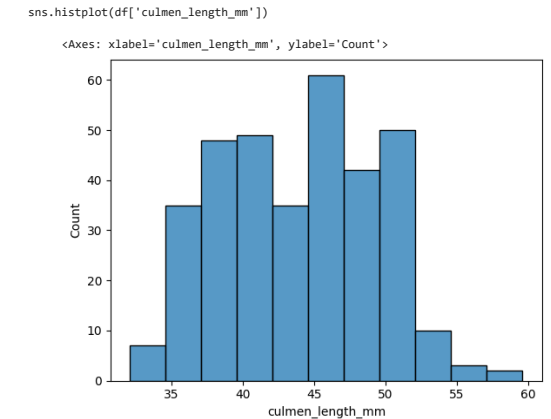
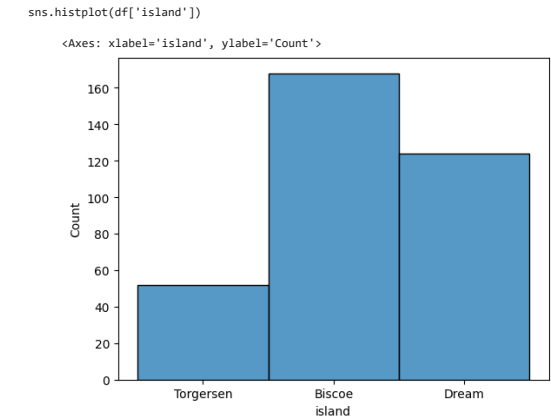
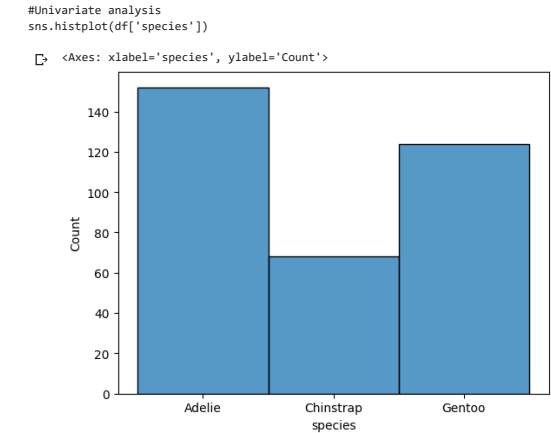


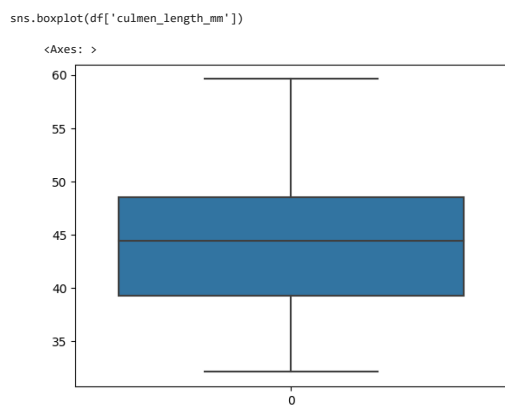
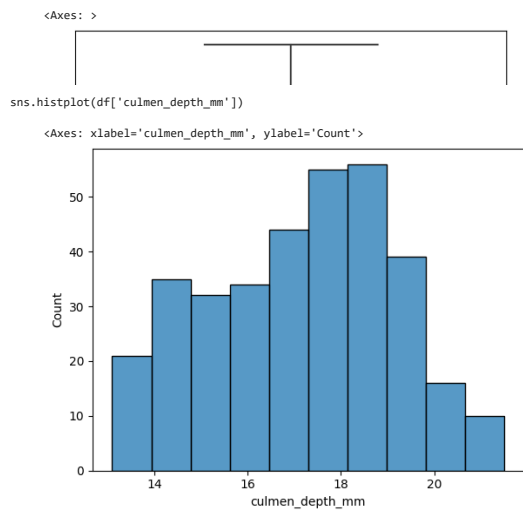
```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns
import numpy as np
from sklearn.preprocessing import LabelEncoder

df=pd.read_csv('penguins_size.csv')
df.head()
```

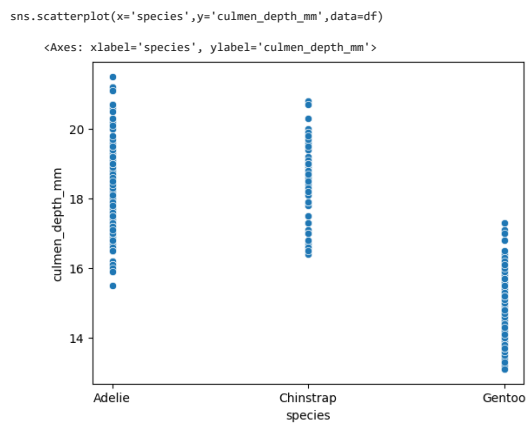
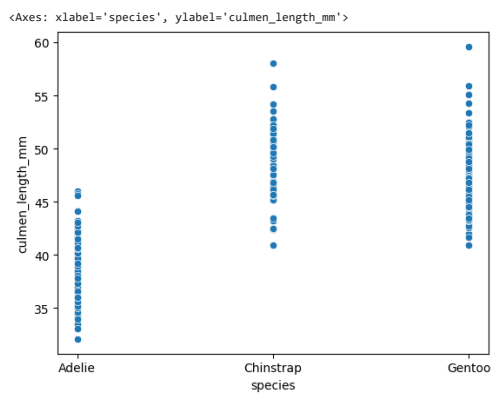
	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE



```
sns.boxplot(df['culmen_depth_mm'])
```



```
#bivariate analysis
sns.scatterplot(x='species', y='culmen_length_mm', data=df)
```



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

```
<Axes: xlabel='species', ylabel='flipper_length_mm'>
230
220
210
200
190
flipper_length_mm
sns.scatterplot(x='species',y='body_mass_g',data=df)
<Axes: xlabel='species', ylabel='body_mass_g'>
6000
5500
5000
4500
4000
3500
3000
body_mass_g
Adelie Chinstrap species Gentoo

#Multivariate analysis
sns.heatmap(df.corr(),annot=True)

<ipython-input-40-488cf15a4e58>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future versio
sns.heatmap(df.corr(),annot=True)
<Axes: >
culmen_length_mm 1 -0.24 0.66 0.6
culmen_depth_mm -0.24 1 -0.58 -0.47
flipper_length_mm 0.66 -0.58 1 0.87
body_mass_g 0.6 -0.47 0.87 1
culmen_length_mm culmen_depth_mm flipper_length_mm body_mass_g

df.describe()

culmen_length_mm culmen_depth_mm flipper_length_mm body_mass_g
count 342.000000 342.000000 342.000000 342.000000
mean 43.921930 17.151170 200.915205 4201.754386
std 5.459584 1.974793 14.061714 801.954536
min 32.100000 13.100000 172.000000 2700.000000
25% 39.225000 15.600000 190.000000 3550.000000
50% 44.450000 17.300000 197.000000 4050.000000
75% 48.500000 18.700000 213.000000 4750.000000
max 59.600000 21.500000 231.000000 6300.000000

df.isnull().any()

species False
island False
culmen_length_mm True
culmen_depth_mm True
flipper_length_mm True
body_mass_g True
sex True
dtype: bool

df['culmen_length_mm']=df['culmen_length_mm'].fillna(df['culmen_length_mm'].median())
df['culmen_depth_mm']=df['culmen_depth_mm'].fillna(df['culmen_depth_mm'].median())
df['flipper_length_mm']=df['flipper_length_mm'].fillna(df['flipper_length_mm'].median())
df['body_mass_g']=df['body_mass_g'].fillna(df['body_mass_g'].median())
df['sex']=df['sex'].fillna(df['sex'].mode())

df.isnull().any()

species False
island False
culmen_length_mm False
culmen_depth_mm False
flipper_length_mm False
body_mass_g False
sex False
dtype: bool
```

```
le=LabelEncoder()
df.sex=le.fit_transform(df.sex)
df.species=le.fit_transform(df.species)
df.island=le.fit_transform(df.island)

print(df['species'])
```

0	0
1	0
2	0
3	0
4	0
...	
339	2
340	2
341	2
342	2
343	2

Name: species, Length: 344, dtype: int64