In [1]:
```
!jupyter --version
```

```
4.4.0
```

find . | grep csv | while read fullpath; do filename=`echo
`fullpath | sed 's/zip_folder/_/g' | sed 's/data_//g' | sed 's/\.csv//g' | sed 's/\//_/g' | sed 's/\.//g' | sed 's/^_//g'`;
number_of_failures=`cat` fullpath | sed '1d'
| cut -d',' -f5 | grep 1 | wc -l`; number_of_drives=`cat
$fullpath|sed'1d'|cut-d','-f2|sort|uniq|wc-l`; echo filename $number_of_failures number_of_drives >>
failures_vs_drive_count_per_day.dat; done

In [2]:
```
import pandas
print('pandas',pandas.__version__)
import glob
import pickle
import numpy
import datetime
import time
import matplotlib.pyplot as plt
```

```
pandas 0.23.4
```

In [3]:
```
df = pandas.read_csv('data_synthesized_from_csvs/failures_vs_drive_count
_per_day.dat',delimiter=' ',header=None)
df.columns=['date in filename','number of drives removed from service',
'number of unique drives']
df.shape
```

Out[3]:
```
(2092, 3)
```

In [4]:
```
#df.head()
```

In [5]:
```
df['date']=df['date in filename'].apply(lambda x: datetime.datetime.strp
time(x.split('_')[-1],'%Y-%m-%d'))
```

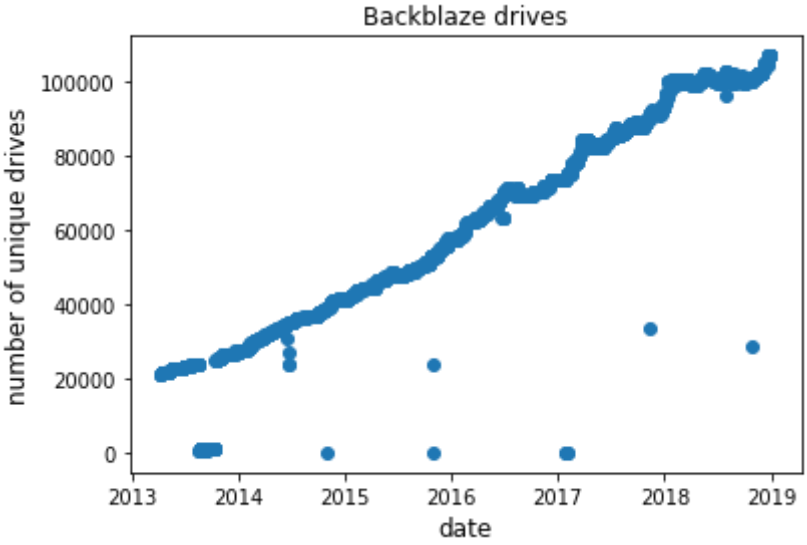
In [6]:
```
#df.head()
```

In [7]:
```
df.drop(['date in filename'], axis=1,inplace=True)
```

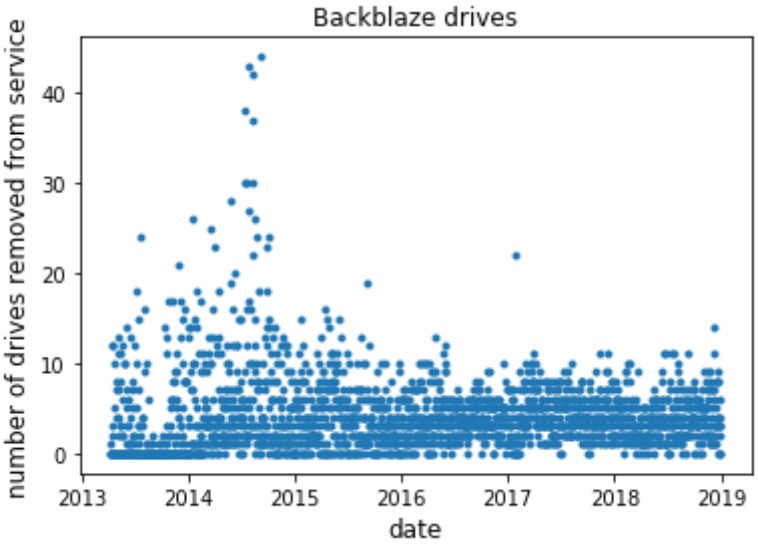
In [8]:
```
df['ratio of drives removed from service to count per day']=df['number o
f drives removed from service']/df['number of unique drives']
```

In [9]:
```
#df.head()
```

In [10]:
```
plt.plot_date(x=df['date'],y=df['number of unique drives'])
plt.xlabel('date',fontsize=12)
plt.ylabel('number of unique drives',fontsize=12);
plt.title('Backblaze drives',fontsize=12);
#plt.xticks(rotation=70);
```



In [11]:
```
plt.plot_date(x=df['date'],y=df['number of drives removed from service'
],markersize=3)
plt.xlabel('date',fontsize=12)
plt.ylabel('number of drives removed from service',fontsize=12);
plt.title('Backblaze drives',fontsize=12);
```



In [12]:
```
plt.plot_date(x=df['date'],y=df['ratio of drives removed from service to count per day'],markersize=3)
plt.xlabel('date',fontsize=12)
plt.ylabel('(removed drive count)/(drive count per day)',fontsize=12);
plt.title('Backblaze drives',fontsize=12);
```