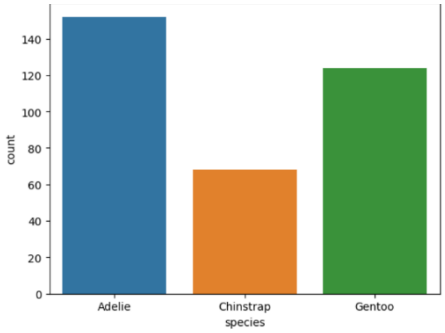
Univariate data analysis

Numerical Data

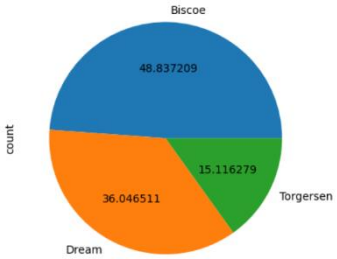
1. **Countplot**

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



2. **Piechart**

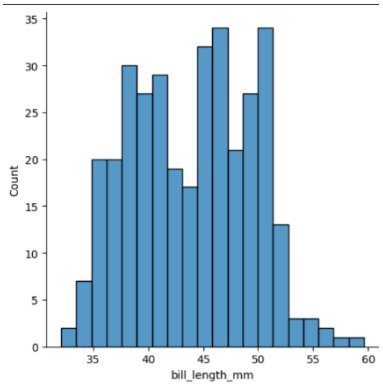
   df['island'].value_counts().plot(kind='pie', autopct='%2f')



Categorical Data

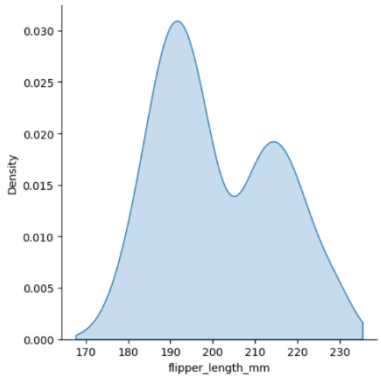
1. **Histogram**

   sns.displot(df, x="bill_length_mm", bins=20)



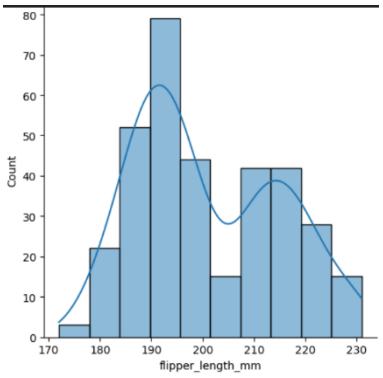
2. **KDE plot**

   sns.displot(penguins, x="flipper_length_mm", kind="kde", fill=True, cut = True)



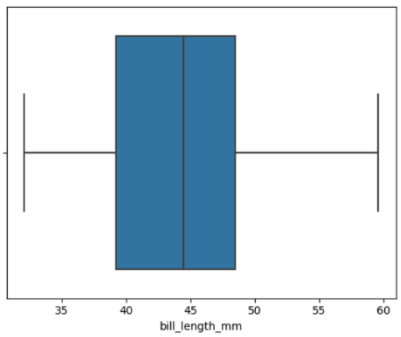
3. **KDE + histogram**

   sns.displot(penguins, x="flipper_length_mm", kde=True)



4. **Boxplot**

   sns.boxplot(data = df, x = 'bill_length_mm')



Bivariate data analysis

Numerical - Numerical Data

3. Scatterplot

Numerical – Numerical

```
sns.scatterplot(data = df, x = 'bill_length_mm', y = 'body_mass_g')
```

Numerical - Numerical – numerical

```
sns.scatterplot(data = df, x = 'bill_length_mm', y = 'bill_depth_mm',
size='body_mass_g')
```

Numerical - Numerical – categorical

```
sns.scatterplot(data = df, x = 'bill_length_mm', y = 'bill_depth_mm', hue='sex')
```

Numerical - Numerical – categorical – numerical

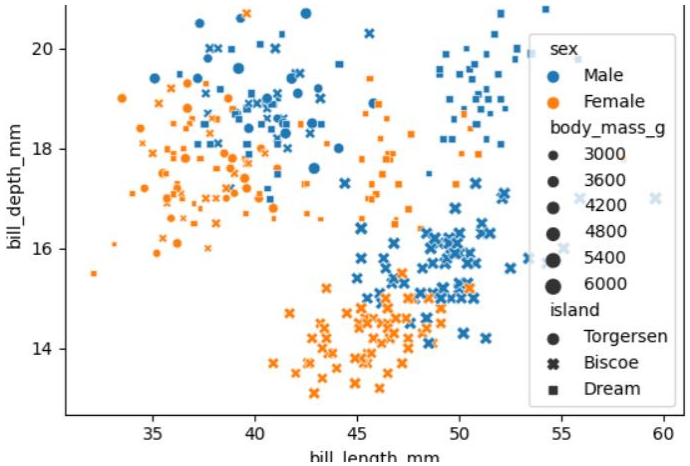
```
sns.scatterplot(data = df, x = 'bill_length_mm', y = 'bill_depth_mm', hue='sex',
size='body_mass_g')
```

Numerical - Numerical – categorical – categorical

```
sns.scatterplot(data = df, x = 'bill_length_mm', y = 'bill_depth_mm', hue='sex',
style='island')
```

Numerical - Numerical – categorical – categorical – numerical

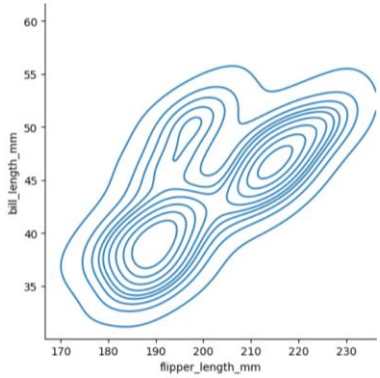
```
sns.scatterplot(data = df, x = 'bill_length_mm', y = 'bill_depth_mm', hue='sex',
style='island',  size='body_mass
```



4. Density Graph

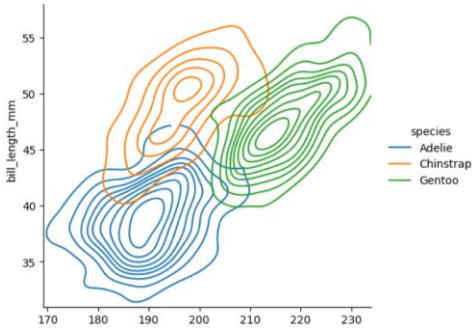
Numerical – Numerical

sns.displot(penguins, x="flipper_length_mm", y='bill_length_mm',kind="kde", cut = True)



Numerical - Numerical – categorical

sns.displot(penguins, x="flipper_length_mm", y='bill_length_mm', hue='species', kind="kde", cut = True)

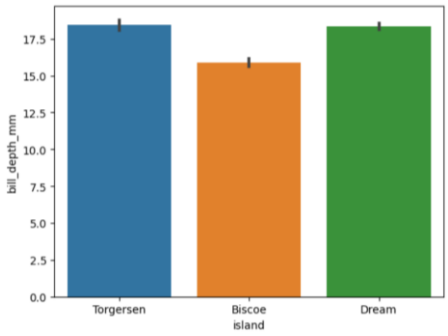


Numerical – Categorical Data

1. Bar Plot

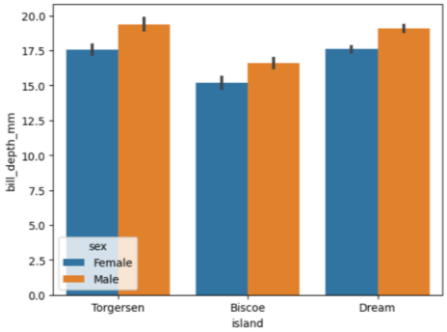
Numerical – Categorical

sns.barplot(data = df, x = 'island', y = 'bill_depth_mm')



Numerical – categorical – Categorical

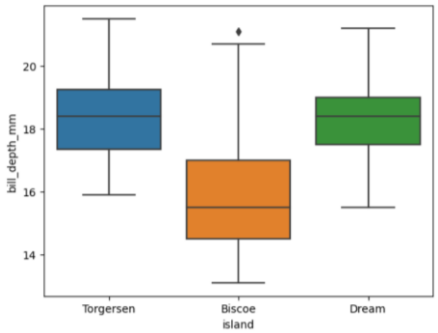
sns.barplot(data = df, x = 'island', y = 'bill_depth_mm', hue='sex', hue_order=['Female', 'Male'])



2. Box Plot

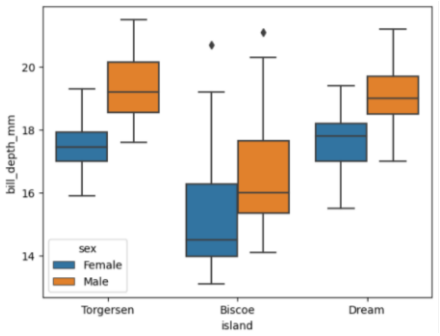
Numerical – Categorical

sns.boxplot(data = df, x = 'island', y = 'bill_depth_mm')



Numerical – categorical – Categorical

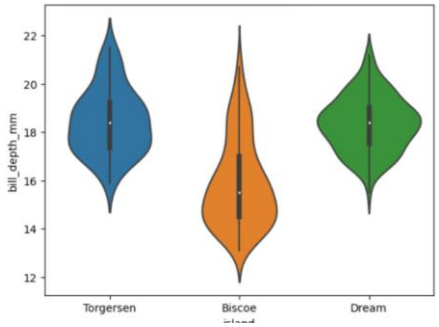
sns.boxplot (data = df, x = 'island', y = 'bill_depth_mm', hue='sex', hue_order=['Female', 'Male'])



3. Violin Plot

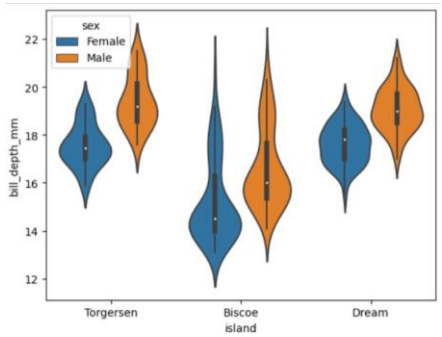
Numerical – Categorical

sns.violinplot(data = df, x = 'island', y = 'bill_depth_mm')



Numerical – Categorical – Categorical

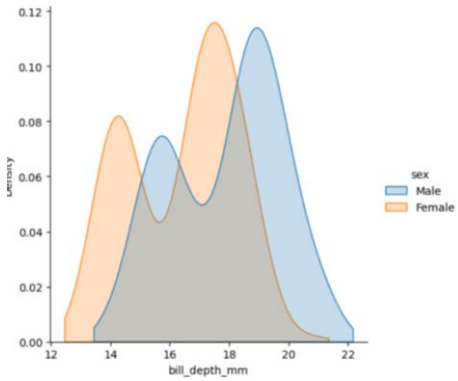
sns.violinplot (data = df, x = 'island', y = 'bill_depth_mm', hue='sex', hue_order=['Female', 'Male'])



4. Distribution Graph

Numerical – Categorical

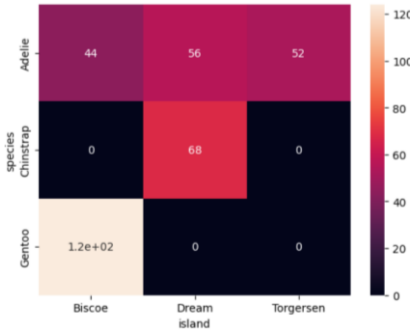
sns.displot(data= df, x="bill_depth_mm", hue='sex', kind="kde", fill=True, cut = True)


0.12
0.10
0.08
0.06
0.04
0.02
0.00
density
12
14
16
18
20
22
bill_depth_mm
sex
Male
Female

Categorical – Categorical Data

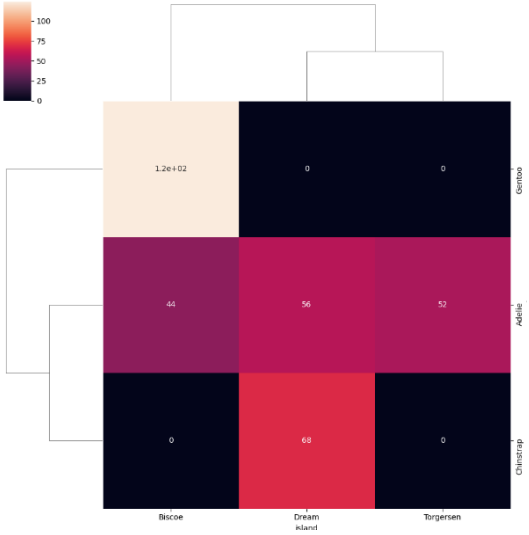
1. **Heatmap**

   sns.heatmap(pd.crosstab(df['species'], df['island']), annot=True)



2. **Clustermap**

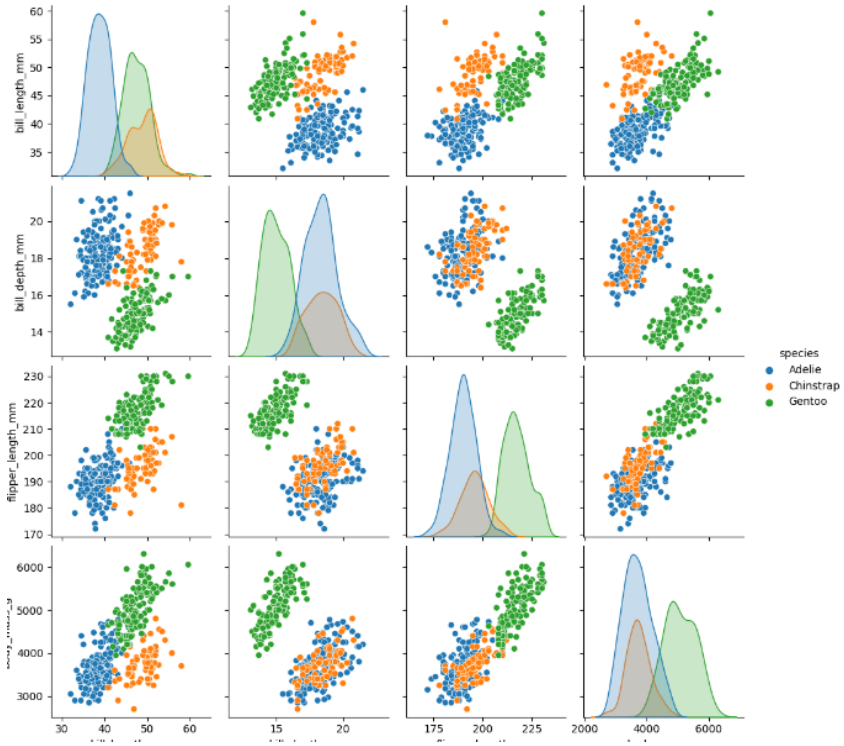
   sns.clustermap(pd.crosstab(df['species'], df['island']), annot=True)



EDA ON GROUP OF VARIABLES

Pairplot

sns.pairplot(df, hue='species')



Heatmap

sns.heatmap(df.select_dtypes(include=[np.number]).corr(), annot=True, fmt=".2f")