

EDAV Project 3 - Alcohol-related traffic violations in Montgomery county of Maryland

Team DC

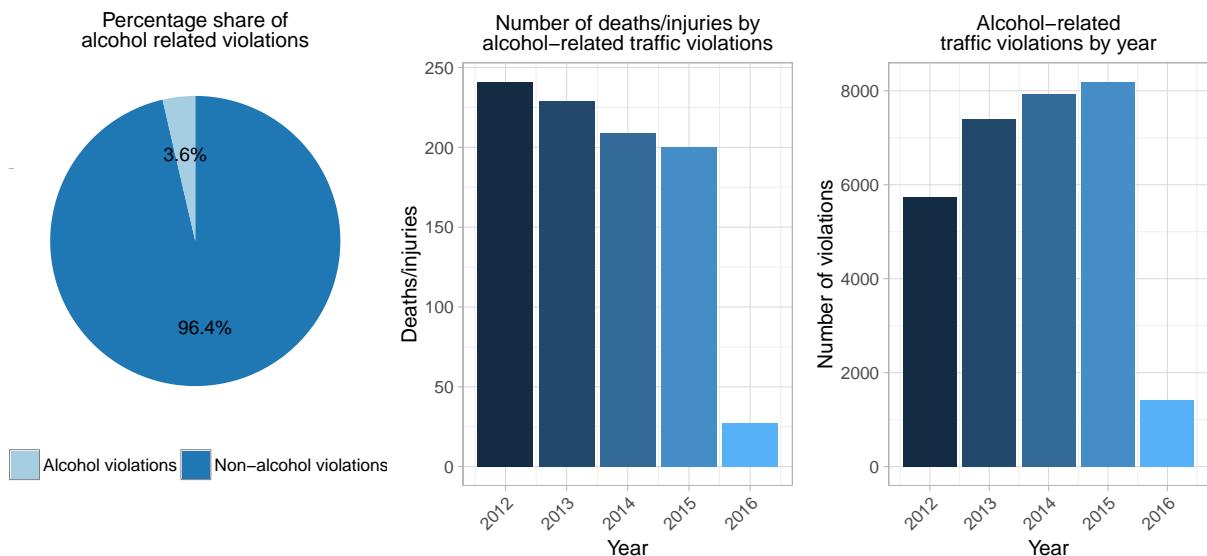
March 23, 2016

With this project, we aim to present a simple yet cohesive and concluding approach to one of the most relevant application fields in Data Science: smart city planning. For this purpose, we targeted a relatively small and simple dataset that contains all traffic violations since 2012 in Montgomery County, Maryland. This dataset, though simple, provides very accurate and descriptive information on the nature of the traffic violations. In particular, we will focus on traffic violations with a specific nature: alcohol consumption-driven traffic violations. We chose this particular subset in order to provide sound and conclusive insight on how to potentially reduce those traffic violations, which are responsible for a significant amount of deaths and injuries.

This project is divided in three sections: in the first one, we will provide some preliminary information in order to describe the dataset and explore how traffic violations are distributed considering different dimensions. Subsequently, we will proceed to overlay traffic violations with bars serving alcohol, as a means to show potential explanations to the nature and number of these traffic violations. Similarly, we will also overlay metropolitan transportation stops in order to assess the relative proximity of these stops to the bars whose attendants seem to incur in a high number of traffic violations. Finally, we will provide conclusions and guidelines on possible means to optimise the public transportation stop layout and transportation frequency in order to possibly reduce the number of traffic violations caused by alcohol consumption.

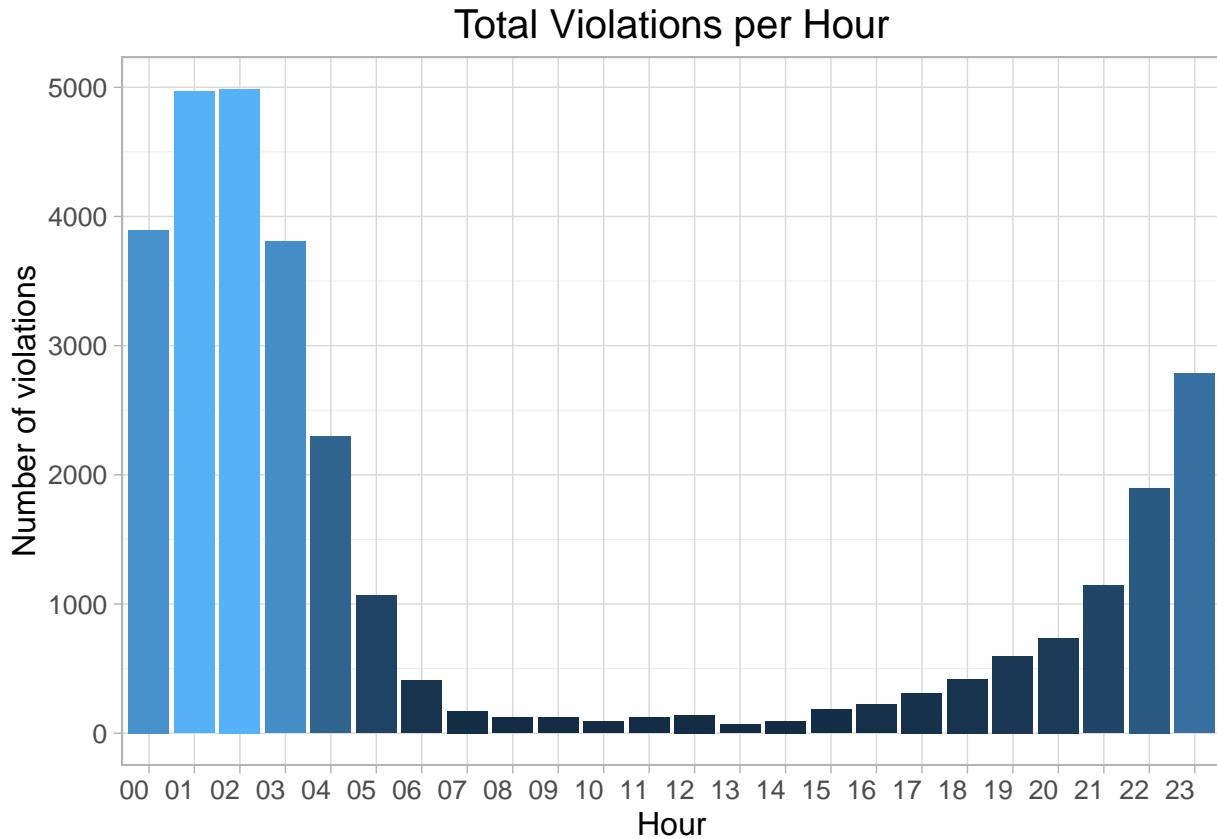
The Montgomery County Data Set

From the entire dataset, and as depicted in the plots below, we will only focus on alcohol-induced traffic violations, which only constitute 3.6% of the entire dataset. Even though this proportion might seem small, in subsequent sections we will show that the volume of data is adequate to provide insight on the current situation in the county of Montgomery.



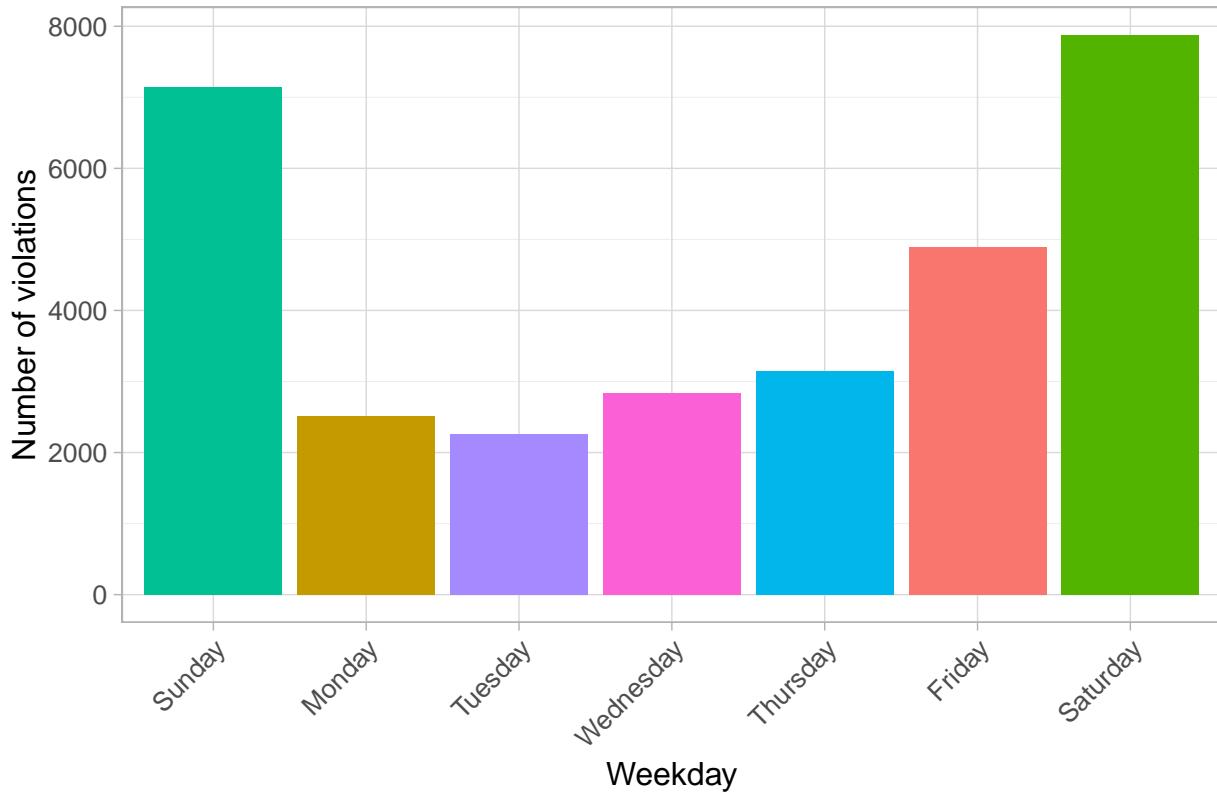
This family of traffic violations actually accounted for 9 deaths and almost 900 injured people since 2012, as the second plot above shows. This decrease in the number of injured might suggest a reduction on the number of accidents. However, as the third plot shows, the number of traffic violations triggered by alcohol consumption has been steadily increasing over year, and the trend for 2016 seems to go in the same direction. Consequently, we consider that tackling ways to reduce this number is not only reasonable but also desired, as the number of injured people remains high.

In order to understand the nature of these traffic violations, we decided to analyse the time of occurrence of this violations, considering three different axes: **day**, **time of the day** and the combination of both axes (**time of the day over each day**, displayed as a trellis plot). All three plots are displayed below:

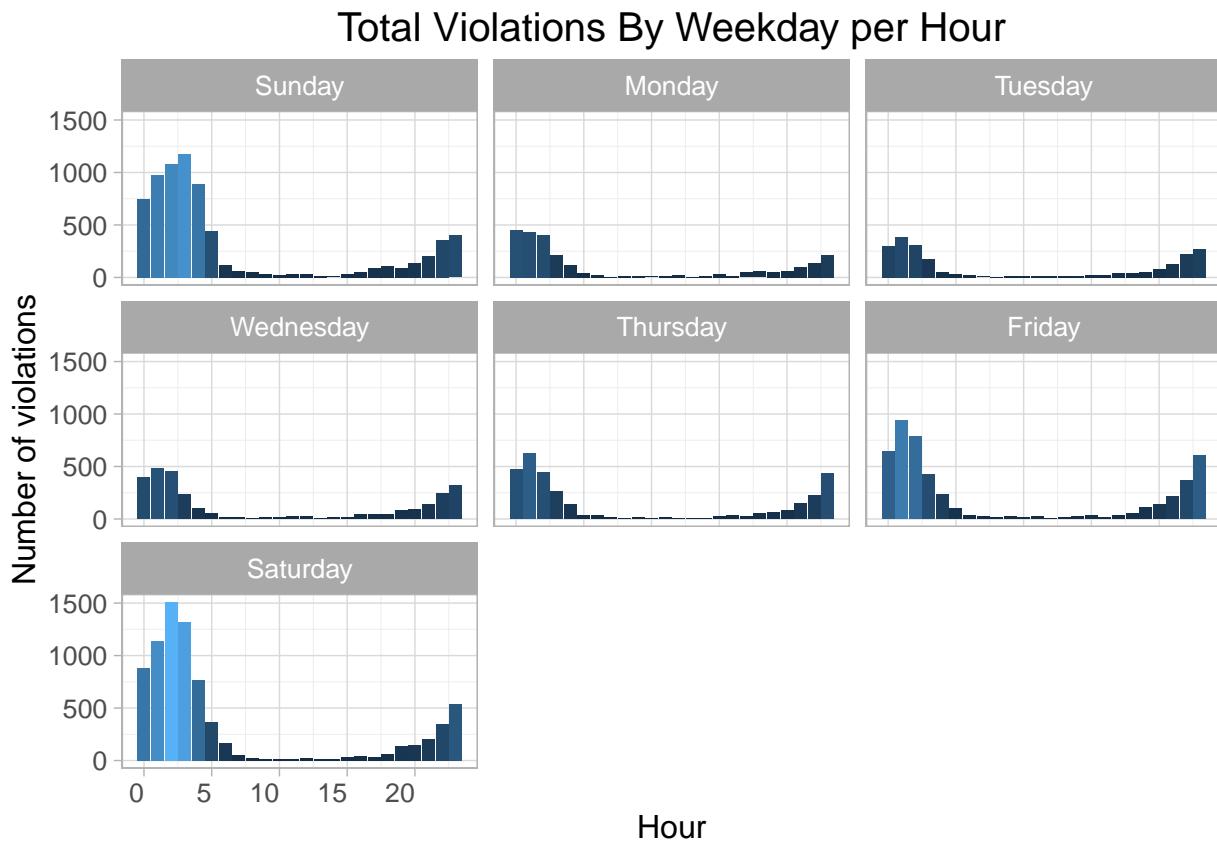


It can clearly be seen that late night and early morning hours present the highest proportion of traffic violations, which immediately suggest nightlife activity as the main root for these traffic violations. This is also supported by the following plot, which shows that most traffic violations occur during weekend days (that is, Friday, Saturday and Sunday).

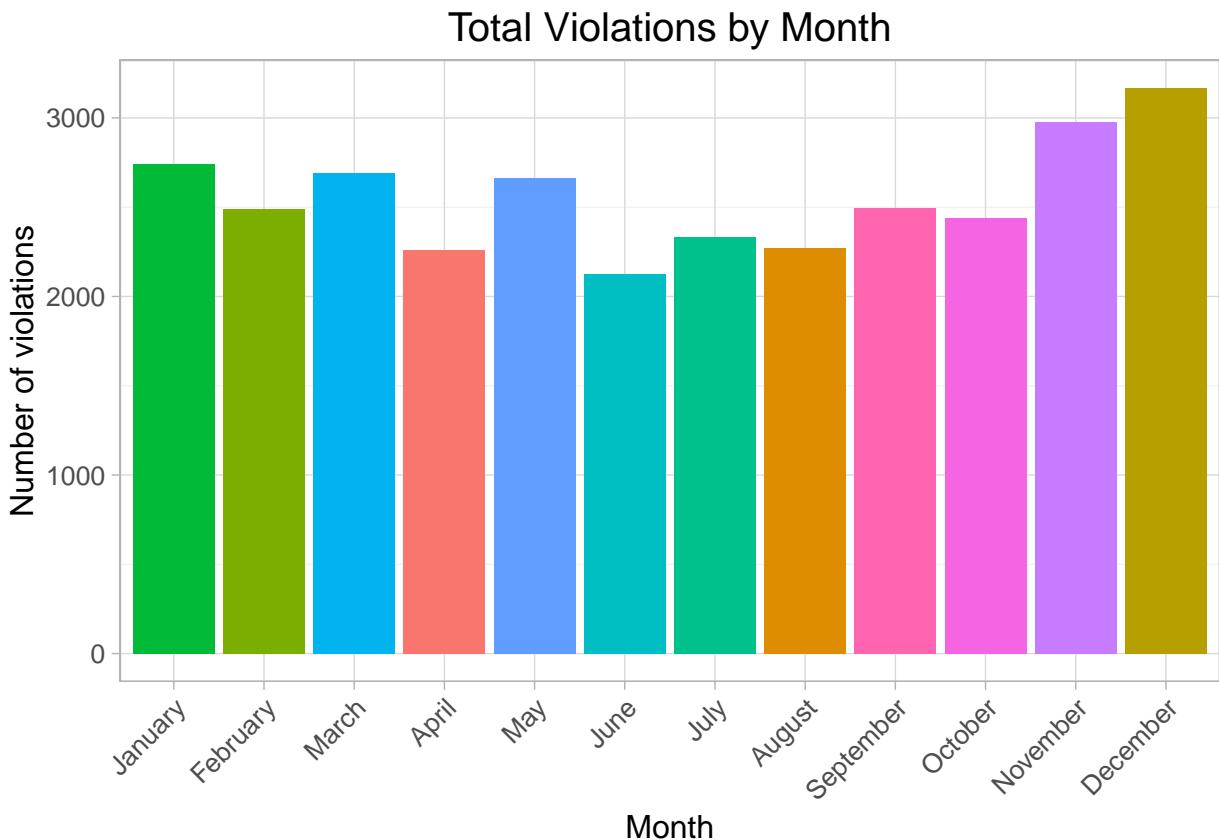
Total Violations by Weekday



Since this information is not enough to actually conclude that this global pattern is also local, that is, that there is no special day where traffic violations occur at night, we decided to display a Trellis plot that breaks the previous information on a day-by-day basis:



We can clearly see, then, that this behaviour pattern (traffic violations occurring during late night and early morning hours) is repeated throughout the entire week, almost the highest proportion can be found in weekend days. As a final step, it is important to be able to discern if the pattern occurs during the entire year. If so, then we can actually conclude that nightlife during weekends is indeed the main root of these violations.

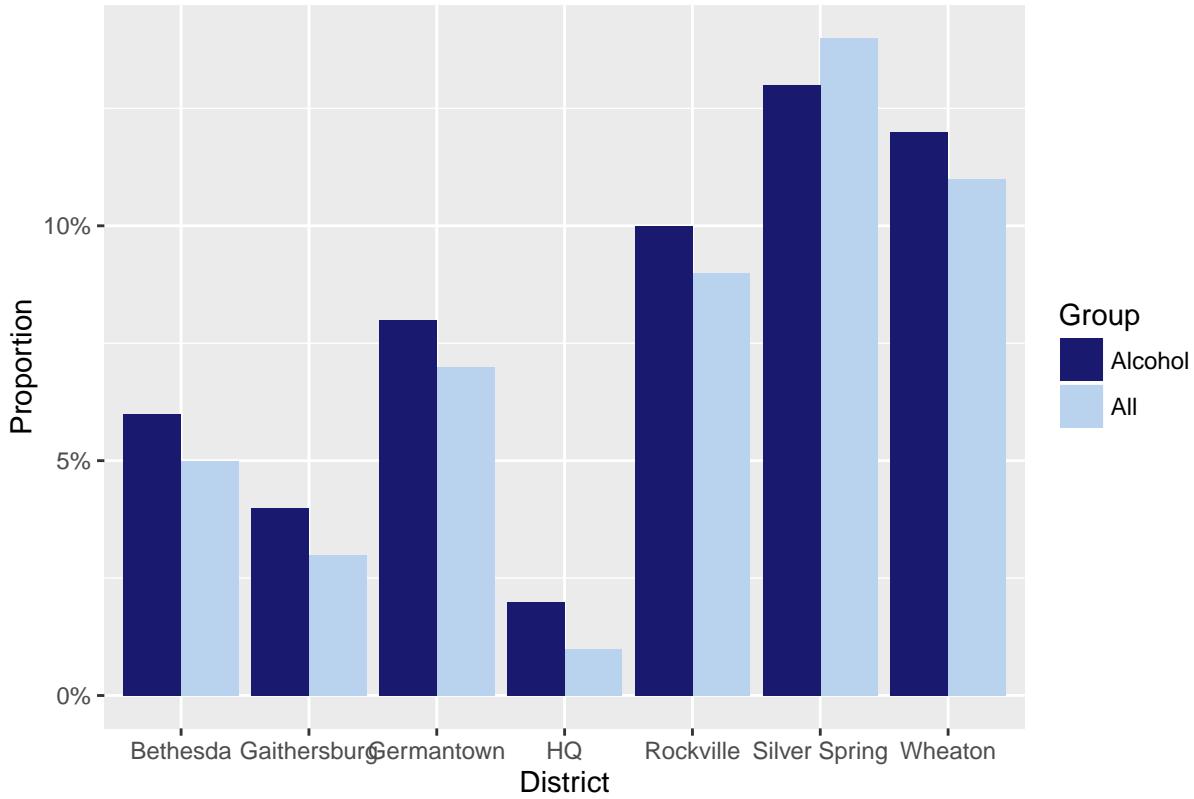


In the plot above, even if we see a slight increase in traffic violations during the months of November and December, the number of violations per month do not differ significantly. Hence, we can conclude that applying measures during the weekends will take effect the entire year, which is a more than desirable characteristic for any measures we can suggest.

Geographical Exploration of Traffic Violations

The second part of our story focuses on geographical aspects of the phenomenon we are exploring. At first, we wanted to take a look at the distribution of the traffic violations among the administrative territorial districts that are called **police districts** in the dataset we are working with. There are 7 main districts in Montgomery county and, obviously, they demonstrate different frequency of violations ?????both overall and alcohol-related ones.

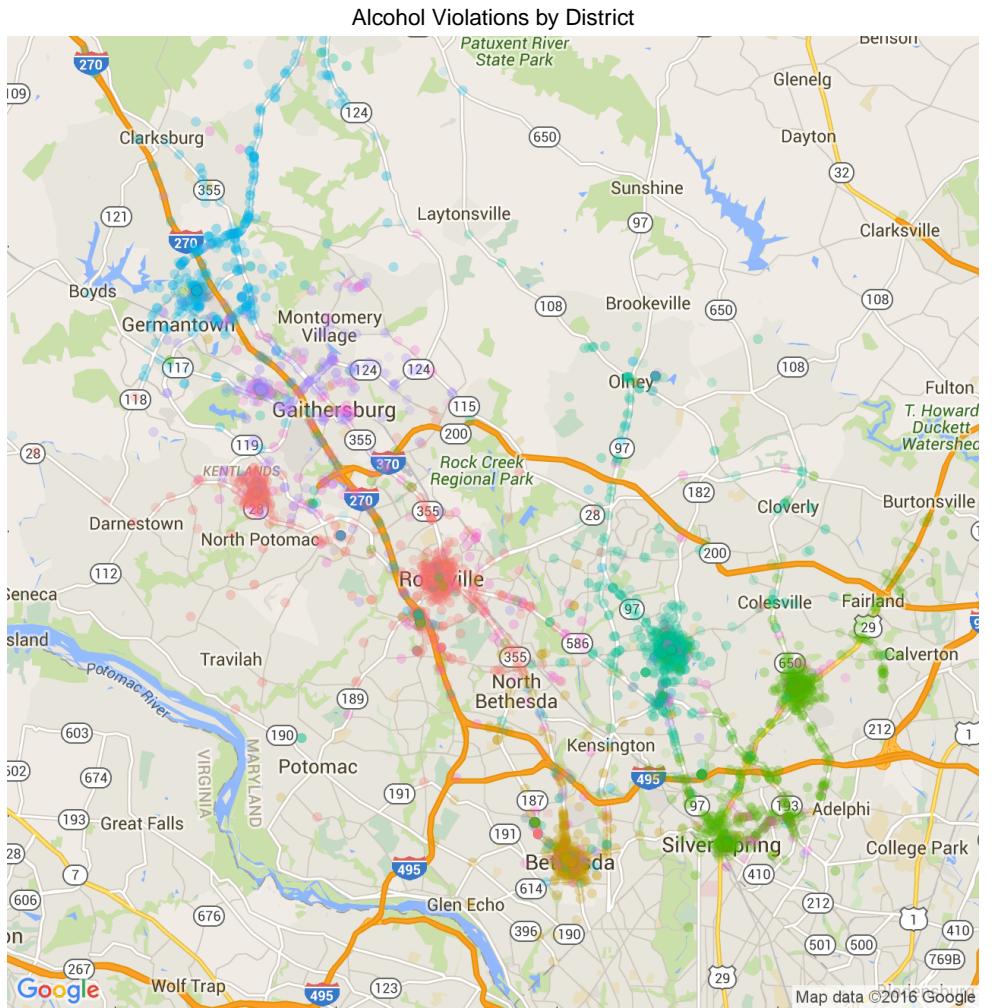
Violations by Police District



The direct comparison displayed by the chart above demonstrates two main take-aways that we are going to use later on:

- Some districts indeed have higher frequency of traffic violations, and alcohol-related violations follow the same distribution across districts in general. This might be an indicator of their relative size, meaning population and traffic in these districts, but it also may point out certain public places that affect this distribution, as well as their respective locations within the districts.
- *Silver Spring* is the only district that had smaller proportion of alcohol-related violations than the overall pool of observed police records. All other districts show relatively similar ratios, and alcohol violations are seemingly redistributed from *Silver Spring* to all other districts. This fact leads us to two more hypotheses we would like to check: firstly, whether *Silver Spring* has indeed fewer public places, eventually “losing” alcohol violations to neighboring districts. Secondly, whether this district has higher traffic density, which causes more general violations, unrelated to driving drunk in general.

We dive deeper into exploring the geographical structure of alcohol violations distribution by plotting them on a map:

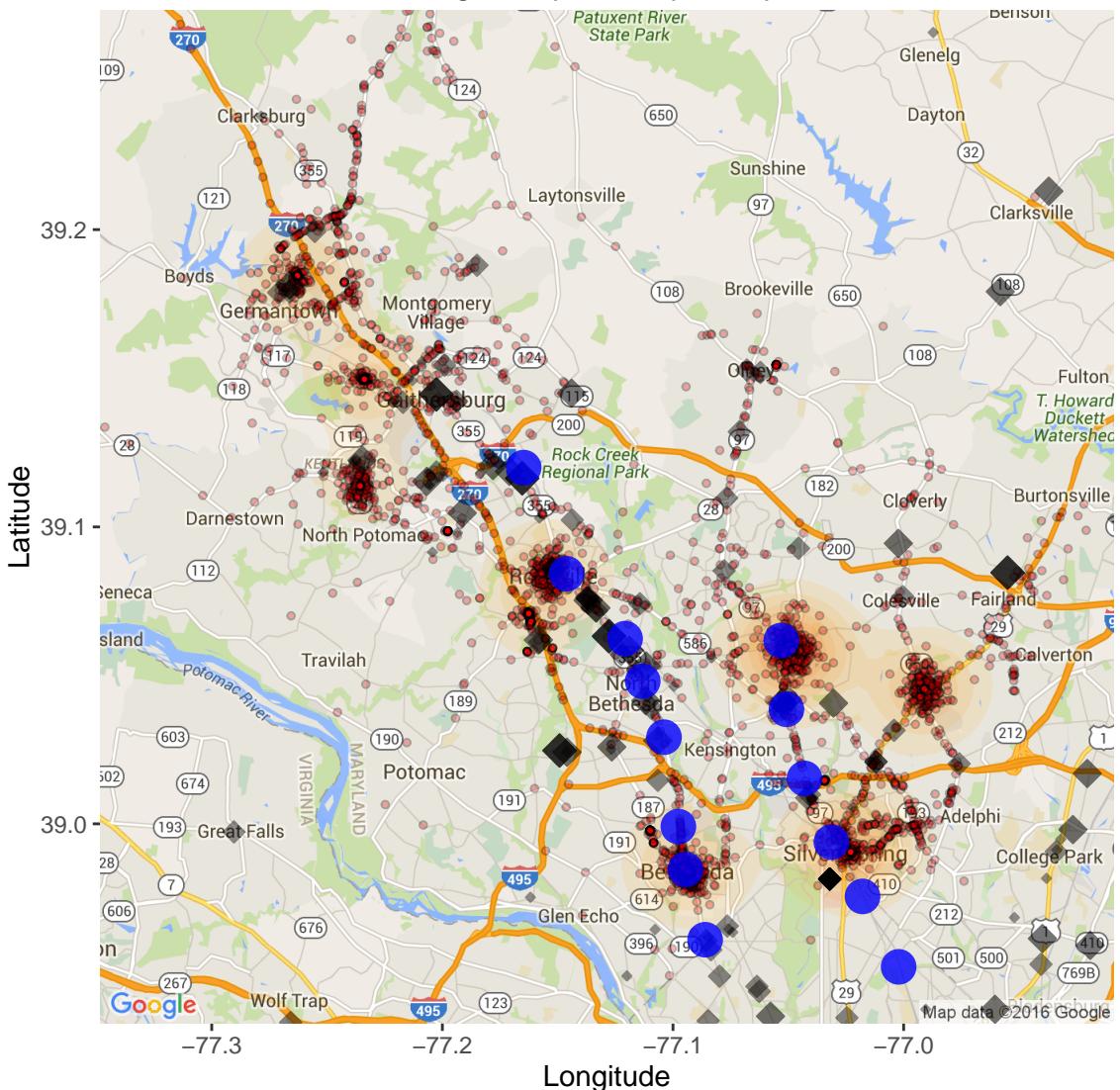


The very first thing that we directly observe from this map is the fact that **violations tend to cluster around certain points**. What is more, **each district has their own centers of gravity**, which we will try to discover further on.

As a possible preliminary explanation, this clustering of traffic violations could be related to the locations of bars, pubs and other drinking houses in the area. As a sidenote, we obtained these locations by scraping the public *Yelp API*. Especifically, we queries for bars and restaurants in Montgomery county, Maryland, with *alcohol* as a keyword. Additionally, subway station locations have been added to the map in order to better understand commuting patterns in the area.

Black diamonds represent public places, where the size of each diamond stands for the rating of this place on *Yelp* ??? the proxy variable for the popularity of the place that we decided to use in our analysis.

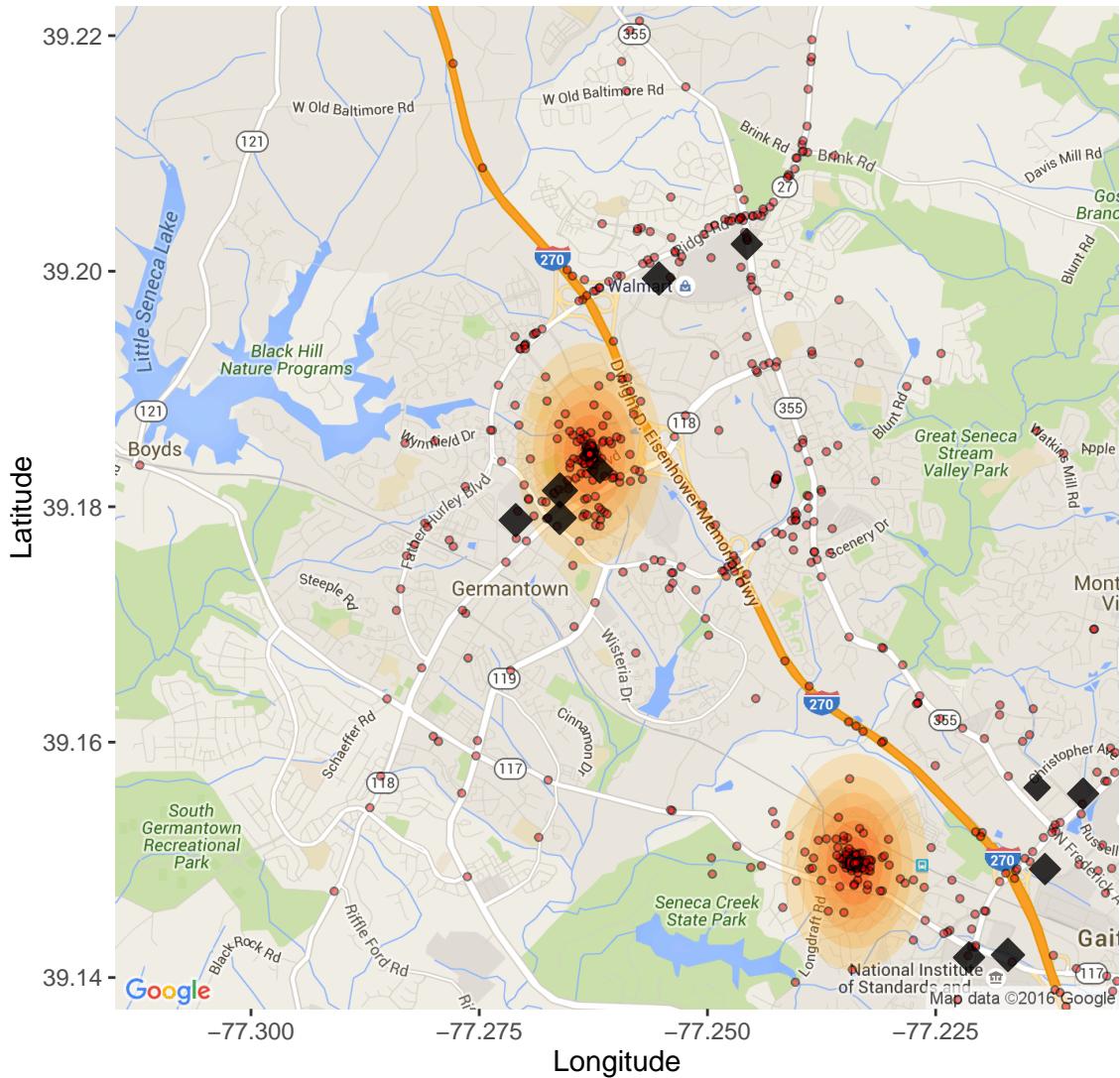
Bars, metro stations and alcohol violations in Montgomery County, Maryland



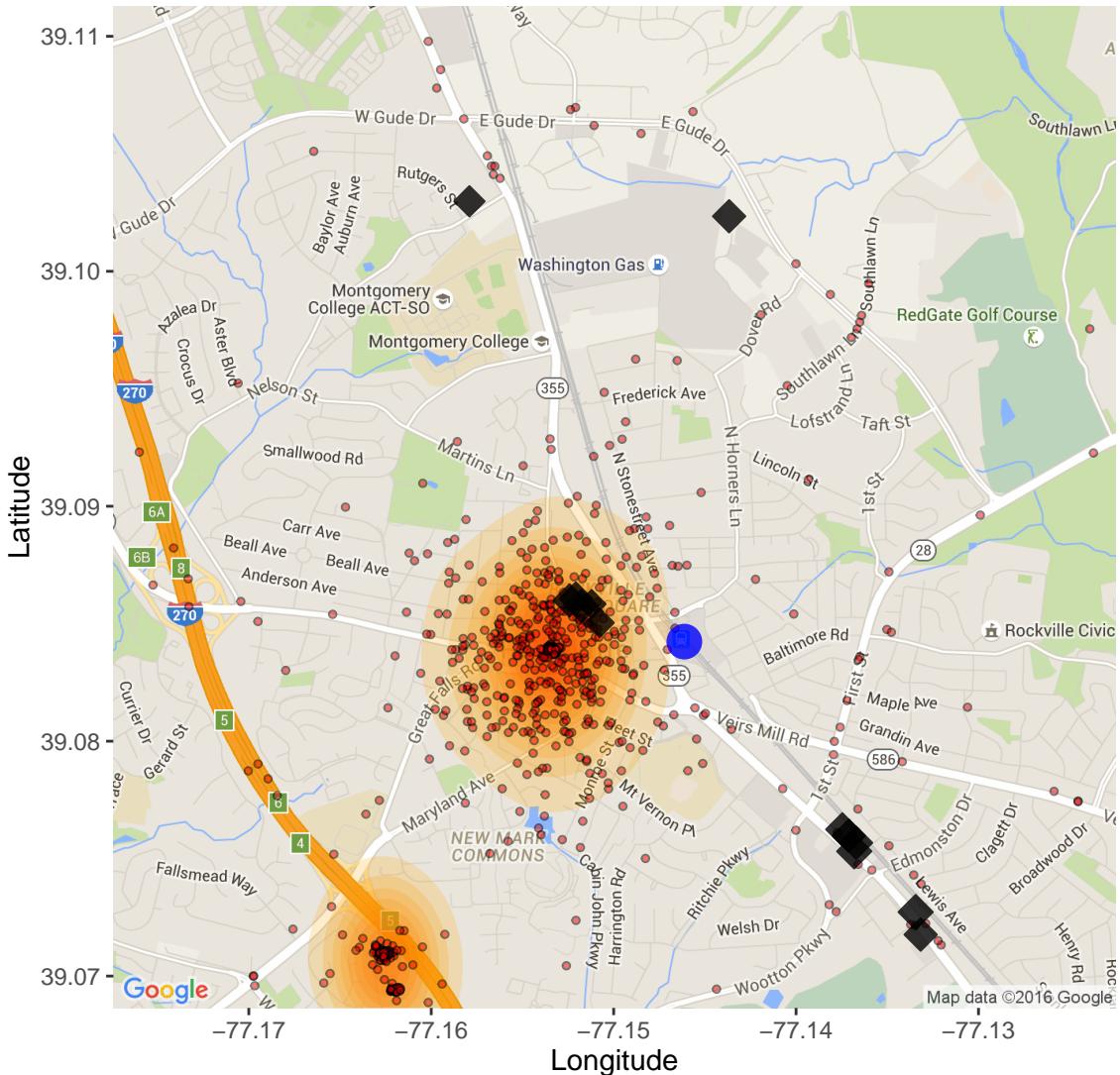
● Incident ♦ Bar ● Metro Station

In general, the map suggests that those violation clusters indeed correlate with certain popular public places and transportation stations. A closer look at different areas provides additional insight:

Bars, metro stations and alcohol violations in Germantown, Maryland

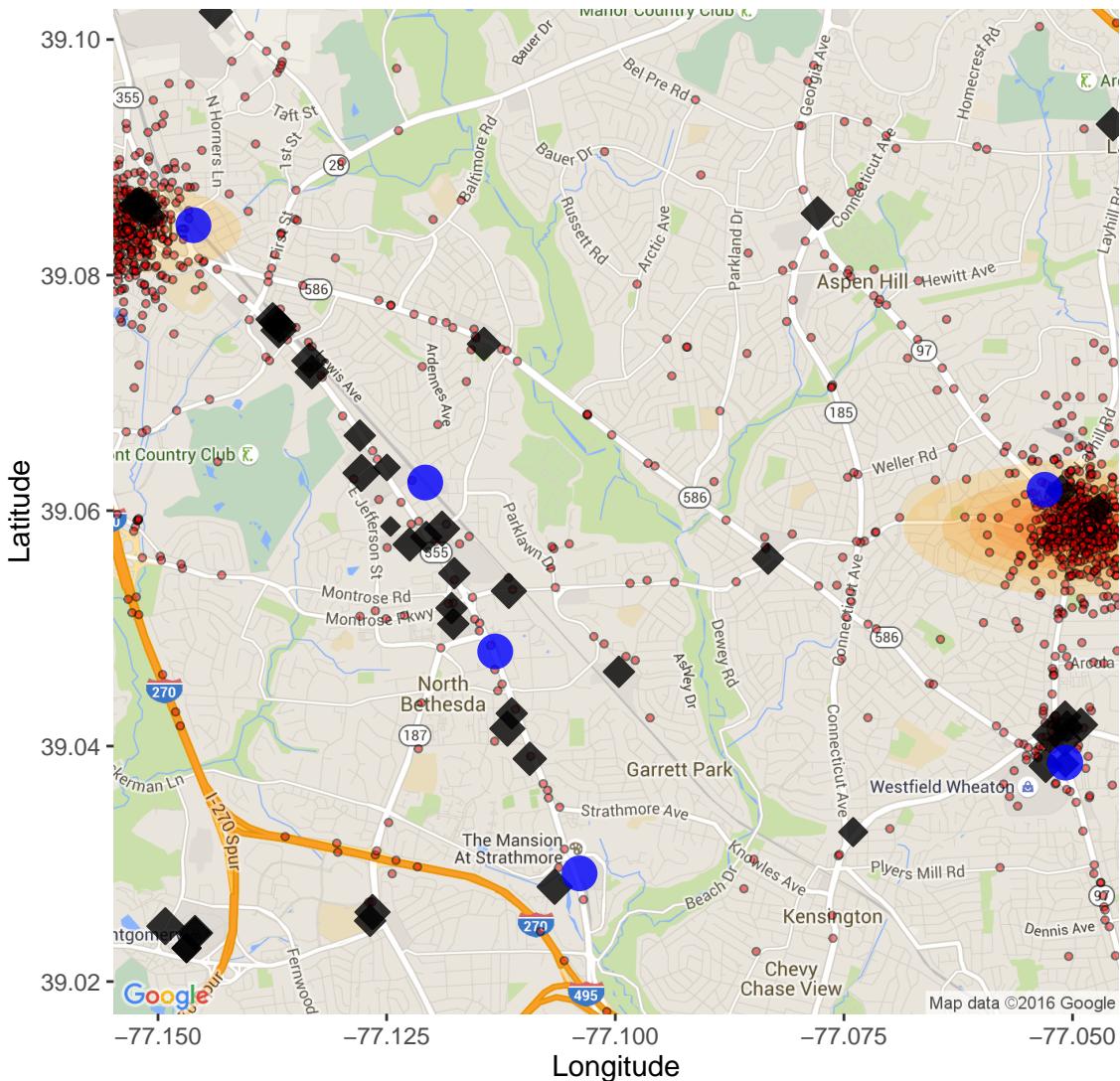


Bars, metro stations and alcohol violations in Rockville, Maryland

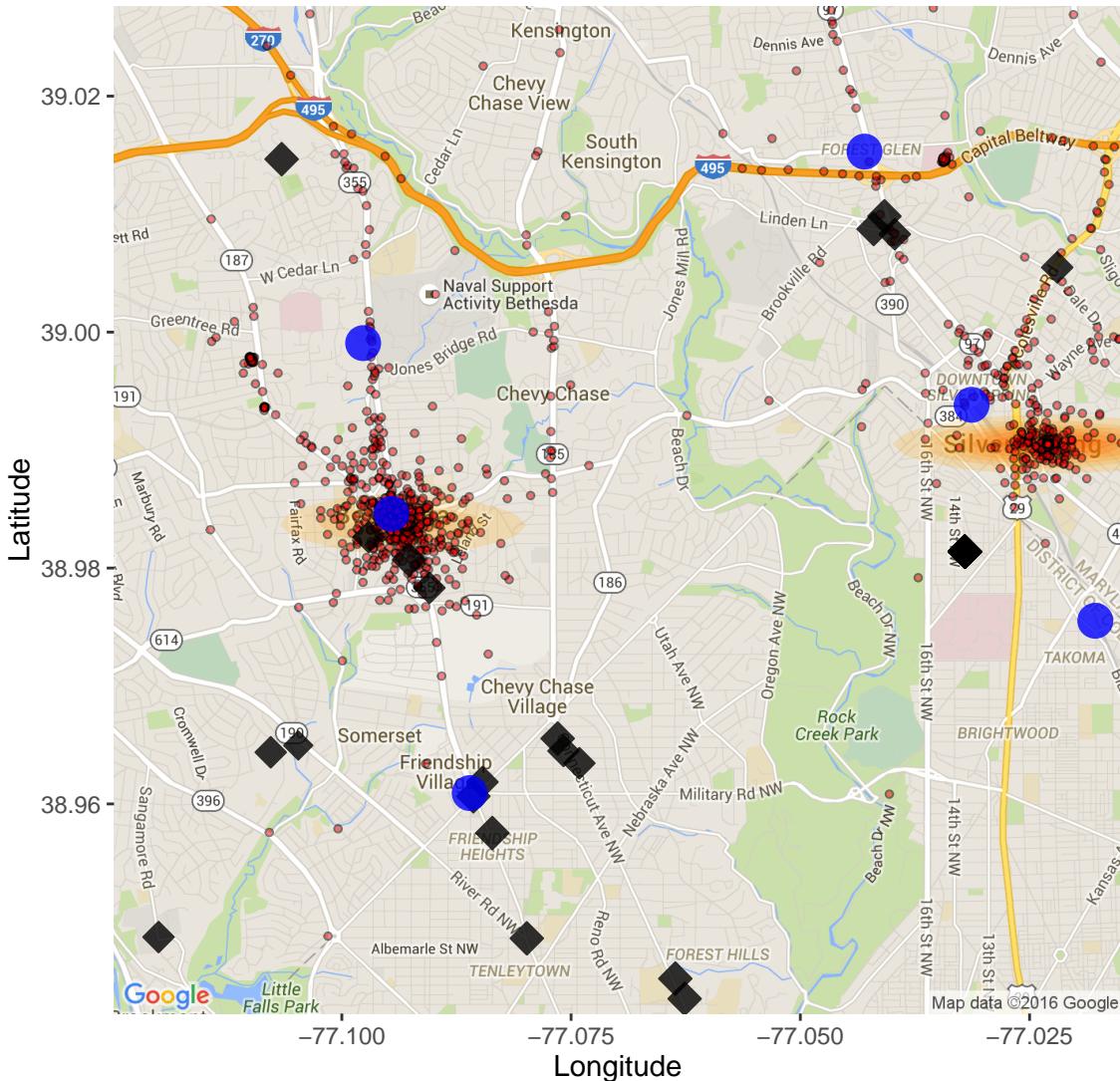


● Incident ♦ Bar ● Metro Station

Bars, metro stations and alcohol violations in North Bethesda, Maryland



Bars, metro stations and alcohol violations in Bethesda, Maryland



● Incident ♦ Bar ● Metro Station

Throughout these maps we actually find 3 different types of clusters based on the objects around them:

1. **Clusters with bars around:** the most common pattern on the map, which supports our hypothesis that the majority of alcohol-related incidents are registered in the immediate neighborhoods of public drinking places. As expected, bars with higher rating attract more people and are placed in the more convenient locations. Consequently, they attract more people who may, eventually, violate traffic laws by driving back home after drinking.
2. **Clusters with bars and subway stops around:** when public commute stations and bar places coincide in the same area, the cluster of alcohol violations becomes heavier, following what was predicted by the previous hypothesis. The explanation for this pattern could be the fact that a considerable amount of people who reside in the county actually work in *Washington, DC* and commute there on a regular basis. Their most probable route involves driving from their homes to the nearest railroad stations towards *DC*, leaving their cars on nearby parking lots and changing for a train to the city. Hence, once they are coming back after a night out in *Washington*, they take cars and drive home with alcohol in their blood, violating traffic laws. Police is certainly aware of this pattern and keeps patrolling these spots as the size of this family of cluster is very significant.
3. **Unidentified Clusters:** finally, there is a cluster in *Germantown* with seemingly few bars in the nearest surroundings, though several of them are located East of the road where these violations were registered. We do not have data to support this hypothesis, but the most reasonable explanation is that those drivers were heading out from these bar places towards the nearest highway to reach their homes after their time out. Road police was patrolling this road as it could be presumably the most convenient patrolling spot for them to prevent traffic violations in this district.

All these suggestions are coming directly from the visual facts that are discovered through this brief analysis. However, proving or disproving them requires more complex analysis which goes beyond the exploratory analysis covered in this paper.

Conclusion

Throughout the analysis we revealed a set of certain spatial and temporal patterns in alcohol-related traffic violations observed in residential counties in the US based on the example of Montgomery county, Maryland.

Main conclusions contain the following points:

- There is little variation of alcohol violations registered per month, and all the increases, though slight, are observed during the months with public holidays in them.
- The distribution of the violations by day of the week and by hour is easily explained by the most common social behavior during the week span – citizens tend to go out more in the night hours on Fridays and Saturdays.
- Bar locations are indeed the places where the biggest portion of alcohol violations are caught, and police is perfectly aware of these hot points and has set up patrolling units in accordance to this information. Subway/railroad stops are also the points of attention for police patrols. However, these stops almost always overlay with bar locations and places of public attraction, thus increasing an already high number of traffic violations in those areas.
- There is a lack of public transportation from 12AM to 5AM. This time period actually coincides with the hours where most tickets are issued. Though on the previous section we fine-grained the analysis on the clusters so as to show that this lack of public transportation might not be the exclusive reason for the occurrence of alcohol-related violations, we still consider that providing night-shift service on those routes will certainly diminish the amount of incidents.