In [ ]:

```
#import libraries
```

In [28]:

```
import pandas as pd  #for dataset
import numpy as np   #for numerical operations
import seaborn as sbn #for visualization --> built on matplotlib
import matplotlib.pyplot as plt
```

In [73]:

```
df=sbn.load_dataset('penguins')  #to load one of the dataset present in seaborn
```

In [4]:

```
df
```

Out[4]:

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	s
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	M
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Fem
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Fem
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	N
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Fem
...	...	...	...	...	...	...	
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	N
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	Fem
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	M
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	Fem
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	M

344 rows × 7 columns

In [9]:

```
df.shape #represents total no. of records and columns in daatframe
```

Out[9]:

```
(344, 7)
```

In [10]:

```
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   bill_length_mm     342 non-null    float64
 3   bill_depth_mm      342 non-null    float64
 4   flipper_length_mm  342 non-null    float64
 5   body_mass_g        342 non-null    float64
 6   sex                333 non-null    object 
dtypes: float64(4), object(3)
memory usage: 14.8+ KB
```

In [5]:

```
df.isnull().sum()
```

Out[5]:

```
species              0
island               0
bill_length_mm       2
bill_depth_mm        2
flipper_length_mm    2
body_mass_g          2
sex                 11
dtype: int64
```

In [68]:

```
df.dropna()# null values are droped temporarily
```

Out[68]:

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	s
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	M
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Fem
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Fem
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Fem
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	M
...	...	...	...	...	...	...	
338	Gentoo	Biscoe	47.2	13.7	214.0	4925.0	Fem
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	Fem
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	M
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	Fem
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	M

333 rows × 7 columns

In [21]:

```
df.shape
```

Out[21]:

```
(344, 7)
```

In [74]:

```
df.dropna(inplace=True) #to drop the records with null values and inplace to make this chan
```

In [16]:

```
df.shape
```

Out[16]:

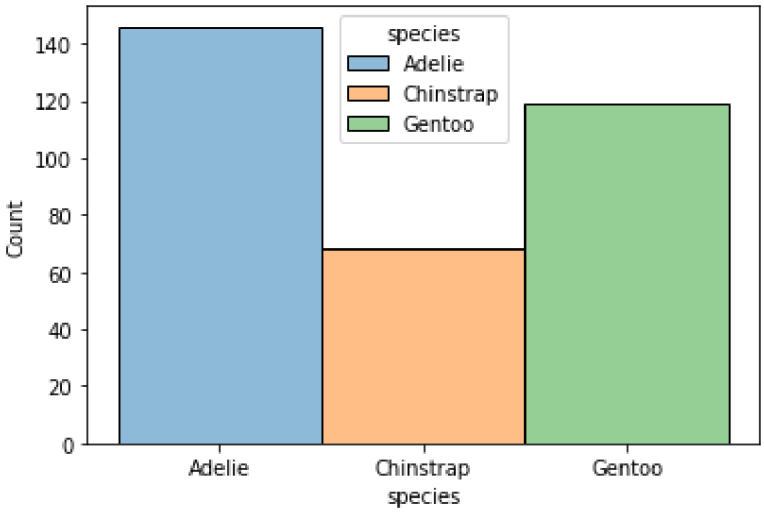
```
(333, 7)
```

In [27]:

```
sbn.histplot(data=df,x=df['species'],hue='species')
```

Out[27]:

```
<AxesSubplot:xlabel='species', ylabel='Count'>
```

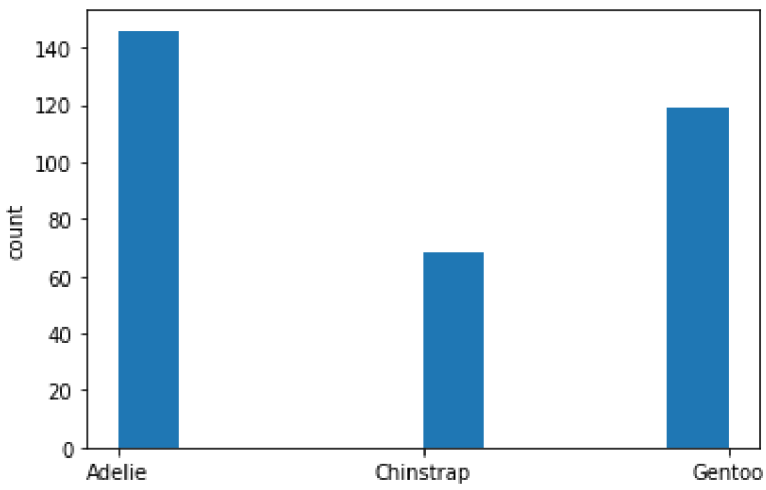


In [37]:

```
#histogram using matplotlib
plt.hist(df['species'])
plt.ylabel('count')
```

Out[37]:

```
Text(0, 0.5, 'count')
```

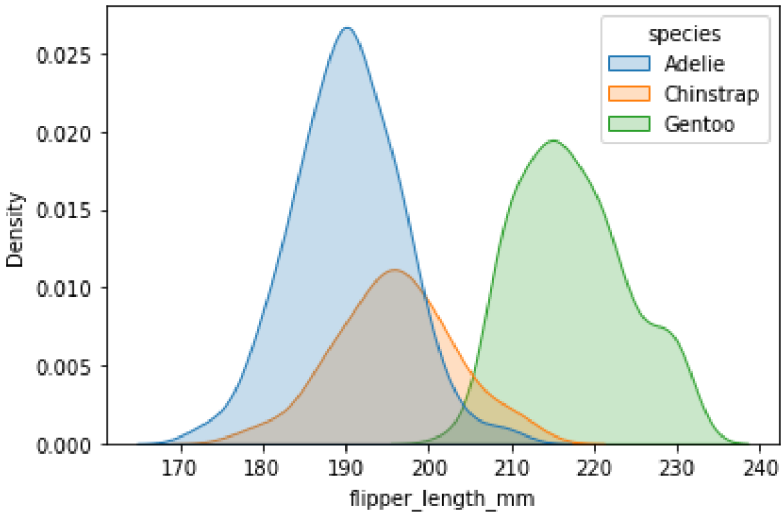


In [42]:

```
#kernal density estimation plot in seaborn
sbn.kdeplot(data=df,x=df['flipper_length_mm'],hue=df['species'],fill=True)
```

Out[42]:

```
<AxesSubplot:xlabel='flipper_length_mm', ylabel='Density'>
```



In [43]:

```
df['sex'].unique()
```

Out[43]:

```
array(['Male', 'Female'], dtype=object)
```

In [55]:

```
x=len(df)
```

In [75]:

```
for i in range(len(df)):
    if df.iloc[i,6]=='Male':
        df.iloc[i,6]=0
    else:
        df.iloc[i,6]=1
df
```

Out[75]:

	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	0
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	1
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	1
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	1
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	0
...	...	...	...	...	...	...	...
338	Gentoo	Biscoe	47.2	13.7	214.0	4925.0	1
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	1
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	0
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	1
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	0

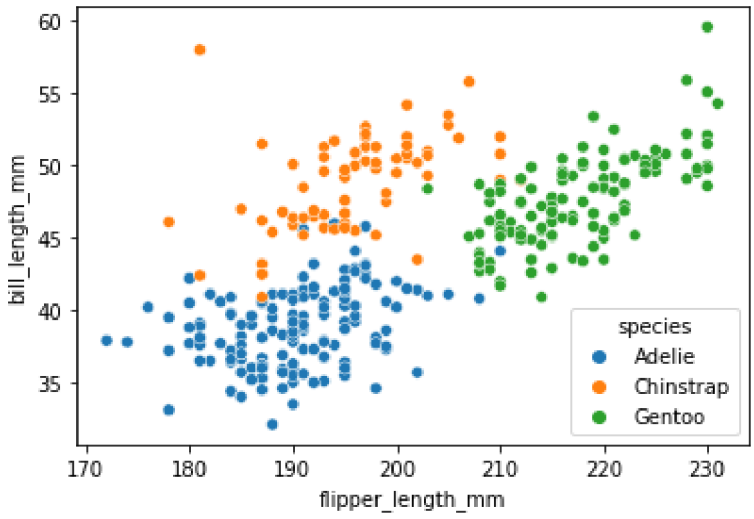
333 rows × 7 columns

In [93]:

```
sbn.scatterplot(data=df,x='flipper_length_mm',y='bill_length_mm',hue='species')
```

Out[93]:

```
<AxesSubplot:xlabel='flipper_length_mm', ylabel='bill_length_mm'>
```

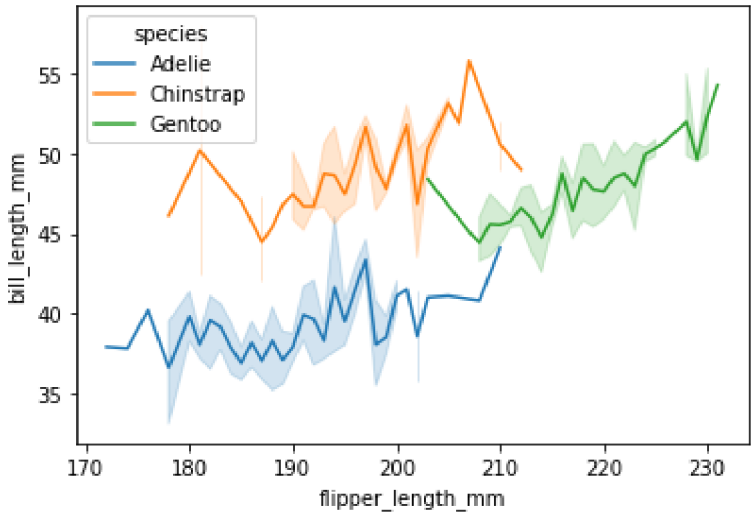


In [92]:

```
sbn.lineplot(data=df,x='flipper_length_mm',y='bill_length_mm',hue='species')
```

Out[92]:

```
<AxesSubplot:xlabel='flipper_length_mm', ylabel='bill_length_mm'>
```

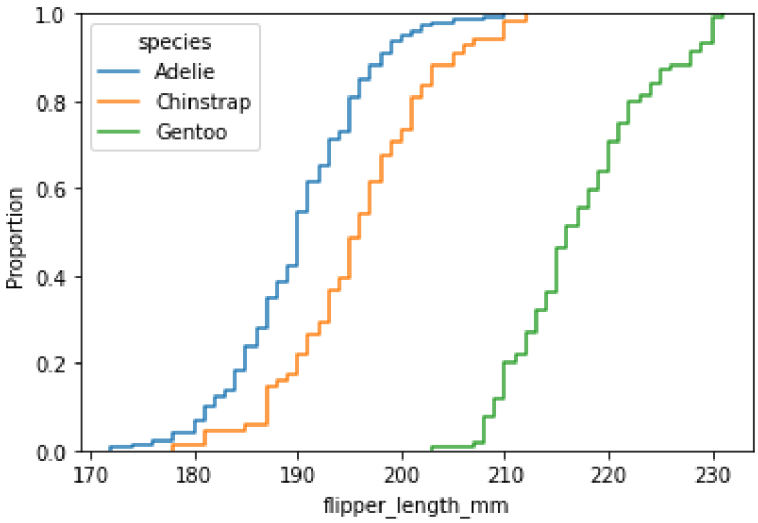


In [94]:

```
sbn.ecdfplot(data=df,x='flipper_length_mm',hue='species')
```

Out[94]:

```
<AxesSubplot:xlabel='flipper_length_mm', ylabel='Proportion'>
```

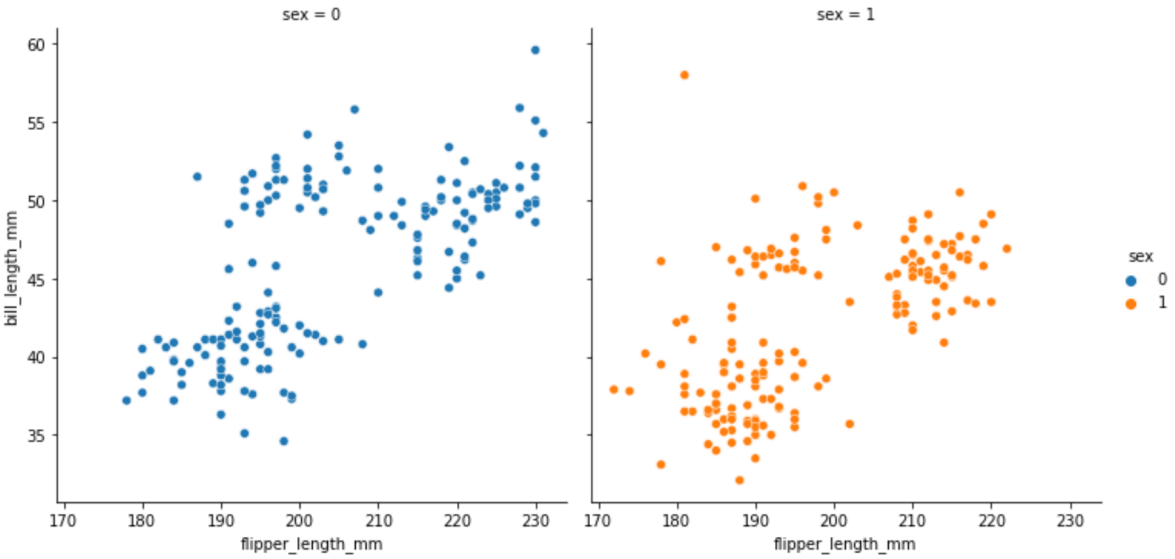


In [101]:

```
sbn.relplot(data=df,x='flipper_length_mm',y='bill_length_mm',col='sex',hue='sex')
```

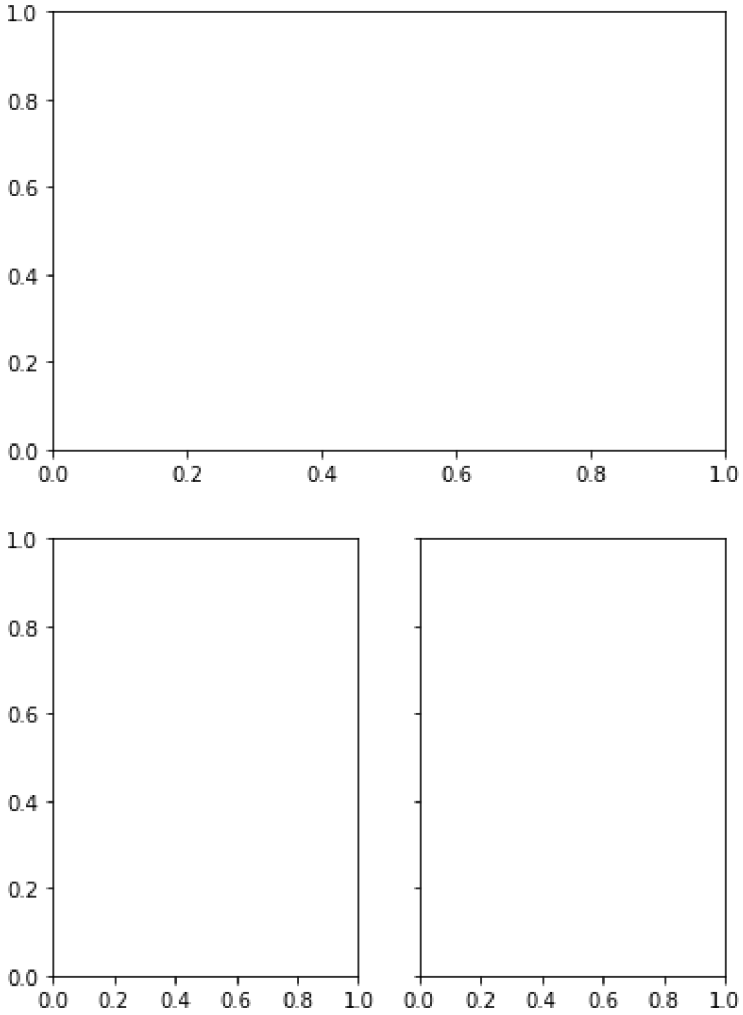
Out[101]:

```
<seaborn.axisgrid.FacetGrid at 0xc585310>
```



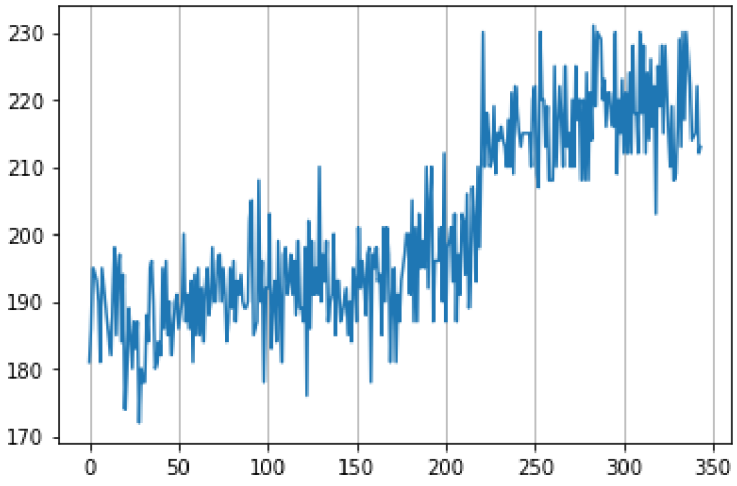
In [111]:

```
x,y=plt.subplots()
x,y=plt.subplots(1,2,sharey=True)
```



In [124]:

```
plt.plot(df['flipper_length_mm'])
plt.grid(axis='x')
```

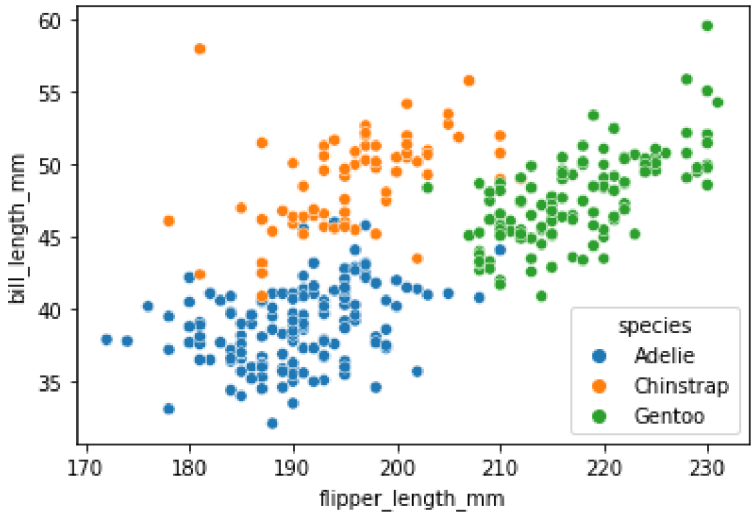


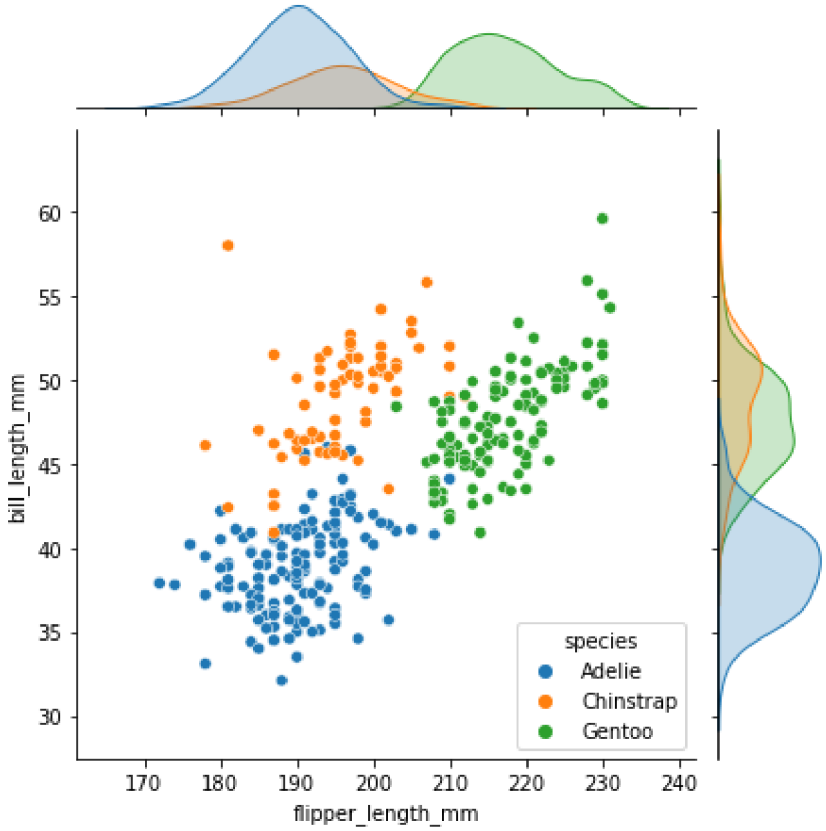
In [142]:

```
sbn.scatterplot(data=df,x='flipper_length_mm',y='bill_length_mm',hue='species')
sbn.jointplot(data=df,x='flipper_length_mm',y='bill_length_mm',hue='species')
```

Out[142]:

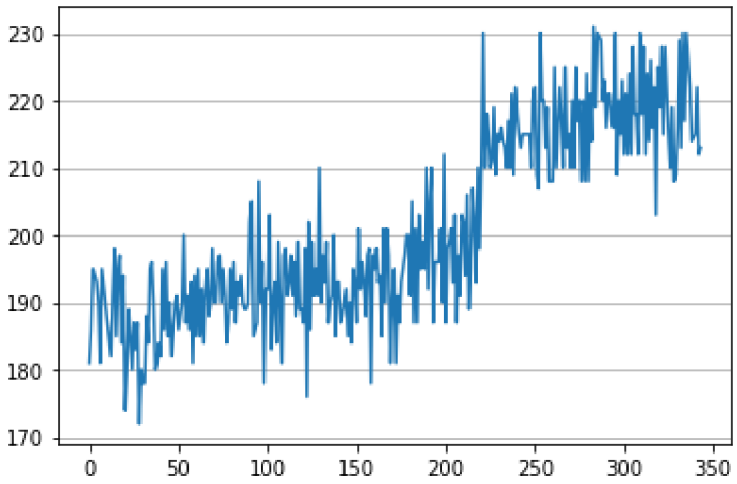
```
<seaborn.axisgrid.JointGrid at 0x10696a48>
```





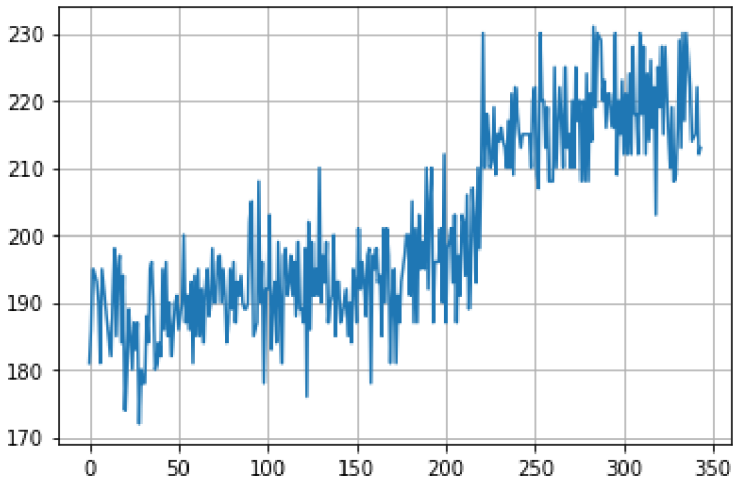
In [122]:

```
plt.plot(df['flipper_length_mm'])
plt.grid(axis='y')
```



In [123]:

```
plt.plot(df['flipper_length_mm'])
plt.grid()
```



In [146]:

```
df
```

Out[146]:

	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	0
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	1
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	1
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	1
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	0
...	...	...	...	...	...	...	...
338	Gentoo	Biscoe	47.2	13.7	214.0	4925.0	1
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	1
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	0
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	1
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	0

333 rows × 7 columns

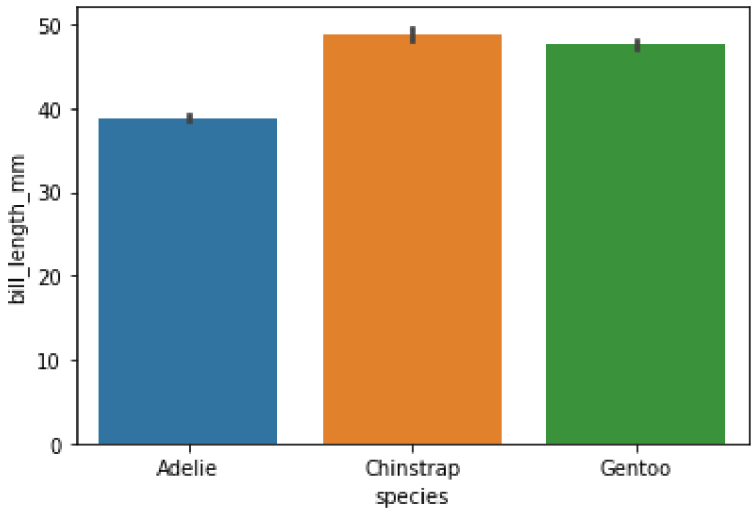
In [148]:

```
sbn.barplot(df['species'],df['bill_length_mm'])
```

```
C:\Users\Pratiksha\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Fu
tureWarning: Pass the following variables as keyword args: x, y. From versio
n 0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or misinterpre
tation.
  warnings.warn(
```

Out[148]:

```
<AxesSubplot:xlabel='species', ylabel='bill_length_mm'>
```

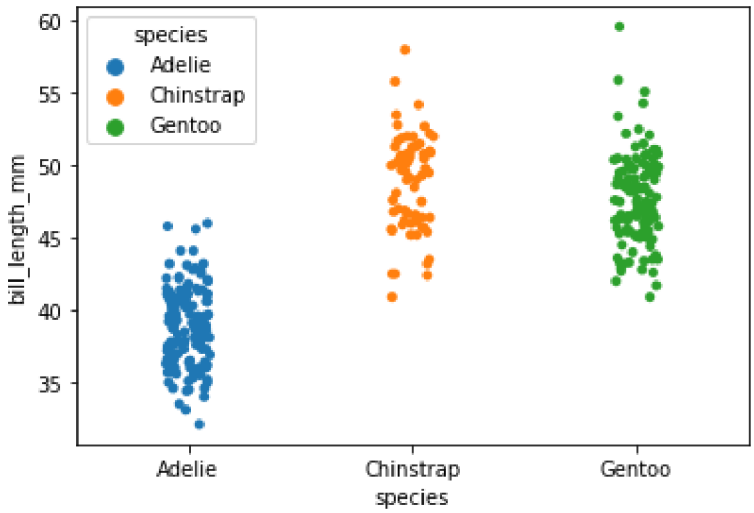


In [150]:

```
sbn.stripplot(df['species'],df['bill_length_mm'],hue=df['species'])
```

Out[150]:

```
<AxesSubplot:xlabel='species', ylabel='bill_length_mm'>
```

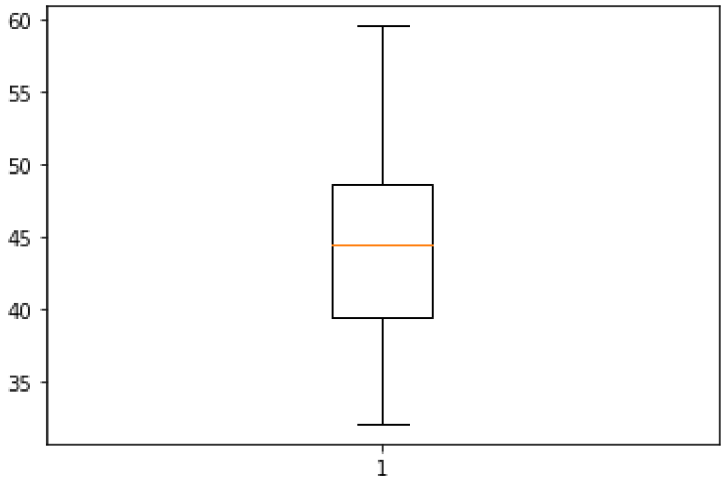


In [153]:

```
plt.boxplot(df['bill_length_mm'])
```

Out[153]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0xe012760>,
  <matplotlib.lines.Line2D at 0xe012bc8>],
 'caps': [<matplotlib.lines.Line2D at 0xe012208>,
  <matplotlib.lines.Line2D at 0xe012fd0>],
 'boxes': [<matplotlib.lines.Line2D at 0xe0127a8>],
 'medians': [<matplotlib.lines.Line2D at 0xe012490>],
 'fliers': [<matplotlib.lines.Line2D at 0xe012d30>],
 'means': []}
```



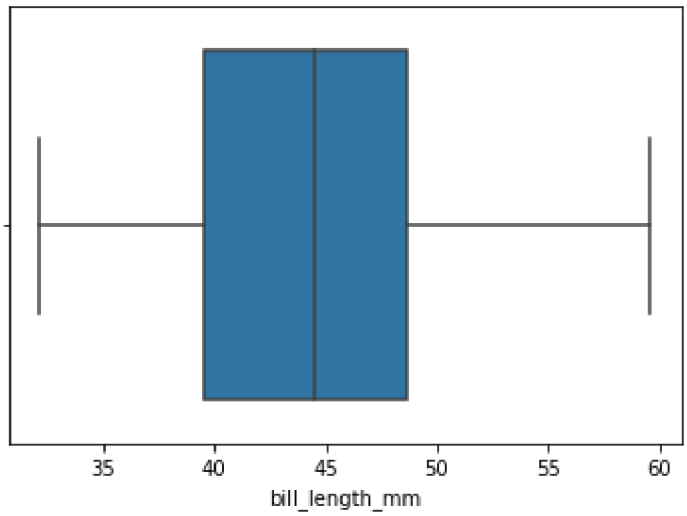
In [160]:

```
sbn.boxplot(df['bill_length_mm'])
```

```
C:\Users\Pratiksha\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Fu
tureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other a
rguments without an explicit keyword will result in an error or misinterpret
ation.
  warnings.warn(
```

Out[160]:

```
<AxesSubplot:xlabel='bill_length_mm'>
```

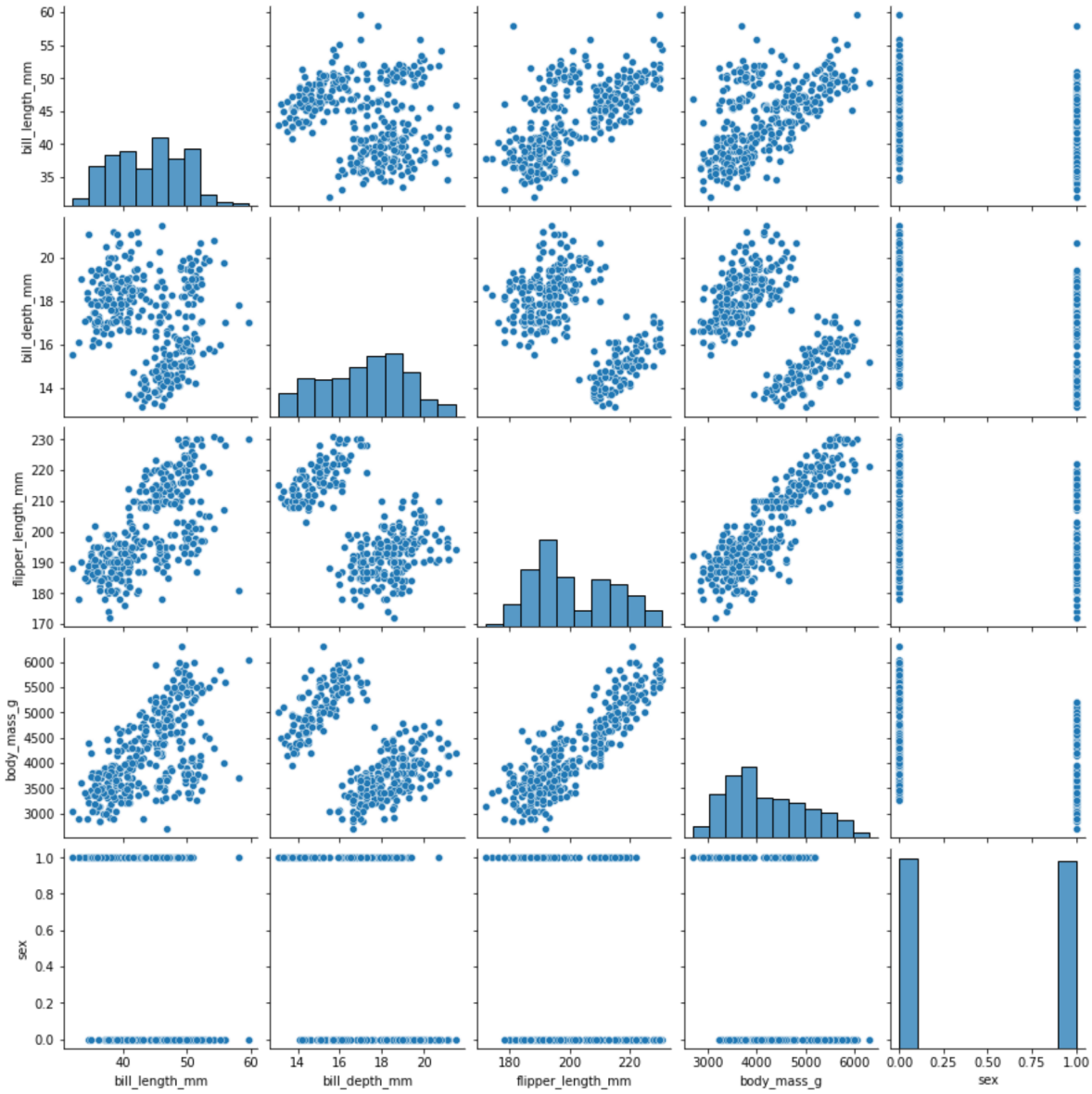


In [162]:

```
sbn.pairplot(df,)
```

Out[162]:

```
<seaborn.axisgrid.PairGrid at 0xf29e5c8>
```



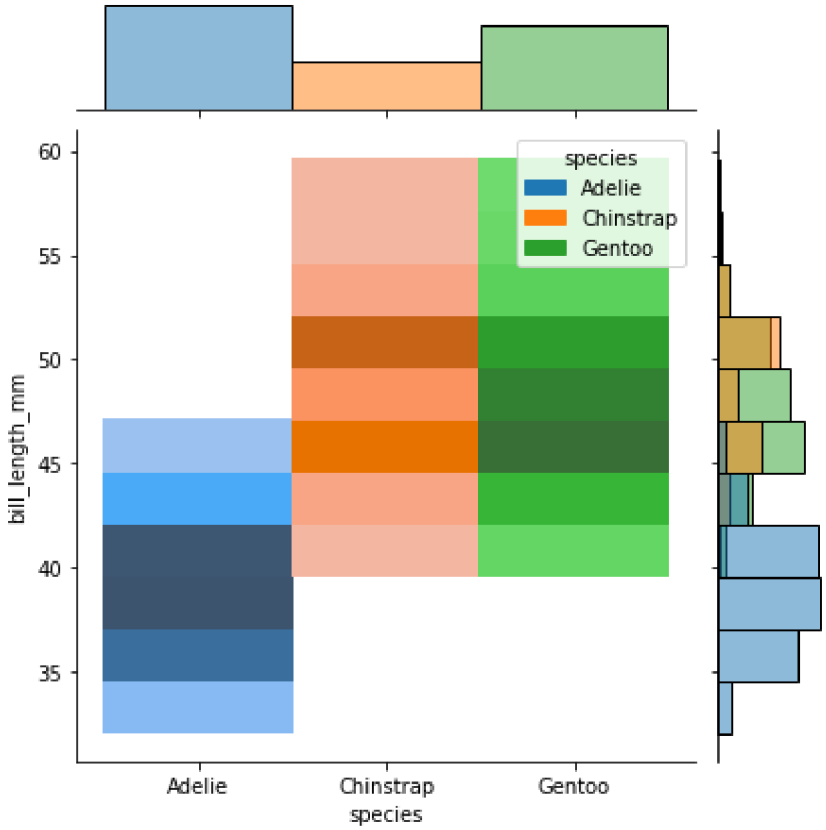
In [163]:

```
sbn.jointplot(df['species'],df['bill_length_mm'],hue=df['species'],kind='hist')
```

```
C:\Users\Pratiksha\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Fu
tureWarning: Pass the following variables as keyword args: x, y. From versio
n 0.12, the only valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error or misinterpre
tation.
  warnings.warn(
```

Out[163]:

```
<seaborn.axisgrid.JointGrid at 0x11910628>
```



In [164]:

```
sbn.swarmplot(df['bill_length_mm'])
```

```
C:\Users\Pratiksha\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Fu
tureWarning: Pass the following variable as a keyword arg: x. From version
0.12, the only valid positional argument will be `data`, and passing other a
rguments without an explicit keyword will result in an error or misinterpret
ation.
  warnings.warn(
```

Out[164]:

```
<AxesSubplot:xlabel='bill_length_mm'>
```