In [143]:
```
%load_ext sql
```

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

In [144]:
```
%sql mysql://prod:nerd@52.2.153.189/rental_nerd
```

Out[144]:
```
u'Connected: prod@rental_nerd'
```

In [145]:
```
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()
```

```
560 rows affected.
```

In [146]:
```
result.csv(filename="SQLdump.csv")
```

Out[146]: CSV results (./files/SQLdump.csv)

In [147]:
```
# 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

# read data into a DataFrame
data.head()
```

Out[147]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	id	price	tra
0	1	1	1	567 Vallejo Street #PH500	San Francisco (North Beach)	3	3	2081	climbsf_renting	http://www.climbsf.com/for-rent/567-vallejo-st...	...	1	12000	op
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	2	3950	op
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	3	5400	op
3	4	4	4	333 Fremont Street #705	San Francisco (South Beach)	1	1	0	climbsf_renting	http://www.climbsf.com/for-rent/333-fremont-st...	...	4	3600	op
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	5	3975	op

5 rows × 26 columns

In [148]:
```
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[148]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	price	transa
0	1	1	1	567 Vallejo Street #PH500	San Francisco (North Beach)	3	3	2081	climbsf_renting	http://www.climbsf.com/for-rent/567-vallejo-st...	...	12000	open
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	3950	open
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	5400	open
3	4	4	4	333 Fremont Street #705	San Francisco (South Beach)	1	1	0	climbsf_renting	http://www.climbsf.com/for-rent/333-fremont-st...	...	3600	open
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	3975	open

5 rows × 27 columns

In [149]:
```
# create neighborhoods from lat/long coordinates
import fiona
import shapely as shapely
from shapely.geometry import asShape
```

In [150]:
```
shaped_neighborhood = ['None'] * len(data)

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 = shapely.geometry.Point([data['longitude'][row], data['latitude'][row]]) # longitude, latitude

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

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

checking for listings in: Alamo Square

```
checking for listings in: Alamo Square
105: Found 814 Hayes Street #2 in hood Alamo Square
414: Found Webster St San Francisco, CA 94117 in hood Alamo Square
checking for listings in: Anza Vista
523: Found 965 Baker St, San Francisco, CA 94115 in hood Anza Vista
checking for listings in: Balboa Terrace
checking for listings in: Bayview
14: Found 1766 Oakdale Avenue in hood Bayview
216: Found 5800 Third Street #1109 in hood Bayview
241: Found 5800 Third Street #1204 in hood Bayview
345: Found 5800 Third Street #1411 in hood Bayview
checking for listings in: Bernal Heights
145: Found 723 Ellsworth Street in hood Bernal Heights
176: Found 501 Crescent Way #5107 in hood Bernal Heights
188: Found 301 Crescent Court #3107 in hood Bernal Heights
367: Found 1 College Avenue in hood Bernal Heights
checking for listings in: Buena Vista Park/Ashbury Heights
78: Found 355 Buena Vista Avenue East #305W in hood Buena Vista Park/Ashbury Heights
checking for listings in: Central Richmond
11: Found 655 26th Avenue in hood Central Richmond
131: Found 657 26th Avenue in hood Central Richmond
167: Found 374 27th Avenue in hood Central Richmond
436: Found 735 Market St # 1, San Francisco, CA 94103 in hood Central Richmond
480: Found 3526 Anza St, San Francisco, CA 94121 in hood Central Richmond
checking for listings in: Central Sunset
392: Found 1384 29th Avenue in hood Central Sunset
checking for listings in: Clarendon Heights
351: Found 100 Palo Alto Avenue in hood Clarendon Heights
478: Found 193 Saint Germain Ave, San Francisco, CA 94114 in hood Clarendon Heights
checking for listings in: Corona Heights
477: Found States St San Francisco, CA 94114 in hood Corona Heights
checking for listings in: Cow Hollow
413: Found Greenwich St San Francisco, CA 94123 in hood Cow Hollow
checking for listings in: Crocker Amazon
checking for listings in: Diamond Heights
166: Found 970 Duncan Street #304F in hood Diamond Heights
483: Found Crags Ct San Francisco, CA 94131 in hood Diamond Heights
checking for listings in: Downtown
128: Found 900 Bush Street #410 in hood Downtown
194: Found 181 O'Farrell Street #502 in hood Downtown
240: Found 637 Powell Street #202 in hood Downtown
266: Found 1001 Pine Street #1204 in hood Downtown
269: Found 1299 Bush Street #303 in hood Downtown
275: Found 735 Geary Street #303 in hood Downtown
503: Found 10 Cyril Magnin St UNIT 404, San Francisco, CA 94102 in hood Downtown
checking for listings in: Duboce Triangle
143: Found 997 14th Street in hood Duboce Triangle
181: Found 8 Buchanan Street #600 in hood Duboce Triangle
355: Found 8 Buchanan Street #214 in hood Duboce Triangle
checking for listings in: Eureka Valley / Dolores Heights
82: Found 4517 18th Street in hood Eureka Valley / Dolores Heights
155: Found 225 Grandview Avenue in hood Eureka Valley / Dolores Heights
191: Found 306 Eureka Street in hood Eureka Valley / Dolores Heights
255: Found 55 Hancock Street in hood Eureka Valley / Dolores Heights
263: Found 163 Liberty Street in hood Eureka Valley / Dolores Heights
272: Found 3065 Market Street in hood Eureka Valley / Dolores Heights
283: Found 234 Grand View Avenue in hood Eureka Valley / Dolores Heights
454: Found Cumberland St San Francisco, CA 94114 in hood Eureka Valley / Dolores Heights
458: Found 18th St San Francisco, CA 94114 in hood Eureka Valley / Dolores Heights
460: Found 3745 21st St, San Francisco, CA 94114 in hood Eureka Valley / Dolores Heights
519: Found 135 Yukon St, San Francisco, CA 94114 in hood Eureka Valley / Dolores Heights
525: Found 20th St San Francisco, CA 94114 in hood Eureka Valley / Dolores Heights
checking for listings in: Excelsior
289: Found 271 Paris Street in hood Excelsior
502: Found Athens St San Francisco, CA 94112 in hood Excelsior
checking for listings in: Financial District/Barbary Coast
474: Found Stockton St San Francisco, CA 94108 in hood Financial District/Barbary Coast
checking for listings in: Yerba Buena
6: Found 199 New Montgomery Street #402 in hood Yerba Buena
12: Found 301 Mission Street #22H in hood Yerba Buena
18: Found 301 Mission Street #35A in hood Yerba Buena
19: Found 301 Mission St. #5A in hood Yerba Buena
21: Found 301 Mission St. #53D in hood Yerba Buena
49: Found 74 New Montgomery #412 in hood Yerba Buena
58: Found 16 Jessie St #407 in hood Yerba Buena
68: Found 16 Jessie Street #311 in hood Yerba Buena
83: Found 300 Third Street #1114 in hood Yerba Buena
88: Found 199 New Montgomery Street #208 in hood Yerba Buena
111: Found 821 Folsom Street #310 in hood Yerba Buena
113: Found 301 Mission Street #701 in hood Yerba Buena
```

```
138: Found One Hawthorne Street #2H in hood Yerba Buena
157: Found One Hawthorne Street #2J in hood Yerba Buena
165: Found 74 New Montgomery Street #515 in hood Yerba Buena
173: Found 74 New Montgomery Street #416 in hood Yerba Buena
184: Found 301 Mission Street #14D in hood Yerba Buena
273: Found 301 Mission Street #29F in hood Yerba Buena
298: Found 870 Harrison Street #603 in hood Yerba Buena
324: Found One Hawthorne Street #11F in hood Yerba Buena
339: Found One Hawthorne Street #16B in hood Yerba Buena
342: Found 74 New Montgomery Street #309 in hood Yerba Buena
349: Found 74 New Montgomery Street #709 in hood Yerba Buena
372: Found 77 Dow Place #707 in hood Yerba Buena
378: Found 246 Second Street #1301 in hood Yerba Buena
397: Found 16 Jessie Street #209 in hood Yerba Buena
431: Found Hawthorne St San Francisco, CA 94105 in hood Yerba Buena
551: Found Mission St San Francisco, CA 94105 in hood Yerba Buena
checking for listings in: Forest Hill
checking for listings in: Forest Hills Extension
545: Found 107 Ulloa St, San Francisco, CA 94127 in hood Forest Hills Extension
checking for listings in: Forest Knolls
checking for listings in: Glen Park
215: Found 205 Beacon Street in hood Glen Park
222: Found 115 Farnum Street in hood Glen Park
245: Found 101 Mateo Street in hood Glen Park
292: Found 149 Chenery Street in hood Glen Park
checking for listings in: Golden Gate Heights
277: Found 2035 9th Avenue in hood Golden Gate Heights
checking for listings in: Golden Gate Park
checking for listings in: Haight Ashbury
checking for listings in: Hayes Valley
66: Found 55 Page Street #814 in hood Hayes Valley
136: Found 300 Ivy Street #403 in hood Hayes Valley
168: Found 300 Ivy Street #503 in hood Hayes Valley
228: Found 539 Octavia Street #13 in hood Hayes Valley
261: Found 539 Octavia Street #17 in hood Hayes Valley
301: Found 539 Octavia Street #11 in hood Hayes Valley
302: Found 539 Octavia Street #12 in hood Hayes Valley
304: Found 539 Octavia Street #14 in hood Hayes Valley
325: Found 539 Octavia Street #9 in hood Hayes Valley
328: Found 300 Ivy Street #500 in hood Hayes Valley
331: Found 539 Octavia Street #4 in hood Hayes Valley
388: Found 539 Octavia Street #8 in hood Hayes Valley
389: Found 539 Octavia Street #3 in hood Hayes Valley
394: Found 55 Page Street #518 in hood Hayes Valley
407: Found Buchanan St San Francisco, CA 94102 in hood Hayes Valley
427: Found Scott St San Francisco, CA 94117 in hood Hayes Valley
455: Found 732 Haight St, San Francisco, CA 94117 in hood Hayes Valley
checking for listings in: Hunters Point
checking for listings in: Ingleside
1: Found 252 Granada Avenue in hood Ingleside
checking for listings in: Ingleside Heights
checking for listings in: Ingleside Terrace
checking for listings in: Inner Mission
8: Found 555 Bartlett Street #409 in hood Inner Mission
20: Found 2735 Folsom Street in hood Inner Mission
97: Found 2125 Bryant Street #102 in hood Inner Mission
126: Found 555 Bartlett Street #413 in hood Inner Mission
217: Found 2125 Bryant Street #303 in hood Inner Mission
239: Found 720 York Street #214 in hood Inner Mission
247: Found 1515 15th Street in hood Inner Mission
251: Found 2633 Harrison St. in hood Inner Mission
311: Found 1875 Mission Street #407 in hood Inner Mission
312: Found 45 Bartlett Street #705 in hood Inner Mission
327: Found 45 Bartlett Street #315 in hood Inner Mission
330: Found 45 Bartlett Street #302 in hood Inner Mission
335: Found 45 Bartlett Street #706 in hood Inner Mission
360: Found 891 York Street in hood Inner Mission
370: Found 2101 Bryant Street #205 in hood Inner Mission
386: Found 163 Capp Street in hood Inner Mission
409: Found 20th St San Francisco, CA 94110 in hood Inner Mission
410: Found 20th St San Francisco, CA 94110 in hood Inner Mission
497: Found Hoff St San Francisco, CA 94110 in hood Inner Mission
517: Found 20th St San Francisco, CA 94110 in hood Inner Mission
checking for listings in: Inner Parkside
checking for listings in: Inner Richmond
412: Found 11th Ave San Francisco, CA 94118 in hood Inner Richmond
432: Found 11th Ave San Francisco, CA 94118 in hood Inner Richmond
checking for listings in: Inner Sunset
521: Found (Undisclosed Address) San Francisco, CA 94122 in hood Inner Sunset
550: Found 168 18th Ave, San Francisco, CA 94121 in hood Inner Sunset
```

```
checking for listings in: Jordan Park / Laurel Heights
checking for listings in: Lake Street
checking for listings in: Lake Shore
552: Found 136 Meadowbrook Dr, San Francisco, CA 94132 in hood Lake Shore
checking for listings in: Lakeside
checking for listings in: Lone Mountain
108: Found 3156 Turk Street #A in hood Lone Mountain
180: Found 155 Wood Street #155 in hood Lone Mountain
219: Found 184 Cook Street in hood Lone Mountain
481: Found 343 Parker Ave, San Francisco, CA 94118 in hood Lone Mountain
498: Found Mcallister St San Francisco, CA 94118 in hood Lone Mountain
checking for listings in: Lower Pacific Heights
64: Found 2712A Pine Street in hood Lower Pacific Heights
101: Found 1600 Webster Street #308 in hood Lower Pacific Heights
140: Found 2075 Sutter Street #521 in hood Lower Pacific Heights
262: Found 1521 Sutter Street #206 in hood Lower Pacific Heights
checking for listings in: Marina
119: Found 1837 Jefferson Street in hood Marina
162: Found 1921 Jefferson Street #302 in hood Marina
192: Found 101 Marina Boulevard in hood Marina
204: Found 1839 Jefferson Street in hood Marina
212: Found 1735 Beach Street in hood Marina
264: Found 1737 Beach Street in hood Marina
482: Found 121 Avila St, San Francisco, CA 94123 in hood Marina
516: Found 1790 Beach St, San Francisco, CA 94123 in hood 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
475: Found 566 Teresita Blvd, San Francisco, CA 94127 in hood Miraloma Park
checking for listings in: Mission Bay
4: Found 420 Mission Bay Boulevard North #121 in hood Mission Bay
118: Found 435 China Basin Street #634 in hood Mission Bay
123: Found 480 Mission Bay Boulevard North #1608 in hood Mission Bay
169: Found 435 China Basin Street #441 in hood Mission Bay
211: Found 420 Mission Bay Boulevard North #308 in hood Mission Bay
234: Found 325 China Basin Street #607 in hood Mission Bay
237: Found 435 China Basin Street #118 in hood Mission Bay
274: Found 435 China Basin Street #639 in hood Mission Bay
382: Found 480 Mission Bay Boulevard North #PH1606 in hood Mission Bay
384: Found 480 Mission Bay Blvd. North #1007 in hood Mission Bay
checking for listings in: Mission Dolores
248: Found 3664 19th Street in hood Mission Dolores
271: Found 198 Dolores Street #6 in hood Mission Dolores
316: Found 35 Dolores Street #410 in hood Mission Dolores
377: Found 49A Elgin Park in hood Mission Dolores
checking for listings in: Mission Terrace
207: Found 217 Theresa Street in hood Mission Terrace
258: Found 55 Junior Terrace in hood Mission Terrace
checking for listings in: Monterey Heights
checking for listings in: Mount Davidson Manor
318: Found 810 Faxon Avenue in hood Mount Davidson Manor
checking for listings in: Noe Valley
2: Found 460 Valley Street in hood Noe Valley
60: Found 682 28th Street in hood Noe Valley
364: Found 1414 Douglass Street in hood Noe Valley
428: Found Dolores St San Francisco, CA 94110 in hood Noe Valley
429: Found 23rd St San Francisco, CA 94114 in hood Noe Valley
checking for listings in: North Beach
67: Found 1703 Powell Street in hood North Beach
115: Found 1940 Stockton Street #104 in hood North Beach
294: Found 727 Green Street in hood North Beach
411: Found Vallejo St San Francisco, CA 94133 in hood North Beach
435: Found Vallejo St San Francisco, CA 94133 in hood North Beach
476: Found 445 Francisco Street # FL 6TH, San Francisco, CA 94133 in hood North Beach
checking for listings in: North Panhandle
246: Found 136 Cole Street in hood North Panhandle
checking for listings in: North Waterfront
0: Found 567 Vallejo Street #PH500 in hood North Waterfront
186: Found 240 Lombard Street #435 in hood North Waterfront
checking for listings in: Oceanview
549: Found Broad St San Francisco, CA 94112 in hood Oceanview
checking for listings in: Outer Mission
checking for listings in: Outer Parkside
522: Found 2042 46th Ave, San Francisco, CA 94116 in hood Outer Parkside
524: Found 2759 38th Ave, San Francisco, CA 94116 in hood Outer Parkside
checking for listings in: Outer Richmond
340: Found 4442 Cabrillo Street in hood Outer Richmond
400: Found 459 33rd Avenue in hood Outer Richmond
520: Found 402 43rd Ave # 2, San Francisco, CA 94121 in hood Outer Richmond
```

```
checking for listings in: Outer Sunset
checking for listings in: Pacific Heights
59: Found 1979 Clay Street in hood Pacific Heights
106: Found 3009 Sacramento Street #2 in hood Pacific Heights
199: Found 1945 Jackson Street in hood Pacific Heights
232: Found 2560 Vallejo Street in hood Pacific Heights
235: Found 1880 Jackson Street in hood Pacific Heights
243: Found 2072 Vallejo Street in hood Pacific Heights
405: Found Vallejo St San Francisco, CA 94123 in hood Pacific Heights
500: Found 1966 Pacific Ave 201 Short Term Furnished Rental, San Francisco, CA 94109 in hood Pacific Heights
547: Found Clay St San Francisco, CA 94115 in hood Pacific Heights
checking for listings in: Parkside
451: Found Ulloa St San Francisco, CA 94116 in hood Parkside
checking for listings in: Cole Valley/Parnassus Heights
314: Found 656 Clayton Street #B in hood Cole Valley/Parnassus Heights
433: Found Beulah St San Francisco, CA 94117 in hood Cole Valley/Parnassus Heights
548: Found 487 Belvedere St, San Francisco, CA 94117 in hood Cole Valley/Parnassus Heights
checking for listings in: Pine Lake Park
checking for listings in: Portola
315: Found 7 Colby Street in hood Portola
checking for listings in: Potrero Hill
10: Found 451 Kansas Street #616 in hood Potrero Hill
73: Found 999 16th Street #15 in hood Potrero Hill
125: Found 451 Kansas Street #489 in hood Potrero Hill
183: Found 1661 18th Street in hood Potrero Hill
187: Found 330 Mississippi Street in hood Potrero Hill
200: Found 1047 Mississippi Street #7 in hood Potrero Hill
208: Found 717 San Bruno Avenue in hood Potrero Hill
209: Found 451 Kansas Street #362 in hood Potrero Hill
223: Found 451 Kansas Street #571 in hood Potrero Hill
236: Found 223 Mississippi Street #2 in hood Potrero Hill
256: Found 743 Rhode Island Street in hood Potrero Hill
280: Found 451 Kansas Street #436 in hood Potrero Hill
290: Found 634 Missouri Street in hood Potrero Hill
322: Found 451 Kansas Street #520 in hood Potrero Hill
323: Found 451 Kansas Street #480 in hood Potrero Hill
338: Found 451 Kansas Street #356 in hood Potrero Hill
348: Found 451 Kansas Street #488 in hood Potrero Hill
354: Found 401 Arkansas Street in hood Potrero Hill
406: Found San Bruno Ave San Francisco, CA 94107 in hood Potrero Hill
453: Found Mississippi St San Francisco, CA 94107 in hood Potrero Hill
518: Found 1143 Rhode Island St, San Francisco, CA 94107 in hood Potrero Hill
checking for listings in: Presidio
checking for listings in: Presidio Heights
checking for listings in: Russian Hill
112: Found 26 Waldo Alley in hood Russian Hill
202: Found 935 Union Street in hood Russian Hill
252: Found 2539 Larkin Street in hood Russian Hill
361: Found 24 Waldo Alley in hood Russian Hill
434: Found 748 Bay St, San Francisco, CA 94109 in hood Russian Hill
456: Found 35 Bret Harte Ter, San Francisco, CA 94133 in hood Russian Hill
457: Found Lombard St San Francisco, CA 94133 in hood Russian Hill
checking for listings in: Saint Francis Wood
checking for listings in: Sea Cliff
checking for listings in: Silver Terrace
201: Found 1782 Quint Street in hood Silver Terrace
checking for listings in: South Beach
3: Found 333 Fremont Street #705 in hood South Beach
5: Found 325 Berry Street #520 in hood South Beach
15: Found 333 Fremont Street #810 in hood South Beach
16: Found 338 Spear Street #24H in hood South Beach
17: Found 425 First Street #1806 in hood South Beach
28: Found 555 4th St # 803 in hood South Beach
56: Found 260 King St in hood South Beach
57: Found 235 Berry Street #102 in hood South Beach
61: Found 425 1st Street #907 in hood South Beach
62: Found 555 4th St # 734 in hood South Beach
63: Found 201 Harrison Street #711 in hood South Beach
70: Found 338 Spear Street #18B in hood South Beach
71: Found 400 Beale Street #406 in hood South Beach
72: Found 235 Berry Street #107 in hood South Beach
74: Found 403 Main Street #410 in hood South Beach
75: Found 333 Main Street #4D in hood South Beach
76: Found 50 Lansing Street #209 in hood South Beach
77: Found 425 1st Street #2802 in hood South Beach
79: Found 338 Spear Street #3C in hood South Beach
80: Found 333 Main Street #7E in hood South Beach
84: Found 400 Beale Street #506 in hood South Beach
85: Found 333 Main Street #4L in hood South Beach
89: Found 88 King Street #904 in hood South Beach
```

```
90: Found 325 Berry Street #317 in hood South Beach
91: Found 333 Main Street #9D in hood South Beach
92: Found 333 Main Street #7B in hood South Beach
94: Found 338 Spear Street #18C in hood South Beach
95: Found 338 Spear Street #21G in hood South Beach
96: Found 501 Beale Street #8D in hood South Beach
100: Found 403 Main Street #515 in hood South Beach
103: Found 400 Beale Street #911 in hood South Beach
104: Found 50 Lansing St. #406 in hood South Beach
109: Found 229 Brannan Street #12J in hood South Beach
110: Found 403 Main Street #416 in hood South Beach
114: Found 425 1st Street #4501 in hood South Beach
116: Found 333 Main Street #4C in hood South Beach
117: Found 200 Brannan Street #314 in hood South Beach
120: Found 200 Brannan Street #323 in hood South Beach
121: Found 260 King Street #479 in hood South Beach
122: Found 301 Main Street #4G in hood South Beach
127: Found 325 Berry Street #614 in hood South Beach
129: Found 555 4th Street #417 in hood South Beach
130: Found 425 1st Street #3503 in hood South Beach
132: Found 250 King Street #536 in hood South Beach
133: Found 501 Beale Street #4D in hood South Beach
134: Found 301 Main Street #25E in hood South Beach
137: Found 88 King Street #106 in hood South Beach
144: Found 301 Main Street #6B in hood South Beach
146: Found 18 Lansing Street #201 in hood South Beach
147: Found 325 Berry Street #506 in hood South Beach
148: Found 49 Zoe Street #4 in hood South Beach
151: Found 425 1st Street #1608 in hood South Beach
152: Found 301 Main Street #10C in hood South Beach
153: Found 301 Main Street #23F in hood South Beach
158: Found 338 Spear Street #39A in hood South Beach
160: Found 325 Berry Street #701 in hood South Beach
161: Found 219 Brannan Street #7E in hood South Beach
163: Found 318 Spear Street #5C in hood South Beach
164: Found 425 1st Street #3704 in hood South Beach
171: Found 325 Berry Street #502 in hood South Beach
172: Found 301 Main Street #3B in hood South Beach
174: Found 301 Main Street #14F in hood South Beach
175: Found 250 King Street #553 in hood South Beach
177: Found 338 Spear Street #35C in hood South Beach
178: Found 425 First Street #3807 in hood South Beach
179: Found 325 Berry Street #710 in hood South Beach
185: Found 260 King Street #907 in hood South Beach
189: Found 555 Fourth Street #928 in hood South Beach
193: Found 425 1st Street #4404 in hood South Beach
195: Found 400 Beale Street #1312 in hood South Beach
196: Found 400 Beale Street #706 in hood South Beach
197: Found 229 Brannan Street #17C in hood South Beach
203: Found 461 2nd St. #557T in hood South Beach
206: Found 338 Spear Street #17G in hood South Beach
210: Found 425 1st Street #3205 in hood South Beach
213: Found 250 King Street #604 in hood South Beach
220: Found 325 Berry Street #416 in hood South Beach
221: Found 50 Lansing Street #108 in hood South Beach
224: Found 318 Spear Street #5A in hood South Beach
225: Found 301 Main Street #18A in hood South Beach
226: Found 400 Beale Street #1003 in hood South Beach
227: Found 18 Lansing Street #308 in hood South Beach
229: Found 555 4th Street #930 in hood South Beach
233: Found 338 Spear Street #39E in hood South Beach
238: Found 325 Berry Street #422 in hood South Beach
242: Found 333 Fremont Street #102 in hood South Beach
244: Found 301 Main Street #22A in hood South Beach
250: Found 425 First Street #3307 in hood South Beach
254: Found 88 Townsend Street #426 in hood South Beach
257: Found 219 Brannan Street #16K in hood South Beach
259: Found 333 Fremont Street #504 in hood South Beach
260: Found 301 Main Street #16E in hood South Beach
265: Found 301 Main Street #4F in hood South Beach
267: Found 200 Townsend Street #47 in hood South Beach
268: Found 338 Spear Street #28B in hood South Beach
270: Found 333 1st Street #N305 in hood South Beach
276: Found 219 Brannan St. #14K in hood South Beach
278: Found 325 Berry Street #314 in hood South Beach
279: Found 250 King Street #776 in hood South Beach
281: Found 50 Lansing Street #305 in hood South Beach
282: Found 301 Main Street #5C in hood South Beach
284: Found 260 King Street #605 in hood South Beach
286: Found 425 1st Street #2005 in hood South Beach
```

```
287: Found 555 Fourth Street #912 in hood South Beach
293: Found 425 1st Street #3402 in hood South Beach
296: Found 355 Bryant Street #401 in hood South Beach
297: Found 81 Lansing Street #408 in hood South Beach
299: Found 401 Harrison Street #3803 in hood South Beach
300: Found 260 King Street #1413 in hood South Beach
303: Found 333 Fremont Street #405 in hood South Beach
306: Found 425 First Street #3308 in hood South Beach
307: Found 333 Fremont Street #808 in hood South Beach
308: Found 461 Second Street #102 in hood South Beach
309: Found 318 Spear Street #6K in hood South Beach
310: Found 601 Fourth Street #324 in hood South Beach
317: Found 300 Berry Street #510 in hood South Beach
319: Found 333 Fremont Street #203 in hood South Beach
321: Found 333 Fremont Street #809 in hood South Beach
326: Found 601 Fourth Street #226 in hood South Beach
333: Found 333 Fremont Street #709 in hood South Beach
334: Found 338 Spear Street #38F in hood South Beach
336: Found 425 First Street #1002 in hood South Beach
337: Found 318 Spear Street #8E in hood South Beach
341: Found 300 Berry Street #1313 in hood South Beach
343: Found 300 Berry Street #523 in hood South Beach
344: Found 555 Fourth Street #312 in hood South Beach
346: Found 18 Lansing Street #102 in hood South Beach
347: Found 333 Fremont Street #603 in hood South Beach
350: Found 501 Beale St. #9E in hood South Beach
352: Found 239 Brannan Street #14A in hood South Beach
353: Found 425 First Street #1308 in hood South Beach
359: Found 333 Fremont Street #412 in hood South Beach
362: Found 260 King Street #1013 in hood South Beach
365: Found 88 Townsend Street #402 in hood South Beach
366: Found 333 Fremont Street #708 in hood South Beach
368: Found 338 Spear Street #6A in hood South Beach
369: Found 333 Fremont Street #802 in hood South Beach
371: Found 18 Lansing Street #306 in hood South Beach
379: Found 301 Bryant Street #102 in hood South Beach
380: Found 325 Berry Street #316 in hood South Beach
381: Found 301 Main Street #35F in hood South Beach
385: Found 425 1st Street #3305 in hood South Beach
390: Found 333 Fremont Street #205 in hood South Beach
391: Found 333 Fremont Street #110 in hood South Beach
396: Found 333 Fremont Street #806 in hood South Beach
398: Found 333 1st Street #N102 in hood South Beach
399: Found 333 Fremont Street #701 in hood South Beach
430: Found 501 Beale St, San Francisco, CA 94105 in hood South Beach
452: Found 318 Main St # 5K, San Francisco, CA 94105 in hood South Beach
479: Found 177 Townsend St UNIT 1131, San Francisco, CA 94107 in hood South Beach
495: Found 260 King St # 1000, San Francisco, CA 94107 in hood South Beach
546: Found 301 Main St UNIT 35A, San Francisco, CA 94105 in hood South Beach
checking for listings in: South of Market
7: Found 1160 Mission Street #1112 in hood South of Market
9: Found 655 Fifth Street #12 in hood South of Market
81: Found 1097 Howard Street #304 in hood South of Market
86: Found 1160 Mission Street #1905 in hood South of Market
87: Found 140 South Van Ness Avenue #922 in hood South of Market
93: Found 1160 Mission Street #1611 in hood South of Market
150: Found 1160 Mission Street #901 in hood South of Market
154: Found 388 Townsend Street #4 in hood South of Market
156: Found 1160 Mission Street #2303 in hood South of Market
170: Found 655 Fifth Street #1 in hood South of Market
182: Found 767 Bryant Street #403 in hood South of Market
190: Found 1328 Mission Street #5 in hood South of Market
198: Found 2 Mint Plaza #203 in hood South of Market
285: Found 1160 Mission Street #1212 in hood South of Market
291: Found 10 Mint Plaza #1 in hood South of Market
295: Found 520 Sixth Street #11 in hood South of Market
305: Found 48 Rausch Street in hood South of Market
320: Found 520 6th Street #16 in hood South of Market
356: Found 2 Mint Plaza #805 in hood South of Market
383: Found 631 Folsom Street #PHC in hood South of Market
459: Found Tehama St San Francisco, CA 94103 in hood South of Market
499: Found (Undisclosed Address) San Francisco, CA 94103 in hood South of Market
501: Found Market St San Francisco, CA 94103 in hood South of Market
544: Found Gilbert St San Francisco, CA 94103 in hood South of Market
checking for listings in: Stonestown
checking for listings in: Sunnyside
checking for listings in: Telegraph Hill
149: Found 373 Green Street in hood Telegraph Hill
253: Found 567 Vallejo Street #301 in hood Telegraph Hill
357: Found 296 Francisco Street in hood Telegraph Hill
```

```
checking for listings in: Twin Peaks
checking for listings in: Van Ness/Civic Center
99: Found 650 Turk Street #501 in hood Van Ness/Civic Center
231: Found 750 Van Ness Avenue #203 in hood Van Ness/Civic Center
288: Found One Daniel Burnham Court #314 in hood Van Ness/Civic Center
313: Found 720 Gough Street #45 in hood Van Ness/Civic Center
408: Found Van Ness Ave San Francisco, CA 94102 in hood Van Ness/Civic Center
checking for listings in: Visitacion Valley
65: Found 5 Sparta Street in hood Visitacion Valley
504: Found Ervine St San Francisco, CA 94134 in hood Visitacion Valley
543: Found 126 Ward St, San Francisco, CA 94134 in hood Visitacion Valley
checking for listings in: West Portal
checking for listings in: Western Addition
98: Found 66 Cleary Court #306 in hood Western Addition
107: Found 38 Beideman Street in hood Western Addition
124: Found 1288 McAllister Street #201 in hood Western Addition
363: Found 1168 Eddy Street #D4 in hood 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
135: Found 701 Minnesota Street #119 in hood Central Waterfront/Dogpatch
139: Found 1325 Indiana Street #308 in hood Central Waterfront/Dogpatch
141: Found 755 Tennessee Street #2 in hood Central Waterfront/Dogpatch
142: Found 868 Minnesota Street #313 in hood Central Waterfront/Dogpatch
159: Found 2030 Third Street #6 in hood Central Waterfront/Dogpatch
214: Found 863 Indiana Street in hood Central Waterfront/Dogpatch
230: Found 1310 Minnesota Street #203 in hood Central Waterfront/Dogpatch
332: Found 1310 Minnesota Street #101 in hood Central Waterfront/Dogpatch
358: Found 1011 23rd Street #18 in hood Central Waterfront/Dogpatch
395: Found 701 Minnesota Street #202 in hood Central Waterfront/Dogpatch
checking for listings in: Candlestick Point
496: Found Otter Cove Ter San Francisco, CA 94134 in hood Candlestick Point
checking for listings in: Bayview Heights
checking for listings in: Little Hollywood
checking for listings in: Nob Hill
69: Found 1340 Clay Street #402 in hood Nob Hill
102: Found 1536 Pacific Avenue #5 in hood Nob Hill
205: Found 1788 Clay Street #701 in hood Nob Hill
218: Found 1625 Hyde Street in hood Nob Hill
249: Found 1260 Clay Street #208 in hood Nob Hill
329: Found 1800 Van Ness Avenue #505 in hood Nob Hill
```

Out[150]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	transaction_s
0	1	1	1	567 Vallejo Street #PH500	San Francisco (North Beach)	3	3	2081	climbsf_renting	http://www.climbsf.com/for-rent/567-vallejo-st...	...	open
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	open
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	open
3	4	4	4	333 Fremont Street #705	San Francisco (South Beach)	1	1	0	climbsf_renting	http://www.climbsf.com/for-rent/333-fremont-st...	...	open
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	open

5 rows × 28 columns

In [151]:
```
# 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.shaped_neighborhood != 'None') & (data.sqft <= 2500) & (data.price <= 8000) & (data.price != 0) & (data.
bedrooms <= 4) & (data.bathrooms <= 3) & (data.sqft != 0)]

# filter out listings over one month old


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

```
Entries before filter: 560
Entries after filter: 304
```

In [152]:
```
# 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[152]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	neighborho Hill
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	0
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	0
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	0
7	8	8	8	1160 Mission Street #1112	San Francisco (SOMA)	1	1	664	climbsf_renting	http://www.climbsf.com/for-rent/1160-mission-s...	...	0
11	12	12	12	655 26th Avenue	San Francisco (Central Richmond)	2	1	1300	climbsf_renting	http://www.climbsf.com/for-rent/655-26th-ave/	...	0

5 rows × 83 columns

In [153]:
```
# 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()
```

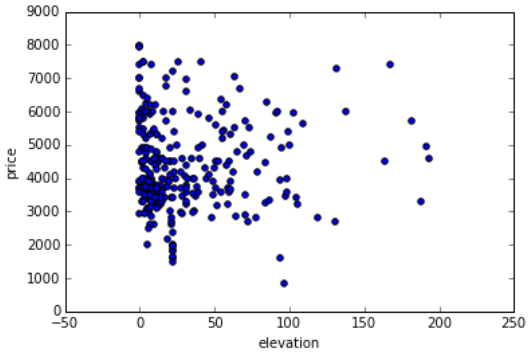
Out[153]:

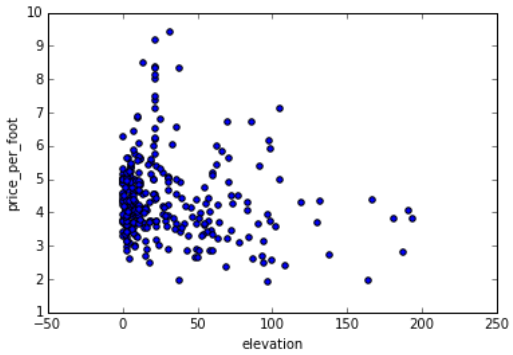
	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	neighborho Ness/Civic
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	0
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	0
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	0
7	8	8	8	1160 Mission Street #1112	San Francisco (SOMA)	1	1	664	climbsf_renting	http://www.climbsf.com/for-rent/1160-mission-s...	...	0
11	12	12	12	655 26th Avenue	San Francisco (Central Richmond)	2	1	1300	climbsf_renting	http://www.climbsf.com/for-rent/655-26th-ave/	...	0

5 rows × 84 columns

In [154]:
```
# 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')
```

Out[154]: <matplotlib.axes._subplots.AxesSubplot at 0x10dc8a750>





In [155]:
```
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 [156]:
```
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: 0.82
Variance score: 0.44
Base area is neighborhood_Alamo Square: $4.17
```

Out[156]:

Neighborhood	$ per square (+/-)
neighborhood_Mount Davidson Manor	-1.75
neighborhood_Ingleside	-1.7
neighborhood_Visitacion Valley	-1.6
neighborhood_Portola	-1.52
neighborhood_Glen Park	-1.44
neighborhood_Bernal Heights	-1.38
neighborhood_Diamond Heights	-1.36
neighborhood_Silver Terrace	-1.33
neighborhood_Lake Shore	-1.3
neighborhood_Central Richmond	-1.05
neighborhood_Anza Vista	-1.03
neighborhood_Bayview	-0.94
neighborhood_Excelsior	-0.94
neighborhood_Cole Valley/Parnassus Heights	-0.79
neighborhood_Downtown	-0.78
neighborhood_Outer Parkside	-0.64
neighborhood_Outer Richmond	-0.57
neighborhood_Buena Vista Park/Ashbury Heights	-0.47
neighborhood_Western Addition	-0.4
neighborhood_Forest Hills Extension	-0.37
neighborhood_Golden Gate Heights	-0.33
neighborhood_Mission Bay	-0.32

neighborhood_Oceanview	-0.21
neighborhood_Central Waterfront/Dogpatch	-0.13
neighborhood_Van Ness/Civic Center	-0.13
neighborhood_North Panhandle	-0.12
neighborhood_Miraloma Park	-0.1
neighborhood_South of Market	-0.08
neighborhood_Telegraph Hill	-0.02
neighborhood_Inner Richmond	-0.0
neighborhood_Potrero Hill	0.07
neighborhood_Marina	0.1
neighborhood_Noe Valley	0.18
neighborhood_South Beach	0.19
neighborhood_Lower Pacific Heights	0.2
neighborhood_Lone Mountain	0.34
neighborhood_Yerba Buena	0.48
neighborhood_Eureka Valley / Dolores Heights	0.5
neighborhood_Pacific Heights	0.5
neighborhood_Inner Mission	0.6
neighborhood_Nob Hill	0.64
neighborhood_Duboce Triangle	1.04
neighborhood_Russian Hill	1.05
neighborhood_North Waterfront	1.07
neighborhood_Mission Dolores	1.22
neighborhood_Inner Sunset	1.69
neighborhood_North Beach	2.46
neighborhood_Hayes Valley	2.64
neighborhood_Financial District/Barbary Coast	4.17

In [157]:
```
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[157]:

Neighborhood	$ per sqft
neighborhood_Mount Davidson Manor	2.41970021413
neighborhood_Ingleside	2.46875
neighborhood_Visitacion Valley	2.56314257913
neighborhood_Portola	2.64285714286
neighborhood_Glen Park	2.72727272727

neighborhood_Bernal Heights	2.78200061463
neighborhood_Diamond Heights	2.8085106383
neighborhood_Silver Terrace	2.83464566929
neighborhood_Lake Shore	2.86666666667
neighborhood_Central Richmond	3.11732711733
neighborhood_Anza Vista	3.13581037796
neighborhood_Excelsior	3.22222222222
neighborhood_Bayview	3.22391991699
neighborhood_Cole Valley/Parnassus Heights	3.3732856291
neighborhood_Downtown	3.38847472785
neighborhood_Outer Parkside	3.52669238052
neighborhood_Outer Richmond	3.59375
neighborhood_Buena Vista Park/Ashbury Heights	3.69863013699
neighborhood_Western Addition	3.76897132069
neighborhood_Forest Hills Extension	3.8
neighborhood_Golden Gate Heights	3.83333333333
neighborhood_Mission Bay	3.85094171788
neighborhood_Oceanview	3.95238095238
neighborhood_Van Ness/Civic Center	4.03577014404
neighborhood_Central Waterfront/Dogpatch	4.04075057006
neighborhood_North Panhandle	4.04545454545
neighborhood_Miraloma Park	4.07072368421
neighborhood_South of Market	4.08636554466
neighborhood_Telegraph Hill	4.14764859845
neighborhood_Inner Richmond	4.16363636364
neighborhood_Potrero Hill	4.2321829593
neighborhood_Marina	4.27103404056
neighborhood_Noe Valley	4.35031959807
neighborhood_South Beach	4.35394300952
neighborhood_Lower Pacific Heights	4.36170254619
neighborhood_Lone Mountain	4.50186409772
neighborhood_Yerba Buena	4.64558173282
neighborhood_Pacific Heights	4.66270219935
neighborhood_Eureka Valley / Dolores Heights	4.66682988664
neighborhood_Inner Mission	4.76632367548
neighborhood_Nob Hill	4.80566691815
neighborhood_Duboce Triangle	5.20254134584
neighborhood_Russian Hill	5.22053967721
neighborhood_North Waterfront	5.24109014675
neighborhood_Mission Dolores	5.38924963925
neighborhood_Inner Sunset	5.85714285714
neighborhood_North Beach	6.62254901961
neighborhood_Hayes Valley	6.80851715243
neighborhood_Financial District/Barbary Coast	8.33333333333

In [158]:
```
# 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[158]:

Neighborhood	Multiplier
neighborhood_Alamo Square	0
neighborhood_Mount Davidson Manor	-0.419271948608
neighborhood_Ingleside	-0.4075
neighborhood_Visitacion Valley	-0.384845781009
neighborhood_Portola	-0.365714285714
neighborhood_Glen Park	-0.345454545455
neighborhood_Bernal Heights	-0.332319852489
neighborhood_Diamond Heights	-0.325957446809
neighborhood_Silver Terrace	-0.31968503937
neighborhood_Lake Shore	-0.312
neighborhood_Central Richmond	-0.251841491841
neighborhood_Anza Vista	-0.247405509289
neighborhood_Excelsior	-0.226666666667
neighborhood_Bayview	-0.226259219923
neighborhood_Cole Valley/Parnassus Heights	-0.190411449016
neighborhood_Downtown	-0.186766065317
neighborhood_Outer Parkside	-0.153593828675
neighborhood_Outer Richmond	-0.1375
neighborhood_Buena Vista Park/Ashbury Heights	-0.112328767123
neighborhood_Western Addition	-0.0954468830351
neighborhood_Forest Hills Extension	-0.088
neighborhood_Golden Gate Heights	-0.08
neighborhood_Mission Bay	-0.0757739877095
neighborhood_Oceanview	-0.0514285714286
neighborhood_Van Ness/Civic Center	-0.0314151654308
neighborhood_Central Waterfront/Dogpatch	-0.0302198631857
neighborhood_North Panhandle	-0.0290909090909
neighborhood_Miraloma Park	-0.0230263157895
neighborhood_South of Market	-0.0192722692805
neighborhood_Telegraph Hill	-0.00456433637285
neighborhood_Inner Richmond	-0.000727272727277

neighborhood_Potrero Hill	0.0157239102317
neighborhood_Marina	0.0250481697352
neighborhood_Noe Valley	0.0440767035374
neighborhood_South Beach	0.0449463222858
neighborhood_Lower Pacific Heights	0.0468086110849
neighborhood_Lone Mountain	0.0804473834535
neighborhood_Yerba Buena	0.114939615876
neighborhood_Pacific Heights	0.119048527843
neighborhood_Eureka Valley / Dolores Heights	0.120039172793
neighborhood_Inner Mission	0.143917682115
neighborhood_Nob Hill	0.153360060357
neighborhood_Duboce Triangle	0.248609923
neighborhood_Russian Hill	0.252929522531
neighborhood_North Waterfront	0.25786163522
neighborhood_Mission Dolores	0.29341991342
neighborhood_Inner Sunset	0.405714285714
neighborhood_North Beach	0.589411764706
neighborhood_Hayes Valley	0.634044116583
neighborhood_Financial District/Barbary Coast	1.0

In [159]:
```
# 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[159]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	neighborho Addition
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	0
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	0
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	0
7	8	8	8	1160 Mission Street #1112	San Francisco (SOMA)	1	1	664	climbsf_renting	http://www.climbsf.com/for-rent/1160-mission-s...	...	0
11	12	12	12	655 26th Avenue	San Francisco (Central Richmond)	2	1	1300	climbsf_renting	http://www.climbsf.com/for-rent/655-26th-ave/	...	0

5 rows × 86 columns

In [160]:
```
# 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: 1581.13
Residual sum of squares: 587671.58
Variance score: 0.67
```

Out[160]:
```
[]
```

In [161]:
```
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: 1.32
Variance score: 0.11
4.79633313103
```

Out[161]:
```
[(u'Year_1969', -1.9410031817959141),
 (u'Year_2011', -0.62577229712902971),
 (u'Year_2012', -1.3700383103173595),
 (u'Year_2013', -0.98855679368271021),
 (u'Year_2014', -0.60179239513716498)]
```

In [162]:
```
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': 2.8553299492385782, u'neighborhood_Alamo Square': 4.7963331310344923, u'Year_2012': 3.4262948207171329, u'Y
ear_2013': 3.8077763373517821, u'Year_2011': 4.1705608339054629, u'Year_2014': 4.194540735897327}
```

In [163]:
```
# 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[163]:
```
[(u'Year_1969', -0.40468481416287083),
 (u'Year_2011', -0.13046889780861826),
 (u'Year_2012', -0.28564285942787804),
 (u'Year_2013', -0.2061067833854765),
 (u'Year_2014', -0.12546926551937987)]
```

In [164]:
```
# 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[164]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	Year_1969
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	0
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	0
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	0
7	8	8	8	1160 Mission Street #1112	San Francisco (SOMA)	1	1	664	climbsf_renting	http://www.climbsf.com/for-rent/1160-mission-s...	...	0
11	12	12	12	655 26th Avenue	San Francisco (Central Richmond)	2	1	1300	climbsf_renting	http://www.climbsf.com/for-rent/655-26th-ave/	...	0

5 rows × 88 columns

In [165]:
```
# 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)
```

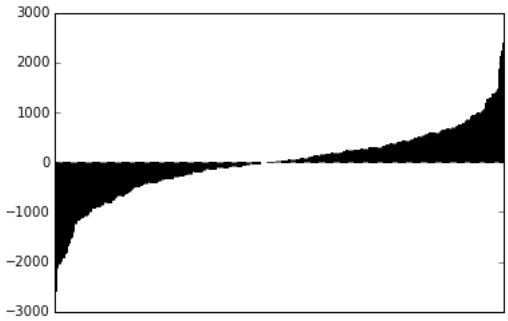
```
1510.45714164
Residual sum of squares: 547156.69
Variance score: 0.69
[('adj_sqft', 2.8457143729597223)]
```

Out[165]:
```
326    1351.485576
243    1371.524527
328    1609.231682
236    1624.388449
108    1636.430154
66     1807.945788
427    1820.489102
294    2338.071426
60     2435.630745
455    3240.446932
Name: Error, dtype: float64
```

In [166]:
```
# 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([])
```

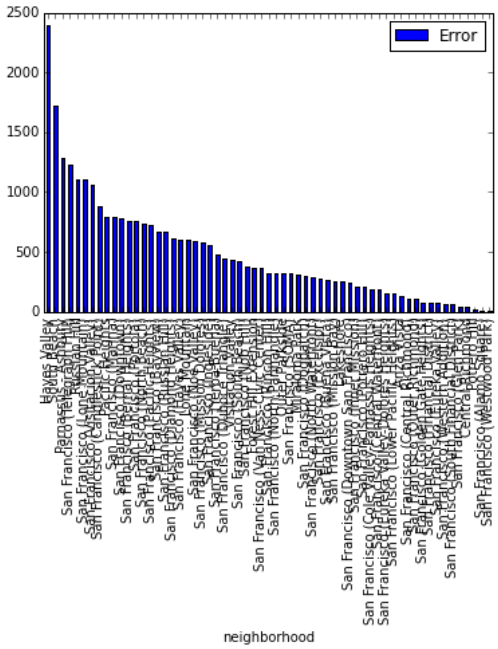
```
Intercept: 1229.39
Residual sum of squares: 489125.54
Variance score: 0.73
[('adj_sqft', 2.323433447134625), ('bedrooms', 194.36513659462096), ('bathrooms', 341.12870661669922)]
```

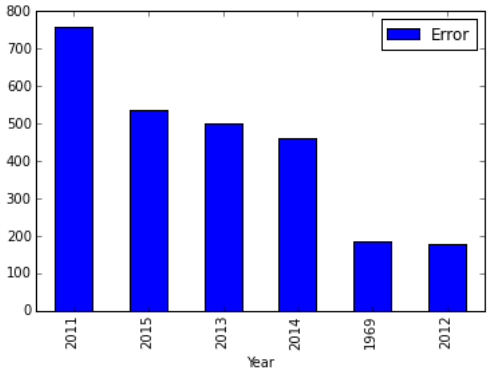
Out[166]:
```
[]
```

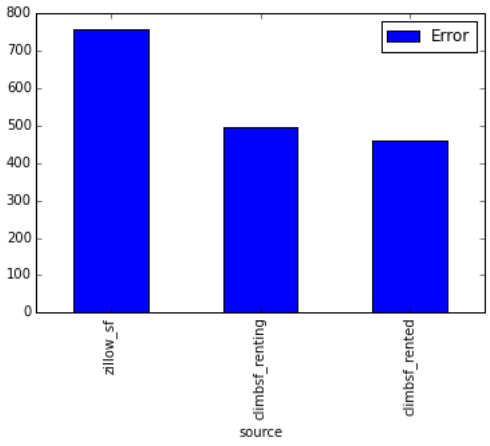


In [167]:
```
# show errors by neighborhood to see if there are any neighborhoods with funky differences

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[167]: <matplotlib.axes._subplots.AxesSubplot at 0x1122bc650>







In [168]:
```
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', 2.323433447134625), ('bedrooms', 194.36513659462096), ('bathrooms', 341.12870661669922)]
```

Out[168]:

Effect	Coefficient
adj_sqft	2.323433
bedrooms	194.365137
bathrooms	341.128707
base_rent	1229.39138178

In [169]:
```
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:
        table.append(i)
        hoodwriter.writerow(i)

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


table
```

Out[169]:

Neighborhood	Multiplier
neighborhood_Mount Davidson Manor	0.580728
neighborhood_Ingleside	0.5925
neighborhood_Visitacion Valley	0.615154
neighborhood_Portola	0.634286

neighborhood_Glen Park	0.654545
neighborhood_Bernal Heights	0.66768
neighborhood_Diamond Heights	0.674043
neighborhood_Silver Terrace	0.680315
neighborhood_Lake Shore	0.688
neighborhood_Central Richmond	0.748159
neighborhood_Anza Vista	0.752594
neighborhood_Excelsior	0.773333
neighborhood_Bayview	0.773741
neighborhood_Cole Valley/Parnassus Heights	0.809589
neighborhood_Downtown	0.813234
neighborhood_Outer Parkside	0.846406
neighborhood_Outer Richmond	0.8625
neighborhood_Buena Vista Park/Ashbury Heights	0.887671
neighborhood_Western Addition	0.904553
neighborhood_Forest Hills Extension	0.912
neighborhood_Golden Gate Heights	0.92
neighborhood_Mission Bay	0.924226
neighborhood_Oceanview	0.948571
neighborhood_Van Ness/Civic Center	0.968585
neighborhood_Central Waterfront/Dogpatch	0.96978
neighborhood_North Panhandle	0.970909
neighborhood_Miraloma Park	0.976974
neighborhood_South of Market	0.980728
neighborhood_Telegraph Hill	0.995436
neighborhood_Inner Richmond	0.999273
neighborhood_Potrero Hill	1.015724
neighborhood_Marina	1.025048
neighborhood_Noe Valley	1.044077
neighborhood_South Beach	1.044946
neighborhood_Lower Pacific Heights	1.046809
neighborhood_Lone Mountain	1.080447
neighborhood_Yerba Buena	1.11494
neighborhood_Pacific Heights	1.119049
neighborhood_Eureka Valley / Dolores Heights	1.120039
neighborhood_Inner Mission	1.143918
neighborhood_Nob Hill	1.15336
neighborhood_Duboce Triangle	1.24861
neighborhood_Russian Hill	1.25293
neighborhood_North Waterfront	1.257862
neighborhood_Mission Dolores	1.29342
neighborhood_Inner Sunset	1.405714
neighborhood_North Beach	1.589412
neighborhood_Hayes Valley	1.634044
neighborhood_Financial District/Barbary Coast	2.0
neighborhood_Alamo Square	1

In [170]:
```
# 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[170]:

	property_id	transaction_log_id	id	address	neighborhood	bedrooms	bathrooms	sqft	source	origin_url	...	Year_2012
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
1	2	2	2	252 Granada Avenue	San Francisco (Ingleside)	2	2	1600	climbsf_renting	http://www.climbsf.com/for-rent/252-granada-ave/	...	0
2	3	3	3	460 Valley Street	San Francisco (Noe Valley)	2	2	1446	climbsf_renting	http://www.climbsf.com/for-rent/460-valley-st/	...	0
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
4	5	5	5	420 Mission Bay Boulevard North #121	San Francisco (Mission Bay)	1	1	980	climbsf_renting	http://www.climbsf.com/for-rent/420-mission-ba...	...	0

5 rows × 90 columns

In [171]:
```
# 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[171]:

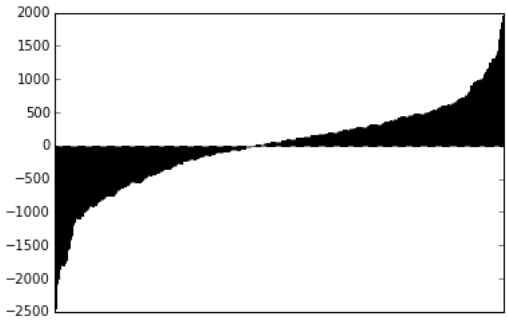
	address	neighborhood	origin_url	latitude	longitude	price	error
546	301 Main St UNIT 35A, San Francisco, CA 94105	South Beach	http://www.zillow.com/homedetails/301-Main-St-...	37.7894	-122.391	7950	-2571.090349
233	338 Spear Street #39E	San Francisco (South Beach)	http://www.climbsf.com/for-rent/338-spear-st-39e/	37.7894	-122.391	7975	-2121.288504
273	301 Mission Street #29F	San Francisco (South Beach)	http://www.climbsf.com/for-rent/301-mission-st...	37.7905	-122.396	7975	-2047.617280
517	20th St San Francisco, CA 94110	None	http://www.zillow.com/homedetails/20th-St-San-...	37.7588	-122.416	6200	-2043.079832
434	748 Bay St, San Francisco, CA 94109	Russian Hill	http://www.zillow.com/homedetails/748-Bay-St-S...	37.8049	-122.419	7500	-1982.857244
158	338 Spear Street #39A	San Francisco (South Beach)	http://www.climbsf.com/for-rent/338-spear-st-39a/	37.7894	-122.391	6700	-1893.917532
382	480 Mission Bay Boulevard North #PH1606	San Francisco (Mission Bay)	http://www.climbsf.com/for-rent/480-mission-ba...	37.7731	-122.393	7500	-1815.353787
89	88 King Street #904	San Francisco (South Beach)	http://www.climbsf.com/for-rent/88-king-st-904/	37.7807	-122.389	6250	-1808.208257
299	401 Harrison Street #3803	San Francisco (Rincon Hill)	http://www.climbsf.com/for-rent/401-harrison-s...	37.7864	-122.392	7225	-1673.712059
232	2560 Vallejo Street	San Francisco (Pacific Heights)	http://www.climbsf.com/for-rent/2560-vallejo-st/	37.7950	-122.439	7050	-1599.302373
525	20th St San Francisco, CA 94114	None	http://www.zillow.com/homedetails/20th-St-San-...	37.7578	-122.432	5700	-1528.763634
381	301 Main Street #35F	San Francisco (South Beach)	http://www.climbsf.com/for-rent/301-main-st-35f/	37.7894	-122.391	7000	-1487.557871
457	Lombard St San Francisco, CA 94133	None	http://www.zillow.com/homedetails/Lombard-St-S...	37.8021	-122.419	6700	-1392.986525
357	296 Francisco Street	San Francisco (Telegraph	http://www.climbsf.com/for-rent/296-	37.8053	-122.410	5475	-1232.987894

		Hill)	francisco-st/				
459	Tehama St San Francisco, CA 94103	None	http://www.zillow.com/homedetails/Tehama-St-Sa...	37.7793	-122.407	6000	-1225.930548
203	461 2nd St. #557T	San Francisco (South Beach)	http://www.climbsf.com/for-rent/461-2nd-st-557t/	37.7838	-122.394	6750	-1137.802049
293	425 1st Street #3402	San Francisco (Rincon Hill)	http://www.climbsf.com/for-rent/425-1st-st-3402/	37.7858	-122.392	6600	-1121.547956
119	1837 Jefferson Street	San Francisco (Marina)	http://www.climbsf.com/for-rent/1837-jefferson...	37.8045	-122.443	6200	-1098.241906
408	Van Ness Ave San Francisco, CA 94102	None	http://www.zillow.com/homedetails/Van-Ness-Ave...	37.7767	-122.419	4500	-1097.671345
113	301 Mission Street #701	San Francisco (SOMA)	http://www.climbsf.com/for-rent/301-mission-st...	37.7905	-122.396	7400	-1079.024926
204	1839 Jefferson Street	San Francisco (Marina)	http://www.climbsf.com/for-rent/1839-jefferson...	37.8048	-122.443	6400	-1058.718792
411	Vallejo St San Francisco, CA 94133	None	http://www.zillow.com/homedetails/Vallejo-St-S...	37.7985	-122.410	4000	-1056.922032
405	Vallejo St San Francisco, CA 94123	None	http://www.zillow.com/homedetails/Vallejo-St-S...	37.7952	-122.435	4200	-1037.812684
134	301 Main Street #25E	San Francisco (South Beach)	http://www.climbsf.com/for-rent/301-main-st-25e/	37.7894	-122.391	5800	-994.196530
283	234 Grand View Avenue	San Francisco (Noe Valley)	http://www.climbsf.com/for-rent/234-grand-view...	37.7545	-122.441	7300	-977.321123
282	301 Main Street #5C	San Francisco (South Beach)	http://www.climbsf.com/for-rent/301-main-st-5c/	37.7894	-122.391	7000	-920.251919
109	229 Brannan Street #12J	San Francisco (South Beach)	http://www.climbsf.com/for-rent/229-brannan-st...	37.7826	-122.390	5950	-902.984513
123	480 Mission Bay Boulevard North #1608	San Francisco (Mission Bay)	http://www.climbsf.com/for-rent/480-mission-ba...	37.7711	-122.389	5475	-883.385142
430	501 Beale St, San Francisco, CA 94105	South Beach	http://www.zillow.com/homedetails/501-Beale-St...	37.7863	-122.389	6000	-883.299578
406	San Bruno Ave San Francisco, CA 94107	None	http://www.zillow.com/homedetails/San-Bruno-Av...	37.7621	-122.405	4900	-876.886047

In [172]:
```
data = data[(data.sqft <= 2500) & (data.price <= 8000) & (data.price != 0) & (data.bedrooms <= 4) & (data.bathrooms <= 3)
& (data.sqft != 0)]



# 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: 96.06
Residual sum of squares: 434859.13
Variance score: 0.76
[('adj_sqft', 5.0230290888950986), ('bedrooms', 13.092104186543915), ('bathrooms', 257.04176156353 85), ('sqft_squared', -0
.0010601887998804621), ('beds_squared', 21.146936417757299)]
```

Out[172]:
```
[]
```



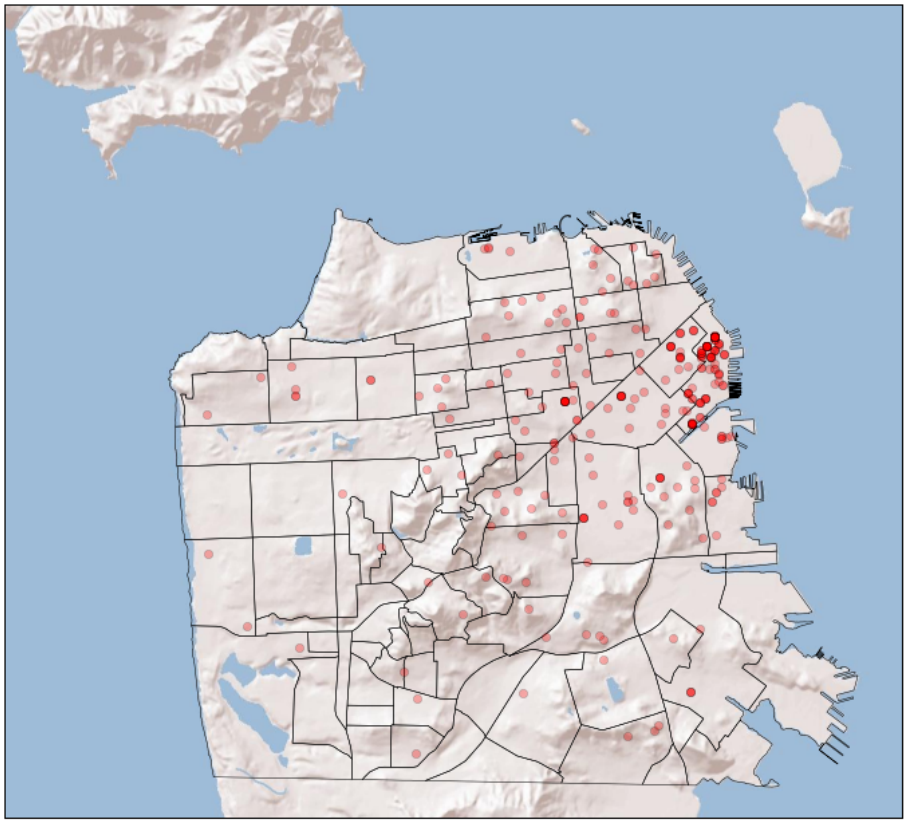
In [173]:
```
import statsmodels.formula.api as sm
result = sm.ols(formula="price ~ adj_sqft + bedrooms + bathrooms", data=data).fit()
print result.params
print result.summary()
```

```
Intercept    1229.391382
adj_sqft        2.323433
bedrooms      194.365137
bathrooms     341.128707
dtype: float64
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.725
Model:                            OLS   Adj. R-squared:                  0.722
Method:                 Least Squares   F-statistic:                     263.9
Date:                Sun, 16 Aug 2015   Prob (F-statistic):           8.33e-84
Time:                        12:11:00   Log-Likelihood:                -2422.6
No. Observations:                 304   AIC:                             4853.
Df Residuals:                     300   BIC:                             4868.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept   1229.3914    127.307      9.657      0.000       978.863  1479.920
adj_sqft       2.3234      0.139     16.768      0.000         2.051     2.596
bedrooms     194.3651     70.364      2.762      0.006        55.896   332.834
bathrooms    341.1287     96.918      3.520      0.000       150.404   531.853
==============================================================================
Omnibus:                       19.929   Durbin-Watson:                   1.844
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               42.734
Skew:                           0.316   Prob(JB):                     5.25e-10
Kurtosis:                       4.725   Cond. No.                     3.75e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.75e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [174]:
```
from mpl_toolkits.basemap import Basemap
import fiona
```

In [175]:
```
plt.figure(figsize=(12,12))


# 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)

# 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)
```

```
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 [ ]:

In [ ]: