

Homework 1: Rats!

Part 1: Descriptive Statistics and Figures

- a) We can see that rat sightings have remained relatively steady after a peak in late 2009. There is clear seasonal cyclicity, with rat sightings dropping during the winter. There seems to be a decrease in sightings immediately after Hurricane Sandy. In Manhattan, there does seem to be a decrease in sightings over the next year. However, sightings have slowly been normalizing and there doesn't seem to have been a major increase.

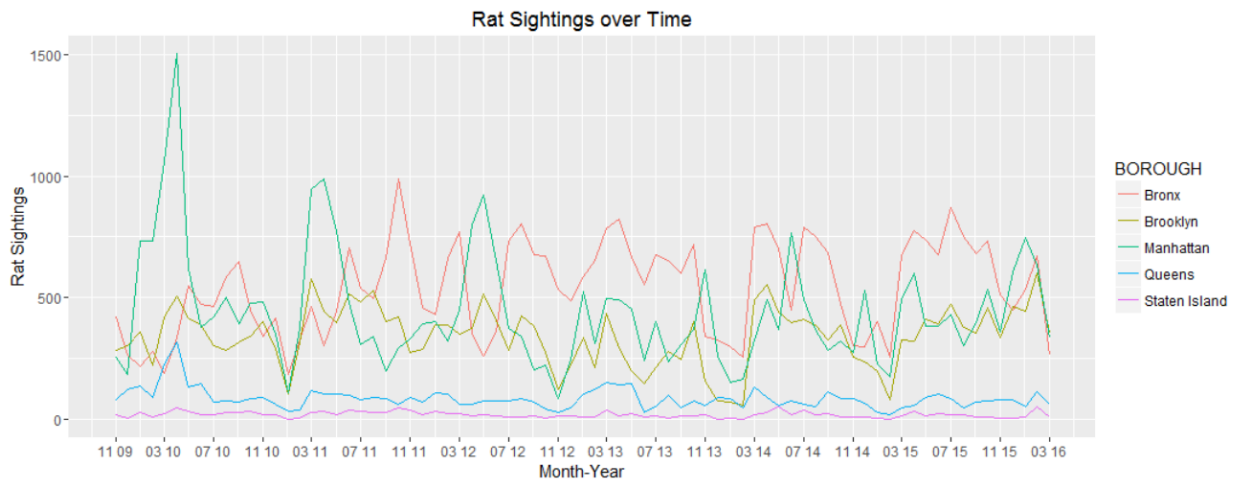


Figure 1: Rat Sightings from Nov. 2009 - Mar. 2016

- b) An interesting insight from this graph is that despite total rat sightings going down each winter, the overall efficiency of rat inspections goes up. This suggests that in the winter time, rats are more prone to show signs during inspections.

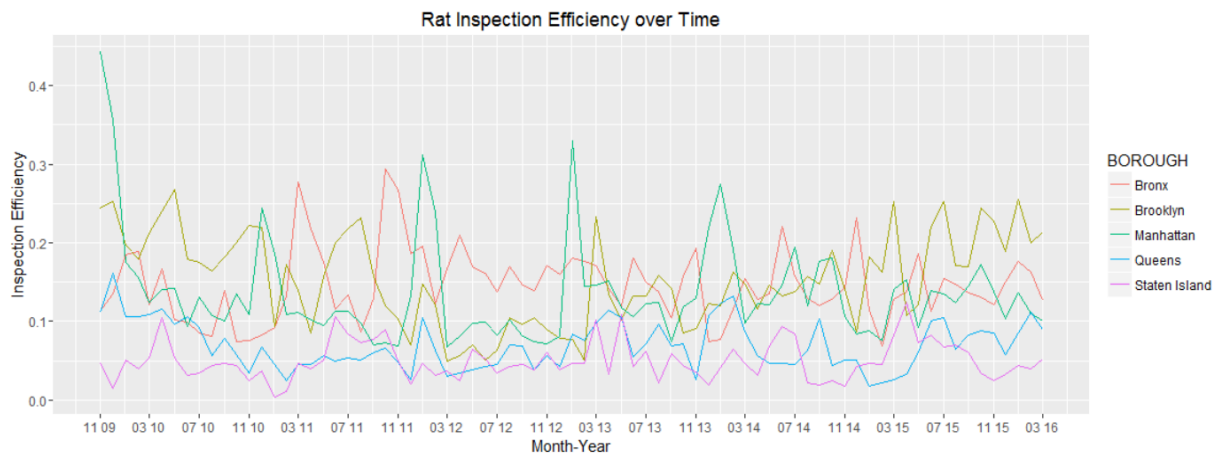


Figure 2: Inspection Efficiency from Nov. 2009 - Mar. 2016

- c) It seems that the zip codes with the most sightings are in the Bronx and in certain areas of Brooklyn.

	ZIP_CODE	BOROUGH	freq
1	10457	Bronx	6717
2	10458	Bronx	6205
3	10456	Bronx	4788
4	10468	Bronx	3952
5	11221	Brooklyn	3834
6	10453	Bronx	3514
7	11237	Brooklyn	2974
8	10452	Bronx	2821
9	10467	Bronx	2494
10	10009	Manhattan	2387

Figure 3: Top 10 Zip Codes for Rat Sightings

- d) We see that inspections cover much of the five boroughs.

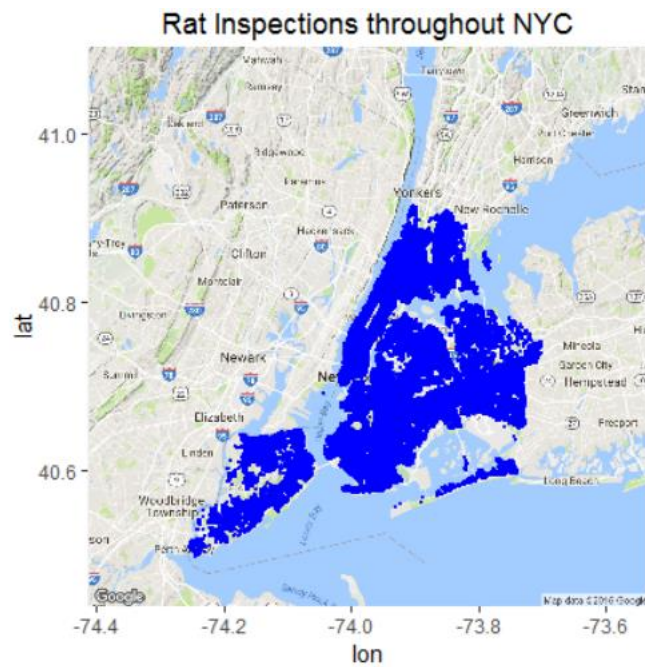


Figure 4: Rat Inspections by Geography

Part 2: Geographic Patterns Before and After Hurricane Sandy

- a) I do not believe Hurricane Sandy had a lasting impact on rat distribution because the top zip codes both before and after Sandy are relatively similar. The Sandy database is too limited to draw any real inferences about distribution during the Hurricane. It is likely that people were more preoccupied with reporting storm damage rather than rodent sightings during that period.

	ZIP_CODE	freq
1	10457	1771
2	10458	1541
3	10456	1457
4	11221	1218
5	10468	991
6	10453	900
7	11237	869
8	10031	851
9	10002	830
10	10009	827
11	11206	782
12	10472	773
13	10459	763
14	10032	761
15	11216	705
16	11238	704
17	10013	687
18	11211	651
19	11217	641
20	10027	598

Figure 5: Pre-Sandy Top Zip Codes

	ZIP_CODE	freq
1	10457	2355
2	10458	2195
3	10468	1521
4	10456	1433
5	10453	1385
6	10452	1158
7	11221	1070
8	10467	892
9	10033	775
10	11206	772
11	11237	663
12	10032	659
13	10460	649
14	10472	632
15	11385	627
16	10027	619
17	10002	583
18	10025	571
19	10009	563
20	11216	562

Figure 6: Post-Sandy Top Zip Codes

	Incident.Zip	freq
1	10025	14
2	10456	7
3	11207	7
4	11237	6
5	11221	5
6	10027	4
7	10453	4
8	10458	4
9	10467	4
10	11222	4
11	11368	4
12	10014	3
13	10022	3
14	10023	3
15	10031	3
16	10452	3
17	10460	3
18	11201	3
19	11212	3
20	11216	3

Figure 7: Sandy Top Zip Codes

Part 3: Rodent and Restaurant Inspection

```
glm(formula = RAT_VIOLATION ~ ActiveRatSightings + MONTH + YEAR,
     family = binomial(link = "logit"), data = restaurant_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.6562	-0.6311	-0.6100	-0.5854	1.9706

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-8.435451	17.929324	-0.470	0.63801
ActiveRatSightings	-0.072964	0.025014	-2.917	0.00353 **
MONTH2	-0.033616	0.018287	-1.838	0.06603 .
MONTH3	-0.092789	0.018573	-4.996	5.86e-07 ***
MONTH4	-0.150625	0.019409	-7.761	8.45e-15 ***
MONTH5	-0.220358	0.019674	-11.201	< 2e-16 ***

MONTH6	-0.201618	0.019766	-10.200	< 2e-16	***
MONTH7	-0.178931	0.020421	-8.762	< 2e-16	***
MONTH8	-0.106843	0.020009	-5.340	9.31e-08	***
MONTH9	-0.080048	0.019728	-4.058	4.96e-05	***
MONTH10	-0.039093	0.019256	-2.030	0.04234	*
MONTH11	0.006169	0.020227	0.305	0.76039	
MONTH12	0.003697	0.019400	0.191	0.84888	
YEAR2011	6.989091	17.930501	0.390	0.69669	
YEAR2012	6.950398	17.929325	0.388	0.69827	
YEAR2013	6.941932	17.929319	0.387	0.69862	
YEAR2014	7.002102	17.929319	0.391	0.69614	
YEAR2015	7.003069	17.929319	0.391	0.69610	
YEAR2016	6.925930	17.929324	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: 421831 on 456641 degrees of freedom
 Residual deviance: 421385 on 456623 degrees of freedom
 AIC: 421423

Number of Fisher Scoring iterations: 7

Based on this regression, it seems that with lower rates of rat sightings, there is a higher chance of a rat violation in the restaurant. It is interesting to note that rat violations increase in November and December (although these p-values are small). However, because the coefficient for this is quite small (.073), I would not say it is too useful in predicting rat violations in restaurants. There would likely be several other restaurant-specific variables that would affect the likelihood of rats.