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

Andrew Barfield

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.

```
knitr::opts_chunk$set(echo = TRUE)
#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
```

```
here()
```

```
## [1] "/home/guest/EDE_Fall2024"
```

```
Lake.Chemistry.Physics.Raw <-
  read.csv(here("Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv"),
           stringsAsFactors = TRUE)

Lake.Chemistry.Physics.Raw$sampledate <- ymd(Lake.Chemistry.Physics.Raw$sampledate)
```

```
## Warning: 33138 failed to parse.
```

```
#2

my_theme <- theme_classic() +
  theme(
  legend.position = 'bottom',
  plot.background = element_rect(fill = 'lightgrey'),
  plot.title = element_text(size = 15, color = 'black',
  hjust=0.5),
  axis.text = element_text(color = 'black', size = 8),
  axis.title.x = element_text(size = 14),
  axis.title.y = element_text(size = 14),
  legend.title = element_blank()
  )
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: mean lake temperature during July does not change with depth across all lakes Ha: mean lake temperature 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.

```
#4
Lake.Chemistry.Physics.Pipe <-
  Lake.Chemistry.Physics.Raw %>%
  filter(daynum > 182 & daynum < 214) %>%
```

```r
  select(lakename, year4, daynum, depth, temperature_C) %>%
  na.omit()

#5
Lake.Scatterplot <-
  ggplot(Lake.Chemistry.Physics.Pipe, aes(
    x = temperature_C, y = depth)) +
  ylim(0, 35) +
  geom_point(color = "red") +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  labs(x = "Temperature (C)", y = "Depth (Meters)") +
  ggtitle("Temperature vs. Depth Plot (All Lakes)")
  my_theme
```

```
## List of 136
##  $ line                            :List of 6
##   ..$ colour       : chr "black"
##   ..$ linewidth    : num 0.5
##   ..$ 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 0.5
##   ..$ 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 11
##   ..$ 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         : num 14
##   ..$ hjust        : NULL
```

```
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 2.75points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- 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 2.75points 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         : num 14
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 2.75points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- 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 2.75points
##   .. ..- 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 "black"
##   ..$ size         : num 8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- 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] 2.2points 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 2.2points 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 2.2points 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 2.2points
##   .. ..- 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 2.2points 0points 2.2points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.ticks                      :List of 6
##   ..$ colour       : chr "grey20"
##   ..$ linewidth    : NULL
##   ..$ linetype     : NULL
##   ..$ lineend      : NULL
##   ..$ arrow        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "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 2.75points
##   ..- attr(*, "unit")= int 8
```

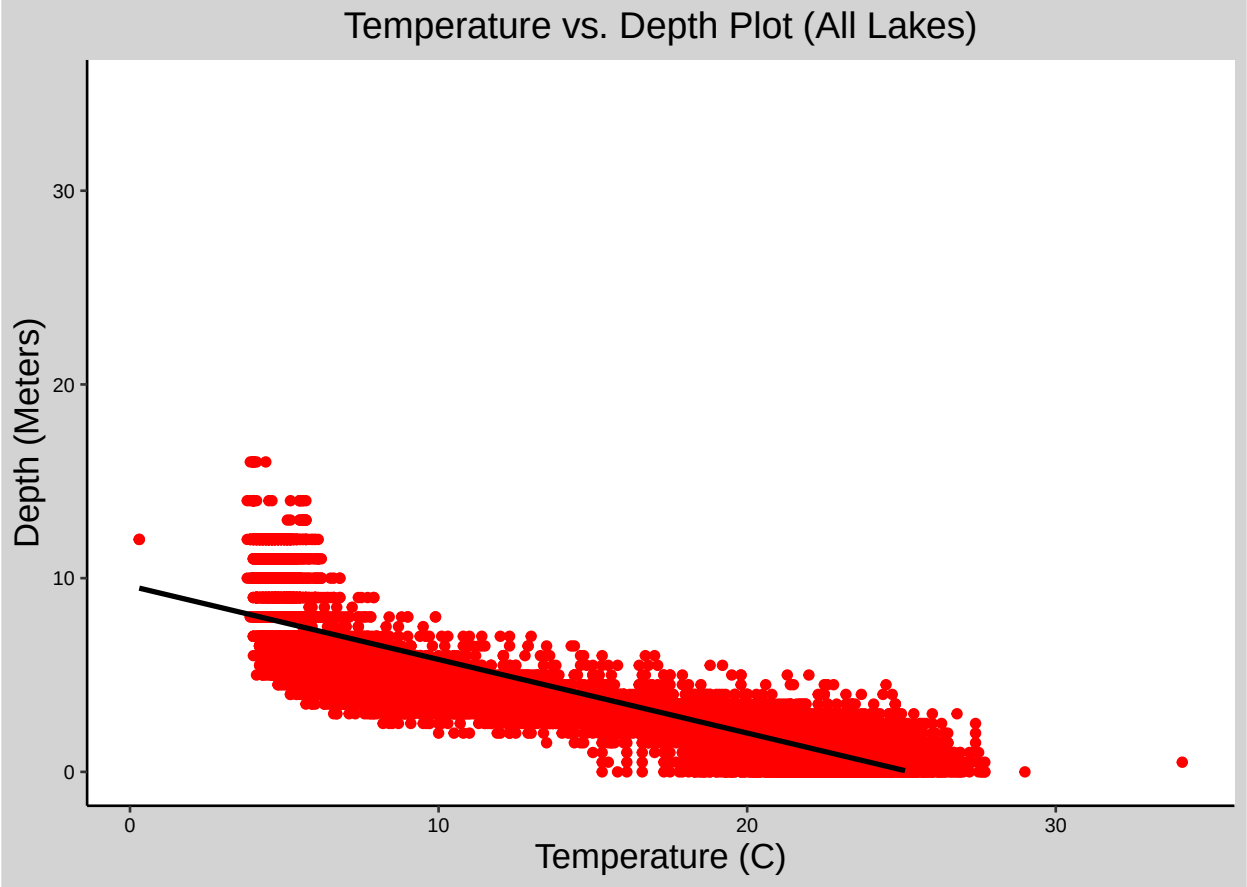
```
##  $ axis.ticks.length.x             : NULL
##  $ axis.ticks.length.x.top         : NULL
##  $ axis.ticks.length.x.bottom      : NULL
##  $ axis.ticks.length.y             : NULL
##  $ axis.ticks.length.y.left        : NULL
##  $ axis.ticks.length.y.right       : NULL
##  $ axis.ticks.length.theta         : 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 of 6
##   ..$ colour       : chr "black"
##   ..$ linewidth    : 'rel' num 1
##   ..$ linetype     : NULL
##   ..$ lineend      : NULL
##   ..$ arrow        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "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 of 5
##   ..$ fill         : NULL
##   ..$ colour       : logi NA
##   ..$ linewidth    : NULL
##   ..$ linetype     : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
##  $ legend.margin                   : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5points
##   ..- attr(*, "unit")= int 8
##  $ legend.spacing                  : 'simpleUnit' num 11points
##   ..- attr(*, "unit")= int 8
##  $ legend.spacing.x                : NULL
##  $ legend.spacing.y                : NULL
##  $ legend.key                      : NULL
##  $ legend.key.size                 : 'simpleUnit' num 1.2lines
##   ..- attr(*, "unit")= int 3
##  $ legend.key.height               : NULL
##  $ legend.key.width                : NULL
##  $ legend.key.spacing              : 'simpleUnit' num 5.5points
##   ..- 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()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.title.position          : NULL
##  $ legend.position                : chr "bottom"
##  $ 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 11points
##   ..- attr(*, "unit")= int 8
##   [list output truncated]
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi TRUE
##  - attr(*, "validate")= logi TRUE
```

```
print(Lake.Scatterplot)
```

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

```
## Warning: Removed 21 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: The figure suggests that temperature increases when depth decreases, showing an inverse relationship between the two variables. The distribution of points suggest that this trend follows a linear pattern.

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

```
#7
Lake.Regression <-
  lm(data = Lake.Chemistry.Physics.Pipe,
    temperature_C ~ depth)
  summary(Lake.Regression)
```

```
##
## Call:
## lm(formula = temperature_C ~ depth, data = Lake.Chemistry.Physics.Pipe)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5606 -3.0380  0.0872  2.9872 13.4706
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.98318    0.06840   321.4   <2e-16 ***
## depth       -1.94086    0.01179  -164.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.852 on 9671 degrees of freedom
## Multiple R-squared:  0.7371, Adjusted R-squared:  0.7371
## F-statistic: 2.712e+04 on 1 and 9671 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 R-squared value is 0.7371, meaning that around 74% of the variability in temperature can be explained by changes in depth across the lake sites. The degrees of freedom on which these findings are based is 9671, meaning that there were a good amount of data points (or observations) that went into this model. The p-value is also much less than 0.05, indicating statistical significance and that the null hypothesis of lake temperature not changing with depth should be rejected. Lastly, the temperature is predicted to change about 1.9 degrees per every 1m change 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.

```
#9

Lake.AIC <- lm(data = Lake.Chemistry.Physics.Pipe, temperature_C ~
                 depth + daynum + year4)
step(Lake.AIC)
```

```
## Start:  AIC=25998.22
## temperature_C ~ depth + daynum + year4
##
##          Df Sum of Sq    RSS   AIC
## <none>                 142056 25998
## - year4   1       201 142257 26010
## - daynum  1      1237 143293 26080
## - depth   1    402549 544605 38995
```

```
##
## Call:
## lm(formula = temperature_C ~ depth + daynum + year4, data = Lake.Chemistry.Physics.Pipe)
##
## Coefficients:
## (Intercept)        depth       daynum        year4
##   -18.19700     -1.94133      0.04024      0.01611
```

```
#10

Lake.Multi.Model <- lm(data = Lake.Chemistry.Physics.Pipe, temperature_C ~
                         depth + daynum + year4)
summary(Lake.Multi.Model)
```

```
##
## Call:
## lm(formula = temperature_C ~ depth + daynum + year4, data = Lake.Chemistry.Physics.Pipe)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.6857 -3.0267  0.1055  2.9937 13.6038
##
## Coefficients:
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept) -18.196998   8.741236   -2.082 0.037392 *
## depth        -1.941328   0.011728 -165.528  < 2e-16 ***
## daynum        0.040237   0.004385    9.176  < 2e-16 ***
## year4         0.016113   0.004353    3.701 0.000216 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.833 on 9669 degrees of freedom
## Multiple R-squared:  0.7398, Adjusted R-squared:  0.7397
## F-statistic:  9162 on 3 and 9669 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 final set of explanatory variables that the AIC method suggests we use to predict temperature in the multi-regression model are all three of the variables in Question 9 (year4, daynum, and depth). In other words, the model is suggesting that the year, day, and depth values can all be used to give us the most accurate prediction of mean temperature. The model explains about 74% of observed variance, or 73.98% to be exact. This is slightly an improvement from the the model using only depth as an explanatory variable, which explained 73.71% of 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
Lakes.AOV <- aov(data = Lake.Chemistry.Physics.Pipe, temperature_C ~ lakename)
summary(Lakes.AOV)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## lakename       8  22188  2773.5   51.18 <2e-16 ***
## Residuals   9664 523706    54.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Lakes.AOV2.LM <- lm(data = Lake.Chemistry.Physics.Pipe, temperature_C ~ lakename)
summary(Lakes.AOV2.LM)
```

```
##
## Call:
## lm(formula = temperature_C ~ lakename, data = Lake.Chemistry.Physics.Pipe)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.773  -6.612  -2.673   7.657  23.813
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                17.6664     0.6507  27.151  < 2e-16 ***
## lakenameCrampton Lake      -2.1851     0.7565  -2.889 0.003879 **
## lakenameEast Long Lake     -7.3795     0.6915 -10.671  < 2e-16 ***
## lakenameHummingbird Lake   -6.6828     0.9571  -6.982 3.09e-12 ***
## lakenamePaul Lake          -3.8234     0.6666  -5.735 1.00e-08 ***
## lakenamePeter Lake         -4.3162     0.6652  -6.489 9.08e-11 ***
## lakenameTuesday Lake       -6.5937     0.6777  -9.730  < 2e-16 ***
## lakenameWard Lake          -3.2078     0.9437  -3.399 0.000679 ***
## lakenameWest Long Lake     -6.0542     0.6893  -8.783  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.361 on 9664 degrees of freedom
## Multiple R-squared:  0.04064,    Adjusted R-squared:  0.03985
## F-statistic: 51.18 on 8 and 9664 DF,  p-value: < 2.2e-16
```

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

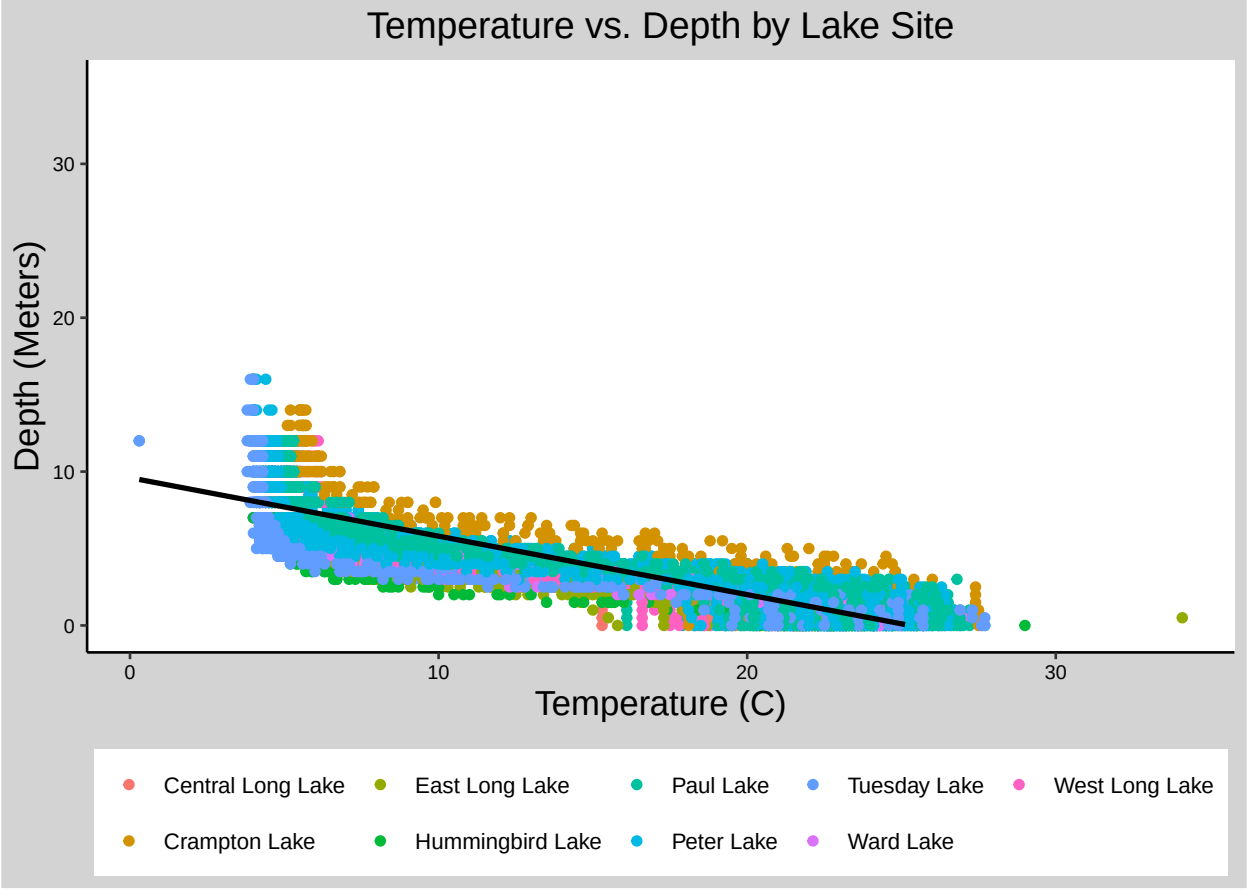
    Answer: Since the p-value for our model is less than 0.05, that means that the null hypothesis, which states that mean temperature is the same across all of our different sites in the month of July, should be rejected. Therefore, there is statistical difference in mean temperature among the different lake sites. However, the results of the ANOVA test reveal that only 4.1% of the variation in mean temperature during July can be explained by difference in lake sites.

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.
Lake.Scatterplot2 <-
  ggplot(Lake.Chemistry.Physics.Pipe, aes(
    x = temperature_C, y = depth, color = lakename)) +
  ylim(0, 35) +
  geom_point() +
  geom_smooth (method = "lm", se = FALSE, color = "black") +
  labs (x = "Temperature (C)", y = "Depth (Meters)") +
  ggtitle ("Temperature vs. Depth by Lake Site") +
  my_theme
print(Lake.Scatterplot2)
```

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

```
## Warning: Removed 21 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(Lakes.AOV)
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
##
## Fit: aov(formula = temperature_C ~ lakename, data = Lake.Chemistry.Physics.Pipe)
##
## $lakename
##                                          diff        lwr        upr     p adj
## Crampton Lake-Central Long Lake    -2.18508757 -4.5319912  0.1618160 0.0915179
## East Long Lake-Central Long Lake   -7.37946293 -9.5249061 -5.2340198 0.0000000
## Hummingbird Lake-Central Long Lake -6.68276989 -9.6520798 -3.7134600 0.0000000
## Paul Lake-Central Long Lake        -3.82336936 -5.8915944 -1.7551443 0.0000004
## Peter Lake-Central Long Lake       -4.31624752 -6.3799766 -2.2525184 0.0000000
## Tuesday Lake-Central Long Lake     -6.59373914 -8.6961647 -4.4913136 0.0000000
## Ward Lake-Central Long Lake        -3.20778556 -6.1355040 -0.2800671 0.0195251
## West Long Lake-Central Long Lake   -6.05416916 -8.1927792 -3.9155591 0.0000000
## East Long Lake-Crampton Lake       -5.19437536 -6.5946962 -3.7940545 0.0000000
## Hummingbird Lake-Crampton Lake     -4.49768232 -6.9825914 -2.0127732 0.0000007
## Paul Lake-Crampton Lake            -1.63828179 -2.9171590 -0.3594045 0.0023129
## Peter Lake-Crampton Lake           -2.13115995 -3.4027534 -0.8595665 0.0000072
## Tuesday Lake-Crampton Lake         -4.40865157 -5.7421303 -3.0751729 0.0000000
## Ward Lake-Crampton Lake            -1.02269799 -3.4577560  1.4123600 0.9307880
## West Long Lake-Crampton Lake       -3.86908159 -5.2589107 -2.4792525 0.0000000
## Hummingbird Lake-East Long Lake     0.69669304 -1.5988991  2.9922852 0.9905616
## Paul Lake-East Long Lake            3.55609357  2.7014024  4.4107847 0.0000000
## Peter Lake-East Long Lake           3.06321541  2.2194620  3.9069688 0.0000000
## Tuesday Lake-East Long Lake         0.78572379 -0.1486934  1.7201409 0.1828556
## Ward Lake-East Long Lake            4.17167737  1.9301428  6.4132120 0.0000003
## West Long Lake-East Long Lake       1.32529377  0.3120836  2.3385039 0.0016418
## Paul Lake-Hummingbird Lake          2.85940053  0.6358062  5.0829949 0.0021745
## Peter Lake-Hummingbird Lake         2.36652237  0.1471092  4.5859355 0.0263810
## Tuesday Lake-Hummingbird Lake       0.08903074 -2.1664094  2.3444709 1.0000000
## Ward Lake-Hummingbird Lake          3.47498433  0.4355186  6.5144501 0.0117238
## West Long Lake-Hummingbird Lake     0.62860073 -1.6606065  2.9178079 0.9952002
## Peter Lake-Paul Lake               -0.49287816 -1.1146082  0.1288519 0.2521516
## Tuesday Lake-Paul Lake             -2.77036979 -3.5104805 -2.0302590 0.0000000
## Ward Lake-Paul Lake                 0.61558380 -1.5521583  2.7833259 0.9939613
## West Long Lake-Paul Lake           -2.23079980 -3.0681906 -1.3934090 0.0000000
## Tuesday Lake-Peter Lake            -2.27749162 -3.0049438 -1.5500394 0.0000000
## Ward Lake-Peter Lake                1.10846196 -1.0549910  3.2719149 0.8108720
## West Long Lake-Peter Lake          -1.73792164 -2.5641457 -0.9116976 0.0000000
## Ward Lake-Tuesday Lake              3.38595358  1.1855572  5.5863500 0.0000641
## West Long Lake-Tuesday Lake         0.53956999 -0.3790495  1.4581895 0.6673292
## West Long Lake-Ward Lake           -2.84638360 -5.0813788 -0.6113884 0.0025399
```

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: Paul Lake and Ward Lake have the same mean temperatures (statistically speaking) as Peter Lake since the resultant p-values for those comparisons to Peter Lake are greater than 0.05. This means that we fail to reject the null hypothesis, which states the mean temperture between sites is the same. None of the lakes have a mean temperature that is statistically distinct from all other lakes, meaning that for each lake site, there is at least one other site where the mean temperatures are statistically considered to be the same. In other words, every lake site has a relationship with at least one other site where the p-value is greater than 0.05 and we fail to reject the null hypothesis.

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: Another test that we could employ to see whetther Peter and Paul Lake have distinct mean temperature is a two-sample T-test, which is used to test the hypothesis of whether or not the mean of two samples are equivalent.

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.Lake <-
  Lake.Chemistry.Physics.Pipe %>%
  filter(lakename == "Crampton Lake" | lakename == "Ward Lake")

Two.Sample.Test <-
  t.test(Crampton.Ward.Lake$temperature_C ~ Crampton.Ward.Lake$lakename)
  Two.Sample.Test
```

```
##
##  Welch Two Sample t-test
##
## data:  Crampton.Ward.Lake$temperature_C by Crampton.Ward.Lake$lakename
## t = 1.2972, df = 192.4, p-value = 0.1961
## alternative hypothesis: true difference in means between group Crampton Lake and group Ward Lake is n
## 95 percent confidence interval:
##  -0.5323014  2.5776973
## sample estimates:
## mean in group Crampton Lake     mean in group Ward Lake
##                    15.48132                    14.45862
```

Answer: The two-sample T-test says that the mean temperature between Crampton and Ward Lake is statistically the same (p-value of 0.1961). The alternative hypothesis given by the test is that the difference in means between the two lake sites is not equal to 0, but since the p-value is greater than 0.05, we fail to reject the null hypothesis and therefore don't accept the alternative. This matches our Tukey HSD test from above and the findings in Question 16. In the Tukey Test, the relationship between Crampton and Ward lake had a p-value of 0.93, also showing that we failed to reject the null hypothesis and that statistically the two sites have the same mean temperature.