Movie Rating Analysis(Advanced Visualization)

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

In [2]:
```
os.getcwd()
```

Out[2]:
```
'C:\\Users\\hp'
```

In [3]:
```
movies=pd.read_excel(r"C:\Users\hp\Downloads\Movie-Rating1.xlsx")
```

In [4]:
```
movies
```

Out[4]:

	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 [5]:
```
len(movies)
```

Out[5]:
```
559
```

In [6]:
```
movies.columns
```

Out[6]:
```
Index(['Film', 'Genre', 'Rotten Tomatoes Ratings %', 'Audience Ratings %',
       'Budget (million $)', 'Year of release'],
      dtype='object')
```

In [7]:
```
movies.columns=['Film','Genre','CriticRating','AudienceRating','BudgetMillions','Year']
```

In [8]:
```
movies.head()
```

Out[8]:

	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 [9]:
```
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 [10]:
```
movies.describe()
```

Out[10]:

	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 [11]:
```
movies['Film']
```

Out[11]:
```
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 [12]:
```
movies.Film
```

Out[12]:
```
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 [13]:
```
movies
```

Out[13]:

	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
...	...	...	...	...	...	...
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 [14]:
```
movies.Film=movies.Film.astype("category")
movies.Genre=movies.Genre.astype("category")
movies.Year=movies.Year.astype("category")
```

In [15]:
```
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 [16]:
```
movies.Year
```

Out[16]:
```
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 [17]:
```
movies.describe()
```

Out[17]:

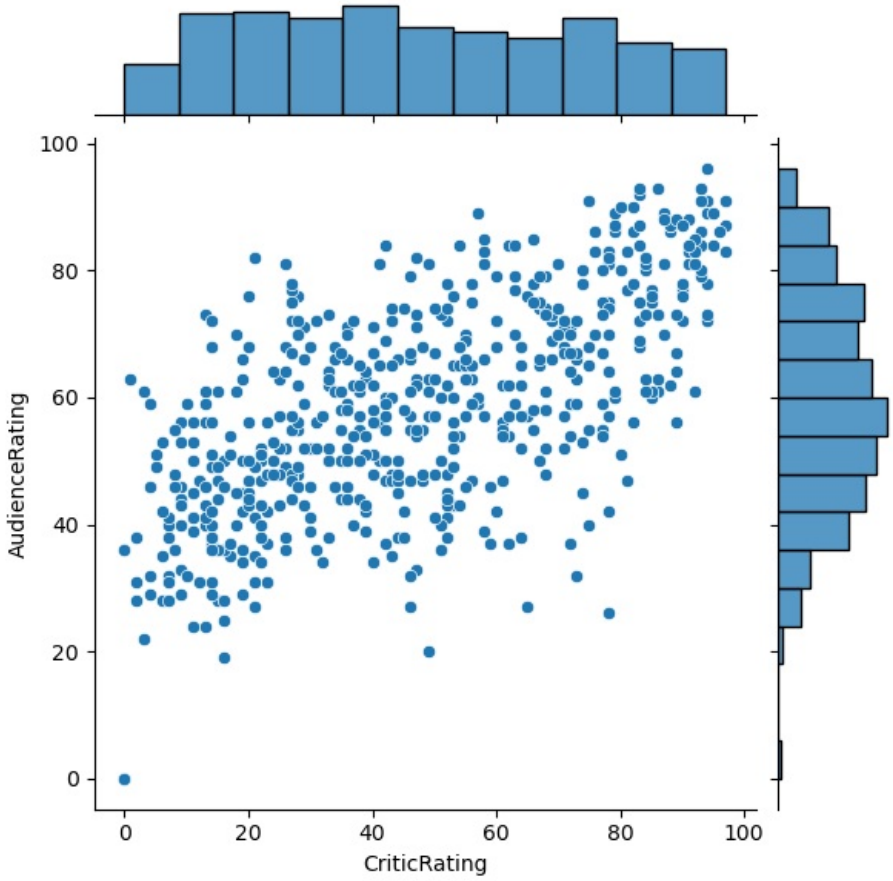
	CriticRating	AudienceRating	BudgetMillions
count	559.000000	559.000000	559.000000
mean	47.309481	58.744186	50.236136
std	26.413091	16.826887	48.731817
min	0.000000	0.000000	0.000000
25%	25.000000	47.000000	20.000000
50%	46.000000	58.000000	35.000000
75%	70.000000	72.000000	65.000000
max	97.000000	96.000000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

In [18]:
```
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

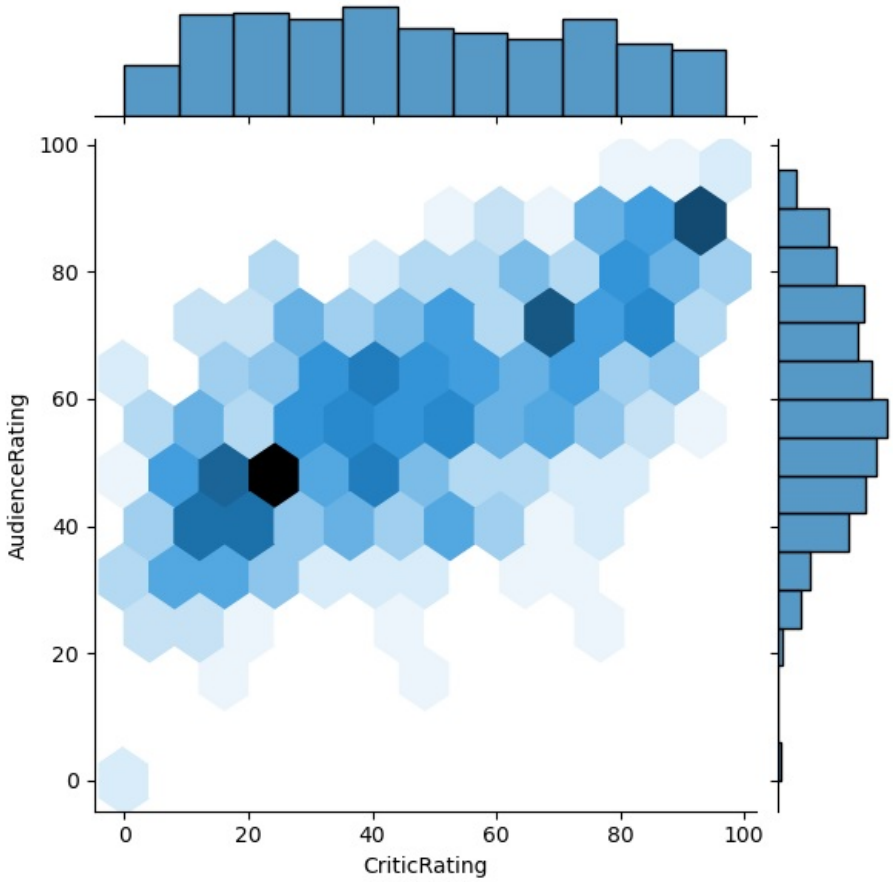
In [19]:
```
import warnings
warnings.filterwarnings('ignore')
```

JOINTPLOT--- basically joint plot is a scatter plot & it find the relation b/w audiene & critics * also if you look up you can find the uniform distribution (critics)and normal distriution (audience)

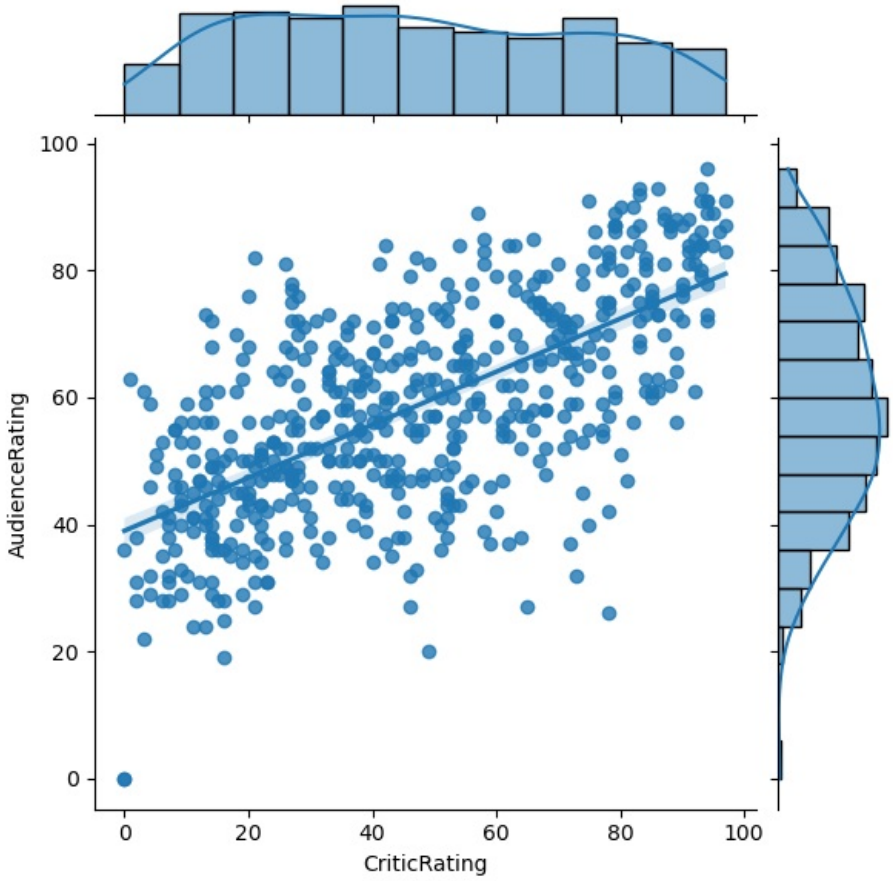
In [20]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating')
#AudienceRating is more dominant than criticRating
# Based on this we find out as most people are most liklihood to watch audience rating & less likely to wathc c
# let me explain the excel - if you filter audience rating & critic rating. critic rating has very low values c
```



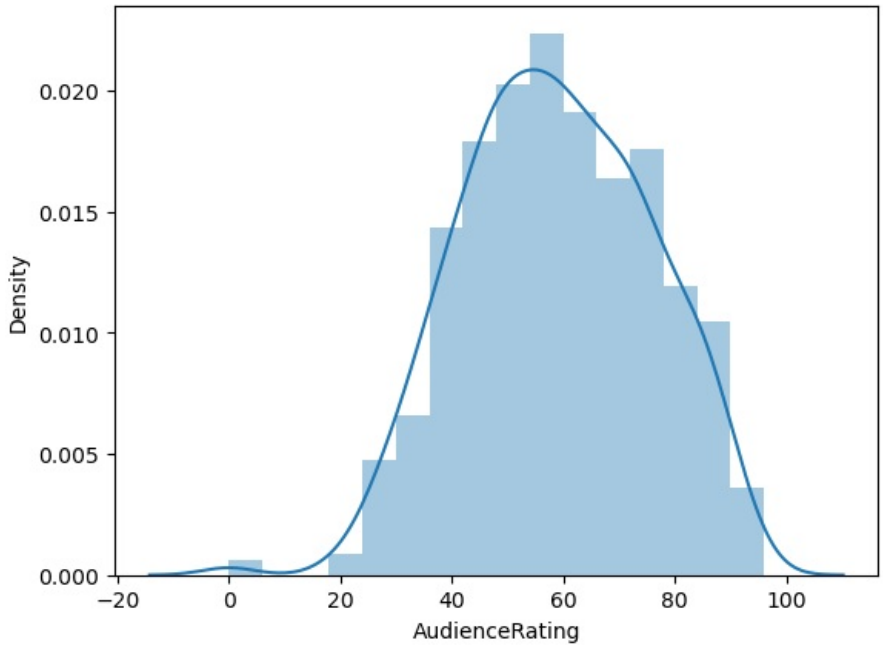
In [21]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='hex')
```



In [23]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='reg')
```

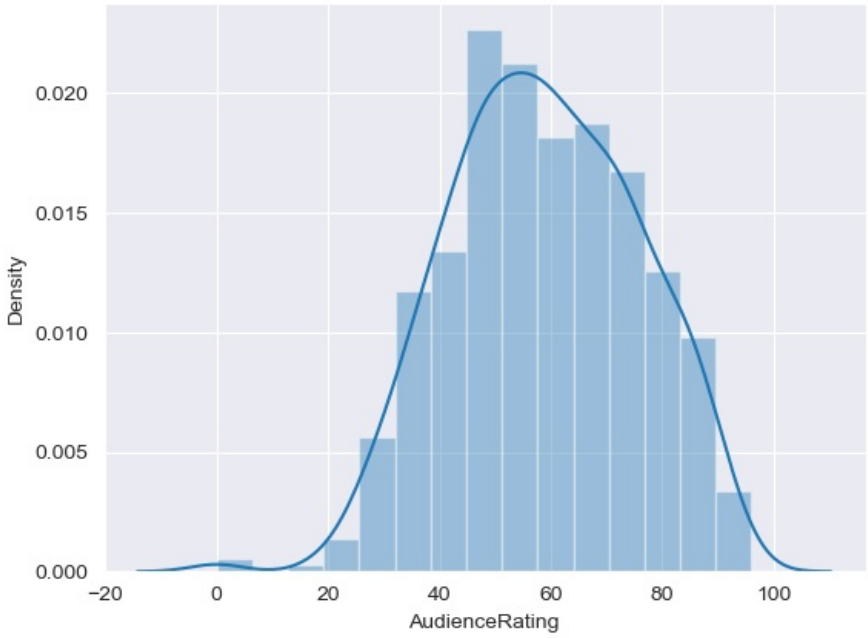


In [24]:
```
#Histogram
m1=sns.distplot(movies.AudienceRating)
```

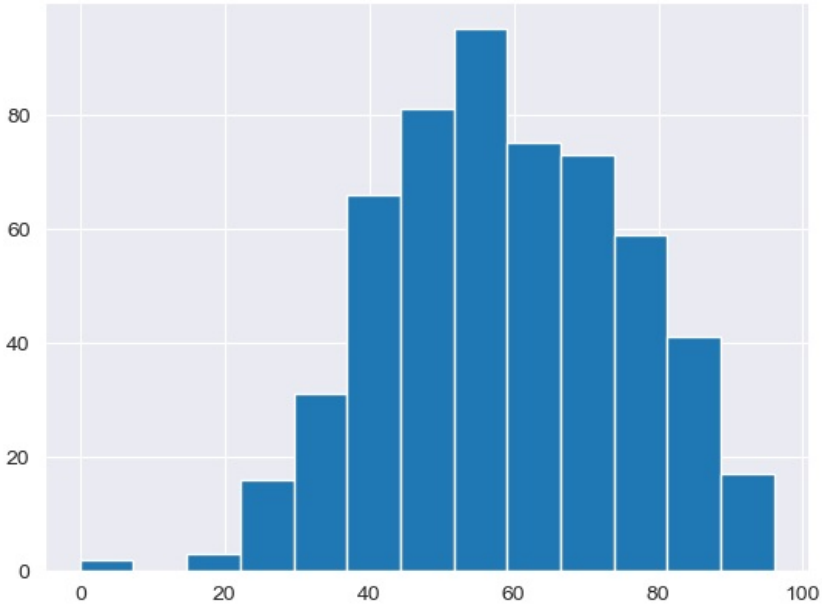


In [25]:
```
sns.set_style('darkgrid')
```

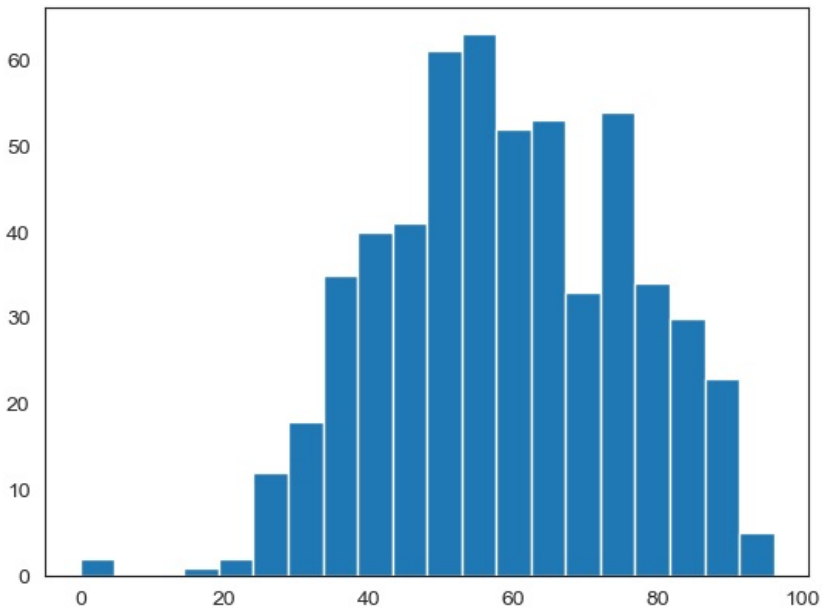
In [26]:
```
m2=sns.distplot(movies.AudienceRating,bins=15)
```



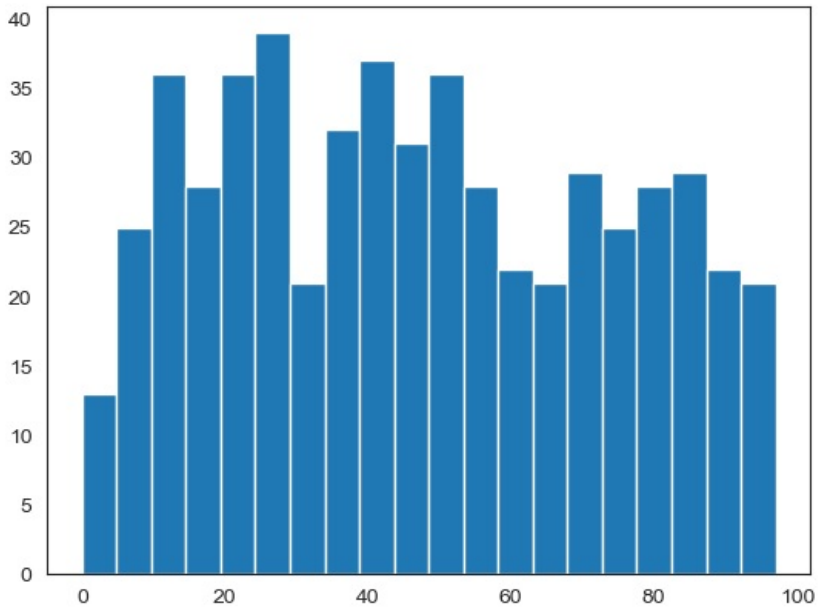
In [27]:
```
#HistogramPlot
n1 = plt.hist(movies.AudienceRating, bins=13)
```



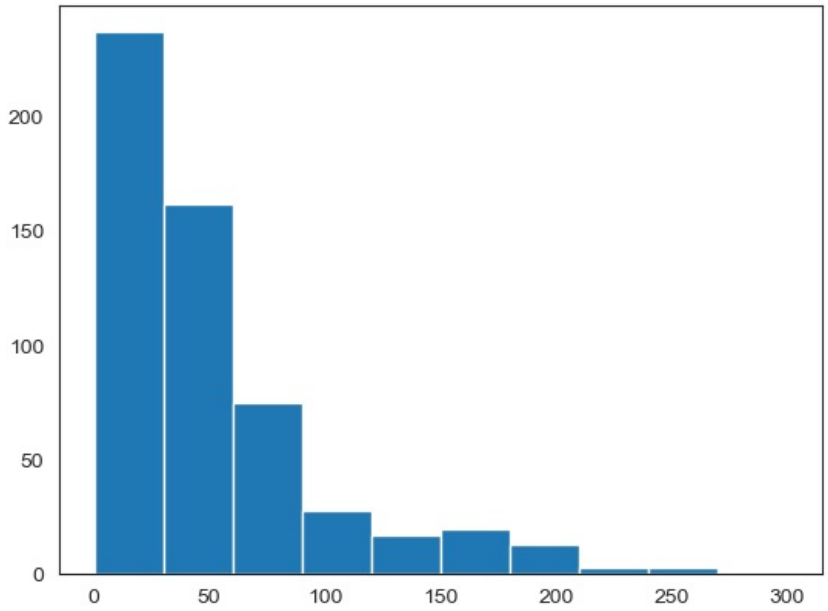
In [28]:
```
sns.set_style('white') #normal distribution / bell curve
n1 = plt.hist(movies.AudienceRating, bins=20)
```



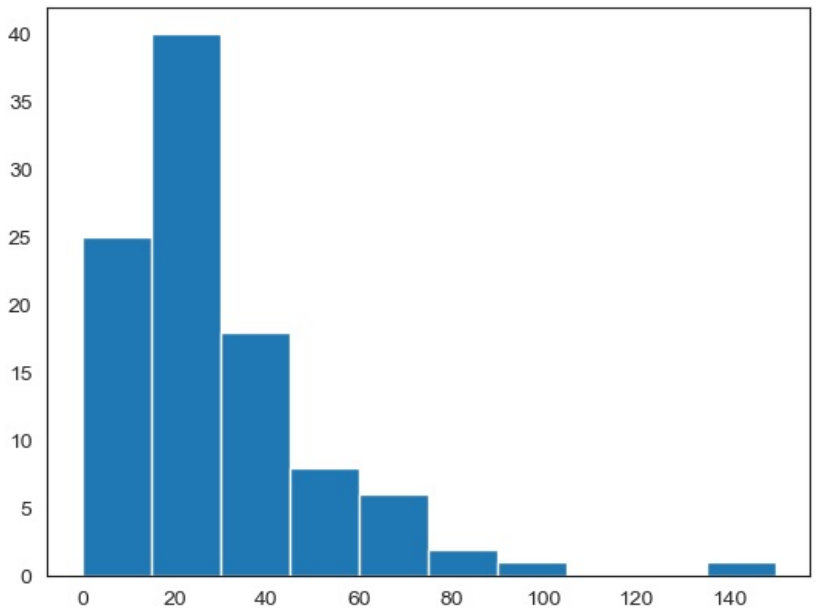
In [29]:
```
n1 = plt.hist(movies.CriticRating, bins=20)
```



In [30]:
```
n2=plt.hist(movies.BudgetMillions)
```



In [31]:
```
plt.hist(movies[movies.Genre=='Drama'].BudgetMillions)
plt.show()
```

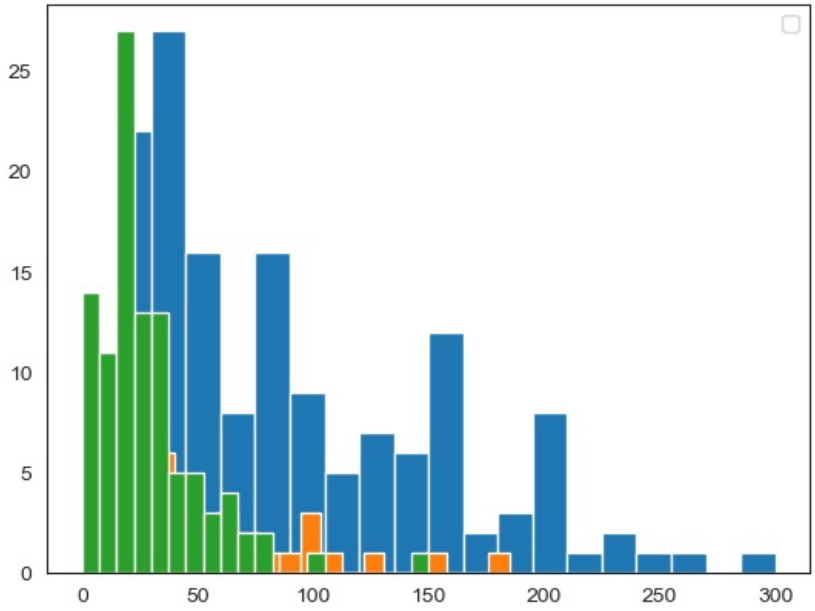


In [32]:
```
movies.Genre.unique()
```

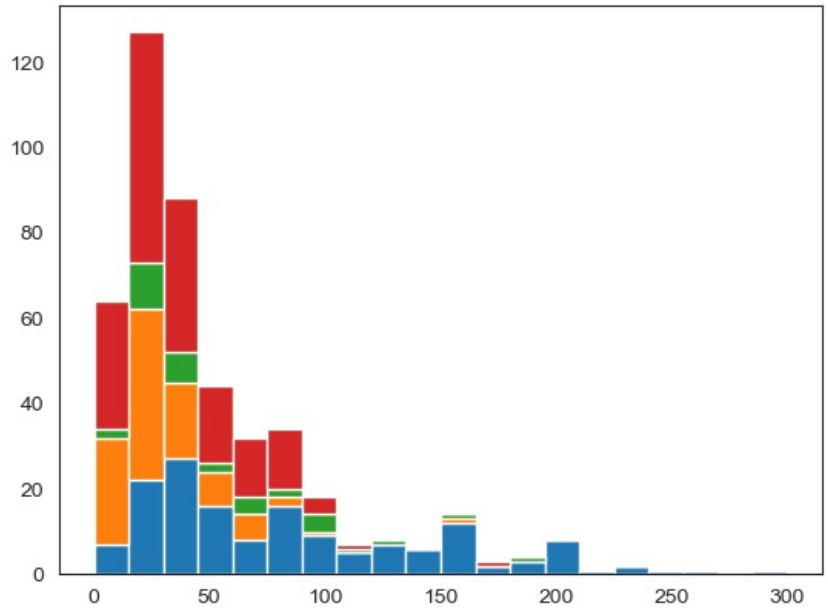
Out[32]:
```
['Comedy', 'Adventure', 'Action', 'Horror', 'Drama', 'Romance', 'Thriller']
Categories (7, object): ['Action', 'Adventure', 'Comedy', 'Drama', 'Horror', 'Romance', 'Thriller']
```

In [33]:
```
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()
```

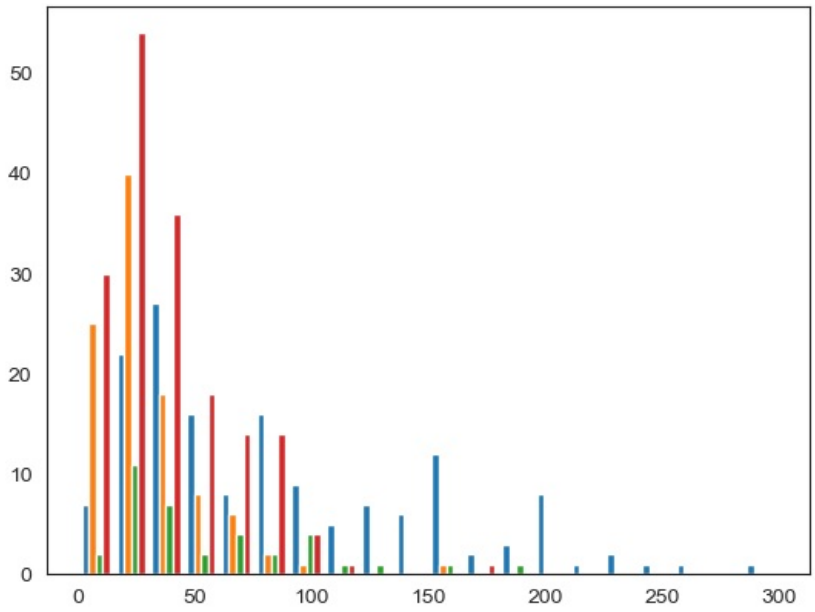
```
No artists with labels found to put in legend.  Note that artists whose label start with an underscore are igno
red when legend() is called with no argument.
```



In [34]:
```
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)
plt.show()
```



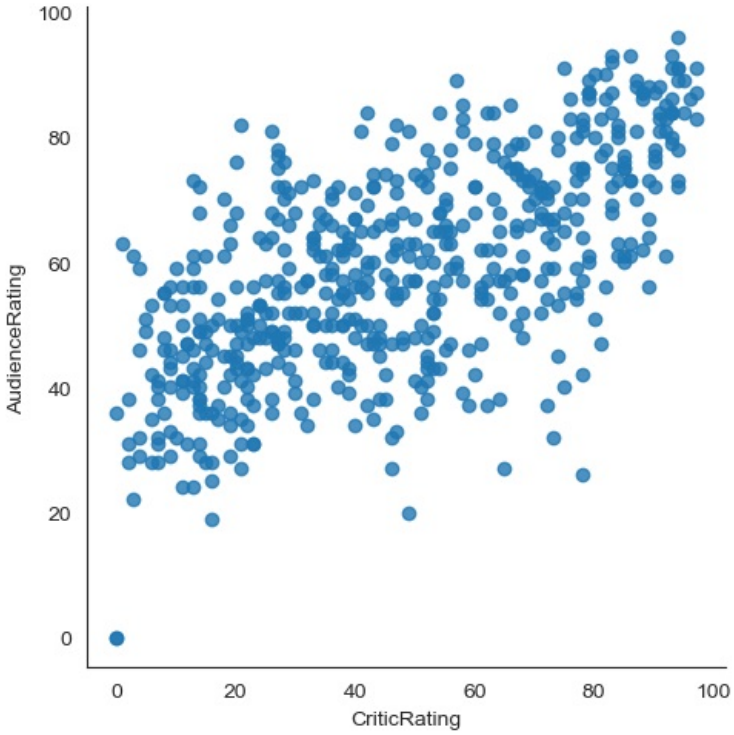
In [35]:
```
plt.hist([movies[movies.Genre=='Action'].BudgetMillions,\
          movies[movies.Genre == 'Drama'].BudgetMillions, \
          movies[movies.Genre == 'Thriller'].BudgetMillions, \
          movies[movies.Genre == 'Comedy'].BudgetMillions],
         bins=20,stacked=False)
plt.show()
```



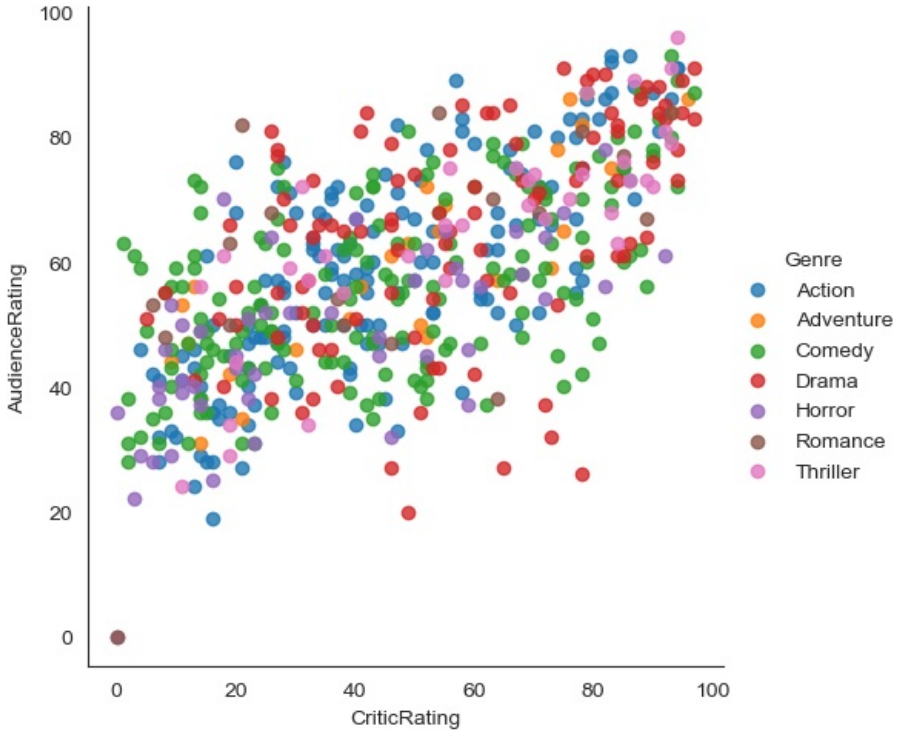
In [36]:
```
for gen in movies.Genre.cat.categories:
    print(gen)
```

```
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller
```

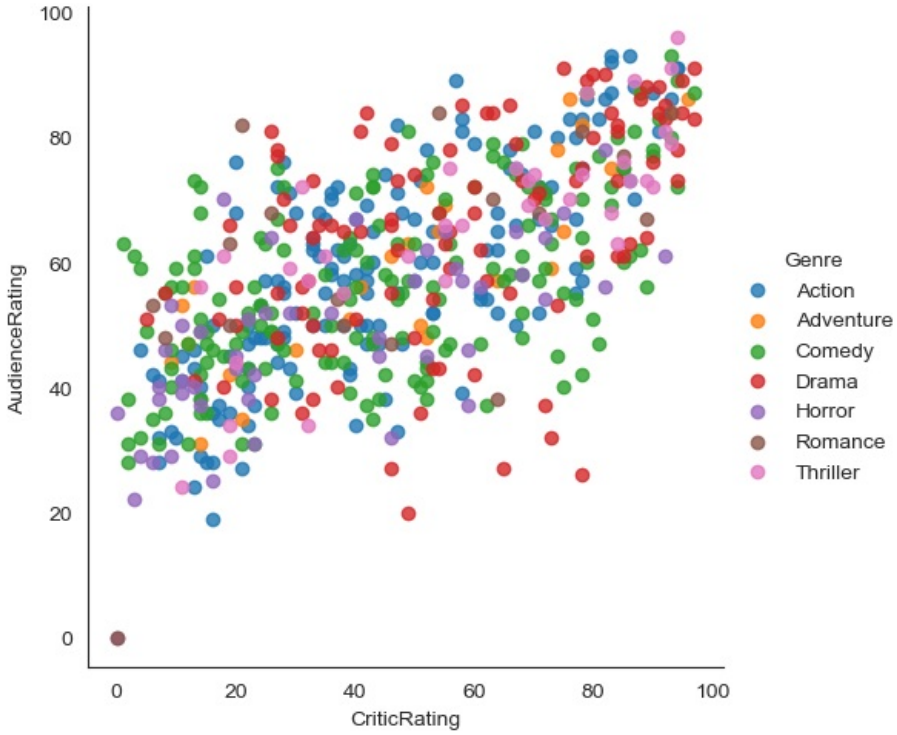
In [37]:
```
vis1=sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',fit_reg=False)
```



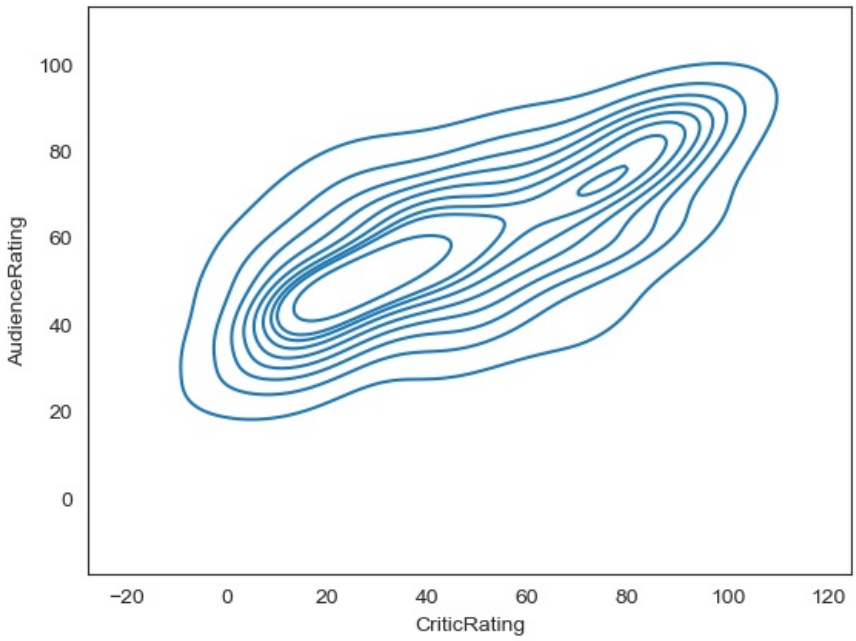
In [38]:
```
vis1=sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',fit_reg=False,hue='Genre')
```



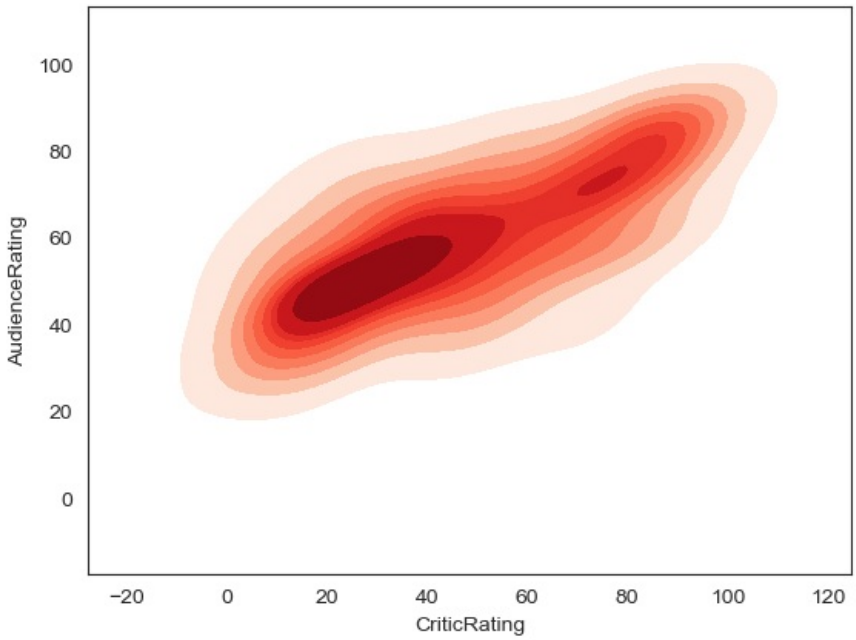
In [39]:
```
vis1=sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',fit_reg=False,hue='Genre',aspect=1)
```



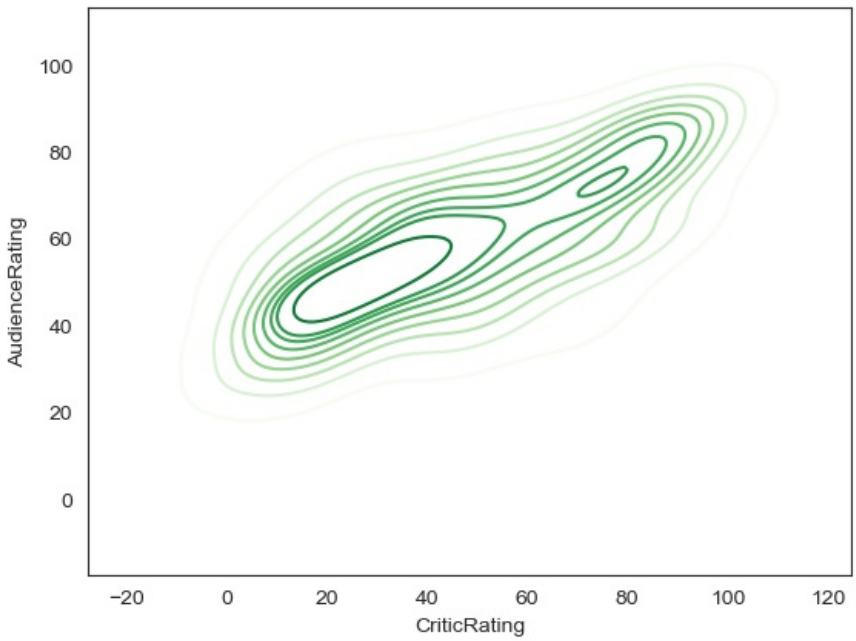
In [40]:
```
k1=sns.kdeplot(x='CriticRating', y='AudienceRating',data=movies)
```



In [45]:
```
k1=sns.kdeplot(x='CriticRating', y='AudienceRating',data=movies,cmap='Reds',shade=True,shade_lowest=False)
```

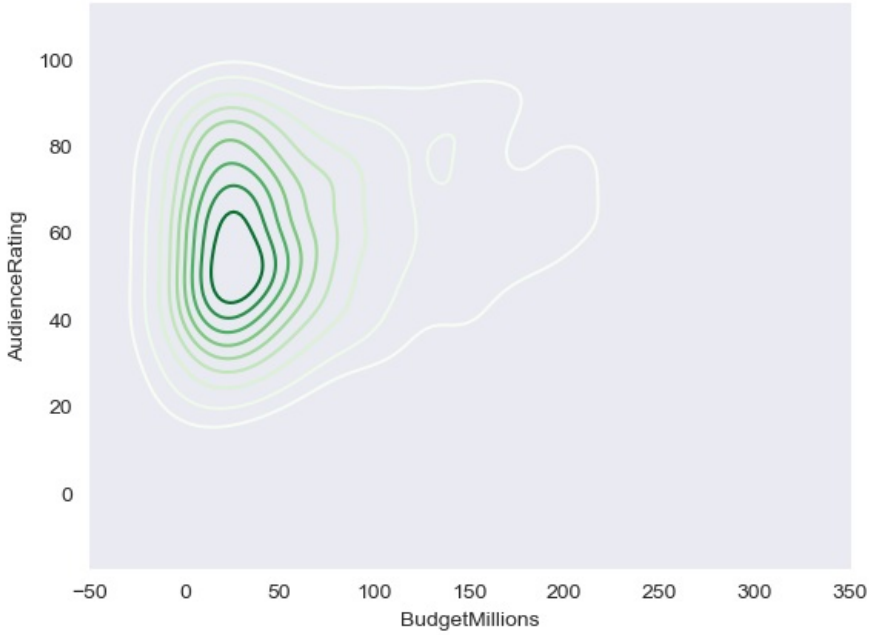


In [49]:
```
k2 = sns.kdeplot(x='CriticRating', y='AudienceRating', data=movies, cmap='Greens', shade_lowest=False)
```

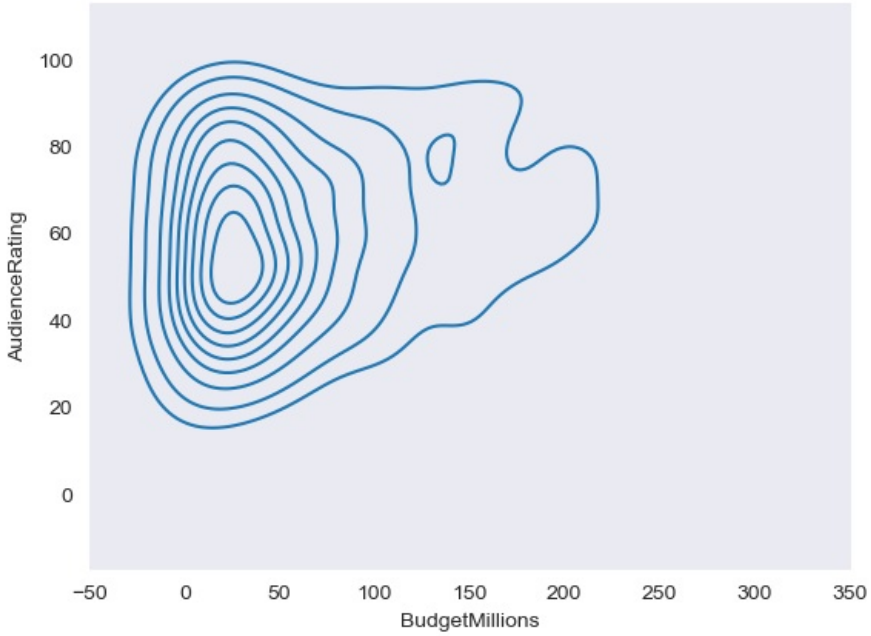


In [50]:
```
sns.set_style('dark')
```

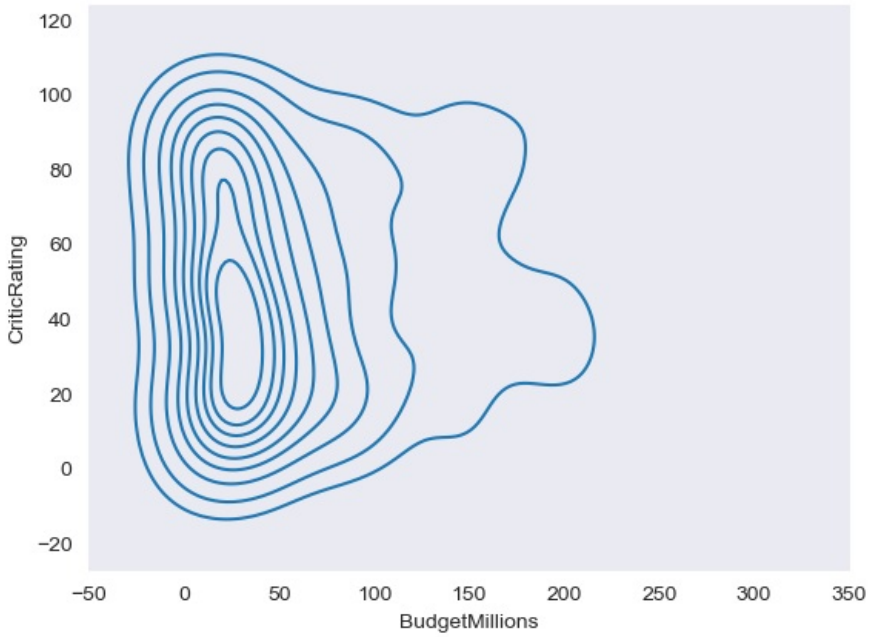
```
k1=k2 = sns.kdeplot(x='BudgetMillions', y='AudienceRating', data=movies, cmap='Greens', shade_lowest=False)
```



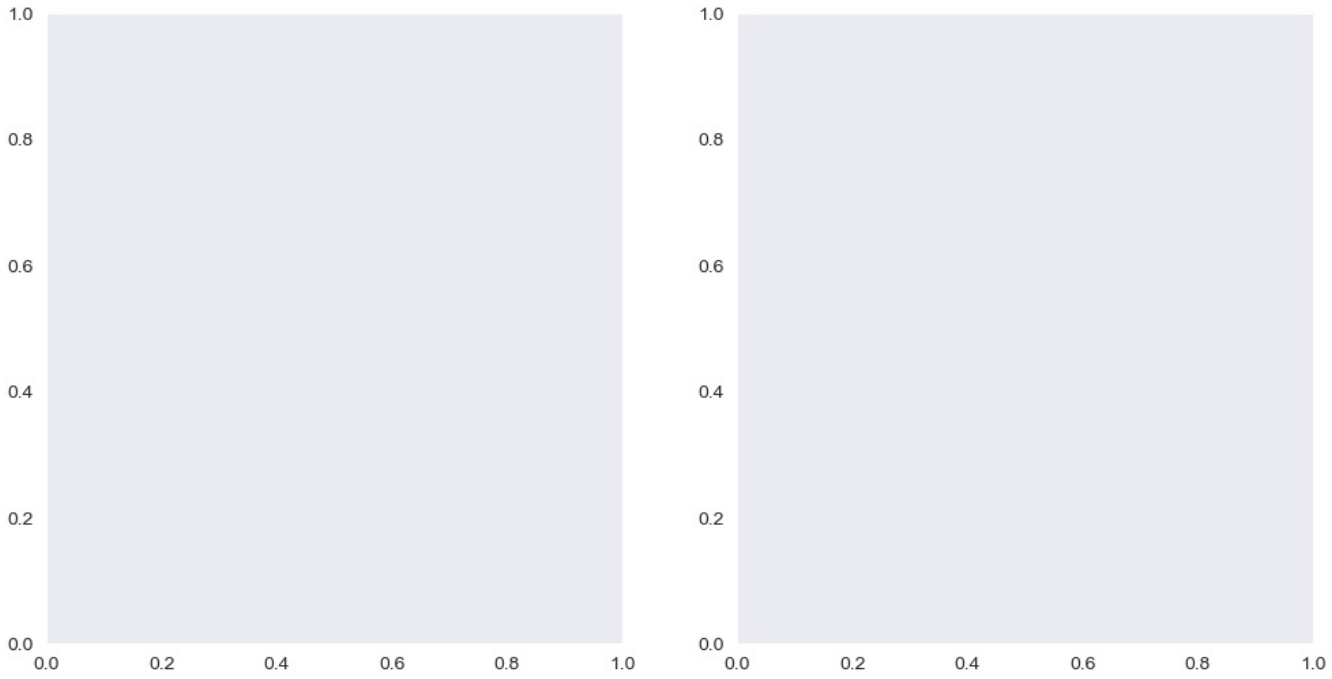
In [53]:
```
sns.set_style('dark')
k2= sns.kdeplot(x='BudgetMillions', y='AudienceRating', data=movies)
```



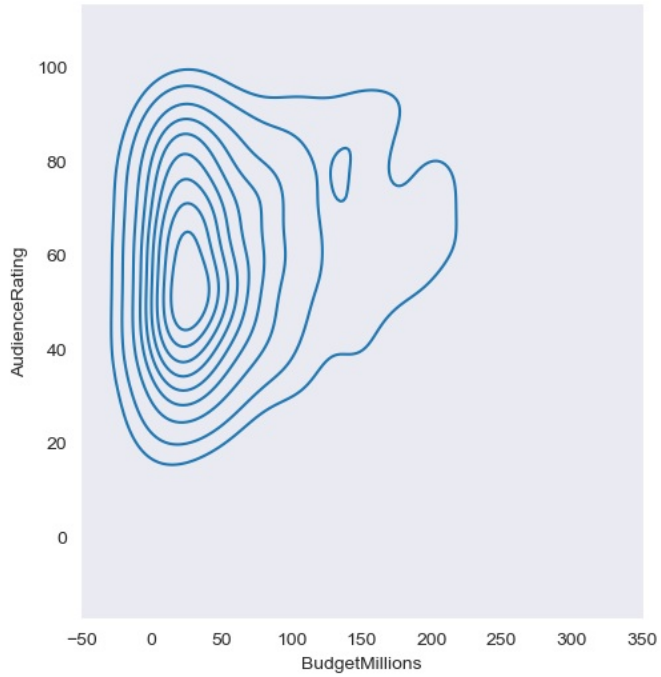
In [56]:
```
k2 = sns.kdeplot( x='BudgetMillions',y='CriticRating', data=movies)
```

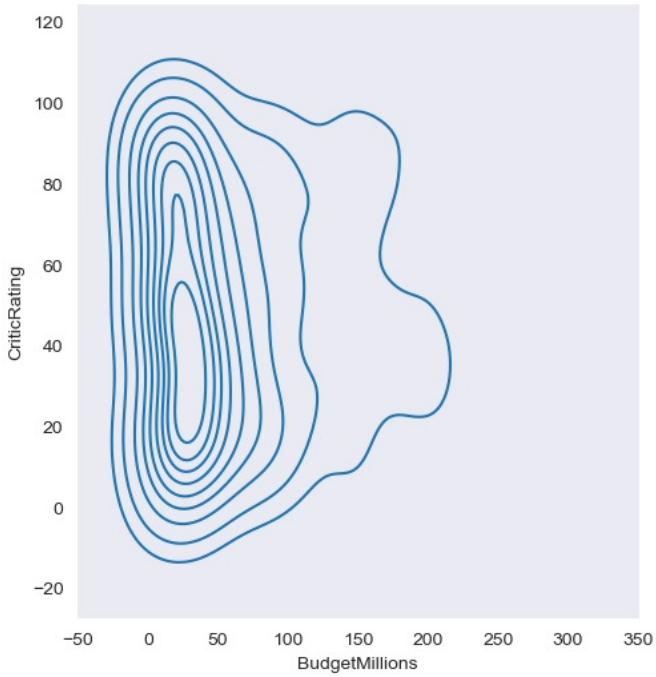


In [62]:
```
#subplots
f,ax=plt.subplots(1,2 ,figsize=(12,6))
```



In [66]:
```
f,axes=plt.subplots(1,2 ,figsize=(12,6))
k1=sns.kdeplot(x='BudgetMillions',y='AudienceRating',data=movies,ax=axes[0])
k1=sns.kdeplot(x='BudgetMillions',y='CriticRating',data=movies,ax=axes[1])
```



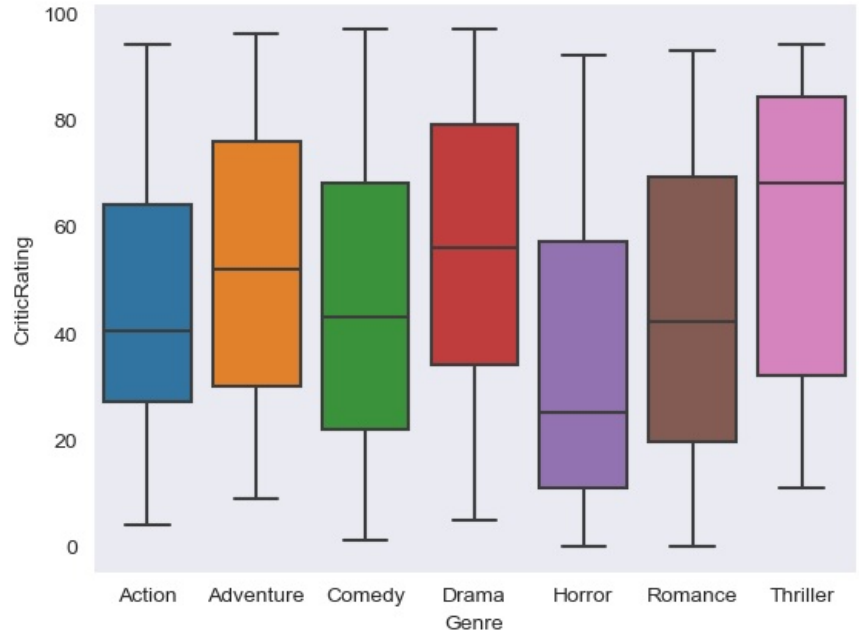


In [67]:
```
axes
```

Out[67]:
```
array([<Axes: xlabel='BudgetMillions', ylabel='AudienceRating'>,
       <Axes: xlabel='BudgetMillions', ylabel='CriticRating'>],
      dtype=object)
```

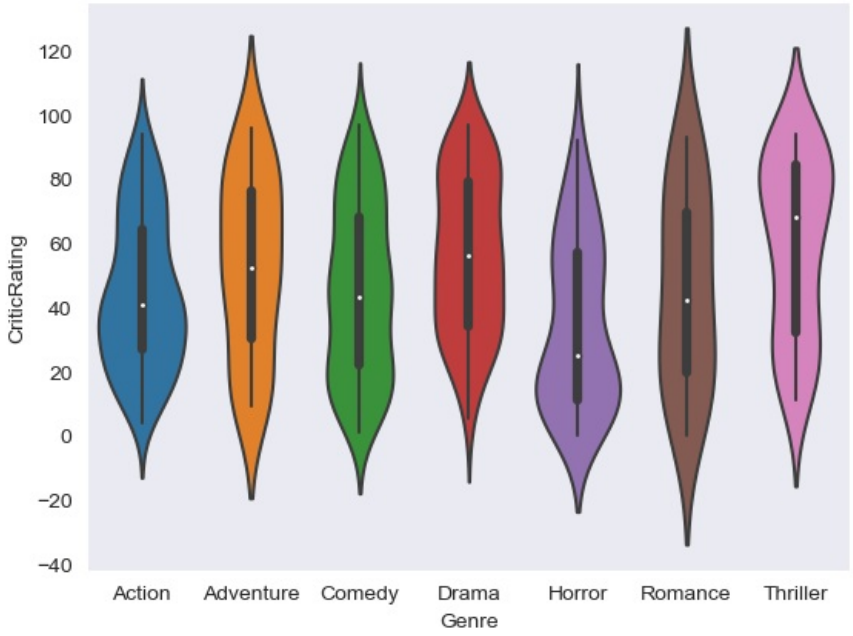
In [ ]:
```
#Boxplot
```

In [69]:
```
w=sns.boxplot(x='Genre',y='CriticRating',data=movies)
```

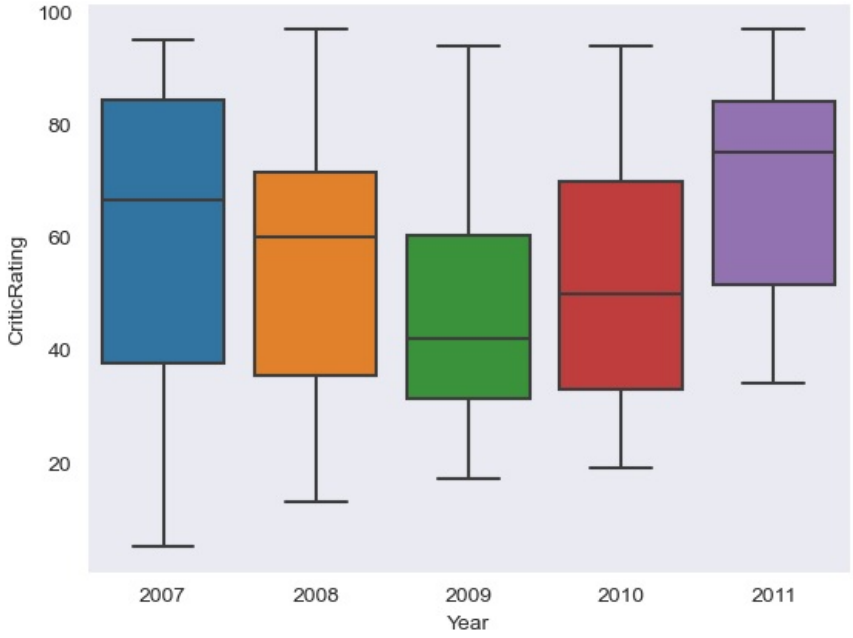


In [ ]:
```
#Violin Plot
```

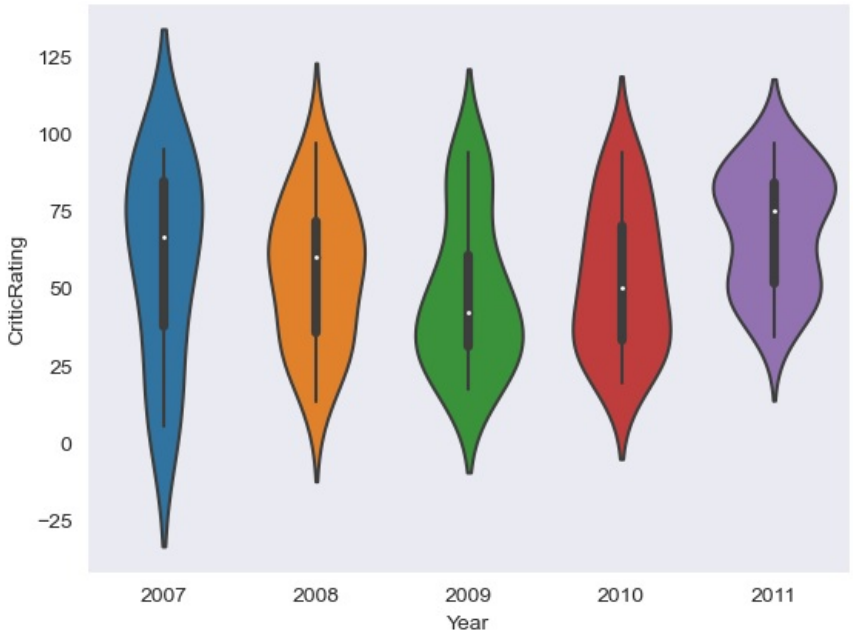
In [70]:
```
z=sns.violinplot(x='Genre',y='CriticRating',data=movies)
```



In [71]:
```
w1 = sns.boxplot(data=movies[movies.Genre == 'Drama'], x='Year', y = 'CriticRating')
```

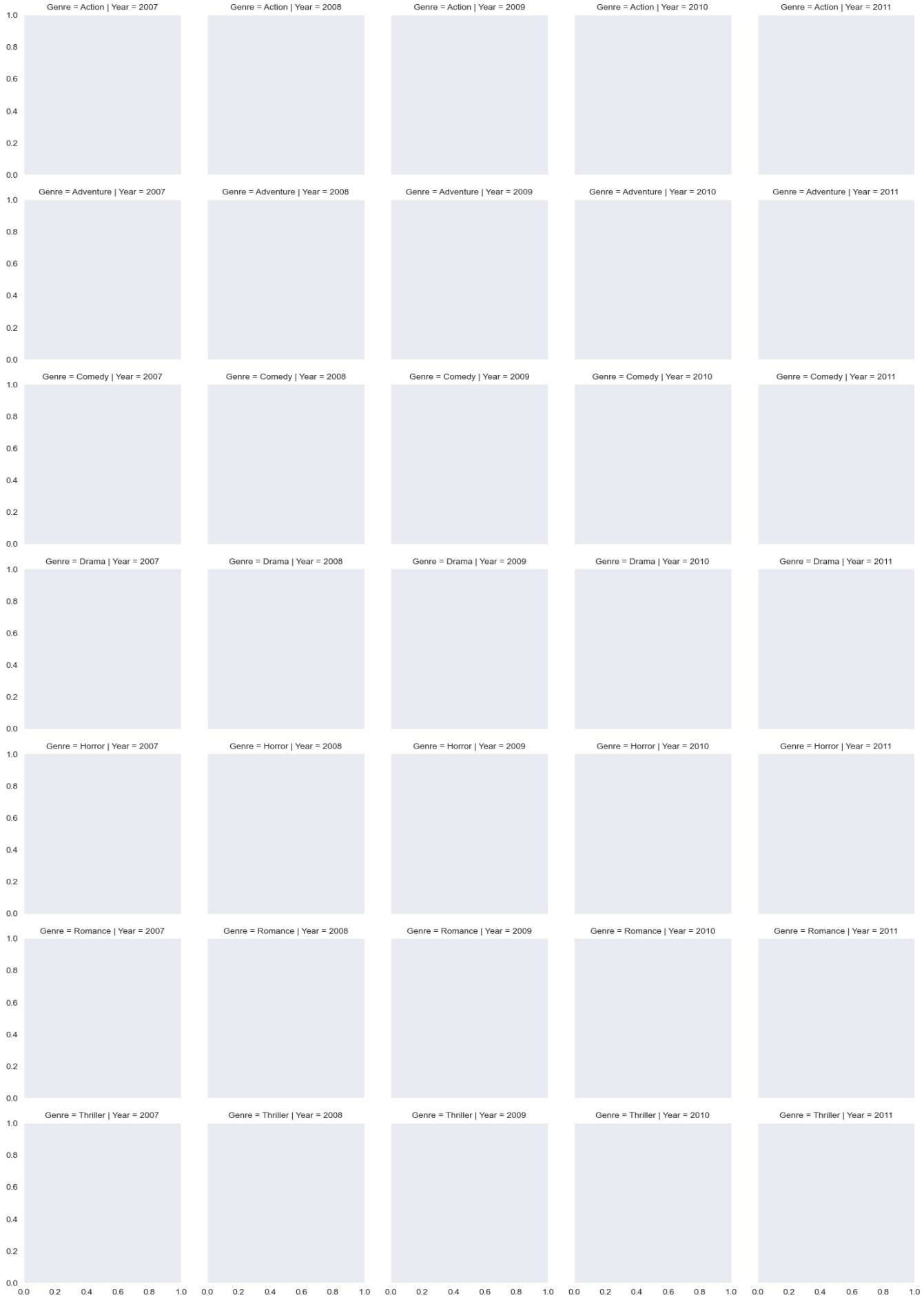


In [73]:
```
z=sns.violinplot(data=movies[movies.Genre=='Drama'], x='Year', y = 'CriticRating')
```



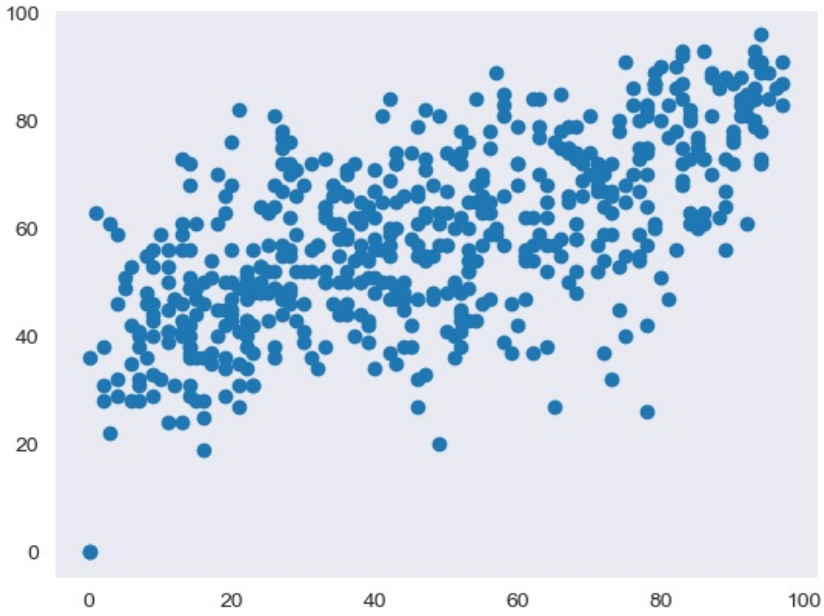
In [ ]:
```
#creating a facet grid
```

In [74]:
```
g=sns.FacetGrid(movies, row='Genre',col='Year',hue='Genre')
```

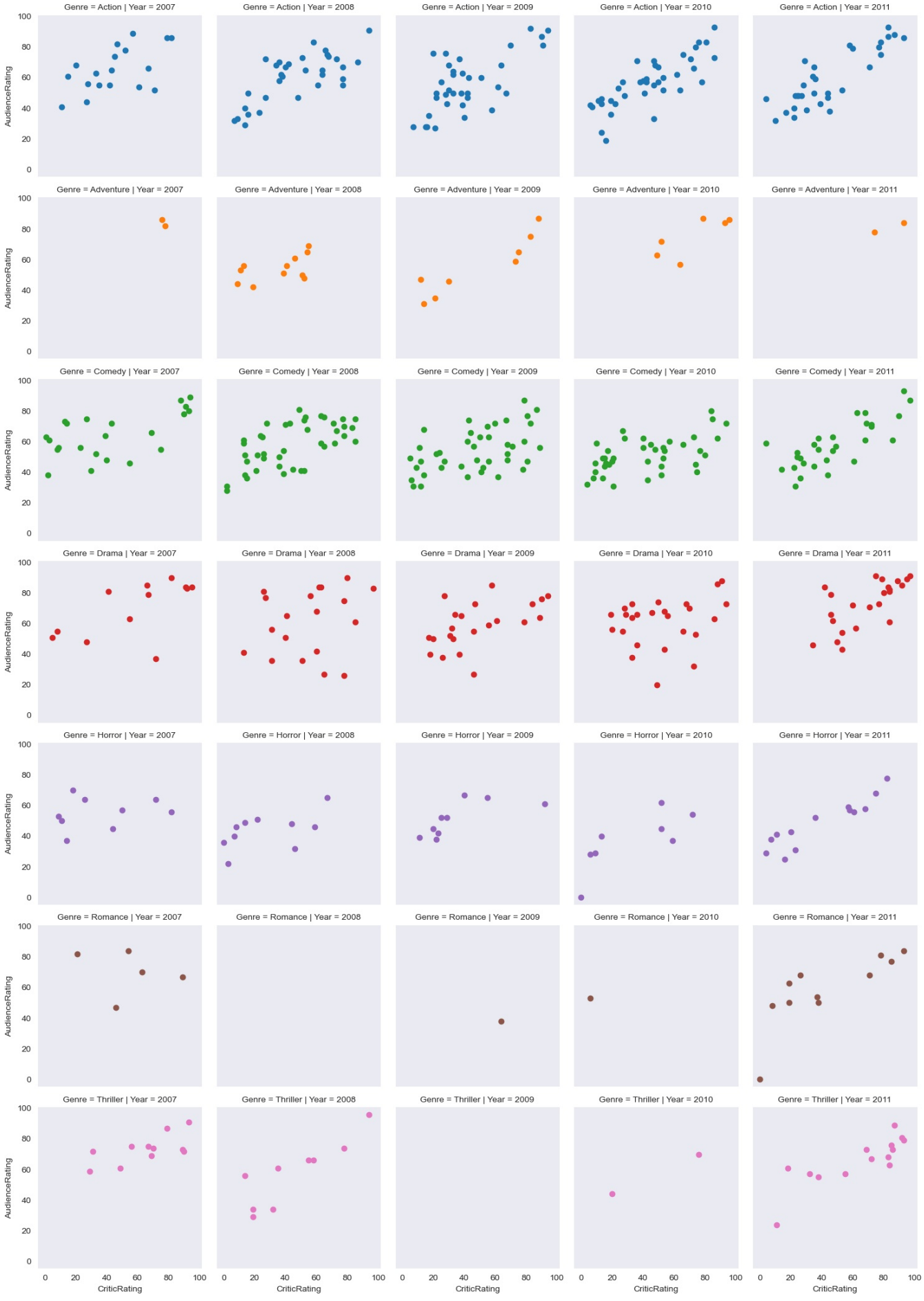


In [75]:
```
plt.scatter(x='CriticRating',y='AudienceRating',data=movies)
```

Out[75]:
```
<matplotlib.collections.PathCollection at 0x2bc812a4690>
```



In [76]:
```
plt.scatter(x='CriticRating',y='AudienceRating',data=movies)
g=g.map(plt.scatter,'CriticRating', 'AudienceRating')
#Scatterplots are mapped in facetgrid
```

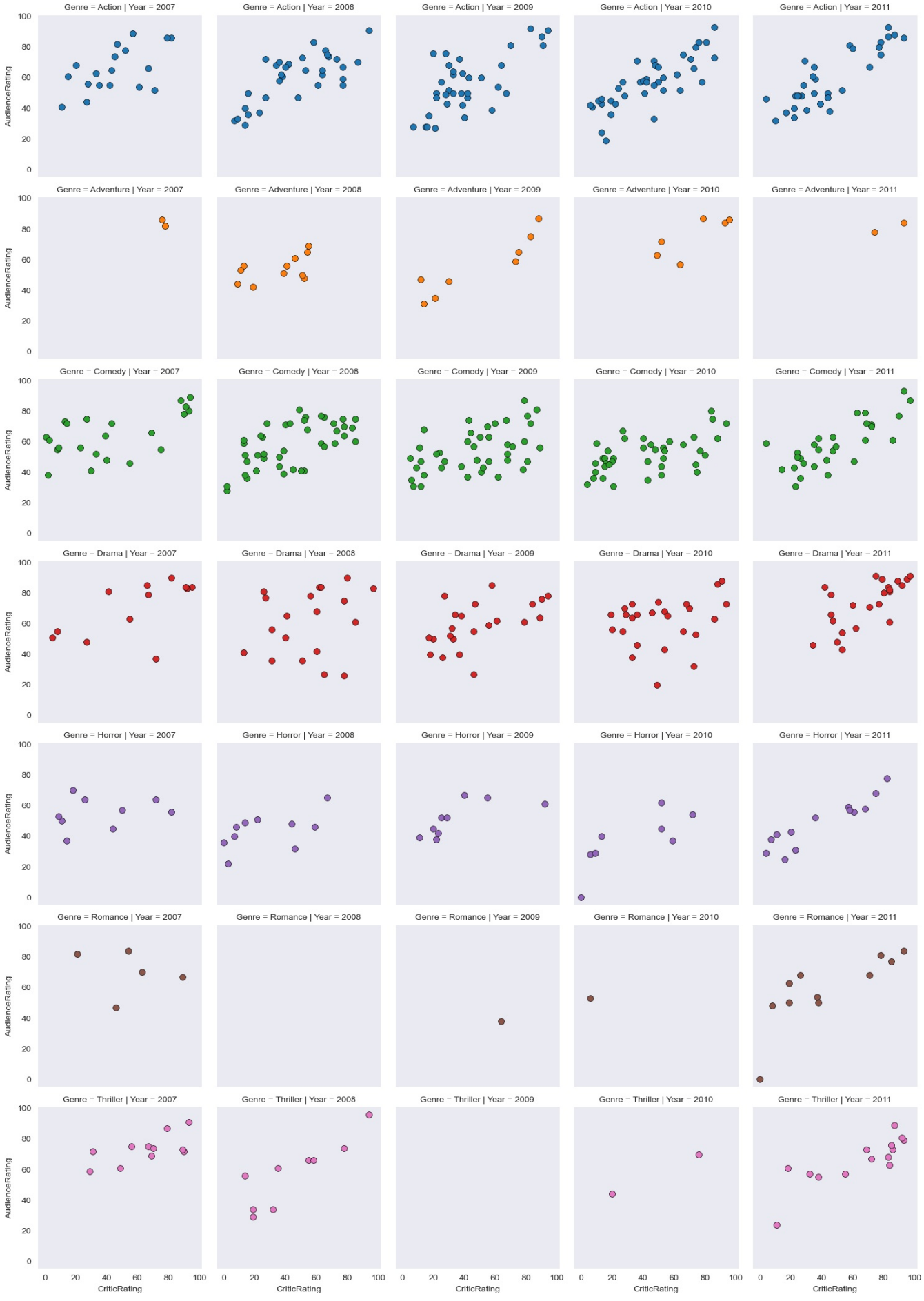


In [77]:

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

In [78]:

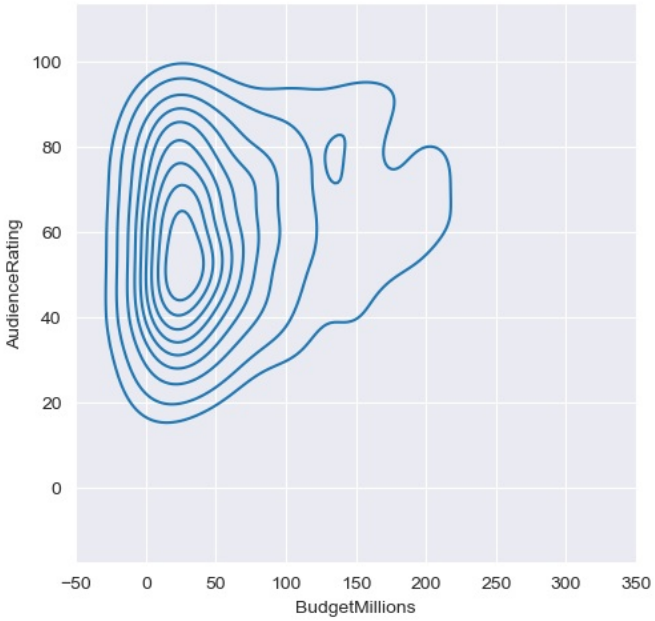
```
g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
kws = dict(s=50, linewidth=0.5,edgecolor='black')
g = g.map(plt.scatter, 'CriticRating', 'AudienceRating' ,**kws ) #scatterplots are mapped in facetgrid
```

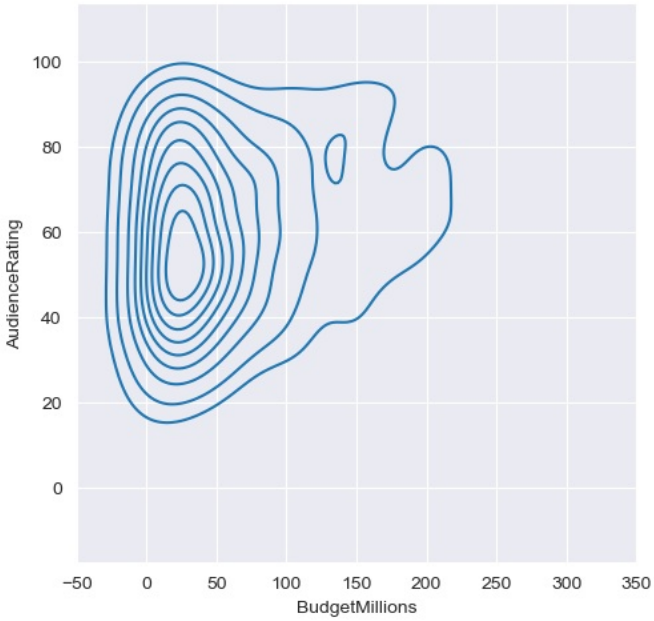


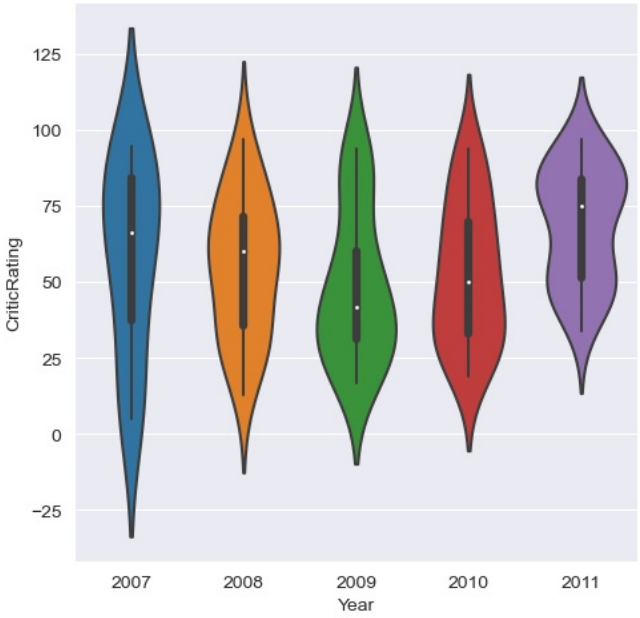
In [86]:

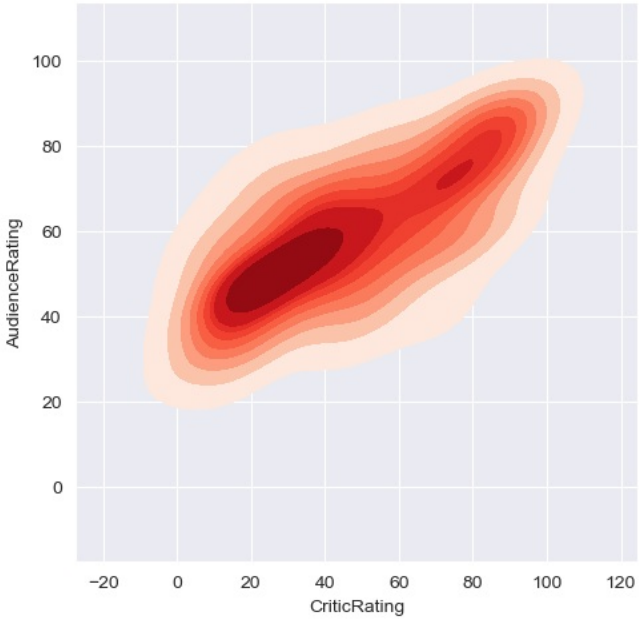
```
sns.set_style('darkgrid')
f,axes=plt.subplots(2,2,figsize=(12,12))

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

```
k3=sns.kdeplot(x='CriticRating', y='AudienceRating',data=movies,cmap='Reds',shade=True,shade_lowest=False,ax=ax
```



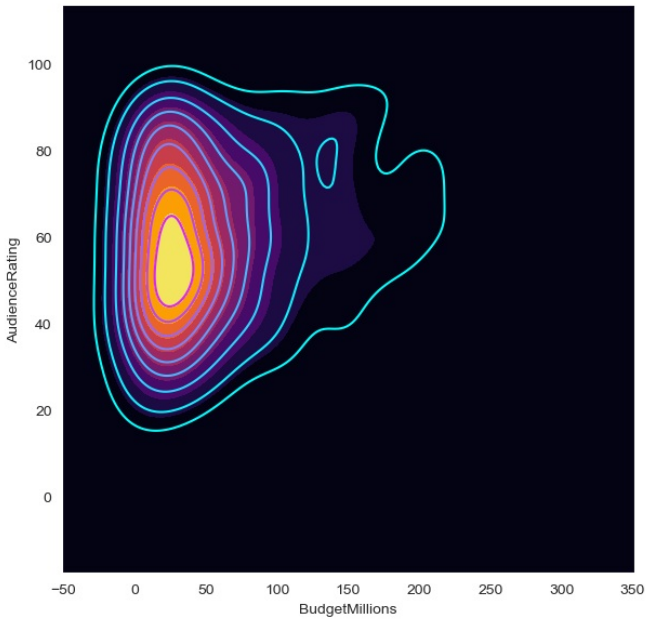


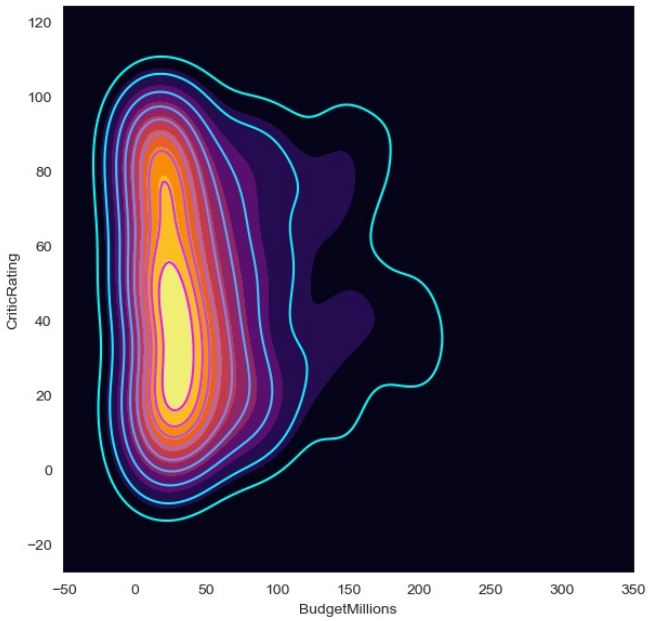


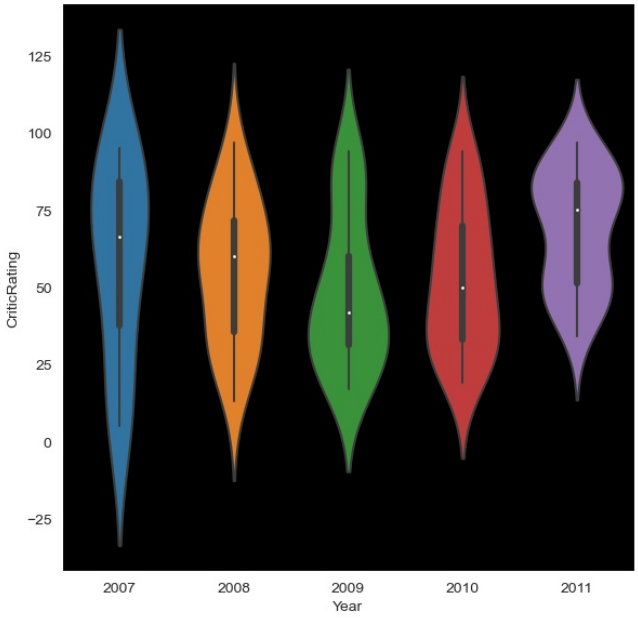


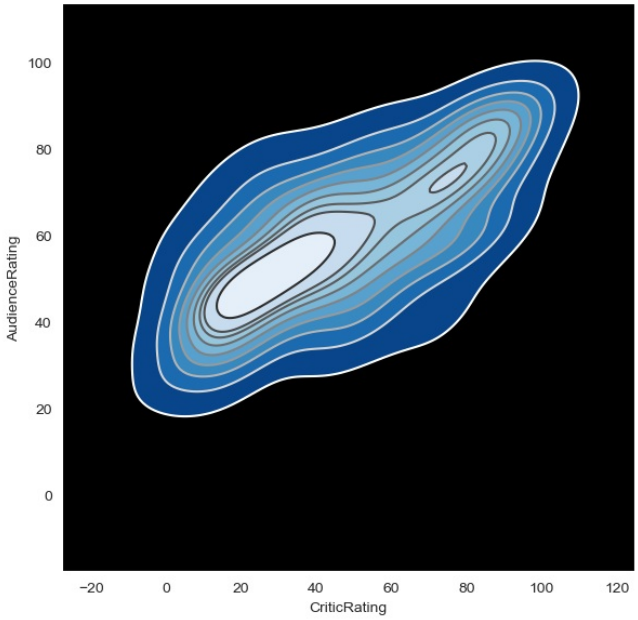
In [93]:

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

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

# Plot [0,1]
k2 = sns.kdeplot(x='BudgetMillions', y='CriticRating', data=movies,
                 shade=True, shade_lowest=True, cmap='inferno',
                 ax=axes[0,1])
k2b = sns.kdeplot(x='BudgetMillions', y='CriticRating', data=movies,
                  cmap='cool', 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='CriticRating', y='AudienceRating', data=movies,
                 shade=True, shade_lowest=False, cmap='Blues_r',
                 ax=axes[1,1])
k4b = sns.kdeplot(x='CriticRating', y='AudienceRating', data=movies,
                  cmap='gist_gray_r', ax=axes[1,1])
```









In [ ]: