

# warmXtrophic Project: Greenup Analyses

March 01, 2021

COLLABORATORS: Moriah Young, Mark Hammond, Pat Bills

DATA INPUT: Data imported as csv files from shared Google drive “SpaCE\_Lab\_warmXtrophic” plant comp folder

DATA OUTPUT: a brief description of the data output from through the script, including what format it’s in

OVERVIEW: This script explores and analyses the greenup data from the KBS and UMBS stations. Greenup was recorded as percent cover of each species in each plot every 3 days in the beginning of the field season. Percent cover continued to be measured until the end of the field season. For calculating greenup dates, we use the date at which 50% of a species maximum percent cover was reached for each plot, to account for variation in litter depth affecting the early season’s first greenup date. REQUIRES: any scripts or code sources that are required NOTES: The green\_kbs and green\_umbs dataframes contain data for greenup at each site.

“half\_cover\_date” is the date at which 50% of a species max cover was reached (per plot, per year)

“state” describes each treatment - warmed or ambient

```
# Clear all existing data
rm(list=ls())

#Load packages
library(tidyverse)
library(lme4)
library(olsrr)
library(predictmeans)
library(car)
library(fitdistrplus)
library(ggpubr)
library(rstatix)
library(vegan)

# Set working directory to Google Drive
# **** Update with the path to your Google drive on your computer
setwd("/Volumes/GoogleDrive/Shared drives/SpaCE_Lab_warmXtrophic/data/")

# Read in plant comp data
greenup <- read.csv("L1/greenup/final_greenup_L1.csv")
# check variable types
str(greenup)
```

```
## 'data.frame':   3419 obs. of  13 variables:
## $ X             : int  1 2 3 4 5 6 7 8 9 10 ...
## $ species       : chr  "Acru" "Acru" "Acru" "Acru" ...
## $ origin        : chr  "Native" "Native" "Native" "Native" ...
## $ duration      : chr  "Perennial" "Perennial" "Perennial" "Perennial" ...
## $ growth_habit  : chr  "Tree" "Tree" "Tree" "Tree" ...
## $ plot          : chr  "A4" "B2" "B2" "B2" ...
## $ state         : chr  "warmed" "warmed" "warmed" "warmed" ...
## $ treatment_key : chr  "W0" "W0" "W0" "W0" ...
```

```
## $ insecticide : chr "no_insects" "no_insects" "no_insects" "no_insects" ...
## $ half_cover_date: int 167 135 145 167 175 144 167 141 202 197 ...
## $ site : chr "umbs" "umbs" "umbs" "umbs" ...
## $ year : int 2017 2016 2019 2017 2015 2018 2017 2015 2015 2016 ...
## $ min_emerg_date : int 167 112 112 112 112 112 167 141 80 80 ...
```

```
# adding sequential year variable starting at 1
```

```
greenup$year1<-greenup$year
greenup$year[greenup$year == 2015] <- 1
greenup$year[greenup$year == 2016] <- 2
greenup$year[greenup$year == 2017] <- 3
greenup$year[greenup$year == 2018] <- 4
greenup$year[greenup$year == 2019] <- 5
greenup$year[greenup$year == 2020] <- 6
```

```
# create dataframes for kbs and umbs
```

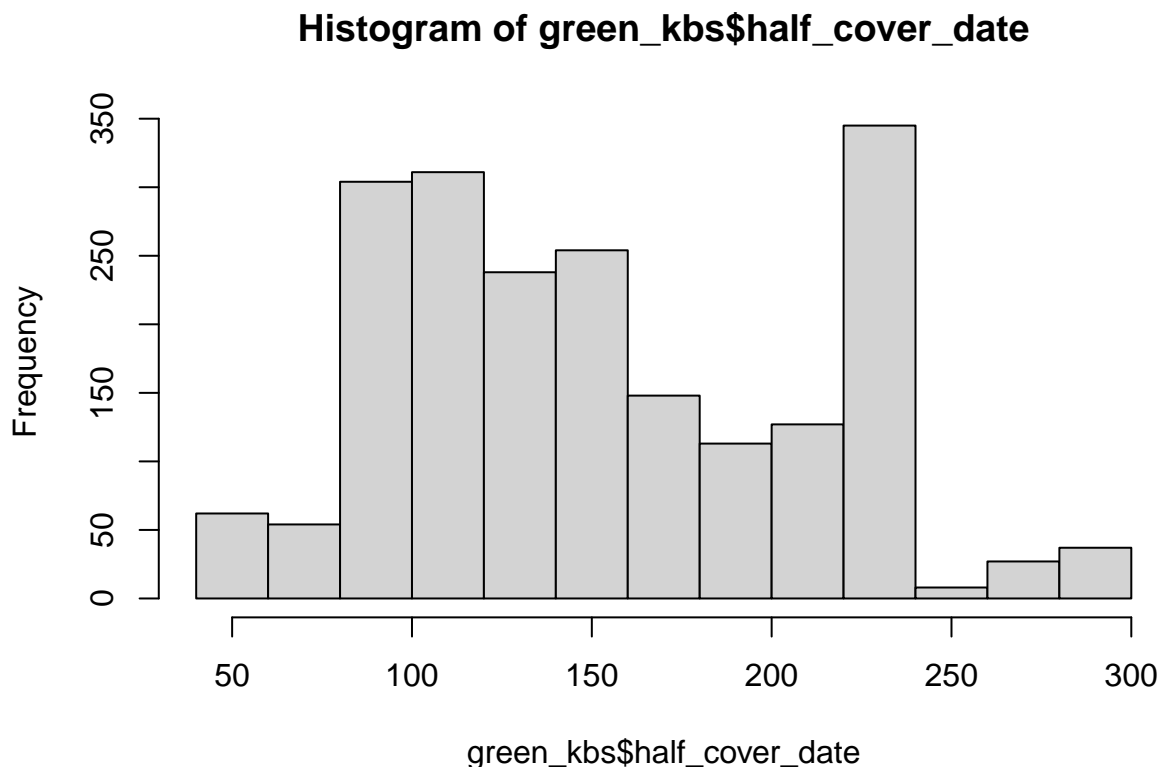
```
green_kbs <- subset(greenup, site == "kbs")
green_umbs <- subset(greenup, site == "umbs")
```

**Data Exploration: are there differences between warmed vs. ambient plots?**

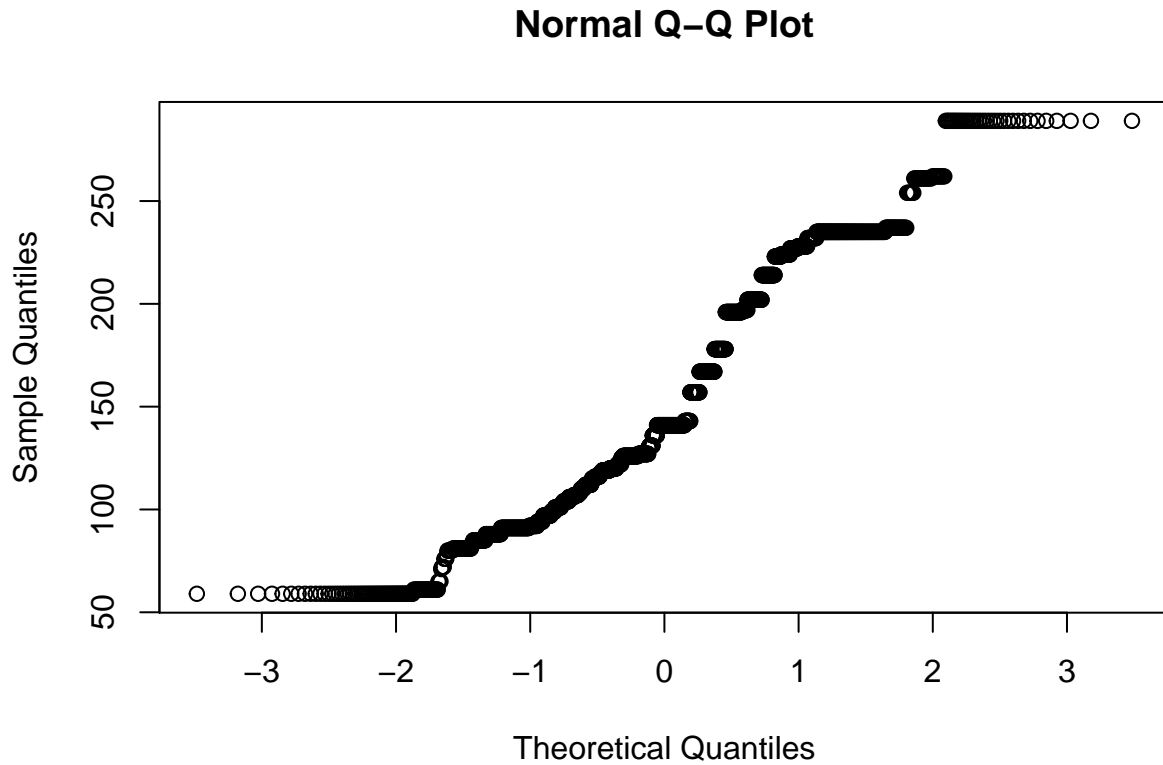
### *Starting with KBS*

First, checking for normality in raw data

```
hist(green_kbs$half_cover_date)
```



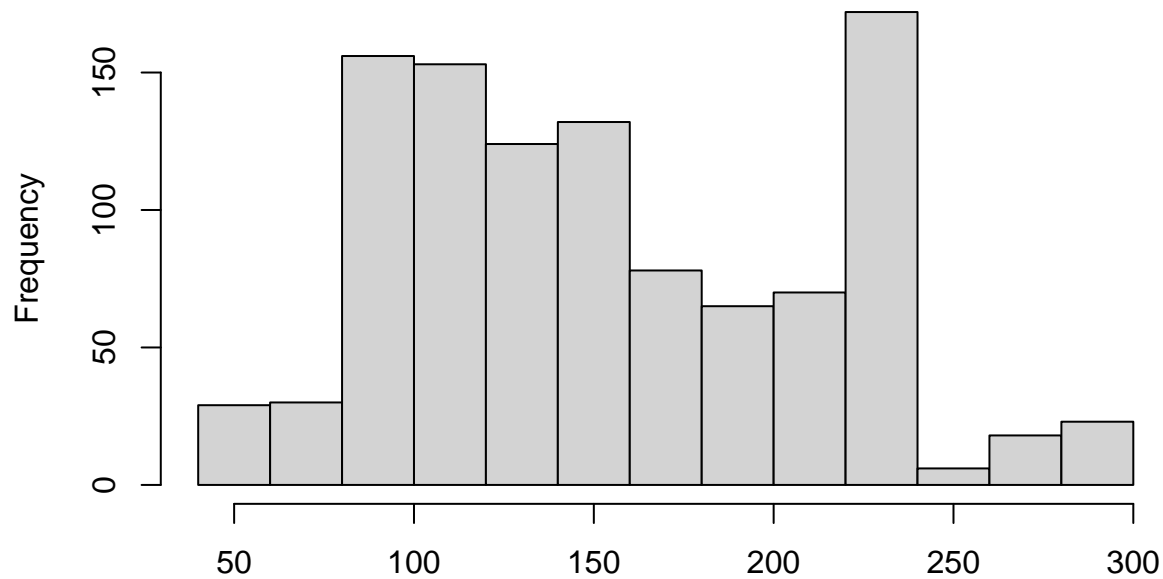
```
qqnorm(green_kbs$half_cover_date)
```



```
shapiro.test(green_kbs$half_cover_date)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  green_kbs$half_cover_date  
## W = 0.94273, p-value < 2.2e-16  
  
# histograms for each treatment separately - look almost identical  
hist(green_kbs$half_cover_date[green_kbs$state == "ambient"])
```

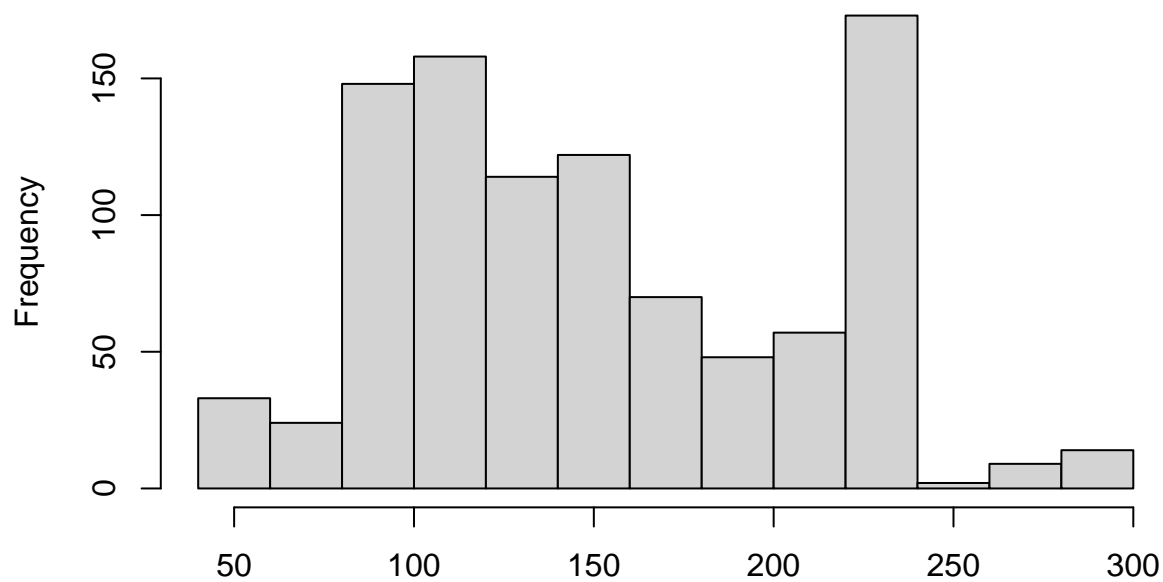
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$state == "ambient"]**



```
green_kbs$half_cover_date[green_kbs$state == "ambient"]
```

```
hist(green_kbs$half_cover_date[green_kbs$state == "warmed"])
```

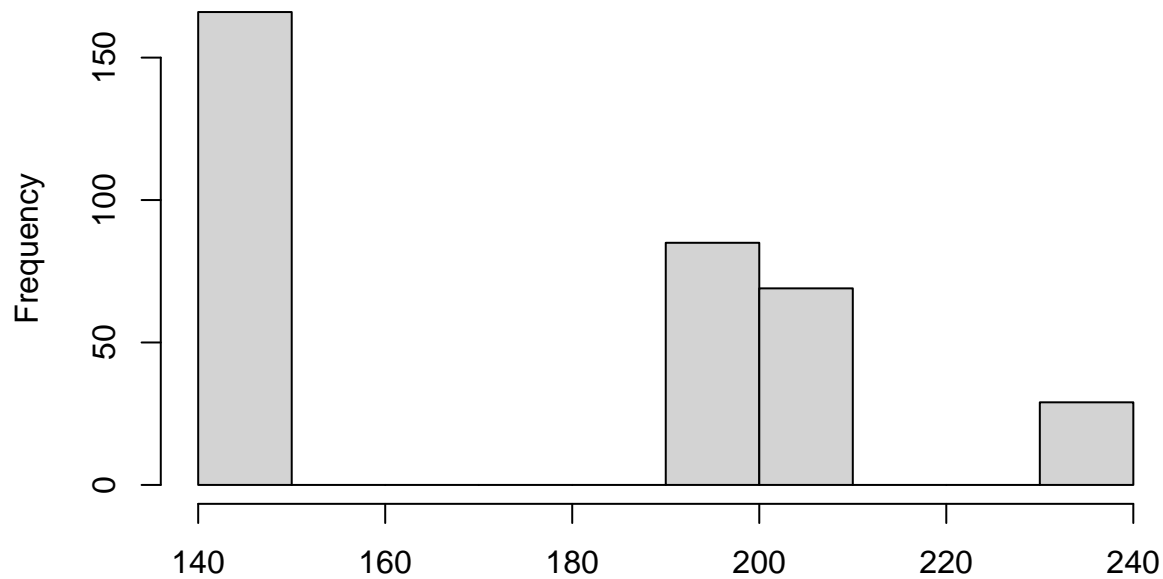
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$state == "warmed"]**



```
green_kbs$half_cover_date[green_kbs$state == "warmed"]
```

```
# histograms for each year
hist(green_kbs$half_cover_date[green_kbs$year == 1])
```

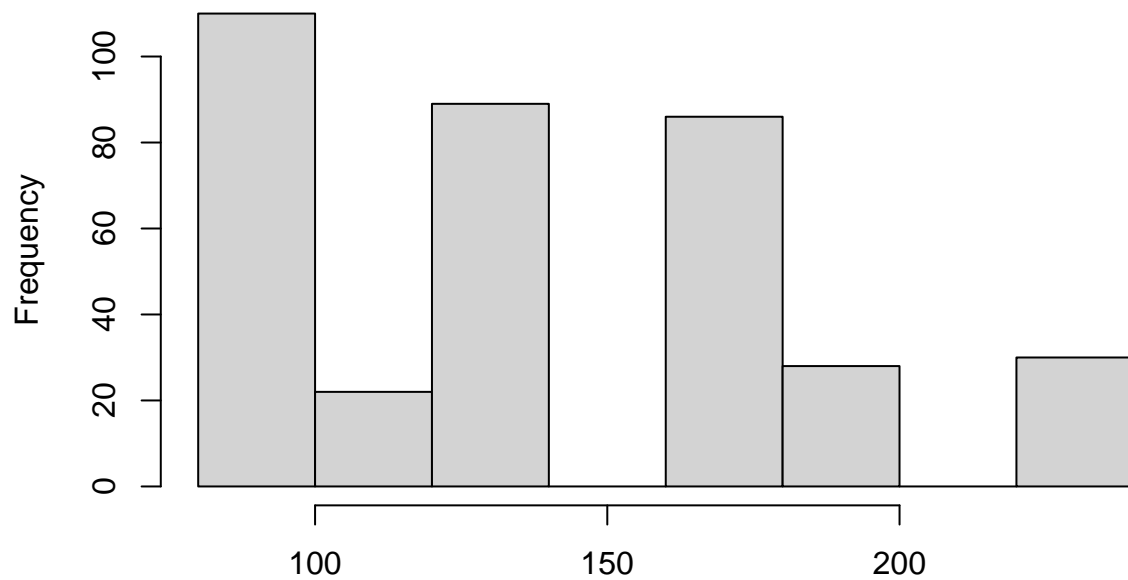
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$year == 1]**



green\_kbs\$half\_cover\_date[green\_kbs\$year == 1]

```
hist(green_kbs$half_cover_date[green_kbs$year == 2])
```

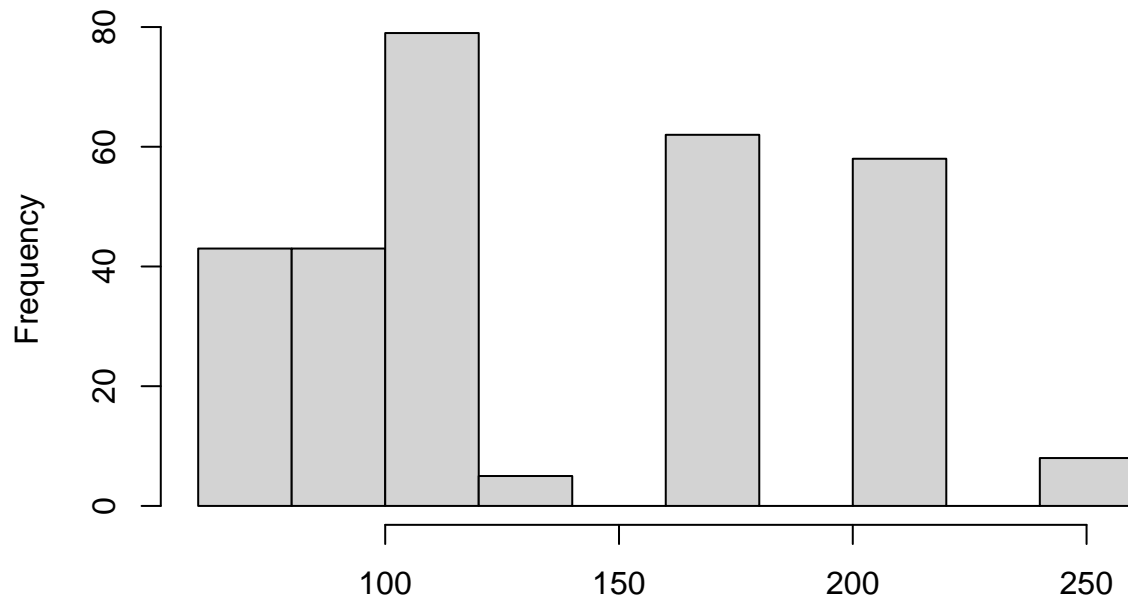
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$year == 2]**



green\_kbs\$half\_cover\_date[green\_kbs\$year == 2]

```
hist(green_kbs$half_cover_date[green_kbs$year == 3])
```

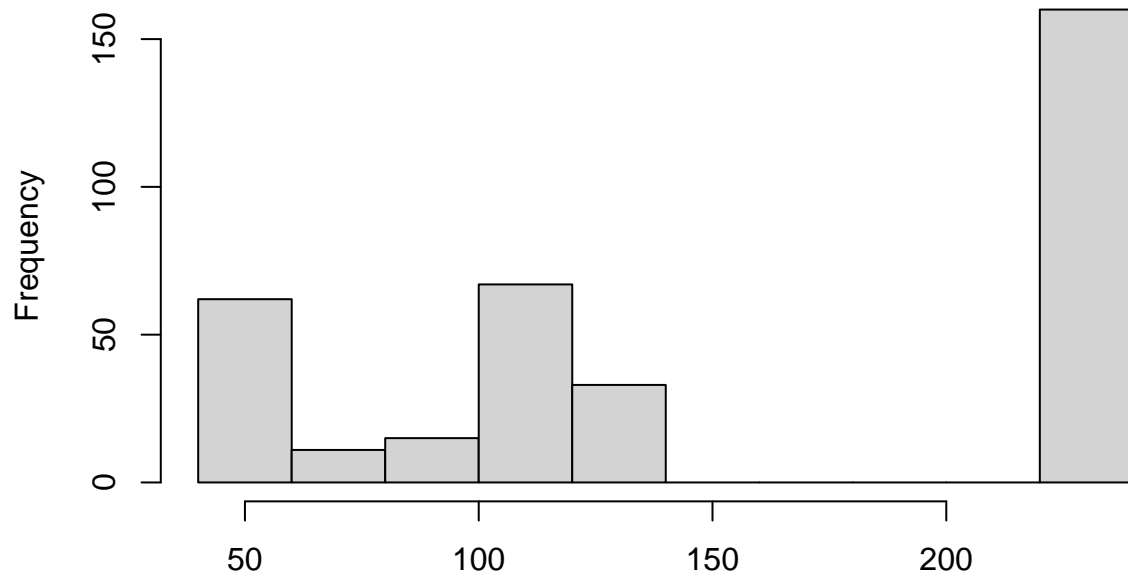
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$year == 3]**



green\_kbs\$half\_cover\_date[green\_kbs\$year == 3]

```
hist(green_kbs$half_cover_date[green_kbs$year == 4])
```

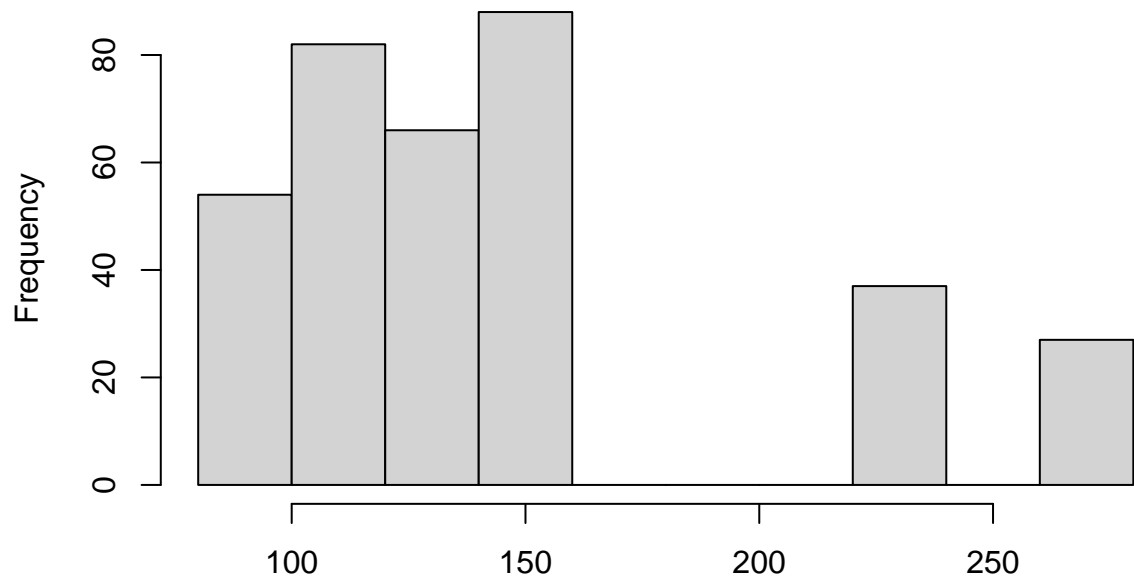
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$year == 4]**



green\_kbs\$half\_cover\_date[green\_kbs\$year == 4]

```
hist(green_kbs$half_cover_date[green_kbs$year == 5])
```

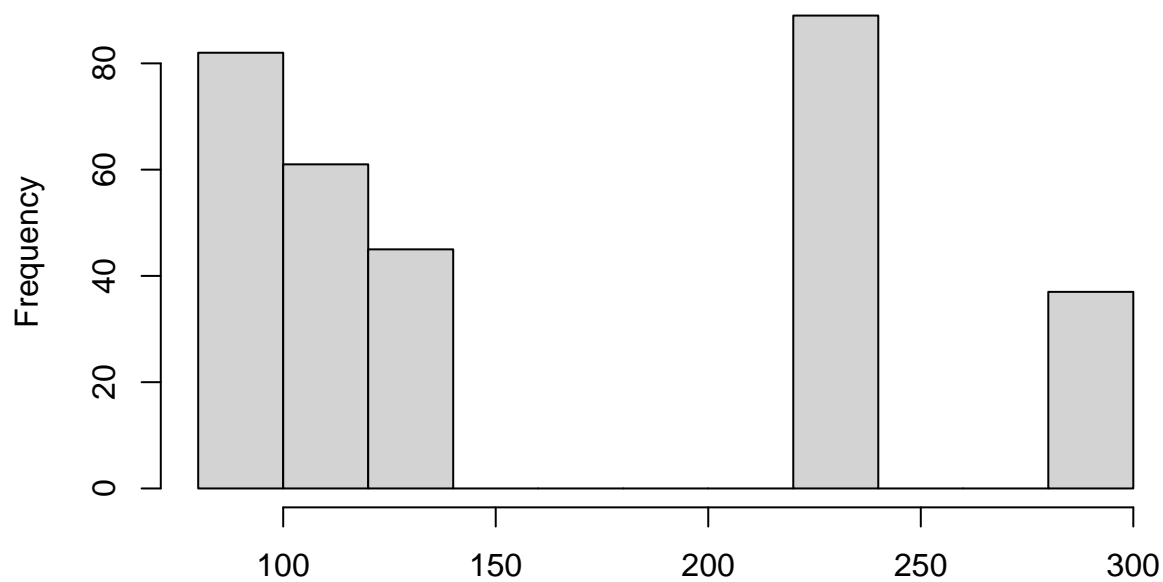
**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$year == 5]**



green\_kbs\$half\_cover\_date[green\_kbs\$year == 5]

```
hist(green_kbs$half_cover_date[green_kbs$year == 6])
```

**Histogram of green\_kbs\$half\_cover\_date[green\_kbs\$year == 6]**



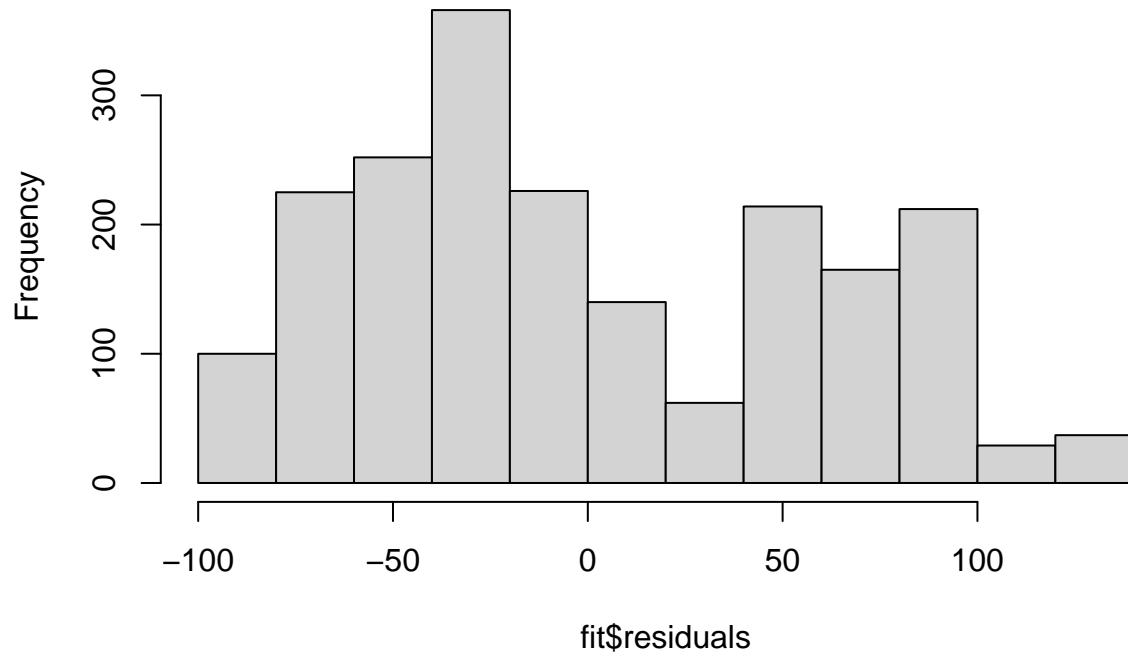
green\_kbs\$half\_cover\_date[green\_kbs\$year == 6]

```
# looks like the 225 spike is from 2018 and 2020
kbs_2018 <- subset(green_kbs, year == 4) # many records on 235
kbs_2020 <- subset(green_kbs, year == 6) # records from 227 & 228
```

Leverage plots? Not sure what these show exactly, but we discussed them previously

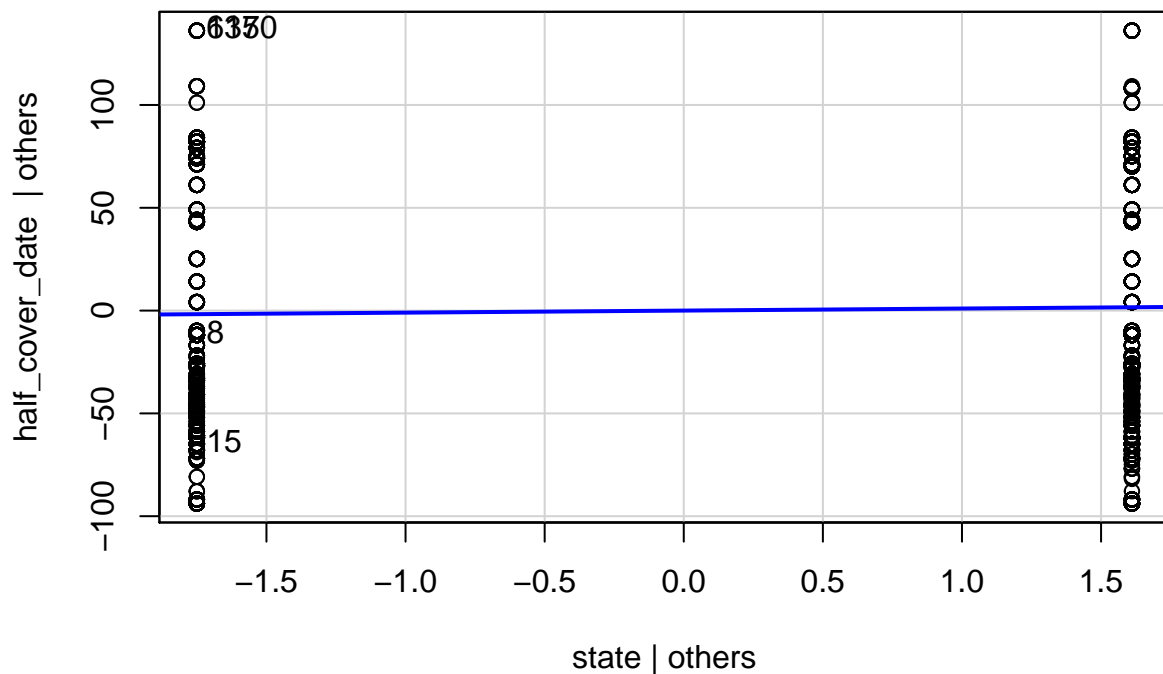
```
# checking fit for date as a function of state  
fit <- lm(half_cover_date~state, data = green_kbs)  
hist(fit$residuals)
```

**Histogram of fit\$residuals**



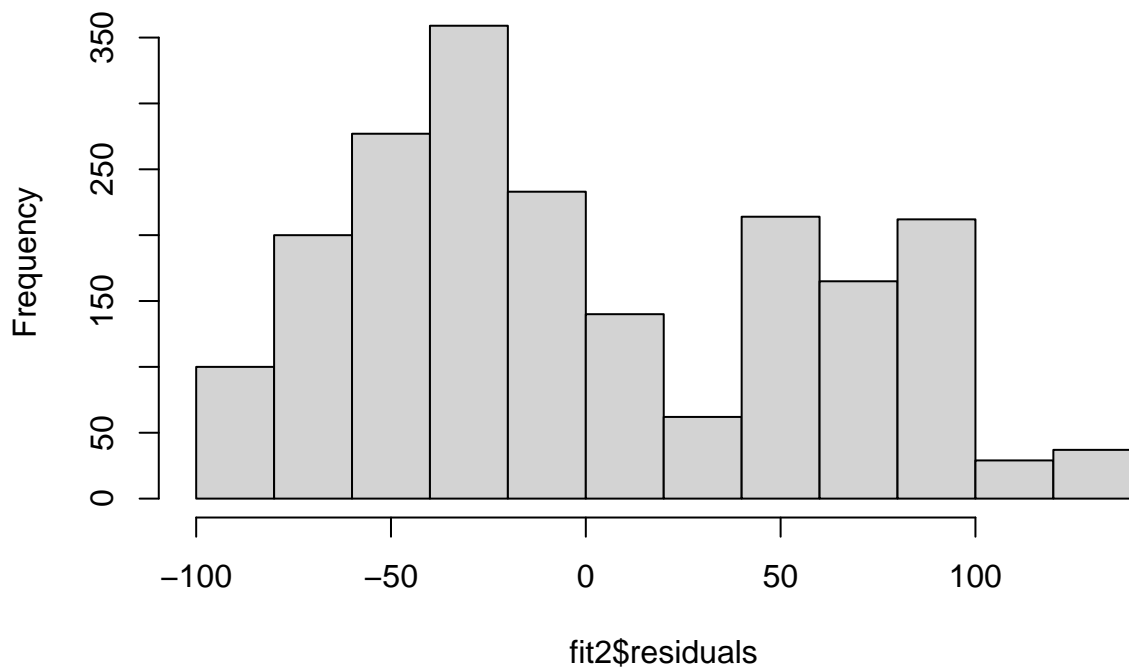
```
leveragePlots(fit)
```





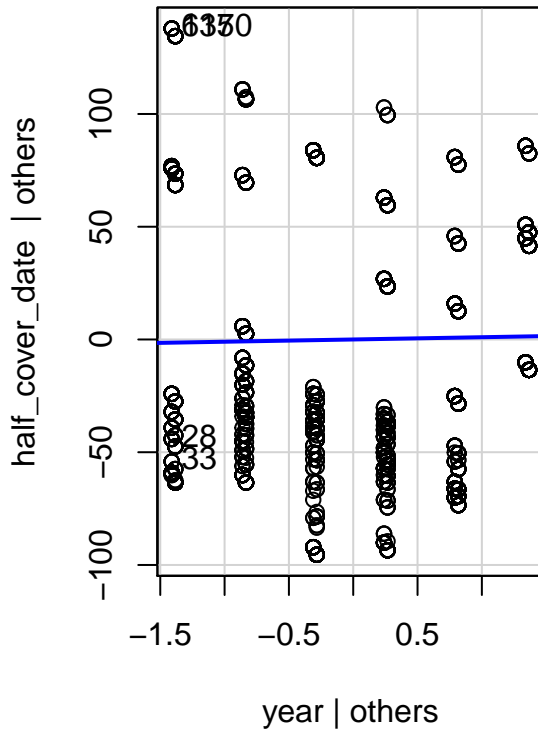
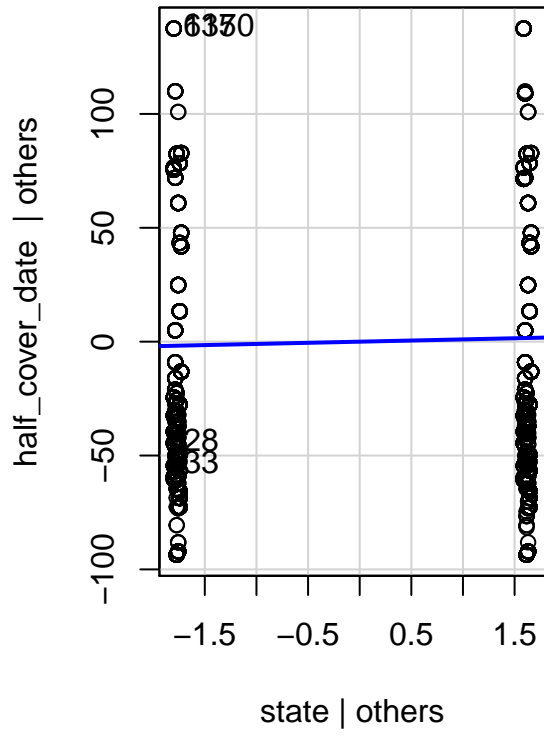
```
# checking fit for date as a function of state and year
fit2 <- lm(half_cover_date~state+year, data = green_kbs)
hist(fit2$residuals)
```

**Histogram of fit2\$residuals**



```
leveragePlots(fit2)
```

## Leverage Plots

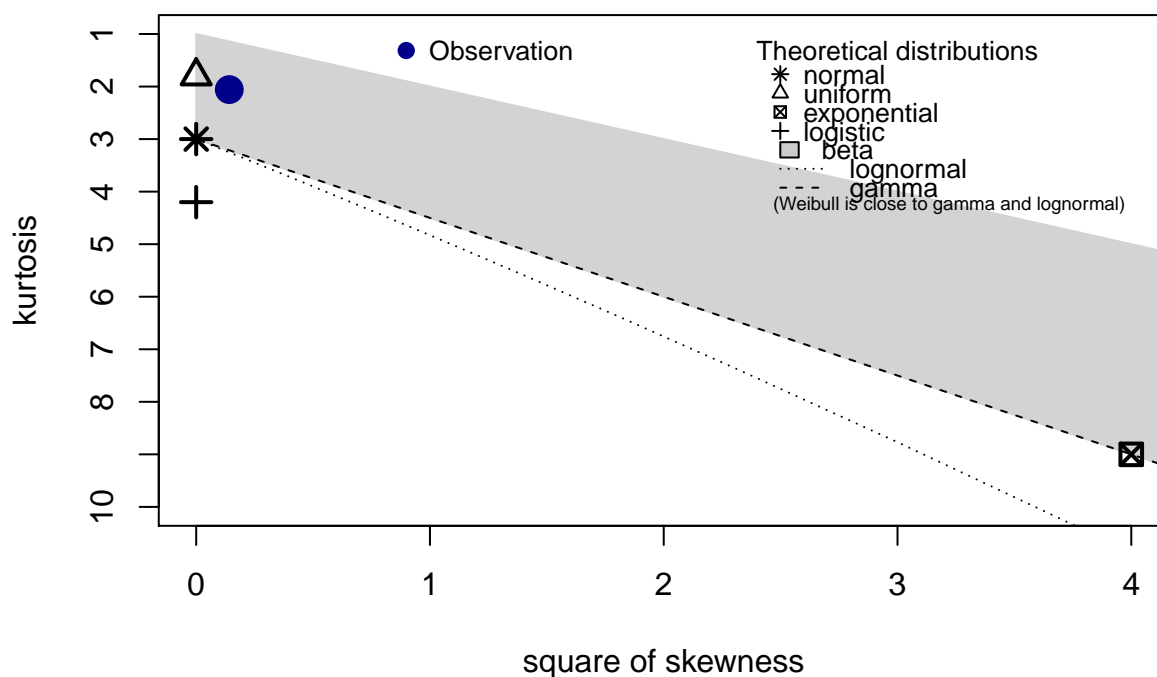


Not normal, and previously attempted transformations don't help  
(in R script)

Seeing what other distribution could fit

```
descdist(green_kbs$half_cover_date, discrete = FALSE)
```

### Cullen and Frey graph

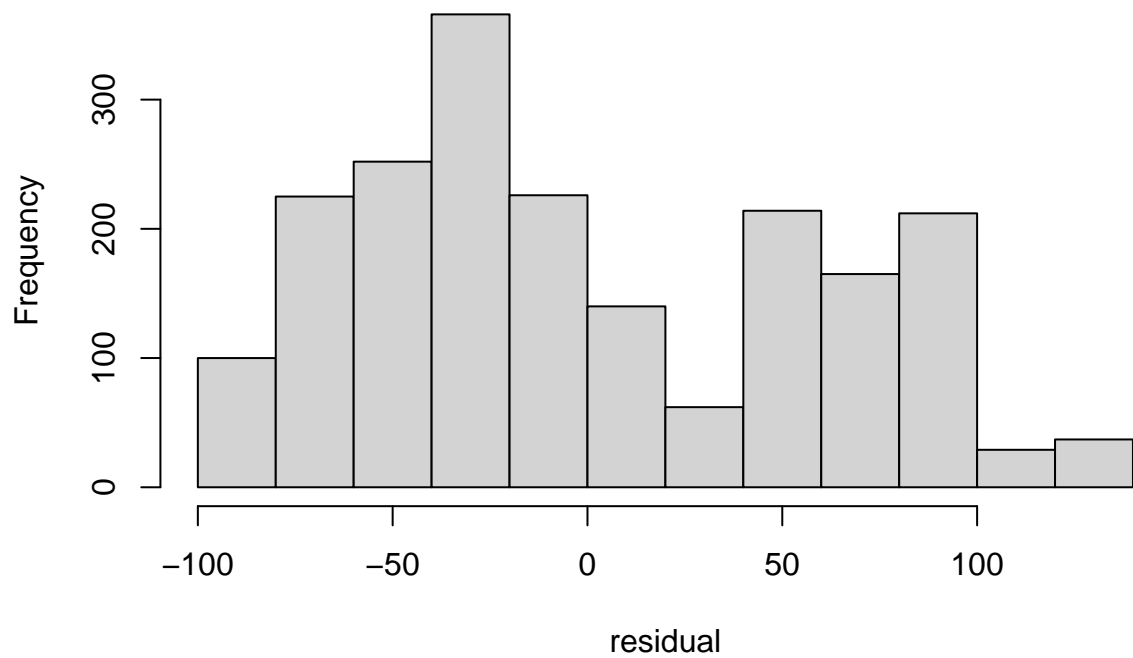


```
## summary statistics
## -----
## min: 59    max: 289
## median: 141
## mean: 152.8644
## estimated sd: 57.73216
## estimated skewness: 0.3763172
## estimated kurtosis: 2.058193
```

While uniform looks the closest, I'll try poisson

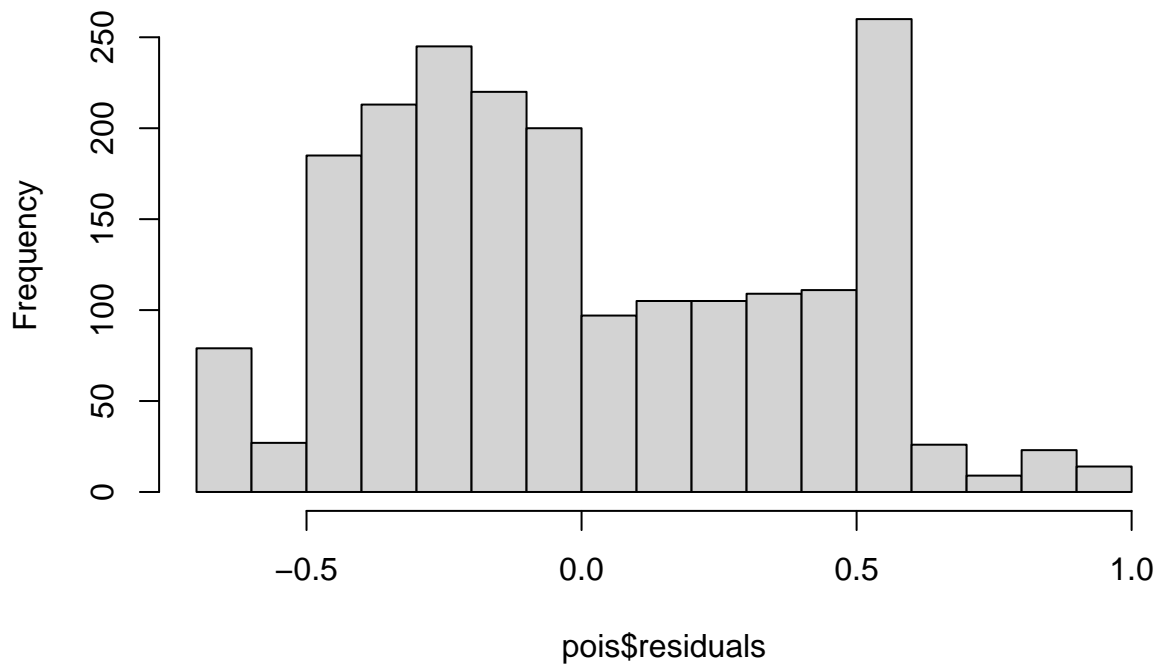
```
fit <- lm(half_cover_date~state, data = green_kbs)
residual <- fit$residuals
hist(residual, main="Raw residuals")
```

## Raw residuals



```
pois <- glm(half_cover_date~state, data = green_kbs, family="poisson")  
hist(pois$residuals, main="Poisson glm residuals")
```

## Poisson glm residuals



Below I try a few different generalized linear models with poisson distribution:

An interaction between state and year, plus insecticide as a fixed effect and species and plot as random effects

```
moda <- glmer(half_cover_date ~ state*year + insecticide + (1|species) + (1|plot),
              data=green_kbs, family = poisson)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable:
## - Rescale variables?
```

```
summary(moda)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: half_cover_date ~ state * year + insecticide + (1 | species) +
## (1 | plot)
## Data: green_kbs
##
##          AIC          BIC    logLik deviance df.resid
## 42924.6   42963.9 -21455.3  42910.6      2021
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -9.0161 -2.8352 -0.6425  2.3763 14.4656
##
## Random effects:
## Groups Name          Variance Std.Dev.
## species (Intercept) 0.069953 0.26449
## plot (Intercept)    0.001848 0.04299
## Number of obs: 2028, groups: species, 55; plot, 24
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    5.020672   0.039631 126.686 < 2e-16 ***
## statewarmed     0.009389   0.019347   0.485  0.6275
## year            0.003600   0.001520   2.368  0.0179 *
## insecticideno_insects 0.028935   0.017936   1.613  0.1067
## statewarmed:year -0.009698   0.002116 -4.584 4.57e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) sttwrm year  insct_
## statewarmed -0.243
## year        -0.130  0.249
## insctcdn_ns -0.227 -0.002  0.008
## statwrmd:yr  0.089 -0.375 -0.648  0.006
## convergence code: 0
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?
```

No interaction between state and year, but with state and insecticide as fixed effects and species and plot as random effects

```
modb <- glmer(half_cover_date ~ state + year + insecticide + (1|species) + (1|plot),
              data=green_kbs, family = poisson)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable:
## - Rescale variables?
```

```
summary(modb)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: half_cover_date ~ state + year + insecticide + (1 | species) +
## (1 | plot)
## Data: green_kbs
##
##      AIC      BIC   logLik deviance df.resid
## 42943.6 42977.3 -21465.8 42931.6      2022
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -9.0054 -2.8448 -0.6575  2.3929 14.2130
##
## Random effects:
## Groups Name          Variance Std.Dev.
## species (Intercept) 0.069954 0.26449
## plot (Intercept)    0.001847 0.04297
## Number of obs: 2028, groups: species, 55; plot, 24
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.0367370  0.0394718 127.603  <2e-16 ***
## statewarmed      -0.0238778  0.0179289  -1.332   0.183
## year             -0.0009129  0.0011579  -0.788   0.430
## insecticideno_insects 0.0294349  0.0179300   1.642   0.101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) sttwrm year
## statewarmed -0.227
## year        -0.096  0.007
## insctcdn_ns -0.228  0.001  0.015
## convergence code: 0
## Model is nearly unidentifiable: very large eigenvalue
## - Rescale variables?
```

## State and insecticide as fixed effects & year, species and plot as random effects

```
modc <- glmer(half_cover_date ~ state + insecticide + (1|year) + (1|species) + (1|plot),
              data=green_kbs, family = poisson)
summary(modc)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: half_cover_date ~ state + insecticide + (1 | year) + (1 | species) +
##   (1 | plot)
## Data: green_kbs
##
##      AIC      BIC   logLik deviance df.resid
## 41436.8 41470.5 -20712.4 41424.8      2022
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -9.3418 -2.8134 -0.5215  2.1783 13.1976
##
## Random effects:
## Groups Name          Variance Std.Dev.
## species (Intercept) 0.068408 0.26155
## plot (Intercept) 0.001820 0.04266
## year (Intercept) 0.005867 0.07659
## Number of obs: 2028, groups: species, 55; plot, 24; year, 6
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.01945    0.04991 100.577 <2e-16 ***
## statewarmed      -0.02683    0.01781  -1.507  0.1319
## insecticideno_insects 0.03093    0.01781   1.737  0.0824 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) sttwrm
## statewarmed -0.178
## insctcdn_ns -0.178 0.001
```

Because no distributions seems to match well, I'll try a Friedman's test

```
#friedman_kbs <- green_kbs %>%  
# friedman_test(half_cover_date ~ state)
```

Error: Must extract column with a single valid subscript. x Subscript var can't be NA

Can't figure out what this means

If I include the blocks portion of the formula (from the documentation) I get this error

```
#friedman_kbs <- green_kbs %>%  
# friedman_test(half_cover_date ~ state | plot)
```

Error in friedman.test.default(c(141L, 202L, 122L, 101L, 127L, 120L, 197L, : not an unreplicated complete block design

Permanova?

```
per1 <- adonis2(green_kbs$half_cover_date ~ state*year + insecticide, data = green_kbs)  
per1
```

```
## Permutation test for adonis under reduced model  
## Terms added sequentially (first to last)  
## Permutation: free  
## Number of permutations: 999  
##  
## adonis2(formula = green_kbs$half_cover_date ~ state * year + insecticide, data = green_kbs)  
##           Df SumOfSqs      R2      F Pr(>F)  
## state      1    0.051 0.00074  1.5046  0.213  
## year      1    0.487 0.00698 14.2700  0.001 ***  
## insecticide 1    0.168 0.00241  4.9326  0.019 *  
## state:year  1    0.025 0.00036  0.7290  0.375  
## Residual   2023   69.058 0.98951  
## Total      2027   69.790 1.00000  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
per2 <- adonis(formula = green_kbs$half_cover_date ~ state*year + insecticide, strata = green_kbs$plot,  
per2
```

```
##  
## Call:  
## adonis(formula = green_kbs$half_cover_date ~ state * year + insecticide, data = green_kbs, strata = green_kbs$plot,  
##  
## Blocks: strata  
## Permutation: free  
## Number of permutations: 999  
##  
## Terms added sequentially (first to last)
```



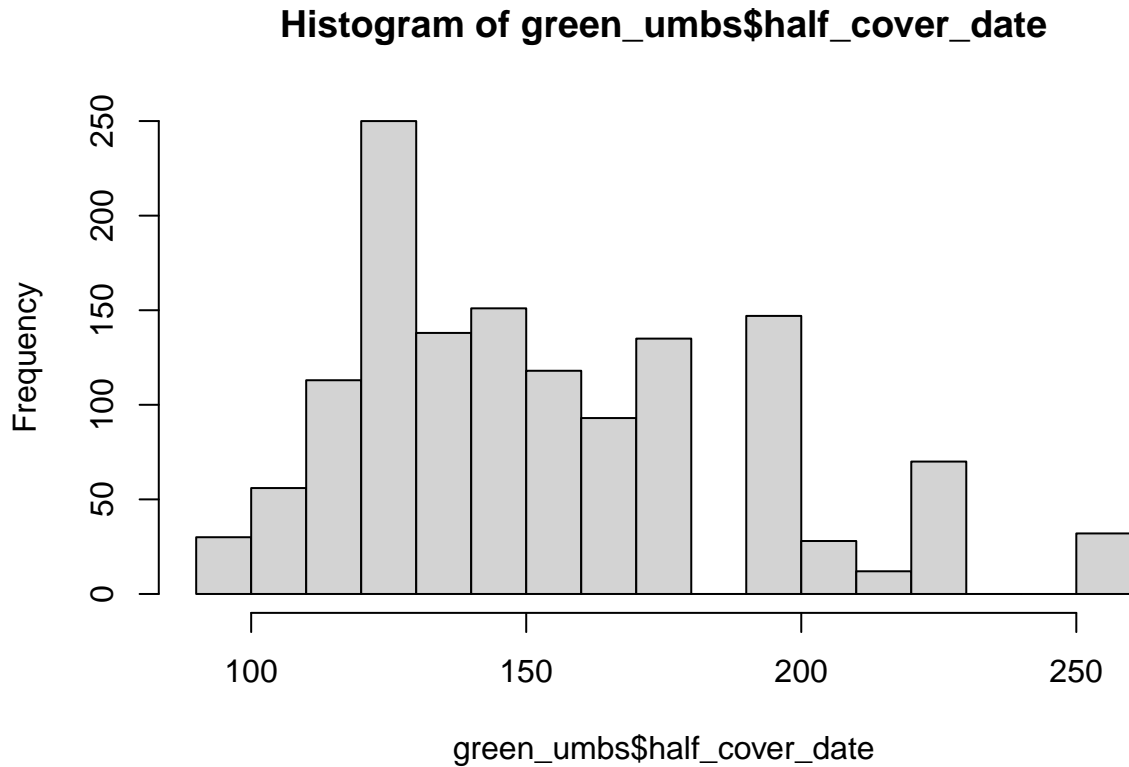
```
##
##              Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## state         1      0.051 0.05136  1.5046 0.00074  0.001 ***
## year          1      0.487 0.48713 14.2700 0.00698  0.001 ***
## insecticide    1      0.168 0.16838  4.9326 0.00241  0.988
## state:year     1      0.025 0.02489  0.7290 0.00036  0.396
## Residuals    2023     69.058 0.03414          0.98951
## Total        2027     69.790          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With per2, when controlling for “plot”, there is a difference btwn treatments

## UMBS

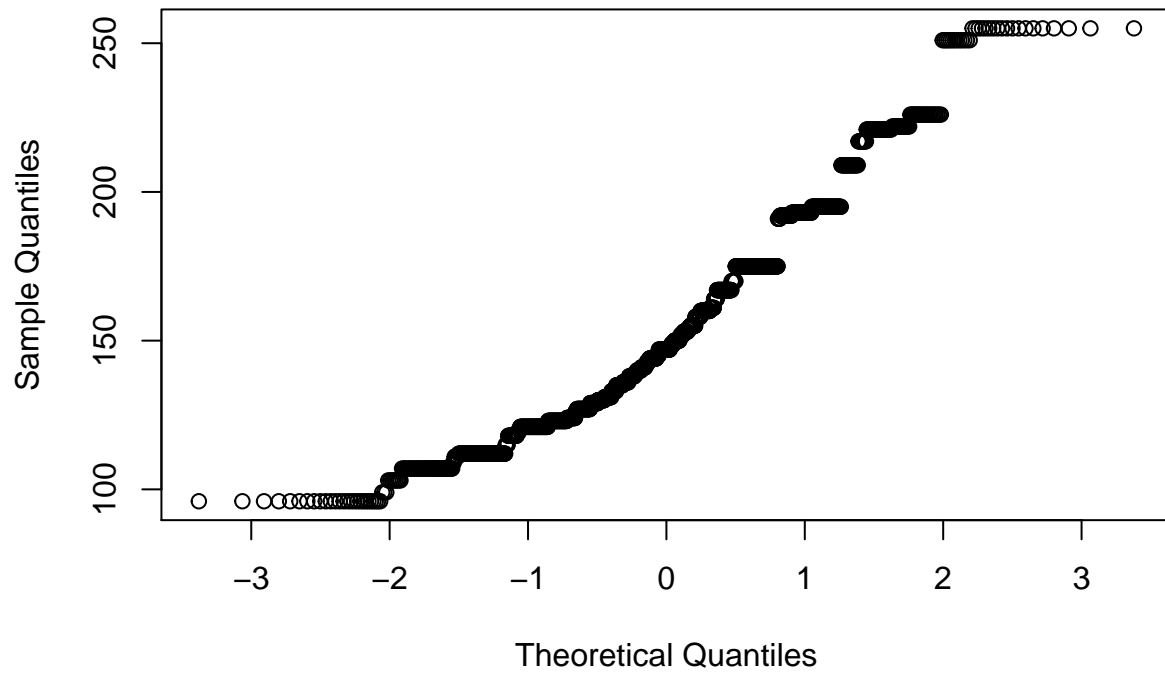
### Checking for normality

```
hist(green_umbs$half_cover_date)
```



```
qqnorm(green_umbs$half_cover_date)
```

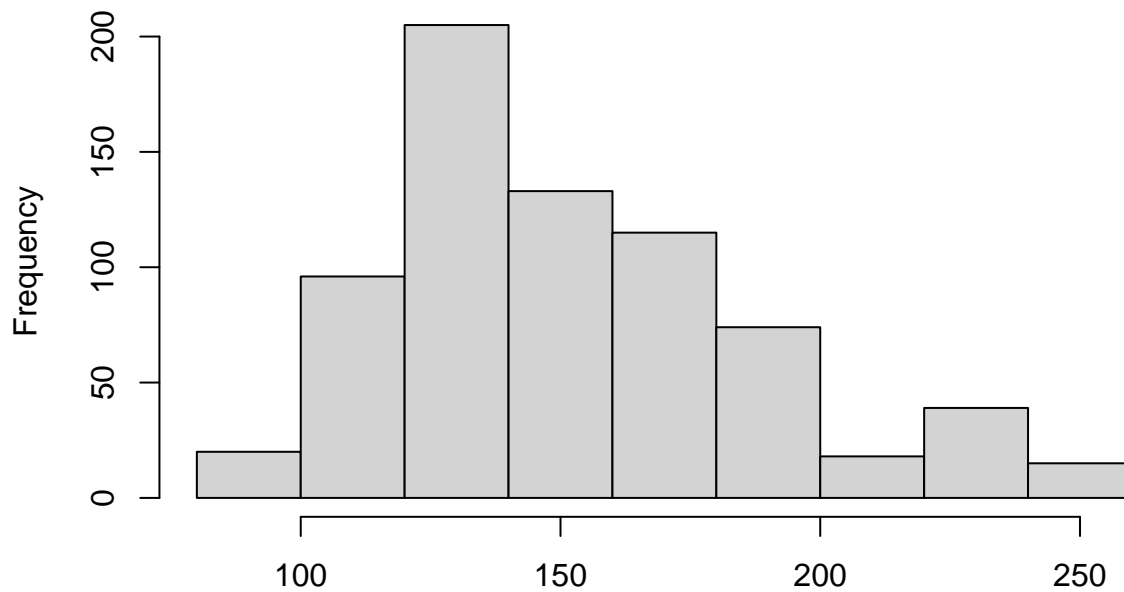
## Normal Q-Q Plot



```
shapiro.test(green_umbs$half_cover_date)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  green_umbs$half_cover_date  
## W = 0.94754, p-value < 2.2e-16  
hist(green_umbs$half_cover_date[green_kbs$state == "ambient"])
```

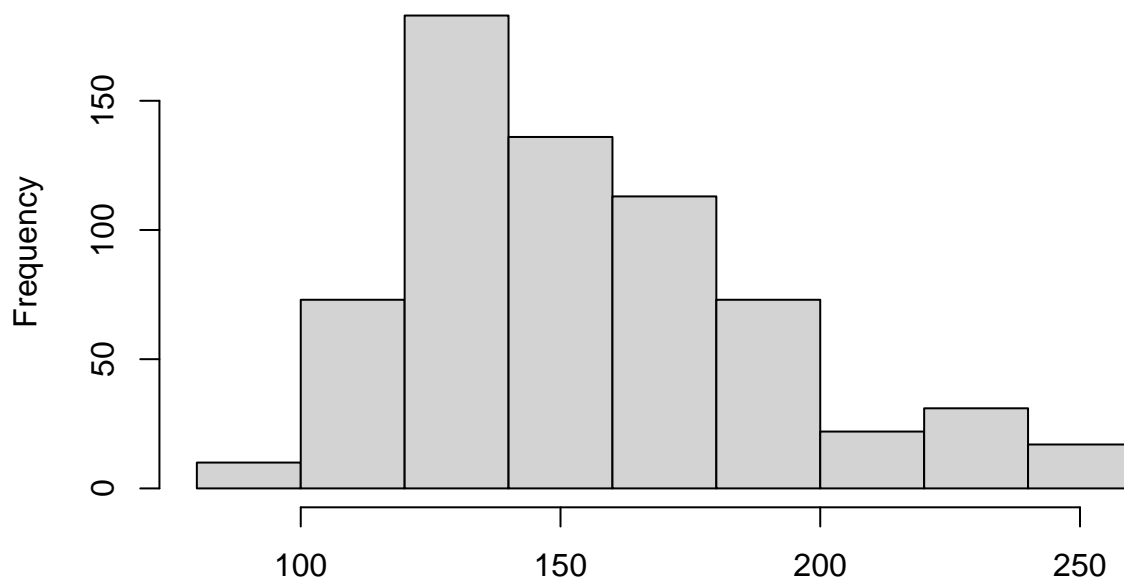
**Histogram of `green_umbs$half_cover_date[green_kbs$state == "ambient"]`**



`green_umbs$half_cover_date[green_kbs$state == "ambient"]`

```
hist(green_umbs$half_cover_date[green_kbs$state == "warmed"])
```

**Histogram of `green_umbs$half_cover_date[green_kbs$state == "warmed"]`**



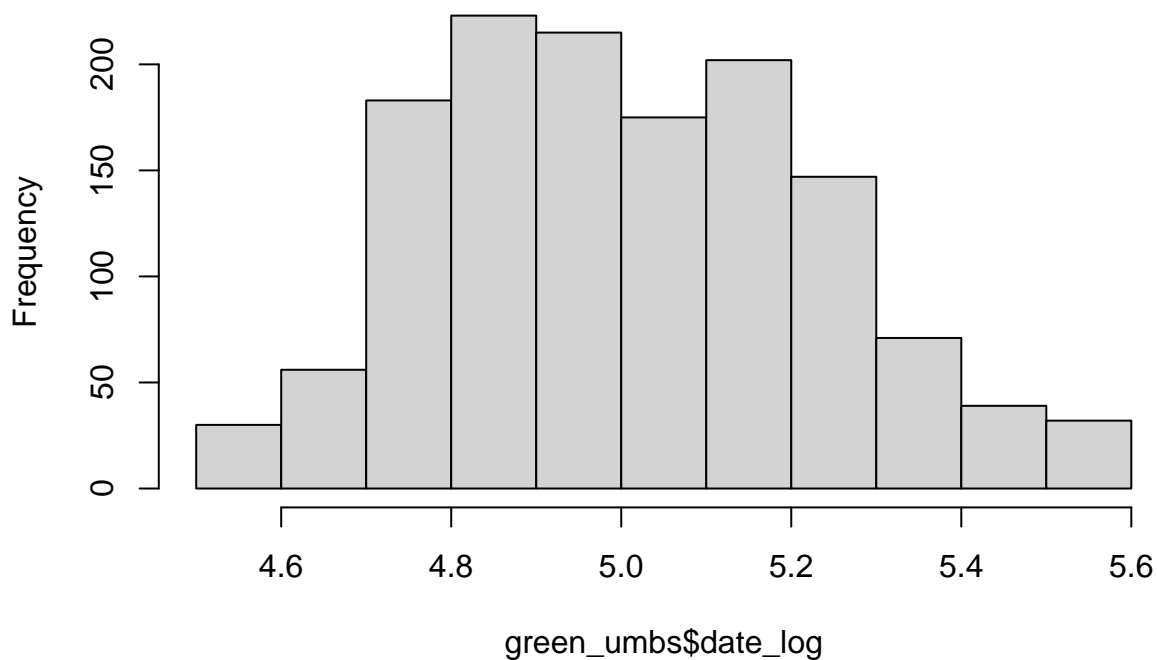
`green_umbs$half_cover_date[green_kbs$state == "warmed"]`

These look pretty good

## Trying log transformation

```
green_umbs$date_log <- log(green_umbs$half_cover_date)
hist(green_umbs$date_log)
```

**Histogram of green\_umbs\$date\_log**



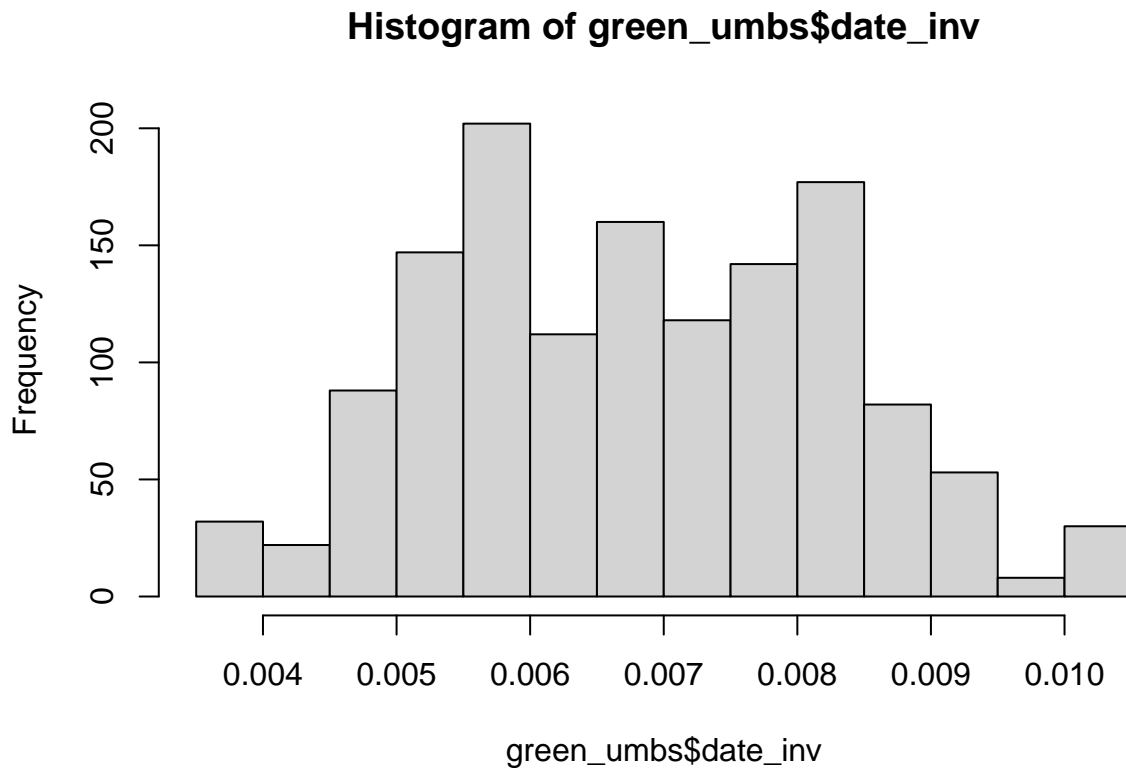
```
shapiro.test(green_umbs$date_log)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  green_umbs$date_log
## W = 0.97728, p-value = 6.765e-14
```

I think this looks good but shapiro-wilk is lower than 0.05

## Trying inverse tranformation

```
green_umbs$date_inv <- 1/(green_umbs$half_cover_date)
hist(green_umbs$date_inv)
```



```
shapiro.test(green_umbs$date_inv)

##
##  Shapiro-Wilk normality test
##
## data:  green_umbs$date_inv
## W = 0.97993, p-value = 6.679e-13
```

This also looks good but is also still low for shapiro-wilk