In [8]:
```
%matplotlib inline

from numpy import arange
from matplotlib import pyplot as plt
from scipy.stats import norm
import pandas as pd

plt.rcParams['figure.figsize'] = [16, 7]
```

In [9]:
```
columns = ['anime_id', 'name', 'genre', 'type', 'episodes', 'rating', 'members']

df = pd.read_csv("anime.csv")
df.head()
```

Out[9]:

	anime_id	name	genre	type	episodes	rating	members
0	32281	Kimi no Na wa.	Drama, Romance, School, Supernatural	Movie	1	9.37	200630
1	5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Mili...	TV	64	9.26	793665
2	28977	Gintama°	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.25	114262
3	9253	Steins;Gate	Sci-Fi, Thriller	TV	24	9.17	673572
4	9969	Gintama'	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.16	151266

In [10]:
```
df.dtypes
```

Out[10]:
```
anime_id      int64
name         object
genre        object
type         object
episodes     object
rating      float64
members       int64
dtype: object
```

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12294 entries, 0 to 12293
Data columns (total 7 columns):
anime_id    12294 non-null int64
name        12294 non-null object
genre       12232 non-null object
type        12269 non-null object
episodes    12294 non-null object
rating      12064 non-null float64
members     12294 non-null int64
dtypes: float64(1), int64(2), object(4)
memory usage: 672.4+ KB
```

In [12]:
```
df.memory_usage()
```

Out[12]:
```
Index          80
anime_id    98352
name        98352
genre       98352
type        98352
episodes    98352
rating      98352
members     98352
dtype: int64
```

In [13]:
```
df.memory_usage().sum()
```

Out[13]:
```
688544
```

In [14]:
```
df.describe()
```

Out[14]:

	anime_id	rating	members
count	12294.000000	12064.000000	1.229400e+04
mean	14058.221653	6.473902	1.807134e+04
std	11455.294701	1.026746	5.482068e+04
min	1.000000	1.670000	5.000000e+00
25%	3484.250000	5.880000	2.250000e+02
50%	10260.500000	6.570000	1.550000e+03
75%	24794.500000	7.180000	9.437000e+03
max	34527.000000	10.000000	1.013917e+06

In [15]:
```
df.mean()
```

Out[15]:
```
anime_id    14058.221653
rating          6.473902
members     18071.338864
dtype: float64
```

In [16]:
```
df['rating'].mean()
```

Out[16]:
```
6.473901690981445
```

In [17]:
```
df.var()
```

Out[17]:
```
anime_id    1.312238e+08
rating      1.054208e+00
members     3.005307e+09
dtype: float64
```

In [18]:
```
df.skew()
```

Out[18]:
```
anime_id    0.441550
rating     -0.543570
members     6.682934
dtype: float64
```

In [19]:
```
df.kurtosis()
```

Out[19]:
```
anime_id    -1.312915
rating       0.507530
members     62.856097
dtype: float64
```

In [20]:
```
df.min()
```

Out[20]:
```
anime_id                1
name        &amp;quot;0&amp;quot;
episodes                1
rating               1.67
members                 5
dtype: object
```

In [21]:
```
df.max()
```

Out[21]:
```
anime_id      34527
name              ◯
episodes    Unknown
rating           10
members     1013917
dtype: object
```

In [22]:
```
df.median()
```

Out[22]:
```
anime_id    10260.50
rating          6.57
members      1550.00
dtype: float64
```

In [23]:
```
df.corr()
```

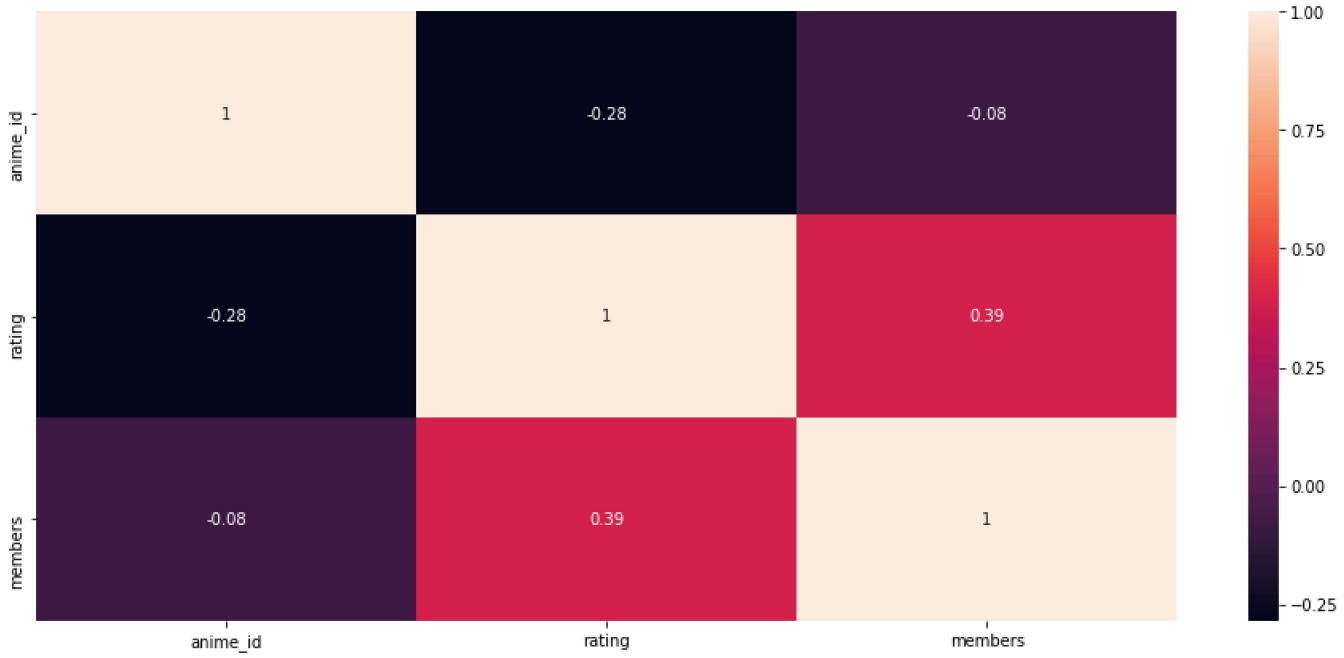
Out[23]:

	anime_id	rating	members
anime_id	1.000000	-0.284625	-0.080071
rating	-0.284625	1.000000	0.387979
members	-0.080071	0.387979	1.000000

In [24]:
```
import seaborn as sns
```

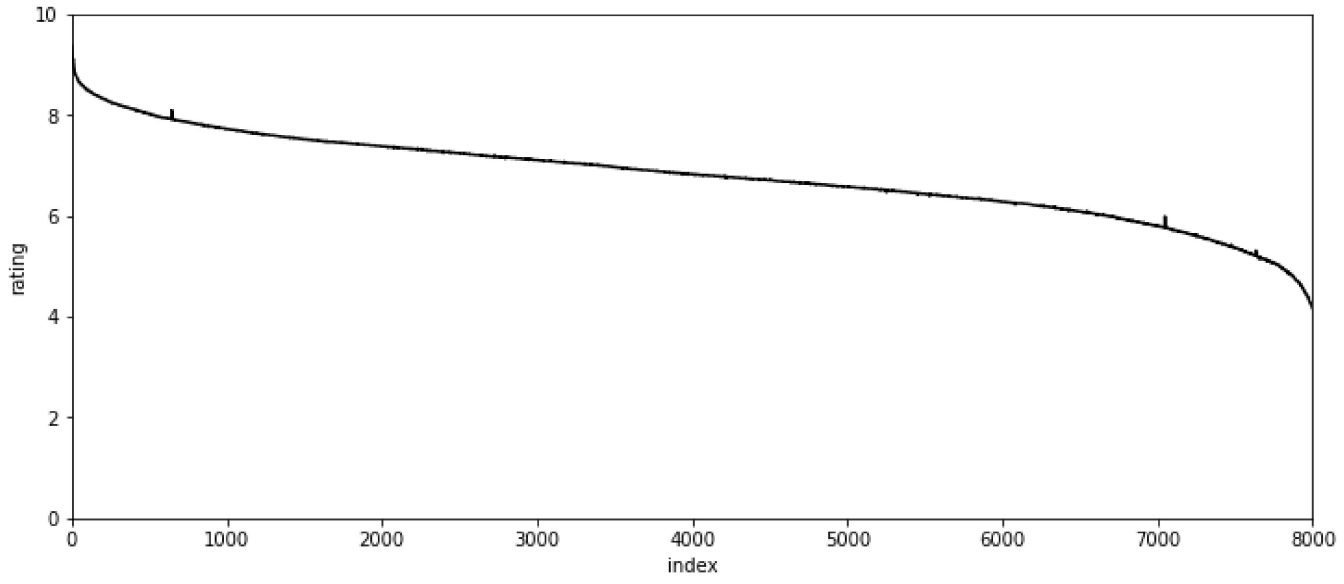
In [25]:
```
sns.heatmap(df.corr(), annot=True)
```

Out[25]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1dab0e48b38>
```



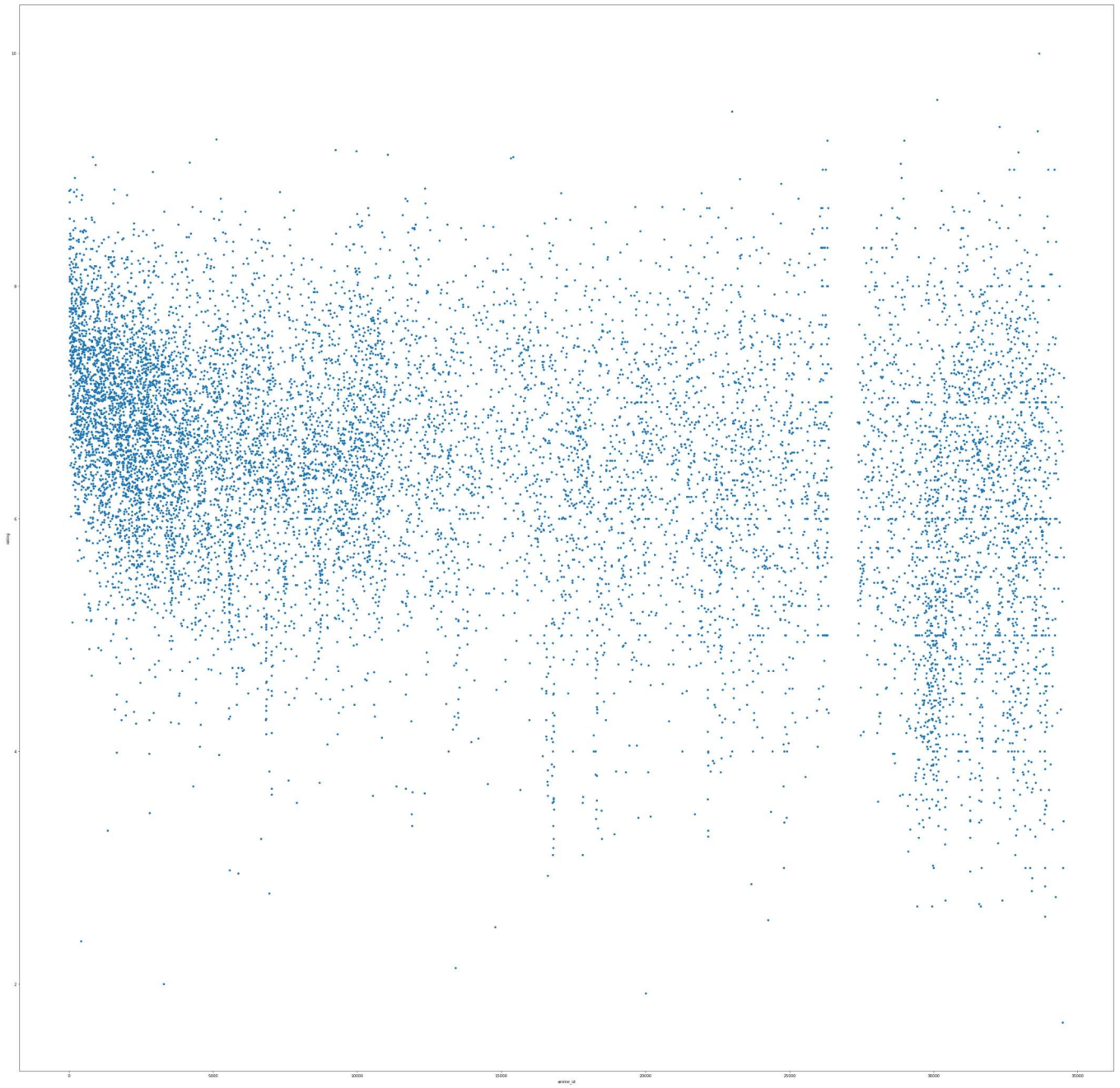
In [26]:
```
# Plotting with index along the x-axis
df['rating'].plot(figsize=(12, 5), color='black') # color and figsize changed

plt.xlim(0, 8000) # range for x-axis
plt.ylim(0, 10) # range for x-axis
plt.xlabel('index')
plt.ylabel('rating'); # ";" prevents object info from displaying
```

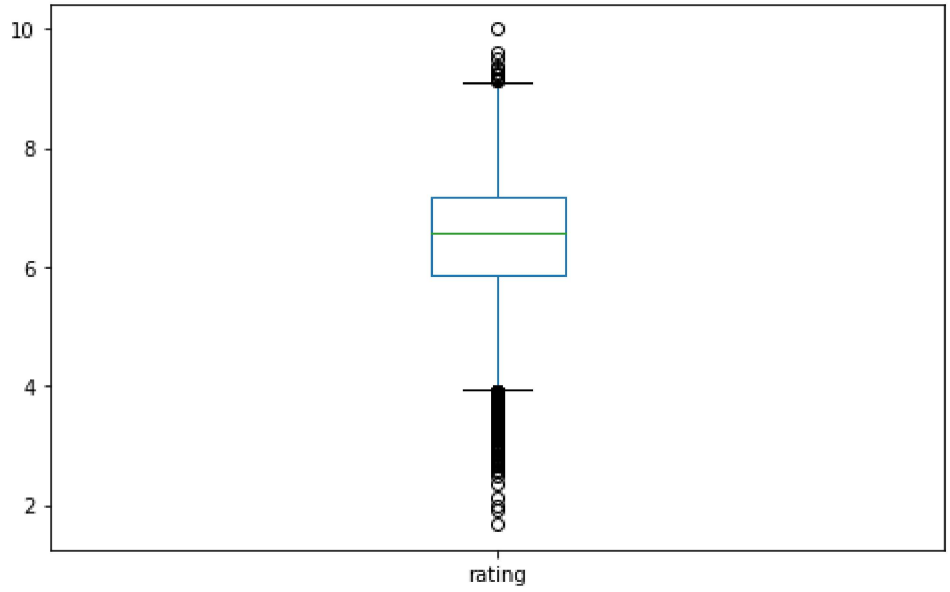


In [27]:
```
df.plot.scatter('anime_id', 'rating', figsize=(50, 50))
```

Out[27]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1dab0fc0630>
```

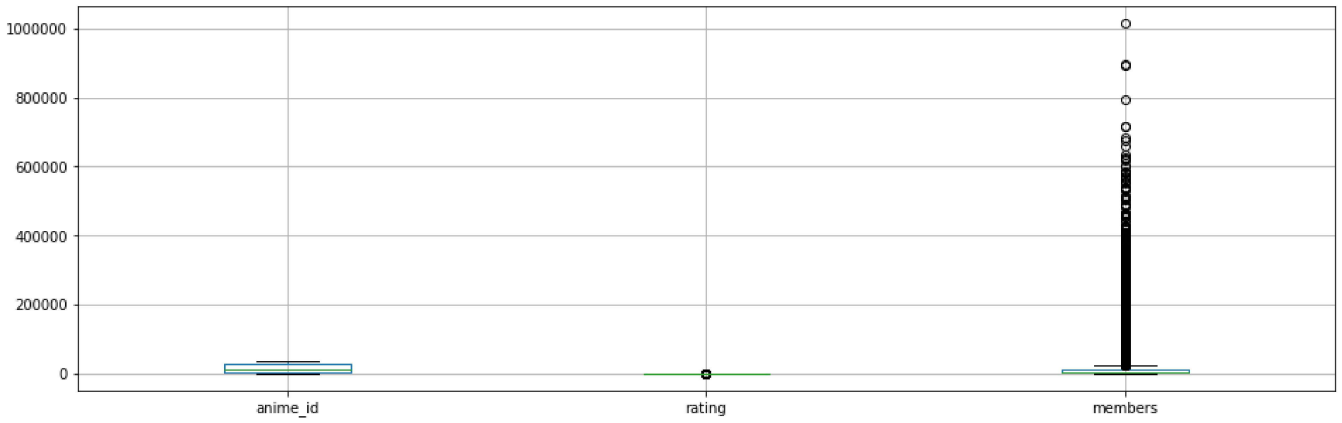


In [28]:
```
df['rating'].plot.box(figsize=(8, 5));
```

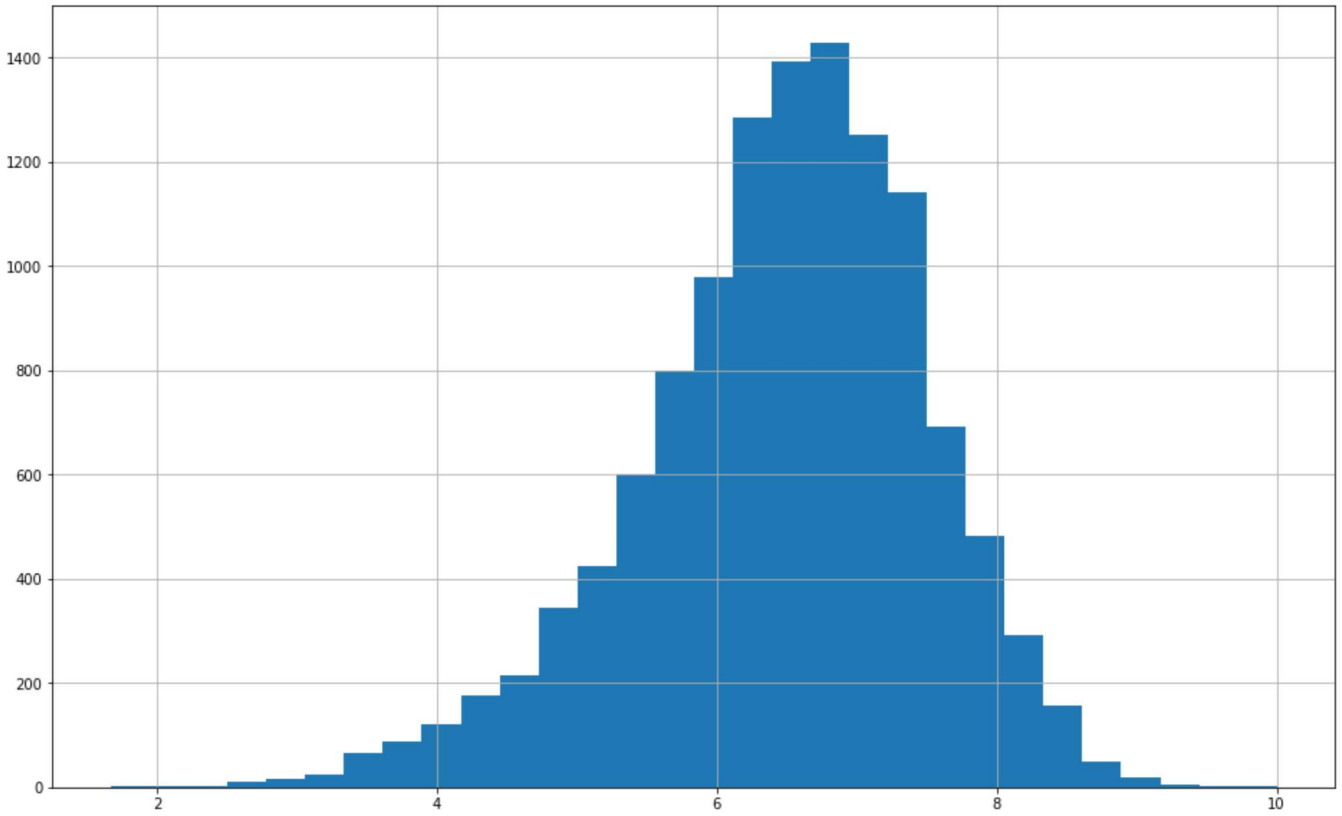


In [29]:
```
df.boxplot(figsize=(16, 5))
```

Out[29]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1dab103f390>
```



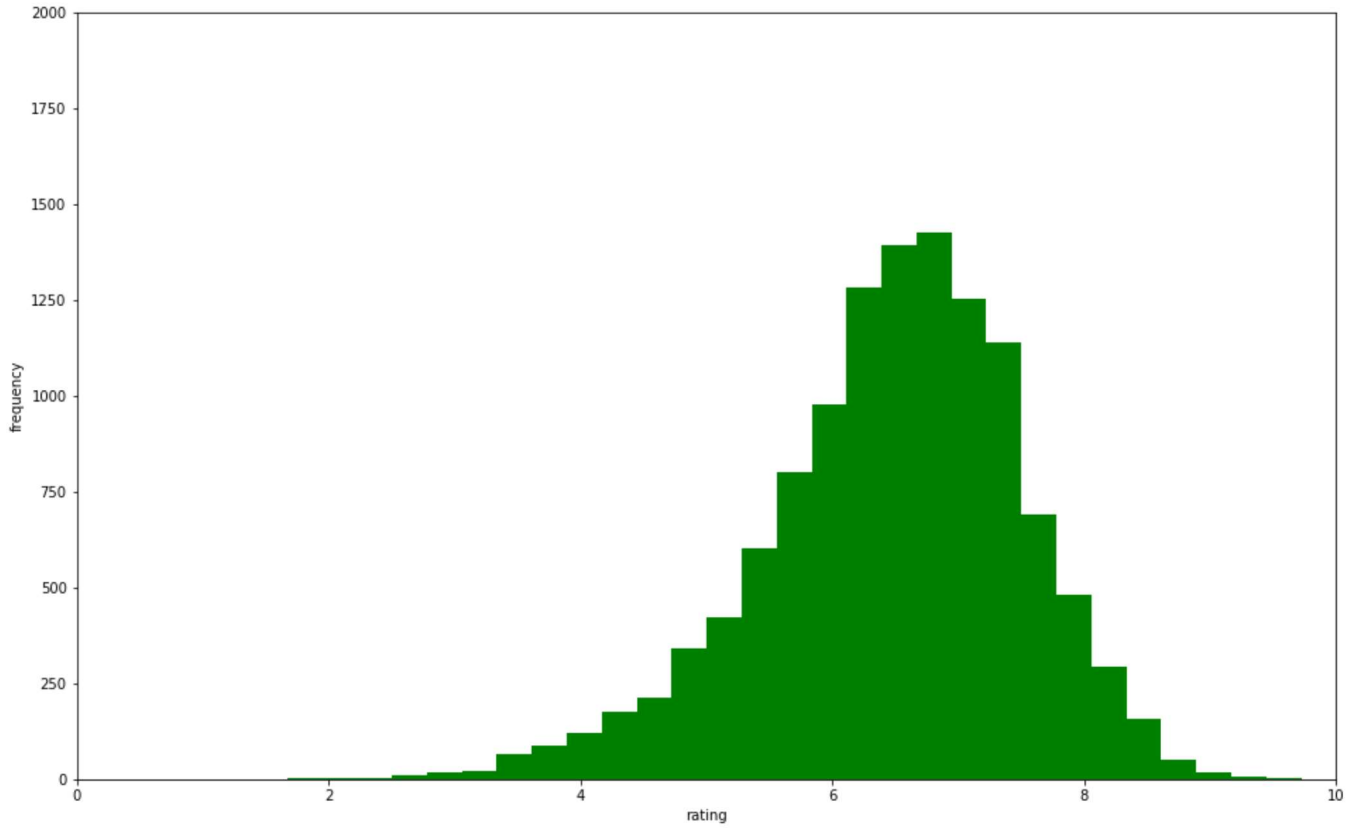
In [30]:
```
df['rating'].hist(bins=30, figsize=(16, 10));
```



In [31]:
```
ax = df['rating'].hist(bins=30, grid=False, color='green', figsize=(16, 10)) # grid turned off and color changed

ax.set_xlabel('rating')
ax.set_ylabel('frequency')

ax.set_xlim(0, 10) # limiting display range to 0-70 for the x-axis
ax.set_ylim(0, 2000); # limiting display range to 0-120 for the y-axis
```



In [32]:
```
ax = df[['rating', 'episodes']][:20].plot.bar(stacked=True)

ax.set_xlabel("Index")
ax.set_ylabel("Count");
```