Rats!

Karl Gourgue

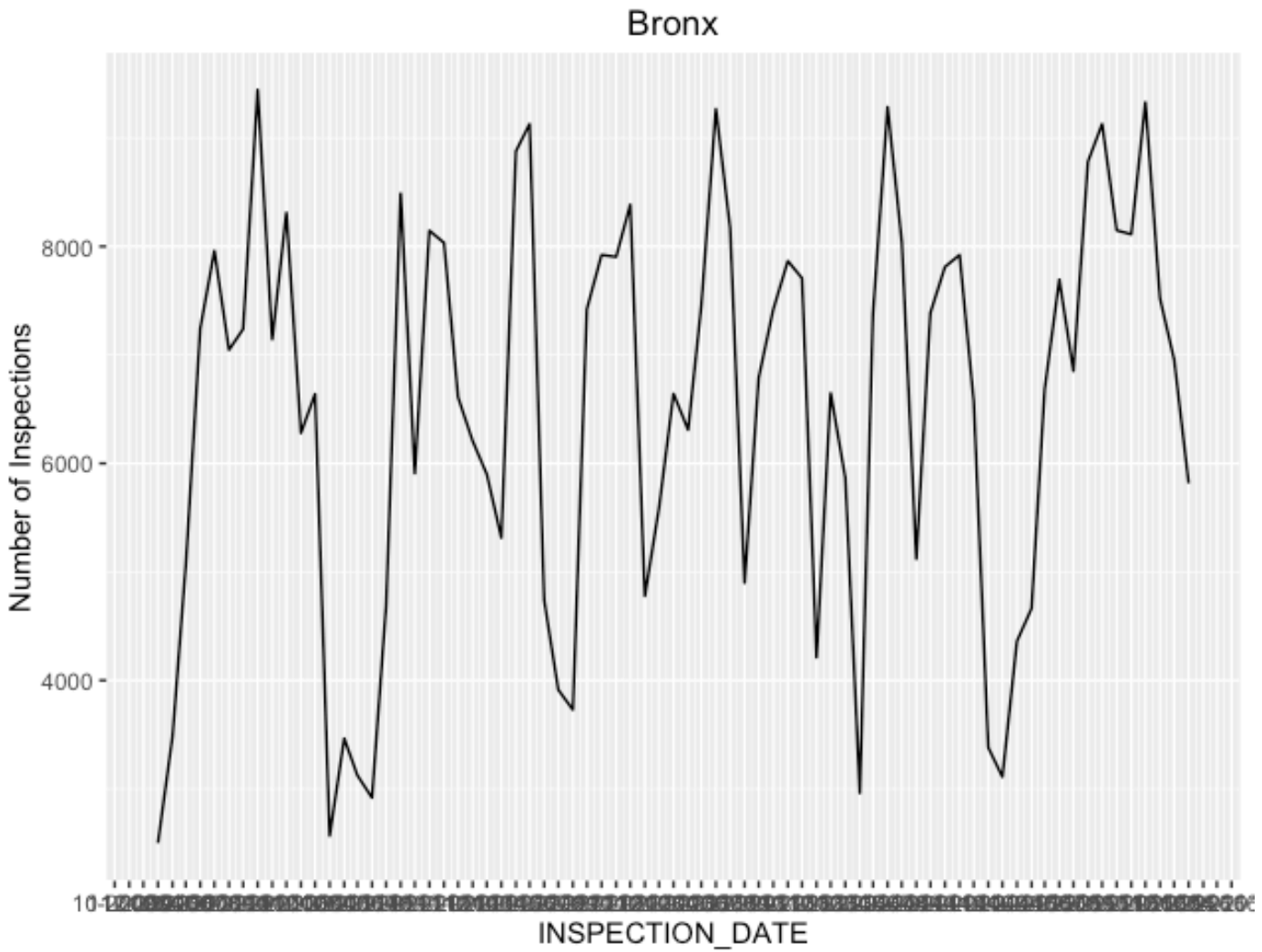
Analytics in the Digital Economy – Individual Homework 1

PART 1.
DESCRIPTIVE STATISTICS AND FIGURES

Bronx

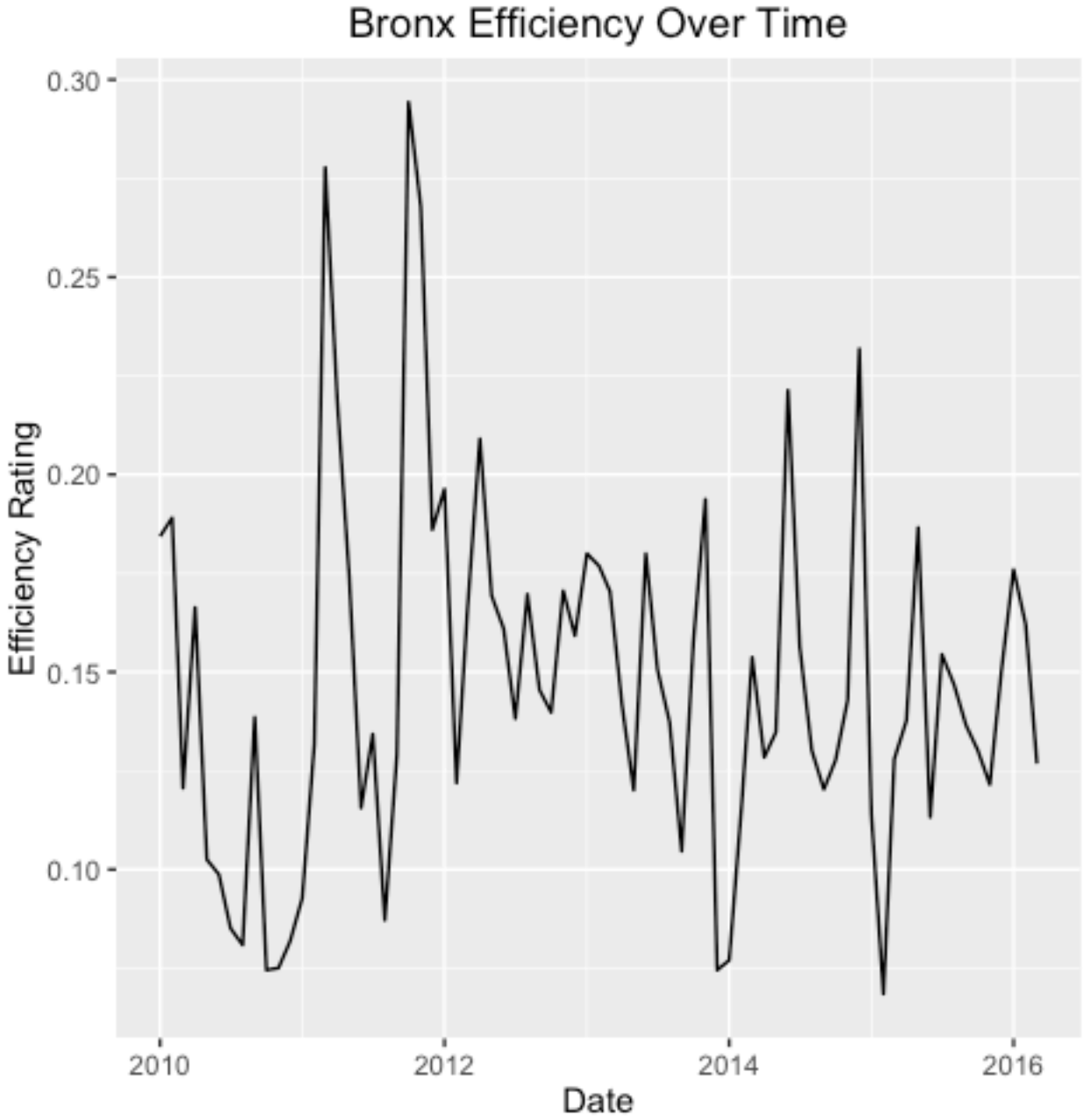
Rat sightings have been decreasing lately but are largely up since January 2010.

There also appears to be a slight bump in inspections after Hurricane Sandy.



Bronx - Efficiency

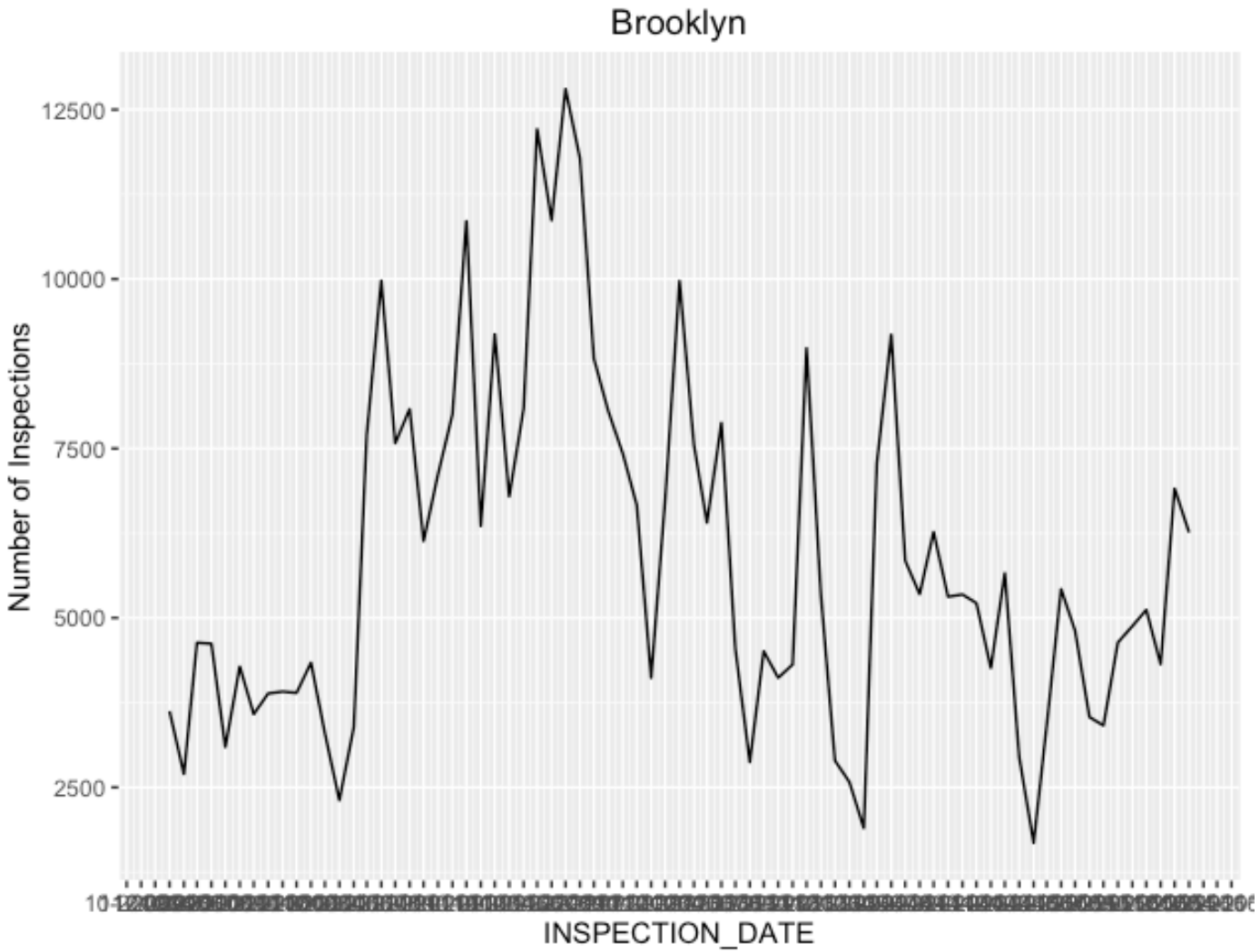
Yield is down overall, with a spike after Sandy.



Brooklyn

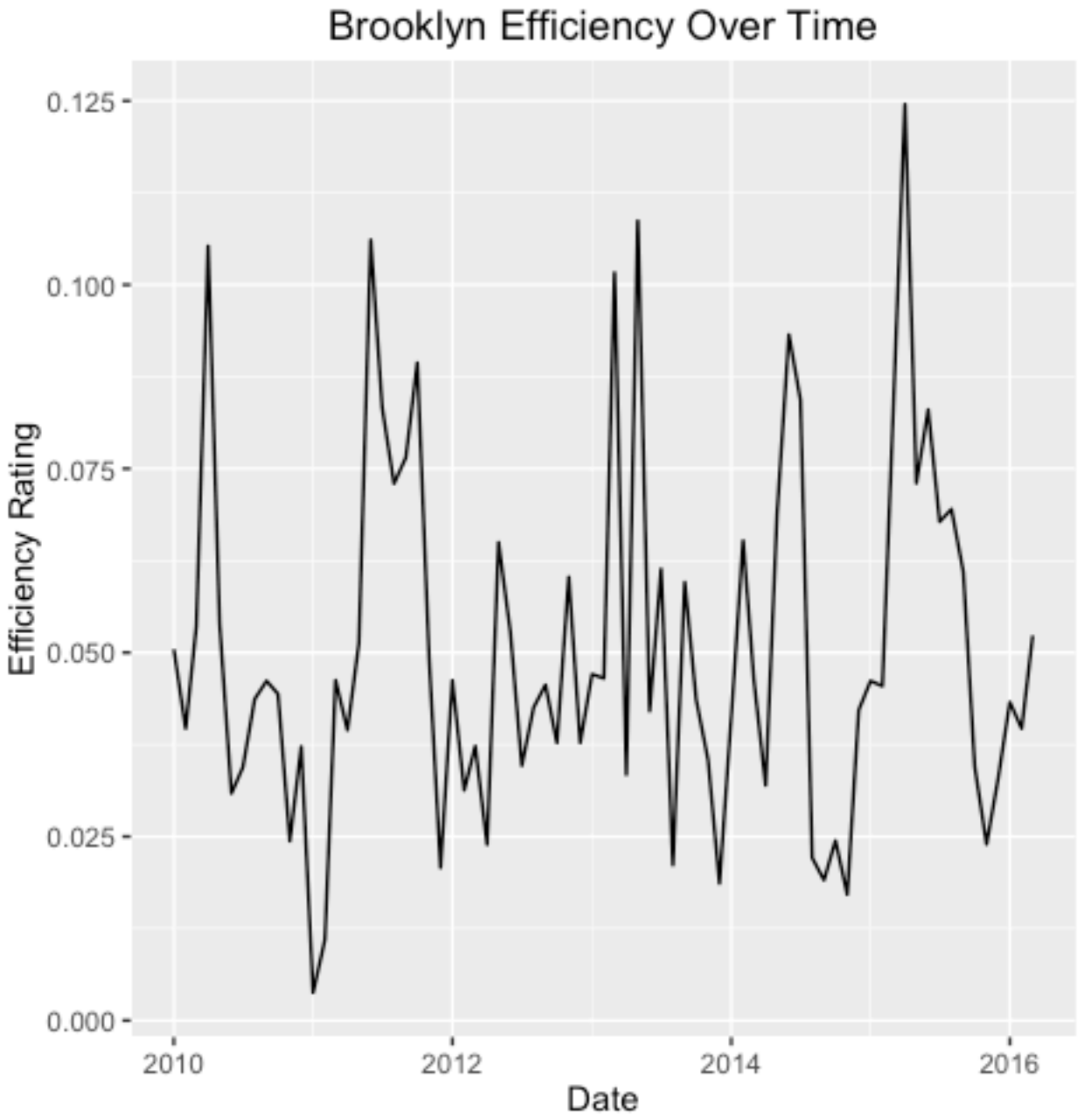
Rat sightings are up since 2010, with a major peak in the beginning of 2012.

There also appears to be a slight bump in inspections after Hurricane Sandy.



Brooklyn - Efficiency

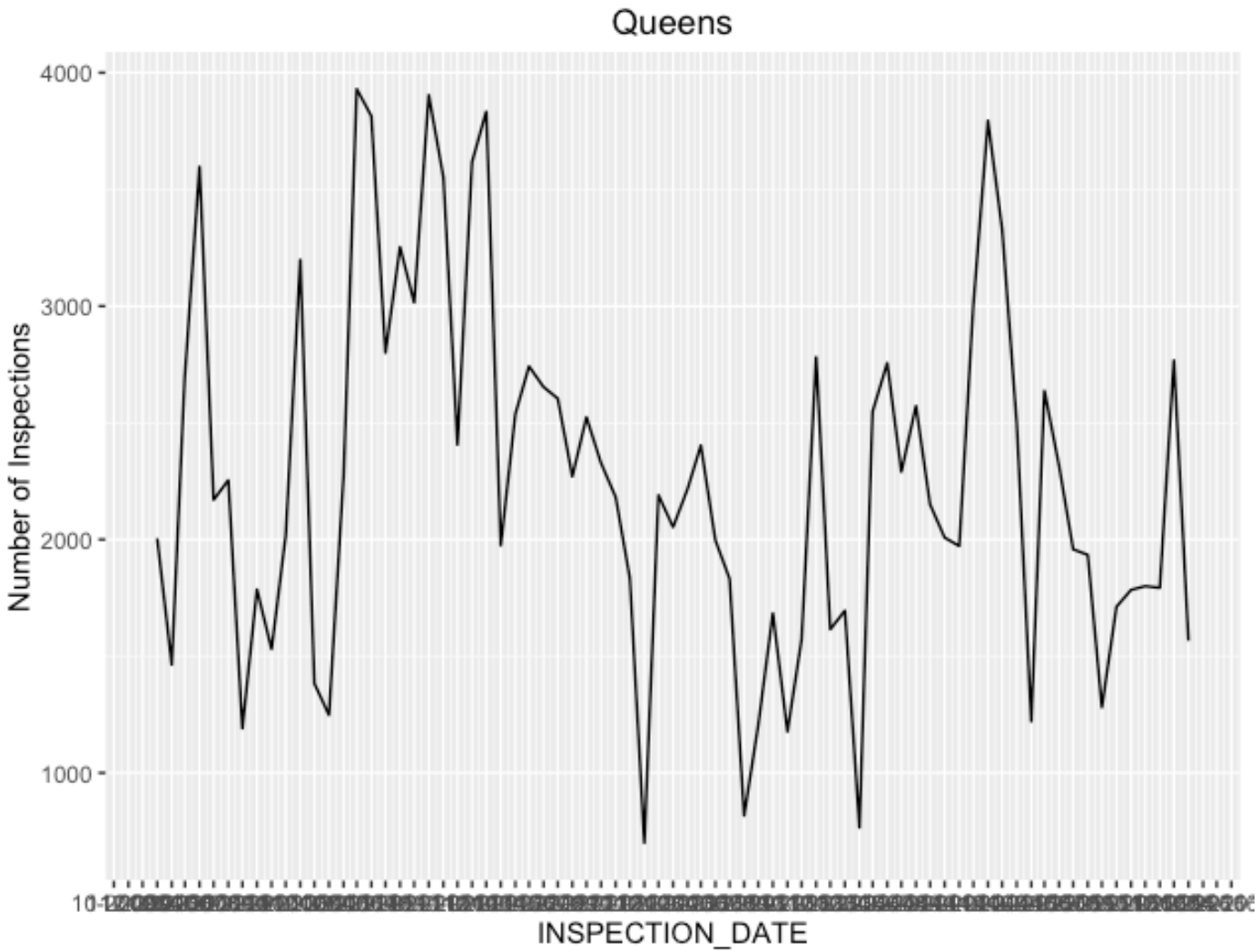
Yield is about the same between 2010 and now, with some obvious volatility over the years.



Queens

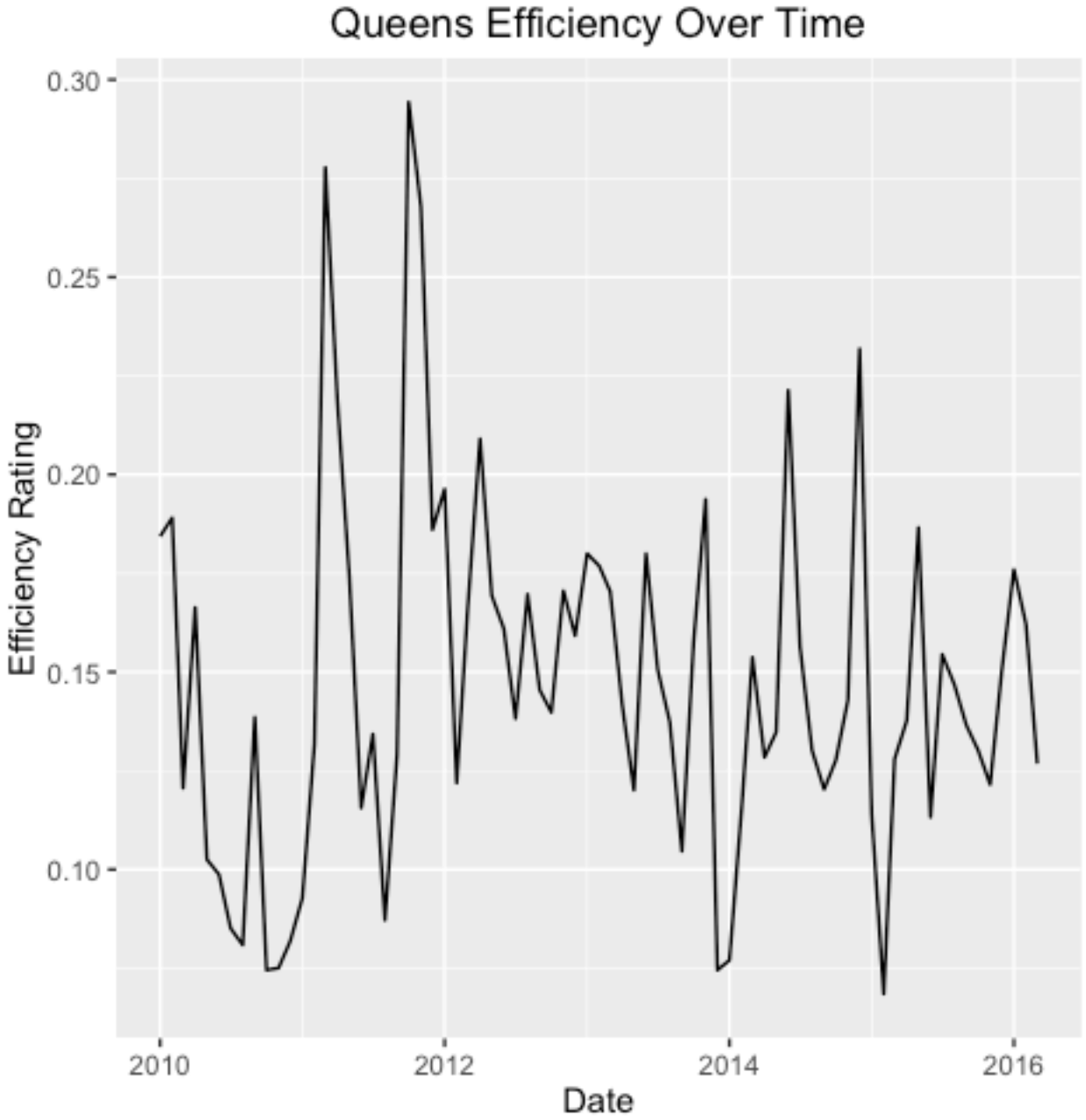
Rat sightings have been decreasing lately overall since the beginning of 2010, and are currently below January 2010 numbers.

There also appears to be a significant bump in inspections after Hurricane Sandy.



Queens - Efficiency

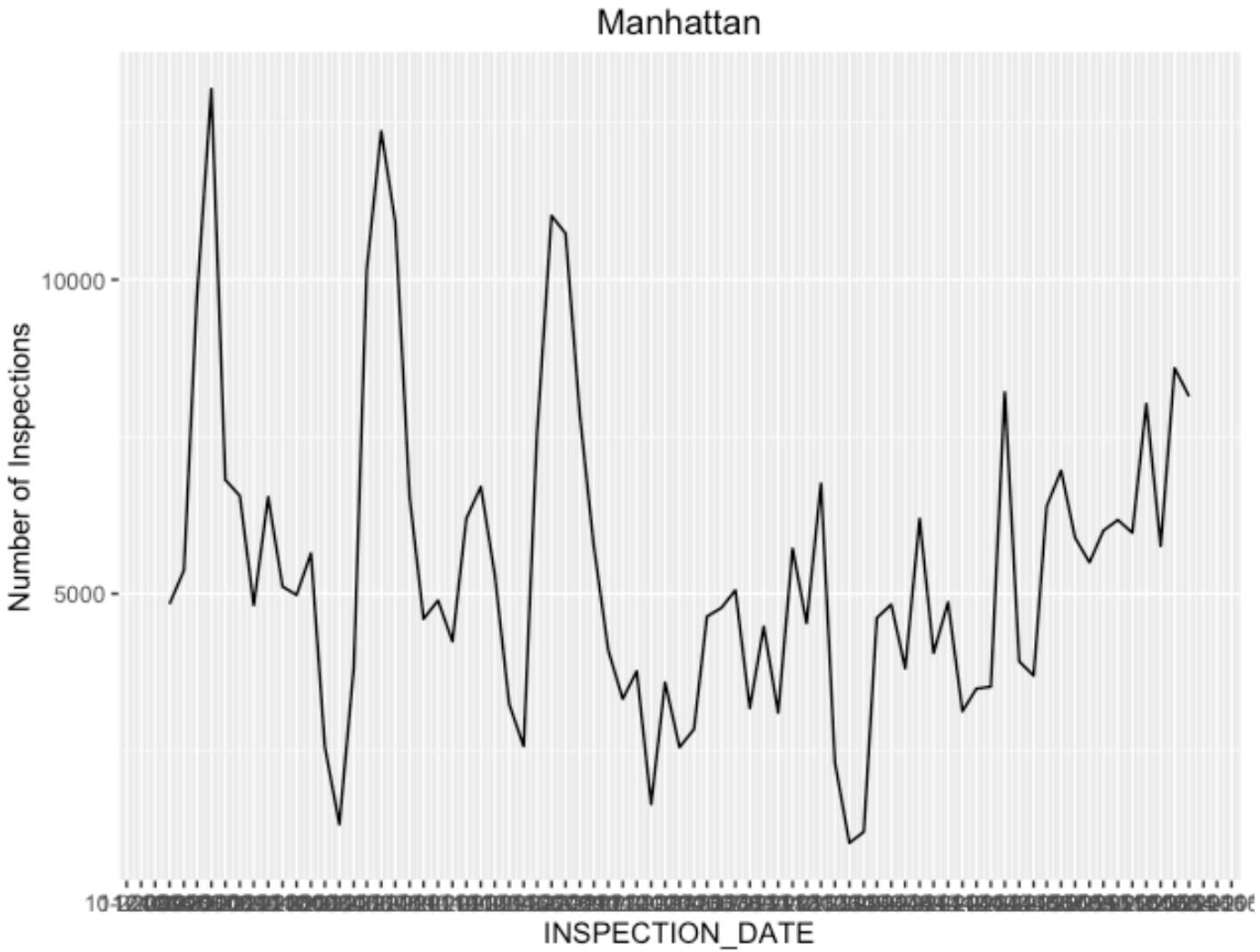
Yield is down overall, with a spike after Sandy.



Manhattan

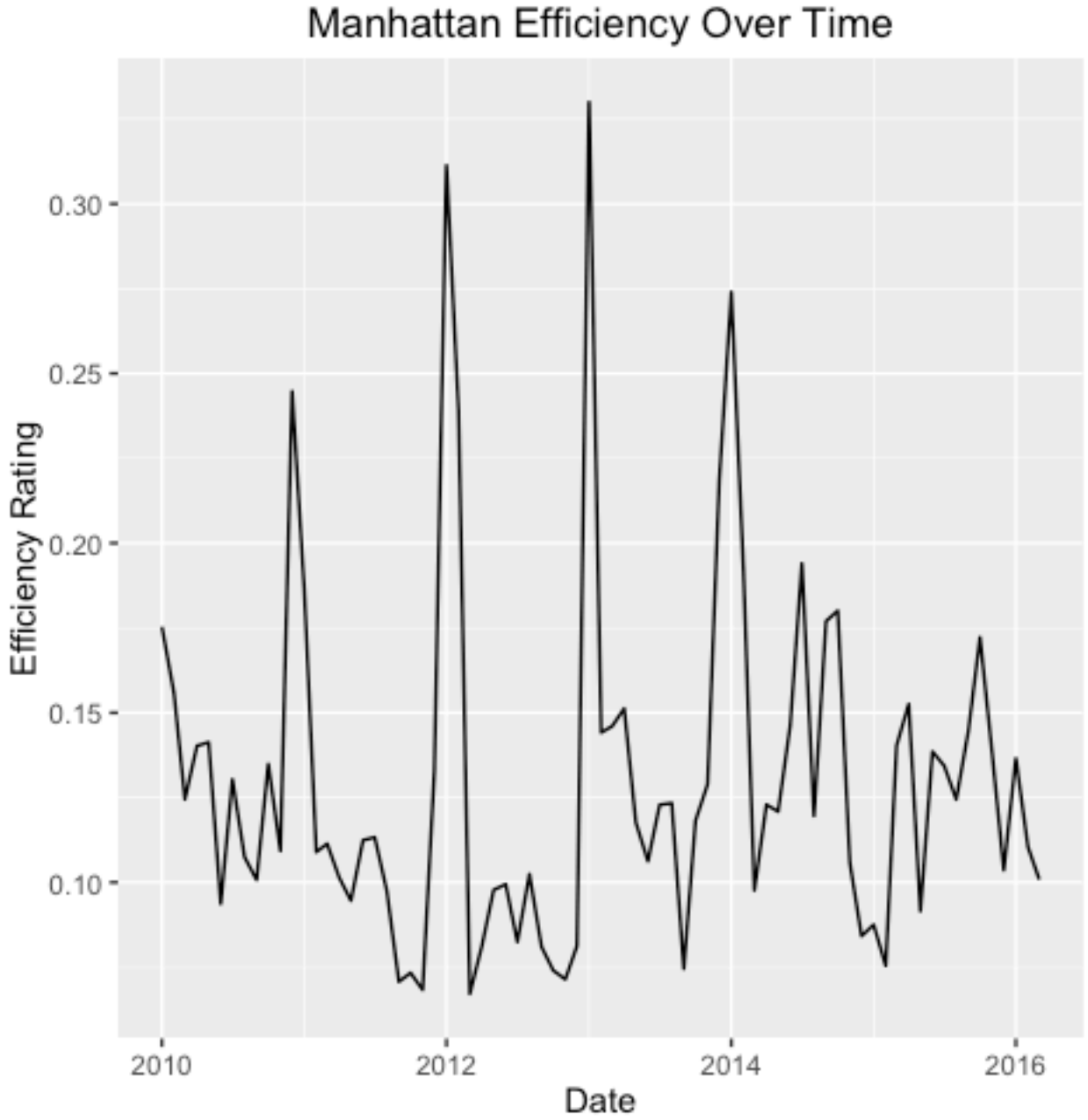
Rat sightings are up overall and showing a steady increase since the end of 2012.

There also appears to be a slight bump in inspections after Hurricane Sandy.



Manhattan - Efficiency

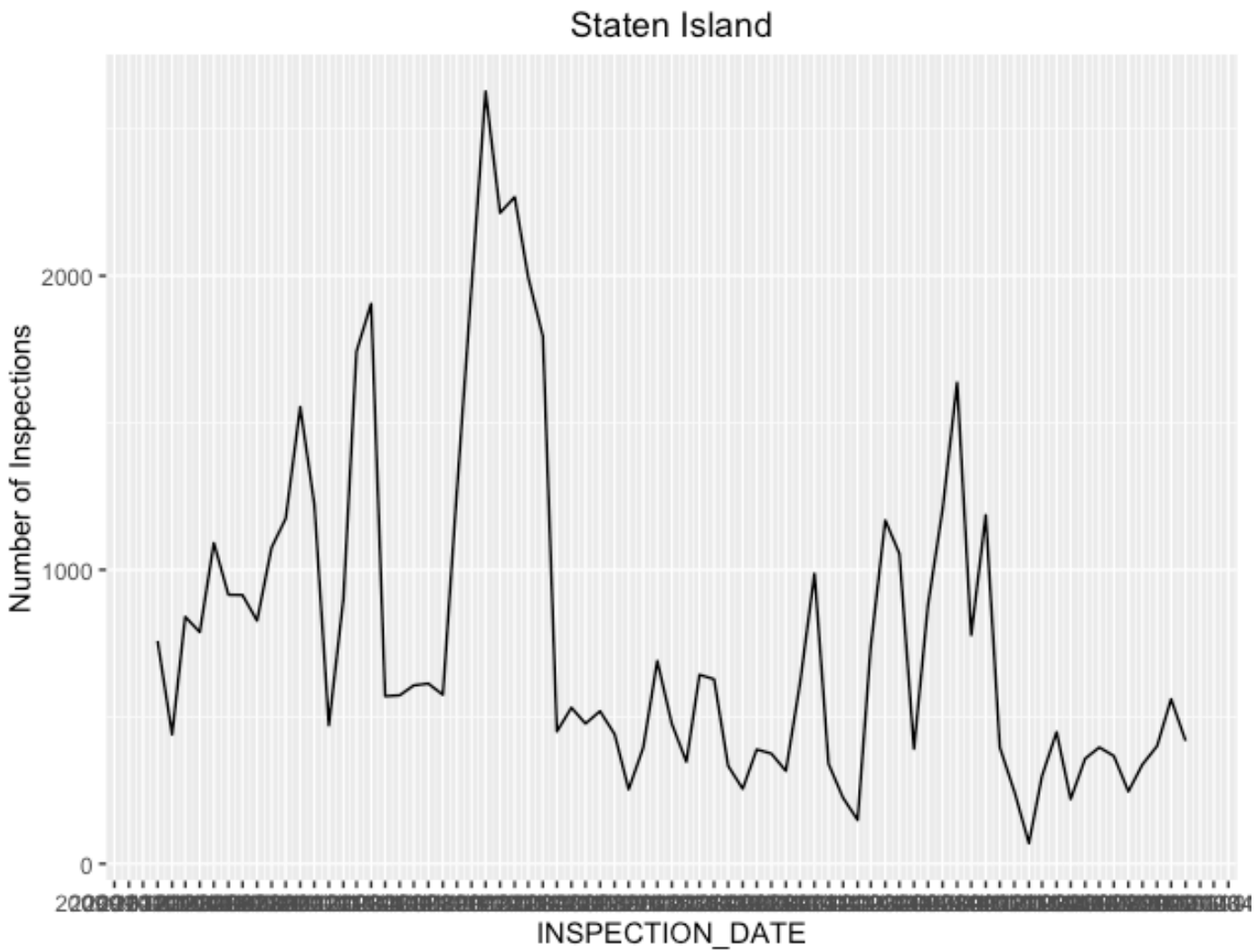
Rat sightings spiked around Sandy (reaching nearly 30% efficiency) but are down overall since 2010.



Staten Island

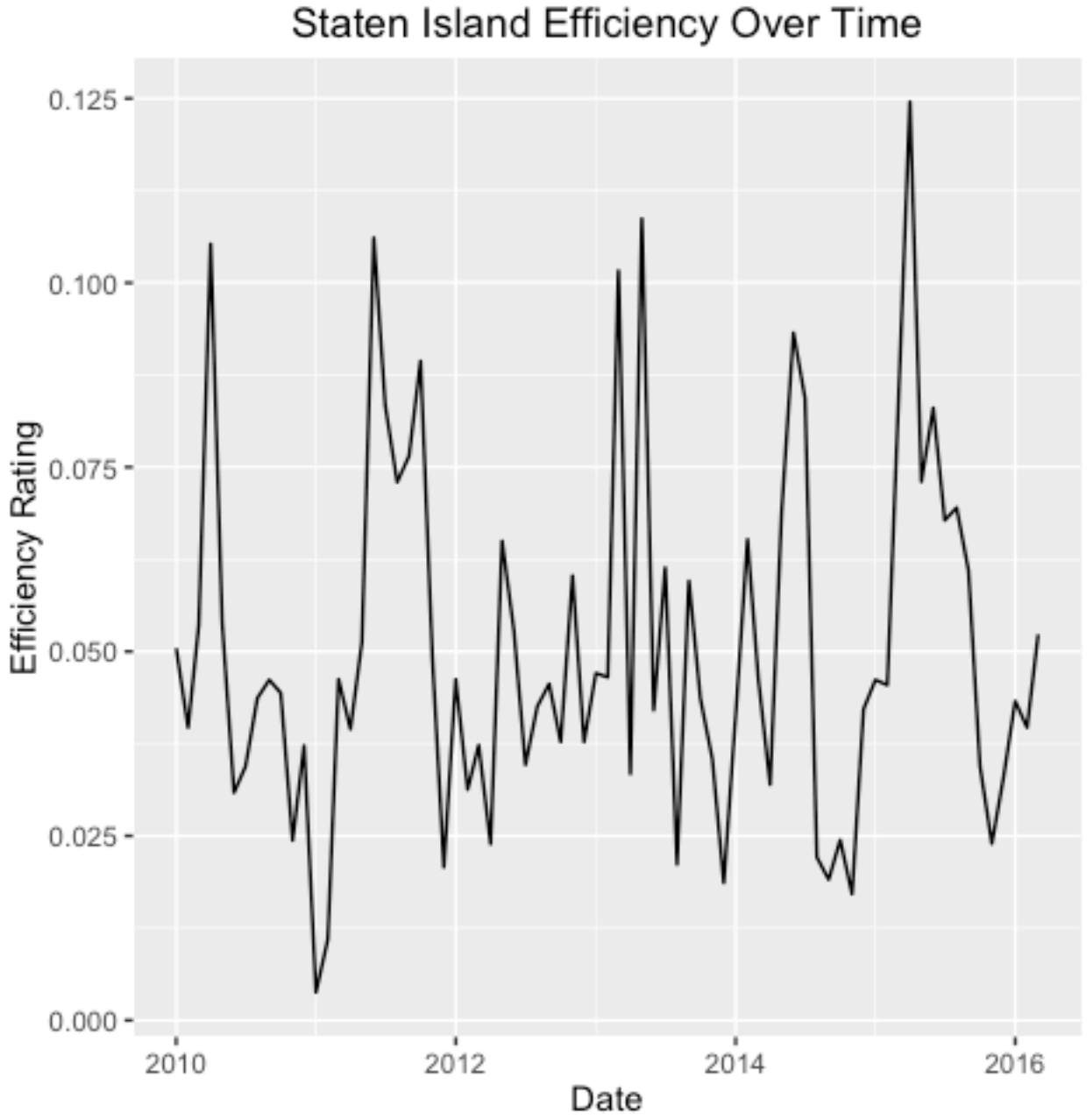
Rat sightings are down slightly since 2010, with a few major peaks before 2012.

There also appears to be a slight bump in inspections after Hurricane Sandy.



Staten Island - Efficiency

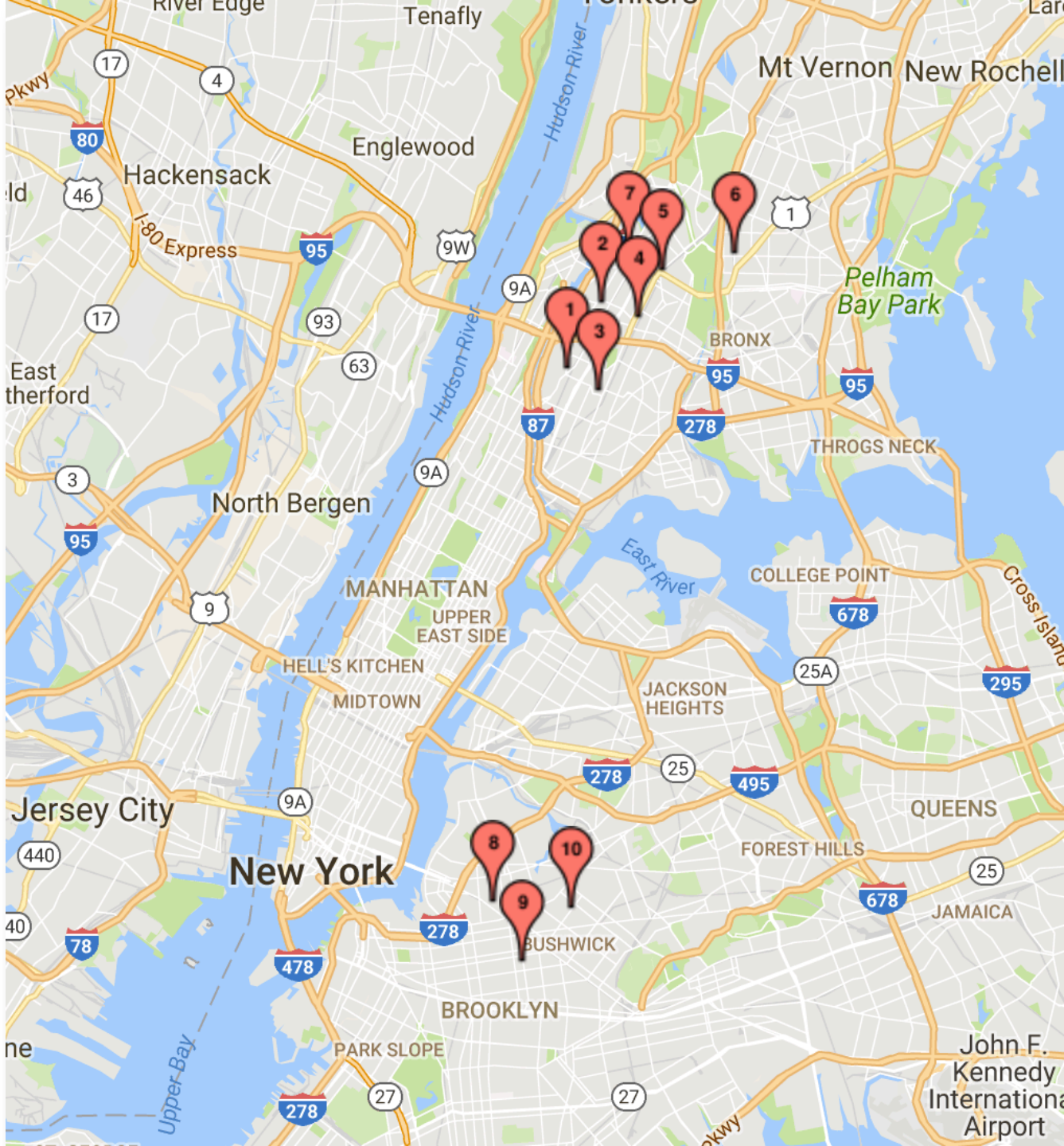
Yield is about the same between 2010 and now, with some obvious volatility over the years.



Top 10 Rat Hotspots

This quick visualization built in ZeeMaps shows where the hotspot zip codes are.

10457 – 10458 – 10456 – 10468 – 11221
10453 – 11237 – 10452 – 10467 – 11206



PART 2.
GEOGRAPHIC PATTERNS BEFORE + AFTER SANDY

14 Zip Codes rank in the top 20 before and after 2012, but not during the year 2012.

It seems as though Hurricane Sandy did not have a lasting effect on the location of rats in NYC.

PART 3.
RODENT AND RESTAURANT INSPECTION

Results of the logistic regression

Active Rat Sightings are statistically significant, as well as March to October.

None of the years are statistically significant.

```
Call:
glm(formula = RatViolation ~ efficiency + Month + Year, family = "binomial",
    data = NewRestData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6563  -0.6311  -0.6101  -0.5853   1.9710

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.4354145 17.9292629  -0.470  0.63801
efficiency  -0.0007362  0.0002502  -2.943  0.00325 **
Month02     -0.0333907  0.0182932  -1.825  0.06795 .
Month03     -0.0928233  0.0185736  -4.998 5.81e-07 ***
Month04     -0.1504877  0.0194097  -7.753 8.96e-15 ***
Month05     -0.2206161  0.0196766 -11.212  < 2e-16 ***
Month06     -0.2016660  0.0197668 -10.202  < 2e-16 ***
Month07     -0.1789740  0.0204220  -8.764  < 2e-16 ***
Month08     -0.1069047  0.0200097  -5.343 9.16e-08 ***
Month09     -0.0802317  0.0197306  -4.066 4.78e-05 ***
Month10     -0.0386062  0.0192586  -2.005  0.04500 *
Month11      0.0062238  0.0202288   0.308  0.75833
Month12      0.0040892  0.0194022   0.211  0.83308
Year2011     6.9889542 17.9304408   0.390  0.69670
Year2012     6.9507066 17.9292647   0.388  0.69826
Year2013     6.9418550 17.9292586   0.387  0.69862
Year2014     7.0023084 17.9292583   0.391  0.69613
Year2015     7.0032502 17.9292582   0.391  0.69609
Year2016     6.9258905 17.9292634   0.386  0.69928
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 421700  on 456479  degrees of freedom
Residual deviance: 421253  on 456461  degrees of freedom
  (15959 observations deleted due to missingness)
AIC: 421291

Number of Fisher Scoring iterations: 7
```

Conclusion

According to the model, when there are rats seen in the neighborhood, the likelihood of a restaurant having a violation goes down by 0.07%.

Overall these data sources are unlikely to be that useful in predicting violations in eating establishments, but they may contain a treasure trove of interesting patterns in other areas.

This regression alone reveals interesting insights in which months have heavier rat activity between years.