Exploratory Data Analysis on the penguins_lter dataset

Report

Introduction

The purpose of this document is to summarise and communicate the insights found in the penguin Lter dataset. According to archive.ics.uci.edu, the dataset is a result of a study of antarctic penguins' behaviour and interaction with their environment. The following columns were contained in the Penguins Lter dataset:

1. Study Name - The name of three different datasets. These datasets are a result of studies done at different times

2. Sample Number - The number assigned to each penguin

3. Species - The species of penguin. These can be bucketed into the Adelie Penguin (Pygoscelis adeliae), Chinstrap penguin (Pygoscelis antarctica) or Gentoo penguin (Pygoscelis papua)

4. Region - This column is 'Anvers' for all the data, indicating where the research was done

5. Island - The Island on which the penguin is found

6. Stage - This column is 'Adult, 1 Egg Stage' for all penguins

7. Individual ID - ID given to each penguin in the study

8. Clutch Completion - According to seaworld.org, this is an indicator for whether the expected number of eggs has been laid

9. Date Egg - Indicates the date at which the nest was observed

10. Culmen Length (mm) - The length of the upper ridge of a bird's beak

11. Culmen Depth (mm) - The vertical measurement of the upper ridge of a penguins beak from the base to the tip

12. Flipper Length (mm) - The length of the penguins' flipper

13. Body Mass (g) - Weight of the penguin in grams

14. Sex - The gender of the penguin

15. Delta 15 N (o/oo) - According to https://pubmed.ncbi.nlm.nih.gov/, delta 15N (o/oo) is a measure of the ratio of stable nitrogen isotopes expressed as parts per thousands. This is used to understand a penguin's diet and trophic level. Higher stable nitrogen isotopes indicates that a penguin is eating prey from a higher position in the food chain suggesting a diet richer in animal protien

16. Delta 13 C (o/oo) - Stable isotope values of carbon provide information of foraging patterns and migratory behaviour in birds

17. Comments - Addiitonal information on penguins in the study where applicable

```
In [34]: import pandas as pd
         import numpy as np
         from scipy import stats
         import seaborn as sns
         import matplotlib.pyplot as plt
         import missingno
         from pandas.plotting import scatter_matrix
```

Data Cleaning

```
In [35]: # read in penguin data
         df = pd.read_csv('penguins_lter.csv')
         df['Count'] = 1
```

The following was observed about the dataset:

The data contains 344 rows and 18 columns. The 'Region' and 'Stage' is the same for all rows and can be removed.

The 'Sample Number' and 'Individual ID' do not provide additional information, and can also be removed.

The 'Date Egg' column has inconsistent date formats and has been cleaned to produce consistent datetimes across all data entries.

```
In [36]: # Removed columns that do not add additional information
         df = df.drop(['Stage', 'Region', 'Sample Number', 'Individual ID'], axis=1)

         # Removed the word penguin and the scientific name from the 'Specie' column
         df['Species'] = df['Species'].str.split(' P').str[0].str.strip()
         df['Species'] = df['Species'].str.split(' p').str[0].str.strip()

         # Changed 'Date Egg' column so that it starts with 20
         df['Date Egg'] = df['Date Egg'].apply(lambda x: '20'+ str(x)
                                     if str(x[0:2]) != 20 else x)

         # Add a padded zero at the start of the month
         df['Date Egg'] = df['Date Egg'].apply(lambda x: str(x[0:5]) + '0'
                                     + str(x[5]) + str(x[6:])
                                     if x[6] == "/" else x)

         # Convert values in the 'Egg Date' column from a string to the datetime format
```

```
df['Date Egg'] = pd.to_datetime(df['Date Egg'],
                                format='%Y/%d/%m', errors="coerce")
```
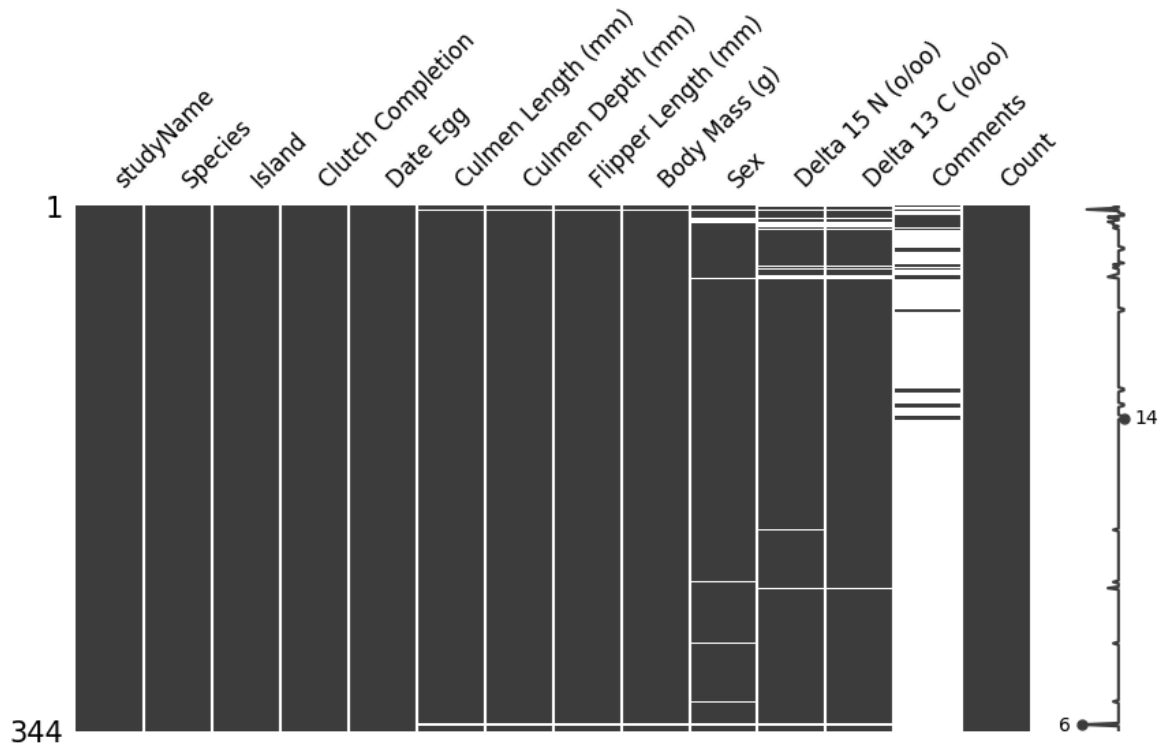
Missing Data

As stated above, the dataset consists of 344 rows and 18 columns. After removing the columns that did not provide any additional information, the 'Date Egg' column was cleaned. I then carried out the following to fill in the missing data:

1. All the columns in the data was checked for missing entries. It was noted that the 'Comments' column should not have an entry for every column

2. For each column, the counts were used to identify missing data. The mean, 25th percentile, 75th percentile minimum and maximum were checked to find any unreasonable values and data errors

3. The missing data in each column was counted and the percentage of missing values (out of the total number of data entries) was calculated to see if the percentage of missing data is small enough to be removed

4. The missing data in the Culmen Length, Culmen Depth, Flipper length and Body Mass was filled by calculating the average of these numbers for the penguins of the same species, study, island, Date Egg and Sex

5. The same methodology (as point 4) was used to fill the missing data in the Delta 15N (o\oo) and Delta 13N (o\oo) columns

6. There was no reliable way to fill the missing values in the 'Sex' column. The rows was therefore removed

The following is a graph of the missing data. The graph indicates that the column with the most missing data is the 'Comments' column. This is expected since a comment is not required for each row in the dataset. The pattern of missing data in the Delta 15N (o/oo) column closely matches the pattern of missing data in the Delta 13C (o/oo) column. There are also a number of missing entries in the 'Sex' column. The distribution of missing data in the 'Culmen Length', 'Culmen Depth', 'Flipper length' and 'Body Mass' seem is very similar indicating that the missing information may result of the same rows of data.

In [5]:
```
# look at graph of missing data
missingno.matrix(df,figsize=(10,5), fontsize=12)
```

Out[5]:   <Axes: >

Below is a summary of the data. The count of the 'Culmen Length (mm)', 'Culmen Depth (mm)', 'Flipper Length (mm)', 'Body Mass (g)', 'Delta 15 N (o/oo)' and 'Delta 13 C (o/oo)' has less then 344 rows. This confirms that there are missing data in these columns. The average culmen length and culmen depth is 43.92mm and 17.15mm respectively. The average flipper length is 200.91mm and the average Body Mass is 4201 grams. The average 'Delta 15N (o/oo)' and 'Delta 13 C (o/oo)' is 8.733 and -25.68. The standard deviation in the body mass of penguins of 801.95.

```
In [6]:  # print out basic statistics of dataset
         pd.set_option('display.float_format', lambda x: '%.0f' % x)
         df.describe()
```

Out[6]:

	Date Egg	Culmen Length (mm)	Culmen Depth (mm)	Flipper Length (mm)	Body Mass (g)	Delta 15 N (o/oo)	Delta 13 C (o/oo)	Count
count	344	342	342	342	342	330	331	344
mean	2011-08-31 10:36:16.744186112	44	17	201	4202	9	-26	1
min	2011-07-09 00:00:00	32	13	172	2700	8	-27	1
25%	2011-07-28 00:00:00	39	16	190	3550	8	-26	1
50%	2011-08-10 12:00:00	44	17	197	4050	9	-26	1
75%	2011-09-17 00:00:00	48	19	213	4750	9	-25	1
max	2012-09-01 00:00:00	60	22	231	6300	10	-24	1
std	NaN	5	2	14	802	1	1	0

Below is the count of the missing data in each column. As seen from the graph above, 'Culmen Length', 'Culmen depth', 'Flipper Length' and 'Body Mass' are all missing two data points. 'Sex', 'Delta 15N (o/oo)' and 'Delta 13C (o/oo)' are also missing 10, 14 and 13 datapoints respectively.

In [7]:
```python
# print number of missing cells per column
missing_values_count = df.isnull().sum()
missing_values_count[0:18]
```

Out[7]:
```
studyName              0
Species                0
Island                 0
Clutch Completion      0
Date Egg               0
Culmen Length (mm)     2
Culmen Depth (mm)      2
Flipper Length (mm)    2
Body Mass (g)          2
Sex                   10
Delta 15 N (o/oo)     14
Delta 13 C (o/oo)     13
Comments             318
Count                  0
dtype: int64
```

In [8]:
```python
# Total number of missing values
total_cells = np.prod(df.shape)
total_missing = missing_values_count.sum()

# Percent of data that is missing
(total_missing/total_cells) * 100
```

Out[8]:   np.float64(7.537375415282392)

The percentage of missing values (out of the total number of cells) is 7.54 percent. This percentage is too large to simply remove the rows with the missing entries and a reasonable method is needed to fill in the missing data. The missing values for Culmen Length (mm), Culmen depth (mm), Flipper Length and Body Mass were filled with the mean value for penguins with the same 'studyName', 'Species', 'Clutch Completion', 'Date Egg' and 'Sex'.

The remaining rows with missing data have been removed since the size of the incomplete dataset is insignificant.

The data now consists of 329 rows and 14 columns

In [26]:
```python
# fill in the Culmen Length, Culmen Depth and Flipper length columns
# using the average for each study name, penguin species, clutch completion and
# egg observation date
df['Culmen Length (mm)'] = df.groupby(
    ['studyName', 'Species', 'Clutch Completion', 'Date Egg', 'Sex']
    )['Culmen Length (mm)'].transform(lambda x: x.fillna(x.mean()))
df['Culmen Depth (mm)'] = df.groupby(
    ['studyName', 'Species', 'Clutch Completion', 'Date Egg', 'Sex']
    )['Culmen Depth (mm)'].transform(lambda x: x.fillna(x.mean()))
df['Flipper Length (mm)'] = df.groupby(
    ['studyName', 'Species', 'Clutch Completion', 'Date Egg', 'Sex']
    )['Flipper Length (mm)'].transform(lambda x: x.fillna(x.mean()))
df['Body Mass (g)'] = df.groupby(
    ['studyName', 'Species', 'Clutch Completion', 'Date Egg', 'Sex']
    )['Body Mass (g)'].transform(lambda x: x.fillna(x.mean()))

# Fill in the Delta 15 N (o/oo) and Delta 13 C (o/oo)
# Using the average for each study name, penguin species, clutch completion,
# egg observation date and gender
df['Delta 15 N (o/oo)'] = df.groupby(
    ['studyName', 'Species', 'Clutch Completion', 'Date Egg', 'Sex']
    )['Delta 15 N (o/oo)'].transform(lambda x: x.fillna(x.mean()))
df['Delta 13 C (o/oo)'] = df.groupby(
    ['studyName', 'Species', 'Clutch Completion', 'Date Egg', 'Sex']
    )['Delta 13 C (o/oo)'].transform(lambda x: x.fillna(x.mean()))

# Removed the observed penguins where the gender is missing
df.dropna(subset=['Sex'], inplace=True)
df.dropna(subset=['Delta 15 N (o/oo)'], inplace=True)
df.drop(df.loc[df['Sex']=='.'].index, inplace=True)
```
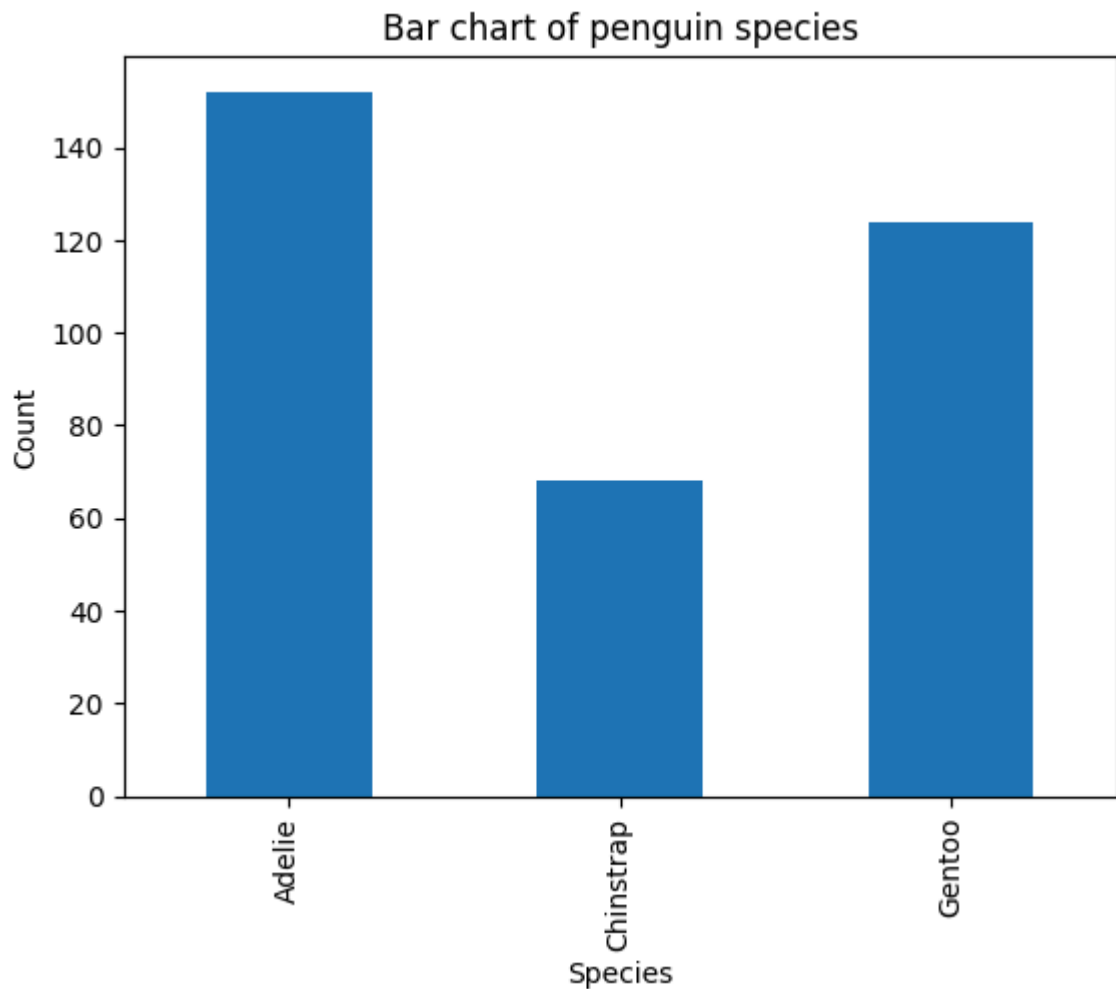
Data Stories and visualisations

The pie chart of the proportion of penguins per species below shows that Adelie penguins are the largest population of penguin species (43.2%), followed by Gentoo (36.2%) and then Chinstrap penguins (20.7%)
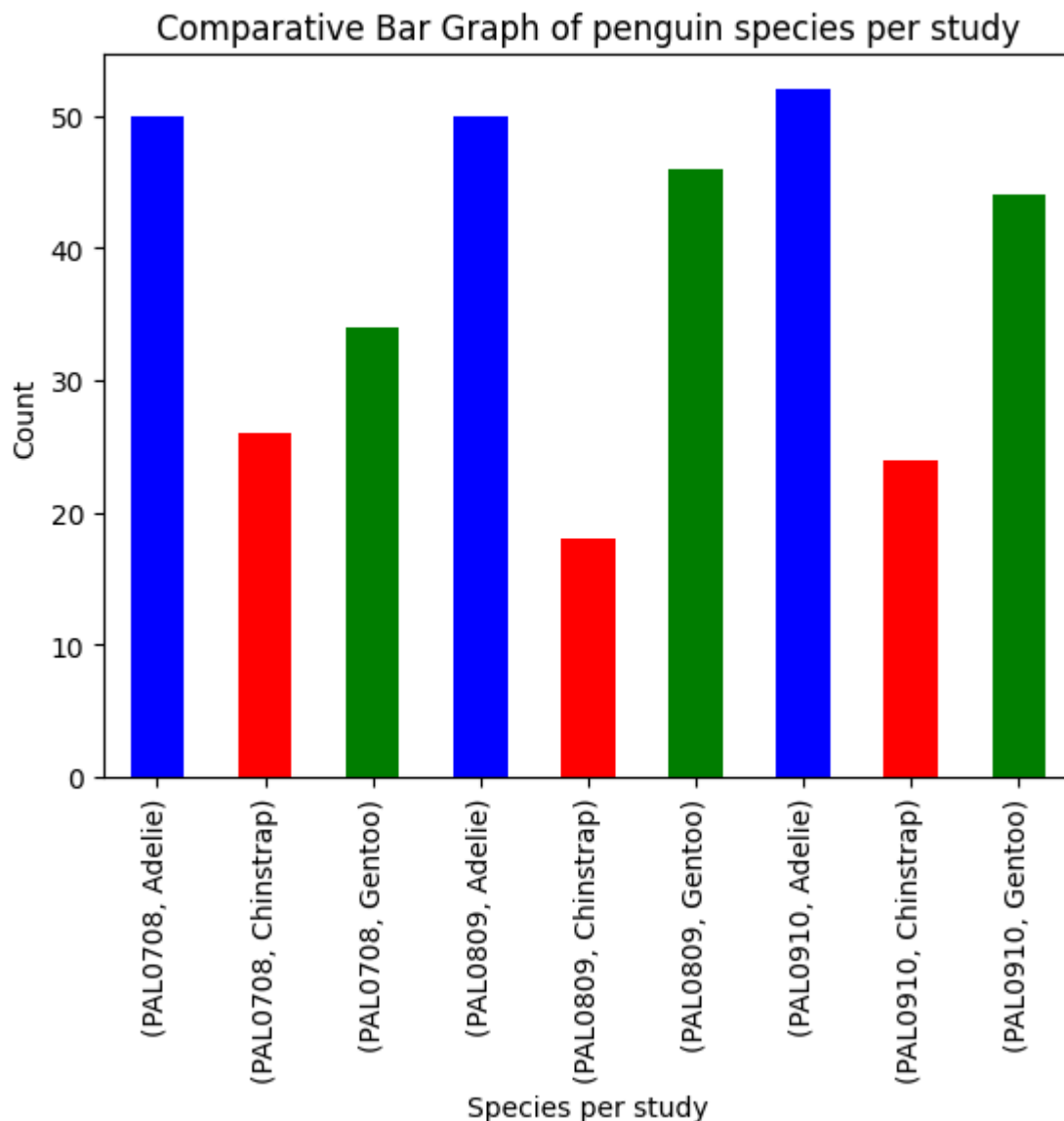
In [37]:
```python
# group data by species and plot on a bar graph
df_species = df.groupby(['Species'], as_index=True)['Count'].sum()
ax = df_species.plot.bar(title='Comparative Bar Graph of count ' \
```

```
'of penguin species', rot = 90, xlabel='Species', ylabel='Count')

plt.title('Bar chart of penguin species')
plt.show()
```
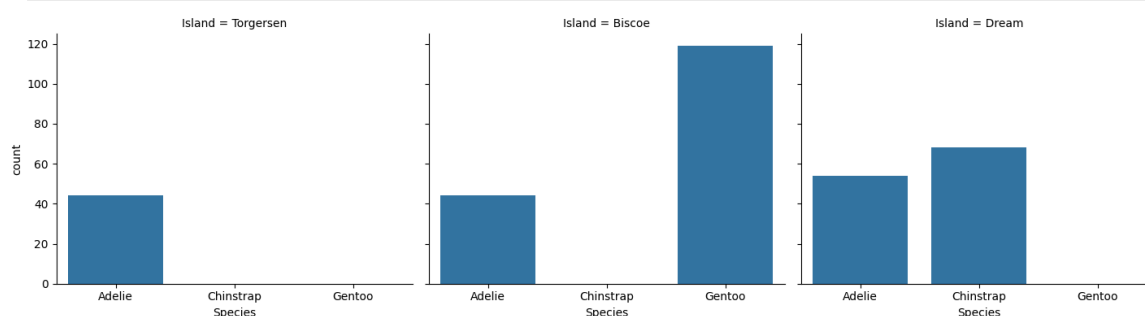

The same pattern can be observed when splitting the data by study. Adelie penguins form the largest portion of penguins in each study, followed by Gentoo, then Chinstrap penguins.

```
In [38]:   # group data by study name and species
           # and plot a bar graph of the count of penguins

           df_species = df.groupby(['studyName', 'Species'])['Count'].sum()
           ax = df_species.plot.bar(color=["Blue", "Red", "Green"],
                                     title='Comparative Bar Graph of ' \
                                     'penguin species per study',
                                     rot = 90, xlabel='Species per study', ylabel='Count')
           ax.set_xlabel = "Species per study"
           plt.show()
```

Comparative Bar Graph of penguin species per study

The data has been split into species and island below. Adelie penguins can be found on all three islands Biscoe, Dream and Torgersen. The Torgensten island only has Adelie penguins. Chinstrap penguins are only found on the Dream island, while Gentoo penguins can only be found on the Biscoe island.

```
In [13]:  # plot bar graph of count of species per island
          sns.catplot(data=df, col='Island', x='Species',
                      kind='count', height=4, aspect=1.2)
          plt.show()
```
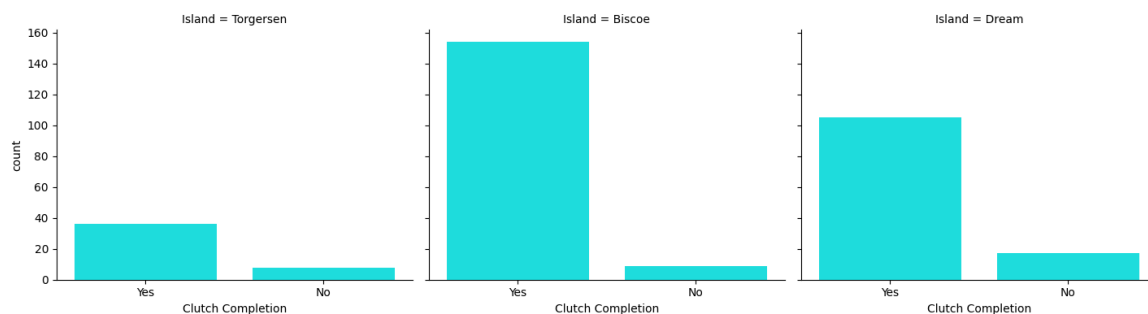

The penguin data was split by Island and Clutch Completion. According to seaworld.org, clutch completion indicates whether the nest contains an expected number of two eggs.

Clutch completion is important because it increases the probability that at least one egg will hatch and survive to adulthood. Clutch completion was achieved for the majority of nests in each island. For Boscoe, clutch completion was completed for 94.5% of nests. For the Dream Island, 86.1% of nests achieved clutch completion and 81.8% of nests on the Torgenston island achieved clutch completion.

In [29]:
```python
# plot bar graph of count of nests achieving clutch completion
# on each island
sns.catplot(data=df, col='Island', x='Clutch Completion',
            kind='count', height=4, aspect=1.2, color="cyan")
```
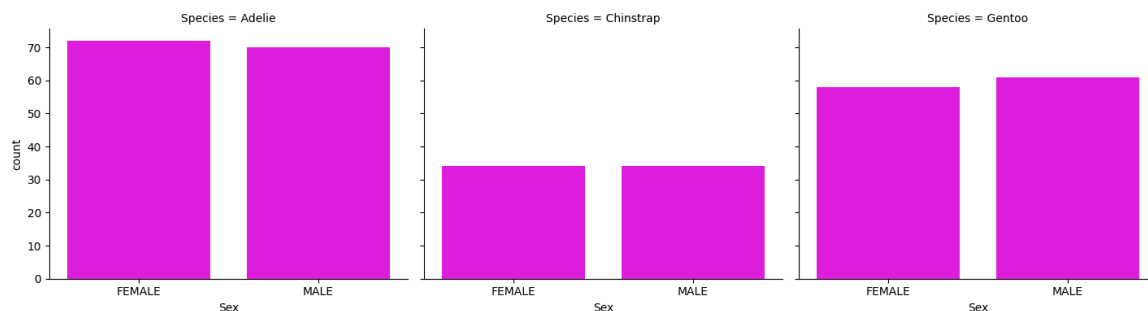
Out[29]:   <seaborn.axisgrid.FacetGrid at 0x1b97e9bb940>

The bar charts below show that there is an approximately equal number of males and female penguins in the study for all penguin species.

In [27]:
```python
# plot a bar graph of count of males and females per species
sns.catplot(data=df, col='Species', x='Sex',
            kind='count', height=4, aspect=1.2, color="magenta")
```

Out[27]:   <seaborn.axisgrid.FacetGrid at 0x1b97d898e80>

The graphs below show the average culmen length, culmen depth, flipper length and body mass per penguin species. Chinstrap penguins have the longest culmen length, followed by Gentoo, then Adelie penguins On average, Chinstrap and Adelie penguins have approximately the same culmen depth while the culmen depth of Gentoo penguins are shorter. Flipper length, on average are longest for the Gentoo penguins, then Chinstrap, followed by the Adelie penguins. Gentoo penguins have a higher average body mass. On average, body mass of Chinstrap and Adelie penguins are similar.

In [16]:
```python
# plot average culmen length, average culmen depth, average
# flipper length and average body mass per species

# specify number of graphs to plot
fig, axes = plt.subplots(nrows=1, ncols=4, figsize=(15, 5))
```

```python
# plot average culmen length per species
df_culmen_len = df.groupby(['Species'], as_index=False)[
    'Culmen Length (mm)'].mean()
axes[0].bar(df_culmen_len['Species'], df_culmen_len[
    'Culmen Length (mm)'] ,color=["Blue", "Red", "Green"])
axes[0].set_title('Average Culmen length per Species')

# plot average culmen depth per species
df_culmen_depth = df.groupby(['Species'], as_index=False)[
    'Culmen Depth (mm)'].mean()
axes[1].bar(df_culmen_depth['Species'], df_culmen_depth[
    'Culmen Depth (mm)'] ,color=["Blue", "Red", "Green"])
axes[1].set_title('Average Culmen Depth (mm) per Species')

# plot average flipper length per species
df_flipper_len = df.groupby(['Species'], as_index=False)[
    'Flipper Length (mm)'].mean()
axes[2].bar(df_flipper_len['Species'], df_flipper_len[
    'Flipper Length (mm)'] ,color=["Blue", "Red", "Green"])
axes[2].set_title('Average Flipper Length (mm) per Species')

# plot average body mass per species
df_Body_Mass = df.groupby(['Species'], as_index=False)[
    'Body Mass (g)'].mean()
axes[3].bar(df_Body_Mass['Species'], df_Body_Mass[
    'Body Mass (g)'] ,color=["Blue", "Red", "Green"])
axes[3].set_title('Average body mass per Species')

# display bar charts
fig.supylabel('Count')
plt.tight_layout()
plt.show()
```
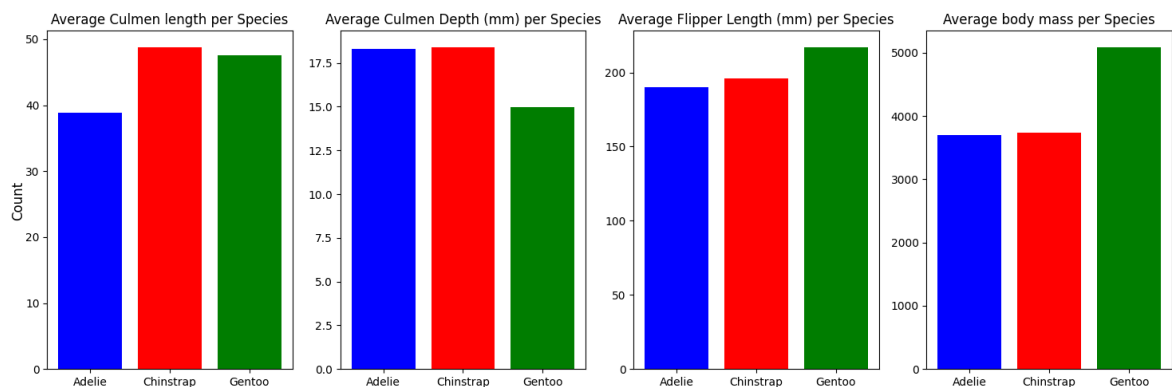


```python
In [18]:   # create dataframes for each species
           df_adelie = df[df['Species']=='Adelie']
           df_gentoo = df[df['Species']=='Gentoo']
           df_chinstrap = df[df['Species']=='Chinstrap']
```

The bar charts below show the distribution of culmen length for Adelie, Gentoo and Chinstrap penguins. The culmen length for Adelie penguins follows a normal distributioin. The culmen length for the Chinstrap penguin follows a normal distribution with two peaks. This indicates that there two distinct groups of penguins in the data. Further investigation will be needed to find the reason for the two peaks. The distribution of culmen length of the Gentoo penguin also follows a normal distribution with two

peaks. These peaks are however, less promounced compared to the distribution of culmen length of the Chinstrap penguin.

```
In [ ]:   # plot distribution of culmen length per penguin species

          # specify the number and size of histograms
          fig, axes = plt.subplots(1, 3, figsize=(15, 5))

          # plot histogram of culmen length for Adelie penguins
          sns.histplot(df_adelie['Culmen Length (mm)'],
                       kde=True, ax=axes[0], color='green')
          axes[0].set_title('Distribution of Adelie penguin culmen length')

          # plot histogram of culmen length for Chinstrap penguins
          sns.histplot(df_chinstrap['Culmen Length (mm)'],
                       kde=True, ax=axes[2], color='blue')
          axes[2].set_title('Distribution of Chinstrap penguin culmen length')

          # plot histogram of culmen length for Gentoo penguins
          sns.histplot(df_gentoo['Culmen Length (mm)'],
                       kde=True, ax=axes[1], color='red')
          axes[1].set_title('Distribution of Gentoo penguin culmen length')

          # display histograms
          plt.tight_layout()
          plt.show()
```
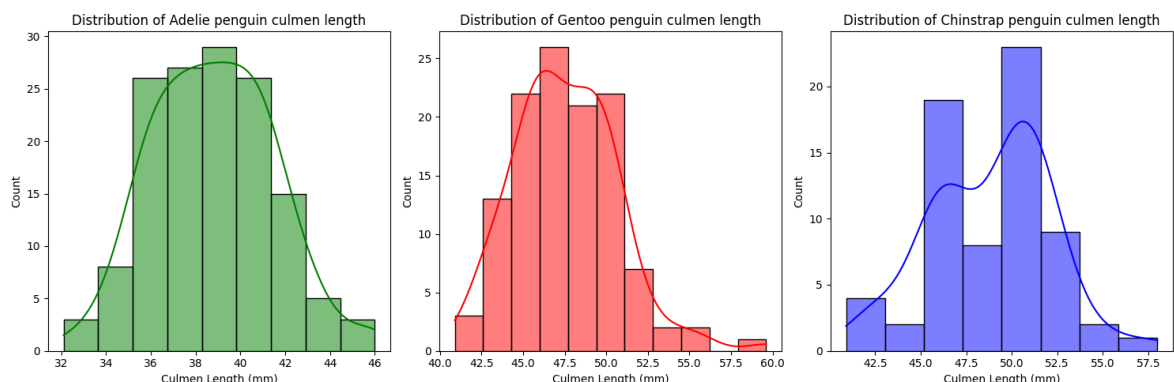

Below are graphs of the distribution of the culmen depth of the Adelie, Gentoo and Chinstrap penguins. The distributions resemble normal distributions.

```
In [39]:  # plot distribution of culmen depth per penguin species

          # specify the number and size of histograms
          fig, axes = plt.subplots(1, 3, figsize=(15, 5))

          # plot histogram of culmen depth for Adelie penguins
          sns.histplot(df_adelie['Culmen Depth (mm)'],
                       kde=True, ax=axes[0], color='green')
          axes[0].set_title('Distribution of Adelie penguin culmen depth')

          # plot histogram of culmen depth for chinstrap penguins
          sns.histplot(df_chinstrap['Culmen Depth (mm)'],
                       kde=True, ax=axes[2], color='blue')
          axes[2].set_title('Distribution of Chinstrap penguin culmen depth')

          # plot histogram of culmen depth for gentoo penguins
```
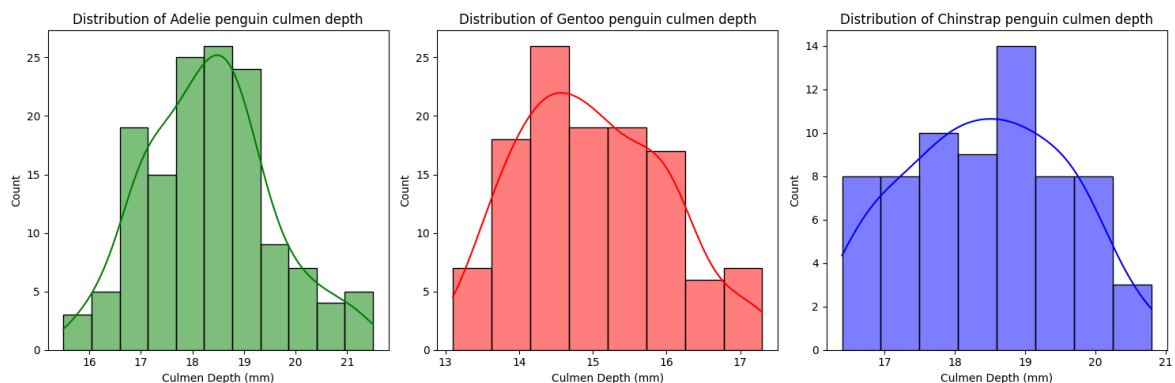
```python
sns.histplot(df_gentoo['Culmen Depth (mm)'],
             kde=True, ax=axes[1], color='red')
axes[1].set_title('Distribution of Gentoo penguin culmen depth')

# display histograms
plt.tight_layout()
plt.show()
```


The distribution of flipper length per penguin species is shown in the graph below. Adelie and Chinstrap penguins' flipper length follow a normal distribution. For Gentoo penguins, the majority of flipper lengths are between 210 and 220mm.

In [40]:
```python
# plot distribution of culmen depth per penguin species

# specify the number and size of histograms
fig, axes = plt.subplots(1, 3, figsize=(15, 5))

# plot histogram of flipper length for Adelie penguins
sns.histplot(df_adelie['Flipper Length (mm)'],
             kde=True, ax=axes[0], color='green')
axes[0].set_title('Distribution of Adelie penguin flipper length')

# plot histogram of flipper length for Gentoo penguins
sns.histplot(df_gentoo['Flipper Length (mm)'],
             kde=True, ax=axes[1], color='red')
axes[1].set_title('Distribution of Gentoo penguin flipper length')

# plot histogram of flipper length for Chinstrap penguins
sns.histplot(df_chinstrap['Flipper Length (mm)'],
             kde=True, ax=axes[2], color='blue')
axes[2].set_title('Distribution of Chinstrap penguin flipper length')

# display histograms
plt.tight_layout()
plt.show()
```
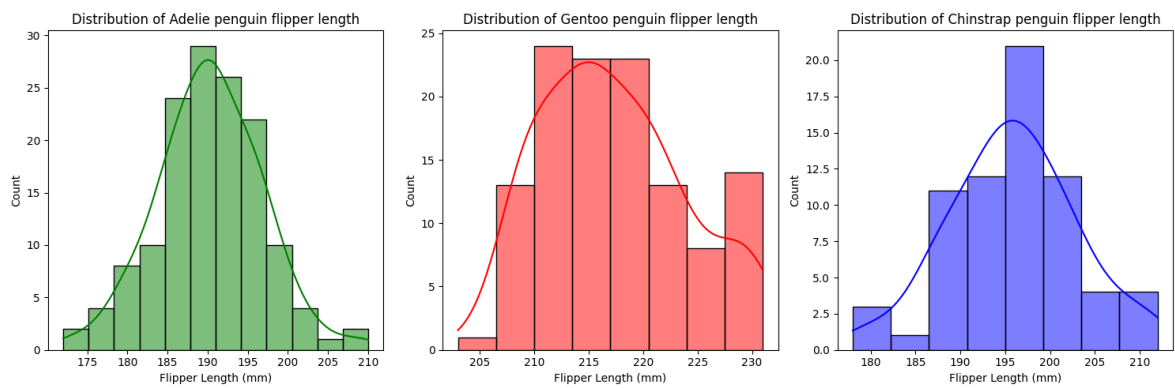
The distribution of the body mass per species is shown below.

The histograms below show that the body mass of the Chinstrap penguins follows a normal distribution. There may be two distinct groups within the Gentoo penguin population due to the two distinct peaks. The body mass of the Adelie penguin is normally distributed with a slight skewness to the right, indicating that Adelie penguin body mass is slightly smaller.

In [41]:
```python
# plot distribution of body mass per penguin species

# specify the number and size of histograms
fig, axes = plt.subplots(1, 3, figsize=(15, 5))

# plot histogram of body mass for Adelie penguin
sns.histplot(df_adelie['Body Mass (g)'],
             kde=True, ax=axes[0], color='green')
axes[0].set_title('Distribution of Adelie penguin body mass')

# plot histogram of body mass for Gentoo penguin
sns.histplot(df_gentoo['Body Mass (g)'],
             kde=True, ax=axes[1], color='red')
axes[1].set_title('Distribution of Gentoo penguin body mass')

# plot histogram of body mass for Chinstrap penguin
sns.histplot(df_chinstrap['Body Mass (g)'],
             kde=True, ax=axes[2], color='blue')
axes[2].set_title('Distribution of Chinstrap penguin body mass')

plt.tight_layout()
plt.show()
```
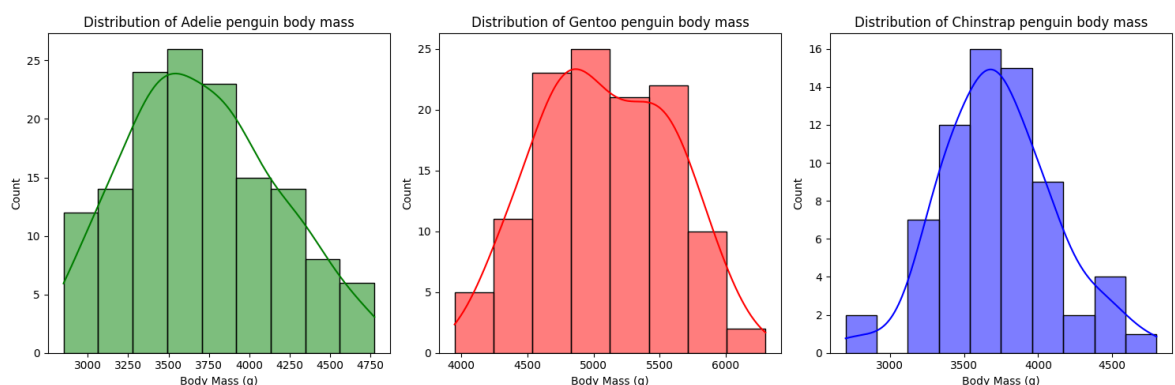
Below is a scatterplot matrix that shows the distribution of culmen length, culmen depth, flipper length, body mass, delta 15N (o/oo) and delta 13C (o/oo) per species

The scatterplot of flipper length and body mass shows a positive correlation between these two variables. This makes sense since larger flippers will be needed to support a larger body regardless of penguin species.

There is also positive correlation between body mass and culmen depth as well as culmen length. Again, this makes sense since a larger penguin should have a bigger beak.
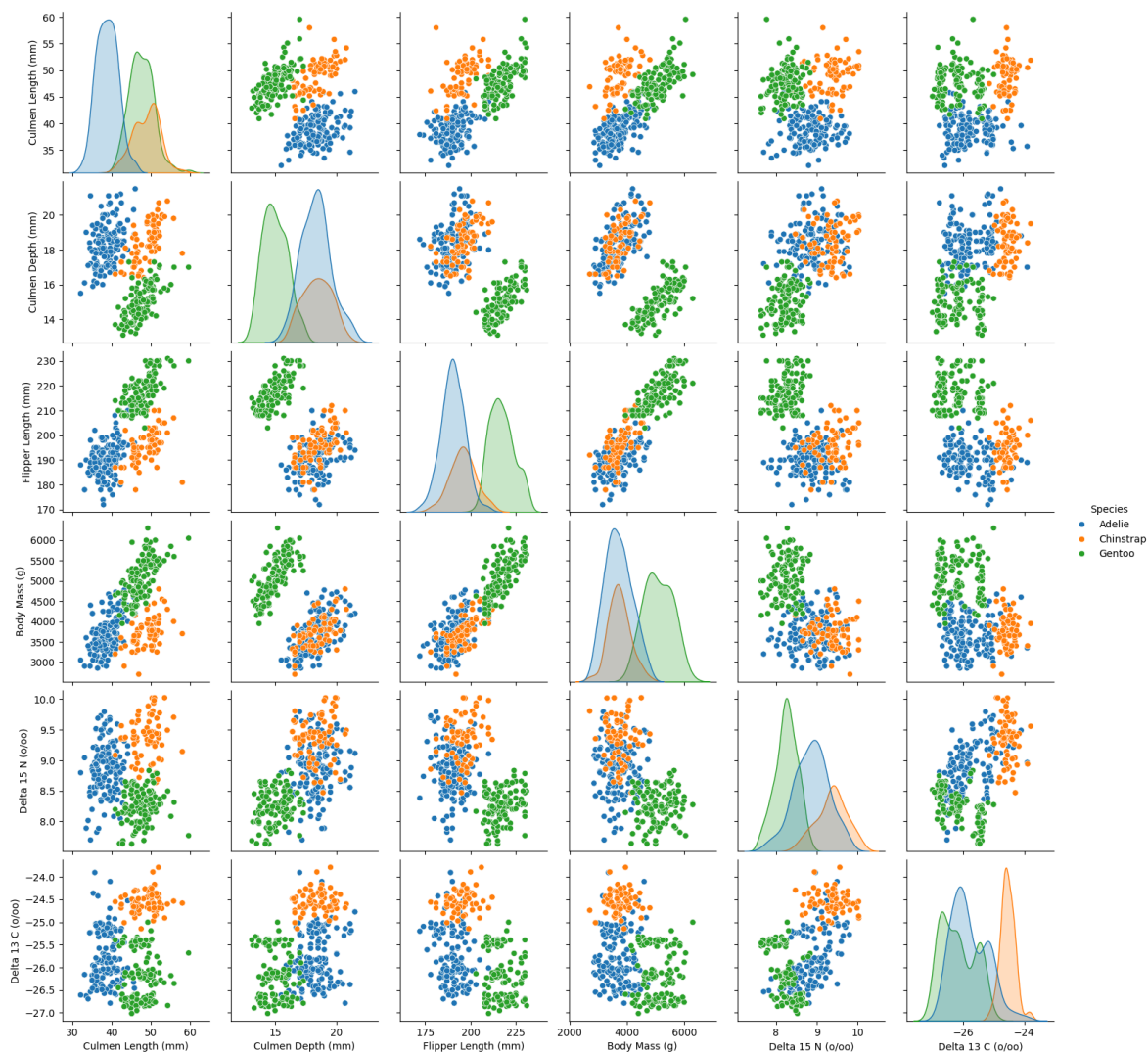
Both the scatterplot matrix below and the average graphs above show that Gentoo penguins have longer flippers and smaller culmen depth compared to Gentoo and Adelie penguins.

According to pubmed.ncbi.nim.nih.gov, The delta 15N (o/oo) and delta 13C (o/oo) is an indication of a penguins diet. A high delta 15N (o/oo) and delta 13C (o/oo) indicates that penguins consume prey higher up on the food chain. Gentoo penguins have lower delta 15 (o/oo) and delta 13C (o/oo) levels. They feed on animals that are fairly low down in the food chain such as krill and squid even though they have a longer flipper length and higher body mass compared to Chinstrap and Adelie penguins.

The correlations will be discussed in further detail in paragraph below.

```python
# plot a scatterplot matrix of culmen length,
# culmen depth, flipper length, body mass
# Delta 15N (o/oo) and Delta 13C (o/oo)
scatter_df=df.drop(['Count'], axis=1)
sns.pairplot(scatter_df, hue="Species")
```

In [24]:

Out[24]:   <seaborn.axisgrid.PairGrid at 0x1b97cb16470>

Below is the heatmap and the correlations between the culmen length, culmen depth, flipper length, body mass, delta 15N (o/oo) and delta 13C (o/oo) per penguin species.

There is a strong correlation between body mass and flipper length. The correlation is strongest for Gentoo penguins (0.71) followed by chinstrap penguins (0.64) then Adelie penguins (0.49).

There is a positive correlation between the culmen length and culmen depth. The correlation is highest for Gentoo and Chinstrap penguins (0.65) followed by Adelie penguins (0.4). This makes sense since a penguin with a longer than average beak should be thicker than average.

Interestingly, the corrolation between Delta 15N (o/oo) and the culmen depth, culmen length, flipper length and body mass is negligible. This indicates that there is no relationship between the penguins diet and the size and length of a penguins beak, flippers and body mass. The same applies to the Delta 13C (o/oo).

```
In [30]:  # create dataframes for correlation calcs
          corrolation_adelie_df = df_adelie[
              ['Flipper Length (mm)', 'Body Mass (g)',
               'Culmen Length (mm)', 'Culmen Depth (mm)',
               'Delta 15 N (o/oo)', 'Delta 13 C (o/oo)']]

          corrolation_gentoo_df = df_gentoo[
```

```python
    ['Flipper Length (mm)', 'Body Mass (g)',
     'Culmen Length (mm)', 'Culmen Depth (mm)',
     'Delta 15 N (o/oo)', 'Delta 13 C (o/oo)']]
corrolation_chinstrap_df = df_chinstrap[
    ['Flipper Length (mm)', 'Body Mass (g)',
     'Culmen Length (mm)', 'Culmen Depth (mm)',
     'Delta 15 N (o/oo)', 'Delta 13 C (o/oo)']]

# Specify number and size of subplots
fig = plt.figure(figsize = (15,15))
adelie_ax = fig.add_subplot(3, 1, 1)
adelie_ax.set_title('Adelie penguins')
gentoo_ax = fig.add_subplot(3, 1, 2)
gentoo_ax.set_title('Gentoo penguins')
chinstrap_ax = fig.add_subplot(3, 1, 3)
chinstrap_ax.set_title('Chinstrap penguins')

# create heatmap for Adelie penguin
sns.heatmap(
    data=corrolation_adelie_df.corr(), ax=adelie_ax,
    square=False, cbar_kws={'shrink': .3}, cmap='coolwarm',
    annot=True, annot_kws={'fontsize': 7}
    )

# create heatmap for Gentoo penguin
sns.heatmap(
    data=corrolation_gentoo_df.corr(), ax=gentoo_ax,
    square=False, cbar_kws={'shrink': .3}, cmap='coolwarm',
    annot=True, annot_kws={'fontsize': 7}
    )

# create heatmap for Chinstrap penguin
sns.heatmap(
    data=corrolation_chinstrap_df.corr(), ax=chinstrap_ax,
    square=False, cbar_kws={'shrink': .3}, cmap='coolwarm',
    annot=True, annot_kws={'fontsize': 7}
    )
```
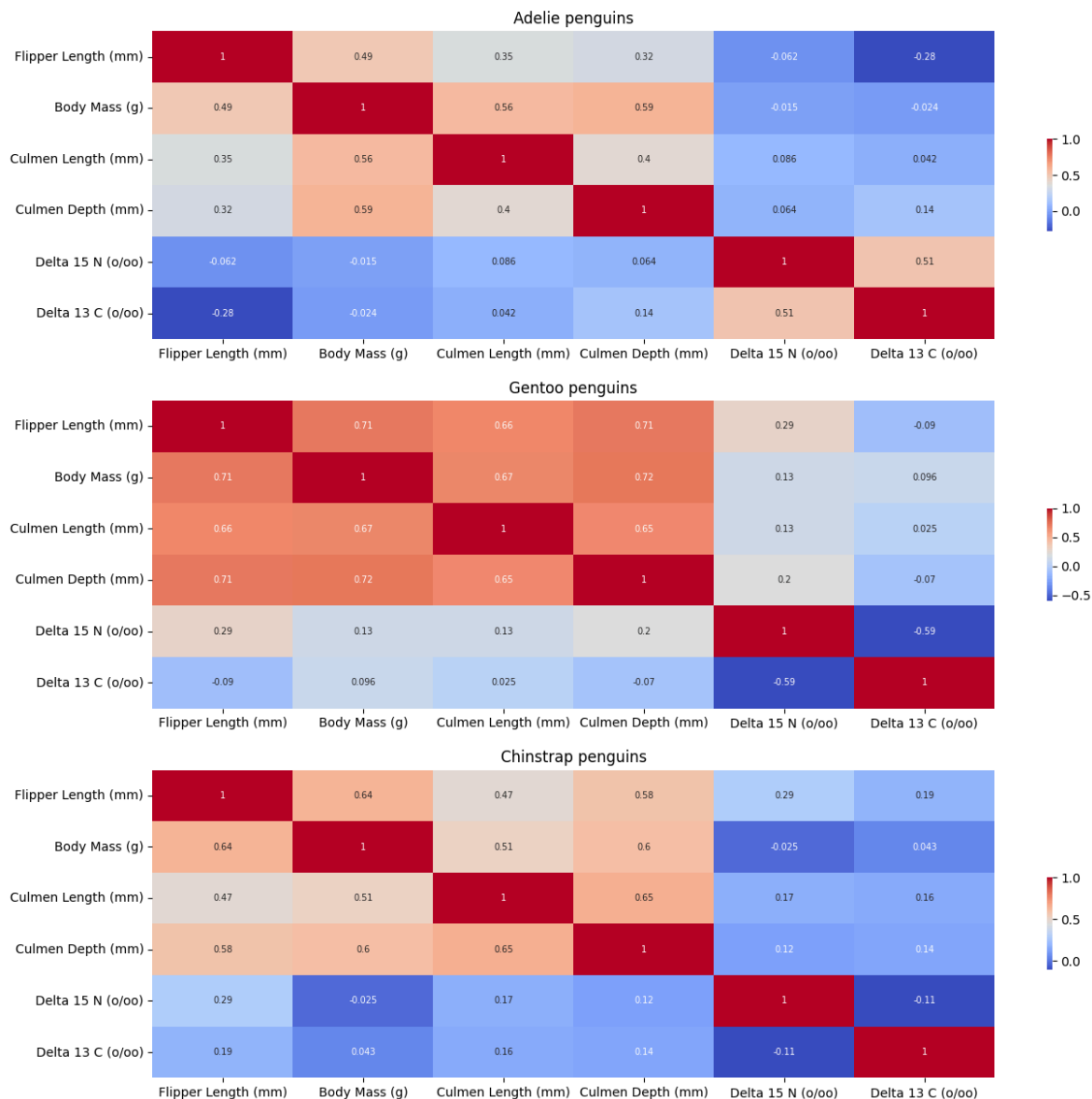
Out[30]:   <Axes: title={'center': 'Chinstrap penguins'}>

Adelie penguins

Gentoo penguins

Chinstrap penguins

Refernces

https://seaworld.org/animals/all-about/penguin/reproduction/#:~:text=A%20nest%20of%20eggs%20is,at%20least%20one%2

https://pubmed.ncbi.nlm.nih.gov/17620620/

https://archive.ics.uci.edu/dataset/690/palmer+penguins-3

The report was written by Charne Munjeri