Assignment 2: Physical Properties of Lakes

Student Name

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)

```
# Check working directory (should be project file location)
getwd()
```

```
## [1] "/Users/katerisalk/Box Sync/Courses/Water Data Analytics/Assignments"
```

```
# 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
```

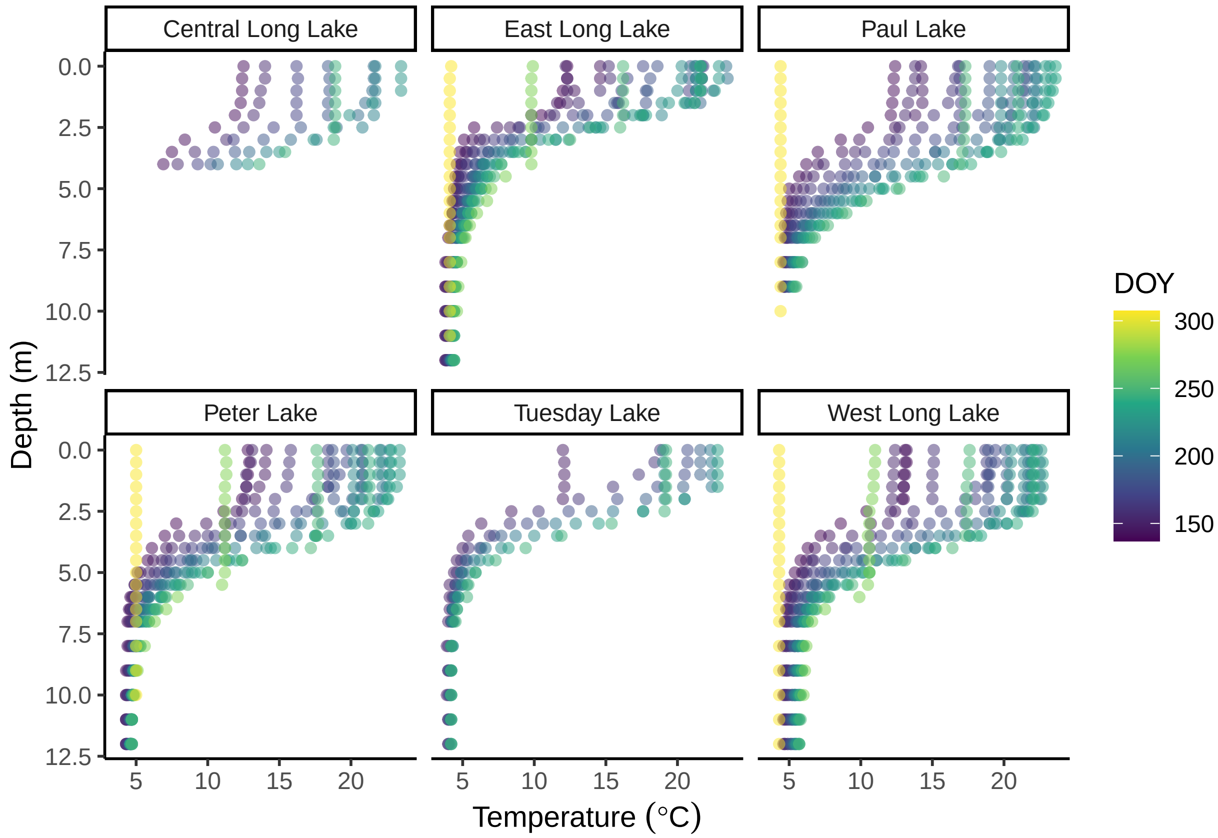
```
library(rLakeAnalyzer)

# Load data
NTLdata <- read.csv("../Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")
NTLdata$sampledate <- as.Date(NTLdata$sampledate, format = "%m/%d/%y")
NTLdata <- NTLdata %>%
  mutate(Month = month(sampledate)) %>%
  drop_na(temperature_C)


# set ggplot theme
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.

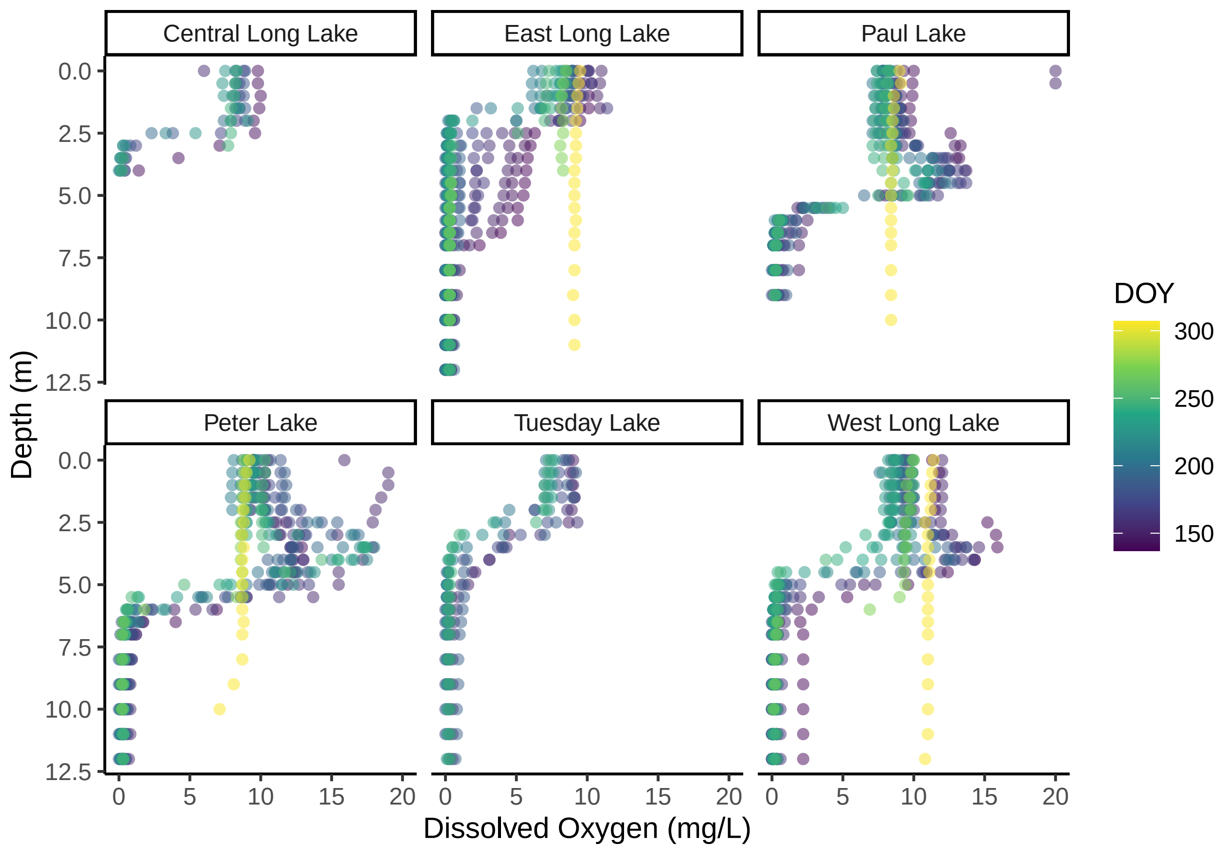
What seasonal trends do you observe, and do these manifest differently in each lake?

```
ggplot(subset(NTLdata, year4 == 1993),
       aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point(alpha = 0.5) +
  scale_y_reverse() +
  facet_wrap(vars(lakename)) +
  scale_color_viridis_c() +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)", color = "DOY")
```




```
ggplot(subset(NTLdata, year4 == 1993),
       aes(x = dissolvedOxygen, y = depth, color = daynum)) +
  geom_point(alpha = 0.5) +
  scale_y_reverse() +
  facet_wrap(vars(lakename)) +
  scale_color_viridis_c() +
  labs(x = "Dissolved Oxygen (mg/L)", y = "Depth (m)", color = "DOY")
```

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



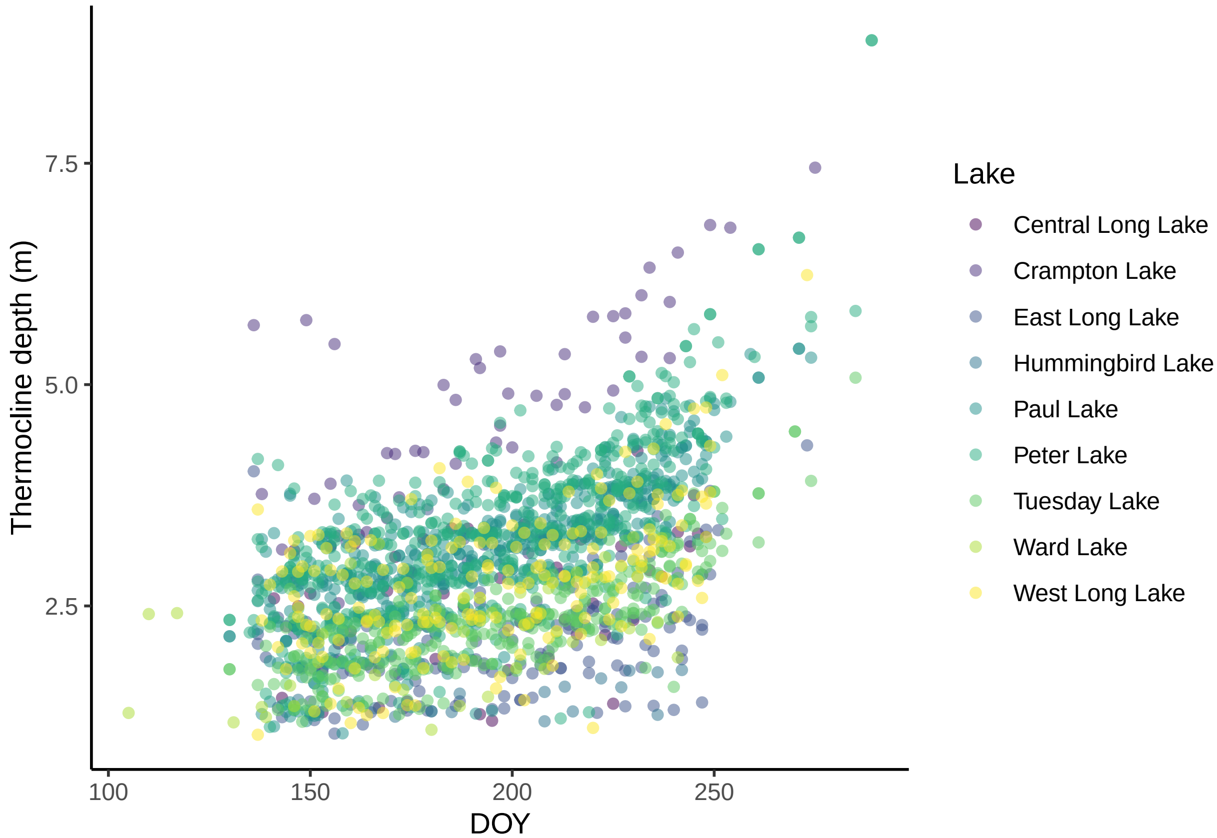
6. Create a new dataset that calculates thermocline depths for all lakes on all dates (hint: you will need group by lake, year, month, DOY, and sample date).
7. Plot thermocline depth by day of year for your newly made dataset. Color each point by lake name, make the points 50% transparent, and choose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

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?

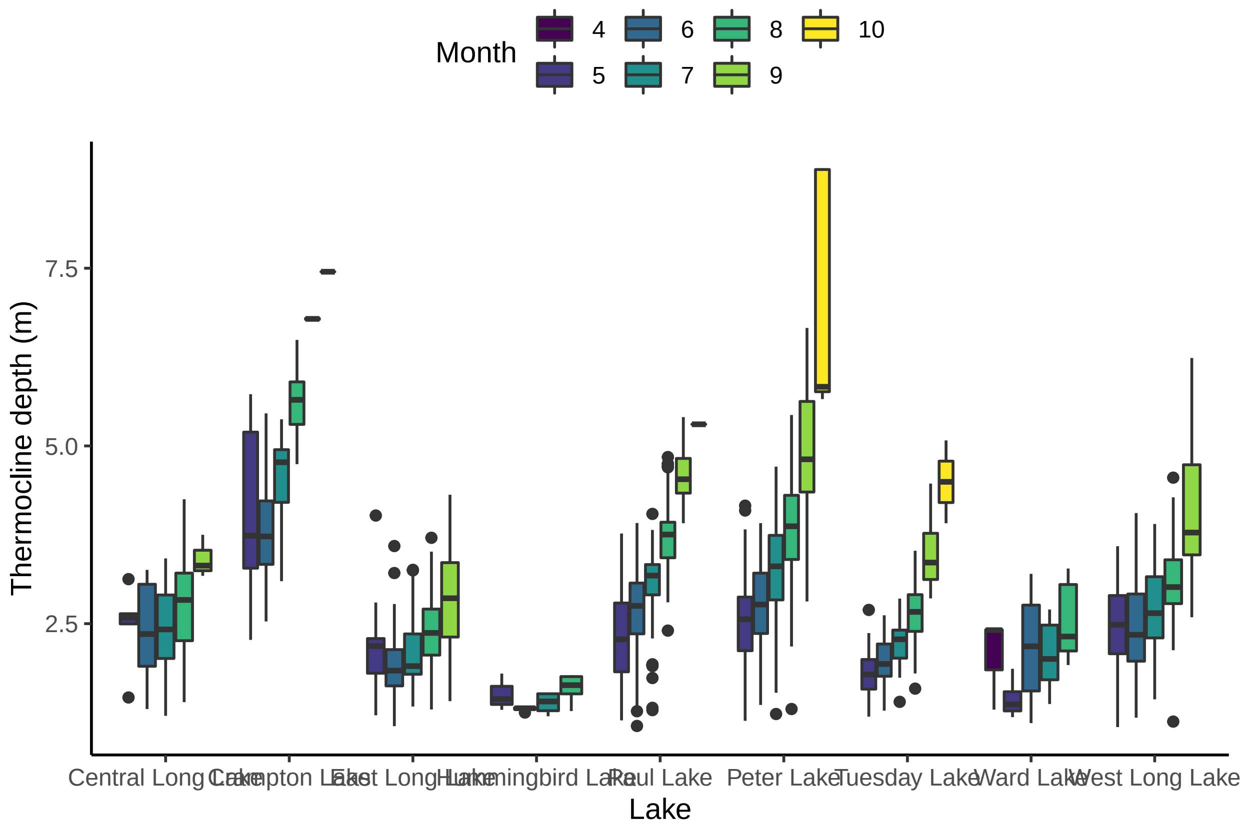
```
NTLdata_thermo <- NTLdata %>%
  group_by(lakename, daynum, Month, year4, sampledate) %>%
  # calculate thermoclines based on temperature profiles.
  # seasonal = FALSE calculates the thermocline as the maximum density gradient
  # seasonal = TRUE calculates the thermocline as the deepest density gradient
  summarise(thermocline = thermo.depth(wtr = temperature_C, depths = depth, seasonal = FALSE)) %>%
  # remove all thermoclines within 1 m of the surface. these can represent transient stratification.
  filter(thermocline > 1)
```

```
## `summarise()` has grouped output by 'lakename', 'daynum', 'Month', 'year4'. You can override using t
```

```
ggplot(NTLdata_thermo, aes(x = daynum, y = thermocline, color = lakename)) +
  geom_point(alpha = 0.5) +
  labs(x = "DOY", y = "Thermocline depth (m)", color = "Lake") +
  scale_color_viridis_d()
```




```
ggplot(NTLdata_thermo, aes(y = thermocline, x = lakename, fill = as.factor(Month))) +
  geom_boxplot() +
  labs(x = "Lake", y = "Thermocline depth (m)", fill = "Month") +
  scale_fill_viridis_d() +
  theme(legend.position = "top")
```



Month
4
5
6
7
8
9
10
Thermocline depth (m)
7.5
5.0
2.5
Central Long Lake
Crampton Lake
East Long Lake
Hummingbird Lake
Paul Lake
Peter Lake
Tuesday Lake
Ward Lake
West Long Lake
Lake