Spatial Models

Morgan Will

2/22/2022

Spatial Modeling Assignment

```r
library(vegan)

library(nlme)
data(BCI)
## UTM Coordinates (in metres)
BCI_xy = data.frame(x = rep(seq(625754, 626654, by=100), each=5),
                    y = rep(seq(1011569,  1011969, by=100), len=50))
```
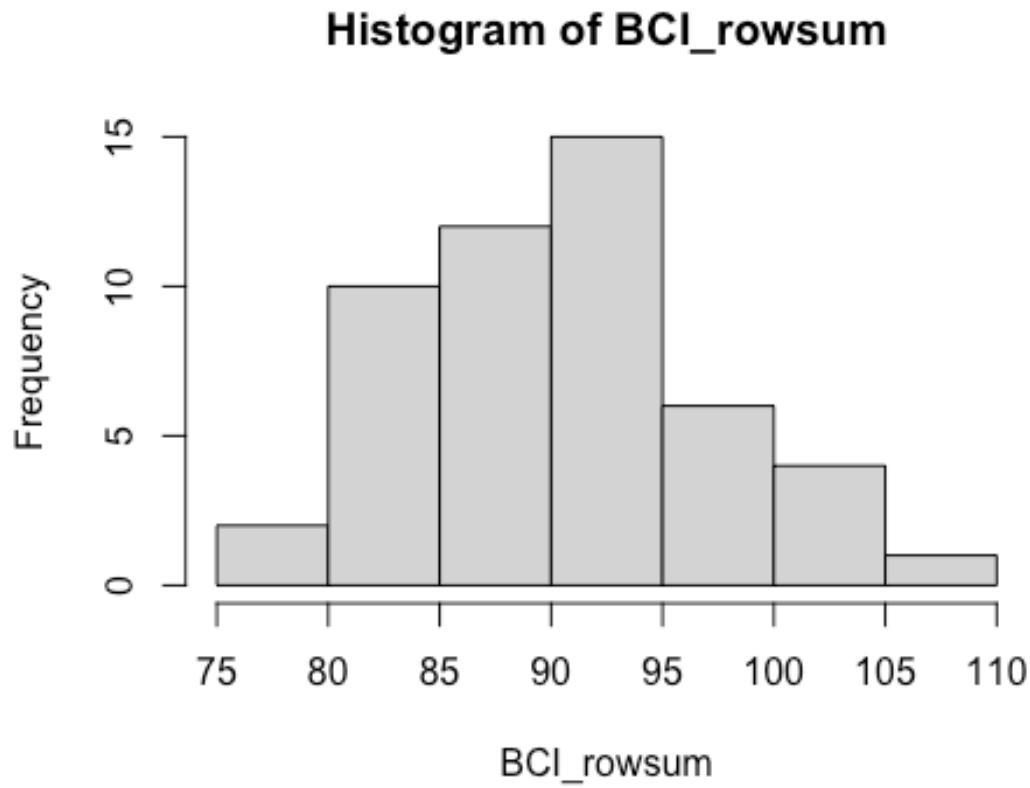
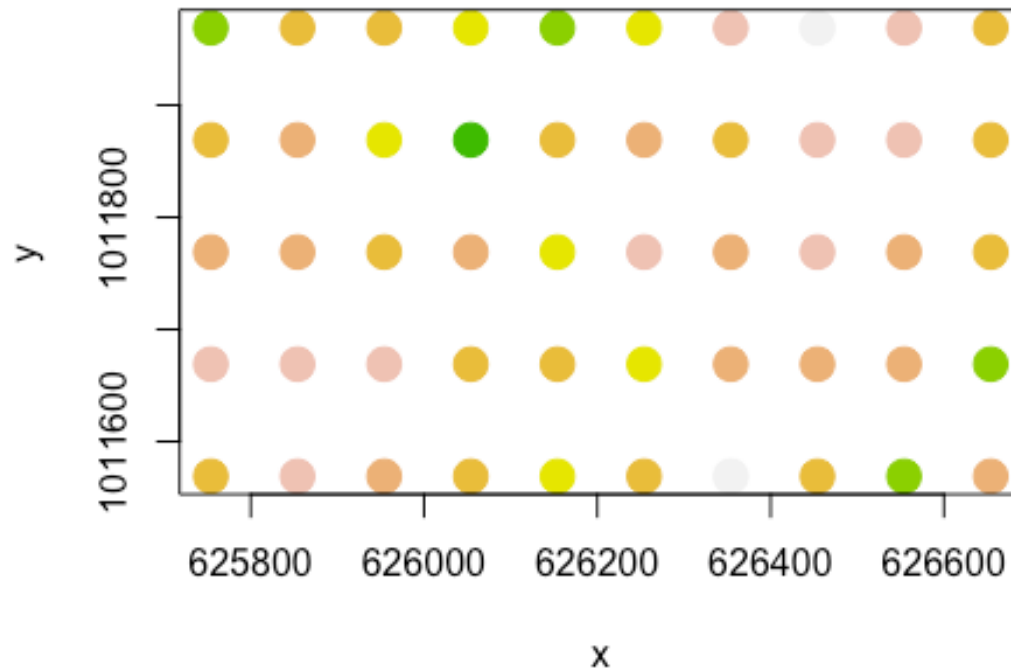1) Examine if there is evidence of spatial dependence in a rare and a common species in the BCI tree dataset

```r
#All Data
BCI_rowsum<-rowSums(BCI>0)
hist(BCI_rowsum)
```
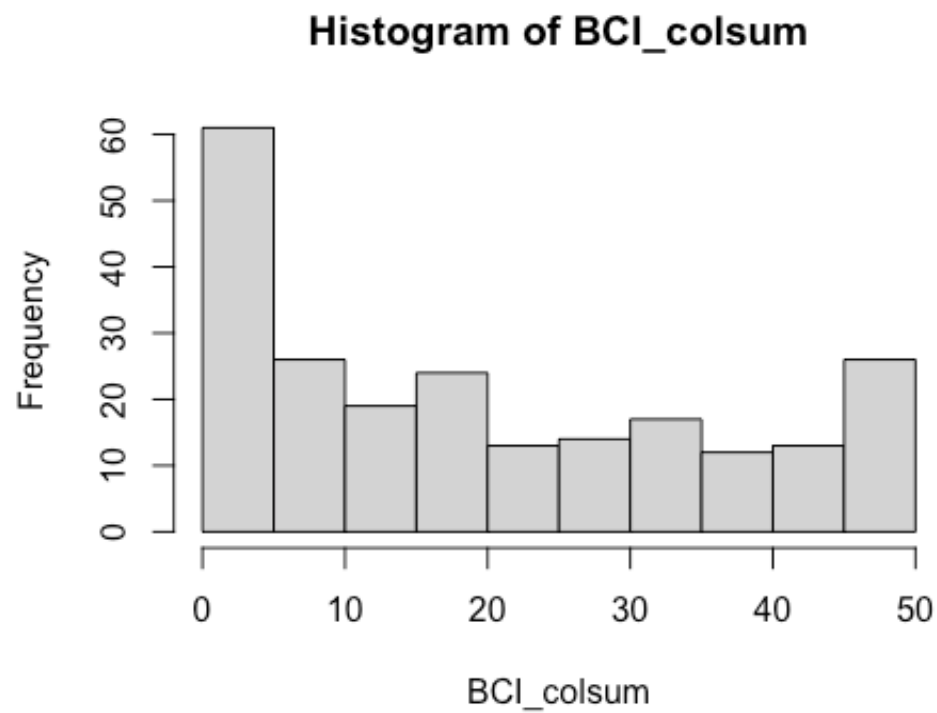
```
BCI_col_brks <- hist(BCI_rowsum, plot=F)$breaks
BCI_col_indices <- as.numeric(cut(BCI_rowsum, BCI_col_brks))
BCI_cols <- rev(terrain.colors(length(BCI_col_brks)))
plot(BCI_xy, cex=2, pch=19, col=BCI_cols[BCI_col_indices])
```
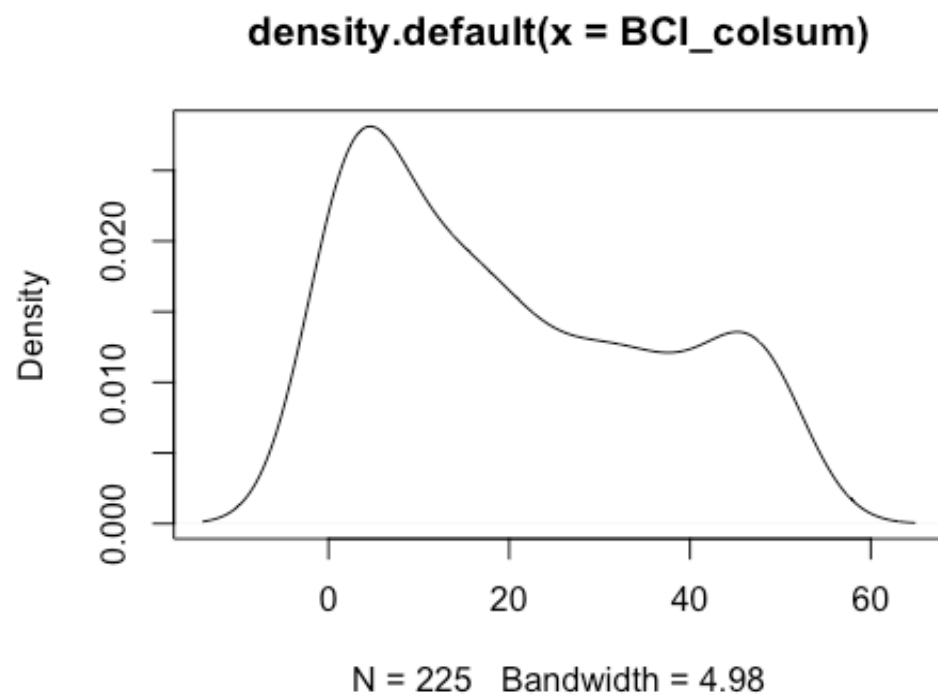


```
#Sum columns to determine number of observations of each species
BCI_colsum <- colSums(BCI > 0)
#Evaluate colsum values
quantile(BCI_colsum)

##   0%  25%  50%  75% 100%
##    1    5   17   34   50

hist(BCI_colsum)
```

Histogram of BCI_colsum


```
plot(density(BCI_colsum))
```

density.default(x = BCI_colsum)

N = 225   Bandwidth = 4.98

```r
#Identify a "rare" and a "common" species
which(BCI_colsum > 10 & BCI_colsum < 14)
```

```
##      Anacardium.excelsum Diospyros.artanthifolia      Hampea.appendiculata
##                      13                       60                        92
##         Inga.acuminata          Inga.spectabilis         Inga.umbellifera
##                     100                      111                      112
##          Laetia.procera          Licania.hypoleuca       Macrocnemum.roseum
##                     116                      119                      124
##      Siparuna.pauciflora          Theobroma.cacao          Trema.micrantha
##                     179                      201                      205
##          Trophis.caucana
##                     211
```

```r
which(BCI_colsum > 37 & BCI_colsum < 40)
```

```
##      Inga.sapindoides        Luehea.seemannii Protium.costaricense
##                   110                     123                  163
##      Sloanea.terniflora   Unonopsis.pittieri
##                   180                     214
```

```r
#Rare Species: Anacardium.excelsum, 13 individuals recorded
#Common Specie: Inga.sapindoides, 38 individuals recorded

#Evaluate spatial dependence in rare and common species
rare <- BCI[,13]
common <- BCI[,110]
#Rare Plots
hist(rare)
```
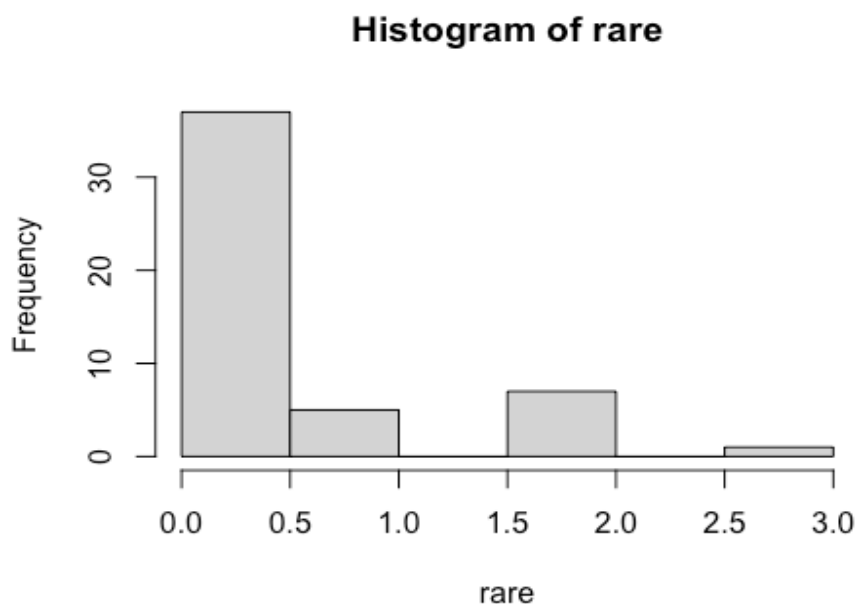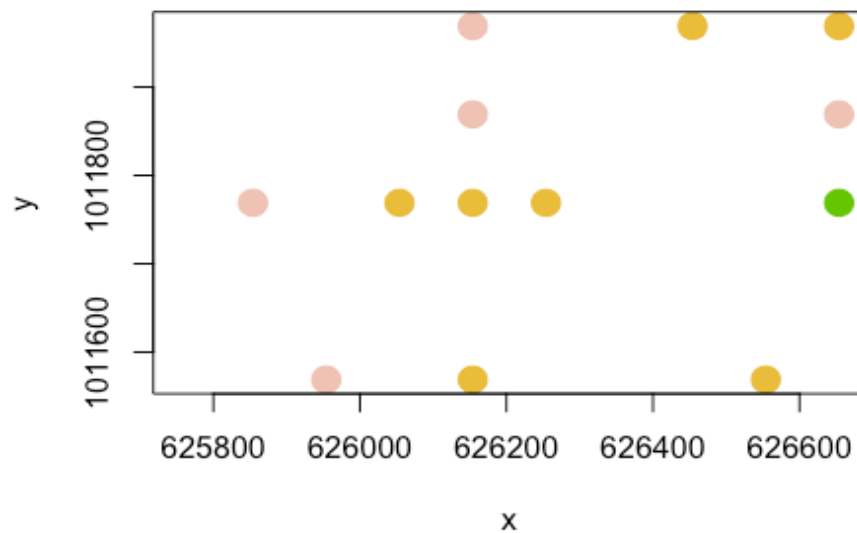

Histogram of rare
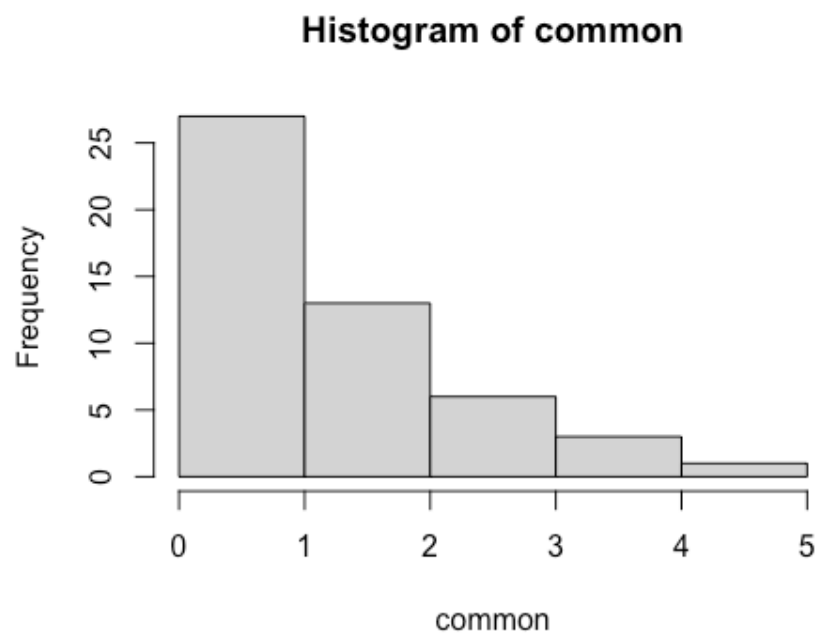
```
rare_col_brks <- hist(rare, plot=F)$breaks
rare_col_indices <- as.numeric(cut(rare, rare_col_brks))
rare_cols <- rev(terrain.colors(length(rare_col_brks)))
plot(BCI_xy, cex=2, pch=19, col=rare_cols[rare_col_indices])
```



```
#Common Plots
hist(common)
```

```
common_col_brks <- hist(common, plot=F)$breaks
common_col_indices <- as.numeric(cut(common, common_col_brks))
common_cols <- rev(terrain.colors(length(common_col_brks)))
plot(BCI_xy, cex=2, pch=19, col=common_cols[common_col_indices])
```
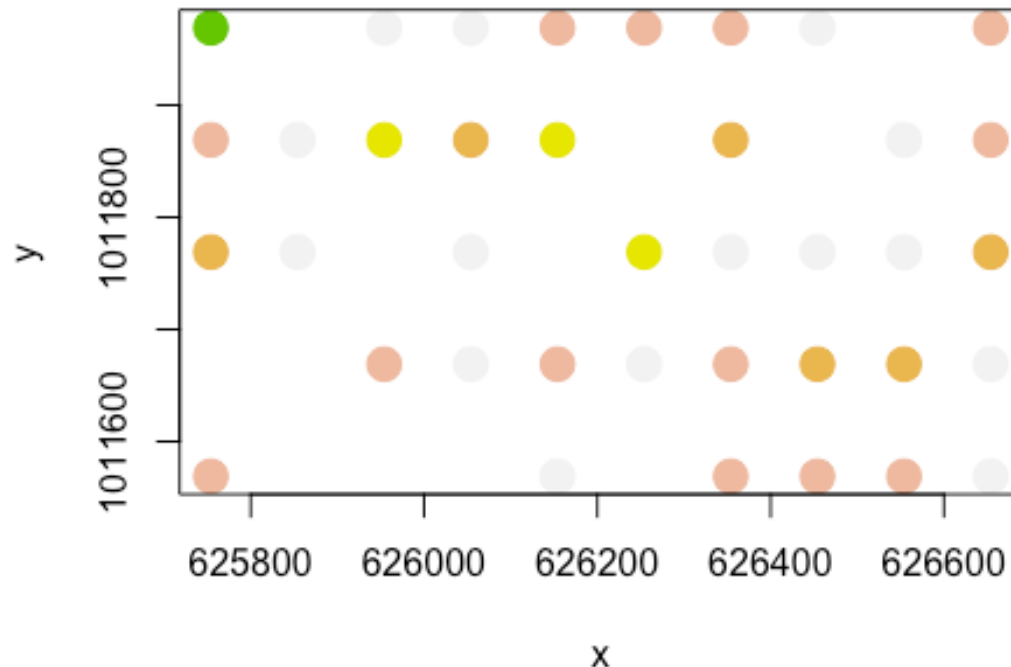

Spatial dependence does not seem to be present in either of the species evaluated (rare and common)

2) Build two generalized linear models to predict the abundance of the species *Drypetes standleyi* using the abundance of other tree species in the study site. Specifically examine the following species as predictor variables:

```
sp_ids = c("Cordia.lasiocalyx", "Hirtella.triandra",
           "Picramnia.latifolia", "Quassia.amara",
           "Tabernaemontana.arborea", "Trattinnickia.aspera",
           "Xylopia.macrantha")
```

Note renaming the species ids to something a little easier to work with like "sp_a", "sp_b" will make model construction a little less cumbersome

```
sp_i<-BCI[,63]  #Drypetes.standleyi
sp_a<-BCI[,51]  #Cordia.lasiocalyx
sp_b<-BCI[,97]  #Hirtella.triandra
sp_c<-BCI[,152] #Picramnia.latifolia
sp_d<-BCI[,171] #Quassia.amara
```

```
sp_e<-BCI[,193] #Tabernaemontana.arborea
sp_f<-BCI[,204] #Trattinnickia.aspera
sp_g<-BCI[,220] #Xylopia.macrantha
```

- Model 1: only include a single species as a predictor variable

```
#Visualize abundance of species i across sites
col_brks<-hist(sp_i, plot=F)$breaks
col_indices<-as.numeric(cut(sp_i, col_brks))
cols<-rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```
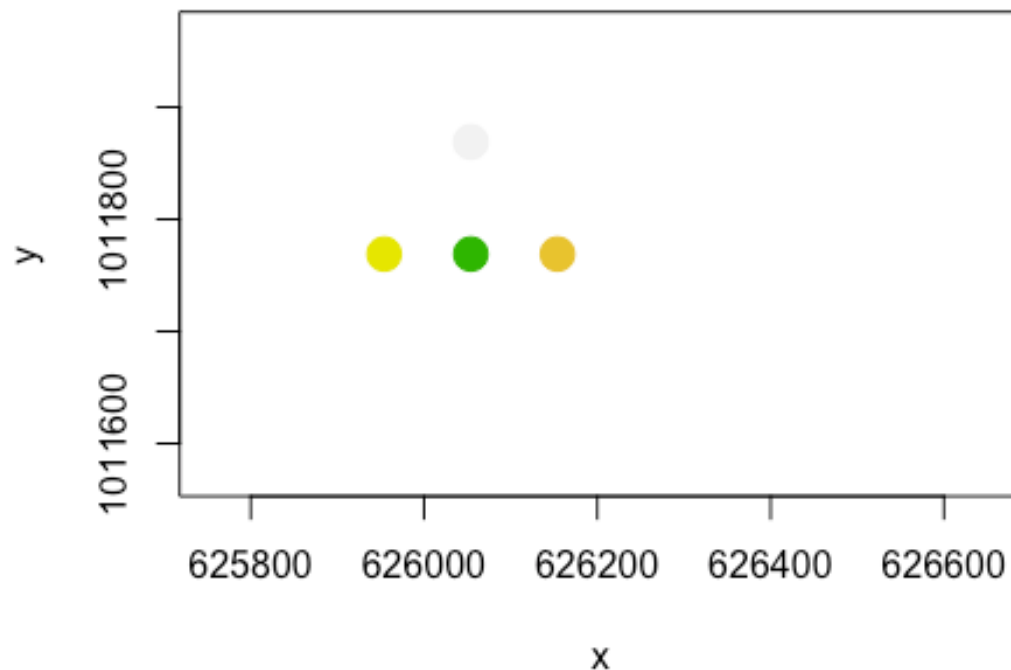


```
#Visualize abundance of species a across sites
col_brks<-hist(sp_a, plot=F)$breaks
col_indices<-as.numeric(cut(sp_a, col_brks))
cols<-rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```
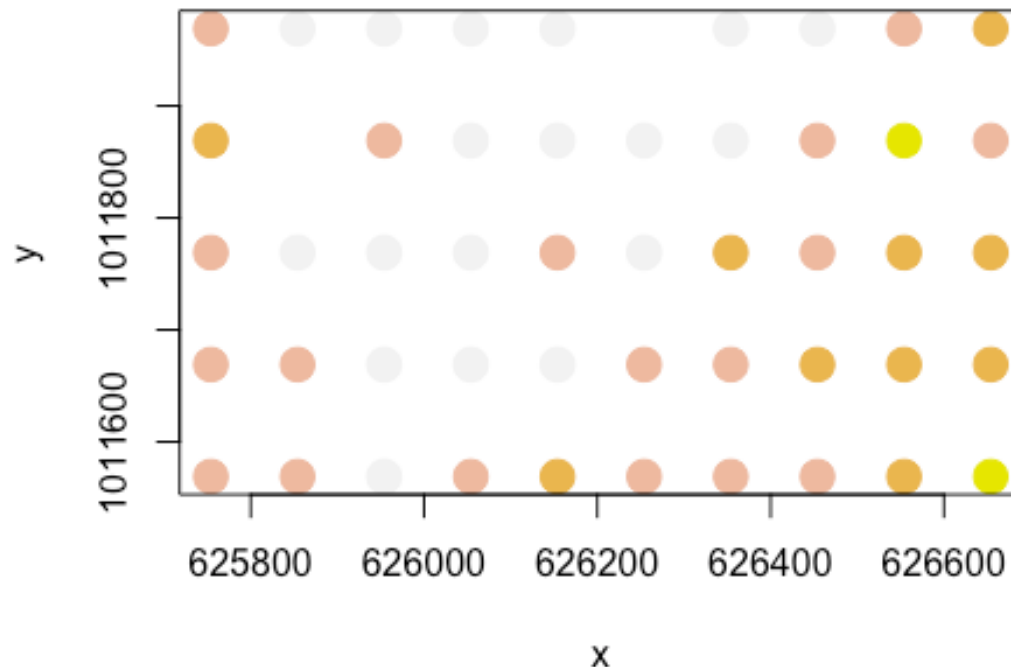
```
#Model comparing abundance of species i to abundance of species a
SingleSpModel<-glm(sp_i~sp_a, family="gaussian")
SingleSpModel

## 
## Call:  glm(formula = sp_i ~ sp_a, family = "gaussian")
## 
## Coefficients:
## (Intercept)          sp_a
##     0.83101       -0.05646
## 
## Degrees of Freedom: 49 Total (i.e. Null);  48 Residual
## Null Deviance:       134.2
## Residual Deviance: 131.5      AIC: 196.2

#abundance of species a at a given site does not correlate with abundance of
species i at a site, so species a is not a useful predictor variable for
modeling abundance of species i

#Is there another species in the group that is a better predictor of species
i abundance?
glm(sp_i~sp_a+sp_b+sp_c+sp_d+sp_e+sp_f+sp_g, family="gaussian")
```
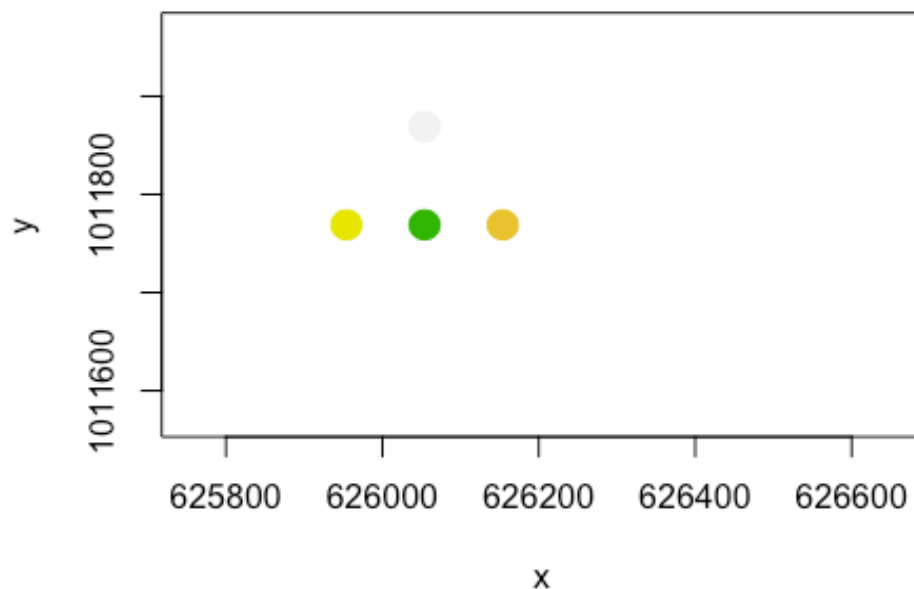
```
## 
## Call:  glm(formula = sp_i ~ sp_a + sp_b + sp_c + sp_d + sp_e + sp_f +
##     sp_g, family = "gaussian")
## 
## Coefficients:
## (Intercept)          sp_a          sp_b          sp_c          sp_d
sp_e
##      0.59551      -0.05833       0.02291      -0.20890       0.85144
0.04449
##          sp_f          sp_g
##      -0.34520       0.01632
## 
## Degrees of Freedom: 49 Total (i.e. Null);  42 Residual
## Null Deviance:        134.2
## Residual Deviance: 124.2      AIC: 205.4
```

#none of the species in the group returned values indicating significant correlation with species i

- Model 2: include all of the species as predictor variables

```
#Visualize abundance of species i across sites
col_brks<-hist(sp_i, plot=F)$breaks
col_indices<-as.numeric(cut(sp_i, col_brks))
cols<-rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```
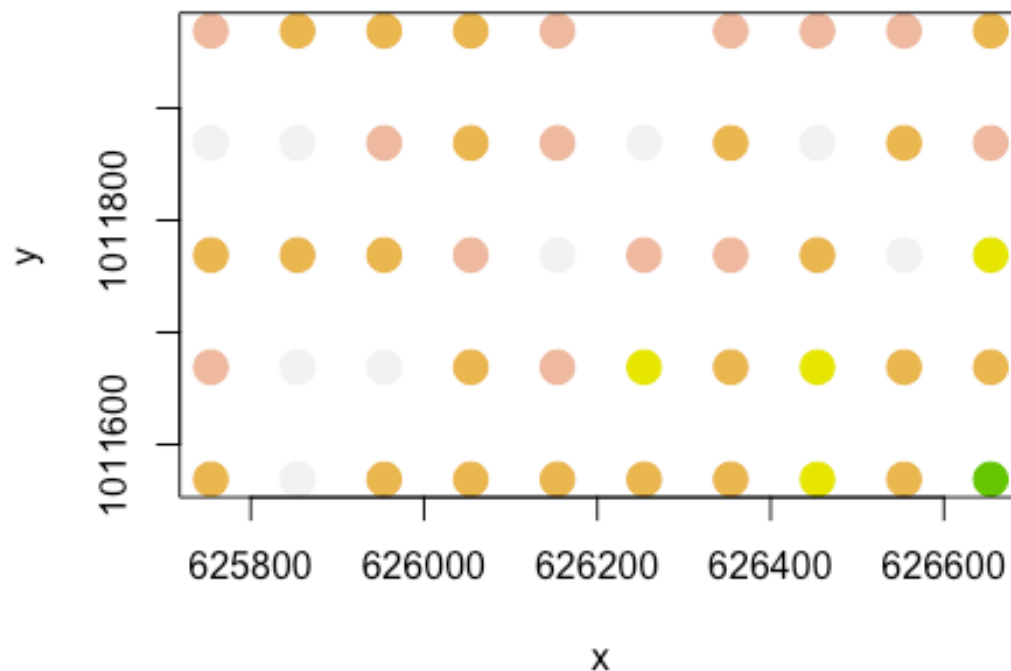
```
#Visualize abundance of group of species (a through g)
pred_sp<-data.frame(sp_a, sp_b, sp_c, sp_d, sp_e, sp_f, sp_g)
pred_sp_rs<-rowSums(pred_sp>0)
col_brks<-hist(pred_sp_rs, plot=F)$breaks
col_indices<-as.numeric(cut(pred_sp_rs, col_brks))
cols<-rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```



```
#Model comparing abundance of species i to abundance of group of species (a
through g)
MultiSpModel<-glm(sp_i~pred_sp_rs, family="gaussian")
MultiSpModel

##
## Call:  glm(formula = sp_i ~ pred_sp_rs, family = "gaussian")
##
## Coefficients:
## (Intercept)    pred_sp_rs
##      1.0588       -0.1439
##
## Degrees of Freedom: 49 Total (i.e. Null);  48 Residual
## Null Deviance:        134.2
## Residual Deviance: 133.1     AIC: 196.9
```
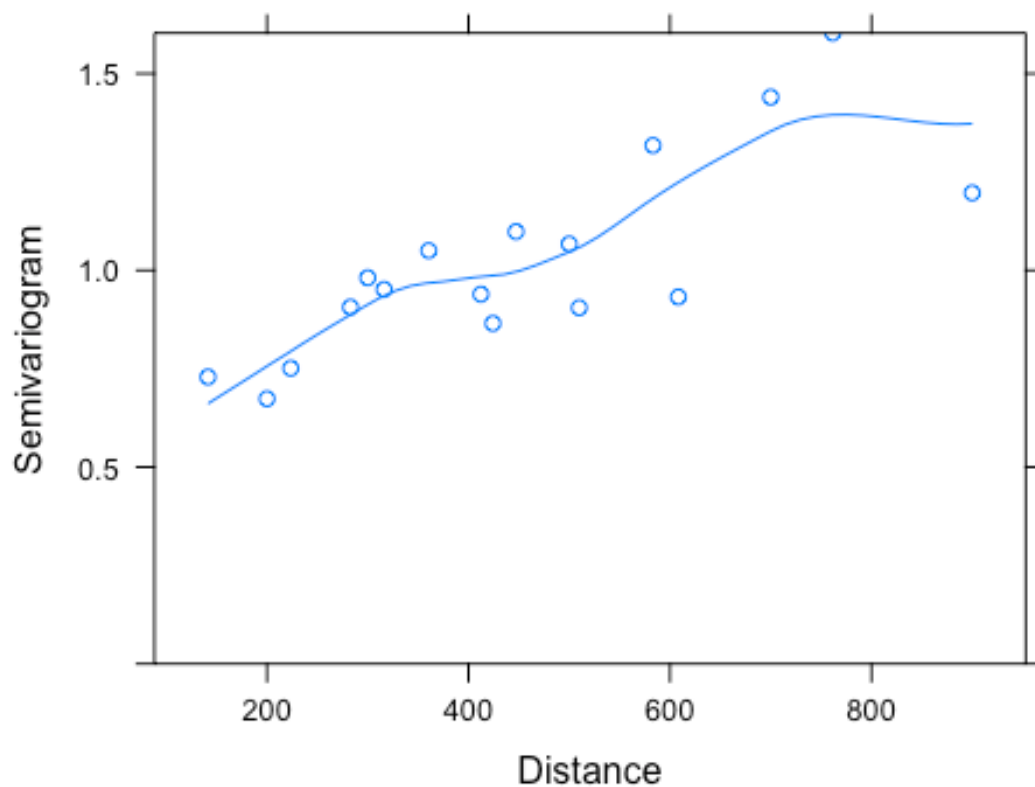
With both models examine the spatial dependence of the residuals using the function Variogram. Model the spatial dependence in the residuals using one of the error structures available.
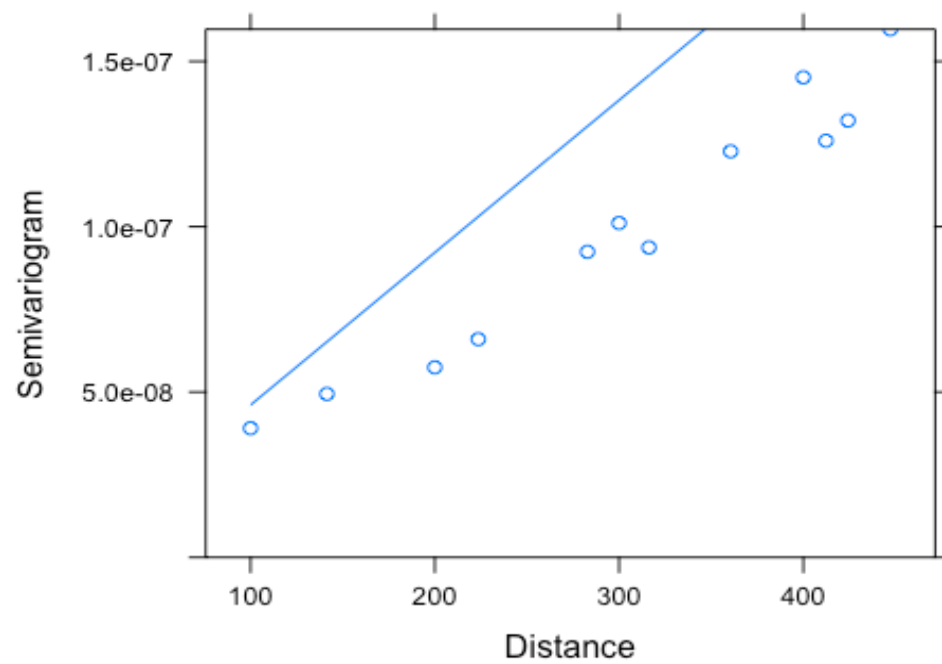
```
?Variogram
#Single Species Model: Initial Variogram
sr_dat <- data.frame(BCI, BCI_xy)
sr_lm <- gls(Drypetes.standleyi~Cordia.lasiocalyx, data=sr_dat)
plot(Variogram(sr_lm, form=~x+y))
```
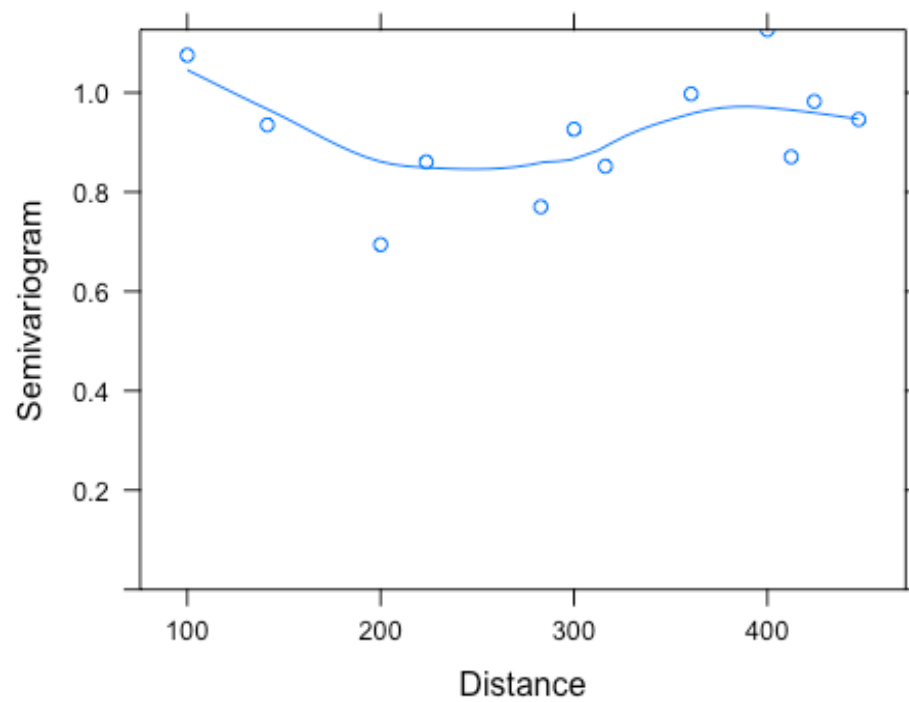


```
#Single Species Model: Exponential Variogram
sr_exp <- update(sr_lm, corr=corExp(form=~x+y))
xy_dist<- dist(BCI_xy)
max_dist <- max(xy_dist)/2
plot(Variogram(sr_exp, maxDist = max_dist))
```
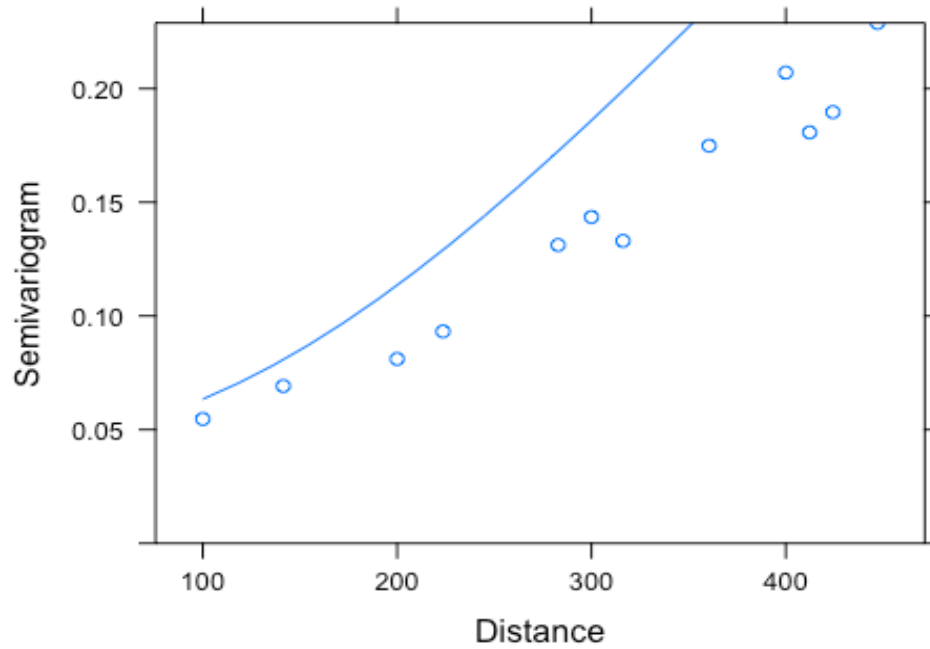
```
#Single Species Model: Normalized Exponential
plot(Variogram(sr_exp, resType='normalized', maxDist=max_dist))
```
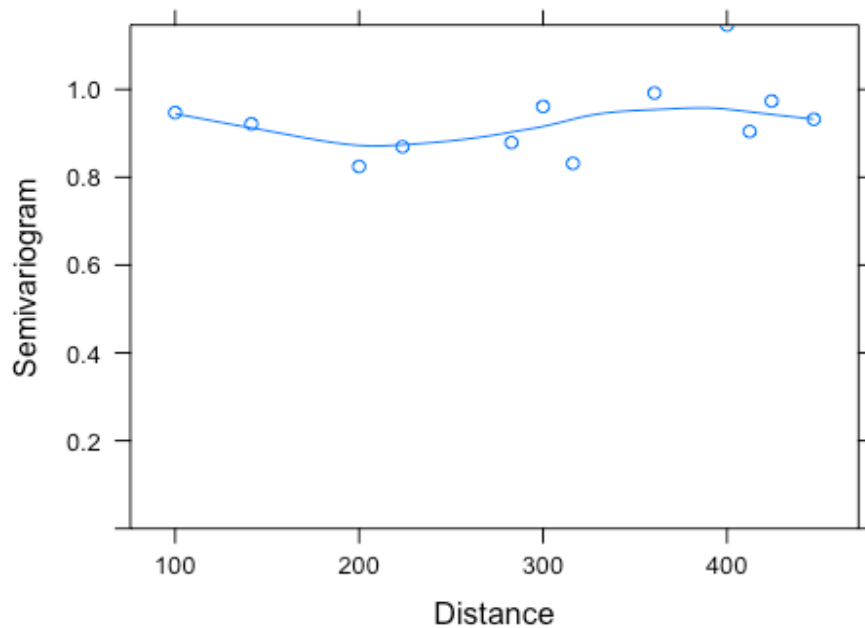
```
#Single Species Model: Rational Quadratic
sr_rat_nug <- update(sr_lm, corr=corRatio(form=~x+y, nugget=T))
plot(Variogram(sr_rat_nug, maxDist=max_dist))
```



```
#Single Species Model: Normalized Rational Quadratic
plot(Variogram(sr_rat_nug, resType='n', maxDist=max_dist))
```

```
#Compare models:
anova(sr_lm, sr_exp, sr_rat_nug)

##              Model df      AIC      BIC     logLik    Test    L.Ratio p-value
## sr_lm            1  3 335.1246 340.7382 -164.5623
## sr_exp           2  4 310.6438 318.1286 -151.3219 1 vs 2 26.480806  <.0001
## sr_rat_nug       3  5 308.6589 318.0149 -149.3295 2 vs 3  3.984871  0.0459

#Multiple Species Model: Initial Variogram
sr_dat <- data.frame(BCI, BCI_xy)
sr_lm_2 <-
gls(Drypetes.standleyi~Cordia.lasiocalyx+Hirtella.triandra+Picramnia.latifoli
a+Quassia.amara+Tabernaemontana.arborea+Trattinnickia.aspera+Xylopia.macranth
a, data=sr_dat)
plot(Variogram(sr_lm_2, form=~x+y))
```
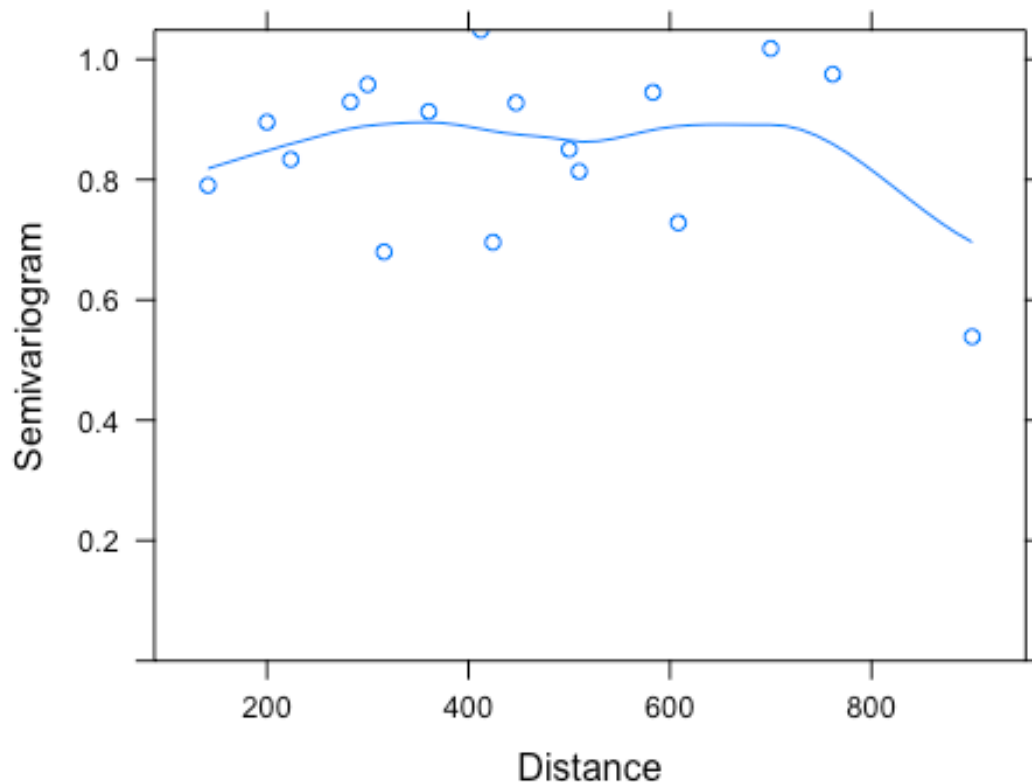


```
#Multiple Species Model: Exponential Variogram
sr_exp_2 <- update(sr_lm_2, corr=corExp(form=~x+y))
xy_dist<- dist(BCI_xy)
max_dist <- max(xy_dist)/2
plot(Variogram(sr_exp_2, maxDist = max_dist))
```
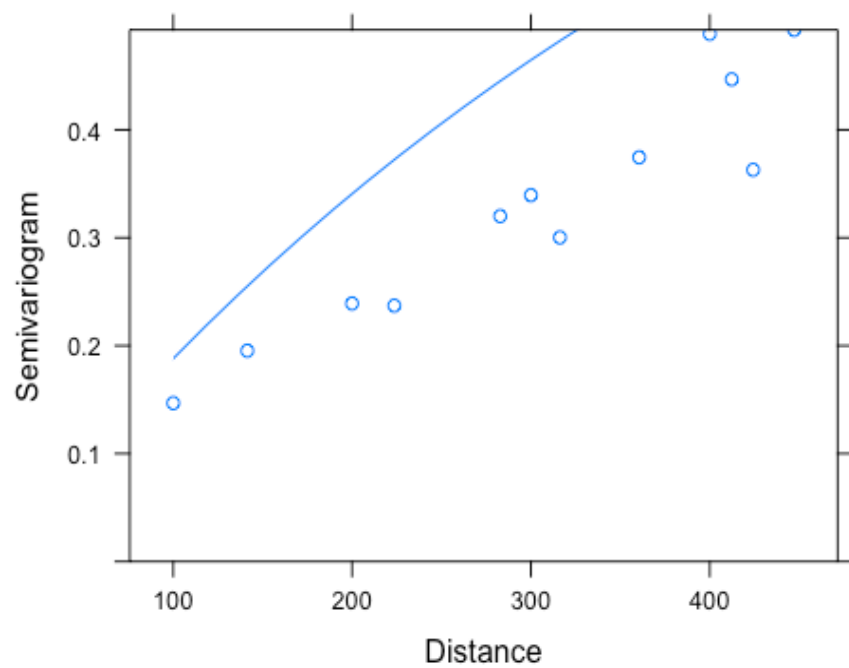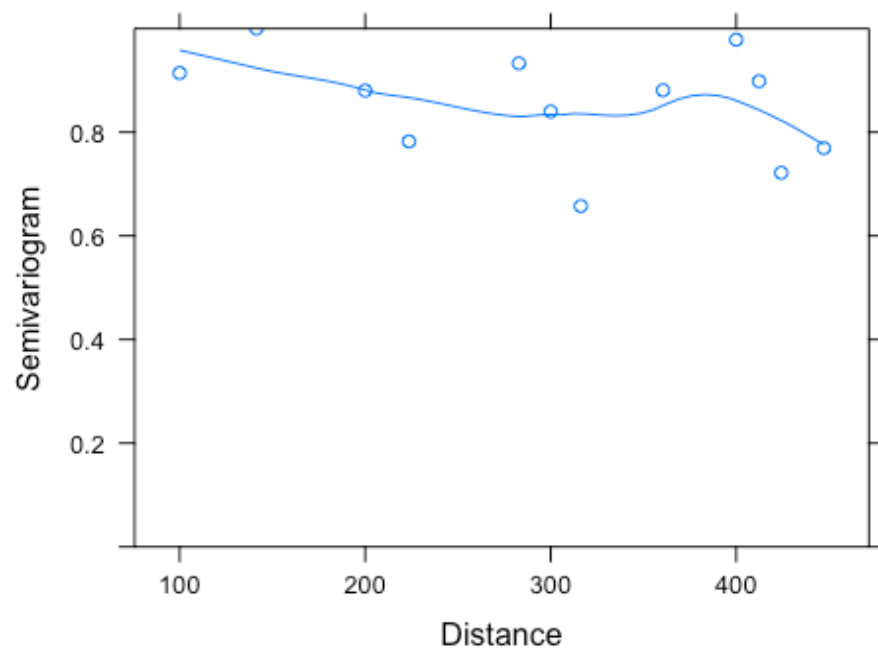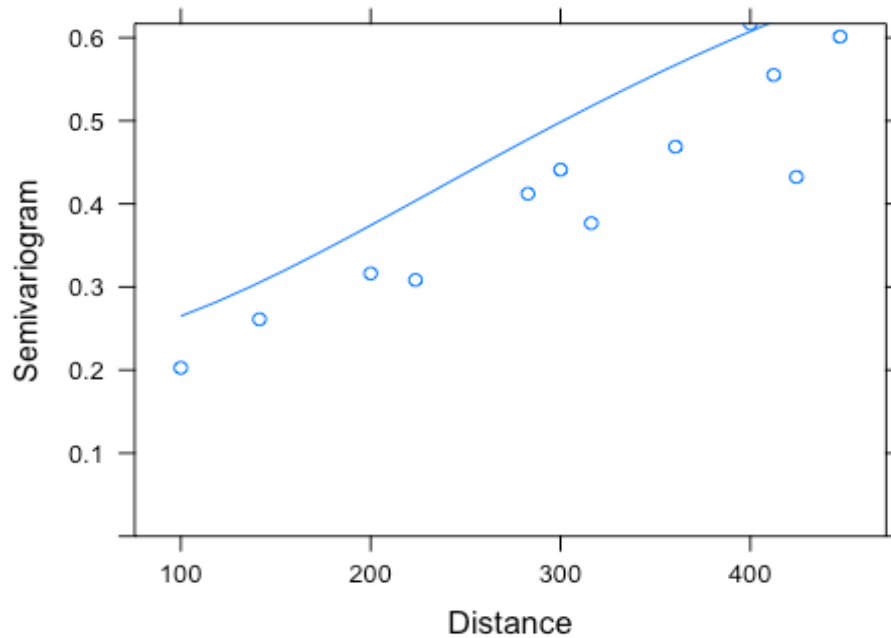
```
#Multiple Species Model: Normalized Exponential
plot(Variogram(sr_exp_2, resType='normalized', maxDist=max_dist))
```
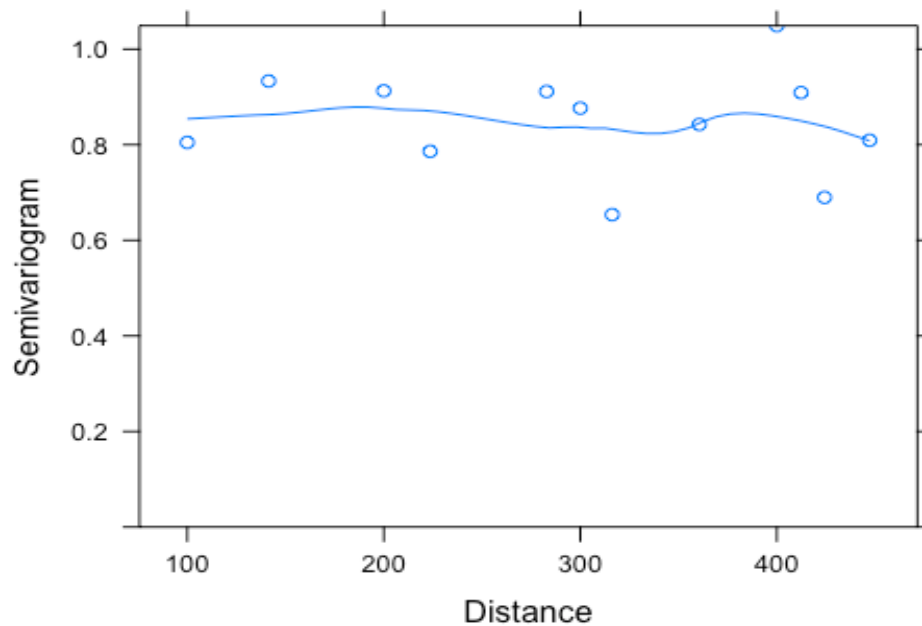
```
#Multiple Species Model: Rational Quadratic
sr_rat_nug_2 <- update(sr_lm_2, corr=corRatio(form=~x+y, nugget=T))
plot(Variogram(sr_rat_nug_2, maxDist=max_dist))
```



```
#Multiple Species Model: Normalized Rational Quadratic
plot(Variogram(sr_rat_nug_2, resType='n', maxDist=max_dist))
```

```
#Compare models:
anova(sr_lm_2, sr_exp_2, sr_rat_nug_2)

##                Model df      AIC      BIC     logLik   Test   L.Ratio p-value
## sr_lm_2           1  9 307.1163 322.7554 -144.5582
## sr_exp_2          2 10 301.6062 318.9829 -140.8031 1 vs 2 7.510175  0.0061
## sr_rat_nug_2      3 11 303.1486 322.2630 -140.5743 2 vs 3 0.457530  0.4988
```

Did including the spatial error term have a large impact on the coefficients of the model?

The t-values in the models changed a fair amount when spatial error was taken into account. Despite this, the single species model still did not show any significant correlation, while the exponential multi-species model resulted in significant results.

Did including the spatial error terms significantly improve model fit (use function anova to carry out model comparison)?

Yes, though the incorporation of spatial error did not make all of the resulting models significant, the models with spatial error considerations (sr_exp and sr_rat_nug) were better fits for the data than those without spatial considerations (sr_lm) in both the single species and multi-species models. This evaluation was made based on the resulting AIC values when evaluated using the anova function (lower AIC indicates better fit).

Explain why you did or did not observe a difference in the influence of adding the spatial error term between the two models.

A difference was seen because taking the spatial value into consideration reduces some of the autocorrelation found in the model. Though explanatory variables can account for large portions of variability in species distribution, there will always be an underlying factor of range and distribution capability to take into account. In other words, species may simply be grouped together because they did not travel far from the parent, not due to an environmental explanation. This is the variability that is taken into account when adding spatial error to the model.