

```
In [1]: import pandas as pd
import os
```

```
In [3]: os.getcwd()
```

```
Out[3]: 'C:\\Users\\DELL'
```

```
In [5]: movies = pd.read_csv(r'C:\Users\DELL\Downloads\Movie-Rating.csv')
```

```
In [7]: movies
```

```
Out[7]:
```

	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

```
In [9]: len(movies)
```

```
Out[9]: 559
```

```
In [11]: movies.head()
```

Out[11]:

	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

In [13]: `movies.tail()`

Out[13]:

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
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

In [15]: `movies.columns`

Out[15]: `Index(['Film', 'Genre', 'Rotten Tomatoes Ratings %', 'Audience Ratings %', 'Budget (million $)', 'Year of release'], dtype='object')`

In [17]: `movies.columns = ['Film', 'Genre', 'CriticRating', 'AudienceRating', 'BudgetMillions', 'Year']`

In [19]: `movies.head()`

Out[19]:

	Film	Genre	CriticRating	AudienceRating	BudgetMillions	Year
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

In [23]: `movies.info()`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Film                   559 non-null   object
1   Genre                  559 non-null   object
2   CriticRating           559 non-null   int64
3   AudienceRating         559 non-null   int64
4   BudgetMillions         559 non-null   int64
5   Year                   559 non-null   int64
dtypes: int64(4), object(2)
memory usage: 26.3+ KB

```

In [29]: `movies.describe()`

Out[29]:

	CriticRating	AudienceRating	BudgetMillions	Year
count	559.000000	559.000000	559.000000	559.000000
mean	47.309481	58.744186	50.236136	2009.152057
std	26.413091	16.826887	48.731817	1.362632
min	0.000000	0.000000	0.000000	2007.000000
25%	25.000000	47.000000	20.000000	2008.000000
50%	46.000000	58.000000	35.000000	2009.000000
75%	70.000000	72.000000	65.000000	2010.000000
max	97.000000	96.000000	300.000000	2011.000000

In [31]: `movies['Film']`

Out[31]:

```

0      (500) Days of Summer
1      10,000 B.C.
2      12 Rounds
3      127 Hours
4      17 Again
...
554     Your Highness
555     Youth in Revolt
556     Zodiac
557     Zombieland
558     Zookeeper
Name: Film, Length: 559, dtype: object

```

In [34]: `movies.Film`

```
Out[34]: 0      (500) Days of Summer
         1      10,000 B.C.
         2      12 Rounds
         3      127 Hours
         4      17 Again
         ...
        554      Your Highness
        555      Youth in Revolt
        556      Zodiac
        557      Zombieland
        558      Zookeeper
Name: Film, Length: 559, dtype: object
```

```
In [41]: movies.Film = movies.Film.astype('category')
```

```
In [43]: movies.Film
```

```
Out[43]: 0      (500) Days of Summer
         1      10,000 B.C.
         2      12 Rounds
         3      127 Hours
         4      17 Again
         ...
        554      Your Highness
        555      Youth in Revolt
        556      Zodiac
        557      Zombieland
        558      Zookeeper
Name: Film, Length: 559, dtype: category
Categories (559, object): ['(500) Days of Summer ', '10,000 B.C.', '12 Rounds', '127 Hours', ..., 'Youth in Revolt', 'Zodiac', 'Zombieland ', 'Zookeeper']
```

```
In [45]: movies.head()
```

```
Out[45]:
```

	Film	Genre	CriticRating	AudienceRating	BudgetMillions	Year
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

```
In [47]: movies.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Film                  559 non-null   category
1   Genre                  559 non-null   object
2   CriticRating          559 non-null   int64
3   AudienceRating        559 non-null   int64
4   BudgetMillions        559 non-null   int64
5   Year                  559 non-null   int64
dtypes: category(1), int64(4), object(1)
memory usage: 43.6+ KB

```

```

In [55]: movies.Genre = movies.Genre.astype('category')
         movies.Year = movies.Year.astype('category')

```

```

In [57]: movies.Genre

```

```

Out[57]: 0      Comedy
         1      Adventure
         2      Action
         3      Adventure
         4      Comedy
         ...
        554     Comedy
        555     Comedy
        556    Thriller
        557     Action
        558     Comedy
        Name: Genre, Length: 559, dtype: category
        Categories (7, object): ['Action', 'Adventure', 'Comedy', 'Drama', 'Horror', 'Romance', 'Thriller']

```

```

In [59]: movies.Year

```

```

Out[59]: 0      2009
         1      2008
         2      2009
         3      2010
         4      2009
         ...
        554     2011
        555     2009
        556     2007
        557     2009
        558     2011
        Name: Year, Length: 559, dtype: category
        Categories (5, int64): [2007, 2008, 2009, 2010, 2011]

```

```

In [61]: movies.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Film                  559 non-null   category
1   Genre                  559 non-null   category
2   CriticRating           559 non-null   int64
3   AudienceRating         559 non-null   int64
4   BudgetMillions         559 non-null   int64
5   Year                   559 non-null   category
dtypes: category(3), int64(3)
memory usage: 36.5 KB

```

```
In [63]: movies.Genre.cat.categories
```

```
Out[63]: Index(['Action', 'Adventure', 'Comedy', 'Drama', 'Horror', 'Romance',
               'Thriller'],
              dtype='object')
```

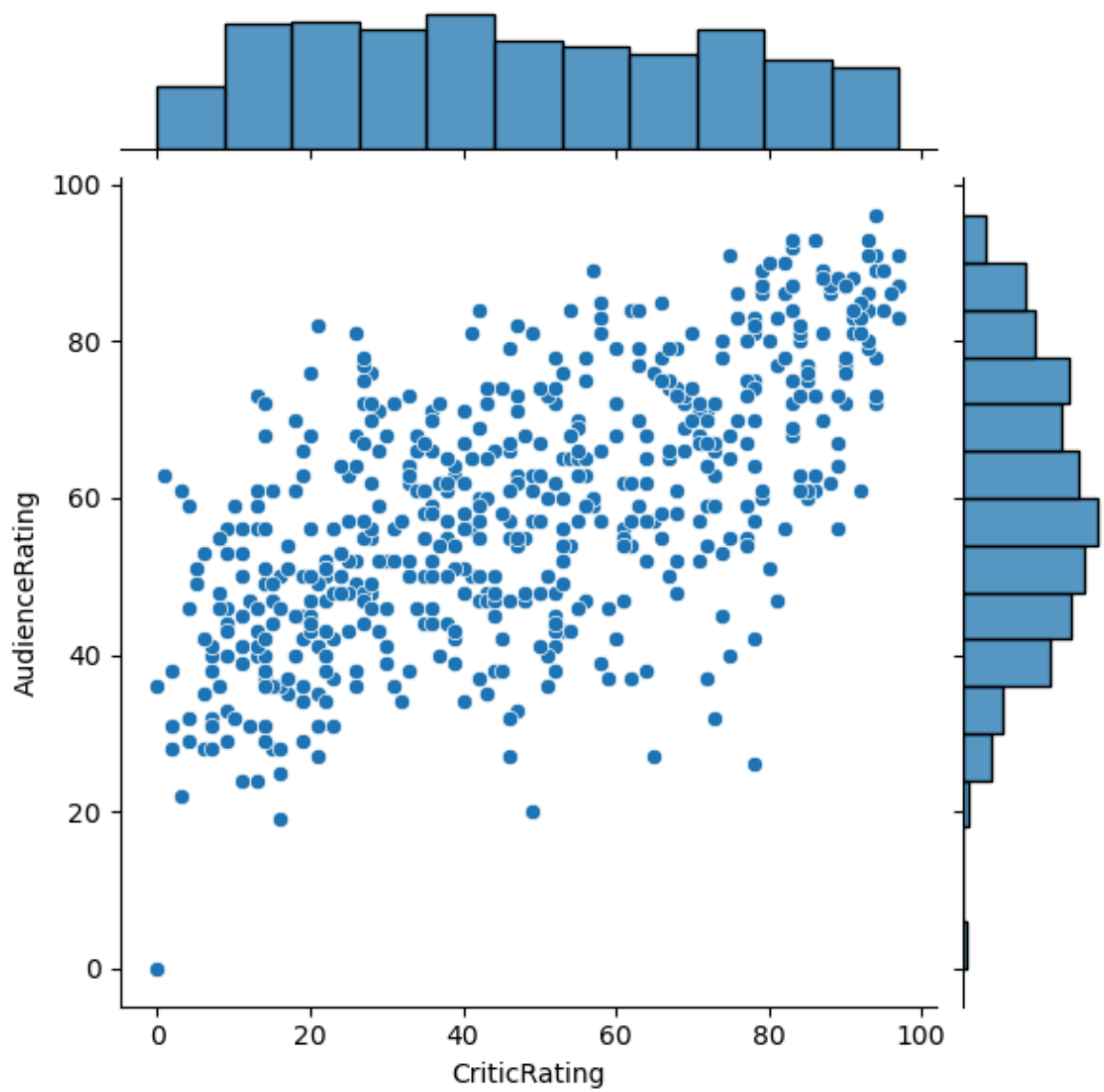
```
In [67]: movies.describe()
```

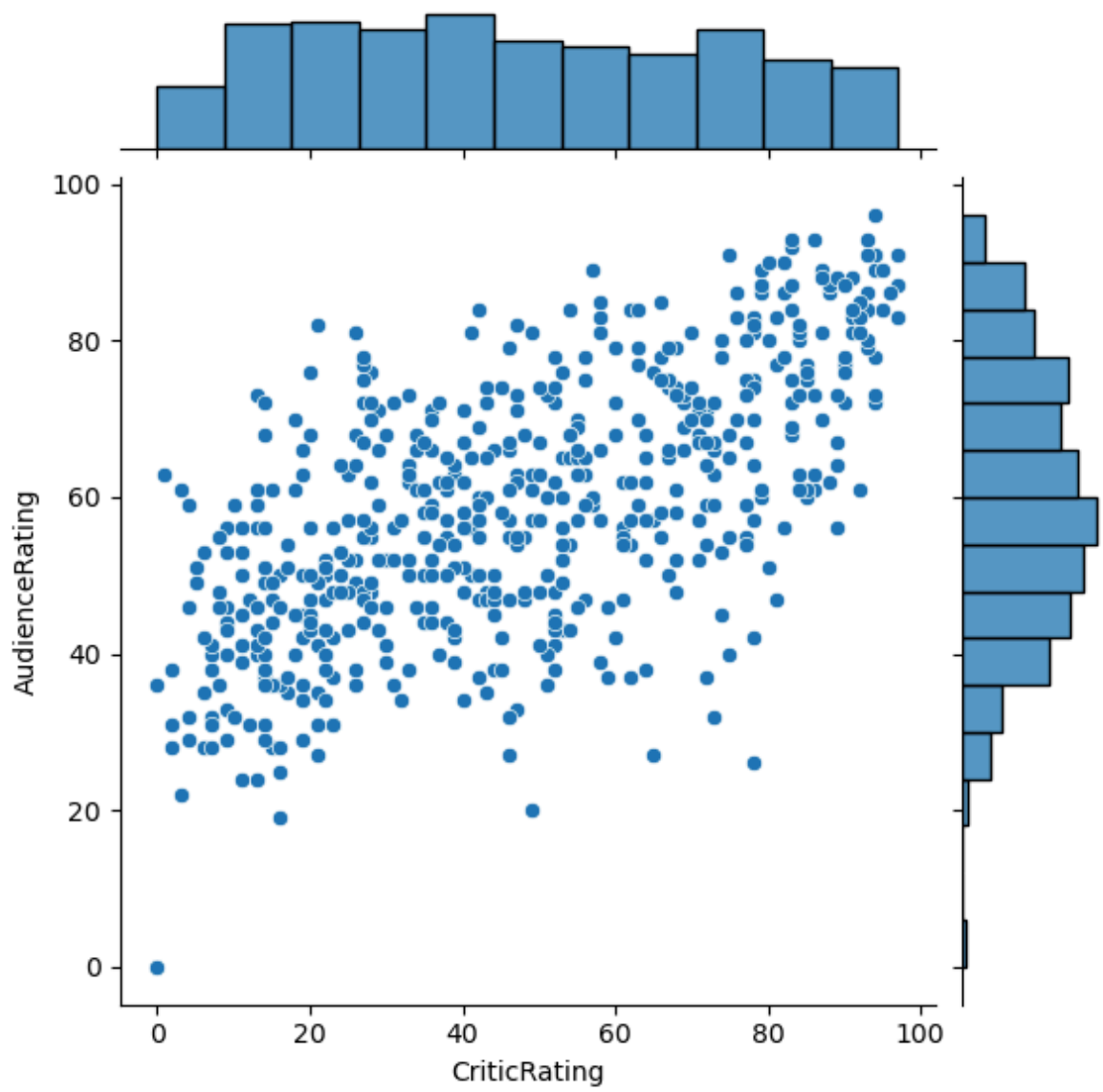
```
Out[67]:
```

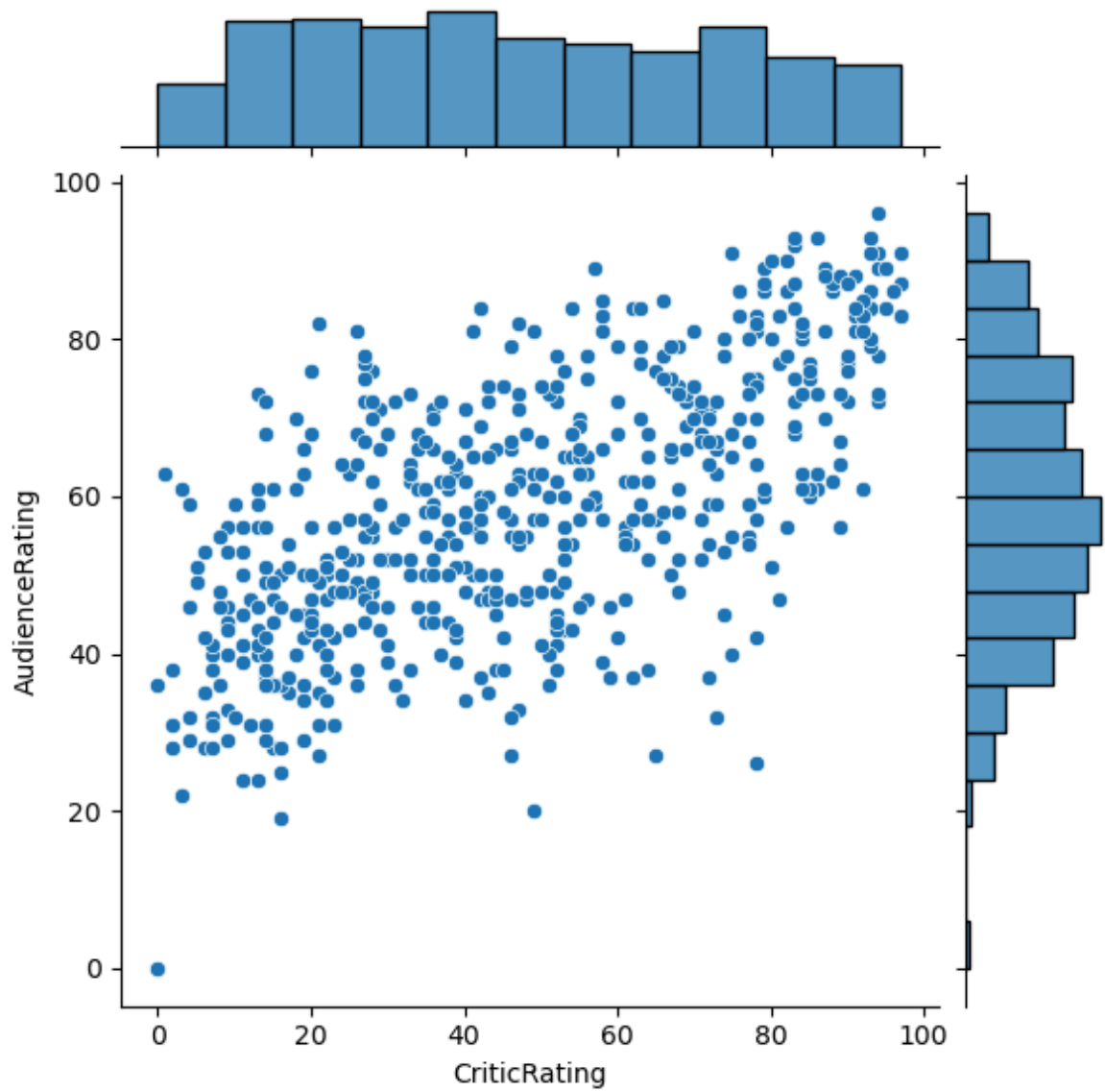
	CriticRating	AudienceRating	BudgetMillions
count	559.000000	559.000000	559.000000
mean	47.309481	58.744186	50.236136
std	26.413091	16.826887	48.731817
min	0.000000	0.000000	0.000000
25%	25.000000	47.000000	20.000000
50%	46.000000	58.000000	35.000000
75%	70.000000	72.000000	65.000000
max	97.000000	96.000000	300.000000

```
In [69]: from matplotlib import pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

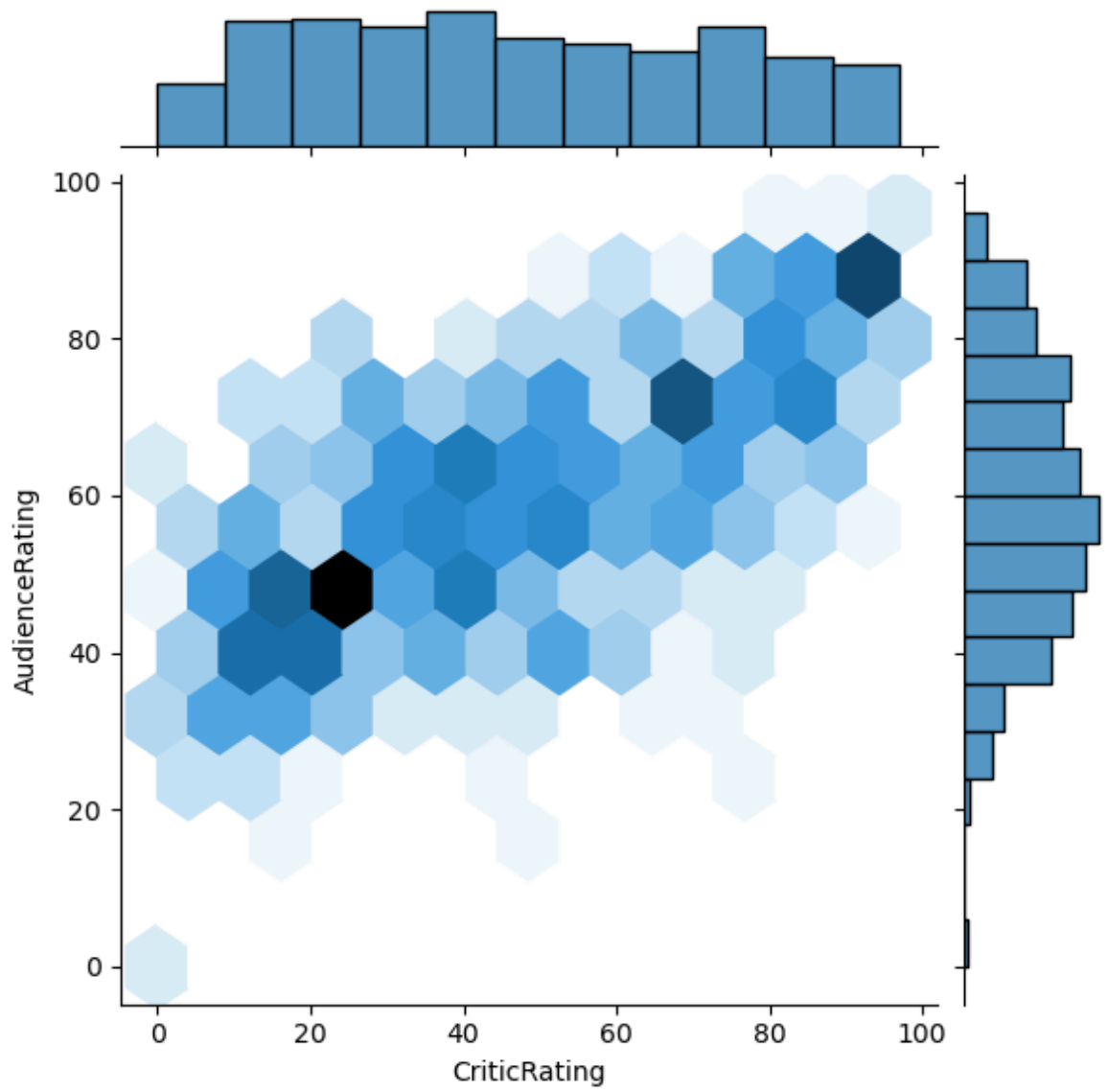
```
In [75]: j = sns.jointplot( data = movies, x = 'CriticRating', y = 'AudienceRating')
plt.show()
```



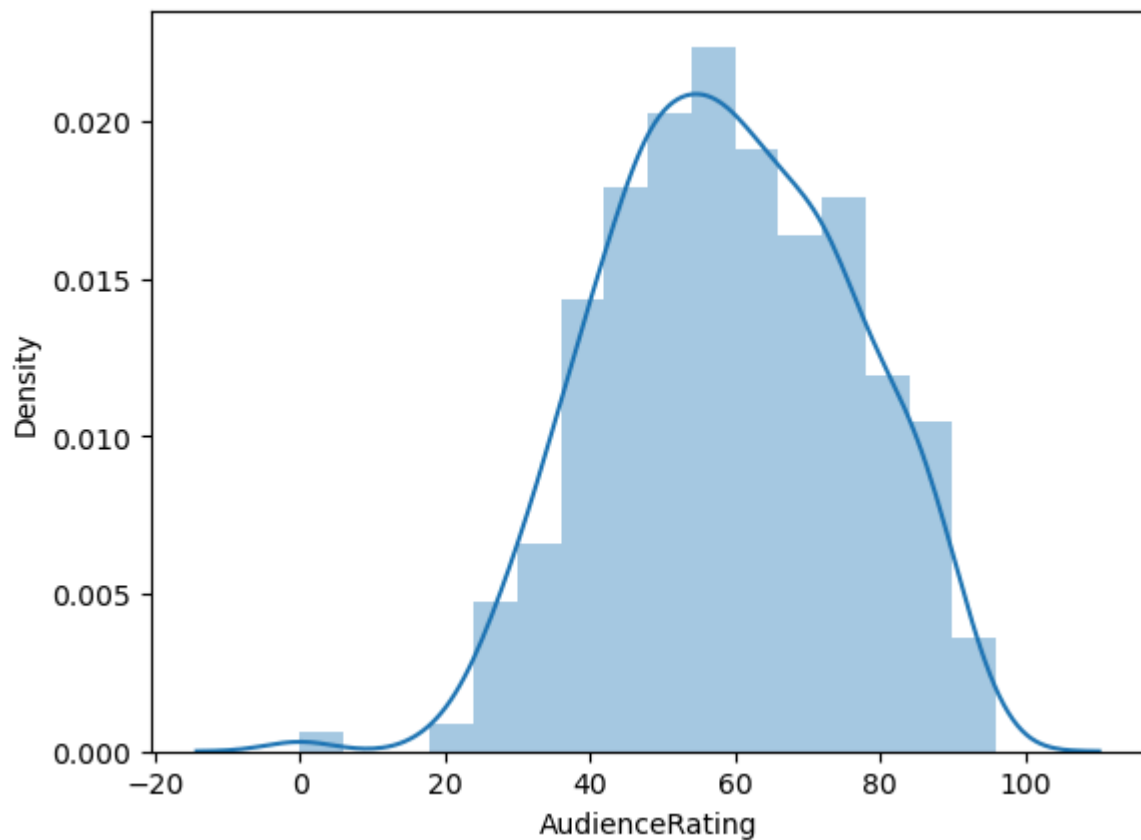




```
In [77]: j = sns.jointplot( data = movies, x = 'CriticRating', y = 'AudienceRating', kind  
plt.show()
```

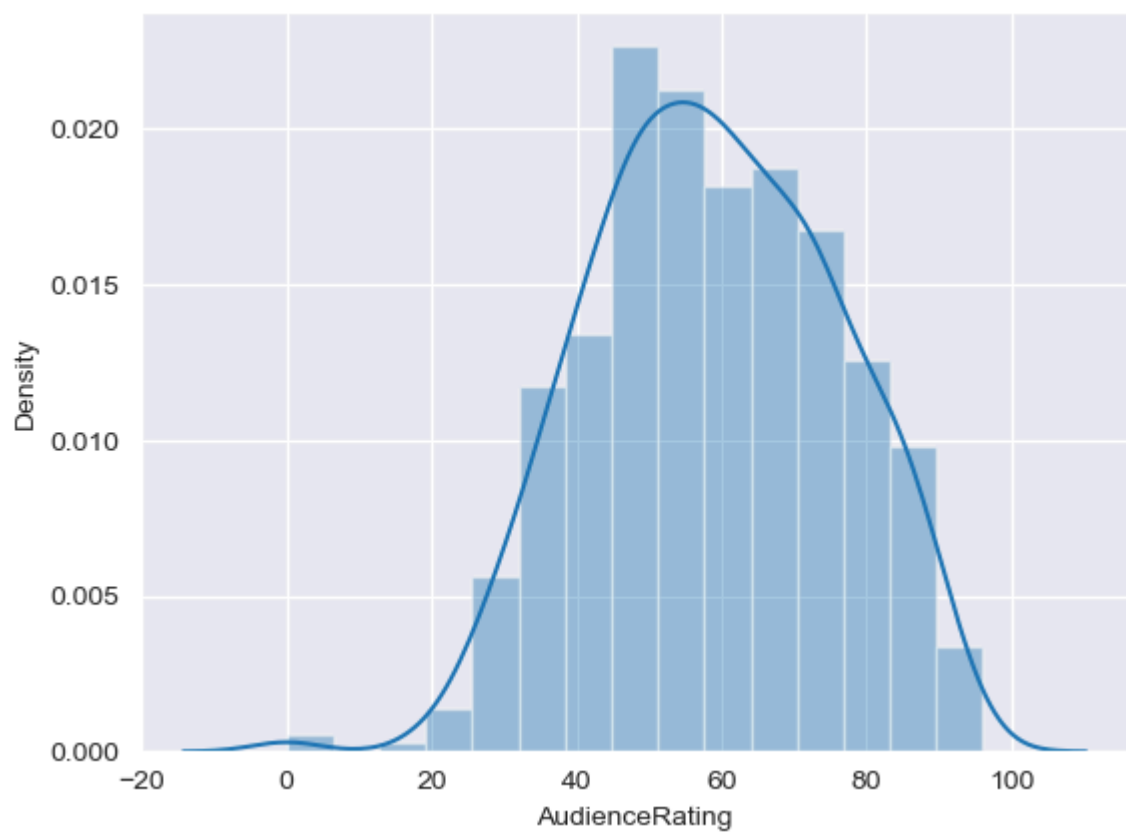


```
In [79]: m1 = sns.distplot(movies.AudienceRating)
plt.show()
```

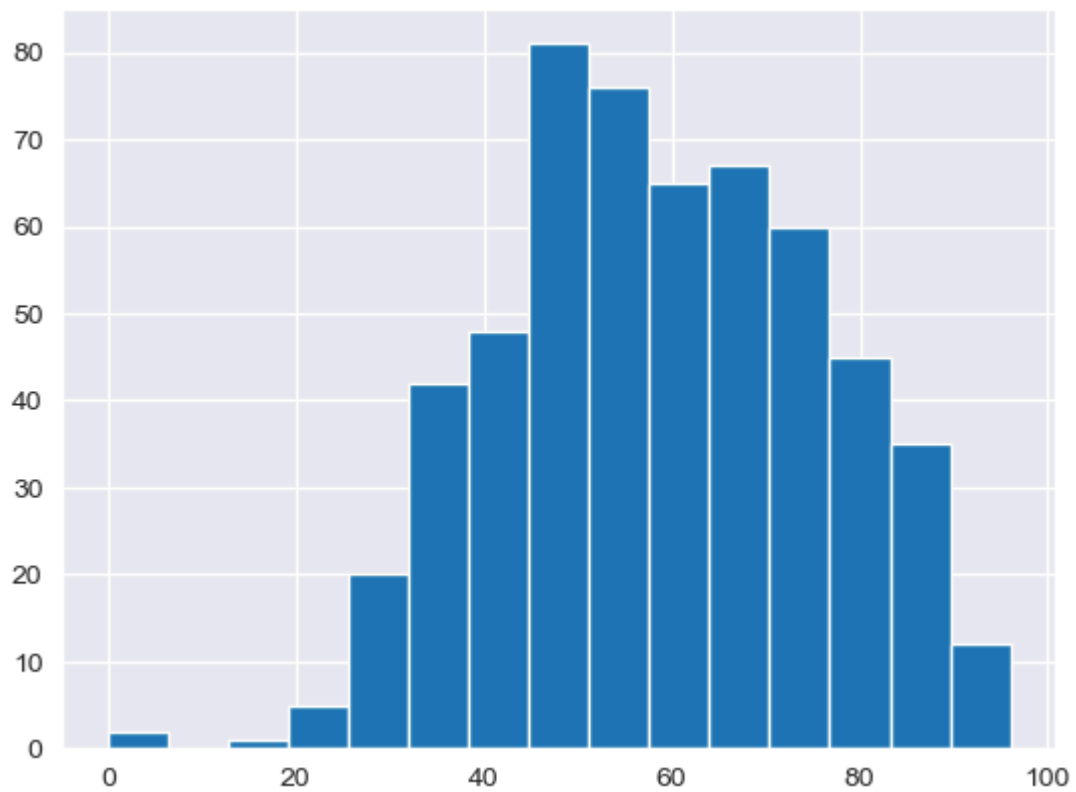


```
In [81]: sns.set_style('darkgrid')
```

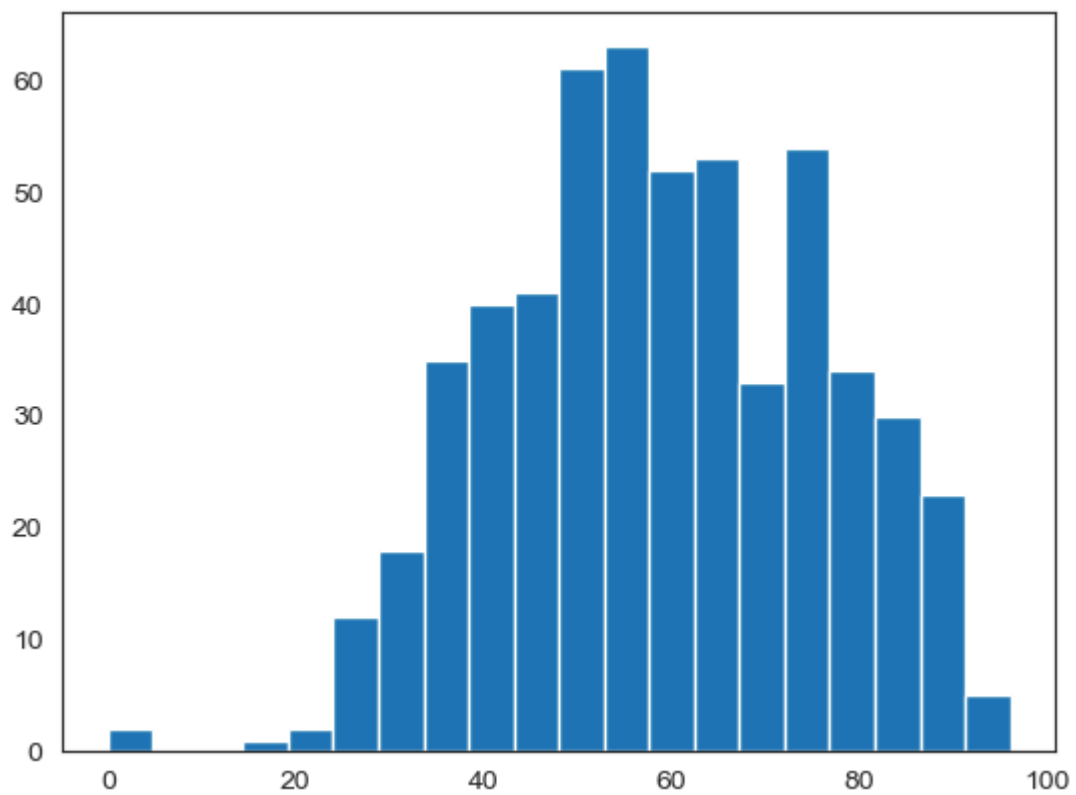
```
In [83]: m2 = sns.distplot(movies.AudienceRating, bins = 15)  
plt.show()
```



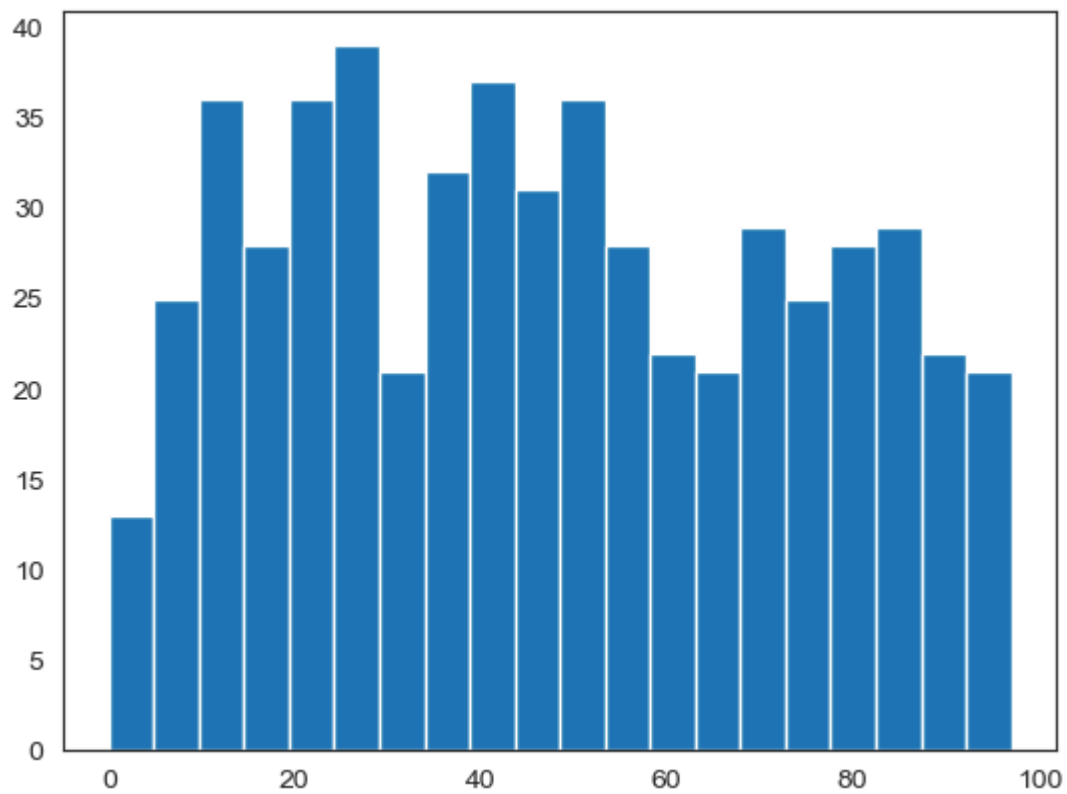
```
In [85]: n1 = plt.hist(movies.AudienceRating, bins=15)  
plt.show()
```



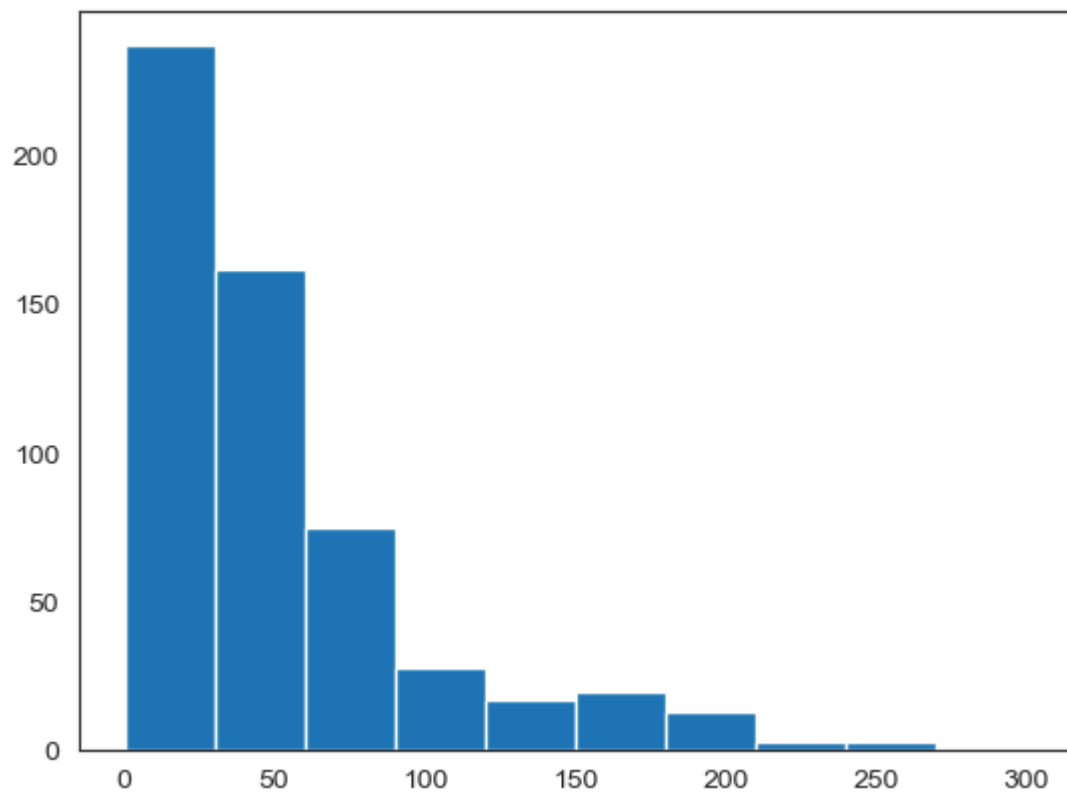
```
In [87]: sns.set_style('white')
n1 = plt.hist(movies.AudienceRating, bins=20)
plt.show()
```



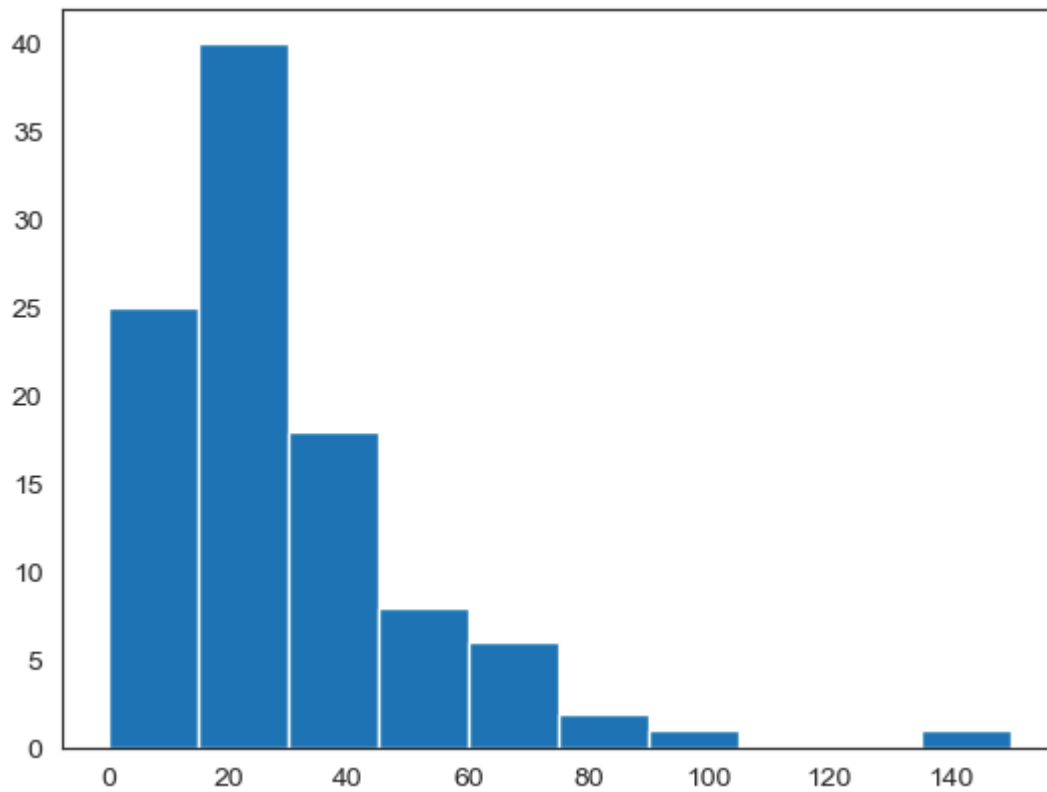
```
In [89]: n1 = plt.hist(movies.CriticRating, bins=20)
plt.show()
```



```
In [91]: plt.hist(movies.BudgetMillions)
plt.show()
```



```
In [93]: plt.hist(movies[movies.Genre == 'Drama'].BudgetMillions)
plt.show()
```

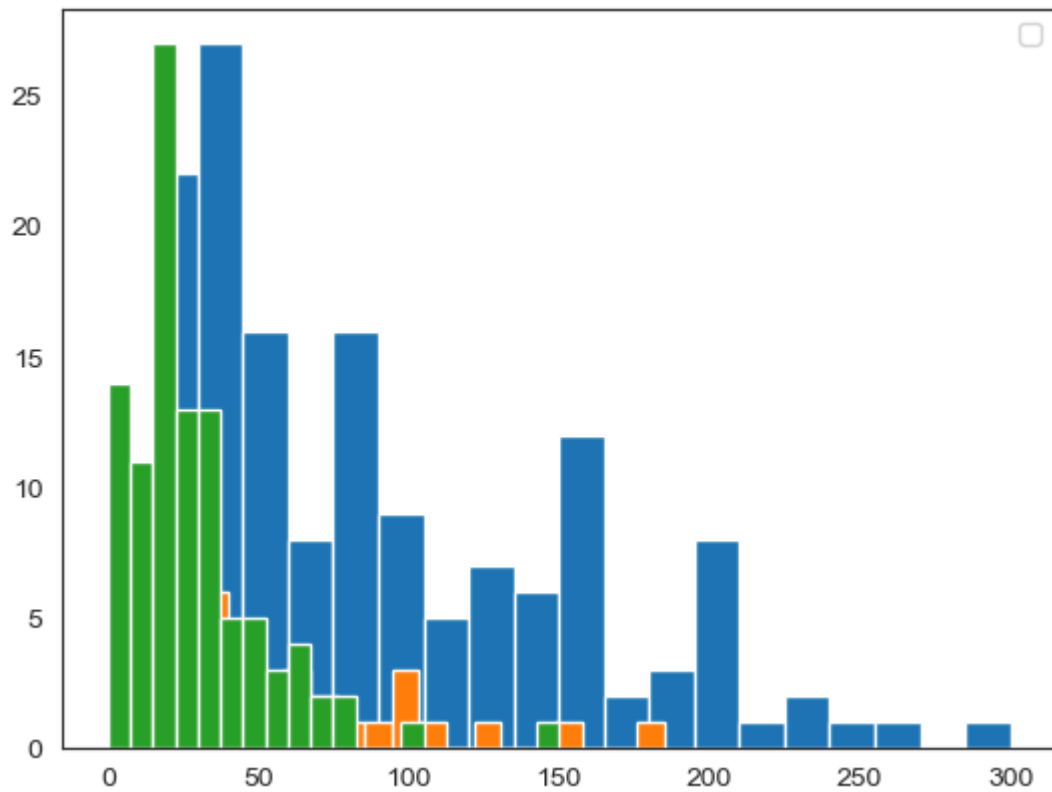


In [95]: `movies.head()`

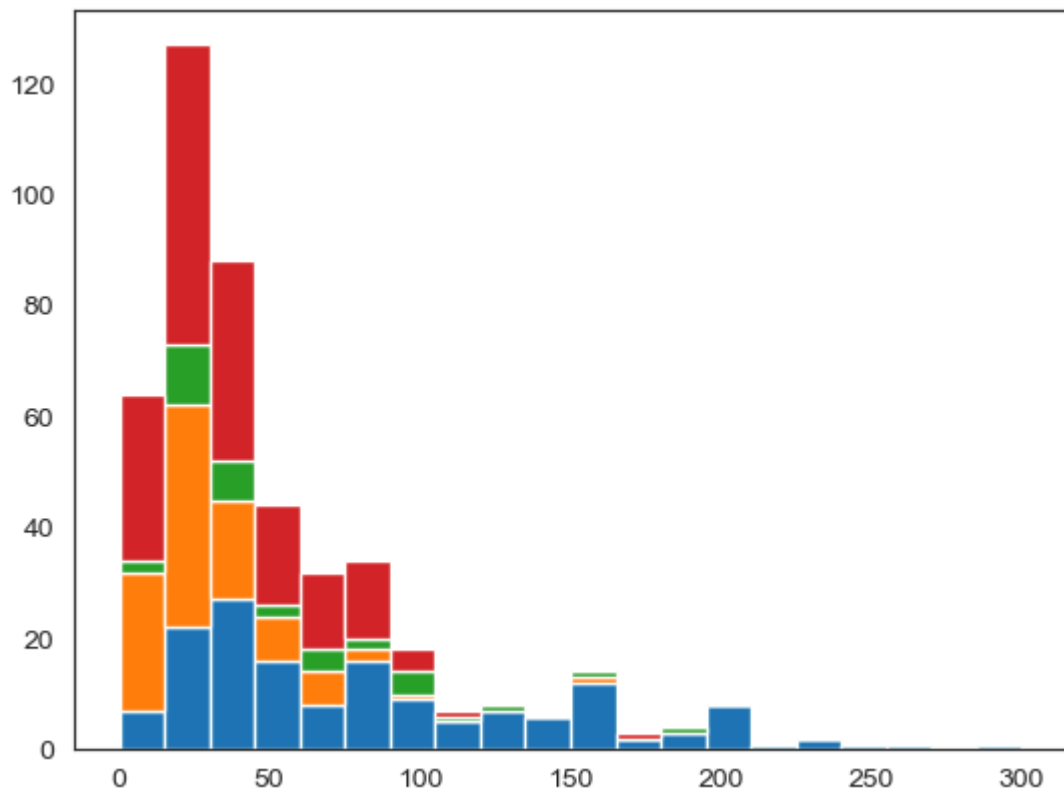
Out[95]:

	Film	Genre	CriticRating	AudienceRating	BudgetMillions	Year
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

In [97]: `plt.hist(movies[movies.Genre == 'Action'].BudgetMillions, bins = 20)`
`plt.hist(movies[movies.Genre == 'Thriller'].BudgetMillions, bins = 20)`
`plt.hist(movies[movies.Genre == 'Drama'].BudgetMillions, bins = 20)`
`plt.legend()`
`plt.show()`



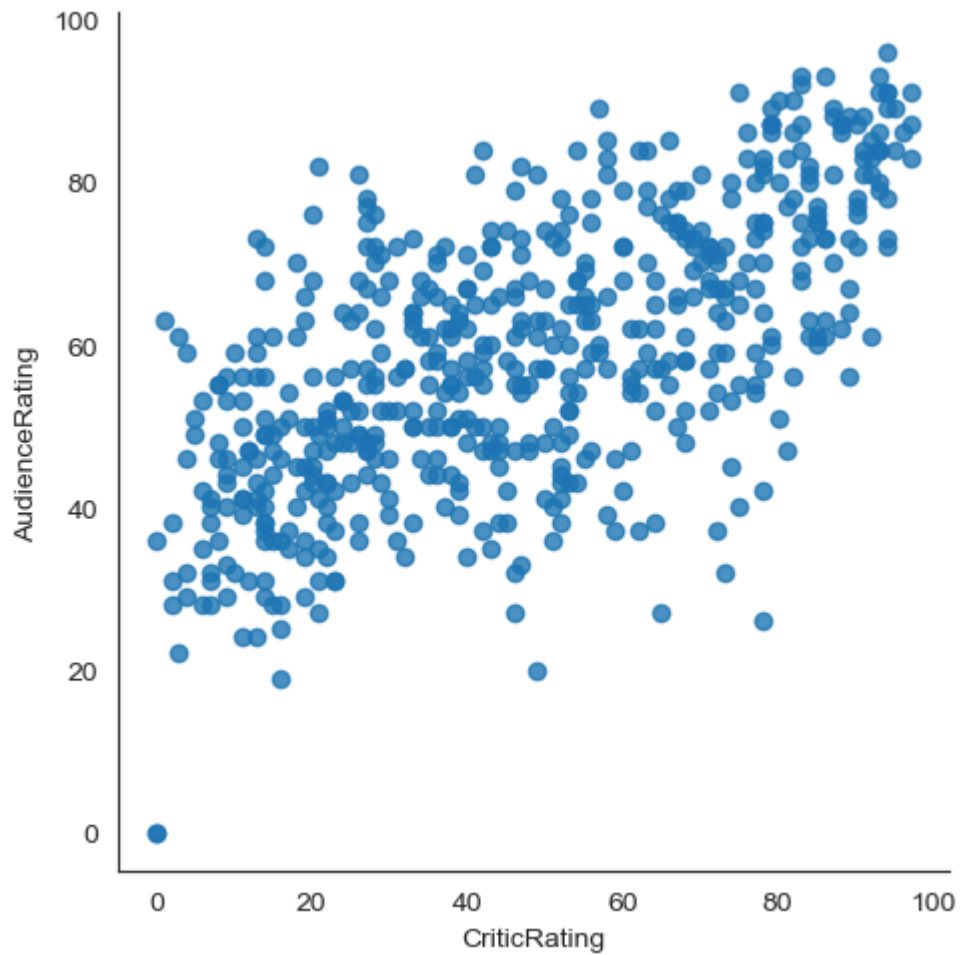
```
In [99]: plt.hist([movies[movies.Genre == 'Action'].BudgetMillions,\
    movies[movies.Genre == 'Drama'].BudgetMillions, \
    movies[movies.Genre == 'Thriller'].BudgetMillions, \
    movies[movies.Genre == 'Comedy'].BudgetMillions],\
    bins = 20, stacked = True)\nplt.show()
```

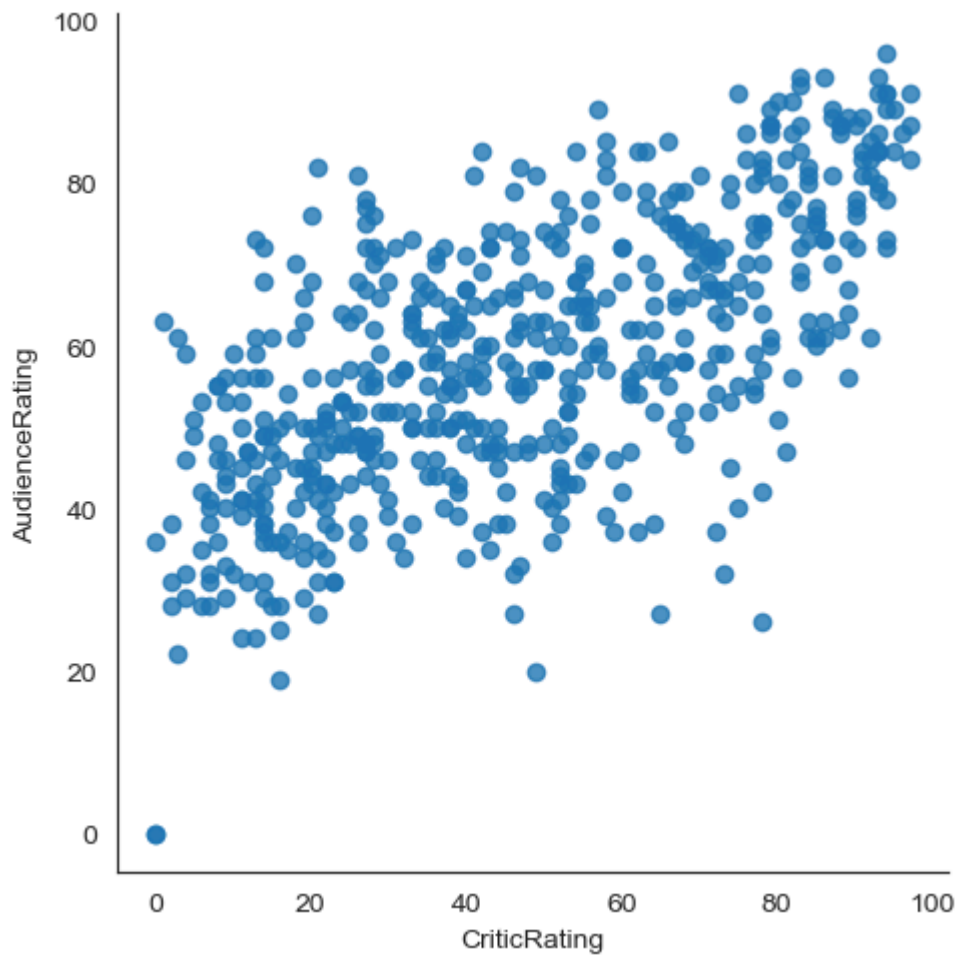


```
In [101... for gen in movies.Genre.cat.categories:\n    print(gen)
```

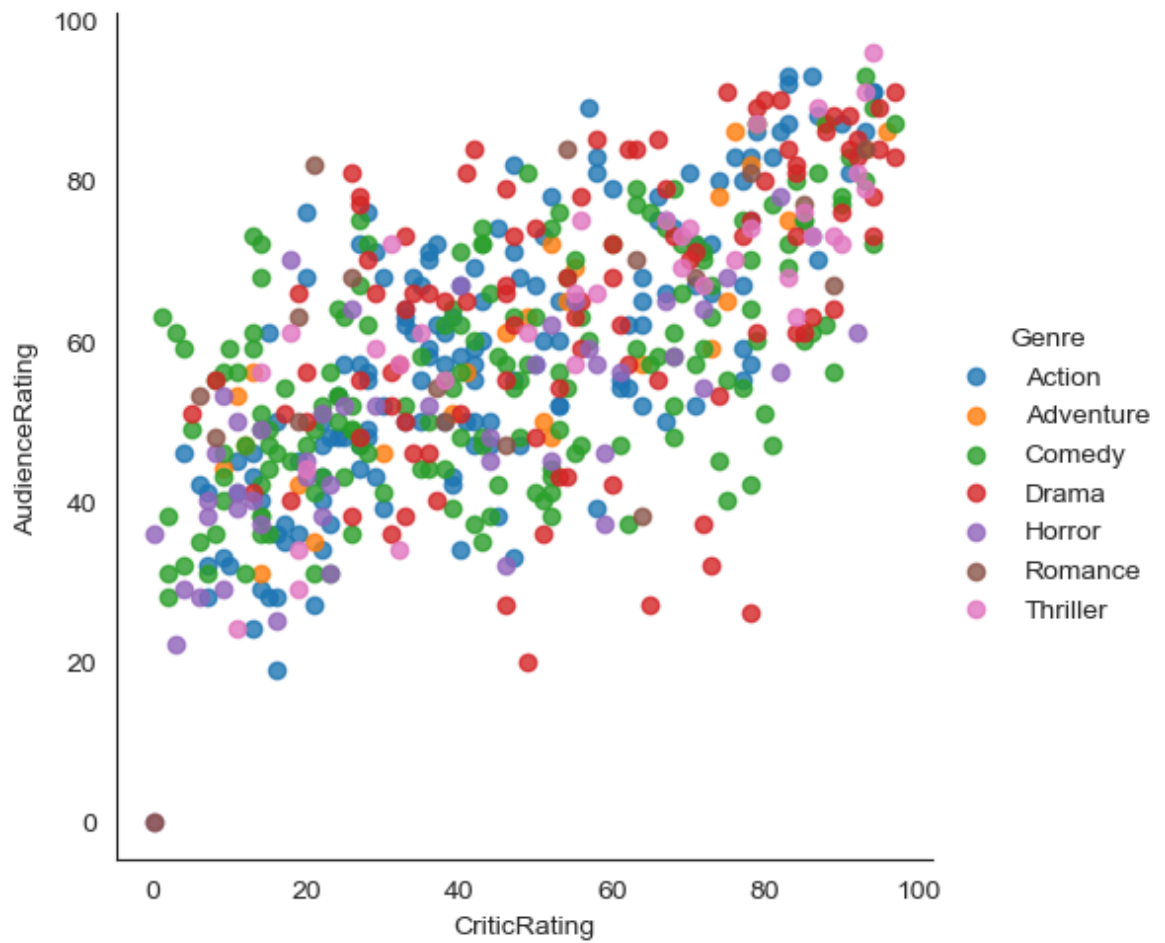
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller

```
In [105... vis1 = sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',\
                    fit_reg=False)
plt.show()
```

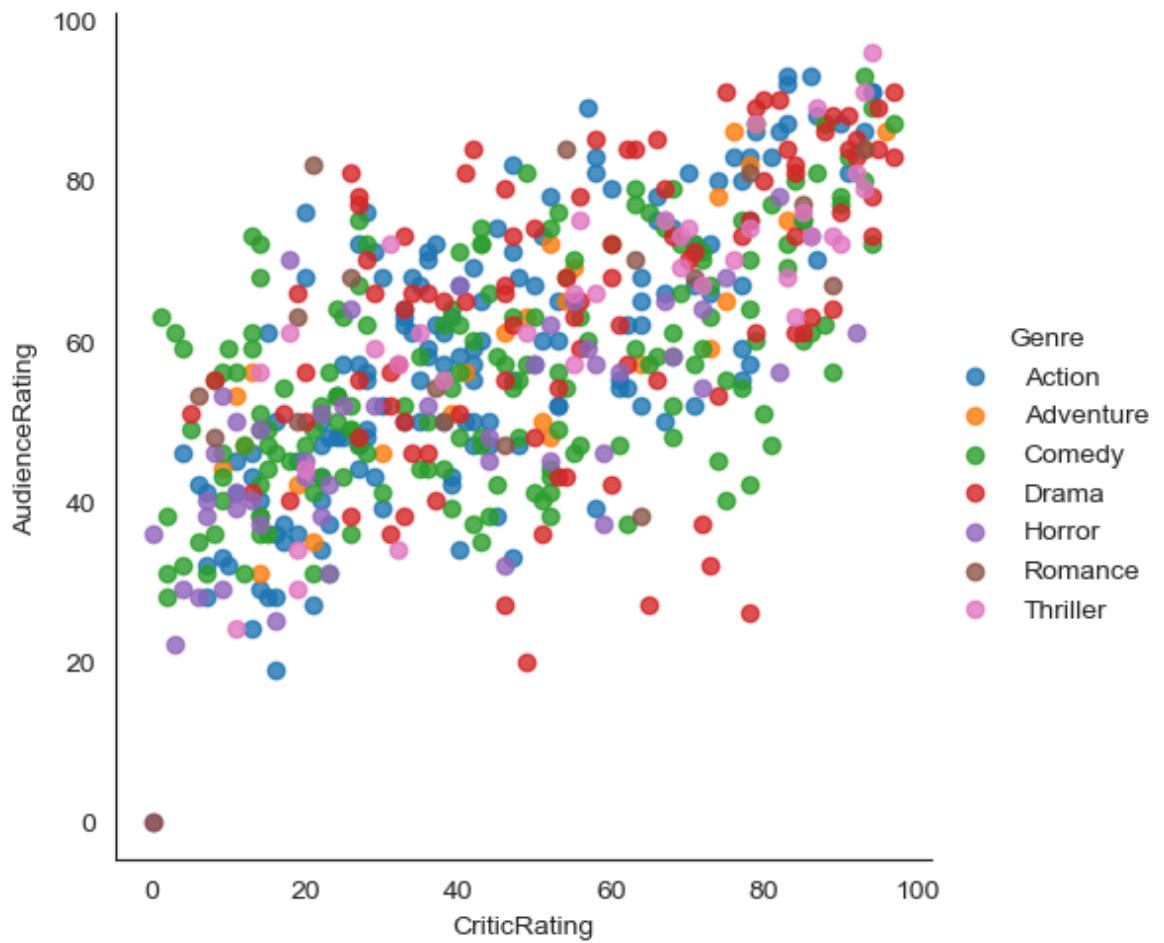




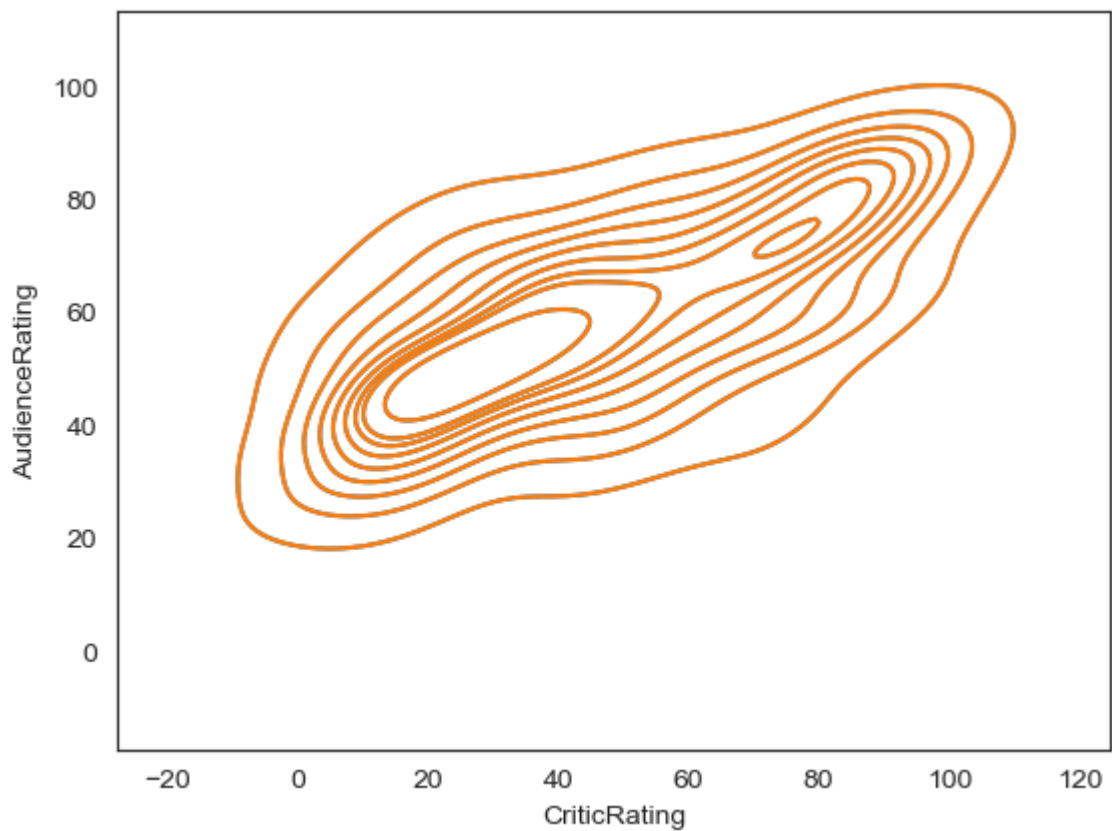
```
In [121... vis1 = sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',\
                    fit_reg=False, hue = 'Genre')
plt.show()
```



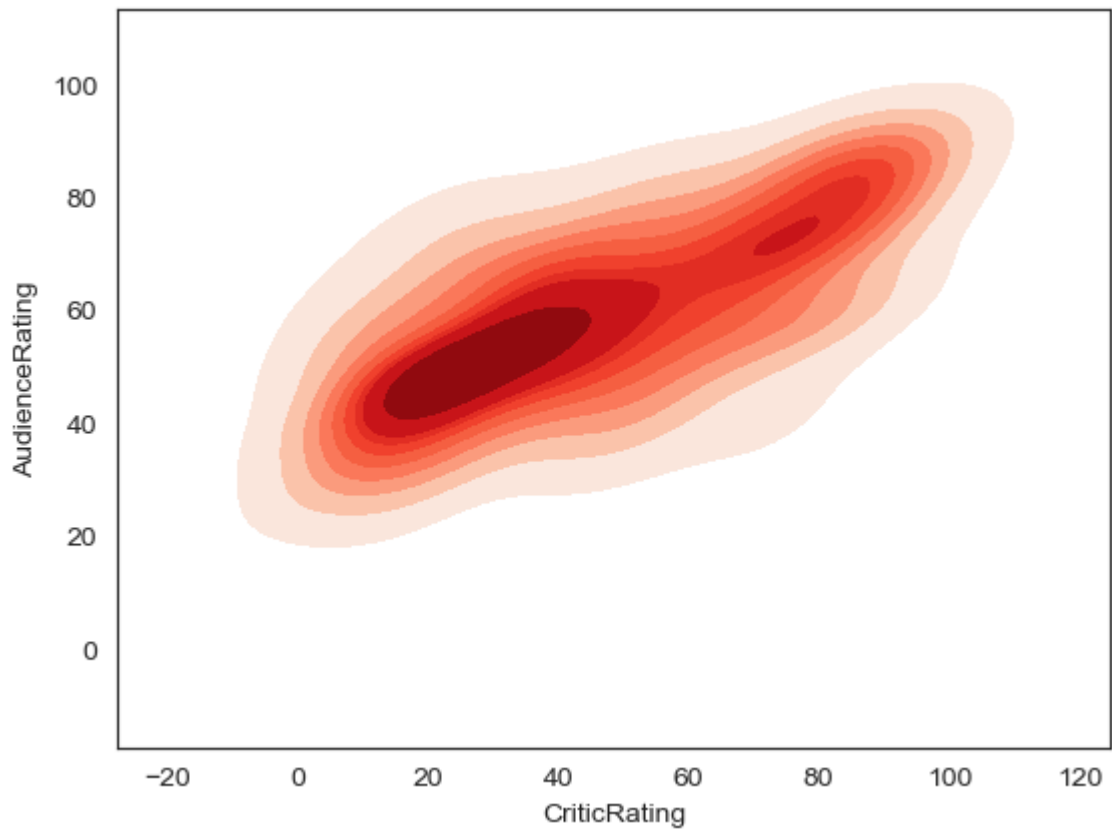
```
In [152... s1 = sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',\
                 fit_reg=False, hue = 'Genre', height =5, aspect=1)
plt.show()
```



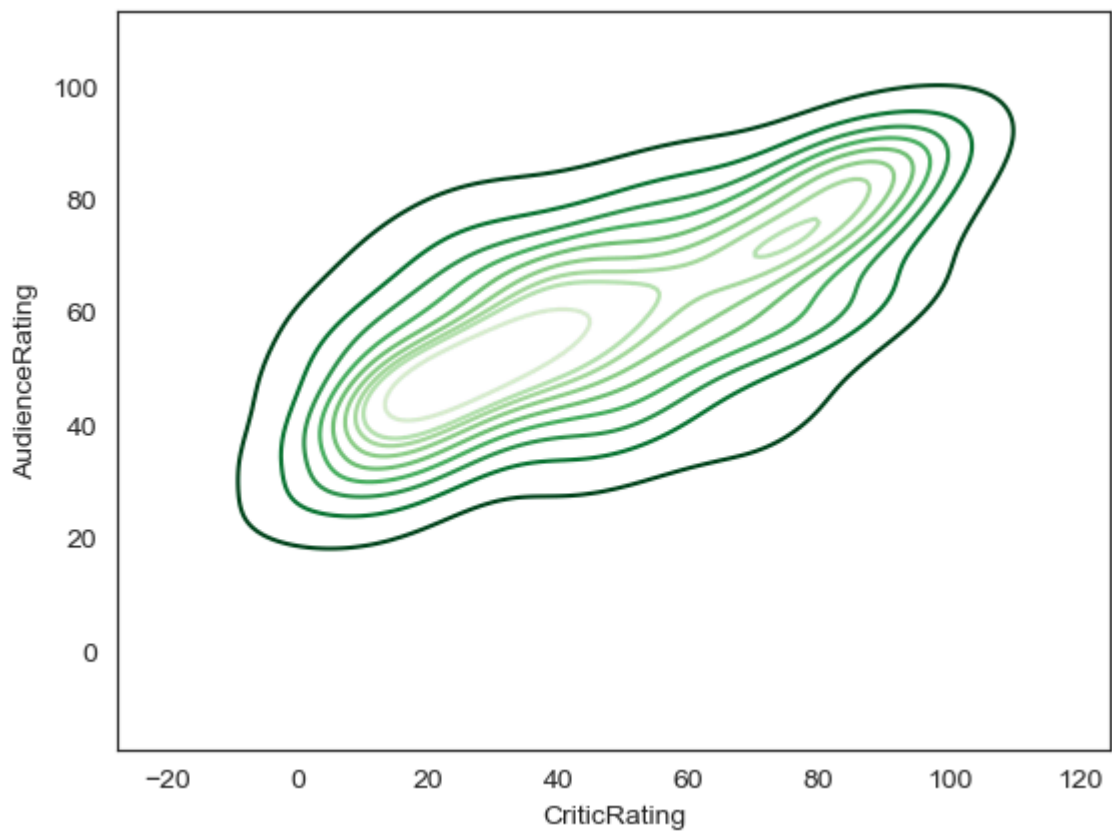
```
In [145... k1 = sns.kdeplot(x=movies.CriticRating, y=movies.AudienceRating)  
plt.show()
```



```
In [156... k1 = sns.kdeplot(x=movies.CriticRating,y=movies.AudienceRating,shade = True,shad
plt.show()
```

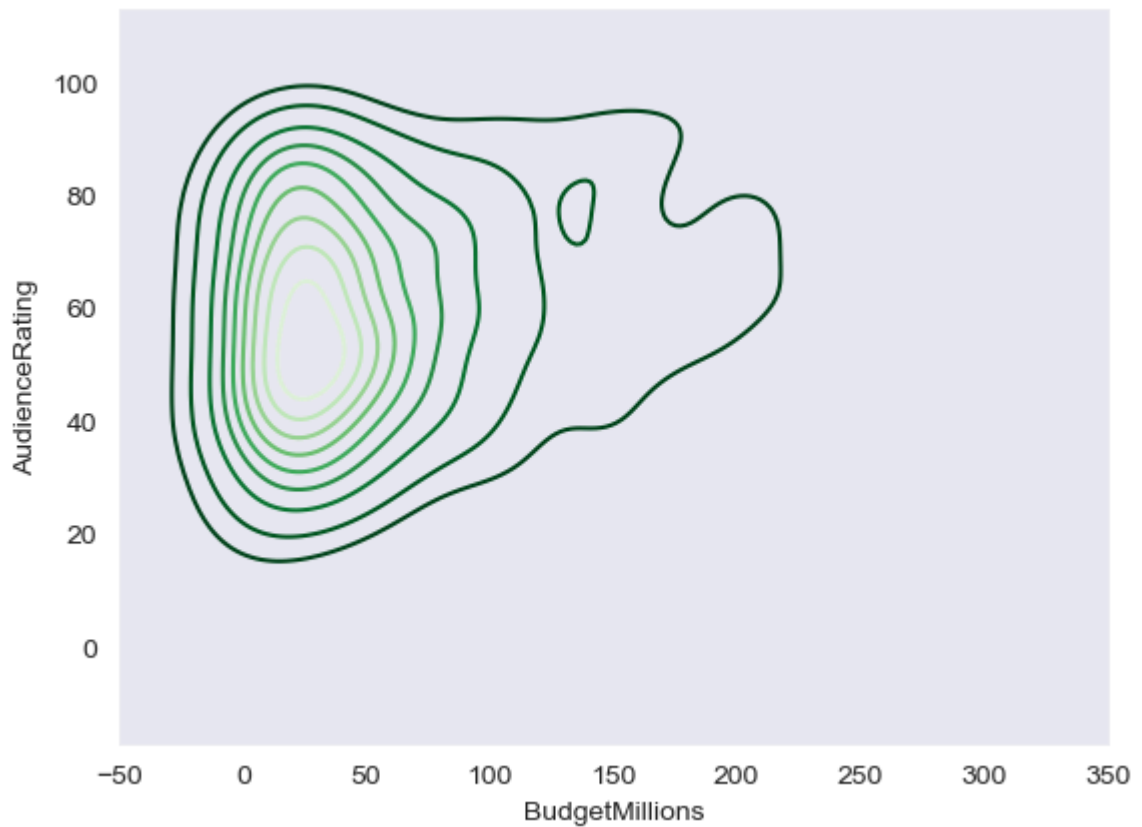


```
In [158... k2 = sns.kdeplot(x=movies.CriticRating,y=movies.AudienceRating,shade_lowest=False
plt.show()
```



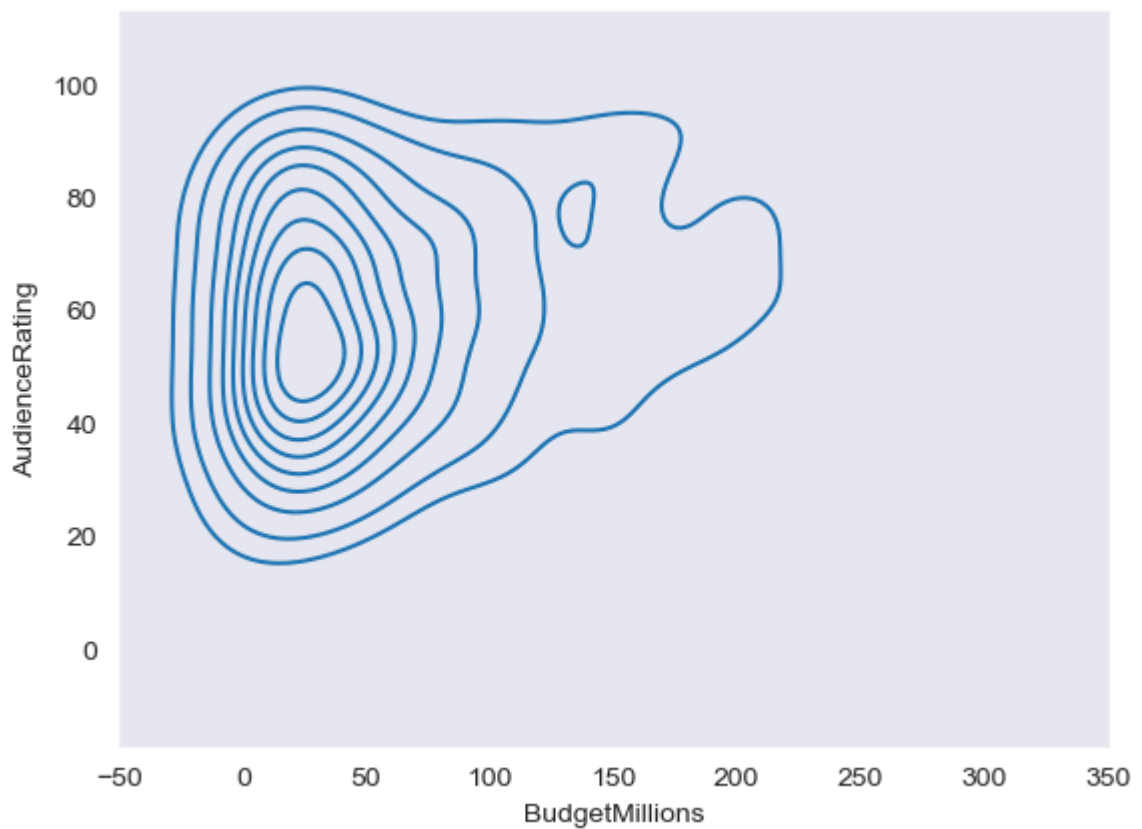
In [162...

```
sns.set_style('dark')
k1 = sns.kdeplot(x=movies.BudgetMillions,y=movies.AudienceRating,shade_lowest=False)
plt.show()
```

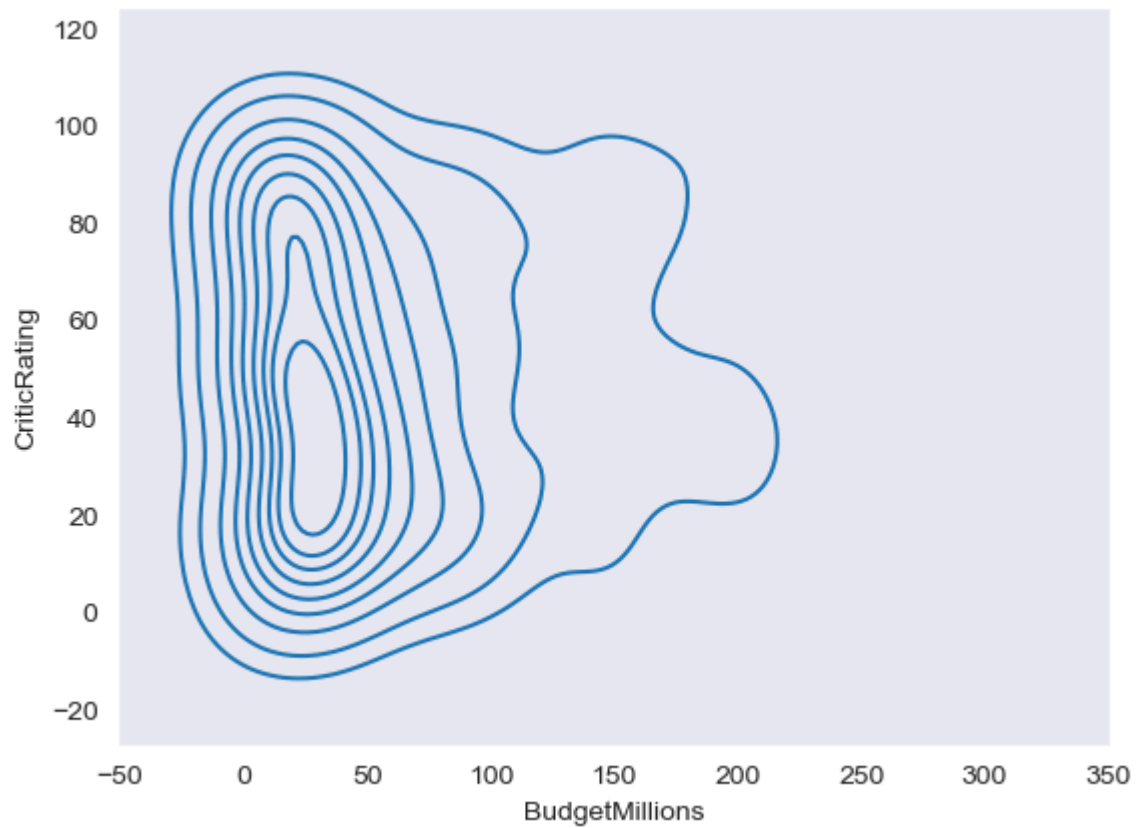


In [164...

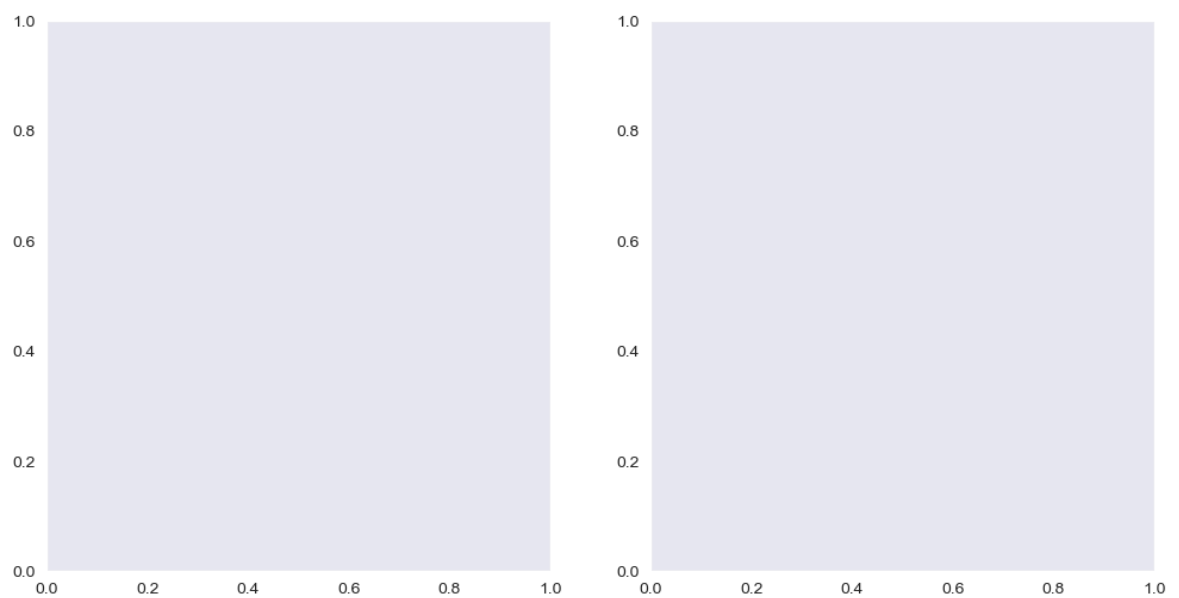
```
sns.set_style('dark')
k1 = sns.kdeplot(x=movies.BudgetMillions,y=movies.AudienceRating)
plt.show()
```

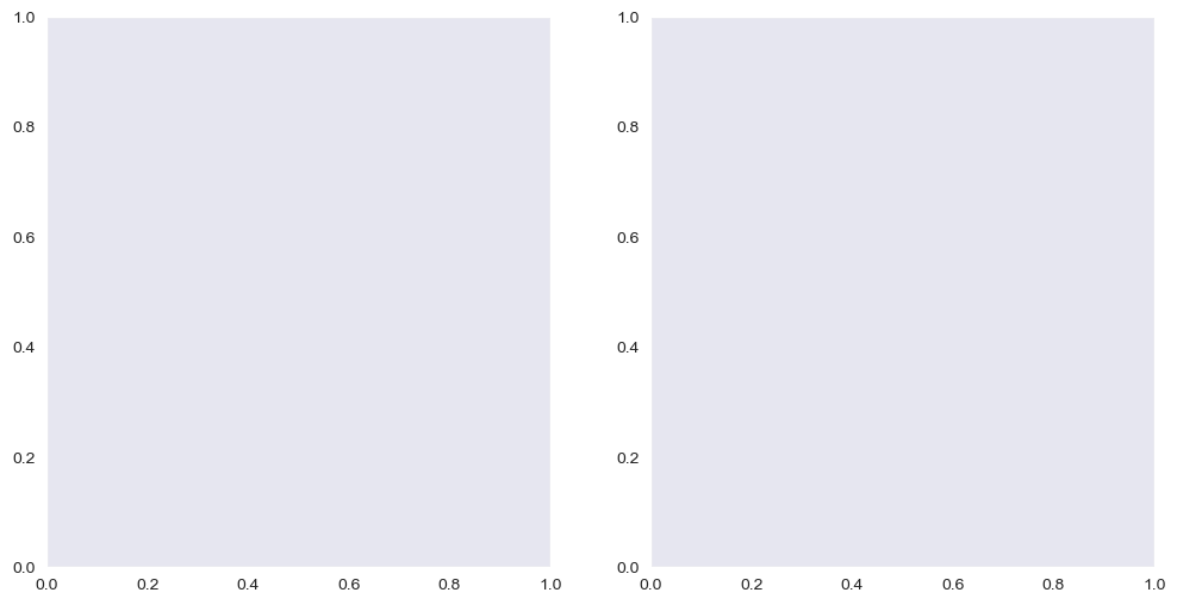


```
In [166... k2 = sns.kdeplot(x=movies.BudgetMillions,y=movies.CriticRating)
plt.show()
```

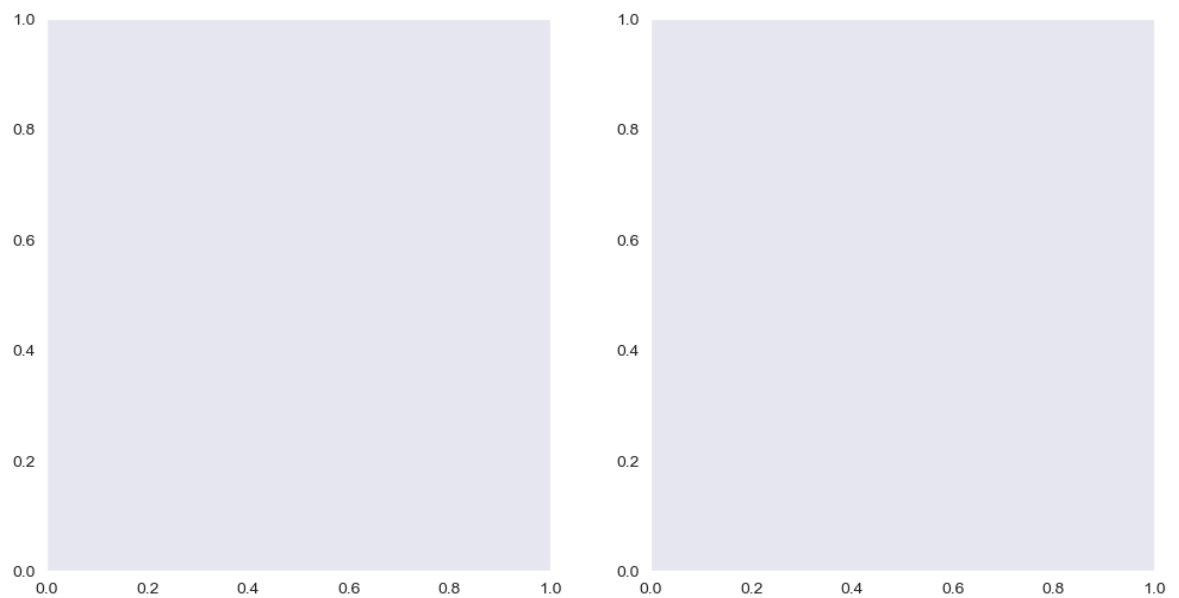


```
In [170... f, ax = plt.subplots(1,2, figsize =(12,6))
plt.show()
```

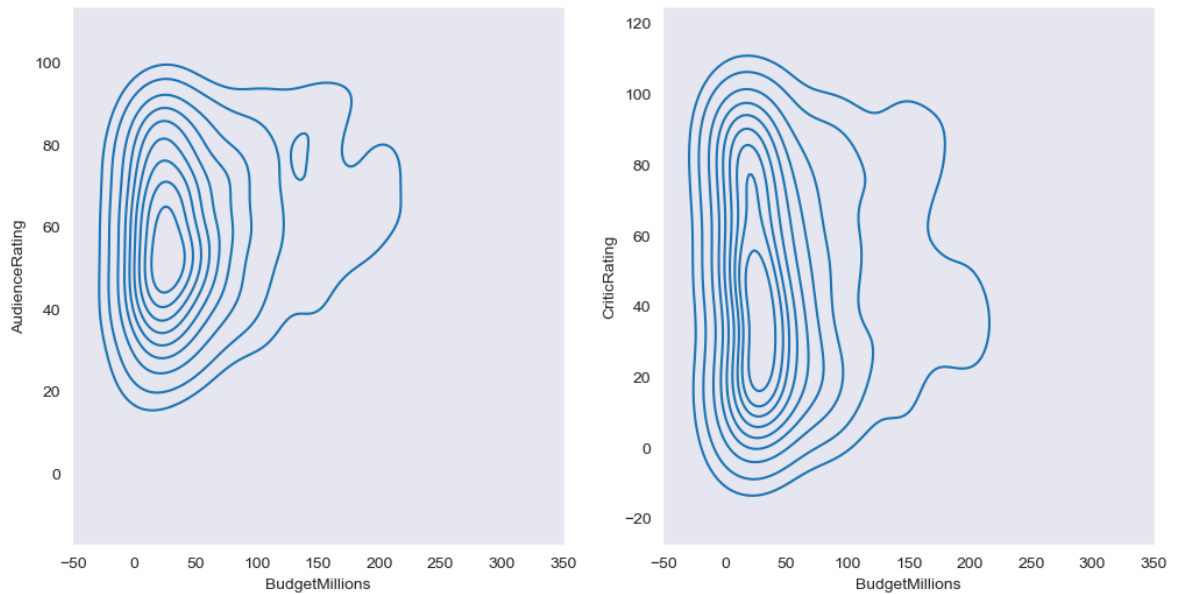




```
In [172... f, axes = plt.subplots(1,2, figsize =(12,6))  
plt.show()
```



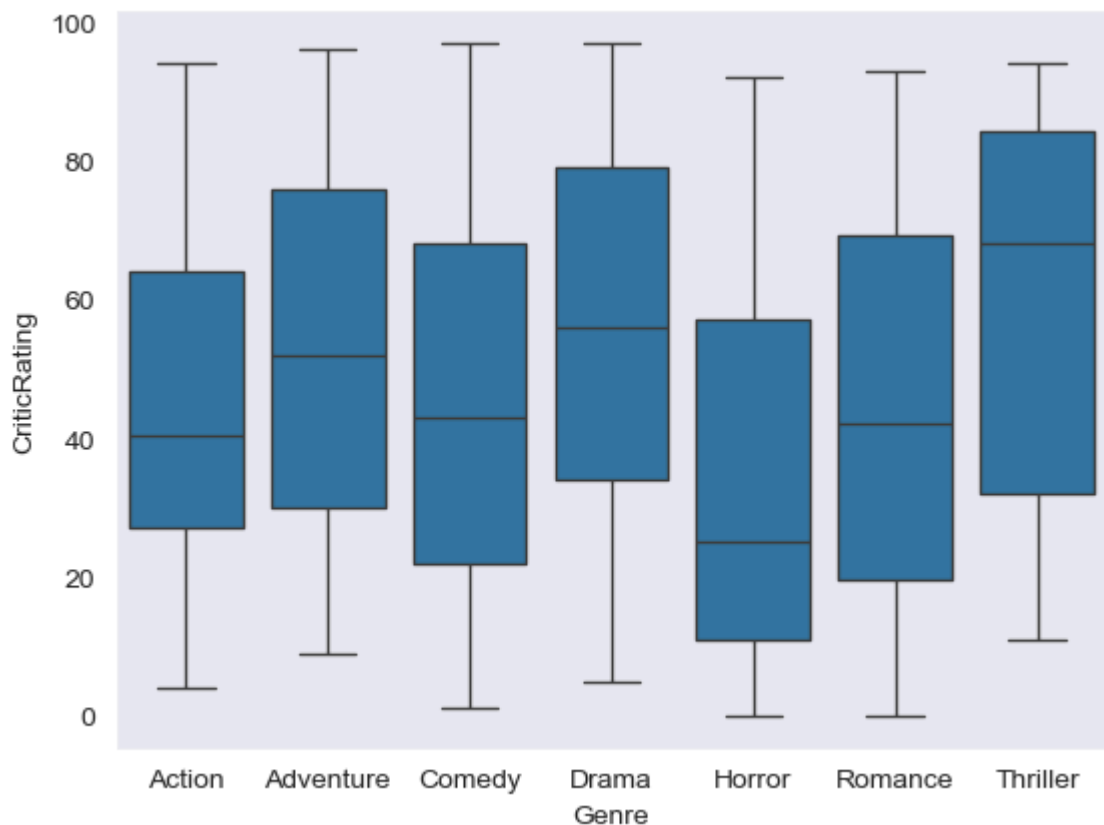
```
In [174... f, axes = plt.subplots(1,2, figsize =(12,6))  
  
k1 = sns.kdeplot(x=movies.BudgetMillions,y=movies.AudienceRating,ax=axes[0])  
k2 = sns.kdeplot(x=movies.BudgetMillions,y=movies.CriticRating,ax = axes[1])  
plt.show()
```



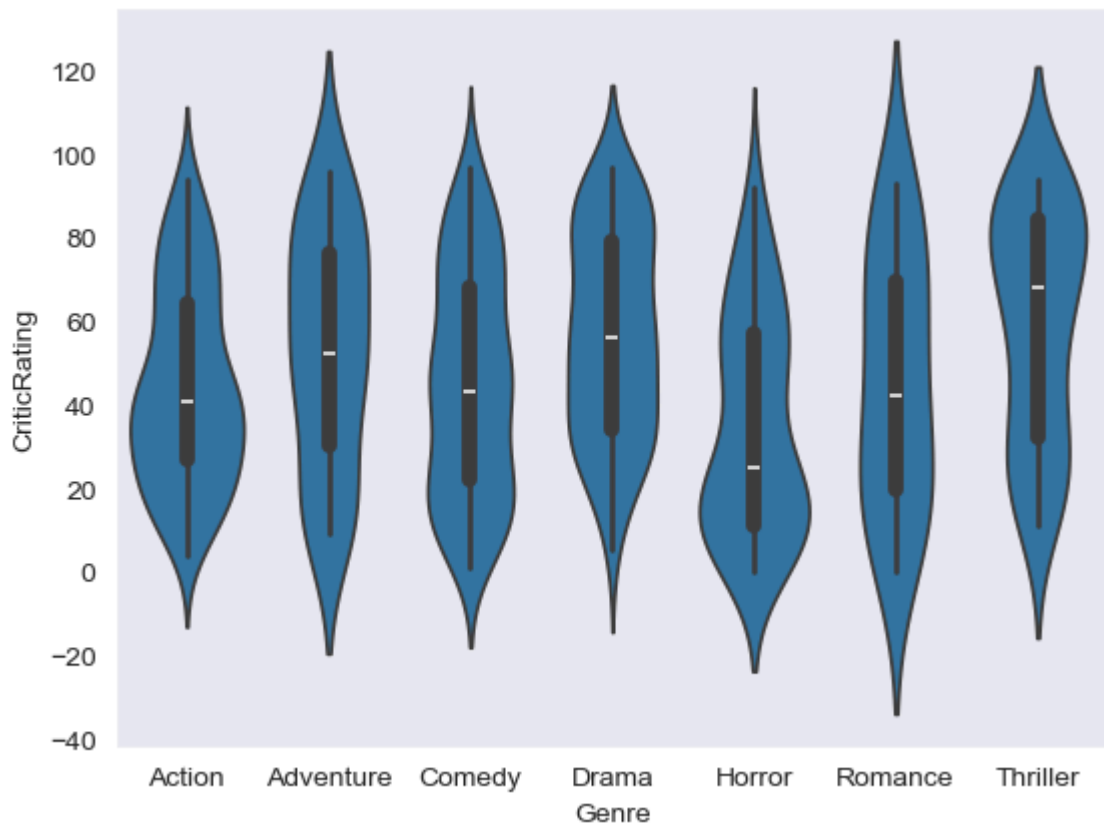
In [176... axes

```
Out[176... array([<Axes: xlabel='BudgetMillions', ylabel='AudienceRating'>,
      <Axes: xlabel='BudgetMillions', ylabel='CriticRating'>],
      dtype=object)
```

```
In [193... w = sns.boxplot(data=movies, x='Genre', y = 'CriticRating')
plt.show()
```

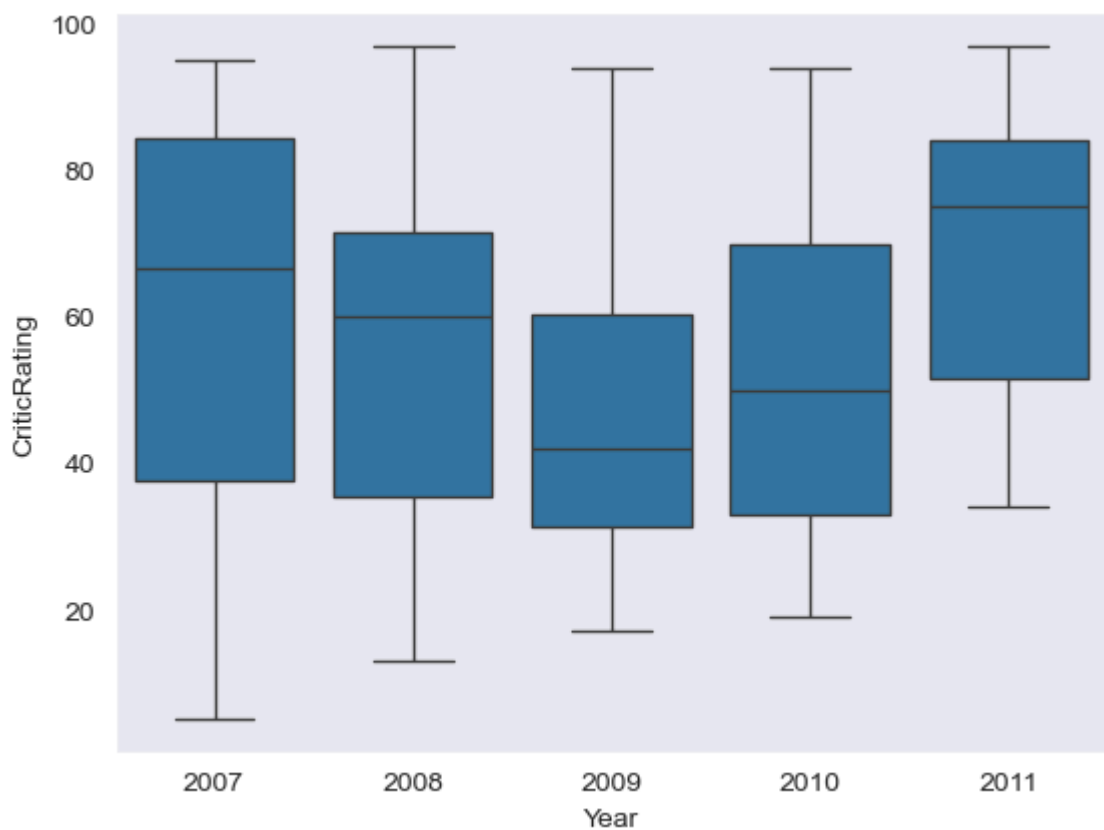


```
In [195... z = sns.violinplot(data=movies, x='Genre', y = 'CriticRating')
plt.show()
```

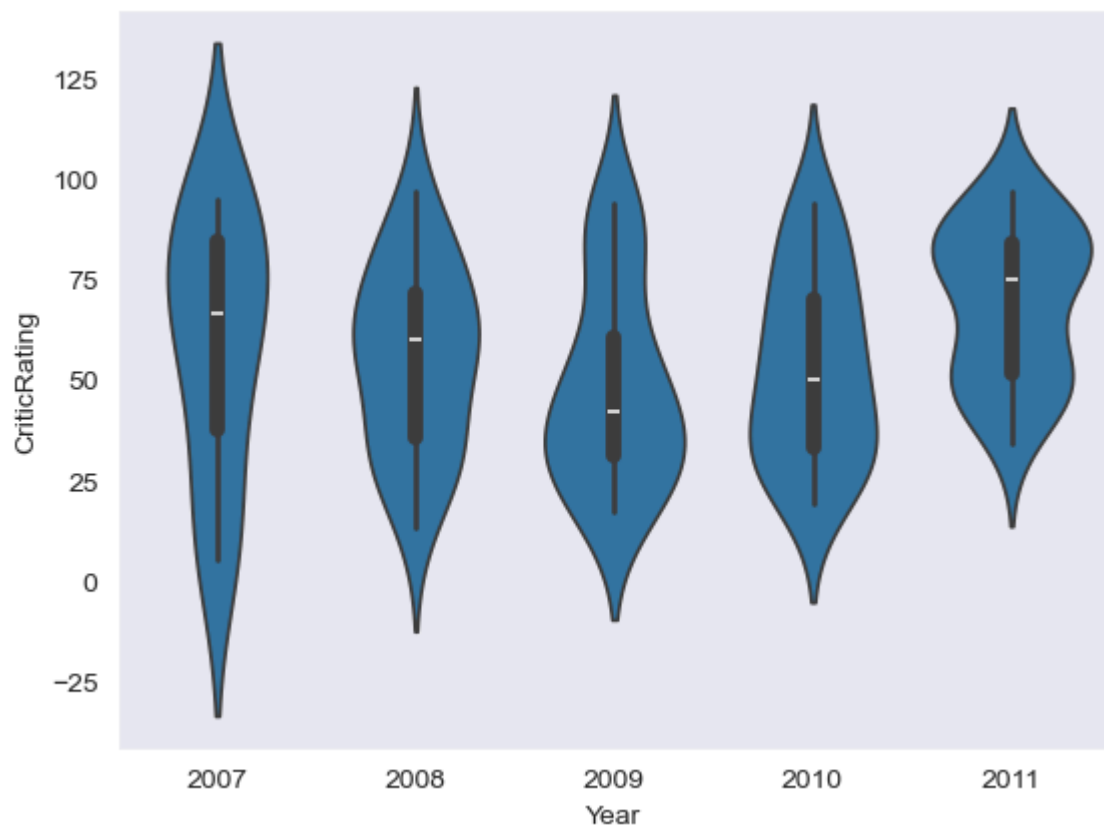
In [204...

```
w1 = sns.boxplot(data=movies[movies.Genre == 'Drama'], x='Year', y = 'CriticRati
plt.show()
```

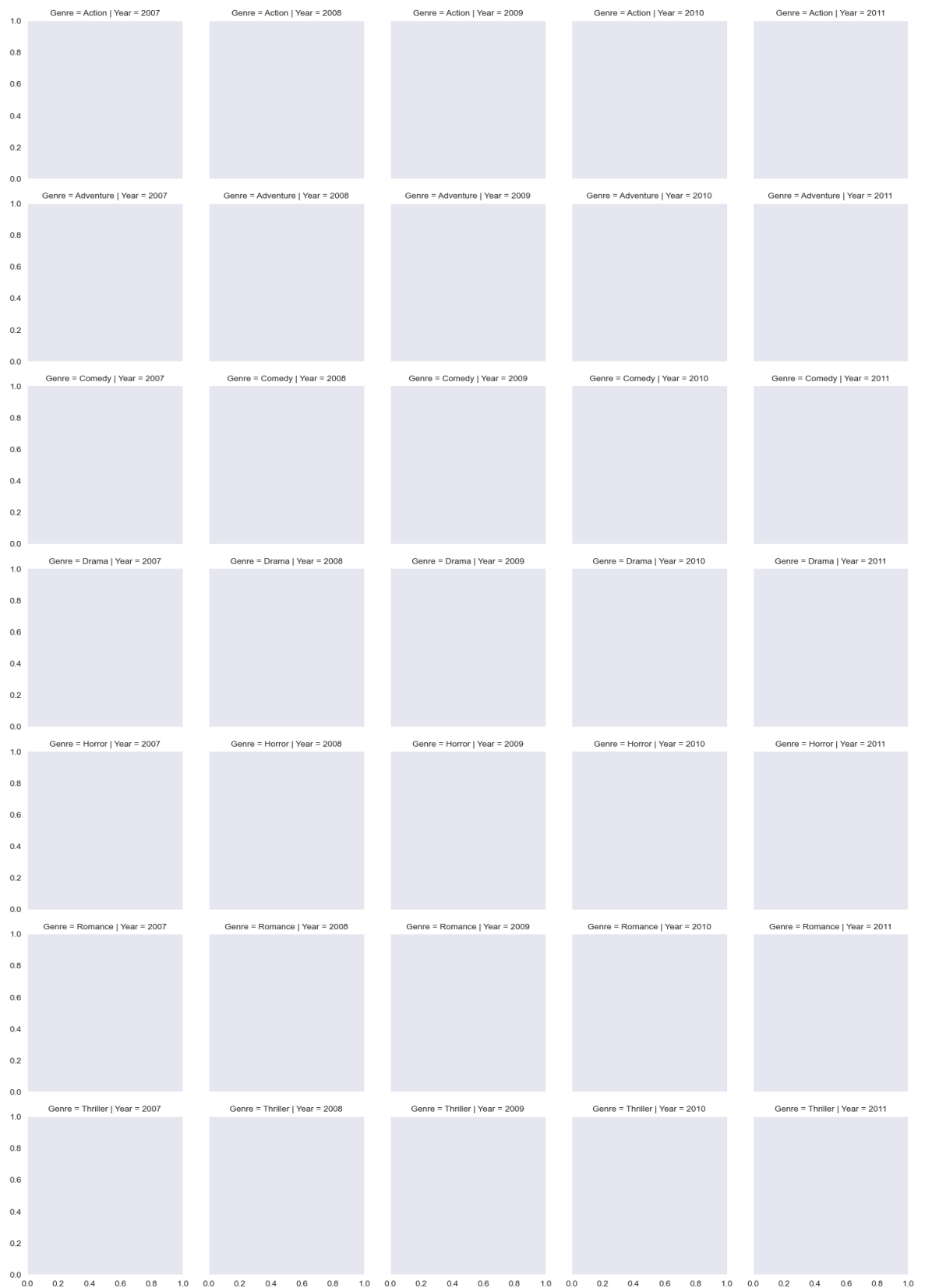


In [206...

```
z = sns.violinplot(data=movies[movies.Genre == 'Drama'], x='Year', y = 'CriticRa
plt.show()
```

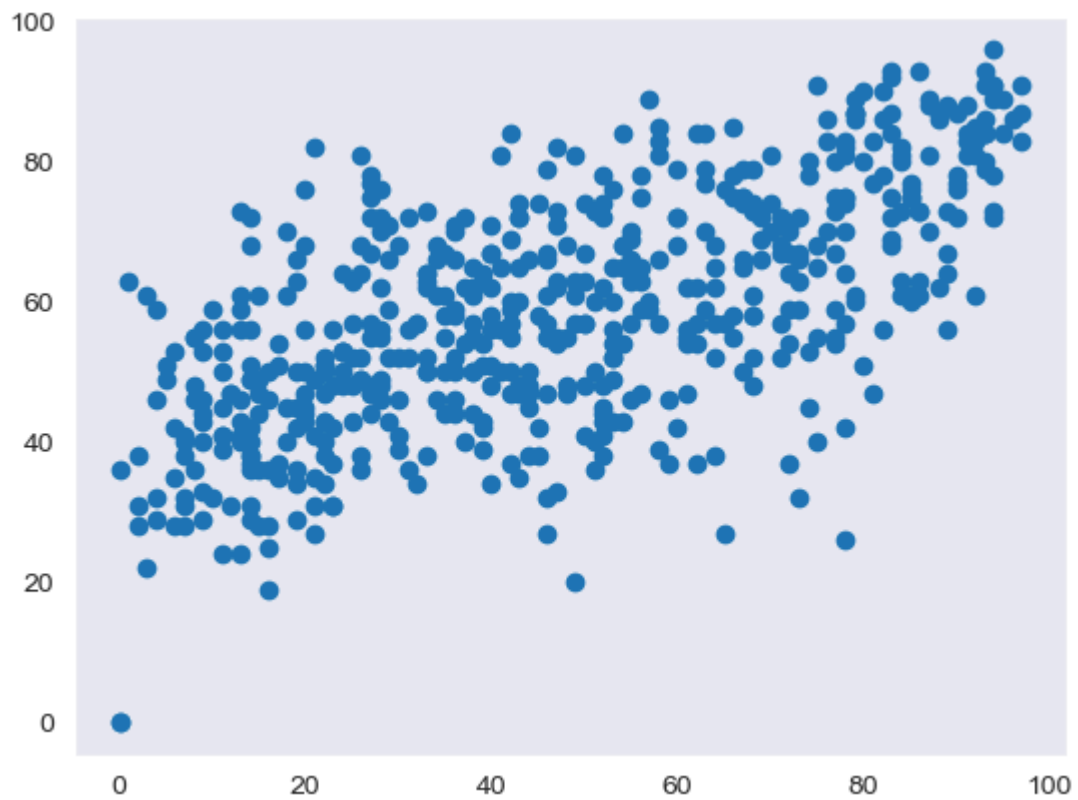


```
In [208... g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
plt.show()
```

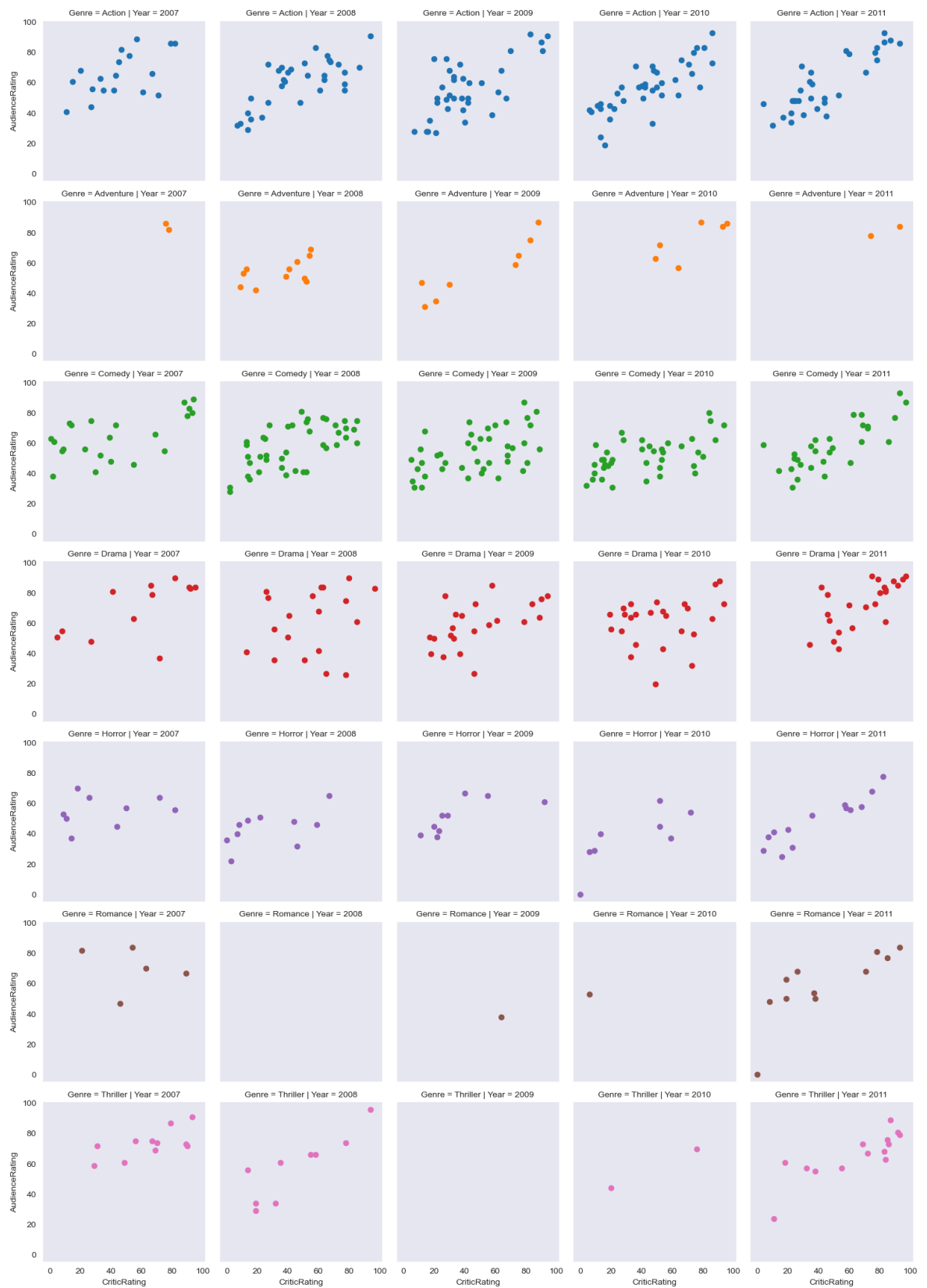


In [210...

```
plt.scatter(movies.CriticRating,movies.AudienceRating)
plt.show()
```



```
In [212... g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
g = g.map(plt.scatter, 'CriticRating', 'AudienceRating' )
plt.show()
```



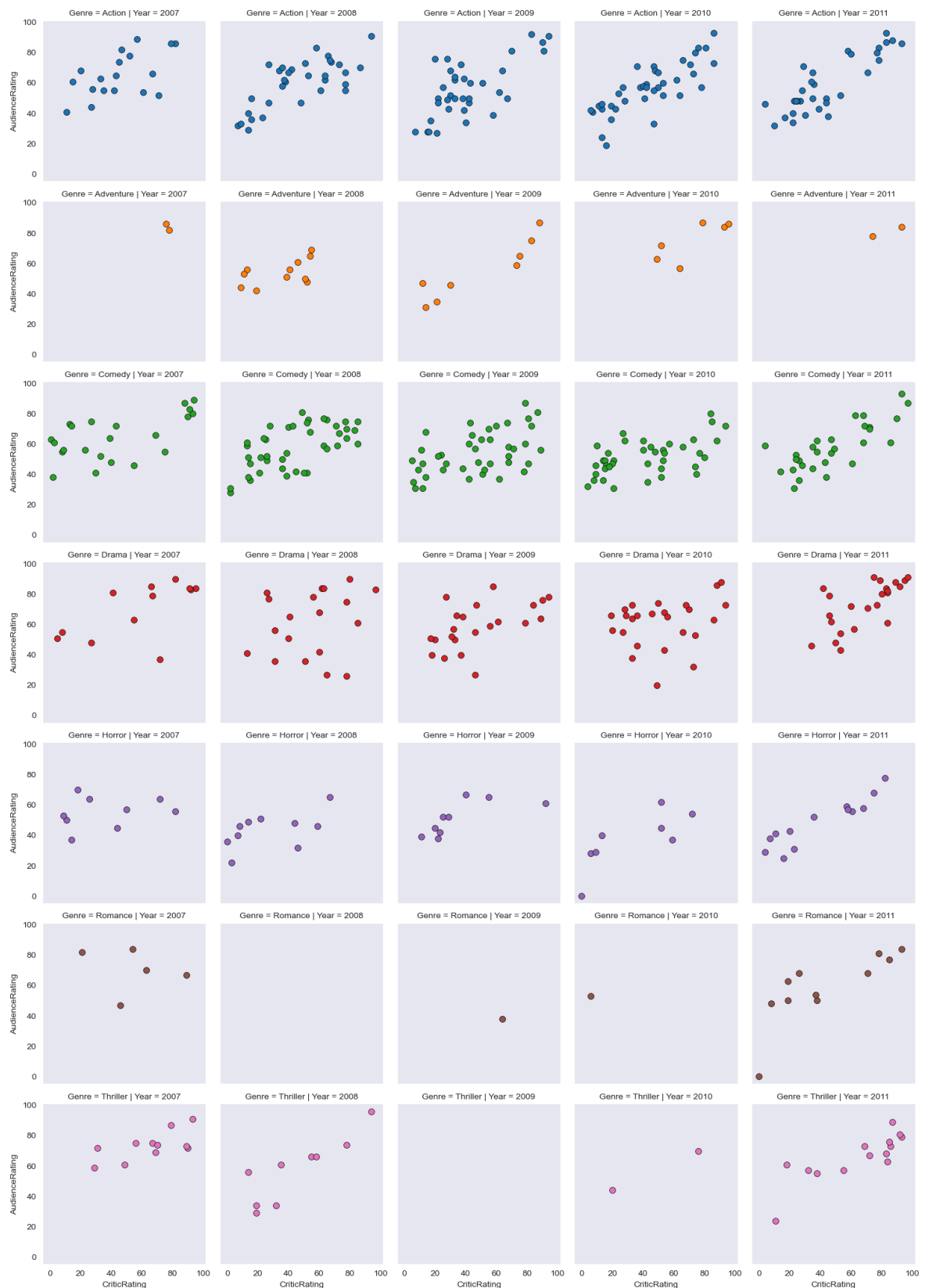
In [214...

```
g = sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
g = g.map(plt.hist, 'BudgetMillions')
plt.show()
```



In [216...

```
g = sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
kws = dict(s=50, linewidth=0.5, edgecolor='black')
g = g.map(plt.scatter, 'CriticRating', 'AudienceRating', **kws )
plt.show()
```



In [220...

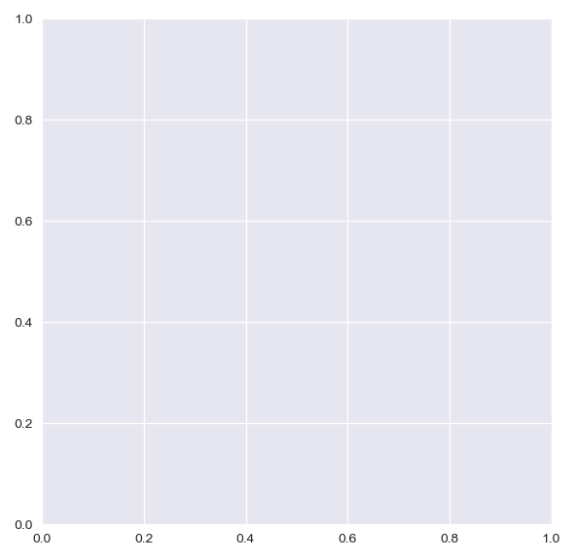
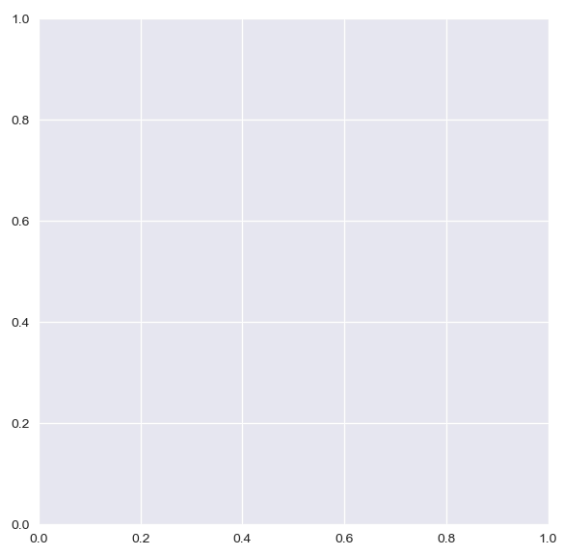
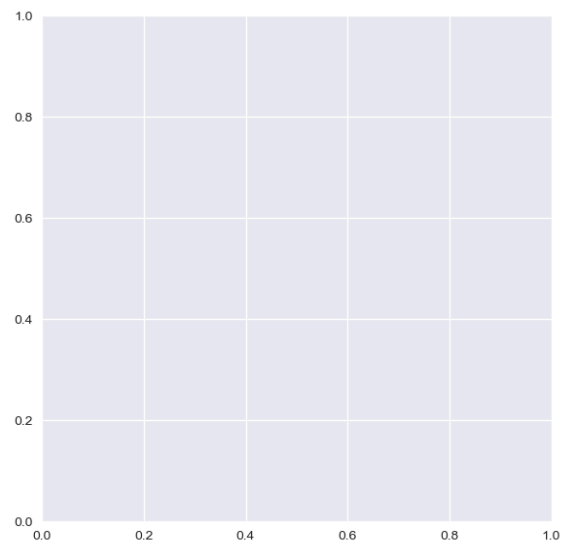
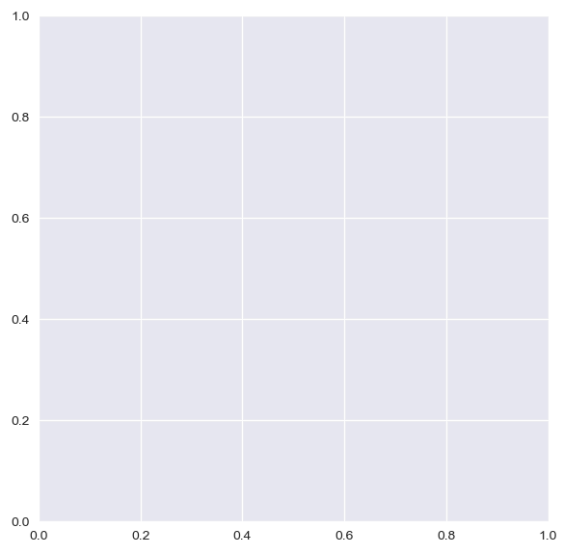
```
sns.set_style('darkgrid')
f, axes = plt.subplots(2,2, figsize = (15,15))

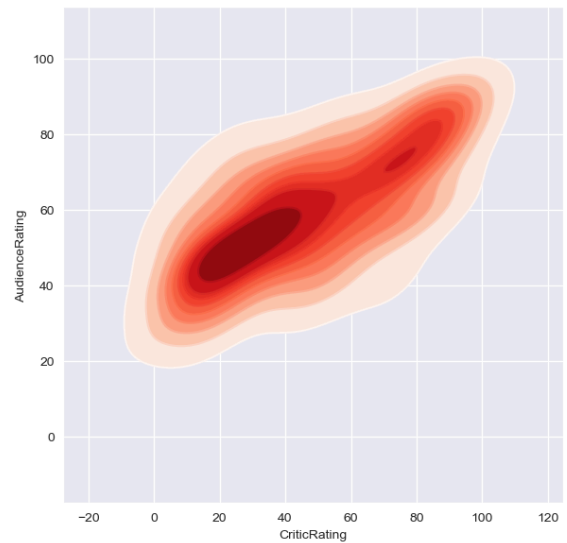
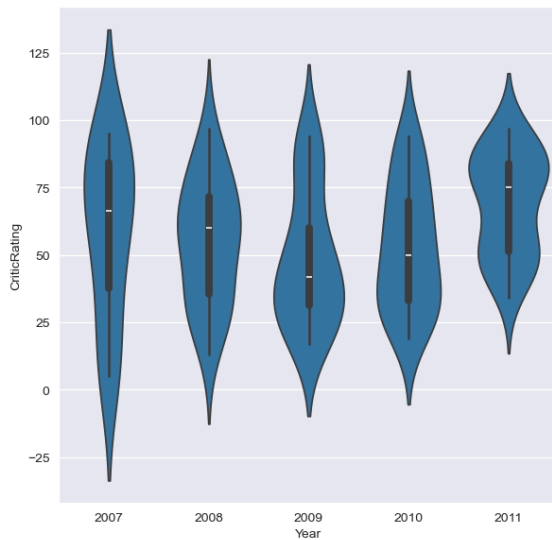
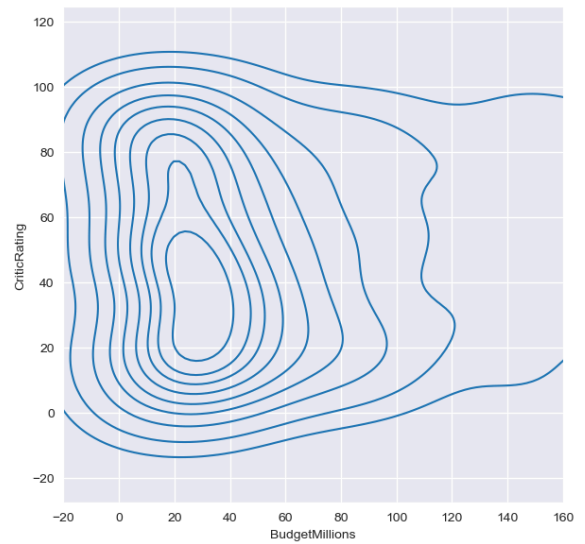
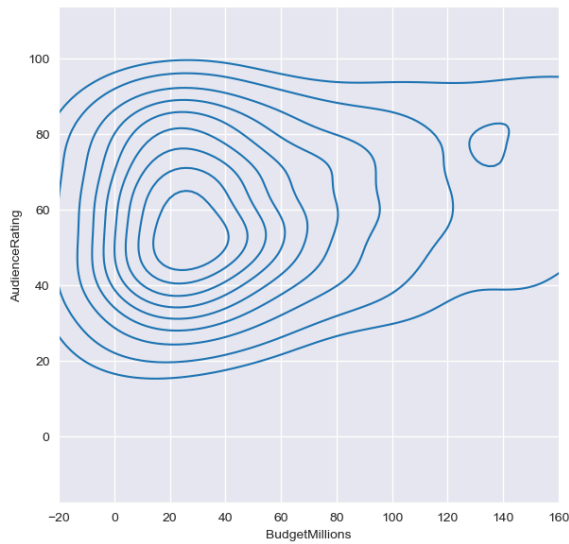
k1 = sns.kdeplot(x=movies.BudgetMillions,y=movies.AudienceRating,ax=axes[0,0])
k2 = sns.kdeplot(x=movies.BudgetMillions,y=movies.CriticRating,ax = axes[0,1])

k1.set(xlim=(-20,160))
k2.set(xlim=(-20,160))

z = sns.violinplot(data=movies[movies.Genre=='Drama'], x='Year', y = 'CriticRating')
```

```
k4 = sns.kdeplot(x=movies.CriticRating,y=movies.AudienceRating,shade = True,shad  
k4b = sns.kdeplot(x=movies.CriticRating, y=movies.AudienceRating,cmap='Reds',ax  
plt.show()
```





In [226...

```
sns.set_style('dark',{'axes.facecolor':'black'})
f, axes = plt.subplots (2,2, figsize = (15,15))

#plot [0,0]
k1 = sns.kdeplot(x=movies.BudgetMillions,y=movies.AudienceRating, \
                 shade = True, shade_lowest=True,\
                 ax = axes[0,0])
k1b = sns.kdeplot(x=movies.BudgetMillions, y=movies.AudienceRating, \
                  cmap = 'cool',ax = axes[0,0])

#plot [0,1]
k2 = sns.kdeplot(x=movies.BudgetMillions,y=movies.CriticRating,\
                 shade=True, shade_lowest=True,\
                 ax = axes[0,1])
k2b = sns.kdeplot(x=movies.BudgetMillions,y=movies.CriticRating,\
                  ax = axes[0,1])

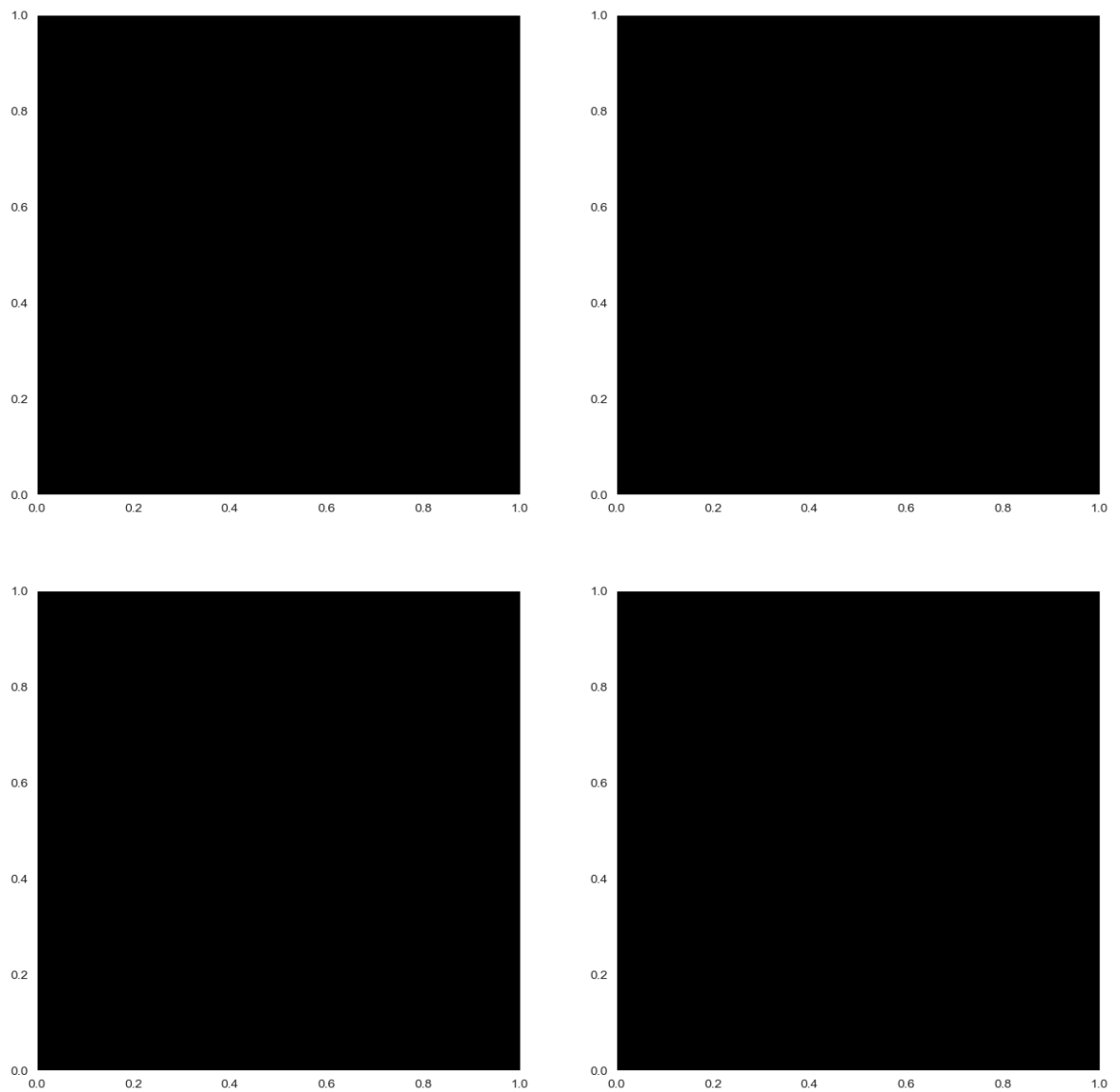
#plot[1,0]
z = sns.violinplot(data=movies[movies.Genre=='Drama'], \
                  x='Year', y = 'CriticRating', ax=axes[1,0])

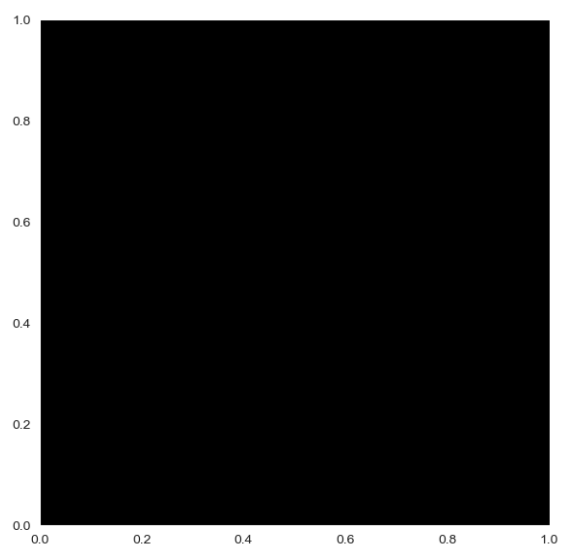
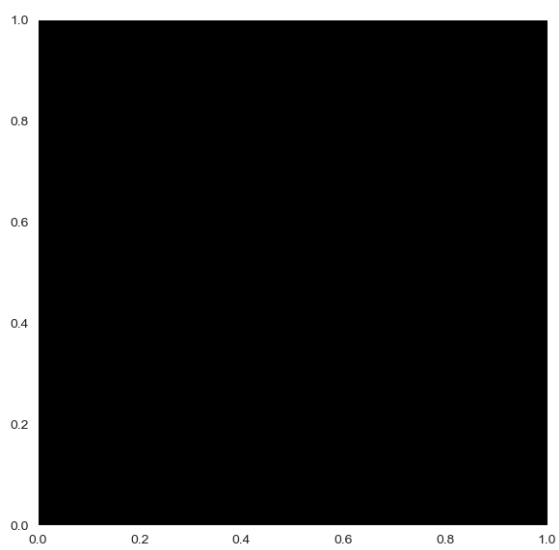
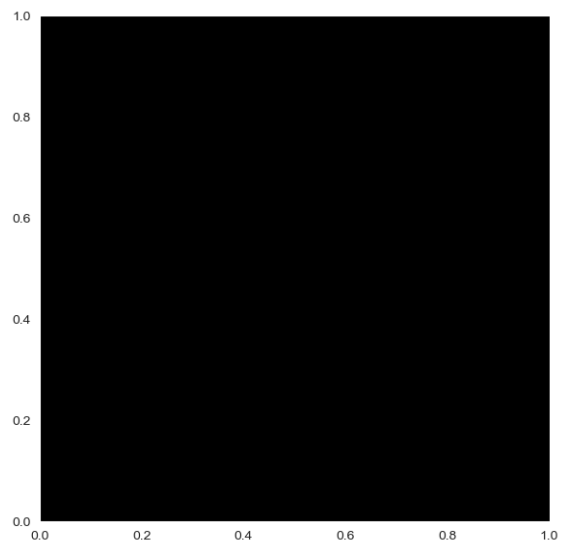
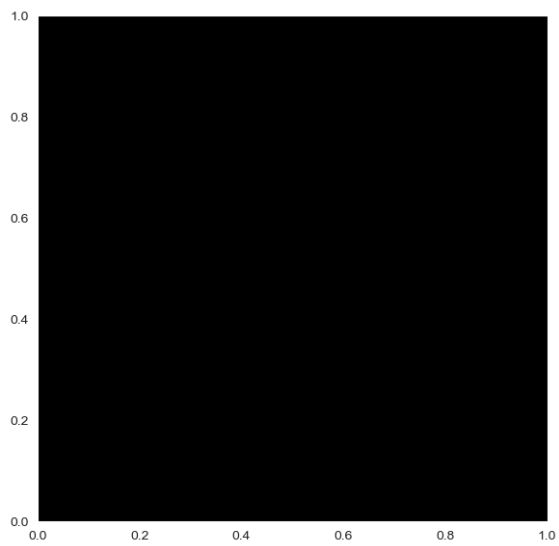
#plot[1,1]
k4 = sns.kdeplot(x=movies.CriticRating,y=movies.AudienceRating, \
                 shade = True,shade_lowest=False,cmap='Blues_r', \
                 ax=axes[1,1])
```

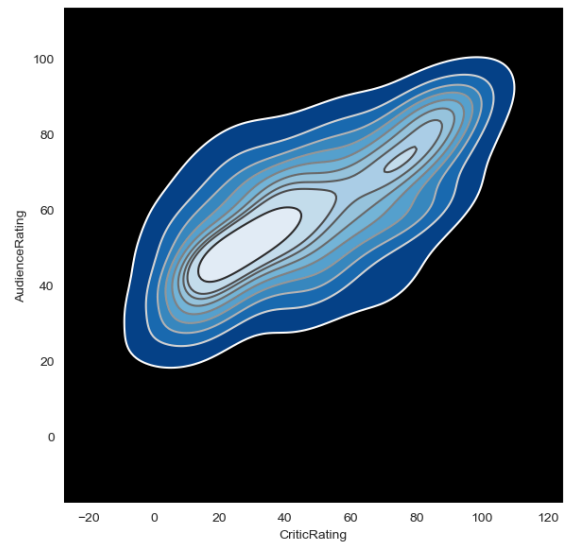
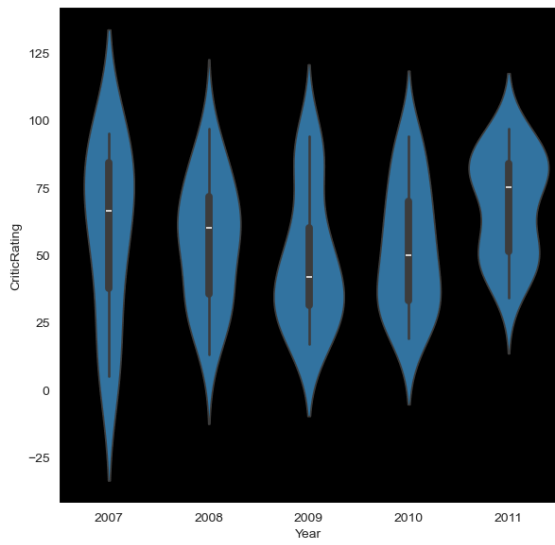
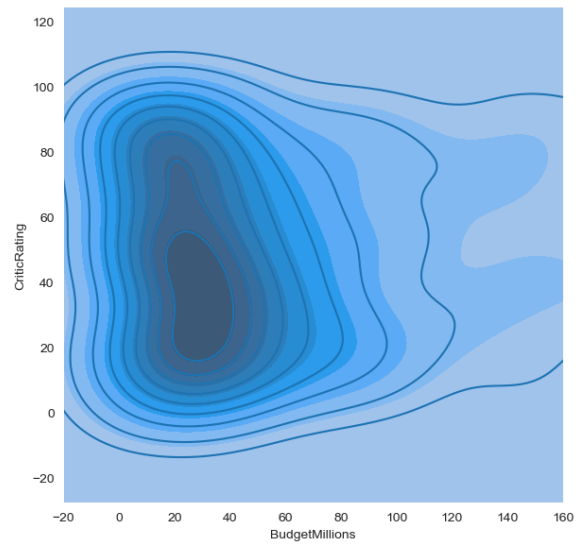
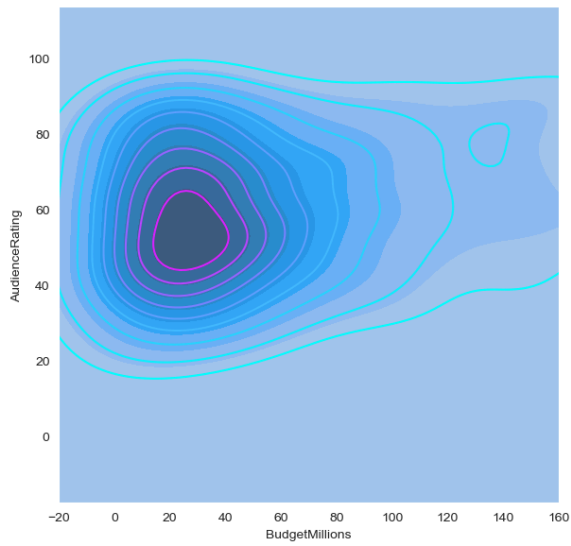
```
k4b = sns.kdeplot(x=movies.CriticRating, y=movies.AudienceRating, \
                  cmap='gist_gray_r', ax = axes[1,1])

k1.set(xlim=(-20,160))
k2.set(xlim=(-20,160))

plt.show()
```







In []: