Appendix

for

'Examining Fouquieria splendens in an environmental and ecological context: Effect of topography and interspecific neighbors on ocotillo morphology and distribution';

Anastasia Bernat, Acacia Tsz So Tang, Allegra Steenson, Eric Larsen, Emma Greig

Contents

1 Details of the Analyses

This document was generated by R Markdown on 2022-12-20 using R version 4.2.1 (2022-06-23). The document provides the step-by-step analytical methods used in the manuscript by Anastasia Bernat (AVB), Acacia Tsz So Tang (ATST), Allegra Steenson (AS), Eric Larsen (EL), and Emma Greig (EG). Draft scripts were written by AVB and ATST between 2019-06-01 and 2021-01-01 until being distilled and complied by AVB at the University of Chicago into this comprehensive script. All draft scripts can be viewed in the GitHub repository, ocotillo-research (https://github.com/avbernat/ocotillo-research).

All code and output from the statistical analyses are shown. Code for data cleaning and the generation of plots is not displayed, but can be viewed in the **appendix.Rmd** file and its accompanying sourced scripts. To repeat analyses and the generation of plots, all data files and sourced scripts should follow the directory structure presented in the ocotillo-research repository.

1.1 Description of the Data

Ocotillos, *Fouquieria splendens*, were measured in Summer 2019 in the Sonoran Desert at Organ Pipe Cactus National Monument

1.2 Abbreviations Used in the Data and Code

1.3 Data Transformations

- `_b` - a column name that ends in `_b` is a column that has been recodified into binary data (0's and 1's). Example columns:
- `_c` - a column name that ends in `_c` is a column that has been centered. Example columns:
- `_s` - a column name that ends in `_s` is a column that has been standardized. Example columns:
- `log` - a column name that starts with `log` is a column that has been log transformed. Example columns:
- `_baj` - a dataset that ends in `_baj` is a dataset that only contains ocotillo measurements from ocotillos across a bajada. Example datasets: ocos_baj, segs_baj

1.4 Read in Libraries

```
library(dplyr)
library(outliers) # dixon.test
require(FactoMineR) # PCA function
library(factoextra) # get_eigenvalue function
library(corrplot) # cor.mtest
library(gridExtra)
library(lme4)    # ** this is read in another function
library(RColorBrewer)
```

1.5 Read Source Scripts

```
source("src/cleaning_data.R")      # clean_ocos_data(), clean_segs_data()
source("src/regression_output.R")  # tidy_regression()
source("src/diagnostics.R")        # plot_diagnostic()
source("src/pretty_reg.R")         # rename_regformula() and rename_regformulaME()
source("src/pca.R")                # PCA_graphs(), run_corrplots()
source("src/model_metrics.R")      # calculate_lk_weights()
```

1.6 Read the Data

```r
ocotillo_data = read.csv("data/General_Oco_Data3.csv",
                         fileEncoding="UTF-8-BOM", stringsAsFactors=TRUE)
branch_data = read.csv("data/branch_lengths_long.csv")
segment_data = read.csv("data/Terminal_5Segs.csv")

branches = branch_data  %>%
            group_by(Tree) %>%
            summarize(Mean_BranchLength = mean(BranchLength, na.rm=TRUE),
                      Median_BranchLength = median(BranchLength, na.rm=TRUE),
                      Max_BranchLength = max(BranchLength, na.rm=TRUE),
                      Min_BranchLength = min(BranchLength, na.rm=TRUE),
                      BranchLength_IQR = IQR(BranchLength, na.rm=TRUE))
segments = segment_data %>%
            group_by(Tree) %>%
            summarize(Mean_Terminal_SegmentLength = mean(Length, na.rm=TRUE),
                      Median_Terminal_SegmentLength = median(Length, na.rm=TRUE),
                      Max_Terminal_SegmentLength = max(Length, na.rm=TRUE),
                      Min_Terminal_SegmentLength = min(Length, na.rm=TRUE),
                      IQR_Terminal_SegmentLength = IQR(Length, na.rm=TRUE))


ocotillo_data$Median_BranchLength = branches$Median_BranchLength # other typos
ocotillo_data$BranchLength_IQR = branches$BranchLength_IQR # other typos
ocotillo_data$Median_TerminalSeg = segments$Median_Terminal_SegmentLength
ocotillo_data$Terminal_SegIQR = segments$IQR_Terminal_SegmentLength

ocos = clean_ocos_data(ocotillo_data)                   # all ocotillos
ocos_baj = ocos[1:20,]                                  # only ocotillos across the bajada
segs = clean_segs_data(segment_data, ocotillo_data) # all ocotillos
segs_baj = segs[1:1000,]                                # only ocotillos across the bajada
```

1.7 Normality

All measurements followed log-normal distributions except for circumference, median branch length, number of nodes, and distance to the nearest arroyo, which were normally distributed. Ocotillo data were log-transformed before analyses to meet assumptions of normality, linear regressions, and homogeneity for parametric analyses.
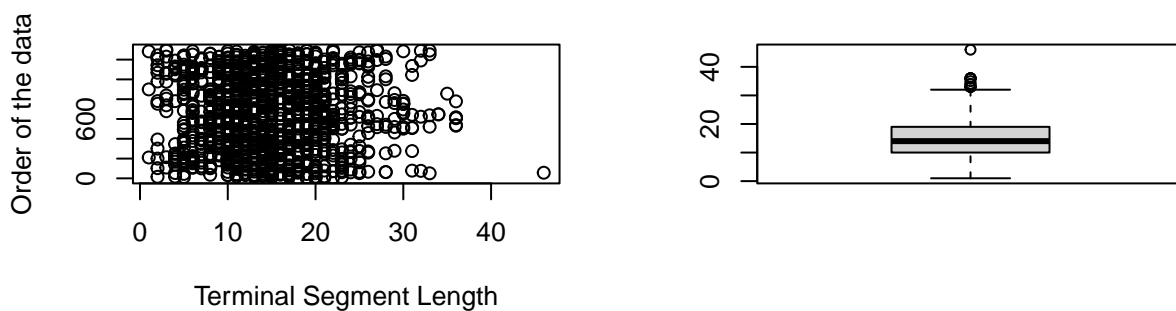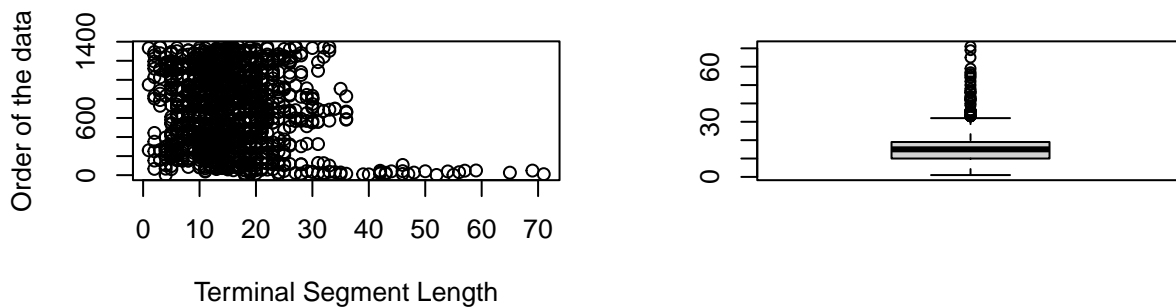
1.8 Outliers

Terminal segment length IQR has an outlier - Ocotillo 1 (see graphs.Rmd). In turn, the outlier was removed and logHeight is predicted with a smaller dataset containing all ocotillos except Ocotillo 1.

```r
# Cleveland Dotplot & Boxplot
par(mfrow=c(2,2))
x = segs$Length
y = seq(1, length(x),)
plot(x,y, ylab="Order of the data", xlab="Terminal Segment Length")
boxplot(segs$Length)

segs = segs[segs$Tree != 1,] # remove Plant 1 outlier
```

3

```
x = segs$Length
y = seq(1, length(x),)
plot(x,y, ylab="Order of the data", xlab="Terminal Segment Length")
boxplot(segs$Length)
```



```
dixon.test(ocos$Terminal_SegIQR)
```

```
##
##   Dixon test for outliers
##
## data:  ocos$Terminal_SegIQR
## Q = 0.67857, p-value < 2.2e-16
## alternative hypothesis: highest value 19 is an outlier
```

```
ocos_data = ocotillo_data[ocotillo_data$Tree != 1,]
segs_data = segment_data[segment_data$Tree != 1,]

segs_data = segs_data[complete.cases(segs_data$Length),] # also need to remove rows with NA Le

# rerun data cleaning to generate newly transformed columns
ocos = clean_ocos_data(ocos_data)
ocos_baj = ocos[1:20,]
segs = clean_segs_data(segs_data, ocos_data)
segs_baj = segs[1:1000,]
```

2   Ocotillo Morphology

Analyses below are multiple variate models of Fouquieria splendens morphology for ocotillos located
on both a bajada and a plain in Organ Pipe National Monument, Arizona. All models were grouped
by their response variable and ordered by their ascending AIC values. Dataset "ocos" indicates all
individuals measured on the bajada and plain while "ocos_baj" indicates only the individuals measured
on the bajada. Ocotillos located on the bajada were encoded with site = 0 while ocotillos on the plain
were encoded with site = 1. Interspecific neighbor group is split between two types – shrub and cactus

– where cactus = 0 and shrub = 1.

2.1 Addressing Multicollinearity

Elevation and site were highly correlated (R = -0.86), which led to multicollinearity. In turn, we removed any interactions between elevation and site in the model comparison process in order to minimize relationships that were spurious.

```
cor(ocos$Site, ocos$Elevation_c)
```

```
## [1] -0.8576868
```

2.2 Principal Component Analysis

2.2.1 Without and With IQR Variables

2.2.1.1 Plain and Bajada

```r
# Without IQR Vars:
d = ocos[, c("Height", "Circumference",
             "Number_Branches", "Median_BranchLength",
             "Num_Nodes", "Median_TerminalSeg")]
abbreviations = c("Height", "Circ", "NB", "M BL", "Nodes", "M TSL")
colnames(d) = abbreviations


par(mfrow=c(2,2))
output_list = PCA_graphs(dataset=d, PCA_title="Without IQR Variables", run_corr=FALSE)
eig_val = output_list[[1]]
var_val = output_list[[2]]
scree = output_list[[3]]
pca_graph = output_list[[5]]
pca_components = output_list[[6]]



# With IQR Vars:
d_IQR = ocos[, c("Height", "Circumference",
             "Number_Branches", "Median_BranchLength",
             "Num_Nodes", "Median_TerminalSeg",
             "BranchLength_IQR",
             "Terminal_SegIQR")]
abbreviations = c("Height", "Circ", "NB", "M BL", "Nodes", "M TSL","BL IQR", "TSL IQR")
colnames(d_IQR) = abbreviations

par(mfrow=c(2,2))
output_list = PCA_graphs(dataset=d_IQR, PCA_title="With IQR Variables", run_corr=FALSE)
eig_val = output_list[[1]]
var_val_IQR = output_list[[2]]
scree = output_list[[3]]
pca_graph_IQR = output_list[[5]]
pca_components_IQR = output_list[[6]]
```
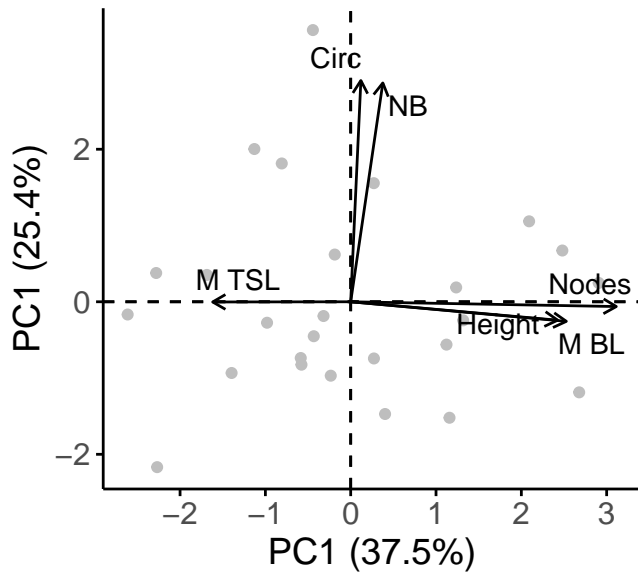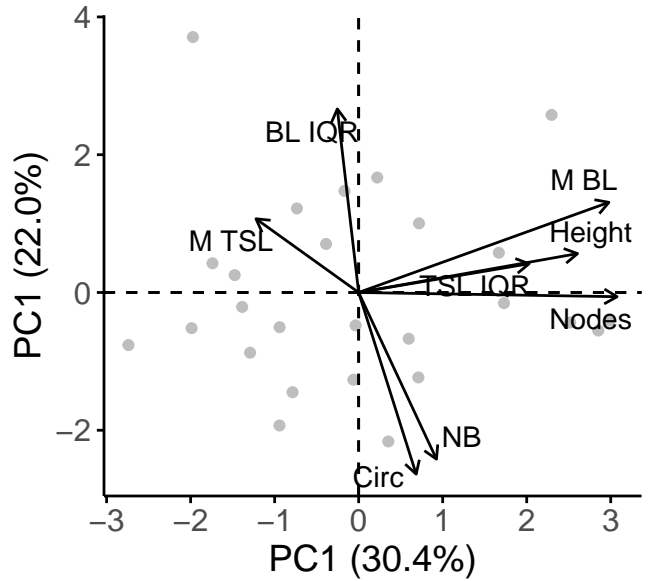
```
grid.arrange(pca_graph, pca_graph_IQR, nrow=1)
```
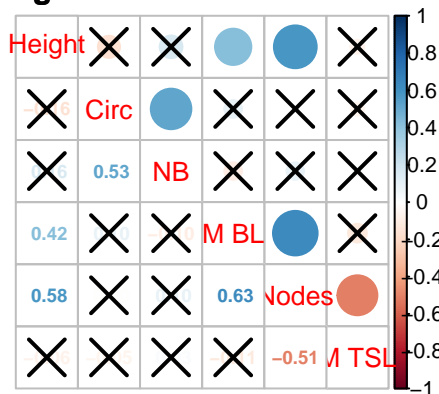
Without IQR Variables

With IQR Variables
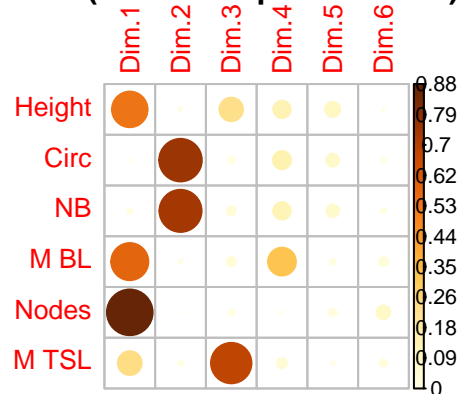


```
par(mfrow=c(2,2))
run_corrplots(dataset=d, pca_vars=var_val)
par(mfrow=c(2,2))
```
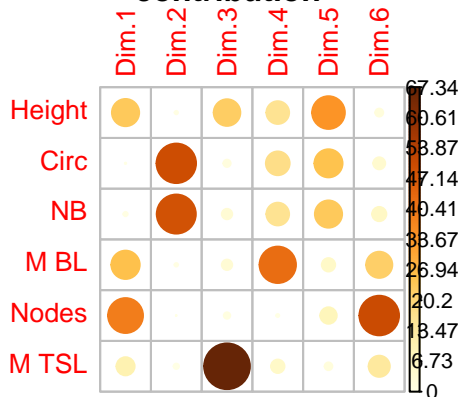
significant correlations
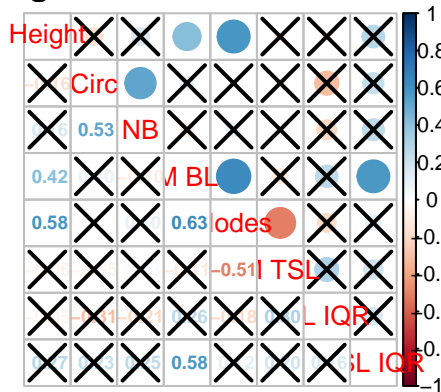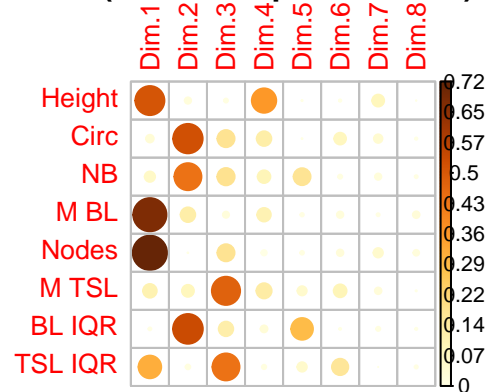

cos2 (variable representation)

contribution


```
run_corrplots(dataset=d_IQR, pca_vars=var_val_IQR)
```

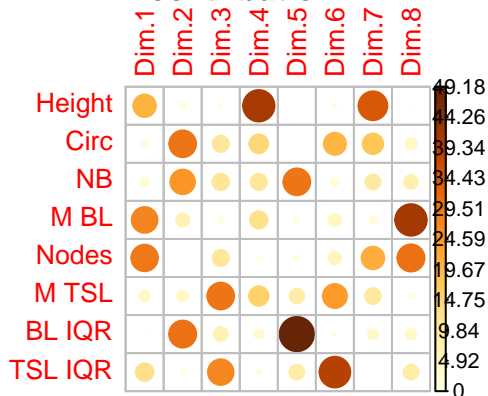**significant correlations**

Height … Circ … NB … M BL … Nodes … M TSL … L IQR … SL IQR

0.53 0.42 0.58 0.63 −0.51 0.58

(correlation scale: 1, 0.8, 0.6, 0.4, 0.2, 0, −0.2, −0.4, −0.6, −0.8, −1)

cos2 (variable representation)

Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7 Dim.8

Height, Circ, NB, M BL, Nodes, M TSL, BL IQR, TSL IQR

(scale: 0.72, 0.65, 0.57, 0.5, 0.43, 0.36, 0.29, 0.22, 0.14, 0.07, 0)

contribution

Dim.1 Dim.2 Dim.3 Dim.4 Dim.5 Dim.6 Dim.7 Dim.8

Height, Circ, NB, M BL, Nodes, M TSL, BL IQR, TSL IQR

(scale: 49.18, 44.26, 39.34, 34.43, 29.51, 24.59, 19.67, 14.75, 9.84, 4.92, 0)

This PCA suggests that there is a lot of variation around range, so much so that it apparently drives the PCA when you add it in.

2.2.1.2   Bajada Only

```
# Without IQR Vars:
d = ocos_baj[, c("Height", "Circumference",
            "Number_Branches", "Median_BranchLength",
            "Num_Nodes", "Median_TerminalSeg")]
abbreviations = c("Height", "Circ", "NB", "M BL", "Nodes", "M TSL")
colnames(d) = abbreviations

par(mfrow=c(2,2))
output_list = PCA_graphs(dataset=d, PCA_title="Without IQR Variables", run_corr=FALSE)
eig_val = output_list[[1]]
var_val_baj = output_list[[2]]
scree = output_list[[3]]
pca_graph = output_list[[5]]
pca_components_baj = output_list[[6]]


# With IQR Vars:
d_IQR = ocos_baj[, c("Height", "Circumference",
            "Number_Branches", "Median_BranchLength",
            "Num_Nodes", "Median_TerminalSeg",
            "BranchLength_IQR",
            "Terminal_SegIQR")]
abbreviations = c("Height", "Circ", "NB", "M BL", "Nodes", "M TSL","BL IQR", "TSL IQR")
```

```
colnames(d_IQR) = abbreviations

par(mfrow=c(2,2))
output_list = PCA_graphs(dataset=d_IQR, PCA_title="With IQR Variables", run_corr=FALSE)
eig_val = output_list[[1]]
var_val_IQR_baj = output_list[[2]]
scree = output_list[[3]]
pca_graph_IQR = output_list[[5]]
pca_components_IQR_baj = output_list[[6]]
```

```
grid.arrange(pca_graph, pca_graph_IQR, nrow=1)
```
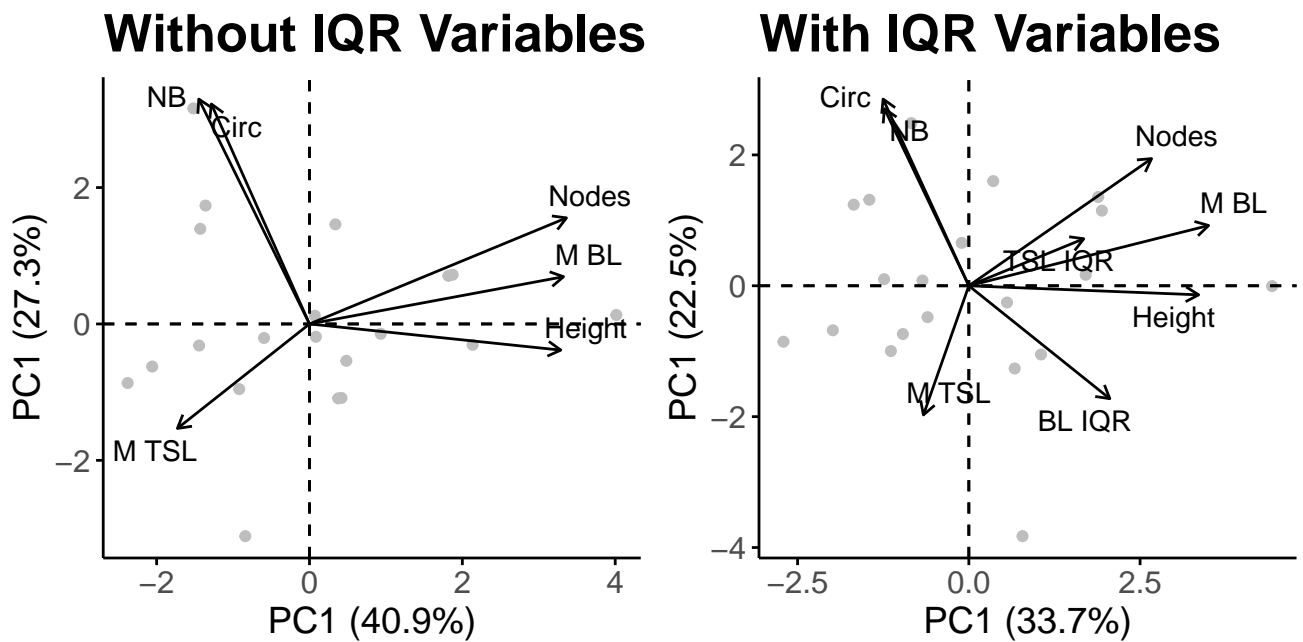


```
par(mfrow=c(2,2))
run_corrplots(dataset=d, pca_vars=var_val_baj)
par(mfrow=c(2,2))
```

significant correlations

cos2 (variable representation)

contribution


```
run_corrplots(dataset=d_IQR, pca_vars=var_val_IQR_baj)
```

significant correlations

cos2 (variable representation)

contribution

9

2.2.2 Multiple Variate Modeling of Principle Components

2.2.2.1 PC1

Bajada and Plain

```
keep_cols = c("Tree", "Elevation", "Intra_Dis", "Inter_Dis", "Inter_Plant_Group",
              "Arroyo_Dis", "X1m_Num1", "X1m_Num2", "X1m_Num3", "X1m_Num4",
              "T1m_NumShrub", "T1m_NumCacti", "R1m_shrub2cactus", "T1m_sum",
              "Latitude", "Longitude", "Elevation_c", "Intra_Dis_c",
              "Arroyo_c", "Inter_Dis_c", "Inter_plant_b", "logElevation",
              "logIntraD", "logInterD", "logArroyo", "logS1m", "Site")
preds = ocos[ , keep_cols]

reg_data = cbind(pca_components$Dim.1, pca_components$Dim.2, pca_components$Dim.3, preds)
```
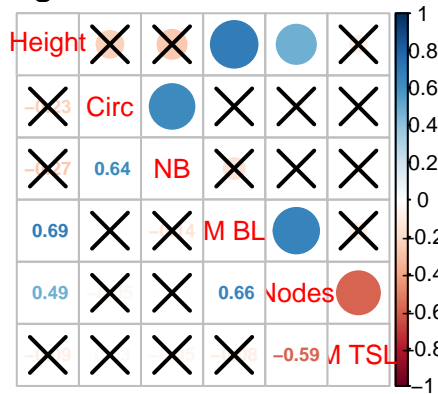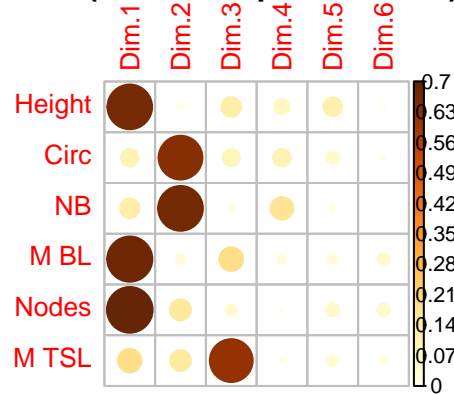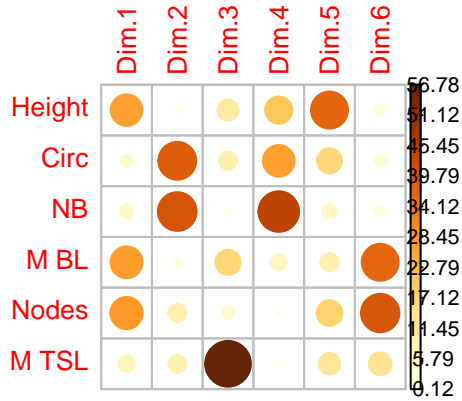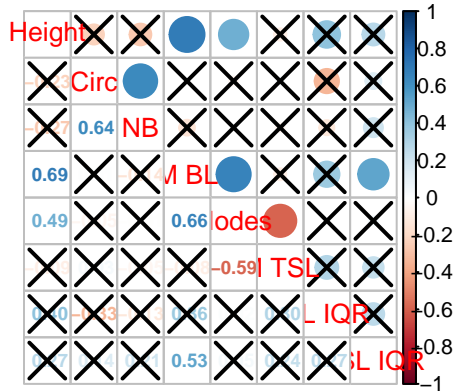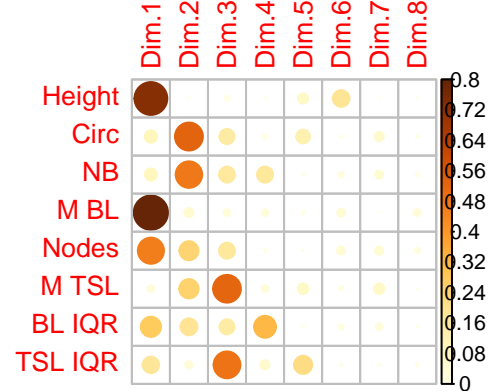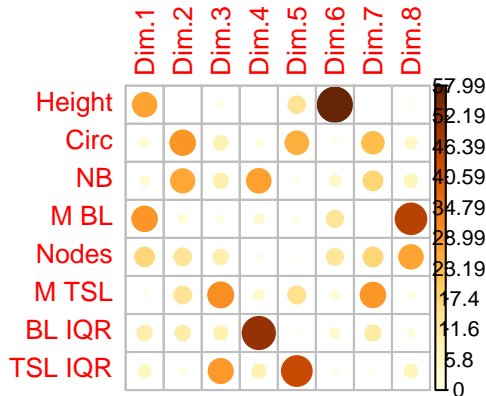
Single Variate Modeling

Only significant models are listed:

```
summary(glm(pca_components$Dim.1 ~ Elevation_c, data=reg_data))$coefficients
summary(glm(pca_components$Dim.1 ~ Site, data=reg_data))$coefficients
```

```
##                 Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) -1.692014e-16 0.26192423 -6.459938e-16 1.00000000
## Elevation_c -2.986636e-02 0.01009293 -2.959136e+00 0.00683707
##                Estimate Std. Error   t value      Pr(>|t|)
## (Intercept) -0.6049919   0.2673699 -2.262753 0.0329793578
## Site          2.2471127   0.5152881  4.360886 0.0002109863
```

Multiple Variate Modeling

```
data<-data.frame(R=pca_components$Dim.1,
                 A=reg_data$Site,
                 B=reg_data$Intra_Dis_c,
                 C=reg_data$Elevation_c,
                 D=reg_data$Arroyo_c)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lm 4-FF.R")
```

```
##         [,1]     [,2]       [,3]
## AICs   85.66475  86.79313   87.16022
## models 1        7          6
## probs  0.1276979 0.07263719 0.06045699
##
## m1   lm(formula = R ~ A, data = data)
## m7   lm(formula = R ~ A + D, data = data)
## m6   lm(formula = R ~ A + C, data = data)
```

```
anova(m0, m1, test="Chisq") # Adding A improves fit
anova(m0, m2, test="Chisq") # Adding B does not improve fit
anova(m0, m3, test="Chisq") # Adding C improves fit
anova(m0, m4, test="Chisq") # Adding D does not improve fit
anova(m1, m5, test="Chisq") # Adding B does not improve fit
```

```
anova(m1, m6, test="Chisq") # Adding C does not improve fit
anova(m1, m7, test="Chisq") # Adding D does not improve fit
```

```
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ A
##   Res.Df    RSS Df Sum of Sq  Pr(>Chi)
## 1     25 58.428
## 2     24 32.598  1     25.83 1.295e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ B
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     25 58.428
## 2     24 52.354  1    6.0746  0.09517 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ C
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     25 58.428
## 2     24 42.809  1    15.619 0.003085 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     25 58.428
## 2     24 55.302  1    3.1264   0.2441
## Analysis of Variance Table
##
## Model 1: R ~ A
## Model 2: R ~ A + B
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     24 32.598
## 2     23 32.581  1  0.016454   0.9142
## Analysis of Variance Table
##
## Model 1: R ~ A
## Model 2: R ~ A + C
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     24 32.598
## 2     23 31.971  1   0.62647    0.502
## Analysis of Variance Table
##
```

```
## Model 1: R ~ A
## Model 2: R ~ A + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     24 32.598
## 2     23 31.523  1    1.0747   0.3759
```

Best Fit

```
M1 = glm(pca_components$Dim.1 ~ Site, data=reg_data)
summary(M1)
```

```
##
## Call:
## glm(formula = pca_components$Dim.1 ~ Site, data = reg_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.00919  -0.72408   0.09496   0.86656   1.92332
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.6050     0.2674  -2.263 0.032979 *
## Site          2.2471     0.5153   4.361 0.000211 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.358246)
##
##     Null deviance: 58.428  on 25  degrees of freedom
## Residual deviance: 32.598  on 24  degrees of freedom
## AIC: 85.665
##
## Number of Fisher Scoring iterations: 2
```

If on the bajada then smaller.

Bajada Only

```
preds = ocos_baj[ , keep_cols]
reg_data_baj = cbind(pca_components_baj$Dim.1, pca_components_baj$Dim.2, pca_components_baj$Di
```

Single Variate Modeling

Only significant models are listed:

```
summary(glm(pca_components_baj$Dim.1 ~ Site, data=reg_data_baj))$coefficients
```

```
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -0.211288  0.3066012 -0.6891296 0.499529345
## Site         4.225760  1.3711625  3.0818813 0.006429015
```

Multiple Variate Modeling

```
data<-data.frame(R=pca_components_baj$Dim.1,
                 A=reg_data_baj$Site,
                 B=reg_data_baj$Intra_Dis_c,
```

```
                C=reg_data_baj$Elevation_c,
                D=reg_data_baj$Arroyo_c)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lm 4-FF.R")
```

```
##          [,1]
## AICs   72.25082
## models 1
## probs  0.05334827
##
## m1    lm(formula = R ~ A, data = data)
```

```
anova(m0, m1, test="Chisq") # Adding A improves fit
anova(m0, m2, test="Chisq") # Adding B does not improve fit
anova(m0, m3, test="Chisq") # Adding C does not improve fit
anova(m0, m4, test="Chisq") # Adding D does not improve fit
```

```
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ A
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     19 49.114
## 2     18 32.149  1    16.964 0.002057 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ B
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     19 49.114
## 2     18 46.912  1    2.2014   0.3581
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ C
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     19 49.114
## 2     18 47.797  1    1.3164   0.4814
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     19 49.114
## 2     18 44.461  1    4.6527   0.1699
```

Best Fit

```
M2 = glm(pca_components_baj$Dim.1 ~ Site, data=reg_data_baj)
summary(M2)
```

```
##
## Call:
## glm(formula = pca_components_baj$Dim.1 ~ Site, data = reg_data_baj)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1678  -1.1655   0.1403   0.6453   2.3460
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2113     0.3066  -0.689  0.49953
## Site          4.2258     1.3712   3.082  0.00643 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.786082)
##
##     Null deviance: 49.114  on 19  degrees of freedom
## Residual deviance: 32.149  on 18  degrees of freedom
## AIC: 72.251
##
## Number of Fisher Scoring iterations: 2
```

2.2.2.2   PC2

Bajada and Plain

Single Variate Modeling

No significant single variate models.

Multiple Variate Modeling

```
data<-data.frame(R=pca_components$Dim.2,
                 A=reg_data$T1m_NumShrub,# X1m_Num2, X1m_Num4 matter too
                 B=reg_data$Inter_Dis_c, # Intra_Dis_c matter too
                 C=reg_data$Elevation_c,
                 D=reg_data$Arroyo_c)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lm 4-FF.R")
```

```
##            [,1]
## AICs   86.73666
## models 7
## probs  0.05986922
##
## m7    lm(formula = R ~ A + D, data = data)
```

14

```
anova(m42, m57, test="Chisq") # Adding D does not improve fit
anova(m42, m28, test="Chisq") # Adding A*B improves fit
anova(m42, m22, test="Chisq") # Adding B*C improves fit
```

```
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * C
## Model 2: R ~ A * B + B * C + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     20 34.892
## 2     19 28.908  1    5.9844  0.04734 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * C
## Model 2: R ~ B * C + A
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     20 34.892
## 2     21 35.561 -1  -0.66862   0.5359
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * C
## Model 2: R ~ A * B + C
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     20 34.892
## 2     21 34.902 -1 -0.010093   0.9394
```

```
M3 = glm(pca_components$Dim.2 ~ T1m_NumShrub * Intra_Dis_c + Intra_Dis_c * Elevation_c, data=r
summary(M3)
```

```
##
## Call:
## glm(formula = pca_components$Dim.2 ~ T1m_NumShrub * Intra_Dis_c +
##     Intra_Dis_c * Elevation_c, data = reg_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.66047  -0.51491   0.05215   0.49819   1.98650
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              0.054140   0.344269   0.157   0.8766
## T1m_NumShrub             0.191080   0.108296   1.764   0.0929 .
## Intra_Dis_c              0.049698   0.081636   0.609   0.5495
## Elevation_c              0.012745   0.011020   1.157   0.2611
## T1m_NumShrub:Intra_Dis_c 0.080901   0.034283   2.360   0.0286 *
## Intra_Dis_c:Elevation_c  0.005019   0.001909   2.630   0.0161 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.129703)
##
##     Null deviance: 39.621  on 25  degrees of freedom
```

```
## Residual deviance: 22.594  on 20  degrees of freedom
## AIC: 84.134
##
## Number of Fisher Scoring iterations: 2
# M3 = glm(pca_components$Dim.2 ~ T1m_NumShrub + Arroyo_c, data=reg_data)
# summary(M3)
```

Bajada Only

Single Variate Modeling

No significant single variate models.

Multiple Variate Modeling

```
data<-data.frame(R=pca_components_baj$Dim.2,
                 A=reg_data_baj$T1m_NumShrub,# X1m_Num2 matter too but more as an interaction
                 B=reg_data_baj$Inter_Dis_c, # Inter_plant_b matter too
                 C=reg_data_baj$Intra_Dis_c,
                 D=reg_data_baj$Arroyo_c)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lm 4-FF.R")
```

```
##           [,1]      [,2]      [,3]
## AICs   64.39599  65.40301  65.86961
## models 104       109       111
## probs  0.1750499 0.1058011 0.08378549
##
## m104     lm(formula = R ~ A * C + B * C + B * D + C * D, data = data)
## m109     lm(formula = R ~ A * B + A * C + B * C + B * D + C * D, data = data)
## m111     lm(formula = R ~ A * C + A * D + B * C + B * D + C * D, data = data)
```

```
anova(m104, m80, test="Chisq") # Adding C*D marginally improves fit
anova(m80, m60, test="Chisq") # Adding B*D improves fit
anova(m80, m47, test="Chisq") # Adding B*C does not improve fit
anova(m47, m24, test="Chisq") # Adding B*D improves fit
anova(m47, m25, test="Chisq") # Adding B*D improves fit
anova(m47, m35, test="Chisq") # Adding B*D improves fit
anova(m47, m38, test="Chisq") # Adding A*C improves fit
```

```
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * C + B * D + C * D
## Model 2: R ~ A * C + B * C + B * D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     11 10.780
## 2     12 14.377 -1   -3.5972  0.05537 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * C + B * D
```

```
## Model 2: R ~ A * C + B * C + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     12 14.377
## 2     13 21.835 -1   -7.4586  0.01259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * C + B * D
## Model 2: R ~ A * C + B * D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     12 14.377
## 2     13 15.129 -1  -0.75264    0.428
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * D
## Model 2: R ~ A * C + B
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     13 15.129
## 2     15 22.261 -2    -7.131  0.04672 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * D
## Model 2: R ~ A * C + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     13 15.129
## 2     15 22.105 -2   -6.9756  0.04994 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * D
## Model 2: R ~ A * C + B + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     13 15.129
## 2     14 21.984 -1   -6.8548  0.01523 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * C + B * D
## Model 2: R ~ B * D + A + C
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     13 15.129
## 2     14 22.908 -1   -7.7788 0.009728 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Best Fit

```
# M4 = glm(pca_components_baj$Dim.2 ~  T1m_NumShrub * Arroyo_c + Inter_Dis_c * Arroyo_c, data=
# summary(M4)

M4 = glm(pca_components_baj$Dim.2 ~ T1m_NumShrub * Intra_Dis_c + Inter_Dis_c * Arroyo_c, data=
summary(M4)
```

```
##
## Call:
## glm(formula = pca_components_baj$Dim.2 ~ T1m_NumShrub * Intra_Dis_c +
##      Inter_Dis_c * Arroyo_c, data = reg_data_baj)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7729  -0.7294   0.1978   0.4786   1.4994
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.401173   0.356519  -1.125   0.2808
## T1m_NumShrub              0.278359   0.124042   2.244   0.0429 *
## Intra_Dis_c              -0.022451   0.076819  -0.292   0.7747
## Inter_Dis_c               0.552194   0.313607   1.761   0.1018
## Arroyo_c                 -0.005005   0.005058  -0.989   0.3405
## T1m_NumShrub:Intra_Dis_c  0.110843   0.042874   2.585   0.0226 *
## Inter_Dis_c:Arroyo_c     -0.011859   0.004887  -2.427   0.0305 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.163804)
##
##     Null deviance: 32.819  on 19  degrees of freedom
## Residual deviance: 15.129  on 13  degrees of freedom
## AIC: 67.176
##
## Number of Fisher Scoring iterations: 2
```

2.2.2.3 PC3

Bajada and Plain

Single Variate Modeling

No significant and marginally significant single variate models.

```
summary(glm(pca_components$Dim.3 ~ Inter_plant_b, data=reg_data))$coefficients
```

```
##                Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)   0.4631920  0.2921763   1.585317 0.12598379
## Inter_plant_b -0.8028661  0.3846679  -2.087167 0.04766647
```

```
summary(glm(pca_components$Dim.3 ~ Inter_Dis_c, data=reg_data))$coefficients
```

```
##                  Estimate Std. Error       t value   Pr(>|t|)
## (Intercept) -9.102906e-16  0.1899667 -4.791843e-15 1.00000000
## Inter_Dis_c  2.740347e-01  0.1309476  2.092704e+00 0.04712566
```

```
summary(glm(pca_components$Dim.3 ~ Arroyo_c, data=reg_data))$coefficients # marginally signifi
```

```
##                    Estimate  Std. Error       t value   Pr(>|t|)
## (Intercept) -1.036526e-15 0.192207834 -5.392736e-15 1.00000000
## Arroyo_c    -4.694600e-03 0.002433559 -1.929108e+00 0.06562374
```

Multiple Variate Modeling

```
data<-data.frame(R=pca_components$Dim.3,
                 A=reg_data$Inter_plant_b,
                 B=reg_data$Inter_Dis_c,
                 C=reg_data$Elevation_c,
                 D=reg_data$T1m_NumShrub) # T1m_NumCacti matter too

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lm 4-FF.R")
```

```
##            [,1]      [,2]      [,3]      [,4]       [,5]       [,6]
## AICs   69.80607  70.52897  71.43361  72.32699   72.36142   72.41337
## models 112       76        100       109        99         96
## probs  0.2249579 0.1567204 0.0996979 0.06378097 0.06269225 0.06108482
##            [,7]
## AICs   72.6049
## models 110
## probs  0.05550633
##
## m112      lm(formula = R ~ A * B + A * C + A * D + B * C + B * D + C *
##     D, data = data)
## m76  lm(formula = R ~ A * B + B * D + C * D, data = data)
## m100      lm(formula = R ~ A * B + B * C + B * D + C * D, data = data)
## m109      lm(formula = R ~ A * B + A * C + B * C + B * D + C * D, data = data)
## m99  lm(formula = R ~ A * B + A * D + B * D + C * D, data = data)
## m96  lm(formula = R ~ A * B + A * C + B * D + C * D, data = data)
## m110      lm(formula = R ~ A * B + A * D + B * C + B * D + C * D, data = data)
```

```
anova(m112, m106, test="Chisq") # Adding C*D improves fit
anova(m112, m107, test="Chisq") # Adding B*D improves fit
anova(m112, m108, test="Chisq") # Adding B*C improves fit
anova(m112, m109, test="Chisq") # Adding A*D marginally improves fit
anova(m112, m110, test="Chisq") # Adding A*C marginally improves fit
anova(m112, m111, test="Chisq") # Adding A*B improves fit

anova(m110, m100, test="Chisq") # Adding A*D does not improve fit
anova(m100, m76, test="Chisq") # Adding B*C does not improve fit
anova(m76, m44, test="Chisq") # Adding B*C improves fit
anova(m76, m66, test="Chisq")  # ADding A*B improves fit
anova(m76, m58, test="Chisq")  # ADding C*D improves fit
```

```
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * C + A * D + B * C + B * D + C * D
## Model 2: R ~ A * B + A * C + A * D + B * C + B * D
##   Res.Df    RSS Df Sum of Sq  Pr(>Chi)
## 1      15  8.8639
```

```
## 2      16 15.8385 -1   -6.9746 0.0005914 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * C + A * D + B * C + B * D + C * D
## Model 2: R ~ A * B + A * C + A * D + B * C + C * D
##   Res.Df     RSS Df Sum of Sq Pr(>Chi)
## 1      15  8.8639
## 2      16 13.0394 -1   -4.1754 0.007857 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * C + A * D + B * C + B * D + C * D
## Model 2: R ~ A * B + A * C + A * D + B * D + C * D
##   Res.Df     RSS Df Sum of Sq Pr(>Chi)
## 1      15  8.8639
## 2      16 11.3041 -1   -2.4401  0.04215 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * C + A * D + B * C + B * D + C * D
## Model 2: R ~ A * B + A * C + B * C + B * D + C * D
##   Res.Df     RSS Df Sum of Sq Pr(>Chi)
## 1      15  8.8639
## 2      16 10.5473 -1   -1.6834  0.09145 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * C + A * D + B * C + B * D + C * D
## Model 2: R ~ A * B + A * D + B * C + B * D + C * D
##   Res.Df     RSS Df Sum of Sq Pr(>Chi)
## 1      15  8.8639
## 2      16 10.6607 -1   -1.7967  0.08121 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * C + A * D + B * C + B * D + C * D
## Model 2: R ~ A * C + A * D + B * C + B * D + C * D
##   Res.Df     RSS Df Sum of Sq Pr(>Chi)
## 1      15  8.8639
## 2      16 14.7587 -1   -5.8948 0.001586 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + A * D + B * C + B * D + C * D
## Model 2: R ~ A * B + B * C + B * D + C * D
##   Res.Df     RSS Df Sum of Sq Pr(>Chi)
```

```
## 1      16 10.661
## 2      17 11.006 -1  -0.34526   0.4716
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * C + B * D + C * D
## Model 2: R ~ A * B + B * D + C * D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1      17 11.006
## 2      18 11.479 -1  -0.47358   0.3924
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * D + C * D
## Model 2: R ~ A * B + C * D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1      18 11.479
## 2      19 15.907 -1   -4.4273 0.008419 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * D + C * D
## Model 2: R ~ B * D + C * D + A
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1      18 11.479
## 2      19 16.916 -1   -5.4368 0.003503 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ A * B + B * D + C * D
## Model 2: R ~ A * B + B * D + C
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1      18 11.479
## 2      19 17.886 -1   -6.4066 0.001527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Best Fit

```
M5 = glm(pca_components$Dim.3 ~ Inter_plant_b * Inter_Dis_c + Inter_Dis_c * T1m_NumShrub + Ele
summary(M5)

##
## Call:
## glm(formula = pca_components$Dim.3 ~ Inter_plant_b * Inter_Dis_c +
##     Inter_Dis_c * T1m_NumShrub + Elevation_c * T1m_NumShrub,
##     data = reg_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1800  -0.3572  -0.0571   0.3261   1.5403
##
## Coefficients:
```

```
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.105242   0.351967   0.299  0.76836
## Inter_plant_b            -0.784347   0.433924  -1.808  0.08742 .
## Inter_Dis_c               0.032020   0.244826   0.131  0.89739
## T1m_NumShrub             -0.013503   0.091318  -0.148  0.88409
## Elevation_c              -0.031995   0.012725  -2.514  0.02165 *
## Inter_plant_b:Inter_Dis_c -0.846730   0.290000  -2.920  0.00915 **
## Inter_Dis_c:T1m_NumShrub  0.208889   0.079281   2.635  0.01682 *
## T1m_NumShrub:Elevation_c  0.015933   0.005027   3.169  0.00531 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.637751)
##
##      Null deviance: 26.628  on 25  degrees of freedom
## Residual deviance: 11.480  on 18  degrees of freedom
## AIC: 70.529
##
## Number of Fisher Scoring iterations: 2
```

Bajada Only

Single Variate Modeling

Only significant and marginally significant single variate models listed:

```
summary(glm(pca_components_baj$Dim.3 ~ T1m_NumCacti, data=reg_data_baj))$coefficients
```

```
##                 Estimate Std. Error    t value    Pr(>|t|)
## (Intercept)    0.1850910  0.2324751  0.7961757  0.43630521
## T1m_NumCacti -0.5288315  0.2385144 -2.2171887  0.03972244
```

Multiple Variate Modeling

```
data<-data.frame(R=pca_components_baj$Dim.3,
                 A=reg_data_baj$Inter_plant_b,
                 B=reg_data_baj$Inter_Dis_c,
                 C=reg_data_baj$T1m_NumShrub,
                 D=reg_data_baj$T1m_NumCacti)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lm 4-FF.R")
```

```
##          [,1]      [,2]        [,3]        [,4]
## AICs   59.44503  59.87995    59.89259    60.17994
## models 4         10          7           13
## probs  0.0865392 0.06962608  0.06918757  0.05992808
##
## m4    lm(formula = R ~ D, data = data)
## m10   lm(formula = R ~ C + D, data = data)
## m7    lm(formula = R ~ A + D, data = data)
## m13   lm(formula = R ~ A + C + D, data = data)
```

```r
anova(m0, m4, test="Chisq") # Adding D improves fit
anova(m4, m10, test="Chisq") # Adding C does not improve fit
anova(m4, m7, test="Chisq") # Adding A does not improve fit
```

```
## Analysis of Variance Table
##
## Model 1: R ~ 1
## Model 2: R ~ D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     19 21.576
## 2     18 16.947  1   4.6284  0.02661 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Analysis of Variance Table
##
## Model 1: R ~ D
## Model 2: R ~ C + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     18 16.947
## 2     17 15.672  1   1.2756   0.2395
## Analysis of Variance Table
##
## Model 1: R ~ D
## Model 2: R ~ A + D
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     18 16.947
## 2     17 15.681  1   1.2657   0.2414
```

Best Fit

```r
M6 = glm(pca_components_baj$Dim.3 ~ T1m_NumCacti, data=reg_data_baj)
summary(M6)
```

```
##
## Call:
## glm(formula = pca_components_baj$Dim.3 ~ T1m_NumCacti, data = reg_data_baj)
##
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7253 -0.6503  0.3021  0.8552  1.1252
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1851     0.2325   0.796   0.4363
## T1m_NumCacti  -0.5288     0.2385  -2.217   0.0397 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9415151)
##
##     Null deviance: 21.576  on 19  degrees of freedom
## Residual deviance: 16.947  on 18  degrees of freedom
## AIC: 59.445
```

```
##
## Number of Fisher Scoring iterations: 2
```

2.3   Mixed Effect, Multiple Variate Modeling

2.3.1   Terminal Segment Lengths

2.3.1.1   Plain and Bajada

Model Comparisons

```
data<-data.frame(R=segs$Length_c,
                 A=segs$T1m_NumCacti,
                 B=segs$Num_Nodes_c,
                 C=segs$BranchLength_IQR_c,
                 X=segs$Tree,
                 Y=segs$seg_num)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lmer 2-RF + 3-FF REMLF.R")
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

##           [,1]      [,2]      [,3]      [,4]      [,5]
## AICs   8257.873  8259.752  8259.872  8259.967  8261.751
## models 11        14        15        8         17
## probs  0.4110724 0.1605991 0.1512525 0.1442369 0.0591095
##
## m11  R ~ A * B + C + (1 | X) + (1 | Y)
## m14  R ~ A * B + A * C + (1 | X) + (1 | Y)
## m15  R ~ A * B + B * C + (1 | X) + (1 | Y)
## m8   R ~ A * B + (1 | X) + (1 | Y)
## m17  R ~ A * B + A * C + B * C + (1 | X) + (1 | Y)
```

```
anova(m17, m15, test="Chisq") # Adding A*C does not improve fit
anova(m15, m13, test="Chisq") # Adding A*B improves fit
anova(m15, m11, test="Chisq") # Adding B*C does not improve fit
# anova(m15, m11, test="Chisq") # Adding B*C improves fit
anova(m11, m7, test="Chisq") # Adding A*B does improve fit
anova(m11, m4, test="Chisq") # Adding C improves fit
```

```
## Data: data
```

```
## Models:
## m15: R ~ A * B + B * C + (1 | X) + (1 | Y)
## m17: R ~ A * B + A * C + B * C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m15    9 8259.9 8306.4 -4120.9   8241.9
## m17   10 8261.8 8313.5 -4120.9   8241.8 0.1209  1     0.7281
## Data: data
## Models:
## m13: R ~ B * C + A + (1 | X) + (1 | Y)
## m15: R ~ A * B + B * C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m13    8 8265.8 8307.2 -4124.9   8249.8
## m15    9 8259.9 8306.4 -4120.9   8241.9 7.9518  1   0.004804 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Data: data
## Models:
## m11: R ~ A * B + C + (1 | X) + (1 | Y)
## m15: R ~ A * B + B * C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## m11    8 8257.9 8299.2 -4120.9   8241.9
## m15    9 8259.9 8306.4 -4120.9   8241.9 4e-04  1     0.9848
## Data: data
## Models:
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
## m11: R ~ A * B + C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m7     7 8266.0 8302.2 -4126.0   8252.0
## m11    8 8257.9 8299.2 -4120.9   8241.9 10.174  1   0.001424 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Data: data
## Models:
## m4: R ~ A + B + (1 | X) + (1 | Y)
## m11: R ~ A * B + C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m4     6 8265.8 8296.9 -4126.9   8253.8
## m11    8 8257.9 8299.2 -4120.9   8241.9 11.969  2   0.002518 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Best Fit

```
# M11 = lmer(Length_c ~ T1m_NumCacti * Num_Nodes_c + Num_Nodes_c * BranchLength_IQR_c +
#                                        (1 | Tree) + (1 | seg_num), data=segs)
# summary(M11)
M11 = lmer(Length_c ~ T1m_NumCacti * Num_Nodes_c + Num_Nodes_c + BranchLength_IQR_c +
                                        (1 | Tree) + (1 | seg_num), data=segs)
summary(M11)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Length_c ~ T1m_NumCacti * Num_Nodes_c + Num_Nodes_c + BranchLength_IQR_c +
```

```
##     (1 | Tree) + (1 | seg_num)
##    Data: segs
##
## REML criterion at convergence: 8261.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0056 -0.6311 -0.0657  0.5834  4.7451
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Tree     (Intercept)  4.8741  2.2077
##  seg_num  (Intercept)  0.5249  0.7245
##  Residual             31.7408  5.6339
## Number of obs: 1300, groups:  Tree, 26; seg_num, 5
##
## Fixed effects:
##                        Estimate Std. Error t value
## (Intercept)            -0.11455    0.62093  -0.184
## T1m_NumCacti           -0.90637    0.55288  -1.639
## Num_Nodes_c            -0.06100    0.01423  -4.286
## BranchLength_IQR_c      0.03217    0.01688   1.906
## T1m_NumCacti:Num_Nodes_c  0.04746    0.01486   3.194
##
## Correlation of Fixed Effects:
##            (Intr) T1m_NC Nm_Nd_ BL_IQR
## T1m_NumCact -0.334
## Num_Nodes_c  0.170 -0.069
## BrnchL_IQR_ -0.098  0.153  0.030
## T1m_NC:N_N_ -0.152 -0.285 -0.528  0.131
```

Likelihood and Weights

```r
# summary_tableSL = calculate_lk_weights(model_list=list(m0, m1, m2, m3, m4, m5, m6, m7, m11,
#                                  best_fit=m15,
#                                  R="Length_c", A="T1m_NumCacti", B="Num_Nodes_c", C="Br
#                                  X="Tree", Y="seg_num", is_lm=FALSE)
summary_tableSL = calculate_lk_weights(model_list=list(m0, m4, m7, m11),
                                 best_fit=m11,
                                 R="Length_c", A="T1m_NumCacti", B="Num_Nodes_c", C="Bran
                                 X="Tree", Y="seg_num", is_lm=FALSE)

summary_tableSL$Equation = gsub("T1m_NumBranchLength_IQLength_c_cacti", "T1m_NumCacti",
                          summary_tableSL$Equation )
summary_tableSL$Equation = gsub("BranchLength_IQLength_c_c", "BranchLength_IQR_c",
                          summary_tableSL$Equation)

summary_tableSL
```

```
##                                                                       Equation
## 1                                       Length_c ~ (1 | Tree) + (1 | seg_num)
## 2               Length_c ~ T1m_NumCacti + Num_Nodes_c + (1 | Tree) + (1 | seg_num)
## 3 Length_c ~ T1m_NumCacti + Num_Nodes_c + BranchLength_IQR_c + (1 | Tree) + (1 | seg_num)
```

```
## 4 Length_c ~ T1m_NumCacti * Num_Nodes_c + BranchLength_IQR_c + (1 | Tree) + (1 | seg_num)
##          AIC      dAIC Likelihood  Weight
## 1 8271.553 13.68091    0.00107 0.00103
## 2 8265.841  7.96895    0.01860 0.01795
## 3 8266.047  8.17404    0.01679 0.01620
## 4 8257.873  0.00000    1.00000 0.96482
```

2.3.1.2 Bajada Only

Model Comparisons

```
data<-data.frame(R=segs_baj$Length_c,
                 A=segs_baj$T1m_NumCacti,
                 B=segs_baj$Num_Nodes_c,
                 C=segs_baj$BranchLength_IQR_c,
                 X=segs_baj$Tree,
                 Y=segs_baj$seg_num)

source("src/compare_models.R")
model_comparisonsAIC("src/generic models-gaussian lmer 2-RF + 3-FF REMLF.R")
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling

##          [,1]       [,2]       [,3]       [,4]       [,5]       [,6]
## AICs   6306.037  6306.425   6307.794   6307.886   6308.032   6308.255
## models 7         11         13         6          12         4
## probs  0.2104824 0.1734001 0.08743109 0.08351391 0.07761663 0.06945919
##          [,7]       [,8]
## AICs   6308.273   6308.339
## models 14         15
## probs  0.06883631 0.06658537
##
## m7   R ~ A + B + C + (1 | X) + (1 | Y)
## m11  R ~ A * B + C + (1 | X) + (1 | Y)
## m13  R ~ B * C + A + (1 | X) + (1 | Y)
## m6   R ~ B + C + (1 | X) + (1 | Y)
## m12  R ~ A * C + B + (1 | X) + (1 | Y)
## m4   R ~ A + B + (1 | X) + (1 | Y)
## m14  R ~ A * B + A * C + (1 | X) + (1 | Y)
## m15  R ~ A * B + B * C + (1 | X) + (1 | Y)
```

```
# anova(m15, m13, test="Chisq") # Adding A*B does not improve fit
# anova(m13, m10, test="Chisq") # Adding A marginally improves fit
# anova(m10, m6, test="Chisq") # Adding B*C improves fit
```

```
anova(m11, m7, test="Chisq") # Adding A*B does not improve fit
anova(m12, m7, test="Chisq") # Adding A*C does not improve fit
anova(m13, m7, test="Chisq") # Adding B*C does not improve fit
anova(m7, m6, test="Chisq") # Adding A improves fit
anova(m7, m5, test="Chisq") # Adding B improves fit
anova(m7, m4, test="Chisq") # Adding C improves fit
```

```
## Data: data
## Models:
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
## m11: R ~ A * B + C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m7     7 6306.0 6340.4 -3146.0   6292.0
## m11    8 6306.4 6345.7 -3145.2   6290.4 1.6124  1     0.2042
## Data: data
## Models:
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
## m12: R ~ A * C + B + (1 | X) + (1 | Y)
##     npar  AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## m7     7 6306 6340.4  -3146     6292
## m12    8 6308 6347.3  -3146     6292 0.0048  1      0.945
## Data: data
## Models:
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
## m13: R ~ B * C + A + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m7     7 6306.0 6340.4 -3146.0   6292.0
## m13    8 6307.8 6347.1 -3145.9   6291.8 0.2429  1     0.6221
## Data: data
## Models:
## m6: R ~ B + C + (1 | X) + (1 | Y)
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
##    npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m6    6 6307.9 6337.3 -3147.9   6295.9
## m7    7 6306.0 6340.4 -3146.0   6292.0 3.8488  1    0.04978 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Data: data
## Models:
## m5: R ~ A + C + (1 | X) + (1 | Y)
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
##    npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m5    6 6317.3 6346.7 -3152.7   6305.3
## m7    7 6306.0 6340.4 -3146.0   6292.0 13.255  1  0.0002718 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Data: data
## Models:
## m4: R ~ A + B + (1 | X) + (1 | Y)
```

```
## m7: R ~ A + B + C + (1 | X) + (1 | Y)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m4     6 6308.3 6337.7 -3148.1   6296.3
## m7     7 6306.0 6340.4 -3146.0   6292.0 4.2173  1    0.04001 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Best Fit

```
# M12 = lmer(Length_c ~ Num_Nodes_c * BranchLength_IQR_c + (1 | Tree) + (1 | seg_num), data=se
# summary(M12)
M12 = lmer(Length_c ~ T1m_NumCacti + Num_Nodes_c + BranchLength_IQR_c + (1 | Tree) + (1 | seg_
summary(M12)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Length_c ~ T1m_NumCacti + Num_Nodes_c + BranchLength_IQR_c +
##     (1 | Tree) + (1 | seg_num)
##    Data: segs_baj
##
## REML criterion at convergence: 6303.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0267 -0.6410 -0.0702  0.6018  4.8337
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Tree     (Intercept)  4.7156  2.1715
##  seg_num  (Intercept)  0.5403  0.7351
##  Residual             30.2246  5.4977
## Number of obs: 1000, groups:  Tree, 20; seg_num, 5
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)        -0.38354    0.68256  -0.562
## T1m_NumCacti       -1.07047    0.57634  -1.857
## Num_Nodes_c        -0.05820    0.01488  -3.912
## BranchLength_IQR_c  0.03532    0.01808   1.954
##
## Correlation of Fixed Effects:
##             (Intr) T1m_NC Nm_Nd_
## T1m_NumCact -0.336
## Num_Nodes_c  0.271 -0.061
## BrnchL_IQR_ -0.205  0.171 -0.027
```

Likelihood and Weights

```
# summary_tableSL2 = calculate_lk_weights(model_list=list(m0, m2, m3, m6, m10), best_fit=m10,
#                                 R="Length_c", A="T1m_NumCacti", B="Num_Nodes_c", C="B
#                                 X="Tree", Y="seg_num", is_lm=FALSE)

summary_tableSL2 = calculate_lk_weights(model_list=list(m0, m1, m2, m3, m4, m5, m7), best_fit=
                                 R="Length_c", A="T1m_NumCacti", C="Bra
                                 X="Tree", Y="seg_num", is_lm=FALSE)
```

```
summary_tableSL2$Equation = gsub("BranchLength_IQLength_c_c", "BranchLength_IQR_c",
                                 summary_tableSL2$Equation)
summary_tableSL2
```

```
##
## 1                                                                Length_c ~ (1 | Tree)
## 2                                               Length_c ~ T1m_NumCacti + (1 | Tree)
## 3                                               Length_c ~ Num_Nodes_c + (1 | Tree)
## 4                                         Length_c ~ BranchLength_IQR_c + (1 | Tree)
## 5                                 Length_c ~ T1m_NumCacti + Num_Nodes_c + (1 | Tree)
## 6            Length_c ~ T1m_NumBranchLength_IQR_cacti + BranchLength_IQR_c + (1 | Tree)
## 7 Length_c ~ T1m_NumBranchLength_IQR_cacti + Num_Nodes_c + BranchLength_IQR_c + (1 | Tree)
##        AIC      dAIC Likelihood  Weight
## 1 6318.499 12.46216    0.00197 0.00137
## 2 6317.343 11.30610    0.00351 0.00243
## 3 6310.653  4.61565    0.09948 0.06902
## 4 6317.902 11.86483    0.00265 0.00184
## 5 6308.255  2.21733    0.33000 0.22898
## 6 6317.292 11.25512    0.00360 0.00250
## 7 6306.037  0.00000    1.00000 0.69387
```

2.3.1.3 Summary: Number of Nodes, Branch Length IQR, and Number of Cacti Affect Terminal Segment Lengths