Advanced Ecology 470 R tutorials; UVic Feb 2015

Created by JPWR, 21-Feb-2015

Diversity statistics

1. Cleaning the data frame

This week we're going to use a NOAA dataset to explore patterns in reef fish diversity at 13 Hawaiian islands.

The data were collected by underwater visual census over 2010-2012 - details on the project can be found here: http://www.pifsc.noaa.gov/cred/fish

First thing, as always, is to load in the data frame and explore its structure. Use the commands you've learned in earlier workshops to help you do this.

Are there any errors in the data frame? Ecological datasets are often messy and require cleaning - we need to examine the levels of factors, and unique values of character strings, to ensure that the data are ready for analysis.

```
unique(cred$island)
levels(cred$trophic)
range(cred$count)
```

You want to go through every variable in the data frame and examine unique values (for words) and outliers (for numbers). In this data frame, some species names have been coded incorrectly - we can fix the data in R and then continue exploring.

```
# change HAWAII to Hawaii
cred$island[cred$island=="HAWAII"]<-"Hawaii"
unique(cred$island) ## Now we have clean island names.
```

For character strings (i.e. words) R remembers past values - these are the levels of a factor or character vector. When you remove typos, it's useful to alsoremove them from the levels. We use droplevels() to reset the levels of a data frame.

```
cred<- droplevels(cred)
unique(cred$island)  ## HAWAII is no longer a level.
```

We also have NA values for some herbivores at Palmyra - have a look using is.na.

```
cred[is.na(cred$count),]
```

We should remove these rows too - we can't use the data for diversity analyses. When working with large dataframes, it's useful to check your edits as you go. You could use dim() to see if the data frame has removed the correct number of rows.

```
dim(cred)
cred<-cred[!is.na(cred$count),]
dim(cred) ## 9 rows removed
cred[is.na(cred$count),]   ### also shows the NAs have been removed
```

2. Species richness as an R function

We're going to explore the diversity of trophic groups across islands. First, we're going to look at patterns in species richness - the total number of species observed. In the full dataset, we can just look at the number of unique species:

```
unique(cred$species)
length(unique(cred$species))
```

The total species richness, in the Pacific, at these 13 islands, is 121. Let's create a function that allows us to estimate species richness quickly, and then explore patterns across islands and trophic groups.

```
richness.func<-function( x){
    species<-unique(x)
    richness<-length(species)
    richness
}
```

Now, if we use the aggregate function, we can estimate species richness for different subsets of the dataset. Remember to check ?aggregate for the arguments.

```
aggregate(species ~ island, cred, richness.func)
aggregate(species ~ island + trophic, cred, richness.func)
aggregate(species ~ region + trophic, cred, richness.func)
```

We can also create tables of species abundances across islands. Tables aren't very easy to read in R, but these kinds of commands help to reveal patterns in the data. We can also wrap the table command in a barplot() to create a figure.

```
table( cred$species, cred$island)
table(cred$trophic, cred$island)
barplot(table(cred$trophic, cred$island))
```

3. Diversity estimates in the vegan package

Instead of manually creating functions each time we want to calculate an ecological statistics, there are often packages in R that have complex functions built in. The vegan package is used for analyses in community ecology - we're going to use its built in diversity functions. You may be asked to choose a 'mirror' - this is where the package will be downloaded from. Simon Fraser University hosts R packages, so you can select Canada (BC) here.

```
install.packages("vegan")
require(vegan)
?vegan
```

vegan works with abundance data in matrix form. Let's convert the dataset into a table of species abundances.

```
abundance<-table( cred$island,cred$species )
```

Now we can estimate the species richness and calculate diversity indices.

```
specnumber(abundance)
aggregate(species ~ island, cred, richness.func)

diversity(abundance, index="shannon")
diversity(abundance, index="simpson")
diversity(abundance, index="invsimpson")
```

Diversity estimates are strongly influenced by sample size (think species-area curves). We should consider rarifying our estimates to account for differences in the numbers of individuals observed.

How many individuals were observed per island?

```
rowSums(abundance)
rarefy(abundance, min(rowSums(abundance)))  ### estimates rarefied by number of individuals
```

We can save these estimates, add in our predictor variables, and begin exploring the drivers of reef fish community diversity.

```
reef<- data.frame(rarefy(abundance, min(rowSums(abundance))))

## add in predictor variables using match()
reef$region<-cred$region[match(rownames(reef), cred$island)]
reef$sst<-cred$sst[match(rownames(reef), cred$island)]
reef$state<-cred$state[match(rownames(reef), cred$island)]
reef$log_population<-cred$log_population[match(rownames(reef), cred$island)]
reef$productivity<-cred$productivity[match(rownames(reef), cred$island)]
## add column name for richness
colnames(reef)[1]<-"richness"
reef

plot(reef$productivity, reef$richness, col=reef$region, pch=as.numeric(reef$state), cex=1.5, xlab="Productivity", ylab="Rarefied ri
legend(0.075,50, legend=c( "MHI","PRIAs", "Remote", "Disturbed"), col=c(1,1,2,1), pch=c(1,2,1,1))
```

Next week, we'll use some statistical tests to examine the drivers of species diversity at Pacific coral reefs.