

# Finding the safest area for the new house in Montgomery County, MD

FINAL PROJECT - DSS 665 - R PROGAMMING

STUDENT: DARKO STANKOVIC

## Introduction

The idea for this project came out of a conversation between me and my girlfriend regarding our plans to purchase a house. We both agreed upon priorities and concluded that determining factor in our decision-making process will be safety of the area. Taking into account that we have to commute to Washington DC every day for our jobs, I decided to start this project and visually present the safest location in Montgomery County.

I started the project by collecting data from the website [Data.gov](#). I derived my data set from reported crimes classified according to the National Incident-Based Reporting System (NIBRS) of the Criminal Justice Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program and documented by approved police incident reports in the Montgomery County, Maryland. The information contained herein includes all founded crimes reported after July 1<sup>st</sup>, 2016.

The project is delivered using R Programming version 3.5.1.

For this project I have selected the following libraries:

1. library(tidyverse) - is the package that load all four libraries:
  - 1.1. library(readr)
  - 1.2. library(tidyr)
  - 1.3. library(dplyr)
  - 1.4. library(ggplot2)
2. library(ggmap) - is the package to pull the real-time roadmaps from GoogleMaps
3. library(caTools) - running easy statistics on the data
4. library(ggthemes) - additional themes, scales and geoms for the ggplot2

## Loading Data

To load my data, I have used the simple command " file.choose()" a select my Crime.csv dataset. The first seven rows of the data show that my dataset contains a different type of data.

```
> crime.data <- read.csv(file.choose(), header = T)
> head(crime.data)
  Incident.ID Offence.Code CR.Number      Dispatch.Date.Time NIBRS.Code Victims          Crime.Name1
1 201087096     2399 16033231 07/01/2016 12:21:16 AM      23H 1 Crime Against Property
2 201087097     2303 16033232 07/01/2016 12:04:56 AM      23C 1 Crime Against Property
3 201087097     5202 16033232 07/01/2016 12:04:56 AM      520 1 Crime Against Society
4 201087097     5707 16033232 07/01/2016 12:04:56 AM      90J 1 Crime Against Society
5 201087100     1315 16033233 07/01/2016 12:32:17 AM      13A 1 Crime Against Person
6 201087102     3562 16033238 07/01/2016 12:58:05 AM      35A 1 Crime Against Society
          Crime.Name2          Crime.Name3 Police.District.Name      Block_Address      City State
1 All other Larceny    LARCENY (DESCRIBE OFFENSE) GERMANTOWN 19600 CRYSTAL ROCK DR GERMANTOWN MD
2 Shoplifting        LARCENY - SHOPLIFTING          WHEATON 13700 CONNECTICUT AVE SILVER SPRING MD
3 Weapon Law Violations      WEAPON - CONCEALED          WHEATON 13700 CONNECTICUT AVE SILVER SPRING MD
4 Trespass of Real Property      TRESPASSING          WHEATON 13700 CONNECTICUT AVE SILVER SPRING MD
5 Aggravated Assault ASSAULT - AGGRAVATED - OTHER GERMANTOWN 19700 FREDERICK RD GERMANTOWN MD
6 Drug/Narcotic Violations DRUGS - MARIJUANA - POSSESS SILVER SPRING 900 SILVER SPRING AVE SILVER SPRING MD
  Zip.Code Agency           Place Sector Beat PRA Address.Number Start.Date.Time End.Date.Time
1 20874 MCPD Residence - Apartment/Condo      N 5N1 595 19600 07/01/2016 12:22:00 AM
2 20906 MCPD Retail - Other                 K 4K2 337 13700 07/01/2016 12:04:00 AM
3 20906 MCPD Retail - Drug Store/Pharmacy   K 4K2 337 13700 07/01/2016 12:04:00 AM
4 20906 MCPD Retail - Drug Store/Pharmacy   K 4K2 337 13700 07/01/2016 12:04:00 AM
5 20876 MCPD Convenience Store               M 5M1 468 19700 07/01/2016 12:32:00 AM
6 20910 MCPD Parking Lot - County            G 3G1 093 900 07/01/2016 12:58:00 AM 07/01/2016 12:58:00 AM
  Latitude Longitude Police.District.Number          Location
1 39.17799 -77.26171 5D (39.177993504815994, -77.261713770478252)
2 39.07891 -77.08079 4D (39.078907345412624, -77.080791908039089)
3 39.07891 -77.08079 4D (39.078907345412624, -77.080791908039089)
4 39.07891 -77.08079 4D (39.078907345412624, -77.080791908039089)
5 39.17854 -77.23849 5D (39.178541261093216, -77.238492567560044)
6 38.99170 -77.02410 3D (38.991700529905174, -77.024095691199292)
> |
```

To find how many variables, what type of data and how big is my dataset I am running glimpse() function on my data.

```
> glimpse(crime.data)
Observations: 118,309
Variables: 26
$ Incident.ID      <int> 201087096, 201087097, 201087097, 201087097, 201087100, 201087102, 201087104, 201087108, ...
$ Offence.Code     <fct> 2399, 2303, 5202, 5707, 1315, 3562, 5311, 1313, 5707, 2602, 9199, 3550, 2903, 3512, 9106, ...
$ CR.Number        <int> 16033231, 16033232, 16033232, 16033233, 16033235, 16033248, 16033252, ...
$ Dispatch.Date.Time <fct> 07/01/2016 12:21:16 AM, 07/01/2016 12:04:56 AM, 07/01/2016 12:04:56 AM, 07/01/2016 12:04:...
$ NIBRS.Code       <fct> 23H, 23C, 520, 90J, 13A, 35A, 90C, 13B, 90J, 26A, 35B, 290, 35A, 90Z, 35A, 90C, ...
$ Victims          <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
$ Crime.Name1      <fct> Crime Against Property, Crime Against Property, Crime Against Society, Crime Against Soci...
$ Crime.Name2      <fct> All other Larceny, Shoplifting, Weapon Law Violations, Trespass of Real Property, Aggra...
$ Crime.Name3      <fct> LARCENY (DESCRIBE OFFENSE), LARCENY - SHOPLIFTING, WEAPON - CONCEALED, TRESPASSING, ASSA...
$ Police.District.Name <fct> GERMANTOWN, WHEATON, WHEATON, WHEATON, GERMANTOWN, SILVER SPRING, ROCKVILLE, GERMANTOWN, ...
$ Block_Address    <fct> 19600 CRYSTAL ROCK DR, 13700 CONNECTICUT AVE, 13700 CONNECTICUT AVE, 13700 CONNECTICU...
$ City              <fct> GERMANTOWN, SILVER SPRING, SILVER SPRING, GERMANTOWN, SILVER SPRING, ROCKVILLE, ...
$ State             <fct> MD, M...
$ Zip.Code          <int> 20874, 20906, 20906, 20906, 20876, 20910, 20850, 20874, 20850, 20852, 20910, 20912...
$ Agency            <fct> MCPD, MCPD, MCPD, MCPD, MCPD, MCPD, RCPD, MCPD, MCPD, MCPD, MCPD, MCPD, RCPD, MCPD, ...
$ Place             <fct> Residence - Apartment/Condo, Retail - Other, Retail - Drug Store/Pharmacy, Retail - Drug ...
$ Sector            <fct> N, K, K, M, G, A, N, A, D, D, G, H, A, K, H, G, G, P, N, A, E, A, N, N, H, N, E, H, ...
$ Beat              <fct> 5N1, 4K2, 4K2, 4K2, 5M1, 3G1, 1A3, 5N1, 1A3, 2D4, 2D4, 3G1, 3H1, 1A2, 4K2, 3H1, 3H1, 3G1, ...
$ PRA               <fct> 595, 337, 337, 337, 468, 093, 277, 702, 277, 187, 187, 094, 129, 299, 334, 126, 126, 671, ...
$ Address.Number   <int> 19600, 13700, 13700, 13700, 19700, 900, 1, 19600, 300, 11800, 11800, 8500, 900, 13000, 33...
$ Start.Date.Time   <fct> 07/01/2016 12:22:00 AM, 07/01/2016 12:04:00 AM, 07/01/2016 12:04:00 AM, 07/01/2016 12:04:...
$ End.Date.Time     <fct> , , , , 07/01/2016 12:58:00 AM, 07/01/2016 01:00:00 AM, , 07/01/2016 05:30:00 AM, , , ...
$ Latitude          <dbl> 39.17799, 39.07891, 39.07891, 39.07891, 39.17854, 38.99170, 39.08047, 39.17689, 39.08644, ...
$ Longitude         <dbl> -77.26171, -77.08079, -77.08079, -77.08079, -77.23849, -77.02410, -77.15281, -77.27311, -...
$ Police.District.Number <fct> 5D, 4D, 4D, 5D, 3D, 1D, 5D, 1D, 2D, 2D, 3D, 3D, 1D, 4D, 3D, 3D, 3D, 3D, 6D, 5D, 1D, 2...
$ Location          <fct> (39.177993504815994, -77.261713770478252), (39.078907345412624, -77.080791908039089), (39...
```

From the given output I can see that my dataset has 26 variables and 118,309 observations (has 4550 rows of data). The majority of my data has a factor data type.

I can see that some of the variables are repeated, and I will choose only variables that will have importance for my project.

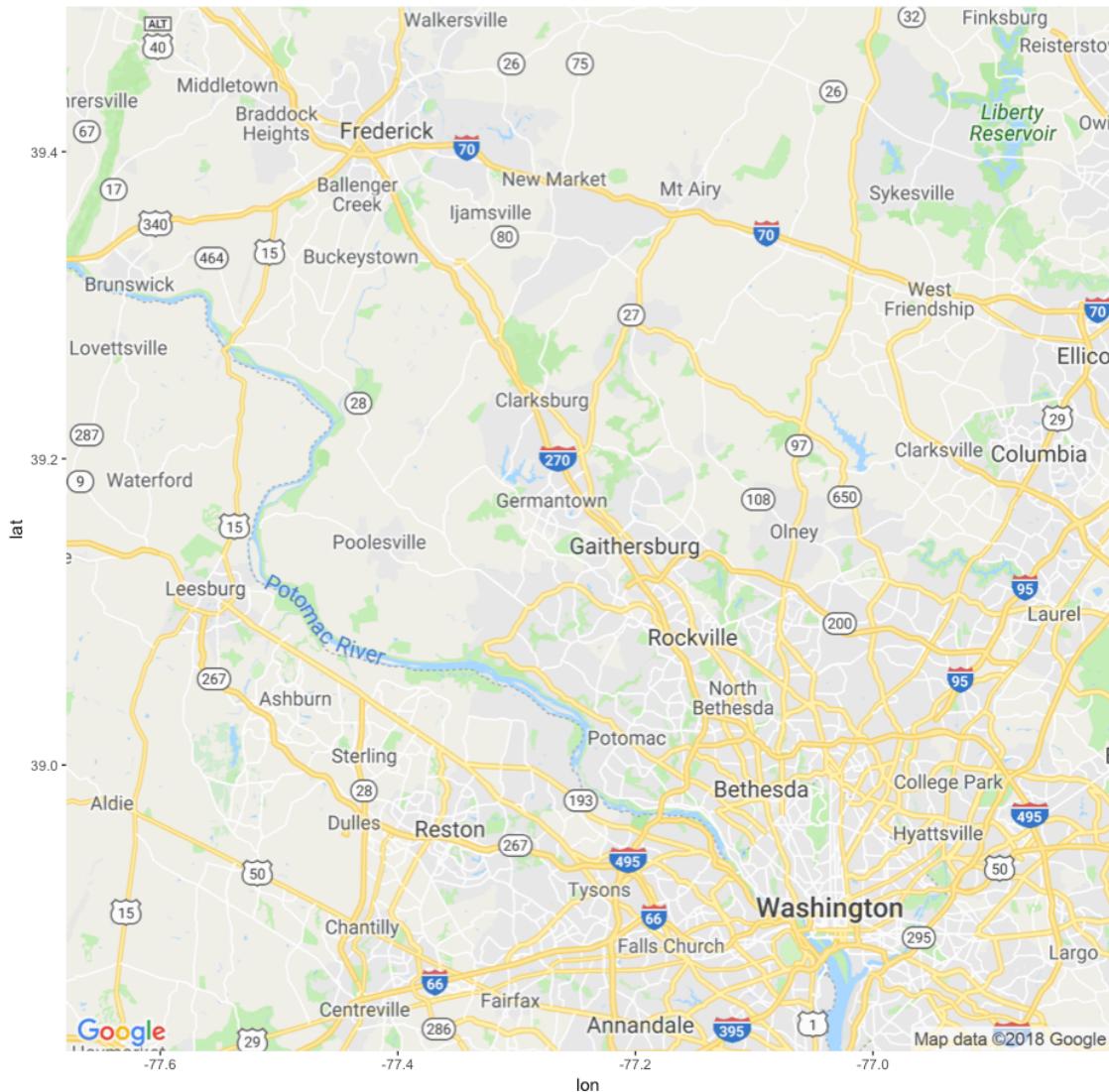
```
> crime.project <- crime.data[c(4, 6, 8, 10, 14,23,24)]
> glimpse(crime.project)
Observations: 118,309
Variables: 7
$ Dispatch.Date.Time   <fct> 07/01/2016 12:21:16 AM, 07/01/2016 12:04:56 AM, 07/01/2016 12:04:56 AM, 07/01/2016 12:04:56...
$ Victims              <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
$ Crime.Name2           <fct> All other Larceny, Shoplifting, Weapon Law Violations, Trespass of Real Property, Aggravate...
$ Police.District.Name <fct> GERMANTOWN, WHEATON, WHEATON, WHEATON, SILVER SPRING, ROCKVILLE, GERMANTOWN, RO...
$ Zip.Code               <int> 20874, 20906, 20906, 20906, 20876, 20910, 20850, 20874, 20850, 20852, 20910, 20912, ...
$ Latitude                <dbl> 39.17799, 39.07891, 39.07891, 39.07891, 39.17854, 38.99170, 39.08047, 39.17689, 39.08664, 3...
$ Longitude              <dbl> -77.26171, -77.08079, -77.08079, -77.08079, -77.23849, -77.02410, -77.15281, -77.27311, -77...
```

Now when I have my data set ready, I can continue the project and start the visualizations of my data.

## Visualization

My visualization starts with finding the coordinates for the area of interest and pulling the map from Google Maps. Montgomery County in Maryland has Longitude = -77.24052 and Latitude = 39.15474. My zoom parameter for this map is set to 10.

```
> geocode("Montgomery County")
Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Montgomery%20County&sensor=false
  lon      lat
1 -77.24052 39.15474
> mapMontCount <- get_map(location = c(-77.24052, 39.15474), source = "google", zoom = 10, maptype = "roadmap")
Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=39.15474,-77.24052&zoom=10&size=640x640&scale=2&maptype=roadmap&language=en-EN&sensor=false
> gmap(mapMontCount)
> |
```



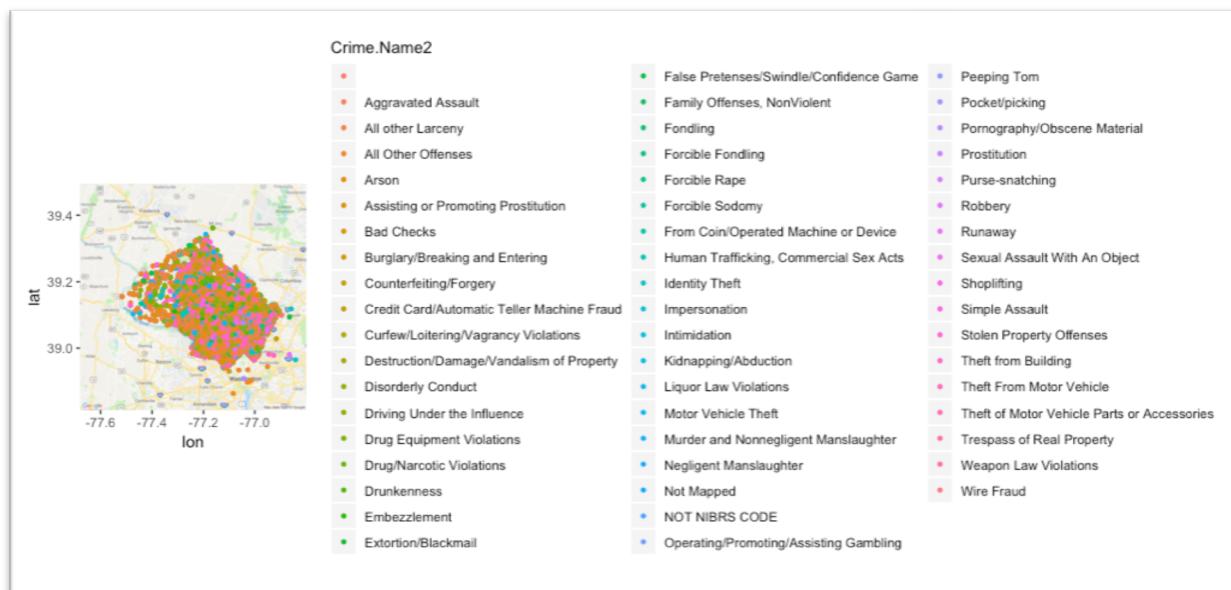
Montgomery County by Google Maps

Now when I have my map, I am going to check what type of crimes I have in my dataset by putting all data into a table. The data is organized from the most repeated offense to the least repetitive offense. Following the crime table, I plugged all crimes on the Montgomery County map.

```
> sort(table(crime.project$Crime.Name2), decreasing = T)
```

All Other Offenses	26206	Drug/Narcotic Violations	12408
Simple Assault	7605	Destruction/Damage/Vandalism of Property	7197
Shoplifting	5575	All other Larceny	5282
Liquor Law Violations	3762	Burglary/Breaking and Entering	3572
Motor Vehicle Theft	1977	False Pretenses/Swindle/Confidence Game	1910
Theft of Motor Vehicle Parts or Accessories	1740	Credit Card/Automatic Teller Machine Fraud	1594
Impersonation	1468	Runaway	1422
Identity Theft	1414	Drug Equipment Violations	1066
Weapon Law Violations	838	Forcible Rape	416
Family Offenses, NonViolent	331	Pocket/picking	265
Forcible Fondling	161	Embezzlement	152
Purse-snatching	145	Not Mapped	126
Forcible Sodomy	122	Stolen Property Offenses	119
Extortion/Blackmail	97	Prostitution	96
Wire Fraud	74	Pornography/Obscene Material	67
Peeping Tom	38	Murder and Nonnegligent Manslaughter	36
Curfew/Loitering/Vagrancy Violations	25	From Coin/Operated Machine or Device	25
Human Trafficking, Commercial Sex Acts	18	Assisting or Promoting Prostitution	15
Operating/Promoting/Assisting Gambling	1		

```
> ggmap(mapMontCount) + geom_point(aes(x = Longitude, y = Latitude, colour = `Crime.Name2`),  
+ size = 1, data = crime.project)
```



All crimes committed in MC after July 1<sup>st</sup>, 2016

Conclusion after this map is that I have a lot of data and a lot of different crimes.

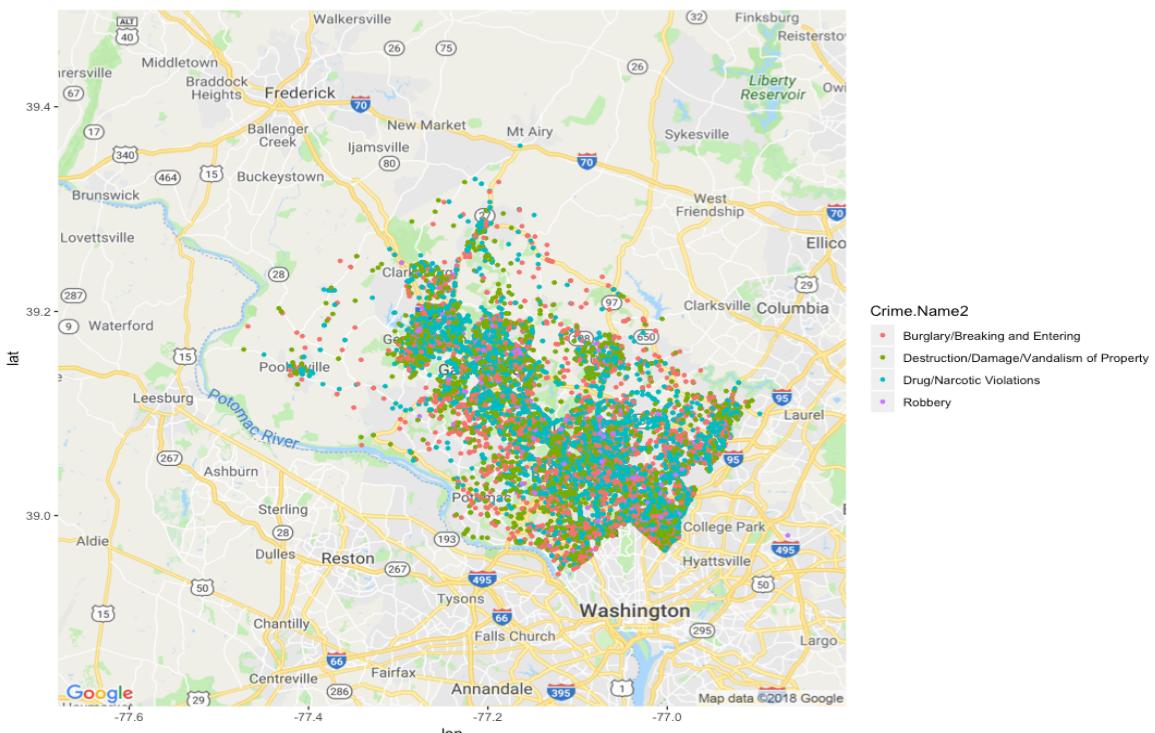
To better present data and achieve desired outcome I am limiting crimes data to four crimes:

1. Drug/Narcotic Violations
2. Destruction/ Damage/ Vandalism of Property
3. Burglary / Breaking and Entering
4. Robbery.

```
filt.crime.project <- filter(crime.project,
  Crime.Name2 == "Drug/Narcotic Violations" |
  Crime.Name2 == "Destruction/Damage/Vandalism of Property" |
  Crime.Name2 == "Burglary/Breaking and Entering" |
  Crime.Name2 == "Robbery")
sort(table(filt.crime.project$Crime.Name2), decreasing = T)
```

Crime Type	Count
Drug/Narcotic Violations	12408
Burglary/Breaking and Entering	3572
Destruction/Damage/Vandalism of Property	7197
Robbery	1484

The rest of the data will remain "zero" in the further analysis. Going back to the map and add only filtered crimes.



Filtered crimes committed in MC after July 1<sup>st</sup>, 2016

I can see on the map that there are some areas where the crime level is on the high level and also some areas where crime level is not that high.

Using our dataset, I am looking for the city that is going to be our possible choice.

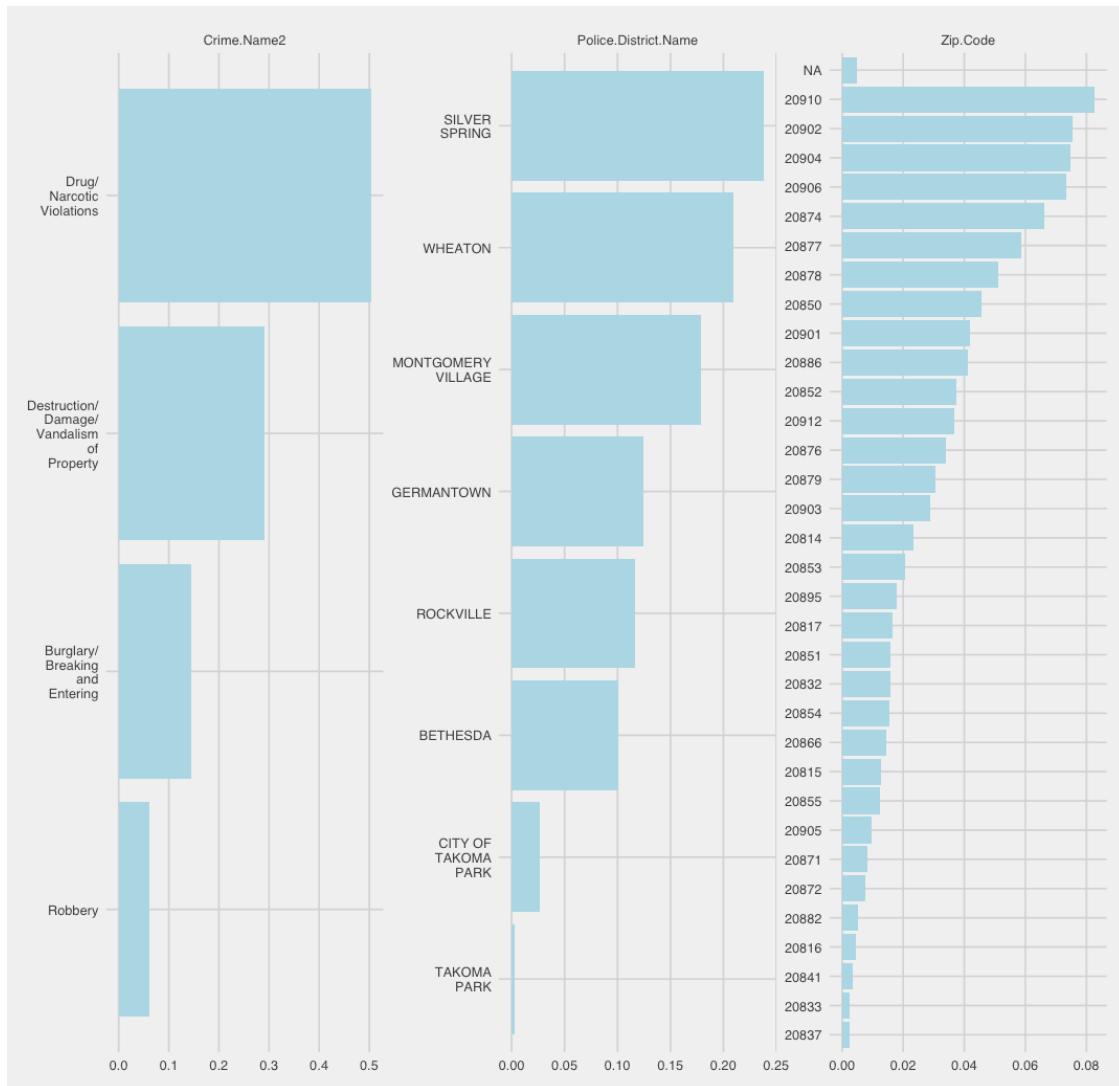
Sorting crime in the table I can see that Takoma Park is the area with only 73 committed crimes. Comparing to Silver Spring with 5874 crimes or Wheaton with 5174 committed crimes, Takoma Park has a much lower crime rate.

```
> sort(table(filt.crime.project$Police.District.Name), decreasing = T)
```

SILVER SPRING	WHEATON	MONTGOMERY VILLAGE	GERMANTOWN	ROCKVILLE
5874	5174	4419	3079	2876
BETHESDA CITY OF TAKOMA PARK		TAKOMA PARK	OTHER	
2484	652	73	30	

Double checking the data by making bar plots to confirm that Takoma Park is the right choice.

```
crime.frequency <- filt.crime.project [c(3,4,5)] %>%  
  gather("var", "value") %>%  
  group_by(var) %>% |  
  count(var, value) %>%  
  mutate(prop = prop.table(n)) %>%  
  filter(prop > .02)  
  
crime.plot <-  
  ggplot(data = crime.frequency,  
         aes(x = reorder(stringr::str_wrap(value, 10), prop),  
              y = prop)) +  
  geom_bar(stat = "identity", fill = "lightblue") +  
  coord_flip() +  
  facet_wrap(~var, ncol = 3, scales = "free") +  
  ggthemes::theme_fivethirtyeight()  
  
crime.plot
```



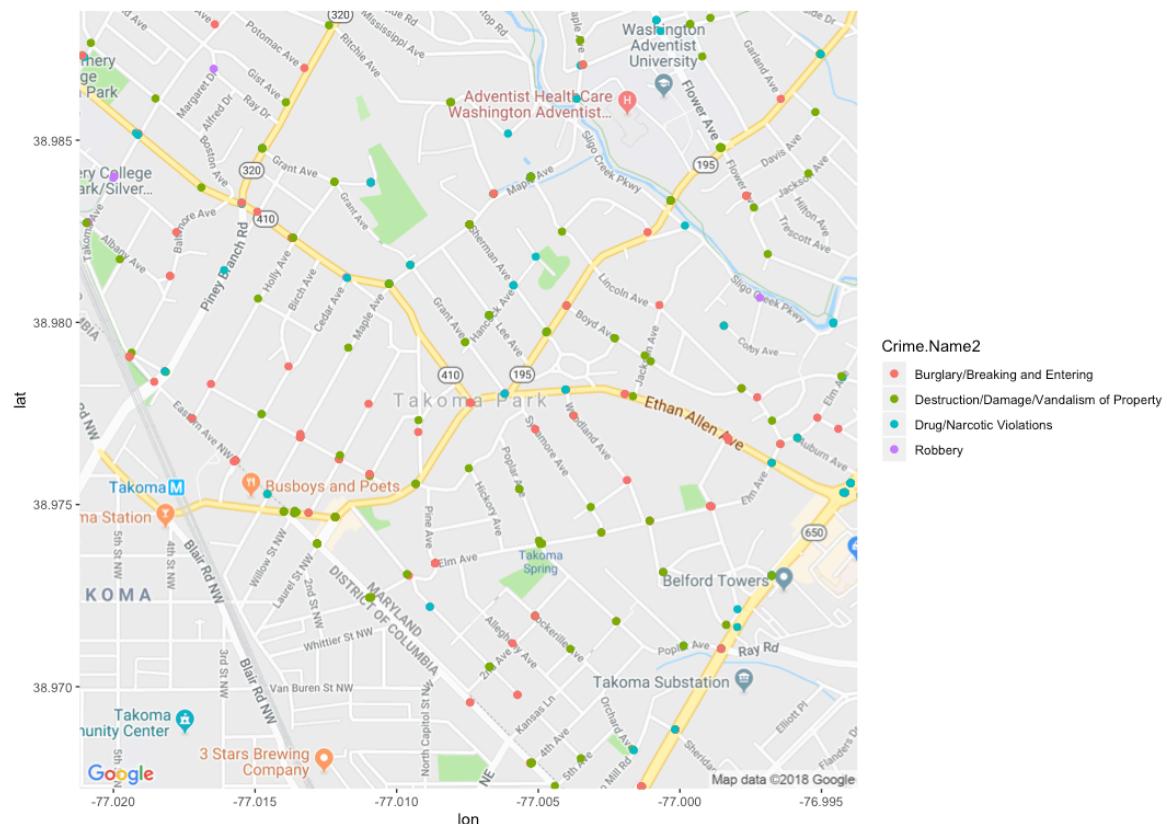
Bar presentation of the crime rate for different variable

Bar presentation confirmed that Takoma Park is for now the best possible choice. I will perform the data visualization to check the area.

## Takoma Park Visualization

First, I found the Longitude and Latitude for Takoma Park area. Next, I imported the crime data in the Takoma Park Google Map to help me visualize the data. The zoom for the map is set to 15 so that I can have a closer look to data.

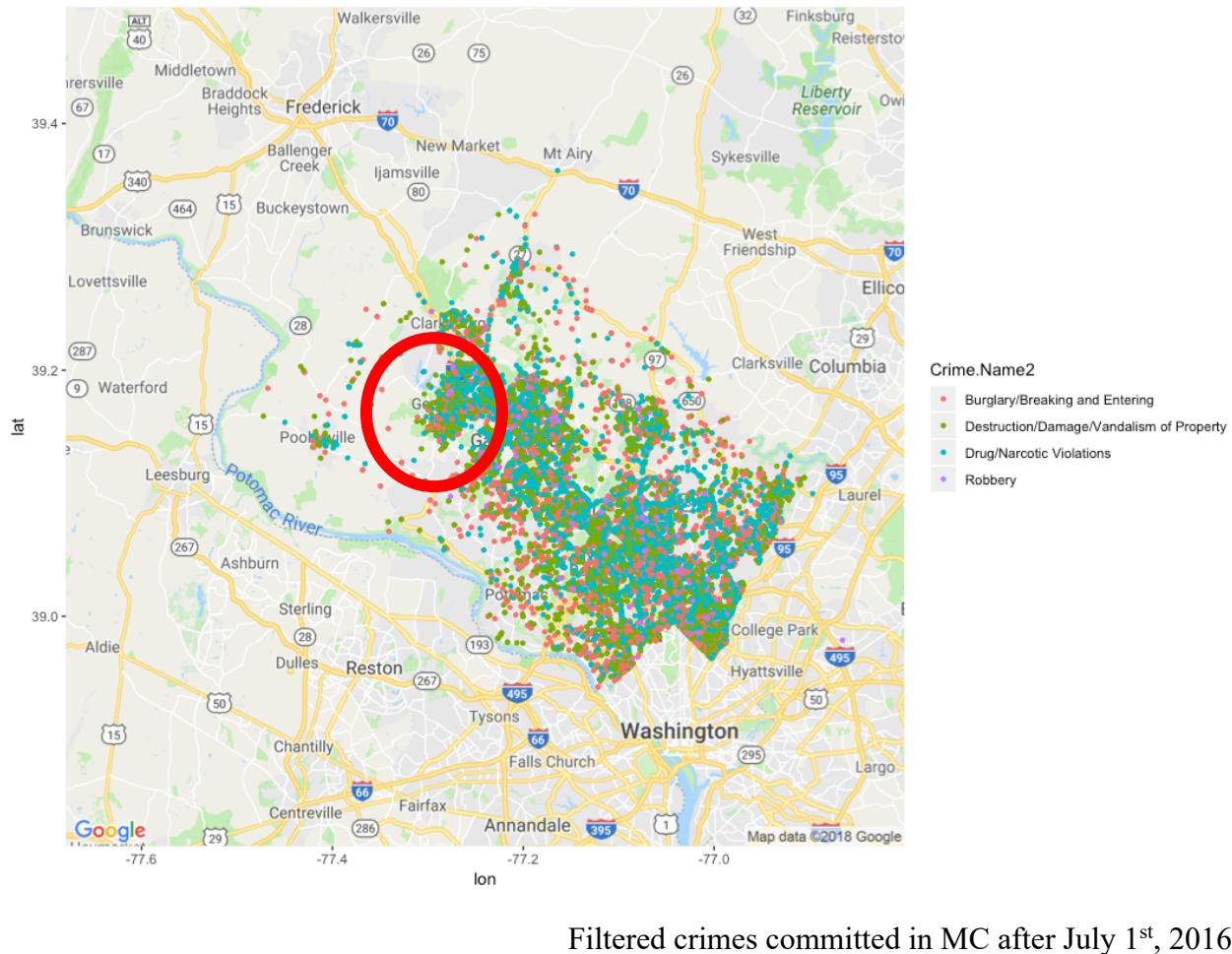
```
> geocode("Takoma Park")
Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Takoma%20Park&sensor=false
  lon      lat
1 -77.00748 38.97789
> mapTakoma <- get_map(location = c(-77.00748, 38.97789), source = "google", zoom = 15, maptype = "roadmap")
Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=38.97789,-77.00748&zoom=15&size=640x640&scale=2&maptype=roadmap&language=en-EN&sensor=false
> ggmap(mapTakoma)
> ggmap(mapTakoma) + geom_point(aes(x = Longitude, y = Latitude, colour = `Crime.Name2`),
+                                   size = 2, data = filt.crime.project)
```



Filtered crimes committed in Takoma Park after July 1<sup>st</sup>, 2016

Get back to my girlfriend with the explanation what this visualization presents, and her first comment was that this area looks like "Crime on every corner". Looking at the map, it seems like that the crime is happening on almost every corner.

From here I decided to go back to the original Montgomery County Crime map and try to find the spot on the map that is "clean" but also not too far from the urban city areas.

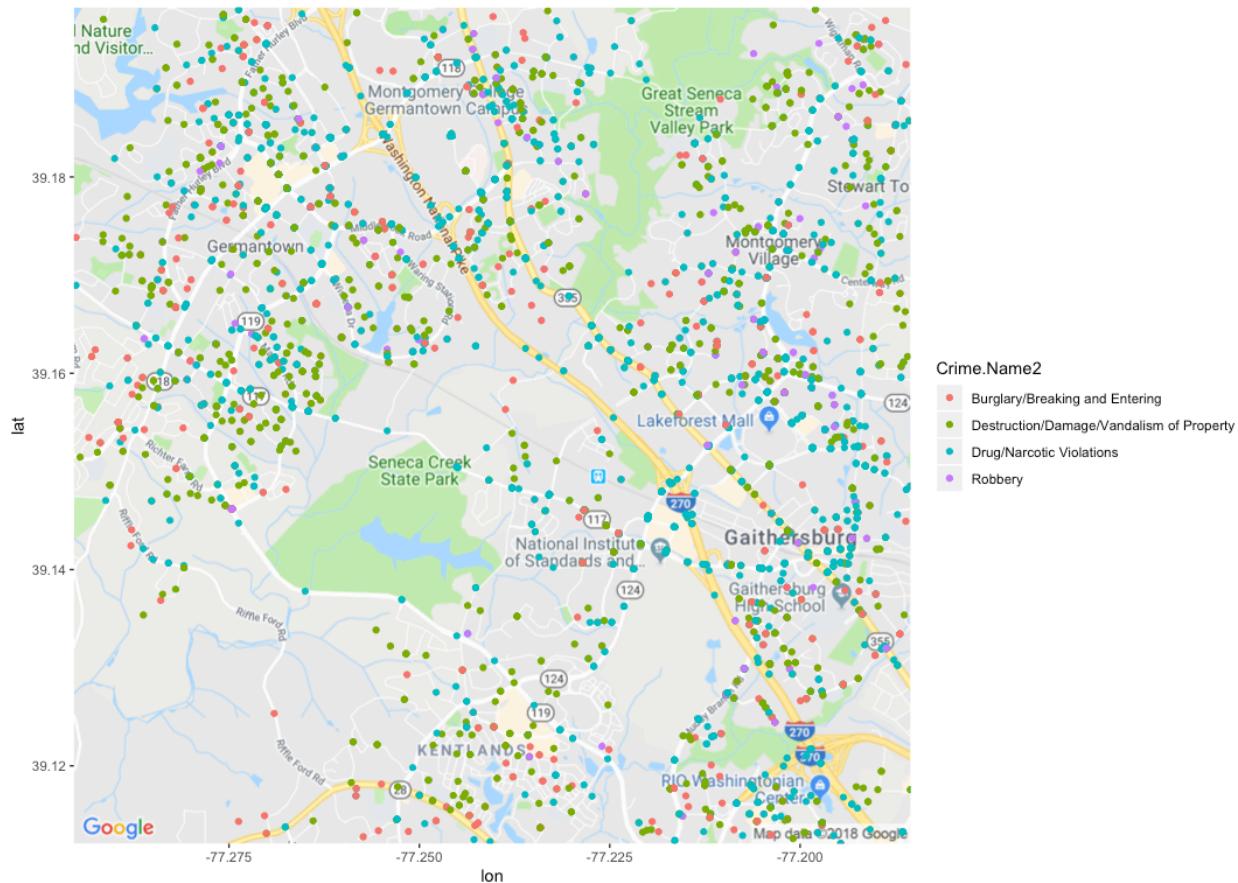


Looking at the map, I decided to check the red circled area because it seems like a part that does not have a lot of dots (crimes). Also, this is close to the urban city area which is one of the priorities in our choice.

## Montgomery County crime on the zoomed map

Changing zoom to 13 allows me to have a better view of this area after I insert the crime data.

```
> mapMontCount <- get_map(location = c(-77.24052, 39.15474), source = "google", zoom = 13, maptype = "roadmap")
Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=39.15474,-77.24052&zoom=13&size=640x640&scale=2&maptype=roadmap&language=en-EN&sensor=false
> ggmap(mapMontCount)
>
> ggmap(mapMontCount) + geom_point(aes(x = Longitude, y = Latitude, colour = `Crime.Name2`),
+                                     size = 1.5, data = filt.crime.project)
```

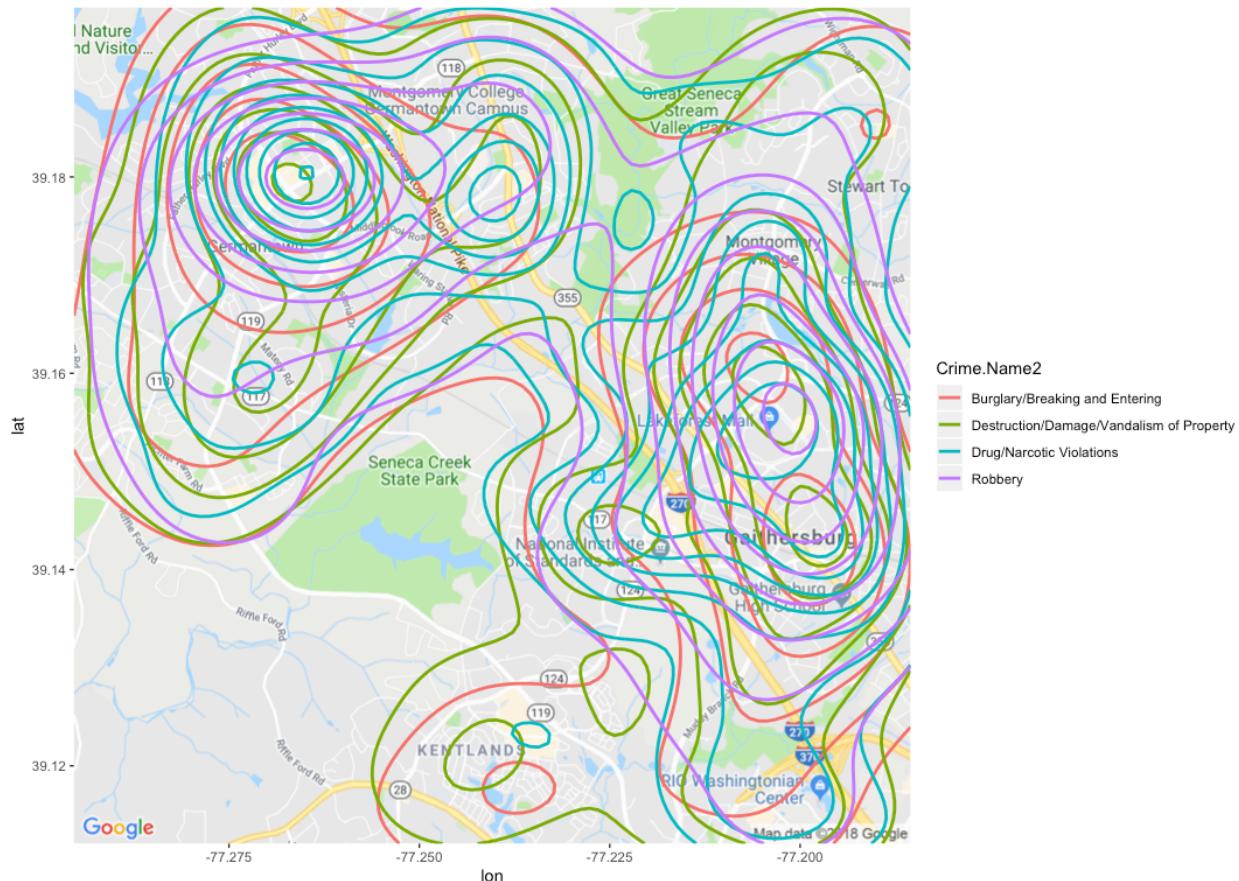


Filtered crimes committed in part of MC after July 1<sup>st</sup>, 2016

On the first look, this area looks like place that has high crime volume. I decided to go into the more in-depth visual analysis and see the density of the crime in the area.

Density of all crime in the chosen area will show me the concentration of the crime.

```
ggmap(mapMontCount) + geom_density2d (aes(x = Longitude, y = Latitude, colour = `Crime.Name2`),  
size = 1, data = filt.crime.project)
```



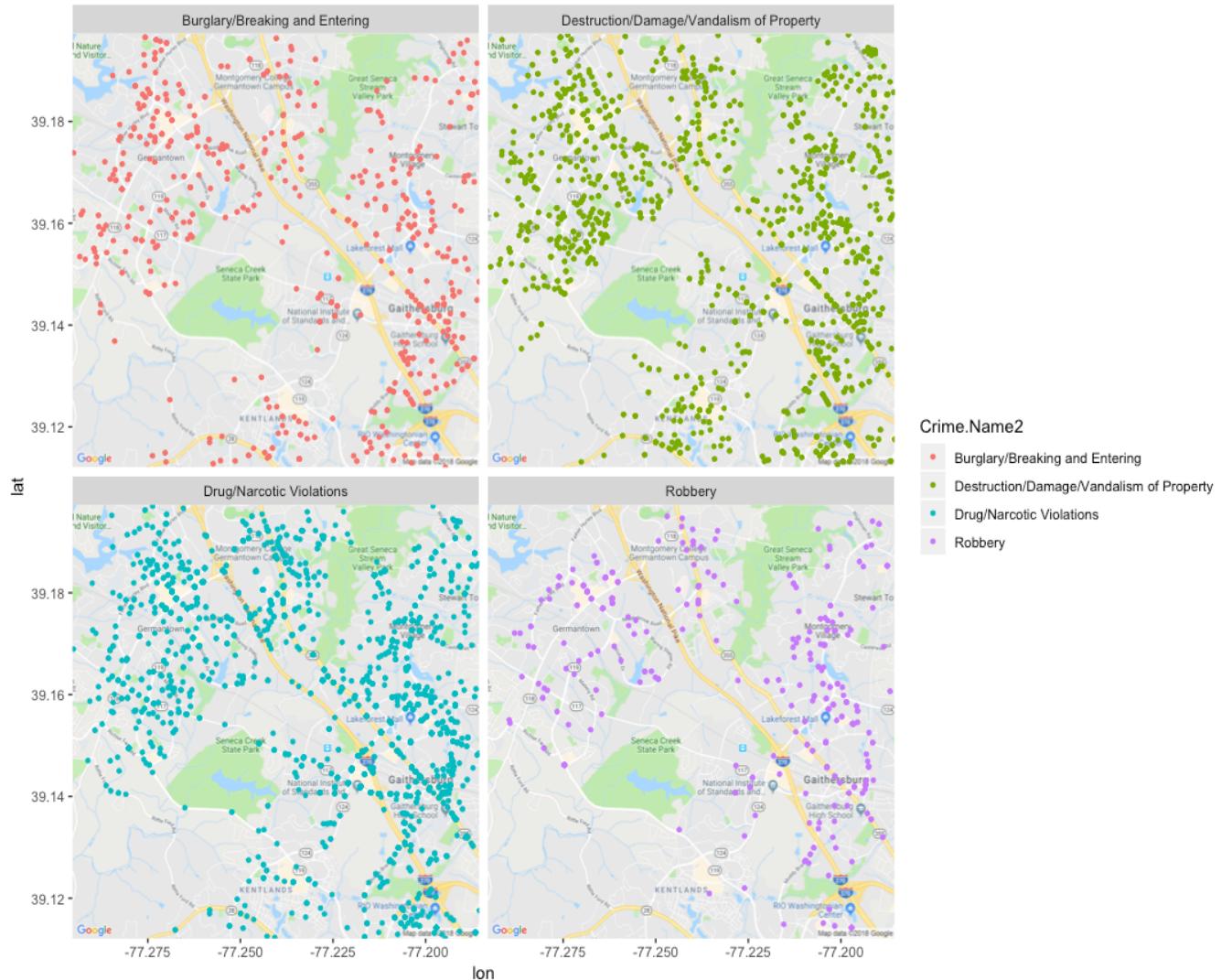
Density of the Filtered crimes after July 1<sup>st</sup>, 2016

The density presentation looks a little bit messy, but I can see that small town called Kentlands has a much lower density of the crime compared to the rest of the map.

## Splitting the map for each crime

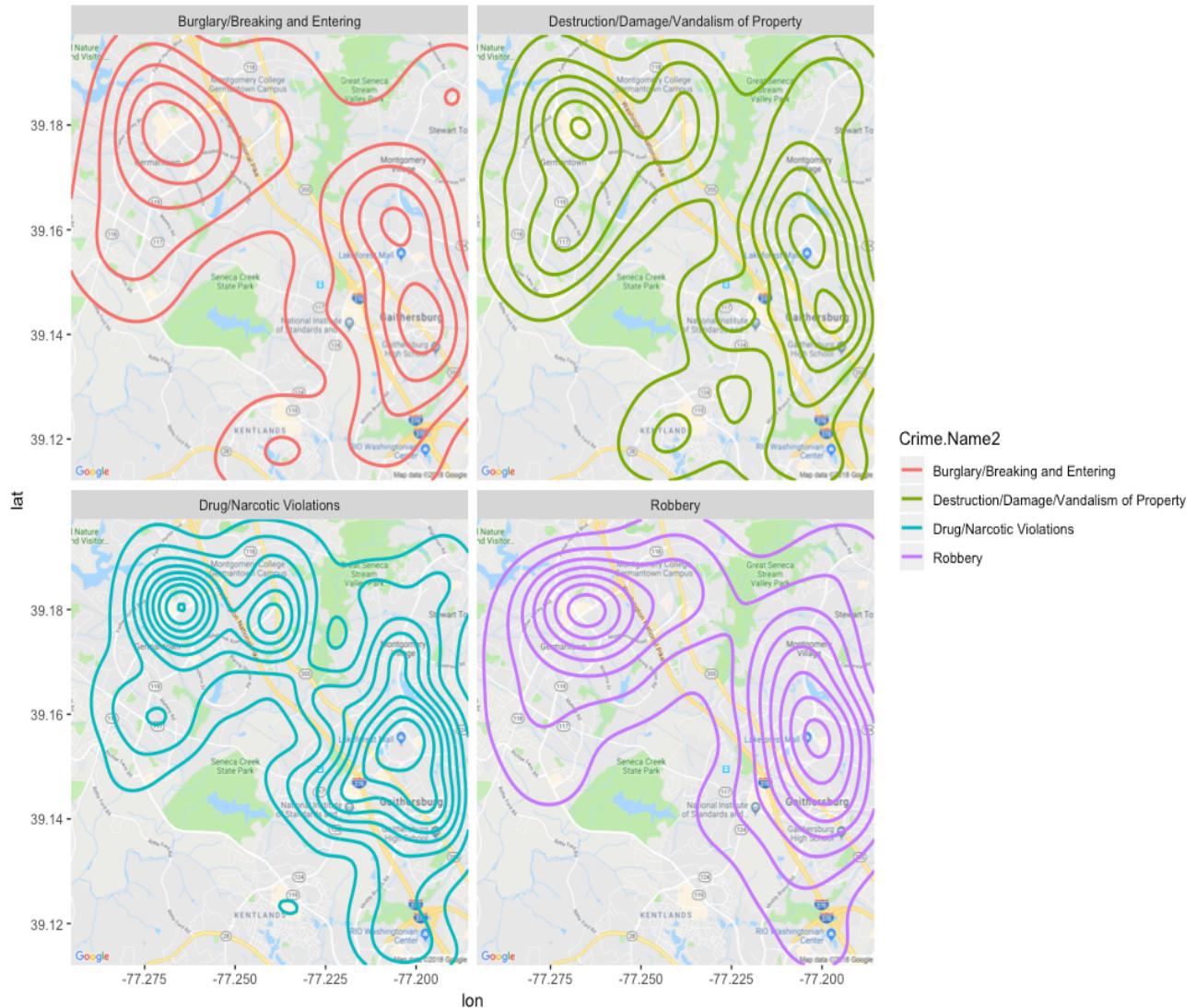
My next goal is to split the map for every crime and check the density of the crime in the area for every crime separately.

```
> ggmap(mapMontCount) + geom_point(aes(x = Longitude, y = Latitude, colour = `Crime.Name2`),  
+                                     size = 1, data = filt.crime.project) +  
+   facet_wrap(~ `Crime.Name2`)
```



Split map for each crime after July 1<sup>st</sup>, 2016

Density for the map by crime shows us that the area around Kentlands has almost no Robbery crime and very low Drug / Narcotic Violations. The Destruction / Damage / Vandalism of Property and Burglary in Kentlands has much lower density comparing to the rest of the urban area of the map.

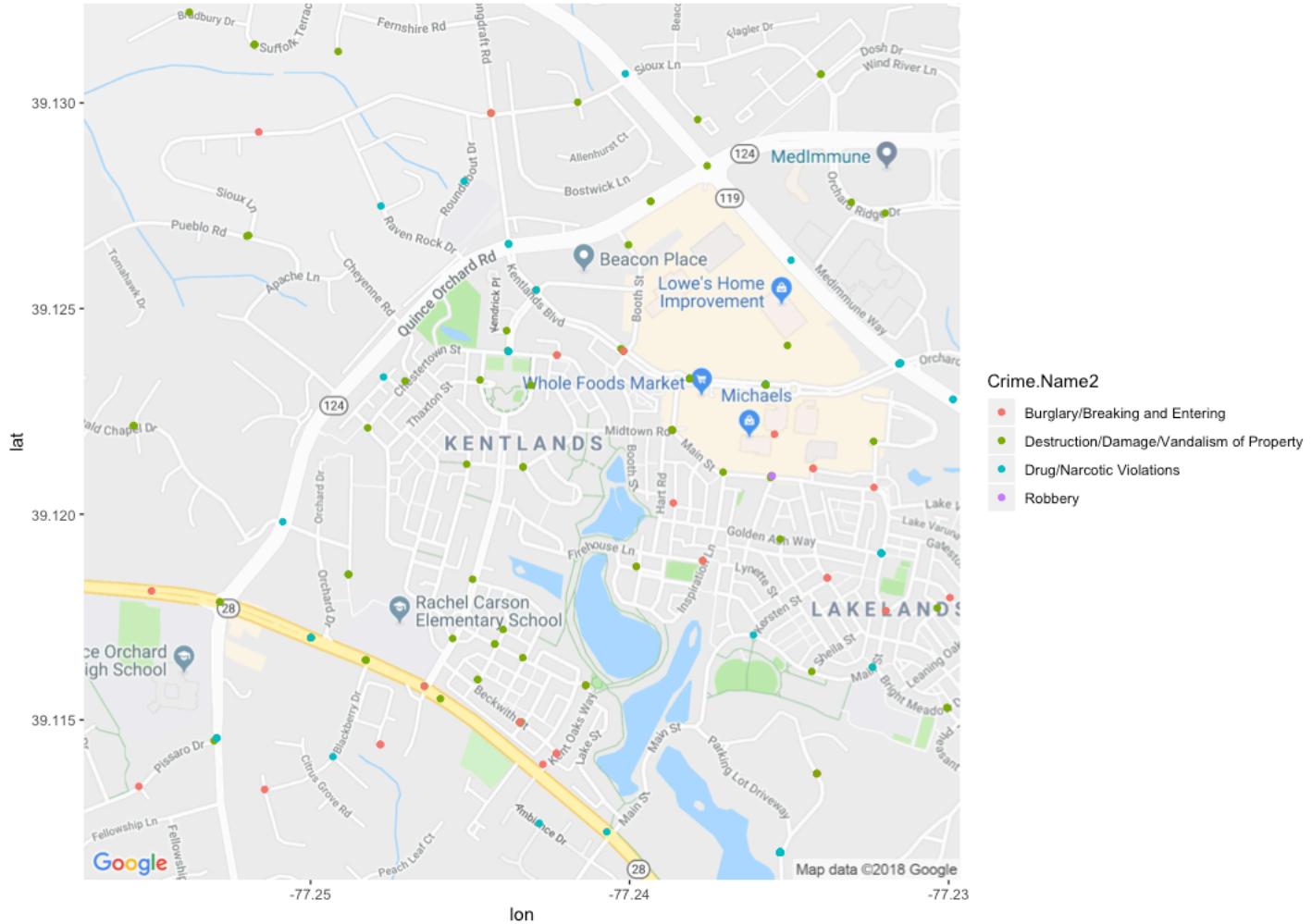


Split map for each crime density after July 1<sup>st</sup>, 2016

## Kentlands Visualization

I plugged the crime data into Kentlands area for the visual comparison with Takoma Park. Note that I am using the same zoom = 15 for the map view as I did in Takoma Park.

```
> geocode("Kentlands")
Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Kentlands&sensor=false
      lon      lat
1 -77.24339 39.12178
> ggmap(mapKentlands) + geom_point(aes(x = Longitude, y = Latitude, colour = `Crime.Name2`),
+                                     size = 1.5, data = filt.crime.project)
+ ...
.
```



Filtered crimes committed in Kentlands after July 1<sup>st</sup>, 2016

## Conclusion

My girlfriend looked at the map and loved the fact that the area has almost everything that we wanted. It has an elementary school, shopping center, small lakes around perfect for a morning run and the most important it has low crime frequency.

We made our decision.

We are going to contact our real estate agent and let him know that our wish is to look for the hour in the Kentlands, MD.

This project shows that R programming is an powerful language with a very broad usage in the real world problems.