Maize Selection Tests

Investigating genome size variation in maize and teosinte. First, I plot the data to visualize the variation across altitudinal clines.

Read in the maize data.

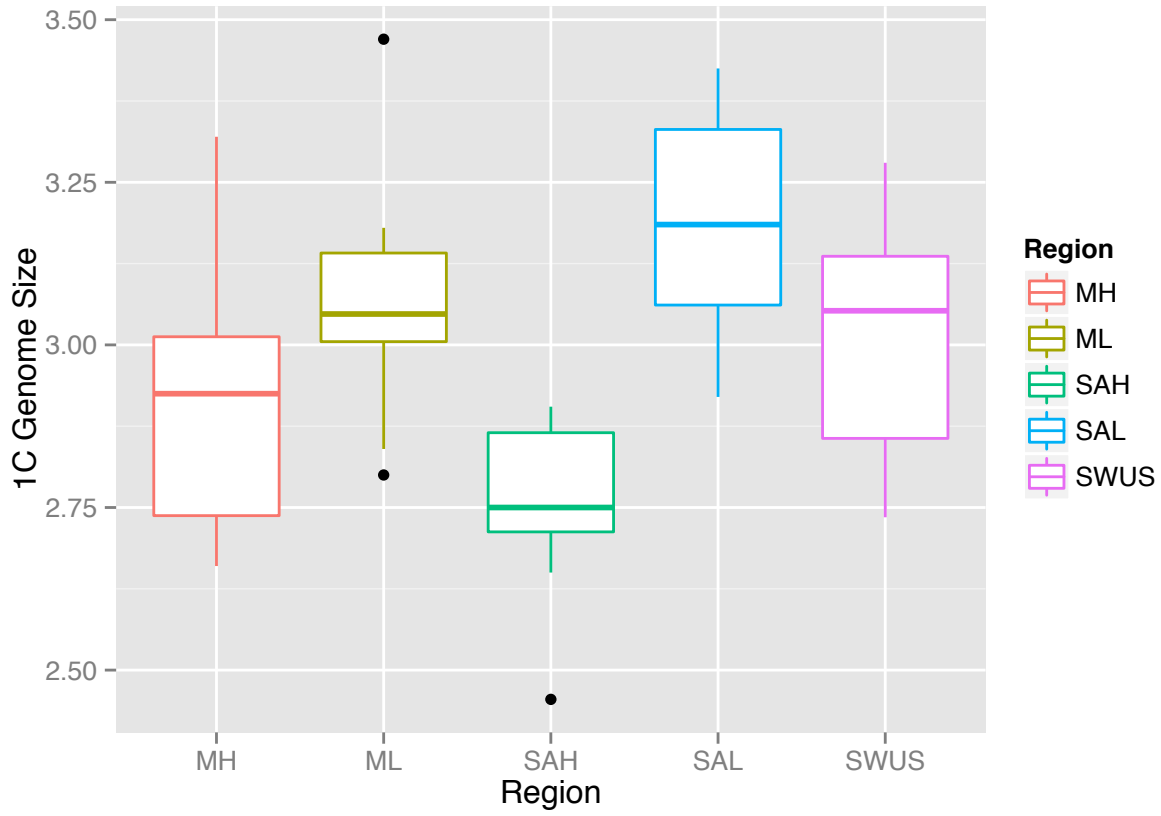
```
setwd("~/Documents/Projects/Genome_Size_Analysis")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.0.2
```

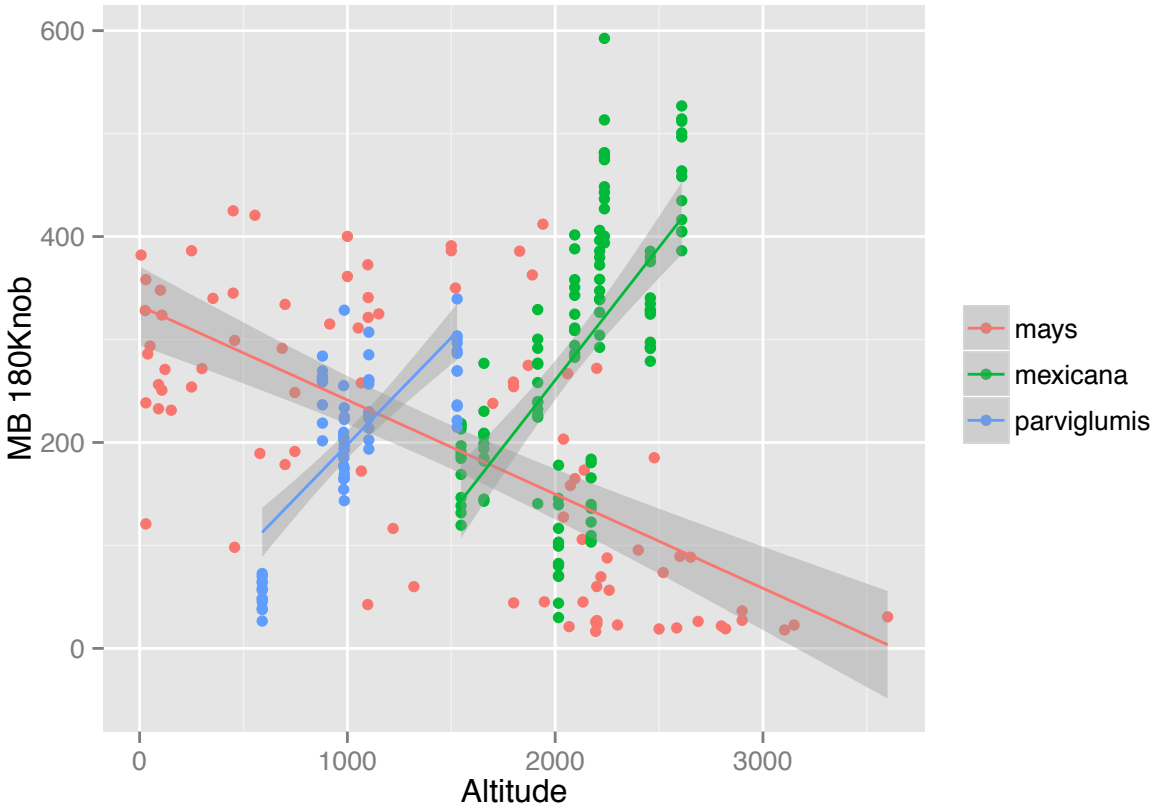
```
dataall <- read.csv("Master_Data_noNA.csv")
dataal <- subset(dataall, dataall$X1C_GS!="NA")
data <- subset(dataal, dataal$X1C_GS<3.6)
dmays <- subset(data, data$Species=="mays")
```

Plot the variation.

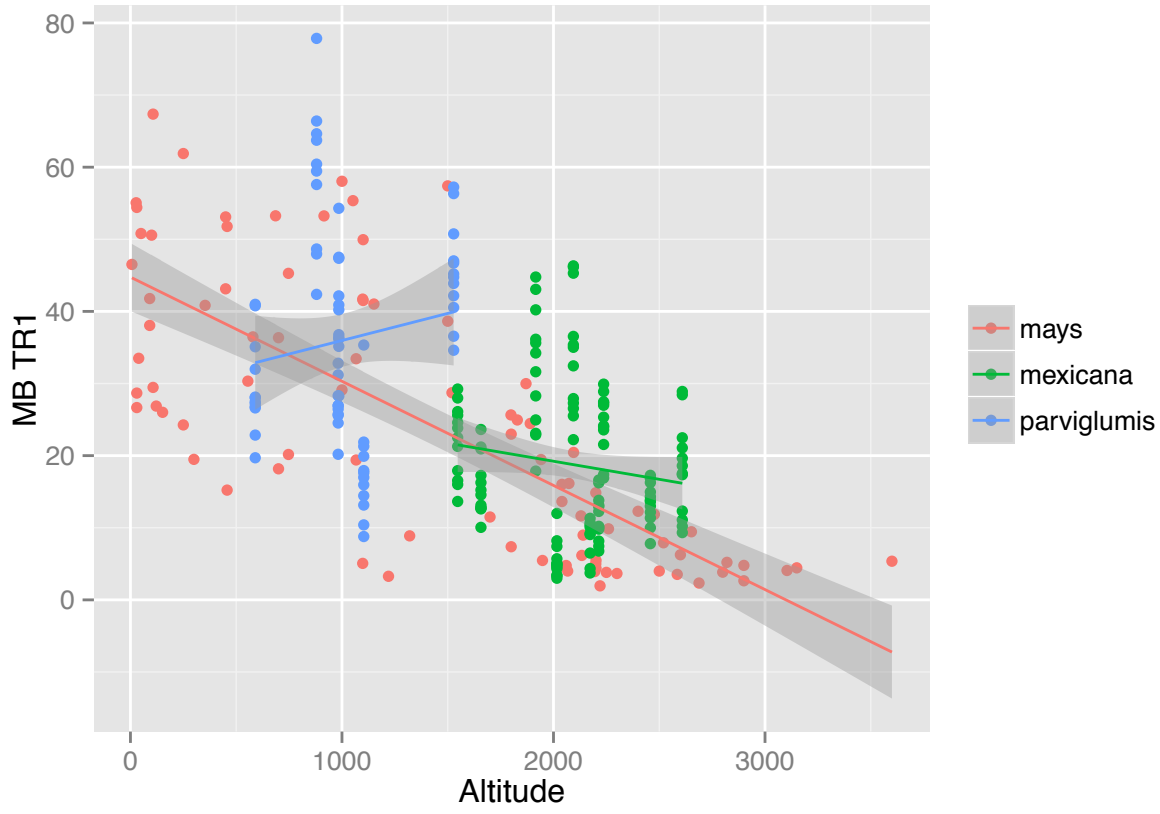
```
p1 <- ggplot(dmays, aes(Region, X1C_GS, color=Region)) + geom_boxplot()+ ylab("1C Genome Size")
p1
```



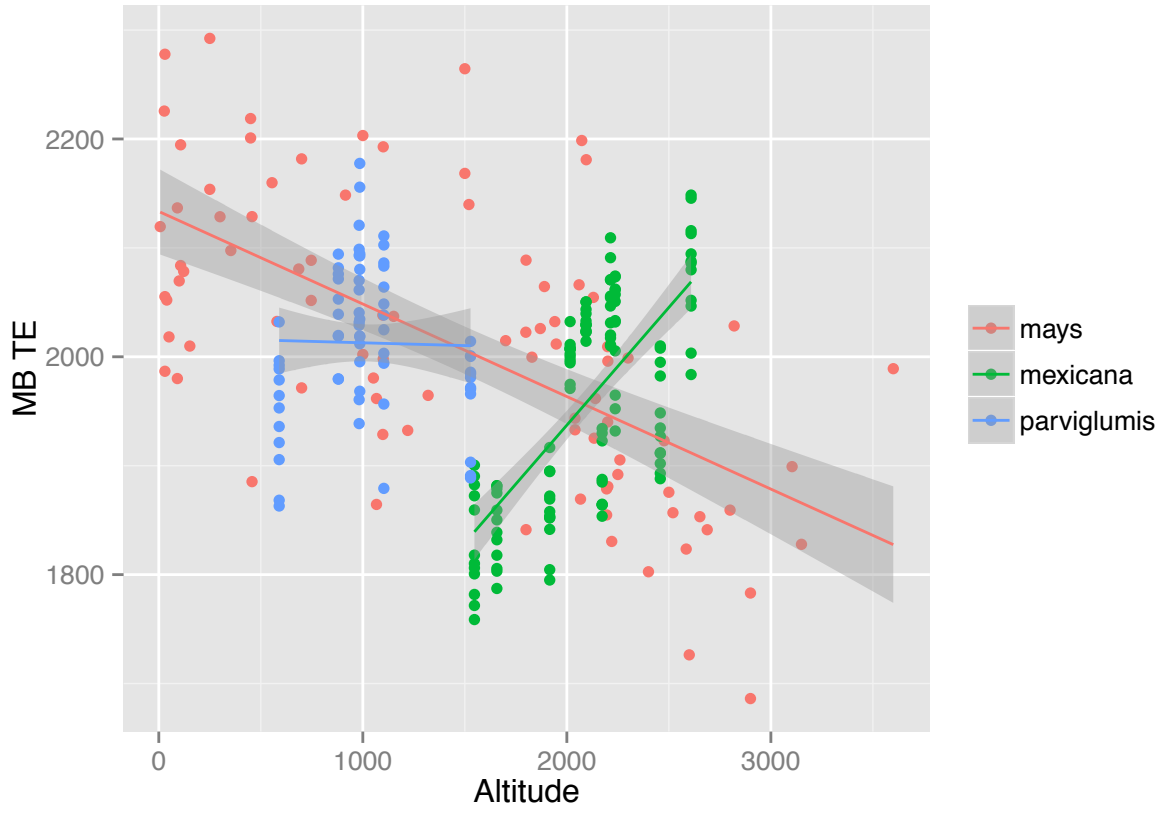

```
p2 <- ggplot(data, aes(Altitude, X180knobMB, color=Species)) + geom_point()+ ylab("MB 180Knob") + theme
p2
```




```
p <- ggplot(data, aes(Altitude, TR1MB, color=Species)) + geom_point()+ ylab("MB TR1") + theme(legend.ti
p
```




```
p <- ggplot(data, aes(Altitude, TotallTeMB, color=Species)) + geom_point()+ ylab("MB TE") + theme(legen
p
```



Next, we set up the data for studying selection on genomic elements. Read in the genotype data with this code, and generate the genetic matrix with the following code.

```
setwd("~/Documents/Projects/Genome_Size_Analysis/Github_ParallelGS/SelectionTests/")
library("rrBLUP")
```

```
## Warning: package 'rrBLUP' was built under R version 3.0.2
```

```
geno <- read.csv("~/Documents/Projects/Genome_Size_Analysis/Github_ParallelGS/SNP_data/Landrace_noSWUS_
dt <-t(geno)
A <- A.mat(dt)
```

Read in the phenotype data, and make sure the order of the samples are the same as order in the genetic matrix.

```
pheno <-read.csv("~/Documents/Projects/Genome_Size_Analysis/Github_ParallelGS/PhenotypeData/Landraces_n

tmp1 <- as.data.frame(colnames(geno))
names(tmp1)[1] <- "names"
tmp2 <- as.data.frame(pheno$FullID)
tmp <- setdiff(tmp1,tmp2)

phenoorder <- merge(tmp1,pheno, by.x="names", by.y="FullID",sort=FALSE)
```

Jeremy Berg's equation for testing selection.

```
library ( mvtnorm )
```

```
## Warning: package 'mvtnorm' was built under R version 3.0.2
```

```
EnvVarTest <- function ( phenos , kinship.mat , test.vector ) {

  # 'phenos' is a vector containing the phenotype
  # (i.e. number of repeats) for each individual; dimensions are N x 1
  # 'kinship.mat' is the kinship matrix; dimensions are N x N;
  # rows and columns need to be in the same order as the phenotypes in the vector
  # test.vector is the environmental factor of interest (in this case altitude)

  eigs <- eigen ( kinship.mat )
  # get eigendecomposition of kinship matrix
  rt.inv <- eigs$vec %*% diag ( sqrt(eigs$val) )
  # calculate inverse of the square root matrix
  rotated.phenos <- t ( rt.inv ) %*% phenos
  # rotate phenotypes from population space into principal component space
  test.vector <- test.vector / (sqrt ( 2 * sum ( test.vector^2 ) ) )
  # scale to be unit length after rotation
  #recover()
  rotated.vector <- rt.inv %*% test.vector
  # rotate environmental variable from population space into principal component space
  model <- lm ( rotated.phenos ~ 1+rotated.vector)
  # fit regression model
  r.sq <- cor.test ( rotated.phenos , rotated.vector )$estimate^2
  # get r^2
  ANOVA <- anova ( model )
  # get p value

  return ( c ( model$coef[2] , r.sq , ANOVA[5][[1]][1] )) # return
}
```

Now run the equations on each of the genetic phenotypes. Altitude is our environmental variable, and we will expand on doing this with other bioclim variables. The MB signifies megabases of the repeat, and the . is a replacement for the % symbol.

```
EnvVarTest(phenoorder$X180knobMB,A,phenoorder$Altitude)
```

```
## rotated.vector            cor
##      994.55885        0.07051        0.01526
```

```
EnvVarTest(phenoorder$X180knob.,A,phenoorder$Altitude)
```

```
## rotated.vector            cor
##       31.06811        0.07232        0.01396
```

```
EnvVarTest(phenoorder$TR1MB,A,phenoorder$Altitude)
```

```
## rotated.vector            cor
##      123.35322        0.05521        0.03250
```

```
EnvVarTest(phenoorder$TR1.,A,phenoorder$Altitude)
```

```
## rotated.vector            cor
##        3.78601        0.05539        0.03221
```

```
EnvVarTest(phenoorder$CentCMB,A,phenoorder$Altitude)
```

```
## rotated.vector            cor
##      -14.38804        0.08506        0.00747
```

```
EnvVarTest(phenoorder$CentC.,A,phenoorder$Altitude)
```

```
## rotated.vector            cor
##      -0.657070       0.088293       0.006373
```