Project Report

Daeshik Park

DATA 205

Professor Michael Iapalucci

**On Montgomery County Population and Crime**

Interactive Visualizations Using Dash

1. **Information of Datasets**

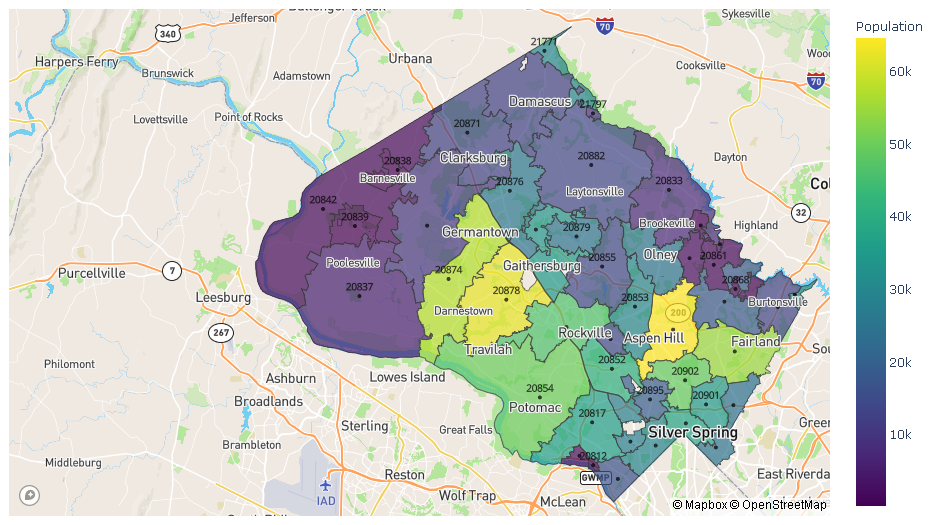
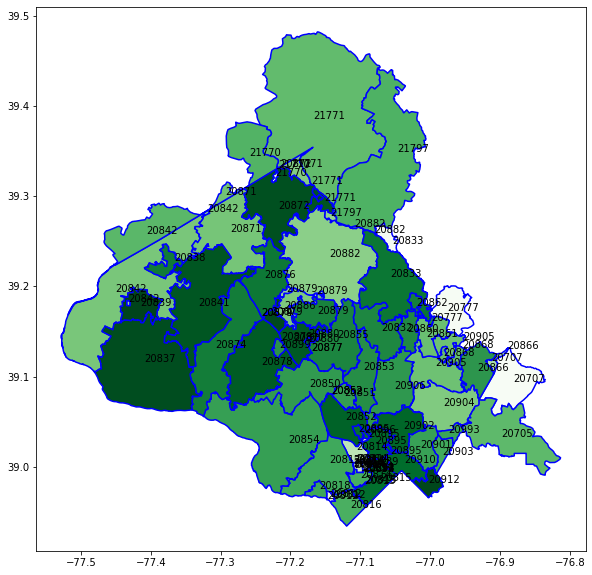
|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Description | Source | Variables of Interest |
| MC Zip Codes GIS (7/30/2020) | Zip codes and their boundaries on a Montgomery County map | Montgomery County GIS Open Data | ZIPCODE |
| ZIP Code Tabulation Areas (ZCTAs, 2010) | General demographic characteristics such as age, race, ethnicity, housing vacancy, tenure, etc. | Maryland Census Data | ZCTA5N, POP100, NHW, NHB, NHAI, NHA, NHNH, NHO, HISP, POP65\_, PCTPOP65\_ |
| Real Property Tax (7/20/2020) | County’s residential real estate properties and the associated tax charges | dataMontgomery | ZIP CODE, ASSESSMENT, COUNTY PROPERTY TAX |
| Crime (7/1/2016 ~) | All founded crimes based on preliminary information refreshed quarterly due to police investigation | dataMontgomery | Zip Code, Victims, Crime Name1, Crime Name2, Crime Name3, Start\_Date\_Time, Year |

1. **Objectives of Project**

* Investigate any relationships between the county’s real property tax and race.
* Does the number of yearly victims follow the same trend as the total number of victims over zip codes?
* Is there any pattern on the normalized number of victims over zip codes?
* Main objective: Construct interactive visualizations on the county’s Population and Crime Victims using Dash
* Provide visualizations about the county’s population over zip codes and the distribution of the population of each zip code.
* Provide visualizations about the county’s normalized number of victims and yearly victims over zip codes and about
* Display graphical information about the number of victims of a crime name for each zip code.

1. **Montgomery County Map for Zip Codes**

From the county’s base map using MC Zip Codes GIS, I manually removed the regions of the zip codes outside the county and the nonstandard zip codes in the county and then combined the polygons of the same zip code into one multi-polygon for a representative point of each zip code. I saved the result as a shape file, named “mcZip\_multiPoly. After creating a geojson file for the shape file, I used Plotly Choropleth Mapbox to visualize a color scaled County’s population with zip codes.



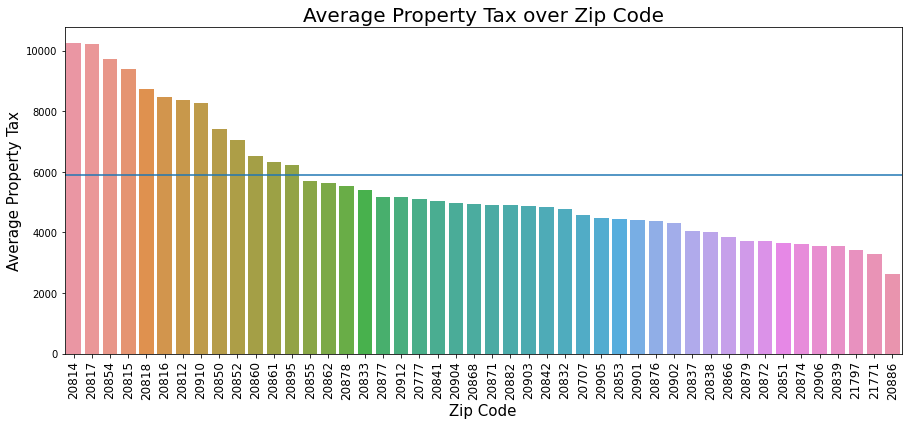
1. **Montgomery County Population**

I extracted the county’s census data for the standard zip codes from Maryland Census Data (ZCTAs). The columns of interest were selected and renamed. I saved the result as “mcPop”.

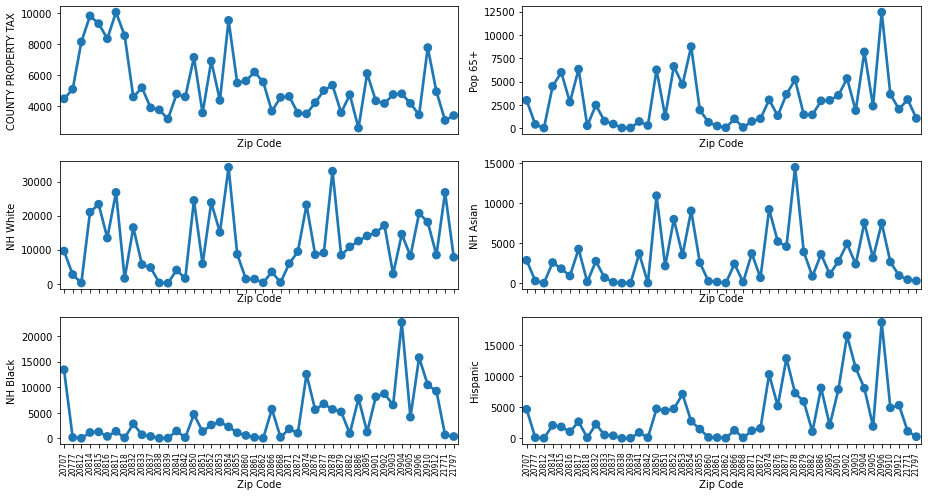
1. **Montgomery County Real Property Tax**

There are 343913 rows with 19788 NAs for zip codes. Among those NAs, there are 17755 NAs for “NOT A PRINCIPAL RESIDENCE”. All of the NAs for zip codes were dropped, and I selected the rows for the standard zip codes. I saved the result as “mcTax”.

Below is the bar graph showing the average county property tax of each zip code with the average ($5906.54) of all the county property tax. About one quarter of the zip codes are above the average county property tax. I constructed a 95% confidence interval about the mean county property tax: (4908.0, 6118.6). It contains the true mean county property tax.



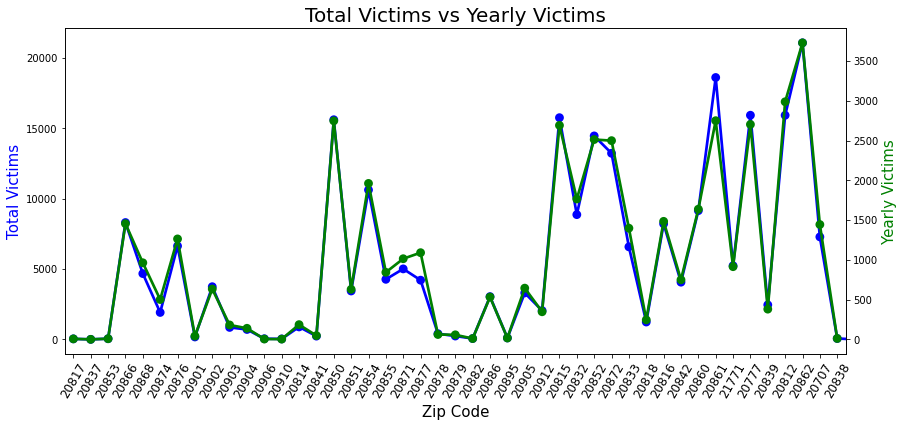
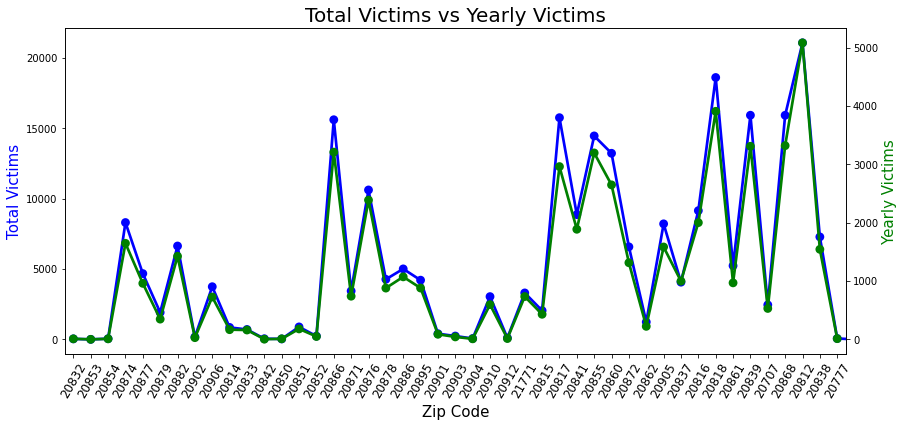
The following line graphs display the county property tax, the population of age 65+, and populations of White, Asian, Black, and Hispanic over zip codes. The county property tax is proportional to the population of age 65+ or the White population and is somewhat inversely related to the populations of Black or Hispanic.



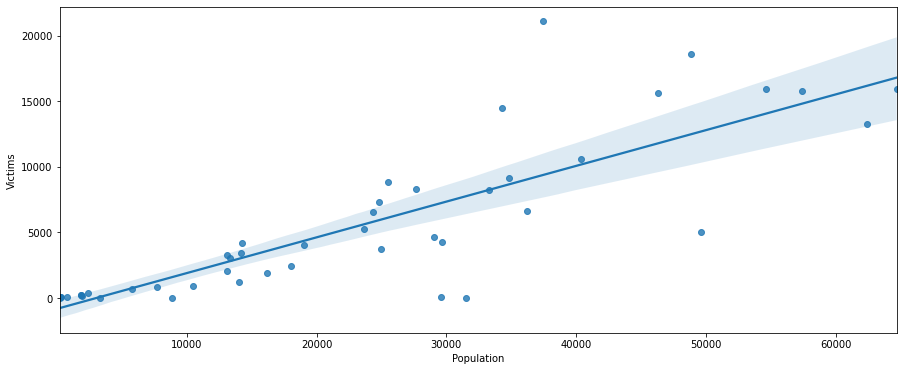
1. **Montgomery County Crime**

I dropped 3197 NAs for Zip Code among 247705 rows and selected the rows for the standard zip codes. To compare the trend of the total number of victims and the number of yearly victims, I had to add rows with zero victims to the zip codes that have no rows of years and saved the result as “mcCrime”.

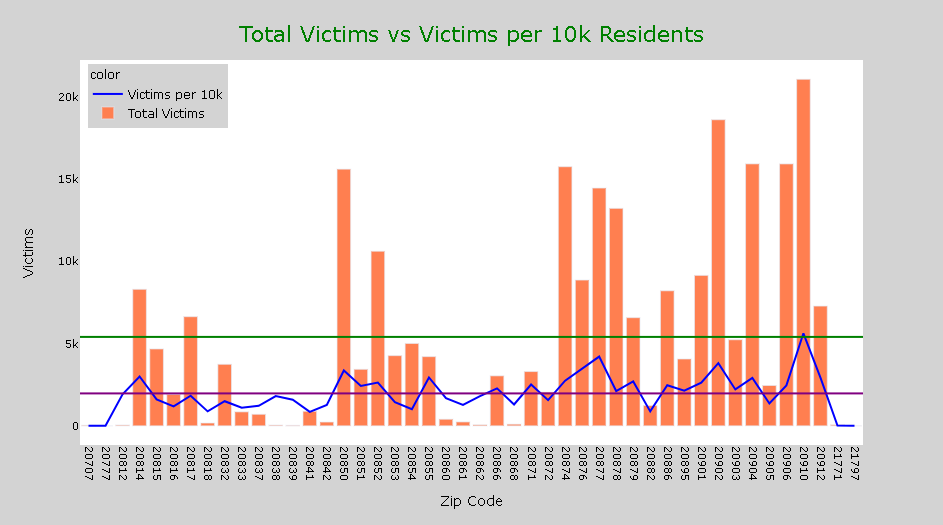
I plotted the total number of victims (blue graphs) with the number of yearly victims (green graphs for the years 2019 and 2020).



Total Victims has almost the same trend as Yearly Victims with Max 5080 and 3730 for the years 2019 and 2020. There is a big decrease (about 1300) for the max yearly victim from 2019 to 2020. The same trend is applied this year with Max 1050 up to April. It seems that there is an impact of the pandemic on the number of yearly victims.  
It is an obvious observation that the higher the population the higher the crime victims. Yes, it is true in the county. Below is the scatterplot between the county’s population and crime victims over zip codes that shows a strong and positive association with a correlation 0.84.

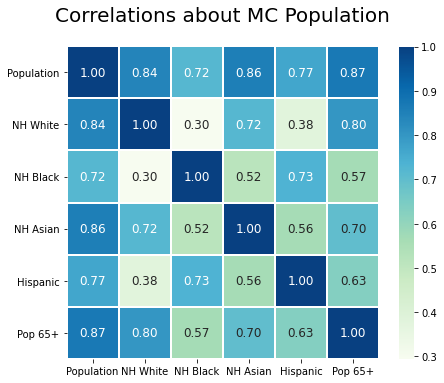


However, it is needed to “normalize” the number of victims by the population of each zip code. The green and purple lines are the average of the total victims and the average of the victims per 10k residents, respectively. Observe and realize that we should not be deceived by the total number of victims.



1. **Statistical Analysis**

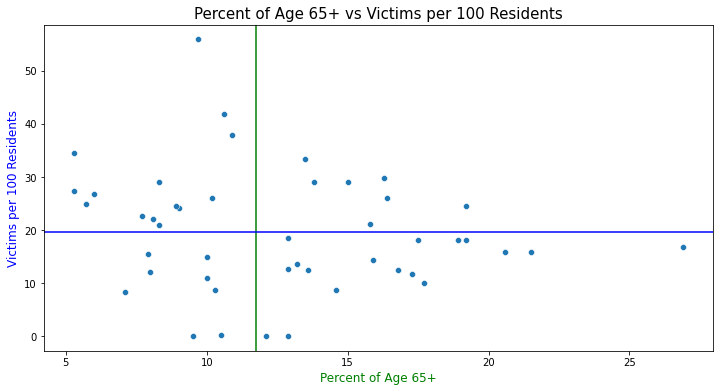
* **Relationship between Race**



The total population is related to the race and the population of age 65+ at about the same strength and direction. However, the population of age 65+ has the strongest association with the white population and the weakest association with the black population. Also, the white population has the lowest association (very weak correlation) with the black or Hispanic populations.

* **Difference of Means of Victims per 100 Residents based on Population of Age 65+**

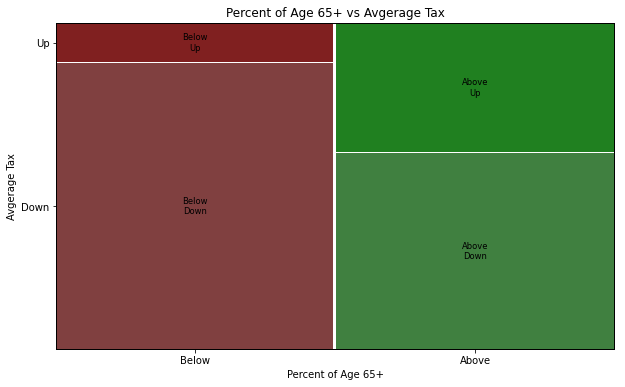
After normalizing the number of victims by 100 residents over zip codes, I divided the dataset into two groups by the average percent (11.7%) of the population of age 65+. Below is the scatterplot between the percent of age 65+ and the victims per 100 residents with the green and blue lines for their averages.



There is not much difference between the two groups. According to a 2-sample t-test for the means of the two groups, the p-value 0.202 and hence there is not enough evidence that the mean number of victims per 100 residents over zip codes between the two groups divided by the mean percent of age 65+ differ from each other. This conclusion is confirmed by a 95% confidence interval (-0.02207, 0.09278) because it contains negative or positive numbers.

* **Association between Population of Age 65+ and County Property Tax**

Divide the percent of the population of age 65+ and the county property tax by their averages. Below is the mosaic plot between the two variables and it looks like they are associated.

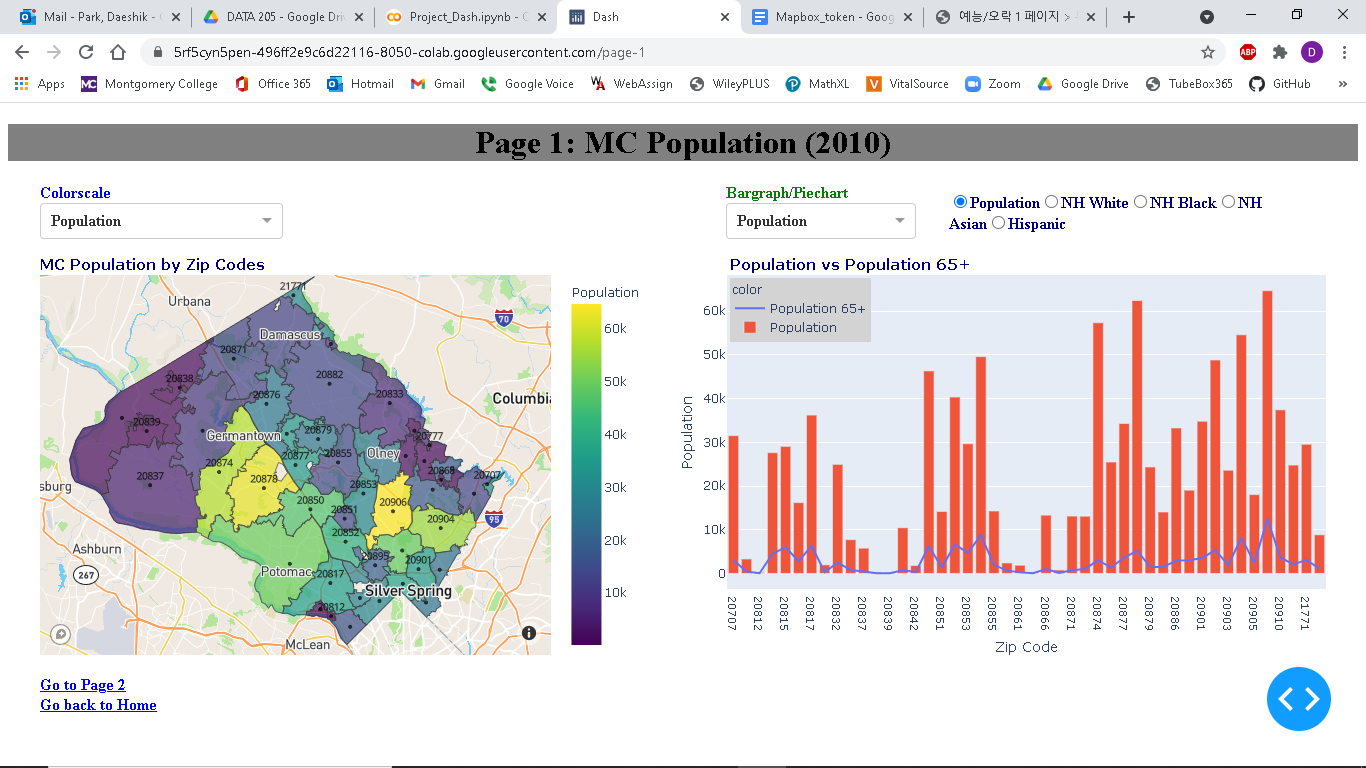


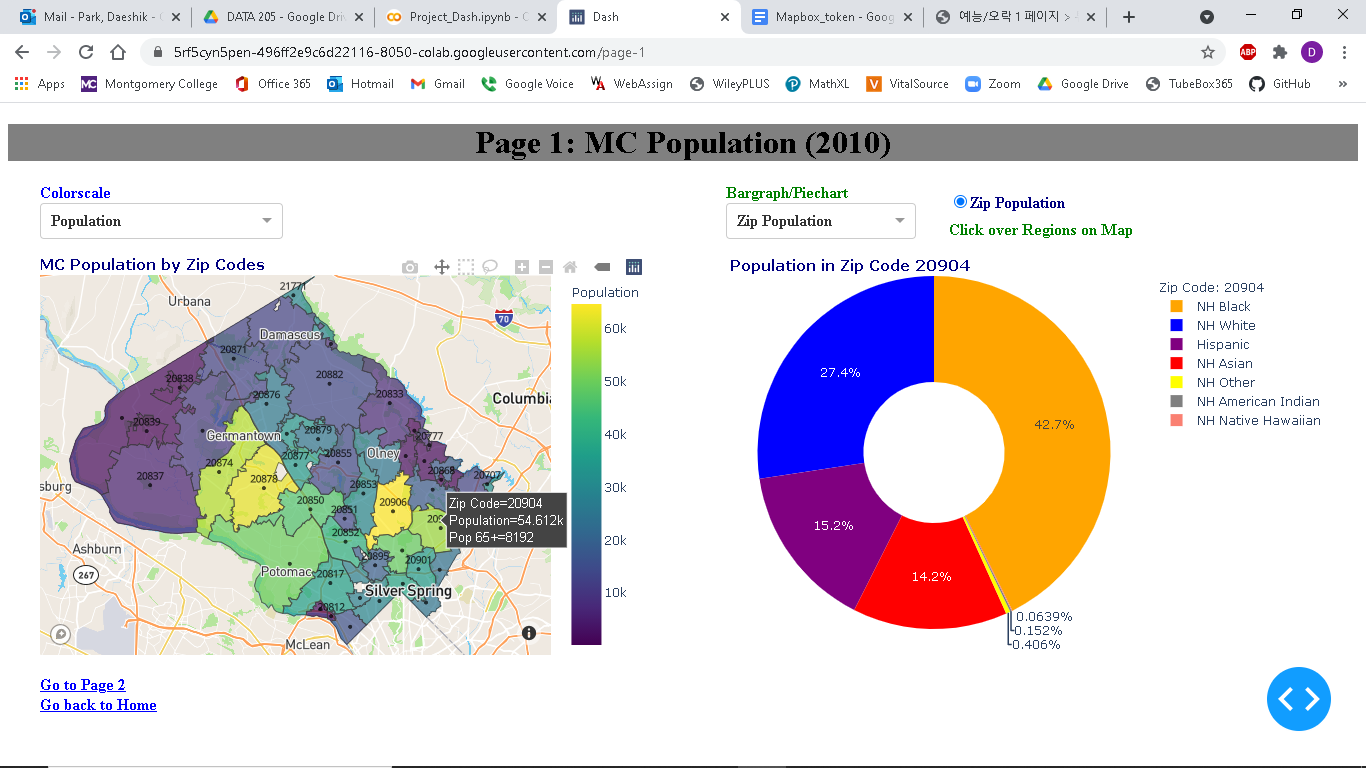
By conducting a chi-square test, I obtained a p-values 0. Thus, there is very strong evidence that the population of age 65+ is associated with the county property tax.

1. **Interactive Visualizations on MC Population and Crime Using Dash**

* **Page 1: MC Population (2010)**

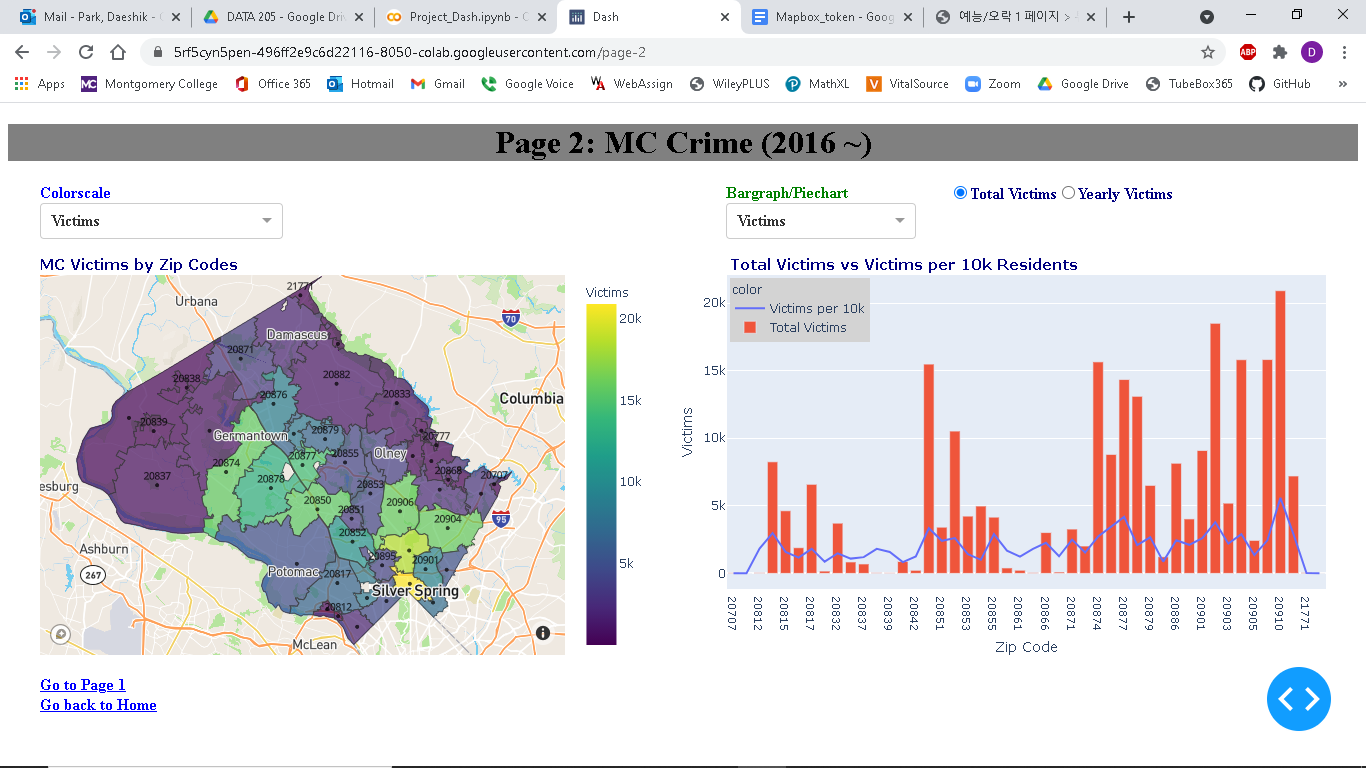
The first dropdown menu Population on the right is for visualizations about Population vs Population of Age 65+ and Populations of NH White, NH Black, NH Asian, and Hispanic over zip codes.

The second dropdown menu Zip Population on the right is for visualizations about the population of each zip code. Upon clicking the region for a zip code on the left, it displays a pie chart about its population for race.

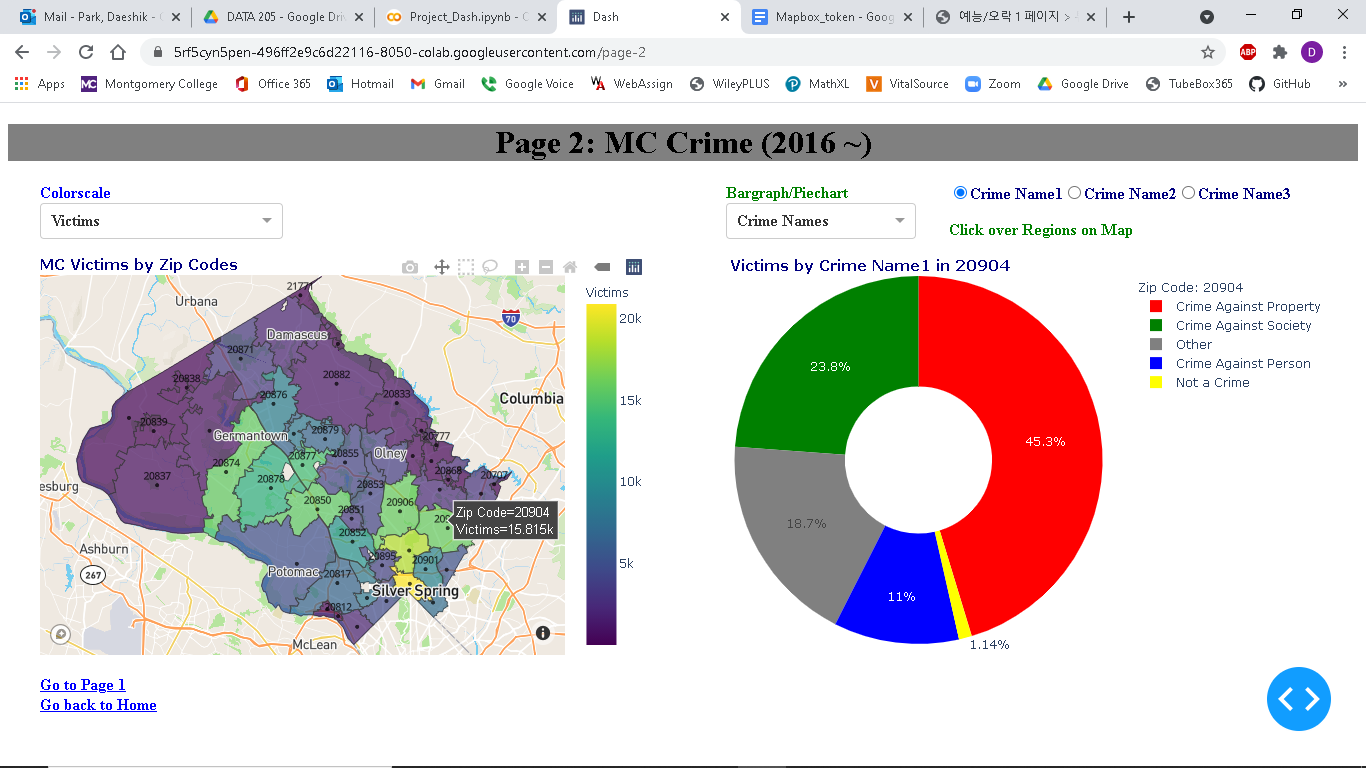


* **Page 2: MC Crime (2016 ~)**

The first dropdown menu Victims on the right is for visualizations about Total Victims vs Victims per 10k Residents and Yearly Victims with a slide for years.



The second dropdown menu Crime Names is for visualizations about Crime Names of each zip code. Upon clicking the region for a zip code on the left, it displays a pie chart about its victims by Crime Name1 and bar graphs about its victims by Crime Name2 and 3.



1. **Recommendations and Acknowledgements**

They should minimize NAs and use uniform city names. Also, I suggest that the county zip codes GIS should be revised to display only the regions in Montgomery County and change the zip code 20993 to the zip code 20903.  
I want to add more filters, such as cities and districts, in the interactive visualizations using Dash and learn how to deploy the visualizations to a web.

I sincerely thank Victoria Lewis, Manager of Data Montgomery Project, for the datasets that I have used in this project and Michael Iapalucci, Professor of DATA 205 at Montgomery College, for helping me figure out my troubles and giving me guidelines in the project.