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 a 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:  speciesin 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 [8]:

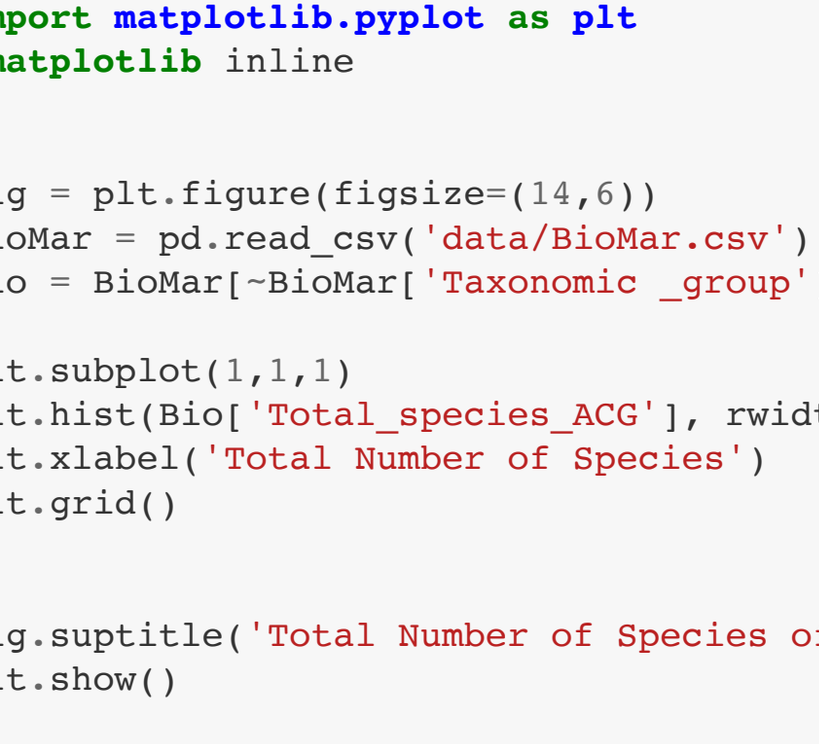
```python
import pandas as pd

BioMar = pd.read_csv("data/BioMar.csv")

BioMar.head()

display(BioMar)
```

	Taxonomic _group	Biomar_samples	Biomar_species	Poss_New_species	New_to_ACG	New_to_Costa_Rica	Total_species_ACG
0	Cyanophyta	11	5	3	4	4	8.0
1	Chlorophyta	45	19	1	15	4	19.0
2	Ochrophyta	37	23	6	17	10	25.0
3	Rhodophyta	137	62	5	59	9	74.0
4	Porifera	167	18	6	18	18	18.0
5	Cnidaria_Hydrozoa	53	13	1	3	2	NaN
6	Cnidaria_Scleractinia	56	15	0	1	0	NaN
7	Cnidaria_Octocorallia	20	16	3	2	3	NaN
8	Cnidaria_Anthipatharia	2	2	1	1	1	NaN
9	Cnidaria_Pennatulacea	1	1	1	1	0	NaN
10	Mollusca	706	166	5	137	8	324.0
11	Annelida	536	70	10	46	10	73.0
12	Nemertea	25	5	5	5	5	6.0
13	Kinorhyncha	7	4	4	4	4	4.0
14	Gastrotrichia	9	9	4	9	9	9.0
15	Brachiopoda	1	1	0	1	0	1.0
16	Phoronida	1	1	0	1	0	1.0
17	Crustacea_Decapoda	2499	200	6	95	8	NaN
18	Crustacea_Stomatopoda	151	9	1	4	1	NaN
19	Echinodermata_Asteriodea	56	7	0	6	0	NaN
20	Echinodermata_Ophiuroidea	159	13	1	13	1	NaN
21	Echinodermata_Holothuroidea	192	26	3	11	3	NaN
22	Echinodermata_Echinoidea	90	11	0	10	0	NaN
23	Chordata_Urochordata	95	11	2	11	9	NaN
24	Chordata_Elasmobranchii	27	14	0	14	1	NaN
25	Chordata_Actinopterygii	2260	402	2	391	8	NaN
26	Cnidaria_TOTAL	132	47	6	8	6	53.0
27	Crustacea_TOTAL	2650	209	6	99	9	292.0
28	Echinodermata_TOTAL	497	57	4	40	4	60.0
29	Chordata_TOTAL	2382	427	4	416	18	449.0

Pie Chart

In [103]:

```python
import matplotlib.pyplot as plt
import pandas as pd

pie = ([3, 1, 6, 5, 6, 5, 10, 5])

plt.title('Possible New Species of 8 Taxonomic Groups')

kind = ('Cyanophyta','Chlorophyta','Ochrophyta','Rhodophyta','Porifera','Mollisca','Annelida', 'Nemertea')



colors = ['violet', 'pink', 'grey','teal','black', 'maroon', 'blue', 'purple']

plt.pie(pie, labels=kind, colors=colors)



plt.show()

plt.savefig('Ohlweiler_Fig_1.pdf')
```

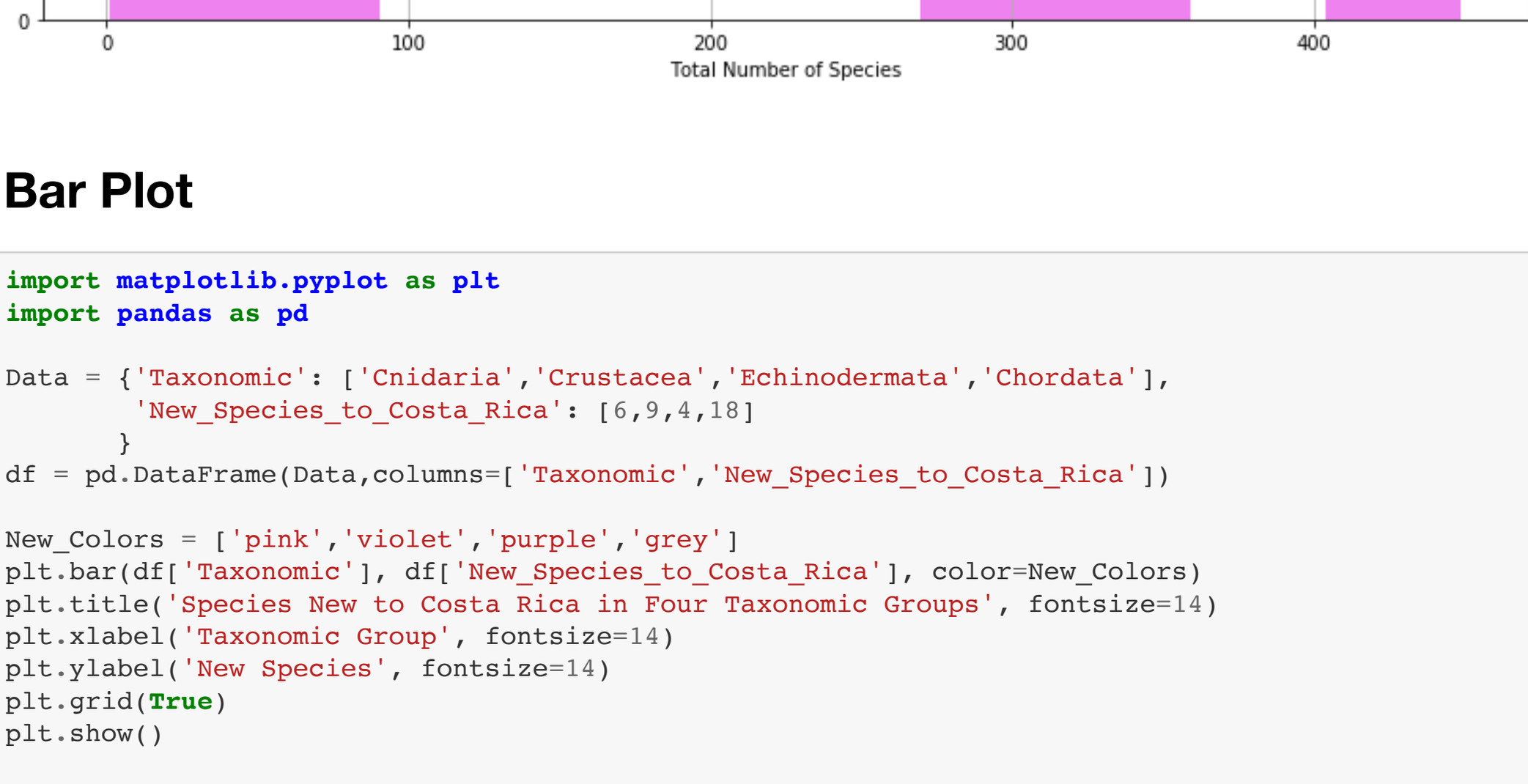



```
<Figure size 432x288 with 0 Axes>
```

Histogram

In [88]:

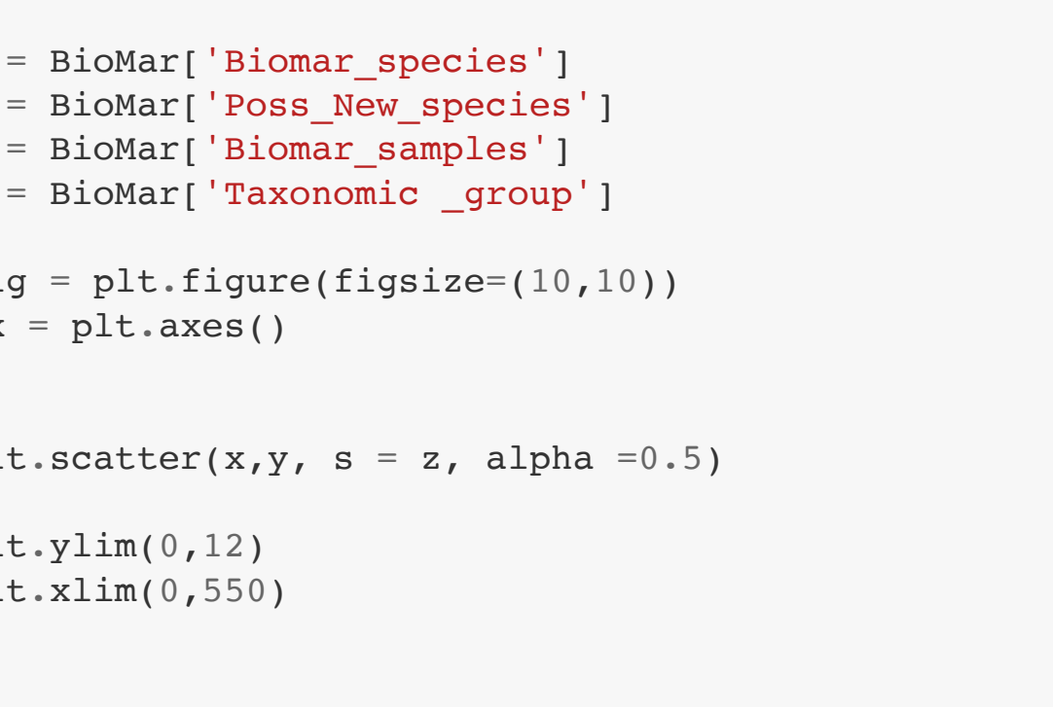
```python
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline


fig = plt.figure(figsize=(14,6))
BioMar = pd.read_csv('data/BioMar.csv')
Bio = BioMar[~BioMar['Taxonomic _group'].str.contains("_") | BioMar['Taxonomic _group'].str.contains("TOTAL")]

plt.subplot(1,1,1)
plt.hist(Bio['Total_species_ACG'], rwidth =1, color = 'violet')
plt.xlabel('Total Number of Species')
plt.grid()


fig.suptitle('Total Number of Species of Taxonomic Groups', size =14)
plt.show()

plt.savefig('Ohlweiler_Fig_2.pdf')
```



Bar Plot

In [101]:

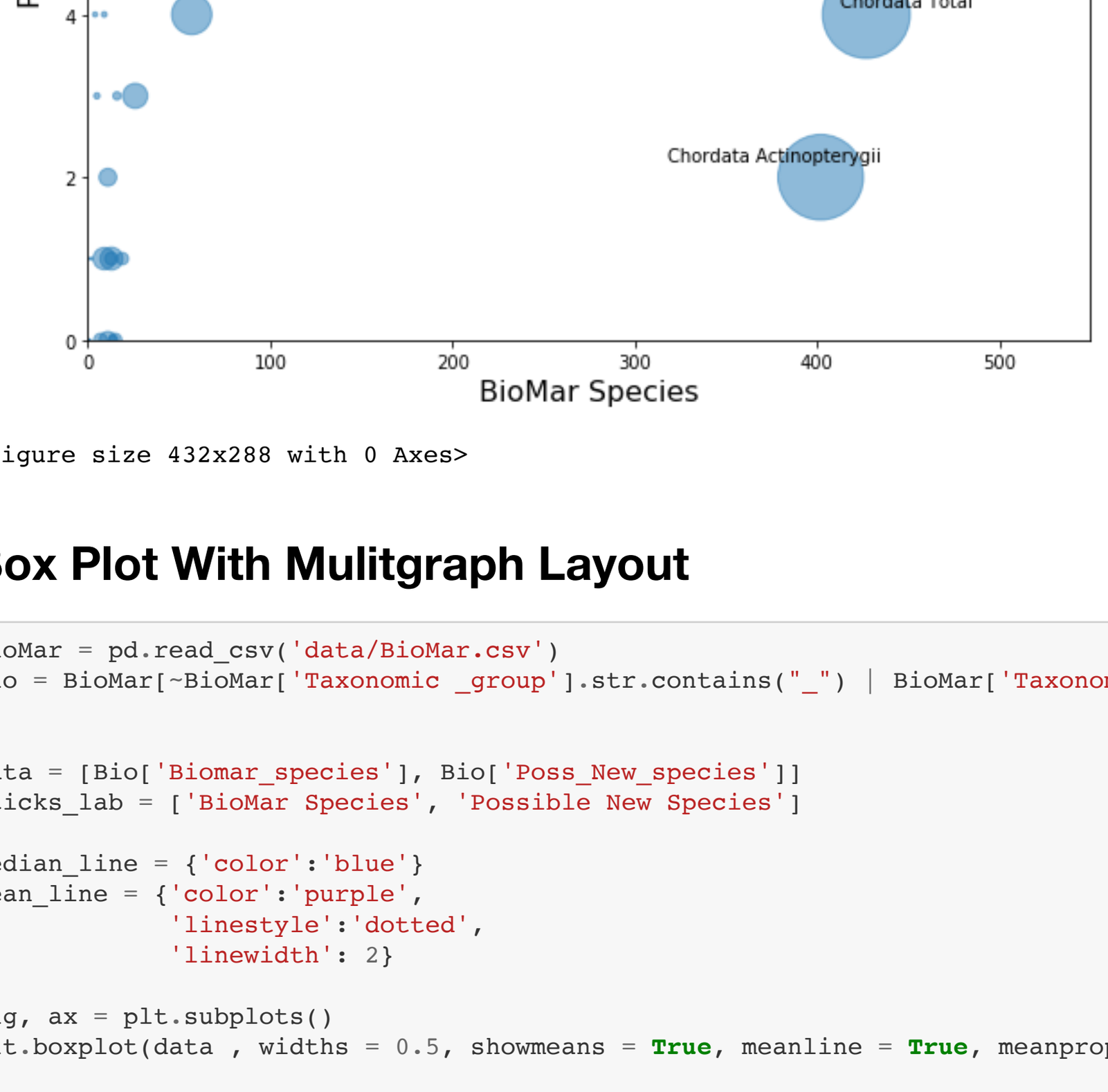
```python
import matplotlib.pyplot as plt
import pandas as pd

Data = {'Taxonomic': ['Cnidaria','Crustacea','Echinodermata','Chordata'],
        'New_Species_to_Costa_Rica': [6,9,4,18]
       }
df = pd.DataFrame(Data,columns=['Taxonomic','New_Species_to_Costa_Rica'])

New_Colors = ['pink','violet','purple','grey']
plt.bar(df['Taxonomic'], df['New_Species_to_Costa_Rica'], color=New_Colors)
plt.title('Species New to Costa Rica in Four Taxonomic Groups', fontsize=14)
plt.xlabel('Taxonomic Group', fontsize=14)
plt.ylabel('New Species', fontsize=14)
plt.grid(True)
plt.show()

plt.savefig('Ohlweiler_Fig_3.pdf')
```



Annotated Scatter Plot

In [134]:

```python
BioMar = pd.read_csv('data/BioMar.csv')

x = BioMar['Biomar_species']
y = BioMar['Poss_New_species']
z = BioMar['Biomar_samples']
c = BioMar['Taxonomic _group']

fig = plt.figure(figsize=(10,10))
ax = plt.axes()


plt.scatter(x,y, s = z, alpha =0.5)

plt.ylim(0,12)
plt.xlim(0,550)


#for i in range(len(x)):
    #plt.text(x[i],y[i],c[i], fontsize = 10)

ax.annotate('point offset from data',
            xy=(65, 10.25), xycoords='data',
            xytext=(-15, 25), textcoords='offset points',
            arrowprops=dict(facecolor='black', shrink=0.05),
            horizontalalignment='right', verticalalignment='bottom')

ax.annotate('Annelida',
            xy=(120, 9.25), xycoords='data',
            xytext=(-15, 25), textcoords='offset points',
            horizontalalignment='right', verticalalignment='bottom')
ax.annotate('Chordata Total',
            xy=(500, 3.5), xycoords='data',
            xytext=(-15, 25), textcoords='offset points',
            horizontalalignment='right', verticalalignment='bottom')

ax.annotate('Chordata Actinopterygii',
            xy=(450, 1.6), xycoords='data',
            xytext=(-15, 25), textcoords='offset points',
            horizontalalignment='right', verticalalignment='bottom')

plt.ylabel('Possible New Species', fontsize = 16)
plt.xlabel('BioMar Species', fontsize = 16)
plt.title('Possible New Species Compared to Known Species', fontsize = 16, fontweight = 'bold')

plt.show()

plt.savefig('Ohlweiler_Fig_4.pdf')
```

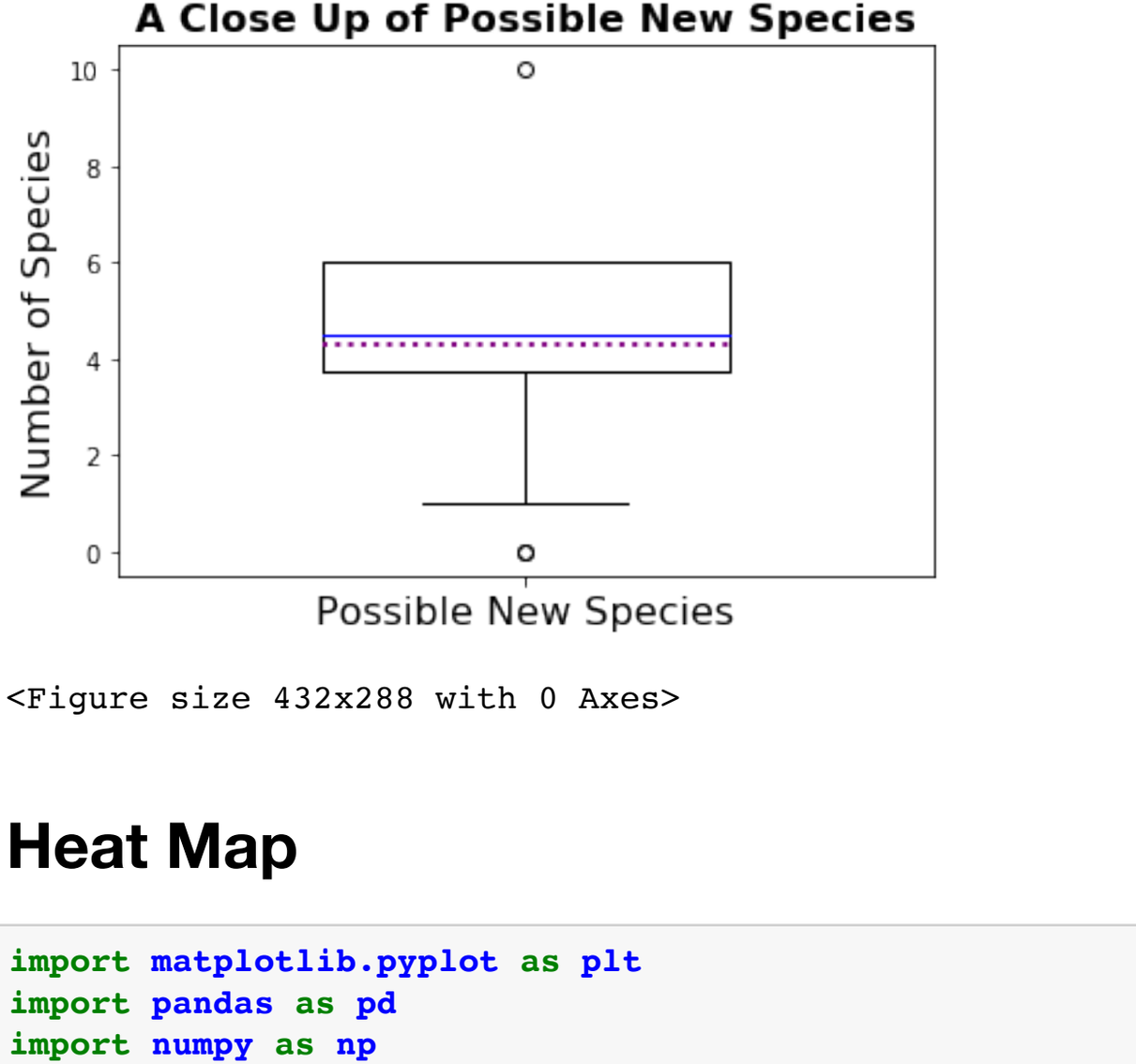


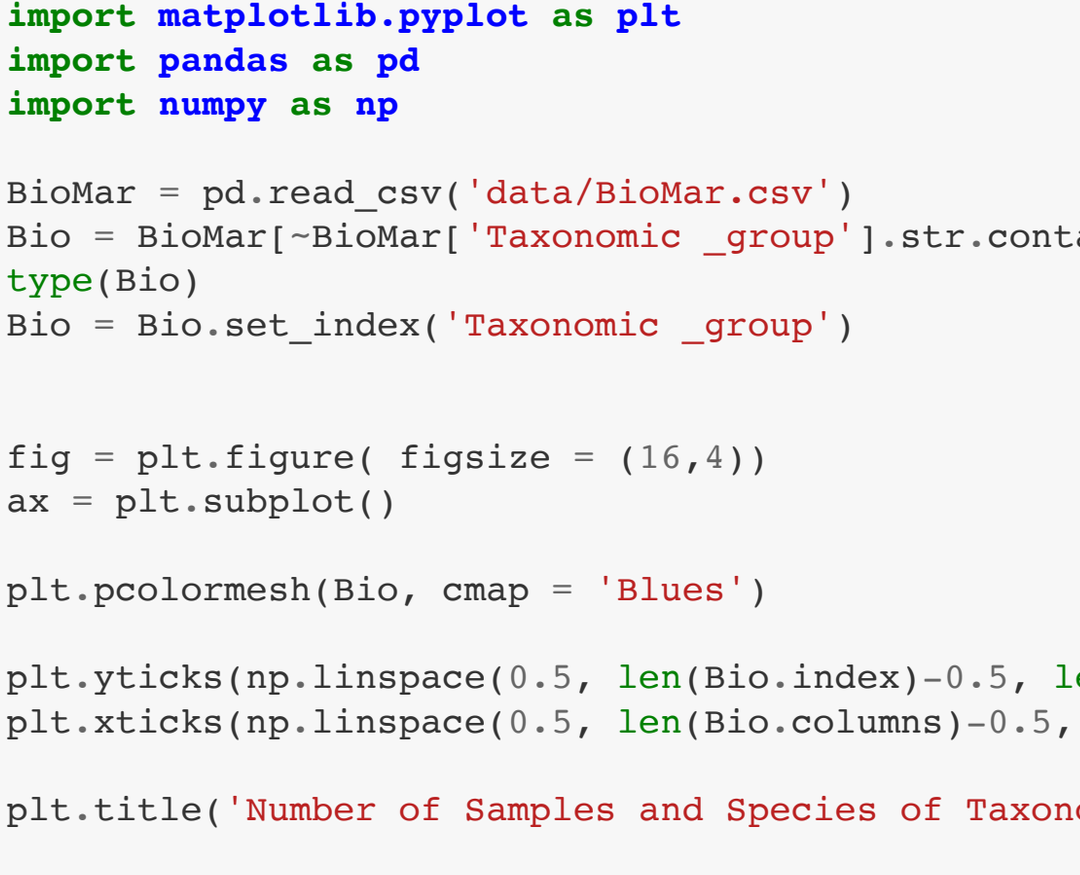

```
<Figure size 432x288 with 0 Axes>
```

Box Plot With Mulitgraph Layout

In [109]:

```python
BioMar = pd.read_csv('data/BioMar.csv')
Bio = BioMar[~BioMar['Taxonomic _group'].str.contains("_") | BioMar['Taxonomic _group'].str.contains("TOTAL")]


data = [Bio['Biomar_species'], Bio['Poss_New_species']]
xticks_lab = ['BioMar Species', 'Possible New Species']

median_line = {'color':'blue'}
mean_line = {'color':'purple',
             'linestyle':'dotted',
             'linewidth': 2}

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



plt.xticks([1,2], xticks_lab, fontsize = 16)


plt.ylabel('Number of Species', fontsize = 16)
plt.title('A Comparison of Possible New Species and Known Species', fontsize = 16, fontweight = 'bold')
plt.show()


data = [Bio['Poss_New_species']]
xticks_lab = ['Possible New Species']

median_line = {'color':'blue'}
mean_line = {'color':'purple',
             'linestyle':'dotted',
             'linewidth': 2}

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



plt.xticks([1,1], xticks_lab, fontsize = 16)

plt.ylabel('Number of Species', fontsize = 16)
plt.title('A Close Up of Possible New Species', fontsize = 16, fontweight = 'bold')
plt.show()

plt.savefig('Ohlweiler_Fig_5.pdf')
```



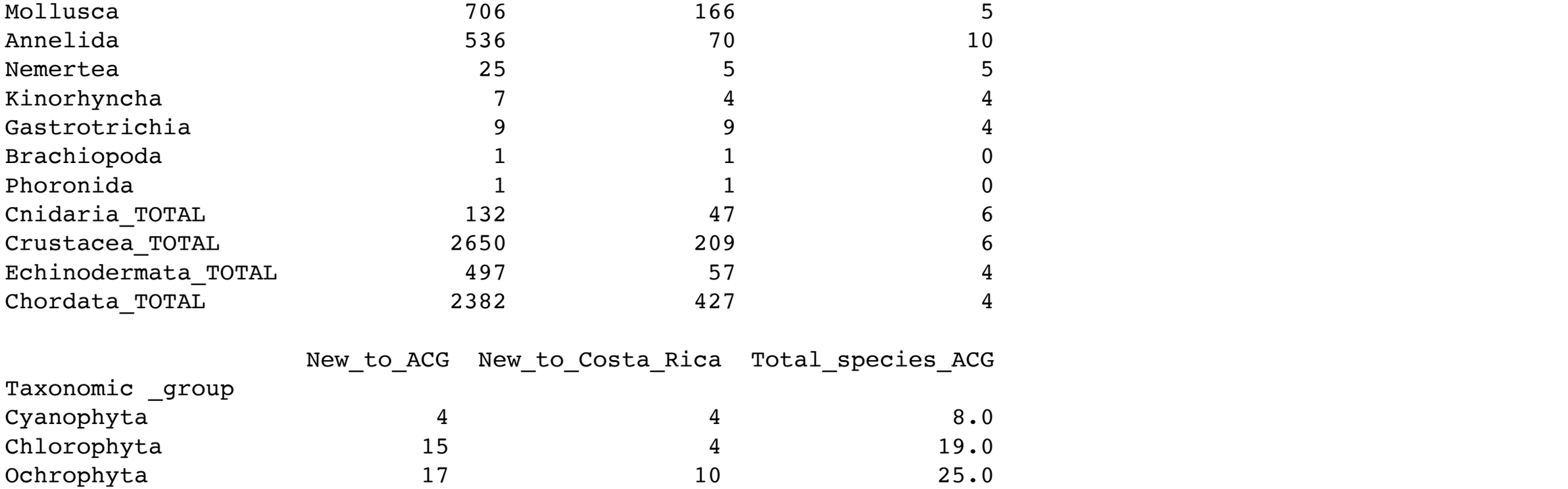



```
<Figure size 432x288 with 0 Axes>
```

Heat Map

In [113]:

```python
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

BioMar = pd.read_csv('data/BioMar.csv')
Bio = BioMar[~BioMar['Taxonomic _group'].str.contains("_") | BioMar['Taxonomic _group'].str.contains("TOTAL")]
type(Bio)
Bio = Bio.set_index('Taxonomic _group')


fig = plt.figure( figsize = (16,4))
ax = plt.subplot()

plt.pcolormesh(Bio, cmap = 'Blues')

plt.yticks(np.linspace(0.5, len(Bio.index)-0.5, len(Bio.index)), Bio.index)
plt.xticks(np.linspace(0.5, len(Bio.columns)-0.5, len(Bio.columns)), Bio.columns)

plt.title('Number of Samples and Species of Taxonomic Groups Across Several Categories')

plt.colorbar()

plt.show()

print('\n',Bio)

plt.savefig('Ohlweiler_Fig_6.pdf')
```




```
                     Biomar_samples  Biomar_species  Poss_New_species  \
Taxonomic _group
Cyanophyta                       11               5                 3
Chlorophyta                      45              19                 1
Ochrophyta                       37              23                 6
Rhodophyta                      137              62                 5
Porifera                        167              18                 6
Mollusca                        706             166                 5
Annelida                        536              70                10
Nemertea                         25               5                 5
Kinorhyncha                       7               4                 4
Gastrotrichia                     9               9                 4
Brachiopoda                       1               1                 0
Phoronida                         1               1                 0
Cnidaria_TOTAL                  132              47                 6
Crustacea_TOTAL                2650             209                 6
Echinodermata_TOTAL             497              57                 4
Chordata_TOTAL                 2382             427                 4

                     New_to_ACG  New_to_Costa_Rica  Total_species_ACG
Taxonomic _group
Cyanophyta                    4                  4                8.0
Chlorophyta                  15                  4               19.0
Ochrophyta                   17                 10               25.0
Rhodophyta                   59                  9               74.0
Porifera                     18                 18               18.0
Mollusca                    137                  8              324.0
Annelida                     46                 10               73.0
Nemertea                      5                  5                6.0
Kinorhyncha                   4                  4                4.0
Gastrotrichia                 9                  9                9.0
Brachiopoda                   1                  0                1.0
Phoronida                     1                  0                1.0
Cnidaria_TOTAL                8                  6               53.0
Crustacea_TOTAL              99                  9              292.0
Echinodermata_TOTAL          40                  4               60.0
Chordata_TOTAL              416                 18              449.0

<Figure size 432x288 with 0 Axes>
```

Jefferson
HOME OF SIDNEY KIMMEL MEDICAL COLLEGE