In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
column_names = ['user_id', 'item_id', 'rating', 'timestamp']
df = pd.read_csv('u.data', sep='\t', names=column_names)
```

In [3]:
```
df.head()
```

Out[3]:

	user_id	item_id	rating	timestamp
0	0	50	5	881250949
1	0	172	5	881250949
2	0	133	1	881250949
3	196	242	3	881250949
4	186	302	3	891717742

In [4]:
```
movie_titles = pd.read_csv('Movie_Id_Titles')
```

In [5]:
```
movie_titles.head()
```

Out[5]:

	item_id	title
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

In [6]:
```
df = pd.merge(df,movie_titles,on='item_id')
```

In [7]:
```
df.head()
```

Out[7]:

	user_id	item_id	rating	timestamp	title
0	0	50	5	881250949	Star Wars (1977)
1	290	50	5	880473582	Star Wars (1977)
2	79	50	4	891271545	Star Wars (1977)
3	2	50	5	888552084	Star Wars (1977)
4	8	50	5	879362124	Star Wars (1977)

In [9]:
```
import matplotlib.pyplot as plt
```

In [10]:
```
import seaborn as sns
```

In [11]:
```
sns.set_style('white')
```

In [12]:
```
%matplotlib inline
```

In [13]:
```
df.groupby('title')['rating'].mean()
```

Out[13]:

title	
'Til There Was You (1997)	2.333333
1-900 (1994)	2.600000
101 Dalmatians (1996)	2.908257
12 Angry Men (1957)	4.344000
187 (1997)	3.024390
2 Days in the Valley (1996)	3.225806
20,000 Leagues Under the Sea (1954)	3.500000
2001: A Space Odyssey (1968)	3.969112
3 Ninjas: High Noon At Mega Mountain (1998)	1.000000
39 Steps, The (1935)	4.050847
8 1/2 (1963)	3.815789
8 Heads in a Duffel Bag (1997)	3.250000
8 Seconds (1994)	3.750000
A Chef in Love (1996)	4.125000
Above the Rim (1994)	3.000000
Absolute Power (1997)	3.370079
Abyss, The (1989)	3.589404
Ace Ventura: Pet Detective (1994)	3.048544
Ace Ventura: When Nature Calls (1995)	2.675676
Across the Sea of Time (1995)	2.750000
Addams Family Values (1993)	2.816092
Addicted to Love (1997)	3.166667
Addiction, The (1995)	2.181818
Adventures of Pinocchio, The (1996)	3.051282
Adventures of Priscilla, Queen of the Desert, The (1994)	3.594595
Adventures of Robin Hood, The (1938)	3.791045
Affair to Remember, An (1957)	4.192308
African Queen, The (1951)	4.184211
Afterglow (1997)	3.111111
Age of Innocence, The (1993)	3.384615
	...
Window to Paris (1994)	4.000000
Wings of Courage (1995)	4.000000
Wings of Desire (1987)	4.000000
Wings of the Dove, The (1997)	3.680000
Winnie the Pooh and the Blustery Day (1968)	3.800000
Winter Guest, The (1997)	3.444444
Wishmaster (1997)	2.444444
With Honors (1994)	3.065217
Withnail and I (1987)	3.230769
Witness (1985)	4.000000
Wizard of Oz, The (1939)	4.077236
Wolf (1994)	2.701493
Woman in Question, The (1950)	1.000000
Women, The (1939)	3.666667
Wonderful, Horrible Life of Leni Riefenstahl, The (1993)	4.000000
Wonderland (1997)	3.200000
Wooden Man's Bride, The (Wu Kui) (1994)	2.666667
World of Apu, The (Apur Sansar) (1959)	4.000000
Wrong Trousers, The (1993)	4.466102
Wyatt Earp (1994)	3.100000
Yankee Zulu (1994)	1.000000
Year of the Horse (1997)	3.285714
You So Crazy (1994)	3.000000

```
Young Frankenstein (1974)                                  3.945000
Young Guns (1988)                                          3.207921
Young Guns II (1990)                                       2.772727
Young Poisoner's Handbook, The (1995)                      3.341463
Zeus and Roxanne (1997)                                    2.166667
unknown                                                    3.444444
Á köldum klaka (Cold Fever) (1994)                         3.000000
Name: rating, Length: 1664, dtype: float64
```

In [15]:
```
df.groupby('title')['rating'].mean().sort_values(ascending=False).head()
```

Out[15]:
```
title
Marlene Dietrich: Shadow and Light (1996)     5.0
Prefontaine (1997)                            5.0
Santa with Muscles (1996)                     5.0
Star Kid (1997)                               5.0
Someone Else's America (1995)                 5.0
Name: rating, dtype: float64
```

In [16]:
```
df.groupby('title')['rating'].count().sort_values(ascending=False).head()
```

Out[16]:
```
title
Star Wars (1977)             584
Contact (1997)               509
Fargo (1996)                 508
Return of the Jedi (1983)    507
Liar Liar (1997)             485
Name: rating, dtype: int64
```

In [18]:
```
ratings = pd.DataFrame(df.groupby('title')['rating'].mean())
ratings.head()
```

Out[18]:

	rating
title	
'Til There Was You (1997)	2.333333
1-900 (1994)	2.600000
101 Dalmatians (1996)	2.908257
12 Angry Men (1957)	4.344000
187 (1997)	3.024390

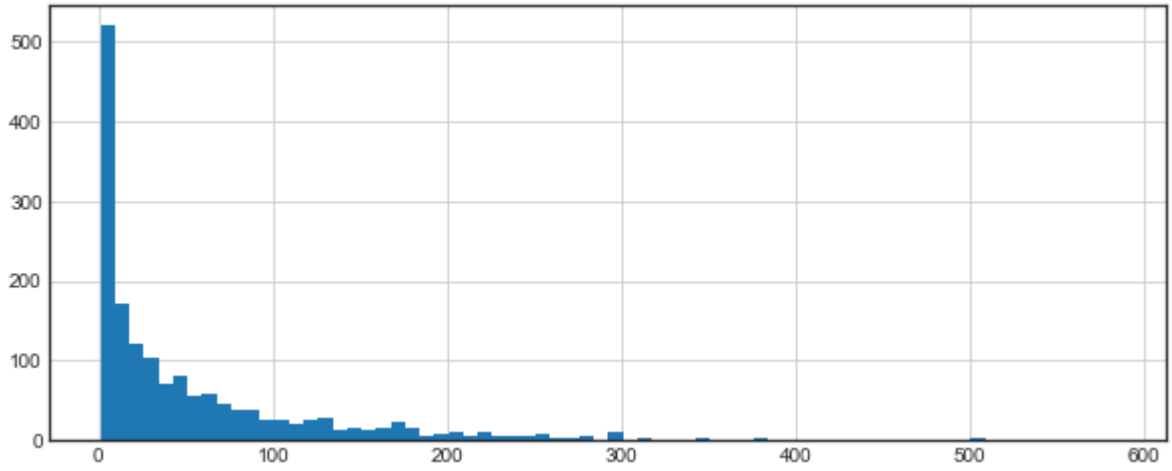
In [19]:
```
ratings['num of ratings'] = pd.DataFrame(df.groupby('title')['rating'].count())
ratings.head()
```

Out[19]:

title	rating	num of ratings
'Til There Was You (1997)	2.333333	9
1-900 (1994)	2.600000	5
101 Dalmatians (1996)	2.908257	109
12 Angry Men (1957)	4.344000	125
187 (1997)	3.024390	41

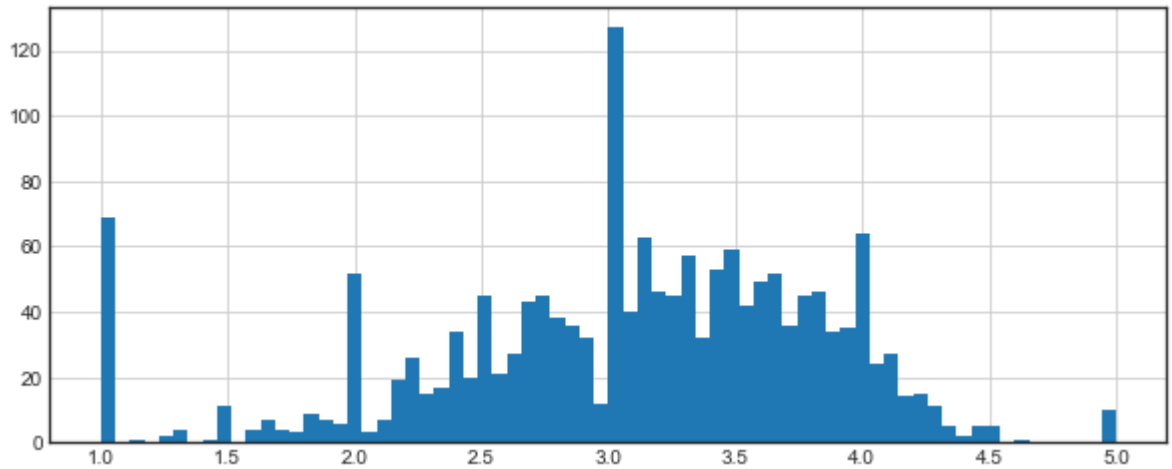
In [20]:
```
plt.figure(figsize=(10,4))
ratings['num of ratings'].hist(bins=70)
```

Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0xb719e48>



In [21]:
```
plt.figure(figsize=(10,4))
ratings['rating'].hist(bins=70)
```

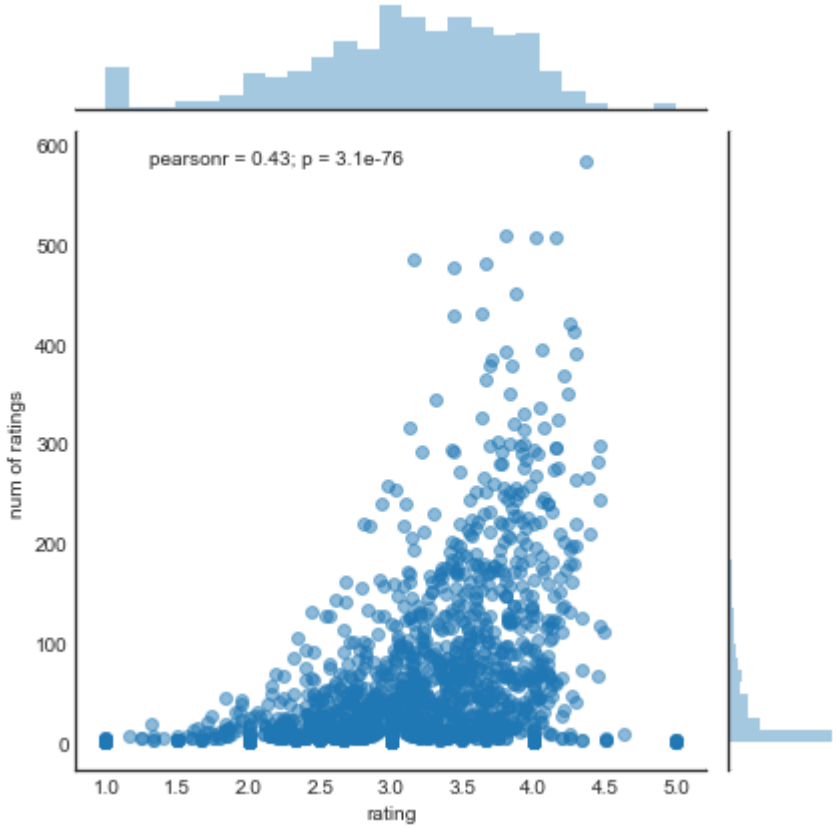
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0xb951c50>



In [22]:
```
sns.jointplot(x='rating',y='num of ratings',data=ratings,alpha=0.5)
```

```
C:\Users\q21\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWar
ning: The 'normed' kwarg is deprecated, and has been replaced by the 'density'
kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
C:\Users\q21\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWar
ning: The 'normed' kwarg is deprecated, and has been replaced by the 'density'
kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```

Out[22]: <seaborn.axisgrid.JointGrid at 0xba5bbe0>



In [23]:
```
# Matrix
```

In [24]:
```
moviemat = df.pivot_table(index='user_id',columns='title',values='rating')
moviemat.head()
```

Out[24]:

user_id	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)	39 Steps, The (1935)	..
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	..
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	..
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	..
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	..
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	..

5 rows × 1664 columns

In [25]:
```
ratings.sort_values('num of ratings',ascending=False).head(10)
```

Out[25]:

title	rating	num of ratings
Star Wars (1977)	4.359589	584
Contact (1997)	3.803536	509
Fargo (1996)	4.155512	508
Return of the Jedi (1983)	4.007890	507
Liar Liar (1997)	3.156701	485
English Patient, The (1996)	3.656965	481
Scream (1996)	3.441423	478
Toy Story (1995)	3.878319	452
Air Force One (1997)	3.631090	431
Independence Day (ID4) (1996)	3.438228	429

In [26]:
```
starwars_user_ratings = moviemat['Star Wars (1977)']
liarliar_user_ratings = moviemat['Liar Liar (1997)']
starwars_user_ratings.head()
```

Out[26]:
```
user_id
0    5.0
1    5.0
2    5.0
3    NaN
4    5.0
Name: Star Wars (1977), dtype: float64
```

In [27]:
```
similar_to_starwars = moviemat.corrwith(starwars_user_ratings)
similar_to_liarliar = moviemat.corrwith(liarliar_user_ratings)
```

```
C:\Users\q21\Anaconda3\lib\site-packages\numpy\lib\function_base.py:3175: Runti
meWarning: Degrees of freedom <= 0 for slice
  c = cov(x, y, rowvar)
C:\Users\q21\Anaconda3\lib\site-packages\numpy\lib\function_base.py:3109: Runti
meWarning: divide by zero encountered in double_scalars
  c *= 1. / np.float64(fact)
```

In [28]:
```
corr_starwars = pd.DataFrame(similar_to_starwars,columns=['Correlation'])
corr_starwars.dropna(inplace=True)
corr_starwars.head()
```

Out[28]:

title	Correlation
'Til There Was You (1997)	0.872872
1-900 (1994)	-0.645497
101 Dalmatians (1996)	0.211132
12 Angry Men (1957)	0.184289
187 (1997)	0.027398

In [29]:
```
corr_starwars.sort_values('Correlation',ascending=False).head(10)
```

Out[29]:

title	Correlation
Hollow Reed (1996)	1.0
Stripes (1981)	1.0
Beans of Egypt, Maine, The (1994)	1.0
Safe Passage (1994)	1.0
Old Lady Who Walked in the Sea, The (Vieille qui marchait dans la mer, La) (1991)	1.0
Outlaw, The (1943)	1.0
Line King: Al Hirschfeld, The (1996)	1.0
Hurricane Streets (1998)	1.0
Good Man in Africa, A (1994)	1.0
Scarlet Letter, The (1926)	1.0

In [30]:
```
corr_starwars = corr_starwars.join(ratings['num of ratings'])
corr_starwars.head()
```

Out[30]:

title	Correlation	num of ratings
'Til There Was You (1997)	0.872872	9
1-900 (1994)	-0.645497	5
101 Dalmatians (1996)	0.211132	109
12 Angry Men (1957)	0.184289	125
187 (1997)	0.027398	41

In [31]:
```
corr_starwars[corr_starwars['num of ratings']>100].sort_values('Correlation',asce
```

Out[31]:

title	Correlation	num of ratings
Star Wars (1977)	1.000000	584
Empire Strikes Back, The (1980)	0.748353	368
Return of the Jedi (1983)	0.672556	507
Raiders of the Lost Ark (1981)	0.536117	420
Austin Powers: International Man of Mystery (1997)	0.377433	130

In [32]:
```
corr_liarliar = pd.DataFrame(similar_to_liarliar,columns=['Correlation'])
corr_liarliar.dropna(inplace=True)
corr_liarliar = corr_liarliar.join(ratings['num of ratings'])
corr_liarliar[corr_liarliar['num of ratings']>100].sort_values('Correlation',asce
```

Out[32]:

title	Correlation	num of ratings
Liar Liar (1997)	1.000000	485
Batman Forever (1995)	0.516968	114
Mask, The (1994)	0.484650	129
Down Periscope (1996)	0.472681	101
Con Air (1997)	0.469828	137

In [ ]:
```

```