

# Spatial\_Models\_CT

Caroline Tribble

```
library(nlme)
# Oribatid mite data. 70 soil cores collected by Daniel Borcard in 1989.
# See Borcard et al. (1992, 1994) for details.
data(mite)
```

```
## Warning in data(mite): data set 'mite' not found
```

```
data(mite.env)
```

```
## Warning in data(mite.env): data set 'mite.env' not found
```

```
data(mite.xy)
```

```
## Warning in data(mite.xy): data set 'mite.xy' not found
```

```
?mite
```

```
## No documentation for 'mite' in specified packages and libraries:
## you could try '??mite'
```

```
library(vegan)
```

```
## Loading required package: permute
```

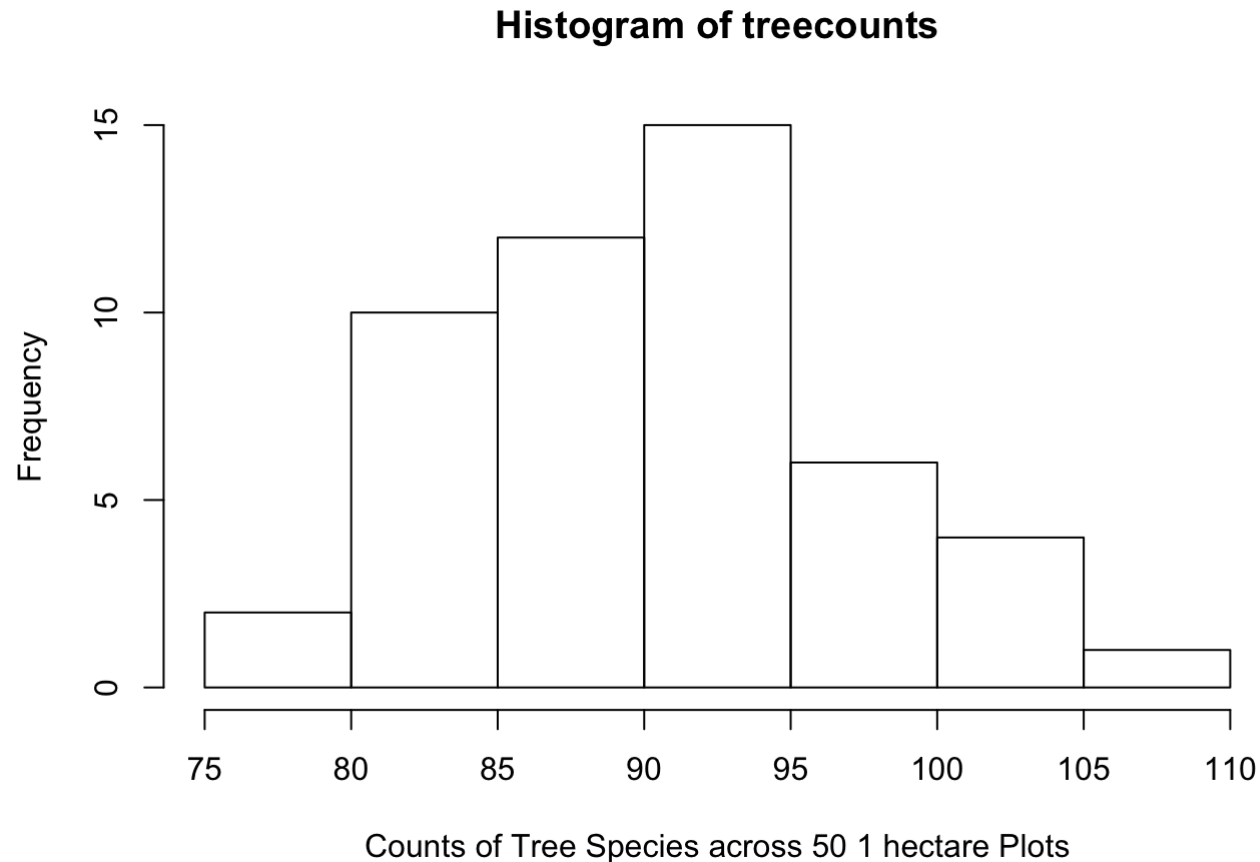
```
## Loading required package: lattice
```

```
## This is vegan 2.5-6
```

```
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

```
treecounts <- rowSums(BCI > 0)
hist(treecounts, xlab = "Counts of Tree Species across 50 1 hectare Plots")
```

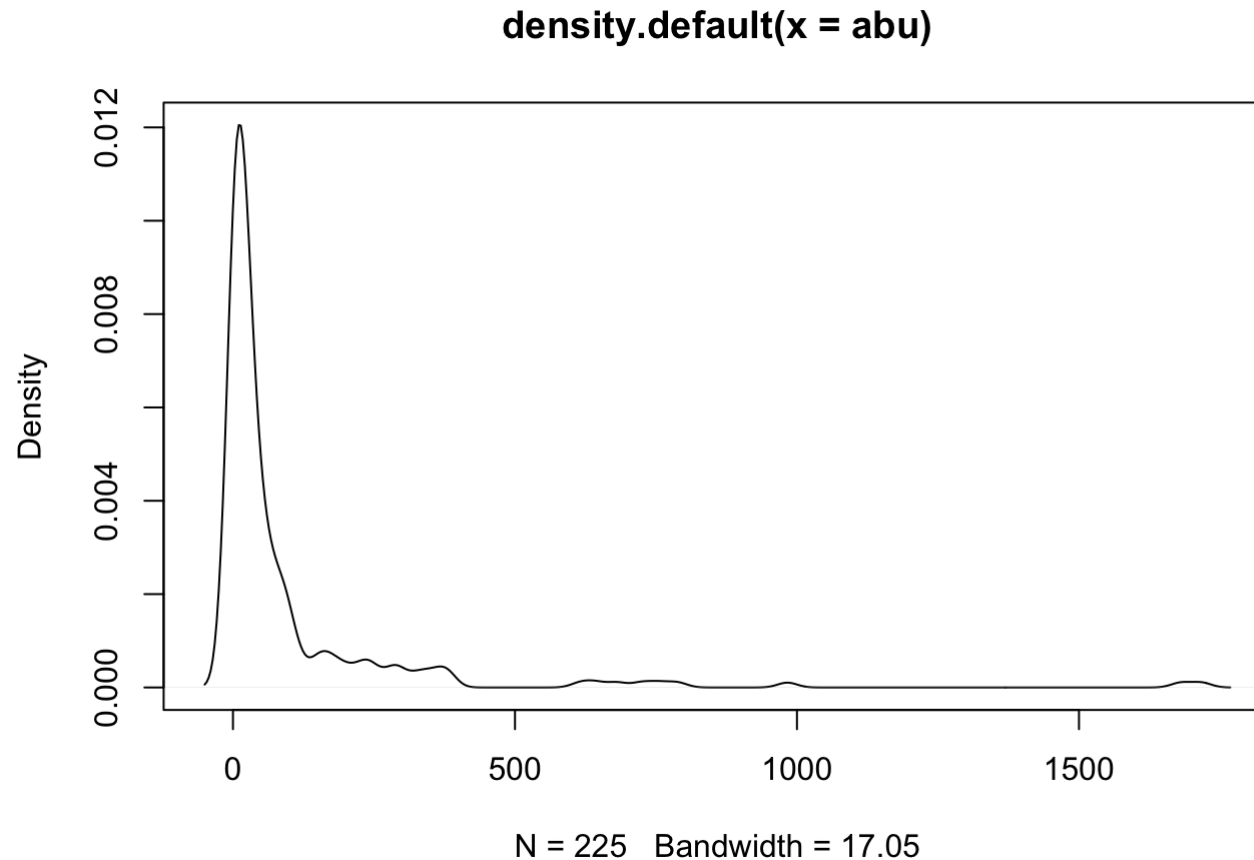


This histogram shows the frequency distribution of tree counts across all 225 species surveyed in these 1 hectare plots in the Barro Colorado Island. The tree count frequency appears to increase exponentially until reaching the majority of plots that have counts between 90-95 trees, from here frequency in counts decreases, so abundance may level out here due to competition for space within a given hectare.

```
abu <- colSums(BCI)
quantile(abu)
```

##	0%	25%	50%	75%	100%
##	1	7	25	82	1717

```
plot(density(abu))
```



Highly skewed, for taking quantiles of this distribution it would make more sense to log transform it first before taking quantiles to get more reasonable lower and upper quantiles.

```
quantile(log10(abu))
```

```
##          0%          25%          50%          75%          100%
## 0.000000 0.845098 1.397940 1.913814 3.234770
```

```
10^0.845
```

```
## [1] 6.99842
```

```
10^1.913814
```

```
## [1] 82.00003
```

```
10^1.397940
```

```
## [1] 25
```

```
which(abu > 25 & abu < 27)
```

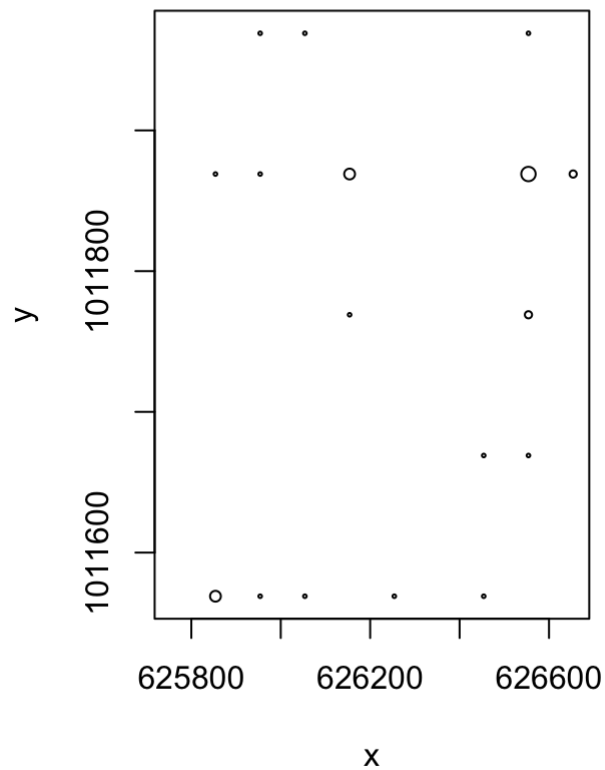
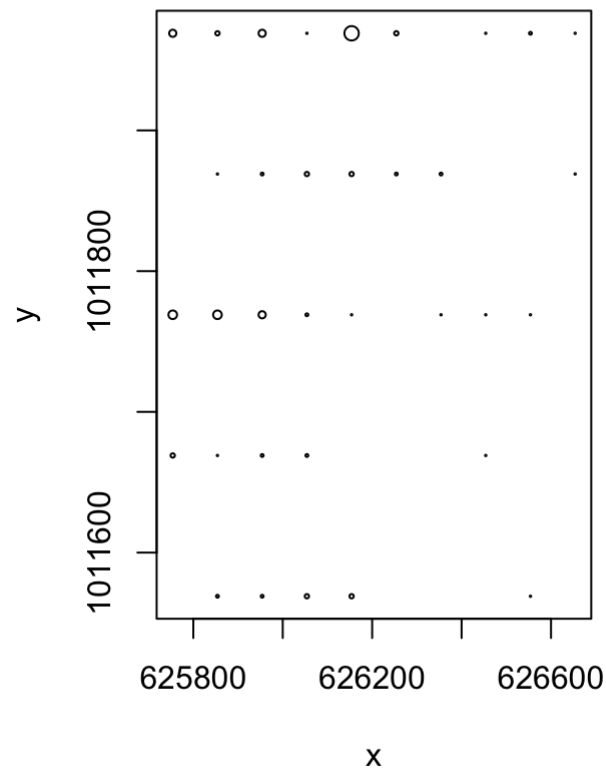
```
## Erythrina.costaricensis      Inga.acuminata      Sterculia.apetala
##                65                100                187
## Symphonia.globulifera
##                190
```

```
which(abu > 10^1.93 & abu < 10^1.95)
```

```
## Cassipourea.guianensis  Dendropanax.arboreus
##                32                58
```

```
rare_sp <- BCI[,65] #E. costaricensis
comm_sp <- BCI[,58] #D. arboreus
```

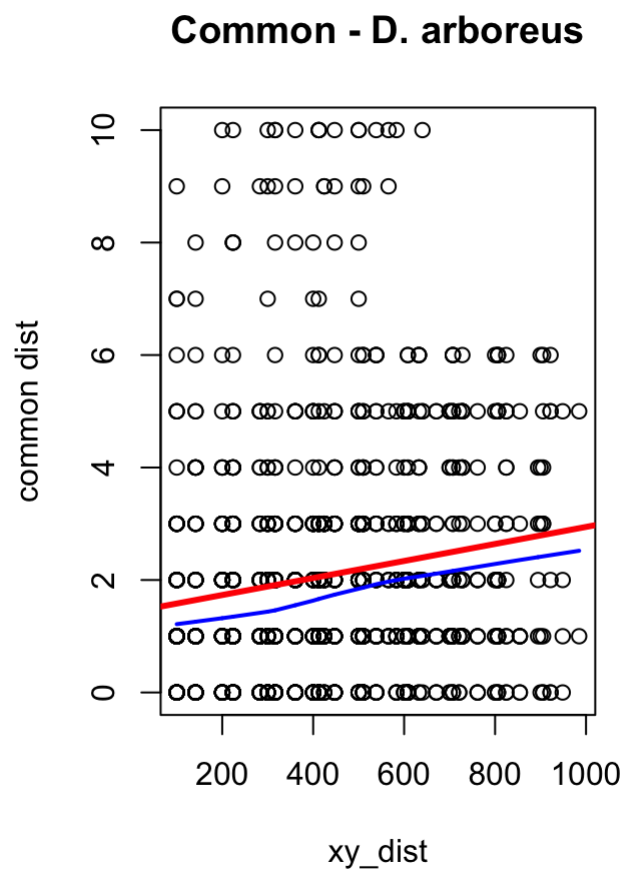
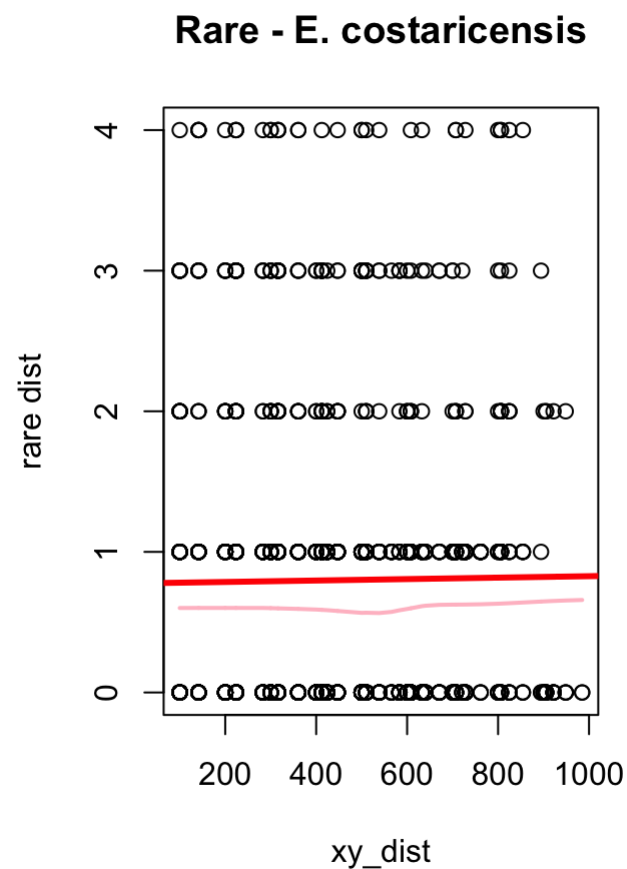
```
par(mfrow=c(1,2))
plot(BCI_xy, cex = rare_sp/ max(rare_sp), main = "E. costaricensis Spatial Dist")
plot(BCI_xy, cex = comm_sp/ max(comm_sp), main = "D. arboreus Spatial Dist")
```

**E. costaricensis Spatial Dist****D. arboreus Spatial Dist**

From these spatial distribution plots it is clear *E. costaricensis* is our rare species, seeming to have random hotspots along similar gradient lines as *D. arboreus*, the common species. *D. arboreus* appears to occur along gradient lines in few numbers, however there is one outlier hotspot of greater tree counts.

```
geod <- dist(BCI_xy)
rared <- dist(rare_sp)
commd <- dist(comm_sp)
```

```
par(mfrow=c(1,2))
plot(geod, rared, main = 'Rare - E. costaricensis', xlab = "xy_dist", ylab = "rare dist")
abline(lm(rared ~ geod), lwd=3, col='red')
lines(lowess(geod, rared), lwd = 2, col = 'pink')
plot(geod, commd, main = 'Common - D. arboreus', xlab = "xy_dist", ylab = "common dist")
abline(lm(commd ~ geod), lwd=3, col='red')
lines(lowess(geod, commd), lwd = 2, col = 'blue')
```



Rare species looks spatially dependent (flat line with correlation near zero) and the common species shows some degree of autocorrelation indicating the variable is spatially non-random in its distribution. Can formally test for the common species.

```
robs_cor <- cor(geod, rared)
cobs_cor <- cor(geod, commd)
robs_cor
```

```
## [1] 0.01075826
```

```
cobs_cor
```

```
## [1] 0.1564616
```

No correlation for the rare species as expected and a very weak one for the common species.

```
mantel(geod, rared)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = geod, ydis = rared)
##
## Mantel statistic r: 0.01076
##      Significance: 0.434
##
## Upper quantiles of permutations (null model):
##      90%      95%   97.5%    99%
## 0.0872 0.1110 0.1290 0.1540
## Permutation: free
## Number of permutations: 999
```

Here the p-value is greater than 0.05 further supporting there is no spatial autocorrelation in our rare species model.

```
mantel(geod, commd)
```

```
##  
## Mantel statistic based on Pearson's product-moment correlation  
##  
## Call:  
## mantel(xdis = geod, ydis = commd)  
##  
## Mantel statistic r: 0.1565  
##      Significance: 0.004  
##  
## Upper quantiles of permutations (null model):  
##      90%      95%     97.5%      99%  
## 0.0813 0.1012 0.1148 0.1334  
## Permutation: free  
## Number of permutations: 999
```

Conversely, the common species displays a p-value of 0.008 indicating a statistically significant positive spatial auto=correlation pattern. From here, distances can be grouped into bins and then estimated at each distance class to understand the pattern.

```
comm_corlog <- mantel.correlog(commd, geod)  
  
comm_corlog
```

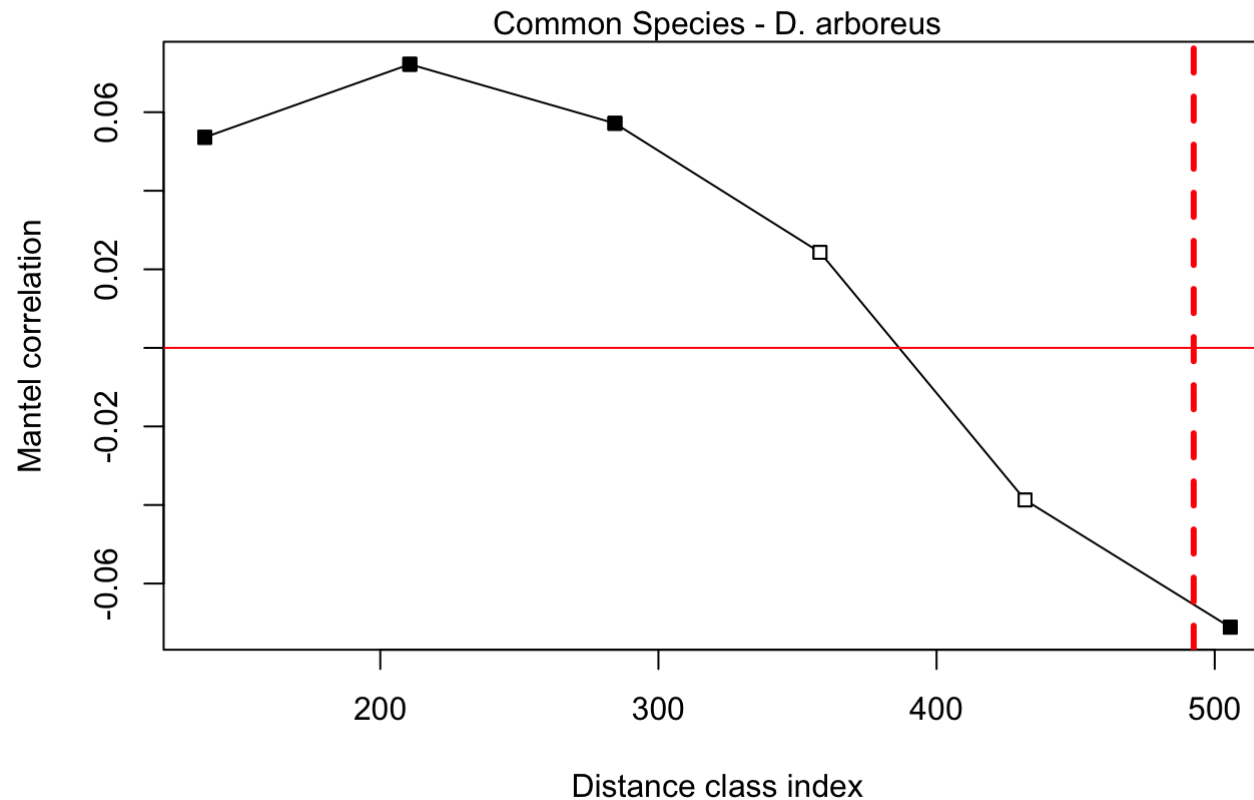


```
##
## Mantel Correlogram Analysis
##
## Call:
## mantel.correlog(D.eco = commd, D.geo = geod)
##
##      class.index      n.dist Mantel.cor Pr(Mantel) Pr(corrected)
## D.cl.1      136.870241 144.000000    0.053599    0.009    0.009 **
## D.cl.2      210.610723 376.000000    0.072199    0.002    0.004 **
## D.cl.3      284.351204 390.000000    0.057169    0.008    0.016 *
## D.cl.4      358.091686 148.000000    0.024322    0.140    0.140
## D.cl.5      431.832168 372.000000   -0.038728    0.105    0.210
## D.cl.6      505.572649 266.000000   -0.071122    0.001    0.006 **
## D.cl.7      579.313131 168.000000         NA         NA         NA
## D.cl.8      653.053613 100.000000         NA         NA         NA
## D.cl.9      726.794094 154.000000         NA         NA         NA
## D.cl.10     800.534576  88.000000         NA         NA         NA
## D.cl.11     874.275058  50.000000         NA         NA         NA
## D.cl.12     948.015539  24.000000         NA         NA         NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With all these distance classes, as we do more and more tests of classes have a multiple comparisons problem, p corrected values correct based on the number of tests we have done.

```
max_dist <- max(geod) / 2

plot(comm_corlog)
mtext(side=3, 'Common Species - D. arboreus')
abline(v = max_dist, col='red', lwd=3, lty=2)
```



Graphical representation of the Mantel correlogram for *D. arboreus*. Black dots represent distance classes that are statistically significant in the model and white dots represent non-significance. Points that are closer in space are more similar and those that are further apart show more dissimilarities with greater distance due to chance.

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:

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

Model 2: include all of the species as predictor variables

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.

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

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

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

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

```
sp_ref <- BCI$Drypetes.standleyi  
sp_a <- BCI$Cordia.lasiocalyx  
sp_b <- BCI$Hirtella.triandra  
sp_c <- BCI$Picramnia.latifolia  
sp_d <- BCI$Quassia.amara  
sp_e <- BCI$Tabernaemontana.arborea  
sp_f <- BCI$Trattinnickia.aspera  
sp_g <- BCI$Xylopia.macrantha
```

```
pred_sp <- BCI[,sp_ids]  
names(pred_sp)
```

```
## [1] "Cordia.lasiocalyx"      "Hirtella.triandra"  
## [3] "Picramnia.latifolia"   "Quassia.amara"  
## [5] "Tabernaemontana.arborea" "Trattinnickia.aspera"  
## [7] "Xylopia.macrantha"
```

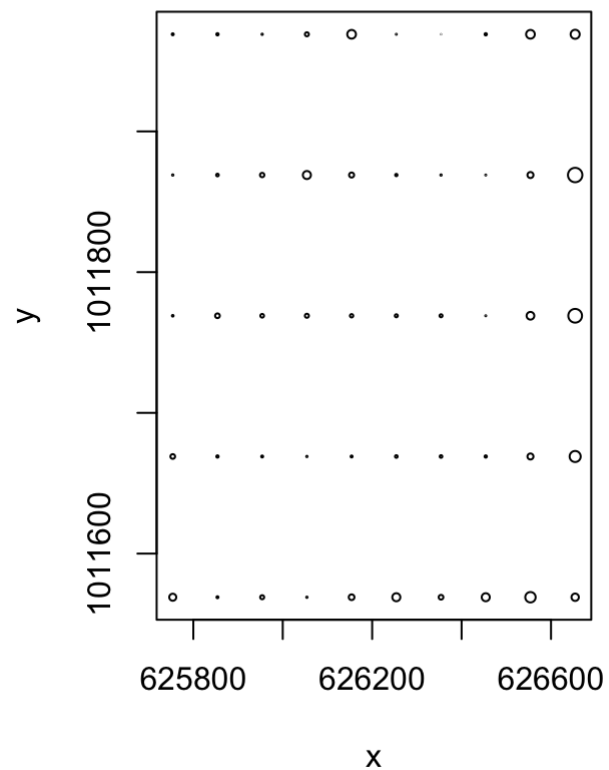
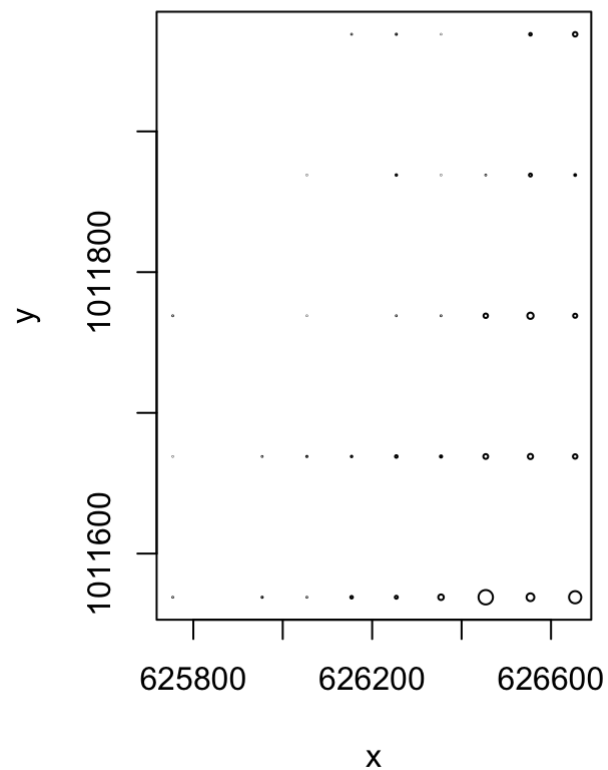
```
head(pred_sp)
```

```
## Cordia.lasiocalyx Hirtella.triandra Picramnia.latifolia Quassia.amara
## 1      8      21      0      0
## 2      6      14      0      0
## 3      6       5      1      0
## 4     11       4      0      0
## 5      7       6      0      0
## 6      6       6      0      0
## Tabernaemontana.arborea Trattinnickia.aspera Xylopia.macrantha
## 1      9       3      1
## 2      5       1      0
## 3      6       1      0
## 4     10       0      0
## 5     16       2      0
## 6     11       0      0
```

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

Influence of *H. triandra* spatial distribution and abundanace on that of *D. standleyi*, another tree species in the BCI dataset.

```
par(mfrow=c(1,2))
plot(BCI_xy, cex = sp_b/ max(sp_b), main = "H. trianda Spatial Dist")
plot(BCI_xy, cex = sp_ref/ max(sp_ref), main = "D. standleyi Spatial Dist")
```

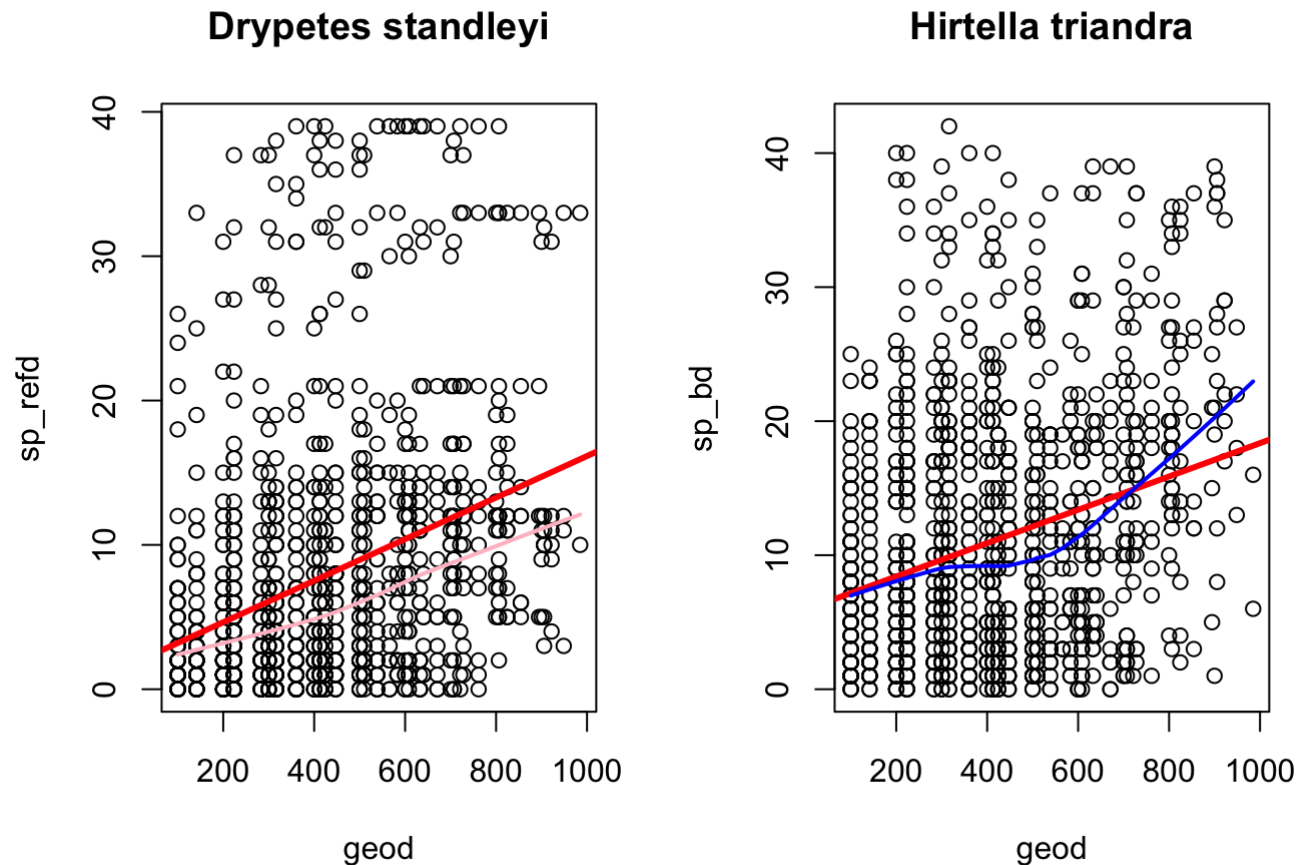
**H. trianda Spatial Dist****D. standleyi Spatial Dist**

```
geod <- dist(BCI_xy)
sp_bd <- dist(sp_b)
sp_refd <- dist(sp_ref)
maxdist <- max(geod) / 2
```

```

par(mfrow=c(1,2))
plot(geod, sp_refd, main = 'Drypetes standleyi')
abline(lm(sp_refd ~ geod), lwd=3, col='red')
lines(lowess(geod, sp_refd), lwd = 2, col = 'pink')
plot(geod, sp_bd, main = 'Hirtella triandra')
abline(lm(sp_bd ~ geod), lwd=3, col='red')
lines(lowess(geod, sp_bd), lwd = 2, col = 'blue')

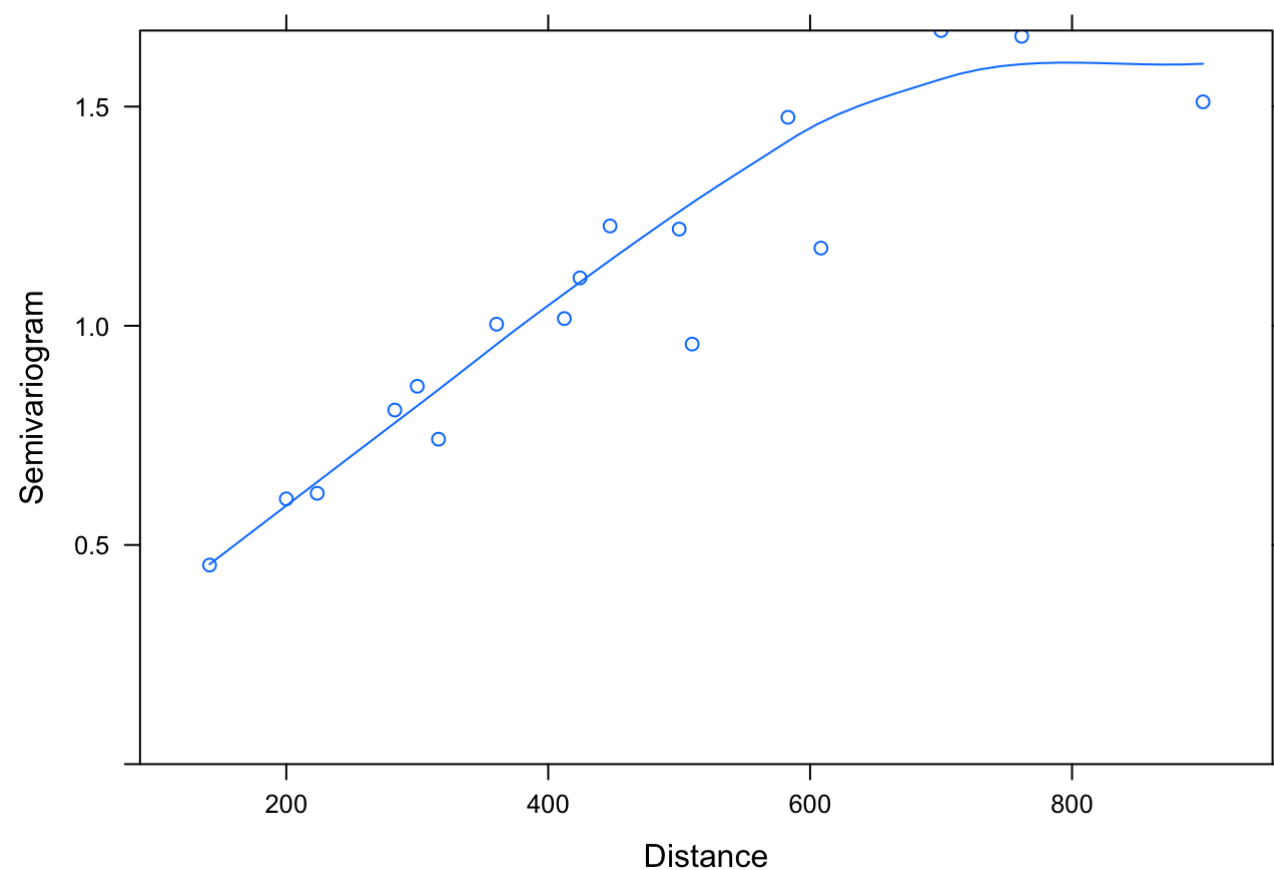
```



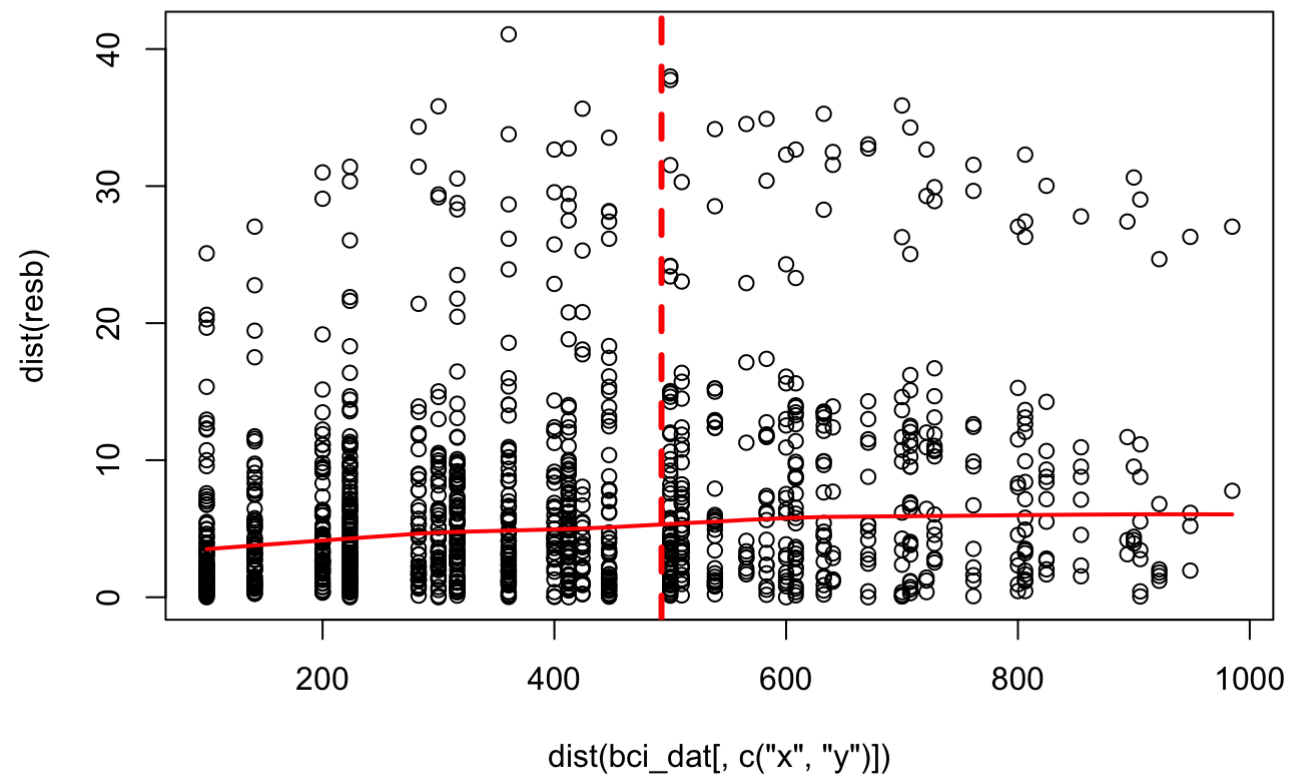
From these plots we can see both tree species exhibit spatial auto-correlation. *H. triandra* has an interesting pattern changing slopes across distances while *D. standleyi* follows the regression line more closely.

```
bci_dat = data.frame(BCI_xy, BCI)
lm_b = gls(sp_ref ~ sp_b, data=bci_dat)
resb = residuals(lm_b)

plot(Variogram(lm_b, form= ~ x + y))
```



```
plot(dist(bci_dat[, c('x', 'y')]), dist(resb))
lines(lowess(dist(bci_dat[, c('x', 'y')]), dist(resb)), col='red', lwd=2)
abline(v = maxdist, col='red', lwd=3, lty=2)
```

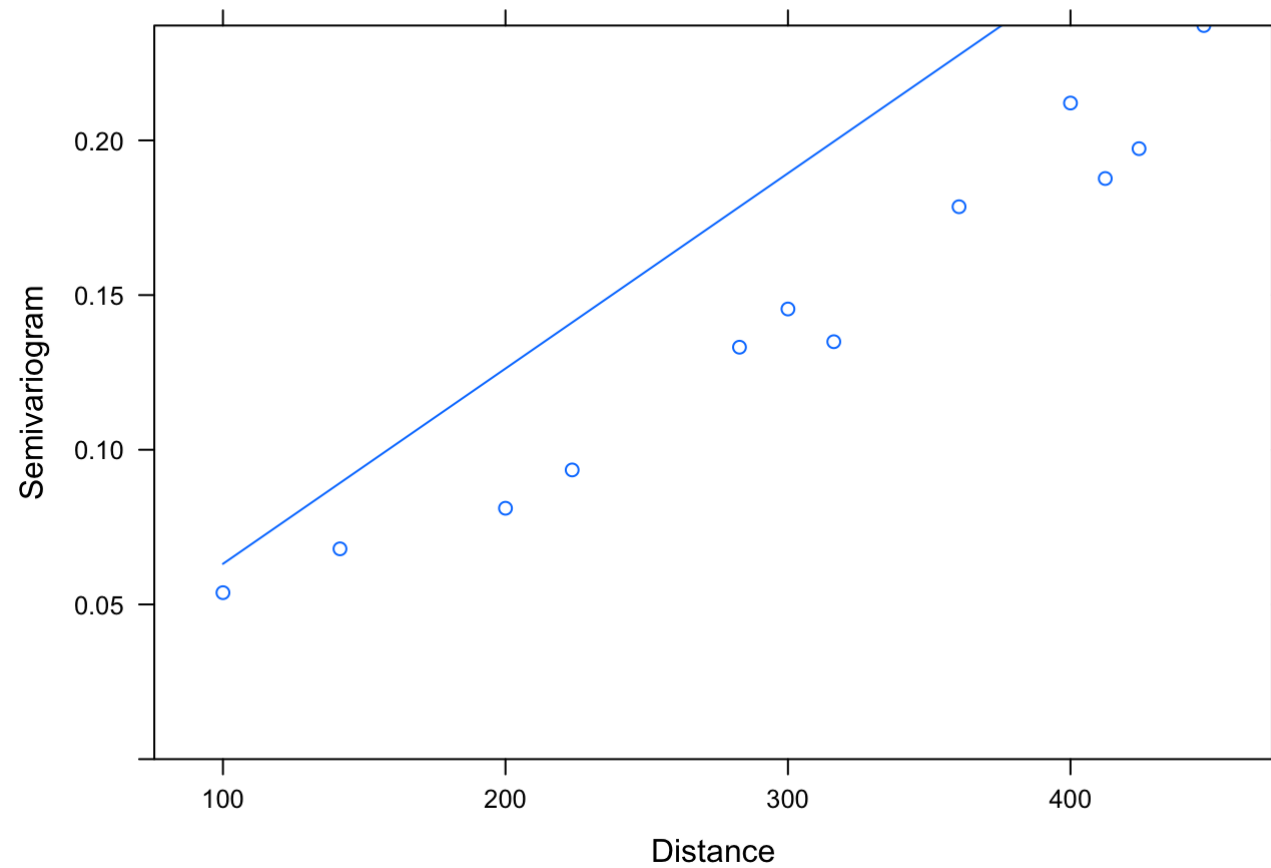


Linear Model -

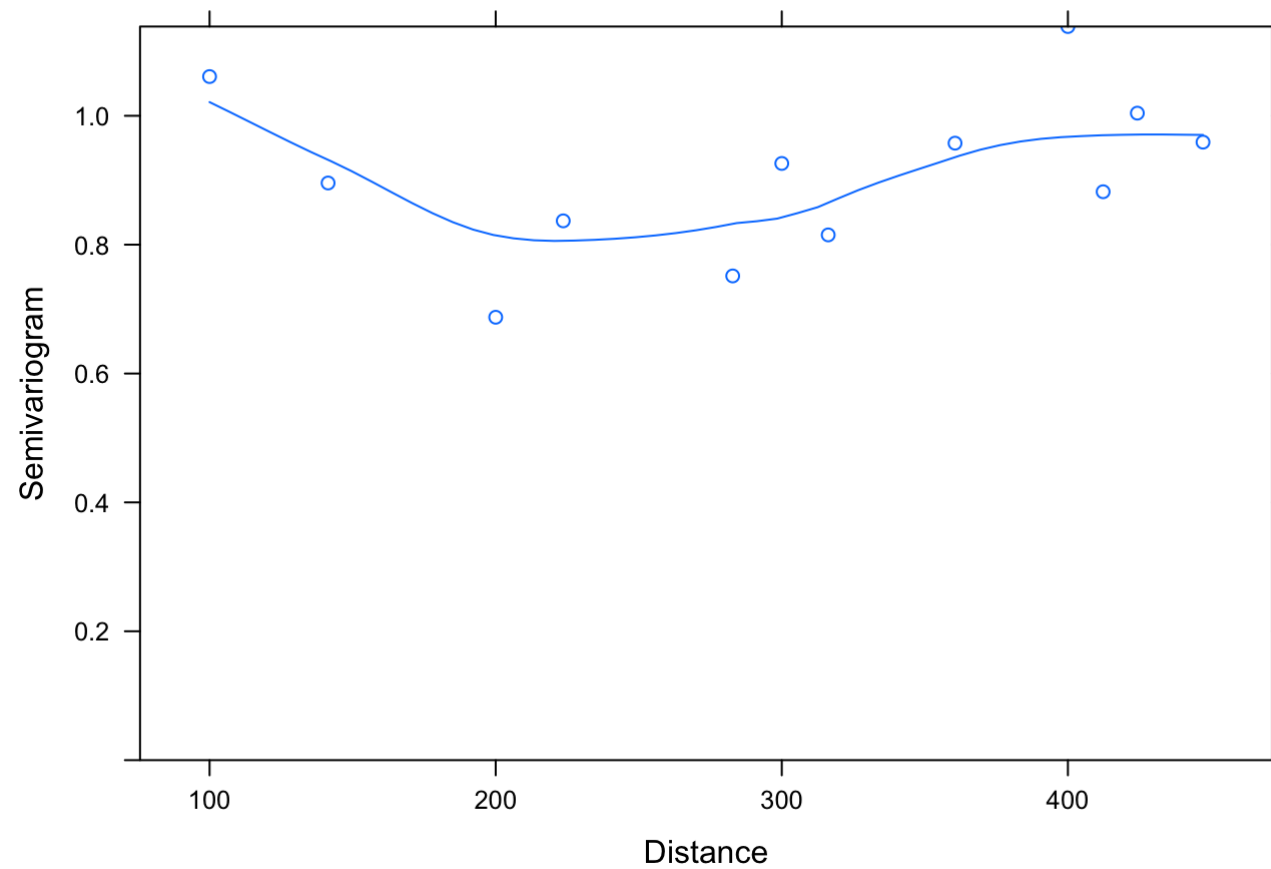
```
b_lin = update(lm_b, corr=corLin(form=~x + y))

plot(Variogram(b_lin, maxDist = maxdist))
```





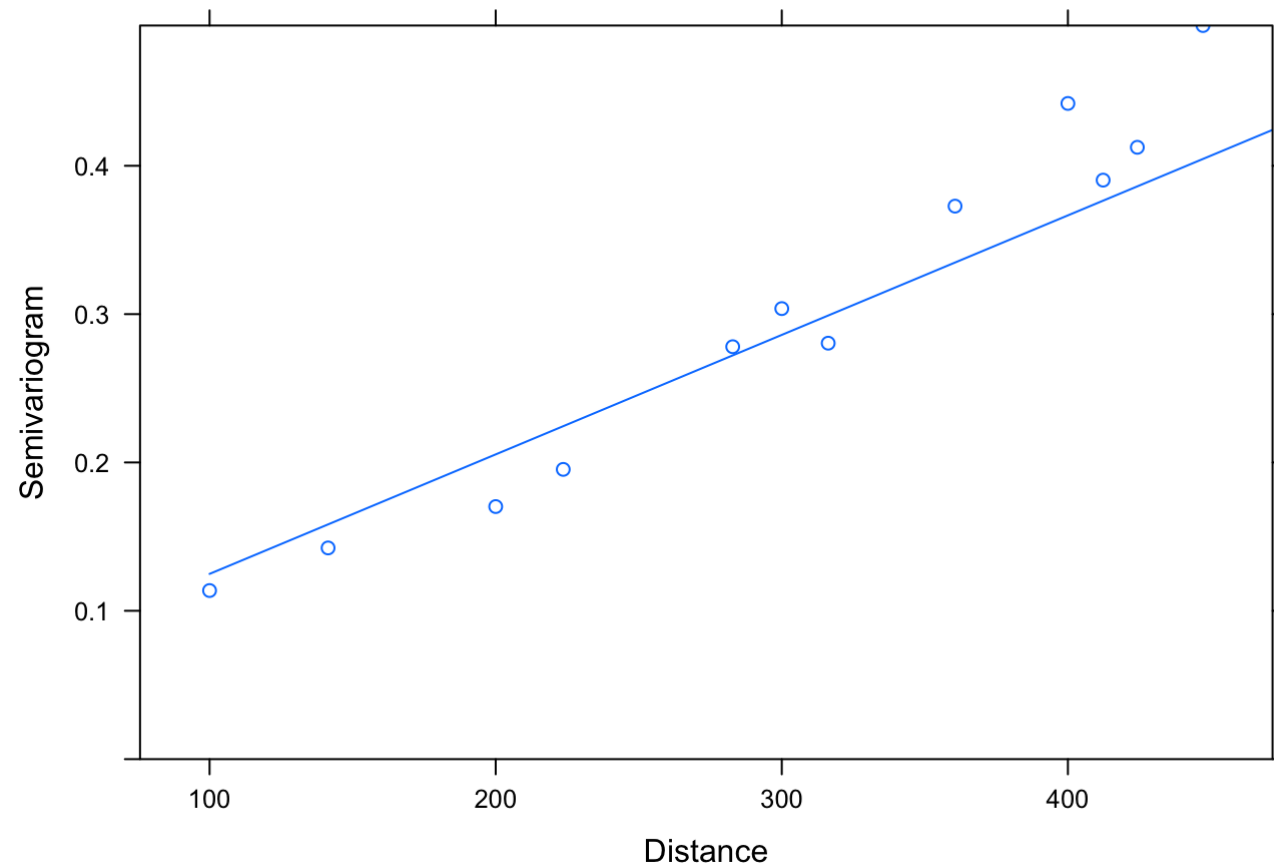
```
plot(Variogram(b_lin, resType='normalized', maxDist = maxdist))
```



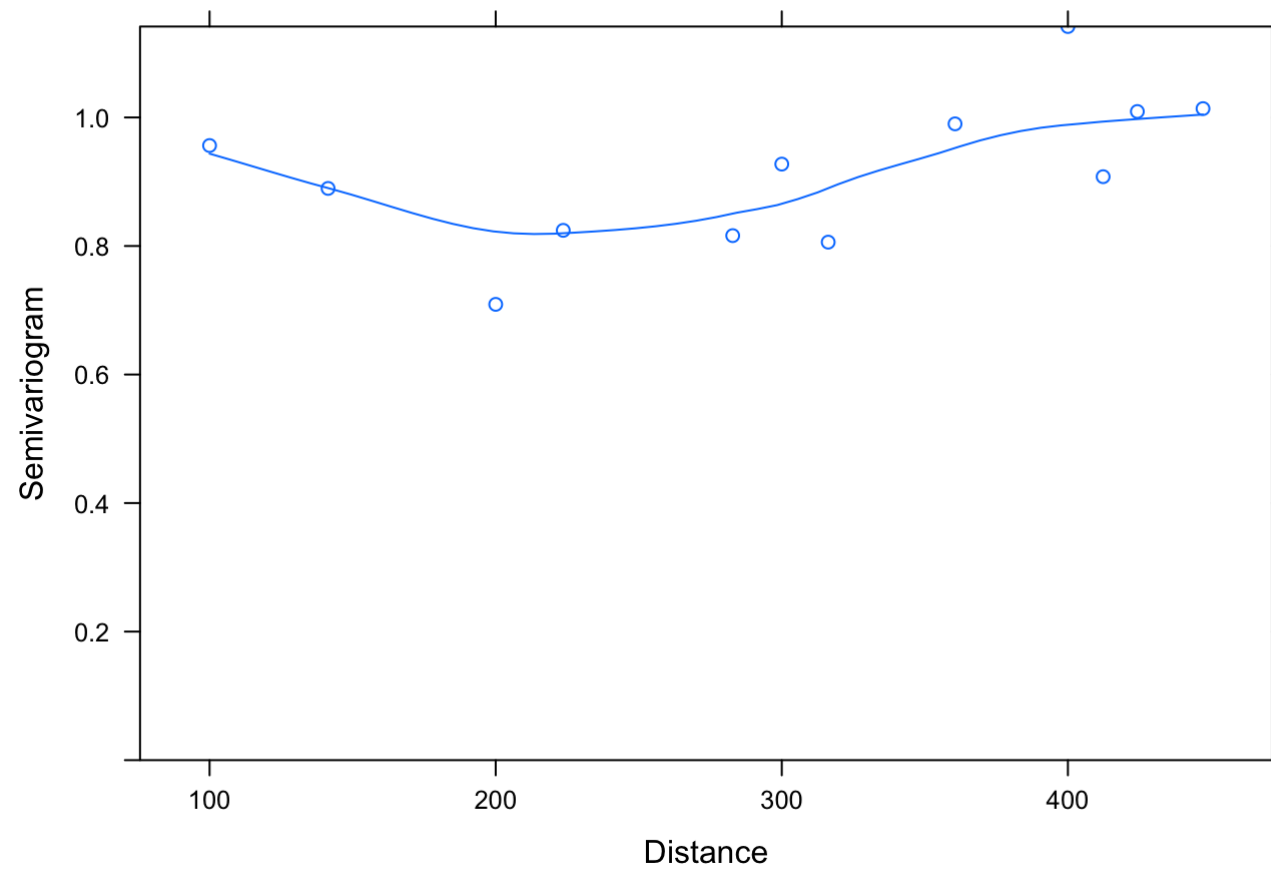
Linear model with nugget (non-zero

y-intercept)

```
b_lin_nug = update(lm_b, corr=corLin(form=~x + y, nugget = T))  
plot(Variogram(b_lin_nug, maxDist = maxdist))
```

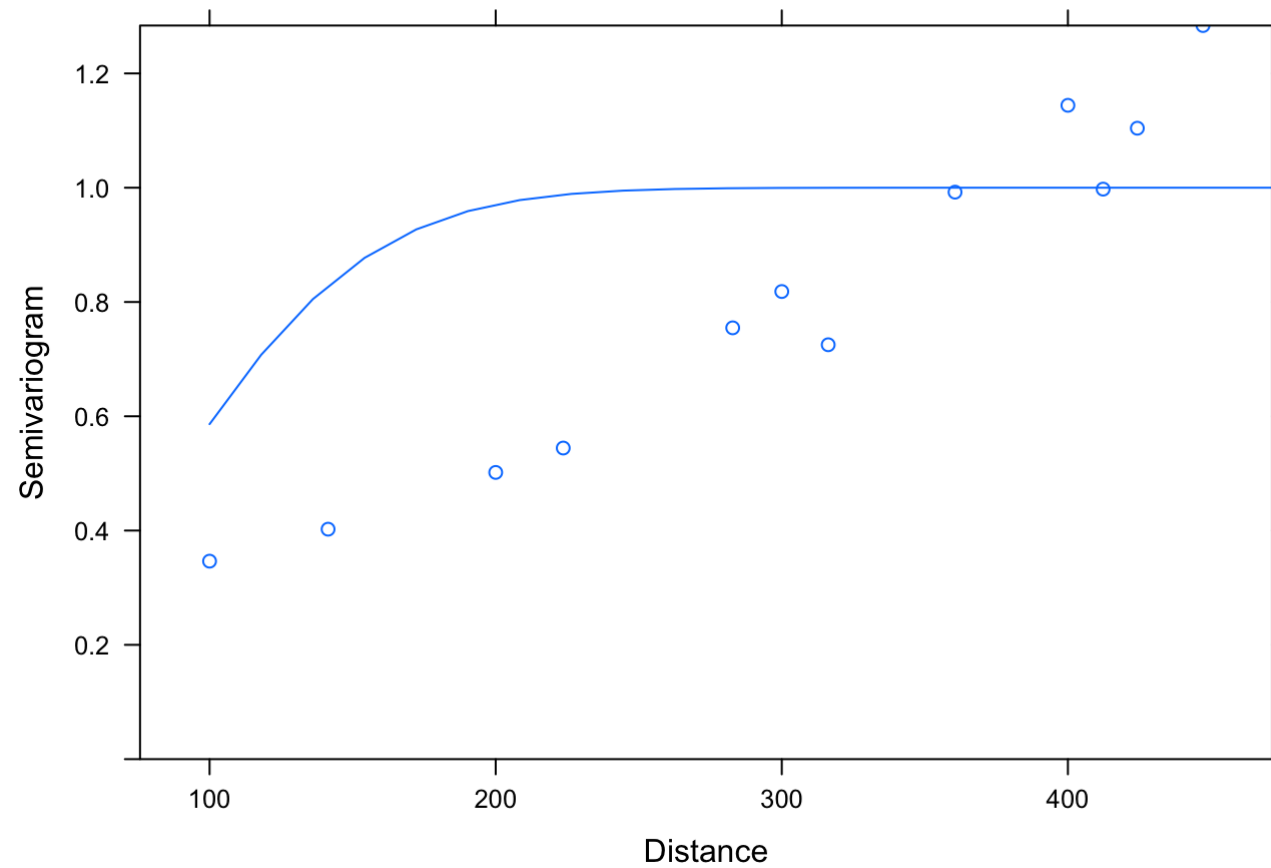


```
plot(Variogram(b_lin_nug, resType='normalized', maxDist = maxdist))
```

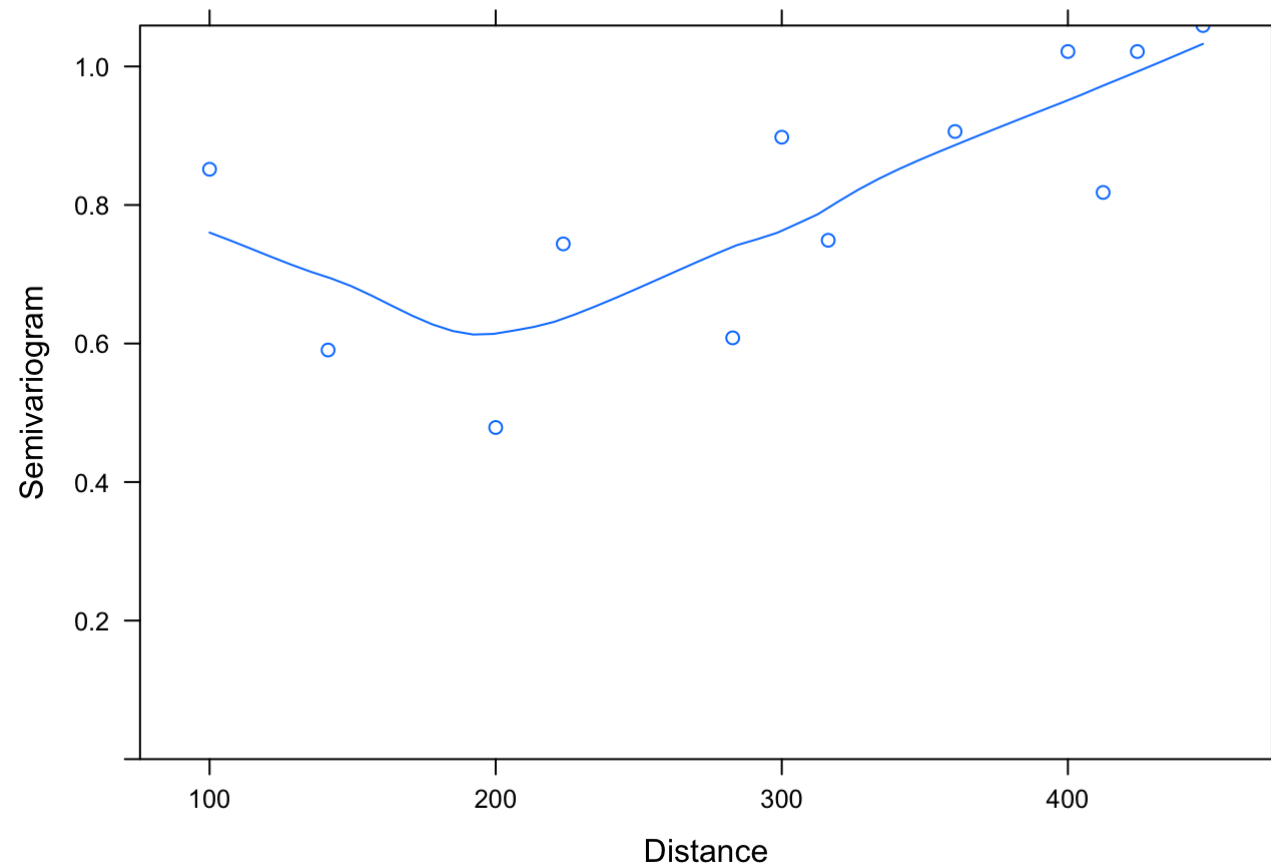


Gaussian Model -

```
b_gaus = update(lm_b, corr=corGaus(form=~x + y))  
  
plot(Variogram(b_gaus, maxDist = maxdist))
```

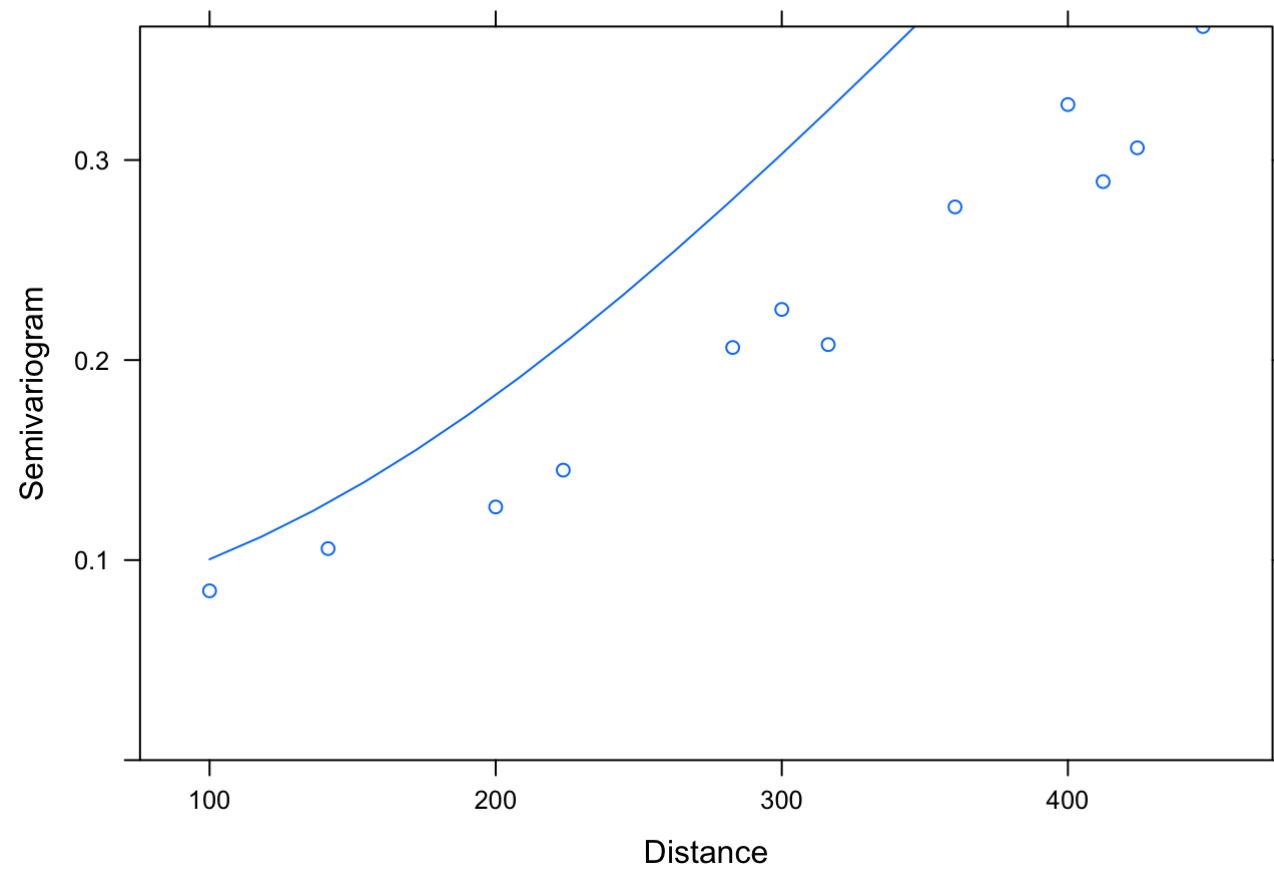


```
plot(Variogram(b_gaus, resType='normalized', maxDist = maxdist))
```

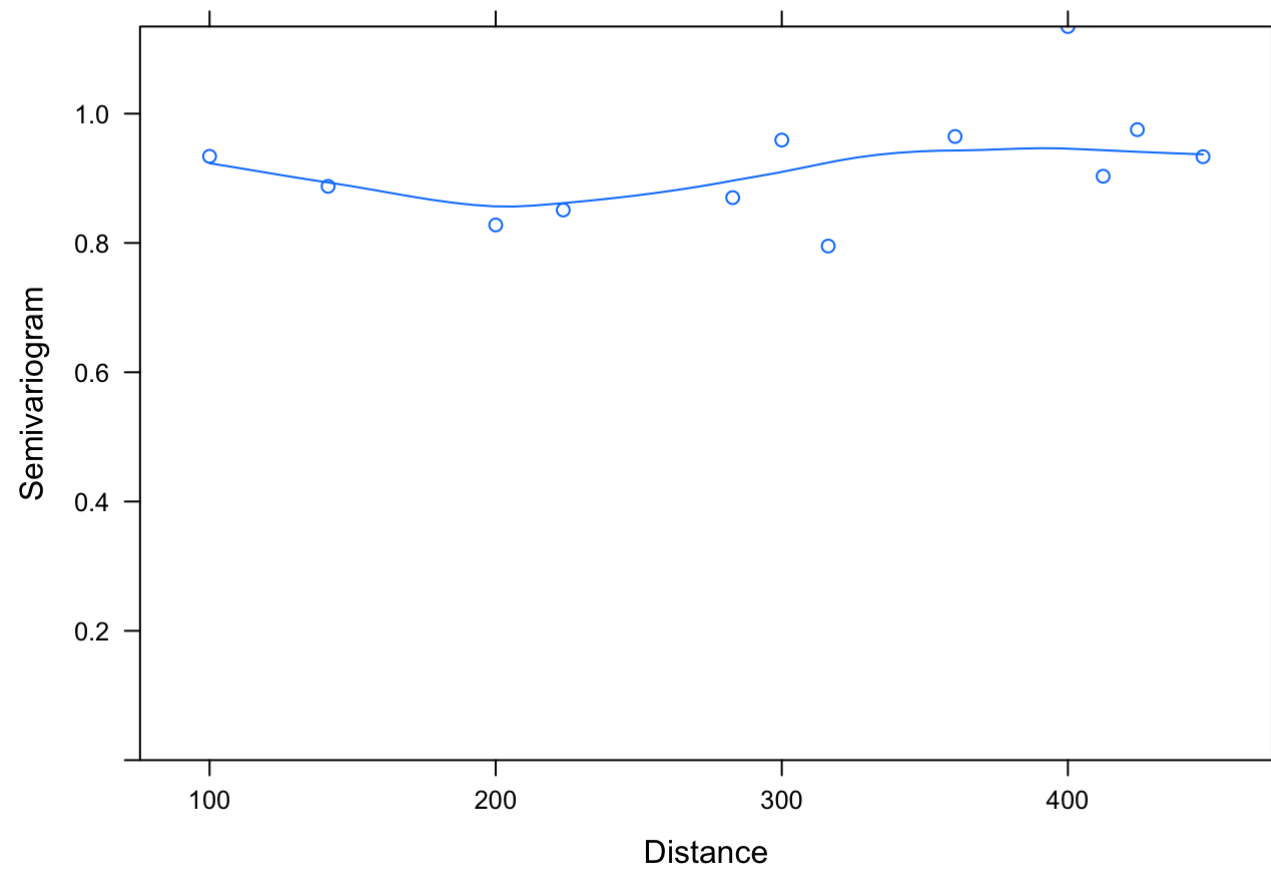


```
b_gaus_nug = update(lm_b, corr=corGaus(form=~x + y, nugget = T))
```

```
plot(Variogram(b_gaus_nug, maxDist = maxdist))
```



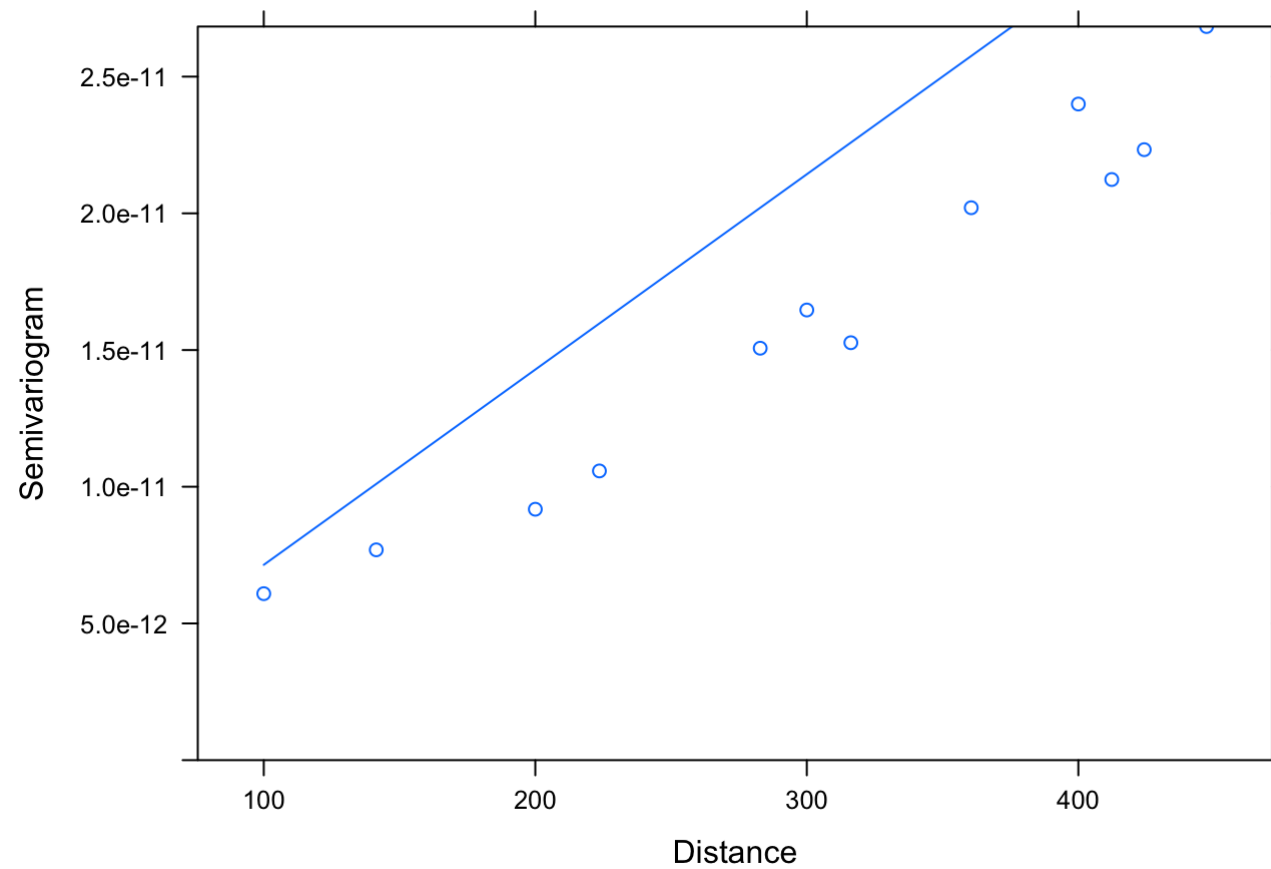
```
plot(Variogram(b_gaus_nug, resType='normalized', maxDist = maxdist))
```



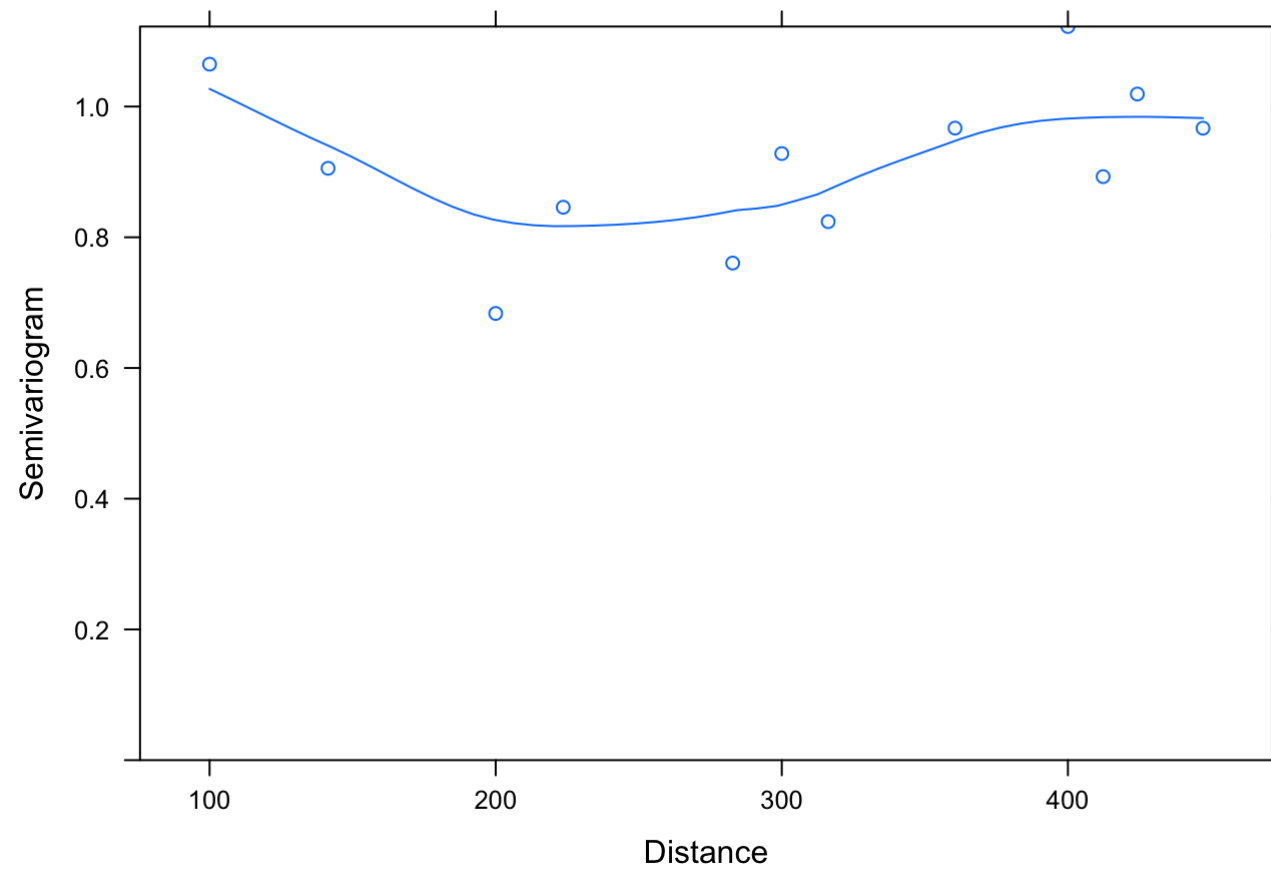
Exponential Model -

```
b_exp = update(lm_b, corr=corExp(form=~x + y))  
plot(Variogram(b_exp, maxDist = maxdist))
```

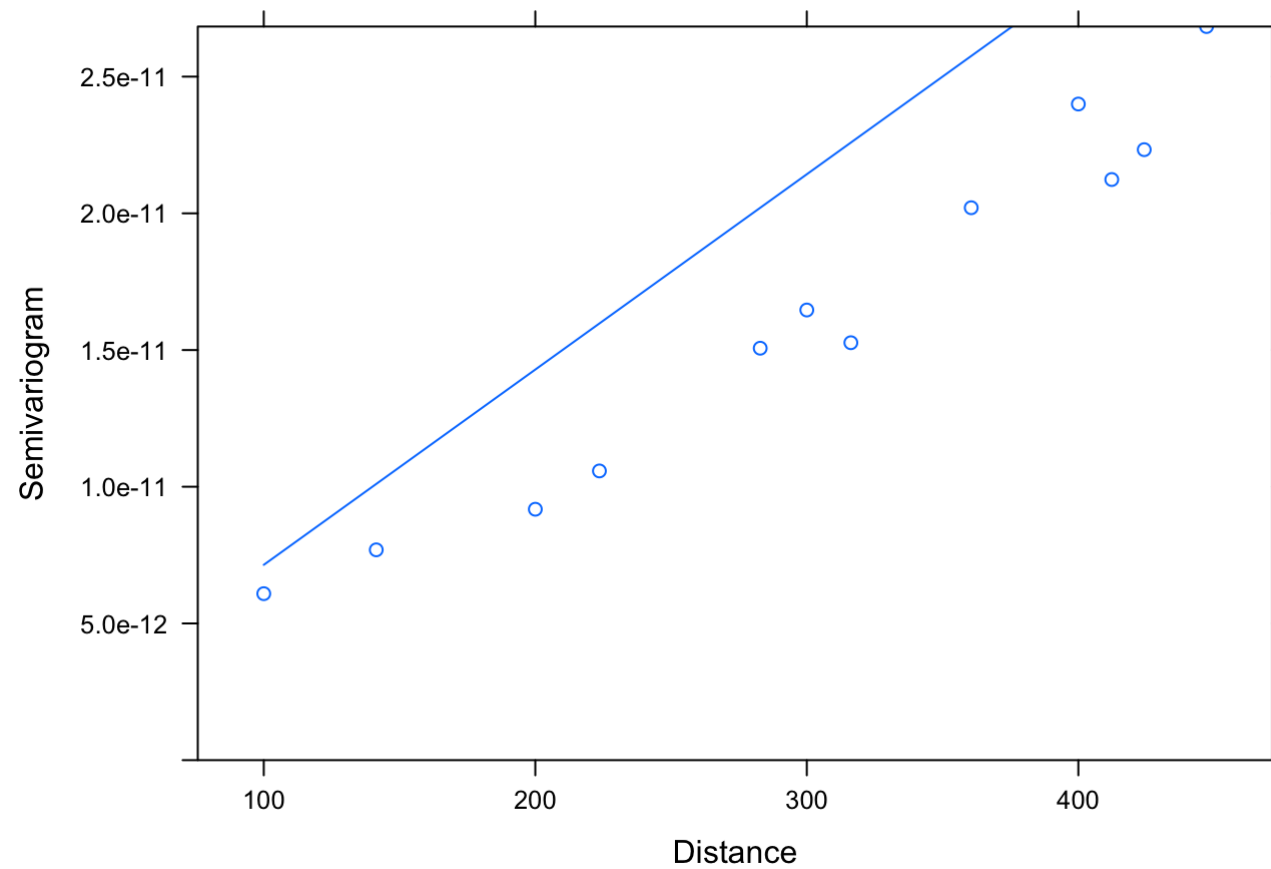




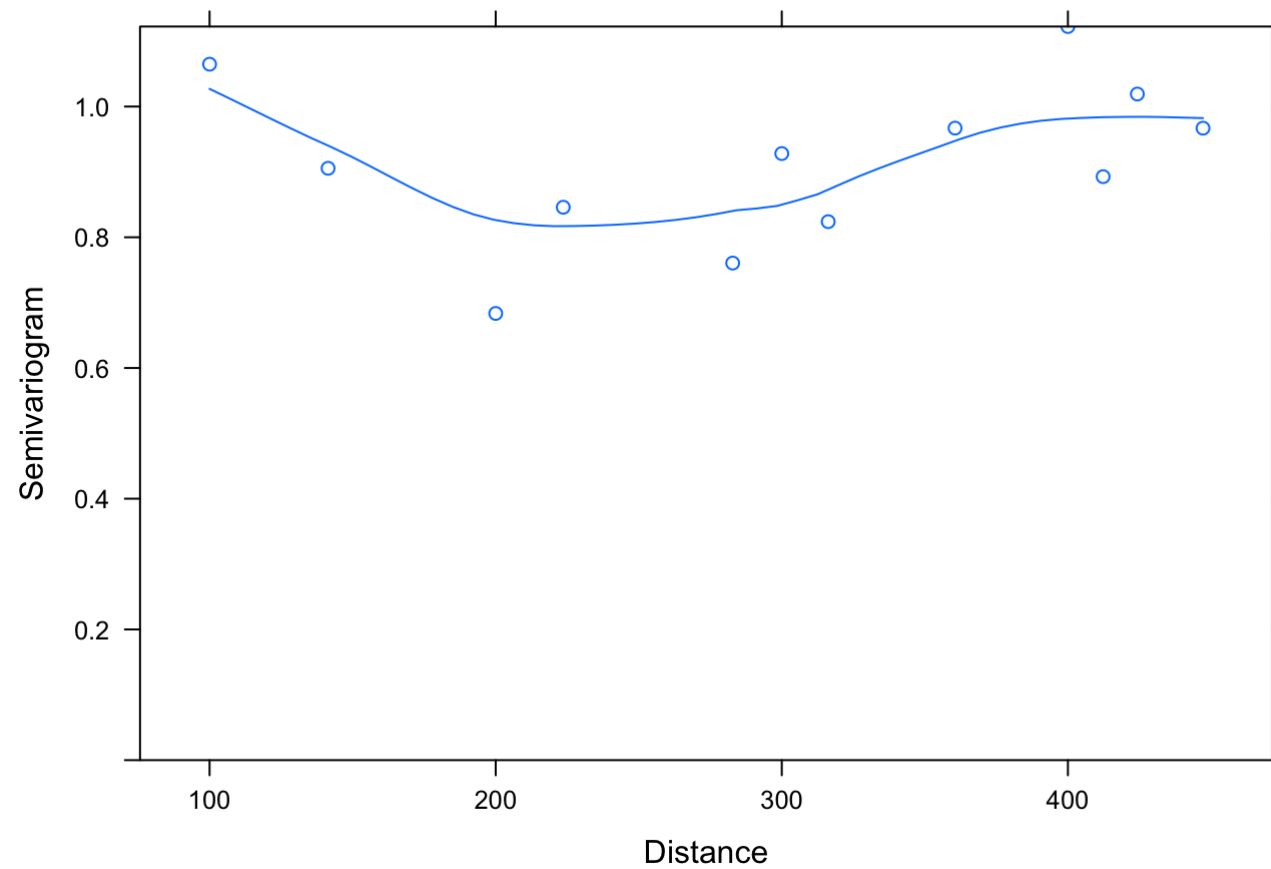
```
plot(Variogram(b_exp, resType='normalized', maxDist = maxdist))
```



```
b_exp_nug = update(lm_b, corr=corExp(form=~x + y))  
plot(Variogram(b_exp_nug, maxDist = maxdist))
```

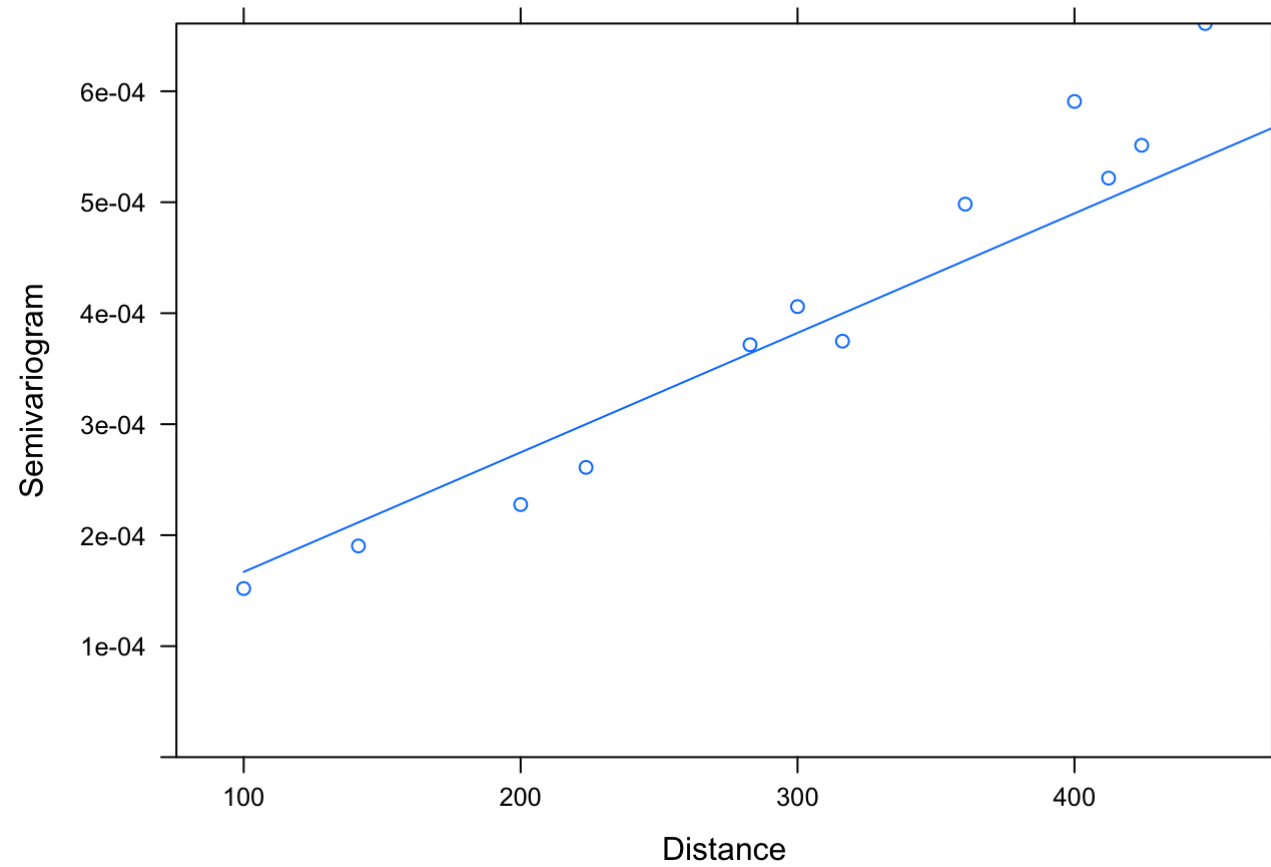


```
plot(Variogram(b_exp_nug, resType='normalized', maxDist = maxdist))
```

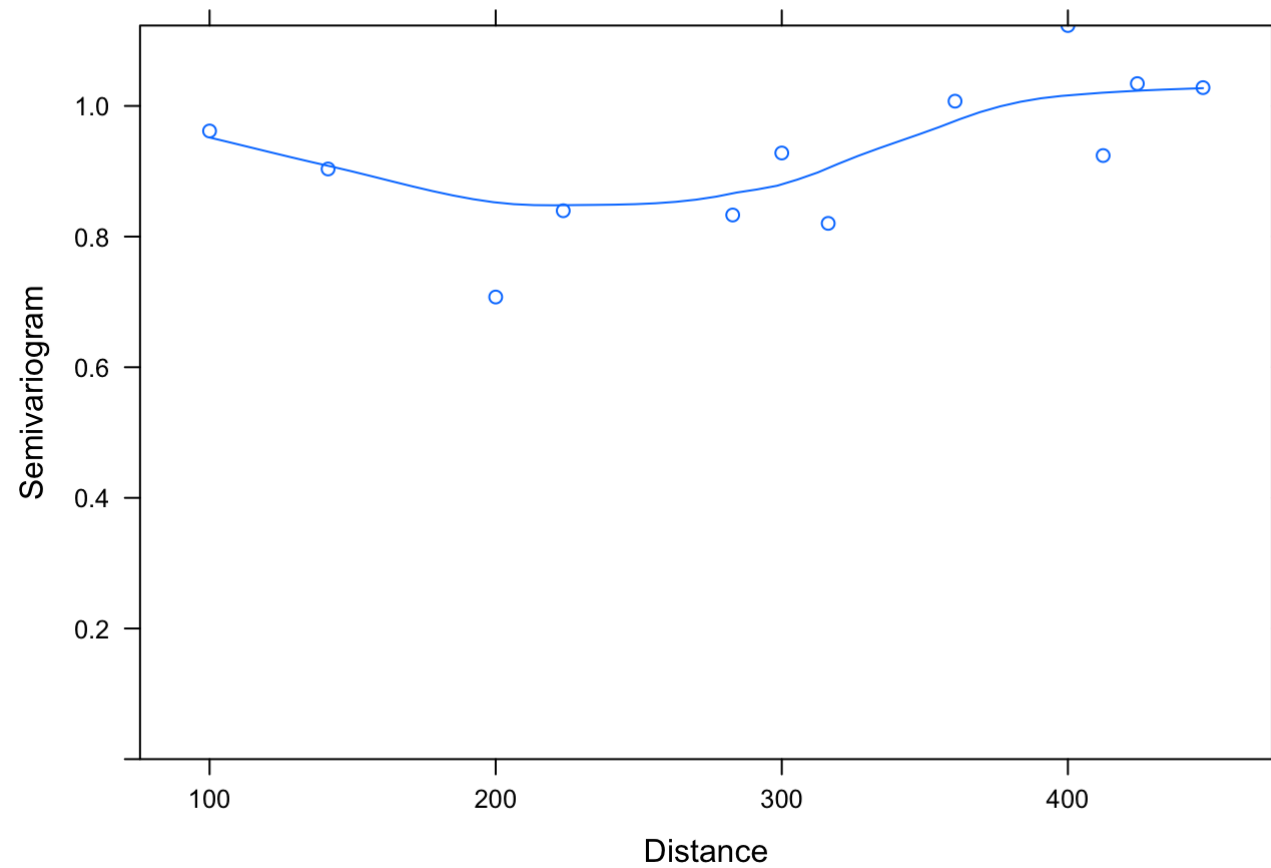


Spherical Model -

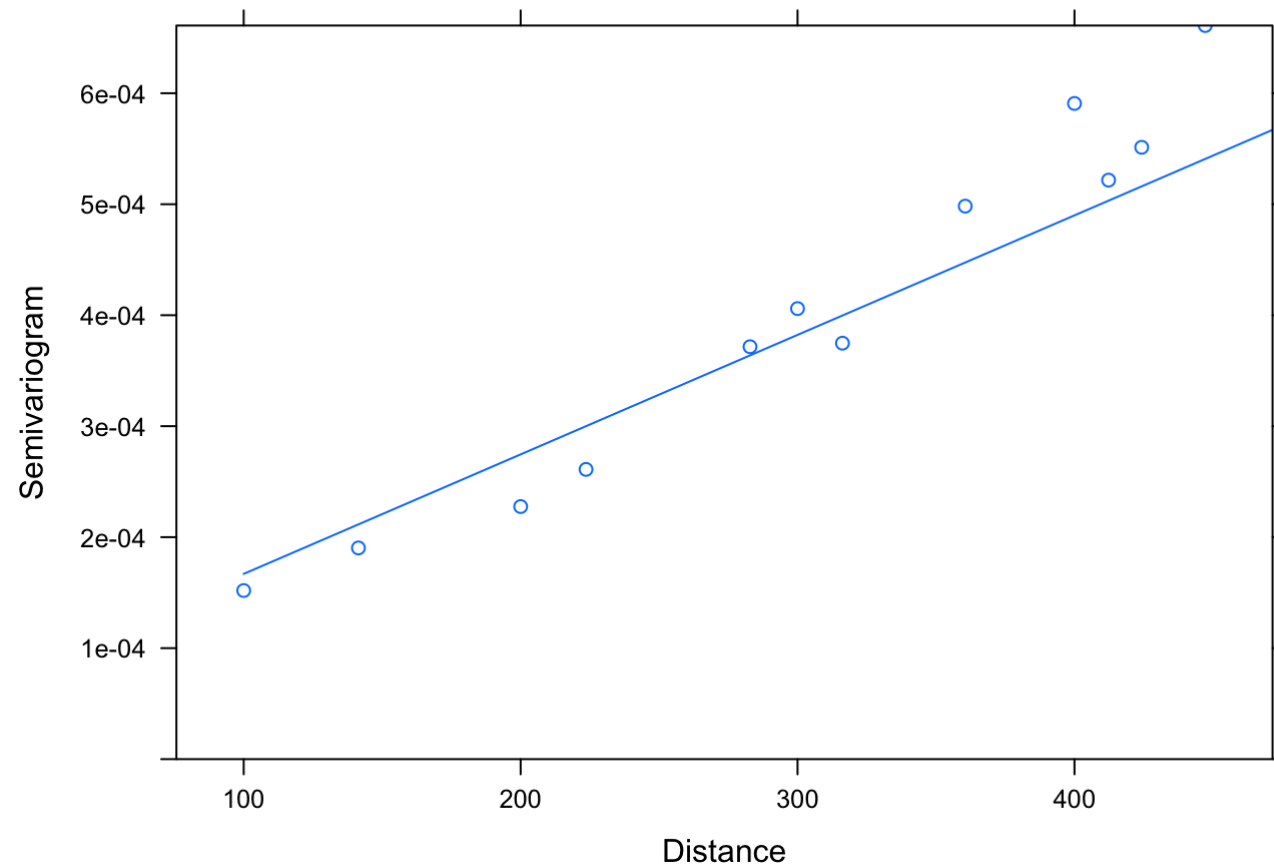
```
b_sph = update(lm_b, corr=corSpher(form=~x + y, nugget = T))  
plot(Variogram(b_sph, maxDist = maxdist))
```



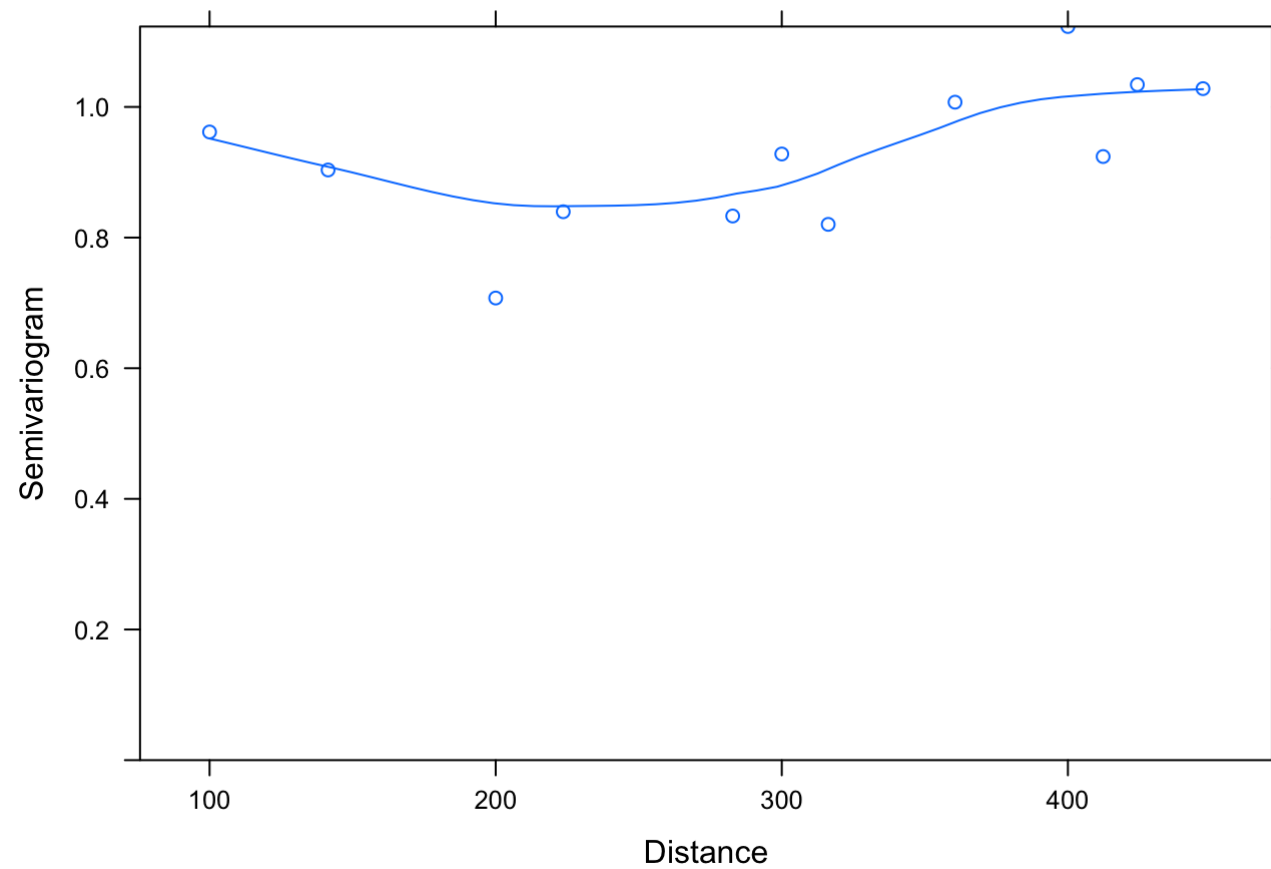
```
plot(Variogram(b_sph, resType = "normalized", maxDist = maxdist))
```



```
b_sph_nug = update(lm_b, corr=corSpher(form=~x + y, nugget = T))  
plot(Variogram(b_sph_nug, maxDist = maxdist))
```



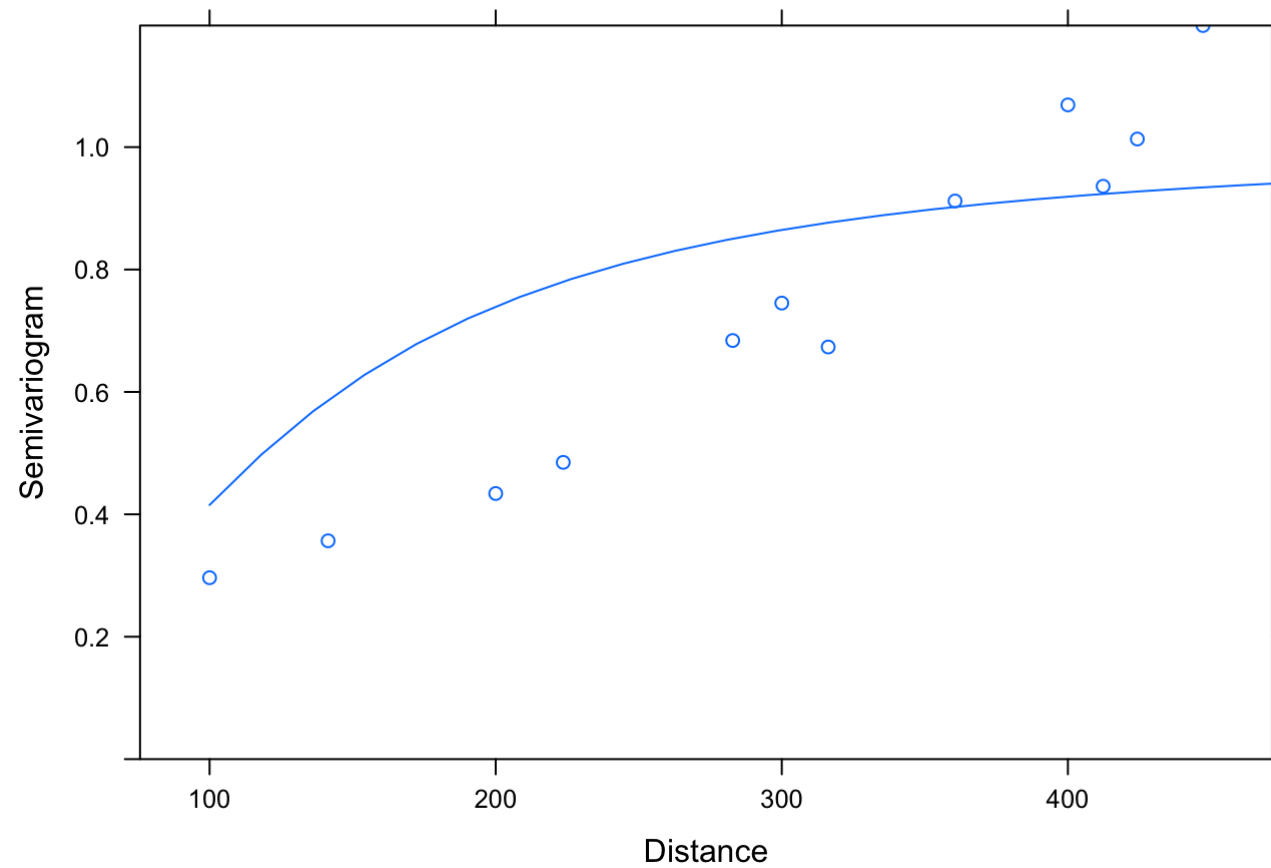
```
plot(Variogram(b_sph_nug, resType = "normalized", maxDist = maxdist))
```



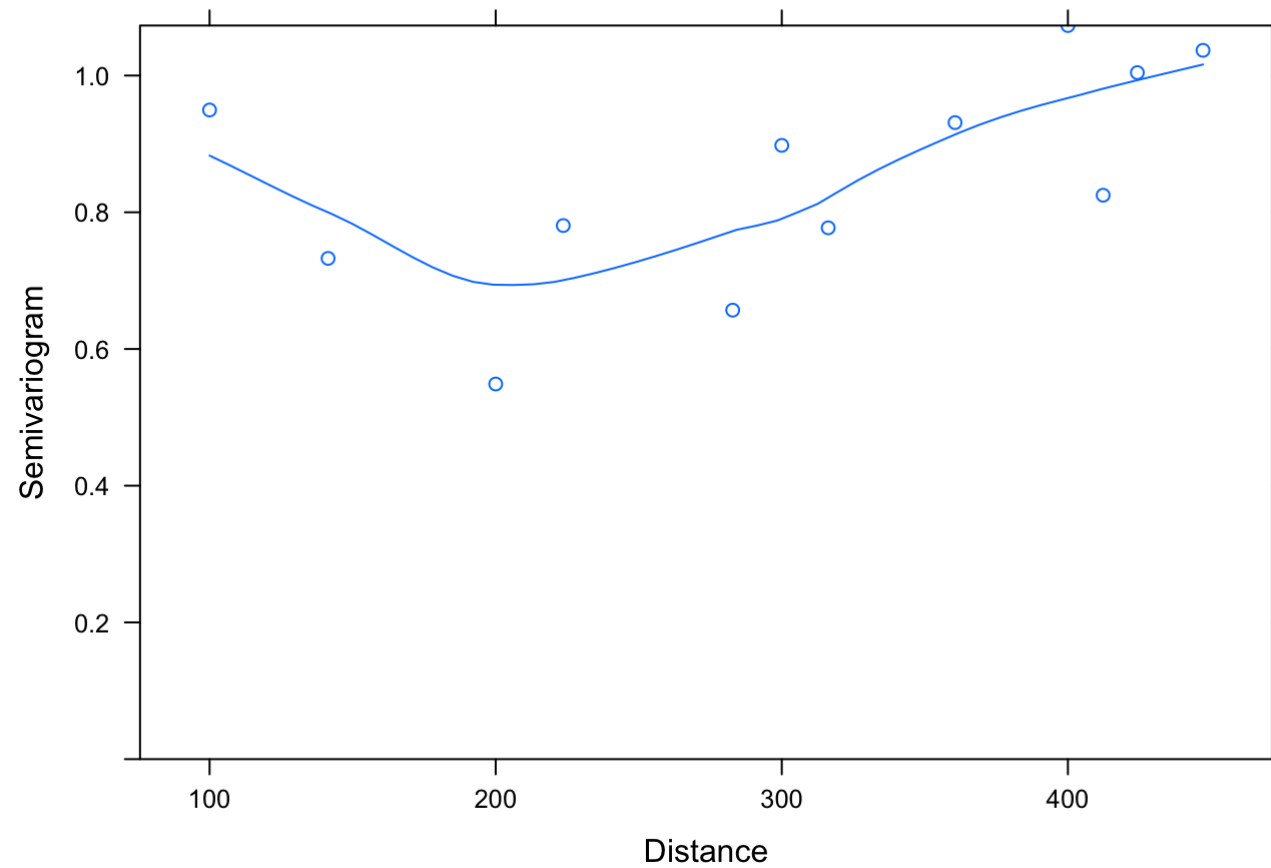
Rational Quadratic Model -

```
b_rat = update(lm_b, corr=corRatio(form=~x + y))  
plot(Variogram(b_rat, maxDist = maxdist))
```

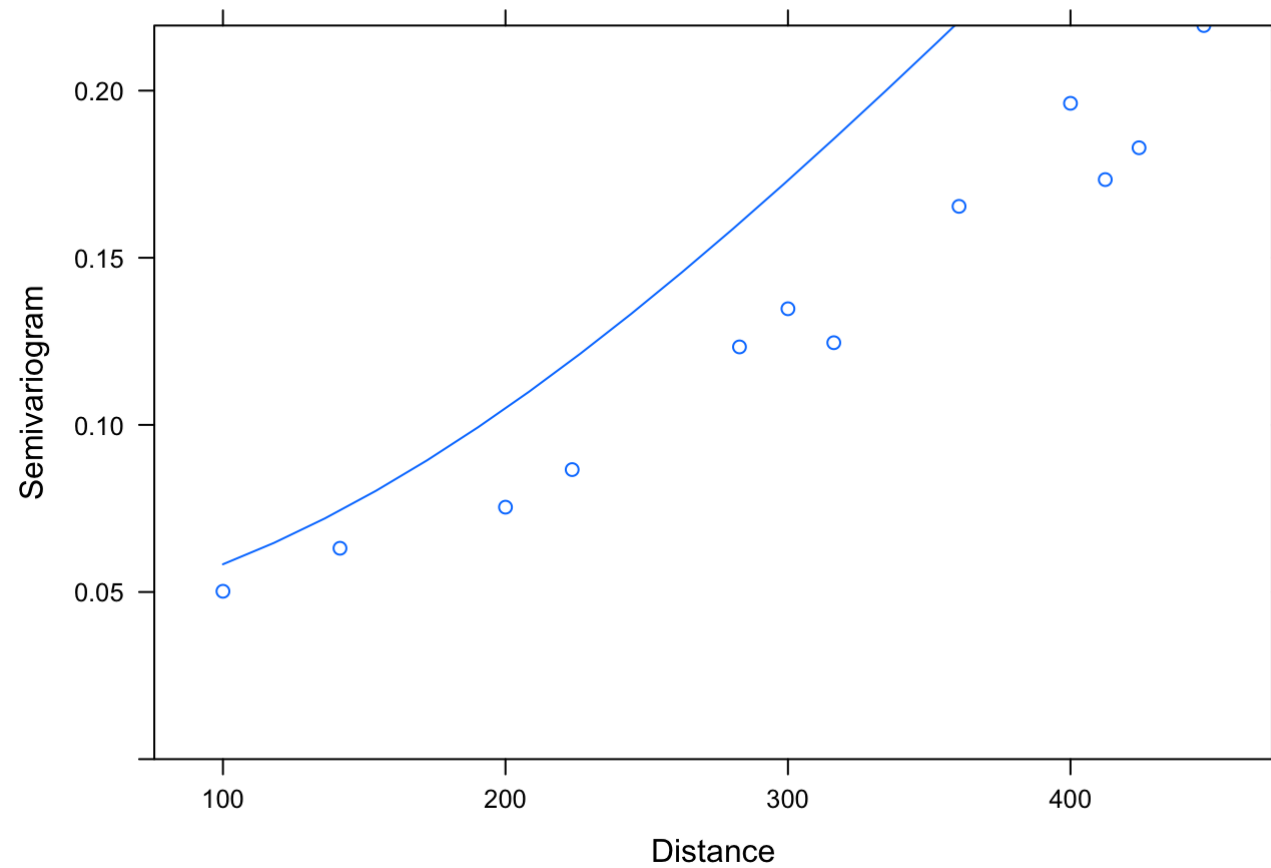




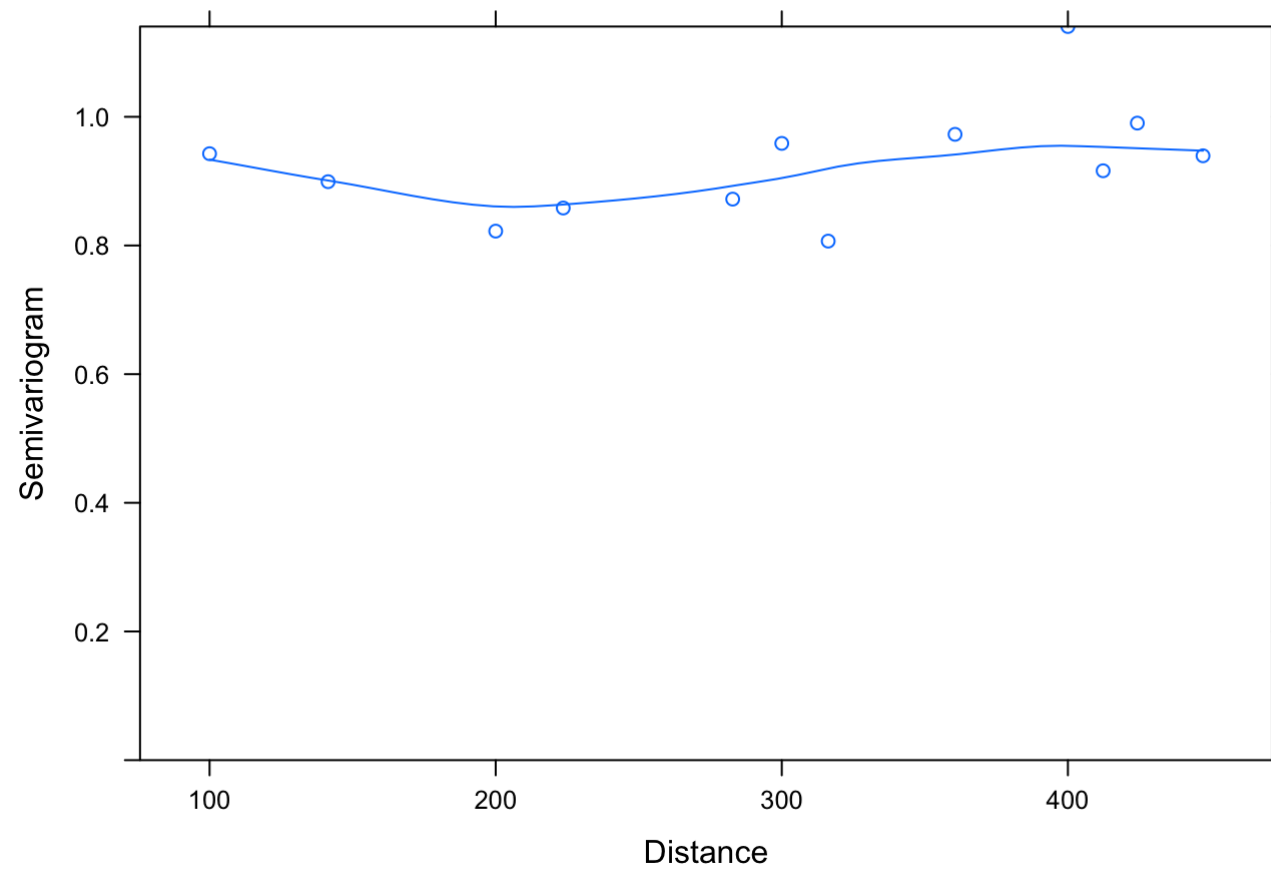
```
plot(Variogram(b_rat, resType = "normalized", maxDist = maxdist))
```



```
b_rat_nug = update(lm_b, corr=corRatio(form=~x + y, nugget = T))  
plot(Variogram(b_rat_nug, maxDist = maxdist))
```



```
plot(Variogram(b_rat_nug, resType = "normalized", maxDist = maxdist))
```



Comparing models -

```
anova(lm_b, b_lin, b_lin_nug, b_gaus, b_gaus_nug, b_exp, b_exp_nug, b_sph, b_sph_nug, b_rat, b_rat_nug, test=F)
```

##	Model	df	AIC	BIC	logLik	
##	lm_b	1	3	346.1696	351.7832	-170.0848
##	b_lin	2	4	312.5378	320.0226	-152.2689
##	b_lin_nug	3	5	313.2119	322.5679	-151.6059
##	b_gaus	4	4	329.3951	336.8799	-160.6975
##	b_gaus_nug	5	5	310.8334	320.1894	-150.4167
##	b_exp	6	4	312.5365	320.0213	-152.2682
##	b_exp_nug	7	4	312.5365	320.0213	-152.2682
##	b_sph	8	5	313.2119	322.5679	-151.6059
##	b_sph_nug	9	5	313.2119	322.5679	-151.6059
##	b_rat	10	4	321.3111	328.7959	-156.6555
##	b_rat_nug	11	5	310.6244	319.9804	-150.3122

From comparing here looking at the AIC values for each model, we can see that the gaussian with nugget and rational quadratic with nugget spatial correlation structures fit the model best.

Based on the graphs for each semivariogram model and their residual distributions, I had expected the spherical model with a nugget to be the best fit. The regression line appeared to fit best and the residuals showed less of a trend compared to the gaussian and rational quadratic models with a nugget.

```
summary(b_gaus)
```

```
## Generalized least squares fit by REML
## Model: sp_ref ~ sp_b
## Data: bci_dat
##      AIC      BIC    logLik
## 329.3951 336.8799 -160.6975
##
## Correlation Structure: Gaussian spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##   range
## 106.4515
##
## Coefficients:
##              Value Std.Error  t-value p-value
## (Intercept) 3.739614 2.4387041 1.533443  0.1317
## sp_b         0.189157 0.1151048 1.643346  0.1068
##
## Correlation:
##      (Intr)
## sp_b -0.684
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.98157029 -0.68280382 -0.39126760  0.07706608  4.14244538
##
## Residual standard error: 7.416059
## Degrees of freedom: 50 total; 48 residual
```

```
summary(b_gaus_nug)
```

```
## Generalized least squares fit by REML
## Model: sp_ref ~ sp_b
## Data: bci_dat
##      AIC      BIC    logLik
## 310.8334 320.1894 -150.4167
##
## Correlation Structure: Gaussian spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##      range      nugget
## 559.91373480  0.07126166
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) 12.35980  9.183772  1.3458304  0.1847
## sp_b         0.02342  0.085167  0.2749965  0.7845
##
## Correlation:
##      (Intr)
## sp_b -0.239
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.8606395 -0.8431963 -0.7085738 -0.3438338  1.7656880
##
## Residual standard error: 14.76937
## Degrees of freedom: 50 total; 48 residual
```

```
summary(b_rat)
```

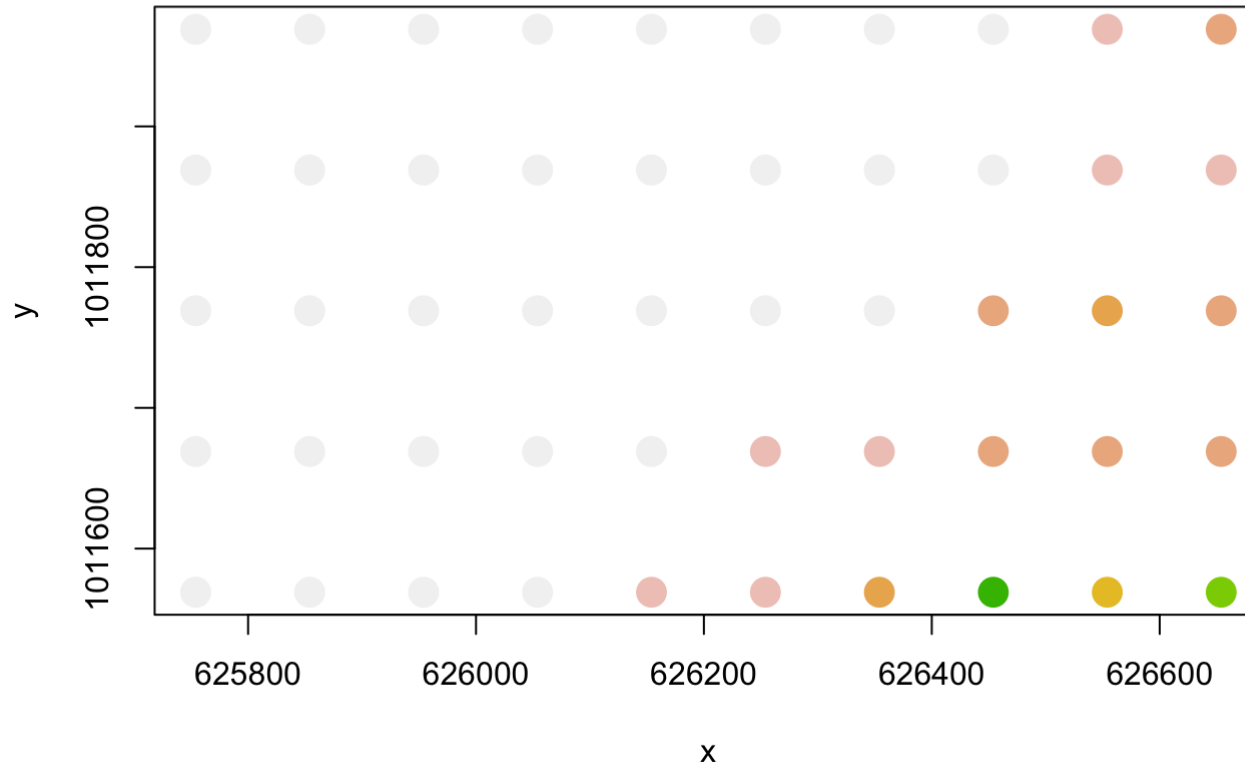
```
## Generalized least squares fit by REML
## Model: sp_ref ~ sp_b
## Data: bci_dat
##      AIC      BIC    logLik
## 321.3111 328.7959 -156.6555
##
## Correlation Structure: Rational quadratic spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##   range
## 118.6667
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) 5.859768  3.536752  1.6568218  0.1041
## sp_b         0.105043  0.112266  0.9356613  0.3541
##
## Correlation:
##      (Intr)
## sp_b -0.505
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.9417905 -0.8220455 -0.5620671  0.0116406  3.8783195
##
## Residual standard error: 7.894969
## Degrees of freedom: 50 total; 48 residual
```

```
summary(b_rat_nug)
```



```
## Generalized least squares fit by REML
## Model: sp_ref ~ sp_b
## Data: bci_dat
##      AIC      BIC    logLik
## 310.6244 319.9804 -150.3122
##
## Correlation Structure: Rational quadratic spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##      range      nugget
## 752.28335795  0.04166127
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) 14.088134 14.874137  0.9471564  0.3483
## sp_b         0.011441  0.087993  0.1300210  0.8971
##
## Correlation:
##      (Intr)
## sp_b -0.167
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.7428201 -0.7362643 -0.6323781 -0.3420863  1.2834076
##
## Residual standard error: 19.19677
## Degrees of freedom: 50 total; 48 residual
```

```
col_brks = hist(residuals(b_rat_nug), plot=F)$breaks
col_indices = as.numeric(cut(residuals(b_rat_nug), col_brks))
cols = rev(terrain.colors(length(col_brks)))
plot(BCI_xy, cex=2, pch=19, col=cols[col_indices])
```



*#how can I add a legend here to represent the dot colors in the graph?*

Both *H. trianda*, the predictor species, and *D. standleyi*, the response species, are displaying spatial auto-correlation. However, it appears *H. trianda* is not a good predictor of the occurrence and abundance of *D. standleyi* within the Barro Colorado Island.

The addition of the spatial error term (nugget =  $T$ ) to correct for the non-zero y-intercept in the models fit much better compared to the models lacking it. Focusing on the models with the best AIC scores compared to their counterpart models lacking the nugget, the regression coefficients and t statistics move towards lesser importance indicating this predictor species alone is not sufficient in understanding abundance and distribution of *D. standleyi*.

The reason we observe a difference when adding the nugget spatial error term is in the model assumes at zero separation distance the semivariogram value equals 0. However, at micro-scale separation distances the semivariogram will exhibit a nugget effect due to having this non-zero y-intercept. This may be due to measurement error and/or the tendency of natural phenomena to occur over a range of spatial scales, at

distances smaller than the sampling interval. Looking at the semivariogram graphs for each model we can see none start at zero and are all just a few points or more away, indicating a nugget may improve the fit of the models.

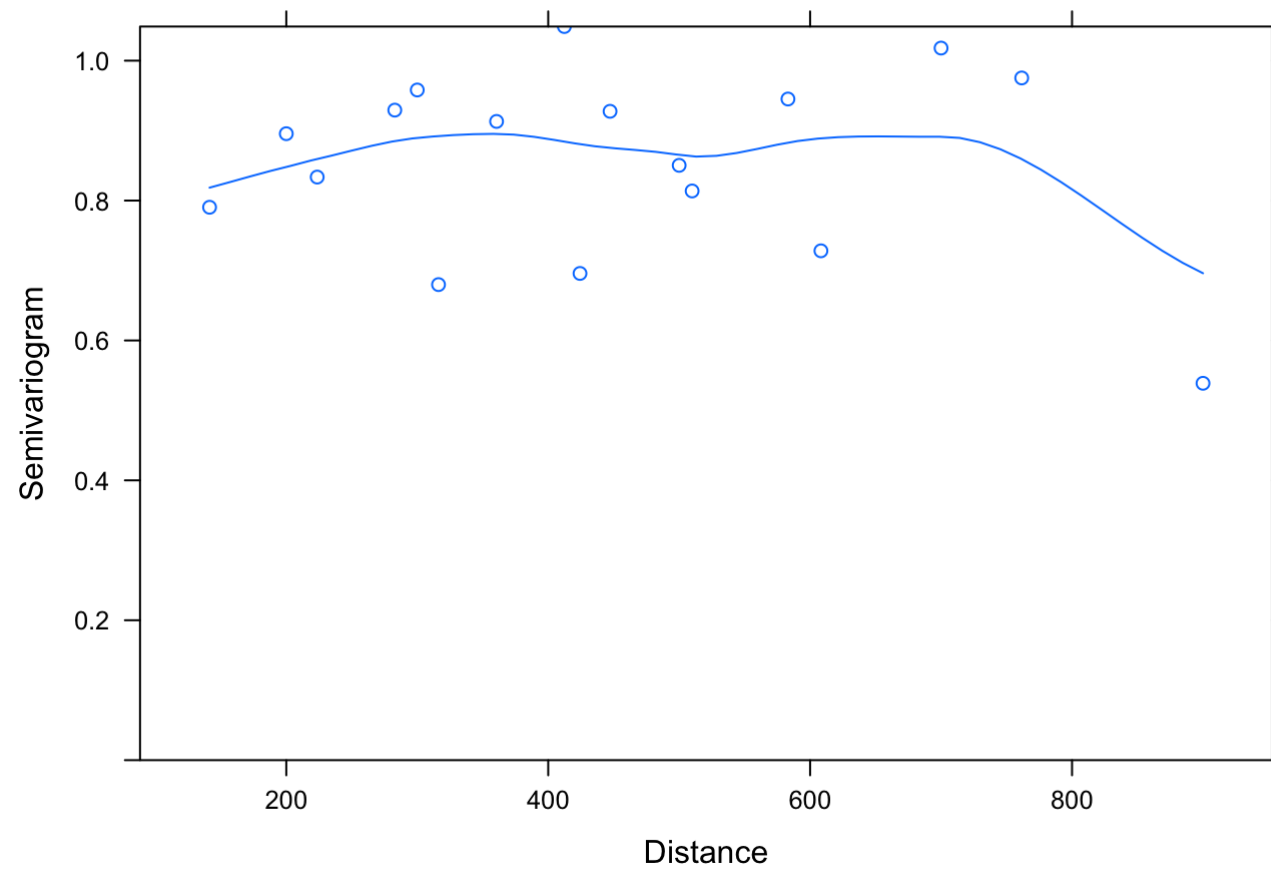
In the last graph showing a spatial map of the residual distribution of our model it appears they do look spatially structured, with clustering in the lower right portion of the area, and we may need to include some more species to better understand spatial patterning of *D. standleyi* in relation to these other co-inhabiting predictor species.

Model 2: include all of the species as predictor variables

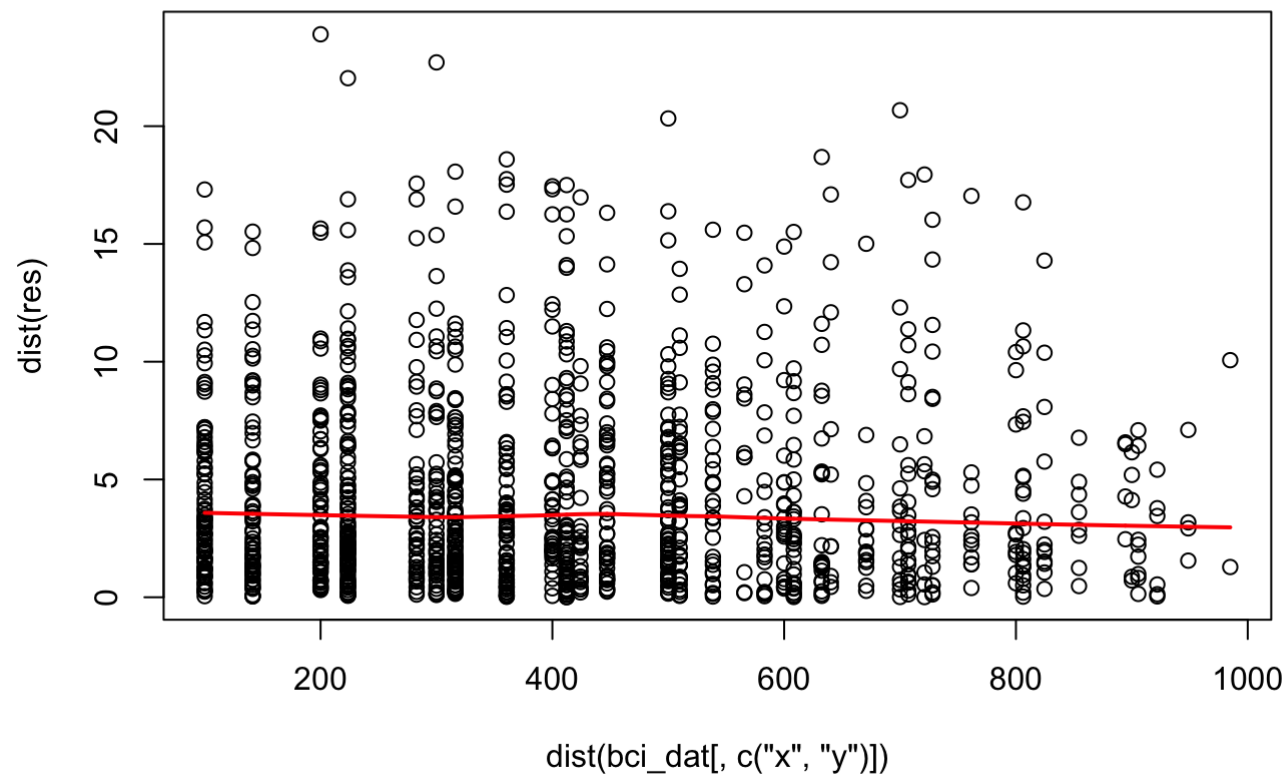
```
bci_dat = data.frame(BCI_xy, BCI)
maxdist <- max(BCI_xy) / 2

lm_all = gls(sp_ref ~ sp_a + sp_b + sp_c + sp_d + sp_e + sp_f + sp_g, data=bci_dat)
res = residuals(lm_all)

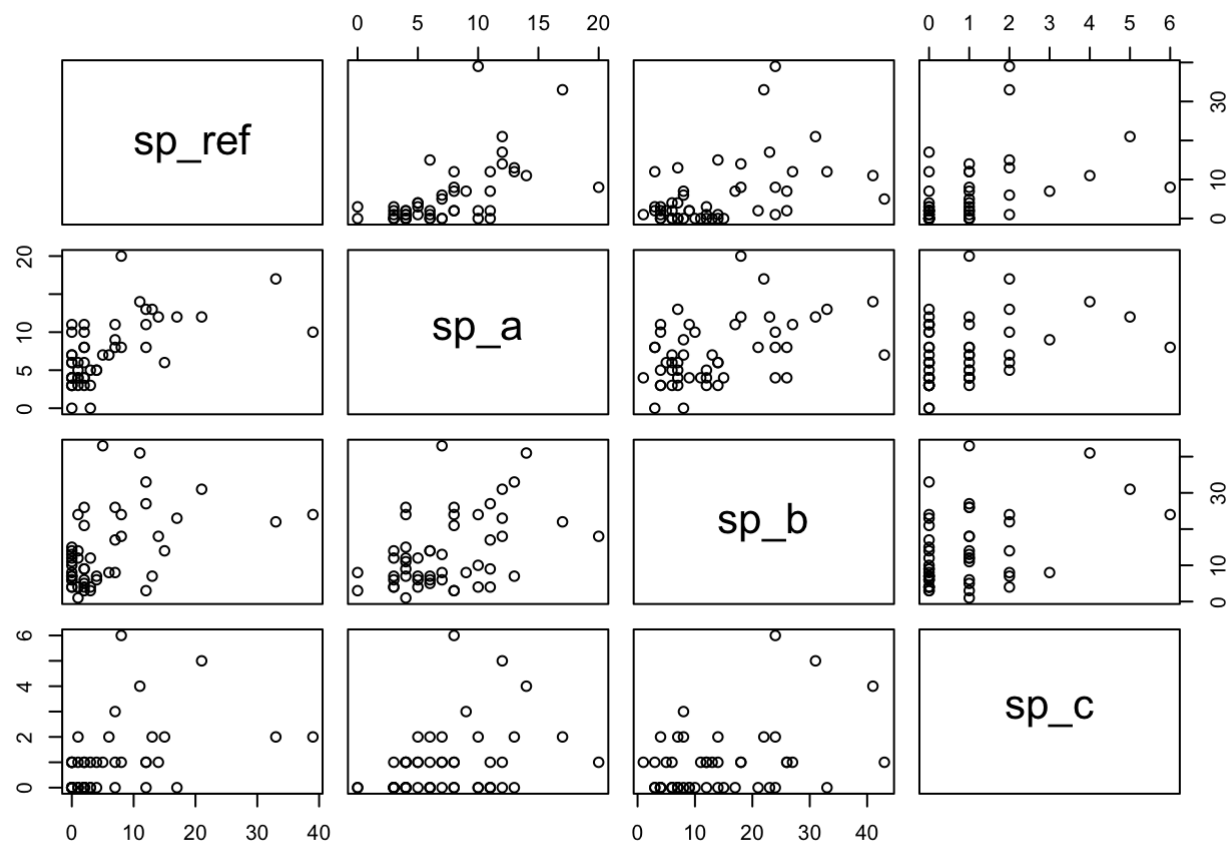
plot(Variogram(lm_all, form= ~ x + y))
```



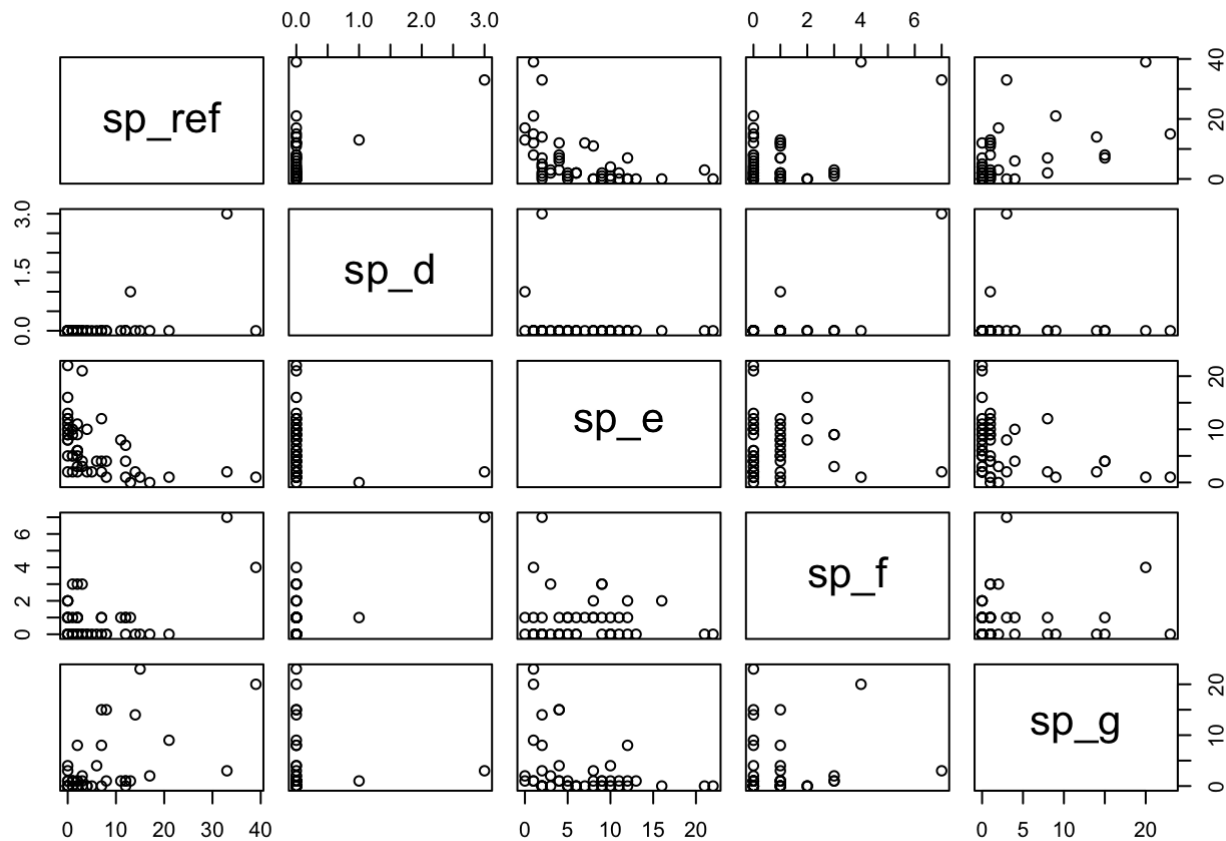
```
plot(dist(bci_dat[, c('x', 'y')]), dist(res))  
lines(lowess(dist(bci_dat[, c('x', 'y')]), dist(res)), col='red', lwd=2)  
abline(v = maxdist, col='red', lwd=3, lty=2)
```



```
pairs(data.frame(sp_ref, sp_a, sp_b, sp_c))
```



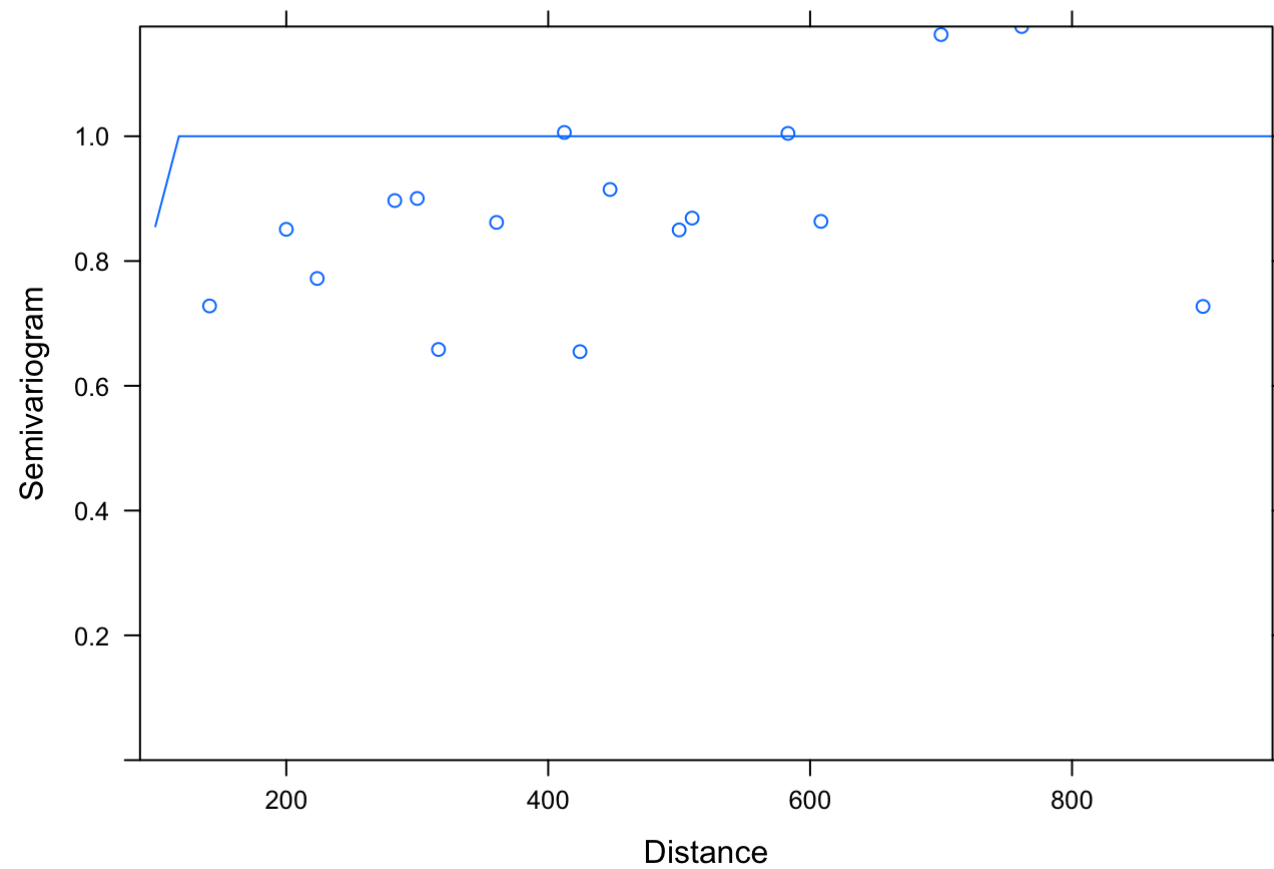
```
pairs(data.frame(sp_ref, sp_d, sp_e, sp_f, sp_g))
```



Comparing *D. standleyi* to the other predictor species with these pairs plots, the main pattern that jumps out is some species follow more clustering patterns in relation to *D. standleyi* abundance while others display high counts of 0 - 2 for the predictor species compared to our reference species creating these gradient looking lines on the pairs plots contrasting species' abundances.

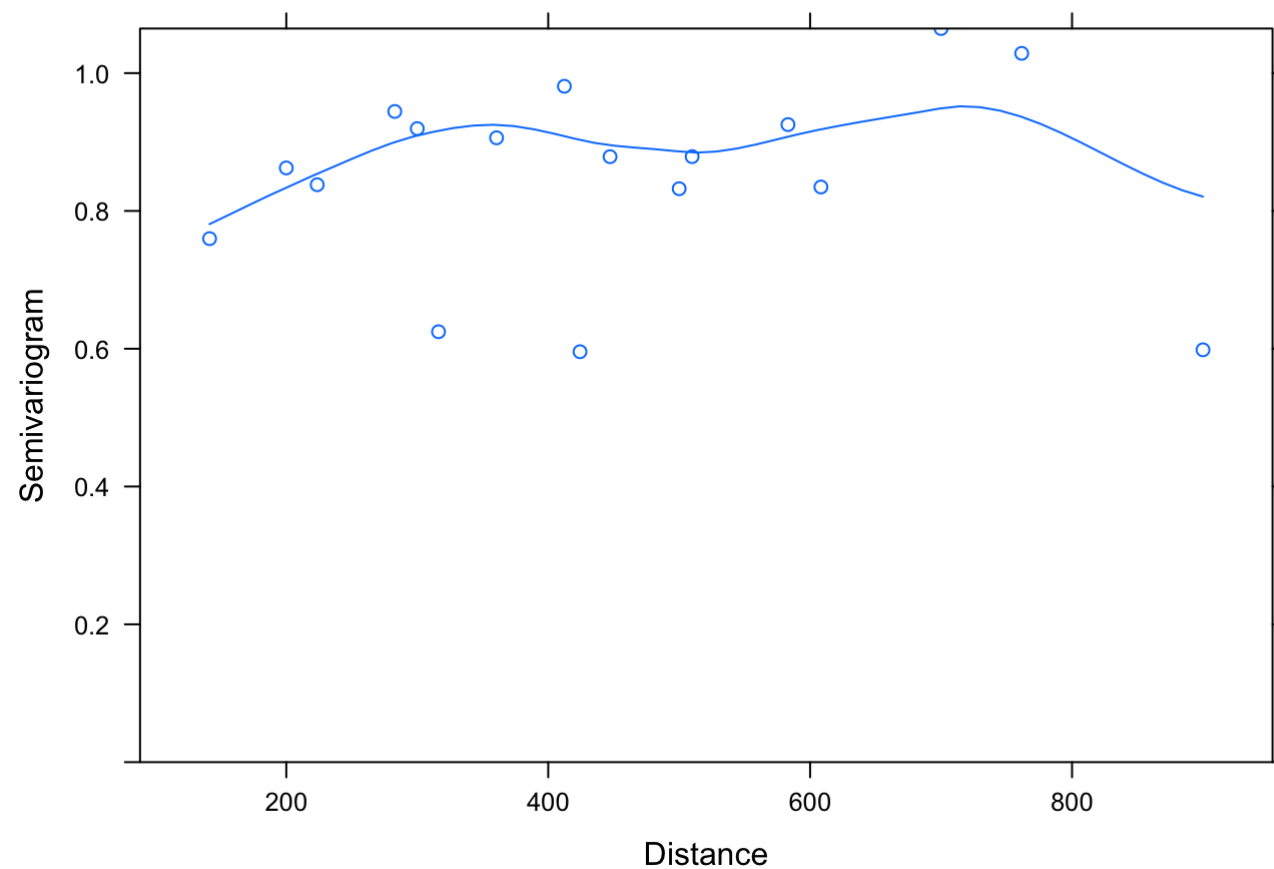
Linear Model -

```
all_lin = update(lm_all, corr=corLin(form=~x + y))
# examine fit of error model to the raw model residuals
# note this function defaults to displaying pearson standardized residuals
# resType='p' or resType='pearson'
plot(Variogram(all_lin, maxDist = maxdist))
```

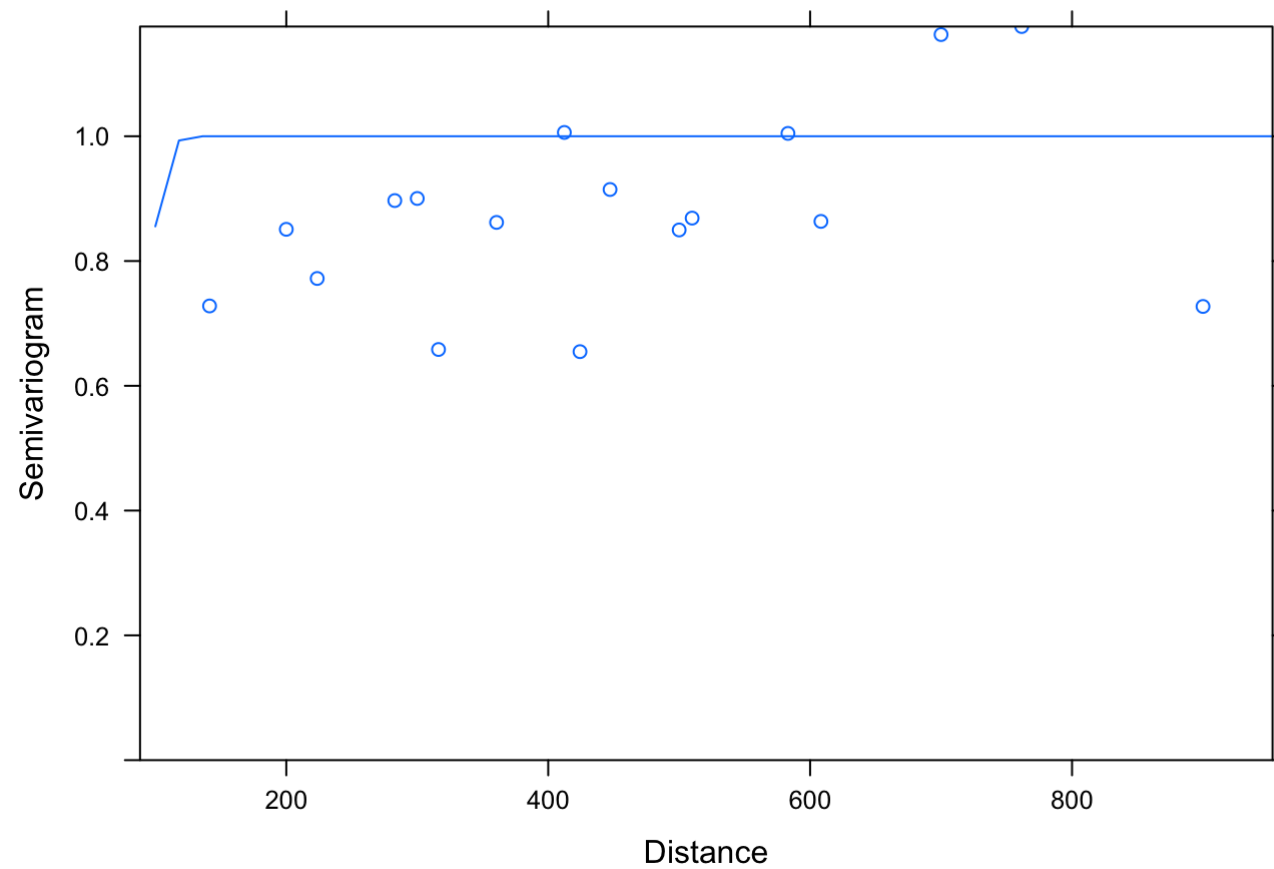


```
plot(Variogram(all_lin, resType='normalized', maxDist = maxdist))
```

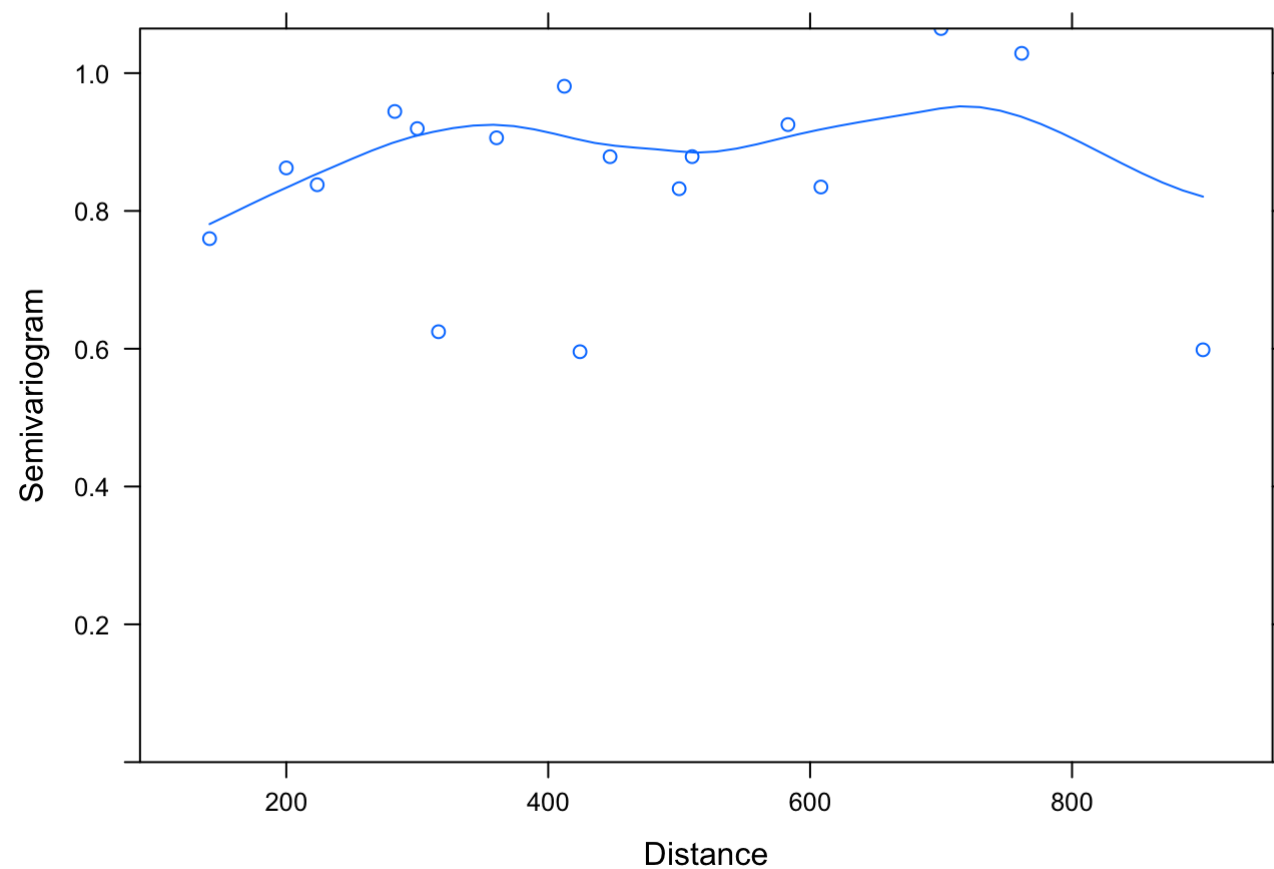




```
all_lin_nug = update(lm_all, corr=corLin(form=~x + y, nugget = T))  
# examine fit of error model to the raw model residuals  
# note this function defaults to displaying pearson standardized residuals  
# resType='p' or resType='pearson'  
plot(Variogram(all_lin_nug, maxDist = maxdist))
```

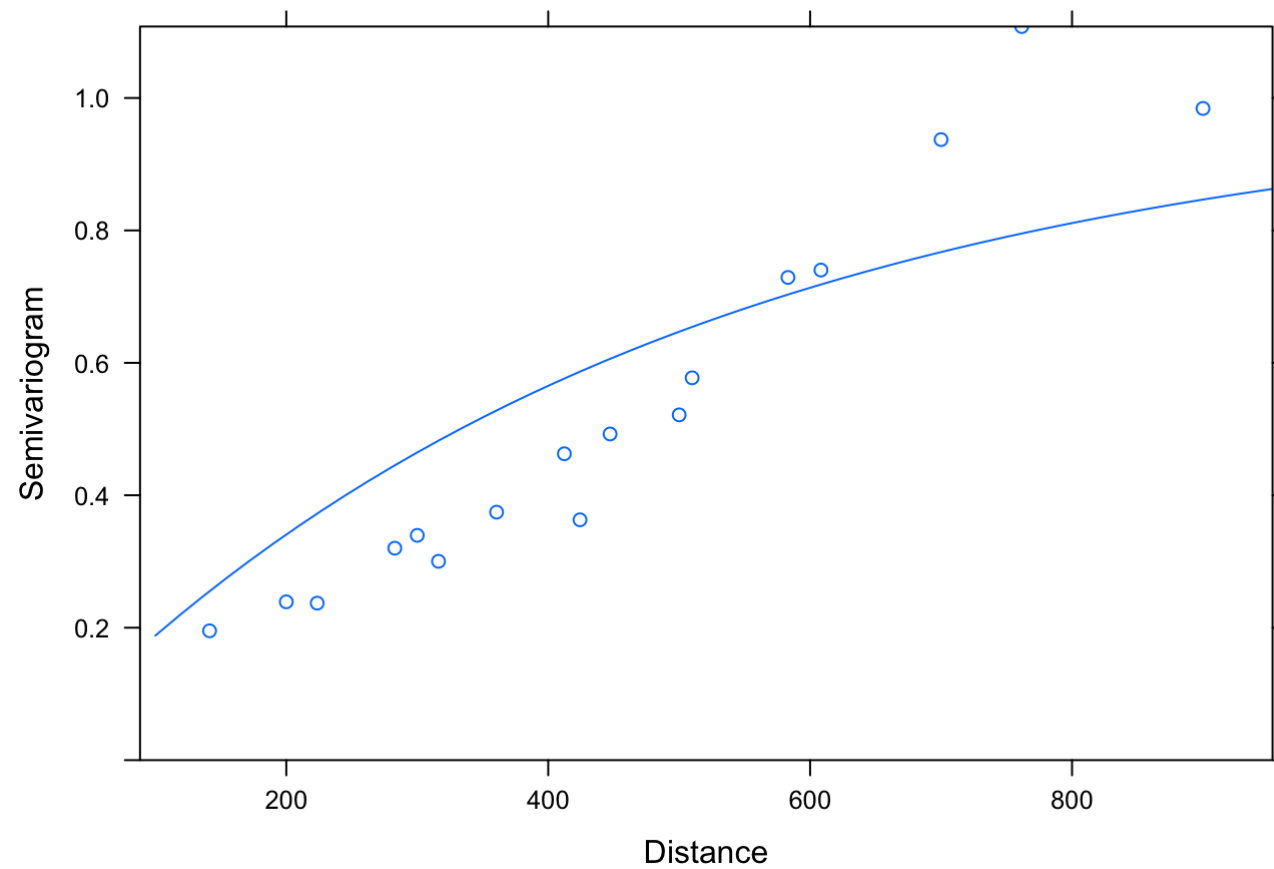


```
plot(Variogram(all_lin_nug, resType='normalized', maxDist = maxdist))
```

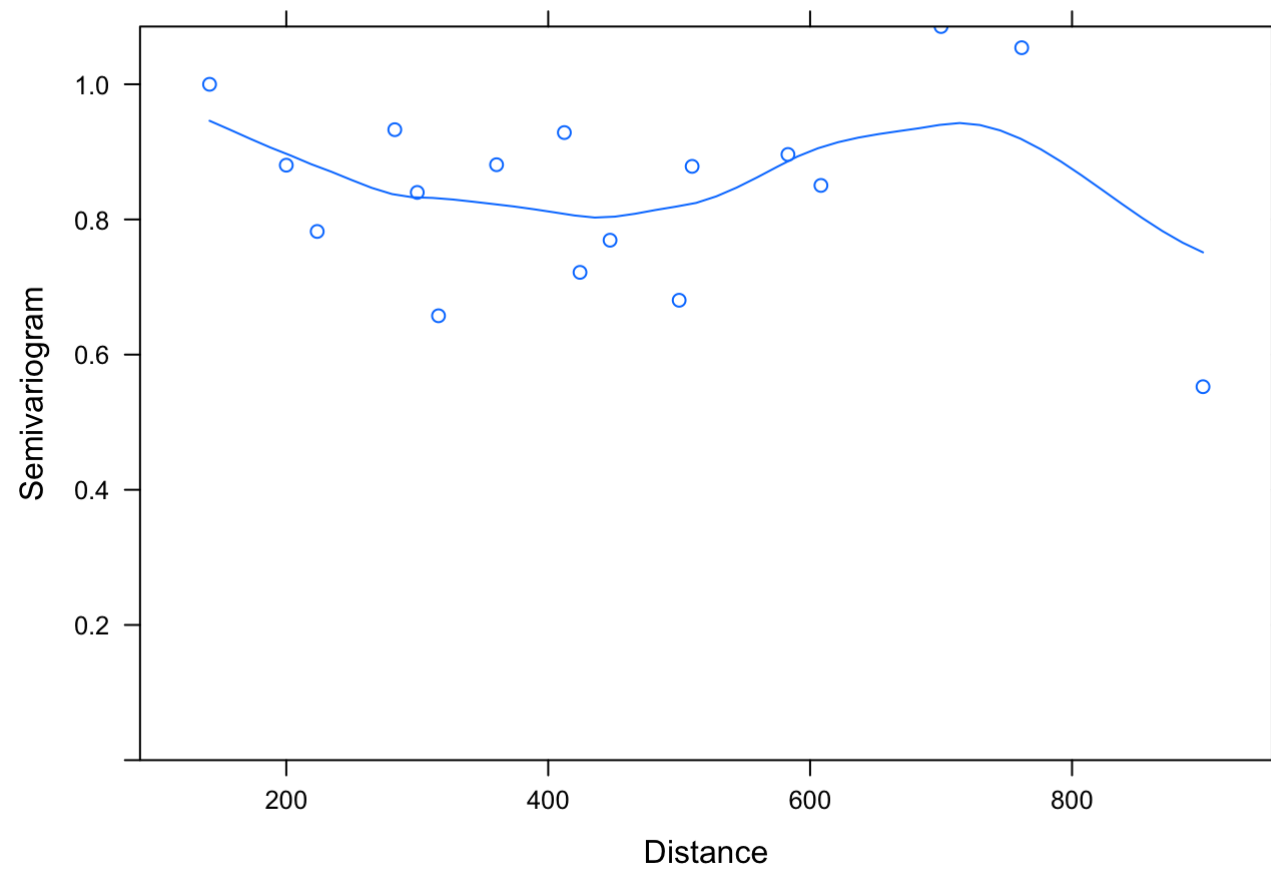


Exponential Model -

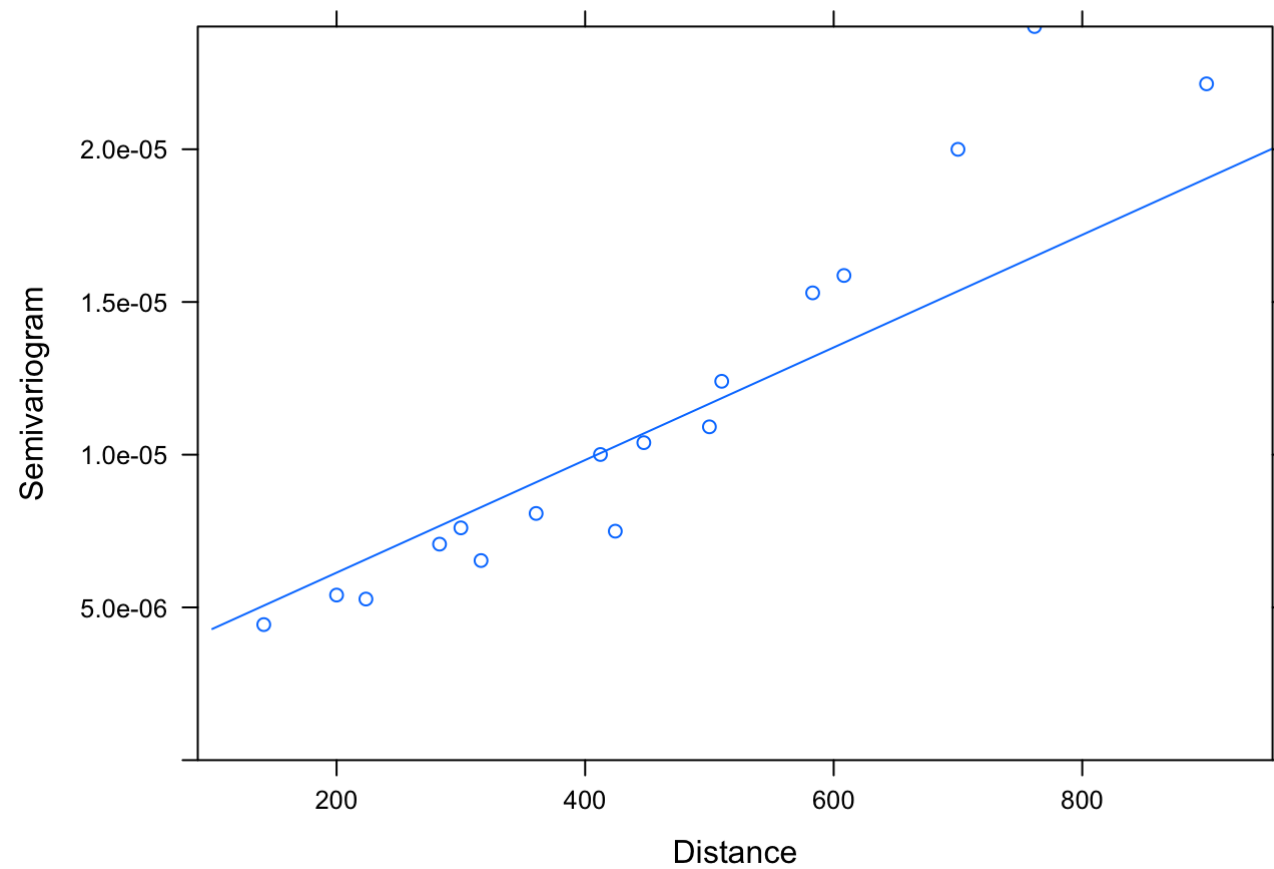
```
all_exp = update(lm_all, corr=corExp(form=~x + y))  
# examine fit of error model to the raw model residuals  
# note this function defaults to displaying pearson standardized residuals  
# resType='p' or resType='pearson'  
plot(Variogram(all_exp, maxDist = maxdist))
```



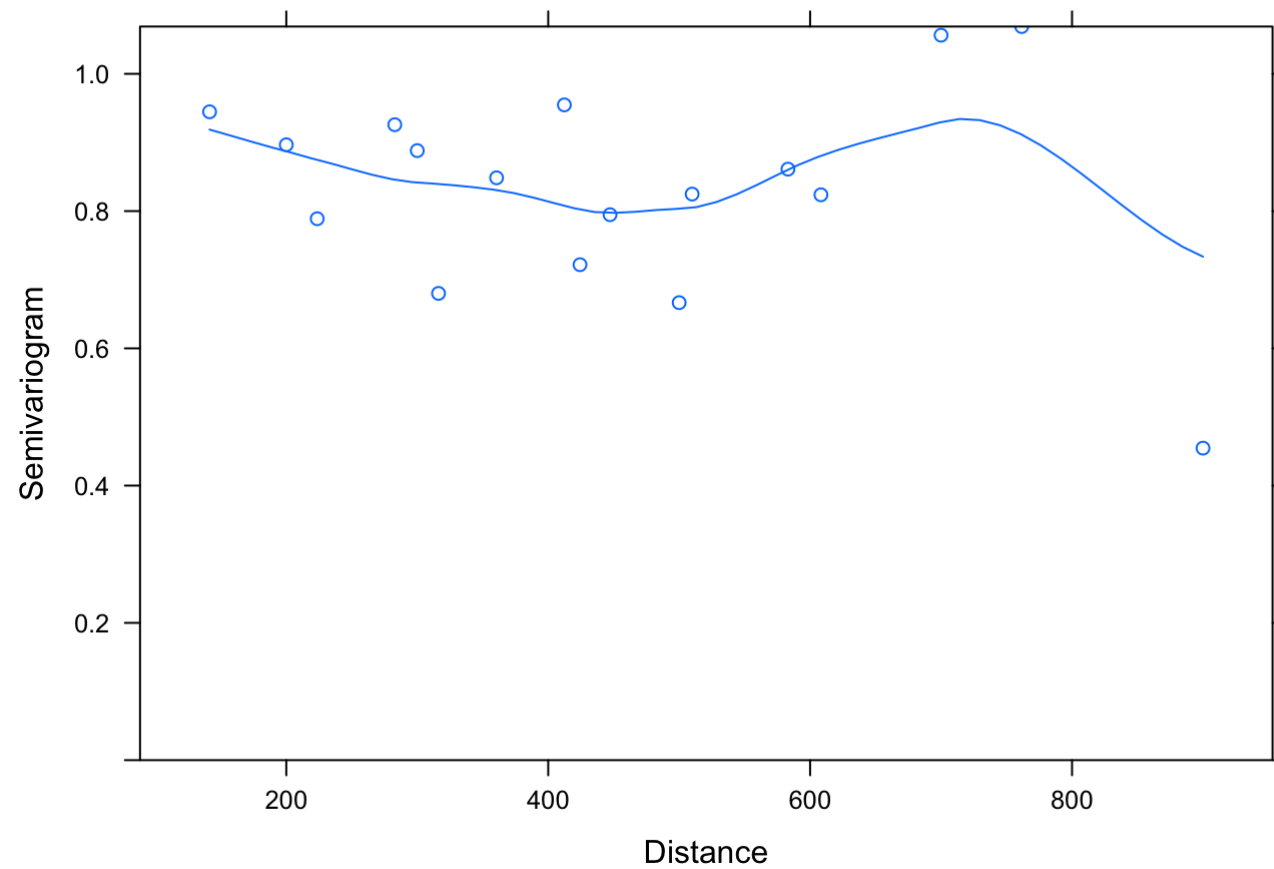
```
plot(Variogram(all_exp, resType='normalized', maxDist = maxdist))
```



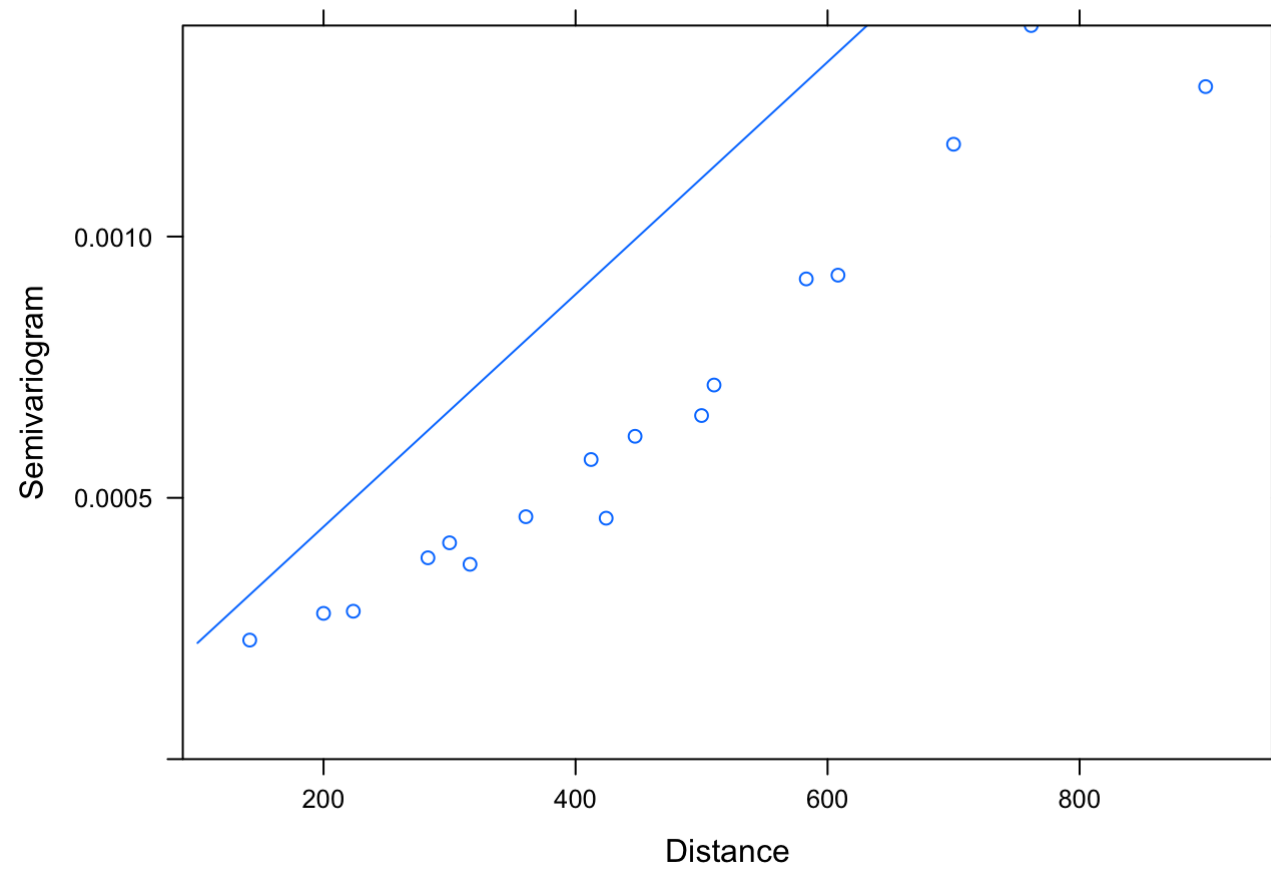
```
all_exp_nug = update(all_exp, corr=corExp(form= ~ x + y, nugget=T))  
plot(Variogram(all_exp_nug, maxDist = maxdist))
```



```
plot(Variogram(all_exp_nug, resType='n', maxDist = maxdist))
```

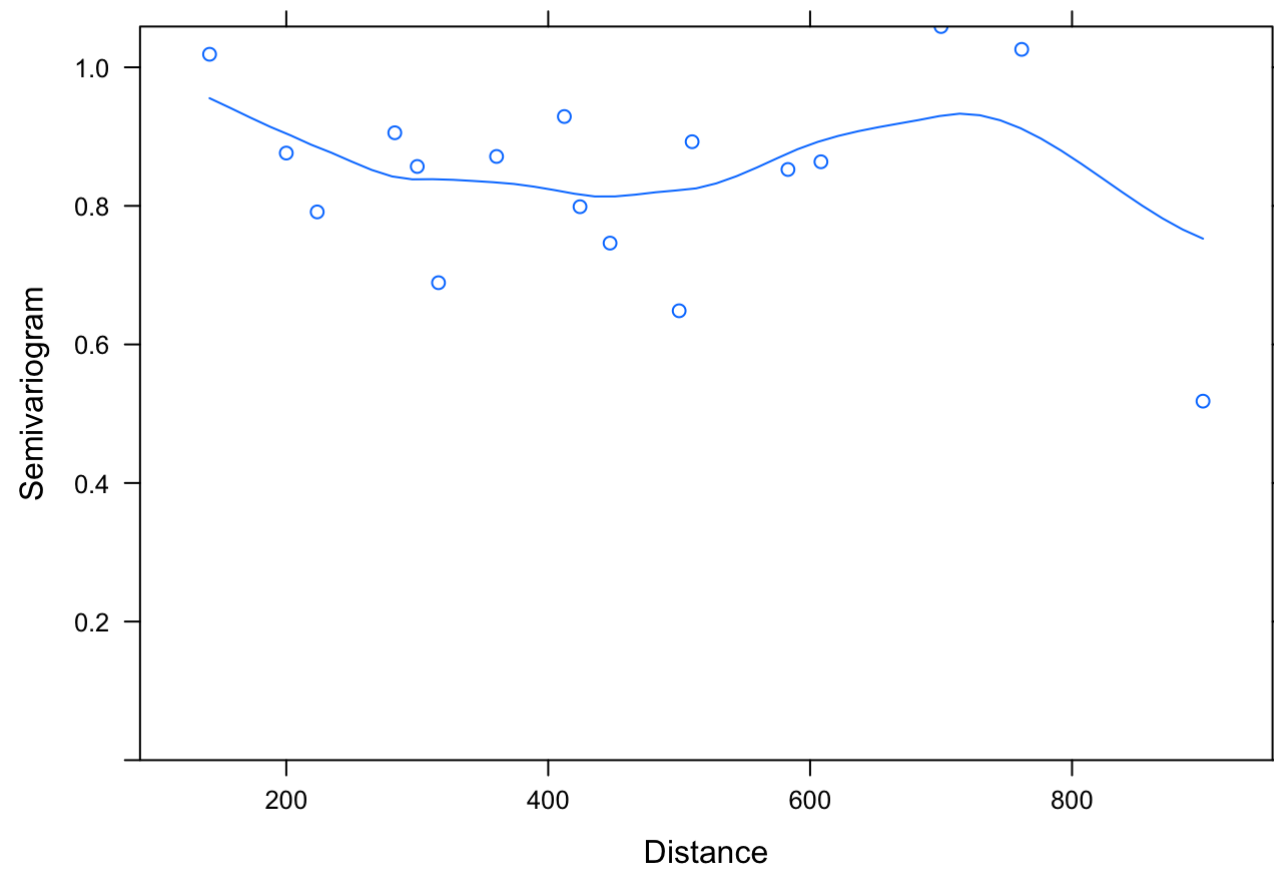


```
all_sph = update(lm_all, corr=corSpher(form=~x + y))  
plot(Variogram(all_sph, maxDist = maxdist))
```

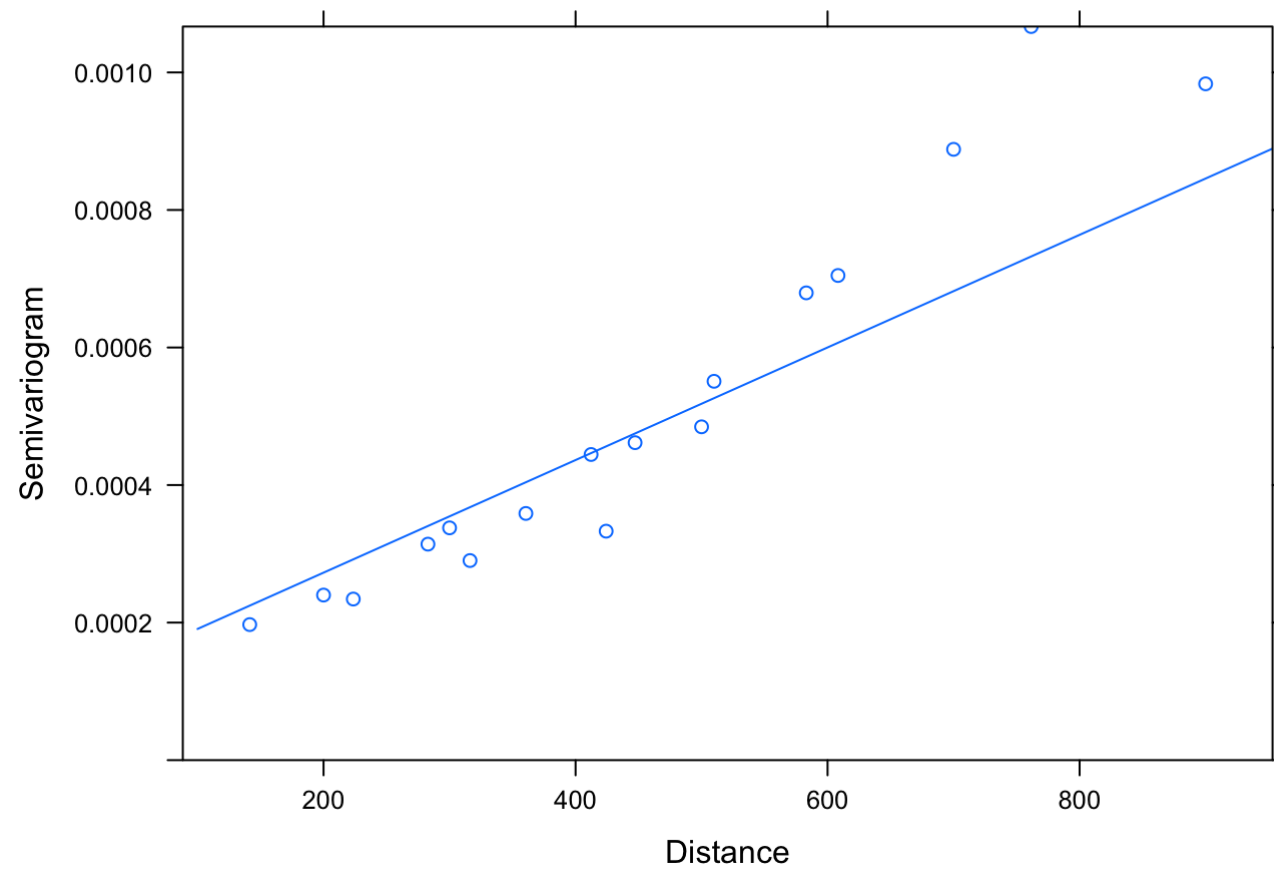


```
plot(Variogram(all_sph, resType='normalized', maxDist = maxdist))
```

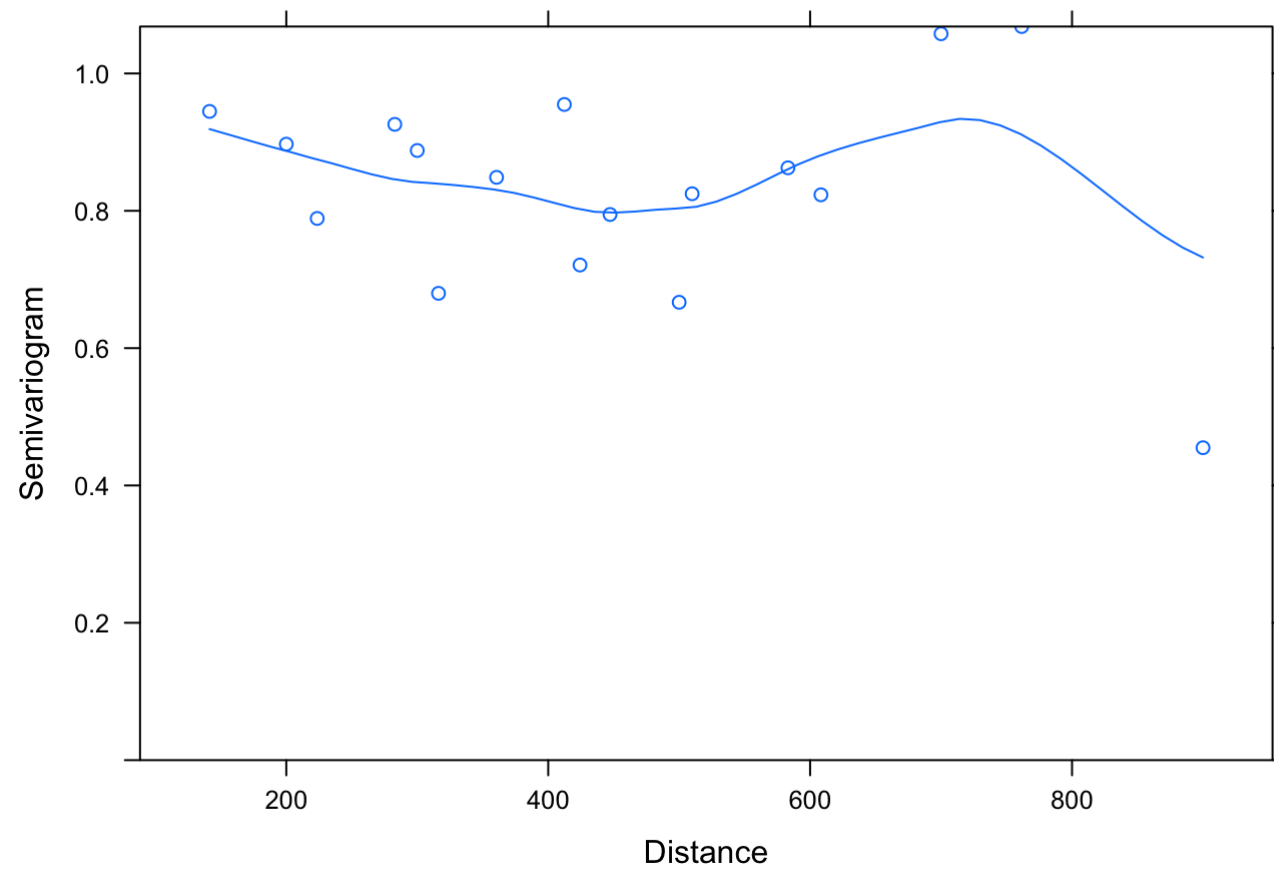




```
all_sph_nug = update(lm_all, corr=corSpher(form=~x + y, nugget = T))  
plot(Variogram(all_sph_nug, maxDist = maxdist))
```

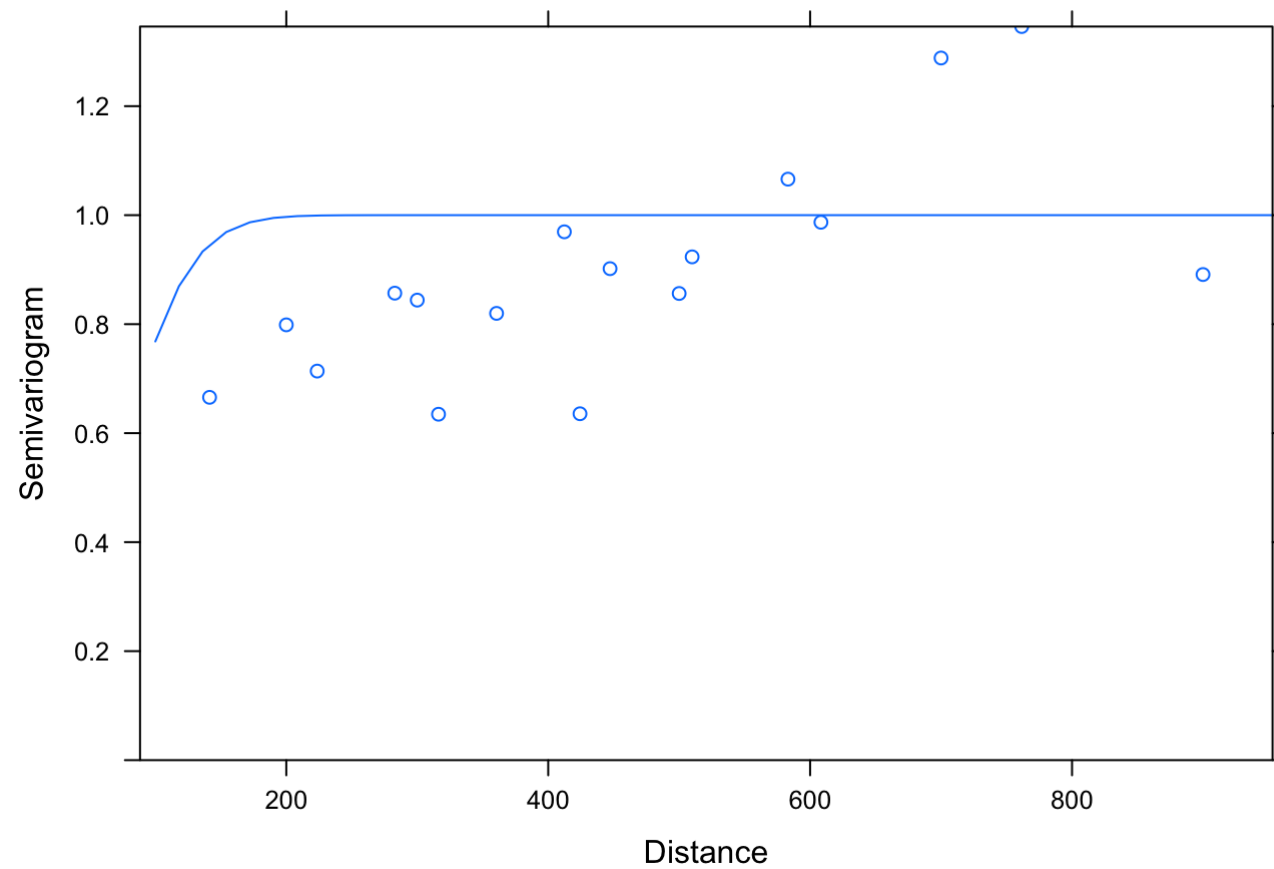


```
plot(Variogram(all_sph_nug, resType='normalized', maxDist = maxdist))
```

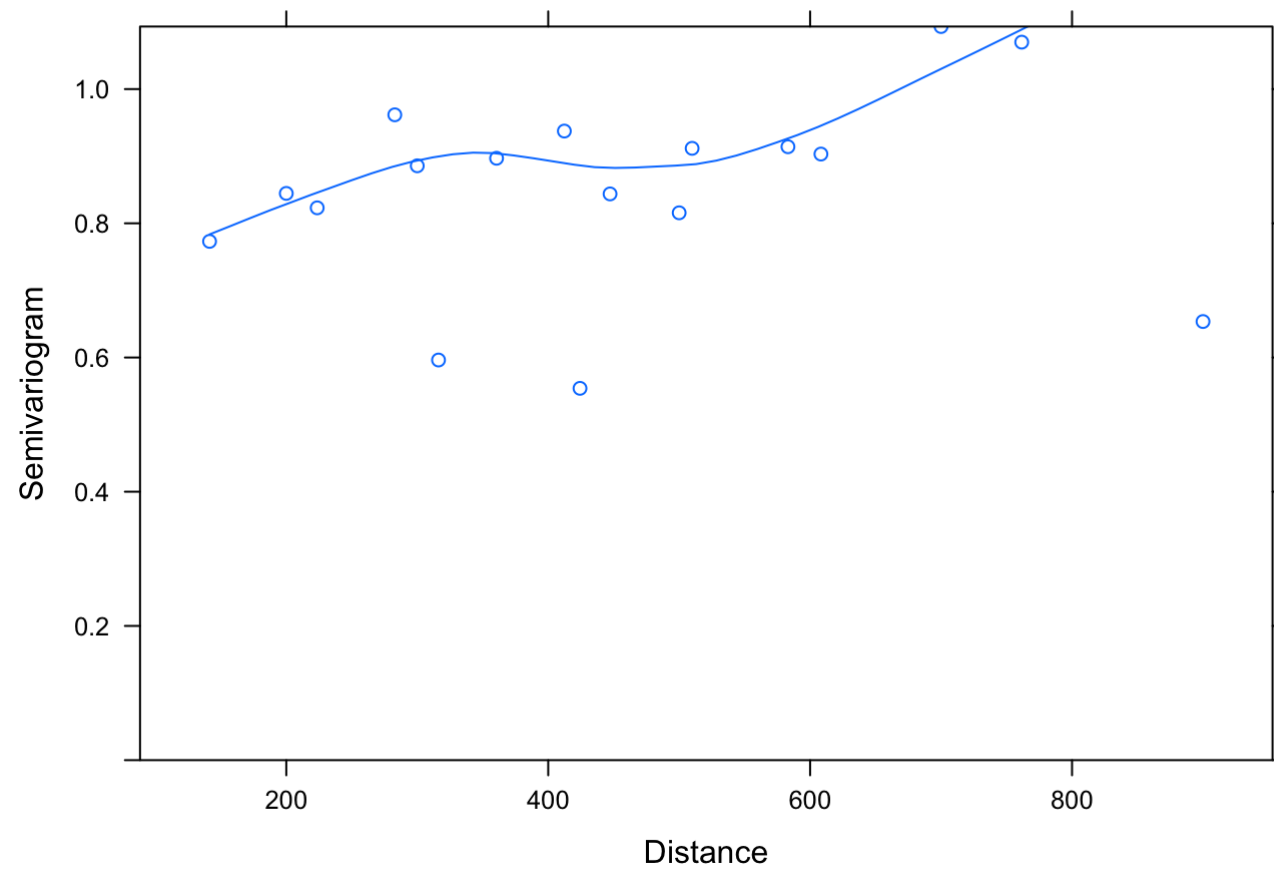


Gaussian Model -

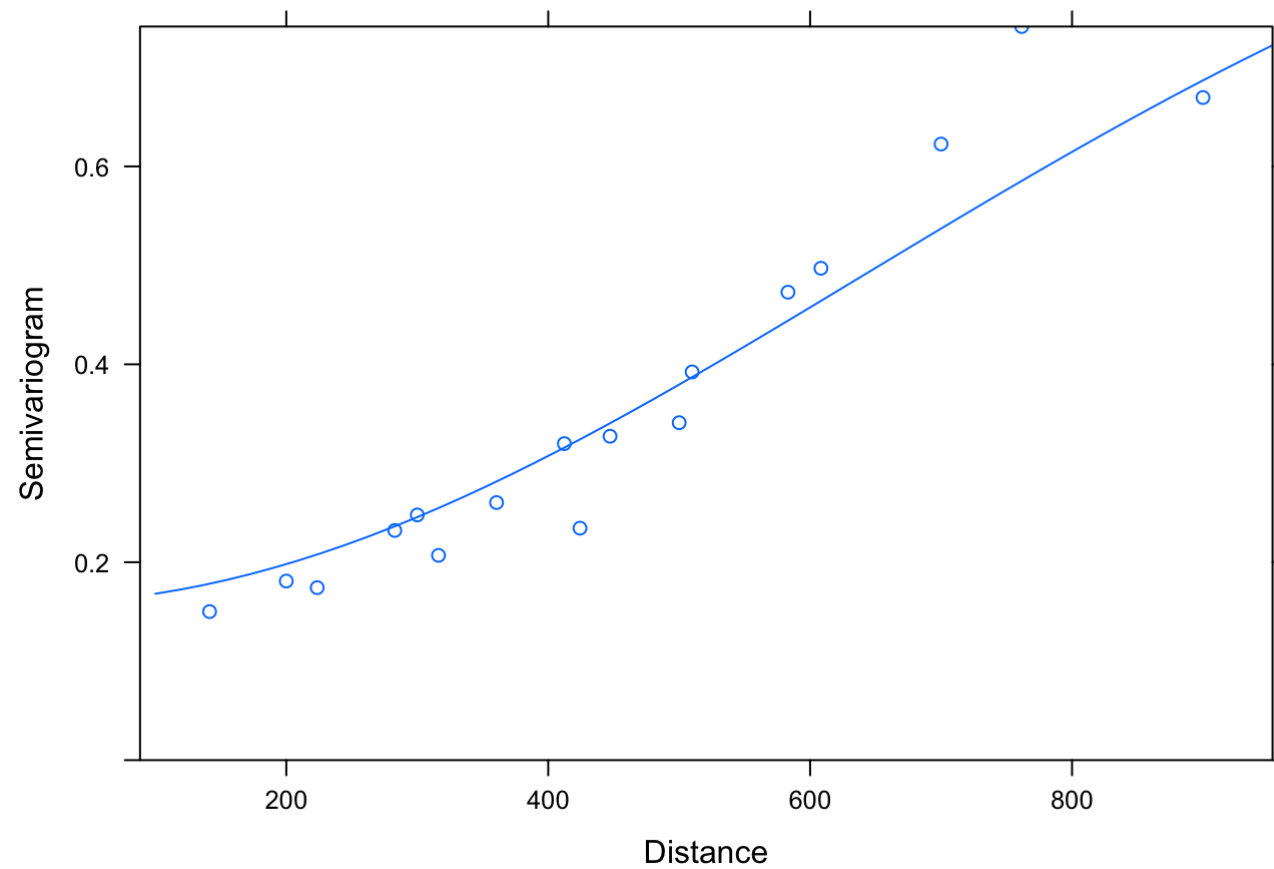
```
all_gaus = update(lm_all, corr=corGaus(form=~x + y))  
  
plot(Variogram(all_gaus, maxDist = maxdist))
```



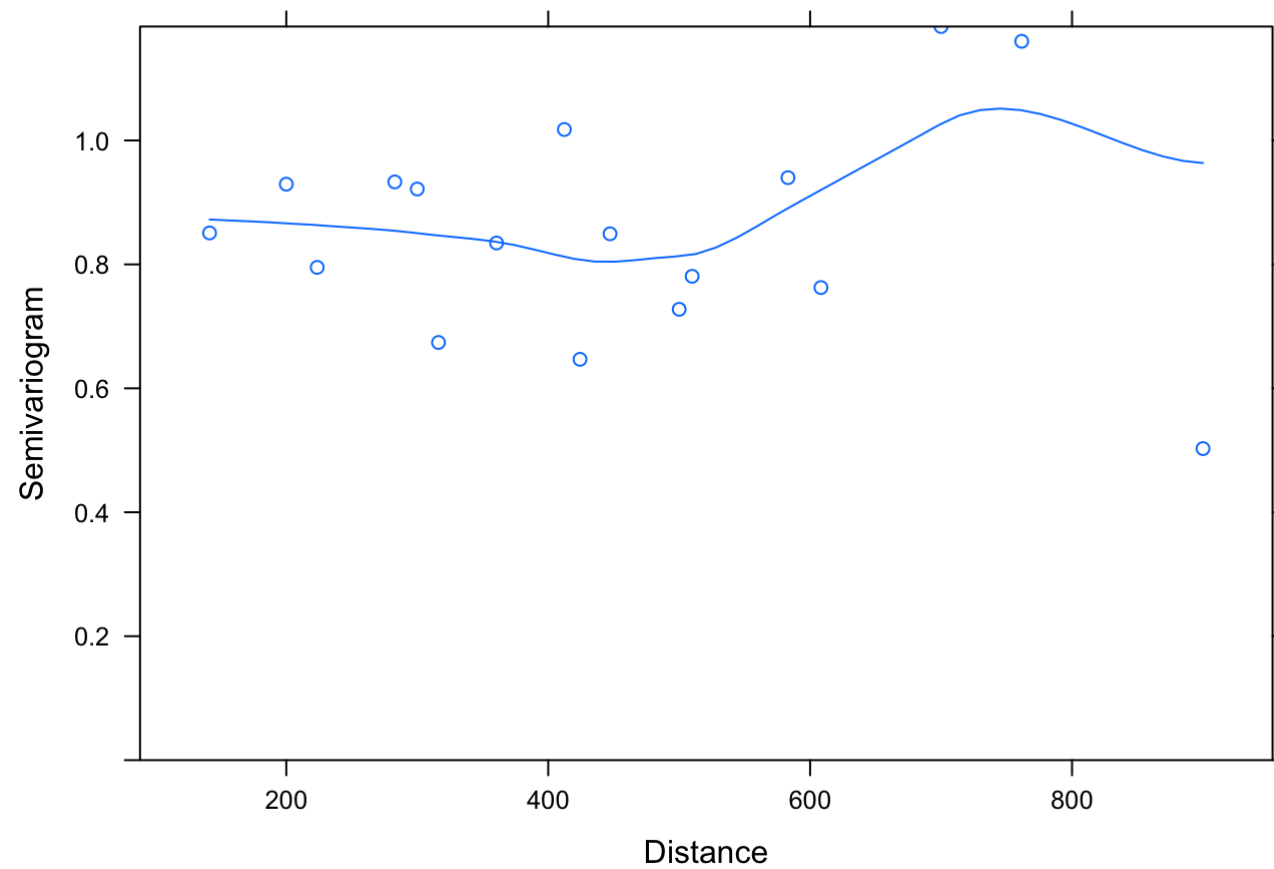
```
plot(Variogram(all_gaus, resType='normalized', maxDist = maxdist))
```



```
all_gaus_nug = update(lm_all, corr=corGaus(form=~x + y, nugget = T))  
plot(Variogram(all_gaus_nug, maxDist = maxdist))
```

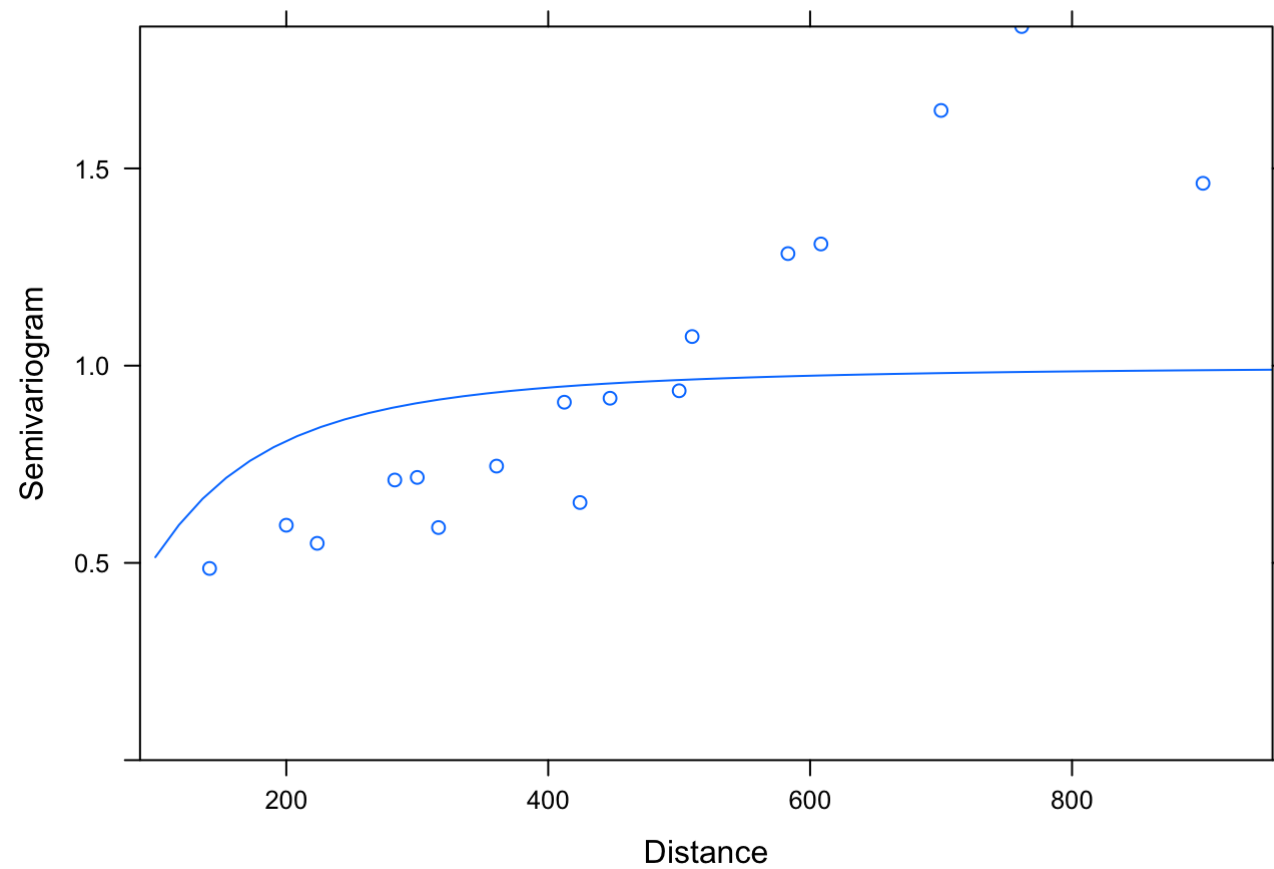


```
plot(Variogram(all_gaus_nug, resType='normalized', maxDist = maxdist))
```



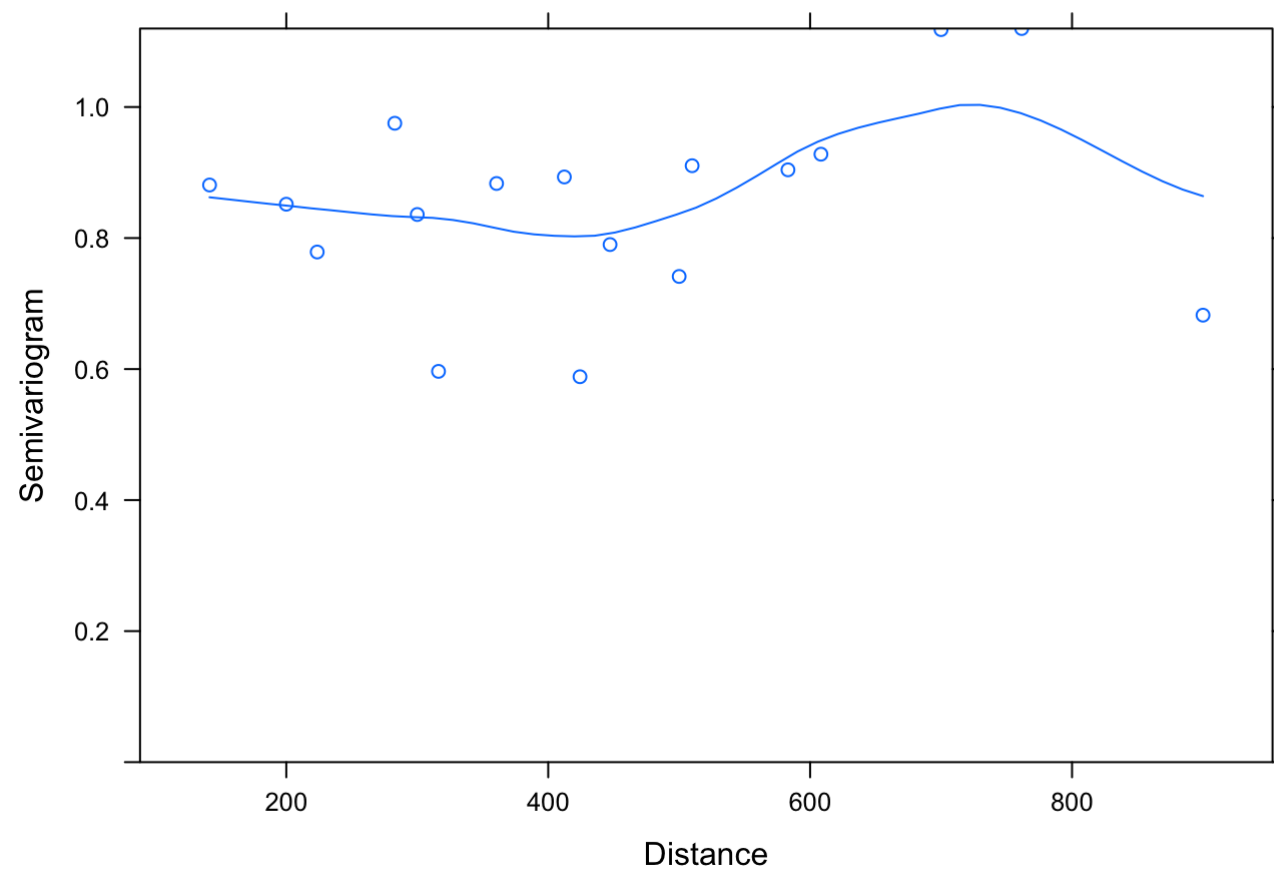
Rational Quadratic Model -

```
all_rat = update(lm_all, corr=corRatio(form=~x + y))  
plot(Variogram(all_rat, maxDist = maxdist))
```

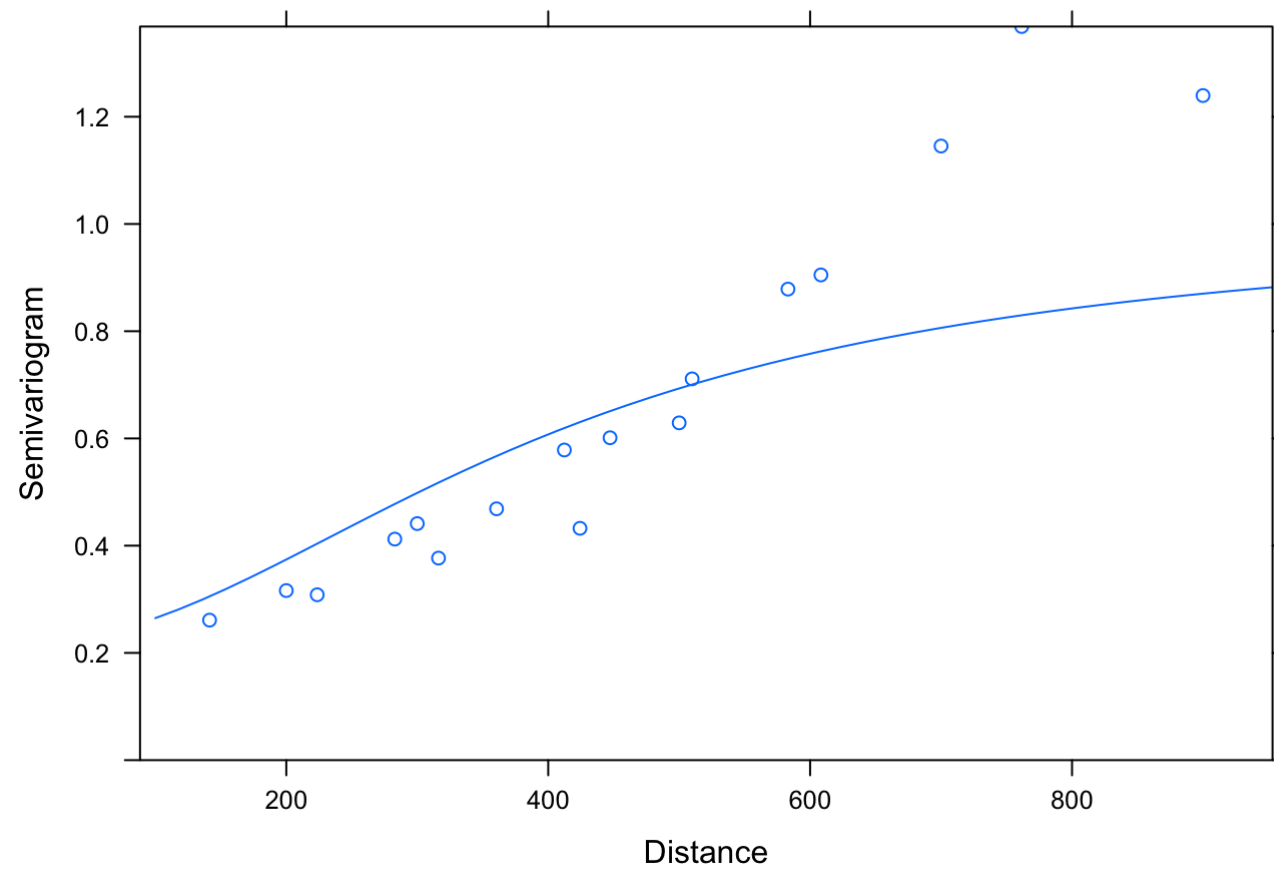


```
plot(Variogram(all_rat, resType='normalized', maxDist = maxdist))
```

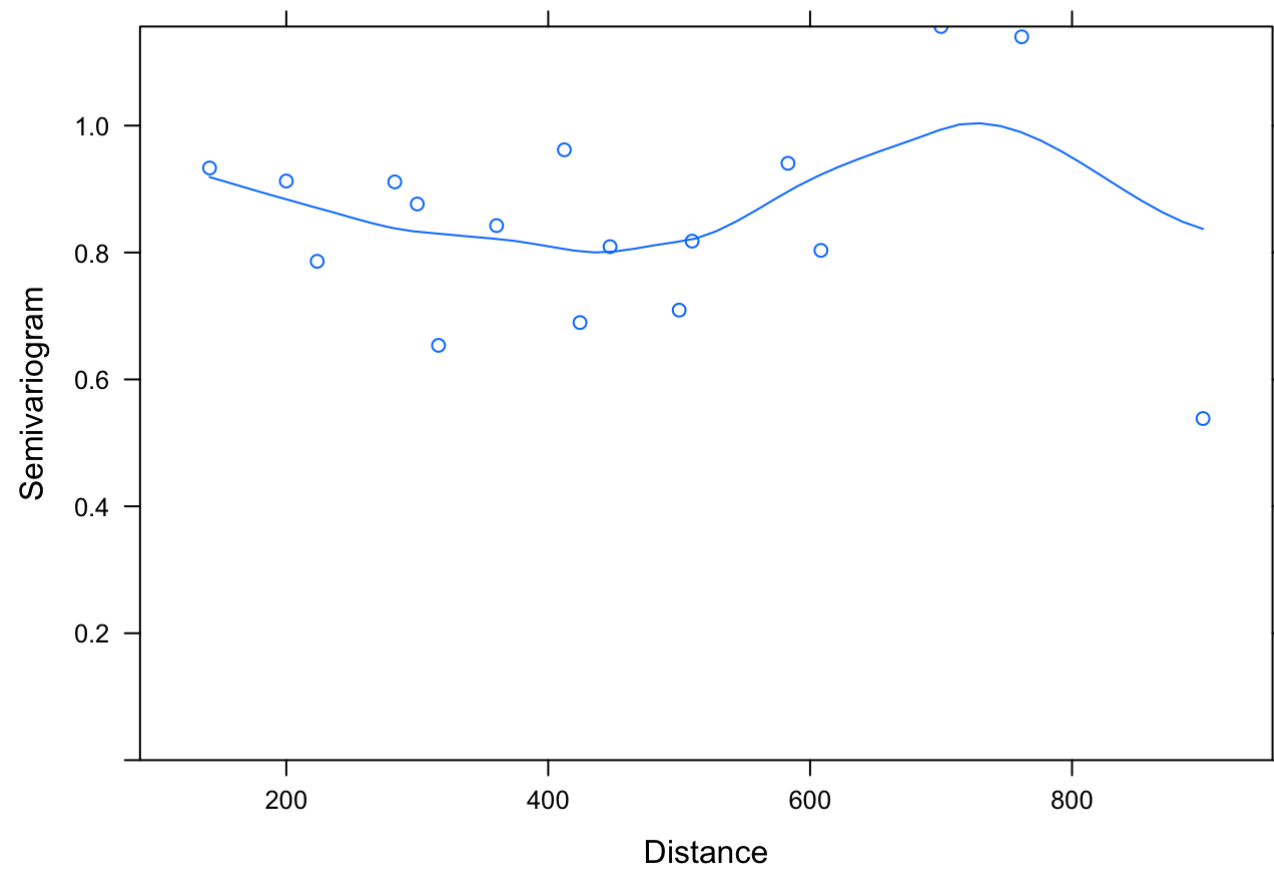




```
all_rat_nug = update(lm_all, corr=corRatio(form=~x + y, nugget=T))  
# examine fit of error model to model residuals  
plot(Variogram(all_rat_nug, maxDist = maxdist))
```



```
plot(Variogram(all_rat_nug, resType='normalized', maxDist = maxdist))
```



Comparing the models -

```
anova(lm_all, all_lin, all_lin_nug, all_gaus, all_gaus_nug, all_exp, all_exp_nug, all_sph, all_sph_nug, all_rat,  
all_rat_nug, test=F)
```

##	Model	df	AIC	BIC	logLik
##	lm_all	1 9	307.1163	322.7554	-144.5582
##	all_lin	2 10	307.7308	325.1075	-143.8654
##	all_lin_nug	3 11	309.7308	328.8452	-143.8654
##	all_gaus	4 10	307.2070	324.5837	-143.6035
##	all_gaus_nug	5 11	303.8653	322.9797	-140.9327
##	all_exp	6 10	301.6062	318.9829	-140.8031
##	all_exp_nug	7 11	301.9592	321.0735	-139.9796
##	all_sph	8 10	301.9254	319.3021	-140.9627
##	all_sph_nug	9 11	301.9592	321.0735	-139.9796
##	all_rat	10 10	303.8542	321.2309	-141.9271
##	all_rat_nug	11 11	303.1486	322.2630	-140.5743

Best fitting models to display the data are the exponential and spherical models with and without the nuggets are equally the best models. It appears the models leaving out the spatial error term are slightly better than those including it. The exponential model is the best by a few decimal points.

```
summary(all_exp)
```

```

## Generalized least squares fit by REML
##   Model: sp_ref ~ sp_a + sp_b + sp_c + sp_d + sp_e + sp_f + sp_g
##   Data: bci_dat
##           AIC      BIC    logLik
##   301.6062 318.9829 -140.8031
##
## Correlation Structure: Exponential spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##   range
## 480.0567
##
## Coefficients:
##               Value Std.Error  t-value p-value
## (Intercept) 2.3485197  6.154919 0.381568  0.7047
## sp_a         0.1208390  0.179811 0.672033  0.5052
## sp_b         0.0191759  0.098501 0.194677  0.8466
## sp_c         0.2014516  0.509196 0.395627  0.6944
## sp_d         1.2792289  1.847570 0.692385  0.4925
## sp_e         0.0674943  0.133782 0.504511  0.6165
## sp_f         1.8115374  0.525147 3.449582  0.0013
## sp_g         0.3388574  0.156874 2.160064  0.0365
##
## Correlation:
##   (Intr) sp_a  sp_b  sp_c  sp_d  sp_e  sp_f
## sp_a -0.226
## sp_b -0.309 -0.022
## sp_c  0.045 -0.066 -0.369
## sp_d -0.059 -0.304  0.321 -0.142
## sp_e -0.240 -0.016  0.288 -0.221  0.112
## sp_f -0.069  0.168 -0.237  0.212 -0.633 -0.041
## sp_g -0.056 -0.137 -0.063  0.109  0.290  0.102 -0.186
##
## Standardized residuals:
##           Min           Q1           Med           Q3           Max
## -1.0051632 -0.5235683 -0.3176178  0.2208753  2.3746027
##
## Residual standard error: 8.628464
## Degrees of freedom: 50 total; 42 residual

```

```
summary(all_exp_nug)
```

```

## Generalized least squares fit by REML
## Model: sp_ref ~ sp_a + sp_b + sp_c + sp_d + sp_e + sp_f + sp_g
## Data: bci_dat
##      AIC      BIC    logLik
## 301.9592 321.0735 -139.9796
##
## Correlation Structure: Exponential spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##      range      nugget
## 5.424635e+07 2.451077e-06
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept) 3.0501045 1785.0597  0.001709  0.9986
## sp_a         0.1426666   0.1895  0.752752  0.4558
## sp_b        -0.0017714   0.0904 -0.019600  0.9845
## sp_c         0.2863335   0.5274  0.542880  0.5901
## sp_d         1.3263643   1.9368  0.684818  0.4972
## sp_e         0.0407530   0.1395  0.292085  0.7717
## sp_f         1.8170749   0.5730  3.171303  0.0028
## sp_g         0.4086700   0.1537  2.659324  0.0110
##
## Correlation:
##      (Intr) sp_a  sp_b  sp_c  sp_d  sp_e  sp_f
## sp_a -0.001
## sp_b -0.001 -0.098
## sp_c  0.000  0.017 -0.360
## sp_d  0.000 -0.292  0.344 -0.193
## sp_e -0.001 -0.020  0.160 -0.197  0.088
## sp_f  0.000  0.165 -0.276  0.255 -0.655 -0.036
## sp_g  0.000 -0.066 -0.037 -0.048  0.306  0.140 -0.183
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.0049325435 -0.0029088009 -0.0020671675  0.0005570041  0.0103697005
##
## Residual standard error: 1785.068
## Degrees of freedom: 50 total; 42 residual

```

```
summary(all_sph)
```



```

## Generalized least squares fit by REML
## Model: sp_ref ~ sp_a + sp_b + sp_c + sp_d + sp_e + sp_f + sp_g
## Data: bci_dat
##      AIC      BIC    logLik
## 301.9254 319.3021 -140.9627
##
## Correlation Structure: Spherical spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
## range
## 674629.1
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  3.988119 256.16657  0.015568  0.9877
## sp_a         0.111421  0.17643  0.631546  0.5311
## sp_b        -0.001071  0.09890 -0.010832  0.9914
## sp_c         0.149191  0.50026  0.298227  0.7670
## sp_d         0.813048  1.83774  0.442417  0.6605
## sp_e         0.075313  0.13096  0.575101  0.5683
## sp_f         1.834240  0.51295  3.575877  0.0009
## sp_g         0.309959  0.15670  1.978002  0.0545
##
## Correlation:
##      (Intr) sp_a  sp_b  sp_c  sp_d  sp_e  sp_f
## sp_a -0.007
## sp_b -0.010 -0.007
## sp_c  0.002 -0.066 -0.361
## sp_d -0.003 -0.290  0.339 -0.127
## sp_e -0.007 -0.017  0.287 -0.227  0.108
## sp_f -0.003  0.171 -0.232  0.207 -0.623 -0.039
## sp_g -0.001 -0.133 -0.061  0.134  0.298  0.089 -0.183
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.038619444 -0.023162234 -0.015811190  0.003689907  0.078083046
##
## Residual standard error: 256.311
## Degrees of freedom: 50 total; 42 residual

```

```
summary(all_sph_nug)
```

```

## Generalized least squares fit by REML
## Model: sp_ref ~ sp_a + sp_b + sp_c + sp_d + sp_e + sp_f + sp_g
## Data: bci_dat
##      AIC      BIC    logLik
## 301.9592 321.0735 -139.9796
##
## Correlation Structure: Spherical spatial correlation
## Formula: ~x + y
## Parameter estimate(s):
##      range      nugget
## 1.831904e+06 1.088636e-04
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  3.0501051 267.79376  0.011390  0.9910
## sp_a         0.1426672  0.18953  0.752754  0.4558
## sp_b        -0.0017713  0.09038 -0.019599  0.9845
## sp_c         0.2863351  0.52743  0.542882  0.5901
## sp_d         1.3263713  1.93681  0.684821  0.4972
## sp_e         0.0407524  0.13952  0.292080  0.7717
## sp_f         1.8170748  0.57298  3.171298  0.0028
## sp_g         0.4086712  0.15367  2.659334  0.0110
##
## Correlation:
##      (Intr) sp_a  sp_b  sp_c  sp_d  sp_e  sp_f
## sp_a -0.006
## sp_b -0.007 -0.098
## sp_c  0.001  0.017 -0.360
## sp_d -0.001 -0.292  0.344 -0.193
## sp_e -0.005 -0.020  0.160 -0.197  0.088
## sp_f -0.002  0.165 -0.276  0.255 -0.655 -0.036
## sp_g -0.001 -0.066 -0.037 -0.048  0.306  0.140 -0.183
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -0.032872366 -0.019385351 -0.013776425  0.003711977  0.069107528
##
## Residual standard error: 267.852
## Degrees of freedom: 50 total; 42 residual

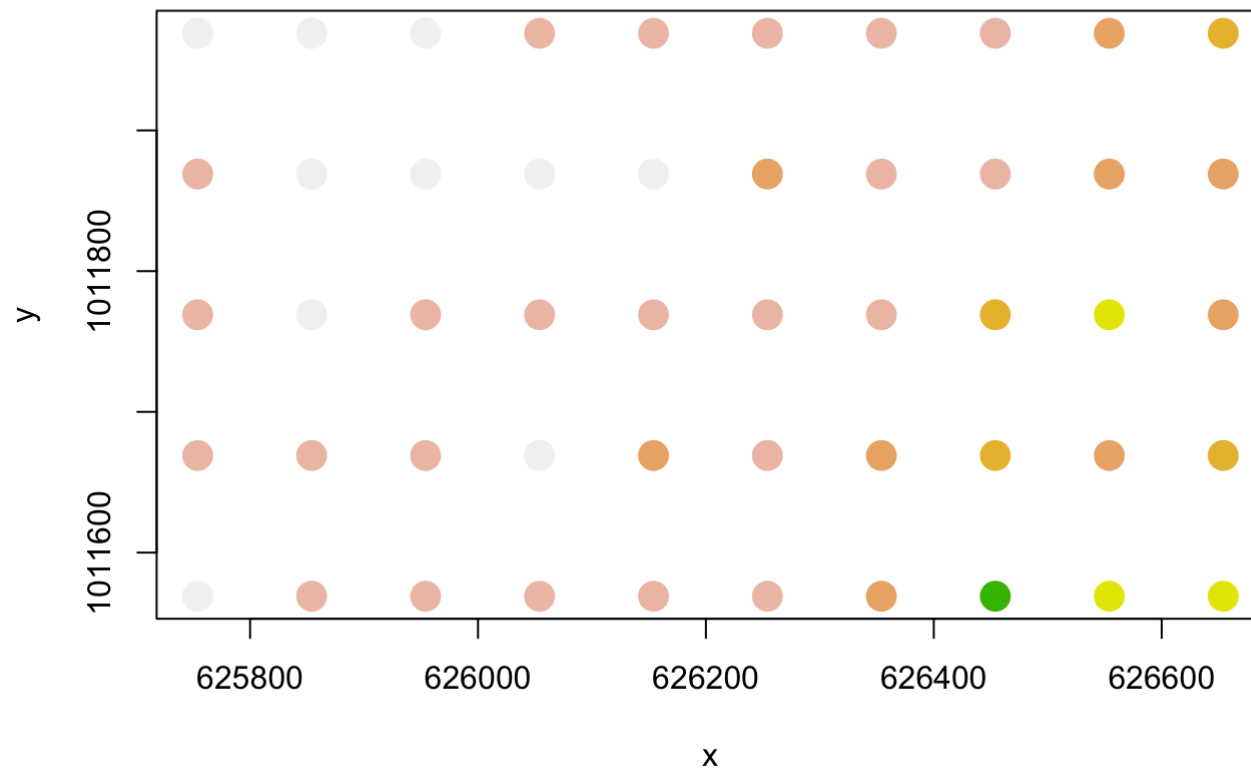
```

Based on the models of best fit, across all the predictor species included there is no significant effect of their abundances' on that of *D. standleyi* across space within the Barro Colorado Island.

The addition of the spatial error term (nugget = T) to correct for a non-zero y-intercept in the models did not significantly differ in goodness of fit compared to the models lacking it. The models with a nugget of 0 were only slightly better in AIC by a few decimal points, not enough to be significantly better. The regression coefficients and t statistics do not change much with/without the nugget included.

The reason we don't observe greater fit of the model with the spatial error term as we did before is likely because of the additional independent variables we have included. Looking at the semivariogram graphs for each model we can for these non-zero y-intercepts, the difference between the y-intercept and 0 is much more minute. While they are still non-zero, we may see less of an effect because the lesser separation distance from 0.

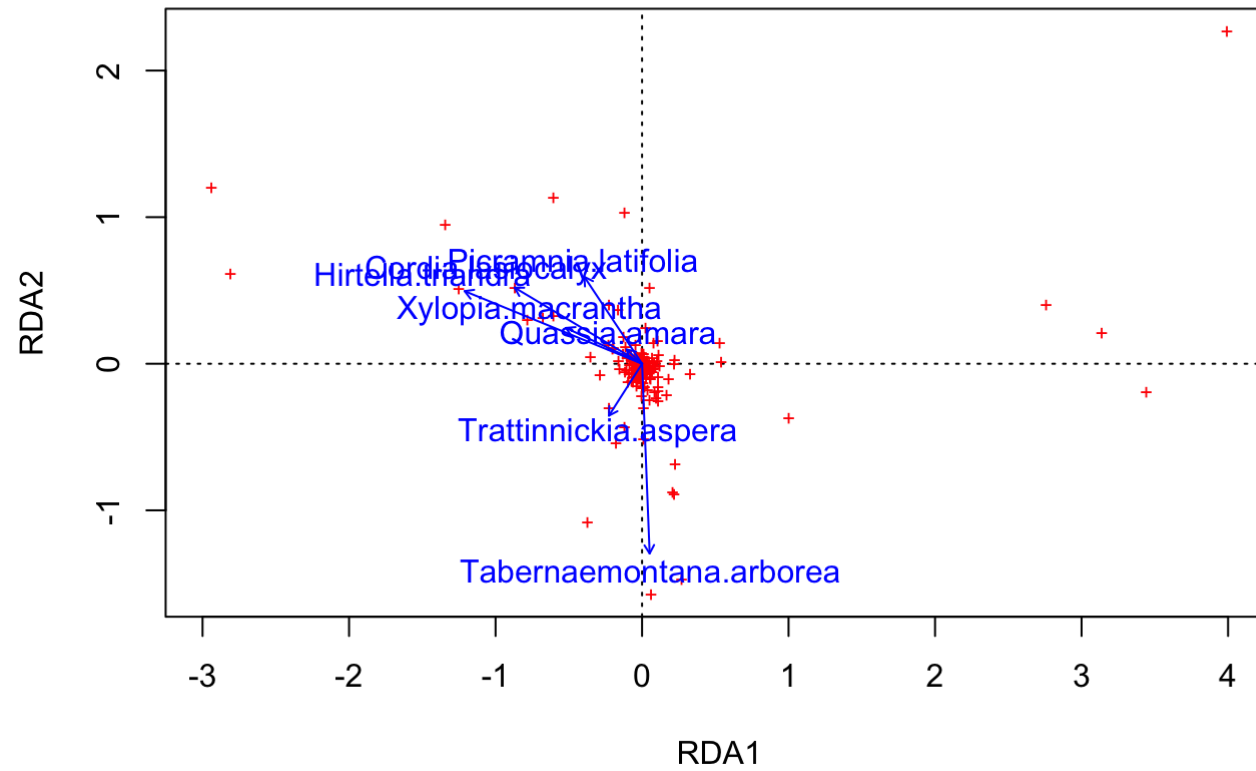
```
col_brksdos = hist(residuals(all_exp), plot=F)$breaks
col_indicesdos = as.numeric(cut(residuals(all_exp), col_brksdos))
colsdos = rev(terrain.colors(length(col_brksdos)))
plot(BCI_xy, cex=2, pch=19, col=colsdos[col_indicesdos])
```



Appears to still be some spatial patterning in the residuals of our model, may need to use a multi-variate approach.

RDA using the predictor species

```
bci_rda = rda(BCI, pred_sp[,])  
  
plot(bci_rda, display=c('sp', 'bp'))
```



```
anova(bci_rda)
```

```
## Permutation test for rda under reduced model
## Permutation: free
## Number of permutations: 999
##
## Model: rda(X = BCI, Y = pred_sp[, ])
##           Df Variance      F Pr(>F)
## Model      7   1034.2 2.5741 0.025 *
## Residual 42    2410.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on this RDA plot it appears a majority of species cluster in the same region, however species *T. aspera* and *T. arborea* show dissimilar patterns. Both species ordinations are pointing in the opposing direction of the other species and *T. arborea*'s arrow is long and parallel to the negative x axis, suggesting it is highly correlated with this axis. The anova shows the rda model is significant so there may be more multi-variate interactions to be explored here.