02a\_beeReproDev\_DataExploration

Jessie Dodge

2024-05-23

# Library

source(file="01i\_beeReproDev\_MasterData.R")

# Questions

Q1) How does landscape and climate variation affect bee parasitism? \* Proposed analyses: MLRs \*\* Parasite abundance ~ Year + Hab + Landscape + Climate + Floral Q2) How does parasitism abundance affect bee fitness traits? \* Proposed analyses: SLRs \*\* bee % ~ parasite % \*\* bee abundance ~ parasite abundance \*\* bee mass ~ parasite abundance (2021) \*\* F% ~ parasite abundance (2021)

# Data org

# omit site with no provisioned cells - zeros omitted (zo)  
zo.data<-filter(data, nTotProvised > 0)  
  
# for count/weight data  
## all  
or.data <- filter(zo.data, oligRatio > 0)  
w.data <- filter(zo.data, avgOligMass\_mg > 0)  
fw.data <- filter(zo.data, avgOligFMass\_mg > 0)  
mw.data <- filter(zo.data, avgOligMMass\_mg > 0)  
lw.data <- filter(zo.data, avgOligLarvMass\_mg > 0)  
  
# remove T. stan outlier  
zo.data.or <- filter(zo.data, nTstanCells < 10)  
#or.data.or <- filter(or.data, nTstanCells < 10)

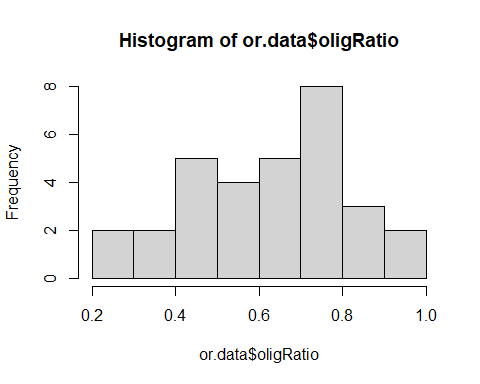
# Data distribution

## O. lignaria

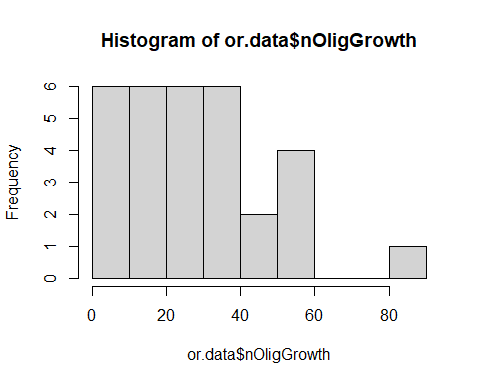
### All sites

### O.lig ratio > 0 Sites

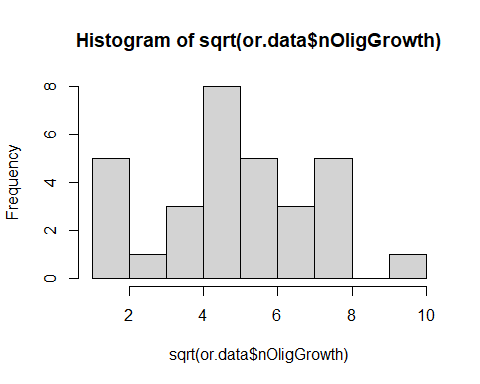
## Ratio  
hist(or.data$oligRatio) # pretty normal



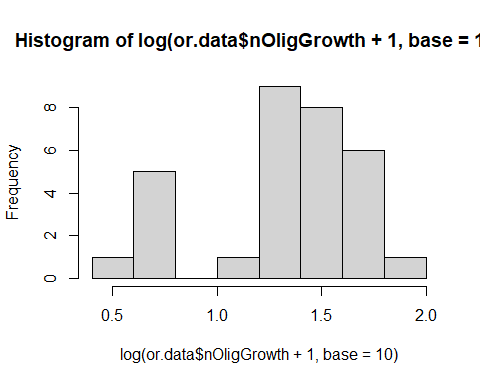
## Counts  
hist(or.data$nOligGrowth) # right skewed



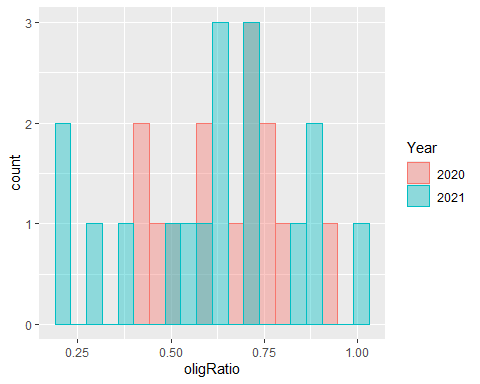
hist(sqrt(or.data$nOligGrowth)) # more normal



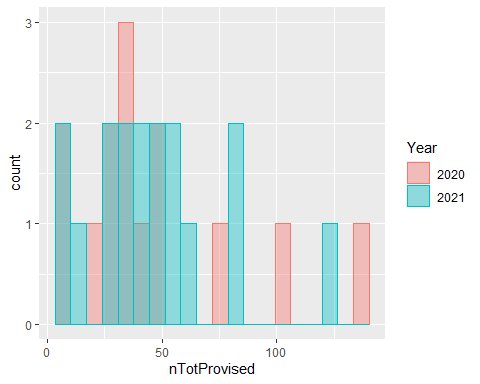
hist(log(or.data$nOligGrowth +1, base=10)) # bi-modal



### Ratio by year  
ggplot(or.data, aes(x = oligRatio)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 20, alpha = 0.4)



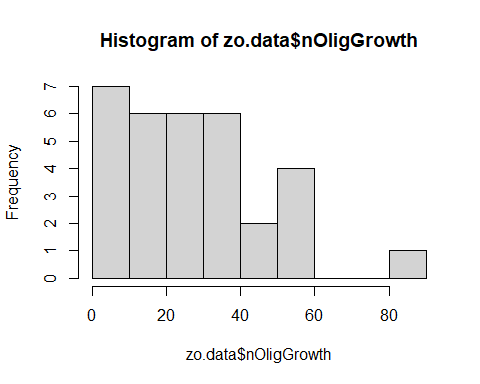
# lower ratios in 2021  
  
### Counts by year  
ggplot(or.data, aes(x = nTotProvised)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 20, alpha = 0.4)



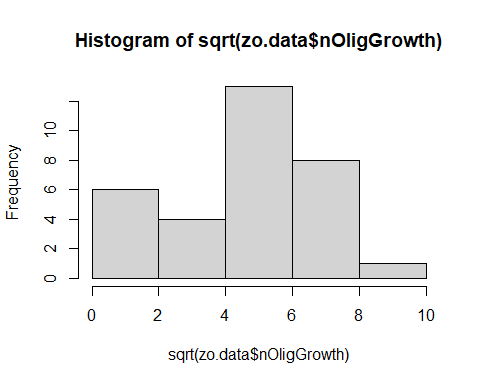
# years are pretty similar

### Provisioned Sites

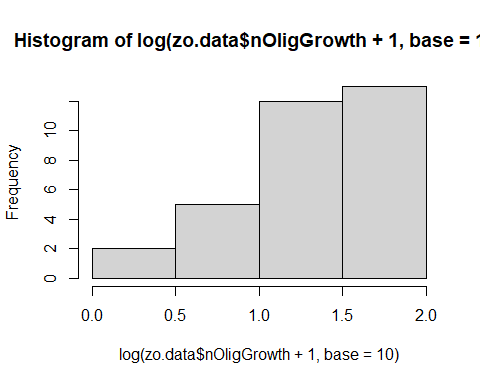
hist(zo.data$nOligGrowth) # right skewed



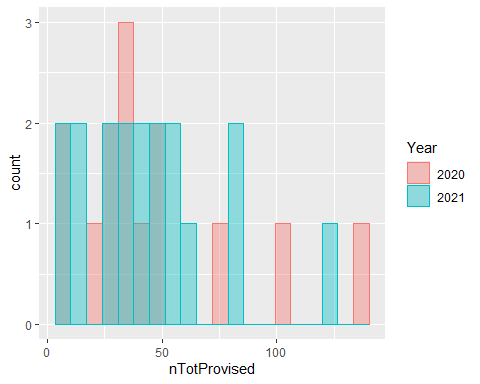
hist(sqrt(zo.data$nOligGrowth)) # pretty normal



hist(log(zo.data$nOligGrowth+1, base=10)) # left skewed



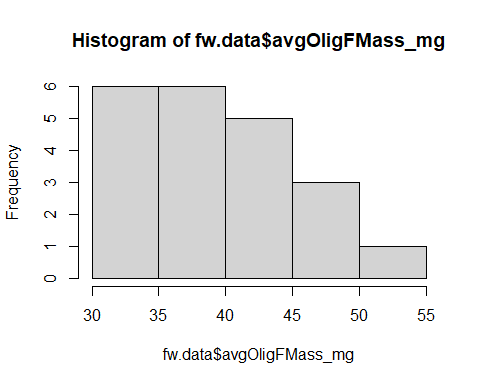
### Counts by year  
ggplot(zo.data, aes(x = nTotProvised)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 20, alpha = 0.4)



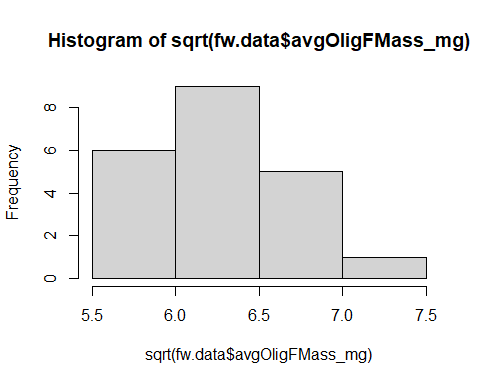
# years are pretty similar

### O.lig weights

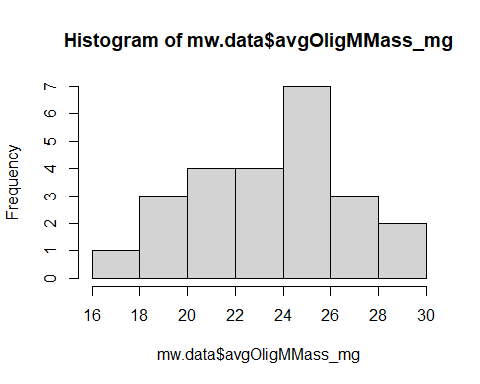
# female  
hist(fw.data$avgOligFMass\_mg) # right skewed



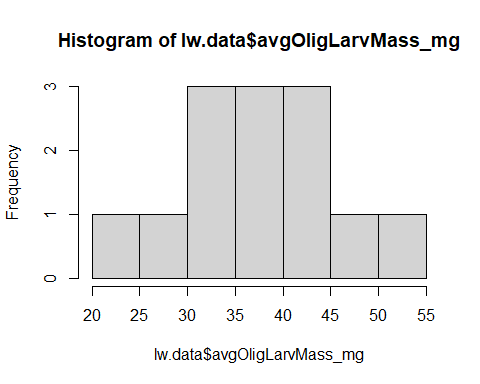
hist(sqrt(fw.data$avgOligFMass\_mg)) # more normal



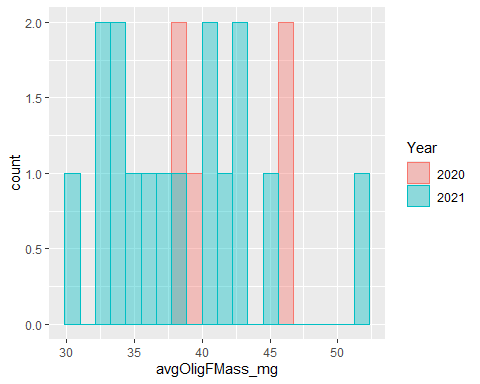
# Male  
hist(mw.data$avgOligMMass\_mg) # fairly normal



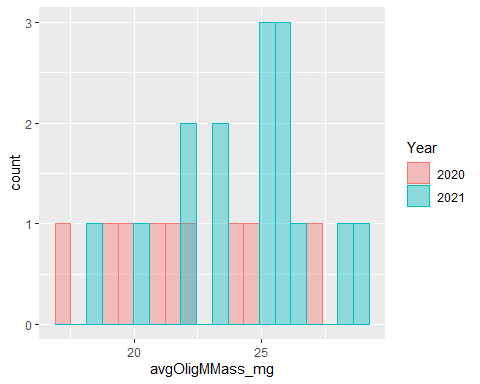
# Larvae  
hist(lw.data$avgOligLarvMass\_mg) # normal



### Counts by year  
ggplot(fw.data, aes(x = avgOligFMass\_mg)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 20, alpha = 0.4)



# only 5 females in 2020, but within the dist. of 2021 weights  
  
ggplot(mw.data, aes(x = avgOligMMass\_mg)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 20, alpha = 0.4)

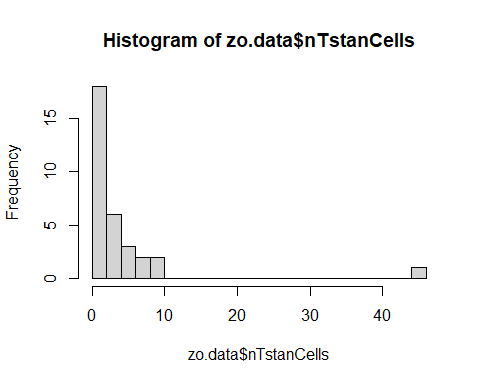


# 2021 is normally distributed. Not many males in 2020

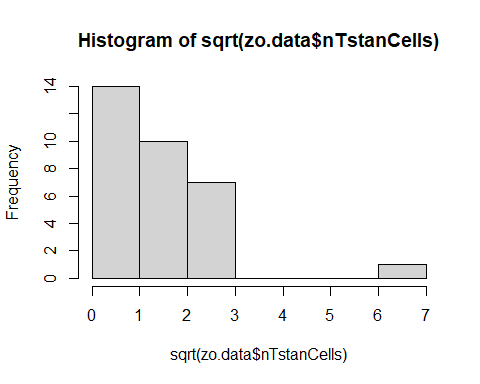
## T.stansburyi

Provisioned Sites - would only be present if there were provisioned cells

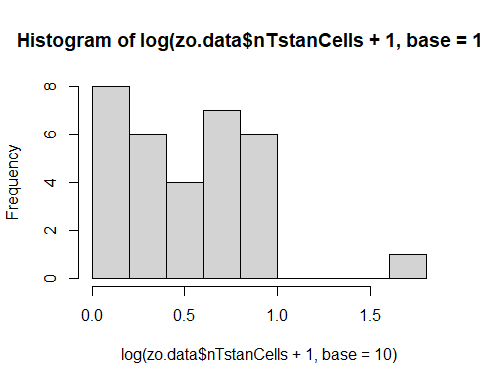
# with outlier  
hist(zo.data$nTstanCells, breaks = 20) # poisson dist - HUGE outlier



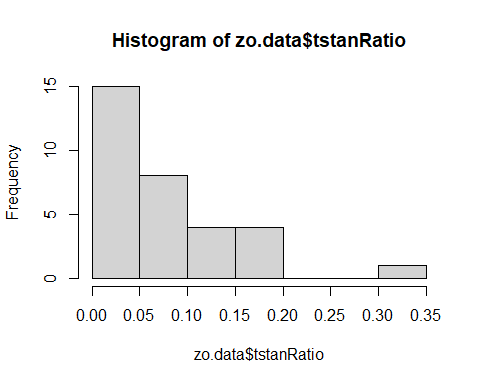
hist(sqrt(zo.data$nTstanCells)) # still skewed



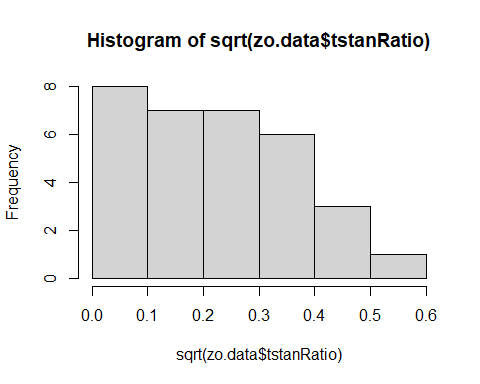
hist(log(zo.data$nTstanCells+1, base=10)) # more poisson shaped now



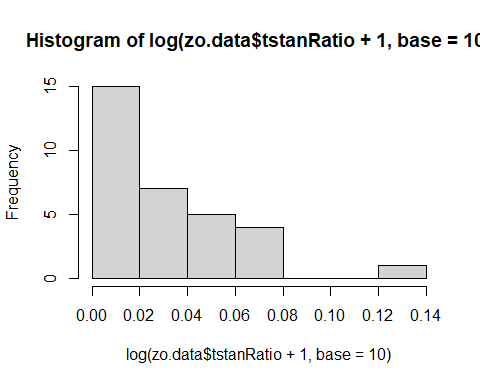
hist(zo.data$tstanRatio) # poisson dist - outlier not as problematic



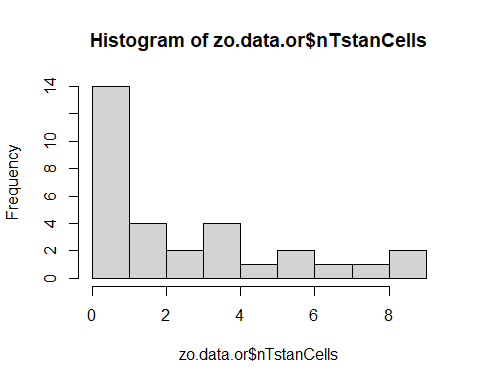
hist(sqrt(zo.data$tstanRatio)) # 0 heavy



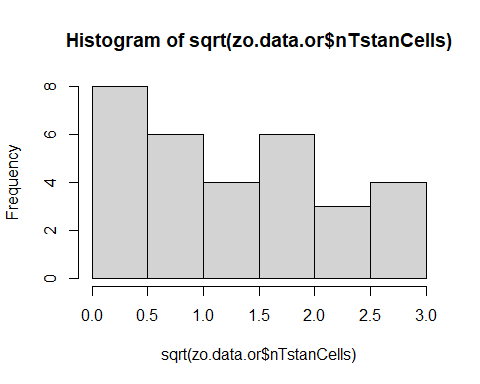
hist(log(zo.data$tstanRatio+1, base=10)) # worse



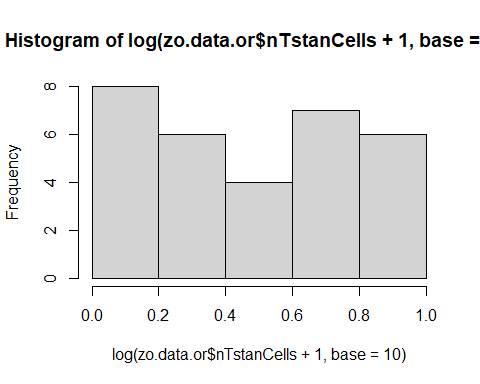
# without outlier  
hist(zo.data.or$nTstanCells, breaks = 10) # very zero heavy



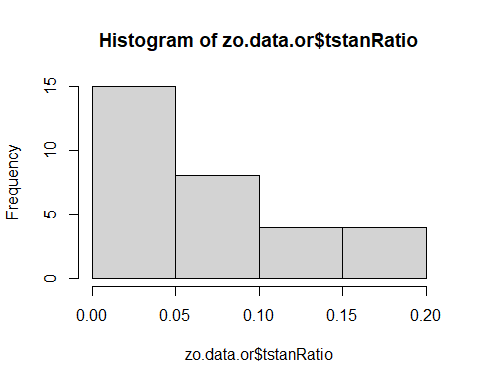
hist(sqrt(zo.data.or$nTstanCells)) # normalish



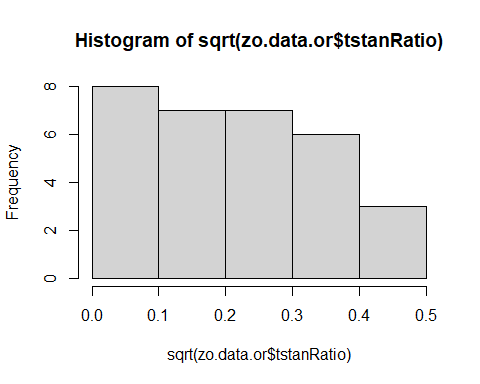
hist(log(zo.data.or$nTstanCells+1,base=10)) # same as above



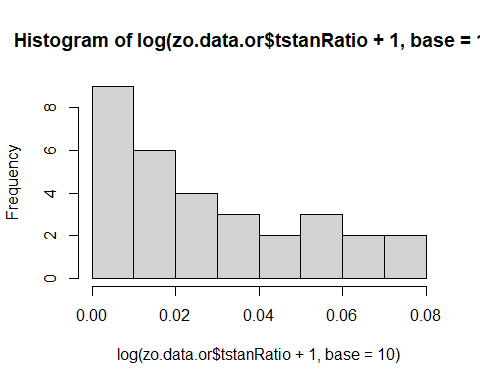
hist(zo.data.or$tstanRatio) # very zero heavy



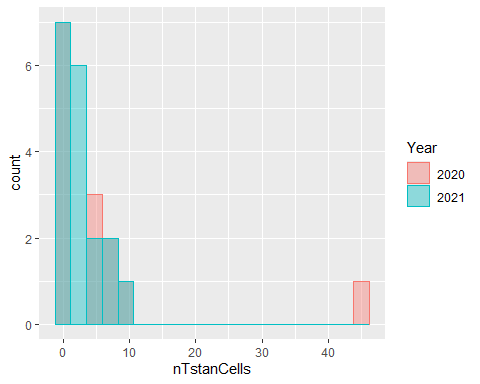
hist(sqrt(zo.data.or$tstanRatio)) # sort of normal



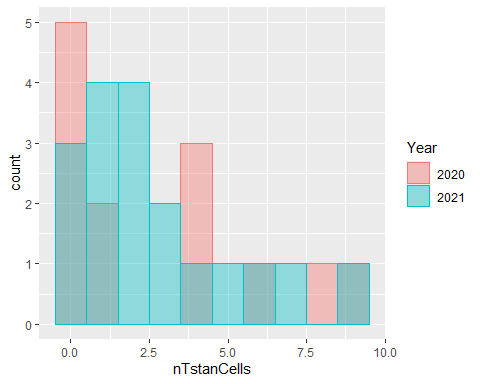
hist(log(zo.data.or$tstanRatio+1, base=10)) # right skewed



### Counts by year  
ggplot(zo.data, aes(x = nTstanCells)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 20, alpha = 0.4)



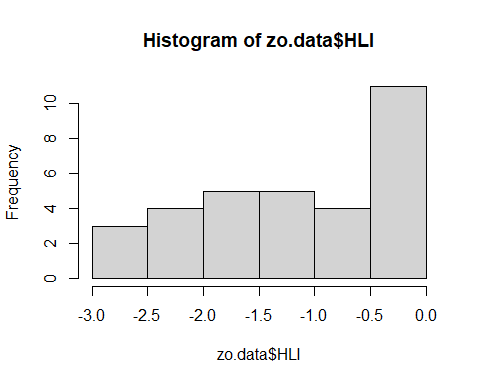
# years are pretty similar  
ggplot(zo.data.or, aes(x = nTstanCells)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 10, alpha = 0.4)



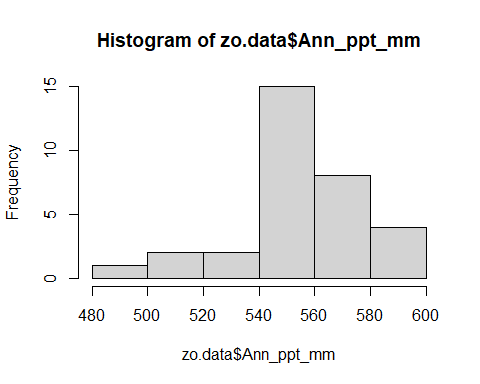
# years are pretty similar

## Climate variables

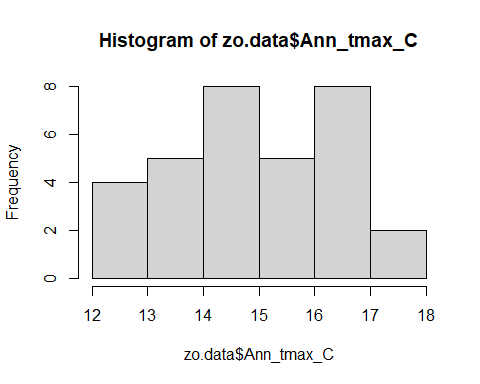
hist(zo.data$HLI)# zero heavy



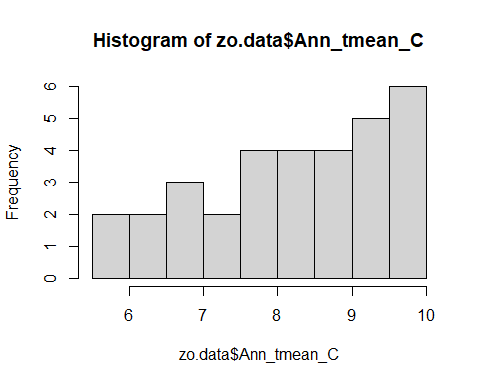
hist(zo.data$Ann\_ppt\_mm) # left skewed



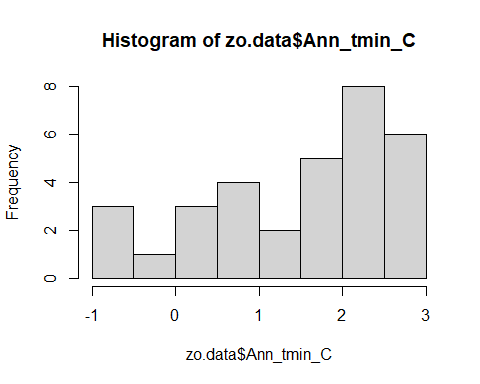
hist(zo.data$Ann\_tmax\_C) # left skewed



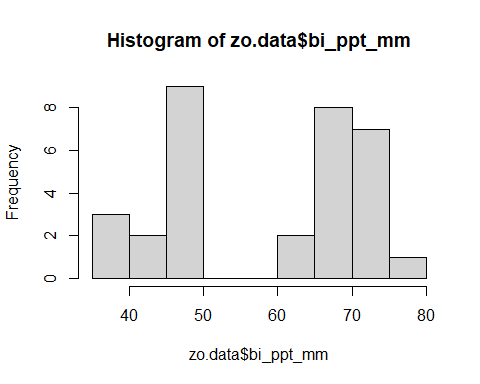
hist(zo.data$Ann\_tmean\_C) # normalish



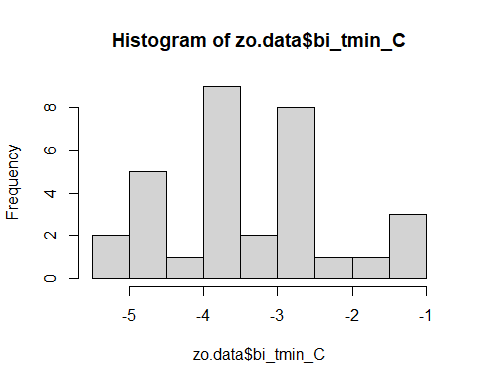
hist(zo.data$Ann\_tmin\_C) # slightly right skewed



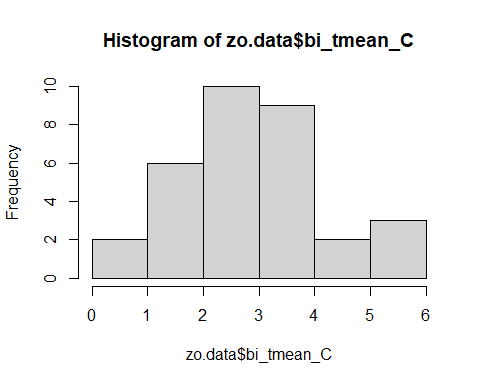
hist(zo.data$bi\_ppt\_mm) # bi-modal



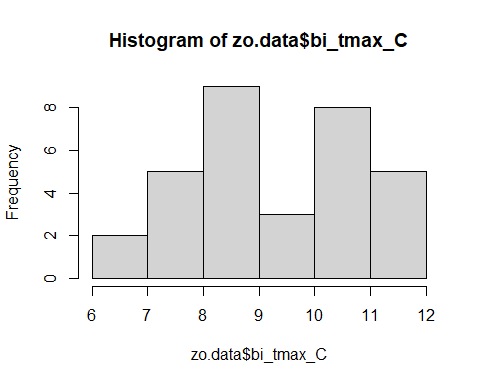
hist(zo.data$bi\_tmin\_C) # normalish?



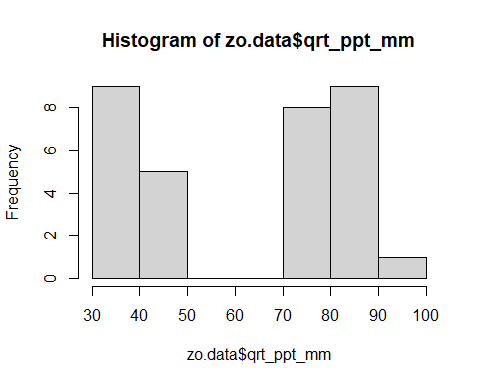
hist(zo.data$bi\_tmean\_C) # normalish



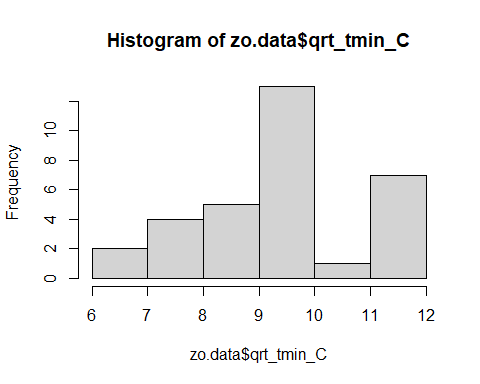
hist(zo.data$bi\_tmax\_C) # normalish



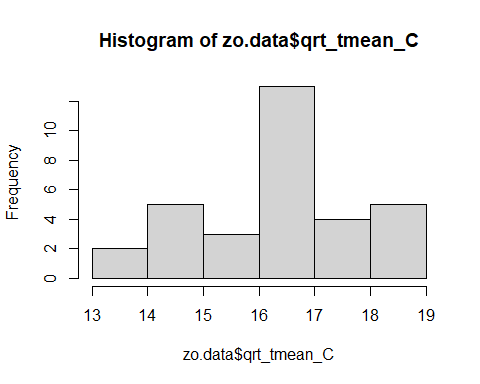
hist(zo.data$qrt\_ppt\_mm) # bi-modal



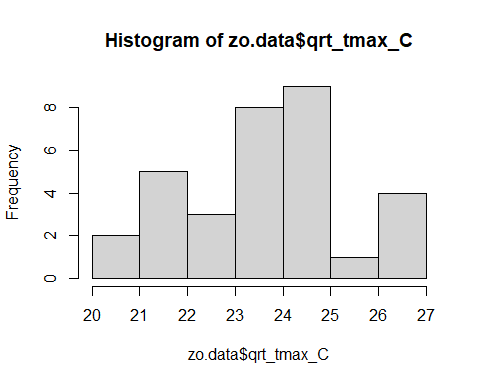
hist(zo.data$qrt\_tmin\_C) # normalish



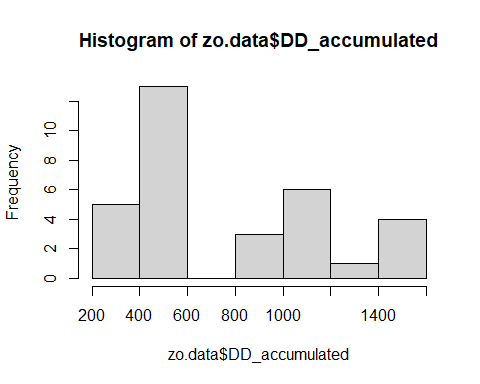
hist(zo.data$qrt\_tmean\_C) # normalish



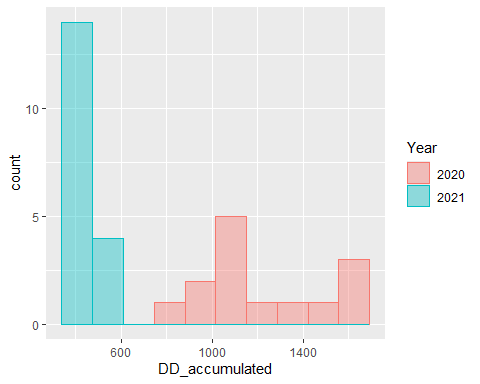
hist(zo.data$qrt\_tmax\_C) # normalish



hist(zo.data$DD\_accumulated) # bi-modal



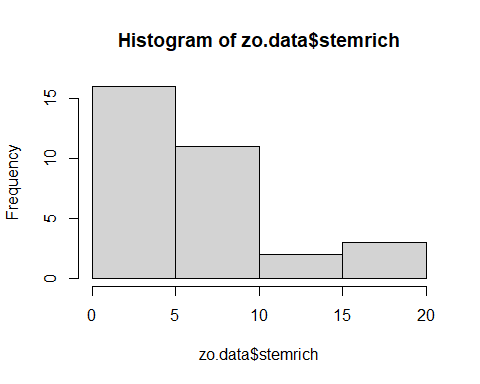
ggplot(zo.data, aes(x = DD\_accumulated)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 10, alpha = 0.4)



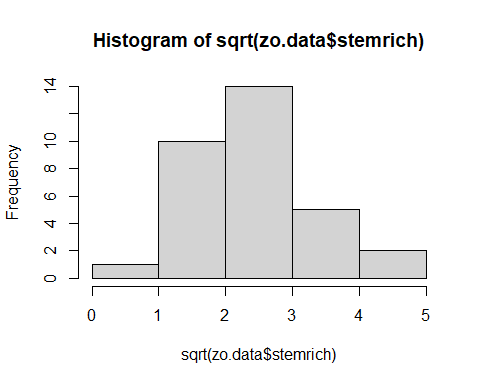
# very different by years

## Floral variables

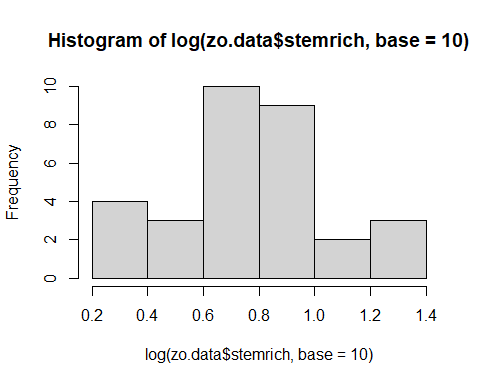
# stemrich  
hist(zo.data$stemrich) # right skewed



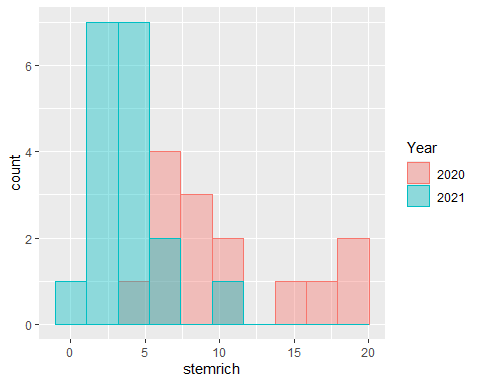
hist(sqrt(zo.data$stemrich)) # more normal



hist(log(zo.data$stemrich, base=10)) # normal



ggplot(zo.data, aes(x = stemrich)) +   
 geom\_histogram(aes(color = Year, fill = Year), position = "identity", bins = 10, alpha = 0.4)

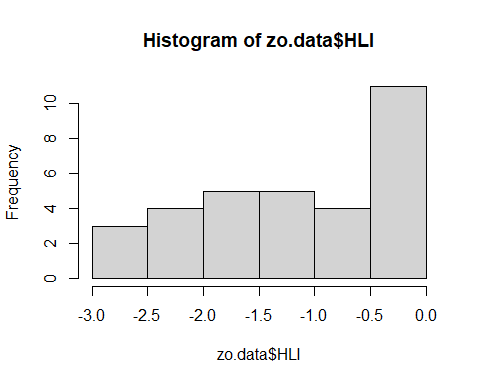


# pretty different by year

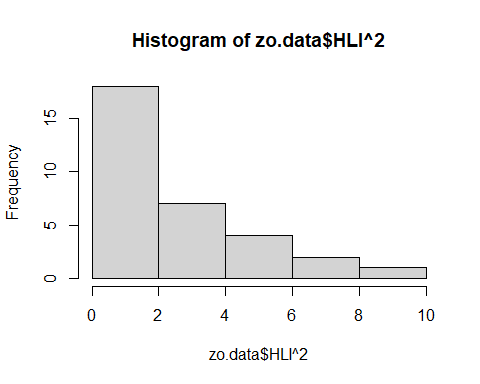
## Landscape variables

Provisioned Sites

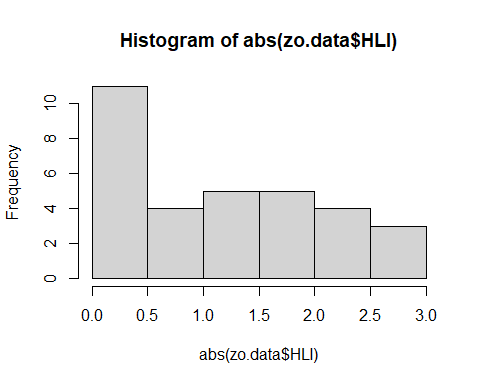
# HLI  
hist(zo.data$HLI) # left skewed - hot and dry sites?



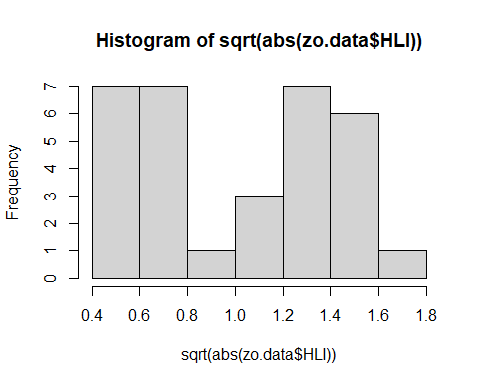
# don't know how to transform  
hist(zo.data$HLI^2) # left skewed



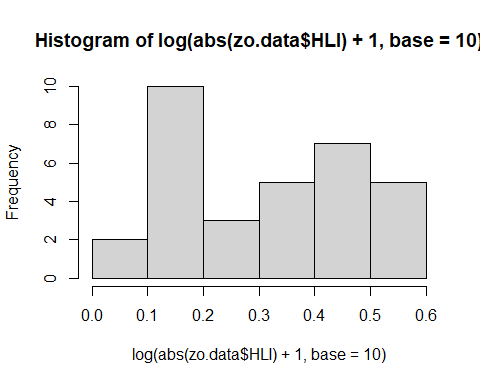
hist(abs(zo.data$HLI)) # zero heavy



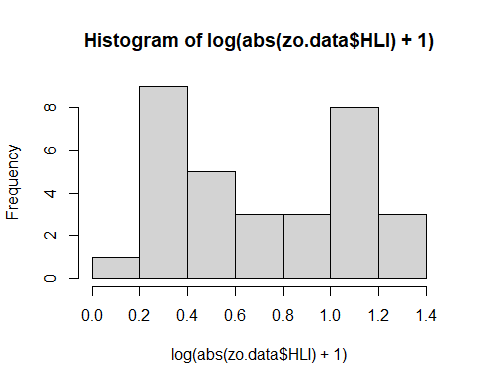
hist(sqrt(abs(zo.data$HLI))) # bimodal



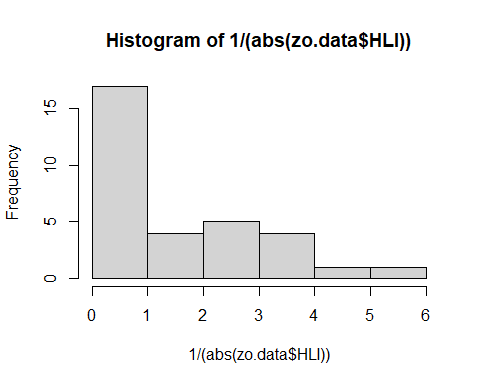
hist(log(abs(zo.data$HLI)+1,base=10))



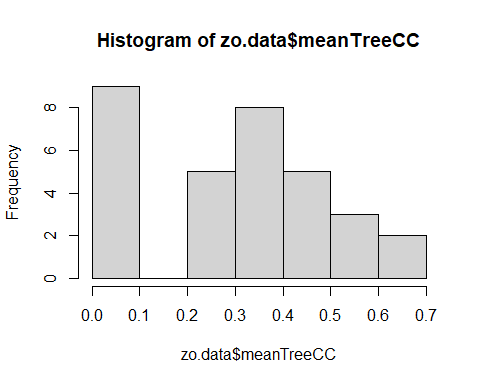
hist(log(abs(zo.data$HLI)+1)) # still left skewed, but not as bad



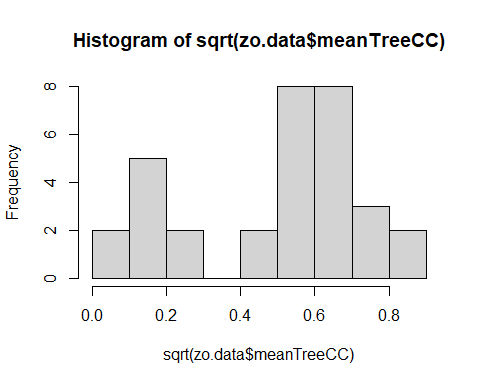
hist(1/(abs(zo.data$HLI)))



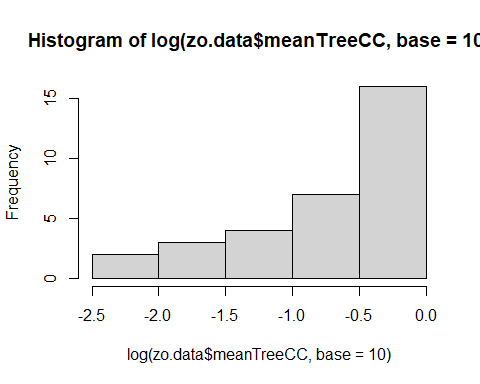
# TCC  
hist(zo.data$meanTreeCC) # zero heavy - burned sites



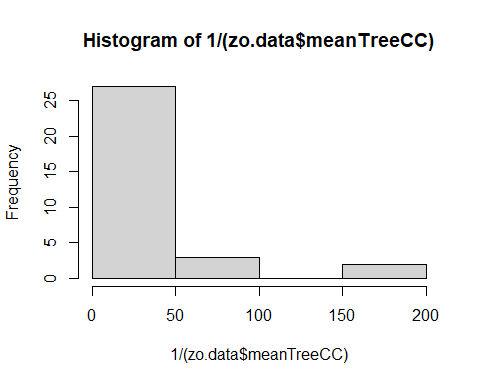
hist(sqrt(zo.data$meanTreeCC)) # still zero heavy



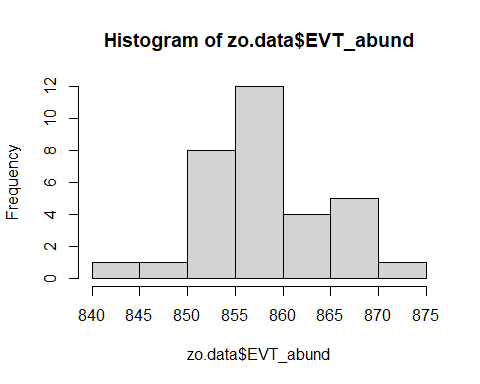
hist(log(zo.data$meanTreeCC, base = 10)) # Left skewed



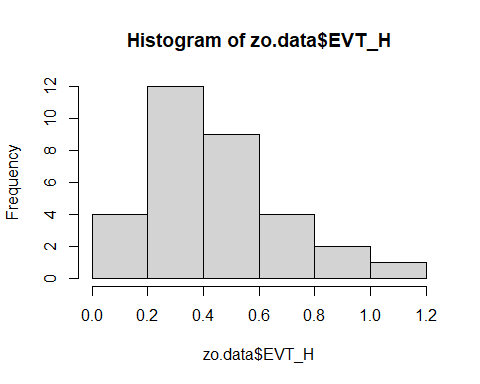
hist(1/(zo.data$meanTreeCC)) # right skewed



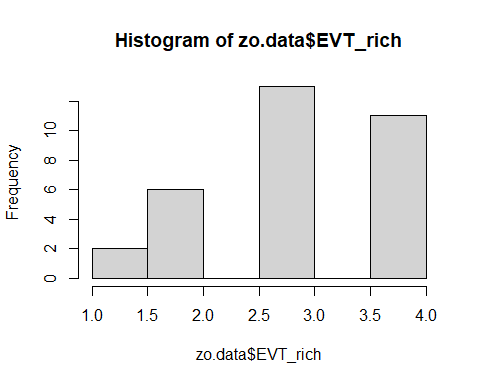
# EVT  
hist(zo.data$EVT\_abund) # normalish



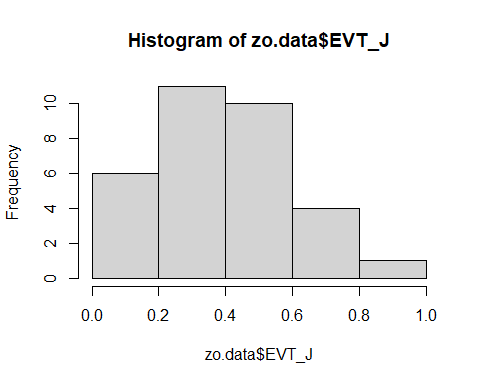
hist(zo.data$EVT\_H) # sort of right skewed



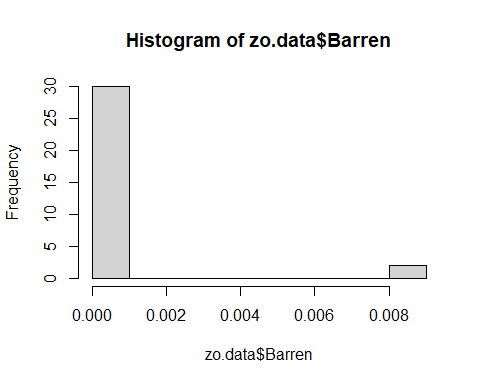
hist(zo.data$EVT\_rich) # normalish, with many breaks



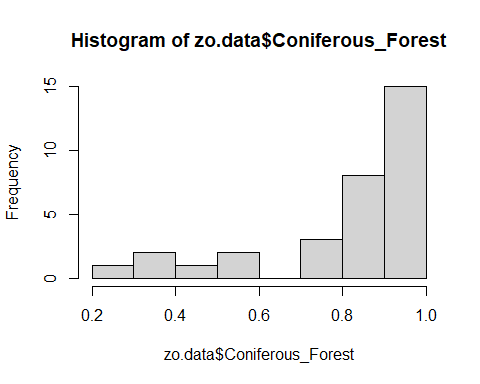
hist(zo.data$EVT\_J) # normalish



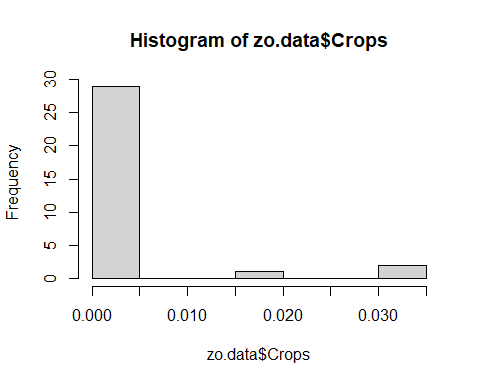
hist(zo.data$Barren) # VERY right skewed



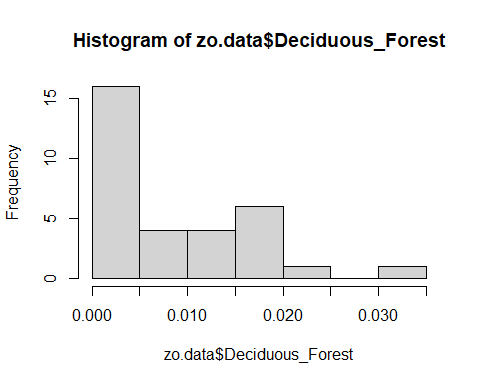
hist(zo.data$Coniferous\_Forest) # VERY left skewed



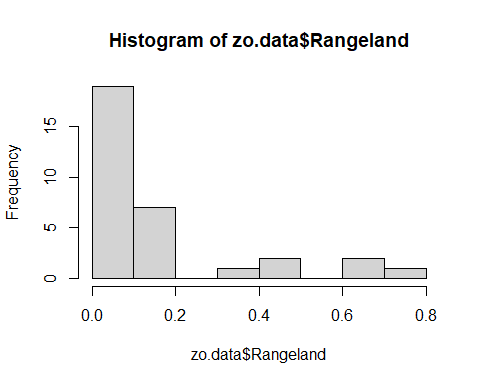
hist(zo.data$Crops) # VERY right skewed



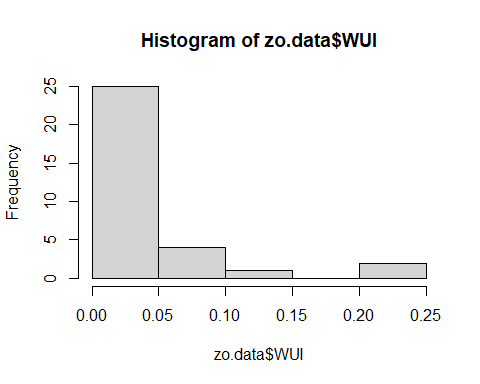
hist(zo.data$Deciduous\_Forest) # VERY right skewed



hist(zo.data$Rangeland) # VERY right skewed



hist(zo.data$WUI) # VERY right skewed

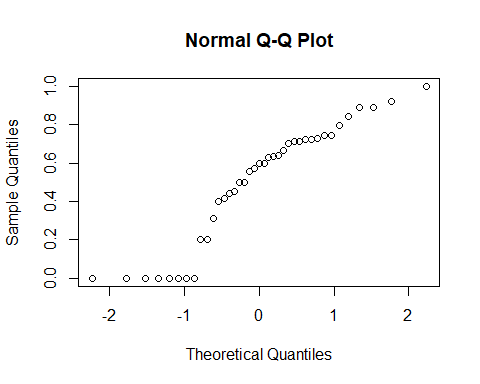


# Normality

## O. lignaria

### All sites

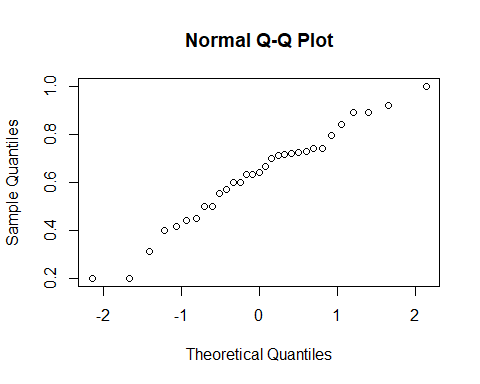
# Ratios  
qqnorm(data$oligRatio) # not quite normal



shapiro.test(data$oligRatio) # technically not sig (p = 0.07)

##   
## Shapiro-Wilk normality test  
##   
## data: data$oligRatio  
## W = 0.89586, p-value = 0.001683

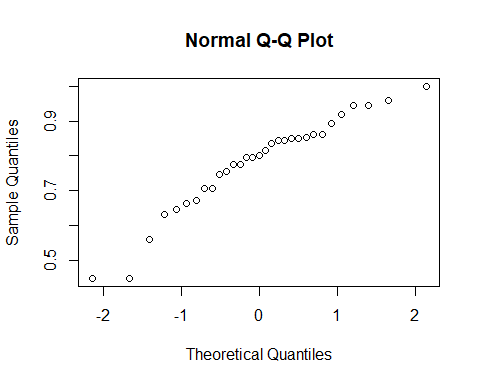
## O.lig ratio > 0 sites  
qqnorm(or.data$oligRatio) # fairly normal



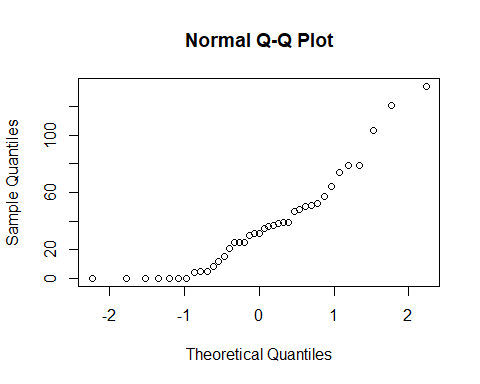
shapiro.test(or.data$oligRatio) # not sig

##   
## Shapiro-Wilk normality test  
##   
## data: or.data$oligRatio  
## W = 0.97229, p-value = 0.584

qqnorm(sqrt(or.data$oligRatio)) # more normal



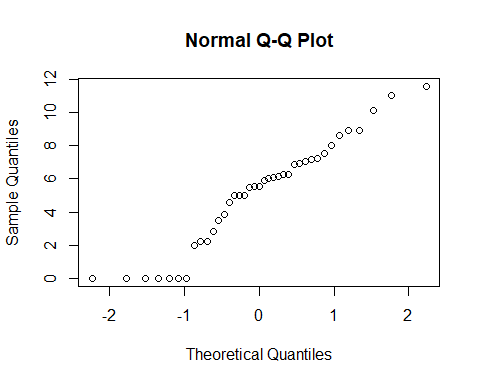
# counts  
qqnorm(data$nTotProvised) # not quite normal



shapiro.test(data$nTotProvised) # sig - not normal

##   
## Shapiro-Wilk normality test  
##   
## data: data$nTotProvised  
## W = 0.88918, p-value = 0.001096

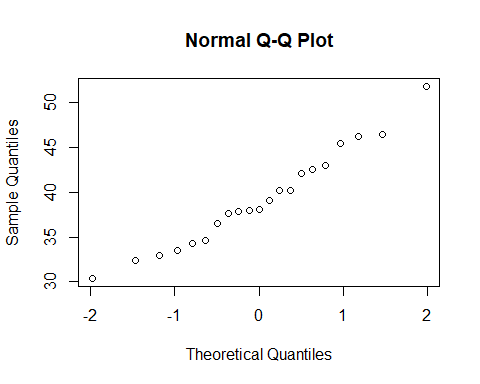
qqnorm(sqrt(data$nTotProvised)) # pretty normal



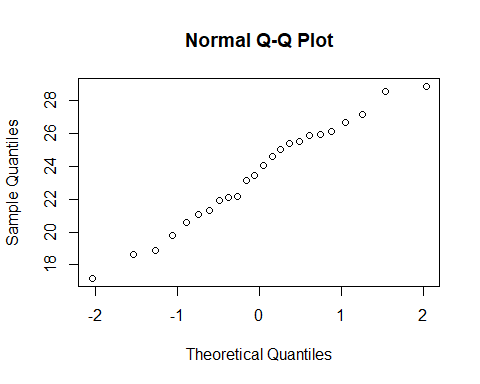
shapiro.test((sqrt(data$nTotProvised))) # not sig - normal

##   
## Shapiro-Wilk normality test  
##   
## data: (sqrt(data$nTotProvised))  
## W = 0.94279, p-value = 0.0471

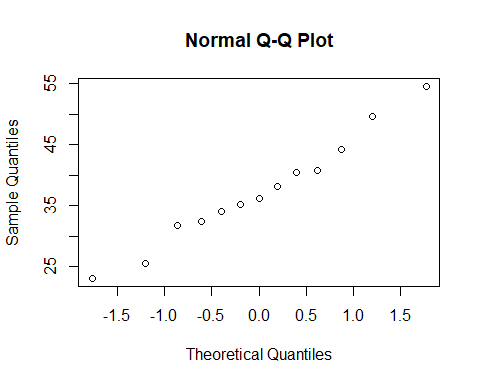
# weights  
## 2020  
qqnorm(fw.data$avgOligFMass\_mg) # too few obs



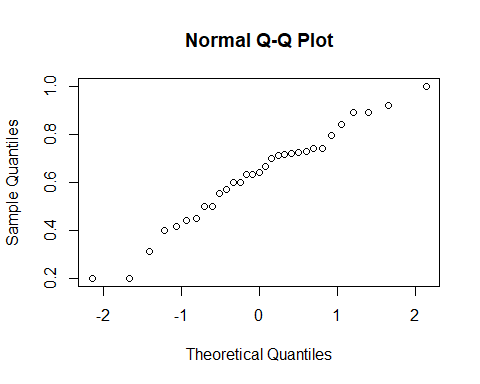
qqnorm(mw.data$avgOligMMass\_mg) # normal



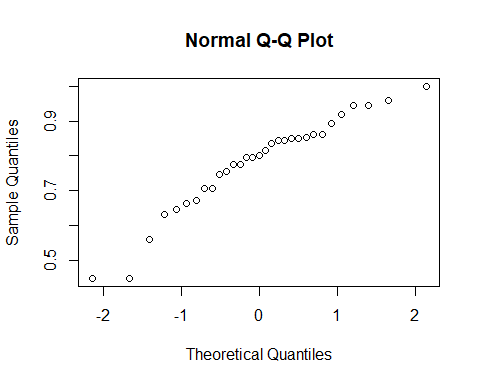
qqnorm(lw.data$avgOligLarvMass\_mg) # normal

 ### O.lig ratio > 0 Sites

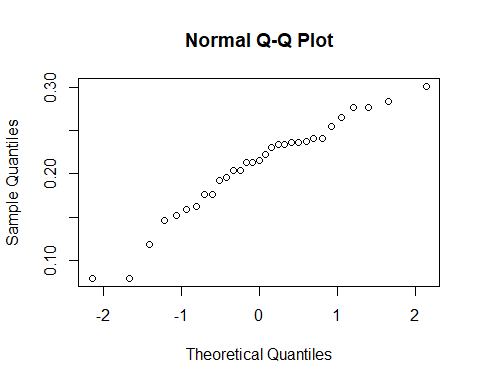
# Ratios  
qqnorm(or.data$oligRatio) # fairly normal



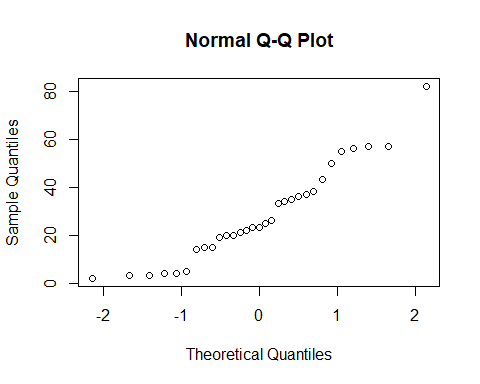
qqnorm(sqrt(or.data$oligRatio)) # more normal



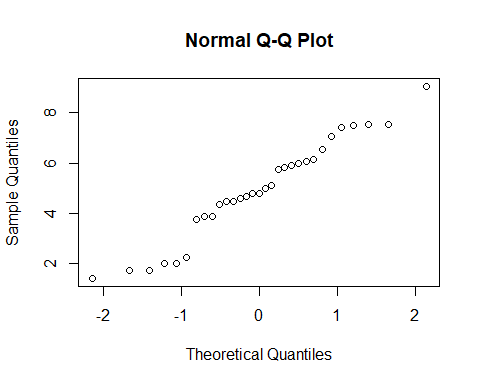
qqnorm(log(or.data$oligRatio+1, base=10)) # sort of normal



## Counts  
qqnorm(or.data$nOligGrowth) # wonky



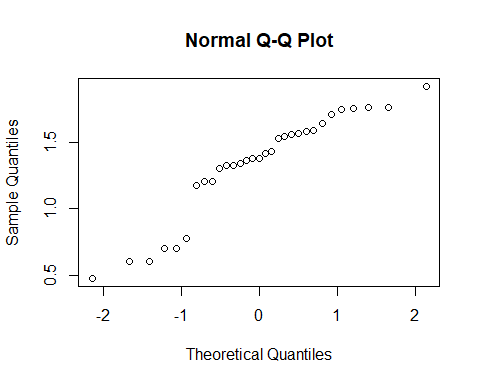
qqnorm(sqrt(or.data$nOligGrowth)) # sort of normal?



shapiro.test(sqrt(or.data$nOligGrowth)) # not sig

##   
## Shapiro-Wilk normality test  
##   
## data: sqrt(or.data$nOligGrowth)  
## W = 0.95984, p-value = 0.2889

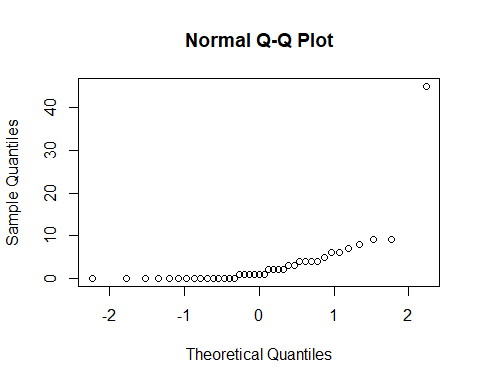
qqnorm(log(or.data$nOligGrowth +1, base=10)) # eehhhh



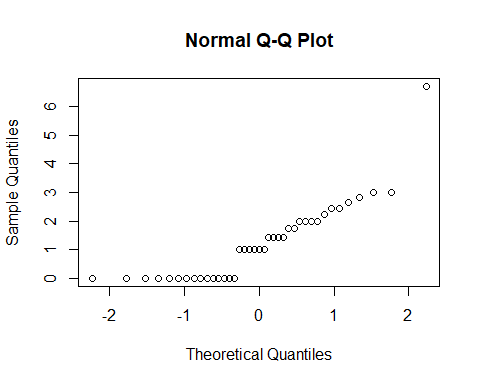
## T. stansburyi data

### All sites

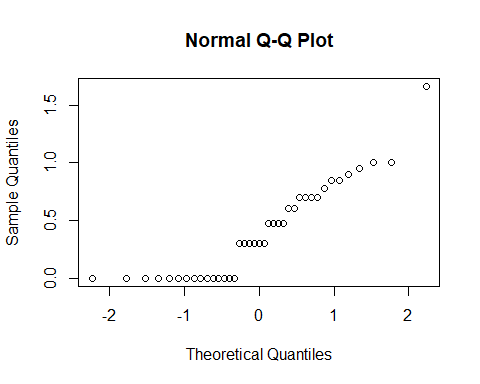
## with outlier  
qqnorm(data$nTstanCells) # not normal - outlier and zero heavy



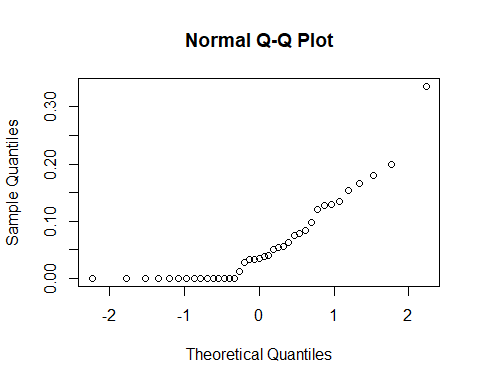
qqnorm(sqrt(data$nTstanCells)) # still zero heavy



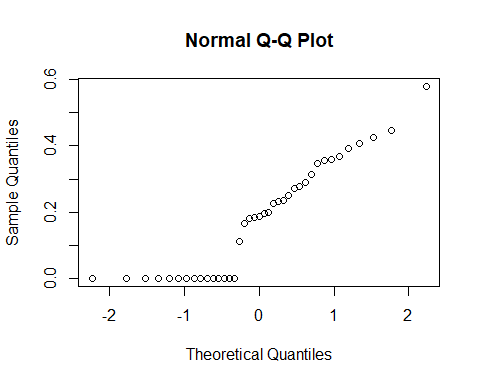
qqnorm(log(data$nTstanCells+1,base = 10)) # still zero heavy



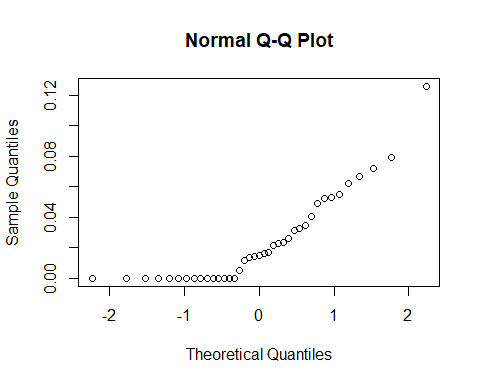
qqnorm(data$tstanRatio) # not normal - outlier and zero heavy



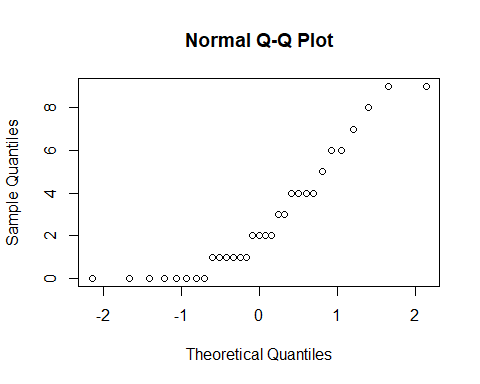
qqnorm(sqrt(data$tstanRatio)) # still zero heavy



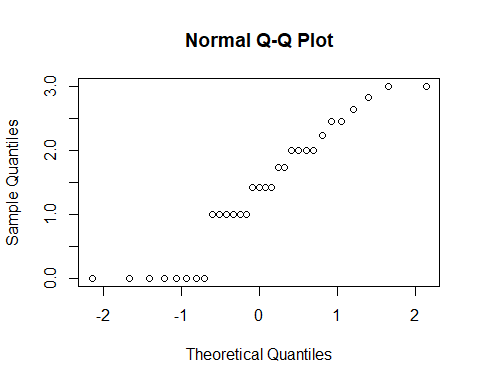
qqnorm(log(data$tstanRatio+1,base = 10)) # still zero heavy



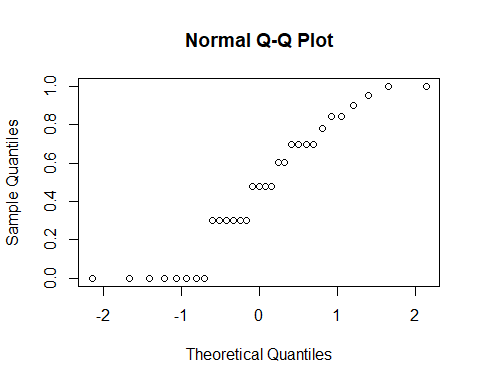
## without outlier  
qqnorm(zo.data.or$nTstanCells) # not normal - outlier and zero heavy



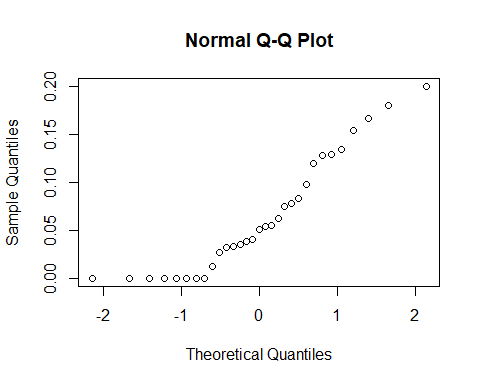
qqnorm(sqrt(zo.data.or$nTstanCells)) # still zero heavy



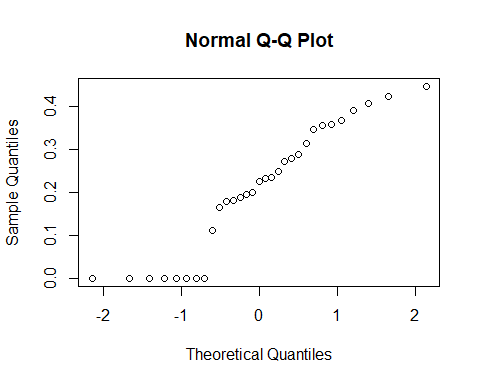
qqnorm(log(zo.data.or$nTstanCells+1,base = 10)) # still zero heavy



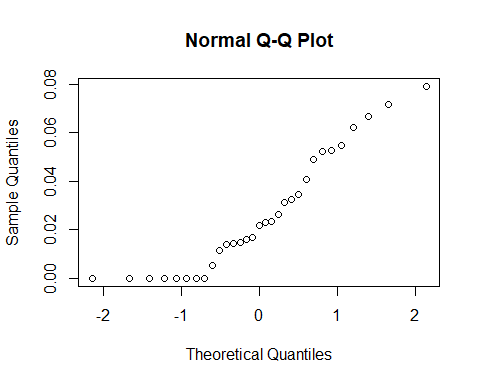
qqnorm(zo.data.or$tstanRatio) # not normal - outlier and zero heavy



qqnorm(sqrt(zo.data.or$tstanRatio)) # still zero heavy

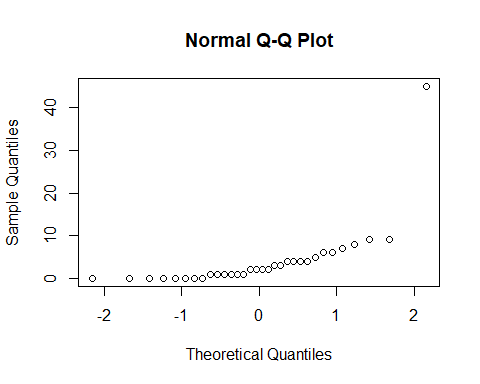


qqnorm(log(zo.data.or$tstanRatio+1,base = 10)) # still zero heavy

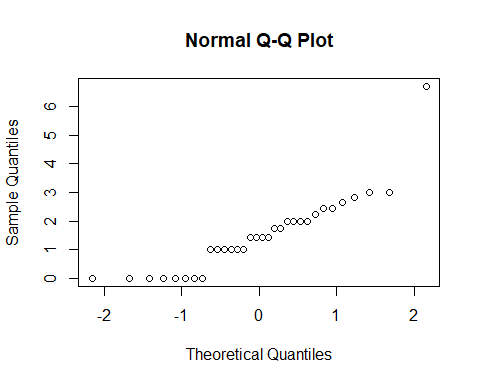


### Provisioned sites

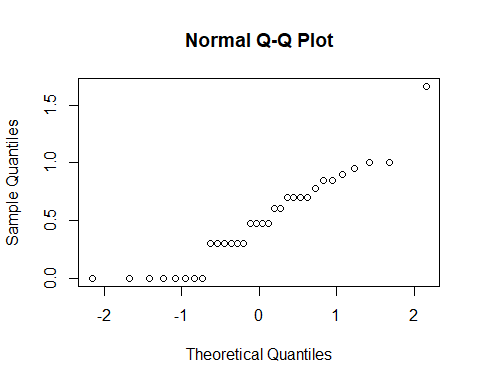
qqnorm(zo.data$nTstanCells) # huge outlier



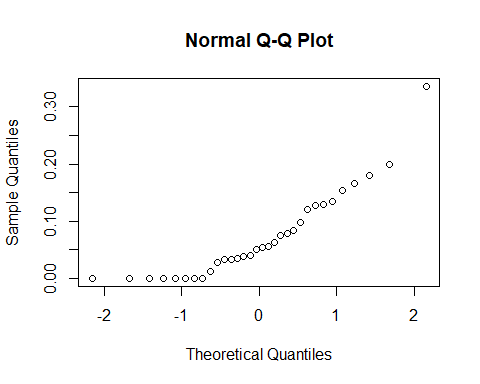
qqnorm(sqrt(zo.data$nTstanCells)) # huge outlier



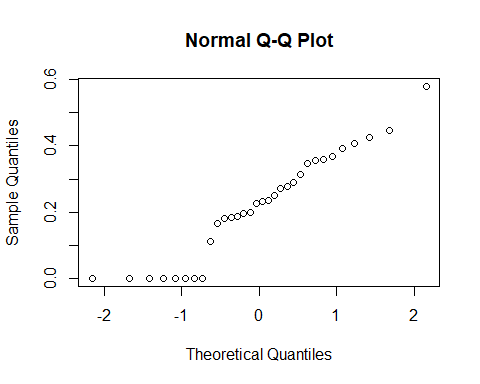
qqnorm(log(zo.data$nTstanCells+1,base = 10)) # still zero heavy



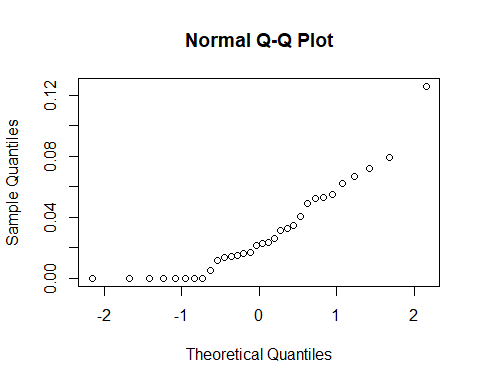
qqnorm(zo.data$tstanRatio) # not normal - outlier and zero heavy



qqnorm(sqrt(zo.data$tstanRatio)) # still zero heavy



qqnorm(log(zo.data$tstanRatio+1,base = 10)) # still zero heavy

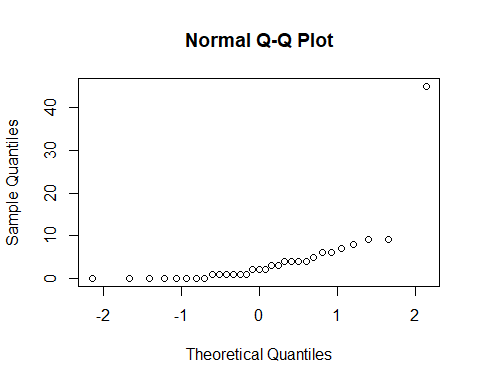


shapiro.test(log(zo.data$tstanRatio+1,base = 10)) # sig

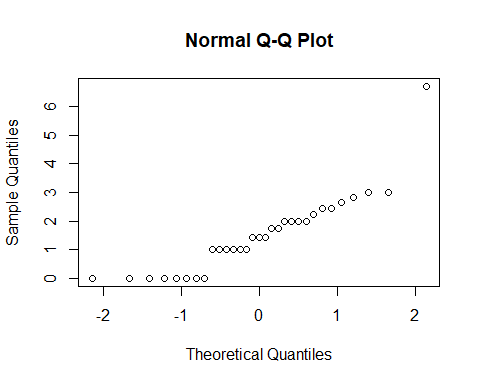
##   
## Shapiro-Wilk normality test  
##   
## data: log(zo.data$tstanRatio + 1, base = 10)  
## W = 0.87097, p-value = 0.001223

### O.lig ratio > 0 Sites

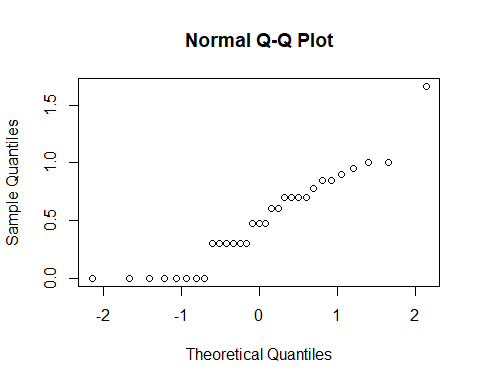
# Ratios  
qqnorm(or.data$nTstanCells) # huge outlier



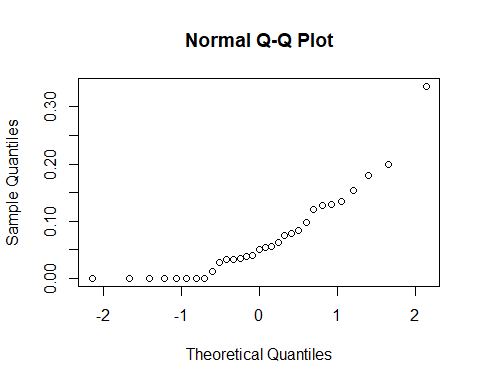
qqnorm(sqrt(or.data$nTstanCells)) # huge outlier



qqnorm(log(or.data$nTstanCells+1,base = 10)) # still zero heavy



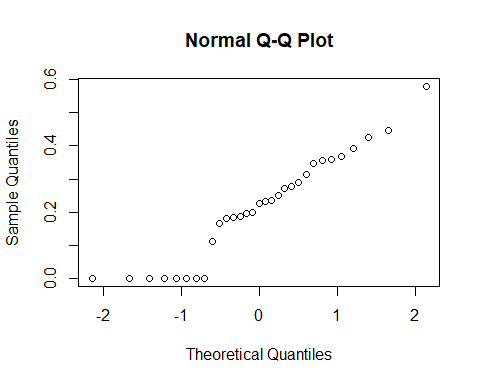
qqnorm(or.data$tstanRatio) # not normal - outlier and zero heavy



shapiro.test(or.data$tstanRatio) # sig

##   
## Shapiro-Wilk normality test  
##   
## data: or.data$tstanRatio  
## W = 0.8326, p-value = 0.0002205

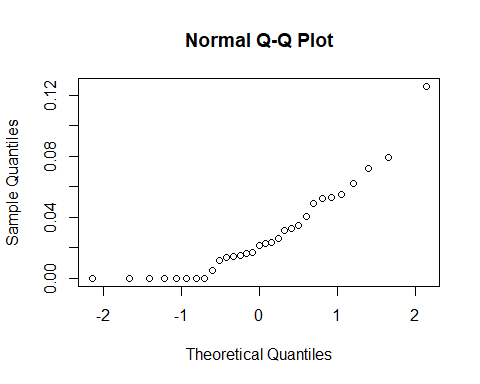
qqnorm(sqrt(or.data$tstanRatio)) # still zero heavy



shapiro.test(sqrt(or.data$tstanRatio)) # not sig

##   
## Shapiro-Wilk normality test  
##   
## data: sqrt(or.data$tstanRatio)  
## W = 0.9295, p-value = 0.04248

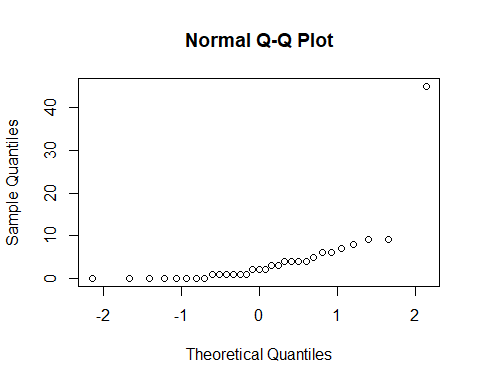
qqnorm(log(or.data$tstanRatio+1,base = 10)) # still zero heavy



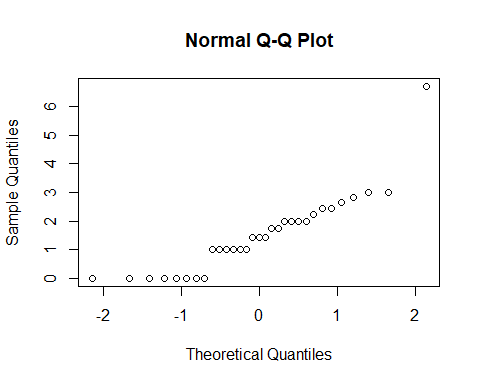
shapiro.test(log(or.data$tstanRatio+1,base = 10)) # sig

##   
## Shapiro-Wilk normality test  
##   
## data: log(or.data$tstanRatio + 1, base = 10)  
## W = 0.85721, p-value = 0.0007245

## Counts  
qqnorm(or.data$nTstanCells) # wonky



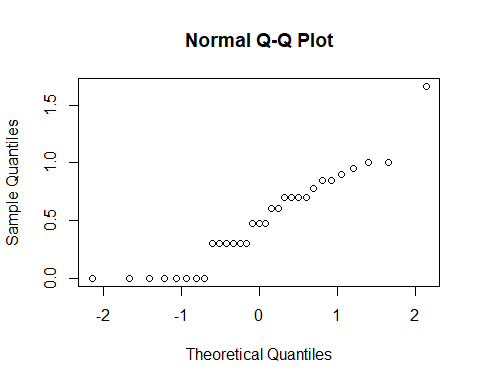
qqnorm(sqrt(or.data$nTstanCells)) # wonky



shapiro.test(sqrt(or.data$nTstanCells)) # sig

##   
## Shapiro-Wilk normality test  
##   
## data: sqrt(or.data$nTstanCells)  
## W = 0.83742, p-value = 0.0002764

qqnorm(log(or.data$nTstanCells +1, base=10)) # zero heavy, then normal?

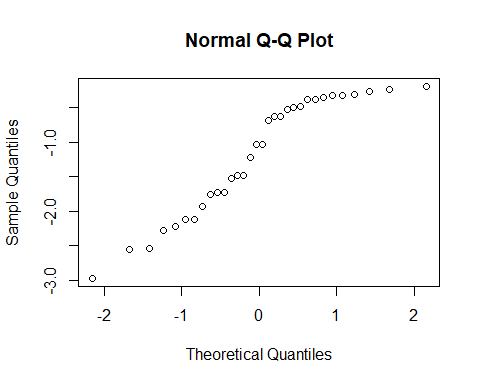


shapiro.test(log(or.data$nTstanCells+1, base=10)) # p = 0.05.

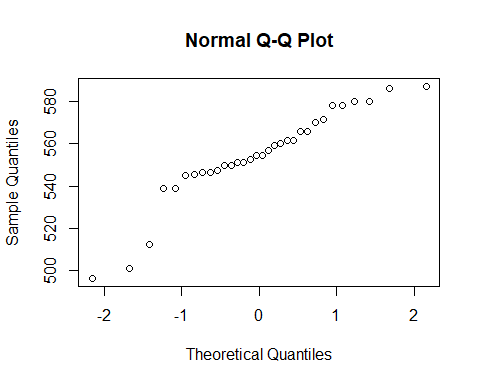
##   
## Shapiro-Wilk normality test  
##   
## data: log(or.data$nTstanCells + 1, base = 10)  
## W = 0.91201, p-value = 0.01457

## Climate variables

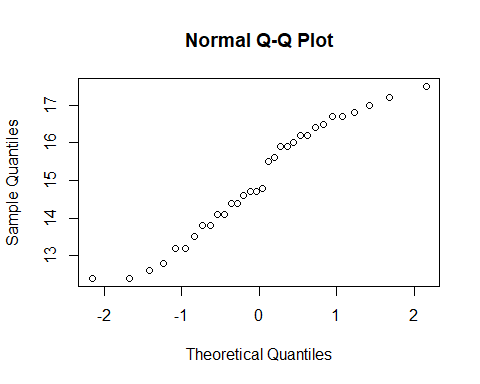
qqnorm(zo.data$HLI)# zero heavy



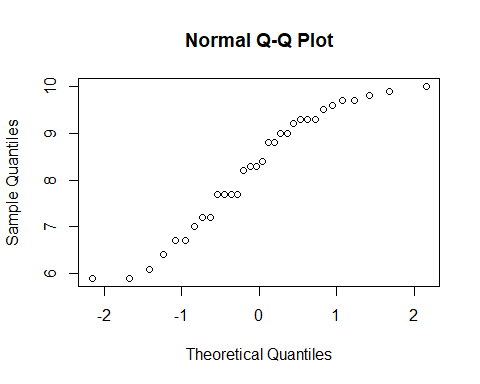
qqnorm(zo.data$Ann\_ppt\_mm) # left skewed



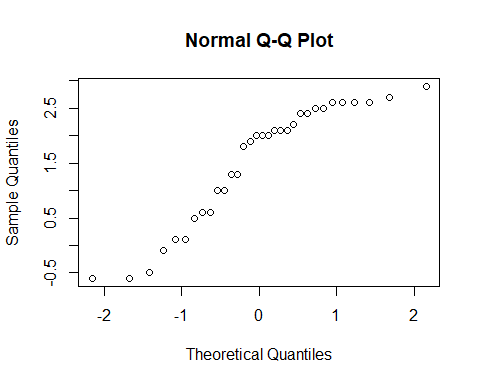
qqnorm(zo.data$Ann\_tmax\_C) # left skewed



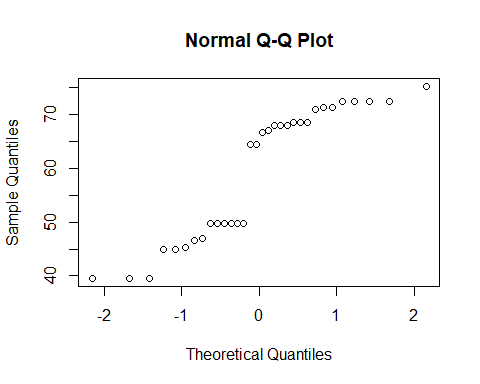
qqnorm(zo.data$Ann\_tmean\_C) # normalish



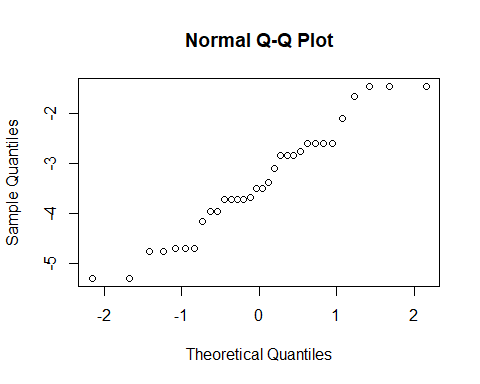
qqnorm(zo.data$Ann\_tmin\_C) # slightly right skewed



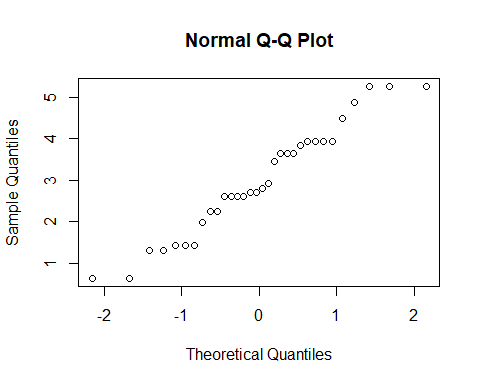
qqnorm(zo.data$bi\_ppt\_mm) # bi-modal



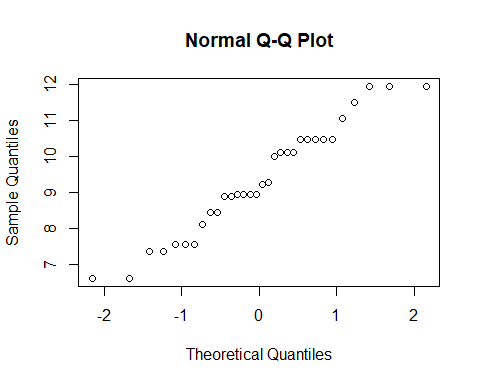
qqnorm(zo.data$bi\_tmin\_C) # normalish?



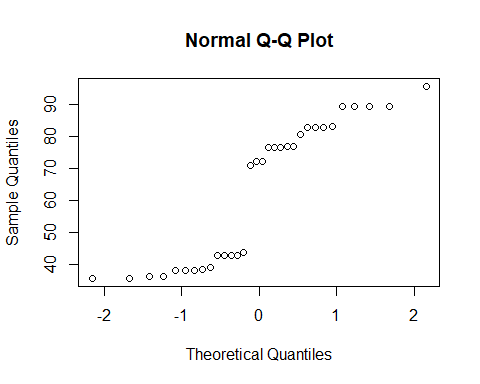
qqnorm(zo.data$bi\_tmean\_C) # normalish



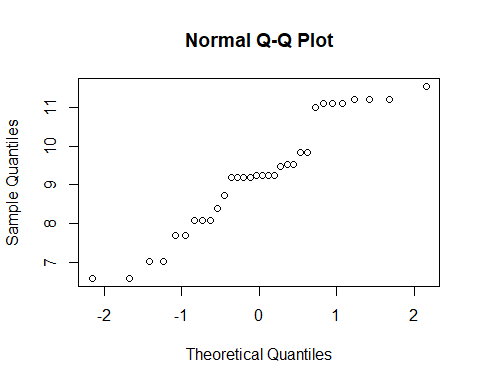
qqnorm(zo.data$bi\_tmax\_C) # normalish



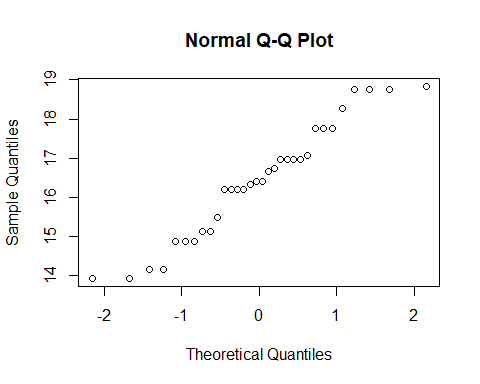
qqnorm(zo.data$qrt\_ppt\_mm) # bi-modal



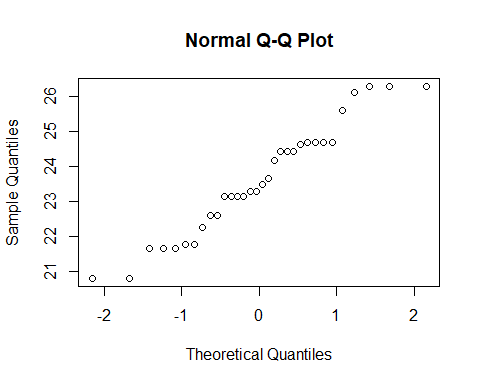
qqnorm(zo.data$qrt\_tmin\_C) # normalish



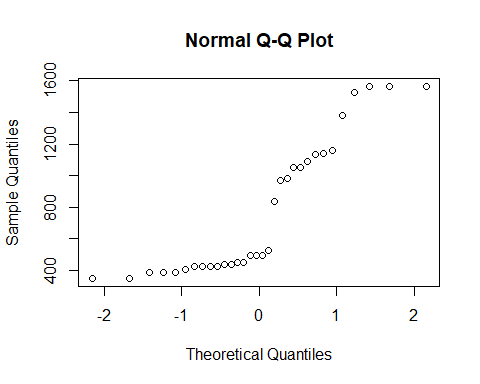
qqnorm(zo.data$qrt\_tmean\_C) # normalish



qqnorm(zo.data$qrt\_tmax\_C) # normalish

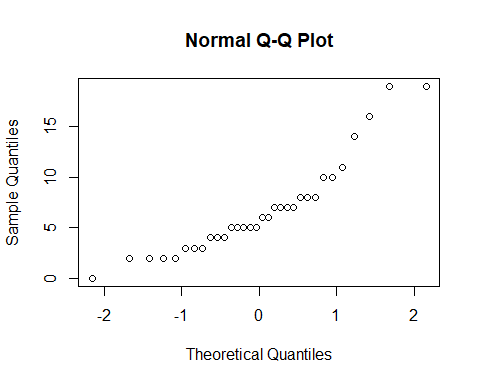


qqnorm(zo.data$DD\_accumulated) # bi-modal

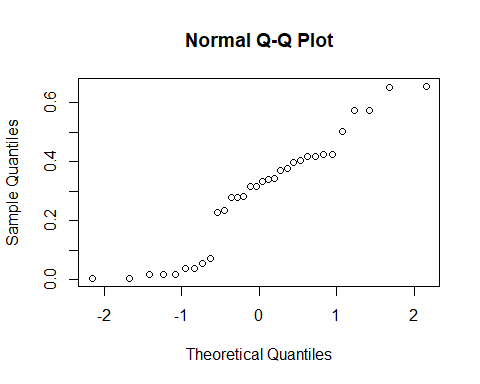


## Floral & Tree variables

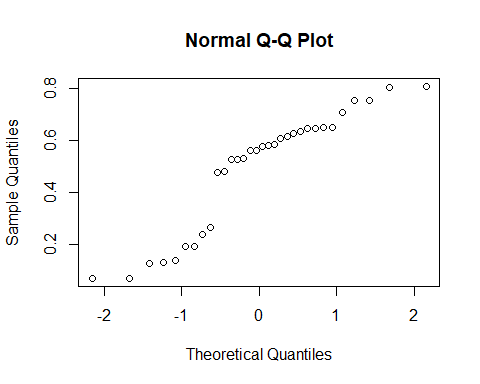
# stemrich  
qqnorm(zo.data$stemrich) # fairly normal



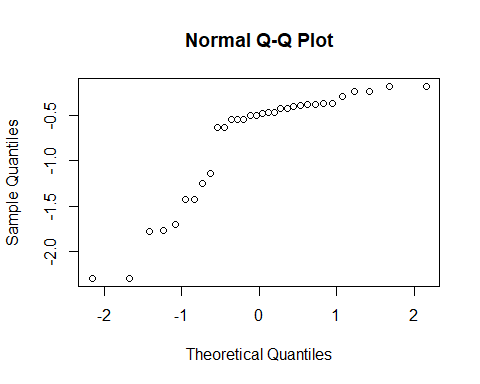
# TCC  
qqnorm(zo.data$meanTreeCC) # zero heavy - burned sites



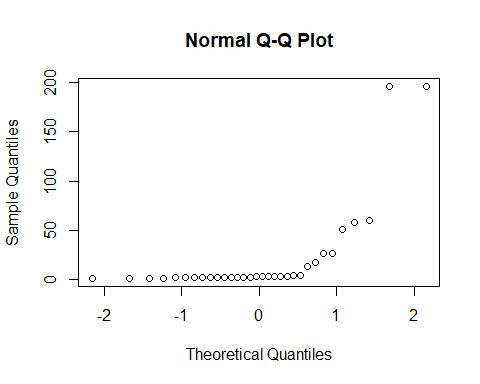
qqnorm(sqrt(zo.data$meanTreeCC)) # still zero heavy



qqnorm(log(zo.data$meanTreeCC, base = 10)) # Left skewed



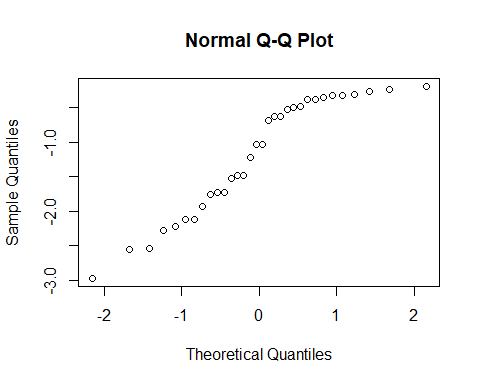
qqnorm(1/(zo.data$meanTreeCC)) # right skewed



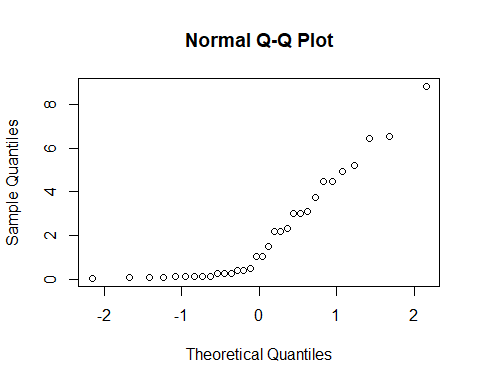
## Landscape variables

Provisioned Sites

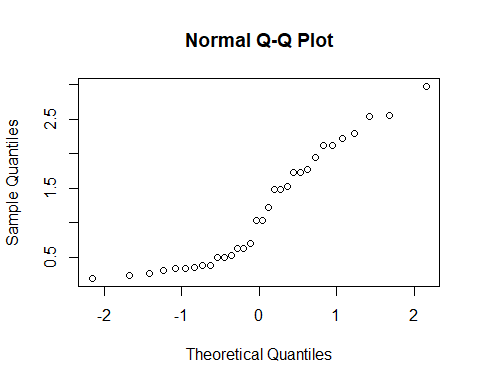
# HLI  
qqnorm(zo.data$HLI) # left skewed - hot and dry sites?



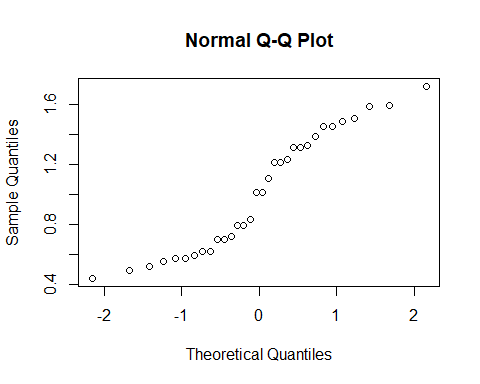
# don't know how to transform  
qqnorm(zo.data$HLI^2) # left skewed



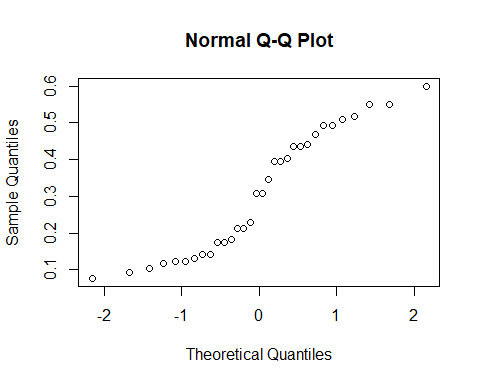
qqnorm(abs(zo.data$HLI)) # zero heavy



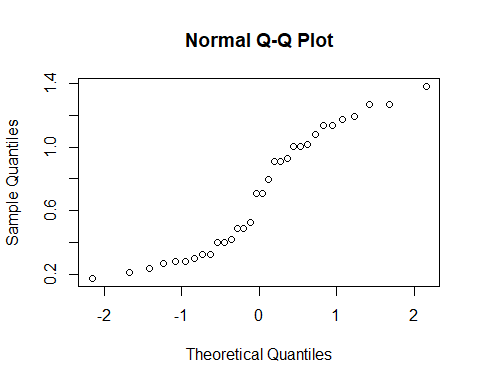
qqnorm(sqrt(abs(zo.data$HLI))) # bimodal



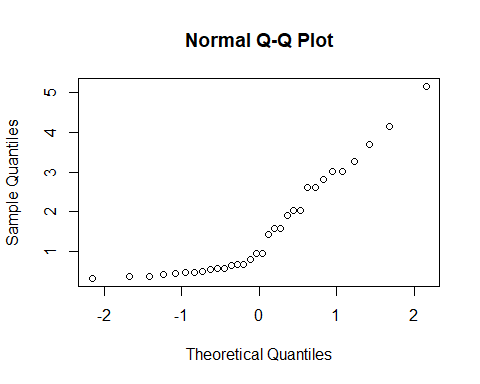
qqnorm(log(abs(zo.data$HLI)+1,base=10))



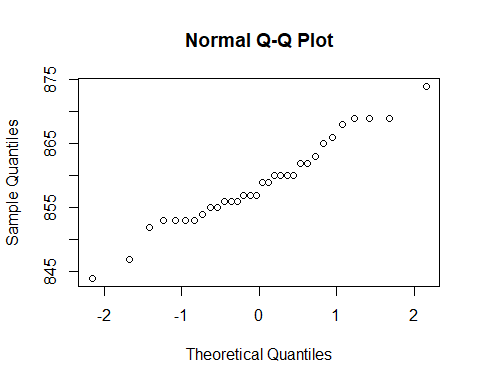
qqnorm(log(abs(zo.data$HLI)+1)) # still left skewed, but not as bad



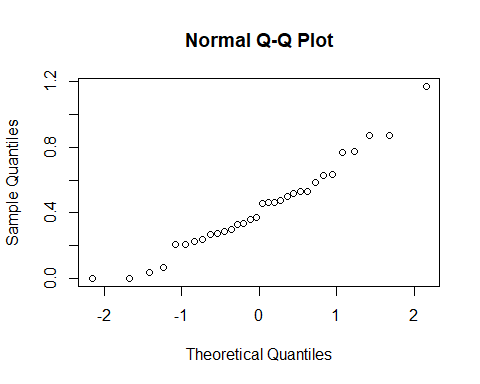
qqnorm(1/(abs(zo.data$HLI)))



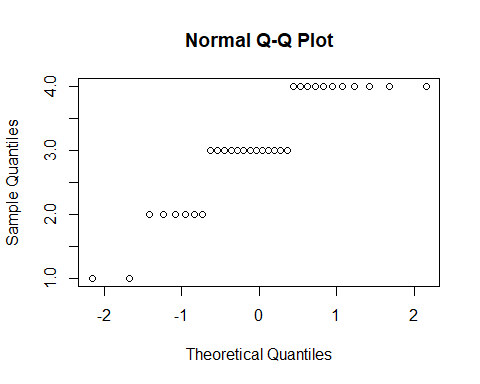
# EVT  
qqnorm(zo.data$EVT\_abund) # normalish



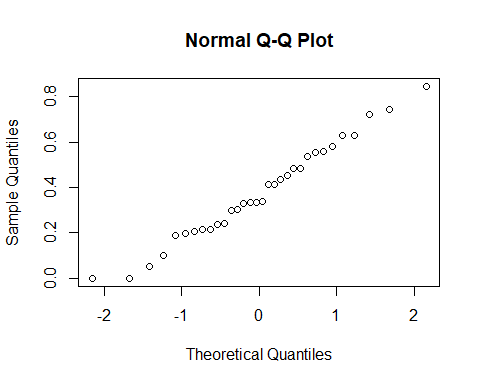
qqnorm(zo.data$EVT\_H) # sort of right skewed



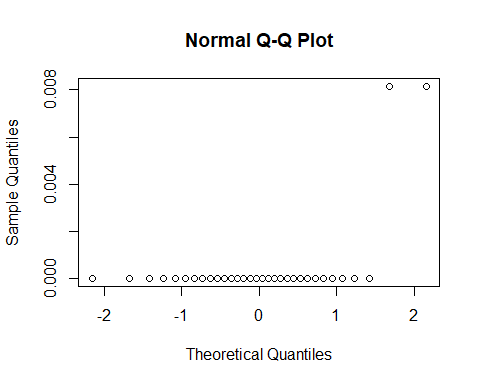
qqnorm(zo.data$EVT\_rich) # normalish, with many breaks



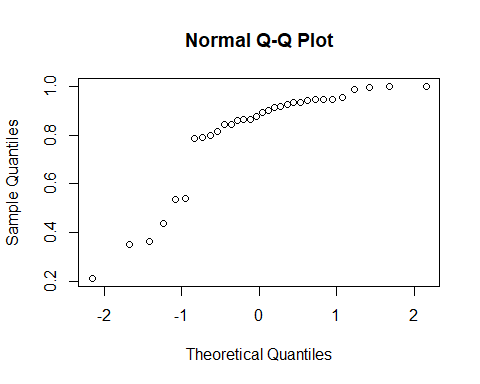
qqnorm(zo.data$EVT\_J) # normalish



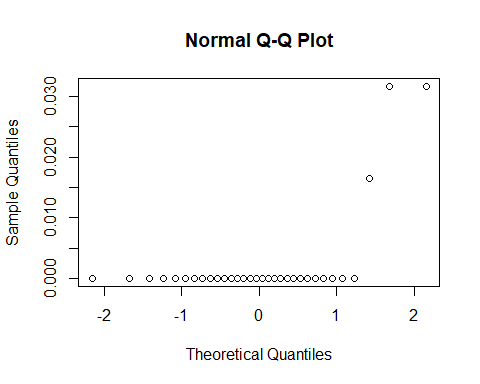
qqnorm(zo.data$Barren) # VERY right skewed



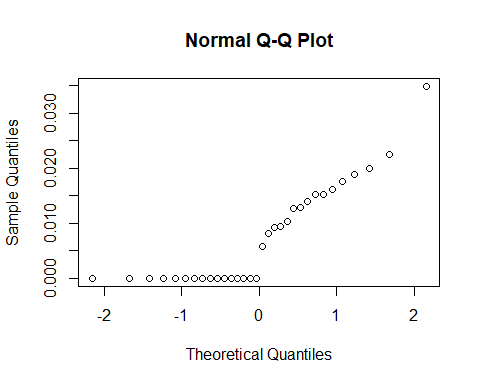
qqnorm(zo.data$Coniferous\_Forest) # VERY left skewed



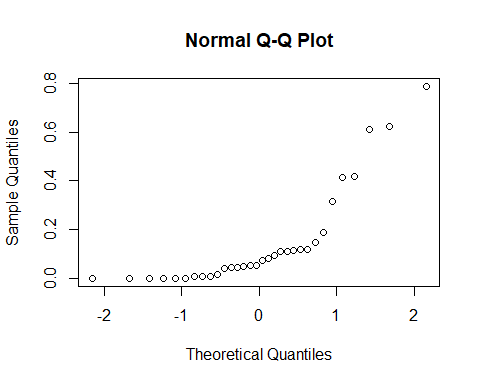
qqnorm(zo.data$Crops) # VERY right skewed



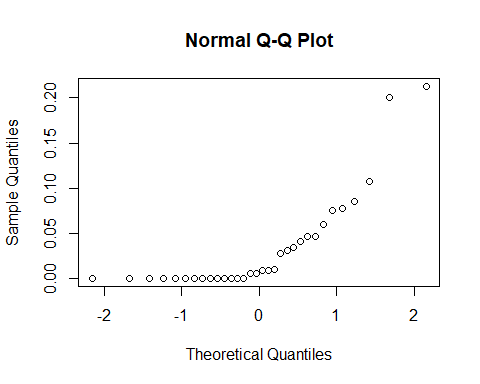
qqnorm(zo.data$Deciduous\_Forest) # VERY right skewed



qqnorm(zo.data$Rangeland) # VERY right skewed

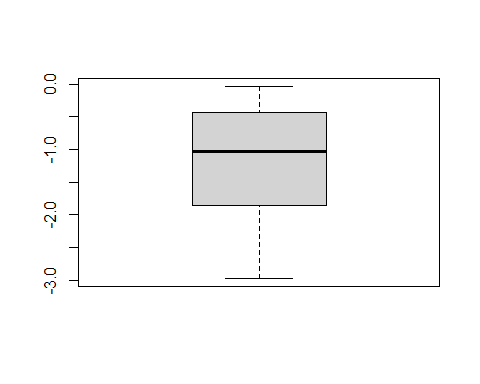


qqnorm(zo.data$WUI) # VERY right skewed

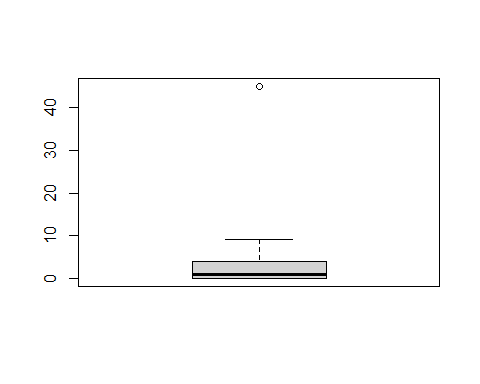


# Boxplots

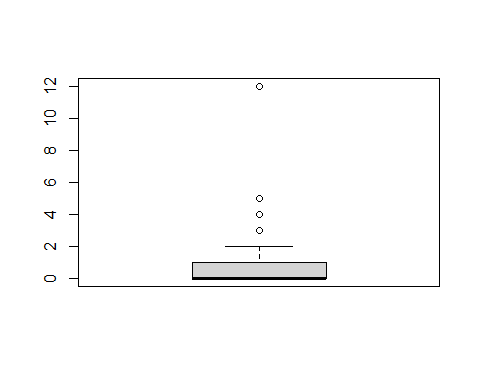
boxplot(data$HLI)



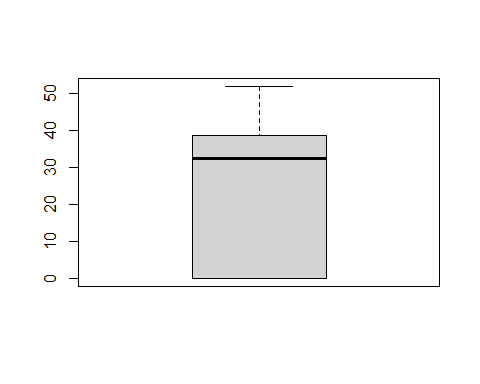
boxplot(data$nTstanCells) # HUGE outlier



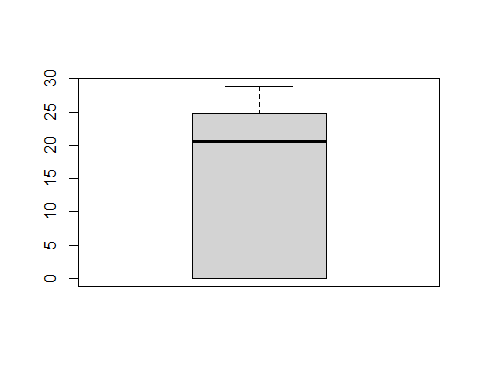
boxplot(data$nParasite)



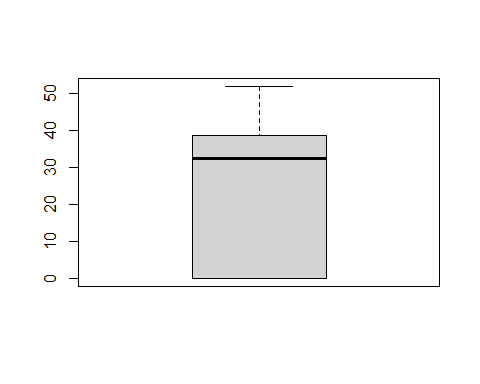
boxplot(data$avgOligFMass\_mg)



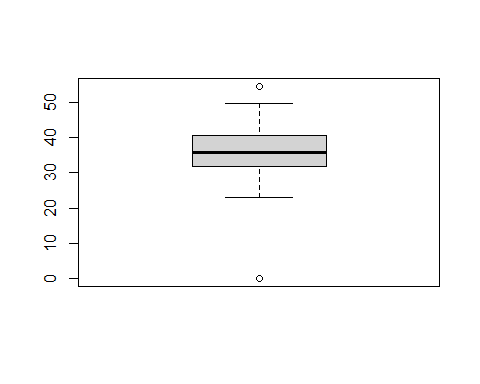
boxplot(data$avgOligMMass\_mg)



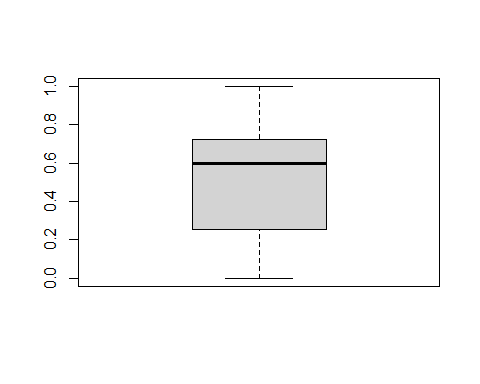
boxplot(data$avgOligFMass\_mg)



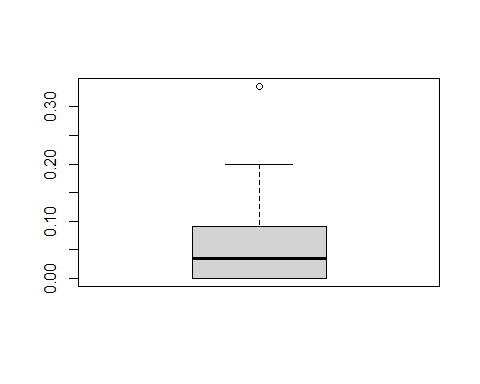
boxplot(data$avgOligLarvMass\_mg) # outlier, but probably more useful than pupae mass



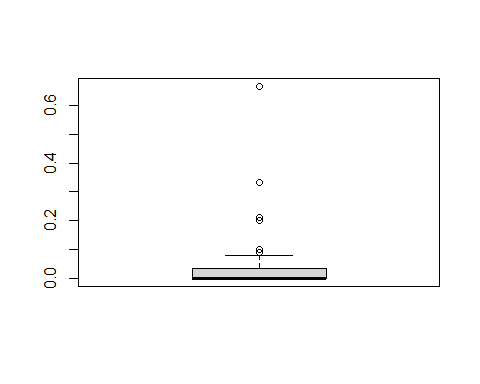
boxplot(data$oligRatio) # outliers (no bees)



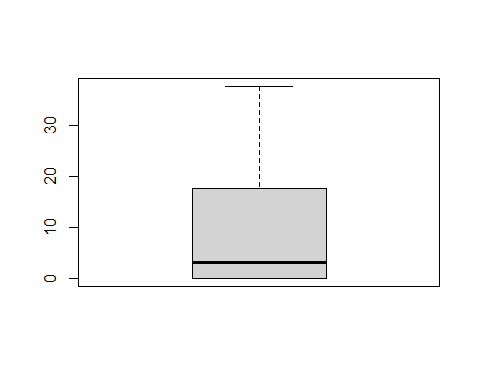
boxplot(data$tstanRatio) # outlier



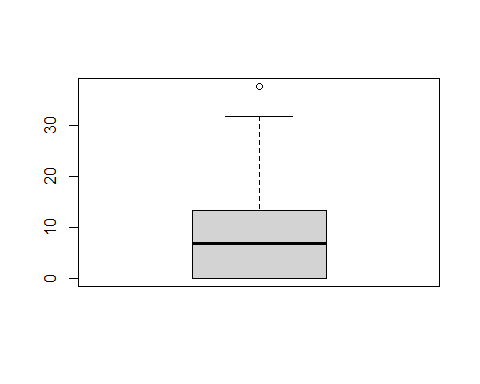
boxplot(data$monoRatio)



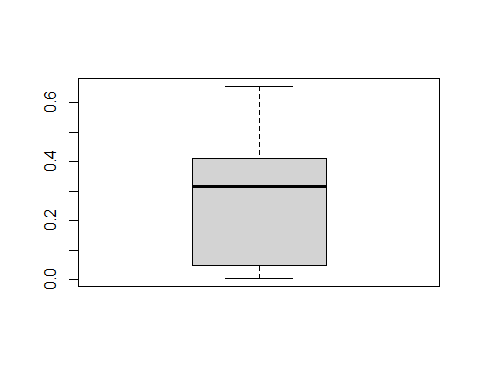
boxplot(data$Dead\_BA\_ha)



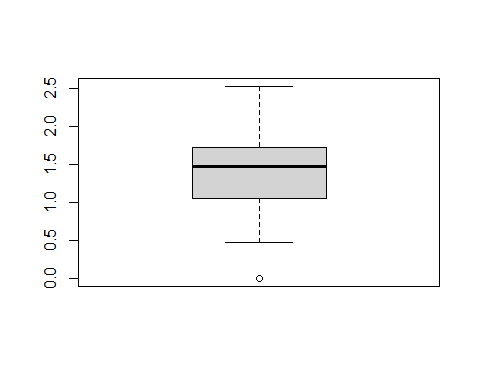
boxplot(data$Live\_BA\_ha)



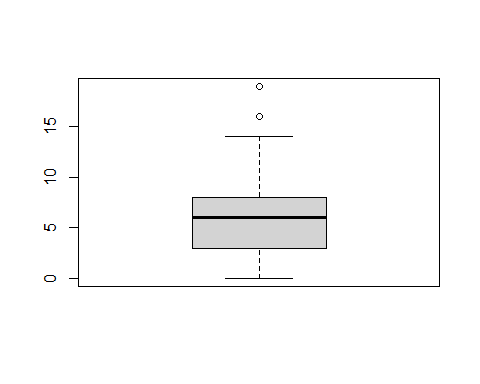
boxplot(data$meanTreeCC)



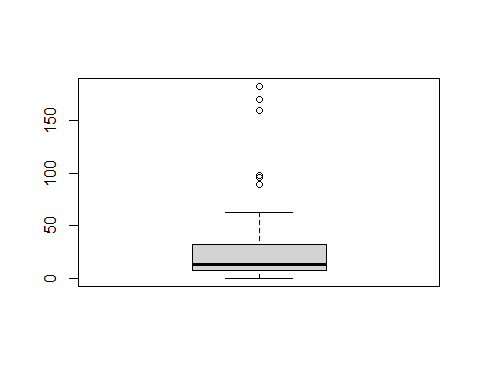
boxplot(data$stemdiv)



boxplot(data$stemrich)

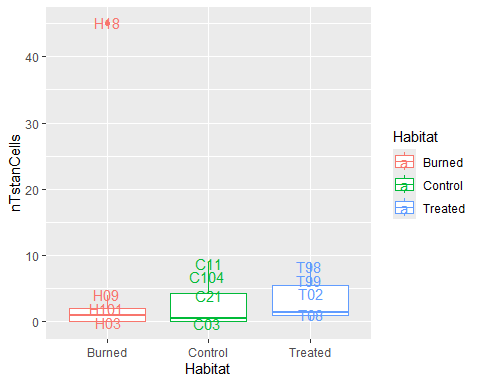


boxplot(data$stemabun)

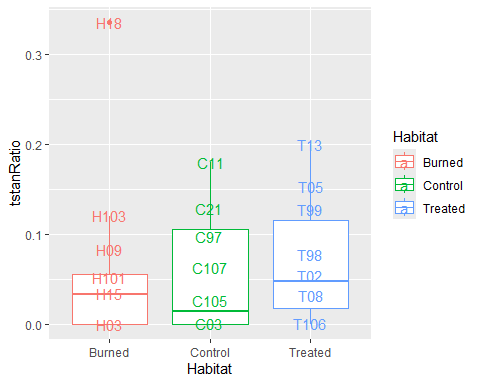


# Identifying outliers

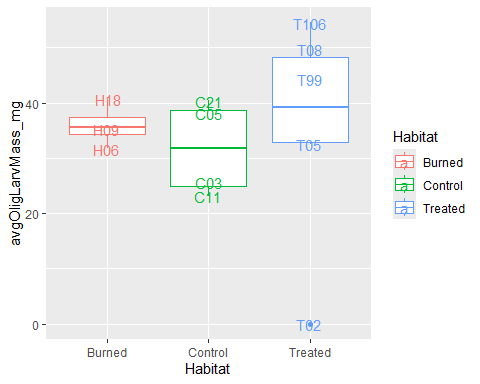
# Outliers:  
## nTstan/TstanRatio  
  
ggplot(data=data, aes(x=Habitat, y=nTstanCells, label = Site, color=Habitat))+  
 geom\_boxplot()+  
 geom\_text(check\_overlap = TRUE, position=position\_jitter(width=0.01))+  
 theme(legend.position="right") # H18 - TON of T. stan



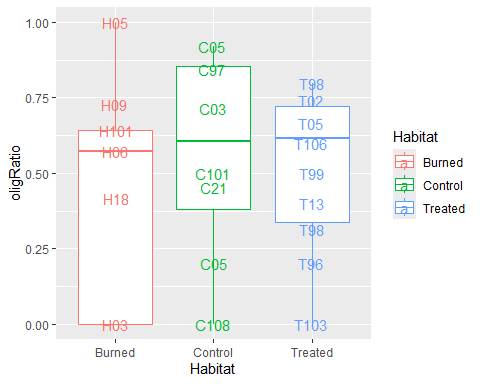
ggplot(data=data, aes(x=Habitat, y=tstanRatio, label = Site, color=Habitat))+  
 geom\_boxplot()+  
 geom\_text(check\_overlap = TRUE, position=position\_jitter(width=0.01))+  
 theme(legend.position="right") # H18 not as extreme



ggplot(data=data, aes(x=Habitat, y=avgOligLarvMass\_mg, label = Site, color=Habitat))+  
 geom\_boxplot()+  
 geom\_text(check\_overlap = TRUE, position=position\_jitter(width=0.01))+  
 theme(legend.position="right") # 2 sites with no larvae (H03 & T02)



ggplot(data=data, aes(x=Habitat, y=oligRatio, label = Site, color=Habitat))+  
 geom\_boxplot()+  
 geom\_text(check\_overlap = TRUE, position=position\_jitter(width=0.01))+  
 theme(legend.position="right") # H03 - no Olig



# Zero inflation?

## All sites

# count data (that you might be using)  
100\*sum(data$nTotProvised == 0)/nrow(data) # 17.9%

## [1] 17.94872

100\*sum(data$nTstanCells == 0)/nrow(data) # 38.5%

## [1] 38.46154

### not too bad, as far as zero inflation, goes  
  
# ratio data  
100\*sum(data$oligRatio == 0)/nrow(data) # 20.5%

## [1] 20.51282

100\*sum(data$oligFRatio == 0)/nrow(data) # 46.1% - 2020 probably skewing data

## [1] 46.15385

100\*sum(data$tstanRatio == 0)/nrow(data) # 38.5%

## [1] 38.46154

## Provisioned sites

# count data   
100\*sum(zo.data$nTstanCells == 0)/nrow(zo.data) # 25 %

## [1] 25

### not too bad, as far as zero inflation, goes  
  
# ratio data  
100\*sum(zo.data$oligRatio == 0)/nrow(zo.data) # 3.1%

## [1] 3.125

100\*sum(zo.data$oligFRatio == 0)/nrow(zo.data) # 34.4% - 2020 probably skewing data

## [1] 34.375

100\*sum(zo.data$tstanRatio == 0)/nrow(zo.data) # 25%

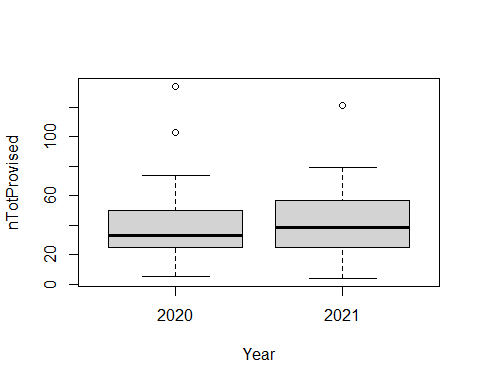
## [1] 25

# Difference between years

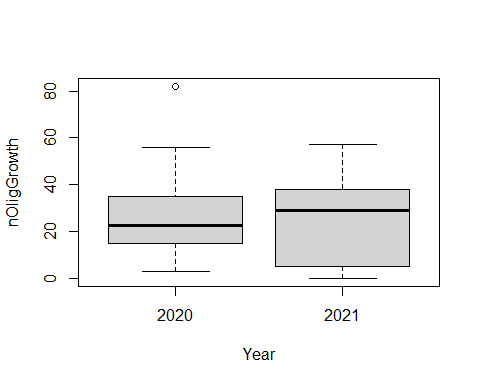
## O. lignaria

### Plots

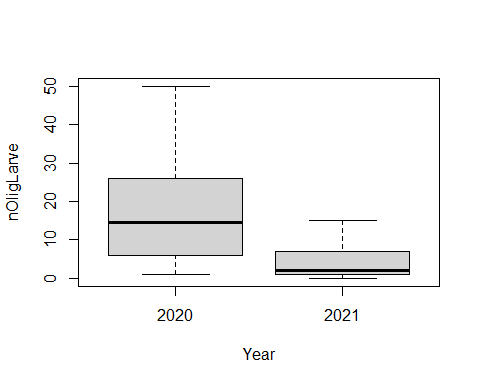
## Count  
boxplot(nTotProvised ~ Year, data = zo.data) # not that diff



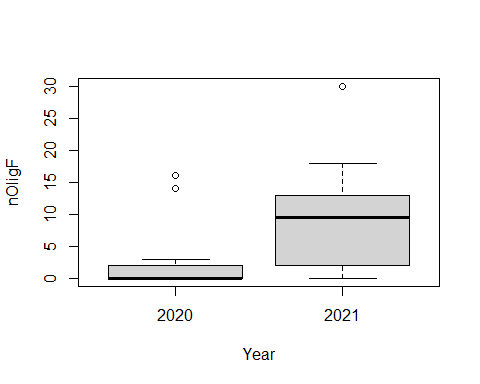
boxplot(nOligGrowth ~ Year, data = zo.data) # no diff



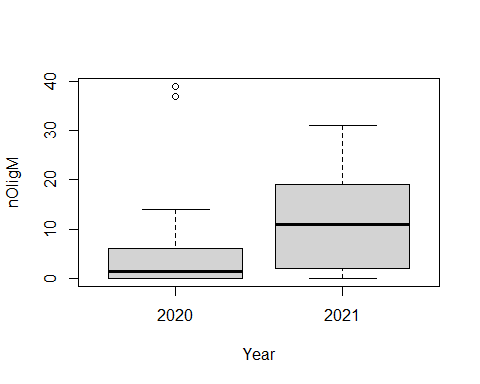
boxplot(nOligLarve ~ Year, data=zo.data) # much more in 2020



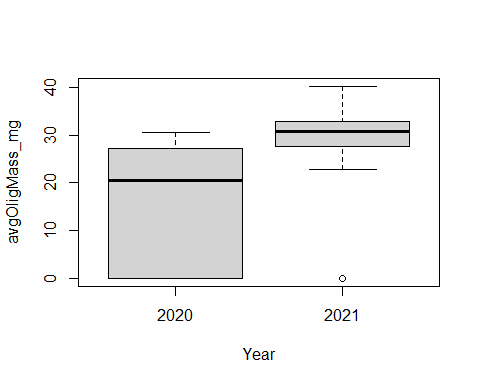
boxplot(nOligF ~ Year, data = zo.data) # much more in 2021



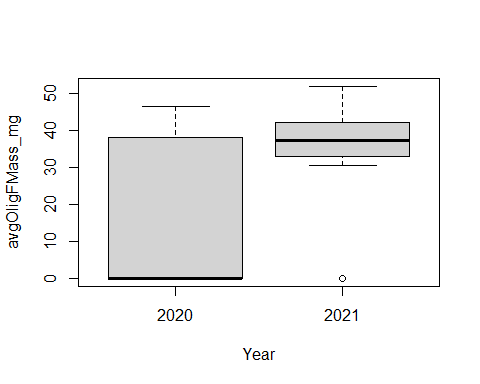
boxplot(nOligM ~ Year, data = zo.data) # much more in 2021



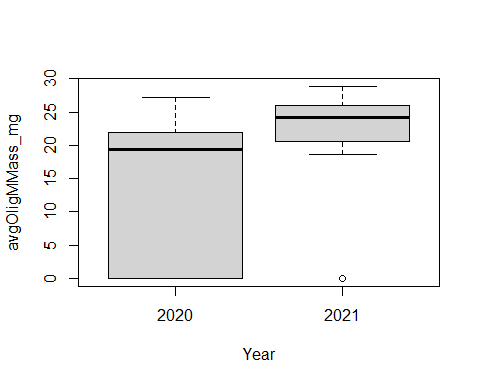
## Mass  
boxplot(avgOligMass\_mg ~ Year, data = zo.data) # higher in 2021



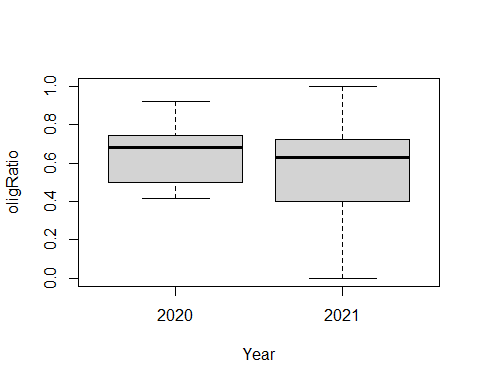
boxplot(avgOligFMass\_mg ~ Year, data = zo.data) # higher in 2021



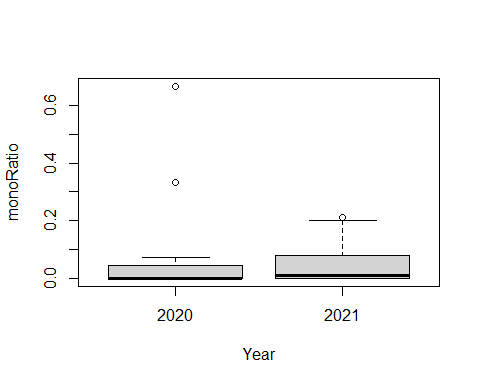
boxplot(avgOligMMass\_mg ~ Year, data = zo.data) # higher in 2021



## Ratio  
boxplot(oligRatio ~ Year, data = zo.data) # similar



boxplot(monoRatio ~ Year, data = zo.data) # similar

 Mass is probably significantly different - different collection methods

### Analyses

## Count  
bmod1<-aov(nTotProvised ~ Year, data = zo.data)   
bmod1<-tidy(bmod1) # no diff  
kable(bmod1, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 20.643 | 20.643 | 0.019 | 0.89 |
| Residuals | 30 | 31806.857 | 1060.229 | NA | NA |

bmod2<-aov(nOligGrowth ~ Year, data = zo.data)  
bmod2<-tidy(bmod2) # no diff  
kable(bmod2, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 29.774 | 29.774 | 0.072 | 0.79 |
| Residuals | 30 | 12415.944 | 413.865 | NA | NA |

bmod3<-aov(nOligLarve ~ Year, data=zo.data)   
bmod3<-tidy(bmod3) # very diff  
kable(bmod3, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 1353.722 | 1353.722 | 14.413 | 0.001 |
| Residuals | 30 | 2817.778 | 93.926 | NA | NA |

bmod4<-aov(nOligF ~ Year, data = zo.data)   
bmod4<-tidy(bmod4) # very diff  
kable(bmod4, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 366.012 | 366.012 | 8.208 | 0.008 |
| Residuals | 30 | 1337.706 | 44.590 | NA | NA |

bmod5<-aov(nOligM ~ Year, data = zo.data)   
bmod5<-tidy(bmod5) # no diff  
kable(bmod5, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 92.571 | 92.571 | 0.717 | 0.404 |
| Residuals | 30 | 3871.429 | 129.048 | NA | NA |

## Mass  
bmod6<-aov(avgOligMass\_mg ~ Year, data = zo.data)   
bmod6<-tidy(bmod6) # diff  
kable(bmod6, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 1500.114 | 1500.114 | 13.412 | 0.001 |
| Residuals | 30 | 3355.543 | 111.851 | NA | NA |

bmod7<-aov(avgOligFMass\_mg ~ Year, data = zo.data)   
bmod7<-tidy(bmod7) # diff  
kable(bmod7, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 2956.044 | 2956.044 | 10.158 | 0.003 |
| Residuals | 30 | 8730.568 | 291.019 | NA | NA |

bmod8<-aov(avgOligMMass\_mg ~ Year, data = zo.data)   
bmod8<-tidy(bmod8) # marginally diff (p = 0.09)  
kable(bmod8, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 327.232 | 327.232 | 3.049 | 0.091 |
| Residuals | 30 | 3219.265 | 107.309 | NA | NA |

## Ratio  
bmod9<-aov(oligRatio ~ Year, data = zo.data)   
bmod9<-tidy(bmod9) # no diff  
kable(bmod9, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.030 | 0.030 | 0.58 | 0.452 |
| Residuals | 30 | 1.534 | 0.051 | NA | NA |

bmod10<-aov(monoRatio ~ Year, data = zo.data)   
bmod10<-tidy(bmod10) # no diff  
kable(bmod10, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

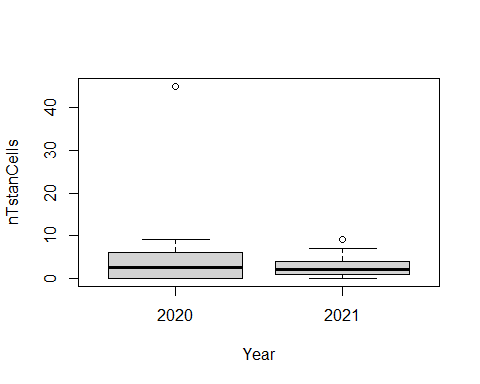
| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.009 | 0.009 | 0.48 | 0.494 |
| Residuals | 30 | 0.552 | 0.018 | NA | NA |

Significant differences: Larvae count, F count, brood mass, F mass, M mass is marginal

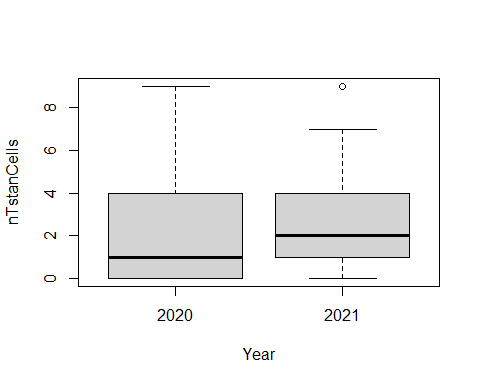
## T. stansburyi

### Plots

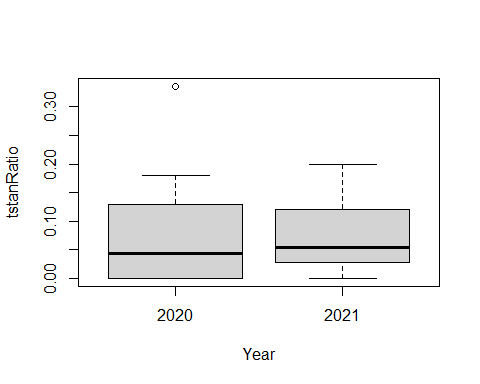
## Count  
boxplot(nTstanCells ~ Year, data = zo.data) # no diff, other than outlier



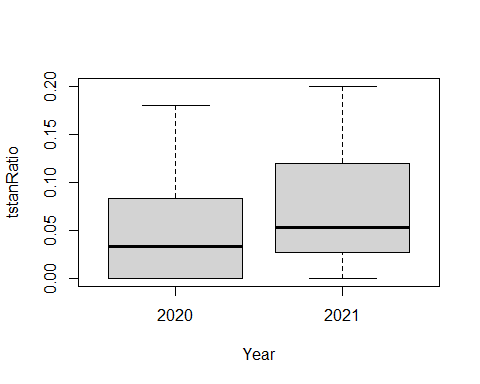
boxplot(nTstanCells ~ Year, data = zo.data.or) # no diff



## Ratio  
boxplot(tstanRatio ~ Year, data = zo.data) # similar



boxplot(tstanRatio ~ Year, data = zo.data.or) # similar



### Analyses

tmod1<-aov(nTstanCells ~ Year, data = zo.data)  
tmod1<-tidy(tmod1) # no diff  
kable(tmod1, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 77.393 | 77.393 | 1.23 | 0.276 |
| Residuals | 30 | 1887.325 | 62.911 | NA | NA |

tmod2<-aov(nTstanCells ~ Year, data = zo.data.or)   
tmod2<-tidy(tmod2) # no diff  
kable(tmod2, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.116 | 0.116 | 0.014 | 0.906 |
| Residuals | 29 | 237.303 | 8.183 | NA | NA |

tmod3<-aov(tstanRatio ~ Year, data = zo.data.or)   
tmod3<-tidy(tmod3) # no diff  
kable(tmod3, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

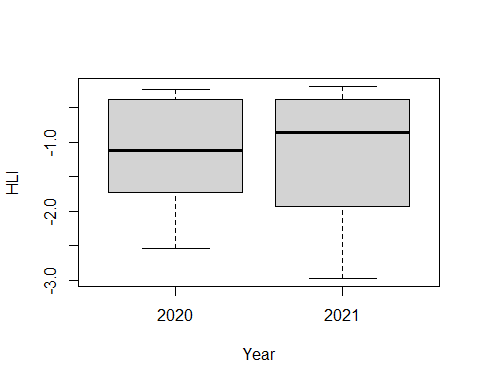
| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.001 | 0.001 | 0.284 | 0.598 |
| Residuals | 29 | 0.109 | 0.004 | NA | NA |

No differences

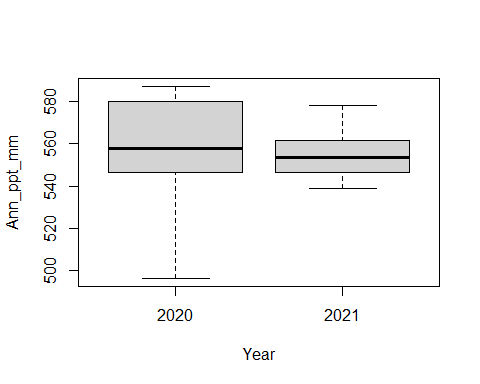
## Climate variables

### Plots

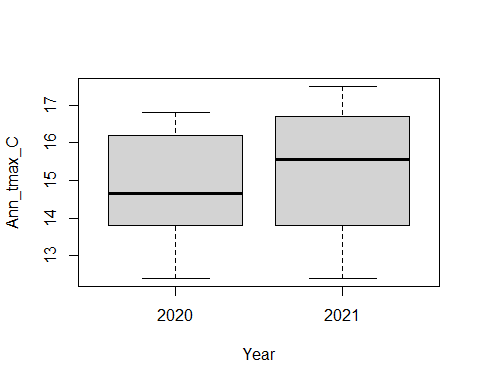
boxplot(HLI ~ Year, data=zo.data) # no diff



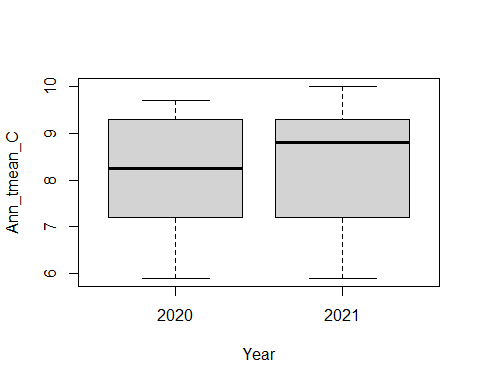
boxplot(Ann\_ppt\_mm ~ Year, data=zo.data) # no diff



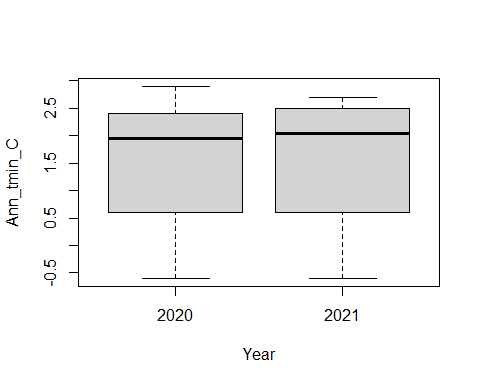
boxplot(Ann\_tmax\_C ~ Year, data=zo.data) # no diff



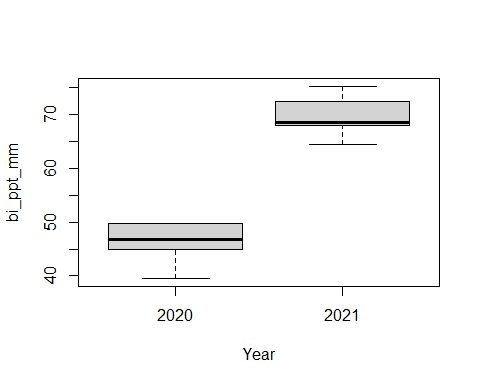
boxplot(Ann\_tmean\_C ~ Year, data=zo.data) # no diff



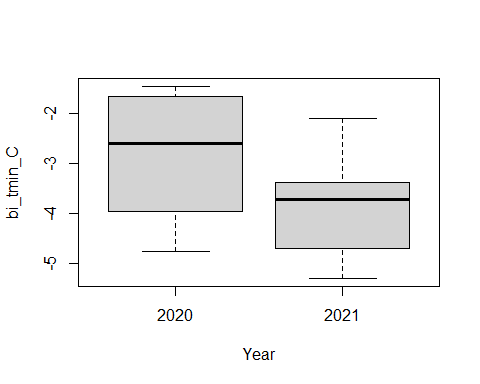
boxplot(Ann\_tmin\_C ~ Year, data=zo.data) # no diff



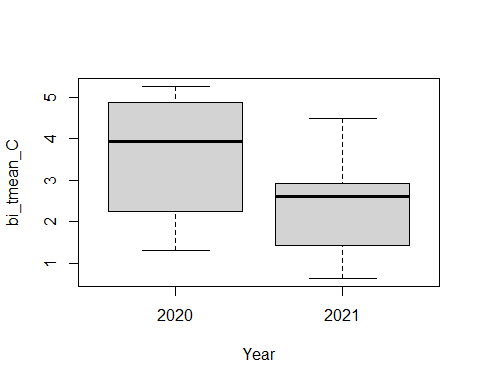
boxplot(bi\_ppt\_mm ~ Year, data = zo.data) # diff



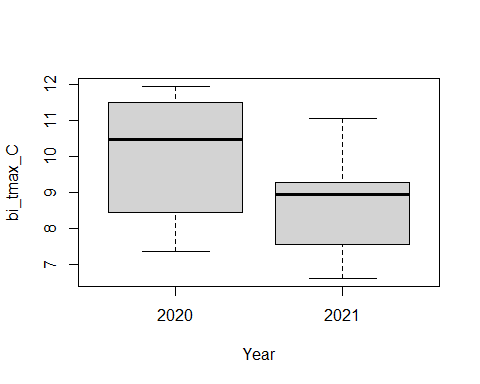
boxplot(bi\_tmin\_C ~ Year, data = zo.data) # diff



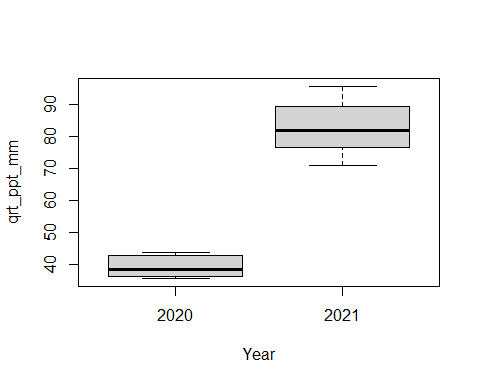
boxplot(bi\_tmean\_C ~ Year, data = zo.data) # diff



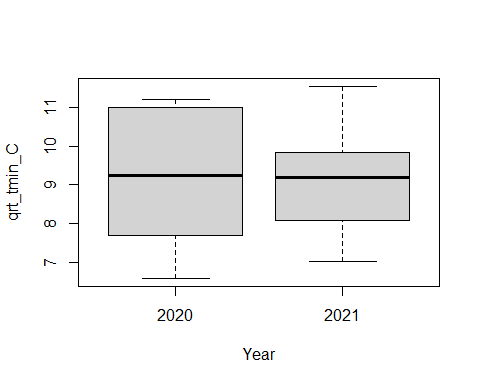
boxplot(bi\_tmax\_C ~ Year, data = zo.data) # diff



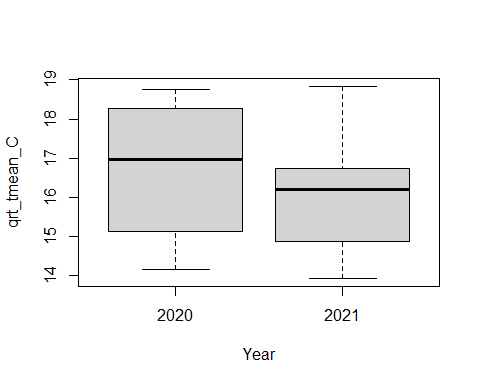
boxplot(qrt\_ppt\_mm ~ Year, data = zo.data) # diff



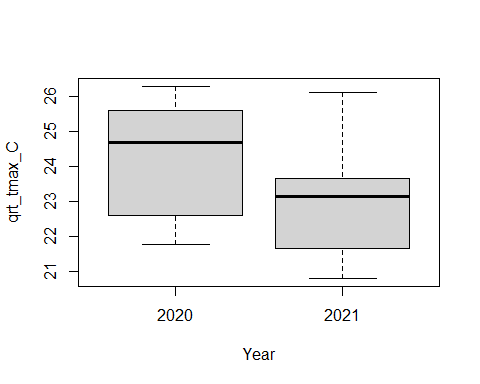
boxplot(qrt\_tmin\_C ~ Year, data = zo.data) # no diff



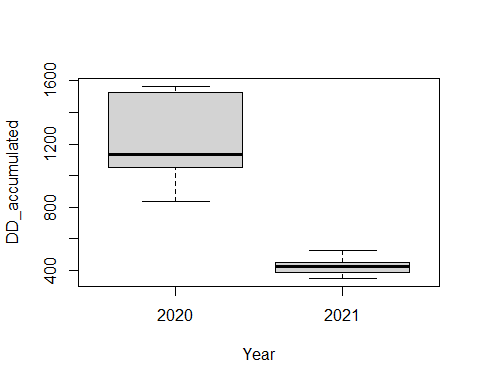
boxplot(qrt\_tmean\_C ~ Year, data = zo.data) # no diff



boxplot(qrt\_tmax\_C ~ Year, data = zo.data) # diff



boxplot(DD\_accumulated ~ Year, data=zo.data) # diff

 Precipitation and temperatures from Jan-Jun will probably be different. ADD is also MUCH higher in 2020 - boxes were left out longer and temperatures were higher in 2020

### Analyses

cmod1<-aov(HLI ~ Year, data=zo.data)   
cmod1<-tidy(cmod1) # no diff  
kable(cmod1, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.009 | 0.009 | 0.012 | 0.914 |
| Residuals | 30 | 21.911 | 0.730 | NA | NA |

cmod2<-aov(Ann\_ppt\_mm ~ Year, data=zo.data)   
cmod2<-tidy(cmod2) # no diff  
kable(cmod2, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 18.331 | 18.331 | 0.039 | 0.846 |
| Residuals | 30 | 14279.308 | 475.977 | NA | NA |

cmod3<-aov(Ann\_tmax\_C ~ Year, data=zo.data)   
cmod3<-tidy(cmod3) # no diff  
kable(cmod3, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.875 | 0.875 | 0.365 | 0.551 |
| Residuals | 30 | 72.000 | 2.400 | NA | NA |

cmod4<-aov(Ann\_tmean\_C ~ Year, data=zo.data)   
cmod4<-tidy(cmod4) # no diff  
kable(cmod4, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.249 | 0.249 | 0.148 | 0.703 |
| Residuals | 30 | 50.371 | 1.679 | NA | NA |

cmod5<-aov(Ann\_tmin\_C ~ Year, data=zo.data)   
cmod5<-tidy(cmod5) # no diff  
kable(cmod5, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.000 | 0.00 | 0 | 0.998 |
| Residuals | 30 | 36.915 | 1.23 | NA | NA |

cmod6<-aov(bi\_ppt\_mm ~ Year, data = zo.data)   
cmod6<-tidy(cmod6) # diff  
kable(cmod6, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 4306.538 | 4306.538 | 358.086 | 0 |
| Residuals | 30 | 360.796 | 12.027 | NA | NA |

cmod7<-aov(bi\_tmin\_C ~ Year, data = zo.data)   
cmod7<-tidy(cmod7) # diff  
kable(cmod7, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 7.296 | 7.296 | 7.199 | 0.012 |
| Residuals | 30 | 30.402 | 1.013 | NA | NA |

cmod8<-aov(bi\_tmean\_C ~ Year, data = zo.data)   
cmod8<-tidy(cmod8) # diff  
kable(cmod8, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 10.510 | 10.510 | 7.23 | 0.012 |
| Residuals | 30 | 43.612 | 1.454 | NA | NA |

cmod9<-aov(bi\_tmax\_C ~ Year, data = zo.data)   
cmod9<-tidy(cmod9) # diff  
kable(cmod9, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 14.271 | 14.271 | 7.281 | 0.011 |
| Residuals | 30 | 58.802 | 1.960 | NA | NA |

cmod10<-aov(qrt\_ppt\_mm ~ Year, data = zo.data)   
cmod10<-tidy(cmod10) # diff  
kable(cmod10, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 13870.106 | 13870.106 | 426.758 | 0 |
| Residuals | 30 | 975.032 | 32.501 | NA | NA |

cmod11<-aov(qrt\_tmin\_C ~ Year, data = zo.data)   
cmod11<-tidy(cmod11) # no diff  
kable(cmod11, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.204 | 0.204 | 0.095 | 0.76 |
| Residuals | 30 | 64.272 | 2.142 | NA | NA |

cmod12<-aov(qrt\_tmean\_C ~ Year, data = zo.data)   
cmod12<-tidy(cmod12) # no diff  
kable(cmod12, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 2.586 | 2.586 | 1.203 | 0.282 |
| Residuals | 30 | 64.502 | 2.150 | NA | NA |

cmod13<-aov(qrt\_tmax\_C ~ Year, data = zo.data)   
cmod13<-tidy(cmod13) # diff  
kable(cmod13, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 13.554 | 13.554 | 6.207 | 0.018 |
| Residuals | 30 | 65.504 | 2.183 | NA | NA |

cmod14<-aov(DD\_accumulated ~ Year, data=zo.data)   
cmod14<-tidy(cmod14) # diff  
kable(cmod14, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

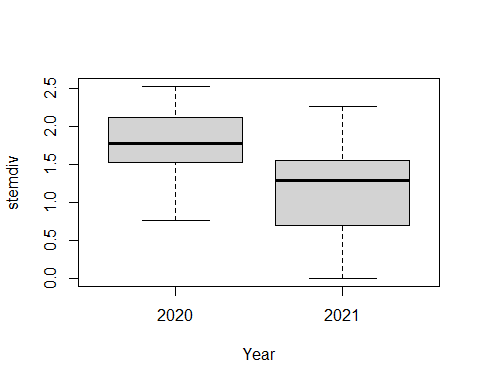
| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 4865801.8 | 4865801.80 | 166.335 | 0 |
| Residuals | 30 | 877590.6 | 29253.02 | NA | NA |

Significant differences: bi\_ppt\_mm, bi\_tmin\_C, bi\_tmean\_C, bi\_tmax\_C, qrt\_ppt\_mm, qrt\_tmax\_C, and DD\_accumulated

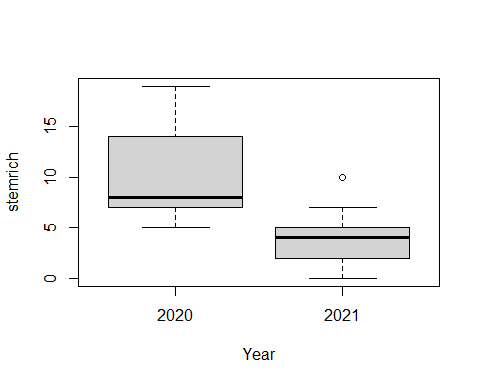
## Floral variables

### Plots

boxplot(stemdiv ~ Year, data=zo.data) # higher in 2020



boxplot(stemrich ~ Year, data=zo.data) # higher in 2020



### Analysis

fmod1<-aov(stemdiv ~ Year, data=zo.data)   
fmod1<-tidy(fmod1) # diff  
kable(fmod1, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 3.103 | 3.103 | 12.08 | 0.002 |
| Residuals | 30 | 7.705 | 0.257 | NA | NA |

fmod2<-aov(stemrich ~ Year, data=zo.data)  
fmod2<-tidy(fmod2) # diff  
kable(fmod2, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

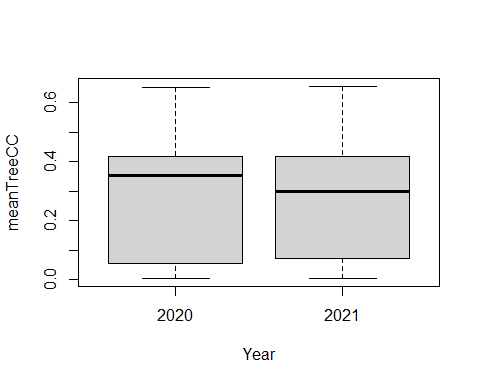
| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 305.667 | 305.667 | 23.286 | 0 |
| Residuals | 30 | 393.802 | 13.127 | NA | NA |

Significant vars: stemdiv, and stemrich

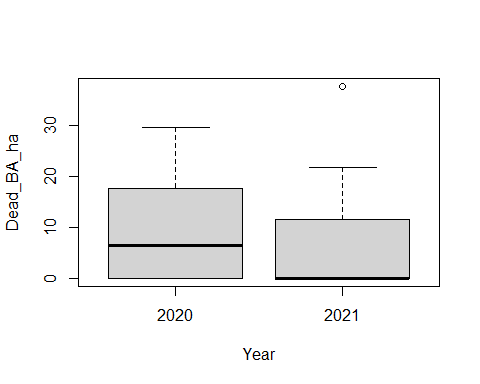
## Trees

### Plots

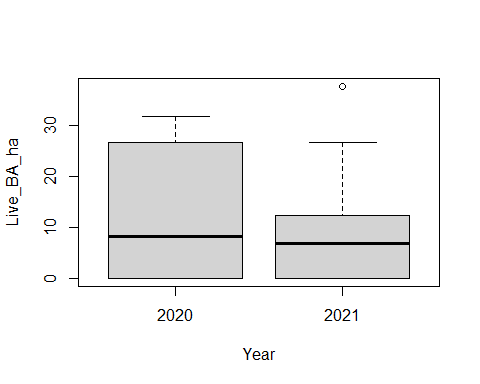
boxplot(meanTreeCC ~ Year, data=zo.data) # no diff



boxplot(Dead\_BA\_ha ~ Year, data=zo.data) # slight diff in median, probably not sig



boxplot(Live\_BA\_ha ~ Year, data=zo.data) # no diff



### Analysis

tmod1<-aov(meanTreeCC ~ Year, data=zo.data)   
tmod1<-tidy(tmod1) # no diff  
kable(tmod1, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.004 | 0.004 | 0.088 | 0.769 |
| Residuals | 30 | 1.201 | 0.040 | NA | NA |

tmod2<-aov(Dead\_BA\_ha ~ Year, data=zo.data)   
tmod2<-tidy(tmod2) # no diff  
kable(tmod2, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 45.514 | 45.514 | 0.413 | 0.526 |
| Residuals | 28 | 3087.580 | 110.271 | NA | NA |

tmod3<-aov(Live\_BA\_ha ~ Year, data=zo.data)   
tmod3<-tidy(tmod3) # no diff  
kable(tmod3, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

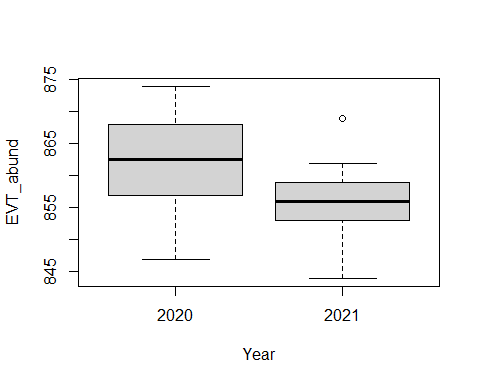
| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 43.323 | 43.323 | 0.31 | 0.582 |
| Residuals | 28 | 3908.357 | 139.584 | NA | NA |

No significant differences

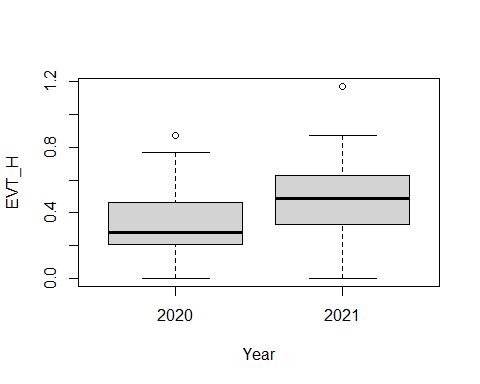
## Landscape variables

### Plots

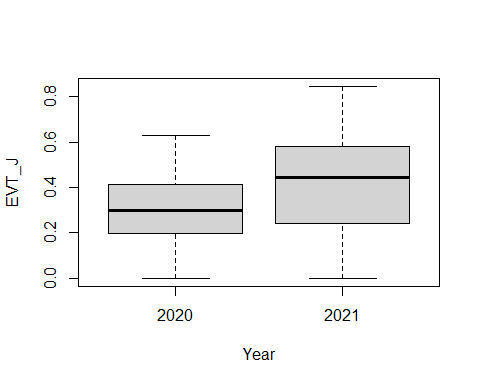
boxplot(EVT\_abund ~ Year, data=zo.data) # higher in 2020



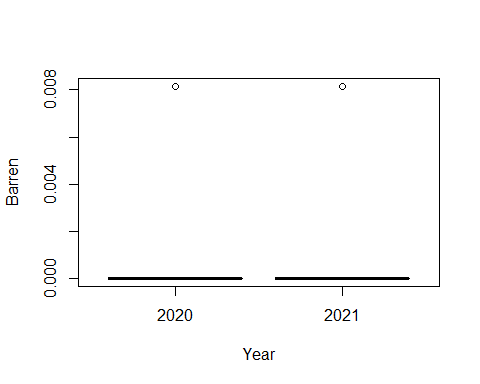
boxplot(EVT\_H ~ Year, data=zo.data) # higher in 2021



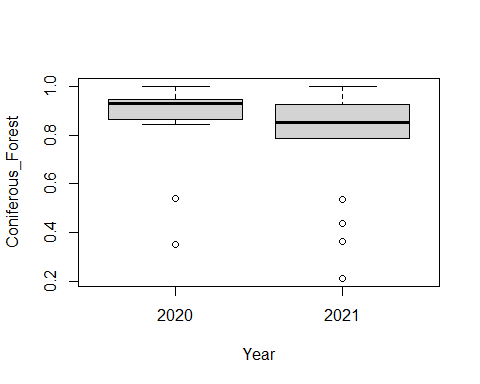
boxplot(EVT\_J ~ Year, data=zo.data) # slightly higher in 2020



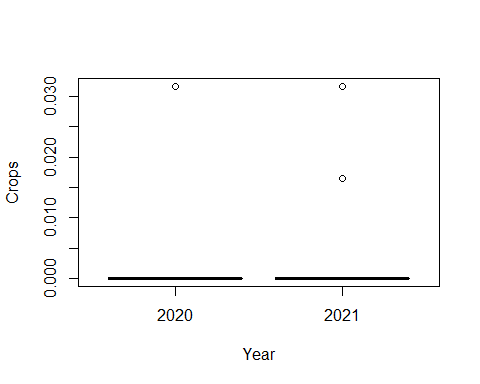
boxplot(Barren ~ Year, data=zo.data) # somewhat higher in 2021



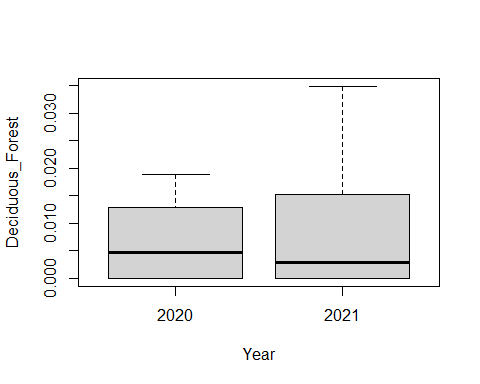
boxplot(Coniferous\_Forest ~ Year, data=zo.data) # no diff



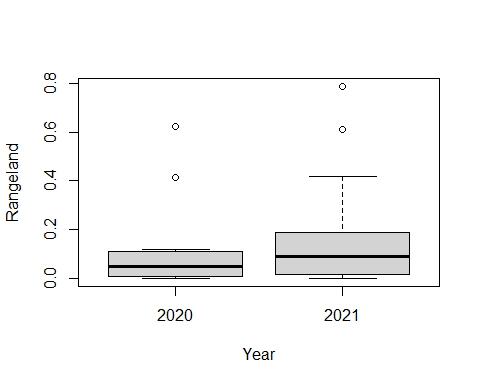
boxplot(Crops ~ Year, data=zo.data) # no diff



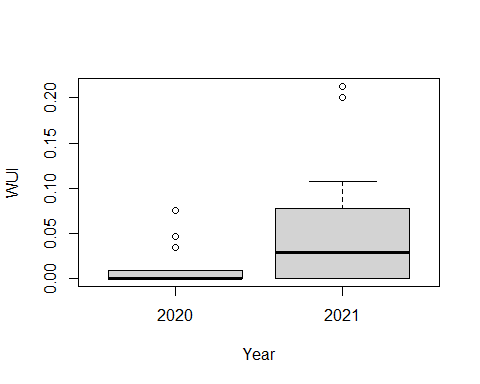
boxplot(Deciduous\_Forest ~ Year, data=zo.data) # no diff



boxplot(Rangeland ~ Year, data=zo.data) # no diff



boxplot(WUI ~ Year, data=zo.data) # higher in 2021

 Probably some differences in EVT\_abudn, EVT\_H, EVT\_J, Barren, and WUI

### Analysis

lmod1<-aov(EVT\_abund ~ Year, data=zo.data)   
lmod1<-tidy(lmod1) # diff  
kable(lmod1, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 259.290 | 259.290 | 6.903 | 0.013 |
| Residuals | 30 | 1126.929 | 37.564 | NA | NA |

lmod2<-aov(EVT\_H ~ Year, data=zo.data)   
lmod2<-tidy(lmod2) # marginally diff  
kable(lmod2, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.229 | 0.229 | 3.369 | 0.076 |
| Residuals | 30 | 2.041 | 0.068 | NA | NA |

lmod3<-aov(EVT\_J ~ Year, data=zo.data)   
lmod3<-tidy(lmod3) # marginally diff  
kable(lmod3, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.170 | 0.170 | 4.03 | 0.054 |
| Residuals | 30 | 1.269 | 0.042 | NA | NA |

lmod4<-aov(Barren ~ Year, data=zo.data)   
lmod4<-tidy(lmod4) # marginally diff  
kable(lmod4, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0 | 0 | 0.031 | 0.86 |
| Residuals | 30 | 0 | 0 | NA | NA |

lmod5<-aov(Coniferous\_Forest ~ Year, data=zo.data)   
lmod5<-tidy(lmod5) # no diff  
kable(lmod5, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.070 | 0.070 | 1.606 | 0.215 |
| Residuals | 30 | 1.316 | 0.044 | NA | NA |

lmod6<-aov(Crops ~ Year, data=zo.data)   
lmod6<-tidy(lmod6) # no diff  
kable(lmod6, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.000 | 0 | 0.019 | 0.891 |
| Residuals | 30 | 0.002 | 0 | NA | NA |

lmod7<-aov(Deciduous\_Forest ~ Year, data=zo.data)   
lmod7<-tidy(lmod7) # no diff  
kable(lmod7, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.000 | 0 | 0.173 | 0.68 |
| Residuals | 30 | 0.003 | 0 | NA | NA |

lmod8<-aov(Rangeland ~ Year, data=zo.data)   
lmod8<-tidy(lmod8) # no diff  
kable(lmod8, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.024 | 0.024 | 0.57 | 0.456 |
| Residuals | 30 | 1.287 | 0.043 | NA | NA |

lmod9<-aov(WUI ~ Year, data=zo.data)   
lmod9<-tidy(lmod9) # marginally diff  
kable(lmod9, col.names = c("Term","Df", "Sum Sq", "Mean Sq", "\*t-statistic\*", "\*P-value\*"), digits = 3 )

| Term | Df | Sum Sq | Mean Sq | *t-statistic* | *P-value* |
| --- | --- | --- | --- | --- | --- |
| Year | 1 | 0.011 | 0.011 | 4.054 | 0.053 |
| Residuals | 30 | 0.081 | 0.003 | NA | NA |

Significant vars: EVT\_abund Marginal sig vars: EVT\_H, EVT\_J, Barren, WUI

## Sum stats

For significantly different variables between years ### O. lignaria Significant differences: Larvae count, F count, F mass, M mass is marginal

zo.data %>% group\_by(Year) %>%   
 summarise\_at(c("nOligF", "nOligLarve", "avgOligFMass\_mg", "avgOligMMass\_mg"),   
 funs(mean=round(mean(.), 3), se=round(sd(.), 3))) %>% print.data.frame()

## Year nOligF\_mean nOligLarve\_mean avgOligFMass\_mg\_mean avgOligMMass\_mg\_mean  
## 1 2020 2.571 17.500 14.833 14.004  
## 2 2021 9.389 4.389 34.207 20.450  
## nOligF\_se nOligLarve\_se avgOligFMass\_mg\_se avgOligMMass\_mg\_se  
## 1 5.360 13.449 20.795 11.099  
## 2 7.531 5.237 13.523 9.755

### Climate

Significant differences: bi\_ppt\_mm, bi\_tmin\_C, bi\_tmean\_C, bi\_tmax\_C, and DD\_accumulated

# Field days between years  
zo.data %>% group\_by(Year) %>%   
 summarise\_at(c("bi\_ppt\_mm", "bi\_tmin\_C", "bi\_tmean\_C", "bi\_tmax\_C", "FieldDay", "DD\_accumulated"),   
 funs(mean=round(mean(.), 3), se=round(sd(.), 3))) %>% print.data.frame()

## Year bi\_ppt\_mm\_mean bi\_tmin\_C\_mean bi\_tmean\_C\_mean bi\_tmax\_C\_mean  
## 1 2020 46.105 -2.839 3.629 10.083  
## 2 2021 69.490 -3.801 2.473 8.737  
## FieldDay\_mean DD\_accumulated\_mean bi\_ppt\_mm\_se bi\_tmin\_C\_se bi\_tmean\_C\_se  
## 1 62.643 1214.964 4.035 1.153 1.379  
## 2 36.000 428.911 2.962 0.878 1.054  
## bi\_tmax\_C\_se FieldDay\_se DD\_accumulated\_se  
## 1 1.602 9.476 253.426  
## 2 1.223 0.000 50.099

### Floral

Significant vars: stemdiv, and stemrich

zo.data %>% group\_by(Year) %>%   
 summarise\_at(c("stemdiv", "stemrich"),   
 funs(mean=round(mean(.), 3), se=round(sd(.), 3))) %>% print.data.frame()

## Year stemdiv\_mean stemrich\_mean stemdiv\_se stemrich\_se  
## 1 2020 1.792 10.286 0.452 4.811  
## 2 2021 1.164 4.056 0.545 2.338

### Landscape

Significant vars: EVT\_abund Marginal sig vars: EVT\_H, EVT\_J, Barren, WUI

zo.data %>% group\_by(Year) %>%   
 summarise\_at(c("EVT\_abund", "EVT\_H", "EVT\_J", "Barren", "WUI"),   
 funs(mean=round(mean(.), 3), se=round(sd(.), 3))) %>% print.data.frame()

## Year EVT\_abund\_mean EVT\_H\_mean EVT\_J\_mean Barren\_mean WUI\_mean EVT\_abund\_se  
## 1 2020 862.071 0.334 0.295 0.001 0.013 7.227  
## 2 2021 856.333 0.504 0.442 0.000 0.050 5.134  
## EVT\_H\_se EVT\_J\_se Barren\_se WUI\_se  
## 1 0.255 0.184 0.002 0.023  
## 2 0.265 0.221 0.002 0.066

# How to analyze data

<https://drizopoulos.github.io/GLMMadaptive/articles/ZeroInflated_and_TwoPart_Models.html#two-parthurdle-poisson-mixed-effects-model>

* Weight data \*\* is continuous data with lots of zeros, or semi-continuous data. Therefore, a *Two-Part Mixed Effects Model for Semi-Continuous Data* might be appropriate \*\*\* this is a two step model: \*\*\*\* 1) uses a logistic regression to determine if the outcome is zero or not (dichotomous) \*\*\*\* 2) if data is not zero, uses a standard linear mixed model \*\* scratch that - it doesn’t make since to include zeros for weight data, only count data. Therefore, just do a glm. \*\*\* Note: Do NOT want to use a Poisson dist (only useful for integer data)
* Count data \*\* is discrete data and we want to know the distribution. Commonly, a Poisson distribution (generalized linear model family) is used for count data. \*\* However, our data is very zero heavy \*\* Can use a *Zero-inflated Poisson Mixed Effects Model* \*\*\* user has the option to leave **zi\_random** set to **NULL**, in which case for the zero-part we have a logistic regression with only fixed effects and no random effects \*\* Another option is the *Two-Part/Hurdle Poisson Mixed Effects Model* similar to the semi-continuous data data (above) but for discrete data (i.e. counts) instead of continuous data \*\*\* 1) uses logistic regression to determine if data is zero or not \*\*\* 2) if value is positive, used a truncated zero-inflated Poisson mixed effects model

## Q1

* T. stansburyi abundance by landscape, climate, and floral variables \*\* What variables to include? \*\* Predictor vars to include in stp-wise model: <https://journals.lww.com/picp/fulltext/2017/08030/Common_pitfalls_in_statistical_analysis__Logistic.9.aspx>
  1. What vars have a sig, univariate relationship (P < 0.1)
  2. Avoid highly correlated variables

## Q2

* T. stansburyi presence depends on O.lig provisions - \*\* Omit sites that don’t have provisions \*\* Note: different than sites with no O.lig ratios \*\*\* 1 site in 2021 was completely parasitized by T. stan
* 2020 \*\* ntstanCells \*\*\* is zero influenced (zi), has a HUGE outlier, and not normal \*\*\*\* log + 1 transformation seems to help, but is still zero heavy, then has a normal dist \*\*\*\* ZINB or ZIP might also be required \*\*\*\*\* Or can we just do a ZINB/ZIP without transforming? \*\* nstan ratio \*\*\* not zi, but is zero heavy and not normal \*\*\* transformations don’t seem to help \*\*\*\*\* glm with poisson dist?
* 2021 \*\* ntstanCells \*\*\* not zi, but is zero heavy and not normal \*\*\* sqrt and log+1 transformations help \*\*\*\* lm/glm \*\* tstan ratio \*\*\* not zi, but is zero heavy and not normal \*\*\* sqrt transformation helps \*\*\*\* lm/glm