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

In [2]:
```
import numpy as np
```

In [3]:
```
os.getcwd() # if you want to change the working directory
```

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

In [4]:
```
movies=pd.read_csv(r'C:\DS & GEN AI\Seaborn\Class Notes\Movie-Rating.csv')
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.head()
```

Out[6]:

	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 [7]:
```
movies.tail()
```

Out[7]:

	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 [8]:
```
movies.columns
```

Out[8]:
```
Index(['Film', 'Genre', 'Rotten Tomatoes Ratings %', 'Audience Ratings %',
       'Budget (million $)', 'Year of release'],
      dtype='object')
```

In [9]:
```
movies.columns=['Film','Genre','CriticRating','AudienceRating','BudgetMillions','Ye
```

In [10]:
```
movies.head()
```

Out[10]:

	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 [11]:
```
movies.head(1)
```

Out[11]:

	Film	Genre	CriticRating	AudienceRating	BudgetMillions	Year
0	(500) Days of Summer	Comedy	87	81	8	2009

In [12]:
```
movies.describe()
```

Out[12]:

	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 [13]:
```
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 [14]:
```
movies['Film']
```

Out[14]:
```
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 [15]:
```
movies.Film
```

Out[15]:
```
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 [16]:
```
movies.Film=movies.Film.astype('category')
movies.Film
```

Out[16]:
```
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 [17]:
```
movies.Genre=movies.Genre.astype('category')
movies.Genre
```

Out[17]:
```
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', 'Roma
nce', 'Thriller']
```

In [18]:
```
movies.Year=movies.Year.astype('category')
movies.Year
```

Out[18]:
```
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 [19]:
```
movies.describe()
```

Out[19]:

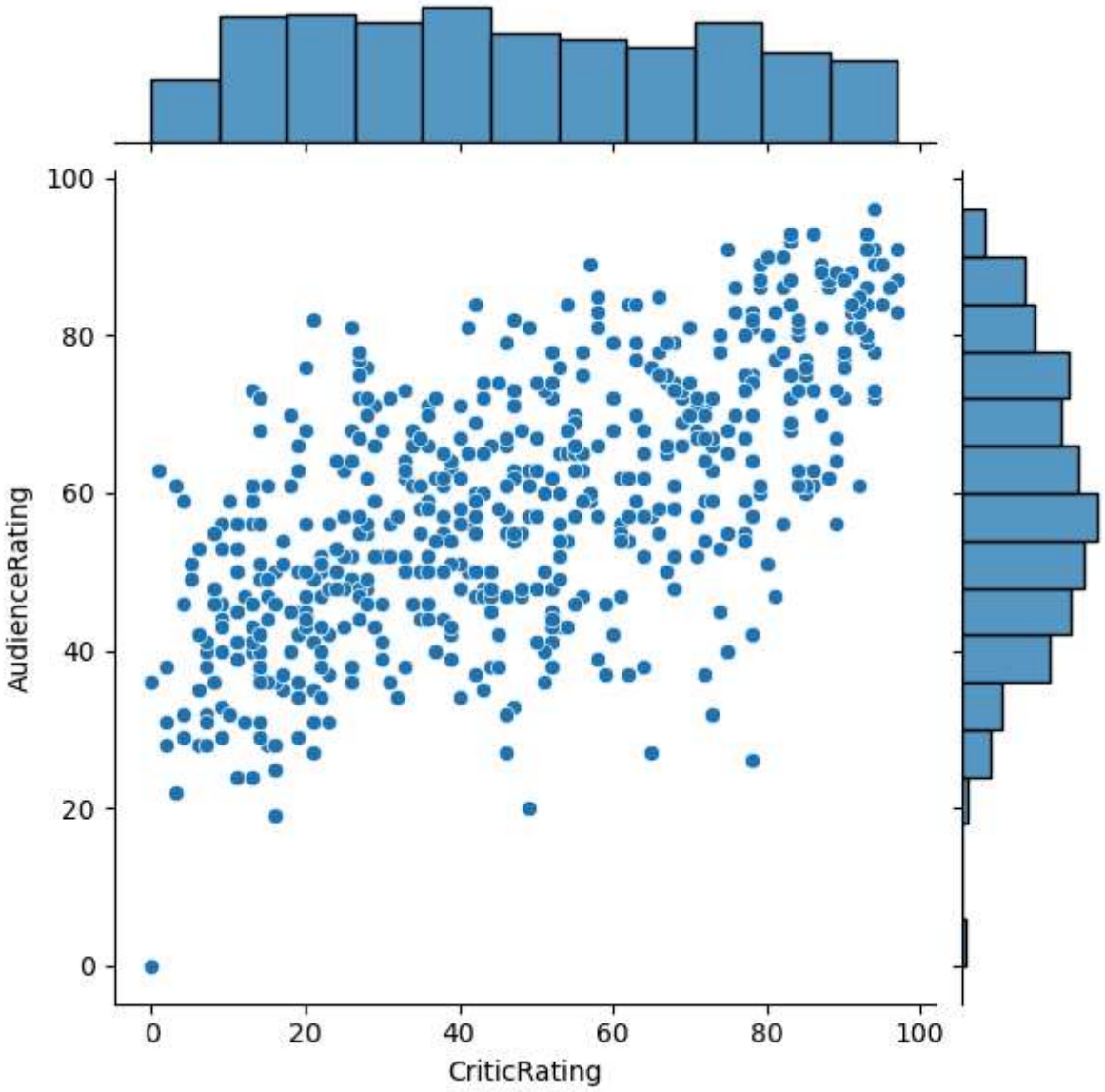
	CriticRating	AudienceRating	BudgetMillions
count	559.000000	559.000000	559.000000
mean	47.309481	58.744186	50.236136
std	26.413091	16.826887	48.731817
min	0.000000	0.000000	0.000000
25%	25.000000	47.000000	20.000000
50%	46.000000	58.000000	35.000000
75%	70.000000	72.000000	65.000000
max	97.000000	96.000000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

In [20]:
```
movies.info()
```

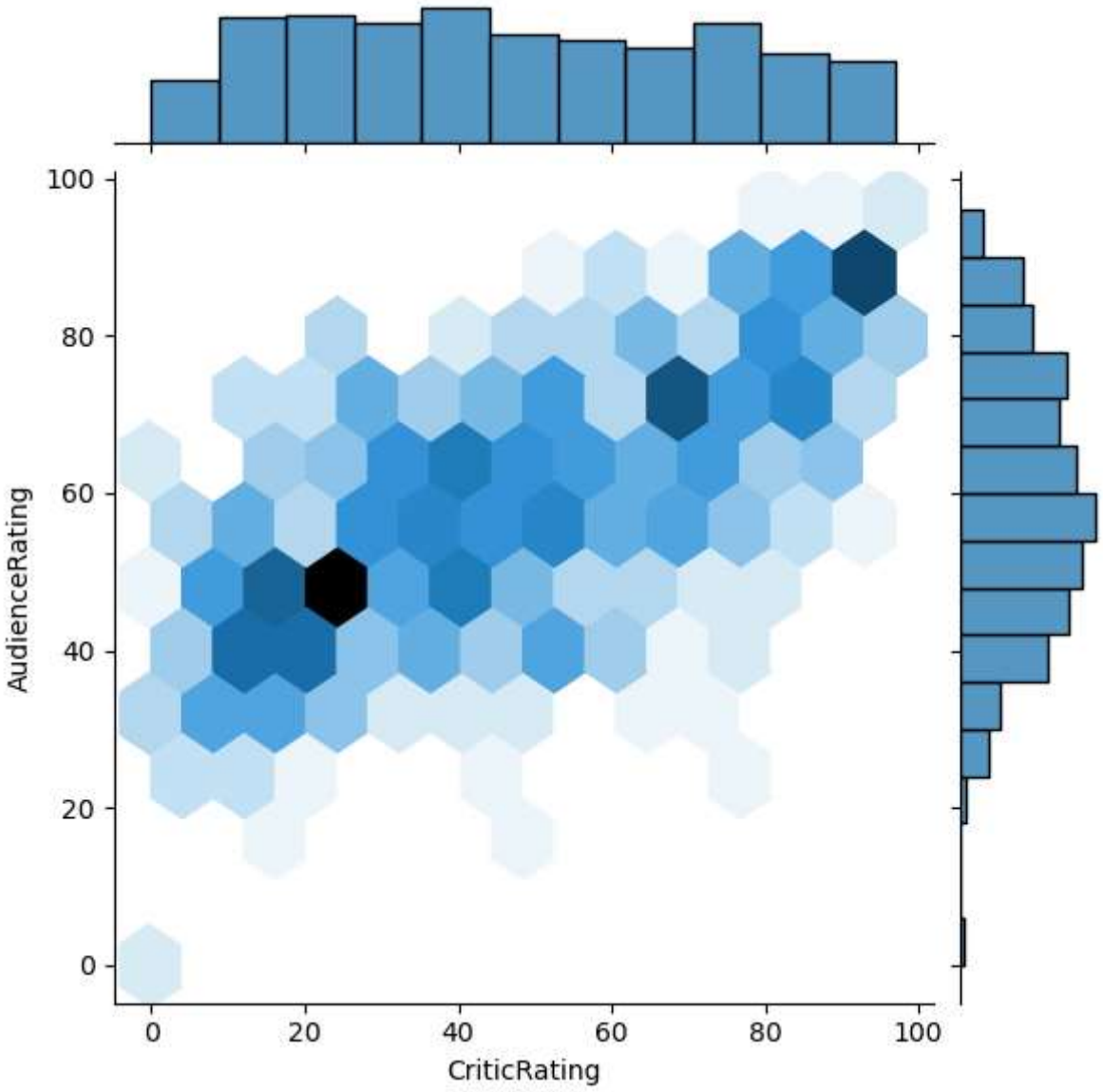
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 559 entries, 0 to 558
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Film            559 non-null    category
 1   Genre           559 non-null    category
 2   CriticRating    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 [21]:
```
from matplotlib import pyplot as plt
import seaborn as sns

%matplotlib inline

import warnings
warnings.filterwarnings('ignore')
```

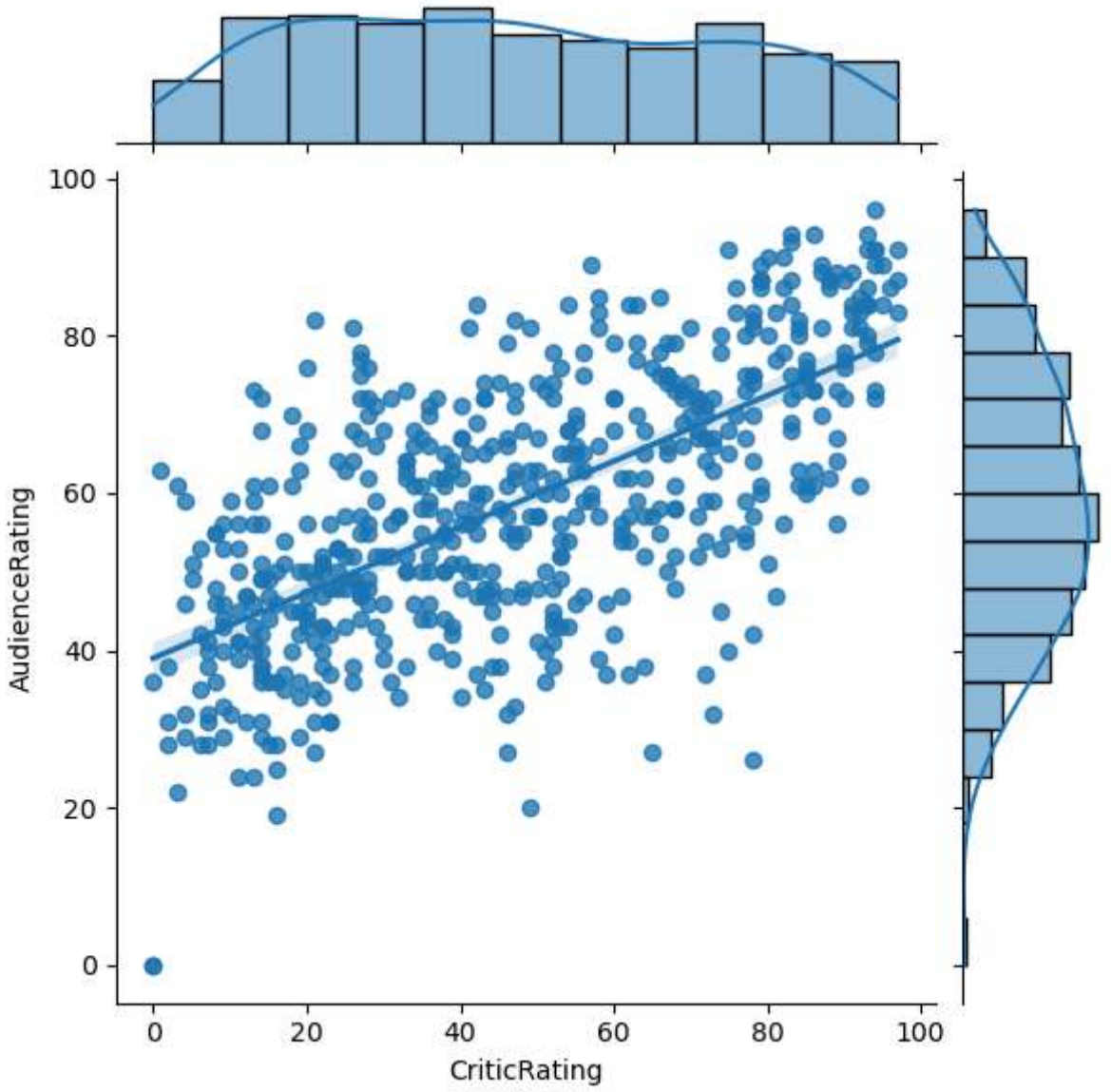
In [22]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating')
plt.show()
```



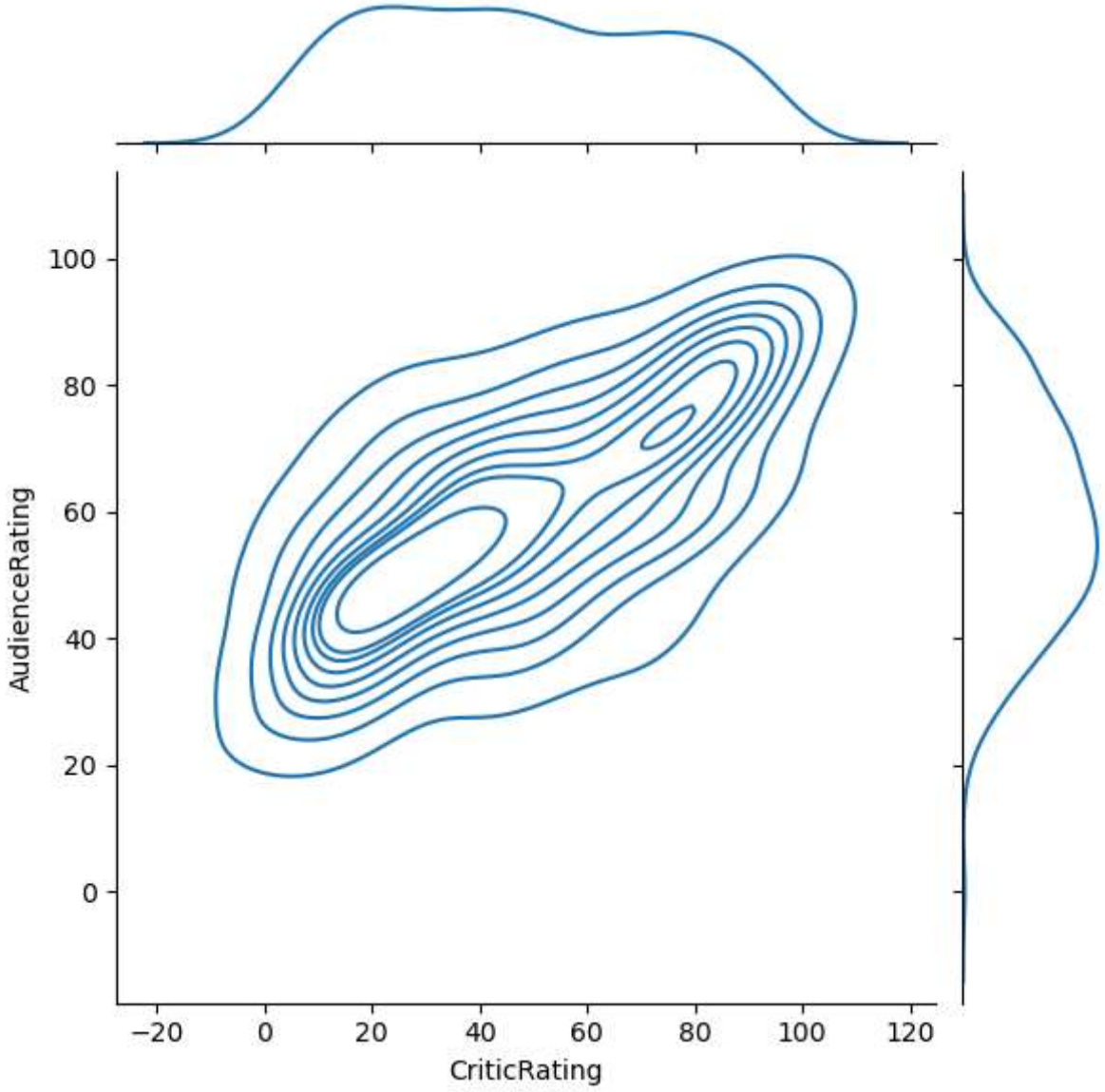
In [23]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='hex')
plt.show()
```



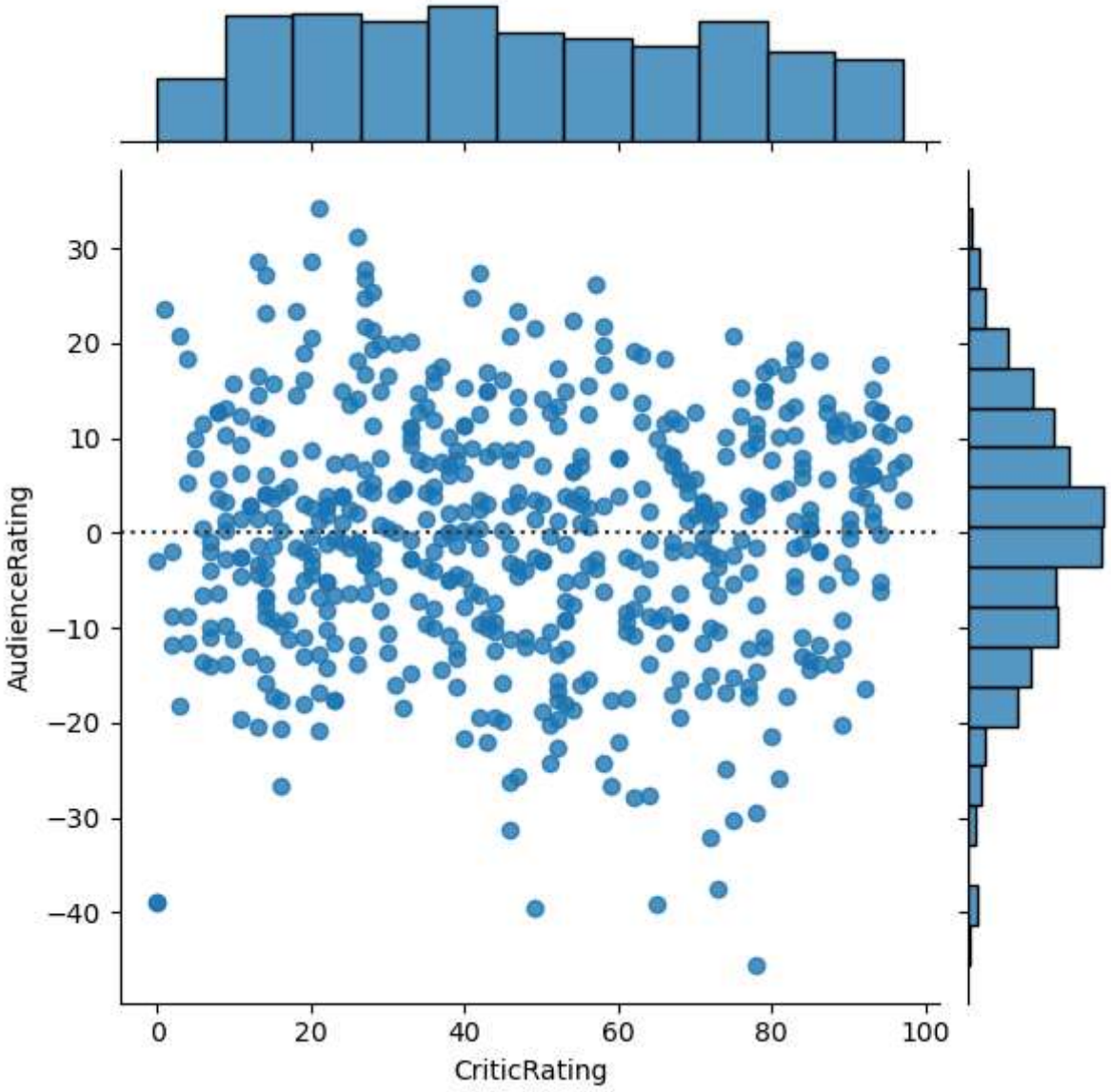
In [24]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='reg')
plt.show()
```



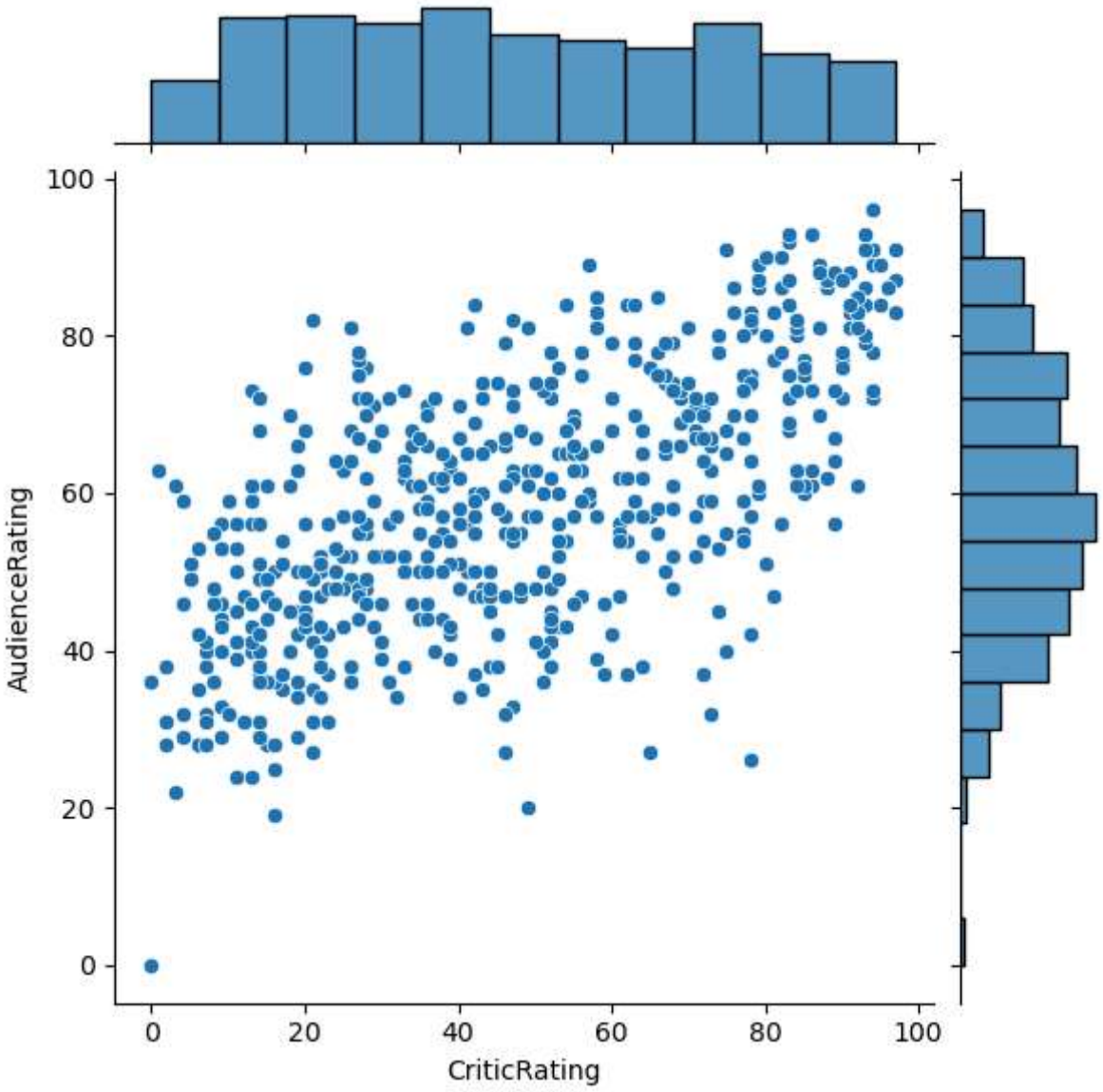
In [25]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='kde')
plt.show()
```



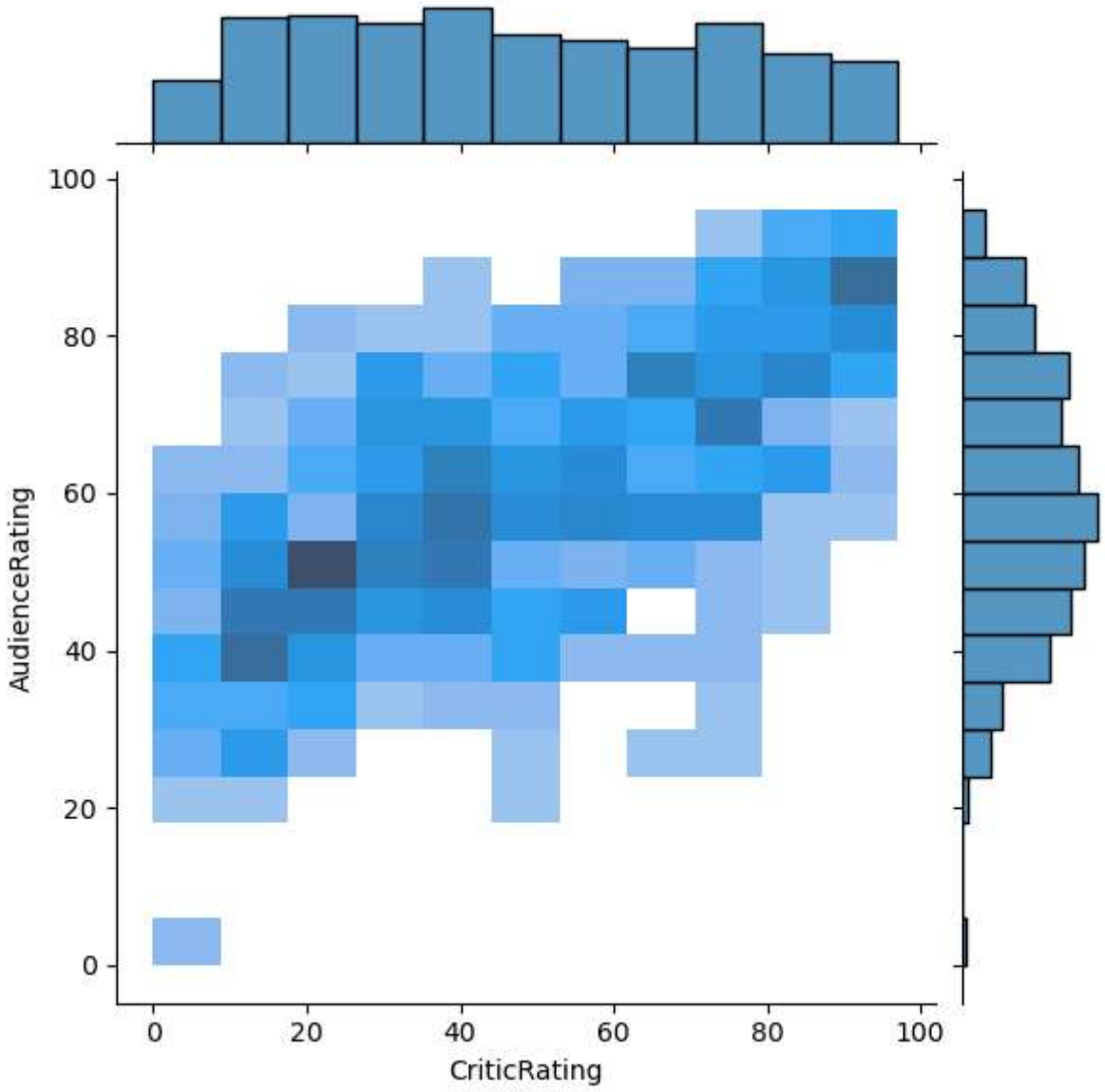
In [26]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='resid')
plt.show()
```



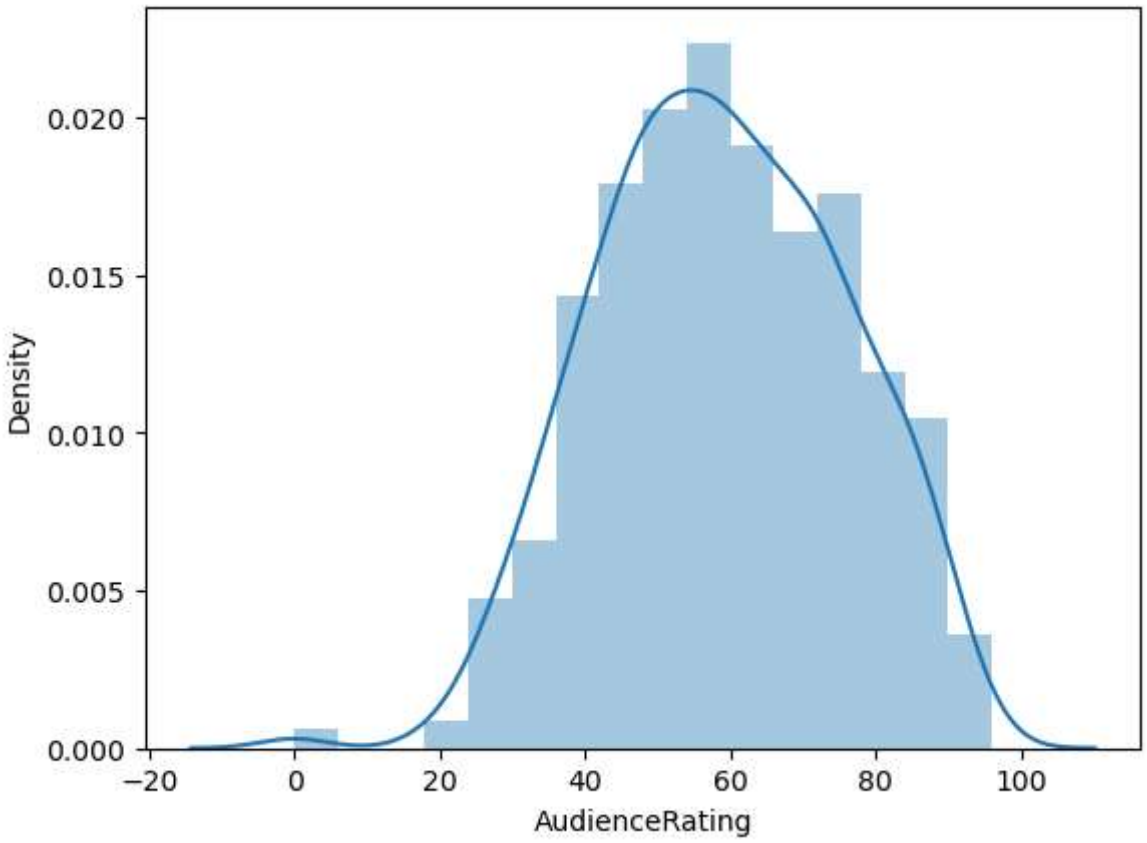
In [27]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='scatter')
plt.show()
```



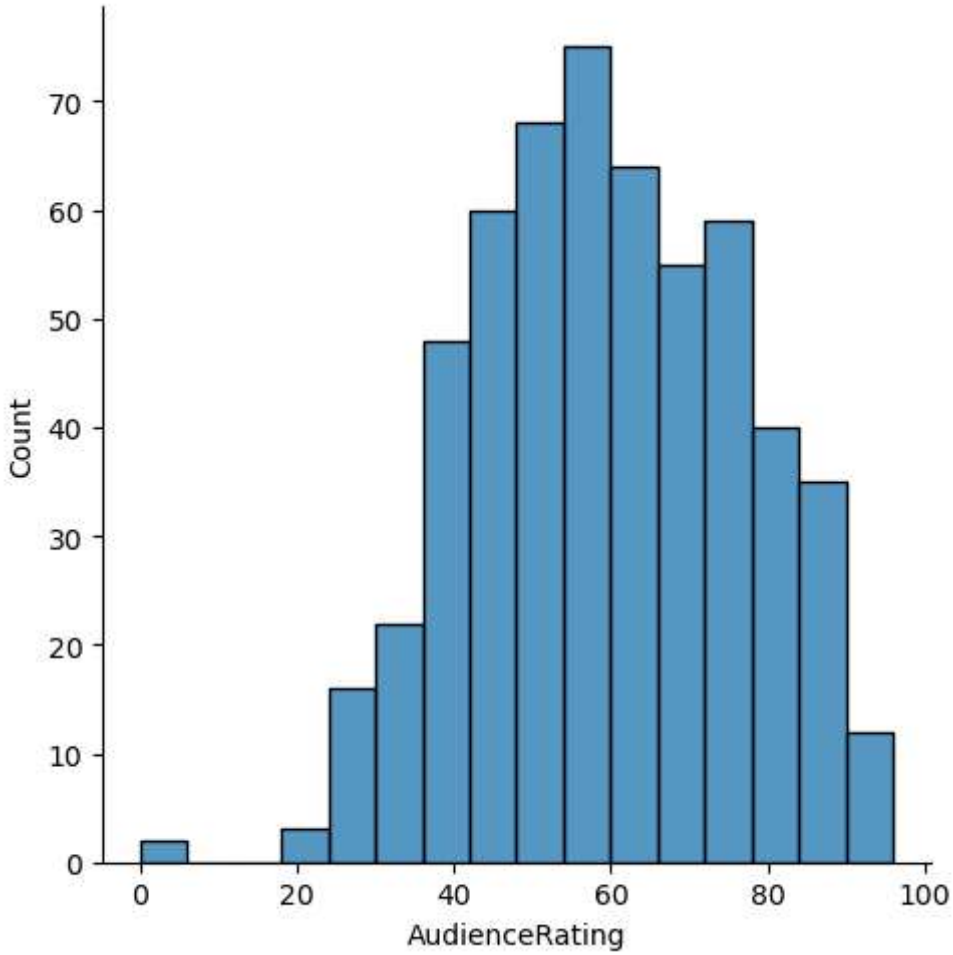
In [28]:
```
j=sns.jointplot(data=movies,x='CriticRating',y='AudienceRating',kind='hist')
plt.show()
```



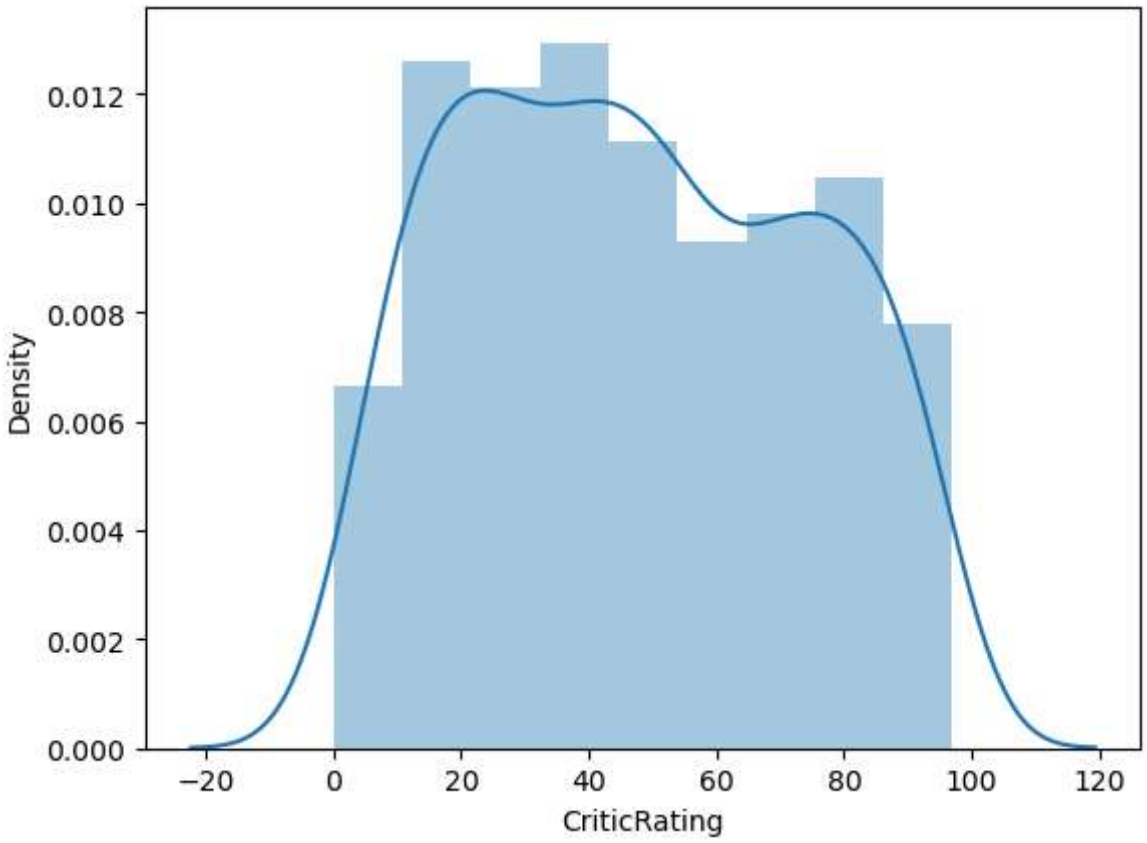
In [29]:
```
m1=sns.distplot(movies.AudienceRating)      #Histograms
plt.show()
```



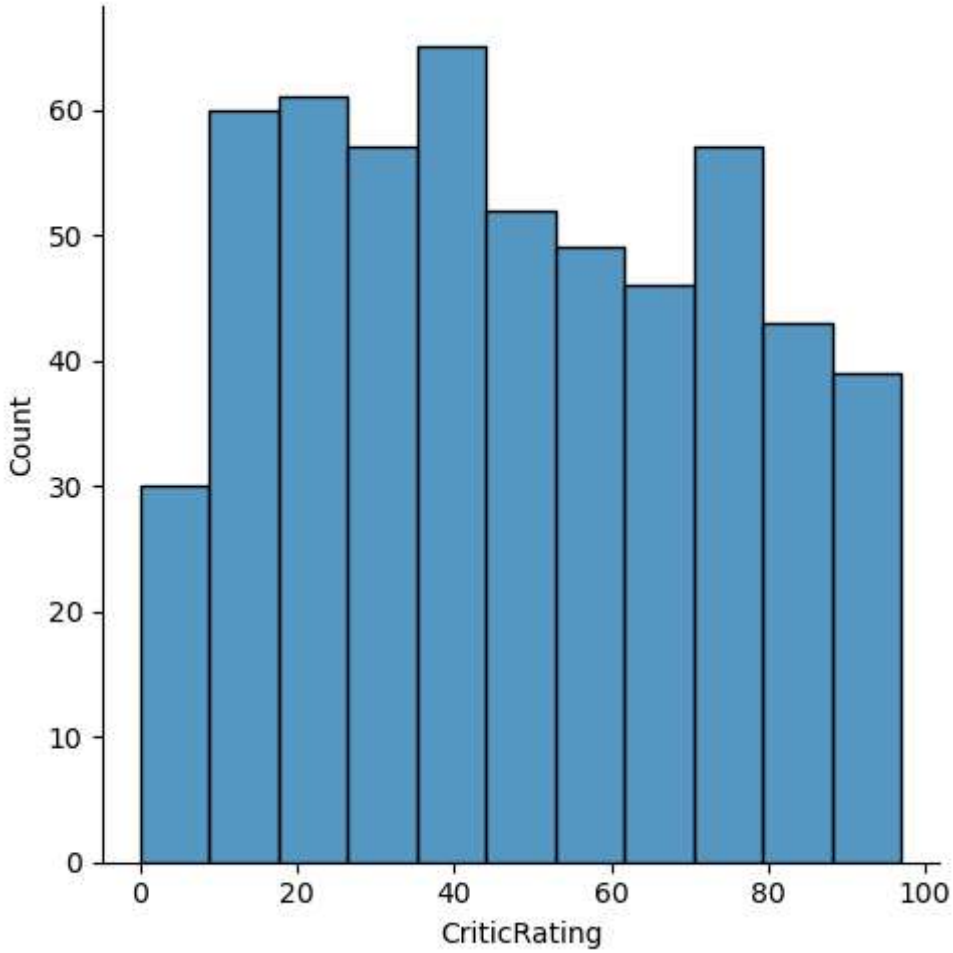
In [30]:
```
m1=sns.displot(movies.AudienceRating)
plt.show()
```



In [31]:
```
m1=sns.distplot(movies.CriticRating)
plt.show()
```

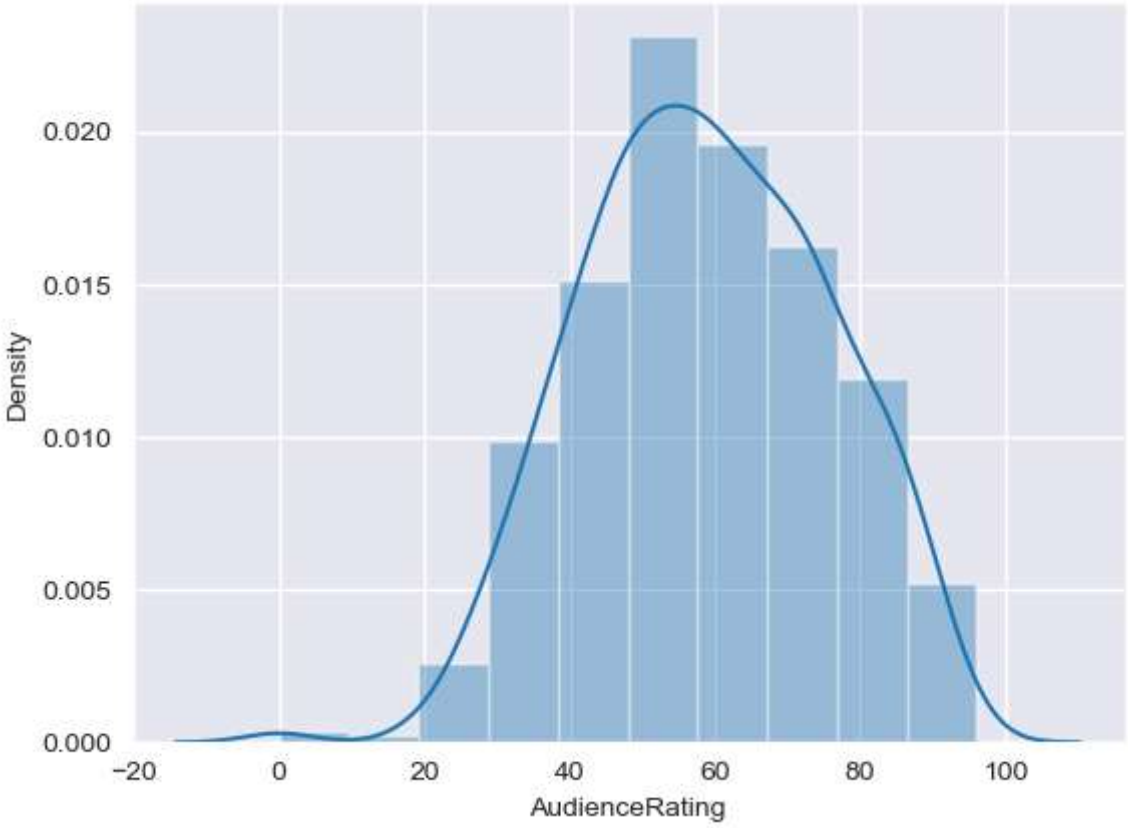


In [32]:
```
m1=sns.displot(movies.CriticRating)
plt.show()
```

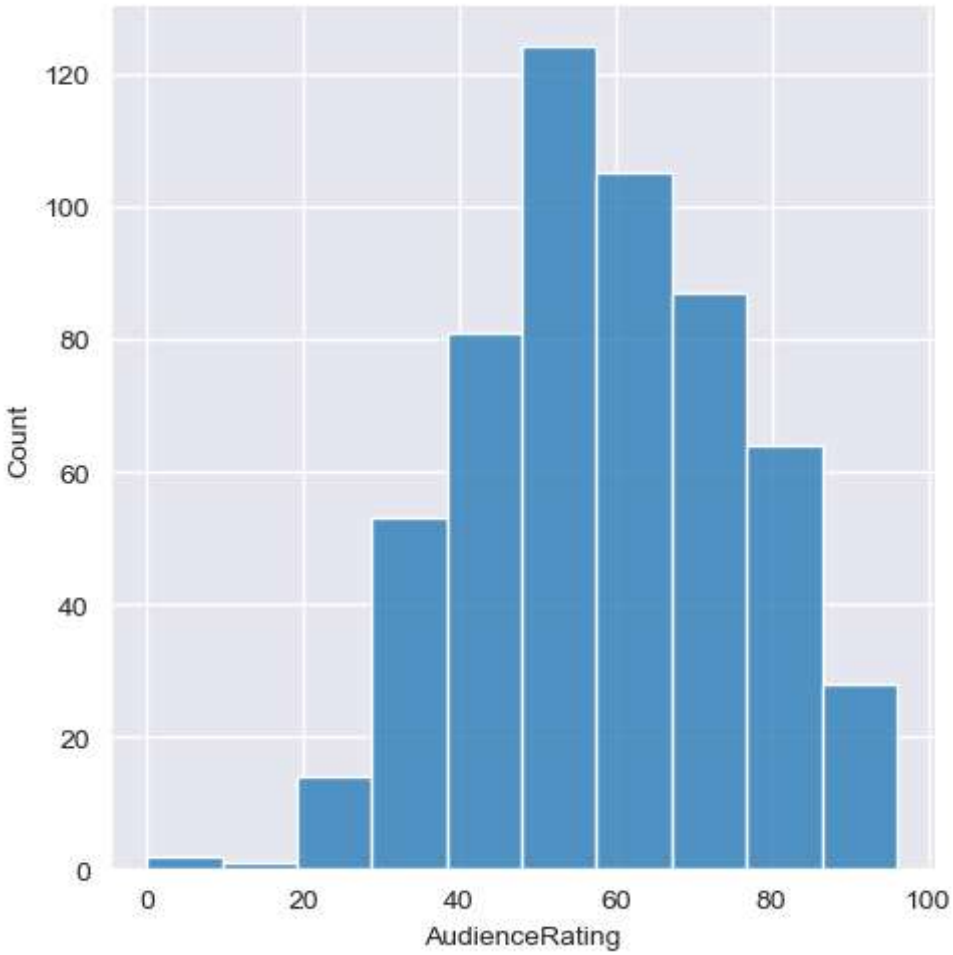


In [33]:
```
sns.set_style('darkgrid')
```

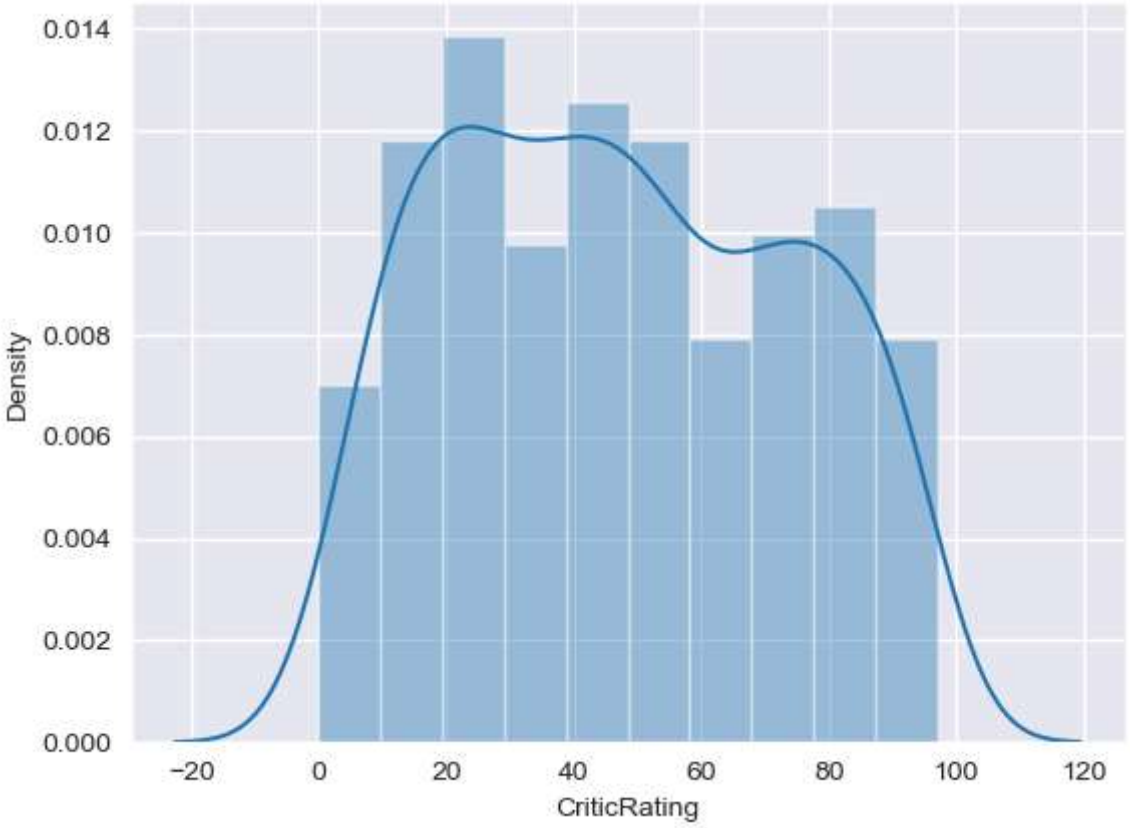
In [34]:
```
m2=sns.distplot(movies.AudienceRating,bins=10)
plt.show()
```



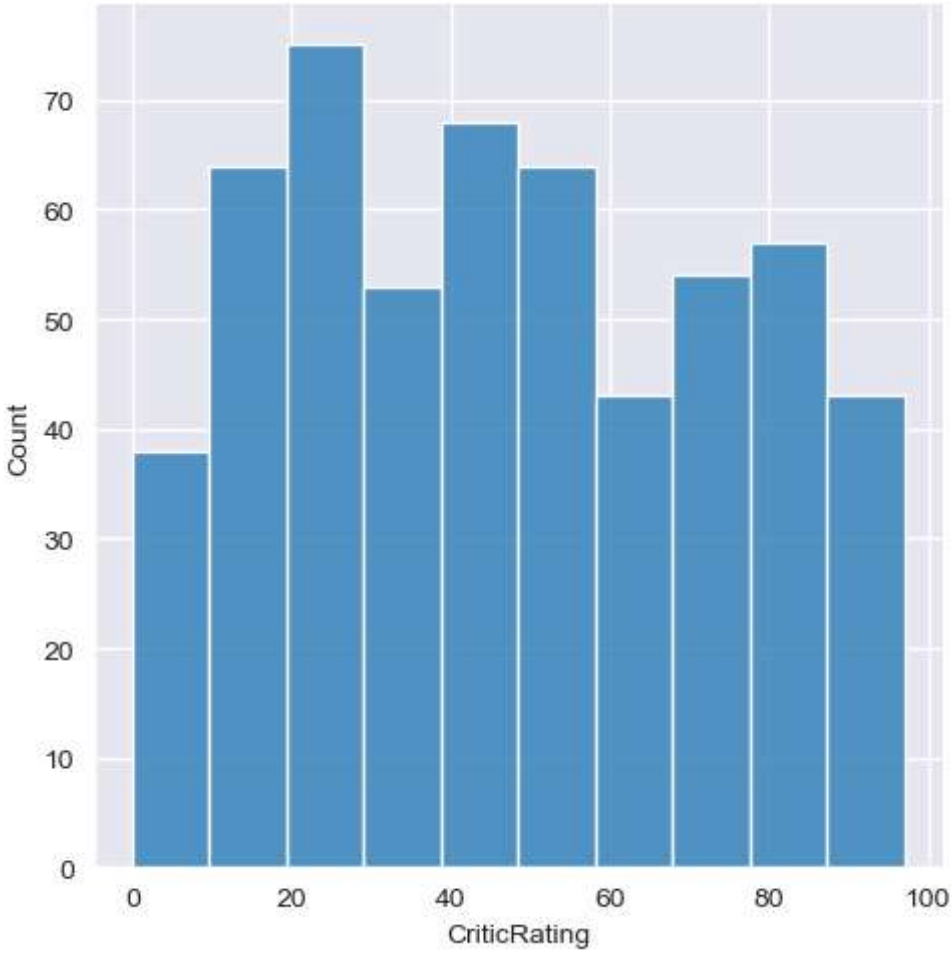
In [35]:
```
m2=sns.displot(movies.AudienceRating,bins=10)
plt.show()
```



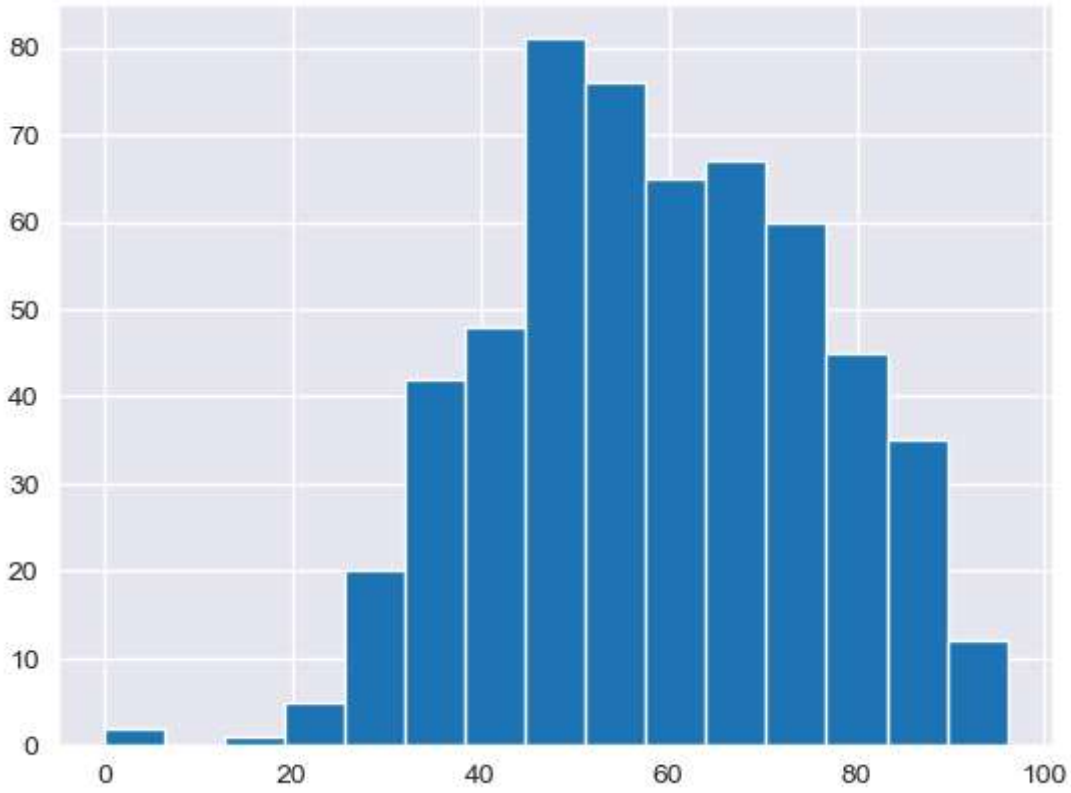
In [36]:
```
m2=sns.distplot(movies.CriticRating,bins=10)
plt.show()
```



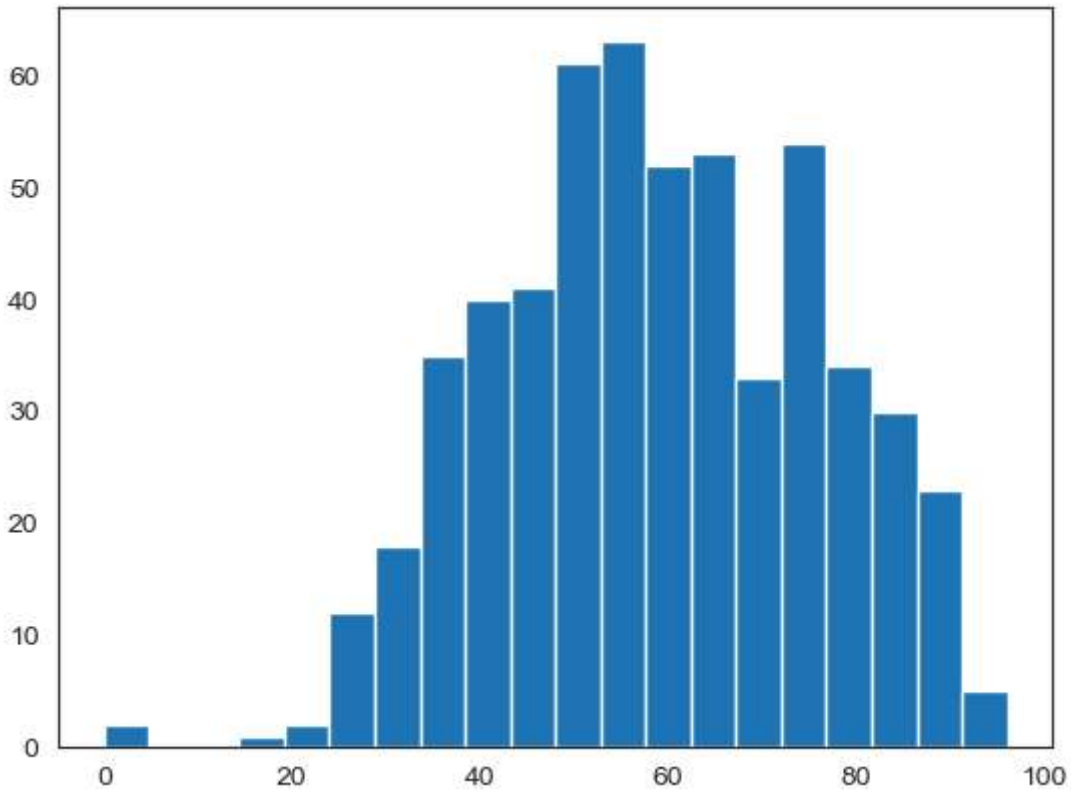
In [37]:
```
m2=sns.displot(movies.CriticRating,bins=10)
plt.show()
```



In [38]:
```
#sns.set_style('darkgrid')
n1 = plt.hist(movies.AudienceRating, bins=15)
plt.show()
```

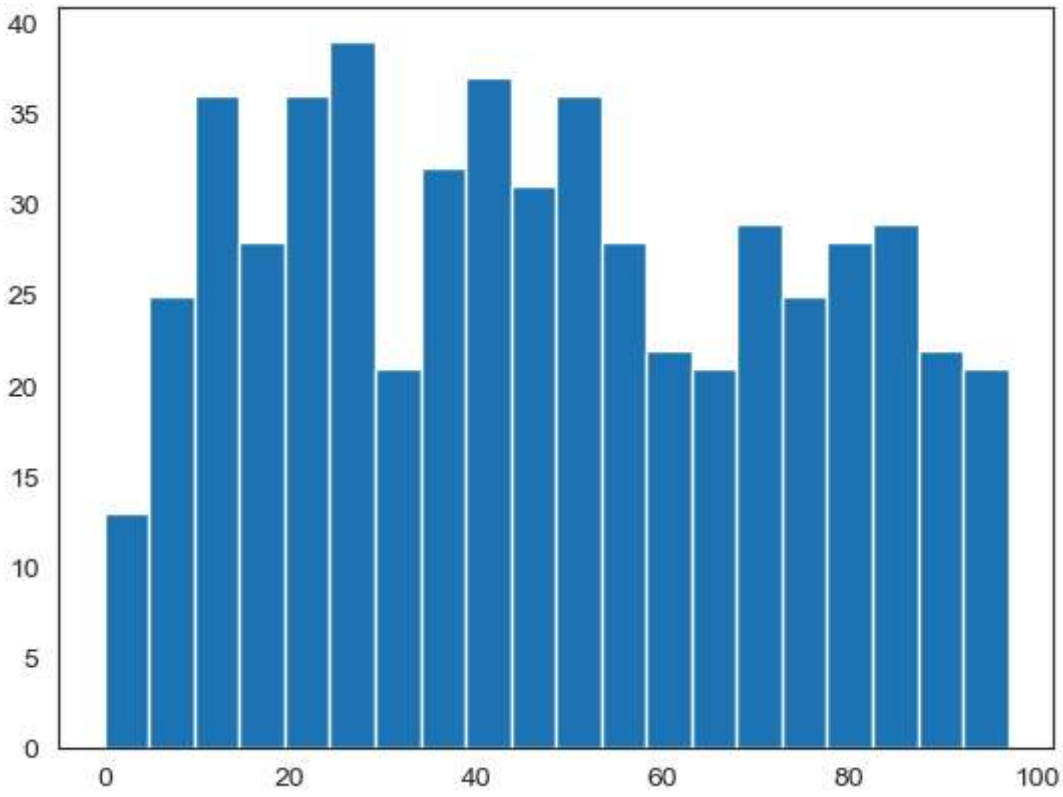


In [39]:
```
sns.set_style('white') #normal distribution & called as bell curve
n1 = plt.hist(movies.AudienceRating, bins=20)
plt.show()
```

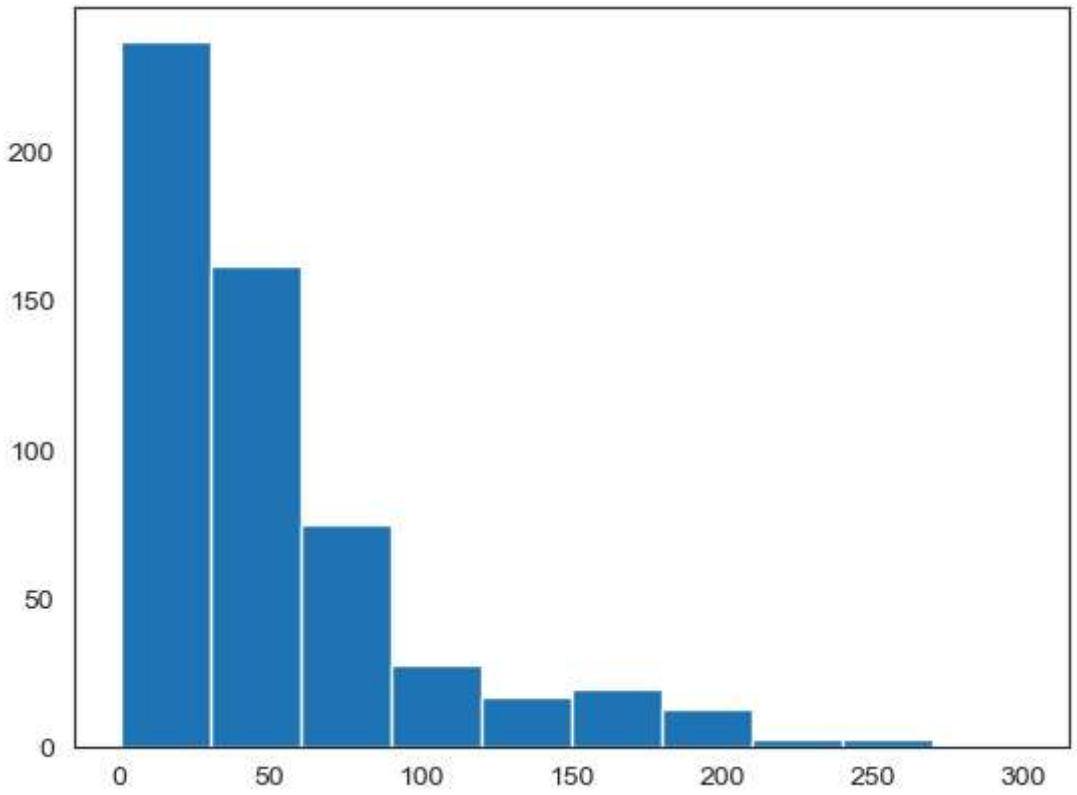


In [40]:
```
n1 = plt.hist(movies.CriticRating, bins=20) #uniform distribution
```

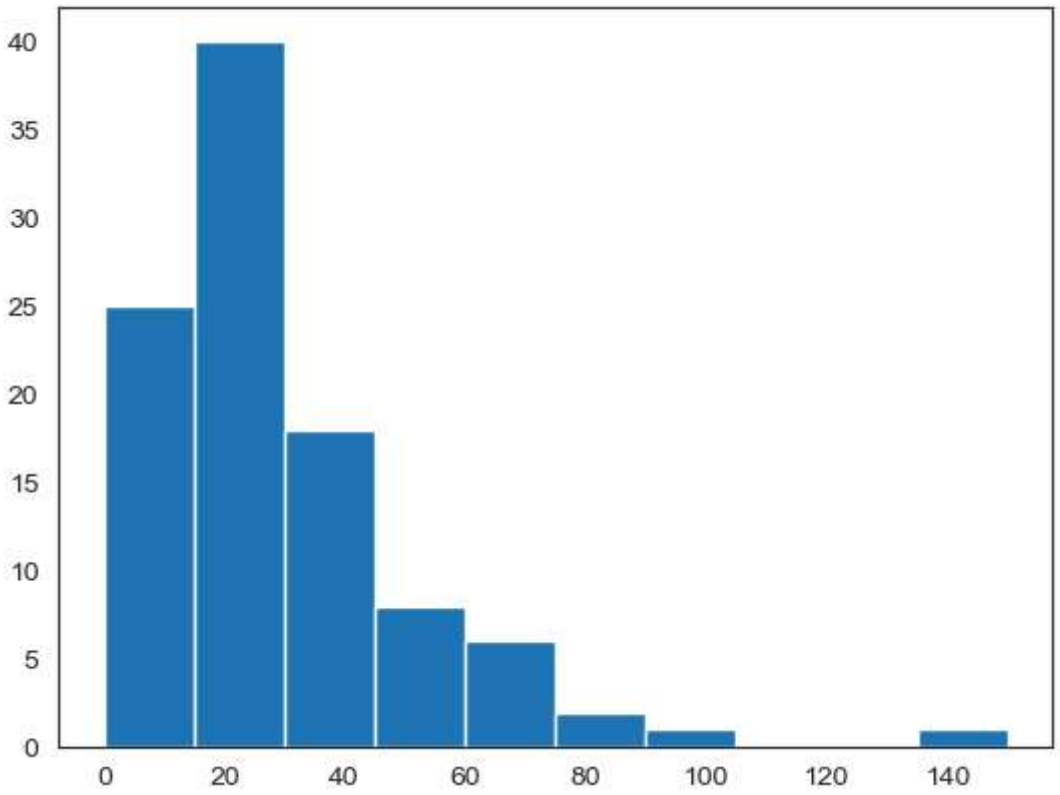
```
plt.show()
```



In [41]:
```
#h1 = plt.hist(movies.BudgetMillions)

plt.hist(movies.BudgetMillions)
plt.show()
```



In [42]:
```
plt.hist(movies[movies.Genre == 'Drama'].BudgetMillions)
plt.show()
```

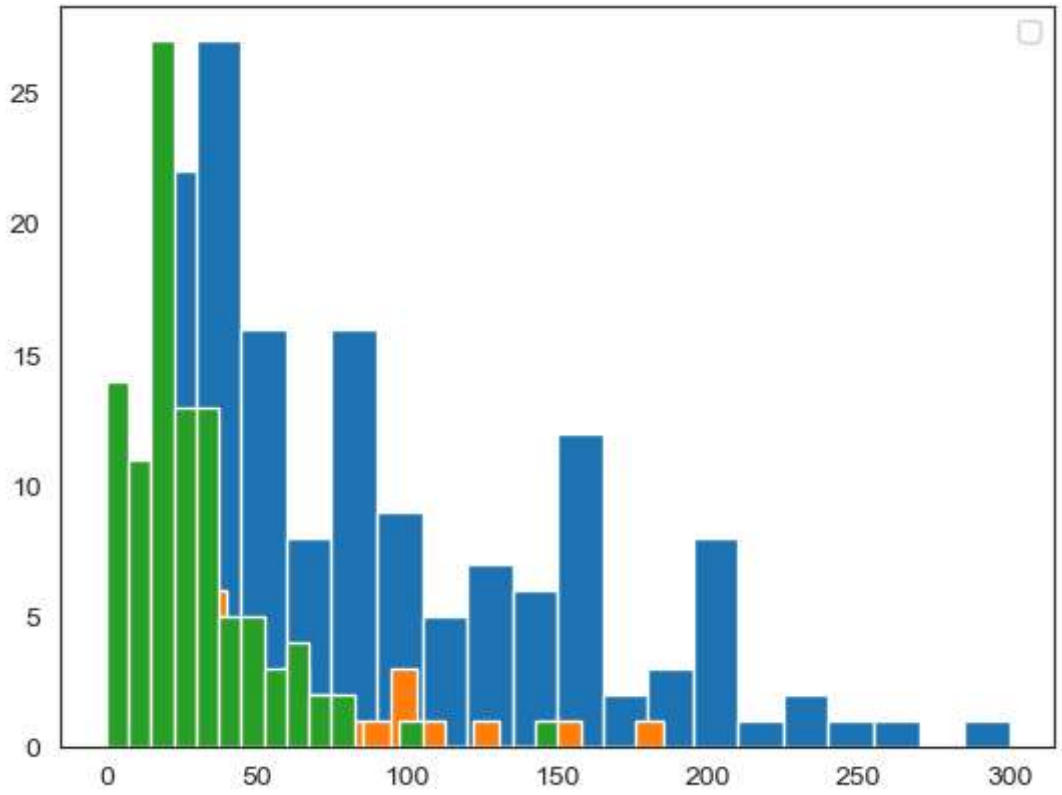


In [43]:
```
movies.head()
```

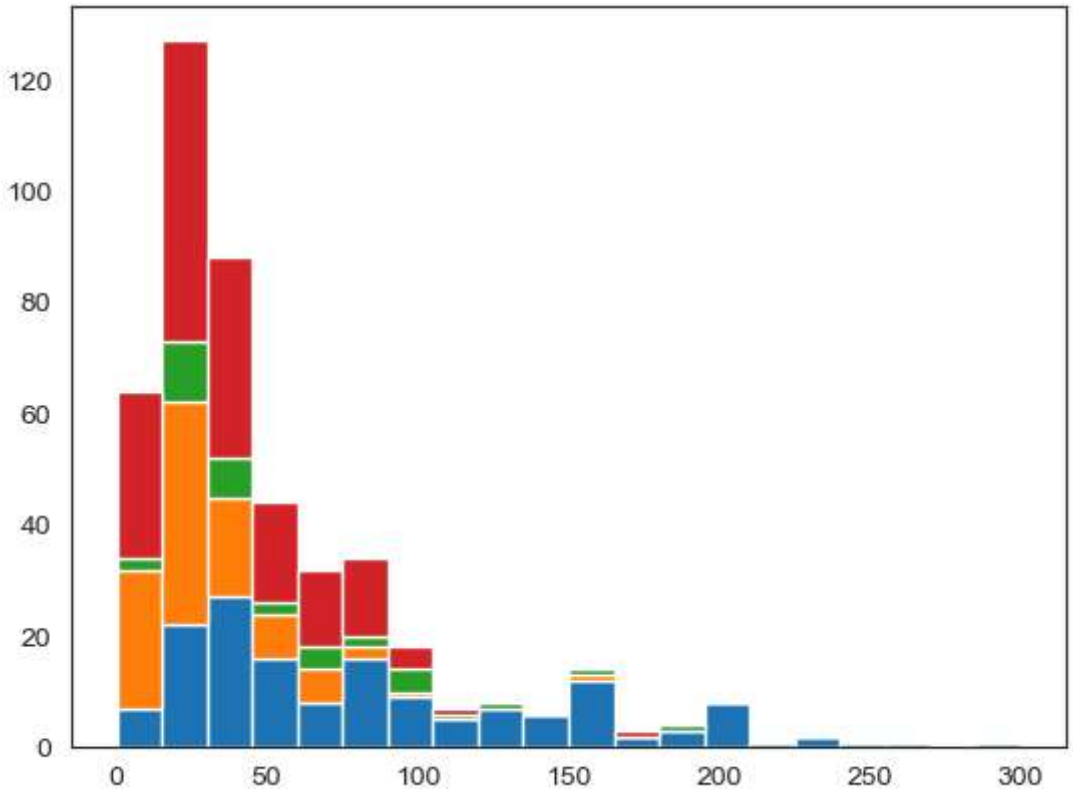
Out[43]:

	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 [44]:
```
# Below plots are stacked histogram becuase overlaped

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()
```



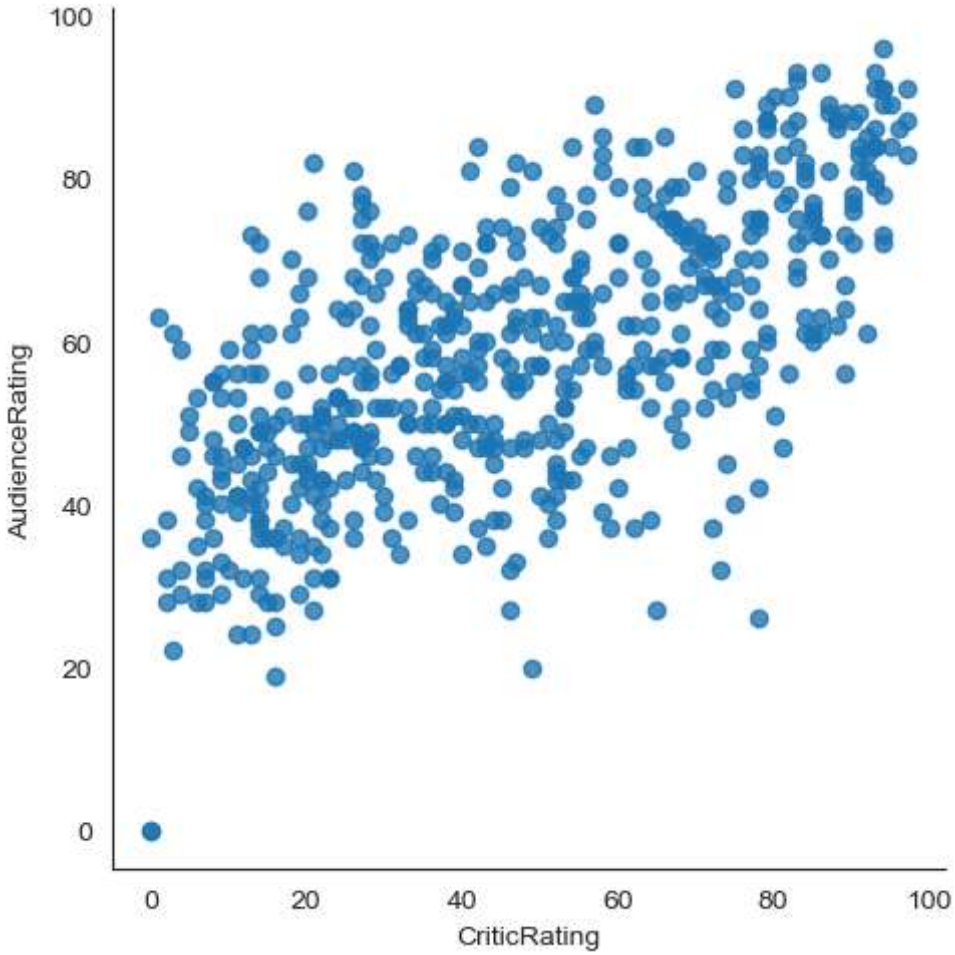
In [45]:
```
plt.hist([movies[movies.Genre == 'Action'].BudgetMillions,\
          movies[movies.Genre == 'Drama'].BudgetMillions, \
          movies[movies.Genre == 'Thriller'].BudgetMillions, \
          movies[movies.Genre == 'Comedy'].BudgetMillions],
         bins = 20, stacked = True)
plt.show()
```



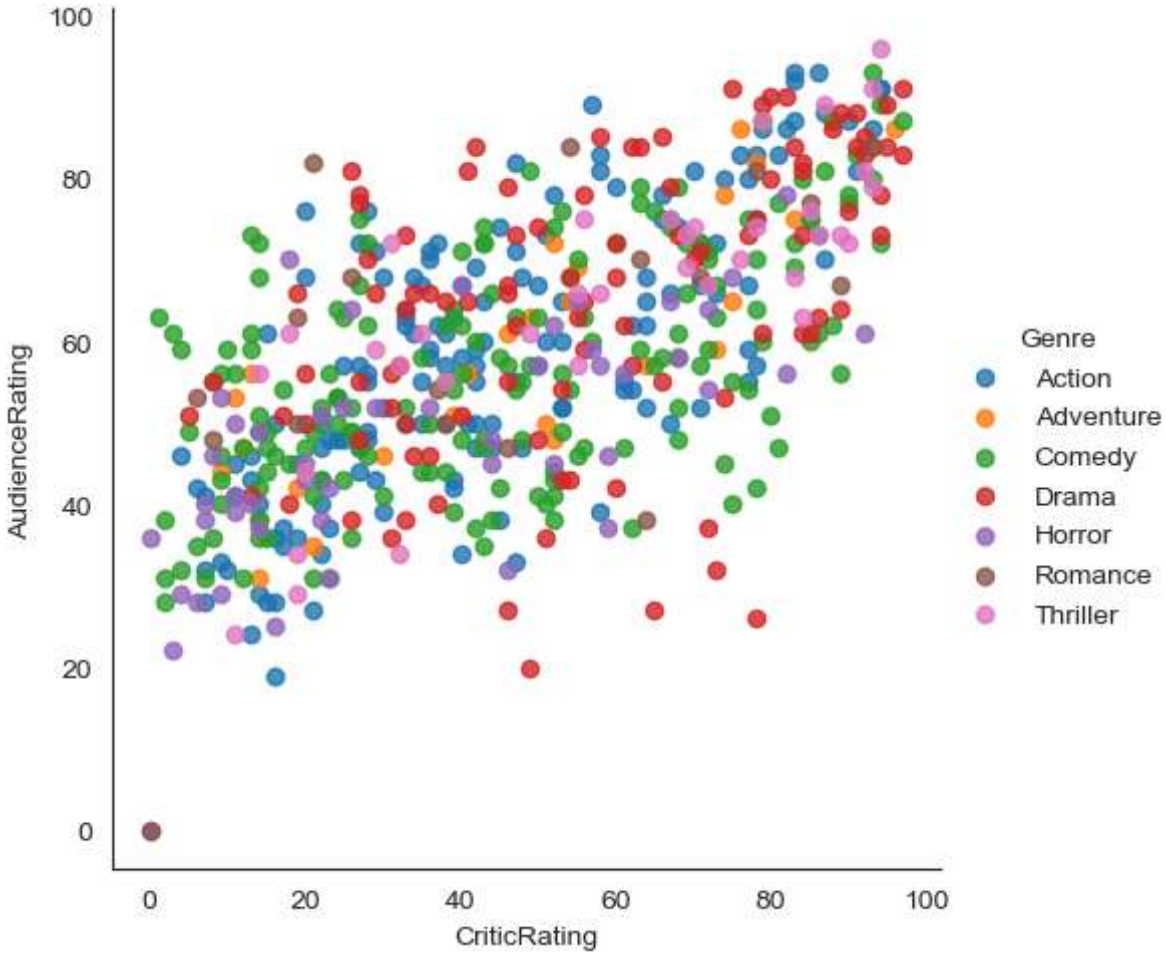
In [46]:
```
# if you have 100 categories you cannot copy & paste all the things

for gen in movies.Genre.cat.categories:
    print(gen)
```

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

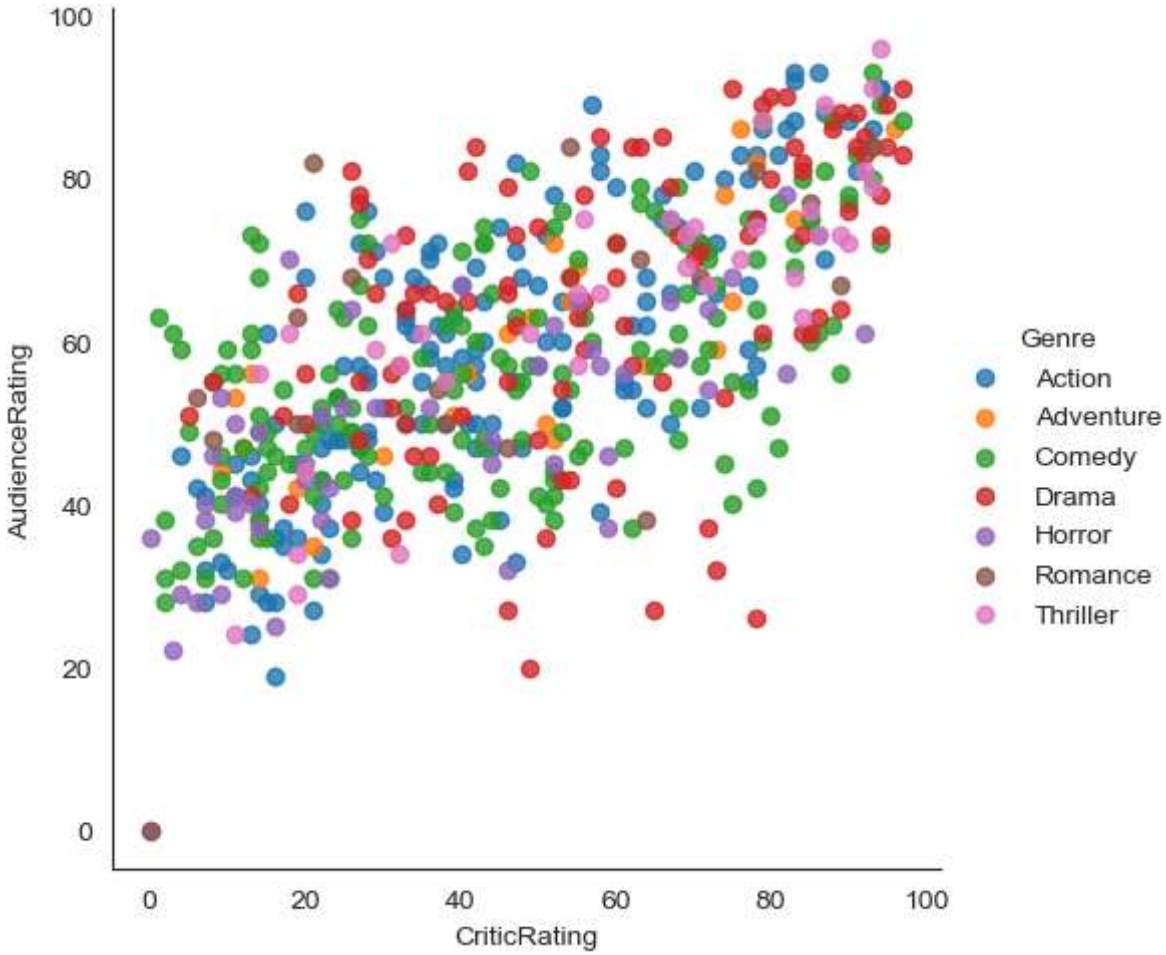
In [47]:
```
vis1 = sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',\
                  fit_reg=False)
plt.show()
```



In [48]:
```
vis1 = sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',\
                  fit_reg=False, hue = 'Genre')
plt.show()
```

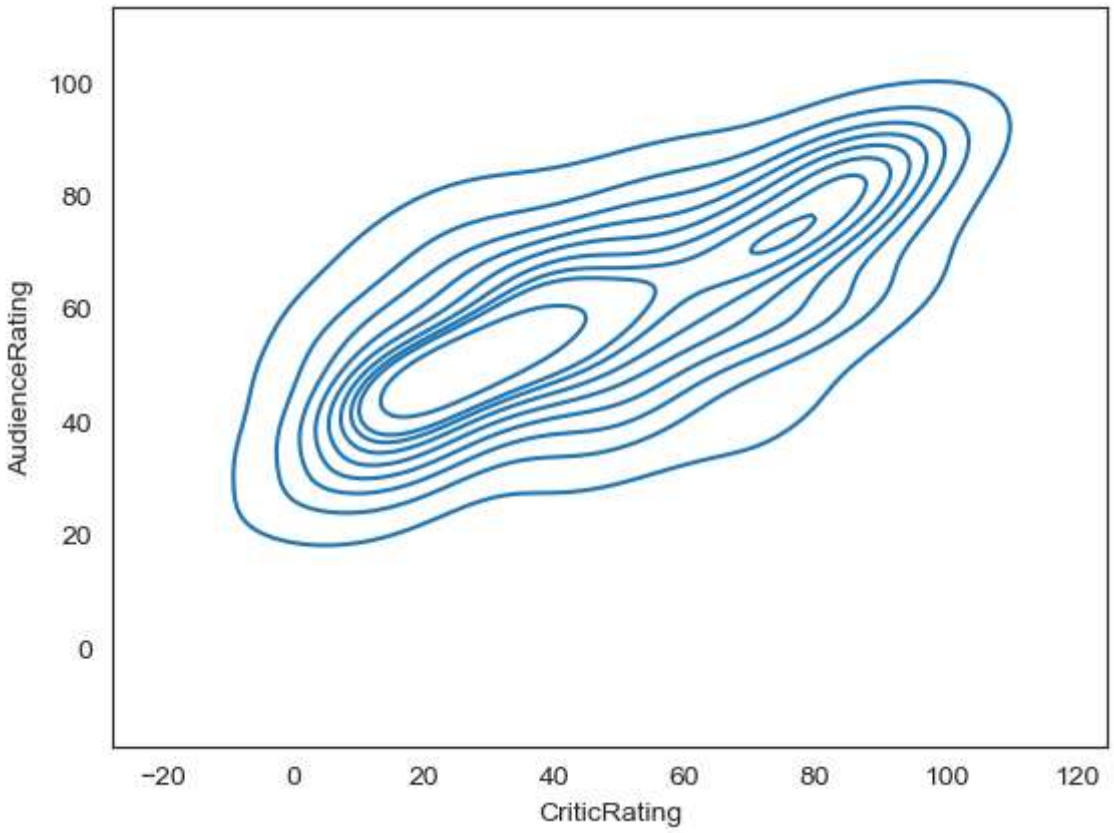


In [49]:
```
vis1 = sns.lmplot(data=movies, x='CriticRating', y='AudienceRating',\
                  fit_reg=False, hue = 'Genre',aspect=1)
plt.show()
```

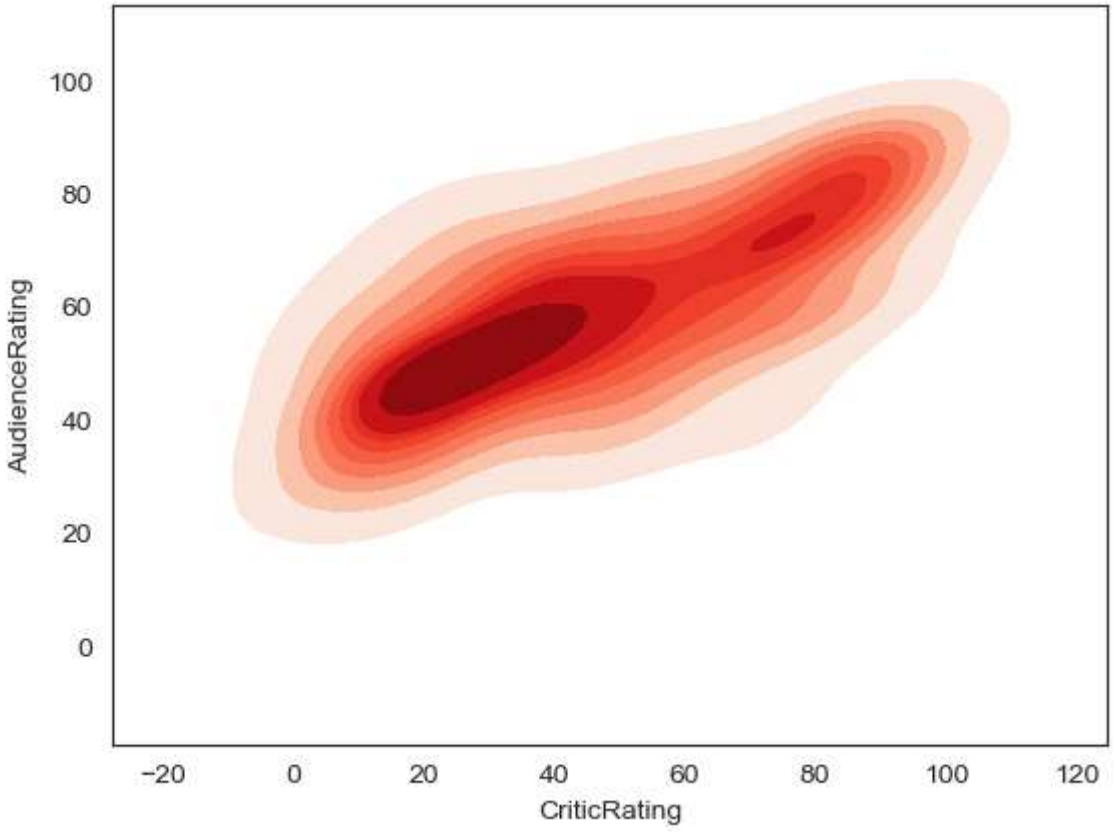


In [50]:
```
# Kernal Density Estimate plot ( KDE PLOT)
# how can i visulize audience rating & critics rating . using scatterplot
```

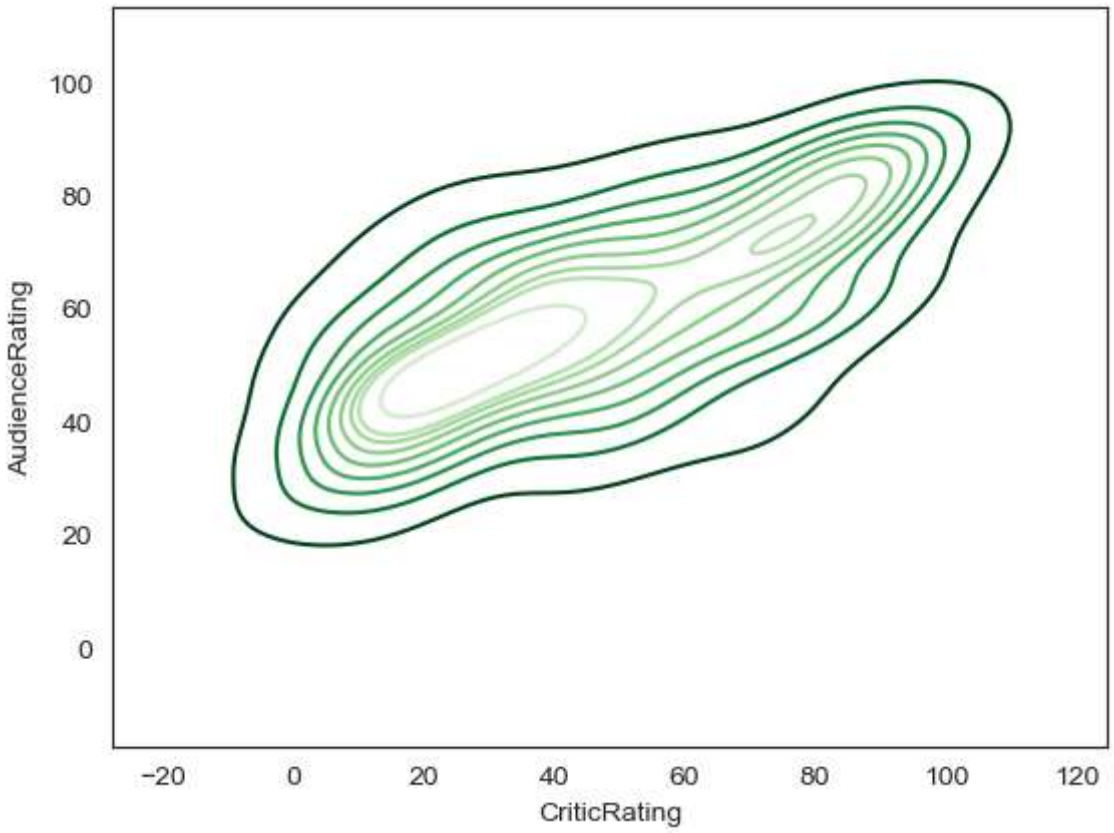
In [51]:
```
k1 = sns.kdeplot(data=movies,x='CriticRating',y='AudienceRating')
plt.show()

# where do u find more density and how density is distibuted across from the the ch
# center point is kernal this is calld KDE & insteade of dots it visualize like thi
# we can able to clearly see the spread at the audience ratings
```



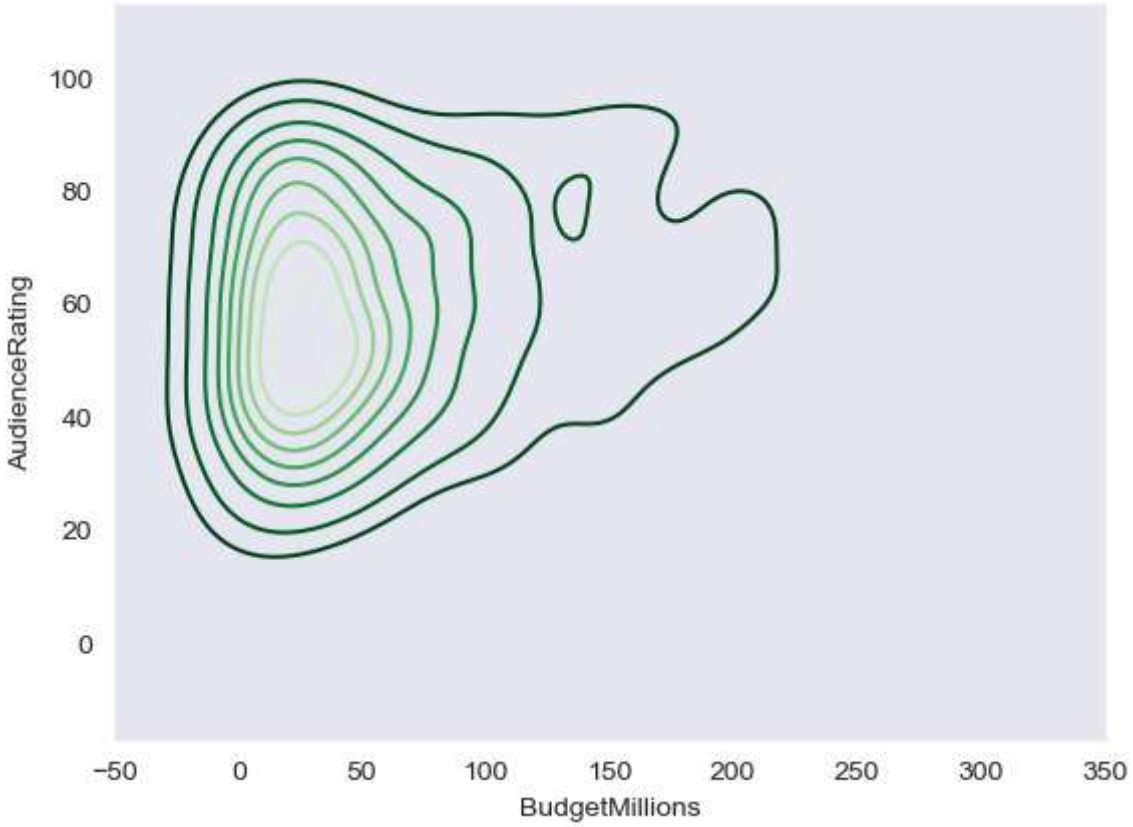
In [52]:
```
k1 = sns.kdeplot(data=movies,x='CriticRating',y='AudienceRating',shade = True,shade
plt.show()
```



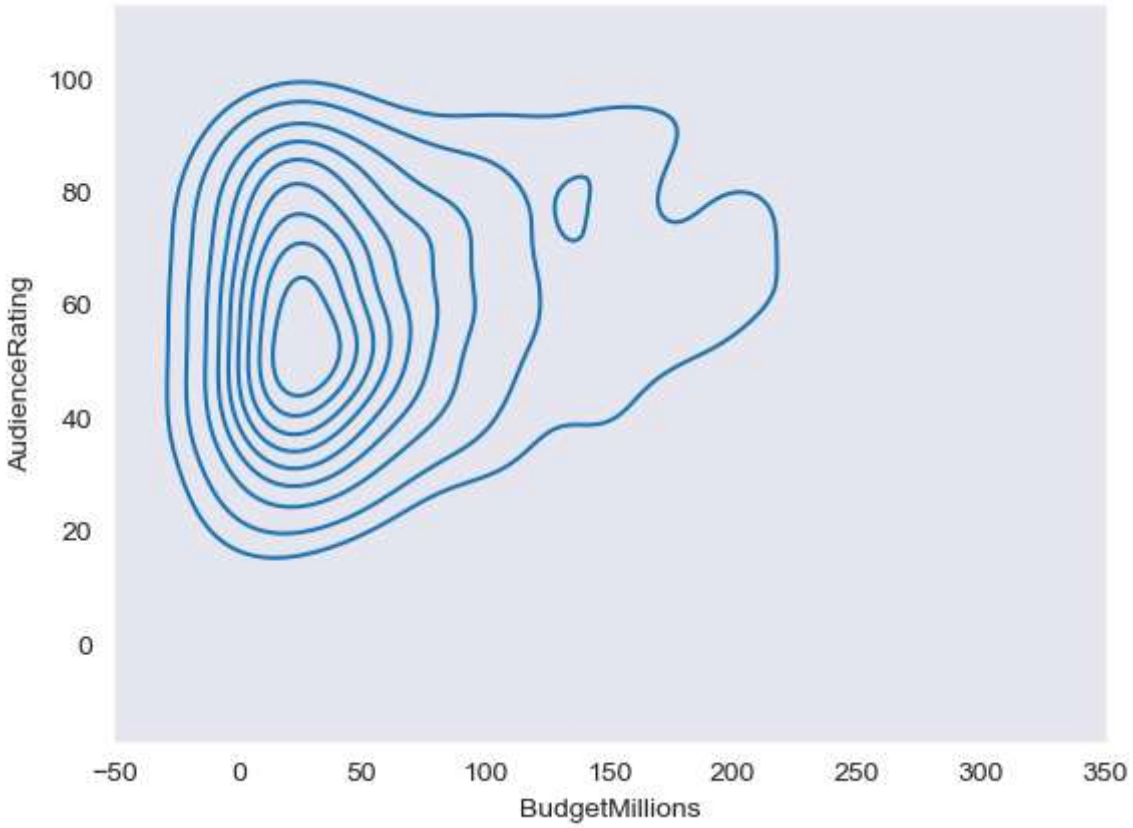
In [53]:
```
k2 = sns.kdeplot(data=movies,x='CriticRating',y='AudienceRating',shade_lowest=False
plt.show()
```



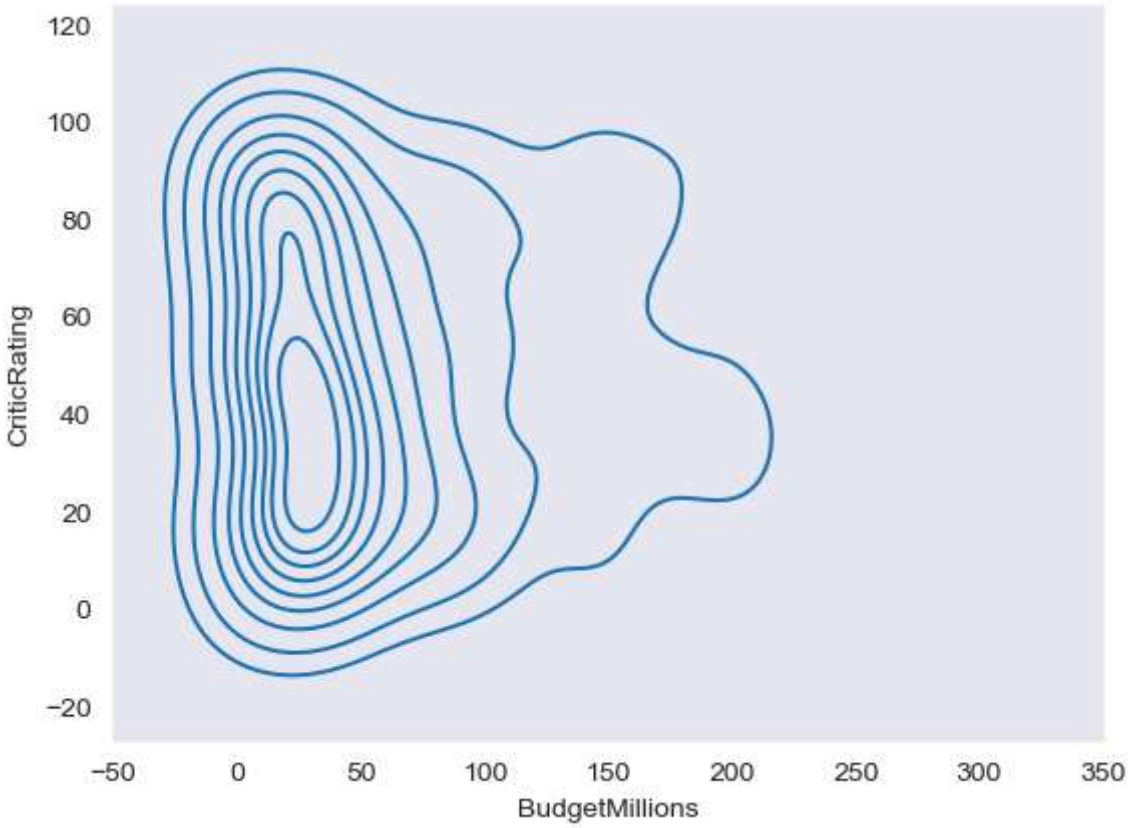
In [54]:
```
sns.set_style('dark')
k1 = sns.kdeplot(data=movies,x='BudgetMillions',y='AudienceRating',shade_lowest=Fal
plt.show()
```



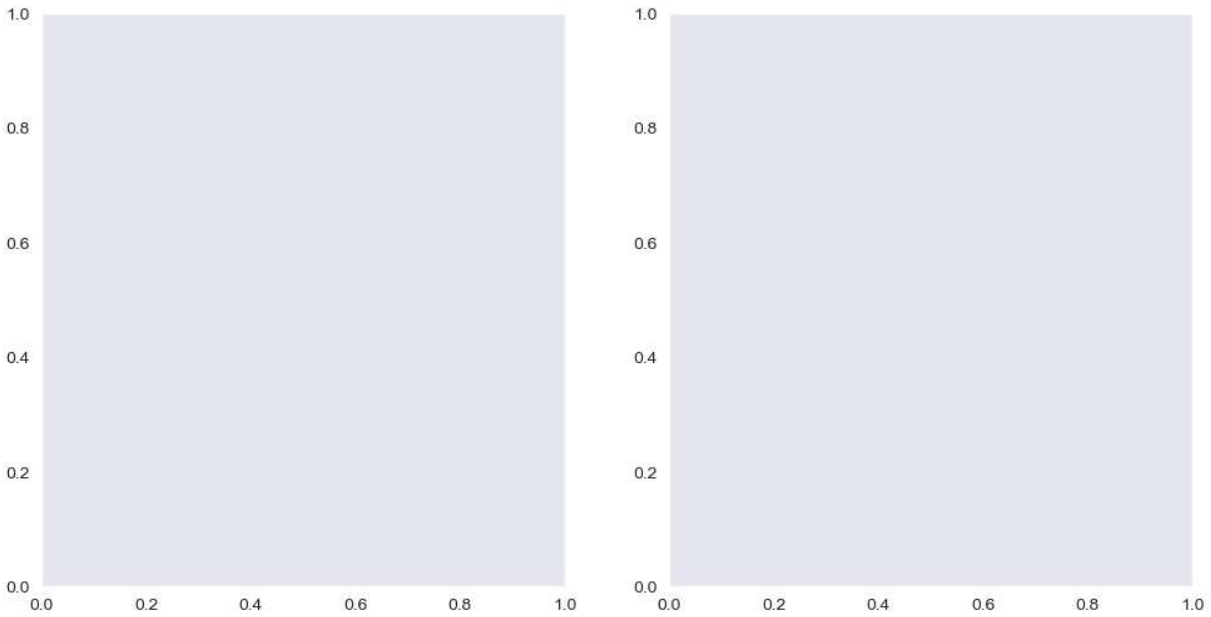
In [55]:
```
sns.set_style('dark')
k1 = sns.kdeplot(data=movies,x='BudgetMillions',y='AudienceRating')
plt.show()
```



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



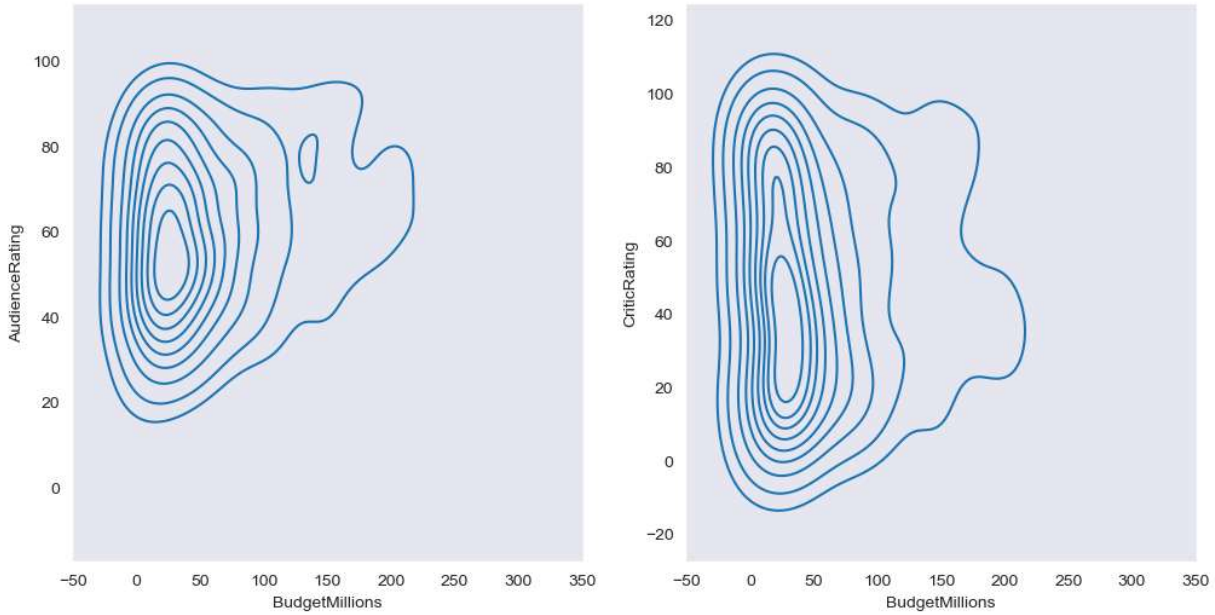
In [57]:
```
#subplots
f, ax = plt.subplots(1,2, figsize =(12,6))    #f, ax = plt.subplots(3,3, figsize =(
plt.show()
```



In [58]:
```
f, axes = plt.subplots(1,2, figsize =(12,6))
```

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

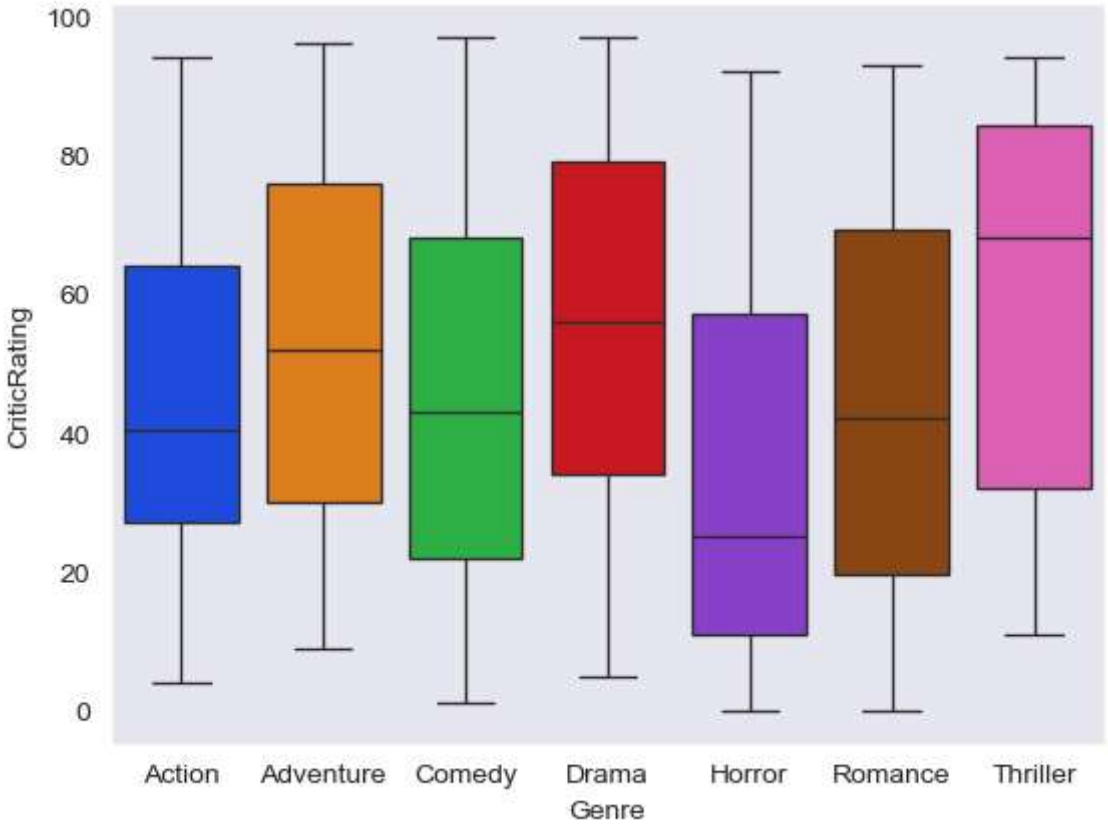
In [59]:
```
plt.show()
```



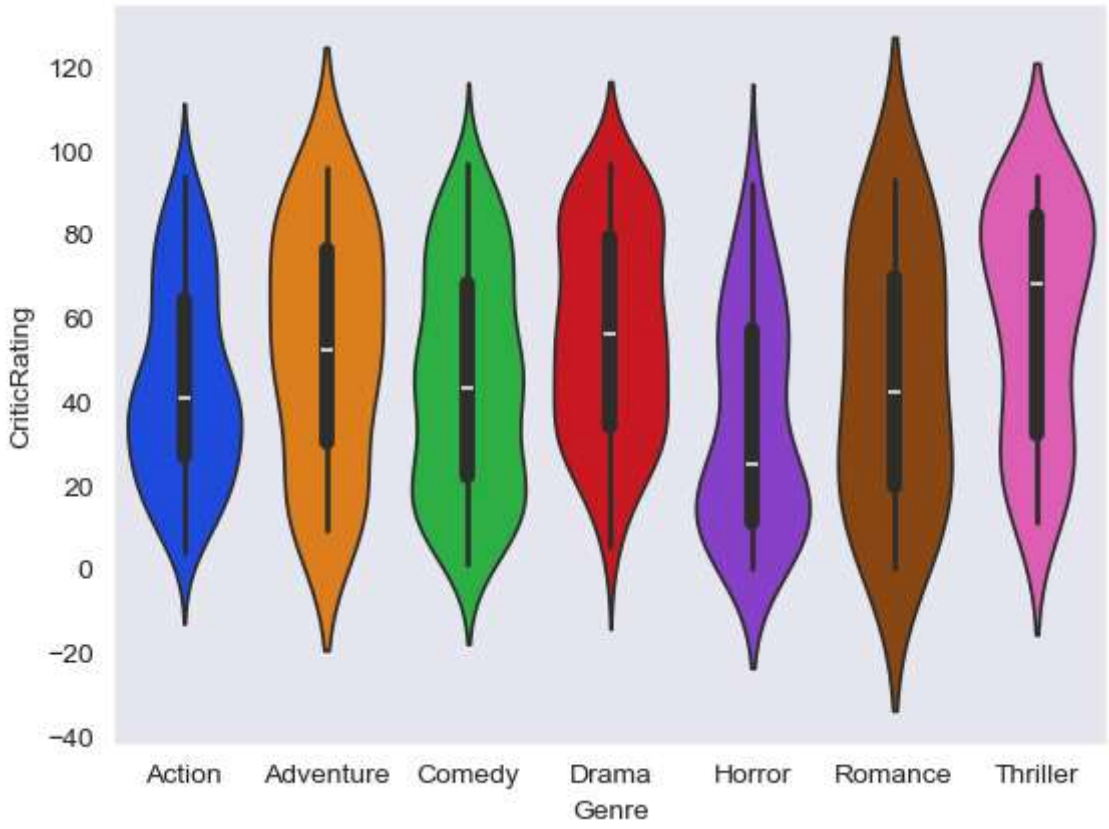
In [60]:
```
axes
```

Out[60]:
```
array([<Axes: xlabel='BudgetMillions', ylabel='AudienceRating'>,
       <Axes: xlabel='BudgetMillions', ylabel='CriticRating'>],
      dtype=object)
```

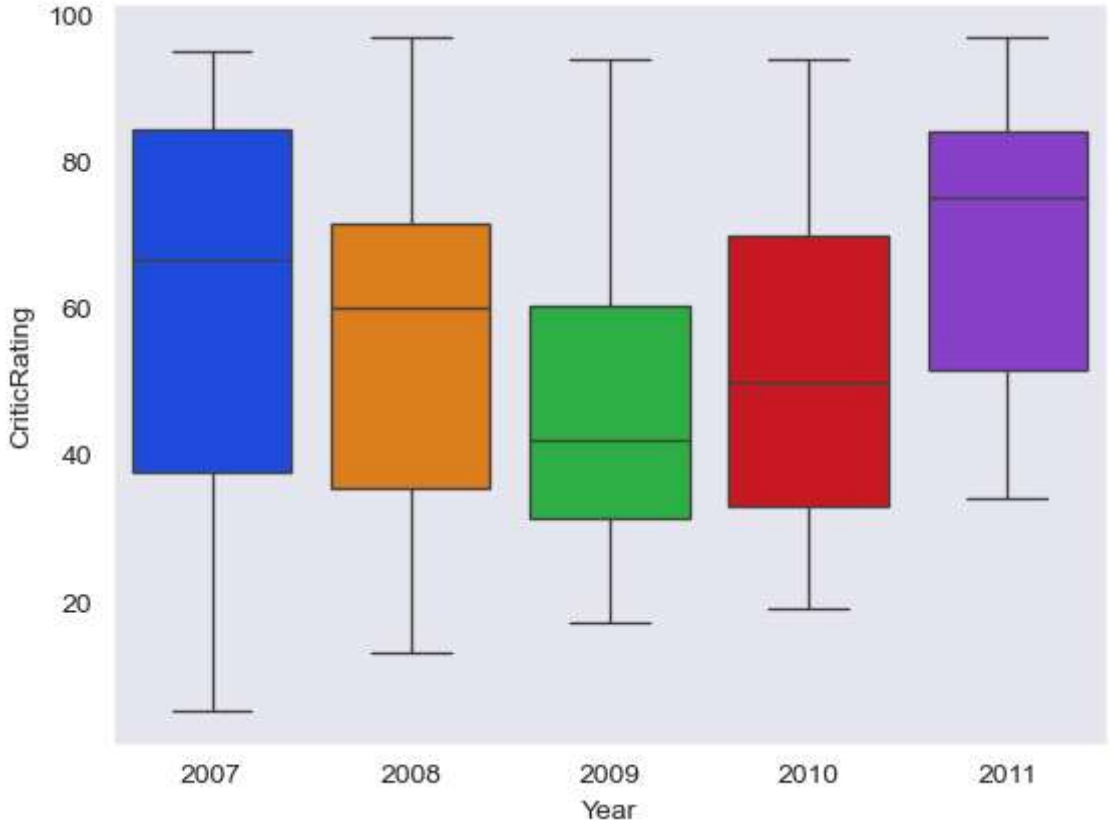
In [61]:
```
#Box plots -

w = sns.boxplot(data=movies, x='Genre', y = 'CriticRating',palette='bright')
plt.show()
```



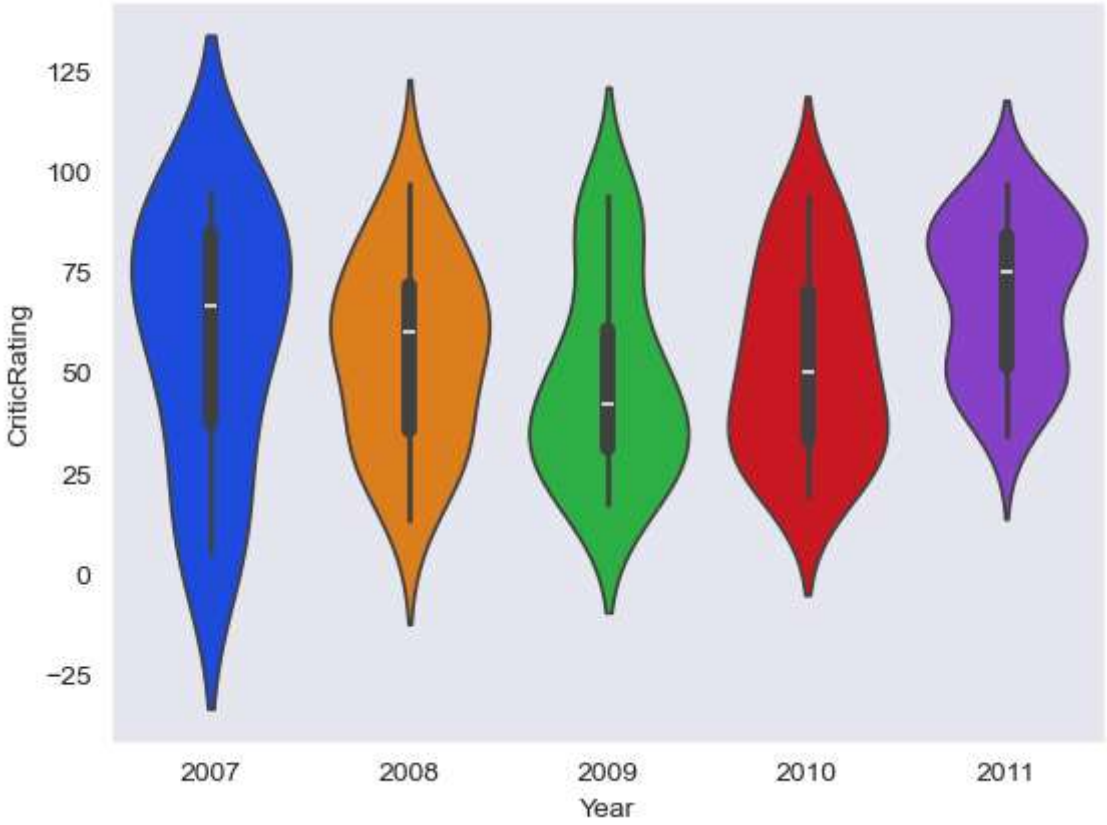
In [62]:
```
#violin plot

z = sns.violinplot(data=movies, x='Genre', y = 'CriticRating',palette='bright')
plt.show()
```



In [63]:
```
w1 = sns.boxplot(data=movies[movies.Genre == 'Drama'], x='Year', y = 'CriticRating'
plt.show()
```

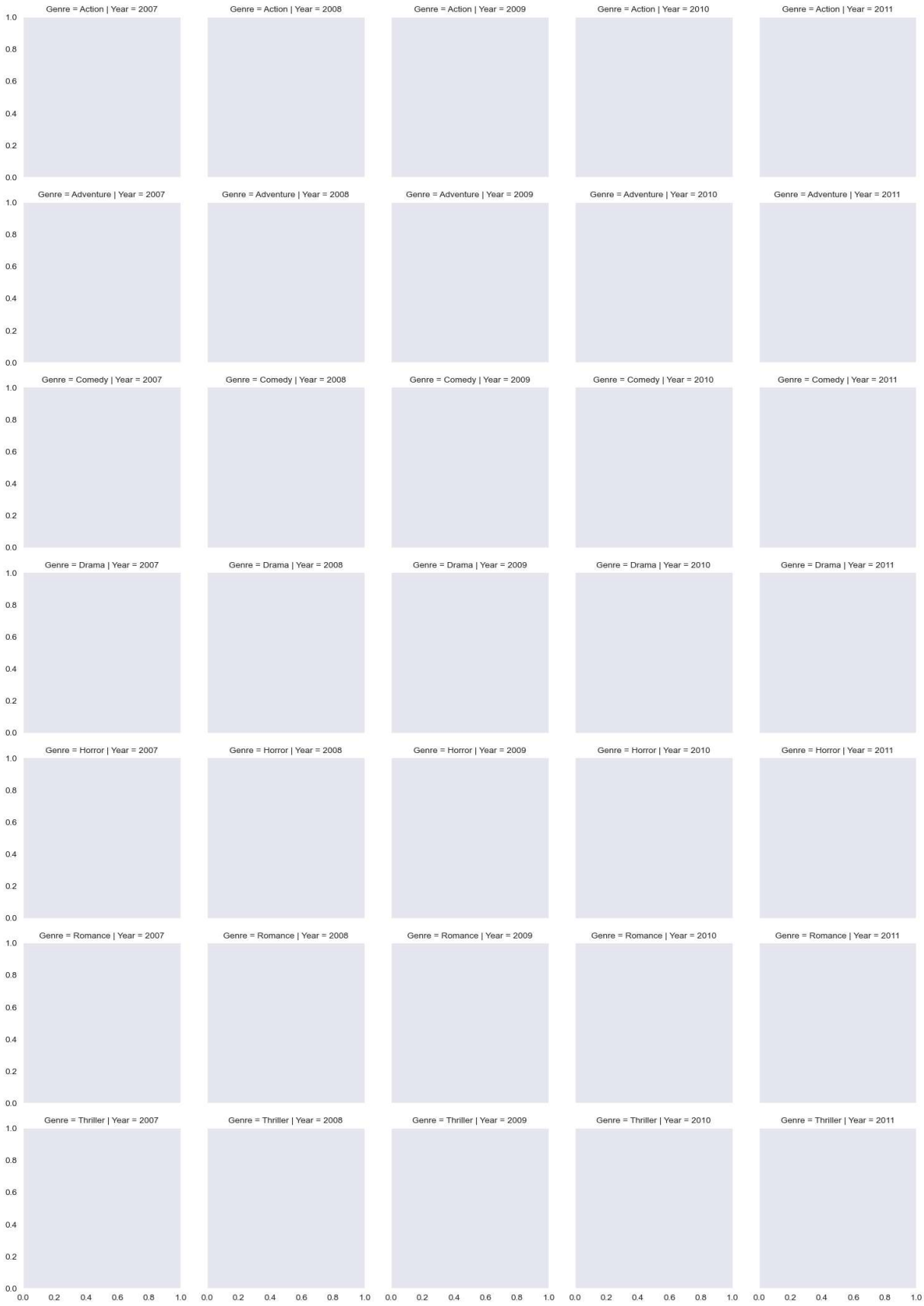


In [64]:
```
z = sns.violinplot(data=movies[movies.Genre == 'Drama'], x='Year', y = 'CriticRatin
plt.show()
```

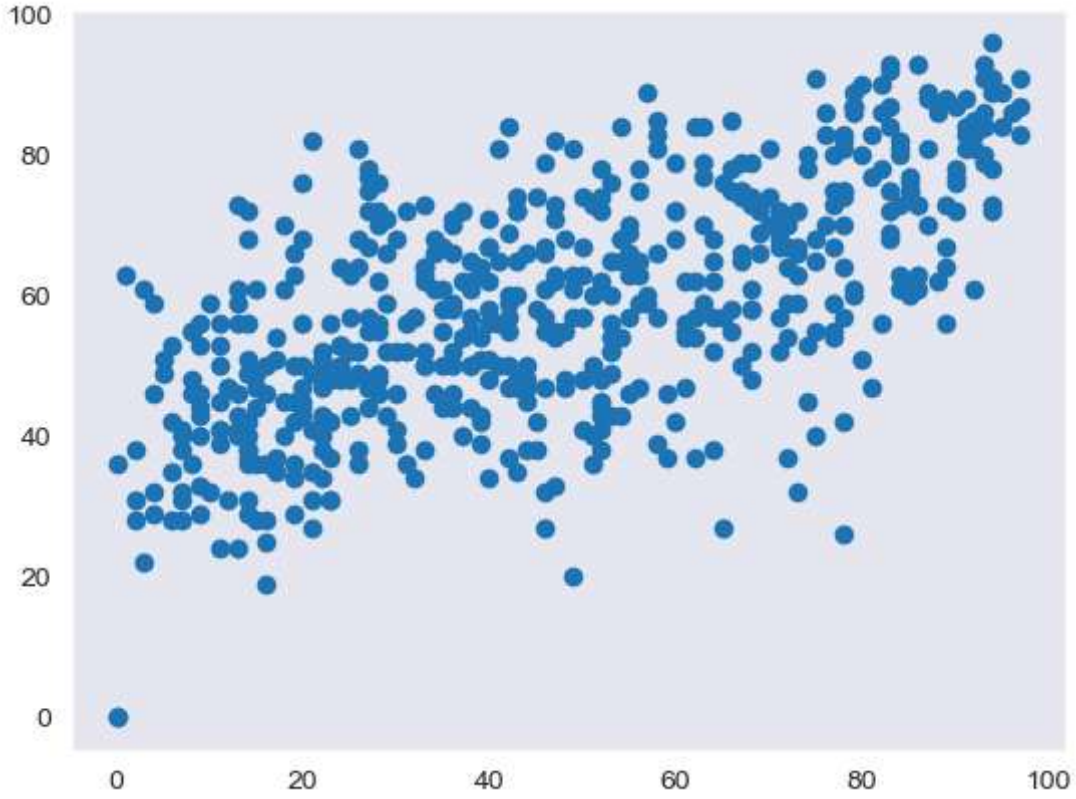


In [65]:
```
# Createing a Facet grid
```

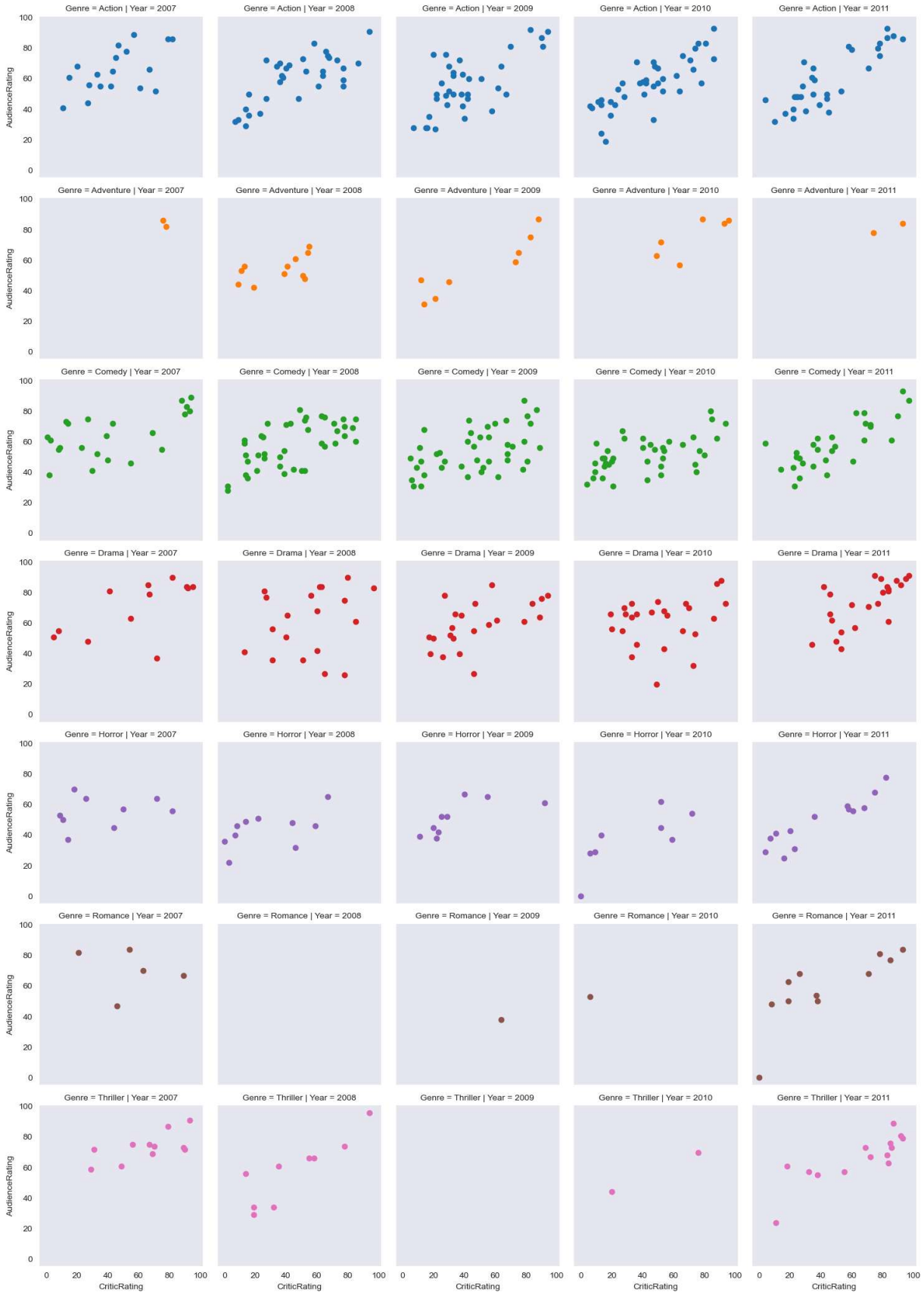
In [66]:
```
g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre') #kind of subp
plt.show()
```



In [67]:
```
plt.scatter(movies.CriticRating,movies.AudienceRating)
plt.show()
```



In [68]:
```
g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
g = g.map(plt.scatter, 'CriticRating', 'AudienceRating' ) #scatterplots are mapped
plt.show()
```

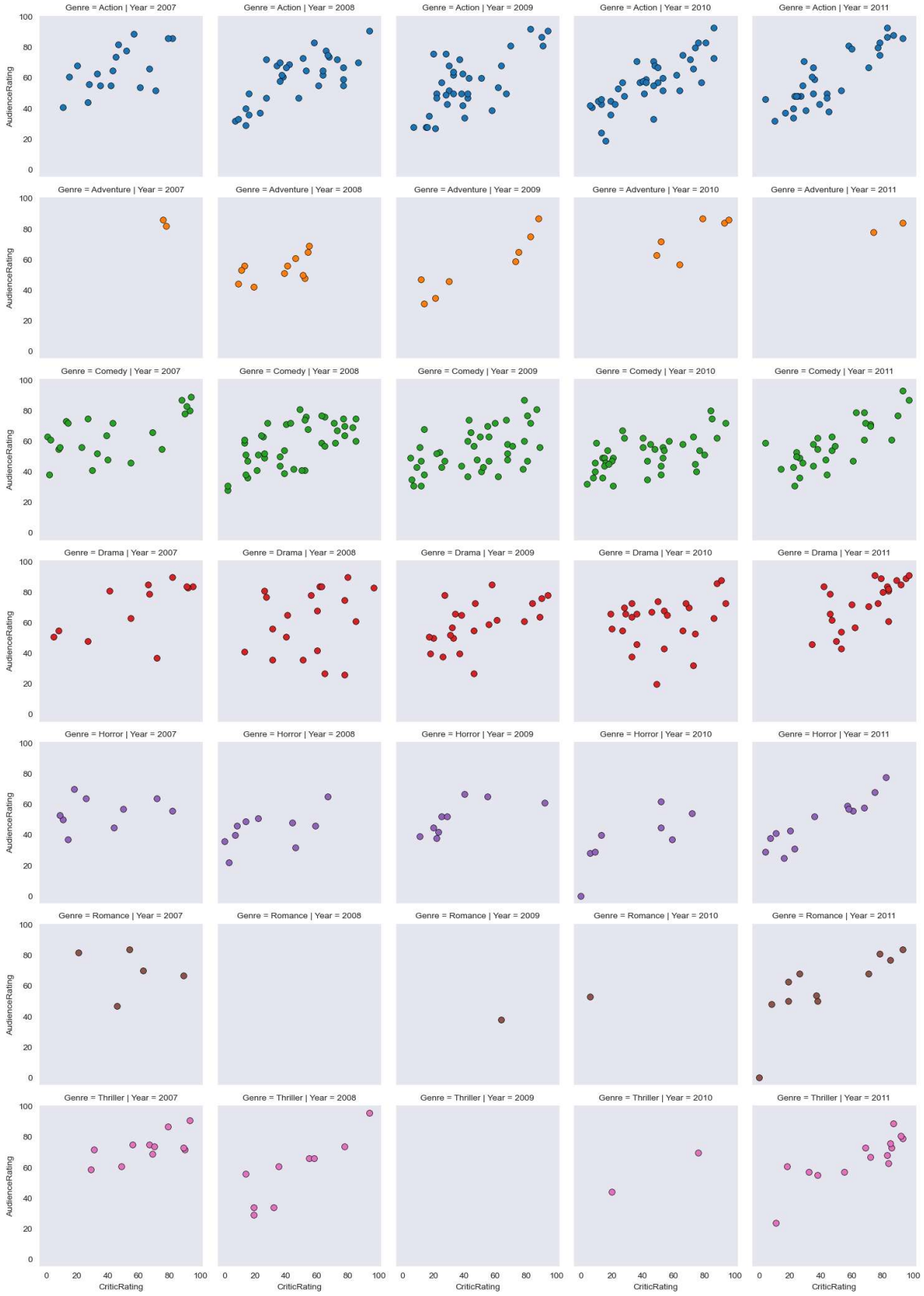


In [69]:
```
# you can populated any type of chat.

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

```
g = g.map(plt.hist, 'BudgetMillions') #scatterplots are mapped in facetgrid
plt.show()
```

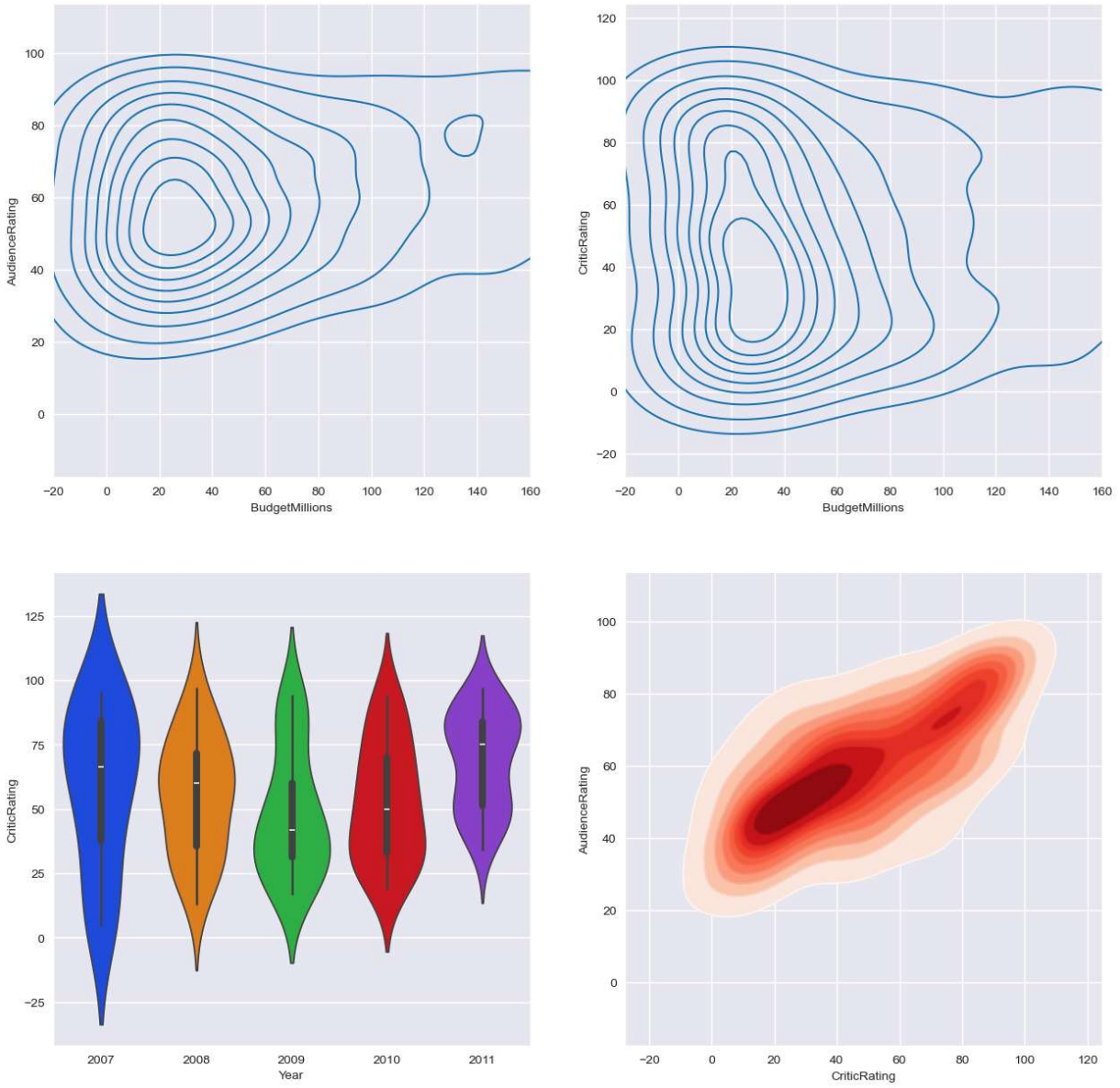
In [70]:
```
g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
kws = dict(s=50, linewidth=0.5,edgecolor='black')
```

```
g = g.map(plt.scatter, 'CriticRating', 'AudienceRating',**kws ) #scatterplots are m
plt.show()
```



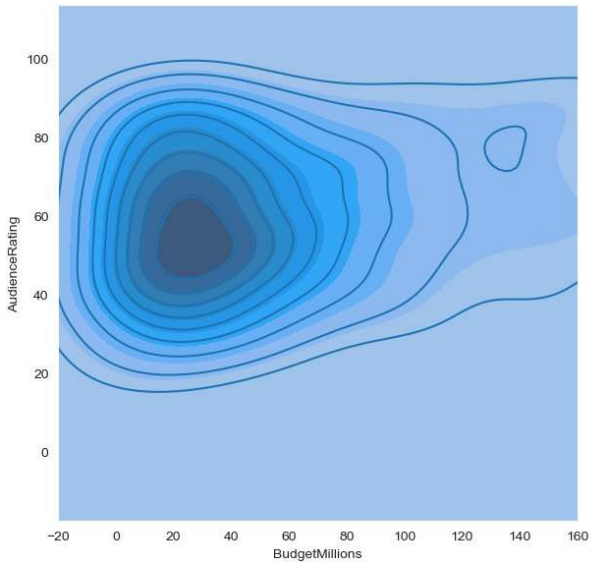
In [71]:
```
# python is not vectorize programming language
# Building dashboards (dashboard - combination of chats)
```

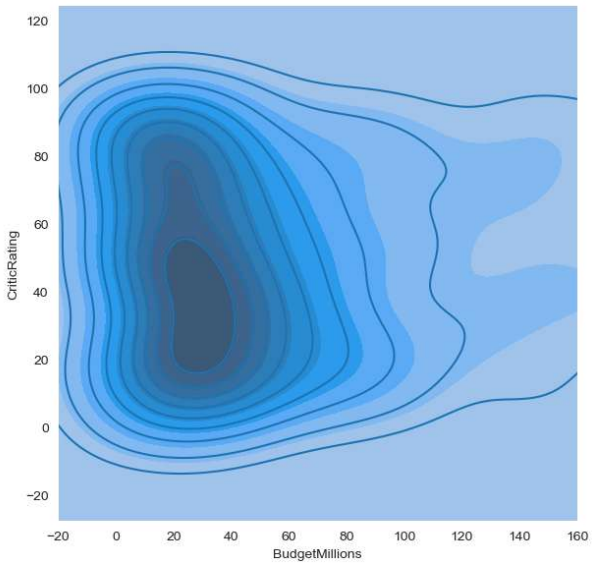
```
sns.set_style('darkgrid')
f, axes = plt.subplots (2,2, figsize = (15,15))

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

k4b = sns.kdeplot(data=movies,x='CriticRating',y='AudienceRating',cmap='Reds',ax =

plt.show()
```

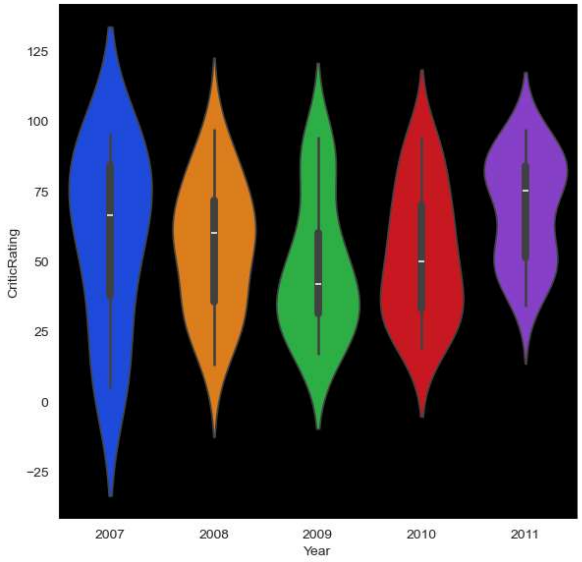


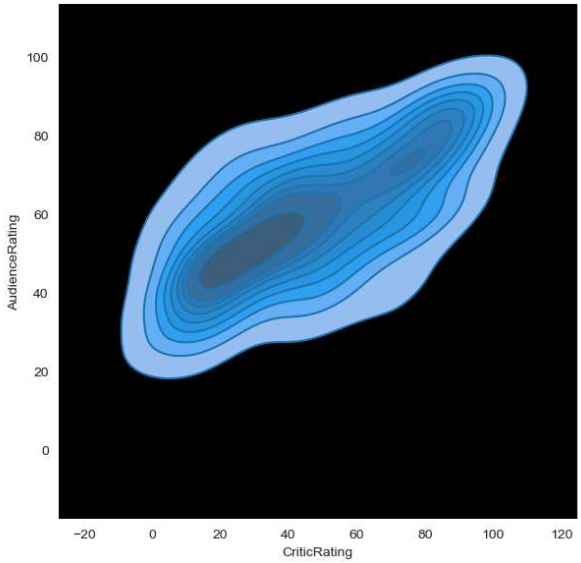
In [85]:
```
# How can you style your dashboard  using different color map
```

```
# python is not vectorize programming language
# Building dashboards (dashboard - combination of chats)

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

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

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

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

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

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


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

plt.show()
```


AudienceRating
BudgetMillions


CriticRating
BudgetMillions


CriticRating
Year


AudienceRating
CriticRating

In [ ]:

In [ ]:

In [ ]: