

Co-occurrence of tree species in a small tropical island

Venus Kuo and Michelle Benavidez, Quantitative Biodiversity, Indiana University

09 March, 2017

Directions:

1. Change “Add title of your project” on line 2 (above) to the name of your project.
2. Change “Add names of team members” on line 3 (above) to the names of team members in your project.
3. Write a <250 word Abstract describing your team’s research.
4. Write a 1-2 paragraph Introduction providing some background and context for your project.
5. Set up your working environment (set directory, load packages) using the `/Week8-Projects` folder.
6. Provide a brief (~1 paragraph) description of the dataset.
7. Load the dataset and process in ways necessary for any initial analyses.
8. Use well annotated R chunks to analyze data, make figures, and perform statistics needed to address your questions and hypotheses.
9. Write a Discussion and Conclusion section to address the questions and hypotheses that you outlined in your Abstract and Introduction.
10. Include citations in the Reference section.
11. Delete the text in this Directions section.
12. By Thursday March 9, 23:59, please submit the completed assignment by creating a **pull request** via GitHub. Your pull request should include this file `Project_assignment.Rmd` and the PDF output of `Knitr` (`Project_assignment.pdf`).

ABSTRACT

Perhaps the most well-documented neutral model of biodiversity is the Unified Neutral Theory of Biodiversity (UNTB), however there is some controversy concerning the utility of UNTB compared to other null models of biodiversity. UNTB postulates that species richness and abundance is generated exclusively through genetic drift between species. Consequently, one should expect that a community governed by UNTB would have predominantly random species associations at any given point in time. For this study we analyzed tree species associations on Barro Colorado Island in Panama. Rank abundance of all individuals in the study site followed a log normal distribution. Among sites, we found that only 21.3% of tree associations were nonrandom, of which 64.64% were positive associations and 35.36% were negative associations. The frequency of this distribution of associations was not significantly different from the frequencies of a randomly generated null distribution ($\chi^2 = 0.36$, $df = 2$, $p = 0.83$). Future research should focus on constructing and comparing a randomized null distribution model with empirical data to see if UNTB best explains this seemingly random distribution.

1) INTRODUCTION

Hubbell (2001) developed the unified Neutral Theory of Biodiversity (UNTB) to explain the diversity and abundances of species in ecological communities by assuming per capita ecological equivalence among all individuals of every species in the same trophic level. UNTB acts as a null hypothesis for niche-theory by stripping away the complexity and asking: How much of the patterns of ecological communities assembly is explained by species similarity rather than differences (Bell 2000; Hubbell 2001). Hubbell (2006) argues that ecological equivalence can easily evolve in high diversity, dispersal and recruitment limited communities, like in Barro Colorado Island in Panama. In a highly complex environment, species do not evolve or evolves slowly due to low pairwise encounter frequencies between any two species (Hubbell 2006). This idea of species co-occurrence is central to the fundamental assumptions of UNTB, and we aim to explore if indeed

this co-occurrences between species on BCI is just by random chance. We hypothesize that if community composition in highly diverse ecosystems is primarily determined by neutral processes, then on Barro Colorado Island we expect to observe predominantly random associations for tree species present among sites.

2) SETUP

```
## Setup work environment ##
rm(list = ls())

## Load packages ##
require("vegan")

## Loading required package: vegan
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.4-2
require("tidyverse")

## Loading required package: tidyverse
require("dplyr")

## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
## 
##     filter, lag
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
require("cooccur")

## Loading required package: cooccur
## Warning: package 'cooccur' was built under R version 3.3.3
require("spatstat")

## Loading required package: spatstat
## Warning: package 'spatstat' was built under R version 3.3.3
## Loading required package: nlme
##
## Attaching package: 'nlme'
## 
## The following object is masked from 'package:dplyr':
## 
##     collapse
## Loading required package: rpart
```

```

## 
## spatstat 1.49-0      (nickname: 'So-Called Software')
## For an introduction to spatstat, type 'beginner'

##
## Attaching package: 'spatstat'

## The following object is masked from 'package:lattice':
##
##     panel.histogram
library("RColorBrewer")

```

3) DESCRIPTION OF DATA

Barro Colorado Island (BCI) was formed in the middle of Gatun Lake (9.1521° N, 79.8465° W) as a result of the construction of the Panama Canal in Panama in the early 20th century. Since 1923 this island has been the focus of extensive ecological research. In 1980, a 50 hectare Forest Dynamics Plot was established for the purpose of studying a variety of aspects of tropical ecology. The first complete tree census was completed in 1982 and have been repeated every 10 years since. All trees and shrubs > 1 cm in diameter are identified and counted for this census.

We took 2010 census BCI 50ha plot dataset that contains long form individual ID number, Species name, Quadrat, PX and PY 50 ha coordinate grid, date censused, status, and DBH. I subsetted the dataset to just include the individual ID number, species name (Latin), Quadrat, and the PX and PY coordinates.

4) LOAD THE DATA

```

setwd("C:/Users/Michelle/Github/QB2017BenavidezKuo/data")
# From BCI, Panama from year 2010 #
BCI <- read.delim("BCI2010.txt", header=T)
# Slim down BCI data frame #
BCI.df <- subset(BCI, select = c(Latin, Quadrat))

```

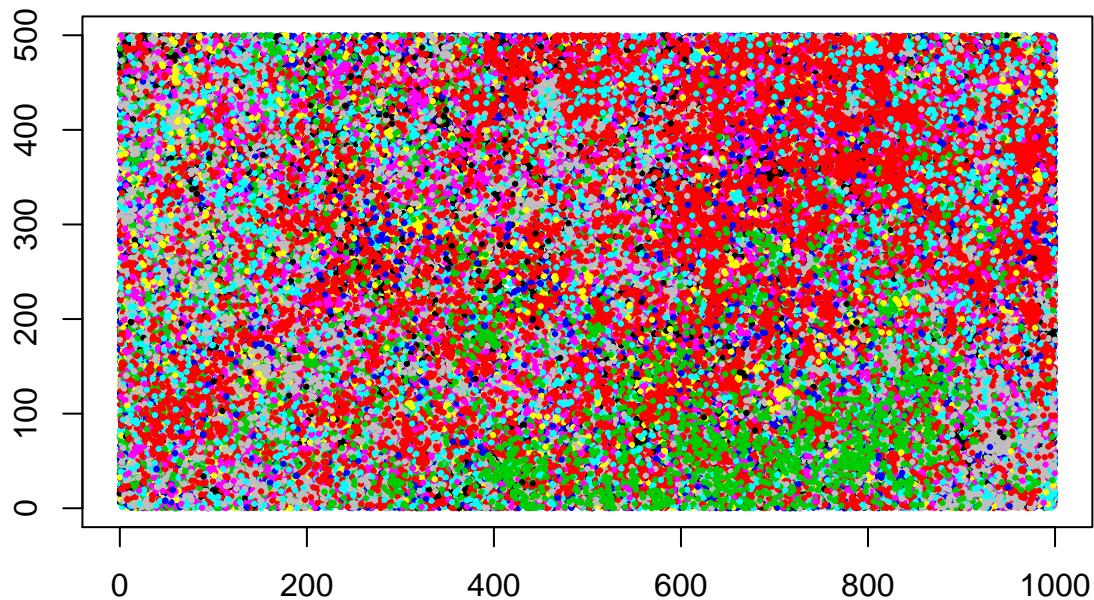
5) ANALYSIS: FIGURES AND STATISTICS

```

#pretty little chart of trees plotted of the plot
plot(BCI$PX,BCI$PY, col=as.numeric(BCI$Latin),pch = 19, cex = 0.3, xlab = "", ylab = "", main = "Tree S"

```

Tree Species on the 50 Hectare Plot in Barro Colorado Island



```
#Good's coverage
C <- function(x = ""){
  1 - (sum(x == 1) / rowSums(x))
}

#create species by site incidence matrix
BCI.SbyS <- group_by(BCI, Quadrat) %>% count(Latin) %>% spread(key=Latin, value=n , fill=0)
BCI.SBS <- BCI.SbyS[, -1]
BCI.SpbS=t(BCI.SBS)
BCI.SpbS[BCI.SpbS > 0] <- 1

#creating cooccur dataframe
BCIcooc <- cooccur::cooccur(mat = BCI.SpbS, type = "spp_site", thresh = T, spp_names = T)

##  
| | 0%  
| | 1%  
| | 0%  
| |= 1%  
|= 2%
```

==		2%
==		3%
==		4%
==		4%
==		5%
====		5%
====		6%
====		7%
=====		7%
=====		8%
=====		8%
=====		9%
=====		10%
=====		10%
=====		11%
=====		12%
=====		12%
=====		13%
=====		13%
=====		14%
=====		15%
=====		15%
=====		16%
=====		16%
=====		17%
=====		18%
=====		18%

-----	19%
-----	19%
-----	20%
-----	21%
-----	21%
-----	22%
-----	22%
-----	23%
-----	24%
-----	24%
-----	25%
-----	25%
-----	26%
-----	27%
-----	27%
-----	28%
-----	28%
-----	29%
-----	30%
-----	30%
-----	31%
-----	32%
-----	32%
-----	33%
-----	33%
-----	34%
-----	35%

-----	35%
-----	36%
-----	36%
-----	37%
-----	38%
-----	38%
-----	39%
-----	39%
-----	40%
-----	41%
-----	41%
-----	42%
-----	42%
-----	43%
-----	44%
-----	44%
-----	45%
-----	45%
-----	46%
-----	47%
-----	47%
-----	48%
-----	48%
-----	49%
-----	50%
-----	50%
-----	51%

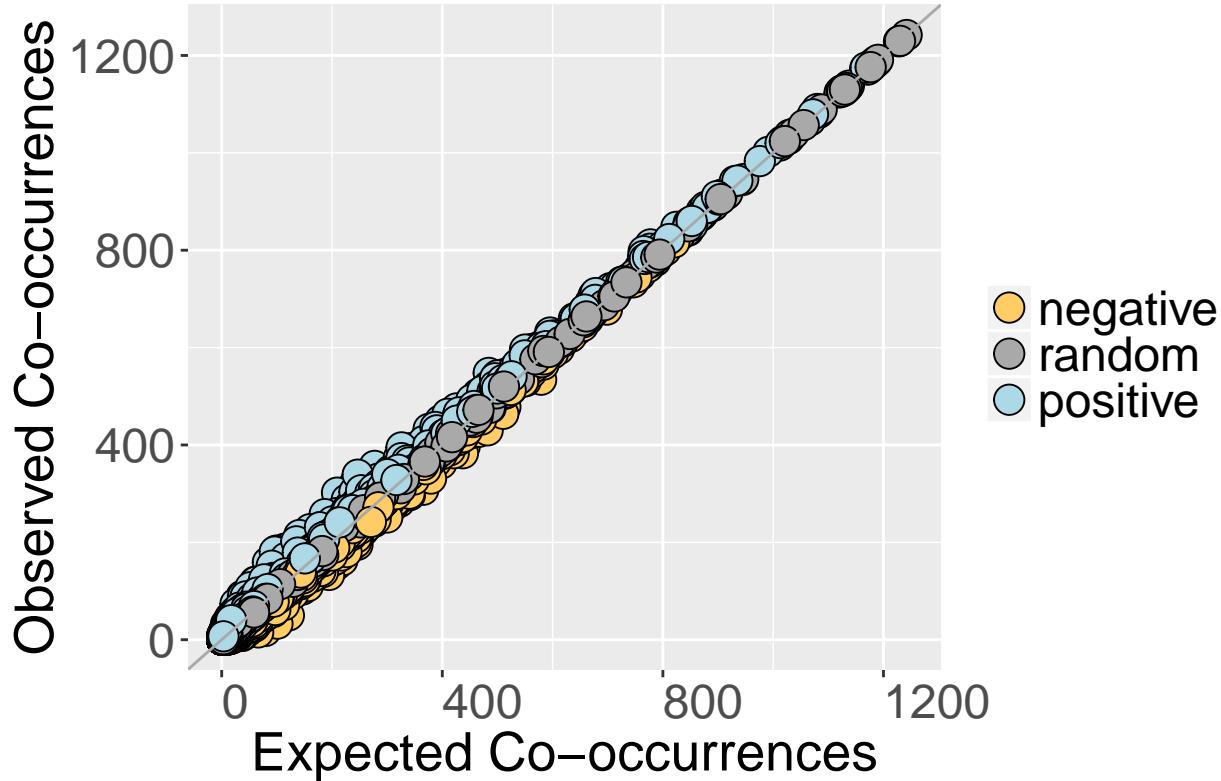
=====	52%
=====	52%
=====	53%
=====	53%
=====	54%
=====	55%
=====	55%
=====	56%
=====	56%
=====	57%
=====	58%
=====	58%
=====	59%
=====	59%
=====	60%
=====	61%
=====	61%
=====	62%
=====	62%
=====	63%
=====	64%
=====	64%
=====	65%
=====	65%
=====	66%
=====	67%
=====	67%

====	68%
====	68%
====	69%
====	70%
====	70%
====	71%
====	72%
====	72%
====	73%
====	73%
====	74%
====	75%
====	75%
====	76%
====	76%
====	77%
====	78%
====	78%
====	79%
====	79%
====	80%
====	81%
====	81%
====	82%
====	82%
====	83%
====	84%

=====	84%
=====	85%
=====	85%
=====	86%
=====	87%
=====	87%
=====	88%
=====	88%
=====	89%
=====	90%
=====	90%
=====	91%
=====	92%
=====	92%
=====	93%
=====	93%
=====	94%
=====	95%
=====	95%
=====	96%
=====	96%
=====	97%
=====	98%
=====	98%
=====	99%
=====	99%
=====	100%

```
#observed vs expected line graph - for observed distributions
cooccur::obs.v.exp(BCIcooc)
```

Observed–Expected Plot



```
#Random matrix assortment and cooccur dataframe
#NOTE: The results we present and report are representative of 99 sets of ranom associations. This c
#exp.df <- rep(NA, (3*99))
#exp.df <- matrix(exp.df, 3, 99)
#N <- 99
#for(i in 1:N) {
#  df <- NA
#  BCI.df.random <- transform(BCI.df, Quadrat = sample(Quadrat))
#  BCI.SbyS.r <- group_by(BCI.df.random, Quadrat) %>% count(Latin) %>% spread(key=Latin, value=n , fill=0)
#  BCI.SBS.r <- BCI.SbyS.r[, -1]
#  BCI.SpbS.r <- t(BCI.SBS.r)
#  BCI.SpbS.r[BCI.SpbS.r > 0] <- 1
#  BCIcooc.rand <- cooccur::cooccur(mat = BCI.SpbS.r, type = "spp_site", thresh = T, spp_names = T)
#  exp.neg <- BCIcooc.rand$negative
#  exp.pos <- BCIcooc.rand$positive
#  exp.rand <- BCIcooc.rand$random
#  exp.df[i] <- cbind(exp.neg, exp.pos, exp.rand)
#}

#chisquare GOF between frequencies of random results and observed results (Results are calculated from .
```

```

#rowMeans(exp.df, na.rm = FALSE, dims = 1)
summary(BCIcooc)

## Call:
## cooccur::cooccur(mat = BCI.SpbS, type = "spp_site", thresh = T,
## spp_names = T)
##
## Of 43956 species pair combinations, 16754 pairs (38.12 %) were removed from the analysis because exp
##
## Cooccurrence Summary:
##          Species      Sites    Positive    Negative     Random
##          297.0       1250.0     3742.0     2047.0     21413.0
## Unclassifiable Non-random (%)
##          0.0        21.3
## attr(),"class")
## [1] "summary.cooccur"
29457.5152+1628.4747+544.1414

## [1] 31630.13
exp.pos <- 1628.4747/31630.13
exp.neg <- 544.1414/31630.13
exp.rand <- 29457.5152/31630.13
obs.pos <- 3742/27202
obs.neg <- 2047/27202
obs.rand <- 21413/27202
#Test statistic value
x2 <- (((obs.pos-exp.pos)^2)/exp.pos) + (((obs.neg-exp.neg)^2)/exp.neg) + (((obs.rand-exp.rand)^2)/exp.
x2

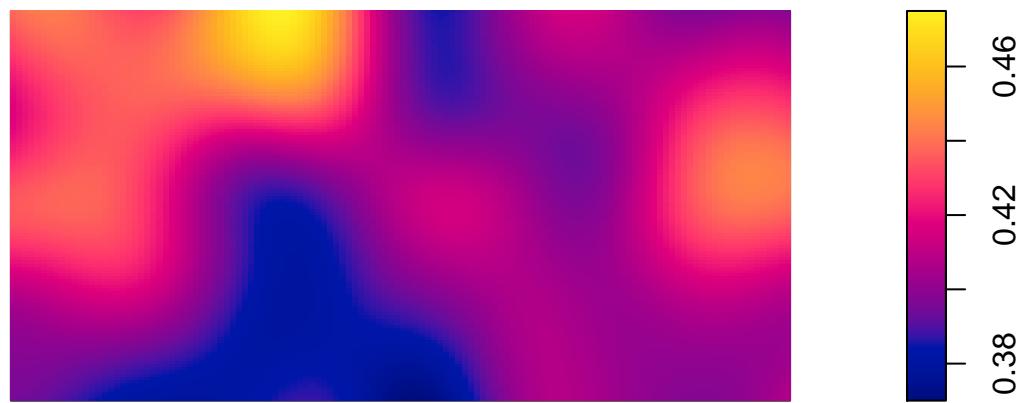
## [1] 0.3620926
# P-value on 2 df
1-pchisq(x2,2)

## [1] 0.8343967
#pretty pictures of what overall tree density looks like
x <- BCI$PX
y <- BCI$PY
p <- ppp(x,y,c(0,1000),c(0,500))

## Warning: data contain duplicated points
p2 <- unique(p)
den <- density(p2, sigma = 70)
plot(den)
plot(den, add = T, cex = 0.5)

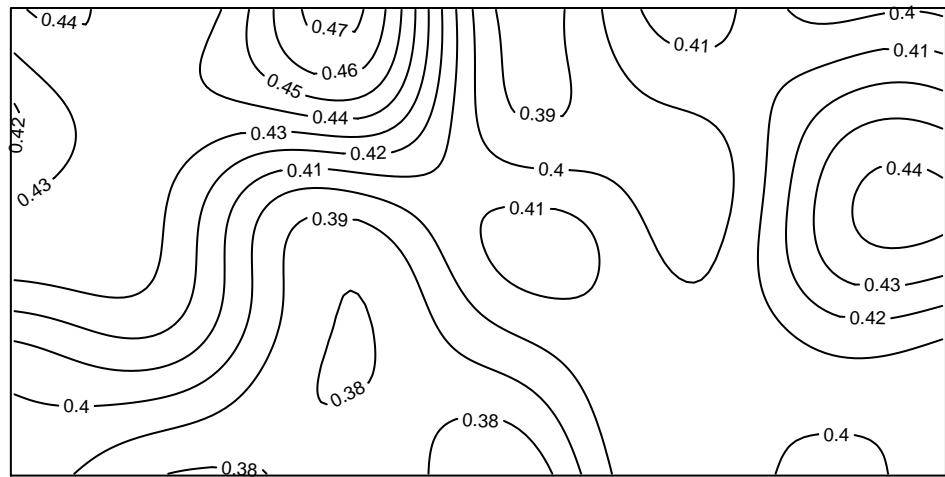
```

den

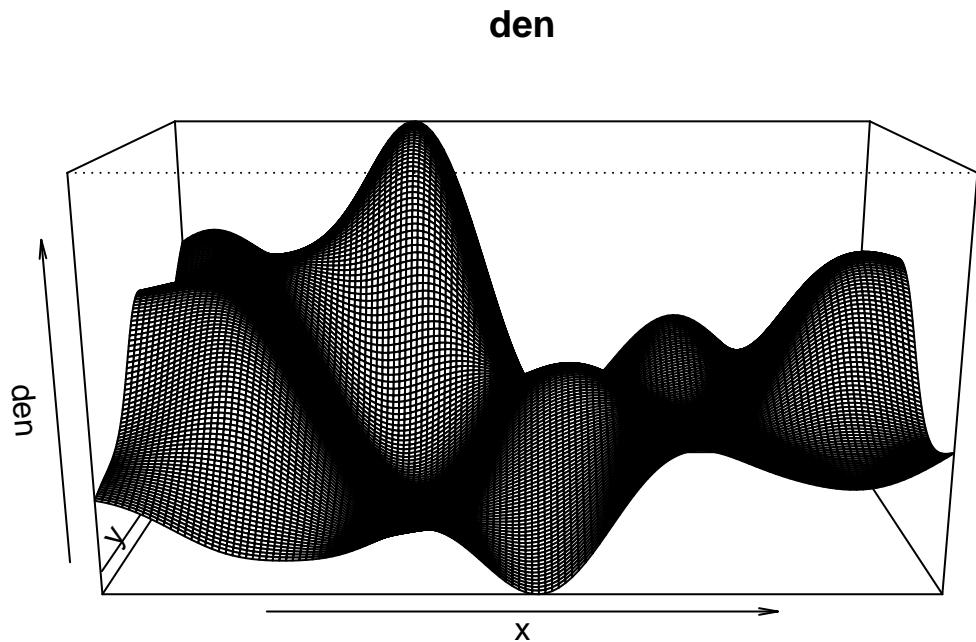


```
contour(den)
```

den



`persp(den)`

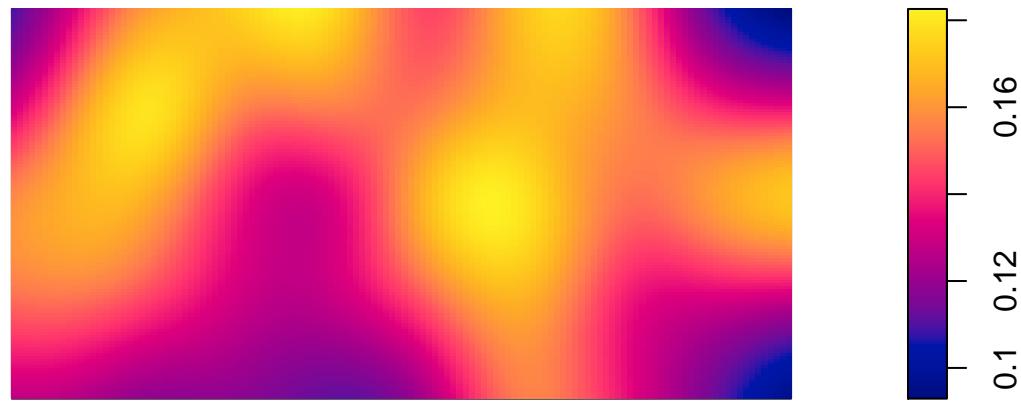


```
#map and density visualization for four most abundant trees
color <- brewer.pal(4,"Set1")
BCI.xy <- subset(BCI, select = c(Latin, PX, PY))
FA <- c("Hybanthus prunifolius","Faramea occidentalis","Desmopsis panamensis","Trichilia tuberculata")
Four <- BCI.xy[BCI.xy$Latin %in% FA,]

x2 <- Four$PX
y2 <- Four$PY
p2 <- ppp(x2,y2,c(0,1000),c(0,500))

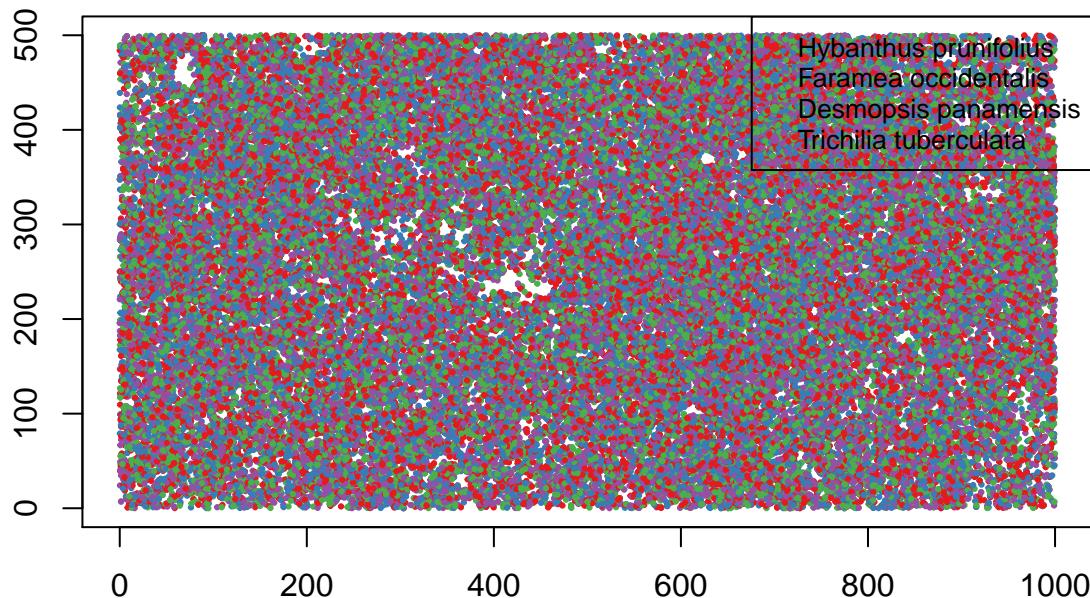
## Warning: data contain duplicated points
p22 <- unique(p2)
den2 <- density(p22, sigma = 70)
plot(den2)
plot(den2, add = T, cex = 0.5)
```

den2



```
plot(Four$PX,Four$PY, col=color,pch = 19, cex = 0.3, xlab = "", ylab = "", main = "Four most Abundant Sp  
legend('topright', legend = c("Hybanthus prunifolius","Faramea occidentalis","Desmopsis panamensis","Tr
```

our most Abundant Species on the 50 Hectare Plot in Barro Colorado I



6) DISCUSSION AND CONCLUSION

Based on the results of this study we have accepted our alternative hypothesis. Tree community composition on Barro Colorado Island in Panama consists of predominantly random associations among sites within the 50 hectare plot. This conclusion was further supported by the random distribution of community indicicence following a similar pattern. These results justify a more in-depth analysis of the mechanisms that may drive the neutral community arrangement of trees in this area. A limitation of this study was the lack of investigation into associations based on species abundance within sites. Instead, this study focused on the incidence of species among sites. In the future it will be interesting to study associations between tree species on the individual level using abundance data that includes an analysis that estimates the likilhood of homospecific associations. Creating a randomized null model for tree individual in the 50 ha to test against observed empirical results would provide greater clarity to the co-occurrence patterns observed in BCI. Furthermore, a more in-depth study of mechanisms that are driving community composition would provide a better indiciation of whether tree community composition on BCI is regulated by the assumptions of the Unified Neutral Theory of Biodiversity. Overall, since the need for tropical forest conservation is increasing in the face of global change, understading more about the process of community stucture will aid in developing more robust mitigation stratigies for maximizing diversity in tropical ecosystems.

7) REFERENCES

Bell, Graham. "The distribution of abundance in neutral communities." *The American Naturalist* 155.5 (2000): 606-617.

Griffith, Veech, and Marsh (2016). cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software*, 69(2), 1-17. doi:10.18637/jss.v069.c02

Hubbell, Stephen P. "Neutral theory and the evolution of ecological equivalence." *Ecology* 87.6 (2006): 1387-1398.

Hubbell, S.P. (2001). A Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press. Princeton NJ.

Veech (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2012.00789.