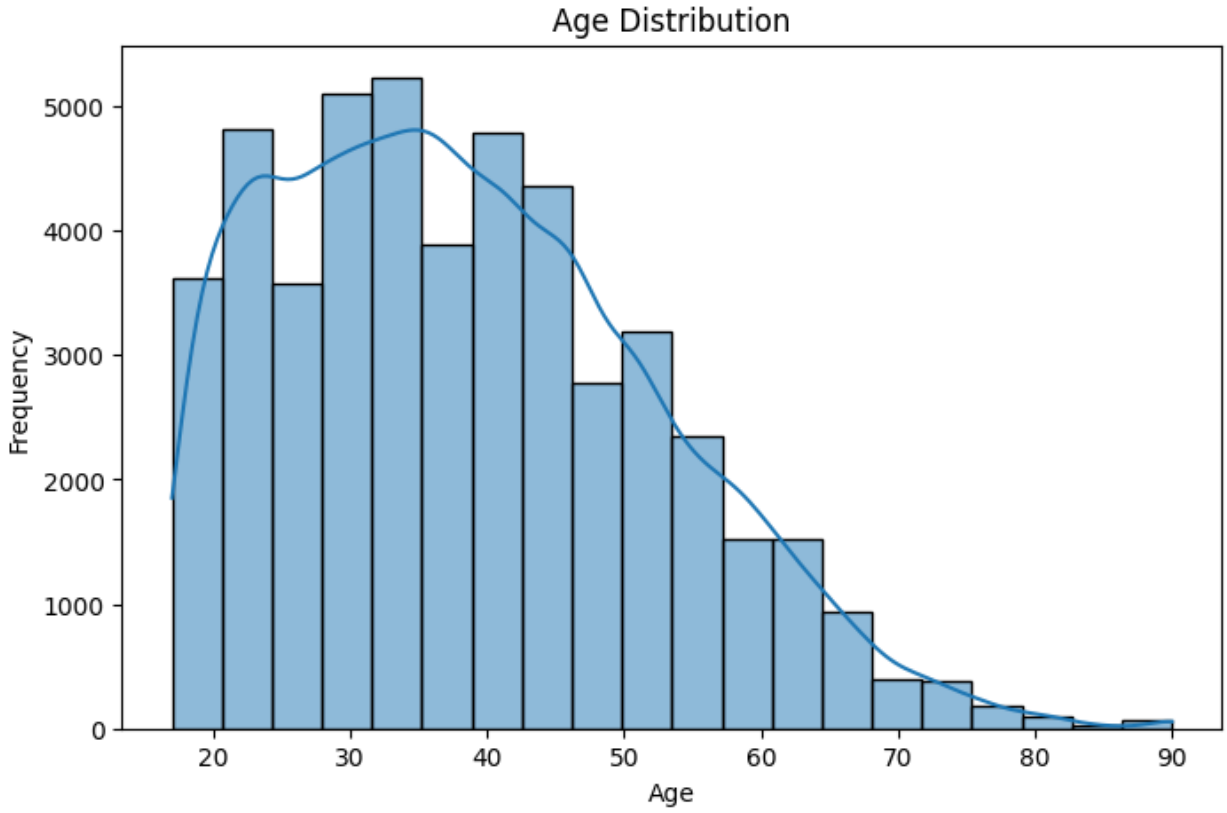
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('adult_dataset.csv')

df.head()
```

```
   age  workclass  fnlwgt     education  educational-num      marital-
status  \
0   25    Private  226802          11th                7       Never-
married
1   38    Private   89814       HS-grad                9  Married-civ-
spouse
2   28  Local-gov  336951    Assoc-acdm               12  Married-civ-
spouse
3   44    Private  160323  Some-college               10  Married-civ-
spouse
4   18          ?  103497  Some-college               10       Never-
married

          occupation relationship   race  gender  capital-gain
capital-loss  \
0  Machine-op-inspct    Own-child  Black    Male             0
0
1    Farming-fishing      Husband  White    Male             0
0
2    Protective-serv      Husband  White    Male             0
0
3  Machine-op-inspct      Husband  Black    Male          7688
0
4                  ?    Own-child  White  Female             0
0

   hours-per-week native-country income
0              40  United-States  <=50K
1              50  United-States  <=50K
2              40  United-States   >50K
3              40  United-States   >50K
4              30  United-States  <=50K
```

a. Histogram Objective: Visualize the distribution of age across the dataset to understand the spread and common age ranges among individuals.

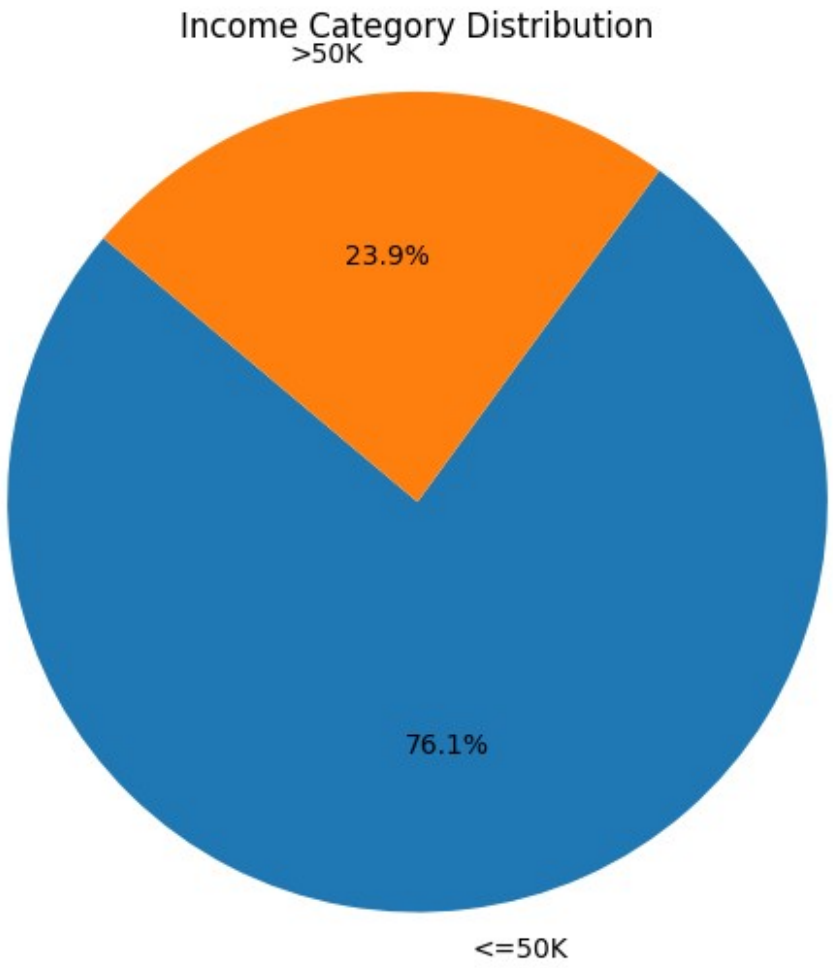
```
plt.figure(figsize=(8, 5))
sns.histplot(df['age'], bins=20, kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
```

```
plt.ylabel('Frequency')
plt.show()
```



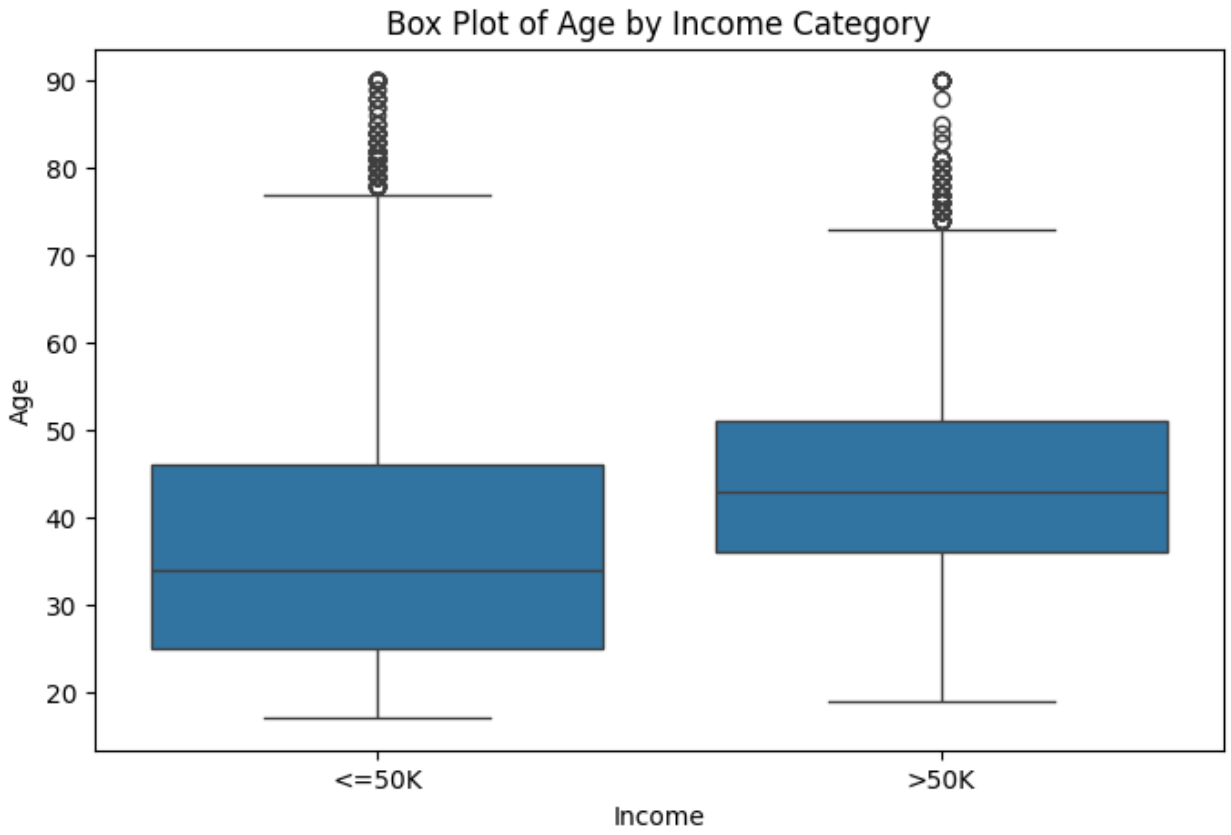
b. Pie Chart Objective: Show the proportion of individuals in each income category (<=50K or >50K).

```
income_counts = df['income'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(income_counts, labels=income_counts.index, autopct='%1.1f%%',
startangle=140)
plt.title('Income Category Distribution')
plt.axis('equal')
plt.show()
```



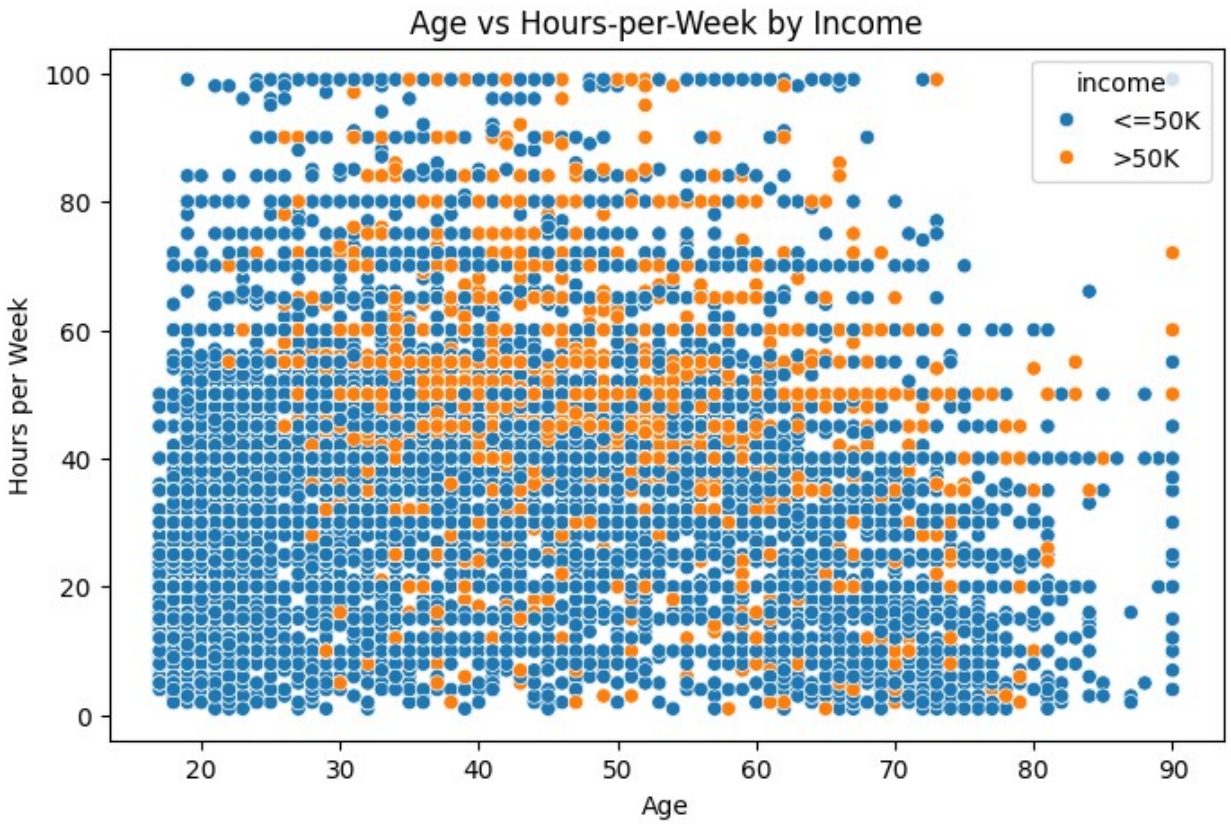
c. Box Plot Objective: Compare the age distributions of individuals based on their income category to detect outliers and medians.

```
plt.figure(figsize=(8, 5))
sns.boxplot(x='income', y='age', data=df)
plt.title('Box Plot of Age by Income Category')
plt.xlabel('Income')
plt.ylabel('Age')
plt.show()
```



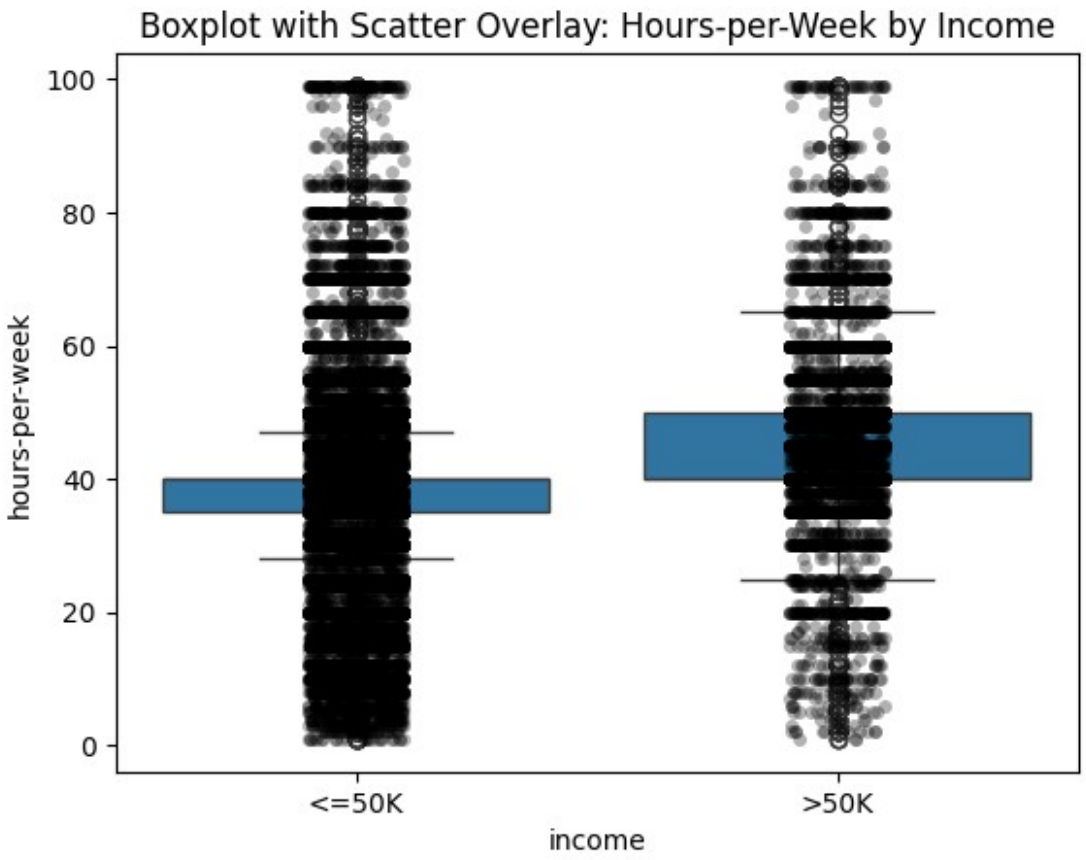
d. Scatter Plot Objective: Visualize the relationship between age and hours-per-week to explore working patterns by age.

```
plt.figure(figsize=(8, 5))
sns.scatterplot(data=df, x='age', y='hours-per-week', hue='income')
plt.title('Age vs Hours-per-Week by Income')
plt.xlabel('Age')
plt.ylabel('Hours per Week')
plt.show()
```



e. Add Boxplots to a Scatterplot Objective: Use a jointplot to show the scatter distribution of age vs hours-per-week, while also visualizing the marginal boxplots to understand distributions on each axis.

```
sns.boxplot(data=df, x='income', y='hours-per-week', whis=1.5)
sns.stripplot(data=df, x='income', y='hours-per-week', color='black',
alpha=0.3)
plt.title('Boxplot with Scatter Overlay: Hours-per-Week by Income')
plt.show()
```


Boxplot with Scatter Overlay: Hours-per-Week by Income
hours-per-week
100
80
60
40
20
0
<=50K
>50K
income

OR