In [96]:
```
%load_ext sql
import statsmodels.api as sma
from statsmodels.sandbox.regression.predstd import wls_prediction_std
import datetime
from scipy import stats
import math

import fiona
import shapely as shapely
from geopandas import GeoSeries, GeoDataFrame
from shapely.geometry import Point
from shapely.geometry import asShape
from time import gmtime, strftime
from array import array


# imports
import pandas as pd
import matplotlib.pyplot as plt
import csv

# follow the usual sklearn pattern: import, instantiate, fit
from sklearn.linear_model import LinearRegression
import numpy as np

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

import statsmodels.formula.api as sm

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

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

In [97]:
```
%sql mysql://prod:nerd@52.2.153.189/rental_nerd
```

Out[97]:
```
u'Connected: prod@rental_nerd'
```

In [98]:
```
result = %sql (\
SELECT  \
properties.id as 'property_id', \
properties.address,  \
properties.bedrooms,  \
properties.bathrooms,  \
properties.sqft,  \
properties.source,  \
properties.origin_url,  \
properties.longitude,  \
properties.latitude,  \
properties.elevation,  \
properties.year_built,  \
properties.garage,  \
properties.level,  \
properties.luxurious,  \
properties.dist_to_park,  \
property_transaction_logs.id 'ptl_id',  \
property_transaction_logs.transaction_type,  \
property_transaction_logs.price,  \
property_transaction_logs.transaction_status,  \
property_transaction_logs.days_on_market,  \
property_transaction_logs.date_closed,  \
property_transaction_logs.date_listed,  \
neighborhoods.name as 'neighborhood',  \
neighborhoods.id as 'nid',  \
neighborhoods.shapefile_source,  \
prediction_results.pred_std as 'pred_std' \
FROM  \
properties,  \
property_transaction_logs,  \
property_neighborhoods,  \
neighborhoods,  \
prediction_results \
WHERE  \
property_transaction_logs.property_id = properties.id AND  \
property_transaction_logs.transaction_type = "rental" AND  \
neighborhoods.shapefile_source = "PH" AND  \
properties.id = property_neighborhoods.property_id AND  \
property_neighborhoods.neighborhood_id = neighborhoods.id AND \
property_transaction_logs.date_closed is not null AND \
prediction_results.property_transaction_log_id = property_transaction_
logs.id AND \
properties.sqft > 0 AND \
prediction_results.pred_std > 0 AND \
property_transaction_logs.price > 0 )

data = result.DataFrame()
```

```
6842 rows affected.
```

In [99]:
```
result.csv(filename=strftime("%Y%m%d")+ " rentals.csv")
```

Out[99]: CSV results (./files/20160313 rentals.csv)

In [100]:
```
plot_data = data #[(data.neighborhood == 'Deer Valley')]
len(plot_data)
plot_data.describe()
```

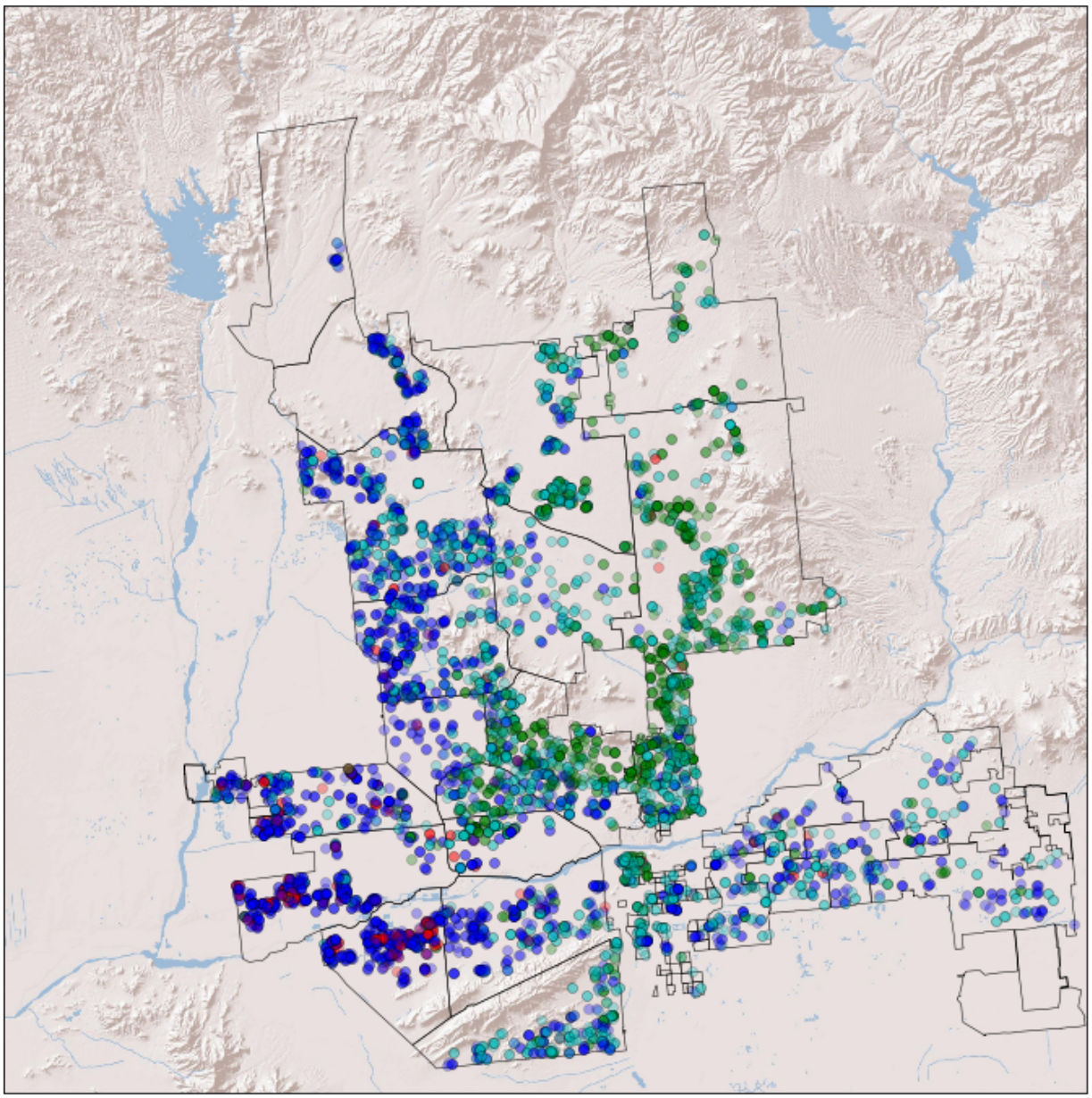
Out[100]:

	property_id	bedrooms	bathrooms	sqft	longitude	latitude
count	6842.000000	6842.000000	6842.000000	6842.000000	6842.000000	6842.000(
mean	17616.179918	3.088278	2.077024	1788.656533	-112.038399	33.52154:
std	8655.553708	0.942328	0.665565	817.102498	0.137573	0.129048
min	660.000000	0.000000	0.000000	1.000000	-112.300000	33.29220(
25%	8512.250000	3.000000	2.000000	1281.250000	-112.140000	33.41640(
50%	20670.000000	3.000000	2.000000	1622.000000	-112.053000	33.49300(
75%	24287.000000	4.000000	2.000000	2104.500000	-111.932000	33.62280(
max	30600.000000	8.000000	8.000000	11631.000000	-111.585000	33.88170(

In [101]:
```
from mpl_toolkits.basemap import Basemap
import fiona
from matplotlib.patches import Polygon
from matplotlib.collections import PatchCollection
```

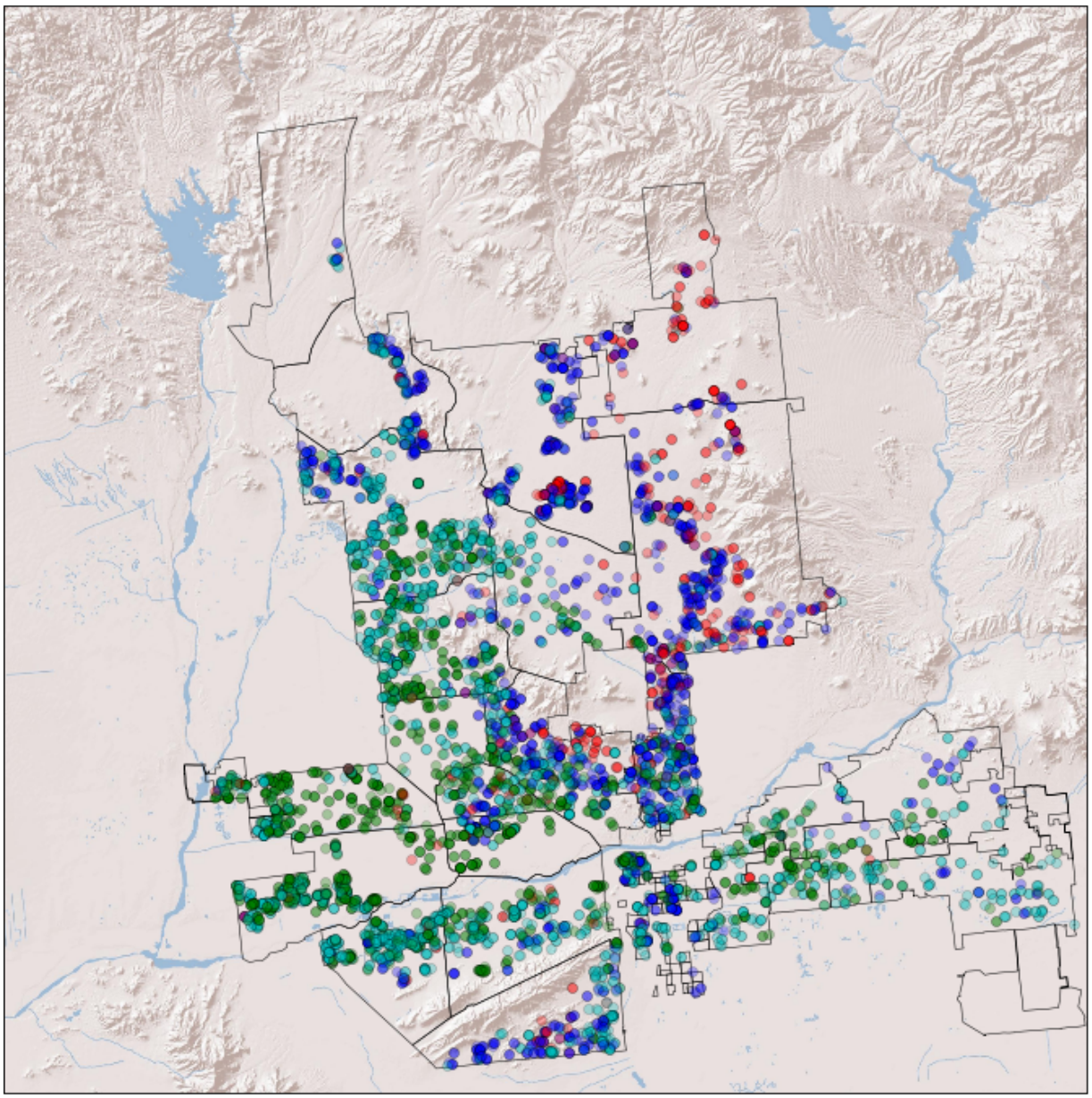
In [102]:
```
fig = plt.figure(figsize=(12,12))
ax = fig.add_subplot(111)

# Create the Basemap
event_map = Basemap(projection='merc',
                    resolution='h', epsg=2227,
                    lat_0 = 33.65, lon_0=-112, # Map center
                    llcrnrlon=-112.5, llcrnrlat=33.3, # Lower left cor
ner
                    urcrnrlon=-111.5, urcrnrlat=34) # Upper right corn
er

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

# add neighborhoods
event_map.readshapefile(
   'data/ZillowNeighborhoods-AZ/ZillowNeighborhoods-AZ', 'PHX', color=
'black', zorder=2)

# create array storing lats and longs
listing_coords = zip(plot_data.latitude,plot_data.longitude, plot_data
.sqft, plot_data.price, plot_data.pred_std)

# Draw the points on the map:
for longitude, latitude, sqft, price, pred_std in listing_coords:
    x, y = event_map(latitude, longitude) # Convert lat, long to y,x

    if((1.0 * price/sqft) < 0.5):
        color = 'ro'
    elif ((1.0 * price/sqft) < 0.8):
        color = 'bo'
    elif ((1.0 * price/sqft) <1.1):
        color = 'co'
    else:
        color = 'go'


    event_map.plot(x,y, color, alpha=0.3)

plt.show()
```

http://server.arcgisonline.com/ArcGIS/rest/services/World_Shaded_Relief/MapServer/export?bbox=9010228.80241,578944.092374,9290888.52907,860909.821619&bboxSR=2227&imageSR=2227&size=1500,1506&dpi=96&format=png32&f=image



In [103]:
```
from decimal import *

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 = 33.65, lon_0=-112, # Map center
                    llcrnrlon=-112.5, llcrnrlat=33.3, # Lower left cor
ner
                    urcrnrlon=-111.5, urcrnrlat=34) # Upper right corn
er


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

# add neighborhoods
event_map.readshapefile(
   'data/ZillowNeighborhoods-AZ/ZillowNeighborhoods-AZ', 'PHX', color=
'black', zorder=2)

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

# Draw the points on the map:
for longitude, latitude, price, pred_std in listing_coords:
    x, y = event_map(latitude, longitude) # Convert lat, long to y,x

    price_d = Decimal(price)
    pred_d = Decimal(pred_std)
    if((pred_d/price_d) < Decimal(0.1)):
        color = 'ro'
    elif ((pred_d/price_d) < Decimal(0.2)):
        color = 'bo'
    elif ((pred_d/price_d) < Decimal(0.3)):
        color = 'co'
    else:
        color = 'go'


    event_map.plot(x,y, color, alpha=0.3)

plt.show()
```

http://server.arcgisonline.com/ArcGIS/rest/services/World_Shaded_Relief/MapServer/export?bbox=9010228.80241,578944.092374,9290888.52907,860909.821619&bboxSR=2227&imageSR=2227&size=1500,1506&dpi=96&format=png32&f=image



In [104]:
```
# fill NaN values with some reasonable defaults
data.year_built = data.year_built.fillna(1970)
data.head()
```

Out[104]:

	property_id	address	bedrooms	bathrooms	sqft	source	origin_url
0	2514	11515 N 91st St UNIT 140, Scottsdale, AZ 85260	2	2	1270	zillow_ph	http://www.zillow.cor N-91st...
1	2543	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	2	2	1279	zillow_ph	http://www.zillow.cor N-94th...
2	2600	7979 E Princess Dr, Scottsdale, AZ 85255	3	3	2253	zillow_ph	http://www.zillow.cor E-Princ...
3	2684	9990 N Scottsdale Rd APT 1005, Scottsdale, AZ ...	2	2	973	zillow_ph	http://www.zillow.cor N-Scott...
4	2719	8027 E Del Timbre Dr, Scottsdale, AZ 85258	3	3	2778	zillow_ph	http://www.zillow.cor E-Del-T...

5 rows × 26 columns

In [105]:
```
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[105]:

	property_id	address	bedrooms	bathrooms	sqft	source	origin_url
0	2514	11515 N 91st St UNIT 140, Scottsdale, AZ 85260	2	2	1270	zillow_ph	http://www.zillow.cor N-91st...
1	2543	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	2	2	1279	zillow_ph	http://www.zillow.cor N-94th...
2	2600	7979 E Princess Dr, Scottsdale, AZ 85255	3	3	2253	zillow_ph	http://www.zillow.cor E-Princ...
3	2684	9990 N Scottsdale Rd APT 1005, Scottsdale, AZ ...	2	2	973	zillow_ph	http://www.zillow.cor N-Scott...
4	2719	8027 E Del Timbre Dr, Scottsdale, AZ 85258	3	3	2778	zillow_ph	http://www.zillow.cor E-Del-T...

5 rows × 27 columns

In [106]:
```
# 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 [107]:
```
# create year dummy variables (because date isn't very intuitive variable)
data["year"] = pd.DatetimeIndex(data["date"]).to_period('A')
```

In [108]:
```
data.describe() #identify filtering tresholds to clean up bad data
```

Out[108]:

	property_id	bedrooms	bathrooms	sqft	longitude	latitude
count	6842.000000	6842.000000	6842.000000	6842.000000	6842.000000	6842.000(
mean	17616.179918	3.088278	2.077024	1788.656533	-112.038399	33.52154:
std	8655.553708	0.942328	0.665565	817.102498	0.137573	0.129048
min	660.000000	0.000000	0.000000	1.000000	-112.300000	33.29220(
25%	8512.250000	3.000000	2.000000	1281.250000	-112.140000	33.41640(
50%	20670.000000	3.000000	2.000000	1622.000000	-112.053000	33.49300(
75%	24287.000000	4.000000	2.000000	2104.500000	-111.932000	33.62280(
max	30600.000000	8.000000	8.000000	11631.000000	-111.585000	33.88170(

In [109]:
```
# filter out any outliers, defined as rent >$10k or >2,500 sq ft, or n
ot in SF

print "Entries before filter: " + `len(data)`
data = data[  (data.sqft <= 10000)
            & (data.price <= 10000)
            & (data.price != 0)
            & (data.neighborhood == 'South Scottsdale')
            #& (data.transaction_status == 'closed')
            & (data.bedrooms <= 6)
            & (data.bathrooms <= 6)
            & (data.sqft != 0)
            & (data.year != pd.Period('1969', freq='A-DEC'))] #include
everything closed

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

```
Entries before filter: 6842
Entries after filter: 482
```

Out[109]:

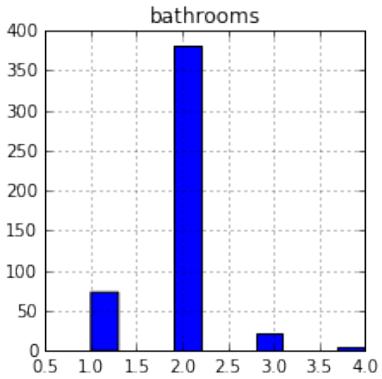
	property_id	address	bedrooms	bathrooms	sqft	source	origin_url
1376	2563	8120 E Valley View Rd, Scottsdale, AZ 85250	3	2	2100	zillow_ph	http://www.zillow E-Valle...
1377	2606	6620 E Earll Dr APT 2, Scottsdale, AZ 85251	1	1	600	zillow_ph	http://www.zillow E-Earll...
1378	2641	7777 E Main St UNIT 153, Scottsdale, AZ 85251	2	2	1130	zillow_ph	http://www.zillow E-Main-...
1379	2664	7777 E Main St UNIT 304, Scottsdale, AZ 85251	2	2	1130	zillow_ph	http://www.zillow E-Main-...
1380	2674	7777 E Main St UNIT 124, Scottsdale, AZ 85251	2	2	1130	zillow_ph	http://www.zillow E-Main-...

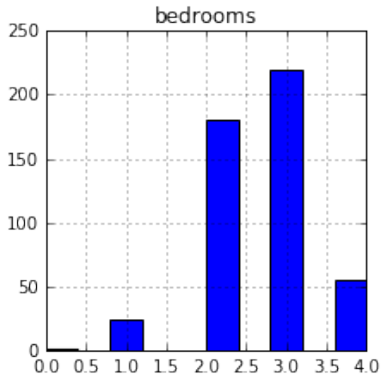
5 rows × 29 columns

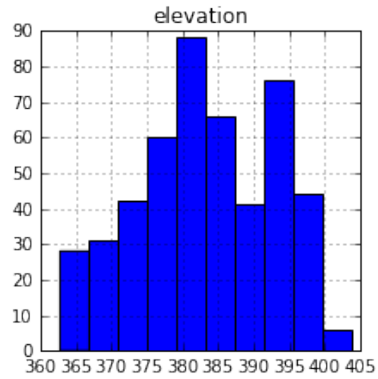
In [ ]:

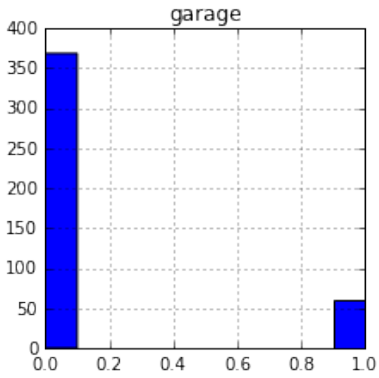
In [110]:
```
data.hist(column=['bathrooms','bedrooms','price','garage','level','yea
r_built','sqft','elevation','luxurious','dist_to_park'],figsize=(12,12
))
```

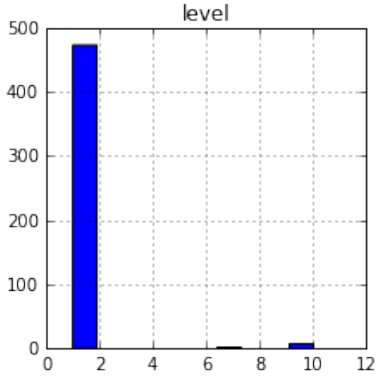
```
Out[110]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x121c3a350
          >,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x1220937d0
          >,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x122245a50
          >],
                 [<matplotlib.axes._subplots.AxesSubplot object at 0x121c2ee10
          >,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x1222bbe90
          >,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x121578750
          >],
                 [<matplotlib.axes._subplots.AxesSubplot object at 0x1218bca90
          >,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x121910e10
          >,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x121a05cd0
          >]], dtype=object)
```

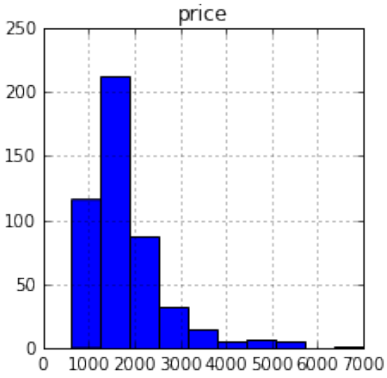

bathrooms
400
350
300
250
200
150
100
50
0
0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0

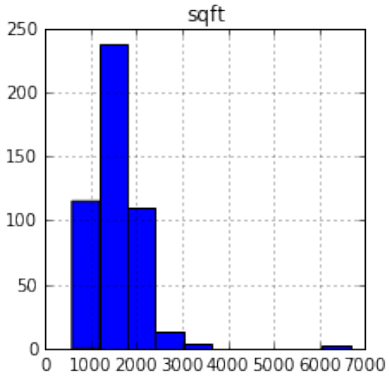

bedrooms
250
200
150
100
50
0
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0

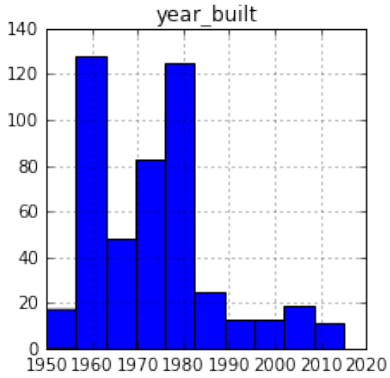

elevation
90
80
70
60
50
40
30
20
10
0
360 365 370 375 380 385 390 395 400 405


garage
400
350
300
250
200
150
100
50
0
0.0 0.2 0.4 0.6 0.8 1.0


level
500
400
300
200
100
0
0 2 4 6 8 10 12


price
250
200
150
100
50
0
0 1000 2000 3000 4000 5000 6000 7000


sqft
250
200
150
100
50
0
0 1000 2000 3000 4000 5000 6000 7000


year_built
140
120
100
80
60
40
20
0
1950 1960 1970 1980 1990 2000 2010 2020

In [111]:
```
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 [112]:
```
age = 2016 - data.year_built
age.name = 'age'
data = pd.concat([data,age],axis=1)
```

In [113]:
```
result = sm.ols(formula="price ~ bedrooms + bathrooms + age + elevatio
n +\
neighborhood:sqft:year -1", data=data).fit()
result.summary()
```

Out[113]:

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.427
Model:	OLS	**Adj. R-squared:**	0.413
Method:	Least Squares	**F-statistic:**	29.18
Date:	Sun, 13 Mar 2016	**Prob (F-statistic):**	1.48e-49
Time:	13:55:00	**Log-Likelihood:**	-3814.5
No. Observations:	482	**AIC:**	7655.
Df Residuals:	469	**BIC:**	7709.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf.**

					Int.]
Intercept	-1.17e+04	1393.105	-8.399	0.000	-1.44e+04 -8962.544
bedrooms	105.3274	55.807	1.887	0.060	-4.336 214.991
bathrooms	377.1120	77.044	4.895	0.000	225.717 528.507
age	-11.4125	2.396	-4.763	0.000	-16.121 -6.704
elevation	32.2234	3.599	8.954	0.000	25.151 39.295
neighborhood[South Scottsdale]:sqft:year[Period('2009', 'A-DEC')]	1.4258	0.296	4.824	0.000	0.845 2.007
neighborhood[South Scottsdale]:sqft:year[Period('2010', 'A-DEC')]	0.1612	0.163	0.986	0.324	-0.160 0.482
neighborhood[South Scottsdale]:sqft:year[Period('2011', 'A-DEC')]	0.1912	0.131	1.455	0.146	-0.067 0.449
neighborhood[South Scottsdale]:sqft:year[Period('2012', 'A-DEC')]	0.3346	0.095	3.512	0.000	0.147 0.522
neighborhood[South Scottsdale]:sqft:year[Period('2013', 'A-DEC')]	0.3944	0.103	3.844	0.000	0.193 0.596
neighborhood[South Scottsdale]:sqft:year[Period('2014', 'A-DEC')]	0.3399	0.086	3.931	0.000	0.170 0.510
neighborhood[South Scottsdale]:sqft:year[Period('2015', 'A-DEC')]	0.4216	0.065	6.466	0.000	0.293 0.550
neighborhood[South Scottsdale]:sqft:year[Period('2016', 'A-DEC')]	0.4735	0.076	6.201	0.000	0.323 0.624

Omnibus:	186.955	**Durbin-Watson:**	1.512
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1172.607
Skew:	1.552	**Prob(JB):**	2.35e-255
Kurtosis:	9.982	**Cond. No.**	4.80e+04

In [114]:
```
output = [['neighborhood','rent per foot']]
table = ListTable()
table.append(output[0])

for row in data.neighborhood.unique():
    output_row = [row, '99']
    for i in result.params.index:
        if 'neighborhood' not in i: continue

        if '2015' in i:
            if 'neighborhood[' + row + ']' in i:
                output_row[1] = `result.params[i]`
                output.append(output_row)
                table.append(output_row)

table
```

Out[114]:

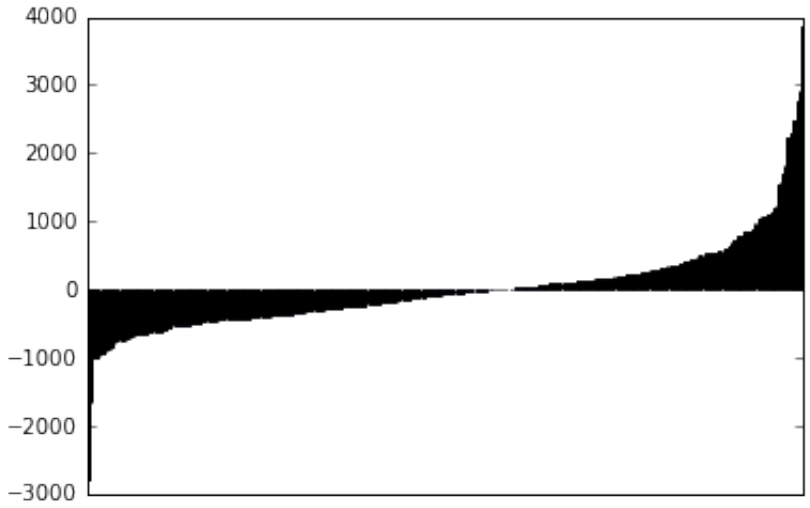
neighborhood	rent per foot
South Scottsdale	0.42162683752082708

In [115]:
```
path = 'rentalnerd_importer/lib/tasks/model_files/'

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

with open(path + 'model_features_ph.csv', 'wb') as csvfile:
    modelwriter = csv.writer(csvfile, delimiter=',', quotechar='|', qu
oting=csv.QUOTE_MINIMAL)

    header = ['Effect','Coefficient']
    table.append(header)
    modelwriter.writerow(header)
    modelwriter.writerow(['base_rent', 0]) # result.params.Intercept])
hardcode 0 as base rent
    modelwriter.writerow(['bedrooms', result.params.bedrooms])
    modelwriter.writerow(['bathrooms', result.params.bathrooms])
    modelwriter.writerow(['dist_to_park', 0])
    modelwriter.writerow(['elevation', result.params.elevation])
    modelwriter.writerow(['level', 0])
    modelwriter.writerow(['age', result.params.age])
    modelwriter.writerow(['garage', 0])
    modelwriter.writerow(['mean square error of residuals', result.mse
_resid])

result.cov_params().to_csv(path + 'model_covs_ph.csv')
```

In [116]:
```
with open(path + 'model_hoods_ph.csv', 'wb') as csvfile:
    hoodwriter = csv.writer(csvfile, delimiter=',', quotechar='|', quo
ting=csv.QUOTE_MINIMAL)

    for i in output:
        hoodwriter.writerow(i)
```

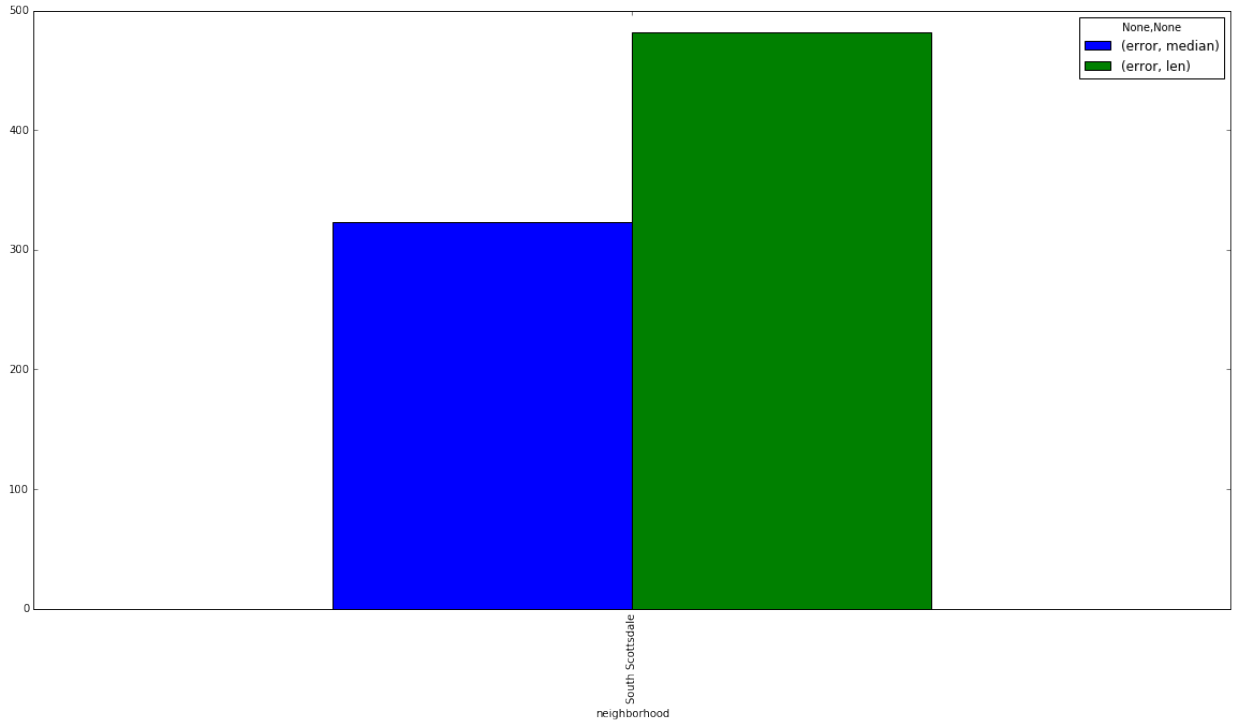
In [117]:
```
hood = "North Scottsdale"
subdata = data[(data.neighborhood == hood) & (data.date > datetime.dat
e(2011, 5, 1))]
plt.plot_date(x=subdata.date,y=subdata.price)
plt.title(hood + " Rental Listings")
plt.grid(True)
```

In [118]:
```
errors = result.resid
errors.name = 'error'
pprice = errors + data.price
pprice.name = "prediction"
data = pd.concat([data, errors], axis=1)
data = pd.concat([data, pprice], axis=1)

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

# show errors by neighborhood to see if there are any neighborhoods wi
th funky differences

hooderrors = data[['neighborhood']]

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

hood_group = hooderrors.groupby('neighborhood')
```



In [119]:
```
error_avg = hood_group.agg([np.median,len])
error_avg.sort_values(by=('error','median'),ascending=False,inplace=True)
error_avg.plot(kind='bar',figsize=(20,10))
```

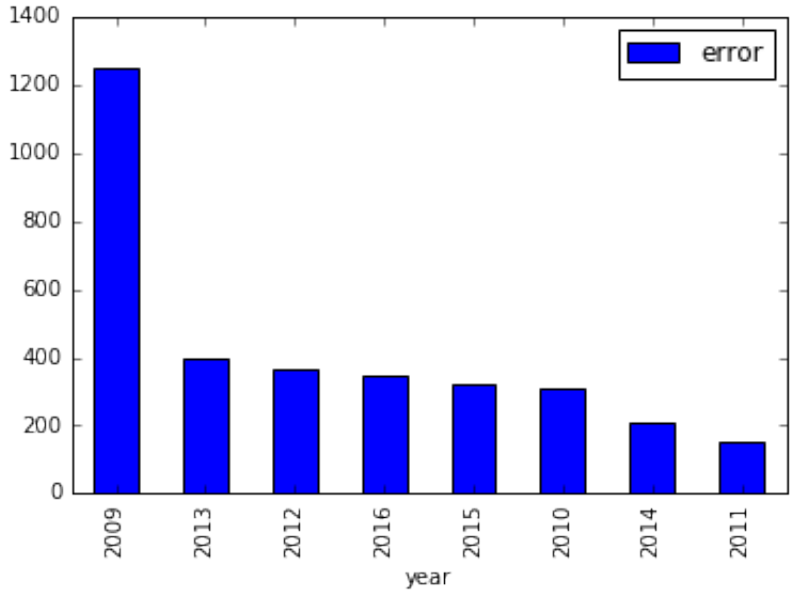
Out[119]: <matplotlib.axes._subplots.AxesSubplot at 0x121a91650>

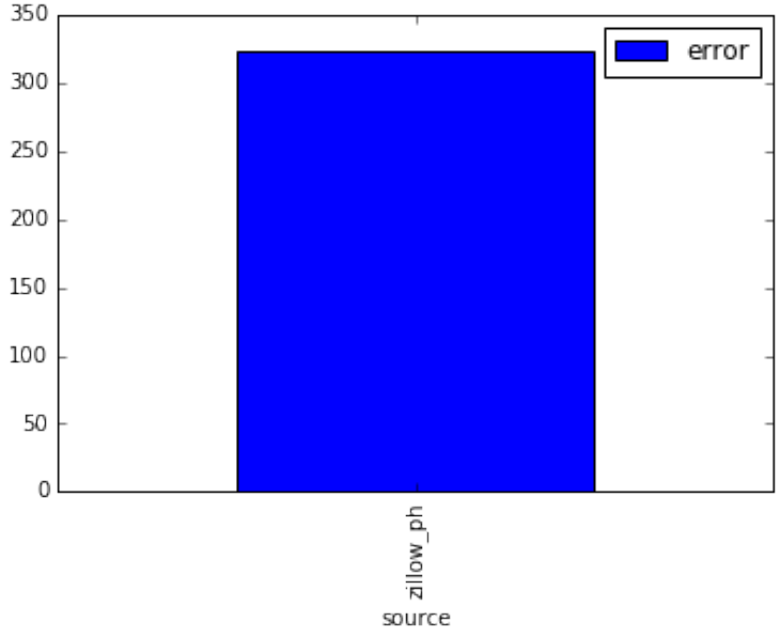


In [ ]:

In [120]:
```
# 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.median()
error_avg.sort_values(by='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.median()
error_avg.sort_values(by='error',ascending=False).plot(kind='bar')
```

Out[120]: <matplotlib.axes._subplots.AxesSubplot at 0x12528e590>





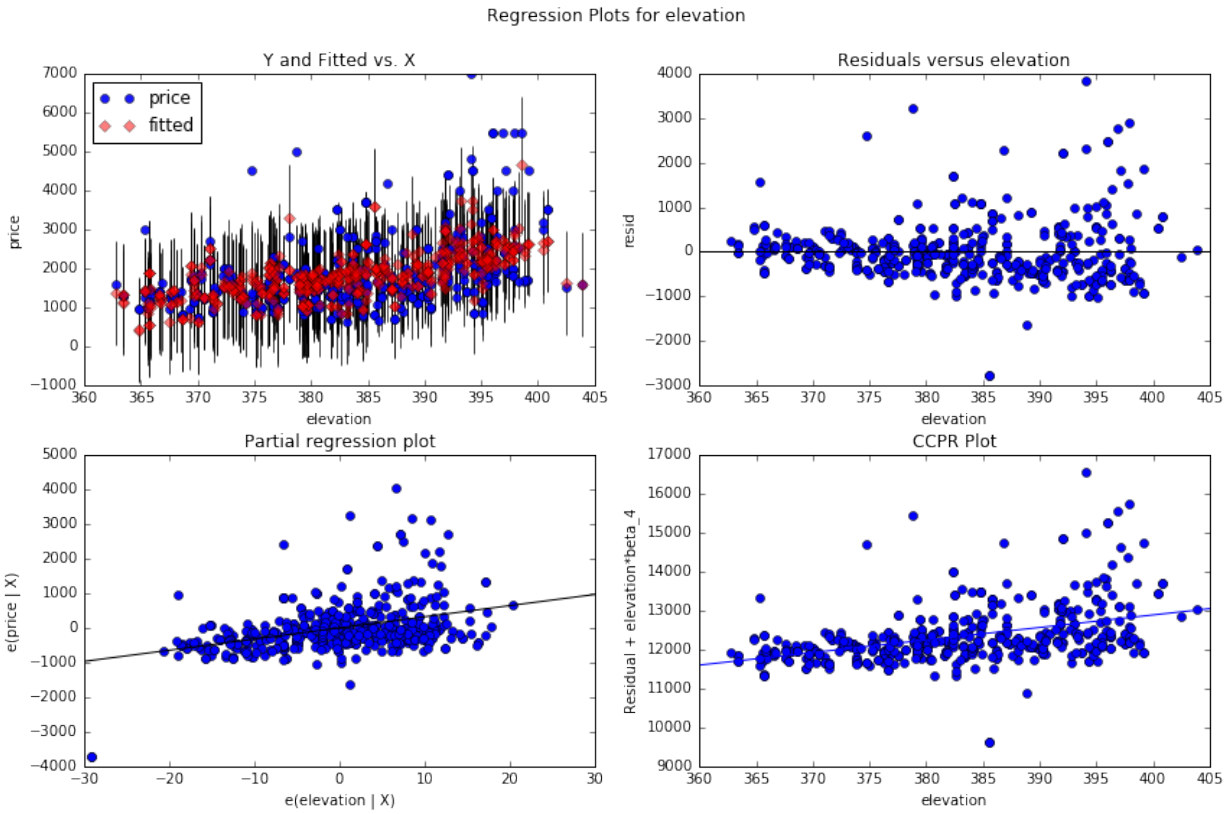
In [121]:
```
data['price'].mean
```

Out[121]:
```
<bound method Series.mean of 1376    2500
1377     625
1378    2100
1379    2100
1380    1400
1381     875
1382    2800
1383    1600
1384    2750
```

1385	1975
1386	1550
1387	1550
1388	690
1389	690
1390	1600
1391	1600
1392	1175
1393	1175
1394	1495
1395	1495
1396	1750
1397	1750
1398	695
1399	695
1400	2800
1401	2800
1402	1600
1403	1600
1404	1550
1405	3275
	...
1840	4000
1841	1150
1842	1150
1843	2100
1844	2500
1845	1095
1846	1500
1847	1400
1848	900
1849	975
1850	2500
1851	5500
1852	1795
1853	3400
1854	1150
1855	1850
1856	2200
1857	2200
1858	2450
1859	2000
1860	1800
1861	1250
1862	1250
1863	1745
1864	1600
1865	2600
1866	1600
1867	1600

```
1868    5500
1869    1650
Name: price, dtype: int64>
```

In [122]:
```
fig = plt.figure(figsize=(12,8))
fig = sma.graphics.plot_regress_exog(result, "elevation", fig=fig)
```



In [123]:
```
# prstd, iv_l, iv_u = wls_prediction_std(result)

# zip(data.address, data.price, data.prediction, prstd, iv_l, iv_u)
```

In [132]:
```
# Artnet white paper index converted to our dataset

# create year dummy variables (because date isn't very intuitive varia
ble)
f = 'M'
data["period"] = pd.DatetimeIndex(data["date"]).to_period(f)
```

In [133]:
```
data[['address','price','period','neighborhood']]
```

Out[133]:

	address	price	period	neighborhood
1376	8120 E Valley View Rd, Scottsdale, AZ 85250	2500	2015-09	South Scottsdale

1377	6620 E Earll Dr APT 2, Scottsdale, AZ 85251	625	2015-09	South Scottsdale
1378	7777 E Main St UNIT 153, Scottsdale, AZ 85251	2100	2015-09	South Scottsdale
1379	7777 E Main St UNIT 304, Scottsdale, AZ 85251	2100	2015-09	South Scottsdale
1380	7777 E Main St UNIT 124, Scottsdale, AZ 85251	1400	2015-09	South Scottsdale
1381	2614 N 72nd Pl UNIT 3, Scottsdale, AZ 85257	875	2015-09	South Scottsdale
1382	6711 E Camelback Rd UNIT 36, Scottsdale, AZ 85251	2800	2015-09	South Scottsdale
1383	5101 N Casa Blanca Dr UNIT 3, Paradise Valley,...	1600	2015-10	South Scottsdale
1384	7127 E Rancho Vista Dr # 3005, Scottsdale, AZ ...	2750	2015-10	South Scottsdale
1385	4020 N Scottsdale Rd # 3010, Scottsdale, AZ 85251	1975	2015-08	South Scottsdale
1386	7940 E Camelback Rd UNIT 501, Scottsdale, AZ 8...	1550	2015-10	South Scottsdale
1387	7940 E Camelback Rd UNIT 501, Scottsdale, AZ 8...	1550	2015-10	South Scottsdale
1388	4525 N 74th St APT 8, Scottsdale, AZ 85251	690	2015-10	South Scottsdale
1389	4525 N 74th St APT 8, Scottsdale, AZ 85251	690	2015-10	South Scottsdale
1390	6824 E 2nd St # 2, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1391	6824 E 2nd St # 2, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1392	3500 N Hayden Rd APT 702, Scottsdale, AZ 85251	1175	2015-10	South Scottsdale
1393	3500 N Hayden Rd APT 702, Scottsdale, AZ 85251	1175	2015-10	South Scottsdale
			2015-	South

1394	6532 E Cypress Cir, Scottsdale, AZ 85257	1495	10	Scottsdale
1395	6532 E Cypress Cir, Scottsdale, AZ 85257	1495	2015-10	South Scottsdale
1396	3508 N 81st St, Scottsdale, AZ 85251	1750	2015-10	South Scottsdale
1397	3508 N 81st St, Scottsdale, AZ 85251	1750	2015-10	South Scottsdale
1398	3126 N 68th St APT 1002, Scottsdale, AZ 85251	695	2015-10	South Scottsdale
1399	3126 N 68th St APT 1002, Scottsdale, AZ 85251	695	2015-10	South Scottsdale
1400	6565 E Thomas Rd #F1041, Scottsdale, AZ 85251	2800	2015-10	South Scottsdale
1401	6565 E Thomas Rd #F1041, Scottsdale, AZ 85251	2800	2015-10	South Scottsdale
1402	3914 N 83rd St, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1403	3914 N 83rd St, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1404	8353 E Granada Rd, Scottsdale, AZ 85257	1550	2015-10	South Scottsdale
1405	4724 N 69th St, Scottsdale, AZ 85251	3275	2014-06	South Scottsdale
...	...	...	...	...
1840	7822 E Sorrel Wood Ct, Scottsdale, AZ 85258	4000	2016-01	South Scottsdale
1841	7901 E Mckinley St, Scottsdale, AZ 85257	1150	2016-02	South Scottsdale
1842	7901 E Mckinley St, Scottsdale, AZ 85257	1150	2016-03	South Scottsdale
1843	8101 E Via Del Desierto, Scottsdale, AZ 85258	2100	2016-02	South Scottsdale
1844	8225 E Via De La Escuela, Scottsdale, AZ 85258	2500	2011-11	South Scottsdale

1845	8343 E Cypress St, Scottsdale, AZ 85257	1095	2011-04	South Scottsdale
1846	8437 E Bonnie Rose Ave, Scottsdale, AZ 85250	1500	2014-09	South Scottsdale
1847	8544 E Plaza Ave, Scottsdale, AZ 85250	1400	2015-03	South Scottsdale
1848	8625 E Belleview Pl UNIT 1053, Scottsdale, AZ ...	900	2014-04	South Scottsdale
1849	8625 E Belleview Pl UNIT 1053, Scottsdale, AZ ...	975	2016-02	South Scottsdale
1850	3130 N 62nd St, Scottsdale, AZ 85251	2500	2016-03	South Scottsdale
1851	8355 E Via De La Luna, Scottsdale, AZ 85258	5500	2016-03	South Scottsdale
1852	2018 N 81st Pl, Scottsdale, AZ 85257	1795	2016-03	South Scottsdale
1853	7625 E Indian Bend Rd, Scottsdale, AZ 85250	3400	2013-12	South Scottsdale
1854	6945 E 2nd St, Scottsdale, AZ 85251	1150	2016-03	South Scottsdale
1855	7157 E Rancho Vista Dr UNIT 5008, Scottsdale, ...	1850	2010-06	South Scottsdale
1856	7157 E Rancho Vista Dr UNIT 5008, Scottsdale, ...	2200	2016-01	South Scottsdale
1857	7157 E Rancho Vista Dr UNIT 5008, Scottsdale, ...	2200	2013-09	South Scottsdale
1858	7157 N Via De La Campana, Scottsdale, AZ 85258	2450	2016-03	South Scottsdale
1859	8121 E Via De Viva, Scottsdale, AZ 85258	2000	2016-03	South Scottsdale
1860	8121 E Via De Viva, Scottsdale, AZ 85258	1800	2016-03	South Scottsdale
1861	6350 N 78th St UNIT 267, Scottsdale, AZ 85250	1250	2010-01	South Scottsdale
			2009-	South

1862	6350 N 78th St UNIT 267, Scottsdale, AZ 85250	1250	10	Scottsdale
1863	8356 E Palm Ln, Scottsdale, AZ 85257	1745	2016-03	South Scottsdale
1864	7610 E Minnezona Ave, Scottsdale, AZ 85251	1600	2010-02	South Scottsdale
1865	7812 N Via Del Sol, Scottsdale, AZ 85258	2600	2016-03	South Scottsdale
1866	4525 N 66th St, Scottsdale, AZ 85251	1600	2016-03	South Scottsdale
1867	4525 N 66th St, Scottsdale, AZ 85251	1600	2016-03	South Scottsdale
1868	7844 E Via Costa, Scottsdale, AZ 85258	5500	2009-12	South Scottsdale
1869	8154 E Via De La Escuela, Scottsdale, AZ 85258	1650	2016-03	South Scottsdale

482 rows × 4 columns

In [134]:
```
paired = data[['property_id','address','price','period','neighborhood'
]]

# identify the earliest date, number of periods, and number of pairs
base_period = paired.period.min()
num_periods = paired.period.max() - paired.period.min()
print "base period: " + `base_period` + " end period: " + `paired.peri
od.max()` + " and number of periods: " + `num_periods`

paired.head()
```

```
base period: Period('2009-10', 'M') end period: Period('2016-03', 'M
') and number of periods: 77
```

Out[134]:

	property_id	address	price	period	neighborhood
1376	2563	8120 E Valley View Rd, Scottsdale, AZ 85250	2500	2015-09	South Scottsdale
1377	2606	6620 E Earll Dr APT 2, Scottsdale, AZ 85251	625	2015-09	South Scottsdale
1378	2641	7777 E Main St UNIT 153, Scottsdale, AZ 85251	2100	2015-09	South Scottsdale
1379	2664	7777 E Main St UNIT 304, Scottsdale, AZ 85251	2100	2015-09	South Scottsdale
1380	2674	7777 E Main St UNIT 124, Scottsdale, AZ 85251	1400	2015-09	South Scottsdale

In [135]:
```
paired
```

Out[135]:

	property_id	address	price	period	neighborhood
1376	2563	8120 E Valley View Rd, Scottsdale, AZ 85250	2500	2015-09	South Scottsdale
1377	2606	6620 E Earll Dr APT 2, Scottsdale, AZ 85251	625	2015-09	South Scottsdale
1378	2641	7777 E Main St UNIT 153, Scottsdale, AZ 85251	2100	2015-09	South Scottsdale
1379	2664	7777 E Main St UNIT 304, Scottsdale, AZ 85251	2100	2015-09	South Scottsdale
1380	2674	7777 E Main St UNIT 124, Scottsdale, AZ 85251	1400	2015-09	South Scottsdale

1381	2724	2614 N 72nd Pl UNIT 3, Scottsdale, AZ 85257	875	2015-09	South Scottsdale
1382	2725	6711 E Camelback Rd UNIT 36, Scottsdale, AZ 85251	2800	2015-09	South Scottsdale
1383	2857	5101 N Casa Blanca Dr UNIT 3, Paradise Valley,...	1600	2015-10	South Scottsdale
1384	3009	7127 E Rancho Vista Dr # 3005, Scottsdale, AZ ...	2750	2015-10	South Scottsdale
1385	3010	4020 N Scottsdale Rd # 3010, Scottsdale, AZ 85251	1975	2015-08	South Scottsdale
1386	3135	7940 E Camelback Rd UNIT 501, Scottsdale, AZ 8...	1550	2015-10	South Scottsdale
1387	3135	7940 E Camelback Rd UNIT 501, Scottsdale, AZ 8...	1550	2015-10	South Scottsdale
1388	3184	4525 N 74th St APT 8, Scottsdale, AZ 85251	690	2015-10	South Scottsdale
1389	3184	4525 N 74th St APT 8, Scottsdale, AZ 85251	690	2015-10	South Scottsdale
1390	3224	6824 E 2nd St # 2, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1391	3224	6824 E 2nd St # 2, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1392	3257	3500 N Hayden Rd APT 702, Scottsdale, AZ 85251	1175	2015-10	South Scottsdale
1393	3257	3500 N Hayden Rd APT 702, Scottsdale, AZ 85251	1175	2015-10	South Scottsdale
1394	3274	6532 E Cypress Cir, Scottsdale, AZ 85257	1495	2015-10	South Scottsdale
1395	3274	6532 E Cypress Cir, Scottsdale, AZ 85257	1495	2015-10	South Scottsdale
1396	3563	3508 N 81st St, Scottsdale, AZ 85251	1750	2015-10	South Scottsdale
1397	3563	3508 N 81st St, Scottsdale, AZ 85251	1750	2015-10	South Scottsdale
		3126 N 68th St APT 1002, Scottsdale,		2015-	South

1398	3610	AZ 85251	695	10	Scottsdale
1399	3610	3126 N 68th St APT 1002, Scottsdale, AZ 85251	695	2015-10	South Scottsdale
1400	3643	6565 E Thomas Rd #F1041, Scottsdale, AZ 85251	2800	2015-10	South Scottsdale
1401	3643	6565 E Thomas Rd #F1041, Scottsdale, AZ 85251	2800	2015-10	South Scottsdale
1402	3759	3914 N 83rd St, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1403	3759	3914 N 83rd St, Scottsdale, AZ 85251	1600	2015-10	South Scottsdale
1404	3777	8353 E Granada Rd, Scottsdale, AZ 85257	1550	2015-10	South Scottsdale
1405	3835	4724 N 69th St, Scottsdale, AZ 85251	3275	2014-06	South Scottsdale
...	...	...	...	...	...
1840	26333	7822 E Sorrel Wood Ct, Scottsdale, AZ 85258	4000	2016-01	South Scottsdale
1841	26335	7901 E Mckinley St, Scottsdale, AZ 85257	1150	2016-02	South Scottsdale
1842	26335	7901 E Mckinley St, Scottsdale, AZ 85257	1150	2016-03	South Scottsdale
1843	26338	8101 E Via Del Desierto, Scottsdale, AZ 85258	2100	2016-02	South Scottsdale
1844	26339	8225 E Via De La Escuela, Scottsdale, AZ 85258	2500	2011-11	South Scottsdale
1845	26342	8343 E Cypress St, Scottsdale, AZ 85257	1095	2011-04	South Scottsdale
1846	26343	8437 E Bonnie Rose Ave, Scottsdale, AZ 85250	1500	2014-09	South Scottsdale
1847	26345	8544 E Plaza Ave, Scottsdale, AZ 85250	1400	2015-03	South Scottsdale
1848	26347	8625 E Belleview Pl UNIT 1053, Scottsdale, AZ ...	900	2014-04	South Scottsdale

1849	26347	8625 E Belleview Pl UNIT 1053, Scottsdale, AZ ...	975	2016-02	South Scottsdale
1850	26906	3130 N 62nd St, Scottsdale, AZ 85251	2500	2016-03	South Scottsdale
1851	26928	8355 E Via De La Luna, Scottsdale, AZ 85258	5500	2016-03	South Scottsdale
1852	27615	2018 N 81st Pl, Scottsdale, AZ 85257	1795	2016-03	South Scottsdale
1853	27619	7625 E Indian Bend Rd, Scottsdale, AZ 85250	3400	2013-12	South Scottsdale
1854	27807	6945 E 2nd St, Scottsdale, AZ 85251	1150	2016-03	South Scottsdale
1855	27808	7157 E Rancho Vista Dr UNIT 5008, Scottsdale, ...	1850	2010-06	South Scottsdale
1856	27808	7157 E Rancho Vista Dr UNIT 5008, Scottsdale, ...	2200	2016-01	South Scottsdale
1857	27808	7157 E Rancho Vista Dr UNIT 5008, Scottsdale, ...	2200	2013-09	South Scottsdale
1858	27809	7157 N Via De La Campana, Scottsdale, AZ 85258	2450	2016-03	South Scottsdale
1859	27811	8121 E Via De Viva, Scottsdale, AZ 85258	2000	2016-03	South Scottsdale
1860	27811	8121 E Via De Viva, Scottsdale, AZ 85258	1800	2016-03	South Scottsdale
1861	27987	6350 N 78th St UNIT 267, Scottsdale, AZ 85250	1250	2010-01	South Scottsdale
1862	27987	6350 N 78th St UNIT 267, Scottsdale, AZ 85250	1250	2009-10	South Scottsdale
1863	28797	8356 E Palm Ln, Scottsdale, AZ 85257	1745	2016-03	South Scottsdale
1864	29123	7610 E Minnezona Ave, Scottsdale, AZ 85251	1600	2010-02	South Scottsdale
1865	29124	7812 N Via Del Sol, Scottsdale, AZ 85258	2600	2016-03	South Scottsdale
				2016-	South

1866	29721	4525 N 66th St, Scottsdale, AZ 85251	1600	03	Scottsdale
1867	29722	4525 N 66th St, Scottsdale, AZ 85251	1600	2016-03	South Scottsdale
1868	29895	7844 E Via Costa, Scottsdale, AZ 85258	5500	2009-12	South Scottsdale
1869	29896	8154 E Via De La Escuela, Scottsdale, AZ 85258	1650	2016-03	South Scottsdale

482 rows × 5 columns

In [136]:
```
len(paired.groupby("address").filter(lambda x: len(x['address']) >1).groupby('property_id'))
```

Out[136]: 146

In [137]:
```
paired.groupby("address").filter(lambda x: len(x['address']) >1).groupby('address').get_group('210 W Helena Dr, Phoenix, AZ 85023')
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-137-1f59fbcdada3> in <module>()
----> 1 paired.groupby("address").filter(lambda x: len(x['address'])>1).groupby('address').get_group('210 W Helena Dr, Phoenix, AZ 85023')

/usr/local/lib/python2.7/site-packages/pandas/core/groupby.pyc in get_group(self, name, obj)
    644         inds = self._get_index(name)
    645         if not len(inds):
--> 646             raise KeyError(name)
    647
    648         return obj.take(inds, axis=self.axis, convert=False)

KeyError: '210 W Helena Dr, Phoenix, AZ 85023'
```

In [138]:
```
# group data into Sets and calc Y_ist of each item
paired = paired.drop_duplicates().groupby("address").filter(lambda x:
len(x) >1)
paired.sort_values(['address','period'],inplace=True)
paired_grp = paired.groupby('address')
print 'number of paired transactions in the data: ' + `paired.shape[0]
`
```

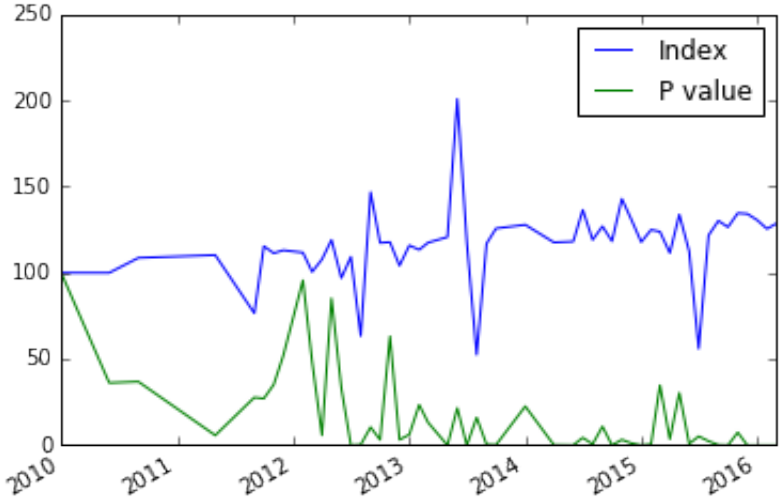
```
number of paired transactions in the data: 242
```

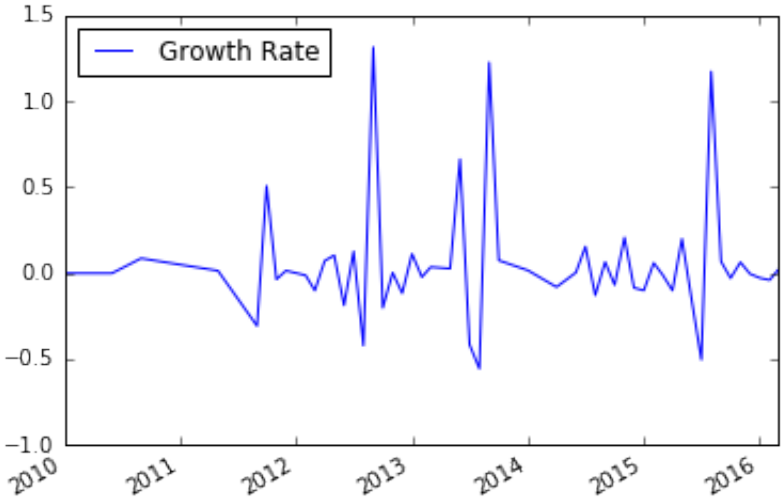
In [139]:
```
d = paired

res = sm.ols(formula="np.log(price) ~ period + address", data=d).fit()

#calculate index
linked = res.params[res.params.index.str.contains('Period')]
linked.name = "Index"
linked[0] = 100
growth = pd.Series(linked, copy=True)
growth.name = "Growth Rate"
growth[0] = 0
for i in range(1,len(linked)):
    linked[i] = (np.exp(res.params[i]))*100
    growth[i] = linked[i]/linked[i-1] - 1

# add P values of each prediction
p = res.pvalues[res.params.index.str.contains('Period')] * 100
p.name = "P value"
index = pd.concat([linked, growth, p], axis=1)
index.index = pd.to_datetime(index.index.str.split("'").str.get(1))

print index

index[['Index','P value']].plot()
index[['Growth Rate']].plot()
```

```
                Index   Growth Rate       P value
2010-01-01  100.000000  0.000000e+00  1.000000e+02
2010-06-01  100.000000  8.881784e-16  3.622030e+01
2010-09-01  108.691411  8.691411e-02  3.688223e+01
2011-05-01  110.314696  1.493481e-02  5.747308e+00
2011-09-01   76.581135 -3.057939e-01  2.772605e+01
2011-10-01  115.446650  5.075077e-01  2.712512e+01
2011-11-01  111.388555 -3.515126e-02  3.517148e+01
2011-12-01  113.090596  1.528022e-02  5.191559e+01
2012-02-01  111.765434 -1.171771e-02  9.564484e+01
2012-03-01  100.569434 -1.001741e-01  4.843830e+01
2012-04-01  107.855475  7.244787e-02  5.811828e+00
2012-05-01  119.124377  1.044815e-01  8.516324e+01
```

```
2012-06-01   96.985499 -1.858467e-01  3.309599e+01
2012-07-01  109.287649  1.268452e-01  2.006762e-02
2012-08-01   63.431659 -4.195899e-01  3.564169e-01
2012-09-01  146.770252  1.313833e+00  1.047530e+01
2012-10-01  117.372414 -2.002983e-01  3.155620e+00
2012-11-01  117.806663  3.699759e-03  6.314744e+01
2012-12-01  104.167677 -1.157743e-01  3.189437e+00
2013-01-01  115.920789  1.128288e-01  6.225652e+00
2013-02-01  113.391113 -2.182246e-02  2.347135e+01
2013-03-01  117.516064  3.637808e-02  1.320041e+01
2013-05-01  120.774554  2.772804e-02  3.422314e-03
2013-06-01  200.849120  6.630086e-01  2.156424e+01
2013-07-01  117.802704 -4.134766e-01  3.442830e-03
2013-08-01   52.556392 -5.538609e-01  1.610047e+01
2013-09-01  116.943043  1.225096e+00  4.607189e-01
2013-10-01  125.755120  7.535358e-02  2.291892e-01
2014-01-01  127.883894  1.692793e-02  2.258765e+01
2014-04-01  117.665216 -7.990591e-02  1.912744e-01
2014-06-01  118.183846  4.407675e-03  4.838938e-02
2014-07-01  136.506635  1.550363e-01  4.396878e+00
2014-08-01  119.127167 -1.273159e-01  6.611467e-01
2014-09-01  126.963386  6.578028e-02  1.088221e+01
2014-10-01  118.344607 -6.788397e-02  1.986690e-03
2014-11-01  142.943126  2.078550e-01  3.233530e+00
2014-12-01  130.872481 -8.444369e-02  1.184156e+00
2015-01-01  117.859200 -9.943481e-02  1.400757e-02
2015-02-01  124.958202  6.023290e-02  6.065300e-01
2015-03-01  123.894823 -8.509877e-03  3.480578e+01
2015-04-01  111.510605 -9.995751e-02  3.594690e+00
2015-05-01  133.806206  1.999415e-01  3.039377e+01
2015-06-01  112.467836 -1.594722e-01  8.941176e-01
2015-07-01   56.060151 -5.015450e-01  5.175319e+00
2015-08-01  121.785713  1.172411e+00  2.403115e+00
2015-09-01  130.180999  6.893490e-02  1.766552e-01
2015-10-01  126.507744 -2.821652e-02  3.576020e-05
2015-11-01  134.628980  6.419557e-02  7.439105e+00
2015-12-01  134.221442 -3.027116e-03  7.225068e-04
2016-01-01  130.605078 -2.694327e-02  7.210915e-06
2016-02-01  125.510646 -3.900638e-02  8.313595e-14
2016-03-01  128.543437  2.416361e-02  2.209344e-09
```

Out[139]: <matplotlib.axes._subplots.AxesSubplot at 0x125219750>


Index
P value
250
200
150
100
50
0
2010
2011
2012
2013
2014
2015
2016


Growth Rate
1.5
1.0
0.5
0.0
−0.5
−1.0
2010
2011
2012
2013
2014
2015
2016

In [ ]:
```
table = ListTable()
table.append(['Neighborhood','Period','Growth Rate','P Value'])

for hood in paired.neighborhood.unique():
    d = paired[paired.neighborhood == hood]
    if len(d) < 10:
        continue

    res = sm.ols(formula="np.log(price) ~ period + address", data=d).f
it()

    #calculate index
    linked = res.params[res.params.index.str.contains('Period')]
    linked.name = "Index"
    linked[0] = 100
    growth = pd.Series(linked, copy=True)
    growth.name = "Growth Rate"
    growth[0] = 0
    for i in range(1,len(linked)):
        linked[i] = (np.exp(res.params[i]))*100
        growth[i] = linked[i]/linked[i-1] - 1

    # add P values of each prediction
    p = res.pvalues[res.params.index.str.contains('Period')] * 100
    p.name = "P value"
    index = pd.concat([linked, growth, p], axis=1)
    index.index = pd.to_datetime(index.index.str.split("'").str.get(1)
)

    last = index.tail(1)
    table.append([hood
                  ,last.index[0]
                  ,round(last.iloc[0]['Growth Rate'] * 100,2)
                  ,round(last.iloc[0]['P value'], 2)])

    index[['Index','P value']].plot(title=hood)
    index[['Growth Rate']].plot()

table
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: