September", "October"))
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```



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 process of thermocline deepening varies by lake. Some lakes show a deepening over time from spring-fall, whereas others break that pattern. For example, Paul Lake and Tuesday Lake deepen gradually through September and then their thermoclines become more shallow in October. Central Long Lake, which is more shallow, shows a more narrow thermocline range than some of the other lakes. The lakes with the deepest thermoclines (Crampton and Peter) are likely to have a larger maximum depth.