Pollution Mapping:  
Montgomery County

By Miguel Mayo in partnership with Data Montgomery

Montgomery College, Rockville MD

**Data:**

In partnership with Montgomery County, all the data sets used were taken from Data Montgomery. The data is available to the public through their website. I chose to use the **Water Stream Quality Map** as the foundation of my project and mapping. I then chose three other data sets that had geolocation data and that recorded data on what could plausibly be a source of pollution.

1. **Water Stream Quality**: Map of Montgomery County divided into its watersheds and sub watersheds. It then color coats them by their overall stream score. The stream score is based on a mixture of water quality, fish health, and benthic health. Also contains data on which tributaries feed into which watersheds and other bodies of water.
2. **Reported Illegal Dumping**: Map Displaying all reported cases of illegal dumping in the county since 2007. Tracks location, type, and subtype. It also tracks the open and close date for each case.
3. **Reported Sanitary Sewer Overflows**: Data Frame containing open and close dates of reported Sanitary Sewer Overflows. Contains a large amount of geographic data but it is unusable due to the input type for the data collection.
4. **Rainscapes Rewards Program**: Map displaying the locations of County funded rainscapes projects**.** Shows the project type, geolocation, and how much water was treated. The project also tracks how it was installed, and for some cases, more details about the trees used.

**Insights:**

The initial project idea was to construct a large shared interactive map to attempt to visually correlate the cases with tributaries and watersheds. Due to project size, time, and data limitations, the data sets were examined individually using location as a common component. After running several correlation test and a chi- square test on the data, its variables displayed little correlation suggesting external factors. The only variable with a strong influence was latitude, but this is explained by downstream effect

To me the most logical source of water pollution out of my data was the sanitary sewer overflow set. To my surprise however, sewer overflows are practically nonexistent compared to what they used to be in 2007. It should be noted that in almost every year, March had the highest number of overflows, usually by a good lead.

The next dataset that jumped out at me was the Rainscapes Rewards Program because there appeared to be a lot of pins around the same areas that were marked yellow or red on the water quality map. Of the types of rain scape projects in the county, the most popular type was the rain barrel with 746 constructed treating 170,585 gallons. Despite over 2000 recorded projects in the county, and one project single handedly treating 14,000 gallons, chemicals and an excess of bio organic compounds are making it into the water. A clear geographical correlation between the rain barrel locations and the unhealthy water samples appeared in the data. This can likely be contributed to the geo-economic dispersion in the County.

The final data set was illegal dumping. Like the Rainscapes Data, there was a clear concentration in the south. Filtering them by city, and even zip code, Silver Spring was consistently the largest contributor of illegal dumping each year. Surprisingly, when done by zip code the second largest contributor of illegal dumping was Leisure World, a retirement community off Georgia Ave. In 2016 under 50 cases were reported. The next year it skyrocketed to around 300 every year since. Using k-means clustering, we identified three main clusteroids that captured most of the data. When re arranging the data into these clusters the percentage contributed by each city changes drastically. Two of the clusters are much tighter than the other and were centered around Wheaton and Montgomery village.

**Operationalization**:

1. If the stream quality data were updated, a great in depth study could be done into the true effectiveness of the Rainscapes Rewards Program with before and after comparisons. One could also look into which system treats the most effectively and which is the most efficient to treatment efficient.
2. Further and more focused study could be conducted on the illegal dumping data set. One could combine it with other data sets and try to break down factors that contribute to illegal dumping. Census data would be a great companion for the set. One could examine if there’s more reported dumping in communities with higher than average concentrations of certain ethnicities. Socio economic factors could also be examined to see how strong a correlation, if any, existed.
3. The data set on sanitary sewer overflows would grow immensely more interesting and useful if a the desc column were to have some sort of format for reporting. The descriptions varied in the order and amount of data given in this column. A large amount of geographical data was inaccessible to mapping engines because of how it was entered.

**Acknowladgements:**

I would like to take the time to thank everyone in the Montgomery County Data Science Department, and specifically my professors:  
 -Kathryn Linehan

-Rachel Saidi

-Mohamed Abdirisak

I would also like to thank

-Victoria Lewis

-Dennis Linders

Without whom my project would not be possible.