Assignment 2: Physical Properties of Lakes

Gaby Garcia

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 (should be project file location)

```
setwd("/Users/gabrielagarcia/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Data_Raw")
```

Load packages

```
library(tidyverse)
library(lubridate)
library(cowplot)
library(jcolors)
library(scico)
library(pals)
```

###Import Dataset and set date column to date format

```
setwd("/Users/gabrielagarcia/Desktop/Hydrologic Data Analysis/Hydrologic_Data_Analysis/Data_Raw")


NTLdata <-read.csv("NTL-LTER_Lake_ChemistryPhysics_Raw.csv")

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

###Check to see if sampledate is no longer a factor

```
class(NTLdata$sampledate)
```

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

###Set GGPlot Theme

```
gabytheme <- theme_bw(base_size = 14) +
  theme(plot.title=element_text(face="bold", size=16, color="hotpink4", hjust=0.5),
        axis.title=element_text(size=16, color="black"),
axis.text = element_text(size=10, color = "black"),
panel.background=element_rect(fill="gray88", color="darkblue"),
panel.border = element_rect(color = "black", size = 2),
legend.position = "top", legend.background = element_rect(fill="white", color="black"),
            legend.key = element_rect(fill="transparent", color="NA"))
theme_set(gabytheme)
```

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).

Tuesday Lake

```
TuesdayData<-filter(NTLdata, lakename=="Tuesday Lake", year4==c(1989, 2014))
```

1989 Tuesday Lake Data

```
TuesdayData1989<-filter(TuesdayData, year4==1989)
```

2014 Tuesday Lake Data

```
TuesdayData2014<-filter(TuesdayData, year4==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.

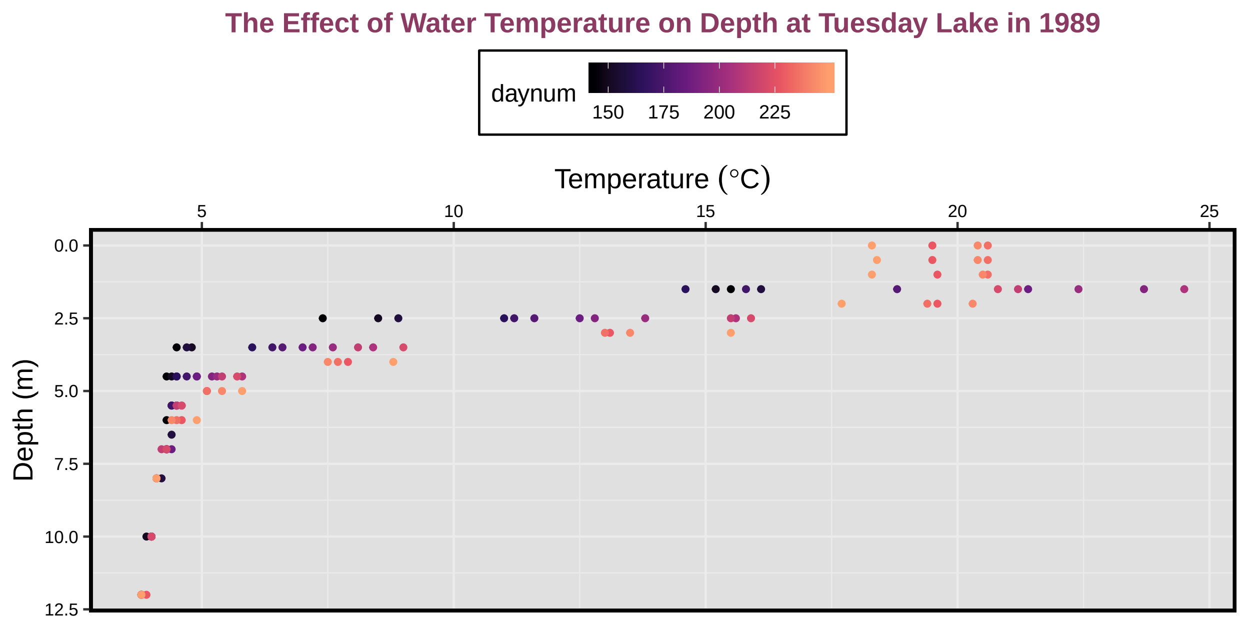
Change axis formatting

```
boldtext <- element_text(face = "bold", color = "black", size=11)
```

1989 Temperature Profile

```
library(viridis)
library(jcolors)
library(ggsci)
```

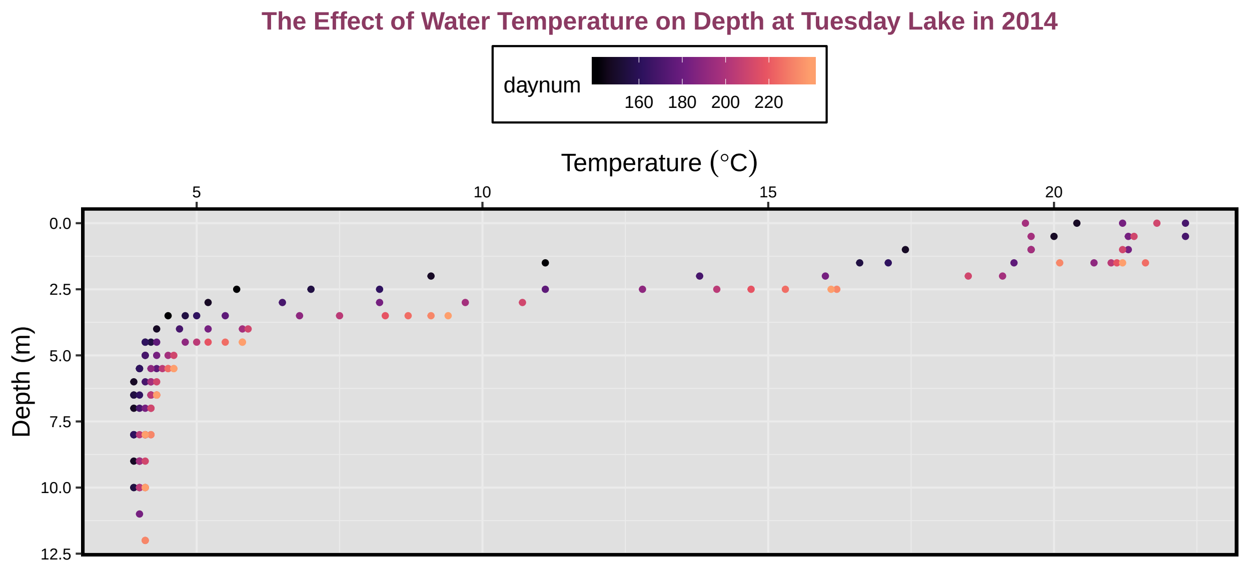
```r
TempvsDepth1989 <-ggplot(TuesdayData1989, aes(x =temperature_C, y =depth, color=daynum)) +
  geom_point() +
    scale_y_reverse() +
  scale_x_continuous(position="top")+
  scale_color_viridis_c(end = 0.8, option = "magma")+
  labs(title="The Effect of Water Temperature on Depth at Tuesday Lake in 1989",
       x = expression("Temperature "(degree*C)), y = "Depth (m)") +
  gabytheme+
    theme(legend.key.width=unit(1,"cm"))

print(TempvsDepth1989)
```



2014 Temperature Profile

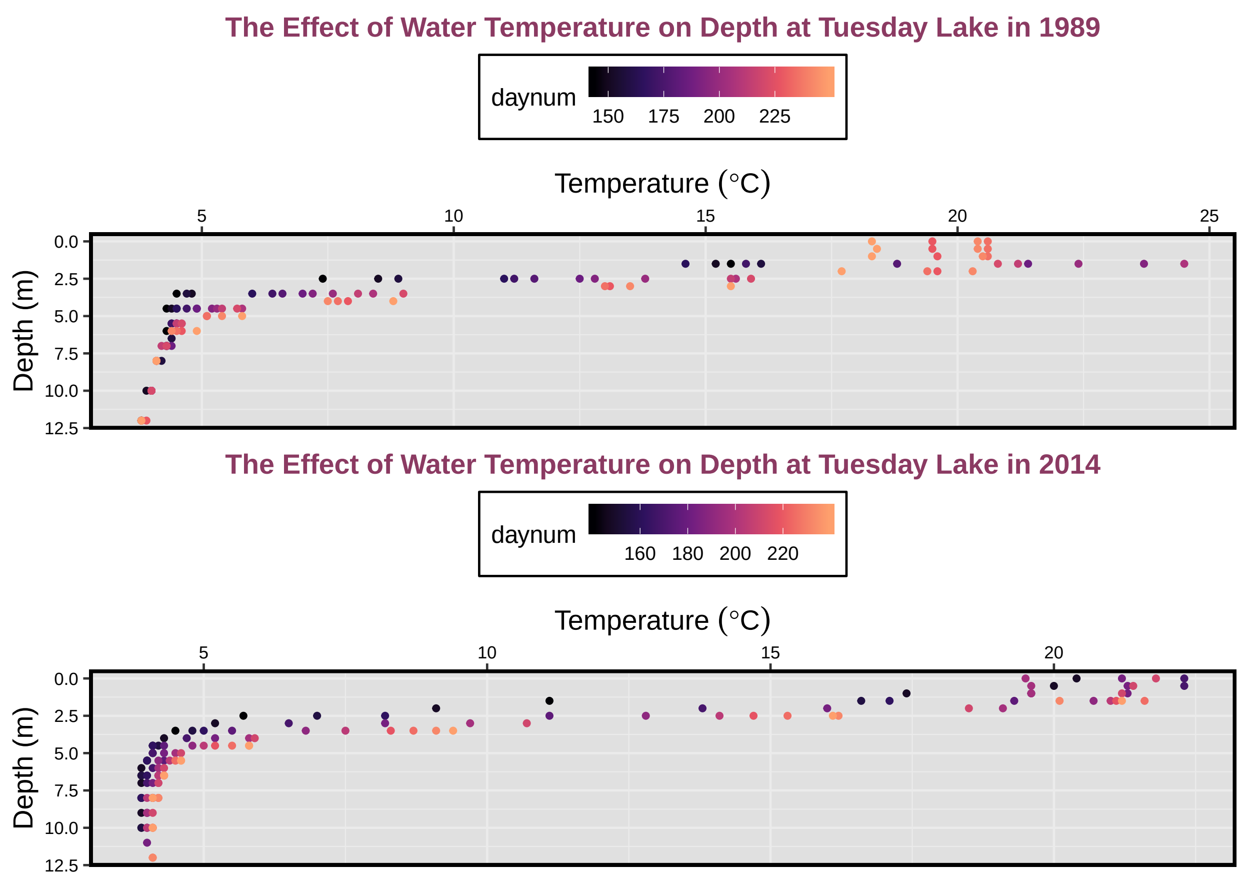
```r
library(viridis)
library(jcolors)
library(ggsci)
TempvsDepth2014 <-
  ggplot(TuesdayData2014, aes(x =temperature_C, y =depth, color=daynum)) +
  geom_point() +
   scale_y_reverse()+
  scale_x_continuous(position="top")+
  scale_color_viridis_c(end = 0.8, option = "magma") +
  labs(title="The Effect of Water Temperature on Depth at Tuesday Lake in 2014",
       x = expression("Temperature "(degree*C)), y = "Depth (m)") +
  gabytheme+
  theme(legend.key.width=unit(1,"cm"))
print(TempvsDepth2014)
```

```
## Warning: Removed 20 rows containing missing values (geom_point).
```



Tuesday Lake 1989 and 2014 Plot Grid Visualization

```
PhysicalProfiles19892014<-plot_grid(TempvsDepth1989, TempvsDepth2014,
            nrow = 2, rel_heights =c(1,1))
print(PhysicalProfiles19892014)
```



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 of my graphs, I see the expected stratification pattern typical of lakes, where the epilimnion or surface layer has the highest temperatures, due to more solar radiation or sunlight being able to penetrate the surface layers. The range of temperatures decreases as depth increases, so less light is able to penetrate the lake with depth because the bottom of the lake isn't interacting with the atmosphere. The data set for both 1989 and 2014 begins in late May, and I noticed that data points with the darkest color symbology (in late May or early June) tend to correspond to the lower water temperatures (less than 5 degrees Celsius), even in water depths at the thermocline and hypolimnion. AKA, the darker data points are to the left of the lighter points. This makes sense because water temperatures will be higher later in the season. Likewise, data points with the lightest color symbology (August through early Semptember) usually correspond to warmer water temperatures, even in depths at the thermocline and the hypolimnion. I do not see much of a difference between the two years.

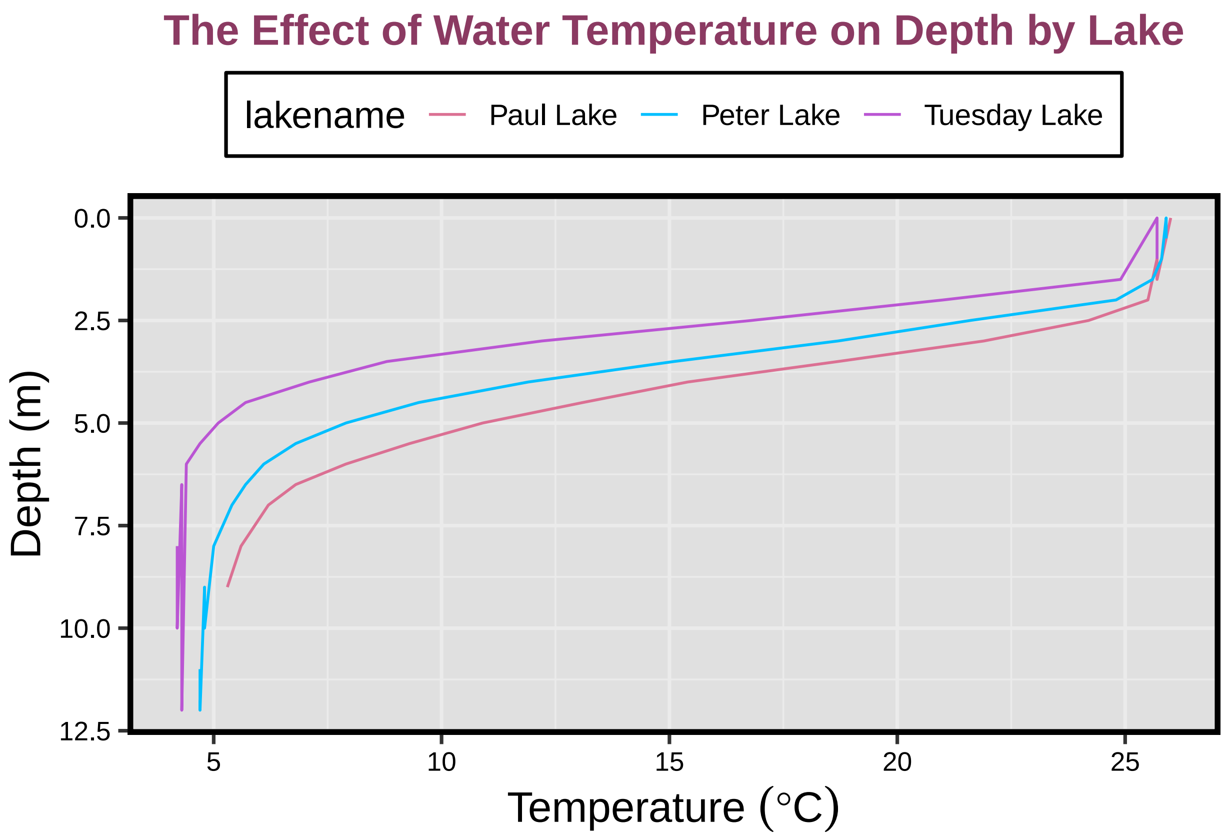
Multiple lakes, single date

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.

```
 PeterPaulTuesdayJuly<-
  NTLdata%>%
  dplyr::filter(lakename == "Paul Lake" | lakename == "Peter Lake" |lakename=="Tuesday Lake") %>%
  dplyr::filter(year4==2016)

PeterPaulTuesdayJulyFinal<-PeterPaulTuesdayJuly[c(615:677),]
```

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

```
PeterPaulTuesdayJulyPlot<-ggplot(PeterPaulTuesdayJulyFinal,
                                 aes(x = temperature_C, y = depth, color=lakename))+
geom_line(data=PeterPaulTuesdayJulyFinal, aes(x = temperature_C, y = depth))+
  gabytheme+
  scale_y_reverse()+
  scale_color_manual(values = c("#DB7093", "#00BFFF", "#BA55D3")) +
   labs(title="The Effect of Water Temperature on Depth by Lake",
        x = expression("Temperature "(degree*C)), y = "Depth (m)")


print(PeterPaulTuesdayJulyPlot)
```



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

The epilimnion is the surface layer in a stratified lake. The epilimnion depth ranges in Tuesday Lake is from 0 meters to about 1.35 meters below the surface. The epilimnion depth range in Peter Lake is from 0 meters to about 1.9 meters below the surface. The epilimnion depth range in Paul Lake is from 0 meters to about 2 meters below the surface.

The thermocline is a layer of the water column in a stratified lake. In that layer, the water temperature changes rapidly and dissolved oxygen levels can be finicky due to decomposition that occurs in every body of water. At the thermocline, water temperature decreases rapidly with increasing depth. At Tuesday Lake, The thermocline is from about 1.35 meters to about 5 meters below the surface. At Peter Lake, the thermocline is from about 1.9 meters to about 6.0 meters below the surface. At Paul Lake, the thermocline is from about 2 meters below the surface to about 6.875 m below surface.

The hypolimnion is the lower layer of water in a stratified lake, typically cooler than the water above and relatively stagnant. The hypolimion in Tuesday lake ranges from 5 meters to 12.0 meters below surface. The hypolimnion in Peter Lake ranges from about 6 meters to 12.0 meters below the surface. The hypolimnion in Pual lake ranges from about 6.875 meters to 12 meters below the surface.

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.

How much have surface temperatures increased over the period of study? Is this increase significant? Isolate surface depths and run the test for May, June, July, and August temperatures separately. Use a linear regression with year as the predictor variable and temperature as the response variable.

Filter Data by Tuesday Lake and add a Month Column

```
Tuesdaydata <- filter(NTLdata, lakename == "Tuesday Lake")
Tuesdaydata$Month<-as.Date(Tuesdaydata $sampledate,format="%m/%d/%y")
Tuesdaydata<-mutate(Tuesdaydata, Month=month(sampledate))
```

Filter Tuesday Lake data frame so it contains only surface depths

```
Tuesdaydatafiltered<-filter(Tuesdaydata, depth==0)
```

Filter Tuesday Lake data frame so it only contains Months 5-8

```
Tuesdaydatafilteredsummer<-filter(Tuesdaydatafiltered, Month==5| Month==6| Month==7| Month==8)
```

Tuesday Lake Data for May

```
TuesdaydatafilteredMay<-filter(Tuesdaydatafilteredsummer, Month=="5")
```

Tuesday Lake Data for June

```
TuesdaydatafilteredJune<-filter(Tuesdaydatafilteredsummer, Month=="6")
```

Tuesday Lake Data for July

```
TuesdaydatafilteredJuly<-filter(Tuesdaydatafilteredsummer, Month=="7")
```

Tuesday Lake Data for August

```
TuesdaydatafilteredAugust<-filter(Tuesdaydatafilteredsummer, Month=="8")
```

Linear Regression for May

```
TuesdayLinearRegressionMay<-lm(TuesdaydatafilteredMay$temperature_C~TuesdaydatafilteredMay$year4)
summary(TuesdayLinearRegressionMay)
```

```
##
## Call:
## lm(formula = TuesdaydatafilteredMay$temperature_C ~ TuesdaydatafilteredMay$year4)
##
## 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
## TuesdaydatafilteredMay$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
```

The Tuesday Lake surface data for the month of May does not display a significant trend because the p-value is 0.556, which is greater than 0.05.

Linear Regression for June

```
TuesdayLinearRegressionJune<-lm(TuesdaydatafilteredJune$temperature_C~TuesdaydatafilteredJune$year4)
summary(TuesdayLinearRegressionJune)
```

```
##
## Call:
## lm(formula = TuesdaydatafilteredJune$temperature_C ~ TuesdaydatafilteredJune$year4)
##
## 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
## TuesdaydatafilteredJune$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
```

The Tuesday Lake surface data for the month of June does not display a significant trend because the p-value is 0.992, which is greater than 0.05.

Linear Regression for July

```
TuesdayLinearRegressionJuly<-lm(TuesdaydatafilteredJuly$temperature_C~TuesdaydatafilteredJuly$year4)
summary(TuesdayLinearRegressionJuly)
```

```
##
## Call:
## lm(formula = TuesdaydatafilteredJuly$temperature_C ~ TuesdaydatafilteredJuly$year4)
##
## 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
## TuesdaydatafilteredJuly$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
```

The Temperature increases 0.04 degrees per year. 0.04*33 years of data=1.32 so there is a 1.32 degree increase over the period of study for July. This is significant because the p-value is 0.0569, which is close to 0.05.

Linear Regression for August

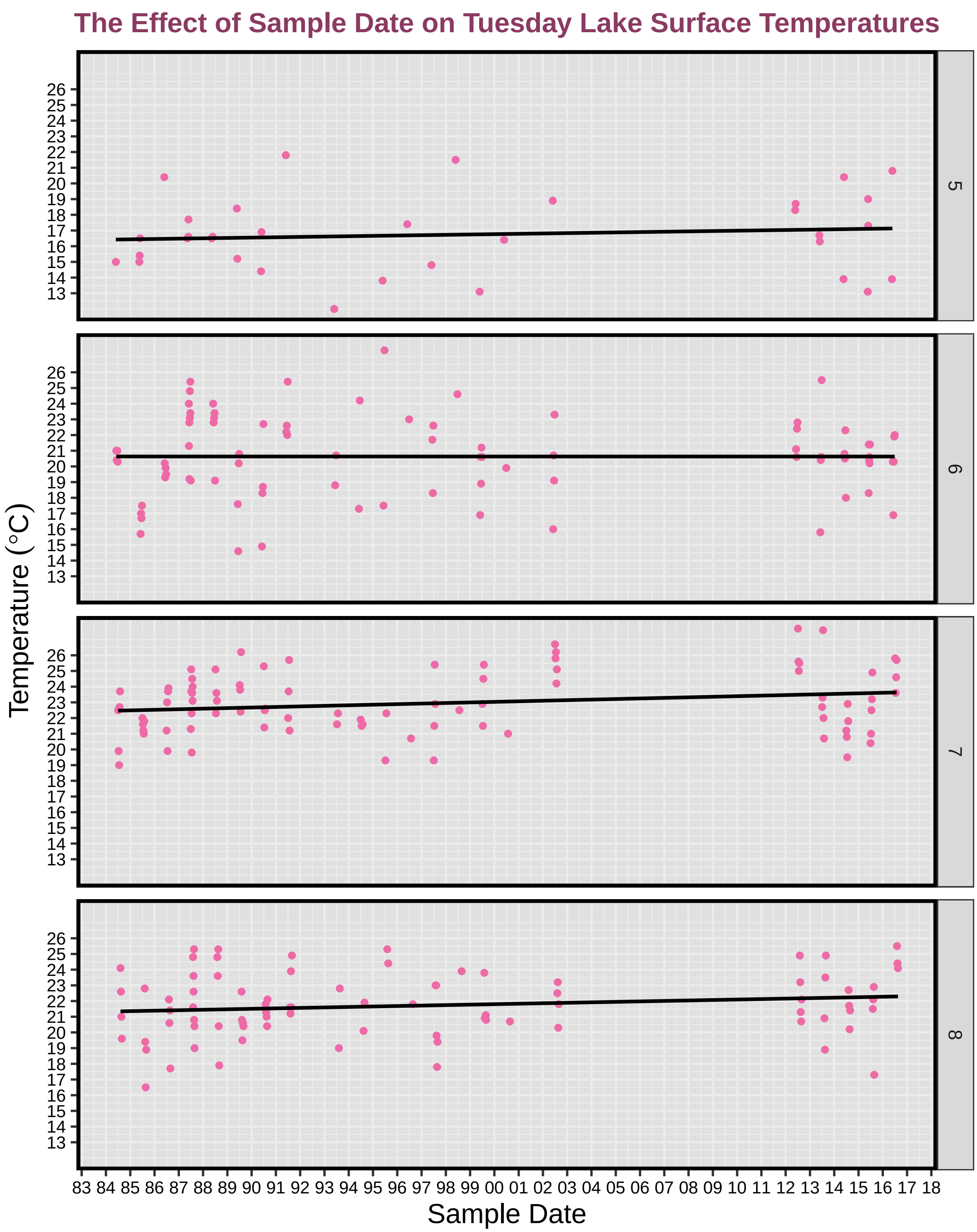
```
TuesdayLinearRegressionAugust<-lm(TuesdaydatafilteredAugust$temperature_C~TuesdaydatafilteredAugust$yea
summary(TuesdayLinearRegressionAugust)
```

```
##
## Call:
## lm(formula = TuesdaydatafilteredAugust$temperature_C ~ TuesdaydatafilteredAugust$year4)
##
## 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
## TuesdaydatafilteredAugust$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
```

The Tuesday Lake surface data for the month of August is not a significant trend because the p-value is 0.15, which is greater than 0.05.

Tuesday Surface Plot

```
library(scales)
TuesdaySurfacePlot<- ggplot(Tuesdaydatafilteredsummer, aes(x = sampledate, y = temperature_C)) +
  geom_point(color="hotpink2")+
  geom_smooth(se=FALSE, method=lm, color="black")+
  gabytheme+
  facet_grid(rows=vars(Month)) +
  labs(title="The Effect of Sample Date on Tuesday Lake Surface Temperatures",
      y= expression("Temperature "(degree*C)), x= "Sample Date")+
   scale_x_date(labels = date_format("%y"),
                breaks = date_breaks("1 year")) +
  scale_y_continuous(breaks=seq(13, 26, by = 1))

print(TuesdaySurfacePlot)
```



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

    I noticed that the geom_smooth layer, or the linear regression line, displayed similar trends to the Paul Lake graph's linear regression line in class. In my Tuesday Lake graph, for May, the linear regression line shows a slightly positive correlation, with it sitting between 16-17 degrees Celsius on the Y-axis. For June, the linear regression line does not have a correlation, with it sitting around 20.5 degrees Celsius on the Y axis. For July, the linear regression line has a slightly positive correlation, with it sitting between 22.5 degrres Celsius and 23.5 on the Y axis. For August, the linear regression line has a slight positive correlation with it sitting between 21.5 and 22.5 on the y-axis. A similar trend was present in the Tuesday Lake graph, where the average temperatures increase from May to July, and then decrease slightly in August. This would follow seasonal trends for solar radiation.