

Intro to Visualization Final Activity

We've made it this far, congratulations! I hope at this point that you are feeling confident in your ability to deploy python and matplotlib to create informative and attractive data visualizations! You should feel like you have a new tool in your scientific toolkit. I also hope that that this semester's activities have helped to reinforce some of your other python skills. The more you practice with things like loops, conditionals, and the various python data structures (lists, dictionaries, and dataframes) the less daunting they seem. As I've said many times, you can always look up the syntax again later, but every time you complete one of these notebooks you are building up base skills in computational thinking that I think will allow you to problem-solve in your own projects someday.

So let's give you a chance to show off!

I've included a dataset in this folder called `BioMar.csv`. Your challenge is to summarize that data set using visualizations to create a set of figures that tell the 'story' of the data as you see it. Within the data folder I have included an article (along with its supplementary data file) that describes this recent data from the marine zone of Area de Conservación Guanacaste (you'll see on the first page of the paper that this is part of a special issue of the journal Biotropica that I was involved in producing). The paper is describing the results of a new biodiversity surveying effort. The main concern is documenting previously undocumented marine species within the conservation area.

Your answer should include AT LEAST one each of:

- pie chart
- histogram
- bar plot
- scatter plot
- box plot
- heat map
- annotated figure
- multi-plot layout

and for each individual plot that you make your code should export the figure as a .pdf file named `LastName_Fig_X.pdf` where `LastName` is your last name and `X` is the number for each figure.

The data that I have given you is drawn from the tables in the paper and the supplementary material.

```
Biomar_samples: the total organisms collected from within the particular taxonomic group

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.

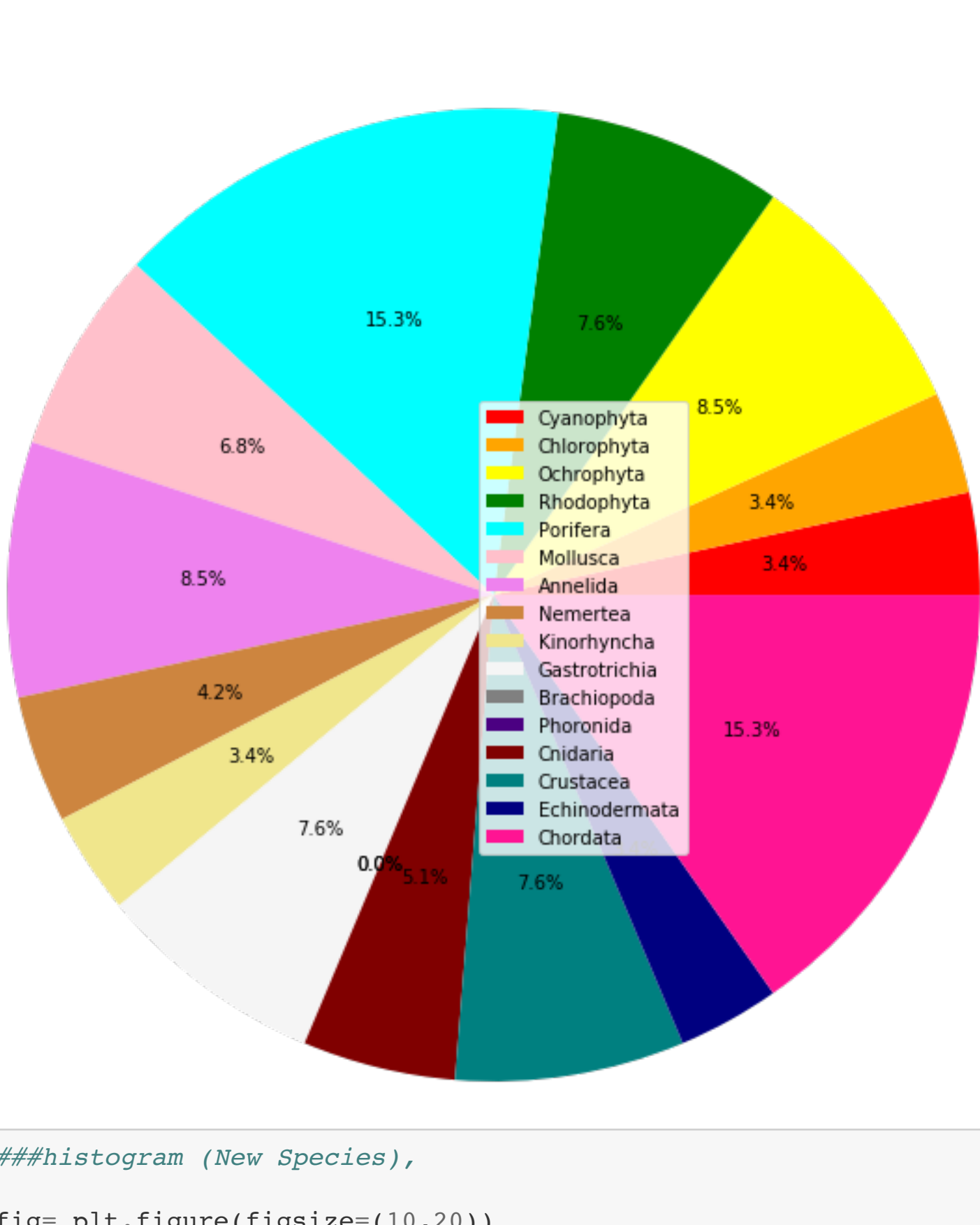
```
In [76]: ##pie chart (New to Costa Rica)
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from matplotlib.backends.backend_pdf import PdfPages
final_df = pd.read_csv(r'C:\Users\renie\Downloads\6_final_project\6_final_project\data\BioMar.csv')

totals= final_df[final_df['Taxonomic_group'].str.contains("_") | final_df['Taxonomic_group'].str.contains("TOTAL")]

#totals['Taxonomic_group']
#totals['New_to_Costa_Rica']

lab= totals['Taxonomic_group'].str.replace('_TOTAL','')
colors =['red','orange','yellow','green','cyan','pink','violet','peru','khaki','whitesmoke','grey','indigo',
'maroon','teal','navy','deeppink']

plt.pie(totals['New_to_Costa_Rica'],colors= colors, autopct= '%.1f%%', radius= 3)
plt.title('Percentage of Species New to Costa Rica per Taxonomic Group', y=2)
plt.legend(loc='best')
plt.show()
fig.savefig('Cooper_Fig_1.pdf')
```

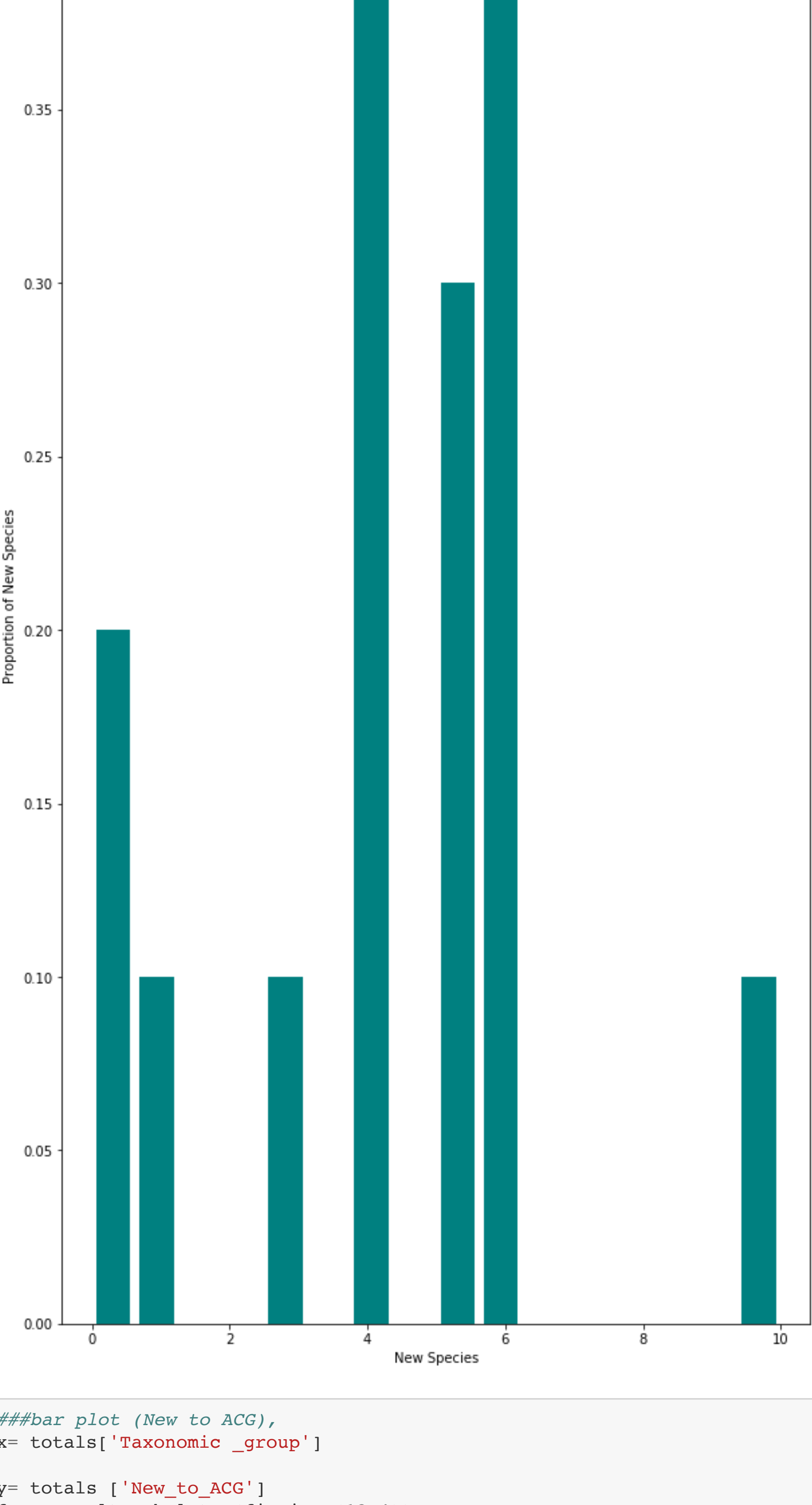


```
In [11]: ##histogram (New Species),
fig= plt.figure(figsize=(10,20))
x= totals['Poss_New_species']

plt.hist(x, 16, rwidth= 0.8, density = True, color= 'teal')

plt.xlabel('New Species')
plt.ylabel('Proportion of New Species')
fig.suptitle('Possible New Species Distribution across Taxonomic Group')
plt.show()
fig.savefig('Cooper_Fig_2.pdf')
```

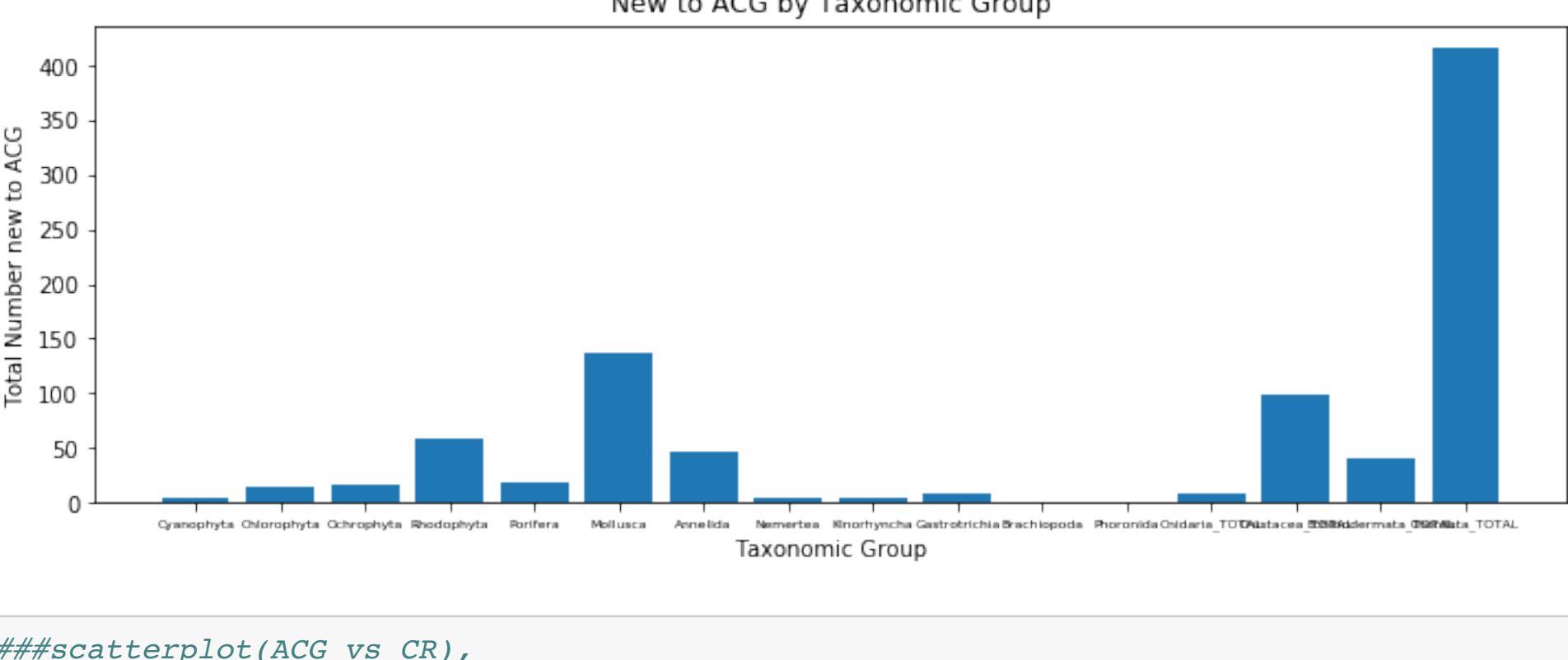
Possible New Species Distribution across Taxonomic Group



```
In [12]: ##bar plot (New to ACG),
x= totals['Taxonomic_group']

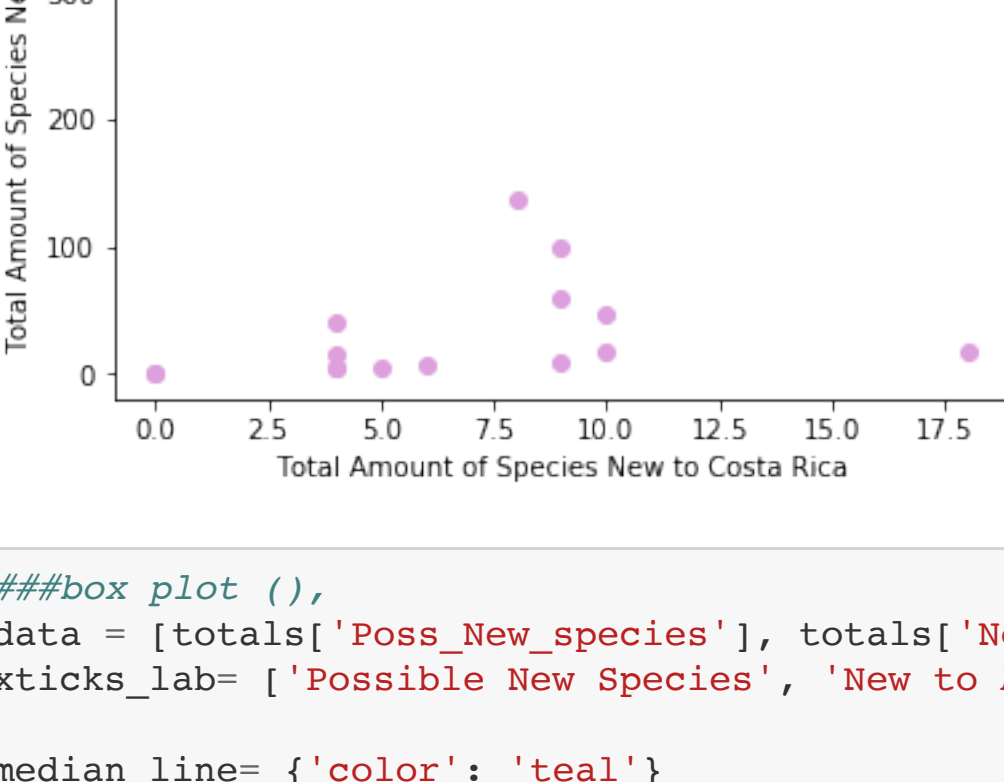
y= totals['New_to_ACG']
f, ax = plt.subplots(figsize=(12,4))
plt.bar(x,y, width = 0.8)
plt.xticks(x, fontsize=6)
plt.title('New to ACG by Taxonomic Group')
plt.xlabel('Taxonomic Group')
plt.ylabel('Total Number new to ACG')

plt.show()
fig.savefig('Cooper_Fig_3.pdf')
#I was trying to figure out a way to view the x labels better by modifying the figsize but it was always hard to s
ee cof
```



```
In [13]: ##scatterplot(ACG vs CR),
y= totals['New_to_ACG']
x= totals['New_to_Costa_Rica']

plt.scatter(x,y, color= 'plum')
plt.xlabel('Total Amount of Species New to Costa Rica')
plt.ylabel('Total Amount of Species New to ACG')
plt.show()
fig.savefig('Cooper_Fig_4.pdf')
```



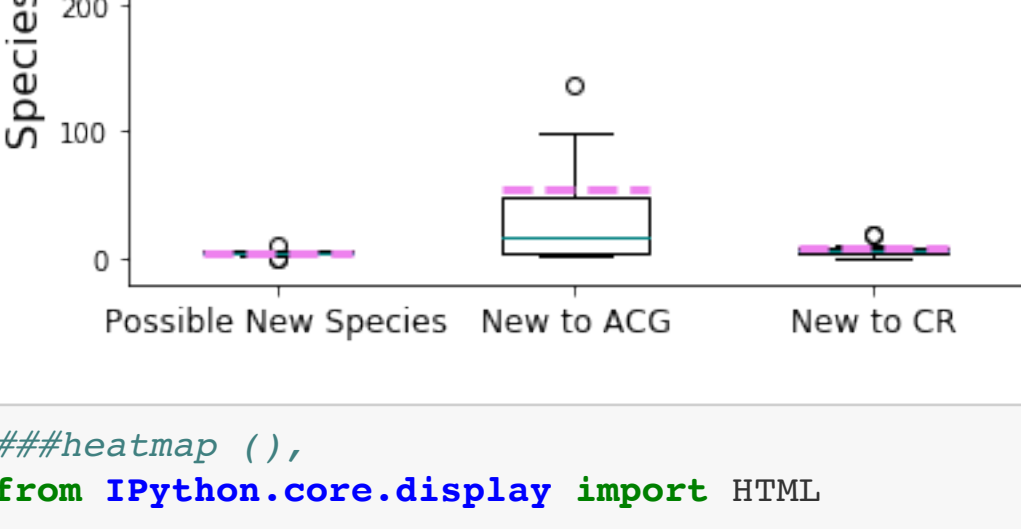
```
In [14]: ##box plot (,),
data = [totals['Poss_New_species'], totals['New_to_ACG'], totals['New_to_Costa_Rica']]
xticks_labels = ['Possible New Species', 'New to ACG', 'New to CR']

median_line= {'color': 'teal'}
mean_line= {'color': 'violet',
'linestyle': 'dashed',
'linewidth': 3}

fig, ax = plt.subplots()
plt.boxplot(data, widths = 0.5, showmeans = True, medianprops = median_line, medianprops = median_line)

plt.xticks([1,2,3], xticks_lab, fontsize = 12)

plt.ylabel('Species Count', fontsize = 16)
plt.xlabel('New Species in ACG and CR', fontsize = 16, fontweight = 'bold')
plt.title('New Species in ACG and CR', fontsize = 16, fontweight = 'bold')
plt.show()
fig.savefig('Cooper_Fig_5.pdf')
```



```
In [88]: ##heatmap (,),
from IPython.core.display import HTML

fig= plt.figure(figsize =(14,8))
ax= plt.subplot()

species= totals.groupby(totals.index).mean()

plt.pcolormesh(species, cmap='YlGn', edgecolor='white', alpha = 0.8 )
plt.colorbar()

ax.spines['bottom'].set_color('white')
ax.spines['left'].set_color('white')
ax.spines['top'].set_color('white')
ax.spines['right'].set_color('white')

# i in range (len(totals[Biomar_species].index)):
# for j in range(len(totals.columns)):
# plt.text(j+0.2, i+ 0.4, totals.iloc[i,j])

plt.yticks(np.linspace(0.5, len(species.index)-0.5, len(species.index)), species.index, size=13)
plt.xticks(np.linspace(0.5, len(species.columns)- 0.5, len(species.columns)), species.columns,size=10)
plt.title('Species Count? Not sure what this reflects', size= 18)

plt.show()
fig.savefig('Cooper_Fig_6.pdf')

##gonna be honest, I'm not sure what this shows, but this code was the only
way I could get a heatmap to come up. I tried to drop columns like before but te code said that was coming back
#as an error, and when I tried to specify columns it wouldn't let me

#
```

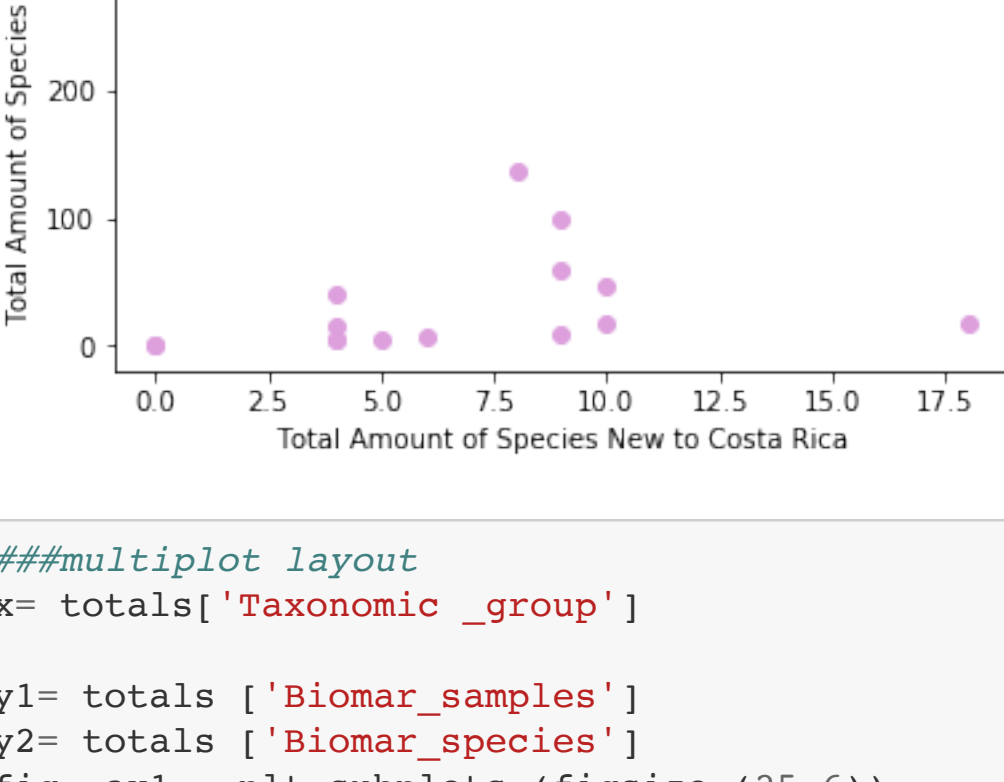


```
In [81]: ##annotated figure,
y= totals['New_to_ACG']
x= totals['New_to_Costa_Rica']

plt.scatter(x,y, color= 'plum')
plt.xlabel('Total Amount of Species New to Costa Rica')
plt.ylabel('Total Amount of Species New to ACG')

plt.annotate('449 New Chordata\nSpecies were found in ACG ', (x[13], y[13]), xytext= (0.6, 0.8), textcoords='axes fra
ction')

plt.show()
fig.savefig('Cooper_Fig_7.pdf')
```



```
In [74]: ##multiplot layout
x= totals['Taxonomic_group']

y1= totals ['Biomar_samples']
y2= totals ['Biomar_species']
fig, ax1 = plt.subplots(figsize=(25,6))
ax1.bar(x,y1, width = 0.8, color='cyan')
ax1.set_title('New to ACG by Taxonomic Group')
ax1.set_xlabel('Taxonomic Group', fontsize= 20)
ax1.set_ylabel('Total Species', fontsize= 20)

ax2 =ax1.twinx()
ax2.set_ylabel('New Species', fontsize= 20)
ax2= plt.plot(x, y2, color='black')

plt.show()
fig.savefig('Cooper_Fig_8.pdf')
```

