Linear Mixed Model for Ponderosa Pine

Courtney Giebink

August 6, 2019

Data Summary

```
load(file = 'C:/Users/clgie/OneDrive/Documents/Masters/Utah/UT_FVS/data/formatted/data_all_pp')

data_all_pp <- data_all_pp %>%
  mutate(dds = (2*RW*0.0393701)^2) #mm to inches
```

My data comes from the U. S. Forest Service's Interior West - Forest Inventory and Analysis (IW-FIA) program (DeRose, Shaw, and Long 2017). The U.S. Forest Service's Forest Inventory and Analysis program has established plots across the nation, which they use to collect data on forest stands. Recently, tree rings associated with plots in the Interior West states were discovered in storage. The IW-FIA program merged these two data sources: inventory data and tree-ring data. The data that I will use for my analysis is inventory and tree-ring data for *Pinus ponderosa* (Ponderosa pine) in Utah. I was able to extract and calculate annual data for 51 trees between the years of 1962 and 1995.

```
#create new dataframe with only variables needed for model
#response: RW, dds
#fixed: SI, ASP, SL, DBH/DIA_C, BAL, CR, CCF, PCCF, climate
#random: TRE_CN, Year
#filter for last 30 years of growth
min(data_all_pp$MEASYEAR) #1992 -> 1962
```

```
## [1] 1992
```

```
glmm_data_pp <- data_all_pp %>%
  select(PLT_CN,TRE_CN, RW, dds, Year, DIA_C,
         SICOND, ASPECT, SLOPE, BAL, CR, CR_weib, PCCF, CCF,
         ppt_pDec, ppt_Jun, ppt_Jul,ppt_pOct,
        ppt_pAugOct, ppt_pOctDec, ppt_pNovJan, ppt_MayJul,
         ppt_pJunJan, wateryr, #JunJan should be AugJan
         tmax_Jun,
         tmax_JunAug) %>%
  filter(Year >= 1962)

#climate
#total ppt
#1 month: ,pDec, Jun, Jul, pOct
#3 month: pAug-pOct, pOct-pDec, pNov-Jan, May-Jul
#6 month + : Jan, wateryr

#average temp
#1 month: Jun tmax
#3 month: tmax_Jun-Aug
```

My data, `glmm_data_pp`, consists of raw radial increment, or ring width value, (`RW`) in mm and annual change in squared inside bark diameter (dds) in inches^2 with corresponding tree, climate, and site variables. Some variables stay constant, such as site index (`SICOND`), aspect (`ASPECT`), and slope (`SLOPE`), while diameter at breast height (`DIA_C`) and crown ratio (CR, CR_weib) vary annually. In addition, density variables, being basal area of trees larger than the subject tree (`BAL`), crown competition factor on the inventory point (`PCCF`), and stand crown competition factor (`CCF`), vary annually. Significant seasonal precipitation and temperature variables were identified based on tradition dendro analysis tools, using R packages `dplR` and `treeclim` (see `Climate-growth.Rmd`)

Exploration

It is necessary to explore the dataset to determine if the data complies with model assumptions (Zuur, Ieno, and Elphick 2010).
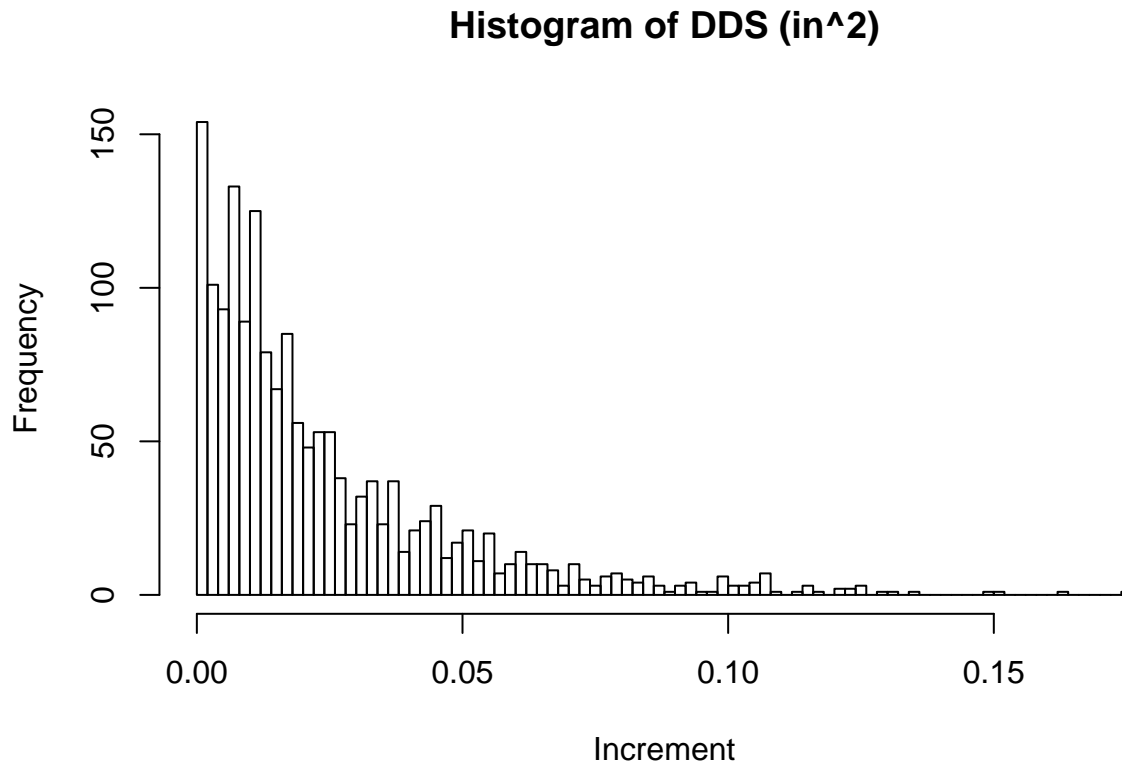
Normality

The response variable, annual change in squared inside bark diameter (DDS), is continuous. It will be explained using a linear mixed model.

```
hist(glmm_data_pp$RW,breaks = 50,
     main = "Histogram of RW (mm)", xlab = "Increment")
```

Histogram of RW (mm)
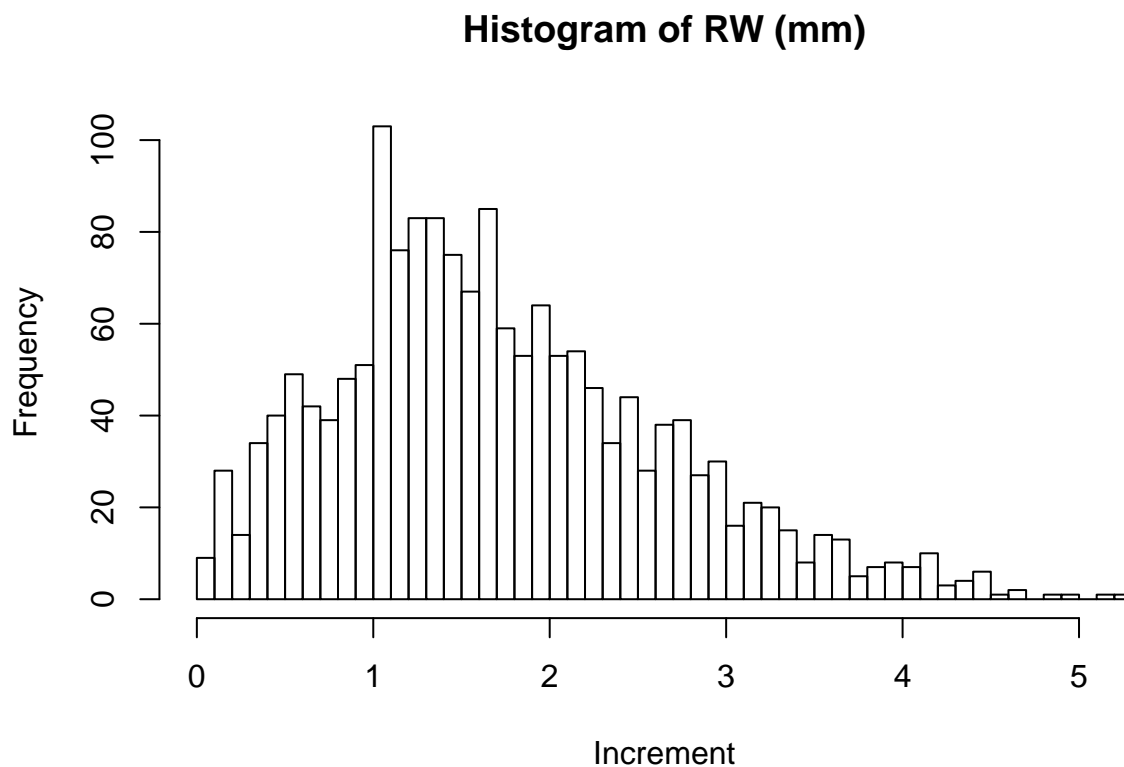
```r
hist(glmm_data_pp$dds,breaks = 100,
     main = "Histogram of DDS (in^2)", xlab = "Increment")
```

Histogram of DDS (in^2)

The distribution of DDS is highly skewed right, whereas the distribution of raw radial increment (RW) is slightly skewed right. The Gamma distribution is often used for skewed data, although Gaussian might still be appropriate and easier to interpret.

```r
which(glmm_data_pp$RW == 0) #0
```
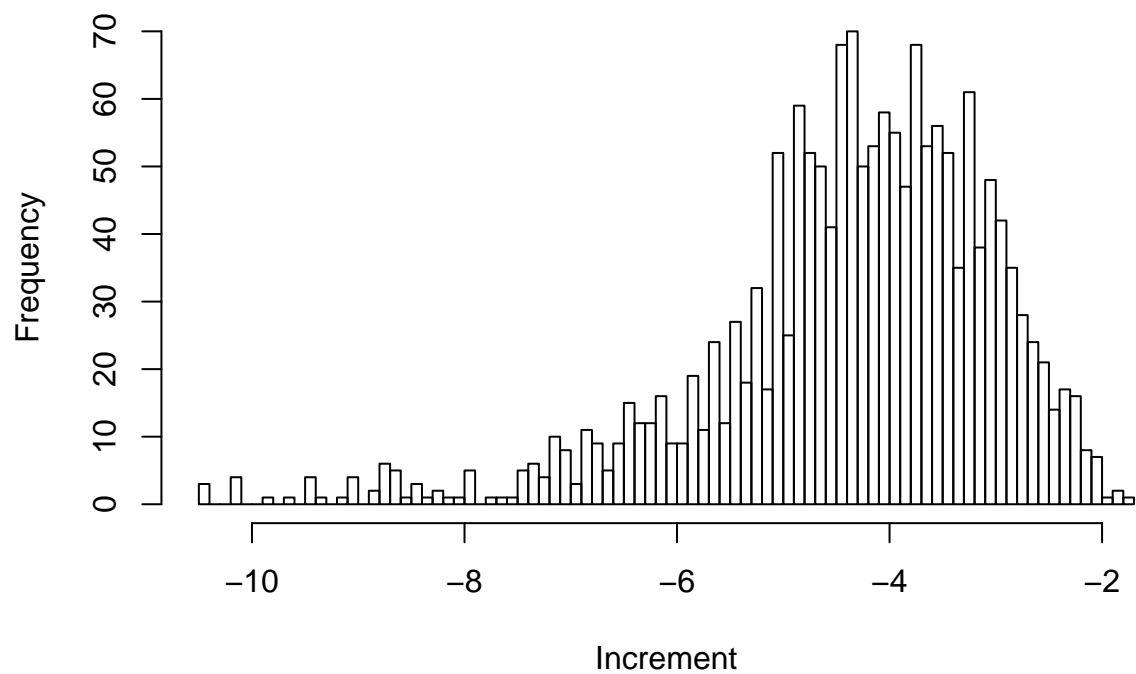
```
## integer(0)
```

```r
#no missing rings
which(is.na(glmm_data_pp$RW)) #0
```
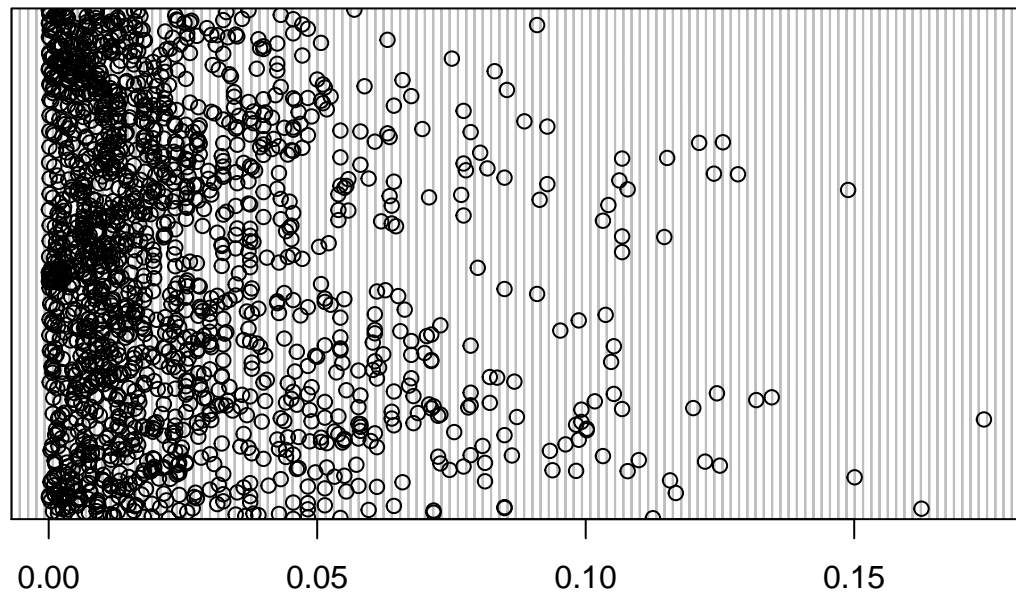
```
## integer(0)
```

```r
hist(log(glmm_data_pp$dds),breaks = 100,
     main = "Histogram of log(DDS)", xlab = "Increment")
```
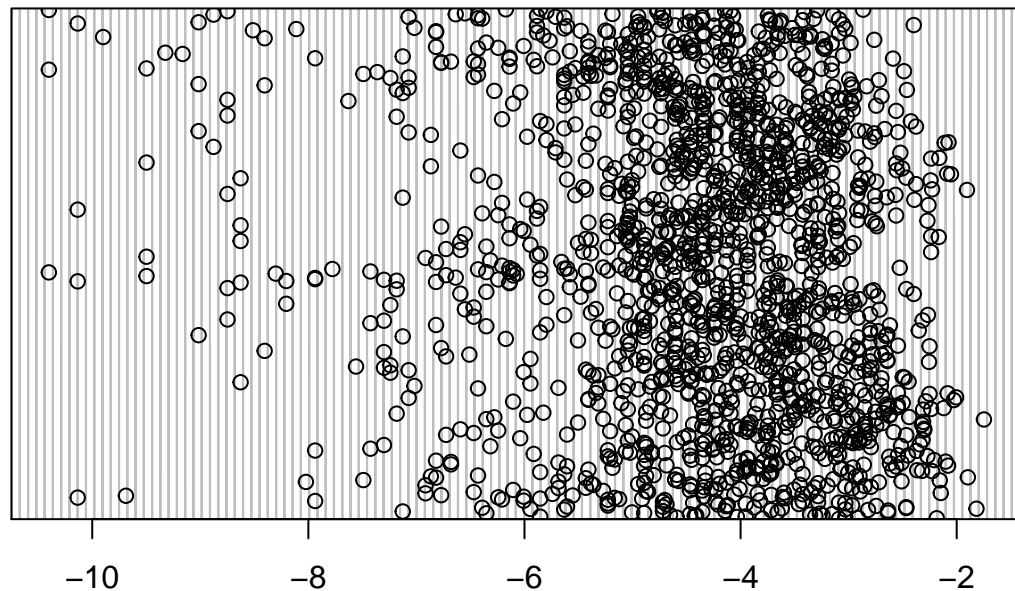
Histogram of log(DDS)


```r
#Cleveland dotplot
dotchart(glmm_data_pp$dds)
```

```r
dotchart(log(glmm_data_pp$dds))
```

A log transformation of dds skews the distribution to the left and doesn't get rid of outliers.

Zeros and Missing Data

```
length(unique(glmm_data_pp$TRE_CN[glmm_data_pp$Year == 1962]))
```

```
## [1] 51
```

```
#[1] 51
```

My data is unbalanced. All trees have a growth measurement for 1962, but the last year of growth is not the same for all trees. Most trees end growth in 1994, but some end in 1992, 1993, or 1995. There are no missing data in between first and last year of growth for each tree, and there are no years of zero growth, indicating a missing ring.

Random and Fixed Effects

In the linear mixed effects model, tree, stand, and climate variables will be fixed effects and a tree identifier and year of growth will be random effects.

```
length(unique(data_all_pp$TRE_CN))
```

```
## [1] 51
```

```
length(unique(data_all_pp$PLT_CN))
```

```
## [1] 43
```

There are almost as many unique plots as unique trees, so it is not necessary to have plots as a random effect as well.

Tree ID and year are appropriate random effects because selected trees and years are a subset from all the possible selected trees and years (Harrison et al. 2018). Each tree has its own intercept of growth due to factors not considered here, such as genetics, microclimate, etc. Each year has its own intercept of growth due to relatively good or poor conditions not accounted for in temperature and precipitation, such as cloud cover, vapor pressure deficit, etc. Each tree and year might also have its own slope, but convergence on a random slope effect can only occur with an adequate amount of data.
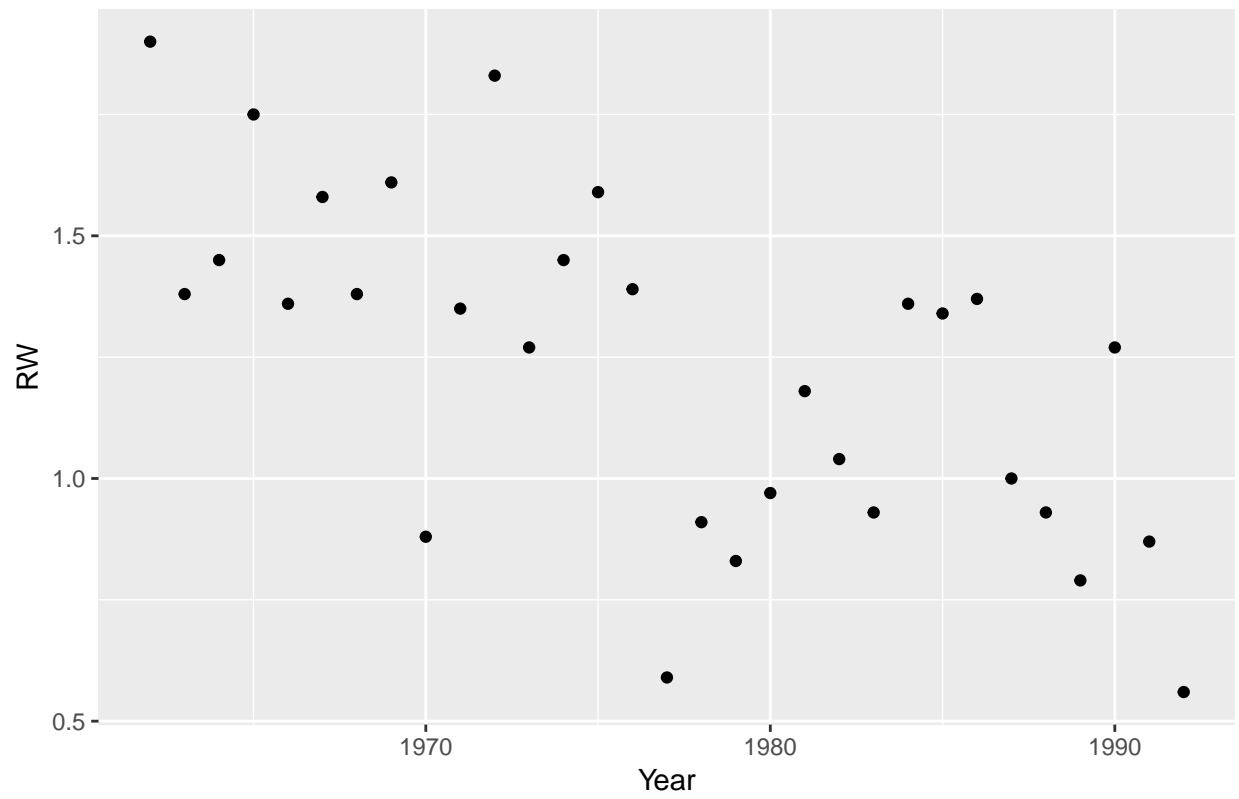
```
#understanding random effects: TRE_CN
growth_yr_plots_pp <- glmm_data_pp %>%
  group_by(TRE_CN) %>%
  do(plots=ggplot(data=.) +
       aes(x=Year, y=RW) + geom_point() + ggtitle(unique(.$TRE_CN)))

growth_yr_plots_pp$plots[[1]]
```
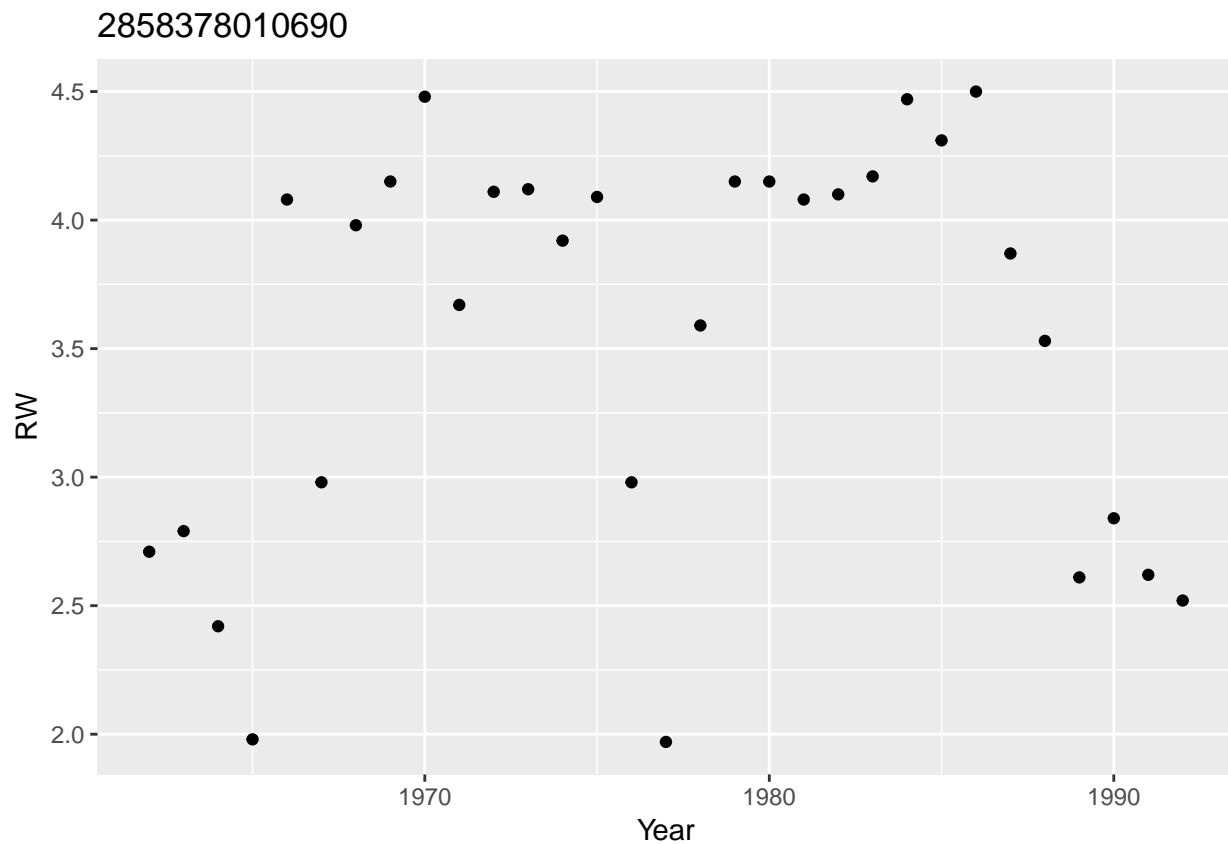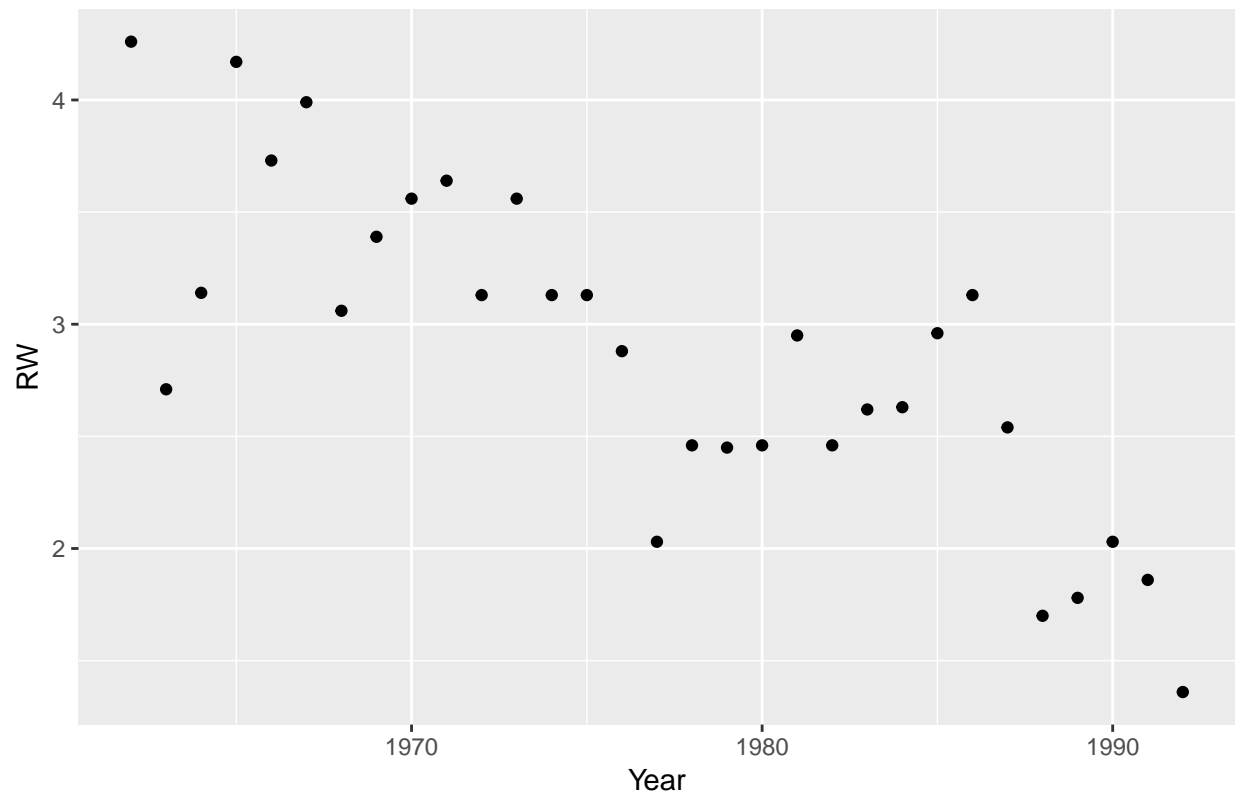


```
growth_yr_plots_pp$plots[[2]]
```

2858391010690


```
growth_yr_plots_pp$plots[[3]]
```
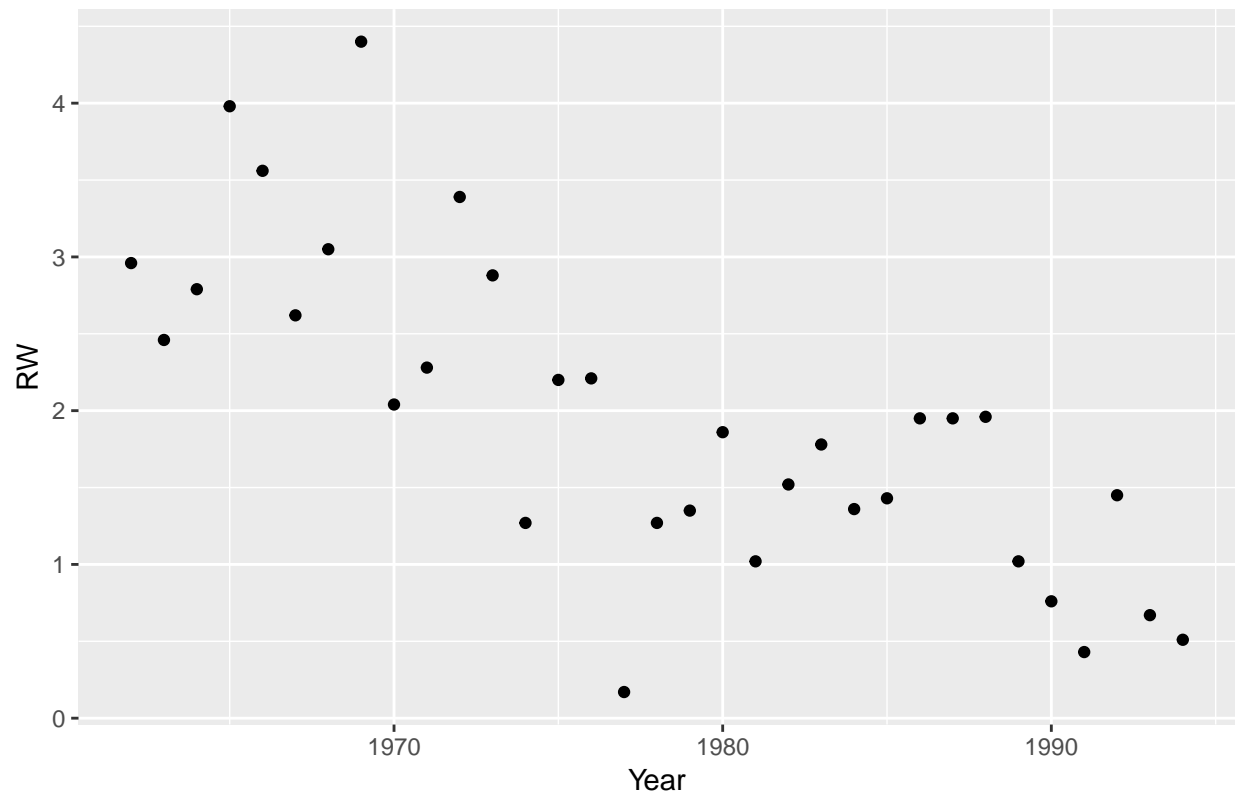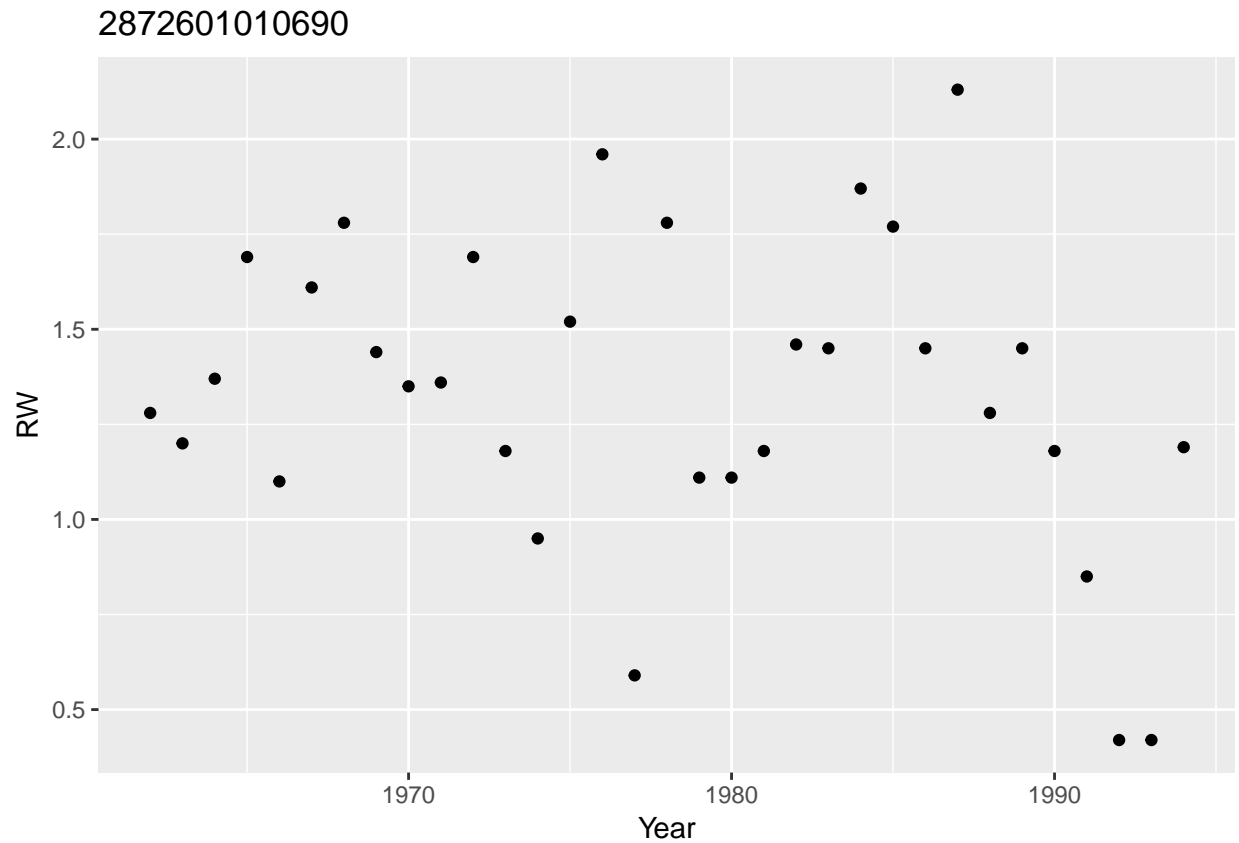
2858393010690
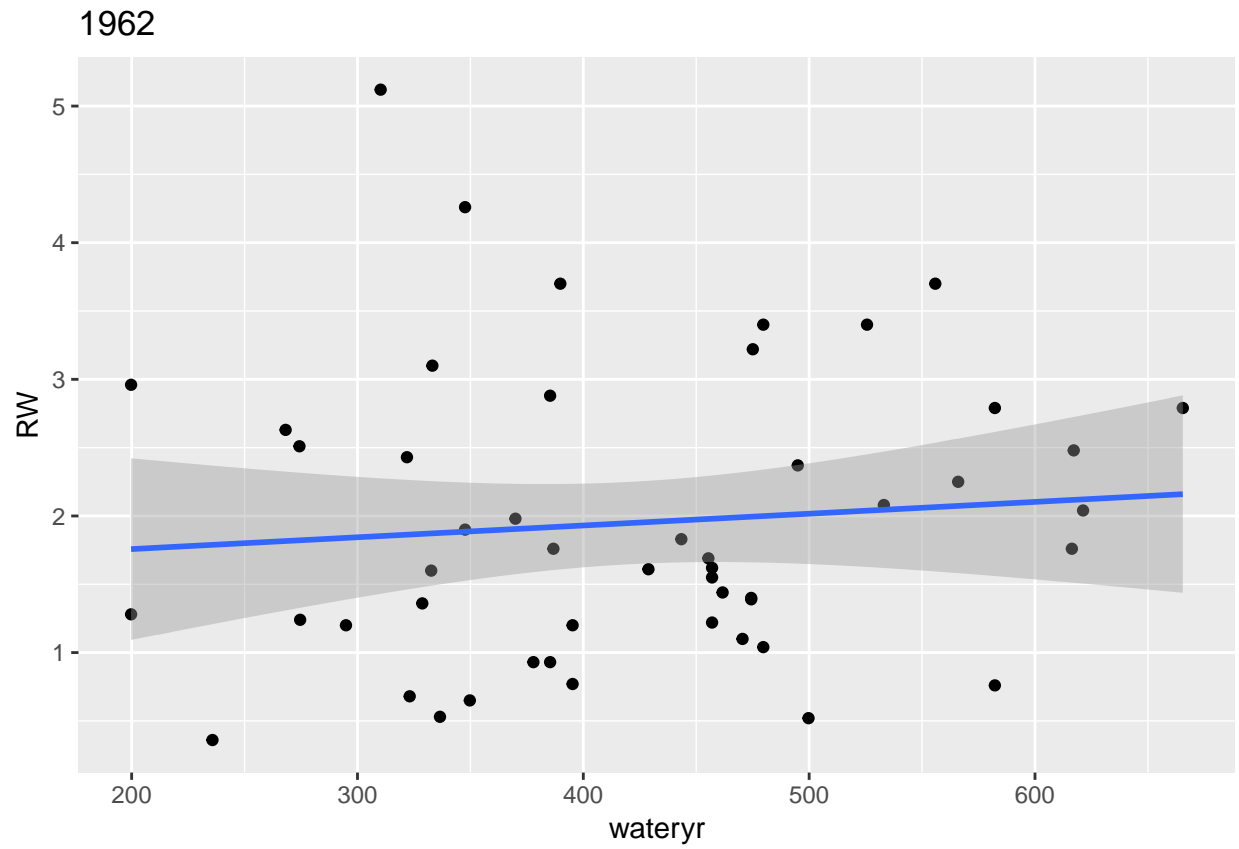


```
growth_yr_plots_pp$plots[[4]]
```

2872599010690

```
growth_yr_plots_pp$plots[[5]]
```
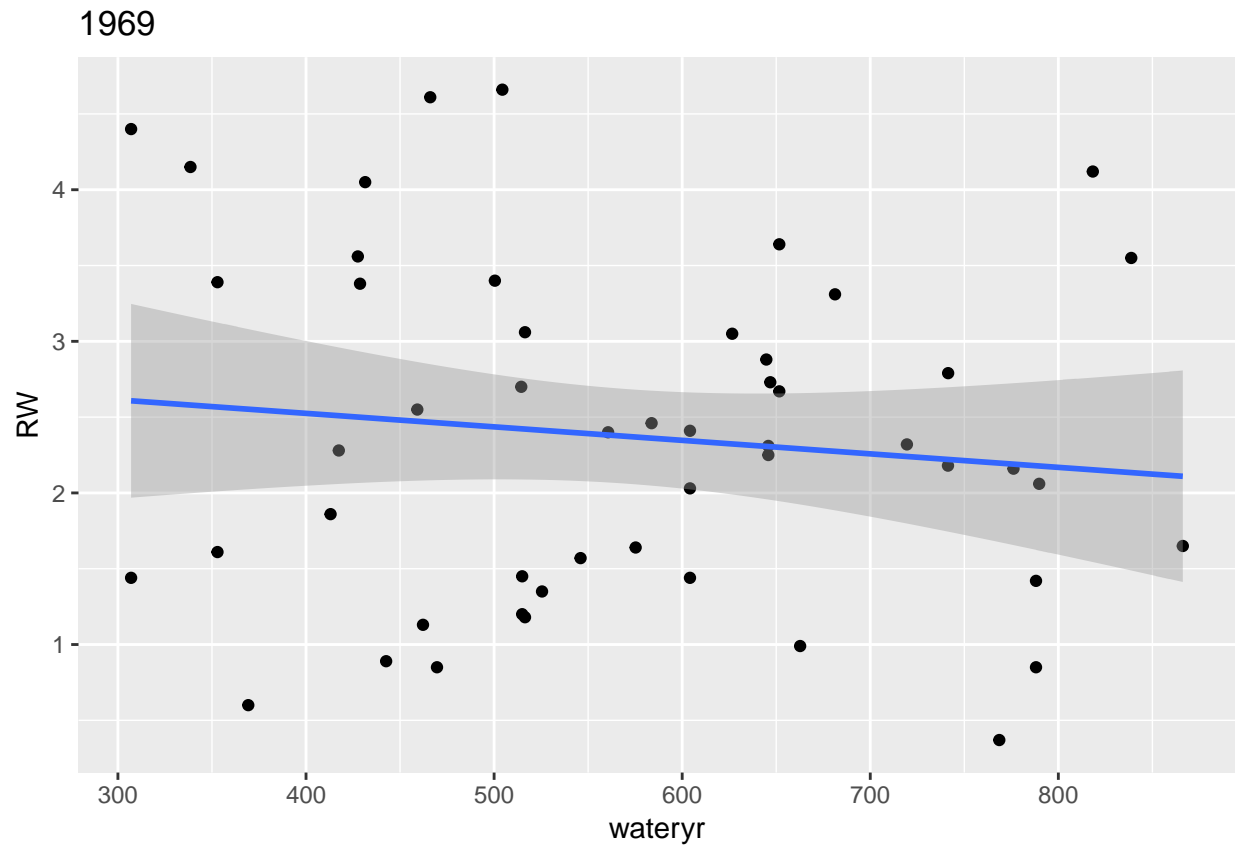
2872601010690

```
#can do more all the way to tree 51
```

```
#understanding random effects: Year
growth_wateryr_plots_pp <- glmm_data_pp %>%
  group_by(Year) %>%
  do(plots=ggplot(data=.) +
     aes(x=wateryr, y=RW) + geom_point() +
     geom_smooth(method = "lm") + ggtitle(unique(.$Year)))

growth_wateryr_plots_pp$plots[[1]]
```
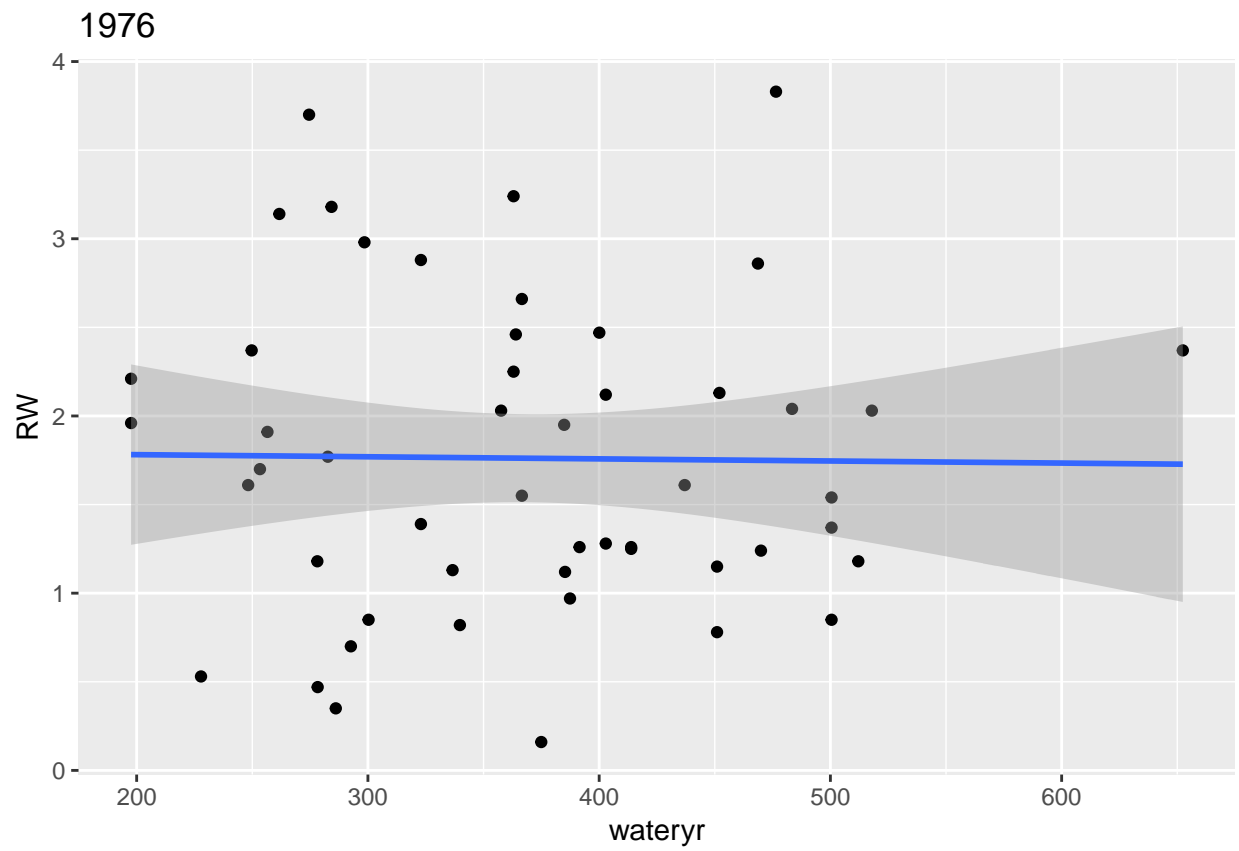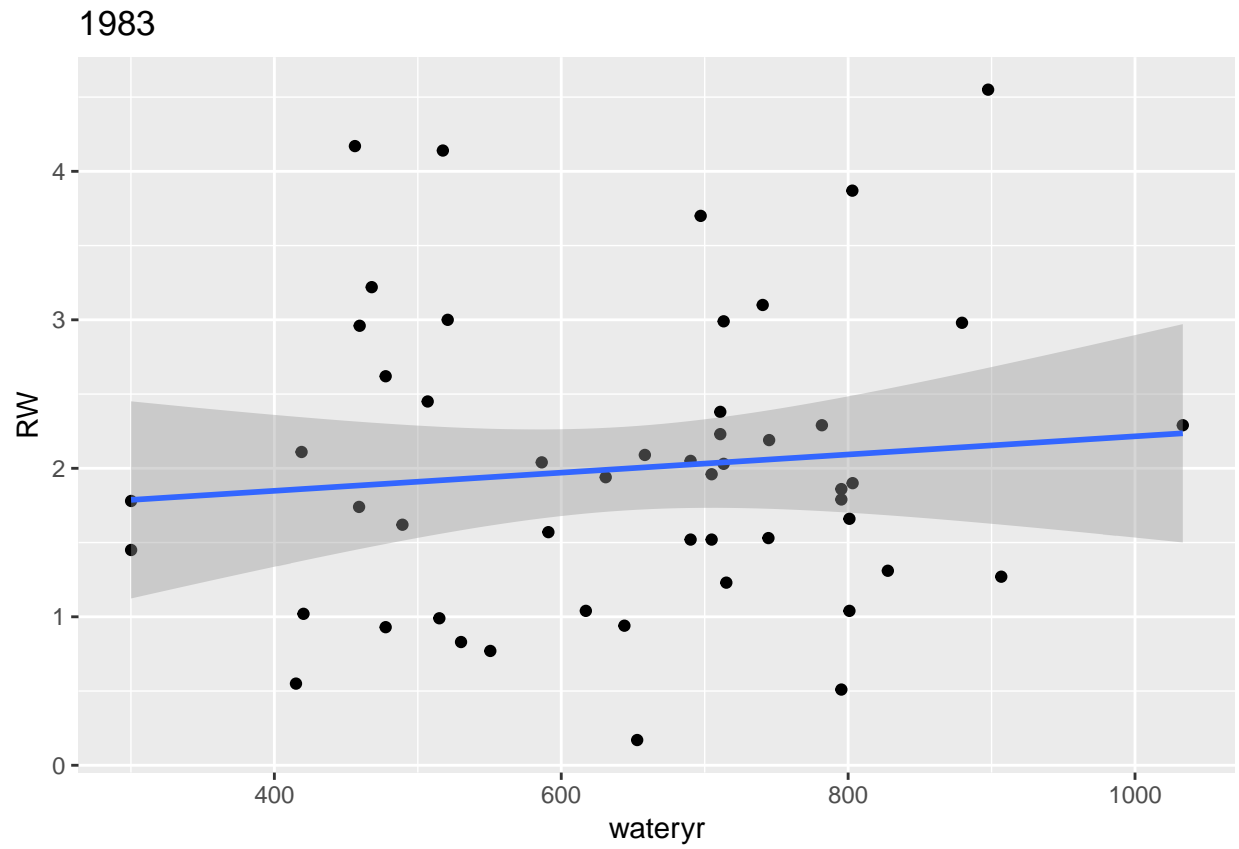
1962

```
growth_wateryr_plots_pp$plots[[8]]
```
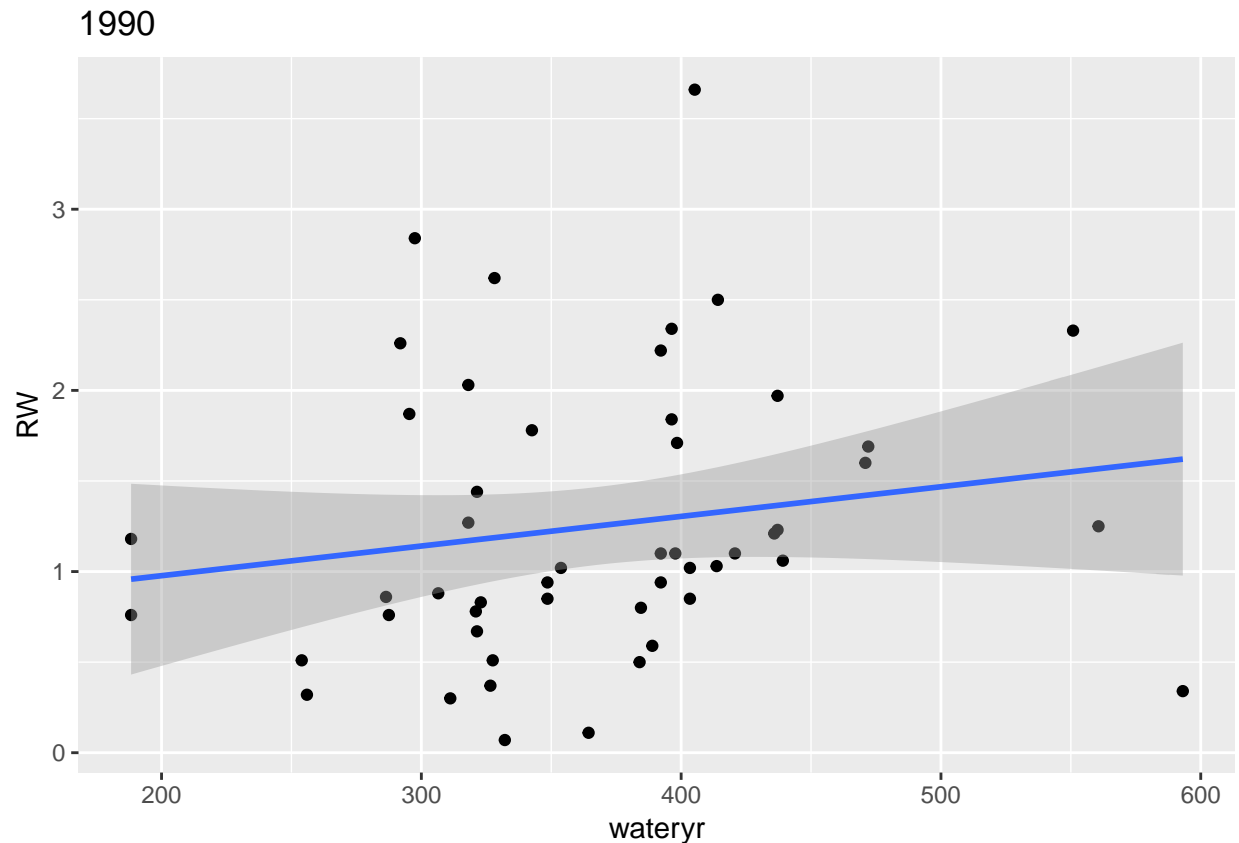
```
growth_wateryr_plots_pp$plots[[15]]
```

1976

```
growth_wateryr_plots_pp$plots[[22]]
```

1983


```
growth_wateryr_plots_pp$plots[[29]]
```

1990

```
#can all years 1962 - 1995
```

A proxy for age (time) is diameter at breast height, which is continuous. As a tree gets older, it grows, puts on biomass, and increases its diameter. While a tree could put on the same amount of biomass each year, the biomass will be stretched over an increasing larger circumference. As such, diameter increment decays as a tree ages, or as diameter at breast height increases. This age trend has been clearly described in tree-ring science. Below is a plot of age (dbh) by my response (incr), which is fit with a linear trend:

```
#detrend
ggplot(data = glmm_data_pp, aes(x = DIA_C, y = RW)) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```
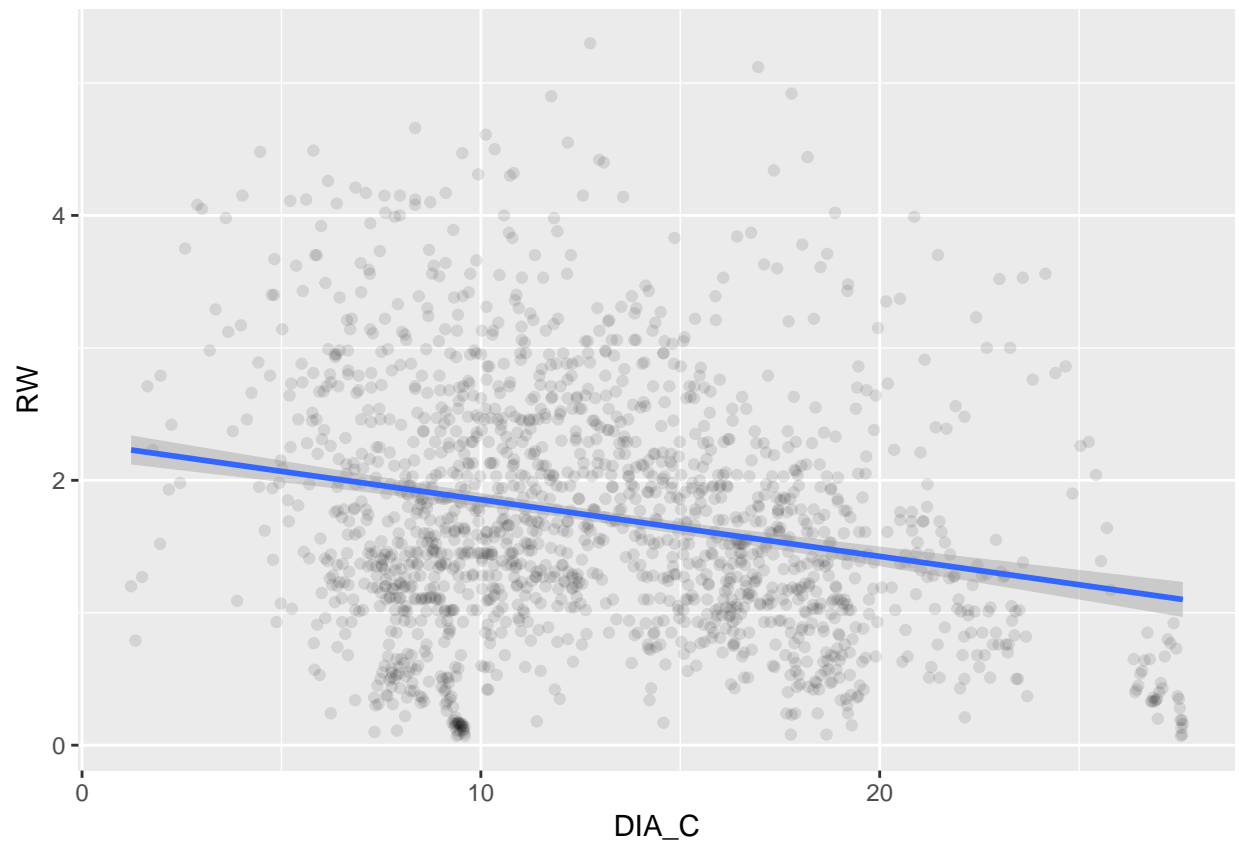
```
ggplot(data = glmm_data_pp, aes(x = DIA_C, y = dds)) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```
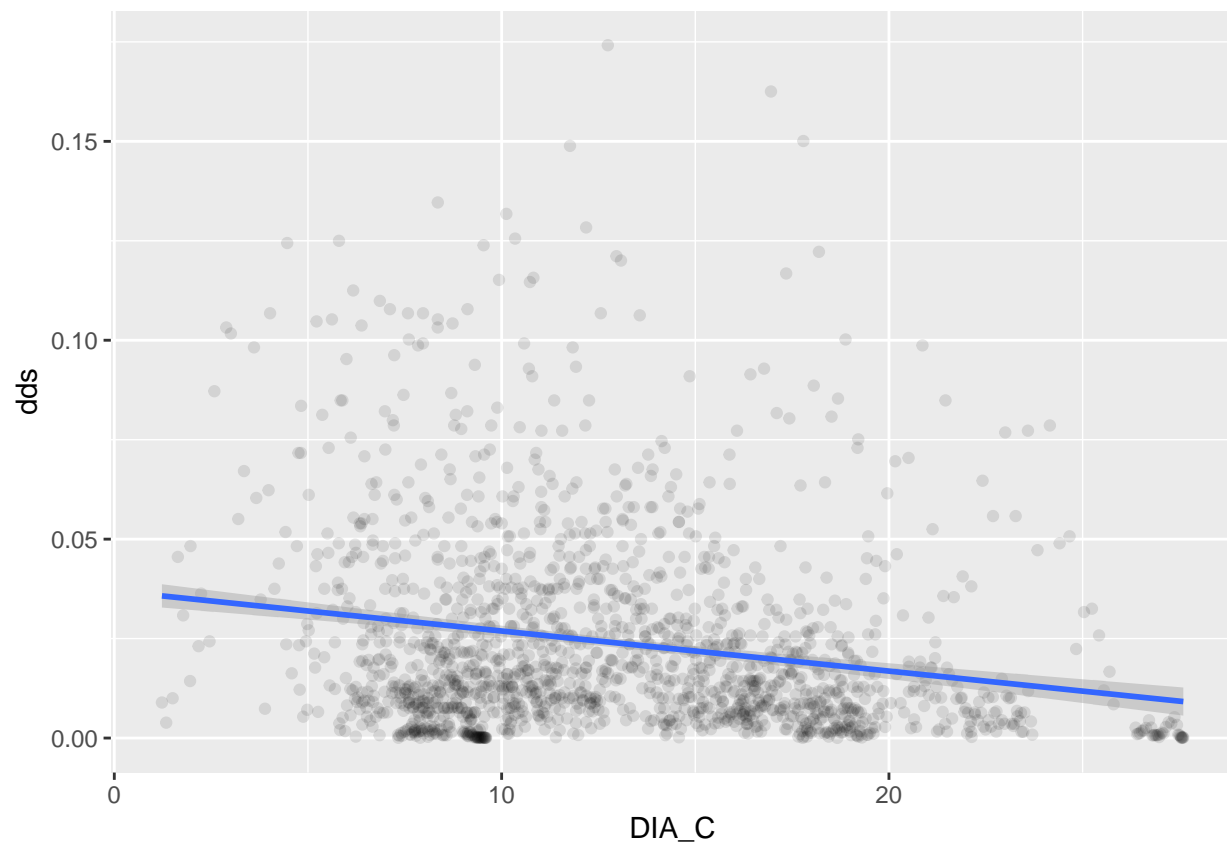
```
#variance relatively constant

ggplot(data = glmm_data_pp, aes(x = DIA_C, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```
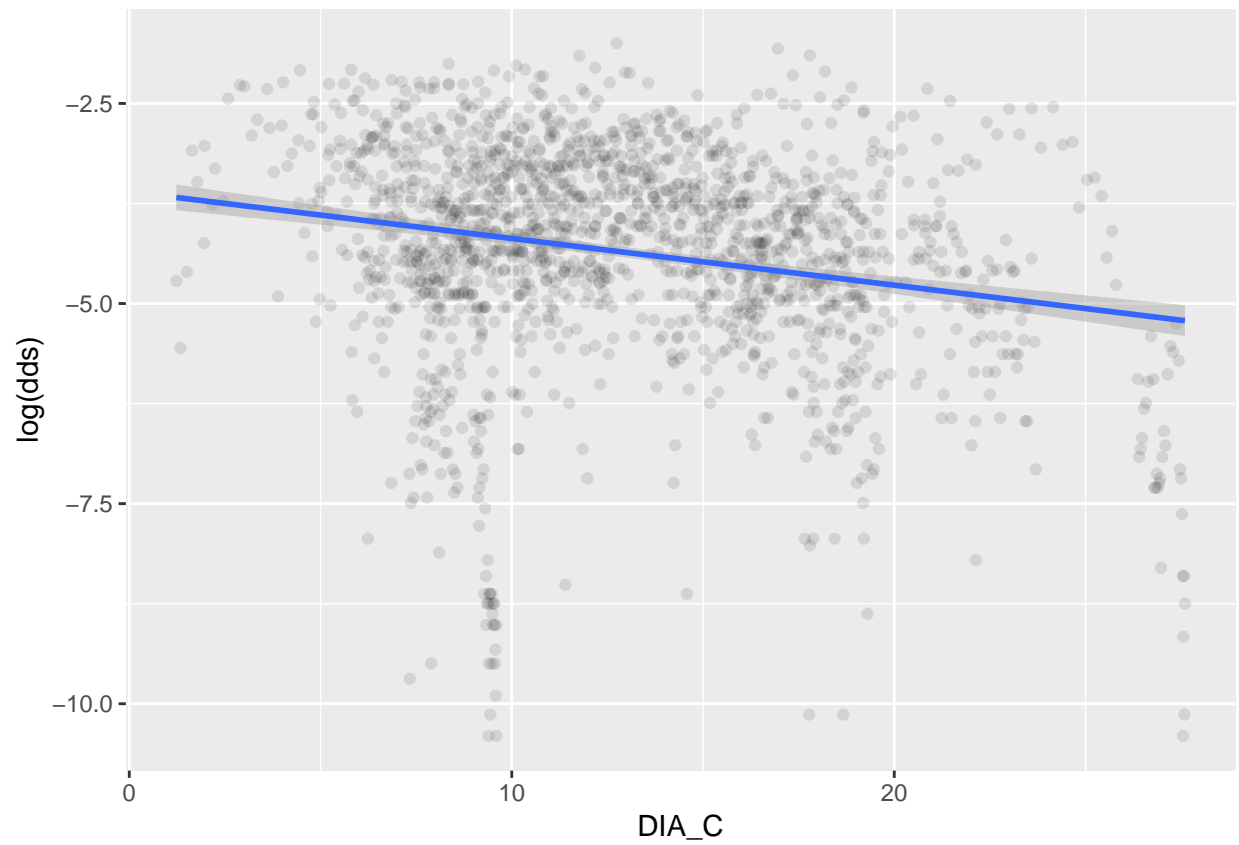
```r
#log to reduce heteroscedasticity
ggplot(data = glmm_data_pp, aes(x = log(DIA_C), y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

```
#log makes it worse
#negative relationship
```

Site index is a measure of site quality. It is the estimated height in feet of a tree on the site at a specific age (FIA Database Description and User Guide for Phase 2).
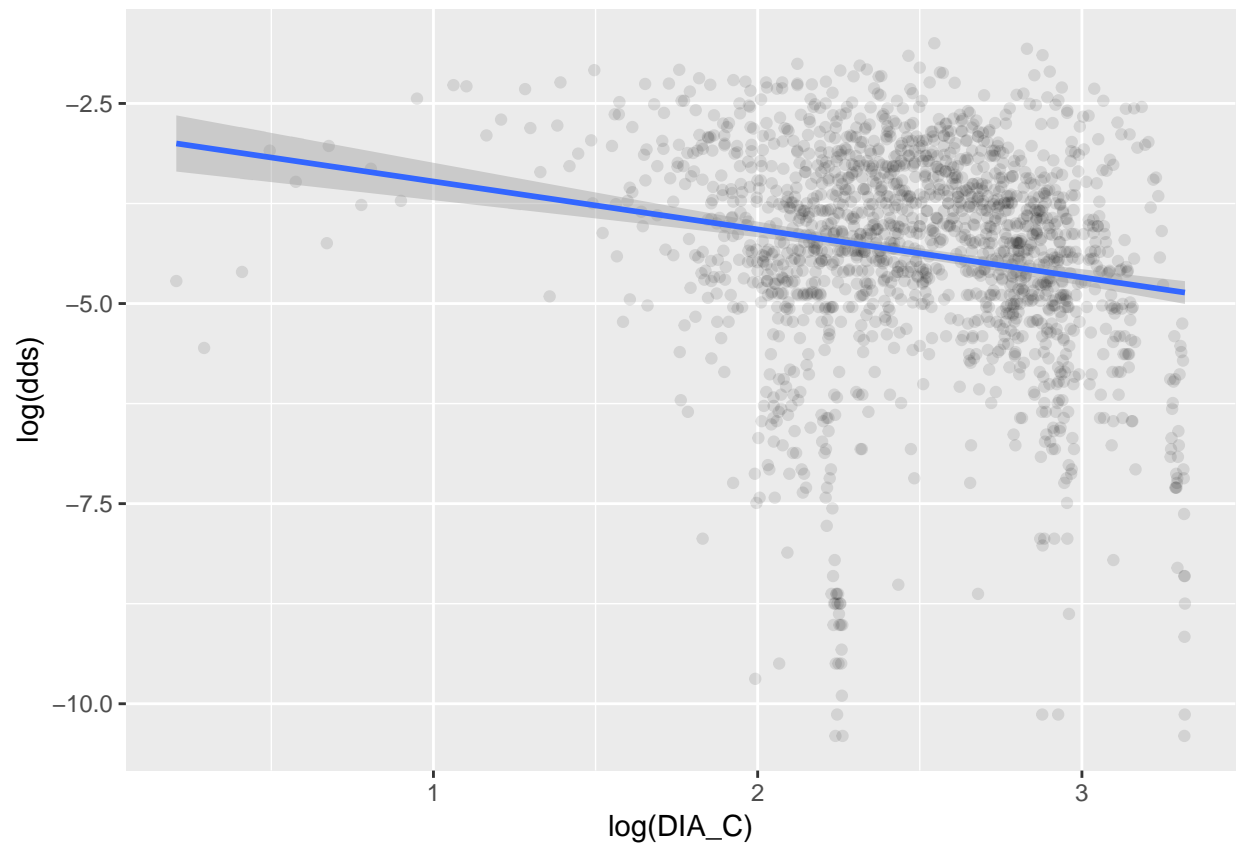
```
ggplot(data = glmm_data_pp, aes(x = SICOND, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

There is a slight postive relationship between growth and site index, meaning growth increases with increasing site quality. There is also relatively constant variance.

Slope is an estimated average at the subplot level, or condition level (FIA Database Description and User Guide for Phase 2).

```
ggplot(data = glmm_data_pp, aes(x = SLOPE, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

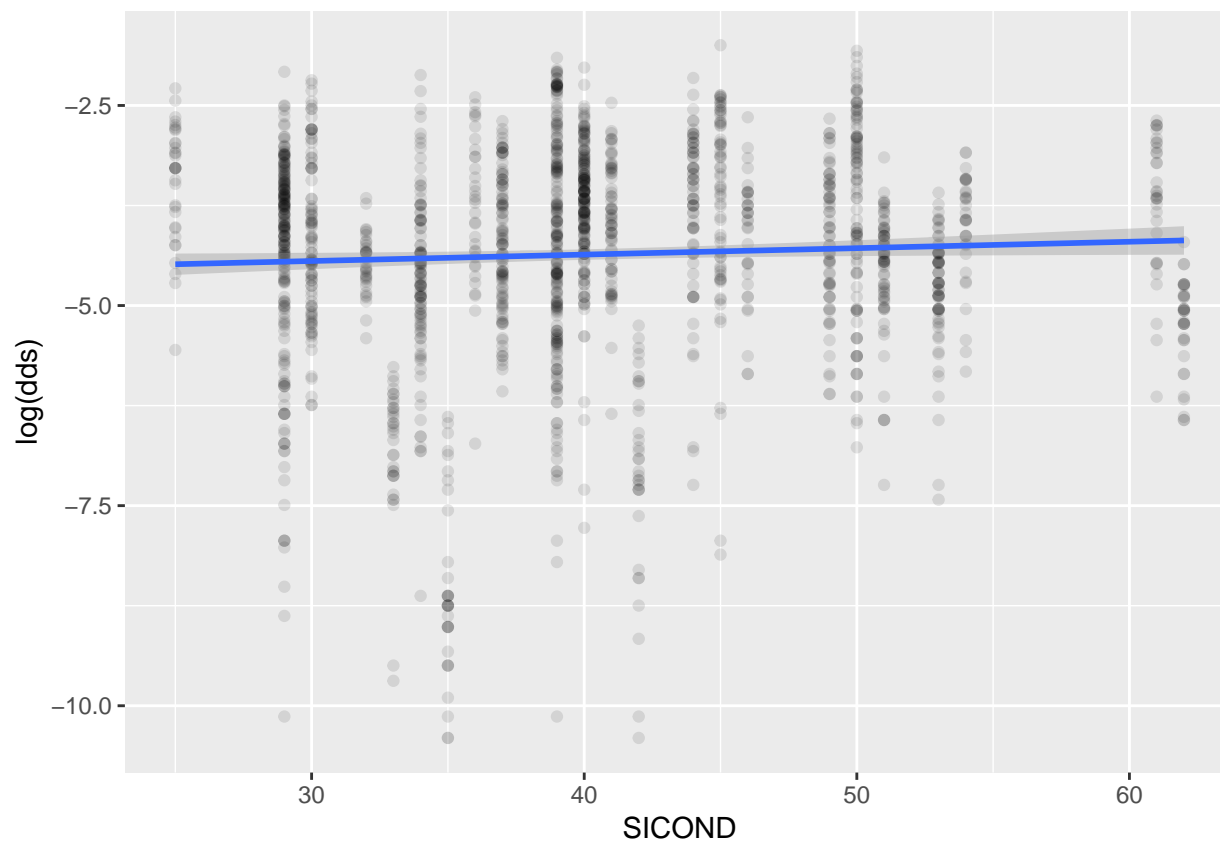There is a negative relationship between growth and slope, meaning growth decreases with increasing slope. There is also relatively constant variance.

Basal area of trees larger than the subject tree (BAL) is a calculated covariate at the plot level.
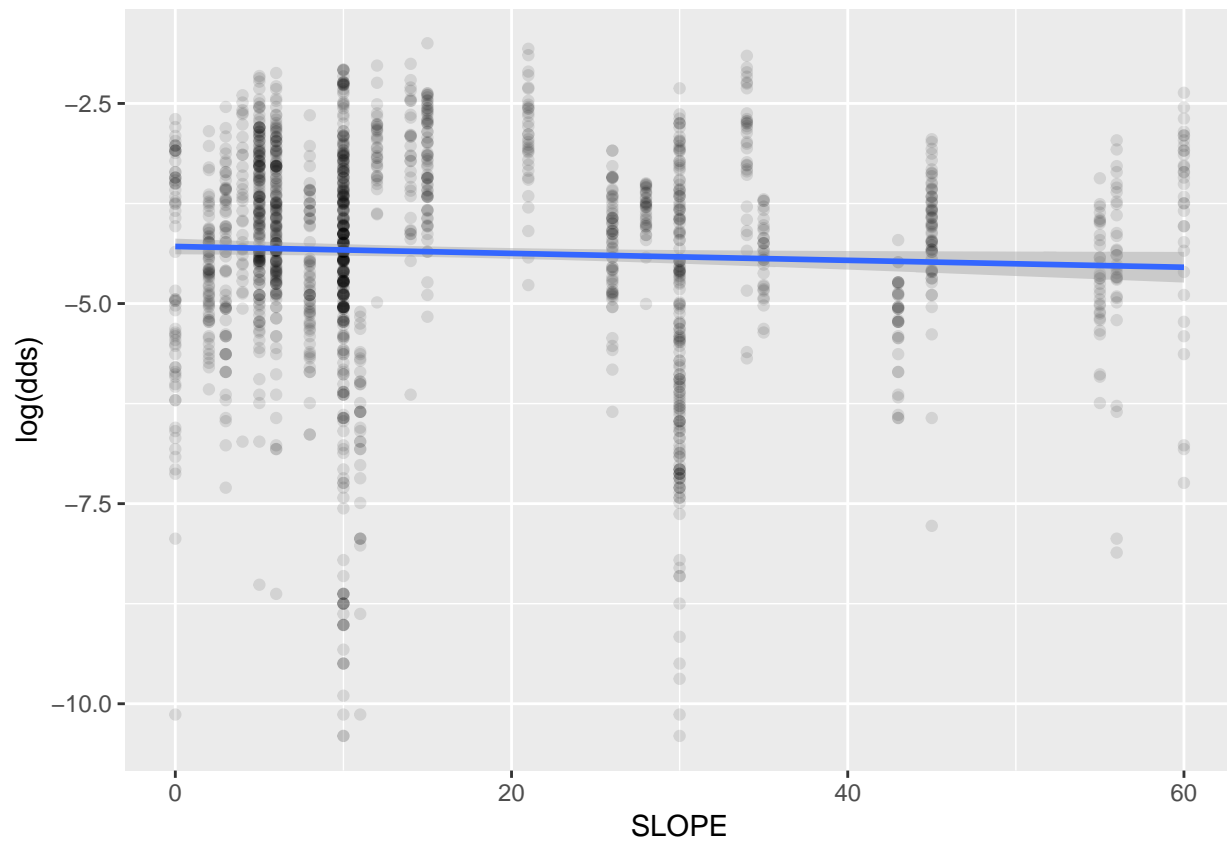
```r
ggplot(data = glmm_data_pp, aes(x = BAL, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

There is a negative relationship between growth and BAL, and there is relatively constant variance.

Crown competition factor is a measure of stand density (Krajicek et al. 1961). Tree values of CCF estimate the percentage of an acre that would be covered by the tree's crown if the tree were open grown (Dixon 2002).

Crown competition factor on the inventory point where the tree is established (PCCF) is a calculated covariate on the plot level.
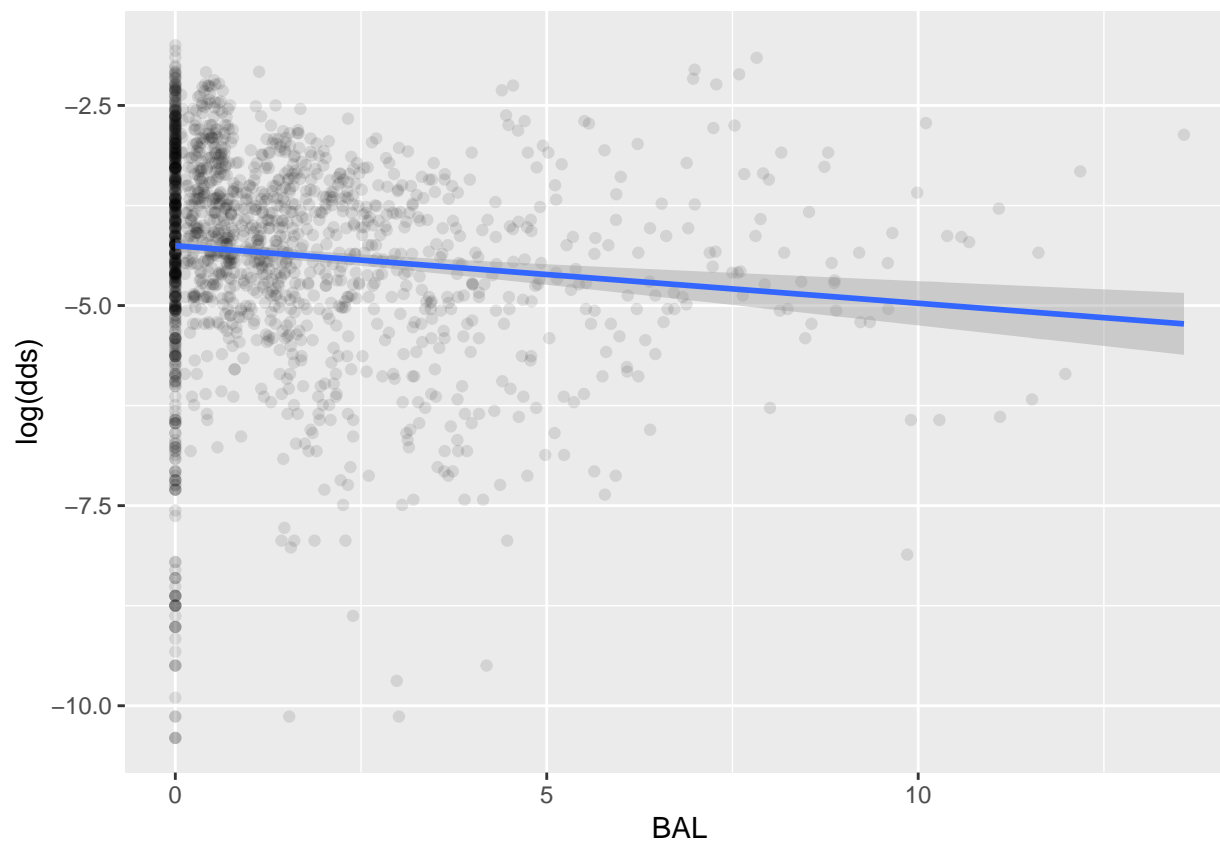
```
ggplot(data = glmm_data_pp, aes(x = PCCF, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

There is a negative relationship between growth and PCCF, and there is relatively constant variance.

Stand crown competition factor (CCF) is a calculated covariate on the plot level. It is a summation of PCCF on the plot.

```
ggplot(data = glmm_data_pp, aes(x = CCF, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

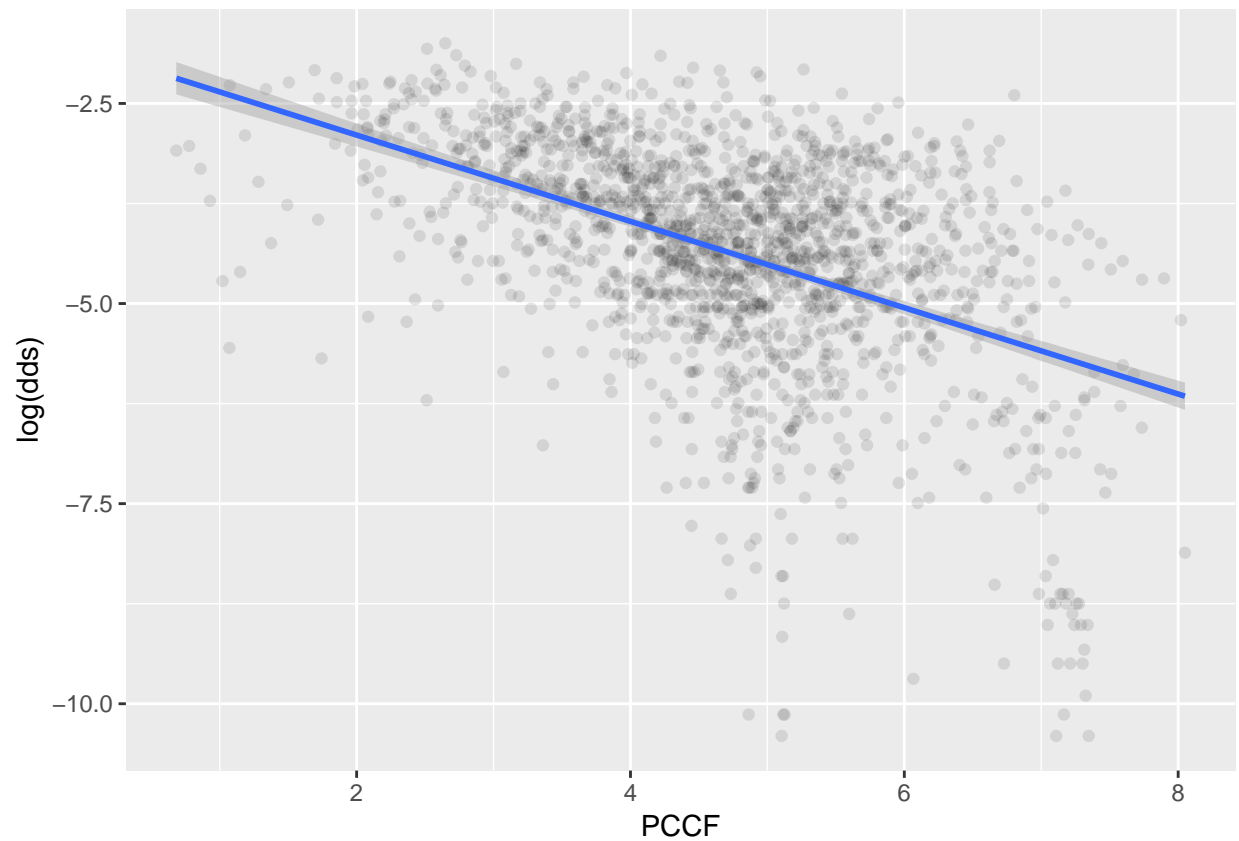There is a negative relationship between growth and CCF. In addition, there is most likely an outlier in the dataset.

Crown ratio (CR/CR_weib) is the percentage of the length of the tree with healthy foliage (FIA Database Description and User Guide for Phase 2). CR was extracted from the FIA Database for the measure year and copied backwards. CR_weib was back calculated using the Weibull distribution (Dixon 1985) on a plot level.

```
ggplot(data = glmm_data_pp, aes(x = CR, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

```r
ggplot(data = glmm_data_pp, aes(x = CR_weib, y = log(dds))) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```
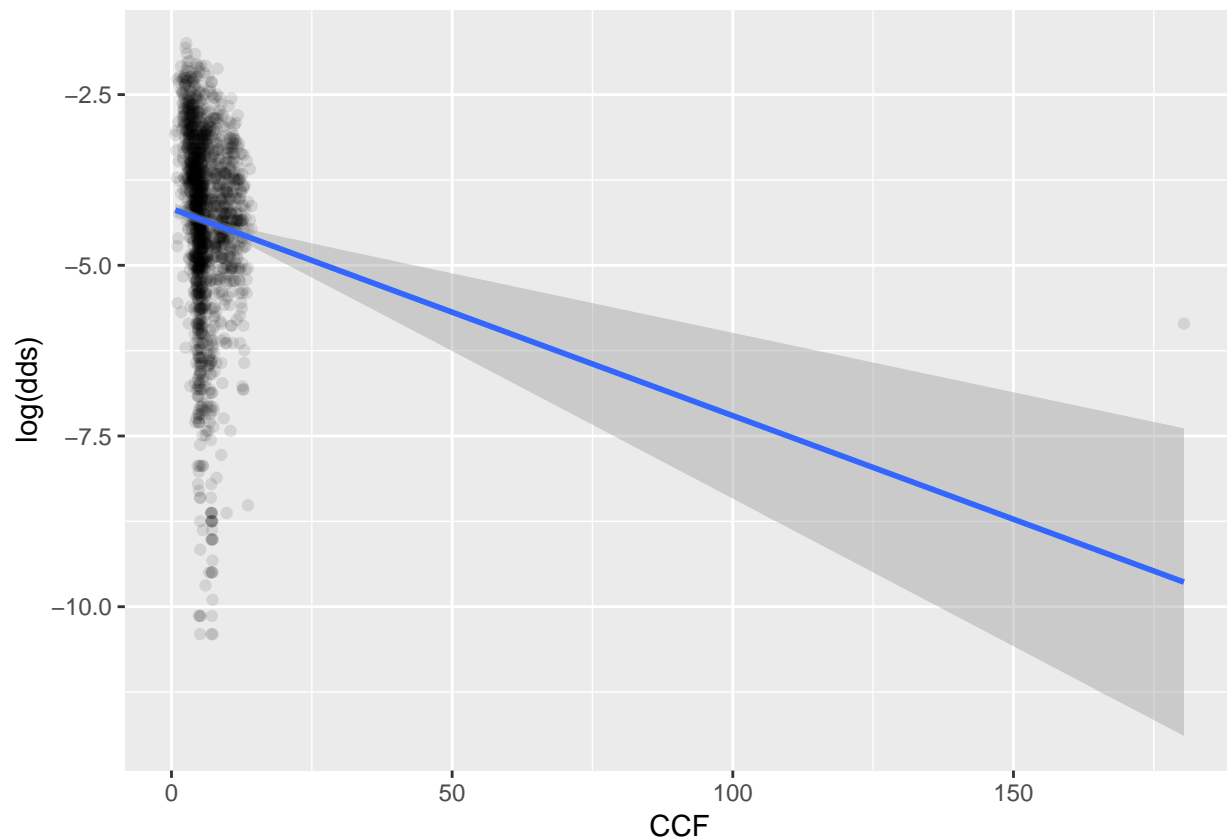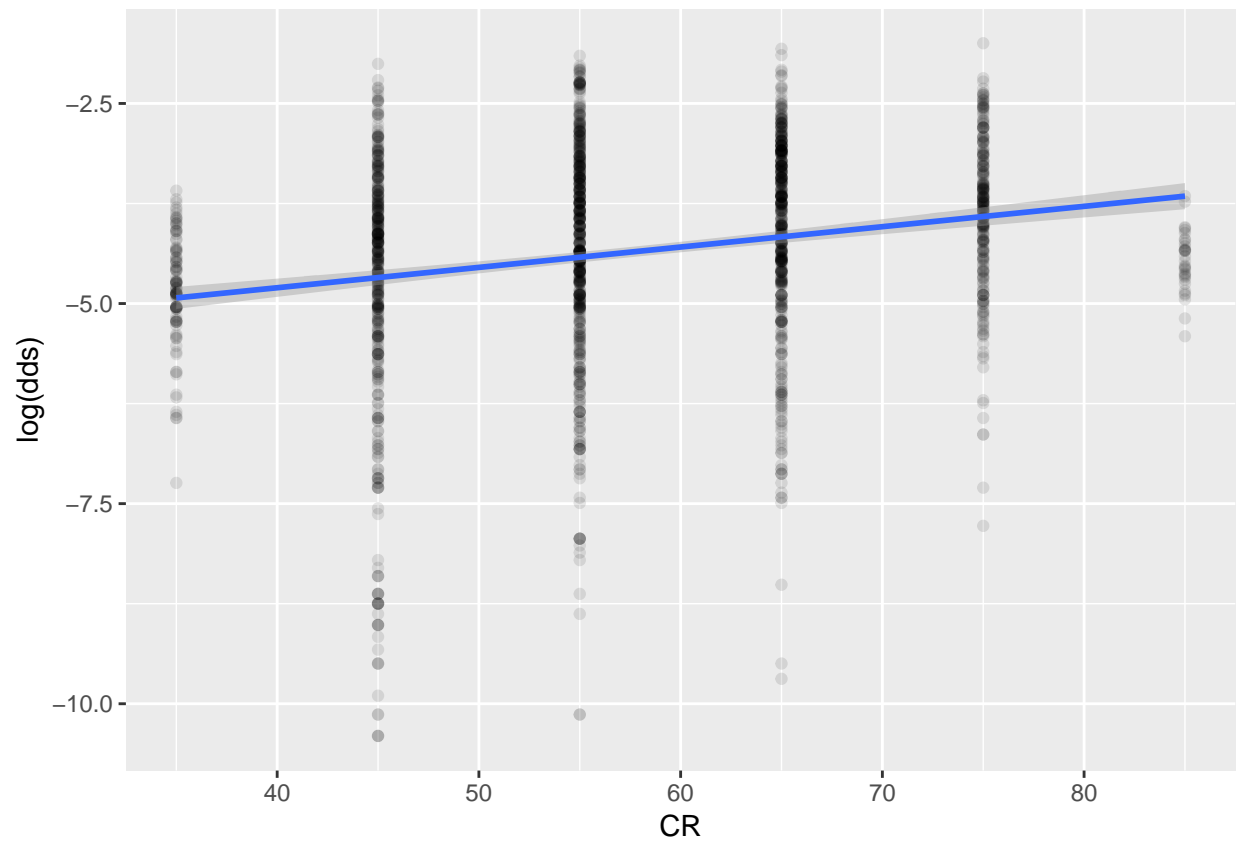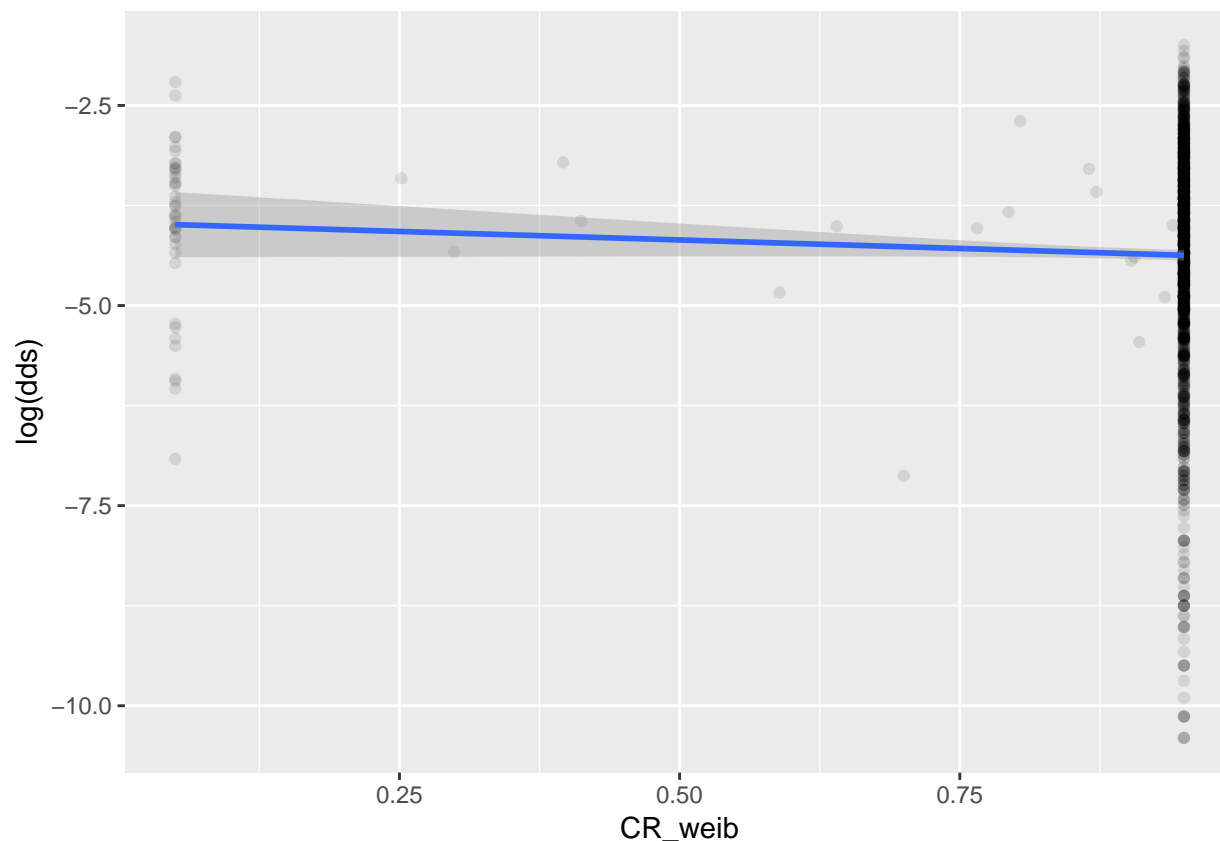
There is a positive relationship between growth and CR and relatively constant variance. However, there is a slight negative relationship between growth and CR_weib and non-constant variance.

Based on the exploration, I will transform the response variable (dds) to satisfy assumptions of normality of the response variable. I will also further explore outliers with Cleveland dotplots. Normality and homogeneity will be further checked using the residuals of the final models.

Model Building

For now, I will use `lme4` to fit models. To fit a covariance-variance structure other than compound symmetry, I can use `nlme`. I may also explore the model fit options with `glmmTMB`.

The current large diameter growth model for Douglas fir in the Utah variant of FVS is a multiple regression.

```
old_fvs <- lm(log(dds)~SICOND+I(sin(ASPECT-0.7854)*SLOPE)+I(cos(ASPECT-0.7854)*SLOPE)+
              SLOPE+I(SLOPE^2)+I(log(DIA_C))+I(BAL/100)+CR+I(CR^2)+I(DIA_C^2)+PCCF,
          data = glmm_data_pp)
summary(old_fvs)$coef
```

```
##                                     Estimate    Std. Error    t value
## (Intercept)                     -8.3098375026 0.8590585748  -9.673191
## SICOND                           0.0227596124 0.0046247176   4.921298
## I(sin(ASPECT - 0.7854) * SLOPE) -0.0047663782 0.0028636426  -1.664446
## I(cos(ASPECT - 0.7854) * SLOPE) -0.0030555854 0.0020027007  -1.525732
## SLOPE                           -0.0229957273 0.0096627865  -2.379824
## I(SLOPE^2)                       0.0003383316 0.0001594854   2.121396
```

```
## I(log(DIA_C))                         1.6564438717 0.1695688680    9.768561
## I(BAL/100)                           10.1004470417 1.9682640978    5.131652
## CR                                     0.1007131667 0.0249869580    4.030629
## I(CR^2)                               -0.0007214952 0.0002028232   -3.557262
## I(DIA_C^2)                            -0.0051417858 0.0004502191  -11.420630
## PCCF                                  -0.7015528374 0.0341211181  -20.560664
##                                           Pr(>|t|)
## (Intercept)                          2.127095e-21
## SICOND                               9.734949e-07
## I(sin(ASPECT - 0.7854) * SLOPE)      9.627199e-02
## I(cos(ASPECT - 0.7854) * SLOPE)      1.273275e-01
## SLOPE                                1.746903e-02
## I(SLOPE^2)                           3.408331e-02
## I(log(DIA_C))                        8.899611e-22
## I(BAL/100)                           3.323236e-07
## CR                                   5.896105e-05
## I(CR^2)                              3.886470e-04
## I(DIA_C^2)                           8.158827e-29
## PCCF                                 3.113125e-81
```

My growth model using tree ring and forest inventory data will be a linear mixed model.

```
full_model <- lmer(log(dds) ~ DIA_C + SICOND + SLOPE + ASPECT +
                   BAL + PCCF + CCF + CR + wateryr + tmax_JunAug +
                   (1|TRE_CN) + (1|Year), data=glmm_data_pp)
summary(full_model)$coef
```

```
##                   Estimate    Std. Error         df    t value     Pr(>|t|)
## (Intercept) -2.944194361 2.3253895861   87.45685 -1.2661080 0.208836557
## DIA_C        0.006152472 0.0284917443  110.62970  0.2159388 0.829433120
## SICOND       0.034937505 0.0240083066   33.04863  1.4552257 0.155040500
## SLOPE       -0.009483526 0.0129036061   32.84779 -0.7349516 0.467584044
## ASPECT      -0.001046331 0.0018807903   32.81113 -0.5563253 0.581763118
## BAL          0.070621906 0.0241512407 1235.48208  2.9241523 0.003516957
## PCCF        -0.021954325 0.0550652803  353.25119 -0.3986963 0.690357857
## CCF         -0.002260774 0.0045712935  610.00375 -0.4945590 0.621089500
## CR           0.025498724 0.0192567624   34.84106  1.3241439 0.194076926
## wateryr      0.000307718 0.0003549295  791.13821  0.8669836 0.386213890
## tmax_JunAug -0.169547047 0.0567013442  260.20478 -2.9901769 0.003055309
```

Collinearity

Using the full model with all the possile covariates, I will test for collinearity between predictors with the `car` package.

```
##      DIA_C     SICOND      SLOPE     ASPECT        BAL       PCCF
##   1.888587   1.283016   1.101705   1.294140   1.170417   1.978524
##        CCF         CR    wateryr tmax_JunAug
##   1.064634   1.494047   1.061257   1.097286
```

The Variance Inflation Factors (VIFs) are relatively low, showing no collinearity.

Model Selection

Compare models via Akaike's Information Criterion. Check for homogeneity and normality using residuals.