Preliminaries

In [10]:
```
import datetime
import pandas as pd
import numpy as np
import sqlite3
import pandas.io.sql as psql
```

In [11]:
```
conn = sqlite3.connect("/home/florian/workspace/VelibApp/velib_2.db")
df = psql.frame_query('select * from data', con=conn)
print 'loaded dataframe from disk. # Records: ', len(df)
```

```
loaded dataframe from disk. # Records:  227492
```

Database description

In [9]:
```
df.describe()
```

Out[9]:
```
<class 'pandas.core.frame.DataFrame'>
Index: 8 entries, count to max
Data columns (total 5 columns):
station_number           8  non-null values
bike_stands              8  non-null values
available_bikes          8  non-null values
available_bike_stands    8  non-null values
last_update              8  non-null values
dtypes: float64(5)
```

In [34]:
```
stations = df.station_number.unique()
print stations
```

```
[ 31705.  10042.   8020. ...,  19121.   1020.  10007.]
```

Converting timestamps to datetime objects

In [44]:
```
def timestamps_2_datetime(ts):
    return map(lambda ts: datetime.datetime.fromtimestamp(ts / 1000.), ts)
```

In [45]:
```
print timestamps_2_datetime(df[df.station_number == 31705].last_update)[:10]
```

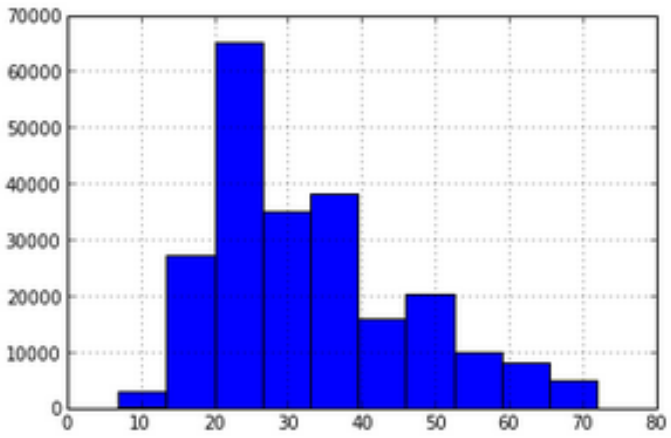
```
[datetime.datetime(2013, 6, 3, 23, 52, 13), datetime.datetime(2013, 6, 3, 23,
52, 13), datetime.datetime(2013, 6, 4, 0, 12, 20), datetime.datetime(2013, 6, 4,
0, 12, 20), datetime.datetime(2013, 6, 4, 0, 22, 24), datetime.datetime(2013, 6,
4, 0, 42, 32), datetime.datetime(2013, 6, 4, 0, 42, 32), datetime.datetime(2013,
6, 4, 17, 36, 36), datetime.datetime(2013, 6, 4, 17, 36, 36),
datetime.datetime(2013, 6, 4, 17, 46, 40)]
```

Global statistics

Below, we plot the histogram for the number of bike stands distribution in the dataset.

In [22]:
```
df.bike_stands.hist()
```

Out[22]:
```
<matplotlib.axes.AxesSubplot at 0xc24a3ac>
```



Statistics for the station down the road

In [39]:
```
my_station = df[df.station_number == 31705]
```

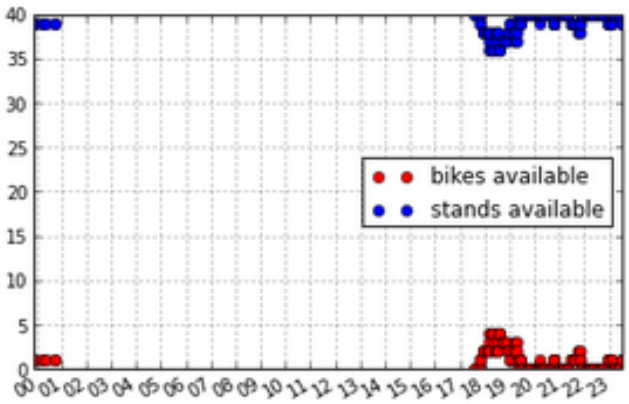
In [40]:
```
print my_station[:10]
```

```
       station_number  bike_stands  available_bikes  available_bike_stands  \
0               31705           50                1                     39
1225            31705           50                1                     39
2450            31705           50                1                     39
3675            31705           50                1                     39
4900            31705           50                1                     39
6125            31705           50                1                     39
7350            31705           50                1                     39
8575            31705           50                0                     40
9798            31705           50                0                     40
11021           31705           50                0                     40

        last_update
0      1.370296e+12
1225   1.370296e+12
2450   1.370298e+12
3675   1.370298e+12
4900   1.370298e+12
6125   1.370299e+12
7350   1.370299e+12
8575   1.370360e+12
9798   1.370360e+12
11021  1.370361e+12
```

In [70]:
```
fig = figure()
t = timestamps_2_datetime(my_station.last_update)
plot_date(t, my_station.available_bikes, 'ro', label="bikes available")
plot_date(t, my_station.available_bike_stands, 'bo', label="stands available")
ax = gca()
ax.xaxis.set_major_locator(mpl.dates.HourLocator())
hoursFmt = DateFormatter('%H')
ax.xaxis.set_major_formatter(hoursFmt)
fig.autofmt_xdate()
grid(True)
legend(loc=5)
```

Out[70]: <matplotlib.legend.Legend at 0xcca48ac>



In [ ]: