Section Assignment 6

Please Read!

This week, your assignment is a series of review problems derived from past weeks. Use this week's assignment as an opportunity to cement what you have learned so far, before we jump into more complex topics next week!

Since we have already downloaded and installed the tidyverse, this week, simply load the tidyverse functions into your working library:

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

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

This line of code above will need to be run each time you start a new R session.

One of the traits of a great programmer is their ability to solve a problem they haven't seen before. One of the best ways to solve a problem you have not seen before is to see if anyone else has. Great programmers are experts of web searching. Do not be afraid to find code on an online help page and run it!

Copying from the internet is one of the foundations of learning how to program, but it only works as learning if you reflect on why the code you used works. For that reason, this week and in future weeks, you will need to annotate your code or answer a specific question about the process. Use the lines that look like this at the end of each question to input your explanation of the code:

```
note<-'your note here'
print(note)
```

You will be graded on not just your code, but your explanation. Remember, in this class, it is okay to copy code, but you still need to demonstrate independent thought.

Question 1 (0.5 point): Read in this week's data set "Week6SoilData.csv". These data represent inorganic soil nitrogen (NO3-) concentrations from 500 gardens collected in both 2018 and 2021. The nitrogen concentration was determined by combustion analysis and is presented as a proportion of total nitrogen (from 0 to 1).

```
W6Data<-read.csv("Week6SoilData.csv")
head(W6Data)
```

```
##   GardenID  SoilN2018  SoilN2021
## 1        1 0.24272954 0.12300013
## 2        2 0.12470401 0.25393789
## 3        3 0.06738397 0.24644464
## 4        4 0.40179051 0.14154592
## 5        5 0.15873990 0.13677533
## 6        6 0.06578679 0.06604907
```

Is this data set tidy? Why or why not? [1] ""

Question 2 (0.5 point): How many gardens in 2018 had more than 20% of their total nitrogen in inorganic nitrogen?

```
High_Inorganic_N<-sum(W6Data$SoilN2018>0.2)
High_Inorganic_N
```

```
## [1] 152
```

There are many ways to answer this question. Explain how your code counts how many gardens had high inorganic nitrogen in 2018 [1] ""

Question 3 (1 point): Convert the 'W6Data' table into a tidy table. Hint:`names_to = "Year",values_to = "InorganicN"`

```
GardenData<-pivot_longer(data = W6Data,cols = 2:3,names_to = "Year",values_to = "InorganicN")
str(GardenData)
```

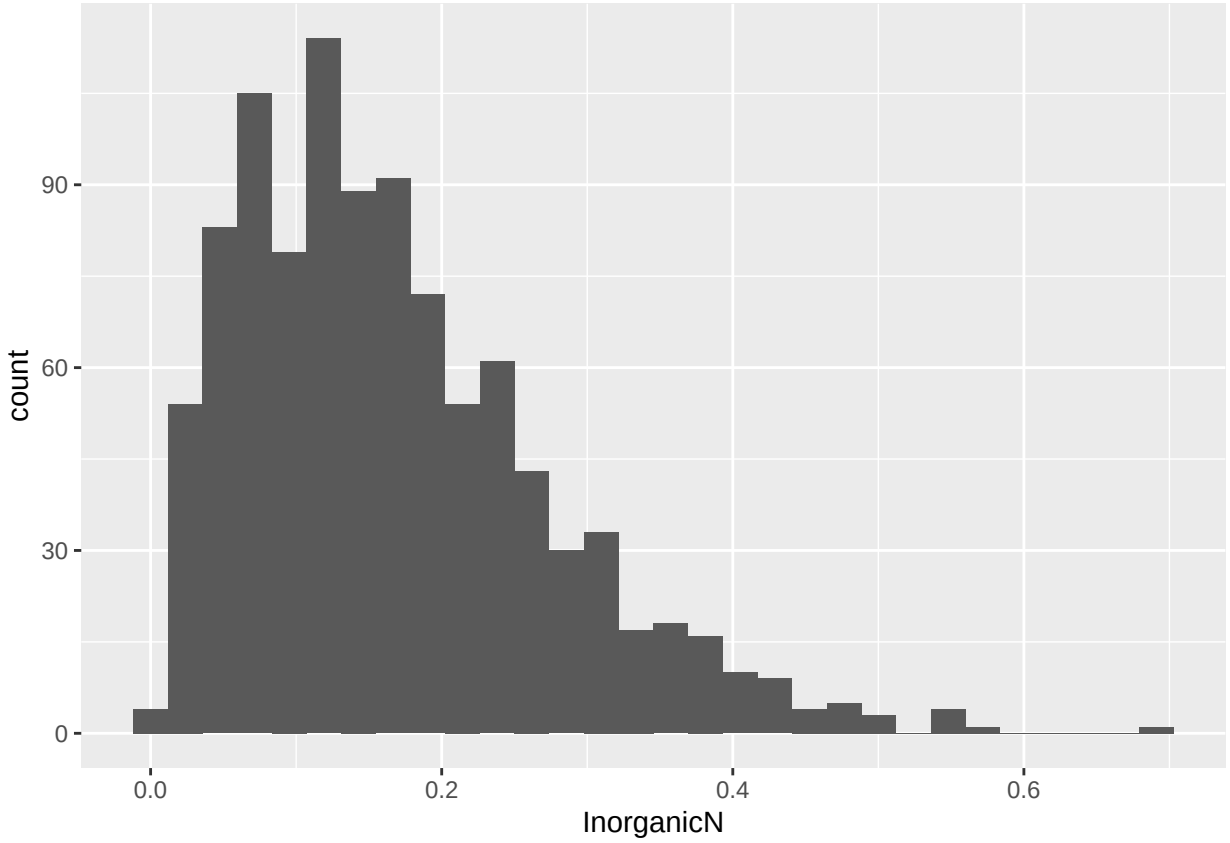
```
## tibble [1,000 x 3] (S3: tbl_df/tbl/data.frame)
##  $ GardenID  : int [1:1000] 1 1 2 2 3 3 4 4 5 5 ...
##  $ Year      : chr [1:1000] "SoilN2018" "SoilN2021" "SoilN2018" "SoilN2021" ...
##  $ InorganicN: num [1:1000] 0.2427 0.123 0.1247 0.2539 0.0674 ...
```

Is this new data set tidy? Why or why not? [1] ""

Question 4 (2 points): Is the Inorganic Nitrogen data normally distributed? Use both a histogram and the ks.test() to support your answer.

```
Histogram1<-ggplot(data = GardenData,mapping = aes(x =InorganicN ))+
  geom_histogram()
Histogram1
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ks.test(x = GardenData$InorganicN,y = "pnorm", mean(GardenData$InorganicN),sd(GardenData$InorganicN))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  GardenData$InorganicN
## D = 0.082731, p-value = 2.27e-06
## alternative hypothesis: two-sided
```

Is the Inorganic Nitrogen data normally distributed? Explain in the note below: [1] ""

Question 5 (1 point): Use the `mutate()` function to create a data column that is transformed InorganicN data. Hint: Try a square root transformation this time.

```
T_Garden_Data<-mutate(.data = GardenData,sqrtN=sqrt(InorganicN))
str(T_Garden_Data)
```

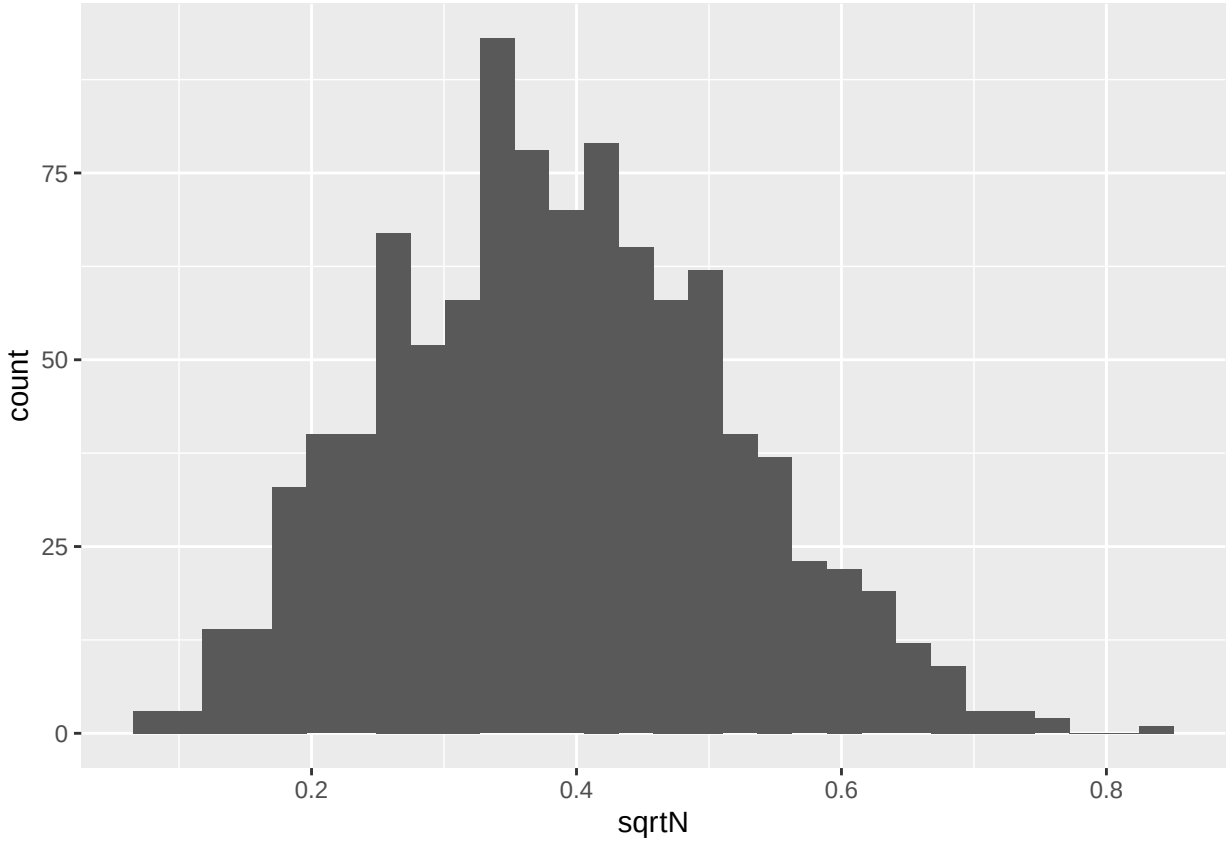
```
## tibble [1,000 x 4] (S3: tbl_df/tbl/data.frame)
##  $ GardenID  : int [1:1000] 1 1 2 2 3 3 4 4 5 5 ...
##  $ Year      : chr [1:1000] "SoilN2018" "SoilN2021" "SoilN2018" "SoilN2021" ...
##  $ InorganicN: num [1:1000] 0.2427 0.123 0.1247 0.2539 0.0674 ...
##  $ sqrtN     : num [1:1000] 0.493 0.351 0.353 0.504 0.26 ...
```

Explain what the mutate function is doing in this case. [1] ""

Question 6 (2 points): Is the transformed Inorganic Nitrogen data normally distributed? Use both a histogram and the ks.test() to support your answer.

```
Histogram2<-ggplot(data = T_Garden_Data,mapping = aes(x =sqrtN ))+
  geom_histogram()
Histogram2
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



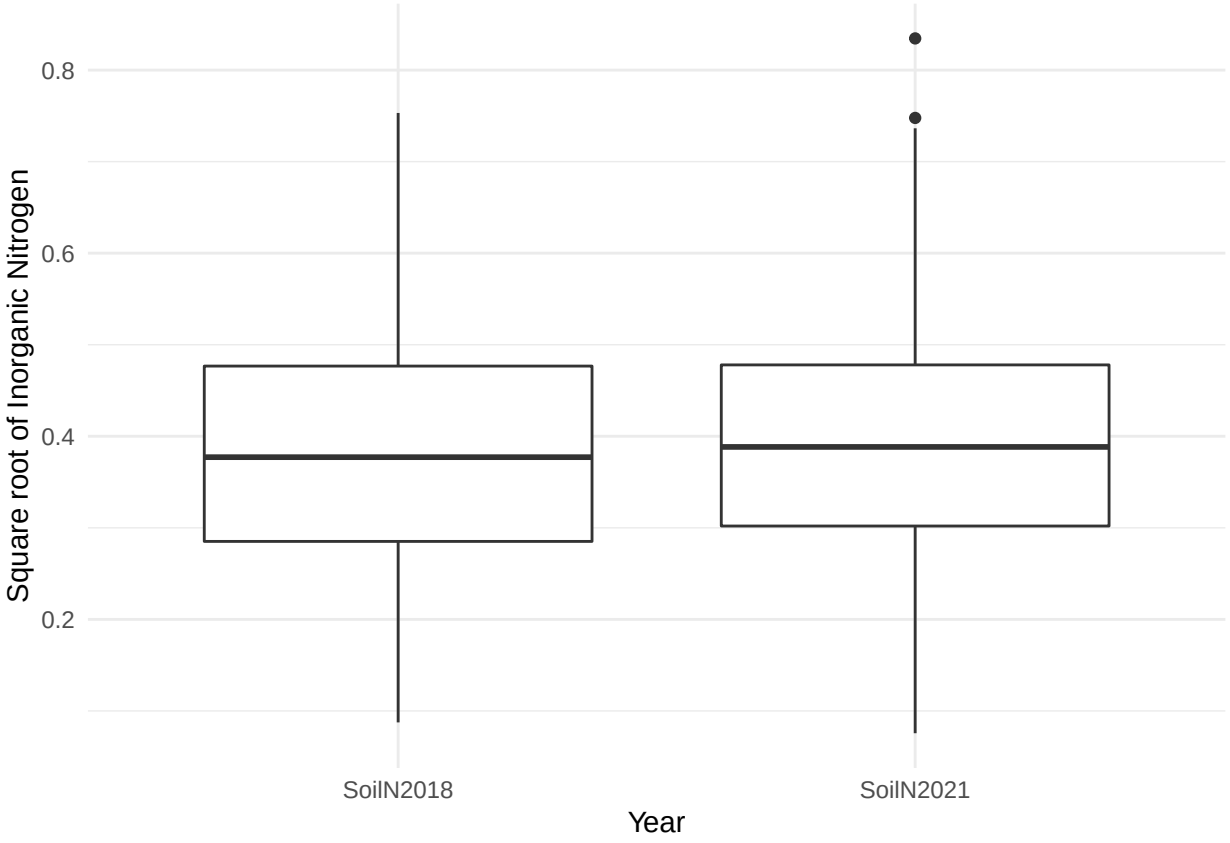

```
ks.test(x = T_Garden_Data$sqrtN,y = "pnorm", mean(T_Garden_Data$sqrtN),sd(T_Garden_Data$sqrtN))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  T_Garden_Data$sqrtN
## D = 0.02534, p-value = 0.542
## alternative hypothesis: two-sided
```

Is the transformed Inorganic Nitrogen data normally distributed? Explain in the note below: [1] ""

Question 7 (1 point): Use ggplot to create a boxplot of the two years of transformed nitrogen data. Do your best to get your graph to look like the example given in the Canvas assignment. Hint 1: `geom_boxplot` Hint 2: If your plot is not looking like the example, go back to your week 3 section assignment and look at how we changed axis labels and the theme.

```
Boxplot1<-ggplot(data = T_Garden_Data,mapping = aes(x = Year,y = sqrtN))+
  geom_boxplot()+
  theme_minimal() +
  ylab("Square root of Inorganic Nitrogen")
Boxplot1
```



Explain how each line in your ggplot code is modifying the plot. Hint: Your code should have used 4 lines to generate a plot similar to the example given on Canvas. [1] ""

Question 8 (2 points): Calculate the overall mean and 95% confidence interval for the inorganic N proportion in all gardens measured in both 2018 and 2021. Report your answers in the original units. Hint 1: The mean only means what we think the mean means when the data are normal! Hint 2: The back transformation of a square root transformation is x^2.

```
t_mean<-mean(T_Garden_Data$sqrtN)
tsd<-sd(T_Garden_Data$sqrtN)
t_upper<-qnorm(0.975)*tsd+t_mean
# If they use 1.96 instead of qnorm here that is fine.
t_lower<-qnorm(0.025)*tsd+t_mean

upperCI<-t_upper^2
meanN<-t_mean^2
lowerCI<-t_lower^2

cbind(upperCI,meanN,lowerCI)
```

```
##        upperCI     meanN   lowerCI
## [1,] 0.4104838 0.1501943 0.0180655
```

Explain how you estimated the mean and confidence intervals of the garden soil data. [1] ""

Once you are done, click the “Knit” button above(It looks like a blue ball of yarn). Save the pdf file with your name and the week number in the file name:

(for example: “Liam_Mueller_Section_6.pdf”).

Then upload this pdf file to canvas/gradescope under the Week 6 Section Assignment before the deadline.

You are now more than half way done, look how much programming you have learned in this short time! Congratulations, and keep up the good work.