Assignment 6: Generalized Linear Models

Felipe Raby Amadori

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.

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE, eval=TRUE)

#1
getwd()
```

```
## [1] "C:/Users/Felipe/OneDrive - Duke University/1. DUKE/1. Ramos 2 Semestre/EOS-872 Env. Data Analyt
```

```
#install.packages('FSA')
library(FSA)
```

```
## ## FSA v0.8.22. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.
```

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------- tidyverse 1.2.1 --
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

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

```
library(RColorBrewer)
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
```

```
library(viridis)
```

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

```
library(colormap)

EPA_Ecotox_Neonicotinoids <- read.csv("./Data/Raw/ECOTOX_Neonicotinoids_Mortality_raw.csv")

NTL_TER_raw_chemistryphysics <- read.csv("./Data/Raw/NTL-LTER_Lake_ChemistryPhysics_Raw.csv")

#2
felipe_theme <- theme_light(base_size = 12) +
  theme(axis.text = element_text(color = "grey8"),
        legend.position = "right", plot.title = element_text(hjust = 0.5))
theme_set(felipe_theme)
```

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
levels(EPA_Ecotox_Neonicotinoids$Chemical.Name)
```

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

```r
length(levels(EPA_Ecotox_Neonicotinoids$Chemical.Name))
```

```
## [1] 9
```

```r
# Numbers of different chemicals listed in the Chemical.Name column.

#4
# To evaluate assumption of normal distribution we run the Shapiro-Wilk test
# for each chemical, where the null hypothesis is that the data come from a normal distribution.

shapiro.test(EPA_Ecotox_Neonicotinoids$Pub..Year[EPA_Ecotox_Neonicotinoids$Chemical.Name
                                                 == "Acetamiprid"])
```

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

```r
# Acetamiprid data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.90191, p-value = 5.706e-08).

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

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

```r
# Clothianidin data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.69577, p-value = 4.287e-11).

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

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

```r
# Dinotefuran data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.82848, p-value = 8.83e-07).

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

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

```r
# Imidacloprid data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.88178, p-value < 2.2e-16).
```

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

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

```r
# Imidaclothiz data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.68429, p-value = 0.00093).

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

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

```r
# Nitenpyram data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.79592, p-value = 0.0005686).

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

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA_Ecotox_Neonicotinoids$Pub..Year[EPA_Ecotox_Neonicotinoids$Chemical.Name ==     "Nithiazin
## W = 0.75938, p-value = 0.0001235
```

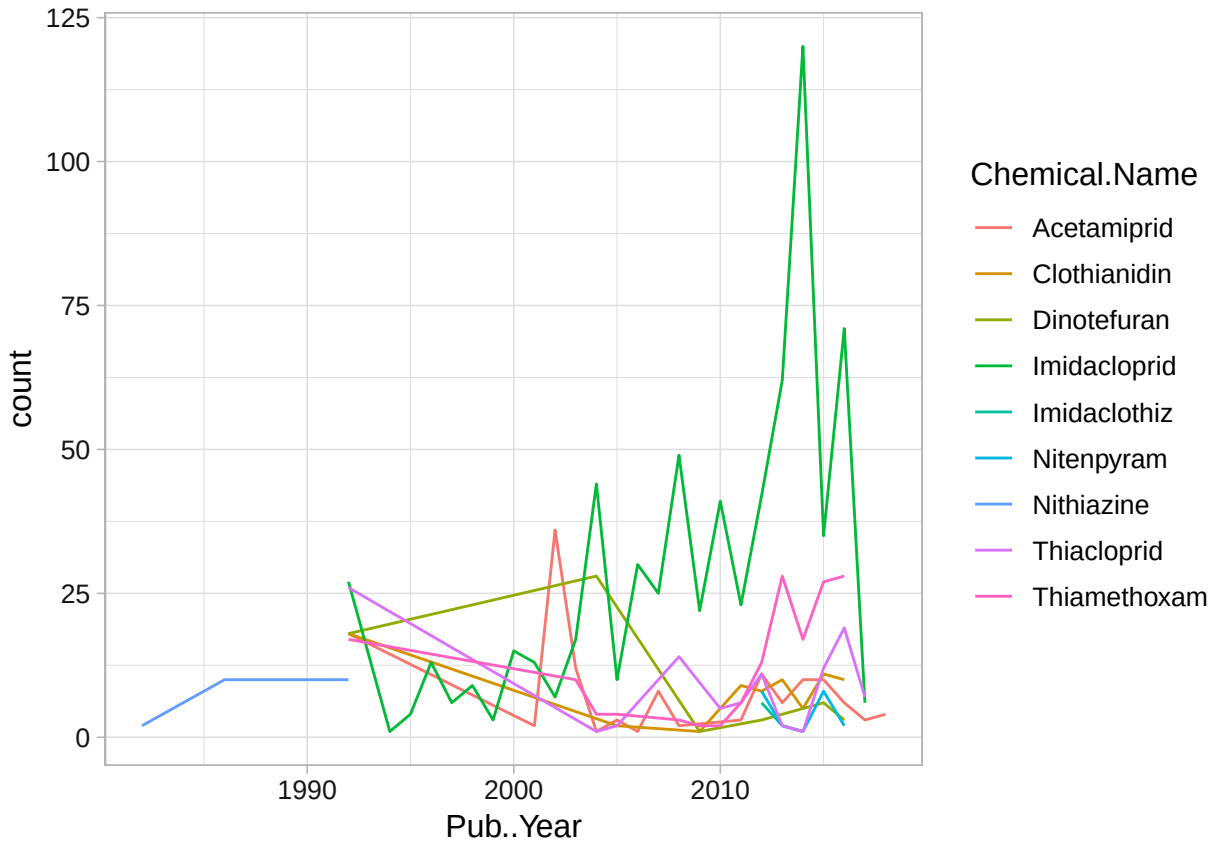
```r
# Nithiazine data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.75938, p-value = 0.0001235).

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

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

```r
# Thiacloprid data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.7669, p-value = 1.118e-11).

shapiro.test(EPA_Ecotox_Neonicotinoids$Pub..Year[EPA_Ecotox_Neonicotinoids$Chemical.Name
                                                 == "Thiamethoxam"])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  EPA_Ecotox_Neonicotinoids$Pub..Year[EPA_Ecotox_Neonicotinoids$Chemical.Name ==     "Thiametho
## W = 0.7071, p-value < 2.2e-16
```

```
# Thiamethoxam data are significantly different from
# a normal distribution (Shapiro-Wilk normality test, W = 0.7071, p-value < 2.2e-16).

#Graphic Analysis
ggplot(EPA_Ecotox_Neonicotinoids, aes(x = Pub..Year, color = Chemical.Name)) +
  geom_freqpoly(stat = "count")
```




```
# The graphic analysis gave as a result that the assumption that the
# data are normally distributed is not met. The data don't have a single peak
# around the mean and has a longer tail towards the earlier years.

#5
# We perform a Bartlett's test of the null hypothesis that the variances in each
# of the groups are the same.
bartlett.test(EPA_Ecotox_Neonicotinoids$Pub..Year ~ EPA_Ecotox_Neonicotinoids$Chemical.Name)
```

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

```
# The variance of the data are significantly different from
# each other (Bartlett test, 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: Considering that the data are not well-approximated by a normal distribution and there

is not equal variance among the publication years for each chemical, we choose the Kruskal-Wallis test, which is the non-parametric counterpart to the one-way ANOVA.

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.

```
#7
PubYear.kW_test <- kruskal.test(
  EPA_Ecotox_Neonicotinoids$Pub..Year ~ EPA_Ecotox_Neonicotinoids$Chemical.Name)
PubYear.kW_test
```

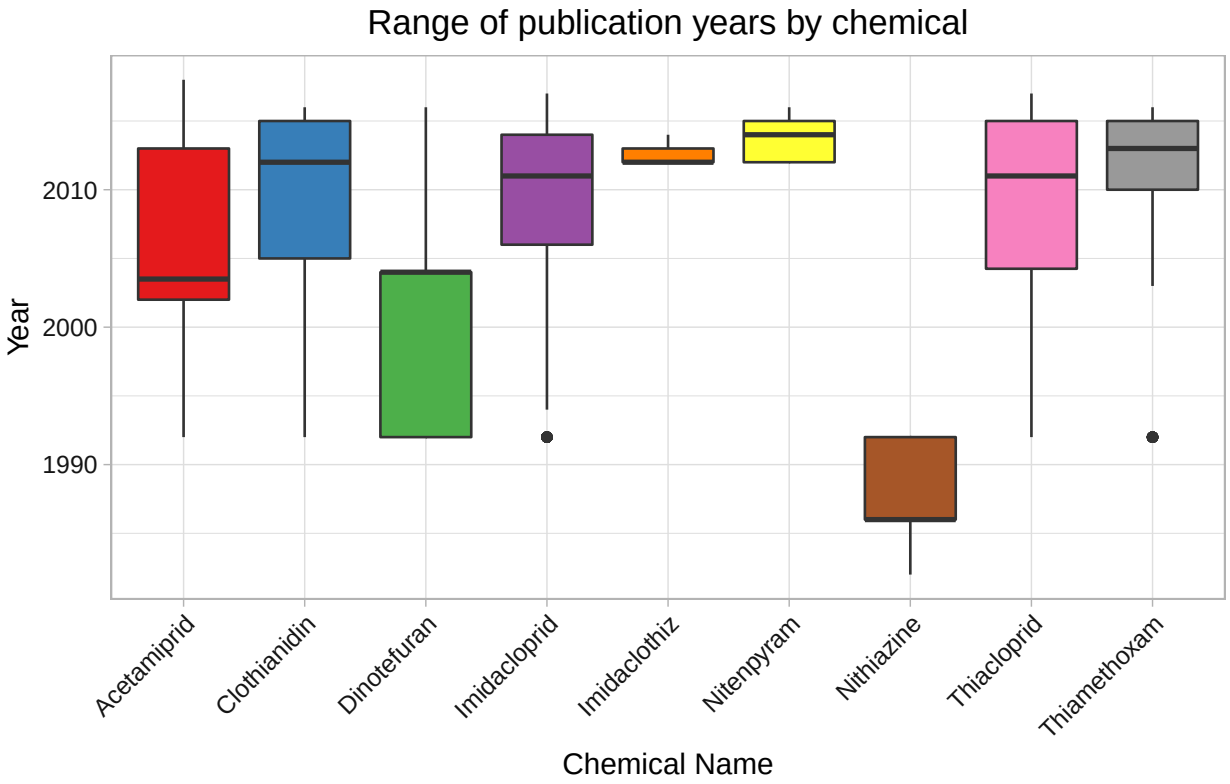
```
##
##  Kruskal-Wallis rank sum test
##
## data:  EPA_Ecotox_Neonicotinoids$Pub..Year by EPA_Ecotox_Neonicotinoids$Chemical.Name
## Kruskal-Wallis chi-squared = 134.15, df = 8, p-value < 2.2e-16
```

```
# There are significant differences between the number of studies on various
# neonicotinoid chemicals conducted in different years (Kruskal-Wallis chi-squared =
# 134.15, df = 8, p-value < 2.2e-16). Nevertheless, we don't know which groups differ
# from each other.

dunnTest(EPA_Ecotox_Neonicotinoids$Pub..Year, EPA_Ecotox_Neonicotinoids$Chemical.Name)
```

```
##                      Comparison          Z      P.unadj        P.adj
## 1   Acetamiprid - Clothianidin -3.0388079 2.375163e-03 4.037777e-02
## 2    Acetamiprid - Dinotefuran  2.1172089 3.424212e-02 4.109054e-01
## 3   Clothianidin - Dinotefuran  4.4060765 1.052598e-05 2.420975e-04
## 4   Acetamiprid - Imidacloprid -4.0204987 5.807507e-05 1.277651e-03
## 5  Clothianidin - Imidacloprid  0.5068899 6.122321e-01 1.000000e+00
## 6   Dinotefuran - Imidacloprid -5.2140290 1.847826e-07 4.989129e-06
## 7   Acetamiprid - Imidaclothiz -1.8052932 7.102881e-02 7.813169e-01
## 8  Clothianidin - Imidaclothiz -0.5166649 6.053901e-01 1.000000e+00
## 9   Dinotefuran - Imidaclothiz -2.6586494 7.845456e-03 1.176818e-01
## 10 Imidacloprid - Imidaclothiz -0.7284284 4.663514e-01 1.000000e+00
## 11    Acetamiprid - Nitenpyram -4.5018639 6.736012e-06 1.616643e-04
## 12   Clothianidin - Nitenpyram -2.4936264 1.264456e-02 1.770238e-01
## 13    Dinotefuran - Nitenpyram -5.4527796 4.958852e-08 1.388479e-06
## 14   Imidacloprid - Nitenpyram -3.0634837 2.187761e-03 3.937970e-02
## 15   Imidaclothiz - Nitenpyram -1.0897204 2.758363e-01 1.000000e+00
## 16    Acetamiprid - Nithiazine  5.6425299 1.675694e-08 4.859513e-07
## 17   Clothianidin - Nithiazine  7.1473251 8.848514e-13 2.831524e-11
## 18    Dinotefuran - Nithiazine  3.8693508 1.091255e-04 2.291636e-03
## 19   Imidacloprid - Nithiazine  7.7286349 1.087060e-14 3.804708e-13
## 20   Imidaclothiz - Nithiazine  4.8473136 1.251445e-06 3.253758e-05
## 21     Nitenpyram - Nithiazine  7.7099812 1.258363e-14 4.278434e-13
## 22   Acetamiprid - Thiacloprid -3.2225618 1.270497e-03 2.413945e-02
## 23  Clothianidin - Thiacloprid  0.1414916 8.874816e-01 8.874816e-01
## 24   Dinotefuran - Thiacloprid -4.6025295 4.173904e-06 1.043476e-04
## 25  Imidacloprid - Thiacloprid -0.3888712 6.973714e-01 1.000000e+00
## 26  Imidaclothiz - Thiacloprid  0.5870686 5.571576e-01 1.000000e+00
## 27    Nitenpyram - Thiacloprid  2.6709745 7.563140e-03 1.210102e-01
## 28    Nithiazine - Thiacloprid -7.3166886 2.541647e-13 8.387437e-12
## 29  Acetamiprid - Thiamethoxam -5.8898861 3.864618e-09 1.159385e-07
## 30 Clothianidin - Thiamethoxam -1.7587256 7.862413e-02 7.862413e-01
```

```
## 31  Dinotefuran - Thiamethoxam -6.6762123 2.451967e-11 7.601098e-10
## 32 Imidacloprid - Thiamethoxam -3.5327039 4.113329e-04 8.226657e-03
## 33 Imidaclothiz - Thiamethoxam -0.1886278 8.503846e-01 1.000000e+00
## 34   Nitenpyram - Thiamethoxam  1.5927766 1.112103e-01 1.000000e+00
## 35   Nithiazine - Thiamethoxam -8.7224129 2.723352e-18 9.804067e-17
## 36  Thiacloprid - Thiamethoxam -2.1461156 3.186376e-02 4.142288e-01
```

```
# Groups that differ
# Acetamiprid - Clothianidin (Dunn Test, Z=-3.0388079, p<0.05)
# Clothianidin - Dinotefuran  (Dunn Test, Z=4.4060765, p<0.05)
# Acetamiprid - Imidacloprid (Dunn Test, Z=-4.0204987, p<0.05)
# Dinotefuran - Imidacloprid (Dunn Test, Z=-5.2140290, p<0.05)
# Acetamiprid - Nitenpyram (Dunn Test, Z=-4.5018639, p<0.05)
# Dinotefuran - Nitenpyram (Dunn Test, Z=-5.4527796, p<0.05)
# Imidacloprid - Nitenpyram (Dunn Test, Z=-3.0634837, p<0.05)
# Acetamiprid - Nithiazine (Dunn Test, Z=5.6425299, p<0.05)
# Clothianidin - Nithiazine (Dunn Test, Z=7.1473251, p<0.05)
# Dinotefuran - Nithiazine (Dunn Test, Z=3.8693508, p<0.05)
# Imidacloprid - Nithiazine (Dunn Test, Z=7.7286349, p<0.05)
# Imidaclothiz - Nithiazine (Dunn Test, Z=4.8473136, p<0.05)
# Nitenpyram - Nithiazine (Dunn Test, Z=7.7099812, p<0.05)
# Acetamiprid - Thiacloprid (Dunn Test, Z=-3.2225618, p<0.05)
# Dinotefuran - Thiacloprid (Dunn Test, Z=-4.6025295, p<0.05)
# Nithiazine - Thiacloprid (Dunn Test, Z=-7.3166886, p<0.05)
# Acetamiprid - Thiamethoxam (Dunn Test, Z=-5.8898861, p<0.05)
# Dinotefuran - Thiamethoxam (Dunn Test, Z=-6.6762123, p<0.05)
# Imidacloprid - Thiamethoxam (Dunn Test, Z=-3.5327039, p<0.05)
# Nithiazine - Thiamethoxam (Dunn Test, Z=-8.7224129, p<0.05)

#8
#Plot
EPA_Ecotox_Neon_Plot <- ggplot(EPA_Ecotox_Neonicotinoids,
                               aes(x = Chemical.Name, y = Pub..Year,
                                   fill = Chemical.Name)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_brewer(palette = "Set1") +
  guides(fill=FALSE) +
  xlab(expression("Chemical Name")) +
  ylab(expression("Year")) +
  ggtitle("Range of publication years by chemical")
print(EPA_Ecotox_Neon_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 are significant differences between the number of studies on various neonicotinoid chemicals conducted in different years (Kruskal-Wallis chi-squared = 134.15, df = 8, p-value < 2.2e-16). the number of studies on various neonicotinoid chemicals conducted in different years that differ from each other are:

Acetamiprid - Clothianidin (Dunn Test, Z=-3.0388079, p<0.05)
Clothianidin - Dinotefuran (Dunn Test, Z=4.4060765, p<0.05)
Acetamiprid - Imidacloprid (Dunn Test, Z=-4.0204987, p<0.05)
Dinotefuran - Imidacloprid (Dunn Test, Z=-5.2140290, p<0.05)
Acetamiprid - Nitenpyram (Dunn Test, Z=-4.5018639, p<0.05)
Dinotefuran - Nitenpyram (Dunn Test, Z=-5.4527796, p<0.05)
Imidacloprid - Nitenpyram (Dunn Test, Z=-3.0634837, p<0.05)
Acetamiprid - Nithiazine (Dunn Test, Z=5.6425299, p<0.05)
Clothianidin - Nithiazine (Dunn Test, Z=7.1473251, p<0.05)
Dinotefuran - Nithiazine (Dunn Test, Z=3.8693508, p<0.05)
Imidacloprid - Nithiazine (Dunn Test, Z=7.7286349, p<0.05)
Imidaclothiz - Nithiazine (Dunn Test, Z=4.8473136, p<0.05)
Nitenpyram - Nithiazine (Dunn Test, Z=7.7099812, p<0.05)
Acetamiprid - Thiacloprid (Dunn Test, Z=-3.2225618, p<0.05)
Dinotefuran - Thiacloprid (Dunn Test, Z=-4.6025295, p<0.05)
Nithiazine - Thiacloprid (Dunn Test, Z=-7.3166886, p<0.05)
Acetamiprid - Thiamethoxam (Dunn Test, Z=-5.8898861, p<0.05)
Dinotefuran - Thiamethoxam (Dunn Test, Z=-6.6762123, p<0.05)
Imidacloprid - Thiamethoxam (Dunn Test, Z=-3.5327039, p<0.05)
Nithiazine - Thiamethoxam (Dunn Test, Z=-8.7224129, p<0.05)

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

NTL_TER_raw_chemistryphysics$sampledate <-
  format(NTL_TER_raw_chemistryphysics$sampledate, "%m")

NTL_LTER_test <- NTL_TER_raw_chemistryphysics %>%
  filter(sampledate == "07") %>%
  select(lakename:daynum, depth, temperature_C) %>%
  na.omit()

#12
NTL_LTER_AIC <- lm(data = NTL_LTER_test, temperature_C ~ year4 + daynum + depth)
step(NTL_LTER_AIC)
```

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

```
# the lower AIC value the better. In this case the better model includes all the explanatory variables.

NTL_LTER_Model <- lm(data = NTL_LTER_test, temperature_C ~ year4 + daynum + depth)
summary(NTL_LTER_Model)
```

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

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 model explain 74.1% of the observed variance. The final linear equation to predict temprature from the explanatory variables is:

$$y = -8.57556 + 0.01135 * year + 0.0398 * day.number - 1.94644 * depth + \epsilon$$

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

```
#14
NTL_LTER_Model2 <- lm(data = NTL_LTER_test, temperature_C ~ depth*lakename)
summary(NTL_LTER_Model2)
```

```
##
## Call:
## lm(formula = temperature_C ~ depth * lakename, data = NTL_LTER_test)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.6470 -2.9129 -0.2949  2.7469 16.3606
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       22.8748     0.5660  40.412  < 2e-16 ***
## depth                             -2.5543     0.2331 -10.956  < 2e-16 ***
## lakenameCrampton Lake              2.2881     0.6634   3.449 0.000565 ***
## lakenameEast Long Lake            -4.3176     0.6002  -7.194 6.76e-13 ***
## lakenameHummingbird Lake          -2.3418     0.8246  -2.840 0.004523 **
## lakenamePaul Lake                  0.7115     0.5786   1.230 0.218863
## lakenamePeter Lake                 0.3884     0.5774   0.673 0.501146
## lakenameTuesday Lake              -2.8656     0.5864  -4.887 1.04e-06 ***
## lakenameWard Lake                  2.4887     0.8302   2.998 0.002728 **
## lakenameWest Long Lake            -2.3819     0.5983  -3.981 6.91e-05 ***
## depth:lakenameCrampton Lake        0.7781     0.2388   3.258 0.001125 **
## depth:lakenameEast Long Lake       0.9189     0.2354   3.903 9.56e-05 ***
## depth:lakenameHummingbird Lake    -0.6303     0.2856  -2.207 0.027334 *
## depth:lakenamePaul Lake            0.3716     0.2342   1.587 0.112592
## depth:lakenamePeter Lake           0.5511     0.2339   2.356 0.018500 *
```

```
## depth:lakenameTuesday Lake       0.6472     0.2347   2.758 0.005826 **
## depth:lakenameWard Lake         -0.7207     0.2797  -2.577 0.009991 **
## depth:lakenameWest Long Lake     0.7892     0.2353   3.354 0.000800 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.476 on 9710 degrees of freedom
## Multiple R-squared:  0.7857, Adjusted R-squared:  0.7853
## F-statistic:  2094 on 17 and 9710 DF,  p-value: < 2.2e-16
```

```
summary.aov(NTL_LTER_Model2)
```

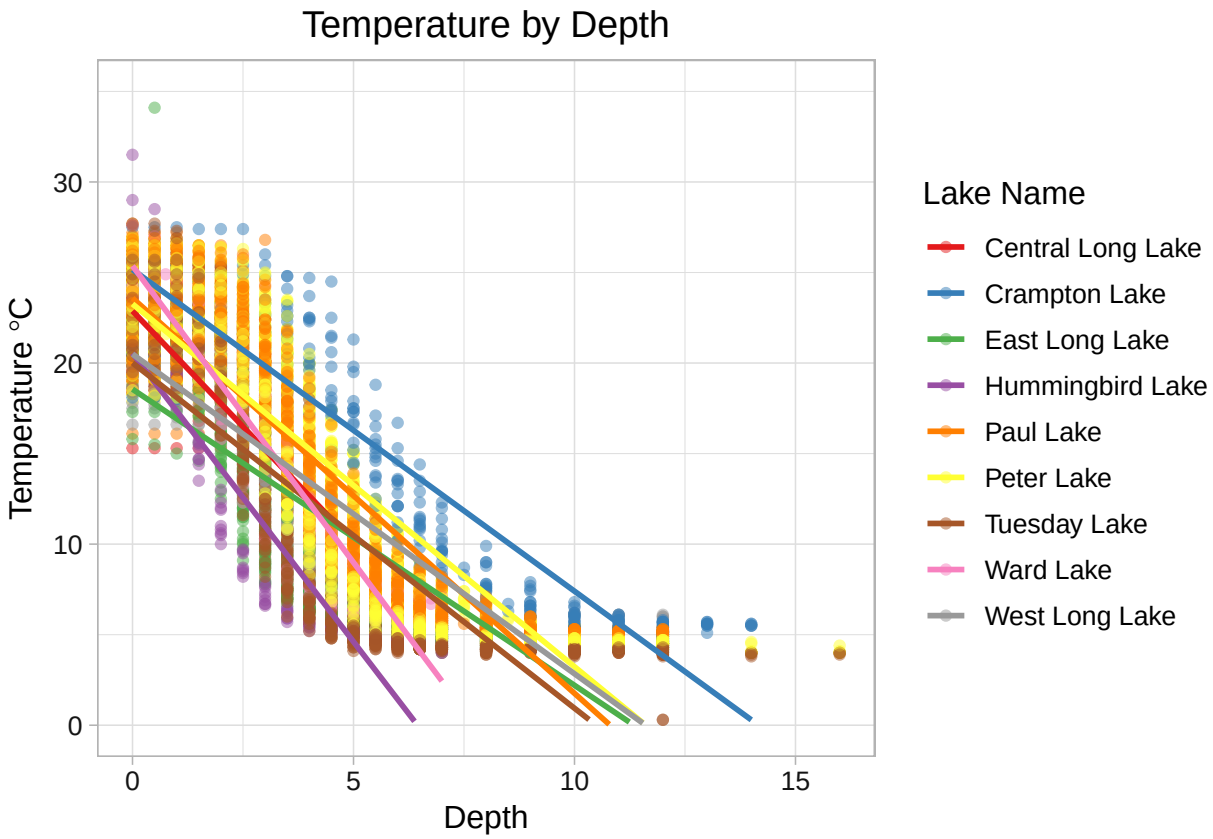
```
##                   Df Sum Sq Mean Sq  F value Pr(>F)
## depth              1 404426  404426 33468.78 <2e-16 ***
## lakename           8  20944    2618   216.66 <2e-16 ***
## depth:lakename     8   4753     594    49.16 <2e-16 ***
## Residuals       9710 117333      12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

    ANSWER: There is a significant interaction between depth and lake. The model explains 78.5% of the total variance (ANCOVA, F-statistic: 2094 on 17 and 9710 DF, p-value: $<$ 2.2e-16).

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
temperaturebydepth <- ggplot(NTL_LTER_test, aes(x = depth, y = temperature_C, color = lakename)) +
  geom_point(alpha = 0.5) +
  geom_smooth(aes(x = depth, y = temperature_C, color = lakename),method = 'lm', se = FALSE) +
  scale_color_brewer(palette = "Set1") +
  xlab(expression("Depth")) +
  ylab(expression(paste("Temperature ",degree,"C"))) +
  ggtitle("Temperature by Depth") +
  labs(color = 'Lake Name') +
  ylim(0, 35)

print(temperaturebydepth)
```


Temperature by Depth
Temperature °C
30
20
10
0
0
5
10
15
Depth
Lake Name
Central Long Lake
Crampton Lake
East Long Lake
Hummingbird Lake
Paul Lake
Peter Lake
Tuesday Lake
Ward Lake
West Long Lake