Colab: https://colab.research.google.com/drive/1QrclVIfhdW2b2IiWKk2atg_c-8e7rri_?
usp=sharing

```
!wget https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/021/299/origin
```

```
--2022-12-21 15:50:24--  https://d2beiqkhq929f0.cloudfront.net/public_assets/a
Resolving d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)... 18.
Connecting to d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)|18
HTTP request sent, awaiting response... 200 OK
Length: 2041483 (1.9M) [text/plain]
Saving to: 'vgsales.csv'

vgsales.csv          100%[===================>]   1.95M  --.-KB/s    in 0.07s

2022-12-21 15:50:25 (29.7 MB/s) - 'vgsales.csv' saved [2041483/2041483]
```

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
data = pd.read_csv('vgsales.csv')
data.head()
```

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales
0	2061	1942	NES	1985.0	Shooter	Capcom	4.569217	3.033887
1	9137	¡Shin Chan Flipa en colores!	DS	2007.0	Platform	505 Games	2.076955	1.493442
2	14279	.hack: Sekai no Mukou ni + Versus	PS3	2012.0	Action	Namco Bandai Games	1.145709	1.762339
		.hack//G.U.			Role	Namco		

```python
top3_pub = data['Publisher'].value_counts().index[:3]
top3_gen = data['Genre'].value_counts().index[:3]
top3_plat = data['Platform'].value_counts().index[:3]
top3_data = data.loc[(data["Publisher"].isin(top3_pub)) & (data["Platform"].isin(to
top3_data
```
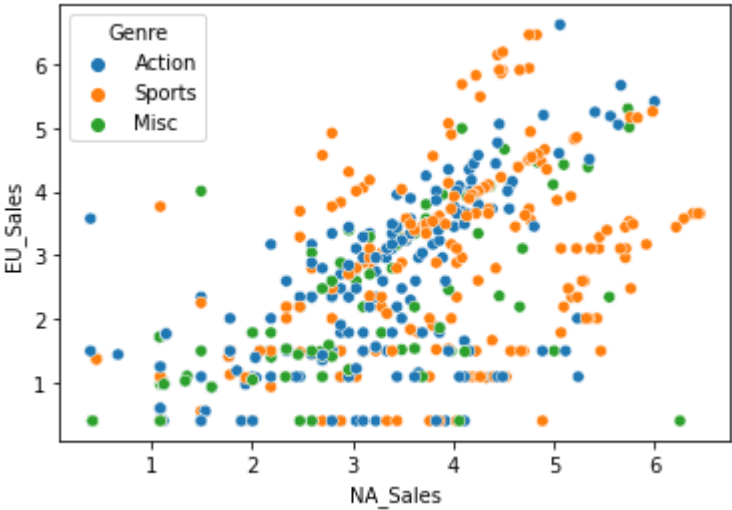
	Rank	Name	Platform	Year	Genre
2	14279	.hack: Sekai no Mukou ni + Versus	PS3	2012.0	Action
13	2742	[Prototype 2]	PS3	2012.0	Action
16	1604	[Prototype]	PS3	2009.0	Action
19	1741	007: Quantum of Solace	PS3	2008.0	Action
21	4501	007: Quantum of Solace	PS2	2008.0	Action
...	...	...	...	...	...

```
# Multivariate - 3 variables
```

```
# NNC
# CCN
# NNN
# CCC
```

```
sns.scatterplot(x="NA_Sales", y="EU_Sales", hue="Genre", data=top3_data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9dad05e6a0>
```



```
# CCN
# Global Sales for each publisher but seperated for different genres?
```
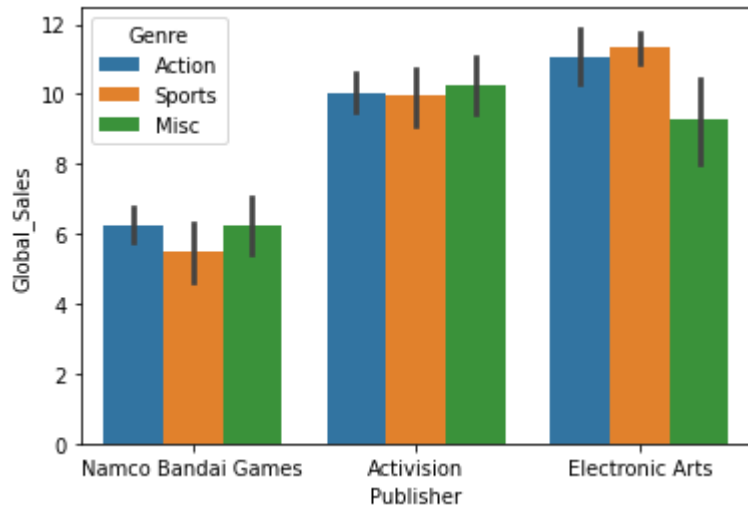
```
sns.boxplot(x="Publisher", y="Global_Sales", hue="Genre", data=top3_data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9dacf322e0>
```



```
sns.barplot(x="Publisher", y="Global_Sales", hue="Genre", data=top3_data, estimator
```
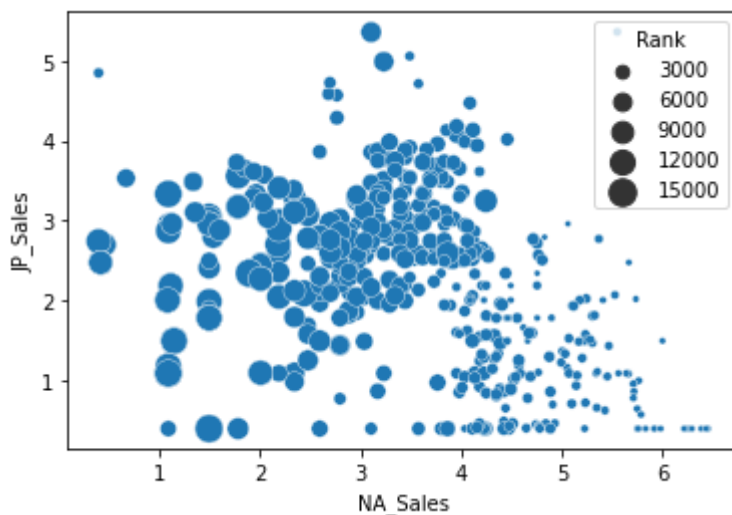
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9dac914f40>
```



```
# NNN
# NA, EU, Rank (Size of dot)
```

```
sns.scatterplot(x="NA_Sales", y="JP_Sales", size="Rank", data=top3_data, sizes=(10,
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9dac68b250>
```



```
# CCC - No generic plot
```
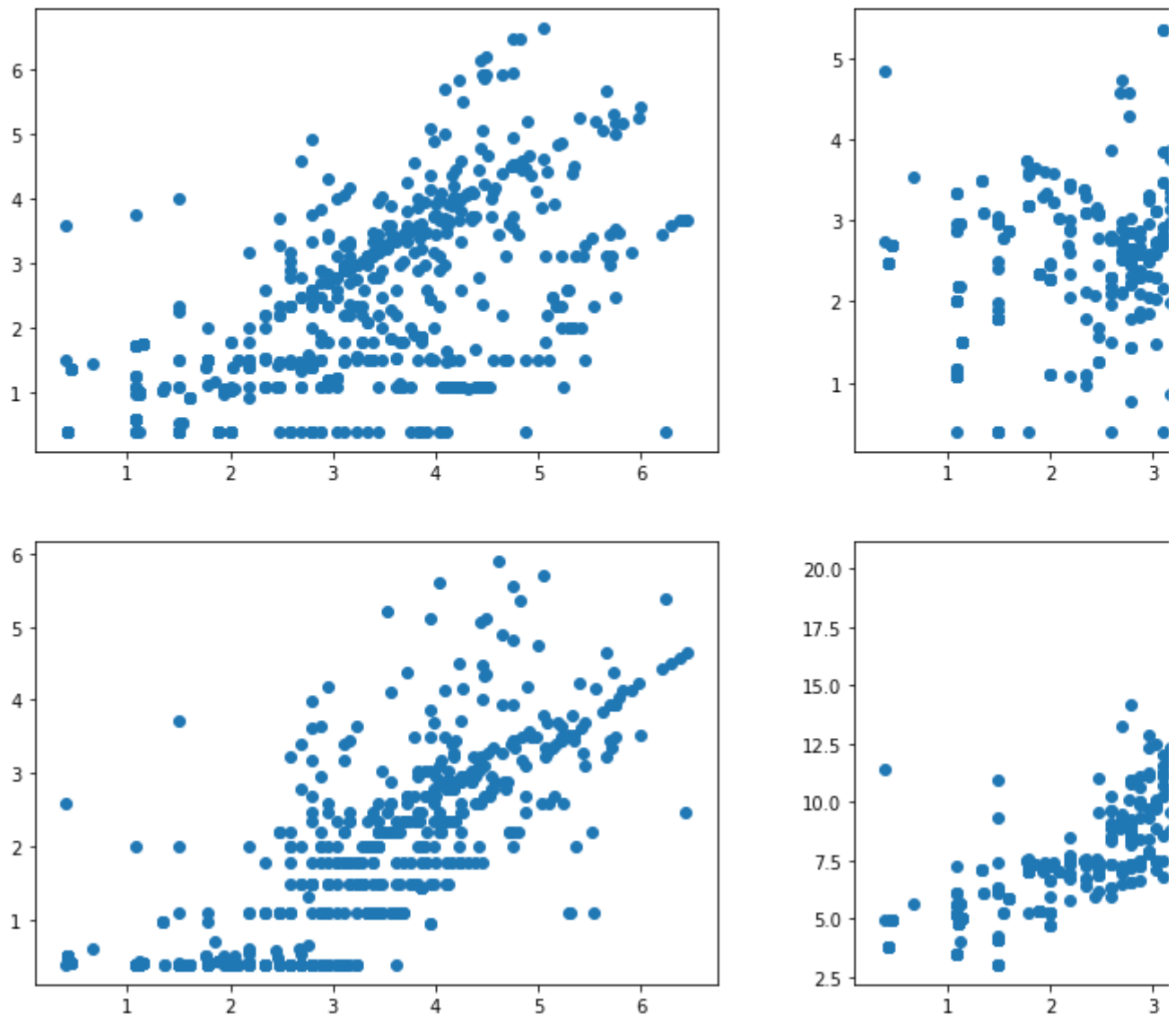
```
# Subplots
```

```
fig, ax = plt.subplots(2, 2, figsize=(15, 10))
ax[0,0].scatter(top3_data['NA_Sales'], top3_data['EU_Sales'])
```

```
ax[0,1].scatter(top3_data['NA_Sales'], top3_data['JP_Sales'])
ax[1,0].scatter(top3_data['NA_Sales'], top3_data['Other_Sales'])
ax[1,1].scatter(top3_data['NA_Sales'], top3_data['Global_Sales'])
fig.suptitle("Relation of NA Sales with other region")
```

```
    Text(0.5, 0.98, 'Relation of NA Sales with other region')
```



```
fig = plt.figure(figsize=(15, 10))

plt.subplot(2, 2, 1)
sns.scatterplot(x='NA_Sales', y='EU_Sales', data=top3_data)

plt.subplot(2, 2, 2)
sns.scatterplot(x='NA_Sales', y='JP_Sales', data=top3_data)
plt.title("Poonam's query")

plt.subplot(2, 2, 3)
sns.scatterplot(x='NA_Sales', y='Other_Sales', data=top3_data)

plt.subplot(2, 2, 4)
```
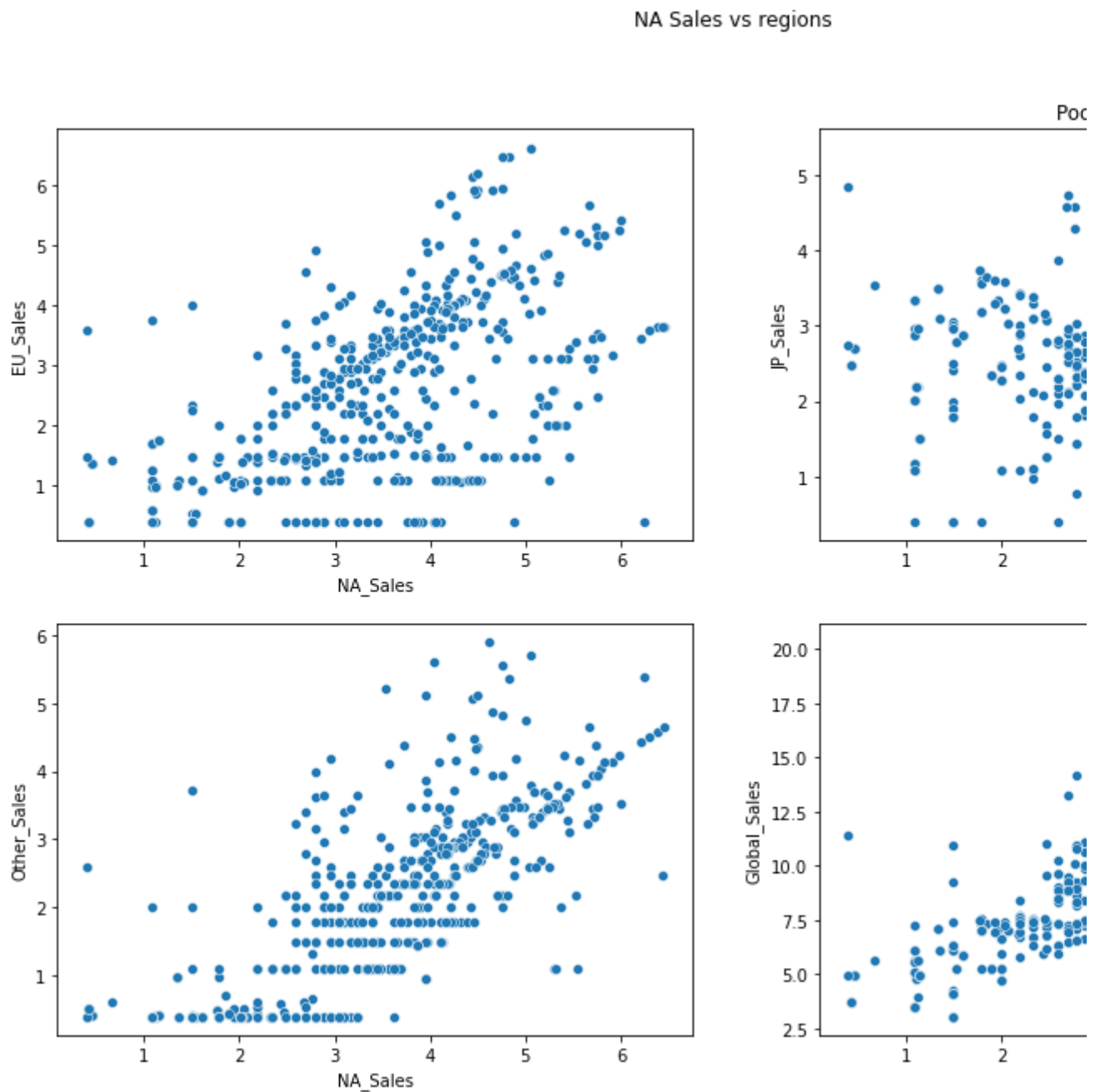
```
sns.scatterplot(x='NA_Sales', y='Global_Sales', data=top3_data)

fig.suptitle("NA Sales vs regions")
plt.show()
```

NA Sales vs regions


```
fig = plt.figure(figsize=(20,12))

# Using a 2x3 subplot
plt.subplot(2, 3, 1)
sns.scatterplot(x='NA_Sales', y='EU_Sales', data=top3_data)
plt.title('NA vs EU Sales', fontsize=12)
plt.xlabel('NA', fontsize=12)
plt.ylabel('EU', fontsize=12)



plt.subplot(2, 3, 3)
sns.scatterplot(x='NA_Sales', y='JP_Sales', data=top3_data, color='red')
```

```python
plt.title('NA vs JP Sales', fontsize=12)
plt.xlabel('NA', fontsize=12)
plt.ylabel('JP', fontsize=12)


plt.subplot(2, 3, 4)
sns.scatterplot(x='NA_Sales', y='Other_Sales', data=top3_data, color='green')
plt.title('NA vs Other Region Sales', fontsize=12)
plt.xlabel('NA', fontsize=12)
plt.ylabel('Other', fontsize=12)


plt.subplot(2, 3, 6)
sns.scatterplot(x='NA_Sales', y='Global_Sales', data=top3_data, color='orange')
plt.title('NA vs Global Sales', fontsize=12)
plt.xlabel('NA', fontsize=12)
plt.ylabel('Global', fontsize=12)



# Countplot of publishers
plt.subplot(1,3,2)
sns.countplot(x='Publisher', data=top3_data)
plt.title('Count of games by each Publisher', fontsize=12)
plt.xlabel('Publisher', fontsize=12)
plt.ylabel('Count of games', fontsize=12)

fig.suptitle("NA Sales vs regions",fontsize=20)
plt.show()
```
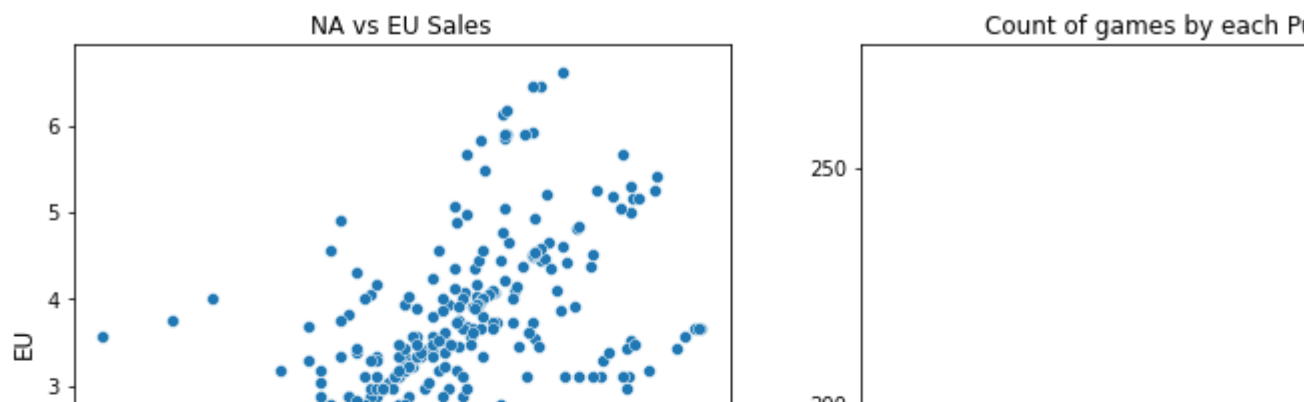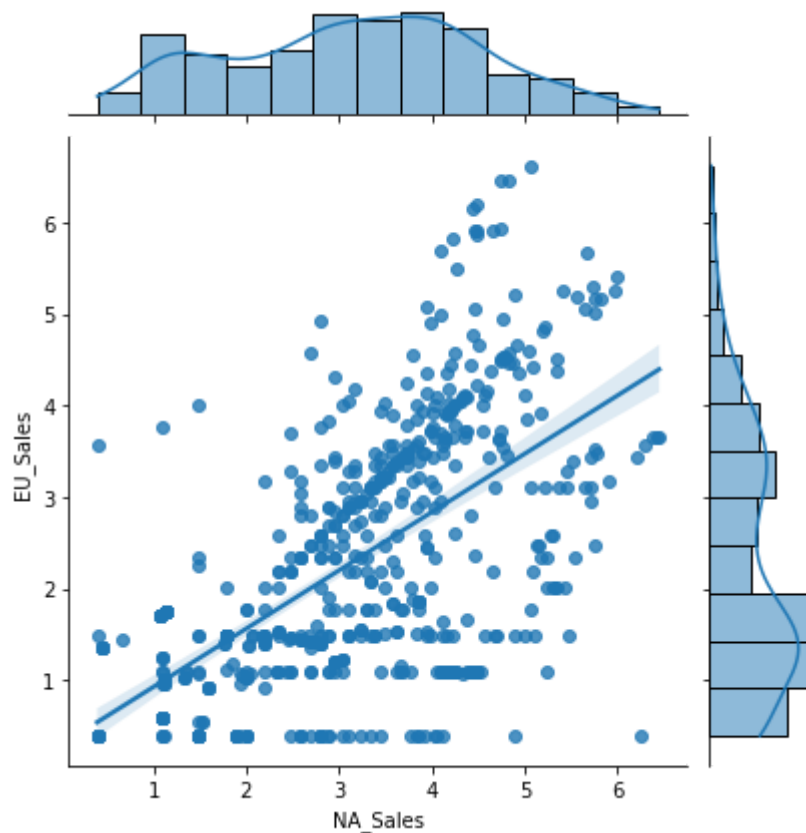
NA Sales vs regio


```
# joint plot
# pairplot
# heatmaps
```
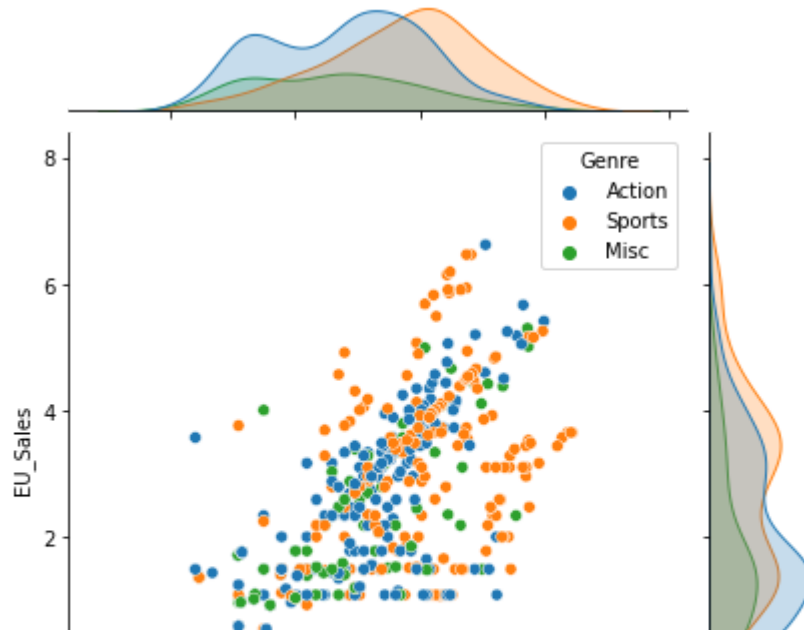
```
sns.jointplot(x="NA_Sales", y="EU_Sales", data=top3_data, kind="reg")
```

    <seaborn.axisgrid.JointGrid at 0x7f9da81f13d0>


```
sns.jointplot(x="NA_Sales", y="EU_Sales", data=top3_data,  hue="Genre")
```
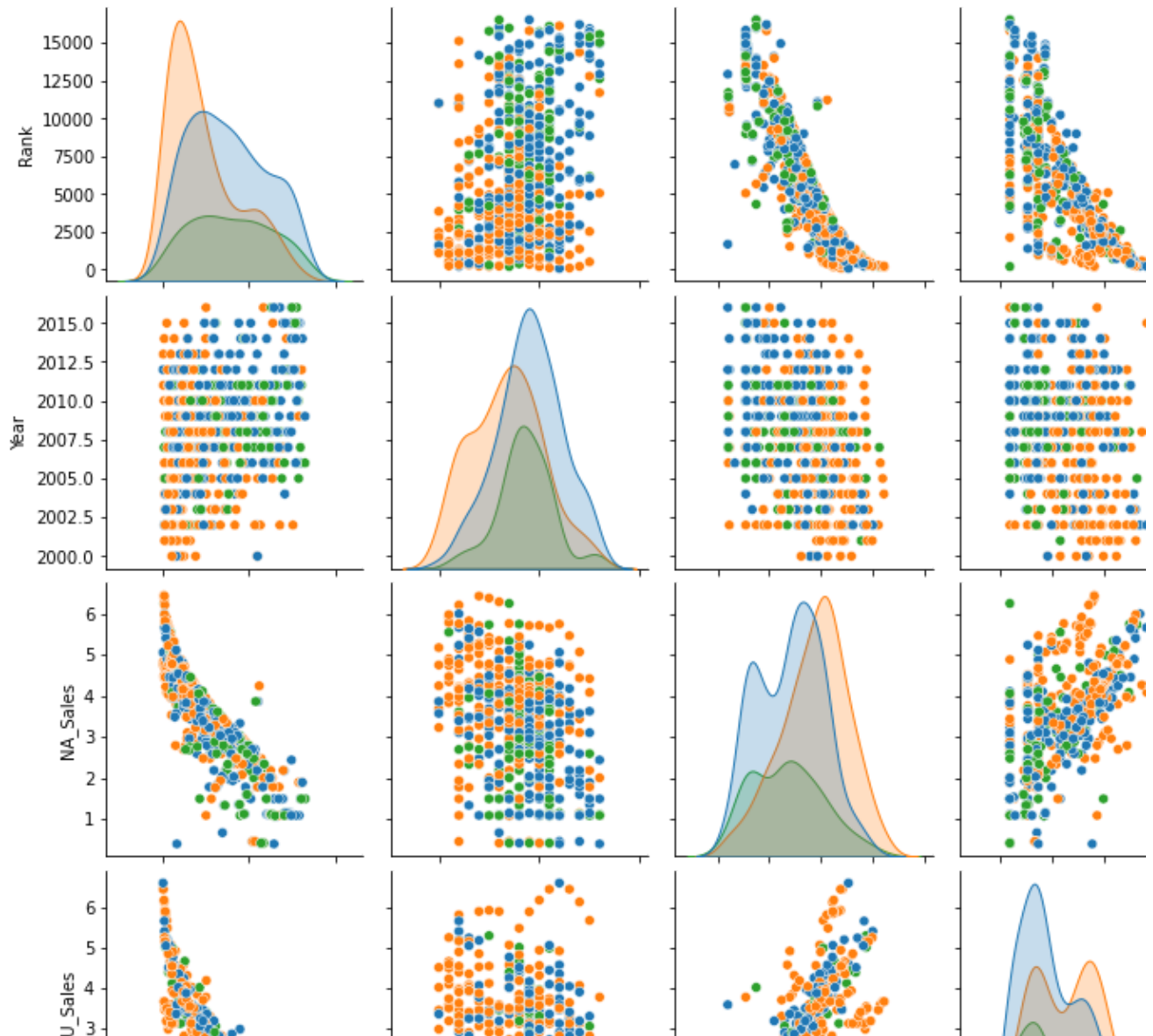
```
<seaborn.axisgrid.JointGrid at 0x7f9da812cf70>
```
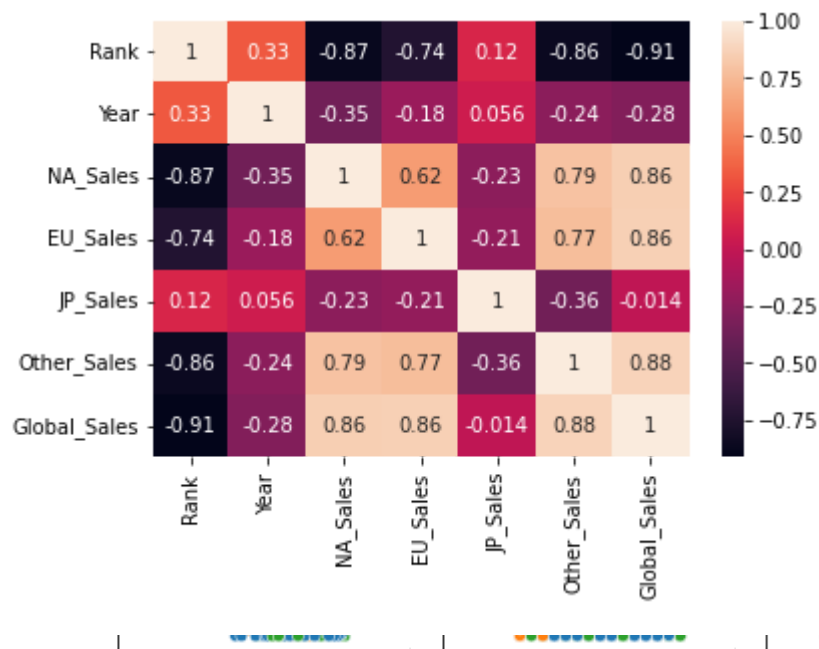


```
sns.pairplot(data=top3_data, hue="Genre")
```
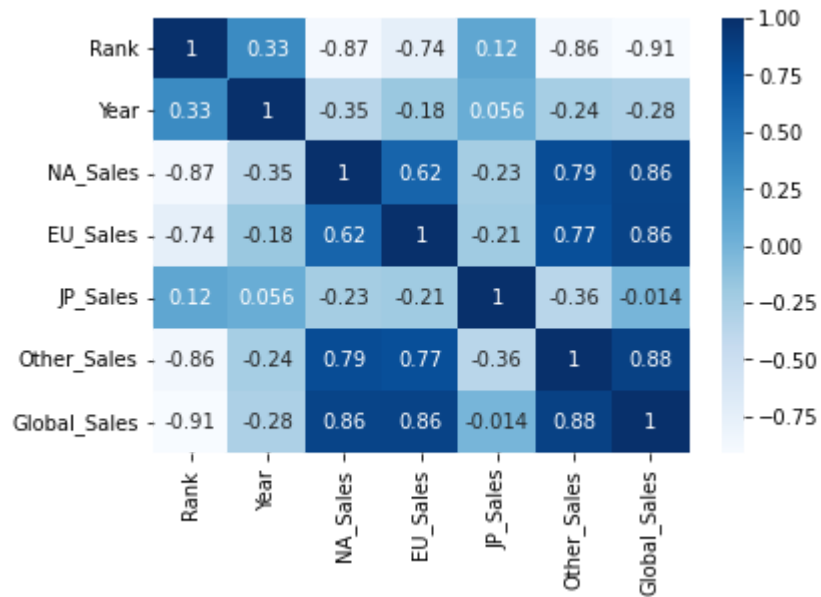
```
<seaborn.axisgrid.PairGrid at 0x7f9da2251e20>
```



```python
sns.heatmap(top3_data.corr(), annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9da1086730>
```



```python
sns.heatmap(top3_data.corr(), annot=True, cmap="Blues")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9da0d53430>
```

✓ 0s   completed at 22:43   ● ✕