

Intro & Data Cleaning

This data comes from the Burlington Police Data Portal found on the City of Burlington website. The data set includes police incident report information on 22 different call types since 2012 up until September 28th 2018. The data is based on reports made to the police by Burlington residents and includes date and hour of call, as well as location. This would be considered an observational study. There may be some sampling bias to this data because it only includes crimes and violations that were reported to police. In reality, some crimes likely go unreported. There may additionally be some bias in the data due to false incident reporting. By analyzing this data, we hope to investigate how the five most common call types vary by location and over time. We use this data to look at how incident reporting changes by time of day as well as across an entire year. We find this topic particularly interesting because crime is a large part of how the safety of cities is defined. It is important to understand where and when crimes are reported in order to put in place initiatives to decrease the number of incidents reported and increase safety. Additionally, studying police incidents can potentially be utilized for predictive modeling algorithms in order to prevent violations before they occur.

Little data cleaning needed to be done. Some variables, like Call Origin and Area, had a handful of NA values, so we used `!is.na()` functions to remove those values. Because of the large size of the data set and the small number of missing values, we felt that this was the best way to clean the data.

Code for plot of all call types in order (suggested by prof. Weaver)
Plot 0

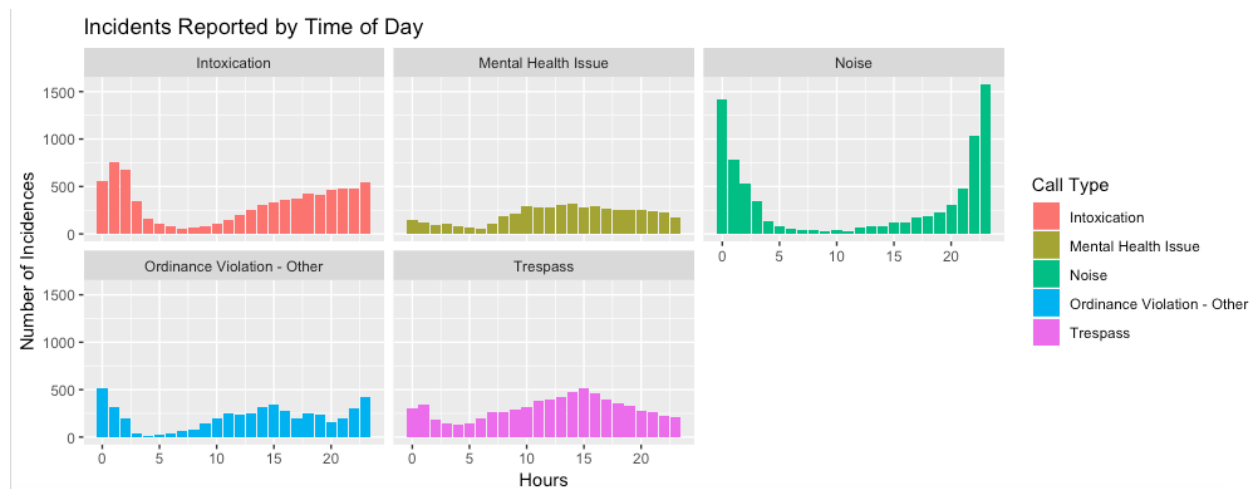
```
callorder <- Police_Incidence %>% group_by(call_type) %>% summarise(freq = n())  
ggplot(callorder, aes(x = reorder(call_type, -freq), y = freq)) + geom_bar(stat = "identity") + coord_flip()  
+ labs(x = "Call Type", title = "Frequency of Call Types")
```

This initial graph presents all of the call types in ascending order of frequency in the entire data set. There are 22 call types total. Based on the large number of call types, we decided to focus the rest of our graphs on just the top five call types. This included: noise, intoxication, trespassing, ordinance violation and mental health, from most cases to least cases. Other frequent incidents included domestic disturbance and vandalism. There were the fewest cases of larceny from a person (ie. pickpocketing) and robbery. Overall, the ordering of incidents was not particularly surprising.

Plot 1

Code for plot of incidents by time of day:

```
ggplot(data = P3, mapping = aes(x = Hour2, fill = call_type)) + geom_bar() + facet_wrap(~call_type,  
nrow = 5) +  
  labs(x = 'Hours', y = 'Number of Incidences', fill = 'Call Type', title = 'Incidents Reported by Time of  
Day')
```

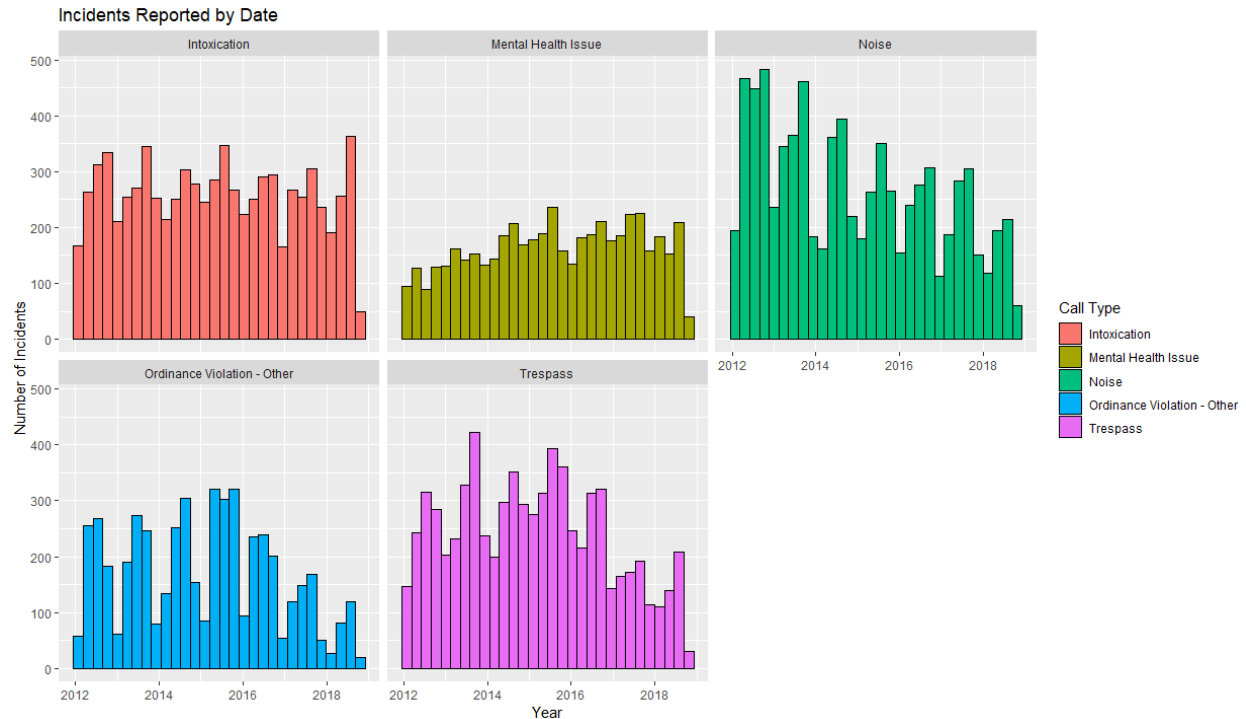


This graph shows the change in the number of incidents reported to the police for the top five call types. The graph shows that for intoxication, noise and ordinance violations, the most likely time of day that reports will be filed to police is between 11pm and 1am. This trend seems logical considering that people are more likely to drink excessively and party at the end of the day when they are done with work or school. However, incidents regarding mental health and trespassing are more likely to be reported during the day. Mental health related reports might be filed more often during the day because people might be more likely to be alone. This would be assuming that those who suffer from severe mental disorders are less likely to work and may be left alone when friends and family are at work during the day. Additionally, crimes that occur in the evening relating to mental illness may be reported as a different call type depending on the violation. It is interesting that the data shows that trespassing is more likely to occur during the day, and may be due to people more likely to be caught trespassing during the day. However, I suspect that trespassing occurs in a similar rate at night, if not more, but reports may not be filed.

Plot 2

Code for plot of Incidents Reported by Date:

```
ggplot(data = P3,
  mapping = aes(x = as.Date(date), fill = call_type)) + facet_grid(rows = vars(call_type)) +
  geom_histogram(color = "Black") + facet_wrap(~ call_type) + labs(x = 'Year', y = 'Number of Incidents',
  fill = 'Call Type', title = 'Incidents Reported by Date')
```



This graph shows the rate of reported incidents from the year 2012 through 2018. It also shows varying amounts within individual years.

The first trend we noticed was the decrease in Noise incidents from 2012 to 2017. This is the most dramatic trend we found in the graph. We suspect this is unrelated to the amount of noise being made in Burlington. Because of how drastic the change is, we suspect it is due to the way police handle these incidents. Perhaps the police have made noise complaints less of a priority in their department, and, as a consequence, the amount of calls and incident reports have declined in turn. We think this is the case, because we do not think that citizens' natural behavior could change so much over such a short period of time.

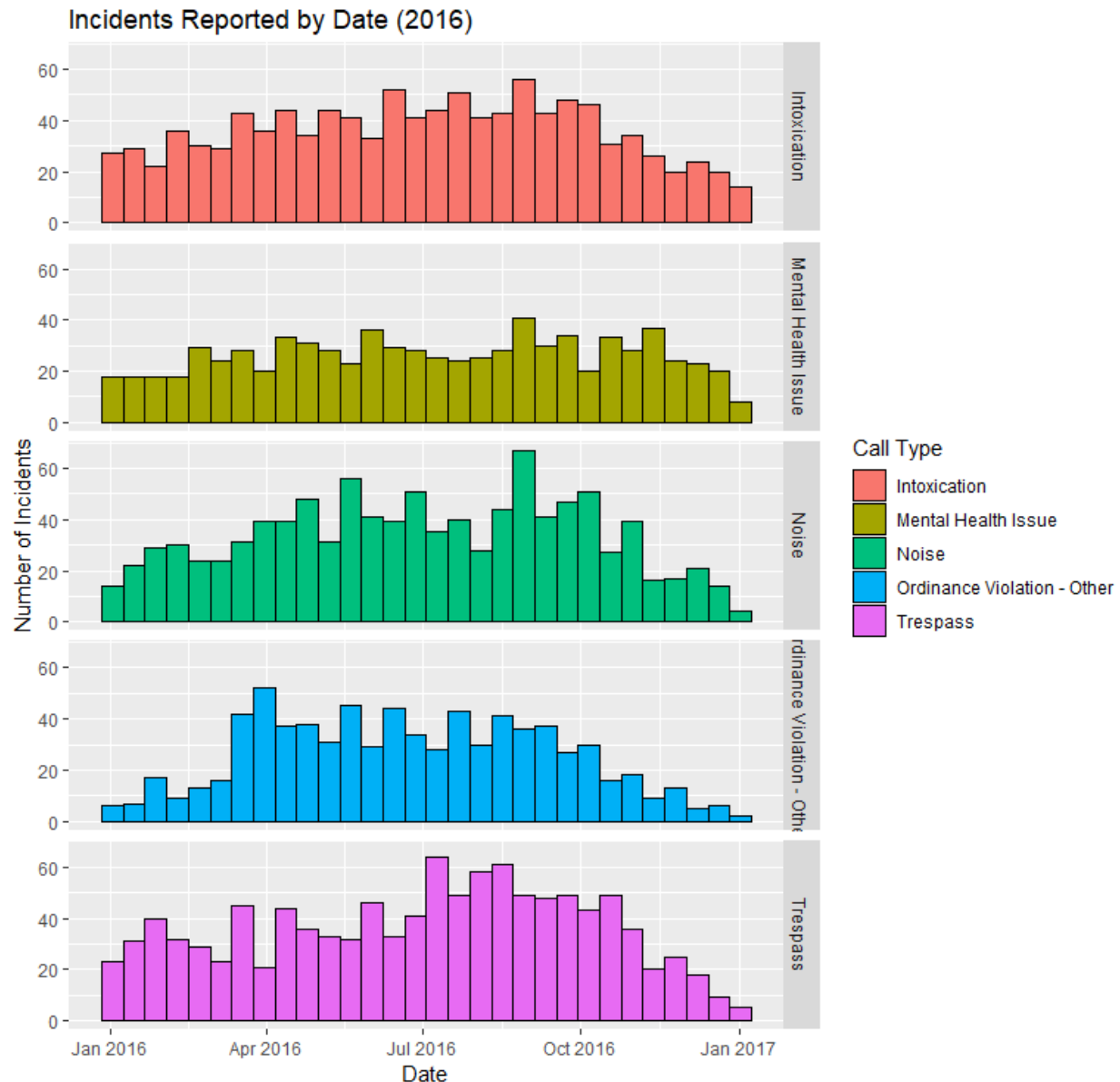
For 2 of the other incident types - Ordinance-Violation Other and Trespass - a slightly different trend can be seen. The rates of these incidents stay fairly uniform from 2012 to 2015, but then decrease over the next 2 years. Intoxication maintained a steady rate across the 7 years.

Mental Health Issue incidents exhibit a different trend. The rate of these incidents increased markedly from 2012 to 2018. In 2012, there were 500 Mental Health Issue incidents. By 2017, there were 853 incidents. The rate of these incidents increased every year, besides 2016. We suspect this could be caused by the city's growing concern about mental health related incidents, especially related to the opioid crisis, which the mainstream media finally began to acknowledge and reckon with in this time period. However, because we only have 7 years worth of data, it is difficult to attribute these trends to any changes in population or public policy.

Finally, one interesting trend we noticed in this data are the 7 peaks that are noticeable in every Call Type besides mental health issue. We were interested in taking a closer look at what these peaks might be caused by.

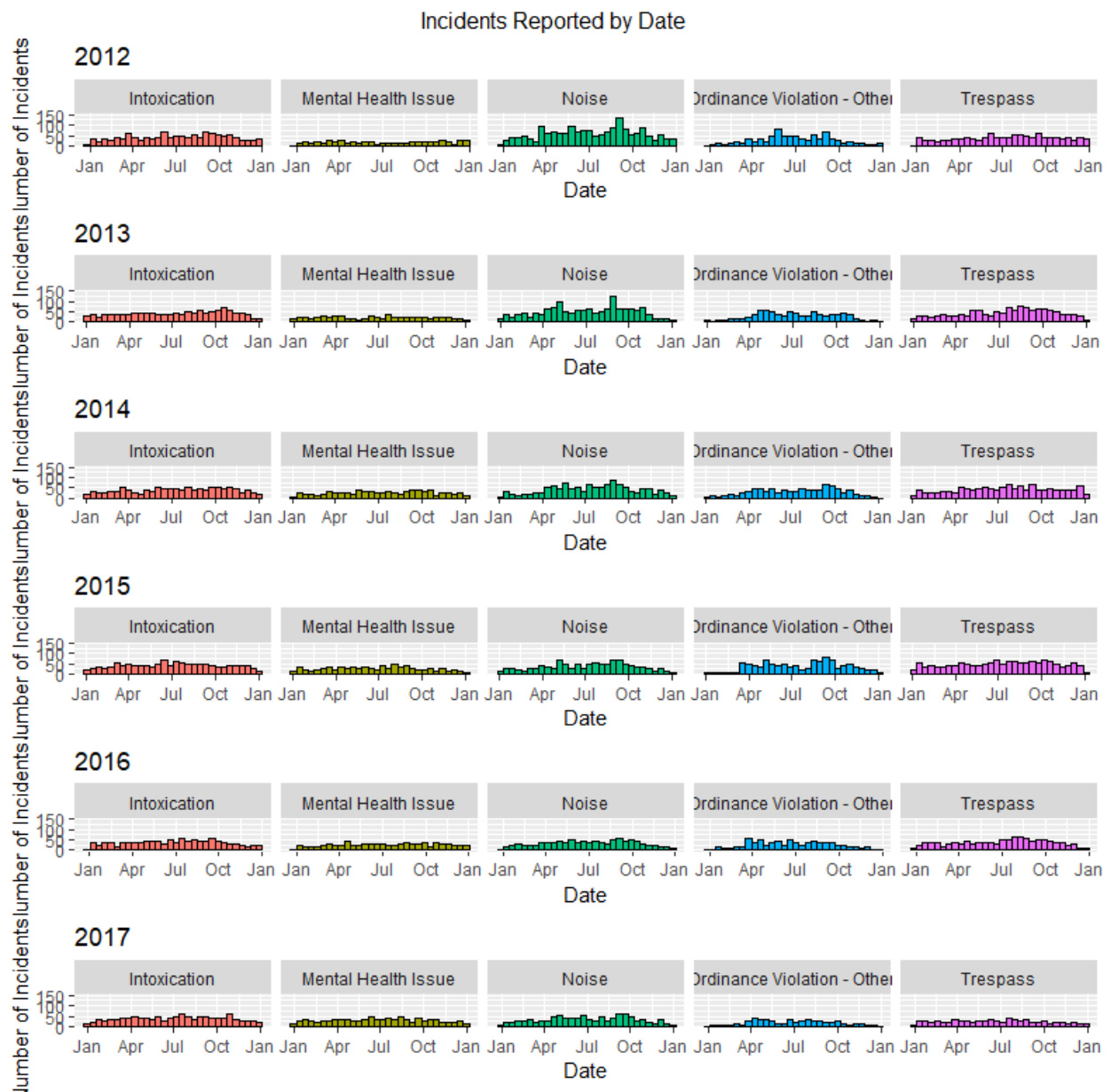
Plot 3

Code for plot of Incidents Reported by Date (2016):



```
ggplot(data = p2016Top5,
mapping = aes(x = as.Date(date), fill = call_type)) + facet_grid(rows = vars(call_type)) +
```

geom_histogram(color = "Black") + labs(x = 'Date', y = 'Number of Incidents', fill = 'Call Type', title = 'Incidents Reported by Date (2016)')

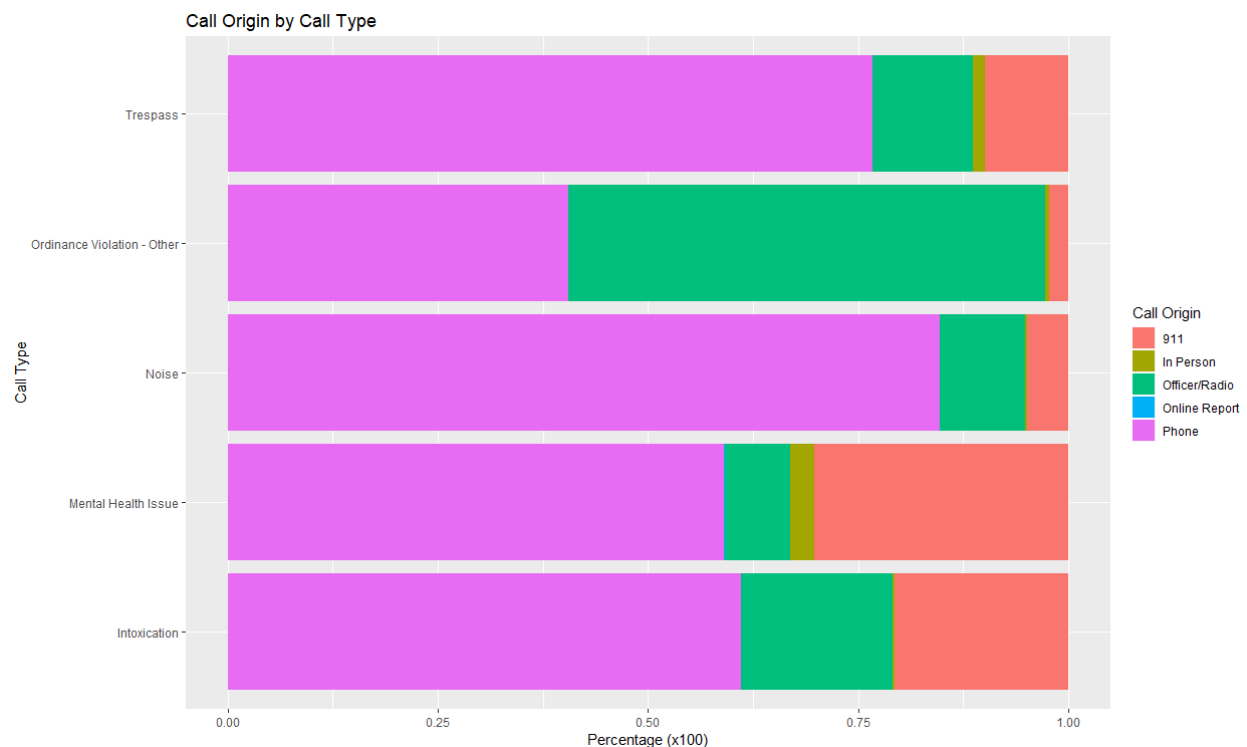


By looking at the trend of each incident type by each calendar year, we were hoping to better understand the intra-year differences in incident rates. We excluded 2018 data here because we do not have data for the whole year yet.

We see that intoxication, noise, ordinance violation, and trespassing calls all seemed to rise in the middle of the year. We researched other analyses of crimes/misdemeanors as they are related to weather. We found is that there is a consistent positive relationship between crimes and temperature/amount of daylight (1). We believe that this to makes sense. The hotter it is outside and the longer the day, greater amounts of people will be outside, and for longer; this increases social interactions, which increases the chances of incidents occurring. Additionally, the noise and ordinance violation plots show what look like smaller peaks within the year: one around May/June, and the other around September. We suspect that this is because Burlington is a college town. In cities without such a significant student population, these plots would probably more closely resemble normal curves, with the highest point in the middle, around July/August (the hottest months). However, because of Burlington's sizeable student population, which greatly decreases during the summer months, the typical seasonal trend is not as clear. Instead, we see this 2-peaked trend, which can be related to warm weather and longer days, but is also influenced by move-in season and graduation. Another factor that could cause the incident increase around September is Burlington's fall visitors. The area is popular for tourism during the fall months because of its foliage. This increase in visitors could possibly cause more behavior that lead to incidents.

However, once again, the mental health issue category did not behave like the other call types. In order to better understand this category, we looked at the Call Type data in relation to Call Origin.

Plot 4

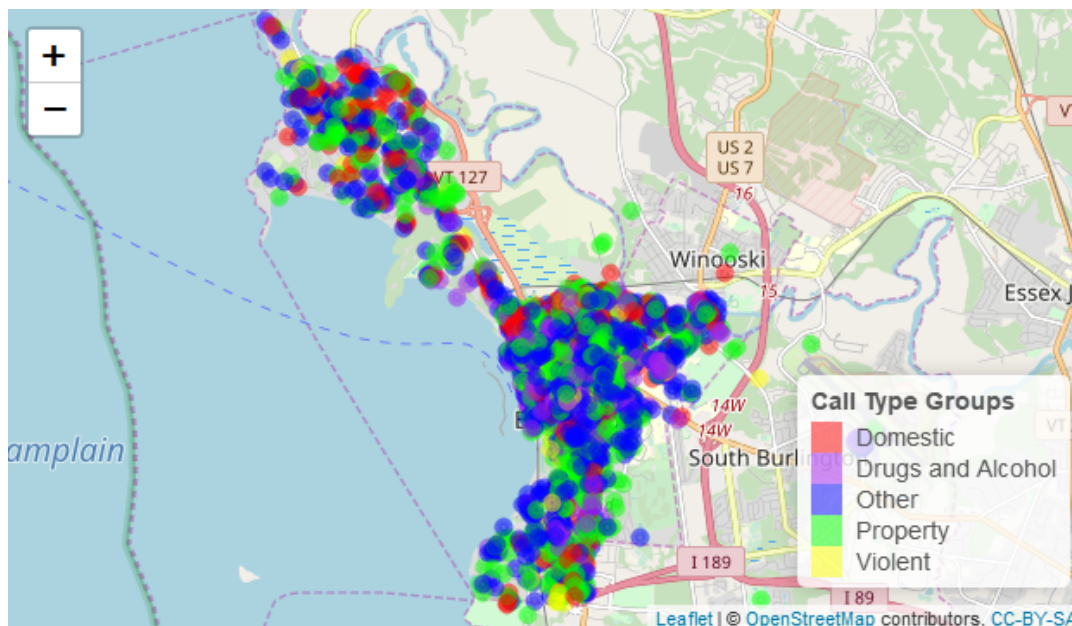


Code for plot of Call Type by Call Origin:

```
ggplot(data = pTop5, mapping = aes(x = call_type, fill = call_origin)) + geom_bar(position = 'fill') +  
coord_flip() + labs(x = "Call Type", y = 'Percentage (x100)', fill = 'Call Origin', title = 'Call Origin by Call  
Type')
```

This plot graphs the percentage of each Call Origin by each Call Type. We thought that this plot might give us more information on the reporting of mental health issue incidents. Two Mental Health related trends are easily noticeable. First, mental health issue incidents come from 911 calls at a much higher rate compared to the other incidents. Additionally, the in-person intervention rate was higher, as well. (Although this only appears to be about 5% of the mental health issue incident origins, it is about twice as much as any other call type.) The high rate of 911 call origins seems to indicate that the incidents being reported demanded urgent assistance. Additionally, the high rate of in-person origins seems to mean that these incidents are occurring in a public place, where a police officer intervenes on their own. These differences indicate that mental health issue incidents are unique compared to the other incidents. Based on the high rate of 911 call origins, they seem to often occur in private places, with a third-party individual feeling the situation demands immediate intervention. On the other hand, the high rate of In Person call origins seems to indicate that there is another kind of mental health issue incident that occurs in public, and is noticeable enough for police to get involved on their own. We believe this data indicates that “Mental Health Issue” is a very broad category. Perhaps there are two common, distinct types of mental health incidents: one occurring in private, and the other occurring in public. Regardless, we feel that the “Mental Health Issue” category is probably distinctly different from incidents like intoxication, noise, and trespassing, which is indicated by the different rates and methods by which it is reported.

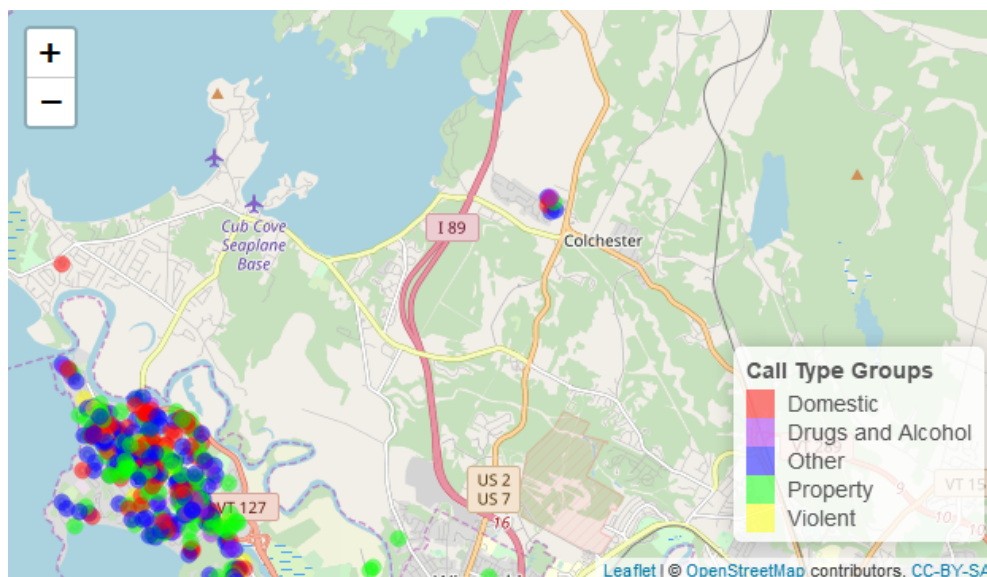
Plot 5



This map is a visualization of police incidents by call type groups. The groups include Domestic, Drugs and Alcohol, Other, Property and Violent. The 'Domestic' group includes call types such as domestic disturbance, domestic assault (misdemeanor) and domestic assault (felony). The 'Drugs and Alcohol' group includes overdose and intoxication calls. The 'Other' group includes calls such as noise, mental health issue, trespass, disorderly conduct and ordinance violation. The 'Property' group includes burglary, vandalism, retail theft, larceny (from building), stolen vehicle, larceny (other) and larceny (motor vehicle) calls. The fifth group is 'Violent' and includes calls such as assault (simple), robbery and assault (aggravated).

The map indicates areas of higher police concern within the Burlington area. Some areas stand out as particularly dense. Church street for example appears to have the highest density of police incidents on the map. The University of Vermont Medical Center has a large amount of incident calls, a large percentage of which falls under the group 'Drugs and Alcohol' and 'Violent'. When examining the map, it appears that there are certain areas with a higher frequency of one type of call. There is a large amount of 'Property' calls between the bottom of Cherry street and Bank street, which lines up with the Square Mall. Domestic calls unfortunately occur throughout Burlington, but it's interesting to see where Red is darkest, perhaps indicating repetitive calls concerning domestic violence. High density of 'Property' calls near most shopping center. Off the Shelburne exit, there is a high density of incidents, all green, indicating a property incident. The points land on a mall strip that includes a T.J. Maxx, Home Good, a market and a Sierra Trading Post. There appears to be more 'Domestic' calls towards the Old North End region. Perhaps this can be correlated to areas of lower income. Not surprisingly, the majority of the incidents reported at the Burlington International Airport are drug and alcohol related.

Plot 6



Corinne Atkinson
Joshua Blouin
Joseph Ortigara
STAT 087 Group Project Draft - Burlington Incidents

This map shows a small cluster separated from the majority of the other points. This small cluster is located at the Lone Pine Campground. There is a variety within the type of those calls, but the majority are 'Domestic' and 'Other' type groups.

Code for mapping call type by groups in Burlington :

```
b2 <- b %>% filter(Longitude < -73.1)

b3 <- b2 %>%
  mutate(Call = mdy_hm(call_time),
         Date = date(Call),
         Month = month(Call, label=TRUE),
         WeekDay = wday(Call, label=TRUE),
         Day = day(Call))

m <- leaflet() %>%
  addProviderTiles(providers$Stamen.Toner) %>%
  setView(lat = 44.475, lng = -73.212, zoom = 11.45)

color_pal <- colorFactor(palette = c('red','orange','blue','green','yellow'), domain = b3$call_type_group,
reverse = F)

m %>% addCircleMarkers(data = b3, lat = ~ Latitude, lng = ~ Longitude, color =
~color_pal(call_type_group),radius = 3)
```

Limitations

We found there to be 2 general limitations with the data: specificity and range.

Categories like "Mental Health Issue" and "Ordinance Violation - Other" are difficult to visualize because they seem to encompass many different types of occurrences. If we had more detailed classifications of Call Type, we could possibly find more subtle trends in the data.

Second, our analysis was limited by the number of years of data included in the data. Including more years of incidents in the data set would allow us to find trends occurring on a larger scale. For example, while it seems that Noise incidents are on a sharp downward trend, it would be useful to see older data, in order to see how long that downward trend has been occurring. There is always a chance, for example, that Noise incidents peaked in 2012, and are now returning to more typical levels. We are not able to tell with only 7 years of data.

We are additionally limited in our accuracy of analysis based on reporting, as mentioned in the introduction. In many cases, crimes go unreported for a variety of reasons. Depending on the crime, there may be differences in the frequency of cases that go unreported which may skew the order of call types

that we presented. For example, domestic violence is often not reported which would mean a much higher rate of incidents than reflected in the data. There could also be the case of misreporting or false reporting as well which could skew the data.

Conclusion

Through analysis of this data we have been able to better understand the distribution of reported incidents in the Burlington area as well as the moderation of reporting due to changes in time on different scales (ie. year, hour). We learned that the most common incidents reported include noise, intoxication, trespass, ordinance violation and mental health issues. An investigation into the reporting of the five most common incidences show that noise, intoxication, and ordinance violations are more likely to be reported in the evening, however, incidents involving mental health and trespassing are more likely to occur during the day. We found that intoxication, noise, and ordinance violation incidents occur most often in the warmer months, and seem to be especially high when college students are in town. This seasonal trend was not found with mental health issue incidents. We found that these incidents seem to be reported on differently than the other incidents; they seem to come from 911 calls and in-person intervention more often.

Reference

1. Government report on relationship between crime and weather.
<https://www.bjs.gov/content/pub/pdf/spcvt.pdf>