---
title: "jess daly spatial hw"
output: html_document

1.

```{r}
library(vegan)
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))
```

First check out quantities for all species present in the dataset.

```{r}
BCIcol<-colSums (BCI, na.rm = FALSE, dims = 1)
```

```{r}
hist(BCIcol)
```

Choosing Alseis blackiana as common Andira inermis as rare species (check out total number of trees to get an idea of what common and rare mean here).

```{r}
sum(BCI [,11])
sum(BCI [,14])
```

```{r}
com<-BCI [,11]
rare<-BCI [,14]
```

First conduct spatial analysis of common species

```{r}
plot(BCI_xy, cex = com/max(com))
```

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

Most of the dots indicate that most patches have few trees, but towards the top right there is a cluster of high density, indicating this species may display spatial dependence.

```{r}
BCI_dist = dist(com)
xy_dist = dist(BCI_xy)
max_dist = max(BCI_xy) / 2
```

```{r}
plot(xy_dist, BCI_dist)
abline(lm(BCI_dist ~ xy_dist), lwd=3, col='red')
lines(lowess(xy_dist, BCI_dist), lwd=3, col='pink')
abline(v = max_dist, col='red', lwd=3, lty=2)1
```

```{r}
obs_cor = cor(xy_dist, BCI_dist)
obs_cor
```

```{r}
nperm = 1000
null_cor = obs_cor
for (i in 2:nperm) {
    tmp_xy = BCI_xy[sample(nrow(BCI_xy)), ]
    null_cor[i] = cor(dist(tmp_xy), BCI_dist)
}

sum(null_cor >= obs_cor) / nperm
```

My p-value indicates there is no significant correlation here. Taking a look at the plot, there doesn't appear to be much there in terms of trends in the data.

```{r}
BCI_mantel = mantel(BCI, BCI_dist)
BCI_mantel
```

The mantel test returns a value very close to 0, indicating a lack of correlation. Overall there seems to be a lack of evidence for spatial dependance in Alseias blackiana.This makes sense because this is a commonly found species, so we would expect to find it everywhere and not in specific patches.

I will perform my analysis again for my rare species.

```{r}
plot(BCI_xy, cex = rare/max(rare))
```

```{r}
col_brks1 = hist(rare, plot=F)$breaks
col_indices1 = as.numeric(cut(rare, col_brks1))
cols1 = rev(terrain.colors(length(col_brks1)))
plot(BCI_xy, cex=2, pch=19, col=cols1[col_indices1])
```
There looks like there might be some pattern in dispersion since some patches have no trees at all, but it is hard to tell.

```{r}
BCI_dist1 = dist(rare)
xy_dist = dist(BCI_xy)
max_dist = max(BCI_xy) / 2
```

```{r}
plot(xy_dist, BCI_dist1)
abline(lm(BCI_dist1 ~ xy_dist), lwd=3, col='red')
lines(lowess(xy_dist, BCI_dist1), lwd=3, col='pink')
abline(v = max_dist, col='red', lwd=3, lty=2)1
```

```{r}
obs_cor1 = cor(xy_dist, BCI_dist1)
obs_cor1
```

```{r}
nperm = 1000
null_cor = obs_cor1
for (i in 2:nperm) {
    tmp_xy = BCI_xy[sample(nrow(BCI_xy)), ]
    null_cor[i] = cor(dist(tmp_xy), BCI_dist1)
}

sum(null_cor >= obs_cor) / nperm
```

The p-value indicates no significant correlation.

```{r}
BCI_mantel1 = mantel(BCI, BCI_dist1)
BCI_mantel1
```

The models returned no evidence of spatial dependence in either species. This makes less sense for the rare species, since you would expect that if it's rare it may only be able to live in very specific condtions and thus certain areas would have higer concentrations. However, the rarer the tree, the harder it is to detect patterns in spatial dependence, which may be the issue here.

2.

I am creating two linear models to predict the abundance of Drypetes standleyi. In the first one I am using a single species, Cordia lasiocalyx (randomly selected) as my predictor.

```{r}
sp_ids = c("Cordia.lasiocalyx", "Hirtella.triandra",
           "Picramnia.latifolia", "Quassia.amara",
           "Tabernaemontana.arborea", "Trattinnickia.aspera",
           "Xylopia.macrantha")
allIDS<-cbind("Cordia.lasiocalyx", "Hirtella.triandra",
           "Picramnia.latifolia", "Quassia.amara",
           "Tabernaemontana.arborea", "Trattinnickia.aspera",
           "Xylopia.macrantha", "Drypetes.standleyi")
```

```{r}
allpre <- apply(allIDS, 1, function(x) sum(x > 0))
pre_dat<-data.frame(allpre, BCI, BCI_xy)
sub_dat<-subset(pre_dat, select=allIDS)
```

```{r}
spe<-BCI$Drypetes.standleyi
pre<-BCI$Cordia.lasiocalyx
```

Variograms show variability between points as a function of distance. I will run a series of variograms to determine which best fits my data.

```{r}
BCI_dat = data.frame(spe, BCI_xy)

BCI_lm = gls(spe ~ pre, data=BCI_dat)

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

Points seem to fit this model quite well.

```{r}
res = residuals(BCI_lm)
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 = max_dist, col='red', lwd=3, lty=2)
```

```{r}
BCI_exp = update(BCI_lm, corr=corExp(form=~x + y))
```

```{r}
```

```
plot(Variogram(BCI_exp, maxDist = max_dist))
```

This one not so good. The model includes a nugget, or non-zero y-intercept. Next examine residuals for patterns.

```{r}
plot(Variogram(BCI_exp, resType='normalized', maxDist = max_dist))
```

The residuals show a positive trend. Ideally they should be randomly disributed and show no pattern, so this model may not be a good fit.

```{r}
BCI_exp_nug = update(BCI_exp, corr=corExp(form=~x + y, nugget=T))
plot(Variogram(BCI_exp_nug, maxDist = max_dist))
```

The same model with a nugget looks to fit this data very well.

```{r}
plot(Variogram(BCI_exp_nug, resType='n', maxDist = max_dist))
```

However, residuals still show a positive trend.

```{r}
BCI_rat_nug = update(BCI_lm, corr=corRatio(form=~x + y, nugget=T))
plot(Variogram(BCI_rat_nug, maxDist = max_dist))
```

The rational quadratic model is not a good fit.

```{r}
plot(Variogram(BCI_rat_nug, resType='n', maxDist = max_dist))
```

```{r}
anova(BCI_lm, BCI_exp, BCI_exp_nug, BCI_rat_nug, test=F)
```

Interestingly according to the ANOVA, all of the models are nearly equal in terms of fitting the data, though lm is a little worse than the others.

```{r}
summary(BCI_exp_nug)
```

```{r}
col_brks2 = hist(residuals(BCI_exp_nug), plot=F)$breaks
col_indices2 = as.numeric(cut(residuals(BCI_exp_nug), col_brks2))
cols2 = rev(terrain.colors(length(col_brks2)))
```

```
plot(BCI_xy, cex=2, pch=19, col=cols2[col_indices2])
```
Judging from the plot it certainly looks like there is a spatial pattern here, with a high concentration in the bottom right corner.

Now I will rerun my variograms using all 7 given species together as a predictor.

```{r}
x<-BCI_xy$x
y<-BCI_xy$y
```

```{r}
BCI_lm1 = gls(Drypetes.standleyi ~ Cordia.lasiocalyx + Hirtella.triandra + Picramnia.latifolia + Quassia.amara + Tabernaemontana.arborea + Trattinnickia.aspera + Xylopia.macrantha, data=sub_dat)

plot(Variogram(BCI_lm1, form= ~ x + y))
```

```{r}
res1 = residuals(BCI_lm1)
plot(dist(BCI_dat[, c('x', 'y')]), dist(res1))
lines(lowess(dist(BCI_dat[, c('x', 'y')]), dist(res1)), col='red', lwd=2)
abline(v = max_dist, col='red', lwd=3, lty=2)
```

```{r}
BCI_exp1 = update(BCI_lm1, corr=corExp(form=~x + y))
```

```{r}
plot(Variogram(BCI_exp1, maxDist = max_dist))
```

```{r}
plot(Variogram(BCI_exp1, resType='normalized', maxDist = max_dist))
```

```{r}
BCI_exp_nug1 = update(BCI_exp1, corr=corExp(form=~x + y, nugget=T))
plot(Variogram(BCI_exp_nug1, maxDist = max_dist))
```

```{r}
plot(Variogram(BCI_exp_nug1, resType='n', maxDist = max_dist))
```

```{r}
BCI_rat_nug1 = update(BCI_lm1, corr=corRatio(form=~x + y, nugget=T))
plot(Variogram(BCI_rat_nug1, maxDist = max_dist))
```

```{r}
plot(Variogram(BCI_rat_nug1, resType='n', maxDist = max_dist))
```

Similarly to when I used only a single species, the rational quadratic error model does not appear to be as useful as the others; however, it is the only model in which the residuals have no obvious pattern.

```{r}
anova(BCI_lm1, BCI_exp1, BCI_exp_nug1, BCI_rat_nug1, test=F)
```

Once again my ANOVA indicates all model are about even when it comes to representing the data at hand.

```{r}
summary(BCI_exp_nug1)
```

```{r}
col_brks3 = hist(residuals(BCI_exp_nug1), plot=F)$breaks
col_indices3 = as.numeric(cut(residuals(BCI_exp_nug1), col_brks3))
cols3 = rev(terrain.colors(length(col_brks3)))
plot(BCI_xy, cex=2, pch=19, col=cols3[col_indices3])
```

Once again there appears to be evidence of a spatial relationship in this plot, with higher concentrations on the far right side.

Using a single vs. multiple species overall produced similar results. Including the nugget (the spatial error term) in analysis did not change the effectivness of the models in either case. If you look at the variograms for the first 2 models, the scale of the y-axis is very small. This may cause the nugget effect to essentally become non-existant, since the value is so close to 0.