Notebook:

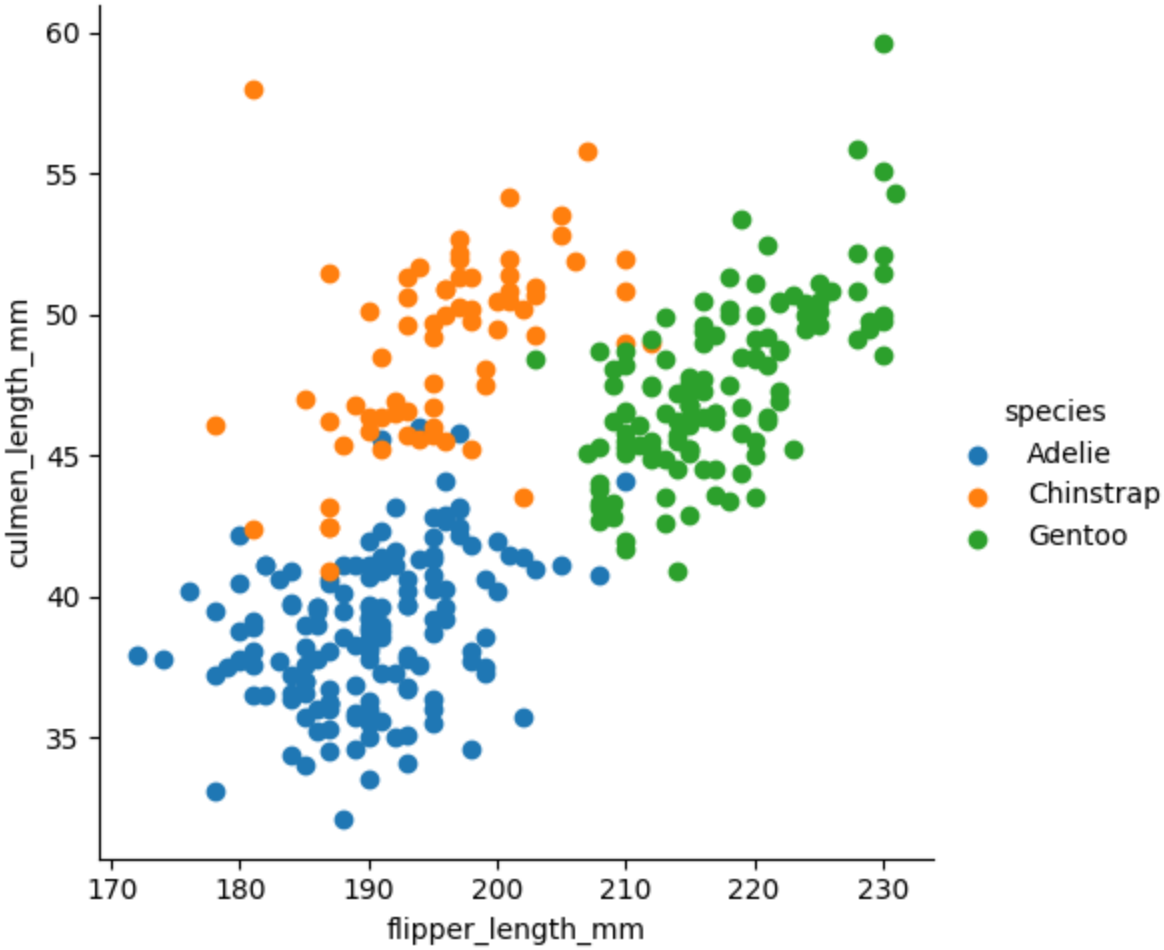
In [ ]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [ ]:
```
df = pd.read_csv('./penguins_size.csv')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 344 entries, 0 to 343
Data columns (total 7 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   species            344 non-null    object 
 1   island             344 non-null    object 
 2   culmen_length_mm   342 non-null    float64
 3   culmen_depth_mm    342 non-null    float64
 4   flipper_length_mm  342 non-null    float64
 5   body_mass_g        342 non-null    float64
 6   sex                334 non-null    object 
dtypes: float64(4), object(3)
memory usage: 18.9+ KB
```

In [ ]:
```
sns.FacetGrid(df, hue="species", height=5) \
.map(plt.scatter, "flipper_length_mm", "culmen_length_mm") \
.add_legend()
```

Out[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x7f849c6bd4d0>
```

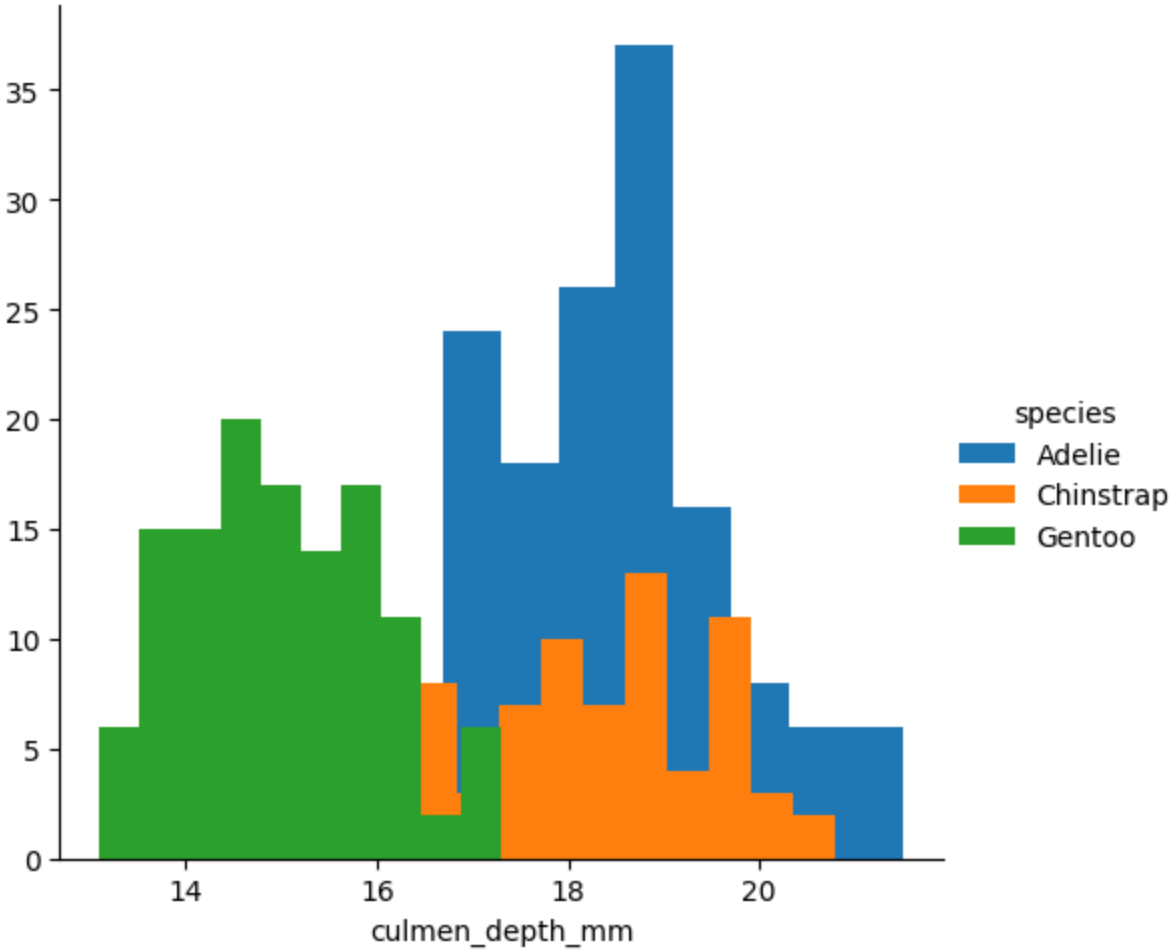


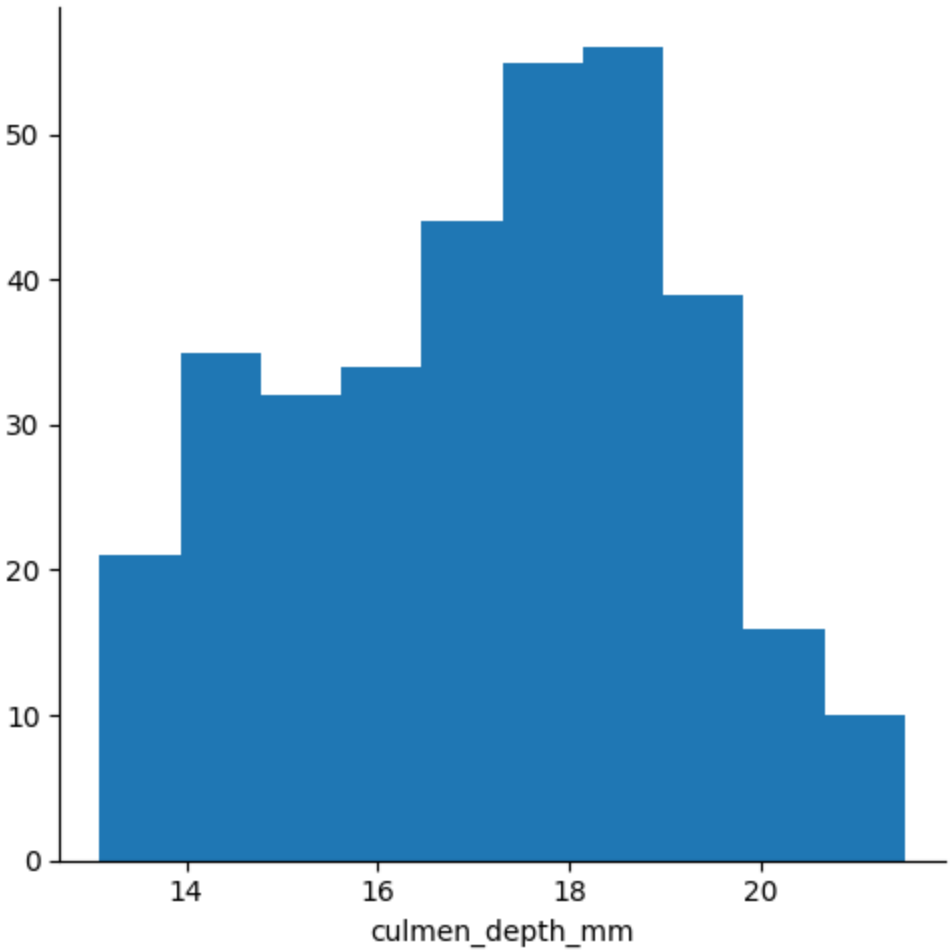
From this graph we can infer the following :

1. The ratio between flipper length and bill length is relatively similar
2. The ratio between flipper length and bill length is approximately 5.14
3. The data is similar to that of a regression problem

In [ ]:
```
sns.FacetGrid(df, hue="species", height=5) \
.map(plt.hist, "culmen_depth_mm") \
.add_legend()
sns.FacetGrid(df, height=5) \
.map(plt.hist, "culmen_depth_mm")
```

Out[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x7f8500237dd0>
```


35
30
25
20
15
10
5
0
14
16
18
20
culmen_depth_mm
species
Adelie
Chinstrap
Gentoo

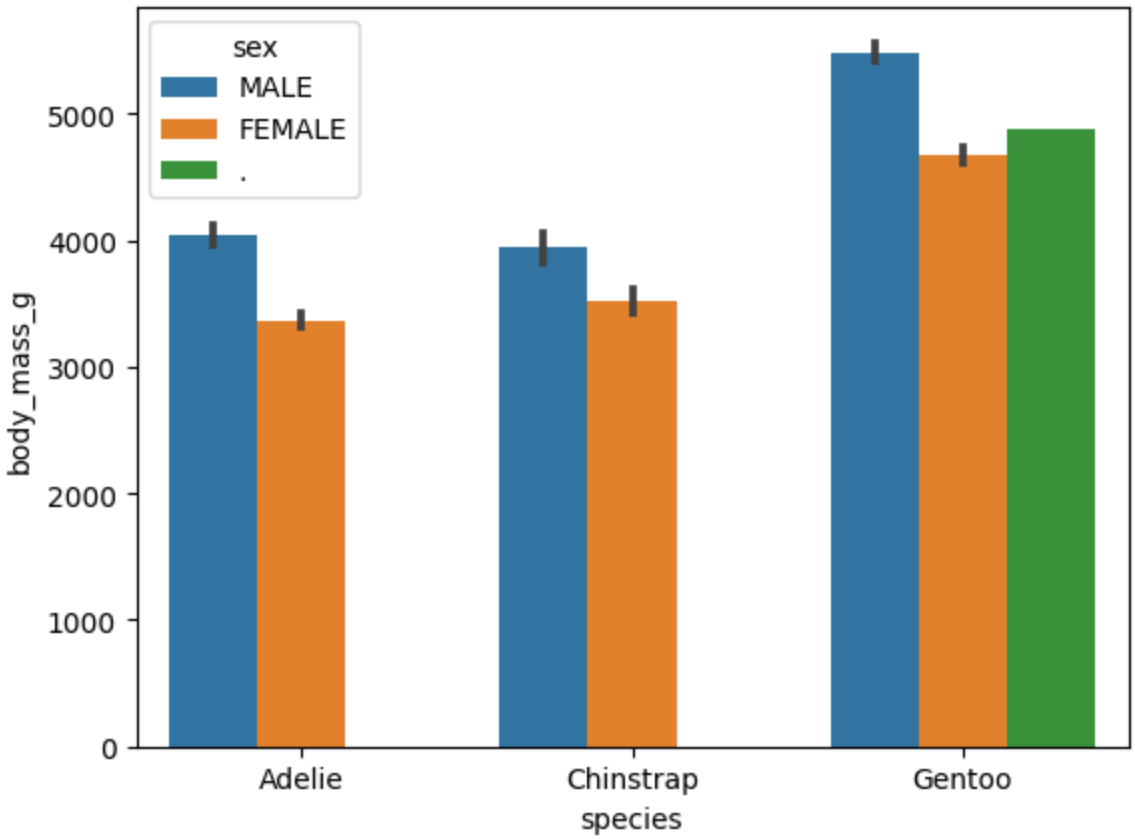


From this graph we can infer the following :

1. The bill depth has a substantial variance
2. The Adelie species has the largest bill depth amongst other species
3. The Chinstrap species has the smallest average bill depth

In [ ]:
```
sns.barplot(df, x="species", y="body_mass_g" ,hue="sex")
```

Out[ ]:
```
<Axes: xlabel='species', ylabel='body_mass_g'>
```

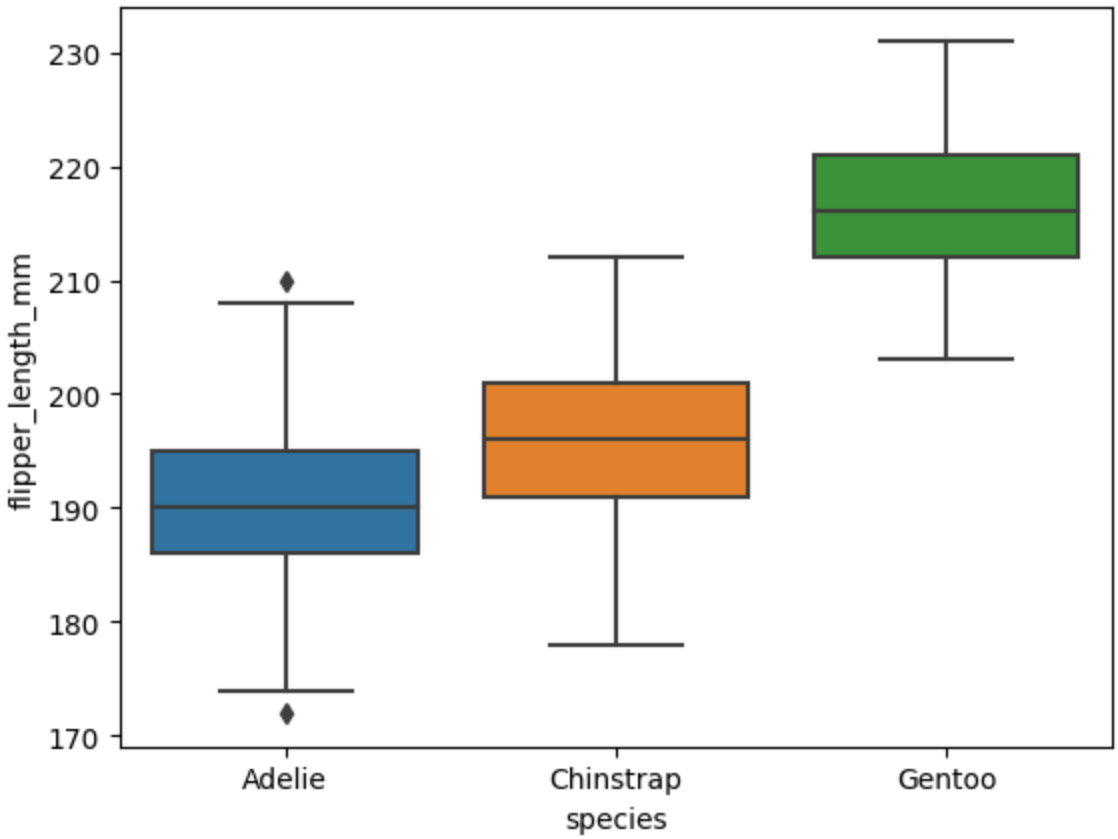


From this graph we can infer the following :

1. The males of all species have larger average body masses
2. The gentoo species have the largest body mass
3. The chinstrap has the smallest body mass on average
4. The gentoo species has 3 genders

In [ ]:
```
sns.boxplot(df, x="species", y="flipper_length_mm")
```

Out[ ]:
```
<Axes: xlabel='species', ylabel='flipper_length_mm'>
```

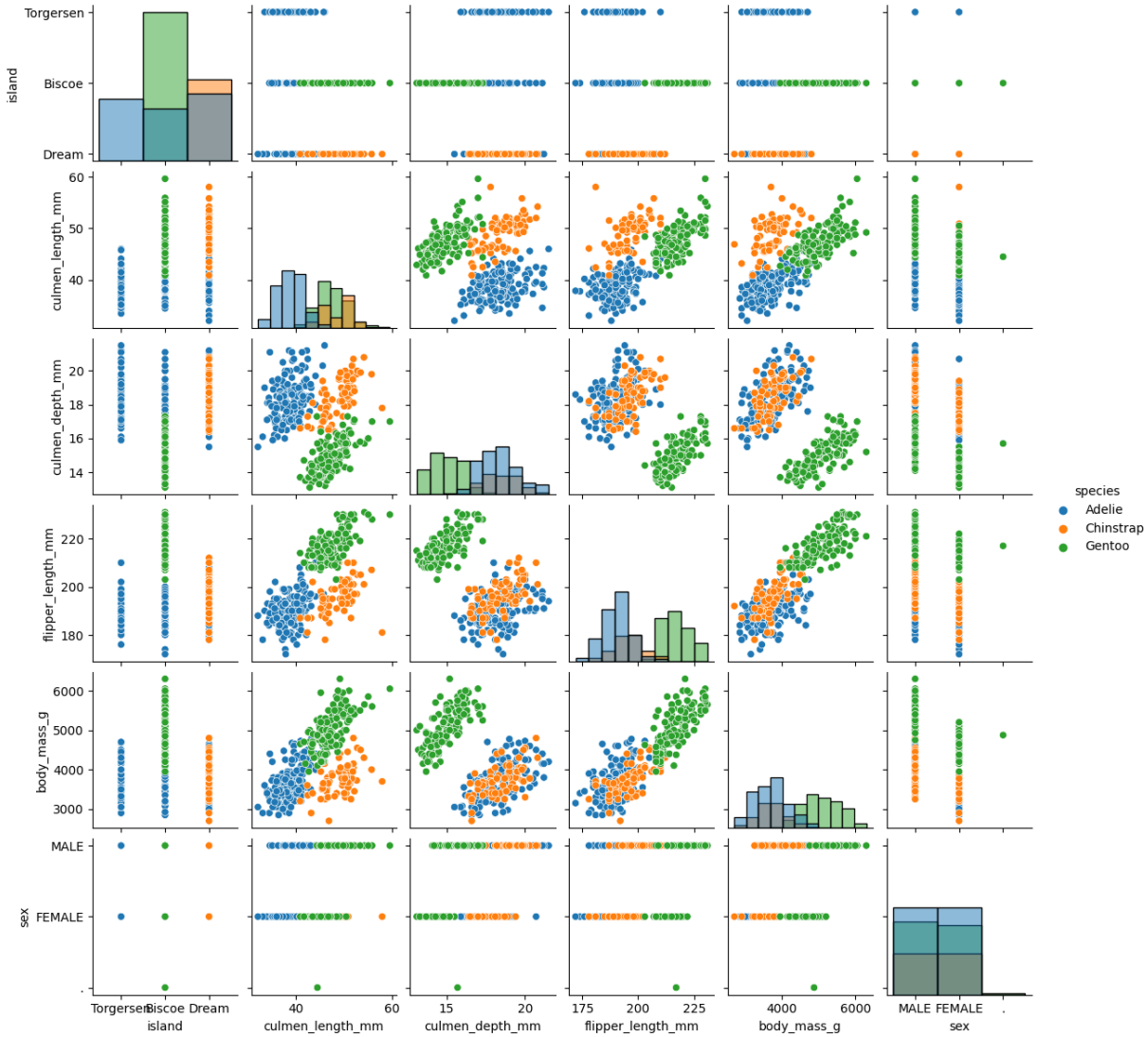


From this graph we can infer the following :

1. The gentoo species have the largest flipper lengths on average
2. The Adelie species has the smallest flipper length on average
3. The chinstrap species has an average flipper length, more on the short side

In [ ]:
```
x=["island","culmen_length_mm","culmen_depth_mm","flipper_length_mm","body_m
y=["island","culmen_length_mm","culmen_depth_mm","flipper_length_mm","body_m
sns.pairplot(df,x_vars=x,y_vars=y,hue="species", height=2, diag_kind="hist")
```

Out[ ]:
```
<seaborn.axisgrid.PairGrid at 0x7f8498e98d10>
```



From this graph we can infer the following :

1. The gentoo species are the largest among the 3
2. The Adelie species are the smallest on average
3. The males of all species are generally bigger