

Assignment 6: GLMs (Linear Regressios, ANOVA, & t-tests)

Nagarajan Vaidya Subramanian

Spring 2023

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on generalized linear models.

Directions

1. Rename this file <FirstLast>_A06_GLMs.Rmd (replacing <FirstLast> with your first and last name).
2. Change "Student Name" on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file.

Set up your session

1. Set up your session. Check your working directory. Load the tidyverse, agricolae and other needed packages. Import the *raw* NTL-LTER raw data file for chemistry/physics (NTL-LTER_Lake_ChemistryPhysics_Raw.csv). Set date columns to date objects.
2. Build a ggplot theme and set it as your default theme.

```
#1
library(tidyverse)

## — Attaching packages ————— tidyverse
1.3.2 —
## ✓ ggplot2 3.4.0      ✓ purrr  1.0.1
## ✓ tibble  3.1.8      ✓ dplyr  1.1.0
## ✓ tidyr   1.3.0      ✓ stringr 1.5.0
## ✓ readr   2.1.3      ✓ forcats 1.0.0
## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()

#install.packages("agricolae")
library(agricolae)
library(lubridate)
```

```

##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

library(here)

## here() starts at X:/ENV 872 Environmental Data Analytics/Git_codes/EDA-
Spring2023

getwd()

## [1] "X:/ENV 872 Environmental Data Analytics/Git_codes/EDA-Spring2023"

ntl_raw <- read.csv(file = here("Data/Raw/NTL-
LTER_Lake_ChemistryPhysics_Raw.csv"), stringsAsFactors = TRUE)
glimpse(ntl_raw) #checking the format of the columns in our data frame

## Rows: 38,614
## Columns: 11
## $ lakeid      <fct> L, L, L, L, L, L, L, L, L, L, L, L, L, L, L, L, L,
R, ...
## $ lakename    <fct> Paul Lake, Paul Lake, Paul Lake, Paul Lake, Paul
Lake,...
## $ year4       <int> 1984, 1984, 1984, 1984, 1984, 1984, 1984, 1984,
1984, ...
## $ daynum      <int> 148, 148, 148, 148, 148, 148, 148, 148, 148, 148,
148,...
## $ sampledate  <fct> 5/27/84, 5/27/84, 5/27/84, 5/27/84, 5/27/84,
5/27/84, ...
## $ depth       <dbl> 0.00, 0.25, 0.50, 0.75, 1.00, 1.50, 2.00, 3.00,
4.00, ...
## $ temperature_C <dbl> 14.5, NA, NA, NA, 14.5, NA, 14.2, 11.0, 7.0, 6.1,
5.5,...
## $ dissolvedOxygen <dbl> 9.5, NA, NA, NA, 8.8, NA, 8.6, 11.5, 11.9, 2.5,
1.6, 0...
## $ irradianceWater <dbl> 1750.0, 1550.0, 1150.0, 975.0, 870.0, 610.0,
420.0, 22...
## $ irradianceDeck <dbl> 1620, 1620, 1620, 1620, 1620, 1620, 1620, 1620,
1620, ...
## $ comments    <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, NA...

ntl_raw$sampledate <- mdy(ntl_raw$sampledate)

#2
#reusing the theme I created in Assignment 5
my_theme <- theme_minimal() +
  theme(

```

```

legend.position = "bottom",
legend.title = element_text(colour = "#440000"),
plot.title = element_text(colour = "blue",
                           family = "Serif",
                           hjust = 0.5),
plot.subtitle = element_text(family = "Serif",
                              hjust = 0.5),
axis.title = element_text(colour = "#4169e1",
                           family = "Serif")
)
theme_set(my_theme)

```

Simple regression

Our first research question is: Does mean lake temperature recorded during July change with depth across all lakes?

3. State the null and alternative hypotheses for this question: > Answer: H0: there is no significant change in the mean lake temperature with depth recorded during July across all lakes Ha: There is a significant change in the mean lake temperature with depth across all lakes
4. Wrangle your NTL-LTER dataset with a pipe function so that the records meet the following criteria:
 - Only dates in July.
 - Only the columns: lakename, year4, daynum, depth, temperature_C
 - Only complete cases (i.e., remove NAs)
5. Visualize the relationship among the two continuous variables with a scatter plot of temperature by depth. Add a smoothed line showing the linear model, and limit temperature values from 0 to 35 °C. Make this plot look pretty and easy to read.

```

#4
ntl_subset <- ntl_raw %>%
  filter(month(sampledate) == 7) %>%
  select('lakename', 'year4', 'daynum', 'depth', 'temperature_C') %>%
  drop_na()

```

```

#5
ggplot(data = ntl_subset, mapping = aes(x = depth,
                                         y = temperature_C,
                                         colour = lakename)) +

  geom_point() +
  geom_smooth(method = 'lm') +
  ylim(0, 35) +
  labs(title = "Plot of temperature against depth across all lakes",
       subtitle = "Data recorded in July of 1984-2016",
       x = "Depth (m)",
       y = "Temperature (deg.C)")

```

```
## `geom_smooth()`` using formula = 'y ~ x'

## Warning: Removed 73 rows containing missing values (`geom_smooth()`).

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family
## not found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family
## not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

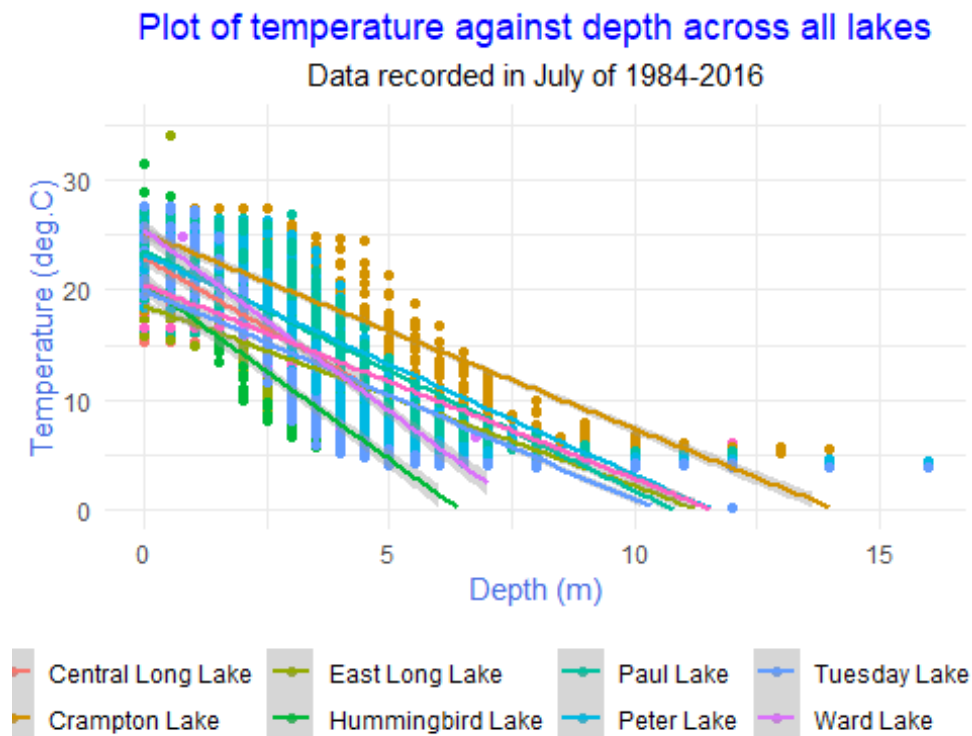
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

## Warning in grid.Call(graphics::C_text, as.graphicsAnnot(x$label), x$x, x$y, : 
:
## font family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : 
font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```

```
font
## family not found in Windows font database
```



6. Interpret the figure. What does it suggest with regards to the response of temperature to depth? Do the distribution of points suggest about anything about the linearity of this trend?

Answer: Broadly speaking, the temperature in the lake decreases with depth, as seen by our trendlines. However, this decrease is not linear; rather is S-shaped. Close to the surface there is very little change in temperature, then as depth increases the temperature drops quickly, then stabilising beyond a certain depth. This seems to indicate that the water in the lake is separated in two layers of different temperatures with a transition zone in between.

7. Perform a linear regression to test the relationship and display the results

```
#7
lm_temp_depth <- lm(formula = temperature_C ~ depth,
                     data = ntl_subset,
                     na.action = NULL)
summary(lm_temp_depth)

##
## Call:
## lm(formula = temperature_C ~ depth, data = ntl_subset, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -9.5173 -3.0192 0.0633 2.9365 13.5834
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.95597    0.06792   323.3  <2e-16 ***
## depth       -1.94621    0.01174  -165.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.835 on 9726 degrees of freedom
## Multiple R-squared:  0.7387, Adjusted R-squared:  0.7387
## F-statistic: 2.75e+04 on 1 and 9726 DF,  p-value: < 2.2e-16
```

8. Interpret your model results in words. Include how much of the variability in temperature is explained by changes in depth, the degrees of freedom on which this finding is based, and the statistical significance of the result. Also mention how much temperature is predicted to change for every 1m change in depth.

Answer: The model of Temperature against Depth has a β of -1.946, with a p-value less than 0.001. This means we have to reject the null hypothesis and accept the alternative hypothesis. The model indicates a statistically significant correlation between temperature and depth, with an increase in depth of 1 m being associated with a β change in temperature. The model has an F-statistic of $2.75e+04$ on 1 explanatory variable and 9726 degrees of freedom, which corresponds to a p-value of less than 0.001. The R^2 value reported is 0.7387. This indicates that the model has some explanatory power, with 73.87% of the variation in Temperature being explained by Depth.

Multiple regression

Let's tackle a similar question from a different approach. Here, we want to explore what might the best set of predictors for lake temperature in July across the monitoring period at the North Temperate Lakes LTER.

9. Run an AIC to determine what set of explanatory variables (year4, daynum, depth) is best suited to predict temperature.
10. Run a multiple regression on the recommended set of variables.

```
#9
lm_temp_year4 <- lm(formula = temperature_C ~ year4,
                    data = ntl_subset,
                    na.action = NULL)
lm_temp_daynum <- lm(formula = temperature_C ~ daynum,
                    data = ntl_subset,
                    na.action = NULL)

AIC(lm_temp_year4, lm_temp_daynum, lm_temp_depth)
```

```
##           df      AIC
## lm_temp_year4    3 66819.14
## lm_temp_daynum   3 66796.54
## lm_temp_depth    3 53762.12

lm_temp_multivar <- lm(formula = temperature_C ~ year4 + daynum + depth,
                        data = ntl_subset,
                        na.action = NULL)
step(lm_temp_multivar)

## Start:  AIC=26065.53
## temperature_C ~ year4 + daynum + depth
##
##           Df Sum of Sq    RSS   AIC
## <none>             141687 26066
## - year4    1         101 141788 26070
## - daynum   1         1237 142924 26148
## - depth    1        404475 546161 39189

##
## Call:
## lm(formula = temperature_C ~ year4 + daynum + depth, data = ntl_subset,
##     na.action = NULL)
##
## Coefficients:
## (Intercept)      year4      daynum      depth
##   -8.57556      0.01134      0.03978     -1.94644

#10
#recommended set of variables is all three of them, i.e. depth, daynum, and
year4. This is because removing any one of them causes the AIC score to
increase, implying a worse model. This means that we can use the model as-is.

summary(lm_temp_multivar)

##
## Call:
## lm(formula = temperature_C ~ year4 + daynum + depth, data = ntl_subset,
##     na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.6536 -3.0000  0.0902  2.9658 13.6123
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.575564   8.630715  -0.994  0.32044
## year4        0.011345   0.004299   2.639  0.00833 **
## daynum       0.039780   0.004317   9.215 < 2e-16 ***
## depth       -1.946437   0.011683 -166.611 < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.817 on 9724 degrees of freedom
## Multiple R-squared:  0.7412, Adjusted R-squared:  0.7411
## F-statistic: 9283 on 3 and 9724 DF,  p-value: < 2.2e-16
```

11. What is the final set of explanatory variables that the AIC method suggests we use to predict temperature in our multiple regression? How much of the observed variance does this model explain? Is this an improvement over the model using only depth as the explanatory variable?

Answer: The AIC score only worsens when we remove any of depth, year, and day number from our model. Therefore, the test implies that we can use our model as-is without removing any variable from it. This model explains 74.12% of the variability in the response variable using the three explanatory variables we have. It is a marginal improvement over the model using only depth as an explanatory variable as that model could explain 73.87% of the variability in temperature. Our multiple regression model has an F-statistic of 9283 on three variables and 9724 degrees of freedom, which corresponds to a p-value much lesser than 0.001. This indicates that our model has some explanatory power. Among the explanatory variables, daynum and depth are the most significant with p-values much lesser than 0.001. Temperature has a negative correlation with depth and a small positive correlation with daynum. 'year4' has a p-value of less than 0.01, and has a small positive correlation with temperature.

Analysis of Variance

12. Now we want to see whether the different lakes have, on average, different temperatures in the month of July. Run an ANOVA test to complete this analysis. (No need to test assumptions of normality or similar variances.) Create two sets of models: one expressed as an ANOVA models and another expressed as a linear model (as done in our lessons).

```
#12
lake_temp_anova <- aov(data = ntl_subset,
                      formula = temperature_C ~ lakename)
lake_temp_lm <- lm(data = ntl_subset,
                  formula = temperature_C ~ lakename)
summary(lake_temp_anova)

##              Df Sum Sq Mean Sq F value Pr(>F)
## lakename      8  21642   2705.2     50 <2e-16 ***
## Residuals  9719 525813     54.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(lake_temp_lm)
```



```
##
## Call:
## lm(formula = temperature_C ~ lakename, data = ntl_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.769  -6.614  -2.679   7.684  23.832
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.6664     0.6501  27.174 < 2e-16 ***
## lakenameCrampton Lake    -2.3145     0.7699   -3.006 0.002653 **
## lakenameEast Long Lake   -7.3987     0.6918 -10.695 < 2e-16 ***
## lakenameHummingbird Lake  -6.8931     0.9429   -7.311 2.87e-13 ***
## lakenamePaul Lake        -3.8522     0.6656   -5.788 7.36e-09 ***
## lakenamePeter Lake       -4.3501     0.6645   -6.547 6.17e-11 ***
## lakenameTuesday Lake     -6.5972     0.6769   -9.746 < 2e-16 ***
## lakenameWard Lake        -3.2078     0.9429   -3.402 0.000672 ***
## lakenameWest Long Lake   -6.0878     0.6895   -8.829 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.355 on 9719 degrees of freedom
## Multiple R-squared:  0.03953,    Adjusted R-squared:  0.03874
## F-statistic:    50 on 8 and 9719 DF,  p-value: < 2.2e-16
```

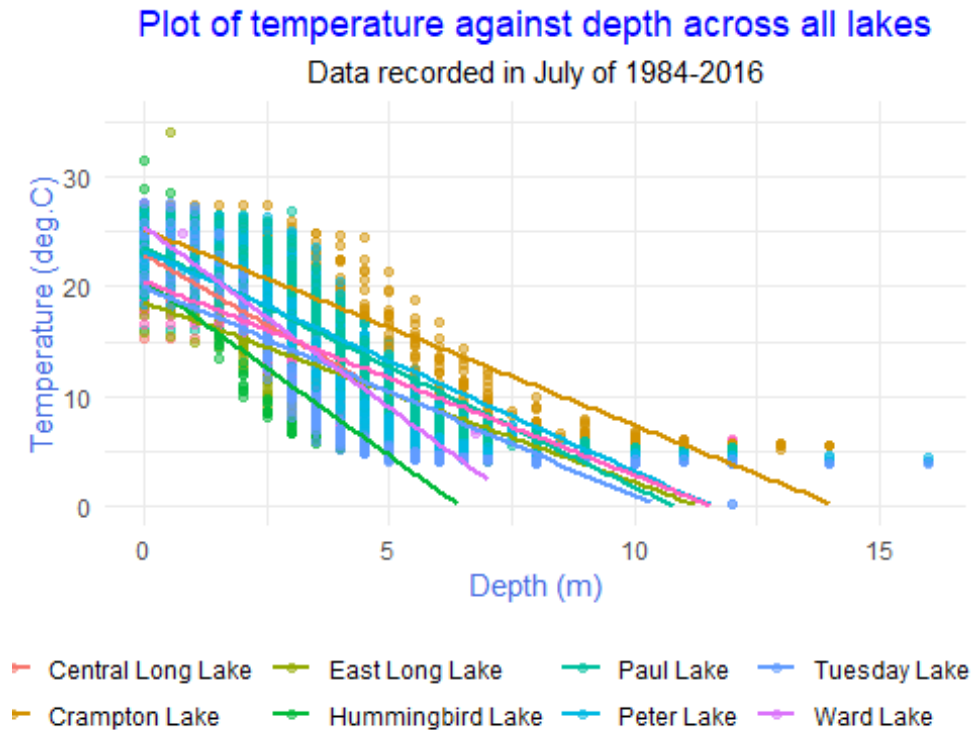
13. Is there a significant difference in mean temperature among the lakes? Report your findings.

Answer: The ANOVA test gives a p-value of less than 0.001 which means we can reject the null hypothesis and infer that the temperatures have significant variability across lakes. However it does not tell us which lakes specifically are significant causes of this variability. This information can be obtained from the linear model. We see that Crampton Lake has a p-value of less than 0.01 and all others have a p-value of less than 0.001. We can infer that all the lakes are statistically significant contributors to the variability of temperature across lakes, with their respective magnitudes being indicated by the β values.

14. Create a graph that depicts temperature by depth, with a separate color for each lake. Add a `geom_smooth` (method = "lm", se = FALSE) for each lake. Make your points 50 % transparent. Adjust your y axis limits to go from 0 to 35 degrees. Clean up your graph to make it pretty.

```
#14.
ggplot(data = ntl_subset, mapping = aes(x = depth,
                                         y = temperature_C,
                                         colour = lakename)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = 'lm', se = FALSE) +
  ylim(0, 35) +
  labs(title = "Plot of temperature against depth across all lakes",
```

```
    subtitle = "Data recorded in July of 1984-2016",  
    x = "Depth (m)",  
    y = "Temperature (deg.C)")  
  
## `geom_smooth()` using formula = 'y ~ x'  
  
## Warning: Removed 73 rows containing missing values (`geom_smooth()`).  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(graphics::C_text, as.graphicsAnnot(x$label), x$x, x$y,  
:  
## font family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
font  
## family not found in Windows font database
```



15. Use the Tukey's HSD test to determine which lakes have different means.

#15

```
aov_temp_depth <- aov(temperature_C ~ lakename, data = ntl_subset)
summary(aov_temp_depth)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## lakename      8  21642   2705.2     50 <2e-16 ***
## Residuals    9719 525813     54.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(aov_temp_depth)
```

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = temperature_C ~ lakename, data = ntl_subset)
##
## $lakename
##              diff              lwr              upr              p
adj
## Crampton Lake-Central Long Lake   -2.3145195 -4.7031913  0.0741524
0.0661566
## East Long Lake-Central Long Lake   -7.3987410 -9.5449411 -5.2525408
0.0000000
## Hummingbird Lake-Central Long Lake -6.8931304 -9.8184178 -3.9678430
0.0000000
```

## Paul Lake-Central Long Lake 0.0000003	-3.8521506 -5.9170942 -1.7872070
## Peter Lake-Central Long Lake 0.0000000	-4.3501458 -6.4115874 -2.2887042
## Tuesday Lake-Central Long Lake 0.0000000	-6.5971805 -8.6971605 -4.4972005
## Ward Lake-Central Long Lake 0.0193405	-3.2077856 -6.1330730 -0.2824982
## West Long Lake-Central Long Lake 0.0000000	-6.0877513 -8.2268550 -3.9486475
## East Long Lake-Crampton Lake 0.0000000	-5.0842215 -6.5591700 -3.6092730
## Hummingbird Lake-Crampton Lake 0.0000004	-4.5786109 -7.0538088 -2.1034131
## Paul Lake-Crampton Lake 0.0127491	-1.5376312 -2.8916215 -0.1836408
## Peter Lake-Crampton Lake 0.0000999	-2.0356263 -3.3842699 -0.6869828
## Tuesday Lake-Crampton Lake 0.0000000	-4.2826611 -5.6895065 -2.8758157
## Ward Lake-Crampton Lake 0.9714459	-0.8932661 -3.3684639 1.5819317
## West Long Lake-Crampton Lake 0.0000000	-3.7732318 -5.2378351 -2.3086285
## Hummingbird Lake-East Long Lake 0.9988050	0.5056106 -1.7364925 2.7477137
## Paul Lake-East Long Lake 0.0000000	3.5465903 2.6900206 4.4031601
## Peter Lake-East Long Lake 0.0000000	3.0485952 2.2005025 3.8966879
## Tuesday Lake-East Long Lake 0.1657485	0.8015604 -0.1363286 1.7394495
## Ward Lake-East Long Lake 0.0000002	4.1909554 1.9488523 6.4330585
## West Long Lake-East Long Lake 0.0022805	1.3109897 0.2885003 2.3334791
## Paul Lake-Hummingbird Lake 0.0004495	3.0409798 0.8765299 5.2054296
## Peter Lake-Hummingbird Lake 0.0080666	2.5429846 0.3818755 4.7040937
## Tuesday Lake-Hummingbird Lake 0.9999752	0.2959499 -1.9019508 2.4938505
## Ward Lake-Hummingbird Lake 0.0043297	3.6853448 0.6889874 6.6817022
## West Long Lake-Hummingbird Lake 0.9717297	0.8053791 -1.4299320 3.0406903
## Peter Lake-Paul Lake 0.2241586	-0.4979952 -1.1120620 0.1160717
## Tuesday Lake-Paul Lake 0.0000000	-2.7450299 -3.4781416 -2.0119182

## Ward Lake-Paul Lake 0.9916978	0.6443651 -1.5200848 2.8088149
## West Long Lake-Paul Lake 0.0000000	-2.2356007 -3.0742314 -1.3969699
## Tuesday Lake-Peter Lake 0.0000000	-2.2470347 -2.9702236 -1.5238458
## Ward Lake-Peter Lake 0.7827037	1.1423602 -1.0187489 3.3034693
## West Long Lake-Peter Lake 0.0000000	-1.7376055 -2.5675759 -0.9076350
## Ward Lake-Tuesday Lake 0.0000609	3.3893950 1.1914943 5.5872956
## West Long Lake-Tuesday Lake 0.7374387	0.5094292 -0.4121051 1.4309636
## West Long Lake-Ward Lake 0.0021080	-2.8799657 -5.1152769 -0.6446546

16. From the findings above, which lakes have the same mean temperature, statistically speaking, as Peter Lake? Does any lake have a mean temperature that is statistically distinct from all the other lakes?

Answer: We see that the Peter Lake - Paul Lake and Ward Lake - Peter Lake pairs have a p value of greater than 0.1, which implies that these three lakes have the same mean temperature. There do not seem to be any lakes that have a mean temperature that is statistically distinct from all other lakes, since each set of pair-wise comparisons has one pair that has a p-value greater than 0.05 implying no statistically significant difference in mean temperatures.

17. If we were just looking at Peter Lake and Paul Lake. What's another test we might explore to see whether they have distinct mean temperatures?

Answer: If we had mean temperatures of two samples, we can compare them using a Student T-Test to explore whether they have a statistically significant difference.

18. Wrangle the July data to include only records for Crampton Lake and Ward Lake. Run the two-sample T-test on these data to determine whether their July temperature are same or different. What does the test say? Are the mean temperatures for the lakes equal? Does that match you answer for part 16?

#the code that is commented out gave an NA result when calculating the mean. the warning code was "Warning: argument is not numeric or logical: returning NA". So i'm having to use a different way to calculate the mean of the temperatures.

```
#ntl_crampton_ward <- ntl_subset %>%
# filter(lakename == 'Crampton Lake' | lakename == 'Ward Lake')
#mean_temp_crampton <- ntl_crampton_ward %>%
# filter(lakename == 'Crampton Lake') %>%
# mean(temperature_C)
#mean_temp_ward <- ntl_crampton_ward %>%
# filter(lakename == 'Ward Lake') %>%
```

```

# mean(temperature_C)

ntl_crampton <- ntl_subset %>%
  filter(lakename == 'Crampton Lake')

ntl_ward <- ntl_subset %>%
  filter(lakename == 'Ward Lake')

ntl_crampton_ward <- rbind(ntl_crampton, ntl_ward)
#mean_temp_crampton <- mean(ntl_crampton$temperature_C)
#mean_temp_ward <- mean(ntl_ward$temperature_C)
t.test(x = ntl_crampton$temperature_C,
       y = ntl_ward$temperature_C,
       alternative = "two.sided")

##
## Welch Two Sample t-test
##
## data: ntl_crampton$temperature_C and ntl_ward$temperature_C
## t = 1.1181, df = 200.37, p-value = 0.2649
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.6821129 2.4686451
## sample estimates:
## mean of x mean of y
## 15.35189 14.45862

```

Answer: The two-sample t-test gave a T score of 1.1181 on 200.37 degrees of freedom, which corresponds to a p-value of 0.2649 which is greater than 0.025. Assuming a 95% confidence interval, we infer that the mean of the temperatures in the two lakes in the month of July is not significantly different.