Assignment 7: GLMs (Linear Regressions, ANOVA, & T-tests)

Damon L. Daniels

Fall 2024

OVERVIEW

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

Directions

1. Rename this file `<FirstLast>_A07_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
getwd()
```

```
## [1] "/home/guest/EDA_Spring2024"
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
library(agricolae)
library(ggthemes)
library(here)
```

```
## here() starts at /home/guest/EDA_Spring2024
```

```r
here()
```

```
## [1] "/home/guest/EDA_Spring2024"
```

```r
NTL_LTER <- read.csv(
  here("Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv"), stringsAsFactors = TRUE)

NTL_LTER$sampledate <- as.Date(NTL_LTER$sampledate, format = "%m/%d/%y")


#2

mytheme <- theme_economist_white(
  base_size = 11,
  base_family = "sans",
  gray_bg = TRUE,
  horizontal = TRUE)

theme_set(mytheme)
```

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: Mean lake temperature recorded during July *does not* change with depth across all lakes.
   >
   > Ha: Mean lake temperature recorded during July *does* change with depth across all lakes.

4. Wrangle your NTL-LTER dataset with a pipe function so that the records meet the following criteria:

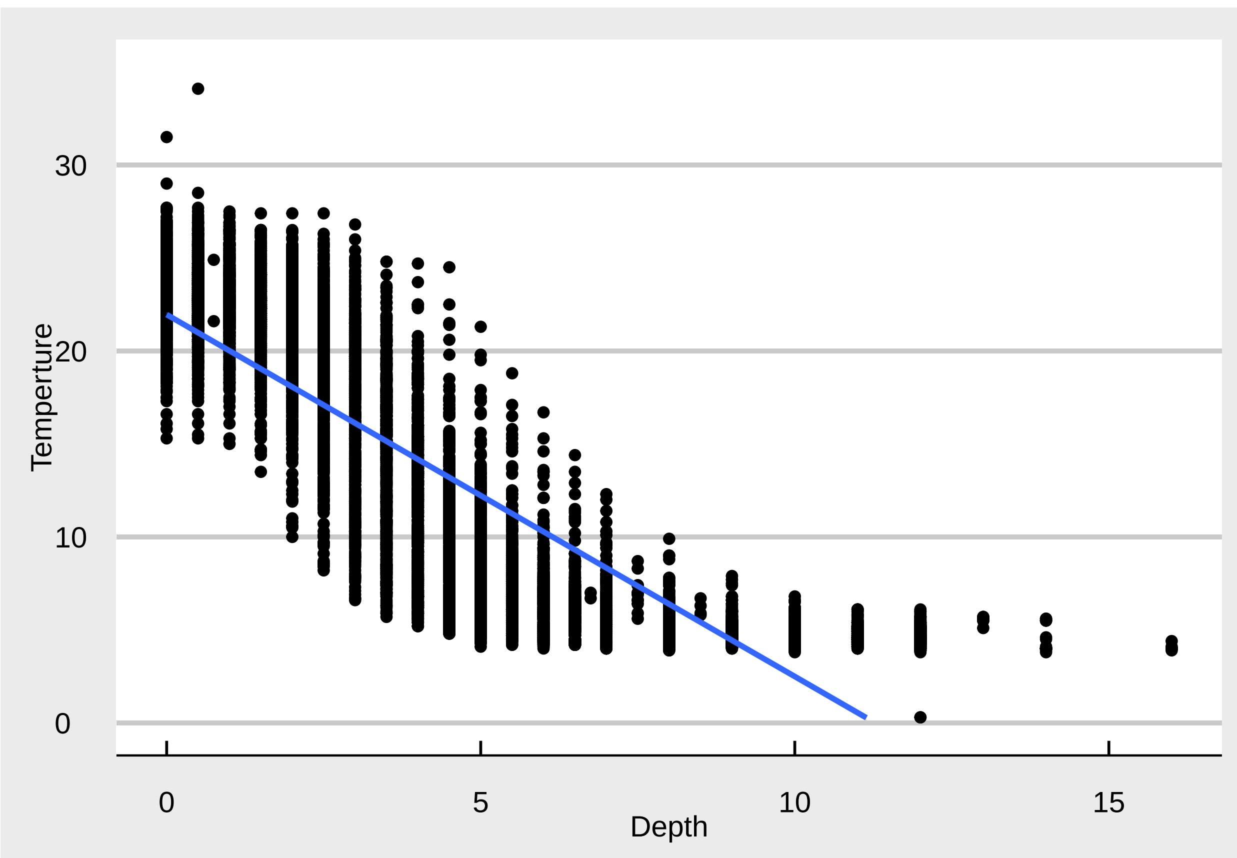
- Only dates in July.
- Only the columns: `lakename`, `year4`, `daynum`, `depth`, `temperature_C`
- Only complete cases (i.e., remove NAs)

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.

```r
#4

NTL_LTER_July <- NTL_LTER %>%
  filter(month(sampledate) == 7) %>%
  select(lakename:temperature_C) %>%
  drop_na(lakename:temperature_C)

#5

library(ggplot2)

NTL_LTER_July_scattplot <-
  ggplot(NTL_LTER_July, aes(x = depth, y = temperature_C)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(x = "Depth", y = "Temperture") +
  ylim(0, 35)
print(NTL_LTER_July_scattplot)
```

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

```
## Warning: Removed 24 rows containing missing values or values outside the scale range
## (`geom_smooth()`).
```



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:

   > As depth increases, temperature decreases, i.e., there's a negative correlation between the two. The relationship does not appear to be entirely linear, given the values that do not fit the line as depth continues to increase. These values appear to indicate that at eventually temperatures remain relative steady beyond a certain depth point.

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

```r
#7

NTL_LTER_July_regression <-
  lm(NTL_LTER_July$temperature_C ~ NTL_LTER_July$depth)
summary(NTL_LTER_July_regression)
```

```
##
## Call:
## lm(formula = NTL_LTER_July$temperature_C ~ NTL_LTER_July$depth)
##
## 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 ***
## NTL_LTER_July$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 slope estimate (NTL_LTER_July$depth) affirms the negative relationship between depth and temperature, as the temperature is expected to decrease by approximately 1.94621 degrees for every 1-meter increase in depth. The p-values (2.2e-16) are less than 0.05, therefore the results are statistically significant. The R-squared values indicate that approximately 73.87% (at a value of 0.7387) of the variability in temperature can be explained by changes in 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.

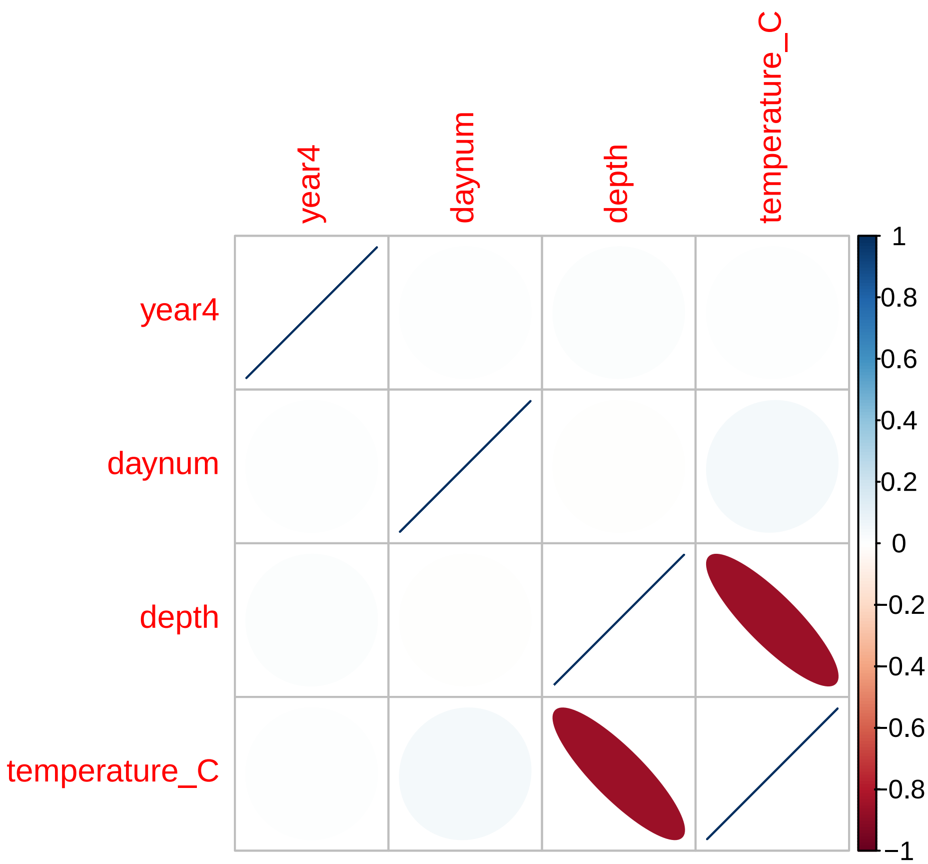
9. Run an AIC to determine what set of explanatory variables (year4, daynum, depth) is best suited to predict temperature.

10. Run a multiple regression on the recommended set of variables.

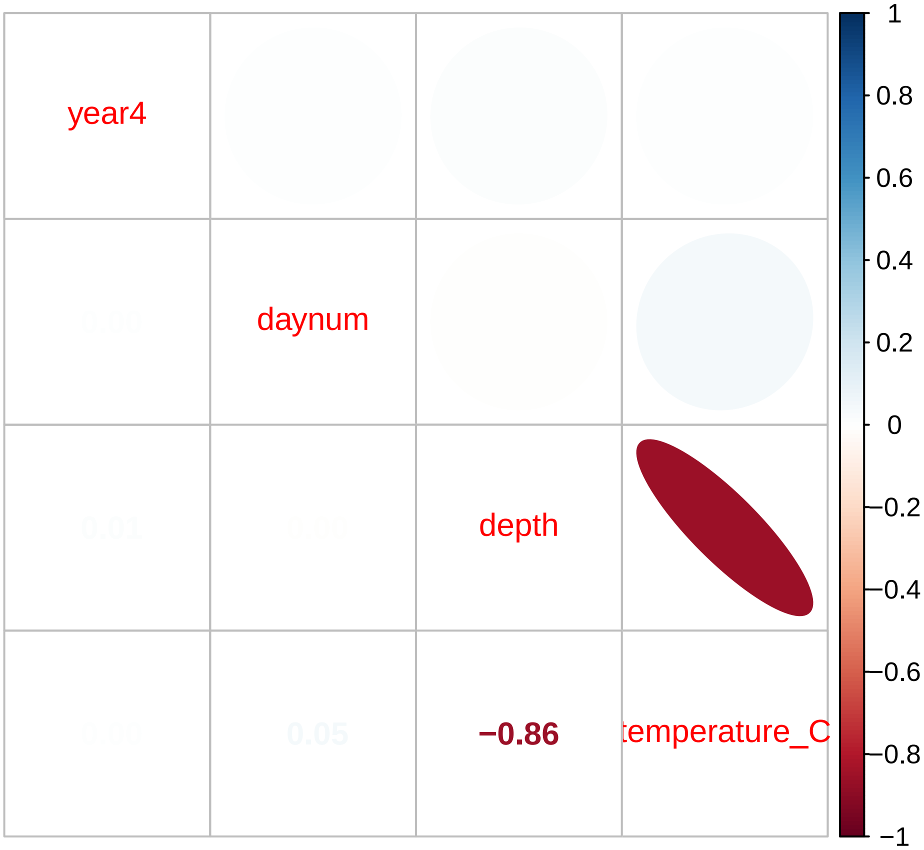
```r
#9

library(corrplot)
```

```
## corrplot 0.95 loaded
```

```r
NTL_LTER_July_subset <-
  NTL_LTER_July %>%
  select(year4, daynum, depth, temperature_C) %>%
  na.omit()

corr.NTL_LTER_July_subset <-
  cor(NTL_LTER_July_subset)

corrplot(corr.NTL_LTER_July_subset, method = "ellipse")
```



```r
corrplot.mixed(corr.NTL_LTER_July_subset, upper = "ellipse")
```




```r
TPAIC <- lm(data = NTL_LTER_July_subset, temperature_C ~ year4 + daynum + depth)
step(TPAIC)
```

```
## 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_LTER_July_subset)
##
## Coefficients:
## (Intercept)        year4       daynum        depth
##    -8.57556      0.01134      0.03978     -1.94644
```

```r
summary(TPAIC)
```

```
##
## Call:
## lm(formula = temperature_C ~ year4 + daynum + depth, data = NTL_LTER_July_subset)
##
## 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
```

```r
#10

NTL_LTER_July_multiple_regression <- lm(data = NTL_LTER_July_subset, temperature_C ~ daynum + depth)
summary(NTL_LTER_July_multiple_regression)
```

```
##
## Call:
## lm(formula = temperature_C ~ daynum + depth, data = NTL_LTER_July_subset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6174 -2.9809  0.0845  2.9681 13.4406
##
## Coefficients:
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 14.088588   0.855505   16.468   <2e-16 ***
## daynum       0.039836   0.004318    9.225   <2e-16 ***
## depth       -1.946111   0.011685 -166.541   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.818 on 9725 degrees of freedom
## Multiple R-squared:  0.741,  Adjusted R-squared:  0.741
## F-statistic: 1.391e+04 on 2 and 9725 DF,  p-value: < 2.2e-16
```

```r
NTL_LTER_July_multiple_regression2 <- lm(data = NTL_LTER_July_subset, temperature_C ~ depth)
summary(NTL_LTER_July_multiple_regression2)
```

```
##
```

```
## Call:
## lm(formula = temperature_C ~ depth, data = NTL_LTER_July_subset)
##
## 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
```

```
## I was also curious to use "year4" and "depth" as explanatory variables:
## NTL_LTER_July_multiple_regression3 <-
##     lm(data = NTL_LTER_July_subset, temperature_C ~ year4 + depth)
## summary(NTL_LTER_July_multiple_regression3)
```

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:

    > Because the AIC value is lowest when all three explanatory variables are used, none of them need to be dropped.
    >
    > I initially used only *daynum* and *depth* as explanatory variables to predict temperature in the multiple regression model. As such, the R-squared values increased from approximately 73.87% (at a value of 0.7387) to approximately 74.1% (at a value of 0.741). This is approximately 0.23 percentage points higher (+0.0023 R-squared value points higher), based on 9725 degrees of freedom. Meaning, *daynum* and *depth* can account for approximately 74% of the temperature variability. The p-values remain less than 0.05, meaning that the results are statistically sigificaht. These results are only slightly higher than under the condition where *depth* is the sole explanatory variable.
    >
    > I also tested the explantory variables of *year4* and *depth* to predict temperature in the multiple regression model. As such, the R-squared values increased from approximately 73.87% (at a value of 0.7387) to approximately 78.89% (at a value of 0.7389). This is approximately 0.02 percentage points higher (+0.0002 R-squared value points higher), based on 9725 degrees of freedom. Meaning, *year4* and *depth* can account for approximately 74% of the temperature variability. The p-value for the intercept is 0.8913, which is not less than 0.05, meaning that this result is not statistically sigificant. However, both the year4 value (0.00754) and the depth value (2e-16) are less than 0.05, meaning that these values are statistically significant. These results are slightly lower than under the condition where *depth* is the sole explanatory variable.

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

NTL_LTER_July_subset.lakes <- NTL_LTER_July %>%
  group_by(lakename, sampledate, temperature_C) %>%
  summarise(temperature_C = mean(temperature_C))
```

```
## `summarise()` has grouped output by 'lakename', 'sampledate'. You can override
## using the `.groups` argument.
```

```
anova.model_lakes_July <- aov(data = NTL_LTER_July_subset.lakes, temperature_C ~ lakename)
summary(anova.model_lakes_July)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## lakename        8  12361    1545   31.55 <2e-16 ***
## Residuals    8090 396256      49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova.model_lakes_July.lm <- lm(data = NTL_LTER_July_subset.lakes, temperature_C ~ lakename)
summary(anova.model_lakes_July.lm)
```

```
##
## Call:
## lm(formula = temperature_C ~ lakename, data = NTL_LTER_July_subset.lakes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.722  -6.219  -2.691   6.881  24.009
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 17.1388     0.6896  24.853  < 2e-16 ***
## lakenameCrampton Lake       -3.2021     0.8180  -3.914 9.13e-05 ***
## lakenameEast Long Lake      -7.0479     0.7317  -9.632  < 2e-16 ***
## lakenameHummingbird Lake    -6.3994     0.9617  -6.654 3.04e-11 ***
## lakenamePaul Lake           -3.9202     0.7049  -5.561 2.76e-08 ***
## lakenamePeter Lake          -4.6497     0.7044  -6.601 4.34e-11 ***
## lakenameTuesday Lake        -6.1172     0.7181  -8.518  < 2e-16 ***
## lakenameWard Lake           -2.8692     0.9554  -3.003  0.00268 **
## lakenameWest Long Lake      -5.6989     0.7319  -7.786 7.76e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.999 on 8090 degrees of freedom
## Multiple R-squared:  0.03025,    Adjusted R-squared:  0.02929
## F-statistic: 31.55 on 8 and 8090 DF,  p-value: < 2.2e-16
```

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

    Answer:

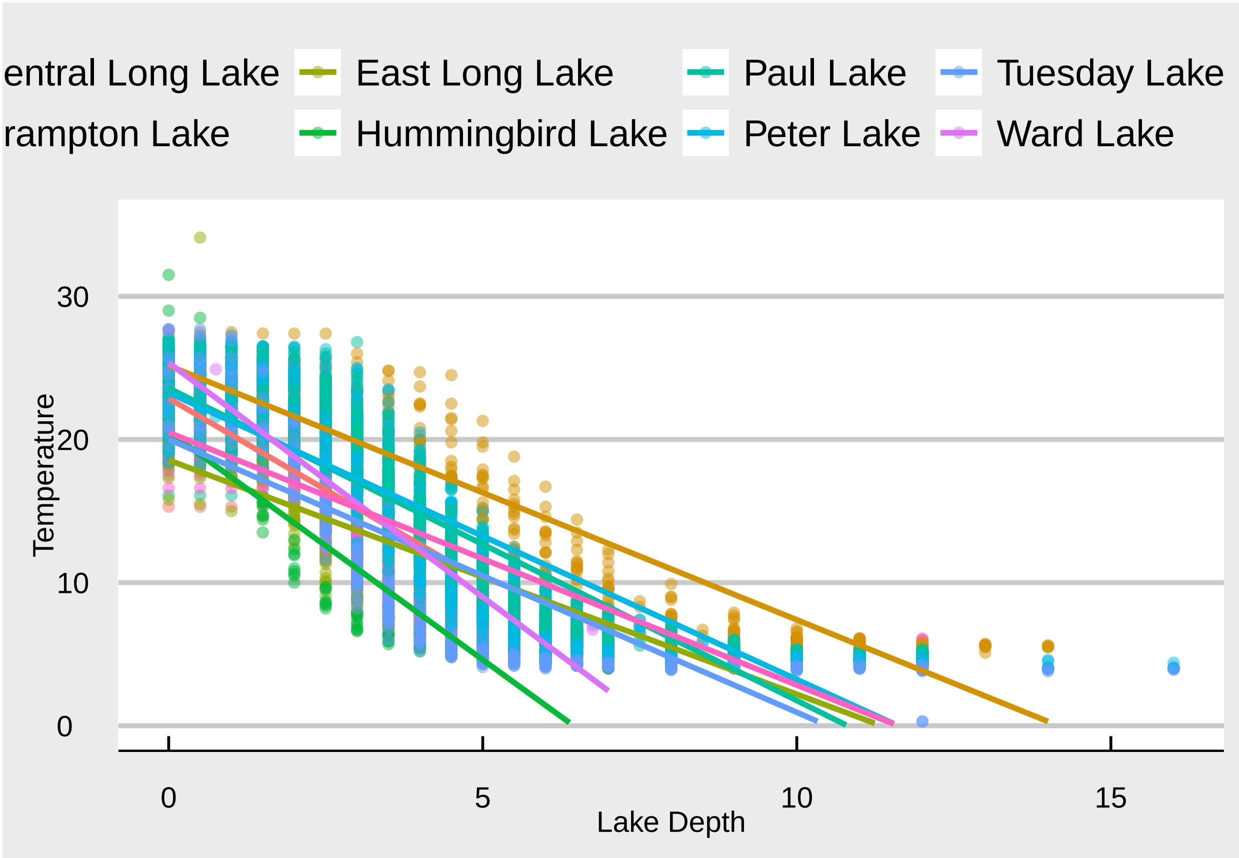
    > There are significant differences in mean temperature among the lakes, and these are evident in both the ANOVA and lm models. in the ANOVA model that has 8 degrees of freedom, the Sum of Squares is 12,361, the Mean-Squared value is 1545, the F-value is 31.55, and the p-value is less than 2e-16, which indicates a rejection of the Null Hypothesis and results that are statistically significant. However, this model does not indicate which lakes have differences in temperature, and by which amounts. In the lm model, which has 8090 degrees of freedom, each of the p-values for each of the different lakes are less than 0.05, which indicates as well that the Null Hypothesis is to be rejected, and the results of each are statistically signifcant. Among both models, the differences in mean temperature are significant.

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.

NTL_LTER.tempbydepth <- NTL_LTER_July %>%
  ggplot(aes(x = depth, y = temperature_C, color = lakename)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = lm, se = FALSE) +
  labs(x = "Lake Depth", y = "Temperature") +
  ylim(0, 35)
print(NTL_LTER.tempbydepth)
```

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

```
## Warning: Removed 73 rows containing missing values or values outside the scale range
## (`geom_smooth()`).
```



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

```r
#15

TukeyHSD(anova.model_lakes_July)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = temperature_C ~ lakename, data = NTL_LTER_July_subset.lakes)
##
## $lakename
##                                         diff         lwr          upr     p adj
## Crampton Lake-Central Long Lake    -3.2020761 -5.74001952 -0.66413259 0.0029444
## East Long Lake-Central Long Lake   -7.0478814 -9.31811407 -4.77764874 0.0000000
## Hummingbird Lake-Central Long Lake -6.3993854 -9.38320037 -3.41557045 0.0000000
## Paul Lake-Central Long Lake        -3.9201851 -6.10717306 -1.73319719 0.0000010
## Peter Lake-Central Long Lake       -4.6496560 -6.83508172 -2.46423029 0.0000000
## Tuesday Lake-Central Long Lake     -6.1171779 -8.34525939 -3.88909632 0.0000000
## Ward Lake-Central Long Lake        -2.8691921 -5.83352807  0.09514388 0.0667707
## West Long Lake-Central Long Lake   -5.6989087 -7.96976506 -3.42805227 0.0000000
## East Long Lake-Crampton Lake       -3.8458053 -5.40785071 -2.28375999 0.0000000
## Hummingbird Lake-Crampton Lake     -3.1973094 -5.68511583 -0.70950288 0.0021963
## Paul Lake-Crampton Lake            -0.7181091 -2.15649204  0.72027390 0.8321368
## Peter Lake-Crampton Lake           -1.4475799 -2.88358653 -0.01157336 0.0463992
```

```
## Tuesday Lake-Crampton Lake         -2.9151018 -4.41522735 -1.41497624 0.0000001
## Ward Lake-Crampton Lake             0.3328840 -2.13152613  2.79729405 0.9999745
## West Long Lake-Crampton Lake       -2.4968326 -4.05978433 -0.93388088 0.0000258
## Hummingbird Lake-East Long Lake     0.6484960 -1.56554564  2.86253763 0.9925521
## Paul Lake-East Long Lake            3.1276963  2.24353289  4.01185967 0.0000000
## Peter Lake-East Long Lake           2.3982254  1.51793327  3.27851753 0.0000000
## Tuesday Lake-East Long Lake         0.9307036 -0.05071176  1.91211886 0.0791210
## Ward Lake-East Long Lake            4.1786893  1.99096986  6.36640877 0.0000001
## West Long Lake-East Long Lake       1.3489727  0.27397627  2.42396922 0.0032020
## Paul Lake-Hummingbird Lake          2.4792003  0.35059975  4.60780082 0.0092415
## Peter Lake-Hummingbird Lake         1.7497294 -0.37726603  3.87672484 0.2073532
## Tuesday Lake-Hummingbird Lake       0.2822076 -1.88859215  2.45300726 0.9999811
## Ward Lake-Hummingbird Lake          3.5301933  0.60866773  6.45171890 0.0055972
## West Long Lake-Hummingbird Lake     0.7004767 -1.51420444  2.91515794 0.9876078
## Peter Lake-Paul Lake               -0.7294709 -1.36498178 -0.09395998 0.0111516
## Tuesday Lake-Paul Lake             -2.1969927 -2.96649639 -1.42748907 0.0000000
## Ward Lake-Paul Lake                 1.0509930 -1.05021526  3.15220132 0.8306572
## West Long Lake-Paul Lake           -1.7787235 -2.66448722 -0.89295986 0.0000000
## Tuesday Lake-Peter Lake            -1.4675219 -2.23257427 -0.70246943 0.0000001
## Ward Lake-Peter Lake                1.7804639 -0.31911834  3.88004616 0.1738206
## West Long Lake-Peter Lake          -1.0492527 -1.93115210 -0.16735322 0.0069477
## Ward Lake-Tuesday Lake              3.2479858  1.10403908  5.39193245 0.0000920
## West Long Lake-Tuesday Lake         0.4182692 -0.56458807  1.40112646 0.9255550
## West Long Lake-Ward Lake           -2.8297166 -5.01808327 -0.64134987 0.0019881
```

```
NTL_LTER_July.groups <- HSD.test(anova.model_lakes_July, "lakename", group = TRUE)
NTL_LTER_July.groups
```

```
## $statistics
##    MSerror   Df     Mean       CV
##   48.98091 8090 12.23285 57.21184
##
## $parameters
##    test   name.t ntr StudentizedRange alpha
##   Tukey lakename   9         4.387705  0.05
##
## $means
##                   temperature_C      std    r         se Min  Max  Q25   Q50
## Central Long Lake      17.13883 4.266473  103 0.6895961 8.9 26.8 13.6 17.30
## Crampton Lake          13.93676 6.820385  253 0.4400005 5.0 27.5  6.8 13.40
## East Long Lake         10.09095 6.406766  818 0.2447016 4.2 34.1  5.2  6.75
## Hummingbird Lake       10.73945 6.993119  109 0.6703478 4.0 31.5  5.2  7.10
## Paul Lake              13.21865 7.013441 2296 0.1460587 4.7 27.7  6.5 11.20
## Peter Lake             12.48918 7.240275 2375 0.1436090 4.0 27.0  5.6 10.10
## Tuesday Lake           11.02166 7.321204 1219 0.2004525 0.3 27.7  4.7  7.50
## Ward Lake              14.26964 7.341905  112 0.6613090 5.7 27.6  7.2 12.50
## West Long Lake         11.43993 6.585183  814 0.2453021 4.0 25.7  5.6  8.50
##                     Q75
## Central Long Lake 20.55
## Crampton Lake     20.50
## East Long Lake    14.50
## Hummingbird Lake  15.50
## Paul Lake         20.30
## Peter Lake        20.00
```

```
## Tuesday Lake      18.30
## Ward Lake         23.10
## West Long Lake    17.70
##
## $comparison
## NULL
##
## $groups
##                   temperature_C groups
## Central Long Lake      17.13883      a
## Ward Lake              14.26964     ab
## Crampton Lake          13.93676      b
## Paul Lake              13.21865      b
## Peter Lake             12.48918      b
## West Long Lake         11.43993      c
## Tuesday Lake           11.02166     cd
## Hummingbird Lake       10.73945     cd
## East Long Lake         10.09095      d
##
## attr(,"class")
## [1] "group"
```

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:
>
> > By their statistical groupings, Crampton Lake and Paul Lake can be considered to have the same mean temperature as Peter Lake. No lake has a mean temperature that is statistically distinct from all of the other lakes, given that no one lake (by lakename) is in a grouping by itself.

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:

    > A two-sample T-Test could is another test that could be conducted, given that we have two different samples. In addition to the hypothesis test to assess whether the means of the two samples are equivalent or different, two-sample T-tests carry the assumption that the variance values of each sample are also equal.

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?

```
NTL_LTER_July.Crampton.and.Ward <- NTL_LTER_July %>%
  filter(lakename %in% c("Crampton Lake", "Ward Lake"))

Crampton.and.Ward_ttest <- t.test(NTL_LTER_July.Crampton.and.Ward$temperature_C ~ NTL_LTER_July.Crampton
print(Crampton.and.Ward_ttest)
```

```
##
##  Welch Two Sample t-test
##
## data:  NTL_LTER_July.Crampton.and.Ward$temperature_C by NTL_LTER_July.Crampton.and.Ward$lakename
## t = 1.1181, df = 200.37, p-value = 0.2649
## alternative hypothesis: true difference in means between group Crampton Lake and group Ward Lake is 
## 95 percent confidence interval:
##  -0.6821129  2.4686451
## sample estimates:
## mean in group Crampton Lake     mean in group Ward Lake
##                    15.35189                    14.45862
```

Answer:

> H0: mu (Crampton Lake) = mu (Ward Lake); or, H0: mean temperature difference b/w lakes = 0 Ha: mu (Crampton Lake) != mu (Ward Lake) or, H0: mean temperature difference b/w lakes != 0
>
> The two-sample T-test result shows that there is no statistical difference between the respective means of Crampton Lake and Ward Lakes. Given that the reported p-value is 0.2649, which is greater than 0.05, so we *fail to reject* the Null Hypothesis (H0). [Alternatively, we could also state that we *reject the Alternative Hypothesis (Ha)]. The respective mean temperatures from Crampton Lake and Ward Lake are similar and could (roughly) be considered equal, which matches our findings in Question 16.
>
> These results are expressed with 200.37 degrees of freedom.