Spatial models assignment

Dan McGlinn

2025-03-24

Spatial Modeling Assignment

```
library(nlme)
library(vegan)
```

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

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

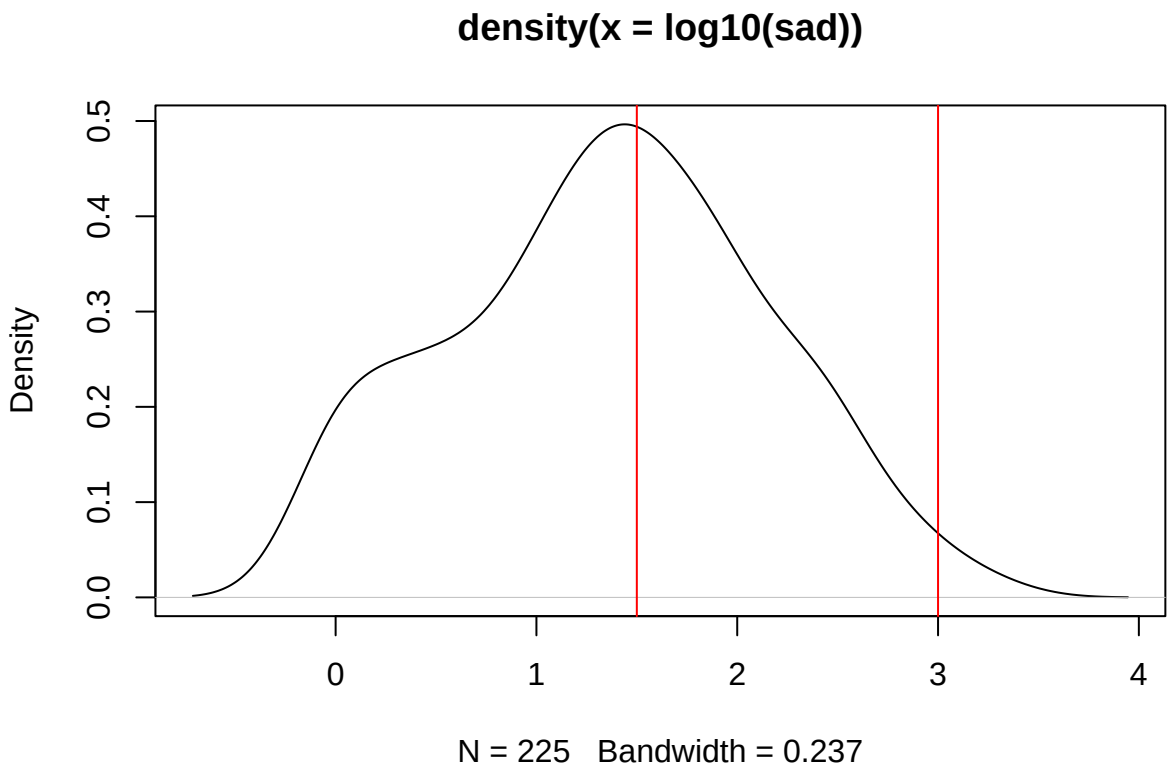
```
## This is vegan 2.6-4
```

```
source('./scripts/utility_functions.R')
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

The first thing we need to resolve is what would we consider as an example of a rare or a common species? The rareest species have only a single individual but clearly there will be no spatial pattern in that. Similarly the most common species in the dataset likely occurs in ever single quadrat so this species may also not yield any interesting spatial information. Let's examine the distribution of abundances to see what would be reasonable to consider in this particular dataset.

```
# compute the species-abundance distribution
sad <- apply(BCI, 2, sum)
# examine log transformed frequency distribution
plot(density(log10(sad)))
# visually define a two cutoffs for rare and common
abline(v=1.5, col='red')
abline(v=3, col='red')
```



The species abundance distribution (`sad`) indicates some relatively clear breaks to consider for a rare and a common species. Notice that I used a log transformation on abundance because a small number of species have very large abundances but most things are rare. We can use the sad to help us define what a rare and common species are (see vertical red lines in above graphic).

Now let's pull one rare species and one common species that are near our thresholds we defined.

```
sp_names <- names(BCI)[order(sad)]
sad_ord <- sad[order(sad)]
rare_sp <- sp_names[sad_ord > 10^1.5][1]
comm_sp <- sp_names[sad_ord > 10^3][1]
print('Example rare species is:')
```

```
## [1] "Example rare species is:"
```

```
rare_sp
```

```
## [1] "Trophis.racemosa"
```

```
print('Example common species is:')
```

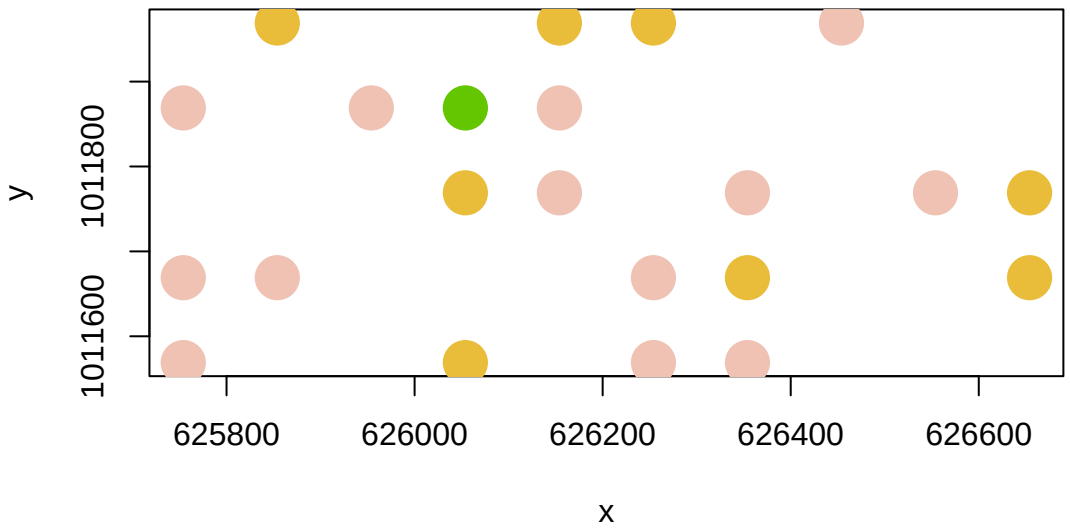
```
## [1] "Example common species is:"
```

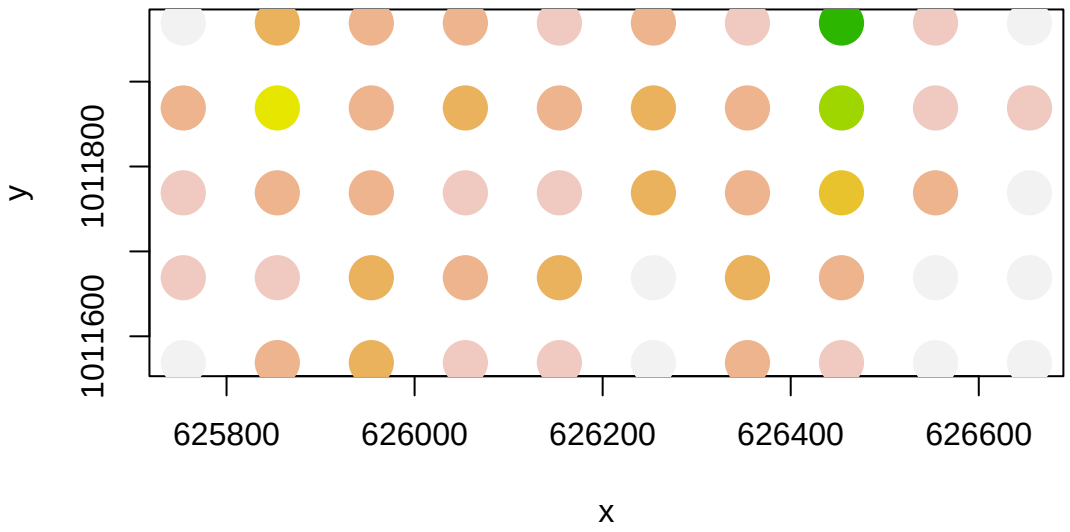
```
comm_sp
```

```
## [1] "Trichilia.tuberculata"
```

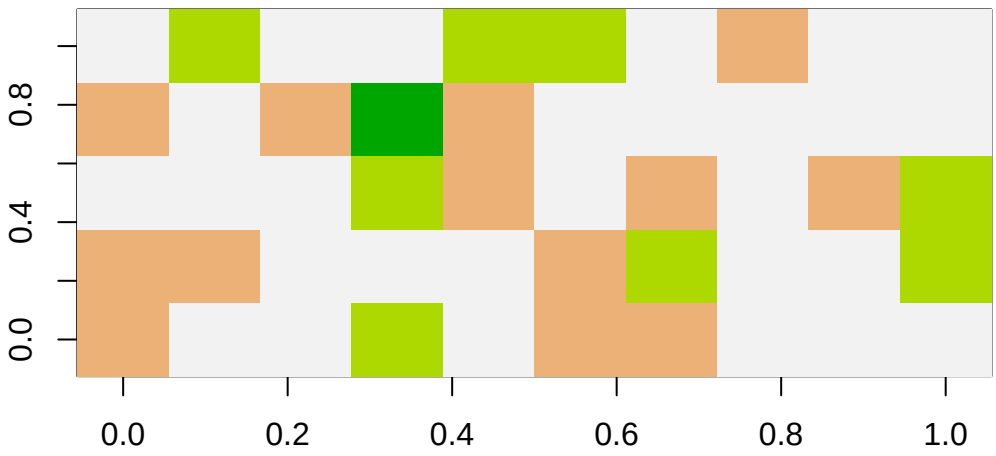
Now we simply need to plot the results. We'll start just by mapping the patterns to try to judge visually if their is a spatial pattern then we'll examine the bivariate patterns.

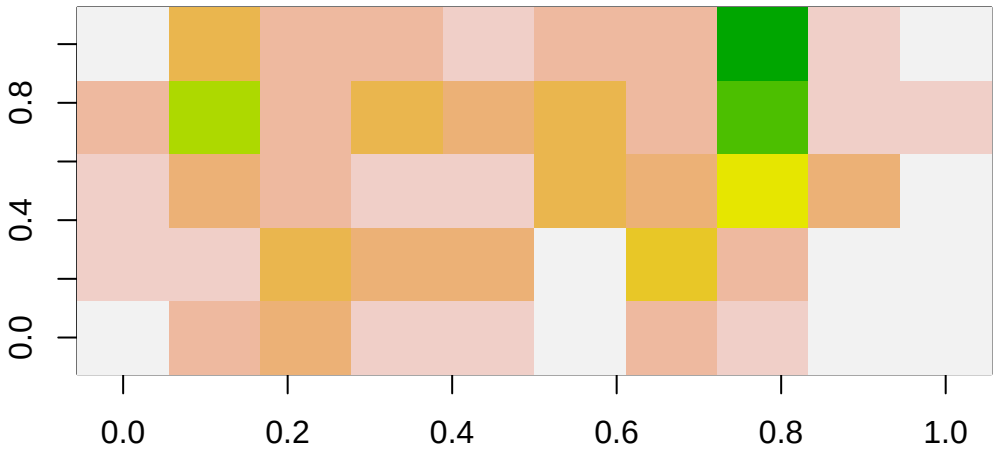
```
get_col_brks <- function(x) {
    col_brks <- hist(x, plot=F)$breaks
    col_indices <- as.numeric(cut(x, col_brks))
    cols <- rev(terrain.colors(length(col_brks)))
    cols[col_indices]
}

# one way to show the spatial patterns
par(mfrow=c(2,1))
plot(BCI_xy, type='n')
points(BCI_xy, col=get_col_brks(BCI[ , rare_sp]), pch=19, cex=3)
plot(BCI_xy, type='n')
points(BCI_xy, col=get_col_brks(BCI[ , comm_sp]), pch=19, cex=3)
```






```
# here is a better way that doesn't require guessing on the point size
par(mfrow=c(2,1))
image(matrix(BCI[ , rare_sp], nrow=10, ncol=5, byrow = TRUE),
      col=rev(terrain.colors(12)))
image(matrix(BCI[ , comm_sp], nrow=10, ncol=5, byrow = TRUE),
      col=rev(terrain.colors(12)))
```

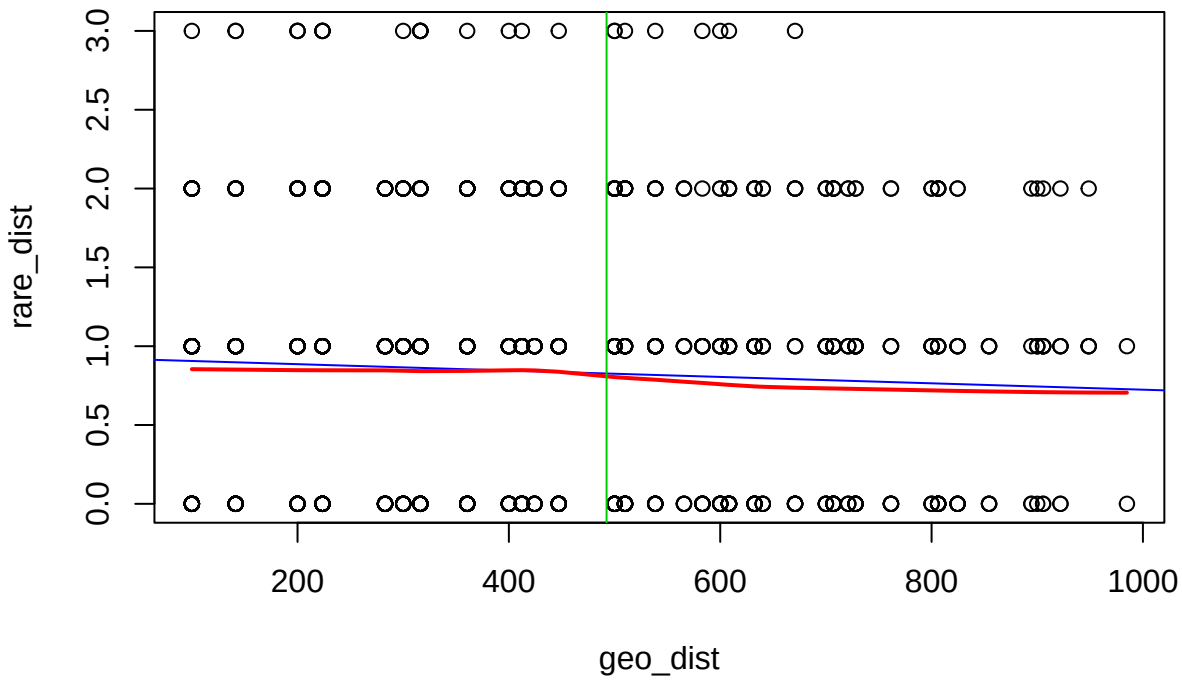




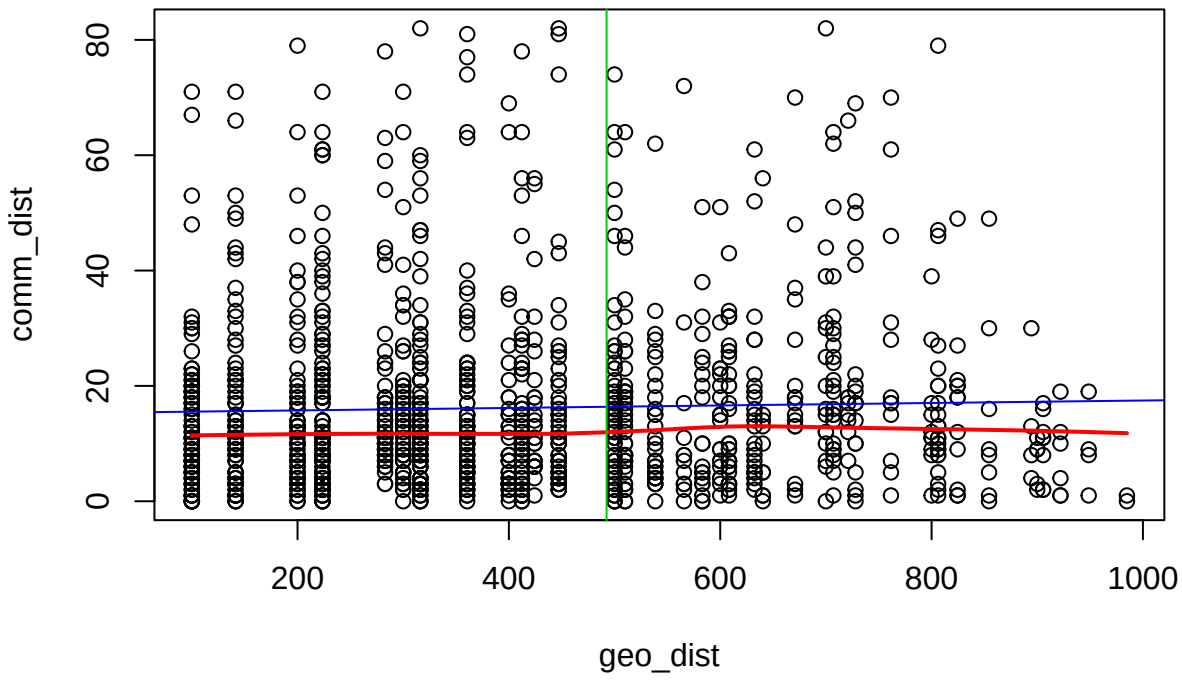
Based on these simple maps it does not look like there is a strong spatial signal in either the rare or common species. Let's look at bivariate relationships between variance in abundance and geographic distance. If the relationship is positive then this would be a signal of spatial autocorrelation.

```
# Now let's examine the bivariate relationship between spatial and ecological
# distance
rare_dist <- dist(BCI[ , rare_sp])
comm_dist <- dist(BCI[ , comm_sp])
```

```r
geo_dist <- dist(BCI_xy)

rare_lm <- lm(rare_dist ~ geo_dist)
comm_lm <- lm(comm_dist ~ geo_dist)

max_dist <- max(geo_dist) / 2

plot(geo_dist, rare_dist)
abline(rare_lm, col='blue')
lines(lowess(as.vector(geo_dist), as.vector(rare_dist)), col='red', lwd=2)
abline(v=max_dist, col='green3')
```



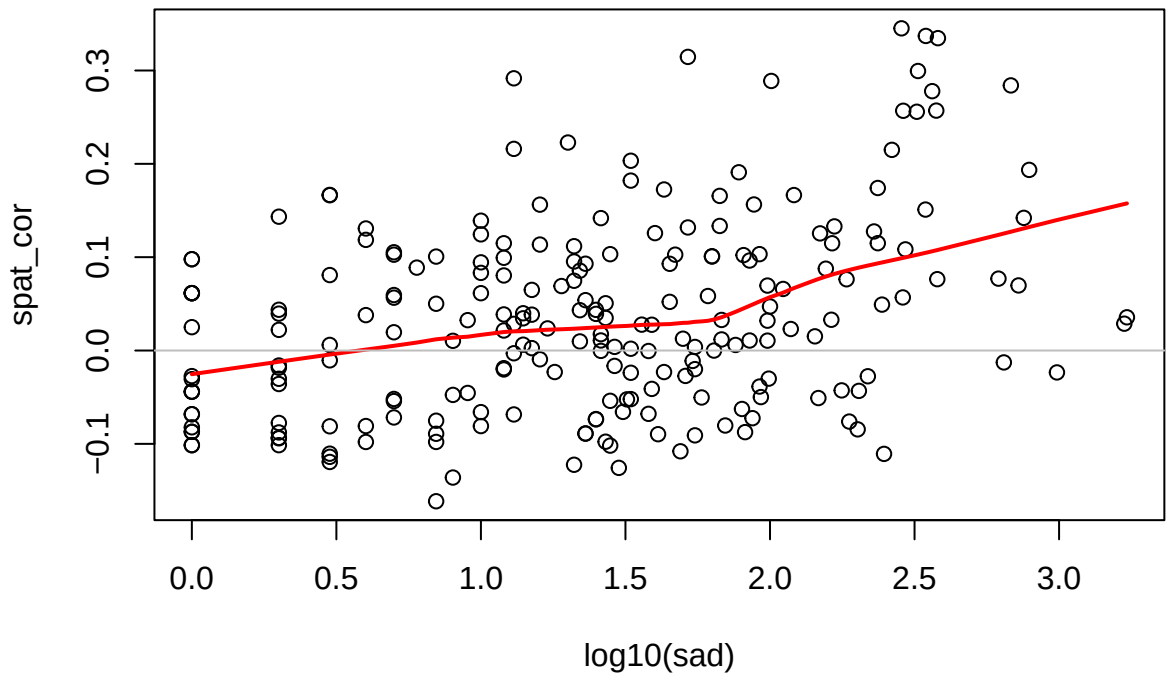

```r
plot(geo_dist, comm_dist)
abline(comm_lm, col='blue')
lines(lowess(geo_dist, comm_dist), col='red', lwd=2)
abline(v=max_dist, col='green3')
```



Those lines are not very different from the null model so for these two specific species it looks like neither one has a strong spatial signal. We could test these correlations using a Mantel test but it really seems like that is not necessary based on how flat those lines appear.

Optional analysis To examine for general relationship between abundance and spatial patterning we can easily compute the correlation of every species with spatial distance

```
spat_cor <- apply(BCI, 2, function(x)
                  cor(geo_dist, dist(x)))

par(mfrow=c(1,1))
plot(log10(sad), spat_cor)
lines(lowess(log10(sad), spat_cor), col='red', lwd=2)
abline(h=0, col='grey')
```



It appears there is a weak positive correlation between abundance and the correlation between spatial distance and species distance. This analysis also reveals that this slope is typically positive (i.e., increasing spatial distance decreases the chance of encountering a conspecific) which is a signature of spatial dependence / auto-correlation.

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:

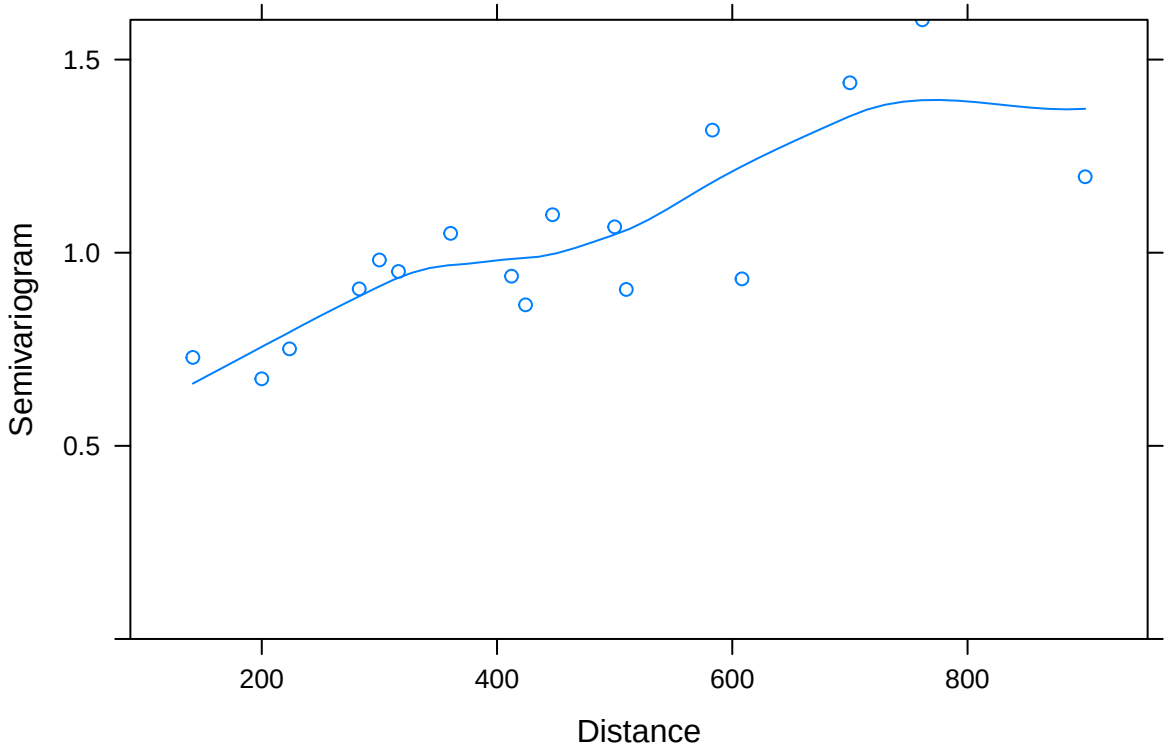
```
sp_ids <- c("Cordia.lasiocalyx", "Hirtella.triandra",
            "Picramnia.latifolia", "Quassia.amara",
            "Tabernaemontana.arborea", "Trattinnickia.aspera",
            "Xylopia.macrantha")
```

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

- 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.

```
# put the species data together with the coordinate
# data so that the model can easily have access to both
# kinds of variables.
bci_dat <- data.frame(BCI, BCI_xy)

# specify single and all species models
# arbitrarily chose the first species in sp_ids as the single variable
# predictor
sing_mod_formula <- as.formula(paste("Drypetes.standleyi ~ ",
                                     paste(sp_ids[1], collapse= "+")))
full_mod_formula <- as.formula(paste("Drypetes.standleyi ~ ",
                                     paste(sp_ids, collapse= "+")))

sing_ns_mod <- gls(sing_mod_formula, data=bci_dat)

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



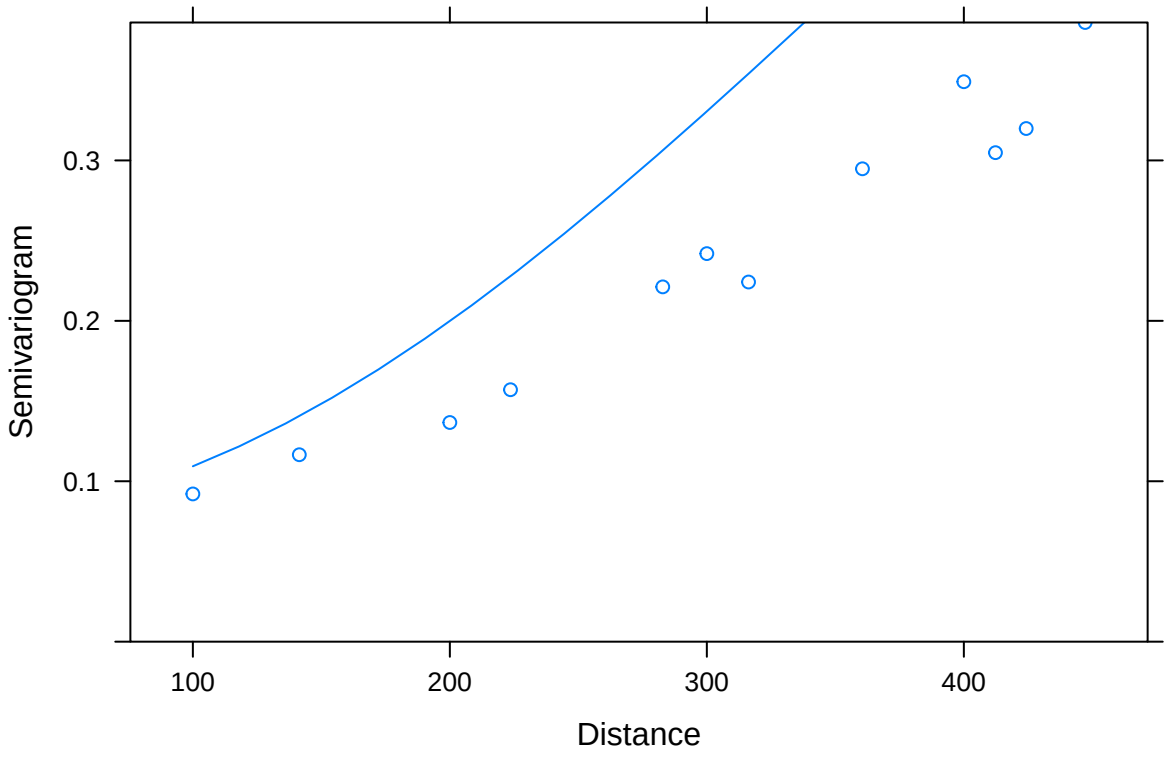

```
sing_sp_mods <- get_spat_mods(sing_ns_mod)
```

```
## Error in gls(model = sing_mod_formula, data = bci_dat, correlation = corExp(form = ~x +  :
##   false convergence (8)
```

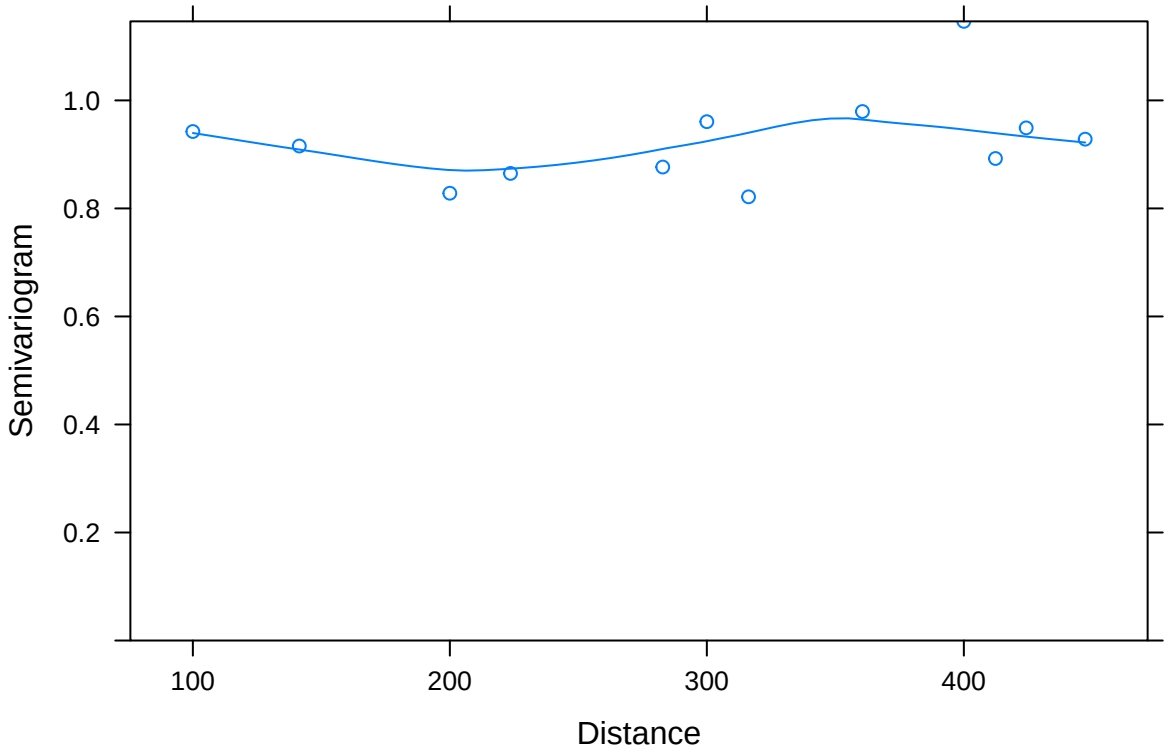
```
get_spat_AIC(sing_sp_mods)
```

```
##    mods AIC_no_nug  AIC_nug
## 1   Exp   310.6438       NA
## 2  Gaus   326.5514 308.8879
## 3   Lin   326.0101 311.0547
## 4 Ratio   318.9170 308.6589
## 5 Spher   310.6438 311.0547
```

```
# Quite a few models cannot be eliminated b/c AIC's are within two.
# Now one should examine the 'best' models visually.
# The spherical model hits the bulk of the points the best it appears

plot(Variogram(sing_sp_mods$Gaus$nug, form = ~ x + y, maxDist = max_dist))
```




```
plot(Variogram(sing_sp_mods$Gaus$nug, form = ~ x + y, resType = 'n',
               maxDist = max_dist))
```




```
anova(sing_ns_mod, sing_sp_mods$Spher$nug)
```

```
##                         Model df      AIC      BIC    logLik   Test  L.Ratio
## sing_ns_mod                 1  3 335.1246 340.7382 -164.5623
## sing_sp_mods$Spher$nug      2  5 311.0547 320.4107 -150.5274 1 vs 2 28.06989
##                         p-value
## sing_ns_mod
## sing_sp_mods$Spher$nug  <.0001
```

```
round(summary(sing_ns_mod)$tTable, 2)
```

```
##                   Value Std.Error t-value p-value
## (Intercept)       -2.78      1.92   -1.45    0.15
## Cordia.lasiocalyx  1.17      0.23    5.08    0.00
```

```
round(summary(sing_sp_mods$Spher$nug)$tTable, 2)
```

```
##                   Value Std.Error t-value p-value
## (Intercept)       10.71    340.27    0.03    0.98
## Cordia.lasiocalyx  0.17      0.21    0.81    0.42
```

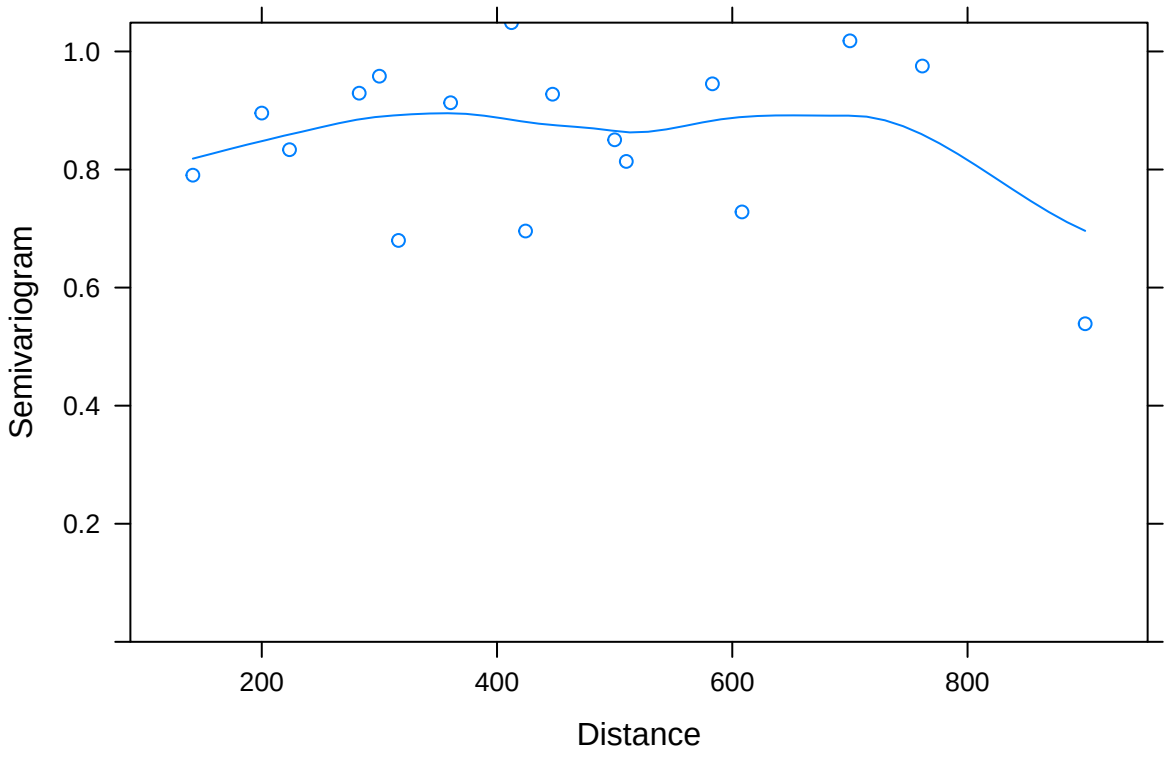
```
# we can also look at pseudo R2 values
pseudo_r2(sing_ns_mod)
```

```
## [1] 0.05724602
```

```
pseudo_r2(sing_sp_mods$Spher$nug, update(sing_ns_mod, . ~ + 1))
```

```
## [1] 0.1376502
```

```
full_ns_mod <- gls(full_mod_formula, data=bci_dat)

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



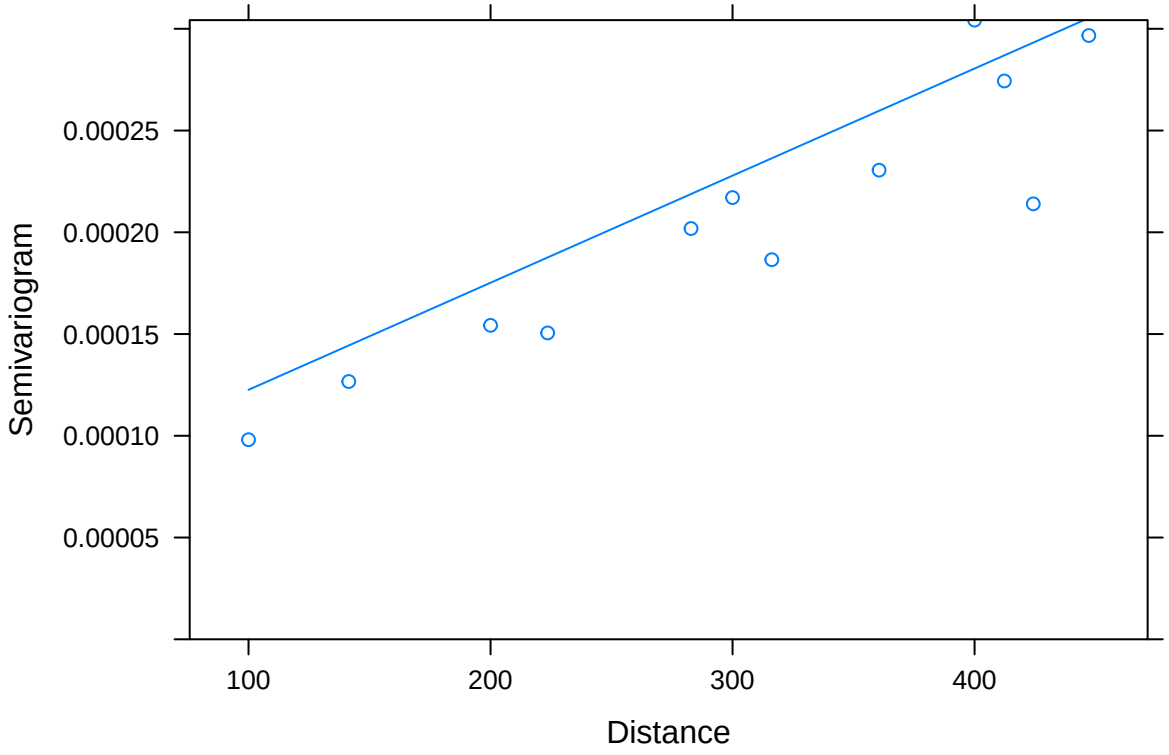

```
#spatial dependence appears non-existant

full_sp_mods <- get_spat_mods(full_ns_mod)
```

```
## Error in gls(model = full_mod_formula, data = bci_dat, correlation = corExp(form = ~x +  :
##   false convergence (8)
```

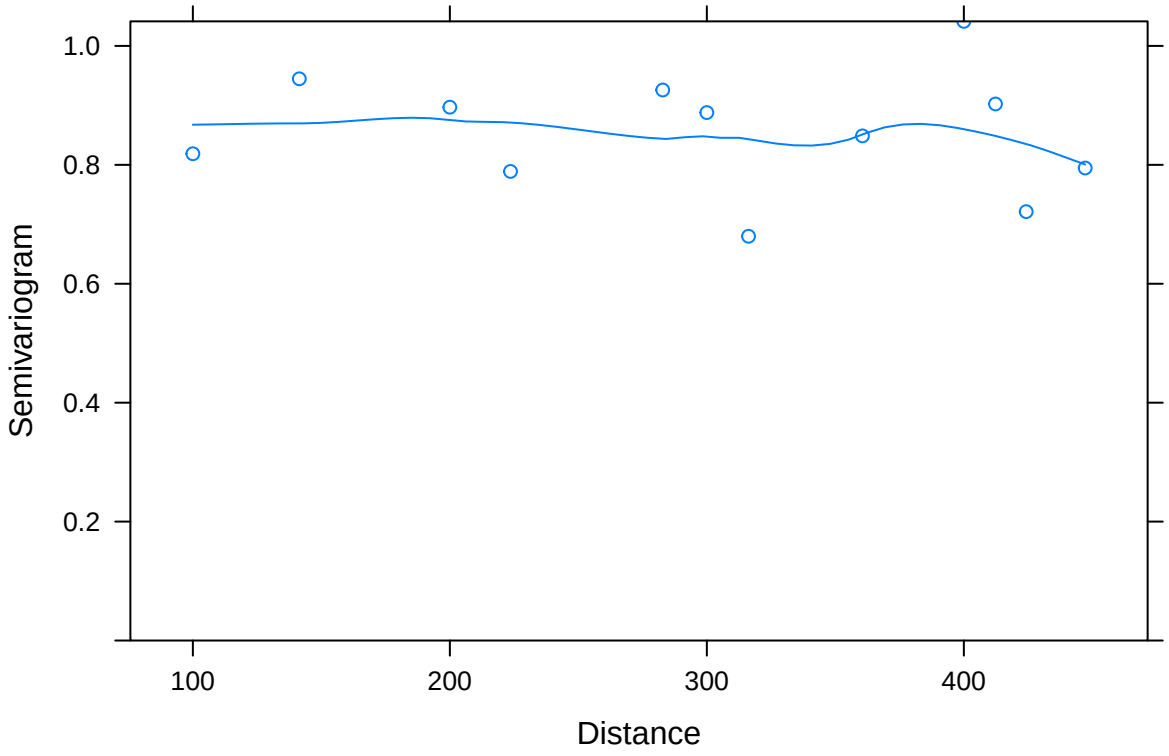
```
get_spat_AIC(full_sp_mods)
```

```
##    mods AIC_no_nug  AIC_nug
## 1   Exp   301.6062       NA
## 2  Gaus   307.2070 303.8654
## 3   Lin   307.7308 309.7308
## 4 Ratio   303.8542 303.1486
## 5 Spher   301.9254 301.9592
```

```
plot(Variogram(full_sp_mods$Spher$nug, form = ~ x + y, maxDist = max_dist))
```




```
plot(Variogram(full_sp_mods$Spher$nug, form = ~ x + y, resType = 'n',
               maxDist = max_dist))
```




```
anova(full_ns_mod, full_sp_mods$Spher$nug)
```

```
##                         Model df      AIC      BIC    logLik   Test  L.Ratio
## full_ns_mod                 1  9 307.1163 322.7554 -144.5582
## full_sp_mods$Spher$nug      2 11 301.9592 321.0735 -139.9796 1 vs 2 9.157175
##                         p-value
## full_ns_mod
## full_sp_mods$Spher$nug  0.0103
```

```
round(summary(full_ns_mod)$tTable, 2)
```

```
##                          Value Std.Error t-value p-value
## (Intercept)              -1.05      2.12   -0.50    0.62
## Cordia.lasiocalyx         0.43      0.20    2.10    0.04
## Hirtella.triandra         0.12      0.08    1.52    0.14
## Picramnia.latifolia       0.66      0.64    1.04    0.30
## Quassia.amara             4.09      2.28    1.79    0.08
## Tabernaemontana.arborea -0.25      0.15   -1.67    0.10
## Trattinnickia.aspera      1.35      0.71    1.89    0.07
## Xylopia.macrantha         0.55      0.15    3.74    0.00
```

```
round(summary(full_sp_mods$Spher$nug)$tTable, 2)
```

```
##                          Value Std.Error t-value p-value
```

```
## (Intercept)              3.05    334.06    0.01    0.99
## Cordia.lasiocalyx        0.14      0.19    0.75    0.46
## Hirtella.triandra        0.00      0.09   -0.02    0.98
## Picramnia.latifolia      0.29      0.53    0.54    0.59
## Quassia.amara            1.33      1.94    0.68    0.50
## Tabernaemontana.arborea  0.04      0.14    0.29    0.77
## Trattinnickia.aspera     1.82      0.57    3.17    0.00
## Xylopia.macrantha        0.41      0.15    2.66    0.01
```

```
# examine pseudo R^2 values
pseudo_r2(full_ns_mod)
```

```
## [1] 0.1718468
```

```
pseudo_r2(full_sp_mods$Spher$nug, update(full_ns_mod, . ~ + 1))
```

```
## [1] 0.1980769
```

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

In the case of the single species model including the spatial term had a large influence on the estimated beta coefficients and effect sizes. The effect went from highly significant to not significant. The spatial model did not have very much influence on the coefficients of the model that included all the species as predictors.

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

The answer to this is technically yes for both models, but the spatial model is relatively much more favored in the single predictor modeling context.

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

We have already verified that many of the species in the dataset are spatially structured. If we include additional spatially structured predictor variables to model a spatially structured response then it should come as no surprise that when we include more predictors in our model that we will observe a decrease in not just the error variance but the spatially structured component of the error variance. This should not necessarily be interpreted as evidence of that the more complex model is doing a better job capturing the processes underlying the spatial dependence in the response.