Assignment 6: GLMs week 1 (t-test and ANOVA)

Xiangtian Wang

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on t-tests and ANOVAs.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Work through the steps, **creating code and output** that fulfill each instruction.
3. Be sure to **answer the questions** in this assignment document.
4. When you have completed the assignment, **Knit** the text and code into a single PDF file.
5. After Knitting, submit the completed exercise (PDF file) to the dropbox in Sakai. Add your last name into the file name (e.g., "Salk_A06_GLMs_Week1.Rmd") prior to submission.

The completed exercise is due on Tuesday, February 18 at 1:00 pm.

Set up your session

1. Check your working directory, load the `tidyverse`, `cowplot`, and `agricolae` packages, and import the NTL-LTER_Lake_Nutrients_PeterPaul_Processed.csv dataset.

2. Change the date column to a date format. Call up `head` of this column to verify.

```
#1
library(tidyverse)
library(cowplot)
library(agricolae)
NL.Nutrients.PPLake <- read.csv("../Data/Processed/NTL-LTER_Lake_Chemistry_Nutrients_PeterPaul_Processe
# Set theme
mytheme <- theme_classic(base_size = 12) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "right")
theme_set(mytheme)
#2
NL.Nutrients.PPLake$sampledate <- as.Date(NL.Nutrients.PPLake$sampledate,"%Y-%m-%d")
head(NL.Nutrients.PPLake)
```

```
##    lakename year4 daynum month sampledate depth temperature_C dissolvedOxygen
## 1 Paul Lake  1984    148     5 1984-05-27  0.00          14.5             9.5
## 2 Paul Lake  1984    148     5 1984-05-27  0.25            NA              NA
## 3 Paul Lake  1984    148     5 1984-05-27  0.50            NA              NA
## 4 Paul Lake  1984    148     5 1984-05-27  0.75            NA              NA
## 5 Paul Lake  1984    148     5 1984-05-27  1.00          14.5             8.8
## 6 Paul Lake  1984    148     5 1984-05-27  1.50            NA              NA
##   irradianceWater irradianceDeck tn_ug tp_ug nh34 no23 po4
## 1            1750           1620    NA    NA   NA   NA  NA
## 2            1550           1620    NA    NA   NA   NA  NA
## 3            1150           1620    NA    NA   NA   NA  NA
## 4             975           1620    NA    NA   NA   NA  NA
## 5             870           1620    NA    NA   NA   NA  NA
## 6             610           1620    NA    NA   NA   NA  NA
```

Wrangle your data

3. Wrangle your dataset so that it contains only surface depths and only the years 1993-1996, inclusive. Set month as a factor.

```
my.NL.Nutrients.PPLake<- NL.Nutrients.PPLake %>%
  filter(year4 > 1992 & year4 < 1997) %>%
  filter(depth == 0)
my.NL.Nutrients.PPLake$month<- as.factor(my.NL.Nutrients.PPLake$month)
```

Analysis

Peter Lake was manipulated with additions of nitrogen and phosphorus over the years 1993-1996 in an effort to assess the impacts of eutrophication in lakes. You are tasked with finding out if nutrients are significantly higher in Peter Lake than Paul Lake, and if these potential differences in nutrients vary seasonally (use month as a factor to represent seasonality). Run two separate tests for TN and TP.

4. Which application of the GLM will you use (t-test, one-way ANOVA, two-way ANOVA with main effects, or two-way ANOVA with interaction effects)? Justify your choice.

   Answer: I choose two-way ANOVA with interaction effects because we need to consider the lakename, month and the interaction of them. The results of the test are: there are significant differences between two lakes (Tn and Tp, p-value < 0.001), and Tp with the interaction of lakename and month(p-value < 0.05). There are no significant differences between the month and the interaction of month and lake by total nitrigen.

5. Run your test for TN. Include examination of groupings and consider interaction effects, if relevant.

6. Run your test for TP. Include examination of groupings and consider interaction effects, if relevant.

```
#5
TN.aov <- aov(tn_ug ~ lakename * month, my.NL.Nutrients.PPLake)
summary(TN.aov)
```

```
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## lakename         1 2468595 2468595  36.414 2.91e-08 ***
## month            4  459542  114885   1.695    0.157
## lakename:month   4  288272   72068   1.063    0.379
## Residuals       97 6575834   67792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 27 observations deleted due to missingness
```

```
#6
TP.aov <- aov(tp_ug ~ lakename * month, my.NL.Nutrients.PPLake)
summary(TP.aov)
```

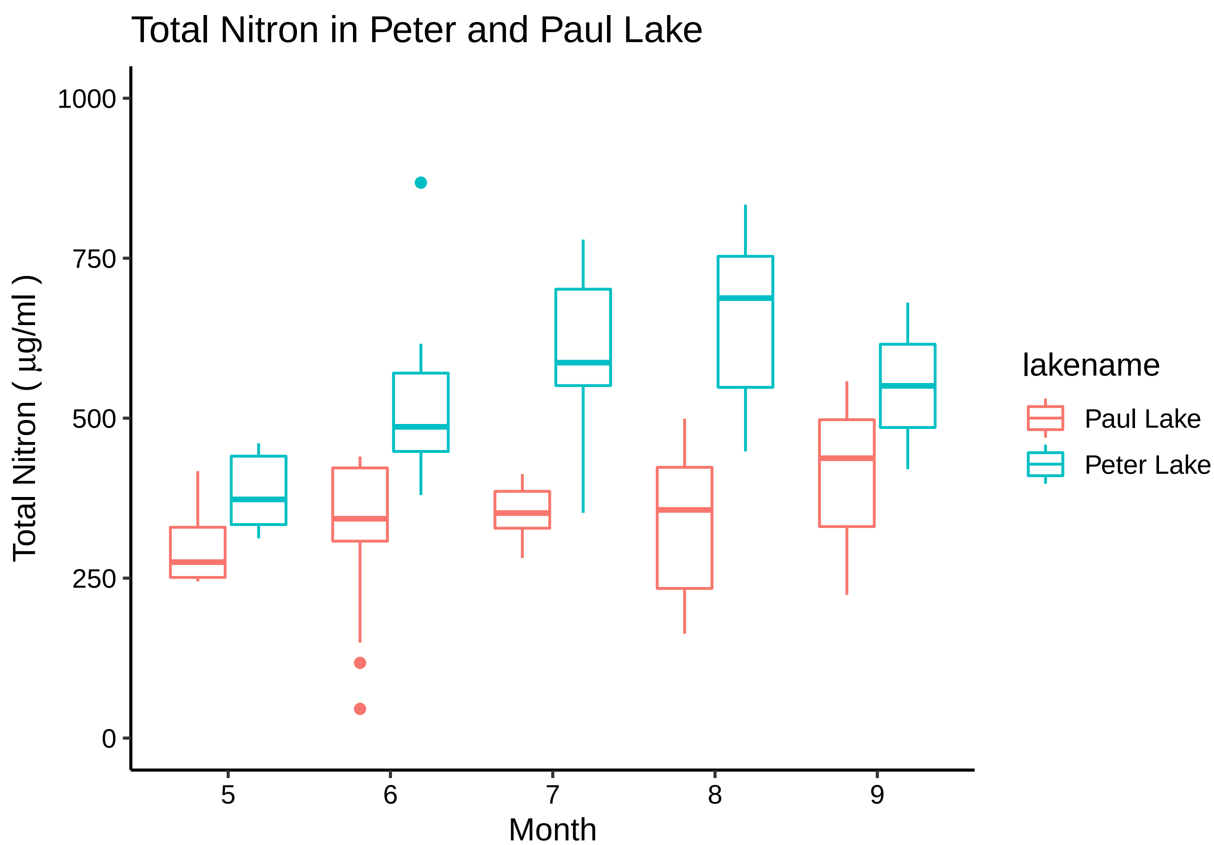
```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## lakename          1  10228   10228  98.914 <2e-16 ***
## month             4    813     203   1.965 0.1043
## lakename:month    4   1014     254   2.452 0.0496 *
## Residuals       119  12305     103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 5 observations deleted due to missingness
```

7. Create two plots, with TN (plot 1) or TP (plot 2) as the response variable and month and lake as the predictor variables. Hint: you may use some of the code you used for your visualization assignment.

Assign groupings with letters, as determined from your tests. Adjust your axes, aesthetics, and color palettes in accordance with best data visualization practices.

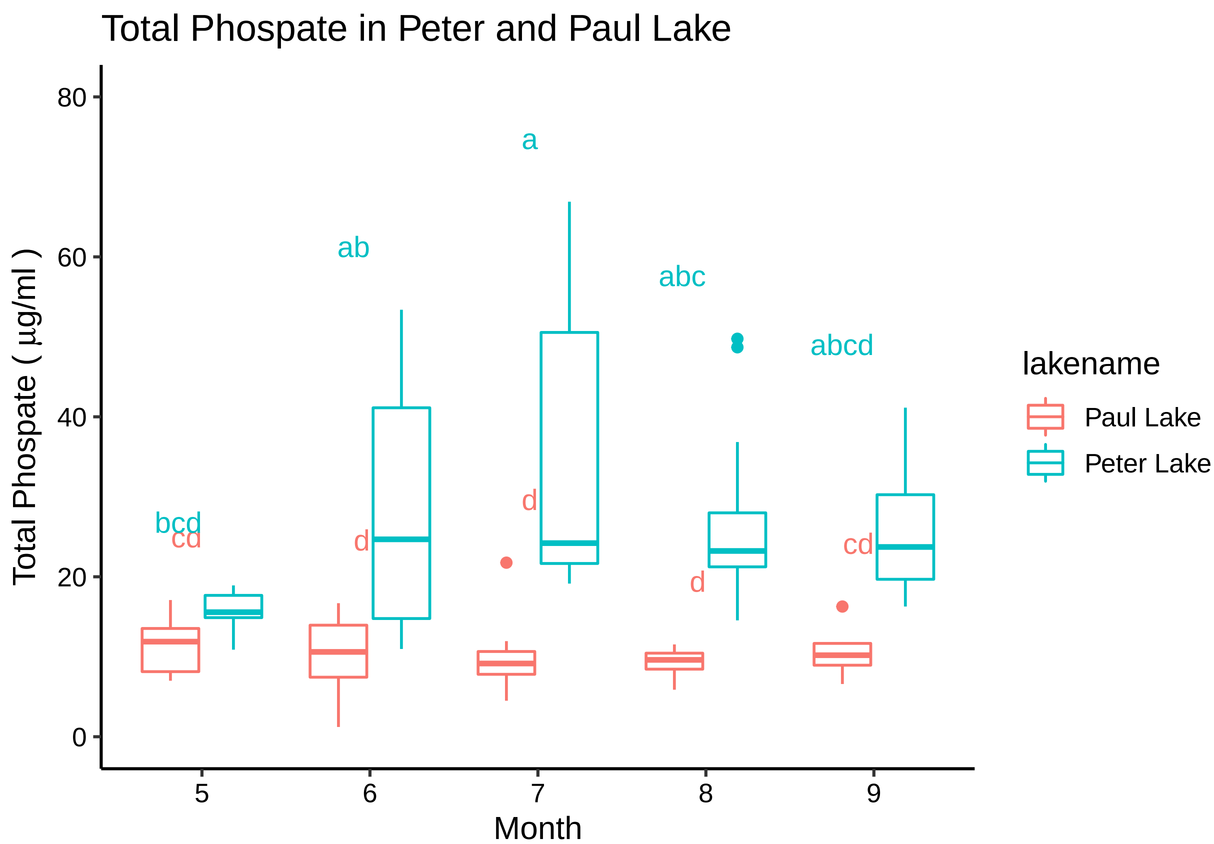
8. Combine your plots with cowplot, with a common legend at the top and the two graphs stacked vertically. Your x axes should be formatted with the same breaks, such that you can remove the title and text of the top legend and retain just the bottom legend.

```
#7
# Total Nitrigen
TN.groups <- HSD.test(TN.aov, "month", group = TRUE)
TN.plot <- ggplot(my.NL.Nutrients.PPLake, aes(month, tn_ug, color=lakename)) +
  labs( title = "Total Nitron in Peter and Paul Lake",x = "Month",
        y = expression("Total Nitron ( "*mu*"g/ml )")) +
  ylim(0,1000)+ scale_x_discrete(limits=c("5","6","7","8","9")) +
  geom_boxplot()
print(TN.plot)
```

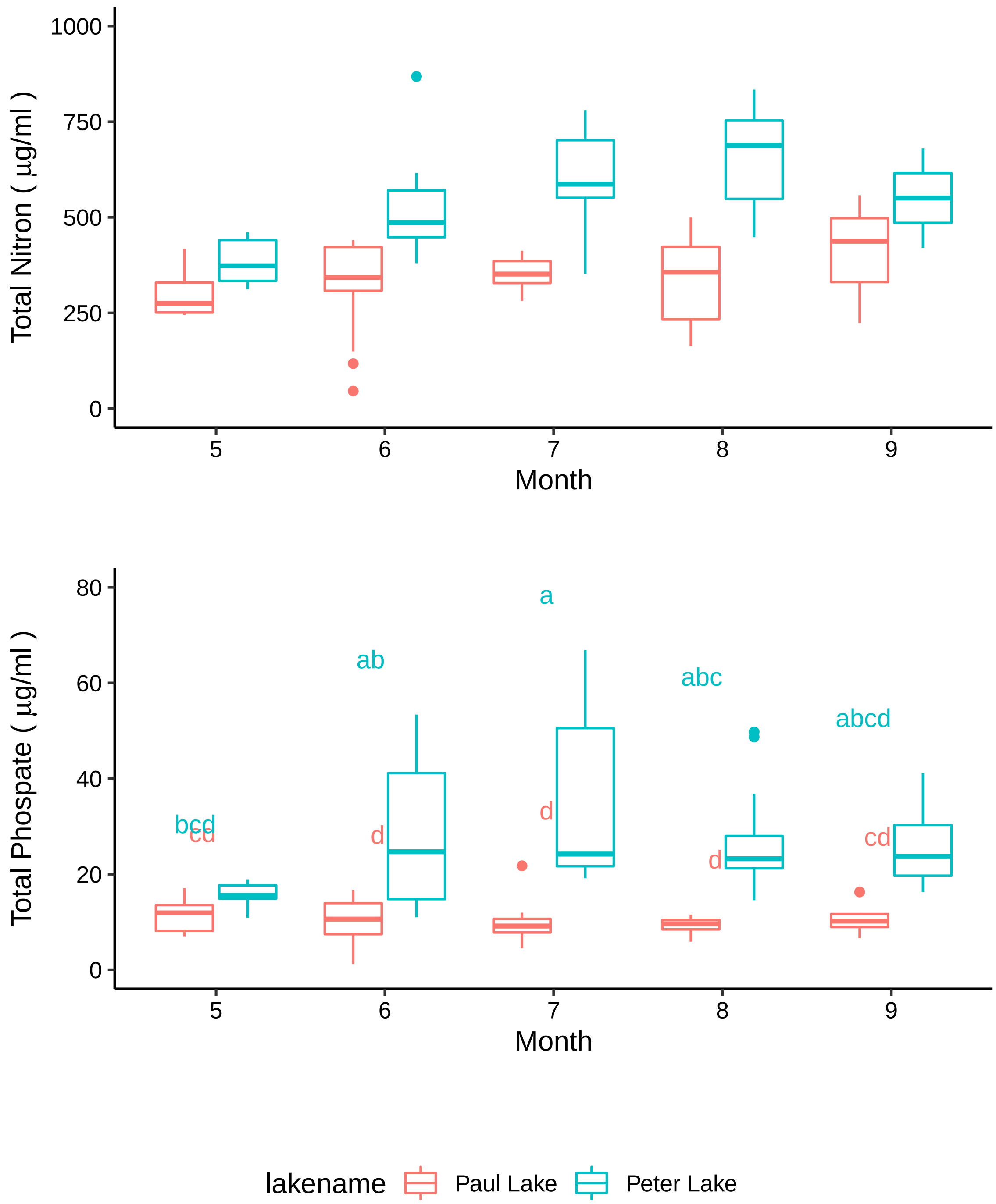



```
# Total Phospate
TP.in <- with(my.NL.Nutrients.PPLake,interaction(lakename,month))
TP.aov <- aov(tp_ug ~ TP.in, my.NL.Nutrients.PPLake)
TP.groups <- HSD.test(TP.aov,"TP.in", group = TRUE)
TP.plot <- ggplot(my.NL.Nutrients.PPLake, aes(month, tp_ug, color=lakename)) +
  labs( title = "Total Phospate in Peter and Paul Lake",x = "Month",
        y = expression("Total Phospate ( "*mu*"g/ml )")) +
  ylim(0,80)+ scale_x_discrete(limits=c("5","6","7","8","9")) +
  stat_summary(fun.y = max, geom="text", vjust= -2.5,hjust= 1,size = 4,
               label =c("cd","bcd","d","ab","d","a","d","abc","cd","abcd")) +
  geom_boxplot()
```

```r
print(TP.plot)
```




```r
#8
all <- plot_grid(TN.plot + labs(title="")+ theme(legend.position="none"),
          TP.plot + labs(title="") + theme(legend.position="none"),
          align = 'v', nrow = 2)
legend_box <- get_legend(TN.plot +
        theme(legend.position = "bottom"))
plot_grid(all, legend_box, ncol = 1, rel_heights = c(1, .2 ))
```


Total Nitron (µg/ml)
1000
750
500
250
0
5
6
7
8
9
Month
Total Phospate (µg/ml)
80
60
40
20
0
a
ab
abc
abcd
bcd
cd
d
d
d
cd
5
6
7
8
9
Month
lakename
Paul Lake
Peter Lake