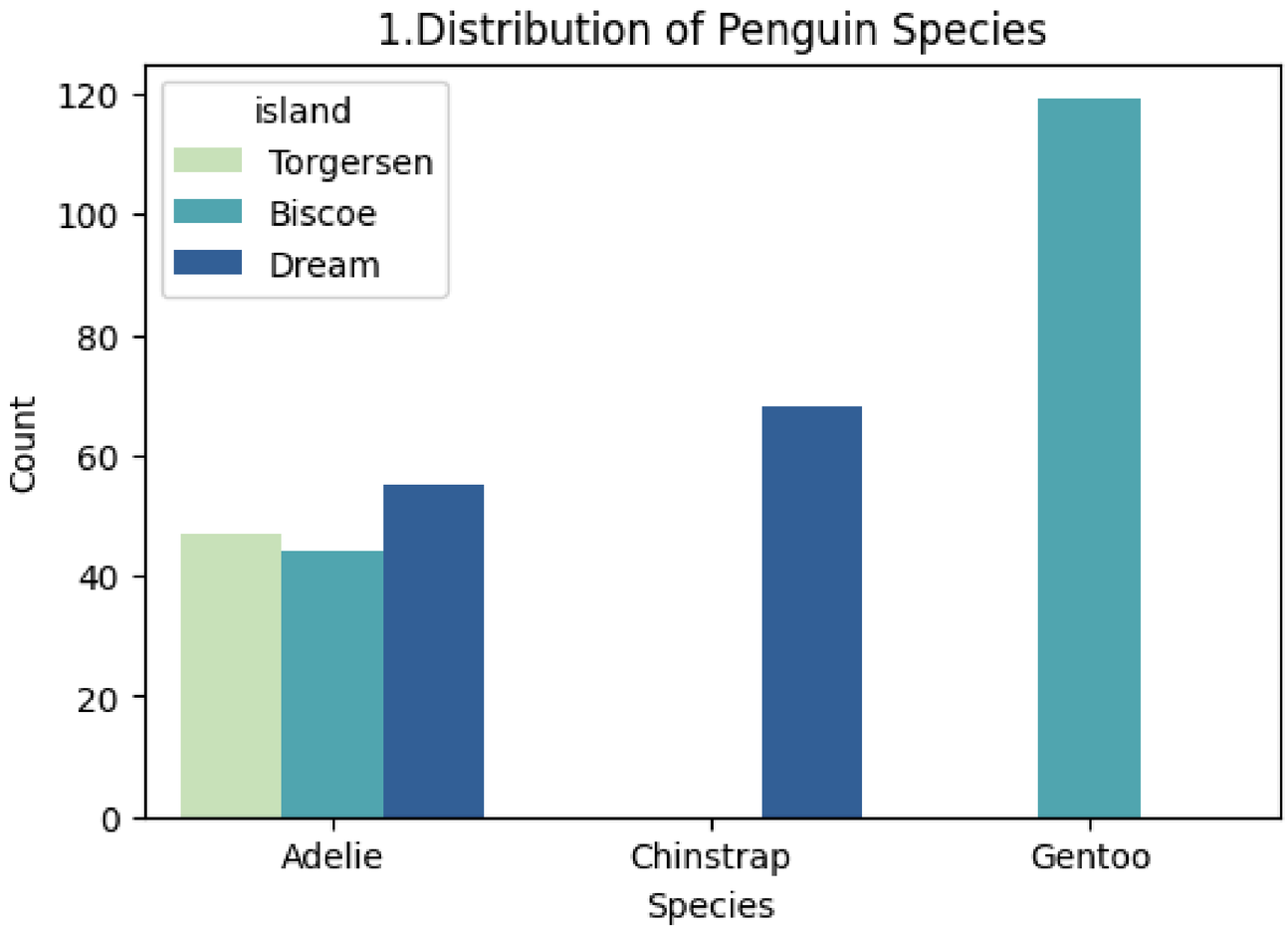
PENGUIN DATASET

What does it tell us?

Liam O’Connor

Distribution of penguins by species and island



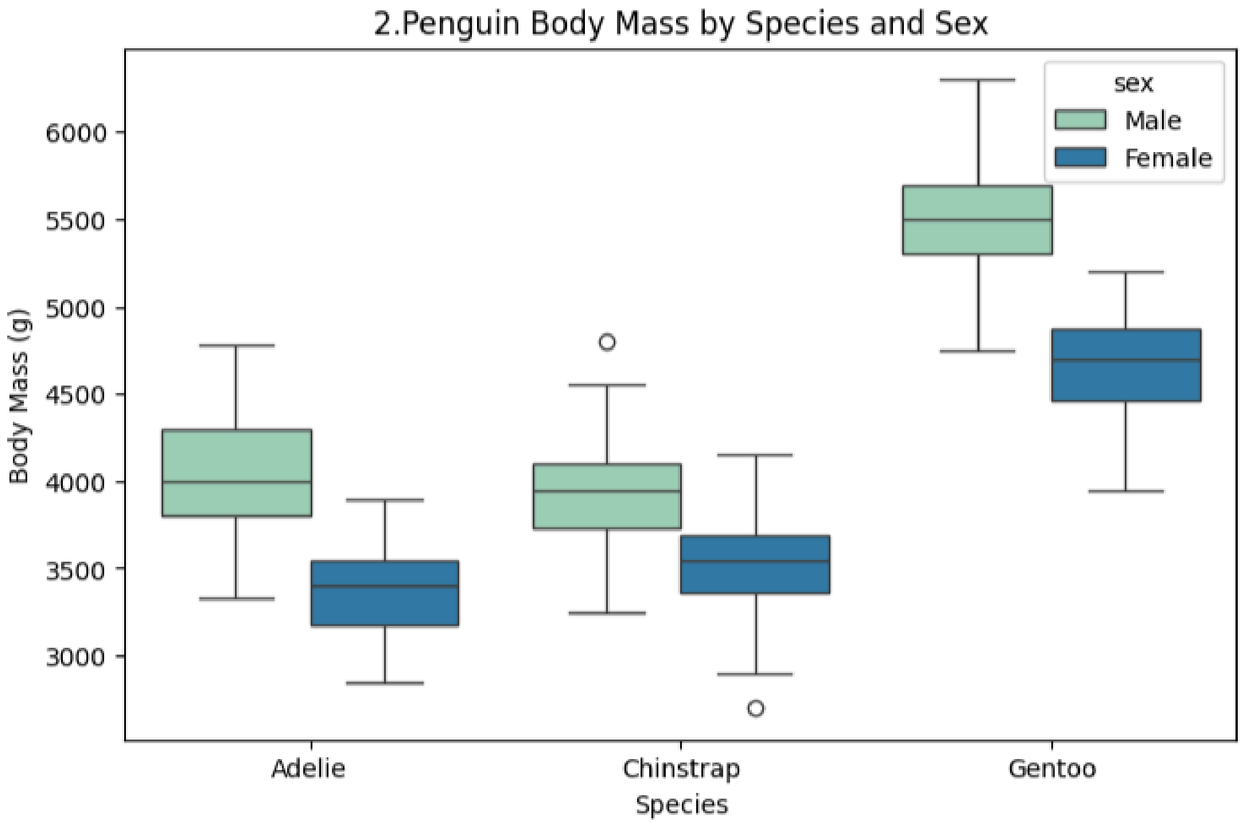
The Countplot shows the distribution of penguin by species and island. This tells us that the Adelie penguin (n=146) is most represented, and most widely spread amongst the islands in the sample. The Gentoo (n=119) are the second most common and are endemic to Biscoe Island. Whereas the Chinstrap (n=68) are endemic to Dream Island. The fact that the Adelie are more disparate amongst the islands could indicate they are more adaptable to different environments.

```
# Distribution of penguins by species and island
# Countplot
plt.figure(figsize=(6,4)) #Set plot canvas size
sns.countplot(data=penguins, x="species", hue="island", palette="YlGnBu") # Count of species split by island
plt.title("1.Distribution of Penguin Species") # Plot title
plt.xlabel("Species") # x-axis label
plt.ylabel("Count")   # Y-axis label
plt.show() # Display plot
```

PENGUIN DATASET What does it tell us?

Liam O’Connor

Penguin by body mass and sex



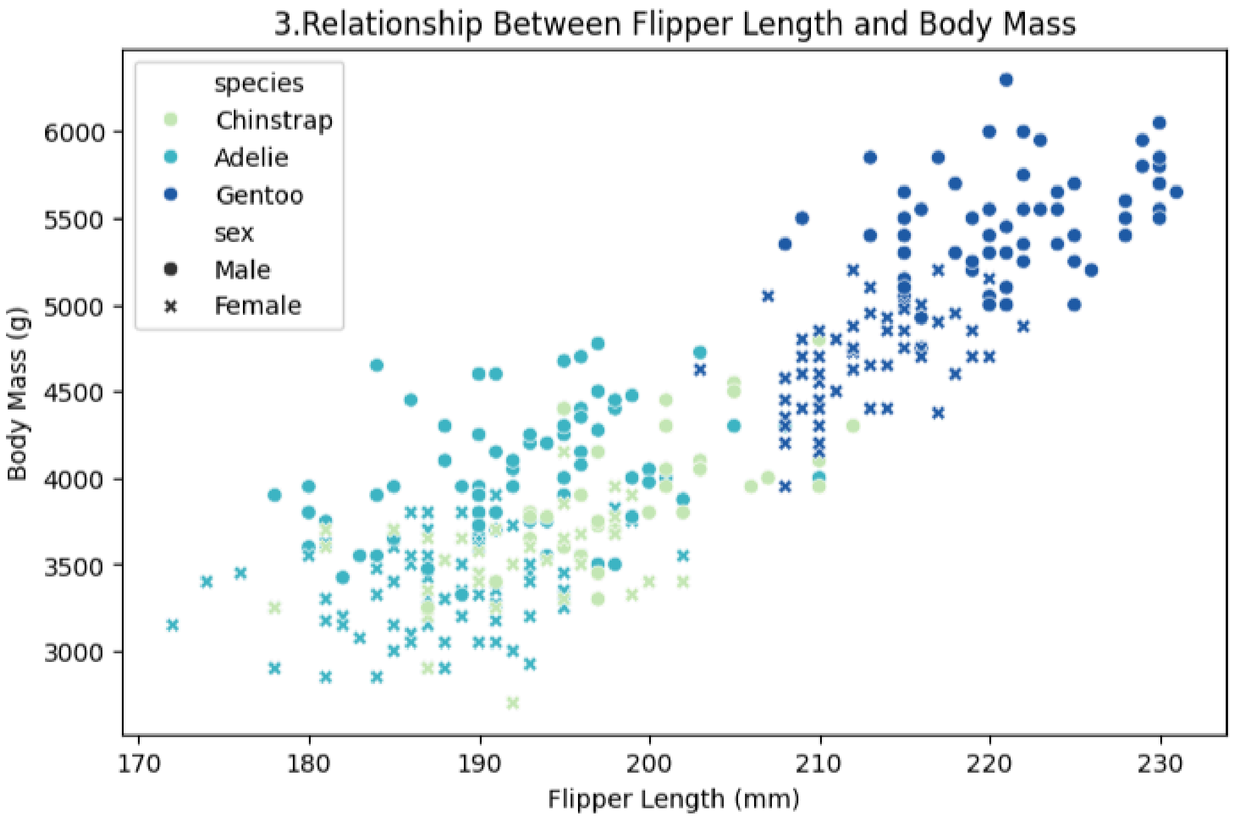
The Boxplot indicates that the Gentoo are heavier than the other penguin species in the sample. Both the male and the female Gentoo have a larger average body mass. There is a small overlap between the male Chinstrap and the female Gentoo. The difference can help in identifying Gentoo compared to the Adelie and Chinstrap. The results show there is sexual dimorphism with the males being heavier across the species than the females. Males are also heavier than females by species.

```
plt.figure(figsize=(8,5)) # Plot size
sns.boxplot(data=penguins, x="species", y="body_mass_g", hue="sex", palette="YlGnBu") # species and body mass count split by sex
plt.title("2.Penguin Body Mass by Species and Sex") # Plot title
plt.xlabel("Species") # x-axis label
plt.ylabel("Body Mass (g)") # y-axis label
plt.show() # Display plot
```

PENGUIN DATASET What does it tell us?

Liam O'Connor

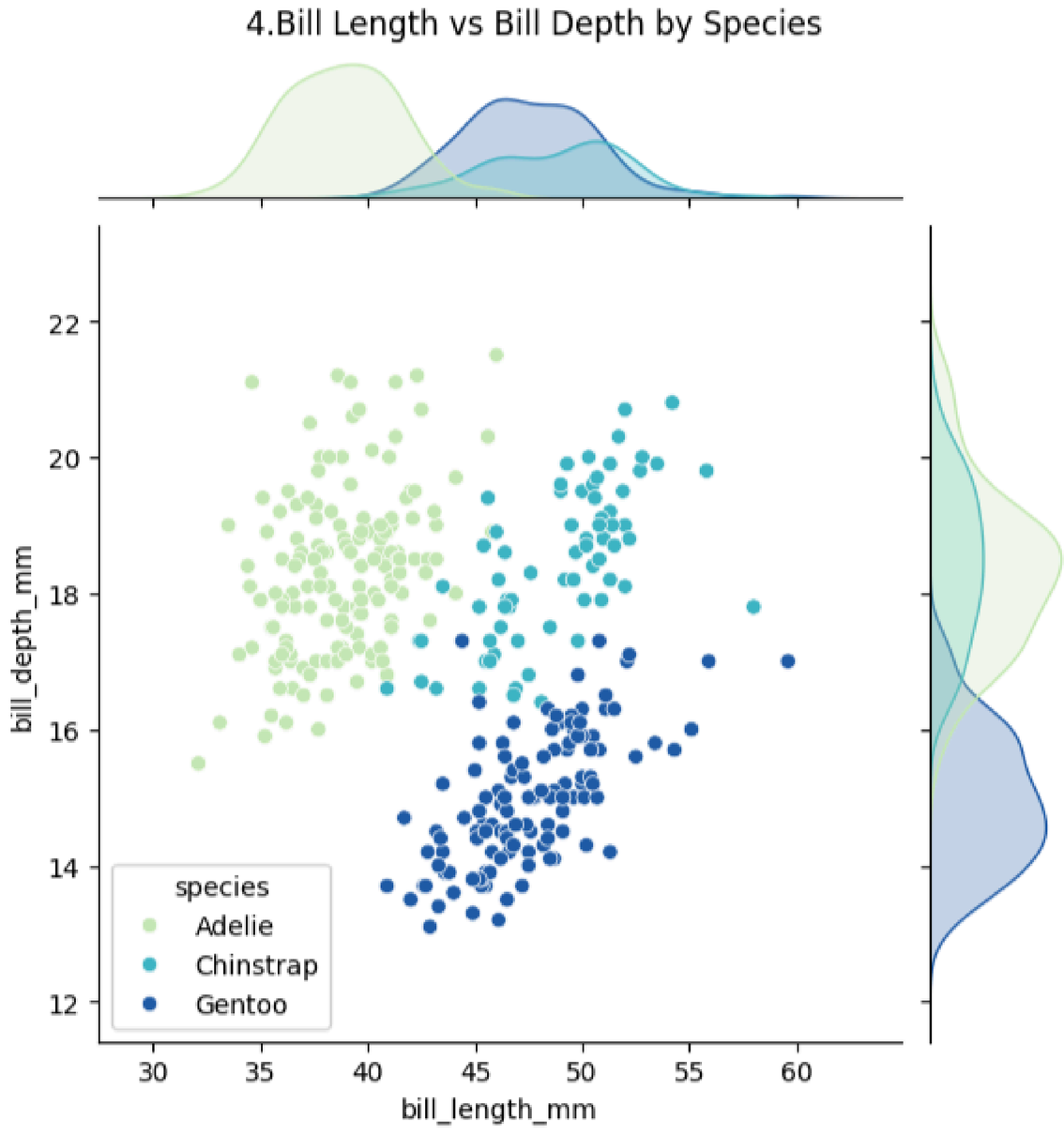
Relationship between flipper length and body mass



The Scatterplot illustrates that the larger Gentoo have longer flipper length than the Chinstrap and the Adelie. And this includes the female Gentoo compared with both male and female of the other two species. There is some overlap between the female Gentoo and the Chinstrap males. But it can be concluded that the Gentoo predominantly have a longer flipper length and that body mass is a good indicator for this. This is a positive correlation.

```
#scatterplot - Relationship between flipper length and body mass
plt.figure(figsize=(8,5))
order = ["Chinstrap", "Adelie", "Gentoo"] # List of species
# Flipper length and body mass relations split by species and styled by sex
sns.scatterplot(data=penguins,  x="flipper_length_mm",  y="body_mass_g", hue="species",  hue_order=order, style="sex",  palette="YlGnBu")
plt.title("3.Relationship Between Flipper Length and Body Mass") # Plot title
plt.xlabel("Flipper Length (mm)") # x-axis label
plt.ylabel("Body Mass (g)")      # y-axis label
plt.show() # Display plot
```

Bill Length vs Bill Depth by Species



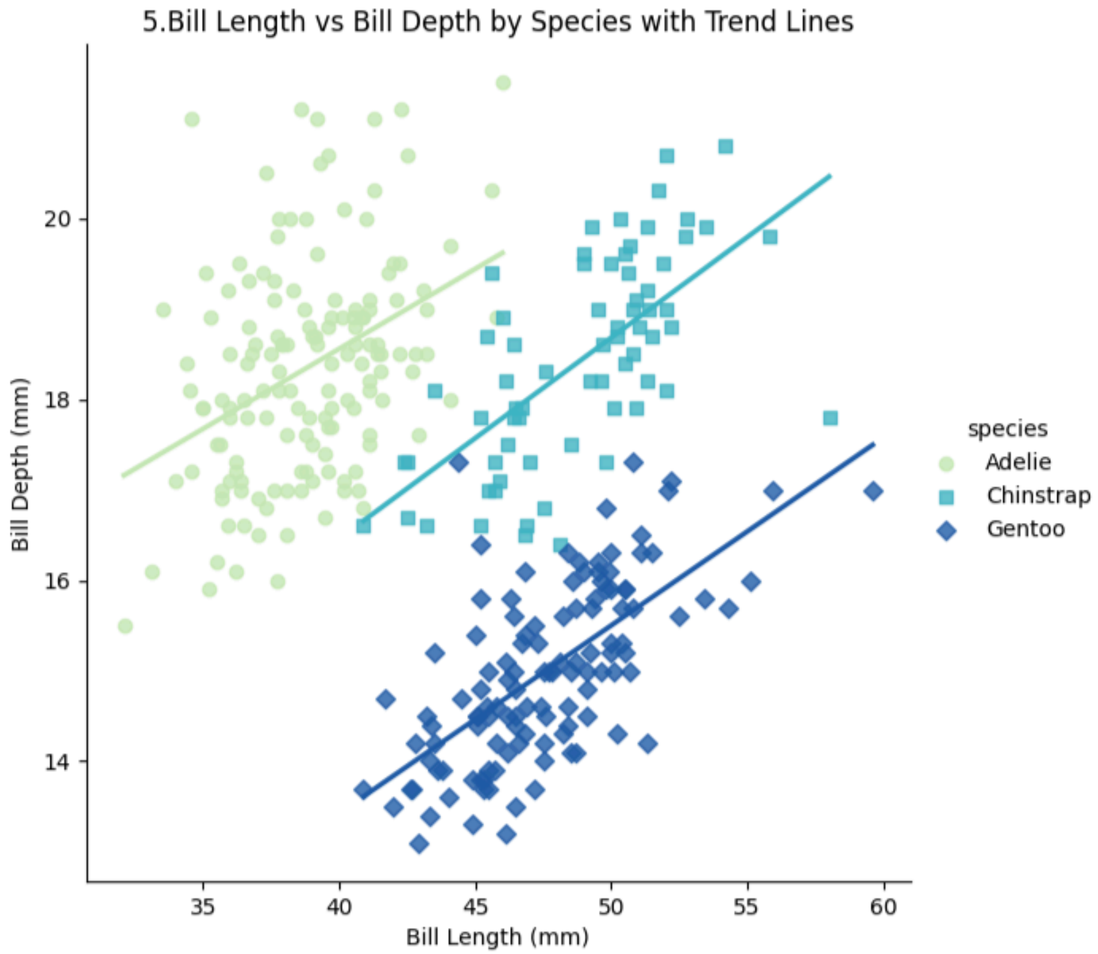
The Jointplot indicates that the Gentoo penguin has an average longer bill length, with a less bill depth, in comparison to the Chinstrap. The Chinstrap, on average, has a shorter bill length, but deeper bill depth, than the Gentoo. The Adelie penguin has a shorter, and less deep bill on average. This could be a biological trait of the penguins' habitat and feeding environment.

```
# Bill length and bill depth split by species scatterplots and histplots
sns.jointplot(
   data=penguins,
   x="bill_length_mm",
   y="bill_depth_mm",
   hue="species",
   palette="YlGnBu",
   height=6
)
plt.suptitle("4.Bill Length vs Bill Depth by Species", y=1.02) #Plot title (suptitle used with jointplot for clarity)
plt.show() # Display plot
```

PENGUIN DATASET What does it tell us?

Liam O'Connor

Bill Length vs Bill Depth by Species with Trend Lines



The Implot trend lines support the Jointplot results. With the Adelie clustered to the left indicating a shorter, deeper bill. The Chinstrap have a correlation between longer and deeper bills. And the Gentoo, the largest penguin, on average have a longer, more slender bill. This leads to the assumption that the Adelie penguin, which is widely spread across the sample islands, has a more adaptable bill for feeding. Whereas the Gentoo, (Biscoe Island), and Chinstrap, (Dream Island), have more specialised bills adapted for the island habitats they are restricted to.

```
# Bill length and depth with scatterplot and trend lines
sns.lmplot(
  data=penguins,
x="bill_length_mm",
y="bill_depth_mm",
hue="species",
palette="YlGnBu",
  height=6,
markers=["o", "s", "D"], # species markers
  ci=None              # removes confidence interval shading
)
```

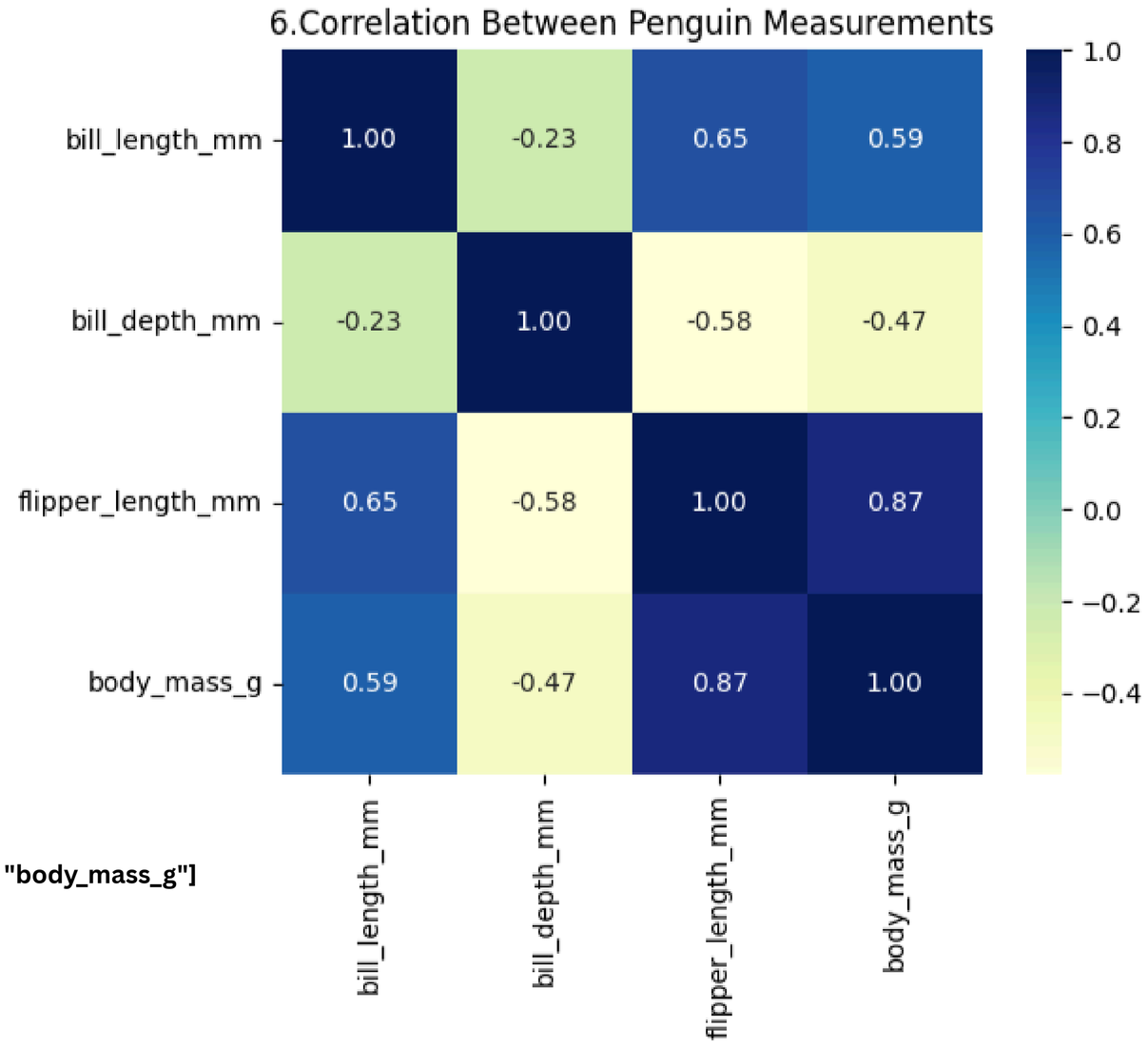
```
plt.title("5.Bill Length vs Bill Depth by Species with Trend Lines") # Plot title
plt.xlabel("Bill Length (mm)") # x-axis label
plt.ylabel("Bill Depth (mm)") # y-axis label
plt.show() # Display plot
```

PENGUIN DATASET What does it tell us?

Liam O’Connor

Correlation Between Penguin Measurements

The Heatmap shows the positive correlation between a penguin's body mass and flipper length, e.g. (0.87 blue) and the negative correlation between the body mass and the bill depth e.g. (-0.47 green). This is borne out by the Gentoo's longer but slender bill.




```
# Correlation Between Penguin Measurements
#Heatmap
import seaborn as sns
import matplotlib.pyplot as plt

# Numeric columns
numeric_cols = ["bill_length_mm", "bill_depth_mm", "flipper_length_mm", "body_mass_g"]
penguin_numeric = penguins[numeric_cols]

# Correlate data columns
corr = penguin_numeric.corr()

# Plot heatmap
plt.figure(figsize=(6,5)) # Plot size
sns.heatmap(corr, annot=True, cmap="YlGnBu", fmt=".2f") # Heatmap correlation mattrix(DataFrame), numbers in cells, colour map, format to 2 decimal places
plt.title("6.Correlation Between Penguin Measurements") #Plot title
plt.show() # Display plot
```

PENGUIN DATASET What does it tell us?

Liam O'Connor