



Biomar_species: the number of unique species found in the sample

Poss_New_species: species in the sample that are undescribed (new to science)

New_to_ACG: species in the sample not previously known to occur in ACG

New_to_Costa_Rica: species in the sample not previously known to occur in Costa Rica

Total_species_ACG: previous species count know to occur in ACG plus 'New_to_ACG'

Two further notes:

1. You don't have all the variables that you need to plot in order to answer some obvious questions. For example, you don't know how many species had been documented for the ACG before this effort, *but* you can calculate it given what you have.
2. There are a few rows in the data that need to be handled differently than others. For example, there is a row for the TOTAL number of Crustaceans. We need that because the data on Total_species_ACG isn't broken down into subgroups. So for some comparisons you will have to use some subsets of the rows and for other comparisons you will have to add up different subsets. I'll let you figure out how to do this, but do this subsetting within python to practice your pandas skills (in other words, don't just go into Excel and make two separate datasets -- even this would probably be the easiest thing to do).

* The box plot will be the hardest thing to figure out. For this in particular you may want to think about how you can create different values by combining the existing columns in different ways. For example, you may want to plot some variables that aren't just driven by the large differences in sample size.

```
In [18]: # pie chart
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline

pie = ([3,1,6,5,6,5,10,5])
lab = ('Cyanophyta', 'Chlorophyta', 'Ochrophyta', 'Rhodophyta', 'Porifera', 'Mollisca', 'Annelida', 'Nemertea')
exp = [0,0,0,0,0,0,.15,0]
```