In [8]:

```python
%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]:

```python
,'Age','Category','length','views','rate','ratings','comment','related id1','related id2','r
```

In [10]:

```python
data.Age
```

Out[10]:

```
0        745
1        744
2        744
3        745
4        745
        ...
162      744
163      746
164      745
165      744
166      744
Name: Age, Length: 167, dtype: int64
```

In [11]:

```python
data.length
```

Out[11]:

```
0        192
1        206
2        318
3         34
4         52
        ...
162      218
163      190
164      552
165      867
166      334
Name: length, Length: 167, dtype: int64
```

In [12]:

```python
data.memory_usage()
```

Out[12]:

```
Index            128
Video ID        1336
Uploader        1336
Age             1336
Category        1336
length          1336
views           1336
rate            1336
ratings         1336
comment         1336
related id1     1336
related id2     1336
related id3     1336
related id4     1336
related id5     1336
related id6     1336
related id7     1336
related id8     1336
related id9     1336
related id10    1336
related id11    1336
related id12    1336
related id13    1336
related id14    1336
related id15    1336
related id16    1336
related id17    1336
related id18    1336
related id19    1336
related id20    1336
dtype: int64
```

In [13]:

```python
data.memory_usage().sum()
```

Out[13]:

```
38872
```

In [14]:

```python
data.mean()
```

Out[14]:

```
Age           744.856287
length        285.622754
views       13753.574850
rate            4.352874
ratings       108.167665
comment        87.161677
dtype: float64
```

In [15]:

```python
data.var()
```

Out[15]:

```
Age          6.418729e-01
length       5.071002e+04
views        6.978418e+08
rate         6.128314e-01
ratings      4.524408e+04
comment      2.953875e+04
dtype: float64
```

In [16]:

```python
data.skew()
```

Out[16]:

```
Age         -2.010436
length       1.620262
views        3.514258
rate        -1.829530
ratings      9.198753
comment      6.833455
dtype: float64
```

In [17]:

```python
data.kurtosis()
```

Out[17]:

```
Age           8.212487
length        5.221875
views        13.148237
rate          2.738035
ratings     102.336580
comment      61.101157
dtype: float64
```

In [18]:

```python
data.min()
```

Out[18]:

```
Video ID           -rE3zAPnU8M
Uploader              60555264
Age                        741
Category       Autos & Vehicles
length                       7
views                      164
rate                      1.29
ratings                      2
comment                      0
dtype: object
```

In [19]:

```python
data.max()
```

Out[19]:

```
Video ID          zz8pryU9GMc
Uploader             yadddle
Age                      746
Category       Travel & Places
length                  1384
views                 159108
rate                       5
ratings                 2530
comment                 1794
dtype: object
```

In [20]:

```python
data.median()
```

Out[20]:

```
Age        745.00
length     228.00
views     6633.00
rate         4.67
ratings     73.00
comment     52.00
dtype: float64
```

In [21]:

```python
data.corr()
```

Out[21]:

	Age	length	views	rate	ratings	comment
Age	1.000000	-0.102377	-0.447826	-0.029305	0.058328	0.050525
length	-0.102377	1.000000	-0.112107	0.237990	-0.055811	-0.078501
views	-0.447826	-0.112107	1.000000	-0.238096	0.305663	0.364266
rate	-0.029305	0.237990	-0.238096	1.000000	-0.026713	-0.129262
ratings	0.058328	-0.055811	0.305663	-0.026713	1.000000	0.858557
comment	0.050525	-0.078501	0.364266	-0.129262	0.858557	1.000000

In [22]:

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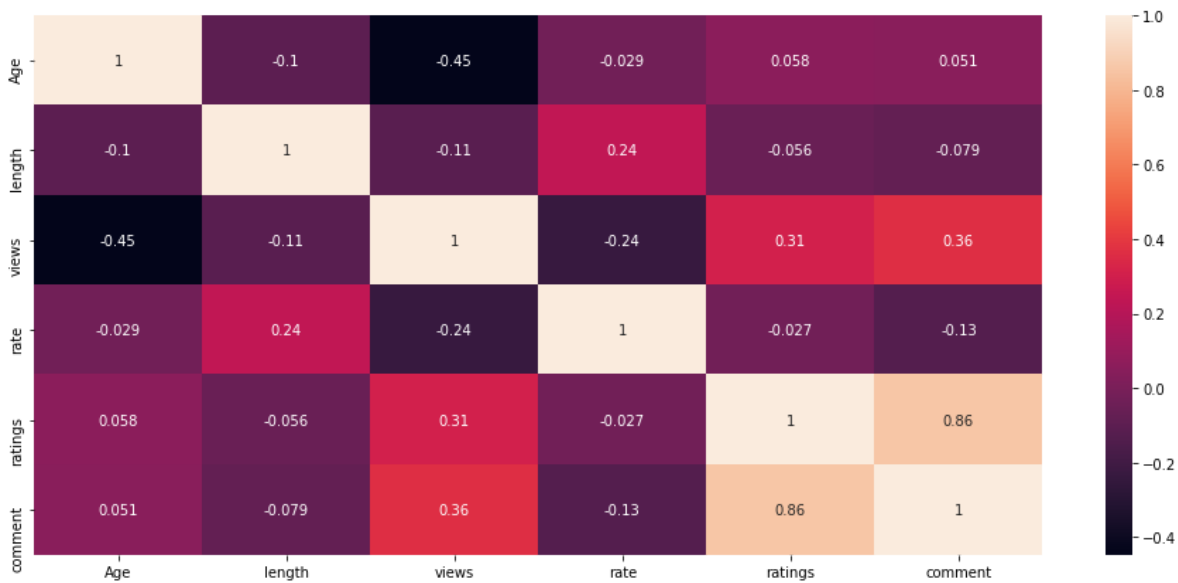
In [24]:

```python
sns.heatmap(data.corr(), annot=True)
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ef236f1c08>
```
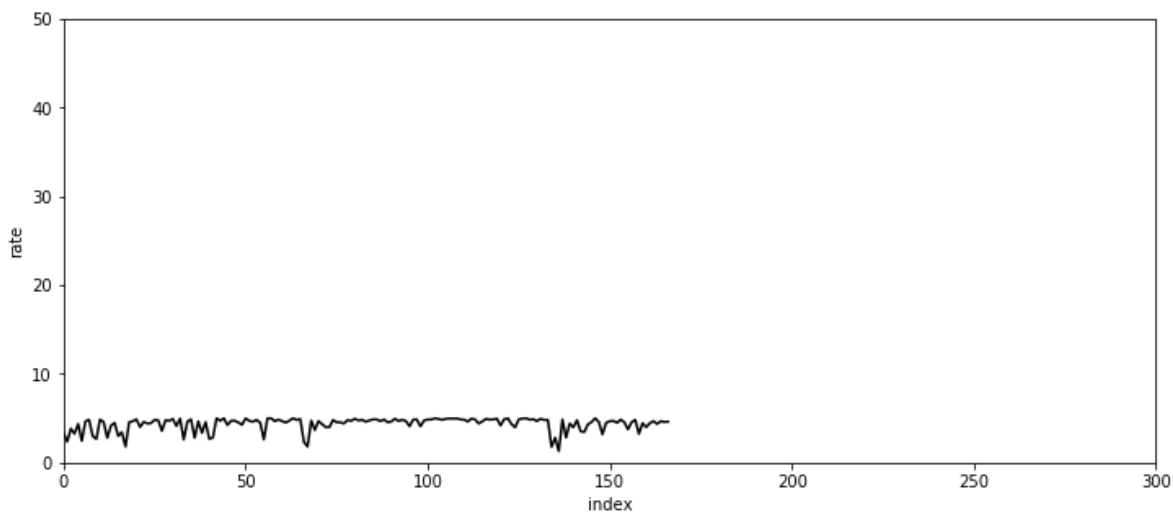

In [28]:

```python
# Plotting with index along the x-axis
data['rate'].plot(figsize=(12, 5), color='black') # color and figsize changed

plt.xlim(0, 300) # range for x-axis
plt.ylim(0, 50) # range for x-axis
plt.xlabel('index')
plt.ylabel('rate'); # ";" prevents object info from displaying
```
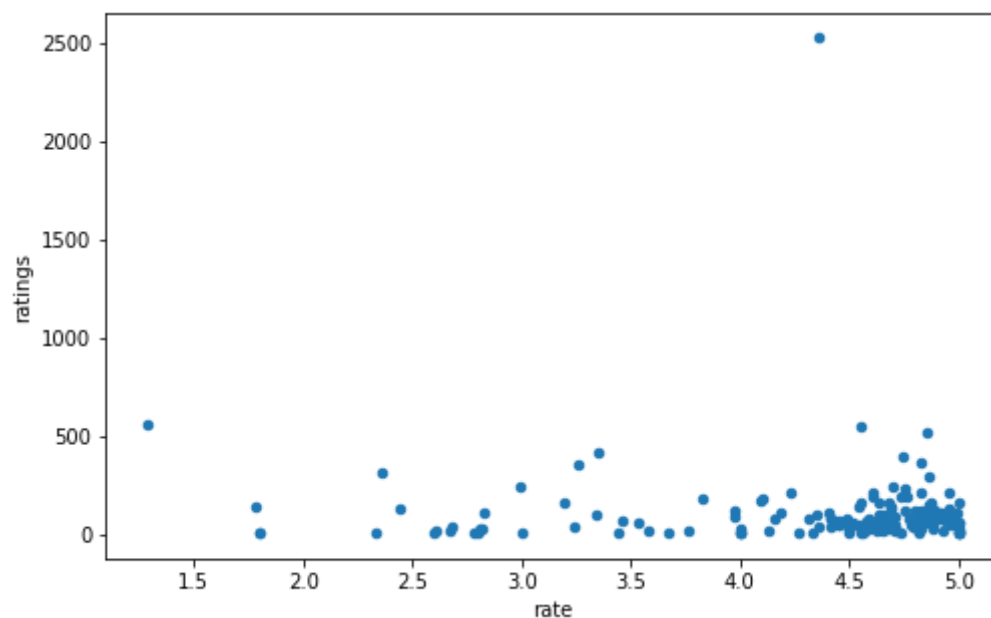
In [29]:

```python
# plotting one variable against the other
data.plot.scatter('rate', 'ratings', figsize=(8, 5))

# The x and y labels are automatically taken from the column names
```

Out[29]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ef2403c3c8>
```


In [30]:
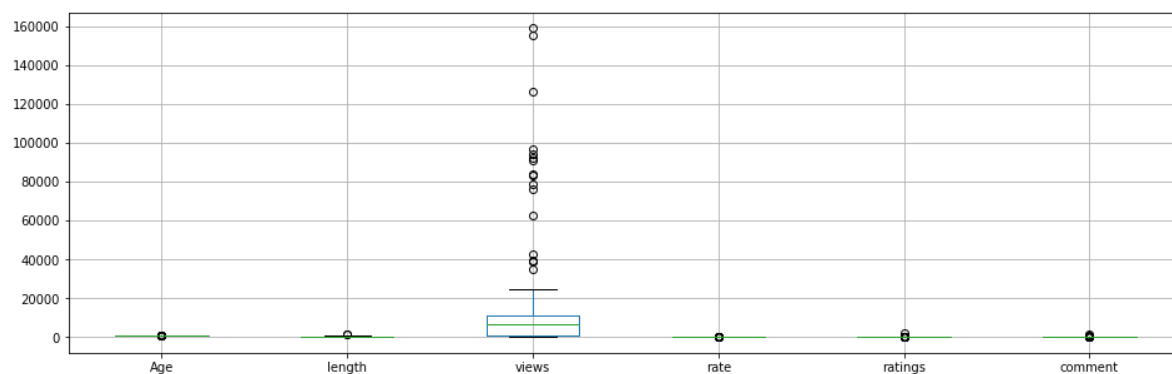
```python
# Box plot of a column
data['rate'].plot.box(figsize=(8, 5));
```
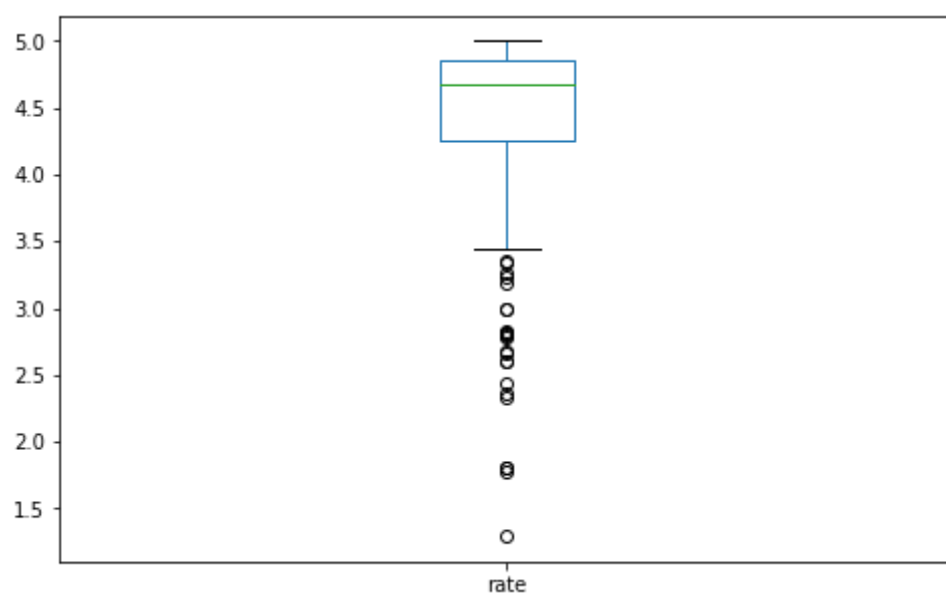
In [31]:

```python
# Box plot of all the columns with numerical data
data.boxplot(figsize=(16, 5)) # or data.plot.box()
```
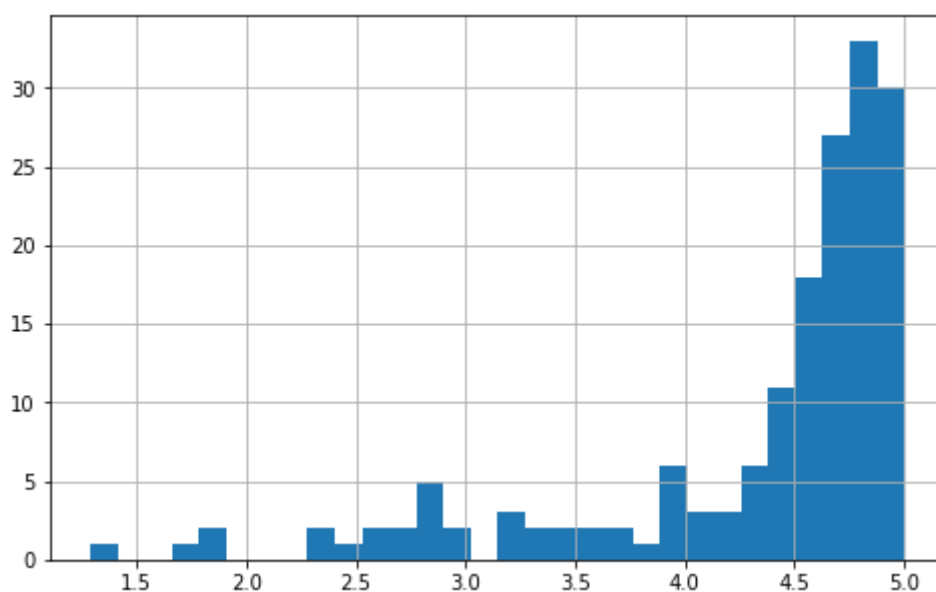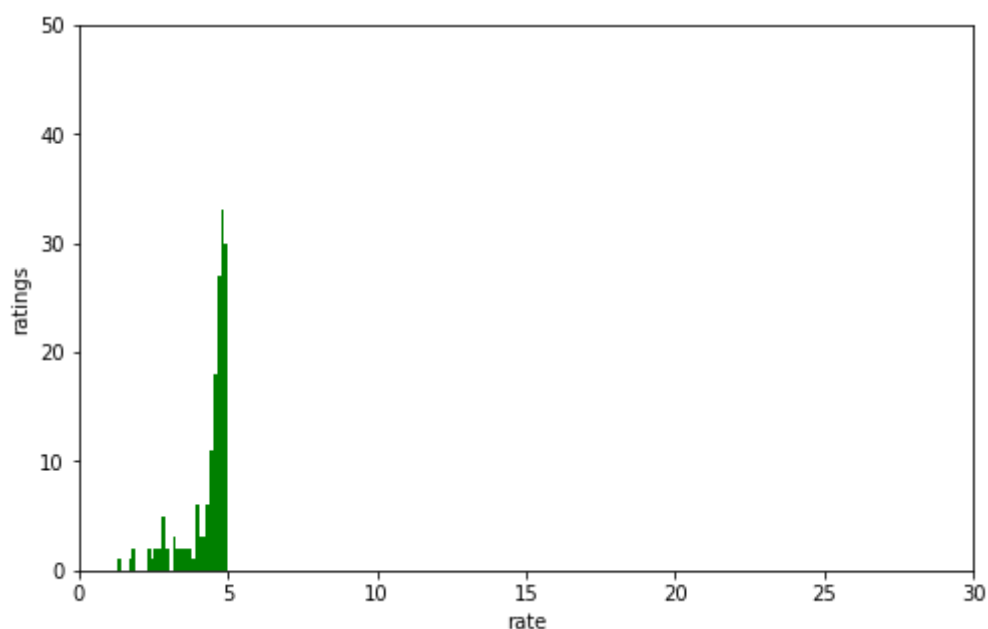
Out[31]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2ef2435e088>
```


In [33]:

```python
data['rate'].hist(bins=30, figsize=(8, 5)); # we can specify the number of bins
```
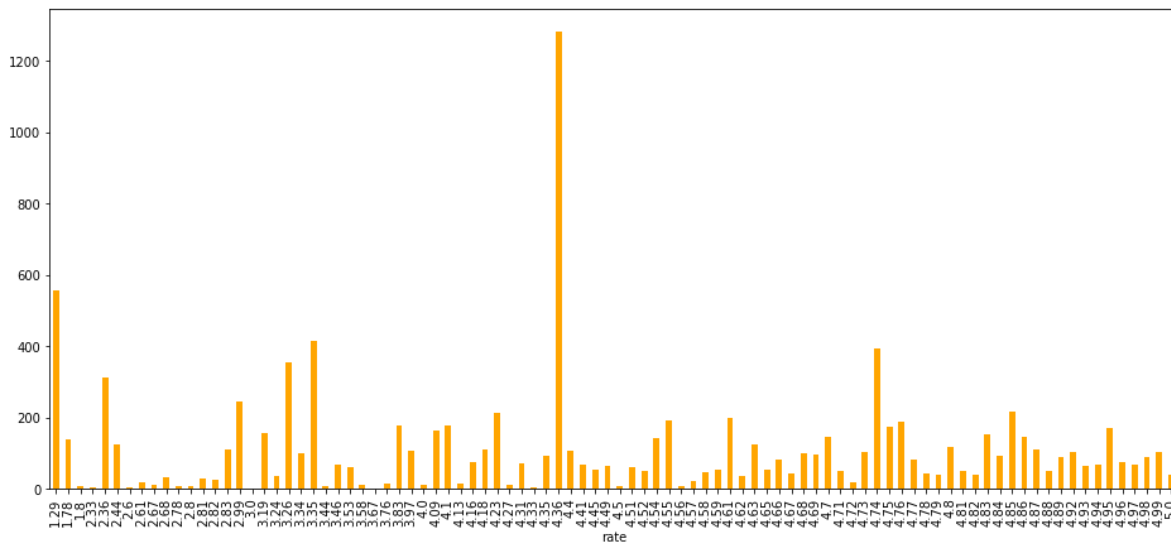
In [35]:

```python
ax = data['rate'].hist(bins=30, grid=False, color='green', figsize=(8, 5)) # grid turned of

ax.set_xlabel('rate')
ax.set_ylabel('ratings')

ax.set_xlim(0, 30) # limiting display range to 0-70 for the x-axis
ax.set_ylim(0, 50); # limiting display range to 0-120 for the y-axis
```
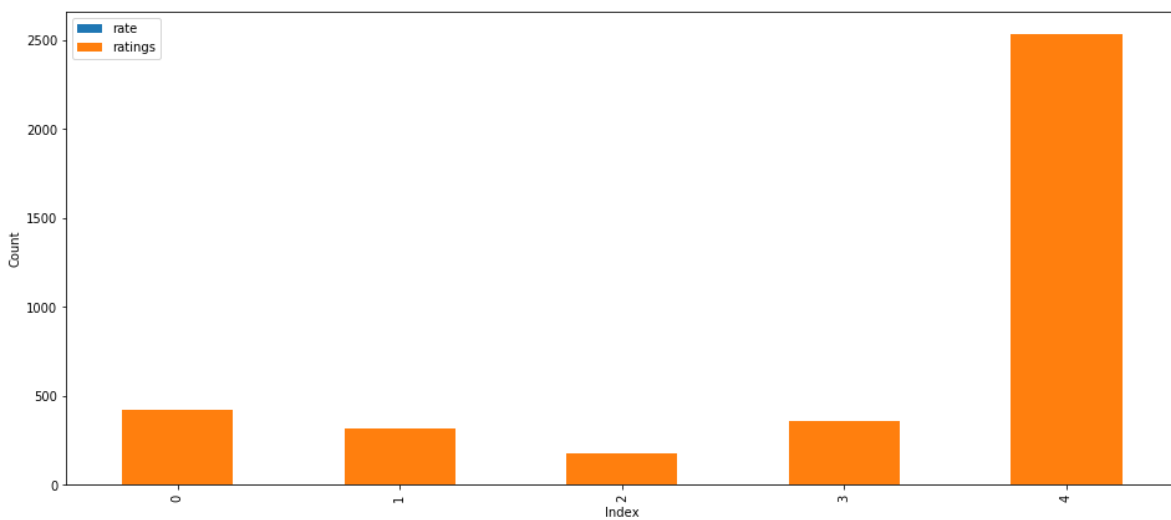
In [37]:

```python
data_avg_BP = data.groupby('rate')['ratings'].mean()
data_avg_BP[:10].plot.bar(color='orange');
```


In [38]:

```python
# To stack values from multiple columns
ax = data[['rate', 'ratings']][:5].plot.bar(stacked=True)

ax.set_xlabel("Index")
ax.set_ylabel("Count");
```
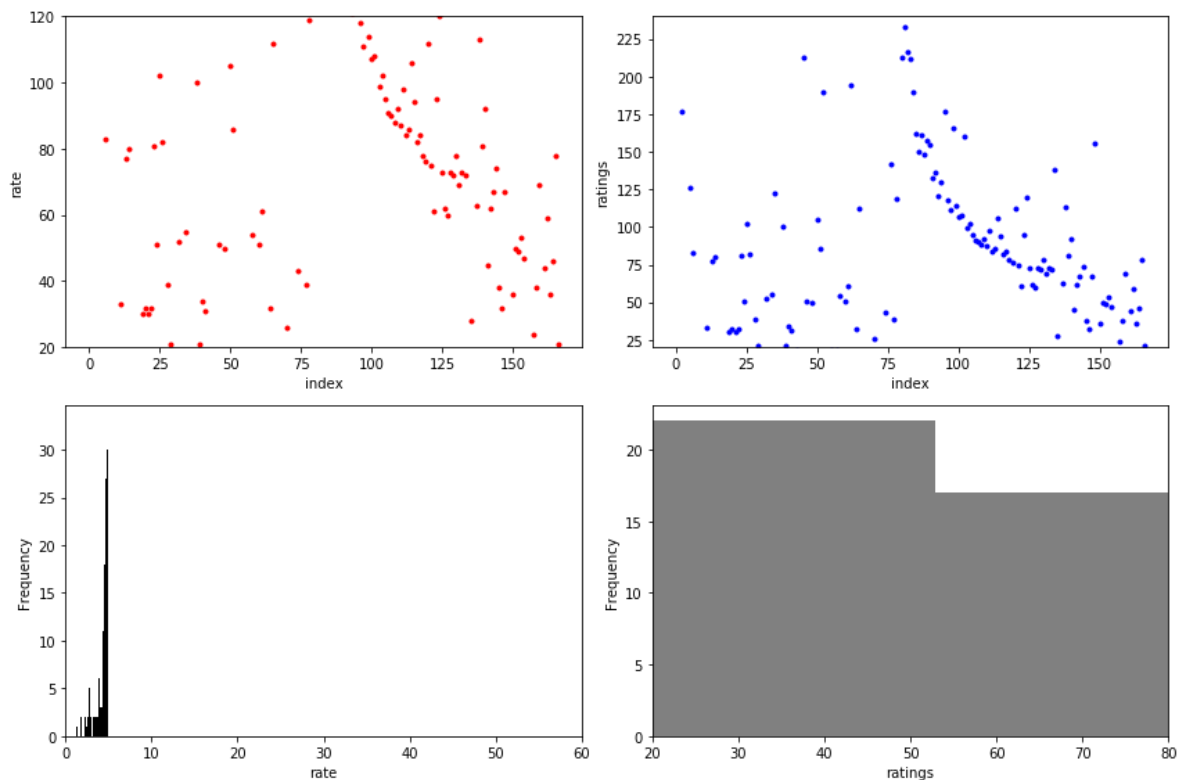
In [44]:

```python
fig, axes = plt.subplots(2, 2, figsize=(12, 8))
# or fig, (ax1, ax2, ax3, ax4) = plt.subplots(2, 2, figsize=(12, 8))

# axes is the axes object(s). It can be a single object or an array of objects.
# In this case, it is an array of dimension 2-by-2

data['ratings'].plot(ax = axes[0][0], style='.', color='red') # top left
data['ratings'].plot(ax = axes[0][1], style='.', color='blue') # top right

data['rate'].plot.hist(bins=30, ax = axes[1][0], color='black') # bottom left
data['length'].plot.hist(bins=30, ax = axes[1][1], color='gray') # bottom right

axes[0][0].set_xlabel('index')
axes[0][1].set_xlabel('index')
axes[1][0].set_xlabel('rate')
axes[1][1].set_xlabel('ratings')

axes[0][0].set_ylabel('rate')
axes[0][1].set_ylabel('ratings')

axes[0][0].set_ylim(20, 120)
axes[0][1].set_ylim(20, 240)

axes[1][0].set_xlim(0, 60)
axes[1][1].set_xlim(20, 80)

fig.tight_layout()
```

In [ ]: