Assignment 2: Physical Properties of Lakes

Felipe Raby Amadori

OVERVIEW

This exercise accompanies the lessons in Hydrologic Data Analysis 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 Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A02_LakePhysical.Rmd") prior to submission.

The completed exercise is due on 11 September 2019 at 9:00 am.

Setup

1. Verify your working directory is set to the R project file,
2. Load the tidyverse, lubridate, and cowplot packages
3. Import the NTL-LTER physical lake dataset and set the date column to the date format
4. Set your ggplot theme (can be theme_classic or something else)

```
knitr::opts_chunk$set(message = FALSE, warning = FALSE)

getwd()
```

```
## [1] "C:/Users/Felipe/OneDrive - Duke University/1. DUKE/Ramos 3 Semestre/Hydrologic_Data_Analysis"
```

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

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

```
library(cowplot)
```

```
##
## ********************************************************

## Note: As of version 1.0.0, cowplot does not change the

##   default ggplot2 theme anymore. To recover the previous

##   behavior, execute:
##   theme_set(theme_cowplot())
```

```
## ********************************************************
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:cowplot':
##
##     stamp
## The following object is masked from 'package:base':
##
##     date
```

```r
NTL_data <- read.csv("./Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")

theme_set(theme_classic())
```

Creating and analyzing lake temperature profiles

Single lake, multiple dates

5. Choose either Peter or Tuesday Lake. Create a new data frame that wrangles the full data frame so that it only includes that lake during two different years (one year from the early part of the dataset and one year from the late part of the dataset).

```r
str(NTL_data)
```

```
## 'data.frame':    38614 obs. of  11 variables:
##  $ lakeid         : Factor w/ 9 levels "C","E","H","L",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ lakename       : Factor w/ 9 levels "Central Long Lake",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ year4          : int  1984 1984 1984 1984 1984 1984 1984 1984 1984 1984 ...
##  $ daynum         : int  148 148 148 148 148 148 148 148 148 148 ...
##  $ sampledate     : Factor w/ 1712 levels "10/1/07","10/1/93",..: 134 134 134 134 134 134 134 134 13
##  $ depth          : num  0 0.25 0.5 0.75 1 1.5 2 3 4 5 ...
##  $ temperature_C  : num  14.5 NA NA NA 14.5 NA 14.2 11 7 6.1 ...
##  $ dissolvedOxygen: num  9.5 NA NA NA 8.8 NA 8.6 11.5 11.9 2.5 ...
##  $ irradianceWater: num  1750 1550 1150 975 870 610 420 220 100 34 ...
##  $ irradianceDeck : num  1620 1620 1620 1620 1620 1620 1620 1620 1620 1620 ...
##  $ comments       : Factor w/ 2 levels "DO Probe bad - Doesn't go to zero",..: NA NA NA NA NA NA NA N
```

```r
NTL_data$sampledate <- as.Date(NTL_data$sampledate, "%m/%d/%y")
NTLdataTuesday <- filter(NTL_data, lakename == "Tuesday Lake")

str(NTLdataTuesday)
```

```
## 'data.frame':    6107 obs. of  11 variables:
##  $ lakeid         : Factor w/ 9 levels "C","E","H","L",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ lakename       : Factor w/ 9 levels "Central Long Lake",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ year4          : int  1984 1984 1984 1984 1984 1984 1984 1984 1984 1984 ...
##  $ daynum         : int  150 150 150 150 150 150 150 150 150 150 ...
##  $ sampledate     : Date, format: "1984-05-29" "1984-05-29" ...
##  $ depth          : num  0 0.25 0.5 0.75 1 1.5 2 3 4 5 ...
##  $ temperature_C  : num  15 NA NA NA 14.5 14 10.5 6.8 5.3 5 ...
##  $ dissolvedOxygen: num  9.5 NA NA NA 9.1 8.2 9.8 5.7 3 1.1 ...
##  $ irradianceWater: num  1850 1150 760 480 320 140 63 9.5 2 0.6 ...
```

```
##  $ irradianceDeck : num  1960 1960 1960 1960 1960 1960 1960 1960 1960 1960 ...
##  $ comments       : Factor w/ 2 levels "DO Probe bad - Doesn't go to zero",..: NA NA NA NA NA NA NA N
```

```
summary(NTLdataTuesday)
```

```
##      lakeid                   lakename        year4          daynum
##  T      :6107   Tuesday Lake     :6107   Min.   :1984   Min.   :130.0
##  C      :   0   Central Long Lake:   0   1st Qu.:1988   1st Qu.:168.0
##  E      :   0   Crampton Lake    :   0   Median :1994   Median :196.0
##  H      :   0   East Long Lake   :   0   Mean   :1997   Mean   :196.5
##  L      :   0   Hummingbird Lake :   0   3rd Qu.:2002   3rd Qu.:224.0
##  M      :   0   Paul Lake        :   0   Max.   :2016   Max.   :306.0
##  (Other):   0   (Other)          :   0
##    sampledate             depth         temperature_C   dissolvedOxygen
##  Min.   :1984-05-29   Min.   : 0.000   Min.   : 0.30   Min.   :  0.000
##  1st Qu.:1988-06-01   1st Qu.: 1.500   1st Qu.: 4.40   1st Qu.:  0.200
##  Median :1994-07-01   Median : 4.000   Median : 6.40   Median :  0.800
##  Mean   :1997-08-24   Mean   : 4.339   Mean   :10.35   Mean   :  3.741
##  3rd Qu.:2002-11-02   3rd Qu.: 6.500   3rd Qu.:17.00   3rd Qu.:  7.000
##  Max.   :2016-08-17   Max.   :16.000   Max.   :27.70   Max.   :802.000
##                                        NA's   :604     NA's   :668
##  irradianceWater  irradianceDeck
##  Min.   :   0.0   Min.   :   2.5
##  1st Qu.:   9.1   1st Qu.: 345.6
##  Median :  59.9   Median : 740.0
##  Mean   : 205.9   Mean   : 737.9
##  3rd Qu.: 266.0   3rd Qu.:1085.0
##  Max.   :2000.0   Max.   :2100.0
##  NA's   :2914     NA's   :3192
##                                 comments
##  DO Probe bad - Doesn't go to zero:  74
##  DO taken with Jones Lab Meter    :  50
##  NA's                             :5983
##
##
##
##
```

```
unique(NTLdataTuesday$year4)
```

```
##  [1] 1984 1985 1986 1987 1988 1989 1990 1991 1993 1994 1995 1996 1997 1998
## [15] 1999 2000 2002 2007 2012 2013 2014 2015 2016
```

```
#Chose 1985 and 2015

NTLdataTuesday_2years <- filter(NTLdataTuesday, year4 == 1985 | year4 == 2015)
```

6. Create three graphs: (1) temperature profiles for the early year, (2) temperature profiles for the late year, and (3) a `plot_grid` of the two graphs together. Choose `geom_point` and color your points by date.

Remember to edit your graphs so they follow good data visualization practices.

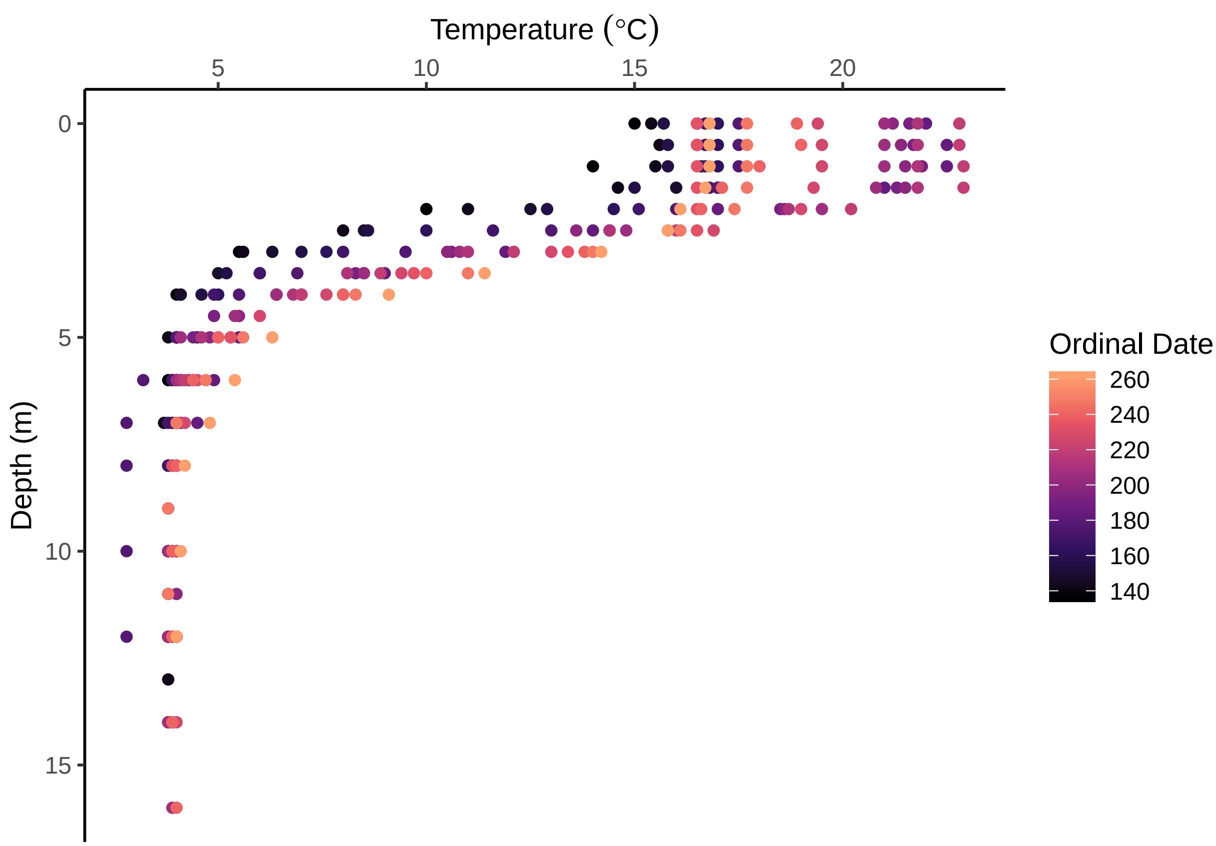
```
NTLdataTuesday_1985 <- filter(NTLdataTuesday, year4 == 1985)
NTLdataTuesday_2015 <- filter(NTLdataTuesday, year4 == 2015)

summary(NTLdataTuesday_1985$daynum)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   137.0   164.8   198.0   196.9   226.0   261.0
```

```
summary(NTLdataTuesday_2015$daynum)
```

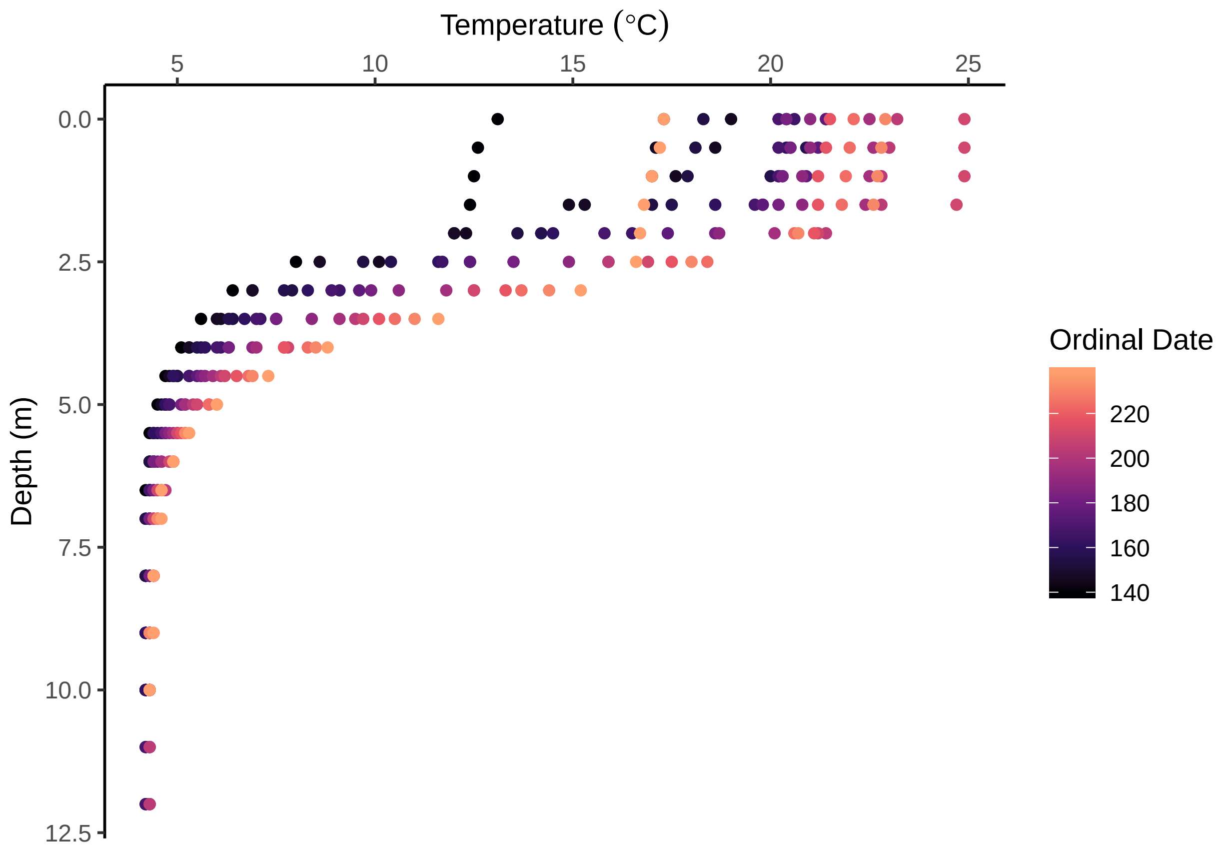
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   140.0   156.0   182.0   184.1   210.0   238.0
```

```
Tempprofiles_1985 <-
  ggplot(NTLdataTuesday_1985,
         aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "magma") +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)",
       color = "Ordinal Date")
print(Tempprofiles_1985)
```



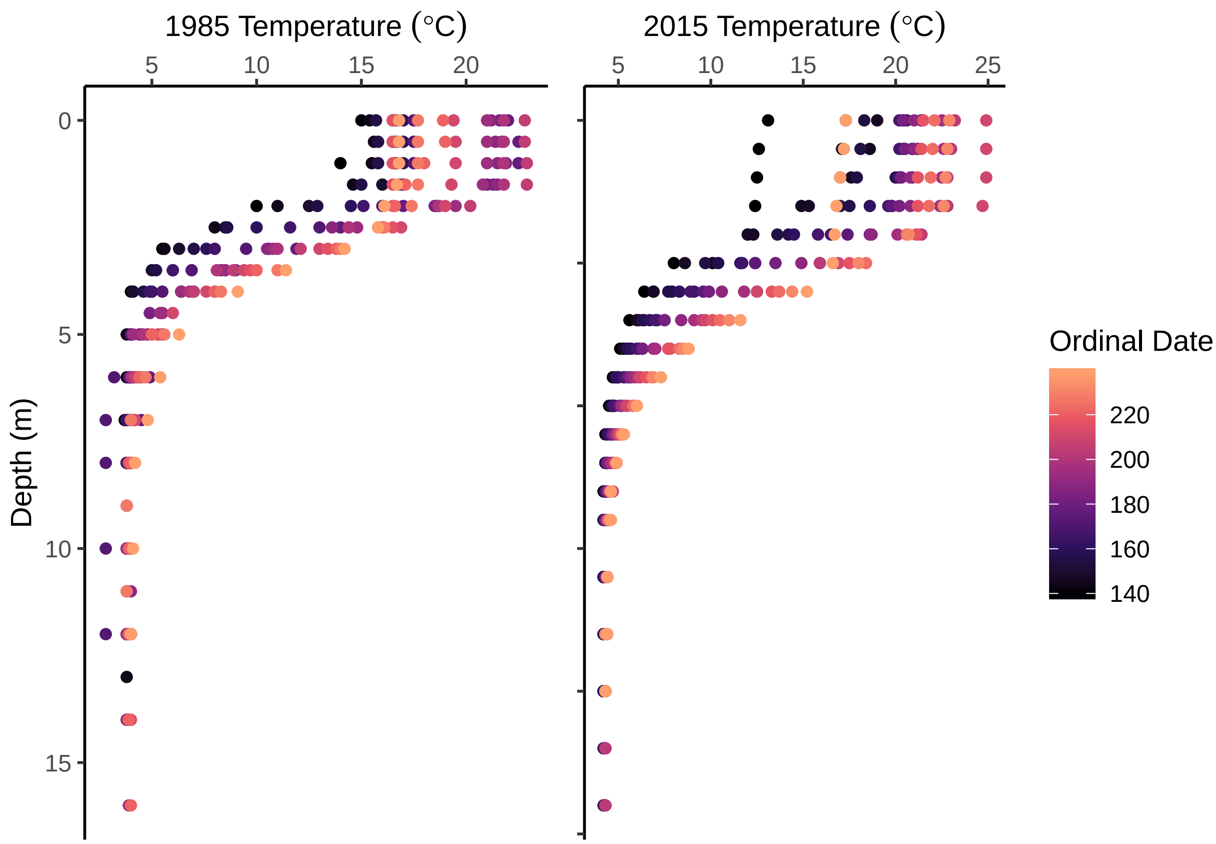

```
Tempprofiles_2015 <-
  ggplot(NTLdataTuesday_2015,
         aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "magma") +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)",
       color = "Ordinal Date")
```

```r
print(Tempprofiles_2015)
```




```r
#For the grid plot

Tempprofiles_1985_grid <-
  ggplot(NTLdataTuesday_1985,
         aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "magma") +
  labs(x = expression("1985 Temperature "(degree*C)), y = "Depth (m)") +
  theme(legend.position = "none")

Tempprofiles_2015_grid <-
  ggplot(NTLdataTuesday_2015,
         aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "magma") +
  labs(x = expression("2015 Temperature "(degree*C)), y = "Depth (m)",
       color = "Ordinal Date") +
  theme(axis.text.y = element_blank(), axis.title.y = element_blank())

TempProfilesAll <-
```

```r
  plot_grid(Tempprofiles_1985_grid, Tempprofiles_2015_grid,
            ncol = 2, rel_widths = c(1.25,1.5))
print(TempProfilesAll)
```



7. Interpret the stratification patterns in your graphs in light of seasonal trends. In addition, do you see differences between the two years?

   In both years the temperature in the water column goes down with depth, showing a higher range of temperatures during summer months. The epilimnion is the well mixed (homogeneous temperature) surface layer. This layer presents high variability of temperatures during the year, varying from approx. 15°C to 25°C for both years. The higher water temperature days coincide with the highest atmospheric temperature season. The lowest water temperature was measured durin the month of May, influenced by winter weather. The metalimnion is the middle layer that shows a rapid decrease of temperature during the year leading to the bottom layer called hypolimnion, which shows very little temperature variability during the year, presenting temperatures of approx. 4-5°C for all the data collected. In 1985 the epilimnion had a length of approx. 2.5 meters. In 2015 the epilimnion was a little bit bigger with a length of approx. 3 meters. It can be observed that the epilimnion presents a similar behavior, starting at a depth of approx. 6 meters in 1985 and at a depth of 7 meters in 2015.

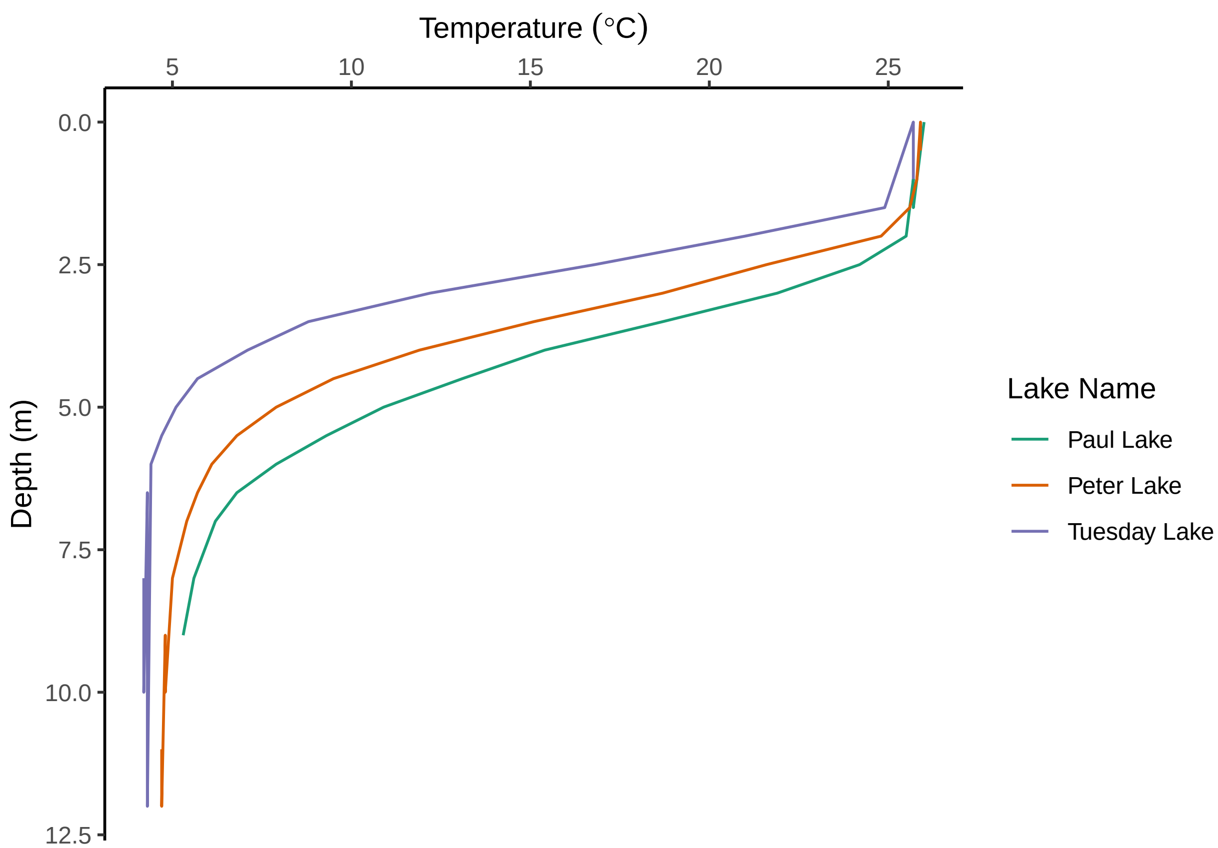
Multiple lakes, single date

8. On July 25, 26, and 27 in 2016, all three lakes (Peter, Paul, and Tuesday) were sampled. Wrangle your data frame to include just these three dates.

```
NTLdataTuesdayPeterPaul <- filter(NTL_data,
                                  lakename == "Tuesday Lake" | lakename == "Paul Lake" |
                                    lakename == "Peter Lake")
NTLdataTuesdayPeterPaul_Days <- filter(NTLdataTuesdayPeterPaul,
                                       sampledate == "2016-07-25" |
                                         sampledate == "2016-07-26" |
                                         sampledate == "2016-07-27")
```

9. Plot a profile line graph of temperature by depth, one line per lake. Each lake can be designated by a separate color.

```
Tempprofiles_July <-
  ggplot(NTLdataTuesdayPeterPaul_Days,
         aes(x = temperature_C, y = depth, color = lakename)) +
  geom_line() +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_brewer(palette = "Dark2") +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)",
       color = "Lake Name")
print(Tempprofiles_July)
```



10. What is the depth range of the epilimnion in each lake? The thermocline? The hypolimnion?

According to the data presented in the graph of temperature by depth for Tuesday Lake the depth range of the Epilimnion, Thermocline, and Hypolimnion are:

Lake	Epilimnion depth range (m)	Thermocline depth range (m)	Hypolimnion depth range (m)
Tuesday Lake	0 - 2	2 - 6	6 - bottom
Paul Lake	0 - 2.5	2.5 - 9	9 - bottom
Peter Lake	0 - 2.4	2.4 - 8	8 - bottom

Trends in surface temperatures over time.

11. Run the same analyses we ran in class to determine if surface lake temperatures for a given month have increased over time ("Long-term change in temperature" section of day 4 lesson in its entirety), this time for either Peter or Tuesday Lake.

```r
# Steps:
#
# 1. Add a column named "Month" to the data frame (hint: lubridate package)
NTLdataTuesday$Month <- month(NTLdataTuesday$sampledate)
# 2. Filter your data frame so that it only contains surface depths and months 5-8
NTLdataTuesday.Summer <- filter(NTLdataTuesday,
                                Month == 5 | Month == 6 | Month == 7 | Month == 8)
NTLdataTuesday.Summer.Surface <- filter(NTLdataTuesday.Summer, depth == 0)
# 3. Create 4 separate data frames, one for each month
NTLdataTuesday.Summer.Surface_May <- filter(NTLdataTuesday.Summer.Surface, Month == 5)
NTLdataTuesday.Summer.Surface_June <- filter(NTLdataTuesday.Summer.Surface, Month == 6)
NTLdataTuesday.Summer.Surface_July <- filter(NTLdataTuesday.Summer.Surface, Month == 7)
NTLdataTuesday.Summer.Surface_August <- filter(NTLdataTuesday.Summer.Surface, Month == 8)
# 4. Run a linear regression for each data frame
lm.Tuesday.May <- lm(temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_May)
summary(lm.Tuesday.May)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_May)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6223 -1.4411  0.0314  1.5604  5.2216
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -27.15303   73.73032  -0.368    0.715
## year4         0.02196    0.03689   0.595    0.556
##
## Residual standard error: 2.522 on 32 degrees of freedom
## Multiple R-squared:  0.01095,    Adjusted R-squared:  -0.01995
## F-statistic: 0.3544 on 1 and 32 DF,  p-value: 0.5558
```

```r
lm.Tuesday.June <- lm(temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_June)
summary(lm.Tuesday.June)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_June)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -6.0339 -1.5343 -0.0279  1.9180  6.7676
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.1373460 50.7897026   0.416    0.678
## year4       -0.0002531  0.0254253  -0.010    0.992
##
## Residual standard error: 2.621 on 80 degrees of freedom
## Multiple R-squared:  1.239e-06,  Adjusted R-squared:  -0.0125
## F-statistic: 9.912e-05 on 1 and 80 DF,  p-value: 0.9921
```

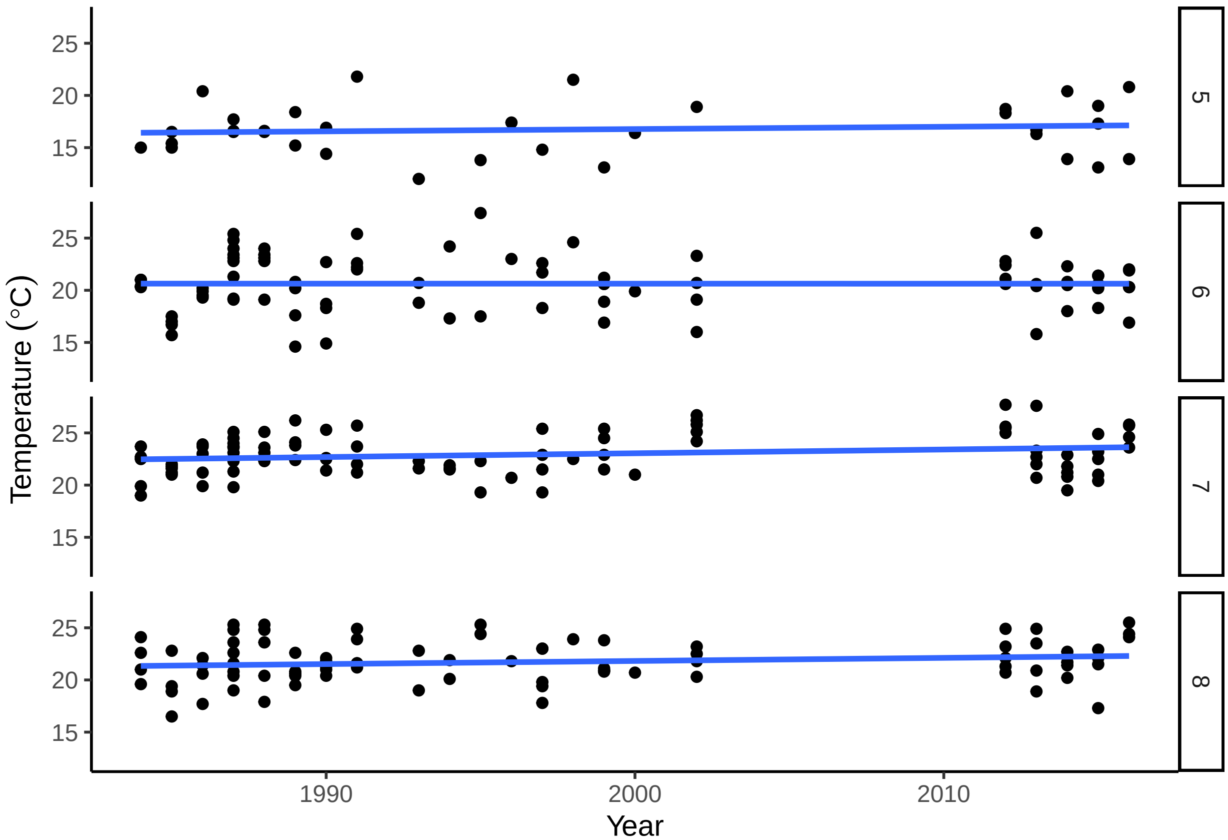
```r
lm.Tuesday.July <- lm(temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_July)
summary(lm.Tuesday.July)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_July)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0561 -1.3275 -0.2047  1.4031  4.2161
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -49.18776   37.36614  -1.316   0.1916
## year4         0.03612    0.01871   1.931   0.0569 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.953 on 84 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.04248,    Adjusted R-squared:  0.03109
## F-statistic: 3.727 on 1 and 84 DF,  p-value: 0.05691
```

```r
lm.Tuesday.August <- lm(temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_August)
summary(lm.Tuesday.August)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = NTLdataTuesday.Summer.Surface_August)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9656 -1.1055 -0.0787  1.2820  3.8677
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -37.70343   41.36954  -0.911    0.365
## year4         0.02976    0.02072   1.436    0.155
##
## Residual standard error: 2.025 on 81 degrees of freedom
## Multiple R-squared:  0.02484,    Adjusted R-squared:  0.0128
## F-statistic: 2.063 on 1 and 81 DF,  p-value: 0.1547
```

```r
Tempchange.plot <-
  ggplot(NTLdataTuesday.Summer.Surface, aes(x = year4, y = temperature_C)) +
  geom_point() +
  geom_smooth(se=FALSE, method =lm) +
  facet_grid(rows=vars(Month)) +
  labs(x = "Year", y = expression("Temperature "(degree*C)))
  print(Tempchange.plot)
```



12. How do your results compare to those we found in class for Paul Lake? Do similar trends exist for both lakes?

    A linear regression was performed to analyze if surface lake temperatures for a given month have increased over time for Tuesday Lake. The null hypothesis is: There has been no variation of Tuesday Lake's surface temperatures in a given month from 1984 to 2016. The alternate hypothesis is: There has been a variation of Tuesday Lake's surface temperatures in a given month from 1984 to 2016.

    According to the results for Tuesday Lake:

    For the month of May we get a p-value of $0.556 > 0.05$; therefore, we don't reject the null hypothesis with a 5% level of significance (F-statistic: 0.3544 on 1 and 32 DF, p-value: 0.5558). There has been no variation of Tuesday Lake's surface temperatures in May from 1984 to 2016.

    For the month of June we get a p-value of $0.992 > 0.05$; therefore, we don't reject the null hypothesis with a 5% level of significance (F-statistic: 9.912e-05 on 1 and 80 DF, p-value: 0.9921). There has been no variation of Tuesday Lake's surface temperatures in June from 1984 to 2016.

    For the month of July we get a p-value of $0.0569 > 0.05$; therefore, we don't reject the null hypothesis with a 5% level of significance (F-statistic: 3.727 on 1 and 84 DF, p-value: 0.05691).

There has been no variation of Tuesday Lake's surface temperatures in July from 1984 to 2016. Nevertheless, the p-value is close to the level of significance, so it could be argued that it is worth studying the results of the linear regression. The coefficient of the linear regression is equal to 0.03612 which means that for every year there is an increase of 0.03612°C. This means that over the period of study, the lake has warmed 0.03612°C * 33 years = 1.2°C.

For the month of August we get a p-value of $0.1547 > 0.05$; therefore, we don't reject the null hypothesis with a 5% level of significance (F-statistic: 2.063 on 1 and 81 DF, p-value: 0.1547). There has been no variation of Tuesday Lake's surface temperatures in August from 1984 to 2016.

The same analysis was performed for Paul Lake in the "Long-term change in temperature" section of day 4 lesson. Like Tuesday lake, for the months of May and June the p-values are > than 0.05. For July and August the p-values are < than 0.05 (For July F-statistic: 13.2 on 1 and 148 DF, p-value: 0.0003852 and for August F-statistic: 6.521 on 1 and 137 DF, p-value: 0.01176). The coefficient for July is 0.06007 which means that for every year there is an increase of 0.06007°C. This means that over the period of study, the lake has warmed 0.06007°C * 33 years = 2.18°C. The coefficient for August is 0.04051 which means that for every year there is an increase of 0.04051°C. This means that over the period of study, the lake has warmed 0.04051°C * 33 years = 1.33°C. Paul lake shows significant trends for July and August. Tuesday lake shows only an almost significant trend for July. In both lakes the significant (or almost significant) trends are positive (Temperature is increasing with time) and with similar values although Paul lake shows higher trend values.