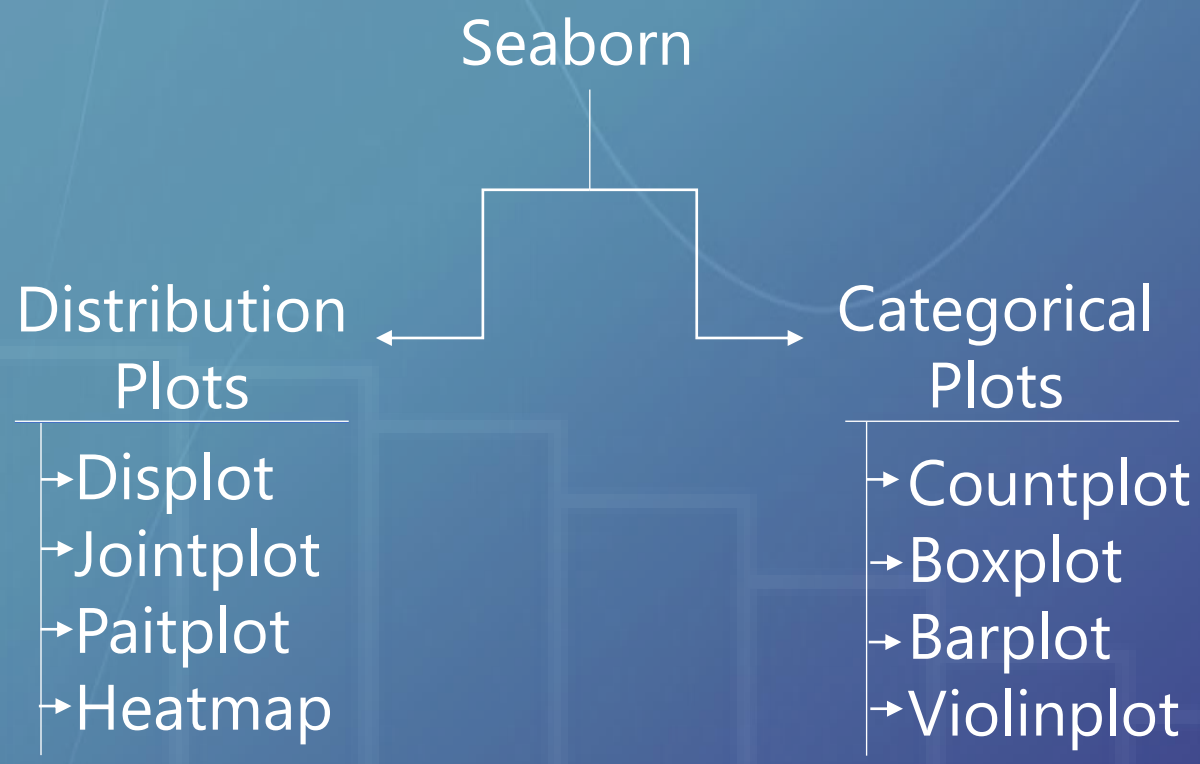
DATA VISUALISATION USING SEABORN

Sneha Mandal

INTRODUCTION

Seaborn is a Python library for 2D and advanced visualization of certain dataframe. It helps to explore and analyse a dataset in more in-depth manner.

There are two varieties of plotting that can be done in seaborn- 1)Distribution Plotting and 2)Categorical Plotting.



Sneha Mandal

Importing Library

In [3]:
```
import seaborn as sns
import warnings

warnings.filterwarnings('ignore')
```

In [4]:
```
# Loading inbuilt dataframe from seaborn module

df=sns.load_dataset("penguins")

df.head()
```

Out[4]:

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female

Distribution Plots

Displot

Displot is used to analyse distribution of column features. It will give a histogram of column parameter. kde here means Kernel Density plot over histogram, it is set to True by default. Bins are the number of buckets you want to create for the column distribution. discrete is used to get rid of unnecessary gaps in the plot. You can also use the kind variable inside this to change the kind of plot. Also there are some other variables like hue, pallette, legend, rug, etc.

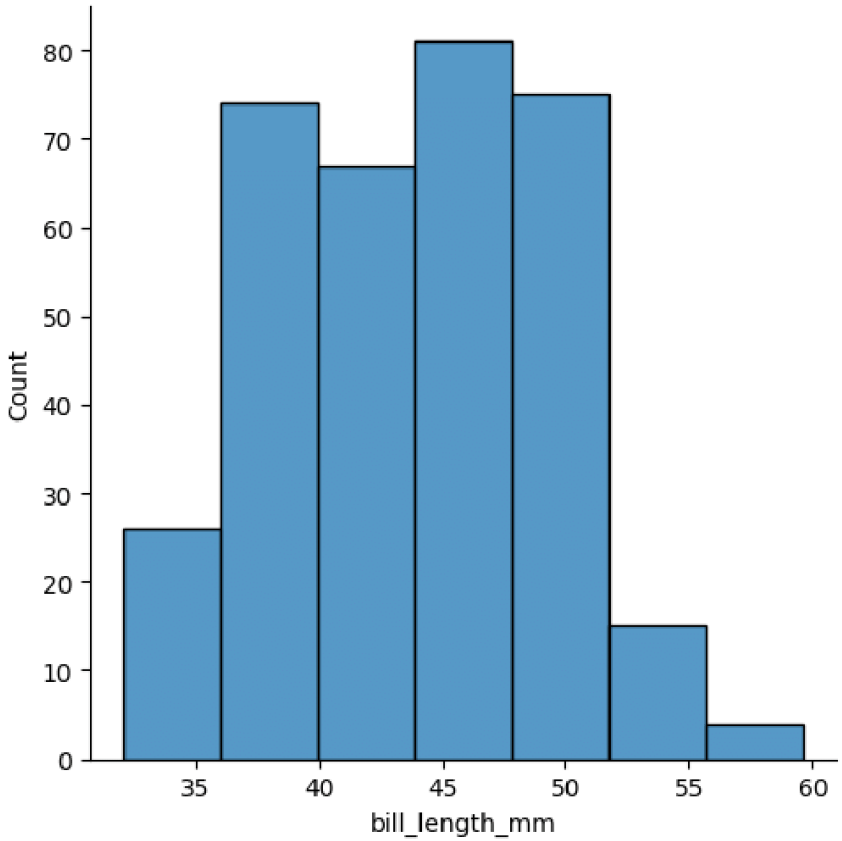
In [5]:
```
#distplot(dataframe, kde=True/False, bins, stat, ... )

sns.displot(df['bill_length_mm'], kde=False, bins=7)
```

Out[5]:
```
<seaborn.axisgrid.FacetGrid at 0x1d0511e10d0>
```

nts/TeX/fontdata.js

Sneha Mandal

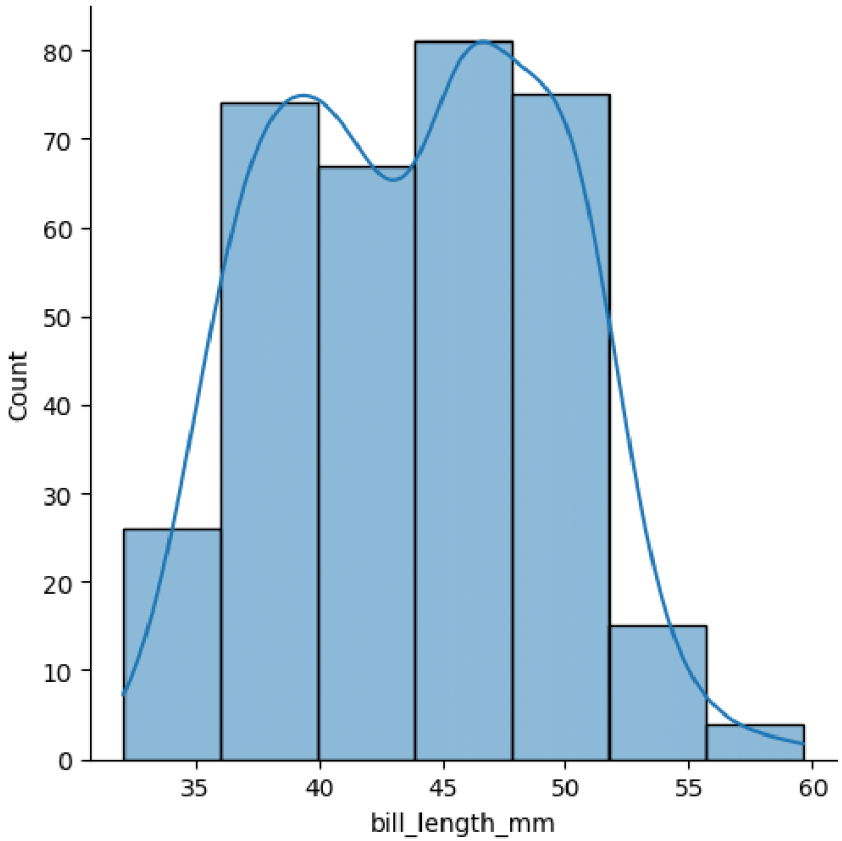


In [6]:
```
sns.displot(df['bill_length_mm'], kde=True, bins=7)
```

Out[6]:
```
<seaborn.axisgrid.FacetGrid at 0x1d0512082d0>
```

nts/TeX/fontdata.js

Sneha Mandal



Joinplot

A join plot allows to study the relationship between multiple features in comparison to each other. The central chart display the correlation of the columns. It is usually a scatterplot, a hexbin plot, a 2D histogram or a 2D density plot.

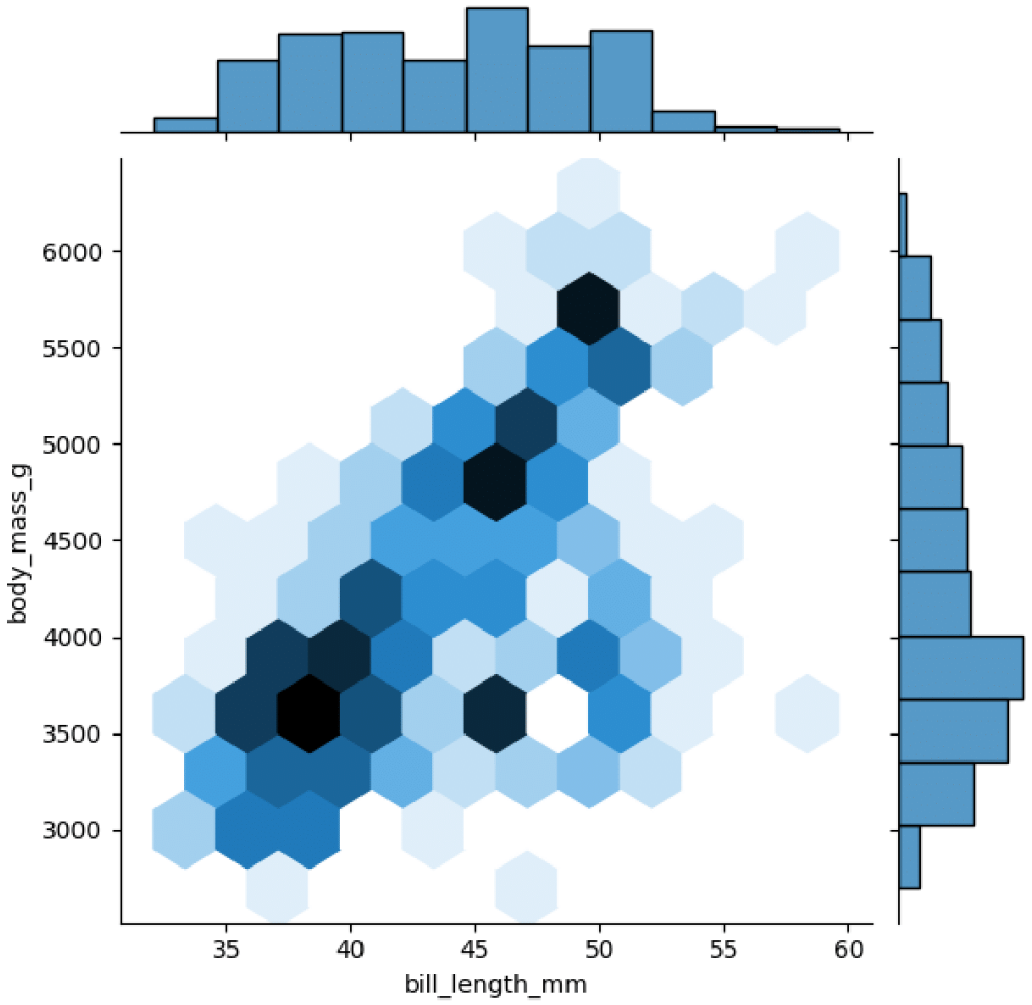
In [7]:
```
#jointplot(data, x_axis, y_axis, hue, kind, palette,...)

sns.jointplot(x='bill_length_mm', y='body_mass_g', data=df, kind='hex')
```

Out[7]:
```
<seaborn.axisgrid.JointGrid at 0x1d05131fed0>
```

nts/TeX/fontdata.js

Sneha Mandal

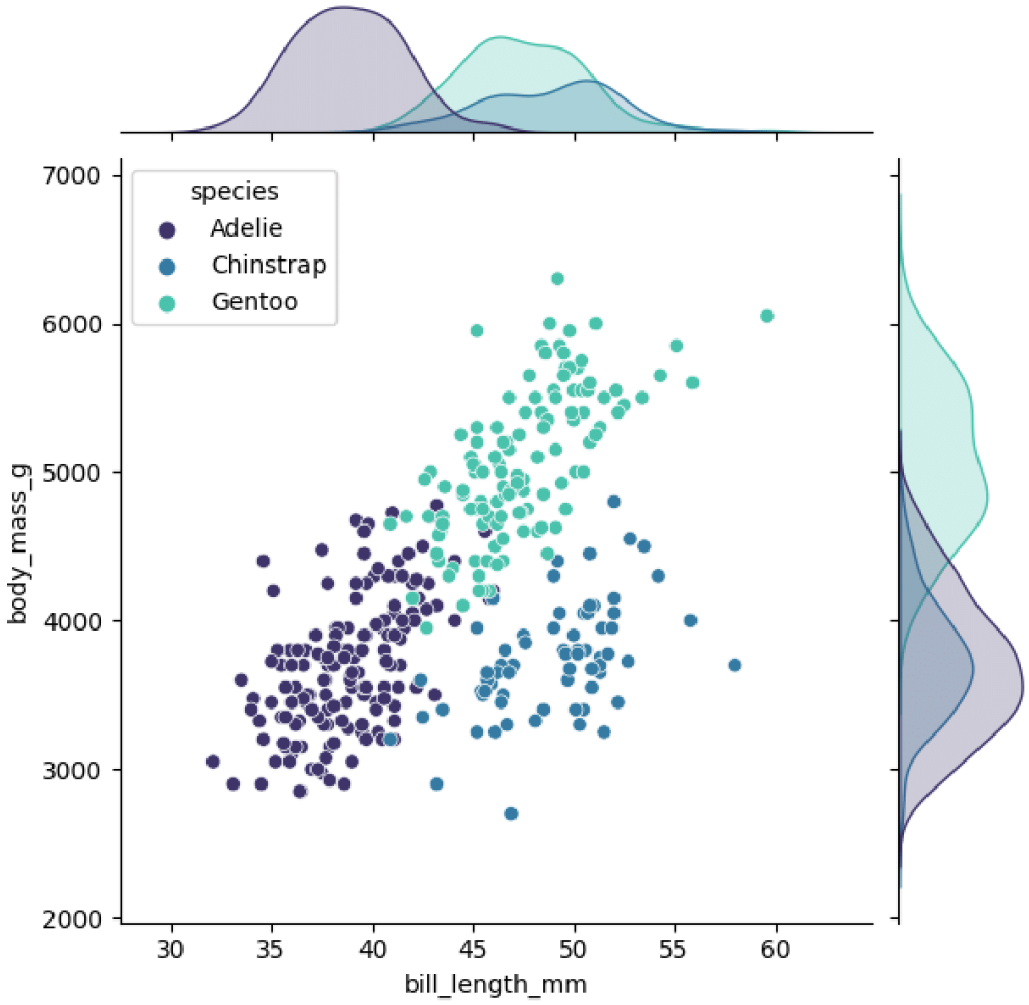


In [8]:
```
sns.jointplot(x='bill_length_mm', y='body_mass_g', data=df, hue='species', p
```

Out[8]:
```
<seaborn.axisgrid.JointGrid at 0x1d053a35890>
```

nts/TeX/fontdata.js

Sneha Mandal

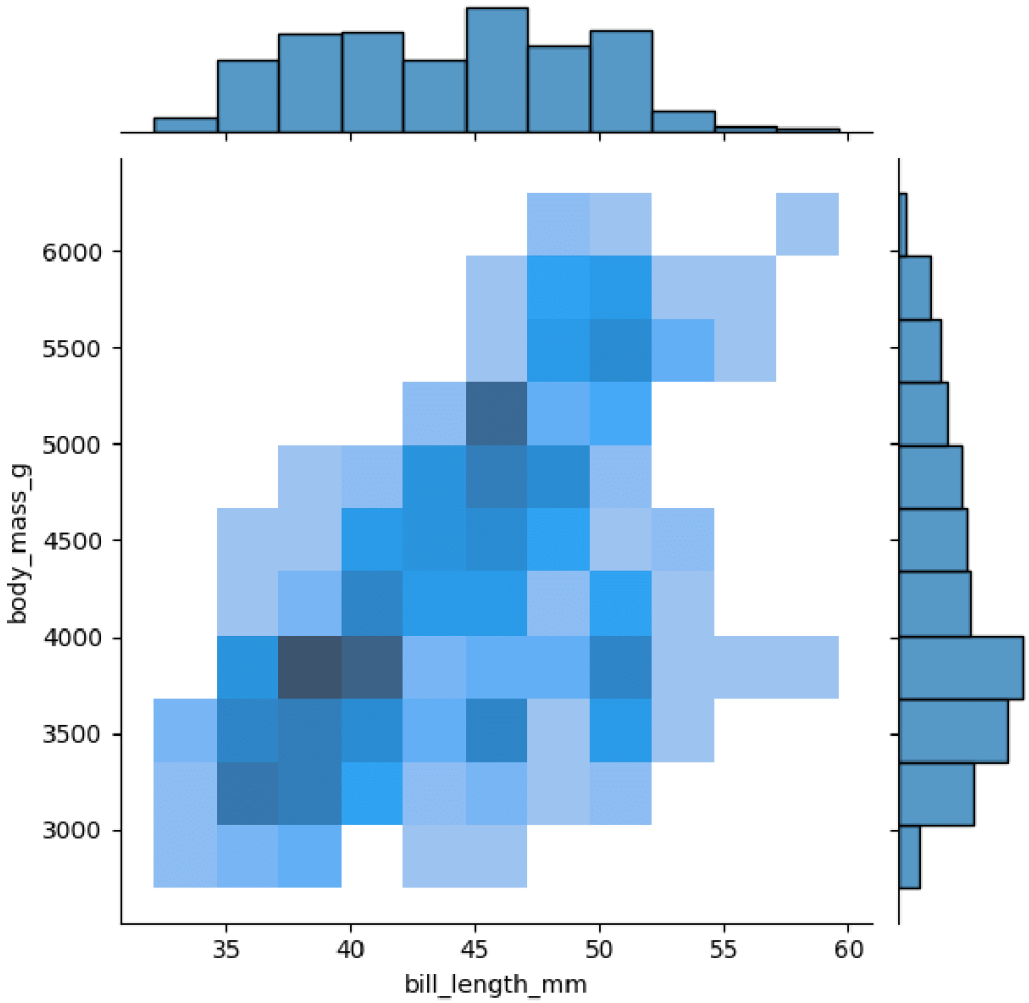


In [9]:
```
sns.jointplot(x='bill_length_mm', y='body_mass_g', kind='hist', data=df)
```

Out[9]:
```
<seaborn.axisgrid.JointGrid at 0x1d053c94cd0>
```

nts/TeX/fontdata.js

Sneha Mandal



Pairplot

Pairs plots shows the variations of all the parameters with respect to each other.

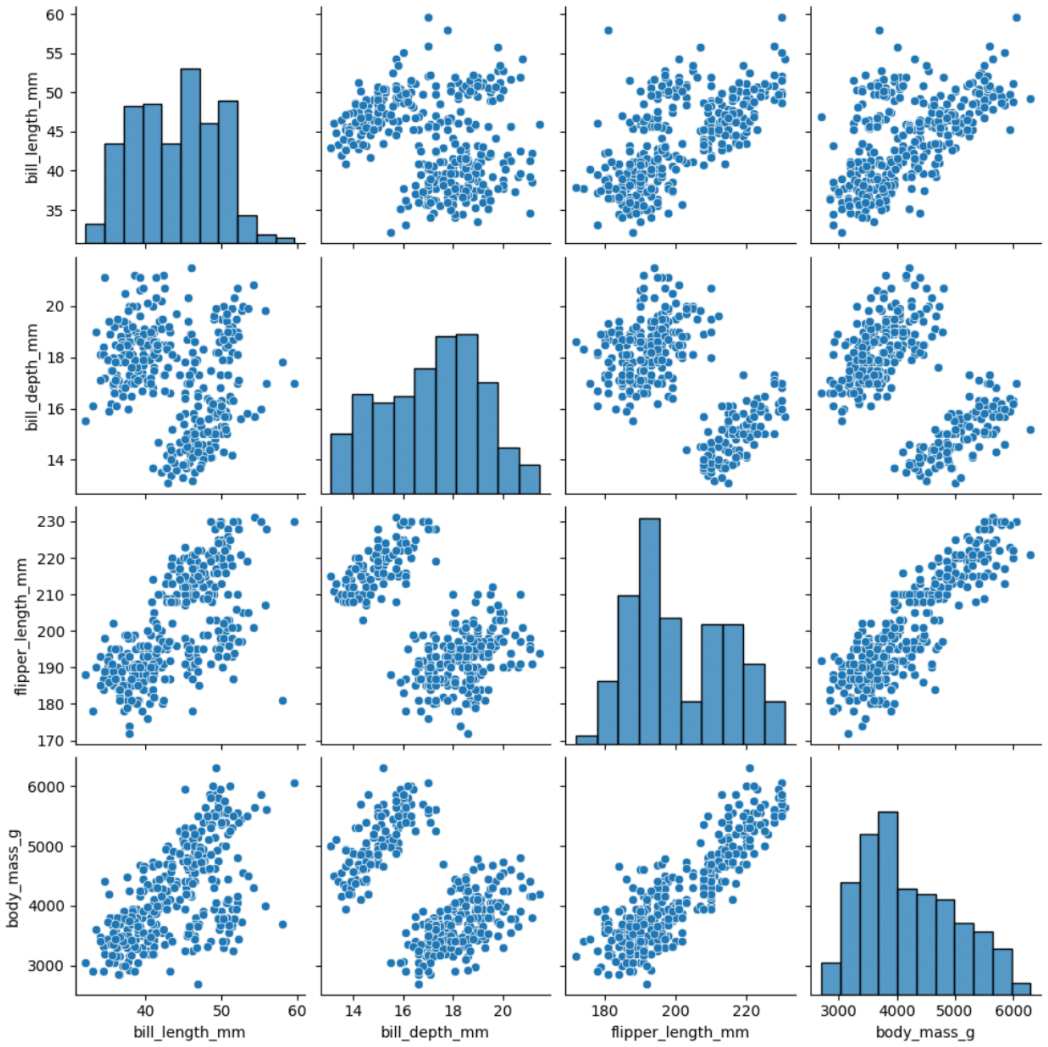
In [10]:
```
#pairplot(data, hue, palette, vars, kind, ...)

sns.pairplot(df)
```

Out[10]:
```
<seaborn.axisgrid.PairGrid at 0x1d053c92910>
```

nts/TeX/fontdata.js

Sneha Mandal

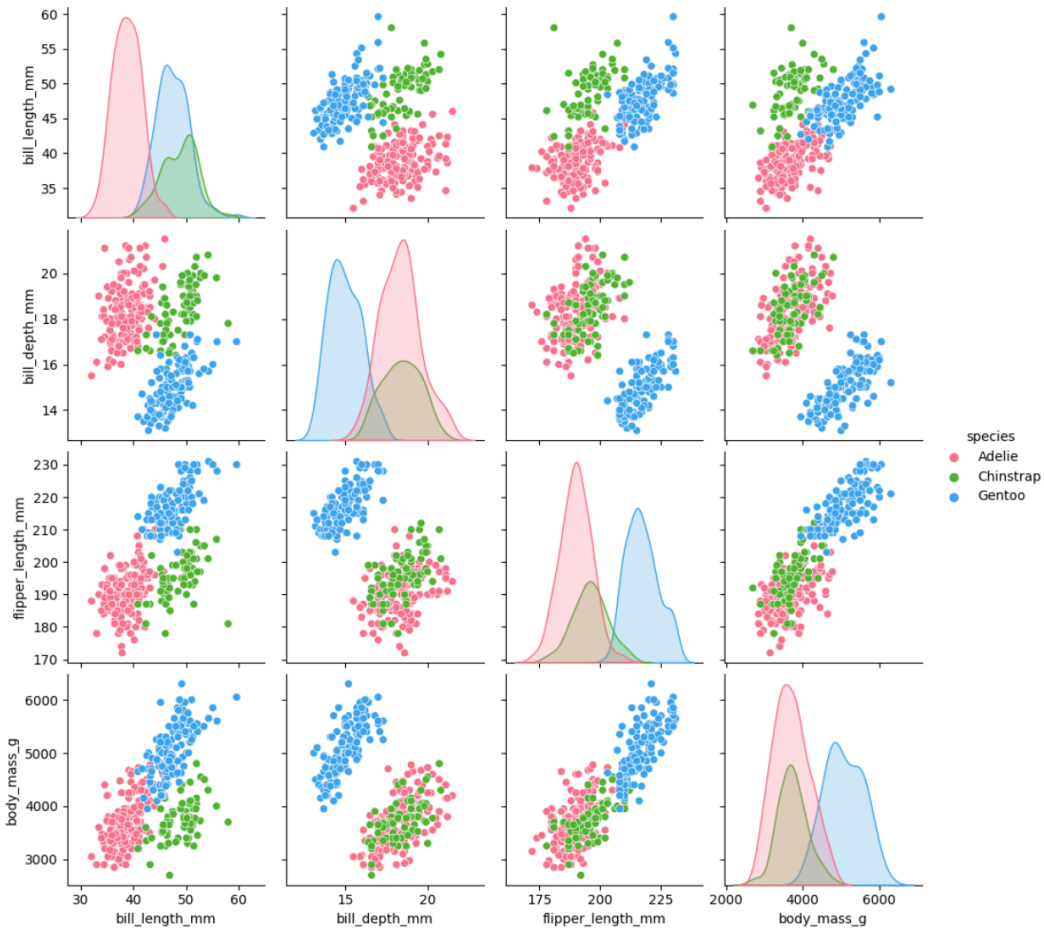


In [11]:
```
sns.pairplot(df, palette='husl', hue='species')
```

Out[11]:
```
<seaborn.axisgrid.PairGrid at 0x1d055846f50>
```

nts/TeX/fontdata.js

Sneha Mandal



Correlation / Heatmap

A correlation is a heatmap that uses colored cellscorrelation matrix (table) between two discrete dimensions or event types. It is very important in Feature Selection

In [12]:
```
df.corr()
```

Out[12]:

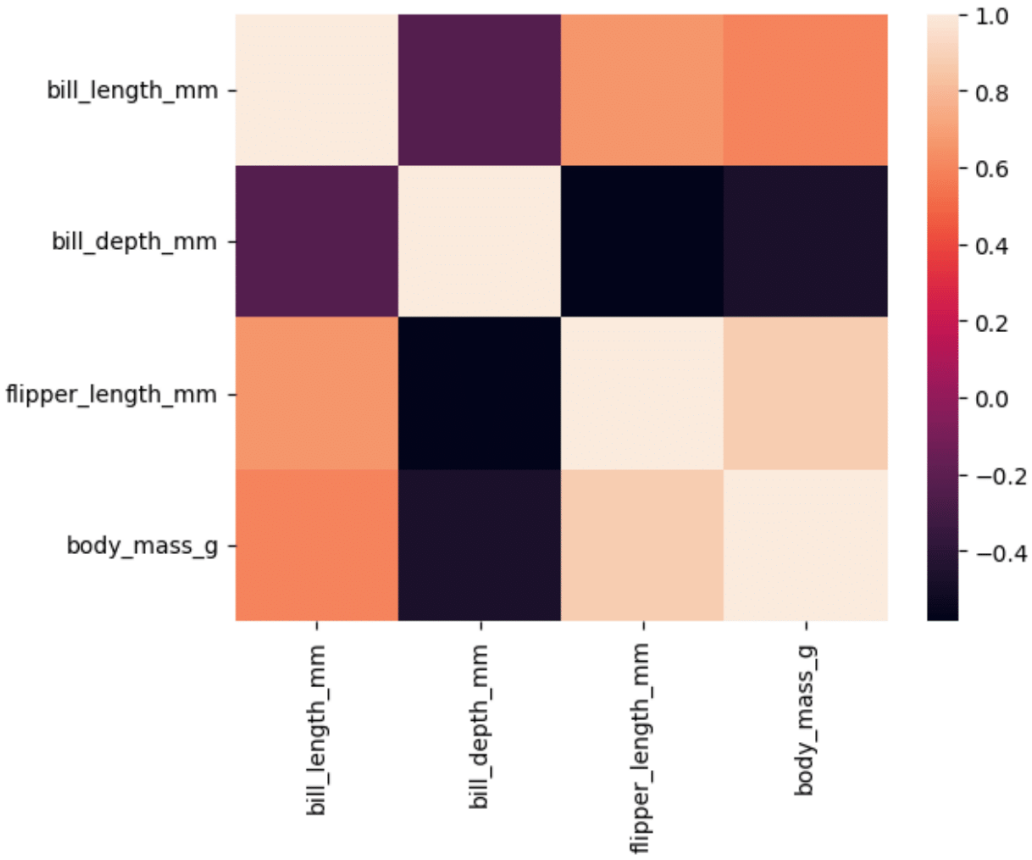
	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g
bill_length_mm	1.000000	-0.235053	0.656181	0.595110
bill_depth_mm	-0.235053	1.000000	-0.583851	-0.471916
flipper_length_mm	0.656181	-0.583851	1.000000	0.871202
body_mass_g	0.595110	-0.471916	0.871202	1.000000

In [22]:
```
sns.heatmap(df.corr())
```

Out[22]: <Axes: >

nts/TeX/fontdata.js

Sneha Mandal



Categorical Plots

Countplot

Count plot shows count of each variable in form of bars.

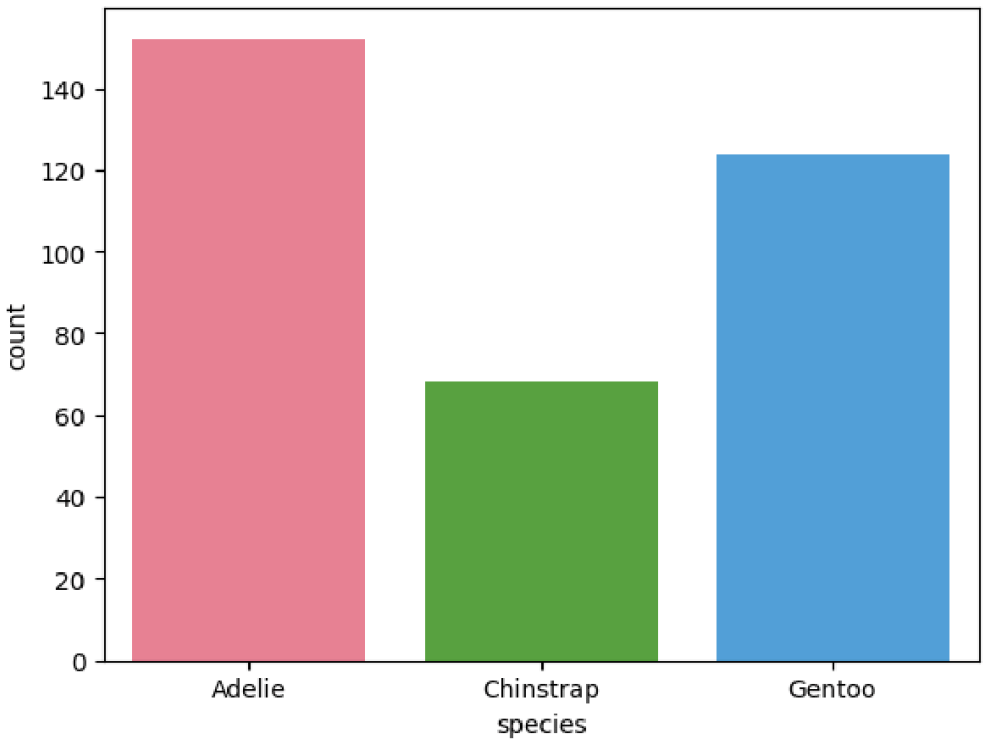
In [20]:
```
#countplot(data, x_axis, y_axis, hue, order, orient, palette, ...)

sns.countplot(data=df, x='species', palette='husl')
```

Out[20]:
```
<Axes: xlabel='species', ylabel='count'>
```

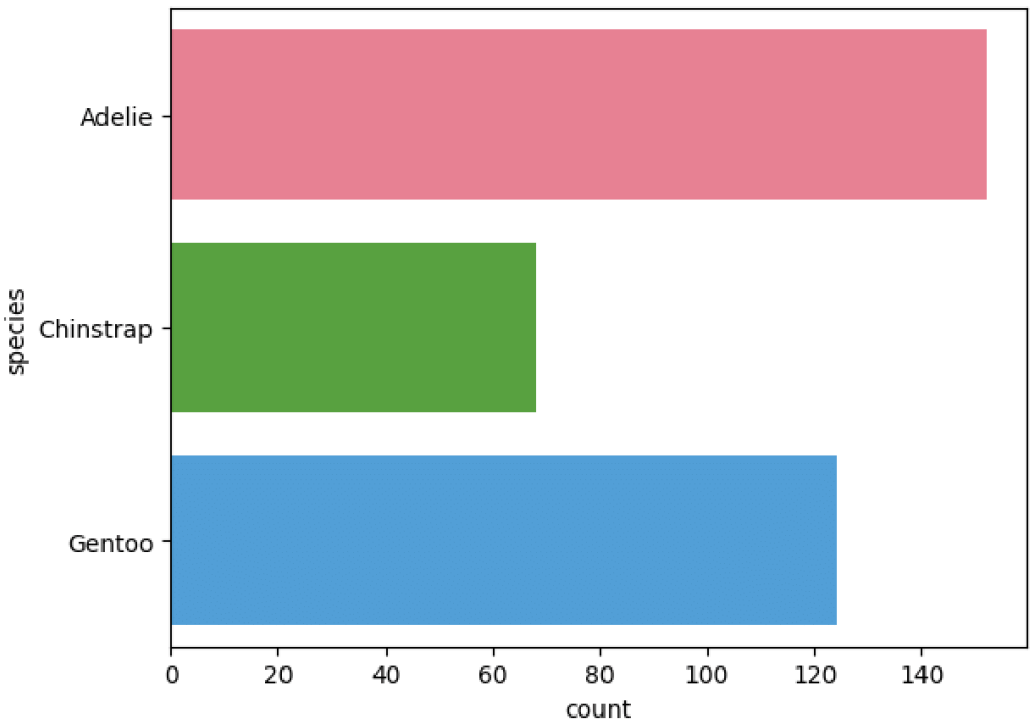
nts/TeX/fontdata.js

Sneha Mandal



In [23]:
```
sns.countplot(data=df, y='species', palette='husl')
```

Out[23]:
```
<Axes: xlabel='count', ylabel='species'>
```

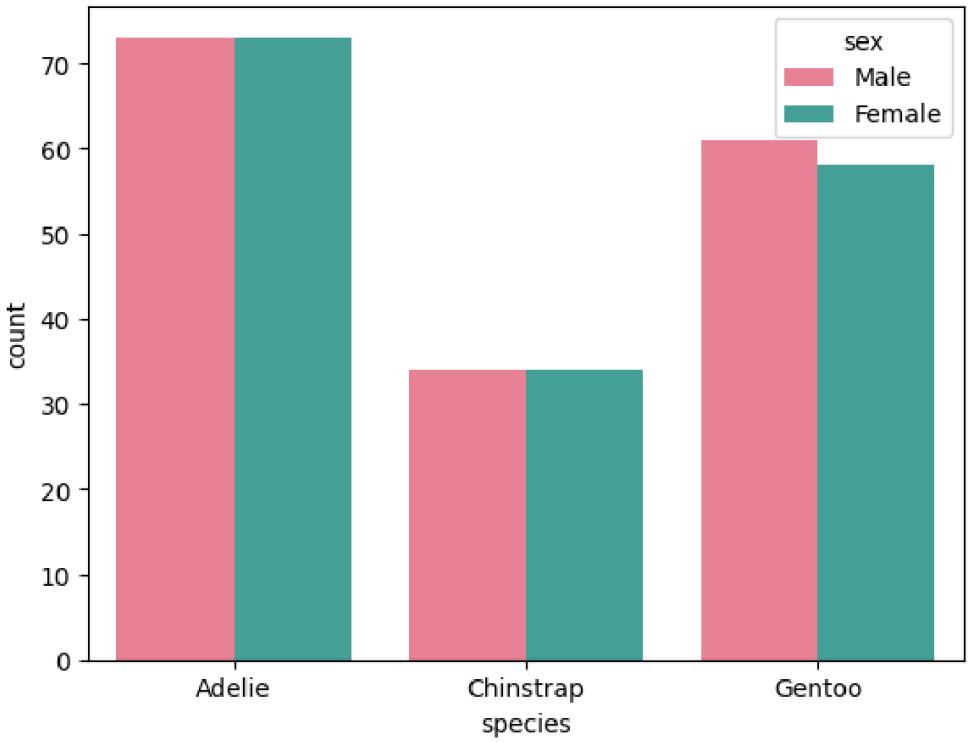


nts/TeX/fontdata.js

Sneha Mandal

In [24]:
```
sns.countplot(data=df, x='species', hue='sex', palette='husl')
```

Out[24]:
```
<Axes: xlabel='species', ylabel='count'>
```



Barplot

It represents an estimate of central tendency in form of bars

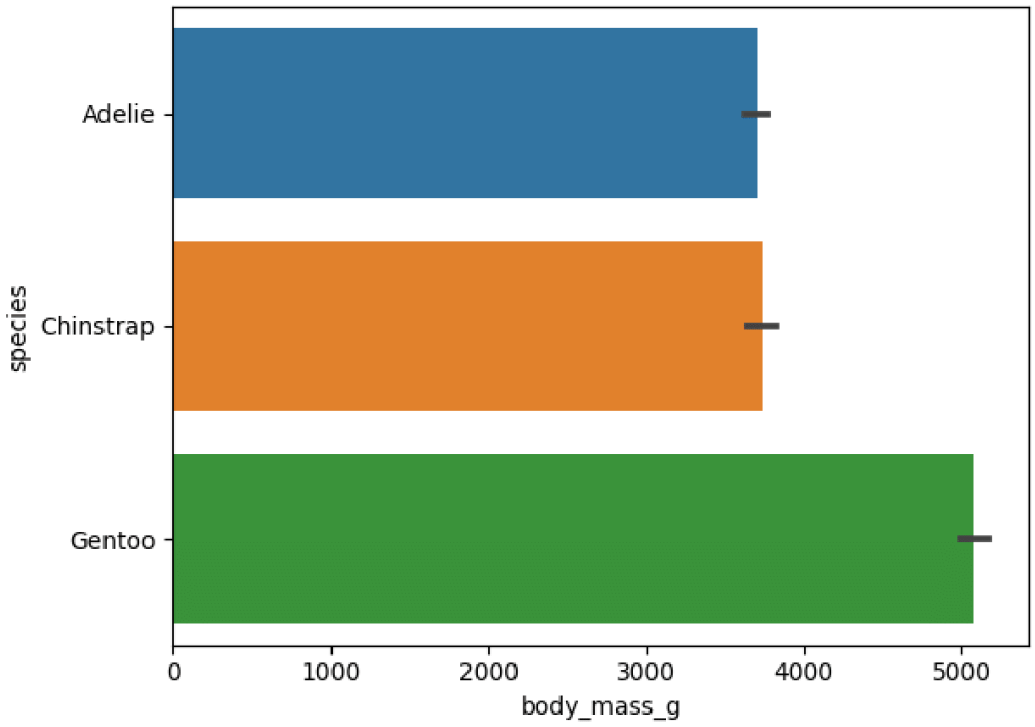
In [26]:
```
#barplot(data, x_axis, y_axis, hue, order, ...)

sns.barplot(x='body_mass_g', y='species', data=df)
```

Out[26]:
```
<Axes: xlabel='body_mass_g', ylabel='species'>
```

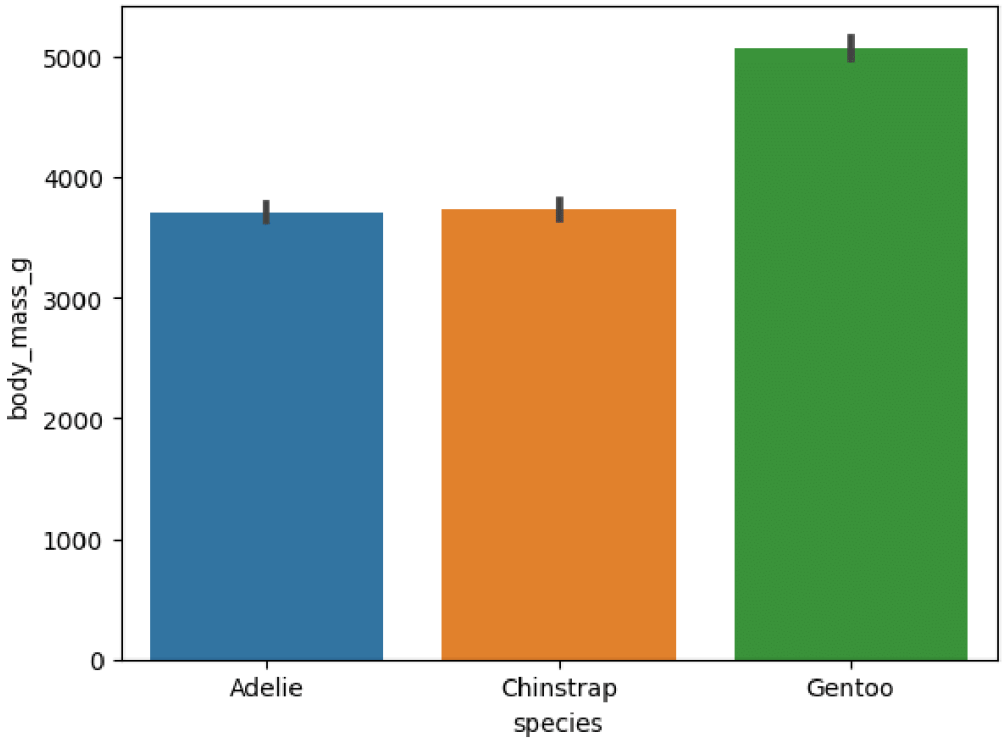
nts/TeX/fontdata.js

Sneha Mandal



In [27]:
```
sns.barplot(y='body_mass_g', x='species', data=df)
```

Out[27]:
```
<Axes: xlabel='species', ylabel='body_mass_g'>
```

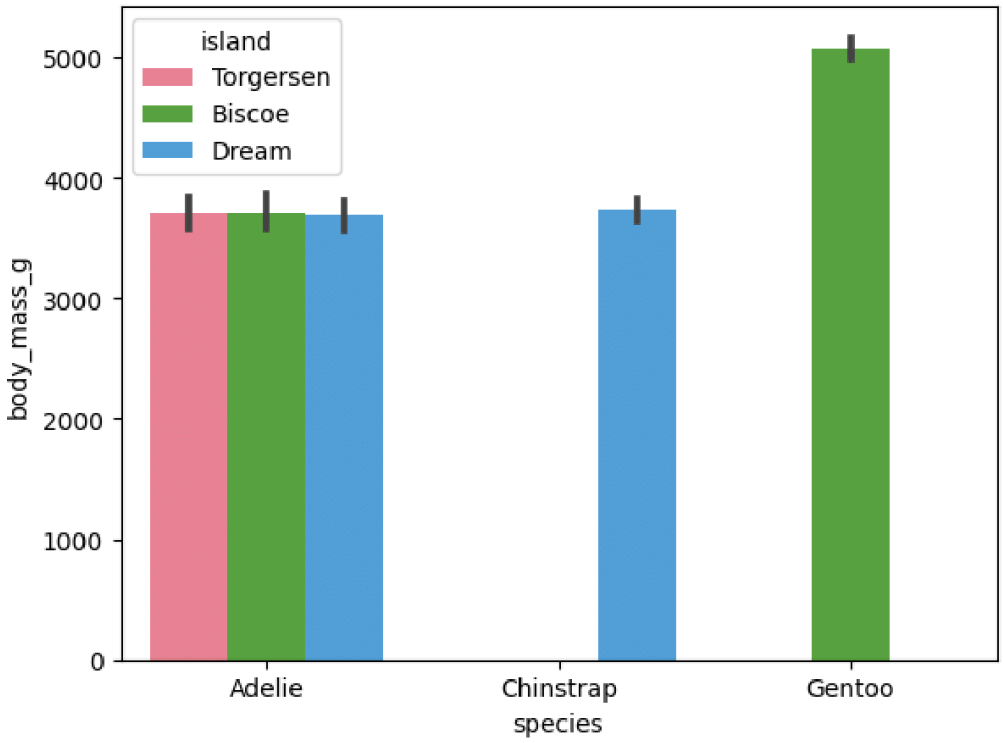


nts/TeX/fontdata.js

Sneha Mandal

In [29]:
```
sns.barplot(y='body_mass_g', x='species', data=df, hue='island', palette='hu
```

Out[29]:
```
<Axes: xlabel='species', ylabel='body_mass_g'>
```



Box plot

A box plot is a graph that presents information from a five-number summary i.e., lower_value, 25%, median, 75%, upper_value. It helps to find out the outliers in a dataframe.

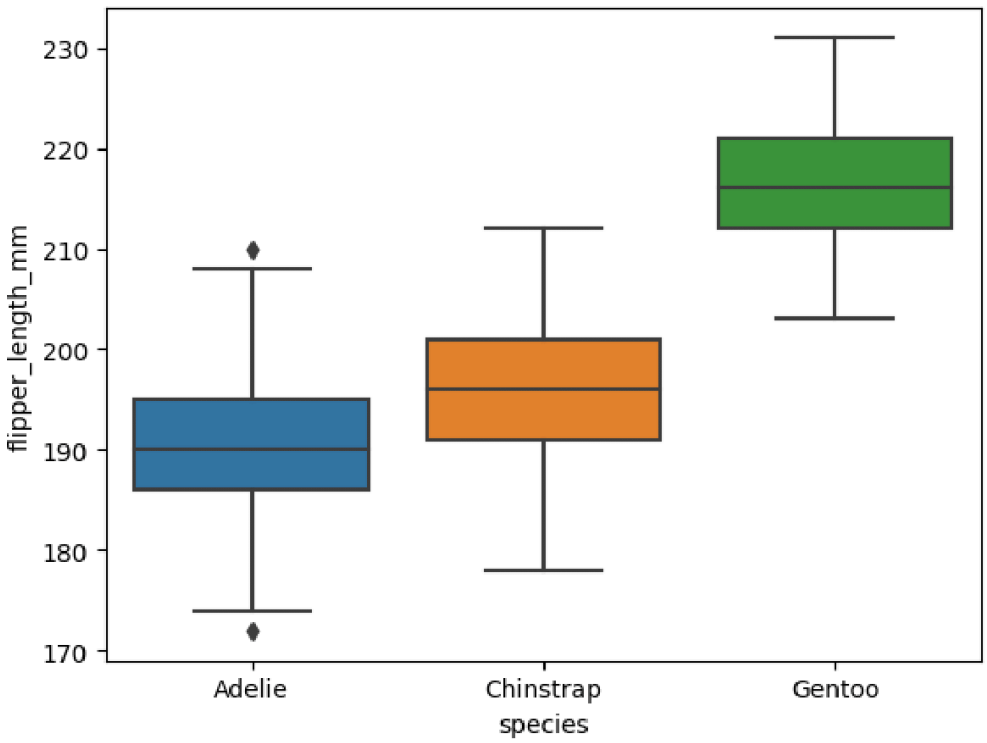
In [30]:
```
#boxplot(data, x_axis, y_axis, hue, order, orient, color, palette, ...)

sns.boxplot(x='species', y='flipper_length_mm', data=df)
```

Out[30]:
```
<Axes: xlabel='species', ylabel='flipper_length_mm'>
```

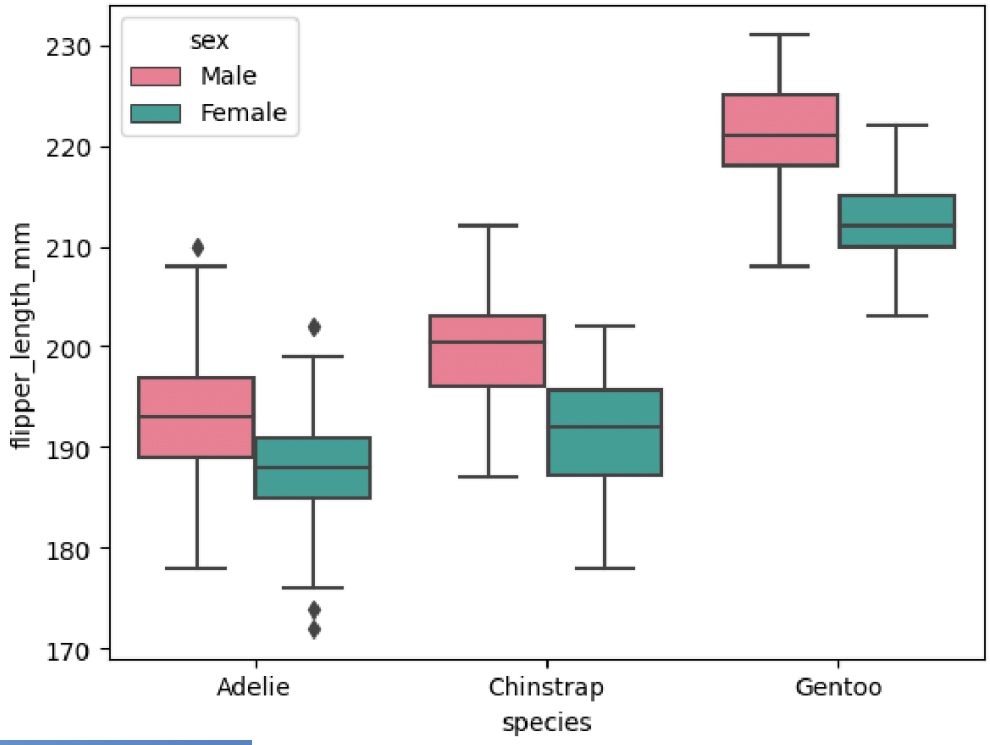
nts/TeX/fontdata.js

Sneha Mandal



In [32]:
```
sns.boxplot(x='species', y='flipper_length_mm', data=df, hue='sex', palette=
```

Out[32]:
```
<Axes: xlabel='species', ylabel='flipper_length_mm'>
```

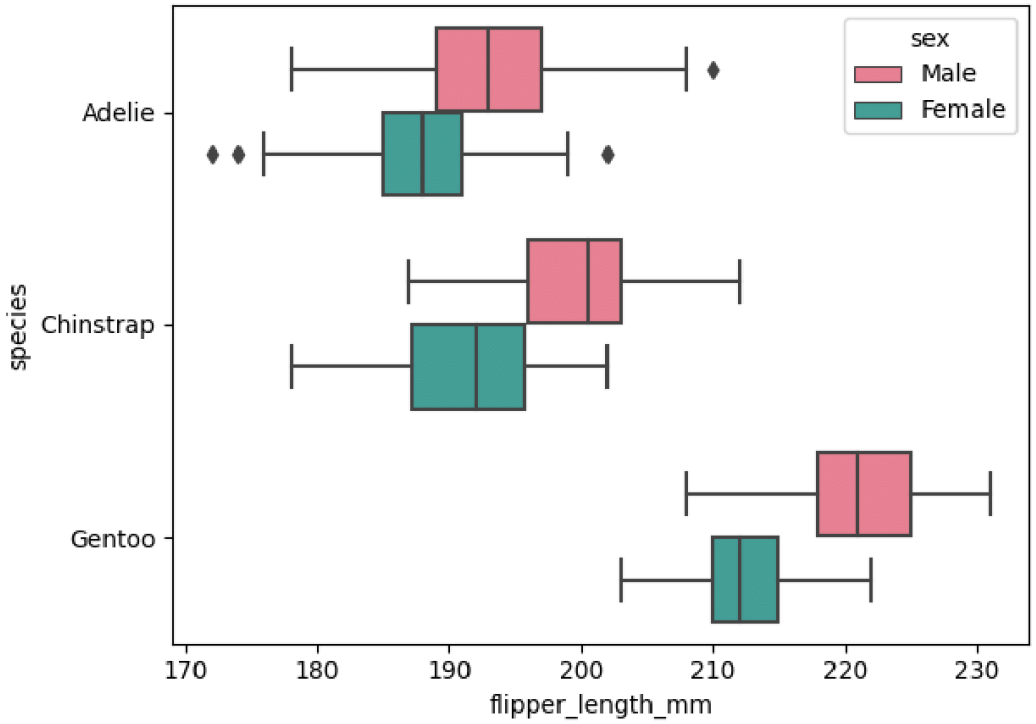


nts/TeX/fontdata.js

Sneha Mandal

In [33]:
```
sns.boxplot(y='species', x='flipper_length_mm', data=df, hue='sex', palette=
```

Out[33]:
```
<Axes: xlabel='flipper_length_mm', ylabel='species'>
```



Violin Plot

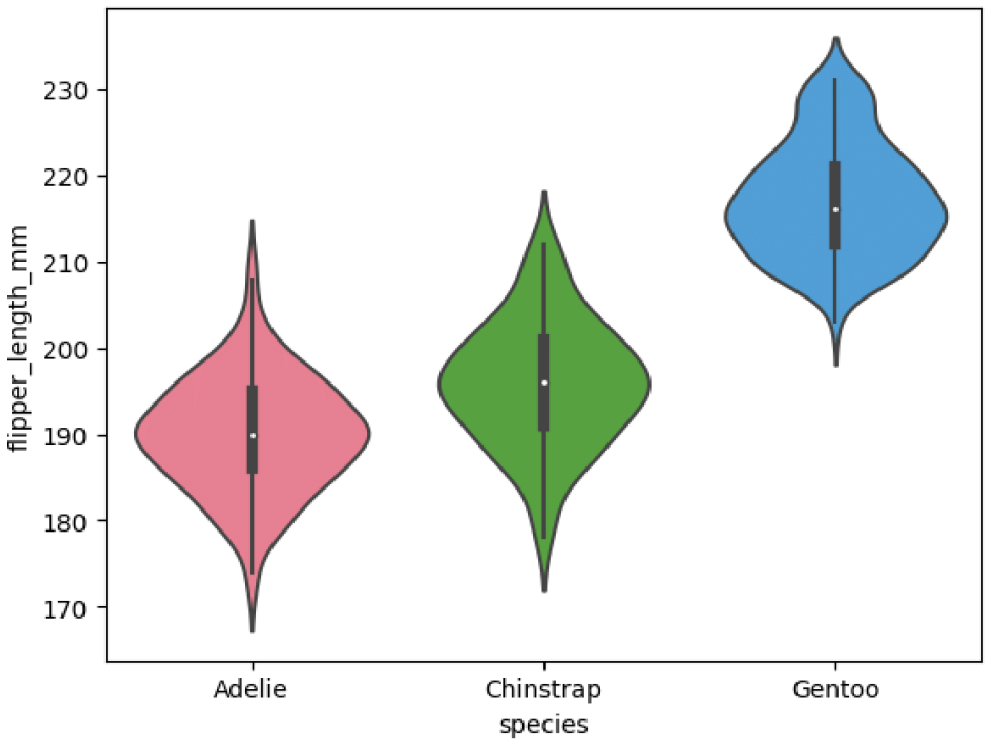
Violin plot helps us to see both the distribution of data in terms of Kernel density estimation and the box plot

In [35]:
```
sns.violinplot(x='species', y='flipper_length_mm', data=df, palette='husl' )
```

Out[35]:
```
<Axes: xlabel='species', ylabel='flipper_length_mm'>
```

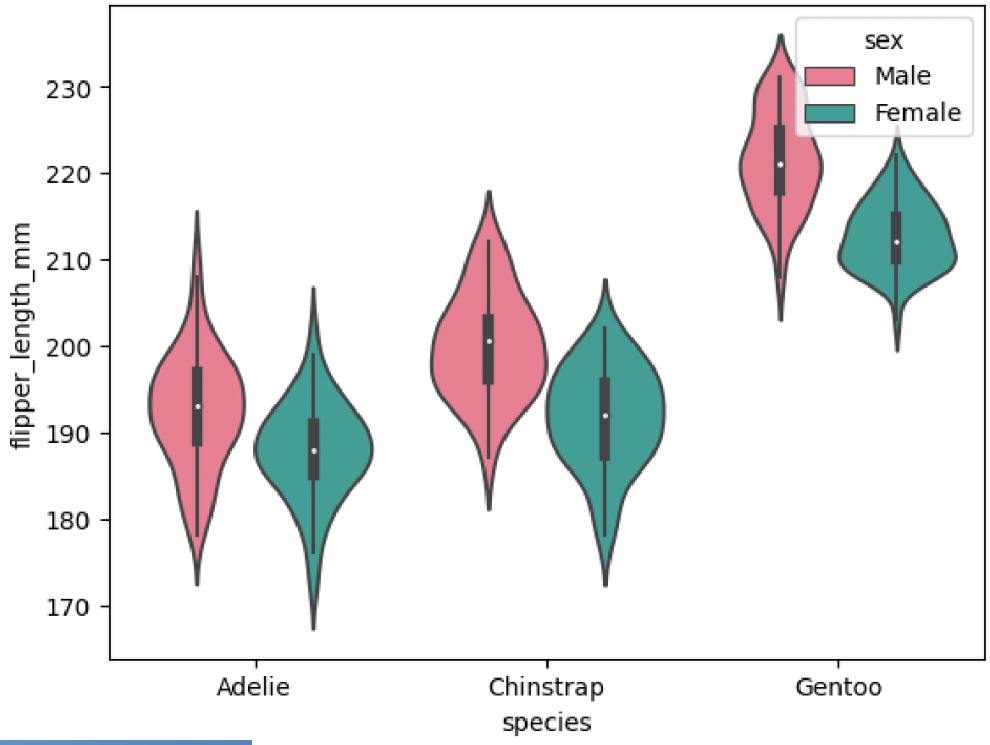
nts/TeX/fontdata.js

Sneha Mandal



In [36]:
```
sns.violinplot(x='species', y='flipper_length_mm', data=df, hue='sex', palet
```

Out[36]:
```
<Axes: xlabel='species', ylabel='flipper_length_mm'>
```



nts/TeX/fontdata.js

Sneha Mandal

What's your take from this???

Sneha Mandal