

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats
%matplotlib inline
```

```
In [2]: df = pd.read_csv('Movie-Ratings.csv')
```

Given Dataset

```
In [3]: df
```

```
Out[3]:
```

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
0	(500) Days of Summer	Comedy	87	81	8	2009
1	10,000 B.C.	Adventure	9	44	105	2008
2	12 Rounds	Action	30	52	20	2009
3	127 Hours	Adventure	93	84	18	2010
4	17 Again	Comedy	55	70	20	2009
...
554	Your Highness	Comedy	26	36	50	2011
555	Youth in Revolt	Comedy	68	52	18	2009
556	Zodiac	Thriller	89	73	65	2007
557	Zombieland	Action	90	87	24	2009
558	Zookeeper	Comedy	14	42	80	2011

559 rows × 6 columns

List of Top 10 movie (descending order) in terms of Budget

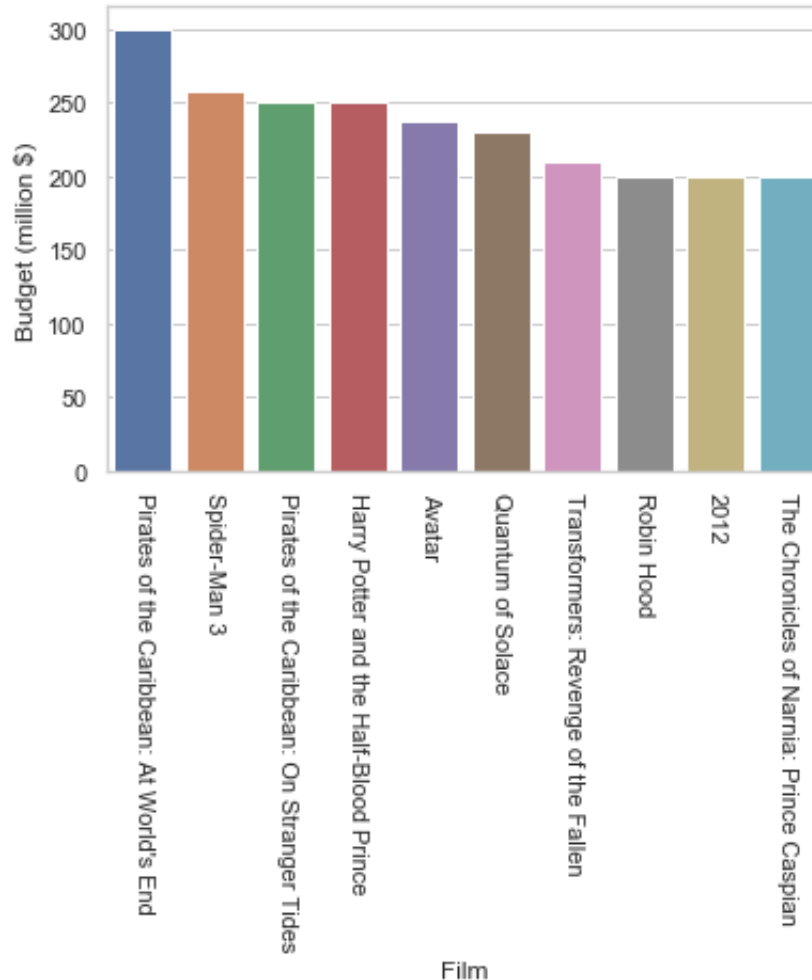
```
In [6]: df_des = df.sort_values('Budget (million $)',ascending = False).head(10)
df_des
```

Out[6]:

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
304	Pirates of the Caribbean: At World's End	Action	45	74	300	2007
360	Spider-Man 3	Action	61	54	258	2007
303	Pirates of the Caribbean: On Stranger Tides	Action	34	61	250	2011
167	Harry Potter and the Half- Blood Prince	Adventure	83	75	250	2009
33	Avatar	Action	83	92	237	2009
314	Quantum of Solace	Action	64	62	230	2008
501	Transformers: Revenge of the Fallen	Action	20	76	210	2009
327	Robin Hood	Action	42	59	200	2010
5	2012	Action	39	63	200	2009
400	The Chronicles of Narnia: Prince Caspian	Action	66	78	200	2008

```
In [51]: sns.set(style="whitegrid")
Descending_list = sns.barplot(x='Film',y='Budget (million $)',data = d
plt.xticks(rotation=-90)
```

```
Out[51]: (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), <a list of 10 Text xticklabel
objects>)
```



List of Top 10 movies (ascending order) in terms of Budget

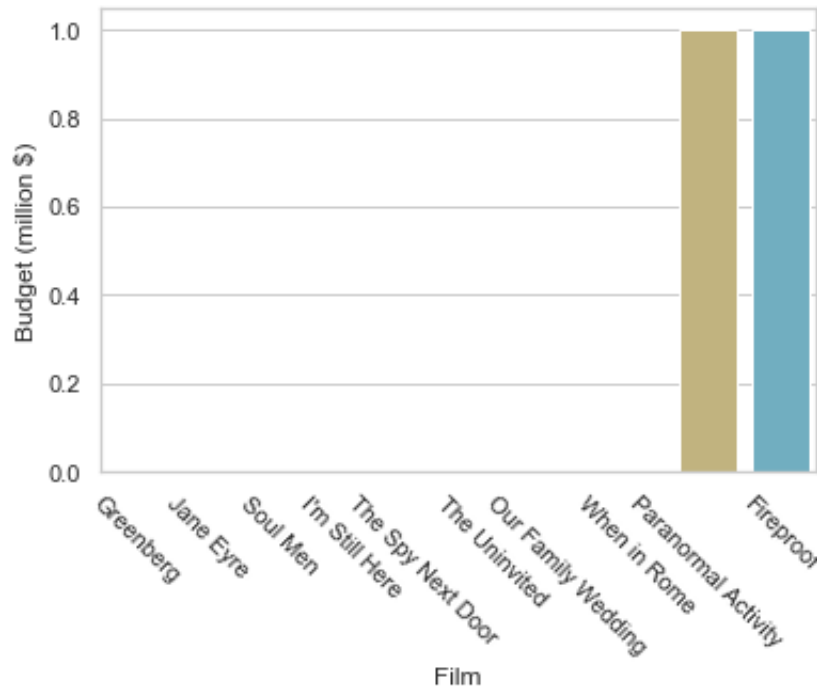
```
In [24]: df_asc = df.sort_values('Budget (million $)',ascending = True).head(10)
df_asc
```

Out[24]:

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
154	Greenberg	Comedy	75	40	0	2010
201	Jane Eyre	Romance	85	77	0	2011
356	Soul Men	Comedy	45	42	0	2008
185	I'm Still Here	Comedy	52	38	0	2010
474	The Spy Next Door	Action	13	46	0	2010
487	The Uninvited	Drama	31	52	0	2009
287	Our Family Wedding	Comedy	14	49	0	2010
539	When in Rome	Comedy	15	44	0	2010
292	Paranormal Activity	Horror	82	56	1	2007
128	Fireproof	Drama	40	51	1	2008

```
In [25]: sns.set(style="whitegrid")
Ascending_list = sns.barplot(x='Film',y='Budget (million $)',data = df
plt.xticks(rotation=-45)
```

```
Out[25]: (array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), <a list of 10 Text xticklabel
objects>)
```



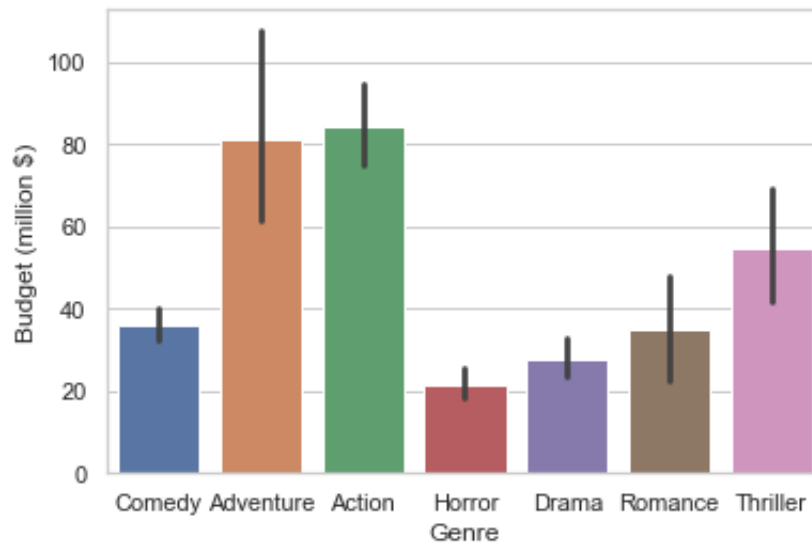
Which Genre is the Best in terms of Price Budget

```
In [11]: df.groupby(['Genre'])['Budget (million $)'].mean().sort_values(ascending=True)
```

```
Out[11]: Genre
Action      84.629870
Adventure   81.482759
Thriller    54.666667
Comedy      36.110465
Romance     35.111111
Drama       27.851485
Horror      21.673469
Name: Budget (million $), dtype: float64
```

```
In [73]: sns.barplot(x='Genre',y = 'Budget (million $)',data=df)
```

```
Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x13b24bc2088>
```



Genre with Highest Budget

```
In [60]: df.groupby(['Genre'])['Budget (million $)'].mean().sort_values(ascending=True)
```

```
Out[60]: Genre
Action    84.62987
Name: Budget (million $), dtype: float64
```

Genre with Lowest Budget

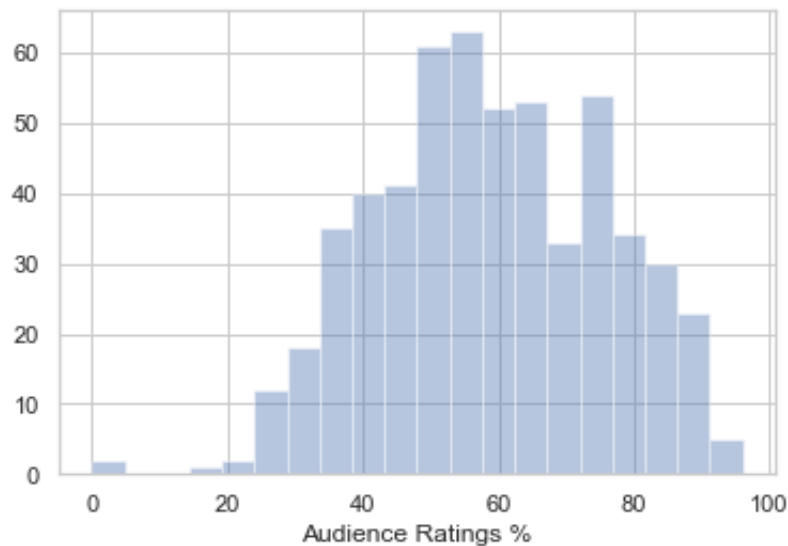
```
In [61]: df.groupby(['Genre'])['Budget (million $)'].mean().sort_values(ascending=True)
```

```
Out[61]: Genre
Horror    21.673469
Name: Budget (million $), dtype: float64
```

Distribution(graph) of Audience Rating

```
In [15]: sns.set_style('whitegrid')
sns.distplot(df['Audience Ratings %'], kde = False, color = 'b', bins =
```

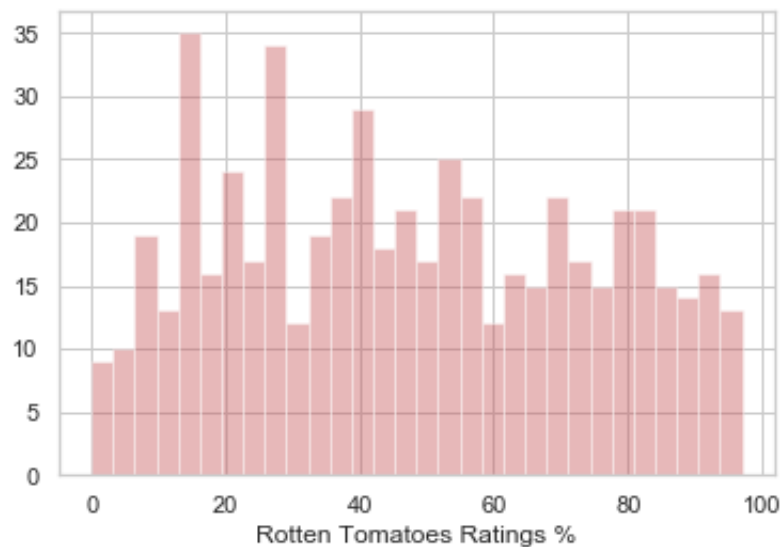
```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1e15f1c8>
```



Distribution(graph) of Critics Rating (Rotten Tomato)

```
In [17]: sns.set_style('whitegrid')
sns.distplot(df['Rotten Tomatoes Ratings %'], kde = False, color = 'r',
```

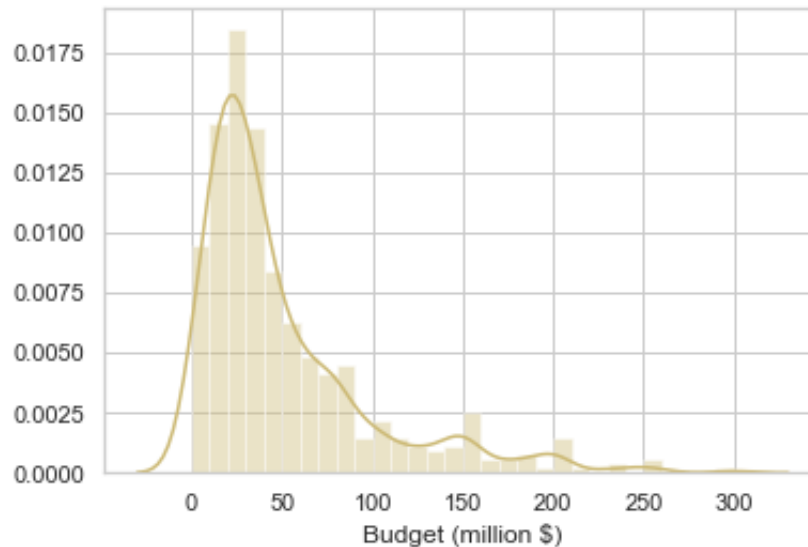
```
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1e200848>
```



Distribution of Budget

```
In [19]: sns.set_style('whitegrid')  
sns.distplot(df['Budget (million $)'], kde = True, color = 'y', bins =
```

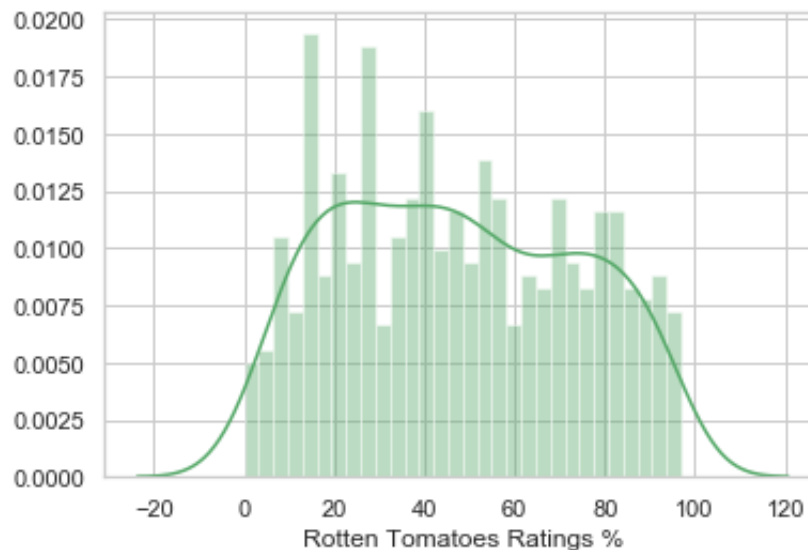
```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1e372148>
```



Distribution of Critics Ratings

```
In [20]: sns.set_style('whitegrid')  
sns.distplot(df['Rotten Tomatoes Ratings %'], kde = True, color = 'g',
```

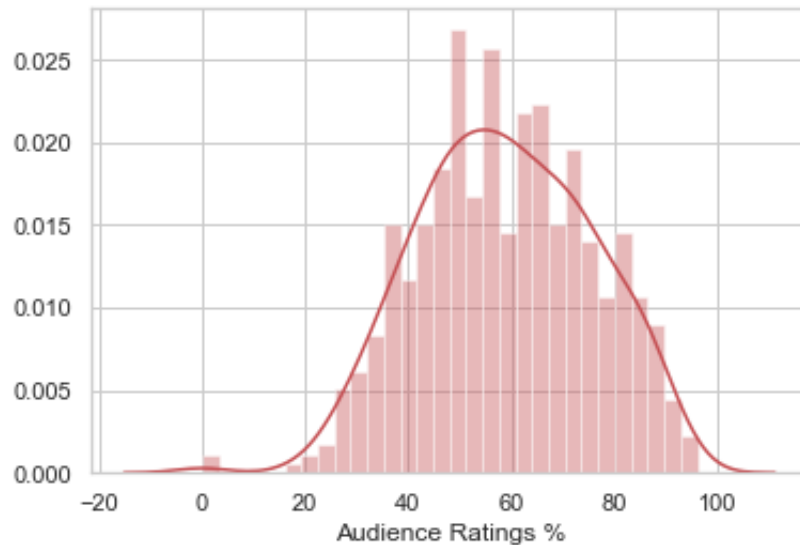
```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1e44e848>
```



Distribution of Audience Ratings

```
In [25]: sns.set_style('whitegrid')
sns.distplot(df['Audience Ratings %'], kde = True, color = 'r', bins =
```

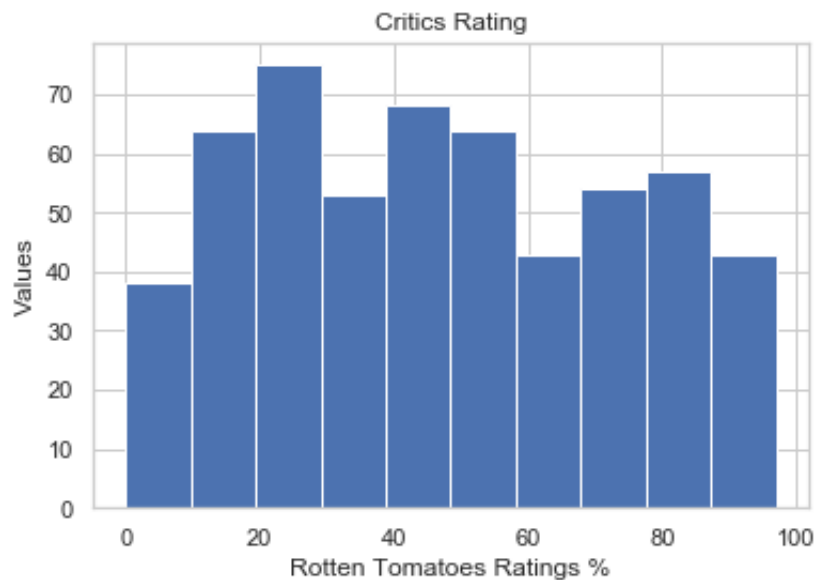
```
Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1e7dcfc8>
```



Histogram of Critics Rating

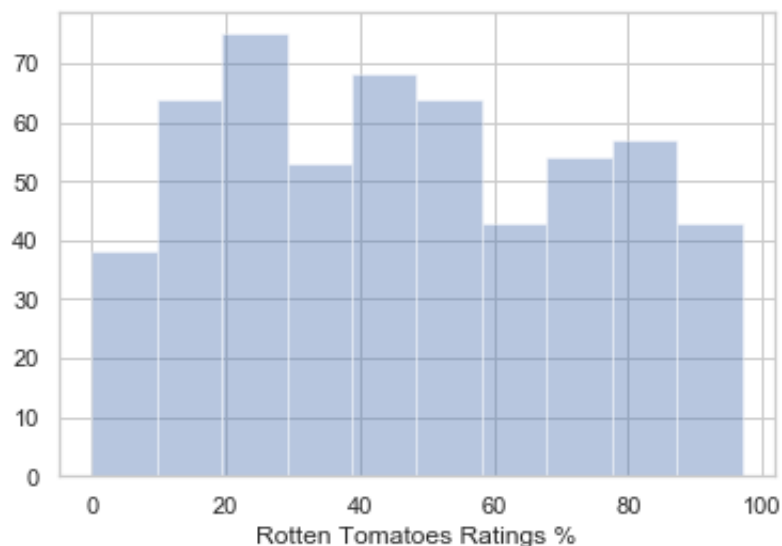
```
In [29]: plt.title('Critics Rating')
plt.xlabel('Rotten Tomatoes Ratings %')
plt.ylabel('Values')
plt.hist(df['Rotten Tomatoes Ratings %'])
```

```
Out[29]: (array([38., 64., 75., 53., 68., 64., 43., 54., 57., 43.]),
array([ 0. ,  9.7, 19.4, 29.1, 38.8, 48.5, 58.2, 67.9, 77.6, 87.3, 97. ]),
<a list of 10 Patch objects>)
```



```
In [35]: sns.distplot(df['Rotten Tomatoes Ratings %'], hist=True, kde=False, bi
```

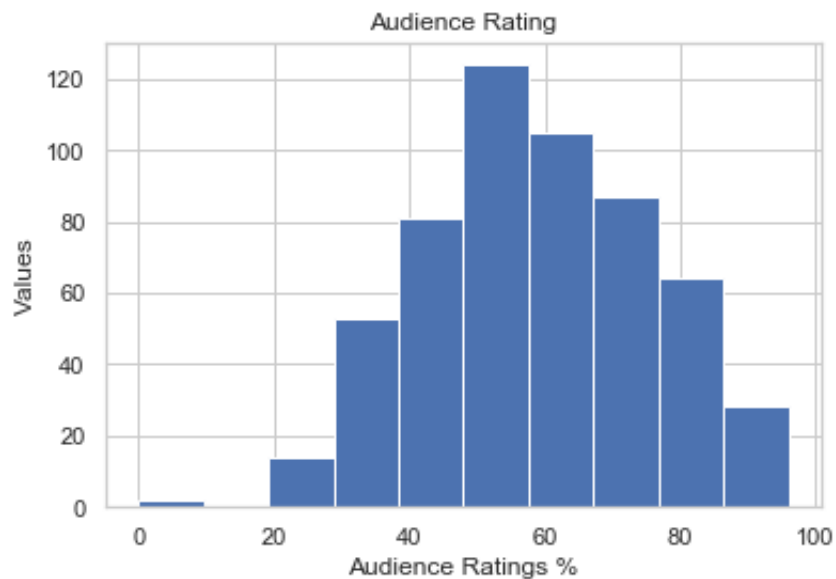
```
Out[35]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1fb72c08>
```



Histogram of Audience Rating

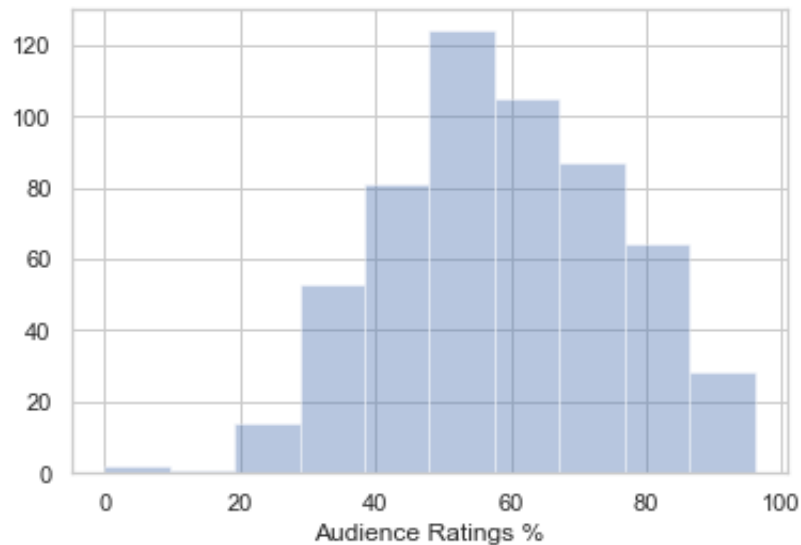
```
In [32]: plt.title('Audience Rating')
plt.xlabel('Audience Ratings %')
plt.ylabel('Values')
plt.hist(df['Audience Ratings %'])
```

```
Out[32]: (array([ 2.,  1., 14., 53., 81., 124., 105., 87., 64., 28.]),
array([ 0. ,  9.6, 19.2, 28.8, 38.4, 48. , 57.6, 67.2, 76.8, 86.4, 96. ],
      <a list of 10 Patch objects>)
```



```
In [34]: sns.distplot(df['Audience Ratings %'], hist=True, kde=False, bins=10,
```

```
Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1fb40708>
```



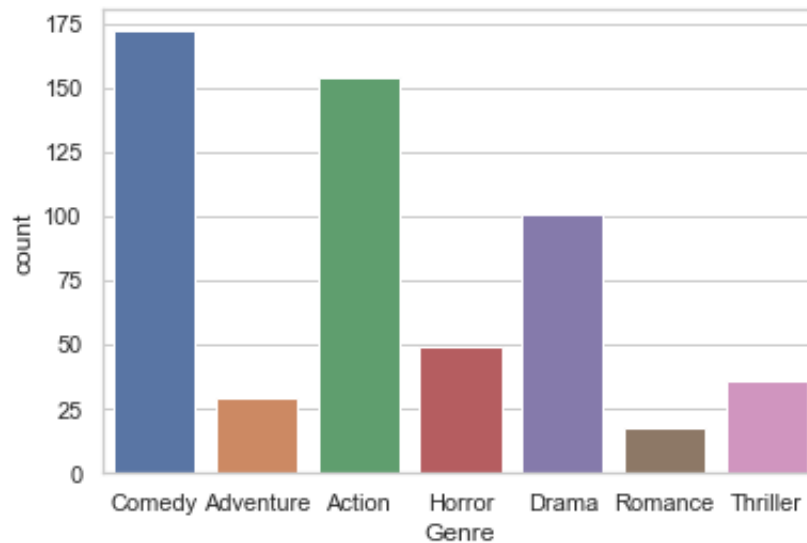
Show the Total count of all the different Genre present in the dataset

```
In [36]: df.groupby('Genre').count()['Film']
```

```
Out[36]: Genre
Action      154
Adventure    29
Comedy      172
Drama       101
Horror       49
Romance      18
Thriller     36
Name: Film, dtype: int64
```

```
In [37]: sns.countplot(x='Genre',data = df)
```

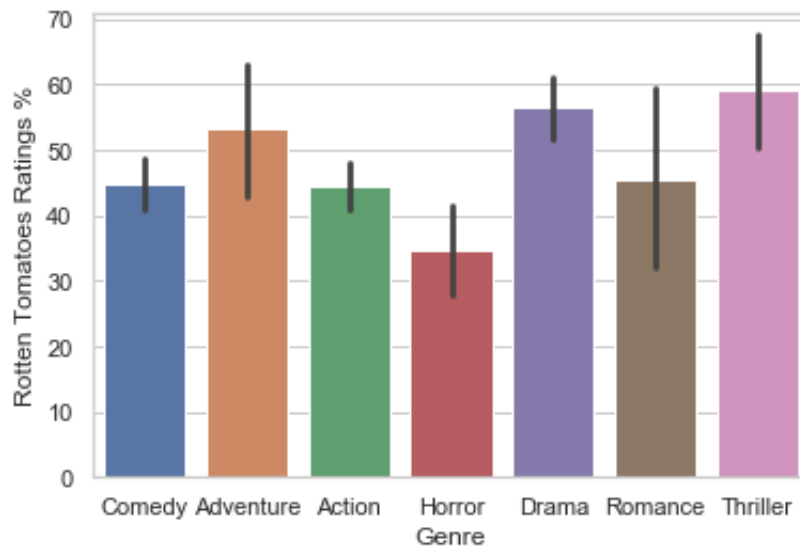
```
Out[37]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1fc7ce08>
```



Show the Graph of Rating based on Genre

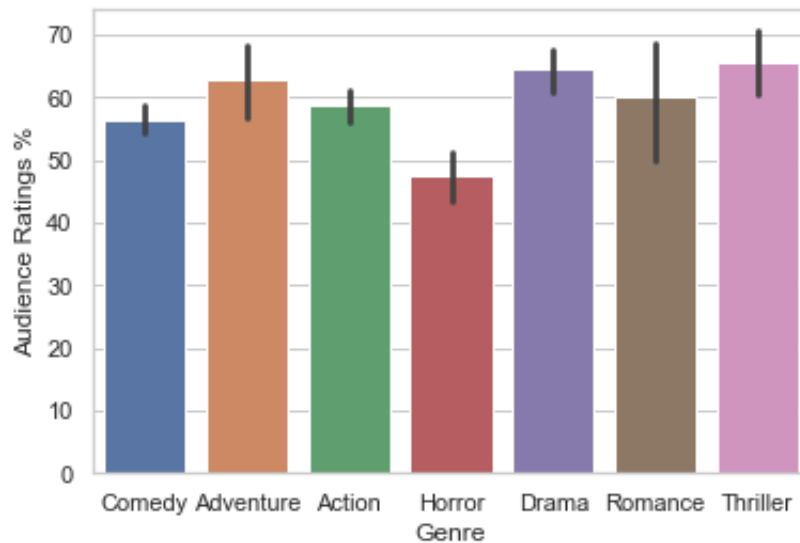
```
In [38]: sns.barplot(x='Genre',y='Rotten Tomatoes Ratings %',data=df)
```

```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1fce4f48>
```



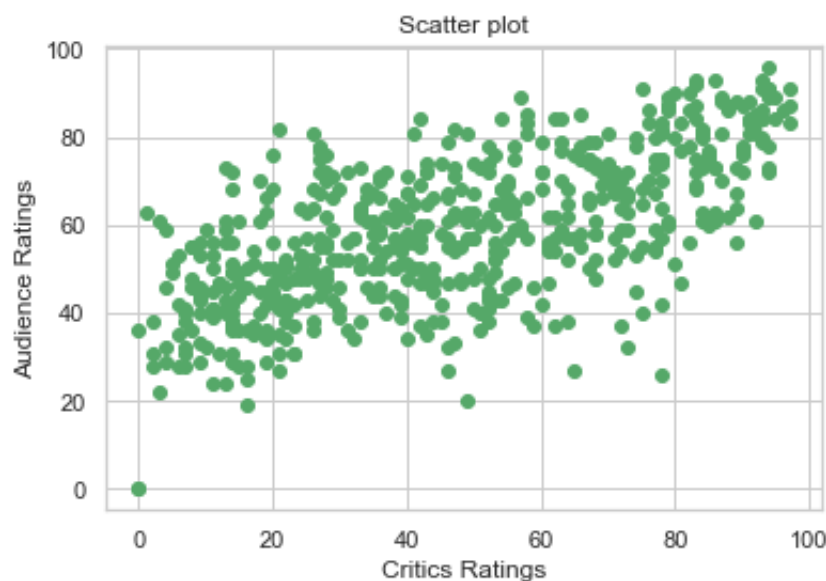
```
In [39]: sns.barplot(x='Genre',y='Audience Ratings %',data=df)
```

```
Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1fd70a08>
```



Scatter plot of Critics Rating vs Audience Rating

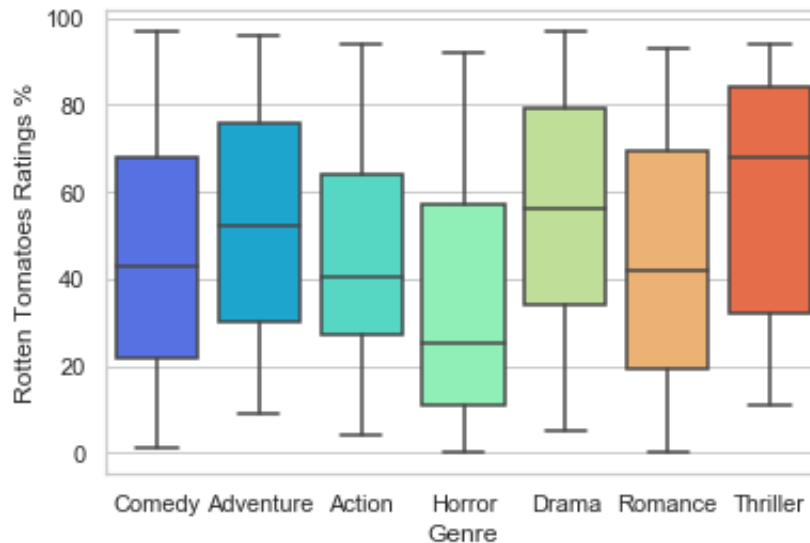
```
In [42]: x = df['Rotten Tomatoes Ratings %']  
y = df['Audience Ratings %']  
plt.scatter(x, y, marker='o', color='g')  
plt.title('Scatter plot')  
plt.xlabel('Critics Ratings')  
plt.ylabel('Audience Ratings')  
plt.show()
```



Box plot based on Genre and Critics Rating

```
In [43]: sns.boxplot(x='Genre', y='Rotten Tomatoes Ratings %', data=df,palette=
```

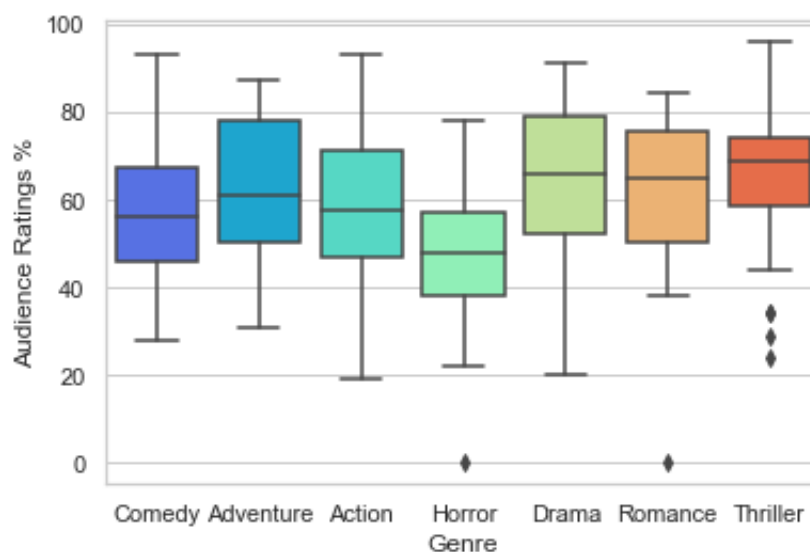
```
Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1e76c948>
```



Box plot based on Genre and Audience Rating

```
In [44]: sns.boxplot(x='Genre', y='Audience Ratings %', data=df,palette='rainbow
```

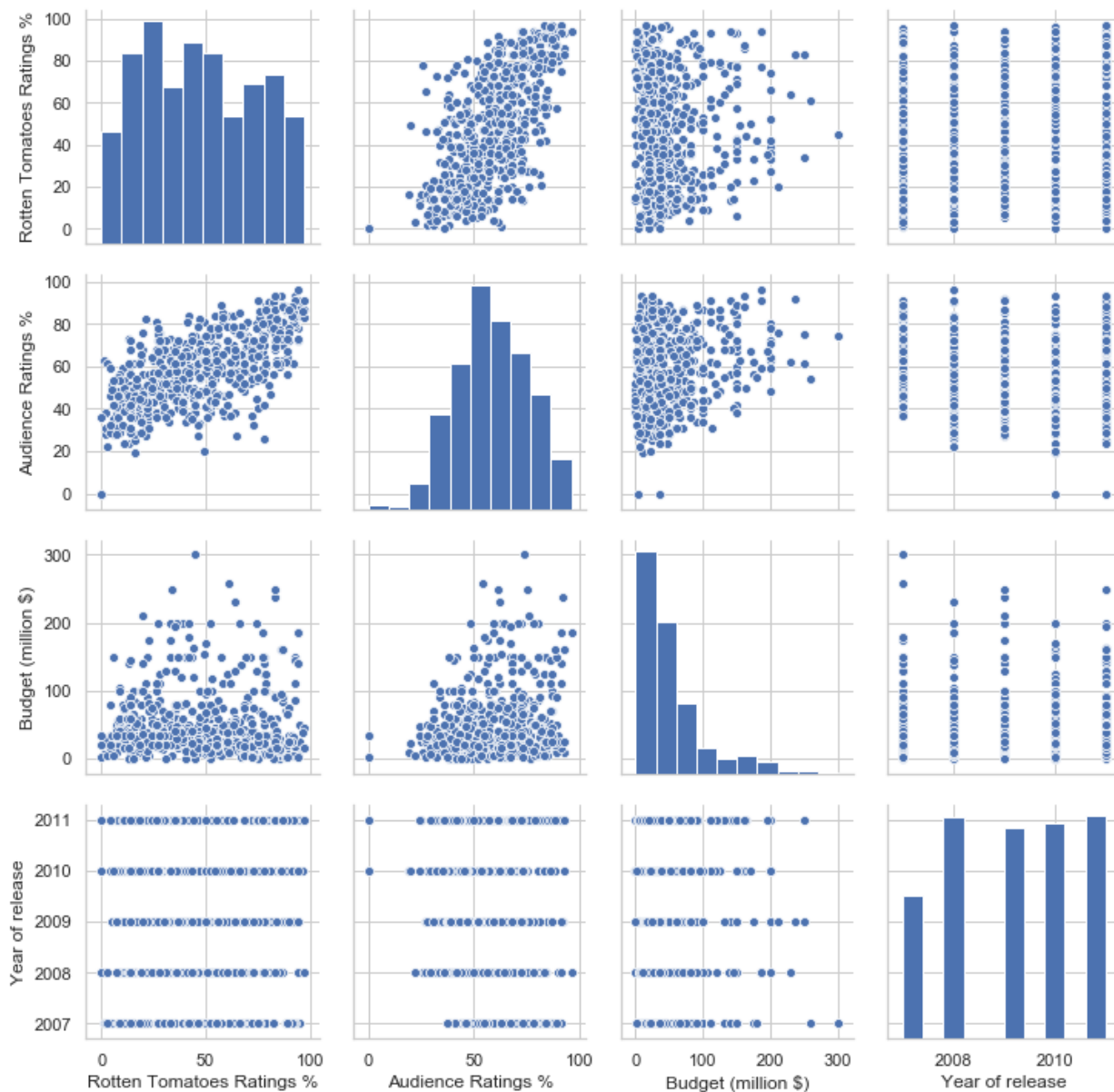
```
Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x13b1ff47f88>
```



Pairplot of the Whole Data Set

In [48]: `sns.pairplot(df)`

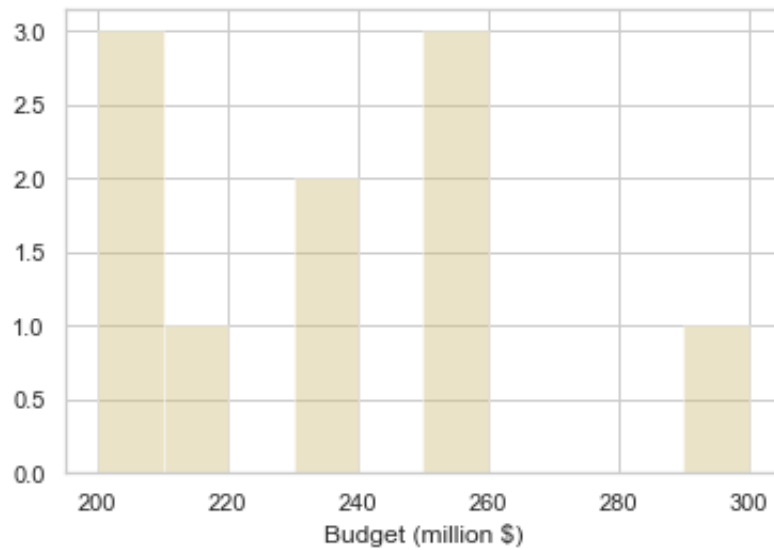
Out [48]: `<seaborn.axisgrid.PairGrid at 0x13b2134fcc8>`



Histogram of Top 10 movies (Descending) in terms of Budget


```
In [69]: sns.distplot(df_des['Budget (million $)'],hist=True, kde=False,bins=10
```

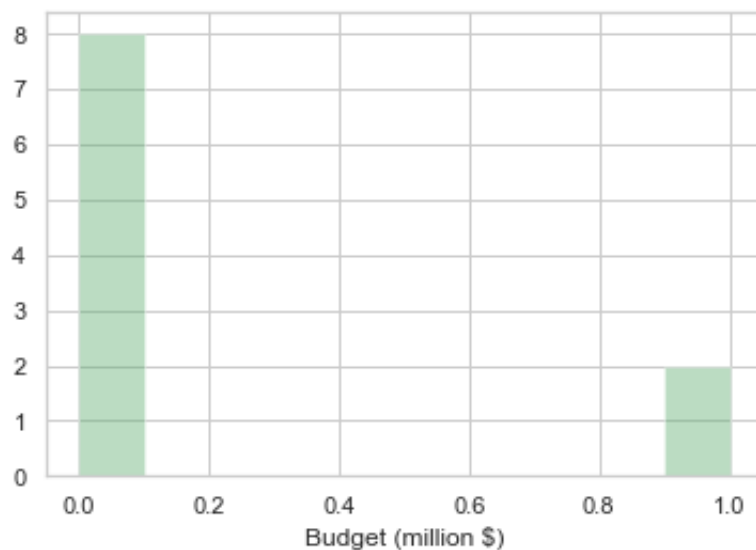
```
Out[69]: <matplotlib.axes._subplots.AxesSubplot at 0x13b24a814c8>
```



Histogram of Top 10 movies (Ascending) in terms of Budget

```
In [63]: sns.distplot(df_asc['Budget (million $)'],hist=True, kde=False,bins=10
```

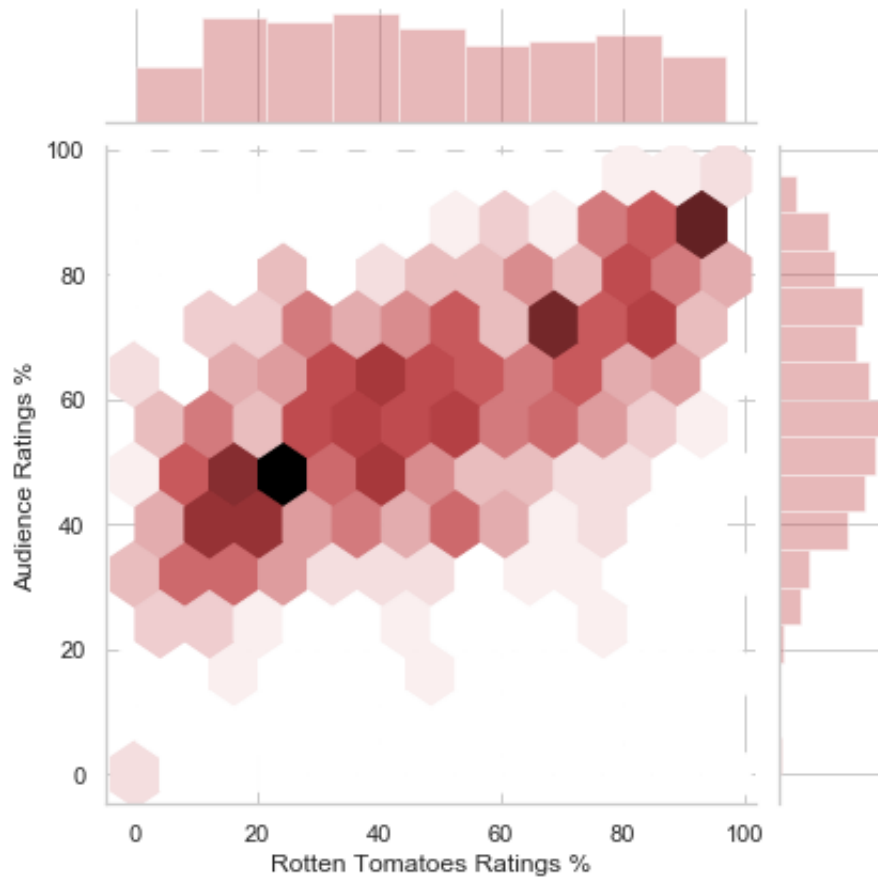
```
Out[63]: <matplotlib.axes._subplots.AxesSubplot at 0x13b236b5b48>
```



Joint Plot of critics rating and audience rating (with type as hexagonal)

```
In [68]: sns.jointplot(x='Rotten Tomatoes Ratings %',y='Audience Ratings %',data=)
```

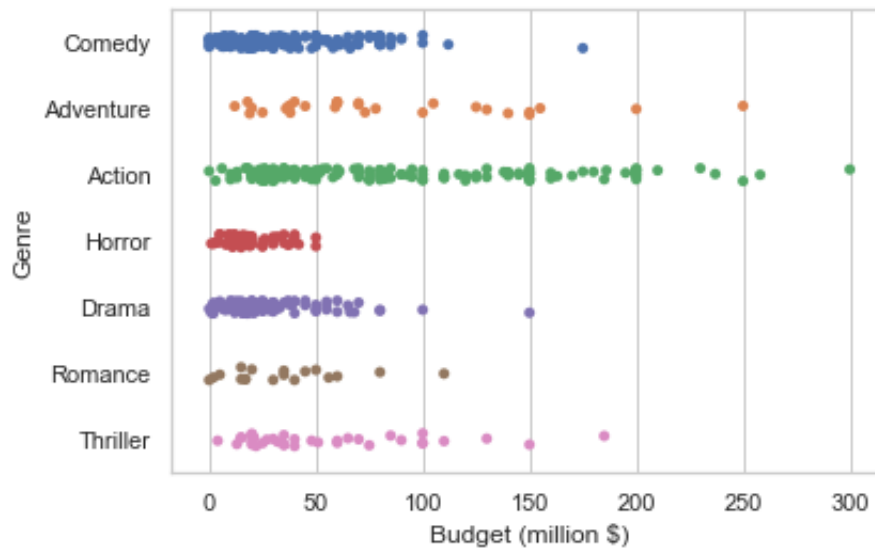
```
Out[68]: <seaborn.axisgrid.JointGrid at 0x13b2397b188>
```



Joint Plot of budget vs genre (type as scatter)

```
In [72]: sns.stripplot(x='Budget (million $)', y='Genre', data=df ,jitter=True)
```

```
Out[72]: <matplotlib.axes._subplots.AxesSubplot at 0x13b24b66948>
```



Extras

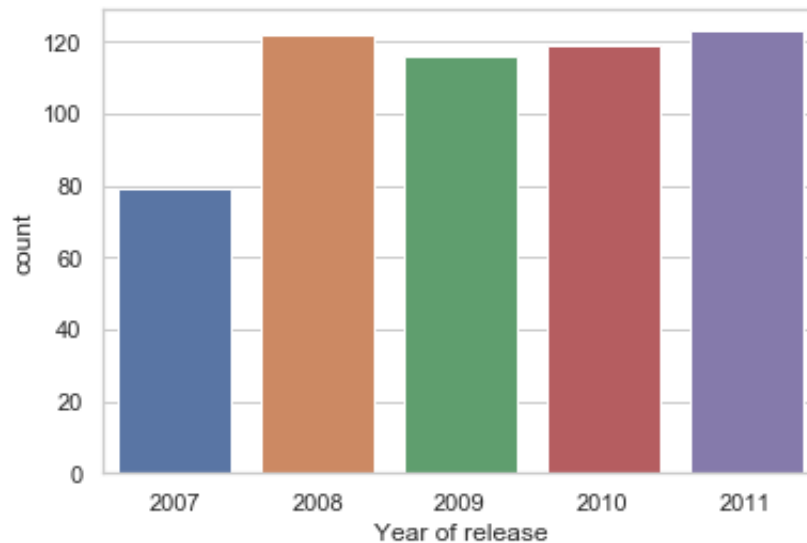
Year wise distribution of the films

```
In [14]: df.groupby('Year of release').count()['Film']
```

```
Out[14]: Year of release
2007      79
2008     122
2009     116
2010     119
2011     123
Name: Film, dtype: int64
```

```
In [8]: sns.countplot(x='Year of release',data = df)
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x22a81badf88>
```



```
In [20]: df.groupby(['Year of release'])['Budget (million $)'].mean().sort_valu
```

```
Out[20]: <bound method Series.sort_values of Year of release
```

```
2007    57.860759
```

```
2008    46.934426
```

```
2009    48.112069
```

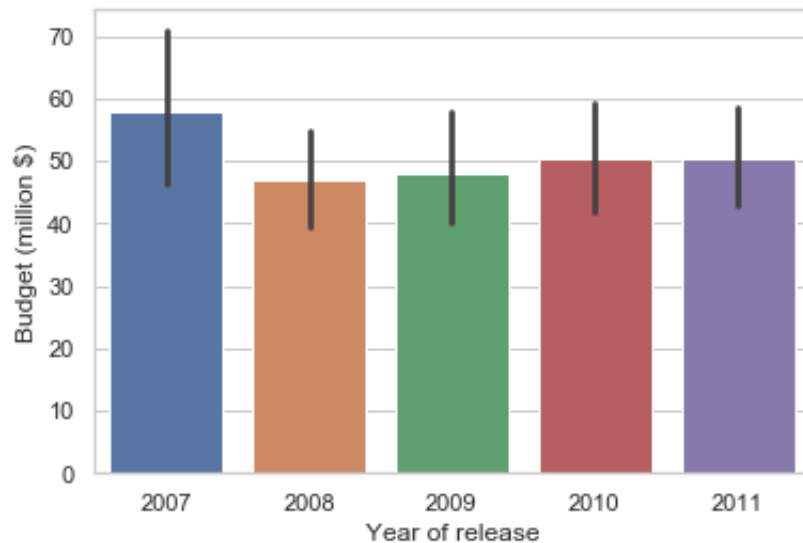
```
2010    50.394958
```

```
2011    50.463415
```

```
Name: Budget (million $), dtype: float64>
```

```
In [21]: sns.barplot(x='Year of release',y = 'Budget (million $)',data=df)
```

```
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x22a81da9688>
```



```
In [45]: df.groupby(['Year of release'])['Audience Ratings %'].mean()
```

```
Out[45]: Year of release
2007      66.417722
2008      57.418033
2009      55.431034
2010      55.462185
2011      61.430894
Name: Audience Ratings %, dtype: float64
```

```
In [46]: df.groupby(['Year of release'])['Rotten Tomatoes Ratings %'].mean()
```

```
Out[46]: Year of release
2007      49.759494
2008      44.385246
2009      45.094828
2010      45.563025
2011      52.414634
Name: Rotten Tomatoes Ratings %, dtype: float64
```

```
In [ ]:
```

