

```
In [88]: %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 [89]: %sql mysql://prod:nerd@52.2.153.189/rental_nerd
```

```
Out[89]: u'Connected: prod@rental_nerd'
```

```
In [90]: 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 \
FROM \
properties, \
property_transaction_logs, \
property_neighborhoods, \
neighborhoods \
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)

data = result.DataFrame()
```

7156 rows affected.

```
In [91]: result.csv(filename=strftime("%Y%m%d")+ " rentals.csv")
```

```
Out[91]: CSV results (./files/20160306 rentals.csv)
```

```
In [92]: # fill NaN values with some reasonable defaults
data.year_built = data.year_built.fillna(1970)
data.head()
```

Out[92]:

	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.com N-91st...
1	2543	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	2	2	1279	zillow_ph	http://www.zillow.com N-94th...
2	2600	7979 E Princess Dr, Scottsdale, AZ 85255	3	3	2253	zillow_ph	http://www.zillow.com E-Princ...
3	2684	9990 N Scottsdale Rd APT 1005, Scottsdale, AZ ...	2	2	973	zillow_ph	http://www.zillow.com N-Scott...
4	2719	8027 E Del Timbre Dr, Scottsdale, AZ 85258	3	3	2778	zillow_ph	http://www.zillow.com E-Del-T...

5 rows × 25 columns

```
In [93]: 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[93]:

	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 [94]: # 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 [95]: # create year dummy variables (because date isn't very intuitive varia
ble)
data["year"] = pd.DatetimeIndex(data["date"]).to_period('A')
```

In [96]: `data.describe()` *#identify filtering thresholds to clean up bad data*

Out[96]:

	property_id	bedrooms	bathrooms	sqft	longitude	latitude
count	7156.000000	7156.000000	7156.000000	7156.000000	7156.000000	7156.000000
mean	16437.613751	3.066657	2.067077	1787.584824	-112.030552	33.520160
std	8737.250668	0.959356	0.673321	847.819737	0.146166	0.136653
min	660.000000	0.000000	0.000000	0.000000	-112.300000	32.227100
25%	6876.000000	3.000000	2.000000	1271.750000	-112.138000	33.416300
50%	20130.500000	3.000000	2.000000	1620.000000	-112.045000	33.494400
75%	23389.250000	4.000000	2.000000	2110.250000	-111.924000	33.621000
max	29592.000000	8.000000	9.000000	17161.000000	-110.839000	33.882100

In [121]: *# filter out any outliers, defined as rent >\$10k or >2,500 sq ft, or not in SF*

```

print "Entries before filter: " + `len(data)`
data = data[ (data.sqft <= 10000)
              & (data.price <= 10000)
              & (data.price != 0)
              & (data.neighborhood == 'North 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: 7002

Entries after filter: 592

Out[121]:

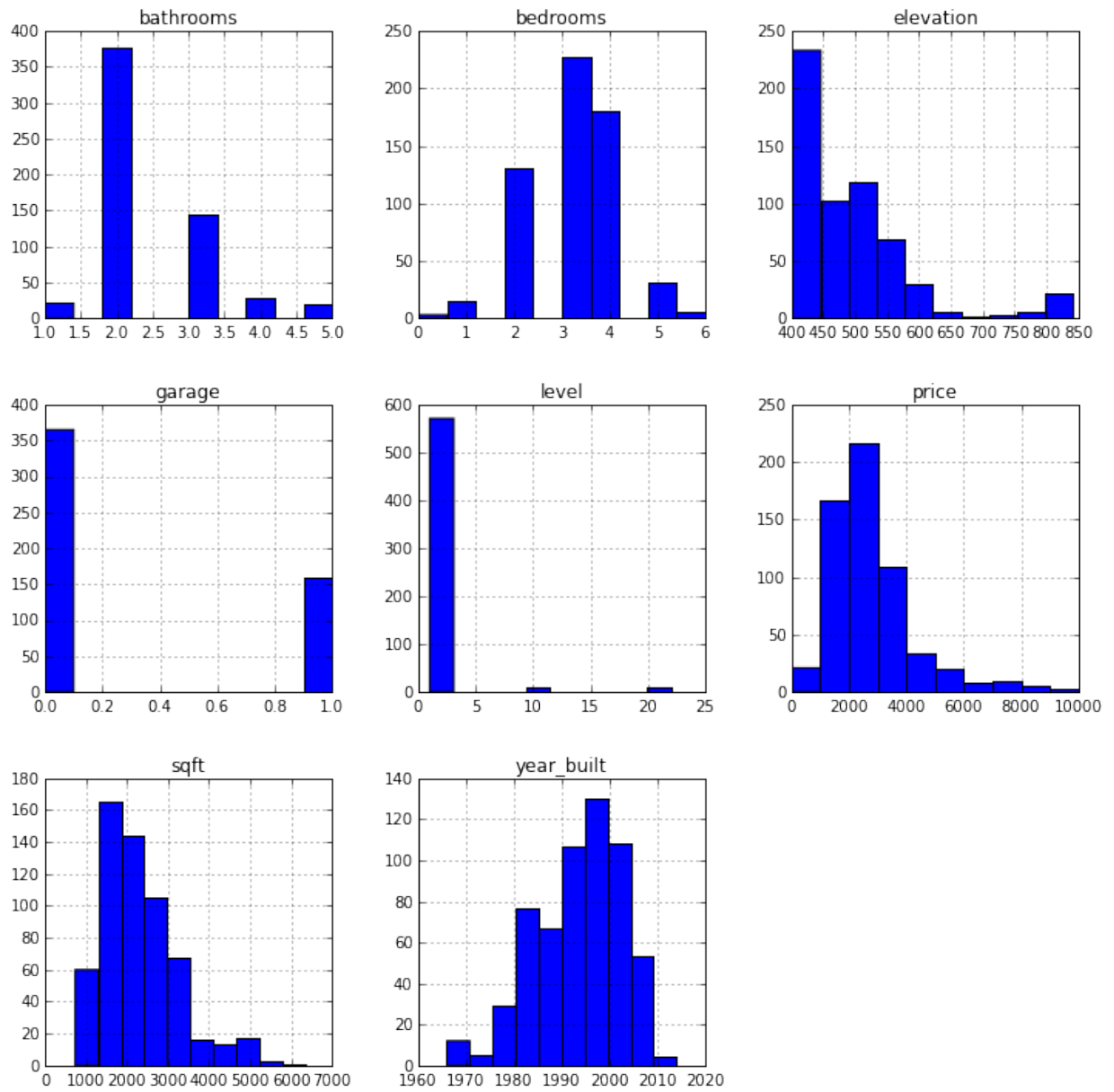
	property_id	address	bedrooms	bathrooms	sqft	source	origin_url
0	2514	11515 N 91st St UNIT 140, Scottsdale, AZ 85260	2	2	1270	zillow_ph	http://www.zillow.cor N-91st...
1	2543	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	2	2	1279	zillow_ph	http://www.zillow.cor N-94th...
2	2600	7979 E Princess Dr, Scottsdale, 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 × 32 columns

In []:

In [122]: `data.hist(column=['bathrooms','bedrooms','price','garage','level','year_built','sqft','elevation','luxurious','dist_to_park'],figsize=(12,12))`

```
Out[122]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x114beeb10
>,
    <matplotlib.axes._subplots.AxesSubplot object at 0x1289ba250
>,
    <matplotlib.axes._subplots.AxesSubplot object at 0x1289e05d0
>],
    [<matplotlib.axes._subplots.AxesSubplot object at 0x12543f890
>,
    <matplotlib.axes._subplots.AxesSubplot object at 0x1254dcf90
>,
    <matplotlib.axes._subplots.AxesSubplot object at 0x125547950
>],
    [<matplotlib.axes._subplots.AxesSubplot object at 0x1255dbbd0
>,
    <matplotlib.axes._subplots.AxesSubplot object at 0x128a15cd0
>,
    <matplotlib.axes._subplots.AxesSubplot object at 0x1256e6250
>]], dtype=object)
```

```
In [123]: 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 [124]: age = 2016 - data.year_built
age.name = 'age'
data = pd.concat([data,age],axis=1)
```

```
In [125]: result = sm.ols(formula="price ~ bedrooms + bathrooms + age + elevatio
n +\
neighborhood:sqft:year -1", data=data).fit()
result.summary()
```

Out[125]: OLS Regression Results

Dep. Variable:	price	R-squared:	0.507
Model:	OLS	Adj. R-squared:	0.496
Method:	Least Squares	F-statistic:	45.75
Date:	Sun, 06 Mar 2016	Prob (F-statistic):	4.83e-80
Time:	11:57:21	Log-Likelihood:	-4962.4
No. Observations:	592	AIC:	9953.
Df Residuals:	578	BIC:	1.001e+04
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[95.0% Conf.
--	------	---------	---	------	-----------------

					Int.]
Intercept	1128.6025	365.847	3.085	0.002	410.050 1847.155
bedrooms	-242.9038	66.346	-3.661	0.000	-373.213 -112.595
bathrooms	-6.7934	96.036	-0.071	0.944	-195.416 181.829
age[0]	-4.7485	2.831	-1.677	0.094	-10.309 0.812
age[1]	-4.7485	2.831	-1.677	0.094	-10.309 0.812
elevation	-0.5265	0.548	-0.960	0.337	-1.604 0.551
neighborhood[North Scottsdale]:sqft:year[Period('2008', 'A- DEC')]	0.8430	0.676	1.246	0.213	-0.486 2.172
neighborhood[North Scottsdale]:sqft:year[Period('2009', 'A- DEC')]	2.6940	0.292	9.239	0.000	2.121 3.267
neighborhood[North Scottsdale]:sqft:year[Period('2010', 'A- DEC')]	1.2872	0.146	8.811	0.000	1.000 1.574
neighborhood[North Scottsdale]:sqft:year[Period('2011', 'A- DEC')]	1.1963	0.123	9.747	0.000	0.955 1.437
neighborhood[North Scottsdale]:sqft:year[Period('2012', 'A- DEC')]	1.1004	0.124	8.866	0.000	0.857 1.344
neighborhood[North Scottsdale]:sqft:year[Period('2013', 'A- DEC')]	1.1686	0.103	11.347	0.000	0.966 1.371
neighborhood[North Scottsdale]:sqft:year[Period('2014', 'A- DEC')]	1.3154	0.105	12.582	0.000	1.110 1.521
neighborhood[North Scottsdale]:sqft:year[Period('2015', 'A- DEC')]	1.2731	0.097	13.135	0.000	1.083

DEC']]					1.464
neighborhood[North Scottsdale]:sqft:year[Period('2016', 'A- DEC']]	1.3683	0.098	13.941	0.000	1.176 1.561

Omnibus:	289.453	Durbin-Watson:	1.429
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1558.080
Skew:	2.178	Prob(JB):	0.00
Kurtosis:	9.648	Cond. No.	4.38e+17

```
In [126]: 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[126]:
```

neighborhood	rent per foot
North Scottsdale	1.2731369353326916

```
In [127]: 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='|', quoting=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')
```

```
-----
-----
AttributeError                                Traceback (most recent call last)
<ipython-input-127-a7d450aaa7b8> in <module>()
      15     modelwriter.writerow(['elevation', result.params.elevation])
      16     modelwriter.writerow(['level', 0])
--> 17     modelwriter.writerow(['age', result.params.age])
      18     modelwriter.writerow(['garage', 0])
      19     modelwriter.writerow(['mean square error of residuals', result.mse_resid])

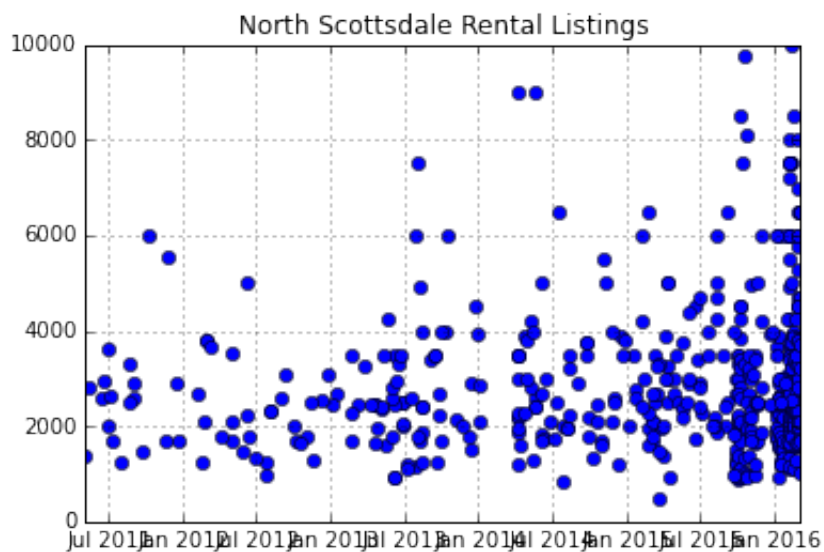
/usr/local/lib/python2.7/site-packages/pandas/core/generic.pyc in __getattr__(self, name)
    2244         return self[name]
    2245         raise AttributeError("%s' object has no attribute '%s'" %
-> 2246                             (type(self).__name__, name))
    2247
    2248     def __setattr__(self, name, value):
```

```
AttributeError: 'Series' object has no attribute 'age'
```

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

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

```
In [128]: hood = "North Scottsdale"
subdata = data[(data.neighborhood == hood) & (data.date > datetime.date(2011, 5, 1))]
plt.plot_date(x=subdata.date, y=subdata.price)
plt.title(hood + " Rental Listings")
plt.grid(True)
```



```
In [129]: 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)

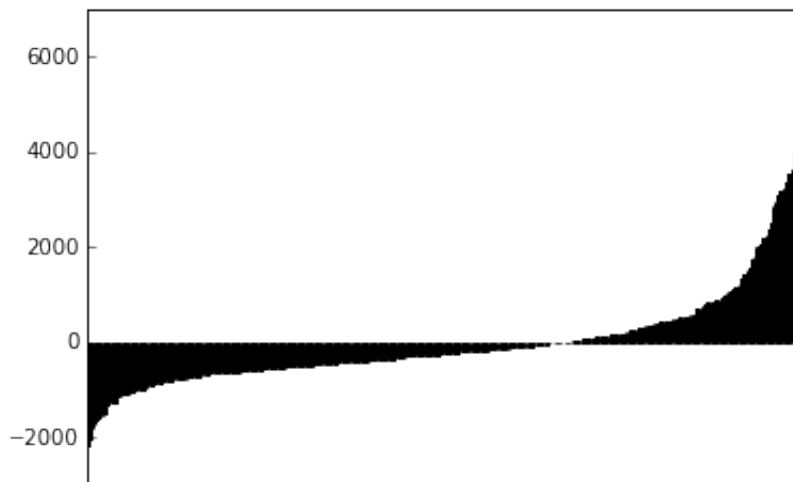
# visualize the relationship between the features and the response using scatterplots
errors.sort_values(inplace=True)
errors.plot(kind='bar').get_xaxis().set_ticks([])

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

hooderrors = data[['neighborhood']]

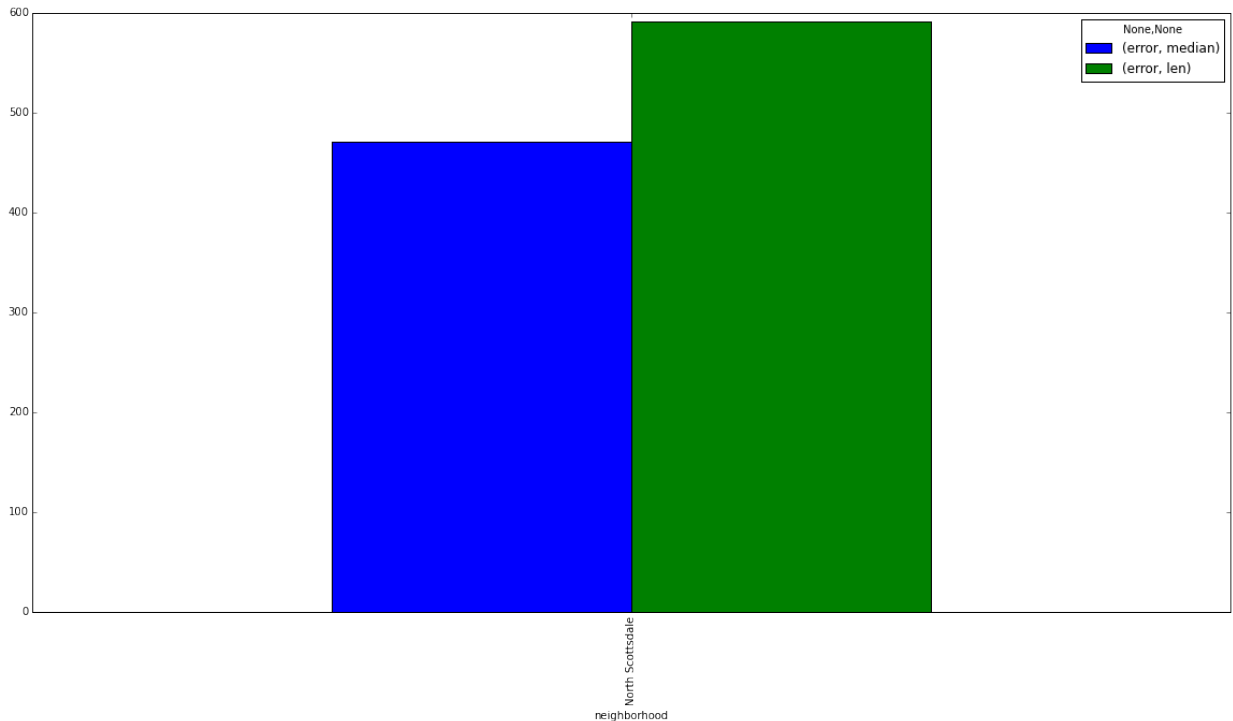
hooderrors = pd.concat([hooderrors, errors.abs()], axis=1)

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



```
In [130]: 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[130]: <matplotlib.axes._subplots.AxesSubplot at 0x1259ba710>



In []:

```
In [131]: # 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)

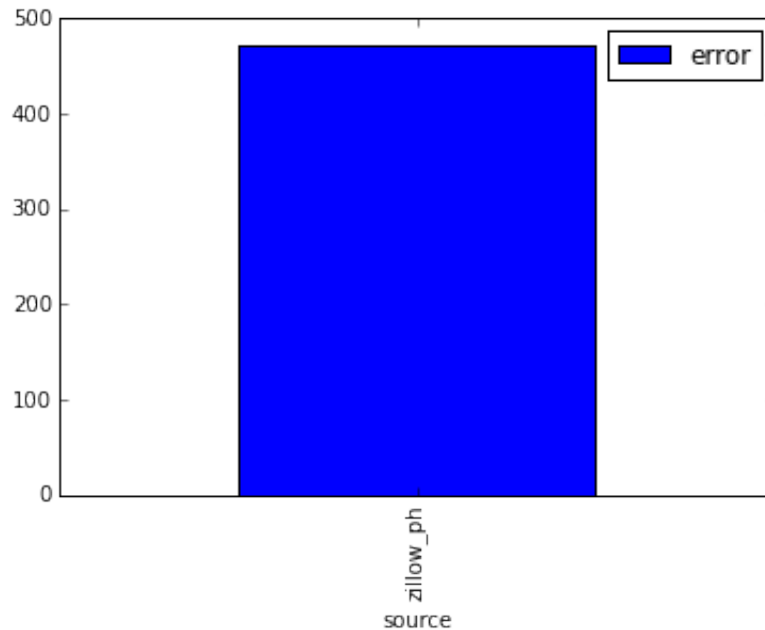
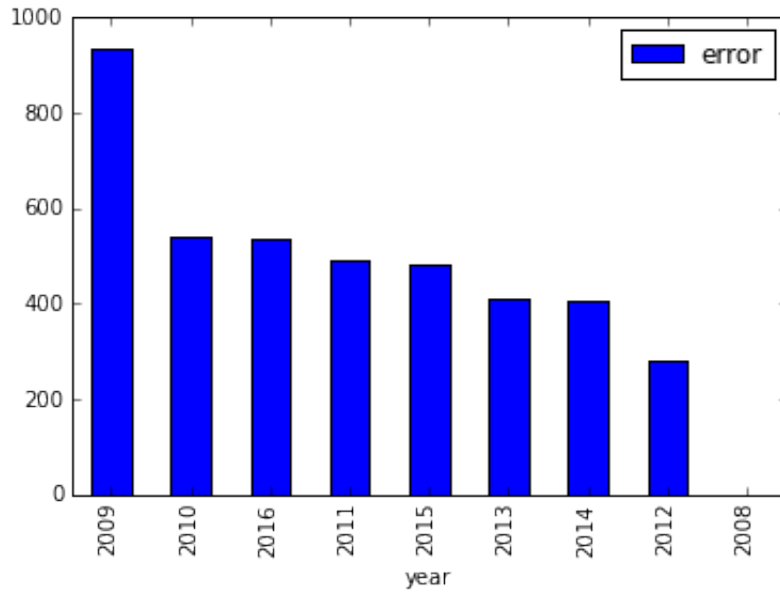
year_group = yearerrors.groupby('year')
error_avg = year_group.median()
error_avg.sort_values(by='error',ascending=False).plot(kind='bar')

# 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[131]: <matplotlib.axes._subplots.AxesSubplot at 0x1141785d0>



In [132]: data['price'].mean

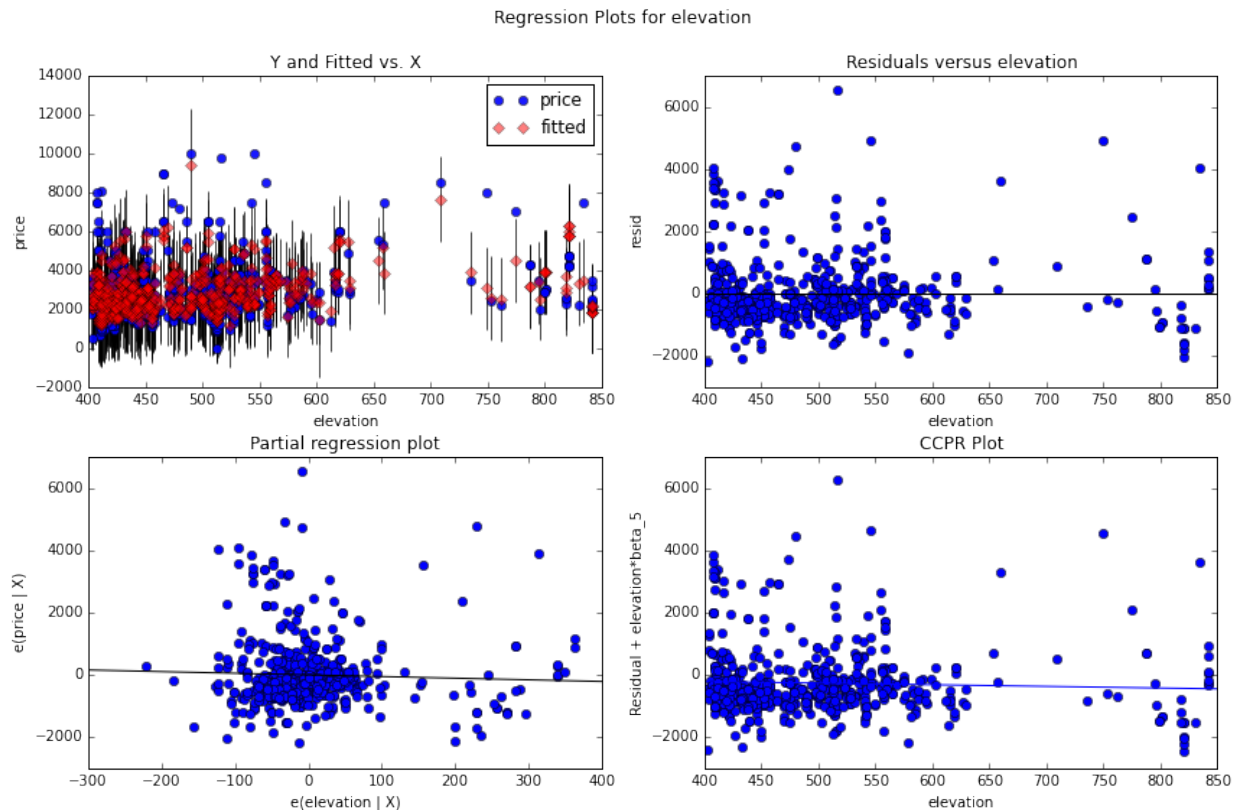
Out[132]: <bound method Series.mean of 0 1395

1	1250
2	3000
3	1195
4	2675
5	1000
6	1500
7	980
8	1150

9	1200
10	1995
11	1995
12	2100
13	1995
14	3500
15	3500
16	900
17	3000
18	1825
19	925
20	3900
22	2895
23	3300
24	1375
25	925
26	925
27	975
28	925
29	1049
30	2595
	...
576	5000
577	2495
578	2295
579	2550
580	1250
581	1300
582	2650
583	2450
584	1350
585	1475
586	4000
587	2500
588	2100
589	2999
590	3100
591	3495
592	3995
593	5000
594	1650
595	1800
596	1700
597	2500
598	2450
599	2295
600	2400
601	2400
602	2100
603	2500

```
604     4995
605     4250
Name: price, dtype: int64>
```

```
In [133]: fig = plt.figure(figsize=(12,8))
fig = sma.graphics.plot_regress_exog(result, "elevation", fig=fig)
```



```
In [134]: prstd, iv_l, iv_u = wls_prediction_std(result)
zip(data.address, data.price, data.prediction, prstd, iv_l, iv_u)
```

```
Out[134]: [('11515 N 91st St UNIT 140, Scottsdale, AZ 85260',
1395,
'prediction',
1080.4940166401011,
-535.66277415011155,
3708.6835179850059),
('14000 N 94th St UNIT 1169, Scottsdale, AZ 85260',
1250,
'prediction',
1080.8177109590981,
-535.86265536701239,
3709.7551577219115)]
```

```
In [135]: # Artnet white paper index converted to our dataset

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

```
In [136]: data[['address', 'price', 'period', 'neighborhood']]
```

```
Out[136]:
```

	address	price	period	neighborhood
0	11515 N 91st St UNIT 140, Scottsdale, AZ 85260	1395	2015Q3	North Scottsdale
1	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	1250	2015Q3	North Scottsdale
2	7979 E Princess Dr, Scottsdale, AZ 85255	3000	2015Q3	North Scottsdale
3	9990 N Scottsdale Rd APT 1005, Scottsdale, AZ ...	1195	2015Q3	North Scottsdale
4	8027 E Del Timbre Dr, Scottsdale, AZ 85258	2675	2015Q3	North Scottsdale
5	10449 N 69th St # 127, Paradise Valley, AZ 85253	1000	2015Q3	North Scottsdale
6	E Redfield Rd Scottsdale, AZ 85260	1500	2015Q4	North Scottsdale
7	11260 N 92nd St APT 2113, Scottsdale, AZ 85260	980	2012Q3	North Scottsdale
8	11260 N 92nd St APT 2113, Scottsdale, AZ 85260	1150	2013Q3	North Scottsdale
9	11260 N 92nd St APT 2113, Scottsdale, AZ 85260	1200	2014Q2	North Scottsdale
10	11675 N 91st Ln, Scottsdale, AZ 85260	1995	2010Q4	North Scottsdale
11	11675 N 91st Ln, Scottsdale, AZ 85260	1995	2015Q4	North Scottsdale
12	11675 N 91st Ln, Scottsdale, AZ 85260	2100	2015Q4	North Scottsdale

13	11675 N 91st Ln, Scottsdale, AZ 85260	1995	2010Q4	North Scottsdale
14	18171 N 93rd St, Scottsdale, AZ 85255	3500	2015Q4	North Scottsdale
15	18171 N 93rd St, Scottsdale, AZ 85255	3500	2015Q4	North Scottsdale
16	20100 N 78th PI APT 2093, Scottsdale, AZ 85255	900	2015Q4	North Scottsdale
17	9551 E Redfield Rd UNIT 1020, Scottsdale, AZ 8...	3000	2015Q4	North Scottsdale
18	6940 E Cochise Rd UNIT 1044, Scottsdale, AZ 85253	1825	2015Q4	North Scottsdale
19	N 78th PI Scottsdale, AZ 85255	925	2015Q4	North Scottsdale
20	9270 E Thompson Peak Pkwy UNIT 301, Scottsdale...	3900	2015Q2	North Scottsdale
22	9270 E Thompson Peak Pkwy UNIT 301, Scottsdale...	2895	2015Q3	North Scottsdale
23	9270 E Thompson Peak Pkwy UNIT 301, Scottsdale...	3300	2015Q4	North Scottsdale
24	11515 N 91st St UNIT 115, Scottsdale, AZ 85260	1375	2015Q4	North Scottsdale
25	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	925	2015Q4	North Scottsdale
26	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	925	2013Q2	North Scottsdale
27	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	975	2015Q4	North Scottsdale
28	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	925	2013Q2	North Scottsdale
29	8787 E Mountain View Rd APT 1064, Scottsdale, ...	1049	2015Q4	North Scottsdale
30	8413 E Belgian Trl, Scottsdale, AZ 85258	2595	2010Q4	North Scottsdale
...

576	11291 E Caribbean Ln, Scottsdale, AZ 85255	5000	2014Q4	North Scottsdale
577	11408 E Running Deer Trl, Scottsdale, AZ 85262	2495	2012Q4	North Scottsdale
578	11408 E Running Deer Trl, Scottsdale, AZ 85262	2295	2013Q1	North Scottsdale
579	11408 E Running Deer Trl, Scottsdale, AZ 85262	2550	2016Q1	North Scottsdale
580	11762 E Becker Ln, Scottsdale, AZ 85259	1250	2012Q1	North Scottsdale
581	11762 E Becker Ln, Scottsdale, AZ 85259	1300	2014Q2	North Scottsdale
582	16526 N 105th St, Scottsdale, AZ 85255	2650	2011Q3	North Scottsdale
583	16526 N 105th St, Scottsdale, AZ 85255	2450	2014Q4	North Scottsdale
584	19475 N Grayhawk Dr UNIT 2112, Scottsdale, AZ ...	1350	2012Q2	North Scottsdale
585	19475 N Grayhawk Dr UNIT 2112, Scottsdale, AZ ...	1475	2015Q1	North Scottsdale
586	7873 E Softwind Dr, Scottsdale, AZ 85255	4000	2014Q2	North Scottsdale
587	9072 E Friess Dr, Scottsdale, AZ 85260	2500	2015Q2	North Scottsdale
588	9127 E Poinsettia Dr, Scottsdale, AZ 85260	2100	2014Q1	North Scottsdale
589	9828 E Bahia Dr, Scottsdale, AZ 85260	2999	2015Q1	North Scottsdale
590	9828 E Bahia Dr, Scottsdale, AZ 85260	3100	2015Q4	North Scottsdale
591	10290 N 117th Pl, Scottsdale, AZ 85259	3495	2013Q2	North Scottsdale
592	10290 N 117th Pl, Scottsdale, AZ 85259	3995	2013Q3	North Scottsdale

593	10290 N 117th Pl, Scottsdale, AZ 85259	5000	2015Q2	North Scottsdale
594	11614 N 88th Pl, Scottsdale, AZ 85260	1650	2012Q4	North Scottsdale
595	20121 N 76th St UNIT 2031, Scottsdale, AZ 85255	1800	2014Q4	North Scottsdale
596	24106 N 72nd Pl, Scottsdale, AZ 85255	1700	2013Q1	North Scottsdale
597	26800 N 57th St, Scottsdale, AZ 85266	2500	2011Q3	North Scottsdale
598	8932 N 86th St, Scottsdale, AZ 85258	2450	2013Q2	North Scottsdale
599	8932 N 86th St, Scottsdale, AZ 85258	2295	2014Q2	North Scottsdale
600	8932 N 86th St, Scottsdale, AZ 85258	2400	2015Q1	North Scottsdale
601	8932 N 86th St, Scottsdale, AZ 85258	2400	2015Q3	North Scottsdale
602	9063 E Maple Dr, Scottsdale, AZ 85255	2100	2014Q2	North Scottsdale
603	9063 E Maple Dr, Scottsdale, AZ 85255	2500	2016Q1	North Scottsdale
604	9621 E Gary Rd, Scottsdale, AZ 85260	4995	2011Q1	North Scottsdale
605	9621 E Gary Rd, Scottsdale, AZ 85260	4250	2013Q2	North Scottsdale

592 rows × 4 columns

```
In [137]: 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.period.max()` + " and number of periods: " + `num_periods`

paired.head()
```

```
base period: Period('2008Q3', 'Q-DEC') end period: Period('2016Q1', 'Q-DEC') and number of periods: 30
```

Out[137]:

	property_id	address	price	period	neighborhood
0	2514	11515 N 91st St UNIT 140, Scottsdale, AZ 85260	1395	2015Q3	North Scottsdale
1	2543	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	1250	2015Q3	North Scottsdale
2	2600	7979 E Princess Dr, Scottsdale, AZ 85255	3000	2015Q3	North Scottsdale
3	2684	9990 N Scottsdale Rd APT 1005, Scottsdale, AZ ...	1195	2015Q3	North Scottsdale
4	2719	8027 E Del Timbre Dr, Scottsdale, AZ 85258	2675	2015Q3	North Scottsdale

In [115]: paired

Out[115]:

	property_id	address	price	period	neighborhood
0	2514	11515 N 91st St UNIT 140, Scottsdale, AZ 85260	1395	2015Q3	North Scottsdale
1	2543	14000 N 94th St UNIT 1169, Scottsdale, AZ 85260	1250	2015Q3	North Scottsdale
2	2600	7979 E Princess Dr, Scottsdale, AZ 85255	3000	2015Q3	North Scottsdale
3	2684	9990 N Scottsdale Rd APT 1005, Scottsdale, AZ ...	1195	2015Q3	North Scottsdale
4	2719	8027 E Del Timbre Dr, Scottsdale, AZ 85258	2675	2015Q3	North Scottsdale

5	2758	10449 N 69th St # 127, Paradise Valley, AZ 85253	1000	2015Q3	North Scottsdale
6	2888	E Redfield Rd Scottsdale, AZ 85260	1500	2015Q4	North Scottsdale
7	2927	11260 N 92nd St APT 2113, Scottsdale, AZ 85260	980	2012Q3	North Scottsdale
8	2927	11260 N 92nd St APT 2113, Scottsdale, AZ 85260	1150	2013Q3	North Scottsdale
9	2927	11260 N 92nd St APT 2113, Scottsdale, AZ 85260	1200	2014Q2	North Scottsdale
10	2945	11675 N 91st Ln, Scottsdale, AZ 85260	1995	2010Q4	North Scottsdale
11	2945	11675 N 91st Ln, Scottsdale, AZ 85260	1995	2015Q4	North Scottsdale
12	2945	11675 N 91st Ln, Scottsdale, AZ 85260	2100	2015Q4	North Scottsdale
13	2945	11675 N 91st Ln, Scottsdale, AZ 85260	1995	2010Q4	North Scottsdale
14	2955	18171 N 93rd St, Scottsdale, AZ 85255	3500	2015Q4	North Scottsdale
15	2955	18171 N 93rd St, Scottsdale, AZ 85255	3500	2015Q4	North Scottsdale
16	2986	20100 N 78th Pl APT 2093, Scottsdale, AZ 85255	900	2015Q4	North Scottsdale
17	3002	9551 E Redfield Rd UNIT 1020, Scottsdale, AZ 8...	3000	2015Q4	North Scottsdale
18	3005	6940 E Cochise Rd UNIT 1044, Scottsdale, AZ 85253	1825	2015Q4	North Scottsdale
19	3061	N 78th Pl Scottsdale, AZ 85255	925	2015Q4	North Scottsdale
20	3088	9270 E Thompson Peak Pkwy UNIT 301, Scottsdale...	3900	2015Q2	North Scottsdale
22	3088	9270 E Thompson Peak Pkwy UNIT 301, Scottsdale...	2895	2015Q3	North Scottsdale
		9270 E Thompson Peak Pkwy UNIT			North

23	3088	301, Scottsdale...	3300	2015Q4	Scottsdale
24	3090	11515 N 91st St UNIT 115, Scottsdale, AZ 85260	1375	2015Q4	North Scottsdale
25	3141	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	925	2015Q4	North Scottsdale
26	3141	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	925	2013Q2	North Scottsdale
27	3141	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	975	2015Q4	North Scottsdale
28	3141	20100 N 78th PI APT 1146, Scottsdale, AZ 85255	925	2013Q2	North Scottsdale
29	3154	8787 E Mountain View Rd APT 1064, Scottsdale, ...	1049	2015Q4	North Scottsdale
30	3169	8413 E Belgian Trl, Scottsdale, AZ 85258	2595	2010Q4	North Scottsdale
...
7124	24347	363 E Palm Ln, Phoenix, AZ 85004	1650	2012Q4	Encanto
7125	24347	363 E Palm Ln, Phoenix, AZ 85004	2500	2016Q1	Encanto
7126	24347	363 E Palm Ln, Phoenix, AZ 85004	2500	2016Q1	Encanto
7127	24348	4210 N 9th Ave, Phoenix, AZ 85013	1225	2016Q1	Encanto
7128	24695	2104 E Sheridan St, Phoenix, AZ 85006	750	2013Q1	Encanto
7129	24695	2104 E Sheridan St, Phoenix, AZ 85006	750	2013Q4	Encanto
7131	24695	2104 E Sheridan St, Phoenix, AZ 85006	850	2012Q3	Encanto
7132	24820	2925 N 21st Ave, Phoenix, AZ 85015	1200	2016Q1	Encanto
7133	24821	3009 N 16th Ave, Phoenix, AZ 85015	1700	2016Q1	Encanto
7134	25032	2546 N 10th St, Phoenix, AZ 85006	1300	2016Q1	Encanto
7135	25240	1136 W Monterosa St, Phoenix, AZ 85013	1000	2015Q1	Encanto
7136	25240	1136 W Monterosa St, Phoenix, AZ 85013	1000	2015Q2	Encanto

7137	25245	321 W Encanto Blvd, Phoenix, AZ 85003	1699	2013Q1	Encanto
7138	25577	1524 W Earll Dr, Phoenix, AZ 85015	1450	2013Q1	Encanto
7139	25866	1017 W Weldon Ave, Phoenix, AZ 85013	990	2011Q4	Encanto
7140	25871	2977 N 19th Ave UNIT 24, Phoenix, AZ 85015	495	2012Q3	Encanto
7141	25871	2977 N 19th Ave UNIT 24, Phoenix, AZ 85015	595	2015Q4	Encanto
7143	26491	4127 N 9th Ave, Phoenix, AZ 85013	775	2015Q2	Encanto
7144	26493	902 W Monterosa St, Phoenix, AZ 85013	1000	2016Q1	Encanto
7145	28013	4258 N 15th Dr, Phoenix, AZ 85015	1300	2016Q1	Encanto
7146	16520	6322 S Colonial Way, Tempe, AZ 85283	1450	2016Q1	Wood Park
7147	22900	6226 S Parkside Dr, Tempe, AZ 85283	1050	2013Q1	Wood Park
7148	22900	6226 S Parkside Dr, Tempe, AZ 85283	1200	2013Q3	Wood Park
7149	22900	6226 S Parkside Dr, Tempe, AZ 85283	1235	2015Q2	Wood Park
7150	22900	6226 S Parkside Dr, Tempe, AZ 85283	1295	2016Q1	Wood Park
7151	28780	6290 S Colonial Way, Tempe, AZ 85283	1195	2016Q1	Wood Park
7152	3548	6609 S Mckemy St, Tempe, AZ 85283	1300	2011Q3	Tempe Royal Estates
7153	3548	6609 S Mckemy St, Tempe, AZ 85283	1200	2013Q1	Tempe Royal Estates
7154	12005	6601 S Mckemy St, Tempe, AZ 85283	1550	2016Q1	Tempe Royal Estates
7155	14430	6847 S Dennis Dr, Tempe, AZ 85283	1550	2016Q1	Tempe Royal Estates

7002 rows × 5 columns

```
In [116]: len(paired.groupby("address").filter(lambda x: len(x['address']) >1).groupby('property_id'))
```

```
Out[116]: 1725
```

```
In [117]: paired.groupby("address").filter(lambda x: len(x['address']) >1).groupby('address').get_group('210 W Helena Dr, Phoenix, AZ 85023')
```

```
Out[117]:
```

	property_id	address	price	period	neighborhood
6927	29261	210 W Helena Dr, Phoenix, AZ 85023	1200	2012Q3	Deer Valley
6928	29261	210 W Helena Dr, Phoenix, AZ 85023	1250	2014Q2	Deer Valley
6929	29261	210 W Helena Dr, Phoenix, AZ 85023	1295	2015Q1	Deer Valley

```
In [118]: # group data into Sets and calc Y_list 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: 4038
```

```
In [119]: 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))

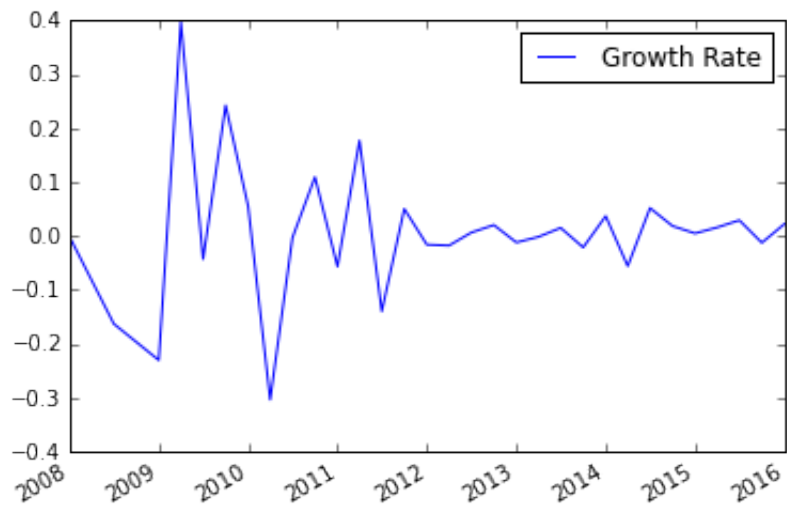
print index

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

	Index	Growth Rate	P value
2008-01-01	100.000000	0.000000	35.264181
2008-07-01	83.733217	-0.162668	2.962797
2009-01-01	64.459050	-0.230185	65.173262
2009-04-01	90.046668	0.396959	52.270681
2009-07-01	86.214997	-0.042552	70.984119
2009-10-01	107.131421	0.242608	50.868457
2010-01-01	113.063703	0.055374	16.494848
2010-04-01	78.757889	-0.303420	14.892371
2010-07-01	78.518922	-0.003034	41.267251
2010-10-01	87.131752	0.109691	24.316948
2011-01-01	82.216244	-0.056415	84.501616
2011-04-01	96.797879	0.177357	27.015211
2011-07-01	83.300391	-0.139440	41.983395
2011-10-01	87.516535	0.050614	36.726766
2012-01-01	86.142728	-0.015698	31.230239
2012-04-01	84.638598	-0.017461	33.159076
2012-07-01	85.206126	0.006705	39.695939
2012-10-01	86.965652	0.020650	35.783715
2013-01-01	85.947200	-0.011711	35.355958
2013-04-01	85.827388	-0.001394	40.393482
2013-07-01	87.147903	0.015386	33.468782
2013-10-01	85.321349	-0.020959	45.825032
2014-01-01	88.494631	0.037192	28.105096
2014-04-01	83.575152	-0.055591	43.436701
2014-07-01	87.932663	0.052139	50.194952
2014-10-01	89.583333	0.018772	52.433595
2015-01-01	90.046668	0.005172	58.920090
2015-04-01	91.497153	0.016108	71.657754
2015-07-01	94.195795	0.029494	66.124243
2015-10-01	93.038307	-0.012288	76.120263
2016-01-01	95.126904	0.022449	81.181726

Out[119]: <matplotlib.axes._subplots.AxesSubplot at 0x13acc2d10>





```

In [120]: 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).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)
    )

    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()

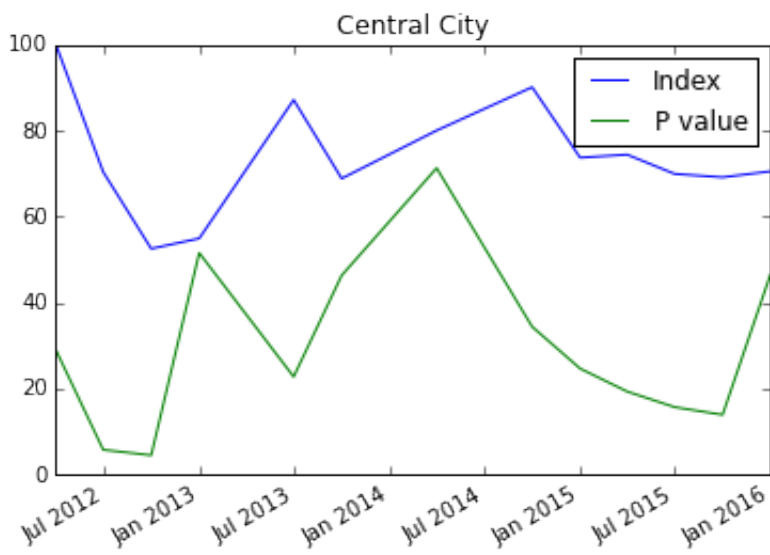
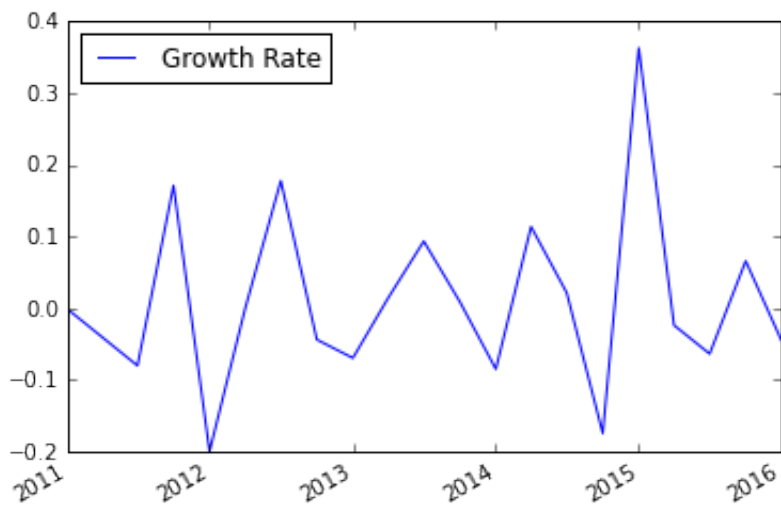
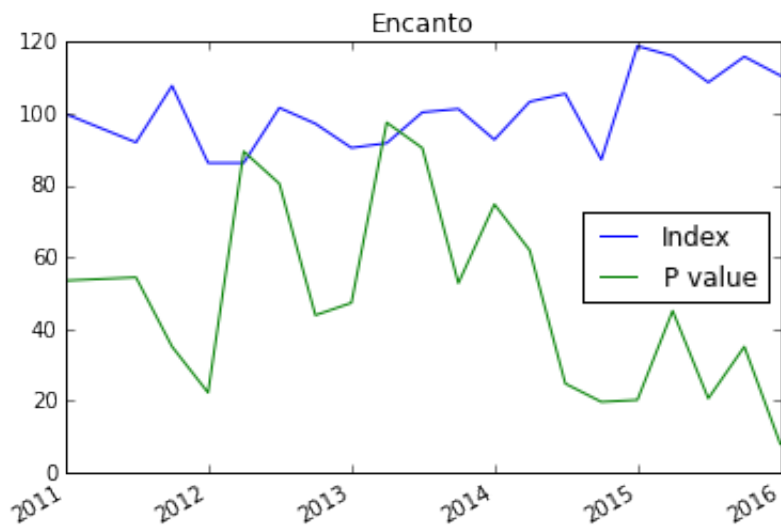
table

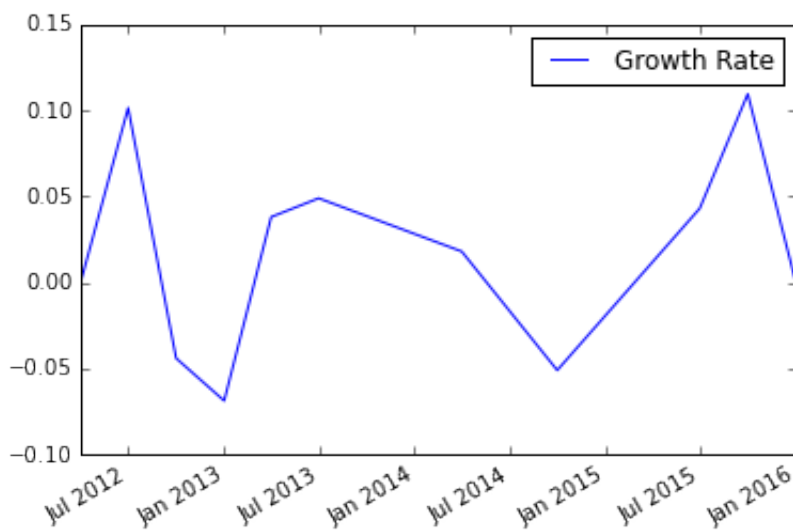
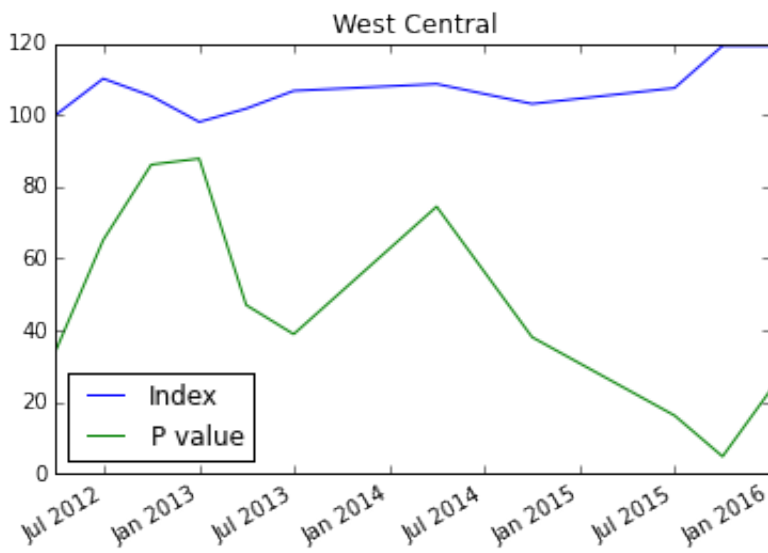
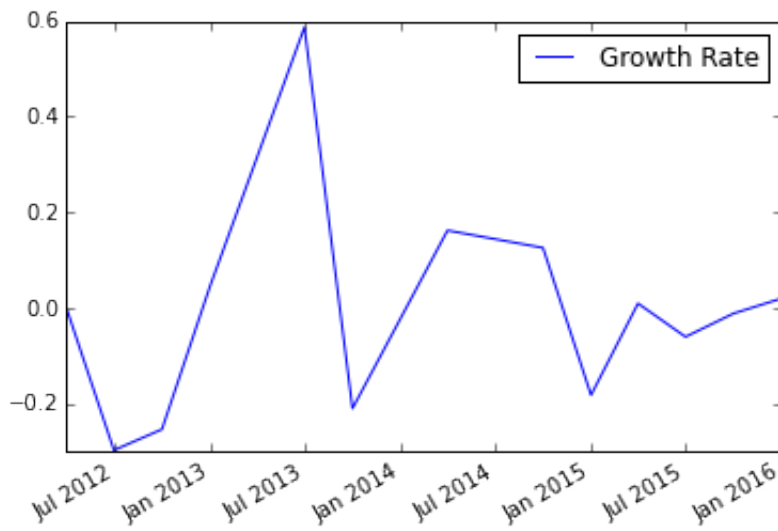
```

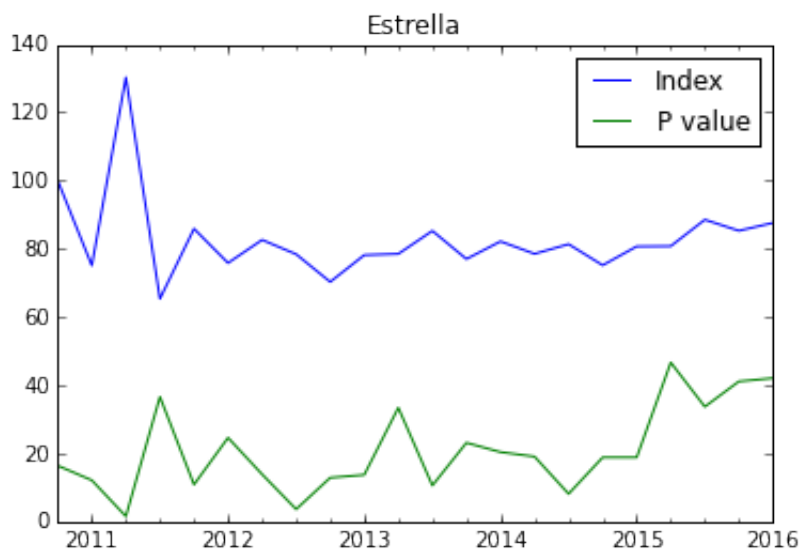
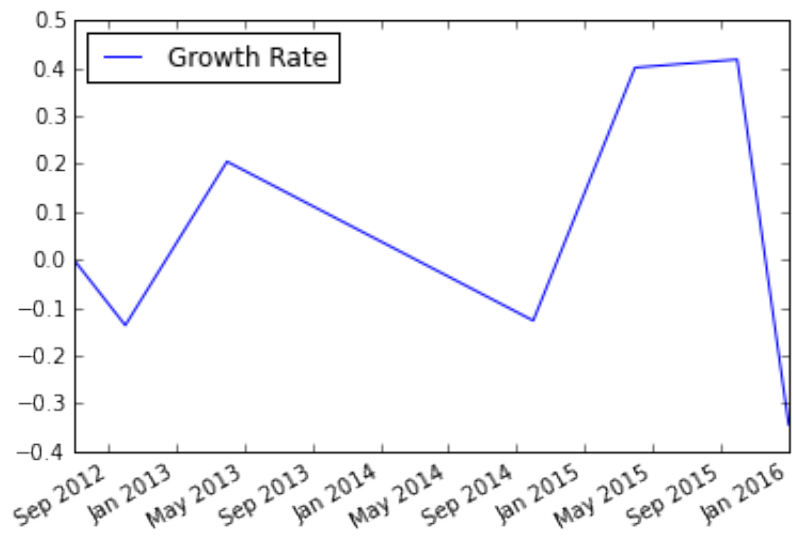
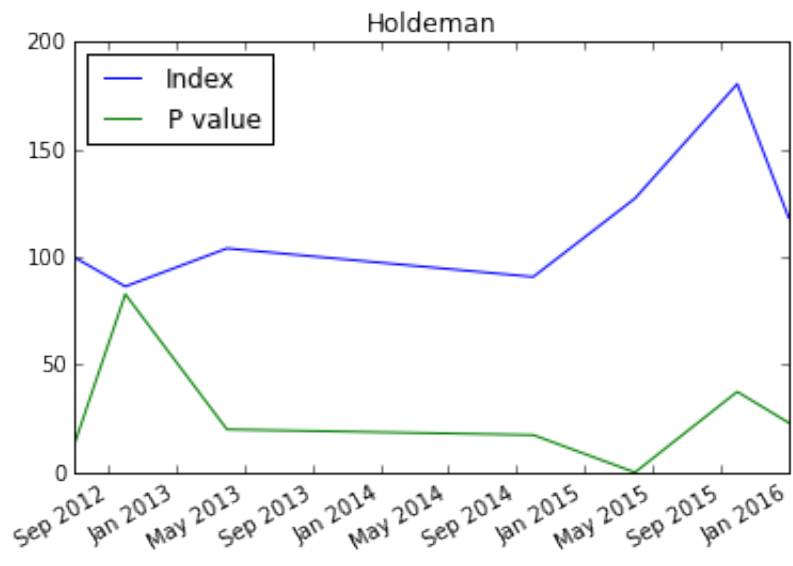
Out[120]:

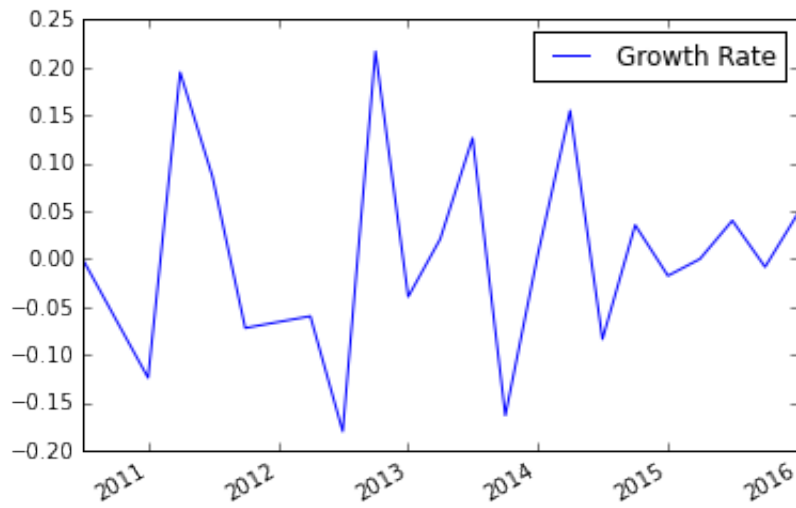
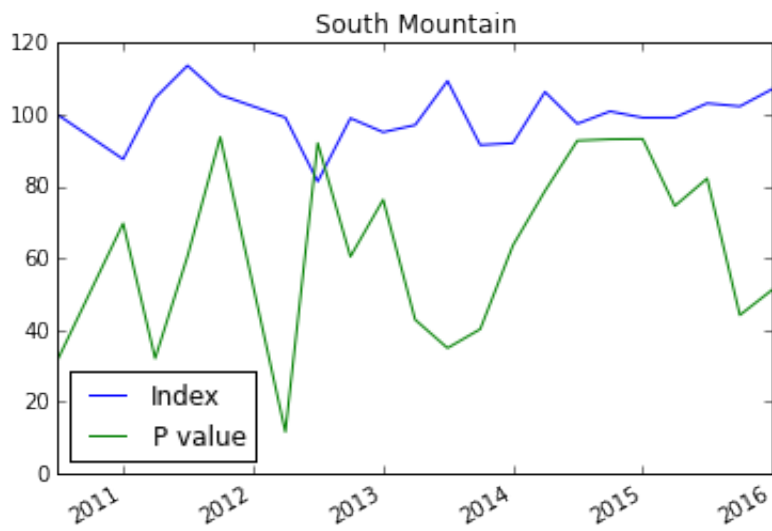
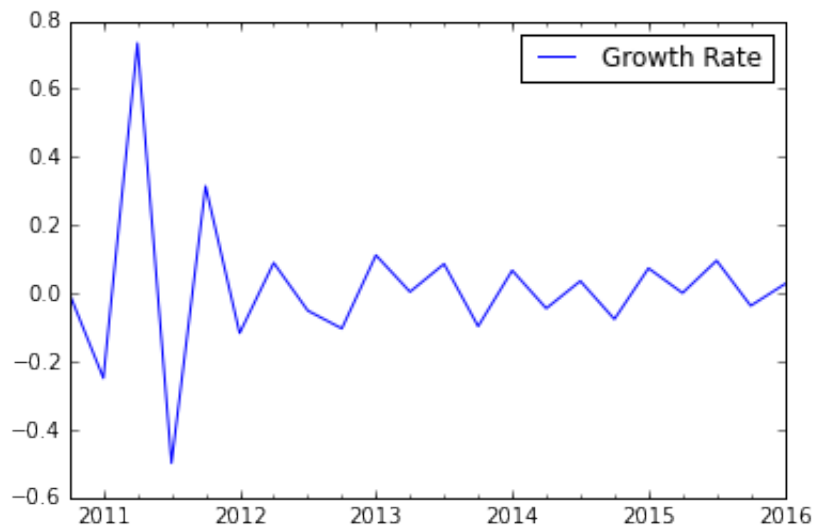
Neighborhood	Period	Growth Rate	P Value
Encanto	2016-01-01 00:00:00	-4.55	8.0
Central City	2016-01-01 00:00:00	2.01	46.6
West Central	2016-01-01 00:00:00	-0.0	23.31
Holdeman	2016-01-01 00:00:00	-34.4	23.12

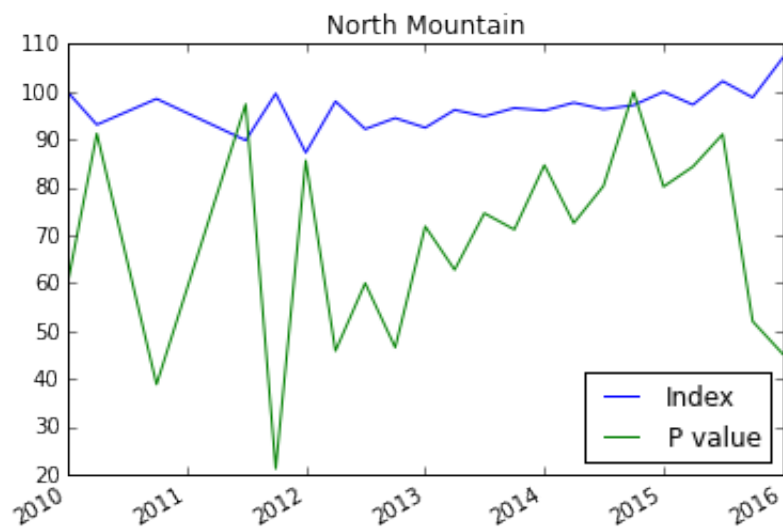
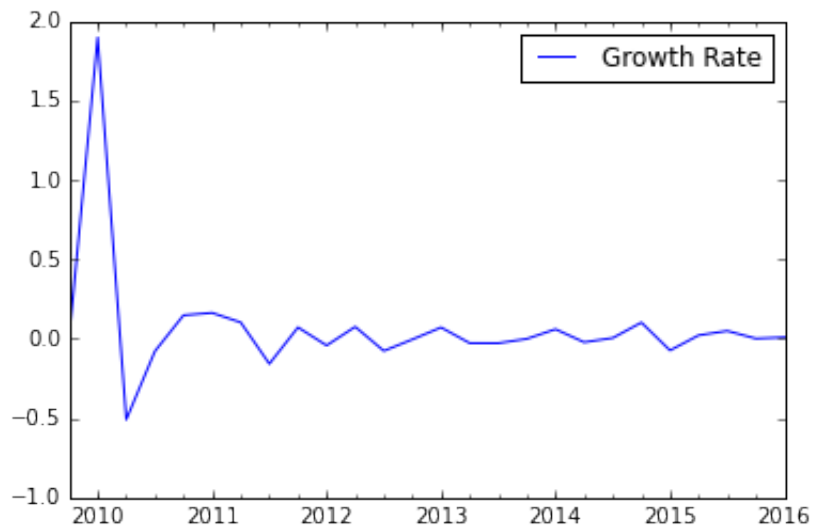
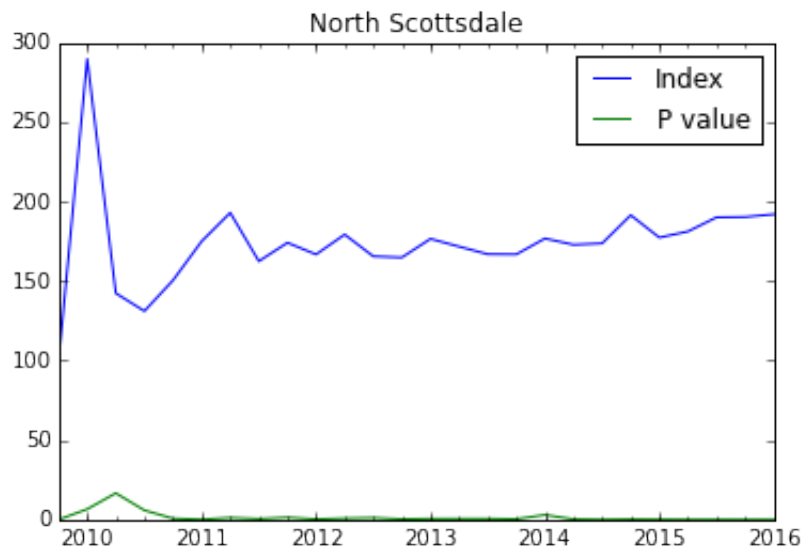
Estrella	2016-01-01 00:00:00	2.7	42.07
South Mountain	2016-01-01 00:00:00	4.74	51.31
North Scottsdale	2016-01-01 00:00:00	0.87	0.02
North Mountain	2016-01-01 00:00:00	8.38	45.26
Deer Valley	2016-01-01 00:00:00	1.07	14.56
Hughes Acres	2016-01-01 00:00:00	-7.86	8.8
Pinnacle Peak	2016-01-01 00:00:00	3.16	74.81
Pepperwood	2016-01-01 00:00:00	13.58	16.74
Gililand	2016-01-01 00:00:00	3.25	84.22
Laveen	2016-01-01 00:00:00	3.23	46.14
Southeast 0	2016-01-01 00:00:00	40.33	77.82
Maryvale	2016-01-01 00:00:00	6.79	28.83
South Scottsdale	2016-01-01 00:00:00	5.73	0.0
Northwest	2016-01-01 00:00:00	5.23	1.18
Camelback East	2016-01-01 00:00:00	5.39	42.76
Ahwatukee Foothills 0	2016-01-01 00:00:00	-1.58	73.37
Riverside	2016-01-01 00:00:00	-51.65	0.01
Sunset	2016-01-01 00:00:00	-21.87	2.38
Alhambra	2015-10-01 00:00:00	-11.59	43.65
Paradise Valley	2016-01-01 00:00:00	-8.43	5.21
Central	2015-10-01 00:00:00	3.06	18.1
Southwest	2016-01-01 00:00:00	-29.29	48.2
Northeast 0	2016-01-01 00:00:00	4.02	86.51
North Gateway	2016-01-01 00:00:00	-7.17	21.44
Desert View	2016-01-01 00:00:00	4.13	52.08
Northeast 3	2016-01-01 00:00:00	-0.0	nan

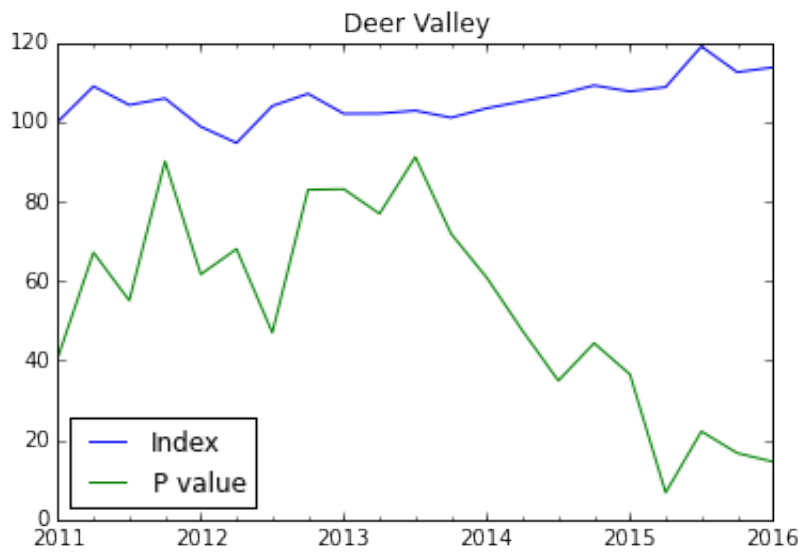
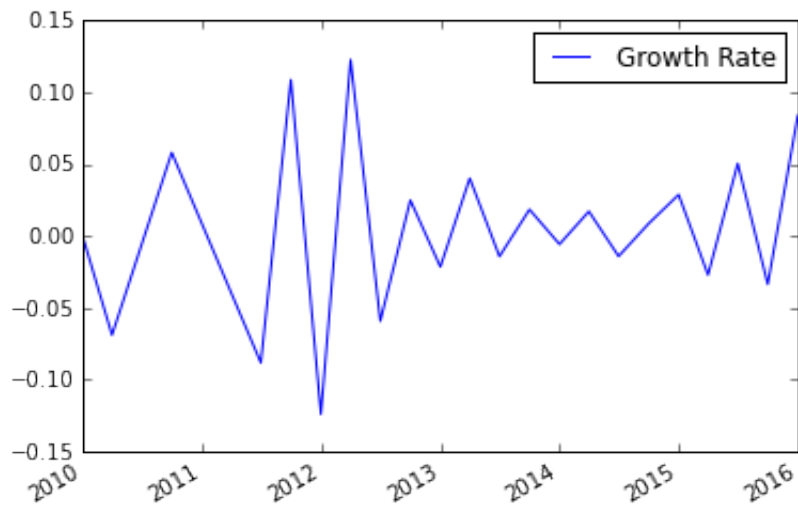


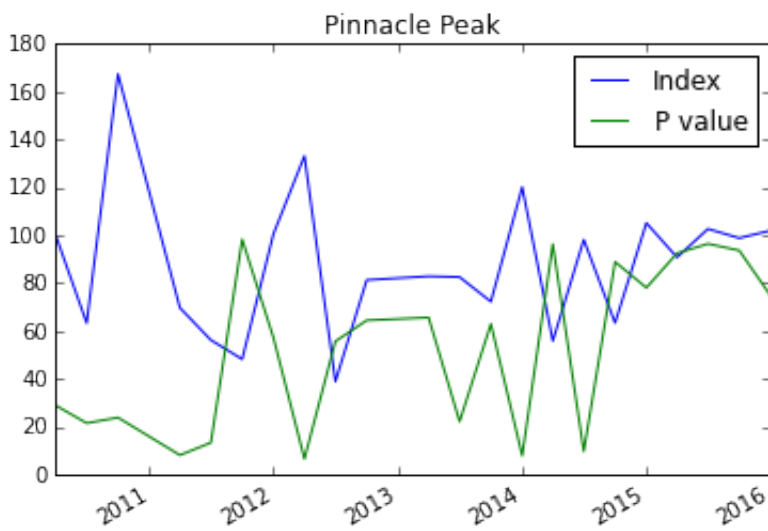
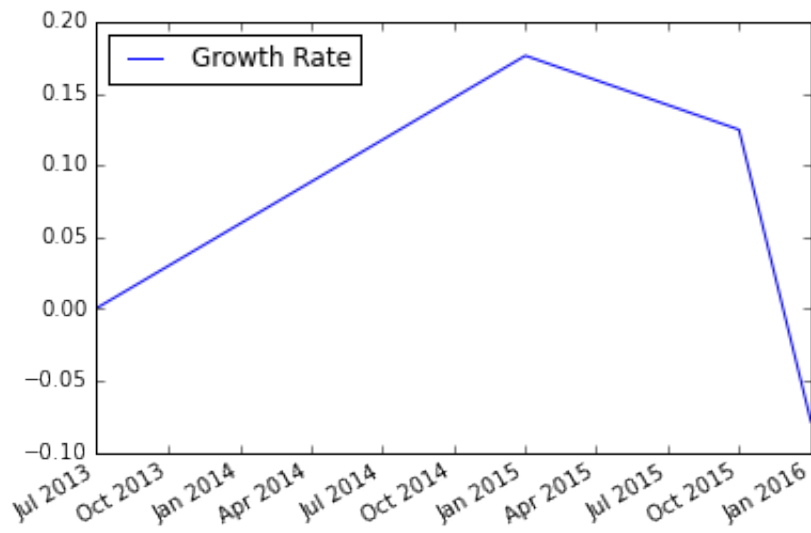
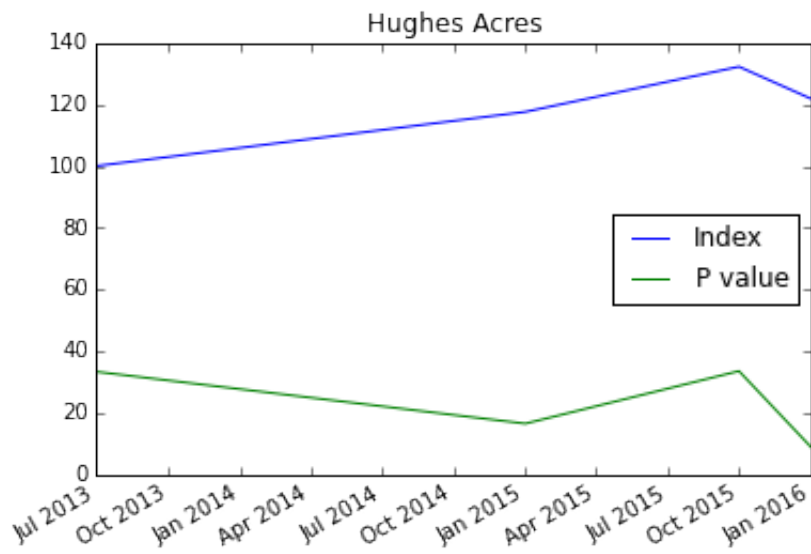


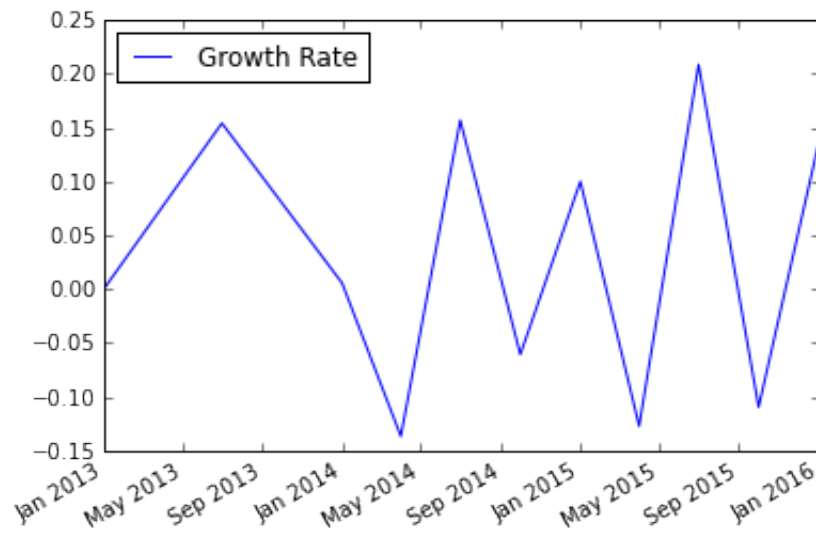
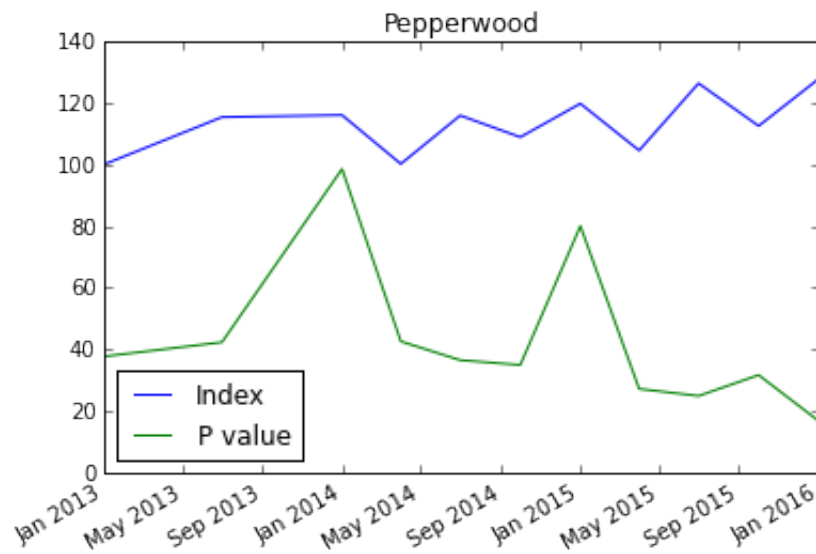
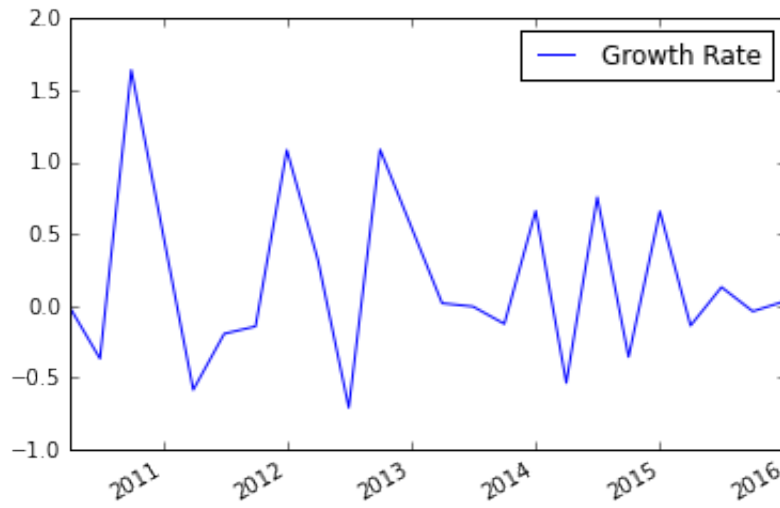


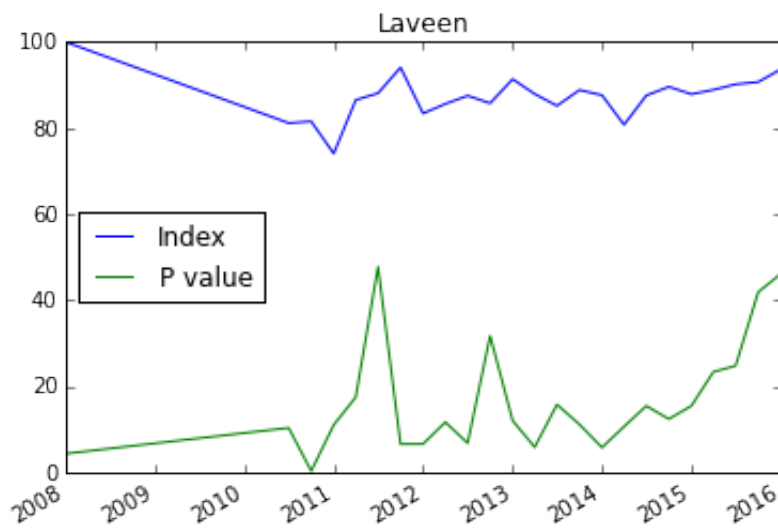
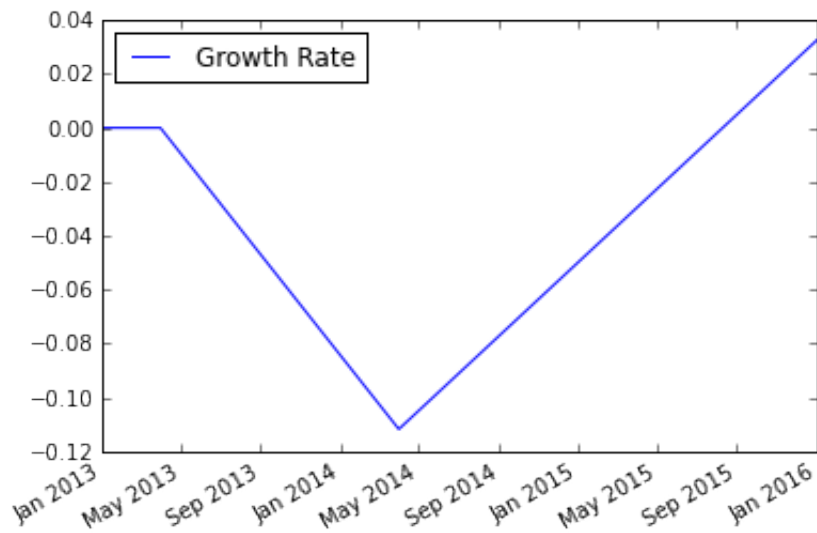
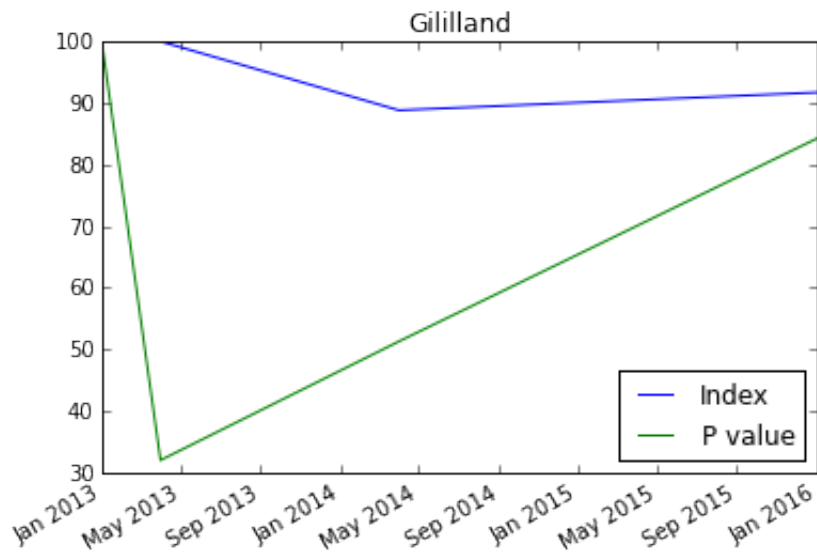


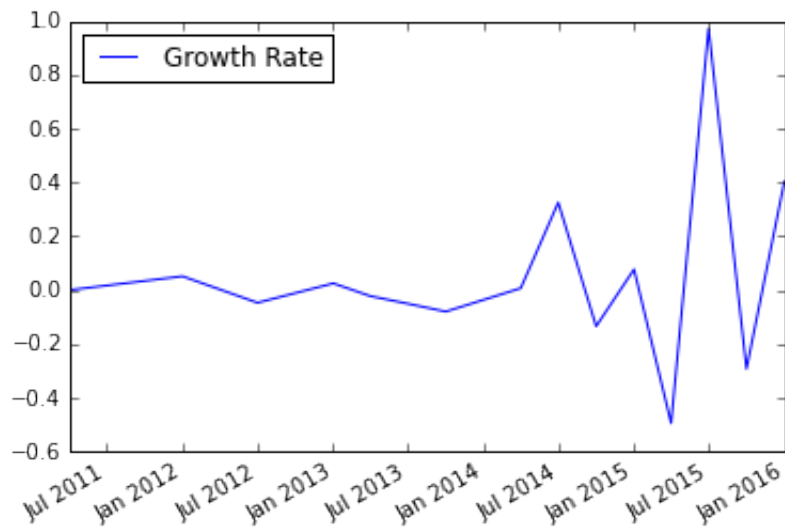
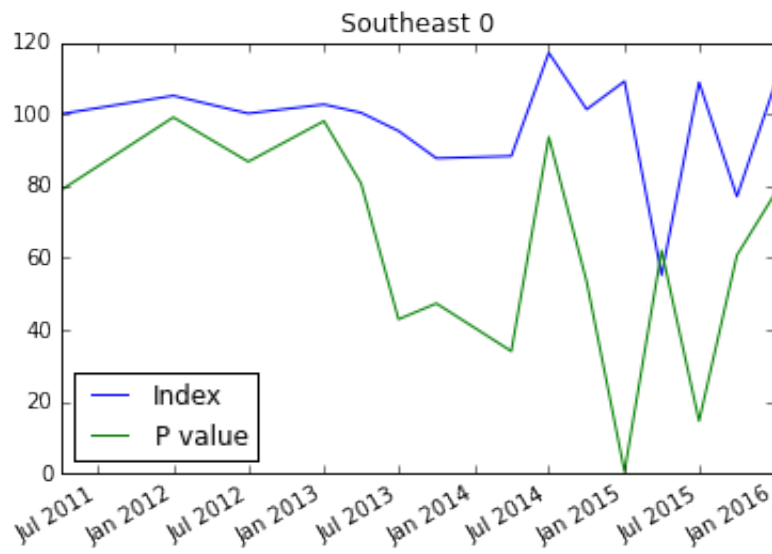
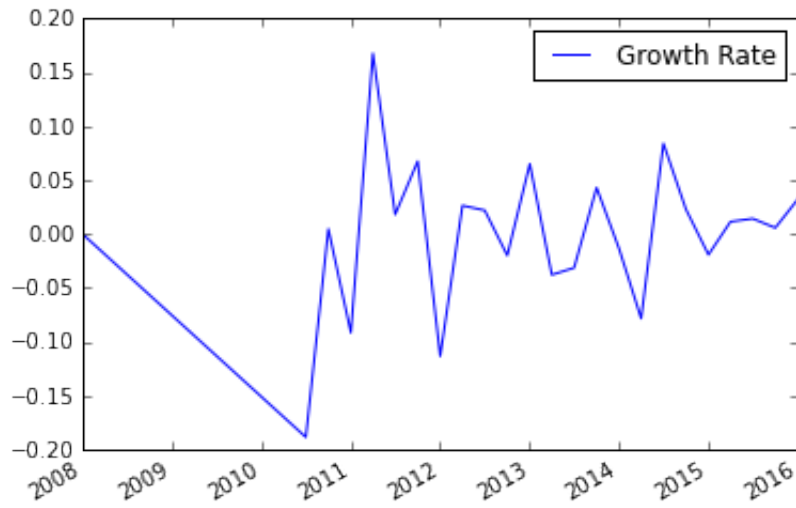


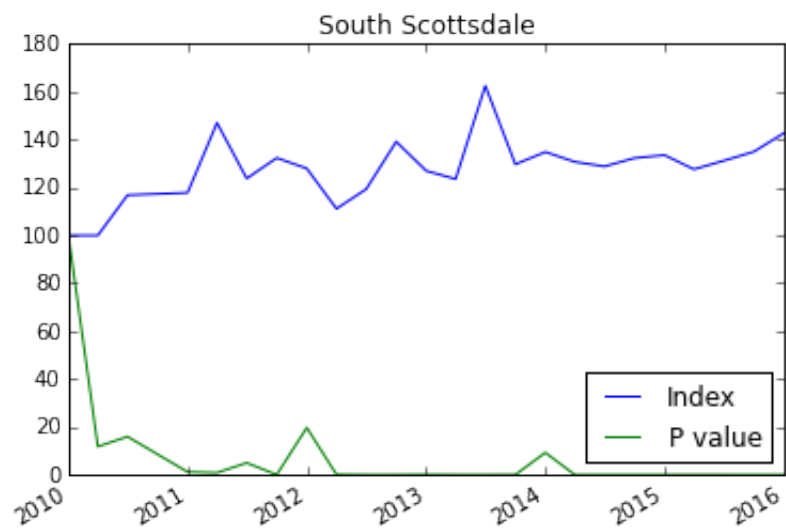
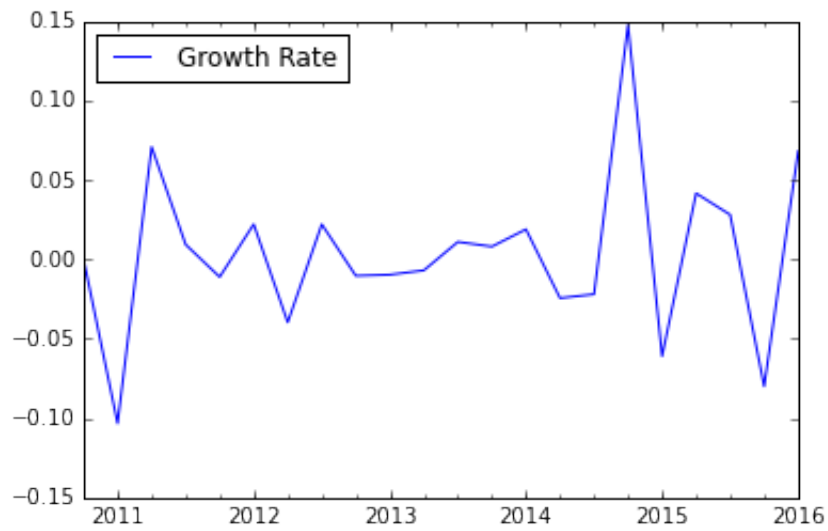
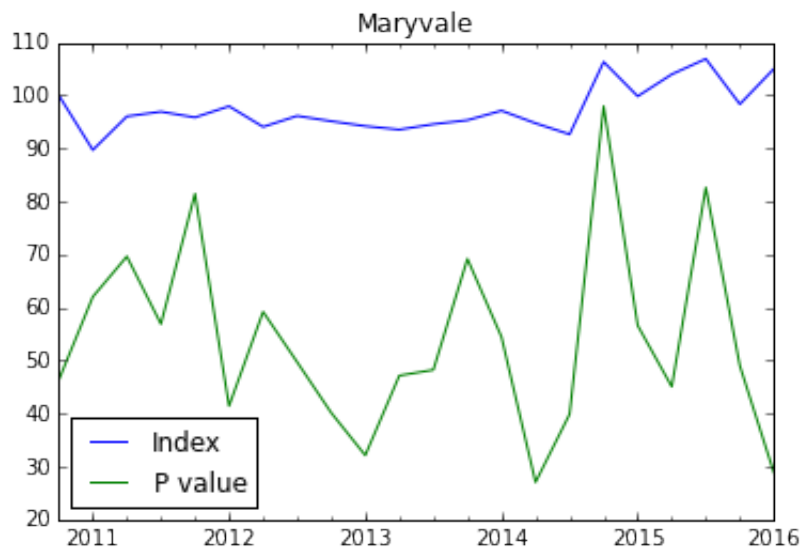


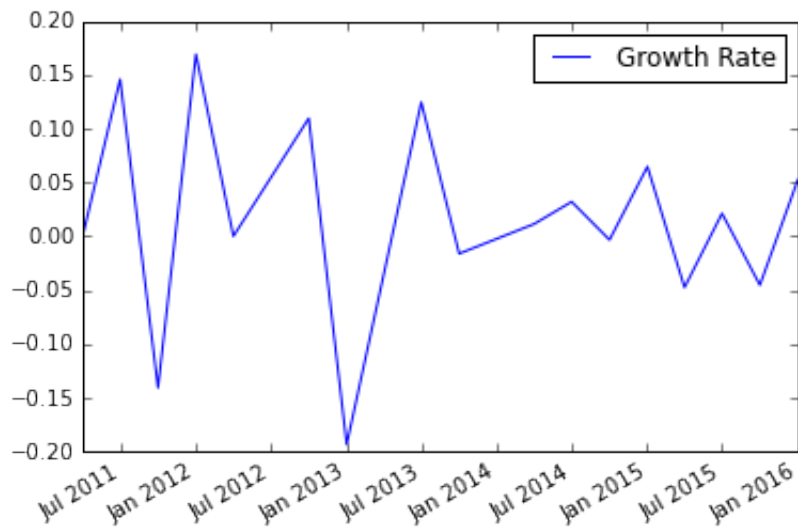
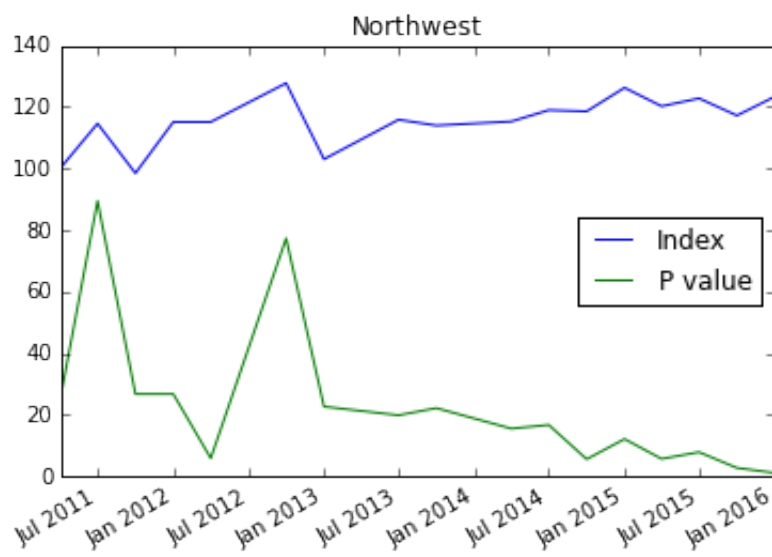
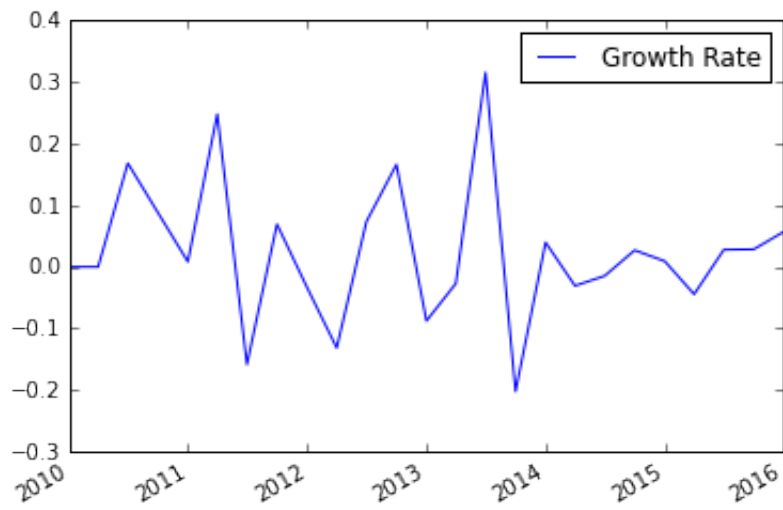


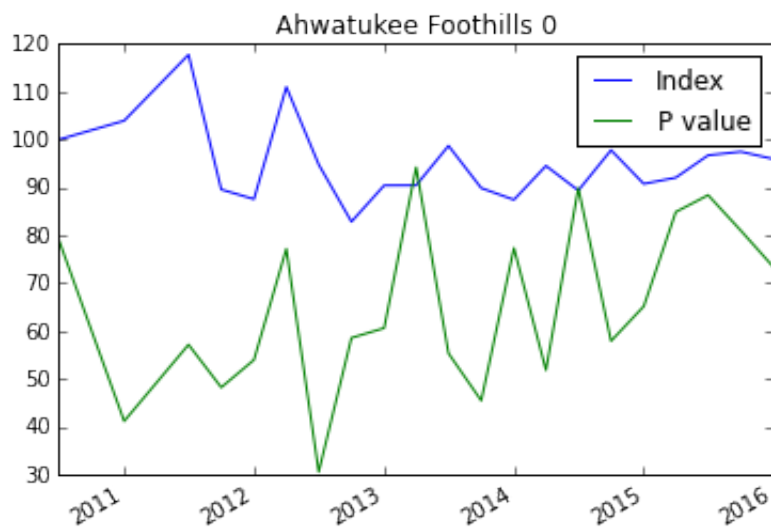
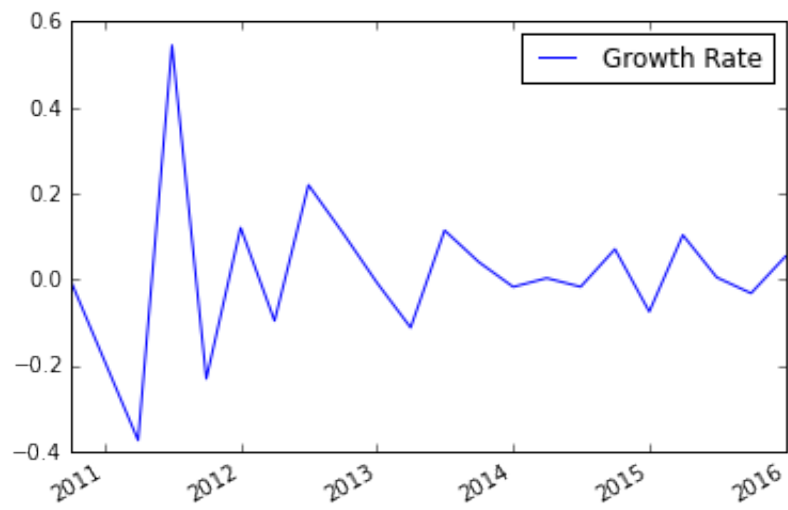
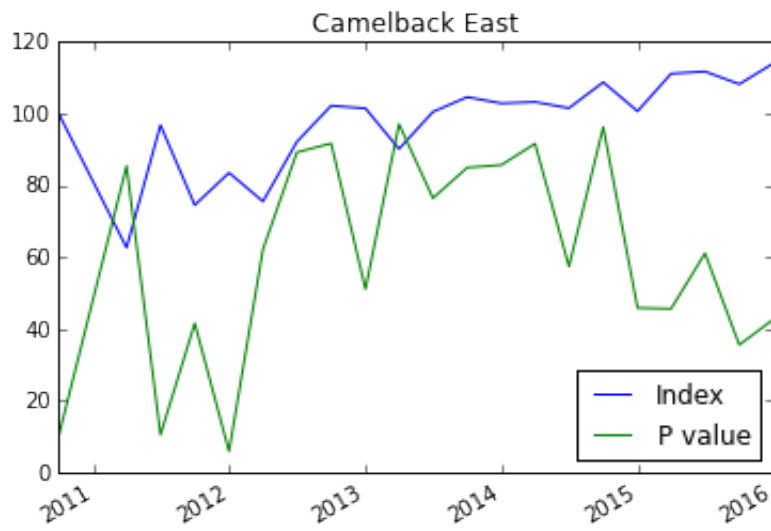


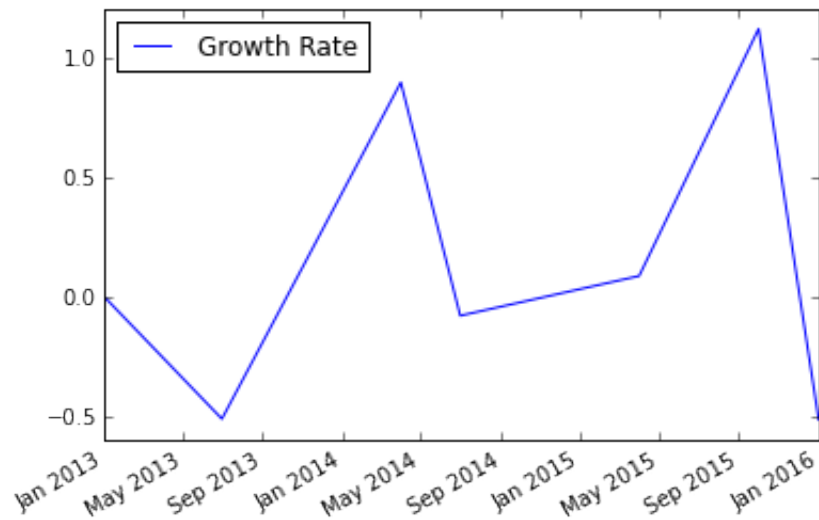
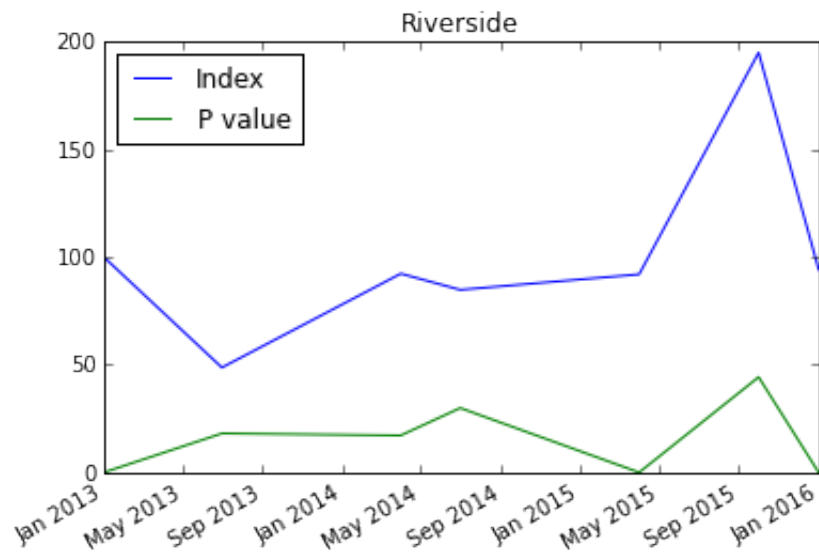
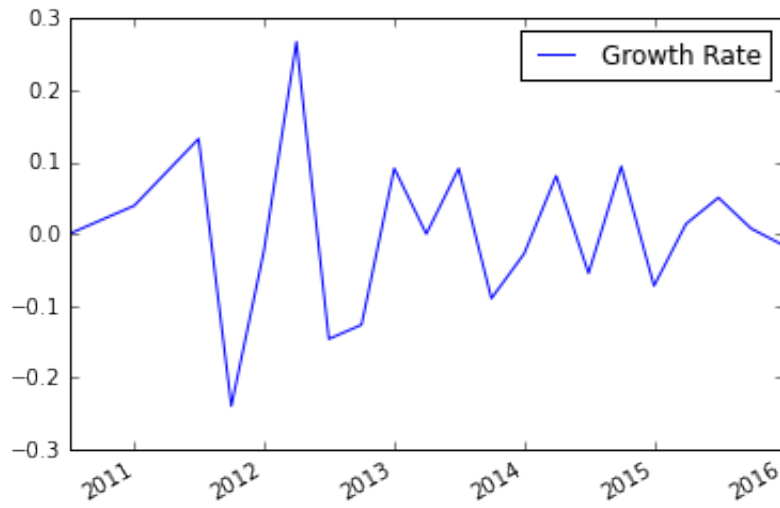


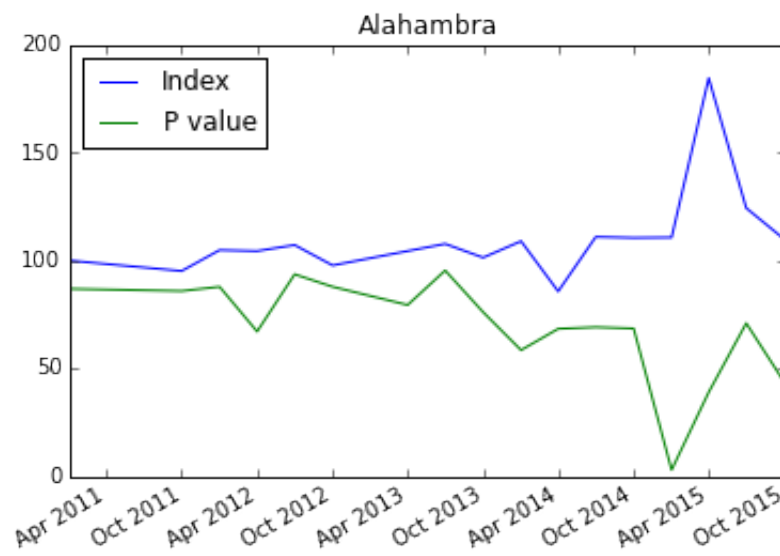
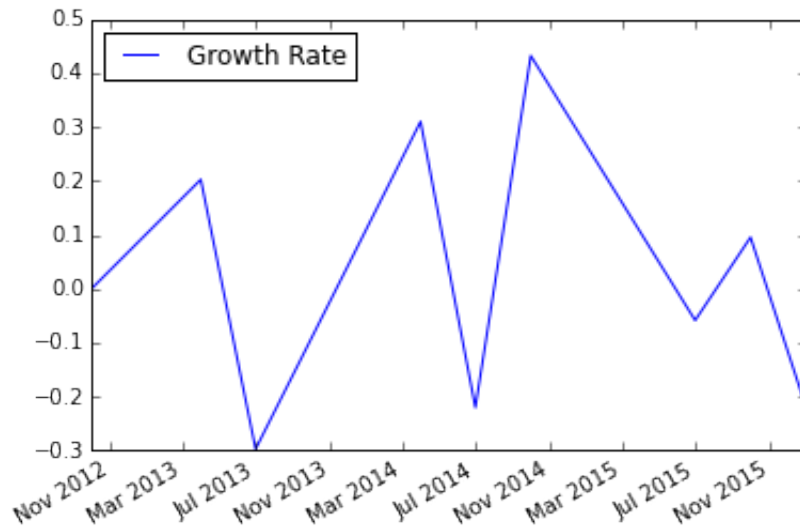
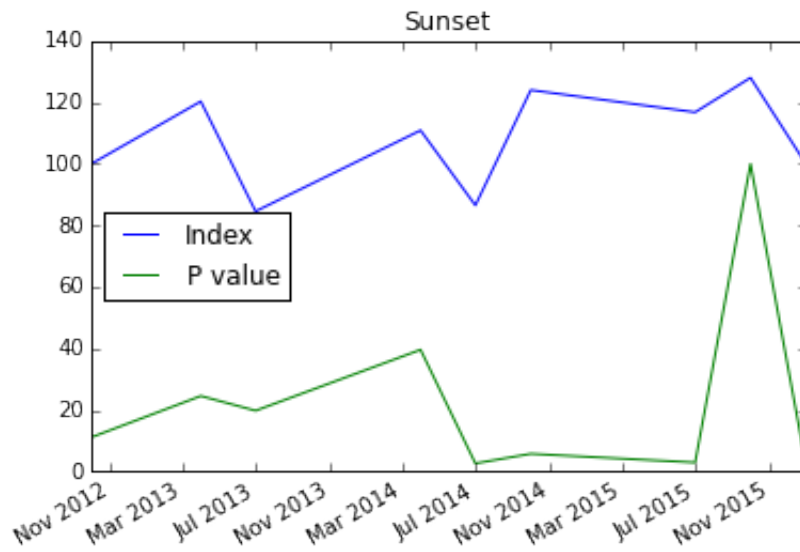


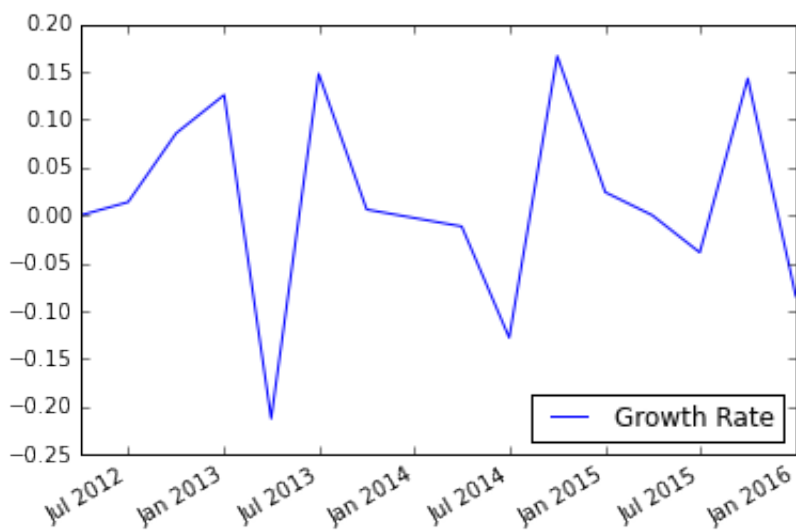
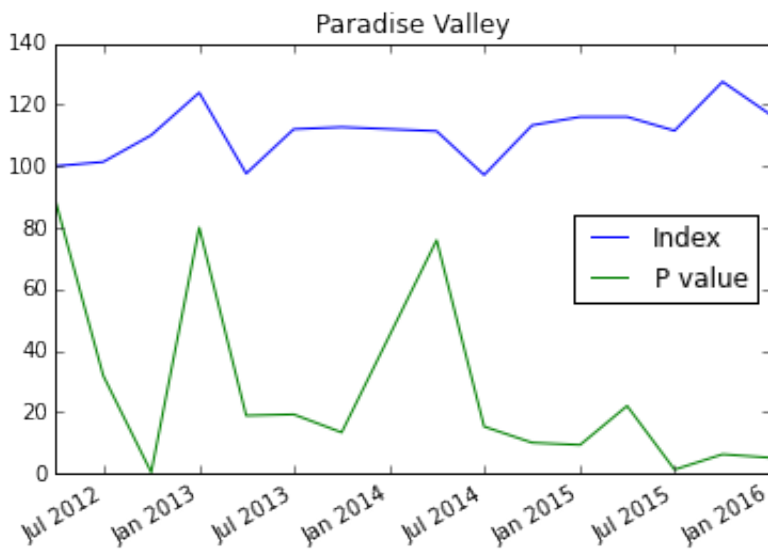
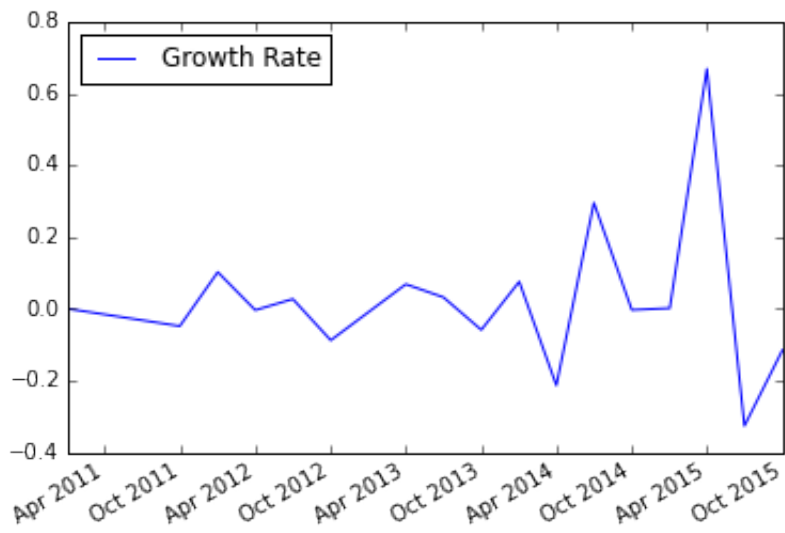


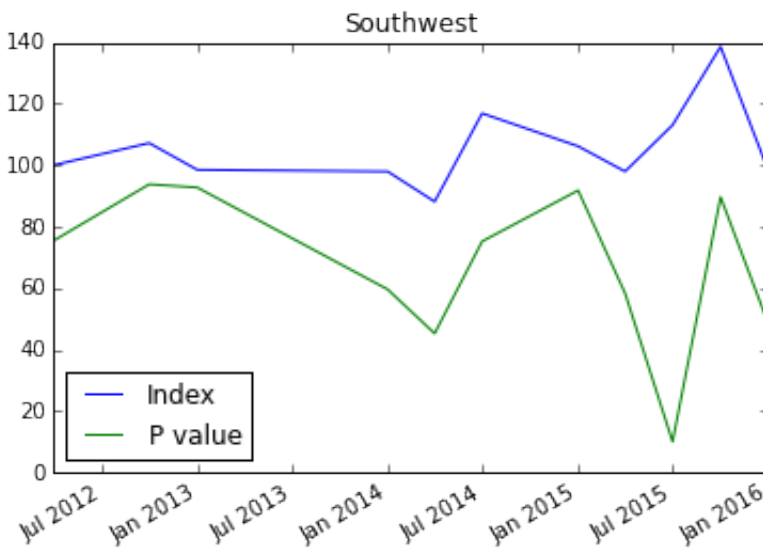
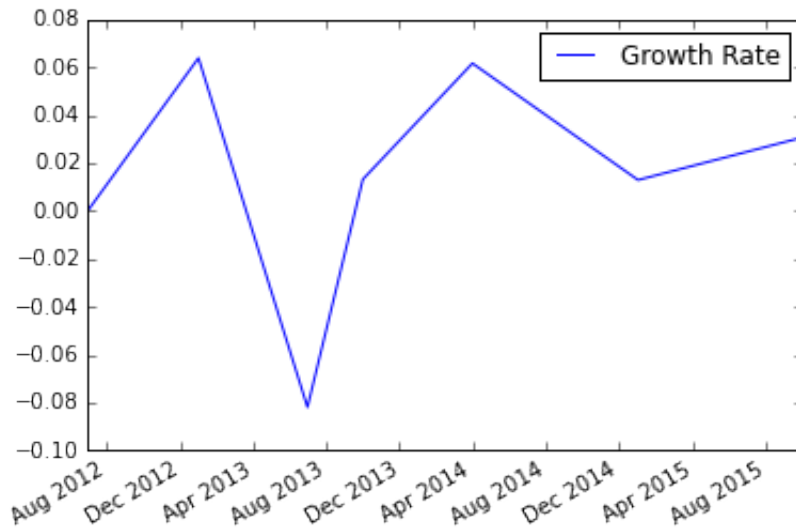
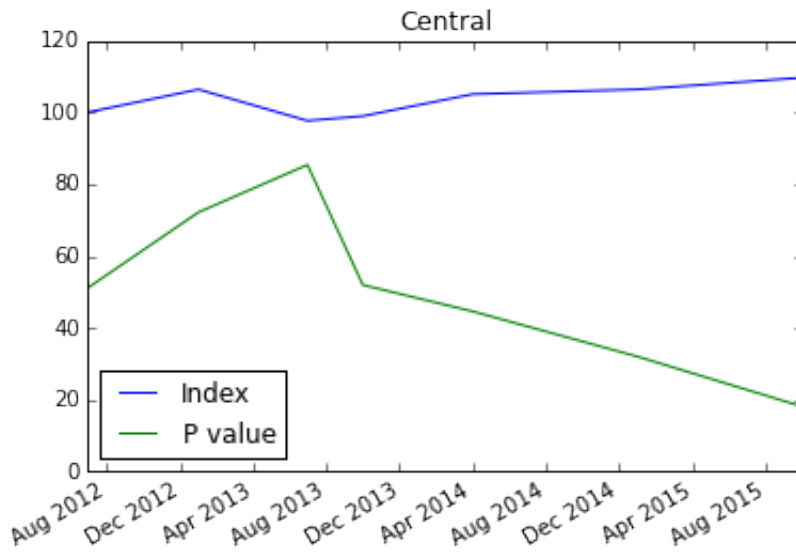


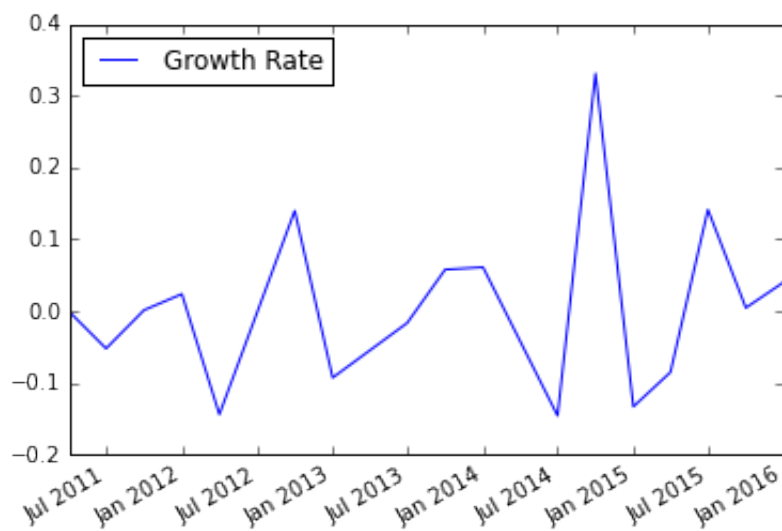
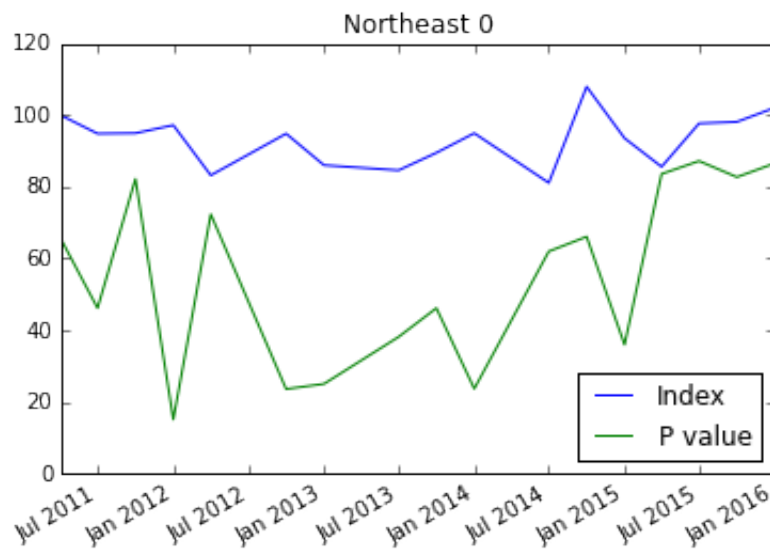
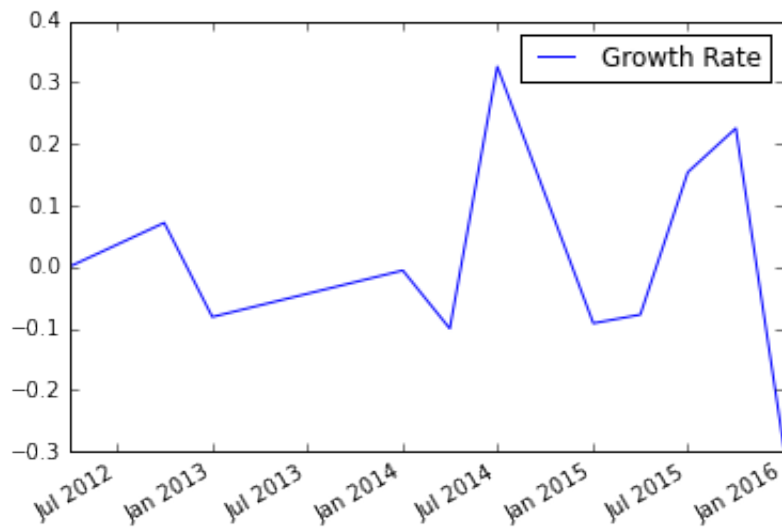


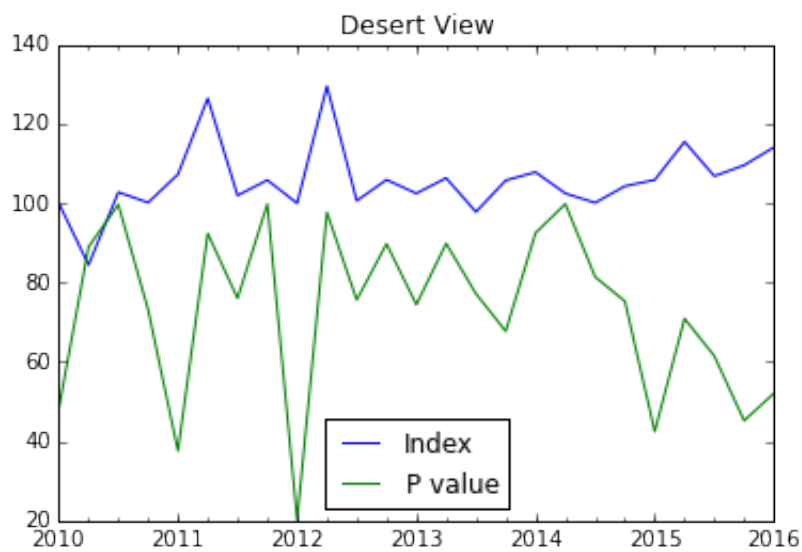
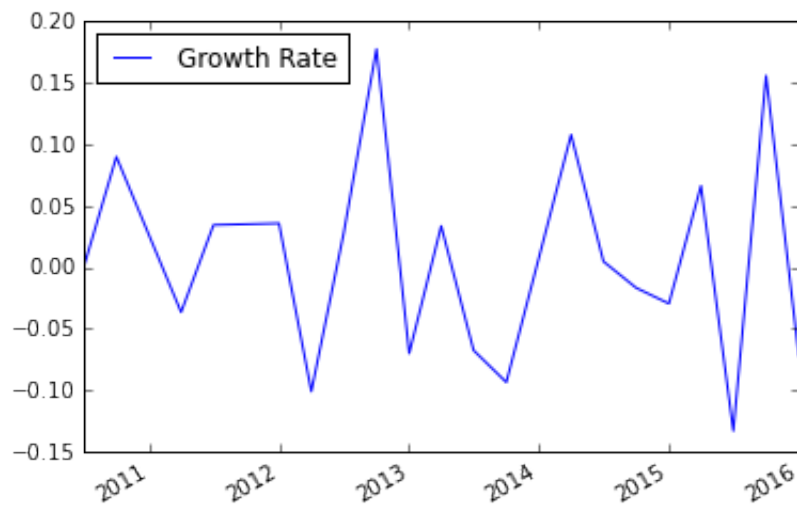
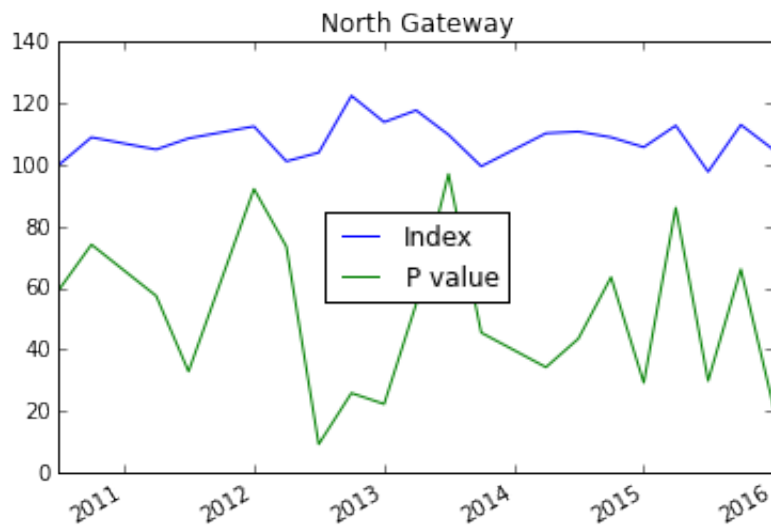


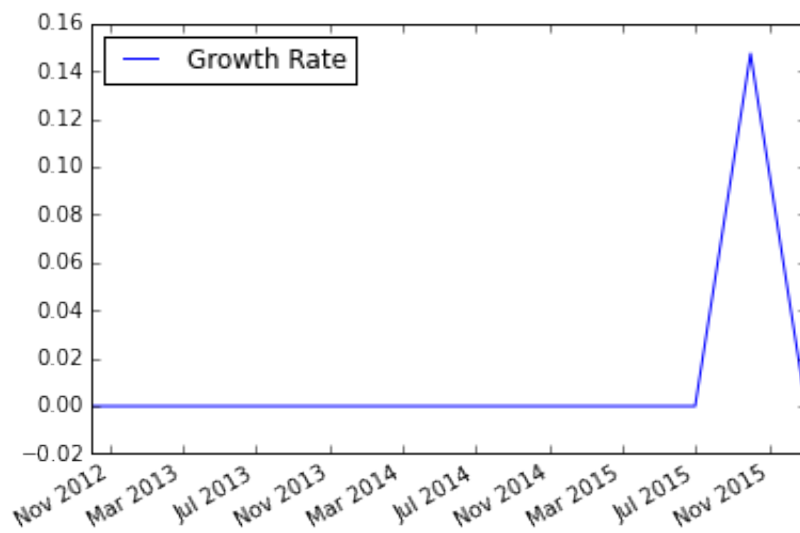
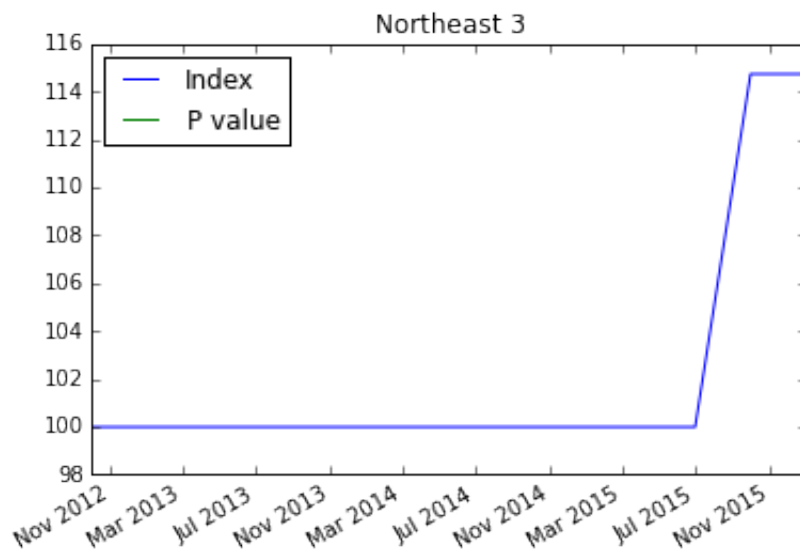
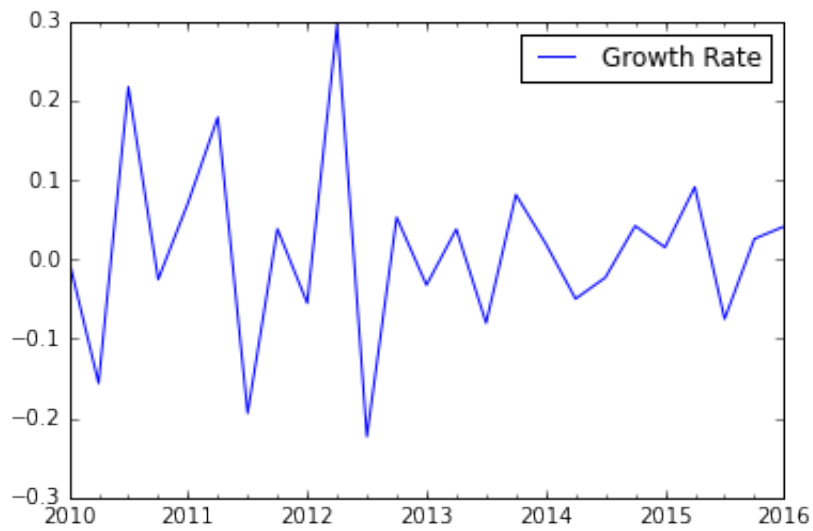












In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []: