**ABSTRACT** 100-250 words

**INTRODUCTION**

Why is this an important question

Set the reader for what’s coming next

Rats have recently been described by Mayor Bill de Blasio as “one New York City institution that we’re happy to get rid of” (1). Residents seem to agree, based on the rising number of rodent-related complaints made to the city’s 311 service in the past few years. Since 2010, there have been close to 140,000 complaints of such type, which include either the sighting of a rat or conditions that attract the animal (2).

The actual dimension of the issue is even greater, as subway complaints are redirected to Metropolitan Transportation Authority and are not recorded by 311 (2). In late September 2015, the video of a rat carrying a whole slice of pizza down the stairs of a subway station, which has been watched over 8.5 million times to date, triggered a heated debate over the cleaning of subway stations (3).

Rodent-related complaints are particularly relevant from a public health standpoint. New York City rats often carry pathogens that may lead to a wide range of diseases, some of which are life-threatening (4). Besides, the exposure to rats’ hair, droppings and urine are linked to an increased risk of allergies and asthma. (4) The damage caused by rats are not limited to the human health, however, as there are reports of rats eating out the electrical cables of cars, for instance (4).

It comes as no surprise, therefore, that the New York City budget agreed upon in June 2015 includes $2.9 million in resources to address the issue, mostly by targeting the so-called rat reservoirs, where there tends to be a concentration of the animal (1). These efforts follow other plans from nearly every of the city’s other 108 mayors, none of which succeeded on this front (1). The New York Times has reported on the city’s struggle against rats at least from as early as 1860 (2).

The subject matter allows some interesting explorations with available open data related to New York City, which will be developed in the final paper. The backbone of the work will be rodent-related complaints to the city’s 311 service, with the analysis focusing on three major fronts:

**Relationship between the amount of complaints and the socioeconomic profile of each city area:** Do poverty-stricken areas tend to complain more about rats than wealthier ones on a per capita basis?

**Evolution of complaints over time:** Has any of the campaigns employed by former New York City mayors have any impact on the amount of rodent-related complaints? Is there any seasonality in these complaints?

**Complaint behavior of city residents:** How do complaints of actual rat sightings relate to complaints about conditions attracting rodents in a given area? In other words, does the perception of an environment favoring rats correspond to the actual presence of the animals?

**LITERATURE REVIEW**

Explain research that’s been done before

Gaps/limitations

A key limitation in this work resides in the potential gap between the actual issue (i.e. rats) and the propensity of city residents to complain about it. It is possible, for instance, that the rats issue has not actually been aggravating in the past few years, as reported by the news (X), but actually that the propensity of people to complain about it has become higher. In any case, residents’ complaints are most likely the best quantifiable proxy for how the issue has evolved. Besides, there is undeniable value in studying the complaints *per se*, as they tend to reflect the collective mood of the city.

**DATA METHODS**

**Data**

Rat-related complaints filed in the New York City’s 311 Service were obtained from the NYC OpenData portal (xx). Between January 1, 2010 and December 9, 2015, there were nearly 140 thousand complaints of such type, which are divided into “rat sightings” (93 thousand), “condition attracting rodents” (39 thousand) and “signs of rodents” (8 thousand). Besides this “descriptor”, the fields extracted from each complaint log were “created date” and “incident ZIP”. The complaint logs were grouped both by ZIP code (“total complaints”, “rat sightings” and “condition attracting rodents” and by date (year and month).

The median household incomes per census tract for the years of 2010 and 2014 were obtained from the American Community Survey, which is conducted by the U.S. Census Bureau (xx). Each census tract was assigned a ZIP Code Tabulation Area (ZCTA), based on the U.S. Census’ 2010 relationship file (xx). Each ZCTA was then assigned a ZIP code, based on a UDS crosswalk file (xx). The household income for each ZIP code, divided into the two years, was calculated as an average of the household incomes of all census tracts pertaining to that ZIP code.

The population of each ZIP code was obtained from the NYC Department of City Planning (xx), which grouped it from the 2010 U.S. Census.

The year in which each property was built was extracted from the NYC Department of City Planning’s MapPLUTO (xx), which gathers land use and geographic data at the tax lot level. The extracted data were restricted to the boroughs of Manhattan and Brooklyn, on the back of computational limitations. The series was grouped through the calculation of the median for each ZIP code.

These indicators at the ZIP code level, namely rat-related 311 complaints, median household incomes for 2009 and 2014, population and average year of construction, were all grouped in a single data frame, which is the base for most numerical analysis in this paper. Based on this data, additional indicators were calculated, such as complaints per 1,000 people and

This data frame was merged into a shapefile of New York City ZIP codes, obtained from NYCPrepared (xx), for visualization purposes.

**Statistical analysis**

The Python programming language was used in all data analysis and visualizations, which were conducted in an IPython notebook publicly available in a GitHub repository (xx). Data manipulation was mostly conducted with the Pandas and Numpy packages.

Heat maps for the total number of rat-related complaints and its two main categories (“rat sightings” and “condition attracting rodents” were developed with the matplotlib Pyplot library.

Linear and second-degree regressions were estimated for (all following series were used on a per ZIP code basis): total rat-related complaints from 2010 to December 2015 and median household income (average of 2009 and 2014); total rat-related complaints per 1,000 people from 2010 to December 2015 and median household income (average of 2009 and 2014); total rat-related complaints per 1,000 people from 2014 to December 2015 and median household income in 2014; total rat-related complaints per 1,000 people from 2014 to December 2015 and growth of median household income in the 2009-2014 period; and total rat-related complaints per 1,000 people from 2010 to December 2015 and median building year of local properties. All regressions were conducted with the Statsmodel package.

The most noticeable results from the regressions observing exclusively the R-squared values, meaning that they explained the largest percentage of the variation in the independent variable, were achieved when (1) total rat-related complaints per 1,000 people is regressed on the median building year of local properties (first-degree adjusted R-squared of X% and second-degree of X%) and (2) when total rat-related complaints are regressed on median household income (first-degree adjusted R-squared of X% and second-degree of X%). The p-value for both independent variables were below 1%. These independent variables also have negative coefficients in common, pointing to a negative correlation between rat-related complaints and both the median building year of local properties and the median household income of each ZIP code.

The use of total complaints in the second regression, however, is problematic given the fact that ZIP codes with a low median household income tend to be more populated(X), which in itself accounts for part of the more numerous complaints. This is addressed with the use of total complaints per 1,000 people, leading to significantly lower adjusted R-squared values in the regressions on the median household income (of X% for the first-degree and of X% for second-degree). The p-value for the remained below 1% for the independent variable under this scenario.

The regression of total complaints per 1,000 people in the past two years on the median household income growth for each ZIP code in the past few years, yields higher adjusted R-squared values (of X% for the first-degree and of X% for second-degree) than the ones observed when only the median household income for one year is considered (first-degree adjusted R-squared of X% and second-degree of X%). The coefficient of the independent variable differ in both their dimension (X in the first regression and X in the second) and sign (positive in the first regression and negative in the second).

**RESULTS**

The socioeconomic dynamics of a given neighborhood tend to impact the amount of rat-related complaints coming from this area.

**Reference:**

1. Flegenheimer, Matt. "New York City Escalates the War on Rats Once Again." *The New York Times*. June 24, 2015. Accessed November 30, 2015. http://www.nytimes.com/2015/06/25/nyregion/new-york-city-escalates-the-war-on-rats-once-again.html.
2. Goldberg, Barbara. "New Yorkers Complain: 'The Rats Are Taking Over'" *The Huffington Post*. Accessed December 1, 2015. http://www.huffingtonpost.com/entry/new-york-rat-problem\_5629452fe4b0aac0b8fc3c24.
3. Fitzsimmons, Emma. "‘Pizza Rat’ Prompts a Collective ‘Ew’ and Debate on Cleaning New York Subway." *The New York Times*. September 22, 2015. Accessed December 1, 2015. http://www.nytimes.com/2015/09/23/nyregion/pizza-rat-spurs-debate-on-how-to-clean-up-new-yorks-subway-system.html.
4. "Rats in New York City." *Wikipedia*. Accessed December 1, 2015. https://en.wikipedia.org/wiki/Rats\_in\_New\_York\_City.

Walsh, MG. "Rat sightings in New York City are associated with neighborhood sociodemographics, housing characteristics, and proximity to open public space." Peerj 2, (n.d.): Science Citation Index, EBSCOhost (accessed November 30, 2015).

"Evaluation of a neighborhood rat-management program - New York City, December 2007-August 2009." MMWR: Morbidity & Mortality Weekly Report 61, no. 37 (September 21, 2012): 733-736 4p. CINAHL Plus, EBSCOhost (accessed November 30, 2015).

Auerbach, Jonathan. "Does New York City really have as many rats as people?" Significance 11, no. 4 (October 2014): 22-27. Health Policy Reference Center, EBSCOhost (accessed November 30, 2015).

Frye, M J, et al. "Preliminary Survey of Ectoparasites and Associated Pathogens from Norway Rats in New York City. "Journal of Medical Entomology 52, no. 2 (March 2015): 253-259. MEDLINE with Full Text, EBSCOhost (accessed November 30, 2015).

Firth, Cadhla, et al. "Detection of zoonotic pathogens and characterization of novel viruses carried by commensal Rattus norvegicus in New York City." Mbio 5, no. 5 (October 14, 2014): e01933-e01914. MEDLINE with Full Text, EBSCOhost (accessed November 30, 2015).