

IE6600_Sec05_Group#04_Hackathon

Abhishek Hemantkumar Taware, Vidyun Akila Sundhara Raaman

08/10/2021

Introduction and Problem Statement

Calling 911 has created a lot of differences between life and death in the United States of America. People find it as a lifeline during difficult situations. The dataset we used in this hackathon contains police responses to 911 calls in Seattle. The calls 911 receive every day is high and analyzing the calls and taking right step to balance the needs of people. The needs of the people can be easily understood by a graphical illustration of the data. We will answer questions related to the part of the city which calls 911 more, which crime takes a lot of time to clear, which crimes are mentioned more during 911 calls. We also included our interpretations, reasoning and suggestions below.

```
Seattle <- read.csv("Seattle_911.csv",na.strings="")

# Storing Location of Police Stations available from data.seattle.gov
Stations<- read.csv("Seattle_Stations.csv",na.strings="")
```

```
# Importing the Libraries
library(dplyr)
library(tidyverse)
library(tidyr)
library(ggplot2)
library(reshape2)
library(lubridate)
library(magrittr)
library(leaflet)
library(leaflet.extras)
library(formattable)
library(treemapify)
library(ggalluvial)
```

```
# Adding new columns namely Date, Month, Day Number, Year, Weekday, Hour, Block, Street
Cleaned_Seattle <- Seattle %>% mutate(Date = mdy_hms(Seattle$Event.Clearance.Date),Month=format((Date),"%B"), Day_n = day(Date), Year = year(Date), Day_Name= weekdays(Date), Hour=hour(Date)) %>%
  separate(Hundred.Block.Location, c("block", "street"), sep=" OF ")
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 394567 rows [71,
## 666, 1409, 1513, 1617, 2139, 2377, 2445, 2553, 2669, 2761, 2915, 3179, 3436,
## 3610, 3650, 4069, 4111, 4349, 4493, ...].
```

```
# Adding different periods of the day
Cleaned_Seattle <- Cleaned_Seattle %>% mutate(Day.Zone = ifelse(as.numeric(Hour) %in% 0:4, "Late Night", ifelse(as.numeric(Hour) %in% 5:8, "Early Morning", ifelse(as.numeric(Hour) %in% 9:11, "Morning", ifelse(as.numeric(Hour) %in% 12:17, "Afternoon", ifelse(as.numeric(Hour) %in% 17:19, "Evening", ifelse(as.numeric(Hour) %in% 20:23, "Night", NA)))))))

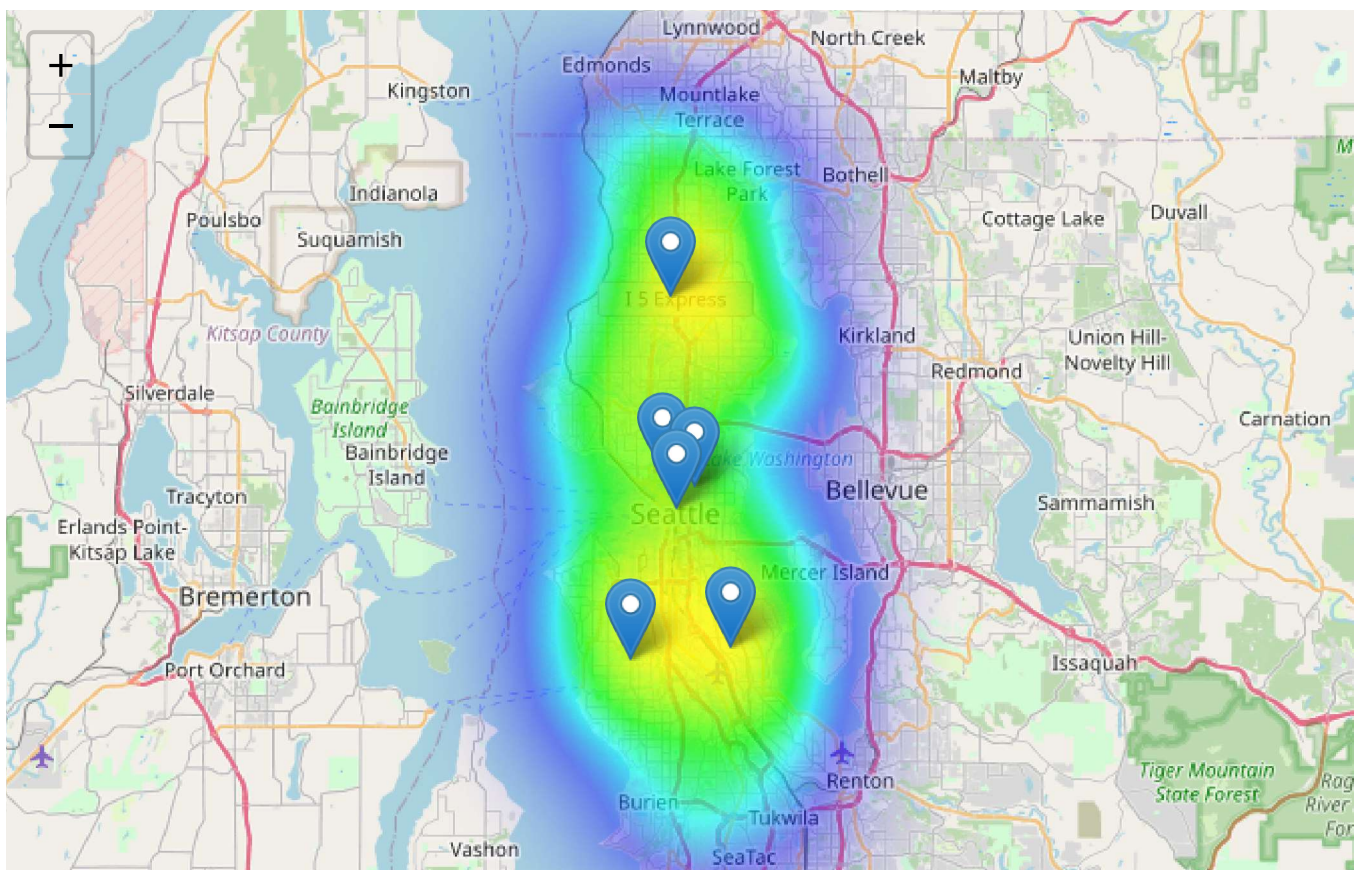
# Saving as a new file
write.csv(Cleaned_Seattle, "Cleaned_Seattle.csv", row.names = FALSE)
Cleaned_Seattle <- read.csv("Cleaned_Seattle.csv", na.strings = "")

# Replacing NA and NULL String values to proper NA Values
Cleaned_Seattle[Cleaned_Seattle == "NA"] <- NA
Cleaned_Seattle[Cleaned_Seattle == "NULL"] <- NA
```

Q1. Is every part of Seattle call 911 equally? Does the location of the police precinct influence the calls and crime?

```
# Subsetting the Latitude and Longitude of all the crime Locations
pick_fast <- Cleaned_Seattle %>%
  select(Longitude, Latitude) %>% na.omit()

# Creating a map of Seattle with heatmap showcasing the concentration of events
leaflet(pick_fast) %>%
  addTiles(group="OSM") %>%
  addHeatmap(group="heat", lng=as.numeric(pick_fast$Longitude), lat=as.numeric(pick_fast$Latitude), max=.6, blur=60) %>%
  #Adding the Markers of the nearby Precincts to the Map
  addMarkers(lng=as.numeric(Stations$Longitude), lat=as.numeric(Stations$Latitude))
```





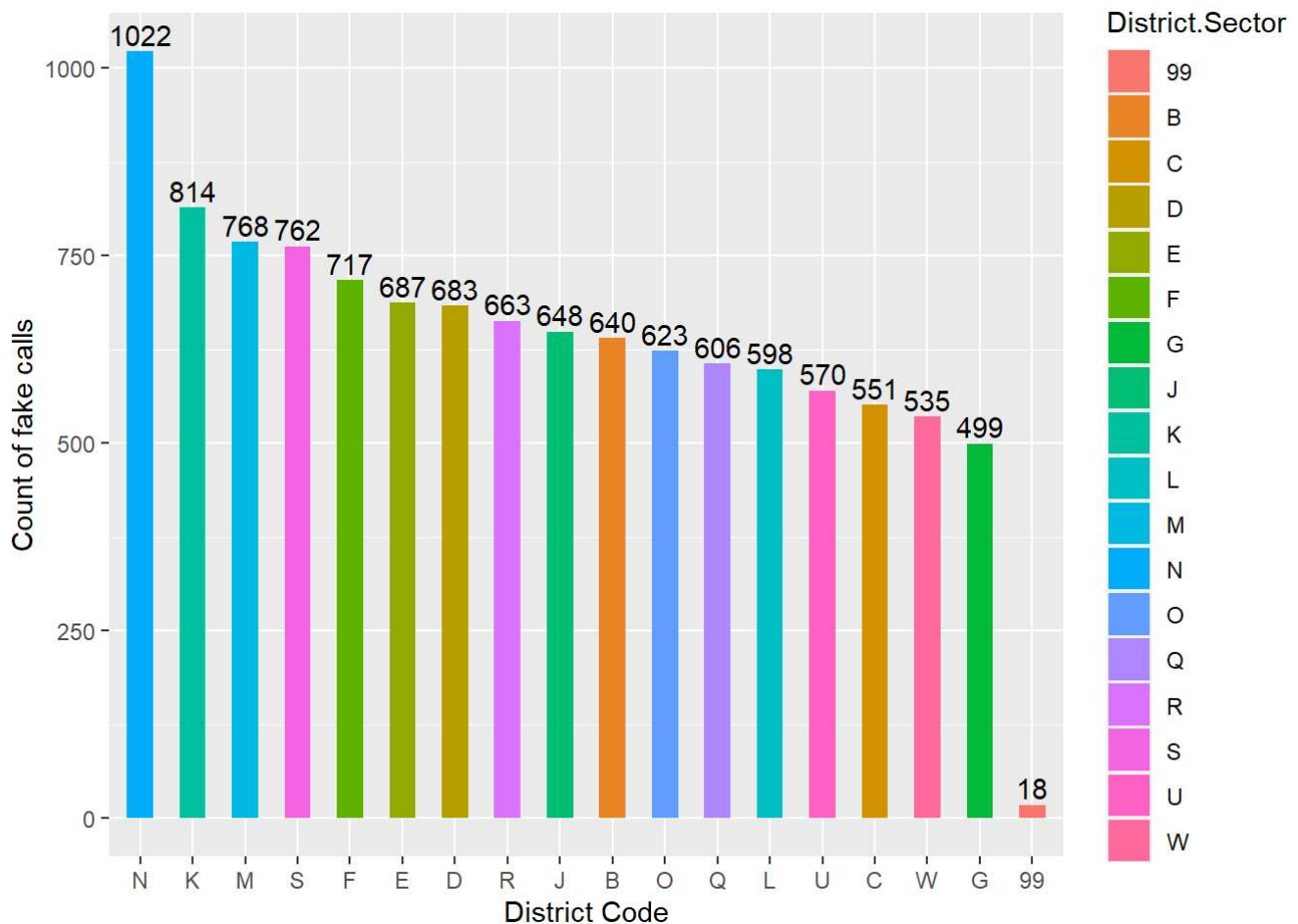
Conclusion

Since the GGmap function has recently become a paid API, we used the Leaflet library to show the heatmap of the number of 911 calls. We also marked the police precincts to understand the correlation with 911 call locations. From the above visualization, it is clear that not all parts of Seattle call 911 equally and there are some locations from which the 911 calls are fewer. This may be because of the people there finding themselves safe. The Police precincts are located around places where 911 calls are high. The police department can increase another precinct down south since the number of 911 calls is high and the police station is distant.

Q2. What was the distribution of fake reportings in each of the sectors?

```
# Filtering the records containing NA Values in Event.Code.Group
District_Data <- Cleaned_Seattle %>% group_by(District.Sector) %>%
  filter(is.na(Event.Clearance.Group)) %>%
  summarise(Count = sum(!is.na(CAD.Event.Number), na.rm = TRUE)) %>%
  arrange(desc(Count)) %>% drop_na()

# Plotting the bar graph showing districts with the highest counts of fake calls to the lowest
t
ggplot(data=District_Data, aes(x=reorder(District.Sector, -Count), y = Count, width = 0.5, fill=District.Sector)) +
  geom_bar(stat="identity") + labs(x = "District Code", y = "Count of fake calls") + geom_text(
    aes(label = Count), vjust = -0.3)
```



Conclusion Based on the data on hand, we wanted to see how many fake calls have been registered in each of the district. For this, we calculated the count of na values in the Event.Clearance.Group column which is used to specify what was the actual emergency at hand for which the 911 call was based. Based on this data, we were able to create a bar chart showing the count of fake calls that have been made in all the districts over the years which basically dwarfs the number of real calls made. Out of all these records, its clearly seen that Sector N is significantly above the other sectors with respect to the number of fake / no reporting calls made.

Q3. What were the reportings of the most number of calls?

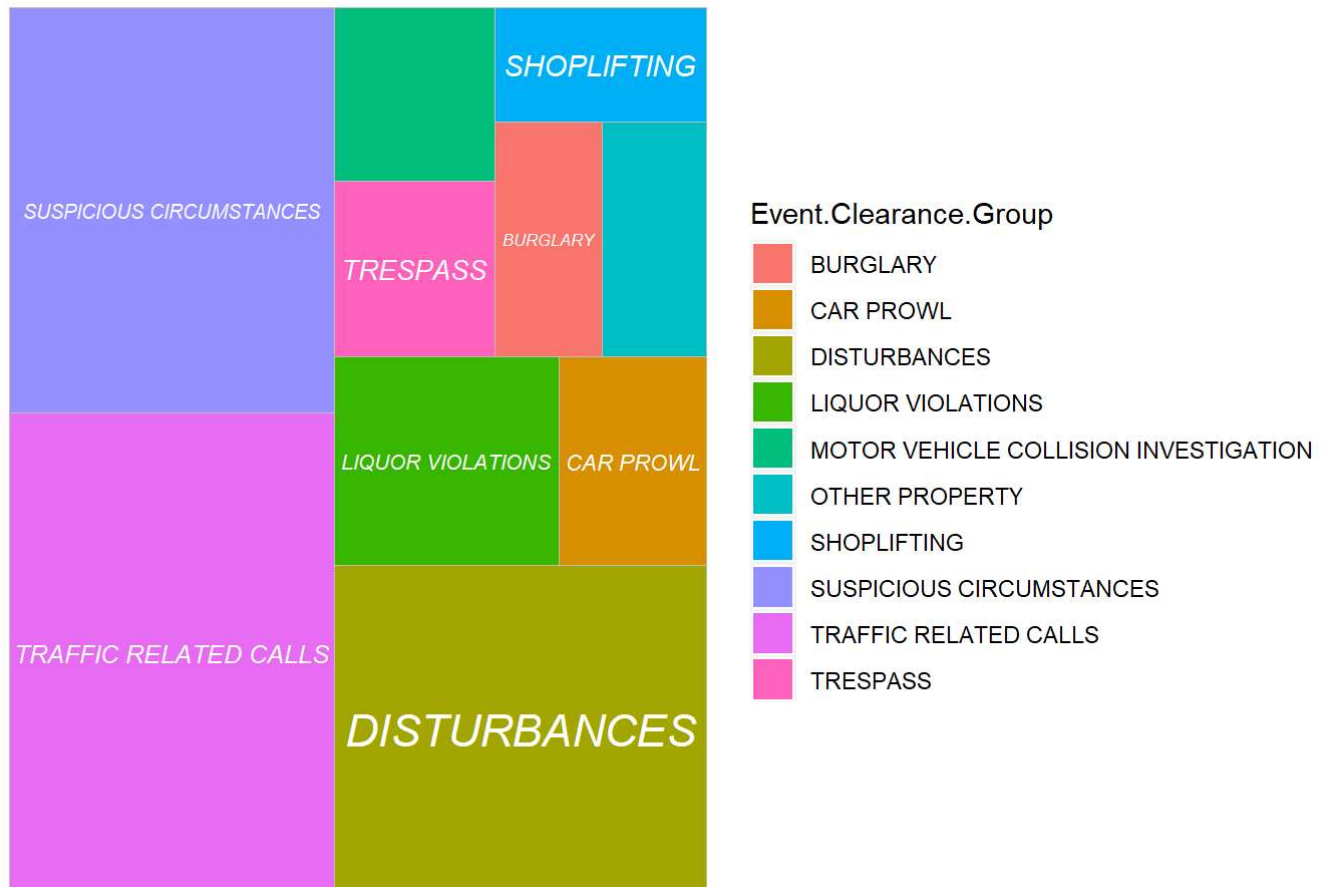
#Grouping and getting Top 10 Crimes that happen in Seattle

```
Tree <- summarise(group_by(Cleaned_Seattle,Event.Clearance.Group), Count.of.crimes = sum(CAD.Event.Number)) %>% arrange(desc(Count.of.crimes)) %>% slice(1:10)
```

#Plotting the Treemap to showcase the crimes

```
ggplot(Tree, aes(area = Count.of.crimes, fill = Event.Clearance.Group, label = Event.Clearance.Group)) +
  geom_treemap() + ggtitle("Top 10 Crimes in Seattle") +
  geom_treemap_text(fontface = "italic", colour = "white", place = "centre", grow =
FALSE) + theme(plot.title = element_text(hjust = 0.5))
```

Top 10 Crimes in Seattle



Conclusion Based on the dataset, we decided to figure out what were the Top 10 number of crimes that have happened in Seattle over the duration of the dataset. We created a Treemap to understand and aesthetically display the Top 10 types of calls that have happened in Seattle based on the number of reporting. Out of this, we can clearly see that Traffic Related Calls, Suspicious Circumstances Calls and Disturbance Calls dwarf the others which are quite generic as compared to the violent crimes in the area which do not even come into the Top 10 Reportings.

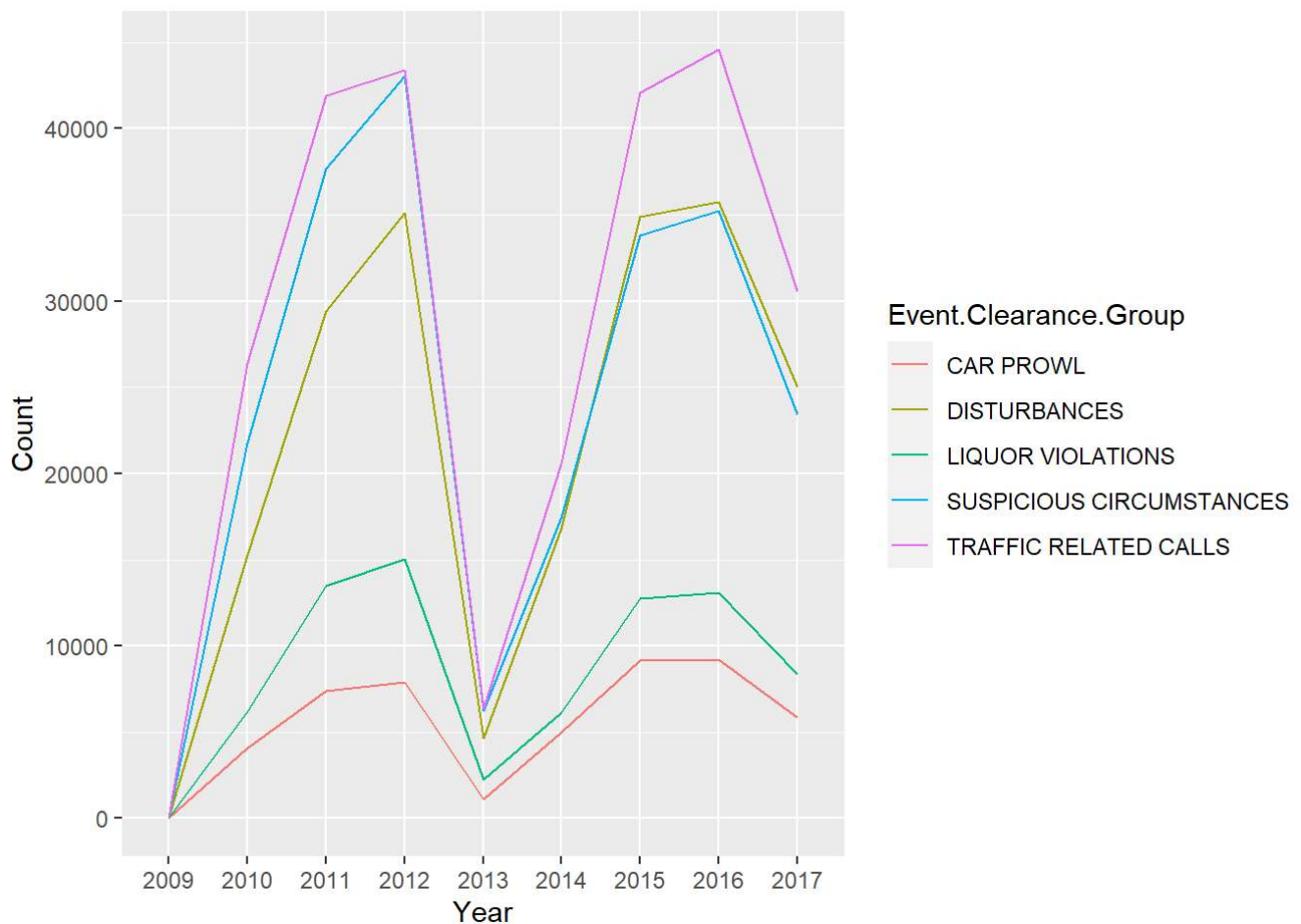
Q4. What was the trend of the Top 5 Reported Call Types in Seattle over the past 9 Years?

```
# Getting the records of most reported crimes
Crime <- Cleaned_Seattle %>% group_by(Event.Clearance.Group) %>% summarize(Count = n()) %>% d
rop_na() %>% arrange(desc(Count))
```

```
# Filtering & Grouping records based on Top 5 Crimes
Crime_over_Years <- Cleaned_Seattle %>%
  group_by(Event.Clearance.Group, Year) %>%
  filter(Event.Clearance.Group %in% Crime$Event.Clearance.Group[1:5]) %>%
  summarize(Count = n()) %>% arrange(desc(Event.Clearance.Group)) %>% drop_na()
```

`summarise()` has grouped output by 'Event.Clearance.Group'. You can override using the `groups` argument.

```
# Plotting the Line graph to showcase the trends of Top 5 crimes over the years
ggplot(data = Crime_over_Years, aes(x=Year, y=Count, colour = Event.Clearance.Group, group =
  Event.Clearance.Group)) + geom_line()
```

Conclusion Based on the data at hand, we wanted to figure if there was a rise / fall in the number of most reported call types. The line chart that is generated above, clearly showcases that there was no type of call has increased over any other type over the past 9 years and at the same time we can also derive that in 2013 there was a significant drop in the number of crimes being reported in Seattle which again got shot up in 2014. At the same time it can also be derived that the number of calls being made to 911 has dropped again to a significantly lower number in 2017 compared to the past couple of years.

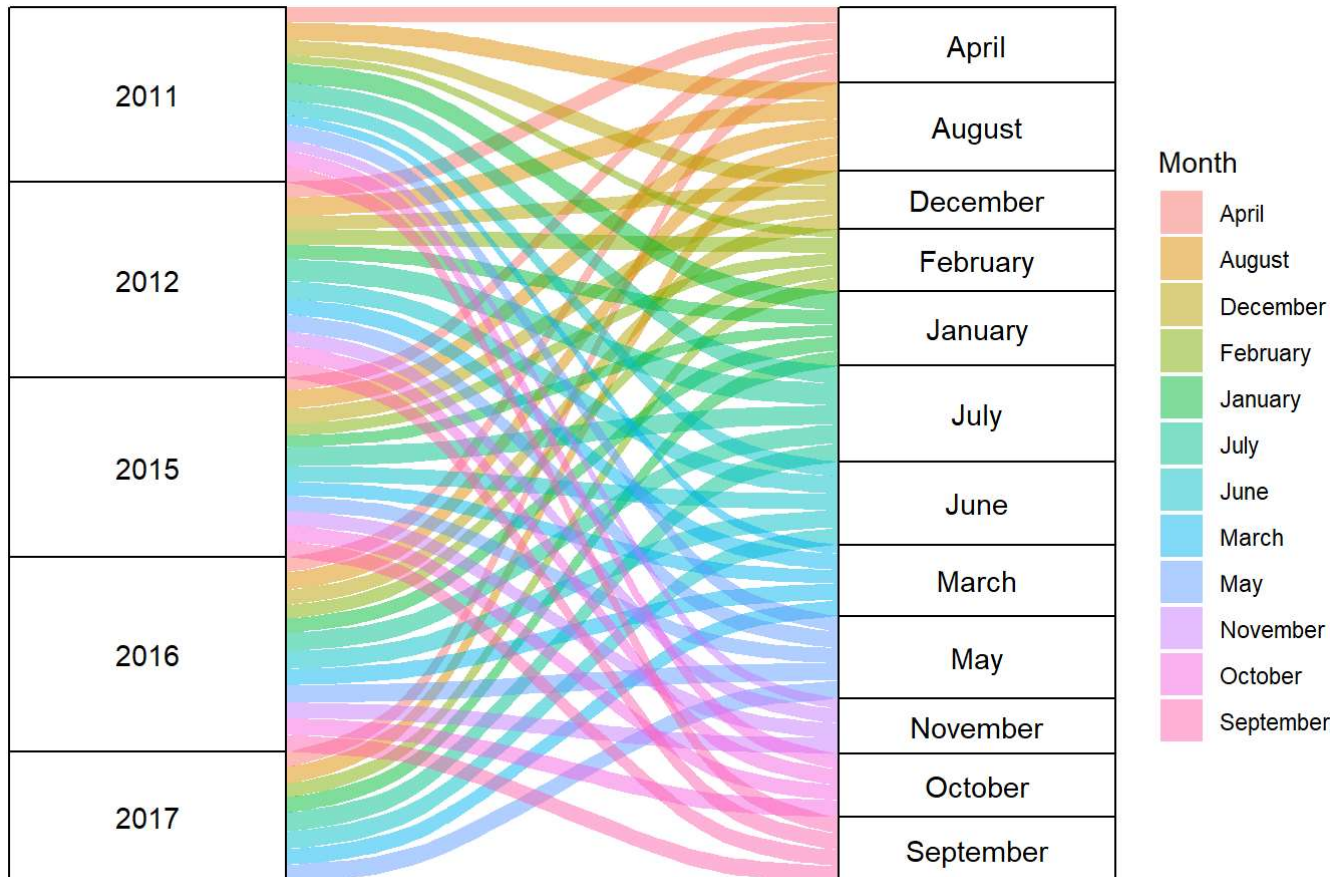
Q4. Every year, how many crimes were distributed during each month?

```
Codes <- Cleaned_Seattle %>% group_by(Year) %>% summarize(Count = n()) %>% arrange(desc(Count)) %>% drop_na() %>% slice(1:5)
```

```
Code_switch <- Cleaned_Seattle %>% group_by(Year, Month) %>% filter(Year %in% Codes$Year) %>% summarize(Count = n()) %>% arrange(desc(Count)) %>% drop_na()
```

`summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.

```
ggplot(data = Code_switch,
       aes(axis1 = Year, axis2 = Month, y = Count)) +
  geom_alluvium(aes(fill = Month)) +
  geom_stratum() +
  geom_text(stat = "stratum",
           aes(label = after_stat(stratum))) +
  scale_x_discrete(limits = c("Year", "Month"),
                  expand = c(0.15, 0.05)) +
  theme_void()
```



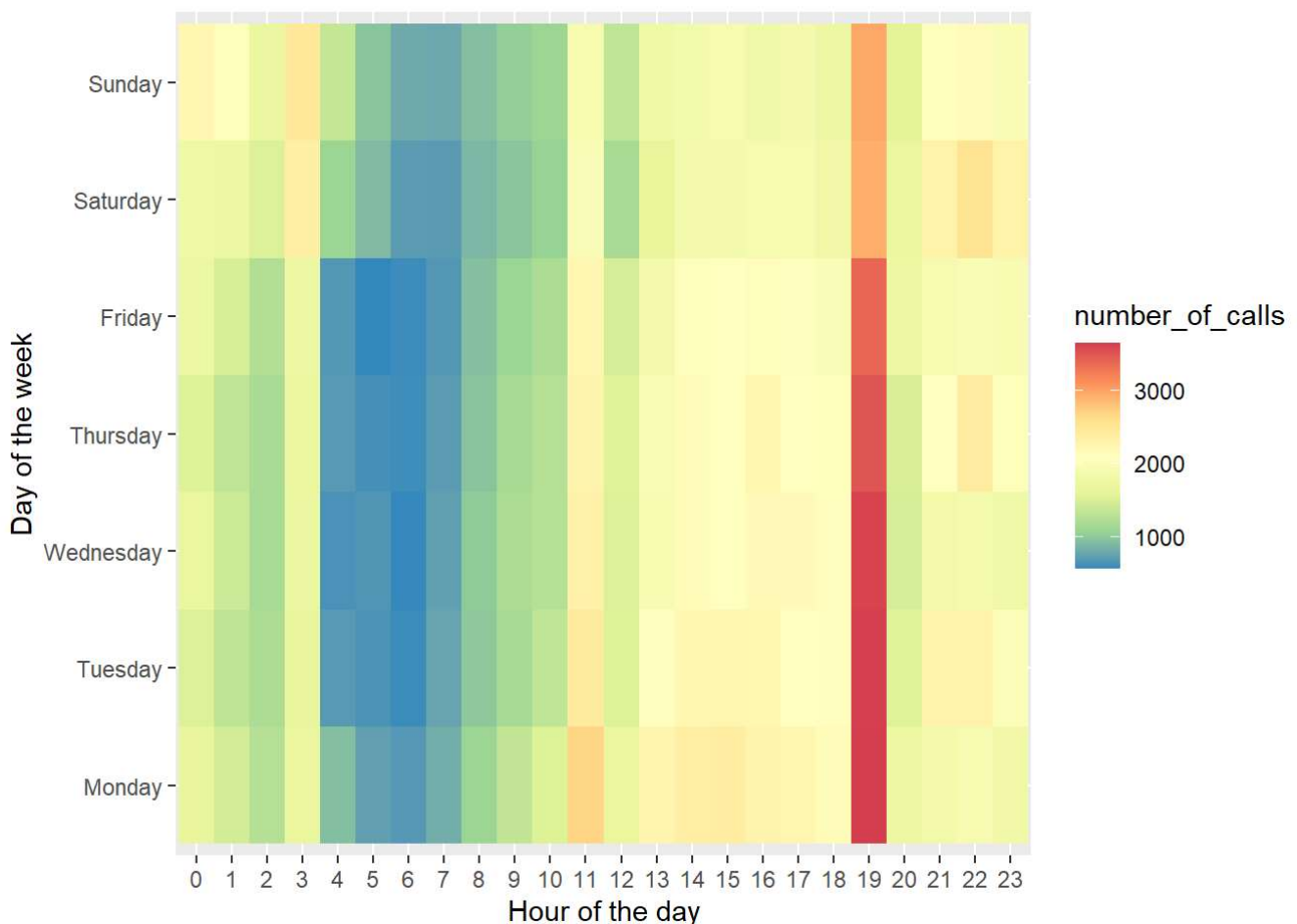
Conclusion Based on the dataset that was given, we wanted to figure out what was the distribution of the number of calls made during different months of the year over the past 9 years of data. Using the alluvial chart, we were able to aesthetically showcase which months have seen a higher number of call reportings as compared to others. of which we can see that July, September sees a higher number of calls as compared to others.

Q6. Is every hour of the day record a the same number of 911 calls? Do weekends influence the call count?

```
C1 <- Cleaned_Seattle %>% na.omit() %>%
  group_by(Day_Name, Hour) %>%
  summarise(number_of_calls = n()) %>% na.omit() %>%
  # Using geom_tile to create a heatmap corresponding to the days and the hours at which calls
  # were made that day
  ggplot(aes(x=factor(Hour,levels=as.character(0:23)), y = factor(Day_Name,levels=c("Monday",
"Tuesday","Wednesday","Thursday","Friday","Saturday","Sunday")), fill = number_of_calls)) +
  geom_tile() +
  #Adding Labels and Colours to the heatmap
  labs(x = "Hour of the day", y = "Day of the week") +
  scale_fill_distiller(palette = "Spectral")
```

`summarise()` has grouped output by 'Day_Name'. You can override using the `.groups` argument.

C1

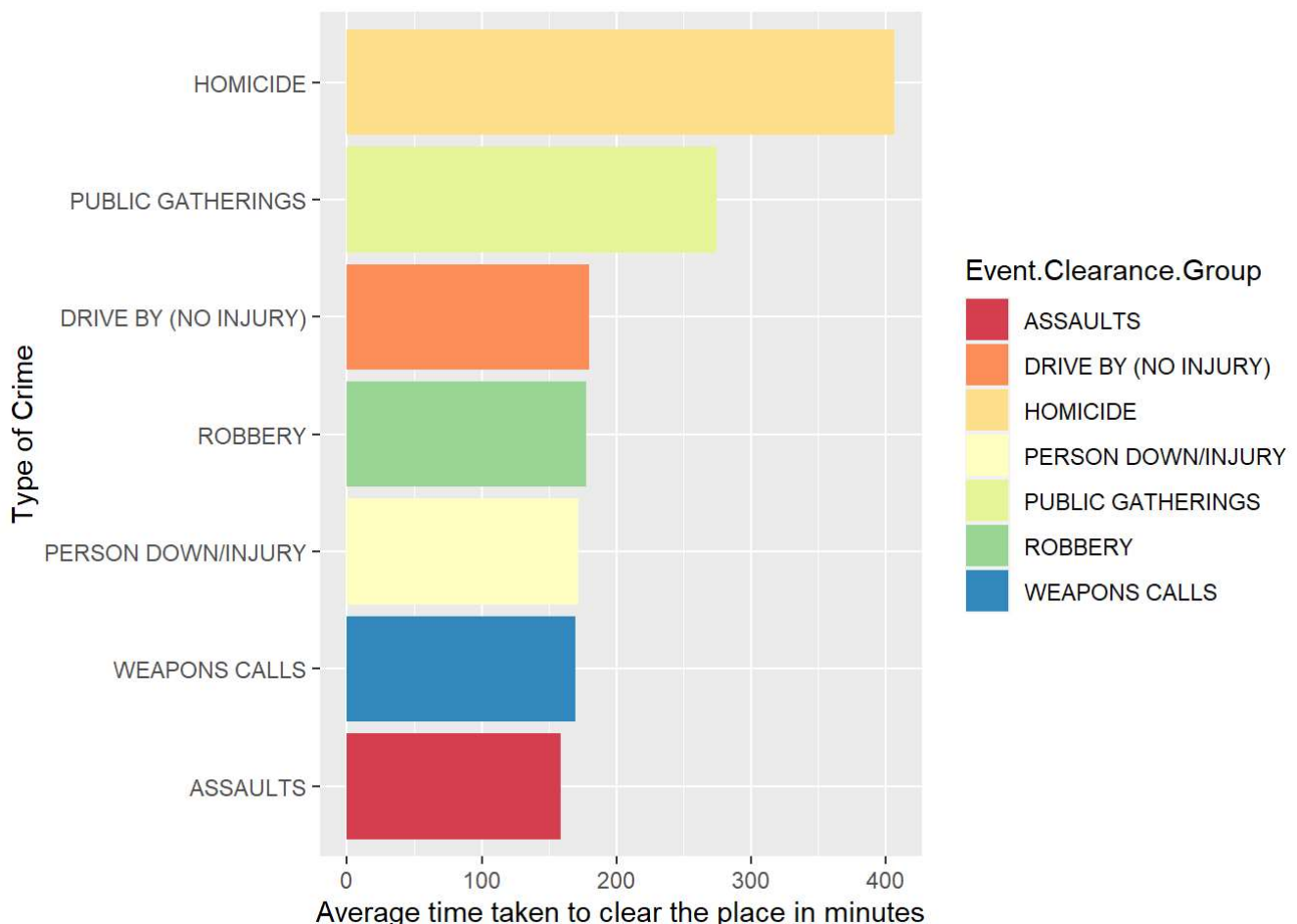


Conclusion We used a tile-based heatmap to understand the distribution of the number of calls every day every hour. We used spectral color to understand the frequency of calls. From this visualization, we found that the number of 911 calls is high during 19.00 PM every day of the week compared to other hours and calls are less from 4.00 AM to 8.00 AM. During the weekend, we found that a similar trend but in different amounts. The number of calls by the weekend is less compared to weekdays. This may be because of people enjoying their weekend and not finding anything suspicious. The blur of red color in the heatmap clearly indicates the above-mentioned trend.

Q7. Is every event mentioned in the 911 calls consume a similar amount of time to sort them off? Do fatal-related 911 calls take a lot of time to clear?

```
Cleaned_Seattle1=Cleaned_Seattle[!is.na(Cleaned_Seattle$At.Scene.Time), ]
Cleaned_Seattle1 <- Cleaned_Seattle1[!is.na(Cleaned_Seattle1$Date), ]

#Creating a new column to find out the difference between the reported time of call and closing time of the case
Cleaned_Seattle1 <- Cleaned_Seattle1 %>% mutate(At.Scene.Time=mdy_hms(At.Scene.Time), Date=ymd_hms(Date),time_dif=round(abs(as.numeric(difftime(At.Scene.Time, Date, units = "mins"))))) %>%
>% group_by(Event.Clearance.Group) %>% summarize(avg_time_taken=mean(time_dif,na.rm=TRUE)) %>%
>% arrange(desc(avg_time_taken)) %>% head(7) %>%
  #Plotting Top 7 crimes consuming the most amount of time per Call
  ggplot(aes(x=reorder(Event.Clearance.Group,avg_time_taken),y=avg_time_taken,fill=Event.Clearance.Group))+ geom_bar(stat="identity") + coord_flip() + scale_fill_brewer(palette = "Spectral") + labs(x="Type of Crime",y="Average time taken to clear the place in minutes")
Cleaned_Seattle1
```



Conclusion There are more than 39 types of crimes mentioned in the 911 calls. The difference between the event clearing time and the crime registered time indicates that the time taken to resolve the problem mentioned in the call. Using barplot we can easily understand the time taken by each event to clear the crime location. From the barplot above, it is clear that each event takes consumes different time amounts and Homicides consume an average of 300 minutes to clear. All the top 5 time-consuming events are related to casualty. Therefore, it is clear that fatal related events consume a lot of time to clear.