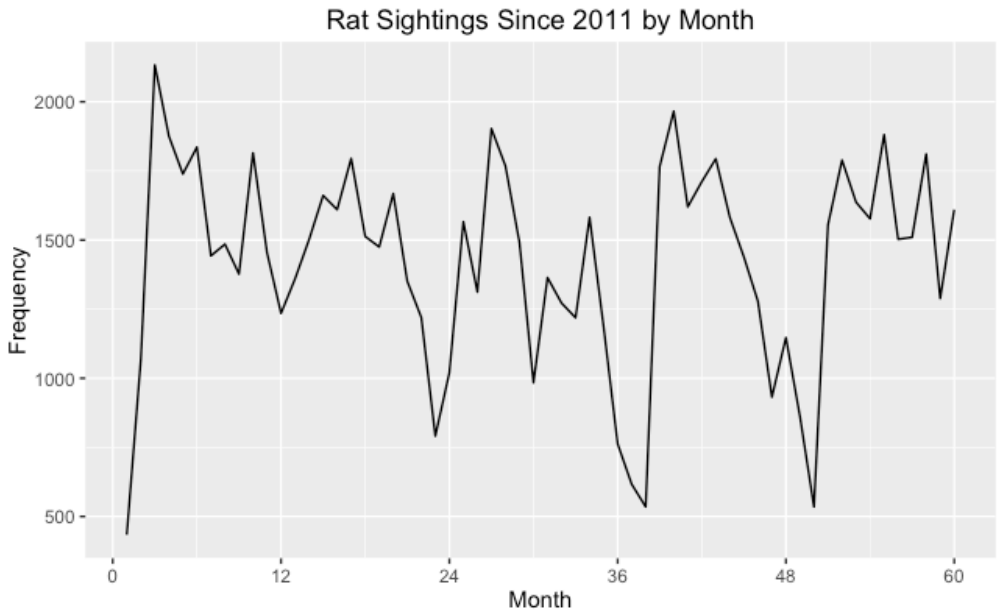
ADE Homework 1– Rats! 🐭

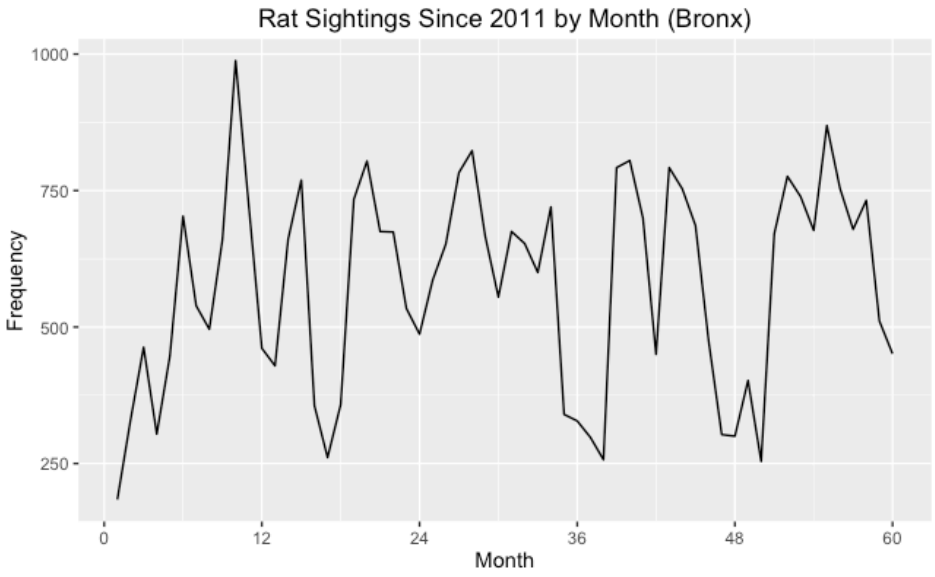
Justin Gage

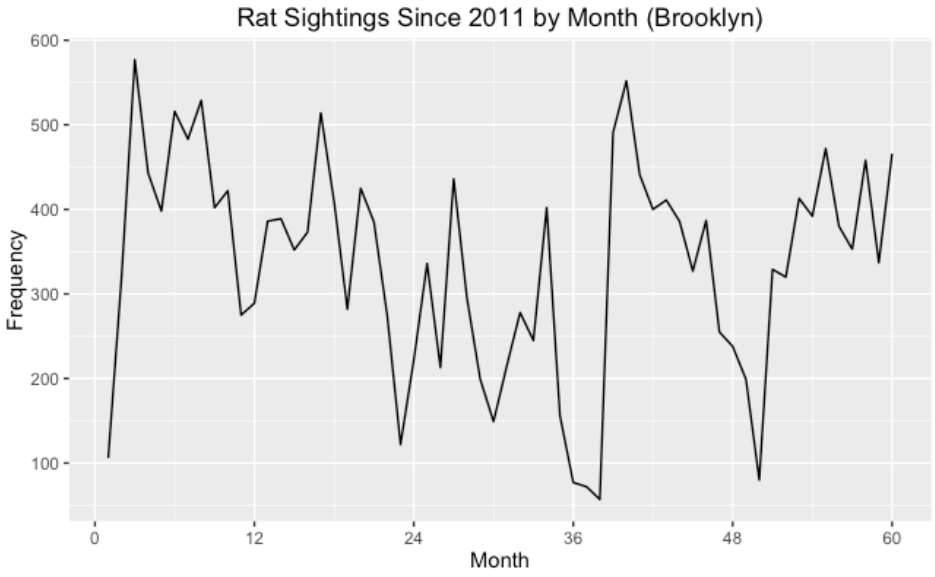
Question 1A.

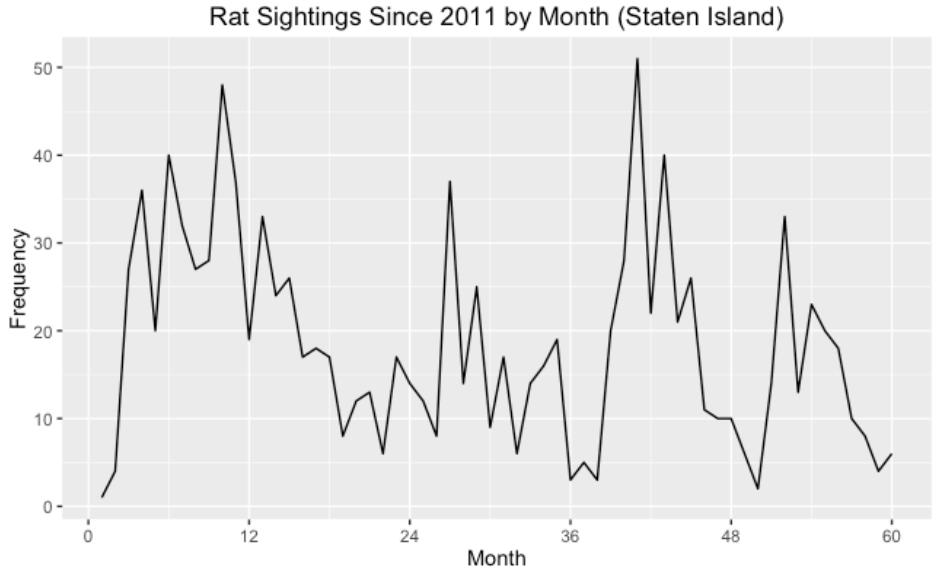
The following are line charts illustrating changes in rat sightings for each of the 5 boroughs. First, the overall trend:

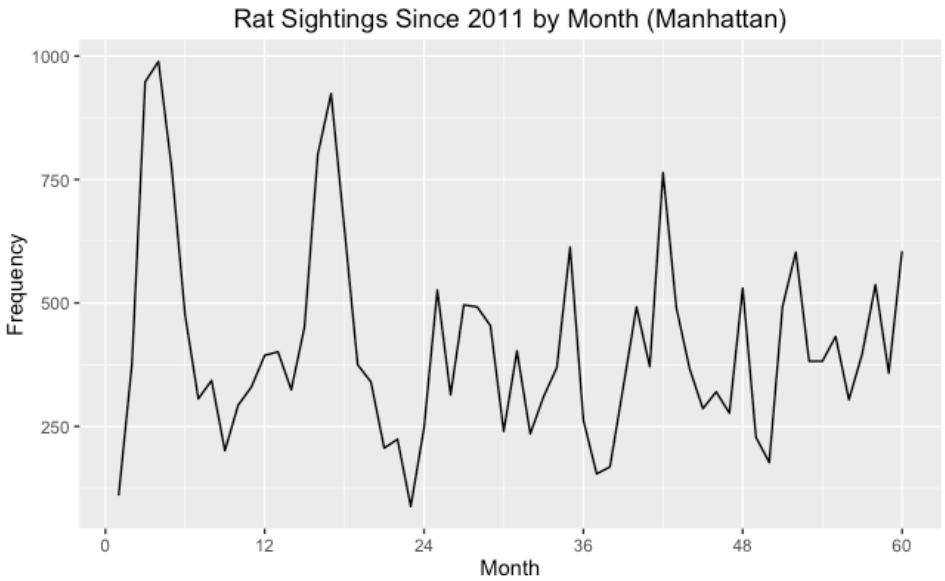


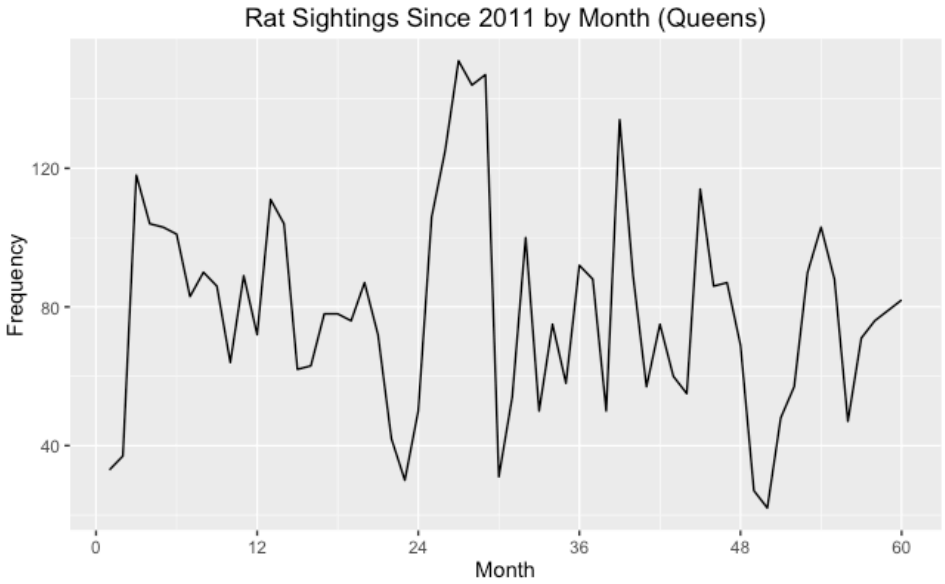
And now, by borough.


Rat Sightings Since 2011 by Month (Bronx)
Frequency
1000
750
500
250
0
12
24
36
48
60
Month


Rat Sightings Since 2011 by Month (Brooklyn)
Frequency
600
500
400
300
200
100
0
12
24
36
48
60
Month


Rat Sightings Since 2011 by Month (Staten Island)
Frequency
50
40
30
20
10
0
0
12
24
36
48
60
Month





Overall, rat sightings don't appear to have increased in the years since 2011. There may be a slightly upward trend in the Bronx and Manhattan, but it isn't obvious. Once important note is that rat sightings appear to be cyclical – they severely increase at the beginning of the year, and then steadily decrease. In almost all of the boroughs and in the overall graph, there are relative minimums towards the end of each year, and then a steady increase.

Hurricane Sandy hit in October of 2012, but doesn't seem to have led to major increases in rat sightings in each borough. In fact, in each of the boroughs, as in the trend identified above, rat sightings *decreased* towards the end of 2012. Various media outlets covered a supposed rat sighting increase after the hurricane, but the data

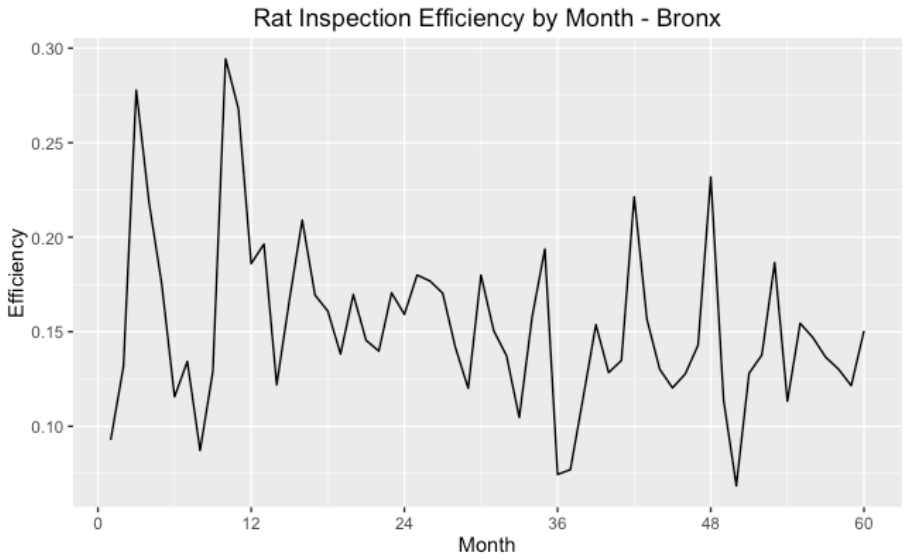
doesn't seem to validate that. In fact, the peak across all boroughs for rat sightings was actually in early 2011.

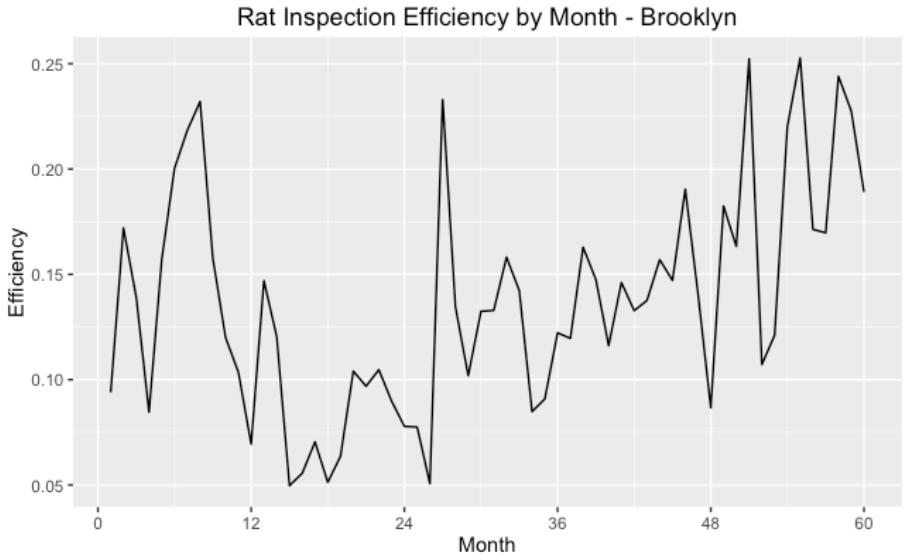
Question 1B.

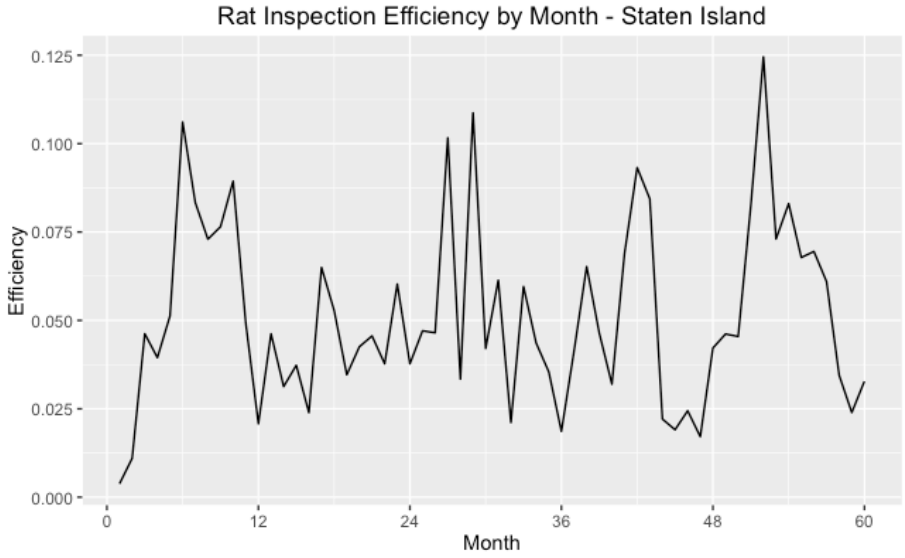
The following is a plot of rat inspection efficiency (sightings divided by total inspections) for the past 5 years.

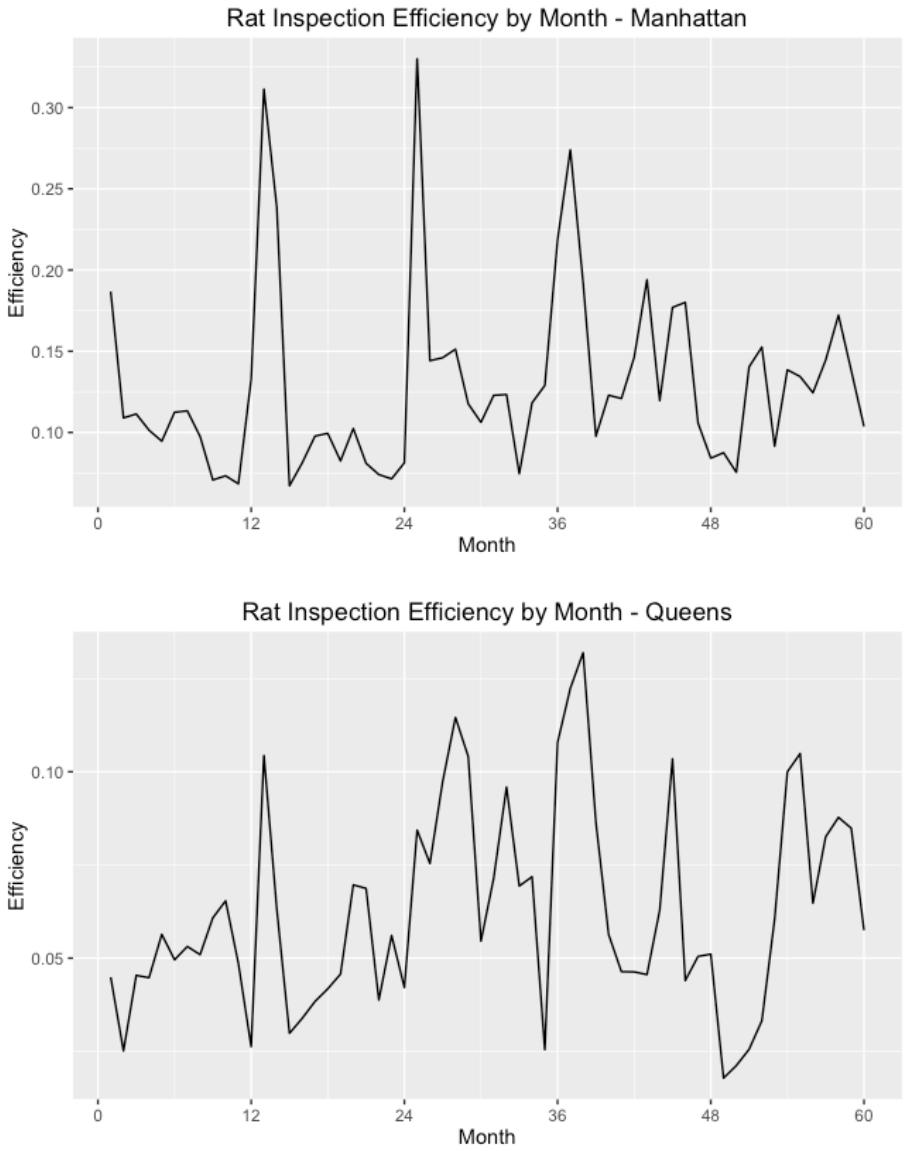
Interestingly, in the second half of 2012, efficiency decreased heavily, and at the beginning of 2013, it was at one of the lowest points in the entire data set. A possible explanation is that in anticipation of Hurricane Sandy and in its aftermath, inspectors increased activity – but as in the above graphs, there weren't that many rats around, so efficiency ratings were low.

Here are the borough specific plots.


Rat Inspection Efficiency by Month - Bronx
Efficiency
0.30
0.25
0.20
0.15
0.10
0
12
24
36
48
60
Month


Rat Inspection Efficiency by Month - Brooklyn
Efficiency
0.25
0.20
0.15
0.10
0.05
0
12
24
36
48
60
Month


Rat Inspection Efficiency by Month - Staten Island
Efficiency
0.125
0.100
0.075
0.050
0.025
0.000
0
12
24
36
48
60
Month



The trends we saw in the aggregate data seem to hold true in each of the boroughs – spikes at the beginning of the year.

Question 1C.

The following is a data frame of the top 10 zip codes with the largest number of rat yielding inspections.

ZIP_CODE <int>	RESULT <fctr>	count <int>
10457	Active Rat Signs	5563
10458	Active Rat Signs	5270
10456	Active Rat Signs	4098
10468	Active Rat Signs	3335
11221	Active Rat Signs	3102
10453	Active Rat Signs	2998
10452	Active Rat Signs	2364
10467	Active Rat Signs	2145
11237	Active Rat Signs	2111
11206	Active Rat Signs	1835

The top two zip codes by far are 10457 and 10458 – they've got more than 1000 more rat sightings than any of the trailing zip codes. These two zip codes border each other in the northern Bronx (Fordham University is actually in 10458). In fact, not only are the top *three* zip codes on the list all next to each other, but *seven* of the top 10 zip codes on this list are in the Bronx. The other three are all in Brooklyn.

Question 2.

The following is a data frame of the top 20 zip codes for active rat signs before Hurricane Sandy, followed by the same list *after* Hurricane Sandy.

ZIP_CODE <int>	RESULT <fctr>	count <int>
10457	Active Rat Signs	1028
10456	Active Rat Signs	955
10458	Active Rat Signs	890
11221	Active Rat Signs	647
10453	Active Rat Signs	595
11216	Active Rat Signs	534
10468	Active Rat Signs	509
11206	Active Rat Signs	428
11237	Active Rat Signs	413
10002	Active Rat Signs	407
11238	Active Rat Signs	383
10031	Active Rat Signs	367
10009	Active Rat Signs	365
10472	Active Rat Signs	363
11211	Active Rat Signs	363
11217	Active Rat Signs	356
10032	Active Rat Signs	346
10459	Active Rat Signs	337
10013	Active Rat Signs	302
10467	Active Rat Signs	277

And here are the top ones following the hurricane.

ZIP_CODE <int>	RESULT <fctr>	count <int>
10457	Active Rat Signs	4535
10458	Active Rat Signs	4380
10456	Active Rat Signs	3143
10468	Active Rat Signs	2826
11221	Active Rat Signs	2455
10453	Active Rat Signs	2403
10452	Active Rat Signs	2088
10467	Active Rat Signs	1868
11237	Active Rat Signs	1698
11206	Active Rat Signs	1407
10460	Active Rat Signs	1383
10009	Active Rat Signs	1364
10002	Active Rat Signs	1345
11216	Active Rat Signs	1299
10032	Active Rat Signs	1242
10025	Active Rat Signs	1185
10027	Active Rat Signs	1170
10029	Active Rat Signs	1161
10033	Active Rat Signs	1136
10031	Active Rat Signs	1073

On the high side of the distribution, the hurricane doesn't seem to have caused any significant displacement of the rat population. The top 10 as analyzed in the previous question is almost identical, and the order is fairly similar. However, there are some notable jumps. 10467 moved from #20 before the hurricane to #8 after, and 10460 moved to #11 from being totally off the pre-Sandy list. There was also a slight shift to Manhattan zip codes – pre-Sandy there were 5 Manhattan zip codes on the top 20 list, but after there are 8.

Another data set that could help answer this question is the database of Sandy 311 calls. Indeed, the leading zip code for 311 calls involving rat sightings was 10025, which jumped from non-listed pre-Sandy to #15 post-Sandy.

Incident.Zip <fctr>	Descriptor <fctr>	count <int>
10025	Rat Sighting	14
10456	Rat Sighting	7
11207	Rat Sighting	7
11237	Rat Sighting	6
11221	Rat Sighting	5
10027	Rat Sighting	4
10453	Rat Sighting	4
10458	Rat Sighting	4
10467	Rat Sighting	4
11222	Rat Sighting	4

That being said, the other top zip codes for 311 calls are mostly Bronx / Brooklyn areas that were already on the top-20 list before the storm.

Question 3A.

To weigh in on the debate about whether more rat sightings in local homes impact the amount of rat-related violations in restaurants, we can run a logistic regression against the efficiency or rat sightings. The following are the printed results from the regression.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6522  -0.6327  -0.6078  -0.5825   1.9528

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.476849   0.206360  -7.157 8.27e-13 ***
Efficiency           -0.008813   0.022568  -0.391 0.696156
INSPECTION_MONTH.f2  -0.027589   0.021153  -1.304 0.192160
INSPECTION_MONTH.f3  -0.097144   0.020496  -4.740 2.14e-06 ***
INSPECTION_MONTH.f4  -0.144680   0.020228  -7.153 8.51e-13 ***
INSPECTION_MONTH.f5  -0.224826   0.020480 -10.978  < 2e-16 ***
INSPECTION_MONTH.f6  -0.203091   0.020539  -9.888  < 2e-16 ***
INSPECTION_MONTH.f7  -0.175705   0.021205  -8.286  < 2e-16 ***
INSPECTION_MONTH.f8  -0.105870   0.020799  -5.090 3.58e-07 ***
INSPECTION_MONTH.f9  -0.075931   0.020522  -3.700 0.000216 ***
INSPECTION_MONTH.f10 -0.037330   0.020094  -1.858 0.063207 .
INSPECTION_MONTH.f11  0.009202   0.020994   0.438 0.661144
INSPECTION_MONTH.f12  0.006024   0.020198   0.298 0.765492
INSPECTION_YEAR.f2012 -0.022235   0.206232  -0.108 0.914141
INSPECTION_YEAR.f2013 -0.035387   0.205823  -0.172 0.863493
INSPECTION_YEAR.f2014  0.028098   0.205791   0.137 0.891396
INSPECTION_YEAR.f2015  0.026271   0.205801   0.128 0.898424
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 402955  on 437434  degrees of freedom
Residual deviance: 402504  on 437418  degrees of freedom
AIC: 402538

Number of Fisher Scoring iterations: 4
```

Not only is the coefficient of the Efficiency variable incredibly low, but the P value is sky high – we can be almost entirely sure that there's no significant relationship between the Efficiency predictor and whether or not there was a rat-related violation in area restaurants.

If we were to infer anything from these results, the coefficient is negative: that means that the higher the efficiency rating is, the more likely that there won't be rat-related violations in area restaurants. However, as noted above, the results are not statistically significant.

Another important note is that Efficiency may not be a good proxy variable for this regression – the fact that inspections were less *efficient* doesn't necessarily mean that there aren't more rats around. Another variable that could be more useful is a weighted value for the amount of actual rat sightings. But in general, the results of this

regression would lead to the recommendation that efficiency ratings are *not* a good predictor variable for rat-related violations in area restaurants.