In [2]:
```
%pylab inline
import pandas as pd
import seaborn as sb
```

```
Populating the interactive namespace from numpy and matplotlib
```

In [4]:
```
data = pd.read_csv('D://Dataset/Poc/uber-raw-data-apr14.csv/uber-raw-data-apr14.csv')
```

In [35]:
```
data.head()
```

Out[35]:

	Time	Lat	Lon	Base
0	4/1/2014 0:11	40.7690	-73.9549	B02512
1	4/1/2014 0:17	40.7267	-74.0345	B02512
2	4/1/2014 0:21	40.7316	-73.9873	B02512
3	4/1/2014 0:28	40.7588	-73.9776	B02512
4	4/1/2014 0:33	40.7594	-73.9722	B02512

In [16]:
```
data['Time'] = pd.to_datetime(data['Time'])
```

In [17]:
```
data.tail()
```

Out[17]:

	Time	Lat	Lon	Base
564511	2014-04-30 23:22:00	40.7640	-73.9744	B02764
564512	2014-04-30 23:26:00	40.7629	-73.9672	B02764
564513	2014-04-30 23:31:00	40.7443	-73.9889	B02764
564514	2014-04-30 23:32:00	40.6756	-73.9405	B02764
564515	2014-04-30 23:48:00	40.6880	-73.9608	B02764

In [19]:
```
data['Time'].dt.day.tail()
```

Out[19]:
```
564511    30
564512    30
564513    30
564514    30
564515    30
Name: Time, dtype: int64
```

In [22]:
```
 #days of month
    def get_dom(dt):
    return dt.day

data['dom'] = data['Time'].map(get_dom)
```

In [23]:
```
data.tail()
```

Out[23]:

	Time	Lat	Lon	Base	dom
564511	2014-04-30 23:22:00	40.7640	-73.9744	B02764	30
564512	2014-04-30 23:26:00	40.7629	-73.9672	B02764	30
564513	2014-04-30 23:31:00	40.7443	-73.9889	B02764	30
564514	2014-04-30 23:32:00	40.6756	-73.9405	B02764	30
564515	2014-04-30 23:48:00	40.6880	-73.9608	B02764	30

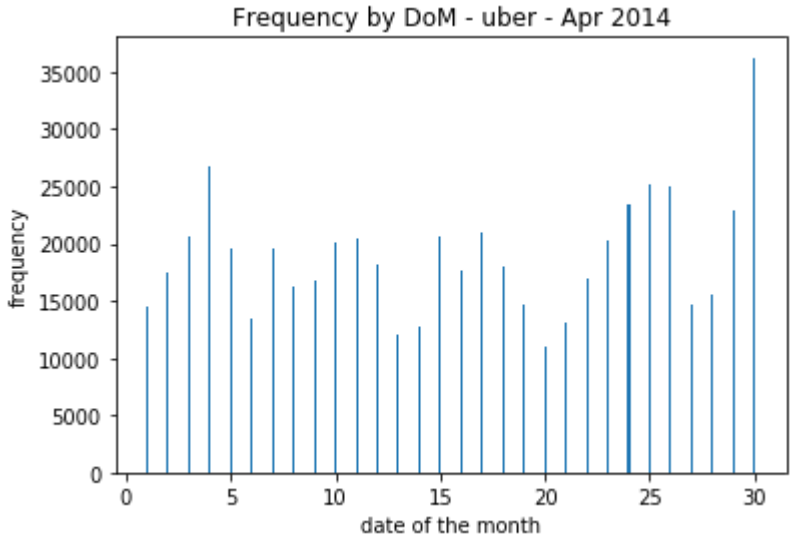
In [31]:
```
def get_weekday(dt):
    return dt.weekday()

data['weekday'] = data['Time'].map(get_weekday)

def get_hour(dt):
    return dt.hour

data['hour'] = data['Time'].map(get_hour)

data.tail()
```

Out[31]:

	Time	Lat	Lon	Base	dom	weekday	hour
564511	2014-04-30 23:22:00	40.7640	-73.9744	B02764	30	2	23
564512	2014-04-30 23:26:00	40.7629	-73.9672	B02764	30	2	23
564513	2014-04-30 23:31:00	40.7443	-73.9889	B02764	30	2	23
564514	2014-04-30 23:32:00	40.6756	-73.9405	B02764	30	2	23
564515	2014-04-30 23:48:00	40.6880	-73.9608	B02764	30	2	23

In [36]:
```
hist(data.dom, bins = 30, rwidth =.10,range=(0.5,30.5))
xlabel('date of the month')
ylabel('frequency')
title('Frequency by DoM - uber - Apr 2014')
```

Out[36]: Text(0.5,1,'Frequency by DoM - uber - Apr 2014')

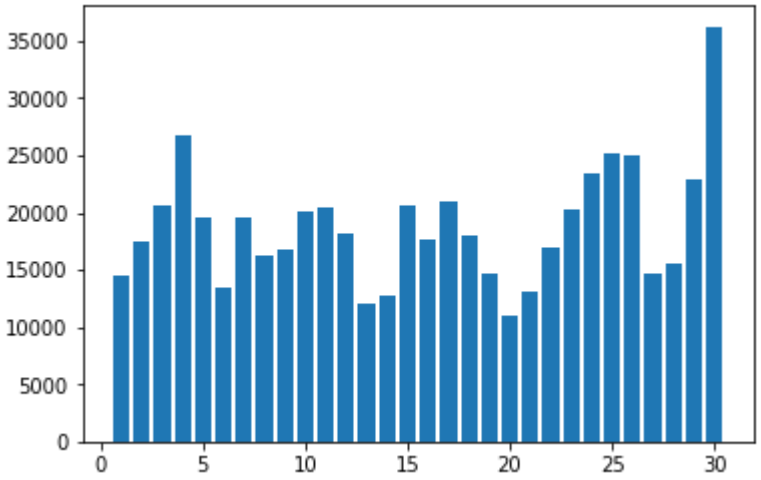


In [41]:
```
def count_rows(rows):
    return len(rows)

by_date = data.groupby('dom').apply(count_rows)
by_date
```

Out[41]:
```
dom
1     14546
2     17474
3     20701
4     26714
5     19521
6     13445
7     19550
8     16188
9     16843
10    20041
11    20420
12    18170
13    12112
14    12674
15    20641
16    17717
17    20973
18    18074
19    14602
20    11017
21    13162
22    16975
23    20346
24    23352
25    25095
26    24925
27    14677
28    15475
29    22835
30    36251
dtype: int64
```

In [43]:
```
bar(range(1,31), by_date)
```

Out[43]: <Container object of 30 artists>

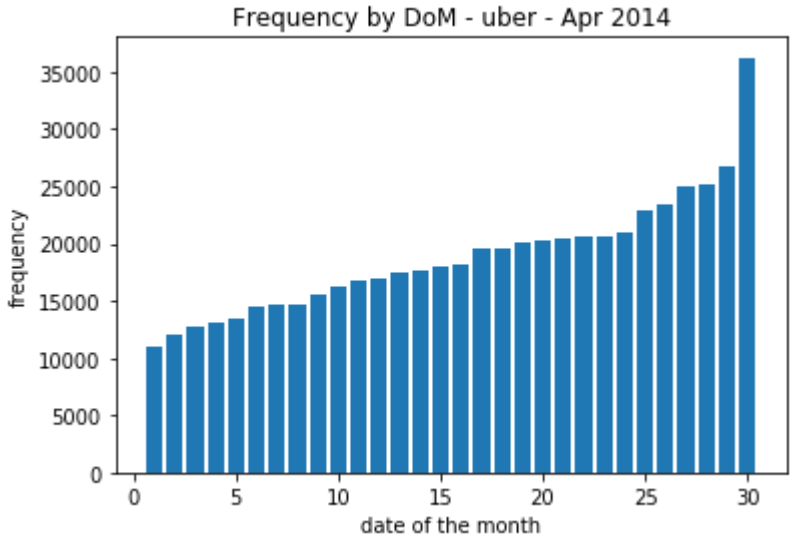


In [44]:
```
by_date_sorted = by_date.sort_values()
by_date_sorted
```

Out[44]:
```
dom
20    11017
13    12112
14    12674
21    13162
6     13445
1     14546
19    14602
27    14677
28    15475
8     16188
9     16843
22    16975
2     17474
16    17717
18    18074
12    18170
5     19521
7     19550
10    20041
23    20346
11    20420
15    20641
3     20701
17    20973
29    22835
24    23352
26    24925
25    25095
4     26714
30    36251
dtype: int64
```

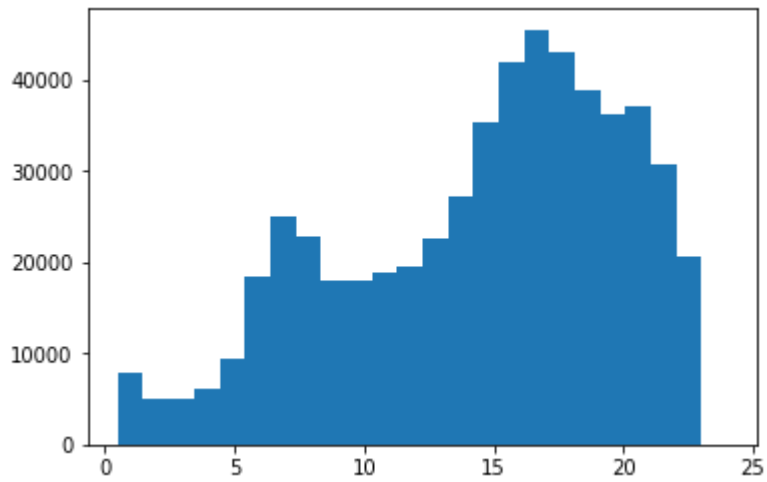
In [46]:
```
bar(range(1, 31), by_date_sorted)
#xticks(range(1,31), by_date_sorted.index)
xlabel('date of the month')
ylabel('frequency')
title('Frequency by DoM - uber - Apr 2014')
```

Out[46]: Text(0.5,1,'Frequency by DoM - uber - Apr 2014')



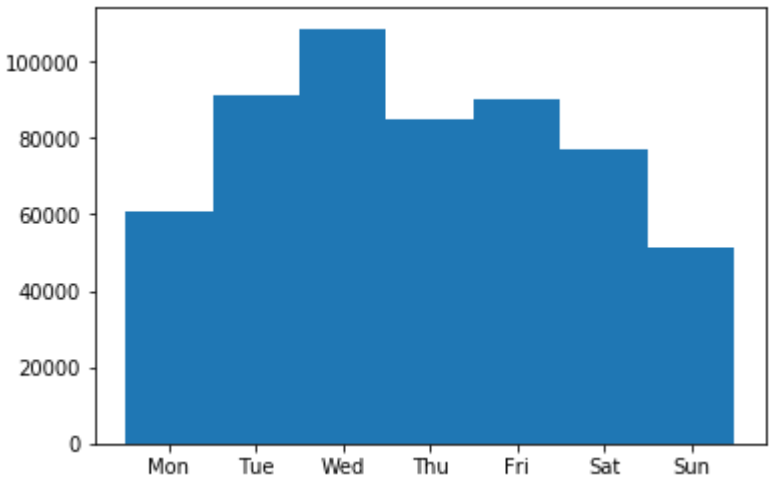
In [47]:
```
hist(data.hour, bins=24, range=(.5, 24))
```

Out[47]:
```
(array([  7769.,   4935.,   5040.,   6095.,   9476.,  18498.,  24924.,
        22843.,  17939.,  17865.,  18774.,  19425.,  22603.,  27190.,
        35324.,  42003.,  45475.,  43003.,  38923.,  36244.,  36964.,
        30645.,  20649.,      0.]),
 array([  0.5       ,   1.47916667,   2.45833333,   3.4375    ,
          4.41666667,   5.39583333,   6.375     ,   7.35416667,
          8.33333333,   9.3125    ,  10.29166667,  11.27083333,
         12.25      ,  13.22916667,  14.20833333,  15.1875    ,
         16.16666667,  17.14583333,  18.125     ,  19.10416667,
         20.08333333,  21.0625    ,  22.04166667,  23.02083333,  24.
]),
 <a list of 24 Patch objects>)
```



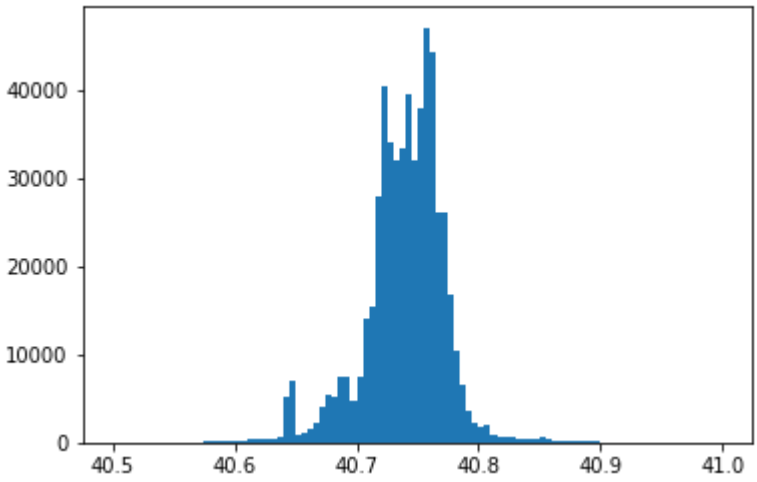
In [49]:
```
hist(data.weekday, bins=7, range =(-.5,6.5))
xticks(range(7), 'Mon Tue Wed Thu Fri Sat Sun'.split())
```

Out[49]:
```
([<matplotlib.axis.XTick at 0x214708fa748>,
  <matplotlib.axis.XTick at 0x214708e9908>,
  <matplotlib.axis.XTick at 0x2147092ca20>,
  <matplotlib.axis.XTick at 0x21470952ef0>,
  <matplotlib.axis.XTick at 0x2147095a588>,
  <matplotlib.axis.XTick at 0x2147095abe0>,
  <matplotlib.axis.XTick at 0x21470961278>],
 <a list of 7 Text xticklabel objects>)
```



In [50]:
```
hist(data['Lat'], bins=100, range = (40.5, 41))
```

```
Out[50]: (array([  4.00000000e+00,   2.00000000e+00,   6.00000000e+00,
                   5.00000000e+00,   5.00000000e+00,   1.00000000e+00,
                   3.00000000e+00,   2.00000000e+00,   4.00000000e+00,
                   1.00000000e+01,   7.00000000e+00,   1.20000000e+01,
                   9.00000000e+00,   7.00000000e+00,   2.00000000e+01,
                   1.05000000e+02,   7.90000000e+01,   1.02000000e+02,
                   9.80000000e+01,   1.78000000e+02,   2.31000000e+02,
                   1.72000000e+02,   3.23000000e+02,   3.54000000e+02,
                   4.21000000e+02,   3.75000000e+02,   4.82000000e+02,
                   5.24000000e+02,   5.25300000e+03,   7.09200000e+03,
                   8.24000000e+02,   1.05900000e+03,   1.59500000e+03,
                   2.28800000e+03,   3.93300000e+03,   5.36000000e+03,
                   5.16700000e+03,   7.42900000e+03,   7.38600000e+03,
                   4.74000000e+03,   7.43400000e+03,   1.39920000e+04,
                   1.54070000e+04,   2.79080000e+04,   4.05000000e+04,
                   3.40080000e+04,   3.19180000e+04,   3.32570000e+04,
                   3.95900000e+04,   3.19070000e+04,   3.79670000e+04,
                   4.70420000e+04,   4.42880000e+04,   2.60320000e+04,
                   2.61140000e+04,   1.67620000e+04,   1.04810000e+04,
                   6.47600000e+03,   3.58600000e+03,   2.12600000e+03,
                   1.82800000e+03,   1.90200000e+03,   8.75000000e+02,
                   6.81000000e+02,   6.05000000e+02,   7.34000000e+02,
                   4.41000000e+02,   3.34000000e+02,   4.27000000e+02,
                   2.99000000e+02,   5.27000000e+02,   3.43000000e+02,
                   2.75000000e+02,   2.49000000e+02,   1.79000000e+02,
                   1.74000000e+02,   1.95000000e+02,   2.26000000e+02,
                   2.88000000e+02,   1.38000000e+02,   6.40000000e+01,
                   4.90000000e+01,   3.50000000e+01,   5.00000000e+01,
                   4.90000000e+01,   3.10000000e+01,   5.40000000e+01,
                   3.20000000e+01,   4.00000000e+01,   2.20000000e+01,
                   2.60000000e+01,   2.20000000e+01,   2.00000000e+01,
                   1.10000000e+01,   2.80000000e+01,   1.60000000e+01,
                   1.60000000e+01,   2.80000000e+01,   1.60000000e+01,
                   2.30000000e+01]),
            array([ 40.5  ,  40.505,  40.51 ,  40.515,  40.52 ,  40.525,  40.53 ,
                    40.535,  40.54 ,  40.545,  40.55 ,  40.555,  40.56 ,  40.565,
                    40.57 ,  40.575,  40.58 ,  40.585,  40.59 ,  40.595,  40.6  ,
                    40.605,  40.61 ,  40.615,  40.62 ,  40.625,  40.63 ,  40.635,
                    40.64 ,  40.645,  40.65 ,  40.655,  40.66 ,  40.665,  40.67 ,
                    40.675,  40.68 ,  40.685,  40.69 ,  40.695,  40.7  ,  40.705,
                    40.71 ,  40.715,  40.72 ,  40.725,  40.73 ,  40.735,  40.74 ,
                    40.745,  40.75 ,  40.755,  40.76 ,  40.765,  40.77 ,  40.775,
                    40.78 ,  40.785,  40.79 ,  40.795,  40.8  ,  40.805,  40.81 ,
                    40.815,  40.82 ,  40.825,  40.83 ,  40.835,  40.84 ,  40.845,
                    40.85 ,  40.855,  40.86 ,  40.865,  40.87 ,  40.875,  40.88 ,
                    40.885,  40.89 ,  40.895,  40.9  ,  40.905,  40.91 ,  40.915,
                    40.92 ,  40.925,  40.93 ,  40.935,  40.94 ,  40.945,  40.95 ,
                    40.955,  40.96 ,  40.965,  40.97 ,  40.975,  40.98 ,  40.985,
                    40.99 ,  40.995,  41.   ]),
            <a list of 100 Patch objects>)
```



In [51]:
```
hist(data['Lon'], bins=100, range = (-74.1, -73.9));
```