Data Preprocessing

In [1]:
```
import pandas as pd
import seaborn as sns
import numpy as np
```

In [2]:
```
df=pd.read_csv('vgsales.csv')
df.shape
```

Out[2]:
```
(16598, 11)
```

In [5]:
```
t2 = pd.read_csv('vgsales.csv')
df=pd.read_csv('vgsales.csv')
df.head()
```

Out[5]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Othe
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	

In [10]:
```
from sklearn.preprocessing import LabelEncoder
labelencoder_df=LabelEncoder()
df['Genre']=labelencoder_df.fit_transform(df['Genre'])
df['Platform']=labelencoder_df.fit_transform(df['Platform'])
df.head()
```

Out[10]:

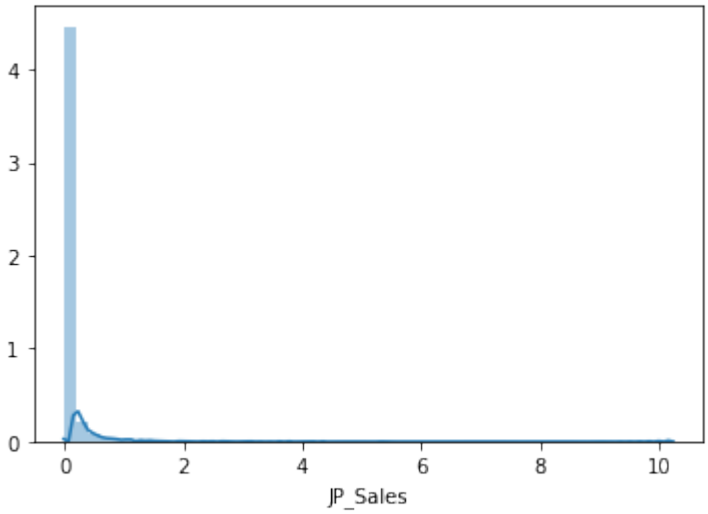
	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_
0	1	Wii Sports	26	2006.0	10	Nintendo	41.49	29.02	3.77	
1	2	Super Mario Bros.	11	1985.0	4	Nintendo	29.08	3.58	6.81	
2	3	Mario Kart Wii	26	2008.0	6	Nintendo	15.85	12.88	3.79	
3	4	Wii Sports Resort	26	2009.0	10	Nintendo	15.75	11.01	3.28	
4	5	Pokemon Red/Pokemon Blue	5	1996.0	7	Nintendo	11.27	8.89	10.22	

In [12]:
```
df.shape
```

Out[12]:
```
(16598, 11)
```

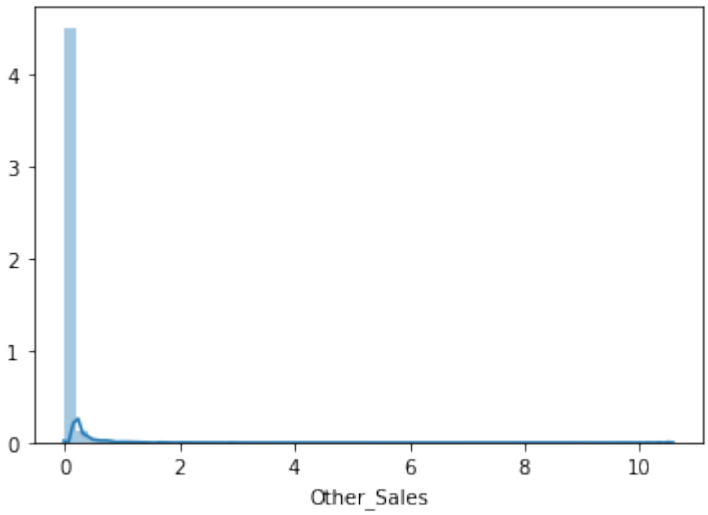
In [13]:
```
sns.distplot(df['JP_Sales'])
```

Out[13]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x238fed0e5f8>
```



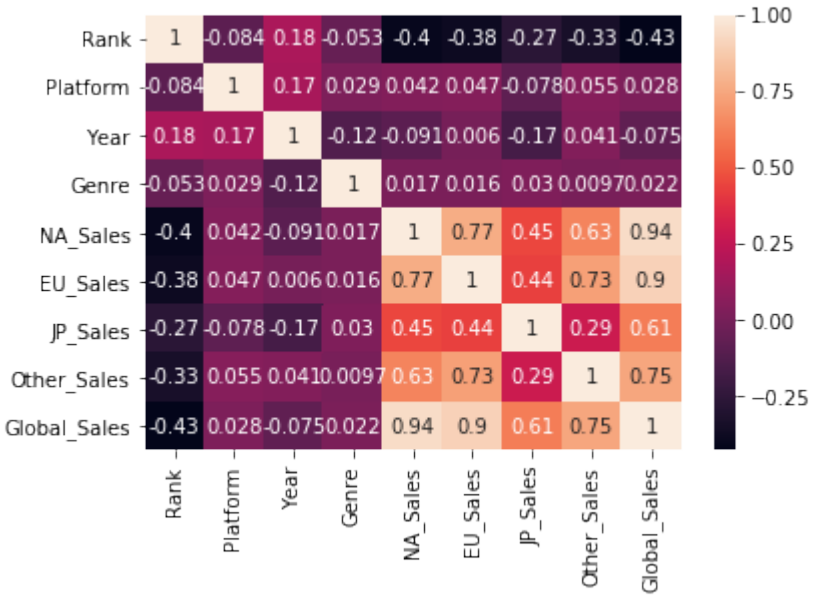
In [14]:
```
sns.distplot(df['Other_Sales'])
```

Out[14]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x238fec22470>
```



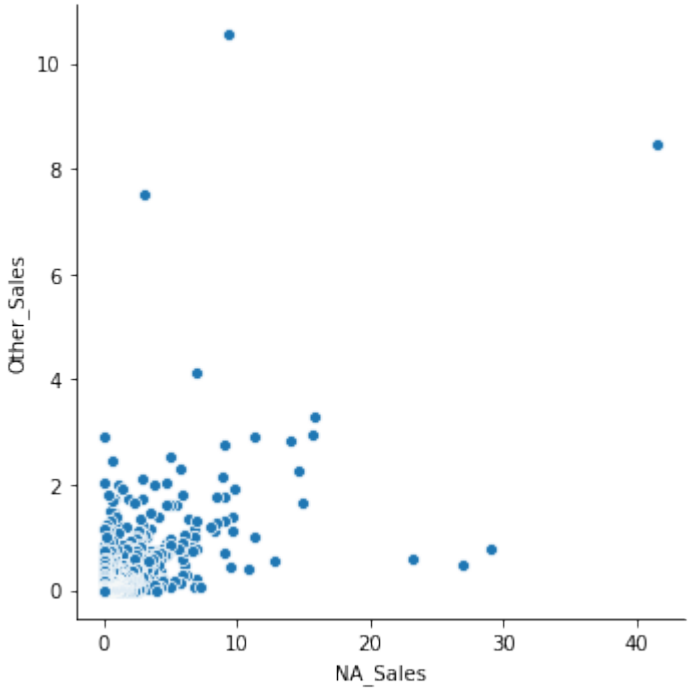
In [15]:
```
corr_matrix=df.corr()
sns.heatmap(data=corr_matrix, annot=True)
```

Out[15]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x238feeedd68>
```



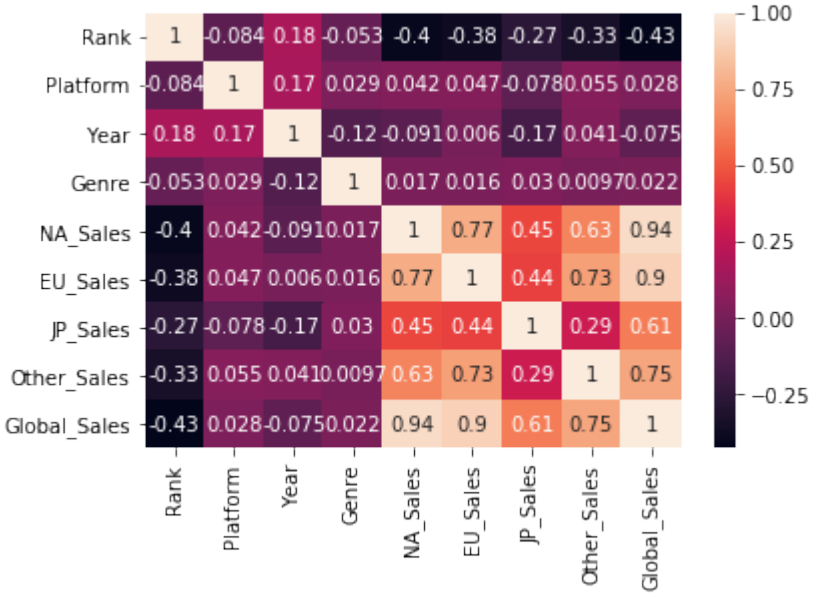
In [16]:
```
sns.relplot(x='NA_Sales', y='Other_Sales', data=df)
```

Out[16]:
```
<seaborn.axisgrid.FacetGrid at 0x238ff054240>
```



In [11]:
```
corr_matrix=df.corr()
sns.heatmap(data=corr_matrix, annot=True)
```

Out[11]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x238fe8f6f98>
```


Rank 1 -0.084 0.18 -0.053 -0.4 -0.38 -0.27 -0.33 -0.43
Platform -0.084 1 0.17 0.029 0.042 0.047 -0.078 0.055 0.028
Year 0.18 0.17 1 -0.12 -0.091 0.006 -0.17 0.041 -0.075
Genre -0.053 0.029 -0.12 1 0.017 0.016 0.03 0.0097 0.022
NA_Sales -0.4 0.042 -0.091 0.017 1 0.77 0.45 0.63 0.94
EU_Sales -0.38 0.047 0.006 0.016 0.77 1 0.44 0.73 0.9
JP_Sales -0.27 -0.078 -0.17 0.03 0.45 0.44 1 0.29 0.61
Other_Sales -0.33 0.055 0.041 0.0097 0.63 0.73 0.29 1 0.75
Global_Sales -0.43 0.028 -0.075 0.022 0.94 0.9 0.61 0.75 1
Rank
Platform
Year
Genre
NA_Sales
EU_Sales
JP_Sales
Other_Sales
Global_Sales
1.00
0.75
0.50
0.25
0.00
−0.25

In [ ]: