Assignment 2: Physical Properties of Lakes

Jack Carpenter

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
getwd()
```

```
## [1] "/Users/Jack/Documents/Duke/Spring 2022/Water Data Analytics/Water_Data_Analytics_2022"
```

```
#load packages
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   1.1.3     v stringr 1.4.0
## v readr   2.0.1     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)

#import dataset
NTL_LTER_data <- read_csv("./Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")
```

```
## Rows: 38614 Columns: 11
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): lakeid, lakename, sampledate, comments
## dbl (7): year4, daynum, depth, temperature_C, dissolvedOxygen, irradianceWat...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
view(NTL_LTER_data)

#set date column to date format
class(NTL_LTER_data$sampledate)
```

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

```
NTL_LTER_data$sampledate <- as.Date(NTL_LTER_data$sampledate, format = "%m/%d/%y")
class(NTL_LTER_data$sampledate)
```

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

```
#now the date is Y-m-d in the spreadsheet despite me asking for m-d-y...
#why is this?

#add a column called month
NTL_LTER_data <- NTL_LTER_data %>%
  mutate(month = month(sampledate))
#for mutate need following syntax:
#mutate(name of new column = type of data(old column))
#Dr. Salk is the above general syntax correct?

#remove temperature NA's
dim(NTL_LTER_data)
```

```
## [1] 38614    12
```

```
#is dim really necessary here?
NTL_LTER_data <- NTL_LTER_data %>%
  drop_na(temperature_C)
dim(NTL_LTER_data)
```

```
## [1] 34756    12
```

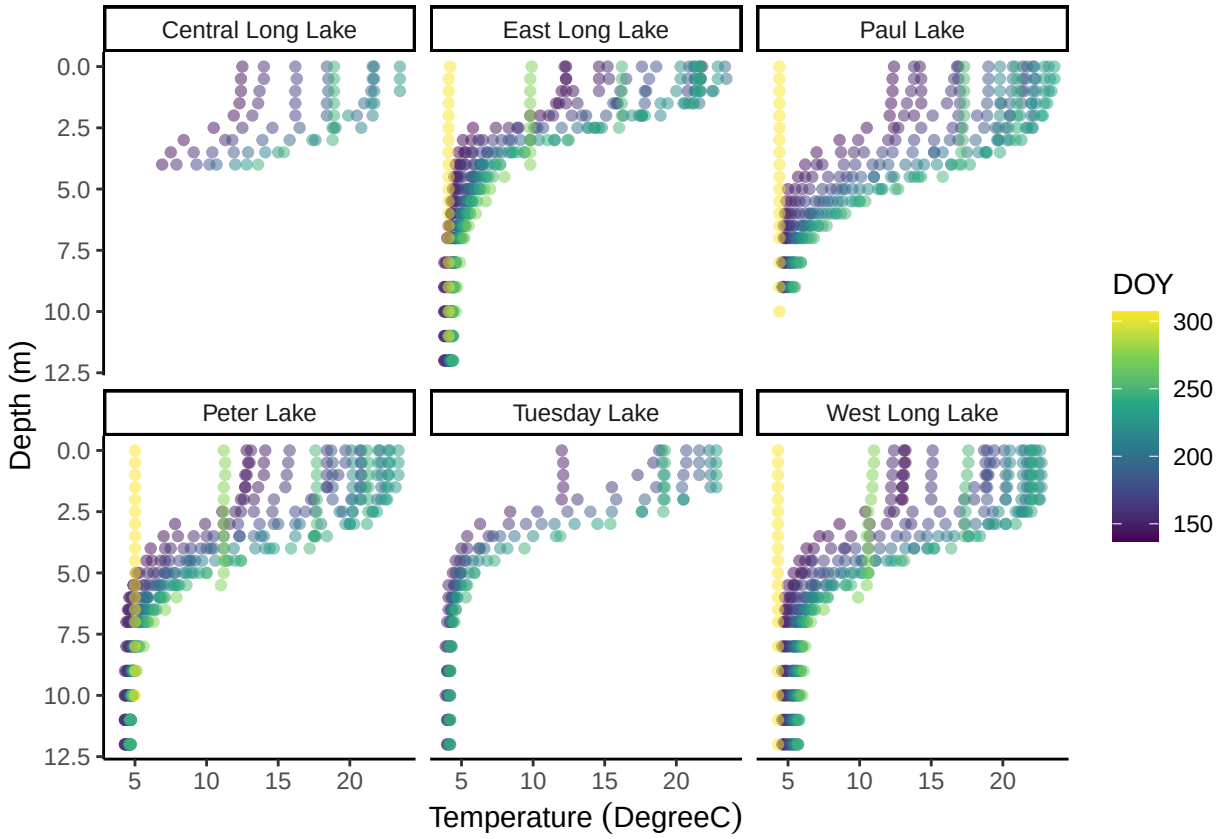
```
#ope looks like fewer rows of data - I guess its a check

#set 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.

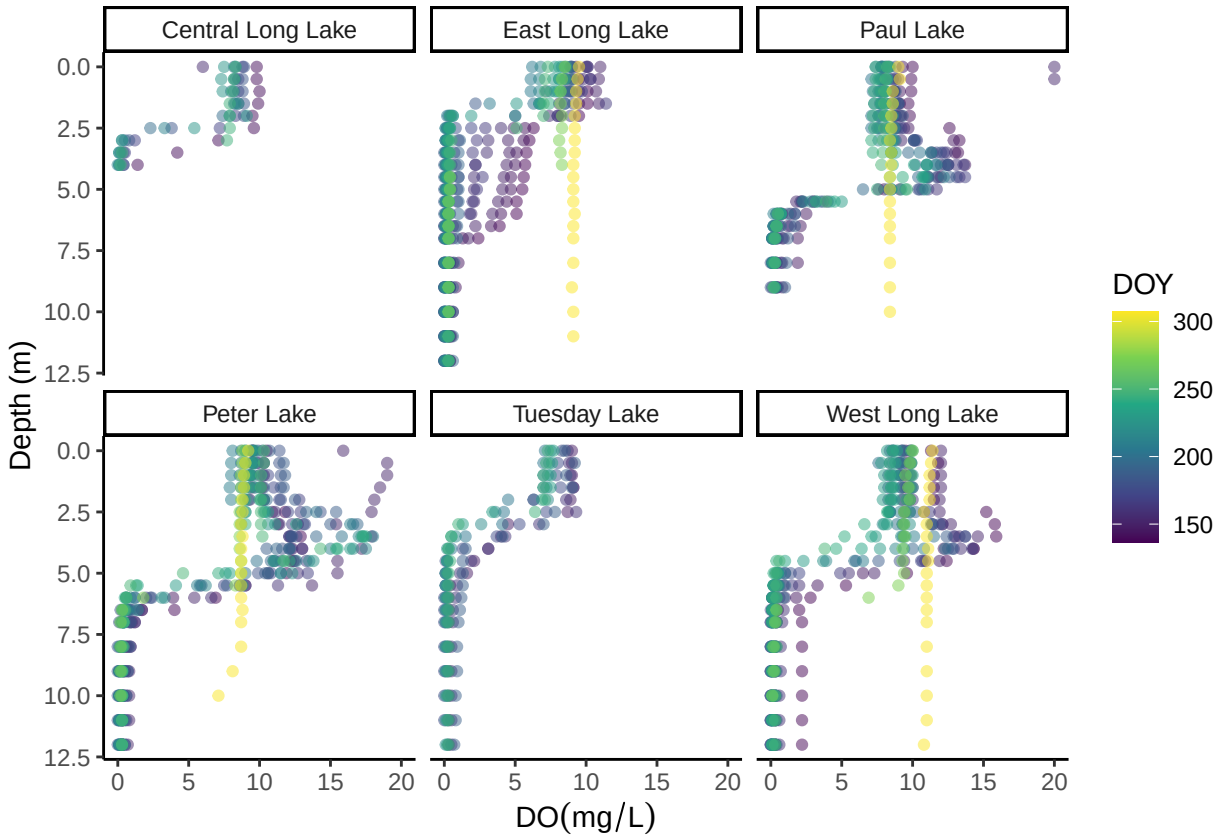
What seasonal trends do you observe, and do these manifest differently in each lake? Answer: In the temperature plot, each lake has a seasonal progression like we talked about in class with increasing temperature and deepening thermocline throughout the summer and mixing in the winter. The DO plots vary more widely between lakes. Paul, Peter, Tuesday,East Long, and West Long lakes show very similar DO profiles that show DO decrease over the summer, hold steady in the winter at all depths, and increase in the spring. I'm not sure what's happening with Central Long, but I think it looks strange from the fact that it has fewer data points rather than becaue it behaves vastly differently.

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




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

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



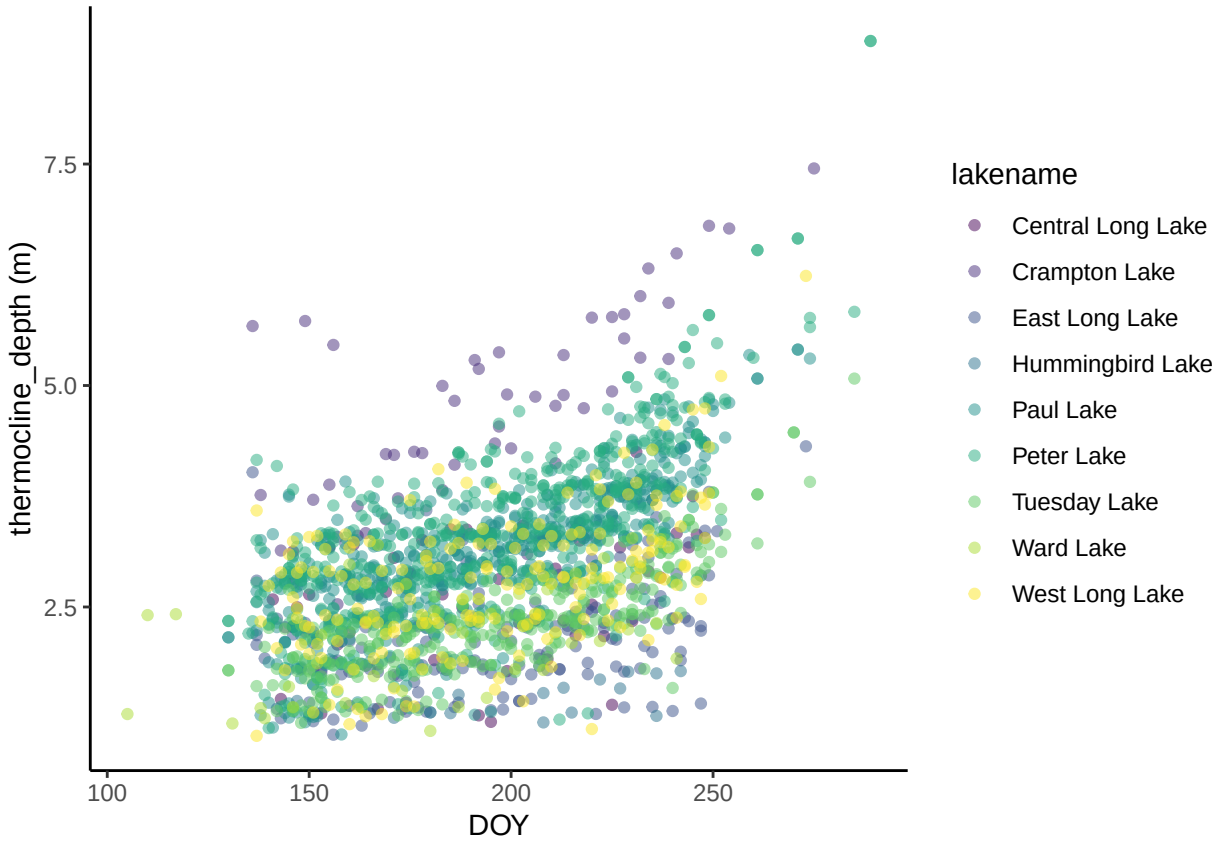
6. Create a new dataset that calculates thermocline depths for all lakes on all dates (hint: you will need group by lake, year, month, DOY, and sample date).
7. Plot thermocline depth by day of year for your newly made dataset. Color each point by lake name, make the points 50% transparent, and choose 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? Answer: All the lakes do appear to show a similar seasonal progression of thermocline deepening. The only one that might not is hummingbird lake, but that also may be a matter of scale - the change is minor compared to others. Peter and Crampton have the deepest thermoclines and both are deep lakes, whereas Hummingbird is a shallow lake and has a shallow thermocline.

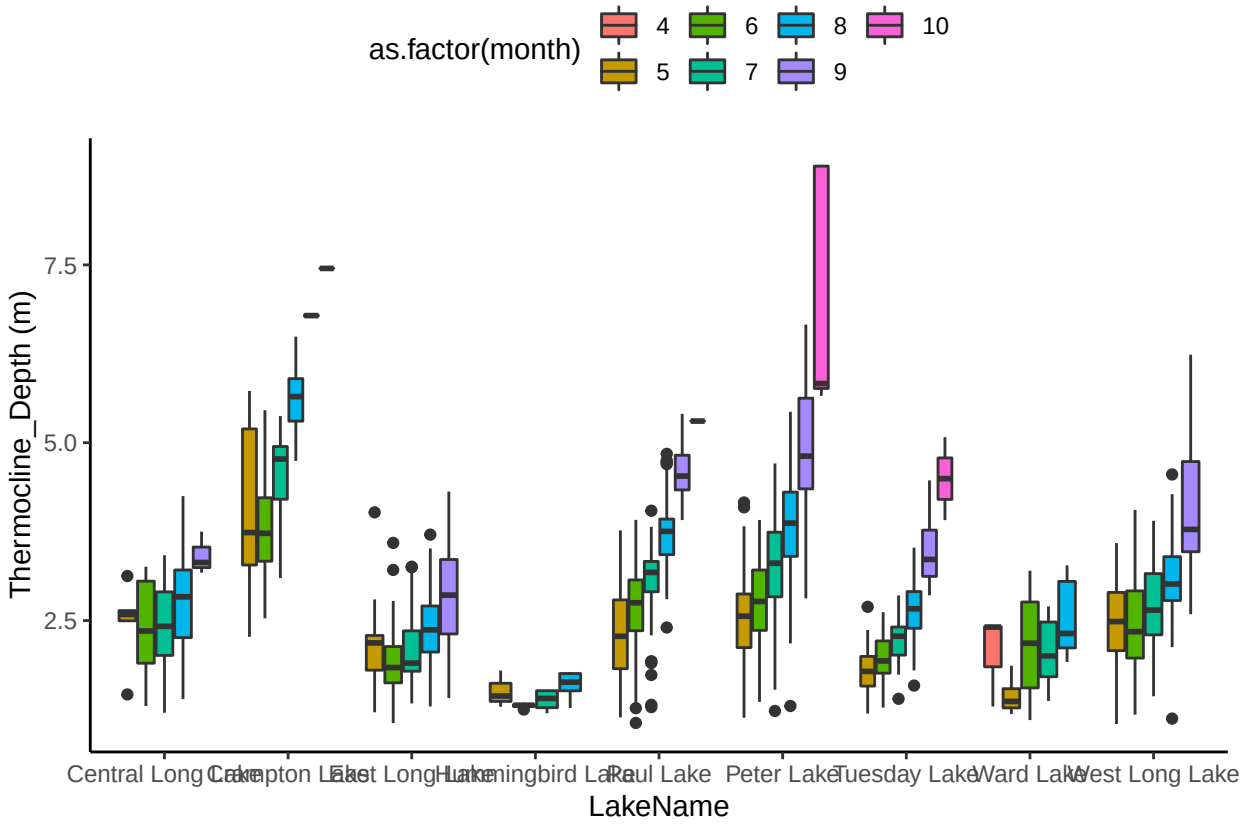
```
# create a new dataset
thermocline_data <- NTL_LTER_data %>%
  group_by(lakename, year4, month, daynum, sampledate) %>%
  summarise(thermocline_depth = thermo.depth(wtr = temperature_C, depths = depth, seasonal = FALSE)) %>%
  filter(thermocline_depth > 1)
```

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

```
#let's plot thermocline depths by lake
ggplot(thermocline_data, aes(x = daynum, y = thermocline_depth, color = lakename)) +
  geom_point(alpha = 0.5) +
  scale_color_viridis_d() +
  labs(x = "DOY", y = "thermocline_depth (m)", color = "lakename")
```




```
#now for the boxplots
ggplot(thermocline_data, aes(x = lakename, y = thermocline_depth, fill = as.factor(month))) +
  geom_boxplot() +
  scale_color_viridis_d() +
  labs(x = "LakeName", y = "Thermocline_Depth (m)", color = "Month") +
  theme(legend.position = "top")
```




```
Max_depths <- NTL_LTER_data %>%
  group_by(lakename) %>%
  summarise(max_depth = max(depth))
```