Assignment 2: Physical Properties of Lakes

Theo Cai

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)

```r
getwd()
```

```
## [1] "Z:/Hydrologic_Data_Analysis"
```

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

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date
```

```r
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())
## *******************************************************
##
## Attaching package: 'cowplot'
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```r
NTLdata <- read.csv("./Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")
NTLdata$sampledate <- as.Date(NTLdata$sampledate, "%m/%d/%y")

theme_set(theme_dark())
```

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).

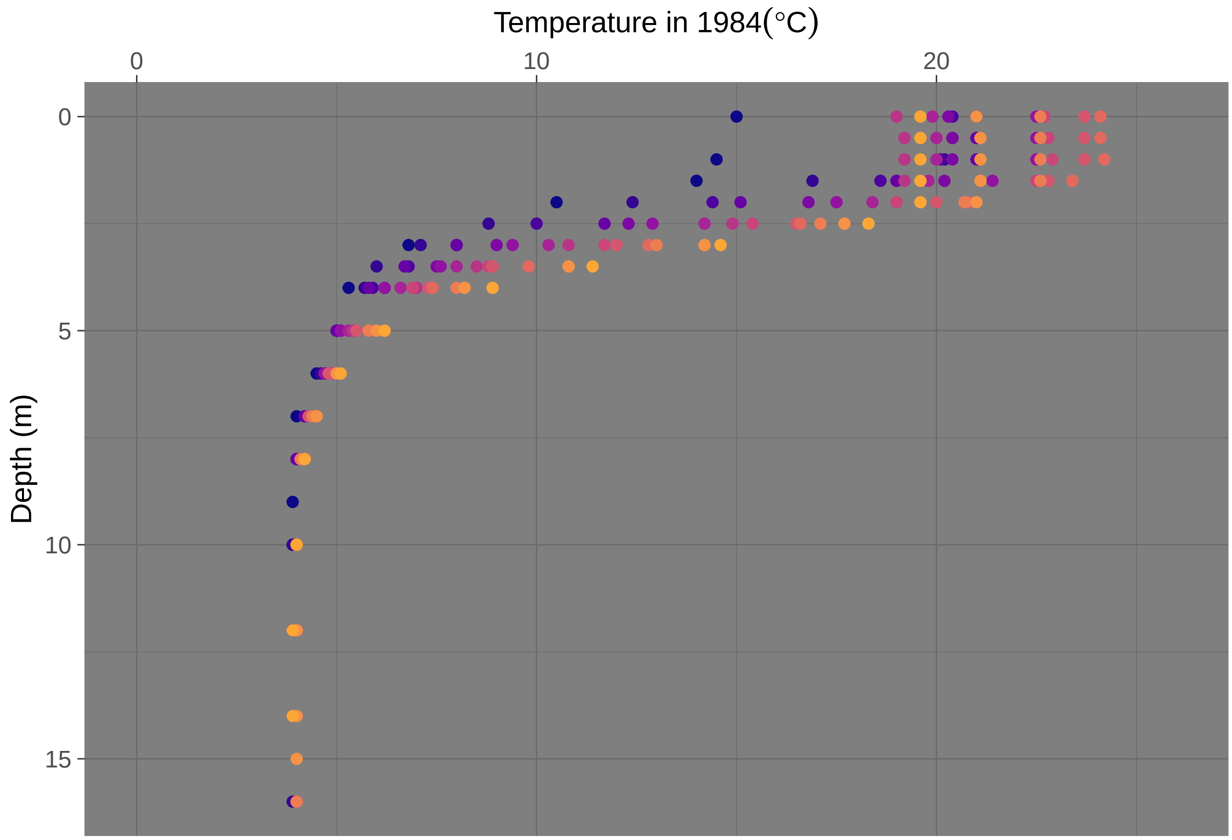
```r
Tuesdaydata <- filter(NTLdata, lakename == "Tuesday Lake")
Tuesdaydata.1984 <- filter(Tuesdaydata, year4 == 1984)
Tuesdaydata.2016 <- filter(Tuesdaydata, year4 == 2016)
```

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.

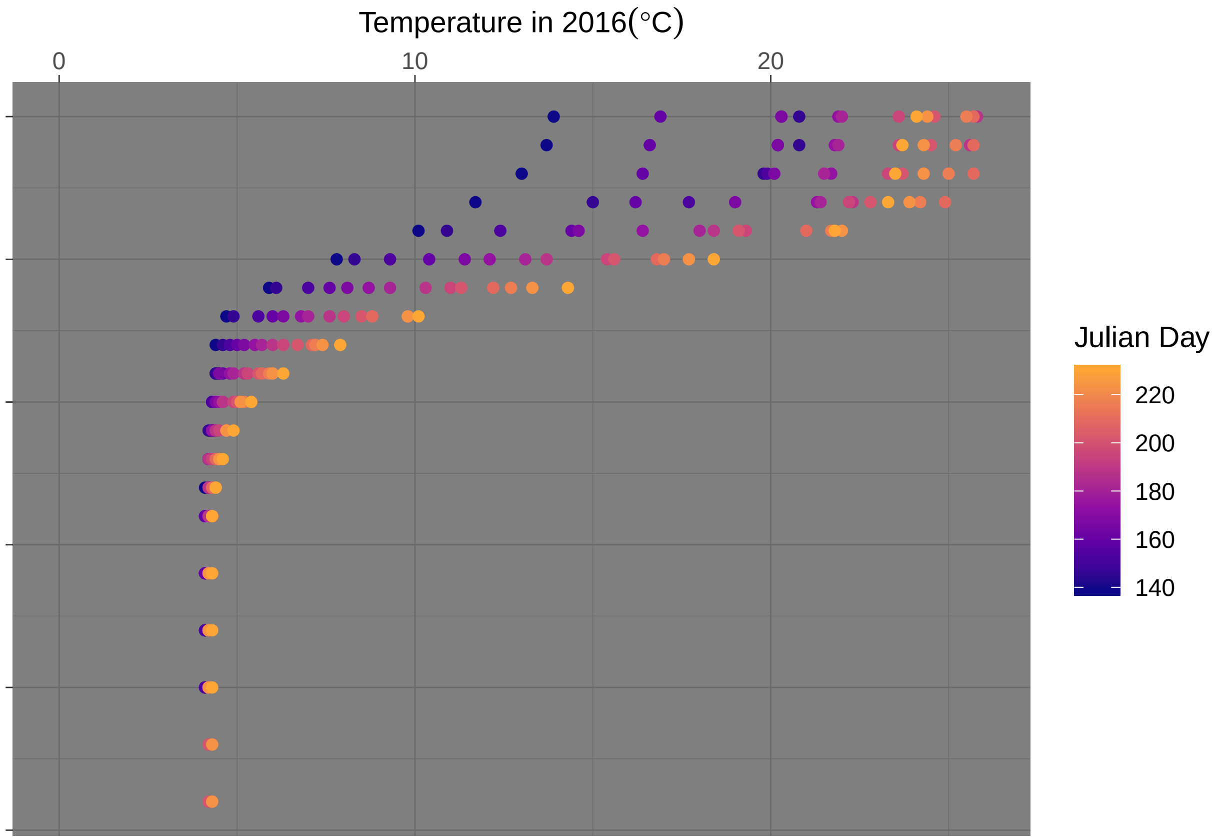
```r
Tempprofiles1984 <-
  ggplot(Tuesdaydata.1984, aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  expand_limits(x = c(0, 26)) +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "plasma") +
  labs(x = expression("Temperature in 1984"(degree*C)), y = "Depth (m)") +
  theme(legend.position = "none")
print(Tempprofiles1984)
```

```
## Warning: Removed 31 rows containing missing values (geom_point).
```




```
Tempprofiles2016 <-
  ggplot(Tuesdaydata.2016, aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  expand_limits(x = c(0, 26)) +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "plasma") +
  labs(x = expression("Temperature in 2016"(degree*C)), y = "Depth (m)", color = "Julian Day") +
  theme(axis.text.y = element_blank(), axis.title.y = element_blank())
print(Tempprofiles2016)
```

```
## Warning: Removed 28 rows containing missing values (geom_point).
```

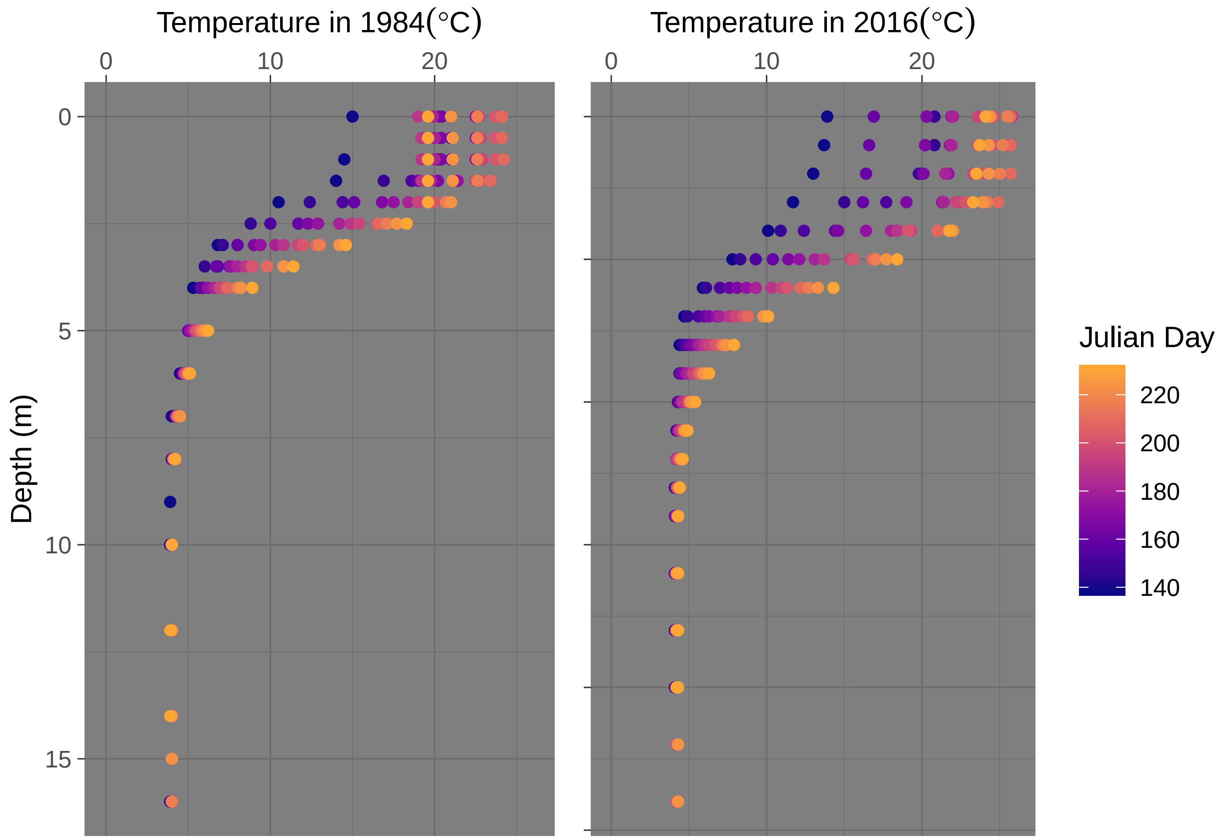



```
Tempprofiles19842016 <-
  plot_grid(Tempprofiles1984, Tempprofiles2016,
            ncol = 2, rel_widths = c(.85, 1))
```

```
## Warning: Removed 31 rows containing missing values (geom_point).
```

```
## Warning: Removed 28 rows containing missing values (geom_point).
```

```
print(Tempprofiles19842016)
```



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

   In general, as the summer season continues, the average temperature of the lake continues to rise, as especially at the surface and in the epilimnion. We do, however, see that the temperature peaks not at the end of that year's season, but somewhere a little before. This implies that the hottest point in the summer drops off somewhere in the month of August, which makes sense considering the transition to fall begins in the same month, thus cooling the lake down.

   I do see a subtle difference between the two years. The lake starts off slightly warmer in early season 1984 than the same time in 2016. However, the max temperature reached in the season is greater in 2016 than in 1984. It seems that the temperature range over the summer season is greater in 2016 than in 1984. Although the peak temperature seemed to have occured around the same time in both years, the subsequent temperature decline is greater and faster in 1984 than in 2016.

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.

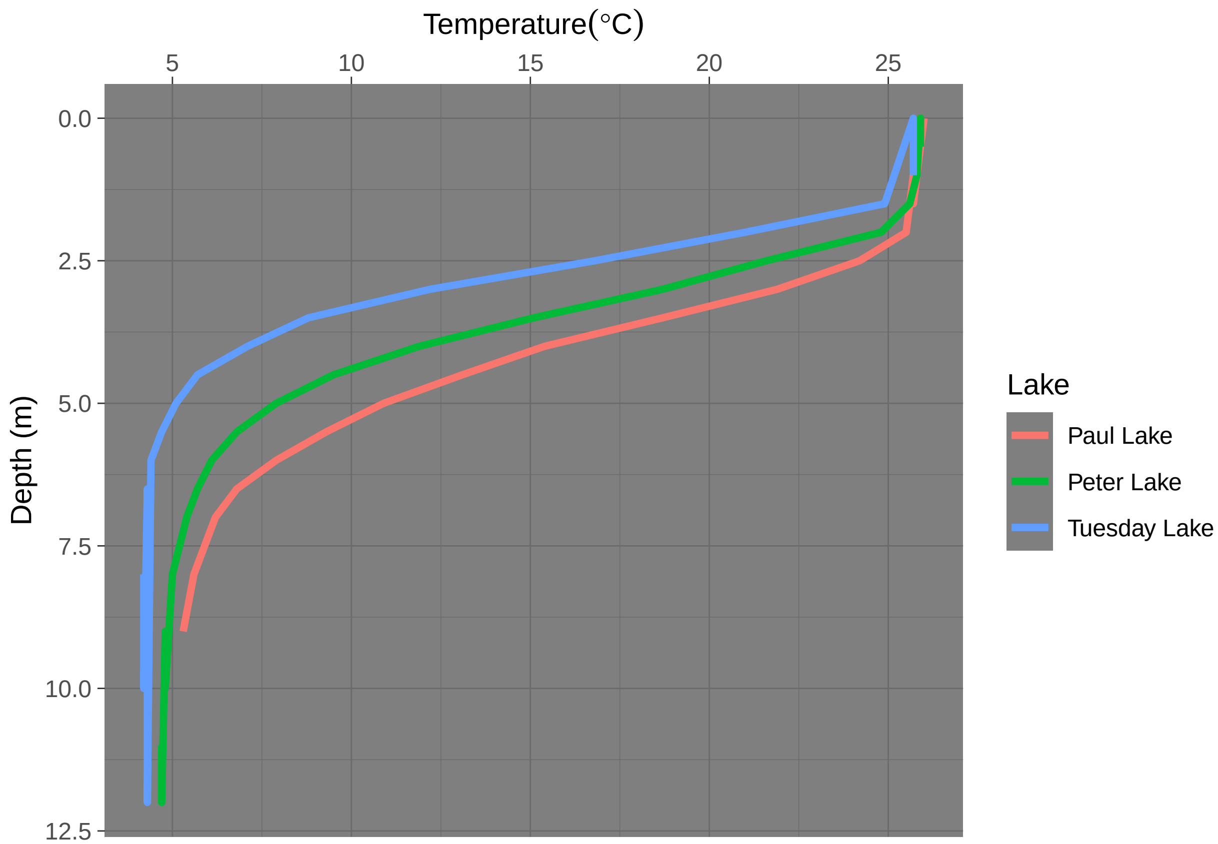
```
NTLJuly <- filter(NTLdata, year4== 2016)
NTLJuly <- filter(NTLJuly, NTLJuly$daynum>206 & NTLJuly$daynum<210)
```

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

```
TempprofilesJuly <-
  ggplot(NTLJuly, aes(x = temperature_C, y = depth, group=lakename, colour=lakename)) +
```

```
  geom_line(size = 1.3) +
  scale_y_reverse() +
  labs(x = expression("Temperature"(degree*C)), y = "Depth (m)", color = "Lake") +
  scale_x_continuous(position = "top")
print(TempprofilesJuly)
```

```
## Warning: Removed 6 rows containing missing values (geom_path).
```



10. What is the depth range of the epilimnion in each lake? The thermocline? The hypolimnion? >Epilimnion > For Paul Lake: 0-2m. For Peter Lake: 0-1.8m. For Tuesday Lake: 0-1.3m.

    Thermocline Paul: 2-6.25m. Peter: 1.8-5m. Tuesday: 1.3-3.3m.

    Hypolimnion Paul: >6.25m. Peter: >5m. Tuesday: >3.3m.

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.

```
Peterdatafilter <- dplyr::filter(NTLdata, lakename == "Peter Lake")

Peterdatafilter2 <- Peterdatafilter %>%
  mutate(month = month(sampledate)) %>%
  filter(depth == 0 & month <9 & month >4)

PeterdatafilterMay <- filter(Peterdatafilter2, month == 5)
```

```r
PeterdatafilterJune <- filter(Peterdatafilter2, month == 6)
PeterdatafilterJuly <- filter(Peterdatafilter2, month == 7)
PeterdatafilterAugust <- filter(Peterdatafilter2, month == 8)

Temptrend.May <- lm(data = PeterdatafilterMay, temperature_C ~ year4)
summary(Temptrend.May)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = PeterdatafilterMay)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2960 -1.4047 -0.0873  1.5605  6.5344
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -35.69596   70.65403  -0.505    0.615
## year4         0.02609    0.03531   0.739    0.463
##
## Residual standard error: 2.895 on 61 degrees of freedom
## Multiple R-squared:  0.008867,   Adjusted R-squared:  -0.007381
## F-statistic: 0.5457 on 1 and 61 DF,  p-value: 0.4629
```

```r
Temptrend.June <- lm(data = PeterdatafilterJune, temperature_C ~ year4)
summary(Temptrend.June)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = PeterdatafilterJune)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1170 -1.5844 -0.0472  1.9932  6.9684
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.696621  43.622161   0.131    0.896
## year4        0.007286   0.021813   0.334    0.739
##
## Residual standard error: 2.572 on 142 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.000785,   Adjusted R-squared:  -0.006252
## F-statistic: 0.1116 on 1 and 142 DF,  p-value: 0.7389
```

```r
Temptrend.July <- lm(data = PeterdatafilterJuly, temperature_C ~ year4)
summary(Temptrend.July)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = PeterdatafilterJuly)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4170 -1.0943  0.0191  1.2393  3.7861
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -104.75590   30.74789  -3.407 0.000846 ***
## year4          0.06392    0.01538   4.157 5.44e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.817 on 148 degrees of freedom
## Multiple R-squared:  0.1045, Adjusted R-squared:  0.09849
## F-statistic: 17.28 on 1 and 148 DF,  p-value: 5.444e-05
```

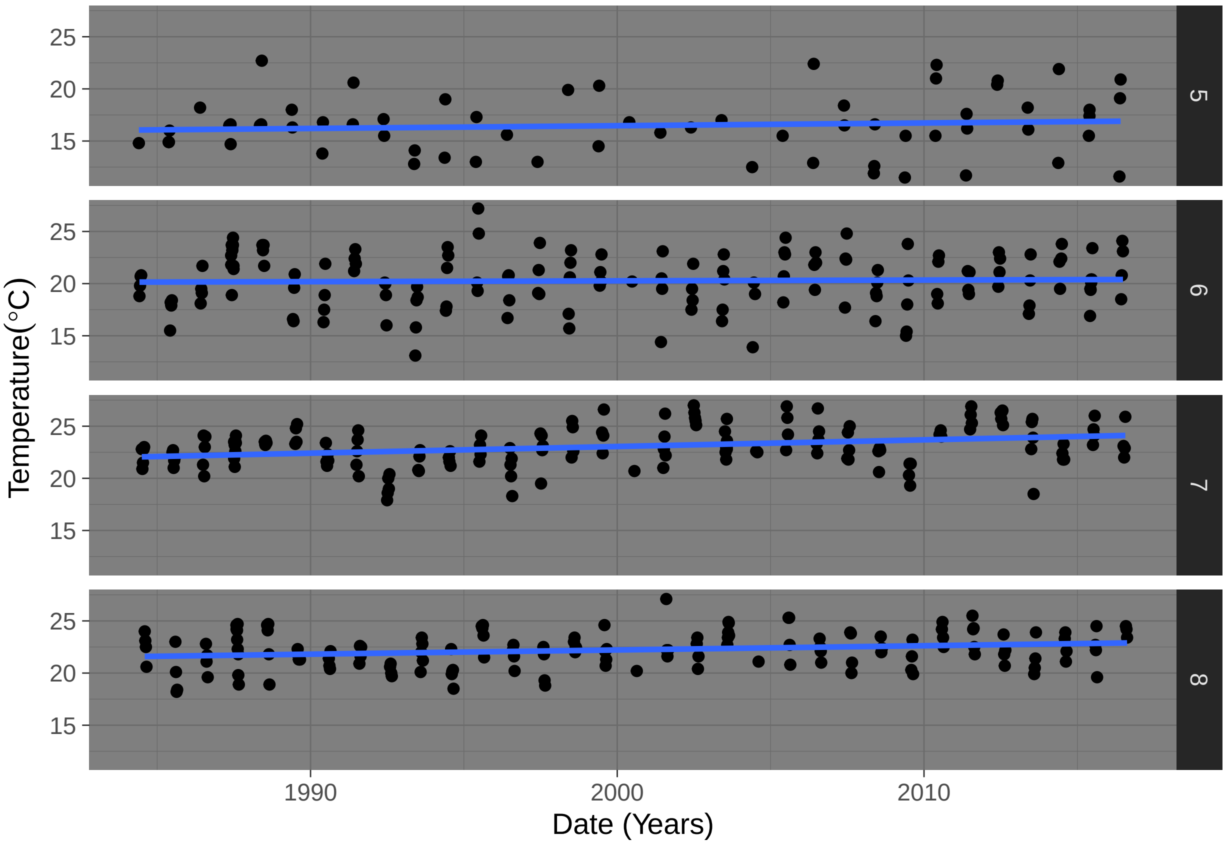
```r
Temptrend.August <- lm(data = PeterdatafilterAugust, temperature_C ~ year4)
summary(Temptrend.August)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = PeterdatafilterAugust)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.497 -1.218  0.016  1.185  4.819
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -59.00883   29.69264  -1.987  0.04888 *
## year4         0.04062    0.01485   2.735  0.00706 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.683 on 137 degrees of freedom
## Multiple R-squared:  0.05178,    Adjusted R-squared:  0.04486
## F-statistic: 7.481 on 1 and 137 DF,  p-value: 0.007061
```

```r
#Peter Lake has warmed an average of 0.034 degrees Celsius, with the most significant increase occuring

ggplot(Peterdatafilter2, aes(x = sampledate, y = temperature_C)) +
  geom_point() +
  facet_grid(rows = vars(month)) +
  geom_smooth(se = FALSE, method = lm) +
  labs(x = "Date (Years)", y =  expression("Temperature"(degree*C)))
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
## Warning: Removed 1 rows containing missing values (geom_point).
```



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 Peter Lake are extremely similar to the results for Paul Lake. Although the average warming was lower for Paul Lake (increase of 0.028 degrees Celsius), it's notable that both lakes saw an increase at all. Additionally, the most increase was seen in the same month for both lakes: July. The significance of the July increase is lower for Paul than for Peter Lake, but both P-values are near zero, making them both significant increases nonetheless.