

Name: Abhay Sharma

Class: BE - C

Moodle Id: 20102065

Roll No.: 28

Sub.: Applied Data Science

Experiment No.: 04

Aim: To explore the data visualization techniques.

Importing Libraries

```
import pandas as pd
import numpy as np
from scipy import stats
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
```

Loading the dataset

```
tips = sns.load_dataset('tips')
df = pd.DataFrame(tips)
df.head()
```

	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	

Preprocessing and Exploratory data analysis

Checking for missing values

```
df.isnull().sum()

total_bill    0
tip           0
sex           0
smoker        0
day           0
time          0
size          0
dtype: int64
```

Viewing the descriptive statistics of the dataset

```
df.describe()
```

	total_bill	tip	size	
count	244.000000	244.000000	244.000000	
mean	19.785943	2.998279	2.569672	
std	8.902412	1.383638	0.951100	
min	3.070000	1.000000	1.000000	
25%	13.347500	2.000000	2.000000	
50%	17.795000	2.900000	2.000000	

75%	24.127500	3.562500	3.000000
max	50.810000	10.000000	6.000000

Get a numerical summary for 'tip'

```
df.tip.describe()

count    244.000000
mean      2.998279
std       1.383638
min       1.000000
25%       2.000000
50%       2.900000
75%       3.562500
max       10.000000
Name: tip, dtype: float64
```

Summary for 'bill' and 'tip'

```
bill = df.total_bill

print("Maximum Bill: ", np.max(bill))
print("Minimum Bill: ", np.min(bill))
print("Standard Deviation: ", np.std(bill))
print("Median: ", np.median(bill))
print("Mean: ", np.mean(bill))

Maximum Bill:  50.81
Minimum Bill:  3.07
Standard Deviation:  8.884150577771132
Median:  17.795
Mean:  19.78594262295082
```

```
tip = df.tip

print("Maximum tip: ", np.max(tip))
print("Minimum tip: ", np.min(tip))
print("Standard Deviation: ", np.std(tip))
print("Median: ", np.median(tip))
print("Mean: ", np.mean(tip))

Maximum tip:  10.0
Minimum tip:  1.0
Standard Deviation:  1.3807999538298954
Median:  2.9
Mean:  2.99827868852459
```

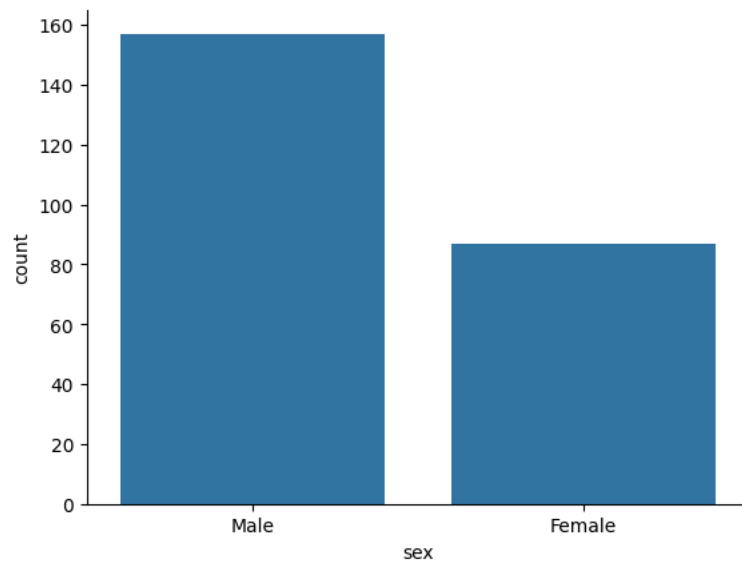
Exploratory Data analysis

Explore if there is any dependency between the variable 'tip' and rest of the variable

```
sns.countplot(x = 'sex', data = tips)
sns.despine() # no top and right axes spine

print(tips.sex.value_counts())
```

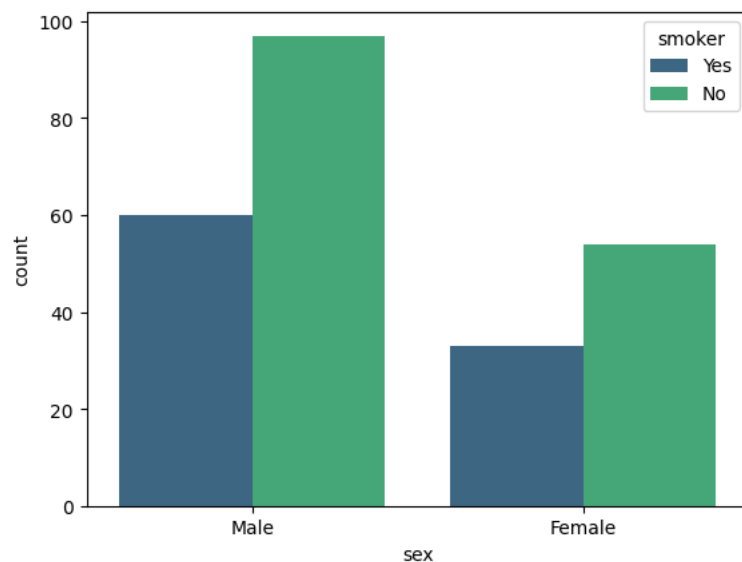
```
Male      157  
Female    87  
Name: sex, dtype: int64
```



Inference: Male customers paid more number of tips than Female

```
sns.countplot(x = 'sex', data = tips, hue = 'smoker', palette = 'viridis')
```

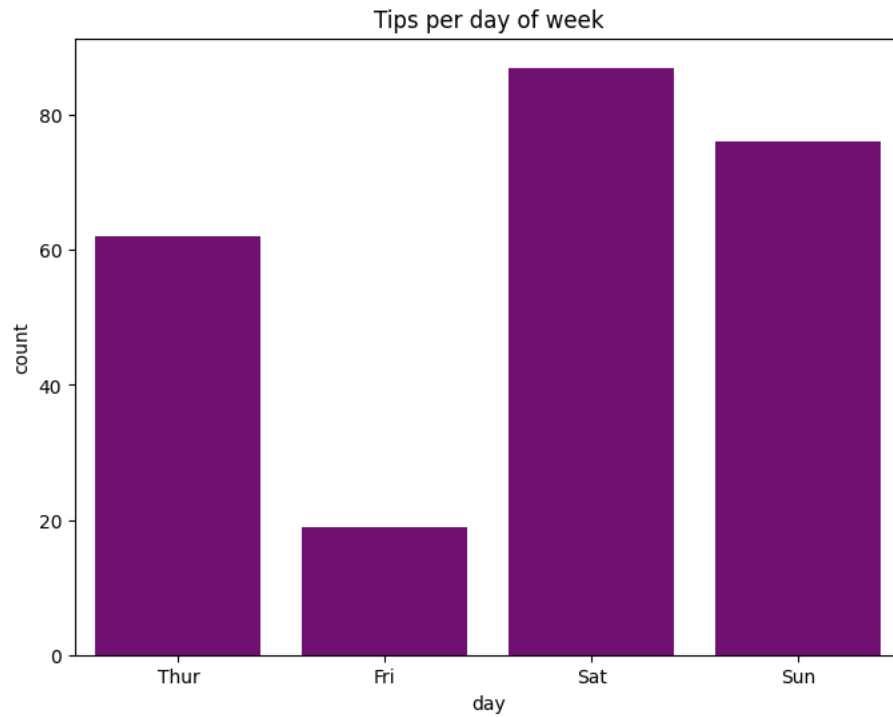
<Axes: xlabel='sex', ylabel='count'>



Inference: More number of tips have been paid by non smoker male and female.

```
plt.figure(figsize = (8,6))  
plt.title("Tips per day of week")  
sns.countplot(x = tips['day'], color = 'purple')
```

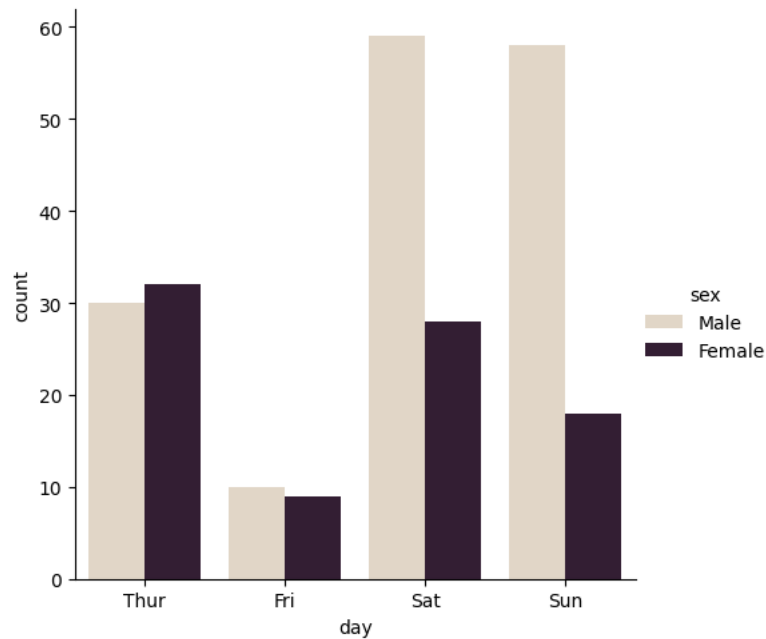
```
<Axes: title={'center': 'Tips per day of week'}, xlabel='day', ylabel='count'>
```



Inference: More number of tips are paid on weekends i.e. on saturday and sunday.

```
sns.catplot(x = 'day', data = tips, hue = 'sex', palette = 'ch:.25', kind = 'count')
```

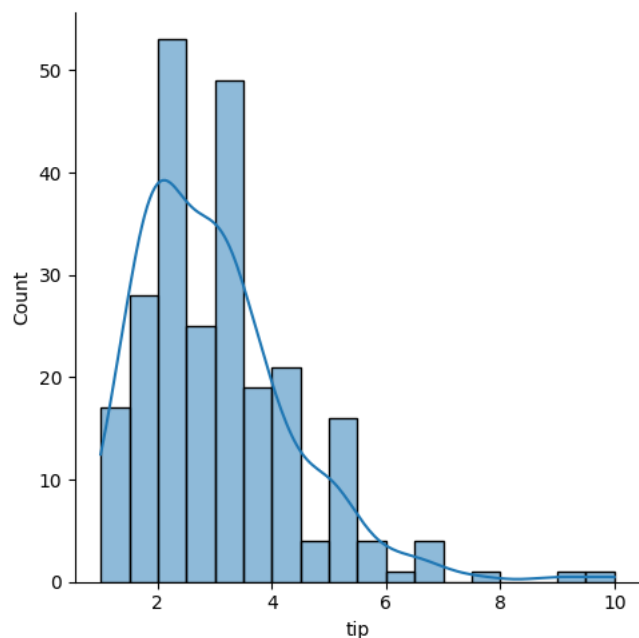
```
<seaborn.axisgrid.FacetGrid at 0x7ba54e3a0490>
```



Inference: On weekends Male customers paid more number of tips than female.

```
sns.displot(df['tip'], kde = 'true')
```

<seaborn.axisgrid.FacetGrid at 0x7ba54c08d210>



Inference: More tips are paid in range of 2 to 4 and there are some Outliers.

```
g = sns.distplot(tips.tip, kde = False)
g.set_title("Tip amount Histogram")
```

<ipython-input-15-c9959e0b8632>:1: UserWarning:

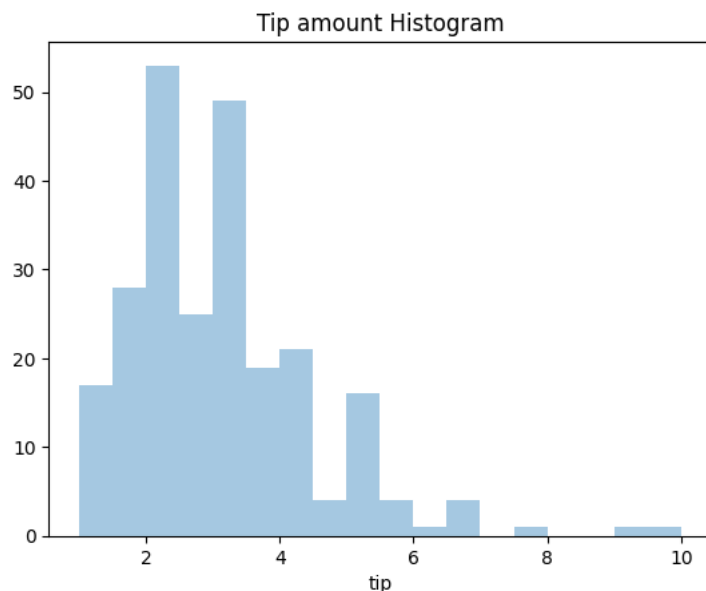
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see

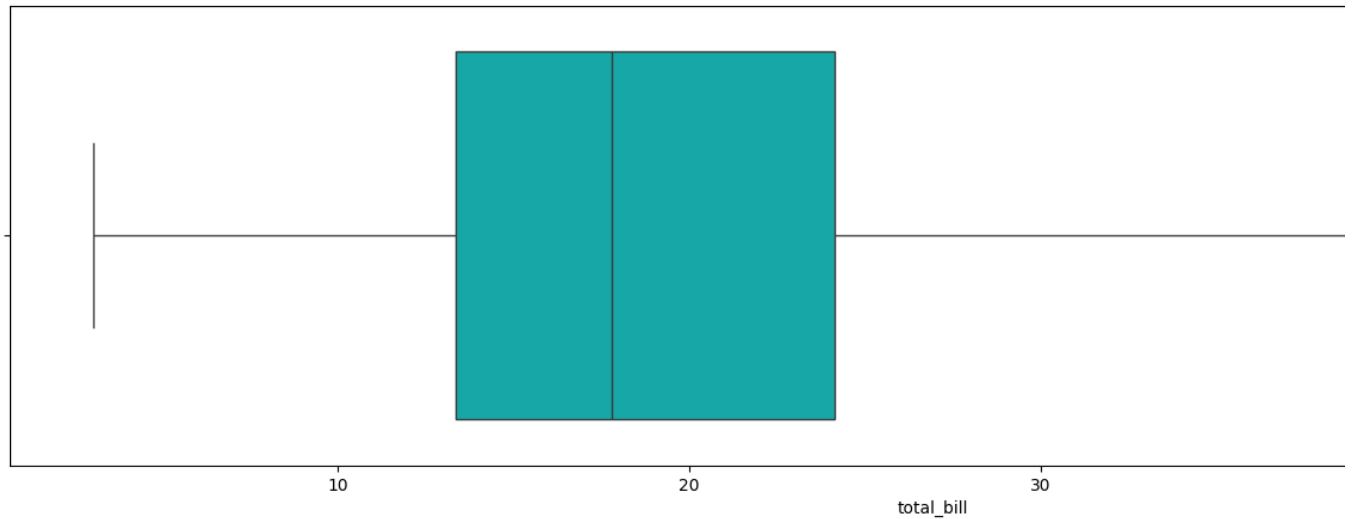
<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
g = sns.distplot(tips.tip, kde = False)
Text(0.5, 1.0, 'Tip amount Histogram')
```



```
plt.figure(figsize=(20,5))
sns.boxplot(x=bill,color="c")
```

<Axes: xlabel='total_bill'>

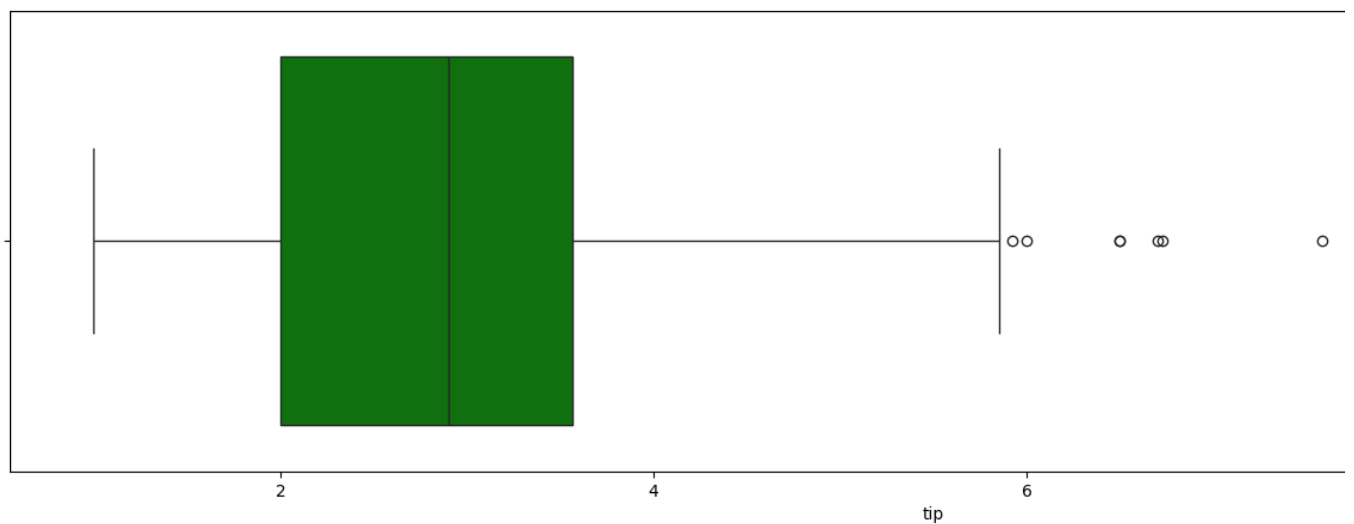


Result : Majority Range=(15,25)

outliers = upto 55

```
plt.figure(figsize=(20,5))
sns.boxplot(x=tip,color="g")
```

<Axes: xlabel='tip'>

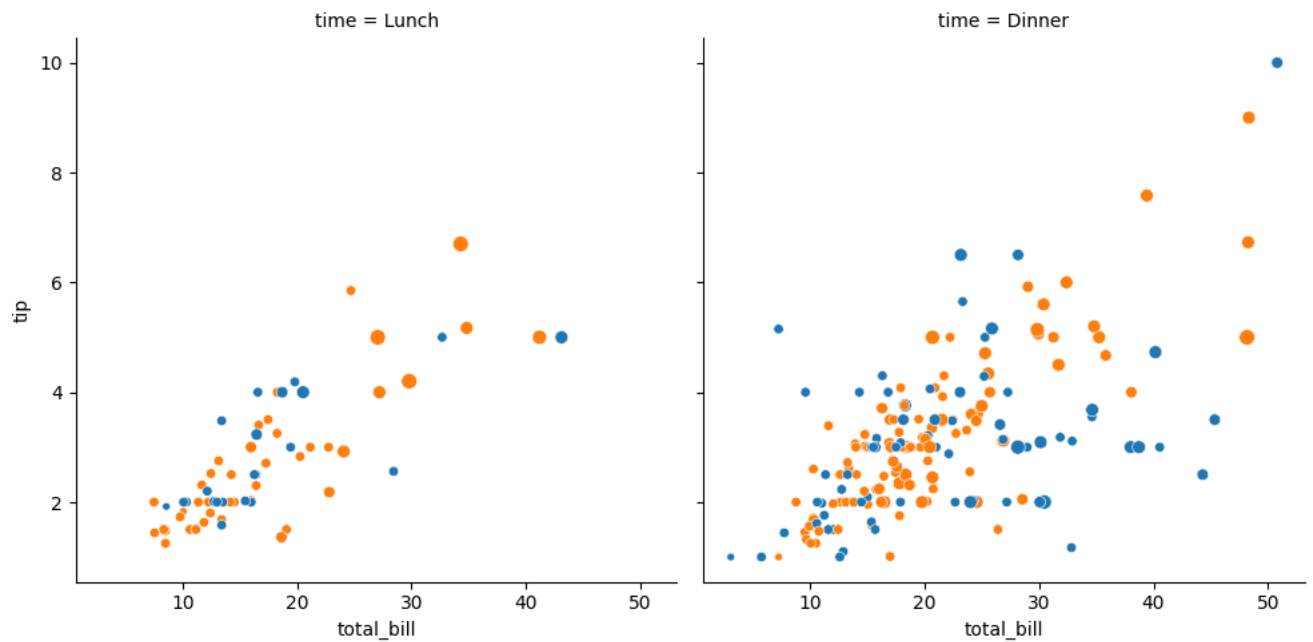


Result : Majority Range=(2,4)

outliers = upto 10

```
sns.relplot(x = "total_bill", y = 'tip', data=df,col = 'time', hue='smoker',size='size')
```

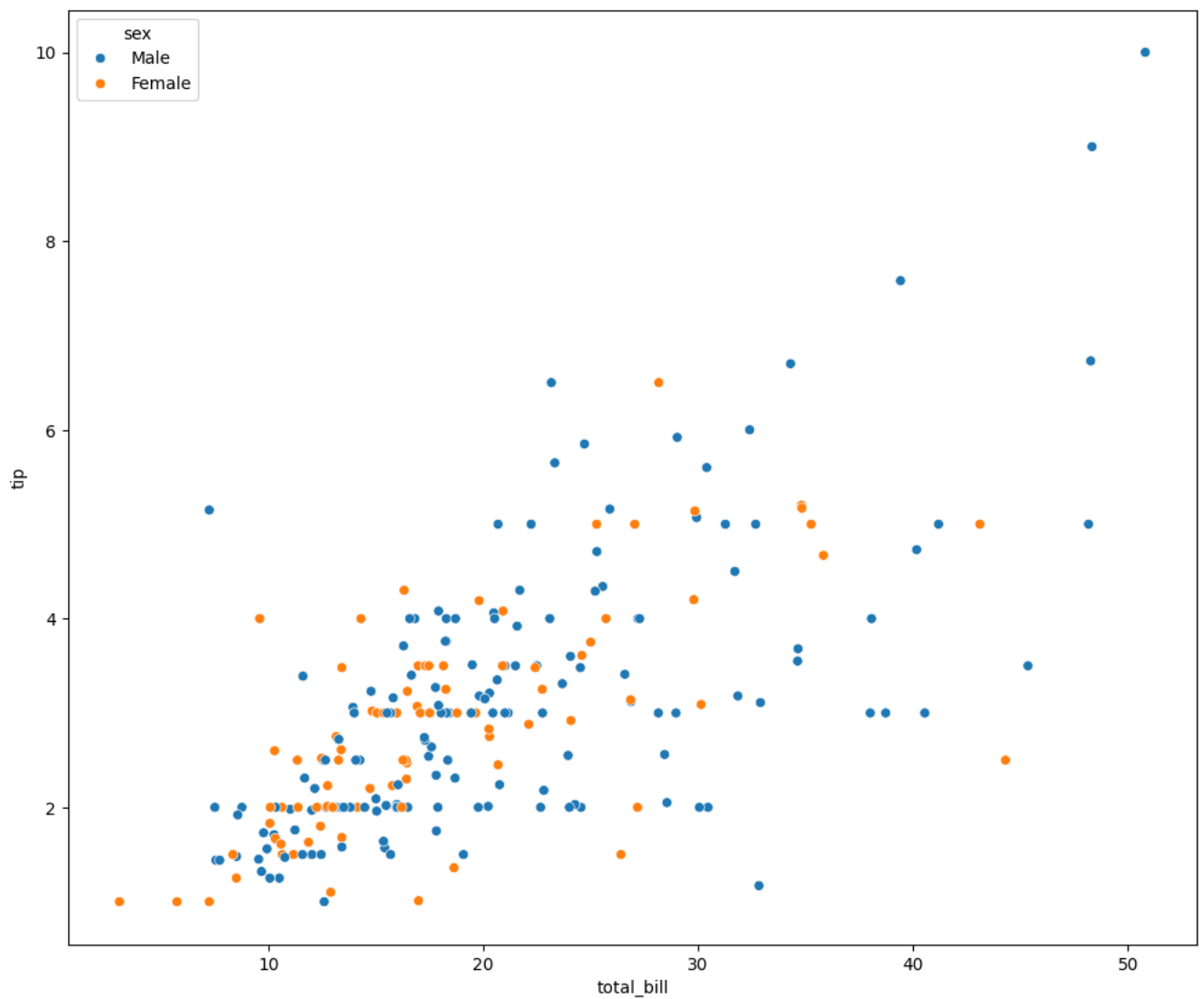
<seaborn.axisgrid.FacetGrid at 0x7ba5451b2320>



Result:

1. lunch time: less tips / Dinner time: More tips
2. more bill amount : Bigger tip
3. Non Smoker pay more tips

```
plt.figure(figsize=(12,10))
sns.scatterplot(x = "total_bill", y = 'tip', data=df, hue='sex');
```



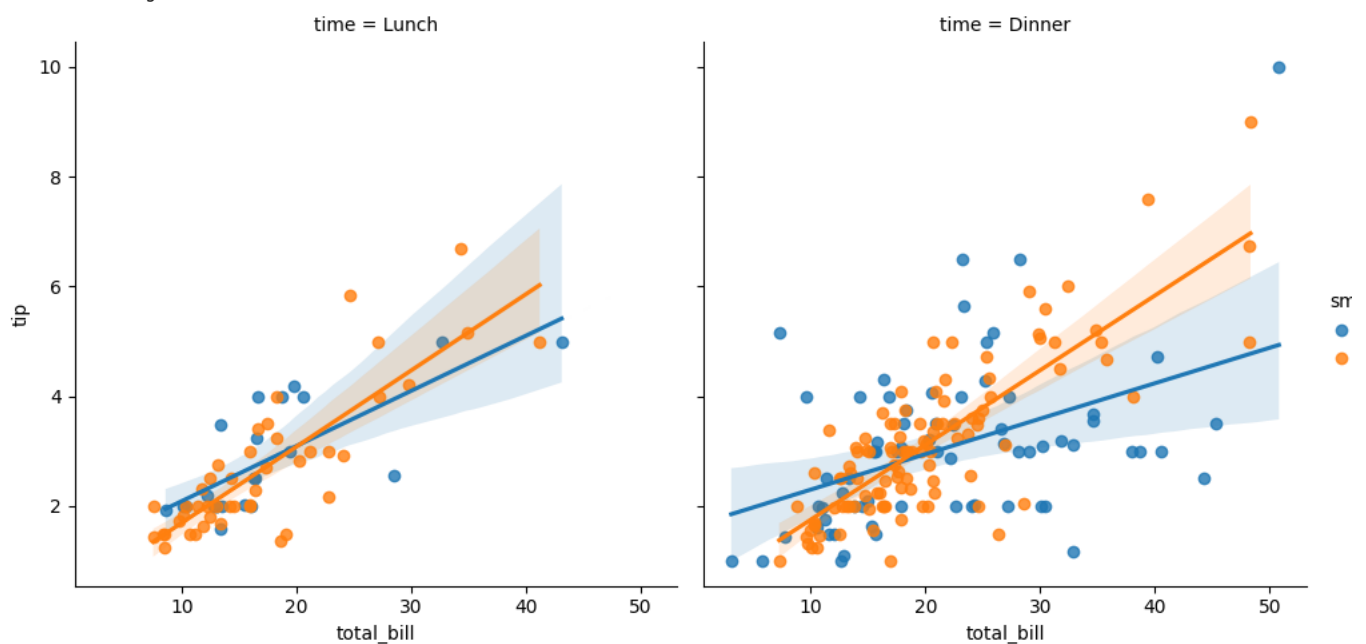
Result:

Male Pay more tips

as amount increases tip incresases

```
sns.lmplot(x = "total_bill", y = 'tip', data=df,col = 'time', hue='smoker')
```

<seaborn.axisgrid.FacetGrid at 0x7ba5452274f0>

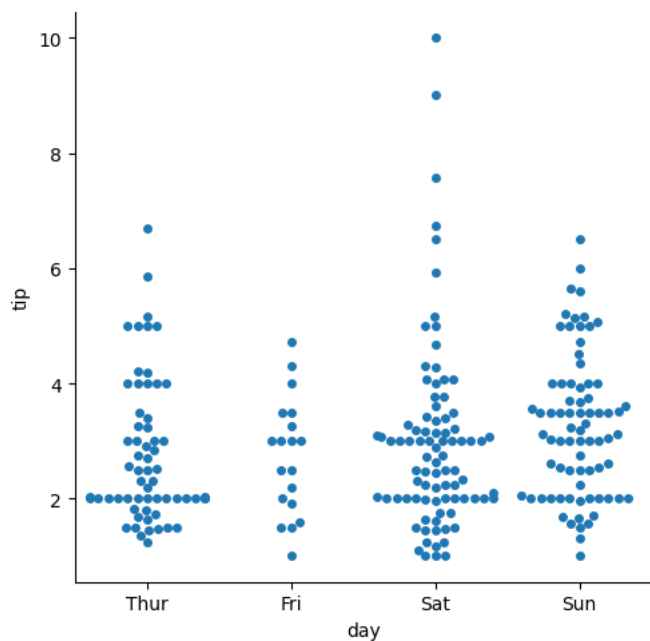


Result:

1. lunch time: less tips / Dinner time: More tips
2. more bill amount : Bigger tip
3. Non Smoker pay more tips

```
sns.catplot(x="day",y="tip",data=df,kind="swarm")
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/categorical.py:3398: UserWarning: 8.1% of the points cannot be placed; y  
warnings.warn(msg, UserWarning)  
<seaborn.axisgrid.FacetGrid at 0x7ba5432a51b0>
```



Result:

More tips on saturday and sunday

Highest on Saturday

```
sns.catplot(x="day",y="tip",data=df,kind="bar",palette = 'muted',height=6)
```

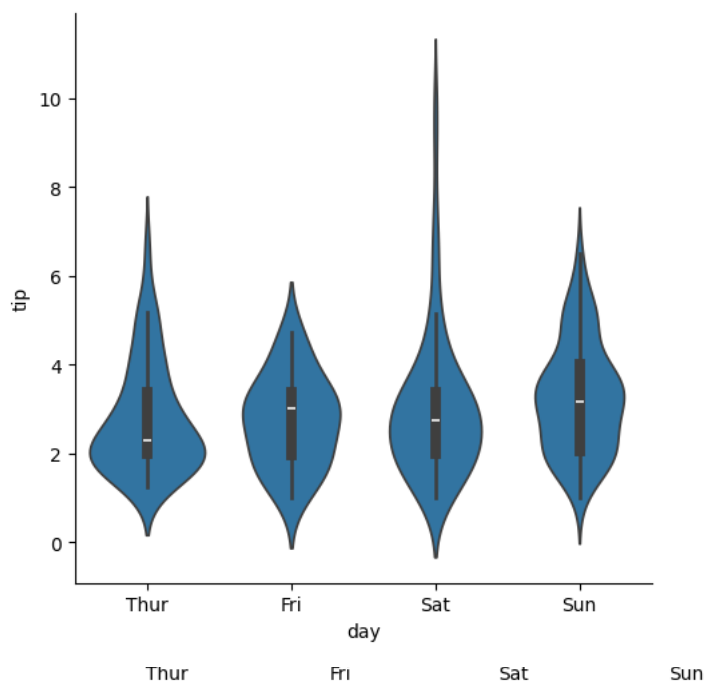
```
<ipython-input-34-6aef28c82619>:1: FutureWarning:
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue`
```

```
sns.catplot(x="day", y="tip", data=df, kind="bar", palette='muted', height=6)
```

```
sns.catplot(x="day", y="tip", data=df, kind="violin")
```

```
<seaborn.axisgrid.FacetGrid at 0x7ba542dcda20>
```



```
sns.pairplot(df, hue='sex')
```

```
<seaborn.axisgrid.PairGrid at 0x7ba542dcd9c0>
```

