

Linear Mixed Model for Douglas Fir

Courtney Giebink

July 31, 2019

Data Summary

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

data_all_df <- data_all_df %>%
  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 *Pseudotsuga menziesii* (Douglas fir) in Utah. I was able to extract and calculate annual data for 59 trees between the years of 1958 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_df$MEASYEAR) #1988 -> 1958
```

```
## [1] 1988
```

```
glmm_data_df <- data_all_df %>%
  select(PLT_CN, TRE_CN, RW, dds, Year, DIA_C,
         SICOND, ASPECT, SLOPE, BAL, CR, CR_weib, PCCF, CCF,
         ppt_pOct, ppt_pDec, ppt_Jun, ppt_Jul,
         ppt_pJunAug, ppt_pAugOct, ppt_MayJul,
         ppt_pJunNov, ppt_FebJul, wateryr,
         tmin_Feb, tmax_Jul,
         tmax_JunAug, tmax_pJulSep, tmin_JanMar,
         tmax_JanJun, tmin_JanJun, tmax_FebJul) %>%
  filter(Year >= 1958)

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

#average temp
#1 month: Feb tmin, Jul tmax
#3 month: tmax_Jun-Aug, tmax_pJul-pSep, tmin_Jan-Mar
#6 month + : Jun, Jul
```

My data, `glmm_data_df`, consists of raw radial increment, or ring width value, (RW) in mm and annual change in squared inside bark diameter (dds) in inches² 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`)

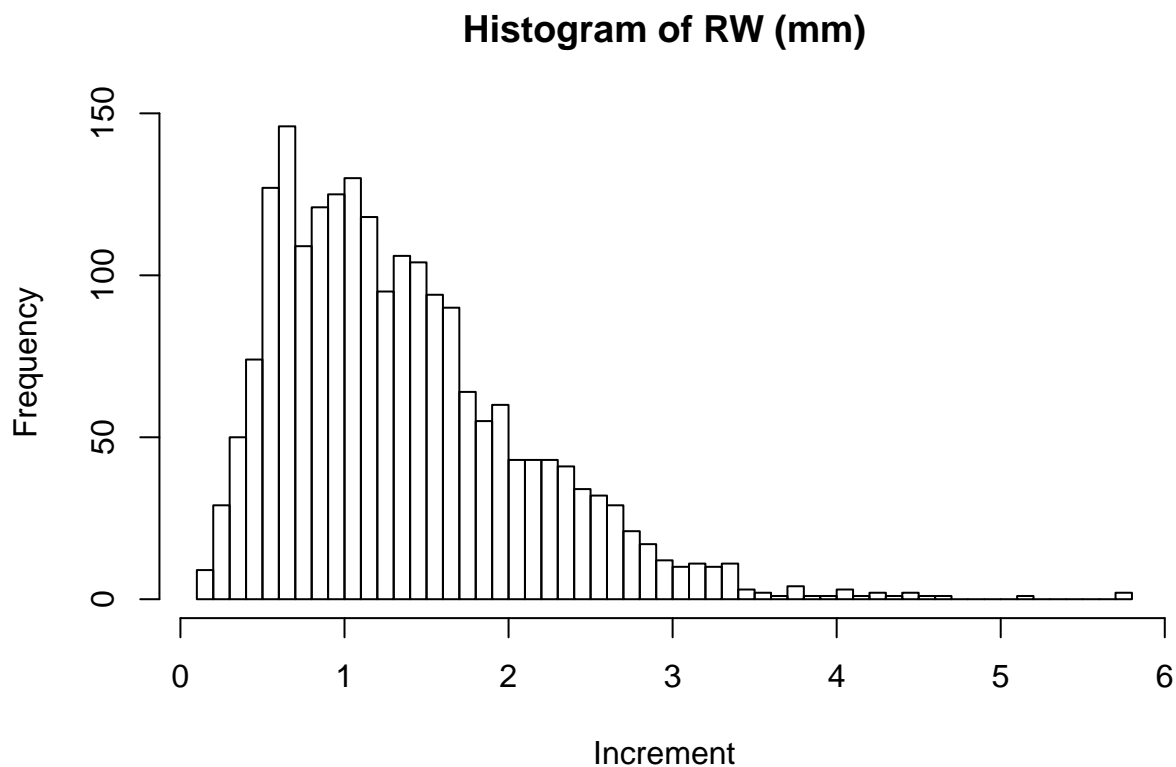
Exploration

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

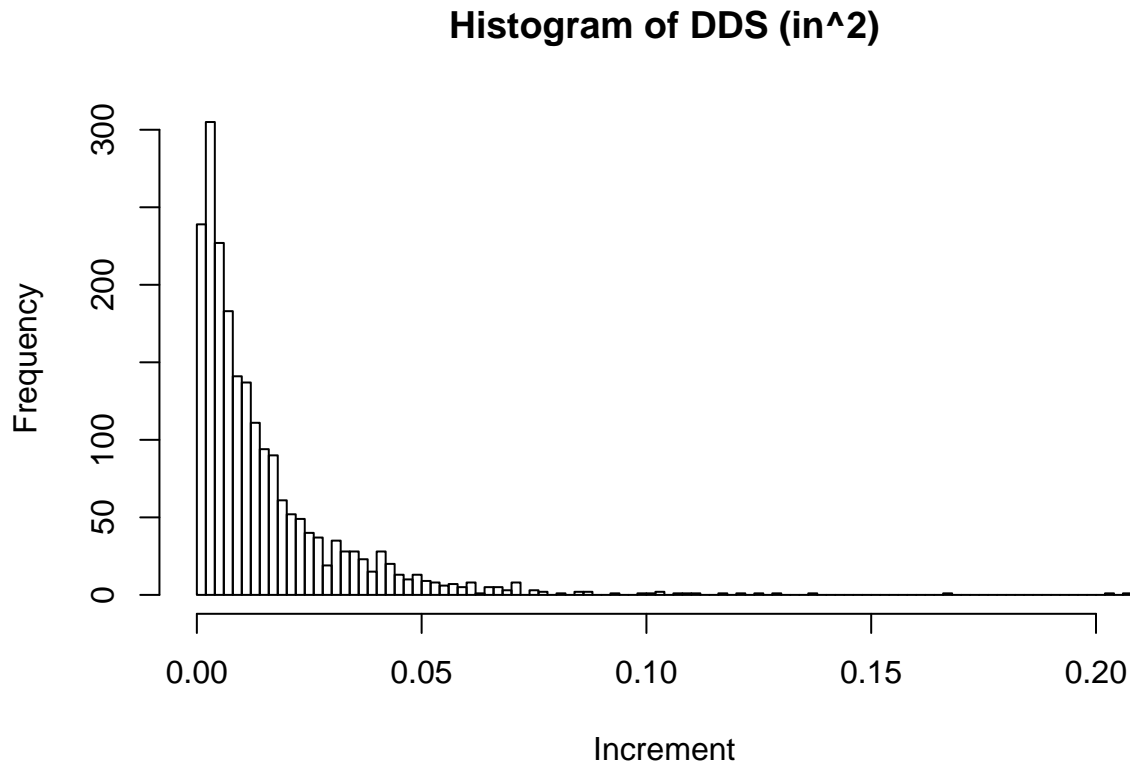
Normality

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



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



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

```
which(glm_data_df$RW == 0) #0
```

```
## integer(0)
```

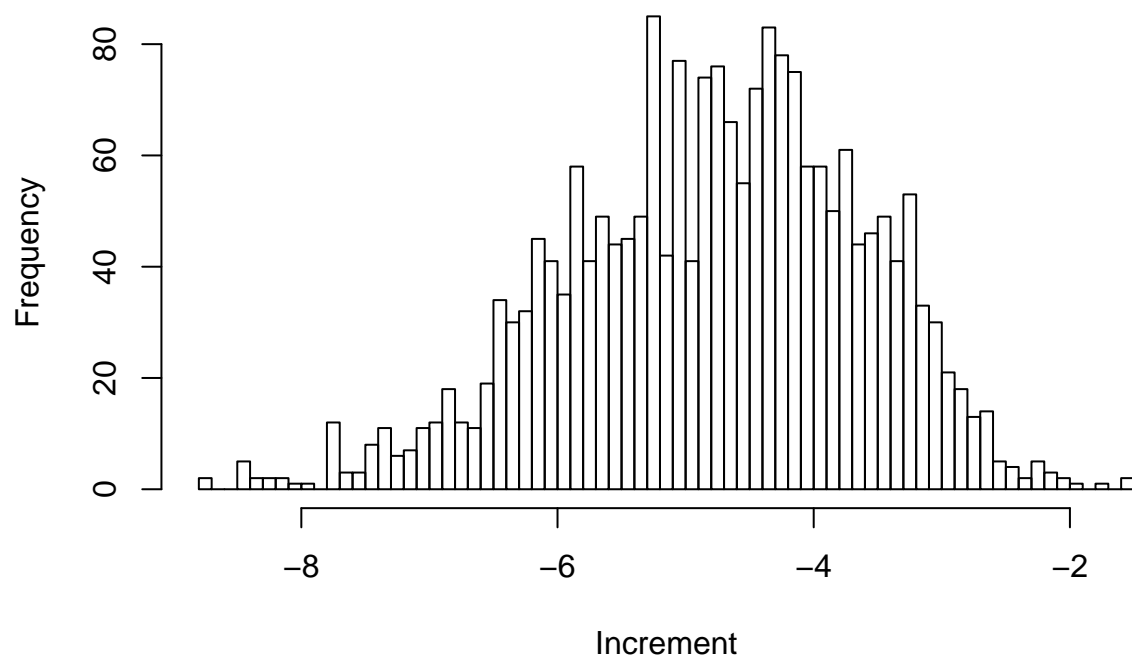
```
#no missing rings
```

```
which(is.na(glm_data_df$RW)) #0
```

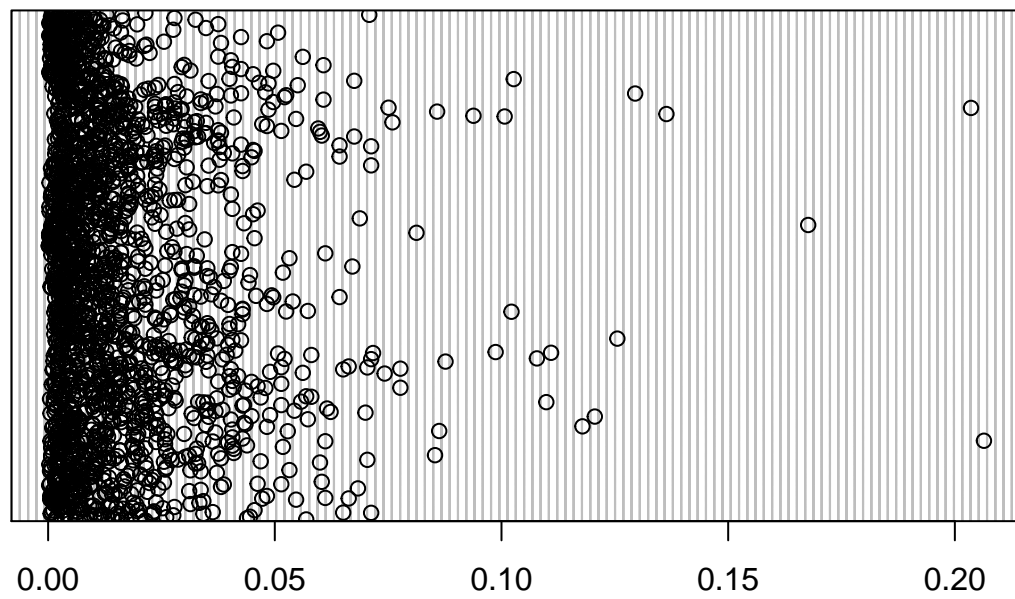
```
## integer(0)
```

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

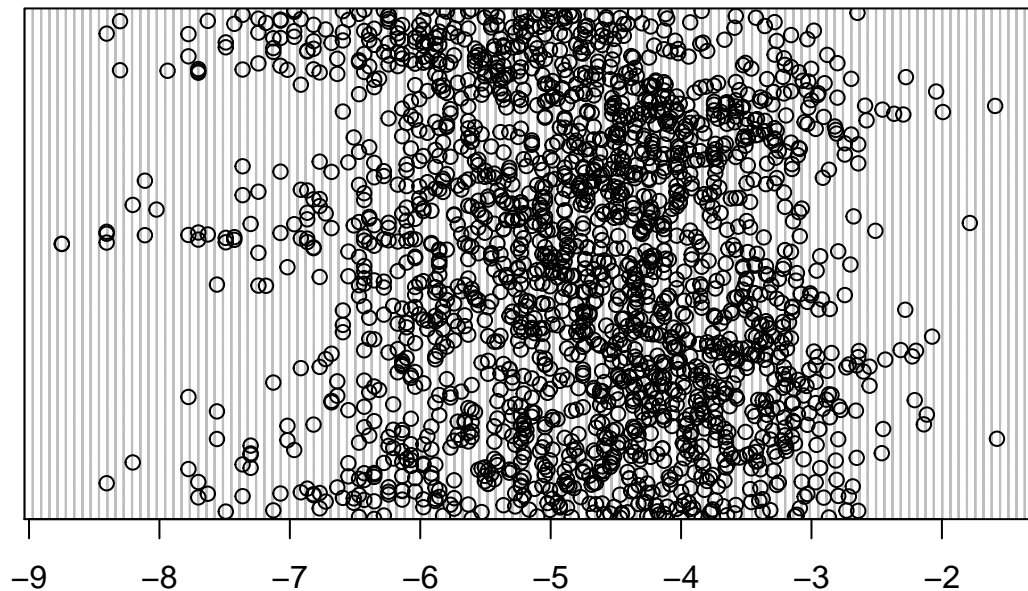
Histogram of log(DDS)



```
#Cleveland dotplot  
dotchart(glm_data_df$dds)
```



```
dotchart(log(glm_data_df$dds))
```



A log transformation normalizes the distribution of DDS and gets rid of outliers.

Zeros and Missing Data

My data is unbalanced. Almost all trees have a growth measurement for 1958, but one tree's growth measurements start several years later. In addition, most trees end growth in 1992, but some end in 1993, 1994, 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_df$TRE_CN))
```

```
## [1] 59
```

```
length(unique(data_all_df$PLT_CN))
```

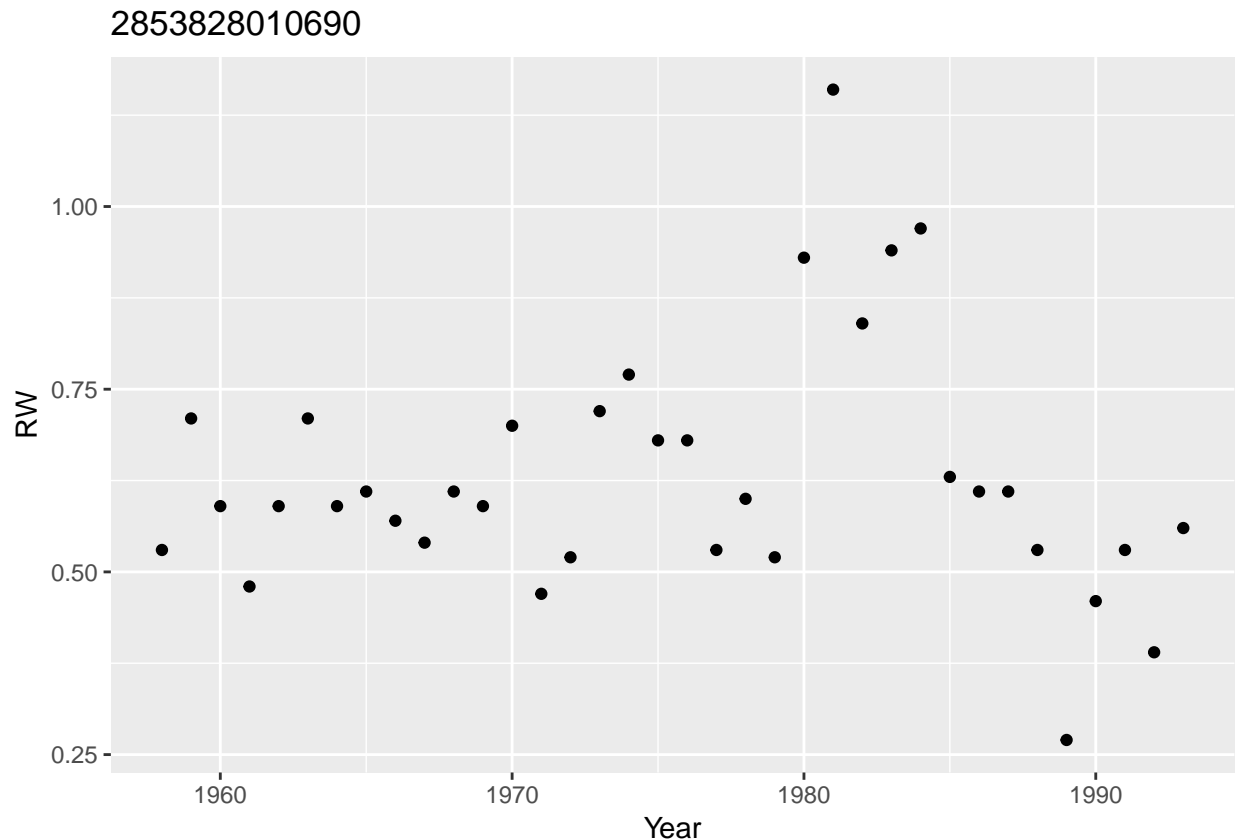
```
## [1] 56
```

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.

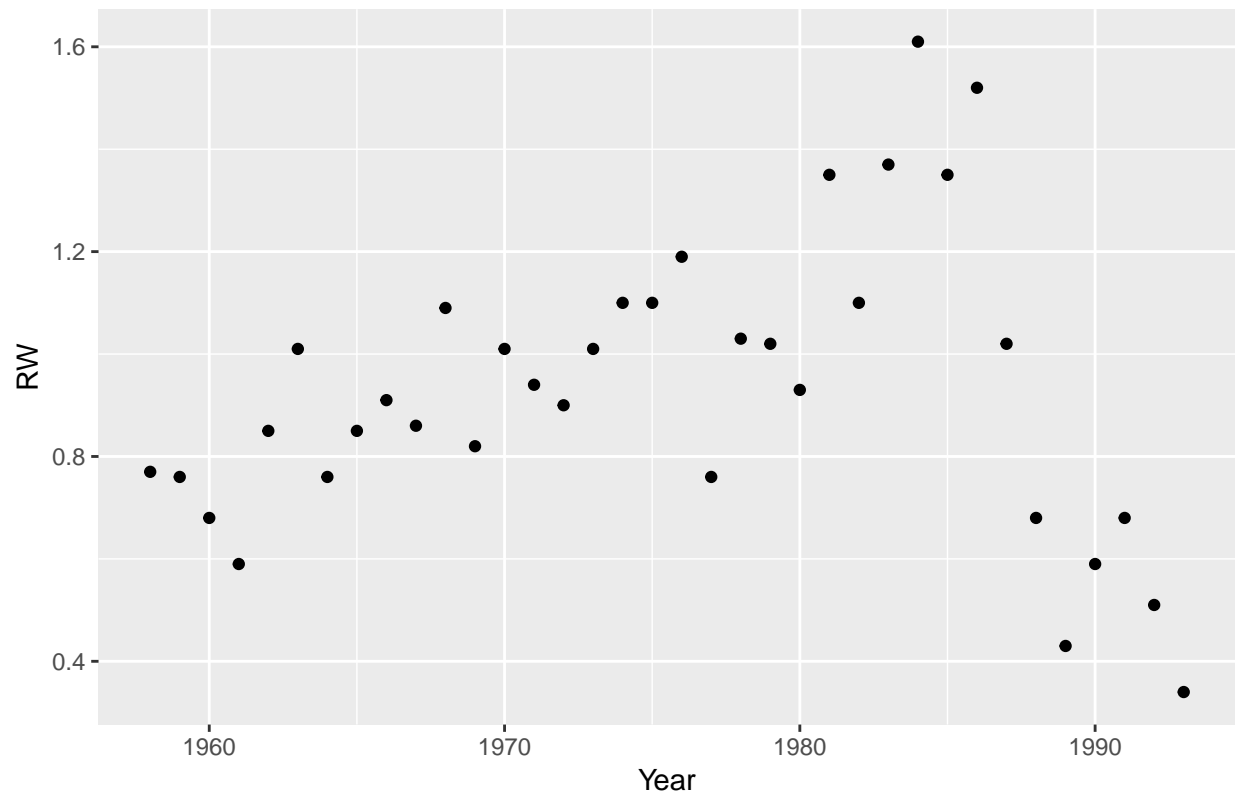
```
#understanding random effects: TRE_CN
growth_yr_plots_df <- glmm_data_df %>%
  group_by(TRE_CN) %>%
  do(plots=ggplot(data=.) +
    aes(x=Year, y=RW) + geom_point() +
    ggtitle(unique(.$TRE_CN)))

growth_yr_plots_df$plots[[1]]
```



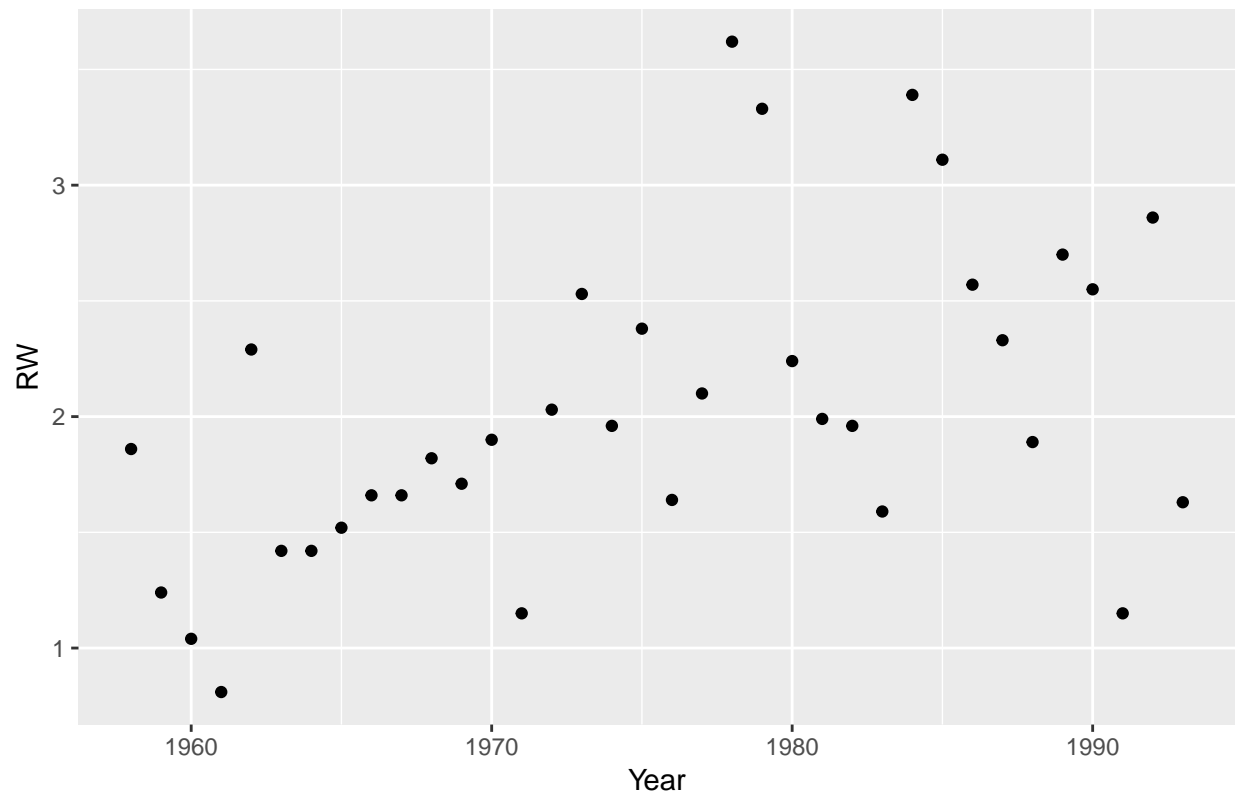
```
growth_yr_plots_df$plots[[2]]
```

2854177010690



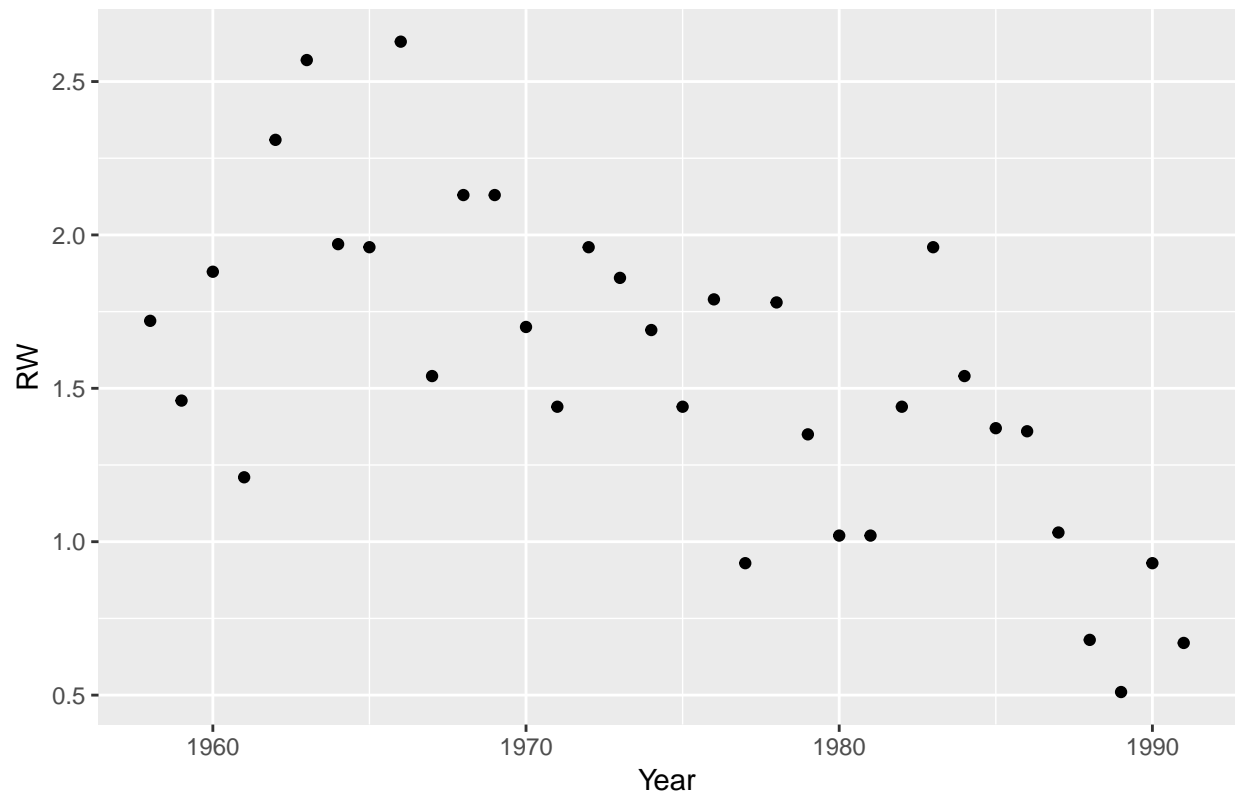
```
growth_yr_plots_df$plots[[3]]
```


2854250010690



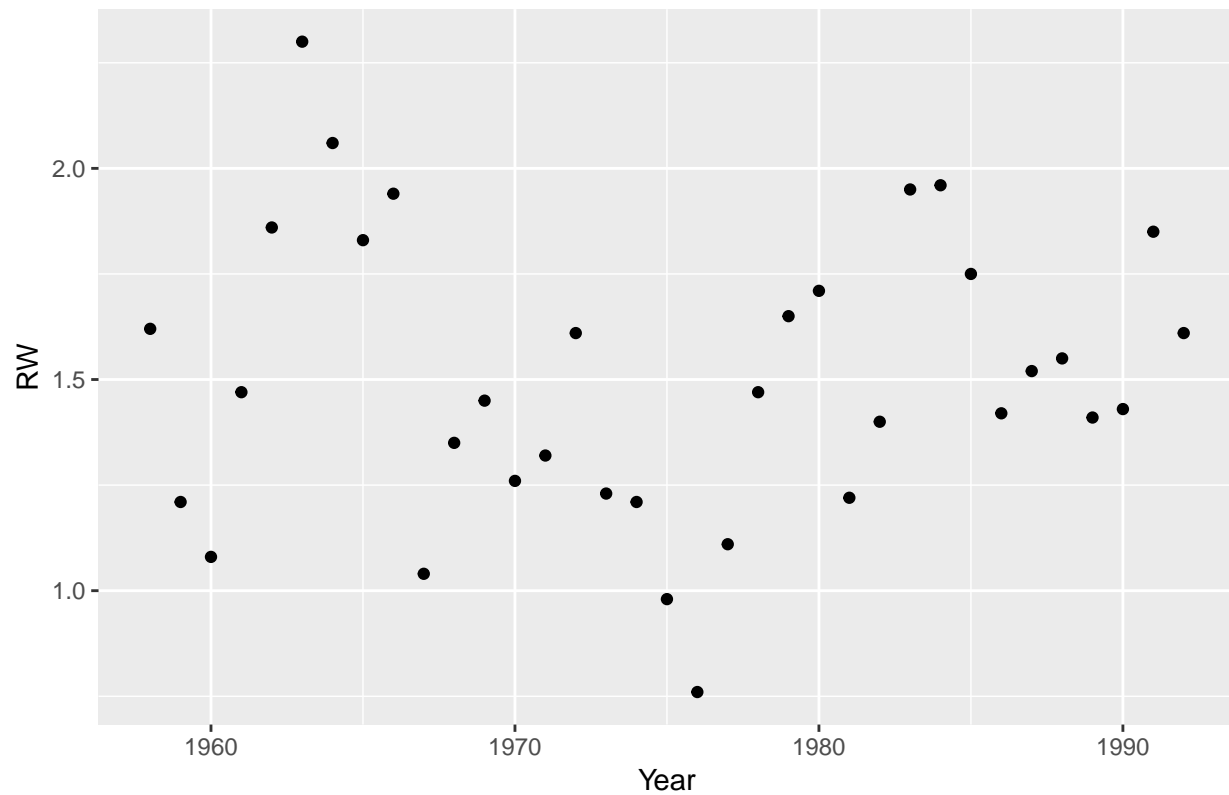
```
growth_yr_plots_df$plots[[4]]
```

2854385010690



```
growth_yr_plots_df$plots[[5]]
```

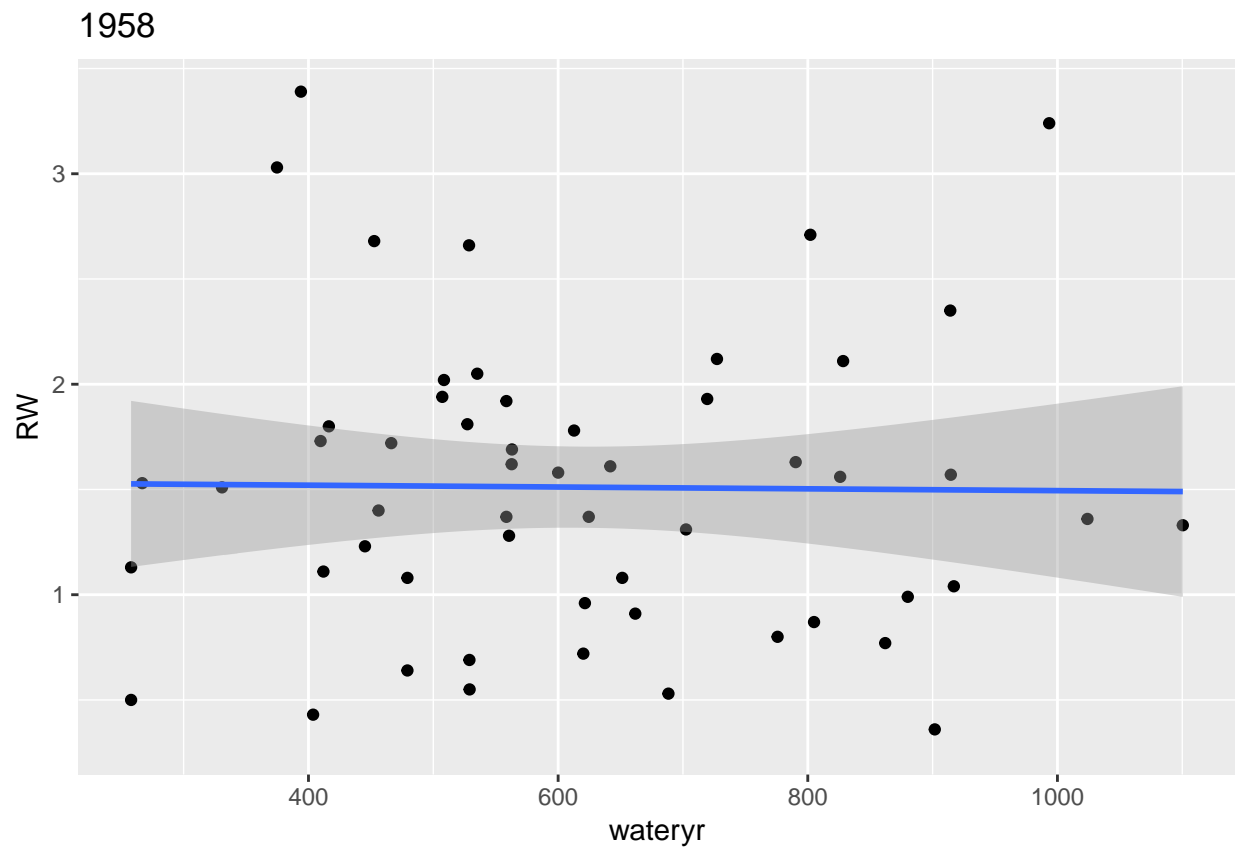
2855103010690



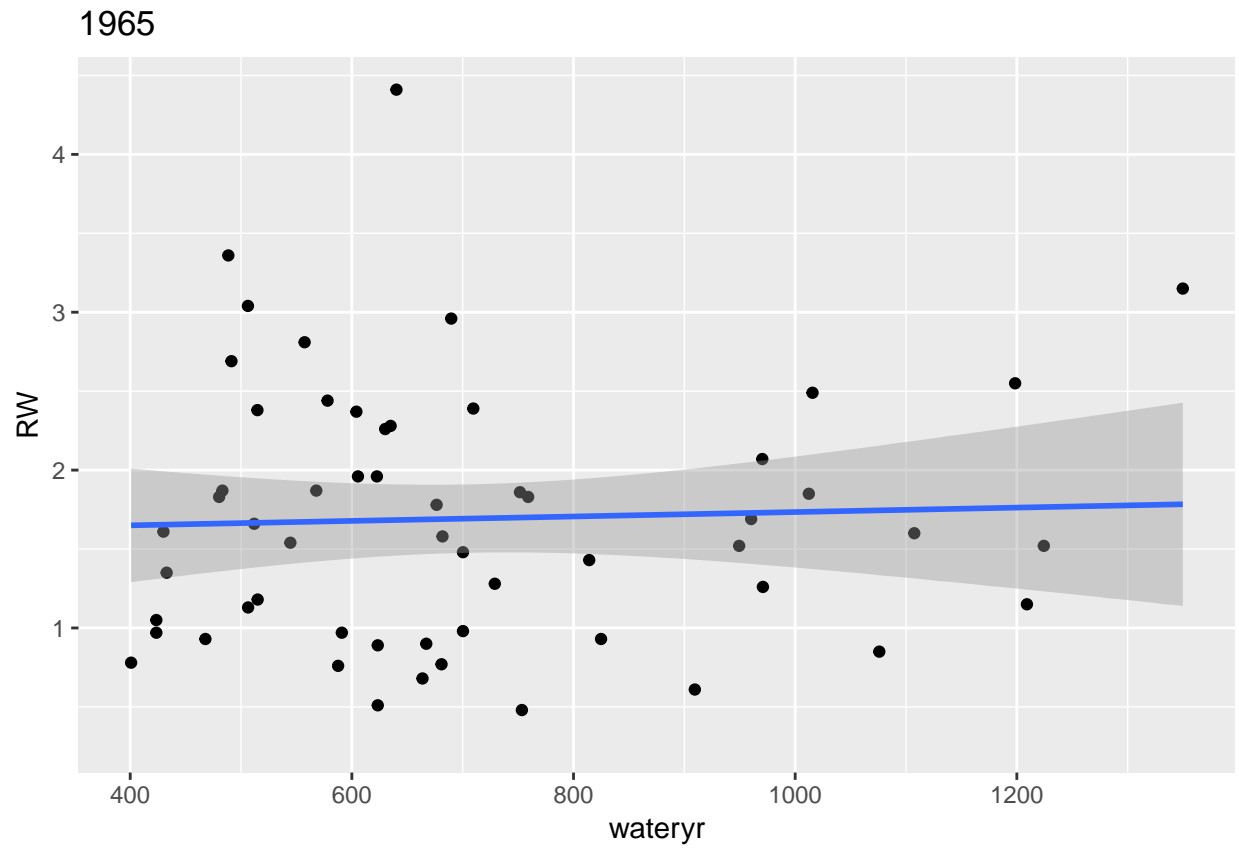
#can do more all the way to tree 59

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

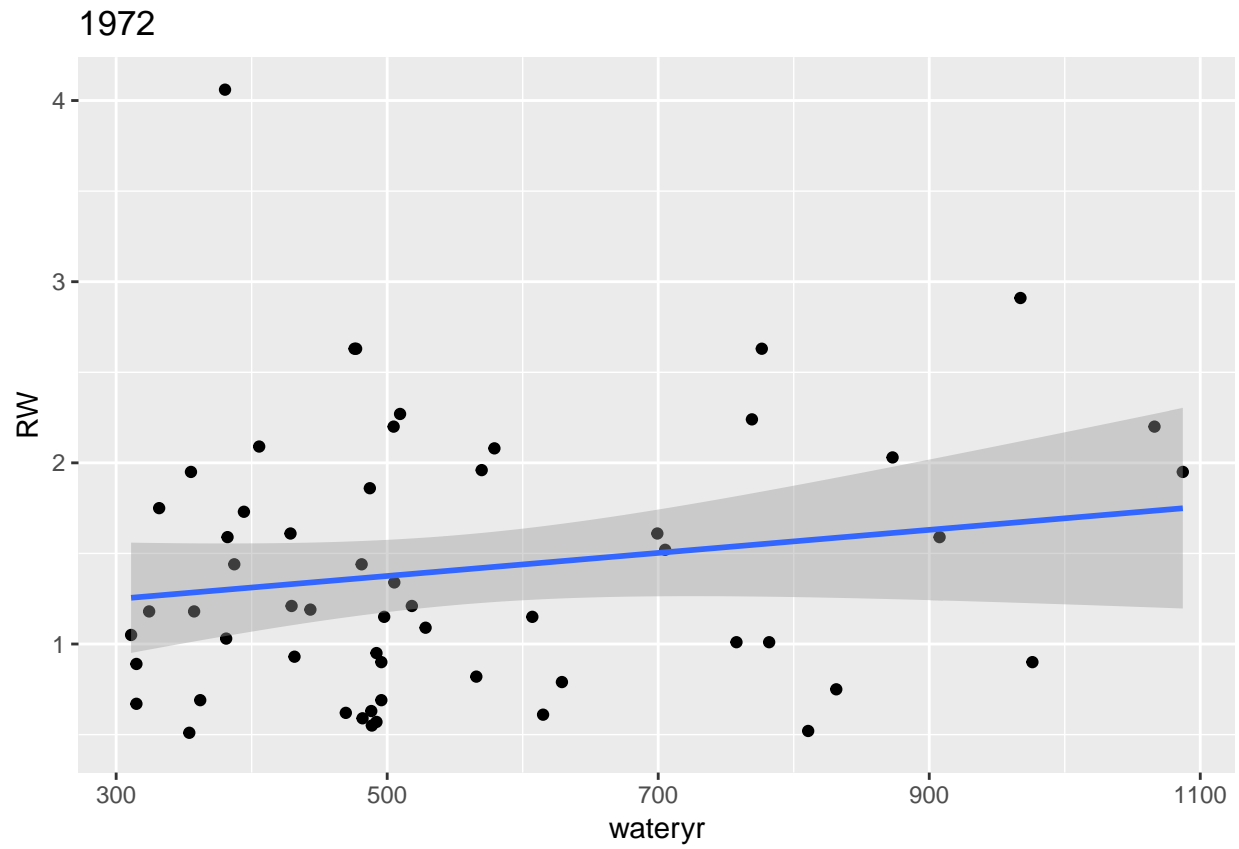
growth_wateryr_plots_df$plots[[1]]
```



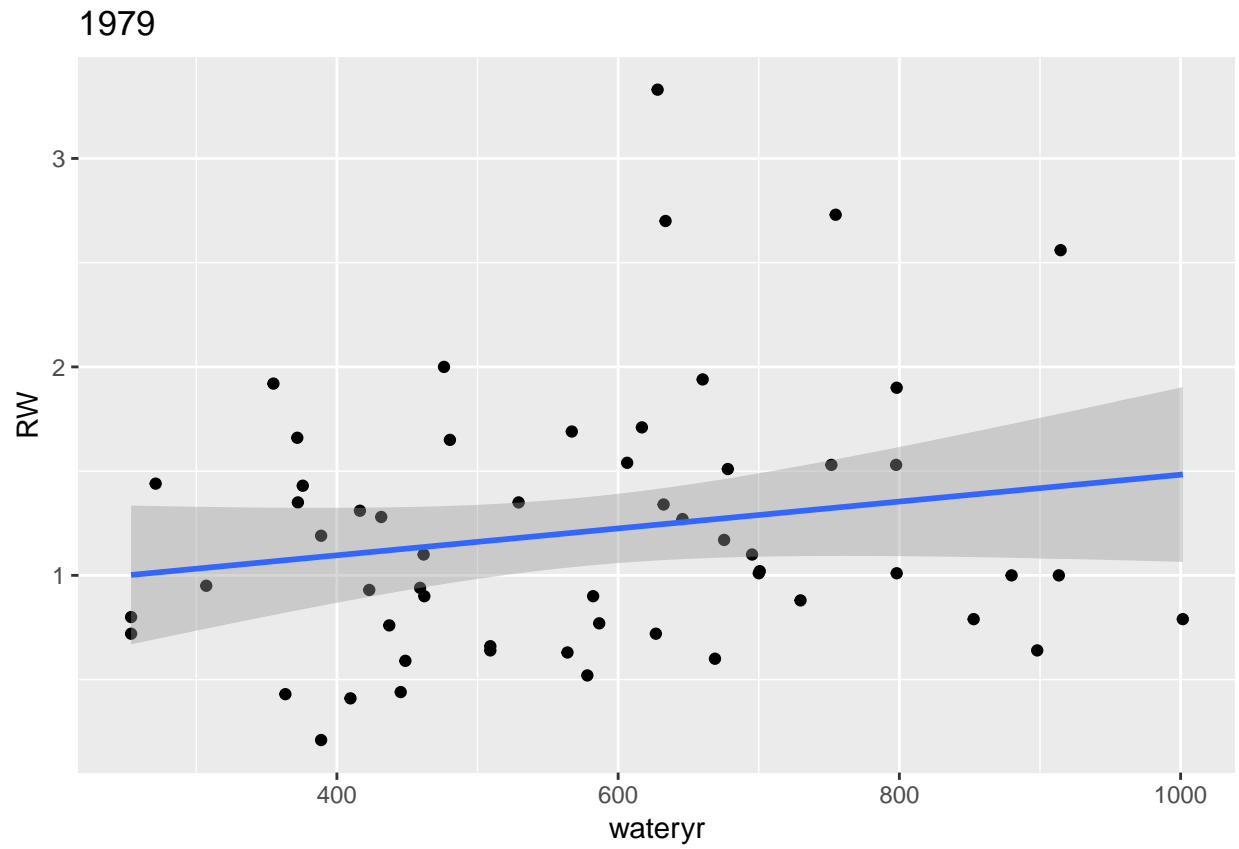
```
growth_wateryr_plots_df$plots[[8]]
```



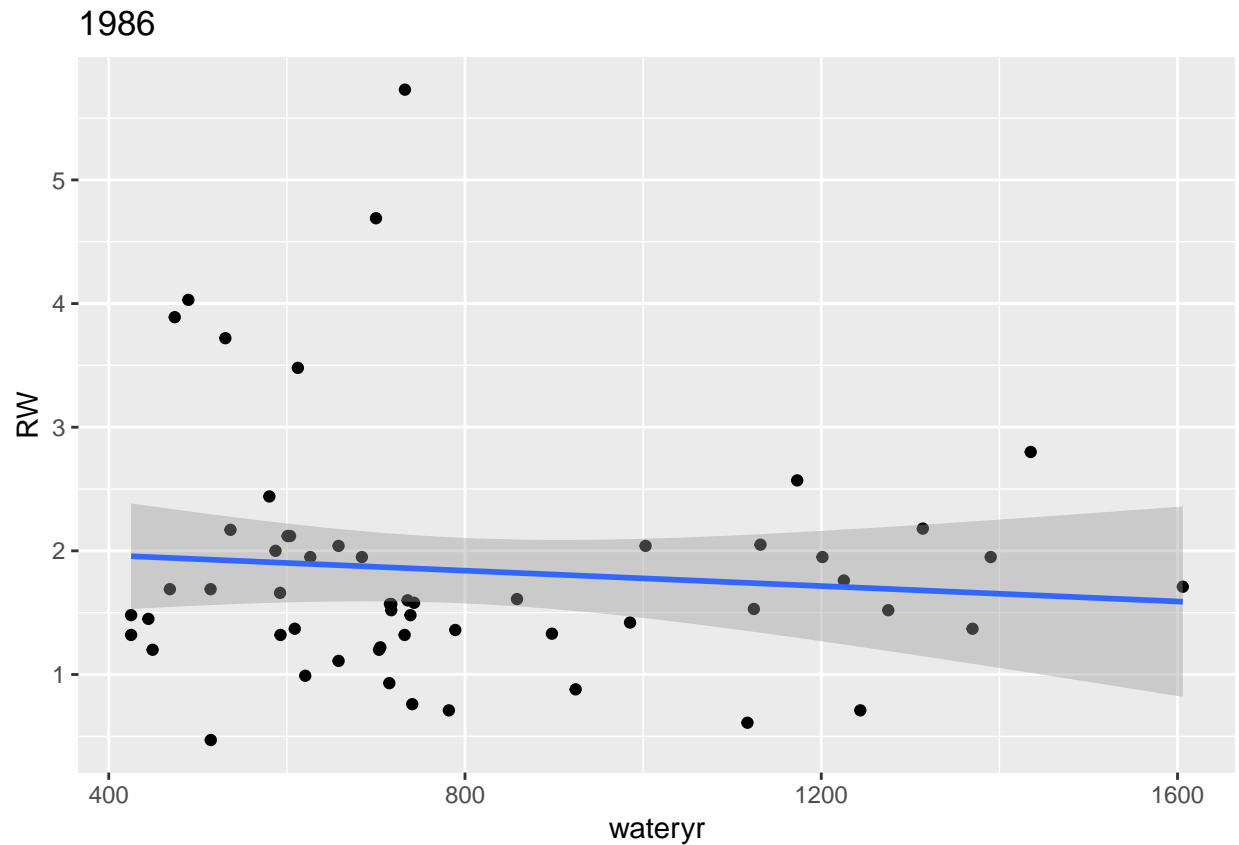
```
growth_wateryr_plots_df$plots[[15]]
```



```
growth_wateryr_plots_df$plots[[22]]
```



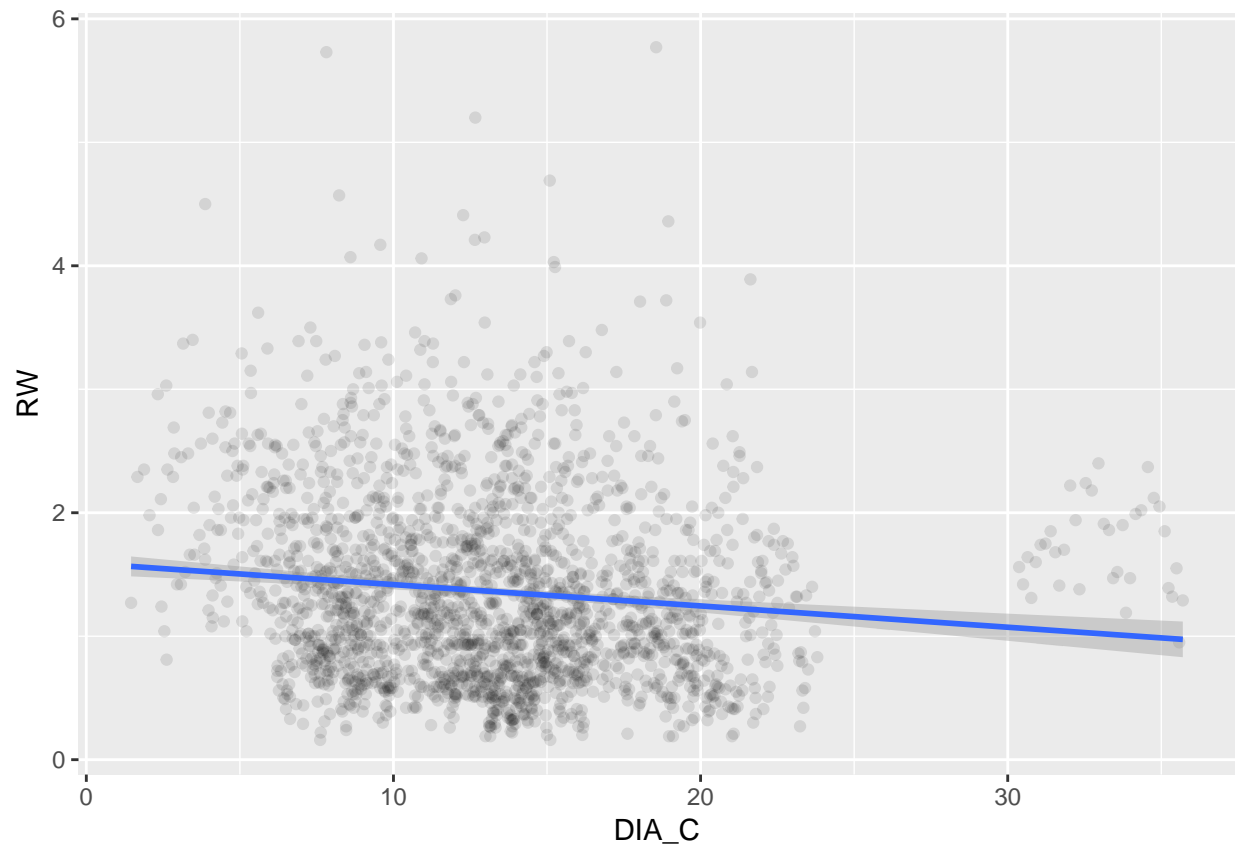
```
growth_wateryr_plots_df$plots[[29]]
```



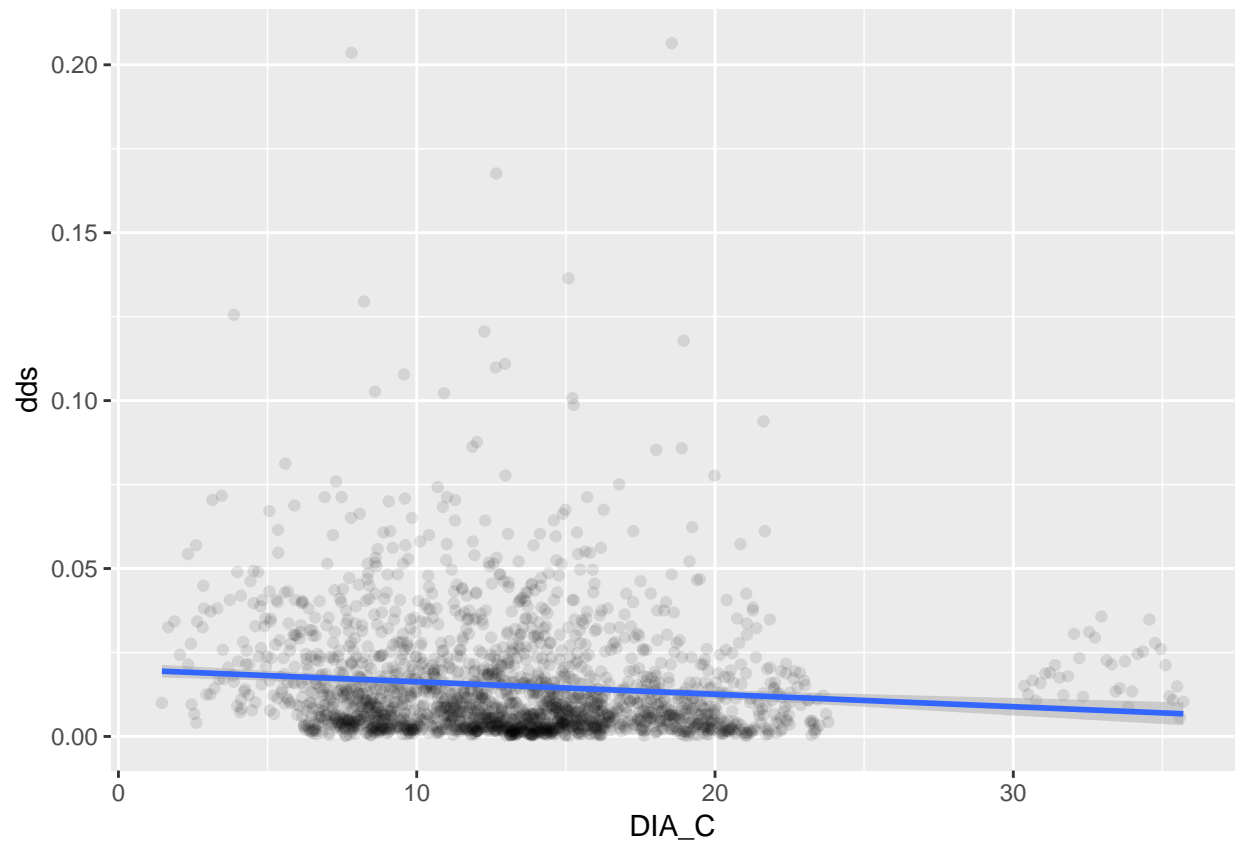
```
#can all years 1958 - 1995
```

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:

```
#detrrend
ggplot(data = glmm_data_df, aes(x = DIA_C, y = RW)) +
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

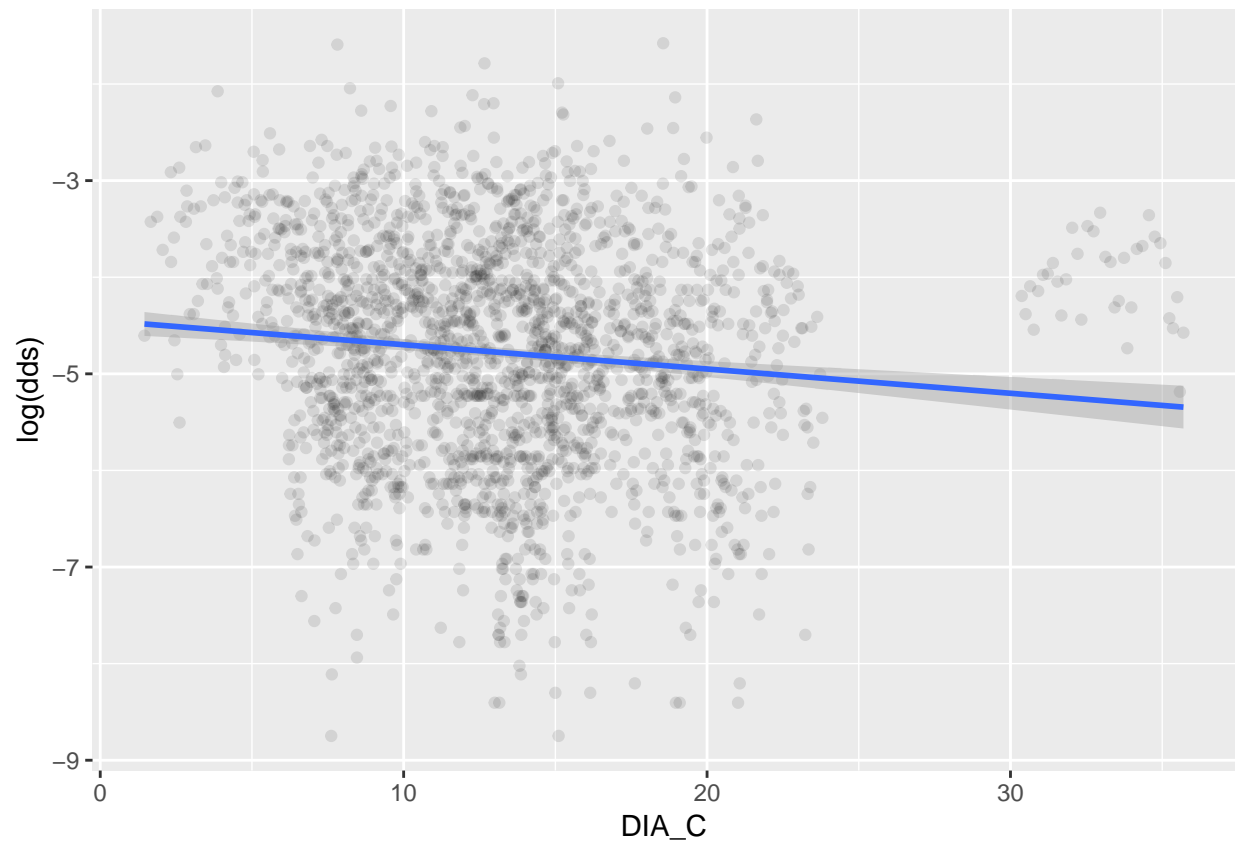



```
ggplot(data = glmm_data_df, aes(x = DIA_C, y = dds)) +  
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

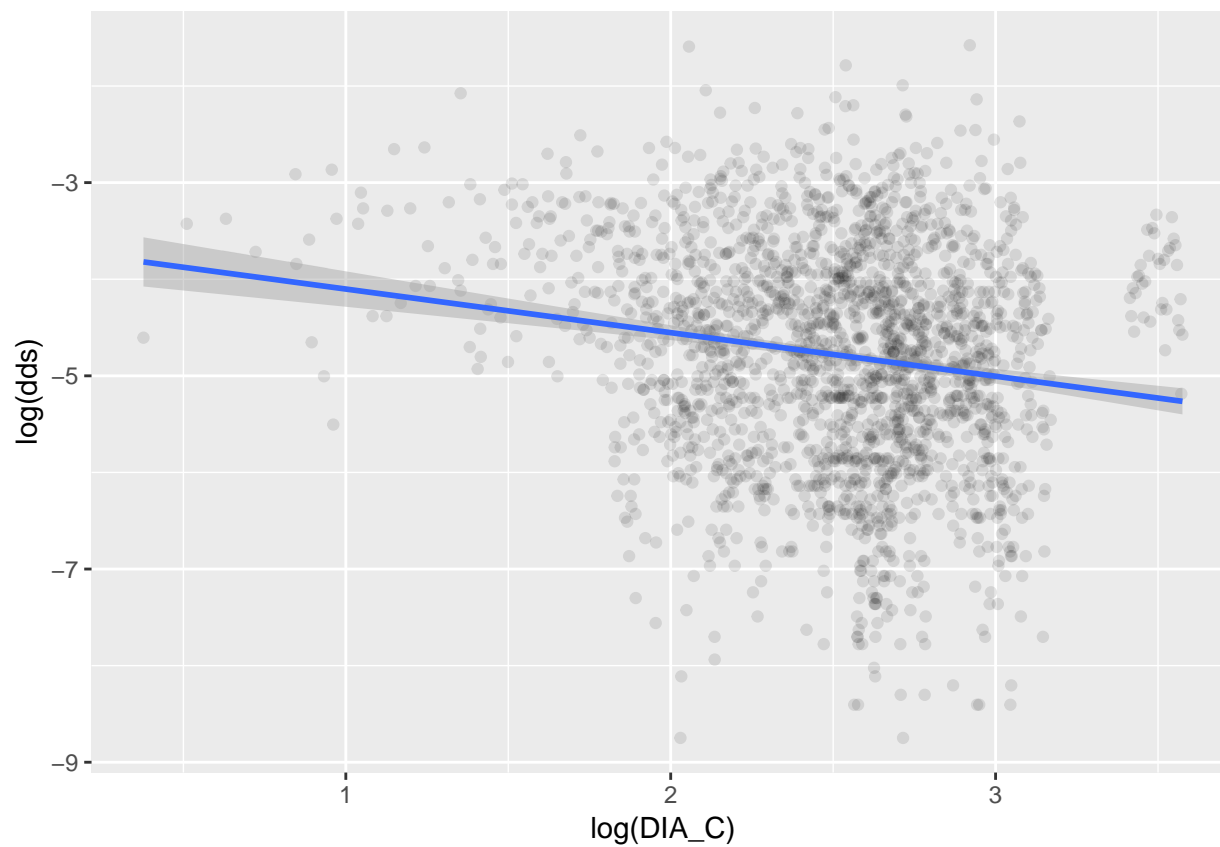


#non constant variance

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



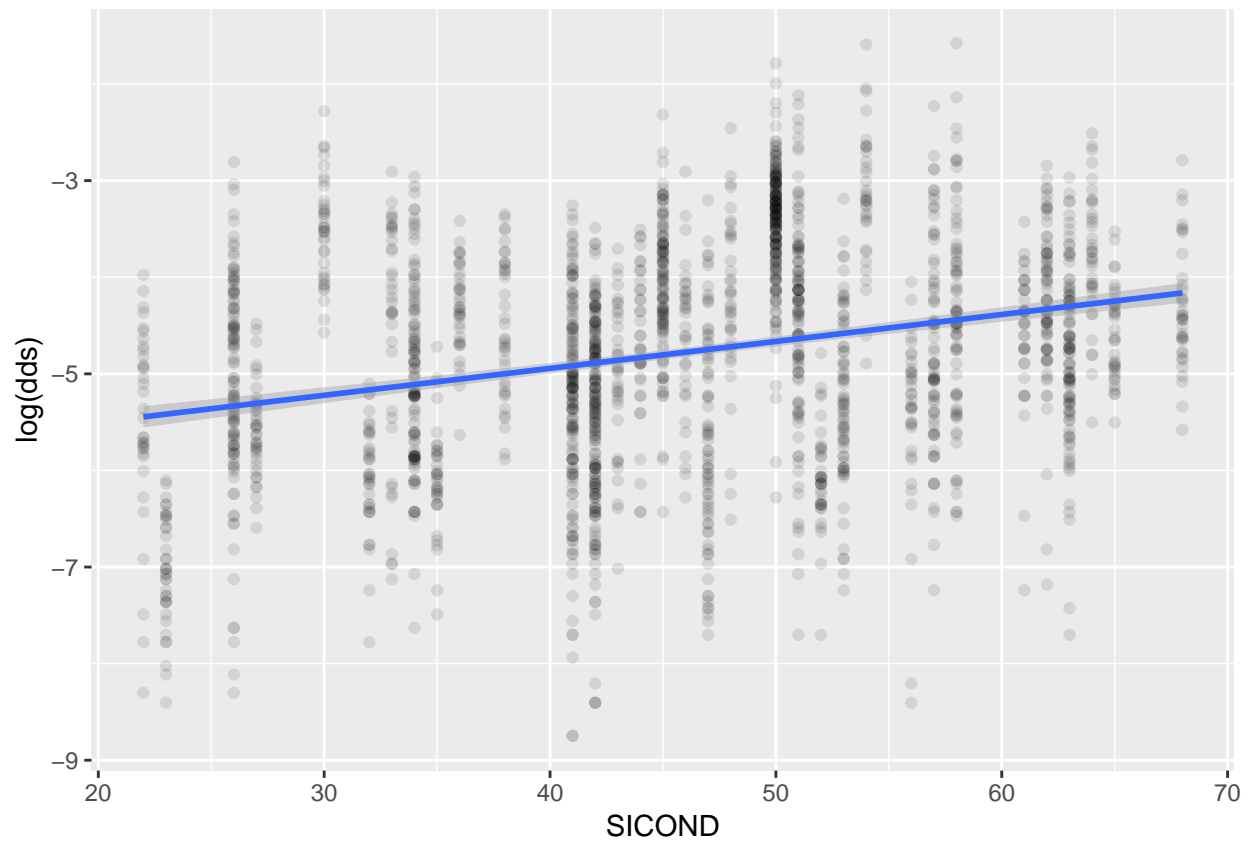
```
#log to reduce heteroscedasticity  
ggplot(data = glmm_data_df, aes(x = log(DIA_C), y = log(dds))) +  
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```



#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).

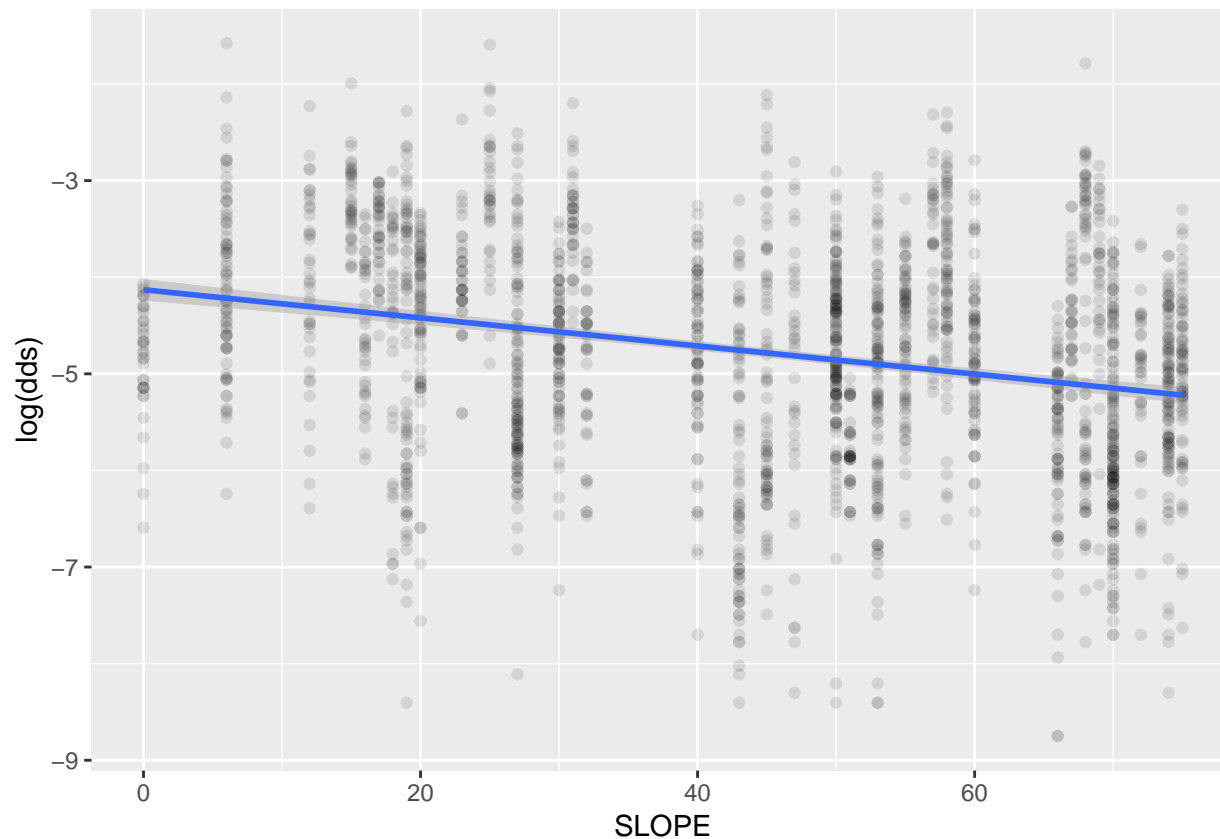
```
ggplot(data = glmm_data_df, aes(x = SICOND, y = log(dds))) +  
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```



There is a positive 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).

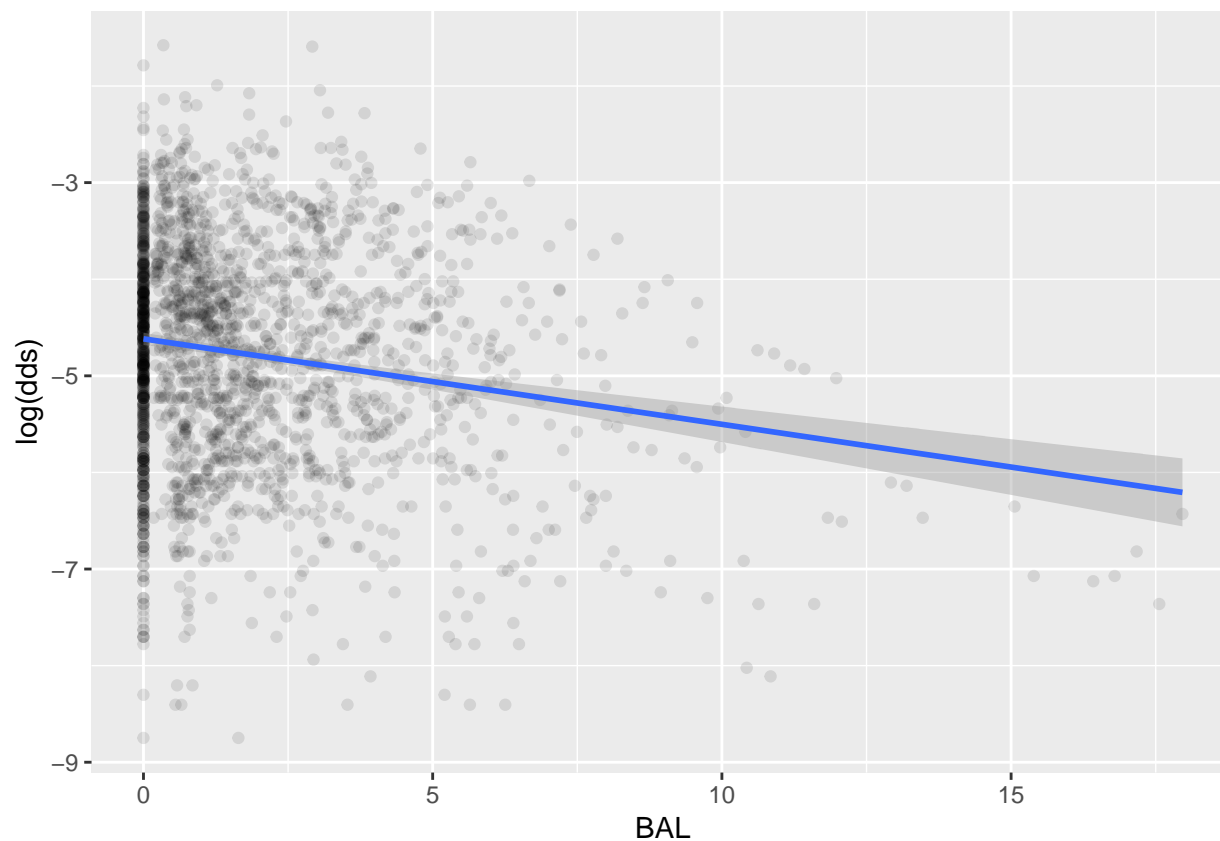
```
ggplot(data = glmm_data_df, aes(x = SLOPE, y = log(dds))) +  
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```



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.

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

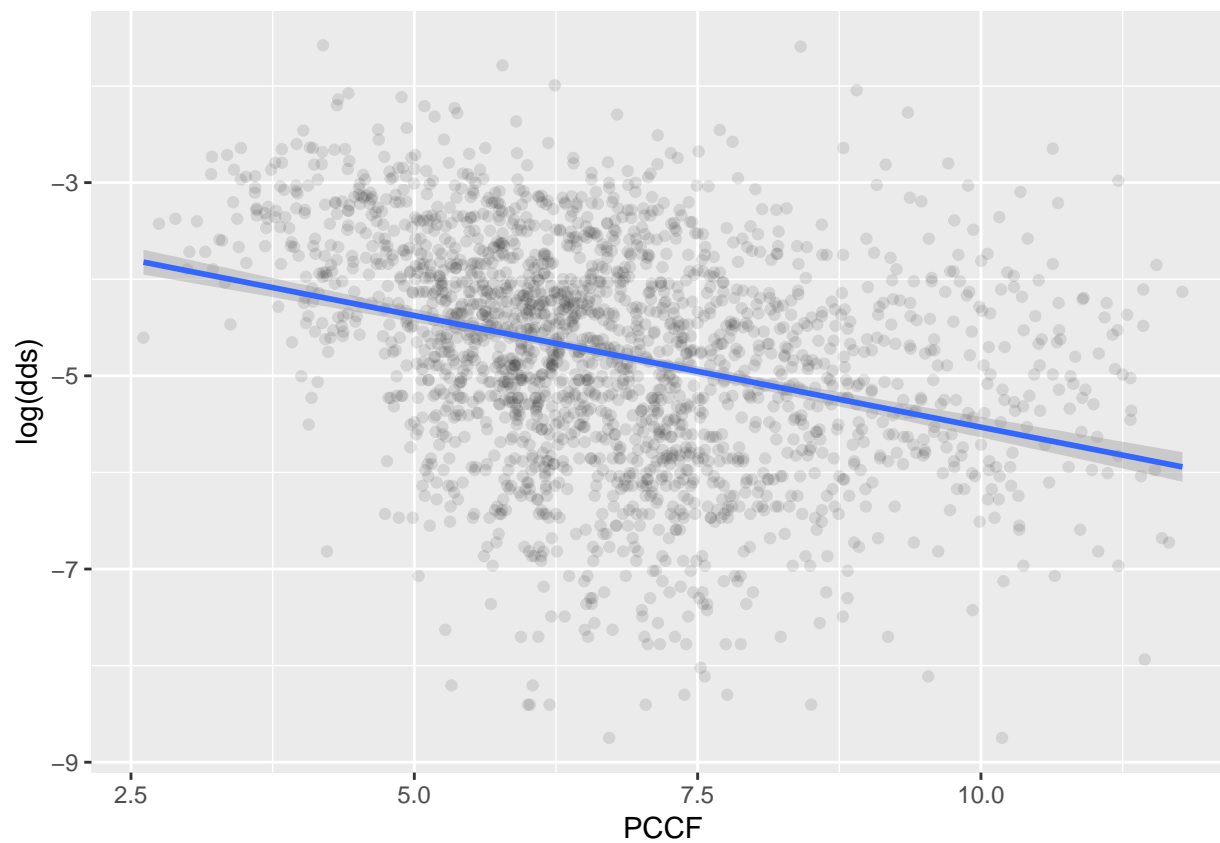


There is a negative relationship between growth and BAL, and there is decreasing variance.

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.

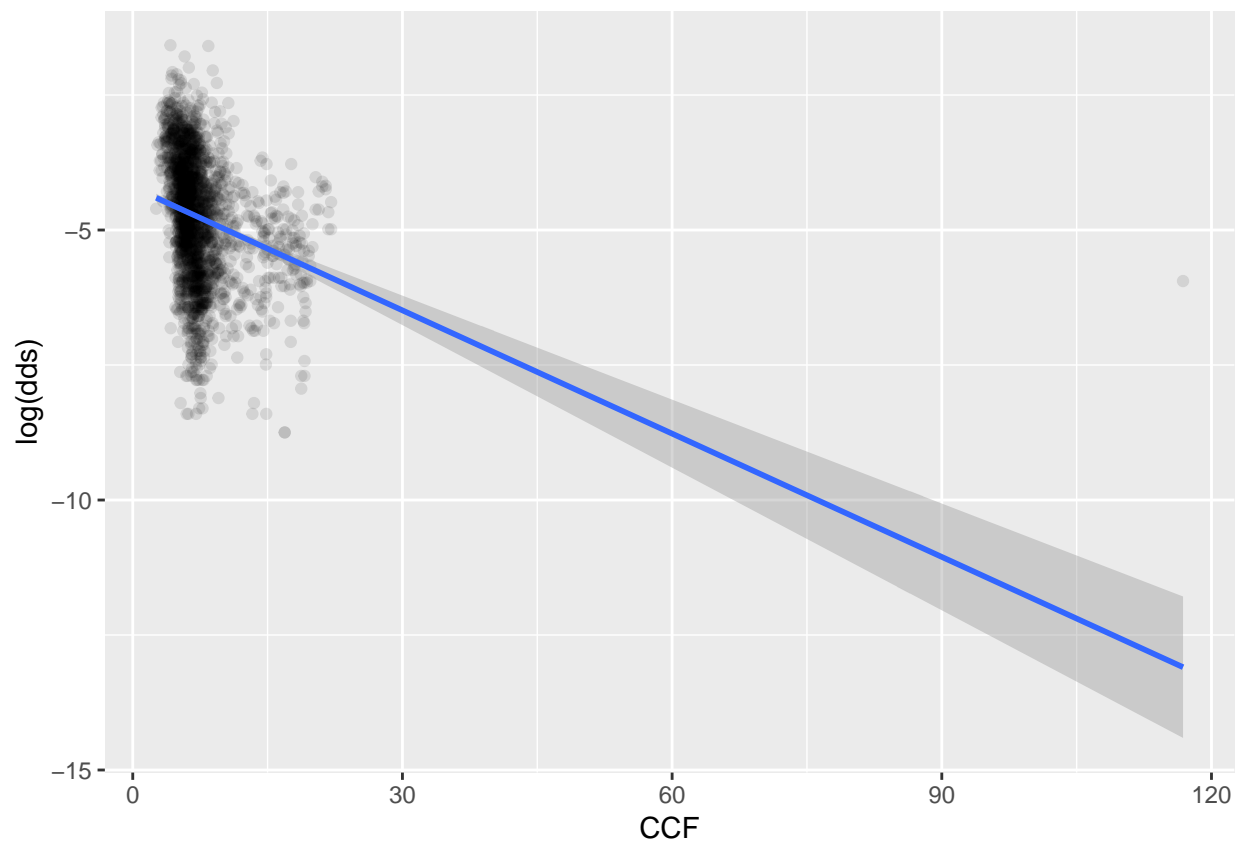
```
ggplot(data = glmm_data_df, aes(x = PCCF, y = log(dds))) +  
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```



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.

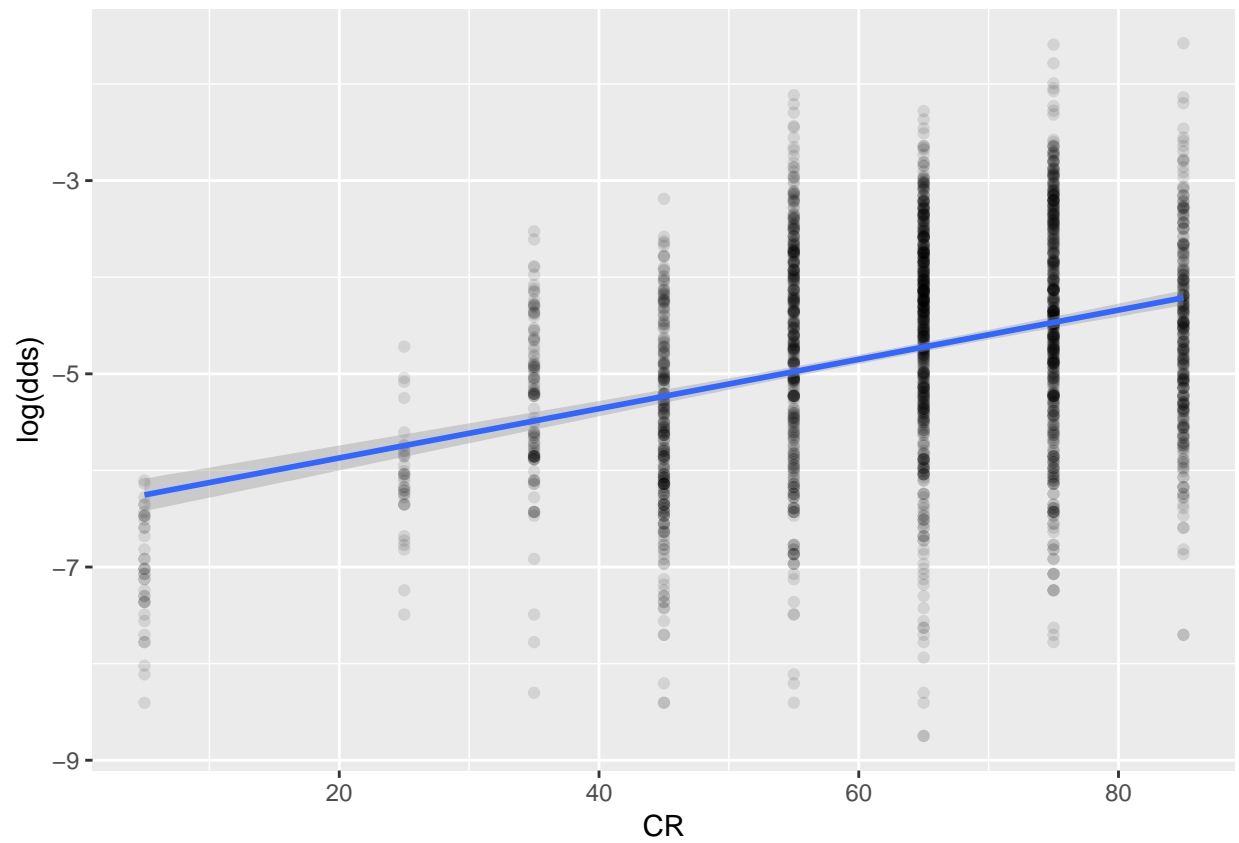
```
ggplot(data = glmm_data_df, aes(x = CCF, y = log(dds))) +  
  geom_point(alpha=0.1) + geom_smooth(method = "lm")
```

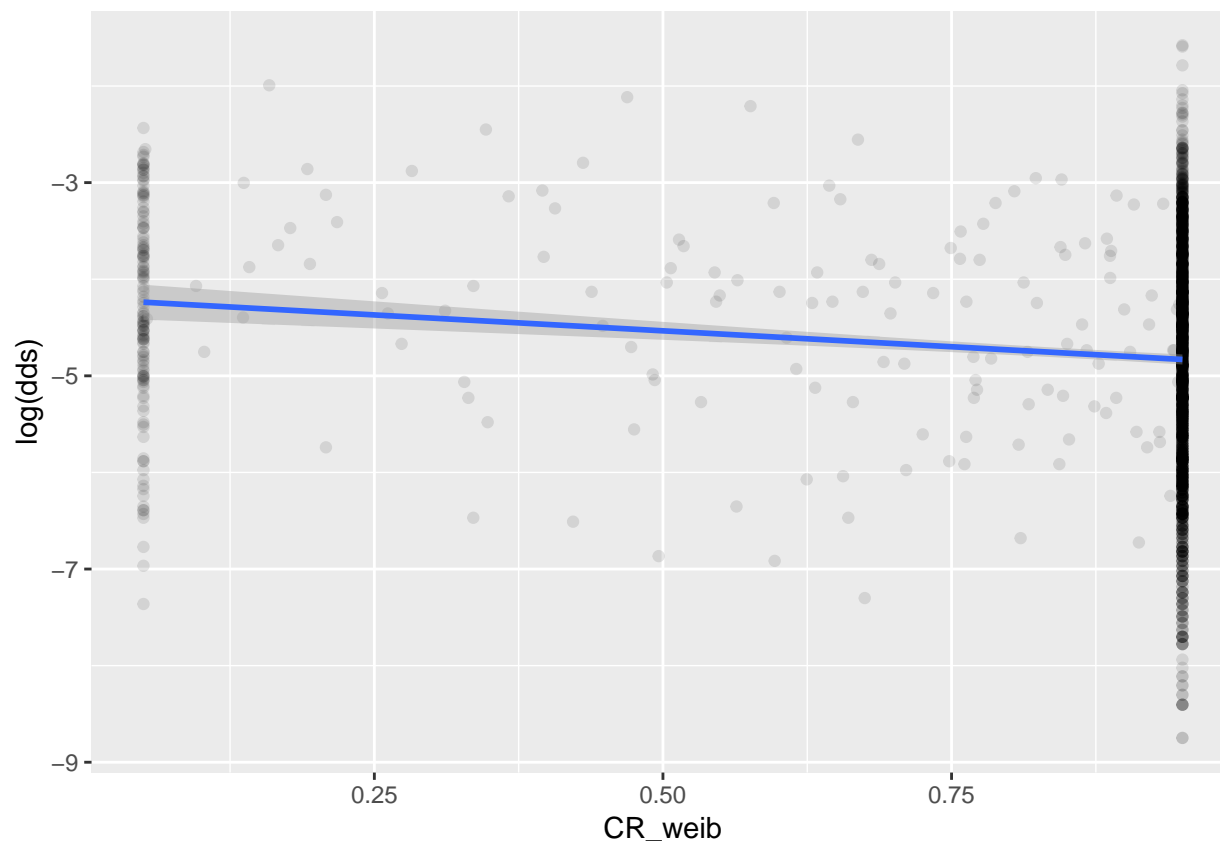
There is a negative relationship between growth and CCF. In addition, there is most likely an outlier in the dataset.

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



```
ggplot(data = glmm_data_df, 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 negative relationship between growth and CR_weib.

Based on the exploration, I will transform the response variable (dds) 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)+PCCF,
              data = glmm_data_df)
summary(old_fvs)$coef
```

	Estimate	Std. Error	t value
## (Intercept)	-4.702211603	3.020644e-01	-15.5669140
## SICOND	0.0113788390	2.076985e-03	5.4785379
## I(sin(ASPECT - 0.7854) * SLOPE)	-0.0001460323	6.519580e-04	-0.2239904
## I(cos(ASPECT - 0.7854) * SLOPE)	-0.0048271135	6.514110e-04	-7.4102427
## SLOPE	0.0025461636	5.405297e-03	0.4710497

```
## I(SLOPE^2) -0.0001619025 6.202065e-05 -2.6104619
## I(log(DIA_C)) -0.5067809300 5.406377e-02 -9.3737621
## I(BAL/100) 0.6507101300 1.256241e+00 0.5179820
## CR 0.0596913110 6.724714e-03 8.8764094
## I(CR^2) -0.0003645740 5.786233e-05 -6.3007151
## PCCF -0.1863145711 1.694702e-02 -10.9939457
## Pr(>|t|)
## (Intercept) 1.237920e-51
## SICOND 4.835768e-08
## I(sin(ASPECT - 0.7854) * SLOPE) 8.227879e-01
## I(cos(ASPECT - 0.7854) * SLOPE) 1.857513e-13
## SLOPE 6.376572e-01
## I(SLOPE^2) 9.110237e-03
## I(log(DIA_C)) 1.844494e-20
## I(BAL/100) 6.045287e-01
## CR 1.511544e-18
## I(CR^2) 3.641568e-10
## PCCF 2.480031e-27
```

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_JanJun +
  (1|TRE_CN) + (1|Year), data=glmm_data_df)
summary(full_model)$coef
```

```
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) -5.0392081023 0.7135247693 84.79158 -7.0624151 4.227119e-10
## DIA_C -0.0275874152 0.0157837169 77.23455 -1.7478402 8.446586e-02
## SICOND 0.0167979800 0.0089500440 47.84300 1.8768600 6.664286e-02
## SLOPE -0.0131235537 0.0048000808 46.21466 -2.7340276 8.839541e-03
## ASPECT -0.0005559801 0.0007195891 45.80089 -0.7726355 4.437071e-01
## BAL -0.0229133843 0.0178426561 1704.25423 -1.2841913 1.992496e-01
## PCCF -0.0614917516 0.0281760060 648.37103 -2.1824155 2.943702e-02
## CCF -0.0048696923 0.0067670397 930.87234 -0.7196193 4.719401e-01
## CR 0.0208494387 0.0066579788 46.24753 3.1314967 3.013141e-03
## wateryr 0.0004065518 0.0001685389 1594.67111 2.4122130 1.596830e-02
## tmax_JanJun -0.0541675071 0.0282433596 624.80209 -1.9178847 5.558117e-02
```

Collinearity

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

```
## DIA_C SICOND SLOPE ASPECT BAL PCCF
## 1.141701 1.122322 1.037901 1.094085 1.091107 1.271624
## CCF CR wateryr tmax_JanJun
## 1.101390 1.039022 1.099324 1.098490
```

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.