Assignment: Temporal Diversity

Alison Partee; Z620: Quantitative Biodiversity, Indiana University

15 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.

```
rm(list=ls())
getwd()
```

```
## [1] "/Users/flopsei/GitHub/QB2017_Partee/Week5-Temporal"
```

```
setwd("/Users/flopsei/GitHub/QB2017_Partee/Week4-Spatial")

package.list <- c('vegan', 'tidyr', 'dplyr', 'codyn', 'ggplot2', 'cowplot', 'MullerPlot', 'RColorBrewer
for (package in package.list) {
```

```
  if (!require(package, character.only = TRUE, quietly = TRUE)) {
    install.packages(package, repos = 'http://cran.us.r-project.org')
    library(package, character.only = TRUE)
  }
}
```

```
## 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
```

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)
str(portal)
```

```
## 'data.frame':    34786 obs. of  13 variables:
##  $ record_id      : int  1 72 224 266 349 363 435 506 588 661 ...
##  $ month          : int  7 8 9 10 11 11 12 1 2 3 ...
##  $ day            : int  16 19 13 16 12 12 10 8 18 11 ...
##  $ year           : int  1977 1977 1977 1977 1977 1977 1977 1978 1978 1978 ...
##  $ plot_id        : int  2 2 2 2 2 2 2 2 2 2 ...
##  $ species_id     : Factor w/ 48 levels "AB","AH","AS",..: 16 16 16 16 16 16 16 16 16 16 ...
##  $ sex            : Factor w/ 3 levels "","F","M": 3 3 1 1 1 1 1 1 3 1 ...
##  $ hindfoot_length: int  32 31 NA NA NA NA NA NA NA NA ...
##  $ weight         : int  NA NA NA NA NA NA NA NA 218 NA ...
##  $ genus          : Factor w/ 26 levels "Ammodramus","Ammospermophilus",..: 13 13 13 13 13 13 13 13
##  $ species        : Factor w/ 40 levels "albigula","audubonii",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ taxa           : Factor w/ 4 levels "Bird","Rabbit",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ plot_type      : Factor w/ 5 levels "Control","Long-term Krat Exclosure",..: 1 1 1 1 1 1 1 1 1 1
```

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: 24 plots ***Answer 1b***: 40 species

3) WRANGLING THE PORTAL DATASET

In the R code chunk below, do the following:

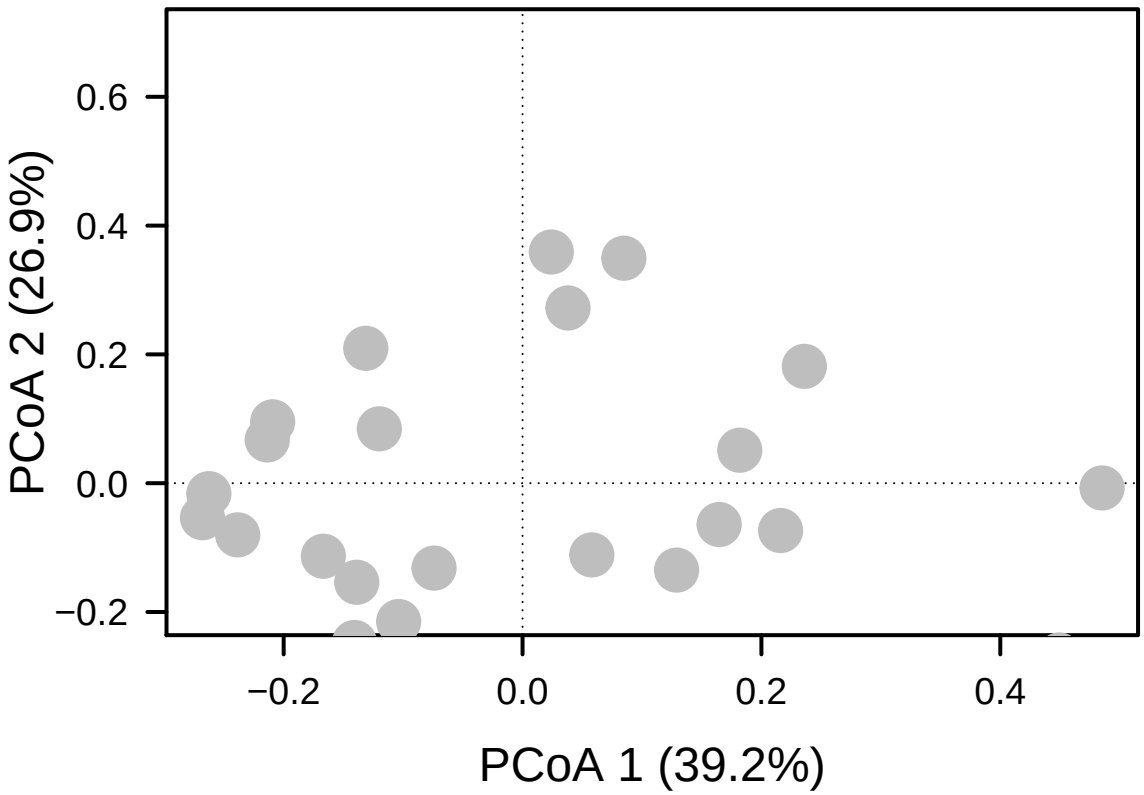
1. Create a site-by-species matrix for any year of your choosing.
2. Create a vector of plot_type for sites in the site-by-species matrix.
3. Analyze alpha diversity (e.g., Shannon/Simpson) across the sites for that year.
4. Create a PCoA ordination of your site-by-species matrix.
5. Using the hypothesis 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).

```
#make a date vector that contains year, month, and day
portal <- unite(portal, col = date, c(year, month, day), sep = "_", remove = FALSE)

#make a taxon vector that contains genus and species names
portal <- 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)

year77 <- dplyr::filter(time.by.species, year == 1977)
year77 <- as.data.frame(year77)
year77plots <- year77$plot_id
year77 <- year77[,3:dim(year77)[2]]


#function to calculate observed richness
S.obs <- function(x = ""){
  # input: site by species matrix
  # output: vector of named nums
  # purpose: sums the amount of cells > 0 for each row
  rowSums(x > 0)
}


#function to calculate simpson's evenness
SimpE <- function(x = "") {
  S <- S.obs(x)
  x = as.data.frame(x)
  D <- diversity(x, "inv")
  E <- (D)/S
  return(E)
}

#analyze alpha div for the different sites
alpha77plots <- c()
n <- 1
while(n <= length(year77plots)) {
  alpha77plots <- c(alpha77plots, SimpE(year77[n,]))
  n <- n + 1
}
alpha77plots
```

```
##         1         2         3         4         5         6         7
## 0.8642857 0.4653869 0.6337719 0.6540541 0.5365079 0.3552632 0.5294118
```

```
##         8         9        10        11        12        13        14
## 0.8653846 0.5126582 0.8000000 0.5235507 0.5632184 0.4830339 0.4709302
##        15        16        17        18        19        20        21
## 0.5614035 0.6436170 0.4049098 0.5044391 0.3600000 0.4860000 0.6598174
##        22        23
## 0.2898551 0.6400000
```

```
year77.db <- vegdist(year77, method = 'bray')

year77.pcoa <- cmdscale(year77.db, eig = TRUE, k = 3)

explainvar1 <- round(year77.pcoa$eig[1] / sum(year77.pcoa$eig), 3) * 100
explainvar2 <- round(year77.pcoa$eig[2] / sum(year77.pcoa$eig), 3) * 100
explainvar3 <- round(year77.pcoa$eig[3] / sum(year77.pcoa$eig), 3) * 100
sum.eig <- sum(explainvar1, explainvar2, explainvar3)

# Define Plot Parameters
par(mar = c(5, 5, 1, 2) + 0.1)
# Initiate Plot
plot(year77.pcoa$points[ ,1], year77.pcoa$points[ ,2], ylim = c(-0.2, 0.7),
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)
# Add Axes
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)
# Add Points & Labels
points(year77.pcoa$points[ ,1], year77.pcoa$points[ ,2],
pch = 19, cex = 3, bg = "gray", col = "gray")
text(year77.pcoa$points[ ,1], year77.pcoa$points[ ,2],
labels = row.names(year77.pcoa$points))
```




```
#hypothesis testing
#permanova
#create factors vector
plot_type = c('spectab','control', 'longkrat', 'control', 'rodent', 'shortkrat','rodent','control', 'sp

#run permanova with adonis function
adonis(year77 ~ plot_type, method = 'bray', permutations = 999)
```

```
##
## Call:
## adonis(formula = year77 ~ plot_type, 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  4    0.9626 0.240649   2.594 0.36566  0.004 **
## Residuals 18    1.6699 0.092772         0.63434
## Total     22    2.6325                 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: Diversity (Simpson's evenness) varies a little between sites, ranging from .4 to .865 .

According to the PERMANOVA results, the treatment type explains about 37% of the squared error values, and the result is significant, meaning the different treatments do have significantly different diversity values, but much of the diversity comes from elsewhere.

Answer 2b: No, treatment type is not a significant predictor of site dissimilarity because the R2 value is .36, meaning .74 of the variation is explained elsewhere.

4) TIME SERIES ANALYSIS

In the R code chunk below, do the following:

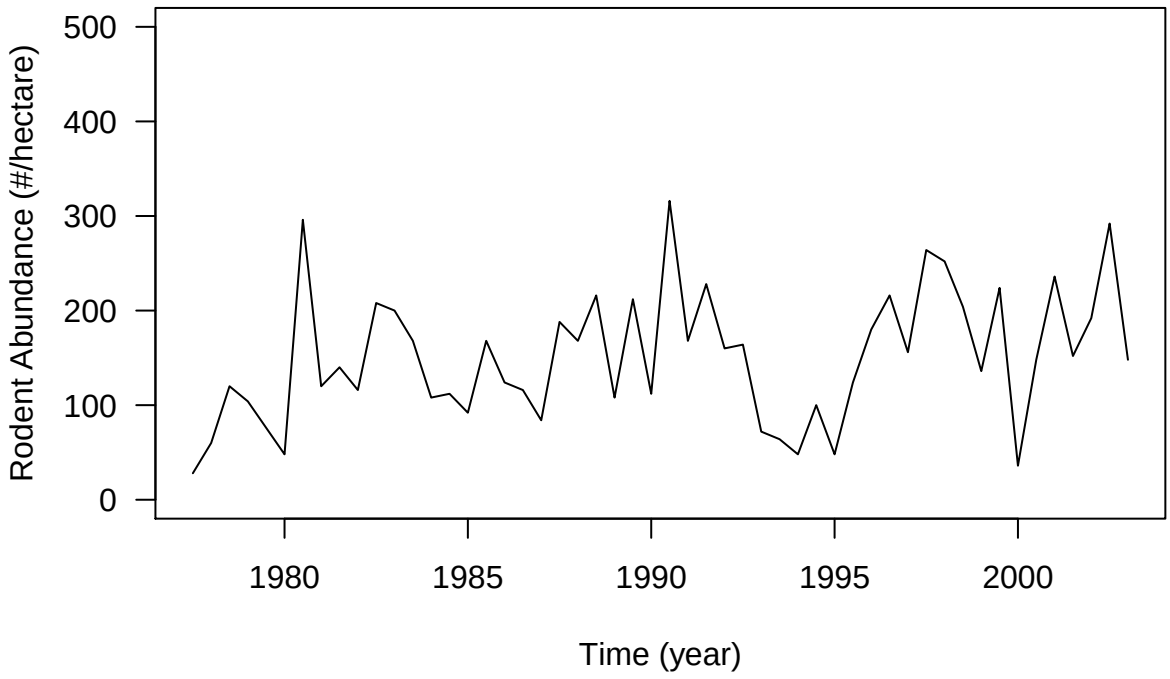
1. Create a time-by-species matrix that includes year, month, and plot_id for a site other than plot_id 2.
2. Examine per-hectare rodent abundance using simple moving average smoothing.
3. Test whether your data meets the assumption of stationarity.
4. If it does not meet this asumption, explore wasy 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.

```
# Create a time-by-species matrix that includes year, month, and plot_id
time.by.spec.2 <- filter(portal, taxa=='Rodent') %>%
  group_by(year, month, plot_id) %>%
  count(taxon)

# Create a seasonality variable using month number (6 - June; 10 = October)
time.by.spec.2$season <- NA
time.by.spec.2$season <- time.by.spec.2$month %in% c(6:10)

#Rainy seasins are June - October
time.by.spec.2$season <- ifelse(time.by.spec.2$season == TRUE, "rain", 'norain')

#group the data by year and season
group_by(time.by.spec.2, year, season)
```

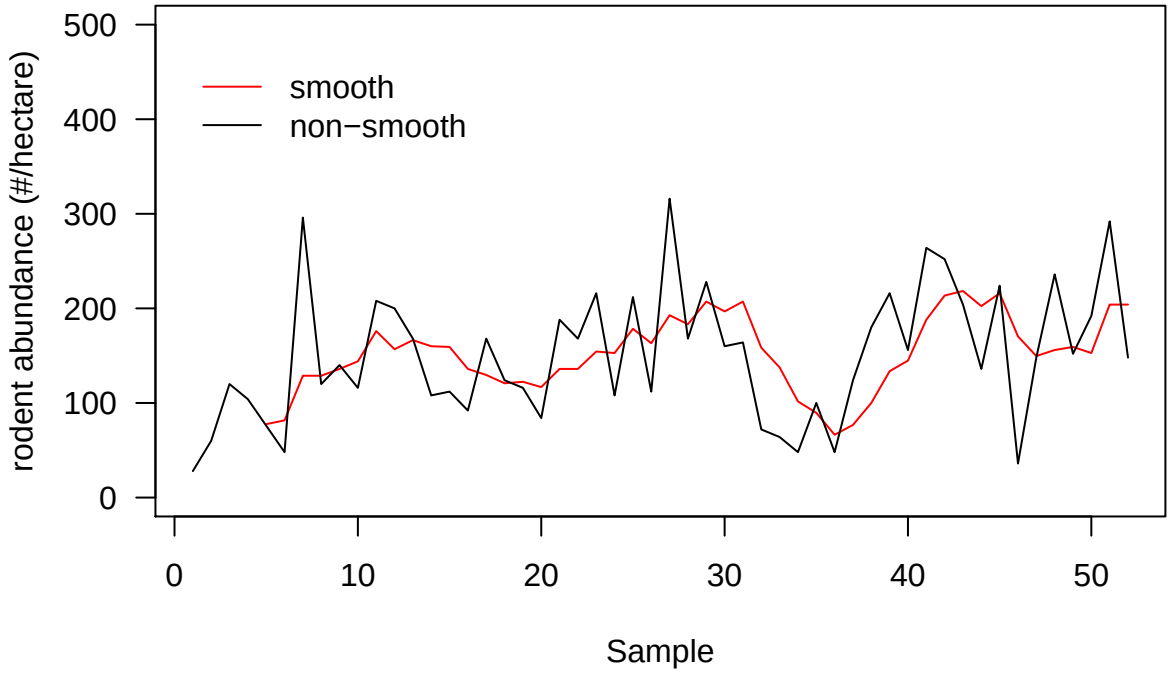
```
## Source: local data frame [16,391 x 6]
## Groups: year, season [52]
##
##     year month plot_id                    taxon     n season
##    <int> <int>   <int>                    <chr> <int>  <chr>
## 1   1977     7       1       Dipodomys_merriami     2   rain
## 2   1977     7       1       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
```

```
abund <- filter(time.by.spec.2, plot_id == 4) %>%
  group_by(year, season) %>%
  count(wt = n)
```

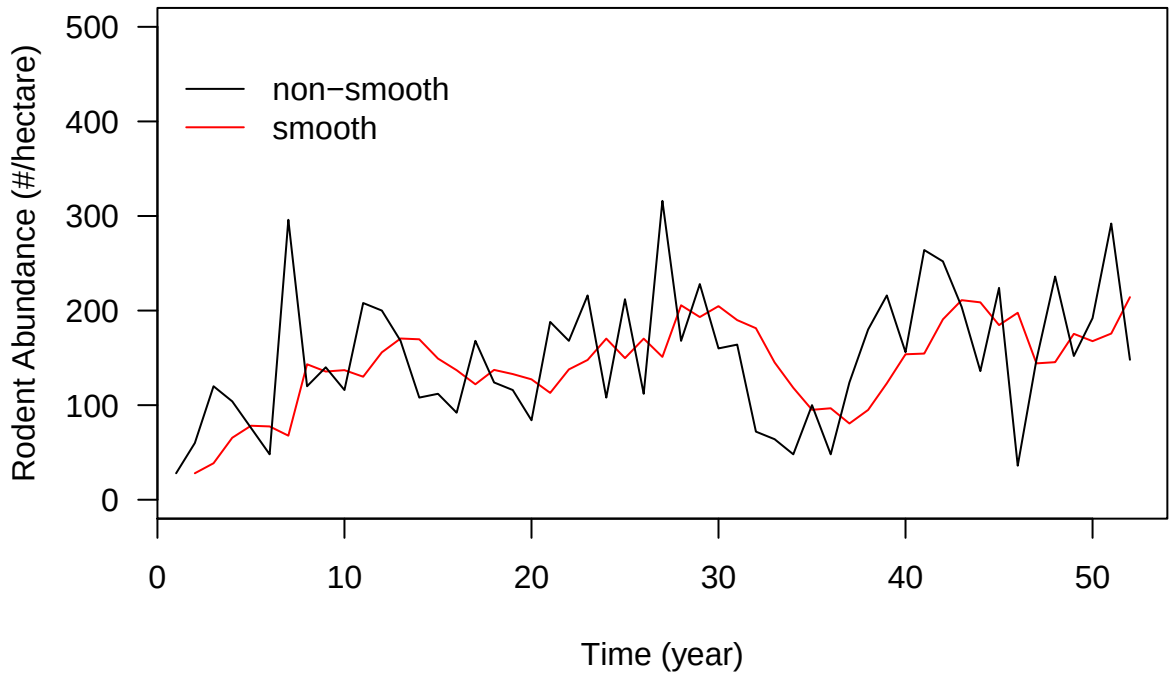
```r
abund$nn <- abund$nn * 4

abund.ts <- ts(abund$nn, frequency = 2, start = c(1977, 2))
plot.ts(abund.ts, type = 'l', ylab = 'Rodent Abundance (#/hectare)', xlab = 'Time (year)', las = 1, yli
```



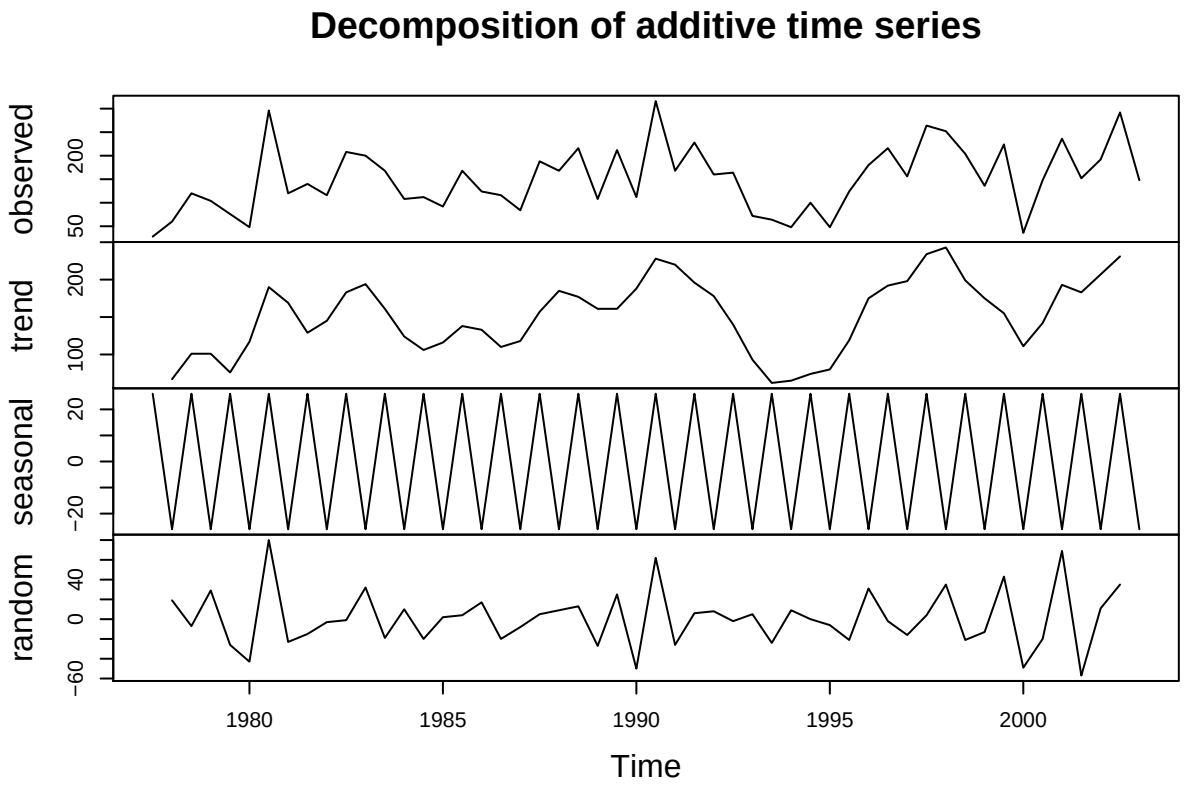

```r
abund.sm <- SMA(abund$nn, n = 5)
plot(abund.sm, type = 'l', col = 'red', ylab = 'rodent abundance (#/hectare)', xlab = 'Sample', las = 1
lines(abund$nn, col = 'black')
legend(0,475,col=c('red', 'black'), lty = c(1,1), c('smooth','non-smooth'), bty = 'n', cex = 1)
```



```r
abund.hw <- HoltWinters(abund$nn, beta = FALSE, gamma = FALSE)
# abund.hw$fitted
plot(abund.hw, xlab = "Time (year)", ylim = c(0,500), ylab = 'Rodent Abundance (#/hectare)', las = 1, ma
legend(0, 475, col = c('black', 'red'), lty = c(1,1), c('non-smooth', 'smooth'), bty = 'n', cex = 1)
```




```r
#decomposition of a time series

#moving average decomposition
abund.comp <- decompose(abund.ts)

#plot decomposition categories
plot(abund.comp)
```



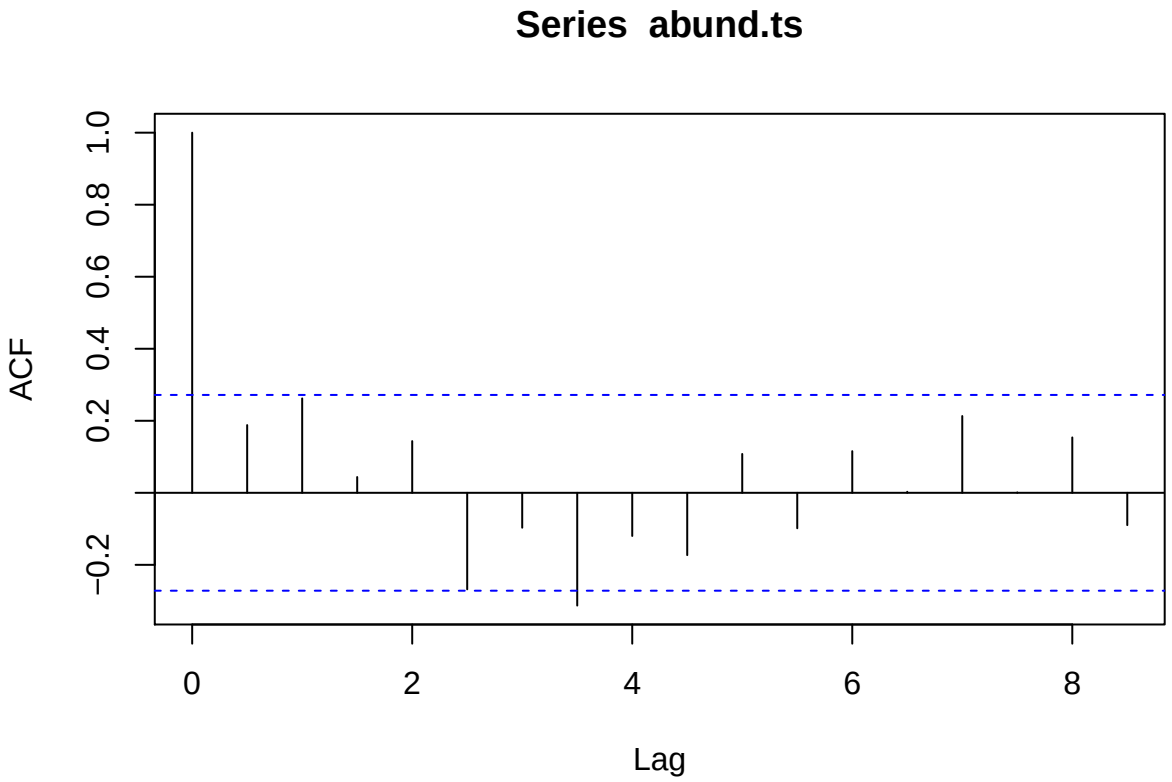

```
#remove seasonality
abund.adj <- abund.ts - abund.comp$seasonal

# ARMA models

#assumptions of stationarity
adf.raw <- adf.test(abund.ts, alternative = 'stationary')
adf.raw$p.value
```

```
## [1] 0.2776163
```

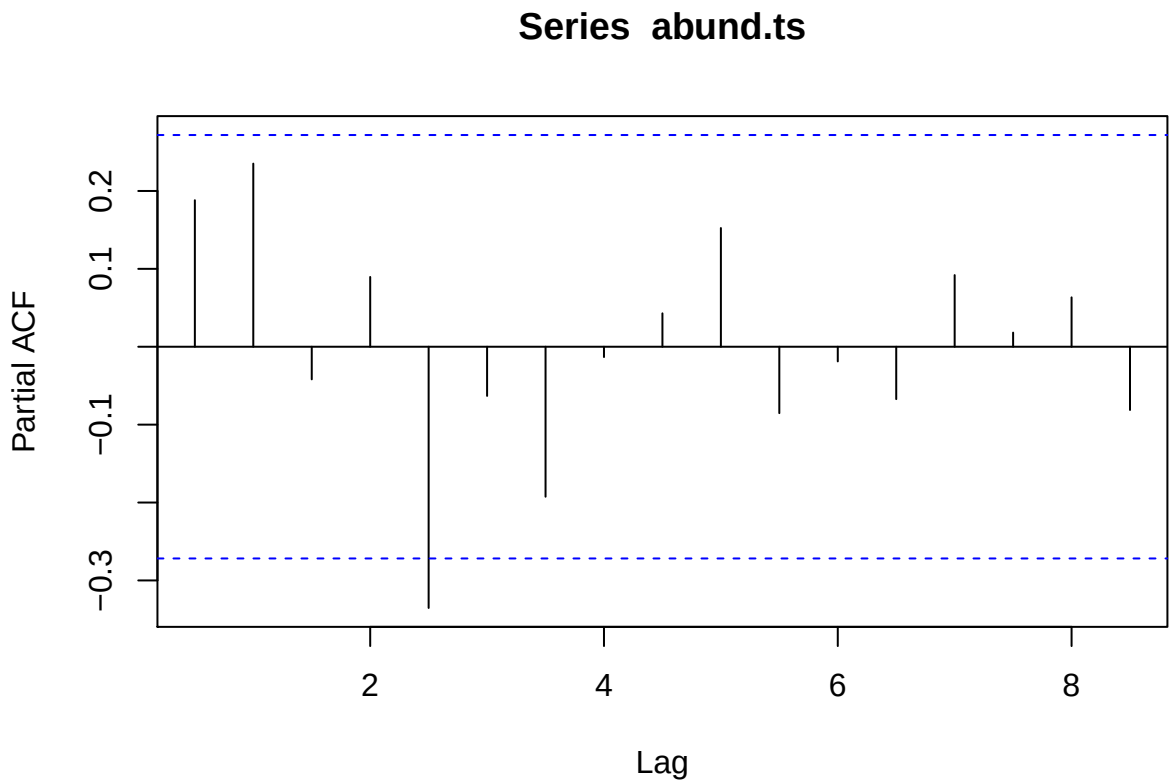
```
#the p value is less than .05, meaning it is not stationary

abund.ts.diff <- diff(abund.ts)
adf.diff <- adf.test(abund.ts.diff, alternative = 'stationary')
adf.diff$p.value
```

```
## [1] 0.03525597
```

```
# this value is under .05

acf(abund.ts)
```

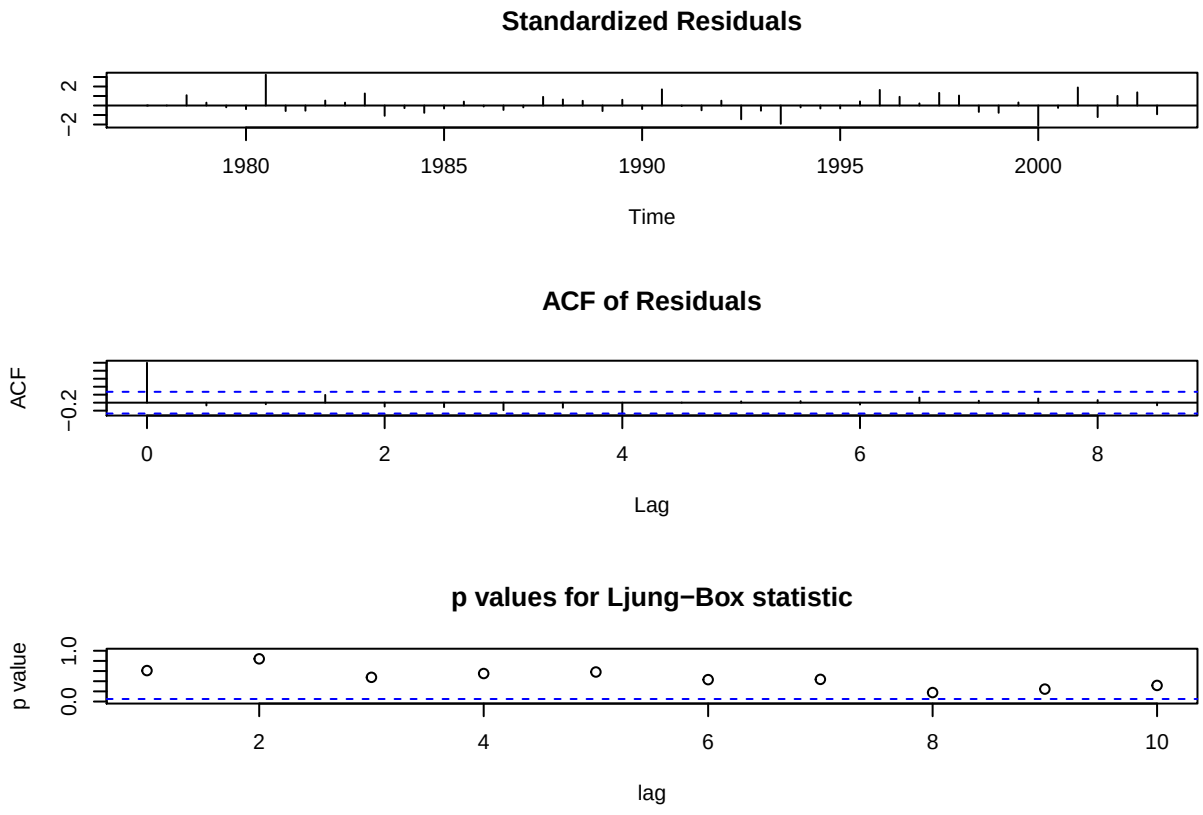



```
pacf(abund.ts)
```

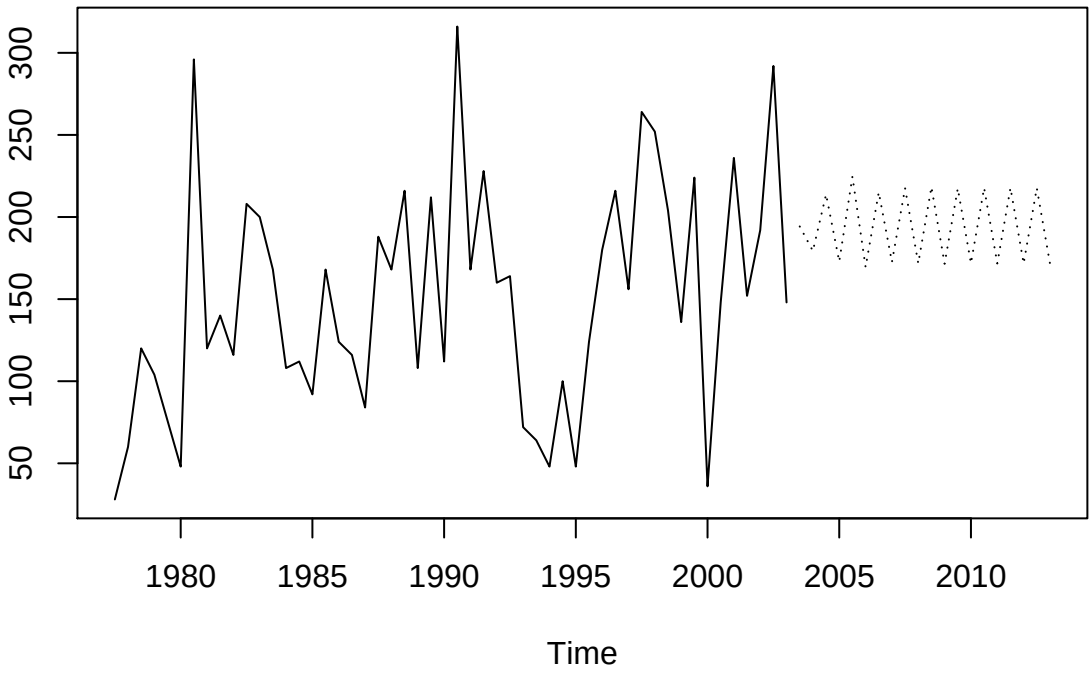



```
abund.arm <- auto.arima(abund.ts)
abund.arm <- arima((abund.ts), c(0,0,1), seasonal = list(order = c(2,1,0), period = 2), include.mean = T
```

```
tsdiag(abund.arm)
```




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



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: My data does not meet the assumption of stationarity, meaning that statistical moments in my time series data are affected by time. However, the time series did meet the assumptions of stationarity after differencing the time series. ***Answer 3b***: The ACF produces plots that help us visualize our data, and tells us about the relationship of moving averages between lagged intervals in our ARMA model. The ACF is different from the ARMA model because it only gives us correlation values between lagged intervals in the ARMA model, it is not a model itself. ***Answer 3c***: The full ARMA model and the other methods in the time series section of the handout tell us that there is seasonality in our time series data, and there is a marginally significant correlation at lag 4 according to the differenced ACF value in the ARIMA model. We predict that the data will continue to have a seasonality component.

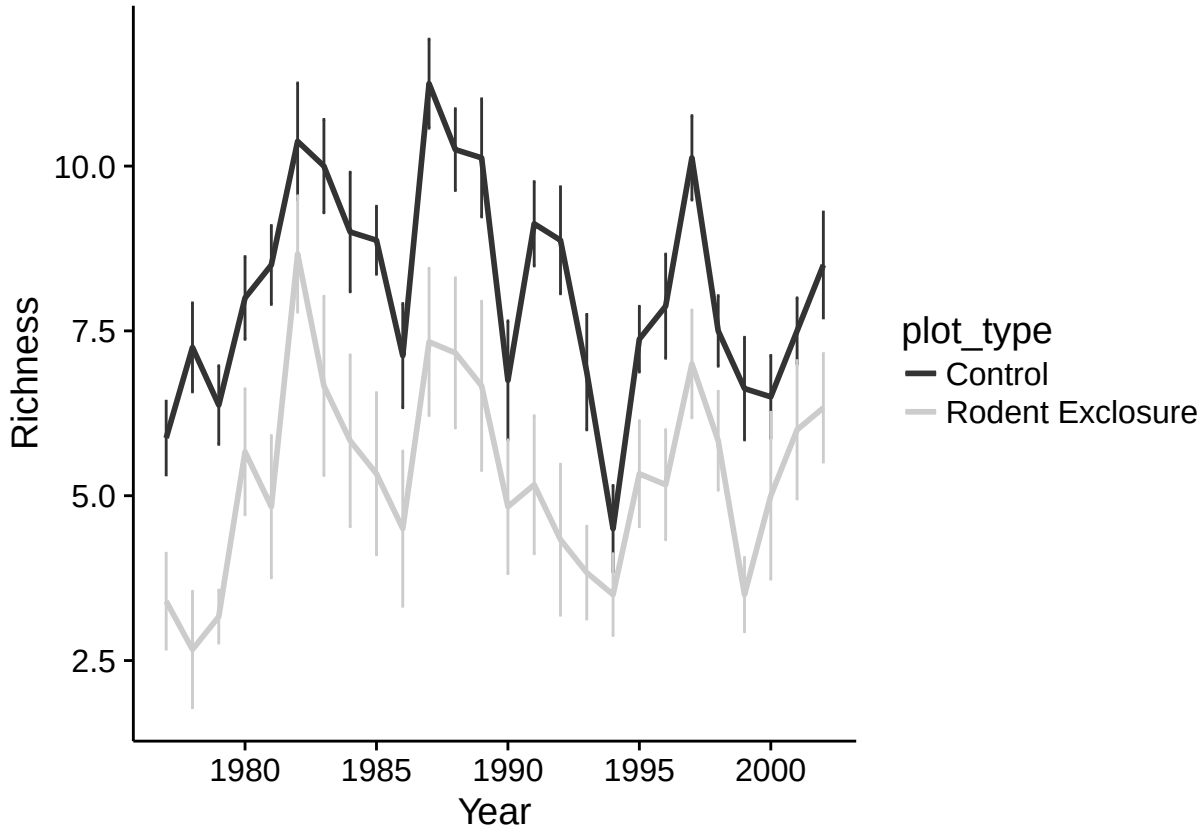
5) REPEATED MEASURES ANALYSIS OF VARIANCE (RM-ANOVA)

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.

```
#construct time-by-species matrix
time.by.species <- group_by(portal, year, plot_id, plot_type) %>% count(taxon) %>% spread(key = taxon, v

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

#create data frame with experimental design and richness data
rich.all <- data.frame(time.by.species[,1:3,],richness)

# rename column
names(rich.all)[4] <- 'richness'

#pull out two of the five Portal treatments
rich.treat <- rich.all[which(rich.all$plot_type == 'Control' | rich.all$plot_type == 'Rodent Exclosure'

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

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)
```




```
#perform rm-ANOVA

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

#look at detailed output
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
```

```
#obtain f-test
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
```

```
#make cleaner ANOVA table
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

```
# use 'lsmeans' package for time-corrected marginal means
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
```

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: An RM-ANOVA test is like an ANOVA test in that it tests for difference between treatment groups. However, it is different in that it accounts for the non-independence of replicate measurements on experimental units. ***Answer 4b***: Simpson's diversity tends to vary a lot from year to year, but I wouldn't call that a noticeable trend. From the graph, it is clear that the plots without rodents consistently had lower richness. ***Answer 4c***: The F-test tells us that our model is a better fit than the null hypothesis, an intercept only model. ***Answer 4d***: The three models are Unstructured, Compound symmetry, and the Autoregressive of order 1 (AR1) model. We used the AR1 model, and this model is probably the most useful for our data because we are not assuming any symmetry between covariance of treatments and there is likely an autoregressive effect in our model, and the other two models do not have such an autoregressive effect.

6) TEMPORAL BETA DIVERSITY

Turnover

In the R code chunk below, do the following:

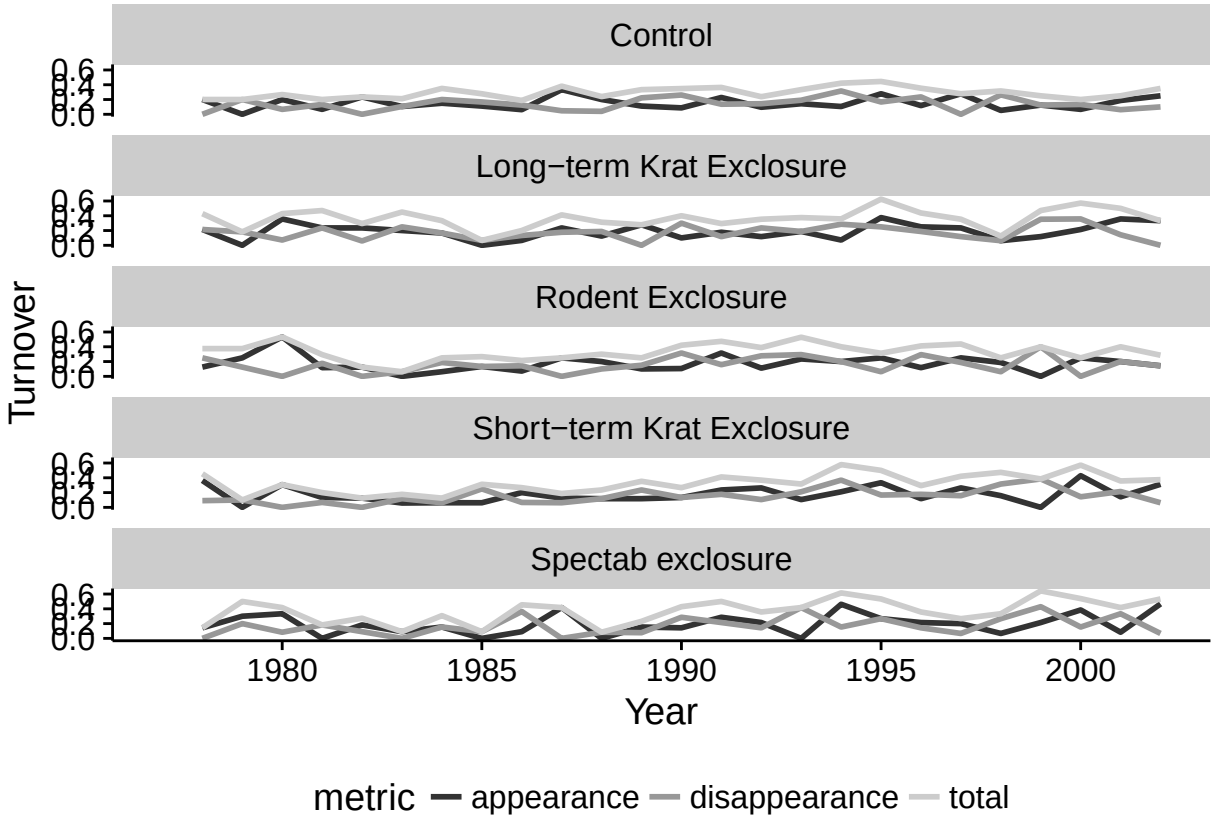
1. Calculate species abundances for each taxonomic group (the `taxa` column).
2. Calculate total turnover and turnover due to the gain/loss of species for each group.
3. Visualize turnover within each group

```
# first, calculate the species abundances from each site over time
portal.species.abunds <- group_by(portal, year, plot_type) %>% count(taxon)

#calculate total turnover
portal.total <- turnover(df = portal.species.abunds,
                         time.var = 'year',
                         species.var = 'taxon',
                         abundance.var = 'n',
                         replicate.var = 'plot_type',
                         metric = 'total')

#calculate species gained
portal.appearance <- turnover(df = portal.species.abunds,
                              time.var = 'year',
                              species.var = 'taxon',
                              abundance.var = 'n',
                              replicate.var = 'plot_type',
                              metric = 'appearance')

#calculate species lost
portal.disappearance <- turnover(df = portal.species.abunds,
                              time.var = 'year',
                              species.var = 'taxon',
                              abundance.var = 'n',
                              replicate.var = 'plot_type',
                              metric = 'disappearance')

#use 'join()' from 'dplyr' to join the columns by shared year & plot type columns
```

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

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

```
# Use 'gather()' from 'tidyr' to convert back to long-form
portal.turnover <- gather(portal.turnover, key = metric, value = turnover, total, appearance, disappear

# Use ggplot to visualize the turnover metrics in all five of the Portal treatments over the entire ann
turn.plot <- ggplot(
  portal.turnover, aes(x = year, y = turnover, color = metric)) +
  geom_line(size = 1, show.legend = T) +
  facet_wrap(~plot_type, 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?
b. Which taxonomic group appears to be the most variable? Which group appears to be the least variable?

Answer 5a: Temporal turnover relates to spatial turnover when temporal turnover changes due to spatial characteristics of species distributions. This can happen in the process of migration.

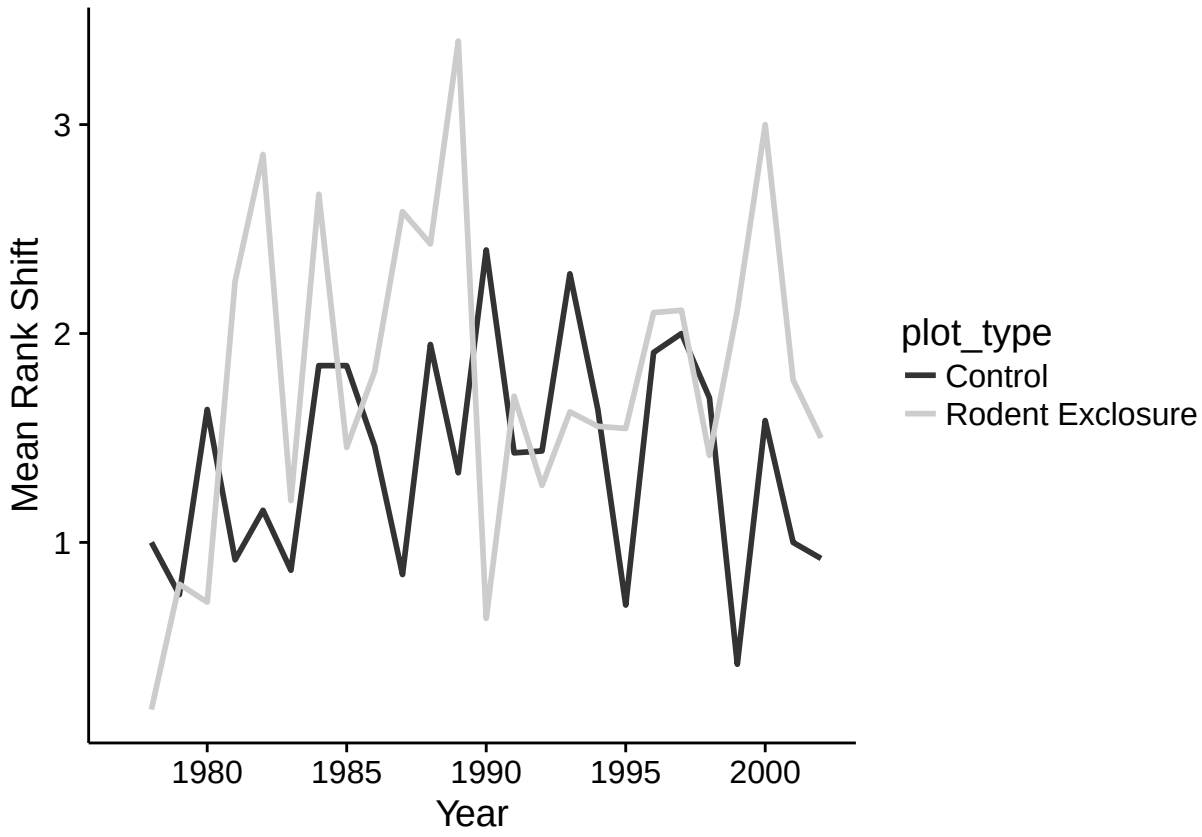
When spatial turnover is high, over time new species can enter the plot via migration due to their close proximity, whereas if spatial turnover is low, temporal turnover will likely not change as fast because new species would have to migrate farther distances to enter the plot. ***Answer 5b***: Based on the visualization, the Spectab exclosure appears to be the most variable and the control seems to be the least variable.

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.

```
#pull out the two treatments we analyzed earlier
portal.abunds.cont.rodent <- filter(portal.species.abunds,
                                    plot_type == 'Control' | plot_type == 'Rodent Exclosure')

# Calculate MRS
portal.rankshift <- rank_shift(
  df = as.data.frame(portal.abunds.cont.rodent),
  time.var = 'year',
  species.var = 'taxon',
  abundance.var = 'n',
  replicate.var = 'plot_type')

# Replace the year range with a single value to plot
portal.rankshift$year <- as.numeric(substr(portal.rankshift$year_pair, 6, 9))

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

plot(rankshift.plot)
```




```
# Does one plot show higher or lower MRS, on average?
group_by(portal.rankshift, plot_type) %>%
  summarise(
    mean = mean(MRS),
    cv = sd(MRS)/mean)
```

```
## # A tibble: 2 × 3
##          plot_type     mean        cv
##              <chr>    <dbl>     <dbl>
## 1          Control 1.400675 0.3759483
## 2 Rodent Exclosure 1.788980 0.4361809
```

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.

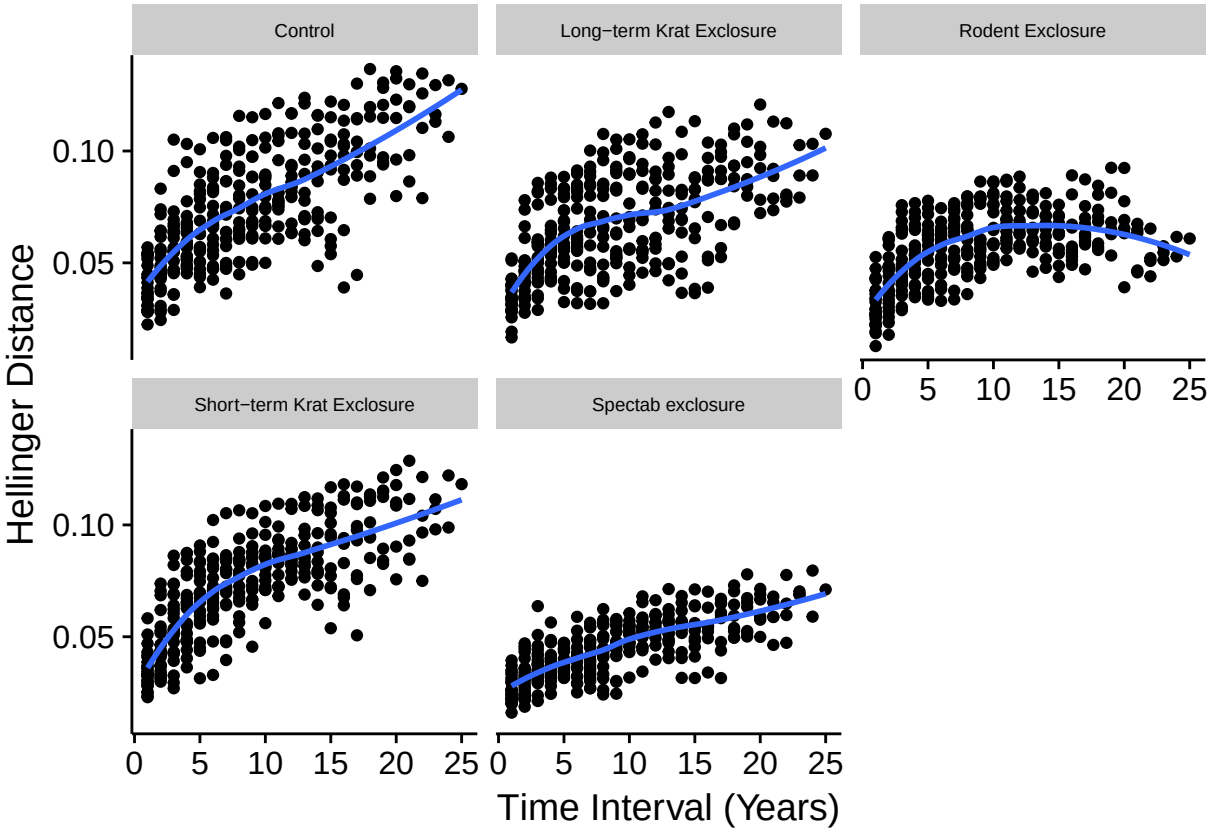
> ***Answer 6a***: A change in rank shift in a community tells us that the commonness and rarity of taxa in our data set have shifted. A higher value indicates a greater shift. ***Answer 6b***: According to the figure we just made, plots tend to go through periods of both large shifts in rank abundances and little shift in rank abundances, depending on the year. The rodent exclosure plot type tends to go through larger rank abundance shifts than the control.

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.

```r
# C. Rate change interval

# In order to calculate relative abundances, we need total abundances
# Let's add a column for total abundances
# We will relativize species abundances across the whole dataset so
# the transformed distances are preserved
portal.species.abunds$tot.abund <- rep(sum(portal.species.abunds$n),
                                       length(portal.species.abunds$n))

# Now, apply Hellinger transformation
portal.hellinger.transf <- portal.species.abunds %>%
  mutate(hellinger.transf = sqrt(n / tot.abund))

# The mutate function creates a new column 'hellinger.transf'
# by taking the square root of species relative abundance

# We can use this new column as our 'abundance' vector
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: Calculating a distance metric across varying time intervals means calculating site dissimilarity at different time intervals, and can be used to determine how fast certain communities diverge over time. ***Answer 7b***: According to our data visualization, it appears that the Spectab exclosure and the rodent exclosure diverge from their original community composition at slower rates than the other treatments, especially the rodent exclosure treatment, which does not seem to be diverging more with increasing time past 10 years. The control, long term krat exclosure, and the short term krat exclosure diverge from their original community compositions quickly and from the given data, they seem to continue to change with time.

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.

```
# A. Community 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

```
# B. Species Synchrony

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"
```

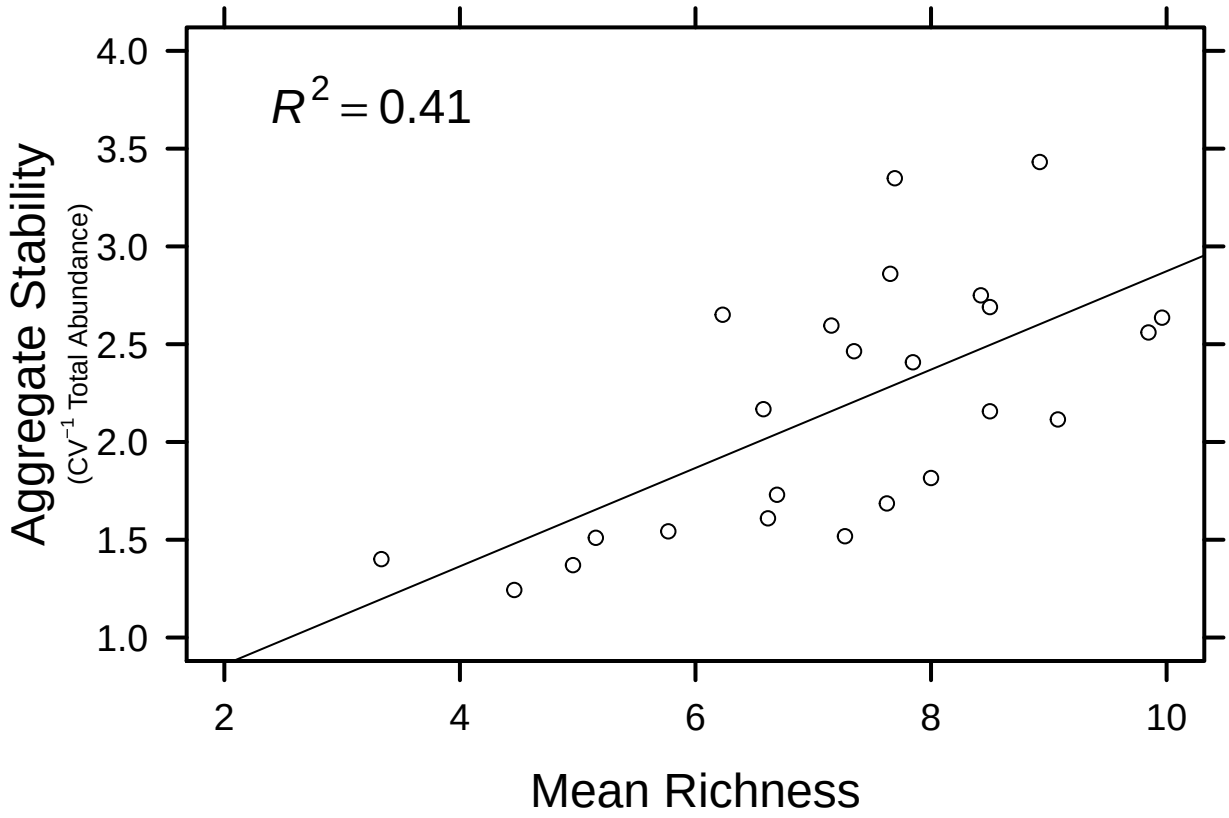
plot_type	loreau	gross
Control	0.1869	0.1263
Long-term Krat Exclosure	0.1578	0.07418
Rodent Exclosure	0.187	0.1423
Short-term Krat Exclosure	0.08773	0.001546
Spectab exclosure	0.1542	0.1393

```
# D. Biodiversity-Stability Relationships

# Recall, earlier we calculated richness in each plot type in each year
# Let's group only by plot_if
# Then, we we summarise average annual richness in each plot type
portal.mean.rich.plot <- rich.all %>%
  group_by(plot_id) %>%
  summarise(mean.rich = mean(richness))

#Let's take a look at how stability metrics relate to 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')

# Join richness and stability
portal.div.stab <- portal.mean.rich.plot %>%
  inner_join(portal.stab.plot)
```

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

```
# Plot the relationship
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)
```



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: According to the stability measure, the control plot type has the highest stability in total abundance. Stability of total abundance is measured using the inverse of the coefficient of variation (CV). The CV is measured by dividing the standard deviation by the mean.

Answer 8b: Species synchrony analyzes whether abundance fluxuations of different species impact each other, i.e., whether the abundances are independent from each other or not. If two species densities are not independent from one another, this means that one impacts the other's density, or they both affect each other's densities.

Answer 8c: The biodiversity-stability relationships we just analyzed tell us that aggregate stability tends to increase with mean richness.

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: There are many ways to look at data that have temporal and spatial components. Autocorrelation can be used when you suspect that your data have time lag effects. When we take multiple measurements of an experimental unit in data sets, we can use RM-ANOVA to account

for the non-independence of those measurements. Relating to spatial diversity, distance decay patterns show the relationship between increasing geographic distance and decreasing similarity between environments. Spatial scale is important to consider because the extent and grain you choose to analyze your data can make your results look differently. Species-area relationships can help determine the rate at which sampling larger areas reveal more species. Overall, these metrics are diverse and use a wide range of methods to conclude facts about biodiversity from data. Studying biodiversity across time and space can be challenging due to the wide range of unknown factors in field data. A large amount of data can also make studying biodiversity relationships overwhelming because of the large amount of relationships that could exist. And of course, these biodiversity metrics are the most useful and interesting with good data, and sometimes it is hard to find something in a data set where there are no obvious relationships between variables.