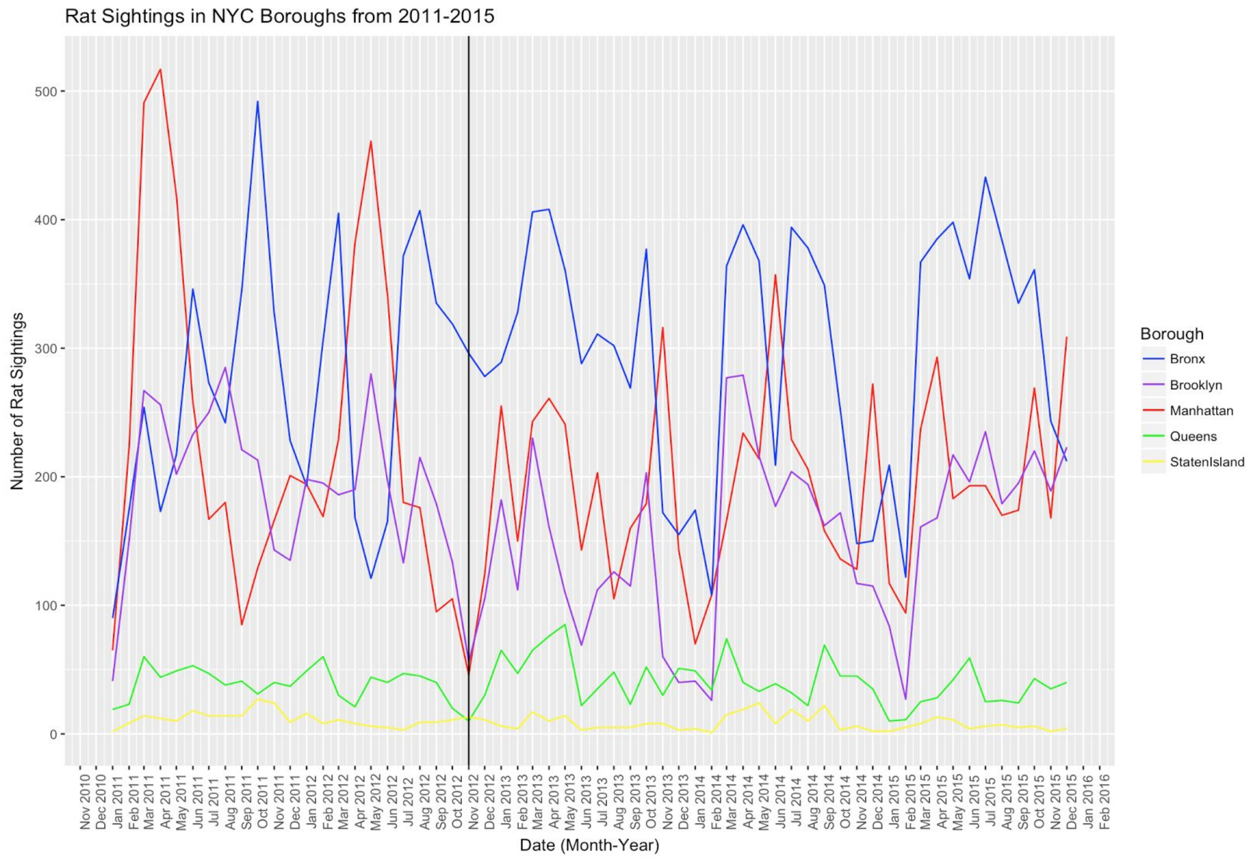
Kelly Xie
Professor Tambe
ADE, Spring 2017
Homework 1: Rats

Descriptive Statistics and Figures

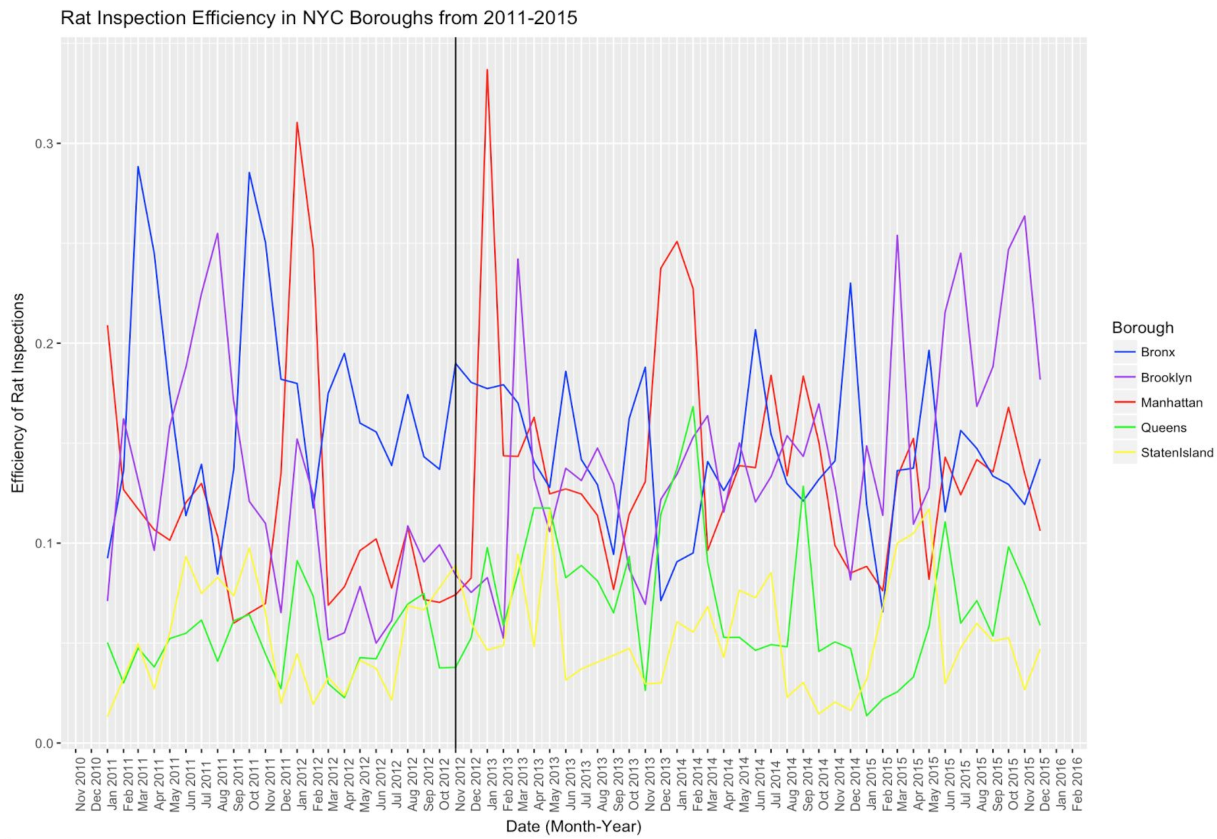
Part 1A.



From the chart, it appears that there are extreme and periodic fluctuations in rat sightings in Manhattan, The Bronx, and Brooklyn over the five year period between 2011 and 2015. In contrast, rat sightings have been relatively infrequent and steady in Queens and Staten Island. Overall, there is also a general downward trend in rat sightings in Brooklyn and Manhattan from 2011 to 2015.

The chart shows a meaningful increase in rat sightings after Hurricane Sandy (Oct-Nov 2012), with the largest spikes occurring in The Bronx, Manhattan, and Brooklyn, and a smaller spike in Queens.

Part 1B.



At a glance, it is clear there was a significant upwards spike in Rat Inspection Efficiency in Manhattan following Hurricane Sandy (Oct-Nov 2012).

Part 1C.

Top 10 “Hot Spot” Zip Codes for Rat Sightings

```
> print(zipRat[1:10,])
Source: local data frame [10 x 3]
Groups: ZIP_CODE [10]

   ZIP_CODE count  BOROUGH
      <dbl> <dbl>   <fctr>
1     10457  2745    Bronx
2     10458  2595    Bronx
3     10456  2081    Bronx
4     10468  1719    Bronx
5     10453  1503    Bronx
6     11221  1494 Brooklyn
7     10452  1157    Bronx
8     10467  1102    Bronx
9     11237  1070 Brooklyn
10    11206   935 Brooklyn
>
```

Most rat sightings are in The Bronx, followed by Brooklyn.

Geographic Patterns

Part 2A.

Top 20 Zip Codes …

Before 2012

```
> print(topBefore[1:20,])
Source: local data frame [20 x 3]
Groups: ZIP_CODE [20]

   ZIP_CODE count   BOROUGH
      <dbl> <dbl>    <fctr>
1     10457   978     Bronx
2     10456   890     Bronx
3     10458   816     Bronx
4     11221   646  Brooklyn
5     10453   551     Bronx
6     10468   534     Bronx
7     10031   506 Manhattan
8     11206   480  Brooklyn
9     10032   477 Manhattan
10    10009   455 Manhattan
11    11237   434  Brooklyn
12    10002   424 Manhattan
13    10472   400     Bronx
14    11238   394  Brooklyn
15    11216   391  Brooklyn
16    11211   390  Brooklyn
17    10029   388 Manhattan
18    11217   385  Brooklyn
19    10013   378 Manhattan
20    10459   362     Bronx
> |
```

During Hurricane Sandy

```
> print(topSandy[1:20,])
Source: local data frame [20 x 3]
Groups: Incident.Zip [20]

   Incident.Zip count   Borough
         <fctr> <dbl>    <fctr>
1         10025    14 MANHATTAN
2         10456    11     BRONX
3         11237    10  BROOKLYN
4         11207     9  BROOKLYN
5         11208     9  BROOKLYN
6         10027     8 MANHATTAN
7         10458     8     BRONX
8         10016     7 MANHATTAN
9         10467     7     BRONX
10        11226     7  BROOKLYN
11        10452     6     BRONX
12        11216     6  BROOKLYN
13        11222     6  BROOKLYN
14        10013     5 MANHATTAN
15        10024     5 MANHATTAN
16        10032     5 MANHATTAN
17        10453     5     BRONX
18        11212     5  BROOKLYN
19        11221     5  BROOKLYN
20        11235     5  BROOKLYN
> |
```

After 2012

```
> print(topAfter[1:20,])
Source: local data frame [20 x 3]
Groups: ZIP_CODE [20]

   ZIP_CODE count   BOROUGH
      <dbl> <dbl>    <fctr>
1     10457  1800     Bronx
2     10458  1757     Bronx
3     10456  1225     Bronx
4     10468  1196     Bronx
5     10453   997     Bronx
6     10452   963     Bronx
7     11221   876  Brooklyn
8     10467   764     Bronx
9     11237   738  Brooklyn
10    10009   637 Manhattan
11    10460   636     Bronx
12    11206   627  Brooklyn
13    10029   556 Manhattan
14    10002   534 Manhattan
15    11216   534  Brooklyn
16    10025   528 Manhattan
17    10033   475 Manhattan
18    10027   457 Manhattan
19    10032   439 Manhattan
20    11385   404    Queens
>
```

Part 2B.

Based on these results, Hurricane Sandy appears to have displaced rats to Manhattan and Brooklyn during the hurricane. In the aftermath, there is a greater number of rat sightings across all boroughs, with the majority of the rat population still residing in The Bronx both before and after 2012. There appears to be significant displacement of rats moving away from Brooklyn, and towards The Bronx and Manhattan.

Rodents and Restaurant Inspection

Part 3.

```
> set.seed(4698)
> fit = glm(ratViolation ~ activeRatSightings + month + year, data=restaurant, family=binomial)
> summary(fit)

Call:
glm(formula = ratViolation ~ activeRatSightings + month + year,
    family = binomial, data = restaurant)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6512  -0.6315  -0.6069  -0.5824   1.9532

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -1.501429   0.022375 -67.103  < 2e-16 ***
activeRatSightings -0.007975   0.020816  -0.383 0.701639
month02            -0.024215   0.021037  -1.151 0.249702
month03            -0.095263   0.020418  -4.666 3.08e-06 ***
month04            -0.143535   0.020148  -7.124 1.05e-12 ***
month05            -0.222871   0.020390 -10.930  < 2e-16 ***
month06            -0.202039   0.020461  -9.875  < 2e-16 ***
month07            -0.172278   0.021081  -8.172 3.03e-16 ***
month08            -0.104337   0.020711  -5.038 4.71e-07 ***
month09            -0.076216   0.020428  -3.731 0.000191 ***
month10            -0.037760   0.020016  -1.886 0.059234 .
month11             0.009284   0.020915   0.444 0.657128
month12             0.004926   0.020104   0.245 0.806452
year2013           -0.014568   0.017582  -0.829 0.407340
year2014            0.048892   0.017331   2.821 0.004786 **
year2015            0.045450   0.017344   2.620 0.008781 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 405664  on 440981  degrees of freedom
Residual deviance: 405218  on 440966  degrees of freedom
  (591 observations deleted due to missingness)
AIC: 405250

Number of Fisher Scoring iterations: 4
```

Note: Data from years 2012-2015.

a) Looking at the logistic regression model summary, the predictor activeRatSightings is associated with a p-value ($p=0.7016$) greater than the threshold $p=0.05$. Thus, the number of Rat Sightings **does not have a statistically significant relationship** with the variable ratViolation (which represents whether a restaurant violation in the same zip code, year, and month is due to rat and mouse problems).

b) A statistically insignificant relationship between active rat sightings and restaurant violations suggests that **neither argument can be supported** by the regression results. From this, we can gather that the rodent inspection data is **not a useful source** for predicting whether restaurants located in a same zip code are likely to have violations related to rat sightings. I would recommend to the city of New York to use other datasets that may be more useful in predicting restaurant violations, such as customer transactional data for rat exterminator companies, or text mining consumer reviews from restaurant-specific sources like Yelp and Open Table.