In [1]:
```
import gzip
from collections import defaultdict as dd
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
url = 'amazon_reviews_us_Musical_Instruments_v1_00.tsv.gz'
fd = gzip.open(url, 'rt',encoding='utf-8')
header = fd.readline().rstrip().split('\t')
print(header)
```

```
['marketplace', 'customer_id', 'review_id', 'product_id', 'product_parent
', 'product_title', 'product_category', 'star_rating', 'helpful_votes',
'total_votes', 'vine', 'verified_purchase', 'review_headline', 'review_bod
y', 'review_date']
```

In [3]:
```
all = []
for line in fd:
    fields = fd.readline().rstrip().split('\t')
    d = dict(zip(header,fields))
    all.append(d)
```

In [4]:
```
df = pd.DataFrame(all)
```

In [5]:
```
df = df[df.star_rating.notna()]
df.shape
```

Out[5]:
```
(452382, 15)
```

In [6]:
```
df['star_rating'] = pd.to_numeric(df.star_rating)
df.total_votes = pd.to_numeric(df.total_votes)
df.helpful_votes = pd.to_numeric(df.helpful_votes)
```

In [7]:
```
df.star_rating.sum()/df.shape[0]
```

Out[7]:
```
4.250752682467472
```

In [8]:
```
pids = df['product_id'].value_counts()
```

In [9]:
```
toppids = list(pids[pids>100].index)
```

In [10]:
```
df.star_rating.std()
dfstar = df.loc[:,['product_id','star_rating','total_votes']]
dfstar
```

Out[10]:

	product_id	star_rating	total_votes
0	B003LRN53I	5	0
1	B002D5SSTRG	5	0
2	B0011HGRG8	5	0
3	B009KI0DA2S	5	0
4	B000IF8D0I	5	1
...	...	...	...
452377	B00002F5ND	4	52
452378	1928571018	5	10
452379	B00002F2IZ	4	47
452380	B00002JV63	5	24
452381	B00002F2IZ	4	49

452382 rows × 3 columns

In [11]:
```
dfpids = dfstar.groupby('product_id').mean()
```

In [12]:
```
dftop = dfpids[(dfpids.total_votes > 100)&(dfpids.star_rating<=5.0)]
```

In [13]:
```
dftitle = df.loc[df.product_id.isin(dftop.index)]
```

Out[13]:

	marketplace	customer_id	review_id	product_id	product_parent	product_title
13845	US	1460376	R3AJRY2JLR87N3	B011QEVO44	585766117	LaluceNatz DJ Lights with 27W 9 Colors Multi-e...
440245	US	49408813	R12JOFD0RANQJU	B001CMEKGK	379568655	Yamaha YPG 235 76-Key Portable Grand Piano Key...
443474	US	16990140	R3HWC8LTCO8596	B001CMEKGK	379568655	Yamaha YPG 235 76-Key Portable Grand Piano Key:...
449975	US	52837047	R1W7SQT0D9SSK9	B0002F7F1K	619660968	AC48S Music Stand
451214	US	18288693	RJMP0CR64O6RD	B0002F7F1K	619660968	AC48S Music Stand
452052	US	34525300	R5FVS6KOB66HZ	B0000AVFB3	304227732	Yamaha EZ-250i Portatone Lighted Musical Keyboard

In [19]:
```
import nltk
#nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer as SIA
sia=SIA()
sia.polarity_scores('this horrible product was very bad, terrible')
```

Out[19]:
```
{'neg': 0.727, 'neu': 0.273, 'pos': 0.0, 'compound': -0.8927}
```

In [23]:
```
li = list(map(sia.polarity_scores,list(df.loc[:25,"review_body"])))
print(df.loc[:25,["star_rating","review_headline","review_body"]])
```

```
    star_rating                                    review_headline  \
0             5                                         Five Stars
1             5  I purchase these for a friend in return for pl...
2             5                                         Five Stars
3             5                                         Five Stars
4             5  I upgraded the power cord to a heavier gauge w...
5             4                      Great pop filter - poor mount
6             2                                          Two Stars
7             5                Great stand... only one little glitch
8             1                                  Poor sound quality
9             5  It really is a musical instrument and the book...
10            5                          Perfect for what I needed
11            5                                     good fit.
12            5                                  Why Read? Buy it!
13            5                                         Five Stars
14            5                                  Satisfied Customer
15            5                                         Five Stars
16            5                  Perfect for my Line 6 Relay 50
17            5                                         Five Stars
18            5                        Just what I was looking for!
19            5                                  She just love it!
20            5                                         Five Stars
21            4                      You get what you pay for
22            5                                         Five Stars
23            5                                         Five Stars
24            5                                  Tuning Simplified
25            5                          A must have guitar

                                                review_body
0                 Nice headphones at a reasonable price.
1   I purchase these for a friend in return for pl...
2   Used to cool equipment inside credenzas.  Work...
3                 Well built Ukulele! My daughter loves it!
4   I've owned multiple fixed boards over the year...
5   by far the best pop filter i have used, extrem...
6   Bridge pickup was broken. I replace d the pick...
7   I love the stand.  I bought two.  The only gli...
8   I was hoping it would work well, but tried a s...
9   Wow! I didn't expect the quality and intonatio...
10  I needed to dim the light from my bedroom cabi...
11  Fits great just as it should. Seemed protectiv...
12  For the price- wait.  Hold on. This is Audio T...
13                Great little set for the money!
14                                      Works fine.
15                                Very easy to use
16  Works as advertised. Perfect for my Line 6 Rel...
17                                    Gift. I don't know
18  I needed a tuner that was easy to use and accu...
19  It's really helpful to my daugther during her ...
20                                              Nice
21  It does what it's supposed to do but it's chea...
22              great little box with a lot of punch!
23                                 Husband loves it!
24  Not sure how I got along without it.  It makes...
25  Killer tone, Light weight, and plays like butt...
```

In [24]:
```
df['compound'] = df['review_body'].apply(lambda x:sia.polarity_scores(x)['compound'])
```

In [25]:
```
dfx = df.loc[:,['product_id','star_rating','compound']]
```

In [28]:
```
dfsummary = dfx.groupby('product_id').mean()
```

In [31]:
```
plt.scatter(dfsummary.star_rating,dfsummary.compound)
```

Out[31]:
```
<matplotlib.collections.PathCollection at 0x7fbee0f79be0>
```

In [32]:
```
dfcount = dfx.groupby('product_id').count()
dfcount
```

Out[32]:

	star_rating	compound
product_id		
0014110024	1	1
0046197141	3	3
0205822908	1	1
0594478944	2	2
0634010263	1	1
...	...	...
B013OL1HK8	1	1
B013XN6B0S	1	1
B0141UGNBE	2	2
B0143H0W3U	2	2
B0143TQTZK	1	1

87939 rows × 2 columns

In [33]:
```
dfsummary['review_count'] = dfcount.compound
```

In [35]:
```
dfsignif = dfsummary.loc[dfsummary['review_count']>=25]
dfsignif = dfsignif.sort_values(by=['star_rating'])
```

In [51]:
```
plt.figure(figsize=(16,9),dpi=96)
x = dfsignif.star_rating
y = dfsignif.compound
plt.scatter(x,y,s=8,color='blue',label='Original Data')
a,b = np.polyfit(x,y,1)
print(a,b)
fun1 = np.polyld((a,b))
plt.plot(x,fun1(x),color='green',label='Linear Fit')
tup = np.polyfit(x,y,2)
print(tup)
#fun2 = np.polyld((tup))
#plt.plot(x,fun2,color='red',label='Quadratic Fit')
plt.legend()
```

```
0.1393234724295763 -0.29328210644917795
[-0.02390052  0.38986601 -0.66637553]
```

Out[51]:
```
<matplotlib.legend.Legend at 0x7fbeb4731940>
```


In [53]:
```
bindata = np.arange(6) + 0.5
plt.hist(df.star_rating,bins=bindata,ec='black',color='wheat',rwidth=0.8)
plt.title('Star Rating Distribution: Musical Instruments')
tix = plt.xticks(range(6))
```


In [ ]:
```

```