

Project Report -

Traffic Violations

November 13, 2016

Group 4 SDM:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gaurav Nagar Himanshu Agrawal |  | U56247298  U80003633 |  | gauravnagar@mail.usf.edu  himanshu2@mail.usf.edu |
| Shifa Chaudhari |  | U23914104 |  | shifa@mail.usf.edu |
| Pratul Agarwal |  | U56790213 |  | pagarwal1@mail.usf.edu |
| Larry Rios |  | U37650809 |  | lrios2@mail.usf.edu |

CONTENTS

[1. ABSTRACT 3](#_Toc466819135)

[2. INTRODUCTION 4](#_Toc466819136)

[3. DATASET 5](#_Toc466819137)

[3.1 General assessment of data quality: 5](#_Toc466819138)

[3.2 Identification of Data Quality Issues 5](#_Toc466819139)

[4. DESCRIPTION OF THE PROBLEM STATEMENT 6](#_Toc466819140)

[4.1 Time-series analysis 6](#_Toc466819141)

[4.2 Impact of Driver’s City 7](#_Toc466819142)

[4.3 Area analysis 7](#_Toc466819143)

[4.4 Risk Analysis 7](#_Toc466819144)

[5. ANALYSIS AND VISUALIZATION 9](#_Toc466819145)

[6. CONCLUSIONS 35](#_Toc466819146)

[7. FUTURE WORK 36](#_Toc466819147)

[8. REFERENCES 37](#_Toc466819148)

# ABSTRACT

In the year of 2015, there were 253 million vehicles in the United States. With so many vehicles in the streets, the number of traffic violations and accidents increased over time. A number of accidents can be very hazardous especially around the high-frequency areas. Such areas witness a huge number of traffic violations which inadvertently tend to lead to more accidents. Hazardous road traffic locations also called black spots defines the intersections or main areas where traffic accidents are relatively high.

Another reason why traffic accidents occur is because traffic policies vary from state to state in the United States. Road traffic safety deals with a complexity of problems since there are various reasons that contribute to its disturbance: blind curve, sharp corner, the width of roads, driver’s behavior in traffic, wear and tear on infrastructure, climate scenarios, and light conditions and intense traffic. With increasing road traffic congestions, safety becomes harder to maintain due to factors such as harsh weather conditions or driving under the influence of alcohol.

# INTRODUCTION

Through this project, our primary objective was to curb these traffic accidents by deriving some important information from the traffic violations data. This project was particularly important for us because it has the potential to touch human lives in every significant manner. For example, analyzing the traffic violations for a city/county/state we can derive some vital information about the violations around schools during prime hours. It is not safe to have high traffic violations around sensitive areas; therefore, the analysis presented in this project allow us to inform the policy makers to alter the current policies if needed.

Another important domain significantly affected by these traffic violations is the nation’s economy. Per a recent study, when a city is experiencing a hard economic cycle, they turn to traffic violations as a source of revenue. A 10% decrease in economic growth leads to an average 6.5% increase in the number of speeding tickets issued during the same period. Major regions like New York State brings in about $76 million in revenue every year in traffic tickets. We think it will be interesting to verify this information by evaluating data from the state of Maryland and reaching conclusions that will proof such findings.

While evaluating the data from the state of Maryland, we can also find out if more tickets are handed out at the end of each month. Also, does the policeman hand out fewer tickets in the rainy season as compared to the other seasons like winter and summer? These kinds of questions can help us derive an analysis to reach conclusions which will help us determine steps to fix a set of problems.



# DATASET

The dataset contains traffic violation information from all electronic traffic violations issued in Montgomery County(Maryland) and was extracted from the Montgomery County of Maryland website (data.montgomerycountymd.gov). It includes all traffic violations since January 1 of 2012, so this analysis is based on 49 months of records. There are 35 attributes and a total 819565 instances on this dataset.

The main attributes that will be used for the data mining process are Location, Geo-location, Accident, Alcohol, Work Zone, Year, Violation Type, Charge, Gender, and Race. The attributes geolocation and location will be used to visualize in a geospatial data visualization.

## **3.1 General assessment of data quality:**

The data obtained from the dataset was of very mediocre quality. Although there weren’t any grave quality issues, a few small to mediocre issues needed to be addressed. Some of the same areas mentioned below:

a)      Missing/blank values

b)      Outliers

c)      Shifted data

# **3.2 Identification of Data Quality Issues**

Data Quality of the obtained dataset was meddling and there were a few actions that were carried out to bring successful analysis on the data. For example, The Date and Time Columns were in a format were in such a way that couldn’t be used to data analysis. Hence, we formed out subsets and worked upon it.

# DESCRIPTION OF THE PROBLEM STATEMENT

When we talk about the road accidents we can see that major reasons which mostly leads to the accidents is the violations in traffic. Danger related with the accidents is not only the property damage as most of the fatal accidents can lead to personal injuries or even the loss of life.

**How can we work on reducing the traffic violation?**

This project is an approach to analyzing the major factors behind most of the traffic violation in an area. We are trying to see the influence of the various demographic factors over the number of violations. So, that based on this study if we can come up with some good suggestions that can help in reducing the number of traffic violations encountered each day.

Below are few of the problem statement on which we have focused on our analysis:

## **4.1 Time-series analysis**

* We can extract the time periods (for a day) during which the traffic violation is highest at a location/s.
  + Try to show the visualization of violations and try to find out during which time of day it is highest and lowest.
  + We can again concentrate it based on area (at what time we have more violations in which area)
* Given a specific month in a year, when is Maryland the most dangerous state for tourist to visit??
* What type of violations is increasing over time? Does this mean that the current Maryland policies are not working towards curbing them?

## **4.2 Impact of Driver’s City**

* Is there a correlation between the driver’s city and the type of violation that the driver was involved in?
* What are the chances of a driver not from Maryland to be involved in a violation in Maryland?
* Drivers from outside state that violated traffic rules?

## **4.3 Area analysis**

* Based on the type and amount of violations, which area in Maryland requires a higher concentration of policeman.
* What sub-Agency in Maryland are the most dangerous?
* Based on the current policies, which areas (sub-Agencies) are softer in their approach towards violations?
* Based on the sub-Agency, we can try to show if any race has more violation in any area.
* Geo-location representation showing point distribution and density of traffic violations incurred.

## **4.4 Risk Analysis**

* Are you more likely to be involved in a violation if you are driving an older vehicle?
* For what kind of violations was there property damage (& its relation with Race and Gender)?
* Which type of vehicle make/model and the color/make was involved in most accidents?
* Will cars with brighter color have more violations?
* Which charge dominates among all violations?
* Based upon the type charge (Ex: Wireless devices), which gender has the most accidents?
* Check the demographic map and its effect on the violations. (Visualization for prevention)

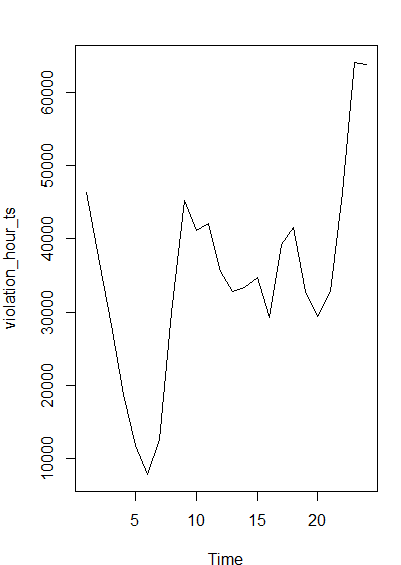
# ANALYSIS AND VISUALIZATION

* **During which hour of the day the traffic violations are highest or lowest**

|  |  |
| --- | --- |
| **Hour** | **Frequency of Violation** |
| Midnight - 1 am | 46270 |
| 1 is - 2 am | 36761 |
| 2 am - 3 am | 28185 |
| 3 am - 4 am | 18741 |
| 4 am - 5 am | 11820 |
| 5 am - 6 am | 7821 |
| 6 am - 7 am | 12671 |
| 7 am - 8 am | 29898 |
| 8 am - 9 am | 45192 |
| 9 am - 10 am | 41190 |
| 10 am - 11 am | 42099 |
| 11 am - Noon | 35483 |
| Noon - 1 pm | 32807 |
| 1 pm - 2 pm | 33360 |
| 2 pm - 3 pm | 34697 |
| 3 pm - 4 pm | 29205 |
| 4 pm - 5 pm | 39211 |
| 5 pm - 6 pm | 41559 |
| 6 pm - 7 pm | 32866 |
| 7 pm - 8 pm | 29387 |
| 8 pm - 9 pm | 32828 |
| 9 pm - 10 pm | 46365 |
| 10 pm - 11 pm | 64097 |
| 11 pm - 12 pm | 63728 |

**Figure: Number of violations in each hour**

From our dataset, we can find that there is a high frequency of violations during 10:00 pm to midnight and after the midnight we can see the number of violations starts to decrease gradually. A minimum number of violations have taken place during between 5:00 am to 6:00 am. We have also noticed that the number of violation have started increasing from morning 8:00 am.



**Figure: Time – Series graph between the Number of violations v/s hours each day**

Adding up to the new analysis we have tried to find the pattern for the above violations in different areas, our analysis on the area along with the time shows that despite more violation during the peak hours there are few areas where the number of violation incidents found is less during the same period. Based on this we can suggest that the same kind of analysis can be performed over all areas and extra policemen can be appointed in those areas during peak hour or there can be the incorporation of the new signals or other safety measures which can help in reducing some amount of violations there.

The table below is used to show the distribution of the number of violation in each area during the different period of the day.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Area/Hours** | **Midnight - 1 am** | **1 am - 2 am** | **2 am - 3 am** | **3 am - 4 am** | **4 am - 5 am** | **5 am - 6 am** |
| **Bethesda** | 7451 | 5366 | 3990 | 2687 | 1823 | 1340 |
| **Gaithersburg / Montgomery Village** | 4794 | 4801 | 3125 | 2087 | 1146 | 750 |
| **Germantown** | 5789 | 5779 | 4489 | 2952 | 1569 | 597 |
| **Headquarters and Special Operations** | 2661 | 2557 | 2192 | 868 | 294 | 194 |
| **Rockville** | 4643 | 4284 | 2986 | 1471 | 835 | 533 |
| **Silver Spring** | 11093 | 7680 | 5927 | 4242 | 3109 | 2201 |
| **Wheaton** | 9839 | 6294 | 5476 | 4434 | 3044 | 2206 |
|  | | | | | | |
| **Area/Hours** | **6 am - 7 am** | **7 am - 8 am** | **8 am - 9 am** | **9 am - 10 am** | **10 am - 11 am** | **11 am - Noon** |
| **Bethesda** | 2475 | 5835 | 9064 | 5679 | 5001 | 4071 |
| **Gaithersburg / Montgomery Village** | 1842 | 4241 | 7061 | 7063 | 7223 | 5660 |
| **Germantown** | 844 | 3115 | 4776 | 4382 | 4202 | 3698 |
| **Headquarters and Special Operations** | 560 | 684 | 802 | 866 | 926 | 1162 |
| **Rockville** | 1416 | 3075 | 5072 | 5394 | 5371 | 4358 |
| **Silver Spring** | 2753 | 5840 | 8354 | 8133 | 8538 | 6975 |
| **Wheaton** | 2781 | 7108 | 10063 | 9673 | 10838 | 9559 |
|  | | | | | | |
| **Area/Hours** | **Noon - 1 pm** | **1 pm - 2 pm** | **2 pm - 3 pm** | **3 pm - 4 pm** | **4 pm - 5 pm** | **5 pm - 6 pm** |
| **Bethesda** | 3669 | 4025 | 4396 | 3731 | 6310 | 6981 |
| **Gaithersburg / Montgomery Village** | 5251 | 4960 | 4123 | 3128 | 3900 | 4483 |
| **Germantown** | 3680 | 3529 | 3375 | 3100 | 4420 | 5289 |
| **Headquarters and Special Operations** | 1225 | 1291 | 1522 | 1447 | 1264 | 945 |
| **Rockville** | 4249 | 4405 | 4485 | 4058 | 4722 | 5011 |
| **Silver Spring** | 6491 | 6628 | 7160 | 6539 | 7800 | 7453 |
| **Wheaton** | 8242 | 8522 | 9636 | 7202 | 10795 | 11397 |
|  | | | | | | |
| **Area/Hours** | **6 pm - 7 pm** | **7 pm - 8 pm** | **8 pm - 9 pm** | **9 pm - 10 pm** | **10 pm - 11 pm** | **11 pm - 12 pm** |
| **Bethesda** | 4254 | 3341 | 3586 | 6381 | 9989 | 10192 |
| **Gaithersburg / Montgomery Village** | 4308 | 4020 | 3620 | 5742 | 7960 | 6641 |
| **Germantown** | 4541 | 3548 | 3516 | 5105 | 7313 | 7506 |
| **Headquarters and Special Operations** | 1073 | 1065 | 1435 | 1849 | 2295 | 2900 |
| **Rockville** | 3868 | 3078 | 3820 | 5430 | 6776 | 6549 |
| **Silver Spring** | 5831 | 6175 | 7526 | 9957 | 14605 | 14975 |
| **Wheaton** | 8991 | 8160 | 9325 | 11901 | 15159 | 14965 |

**Figure: Distribution of number of violation in different area at different time-period**

**Code:**

rxViolations **<-** read.csv**(**file.choose**())**

head**(**rxViolations**)**

rxViolations**$**hour\_of\_day **<-** substr**(**rxViolations**$**Time.Of.Stop,1,regexpr**(**":",rxViolations**$**Time.Of.Stop**)-**1**)**

head**(**rxViolations**$**hour\_of\_day**)**

class**(**rxViolations**$**hour\_of\_day**)**

rxViolations**$**hour\_of\_day **<-** as.numeric**(**rxViolations**$**hour\_of\_day**)**

violation\_hour **<-** as.data.frame**(**table**(**rxViolations**$**hour\_of\_day**))**

View**(**violation\_hour**)**

write.csv**(**violation\_hour, file **=** "violation\_hour.csv"**)**

rxViolations**$**SubAgency **<-** as.character**(**rxViolations**$**SubAgency**)**

rxViolations**$**SubAgency **<-** substr**(**rxViolations**$**SubAgency, regexpr**(**",",rxViolations**$**SubAgency**)+**2, nchar**(**rxViolations**$**SubAgency**))**

head**(**rxViolations**$**SubAgency**)**

violation\_area\_hour **<-** as.data.frame**(**table**(**rxViolations**$**SubAgency, rxViolations**$**hour\_of\_day**))**

View**(**violation\_area\_hour**)**

write.csv**(**violation\_area\_hour, file **=** "violation\_area\_hour.csv"**)**

* **During which month, most of the violations are found in Maryland.**

From our dataset, which include the details of violation for more than 3.5 years we have tried to find out during which month in a year there are more violations. Analysis of this dataset has shown that most of the violations have taken place in the month of May also March and April are the months which has a high probability of violations each year. We have also found that drivers are more careful in the starting month of the year which has resulted in comparatively less number of violation in the month of January and February.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Month/Year** | **2013** | **2014** | **2015** | **2016** | **Total** |
| **January** | 13098 | 15461 | 18256 | 16669 | 63484 |
| **Febuary** | 12909 | 15888 | 16395 | 19823 | 65015 |
| **March** | 15861 | 19572 | 20764 | 22521 | 78718 |
| **April** | 14842 | 21405 | 21421 | 20984 | 78652 |
| **May** | 18417 | 22601 | 20917 | 18734 | 80669 |
| **June** | 13554 | 17588 | 19160 | 20033 | 70335 |
| **July** | 15748 | 19574 | 19431 | 17233 | 71986 |
| **August** | 16892 | 17821 | 21314 | 17643 | 73670 |
| **September** | 18170 | 18415 | 19985 | 16335 | 72905 |
| **October** | 17448 | 20154 | 19936 | 16677 | 74215 |
| **November** | 17027 | 18796 | 19899 | NA | 55722 |
| **December** | 16553 | 16446 | 17871 | NA | 50870 |

**Figure: Number of violations in each month every year**

**Code:**

rxViolations **<-** read.csv**(**file.choose**())**

rxViolations**$**month\_of\_year **<-** substr**(**rxViolations**$**Date.of.Stop, 1, regexpr**(**"/",rxViolations**$**Date.of.Stop**)-**1**)**

head**(**rxViolations**$**month\_of\_year**)**

class**(**rxViolations**$**month\_of\_year**)**

rxViolations**$**month\_of\_year **<-** as.numeric**(**rxViolations**$**month\_of\_year**)**

rxViolations**$**Date.of.Stop **<-** as.character**(**rxViolations**$**Date.of.Stop**)**

rxViolations**$**year **<-** substr**(**rxViolations**$**Date.of.Stop, nchar**(**rxViolations**$**Date.of.Stop**)-**3, nchar**(**rxViolations**$**Date.of.Stop**))**

head**(**rxViolations**$**year**)**

violations\_year\_ts **<-** ts**(**violation\_year**$**Freq**)**

plot**(**violations\_year\_ts**)**

violation\_month **<-** as.data.frame**(**table**(**rxViolations**$**month\_of\_year, rxViolations**$**year**))**

View**(**violation\_month**)**

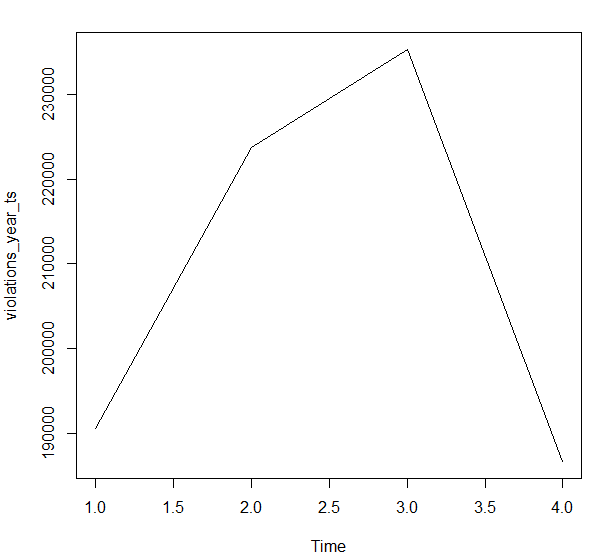
write.csv**(**violation\_month, file **=** "violation\_month.csv”)

* **The trend in a number of violations and type of violations over the year.**

|  |  |
| --- | --- |
| **Year** | **Number of Violations** |
| 2013 | 190519 |
| 2014 | 223721 |
| 2015 | 235349 |
| 2016 | 186652 |

**Figure: Number of violations every year**

An increasing trend is noticed in the number of violations every year with almost a quarter remaining in the year 2016 but a number of violations have already reached close to what registered in last years.



**Figure: Time – Series graph for Number of violations every year**

**Code:**

rxViolations **<-** read.csv**(**file.choose**())**

head**(**rxViolations**)**

rxViolations**$**Date.of.Stop **<-** as.character**(**rxViolations**$**Date.of.Stop**)**

rxViolations**$**year **<-** substr**(**rxViolations**$**Date.of.Stop, nchar**(**rxViolations**$**Date.of.Stop**)-**3, nchar**(**rxViolations**$**Date.of.Stop**))**

head**(**rxViolations**$**year**)**

violation\_year **<-** as.data.frame**(**table**(**rxViolations**$**year**))**

View**(**violation\_year**)**

write.csv**(**violation\_year, file **=** "violation\_year.csv"**)**

violations\_year\_ts **<-** ts**(**violation\_year**$**Freq**)**

plot**(**violations\_year\_ts**)**

* **Let’s try to find out which area is most dangerous in Maryland**

We have done the similar analysis to find that which is the most dangerous area in Maryland with respect to traffic violations.

|  |  |
| --- | --- |
| **Area** | **Number of Violations** |
| **Bethesda** | 121637 |
| **Gaithersburg / Montgomery Village** | 107929 |
| **Germantown** | 97114 |
| **Headquarters and Special Operations** | 32077 |
| **Rockville** | 95889 |
| **Silver Spring** | 175985 |
| **Wheaton** | 205610 |

**Figure: Time – Series graph for Number of violations every year**

In our analysis, we have seen that traffic violations are more dominant and frequent in ‘Wheaton’ are of Maryland while ‘Headquarters and Special Operations’ and Rockville are the areas with least number of violations. So, we feel that the extra rules can be enforced in the are with high violations to deal with this problem.

We can further enhance our analysis by trying to find out the reasons behind the more number of violations in this area as that may help us to find the effective solution for reducing the number of violations.

**Code:**

rxViolations **<-** read.csv**(**file.choose**())**

head**(**rxViolations**)**

rxViolations**$**SubAgency **<-** as.character**(**rxViolations**$**SubAgency**)**

rxViolations**$**SubAgency **<-** substr**(**rxViolations**$**SubAgency, regexpr**(**",",rxViolations**$**SubAgency**)+**2, nchar**(**rxViolations**$**SubAgency**))**

head**(**rxViolations**$**SubAgency**)**

violation\_area **<-** as.data.frame**(**table**(**rxViolations**$**SubAgency**))**

View**(**violation\_area**)**

write.csv**(**violation\_area, file **=** "violation\_area.csv"**)**

* **Analyzing data on districts**

|  |  |  |
| --- | --- | --- |
| **Number** | **Districts** | **Number of Violations** |
| 1 | 1st district, Rockville | 95889 |
| 2 | 2nd district, Bethesda | 121637 |
| 3 | 3rd district, Silver Spring | 175985 |
| 4 | 4th district, Wheaton | 205610 |
| 5 | 5th district, Germantown | 97114 |
| 6 | 6th district, Gaithersburg / Montgomery Village | 107929 |
| 7 | Headquarters and Special Operations | 32077 |

Based on the data collected, we can see that  4th district (Weather) is the most dangerous district in Maryland with a total amount of violations of 205610. It is recommended that the state of Maryland increases the amount of regulations in Wheaton to be able to decrease the number of violations over time.

* **Analyzing data on districts and Charges**

The information below shows the greatest number of violation by the town for a specific charge. For example, Wheaton has the greatest number of violations in Montgomery for the charge 21-801.1. This information is useful because we can provide the police enforcement in Wheaton with this information to be able to lower the number of violations in the future for a specific charge.

|  |  |  |
| --- | --- | --- |
| **Districts** | **Charges** | **Number of Violations** |
| 4th district Wheaton | 21-801.1 | 23439 |
| 1st district Rockville | 21-801.1 | 17660 |
| 5th district Germantown | 21-801.1 | 17264 |
| 2nd district Bethesda | 21-201(a1) | 16086 |
| 6th district Gaithersburg / Montgomery Village | 21-801.1 | 15922 |
| 3rd district Silver Spring | 21-201(a1) | 15397 |
| 3rd district Silver Spring | 21-801.1 | 15328 |
| 2nd district Bethesda | 21-801.1 | 14951 |
| 4th district Wheaton | 21-201(a1) | 14701 |
| 4th district Wheaton | 13-409(b) | 10740 |

* **Analyzing data on districts and Gender**

In the table below, we can see that for Wheaton, Silver Spring, Bethesda, and Gaithersburg males is the gender that generates the most violations. This is information is useful because a policeman can target males more often to prevent future violations. Also, policymakers can create different types of advertisement to aim males in Montgomery to prevent them from causing violations.

|  |  |  |
| --- | --- | --- |
| **Districts** | **Gender** | **Number of Violations** |
| 4th district Wheaton | M | 143091 |
| 3rd district Silver Spring | M | 120356 |
| 2nd district Bethesda | M | 76478 |
| 6th district Gaithersburg / Montgomery Village | M | 70519 |
| 5th district Germantown | M | 63103 |
| 4th district Wheaton | F | 62489 |
| 1st district Rockville | M | 61173 |
| 3rd district Silver Spring | F | 55128 |
| 2nd district Bethesda | F | 44994 |
| 6th district Gaithersburg / Montgomery Village | F | 37349 |
| 1st district Rockville | F | 34698 |
| 5th district Germantown | F | 34004 |
| Headquarters and Special Operations | M | 23801 |
| Headquarters and Special Operations | F | 8245 |

* **Which race is more involved in the violations in different areas.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Area/Race** | **ASIAN** | **BLACK** | **HISPANIC** | **NATIVE AMERICAN** | **WHITE** | **OTHER** | **Total** |
| **Bethesda** | 7258 | 27466 | 16565 | 435 | 63533 | 6380 | 121637 |
| **Gaithersburg / Montgomery Village** | 6960 | 32346 | 22864 | 173 | 38694 | 6892 | 107929 |
| **Germantown** | 5774 | 29361 | 12405 | 295 | 43954 | 5325 | 97114 |
| **Headquarters and Special Operations** | 1924 | 8743 | 7736 | 132 | 12166 | 1376 | 32077 |
| **Rockville** | 9245 | 18575 | 14838 | 169 | 48728 | 4334 | 95889 |
| **Silver Spring** | 7142 | 86114 | 39637 | 498 | 34730 | 7864 | 175985 |
| **Wheaton** | 10910 | 58749 | 61396 | 647 | 62816 | 11092 | 205610 |
| **Total** | 49213 | 261354 | 175441 | 2349 | 304621 | 43263 |  |

**Figure: Number of violations by each race in different areas**

From the dataset, we can figure out that the White race is more involved in the traffic violations. But we feel it will be wrong to generalize anything based on this dataset for this analysis as we don’t have the data of population density of each race in these areas.

**Code:**

rxViolations **<-** read.csv**(**file.choose**())**

head**(**rxViolations**)**

rxViolations**$**SubAgency **<-** as.character**(**rxViolations**$**SubAgency**)**

rxViolations**$**SubAgency **<-** substr**(**rxViolations**$**SubAgency, regexpr**(**",",rxViolations**$**SubAgency**)+**2, nchar**(**rxViolations**$**SubAgency**))**

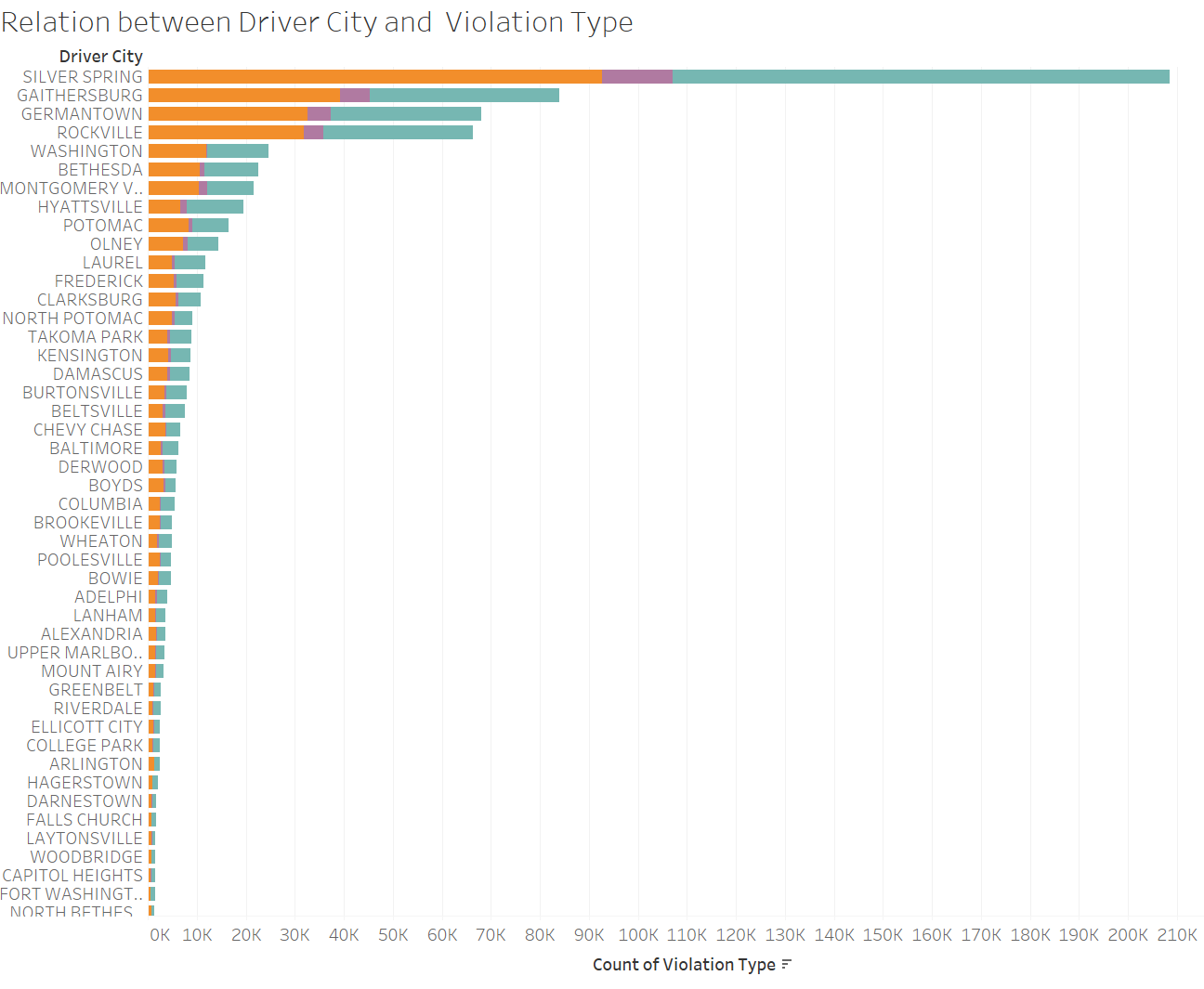
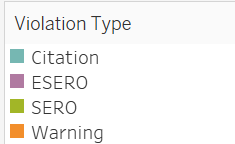
head**(**rxViolations**$**SubAgency**)**

violations\_area\_race **<-** as.data.frame**(**table**(**rxViolations**$**SubAgency, rxViolations**$**Race**))**

View**(**violations\_area\_race**)**

write.csv**(**violations\_area\_race, file **=** "violations\_area\_race.csv"**)**

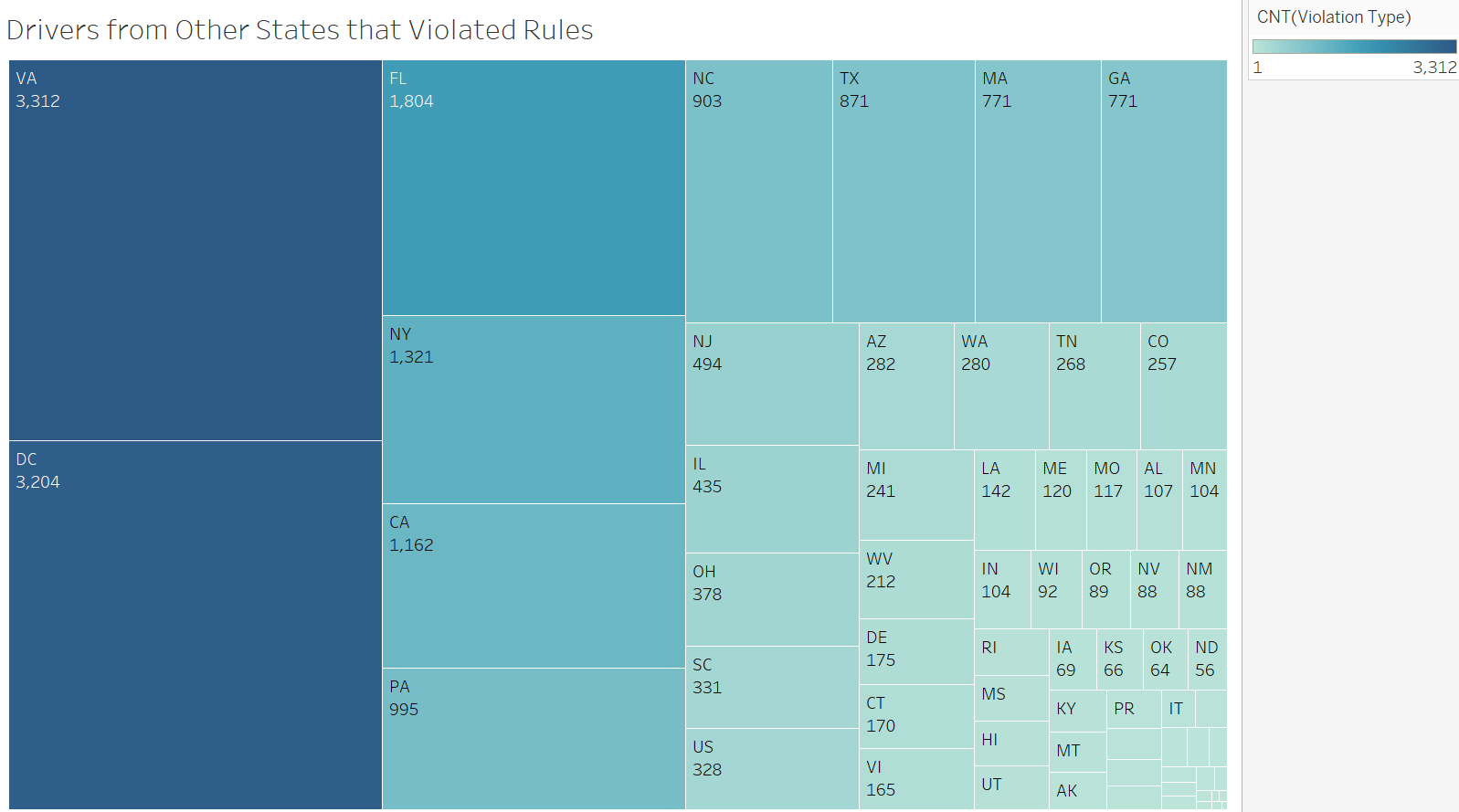
* **Correlation between the driver’s city and the type of violation that the driver was involved.**



**Figure: Driver City and Violation Type**

We can see from the above visualization that drivers from Silver Spring and Gaithersburg had the maximum number of violations. In Silver Springs citations are more as compared to warnings with a very little number of ESEROs also Gaithersburg has a similar trend. If a vehicle is being stopped it has fair, chances of getting a ticket and warning. Germantown and Rockville also have the same trend when it comes to citations. These are the major cities where violations occur.

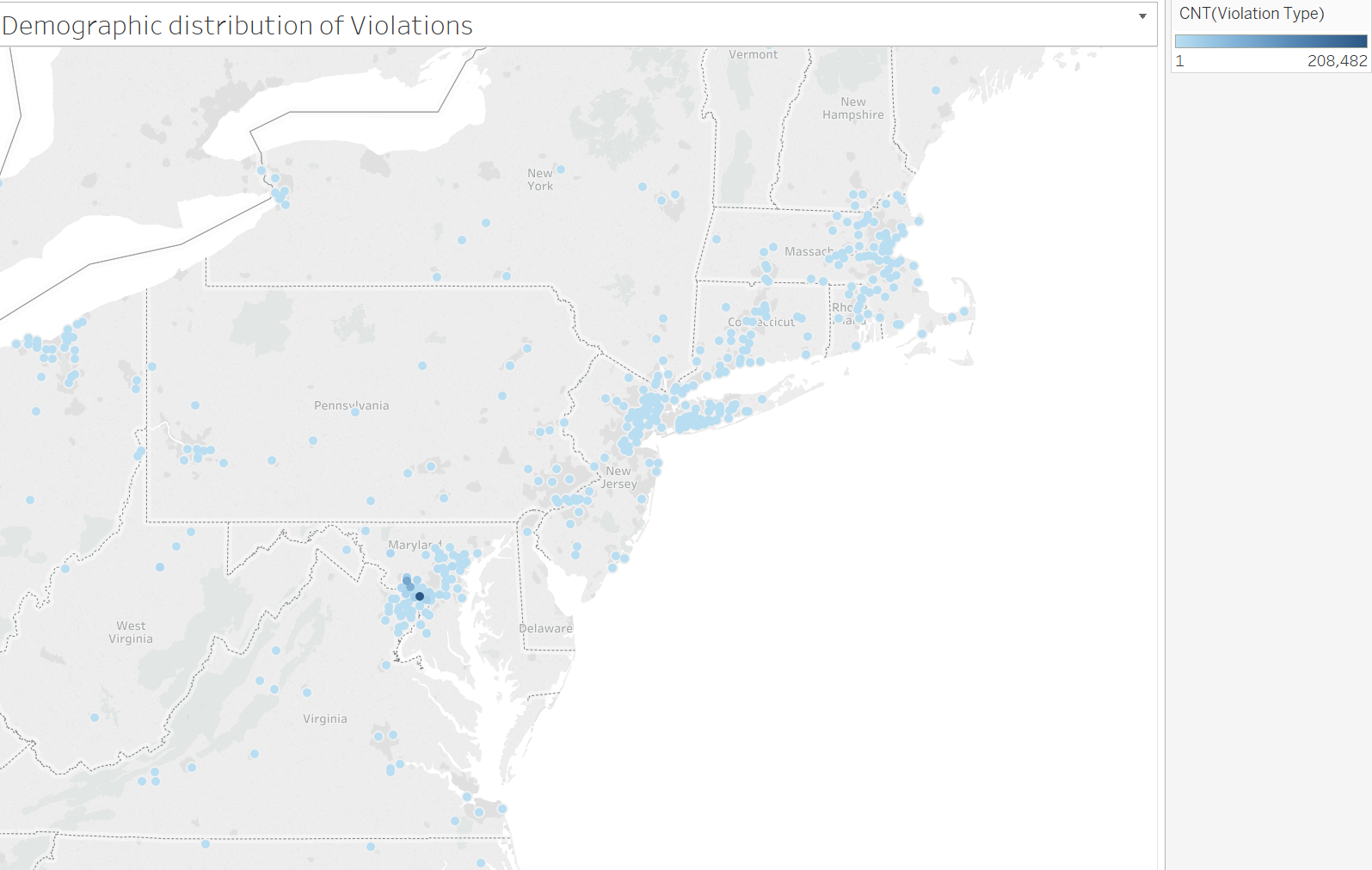
* **Number of drivers from outside state that violated traffic rules.**

**Figure: Driver City and Number of violation done by them in Maryland** 

The following graph shows that drivers from states other than Maryland that violated traffic rules in traffic rules. Drivers from Virginia had the maximum number of violations in Maryland followed by Washington DC. The darker shades of blue represent the higher number in the above representation.

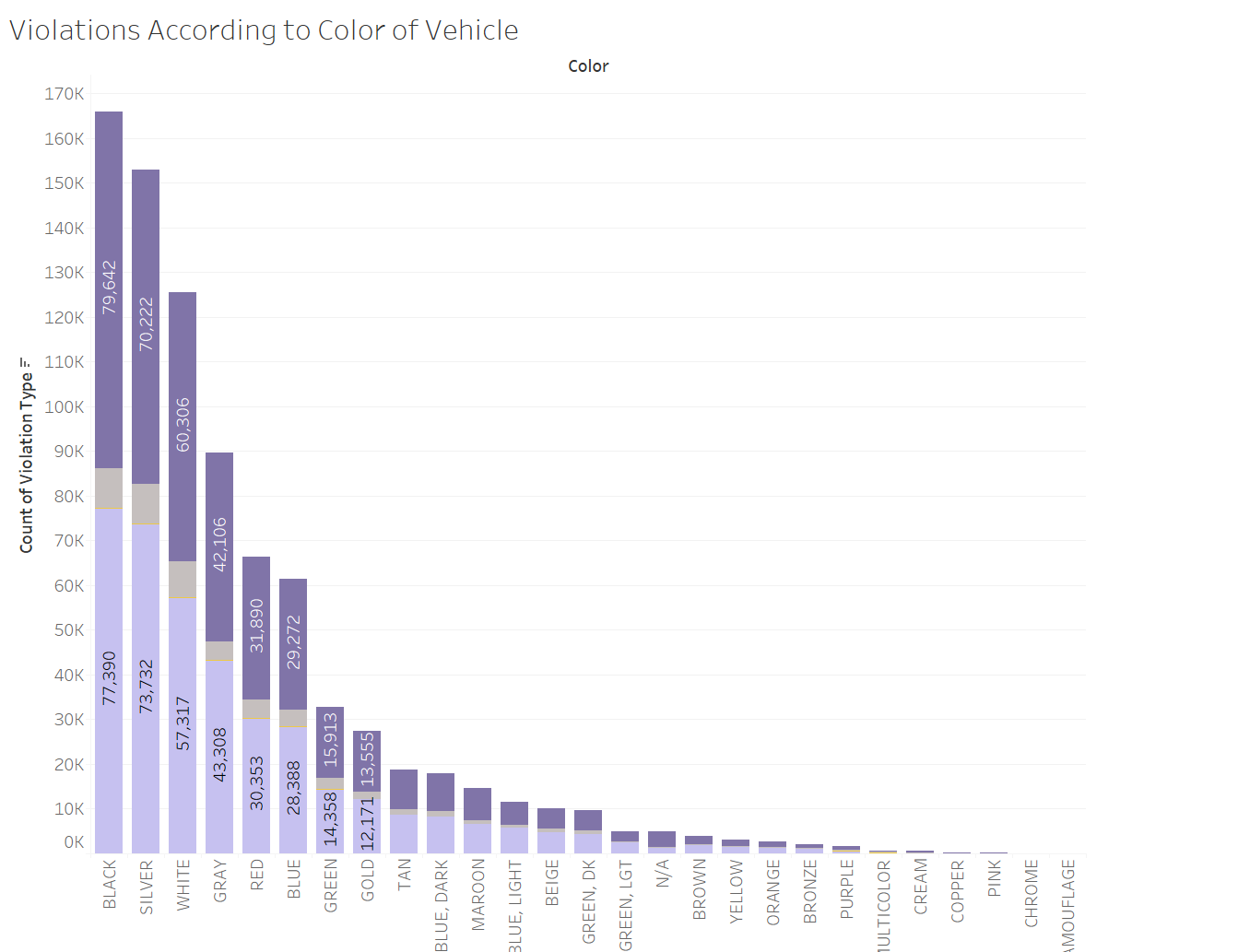
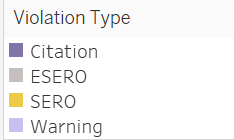
There is another important factor to consider since the traffic rules differ in each state hence drivers can violate traffic rules in Maryland if they are not familiar.

* **Demographic distribution of violations.**



The visualization above shows that drivers from which cities violated traffic rules in Maryland. We can see that the drivers who violated traffic rules were mainly from Maryland in addition to drivers from different states as Connecticut, New Jersy, Massachusetts etc. The color indicates the intensity of violations which is evident in Maryland as rest of the areas are lighter which indicates less number of violations.

* **The relation between the vehicle color and the number of violation of each type.**

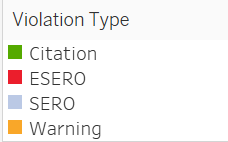
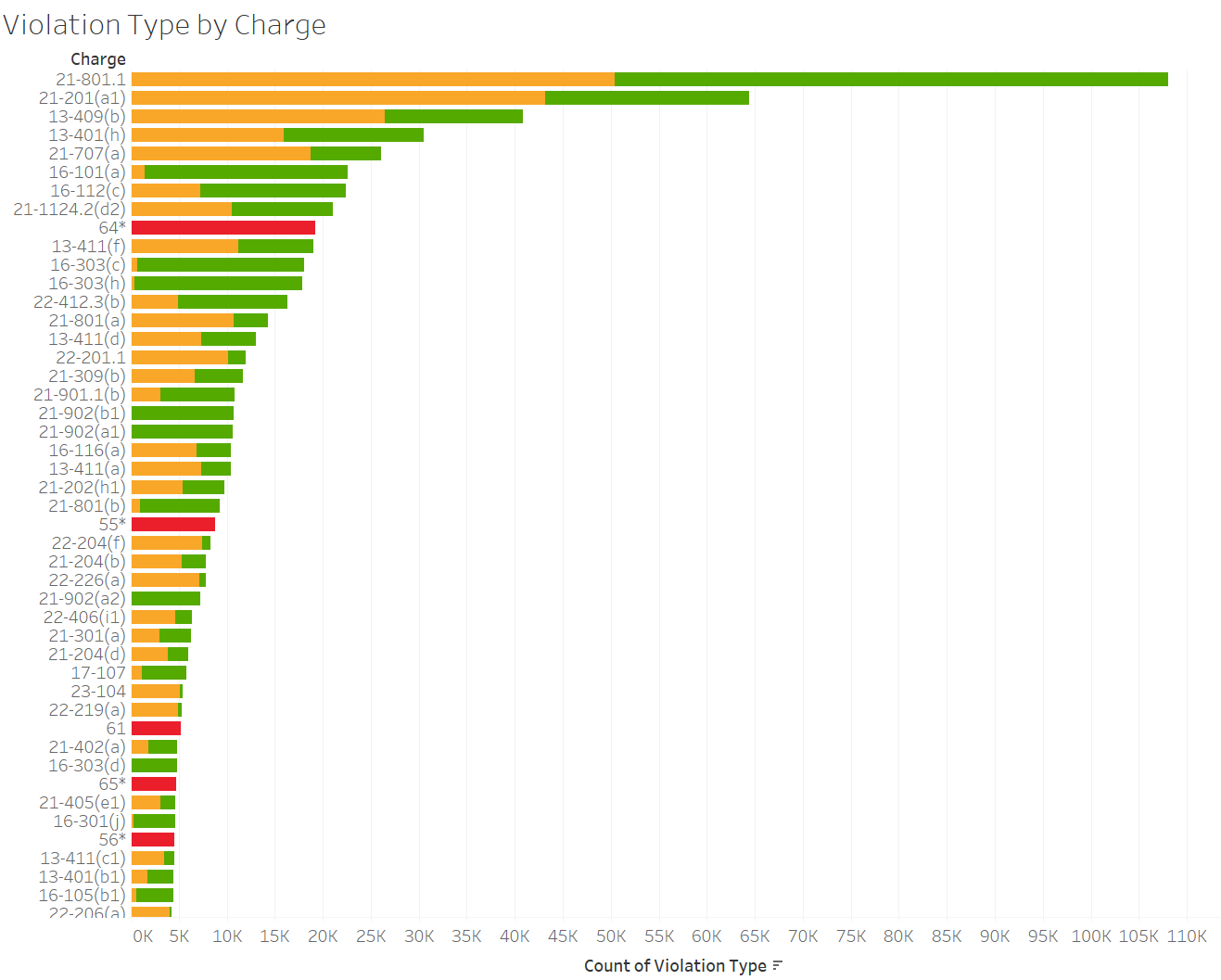


**Figure: Color of cars involved in each violation type**

It seems that drivers with bright colored cars violate less number of traffic rules but it can also be that there are a lesser number of bright colored cars overall in Maryland.

Black colored cars were stopped more times as compared to any other color followed by silver, white and gray. But we need more data regarding the total count of cars that are present in Maryland of each color to better reach a conclusion in this scenario.

* **Analysis of various charge types in each violation.**

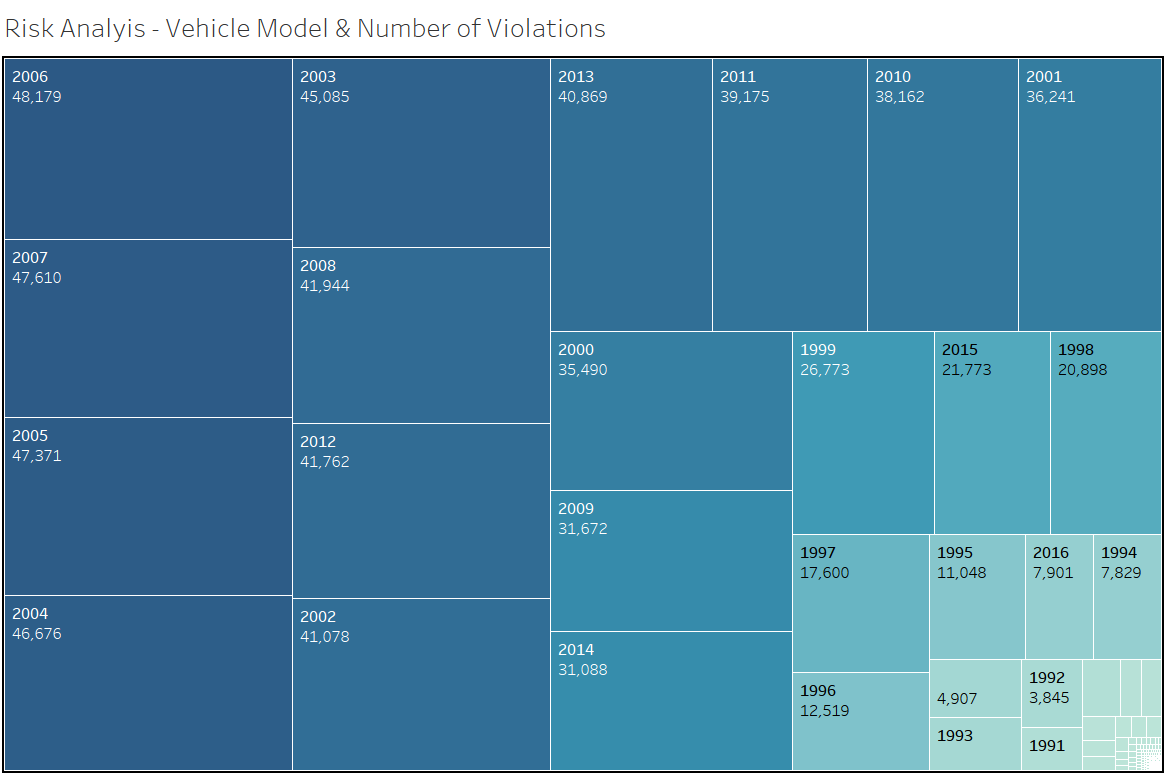


**Figure: Distribution of charge in each violation**

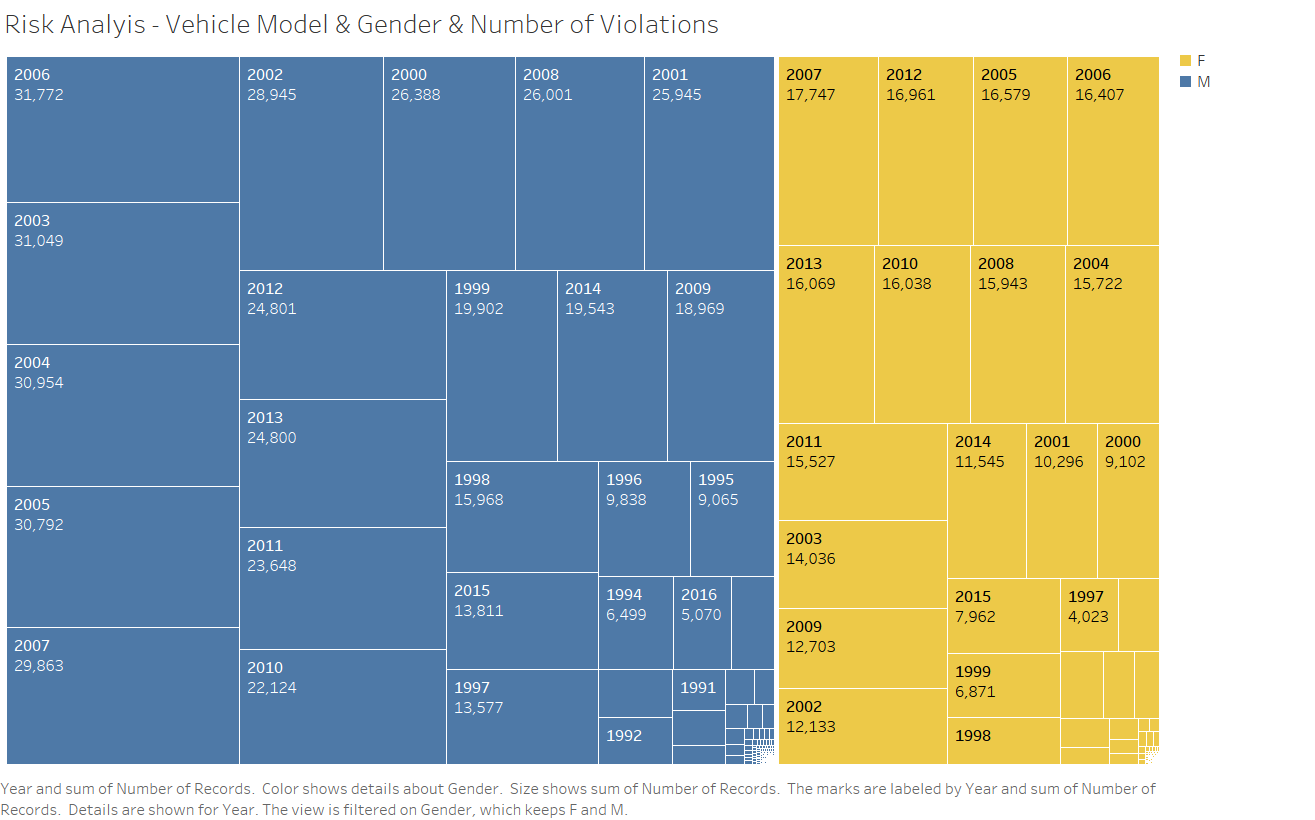
The violation charge that was prevalent in the maximum number was 21-801.1 which corresponds to speeding in the state of Maryland. Hence we can see that maximum number of warnings and citations were issued for speeding followed by 21-201(a1) that is a failure to obey instructions of a traffic control device but in this type of charge number of people were given a warning instead of a citation.

Another peculiar violation charge was 13-409(b) which implies that a large number of people were not able to produce the registration of the vehicle when stopped for a violation.

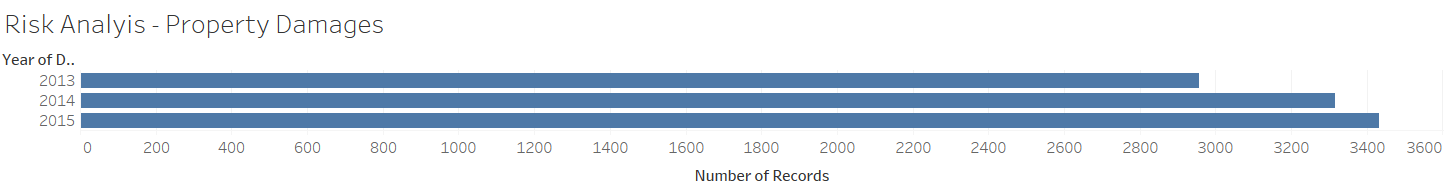
* **Are you more likely to be involved in a violation if you are driving an older vehicle?**



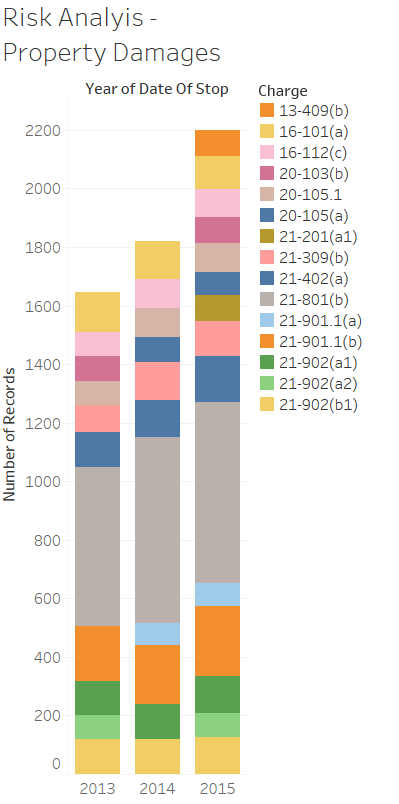
* From above tree map, we can infer that there are more violations from older vehicle models around 11-12 years i.e. 2006, 2007, 2005 & 2004.
* Much older models from 90s have fewer violations this is since the comparatively on-road count of these vehicles are less.
* Another reason for this result can also be the fact that, according to a recent study, the average age of an on-road vehicle in the U.S. is 11.4 years.
* Thus, the violation from vehicles of 2004,2015 is pretty high, but they might get replaced in time.
* As for 1990s vehicles, there are always a set of people who are an innate fan of Vintage cars.
* **Digging further down, we segregated our data on the basis on Gender.**



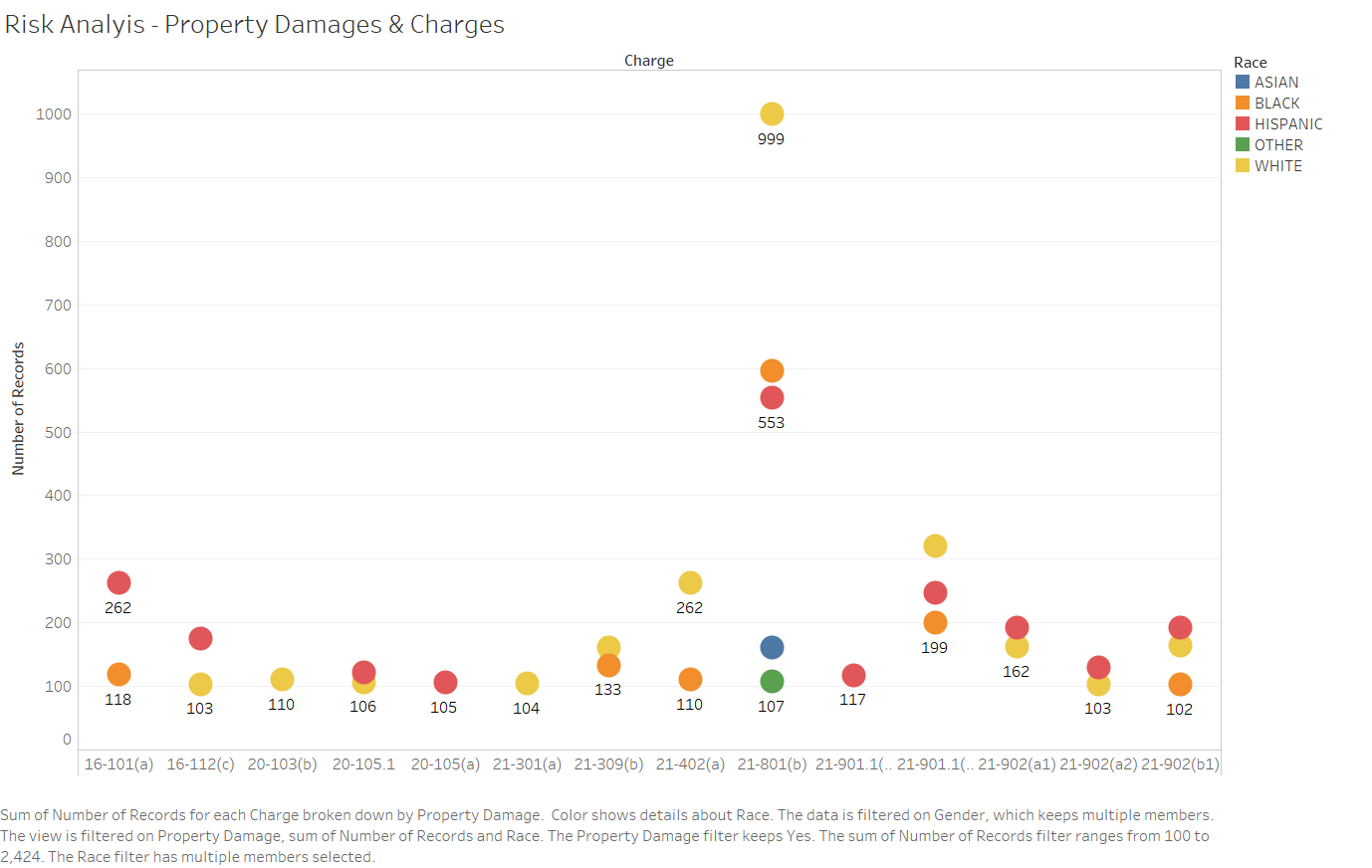
* Again, the pattern from this analysis was in sync to the one carried out before it.
* We can see most violations by males are from vehicle models 2006, 2003 & 2004 whereas by females, 2007 models followed by 2012 vehicle models.
* **For what kind of violations was there property damage (& its relation with Race and Gender)?**



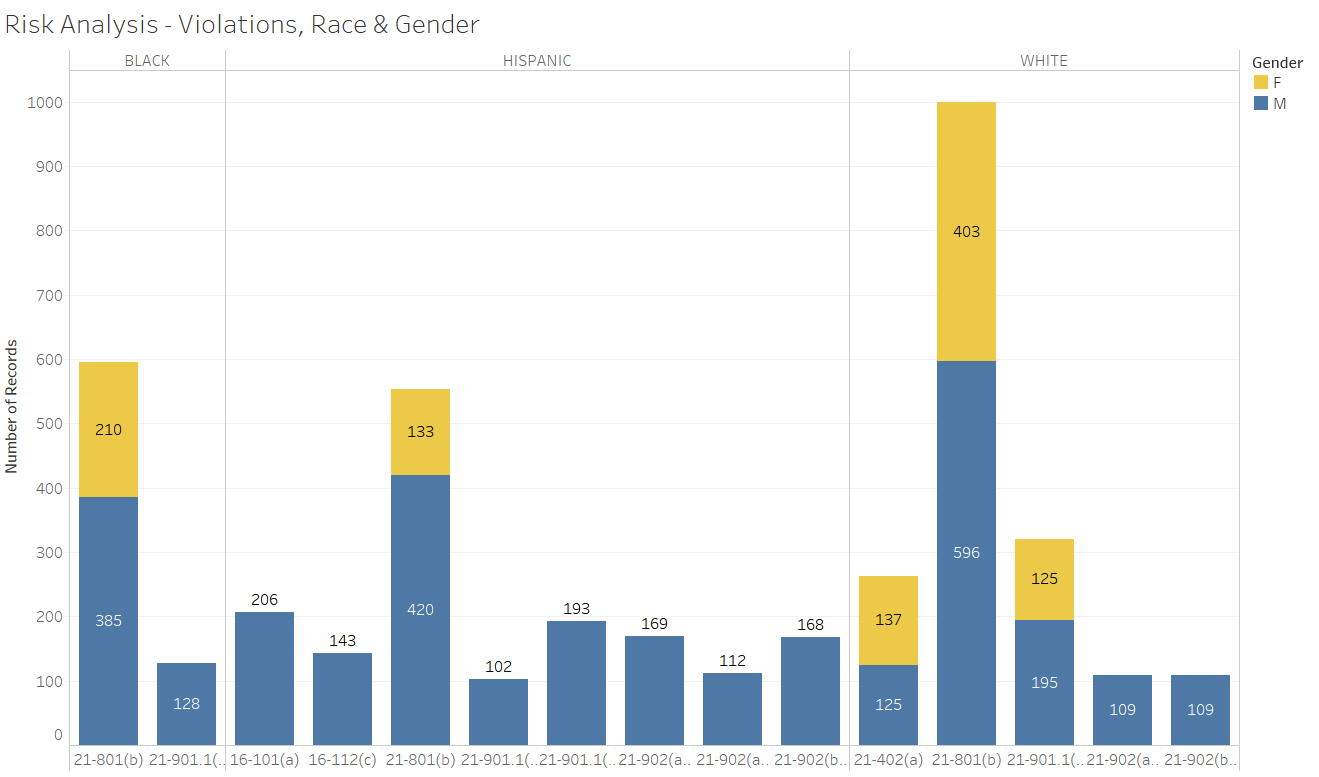
* During our analysis, we saw a gradual increase in property damages due to the traffic violations.
* This also emphasized the need of such an analysis were relatively unfamiliar conclusions could be unlocked so that necessary steps can be taken to curb them.
* **Risk Analysis based on Property damages due to Speeding Violation**



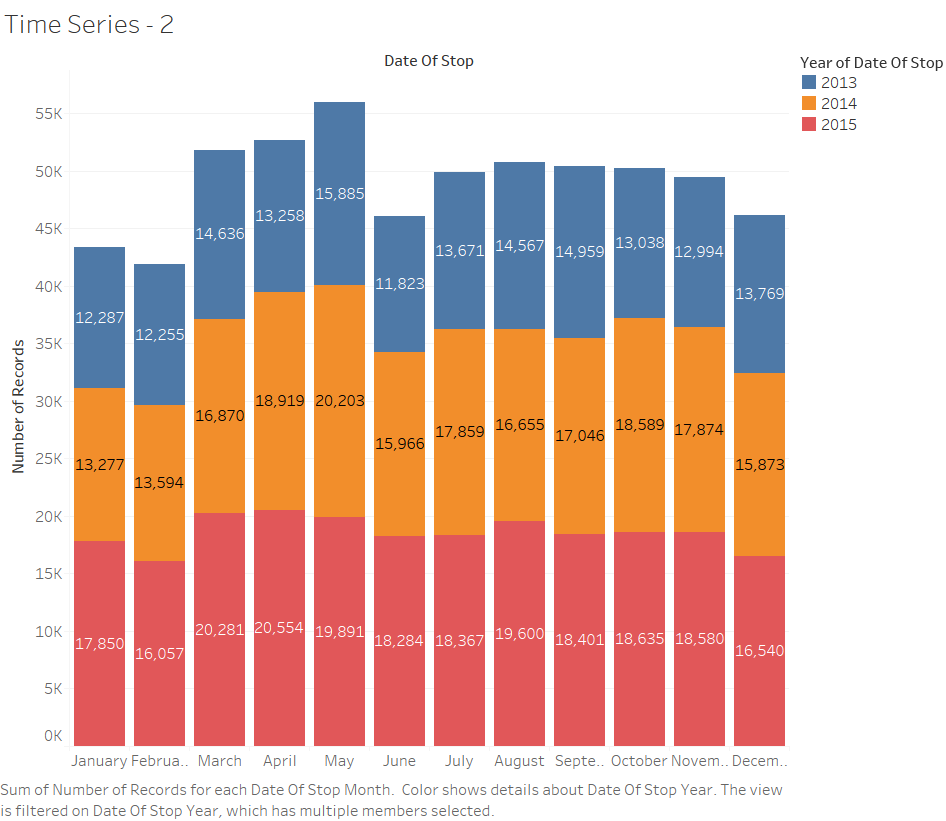
* A simple Risk analysis shows that most of the property damages occur due to speeding violations (21-801).
* **Risk Analysis for property damage by disparate races.**



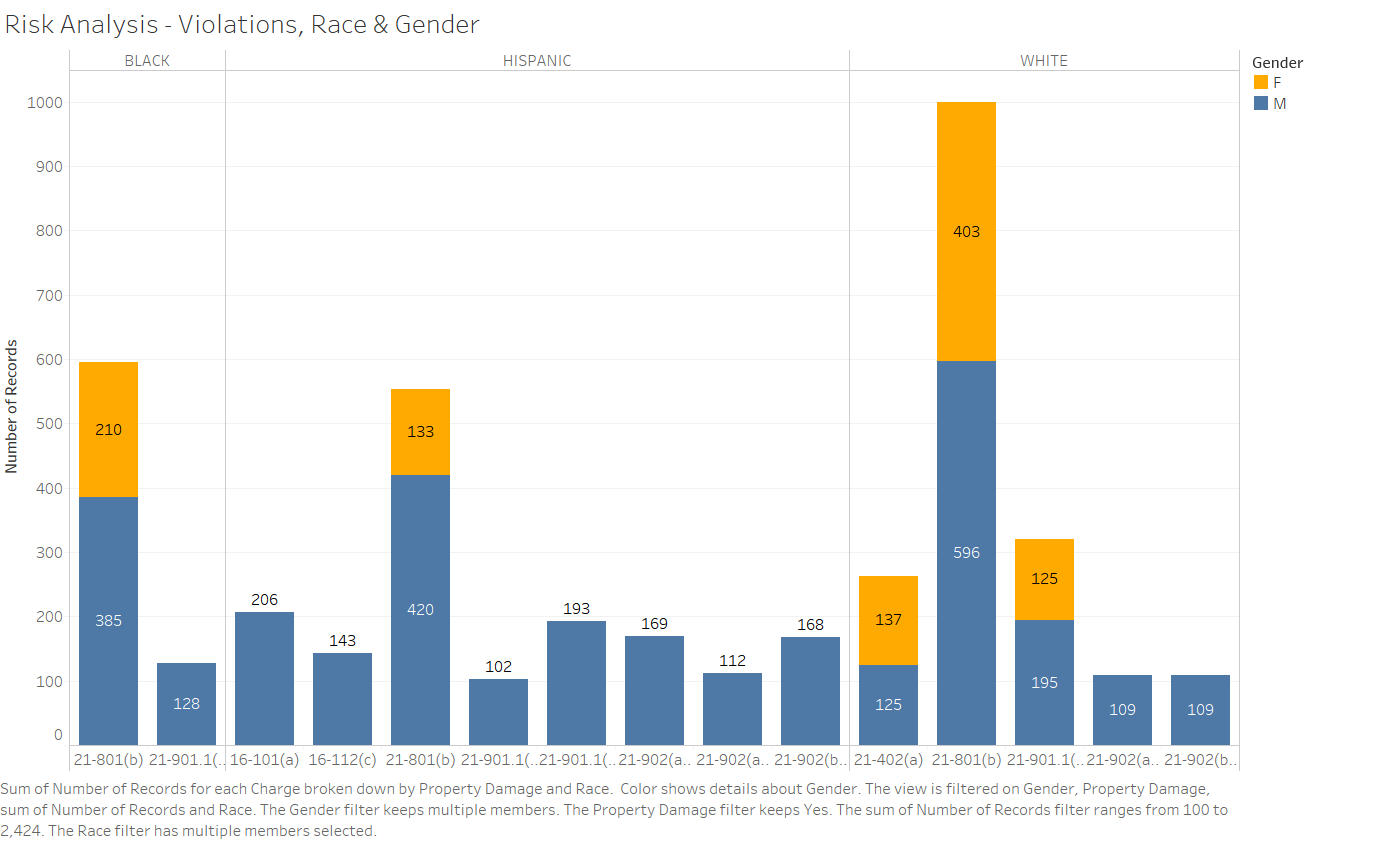
* In above graph, we can see most of the violations leading to property damages are from White folks followed by Black and Hispanic.
* The reason for the same could also be the fact that the demographic population for whites is much more than that of blacks and Hispanics.
* **Risk Analysis showing violations differentiated by Race and Gender.**



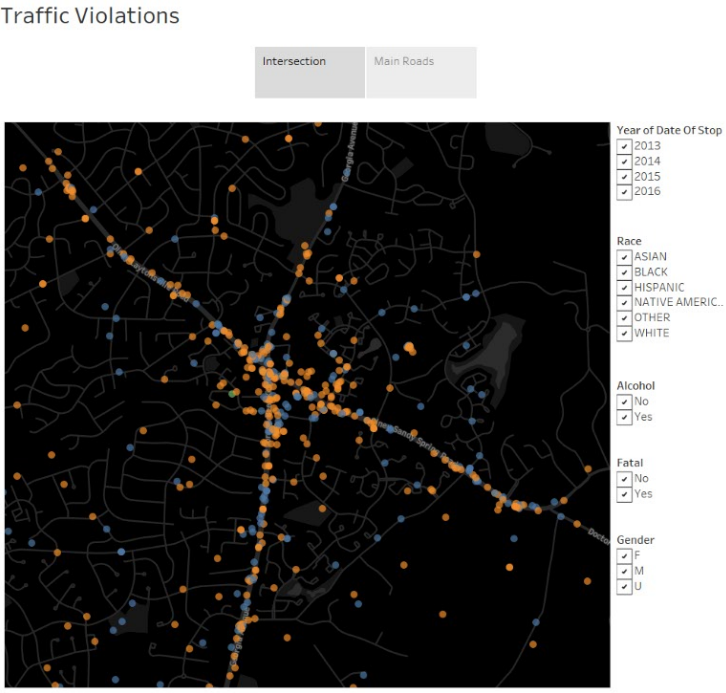
* After Filtering data on Gender, we observe that most of the violations are from males except for charge 21-402.
* Charge 21-402 represents a failure to yield right-of-way when making a left or U-turn.
* Females tend to make this violation frequently, inevitably leading to property damages.
* **Given a specific month in a year, when is Maryland the most dangerous state for tourist to visit.**

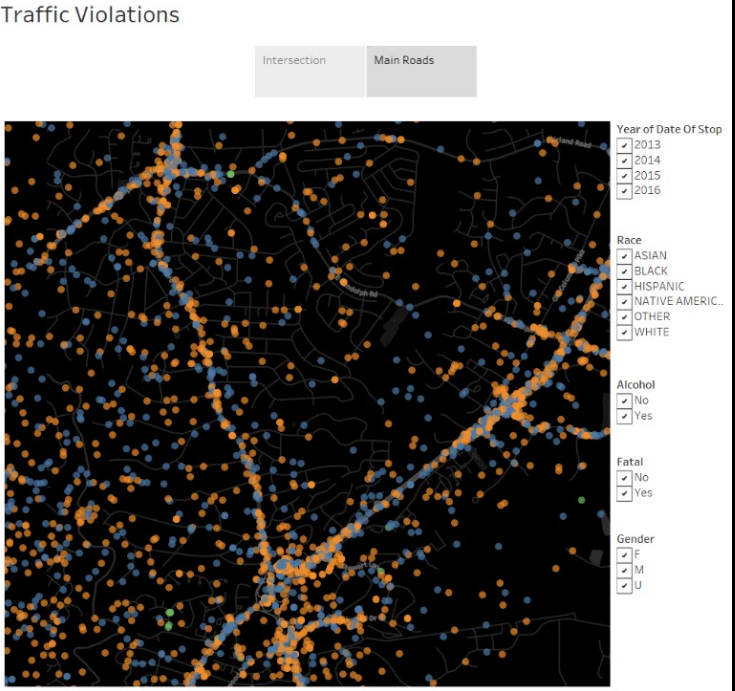


* From above bar graph, we can interpret that most violations occur in the month of May followed by April.
* In May, there is a slight decrease in the violations whereas there is a steady increase in a number of violations in April.
* Although, April accounts for highest number of violations in Year 2015.
* Also, the number of violations have gone on to decrease gradually from 2013 to 2015.
* **Risk Analysis for the number of violations based on Race and Gender.**



* According to recent statistics, around 41 million speeding tickets are issued each year.
* This means every day, 112,328 people are slapped with an average fine of around $150.
* In Montgomery county of Maryland, most violations are of speeding represented by violation code 21-801.
* This is common against all races where violations are higher for Whites which is mainly due to the larger white population.
* **Traffic Violations around Intersections and Main Roads.**





* Looking at the two traffic violations we can observe that the amount traffic violations is very high around Main roads and Intersections which are the New Hampshire Road and Columbia Pike Road.
* Data from different sources can be collected and an efficient path with minimum traffic violation can be concluded.

# CONCLUSIONS

Working with the Traffic Violation data, we have come up with various safety measures that can be implemented and followed by the people and police department in Montgomery(Maryland). The idea of this project is to aid the driver in a way that he can be more aware of his surroundings. The findings of the project will help people be extra careful before entering the accident-prone zones and drive more carefully. For Example, a resident in Montgomery who has traffic violation knowledge of an area (due to the findings of this project) would be aware of an area that is more prone to traffic violations. The resident can be extra careful while driving the concerned area just to make sure that he doesn't receive a ticket (Since policeman would tend to be around areas where violations occur more often). Also, With the help of GPS module, this concept can be regulated into the vehicles as a geo-location representation of black spots. Policymakers in Montgomery county can now change or implement new policies based on our findings. For example, if there is a specific area in Montgomery where drivers are infringing the law, then the policymakers may decide to increase the number of a policeman in that area. On the other hand, due to the findings presented in this project, the policymakers may decide to completely reform the law in Montgomery(Weather) due to a large amount of violations in that area.

# FUTURE WORK

In this project traffic violations, data of a state of Maryland was carried out to curb the number of traffic accidents and observe the data by performing specific data analysis in R and Tableau. As part of the future scope:

* The Same analysis can be carried out for different county/state or an entire country.
* Using weather data to examine its effects on Traffic violations.
* Installation of a blackspot program on the GPS system of Vehicles.
* Building a real-time application that would display most recent traffic violations in the vicinity.

# REFERENCES

1. <http://blog.nycdatascience.com/student-works/visualizing-traffic-violations-local-insight/>
2. https://www.dataquest.io/blog/python-json-tutorial/
3. <http://vizual-statistix.tumblr.com/post/101677467941/i-came-across-a-database-of-traffic-violations>
4. <http://blog.revolutionanalytics.com/2016/04/an-analysis-of-traffic-violation-data-with-sql-server-and-r.html>
5. <https://scholarworks.bridgeport.edu/xmlui/bitstream/handle/123456789/1573/179-term_project___poster.pdf?sequence=1>