Assignment 2: Physical Properties of Lakes

Caroline Watson

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)

```
#checking working directory
getwd()
```

```
## [1] "/Users/carolinewatson/Documents/Fall 2019/Hydrologic_Data_Analysis/Assignments"
```

```
#loading packages
suppressMessages(library(tidyverse))
library(lubridate)
library(cowplot)

#reading in NTL-LTER physical lake dataset
NTLdata <- read.csv("../Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")

#setting date column to date format
NTLdata$sampledate <- as.Date(NTLdata$sampledate, "%m/%d/%y", origin = "1970-01-01")

#setting ggplot theme
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).

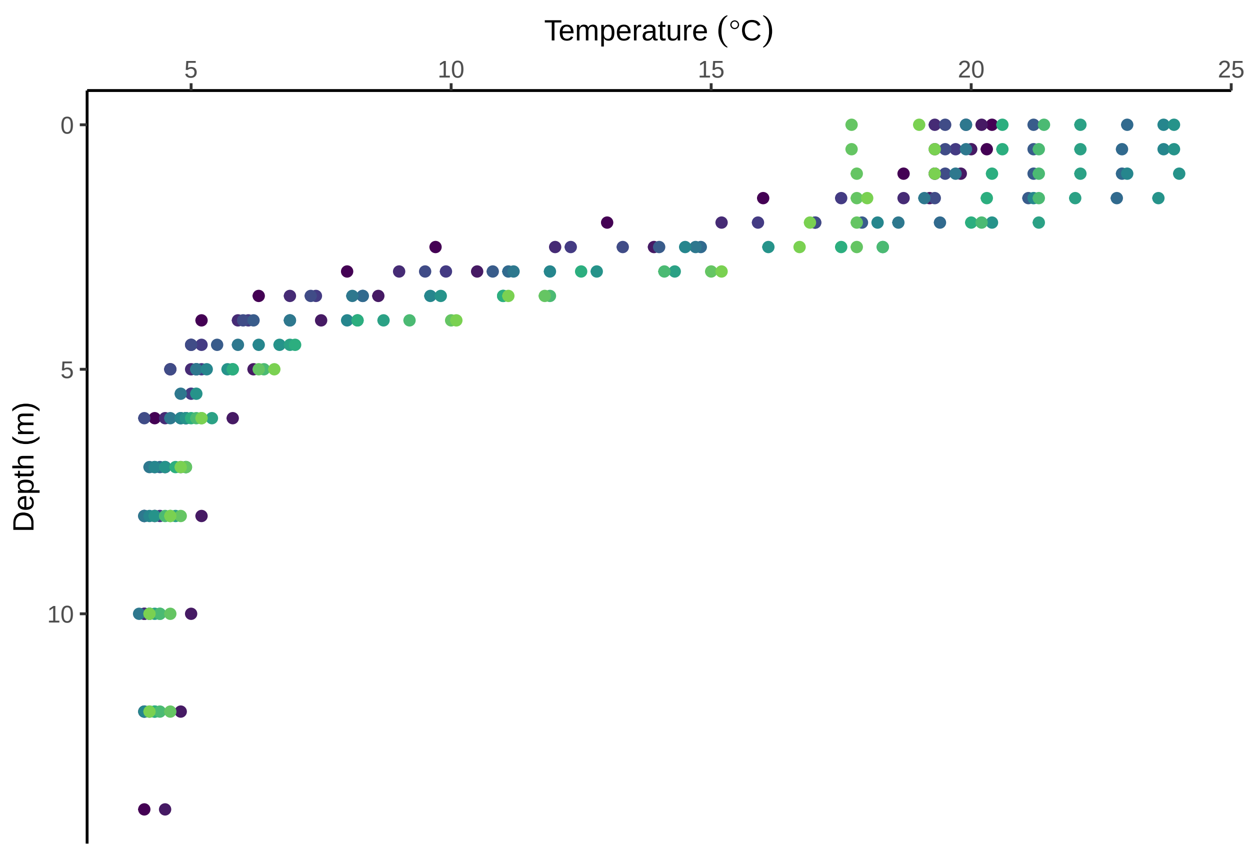
```r
#Wrangling dataset to include only Tuesday Lake from 1986 and 2015
Tuesdaydata <- NTLdata %>%
  filter(lakename == "Tuesday Lake")

Tuesdaydata_1986 <- filter(Tuesdaydata, year4 == 1986)
Tuesdaydata2015 <- filter(Tuesdaydata, 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.

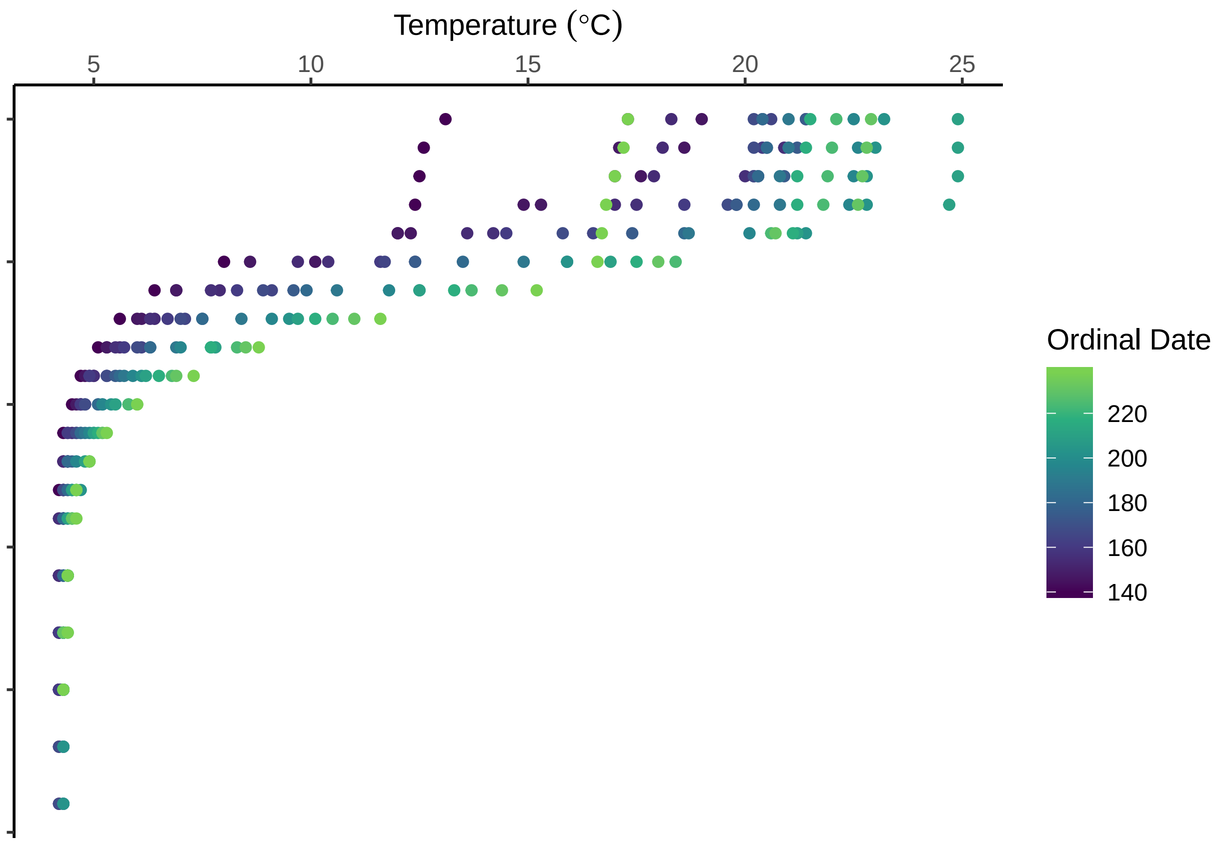
Remember to edit your graphs so they follow good data visualization practices.

```r
#temperature profile for Tuesday Lake in 1986
Temp1986 <-
  ggplot(Tuesdaydata_1986, 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 = "viridis") +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)", color = "Ordinal Date") +
  theme(legend.position = "none")
print(Temp1986)
```

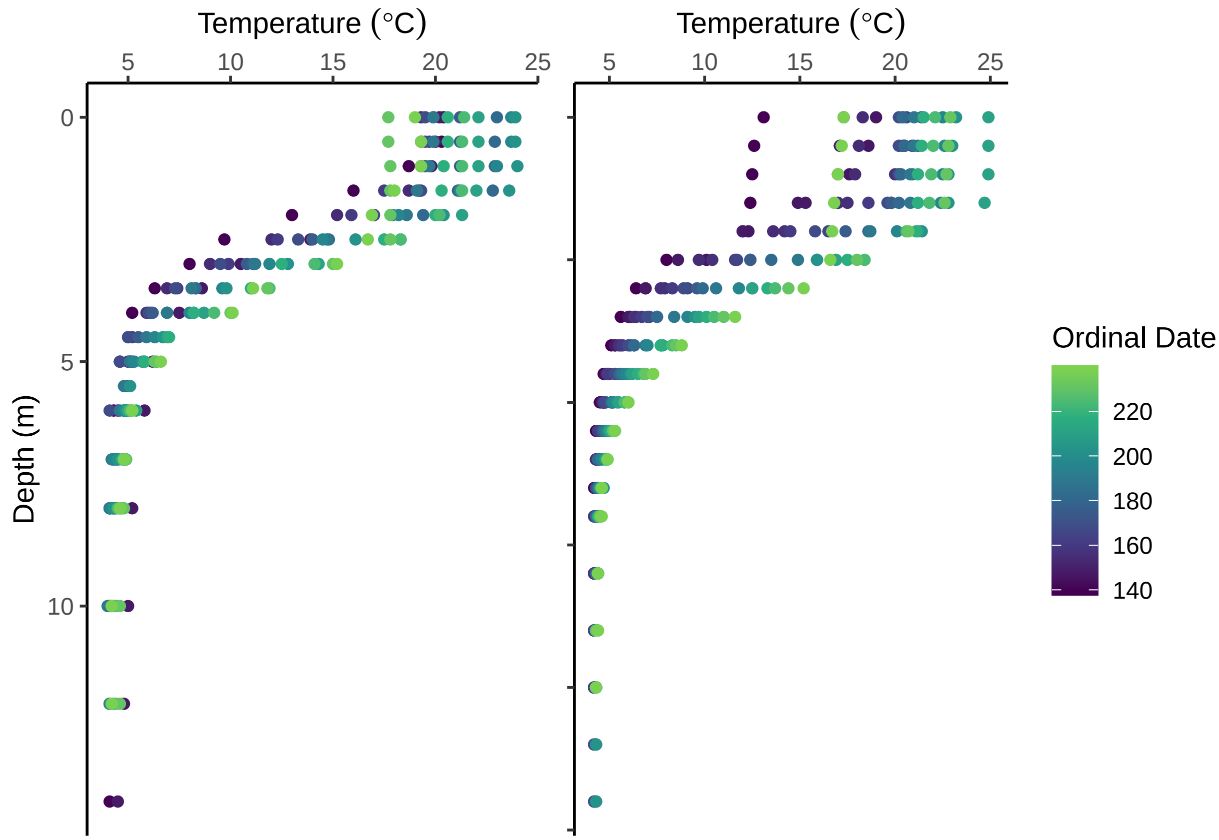



```r
#temperature profile for Tuesday Lake in 2016
Temp2015 <-
  ggplot(Tuesdaydata2015, aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
```

```
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "viridis") +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)", color = "Ordinal Date") +
  theme(axis.text.y = element_blank(), axis.title.y = element_blank()) +
  theme(legend.position = "right")
print(Temp2015)
```




```
#plot_grid of the two temperature profile graphs
TuesdayLake <-
  plot_grid(Temp1986, Temp2015,
            ncol = 2, rel_widths = c(1, 1.25))
print(TuesdayLake)
```



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 graphs the temperature is highest in the lake towards the surface of the lake and as the depth increases, the temperatures decrease as well. Temperatures at the eplimnion were always pretty high throughout the year, but are warmest during summer months. Also, temperature inreases throughout the season and then starts to decrease in the fall and winter. Yes, there are differences between the two years, most notably that in the early part of 2015 (between day 1 and day 160), the water at the eplimnion was much colder compared to 1986. It also looks like there may have been multiple samples taken around the same time frame, but that doesn't explain why some samples were colder than others even when taken at the same time. The one thing that could explain this is if samples were taken at different locations.

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.

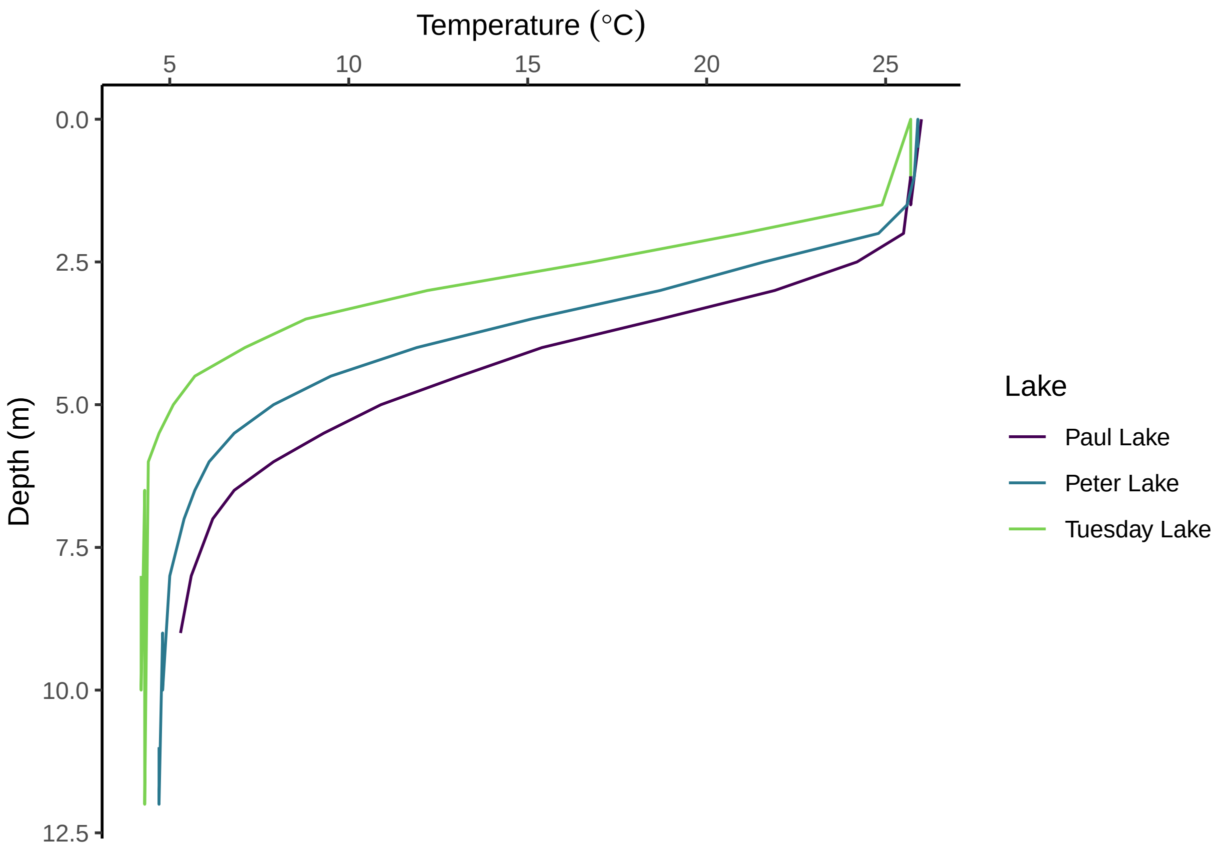
```
#wrangle original dataset to just get data for July 25, 26, 27 in 2016
NTL_July_2016 <- NTLdata %>%
  filter(sampledate == "2016-07-25" | sampledate == "2016-07-26" | sampledate == "2016-07-27")
```

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

```
#temperature profile by depth for each lake
tempJuly2016 <- ggplot(NTL_July_2016, aes (x = temperature_C, y = depth, color = lakename)) +
  geom_line() +
```

```r
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_d(end = 0.8, option = "brewer blues") +
  labs(x = expression("Temperature "(degree*C)), y = "Depth (m)", color = "Lake")


print(tempJuly2016)
```



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

    The depth range of the epilimnion in each lake is: Tuesday Lake: 0.0m - 2.3m, Peter Lake: 0.0m - 2.5m, Paul Lake: 0.0m - 2.6m. For the thermocline, the depth ranges for each lake are: Tuesday Lake: 2.4m - about 4.8m, Peter Lake: 2.5 m - 5.0m, Paul Lake: 2.6m - 5.0m. For the hypolimnion, the depth ranges for each lake are: Tuesday Lake: 4.8m - 12.3m, Peter Lake: 5.0m - 12.3m, and Paul Lake: 7.0m - 9m.

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
#adding a month column to Tuesday Lake dataframe
Tuesdaydata <- Tuesdaydata %>%
  mutate(Month = month(sampledate))

#filtering data frame
```

```r
Tuesdaydata_skinny <- Tuesdaydata %>%
  filter(depth == 0.0 & Month %in% c("5", "6", "7", "8"))

#creating 4 seperate dataframes for each month
Tuesdaydata_skinny_May <- Tuesdaydata_skinny %>%
  filter(Month == "5")

Tuesdaydata_skinny_June <- Tuesdaydata_skinny %>%
  filter(Month == "6")

Tuesdaydata_skinny_July <- Tuesdaydata_skinny %>%
  filter(Month == "7")

Tuesdaydata_skinny_Aug <- Tuesdaydata_skinny %>%
  filter(Month == "8")

#running linear regression on each dataframe
Maytest <- lm(data = Tuesdaydata_skinny_May, temperature_C ~ year4)
summary(Maytest) #no significant temp change in May accross years
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Tuesdaydata_skinny_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
Junetest <- lm(data = Tuesdaydata_skinny_June, temperature_C ~ year4)
summary(Junetest) #no significant temp change in May accross years
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Tuesdaydata_skinny_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
Julytest <- lm(data = Tuesdaydata_skinny_July, temperature_C ~ year4)
summary(Julytest) #no significant temp change
```

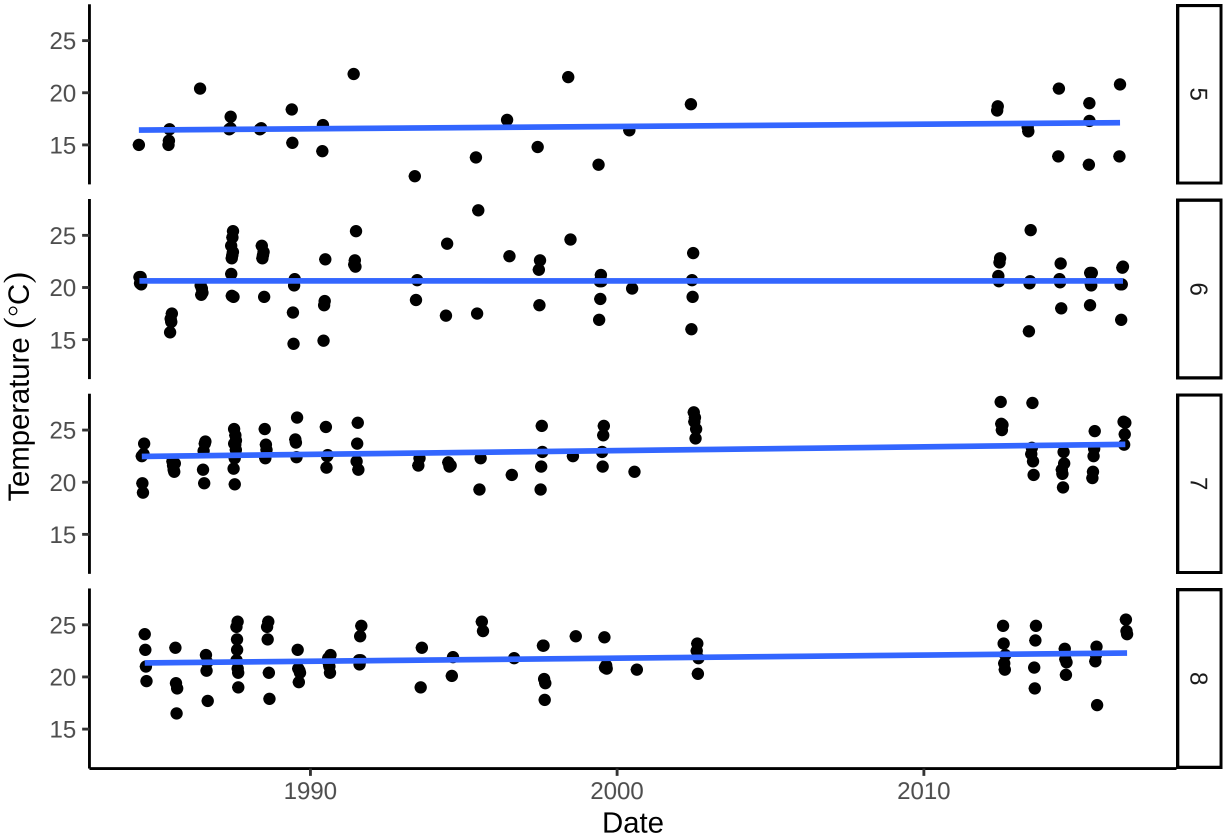
```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Tuesdaydata_skinny_July)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0561 -1.3275 -0.2047  1.4031  4.2161
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -49.18776   37.36614  -1.316   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
Augtest <- lm(data = Tuesdaydata_skinny_Aug, temperature_C ~ year4)
summary(Augtest) #no significant temp change
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Tuesdaydata_skinny_Aug)
##
## 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
#creating ggplot of surface temperature by date
surface_ggplot <- ggplot(Tuesdaydata_skinny, aes(x = sampledate, y = temperature_C)) +
  geom_point() +
  geom_smooth(se = FALSE, method = lm) +
  facet_grid(rows = vars(Month)) +
  labs(x = "Date", y = expression("Temperature "(degree*C)))

print(surface_ggplot)
```



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

    The results for Tuesday Lake are not statistically significant, which is different than the results for Paul Lake. When running this analysis for Paul Lake, it was determined that there was a significant increase in temperature in July of 0.06 degrees each year and August of 0.04 degrees each year. Although the results from Tuesday Lake were not statistically significant, the coefficients for July and August in Tuesday Lake were closest to being statisticallly significant. The fact that these months were closest to being statistically significant is the one similarity surface temperatures of Tuesday Lake have in comparison to surface temperatures from May to August in Paul Lake.