2: Physical Properties of Lakes

Water Data Analytics | Kateri Salk

Spring 2022

Lesson Objectives

1. Investigate the concepts of lake stratification and mixing by analyzing monitoring data
2. Apply data analytics skills to applied questions about physical properties of lakes
3. Communicate findings with peers through oral, visual, and written modes

Opening Discussion

What are the physical properties of lakes?

Session Set Up

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

```
## [1] "/Users/katerisalk/Documents/GitHub_Repos/Water_Data_Analytics_2022"
```

```
# install.packages("tidyverse")
# install.packages("lubridate")
# install.packages("rLakeAnalyzer")

# load packages
library(tidyverse)
library(lubridate)
library(rLakeAnalyzer)

# Load data
NTLdata <- read.csv("./Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")

# set ggplot theme
theme_set(theme_classic())
```

Data Wrangling and Exploration

Investigate structure of dataset

Common steps/checks for data wrangling:

- Date formatting
- Addressing NAs
- Subsetting and filtering

```
# Is the date column perceived as a date by R?
class(NTLdata$sampledate)
```

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

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

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

```r
# What does this column look like now?

# Remove rows that contain NAs in temperature column
dim(NTLdata)
```

```
## [1] 38614    11
```

```r
NTLdata <- NTLdata %>%
  drop_na(temperature_C)
dim(NTLdata)
```

```
## [1] 34756    11
```

```r
# How many observations are there for each lake?
summary(NTLdata$lakename)
```

```
##    Length     Class      Mode 
##     34756 character character
```

```r
summary(as.factor(NTLdata$lakename))
```

```
## Central Long Lake     Crampton Lake    East Long Lake  Hummingbird Lake 
##               443              1108              3550               378 
##         Paul Lake        Peter Lake      Tuesday Lake         Ward Lake 
##              9253             10189              5503               527 
##    West Long Lake 
##              3805
```

```r
# Let's choose the two lakes with the most data
NTLdata_PeterPaul <- NTLdata %>%
  filter(lakename %in% c("Paul Lake", "Peter Lake"))

# What is another way to use the filter command to get to the same result?
NTLdata_PeterPaul <- NTLdata %>%
  filter(lakename == "Paul Lake" | lakename == "Peter Lake")

# Make two data frames, one for each lake
Pauldata <- NTLdata %>%
  filter(lakename == "Paul Lake")
Peterdata <- NTLdata %>%
  filter(lakename == "Peter Lake")

# How long did the monitoring last?
min(Pauldata$sampledate)
```

```
## [1] "1984-05-27"
```

```r
max(Pauldata$sampledate)
```

```
## [1] "2016-08-16"
```

```r
min(Peterdata$sampledate)
```

```
## [1] "1984-05-28"
```

```r
max(Peterdata$sampledate)
```

```
## [1] "2016-08-15"
```

```r
# Which depths are sampled in each lake?
unique(Pauldata$depth)
```

```
##  [1]  0.0  1.0  2.0  3.0  4.0  5.0  6.0  7.0  8.0  9.0 10.0 11.0 12.0  0.5  1.5
## [16]  2.5  3.5 20.0  4.5  5.5  6.5  7.5 11.5 13.0  8.5  9.5
```

```r
unique(Peterdata$depth)
```

```
##  [1]  0.00  1.00  2.00  3.00  4.00  5.00  6.00  7.00  8.00  9.00 10.00 11.00
## [13] 12.00 13.00 14.00 15.00 16.00 17.00  0.50  1.50  2.50  3.50  4.50  5.50
## [25]  6.50  7.50  8.50  0.75
```

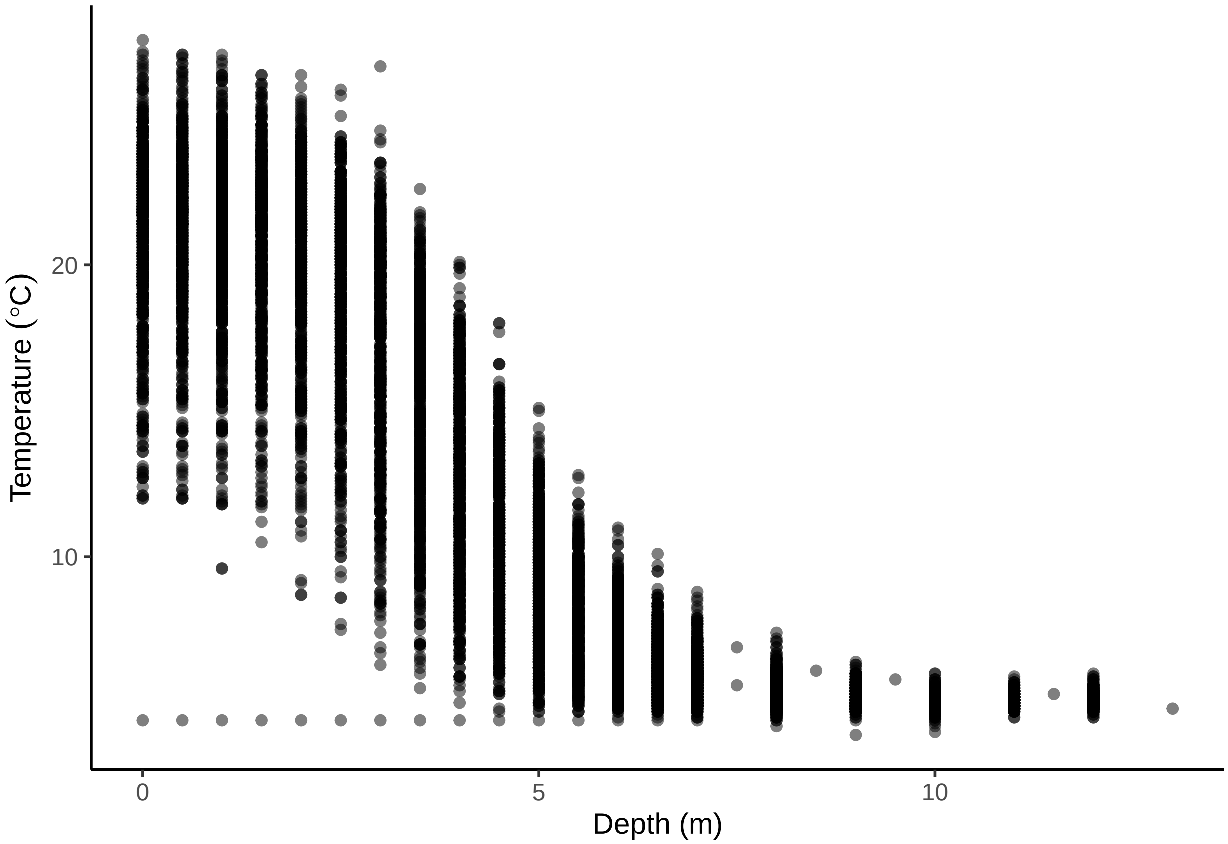
```r
# Why didn't we use the "summary" function here?

# QA the one data point at 20 m depth.
Pauldata <- Pauldata %>%
  filter(depth < 20)
```

Exploratory data visualization

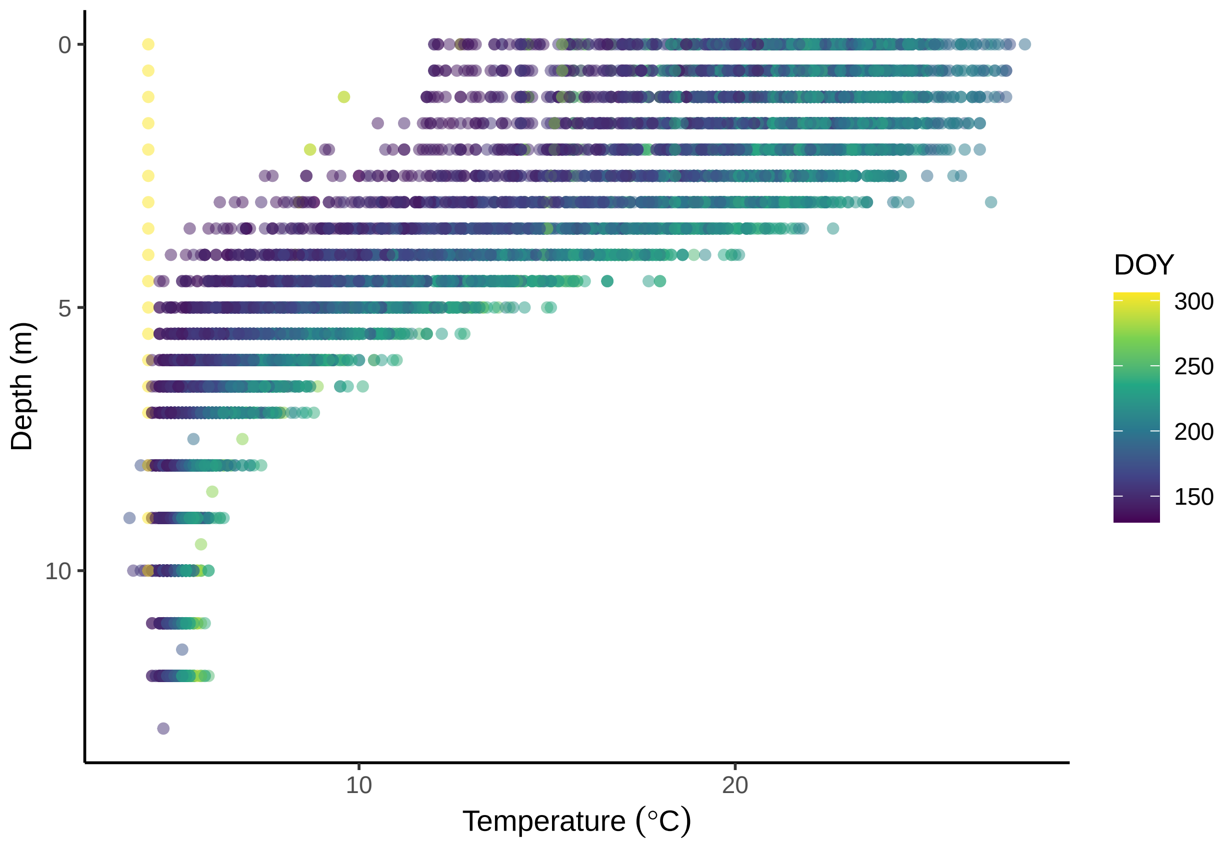
Let's make a plot of temperatures by depth. There are a lot of points, so adding a 50 % transparency to the points helps us see where points are densely clustered together.

Let's also try using the traditional limnological graph type, with depth on the y axis in reverse, to simulate a cross section of a lake. When and where do we usually observe high and low temperatures?

```r
ggplot(Pauldata, aes(x = depth, y = temperature_C)) +
  geom_point(alpha = 0.5) +
  labs(y = expression("Temperature "(degree*C)), x = "Depth (m)")
```




```
ggplot(Pauldata, 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")
```



How do temperatures at the surface compare to temperatures at the mid-depths and at the bottom?

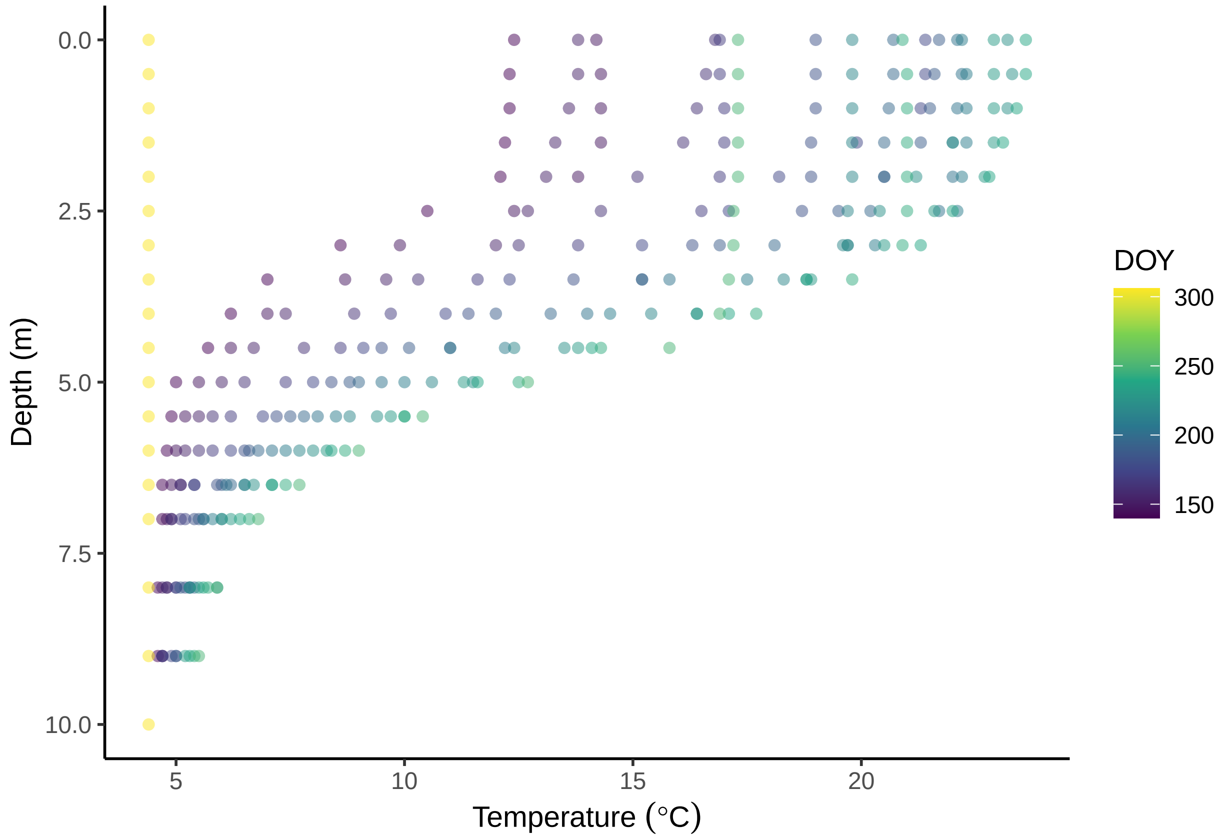
Graphing seasonal water profiles

Lakes in the North Temperate Lakes LTER are dimictic, meaning they mix fully twice per year. When the lakes are not mixed (stratified), the top layer (epilimnion) and the bottom layer (hypolimnion) of the lake are different temperatures.

For background information on lake stratification, see this stratification summary article from the North American Lake Management Society.

In 1993, Paul Lake was sampled from May 20 through November 2. Let's look at temperature profiles of the lake across the year and also explore how this impacts dissolved oxygen.

```
ggplot(subset(Pauldata, 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")
```




```
# Now, make a plot as above but with dissolved oxygen saturation instead of temperature.
```

Why does mixing occur in the spring and fall? What are the mechanisms that make this possible?

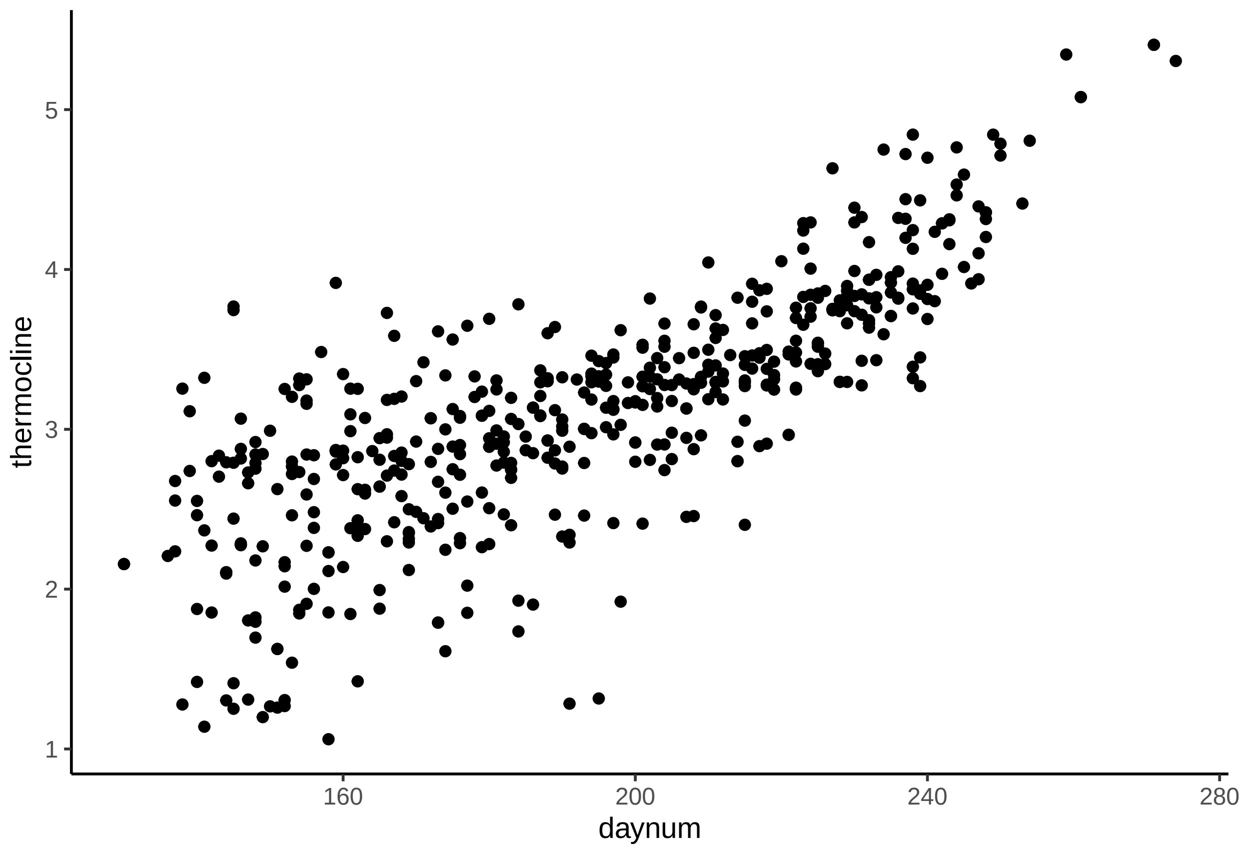
Calculating thermoclines

rLakeAnalyzer documentation

```
Pauldata_thermo <- Pauldata %>%
  group_by(year4, daynum, 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 'year4', 'daynum'. You can override using the `.groups` argument
```

```
ggplot(Pauldata_thermo, aes(x = daynum, y = thermocline)) +
  geom_point()
```



How does thermocline depth change over the course of the year? When do we observe the most variability in thermocline depth?

Climate change

Increases in global temperature are predicted to cause several changes to lake thermal conditions, including:

- Increases in surface temperature
- Increases in surface minimum temperature
- Increases in extent of stratification
- Increases in length of stratification
- Decreases in ice cover

Several studies have addressed this topic, using long-term and spatially diverse datasets:

https://link.springer.com/article/10.1007/s10584-015-1326-1?sa_campaign=email/event/articleAuthor/onlineFirst&error=cookies_not_supported&error=cookies_not_supported&code=2b415e25-de4c-452f-bd02-2cceae08b7a3&code=e63aabb9-76d3-4e49-b36c-e591007a9e9c

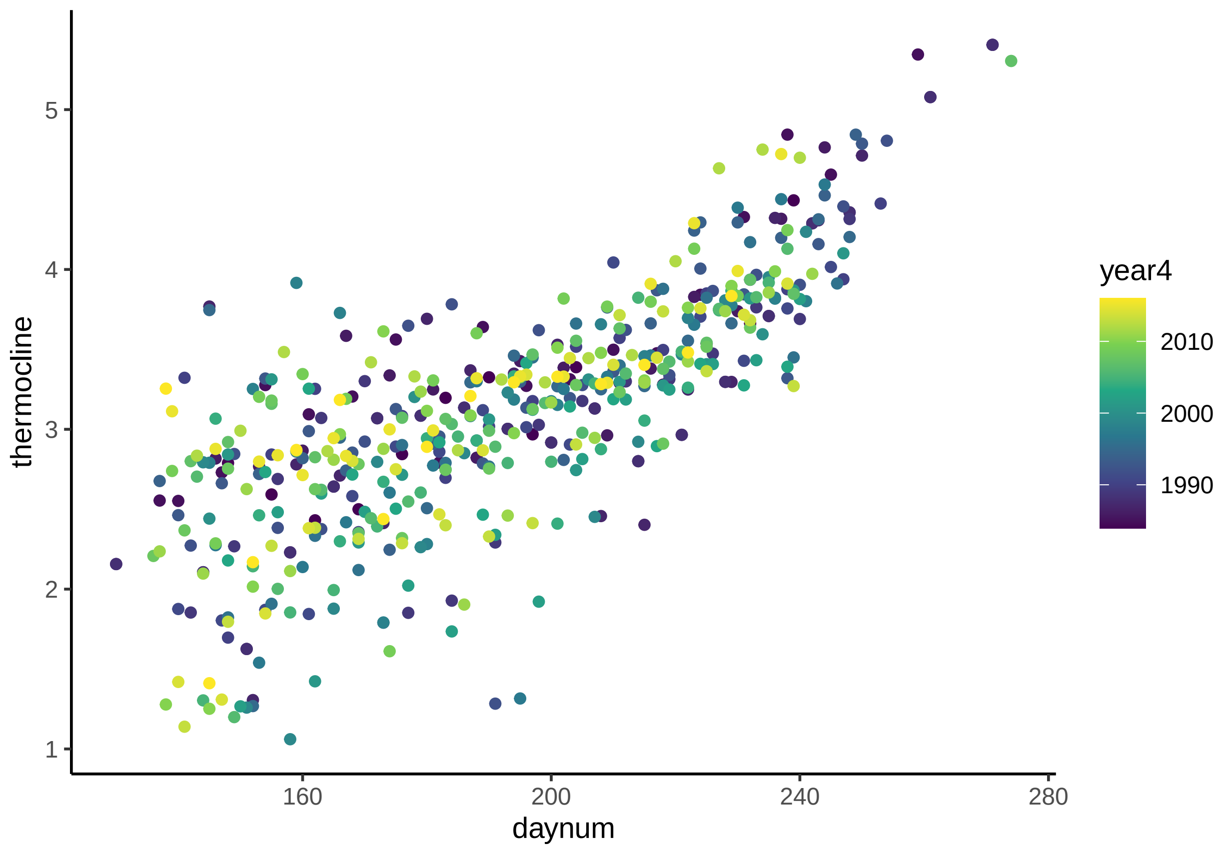
http://hpkx.cnjournals.com/uploadfile/news_images/hpkx/2020-07-15/10.1038-s43017-020-0067-5.pdf

https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015GL064097

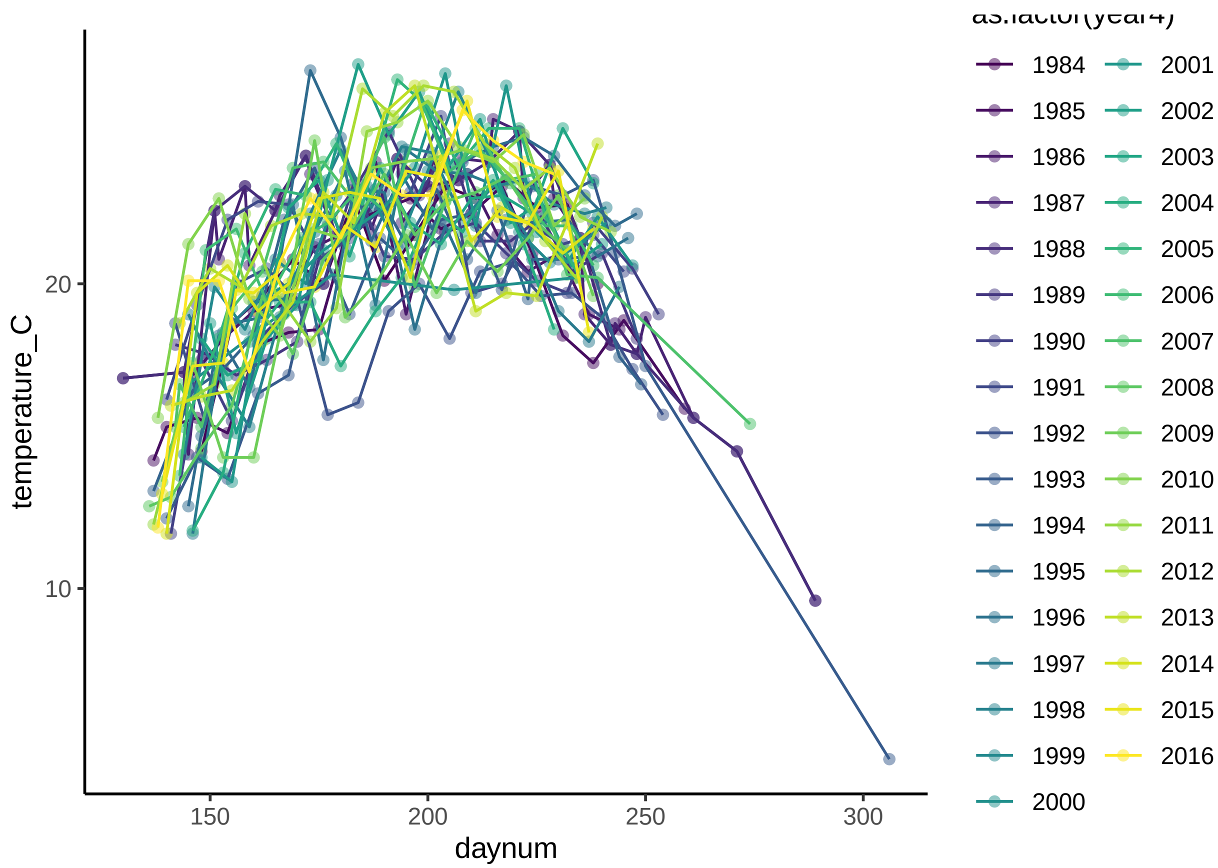
https://link.springer.com/article/10.1007/s10584-019-02465-y

Let's explore how these changes might manifest in Paul Lake.

```
ggplot(Pauldata_thermo, aes(x = daynum, y = thermocline, color = year4)) +
  geom_point() +
  scale_color_viridis_c()
```




```r
# exercise: put the legend on top and relabel the aesthetics

ggplot(subset(Pauldata, depth == 1),
       aes(x = daynum, y = temperature_C, color = as.factor(year4))) +
  geom_point(alpha = 0.5) +
  geom_line() +
  scale_color_viridis_d()
```




```
# exercise: relabel the aesthetics
```

Closing Discussion

What are the main concepts you learned about the physical properties of lakes today? What was the evidence for these concepts in the dataset?