Assignment 7: GLMs (Linear Regressios, ANOVA, & t-tests)

Lauren Ng

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

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 errors
```

```
library(agricolae)
library(here)
```

```
## here() starts at /home/guest/EDE_Fall2024
```

```
library (dplyr)

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

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


#2

mytheme <- theme_minimal(base_size = 14) +
  theme(axis.text = element_text(color = "black"))
```

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: The mean lake temperature in July does not change with depth across all lakes. Ha: The mean lake temperature in July changes with depth across all lakes.
4. Wrangle your NTL-LTER dataset with a pipe function so that the records meet the following criteria:

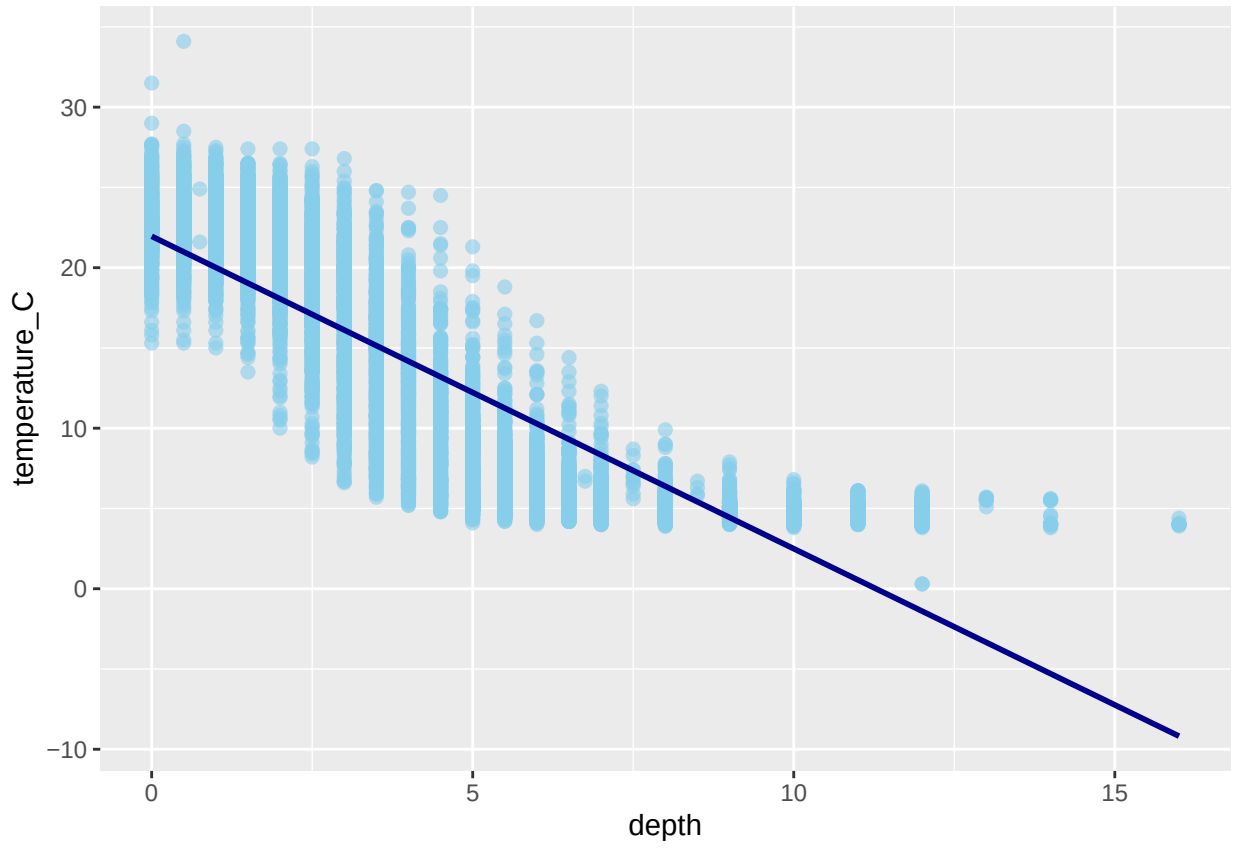
- Only dates in July.
- Only the columns: `lakename`, `year4`, `daynum`, `depth`, `temperature_C`
- Only complete cases (i.e., remove NAs)

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
lakedepth.temp <- NLTERdata %>%
  filter(month(sampledate) == 7) %>%
  select(lakename, year4, daynum, depth, temperature_C) %>%
  drop_na()


#5

ggplot(lakedepth.temp, aes(x = depth, y = temperature_C)) +
  geom_point(color = "skyblue", alpha = 0.6, size = 2) +
  geom_smooth(method = "lm", color = "darkblue",se = FALSE)
```




```
  theme_minimal(base_size = 14)
```

```
## List of 136
##  $ line                            :List of 6
##   ..$ colour       : chr "black"
##   ..$ linewidth    : num 0.636
##   ..$ linetype     : num 1
##   ..$ lineend      : chr "butt"
##   ..$ arrow        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ rect                            :List of 5
##   ..$ fill         : chr "white"
##   ..$ colour       : chr "black"
##   ..$ linewidth    : num 0.636
##   ..$ linetype     : num 1
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
##  $ text                            :List of 11
##   ..$ family       : chr ""
##   ..$ face         : chr "plain"
##   ..$ colour       : chr "black"
##   ..$ size         : num 14
##   ..$ hjust        : num 0.5
##   ..$ vjust        : num 0.5
```

```
##   ..$ angle        : num 0
##   ..$ lineheight   : num 0.9
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ title                          : NULL
##  $ aspect.ratio                   : NULL
##  $ axis.title                     : NULL
##  $ axis.title.x                   :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 3.5points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.x.top               :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 0
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 3.5points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.x.bottom            : NULL
##  $ axis.title.y                   :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 3.5points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.y.left              : NULL
```

```
##  $ axis.title.y.right             :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : num -90
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 3.5points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text                      :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : chr "grey30"
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x                    :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 2.8points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x.top                :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 0
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 2.8points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
```

```
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x.bottom              : NULL
##  $ axis.text.y                     :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 1
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 2.8points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.y.left                : NULL
##  $ axis.text.y.right               :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.8points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.theta                 : NULL
##  $ axis.text.r                     :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0.5
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 2.8points 0points 2.8points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.ticks                      : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ axis.ticks.x                    : NULL
##  $ axis.ticks.x.top                : NULL
##  $ axis.ticks.x.bottom             : NULL
##  $ axis.ticks.y                    : NULL
##  $ axis.ticks.y.left               : NULL
##  $ axis.ticks.y.right              : NULL
```

```
##  $ axis.ticks.theta               : NULL
##  $ axis.ticks.r                   : NULL
##  $ axis.minor.ticks.x.top         : NULL
##  $ axis.minor.ticks.x.bottom      : NULL
##  $ axis.minor.ticks.y.left        : NULL
##  $ axis.minor.ticks.y.right       : NULL
##  $ axis.minor.ticks.theta         : NULL
##  $ axis.minor.ticks.r             : NULL
##  $ axis.ticks.length              : 'simpleUnit' num 3.5points
##   ..- attr(*, "unit")= int 8
##  $ axis.ticks.length.x            : NULL
##  $ axis.ticks.length.x.top        : NULL
##  $ axis.ticks.length.x.bottom     : NULL
##  $ axis.ticks.length.y            : NULL
##  $ axis.ticks.length.y.left       : NULL
##  $ axis.ticks.length.y.right      : NULL
##  $ axis.ticks.length.theta        : NULL
##  $ axis.ticks.length.r            : NULL
##  $ axis.minor.ticks.length        : 'rel' num 0.75
##  $ axis.minor.ticks.length.x      : NULL
##  $ axis.minor.ticks.length.x.top  : NULL
##  $ axis.minor.ticks.length.x.bottom: NULL
##  $ axis.minor.ticks.length.y      : NULL
##  $ axis.minor.ticks.length.y.left  : NULL
##  $ axis.minor.ticks.length.y.right : NULL
##  $ axis.minor.ticks.length.theta   : NULL
##  $ axis.minor.ticks.length.r      : NULL
##  $ axis.line                      : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ axis.line.x                    : NULL
##  $ axis.line.x.top                : NULL
##  $ axis.line.x.bottom             : NULL
##  $ axis.line.y                    : NULL
##  $ axis.line.y.left               : NULL
##  $ axis.line.y.right              : NULL
##  $ axis.line.theta                : NULL
##  $ axis.line.r                    : NULL
##  $ legend.background              : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.margin                  : 'margin' num [1:4] 7points 7points 7points 7points
##   ..- attr(*, "unit")= int 8
##  $ legend.spacing                 : 'simpleUnit' num 14points
##   ..- attr(*, "unit")= int 8
##  $ legend.spacing.x               : NULL
##  $ legend.spacing.y               : NULL
##  $ legend.key                     : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.key.size                : 'simpleUnit' num 1.2lines
##   ..- attr(*, "unit")= int 3
##  $ legend.key.height              : NULL
##  $ legend.key.width               : NULL
##  $ legend.key.spacing             : 'simpleUnit' num 7points
##   ..- attr(*, "unit")= int 8
##  $ legend.key.spacing.x           : NULL
```

```
##  $ legend.key.spacing.y           : NULL
##  $ legend.frame                   : NULL
##  $ legend.ticks                   : NULL
##  $ legend.ticks.length            : 'rel' num 0.2
##  $ legend.axis.line               : NULL
##  $ legend.text                    :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.text.position           : NULL
##  $ legend.title                   :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.title.position          : NULL
##  $ legend.position                : chr "right"
##  $ legend.position.inside         : NULL
##  $ legend.direction               : NULL
##  $ legend.byrow                   : NULL
##  $ legend.justification           : chr "center"
##  $ legend.justification.top       : NULL
##  $ legend.justification.bottom    : NULL
##  $ legend.justification.left      : NULL
##  $ legend.justification.right     : NULL
##  $ legend.justification.inside    : NULL
##  $ legend.location                : NULL
##  $ legend.box                     : NULL
##  $ legend.box.just                : NULL
##  $ legend.box.margin              : 'margin' num [1:4] 0cm 0cm 0cm 0cm
##   ..- attr(*, "unit")= int 1
##  $ legend.box.background          : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.box.spacing             : 'simpleUnit' num 14points
##   ..- attr(*, "unit")= int 8
##   [list output truncated]
##  - attr(*, "class")= chr [1:2] "theme" "gg"
```

```
##  - attr(*, "complete")= logi TRUE
##  - attr(*, "validate")= logi TRUE
```

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: It appears that as the depth increases, the temperature decreases. The trend looks like a negative linear correlation.

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

```
#7
lakedepthtemp.regression <-
  lm(NLTERdata$depth ~
       NLTERdata$temperature_C)

summary(lakedepthtemp.regression)
```

```
##
## Call:
## lm(formula = NLTERdata$depth ~ NLTERdata$temperature_C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.7951 -1.3078 -0.2006  1.1548 12.5604
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                9.453719   0.020170   468.7   <2e-16 ***
## NLTERdata$temperature_C -0.394920   0.001473  -268.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.905 on 34754 degrees of freedom
##   (3858 observations deleted due to missingness)
## Multiple R-squared:  0.6742, Adjusted R-squared:  0.6742
## F-statistic: 7.192e+04 on 1 and 34754 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 p-value is very low, (< 2e-16), which means the results are statistically significant. There is strong evidence that changes in depth are correlated with changes in temp. The R-squared value is 0.6742, which means that 67.42% of the variability in lake temperature can be explained by changes in depth. There are 34,754 degrees of freedom, which is a large sample size. This means we have higher confidence in how robust the statistical findings are. For every 1 m of depth change, the model predicts 0.394920 of decrease in temperature. —

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

temp_model <- lm(temperature_C ~ depth + year4 + daynum, data = lakedepth.temp)

step(temp_model)
```

```
## Start:  AIC=26065.53
## temperature_C ~ depth + year4 + daynum
##
##          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 ~ depth + year4 + daynum, data = lakedepth.temp)
##
## Coefficients:
## (Intercept)        depth        year4       daynum
##    -8.57556     -1.94644      0.01134      0.03978
```

```
best.tempmodel <- lm(data = lakedepth.temp,temperature_C ~ depth + year4 + daynum)
summary(best.tempmodel)
```

```
##
## Call:
## lm(formula = temperature_C ~ depth + year4 + daynum, data = lakedepth.temp)
##
## 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
## depth       -1.946437   0.011683 -166.611  < 2e-16 ***
## year4        0.011345   0.004299    2.639  0.00833 **
## daynum       0.039780   0.004317    9.215  < 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
```

```
#10
#The multiple r squared is 0.7412 and adjusted is 0.7411, which means all three variables are well adju

bestmodel.regression <- lm(temperature_C ~ depth + year4 + daynum, data = lakedepth.temp)
```

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 final set of variables includes depth, year, and day of July. The model explains 74.12% of the observed variance in lake temperature. This is an improvement over the last model with just depth that explained 67.42% of the variance.

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
temp_anova <- aov(temperature_C ~ lakename, data = NLTERdata)
summary(temp_anova)
```

```
##                Df  Sum Sq Mean Sq F value Pr(>F)
## lakename        8   57921    7240   155.7 <2e-16 ***
## Residuals   34747 1615571      46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 3858 observations deleted due to missingness
```

```
temp_anova_lm <- lm(temperature_C ~ lakename, data = NLTERdata)
summary(temp_anova_lm)
```

```
##
## Call:
## lm(formula = temperature_C ~ lakename, data = NLTERdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.436  -5.959  -2.559   6.549  24.321
##
```

```
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 16.7363     0.3240  51.660  < 2e-16 ***
## lakenameCrampton Lake       -2.5443     0.3833  -6.638 3.23e-11 ***
## lakenameEast Long Lake      -6.9570     0.3436 -20.248  < 2e-16 ***
## lakenameHummingbird Lake    -6.6985     0.4775 -14.030  < 2e-16 ***
## lakenamePaul Lake           -3.9441     0.3316 -11.893  < 2e-16 ***
## lakenamePeter Lake          -4.4838     0.3309 -13.549  < 2e-16 ***
## lakenameTuesday Lake        -6.3896     0.3368 -18.974  < 2e-16 ***
## lakenameWard Lake           -4.3083     0.4395  -9.802  < 2e-16 ***
## lakenameWest Long Lake      -5.6778     0.3423 -16.587  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.819 on 34747 degrees of freedom
##   (3858 observations deleted due to missingness)
## Multiple R-squared:  0.03461,    Adjusted R-squared:  0.03439
## F-statistic: 155.7 on 8 and 34747 DF,  p-value: < 2.2e-16
```

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

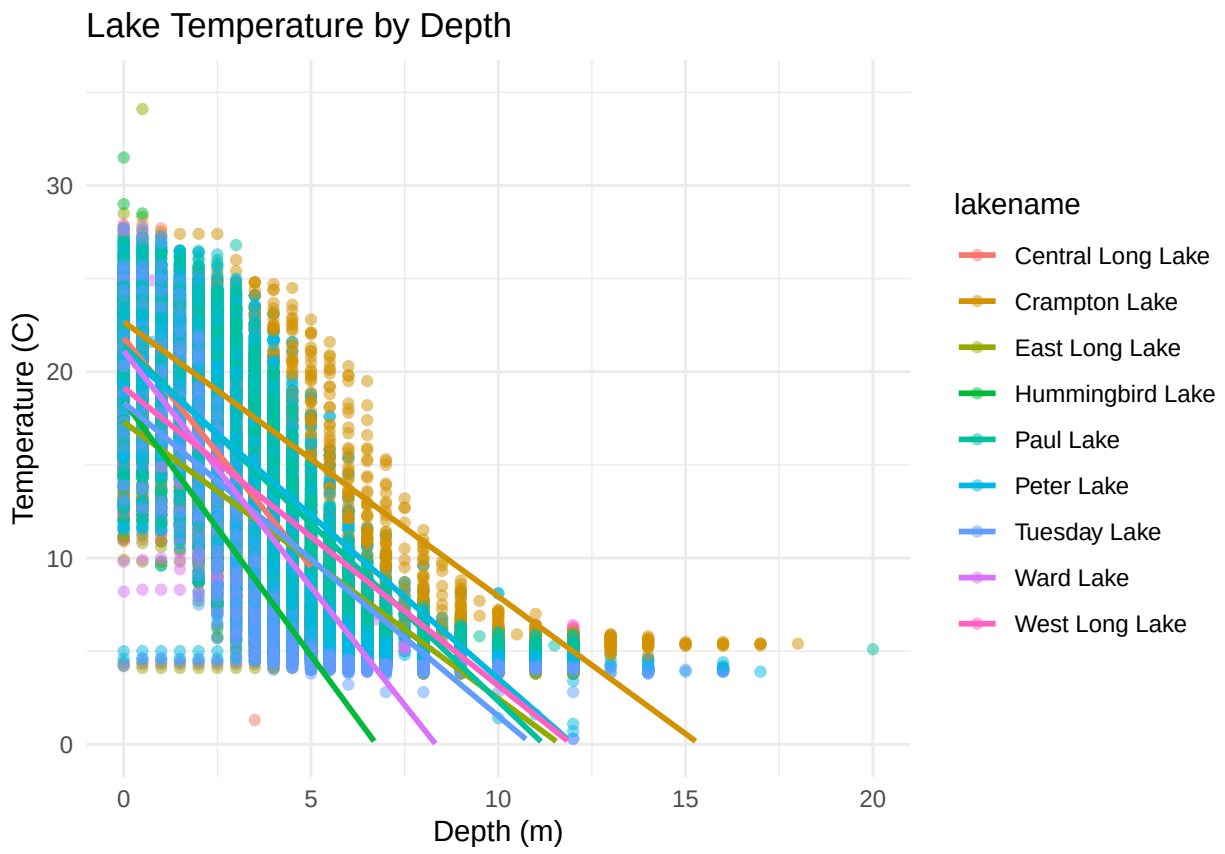
    Answer: The p-value of $< 2.2e\text{-}16$ shows that there is a statistically significant difference in the mean temperature. R-squared value is 0.03461, meaning 3.46% of the variance in temperature can be explained by the lake variable. This means that most of the variability in lake temp is influenced by factors not included in this model.

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.

```r
#14.
tempbydepth <- NLTERdata %>%
  ggplot(aes(x=depth, y=temperature_C, color = lakename))+
  geom_point(alpha=0.5)+
  geom_smooth(method ="lm",se=FALSE)+
  scale_y_continuous(limits = c(0, 35)) +
  labs(title = "Lake Temperature by Depth",
       x = "Depth (m)",
       y = "Temperature (C)") +
  theme_minimal() +
  theme(legend.position = "right")

tempbydepth
```

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



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

```
#15
TukeyHSD(temp_anova)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = temperature_C ~ lakename, data = NLTERdata)
##
## $lakename
##                                         diff        lwr        upr     p adj
## Crampton Lake-Central Long Lake    -2.5442854 -3.7331780 -1.3553927 0.0000000
## East Long Lake-Central Long Lake   -6.9570473 -8.0227648 -5.8913298 0.0000000
## Hummingbird Lake-Central Long Lake -6.6985124 -8.1794348 -5.2175900 0.0000000
## Paul Lake-Central Long Lake        -3.9440682 -4.9727040 -2.9154324 0.0000000
## Peter Lake-Central Long Lake       -4.4837864 -5.5102613 -3.4573116 0.0000000
## Tuesday Lake-Central Long Lake     -6.3896413 -7.4341675 -5.3451152 0.0000000
## Ward Lake-Central Long Lake        -4.3082596 -5.6715463 -2.9449730 0.0000000
## West Long Lake-Central Long Lake   -5.6777623 -6.7395105 -4.6160141 0.0000000
## East Long Lake-Crampton Lake       -4.4127620 -5.1405823 -3.6849417 0.0000000
## Hummingbird Lake-Crampton Lake     -4.1542271 -5.4140285 -2.8944256 0.0000000
## Paul Lake-Crampton Lake            -1.3997828 -2.0721371 -0.7274286 0.0000000
## Peter Lake-Crampton Lake           -1.9395011 -2.6085446 -1.2704575 0.0000000
## Tuesday Lake-Crampton Lake         -3.8453560 -4.5417779 -3.1489341 0.0000000
```

```
## Ward Lake-Crampton Lake            -1.7639743 -2.8831342 -0.6448143 0.0000357
## West Long Lake-Crampton Lake       -3.1334769 -3.8554727 -2.4114812 0.0000000
## Hummingbird Lake-East Long Lake     0.2585349 -0.8857499  1.4028198 0.9987916
## Paul Lake-East Long Lake            3.0129792  2.5954288  3.4305296 0.0000000
## Peter Lake-East Long Lake           2.4732609  2.0610627  2.8854591 0.0000000
## Tuesday Lake-East Long Lake         0.5674060  0.1121132  1.0226989 0.0035472
## Ward Lake-East Long Lake            2.6487877  1.6614645  3.6361109 0.0000000
## West Long Lake-East Long Lake       1.2792850  0.7857610  1.7728091 0.0000000
## Paul Lake-Hummingbird Lake          2.7544443  1.6446129  3.8642756 0.0000000
## Peter Lake-Hummingbird Lake         2.2147260  1.1068972  3.3225548 0.0000000
## Tuesday Lake-Hummingbird Lake       0.3088711 -0.8157039  1.4334461 0.9952041
## Ward Lake-Hummingbird Lake          2.3902528  0.9647057  3.8157999 0.0000071
## West Long Lake-Hummingbird Lake     1.0207501 -0.1198389  2.1613391 0.1224797
## Peter Lake-Paul Lake               -0.5397183 -0.8434372 -0.2359993 0.0000013
## Tuesday Lake-Paul Lake             -2.4455731 -2.8056140 -2.0855323 0.0000000
## Ward Lake-Paul Lake                -0.3641914 -1.3113688  0.5829859 0.9582889
## West Long Lake-Paul Lake           -1.7336941 -2.1410071 -1.3263812 0.0000000
## Tuesday Lake-Peter Lake            -1.9058549 -2.2596746 -1.5520351 0.0000000
## Ward Lake-Peter Lake                0.1755268 -0.7693033  1.1203570 0.9997136
## West Long Lake-Peter Lake          -1.1939759 -1.5958003 -0.7921515 0.0000000
## Ward Lake-Tuesday Lake              2.0813817  1.1169709  3.0457925 0.0000000
## West Long Lake-Tuesday Lake         0.7118790  0.2659563  1.1578017 0.0000259
## West Long Lake-Ward Lake           -1.3695027 -2.3525401 -0.3864652 0.0005266
```

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: Ward Lake and Peter Lake have the same mean temperature, as the p value is 0.9997136 which means there is no significant difference. Central Long Lake is statistically distinct from all other lakes.

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: t-test

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?

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

t_test_lake <- t.test(temperature_C ~ lakename, data = Crampton.Ward)

print(t_test_lake)
```

```
##
##  Welch Two Sample t-test
##
## data:  temperature_C by 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 n
## 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: The p-value of 0.2649 suggests that there is no statistically significant difference. The mean temperatures of the lakes are not equal, but this result is not statistically significant.This does match my result from 16.