```
# Importing required package "pandas" to Load data
 In [1]:
          import pandas as pd
         # importing data using pandas
 In [2]:
          movies = pd.read_csv("movie-Rating.csv")
          movies
 Out[2]:
                                                Rotten
                                                           Audience
                                                                         Budget
                                                                                    Year of
                        Film
                                 Genre
                                              Tomatoes
                                                           Ratings %
                                                                      (million $)
                                                                                    release
                                             Ratings %
                 (500) Days of
            0
                                Comedy
                                                    87
                                                                 81
                                                                              8
                                                                                      2009
                     Summer
                   10,000 B.C. Adventure
                                                     9
                                                                  44
                                                                            105
                                                                                      2008
                                                                             20
            2
                   12 Rounds
                                                                                      2009
                                 Action
                                                    30
                                                                 52
            3
                   127 Hours Adventure
                                                                             18
                                                                                      2010
                                                    93
                                                                 84
                                                                                      2009
            4
                    17 Again
                                Comedy
                                                    55
                                                                 70
                                                                             20
          554
                Your Highness
                                                    26
                                                                             50
                                                                                      2011
                                Comedy
                                                                  36
               Youth in Revolt
          555
                                                                             18
                                                                                      2009
                                Comedy
                                                    68
                                                                  52
                      Zodiac
          556
                                 Thriller
                                                    89
                                                                             65
                                                                                      2007
                                                                 73
                  Zombieland
          557
                                 Action
                                                    90
                                                                 87
                                                                             24
                                                                                      2009
          558
                   Zookeeper
                                                                 42
                                                                             80
                                                                                      2011
                                Comedy
                                                    14
         559 rows × 6 columns
In [57]: movies.columns
Out[57]: Index(['Film', 'Genre', 'Rotten Tomatoes Ratings %', 'Audience Ratings %',
                  'Budget (million $)', 'Year of release'],
                dtype='object')
In [58]: 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
             Genre
         1
                                          559 non-null
                                                           object
         2
             Rotten Tomatoes Ratings % 559 non-null
                                                           int64
                                                           int64
             Audience Ratings %
                                          559 non-null
             Budget (million $)
                                          559 non-null
                                                           int64
             Year of release
                                          559 non-null
                                                           int64
        dtypes: int64(4), object(2)
        memory usage: 26.3+ KB
```

In [59]: type(movies)

Out[59]: pandas.core.frame.DataFrame

In [60]: len(movies)

Out[60]: 559

In [61]: movies.isnull()

Out[61]:

	Film	Genre	Rotten Tomatoes Ratings %	Audience Ratings %	Budget (million \$)	Year of release
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
•••						
554	False	False	False	False	False	False
555	False	False	False	False	False	False
556	False	False	False	False	False	False
557	False	False	False	False	False	False
558	False	False	False	False	False	False

559 rows × 6 columns

```
In [62]: movies.isnull().sum()
```

Out[62]: Film 0
Genre 0
Rotten Tomatoes Ratings % 0
Audience Ratings % 0
Budget (million \$) 0
Year of release 0
dtype: int64

In [63]: movies.describe()

Out[63]:		Rotten Tomatoes Ratings %	Audience Ratin	gs Budget (million % \$)	Year of release
	count	559.000000	559.0000	559.000000	559.000000
	mean	47.309481	58.7441	86 50.236136	2009.152057
	std	26.413091	16.8268	87 48.731817	1.362632
	min	0.000000	0.0000	0.000000	2007.000000
	25%	25.000000	47.0000	20.00000	2008.000000
	50%	46.000000	58.0000	35.000000	2009.000000
	75%	70.000000	72.0000	00 65.000000	2010.000000
	max	97.000000	96.0000	300.00000	2011.000000
In [64]:	movies.dty	pes			
Out[64]:	Film Genre Rotten Tom Audience R	natoes Ratings % natings %	object object int64 int64		

```
In [65]: movies['Genre'][5:10]
```

Budget (million \$)

Year of release

dtype: object

Out[65]: 5 Action 6 Comedy 7 Horror 8 Comedy 9 Comedy

Name: Genre, dtype: object

int64

int64

In [75]: movies

Out[75]:	[75]:	Film	Genre	CriticsRating	AudienceRa	

	Film	Genre	CriticsRating	AudienceRatings	Budget	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 [76]: movies.Genre = movies.Genre.astype('category')
         movies.Film = movies.Film.astype('category')
         movies.Year = movies.Year.astype('category')
         movies
```

Out[76]:

	Film	Genre	CriticsRating	AudienceRatings	Budget	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

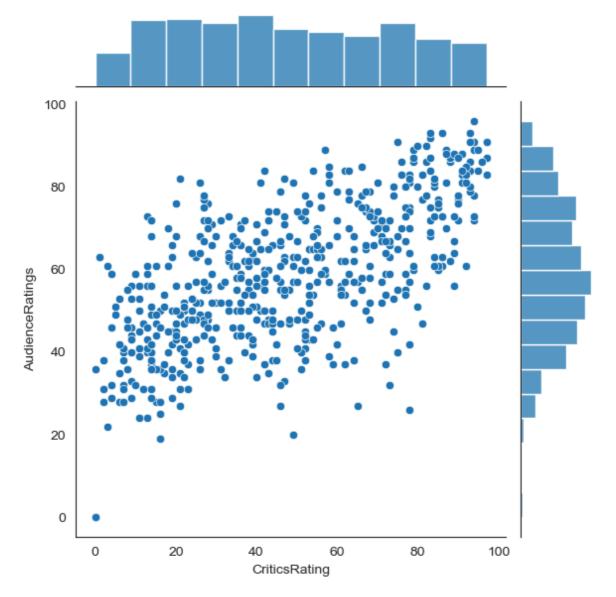
dtype='object')

```
In [77]: movies.Genre.cat.categories
Out[77]: Index(['Action', 'Adventure', 'Comedy', 'Drama', 'Horror', 'Romance',
                 'Thriller'],
```

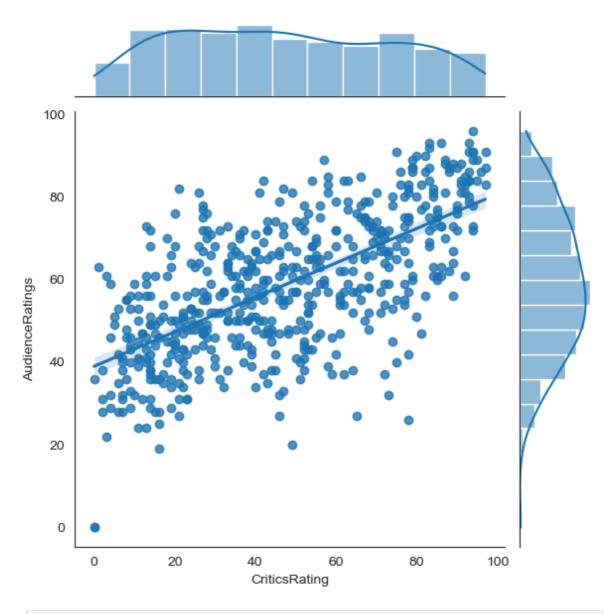
```
In [78]: import numpy as np
   import matplotlib.pyplot as plt
   import seaborn as sns

In [79]: import warnings
   warnings.filterwarnings('ignore')

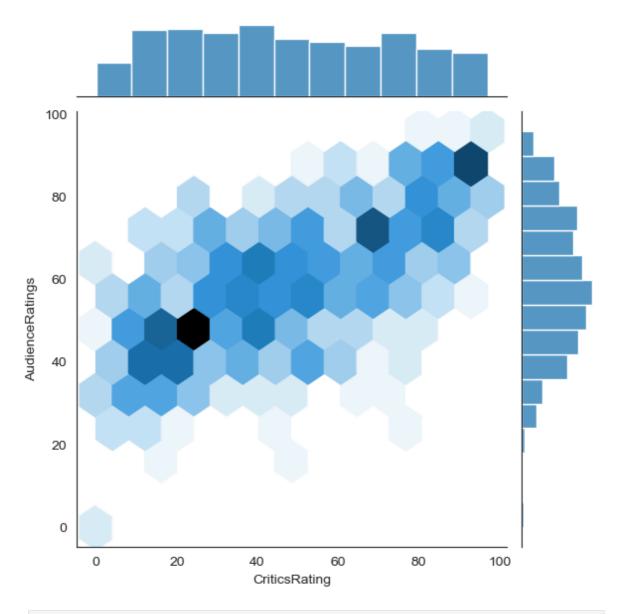
In [80]: j = sns.jointplot(data = movies, y='AudienceRatings', x = 'CriticsRating')
```



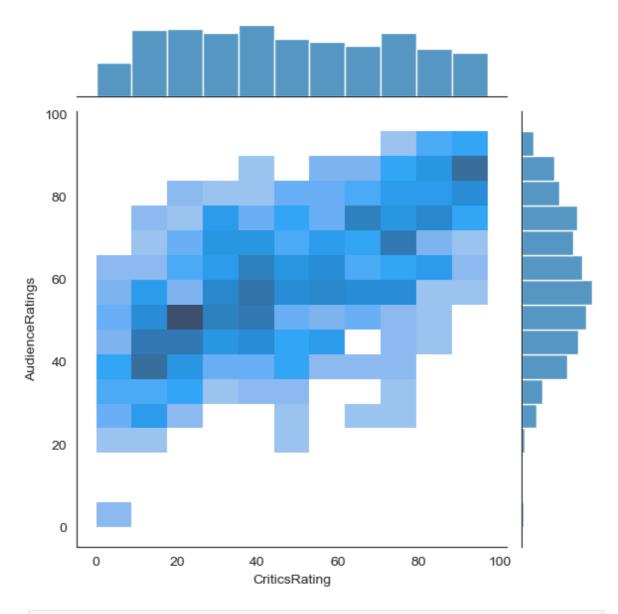
In [81]: j = sns.jointplot(data = movies, y='AudienceRatings', x = 'CriticsRating', kind



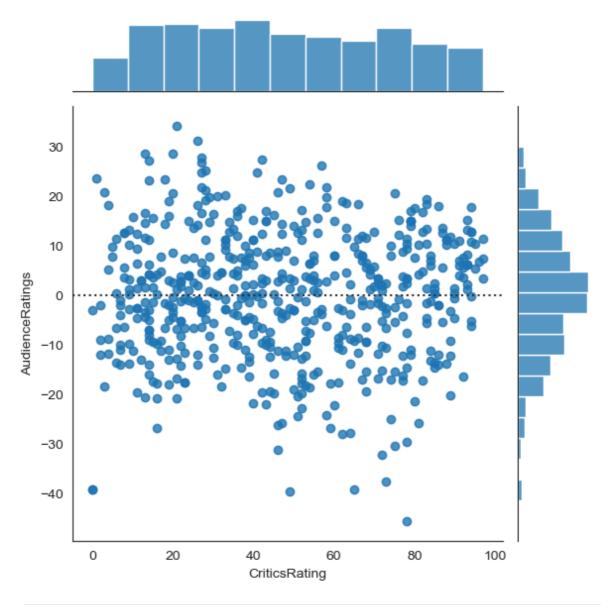
In [82]: j = sns.jointplot(data = movies, y='AudienceRatings', x = 'CriticsRating', kind



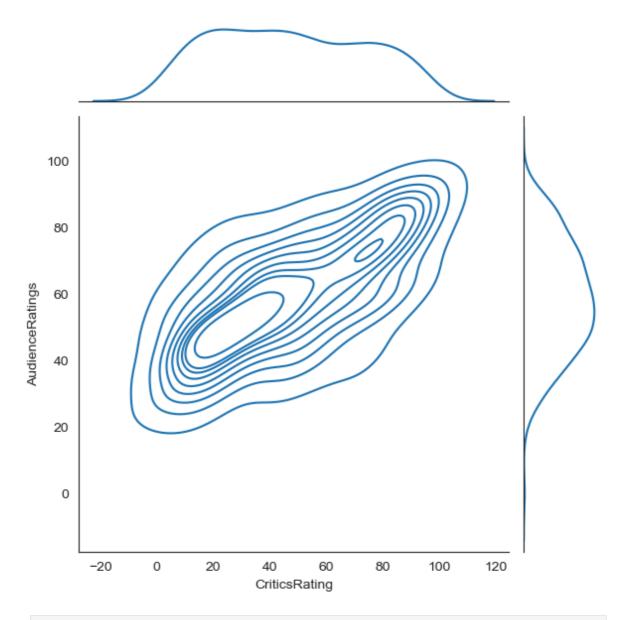
In [83]: j = sns.jointplot(data = movies, y='AudienceRatings', x = 'CriticsRating', kind



In [84]: j = sns.jointplot(data=movies, y='AudienceRatings', x='CriticsRating', kind='res



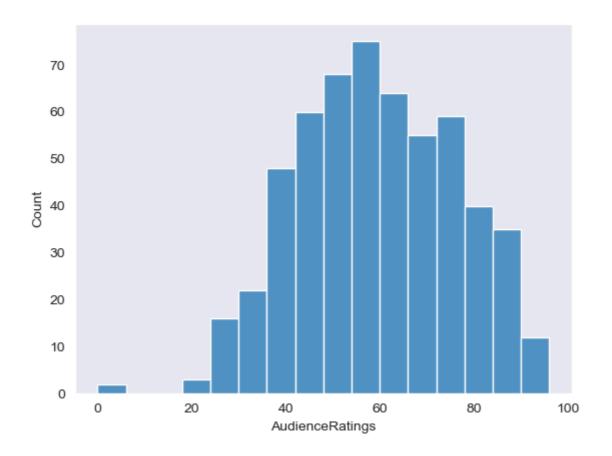
In [85]: j = sns.jointplot(data = movies, y='AudienceRatings', x = 'CriticsRating', kind



In [86]: sns.set_style('dark')

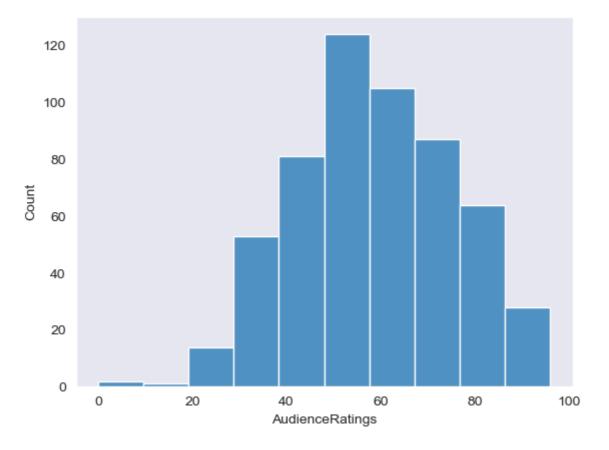
In [87]: sns.histplot(movies.AudienceRatings)

Out[87]: <Axes: xlabel='AudienceRatings', ylabel='Count'>



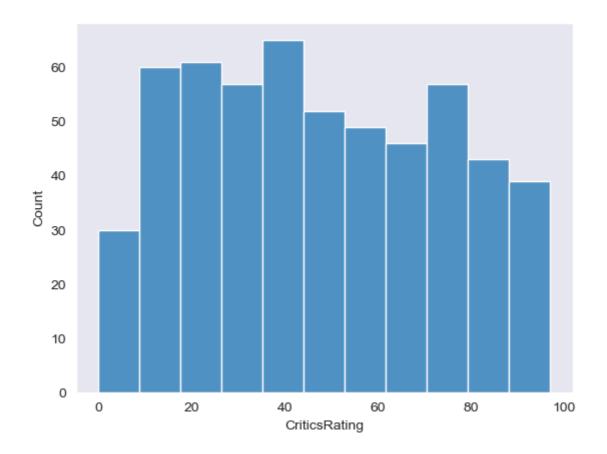
In [88]: sns.histplot(movies.AudienceRatings,bins = 10)

Out[88]: <Axes: xlabel='AudienceRatings', ylabel='Count'>



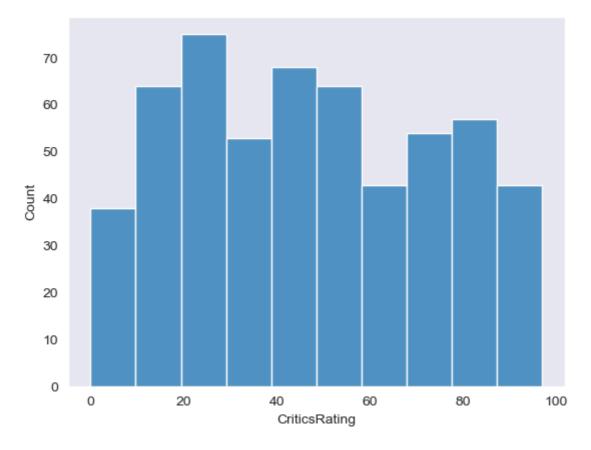
In [89]: sns.histplot(movies.CriticsRating)

Out[89]: <Axes: xlabel='CriticsRating', ylabel='Count'>

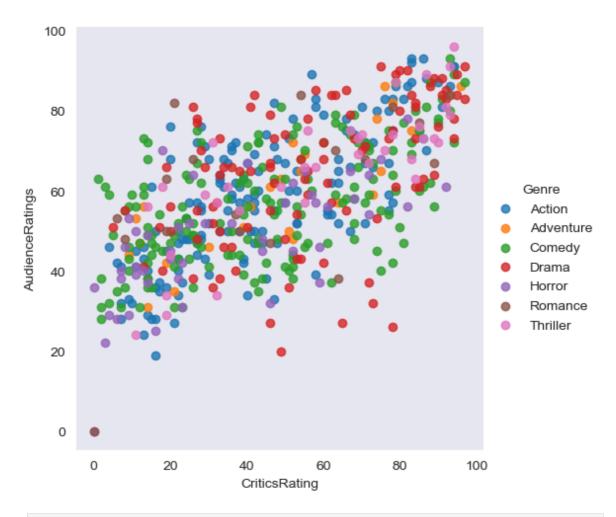


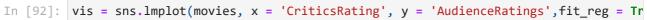
In [90]: sns.histplot(movies.CriticsRating,bins = 10)

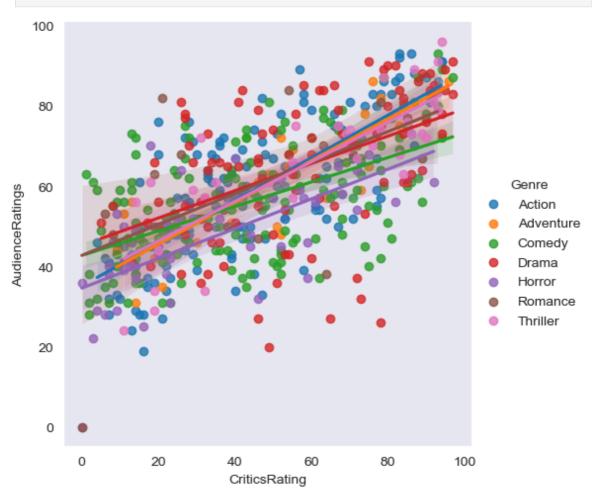
Out[90]: <Axes: xlabel='CriticsRating', ylabel='Count'>

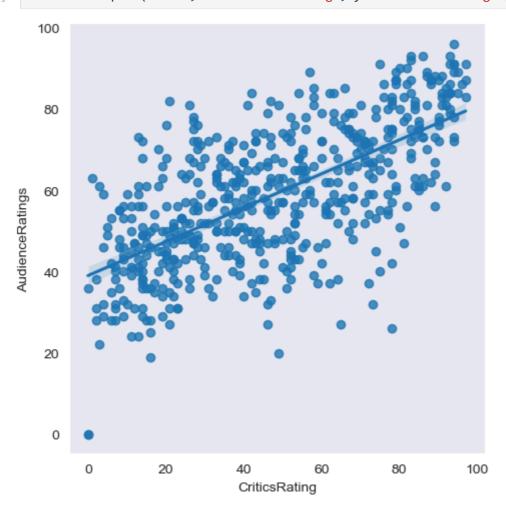


In [91]: vis = sns.lmplot(movies, x = 'CriticsRating', y = 'AudienceRatings',fit_reg = Fa

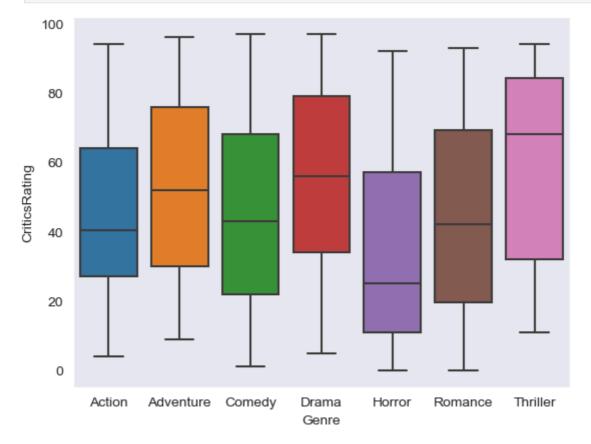




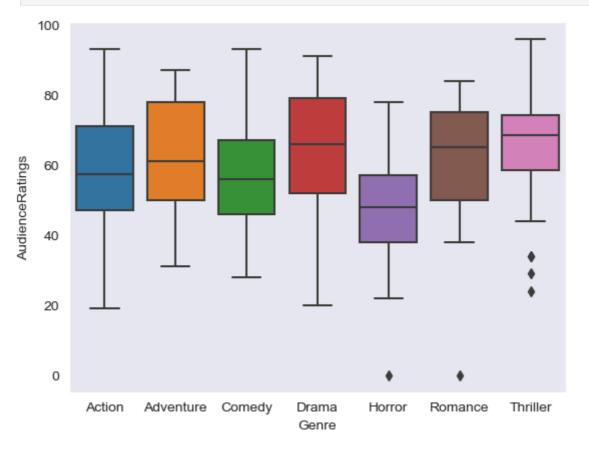




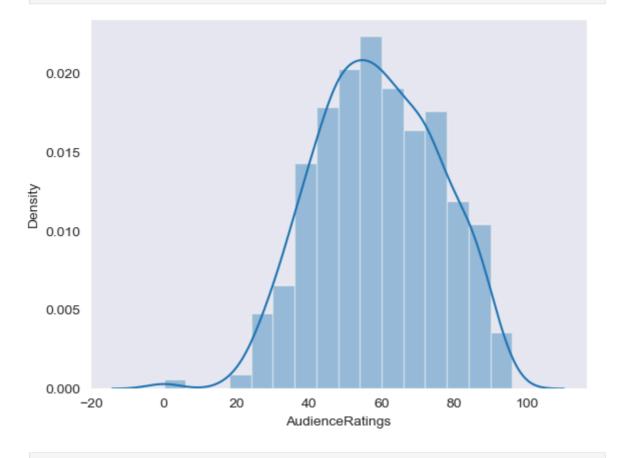
In [94]: ax = sns.boxplot(movies,x ='Genre', y = 'CriticsRating')



In [95]: ax = sns.boxplot(movies,x ='Genre', y = 'AudienceRatings')



In [96]: m1= sns.distplot(movies.AudienceRatings)



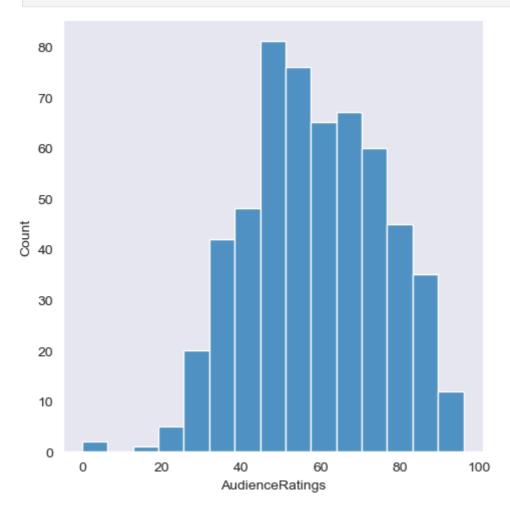
In [97]: movies

0 1	$\Gamma \cap \neg \neg$	
()))	19/	

	Film	Genre	CriticsRating	AudienceRatings	Budget	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

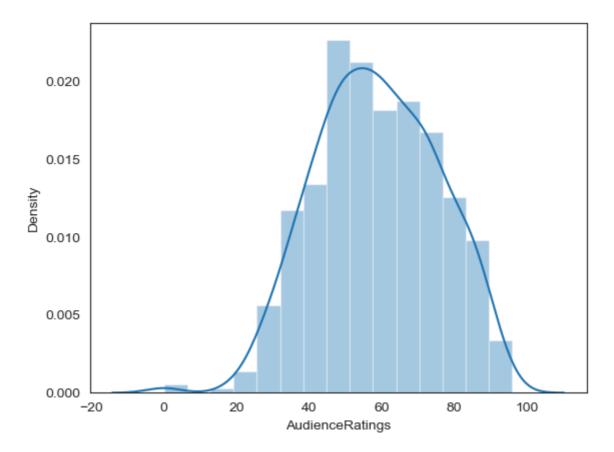
559 rows × 6 columns

In [98]: m2= sns.displot(movies.AudienceRatings,bins=15)

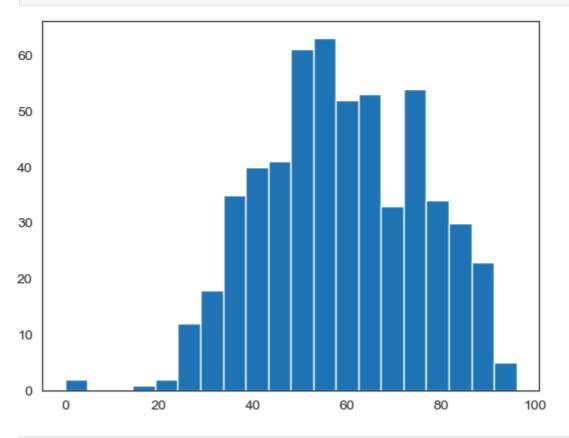


In [99]: sns.set_style('white')

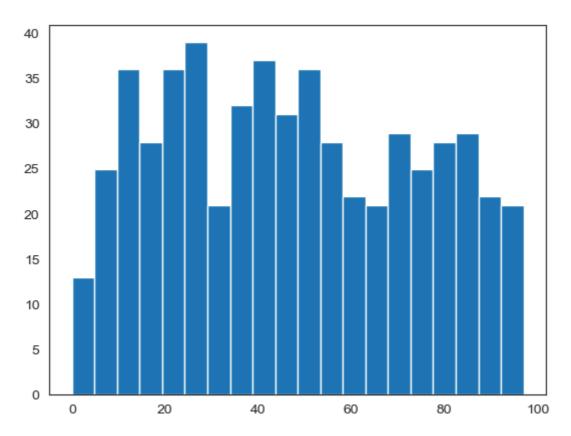
In [100... m2= sns.distplot(movies.AudienceRatings,bins = 15)



In [101... sns.set_style('white')
 n1=plt.hist(movies.AudienceRatings,bins=20) # normal distribution



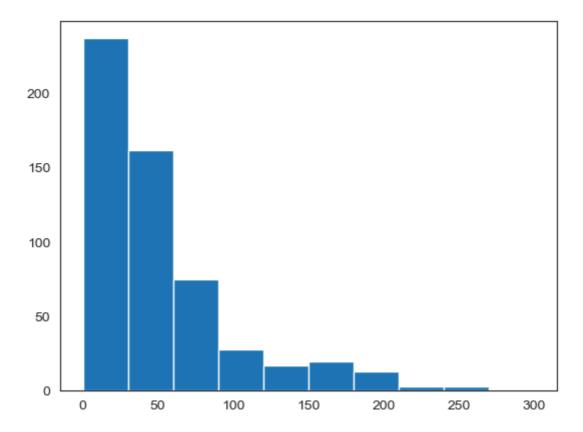
In [102... n1 = plt.hist(movies.CriticsRating, bins= 20) # uniform distribution



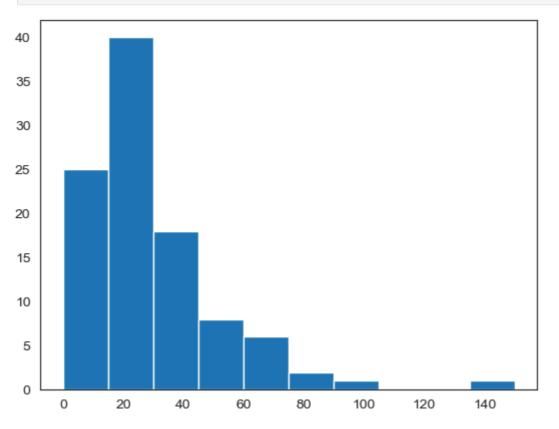
```
movies.Budget
In [103...
Out[103...
            0
                      8
            1
                    105
            2
                     20
            3
                     18
            4
                     20
            554
                     50
            555
                     18
            556
                     65
            557
                     24
            558
                     80
```

In [104... plt.hist(movies.Budget)
 plt.show()

Name: Budget, Length: 559, dtype: int64



In [105... plt.hist(movies[movies.Genre=='Drama'].Budget)
 plt.show()



In [106... movies.head()

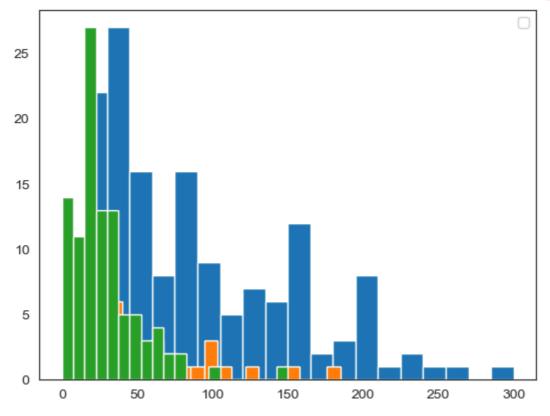
	Film	Genre	CriticsRating	AudienceRatings	Budget	Year
0	(500) Days of Summer	Comedy	87	81	8	2009
1	10,000 B.C.	Adventure	9	44	105	2008
2	12 Rounds	Action	30	52	20	2009
3	127 Hours	Adventure	93	84	18	2010
4	17 Again	Comedy	55	70	20	2009

```
In [107... movies.Genre.unique()
```

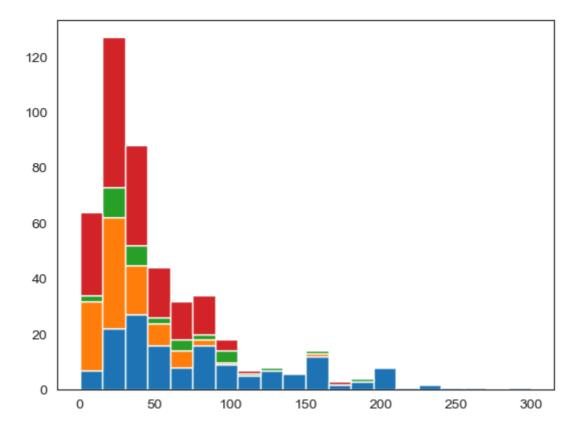
Out[107... ['Comedy', 'Adventure', 'Action', 'Horror', 'Drama', 'Romance', 'Thriller']
Categories (7, object): ['Action', 'Adventure', 'Comedy', 'Drama', 'Horror', 'R
omance', 'Thriller']

```
In [109... plt.hist(movies[movies.Genre=='Action'].Budget,bins=20)
    plt.hist(movies[movies.Genre=='Thriller'].Budget , bins=20)
    plt.hist(movies[movies.Genre=='Drama'].Budget, 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 ignored when legend() is called with no argument.



```
In [110... plt.hist([movies.Genre=='Action'].Budget,\
    movies[movies.Genre=='Drama'].Budget,\
    movies[movies.Genre=='Thriller'].Budget,\
    movies[movies.Genre=='Comedy'].Budget],bins=20, stacked=True)
plt.show()
```



In [111... for gen in movies.Genre.cat.categories:
 print(gen)

Action

Adventure

Comedy

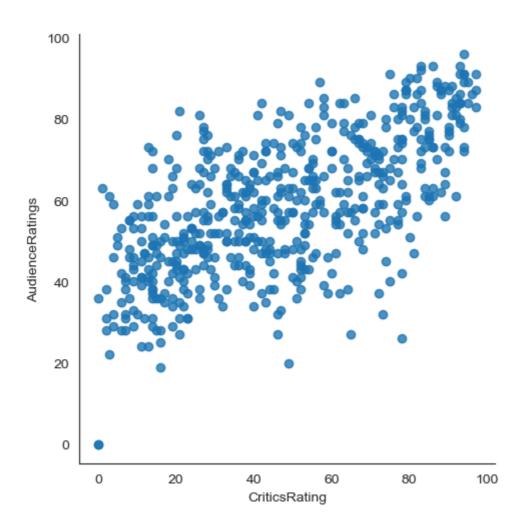
Drama

Horror

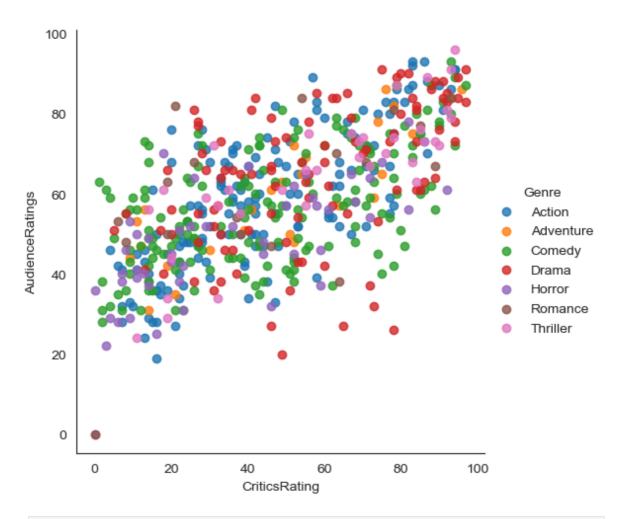
Romance

Thriller

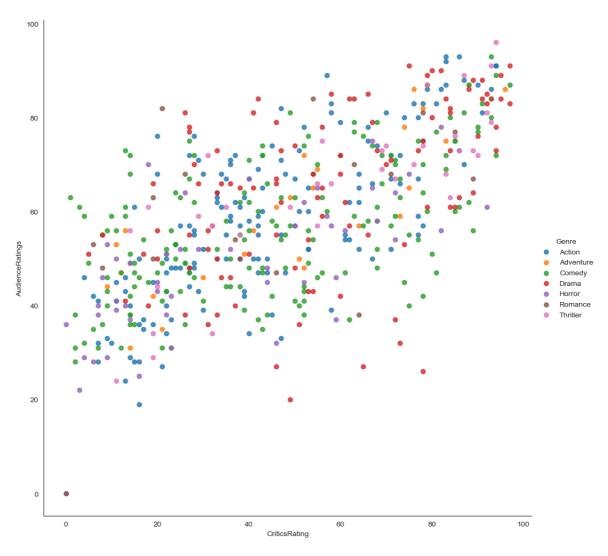
In [113... vis1 = sns.lmplot(data = movies, x = 'CriticsRating' ,y ='AudienceRatings', fit_

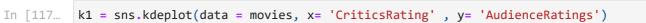


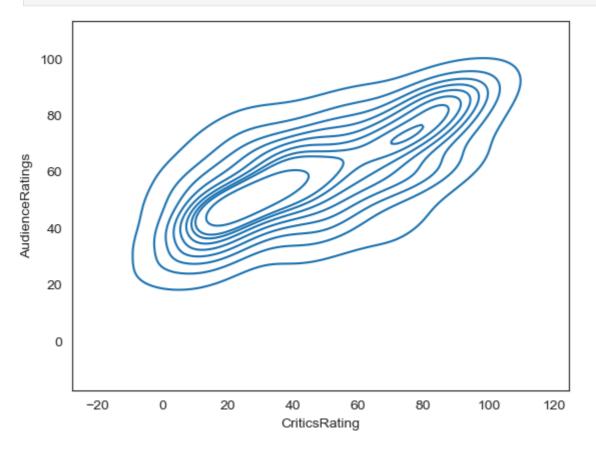
In [114... vis1 = sns.lmplot(data = movies, x = 'CriticsRating' ,y ='AudienceRatings', fit_



In [115... vis1 = sns.lmplot(data = movies, x = 'CriticsRating' ,y ='AudienceRatings', fit_







In [118...

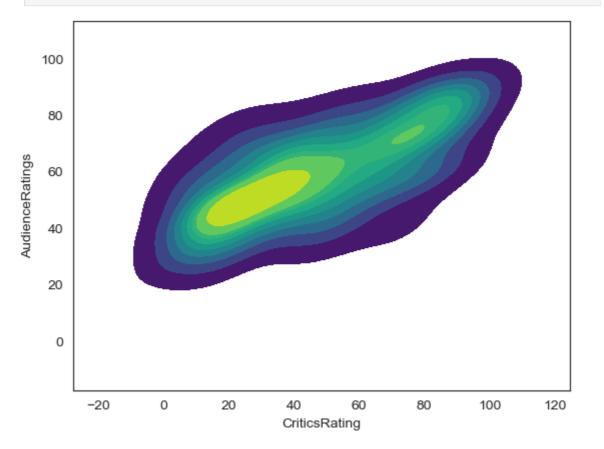
movies

Out[118...

	Film	Genre	CriticsRating	AudienceRatings	Budget	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

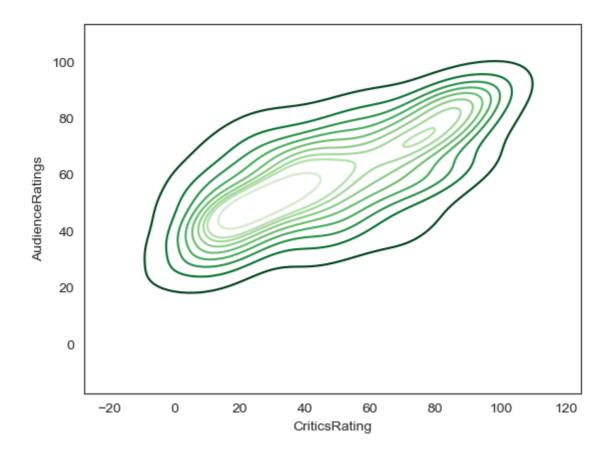
559 rows × 6 columns

In [120... k1 = sns.kdeplot(data = movies, x= 'CriticsRating' , y= 'AudienceRatings', shade=

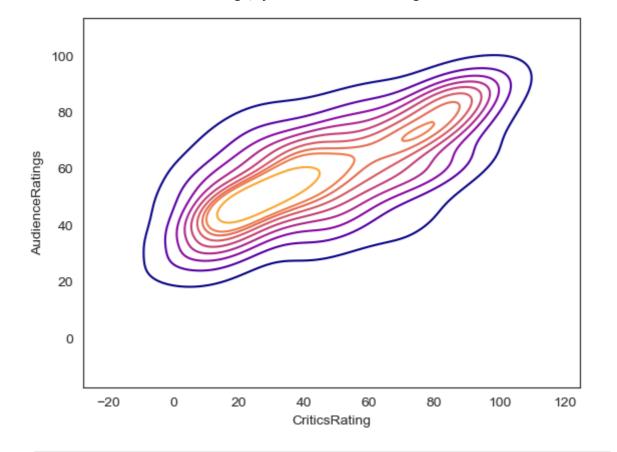


In [123... sns.kdeplot(data = movies, x= 'CriticsRating' , y= 'AudienceRatings', shade_lowes

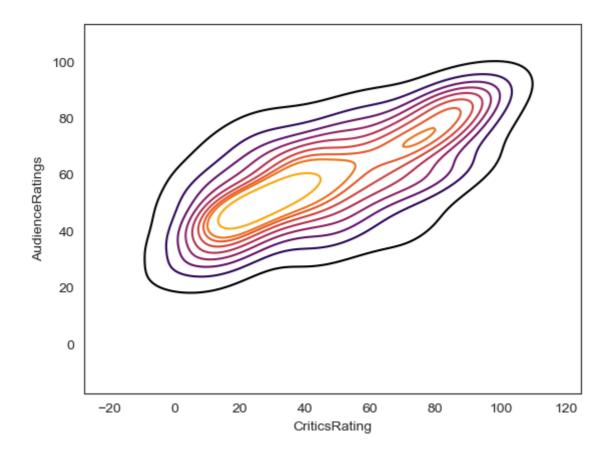
Out[123... <Axes: xlabel='CriticsRating', ylabel='AudienceRatings'>



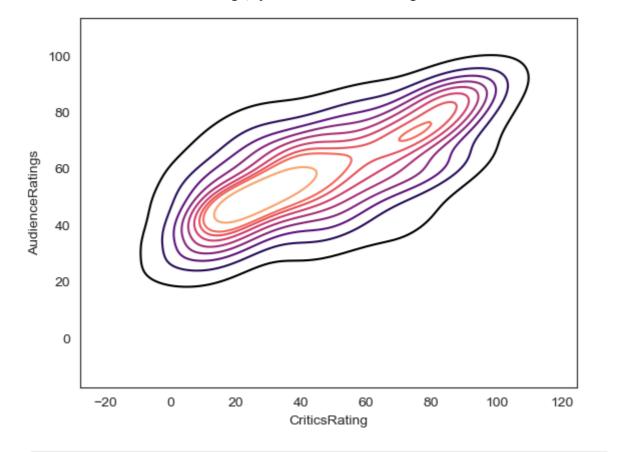
In [124... sns.kdeplot(data = movies, x= 'CriticsRating' , y= 'AudienceRatings', shade_lowes
Out[124... <Axes: xlabel='CriticsRating', ylabel='AudienceRatings'>

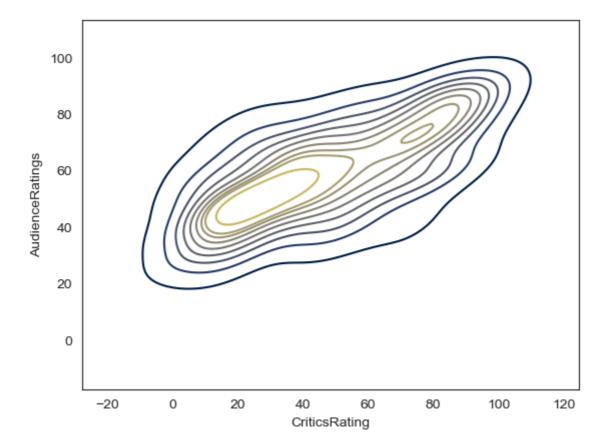


In [125... sns.kdeplot(data = movies, x= 'CriticsRating' , y= 'AudienceRatings', shade_lowes
Out[125... <Axes: xlabel='CriticsRating', ylabel='AudienceRatings'>

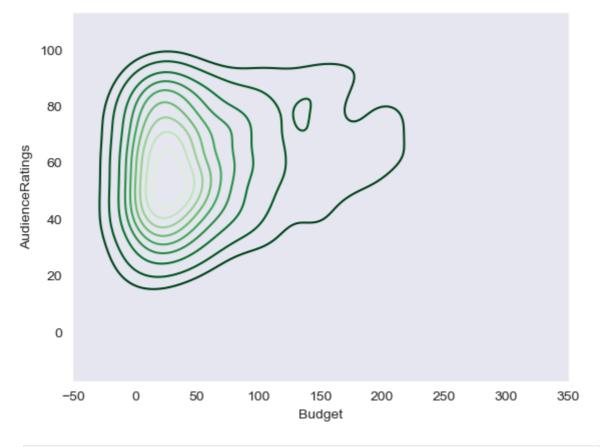


In [126... sns.kdeplot(data = movies, x= 'CriticsRating' , y= 'AudienceRatings', shade_lowes
Out[126... <Axes: xlabel='CriticsRating', ylabel='AudienceRatings'>



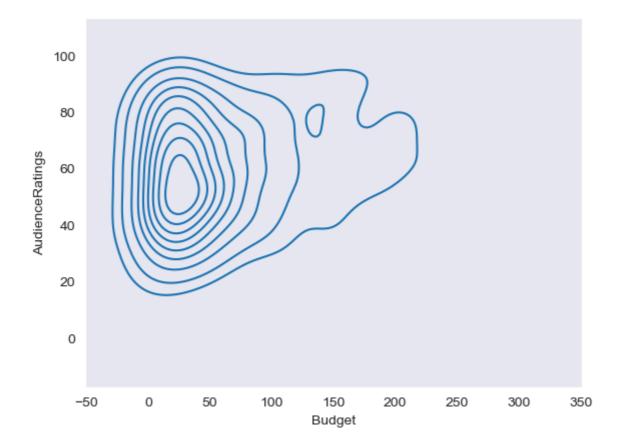


In [128...
sns.set_style('dark')
k1 = sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings',shade_lowest=



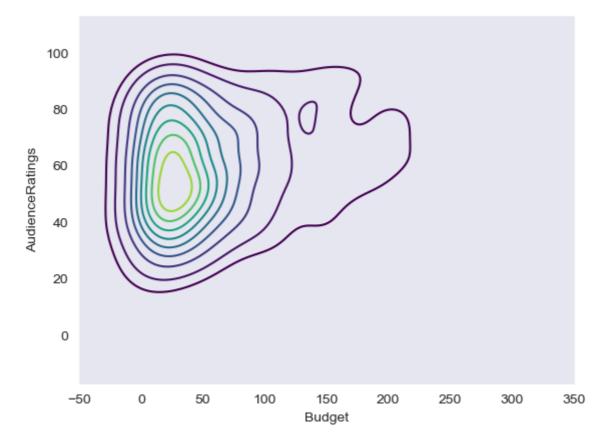
```
In [129... sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings')
```

Out[129... <Axes: xlabel='Budget', ylabel='AudienceRatings'>



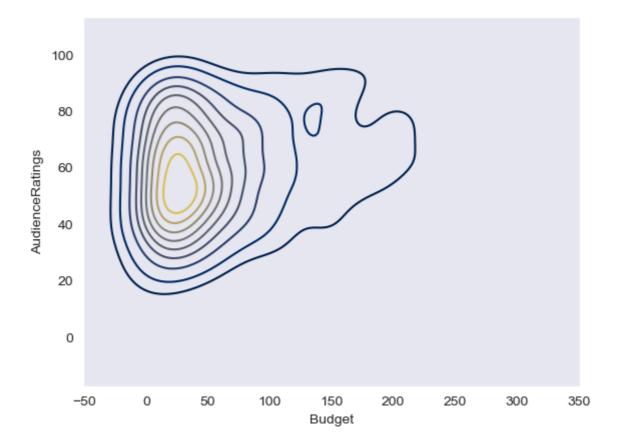
In [130... sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings',cmap='viridis')

Out[130... <Axes: xlabel='Budget', ylabel='AudienceRatings'>



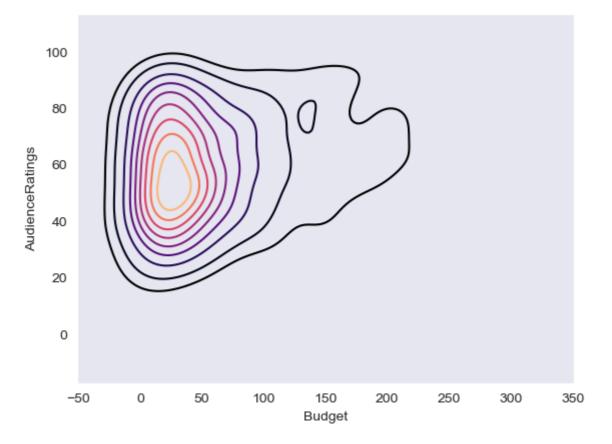
```
In [131... sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings',cmap='cividis')
```

Out[131... <Axes: xlabel='Budget', ylabel='AudienceRatings'>



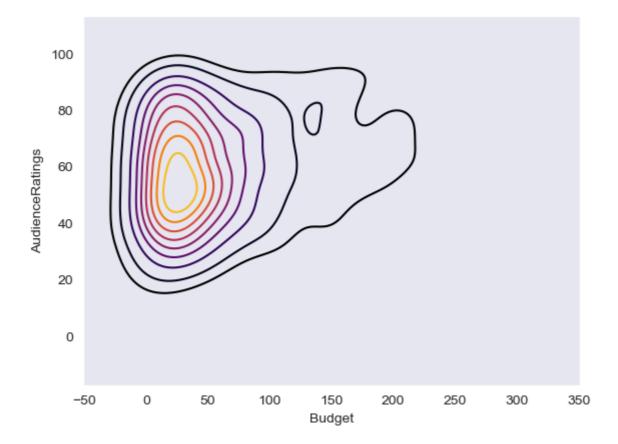
In [132... sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings',cmap='magma')

Out[132... <Axes: xlabel='Budget', ylabel='AudienceRatings'>



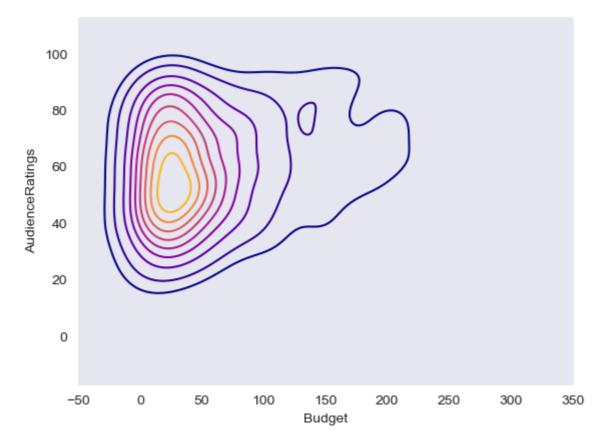
```
In [133... sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings',cmap='inferno')
```

Out[133... <Axes: xlabel='Budget', ylabel='AudienceRatings'>



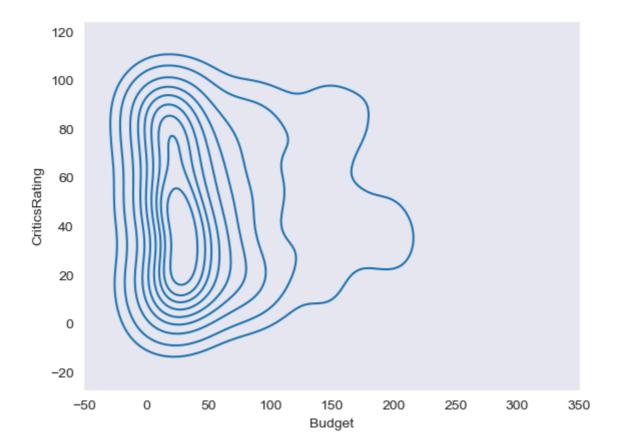
In [141... sns.kdeplot(data = movies, x= 'Budget' , y= 'AudienceRatings',cmap='plasma')

Out[141... <Axes: xlabel='Budget', ylabel='AudienceRatings'>



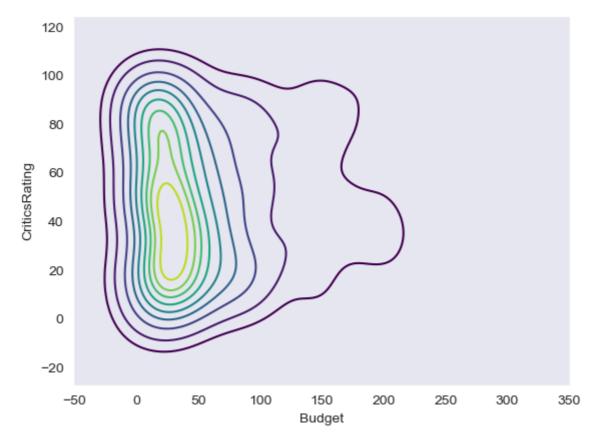
```
In [135... sns.kdeplot(data= movies,x='Budget',y= 'CriticsRating')
```

Out[135... <Axes: xlabel='Budget', ylabel='CriticsRating'>



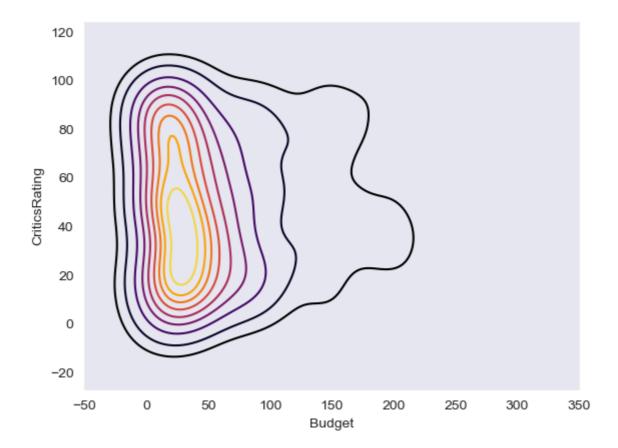
In [136... sns.kdeplot(data= movies,x='Budget',y= 'CriticsRating',cmap='viridis')

Out[136... <Axes: xlabel='Budget', ylabel='CriticsRating'>



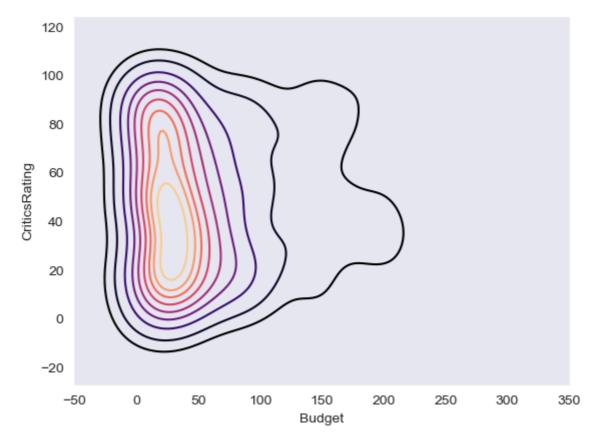
```
In [137... sns.kdeplot(data= movies,x='Budget',y= 'CriticsRating',cmap='inferno')
```

Out[137... <Axes: xlabel='Budget', ylabel='CriticsRating'>



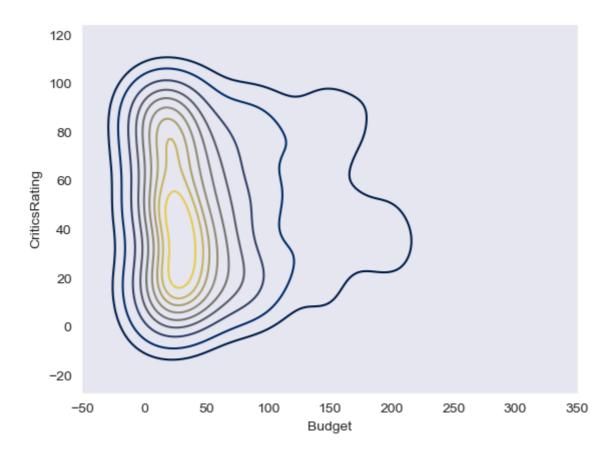
In [138... sns.kdeplot(data= movies,x='Budget',y= 'CriticsRating',cmap='magma')

Out[138... <Axes: xlabel='Budget', ylabel='CriticsRating'>



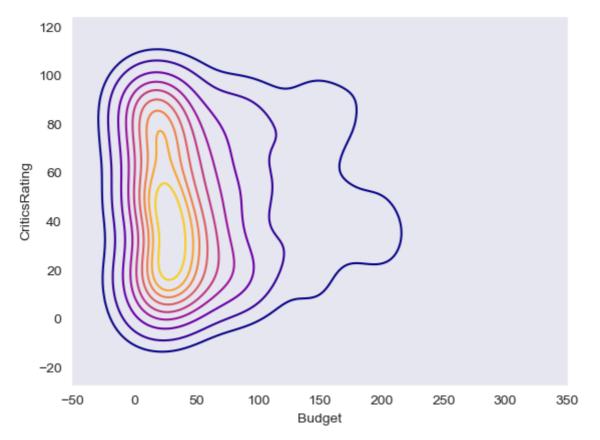
```
In [139... sns.kdeplot(data= movies,x='Budget',y= 'CriticsRating',cmap='cividis')
```

Out[139... <Axes: xlabel='Budget', ylabel='CriticsRating'>

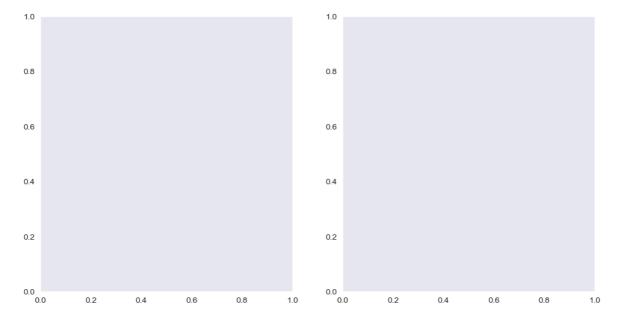


In [140... sns.kdeplot(data= movies,x='Budget',y= 'CriticsRating',cmap='plasma')

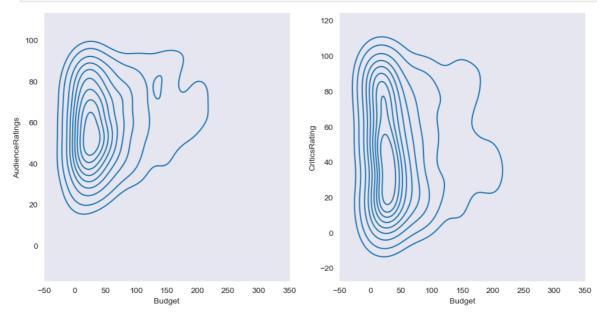
Out[140... <Axes: xlabel='Budget', ylabel='CriticsRating'>



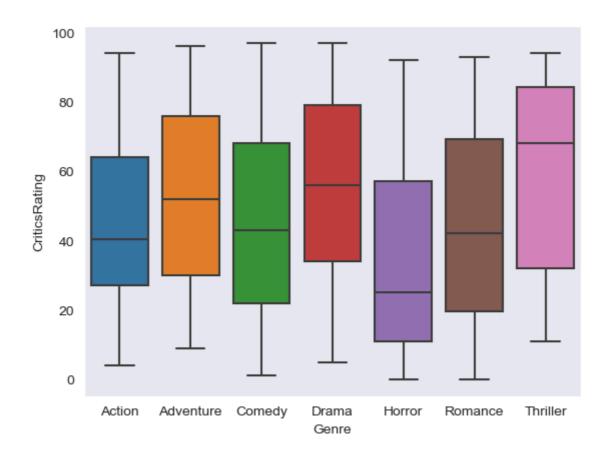
```
In [145... #subplots
f,ax= plt.subplots(1,2,figsize = (12,6))
```



In [146...
f, axes = plt.subplots(1,2, figsize = (12,6))
k1= sns.kdeplot(data=movies, x='Budget',y='AudienceRatings',ax= axes[0])
k1= sns.kdeplot(data=movies, x='Budget',y='CriticsRating',ax= axes[1])

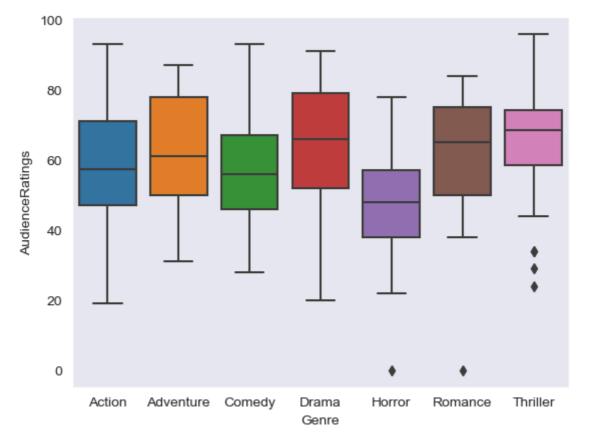


Out[149... <Axes: xlabel='Genre', ylabel='CriticsRating'>

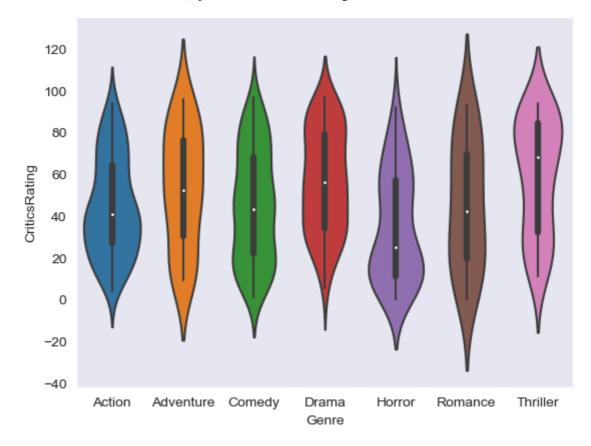


In [150... sns.boxplot(movies,x = 'Genre', y = 'AudienceRatings')

Out[150... <Axes: xlabel='Genre', ylabel='AudienceRatings'>

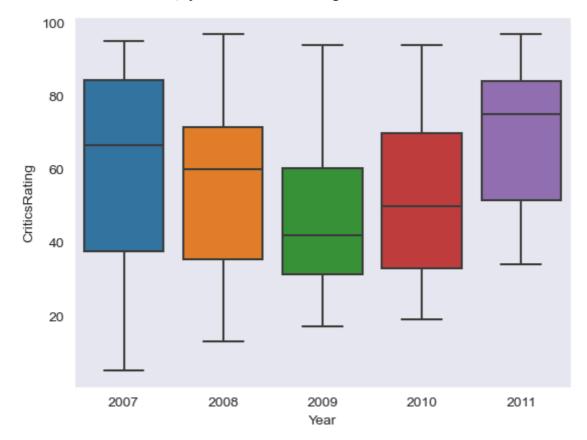


```
In [151... #violin plot
    sns.violinplot(data=movies, x='Genre', y = 'CriticsRating')
```

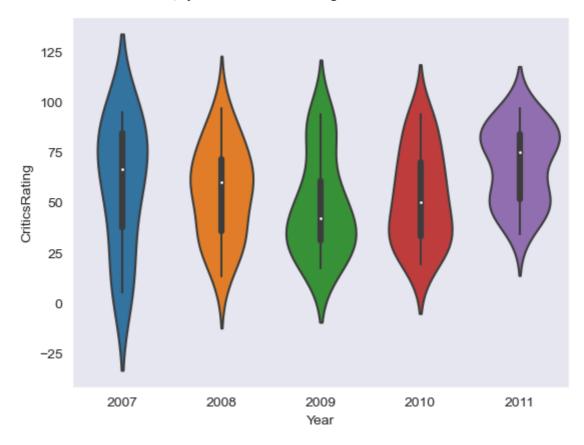


In [152... sns.boxplot(data=movies[movies.Genre=='Drama'], x='Year',y = 'CriticsRating')

Out[152... <Axes: xlabel='Year', ylabel='CriticsRating'>

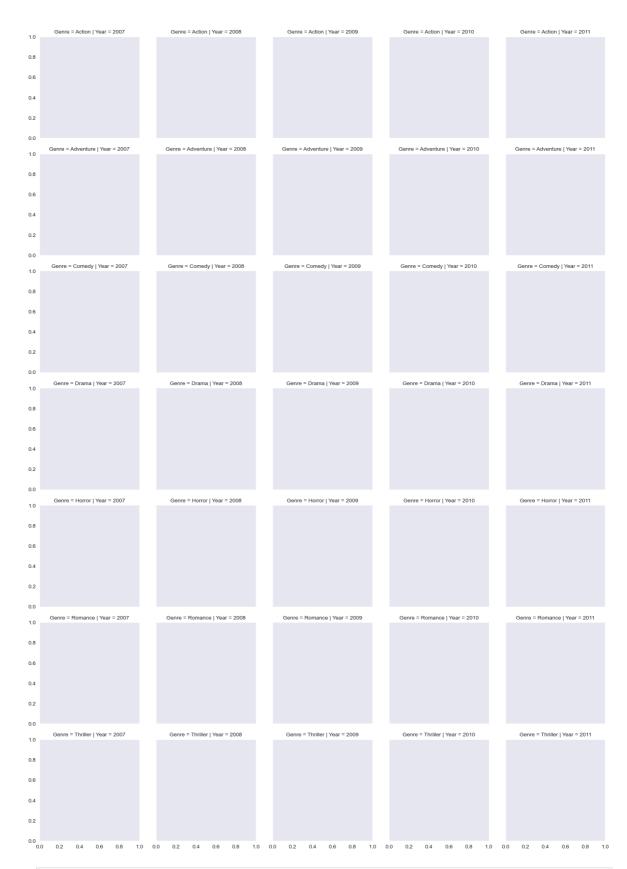


Out[153... <Axes: xlabel='Year', ylabel='CriticsRating'>



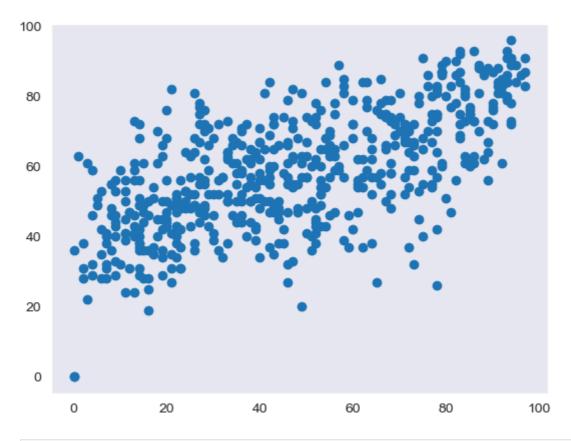
In [154... sns.FacetGrid(movies, row='Genre', col='Year',hue='Genre')

Out[154... <seaborn.axisgrid.FacetGrid at 0x272044e1590>

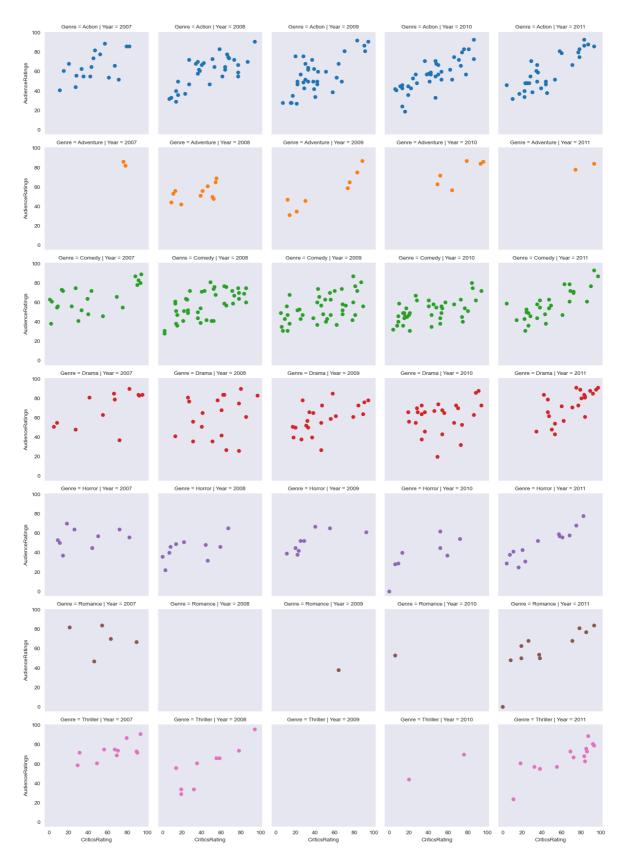


In [156... plt.scatter(movies.CriticsRating,movies.AudienceRatings)

Out[156... <matplotlib.collections.PathCollection at 0x27205d12890>



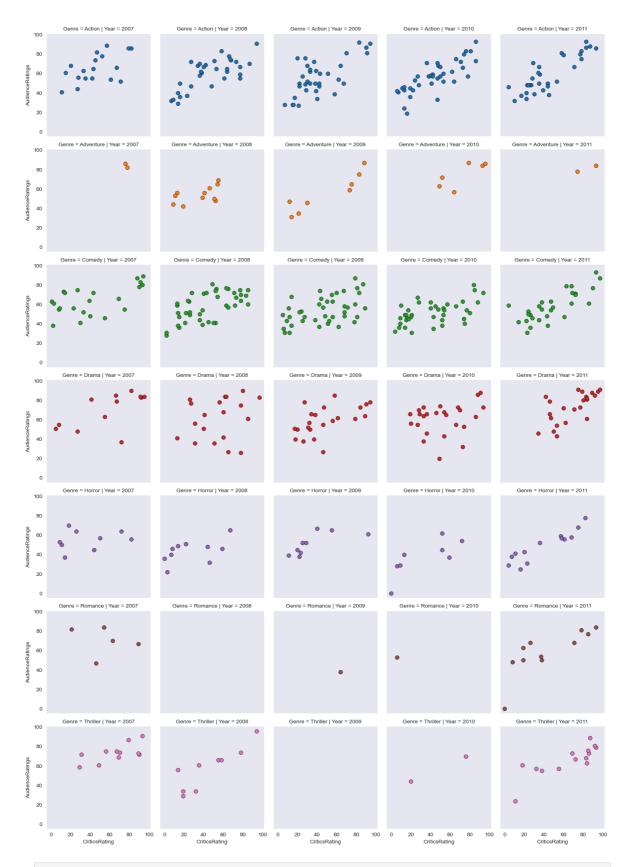
In [159... g = sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
g = g.map(plt.scatter, 'CriticsRating', 'AudienceRatings') #scatterplots are ma



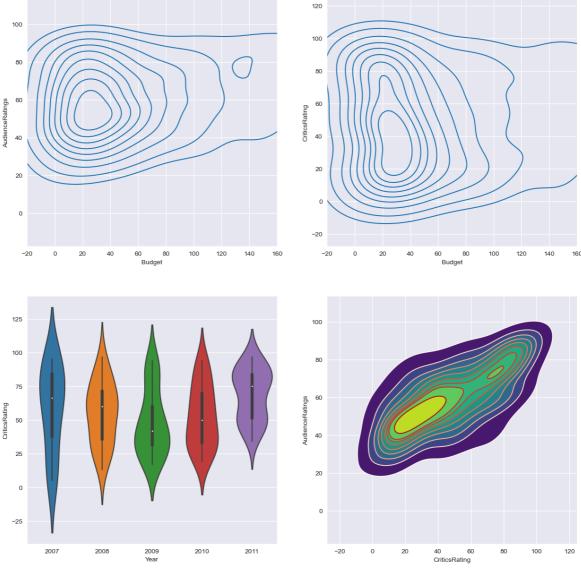
In [160... g= sns.FacetGrid(movies,row='Genre',col='Year', hue='Genre')
g = g.map(plt.hist,'Budget')



g =sns.FacetGrid (movies, row = 'Genre', col = 'Year', hue = 'Genre')
kws = dict(s=50, linewidth=0.5,edgecolor='black')
g = g.map(plt.scatter, 'CriticsRating', 'AudienceRatings',**kws) #scatterplots

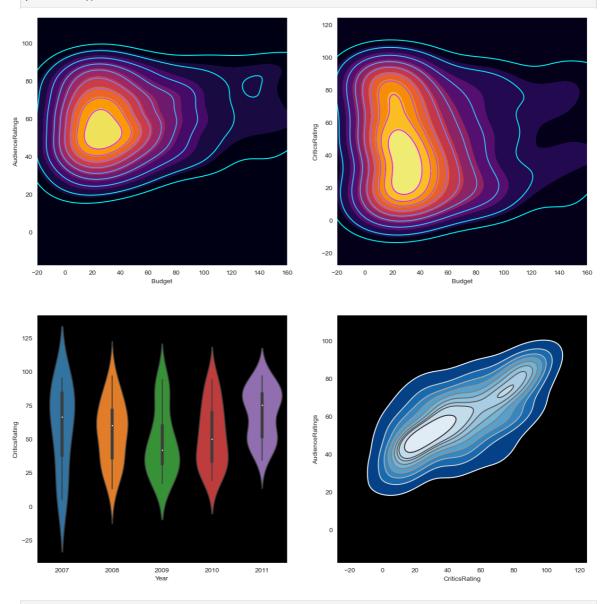


```
In [164...
sns.set_style('darkgrid')
f, axes = plt.subplots (2,2, figsize = (15,15))
k1 = sns.kdeplot(data= movies, x='Budget', y='AudienceRatings', ax=axes[0,0])
k2 = sns.kdeplot(data=movies, x='Budget', y='CriticsRating', ax = axes[0,1])
k1.set(xlim=(-20,160))
k2.set(xlim=(-20,160))
z = sns.violinplot(movies[movies.Genre=='Drama'], x='Year', y = 'CriticsRating',
k4 = sns.kdeplot(data=movies, x='CriticsRating', y='AudienceRatings', shade = True,
k4b = sns.kdeplot(data=movies, x='CriticsRating', y='AudienceRatings', cmap='Reds',
plt.show()
```



```
In [165...
          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='Budget',y='AudienceRatings', \
          shade = True, shade_lowest=True,cmap = 'inferno', \
          ax = axes[0,0])
          k1b = sns.kdeplot(data=movies, x='Budget', y='AudienceRatings', \
          cmap = 'cool', ax = axes[0,0])
          #plot [0,1]
          k2 = sns.kdeplot(data=movies, x='Budget', y='CriticsRating', \
          shade=True, shade_lowest=True, cmap='inferno',\
          ax = axes[0,1]
          k2b = sns.kdeplot(data=movies, x='Budget', y='CriticsRating', \
          cmap = 'cool', ax = axes[0,1])
          #plot[1,0]
          z = sns.violinplot(data=movies[movies.Genre=='Drama'], \
          x='Year', y = 'CriticsRating', ax=axes[1,0])
          #plot[1,1]
          k4 = sns.kdeplot(data=movies,x='CriticsRating',y='AudienceRatings', \
          shade = True, shade_lowest=False, cmap='Blues_r', \
          ax=axes[1,1]
          k4b = sns.kdeplot( data=movies,x='CriticsRating',y='AudienceRatings',\
          cmap='gist_gray_r',ax = axes[1,1])
          k1.set(xlim=(-20,160))
```

k2.set(xlim=(-20,160)) plt.show()



In []: