In [252]:
```
%load_ext sql
```

```
The sql extension is already loaded. To reload it, use:
  %reload_ext sql
```

In [253]:
```
%sql mysql://prod:nerd@52.2.153.189/rental_nerd
```

Out[253]:
```
u'Connected: prod@rental_nerd'
```

In [254]:
```
result = %sql (SELECT \
properties.id as "property_id", \
property_transaction_logs.id as "transaction_log_id", \
properties.*, \
property_transaction_logs.* \
FROM \
properties, \
property_transactions, \
property_transaction_logs \
WHERE \
properties.id = property_transactions.property_id AND \
property_transactions.property_transaction_log_id = property_transaction_logs.id AND \
property_transactions.transaction_type = 'rental')

data = result.DataFrame()
```

```
801 rows affected.
```

In [255]:
```
from time import gmtime, strftime
result.csv(filename=strftime("%Y%m%d")+ " rentals.csv")
```

Out[255]: CSV results (./files/20150907 rentals.csv)

In [256]:
```
# imports
import pandas as pd
import matplotlib.pyplot as plt
# follow the usual sklearn pattern: import, instantiate, fit
from sklearn.linear_model import LinearRegression
import numpy as np

# this allows plots to appear directly in the notebook
%matplotlib inline

data.head()
```

Out[256]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	id	price	tran
0	1	1	1	65 Roan Place	Other (Woodside)	4	3	3500	climbsf_rented	http://www.climbsf.com/for-rent/65-roan-pl/	...	1	10500	clos
1	2	2	2	166 Salada Avenue	Peninsula (Pacifica)	2	2	993	climbsf_rented	http://www.climbsf.com/for-rent/salada-avenue-...	...	2	2400	clos
2	3	3	3	1500 Park Avenue #105	East Bay (Emeryville)	1	1	1300	climbsf_rented	http://www.climbsf.com/for-rent/1500-park-ave-...	...	3	2300	clos
3	4	4	4	1201 Pine Street #257	East Bay (West Oakland)	1	1	815	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-st-257/	...	4	2200	clos
4	5	5	5	1523 Brunswig Lane	East Bay (Emeryville)	2	1	745	climbsf_rented	http://www.climbsf.com/for-rent/1523-brunswig-...	...	5	2500	clos

5 rows × 26 columns

In [257]:
```
import datetime

Date_final = [0.1] * len(data)

for x in range(0,len(data)):
    data
    if data["date_closed"][x] is not None :
        # print " row: "+ `x` + ": using date_rented"
        # data.ix['Date_final',x]
        Date_final[x] = data["date_closed"][x]

    elif data["date_listed"][x] is not None :
        # print " row: "+ `x` + ": using date_listed"
        Date_final[x] = data["date_listed"][x]
    else:
        Date_final[x] = data["date_closed"][2]
        print " row: "+ `x` + ": we are screwed"



data['Date'] = pd.to_datetime(Date_final)

data.head()
```

Out[257]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	price	transac
0	1	1	1	65 Roan Place	Other (Woodside)	4	3	3500	climbsf_rented	http://www.climbsf.com/for-rent/65-roan-pl/	...	10500	closed
1	2	2	2	166 Salada Avenue	Peninsula (Pacifica)	2	2	993	climbsf_rented	http://www.climbsf.com/for-rent/salada-avenue-...	...	2400	closed
2	3	3	3	1500 Park Avenue #105	East Bay (Emeryville)	1	1	1300	climbsf_rented	http://www.climbsf.com/for-rent/1500-park-ave-...	...	2300	closed
3	4	4	4	1201 Pine Street #257	East Bay (West Oakland)	1	1	815	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-st-257/	...	2200	closed
4	5	5	5	1523 Brunswig Lane	East Bay (Emeryville)	2	1	745	climbsf_rented	http://www.climbsf.com/for-rent/1523-brunswig-...	...	2500	closed

5 rows × 27 columns

In [258]:
```
# create neighborhoods from lat/long coordinates
import fiona
import shapely as shapely
from geopandas import GeoSeries, GeoDataFrame
from shapely.geometry import Point
from shapely.geometry import asShape
```

In [259]:
```
# create a column of GeoSeries - each house should be represented by a point
pts = GeoSeries([Point(x, y) for x, y in zip(data['longitude'], data['latitude'])])
data['latlong'] = pts
```

In [260]:
```
# filter out any outliers, defined as rent >$10k or >2,500 sq ft, or not in SF

print "Entries before filter: " + `len(data)`
data = data[  (data.sqft <= 2500)
            & (data.price <= 8000)
            & (data.price != 0)
            & (data.bedrooms <= 4)
            & (data.bathrooms <= 3)
            & (data.sqft != 0)
            & (data.address > '(Undisclosed Address) San Francisco, CA 94999')  # eliminate (Undisclosed)
            & ((data.source == 'climbsf_rented') | (data.Date > datetime.datetime(2015, 7, 1)) )] # eliminate listings old
er than 2 months

print "Entries after filter: " + `len(data)`
data.head()
```

```
Entries before filter: 801
Entries after filter: 525
```

Out[260]:

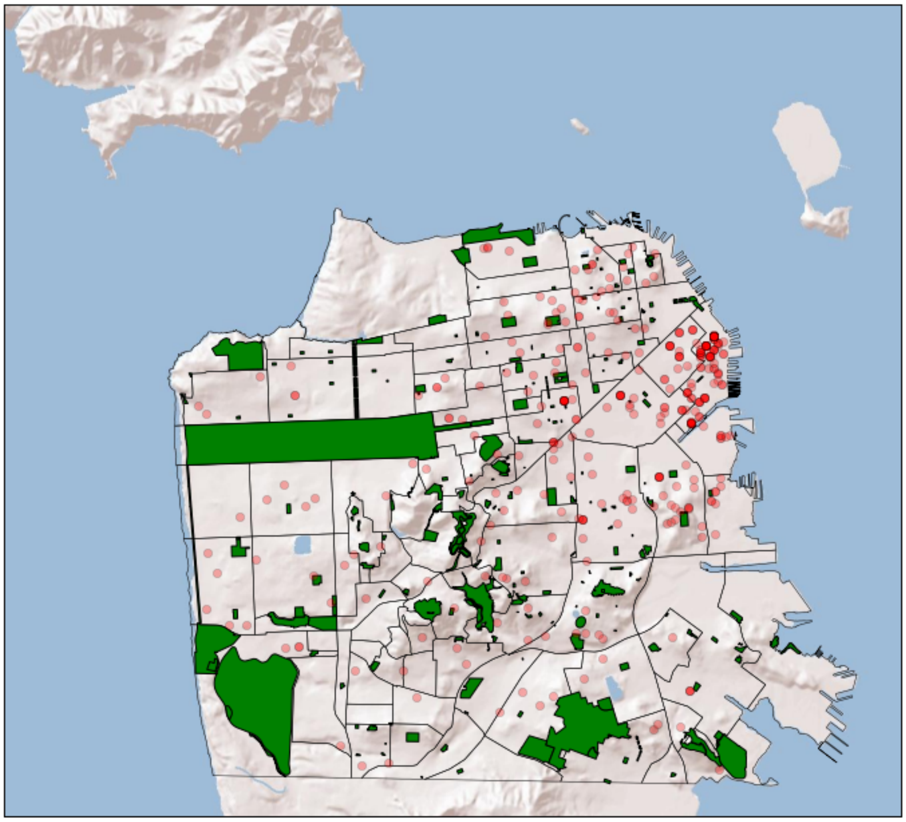
	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	transaction_st
1	2	2	2	166 Salada Avenue	Peninsula (Pacifica)	2	2	993	climbsf_rented	http://www.climbsf.com/for-rent/salada-avenue-...	...	closed
2	3	3	3	1500 Park Avenue #105	East Bay (Emeryville)	1	1	1300	climbsf_rented	http://www.climbsf.com/for-rent/1500-park-ave-...	...	closed
3	4	4	4	1201 Pine Street #257	East Bay (West Oakland)	1	1	815	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-st-257/	...	closed
4	5	5	5	1523 Brunswig Lane	East Bay (Emeryville)	2	1	745	climbsf_rented	http://www.climbsf.com/for-rent/1523-brunswig-...	...	closed
5	6	6	6	1201 Pine Street #132	East Bay (Oakland)	1	1	1072	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-stre...	...	closed

5 rows × 28 columns

In [261]:

```
from mpl_toolkits.basemap import Basemap
import fiona
from matplotlib.patches import Polygon
from matplotlib.collections import PatchCollection

fig = plt.figure(figsize=(12,12))
ax = fig.add_subplot(111)

# Create the Basemap
event_map = Basemap(projection='merc',
                    resolution='h', epsg=2227,
                    lat_0 = 37.7, lon_0=-122.4, # Map center
                    llcrnrlon=-122.55, llcrnrlat=37.7, # Lower left corner
                    urcrnrlon=-122.35, urcrnrlat=37.85) # Upper right corner

# Draw important features
event_map.arcgisimage(service='World_Shaded_Relief', xpixels = 1500, verbose= True)

# add neighborhoods
event_map.readshapefile(
    'data/Realtor_Neighborhoods_4326/hoods_4326', 'SF', color='black', zorder=2)

# add parks
event_map.readshapefile(
    'data/RPD_Parks_4326/parks_4326', 'parks', color='none', zorder=2)

# fill in parks in green
patches   = []

for shape in event_map.parks:
    patches.append( Polygon(np.array(shape), True) )

ax.add_collection(PatchCollection(patches, facecolor= 'green', zorder=2))


# create array storing lats and longs
listing_coords = zip(data.latitude,data.longitude)

# Draw the points on the map:
for longitude, latitude in listing_coords:
    x, y = event_map(latitude, longitude) # Convert lat, long to y,x
    event_map.plot(x,y, 'ro', alpha=0.3)

plt.show()
```

http://server.arcgisonline.com/ArcGIS/rest/services/World_Shaded_Relief/MapServer/export?bbox=5968621.97922,2083843.65958,6027551.68158,2137245.61137&bboxSR=2227&imageSR=2227&size=1500,1359&dpi=96&format=png32&f=image



In [262]:
```
shaped_neighborhood = ['None'] * len(data)
latlong = data['latlong'].values

#with fiona.open('data/Planning_Neighborhoods_4326/planning_hoods_4326.shp') as fiona_collection:
with fiona.open('data/Realtor_Neighborhoods_4326/hoods_4326.shp') as fiona_collection:
    for hood in fiona_collection:
        print "checking for listings in: " + hood["properties"]["nbrhood"]
        # Use Shapely to create the polygon
        shape = asShape( hood['geometry'] )

        for row in range(0,len(data)):
            point = latlong[row] # longitude, latitude

            if shaped_neighborhood[row] != 'None':
                continue

            if shape.contains(point):
                #print `row` + ": Found " + data.address[row] + " in hood " + hood["properties"]["nbrhood"]
                shaped_neighborhood[row] = hood["properties"]["nbrhood"]  # for Planning Neighborhoods, "neighborho"

data['shaped_neighborhood'] = shaped_neighborhood
data.head()
```

```
checking for listings in: Alamo Square
checking for listings in: Anza Vista
checking for listings in: Balboa Terrace
checking for listings in: Bayview
checking for listings in: Bernal Heights
checking for listings in: Buena Vista Park/Ashbury Heights
checking for listings in: Central Richmond
checking for listings in: Central Sunset
checking for listings in: Clarendon Heights
checking for listings in: Corona Heights
checking for listings in: Cow Hollow
checking for listings in: Crocker Amazon
checking for listings in: Diamond Heights
checking for listings in: Downtown
checking for listings in: Duboce Triangle
```

```
checking for listings in: Eureka Valley / Dolores Heights
checking for listings in: Excelsior
checking for listings in: Financial District/Barbary Coast
checking for listings in: Yerba Buena
checking for listings in: Forest Hill
checking for listings in: Forest Hills Extension
checking for listings in: Forest Knolls
checking for listings in: Glen Park
checking for listings in: Golden Gate Heights
checking for listings in: Golden Gate Park
checking for listings in: Haight Ashbury
checking for listings in: Hayes Valley
checking for listings in: Hunters Point
checking for listings in: Ingleside
checking for listings in: Ingleside Heights
checking for listings in: Ingleside Terrace
checking for listings in: Inner Mission
checking for listings in: Inner Parkside
checking for listings in: Inner Richmond
checking for listings in: Inner Sunset
checking for listings in: Jordan Park / Laurel Heights
checking for listings in: Lake Street
checking for listings in: Lake Shore
checking for listings in: Lakeside
checking for listings in: Lone Mountain
checking for listings in: Lower Pacific Heights
checking for listings in: Marina
checking for listings in: Merced Heights
checking for listings in: Merced Manor
checking for listings in: Midtown Terrace
checking for listings in: Miraloma Park
checking for listings in: Mission Bay
checking for listings in: Mission Dolores
checking for listings in: Mission Terrace
checking for listings in: Monterey Heights
checking for listings in: Mount Davidson Manor
checking for listings in: Noe Valley
checking for listings in: North Beach
checking for listings in: North Panhandle
checking for listings in: North Waterfront
checking for listings in: Oceanview
checking for listings in: Outer Mission
checking for listings in: Outer Parkside
checking for listings in: Outer Richmond
checking for listings in: Outer Sunset
checking for listings in: Pacific Heights
checking for listings in: Parkside
checking for listings in: Cole Valley/Parnassus Heights
checking for listings in: Pine Lake Park
checking for listings in: Portola
checking for listings in: Potrero Hill
checking for listings in: Presidio
checking for listings in: Presidio Heights
checking for listings in: Russian Hill
checking for listings in: Saint Francis Wood
checking for listings in: Sea Cliff
checking for listings in: Silver Terrace
checking for listings in: South Beach
checking for listings in: South of Market
checking for listings in: Stonestown
checking for listings in: Sunnyside
checking for listings in: Telegraph Hill
checking for listings in: Twin Peaks
checking for listings in: Van Ness/Civic Center
checking for listings in: Visitacion Valley
checking for listings in: West Portal
checking for listings in: Western Addition
checking for listings in: Westwood Highlands
checking for listings in: Westwood Park
checking for listings in: Lincoln Park
checking for listings in: Sherwood Forest
checking for listings in: Tenderloin
checking for listings in: Central Waterfront/Dogpatch
checking for listings in: Candlestick Point
checking for listings in: Bayview Heights
checking for listings in: Little Hollywood
checking for listings in: Nob Hill
```

Out[262]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	date_listed	da
1	2	2	2	166 Salada Avenue	Peninsula (Pacifica)	2	2	993	climbsf_rented	http://www.climbsf.com/for-rent/salada-avenue-...	...	None	20
2	3	3	3	1500 Park Avenue #105	East Bay (Emeryville)	1	1	1300	climbsf_rented	http://www.climbsf.com/for-rent/1500-park-ave-...	...	None	20
3	4	4	4	1201 Pine Street #257	East Bay (West Oakland)	1	1	815	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-st-257/	...	None	20
4	5	5	5	1523 Brunswig Lane	East Bay (Emeryville)	2	1	745	climbsf_rented	http://www.climbsf.com/for-rent/1523-brunswig-...	...	None	20
5	6	6	6	1201 Pine Street #132	East Bay (Oakland)	1	1	1072	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-stre...	...	None	20

5 rows × 29 columns

In [263]:

```
# calculate distance to the nearest park

dist_to_park = [999999] * len(data)  # fill with dummy values to be filtered out
closest_park = ['None'] * len(data)

latlong = data['latlong'].values

with fiona.open('data/RPD_Parks_4326/parks_4326.shp') as fiona_collection:

    for park in fiona_collection:
        park_name = park["properties"]["map_park_n"]

        # Use Shapely to create the polygon
        shape = asShape( park['geometry'] )

        if shape.area < 0.0000005:
            print park_name + ' is too small at ' + `shape.area`
            continue

        print "checking for proximity to: " + park_name + ' with area: ' + `shape.area`



        for row in range(0,len(data)):
            point = latlong[row] # longitude, latitude

            dist = shape.distance(point)
            if dist < dist_to_park[row]:
                dist_to_park[row] = dist
                closest_park[row] = park_name


data['dist_to_park'] = dist_to_park
data['closest_park'] = closest_park
data.head()
```

```
checking for proximity to: Maritime Plaza with area: 8.333416041071514e-07
checking for proximity to: Victoria Manalo Draves Park with area: 1.0414108652372042e-06
checking for proximity to: Crocker Amazon Playground with area: 2.335089291946154e-05
Dearborn Community Garden is too small at 6.384420351348188e-08
checking for proximity to: Angelo J. Rossi Playground with area: 2.6786514833549074e-06
checking for proximity to: Lake Merced Park with area: 0.0002516841071582249
checking for proximity to: Telegraph Hill/Pioneer Park with area: 2.0250235515636755e-06
Head & Brotherood Mini Park is too small at 2.2842949387512543e-07
checking for proximity to: Balboa Natural Area with area: 7.628411322265208e-07
Broadway Tunnel West Mini Park is too small at 5.138207003055309e-08
checking for proximity to: Alamo Square with area: 5.253994710747828e-06
24th & York Mini Park is too small at 4.939779756558067e-08
checking for proximity to: Mt. Davidson Park with area: 1.684320788581448e-05
checking for proximity to: Union Square with area: 1.076244689464398e-06
checking for proximity to: Midtown Terrace Playground with area: 4.999909744839785e-06
```

```
checking for proximity to: Precita Park with area: 9.150017154907253e-07
checking for proximity to: Brooks Park with area: 1.5354832683227666e-06
checking for proximity to: Adam Rodgers Park with area: 1.1334586337861152e-06
checking for proximity to: John McLaren Park with area: 0.00012926543047010142
Russian Hill Open Space is too small at 3.983418875406988e-07
Beideman & O'Farrell Mini Park is too small at 2.4518591981378515e-08
Mullen & Peralta Mini Park is too small at 1.863497809635393e-07
checking for proximity to: Alta Plaza Park with area: 4.928986328119509e-06
Ina Coolbrith Park is too small at 3.573055370089921e-07
Dogpatch/Miller Memorial Garden is too small at 1.0767516984810728e-07
checking for proximity to: Louis Sutter Playground with area: 5.75807638780225e-06
Joseph Conrad Mini Park is too small at 6.387403297412363e-08
Geneva Avenue Strip is too small at 9.594872996689325e-08
29th & Diamond Open Space is too small at 3.372585423149393e-07
Diamond & Farnum Open Space is too small at 2.875622968890608e-08
Cow Hollow Playground is too small at 6.188292073135964e-08
Coso & Precita Mini Park is too small at 6.226468008590154e-08
Duncan & Castro Open Space is too small at 2.397736033385076e-07
checking for proximity to: Hawk Hill with area: 2.005993246809935e-06
checking for proximity to: Herz Playground with area: 2.760805220617754e-06
checking for proximity to: Dorothy Erskine Park with area: 6.607523788025506e-07
Argonne Playground is too small at 3.393491772166019e-07
Mt. Olympus is too small at 8.490259437786529e-08
checking for proximity to: Edgehill Mountain with area: 9.640723779051192e-07
Alioto Mini Park is too small at 6.69852889445249e-08
checking for proximity to: Kite Hill with area: 1.1875097477699357e-06
Ridgetop Plaza is too small at 1.1515761986200772e-07
Broadway Tunnel East Mini Park is too small at 1.3502462748961951e-08
Page Street Community Garden is too small at 3.1356754823853284e-08
checking for proximity to: Julius Kahn Playground with area: 5.1258299532461146e-06
Turk & Hyde Mini Park is too small at 4.6649126665192646e-08
South Park is too small at 3.538110270782186e-07
checking for proximity to: Alice Chalmers Playground with area: 6.942960445399324e-07
checking for proximity to: Bay View Playground with area: 1.4072676457780056e-06
Allyne Park is too small at 3.121407833365481e-07
checking for proximity to: Kelloch & Velasco Mini Park with area: 7.16399919524842e-07
checking for proximity to: Lower Great Highway with area: 8.729838704318009e-06
checking for proximity to: Lincoln Park with area: 4.662071648425129e-05
checking for proximity to: Palace of Fine Arts with area: 8.020874096457907e-06
Garden for the Environment is too small at 2.0989658853910262e-07
checking for proximity to: O'Shaughnessy Hollow with area: 1.5511591298802525e-06
checking for proximity to: Raymond Kimball Playground with area: 2.2448715899110407e-06
checking for proximity to: Rocky Outcrop with area: 6.655589313730214e-07
Lakeview & Ashton Mini Park is too small at 2.1294079064927143e-07
checking for proximity to: Pine Lake Park with area: 1.2728325681807494e-05
checking for proximity to: Tank Hill with area: 1.1891897857782891e-06
checking for proximity to: Twin Peaks with area: 2.2588218661228207e-05
checking for proximity to: Interior Greenbelt with area: 8.833365824548397e-06
checking for proximity to: Alice Marble Tennis Courts with area: 1.0747651460104084e-06
checking for proximity to: Gilman Playground with area: 2.154401696864766e-06
LeConte Mini Park is too small at 4.236266147561731e-08
Bernal Heights Recreation Center is too small at 3.0815571911415067e-07
Saturn Street Steps is too small at 4.0055327091039384e-08
Parque Ninos Unidos is too small at 2.211990867730097e-07
Brotherhood & Chester Mini Park is too small at 2.433045629588723e-07
checking for proximity to: Margaret S.Hayward Playground with area: 2.0802331148249555e-06
checking for proximity to: Lafayette Park with area: 4.758211976565886e-06
checking for proximity to: West Portal Playground with area: 7.91111267136949e-07
Cabrillo Playground is too small at 3.698138023197936e-07
Fulton Playground is too small at 3.3938951434000644e-07
Muriel Leff Mini Park is too small at 8.489096744844787e-08
DuPont Tennis Courts is too small at 3.4084039892843635e-07
Rochambeau Playground is too small at 3.4198262024003587e-07
Richmond Playground is too small at 3.415126814512935e-07
Richmond Recreation Center is too small at 3.415068303049054e-07
checking for proximity to: Laurel Hill Playground with area: 6.089492550035908e-07
Presidio Library Mini Park is too small at 2.809407061476052e-07
Clipper Terrace Community Garden is too small at 1.4469511568558045e-07
checking for proximity to: Bernal Heights Park with area: 1.0897388388402991e-05
checking for proximity to: Jefferson Square with area: 2.33567063020698e-06
checking for proximity to: Joseph L. Alioto Performing Arts Piazza with area: 1.8344681473706315e-06
Sgt. John Macaulay Park is too small at 8.607810824056863e-08
Tenderloin Recreation Center is too small at 2.5087909111528274e-07
Father Alfred E. Boeddeker Park is too small at 4.0101717135107635e-07
Japantown Peace Plaza is too small at 2.9535521920095474e-07
Cottage Row Mini Park is too small at 6.485418078796215e-08
Golden Gate & Steiner Mini Park is too small at 3.116880401674962e-08
Fillmore & Turk Mini Park is too small at 8.382260549364783e-08
checking for proximity to: Buchanan Street Mall with area: 7.500063481941431e-07
Hayes Valley Playground is too small at 2.5266061456052383e-07
```

```
Bush & Broderick Mini Park is too small at 7.508461101101129e-08
Koshland Park is too small at 3.396203753478794e-07
Page & Laguna Mini Park is too small at 6.266618556463739e-08
checking for proximity to: Duboce Park with area: 1.7829926972418978e-06
checking for proximity to: Mission Dolores Park with area: 5.876475694598816e-06
Corwin Community Garden is too small at 3.8341911275295404e-08
Hooker Alley Community Garden is too small at 2.2836794080212884e-08
Wolfe Lane Community Garden is too small at 2.2563223714512406e-08
Arlington Community Garden is too small at 5.364918180901646e-08
checking for proximity to: McCoppin Square with area: 2.9969047779956e-06
checking for proximity to: Silver Terrace Playground with area: 2.264230334870624e-06
Seward Mini Park is too small at 1.549957407864033e-07
Coleridge Mini Park is too small at 8.566820967602584e-08
Grand View Open Space is too small at 2.6796914466680403e-07
Howard & Langton Mini Park is too small at 9.709752616773353e-08
Kidpower Park is too small at 9.599446011860595e-08
Lessing & Sears Mini Park is too small at 6.02432966499595e-08
Merced Heights Playground is too small at 4.2464806649908347e-07
Michelangelo Playground is too small at 1.8060654444965552e-07
checking for proximity to: Miraloma Playground with area: 9.119819006646205e-07
checking for proximity to: Moscone Recreation Center with area: 4.997183843937509e-06
Noe Valley Courts is too small at 3.830340417178794e-07
Noe & Beaver Mini Park is too small at 2.4478831833529147e-08
Presidio Heights Playground is too small at 1.8150183026832098e-07
checking for proximity to: Portsmouth Square with area: 5.340307854230594e-07
checking for proximity to: Hamilton Playground with area: 1.3341225308880821e-06
checking for proximity to: San Francisco Zoo with area: 5.4405376783965375e-05
checking for proximity to: Potrero del Sol with area: 1.8053735114240483e-06
checking for proximity to: Jackson Playground with area: 1.8242606624131285e-06
checking for proximity to: Ocean View Playground with area: 4.258622772265438e-06
checking for proximity to: Palega Playground with area: 2.054386398670562e-06
checking for proximity to: Excelsior Playground with area: 6.67006055881951e-07
checking for proximity to: Sunset Playground with area: 1.3597577138521218e-06
checking for proximity to: James Rolph Jr. Playground with area: 1.2121405372026234e-06
Jose Coronado Playground is too small at 3.2371151247172425e-07
Mission Recreation Center is too small at 2.608685305998709e-07
Randolph & Bright Mini Park is too small at 3.6029250397106904e-08
SOMA Rec Center is too small at 4.2145503824411936e-07
checking for proximity to: South Sunset Playground with area: 1.5401308973238756e-06
checking for proximity to: Upper Noe Playground with area: 1.0381102355219405e-06
checking for proximity to: Visitation Valley Playground with area: 8.48986899989034e-07
checking for proximity to: Washington Square with area: 9.347217770841064e-07
Woh Hei Yuen is too small at 1.2789996896308982e-07
checking for proximity to: Youngblood-Coleman Playground with area: 2.5357593471459666e-06
checking for proximity to: Yacht Harbor and Marina Green with area: 3.231134582830828e-05
checking for proximity to: White Crane Springs with area: 1.1073737387334295e-06
checking for proximity to: West Sunset Playground with area: 7.133414594268702e-06
Washington & Hyde Mini Park is too small at 6.221538857634999e-08
checking for proximity to: Visitacion Valley Greenway with area: 8.81371971108938e-07
Utah & 18th Mini Park is too small at 4.343266499250789e-08
Topaz Open Space is too small at 3.423480799651553e-07
checking for proximity to: Sunnyside Playground with area: 9.724364900759024e-07
Sunnyside Conservatory is too small at 1.0515769308771232e-07
checking for proximity to: St. Mary's Playground with area: 5.694006549214278e-06
Selby & Palou Mini Park is too small at 1.1872118081607924e-07
Rowing Club/Dolphin Club is too small at 4.60845624293841e-07
Roosevelt & Henry Stairs is too small at 1.4198591221465986e-07
checking for proximity to: Rolph Nicol Playground with area: 1.2593956598252343e-06
Prentiss Mini Park is too small at 1.6313307902247952e-08
checking for proximity to: Parkside Square with area: 3.438322576931479e-06
checking for proximity to: Sigmund Stern Recreation Grove with area: 1.3974558908596191e-05
checking for proximity to: Carl Larsen Park with area: 2.7243147512701698e-06
Chinese Recreation Center is too small at 2.704098381241927e-07
Collins P. Huntington Park is too small at 4.4315716565598836e-07
checking for proximity to: Douglass Playground with area: 3.0823503486130344e-06
checking for proximity to: Esprit Park with area: 7.574665914718511e-07
checking for proximity to: Eureka Valley Playground with area: 7.984097233621879e-07
Fay Park is too small at 1.03866259822834e-07
checking for proximity to: Garfield Square with area: 1.212833616121347e-06
checking for proximity to: Grattan Playground with area: 6.290014799290612e-07
Helen Wills Playground is too small at 3.3299896572400123e-07
checking for proximity to: Hilltop Park with area: 1.4309569233219596e-06
Hyde & Vallejo Mini Park is too small at 4.1737911163458235e-08
checking for proximity to: J.P. Murphy Playground with area: 5.075767457149975e-07
checking for proximity to: Joe DiMaggio Playground with area: 1.0024301532477977e-06
checking for proximity to: Joseph Lee Recreation Center with area: 7.664467655298112e-07
checking for proximity to: Junipero Serra Playground with area: 6.345033624773983e-07
Juri Commons is too small at 1.3412630334481023e-07
Ogden Terrace Community Garden is too small at 6.382151549268159e-08
Good Prospect Community Garden is too small at 4.740625760162154e-08
```

```
checking for proximity to: Potrero Hill Recreation Center with area: 3.954131045330463e-06
Connecticut Friendship Garden is too small at 6.356166227007175e-08
Arkansas Friendship Garden is too small at 6.184298860952733e-08
Park St. Garden is too small at 1.9416889809571332e-08
checking for proximity to: States Street Playground with area: 1.1370414417168124e-06
Peixotto Playground is too small at 3.28064886522635e-07
checking for proximity to: Palou & Phelps Park with area: 1.0879776997433266e-06
checking for proximity to: Mountain Lake Park with area: 5.293425861514069e-06
checking for proximity to: Park Presidio Blvd with area: 6.910526855622263e-06
checking for proximity to: Monster Park with area: 3.521373886029161e-05
checking for proximity to: McKinley Square with area: 1.1253099183935466e-06
Berkeley Way Open Space is too small at 3.4384150803455556e-07
Crags Court Garden is too small at 1.7037270435435663e-07
checking for proximity to: Franklin Square with area: 2.141825398225521e-06
checking for proximity to: Justin Herman/Embarcadero Plaza with area: 1.723382969476932e-06
Fairmont Plaza is too small at 3.0634611233234674e-07
Everson & Digby Lots is too small at 4.977273857460294e-07
checking for proximity to: Corona Heights with area: 5.482770373126953e-06
Willie "Woo Woo" Wong Playground is too small at 2.522925042603624e-07
checking for proximity to: Cayuga Playground with area: 1.0609226465998123e-06
Cayuga & Lamartine Mini Park is too small at 5.0822072730282624e-08
checking for proximity to: Balboa Park with area: 9.827603031314997e-06
15th Avenue Steps is too small at 2.102308171560447e-07
checking for proximity to: Grand View Park with area: 1.619898535879658e-06
Potrero Hill Mini Park is too small at 1.256617518115001e-07
checking for proximity to: Sharp Park with area: 0.0001692401703086459
checking for proximity to: Buena Vista Park with area: 1.4931297203760588e-05
Portola Open Space is too small at 3.362587124559358e-07
Patricia's Green in Hayes Valley is too small at 1.981698977139144e-07
checking for proximity to: Mission Playground with area: 8.406025594465439e-07
Joost & Baden Mini Park is too small at 5.846754171688136e-08
checking for proximity to: India Basin/Shoreline Park with area: 4.7845113132668446e-06
checking for proximity to: Holly Park with area: 3.369833522037716e-06
checking for proximity to: Golden Gate Heights Park with area: 2.8770804247042982e-06
checking for proximity to: George Christopher Playground with area: 2.816520793526006e-06
checking for proximity to: Ferry Park with area: 1.7640843120144974e-06
Chesnut & Kearny Open Space is too small at 3.3511959257647285e-08
checking for proximity to: Bay View Park with area: 1.6605940985886756e-05
10th Ave & Clement Mini Park is too small at 3.4324663732568383e-07
checking for proximity to: Glen Canyon Park with area: 3.2244055895260945e-05
checking for proximity to: Aptos Playground with area: 1.989083799220228e-06
checking for proximity to: Walter Haas Playground with area: 1.9146687587795825e-06
checking for proximity to: Golden Gate Park with area: 0.00042493138816672194
checking for proximity to: Camp Mather with area: 0.00013989822888904908
checking for proximity to: Billy Goat Hill with area: 1.5176999724497963e-06
checking for proximity to: St. Mary's Square with area: 5.87095588332992e-07
checking for proximity to: Little Hollywood Park with area: 5.435273816410738e-07
```

Out[263]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	days_on_mark
1	2	2	2	166 Salada Avenue	Peninsula (Pacifica)	2	2	993	climbsf_rented	http://www.climbsf.com/for-rent/salada-avenue-...	...	NaN
2	3	3	3	1500 Park Avenue #105	East Bay (Emeryville)	1	1	1300	climbsf_rented	http://www.climbsf.com/for-rent/1500-park-ave-...	...	NaN
3	4	4	4	1201 Pine Street #257	East Bay (West Oakland)	1	1	815	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-st-257/	...	NaN
4	5	5	5	1523 Brunswig Lane	East Bay (Emeryville)	2	1	745	climbsf_rented	http://www.climbsf.com/for-rent/1523-brunswig-...	...	NaN
5	6	6	6	1201 Pine Street #132	East Bay (Oakland)	1	1	1072	climbsf_rented	http://www.climbsf.com/for-rent/1201-pine-stre...	...	NaN

5 rows × 31 columns

In [265]:
```
print "Entries before filter: " + `len(data)`


# filter out listings over one month old
data = data[ (data.dist_to_park < 999) & (data.shaped_neighborhood != 'None') ]


print "Entries after filter: " + `len(data)`
```

```
Entries before filter: 525
Entries after filter: 371
```

In [ ]:

In [266]:
```
# create year dummy variables (because date isn't very intuitive variable)
data["Year"] = pd.DatetimeIndex(data["Date"]).to_period('Y')

# create dummy variables using get_dummies, then exclude the first dummy column
year_dummies = pd.get_dummies(data.Year, prefix='Year').iloc[:, :-1]

# print out baseline neighborhood
base_area = pd.get_dummies(data.shaped_neighborhood, prefix='neighborhood').iloc[:, 0:1].columns[0]
print('Base neighborhood: %s' % base_area)

# create dummy variables using get_dummies, then exclude the first dummy column
area_dummies = pd.get_dummies(data.shaped_neighborhood, prefix='neighborhood').iloc[:, 1:]

# concatenate the dummy variable columns onto the original DataFrame (axis=0 means rows, axis=1 means columns)
data = pd.concat([data, area_dummies, year_dummies], axis=1)

data.head()
```

```
Base neighborhood: neighborhood_Alamo Square
```

Out[266]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	neighborl Valley
37	38	38	38	480 Mission Bay Blvd. North #1007	San Francisco (Mission Bay)	2	2	1576	climbsf_rented	http://www.climbsf.com/for-rent/480-mission-ba...	...	0
38	39	39	39	74 New Montgomery #412	San Francisco (Financial District)	1	1	870	climbsf_rented	http://www.climbsf.com/for-rent/74-new-montgom...	...	0
41	42	42	42	16 Jessie St #407	San Francisco (Downtown)	1	1	450	climbsf_rented	http://www.climbsf.com/for-rent/16-jessie-st-407/	...	0
43	44	44	44	55 Page Street #814	San Francisco (Hayes Valley)	1	1	865	climbsf_rented	http://www.climbsf.com/for-rent/55-page-street...	...	0
44	45	45	45	235 Berry Street #102	San Francisco (Mission Bay)	2	2	1255	climbsf_rented	http://www.climbsf.com/for-rent/235-berry-stre...	...	0

5 rows × 99 columns

In [267]:
```
# create "on the park" variable if the house is within 100 yards of the park

def on_the_park (row):
   if row['dist_to_park'] < 0.0008 :
      return True
   return False


data['on_the_park'] = data.apply (lambda row: on_the_park (row),axis=1)


print "Number of houses deemed 'on the park'" + `len(data[data.on_the_park == True])`
```

```
Number of houses deemed 'on the park'24
```

In [268]:
```
# FACTORING BY YEAR AND NEIGHBORHOOD
# Thesis: Neighborhoods influence valuations as a multiplier, rather than a constant.
# a square foot in SOMA is worth more than a square foot in Portrero by X%
# New model will look like this:
#      Price = B_1 x (SOMA Coeff * Year Coeff * Sqft) + intercept
#      $3,900 = B_1 x (1.20% * 1.15% * 2,023 sqft) + intercept
# where B_1 represents the price per square foot in base year and base neighborhood
# I will ignore intercepts for now FIXME
# calculate the coefficients for the following matrix and save them for later regressions
#                  SOMA    Mission    Portrero    Intercept
#  Price/SQFT      $1.23    $0.59       $0.88        $_.__

# create Price per square foot

price_per_foot = data.price / data.sqft
price_per_foot.name = 'price_per_foot'
data = pd.concat([data, price_per_foot], axis=1)

data.head()
```

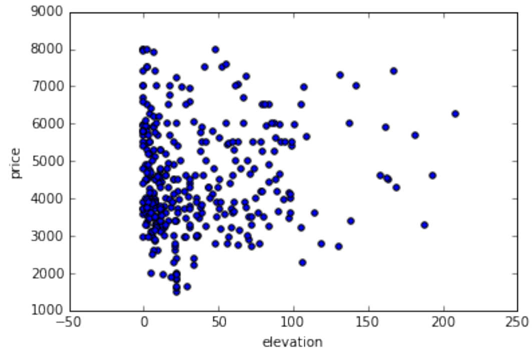
Out[268]:

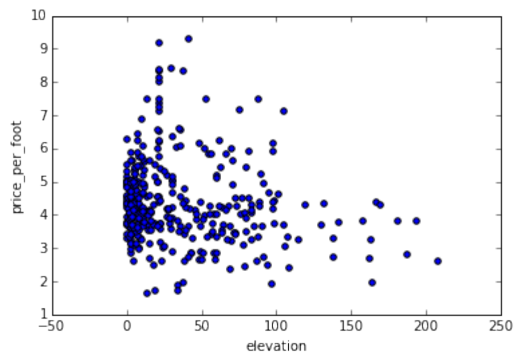
	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	neighborl Addition
37	38	38	38	480 Mission Bay Blvd. North #1007	San Francisco (Mission Bay)	2	2	1576	climbsf_rented	http://www.climbsf.com/for-rent/480-mission-ba...	...	0
38	39	39	39	74 New Montgomery #412	San Francisco (Financial District)	1	1	870	climbsf_rented	http://www.climbsf.com/for-rent/74-new-montgom...	...	0
41	42	42	42	16 Jessie St #407	San Francisco (Downtown)	1	1	450	climbsf_rented	http://www.climbsf.com/for-rent/16-jessie-st-407/	...	0
43	44	44	44	55 Page Street #814	San Francisco (Hayes Valley)	1	1	865	climbsf_rented	http://www.climbsf.com/for-rent/55-page-street...	...	0
44	45	45	45	235 Berry Street #102	San Francisco (Mission Bay)	2	2	1255	climbsf_rented	http://www.climbsf.com/for-rent/235-berry-stre...	...	0

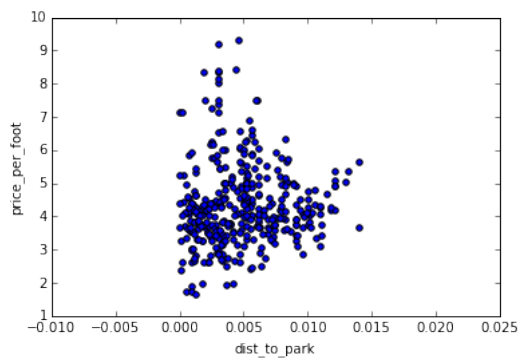
5 rows × 101 columns

In [269]:
```
# visualize the relationship between the features and the response using scatterplots
data.plot(kind='scatter', x='elevation', y='price')
data.plot(kind='scatter', x='elevation', y='price_per_foot')
data.plot(kind='scatter', x='dist_to_park', y='price_per_foot')
```

Out[269]: <matplotlib.axes._subplots.AxesSubplot at 0x11c6f8450>







In [270]:
```
class ListTable(list):
    """ Overridden list class which takes a 2-dimensional list of
        the form [[1,2,3],[4,5,6]], and renders an HTML Table in
        IPython Notebook. """

    def _repr_html_(self):
        html = ["<table>"]
        for row in self:
            html.append("<tr>")

            for col in row:
                html.append("<td>{0}</td>".format(col))

            html.append("</tr>")
        html.append("</table>")
        return ''.join(html)
```

In [271]:
```
feature_cols = area_dummies.columns

X = data[feature_cols]
y = data.price_per_foot

# instantiate, fit
lm = LinearRegression()
lm.fit(X, y)

# print coefficients
# The mean square error
print("Residual sum of squares: %.2f"
      % np.mean((lm.predict(X) - y) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % lm.score(X, y))

# print raw results
print("Base area is %s: $%.2f" % (base_area, lm.intercept_))

zip(feature_cols,lm.coef_)

table = ListTable()

dtype = [('Neighborhood', 'S100'), ('$ per square', float)]

# round to pennies
round_coef = map(round,lm.coef_,[2]*len(lm.coef_))
x = np.array(zip(feature_cols, round_coef),dtype=dtype)
x.T
x = np.sort(x,axis=0,order='$ per square')

table.append(['Neighborhood','$ per square (+/-)'])
for i in x:
    table.append(i)

table
```

```
Residual sum of squares: 0.88
Variance score: 0.43
Base area is neighborhood_Alamo Square: $4.17
```

Out[271]:

Neighborhood	$ per square (+/-)
neighborhood_Candlestick Point	-2.44
neighborhood_Mount Davidson Manor	-1.75
neighborhood_Ingleside	-1.7
neighborhood_Visitacion Valley	-1.66
neighborhood_Miraloma Park	-1.56
neighborhood_Bayview Heights	-1.55
neighborhood_Westwood Highlands	-1.48
neighborhood_Stonestown	-1.45
neighborhood_Central Richmond	-1.38
neighborhood_Diamond Heights	-1.36
neighborhood_Silver Terrace	-1.33
neighborhood_Portola	-1.22
neighborhood_Anza Vista	-1.11
neighborhood_Bayview	-0.99
neighborhood_Excelsior	-0.98
neighborhood_North Beach	-0.98
neighborhood_Outer Richmond	-0.9
neighborhood_West Portal	-0.84
neighborhood_Lake Shore	-0.81
neighborhood_Downtown	-0.78
neighborhood_Central Sunset	-0.74
neighborhood_Inner Parkside	-0.65

neighborhood_Parkside	-0.64
neighborhood_Haight Ashbury	-0.53
neighborhood_Glen Park	-0.5
neighborhood_Ingleside Heights	-0.5
neighborhood_Buena Vista Park/Ashbury Heights	-0.4
neighborhood_Forest Hills Extension	-0.37
neighborhood_Sunnyside	-0.37
neighborhood_Golden Gate Heights	-0.33
neighborhood_Mission Bay	-0.32
neighborhood_Western Addition	-0.31
neighborhood_Noe Valley	-0.2
neighborhood_Bernal Heights	-0.16
neighborhood_Central Waterfront/Dogpatch	-0.16
neighborhood_Inner Sunset	-0.12
neighborhood_South of Market	-0.1
neighborhood_Potrero Hill	-0.05
neighborhood_Van Ness/Civic Center	0.02
neighborhood_Telegraph Hill	0.05
neighborhood_Marina	0.1
neighborhood_Twin Peaks	0.12
neighborhood_Cole Valley/Parnassus Heights	0.14
neighborhood_Eureka Valley / Dolores Heights	0.16
neighborhood_Lone Mountain	0.18
neighborhood_South Beach	0.2
neighborhood_Inner Mission	0.37
neighborhood_Pacific Heights	0.41
neighborhood_Outer Parkside	0.44
neighborhood_North Panhandle	0.53
neighborhood_Yerba Buena	0.57
neighborhood_Outer Sunset	0.61
neighborhood_Lower Pacific Heights	0.79
neighborhood_Nob Hill	1.06
neighborhood_North Waterfront	1.07
neighborhood_Duboce Triangle	1.26
neighborhood_Russian Hill	1.4
neighborhood_Mission Dolores	1.5
neighborhood_Ingleside Terrace	1.82
neighborhood_Tenderloin	2.4
neighborhood_Hayes Valley	2.79
neighborhood_Financial District/Barbary Coast	4.17

In [272]:
```
full_price = [lm.intercept_] * len(lm.coef_)
full_price += lm.coef_

area_price_per_foot = dict(zip(feature_cols,full_price))
area_price_per_foot[base_area] = lm.intercept_

dtype = [('Neighborhood', 'S100'), ('$ per sqft', float)]

# round to pennies
round_coef = map(round,full_price,[2]*len(full_price))
x = np.array(zip(feature_cols, full_price),dtype=dtype)
x.T
x = np.sort(x,axis=0,order='$ per sqft')

table = ListTable()

table.append(['Neighborhood','$ per sqft'])
for i in x:
    table.append(i)

table
```

Out[272]:

Neighborhood	$ per sqft
neighborhood_Candlestick Point	1.72257479601
neighborhood_Mount Davidson Manor	2.41970021413
neighborhood_Ingleside	2.46875
neighborhood_Visitacion Valley	2.51138053536
neighborhood_Miraloma Park	2.60416666667
neighborhood_Bayview Heights	2.61363636364
neighborhood_Westwood Highlands	2.68181818182
neighborhood_Stonestown	2.71739130435
neighborhood_Central Richmond	2.78787138787
neighborhood_Diamond Heights	2.8085106383
neighborhood_Silver Terrace	2.83464566929
neighborhood_Portola	2.94617857143
neighborhood_Anza Vista	3.06
neighborhood_Bayview	3.17631505701
neighborhood_North Beach	3.18401557636
neighborhood_Excelsior	3.19002351914
neighborhood_Outer Richmond	3.26856060606
neighborhood_West Portal	3.33095238095
neighborhood_Lake Shore	3.3597866078
neighborhood_Downtown	3.38847472785
neighborhood_Central Sunset	3.42199446966
neighborhood_Inner Parkside	3.5148488121
neighborhood_Parkside	3.52552552553
neighborhood_Haight Ashbury	3.63636363636
neighborhood_Glen Park	3.66633366633
neighborhood_Ingleside Heights	3.67028270874
neighborhood_Buena Vista Park/Ashbury Heights	3.76598173516
neighborhood_Sunnyside	3.79500437012
neighborhood_Forest Hills Extension	3.8
neighborhood_Golden Gate Heights	3.83333333333
neighborhood_Mission Bay	3.85094171788
neighborhood_Western Addition	3.8613309027
neighborhood_Noe Valley	3.96901651887

neighborhood_Bernal Heights	4.00397485633
neighborhood_Central Waterfront/Dogpatch	4.01161828726
neighborhood_Inner Sunset	4.04938019782
neighborhood_South of Market	4.06832593038
neighborhood_Potrero Hill	4.12067713874
neighborhood_Van Ness/Civic Center	4.1836066825
neighborhood_Telegraph Hill	4.21519334684
neighborhood_Marina	4.27103404056
neighborhood_Twin Peaks	4.29141716567
neighborhood_Cole Valley/Parnassus Heights	4.30769230769
neighborhood_Eureka Valley / Dolores Heights	4.32698780892
neighborhood_Lone Mountain	4.34580489686
neighborhood_South Beach	4.37114241606
neighborhood_Inner Mission	4.53418567996
neighborhood_Pacific Heights	4.57466585316
neighborhood_Outer Parkside	4.60287755338
neighborhood_North Panhandle	4.69627866503
neighborhood_Yerba Buena	4.73620473644
neighborhood_Outer Sunset	4.7775
neighborhood_Lower Pacific Heights	4.95908374874
neighborhood_Nob Hill	5.22213557385
neighborhood_North Waterfront	5.24109014675
neighborhood_Duboce Triangle	5.42339038438
neighborhood_Russian Hill	5.56417772573
neighborhood_Mission Dolores	5.67139509236
neighborhood_Ingleside Terrace	5.99
neighborhood_Tenderloin	6.57142857143
neighborhood_Hayes Valley	6.96084573298
neighborhood_Financial District/Barbary Coast	8.33333333333

In [273]:

```
# calculate the multipliers for each neighborhood relative to base area
# SOMA_mult = SOMA_per_foot / Base_per_foot

area_mults = [lm.intercept_] * len(lm.coef_)
area_mults = full_price / area_mults - [1]*len(lm.coef_)


dtype = [('Neighborhood', 'S100'), ('Multiplier', float)]

# round to pennies
round_coef = map(round,area_mults,[2]*len(area_mults))
x = np.array(zip(feature_cols, area_mults),dtype=dtype)
x.T
x = np.sort(x,axis=0,order='Multiplier')

table = ListTable()

table.append(['Neighborhood','Multiplier'])
table.append([base_area,0])
for i in x:
    table.append(i)

table
```

Out[273]:

Neighborhood	Multiplier
neighborhood_Alamo Square	0
neighborhood_Candlestick Point	-0.586582048957

neighborhood_Mount Davidson Manor	-0.419271948608
neighborhood_Ingleside	-0.4075
neighborhood_Visitacion Valley	-0.397268671514
neighborhood_Miraloma Park	-0.375
neighborhood_Bayview Heights	-0.372727272727
neighborhood_Westwood Highlands	-0.356363636364
neighborhood_Stonestown	-0.347826086957
neighborhood_Central Richmond	-0.330910866911
neighborhood_Diamond Heights	-0.325957446809
neighborhood_Silver Terrace	-0.31968503937
neighborhood_Portola	-0.292917142857
neighborhood_Anza Vista	-0.2656
neighborhood_Bayview	-0.237684386317
neighborhood_North Beach	-0.235836261674
neighborhood_Excelsior	-0.234394355407
neighborhood_Outer Richmond	-0.215545454545
neighborhood_West Portal	-0.200571428571
neighborhood_Lake Shore	-0.193651214128
neighborhood_Downtown	-0.186766065317
neighborhood_Central Sunset	-0.178721327281
neighborhood_Inner Parkside	-0.156436285097
neighborhood_Parkside	-0.153873873874
neighborhood_Haight Ashbury	-0.127272727273
neighborhood_Glen Park	-0.12007992008
neighborhood_Ingleside Heights	-0.119132149901
neighborhood_Buena Vista Park/Ashbury Heights	-0.0961643835616
neighborhood_Sunnyside	-0.0891989511705
neighborhood_Forest Hills Extension	-0.088
neighborhood_Golden Gate Heights	-0.08
neighborhood_Mission Bay	-0.0757739877095
neighborhood_Western Addition	-0.0732805833512
neighborhood_Noe Valley	-0.0474360354719
neighborhood_Bernal Heights	-0.0390460344801
neighborhood_Central Waterfront/Dogpatch	-0.0372116110567
neighborhood_Inner Sunset	-0.0281487525229
neighborhood_South of Market	-0.0236017767097
neighborhood_Potrero Hill	-0.0110374867024
neighborhood_Van Ness/Civic Center	0.00406560379997
neighborhood_Telegraph Hill	0.0116464032422
neighborhood_Marina	0.0250481697352
neighborhood_Twin Peaks	0.0299401197605
neighborhood_Cole Valley/Parnassus Heights	0.0338461538462
neighborhood_Eureka Valley / Dolores Heights	0.0384770741414
neighborhood_Lone Mountain	0.0429931752458
neighborhood_South Beach	0.0490741798535
neighborhood_Inner Mission	0.0882045631893
neighborhood_Pacific Heights	0.0979198047589

neighborhood_Outer Parkside	0.104690612811
neighborhood_North Panhandle	0.127106879607
neighborhood_Yerba Buena	0.136689136745
neighborhood_Outer Sunset	0.1466
neighborhood_Lower Pacific Heights	0.190180099698
neighborhood_Nob Hill	0.253312537723
neighborhood_North Waterfront	0.25786163522
neighborhood_Duboce Triangle	0.30161369225
neighborhood_Russian Hill	0.335402654175
neighborhood_Mission Dolores	0.361134822167
neighborhood_Ingleside Terrace	0.4376
neighborhood_Tenderloin	0.577142857143
neighborhood_Hayes Valley	0.670602975915
neighborhood_Financial District/Barbary Coast	1.0

In [274]:
```
# calculate the adjusted Sqft (Sqft * Area_mult) for the dataset and add it as a new column to data

# for each property, multiplier is sum of array [area_dummies] x [area_mults]

t = data[area_dummies.columns] * area_mults
t = t.T.sum()

t.name = 'area_multiplier'
t = t + 1
data = pd.concat([data, t], axis=1)

adj_sqft = data.sqft * t
adj_sqft.name = 'area_adj_sqft'
data = pd.concat([data, adj_sqft], axis=1)

data.head()
```

Out[274]:

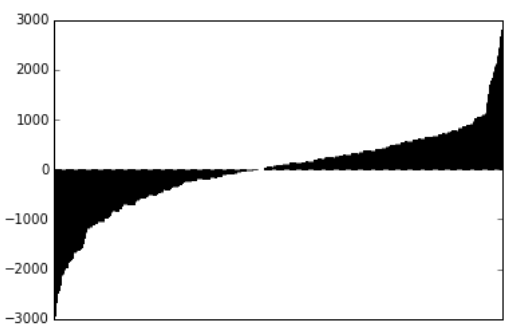
	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	neighborl Buena
37	38	38	38	480 Mission Bay Blvd. North #1007	San Francisco (Mission Bay)	2	2	1576	climbsf_rented	http://www.climbsf.com/for-rent/480-mission-ba...	...	0
38	39	39	39	74 New Montgomery #412	San Francisco (Financial District)	1	1	870	climbsf_rented	http://www.climbsf.com/for-rent/74-new-montgom...	...	1
41	42	42	42	16 Jessie St #407	San Francisco (Downtown)	1	1	450	climbsf_rented	http://www.climbsf.com/for-rent/16-jessie-st-407/	...	1
43	44	44	44	55 Page Street #814	San Francisco (Hayes Valley)	1	1	865	climbsf_rented	http://www.climbsf.com/for-rent/55-page-street...	...	0
44	45	45	45	235 Berry Street #102	San Francisco (Mission Bay)	2	2	1255	climbsf_rented	http://www.climbsf.com/for-rent/235-berry-stre...	...	0

5 rows × 103 columns

In [275]:
```
# run the regression based on area_adj_sqft rather than sqft

# create X and y
feature_cols = [data.area_adj_sqft.name]

X = data[feature_cols]
y = data.price

# instantiate, fit
lm = LinearRegression()
lm.fit(X, y)

# print coefficients
print("Intercept: %.2f" % lm.intercept_)

# The mean square error
print("Residual sum of squares: %.2f"
      % np.mean((lm.predict(X) - y) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % lm.score(X, y))
zip(feature_cols, lm.coef_)

# calculate predictions for the data set and plot errors
predictions = lm.predict(X)
errors = predictions-y
errors.name = 'Error'

# visualize the relationship between the features and the response using scatterplots
errors.sort()
errors.plot(kind='bar').get_xaxis().set_ticks([])
```

```
Intercept: 1484.48
Residual sum of squares: 687494.53
Variance score: 0.64
```

Out[275]: []



In [276]:
```
feature_cols = year_dummies.columns

X = data[feature_cols]
y = data.price_per_foot

# instantiate, fit
lm = LinearRegression()
lm.fit(X, y)

# print coefficients
# The mean square error
print("Residual sum of squares: %.2f"
      % np.mean((lm.predict(X) - y) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % lm.score(X, y))

# print raw results
print lm.intercept_

zip(feature_cols,lm.coef_)
```

```
Residual sum of squares: 1.43
Variance score: 0.07
4.62610888142
```

Out[276]:
```
[(u'Year_1969', -1.7707789321841727),
 (u'Year_2011', -0.45554804751729083),
 (u'Year_2012', -1.1998140607056227),
 (u'Year_2013', -0.81833254407097167),
 (u'Year_2014', -0.52267720979647458)]
```

In [277]:
```
full_price = [lm.intercept_] * len(lm.coef_)
full_price += lm.coef_

year_price_per_foot = dict(zip(feature_cols,full_price))
year_price_per_foot[base_area] = lm.intercept_

print year_price_per_foot
```

```
{u'Year_1969': 2.8553299492385813, u'neighborhood_Alamo Square': 4.6261088814227538, u'Year_2012': 3.4262948207171311, u'Y
ear_2013': 3.8077763373517821, u'Year_2011': 4.1705608339054629, u'Year_2014': 4.1034316716262795}
```

In [278]:
```
# calculate the multipliers for each year relative to base year
# 2014_mult = 2014_per_foot / 2015_per_foot

year_mults = [lm.intercept_] * len(lm.coef_)
year_mults = full_price / year_mults - [1]*len(lm.coef_)

zip(feature_cols, year_mults)
```

Out[278]:
```
[(u'Year_1969', -0.38277934600613461),
 (u'Year_2011', -0.098473265371390784),
 (u'Year_2012', -0.25935707339785341),
 (u'Year_2013', -0.17689435442325663),
 (u'Year_2014', -0.11298419972245133)]
```

In [279]:
```
# calculate the adjusted Sqft (Sqft * Year_mult) for the dataset and add it as a new column to data

# for each property, multiplier is sum of array [year_dummies] x [year_mults]

t = data[year_dummies.columns] * year_mults
t = t.T.sum()

t.name = 'year_multiplier'
t = t + 1
data = pd.concat([data, t], axis=1)

year_adj_sqft = data.area_adj_sqft * t
year_adj_sqft.name = 'adj_sqft'
data = pd.concat([data, year_adj_sqft], axis=1)

data.head()
```

Out[279]:

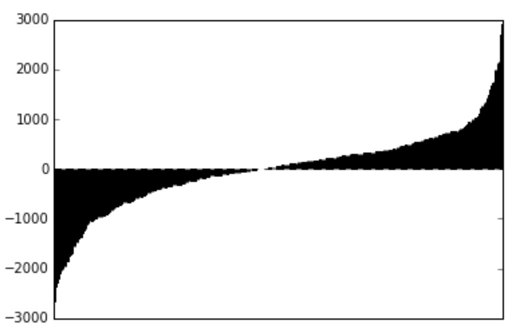
	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	Year_201
37	38	38	38	480 Mission Bay Blvd. North #1007	San Francisco (Mission Bay)	2	2	1576	climbsf_rented	http://www.climbsf.com/for-rent/480-mission-ba...	...	0
38	39	39	39	74 New Montgomery #412	San Francisco (Financial District)	1	1	870	climbsf_rented	http://www.climbsf.com/for-rent/74-new-montgom...	...	1
41	42	42	42	16 Jessie St #407	San Francisco (Downtown)	1	1	450	climbsf_rented	http://www.climbsf.com/for-rent/16-jessie-st-407/	...	1
43	44	44	44	55 Page Street #814	San Francisco (Hayes Valley)	1	1	865	climbsf_rented	http://www.climbsf.com/for-rent/55-page-street...	...	1
44	45	45	45	235 Berry Street #102	San Francisco (Mission Bay)	2	2	1255	climbsf_rented	http://www.climbsf.com/for-rent/235-berry-stre...	...	0

5 rows × 105 columns

In [280]:
```
# run the regression based on year_and_area_adj_sqft rather than area_adj_sqft

# create X and y
feature_cols = ['adj_sqft']

X = data[feature_cols]
y = data.price

# instantiate, fit
lm = LinearRegression()
lm.fit(X, y)

# print coefficients
print lm.intercept_
# The mean square error
print("Residual sum of squares: %.2f"
      % np.mean((lm.predict(X) - y) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % lm.score(X, y))
print zip(feature_cols, lm.coef_)

# calculate predictions for the data set and plot errors
predictions = lm.predict(X)
errors = predictions-y
errors.name = 'Error'

# visualize the relationship between the features and the response using scatterplots
errors.sort(inplace=True)
errors.plot(kind='bar').get_xaxis().set_ticks([])

errors.tail(10)
```

```
1384.73480939
Residual sum of squares: 620084.13
Variance score: 0.68
[('adj_sqft', 2.929422117381312)]
```

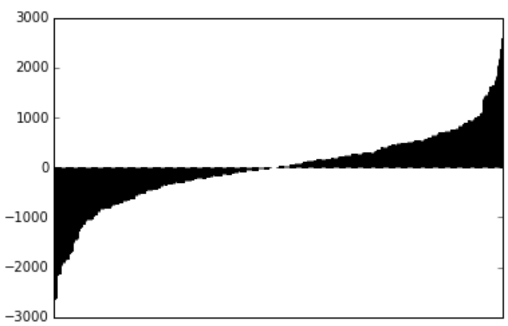
Out[280]:
```
83     1666.003923
524    1721.698995
303    1745.678714
98     1989.616326
43     2001.099988
703    2129.582969
51     2167.478267
536    2435.202459
538    2704.596648
550    2904.596648
Name: Error, dtype: float64
```



In [281]:
```
# create X and y
feature_cols = ['adj_sqft', 'bedrooms', 'bathrooms']

X = data[feature_cols]
y = data.price

# instantiate, fit
lm = LinearRegression()
lm.fit(X, y)

# print coefficients
print("Intercept: %.2f" % lm.intercept_)
# The mean square error
print("Residual sum of squares: %.2f"
      % np.mean((lm.predict(X) - y) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % lm.score(X, y))
print zip(feature_cols, lm.coef_)

# calculate predictions for the data set and plot errors
predictions = lm.predict(X)
errors = predictions-y
errors.name = 'Error'

# visualize the relationship between the features and the response using scatterplots
errors.sort()
errors.plot(kind='bar').get_xaxis().set_ticks([])
```

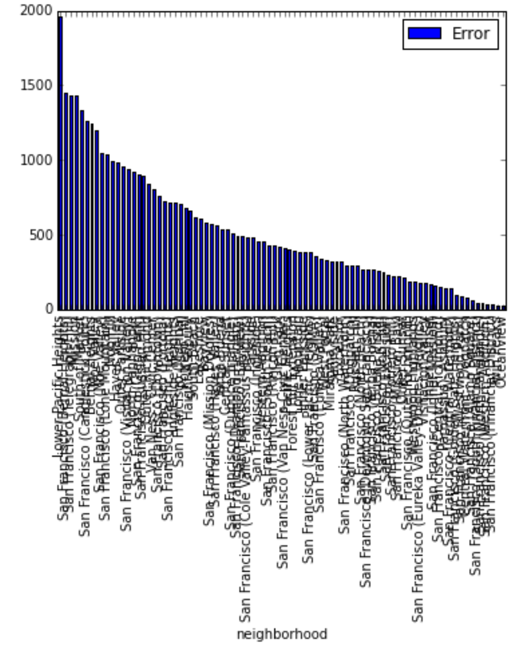
```
Intercept: 1171.61
Residual sum of squares: 579224.66
Variance score: 0.70
[('adj_sqft', 2.4965433868510809), ('bedrooms', 141.4281530806532), ('bathrooms', 292.77503046393372)]
```

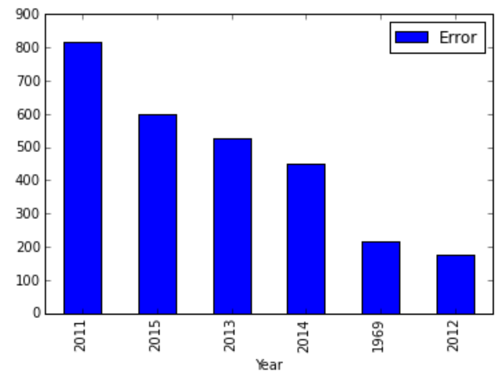
Out[281]: []

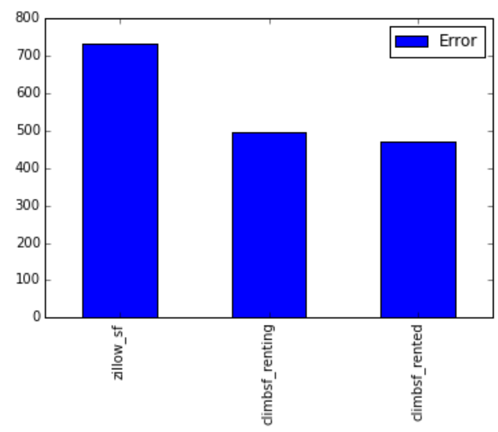


In [282]:
```
# show errors by neighborhood to see if there are any neighborhoods with funky differences

hooderrors = data[['neighborhood']]

errors = predictions-y
errors.name = 'Error'

hooderrors = pd.concat([hooderrors,errors.abs()],axis=1)

hood_group = hooderrors.groupby('neighborhood')

import numpy
def median(lst):
    return numpy.median(numpy.array(lst))

error_avg = hood_group.median()
error_avg.sort(columns='Error',ascending=False).plot(kind='bar')

# show errors by year to see if there are any years with funky differences

yearerrors = data[['Year']]

yearerrors = pd.concat([yearerrors,errors.abs()],axis=1)

year_group = yearerrors.groupby('Year')
error_avg = year_group.mean()
error_avg.sort(columns='Error',ascending=False).plot(kind='bar')

# show errors by source to see if there are any sources have noisy data

srcerrors = data[['source']]

srcerrors = pd.concat([srcerrors,errors.abs()],axis=1)

src_group = srcerrors.groupby('source')
error_avg = src_group.mean()
error_avg.sort(columns='Error',ascending=False).plot(kind='bar')
```

```
Out[282]: <matplotlib.axes._subplots.AxesSubplot at 0x11770e390>
```







In [283]:
```
import csv

table = ListTable()

dtype = [('Effect', 'S100'), ('Coefficient', float)]

# round to pennies
round_coef = map(round,lm.coef_,[6]*len(lm.coef_))
x = np.array(zip(feature_cols, round_coef),dtype=dtype)
x.T
print zip(feature_cols, lm.coef_)
#x = np.sort(x,axis=0,order='Coefficient')

with open('model_features_v1.csv', 'wb') as csvfile:
    modelwriter = csv.writer(csvfile, delimiter=',', quotechar='|', quoting=csv.QUOTE_MINIMAL)

    header = ['Effect','Coefficient']
    table.append(header)
    modelwriter.writerow(header)
    for i in x:
        table.append(i)
        modelwriter.writerow(i)


    table.append(['base_rent', lm.intercept_])


    modelwriter.writerow(['base_rent',lm.intercept_])

table
```

```
[('adj_sqft', 2.4965433868510809), ('bedrooms', 141.4281530806532), ('bathrooms', 292.77503046393372)]
```

Out[283]:

Effect	Coefficient
adj_sqft	2.496543
bedrooms	141.428153
bathrooms	292.77503
base_rent	1171.61068641

In [284]:
```
table = ListTable()

dtype = [('Effect', 'S100'), ('Coefficient', float)]

# round to pennies
round_coef = map(round,(area_mults + [1]*len(area_mults)),[6]*len(area_mults))
x = np.array(zip(area_dummies.columns, round_coef),dtype=dtype)
x.T
x = np.sort(x,axis=0,order='Coefficient')

with open('model_hoods_v1.csv', 'wb') as csvfile:
    hoodwriter = csv.writer(csvfile, delimiter=',', quotechar='|', quoting=csv.QUOTE_MINIMAL)

    header = ['Neighborhood','Multiplier']
    table.append(header)
    hoodwriter.writerow(header)

    for i in x:
        i[0] = i[0][13:]
        table.append(i)
        hoodwriter.writerow(i)

    lastrow = [base_area[13:], 1]
    table.append(lastrow)
    hoodwriter.writerow(lastrow)


table
```

Out[284]:

Neighborhood	Multiplier
Candlestick Point	0.413418
Mount Davidson Manor	0.580728
Ingleside	0.5925
Visitacion Valley	0.602731

Miraloma Park	0.625
Bayview Heights	0.627273
Westwood Highlands	0.643636
Stonestown	0.652174
Central Richmond	0.669089
Diamond Heights	0.674043
Silver Terrace	0.680315
Portola	0.707083
Anza Vista	0.7344
Bayview	0.762316
North Beach	0.764164
Excelsior	0.765606
Outer Richmond	0.784455
West Portal	0.799429
Lake Shore	0.806349
Downtown	0.813234
Central Sunset	0.821279
Inner Parkside	0.843564
Parkside	0.846126
Haight Ashbury	0.872727
Glen Park	0.87992
Ingleside Heights	0.880868
Buena Vista Park/Ashbury Heights	0.903836
Sunnyside	0.910801
Forest Hills Extension	0.912
Golden Gate Heights	0.92
Mission Bay	0.924226
Western Addition	0.926719
Noe Valley	0.952564
Bernal Heights	0.960954
Central Waterfront/Dogpatch	0.962788
Inner Sunset	0.971851
South of Market	0.976398
Potrero Hill	0.988963
Van Ness/Civic Center	1.004066
Telegraph Hill	1.011646
Marina	1.025048
Twin Peaks	1.02994
Cole Valley/Parnassus Heights	1.033846
Eureka Valley / Dolores Heights	1.038477
Lone Mountain	1.042993
South Beach	1.049074
Inner Mission	1.088205
Pacific Heights	1.09792
Outer Parkside	1.104691
North Panhandle	1.127107
Yerba Buena	1.136689

Outer Sunset	1.1466
Lower Pacific Heights	1.19018
Nob Hill	1.253313
North Waterfront	1.257862
Duboce Triangle	1.301614
Russian Hill	1.335403
Mission Dolores	1.361135
Ingleside Terrace	1.4376
Tenderloin	1.577143
Hayes Valley	1.670603
Financial District/Barbary Coast	2.0
Alamo Square	1

In [285]:
```
# show negative errors meaning we expected rents to be higher

error = predictions-y
error.name = 'error'

data = pd.concat([data,error,pd.DataFrame(predictions,columns=['predicted_price'])],axis=1)

data.head()
```

Out[285]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	Year_2013	Year_2014	on_the_park
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

5 rows × 107 columns

In [286]:
```
# filter out overshoot error
overshoot = data[(data.error <= -500)]
columns = data.columns - ['error','latitude', 'longitude', 'address', 'origin_url','price','neighborhood']
overshoot = data.drop(columns,1)
overshoot.sort('error',ascending=True,inplace=True)
overshoot.head(30)
```

Out[286]:

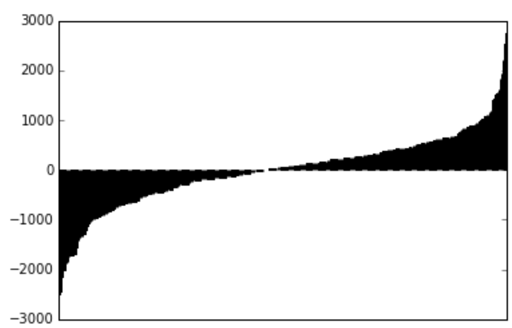
	address	neighborhood	origin_url	latitude	longitude	price	error
544	55 Rodgers St, San Francisco, CA 94103	South of Market	http://www.zillow.com/homedetails/55-Rodgers-S...	37.7750	-122.409	7900	-2618.208212
504	301 Main St UNIT 35A, San Francisco, CA 94105	South Beach	http://www.zillow.com/homedetails/301-Main-St-...	37.7894	-122.391	7950	-2589.015873
487	20th St San Francisco, CA 94110	None	http://www.zillow.com/homedetails/20th-St-San-...	37.7588	-122.416	6200	-2149.111215
465	Grenard Ter San Francisco, CA 94109	None	http://www.zillow.com/homedetails/Grenard-Ter-...	37.8010	-122.423	7500	-2126.092281
212	338 Spear Street #39E	San Francisco (South Beach)	http://www.climbsf.com/for-rent/338-spear-st-39e/	37.7894	-122.391	7975	-2118.336401
496	1840 Broadway, San Francisco, CA 94109	Pacific Heights	http://www.zillow.com/homedetails/1840-Broadwa...	37.7953	-122.428	7580	-1976.677190
724	2434 Pine St, San Francisco, CA 94115	Lower Pacific Heights	http://www.zillow.com/homedetails/2434-Pine-St...	37.7878	-122.437	8000	-1955.817659
243	301 Mission Street #29F	San Francisco (South Beach)	http://www.climbsf.com/for-rent/301-mission-st...	37.7905	-122.396	7975	-1904.997152
463	Taylor St San Francisco, CA 94108	None	http://www.zillow.com/homedetails/Taylor-St-Sa...	37.7919	-122.413	6000	-1891.026828
81	338 Spear Street #39A	San Francisco (South	http://www.climbsf.com/for-rent/338-spear-st-	37.7894	-122.391	6700	-1857.491802

		Beach)	39a/				
384	480 Mission Bay Boulevard North #PH1606	San Francisco (Mission Bay)	http://www.climbsf.com/for-rent/480-mission-ba...	37.7731	-122.393	7500	-1823.567292
72	88 King Street #904	San Francisco (South Beach)	http://www.climbsf.com/for-rent/88-king-st-904/	37.7807	-122.389	6250	-1814.930899
207	2560 Vallejo Street	San Francisco (Pacific Heights)	http://www.climbsf.com/for-rent/2560-vallejo-st/	37.7950	-122.439	7050	-1707.846286
287	401 Harrison Street #3803	San Francisco (Rincon Hill)	http://www.climbsf.com/for-rent/401-harrison-s...	37.7864	-122.392	7225	-1678.062670
522	Bluxome St San Francisco, CA 94107	None	http://www.zillow.com/homedetails/Bluxome-St-S...	37.7760	-122.398	8000	-1662.457957
793	209 Ashbury St, San Francisco, CA 94117	North Panhandle	http://www.zillow.com/homedetails/209-Ashbury-...	37.7736	-122.448	6500	-1657.499628
365	301 Main Street #35F	San Francisco (South Beach)	http://www.climbsf.com/for-rent/301-main-st-35f/	37.7894	-122.391	7000	-1494.967617
700	212 Cortland Ave, San Francisco, CA 94110	Bernal Heights	http://www.zillow.com/homedetails/212-Cortland...	37.7394	-122.419	6500	-1439.678872
346	296 Francisco Street	San Francisco (Telegraph Hill)	http://www.climbsf.com/for-rent/296-francisco-st/	37.8053	-122.410	5675	-1434.971888
715	Columbus Ave San Francisco, CA 94133	None	http://www.zillow.com/homedetails/Columbus-Ave...	37.8016	-122.412	6500	-1430.447478
210	234 Grand View Avenue	San Francisco (Noe Valley)	http://www.climbsf.com/for-rent/234-grand-view...	37.7545	-122.441	7300	-1250.493221
533	1151 Church St, San Francisco, CA 94114	Noe Valley	http://www.zillow.com/homedetails/1151-Church-...	37.7525	-122.427	5475	-1199.552716
511	1st St San Francisco, CA 94105	None	http://www.zillow.com/homedetails/1st-St-San-F...	37.7881	-122.395	5100	-1170.033051
259	425 1st Street #3402	San Francisco (Rincon Hill)	http://www.climbsf.com/for-rent/425-1st-st-3402/	37.7858	-122.392	6600	-1131.634446
454	Lombard St San Francisco, CA 94133	None	http://www.zillow.com/homedetails/Lombard-St-S...	37.8021	-122.419	6700	-1118.783712
136	1837 Jefferson Street	San Francisco (Marina)	http://www.climbsf.com/for-rent/1837-jefferson...	37.8045	-122.443	6200	-1106.911042
701	143 Riverton Dr, San Francisco, CA 94132	Lakeshore	http://www.zillow.com/homedetails/143-Riverton...	37.7316	-122.487	5250	-1070.951624
697	165 Crescent Ave, San Francisco, CA 94110	Bernal Heights	http://www.zillow.com/homedetails/165-Crescent...	37.7353	-122.421	4000	-1050.710700
183	1839 Jefferson Street	San Francisco (Marina)	http://www.climbsf.com/for-rent/1839-jefferson...	37.8048	-122.443	6400	-1045.867719
333	1414 Douglass Street	San Francisco (Noe Valley)	http://www.climbsf.com/for-rent/1414-douglass-st/	37.7452	-122.438	7400	-1041.461640

In [287]:
```
# add squared square footage to the table
squared = data.adj_sqft ** 2
squared.name = 'sqft_squared'

squared_beds = data.bedrooms ** 2
squared_beds.name = 'beds_squared'

data = pd.concat([data, squared, squared_beds], axis=1)
#data = pd.concat([data, squared_beds], axis=1)



# create X and y
feature_cols = ['adj_sqft', 'bedrooms', 'bathrooms','sqft_squared','beds_squared']

X = data[feature_cols]
y = data.price

# instantiate, fit
lm = LinearRegression()
lm.fit(X, y)

# print coefficients
print("Intercept: %.2f" % lm.intercept_)
# The mean square error
print("Residual sum of squares: %.2f"
      % np.mean((lm.predict(X) - y) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % lm.score(X, y))
print zip(feature_cols, lm.coef_)

# calculate predictions for the data set and plot errors
predictions = lm.predict(X)
errors = predictions-y
errors.name = 'Error'

# visualize the relationship between the features and the response using scatterplots
errors.sort()
errors.plot(kind='bar').get_xaxis().set_ticks([])
```

```
Intercept: 216.21
Residual sum of squares: 546455.33
Variance score: 0.71
[('adj_sqft', 4.2383035499430557), ('bedrooms', 243.60686522537156), ('bathrooms', 299.27211690977134), ('sqft_squared', -
0.00071629000021146676), ('beds_squared', -43.438584693259109)]
```

Out[287]: []



In [288]:

```
import statsmodels.formula.api as sm
result = sm.ols(formula="price ~ adj_sqft + bedrooms + bathrooms + dist_to_park + elevation + on_the_park", data=data).fit
()
print result.params
print result.summary()
```

```
Intercept                 988.527703
on_the_park[T.True]       199.671306
adj_sqft                    2.487226
bedrooms                  181.994252
bathrooms                 259.939823
dist_to_park            36055.004600
elevation                  -0.303984
dtype: float64
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.700
Model:                            OLS   Adj. R-squared:                  0.695
Method:                 Least Squares   F-statistic:                     139.6
Date:                Mon, 07 Sep 2015   Prob (F-statistic):           1.14e-90
Time:                        13:51:22   Log-Likelihood:                -2946.2
No. Observations:                 366   AIC:                             5906.
Df Residuals:                     359   BIC:                             5934.
Df Model:                           6
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [95.0% Conf. Int.]
---------------------------------------------------------------------------------------
Intercept             988.5277    161.686      6.114      0.000       670.557  1306.499
on_the_park[T.True]   199.6713    175.378      1.139      0.256      -145.226   544.568
adj_sqft                2.4872      0.138     17.992      0.000         2.215     2.759
bedrooms              181.9943     73.878      2.463      0.014        36.706   327.282
bathrooms             259.9398     97.045      2.679      0.008        69.093   450.787
dist_to_park         3.606e+04   1.55e+04      2.325      0.021      5562.677  6.65e+04
elevation              -0.3040      1.195     -0.254      0.799        -2.654     2.046
==============================================================================
Omnibus:                       20.623   Durbin-Watson:                   2.027
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               40.761
Skew:                           0.309   Prob(JB):                     1.41e-09
Kurtosis:                       4.514   Cond. No.                     4.32e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.32e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [289]:

```
result = sm.ols(formula="price ~ elevation", data=data).fit()
print result.params
print result.summary()
```

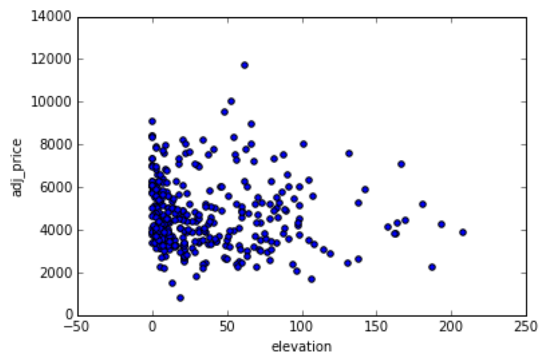
```
Intercept    4322.392306
elevation       3.596286
dtype: float64
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.011
Model:                            OLS   Adj. R-squared:                  0.008
Method:                 Least Squares   F-statistic:                     3.949
Date:                Mon, 07 Sep 2015   Prob (F-statistic):             0.0477
Time:                        13:51:22   Log-Likelihood:                -3164.5
No. Observations:                 366   AIC:                             6333.
Df Residuals:                     364   BIC:                             6341.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept   4322.3923     96.085     44.985      0.000      4133.440  4511.344
elevation      3.5963      1.810      1.987      0.048         0.037     7.155
==============================================================================
Omnibus:                       16.331   Durbin-Watson:                   1.918
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               17.774
Skew:                           0.533   Prob(JB):                     0.000138
Kurtosis:                       2.830   Cond. No.                         70.7
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

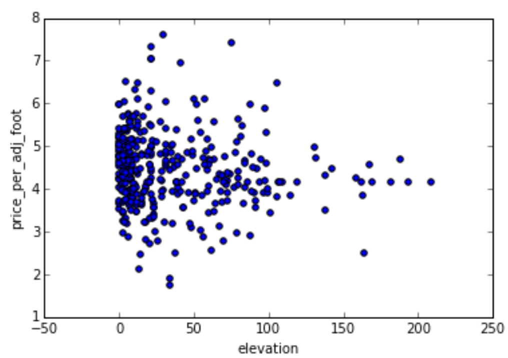
In [ ]:

In [ ]:

In [290]:
```
price_per_adj_foot = data['price'] / data['adj_sqft']
price_per_adj_foot.name = 'price_per_adj_foot'
adj_price = data['price'] * data['area_multiplier']
adj_price.name = 'adj_price'
data = pd.concat([data, price_per_adj_foot, adj_price], axis=1)

# visualize the relationship between the features and the response using scatterplots
data.plot(kind='scatter', x='elevation', y='adj_price')
data.plot(kind='scatter', x='elevation', y='price_per_adj_foot')
```

Out[290]: <matplotlib.axes._subplots.AxesSubplot at 0x11aad03d0>





In [ ]: