Assignment 2: Physical Properties of Lakes

Yixin Wem

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)

```
# Check working directory
getwd()
```

```
## [1] "/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Assignments
```

```
# load packages
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------- tidyverse 1.2.1

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

```
library(lubridate)
```

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

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

```
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
# Load data
NTLdata <-
  read.csv(paste0("/Users/yixinwen/Box/Duke/2019 Fall/Hydrologic Data ", "Analysis/Hydrologic_Data_Anal

readr::read_csv
```

```
## function (file, col_names = TRUE, col_types = NULL, locale = default_locale(),
##     na = c("", "NA"), quoted_na = TRUE, quote = "\"", comment = "",
##     trim_ws = TRUE, skip = 0, n_max = Inf, guess_max = min(1000,
##         n_max), progress = show_progress(), skip_empty_rows = TRUE)
## {
##     tokenizer <- tokenizer_csv(na = na, quoted_na = quoted_na,
##         quote = quote, comment = comment, trim_ws = trim_ws,
##         skip_empty_rows = skip_empty_rows)
##     read_delimited(file, tokenizer, col_names = col_names, col_types = col_types,
##         locale = locale, skip = skip, skip_empty_rows = skip_empty_rows,
##         comment = comment, n_max = n_max, guess_max = guess_max,
##         progress = progress)
## }
## <bytecode: 0x7fe3a1413c08>
## <environment: namespace:readr>
```

```r
# Set the date column to the date format
NTLdata$sampledate <- as.Date(NTLdata$sampledate, "%m/%d/%y") # "%m/%d/%y" dates format

# set 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
# see how many years and categories of lakes included
summary(NTLdata$lakename)
```

```
## Central Long Lake     Crampton Lake    East Long Lake  Hummingbird Lake
##               539              1234              3905               430
##         Paul Lake        Peter Lake      Tuesday Lake         Ward Lake
##             10325             11288              6107               598
```

```
##    West Long Lake
##              4188
```

```r
unique(NTLdata$year4)
```

```
##  [1] 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997
## [15] 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
## [29] 2012 2013 2014 2015 2016
```

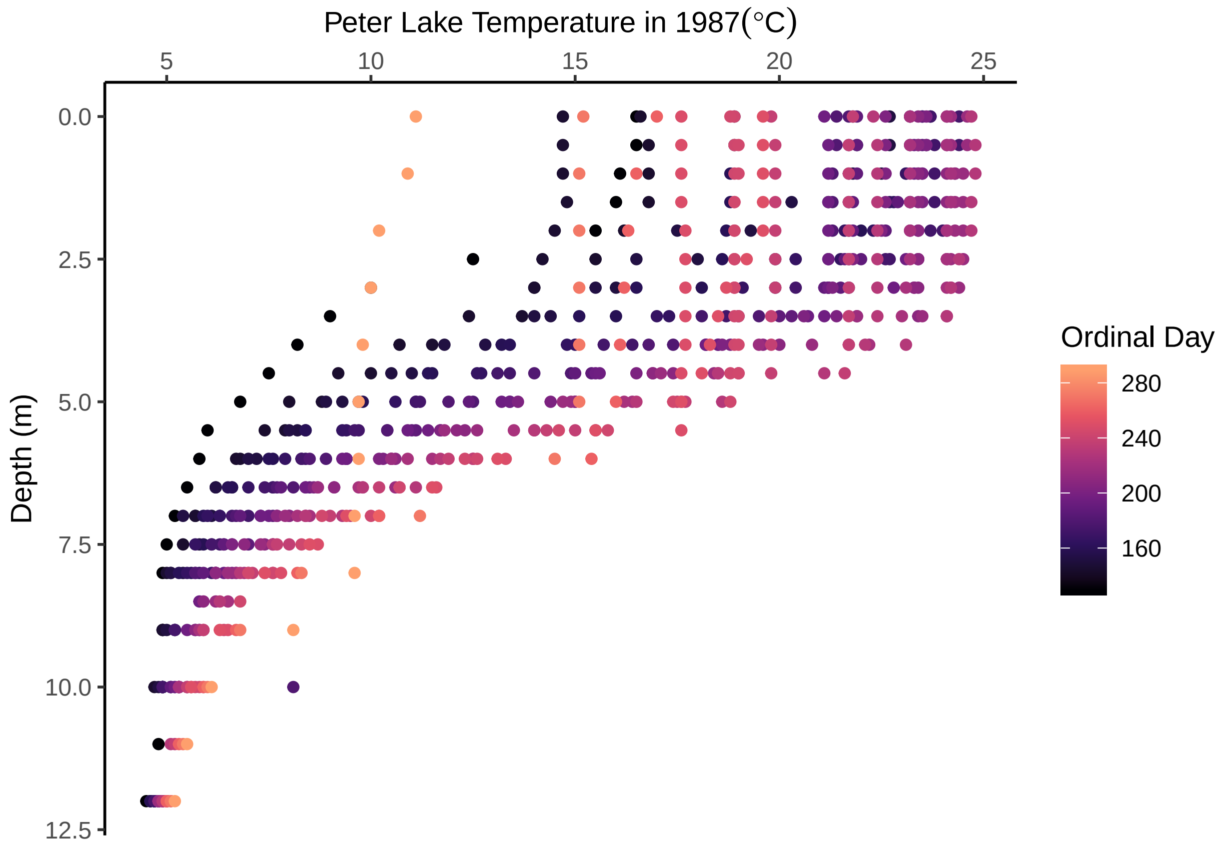
```r
# choose Peter Lake with year 1987 and 2014
Peterlake <- NTLdata %>%
  filter(lakename == "Peter Lake" & year4 %in% c(1987,2014) )
```

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.

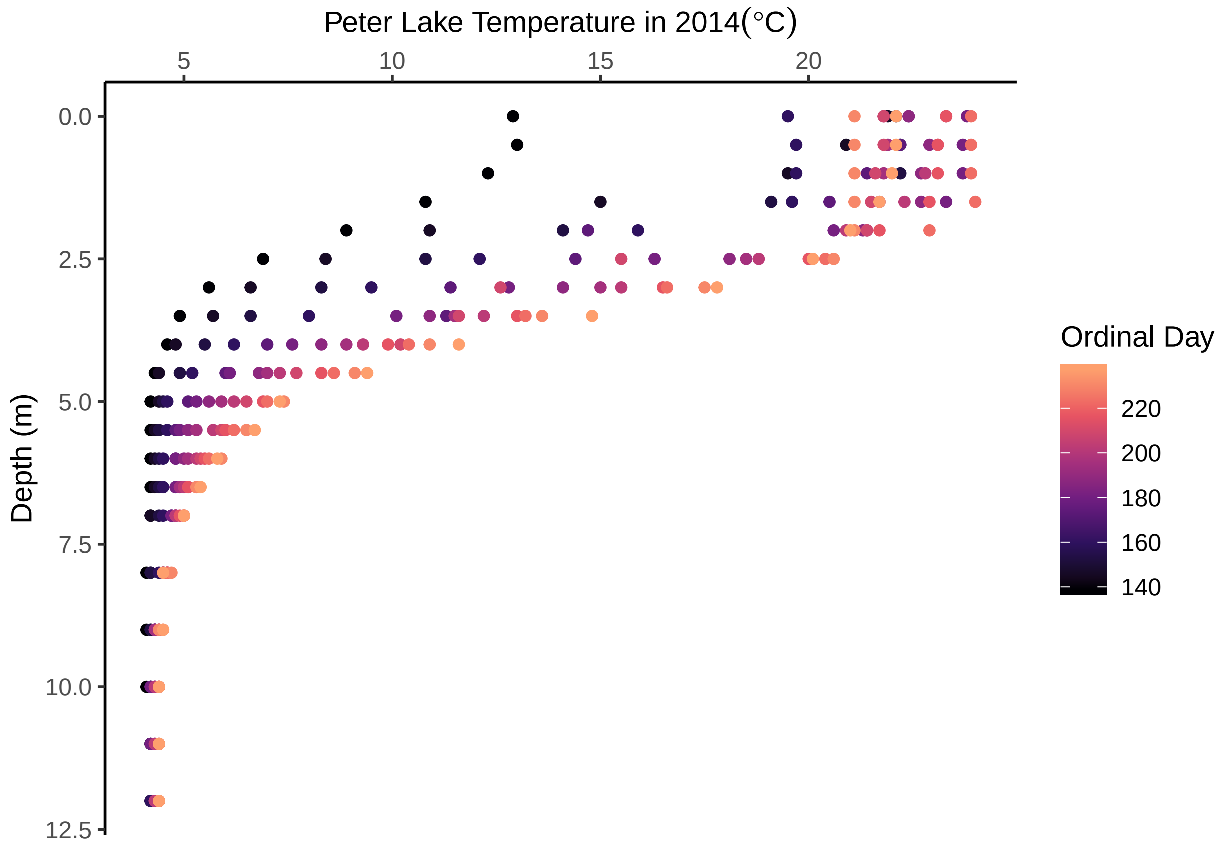
```r
# Temperature profiles for Peter 1987
Peter1987 <- Peterlake %>%
  filter(year4==1987)

Tempprofiles1987 <-
  ggplot(Peter1987, 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("Peter Lake Temperature in 1987"(degree*C)),
       y = "Depth (m)",color = "Ordinal Day")
print(Tempprofiles1987)
```

```
## Warning: Removed 70 rows containing missing values (geom_point).
```




```r
# Temperature profiles for Peter 2014
Peter2014 <- Peterlake %>%
  filter(year4==2014)

Tempprofiles2014 <-
  ggplot(Peter2014, 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("Peter Lake Temperature in 2014"(degree*C)),
       y = "Depth (m)",color = "Ordinal Day")
print(Tempprofiles2014)
```

```
## Warning: Removed 50 rows containing missing values (geom_point).
```

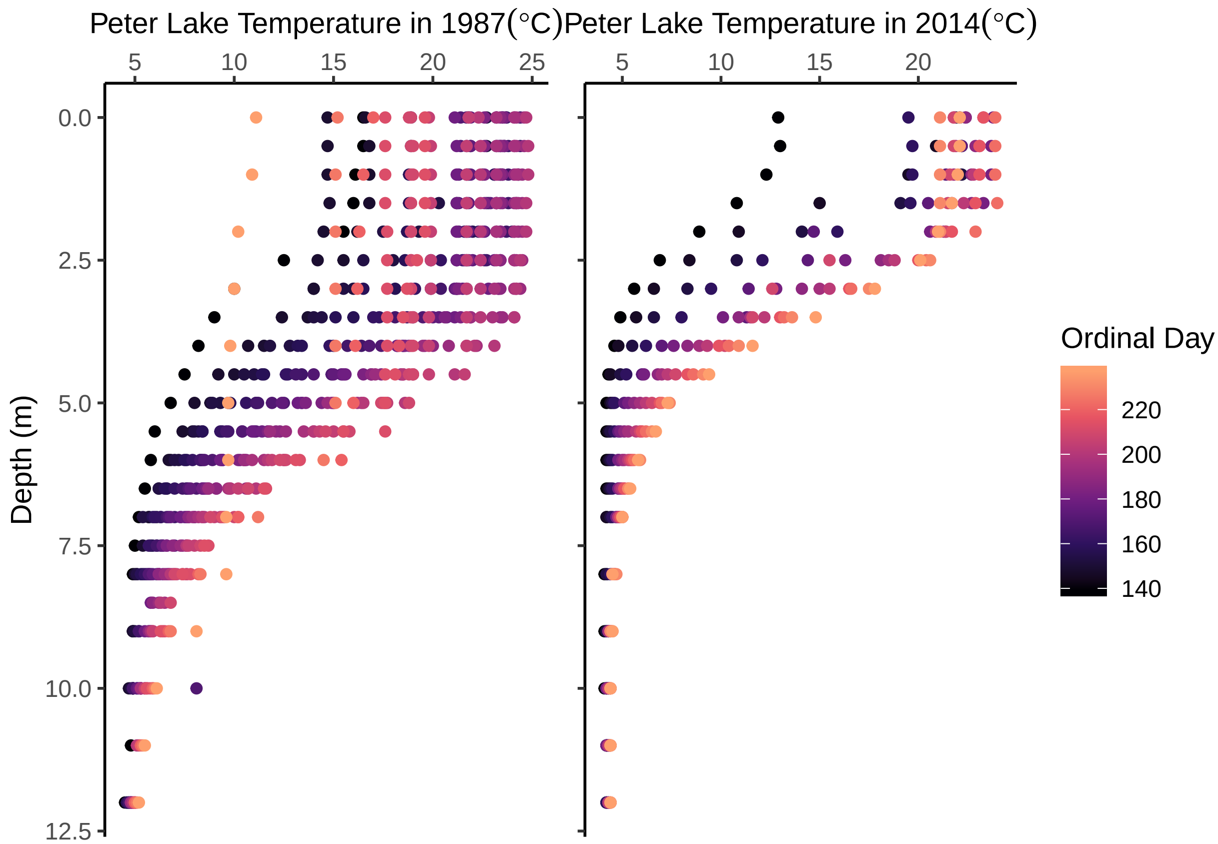



```
# plot_grid of 2 graphs
Tempprofiles1987 <-
  ggplot(Peter1987, aes(x = temperature_C, y = depth, color = daynum)) +
  geom_point() +
  scale_y_reverse() +
  scale_x_continuous(position = "top") +
  scale_color_viridis_c(end = 0.8, option = "magma") +
  labs(x = expression("Peter Lake Temperature in 1987"(degree*C)),
       y = "Depth (m)",color = "Ordinal Day") +
  theme(legend.position = "none")

Tempprofiles2014 <-
  ggplot(Peter2014, 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("Peter Lake Temperature in 2014"(degree*C)),
       y = "Depth (m)",color = "Ordinal Day") +
  theme(axis.text.y = element_blank(), axis.title.y = element_blank())

Peterprofiles <-
  plot_grid(Tempprofiles1987, Tempprofiles2014,
            ncol = 2, rel_widths = c(1,1.2))
```

```
## Warning: Removed 70 rows containing missing values (geom_point).
```

```
## Warning: Removed 50 rows containing missing values (geom_point).
```

```
print(Peterprofiles)
```



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

   In 1987, the stratification of Peter Lake gets stronger from May to June, and becomes weaker from June to August.In May, the surface water becomes warmer, and wind can mix the layers. In September, the epilimnion cools and becomes more dense and sinks to the hypolimnion, which can mix the layer with the help of wind, and thus the temperature difference is small. In 2014, the stratification of Peter Lake became stronger from May to July, and became slightly weaker from July to August. In 1987, the stratification of Peter Lake was weak in July and August due to the wind and water density, while in 2014, the stratification of Peter Lake was much more obvious in July and August compared to that in 1987,this may because the summer lasts longer in 2014 than in 1987.

Multiple lakes, single date

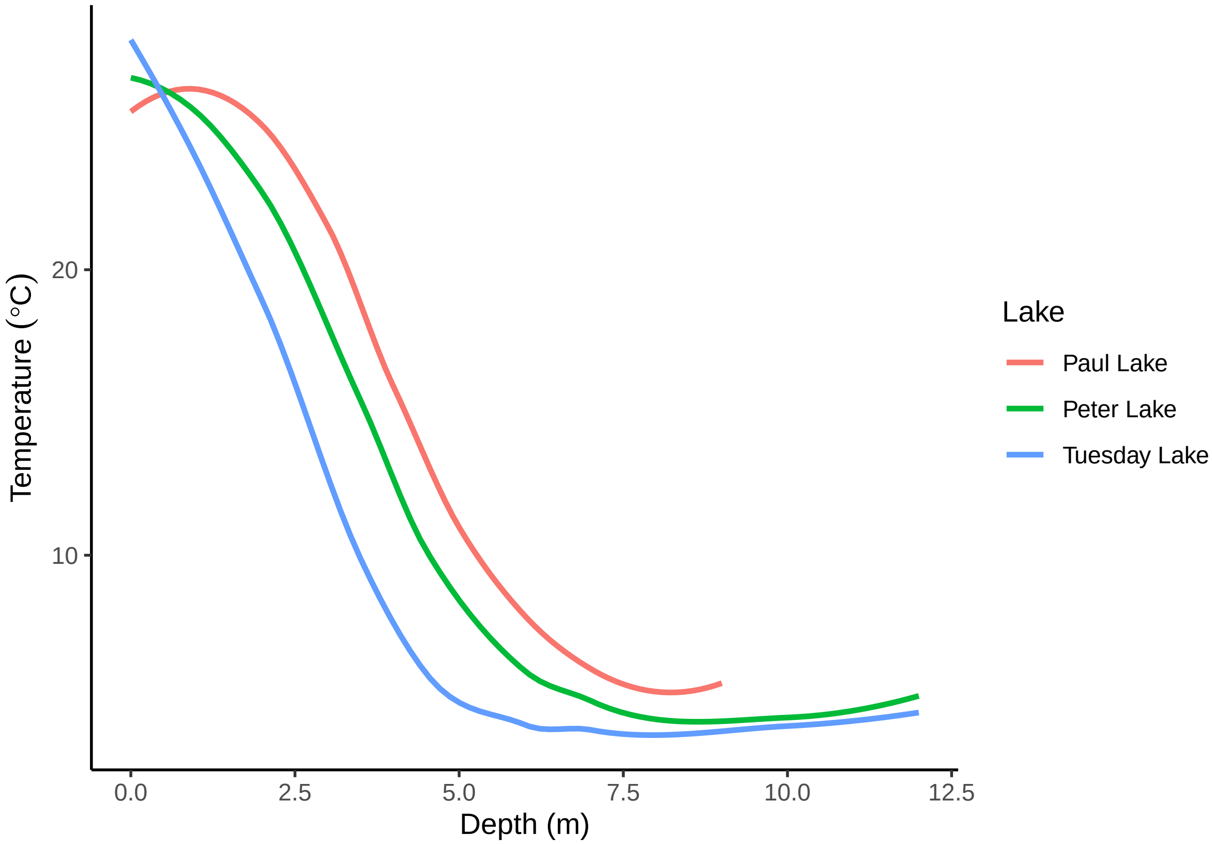
8. On July 25, 26, and 27 in 2016, all three lakes (Peter, Paul, and Tuesday) were sampled. Wrangle your data frame to include just these three dates.

```
Threedays <- NTLdata %>%
  filter(sampledate >= "2016-07-25" & 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.

```r
ggplot(Threedays, aes(x = depth, y = temperature_C, color= lakename)) +
  geom_smooth(se = FALSE)+
  labs(y = expression("Temperature "(degree*C)), x = "Depth (m)",
       color = "Lake")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 6 rows containing non-finite values (stat_smooth).
```



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

    In Paul Lake, the depth range of the epilimnion is about 0-2.5m, the depth of the thermocline is about 2.5-7.5m, the depth range of hypolimnion is above 7.5m. In Peter Lake, the depth range of epilimnion is 0-2m, the depth of the thermocline is about 2-7m, the depth range of hypolimnion is above 7m.In Tuesday Lake, it does not have epilimnion layer, the depth of the thermocline is about 0-5m, the depth range of hypolimnion is above 5m.

Trends in surface temperatures over time.

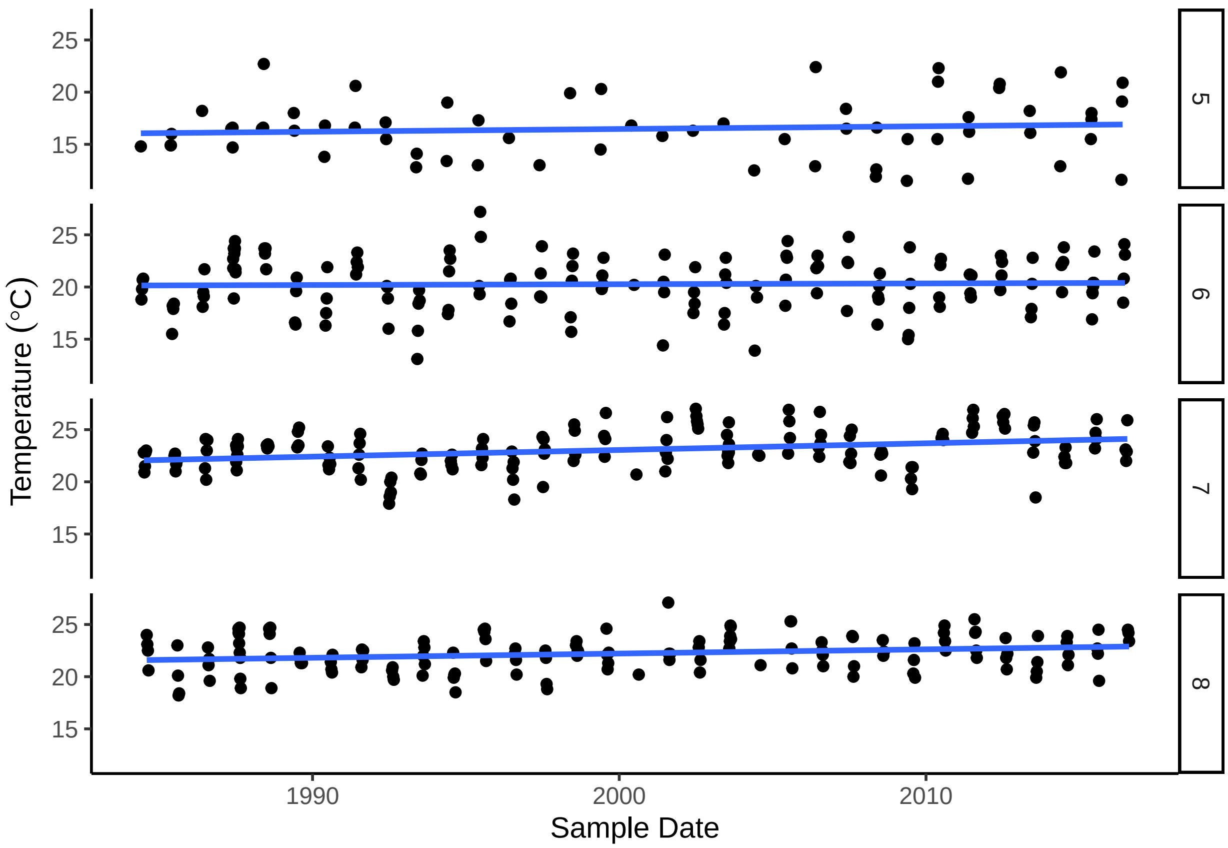
11. Run the same analyses we ran in class to determine if surface lake temperatures for a given month have increased over time ("Long-term change in temperature" section of day 4 lesson in its entirety), this time for either Peter or Tuesday Lake.

```r
Petersurface <- NTLdata %>%
  mutate(Month = month(sampledate)) %>%
  filter(lakename == "Peter Lake" & depth == 0 & Month <= 8 & Month >= 5)

ggplot(Petersurface, aes(x = sampledate, y = temperature_C)) +
```

```r
  geom_point() +
  facet_grid(rows = vars(Month)) +
  geom_smooth(se = FALSE, method = lm)+
  labs(y = expression("Temperature "(degree*C)), x = "Sample Date")
```

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

```
## Warning: Removed 1 rows containing missing values (geom_point).
```




```r
Petersurface.May <- filter(Petersurface, Month == 5)
Petersurface.June <- filter(Petersurface, Month == 6)
Petersurface.July <- filter(Petersurface, Month == 7)
Petersurface.August <- filter(Petersurface, Month == 8)

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

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Petersurface.May)
##
## 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
```

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

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Petersurface.June)
##
## 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
```

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

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Petersurface.July)
##
## 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
```

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

```
##
## Call:
## lm(formula = temperature_C ~ year4, data = Petersurface.August)
```

```
##
## 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
```

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

    From the graph, the overall highest surface water temperature is in July and August, and the lowest is in May in both Paul Lake and Peter Lake. For Peter Lake, the surface water temperature in July and August slightly increase over the period, and in the other months, the surfacewater temperature does not change obviously. In Paul Lake, the surface temperature slightly decreases over the period in June.