Tips

Introduction:

This exercise was created based on the tutorial and documentation from Seaborn
The dataset being used is tips from Seaborn.

Step 1. Import the necessary libraries:

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

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called tips

```
tips = pd.read_csv('tips.csv')
print("Tips dataset:\n", tips.head())
```

```
Tips dataset:
    Unnamed: 0  total_bill   tip     sex smoker  day    time  size
0           0       16.99  1.01  Female     No  Sun  Dinner     2
1           1       10.34  1.66    Male     No  Sun  Dinner     3
2           2       21.01  3.50    Male     No  Sun  Dinner     3
3           3       23.68  3.31    Male     No  Sun  Dinner     2
4           4       24.59  3.61  Female     No  Sun  Dinner     4
```

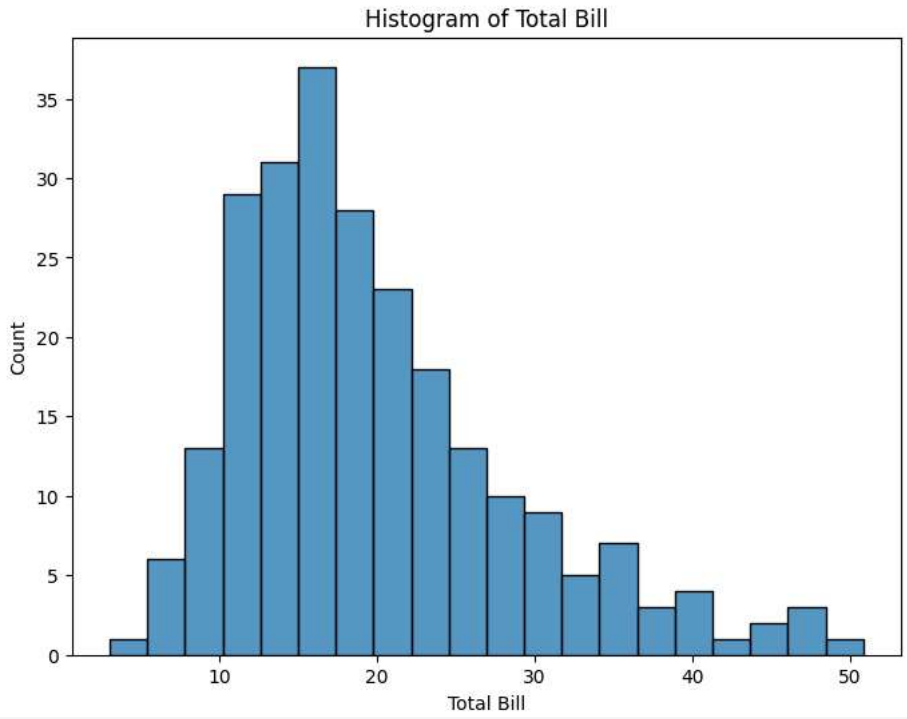
Step 4. Delete the Unnamed 0 column

```
tips = tips.drop(columns=['Unnamed: 0'])
print("Tips after dropping Unnamed: 0:\n", tips.head())
```

```
Tips after dropping Unnamed: 0:
    total_bill   tip     sex smoker  day    time  size
0       16.99  1.01  Female     No  Sun  Dinner     2
1       10.34  1.66    Male     No  Sun  Dinner     3
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4
```

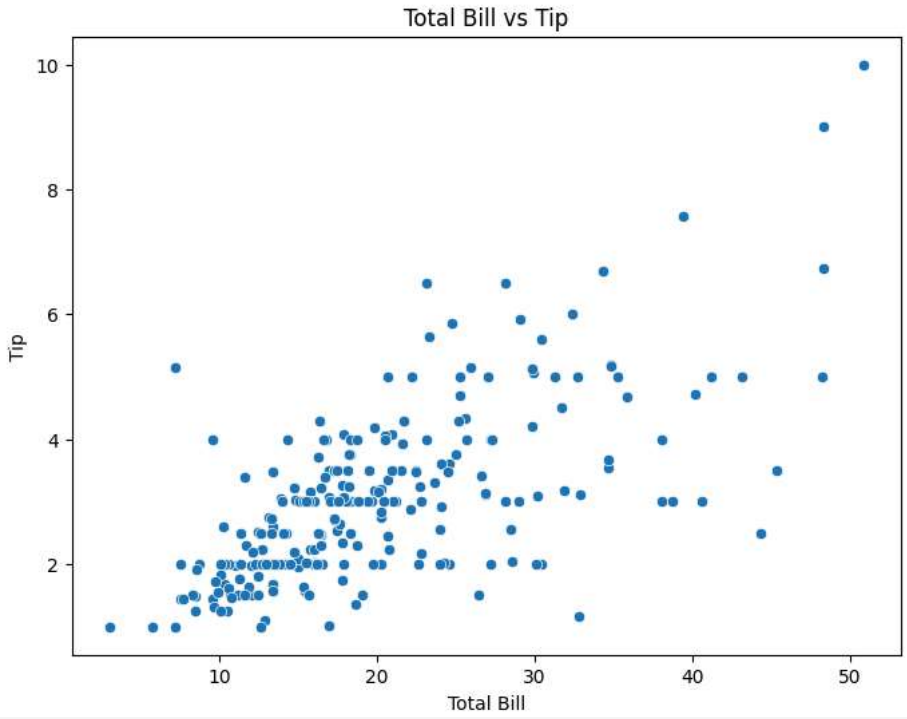
Step 5. Plot the total_bill column histogram

```
plt.figure(figsize=(8, 6))
sns.histplot(data=tips, x='total_bill', bins=20)
plt.title('Histogram of Total Bill')
plt.xlabel('Total Bill')
plt.ylabel('Count')
plt.show()
```



Step 6. Create a scatter plot presenting the relationship between total_bill and tip

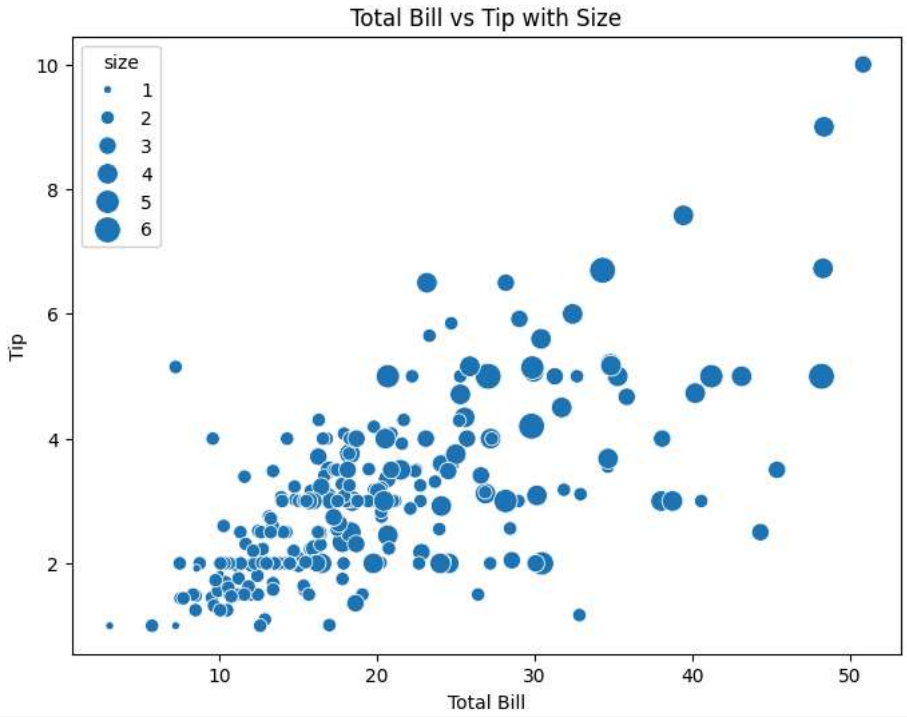
```
plt.figure(figsize=(8, 6))
sns.scatterplot(data=tips, x='total_bill', y='tip')
plt.title('Total Bill vs Tip')
plt.xlabel('Total Bill')
plt.ylabel('Tip')
plt.show()
```



Step 7. Create one image with the relationship of total_bill, tip and size.

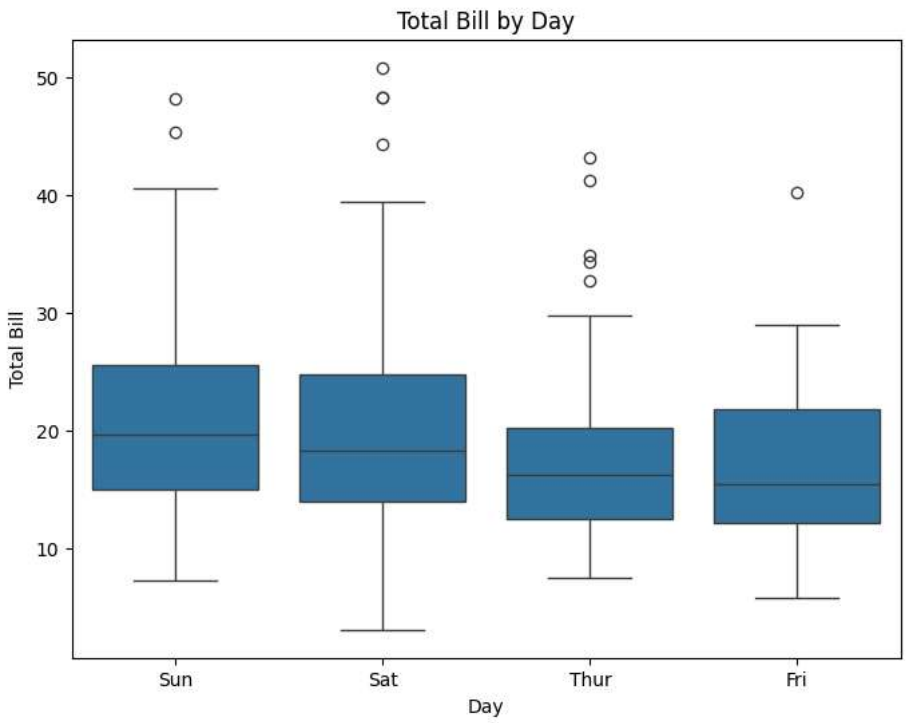
Hint: It is just one function.

```
plt.figure(figsize=(8, 6))
sns.scatterplot(data=tips, x='total_bill', y='tip', size='size', sizes=(20, 200))
plt.title('Total Bill vs Tip with Size')
plt.xlabel('Total Bill')
plt.ylabel('Tip')
plt.show()
```



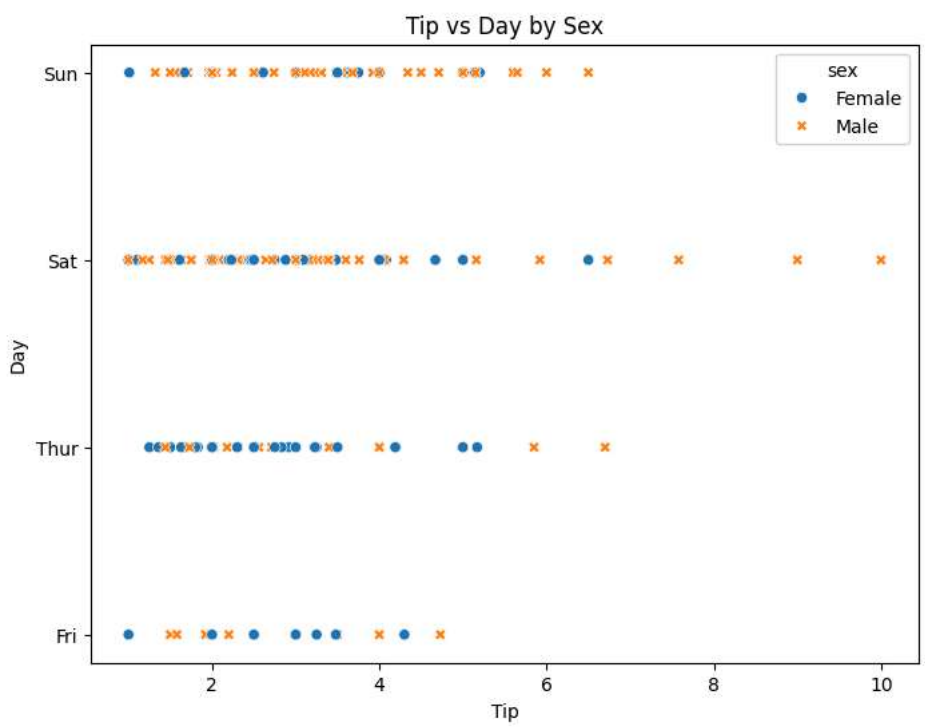
Step 8. Present the relationship between days and total_bill value

```
plt.figure(figsize=(8, 6))
sns.boxplot(data=tips, x='day', y='total_bill')
plt.title('Total Bill by Day')
plt.xlabel('Day')
plt.ylabel('Total Bill')
plt.show()
```



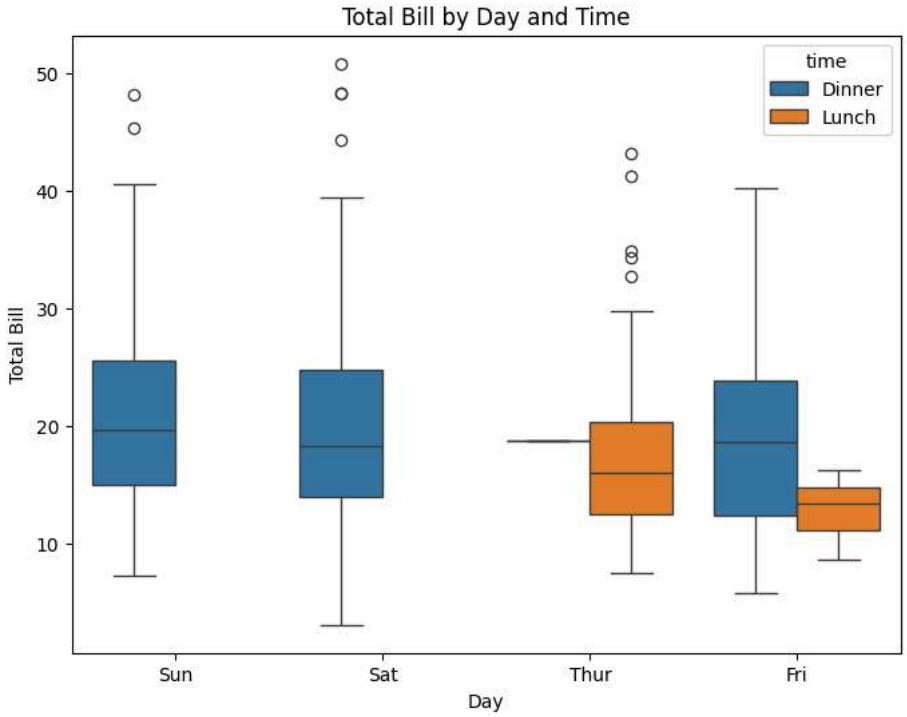
Step 9. Create a scatter plot with the day as the y-axis and tip as the x-axis, differ the dots by sex

```
plt.figure(figsize=(8, 6))
sns.scatterplot(data=tips, x='tip', y='day', hue='sex', style='sex')
plt.title('Tip vs Day by Sex')
plt.xlabel('Tip')
plt.ylabel('Day')
plt.show()
```



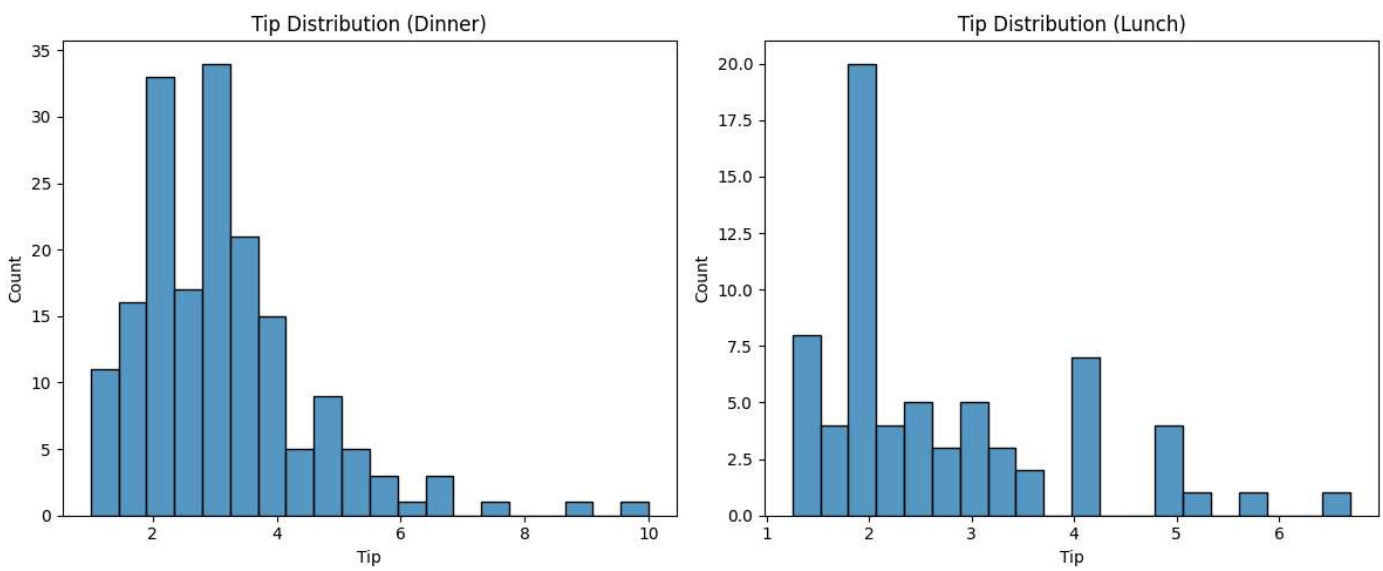
Step 10. Create a box plot presenting the total_bill per day differetiation the time (Dinner or Lunch)

```
plt.figure(figsize=(8, 6))
sns.boxplot(data=tips, x='day', y='total_bill', hue='time')
plt.title('Total Bill by Day and Time')
plt.xlabel('Day')
plt.ylabel('Total Bill')
plt.show()
```



Step 11. Create two histograms of the tip value based for Dinner and Lunch. They must be side by side.

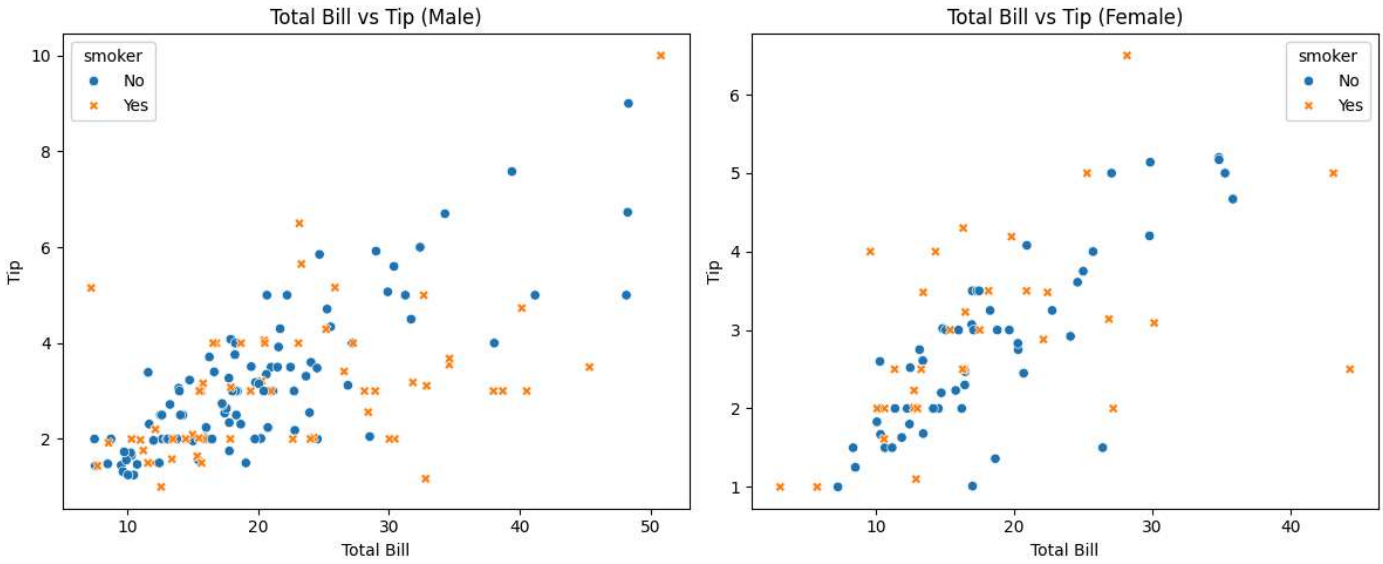
```
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
sns.histplot(data=tips[tips['time'] == 'Dinner'], x='tip', bins=20)
plt.title('Tip Distribution (Dinner)')
plt.xlabel('Tip')
plt.ylabel('Count')
plt.subplot(1, 2, 2)
sns.histplot(data=tips[tips['time'] == 'Lunch'], x='tip', bins=20)
plt.title('Tip Distribution (Lunch)')
plt.xlabel('Tip')
plt.ylabel('Count')
plt.tight_layout()
plt.show()
```



Step 12. Create two scatterplots graphs, one for Male and another for Female, presenting the total_bill value and tip relationship, differing by smoker or no smoker

They must be side by side

```
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
sns.scatterplot(data=tips[tips['sex'] == 'Male'], x='total_bill', y='tip', hue='smoker', style='smoker')
plt.title('Total Bill vs Tip (Male)')
plt.xlabel('Total Bill')
plt.ylabel('Tip')
plt.subplot(1, 2, 2)
sns.scatterplot(data=tips[tips['sex'] == 'Female'], x='total_bill', y='tip', hue='smoker', style='smoker')
plt.title('Total Bill vs Tip (Female)')
plt.xlabel('Total Bill')
plt.ylabel('Tip')
plt.tight_layout()
plt.show()
```



BONUS: Create your own question and answer it using a graph.

```
#vẽ box plot cho tip theo size, phân biệt theo time
plt.figure(figsize=(10, 6))
sns.boxplot(data=tips, x='size', y='tip', hue='time')
plt.title('Tip Distribution by Party Size and Time')
plt.xlabel('Party Size (Number of People)')
plt.ylabel('Tip Amount ($)')
plt.legend(title='Time')
plt.show()
```