MONTGOMERY COUNTY, MD: COUNTING OUR TOWS

Everyone gets annoyed when they find out their car has been towed, but most people understand it is a necessary service in order to keep cars from blocking fire lanes, and to ensure parking is available in residential lots for those that live there and their guests. We expect towing counts to be higher in more populated areas, but in general, assume that average rates and distribution among the community is the same for all county residents. The dataMontgomery website provides a starting point to look into these assumptions.

THE DATA: INGESTION AND WRANGLING

The main dataset examined is at dataMontgomery is the “Trespass Towing Report” found at: <https://data.montgomerycountymd.gov/Consumer-Housing/Trespass-Towing-Report/i6vn-3s6e>. The data contains each instance of towing in Montgomery County that is reported to the police department; by regulation, all non-owner-initiated cars that are towed must be reported by the towing company so that the police do not waste time taking “stolen” car reports from frantic citizens. Variables of interest include:

* Storage Company: name of towing company
* Location: street address where vehicle was towed from
* City
* Geo-Location: point coordinates for mapping where the vehicle was towed from

The other data is from the 2019 American Community Survey (ACS) available at the US Government Census Bureau API (<https://api.census.gov/data/2019/acs>) and the US Government Federal Communications Commission API (<https://geo.fcc.gov/api/census/area>).

From the former, I pulled variables at the census block level for:

* Average Household Income (for tract+block #)
* Total Population (for tract+block #)
* Total Black Population (for tract+block#)

I used the latter to add block numbers (tract+block#) to my dataMontgomery set; this enables me to join the dataMontgomery set to the census data. Block number becomes an additional variable.

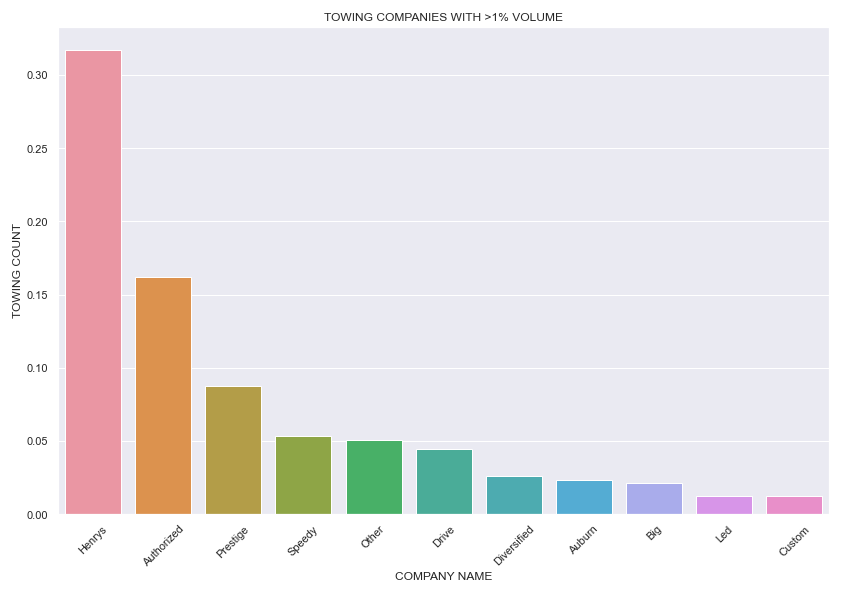
Finally, I added two columns to give a percentage rate within each census block:

* Tow Rate: Tow count/total population within block
* Black Rate: Black population/total population within block

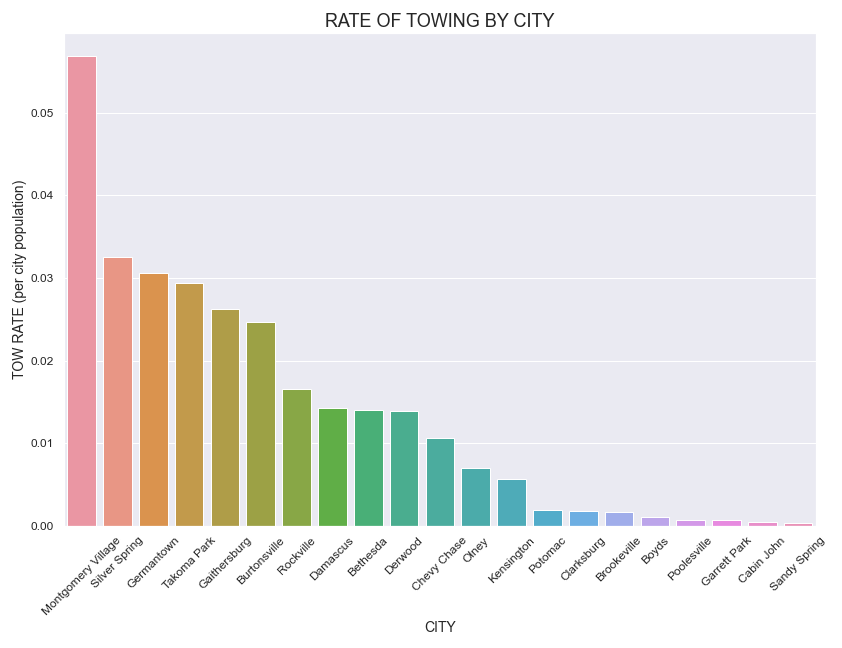
In this context, “total population” of the County means the population sums of the blocks included in the towing dataset (470 of over 600 existing blocks)

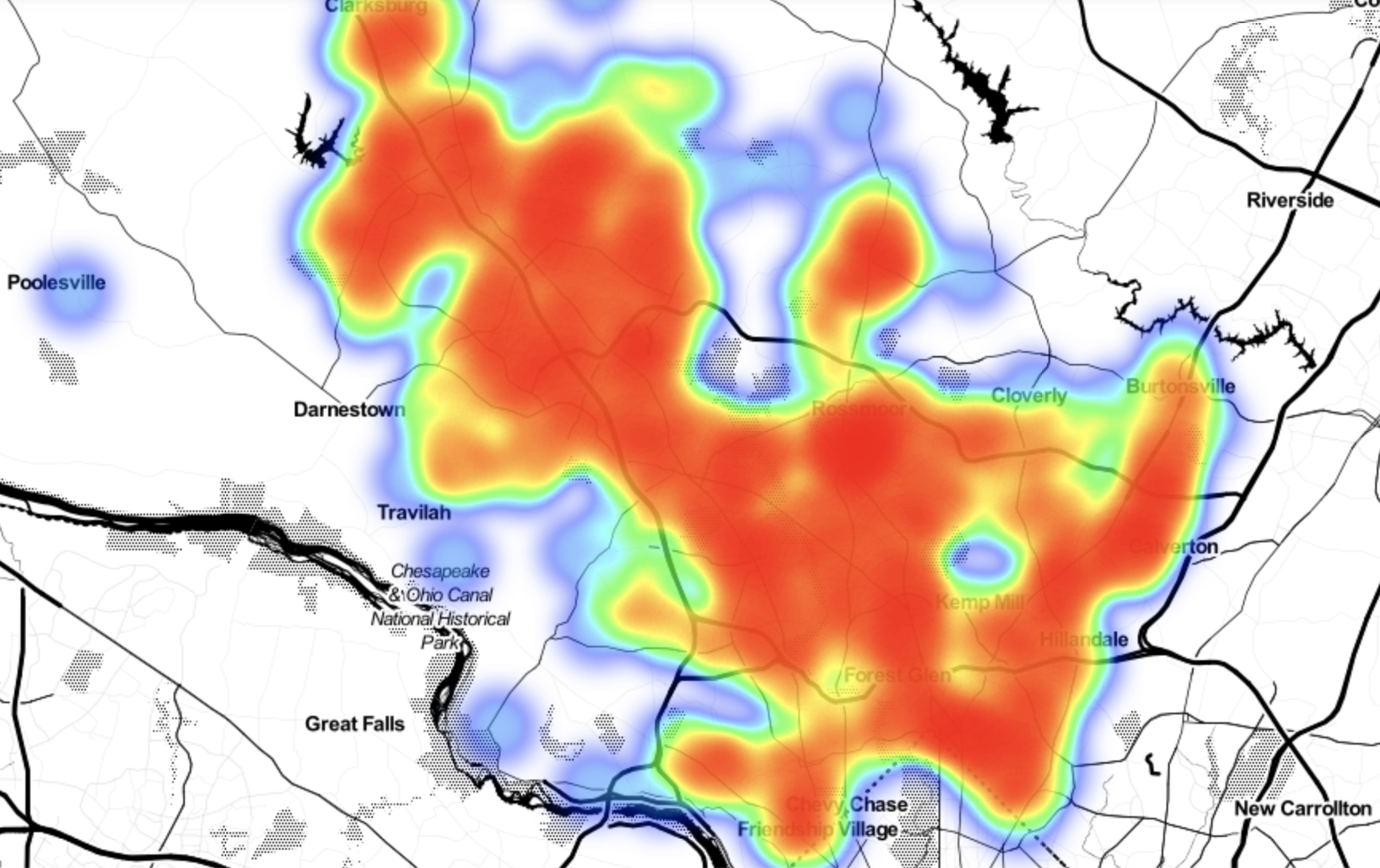
EXPLORATORY DATA ANALYSIS

After doing extensive cleaning of the company names by string-matching in python, a cursory view of the data was obtained by merely obtaining counts and presenting them in a bar chart. For the companies, I looked at those with frequency over 1%:



Next, I began looking at towing volume by city. The following chart shows the towing rates taking block populations into account. So far, it looked like Montgomery Village is a standout, but I do an analysis of the top 5 cities (Montgomery Village, Germantown, Gaithersburg, Rockville and Silver Spring) against each other and also as one group compared to all others outside of these 5 as a group to see if towing rates seem equitable and if there appears to be any relationship with income and/or race in Montgomery County.



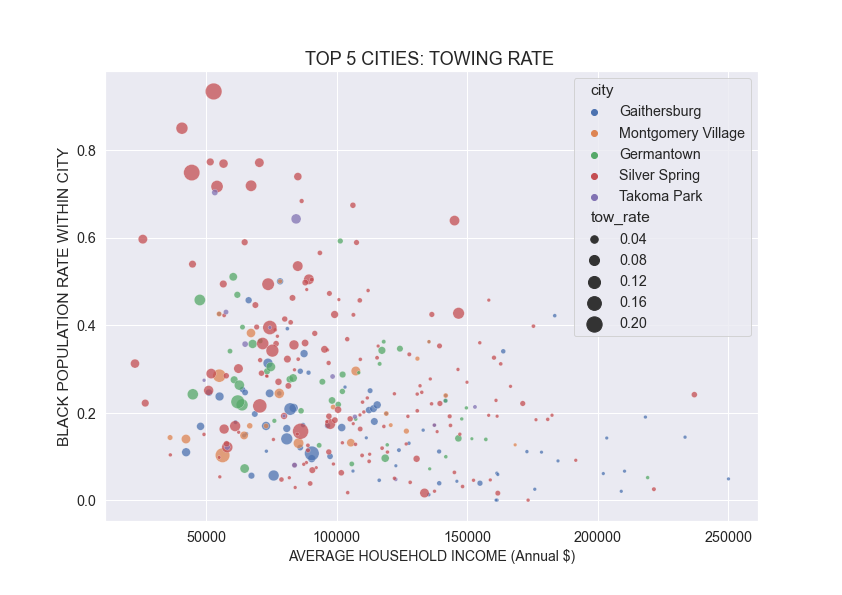


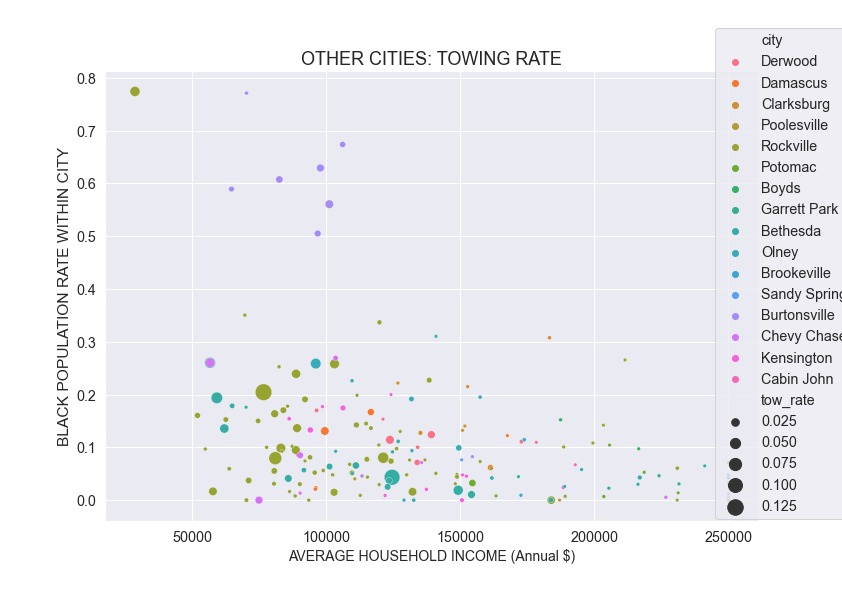
To begin, we look at a heatmap of the original towing dataset as a whole. The rainbow colors indicate lowest towing counts at cool colors of purple, blue and green proceeding up through yellow, orange, then red at the top rates. It’s no surprise that less populous regions indeed show cooler colors, while most large towns and cities have many bright red areas. Parking obviously becomes more of an issue in more crowded areas. Nothing really seems amiss from this map.

STATISTICAL ANALYSIS

I began by hypothesizing the mean and median of the towing rate is not significantly different across the county when population is taken into account. Further, when we examine the distribution of the data, we hypothesize the distribution will be the same across the county.

First, I split the data into two groups: Top 5 Cities and Other Cities. To get an initial look at the demographic variables of income and black population proportion, I produced the two charts below. From these charts, it becomes apparent that lower incomes correlate with higher rates of towing. It’s also clear that race is having some effect, although it’s more complicated to separate out; however, I can see that there are many more areas with the proportion of blacks over 30% in the Top 5 Cities, and some of those are experiencing the highest rates in the County.





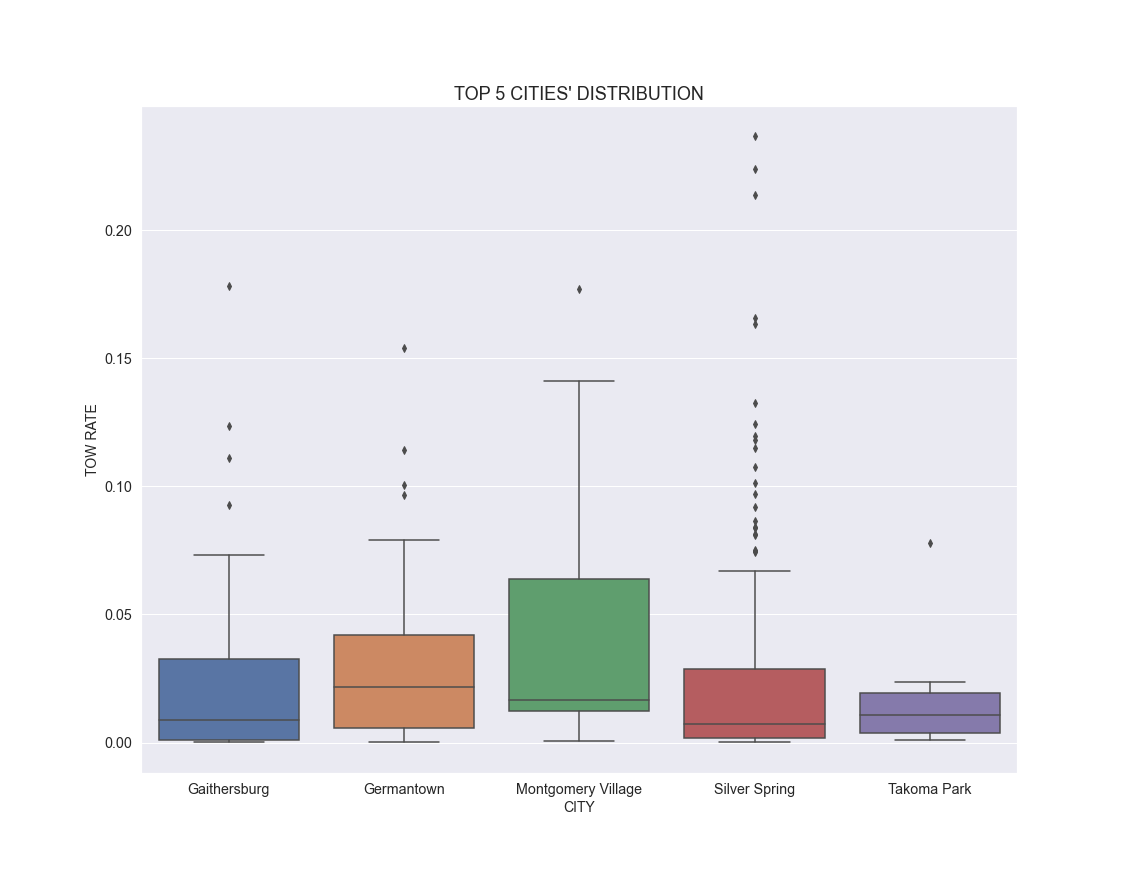
For comparing the distributions of the Top 5 cities group, I did paired testing of each city to each other city. Testing was performed using the Wilcoxon-Mann-Whitney test, a nonparametric test (due to non-normality verified by the Shapiro-Wilkes test and QQPlot):

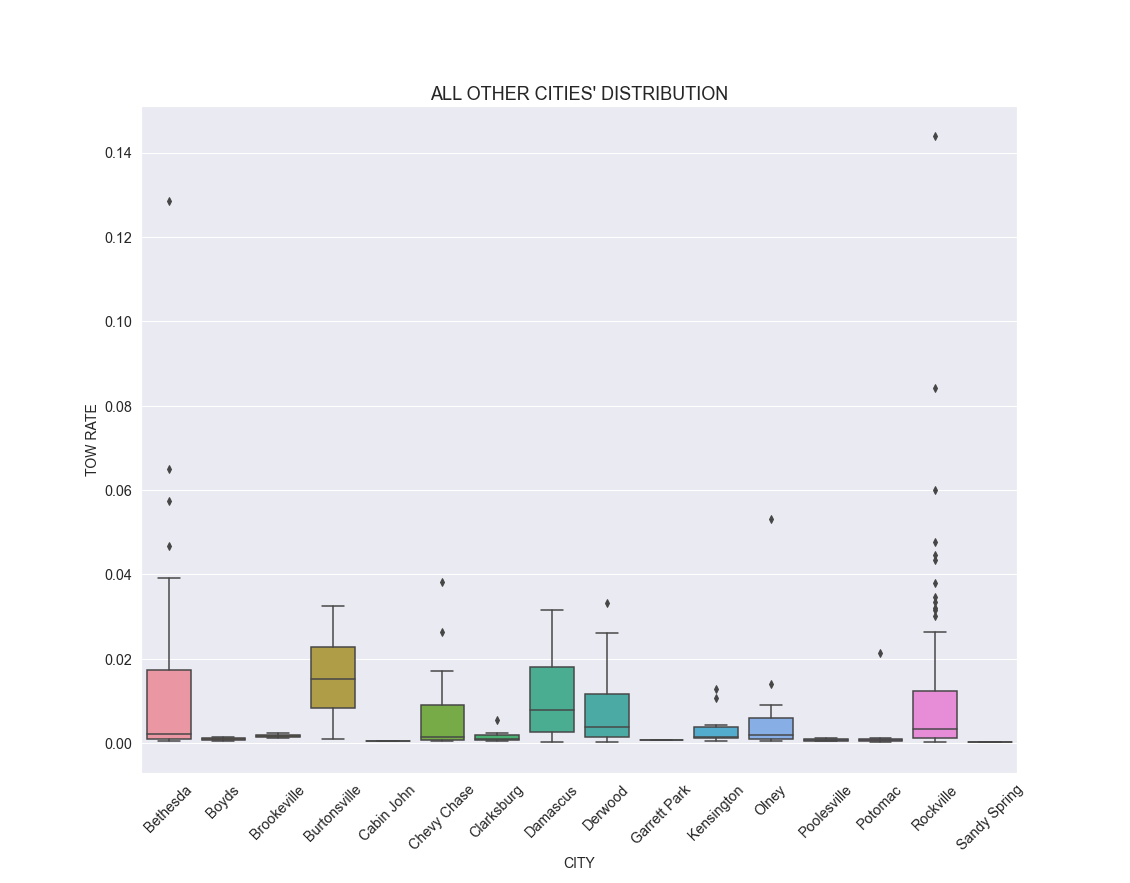
H\_0: Distributions between the two cities’ tow rates are the same

H\_alpha: Distributions between the two cities’ tow rates are different

With alpha=0.05, we cannot reject the null for the Gaithersburg-Takoma Park-Silver Spring combinations (p>0.05 ) and the Montgomery Village-Germantown pairings, so distributions may not be different in these two match-ups; all other pairings between the Top 5 cities group have p<=0.05; therefore, all others appear to have significantly different distributions of towing rates. I also compared the aggregated Top 5 cities and Other cities groups: p<=0.05, so we reject the null there as well and have 95% confidence in significantly different distributions between the two groups, Top5 and Other.

Box plots below seem to confirm these conclusions.





What really jumps out are the number of what appear to be outliers, particularly in Silver Spring. Sometimes these are errors in data entry, but that is doubtful here. In this case, we want to pay special attention to the outliers, because they represent an area in the county that appears to be significantly more impacted by high rates of towing.

I constructed a logistic regression model using the 2 features of income and black proportion to predict the towing rate. Using the IQR upper fence calculation cutoff of .0592, I coded the classes of towing rate as 0 for values under .0592 and as 1 for values over .0592. Using the scikit.learn package, I ran the regression model the first time on imbalanced data. Because the data needed rebalancing, I implemented the SMOTE method using the imblearn package (I had to switch to Google Colab because Anaconda would not recognize the imblearn package). The balanced data run gave good results for recall, the True Positives, which is the at-risk group:

Accuracy = 88%, Precision = 36%, Recall = 92% and ROC Score of 87% were the balanced.

Two methods were used to calculate a cutoff value for tow rate outliers:

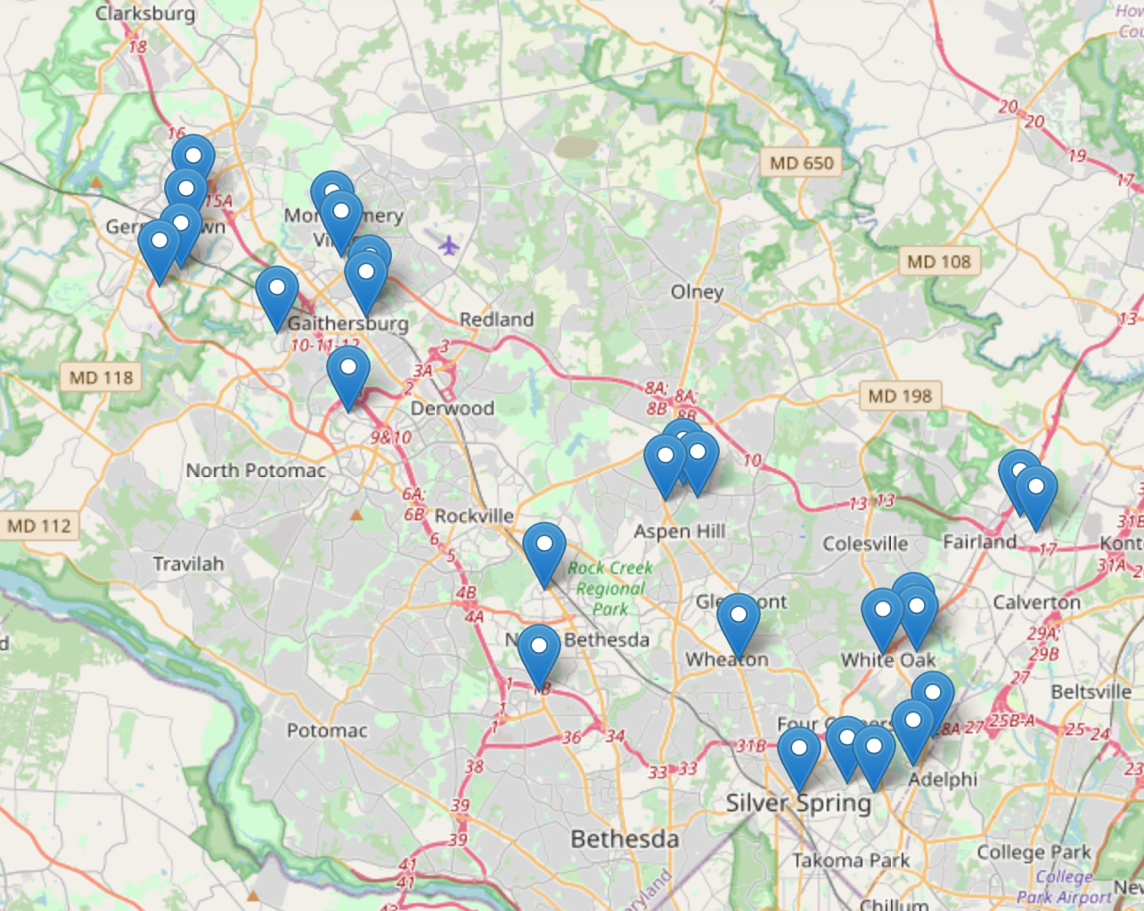
* Upper fence of 0.0529, was calculated using 1.5 x IQR added to the 3rd quartile cutoff.
* Generalized Extreme Test for Outliers\* (Grubb/Rosner): this test gives us cutoff values for extreme outliers which are usually wider than the IQR fences. The cutoff was calculated to be 0.0919. This test requires approximate normality so this cutoff may

not be valid; however, it’s greater than the IQR cutoff calculated rate, so if nothing else, these outliers are the most extreme.

I used the 0.0919 cutoff to produce a map of the outliers, shown as a static image here. To get an interactive image I saved it to an html file as well with a tooltip for each marker.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\*https://github.com/bhattbhavesh91/outlier-detection-grubbs-test-and-generalized-esd-test-python



link: map\_GESD.html

RECOMMENDATIONS

* Data Entry: incorporating spellcheck or a choice menu for names would improve data quality.
* Montgomery County could use the dataMontgomery website to investigate properties experiencing excessive towing using the map produced here.
* Consider policy and towing law, particularly contract towing in apartment and condominium developments. Change the law to be more tenant friendly because some of our most financially vulnerable citizens are experiencing the cost and added stress of the practice as it stands. We must also be on guard against systemic racism in these practices.

ACKNOWLEDGEMENTS

Special Thanks to:

Professor Michael Iapalucci

Professor Abdirisak Mohamed

Professor Rachel Saidi

Professor Celia Evans

Ms. Victoria Lewis

Mr. Hudson Finch-Batista (FCC API assistance)

All my fellow students