Data Story

While scrolling through the data Montgomery site, I came across a data set called illegal dumping activity. I immediately remember a true crime podcast I listened to that discussed the death of Hay Min Lee, a high school student whose remains were found in Leakin Park, in West Baltimore. I opened the data sets, curious to see its content and I found that it lists mundane cases of littering and neighbors using each other's trash cans instead of their own. Slightly relieved (and maybe the tiniest bit disappointed that I didn’t find any murder cases), I was curious to see how many cases were reported for each year. Shocked to see over 7000 cases, but at least that was over the past 21 years, though, I became curious about the overall environmental health of Montgomery County. So I looked for data sets that answered my question.

I ended up finding a reputable weather and air quality data set, but it felt as if it was missing something. So I searched for a water quality data set. Coming back empty-handed, I reached out to the Department of Environmental Protection of Montgomery County. Luckily I got a response and a data set of the water quality in Montgomery County that I can use. Delighted, I started working on the most exhausting part of the project. The data wrangling. Finding that the illegal activity cases had jumbled up city names, and over 2000 cases were missing their longitude and latitude, I pushed through to label everything correctly and geocode all the missing information that I needed. I move on to the other data sets and find some NAs here and there, some mixed-up date and time columns, many unnecessary columns that were removed, and so on.

Finally finished all of that, I began to make basic graphs to see the trends and patterns, this part was existing, from choosing the types of graphs to the colors, and I enjoyed every bit. I try to create all kinds of graphs that I can to visualize the information. I find out that Silver Spring has the most amount of cases, I also learn that most of these cases take on public land and residential areas. The air quality and water quality in Montgomery County are in good shape. I also learned that the temperatures are slowly increasing.

After all the graphs, I was pleasantly surprised to see that the environmental health of Montgomery was in good shape. But I went back to that first data that I found, I wrote a logistic regression model to see whether or not Montgomery County will have over 300 cases in 2024, and the probability came back as 0.92, so yes, chances are 2024 will have over 300 cases of illegal dumping. I also used linear regression models to look a the average temperatures and find that they are increasing. I also wrote another linear regression code to see the trend in water quality scores, and I found that the numbers seem to be on the rise, which is good. So the Water quality is also increasing. Lastly, I run a Pearson’s product moment correlation to see the correlation between the median and maximum Air quality index levels, and I find that there is a positive correlation, which means that if the maximum increase, so will the median.

After all the graphs and statistical analysis that I conducted, I find that Montgomery County is in good shape. The air quality is good, and so is the water quality, the temperatures are increasing, but slowly, which could mean that we can maybe control it. Aside from the 90% chance that 2024 will have over 300 illegal dumping cases, the county is not in bad shape.

here is the link to my presentation : <https://prezi.com/view/eCCCV3ozOH65TscVxb6t/>