[5]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[6]:
```
df = pd.read_csv("Heart.csv")
df.head()
```

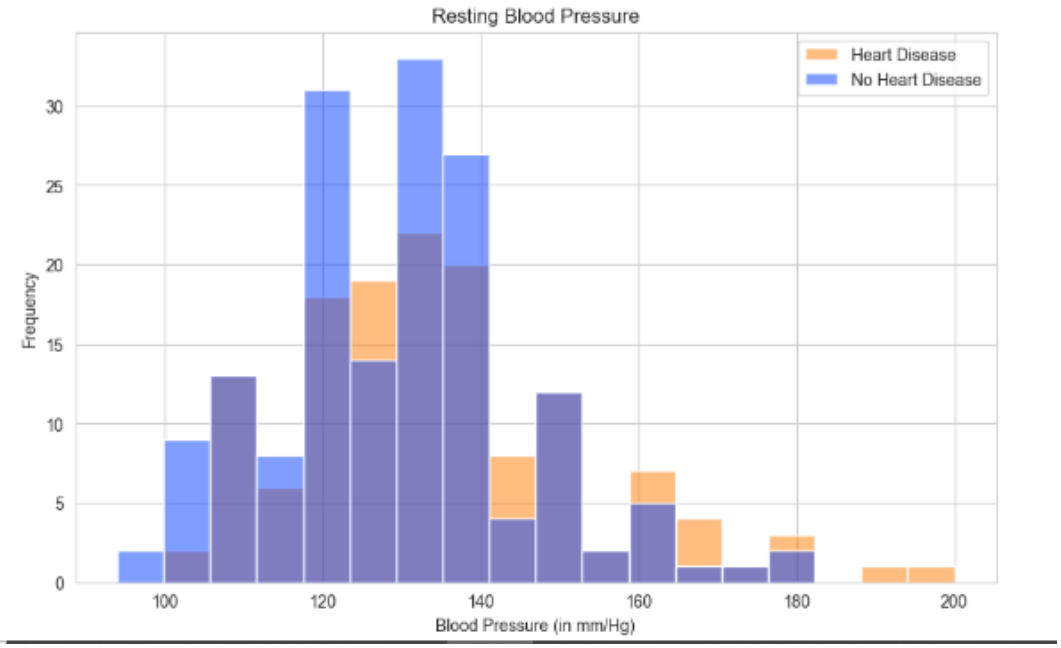
[6]:

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	1	63	1	typical	145	233	1	2	150	0	2.3	3	0.0	fixed	No
1	2	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3.0	normal	Yes
2	3	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2.0	reversable	Yes
3	4	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0.0	normal	No
4	5	41	0	nontypical	130	204	0	2	172	0	1.4	1	0.0	normal	No

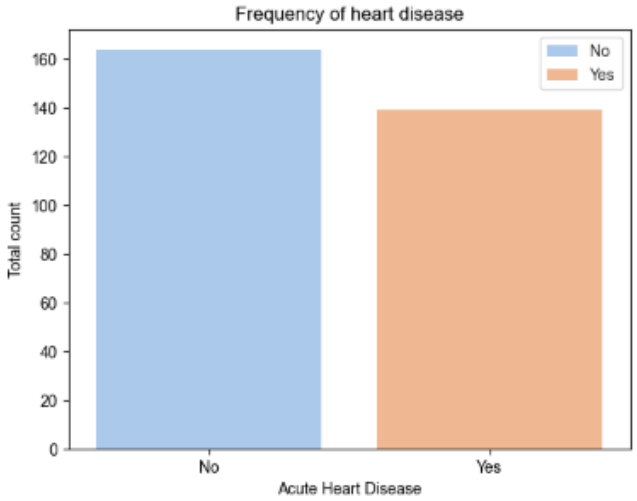
[7]:
```
df.shape
```

[7]:
```
(303, 15)
```

[21]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="RestBP" , data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Resting Blood Pressure")
plt.xlabel("Blood Pressure (in mm/Hg)")
plt.ylabel("Frequency")
plt.legend(["Heart Disease","No Heart Disease"])
plt.show()
```



[17]: <matplotlib.legend.Legend at 0x196492b4860>



[18]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="Age", data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Effect of age on frequency of heart disease")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend(["Heart Disease","No Heart Disease"])
plt.show()
```

[12]:
```
mean_age =df['Age'].mean()
mean_age
```

[12]:
```
54.43894389438944
```

[13]:
```
df.columns
```

[13]:
```
Index(['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs',
       'RestECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
      dtype='object')
```

[15]:
```
df2 = df.filter(['Age','Sex','ChestPain','RestBP','Chol'])
df2
```

[15]:

	Age	Sex	ChestPain	RestBP	Chol
0	63	1	typical	145	233
1	67	1	asymptomatic	160	286
2	67	1	asymptomatic	120	229
3	37	1	nonanginal	130	250
4	41	0	nontypical	130	204
...	...	...	...	...	...
298	45	1	typical	110	264
299	68	1	asymptomatic	144	193
300	57	1	asymptomatic	130	131
301	57	0	nontypical	130	236
302	38	1	nonanginal	138	175

303 rows × 5 columns

[10]:
```
df.dtypes
```

[10]:
```
Unnamed: 0      int64
Age             int64
Sex             int64
ChestPain      object
RestBP          int64
Chol            int64
Fbs             int64
RestECG         int64
MaxHR           int64
ExAng           int64
Oldpeak       float64
Slope           int64
Ca            float64
Thal           object
AHD            object
dtype: object
```

[11]:
```
(df == 0).sum(axis=0)
```

[11]:
```
Unnamed: 0      0
Age             0
Sex            97
ChestPain       0
RestBP          0
Chol            0
Fbs           258
RestECG       151
MaxHR           0
ExAng         204
Oldpeak        99
Slope           0
Ca            176
Thal            0
AHD             0
dtype: int64
```

[25]:
```
X = df[['Age','Sex','ChestPain','RestBP','Chol','RestECG','MaxHR']]
Y= df['AHD']
```

[27]:
```
from sklearn.model_selection import train_test_split

X_train, Y_train, X_test, Y_test =train_test_split(X,Y,test_size=0.25)
```

[ ]: