Assignment: Temporal Diversity

Erik Parker; Z620: Quantitative Biodiversity, Indiana University

14 February, 2017

OVERVIEW

In this Assignment, we extend our understanding of diversity from the spatial dimension to the temporal dimension.

After completing this exercise you will know how to:

1. wrangle a large dataset to visualize and analyze time series data
2. test hypotheses from experiments with temporal data
3. quantify temporal β-diversity and stability

Directions:

1. Change "Student Name" on line 3 (above) with your name.
2. Complete as much of the exercise as possible during class; what you do not complete in class will need to be done on your own outside of class.
3. Use the Handout as a guide; it contains a more complete description of data sets along with the proper scripting needed to carry out the exercise.
4. Be sure to **answer the questions** in this exercise document; they also correspond to the Handout. Space for your answer is provided in this document and indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">".
5. Before you leave the classroom, **push** this file to your GitHub repo.
6. When you are done with the Assignment, **Knit** the text and code into a html file.
7. After Knitting, please submit the completed Assignment by creating a **pull request** via GitHub. Your pull request should include this file *temporal_assignment.Rmd* and the html output of `Knitr` (*temporal_assignment.html*).

1) R SETUP

Typically, the first thing you will do in either an R script or an RMarkdown file is setup your environment. This includes things such as setting the working directory and loading any packages that you will need.

In the R code chunk below, provide the code to:

1. clear your R environment,
2. print your current working directory,
3. set your working directory to your "*/Week5-Temporal*" folder, and
4. load any packages you need to complete the assignment.

```
## [1] "/media/Datas/GitHub/QB2017_Parker/Week5-Temporal"
```

```
## This is vegan 2.4-2
```

```
##
## 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

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggplot2':
##
##     ggsave

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths

##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'timeDate'

## The following object is masked from 'package:xtable':
##
##     align

## This is forecast 7.3
```

```
##
## Attaching package: 'forecast'
```

```
## The following object is masked from 'package:nlme':
##
##     getResponse
```

```
## BiodiversityR 2.8-0: Use command BiodiversityRGUI() to launch the Graphical User Interface and to lea
```

2) LOADING DATA

Load dataset

In the R code chunk below, do the following:

1. load the `portal` dataset from in the "*/Week5/data*" folder, and
2. explore the structure of the dataset.

```
portal <- read.table("data/combined.csv", sep = ",", header = TRUE)

min(portal$plot_id)
```

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

```
max(portal$plot_id)
```

```
## [1] 24
```

```
str(portal$species_id)
```

```
##  Factor w/ 48 levels "AB","AH","AS",..: 16 16 16 16 16 16 16 16 16 16 ...
```

Question 1: Describe some of the attributes of the `portal` dataset.

a. How many plots are in `portal`?
b. How many rodent species are there in the `portal` dataset?

Answer 1a:
There are 24 plots in the portal dataset.

Answer 1b:
There are 48 distinct rodent species in the portal dataset.

3) WRANGLING THE PORTAL DATASET

In the R code chunk below, do the following:

1. Create a site-by-species matrix for any year of your choosing.
2. Create a vector of plot_type for sites in the site-by-species matrix.
3. Analyze alpha diversity (e.g., Shannon/Simpson) across the sites for that year.
4. Create a PCoA ordination of your site-by-species matrix.
5. Using the hypothesis testing tools you learned in the beta-diversity module, test the hypothesis that species abundances across sites vary as a factor of treatment type (i.e., plot_type).

```
#Creating site by species matrix for 1985
site.by.species <- group_by(portal, year, plot_id) %>%
  count(species_id) %>% spread(key = species_id, value = n, fill = 0)

site.by.species.1985 <- dplyr::filter(site.by.species, year == 1985)
```

```
site.by.species.1985 <- site.by.species.1985[,-1]

#Finding plot type for sites in the 1985 matrix
plot.type <- group_by(portal, year, plot_id) %>% filter(year == 1985) %>% count(plot_type) %>% ungroup()

#Adding plot types to matrix
site.by.species.1985.type <- cbind(site.by.species.1985, plot.type)

#Looking at alpha diversity
diversity(site.by.species.1985)
```

```
##  [1] 1.2115088 1.5760344 1.9870671 1.5248765 1.3252108 1.5941474 1.3562335
##  [8] 1.5418309 1.4475500 1.0825349 1.8138726 1.7378588 1.5416549 1.5698005
## [15] 1.3612311 1.5418216 1.6802446 1.4125256 1.7146247 1.7053412 1.6248774
## [22] 1.6980303 0.8855148 1.6204370
```

```
diversity(site.by.species.1985, index = "simpson")
```

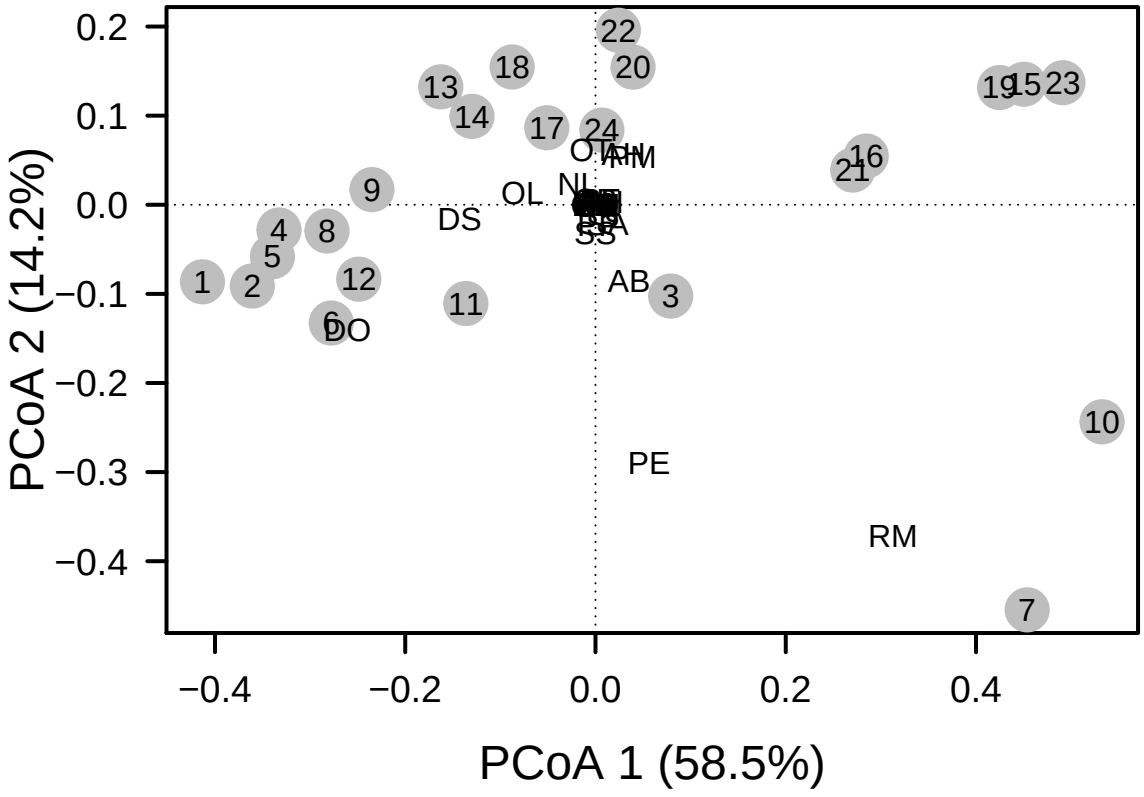
```
##  [1] 0.6496371 0.6637593 0.8304359 0.7171530 0.6726789 0.7216091 0.7138398
##  [8] 0.7350082 0.6635183 0.6157025 0.7700317 0.7292308 0.6957633 0.7129079
## [15] 0.6447188 0.7376562 0.7725181 0.6515705 0.7275000 0.7761111 0.7520827
## [22] 0.7430796 0.4719927 0.7040816
```

```
diversity(site.by.species.1985, index = "invsimpson")
```

```
##  [1] 2.854183 2.974060 5.897474 3.535480 3.055104 3.592072 3.494545
##  [8] 3.773702 2.971930 2.602151 4.348425 3.693182 3.286915 3.483203
## [15] 2.814672 3.811791 4.395954 2.870021 3.669725 4.466501 4.033604
## [22] 3.892256 1.893913 3.379310
```

```
#PCoA ordination of site by species matrix
portal.db <- vegdist(site.by.species.1985, method = "bray")
portal.pcoa <- cmdscale(portal.db, eig = TRUE, k = 3)

explainvar1 <- round(portal.pcoa$eig[1] / sum(portal.pcoa$eig), 3) * 100
explainvar2 <- round(portal.pcoa$eig[2] / sum(portal.pcoa$eig), 3) * 100
explainvar3 <- round(portal.pcoa$eig[3] / sum(portal.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)


par(mar = c(5, 5, 1, 2) + 0.1)

plot(portal.pcoa$points[ ,1], portal.pcoa$points[ ,2],
  xlab = paste("PCoA 1 (", explainvar1, "%)", sep = ""),
  ylab = paste("PCoA 2 (", explainvar2, "%)", sep = ""),
  pch = 16, cex = 2.0, type = "n", cex.lab = 1.5, cex.axis = 1.2, axes =   FALSE)

axis(side = 1, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
axis(side = 2, labels = T, lwd.ticks = 2, cex.axis = 1.2, las = 1)
abline(h = 0, v = 0, lty = 3)
box(lwd = 2)

#Row names didn't show up by themselves, needed to assign them, so assined each row(site) the appropria
row.names(portal.pcoa$points) <- 1:24

points(portal.pcoa$points[ ,1], portal.pcoa$points[ ,2],
```

```
pch = 19, cex = 3, bg = "gray", col = "gray")
text(portal.pcoa$points[ ,1], portal.pcoa$points[ ,2],
labels = row.names(portal.pcoa$points))

portalREL <- site.by.species.1985
for(i in 1:nrow(site.by.species.1985)){
portalREL[i, ] = site.by.species.1985[i, ] / sum(site.by.species.1985[i, ])
}

#Wouldn't work with portalREL as a data frame which it was by default, had to change to matrix.
portalREL <- as.matrix(portalREL)

portal.pcoa <- add.spec.scores(portal.pcoa,portalREL,method = "pcoa.scores")
text(portal.pcoa$cproj[ ,1], portal.pcoa$cproj[ ,2],
  labels = row.names(portal.pcoa$cproj), col = "black")
```




```
#Test hypothesis that species abundances vary as factor of treatment type

#Had to change plot.type to a matrix from a data frame(?) for it to work in the adonis function.
plot.type.perm <- as.matrix(plot.type)

adonis(site.by.species.1985 ~ plot.type.perm, method = "bray", permutations = 999)
```

```
##
## Call:
## adonis(formula = site.by.species.1985 ~ plot.type.perm, permutations = 999,      method = "bray")
##
## Permutation: free
## Number of permutations: 999
```

```
##
## Terms added sequentially (first to last)
##
##                Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## plot.type.perm  4    1.5215 0.38037  3.3543 0.41389  0.003 **
## Residuals      19    2.1545 0.11340         0.58611
## Total          23    3.6760                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 2: Describe how different biodiversity estimates vary among sites.

a. Does diversity vary among sites? Does this correspond to treatment type?
b. Is treatment type a significant predictor of site dissimilarity?

> ***Answer 2a***: From the ordination constructed, and the alpha diversity indices calculated for each site, it is clear that diversity does indeed vary between sites. There is a pretty large range in alpha diversity values across the dataset, and we can also see some good seperation between sites in the ordination with a few distinct clusters forming. Looking at the treatment type of each site, it is clear that while there is some seperation by treatment, especially along the first axis, the clustering isn't perfect and there are some sites of different treatments falling together. This is most clear by example in the top right quadrant: sites 19, 15, and 23 cluster very closely on axis 1 and 2 and these three sites correspond to the long-term Krat exclusure, long-term Krat exclosure, and rodent exclosure respectively. And moving a little to the left we see points 16 and 21 which are rodent and long-term Krat exclosure treatments again respectively. So, in general there is definitely some variation in diversity between sites that corresponds to treatment type, but it isn't necessarily the only thing going on, a conclusion reflected in the answer below...
>
> ***Answer 2b***: When PERMANANOVA was used to test the hypothesis that species abundances vary as a factor of treatment type, we found pretty solid supporting evidence that they do. We see pretty high, significant support (R^2 = .414, p = .003) for our hypothesis that plot type is explaining the variation seen between sites. That said, the R^2 value of .414 isn't amazing, and there is still over half of the variance seen left unexplained by just plot type alone, meaning that there is something else going on with these data that we haven't addressed yet just by looking at plot type.

4) TIME SERIES ANALYSIS

In the R code chunk below, do the following:

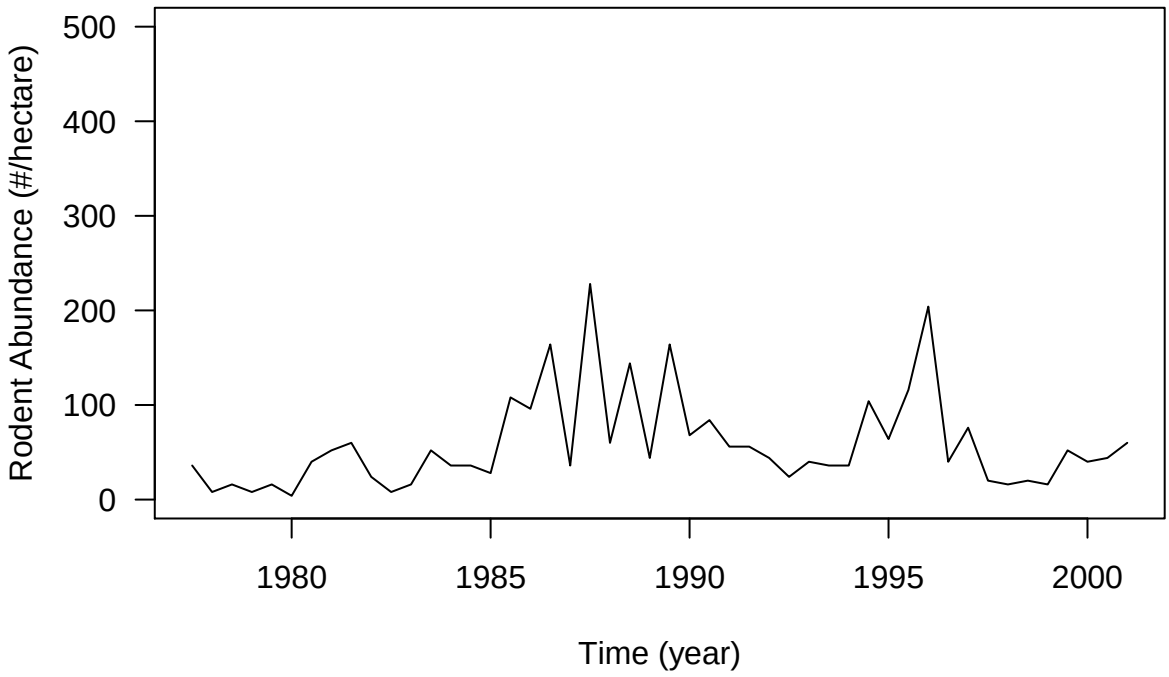
1. Create a time-by-species matrix that includes year, month, and plot_id for a site other than plot_id 2.
2. Examine per-hectare rodent abundance using simple moving average smoothing.
3. Test whether your data meets the assumption of stationarity.
4. If it does not meet this asumption, explore wasy to make your data stationary.
5. Examine and plot time lags using the partial autocorrelation function (PACF) and autocorrelation function (ACR).
6. Use the tools outlined in the Handout to create an ARMA model.

```
portal <- tidyr::unite(portal, col = date, c(year, month, day), sep = "-", remove = FALSE)
portal <- tidyr::unite(portal, col = taxon, c(genus, species), sep = "_", remove = FALSE)

time.by.species <- group_by(portal, year, plot_id) %>%
count(taxon) %>% spread(key = taxon, value = n, fill = 0)

time.by.spec.7 <- filter(portal, taxa=="Rodent") %>%
group_by(year, month, plot_id) %>%
```

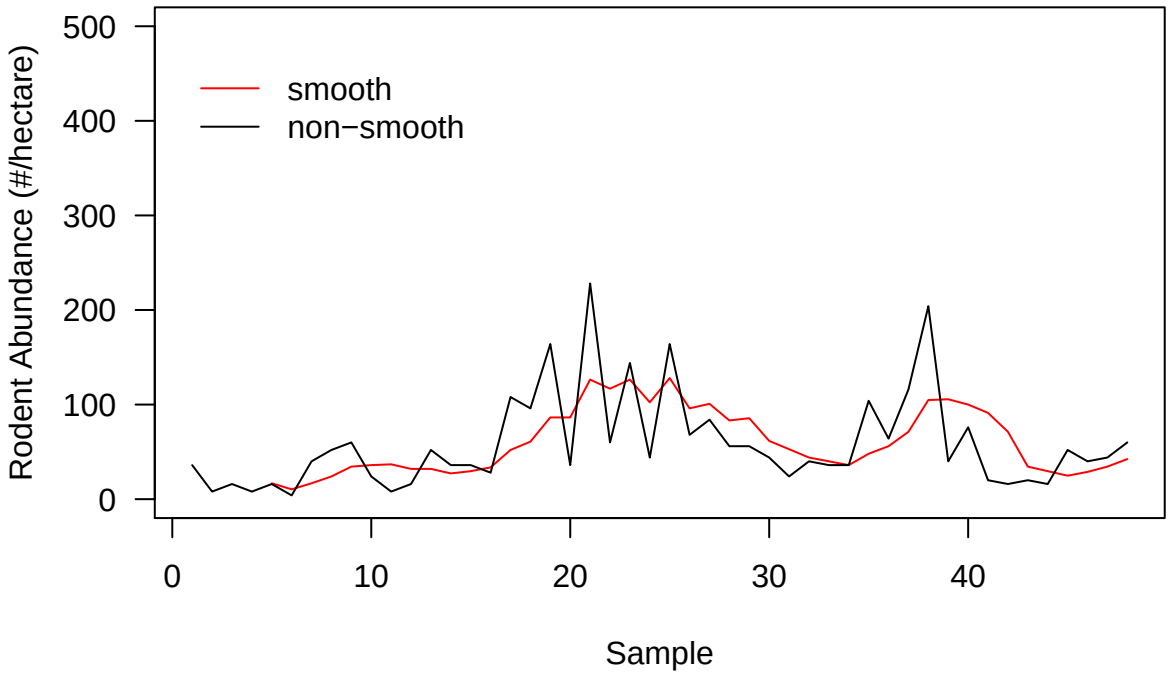
```
count(taxon)

time.by.spec.7$season <- NA
time.by.spec.7$season <- time.by.spec.7$month %in% c(6:10)

time.by.spec.7$season <- ifelse(time.by.spec.7$season == TRUE, "rain", "norain")

group_by(time.by.spec.7, year, season)
```

```
## Source: local data frame [16,391 x 6]
## Groups: year, season [52]
##
##     year month plot_id                   taxon     n season
##    <int> <int>   <int>                   <chr> <int>  <chr>
## 1   1977     7       1       Dipodomys_merriami     2   rain
## 2   1977     7       1       Perognathus_flavus     1   rain
## 3   1977     7       2 Chaetodipus_penicillatus     1   rain
## 4   1977     7       2       Dipodomys_merriami     1   rain
## 5   1977     7       2         Neotoma_albigula     1   rain
## 6   1977     7       2      Peromyscus_eremicus     1   rain
## 7   1977     7       3       Dipodomys_merriami     2   rain
## 8   1977     7       3    Dipodomys_spectabilis     1   rain
## 9   1977     7       3         Neotoma_albigula     1   rain
## 10  1977     7       4       Dipodomys_merriami     1   rain
## # ... with 16,381 more rows
```

```
abundance <- filter(time.by.spec.7, plot_id == 7) %>%
  group_by(year, season) %>%
  count(wt = n)

abundance$nn <- abundance$nn * 4

abundance.ts <- ts(abundance$nn, frequency = 2, start = c(1977,2))

plot.ts(abundance.ts, type = "l", ylab = "Rodent Abundance (#/hectare)",
xlab = "Time (year)", las = 1, ylim = c(0, 500))
```




```
abund.sm <- SMA(abundance$nn, n = 5)
plot(abund.sm, type = "l", col = "red", ylab = "Rodent Abundance (#/hectare)",
xlab = "Sample", las = 1, ylim = c(0, 500))
lines(abundance$nn, col = "black")
legend(0, 475, col = c("red", "black"), lty = c(1,1),
c("smooth", "non-smooth"), bty = "n", cex = 1)
```



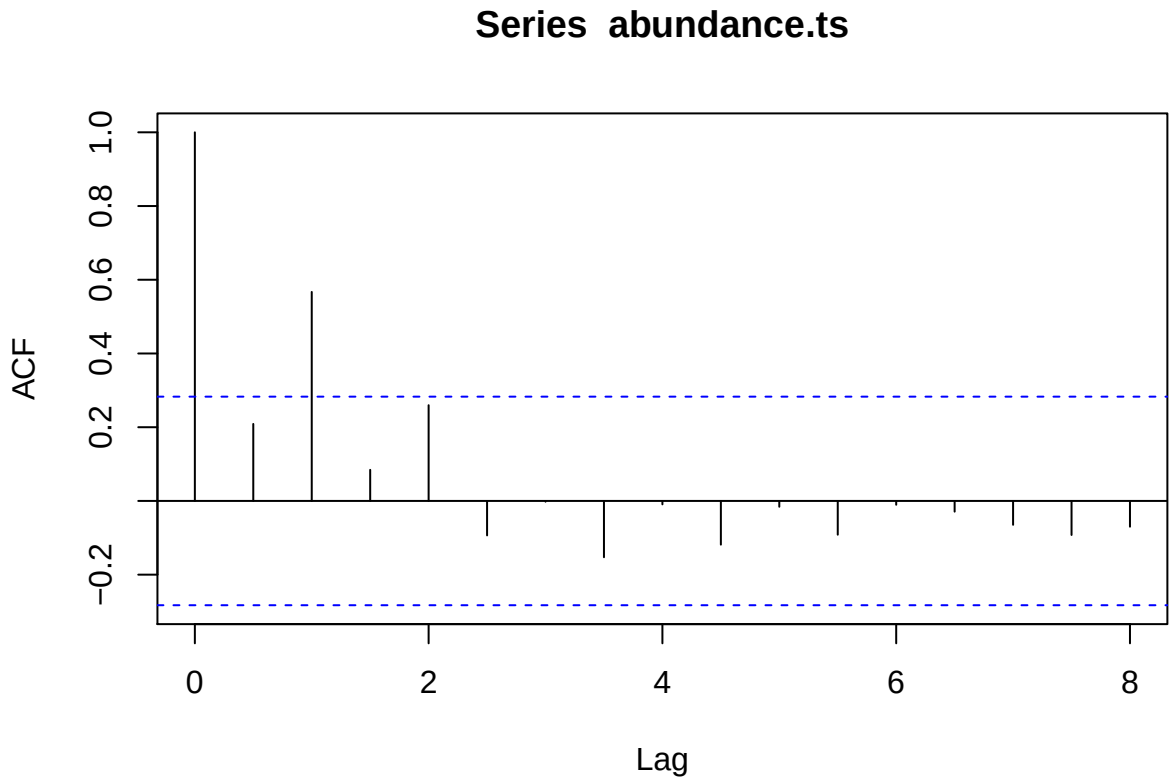

```
#checking for stationarity of data
adf.raw <- adf.test(abundance.ts, alternative = "stationary")
adf.raw$p.value
```

```
## [1] 0.4384658
```

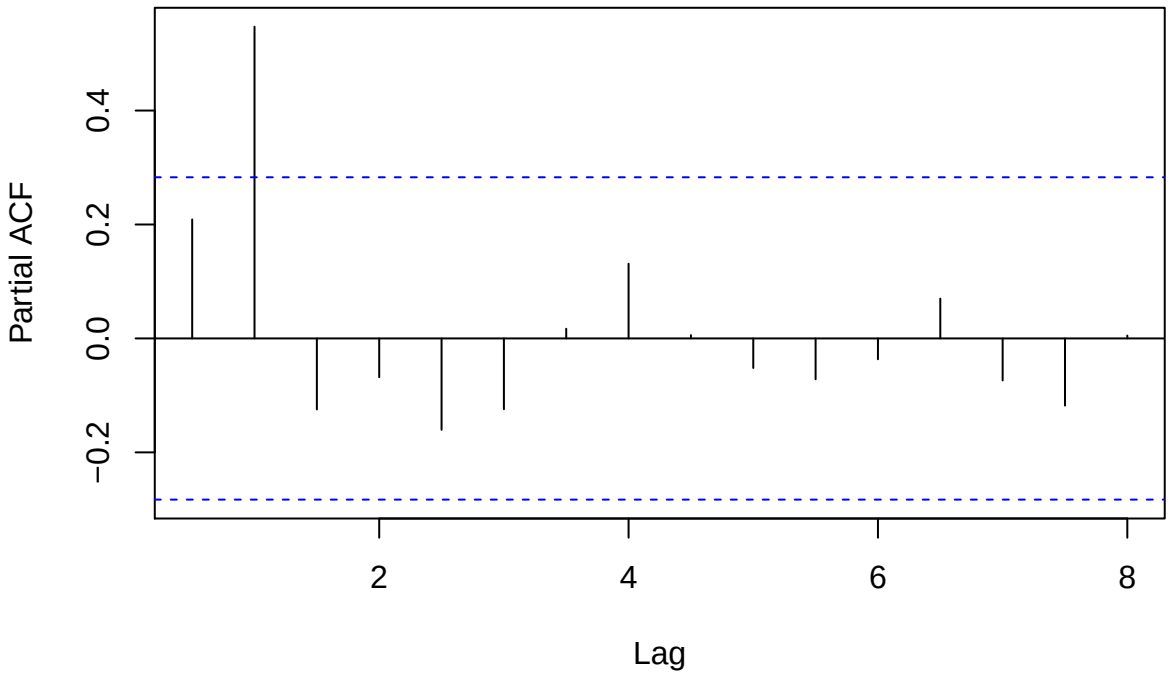
```
#correcting for stationarity using diff()
abundance.ts.diff <- diff(abundance.ts)
adf.diff <- adf.test(abundance.ts.diff, alternative = "stationary")
adf.diff$p.value
```

```
## [1] 0.04498274
```

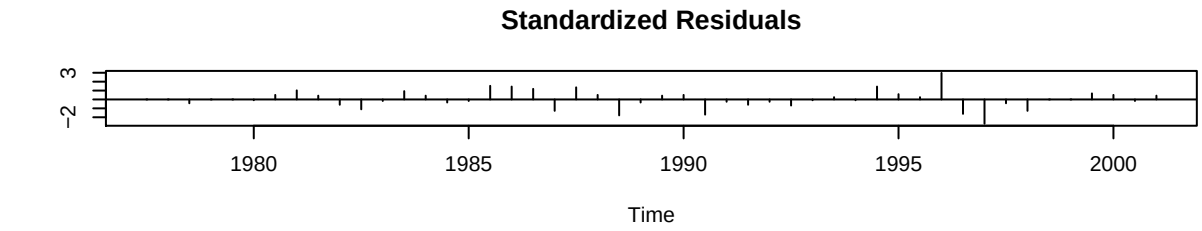
```
#Examining and plotting time lags using PACF and ACR.

acf(abundance.ts)
```

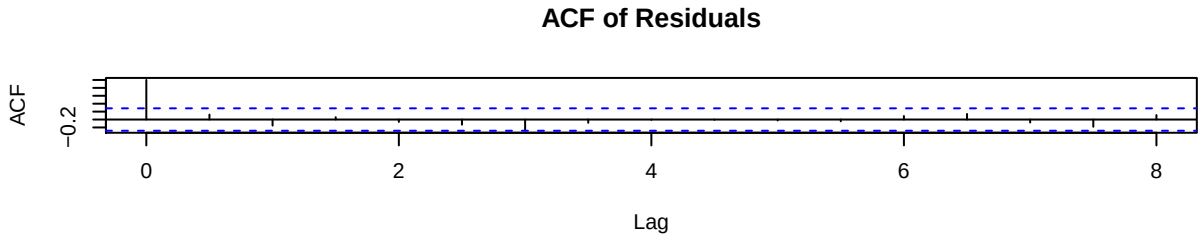


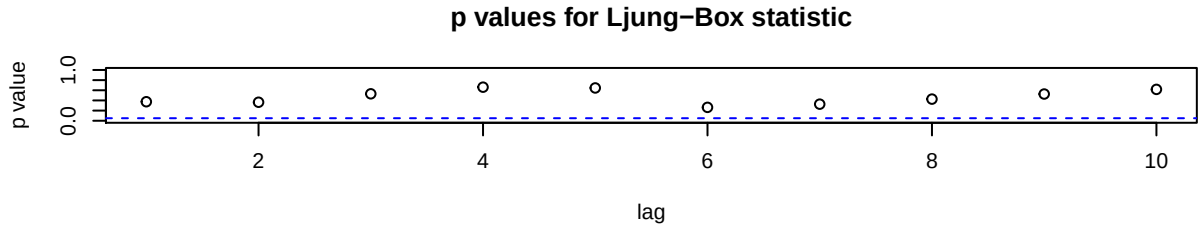

```
pacf(abundance.ts)
```



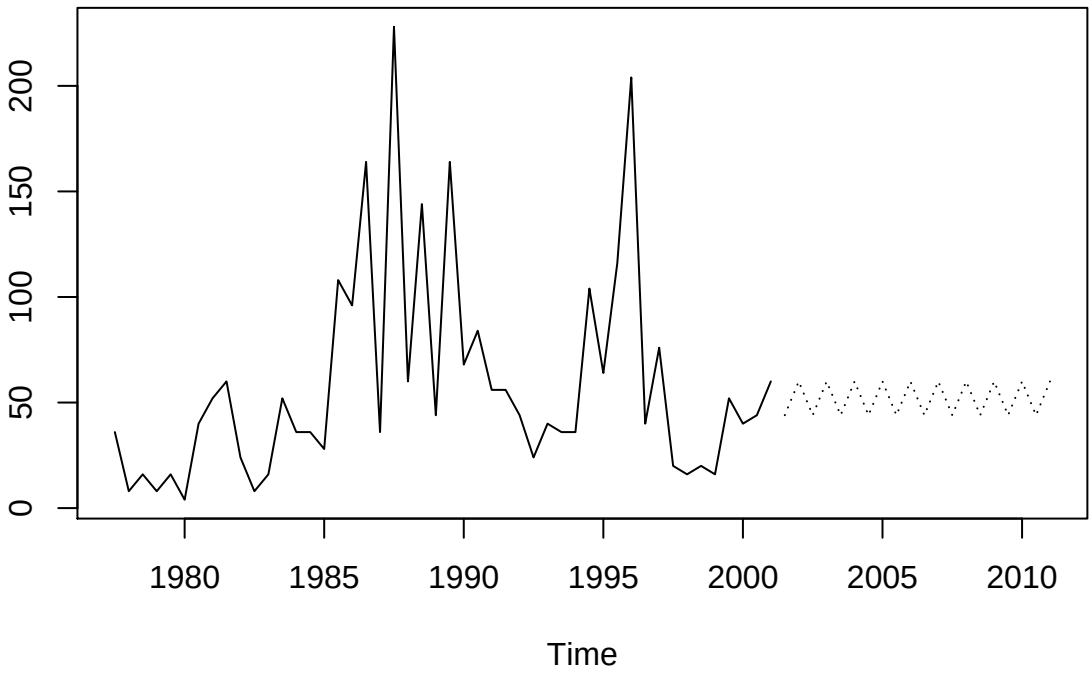

```
#Complete the ARMA model

abundance.arm <- auto.arima(abundance.ts)
tsdiag(abundance.arm)
```



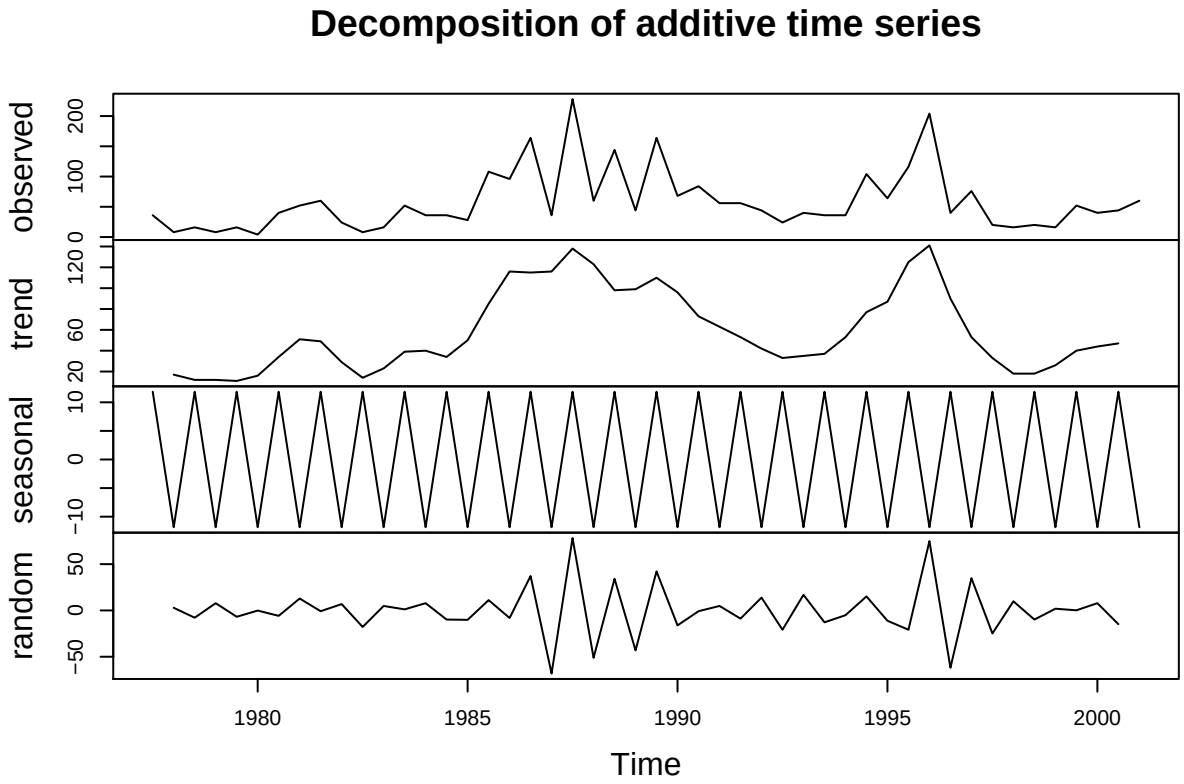





```
pred.arm <- predict(abundance.arm, n.ahead = 20)
ts.plot(abundance.ts, pred.arm$pred, lty = c(1,3))
```




```
abundance.comp <- decompose(abundance.ts)
plot(abundance.comp)
```



Question 3: Describe the results from your time series analysis.

a. Does your data meet the assumption of stationarity? If not, what does this violation imply?
b. What does the ACF function do and how does it relate to the ARMA model? How does this differ from the autocorrelation function (ACF)?
c. What results can you conclude from your full ARMA model along with other methods outlined in the time series setcion of the Handout?

Answer 3a: No, it does not. This violation implies that time has an impact on the data being analyzed, specifically on the mean, variance and covariance of the data. That is, there is a trend to the data and it is changing over time, not remaining stable with oscillations.

Answer 3b: The ACF looks at time lags in the time series data being analyzed by checking for correlations between intervals as defined by our sliding window. So, it is looking for similarities between observations that are just a function of those observations being repititions of the same thing just with time between them. This measure seems as though it would be quite important to ARMA models as they are looking at autoregressive trends to model predictions about future observations, so information about autocorrelations in the past and their time lags would be critical to future prediction.

The partial autocorrelation function (PACF) differs from the ACF described above in that it gives information about correlations between two series, after it has already corrected for other correlations that might exist with another lagged series.

Answer 3c: From the full ARMA model, along with everything else, I have learned here that site 7 abundances have seen two major boom and bust cycles over the years with one falling from ~1985-1990, and another about 8 years later. These cycles are definitely solid trends, yet also contain within them much random noise. These two dramatic cycles dominate much of the timeframe examined here, and seem to make it difficult to get an accurate idea of what an average population level looks like in this site, seemingly complicating the task of future predictions from the ARMA model (which don't seem too good). In general, it is my view that the extremly volatile

nature of the populations in plot 7 over the years examined here has lead to a poor outcome from the ARMA model, really limiting its descriptive and predictive power.

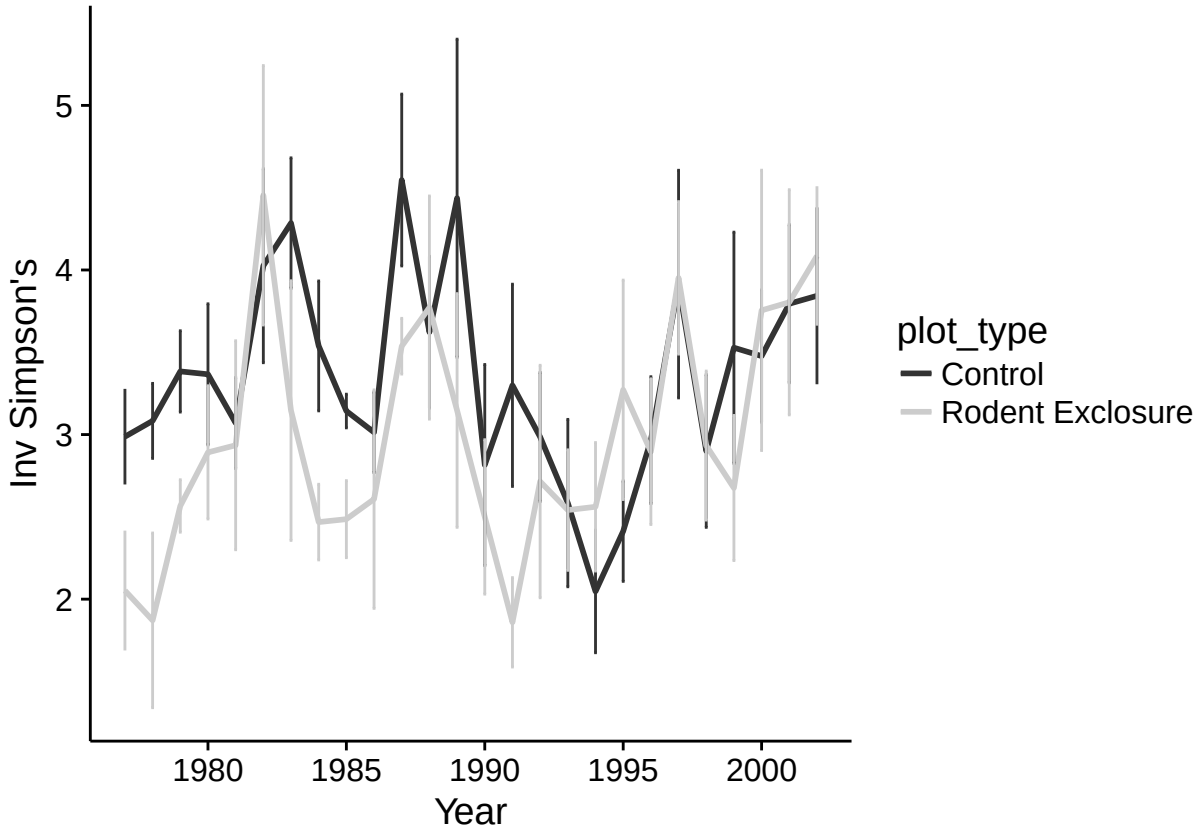
5) REPEATED MEASURES ANALYSIS OF VARIANCE (RM-ANOVA)

In the R code chunk below, do the following:

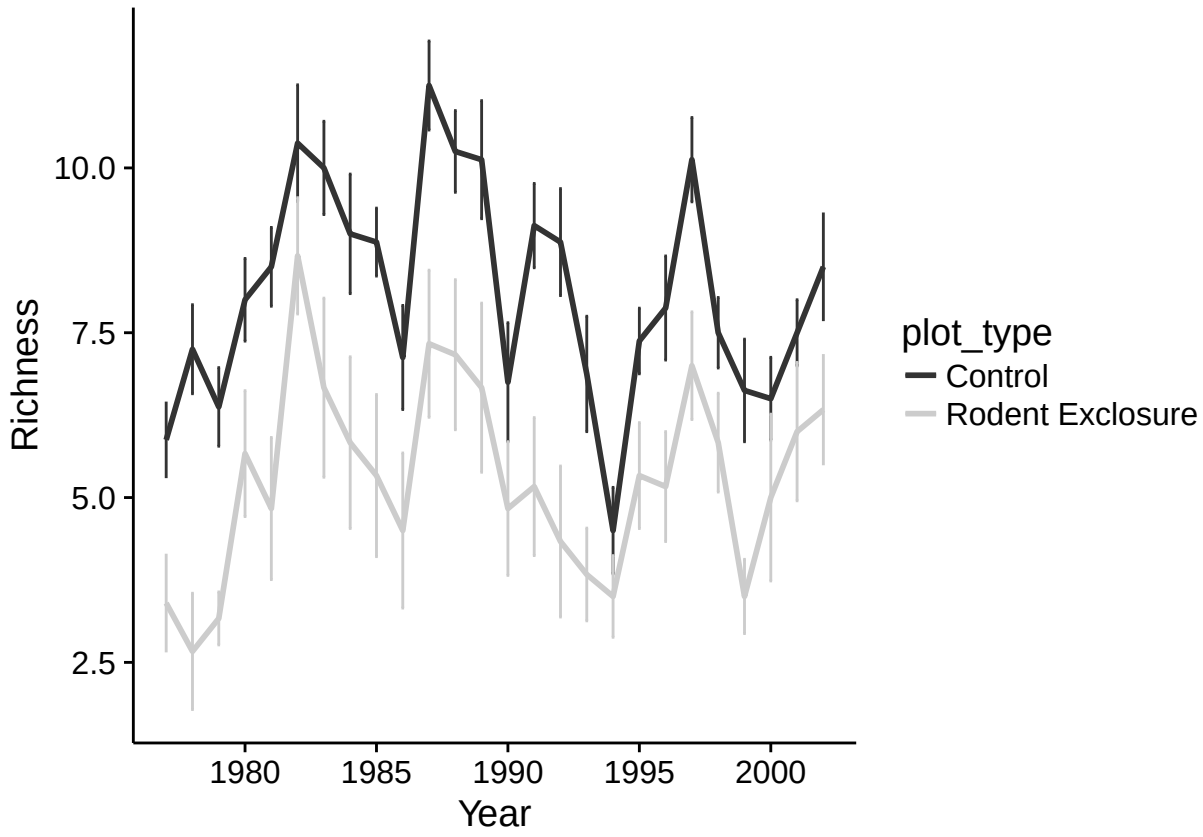
1. Create an appropriate data frame for RM-ANOVA (e.g., yearly species abundance values within plots).
2. Calculate the inverse of Simpson's diversity for each year, and plot it as a function of year for the Control and Rodent Exclosure plots.
3. Perform an RM-ANOVA and construct a F-test using the AR(1), compound symmetery, and unstructured covariance structures.

```
time.by.species <- group_by(portal, year, plot_id,
plot_type) %>% count(taxon) %>% spread(key = taxon, value = n, fill = 0)

inv.simpson <- as.data.frame(diversity((time.by.species[,-c(1:3)]), index = "invsimpson"))

richness <- as.data.frame(rowSums(time.by.species[,-c(1:3)] > 0))

div.all <- data.frame(time.by.species[,1:3,], inv.simpson)

rich.all <- data.frame(time.by.species[,1:3,], richness)

names(div.all)[4] <- "Inversesimp"

names(rich.all)[4] <- "richness"

div.treat <- div.all[which(div.all$plot_type ==
"Control" | div.all$plot_type == "Rodent Exclosure"), ]

rich.treat <- rich.all[which(rich.all$plot_type ==
"Control" | rich.all$plot_type == "Rodent Exclosure"), ]


div.treat.plot <- group_by(div.treat, plot_type, year) %>%
summarise(
mean = mean(Inversesimp),
sd = sd(Inversesimp),
n = n(),
sem = sd/sqrt(n))

rich.treat.plot <- group_by(rich.treat, plot_type, year) %>%
summarise(
mean = mean(richness),
sd = sd(richness),
n = n(),
sem = sd/sqrt(n))

div.plot <- ggplot(div.treat.plot, aes(x = year, y = mean, color = plot_type)) +
geom_line(size = 1, show.legend = T) +
geom_errorbar(aes(ymin = mean - sem, ymax = mean + sem), width = .1) +
xlim(1977, 2002) +
xlab("Year") +
```

```
ylab("Inv Simpson's")+
scale_color_grey()

plot(div.plot)
```




```
rich.plot <- ggplot(rich.treat.plot, aes(x = year, y = mean, color = plot_type)) +
geom_line(size = 1, show.legend = T) +
geom_errorbar(aes(ymin = mean - sem, ymax = mean + sem), width = .1) +
xlim(1977, 2002) +
xlab("Year") +
ylab("Richness")+
scale_color_grey()

plot(rich.plot)
```




```
# RM-ANOVA A1
rich.rm <- lme(richness ~ plot_type * year, random = ~ 1 | plot_id,
correlation = corAR1(form = ~ 1 | plot_id),
data = rich.treat)

summary(rich.rm)
```

```
## Linear mixed-effects model fit by REML
##  Data: rich.treat
##        AIC      BIC    logLik
##   1590.462 1617.567 -788.2309
##
## Random effects:
##  Formula: ~1 | plot_id
##         (Intercept) Residual
## StdDev:    1.296979 2.246238
##
## Correlation Structure: AR(1)
##  Formula: ~1 | plot_id
##  Parameter estimate(s):
##     Phi
## 0.37097
## Fixed effects: richness ~ plot_type * year
##                                    Value Std.Error  DF    t-value p-value
## (Intercept)                     28.62737  57.07705 343  0.5015565  0.6163
## plot_typeRodent Exclosure      -62.03957  89.14103  12 -0.6959710  0.4997
## year                            -0.01033   0.02869 343 -0.3600129  0.7191
## plot_typeRodent Exclosure:year   0.02980   0.04480 343  0.6652702  0.5063
```

```
##  Correlation:
##                                (Intr) plt_RE year
## plot_typeRodent Exclosure      -0.64
## year                           -1.00   0.64
## plot_typeRodent Exclosure:year  0.64  -1.00  -0.64
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.98978687 -0.69549540 -0.07250359  0.68791058  2.73090608
##
## Number of Observations: 359
## Number of Groups: 14
```

```
anova(rich.rm)
```

```
##                numDF denDF  F-value p-value
## (Intercept)        1   343 319.5050  <.0001
## plot_type          1    12  12.2775  0.0044
## year               1   343   0.0074  0.9316
## plot_type:year     1   343   0.4426  0.5063
```

```
set.caption("RMANOVA for Portal")
pander(anova(rich.rm))
```

Table 1: RMANOVA for Portal

	numDF	denDF	F-value	p-value
(Intercept)	1	343	319.5	0
plot_type	1	12	12.28	0.00435
year	1	343	0.007381	0.9316
plot_type:year	1	343	0.4426	0.5063

```
lsmeans(rich.rm, ~plot_type)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
##  plot_type          lsmean        SE df lower.CL upper.CL
##  Control          8.078695 0.5107372 13 6.975315 9.182076
##  Rodent Exclosure 5.338471 0.5916197 12 4.049442 6.627499
##
## Confidence level used: 0.95
```

```
# RM-ANOVA compound symmetry
rich.rm.cs <- lme(richness ~ plot_type * year, random = ~ 1 | plot_id,
correlation = corCompSymm(form = ~ 1 | plot_id),
data = rich.treat)

summary(rich.rm.cs)
```

```
## Linear mixed-effects model fit by REML
##  Data: rich.treat
##        AIC      BIC    logLik
##   1634.628 1661.733 -810.3142
##
## Random effects:
##  Formula: ~1 | plot_id
```

```
##         (Intercept) Residual
## StdDev:    1.395556 2.185797
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | plot_id
##  Parameter estimate(s):
##          Rho
## 6.672013e-18
## Fixed effects: richness ~ plot_type * year
##                                   Value Std.Error  DF   t-value p-value
## (Intercept)                    61.25855  40.20654 343  1.523597  0.1285
## plot_typeRodent Exclosure     -66.04270  62.95273  12 -1.049084  0.3148
## year                           -0.02671   0.02021 343 -1.321744  0.1871
## plot_typeRodent Exclosure:year   0.03181   0.03164 343  1.005341  0.3154
##  Correlation:
##                                (Intr) plt_RE year
## plot_typeRodent Exclosure      -0.639
## year                           -1.000  0.639
## plot_typeRodent Exclosure:year  0.639 -1.000 -0.639
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -3.0806434 -0.6993469 -0.0728110  0.6795500  2.7655354
##
## Number of Observations: 359
## Number of Groups: 14
```

```
anova(rich.rm.cs)
```

```
##                numDF denDF   F-value p-value
## (Intercept)        1   343 315.82853  <.0001
## plot_type          1    12  12.25170  0.0044
## year               1   343   0.78014  0.3777
## plot_type:year     1   343   1.01071  0.3154
```

```
set.caption("RMANOVA for Portal")
pander(anova(rich.rm.cs))
```

Table 2: RMANOVA for Portal

	numDF	denDF	F-value	p-value
(Intercept)	1	343	315.8	0
plot_type	1	12	12.25	0.00438
year	1	343	0.7801	0.3777
plot_type:year	1	343	1.011	0.3154

```
lsmeans(rich.rm.cs, ~plot_type)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
##  plot_type          lsmean        SE df lower.CL upper.CL
##  Control          8.117551 0.5161597 13 7.002456 9.232646
##  Rodent Exclosure 5.357391 0.5968970 12 4.056865 6.657918
##
## Confidence level used: 0.95
```

```r
# RM-ANOVA unstructured
rich.rm.u <- lme(richness ~ plot_type * year, random = ~ 1 | plot_id,
data = rich.treat)

summary(rich.rm.u)
```

```
## Linear mixed-effects model fit by REML
##  Data: rich.treat
##        AIC      BIC    logLik
##   1632.628 1655.861 -810.3142
##
## Random effects:
##  Formula: ~1 | plot_id
##         (Intercept) Residual
## StdDev:    1.395556 2.185797
##
## Fixed effects: richness ~ plot_type * year
##                                   Value Std.Error  DF   t-value p-value
## (Intercept)                    61.25855  40.20654 343  1.523597  0.1285
## plot_typeRodent Exclosure     -66.04270  62.95273  12 -1.049084  0.3148
## year                           -0.02671   0.02021 343 -1.321744  0.1871
## plot_typeRodent Exclosure:year  0.03181   0.03164 343  1.005341  0.3154
##  Correlation:
##                                (Intr) plt_RE year
## plot_typeRodent Exclosure      -0.639
## year                           -1.000  0.639
## plot_typeRodent Exclosure:year  0.639 -1.000 -0.639
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -3.0806434 -0.6993469 -0.0728110  0.6795500  2.7655354
##
## Number of Observations: 359
## Number of Groups: 14
```

```r
anova(rich.rm.u)
```

```
##                numDF denDF   F-value p-value
## (Intercept)        1   343 315.82853  <.0001
## plot_type          1    12  12.25170  0.0044
## year               1   343   0.78014  0.3777
## plot_type:year     1   343   1.01071  0.3154
```

```r
set.caption("RMANOVA for Portal")
pander(anova(rich.rm.u))
```

Table 3: RMANOVA for Portal

	numDF	denDF	F-value	p-value
(Intercept)	1	343	315.8	0
plot_type	1	12	12.25	0.00438
year	1	343	0.7801	0.3777
plot_type:year	1	343	1.011	0.3154

```
lsmeans(rich.rm.u, ~plot_type)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
##  plot_type          lsmean        SE df lower.CL upper.CL
##  Control          8.117551 0.5161597 13 7.002456 9.232646
##  Rodent Exclosure 5.357391 0.5968970 12 4.056865 6.657918
##
## Confidence level used: 0.95
```

Question 4: Describe the results from your RM-ANOVA.

a. In your own words describe what a RM-ANOVA test is doing
b. Is there a noticeable trend in the inverse of Simpson's diversity over time?
c. What does the result of your F-test tell you?
d. Of the three RM-ANOVA models with different covariance structures, which one is best? How does this affect the interpretation of your data?

Answer 4a: Repeated measures analysis of variance is an ANOVA that specifically works around the standard ANOVA assumption of independence of samples and allows for analysis of samples that have been measured multiple times.

Answer 4b: Not particularly. In both control and rodent exclosure plots, there just seems to be a lot of oscillation.

Answer 4c: If I am intrepreting it correctly, the AR(1) F-test seems to be saying that the plot type is far and away the best explanatory variable for the variation seen between sites. The year and plot type by year interaction terms are both hold very poor explanatory value.

Answer 4d: I may have done it wrong, because there is very little difference between the three RM-ANOVA models here. The compound and unstructured covariance structures were identical, and not much different than the AR(1) which was the best by a hair. The fact that the AR(1) model is the best, tells me that it is most accurate to assume that an observation at time t is most influenced by observations at time t-1, which seems very logical and straightforward - an easy intrepretation.

6) TEMPORAL BETA DIVERSITY

Turnover

In the R code chunk below, do the following:

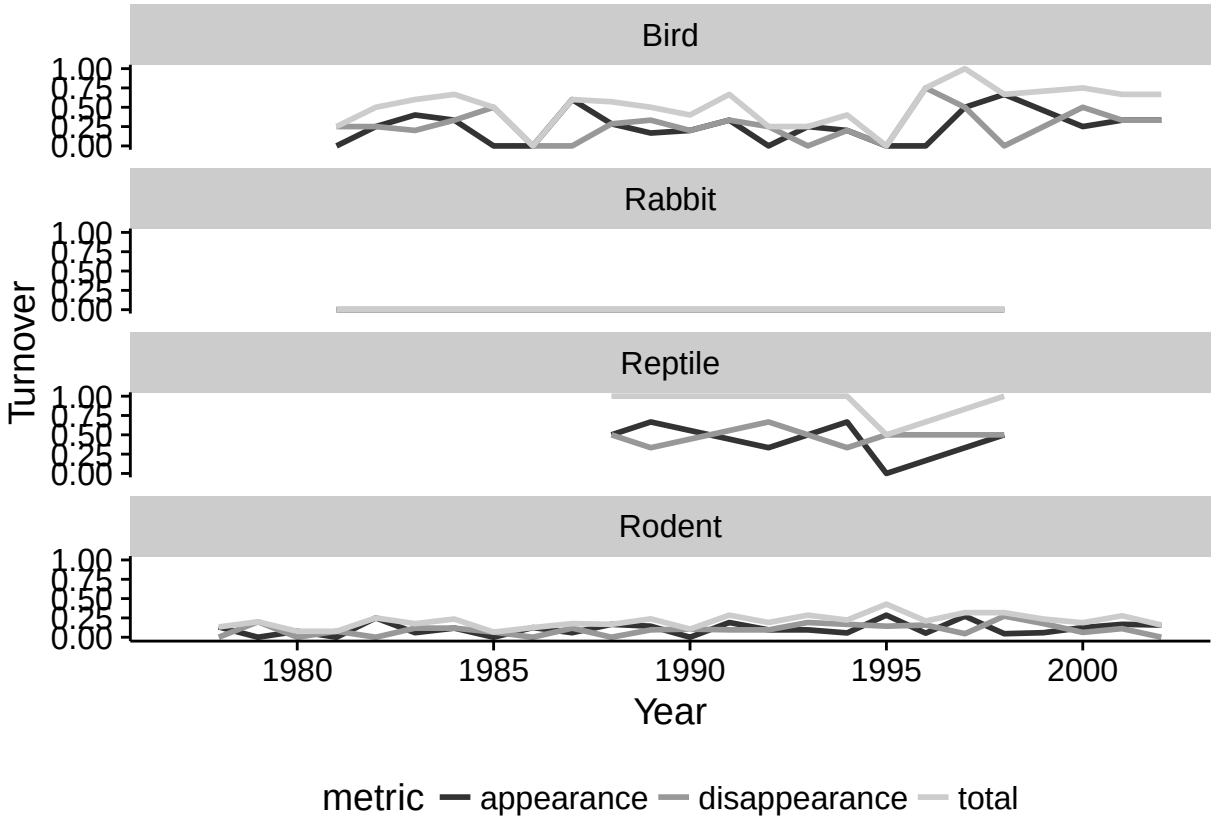
1. Calculate species abundances for each taxonomic group (the `taxa` column).
2. Calculate total turnover and turnover due to the gain/loss of species for each group.
3. Visualize turnover within each group

```
# abundances
portal.species.abunds <- group_by(portal, year, taxa) %>% count(taxon)

# Total turnover and turnover due to loss/gain of species

portal.total <- turnover(df = portal.species.abunds,
time.var = "year",
species.var = "taxon",
abundance.var = "n",
replicate.var = "taxa",
metric = "total")
```

```
portal.appearance <- turnover(df = portal.species.abunds,
time.var = "year",
species.var = "taxon",
abundance.var = "n",
replicate.var = "taxa",
metric = "appearance")

portal.disappearance <- turnover(df = portal.species.abunds,
time.var = "year",
species.var = "taxon",
abundance.var = "n",
replicate.var = "taxa",
metric = "disappearance")

#Visualizing turnover
portal.turnover <- full_join(portal.total, portal.disappearance) %>%
full_join(portal.appearance)
```

```
## Joining, by = c("year", "taxa")
## Joining, by = c("year", "taxa")
```

```
portal.turnover <- gather(portal.turnover, key = metric, value = turnover,
total, appearance, disappearance)

turn.plot <- ggplot(
portal.turnover, aes(x = year, y = turnover, color = metric)) +
geom_line(size = 1, show.legend = T) +
facet_wrap(~taxa, ncol = 1) +
xlim(1977, 2002) +
xlab("Year") +
ylab("Turnover") +
theme(legend.position = "bottom") +
scale_color_grey()
plot(turn.plot)
```



Question 5:

a. How does temporal turnover relate to spatial turnover? > Temporal and spatial turnover both describe patterns of species compositions changing. Though one is change over time and the other change over distance, the core concept seems to remain the same for both.

b. Which taxonomic group appears to be the most variable? Which group appears to be the least variable? > Overall, the reptile group seems to be the most variable, with the birds close behind in later years. The rodent and rabbit groups are very stable, with the rabbit group (oddly? wrongly?) showing a turnover of 0 for every year it is present in the dataset.

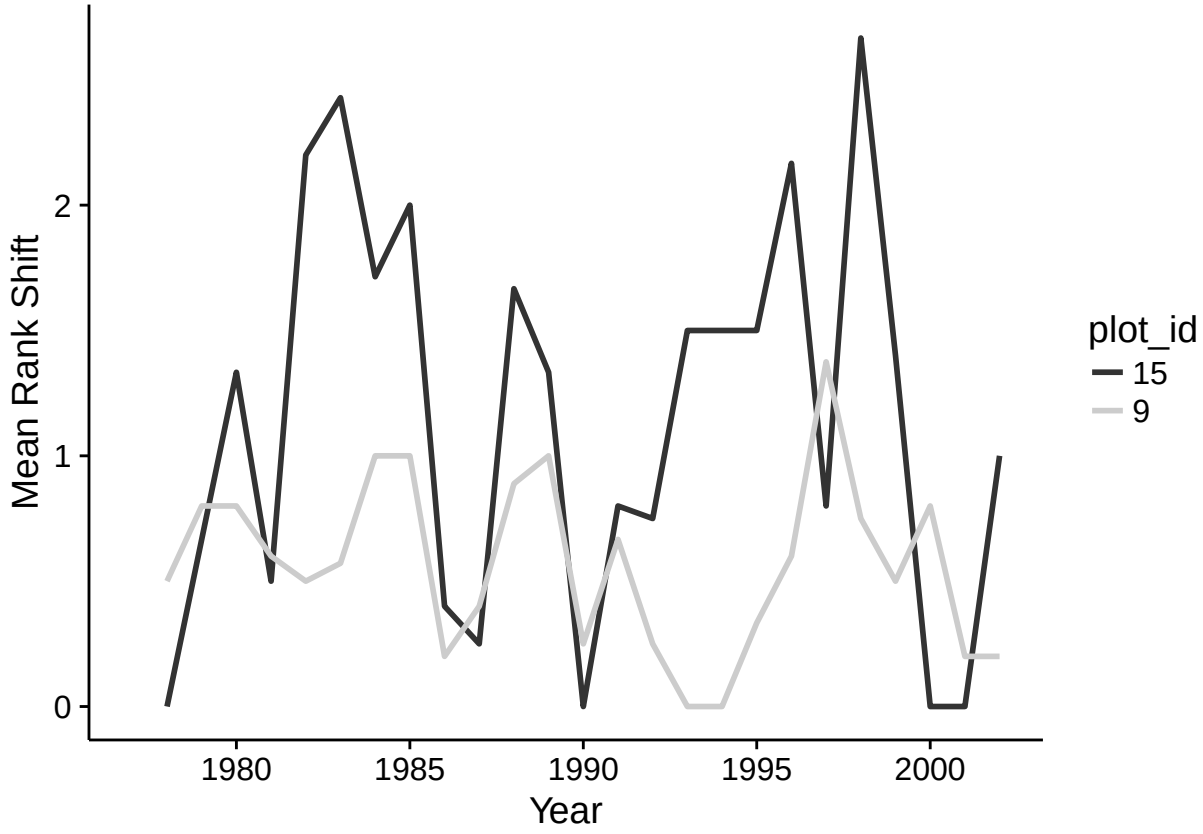
Mean Rank Shift

In the code chunk below, do the following:

1. Choose two plot_types or two plot_ids and compare the mean rank shift (MRS) between them.
2. Plot MRS for each through time.

```
portal.plots.abunds <- group_by(portal, year, plot_id) %>% count(taxon)

portal.abunds.9.15 <- filter(portal.plots.abunds,
plot_id == "9" | plot_id == "15")

portal.rankshift <- rank_shift(
df = as.data.frame(portal.abunds.9.15),
time.var = "year",
species.var = "taxon",
abundance.var = "n",
replicate.var = "plot_id")
```

```
portal.rankshift$year <- as.numeric(substr(portal.rankshift$year_pair, 6, 9))

rankshift.plot <- ggplot(portal.rankshift, aes(x = year, y = MRS, color = plot_id)) +
geom_line(size = 1) +
xlim(1977, 2002) +
xlab("Year") +
ylab("Mean Rank Shift") +
scale_color_grey()
plot(rankshift.plot)
```




```
group_by(portal.rankshift, plot_id) %>%
summarise(
mean = mean(MRS),
cv = sd(MRS)/mean)
```

```
## # A tibble: 2 × 3
##   plot_id      mean        cv
##     <chr>     <dbl>     <dbl>
## 1      15 1.1430476 0.7058073
## 2       9 0.5674127 0.6095967
```

Question 6:

a. What does a change in the rank shift tell you about the community?
b. Interpret the analysis and figure you just made.

Answer 6a: A change in rank shift tells you that the the relative commonness and rarity of taxa within the community is changing. A large MRS value means that there is a lot of change between the level of commonness and rarity of taxa in general, while a low value means that that

level is staying mostly the same. There could still be a lot of change in the community, it is just change that is affecting all taxa - rare and common - equally.

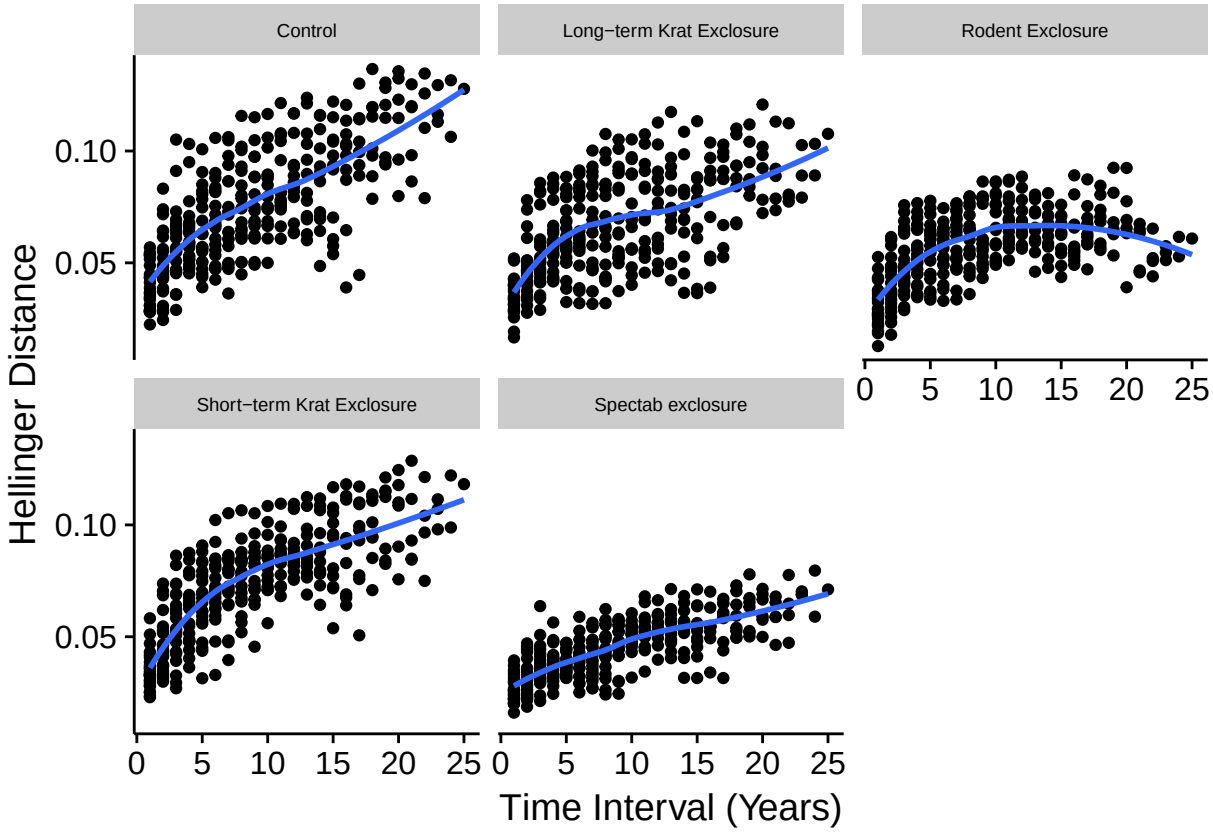
Answer 6b: My analysis above shows that between plots 9 and 15, there is much more change happening in the relative rarity and commonness of taxa found in plot 15. Plot 9 has some change happening, but by comparison is much more stable.

Rate Change Interval

In the R code chunk below, do the following:

1. Calculate the rate change interval using the Hellinger distance.
2. Plot the results.

```
portal.species.abunds <- group_by(portal, year, plot_type) %>% count(taxon)

portal.species.abunds$tot.abund <- rep(sum(portal.species.abunds$n), length(portal.species.abunds$n))

portal.hellinger.transf <- portal.species.abunds %>%
mutate(hellinger.transf = sqrt(n / tot.abund))

portal.change.int <- rate_change_interval(portal.hellinger.transf,
time.var = "year",
species.var = "taxon",
abundance.var = "hellinger.transf",
replicate.var = "plot_type")

rate.plot <- ggplot(portal.change.int, aes(interval, distance)) + geom_point() + facet_wrap(~plot_type)
theme(strip.text.x = element_text(size = 7)) +
stat_smooth(method = "loess", se = F, size = 1) +
ylab("Hellinger Distance") +
xlab("Time Interval (Years)")
rate.plot
```



Question 7:

a. What does it mean to calculate a distance metric across varying time intervals?
b. Interpret the overall results. Develop a hypothesis based on the different responses of each treatment.

Answer 7a: Here it seems to be measuring how different communities can become at a single site, over time. How different is the community at a single site at time t=0 vs t=25 vs t=100?

Answer 7b: Based on the data above, it appears that the presence of rodents, and Banner-tailed kangaroo rats (absent from Spectab exclosures) leads to the largest amounts of community change over time. When these two groups are excluded, the rate of increase for Hellinger distance is much lower than in any of the other treatments. So, I hypothesize that rodents and Banner-tailed kangaroo rats are the groups with the largest overall effect on community structure, due primarily to their feeding behavior.

7) STABILITY

In the R code chunk below, do the following:

1. Using total abundance as your focal variable, calculate stability (i.e., 1/CV) and synchrony for each plot type.
2. Test for a biodiversity-stability relationship by regressing community stability on mean richness.
3. Test for a biodiversity-stability relationship by regressing community stability on mean inverse Simpson's diversity.

```
# Stability
portal.stab <- community_stability(df = as.data.frame(portal.species.abunds),
time.var = "year",
abundance.var = "n",
```

```
replicate.var = "plot_type")

pander(portal.stab)
```

plot_type	stability
Control	3.044
Long-term Krat Exclosure	1.865
Rodent Exclosure	1.864
Short-term Krat Exclosure	2.462
Spectab exclosure	2.911

```
#Synchrony - measure of whether population densities fluctuate independently or not.
portal.loreau <- synchrony(df = as.data.frame(portal.species.abunds),
time.var = "year",
species.var = "taxon",
abundance.var = "n",
replicate.var = "plot_type",
metric = "Loreau")

names(portal.loreau)[2] <- "loreau"

portal.gross <- synchrony(df = as.data.frame(portal.species.abunds),
time.var = "year",
species.var = "taxon",
abundance.var = "n",
replicate.var = "plot_type",
metric = "Gross")
names(portal.gross)[2] <- "gross"

pander(full_join(portal.loreau, portal.gross))
```

```
## Joining, by = "plot_type"
```

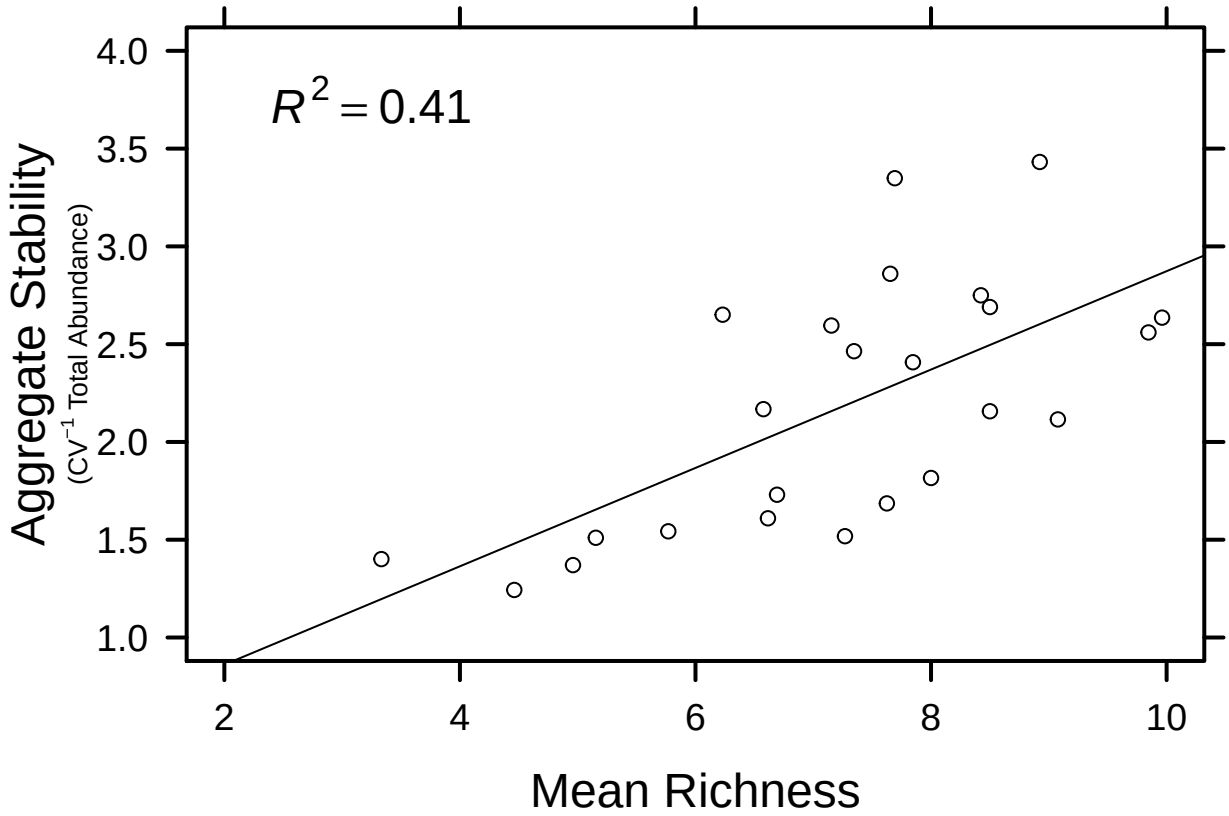
plot_type	loreau	gross
Control	0.1869	0.1263
Long-term Krat Exclosure	0.1578	0.07418
Rodent Exclosure	0.187	0.1423
Short-term Krat Exclosure	0.08773	0.001546
Spectab exclosure	0.1542	0.1393

```
#regressing community stability on mean richness

portal.mean.rich.plot <- rich.all %>%
group_by(plot_id) %>%
summarise(mean.rich = mean(richness))

portal.plot.abunds <- as.data.frame(
group_by(portal, year, plot_id) %>% count(taxon))
portal.stab.plot <- community_stability(df = portal.plot.abunds,
time.var = "year",
abundance.var = "n",
```

```
replicate.var = "plot_id")

portal.div.stab <- portal.mean.rich.plot %>%
inner_join(portal.stab.plot)
```

```
## Joining, by = "plot_id"
```

```
par(mar = c(5,5,1,1))
plot(portal.div.stab$stability ~ portal.div.stab$mean.rich,
xlab = "", ylab = "", yaxt = "n", xaxt = "n",
xlim = c(2,10), ylim = c(1,4))
axis(side = 1, cex.axis = 1.2, lwd.ticks = 2, las = 1)
axis(side = 2, cex.axis = 1.2, lwd.ticks = 2, las = 1)
axis(side = 3, lwd.ticks = 2, las = 1, labels = F)
axis(side = 4, lwd.ticks = 2, las = 1, labels = F)
box(lwd = 2)
mtext("Mean Richness", side = 1, line = 3, cex = 1.5)
mtext("Aggregate Stability", side = 2, line = 3.5, cex = 1.5)
mtext(expression(paste("(CV"^"-1"," Total Abundance)")),
side = 2, line = 2.25, cex = .8)
div.stab.lm <- lm(portal.div.stab$stability ~ portal.div.stab$mean.rich)
abline(div.stab.lm)
r2 <- bquote(italic(R)^2 == .(format(
summary(div.stab.lm)$adj.r.square, digits = 3)))
text(3.25,3.75, cex = 1.5, labels = r2)
```




```
#regressing community stability on mean inverse simpson
```

```
portal.mean.div.plot <- div.all %>%
group_by(plot_id) %>%
summarise(mean.div = mean(Inversesimp))

portal.plot.abunds <- as.data.frame(
group_by(portal, year, plot_id) %>% count(taxon))
portal.stab.plot <- community_stability(df = portal.plot.abunds,
time.var = "year",
abundance.var = "n",
replicate.var = "plot_id")

portal.simp.stab <- portal.mean.div.plot %>%
inner_join(portal.stab.plot)
```

```
## Joining, by = "plot_id"
```

Question 8:

a. Which plot type has the highest stability in total abundance? How is stability of total abundance measured with the function you learned? How does this measure of stability relate to the coefficient of variation?

b. In your own words, describe the concept of synchrony

c. Interpret the results from the biodiversity-stability relationships you analyzed.

> ***Answer 8a***: The control plot type has the highest stability in total abundance. Here stability was calculated as the inverse of the coefficient of variation, which is the standard deviation of a variable divided by its mean value.
>
> ***Answer 8b***: Synchrony is a measure of the strength of dependence of population's density fluctuations. If two populations' have densities that fluctuate together in the same direction (both increase at the same time), they are said to be positively synchronous.
>
> ***Answer 8c***: The first biodiversity-stability plot (stability vs richness) shows that as the mean richness of a site increases, so too does its stability in a relatively predictable (R^2 = .41) manner; it isn't perfect but there is a definite trend there. On the other hand, the second plot clearly shows that there is no real relationship (R^2 = -.039) between mean inverse simpson's diversity and the stability of a population for these particular data.

SYNTHESIS

Compare and contrast the core concepts from temporal and spatial diversity (e.g., autocorrelation, scale, variability, etc.). Identify a few of the major challenges associated with studying biodiversity through time and across space.

> ***Answer***: After giving this question some thought, I struggle to find a concept in either spatial or temporal diversity that has no close analog in the other. Geographic vs time scales (feet-miles vs hours-years), spatial vs temporal autocorrelation, distance-decay relationships in terms of both temporal and geographic distance - all are concepts which apply easily to both types of diversity. It seems to me that, at least in ecology, there is little difference between temporal and geographic distance and how they structure communities; a conclusion I believe is supported well by the similarity of the tools utilized in both units. Contrast gun to my head though: I do have a hard time identifying a spatial for the temporal concept of stability. Perhaps it is just because we didn't cover any in class that I can remember though, but it seems more difficult and less useful to think about community stability over space than over time. Unless you think about a large meta

community and how it changes at the edges relative to the aggregate community composition? Not sure if that's a thing, but it sounds interesting in my head!

As temporal and spatial diversity share many of the same characteristics, it makes sense that they should share some of the same challenges as well. It seems to me that the largest challenge facing studies of both of these types of diversity is that of autocorrelation. When looking at different sites spread across a landscape, how can you be sure that any similarity you are seeing is real and not just a function of the close proximity (geographically and environmentally) of the sites? Likewise, when looking at the same site over many years how does one ensure that they are properly removing the effect of sample points looking similar just because they were taken from the same site in successive years and so are similar just because the same things are still there? Autocorrelation, be it spatial or temporal seems like an obvious thing once you know to look for it, but failure to account for it fully can really ruin an otherwise promising study. A related, but slightly different problem in studying these two aspects of diversity is that of ensuring you sample along an adaquate distance. If we know that autocorrelation exists, we also have to know that a good way to combat it and find an actual signal in the data is to make sure there are samples spread out along a long enough distance (geographical or temporal) to seperate out all possible effects of proximity.