Assignment 6: Generalized Linear Models

Tristen Townsend

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics (ENV872L) on generalized linear models.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Use the lesson as a guide. It contains code that can be modified to complete the assignment.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document. Space for your answers is provided in this document and is indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">". You should notice that the answer is highlighted in green by RStudio.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file. You will need to have the correct software installed to do this (see Software Installation Guide) Press the `Knit` button in the RStudio scripting panel. This will save the PDF output in your Assignments folder.
6. After Knitting, please submit the completed exercise (PDF file) to the dropbox in Sakai. Please add your last name into the file name (e.g., "Salk_A06_GLMs.pdf") prior to submission.

The completed exercise is due on Tuesday, 26 February, 2019 before class begins.

Set up your session

1. Set up your session. Upload the EPA Ecotox dataset for Neonicotinoids and the NTL-LTER raw data file for chemistry/physics.

2. Build a ggplot theme and set it as your default theme.

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------ tidyverse 1.2.1 --

## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0

## Warning: package 'tibble' was built under R version 3.5.2

## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(forcats)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(pander)
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(RColorBrewer)
library(colormap)
library(ggpubr)
```

```
## Loading required package: magrittr

##
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract
```

```
#1
getwd()
```

```
## [1] "/Users/Tristen/OneDrive - Duke University/Spring 2019/Data Analytics/Environmental_Data_Analytic
```

```
NTL <- read.csv("./Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")
EPA.ecotox <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Mortality_raw.csv")

#2
tristentheme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right")
```

Neonicotinoids test

Research question: Were studies on various neonicotinoid chemicals conducted in different years?

3. Generate a line of code to determine how many different chemicals are listed in the Chemical.Name column.

4. Are the publication years associated with each chemical well-approximated by a normal distribution? Run the appropriate test and also generate a frequency polygon to illustrate the distribution of counts for each year, divided by chemical name. Bonus points if you can generate the results of your test from a pipe function. No need to make this graph pretty.

5. Is there equal variance among the publication years for each chemical? Hint: var.test is not the correct function.

```
#3
class(EPA.ecotox$Chemical.Name) #Factor
```

```
## [1] "factor"
```

```r
levels(EPA.ecotox$Chemical.Name) #There are 9 chemicals in this column
```

```
## [1] "Acetamiprid"  "Clothianidin" "Dinotefuran"  "Imidacloprid"
## [5] "Imidaclothiz" "Nitenpyram"   "Nithiazine"   "Thiacloprid"
## [9] "Thiamethoxam"
```

```r
#4
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Acetamiprid"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Acetamiprid"]
## W = 0.90191, p-value = 5.706e-08
```

```r
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Clothianidin"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Clothianidin"]
## W = 0.69577, p-value = 4.287e-11
```

```r
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Dinotefuran"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Dinotefuran"]
## W = 0.82848, p-value = 8.83e-07
```

```r
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Imidacloprid"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Imidacloprid"]
## W = 0.88178, p-value < 2.2e-16
```

```r
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Imidaclothiz"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Imidaclothiz"]
## W = 0.68429, p-value = 0.00093
```

```r
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Nitenpyram"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Nitenpyram"]
## W = 0.79592, p-value = 0.0005686
```

```r
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Nithiazine"])
```

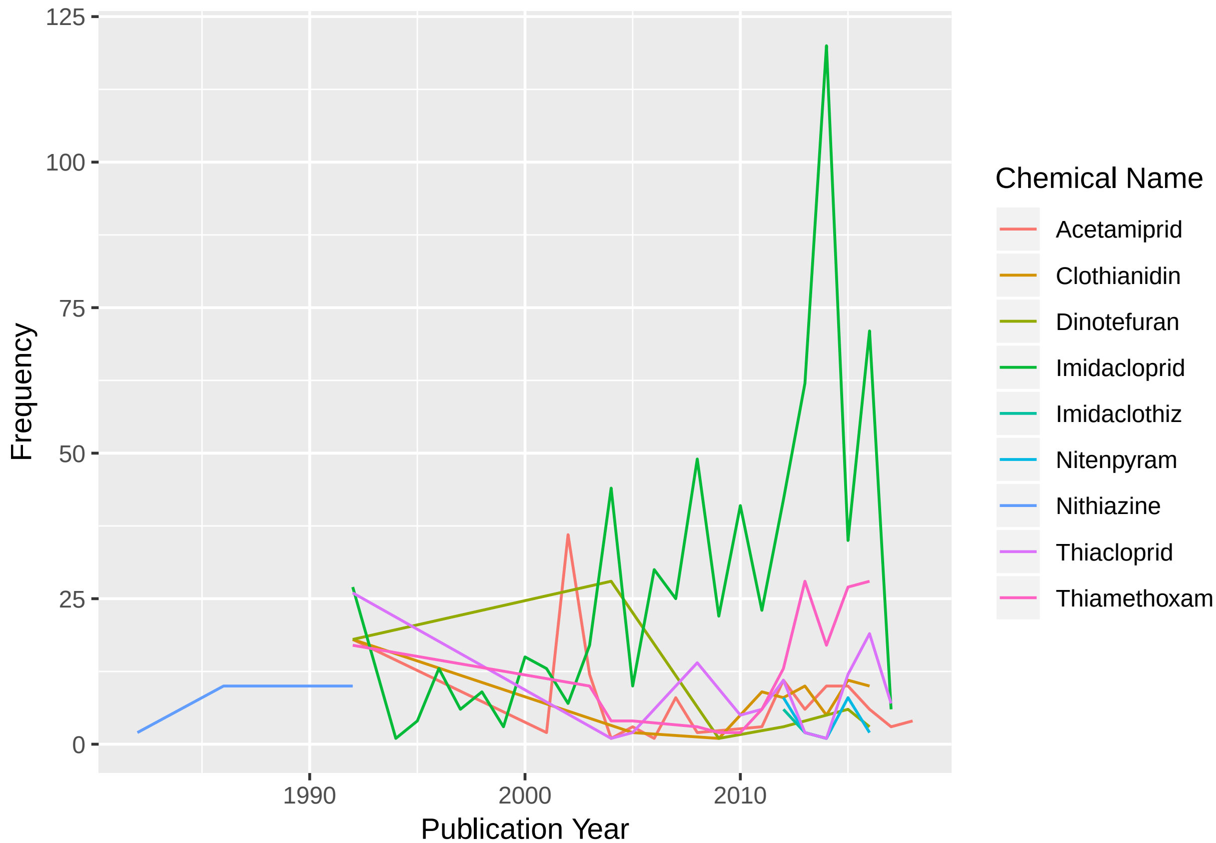
```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Nithiazine"]
## W = 0.75938, p-value = 0.0001235
```

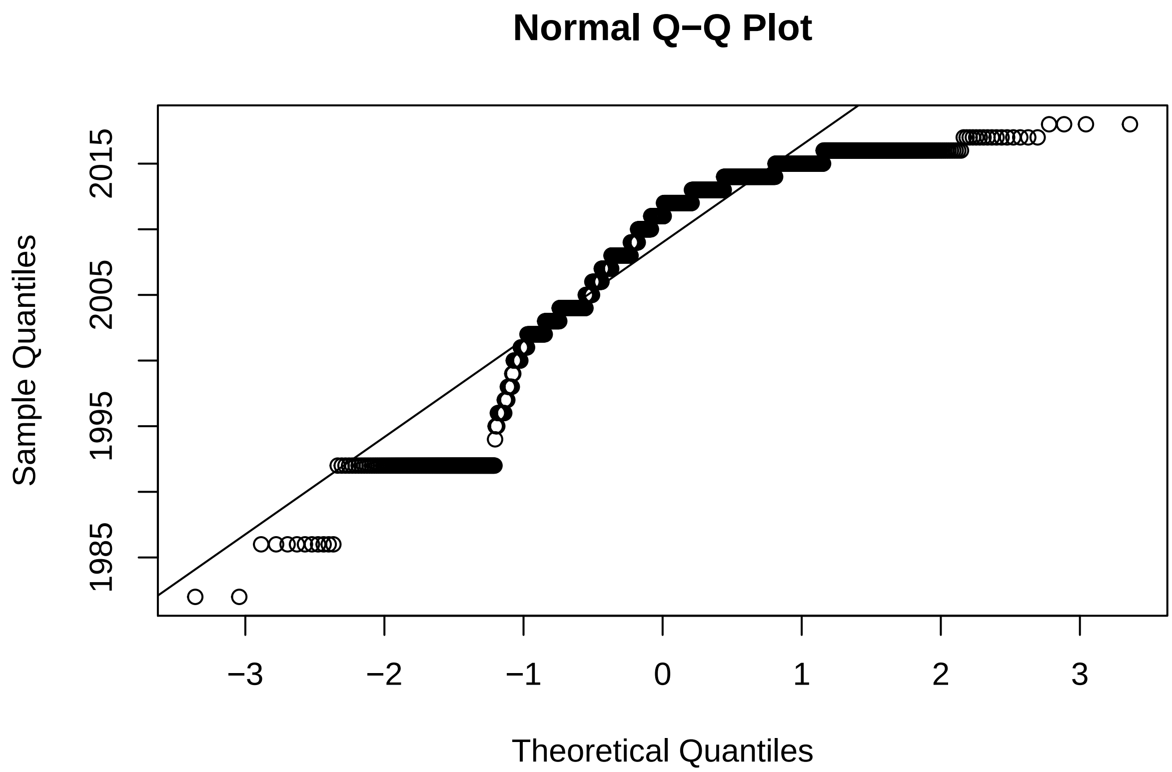
```
shapiro.test(EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Thiacloprid"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA.ecotox$Pub..Year[EPA.ecotox$Chemical.Name == "Thiacloprid"]
## W = 0.7669, p-value = 1.118e-11
```

```
EPA_Ecotox_Frequency <-
  ggplot(EPA.ecotox) +
  geom_freqpoly(aes(x = Pub..Year, color = Chemical.Name),
                stat="count") +
  labs(x = "Publication Year", y = "Frequency",
                color = "Chemical Name")
print(EPA_Ecotox_Frequency)
```




```
qqnorm(EPA.ecotox$Pub..Year); qqline(EPA.ecotox$Pub..Year)
```




```r
#5 Bartlett test of homogeneity of variances
bartlett.test(EPA.ecotox$Pub..Year ~ EPA.ecotox$Chemical.Name)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  EPA.ecotox$Pub..Year by EPA.ecotox$Chemical.Name
## Bartlett's K-squared = 139.59, df = 8, p-value < 2.2e-16
```

```r
#Bartlett's K-squared = 139.59, df = 8, p-value < 2.2e-16
```

6. Based on your results, which test would you choose to run to answer your research question?

   ANSWER: Anova GLM

7. Run this test below.

8. Generate a boxplot representing the range of publication years for each chemical. Adjust your graph to make it pretty.

```r
#7
Chem <- aov(EPA.ecotox$Pub..Year ~ EPA.ecotox$Chemical.Name)
summary(Chem)
```

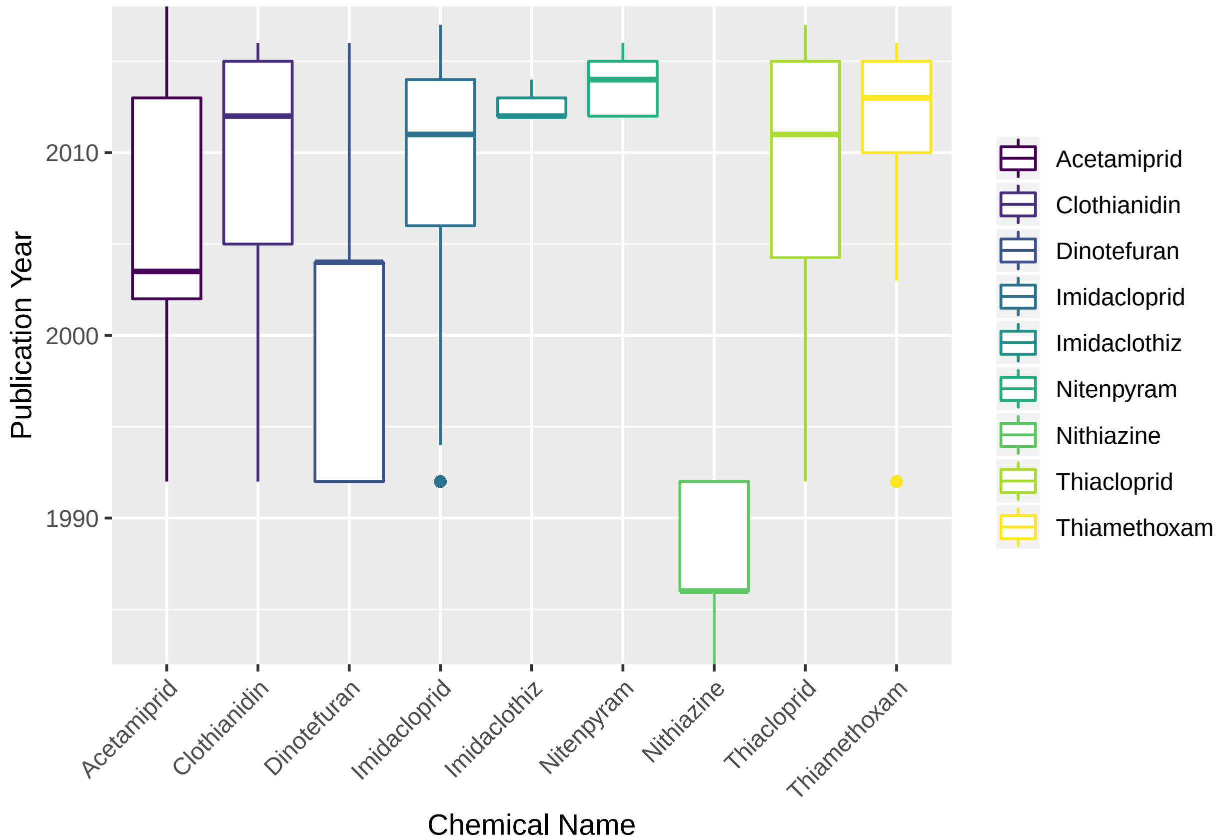
```
##                            Df Sum Sq Mean Sq F value Pr(>F)
## EPA.ecotox$Chemical.Name    8  13365  1670.7   33.21 <2e-16 ***
## Residuals                1274  64093    50.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(Chem)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = EPA.ecotox$Pub..Year ~ EPA.ecotox$Chemical.Name)
##
## $`EPA.ecotox$Chemical.Name`
##                                 diff          lwr          upr     p adj
## Clothianidin-Acetamiprid    2.0478935  -1.13556203   5.231348994 0.5444735
## Dinotefuran-Acetamiprid    -3.4333250  -6.86887521   0.002225165 0.0502982
## Imidacloprid-Acetamiprid    3.1181443   1.05175043   5.184538190 0.0001059
## Imidaclothiz-Acetamiprid    6.4517974  -1.13341497  14.037009746 0.1700689
## Nitenpyram-Acetamiprid      7.7216387   2.55455876  12.888718547 0.0001312
## Nithiazine-Acetamiprid    -17.6290107 -22.69334307 -12.564678323 0.0000000
## Thiacloprid-Acetamiprid     1.6394284  -1.21592420   4.494781028 0.6929485
## Thiamethoxam-Acetamiprid    4.3738126   1.80714109   6.940484045 0.0000050
## Dinotefuran-Clothianidin   -5.4812185  -9.32765358  -1.634783428 0.0003529
## Imidacloprid-Clothianidin   1.0702508  -1.62456668   3.765068330 0.9489438
## Imidaclothiz-Clothianidin   4.4039039  -3.37603853  12.183846336 0.7094335
## Nitenpyram-Clothianidin     5.6737452   0.22482121  11.122669133 0.0338611
## Nithiazine-Clothianidin   -19.6769042 -25.02849460 -14.325313751 0.0000000
## Thiacloprid-Clothianidin   -0.4084651  -3.74689228   2.929962144 0.9999879
## Thiamethoxam-Clothianidin   2.3259191  -0.76921583   5.421054007 0.3218154
## Imidacloprid-Dinotefuran    6.5514693   3.56304877   9.539889900 0.0000000
## Imidaclothiz-Dinotefuran    9.8851224   1.99867071  17.771574107 0.0033119
## Nitenpyram-Dinotefuran     11.1549637   5.55501829  16.754909074 0.0000000
## Nithiazine-Dinotefuran    -14.1956857 -19.70096824  -8.690403099 0.0000000
## Thiacloprid-Dinotefuran     5.0727534   1.49312883   8.652378050 0.0003937
## Thiamethoxam-Dinotefuran    7.8071376   4.45326278  11.161012409 0.0000000
## Imidaclothiz-Imidacloprid   3.3336531  -4.05979665  10.727102808 0.8976481
## Nitenpyram-Imidacloprid     4.6034943  -0.27773162   9.484720319 0.0825706
## Nithiazine-Imidacloprid   -20.7471550 -25.51948303 -15.974826981 0.0000000
## Thiacloprid-Imidacloprid   -1.4787159  -3.77669129   0.819259503 0.5440493
## Thiamethoxam-Imidacloprid   1.2556683  -0.67188326   3.183219776 0.5266734
## Nitenpyram-Imidaclothiz     1.2698413  -7.51035405  10.050036590 0.9999561
## Nithiazine-Imidaclothiz   -24.0808081 -32.80093294 -15.360683218 0.0000000
## Thiacloprid-Imidaclothiz   -4.8123690 -12.46391482   2.839176871 0.5758935
## Thiamethoxam-Imidaclothiz  -2.0779848  -9.62655539   5.470585757 0.9950400
## Nithiazine-Nitenpyram     -25.3506494 -32.07402988 -18.627268825 0.0000000
## Thiacloprid-Nitenpyram     -6.0822102 -11.34618420  -0.818236282 0.0103350
## Thiamethoxam-Nitenpyram    -3.3478261  -8.46096463   1.765312458 0.5194316
## Thiacloprid-Nithiazine     19.2684391  14.10528410  24.431594116 0.0000000
## Thiamethoxam-Nithiazine    22.0028233  16.99353853  27.012107998 0.0000000
## Thiamethoxam-Thiacloprid    2.7343842  -0.02215529   5.490923603 0.0538087
```

```r
#8
Chem_Box_plot <-
  ggplot(EPA.ecotox, aes(x = Chemical.Name, y = Pub..Year, color = Chemical.Name)) +
  geom_boxplot() +
  labs(x= "Chemical Name", y="Publication Year", color = NULL) +
  scale_y_continuous(expand = c(0, 0)) +
  #scale_color_manual(values = c("#7fcdbb", "yellow", "#1d91c0", "black", "green", "violet", "pink", "#
  scale_color_viridis(discrete = TRUE) +
```

```
  theme(axis.text.x = element_text(angle = 45,  hjust = 1))
print(Chem_Box_plot)
```



9. How would you summarize the conclusion of your analysis? Include a sentence summarizing your findings and include the results of your test in parentheses at the end of the sentence.

   ANSWER: There exists significant differences in the mean publication year for various chemicals (acetamiprid, clothianidin, dinotefuran, imidacloprid, imidaclothiz, nitenpyram, nithiazine, thiacloprid, thiamethoxam) (p-value < 0.05, df $= 8$).

NTL-LTER test

Research question: What is the best set of predictors for lake temperatures in July across the monitoring period at the North Temperate Lakes LTER?

11. Wrangle your NTL-LTER dataset with a pipe function so that it contains only the following criteria:

- Only dates in July (hint: use the daynum column). No need to consider leap years.
- Only the columns: lakename, year4, daynum, depth, temperature_C
- Only complete cases (i.e., remove NAs)

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

```
#11
NTL_processed <- NTL %>% filter(daynum == 182:212) %>%
  select(lakename, year4, daynum, depth, temperature_C) %>%
  filter(!is.na(temperature_C) & !is.na(depth))
```

```
## Warning in daynum == 182:212: longer object length is not a multiple of
## shorter object length
```

```r
#12

NTL_1 <- lm(NTL_processed$temperature_C ~ NTL_processed$lakename + NTL_processed$year4 + NTL_processed$d
summary(NTL_1)
```

```
##
## Call:
## lm(formula = NTL_processed$temperature_C ~ NTL_processed$lakename +
##     NTL_processed$year4 + NTL_processed$daynum + NTL_processed$depth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5410 -2.9613  0.1268  2.8153 11.3393
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                       11.6123988 47.3388708   0.245
## NTL_processed$lakenameCrampton Lake     4.1824152  2.7864181   1.501
## NTL_processed$lakenameEast Long Lake   -0.2752175  2.6376999  -0.104
## NTL_processed$lakenameHummingbird Lake -2.5876240  3.0266761  -0.855
## NTL_processed$lakenamePaul Lake         2.2642841  2.5884021   0.875
## NTL_processed$lakenamePeter Lake        3.3398972  2.5846910   1.292
## NTL_processed$lakenameTuesday Lake      0.3166373  2.6138862   0.121
## NTL_processed$lakenameWard Lake        -0.1899035  3.1560069  -0.060
## NTL_processed$lakenameWest Long Lake    1.2167240  2.6295260   0.463
## NTL_processed$year4                     0.0001768  0.0235463   0.008
## NTL_processed$daynum                    0.0413728  0.0225936   1.831
## NTL_processed$depth                    -1.9633331  0.0627321 -31.297
##                                     Pr(>|t|)
## (Intercept)                           0.8064
## NTL_processed$lakenameCrampton Lake     0.1344
## NTL_processed$lakenameEast Long Lake    0.9170
## NTL_processed$lakenameHummingbird Lake  0.3933
## NTL_processed$lakenamePaul Lake         0.3824
## NTL_processed$lakenamePeter Lake        0.1973
## NTL_processed$lakenameTuesday Lake      0.9037
## NTL_processed$lakenameWard Lake         0.9521
## NTL_processed$lakenameWest Long Lake    0.6439
## NTL_processed$year4                     0.9940
## NTL_processed$daynum                    0.0681 .
## NTL_processed$depth                     <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.613 on 300 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.7674
## F-statistic:  94.3 on 11 and 300 DF,  p-value: < 2.2e-16
```

```r
NTL_2 <- lm(NTL_processed$temperature_C ~ NTL_processed$lakename + NTL_processed$daynum + NTL_processed$
summary(NTL_2)
```

```
##
## Call:
## lm(formula = NTL_processed$temperature_C ~ NTL_processed$lakename +
##     NTL_processed$daynum + NTL_processed$depth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5385 -2.9612  0.1263  2.8169 11.3365
##
## Coefficients:
##                                      Estimate Std. Error t value
## (Intercept)                          11.96570    5.09506   2.348
## NTL_processed$lakenameCrampton Lake   4.18412    2.77248   1.509
## NTL_processed$lakenameEast Long Lake -0.27505    2.63322  -0.104
## NTL_processed$lakenameHummingbird Lake -2.58667  3.01896  -0.857
## NTL_processed$lakenamePaul Lake       2.26524    2.58096   0.878
## NTL_processed$lakenamePeter Lake      3.34080    2.57762   1.296
## NTL_processed$lakenameTuesday Lake    0.31715    2.60863   0.122
## NTL_processed$lakenameWard Lake      -0.18697    3.12647  -0.060
## NTL_processed$lakenameWest Long Lake  1.21703    2.62484   0.464
## NTL_processed$daynum                  0.04137    0.02255   1.835
## NTL_processed$depth                  -1.96332    0.06260 -31.361
##                                      Pr(>|t|)
## (Intercept)                            0.0195 *
## NTL_processed$lakenameCrampton Lake    0.1323
## NTL_processed$lakenameEast Long Lake   0.9169
## NTL_processed$lakenameHummingbird Lake 0.3922
## NTL_processed$lakenamePaul Lake        0.3808
## NTL_processed$lakenamePeter Lake       0.1959
## NTL_processed$lakenameTuesday Lake     0.9033
## NTL_processed$lakenameWard Lake        0.9524
## NTL_processed$lakenameWest Long Lake   0.6432
## NTL_processed$daynum                   0.0676 .
## NTL_processed$depth                    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.607 on 301 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.7682
## F-statistic: 104.1 on 10 and 301 DF,  p-value: < 2.2e-16
```

```r
NTL_3 <- lm(NTL_processed$temperature_C ~ NTL_processed$lakename + NTL_processed$depth)
summary(NTL_3)
```

```
##
## Call:
## lm(formula = NTL_processed$temperature_C ~ NTL_processed$lakename +
##     NTL_processed$depth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6035 -2.8985 -0.0256  2.8815 11.9500
##
## Coefficients:
##                                      Estimate Std. Error t value
```

```
## (Intercept)                              20.04471    2.57258   7.792
## NTL_processed$lakenameCrampton Lake     4.17829    2.78331   1.501
## NTL_processed$lakenameEast Long Lake   -0.10632    2.64190  -0.040
## NTL_processed$lakenameHummingbird Lake -2.42447    3.02946  -0.800
## NTL_processed$lakenamePaul Lake         2.37121    2.59040   0.915
## NTL_processed$lakenamePeter Lake        3.38412    2.58759   1.308
## NTL_processed$lakenameTuesday Lake      0.33726    2.61881   0.129
## NTL_processed$lakenameWard Lake         0.06044    3.13577   0.019
## NTL_processed$lakenameWest Long Lake    1.23339    2.63508   0.468
## NTL_processed$depth                    -1.96118    0.06284 -31.210
##                                        Pr(>|t|)
## (Intercept)                            1.07e-13 ***
## NTL_processed$lakenameCrampton Lake       0.134
## NTL_processed$lakenameEast Long Lake      0.968
## NTL_processed$lakenameHummingbird Lake    0.424
## NTL_processed$lakenamePaul Lake           0.361
## NTL_processed$lakenamePeter Lake          0.192
## NTL_processed$lakenameTuesday Lake        0.898
## NTL_processed$lakenameWard Lake           0.985
## NTL_processed$lakenameWest Long Lake      0.640
## NTL_processed$depth                     < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.621 on 302 degrees of freedom
## Multiple R-squared:  0.7732, Adjusted R-squared:  0.7664
## F-statistic: 114.4 on 9 and 302 DF,  p-value: < 2.2e-16
```

```r
AIC(NTL_1, NTL_2, NTL_3) #Choose model 3 - most parsimonious - not much difference between AIC for each
```

```
##       df      AIC
## NTL_1 13 1700.677
## NTL_2 12 1698.677
## NTL_3 11 1700.147
```

13. What is the final linear equation to predict temperature from your multiple regression? How much of the observed variance does this model explain?

    ANSWER: The final linear equation to predict temperature is as follows: Temperature = 20.04471 + coef*(lakename) - 1.96118(depth). This model explains about 77% of the variance.

14. Run an interaction effects ANCOVA to predict temperature based on depth and lakename from the same wrangled dataset.

```r
#14
NTL_interaction <- lm(NTL_processed$temperature_C ~ NTL_processed$lakename * NTL_processed$depth)

summary(NTL_interaction)
```

```
##
## Call:
## lm(formula = NTL_processed$temperature_C ~ NTL_processed$lakename *
##     NTL_processed$depth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6100 -2.7826 -0.2609  2.8225 12.1803
```

```
##
## Coefficients: (1 not defined because of singularities)
##                                                          Estimate
## (Intercept)                                              19.54590
## NTL_processed$lakenameCrampton Lake                       4.35631
## NTL_processed$lakenameEast Long Lake                     -1.26077
## NTL_processed$lakenameHummingbird Lake                   -0.04503
## NTL_processed$lakenamePaul Lake                           3.76866
## NTL_processed$lakenamePeter Lake                          3.98209
## NTL_processed$lakenameTuesday Lake                        0.59795
## NTL_processed$lakenameWard Lake                           8.38017
## NTL_processed$lakenameWest Long Lake                      1.12336
## NTL_processed$depth                                      -1.83647
## NTL_processed$lakenameCrampton Lake:NTL_processed$depth   -0.04629
## NTL_processed$lakenameEast Long Lake:NTL_processed$depth   0.22998
## NTL_processed$lakenameHummingbird Lake:NTL_processed$depth -0.64710
## NTL_processed$lakenamePaul Lake:NTL_processed$depth       -0.31070
## NTL_processed$lakenamePeter Lake:NTL_processed$depth      -0.14529
## NTL_processed$lakenameTuesday Lake:NTL_processed$depth    -0.07844
## NTL_processed$lakenameWard Lake:NTL_processed$depth       -1.91234
## NTL_processed$lakenameWest Long Lake:NTL_processed$depth        NA
##                                                          Std. Error
## (Intercept)                                                 2.66139
## NTL_processed$lakenameCrampton Lake                         3.13662
## NTL_processed$lakenameEast Long Lake                        2.87928
## NTL_processed$lakenameHummingbird Lake                      3.81149
## NTL_processed$lakenamePaul Lake                             2.76530
## NTL_processed$lakenamePeter Lake                            2.73251
## NTL_processed$lakenameTuesday Lake                          2.84690
## NTL_processed$lakenameWard Lake                             5.40085
## NTL_processed$lakenameWest Long Lake                        2.62700
## NTL_processed$depth                                         0.19244
## NTL_processed$lakenameCrampton Lake:NTL_processed$depth     0.36216
## NTL_processed$lakenameEast Long Lake:NTL_processed$depth    0.27078
## NTL_processed$lakenameHummingbird Lake:NTL_processed$depth  0.64119
## NTL_processed$lakenamePaul Lake:NTL_processed$depth         0.23393
## NTL_processed$lakenamePeter Lake:NTL_processed$depth        0.21836
## NTL_processed$lakenameTuesday Lake:NTL_processed$depth      0.25389
## NTL_processed$lakenameWard Lake:NTL_processed$depth         1.01063
## NTL_processed$lakenameWest Long Lake:NTL_processed$depth         NA
##                                                          t value
## (Intercept)                                                7.344
## NTL_processed$lakenameCrampton Lake                        1.389
## NTL_processed$lakenameEast Long Lake                      -0.438
## NTL_processed$lakenameHummingbird Lake                    -0.012
## NTL_processed$lakenamePaul Lake                            1.363
## NTL_processed$lakenamePeter Lake                           1.457
## NTL_processed$lakenameTuesday Lake                         0.210
## NTL_processed$lakenameWard Lake                            1.552
## NTL_processed$lakenameWest Long Lake                       0.428
## NTL_processed$depth                                       -9.543
## NTL_processed$lakenameCrampton Lake:NTL_processed$depth   -0.128
## NTL_processed$lakenameEast Long Lake:NTL_processed$depth   0.849
## NTL_processed$lakenameHummingbird Lake:NTL_processed$depth -1.009
```

```
## NTL_processed$lakenamePaul Lake:NTL_processed$depth         -1.328
## NTL_processed$lakenamePeter Lake:NTL_processed$depth        -0.665
## NTL_processed$lakenameTuesday Lake:NTL_processed$depth      -0.309
## NTL_processed$lakenameWard Lake:NTL_processed$depth         -1.892
## NTL_processed$lakenameWest Long Lake:NTL_processed$depth        NA
##                                                             Pr(>|t|)
## (Intercept)                                                 2.03e-12 ***
## NTL_processed$lakenameCrampton Lake                           0.1659
## NTL_processed$lakenameEast Long Lake                          0.6618
## NTL_processed$lakenameHummingbird Lake                        0.9906
## NTL_processed$lakenamePaul Lake                               0.1740
## NTL_processed$lakenamePeter Lake                              0.1461
## NTL_processed$lakenameTuesday Lake                            0.8338
## NTL_processed$lakenameWard Lake                               0.1218
## NTL_processed$lakenameWest Long Lake                          0.6692
## NTL_processed$depth                                          < 2e-16 ***
## NTL_processed$lakenameCrampton Lake:NTL_processed$depth       0.8984
## NTL_processed$lakenameEast Long Lake:NTL_processed$depth      0.3964
## NTL_processed$lakenameHummingbird Lake:NTL_processed$depth    0.3137
## NTL_processed$lakenamePaul Lake:NTL_processed$depth           0.1851
## NTL_processed$lakenamePeter Lake:NTL_processed$depth          0.5063
## NTL_processed$lakenameTuesday Lake:NTL_processed$depth        0.7576
## NTL_processed$lakenameWard Lake:NTL_processed$depth           0.0594 .
## NTL_processed$lakenameWest Long Lake:NTL_processed$depth          NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.603 on 295 degrees of freedom
## Multiple R-squared:  0.7806, Adjusted R-squared:  0.7687
## F-statistic: 65.59 on 16 and 295 DF,  p-value: < 2.2e-16
```

15. Is there an interaction between depth and lakename? How much variance in the temperature observations does this explain?

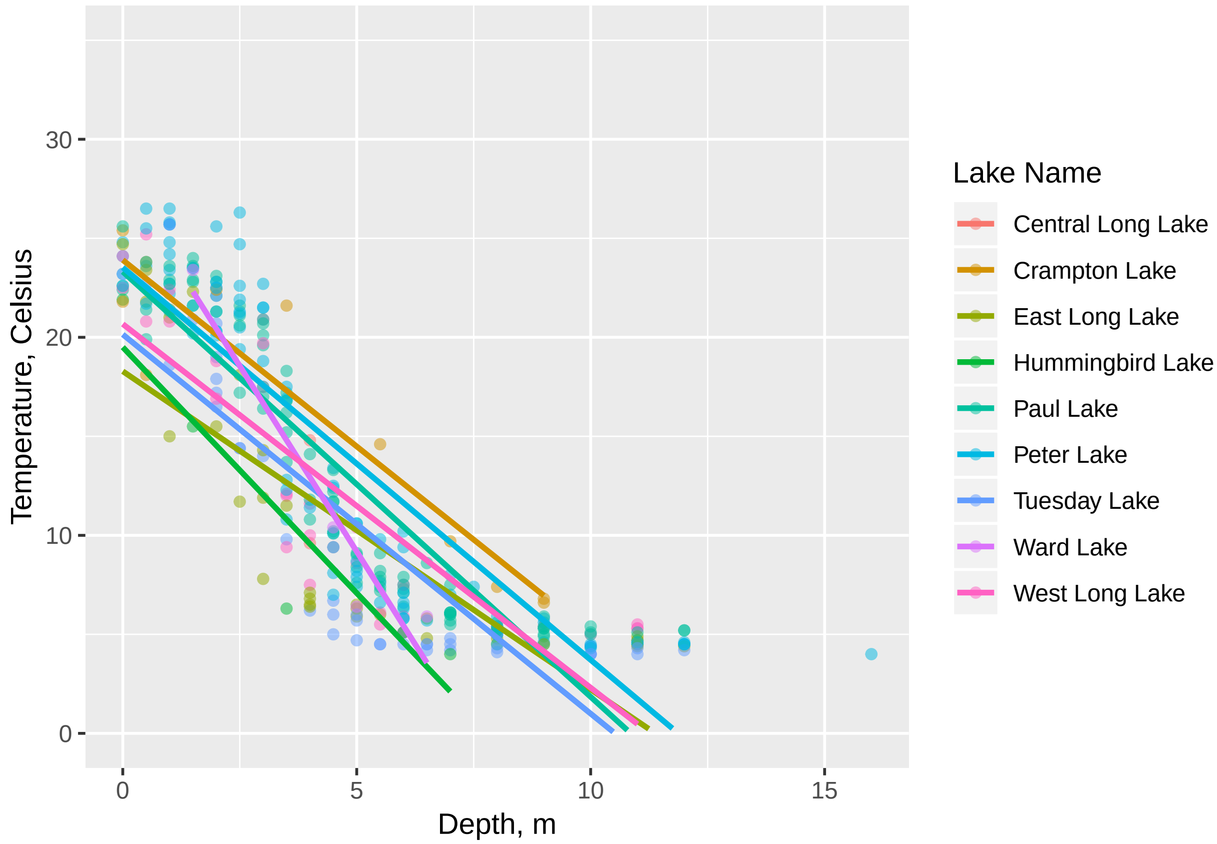
    ANSWER: No. The interaction is not explaining much and rather the depth alone is explaining most variance.

16. 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.

```
#16

NTL_plot <- ggplot(NTL_processed, aes(x = depth, y = temperature_C, color = lakename)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE) +
  ylim(0,35) +
  labs(x = "Depth, m", y = "Temperature, Celsius", color = "Lake Name")
print(NTL_plot)
```

```
## Warning: Removed 44 rows containing missing values (geom_smooth).
```



30
20
10
0
0
5
10
15
Temperature, Celsius
Depth, m
Lake Name
Central Long Lake
Crampton Lake
East Long Lake
Hummingbird Lake
Paul Lake
Peter Lake
Tuesday Lake
Ward Lake
West Long Lake