

```
In [1]: # This is the summary of visualization in Python including Pandas, Matplotlib.pyplot and Seaborn.
```

```
In [2]: ## Dataset of NYC Taxi Records in January,2017
import pandas as pd
df = pd.read_csv('/Users/xiaoruizhang/Desktop/yellow_tripdata_2017-01.csv',
                 sep=',')
df = df.loc[0:5000,]
df.head()
```

Out[2]:

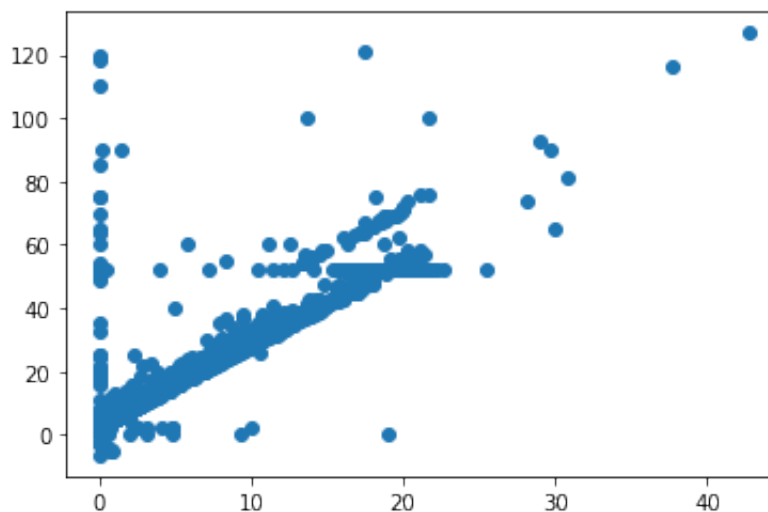
|   | VendorID | tpep_pickup_datetime | tpep_dropoff_datetime | passenger_count | trip_distance | Rat |
|---|----------|----------------------|-----------------------|-----------------|---------------|-----|
| 0 | 1        | 2017-01-09 11:13:28  | 2017-01-09 11:25:45   | 1               | 3.30          |     |
| 1 | 1        | 2017-01-09 11:32:27  | 2017-01-09 11:36:01   | 1               | 0.90          |     |
| 2 | 1        | 2017-01-09 11:38:20  | 2017-01-09 11:42:05   | 1               | 1.10          |     |
| 3 | 1        | 2017-01-09 11:52:13  | 2017-01-09 11:57:36   | 1               | 1.10          |     |
| 4 | 2        | 2017-01-01 00:00:00  | 2017-01-01 00:00:00   | 1               | 0.02          |     |

```
In [3]: ## Matplotlib.pyplot
```

```
In [4]: import matplotlib.pyplot as plt
%matplotlib inline
```

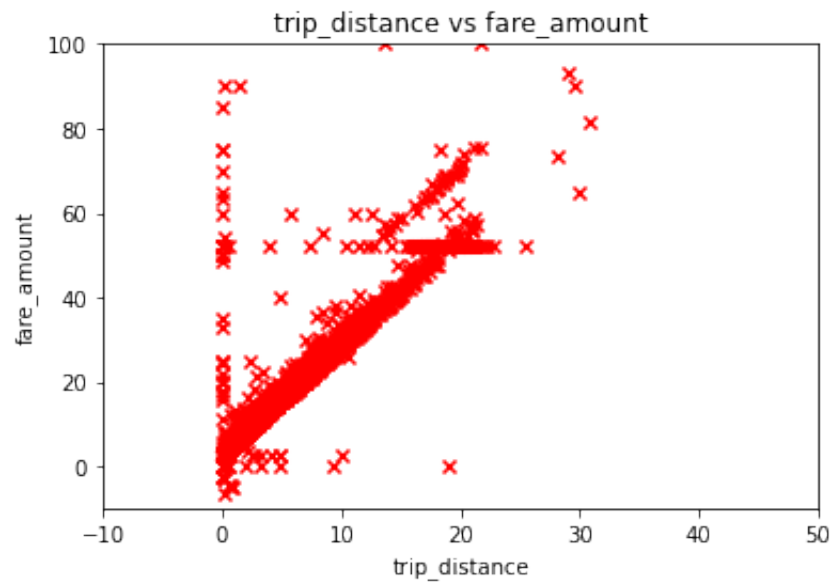
```
In [5]: plt.scatter(df.trip_distance,df.fare_amount)
```

Out[5]: <matplotlib.collections.PathCollection at 0x7ff5b48ba3d0>



```
In [6]: ## Matplotlib Axes
```

```
In [7]: fig = plt.figure(figsize=(6,4))
ax = fig.gca()
ax.scatter(x=df.trip_distance, y=df.fare_amount,marker='x',color='red'
)
ax.set_xlabel('trip_distance')
ax.set_ylabel('fare_amount')
ax.set_xlim([-10,50])
ax.set_ylim([-10,100])
ax.set_title('trip_distance vs fare_amount');
```



```
In [8]: ## Matplotlib: Subplots, Figure and Axis
```

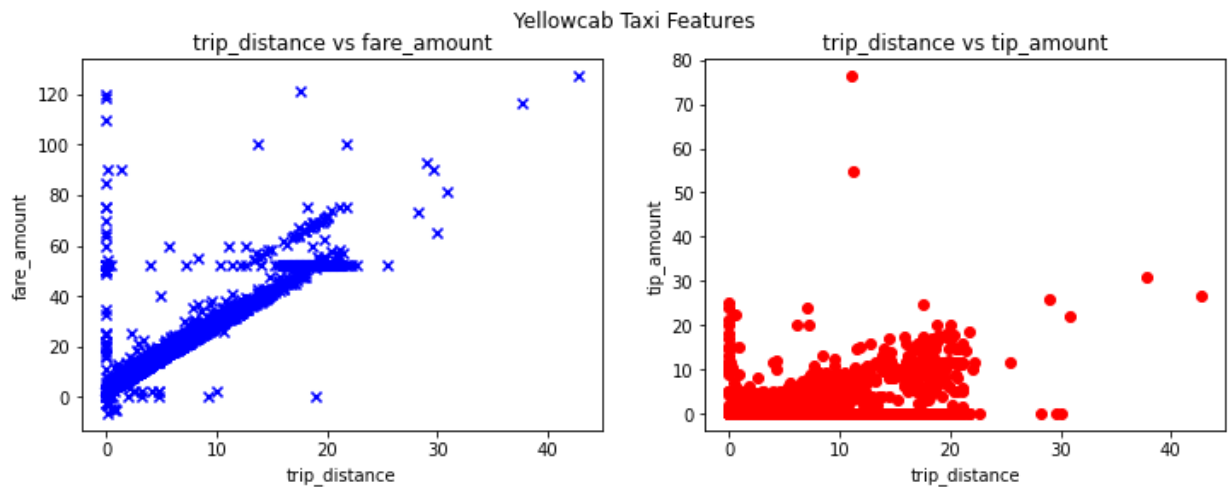
```
In [9]: fig,ax = plt.subplots(1,2,figsize=(12,4))

ax[0].scatter(df.trip_distance,df.fare_amount,marker='x',color='blue')
ax[1].scatter(df.trip_distance,df.tip_amount,color='red');

ax[0].set_xlabel('trip_distance')
ax[1].set_xlabel('trip_distance')

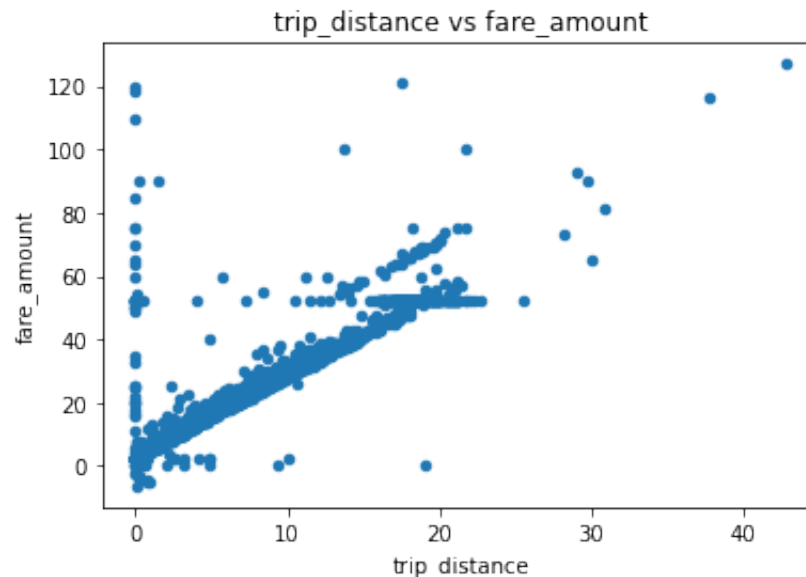
ax[0].set_ylabel('fare_amount'), ax[1].set_ylabel('tip_amount')

ax[0].set_title('trip_distance vs fare_amount')
ax[1].set_title('trip_distance vs tip_amount')
fig.suptitle('Yellowcab Taxi Features');
```



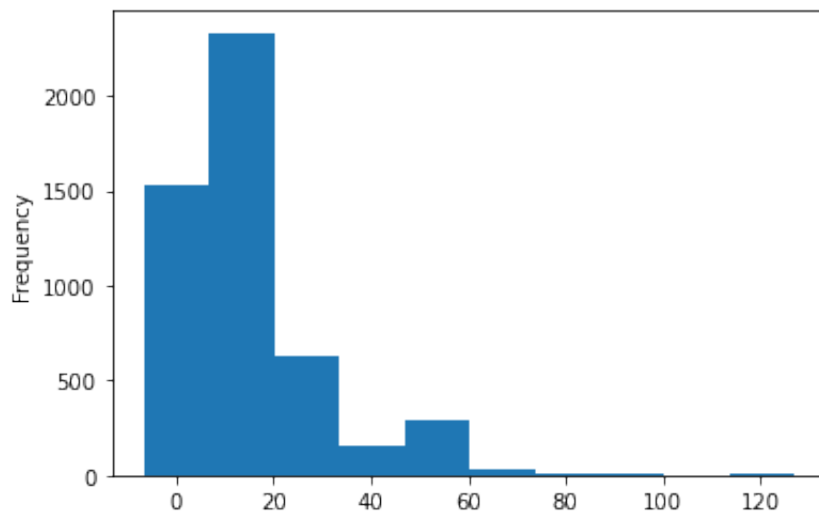
```
In [10]: ## Plotting via Pandas
```

```
In [11]: ax = df.plot.scatter(x='trip_distance',y='fare_amount');
ax.set_title('trip_distance vs fare_amount');
```

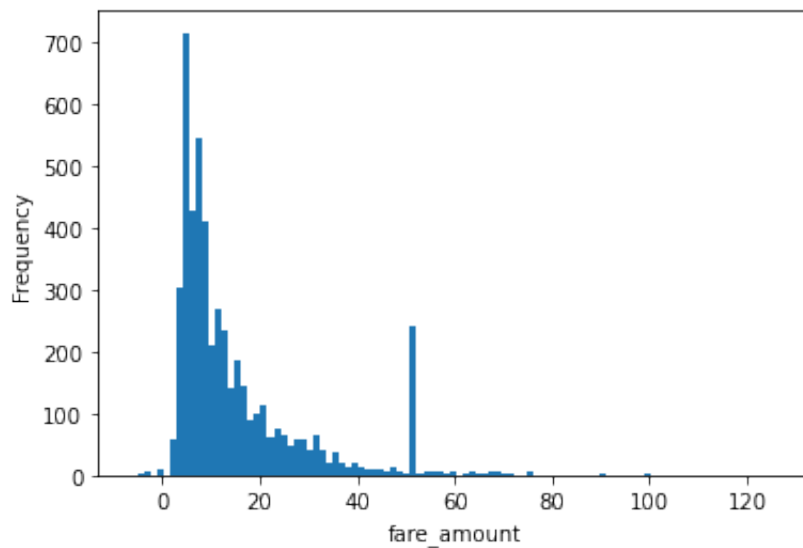


```
In [12]: ## Univariate Distribution: Histogram
```

```
In [13]: df.fare_amount.plot.hist();
```



```
In [14]: ax = df.fare_amount.plot.hist(bins=100)  
ax.set_xlabel('fare_amount');
```



```
In [15]: ## Univariate Distribution: Histogram
```

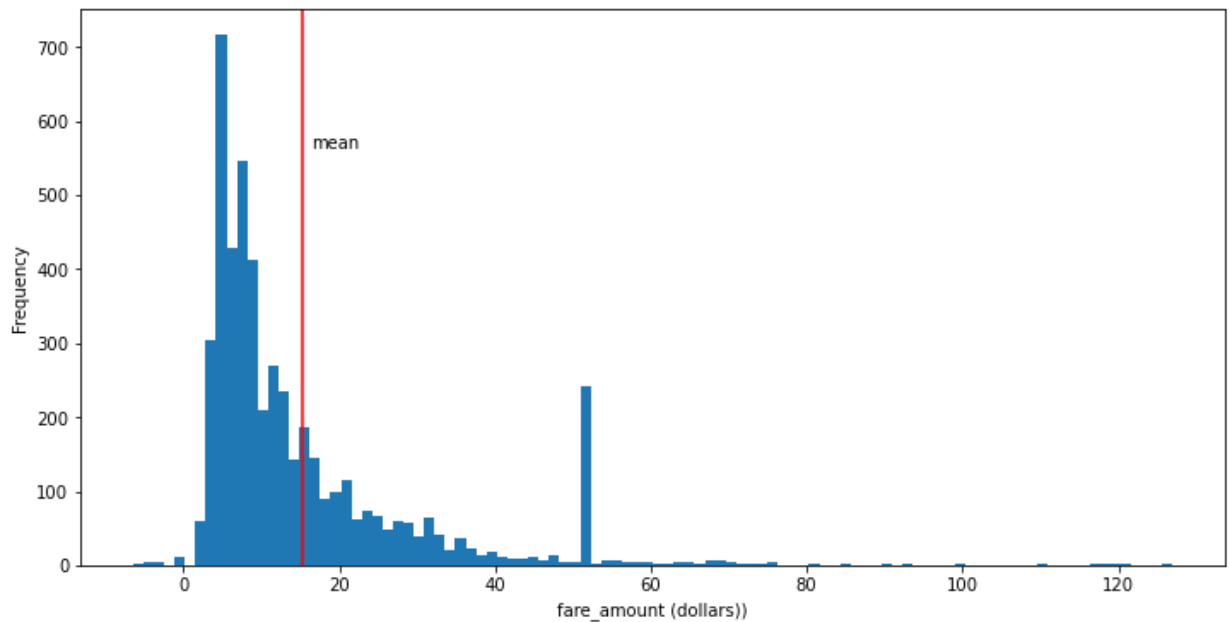
```
In [16]: fig,ax = plt.subplots(1,1,figsize=(12,6));

df.fare_amount.plot.hist(bins=100, ax=ax);
ax.set_xlabel('fare_amount (dollars)');

# add a vertical line
ax.axvline(df.fare_amount.mean(),color='r');

#ax.vlines(df.fare_amount.mean(),*ax.get_ylim(),color='r');

# add some text
ax.text(df.fare_amount.mean()+1,ax.get_ylim()[1]*.75,'mean');
```



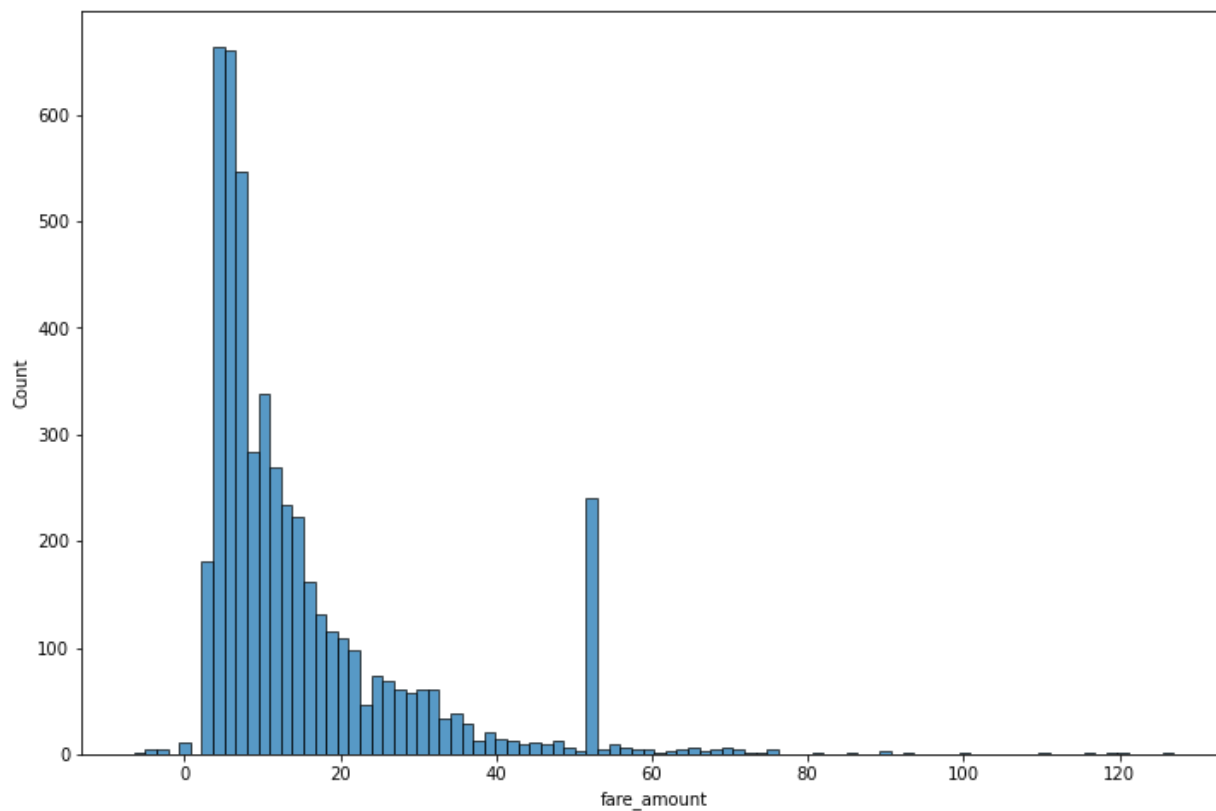
```
In [17]: ## Plotting with Seaborn
```

```
In [18]: import seaborn as sns
sns.__version__
```

```
Out[18]: '0.11.1'
```

```
In [19]: # Univariate Distribution with Seaborn Histplot
```

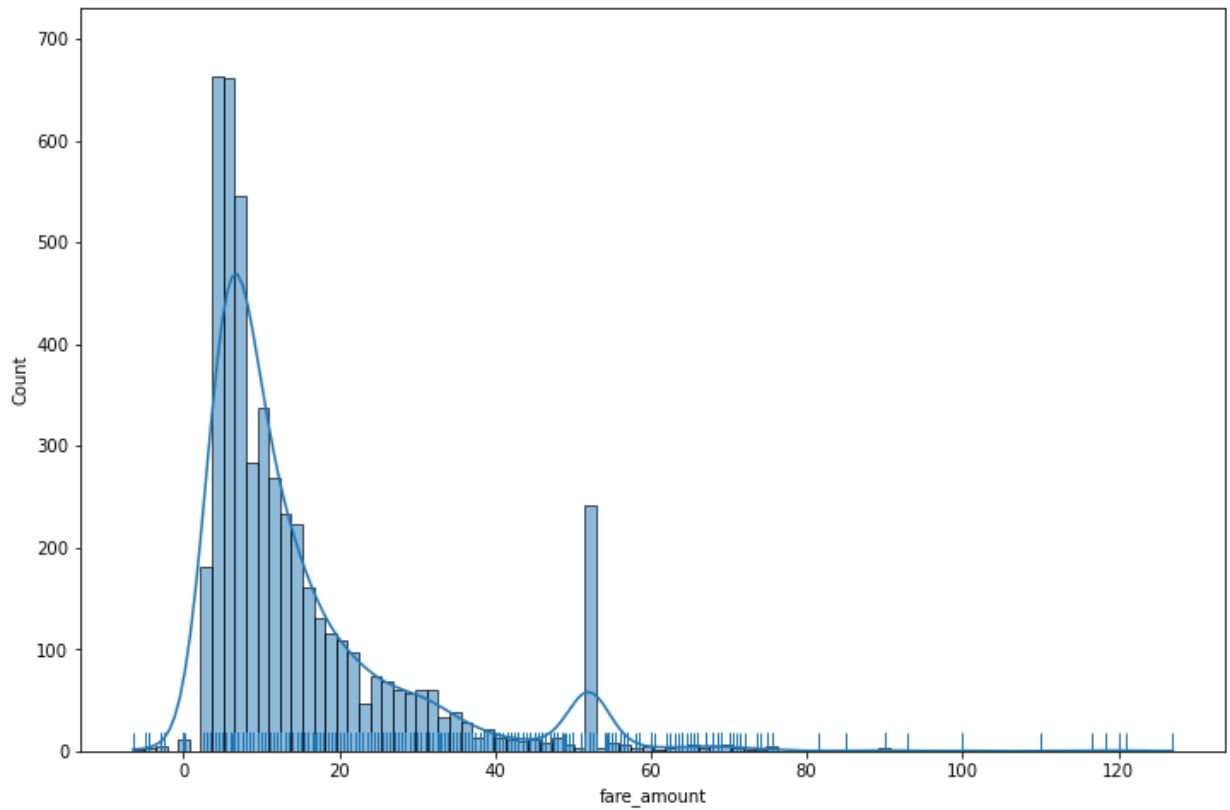
```
In [20]: fig,ax = plt.subplots(1,1,figsize=(12,8))
sns.histplot(df.fare_amount,ax=ax);
```



```
In [21]: # Univariate Distribution: Histogram with KDE and Rugplot
```

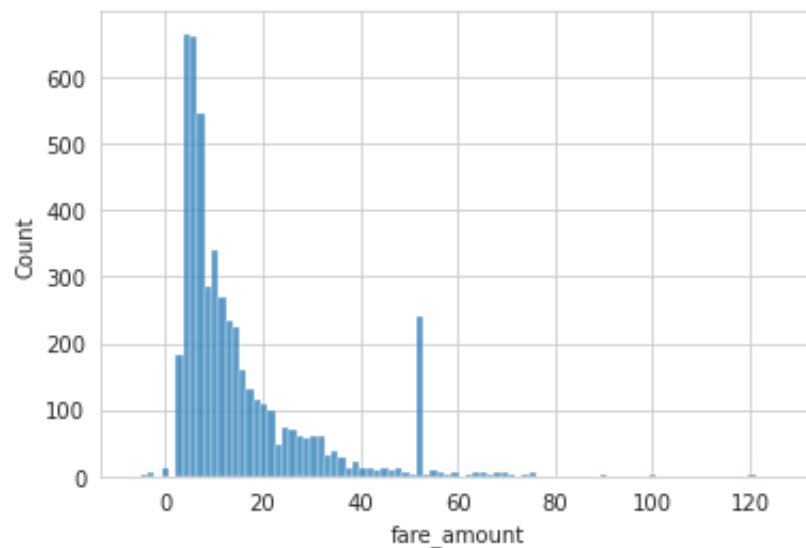
```
In [22]: fig,ax = plt.subplots(1,1,figsize=(12,8))

sns.histplot(df.fare_amount,kde=True,ax=ax);
sns.rugplot(df.fare_amount);
```



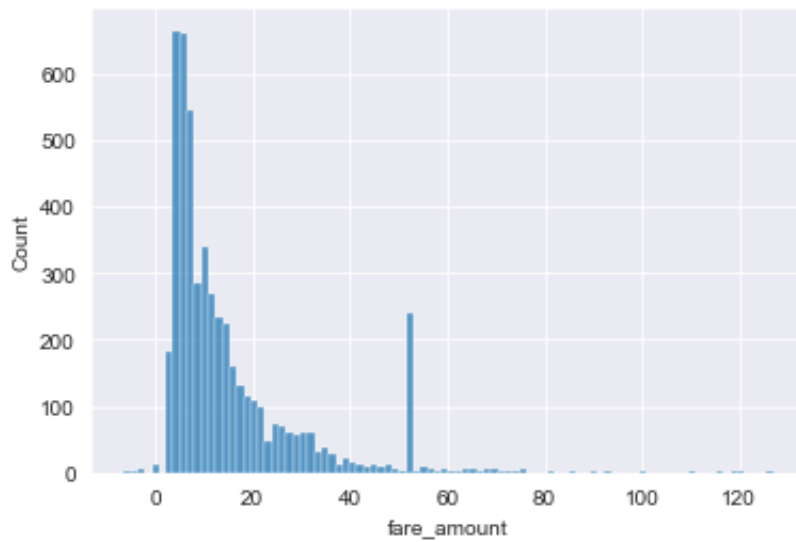
```
In [23]: ## Seaborn Styles
```

```
In [24]: # for a single plot using a context
with sns.axes_style('whitegrid'):
    sns.histplot(df.fare_amount);
```



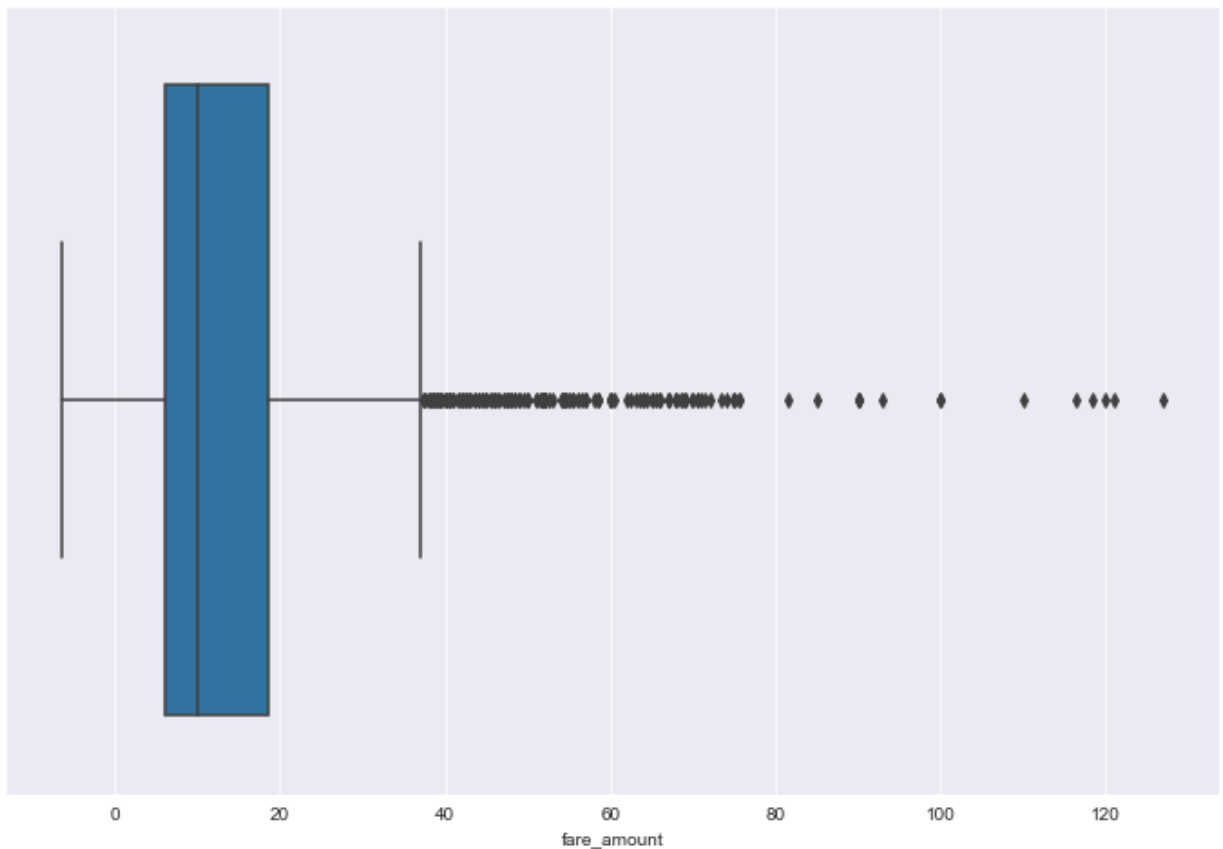
```
In [25]: # set style globally
sns.set_style('darkgrid')
```

```
In [26]: sns.histplot(x=df.fare_amount);
```



```
In [27]: ## Univariate Distributions: Boxplot
```

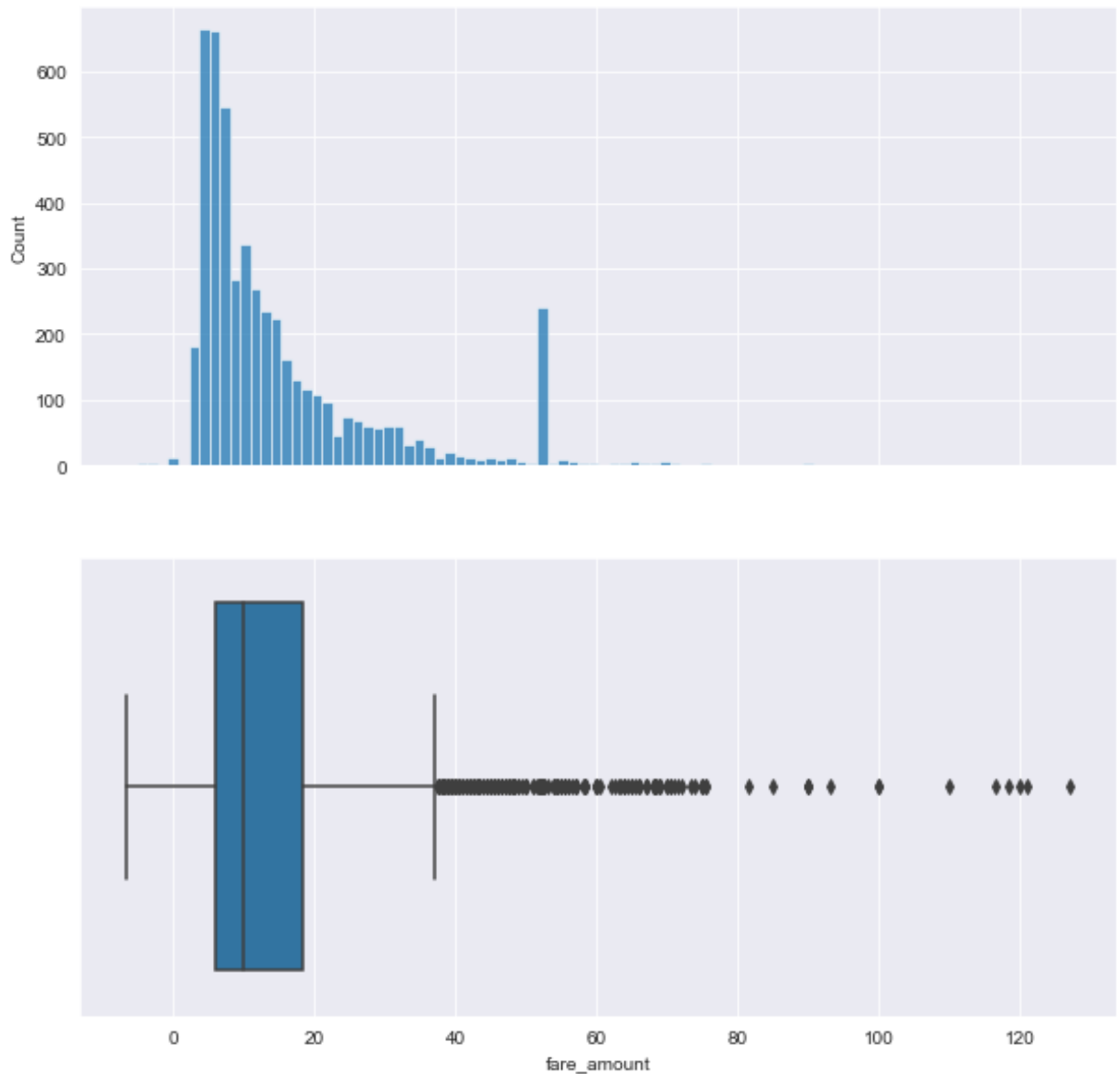
```
In [28]: fig,ax = plt.subplots(1,1,figsize=(12,8))
sns.boxplot(x=df.fare_amount,ax=ax);
```





```
In [29]: ## Combining Plots with Subplots
```

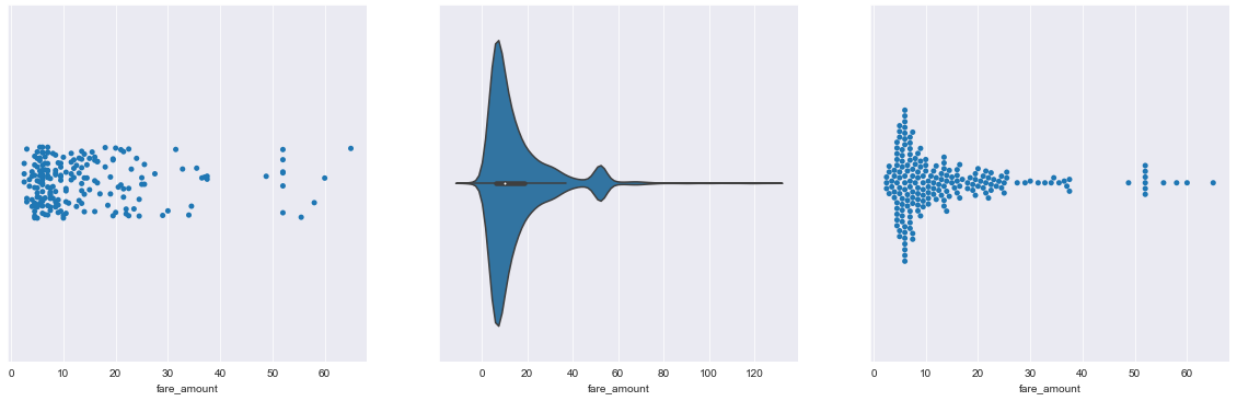
```
In [30]: fig,ax = plt.subplots(2,1,figsize=(10,10), sharex=True)  
  
sns.histplot(x=df.fare_amount, ax=ax[0]);  
sns.boxplot(x=df.fare_amount, ax=ax[1]);
```



```
In [31]: ## Other Univariate Distribution Visualizations
```

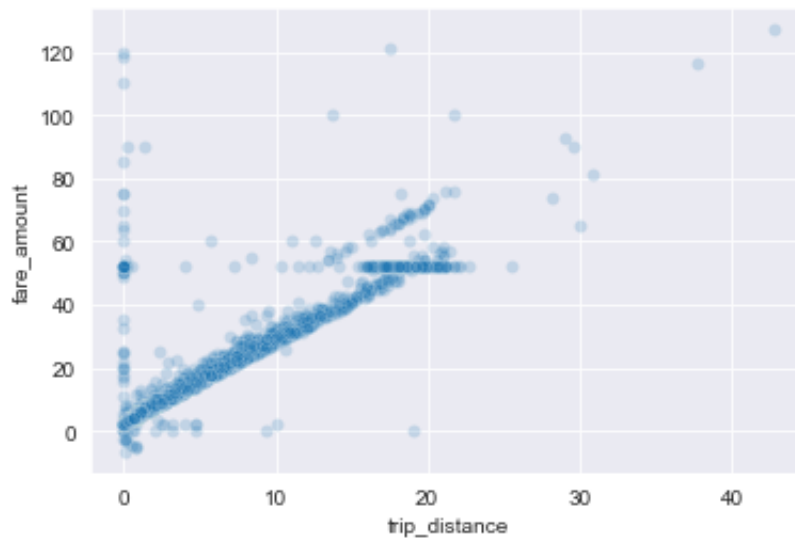
```
In [32]: fig,ax = plt.subplots(1,3,figsize=(20,6))

sns.stripplot(x='fare_amount',data=df[:200],ax=ax[0])
sns.violinplot(x='fare_amount',data=df,ax=ax[1])
sns.swarmplot(x='fare_amount',data=df[:200],ax=ax[2]);
```



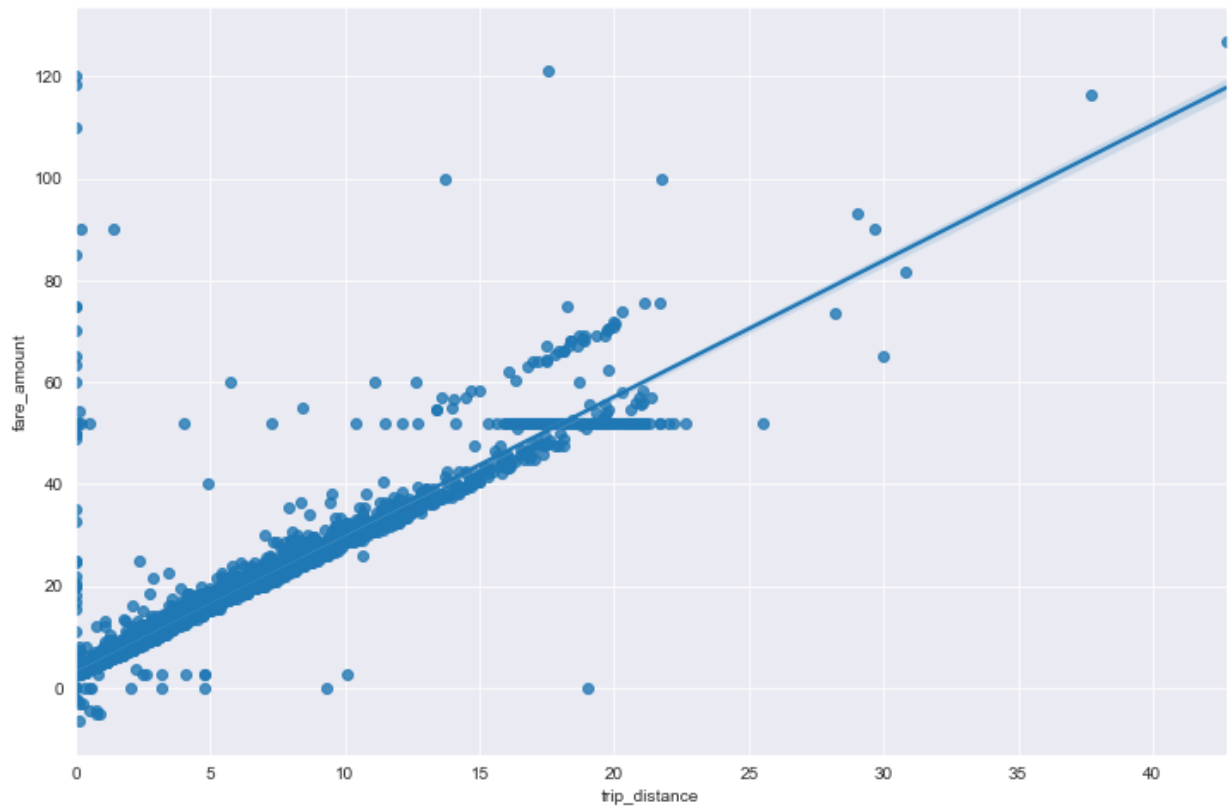
```
In [33]: ## Bivariate: Scatterplot
```

```
In [34]: sns.scatterplot(x='trip_distance',y='fare_amount',data=df,alpha=0.2);
```



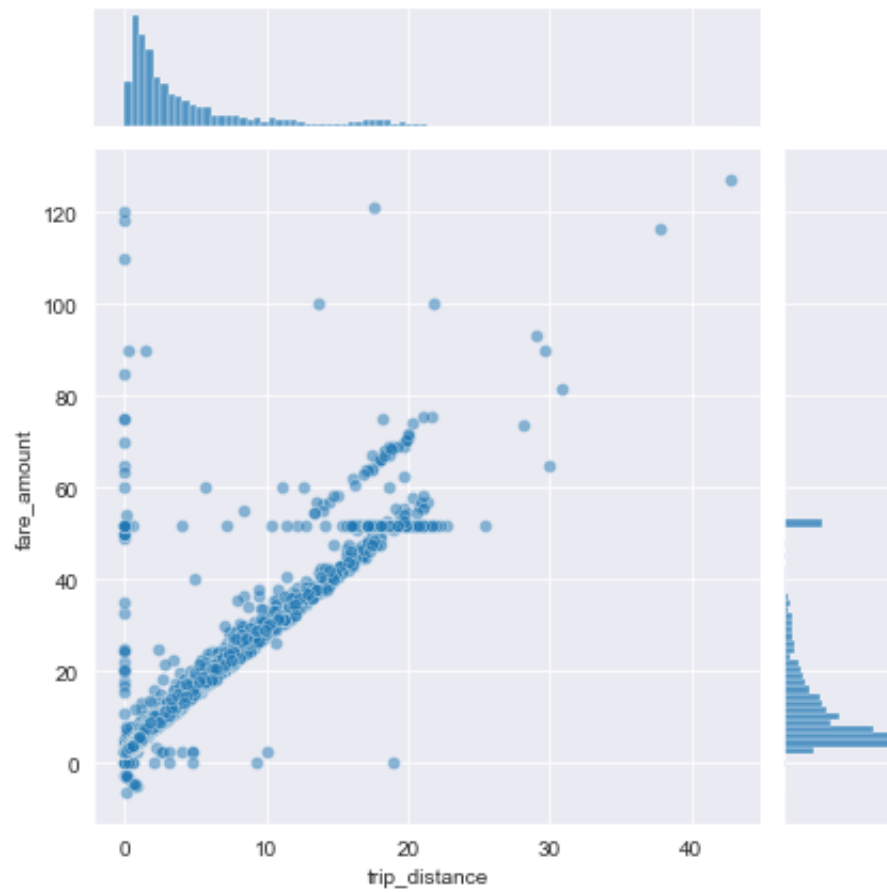
```
In [35]: ## Bivariate: Add Regression Line
```

```
In [36]: fig,ax = plt.subplots(1,1,figsize=(12,8))  
  
sns.regplot(x='trip_distance',y='fare_amount',data=df,ax=ax);
```



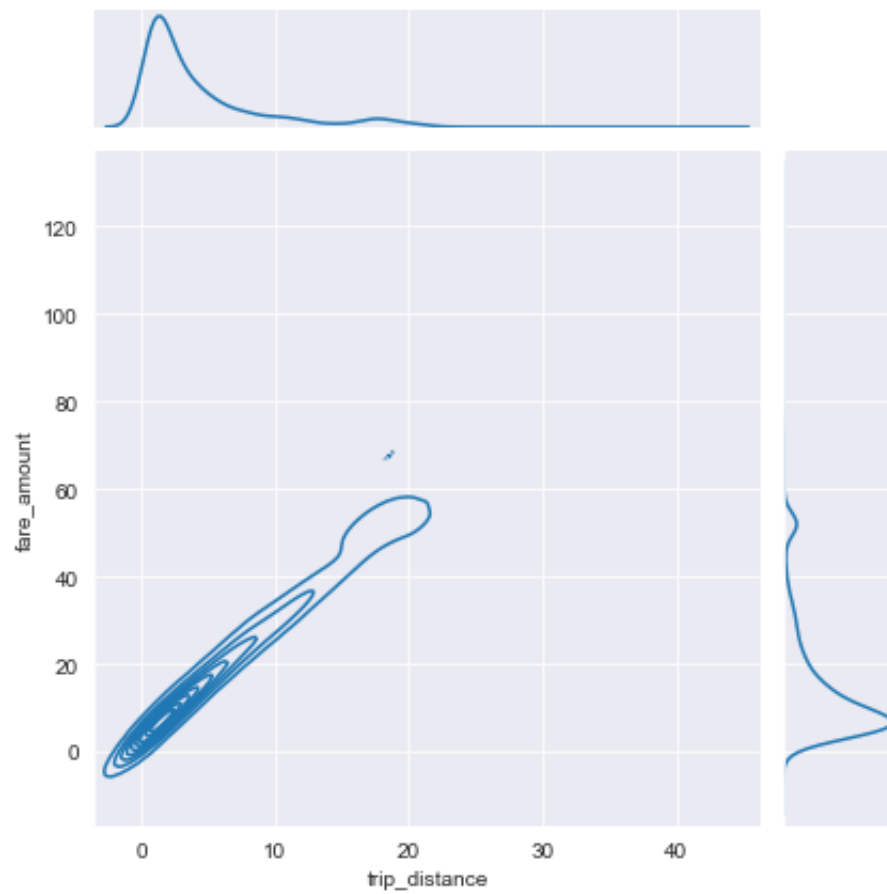
```
In [37]: ## Bivariate: Joint Plot
```

```
In [38]: sns.jointplot(x='trip_distance',y='fare_amount',data=df,alpha=0.5);
```



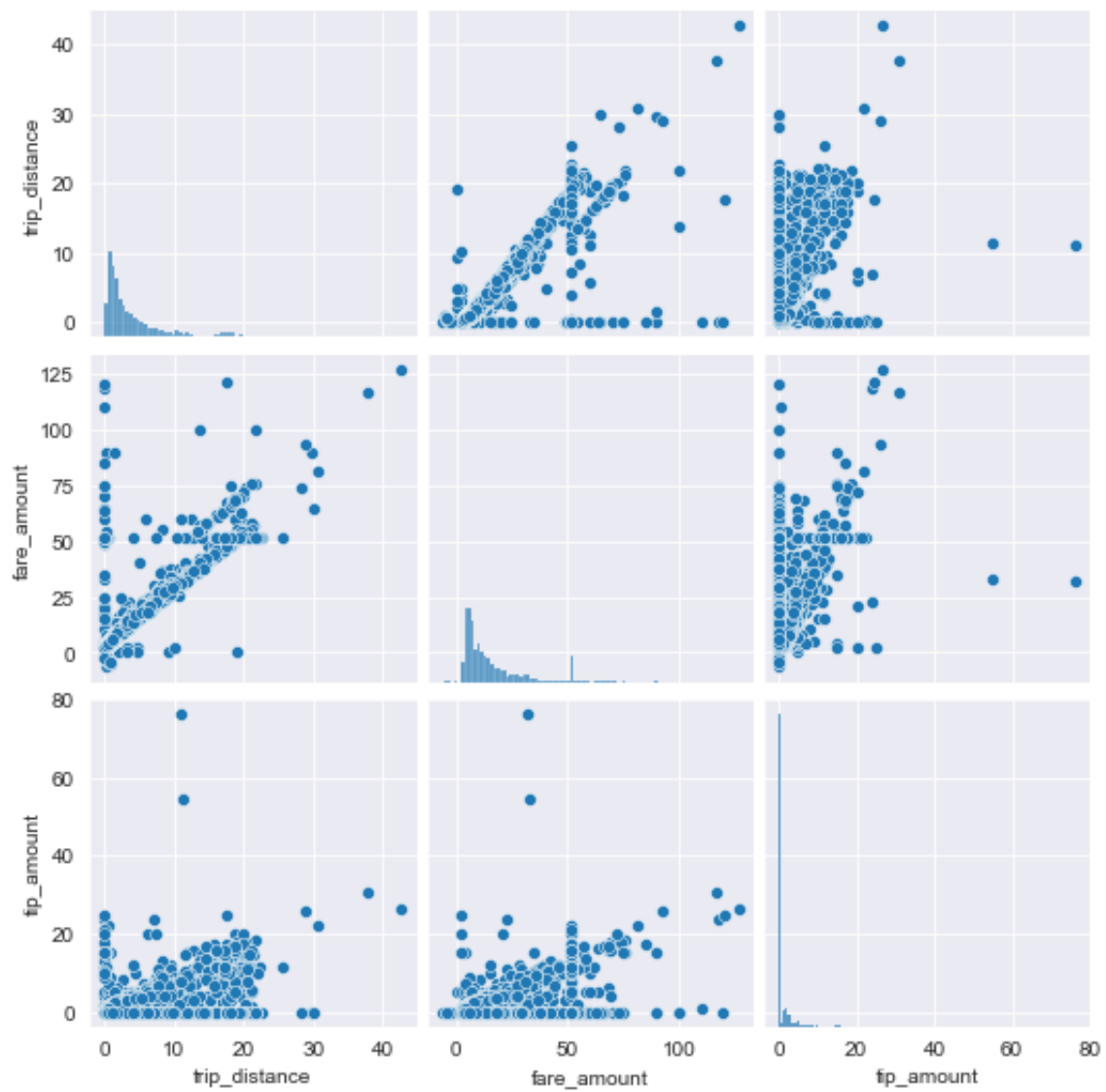
```
In [39]: ## Bivariate: Joint Plot with KDE
```

```
In [40]: sns.jointplot(x='trip_distance', y='fare_amount',  
                      data=df,  
                      kind='kde');
```



```
In [41]: ## Comparing Multiple Variables with pairplot
```

```
In [42]: sns.pairplot(df[['trip_distance', 'fare_amount', 'tip_amount']]);
```



```
In [43]: ## Plotting Numeric and Categorical
```

```
In [44]: fig,ax = plt.subplots(1,1,figsize=(12,8))  
  
sns.barplot(x='payment_type',y='fare_amount',data=df);
```

