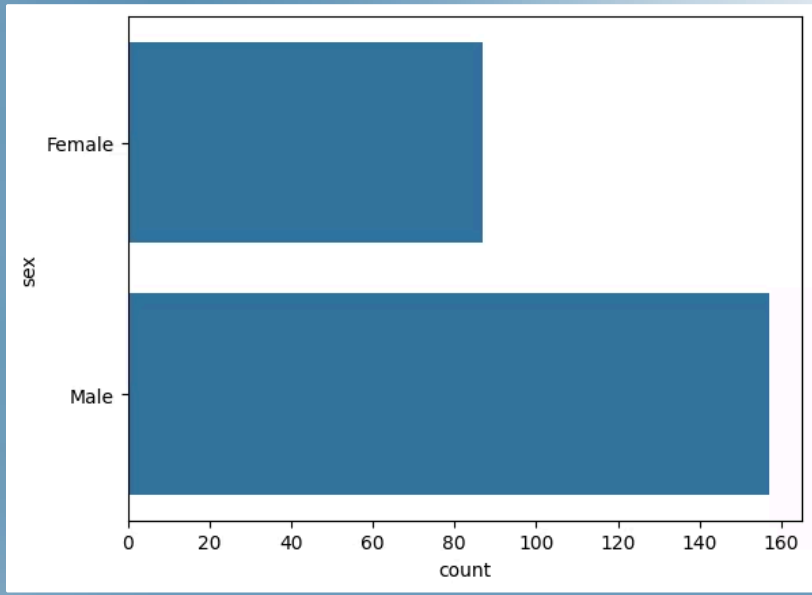


Understanding Seaborn's Countplot for 'Sex' Variable

This presentation explains how to interpret a horizontal bar plot from Seaborn's countplot of the 'sex' column in a dataset.

Reading the Gender Distribution Graph

Graph Axes

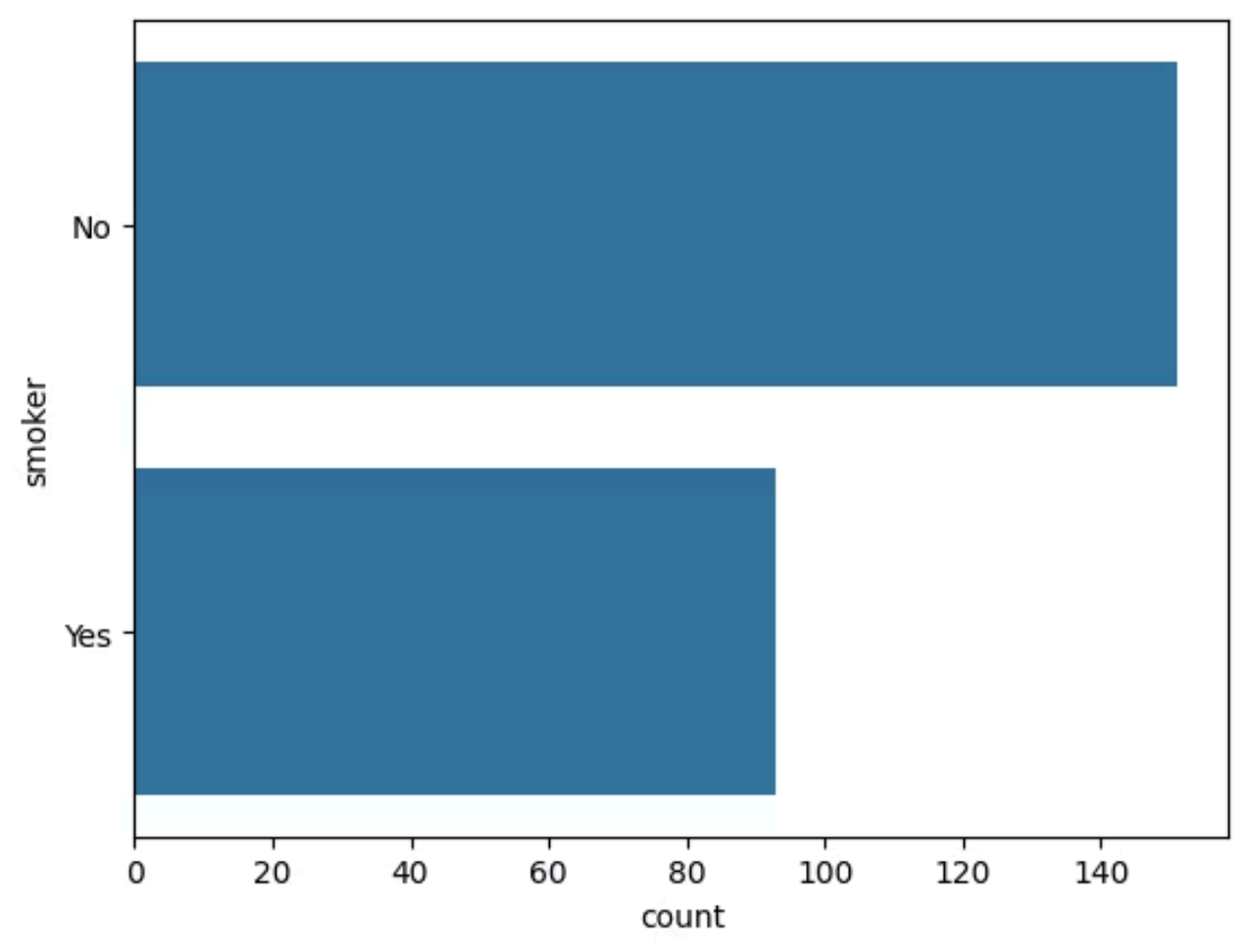
The Y-axis lists categories: Female and Male.

The X-axis shows the count of records for each sex.

Bar Length Interpretation

Longer bars mean higher counts.

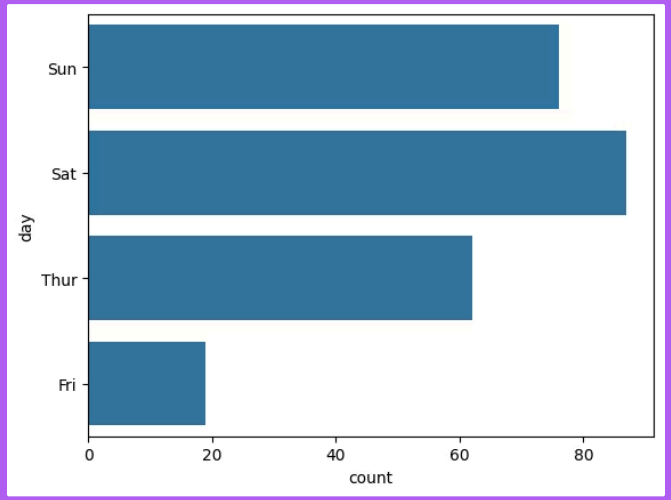
Males appear nearly twice as frequent as females.



sns.countplot(df['smoker']) write a desciption in short

This bar plot shows the distribution of smokers in the dataset.

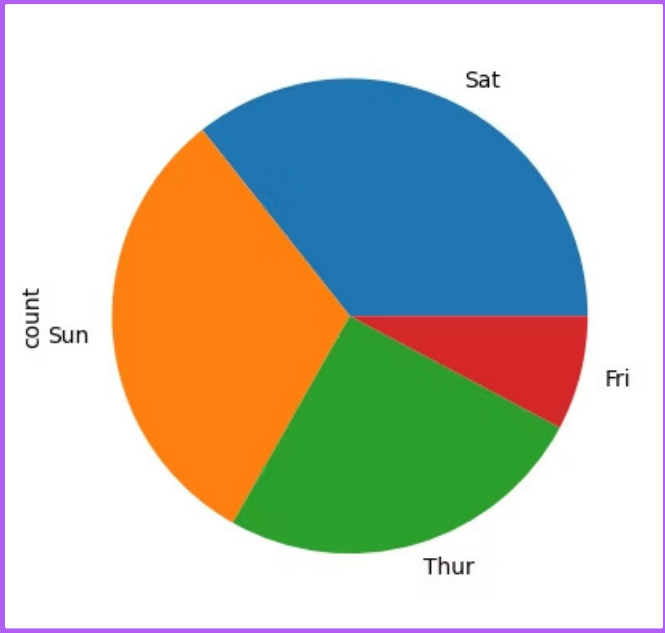
- **"No"** (non-smokers) are more frequent, with around **150** entries.
- **"Yes"** (smokers) have fewer entries, around **90**.
  Thus, there are significantly more non-smokers than smokers in the data.



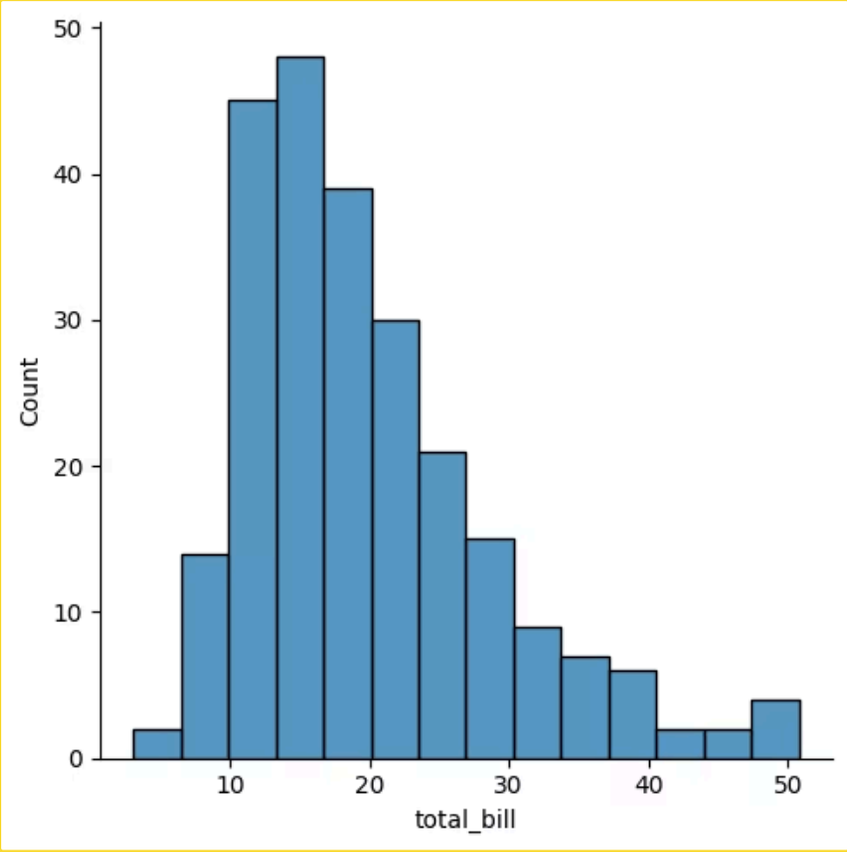

```
sns.countplot(df['day'])
```

This bar plot shows the distribution of entries across different days.

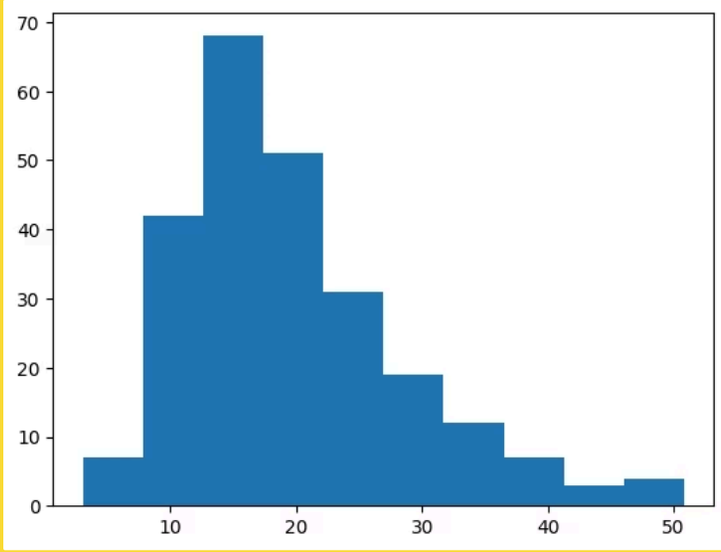
- **Saturday** has the highest count (around 90), followed by **Sunday** and **Thursday**.
- **Friday** has the lowest count, with very few entries (around 20). Thus, most data was recorded during the weekend (Saturday and Sunday).



sns.displot(df['total_bill'])



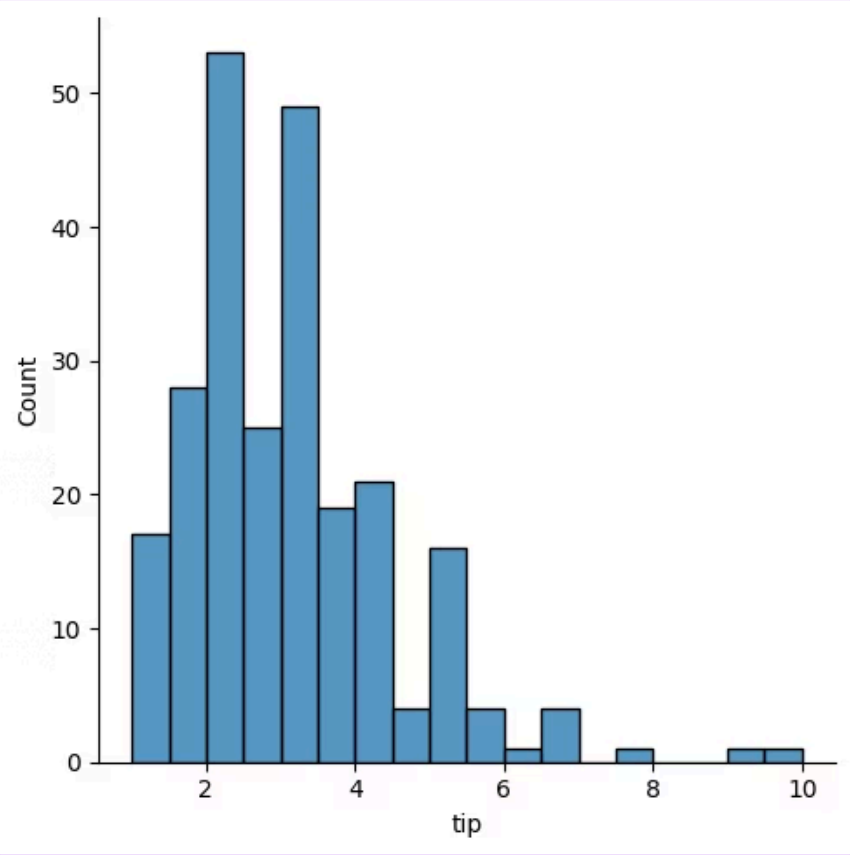

```
plt.hist(df['total_bill'])
```



This histogram shows the distribution of the `total_bill` amounts.

- Most bills are between **10 and 20 units**.
- The frequency drops as the bill amount increases beyond 20.
- The distribution is **right-skewed** (longer tail on the right), meaning a few high-value bills are present.

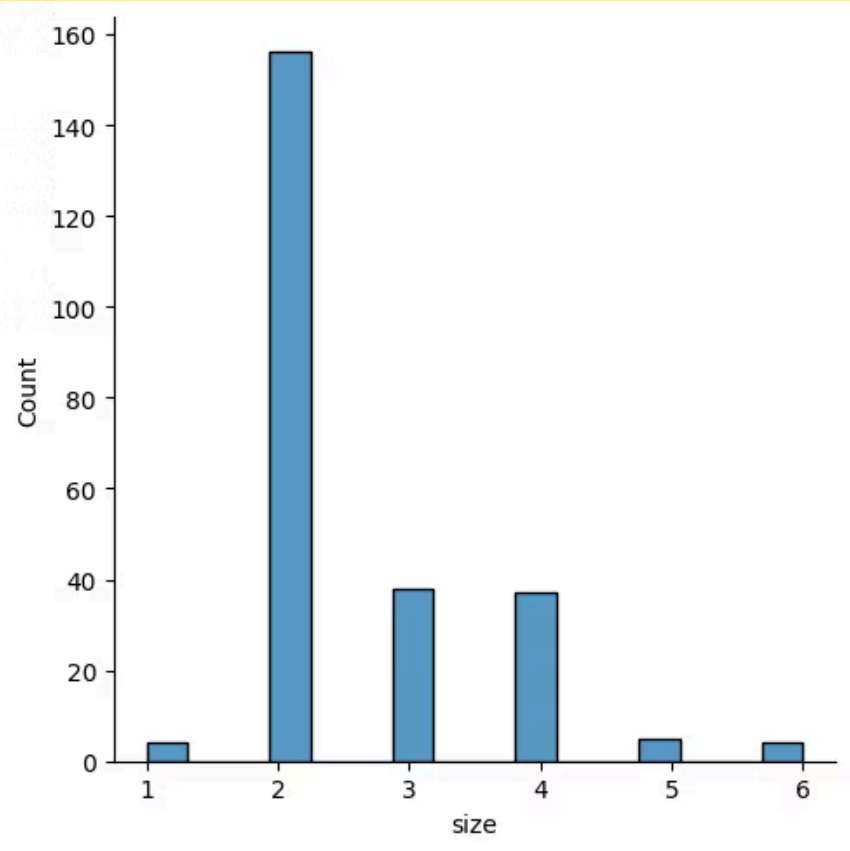
sns.displot(df['tip'])



This histogram shows the distribution of `tip` amounts.

- Most tips are between **2 and 4 units**.
- The frequency sharply decreases for higher tips.
- The distribution is **right-skewed**, indicating that larger tips are rare but present.

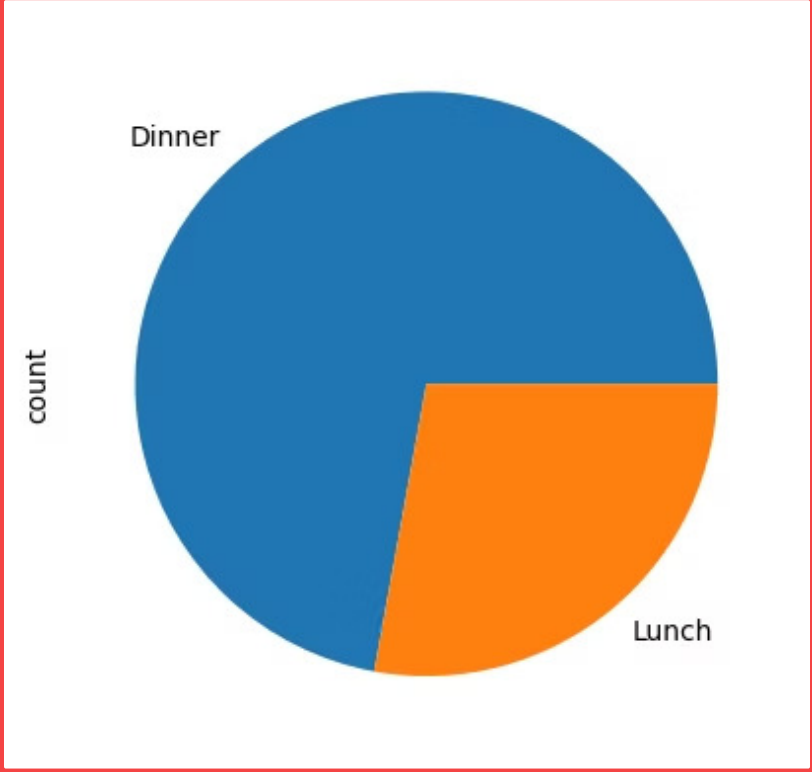
sns.displot(df['size'])



This histogram shows the distribution of the `size` of dining groups.

- Most groups have **2 people**, with around **160 entries**.
- Groups of size **3 and 4** are less frequent but still common.
- Groups of **1, 5, and 6** are very rare.

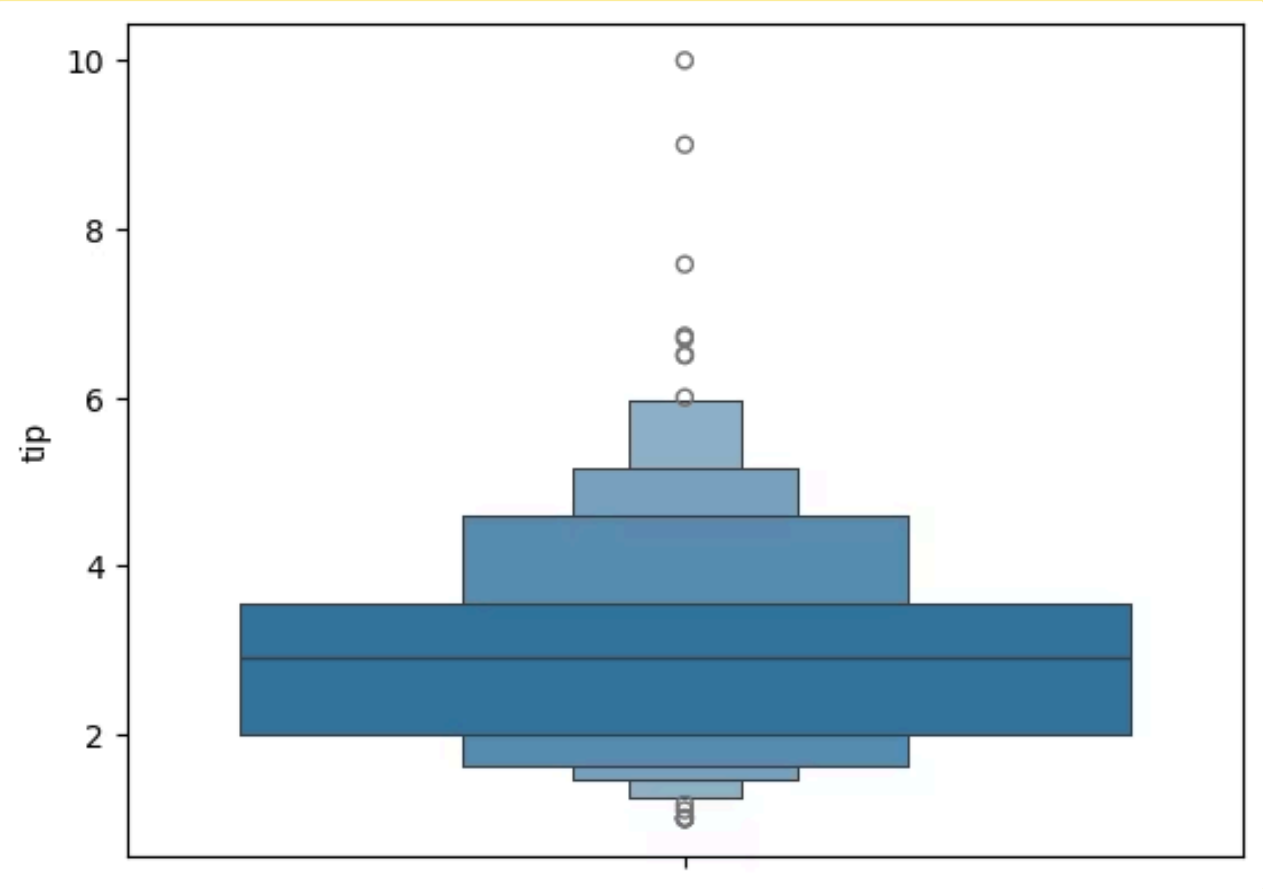
df['time'].value_counts().plot(kind='pie')



This pie chart shows the distribution of meal times.

- **Dinner** accounts for the majority of entries.
- **Lunch** makes up a smaller portion, roughly around one-third of the total.

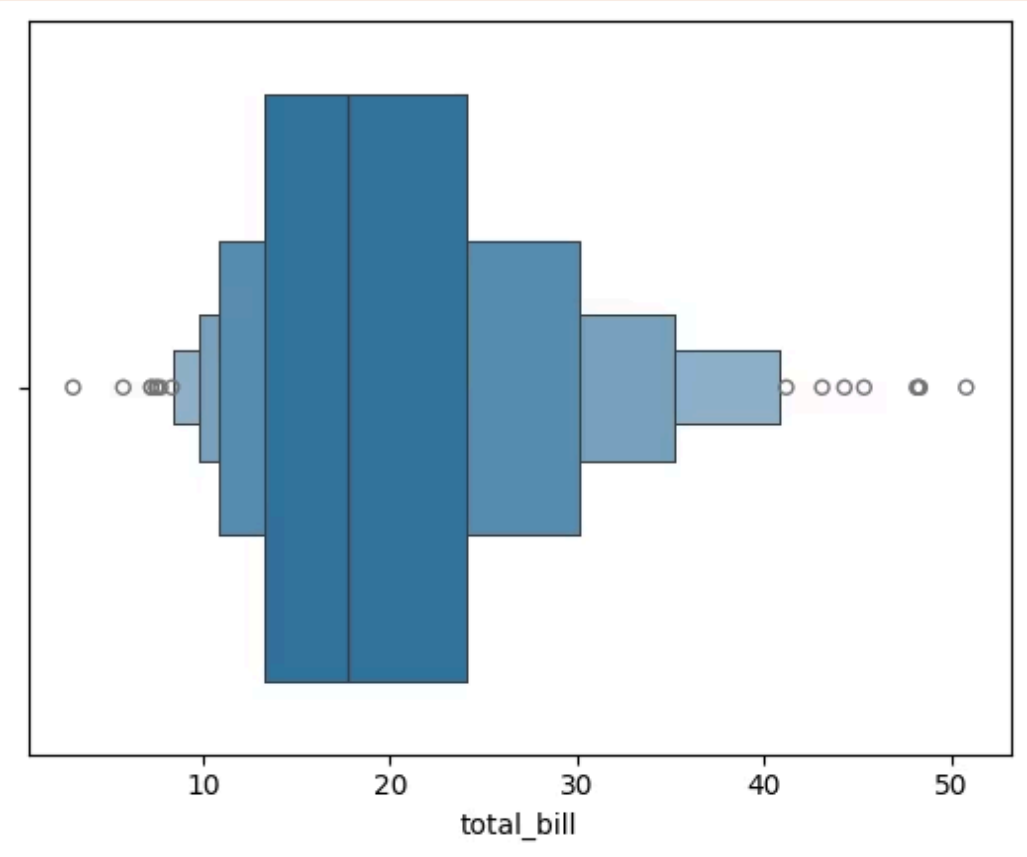
sns.boxenplot(df['tip'])



This boxen plot shows the distribution of `tip` amounts.

- Most tips are concentrated between **2 and 4 units**.
- There are several **outliers** above 6 units, with the highest around **10**.
- The data is **right-skewed**, meaning a few higher tips pull the distribution to the right.

sns.boxenplot(x=df['total_bill'], orient='h')



This boxen plot shows the distribution of `total_bill` amounts.

- Most total bills are between **10 and 25 units**.
- The median bill is around **18 units**.
- There are several **outliers** above 40 units, with the highest reaching nearly **50 units**.
- The distribution is **right-skewed**, indicating that a few high total bills stretch the upper range.