HW6 87 F21

Lucía Carrera

11/5/2021

Burlington Police Department Data

This is a summary of calls to the Burlington Police Department from January 1, 2020 to January 1, 2021.

```
knitr::opts_chunk$set(echo = TRUE)

#  load tidyverse, lubridate, and leaflet
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(leaflet)


#  Use read.csv to read in the data.

b <- read.csv('BPDIncidents2020.csv')
```

Organizing Date/Time data

Note: some data may appear in Spanish due to my computers software.

```
#  Look at the call_time variable:  Print the first few observations, and do class
().   It is in a month/day/year hour/minute format, but R does not recognize it as
date/time data.  Using lubridate functions, we can 'extract' the date from call_tim
e using the function mdy_hm(). (Other extraction functions are: dmy, myd, ymd, ydm,
dym, mdy, ymd_hm)  Run the following code, print the first few observations of the
new variable that has been created, and do class() for this new variable. (If R rec
ognizes it as date/time data, the class should be "POSIXct" "POSIXt")
b <- b %>%
  mutate(Call = mdy_hm(call_time))

tibble(b$Call)
```

```
## # A tibble: 23,584 x 1
##    `b$Call`
##    <dttm>
##  1 2020-01-01 00:04:00
##  2 2020-01-01 00:13:00
##  3 2020-01-01 00:41:00
##  4 2020-01-01 01:06:00
##  5 2020-01-01 01:11:00
##  6 2020-01-01 02:05:00
##  7 2020-01-01 02:13:00
##  8 2020-01-01 03:10:00
##  9 2020-01-01 04:41:00
## 10 2020-01-01 09:36:00
## # ... with 23,574 more rows
```

```
class(b$Call)
```

```
## [1] "POSIXct" "POSIXt"
```

```
#  Then we can use the functions date(), month(), wday(), and day() extract values
from the date. Create a variable for Month, for Weekday, Day, and Date.  Make sure
the Month and Weekday have character values (e.g., Jan, Feb...Mon, Tue).  Print abo
ut 10 observations of the new variables that have been created -- to get a variety,
use sample_n().
b <- b %>%
  mutate(Weekday = wday(Call, label = TRUE),
                   Mon = month(Call, label = TRUE),
                   Date = date(Call),
                   Day = day(Call))

sample_n(b, 10) %>% select(Weekday, Mon, Date, Day)
```

```
##    Weekday Mon       Date Day
## 1    mi\\. abr 2020-04-15  15
## 2    mi\\. abr 2020-04-15  15
## 3    mi\\. sep 2020-09-09   9
## 4    ma\\. ene 2020-01-14  14
## 5    do\\. jun 2020-06-07   7
## 6    do\\. ene 2020-01-12  12
## 7    sá\\. oct 2020-10-10  10
## 8    vi\\. feb 2020-02-21  21
## 9    ma\\. oct 2020-10-06   6
## 10   lu\\. abr 2020-04-06   6
```

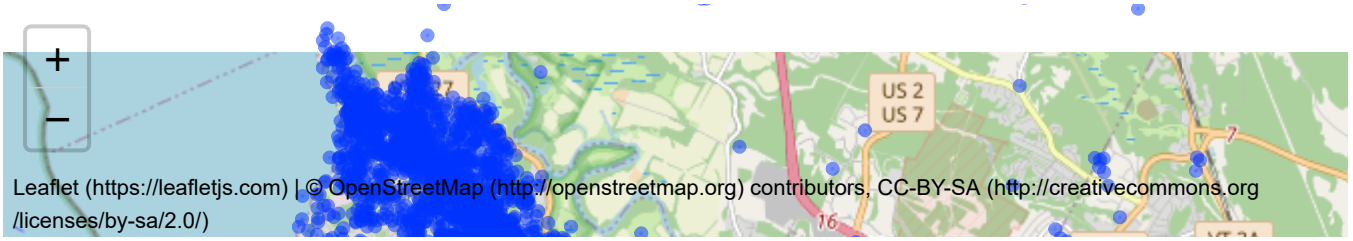
Map of Calls in Burlington

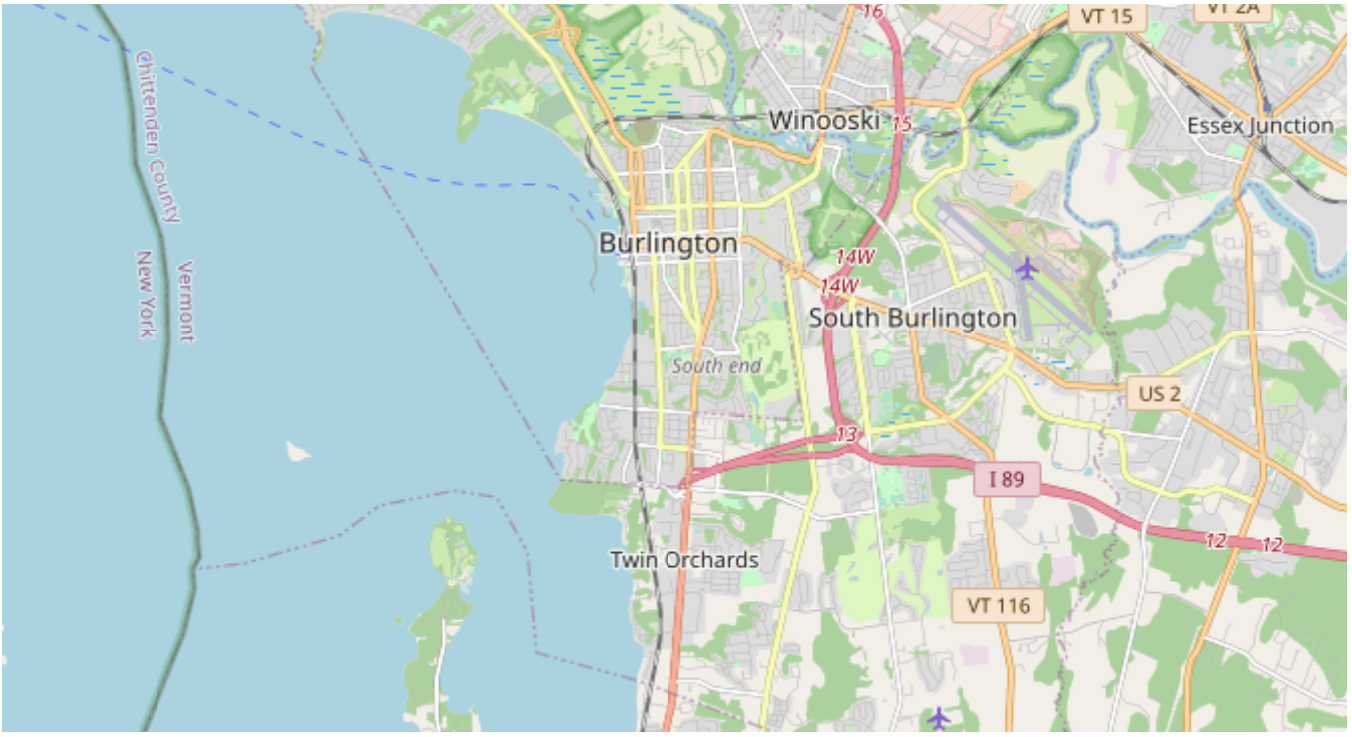
```
#  Do a scatterplot of Latitude by Longitude, to see where calls are made in Burlin
gton. Do you recognize the shape of the data? It looks like Burlington
# ggplot(data = b,
#        mapping = aes(x = Longitude, y = Latitude)) +
#  geom_point()

#  Using leaflet, make a map showing circles on the location of each incident.  Sta
rt here:
m <- leaflet() %>%
  addTiles() %>%
  setView(lat = 44.475, lng = -73.212, zoom = 12)

#  (Once you get the leaflet plot working, remove the scatterplot of Longitude by L
atitude)
m  %>% addCircles(lng = b$Longitude, lat = b$Latitude)
```

```
## Warning in validateCoords(lng, lat, funcName): Data contains 2020 rows with
## either missing or invalid lat/lon values and will be ignored
```

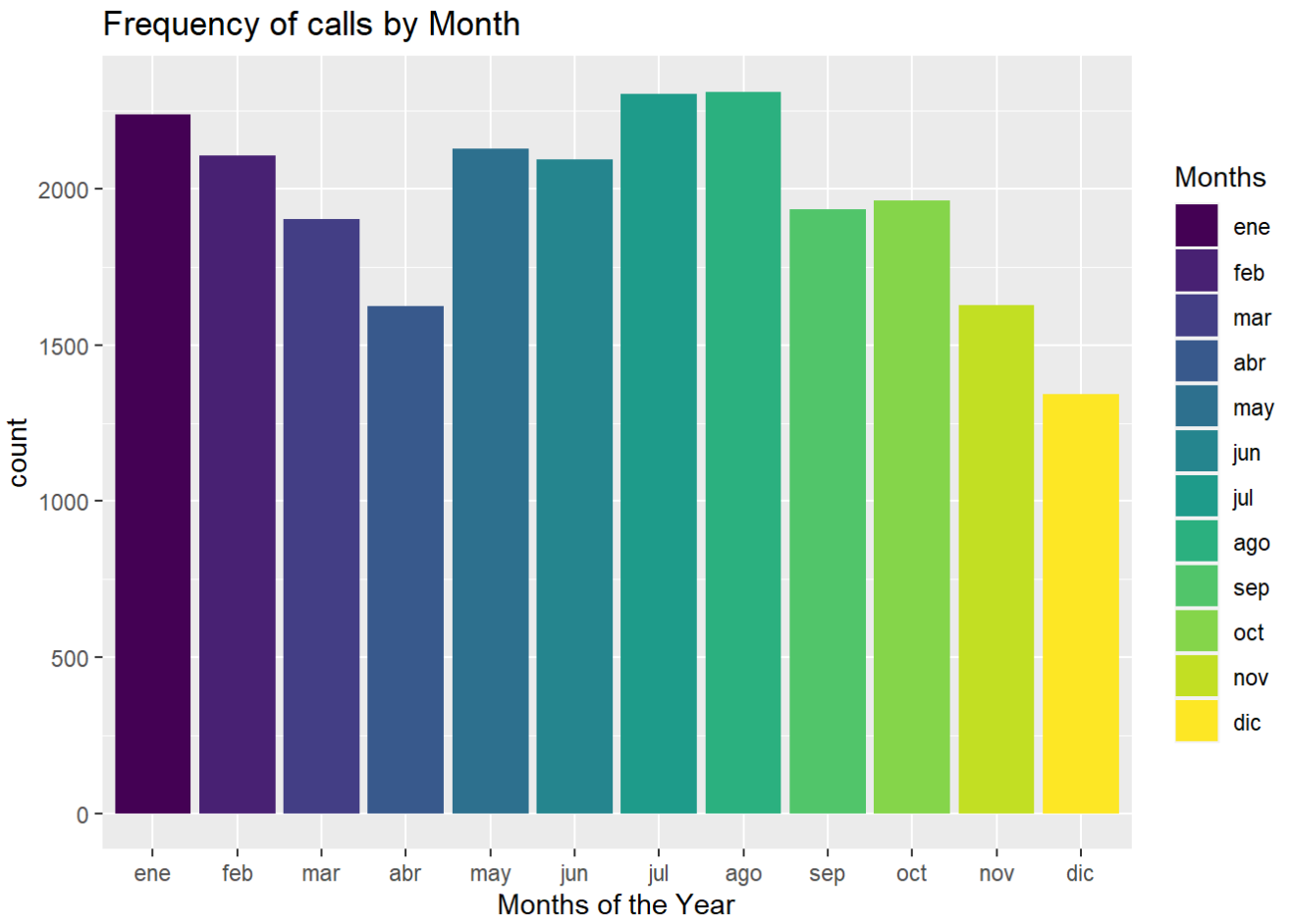




Frequency of Calls by Month

Here we can see the number of calls per month. The most interesting thing about this graph is that there is a decrease of calls in december and abril. I believe that the main reason for this is because students tend to leave Burlington in these months.

```
#  Make a bar graph of the frequency of calls by month.  Use color in your graph.
Comment on the trend you observe in the white space above.
ggplot(data = b,
       mapping = aes(x = Mon, fill = Mon)) +
   geom_bar() +
  labs(x="Months of the Year", title = "Frequency of calls by Month", fill = "Month
s")
```



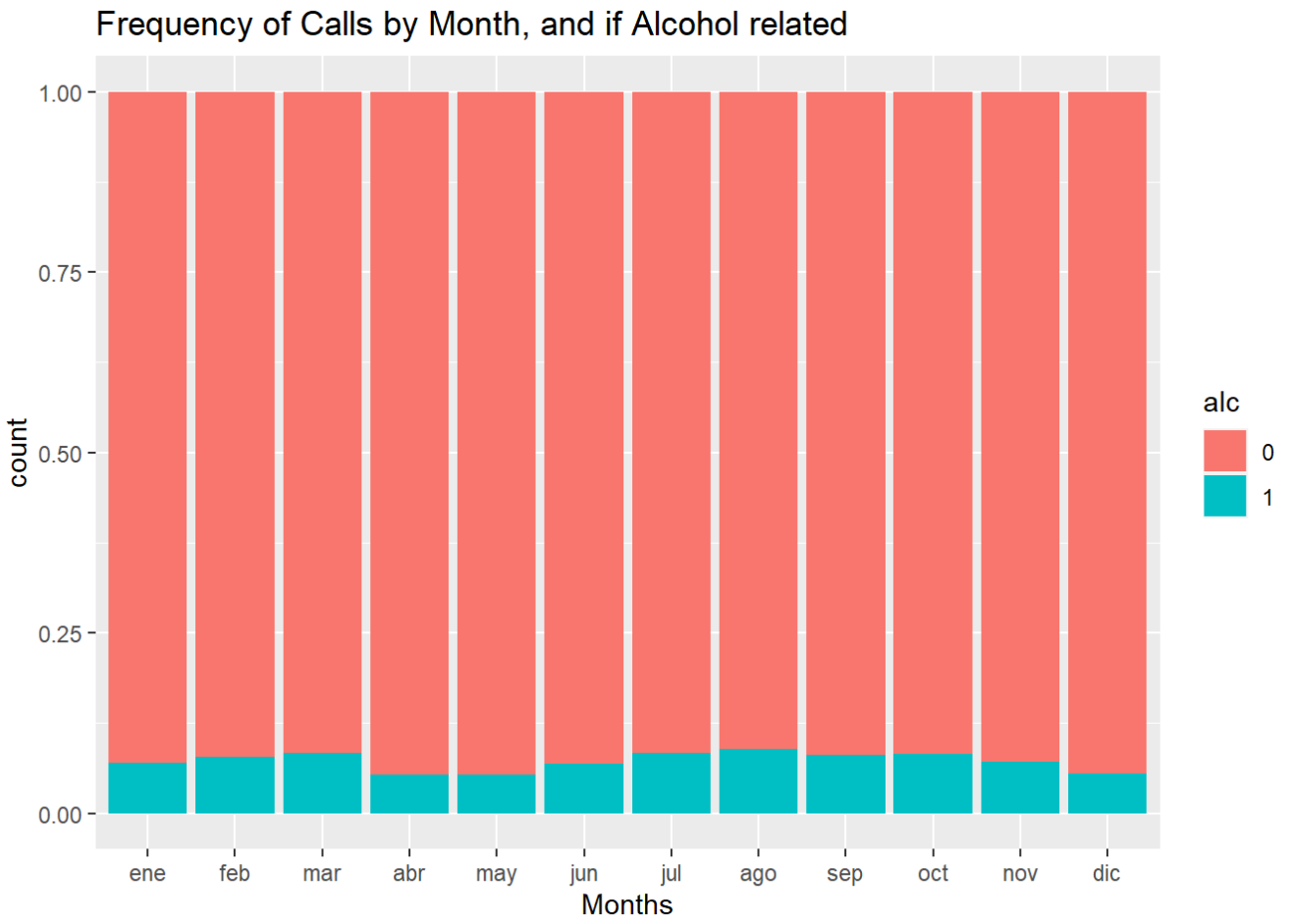
Proportion of Calls related to Alcohol

Proportionately there seems to be at its lowest in december, abril and may. This could be due to students having finals and the holidays. Alcohol related calls also seem to rise in the summer months, peaking in august. This could be due to the hot weather and vacations. January, February & March have also a high proportion of calls, this might be because ski season is in full swing and apres ski become more popular. 0 is not alcohol related and 1 is alcohol related.

Note: position = 'fill' placed under Sheila's approval to better grasp the graphs' information.

```
#  Make a bar graph of the frequency of calls by month.  Use alcohol_related as a f
ill variable.  Note that you may need to treat alcohol_related as a factor to make
the graph work.  Comment on the trend you observe in the white space above.
b <- b %>% mutate(alc = as.factor(alcohol_related))

ggplot(data = b,
       mapping = aes(x = Mon, fill = alc)) +
   geom_bar(position = 'fill') +
  labs(x = "Months", title = "Frequency of Calls by Month, and if Alcohol related")
```

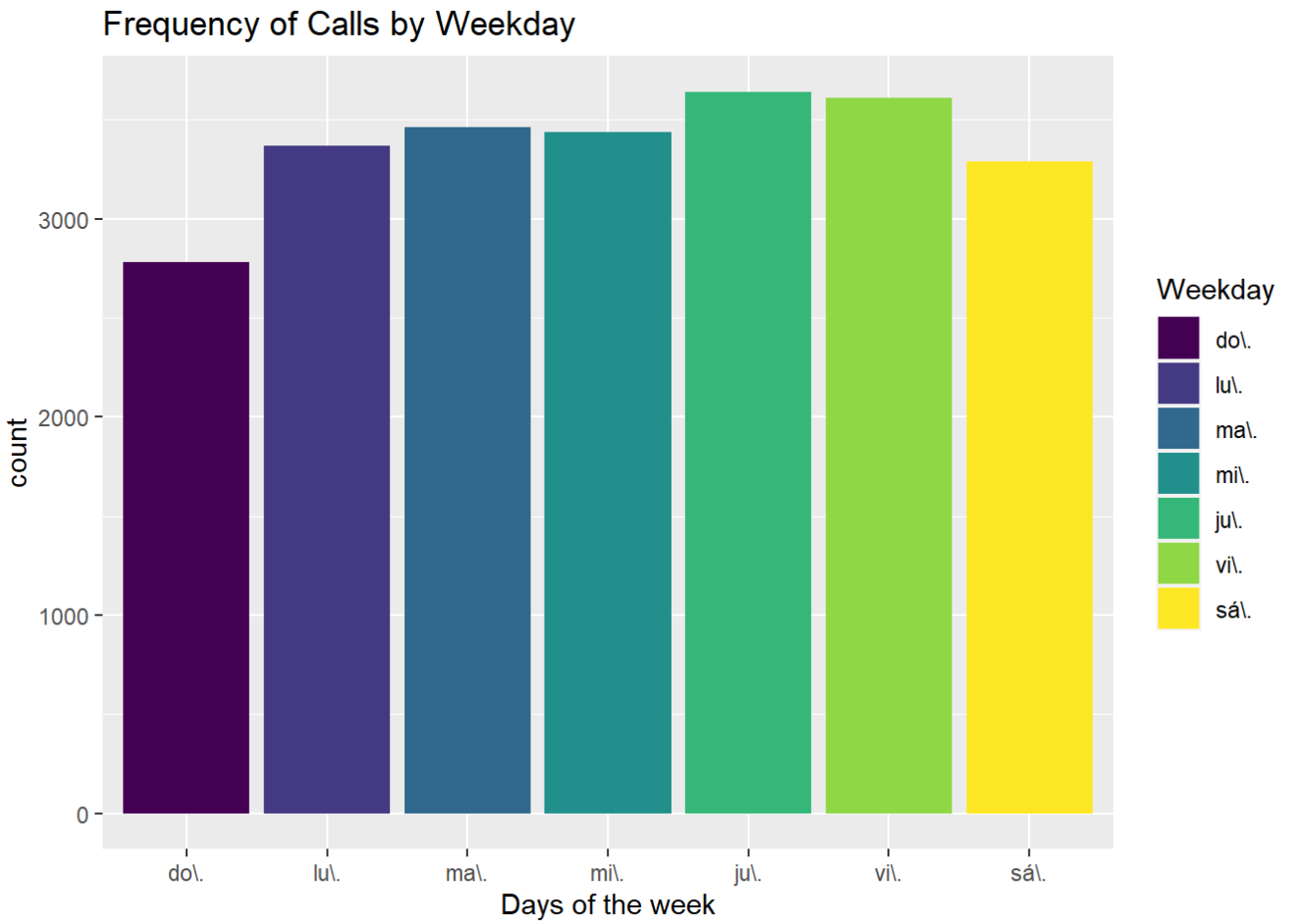


Frequency of Calls by WeekDay

Thursday seems to be the most popular day, while Sunday is the day where less calls are made. I think that Sundays have less calls because usually people prefer to stay in and not do much to prepare for the following week.

```
#  Make a bar graph of the frequency of calls by weekday.  Use color in your graph.
Comment on the trend you observe in the white space above.

ggplot(data = b,
       mapping = aes(x = Weekday, fill = Weekday)) +
   geom_bar() +
  labs(x = "Days of the week", title = "Frequency of Calls by Weekday")
```

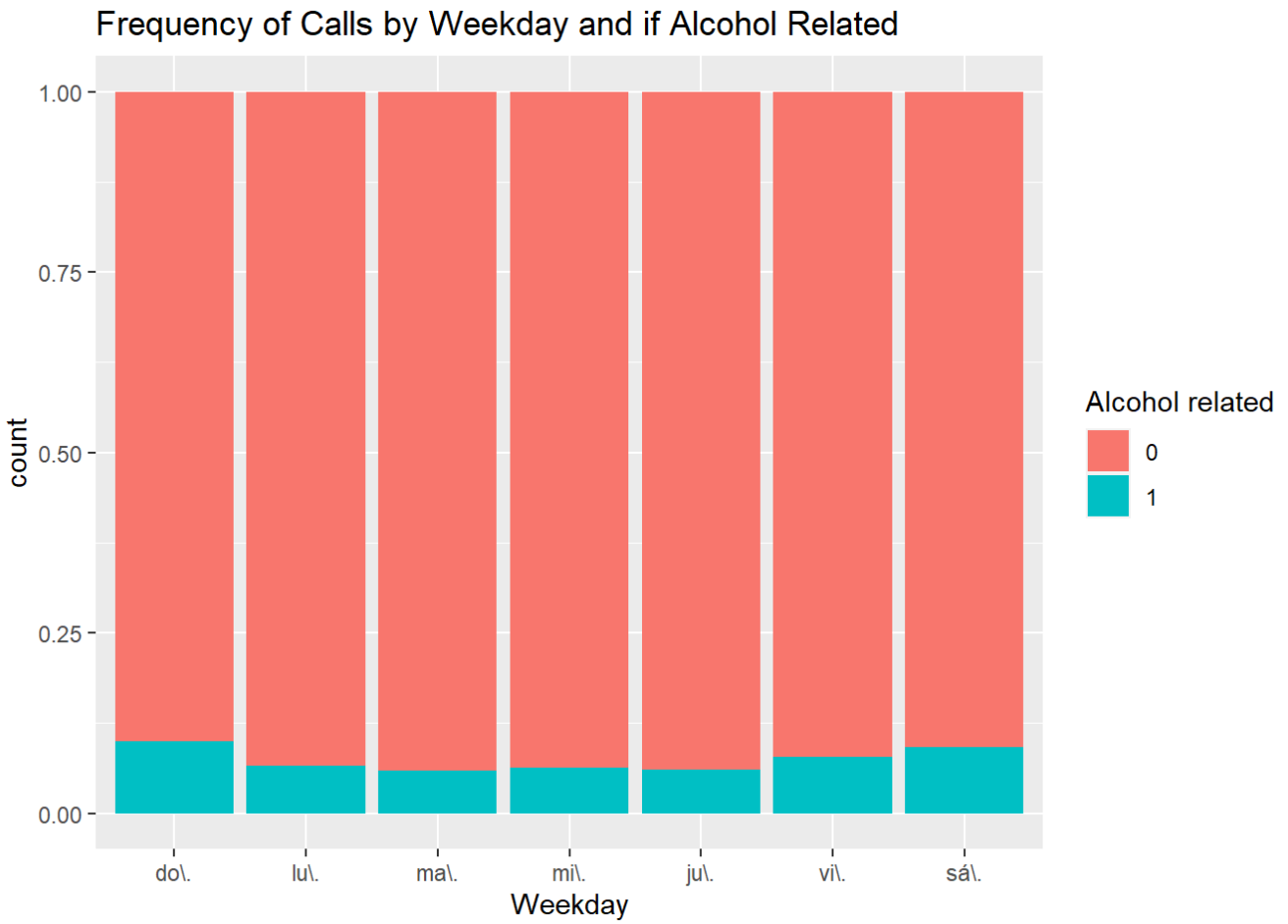


Proportion of Calls related to Alcohol

Alcohol related calls seem to increase in the weekend, which makes sense as fewer people need to go to work or school during these days. I am including Friday as part of the weekend. 0 is not alcohol related and 1 is alcohol related.

Note: position = 'fill' placed under Sheila's approval to better grasp the graphs' information.

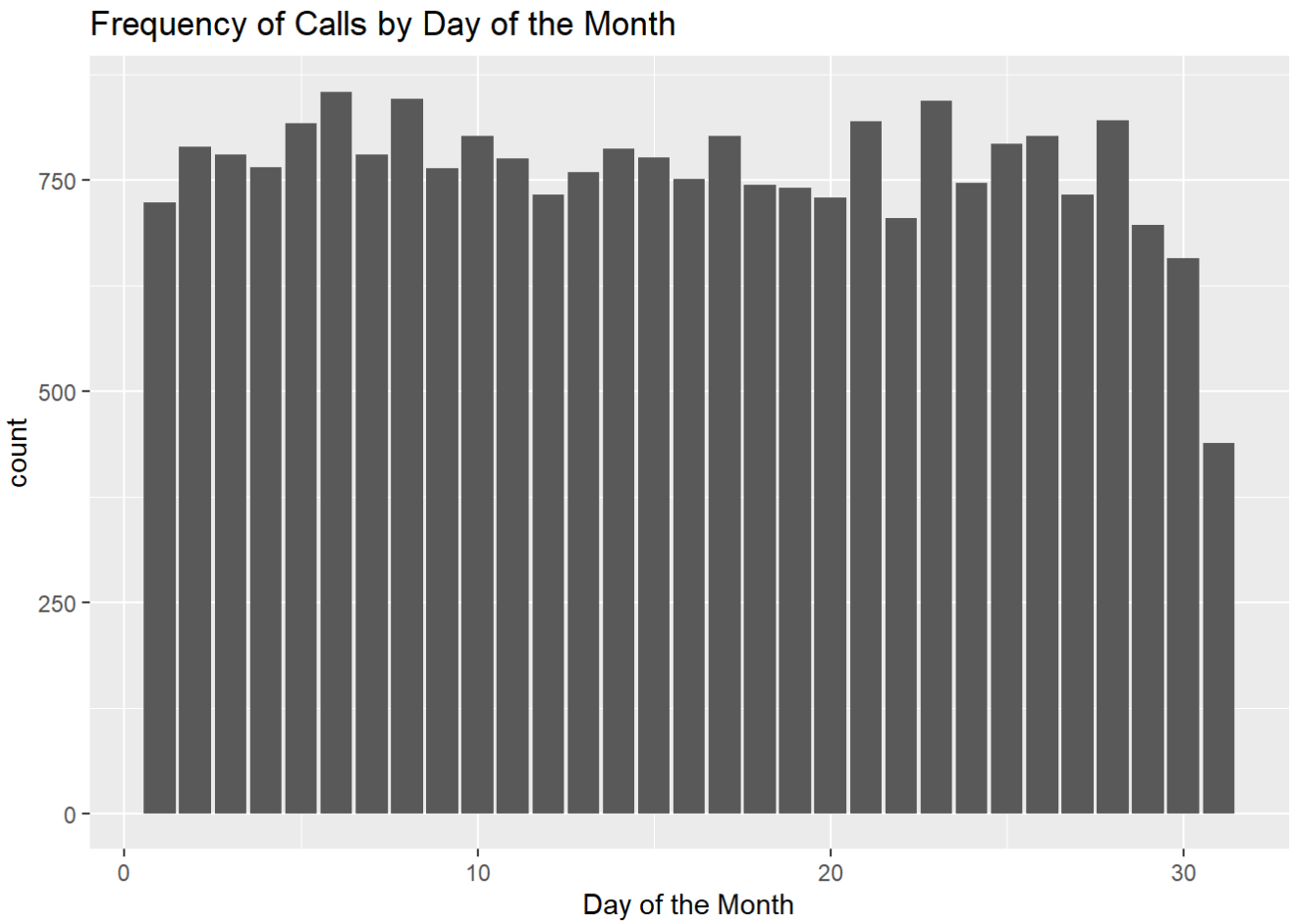
```
#  Make a bar graph of the frequency of calls by weekday.  Use alcohol_related as a fill variable.  As above,  you may need to treat alcohol_related as a factor to make the graph work.  Comment on the trend you observe in the white space above.
ggplot(data = b,
       mapping = aes(x = Weekday, fill = alc)) +
   geom_bar(position = 'fill') +
   labs(x = "Weekday", fill = "Alcohol related", title = "Frequency of Calls by Weekday and if Alcohol Related")
```



Frequency of Calls by Day of the Month

The day with the least amount of calls is the 31st. The logic to this is that 5 months in the year do not have the 31 days.

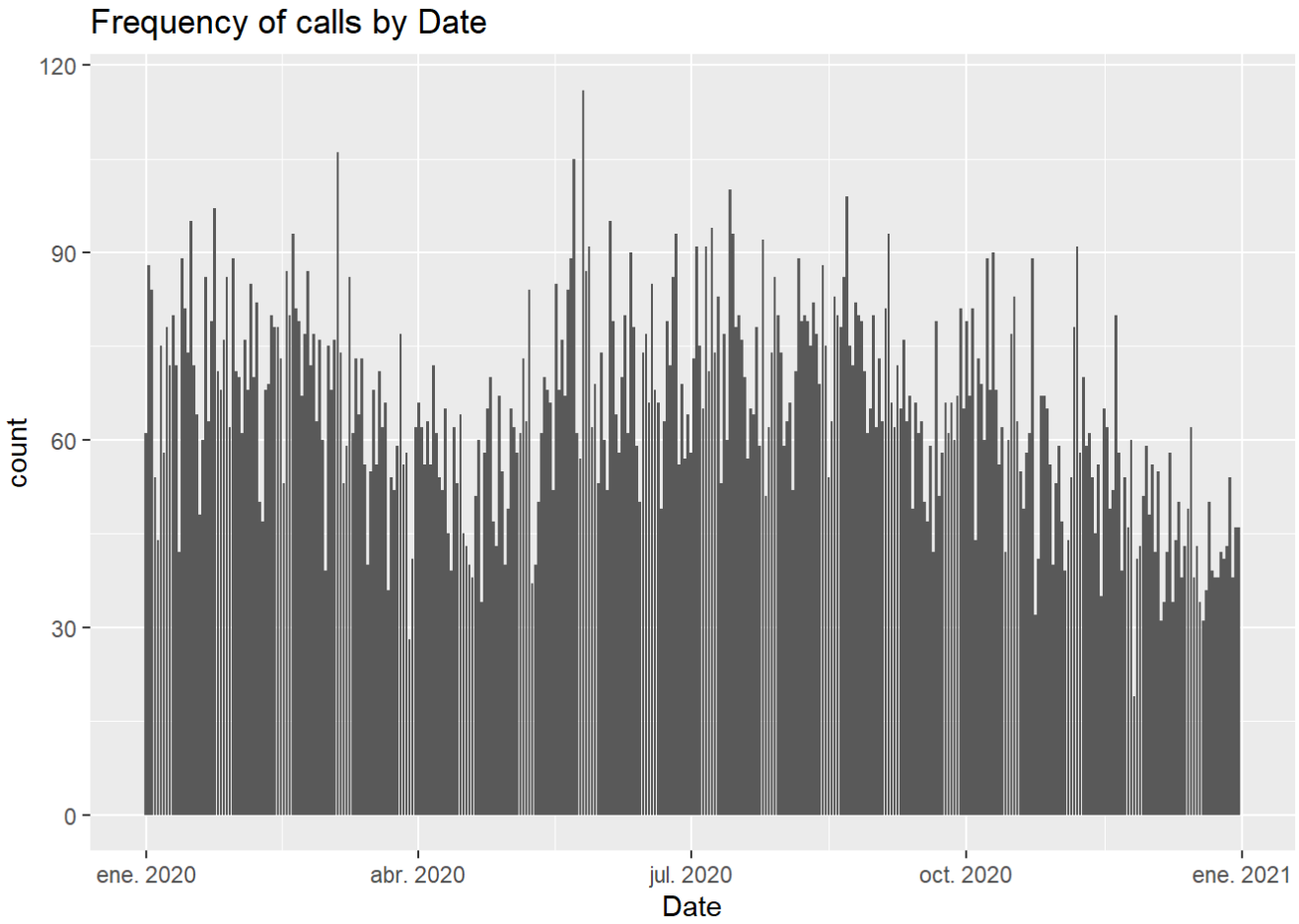
```
#  Make a bar graph of the frequency of calls by day of the month. Comment on the t
rend you observe in the white space above.
ggplot(data = b,
       mapping = aes(x = Day)) +
   geom_bar() +
  labs(x = "Day of the Month", title = "Frequency of Calls by Day of the Month")
```



Frequency of Calls by Date

By looking at the graph below we can see that something big must have happened at the end of may or beginning of june, that made a lot of people call the police department. There seems to be a decrese in calls from March to June, this could be because of the main quarantine due to Covid. Afterwards, there appears to be an increase in calls, that steadily lowers from October.

```
#  Make a bar graph of the frequency of calls by date. Comment on the trend you obs
erve in the white space above.
ggplot(data = b,
       mapping = aes(x = Date)) +
   geom_bar() +
  labs(title = "Frequency of calls by Date")
```



Top 10 reasons for Calls

The top 5 reasons for calls are Public Service, Motor Vehicle, Quality of Life, Other & Property. It seems to be that Public Service are a lot of miscellaneous events and disturbances in public spaces. Motor vehicle being the number 2 makes sense because there are a lot of cars in the US and they can be very dangerous. Quality of Life varies from noise disturbances to intoxications. Other groups in multiple reasons that cannot be placed in different categories, making it a “wildcard” category. Property pertains to stolen or damaged items, vehicles, etc.

```
#  Finally, using dplyr, create a new data set that counts the number of calls of e
ach type.  (hint:  use group_by and use call_type_group).

#  Use dplyr to print the top 10 types of calls (use call_type_group). What are the
top 5 types?  Comment on them in the white space above.


b %>%
  group_by(call_type_group) %>%
  summarize(numCallType = n()) %>%
  arrange(desc(numCallType))%>%
  head(10)
```

```
## # A tibble: 10 x 2
##    call_type_group numCallType
##    <chr>                 <int>
##  1 Public Service         8840
##  2 Motor Vehicle          3709
##  3 Quality of Life        3279
##  4 Other                  1856
##  5 Property               1673
##  6 Assist                 1294
##  7 Domestic                756
##  8 Court Order             753
##  9 Animal                  489
## 10 Assault                 330
```