

CapDAP - Small Mammal Diversity Data in the Konza Prairie

Marliese Baer

11/14/2022

##Introduction:

what is the biological motivation behind your research? What is the scientific justification (i.e., cite some papers) for your hypotheses and expected outcomes?

My overall biological motivation for this research was driven by an interest in small mammals and a desire to learn about habitat factors that could impact populations of these mammals. I also think looking at watershed as an indicator for burn policies and how they could impact habitat and therefore the species that live in those areas is interesting.

For hypothesis 1) My initial thought was that there would be a significant difference found between the number of species between the different watersheds because they have different burn policies, therefore the habitat in these watersheds will be different and support different species. After conducting some preliminary research however, I do not think there will be a significant difference. The understory wasn't remarkably changed by the fires in this study, which altered my hypothesis (Strand, E.K., Satterberg, K.L., Hudak, A.T. et al. Does burn severity affect plant community diversity and composition in mixed conifer forests of the United States Intermountain West one decade post fire?. *fire ecol* 15, 25 (2019). <https://doi.org/10.1186/s42408-019-0038-8>).

For hypothesis 2) I think there will be a significant difference in avg animal weight in different watersheds. Fires can expose food sources that were hidden or not available before, and they can support certain plant species (Fawcett, Moorman, Sharpe. Using Fire to Improve Wildlife Habitat. NC State Extension Publications (April 7 2021)).

For hypothesis 3) I think there will be a small significant difference in animal diversity in different watersheds because, as I said before, fires can expose different food sources and support the growth of different plants (Fawcett, Moorman, Sharpe. Using Fire to Improve Wildlife Habitat. NC State Extension Publications (April 7 2021)). With that being said though, different fire treatments can help maintain growth and production of various plants differently. These different ways may vary in their effect on biodiversity, however while they do have differing effects, they are all positive (Hulbert, Lloyd C. "Fire effects on tallgrass prairie." *Proceedings of the Ninth North American Prairie Conference*. GK Clambey and RH Pemble [eds.], Fargo, North Dakota. 1986.). This could mean that the difference may not be as significant as I originally thought.

##Analysis:

First lets set things up

```
rm(list = ls())
library(tidyverse)
```

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

## v ggplot2 3.3.5      v purrr   0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.0.1      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

library(here)

## here() starts at /home/mebaer20/BioStat/Capstone_Data_Analysis_Project
library(ggfortify)
library(readr)
```

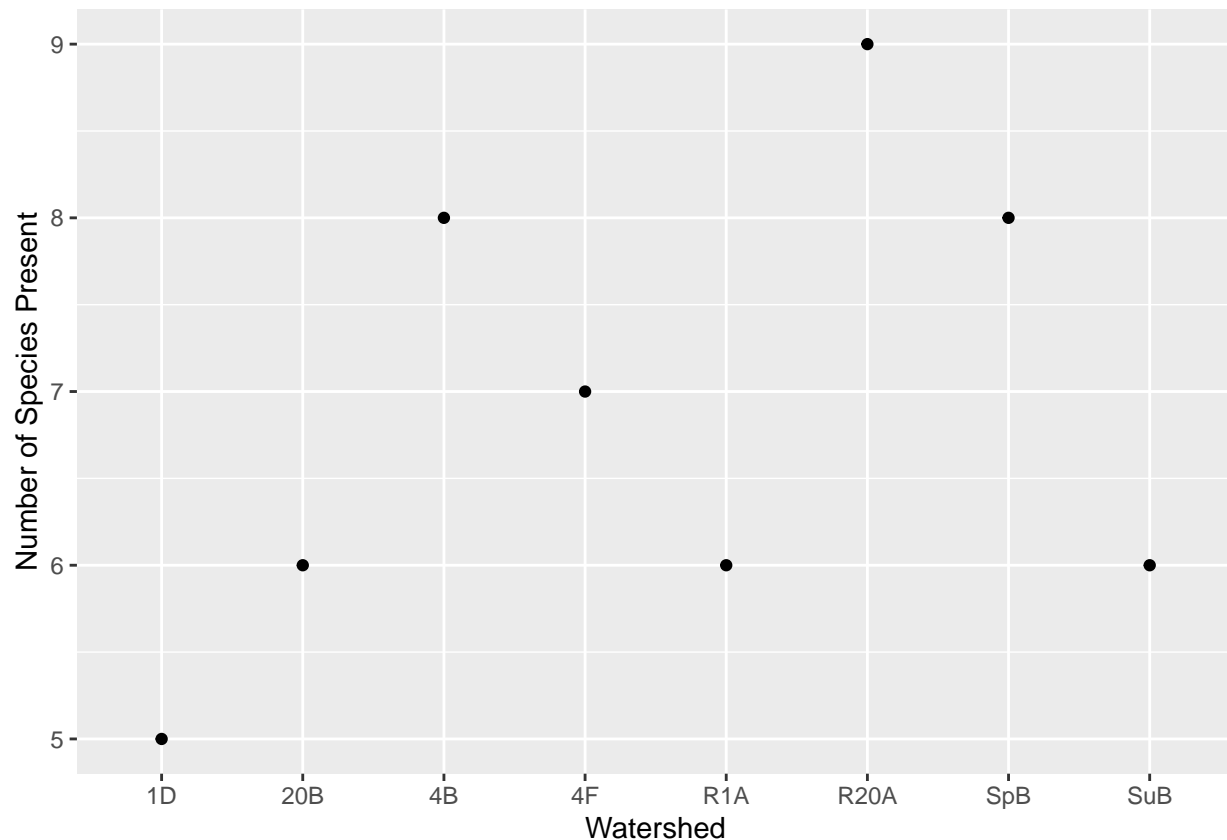
Analysis 1:

1. pull in data

```
A1 <- read.csv(here("Data", "Processed Data", "Ws_Num_Sp.csv"))
```

2. plot expected relationships

```
ggplot(A1, aes(x = Watershed, y=num_species))+
  geom_point()+
  xlab("Watershed")+
  ylab("Number of Species Present")
```



3. guess expected relationships (hypothesis)

I think there will not be a significant difference in the number of species found in different watersheds.

Null hypothesis (Ho): The number of species found in each watershed is equal (no significant difference).

Alternate hypothesis (Ha): The number of species found in each watershed is not equal (significant difference).

4. decide on statistical model and test and explain why

explanatory = Watershed = categorical nominal response = num_species = discrete quantitative

This will be a chi squared test because Watershed (the explanatory or predictor variable) is a categorical nominal variable and num_species (the response variable) is a quantitative discrete variable. (Based on the chart we were provided with.)

5. create model

5.a) make a matrix

```
A1.mat<-xtabs(num_species~Watershed,data=A1)
A1.mat
```

```
## Watershed
##   1D  20B   4B   4F  R1A R20A  SpB  SuB
##    5    6    8    7    6    9    8    6
```

5.b) make the model

```
A1.chi<-chisq.test(A1.mat)
```

6. check model assumptions

The chi-square test assumptions cannot be checked using autoplot. The best way to check the assumptions is just to look at the variables. The explanatory variable is categorical nominal and the response variable is count data. Because it is using count data, it can therefore be a chi-squared test.

7. run stat model

```
names(A1.chi)
```

```
## [1] "statistic" "parameter" "p.value"   "method"     "data.name"  "observed"
## [7] "expected"  "residuals" "stdres"
```

```
A1.chi$expected
```

```
##   1D  20B   4B   4F  R1A R20A  SpB  SuB
## 6.875 6.875 6.875 6.875 6.875 6.875 6.875 6.875
```

```
A1.chi
```

```
##
## Chi-squared test for given probabilities
##
## data:  A1.mat
## X-squared = 1.8727, df = 7, p-value = 0.9665
```

8. interpret model results (were expected relationships supported?)

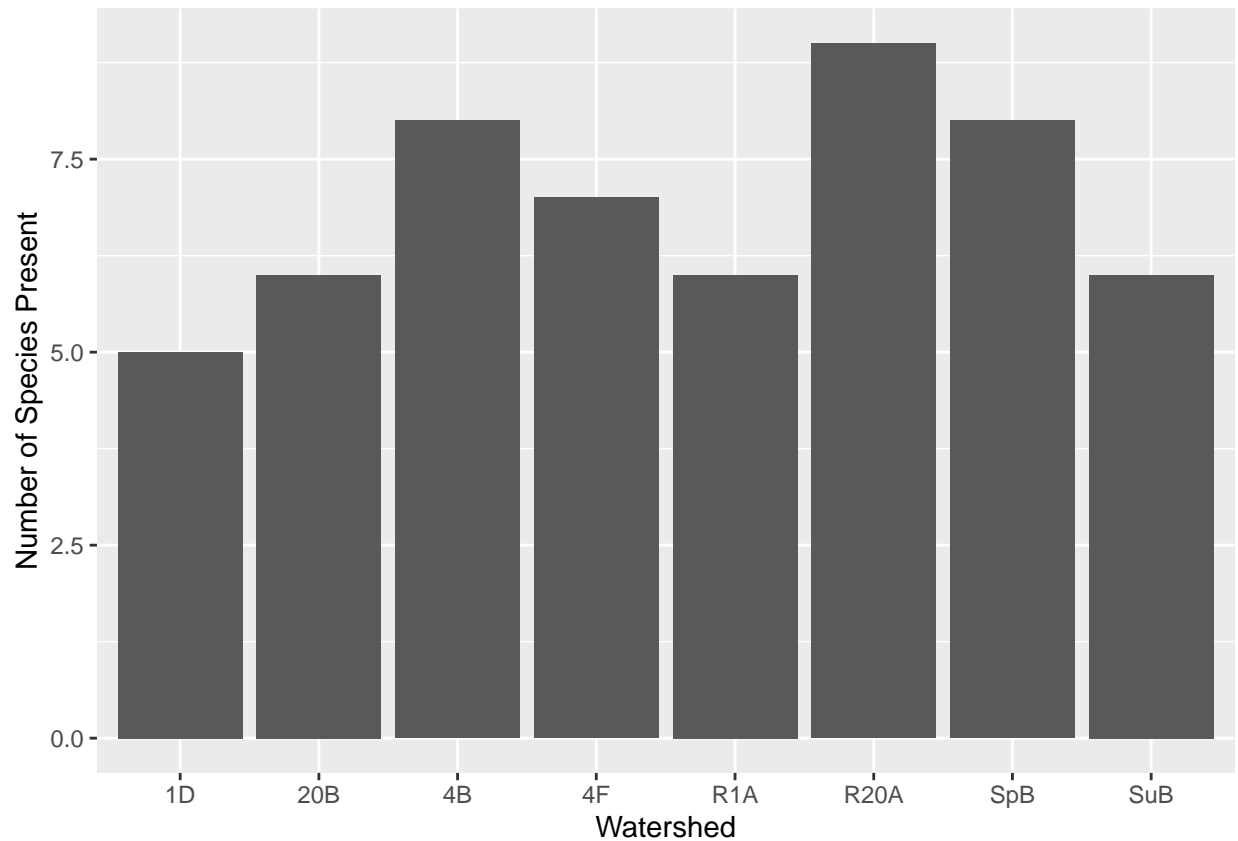
Null hypothesis (Ho): The number of species found in each watershed is equal. Alternate hypothesis (Ha): The number of species found in each watershed is not equal.

p-value=0.9665 0.9665>0.05 fail to reject the null

There is not a significant difference in the number of species found in each watershed based on a p-value of 0.9665.

9. replot data to show key relationships

```
ggplot(A1, aes(x=Watershed,y=num_species),color=blues9,fill=blues9)+
  geom_col()+
  xlab("Watershed")+
  ylab("Number of Species Present")
```



Analysis 2:

1. pull in data for all three species

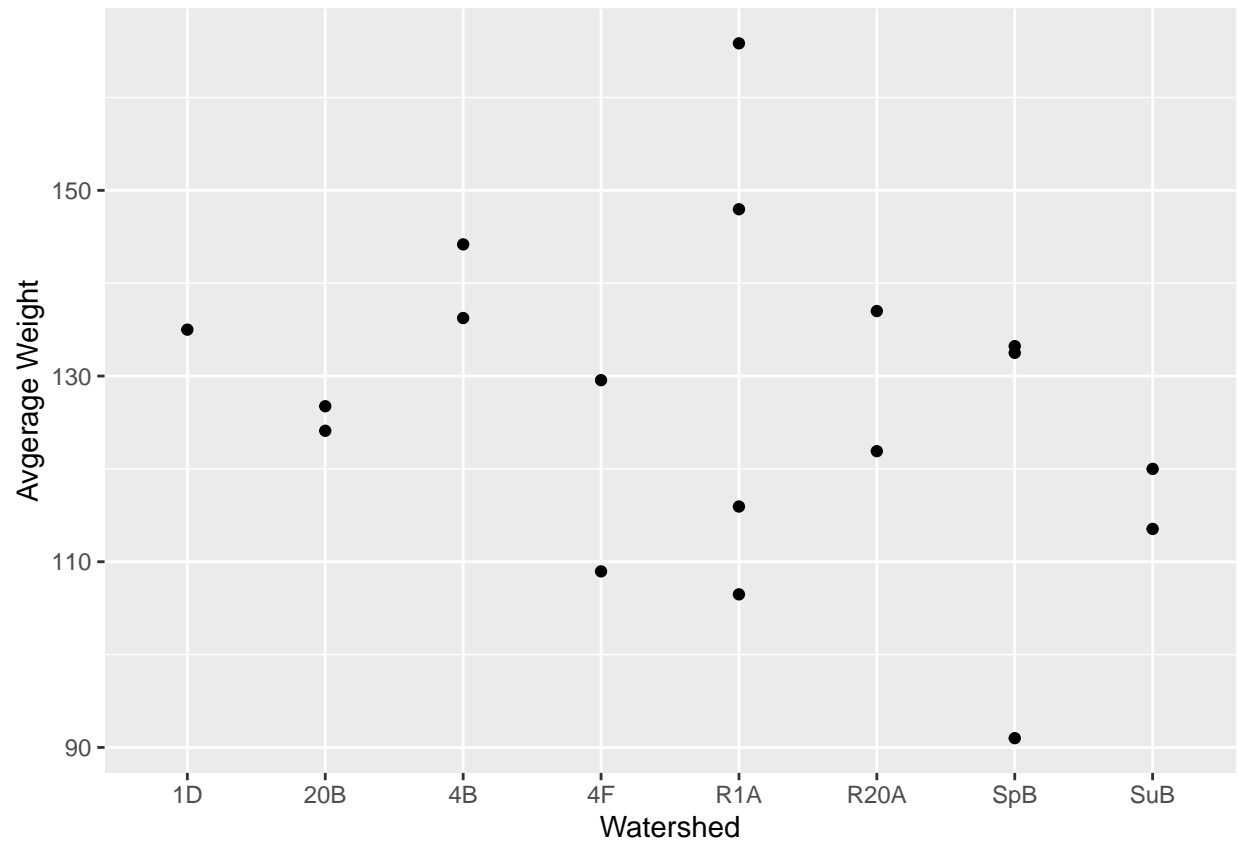
```
A2S <- read.csv(here("Data", "Processed Data", "SIHI_Ws_AvgW.csv"))
A2M <- read.csv(here("Data", "Processed Data", "MIOC_Ws_AvgW.csv"))
A2P <- read.csv(here("Data", "Processed Data", "PMMA_Ws_AvgW.csv"))
```

1.b) make Watershed a factor variable

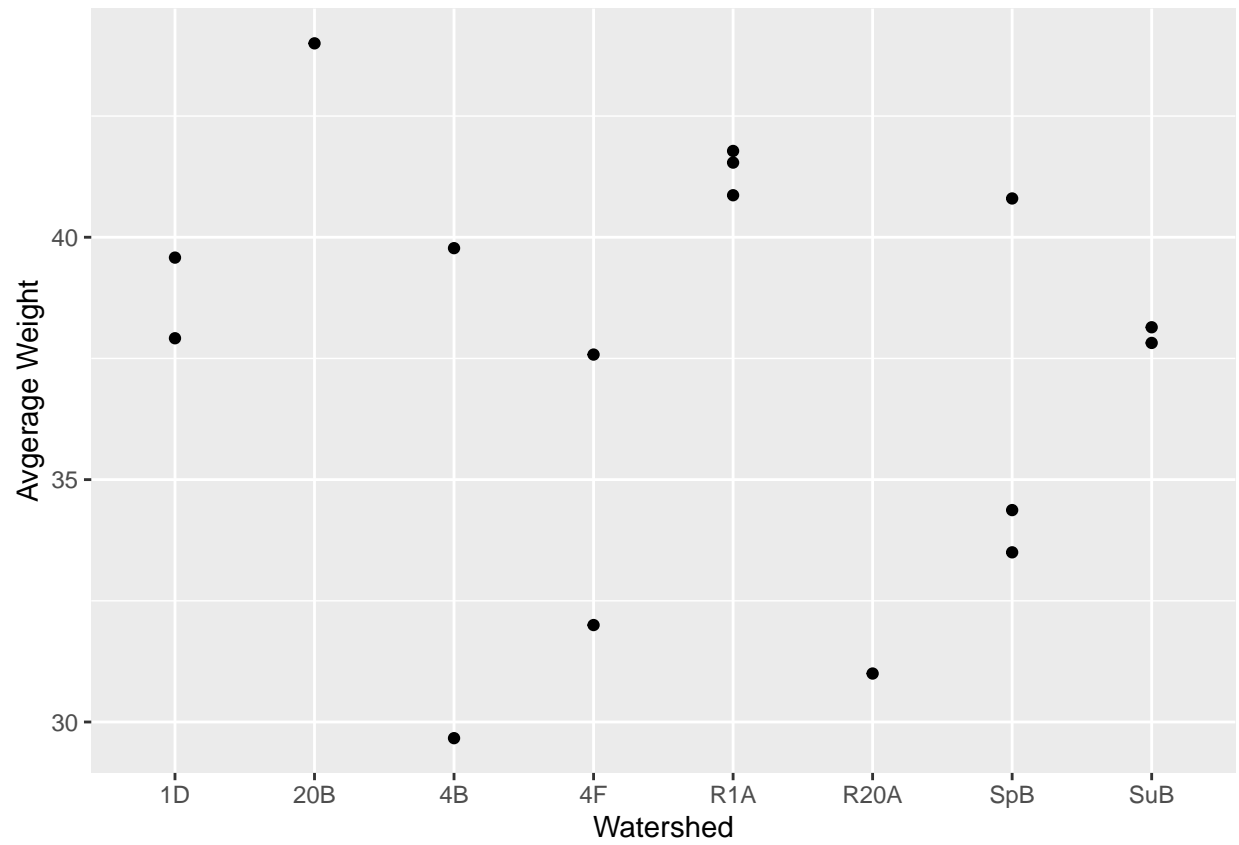
```
A2S$Watershed<-as.factor(A2S$Watershed)
A2M$Watershed<-as.factor(A2M$Watershed)
A2P$Watershed<-as.factor(A2P$Watershed)
```

2. plot expected relationships

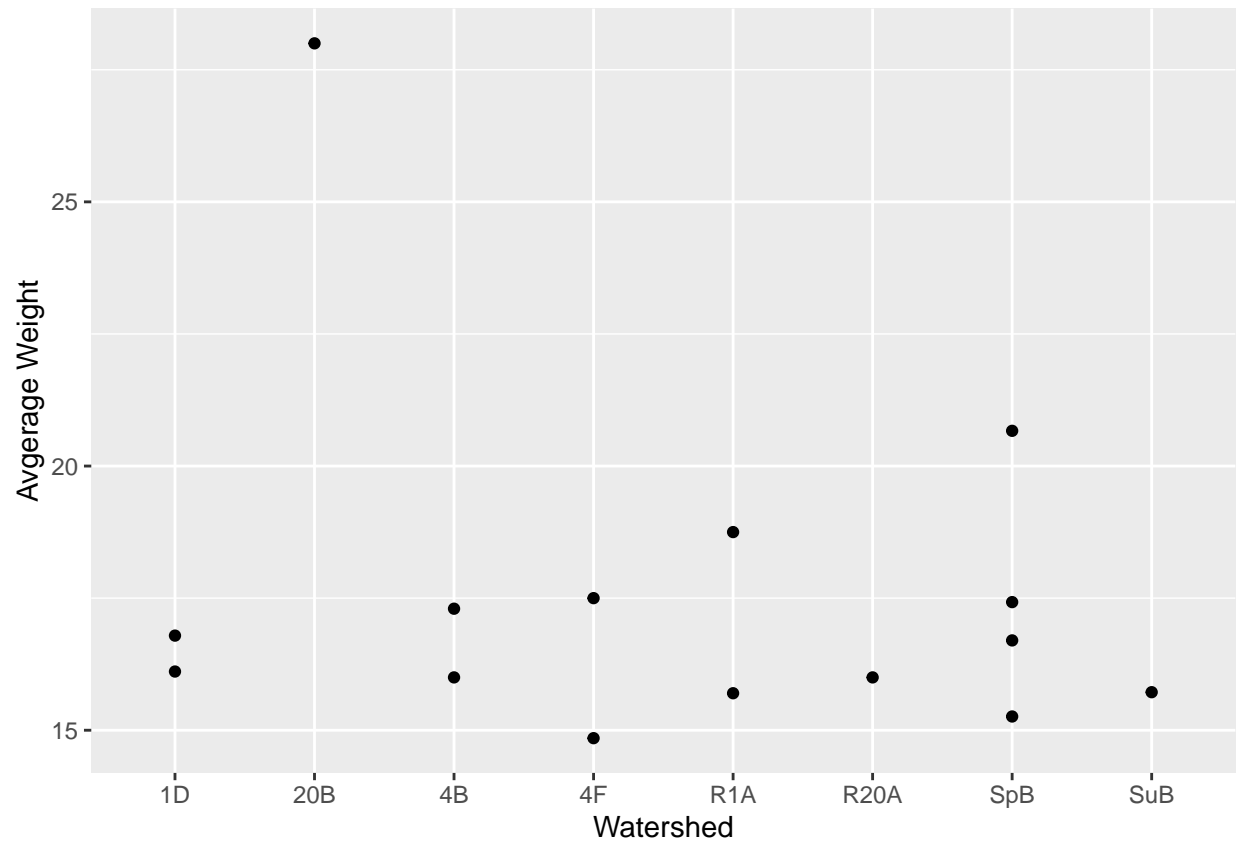
```
ggplot(A2S, aes(x = Watershed, y=avg_Weight))+
  geom_point()+
  xlab("Watershed")+
  ylab("Average Weight")
```



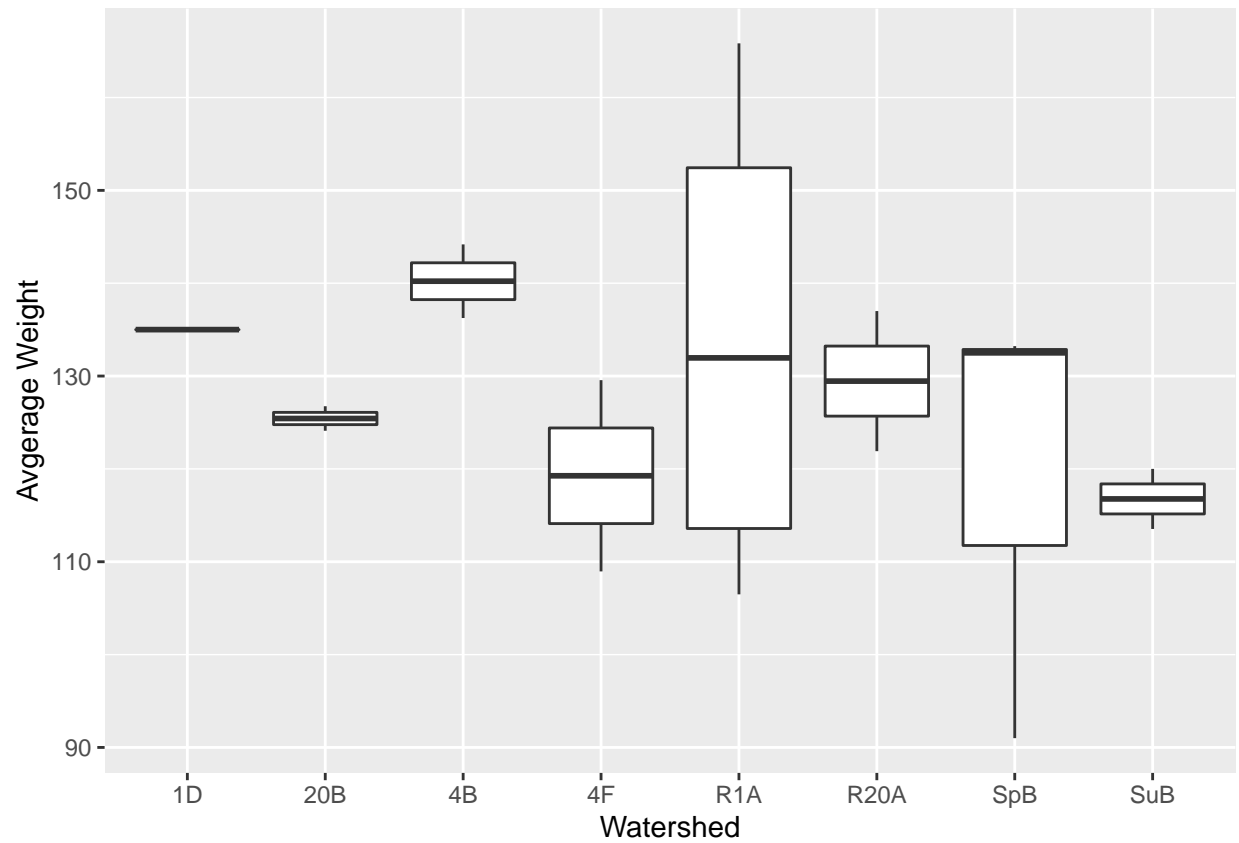
```
ggplot(A2M, aes(x = Watershed, y=avg_Weight))+  
  geom_point()+  
  xlab("Watershed")+  
  ylab("Avgerage Weight")
```



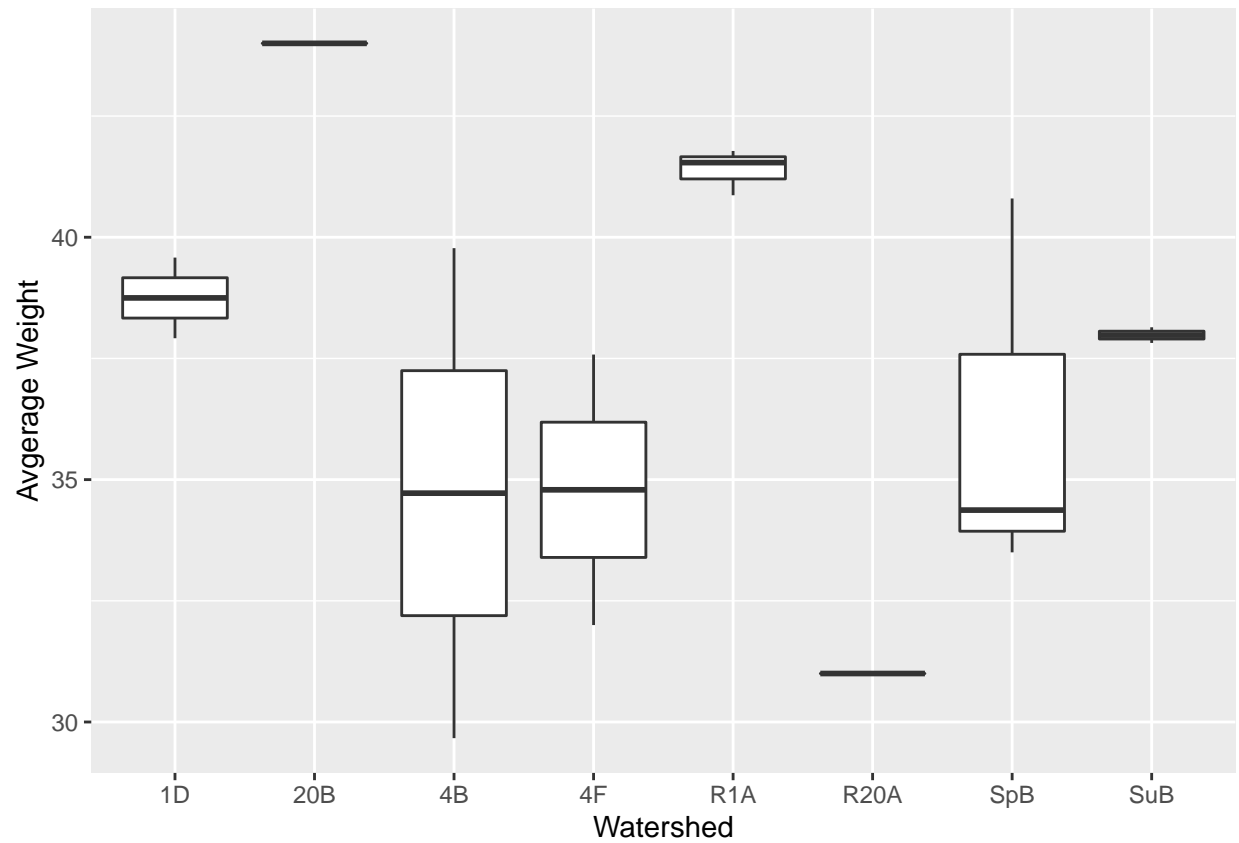
```
ggplot(A2P, aes(x = Watershed, y=avg_Weight))+  
  geom_point()+  
  xlab("Watershed")+  
  ylab("Avgerage Weight")
```



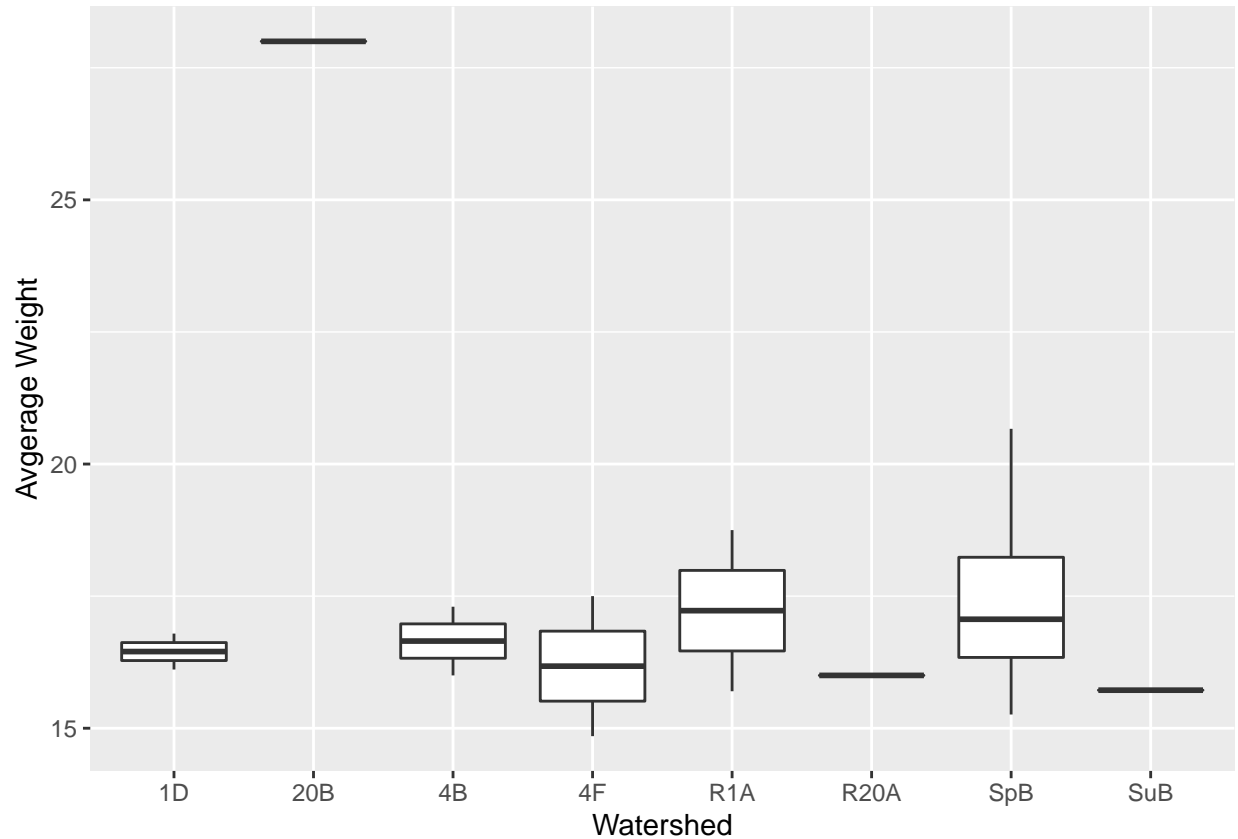
```
ggplot(A2S, aes(x = Watershed, y=avg_Weight))+  
  geom_boxplot()+  
  xlab("Watershed")+  
  ylab("Average Weight")
```



```
ggplot(A2M, aes(x = Watershed, y=avg_Weight))+  
  geom_boxplot()+  
  xlab("Watershed")+  
  ylab("Average Weight")
```

```
ggplot(A2P, aes(x = Watershed, y=avg_Weight))+  
  geom_boxplot()+  
  xlab("Watershed")+  
  ylab("Average Weight")
```



3. guess expected relationships (hypothesis)

Test 1: I think there will not be a significant difference in the average weight of animals (SIHI) found in different watersheds.

Null hypothesis (Ho): The average weight of these animals found in each watershed is equal (no significant difference). Alternate hypothesis (Ha): The average weight of these animals found in each watershed is not equal (significant difference).

Test 2: I think there may be a significant difference in the average weight of animals (MIOC) found in different watersheds, specifically involving watershed 20B and watershed R20A.

Null hypothesis (Ho): The average weight of these animals found in each watershed is equal (no significant difference). Alternate hypothesis (Ha): The average weight of these animals found in each watershed is not equal (significant difference).

Test 3: I think there will be a significant difference in the average weight of animals (PMMA) found in different watersheds, specifically between 20B and all other watersheds.

Null hypothesis (Ho): The average weight of these animals found in each watershed is equal (no significant difference). Alternate hypothesis (Ha): The average weight of these animals found in each watershed is not equal (significant difference).

4. decide on statistical model and test and explain why

This is a one-way anova. This is the correct test because the response variable (or y) is average weight which is a quantitative continuous variable, and the predictor/explanatory (or x) is Watershed which is categorical. I am asking about difference in group means, which makes this an anova. If I wanted this to be a two-way anova, I would have to add another categorical variable (column) to the dataframe that could also impact weight, and could be influenced by or influence watershed.

5. create model

```
modWS<-lm(avg_Weight~Watershed,data=A2S)

modWM<-lm(avg_Weight~Watershed,data=A2M)

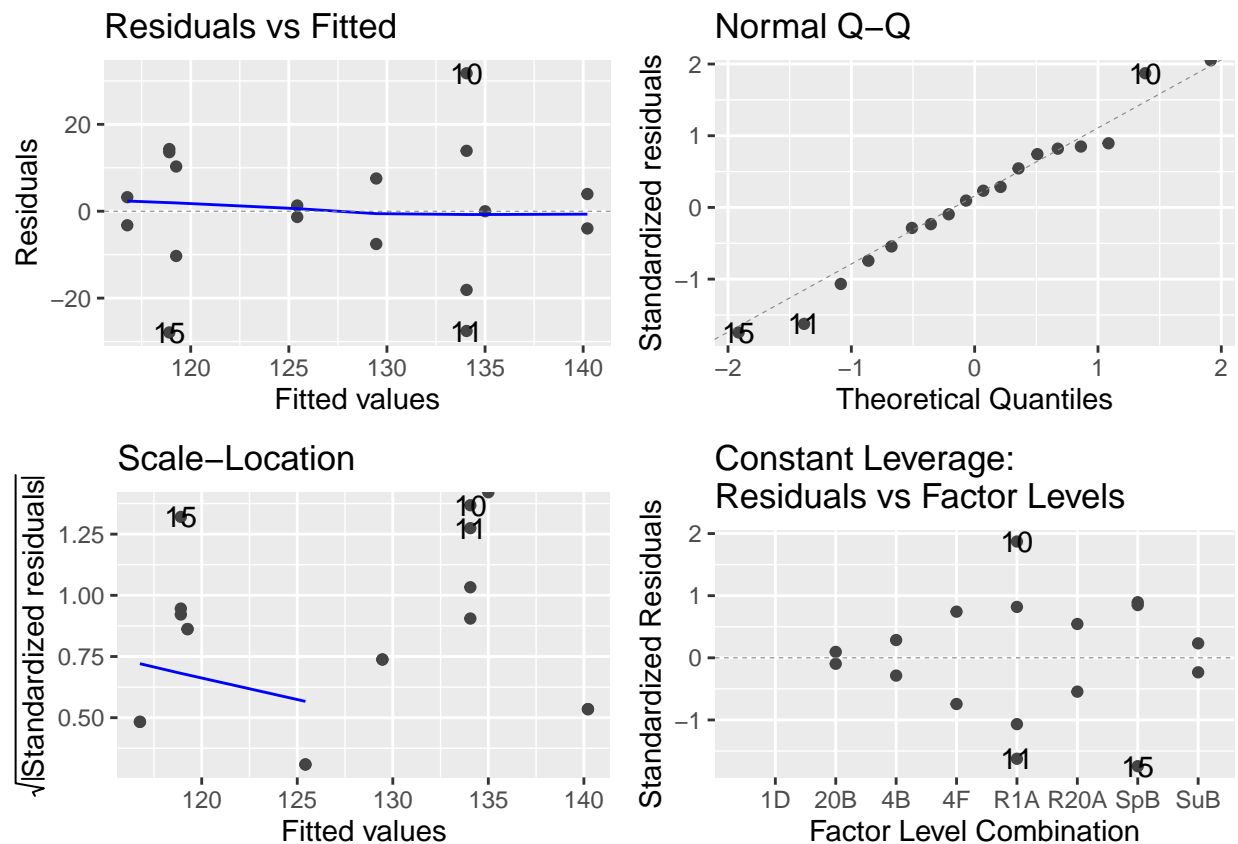
modWP<-lm(avg_Weight~Watershed,data=A2P)
```

6. check model assumptions

```
autoplot(modWS)
```

```
## Warning: Removed 9 row(s) containing missing values (geom_path).
```

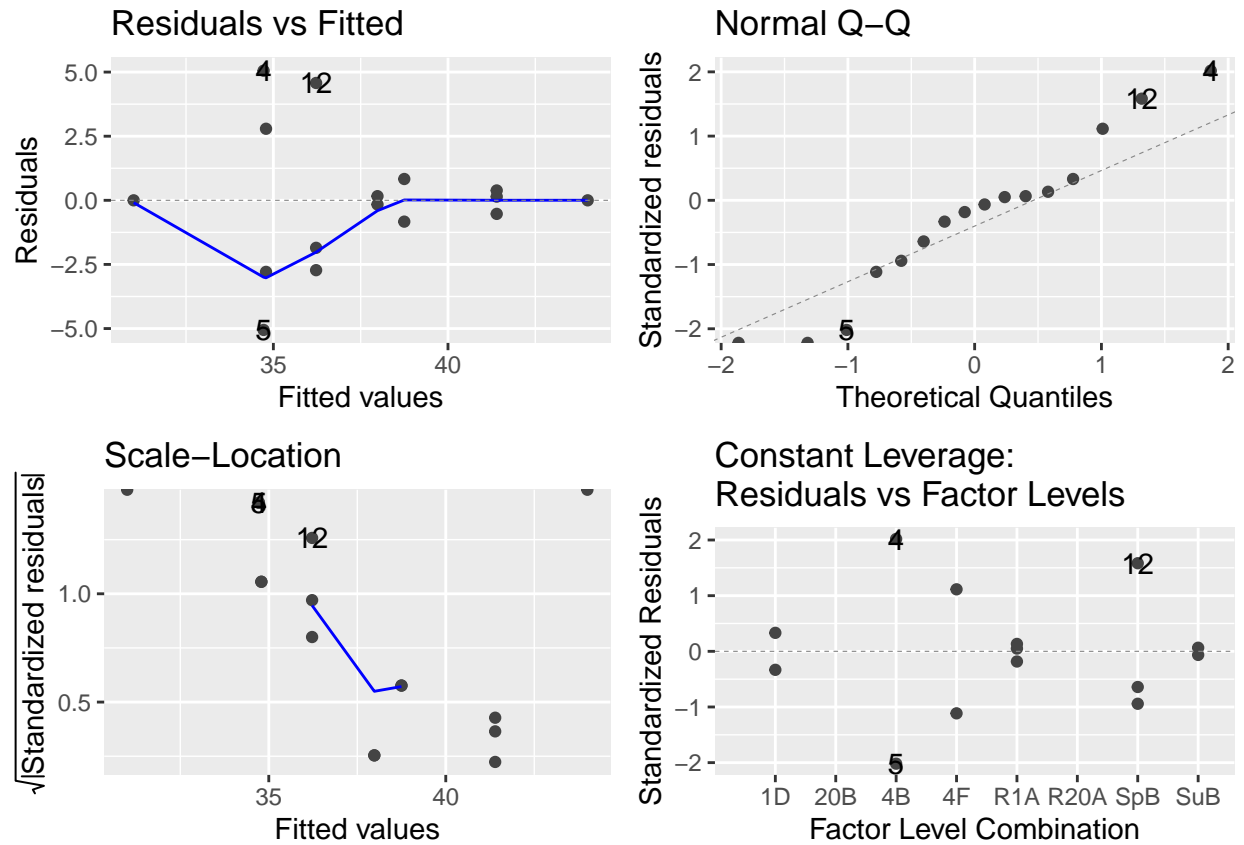
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
autoplot(modWM)
```

```
## Warning: Removed 9 row(s) containing missing values (geom_path).
```

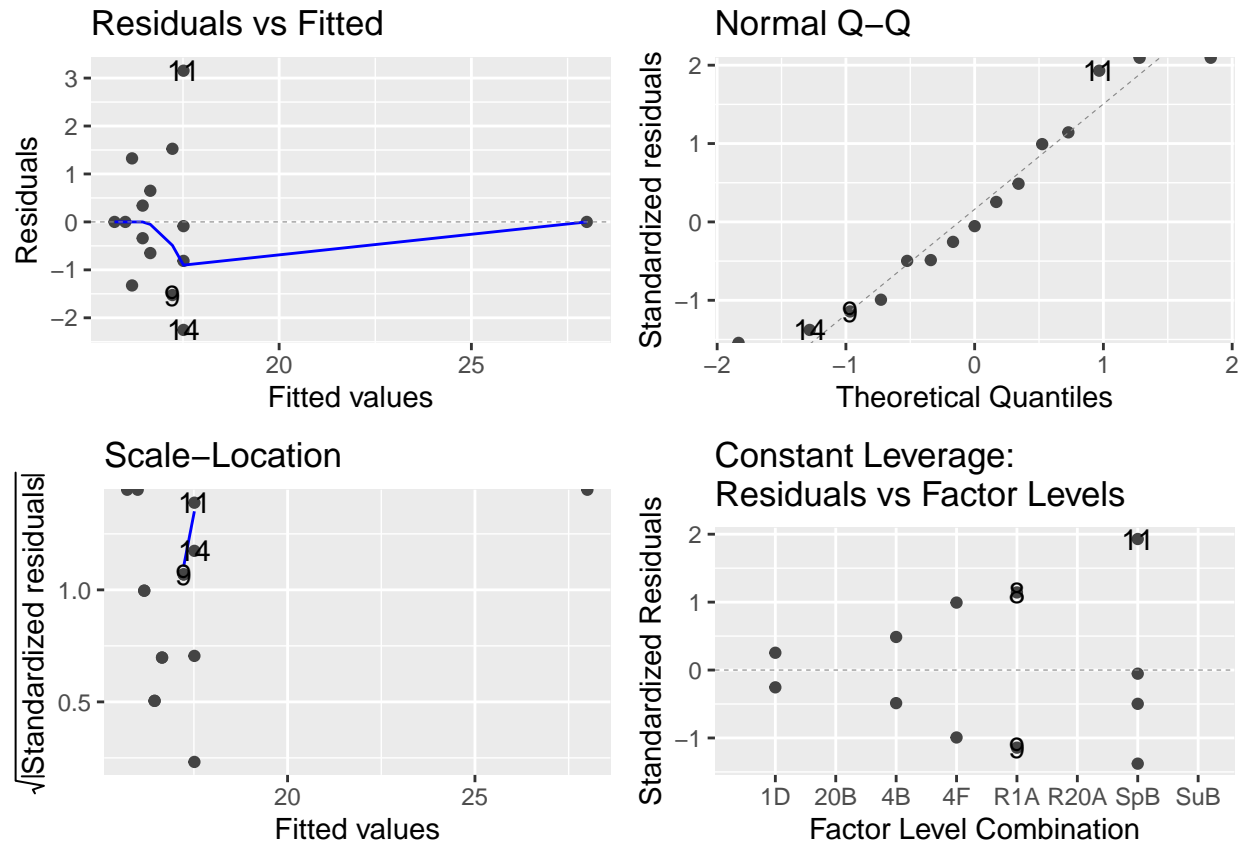
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



```
autoplot(modWP)
```

```
## Warning: Removed 9 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```



These are a little bit funky, but I feel comfortable accepting them and moving forward.

7. run stat model

a) for SIHI

```
aov(modWS)
```

```
## Call:
##   aov(formula = modWS)
##
## Terms:
##               Watershed Residuals
## Sum of Squares  1154.360 3841.545
## Deg. of Freedom      7      10
##
## Residual standard error: 19.59986
## Estimated effects may be unbalanced
```

```
anova(modWS)
```

```
## Analysis of Variance Table
##
## Response: avg_Weight
##           Df Sum Sq Mean Sq F value Pr(>F)
## Watershed  7 1154.4   164.91   0.4293 0.8629
## Residuals 10 3841.5    384.15
```

```
summary(modWS)
```

```
##
## Call:
## lm(formula = avg_Weight ~ Watershed, data = A2S)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.907  -6.650   0.661   9.610  31.775
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  135.0000    19.5999   6.888 4.26e-05 ***
## Watershed20B   -9.5722    24.0048  -0.399   0.698
## Watershed4B     5.2150    24.0048   0.217   0.832
## Watershed4F   -15.7414    24.0048  -0.656   0.527
## WatershedR1A   -0.9414    21.9133  -0.043   0.967
## WatershedR20A  -5.5444    24.0048  -0.231   0.822
## WatershedSpB  -16.0926    22.6320  -0.711   0.493
## WatershedSuB  -18.2333    24.0048  -0.760   0.465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.6 on 10 degrees of freedom
## Multiple R-squared:  0.2311, Adjusted R-squared:  -0.3072
## F-statistic: 0.4293 on 7 and 10 DF,  p-value: 0.8629
```

b) for MIOC

```
aov(modWM)
```

```
## Call:
##      aov(formula = modWM)
##
## Terms:
##              Watershed Residuals
## Sum of Squares  168.6208  100.3332
## Deg. of Freedom       7         8
##
## Residual standard error: 3.54142
## Estimated effects may be unbalanced
```

```
anova(modWM)
```

```
## Analysis of Variance Table
##
## Response: avg_Weight
##              Df Sum Sq Mean Sq F value Pr(>F)
## Watershed    7  168.62  24.089   1.9207   0.19
## Residuals    8  100.33  12.542
```

```
summary(modWM)
```

```
##
## Call:
## lm(formula = avg_Weight ~ Watershed, data = A2M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -5.0542 -1.0868 0.0000 0.4962 5.0542
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   38.7483     2.5042   15.474 3.03e-07 ***
## Watershed20B    5.2517     4.3373    1.211  0.261
## Watershed4B   -4.0275     3.5414   -1.137  0.288
## Watershed4F   -3.9583     3.5414   -1.118  0.296
## WatershedR1A    2.6472     3.2329    0.819  0.437
## WatershedR20A  -7.7483     4.3373   -1.786  0.112
## WatershedSpB   -2.5245     3.2329   -0.781  0.457
## WatershedSuB   -0.7669     3.5414   -0.217  0.834
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.541 on 8 degrees of freedom
## Multiple R-squared:  0.627, Adjusted R-squared:  0.3005
## F-statistic: 1.921 on 7 and 8 DF, p-value: 0.19
```

c) for PMMA

```
aov(modWP)
```

```
## Call:
## aov(formula = modWP)
##
## Terms:
##             Watershed Residuals
## Sum of Squares  122.9723   24.9289
## Deg. of Freedom      7         7
##
## Residual standard error: 1.887133
## Estimated effects may be unbalanced
```

```
anova(modWP)
```

```
## Analysis of Variance Table
##
## Response: avg_Weight
##           Df Sum Sq Mean Sq F value Pr(>F)
## Watershed  7 122.972  17.5675   4.9329 0.02583 *
## Residuals  7   24.929   3.5613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(modWP)
```

```
##
## Call:
## lm(formula = avg_Weight ~ Watershed, data = A2P)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2529 -0.7315  0.0000  0.4950  3.1538
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      16.451      1.334  12.328  5.3e-06 ***
## Watershed20B     11.549      2.311   4.997  0.00157 **
## Watershed4B       0.199      1.887   0.105  0.91898
## Watershed4F      -0.276      1.887  -0.146  0.88784
## WatershedR1A      0.774      1.887   0.410  0.69396
## WatershedR20A    -0.451      2.311  -0.195  0.85083
## WatershedSpB      1.062      1.634   0.650  0.53658
## WatershedSuB     -0.731      2.311  -0.316  0.76101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.887 on 7 degrees of freedom
## Multiple R-squared:  0.8314, Adjusted R-squared:  0.6629
## F-statistic: 4.933 on 7 and 7 DF,  p-value: 0.02583
```

8. interpret model results (were expected relationships supported?)

Test 1: SIHI p-value = 0.8629 > 0.05 = not significant

Based on a p-value of 0.8629 (F-stat=0.4293 on 7 and 10 degrees of freedom) I fail to reject the null hypothesis. There were also no significant differences (p-value < 0.05) within the summary.

Test 2: MIOC p-value = 0.19 > 0.05 = not significant

Based on a p-value of 0.19 (F-stat=0.1.921 on 7 and 8 degrees of freedom) I fail to reject the null hypothesis. There were also no significant differences (p-value < 0.05) within the summary.

Test 3: PMMA p-value = 0.02583 < 0.05 = significant

Based on an overall p-value of 0.02583 (F-stat=4.933 on 7 and 7 degrees of freedom) I reject the null hypothesis, meaning there is a significant difference in mean PMMA animal weights in different watersheds. When looking at the summary, watershed 20B had a significant p-value of 0.00157. This indicates that watershed 20B has significantly different weights of PMMA animals found there than the weights of PMMA animals found in other watersheds, but I am going to examine that further with a Tukey test to see more clearly where the significance lies.

```
ttmodWP<-aov(modWP)
TukeyHSD(ttmodWP, conf.level=.95)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = modWP)
##
## $Watershed
##          diff          lwr          upr          p adj
## 20B-1D    11.5489899    2.044992  21.052987  0.0188494
## 4B-1D      0.1989899   -7.560992   7.958971  1.0000000
## 4F-1D     -0.2760101   -8.035992   7.483971  0.9999998
## R1A-1D     0.7739899   -6.985992   8.533971  0.9997861
## R20A-1D    -0.4510101   -9.955008   9.052987  0.9999986
## SpB-1D     1.0619066   -5.658435   7.782248  0.9961699
## SuB-1D    -0.7310101  -10.235008   8.772987  0.9999619
## 4B-20B   -11.3500000  -20.853998  -1.846002  0.0206572
## 4F-20B   -11.8250000  -21.328998  -2.321002  0.0166217
## R1A-20B   -10.7750000  -20.278998  -1.271002  0.0270290
## R20A-20B  -12.0000000  -22.974271  -1.025729  0.0324341
## SpB-20B  -10.4870833  -19.163006  -1.811160  0.0193843
```

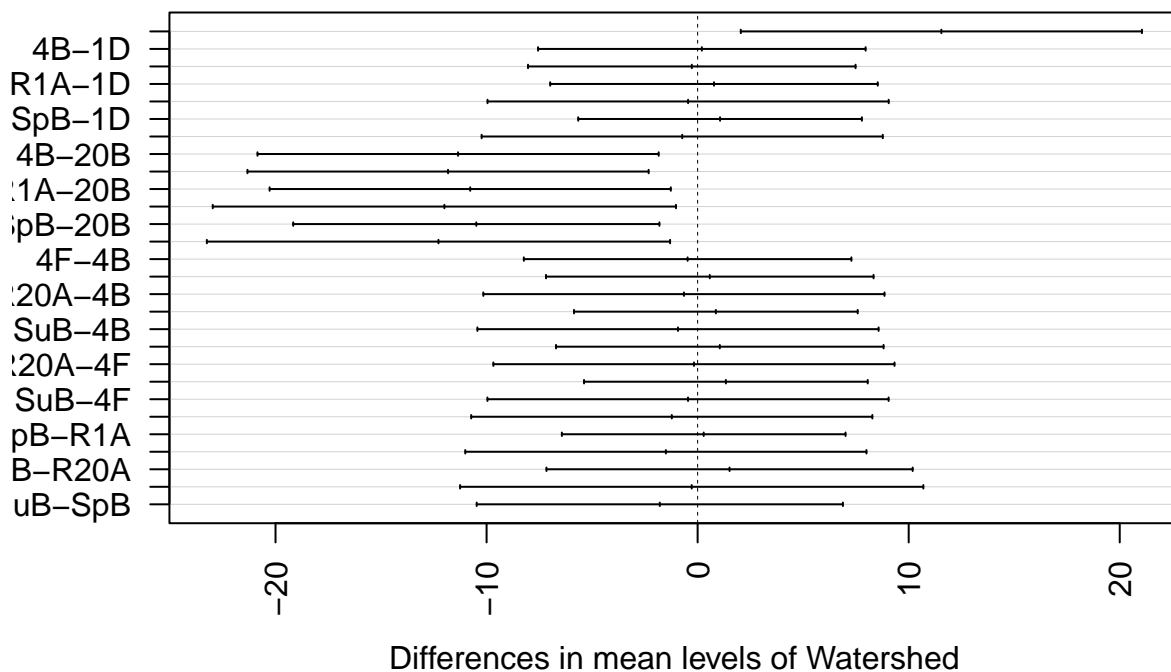


```
## SuB-20B -12.280000 -23.254271 -1.305729 0.0288870
## 4F-4B -0.4750000 -8.234982 7.284982 0.9999919
## R1A-4B 0.5750000 -7.184982 8.334982 0.9999703
## R20A-4B -0.6500000 -10.153998 8.853998 0.9999827
## SpB-4B 0.8629167 -5.857424 7.583258 0.9989207
## SuB-4B -0.9300000 -10.433998 8.573998 0.9998113
## R1A-4F 1.0500000 -6.709982 8.809982 0.9985046
## R20A-4F -0.1750000 -9.678998 9.328998 1.0000000
## SpB-4F 1.3379167 -5.382424 8.058258 0.9857466
## SuB-4F -0.4550000 -9.958998 9.048998 0.9999985
## R20A-R1A -1.2250000 -10.728998 8.278998 0.9988946
## SpB-R1A 0.2879167 -6.432424 7.008258 0.9999993
## SuB-R1A -1.5050000 -11.008998 7.998998 0.9961207
## SpB-R20A 1.5129167 -7.163006 10.188840 0.9931945
## SuB-R20A -0.2800000 -11.254271 10.694271 1.0000000
## SuB-SpB -1.7929167 -10.468840 6.883006 0.9825835
```

This shows a significant difference between watershed 20B and 1D, 20B and 4B, 20B and 4F, 20B and R1A, 20B and R20A, 20B and SpB, and 20B and SuB. This clearly shows that weights of PMMA animals found in watershed 20B were significantly different than the weights of PMMA animals found in all the other watersheds

```
plot(TukeyHSD(ttmodWP, conf.level=.95), las = 2)
```

95% family-wise confidence level



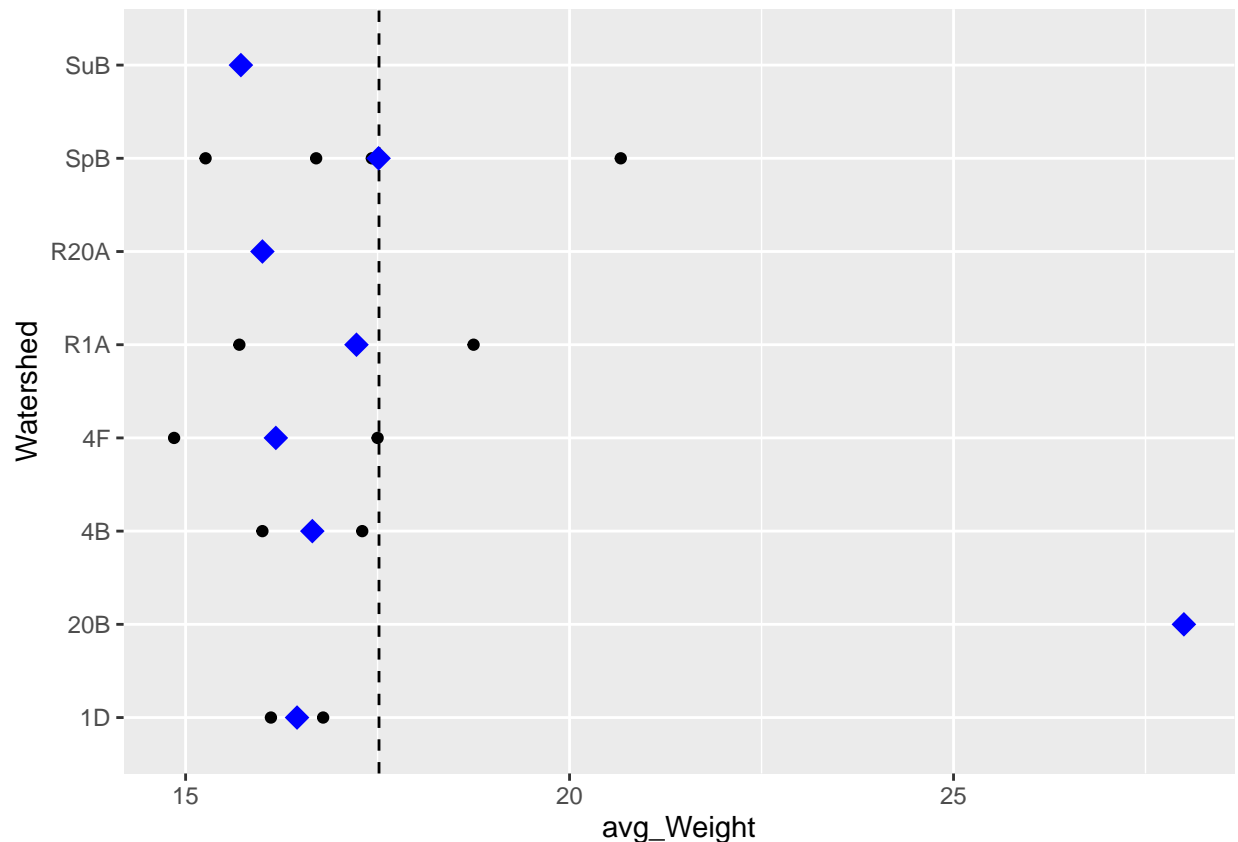
This plot shows that the only significant difference in weights is found between watershed 20B and all other watersheds.

9. replot data to show key relationships

Show only sig differences found in Test 3

```
sumDat<-A2P%>%
  group_by(Watershed)%>%
  summarise(meanWT=mean(avg_Weight))

ggplot(A2P,aes(x=avg_Weight,y=Watershed),colour="black")+
  geom_point()+
  geom_vline(xintercept = mean(A2P$avg_Weight, na.rm=TRUE), colour="black", lty="dashed")+
  geom_point(data=sumDat,aes(x=meanWT,y=Watershed),shape="diamond", colour="blue", size=4)
```



This figure shows the mean weight per watershed (symbolized by the blue diamond) and how far away it is from the overall mean weight for the species PMMA (symbolized by the dotted black line).

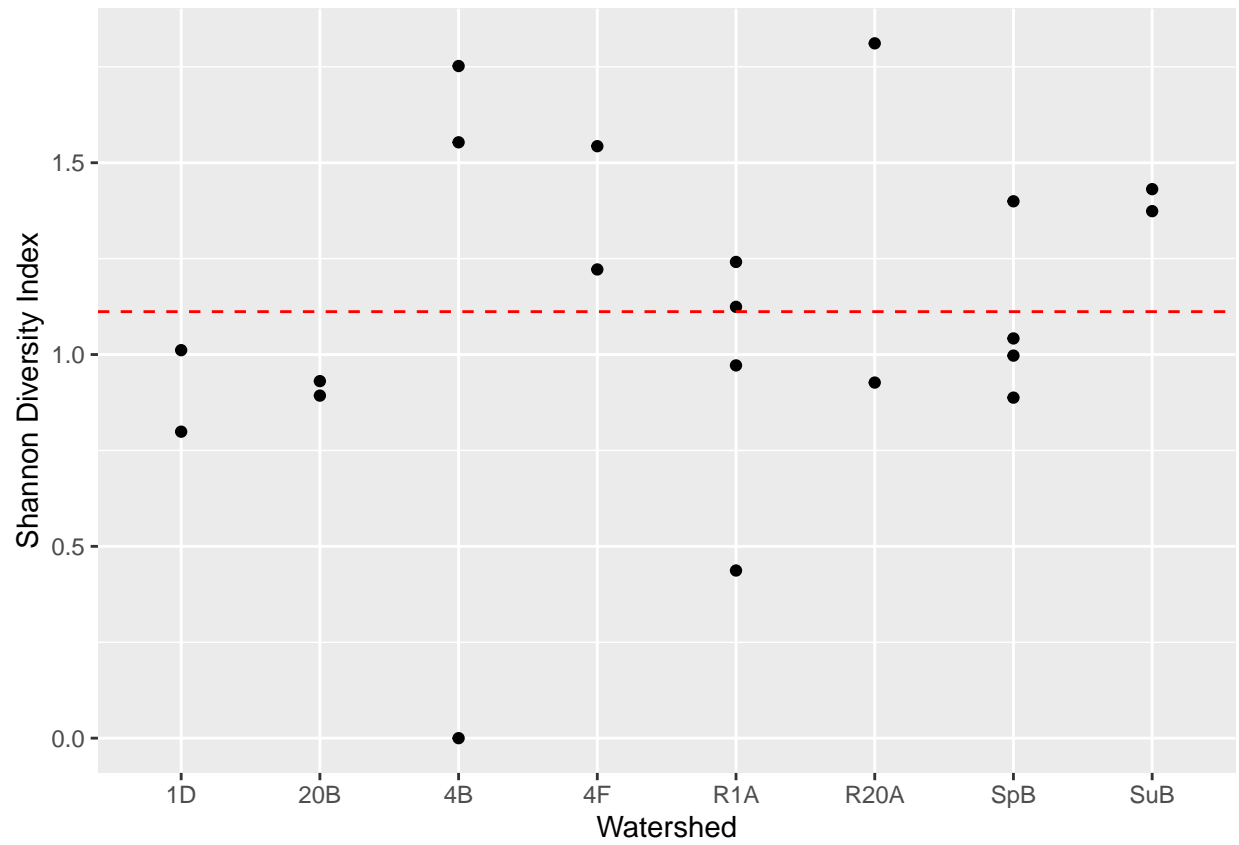
Analysis 3:

1. pull in data

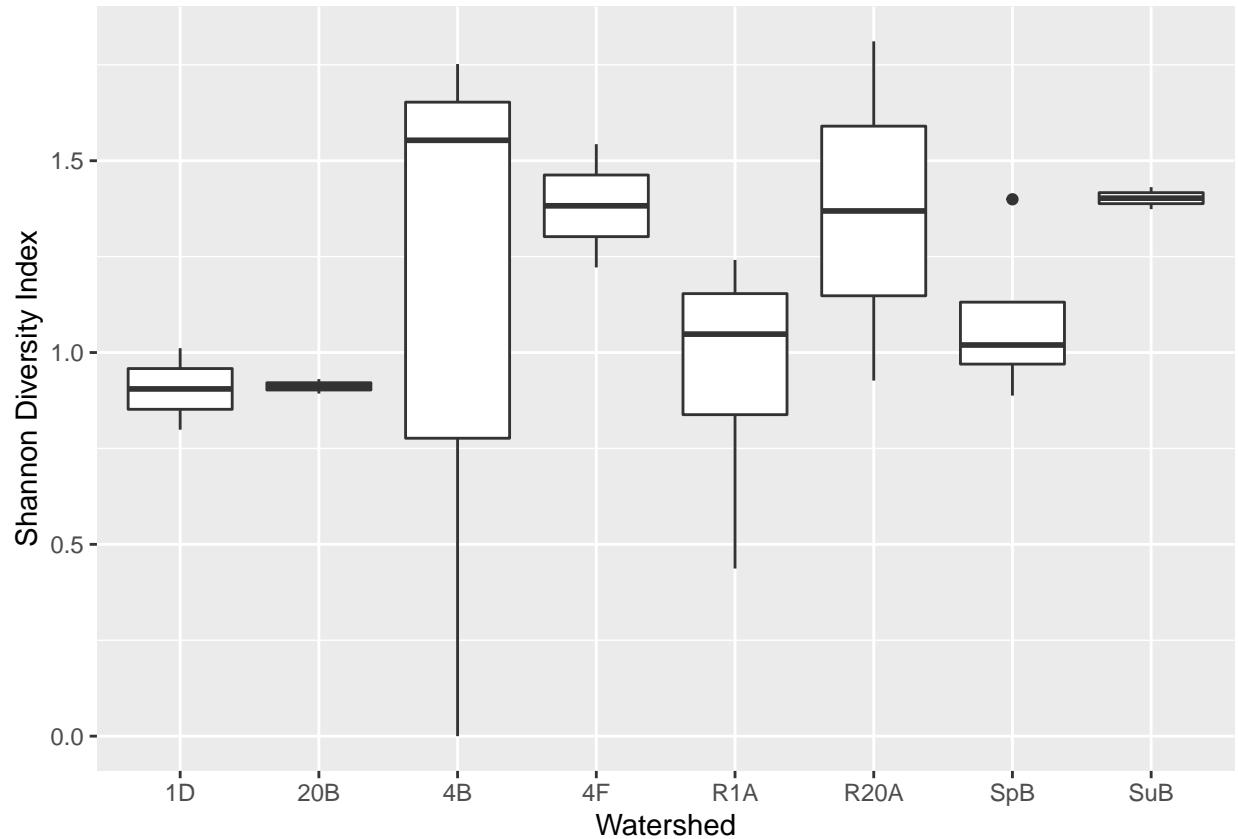
```
DivShan <- read.csv(here("Data", "Processed Data", "Ws_Sp_Shannon.csv"))
```

2. plot expected relationships

```
ggplot(DivShan, aes(x = Watershed, y=Sh_div))+
  geom_point()+
  geom_hline(yintercept = mean(DivShan$Sh_div, na.rm=TRUE), colour="red", lty="dashed")+
  xlab("Watershed")+
  ylab("Shannon Diversity Index")
```



```
ggplot(DivShan, aes(x = Watershed, y=Sh_div))+  
  geom_boxplot()+  
  xlab("Watershed")+  
  ylab("Shannon Diversity Index")
```



3. guess expected relationships (hypothesis)

I expect there will be a small significant difference between the Shannon diversity indexes in different watersheds.

Null hypothesis (Ho): The Shannon diversity indexes calculated for each watershed is equal (no significant difference). Alternate hypothesis (Ha): The Shannon diversity indexes calculated for each watershed is not equal (significant difference).

4. decide on statistical model and test and explain why

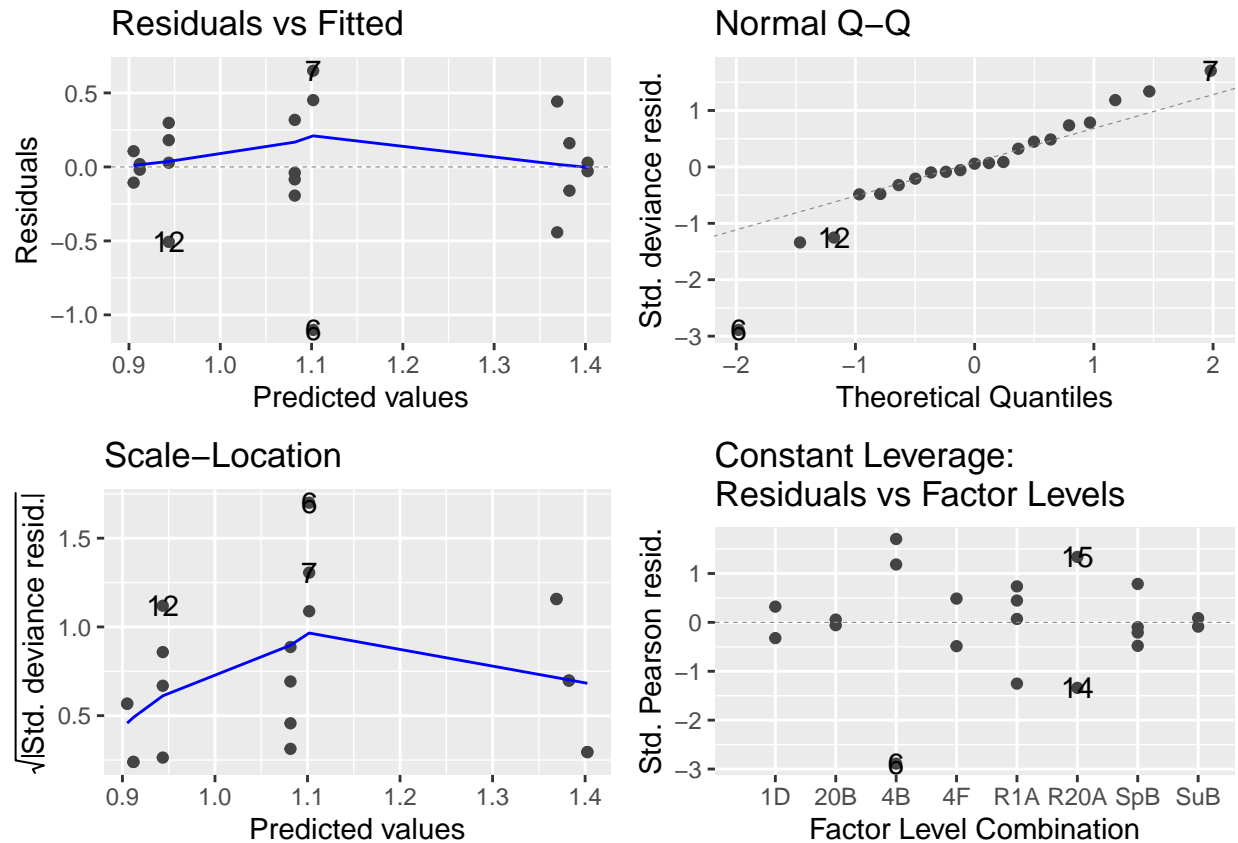
This test is an anova using glm. This is the correct test because the response variable (or y) is the Shannon diversity index which is a quantitative continuous variable, and the predictor/explanatory (or x) is Watershed which is categorical. I am using glm() because this data doesn't fit the linear model (lm()) distribution requirements.

5. create model

```
modDiv<-glm(Sh_div~Watershed,data=DivShan)
```

6. check model assumptions

```
autoplot(modDiv)
```



I feel comfortable accepting these and moving on.

7. run stat model

```
aov(modDiv)
```

```
## Call:
## aov(formula = modDiv)
##
## Terms:
##           Watershed Residuals
## Sum of Squares  0.7303502 2.8343291
## Deg. of Freedom      7      13
##
## Residual standard error: 0.4669318
## Estimated effects may be unbalanced
```

```
summary(modDiv)
```

```
##
## Call:
## glm(formula = Sh_div ~ Watershed, data = DivShan)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10179  -0.10625   0.01888   0.18072   0.65029
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    0.905135    0.330171    2.741    0.0168 *
## Watershed20B   0.006697    0.466932    0.014    0.9888
## Watershed4B    0.196654    0.426248    0.461    0.6522
## Watershed4F    0.477346    0.466932    1.022    0.3253
## WatershedR1A   0.038474    0.404375    0.095    0.9257
## WatershedR20A  0.463923    0.466932    0.994    0.3386
## WatershedSpB   0.176469    0.404375    0.436    0.6697
## WatershedSuB   0.497315    0.466932    1.065    0.3062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2180253)
##
## Null deviance: 3.5647 on 20 degrees of freedom
## Residual deviance: 2.8343 on 13 degrees of freedom
## AIC: 35.538
##
## Number of Fisher Scoring iterations: 2
```

```
anova(modDiv, test="F")
```

```
## Analysis of Deviance Table
##
## Model: gaussian, link: identity
##
## Response: Sh_div
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev      F Pr(>F)
## NULL                20      3.5647
## Watershed  7  0.73035      13      2.8343 0.4785 0.8337
```

8. interpret model results (were expected relationships supported?)

p-value=0.8337>0.05 = not significant

Based on a p-value of 0.8337, I fail to reject the null hypothesis, meaning there is not a significant difference in Shannon diversity indexes between the different watersheds.

9. replot data to show key relationships

There was not a significant difference found, therefore there aren't really key relationships to show.

##Biological Summary:

I found that for my analysis 1 and 3 there were no significant differences found. There was not a significant difference in the number of species found in each watershed (analysis 1) and there was not a significant difference in species diversity in each watershed (analysis 3) either. In analysis 2 I performed 3 identical tests on three different species, the third of which ended up yielding a significant result. There was a significant difference between mean PMMA weights per watersheds, specifically between mean weights found in watershed 20B and all other watersheds. This could be because watershed 20B has a 20 year burn replicate treatment, which is a lot longer between burns than the other burn replicate treatments. This could allow plant litter to accumulate, larger plants to establish, etc. The only watershed that comes close to this timeframe is the R20A watershed, which means they changed the treatment from an annual burn to a 20-year burn replicate. The weights of PMMA animals were not significantly different than the other watersheds however for a few possible reasons. The change from annual burn to 20 year burn may have happened fairly recently, which

would not have allowed time for plant litter to accumulate at the level it has on watershed 20B.

##Challenges:

There were a number of challenges I faced throughout this project, including:

- The error messages have been a struggle in the analysis section.
- Working on VPN is difficult as well because it kicks me off a lot.
- Learning how to make the data frame for analysis 3 and all the work and trial and error that went into that.
 - Shannon diversity index and vegan package
- glm things
- making “beautiful” figures