

Comparing Growth in California Valley Oaks Between Common Gardens

Matthew Muia

Contents

1	Abstract	2
2	Introduction	3
3	Methods	5
3.1	Plotting Current Histograms of the Data	5
3.2	Creating New Data for Controlled Block Effects in Placerville	5
3.3	Creating New Data for Controlled Block Effect in Chico	6
3.4	Creating a New Data Set with Controlled Values Appended	7
3.5	Plotting Histograms of the Data After Block Control	7
3.6	Opening new data set	8
3.7	Making dictionaries of growth data	8
3.8	Writing a function to average growth data	9
4	Results	9
4.1	Plotting Histograms of the Data Before Block Control	9
4.2	Creating New Data for Controlled Block Effect in Placerville	11
4.3	Creating New Data for Controlled Block Effect in Chico	14
4.4	Creating a New Data Set with Controlled Values Appended	16
4.5	Plotting Histograms of the Data After Block Control	17
4.6	Performing 2-Group T-tests on Each Growth Variable	17
5	Discussion	18
5.1	Histograms of Data	18
5.2	2-Group T-tests	18
6	Conclusion	19
7	References	20

1 Abstract

The California Valley Oak is an important oak species endemic to California. Recent climate projections for California predict drastic changes in precipitation and temperature, which could have adverse affects on Valley Oak growth and survival (2). Dr. Sork's lab, where I work as an undergraduate researcher, is performing a study aimed at determining the potential threats climate change has on Valley Oak populations. The study began by identifying seperate populations across California, and compiling climate data unique to each. Seed samples were then taken from individuals within each of these populations (93 in total). These seeds were then planted into two common gardens, one in Chico and the other in Placerville. The common gardens are in different environemnts, so I predict that all three measured growth variables(tree height, leaf number, primary branch number) will be significantly different between the two gardens. After controlling the data for block-effects, the results of my data analysis show that there are significant differences ($p < 0.001$) for all three of the growth variables between the two gardens. Threfore, further analysis should be done on genetic and environemtal data to find which variables are the best predictors of the change in growth.



Figure 1: California Valley Oak Species Range

2 Introduction

The California Valley Oak is an important oak species endemic to California. The Valley Oak is found throughout California, from riparian forests to savannas, and it creates structure within these ecosystems that other species can inhabit (1).

Recent climate projections for California predict drastic changes in precipitation and temperature, which could have adverse effects on Valley Oak growth and survival (2). If Valley Oak populations are affected by these changes in climate, there could be further detrimental changes in the vegetation and animals that live within the Valley Oak's ecological range. There are also genetic components to the possible effects, as different genotypes may confer different phenotypic plasticity in the face of climate change(3).

Dr. Sork's lab, where I work as an undergraduate researcher, is performing a study aimed at determining the potential threats climate change has on Valley Oak populations. The study began by identifying separate populations across California, and compiling climate data unique to each. Seed samples were then taken from individuals within each of these populations (93 in total). These seeds were then planted into two common gardens, one in Chico and the other in Placerville. The gardens contain the same proportions of individuals from each population.

The purpose of these common gardens is to enable our lab to compare the growth of individuals from each population to each other. The climate data from the parent populations then allows us to determine if variations in the climates from which each population is adapted to explains the variation in growth we observe in the common gardens.

Since there is a lot that can be analyzed within this study, my project will be focusing on the differences between the growth variables of the two common gardens.



Figure 2: California Valley Oak

3 Methods

3.1 Plotting Current Histograms of the Data

I will first plot histograms for each growth variable (tree height, number of leaves, number of primary branches) so that I can visualize the distribution of the data. I will make three sets of histograms, one for the entire data set, and the other two for each common garden on its own.

```
# reading in my csv with all the data
data <- read.csv("OakGrowthData1.csv")
head("OakGrowthData1.csv")

# seperate the data between the two gardens, so each can be controlled seperately
IFG <- subset(x = data, Site == "IFG")
Chico <- subset(x = data, Site == "Chico")

# making a 3 by 3 window for all 9 histograms to be viewed in together
par(mfrow = c(3,3))

# histograms for the entire data set
hist(data$Height, xlab = "Height", main = "Data-Height")
hist(data$Leaves, xlab = "Log of Leaves", main = "Data-Leaves")
hist(data$Branches, xlab = "Branches", main = "Data-Branches")

# histograms for Placerville data, colored blue
hist(IFG$Height, col = "blue", xlab = "Height", main = "Placerville-Height")
hist(IFG$Leaves, col = "blue", xlab = "Log of Leaves", main = "Placerville-Leaves")
hist(IFG$Branches, col = "blue", xlab = "Branches", main = "Placerville-Branches")

# histograms for Chico data, colored green
hist(Chico$Height, col = "green", xlab = "Height", main = "Chico-Height")
hist(Chico$Leaves, col = "green", xlab = "Log of Leaves", main = "Chico-Leaves")
hist(Chico$Branches, col = "green", xlab = "Branches", main = "Chico-Branches")
```

3.2 Creating New Data for Controlled Block Effects in Placerville

When the common gardens were planted, large areas were needed to fit all of the trees. Since such large areas were needed, we could not assure that all the trees would receive the exact same environmental conditions throughout the garden (watering, soil quality). To control for this, we drew out blocks within each garden, and trees were randomly placed within one of these blocks. This method ensures that varying micro-environmental conditions are not incorrectly mis-characterized as a result of genetic differences. I will be using the function

shown below to control for block effects in each of the gardens, for each of the growth variables. The function creates Regression Diagnostic Plots, which are used to test the validity of a linear model. The plot of interest for each set of four is the first, which is the residuals plot. The residuals plot uses a linear model to calculate predicted values, then subtracts each observed value by its predicted value, which equals the residual. The residual values are plotted, and a horizontal line is expected within the plot. I will use these residuals to calculate a new data set, which will be controlled for block effects. I did this by taking all the calculated residual values, and adding the mean of the data to each of these values. I then appended the new values to the data sheet as a new column.

```
# Creating graph for Placerville-Height seperated by block
IFGBlockHeight <- lm(IFG$Height ~ as.factor(IFG$Block), data = IFG)
par(mfrow = c(2,2))
plot(IFGBlockHeight)
# appending residuals+mean to Placerville data
IFG[,12] <- ((IFGBlockHeight$residuals)+mean(IFG$Height))

# Creating graph for Placerville-Leaves seperated by block
IFGBlockLeaves <- lm(IFG$Leaves ~ as.factor(IFG$Block), data = IFG)
par(mfrow = c(2,2))
plot(IFGBlockLeaves)
# appending residuals+mean to Placerville data
IFG[,13] <- ((IFGBlockLeaves$residuals)+mean(IFG$Leaves))

# Creating graph for Placerville-Branches seperated by block
IFGBlockBranches <- lm(IFG$Branches ~ as.factor(IFG$Block), data = IFG)
par(mfrow = c(2,2))
plot(IFGBlockBranches)
# appending residuals+mean to Placerville data
IFG[,14] <- ((IFGBlockBranches$residuals)+mean(IFG$Branches))

# re-naming new columns for clarity
colnames(IFG)[colnames(IFG)=="V12"] <- "NewHeight"
colnames(IFG)[colnames(IFG)=="V13"] <- "NewLeaves"
colnames(IFG)[colnames(IFG)=="V14"] <- "NewBranches"
```

3.3 Creating New Data for Controlled Block Effect in Chico

I will perform the same operations with the Chico data as I did with the Placerville data.

```
# Creating graph for Chico-Height seperated by block
ChicoBlockHeight <- lm(Chico$Height ~ as.factor(Chico$Block), data = Chico)
par(mfrow = c(2,2))
plot(ChicoBlockHeight)
# appending residuals+mean to Chico data
```

```

Chico[,12] <- ((ChicoBlockHeight$residuals)+mean(Chico$Height))

# Creating graph for Chico-Leaves seperated by block
ChicoBlockLeaves <- lm(Chico$Leaves ~ as.factor(Chico$Block), data = Chico)
par(mfrow = c(2,2))
plot(ChicoBlockLeaves)
# appending residuals+mean to Chico data
Chico[,13] <- ((ChicoBlockLeaves$residuals)+mean(Chico$Leaves))

# Creating graph for Chico-Branches seperated by block
ChicoBlockBranches <- lm(Chico$Branches ~ as.factor(Chico$Block), data = Chico)
par(mfrow = c(2,2))
plot(ChicoBlockBranches)
# appending residuals+mean to Chico data
Chico[,14] <- ((ChicoBlockBranches$residuals)+mean(Chico$Branches))

# re-naming new columns for clarity
colnames(Chico)[colnames(Chico)=="V12"] <- "NewHeight"
colnames(Chico)[colnames(Chico)=="V13"] <- "NewLeaves"
colnames(Chico)[colnames(Chico)=="V14"] <- "NewBranches"

```

3.4 Creating a New Data Set with Controlled Values Appended

```

# attaching the seperated data into a combined format with controlled values
data2 <- rbind(IFG, Chico)

# now I will output this as a new data file
write.csv(data2, "~/Downloads/Block-Controlled-Data.csv", append = F, quote = F, sep = ",")

```

3.5 Plotting Histograms of the Data After Block Control

I am now plotting the same histograms as before, but with the data that has been controlled for block effects.

```

# seperate the data between the two gardens, so each can be controlled seperately
IFG2 <- subset(x = data2, Site == "IFG")
Chico2 <- subset(x = data2, Site == "Chico")

# making a 3 by 3 window for all 9 histograms to be viewed in together
par(mfrow = c(3,3))

# histograms for the entire data set

```



```

hist(data2$NewHeight, xlab = "Height", main = "Data-Height")
hist(data2$NewLeaves, xlab = "Log of Leaves", main = "Data-Leaves")
hist(data2$NewBranches, xlab = "Branches", main = "Data-Branches")

# histograms for Placerville data, colored blue
hist(IFG2$NewHeight, col = "blue", xlab = "Height", main = "Placerville-Height")
hist(IFG2$NewLeaves, col = "blue", xlab = "Log of Leaves", main = "Placerville-Leaves")
hist(IFG2$NewBranches, col = "blue", xlab = "Branches", main = "Placerville-Branches")

# histograms for Chico data, colored green
hist(Chico2$NewHeight, col = "green", xlab = "Height", main = "Chico-Height")
hist(Chico2$NewLeaves, col = "green", xlab = "Log of Leaves", main = "Chico-Leaves")
hist(Chico2$NewBranches, col = "green", xlab = "Branches", main = "Chico-Branches")

```

3.6 Opening new data set

I will now be utilizing a function that I wrote within python to find averages of the data that can be used to determine effect size between the two common gardens.

```

file = "/home/eeb177-student/Downloads/Block-Controlled-Data.csv"
with open(file, 'r') as ff:
    ff.readline()
    growth_data = ff.readlines()

```

3.7 Making dictionaries of growth data

```

from collections import defaultdict

# initialize dictionaries for growth data
growth_height = {}
growth_leaves = {}
growth_branches = {}

for line in growth_data:
    garden = str(line.split(",")[1])
    new_height = float(line.split(",")[12])
    new_leaves = float(line.split(",")[13])
    new_branches = float(line.split(",")[14])

    #populate dictionaries with garden name as key, and growth data as values
    growth_height[garden] = new_height

```



```

growth_leaves[garden] = new_leaves
growth_branches[garden] = new_branches

```

3.8 Writing a function to average growth data

```

def average_dict(DictName):
    from scipy import mean
    print((list(map(lambda x: mean(DictName[x]), DictName))))

average_dict(growth_height)
average_dict(growth_leaves)
average_dict(growth_branches)

```

4 Results

Each of the below sections displays the results from the code outlined in the above Methods section. Explanations of these results can be found in the Discussion section.

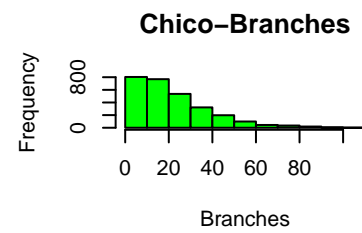
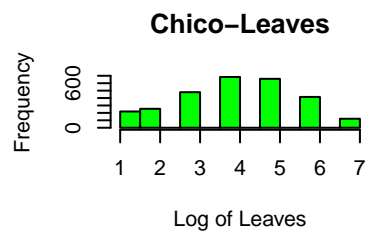
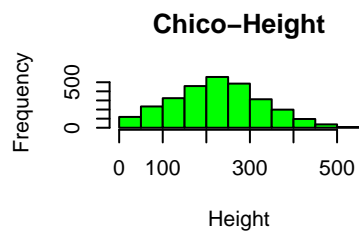
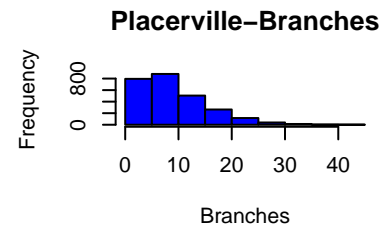
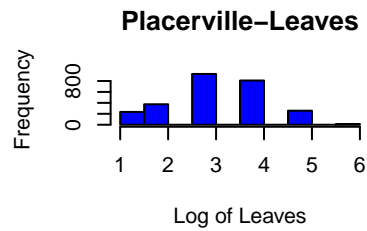
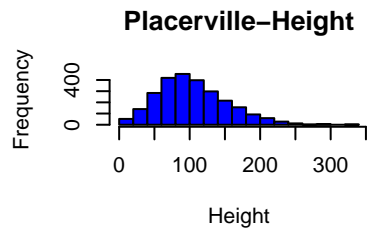
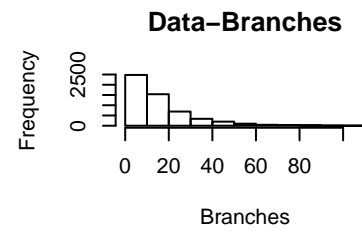
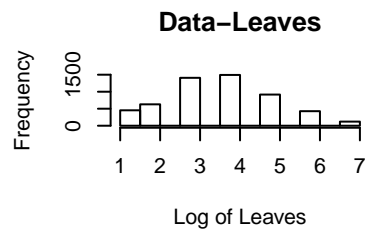
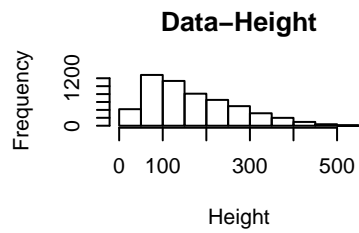
4.1 Plotting Histograms of the Data Before Block Control

```

# reading in my csv with all the data
data <- read.csv("~/Desktop/eeb-177/eeb-177-final-project/OakGrowthData1.csv")
# printing the first few lines of the file
head(data)

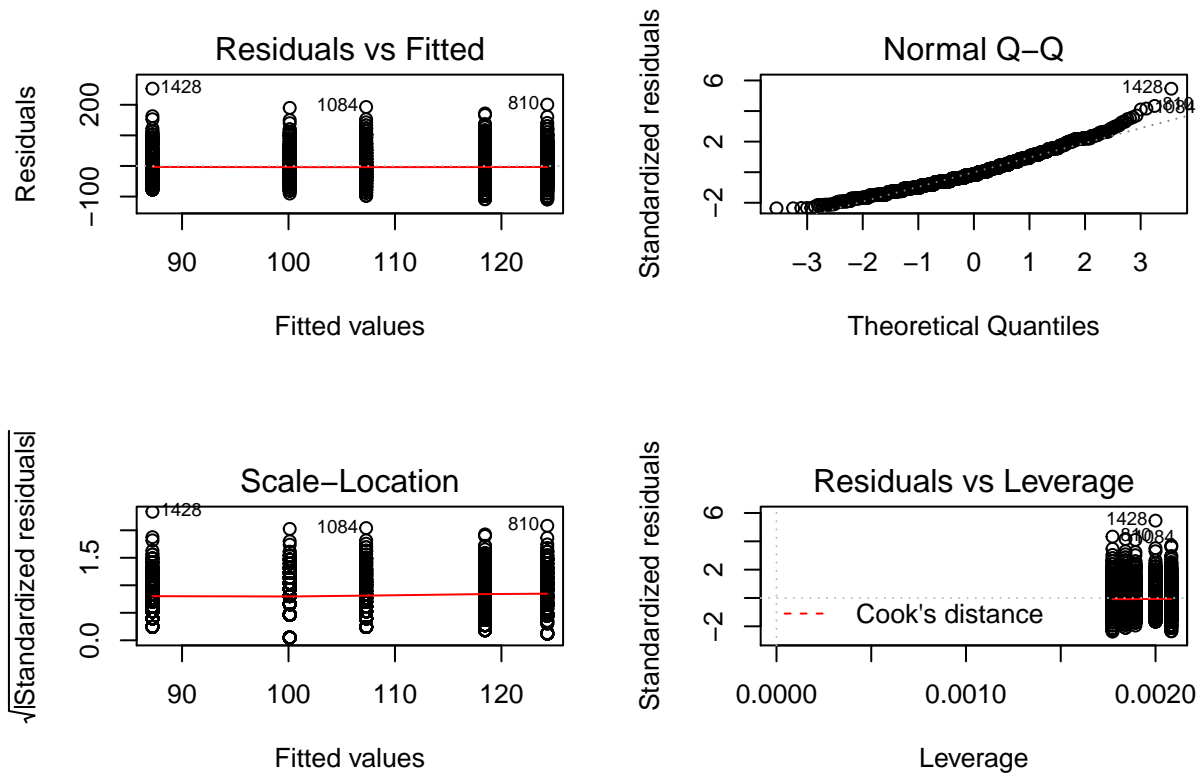
```

##	Site	Section	Block	Provenence	accessID	Height	Leaves	Branches	lat
## 1	IFG		1	ALH	1	165	4	7	37.96990
## 2	IFG		2	ALH	1	155	3	6	37.96990
## 3	IFG		4	ALH	1	90	5	8	37.96990
## 4	IFG		5	ALH	1	70	3	7	37.96990
## 5	IFG		1	ALH	2	165	4	7	37.96964
## 6	IFG		2	ALH	2	105	3	4	37.96964
##		lon	Sorting.Row						
## 1	-122.1697		1						
## 2	-122.1697		2						
## 3	-122.1697		3						
## 4	-122.1697		4						
## 5	-122.1762		5						
## 6	-122.1762		6						

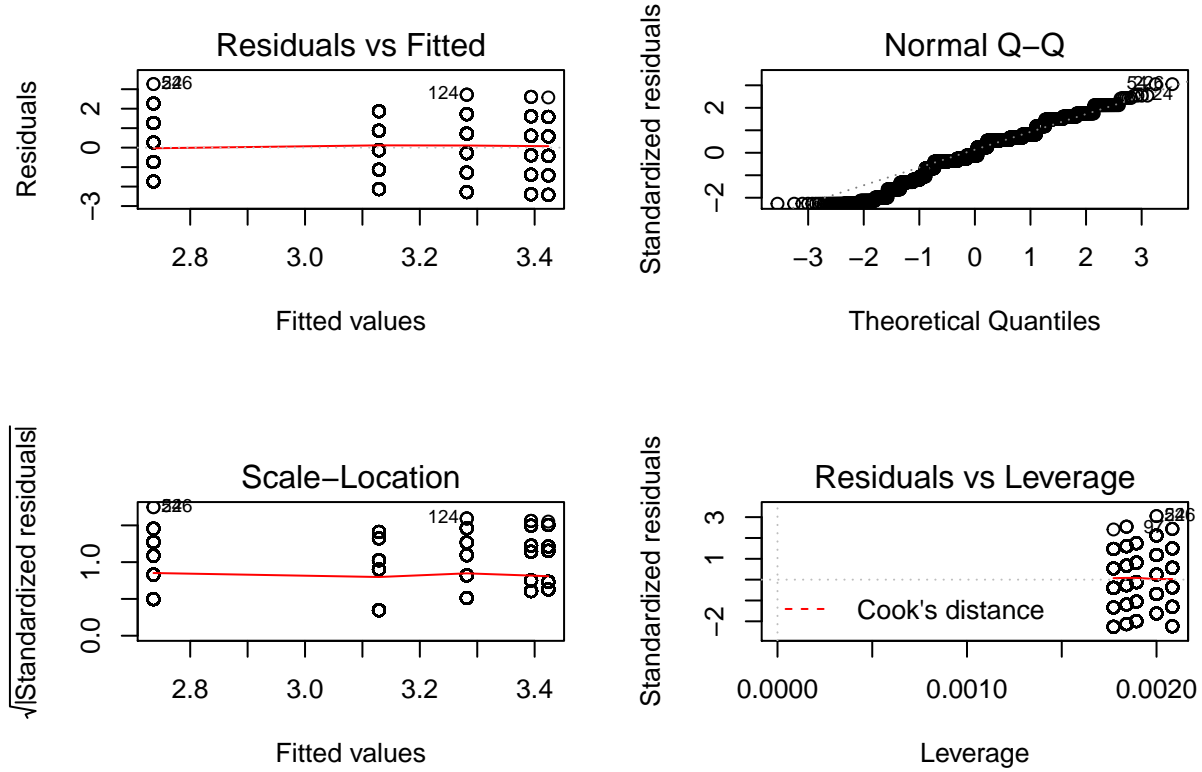


4.2 Creating New Data for Controlled Block Effect in Placerville

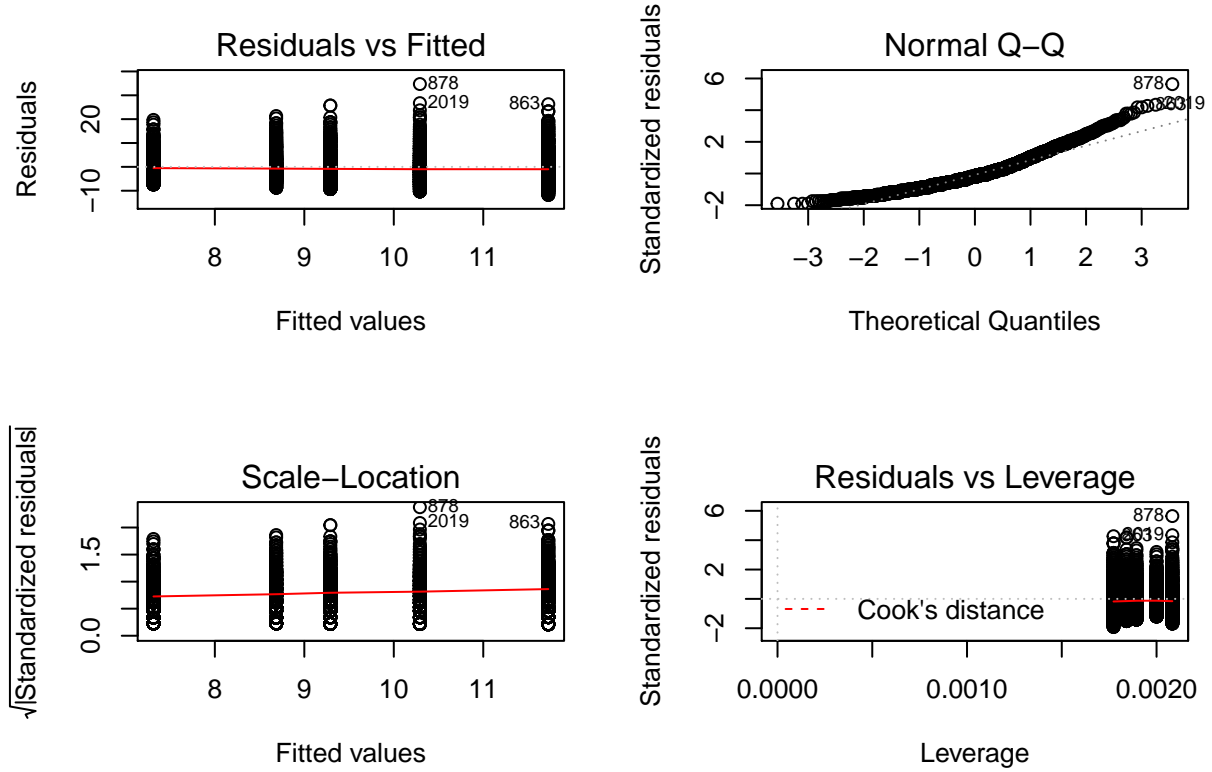
4.2.1 Placerville-Height Block Effect



4.2.2 Placerville-Leaves Block Effect

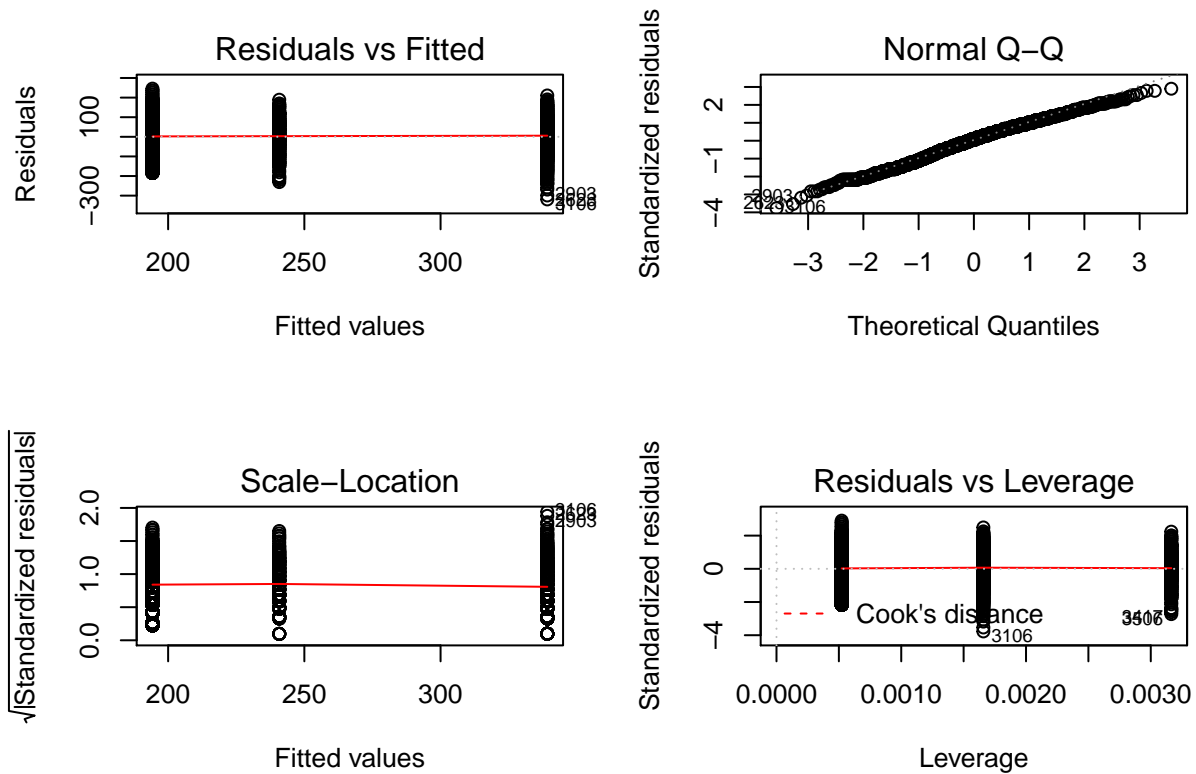


4.2.3 Placerville-Branches Block Effect

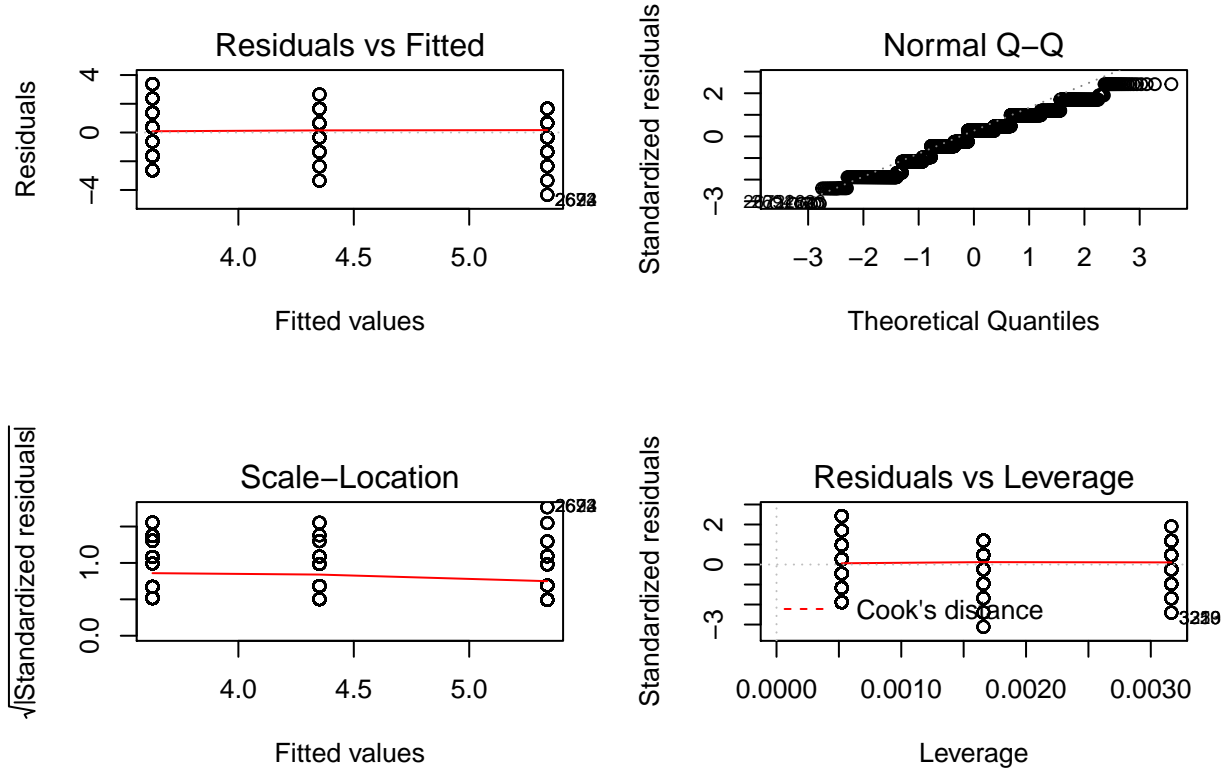


4.3 Creating New Data for Controlled Block Effect in Chico

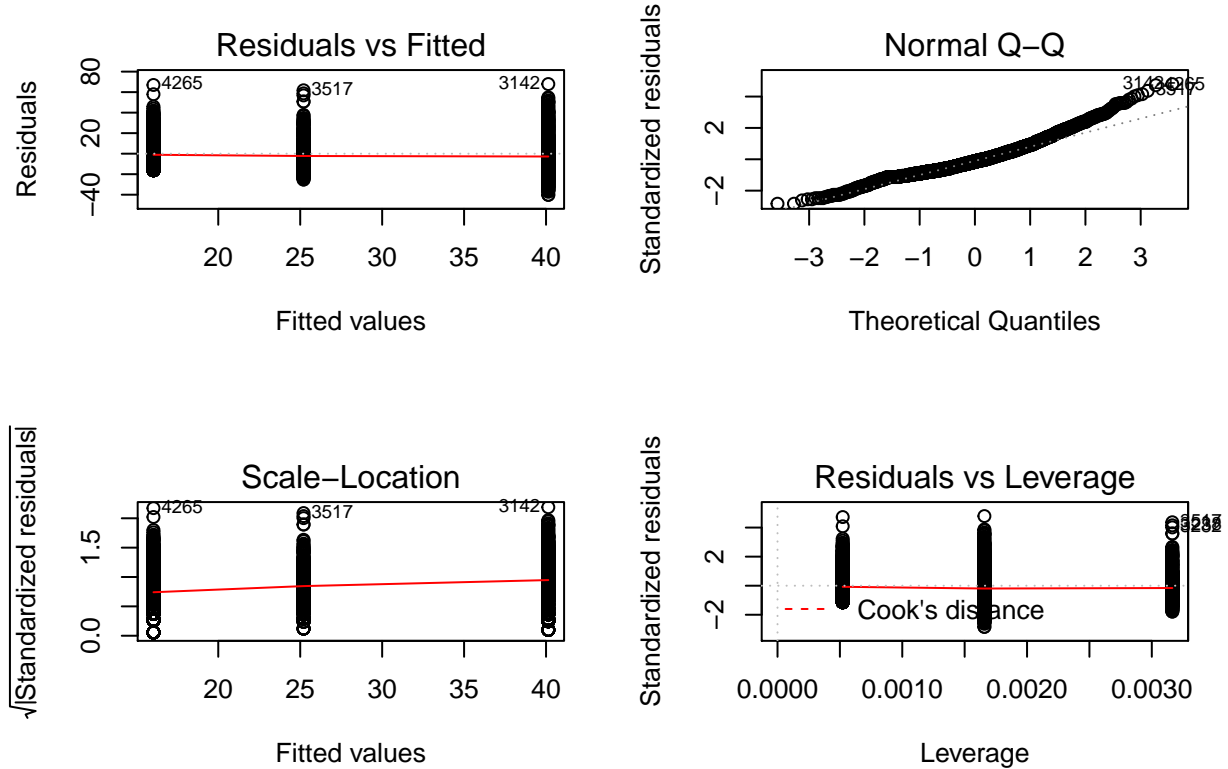
4.3.1 Chico-Height Block Effect



4.3.2 Chico-Leaves Block Effect



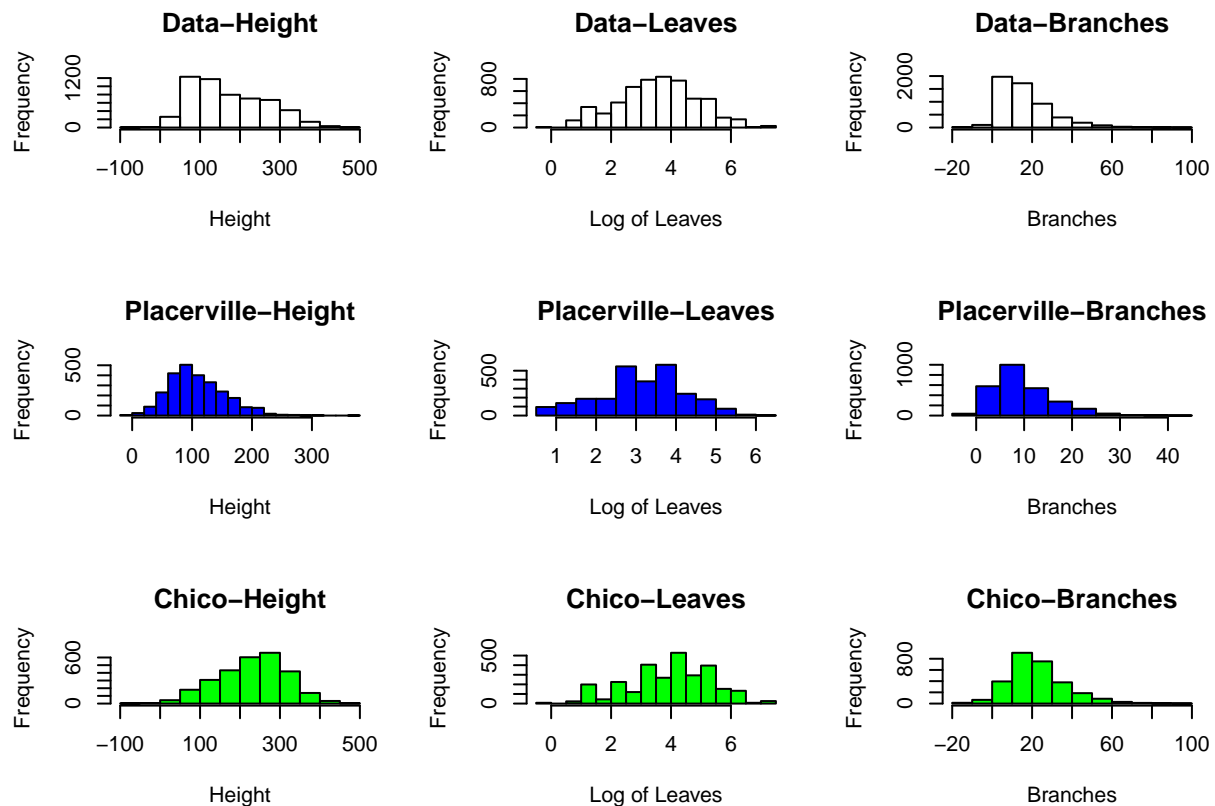
4.3.3 Chico-Branches Block Effect



4.4 Creating a New Data Set with Controlled Values Appended

```
## Site Section Block Provenience accessID Height Leaves Branches lat
## 1 IFG 1 ALH 1 165 4 7 37.96990
## 2 IFG 2 ALH 1 155 3 6 37.96990
## 3 IFG 4 ALH 1 90 5 8 37.96990
## 4 IFG 5 ALH 1 70 3 7 37.96990
## 5 IFG 1 ALH 2 165 4 7 37.96964
## 6 IFG 2 ALH 2 105 3 4 37.96964
## lon Sorting.Row NewHeight NewLeaves NewBranches
## 1 -122.1697 1 148.41615 3.773947 4.772858
## 2 -122.1697 2 144.26633 2.803956 5.208142
## 3 -122.1697 3 90.43185 4.915938 8.208833
## 4 -122.1697 4 77.61103 3.068918 7.814203
## 5 -122.1762 5 148.41615 3.773947 4.772858
## 6 -122.1762 6 94.26633 2.803956 3.208142
```

4.5 Plotting Histograms of the Data After Block Control



4.6 Performing 2-Group T-tests on Each Growth Variable

```
t.test(IFG$NewHeight,Chico$NewHeight,alternative="less", var.equal=TRUE)
```

```
##
## Two Sample t-test
##
## data: IFG$NewHeight and Chico$NewHeight
## t = -65.529, df = 5445, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -119.5002
## sample estimates:
## mean of x mean of y
## 107.7247 230.3023
```

```
t.test(IFG$NewLeaves,Chico$NewLeaves,alternative="less", var.equal=TRUE)
```

```
##
## Two Sample t-test
```

```
##
## data: IFG$NewLeaves and Chico$NewLeaves
## t = -25.844, df = 5445, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.8193333
## sample estimates:
## mean of x mean of y
##  3.197706  4.072740

t.test(IFG$NewBranches,Chico$NewBranches,alternative="less", var.equal=TRUE)

##
## Two Sample t-test
##
## data: IFG$NewBranches and Chico$NewBranches
## t = -42.281, df = 5445, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -12.20454
## sample estimates:
## mean of x mean of y
##  9.499809 22.198446
```

5 Discussion

5.1 Histograms of Data

The histograms of the overall data set appeared to become less scewed and less varied after controlling for block effects. This was expected because controlling for blocking effects has a greater effect on individuals far away from the mean than those closer to the mean. Controlling for the blocking effects on each growth variable allowed me to more accurately compare the differences between the two common gardens.

5.2 2-Group T-tests

The t-test from the height data suggests that there is a significant difference between the Placerville and Chico data ($p < 0.001$) with an effect size of $(230.3 - 107.7) 122.6$ cm. The Chico height data is significantly greater than the Placerville data.

The t-test from the leaf number data suggests that there is a significant difference between the Placerville and Chico data ($p < 0.001$) with an effect size of $(4.07 - 3.20) 0.87$ log valued leaves. The Chico leaf number data is significantly greater than the Placerville data.

The t-test from the branches data suggests that there is a significant difference between the Placerville and Chico data ($p < 0.001$) with an effect size of $(22.20 - 9.50) / 12.7$ main branches. The Chico main branch number data is significantly greater than the Placerville data.

6 Conclusion

From my analysis, I found that all three of the growth variables measured in Chico were significantly greater than those measured in Placerville. Controlling for block effects before this analysis allowed me to perform t-test without the outliers of blocks that grew significantly more because of favorable micro-environments. These results show that the differing environments between the two common gardens had a significant effect on the growth of the oaks. Therefore, there is evidence to suggest that further analysis of this data should be done in order to discover what variables are associated with the change in growth. This analysis would require detailed climate and soil data from each of the gardens, which has not yet been compiled into a workable data set. Further analysis of the differences in growth between individual provenances could inform a possible genetic effect in the growth of the oaks. By identifying which variables (climate and/or genetic) predict changes in growth, conservation efforts can be taken to mitigate the effects of climate change on the species range of the California Valley Oak.

7 References

1. McLaughlin, B. C., & Zavaleta, E. S. (2012). Predicting species responses to climate change: demography and climate microrefugia in California valley oak (*Quercus lobata*). *Global Change Biology*, 2301-2312.
2. Sork, V. L. (2016). Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. *Evolutionary Applications*, 291-310.
3. Sork, V. L., Davis, F. W., Westfall, R., Flint, A., Ikegami, M., Wang, H. F., et al. (2010). Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. *Molecular Ecology*, 19, 3806-3823.