Exploratory Data Analysis of James Bond Movies

In [1]:

```
import pandas as pd
pd.options.mode.chained_assignment = None
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import plotly.express as px
import os

sns.set_style('darkgrid')
matplotlib.rcParams['font.size']=14
matplotlib.rcParams['figure.figsize']=(9,5)
matplotlib.rcParams['figure.facecolor']='#00000000'
```

In [2]:

```
raw_bond_df = pd.read_csv('data.csv')
```

In [3]:

```
raw_bond_df.head()
```

Out[3]:

	Year	Movie	Bond	Director	Composer	Writer	Cinematographer	Depicted_Film_Loc	Shooting_Loc	Bond_Car_MFG	...	Fil
0	1962	Dr. No	Sean Connery	Terence Young	Monty Norman	Richard Maibaum, Johanna Harwood & Berkely Mather	Ted Moore	Great Britain, Jamaic	England, Jamaica	Sunbeam	...	
1	1963	From Russia with Love	Sean Connery	Terence Young	John Barry	Richard Maibaum & Johanna Harwood	Ted Moore	United Kingdom, Great Britain, Turkey, Croatia...	England, Scotland, Italy, Switzerland, Turkey	Bently	...	
2	1964	Goldfinger	Sean Connery	Guy Hamilton	John Barry	Richard Maibaum & Paul Dehn	Ted Moore	United States, Great Britain, Switzerland	England, Switzerland, United States	Aston Martin	...	
3	1965	Thunderball	Sean Connery	Terence Young	John Barry	Richard Maibaum & John Hopkins	Ted Moore	France, Great Britain, Bahamas, United States	England, France, Bahamas, United States	Aston Martin	...	
4	1967	You Only Live Twice	Sean Connery	Lewis Gilbert	John Barry	Roald Dahl	Freddie Young	United States, Russia, Kazakhstan, Norway, Japan	Japan, Spain, Norway	Toyota	...	

5 rows × 27 columns

In [4]:

```
raw_bond_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 27 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Year               24 non-null     int64  
 1   Movie              24 non-null     object 
 2   Bond               24 non-null     object 
 3   Director           24 non-null     object 
 4   Composer           24 non-null     object 
 5   Writer             24 non-null     object 
 6   Cinematographer    24 non-null     object 
 7   Depicted_Film_Loc  24 non-null     object 
 8   Shooting_Loc       24 non-null     object 
 9   Bond_Car_MFG       24 non-null     object 
 10  Bond_Girl_Nat      24 non-null     object 
 11  US_Gross           24 non-null     int64  
 12  US_Adj             24 non-null     int64  
 13  World_Gross        24 non-null     int64  
 14  World_Adj          24 non-null     int64  
 15  Budget             24 non-null     int64  
 16  Budget_Adj         24 non-null     int64  
 17  Film_Length        24 non-null     int64  
 18  Avg_User_IMDB      24 non-null     float64
 19  Avg_User_Rtn_Tom   24 non-null     float64
 20  Conquests          24 non-null     int64  
 21  Martinis           24 non-null     int64  
 22  BJB                24 non-null     int64  
 23  Kills_Bond         24 non-null     int64  
 24  Kills_Others       24 non-null     int64  
 25  Top_100            24 non-null     int64  
 26  Video_Game         24 non-null     int64  
dtypes: float64(2), int64(15), object(10)
memory usage: 5.2+ KB
```

In [5]:

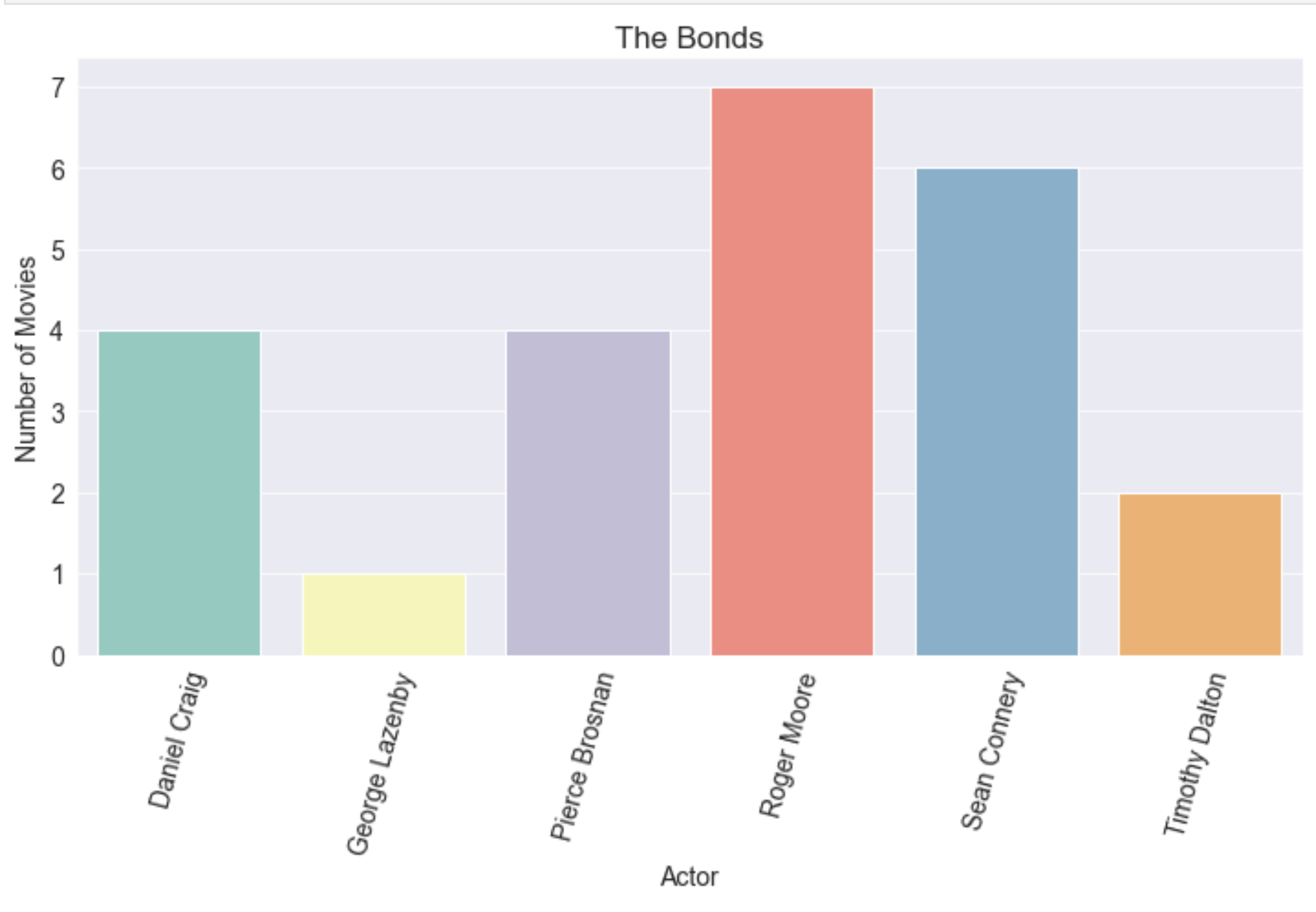
```
#data is already clean
bond_df = raw_bond_df.copy()
```

In [6]:

```
the_bonds = bond_df.groupby('Bond').Movie.count()
```

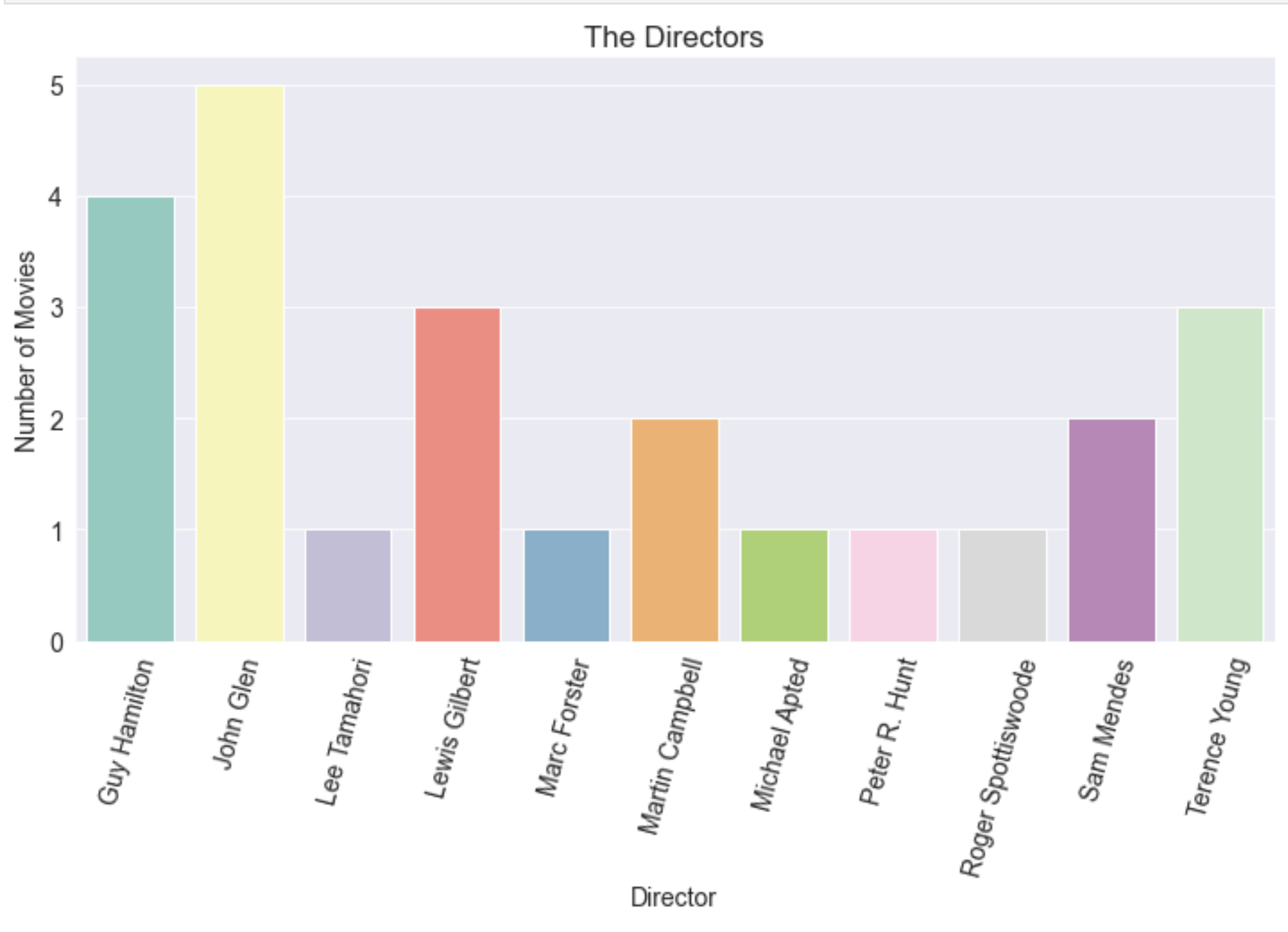
In [7]:

```
plt.figure(figsize=(12,6))
plt.xticks(rotation=75)
plt.title('The Bonds')
g = sns.barplot(x=the_bonds.index, y=the_bonds, palette='Set3')
g.set(xlabel = "Actor", ylabel="Number of Movies");
```



In [8]:

```
the_directors = bond_df.groupby('Director').Movie.count()
plt.figure(figsize=(12,6))
plt.xticks(rotation=75)
plt.title('The Directors')
g = sns.barplot(x=the_directors.index, y=the_directors, palette='Set3')
g.set(xlabel = "Director", ylabel="Number of Movies");
```



In [9]:

```
ratings = bond_df[['Movie', 'Bond', 'Year', 'Avg_User_IMDB', 'Avg_User_Rtn_Tom', 'Budget_Adj', 'World_Adj']]
```

In [10]:

```
ratings.World_Adj = ratings.World_Adj / 1000
```

In [11]:

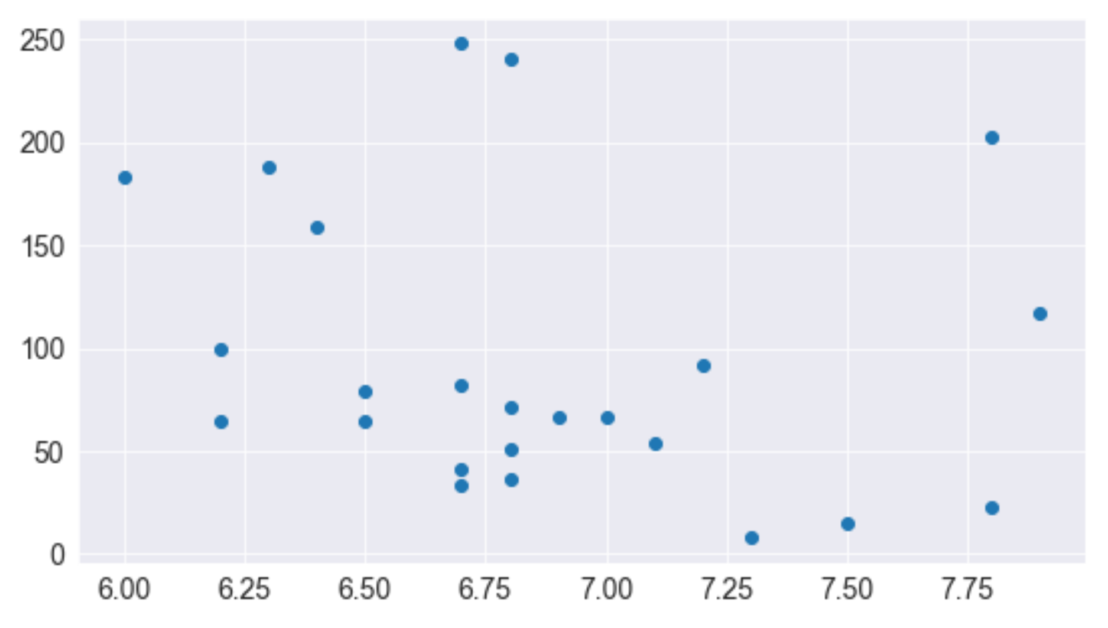
```
ratings.Budget_Adj = ratings.Budget_Adj / 1000
ratings = ratings.rename(columns={'World_Adj':'World_Adj_Thousands', 'Budget_Adj':'Budget_Adj_Thousands'})
ratings.head()
```

Out[11]:

	Movie	Bond	Year	Avg_User_IMDB	Avg_User_Rtn_Tom	Budget_Adj_Thousands	World_Adj_Thousands
0	Dr. No	Sean Connery	1962	7.3	7.7	7.688	457.928
1	From Russia with Love	Sean Connery	1963	7.5	8.0	15.174	598.624
2	Goldfinger	Sean Connery	1964	7.8	8.4	22.468	935.404
3	Thunderball	Sean Connery	1965	7.0	6.8	66.333	1040.693
4	You Only Live Twice	Sean Connery	1967	6.9	6.3	66.035	775.740

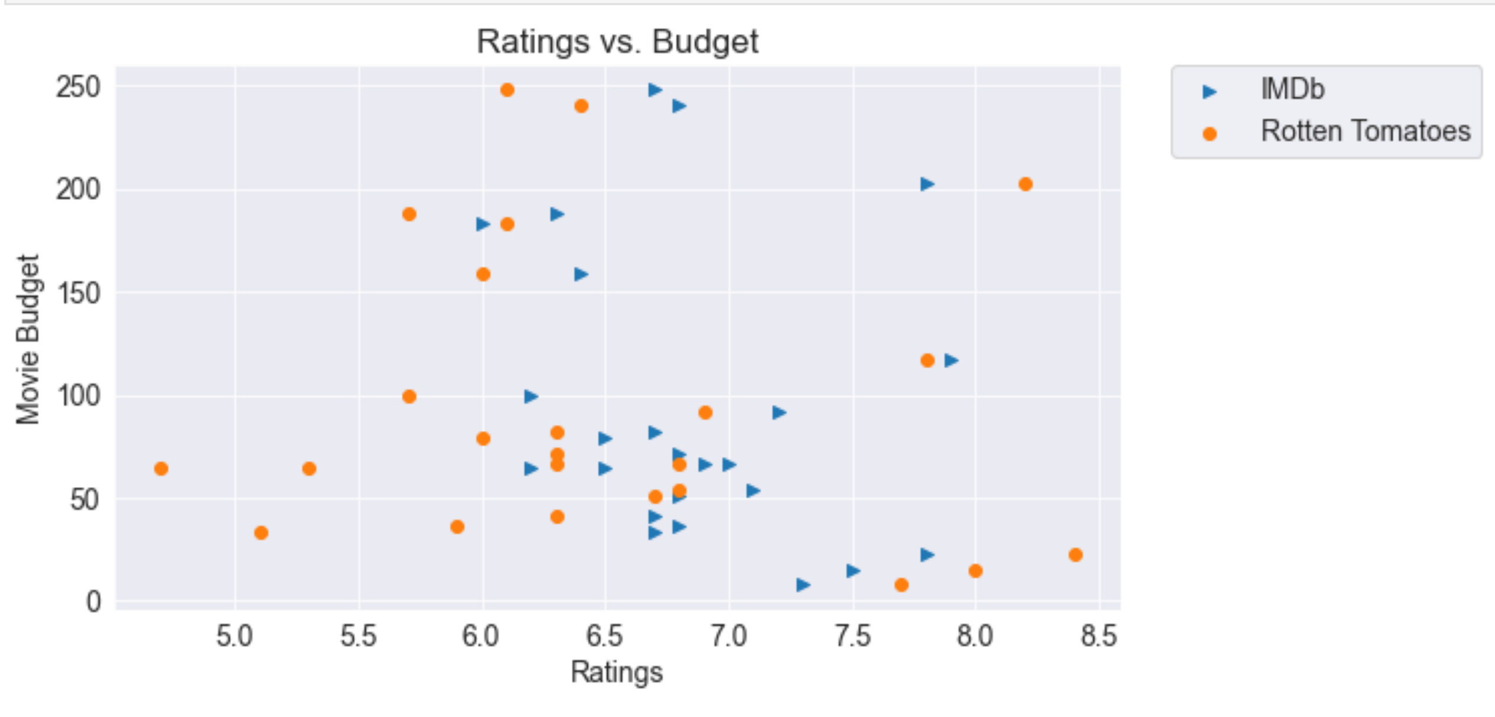
In [12]:

```
plt.scatter('Avg_User_IMDB', 'Budget_Adj_Thousands', data=ratings);
```



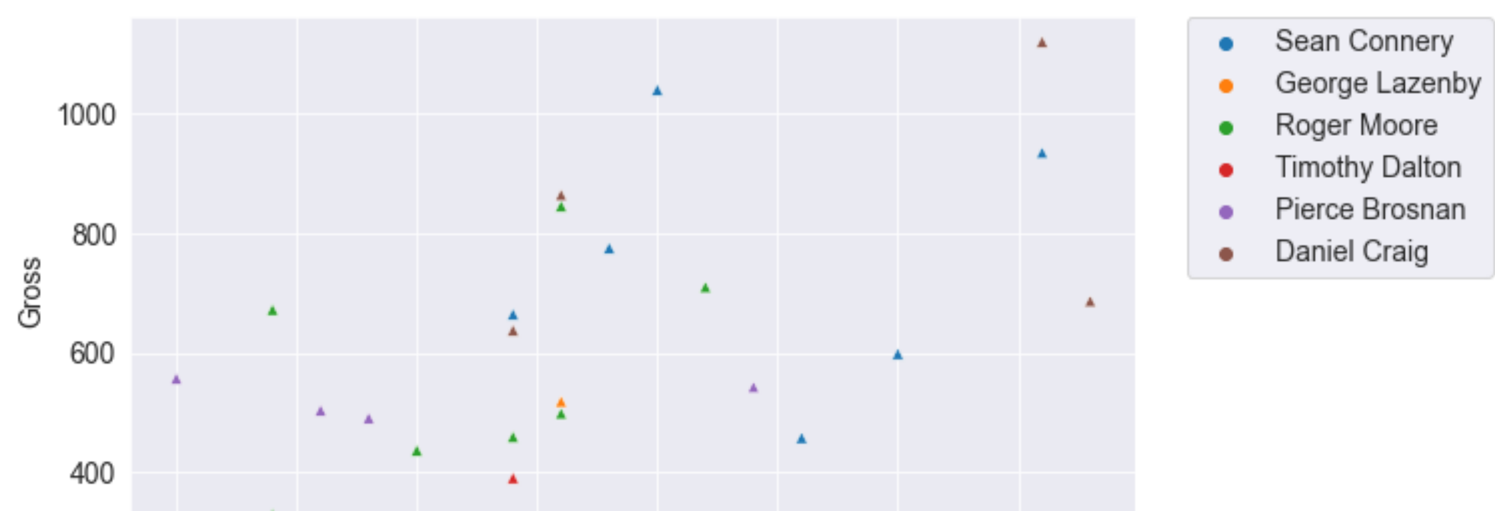
In [13]:

```
plt.scatter('Avg_User_IMDB', 'Budget_Adj_Thousands', data=ratings, marker='>', label='IMDb')
plt.scatter('Avg_User_Rtn_Tom', 'Budget_Adj_Thousands', data=ratings, marker='o', label='Rotten Tomatoes')

plt.xlabel('Ratings')
plt.ylabel('Movie Budget')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)

plt.title('Ratings vs. Budget')
plt.show();
```



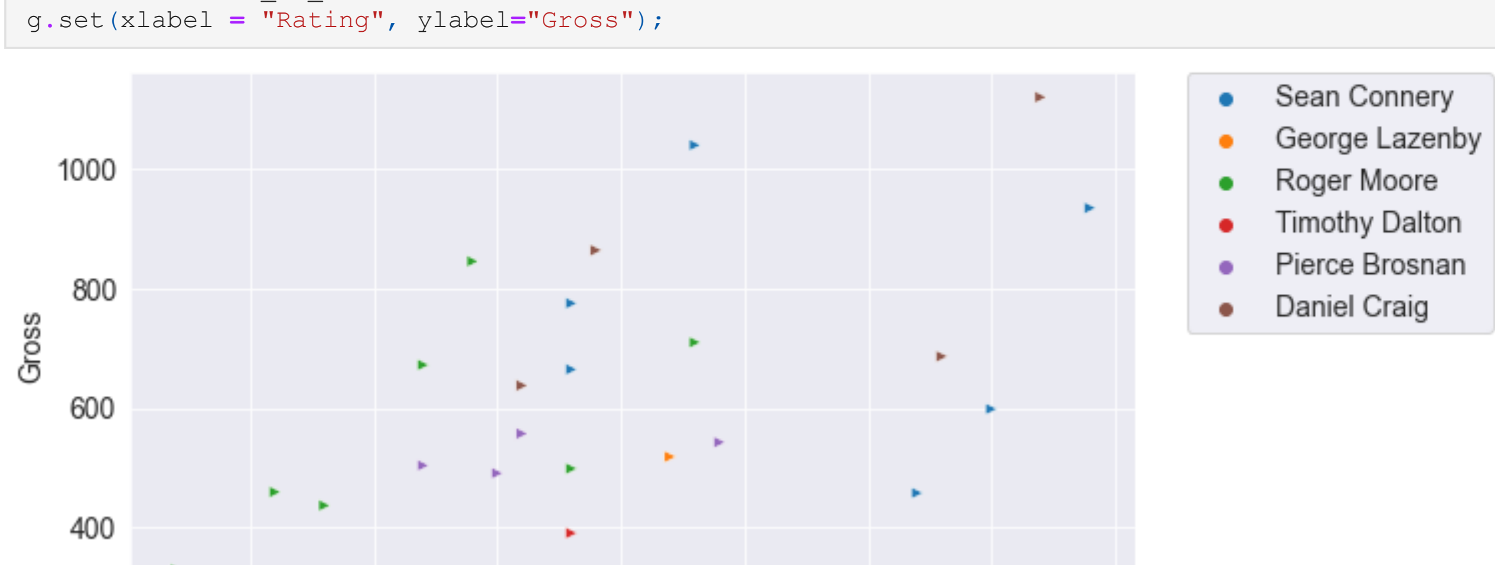
In [14]:

```
#now for seaborn

g = sns.scatterplot(data=ratings, x='Avg_User_IMDB', y='World_Adj_Thousands', hue='Bond', marker="^")
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
g.set(xlabel = "Rating", ylabel="Gross");
```



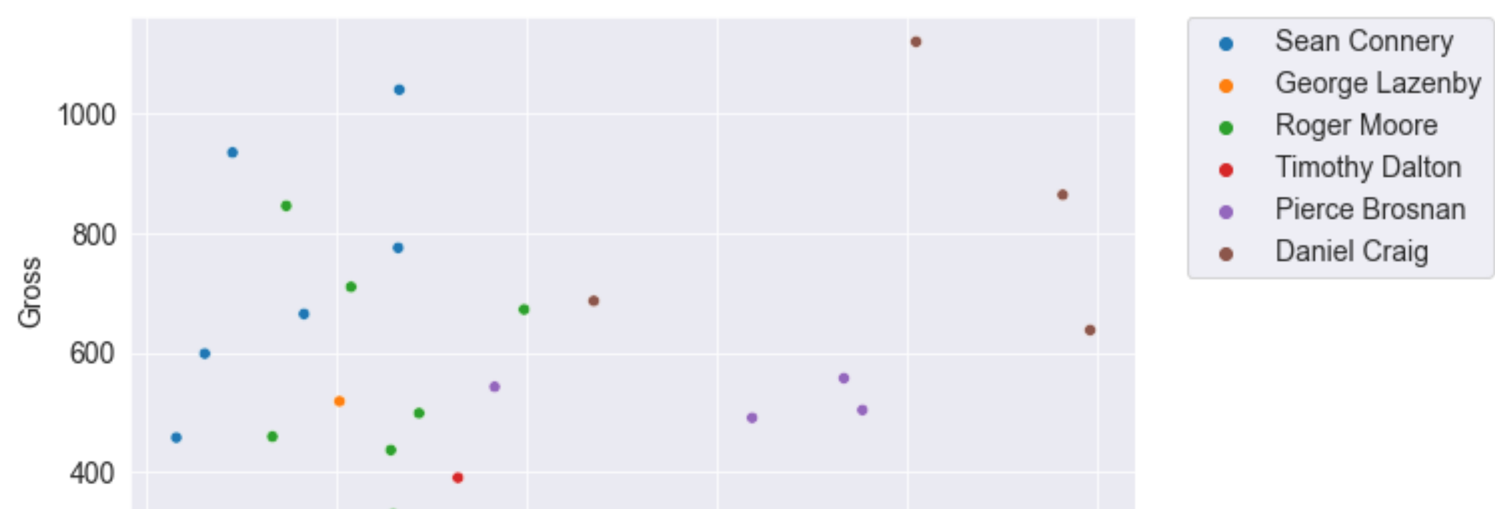
In [15]:

```
g = sns.scatterplot(data=ratings, x='Avg_User_Rtn_Tom', y='World_Adj_Thousands', hue='Bond', marker=">")
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
g.set(xlabel = "Rating", ylabel="Gross");
```



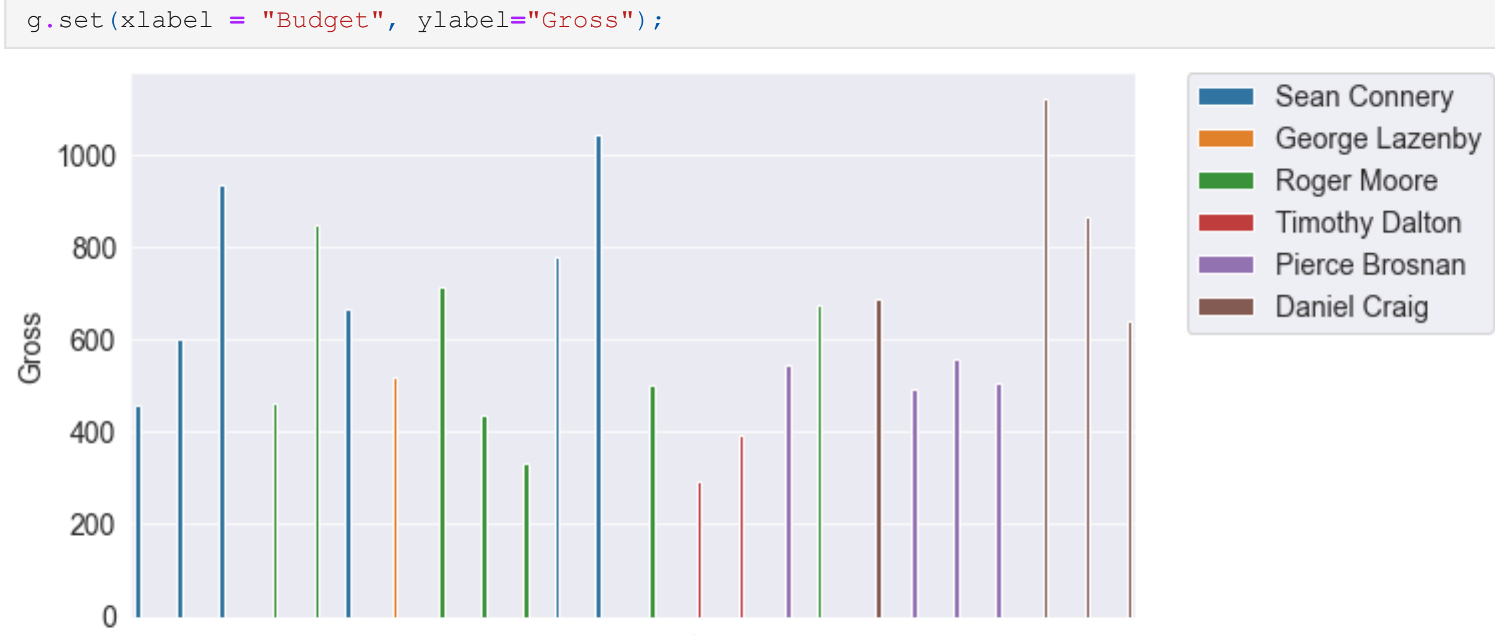
In [16]:

```
g = sns.scatterplot(data=ratings, x='Budget_Adj_Thousands', y='World_Adj_Thousands', hue='Bond', marker="o")
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
g.set(xlabel = "Budget", ylabel="Gross");
```



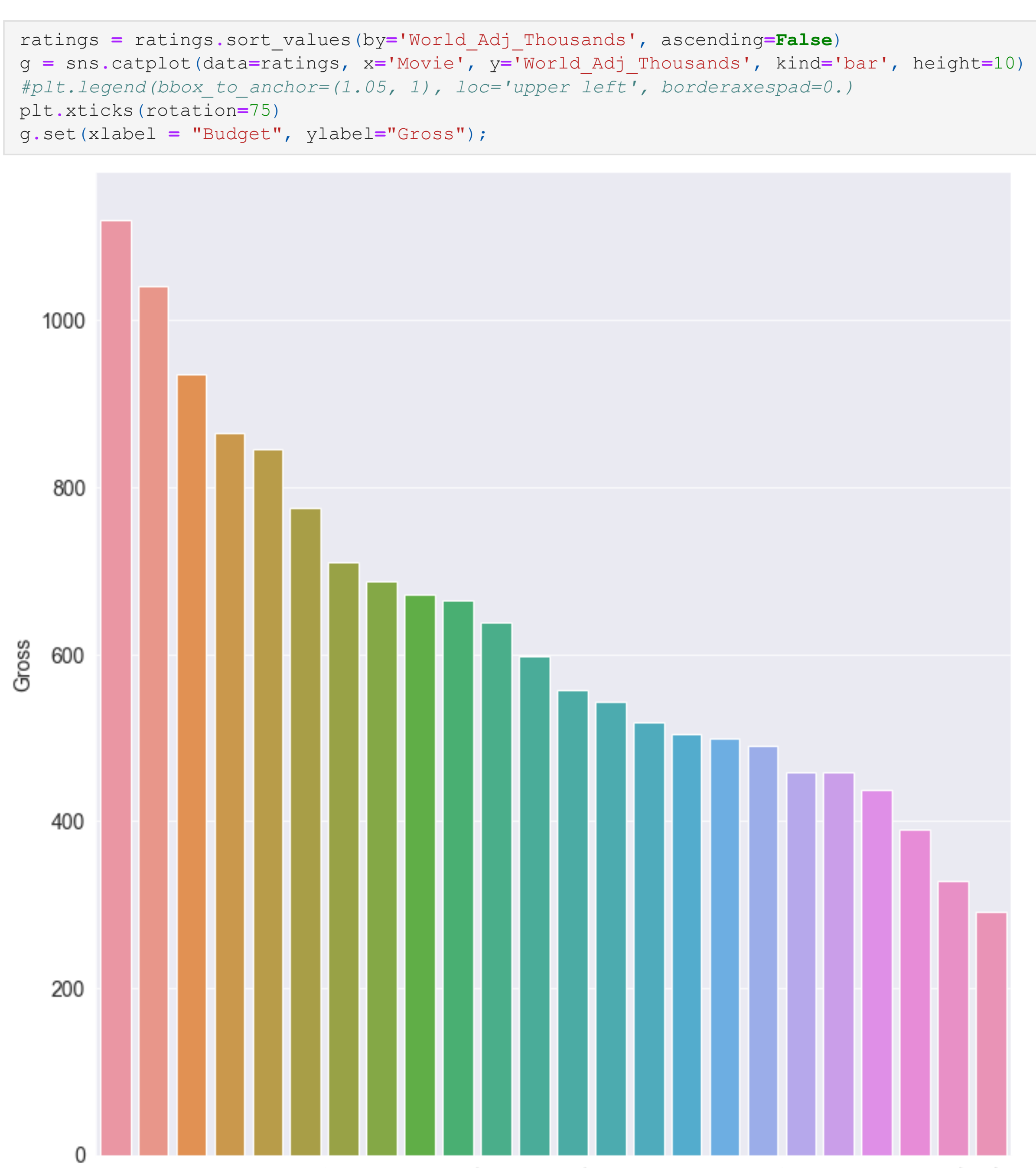
In [17]:

```
g = sns.barplot(data=ratings, x='Budget_Adj_Thousands', y='World_Adj_Thousands', hue='Bond')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
plt.xticks(rotation=75)
g.set(xlabel = "Budget", ylabel="Gross");
```



In [18]:

```
ratings = ratings.sort_values(by='World_Adj_Thousands', ascending=False)
g = sns.catplot(data=ratings, x='Movie', y='World_Adj_Thousands', kind='bar', height=10)
#plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
plt.xticks(rotation=75)
g.set(xlabel = "Budget", ylabel="Gross");
```



In [19]:

```
g = sns.barplot(data=ratings, x='Budget_Adj_Thousands', y='World_Adj_Thousands', palette="Blues_d")
#plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
plt.xticks(rotation=75)
g.set(xlabel = "Budget", ylabel="Gross");
```