

Code

Set up

Sneak peak

```
dim(data)
```

```
## [1] 1445230      19
```

```
names(data)
```

```
## [1] "CAD CDW ID"          "CAD Event Number"
## [3] "General Offense Number" "Event Clearance Code"
## [5] "Event Clearance Description" "Event Clearance SubGroup"
## [7] "Event Clearance Group"    "Event Clearance Date"
## [9] "Hundred Block Location"   "District/Sector"
## [11] "Zone/Beat"              "Census Tract"
## [13] "Longitude"              "Latitude"
## [15] "Incident Location"       "Initial Type Description"
## [17] "Initial Type Subgroup"   "Initial Type Group"
## [19] "At Scene Time"
```

Rename columns

```
names(data)[1] <- "CAD_CDW_ID"
names(data)[2] <- "CAD_EVENT_NUMBER"
names(data)[3] <- "GENERAL_OFFENSE_NUMBER"
names(data)[4] <- "EVENT_CLEARANCE_CODE"
names(data)[5] <- "EVENT_CLEARANCE_DESCRIPTION"
names(data)[6] <- "EVENT_CLEARANCE_SUBGROUP"
names(data)[7] <- "EVENT_CLEARANCE_GROUP"
names(data)[8] <- "EVENT_CLEARANCE_DATE"
names(data)[9] <- "HUNDRED_BLOCK_LOCATION"
names(data)[10] <- "DISTRICT"
names(data)[11] <- "ZONE"
names(data)[12] <- "CENSUS_TRACT"
names(data)[13] <- "LONGITUDE"
names(data)[14] <- "LATITUDE"
names(data)[15] <- "INCIDENT_LOCATION"
names(data)[16] <- "INITIAL_TYPE_DESCRIPTION"
names(data)[17] <- "INITIAL_TYPE_SUBGROUP"
names(data)[18] <- "INITIAL_TYPE_GROUP"
names(data)[19] <- "AT_SCENE_TIME"
```

Reduce data

```
data_v1 <- select(data, c(1,5,6,7,8,10,13,14))
data_v1 <- drop_na(data_v1)
```

```
unique(data_v1$EVENT_CLEARANCE_GROUP)
```

```
## [1] "DISTURBANCES"
## [2] "OTHER PROPERTY"
## [3] "NUISANCE, MISCHIEF"
## [4] "TRAFFIC RELATED CALLS"
## [5] "SUSPICIOUS CIRCUMSTANCES"
## [6] "MENTAL HEALTH"
## [7] "LIQUOR VIOLATIONS"
## [8] "TRESPASS"
## [9] "ASSAULTS"
## [10] "NARCOTICS COMPLAINTS"
## [11] "ACCIDENT INVESTIGATION"
## [12] "SHOPLIFTING"
## [13] "PROWLER"
## [14] "BURGLARY"
## [15] "HAZARDS"
## [16] "PROPERTY DAMAGE"
## [17] "CAR PROWL"
## [18] "BIKE"
## [19] "PROSTITUTION"
## [20] "MISCELLANEOUS MISDEMEANORS"
## [21] "ANIMAL COMPLAINTS"
## [22] "THREATS, HARASSMENT"
## [23] "PERSON DOWN/INJURY"
## [24] "AUTO THEFTS"
## [25] "PERSONS - LOST, FOUND, MISSING"
## [26] "FALSE ALARMS"
## [27] "ROBBERY"
## [28] "ARREST"
## [29] "FALSE ALACAD"
## [30] "PROPERTY - MISSING, FOUND"
## [31] "RECKLESS BURNING"
## [32] "BEHAVIORAL HEALTH"
## [33] "WEAPONS CALLS"
## [34] "LEWD CONDUCT"
## [35] "FRAUD CALLS"
## [36] "HARBOR CALLS"
## [37] "MOTOR VEHICLE COLLISION INVESTIGATION"
## [38] "DRIVE BY (NO INJURY)"
## [39] "OTHER VICE"
## [40] "FAILURE TO REGISTER (SEX OFFENDER)"
## [41] "HOMICIDE"
## [42] "VICE CALLS"
## [43] "PUBLIC GATHERINGS"
```

```
unique(data_v1$DISTRICT)
```

```
## [1] "M" "Q" "D" "R" "U" "K" "N" "S" "F" "J"  
## [11] "L" "W" "E" "C" "G" "B" "O" "99" "NULL" "H"
```

```
#Filter out NULL in districts and 99  
data_v1 <- data_v1 %>%  
  filter(DISTRICT != "99") %>%  
  filter(DISTRICT != "NULL")  
data_v1$DISTRICT <- as.factor(data_v1$DISTRICT)
```

REFORMAT DATE AND TIME, ADDING SEPARATE VALUES FOR EACH

```
data_v1$DATE <- format(as.POSIXct(strptime(data_v1$EVENT_CLEARANCE_DATE, "%m/%d/%Y %I:%M:%S %p")), format = "%m/%d/%Y")  
data_v1$MONTH <- format(as.POSIXct(strptime(data_v1$EVENT_CLEARANCE_DATE, "%m/%d/%Y %I:%M:%S %p")), format = "%m")  
data_v1$YEAR <- format(as.POSIXct(strptime(data_v1$EVENT_CLEARANCE_DATE, "%m/%d/%Y %I:%M:%S %p")), format = "%Y")  
data_v1$HOUR <- format(as.POSIXct(strptime(data_v1$EVENT_CLEARANCE_DATE, "%m/%d/%Y %I:%M:%S %p")), format = "%H")  
data_v1$WEEKDAY <- format(as.POSIXct(strptime(data_v1$EVENT_CLEARANCE_DATE, "%m/%d/%Y %I:%M:%S %p")), format = "%u")
```

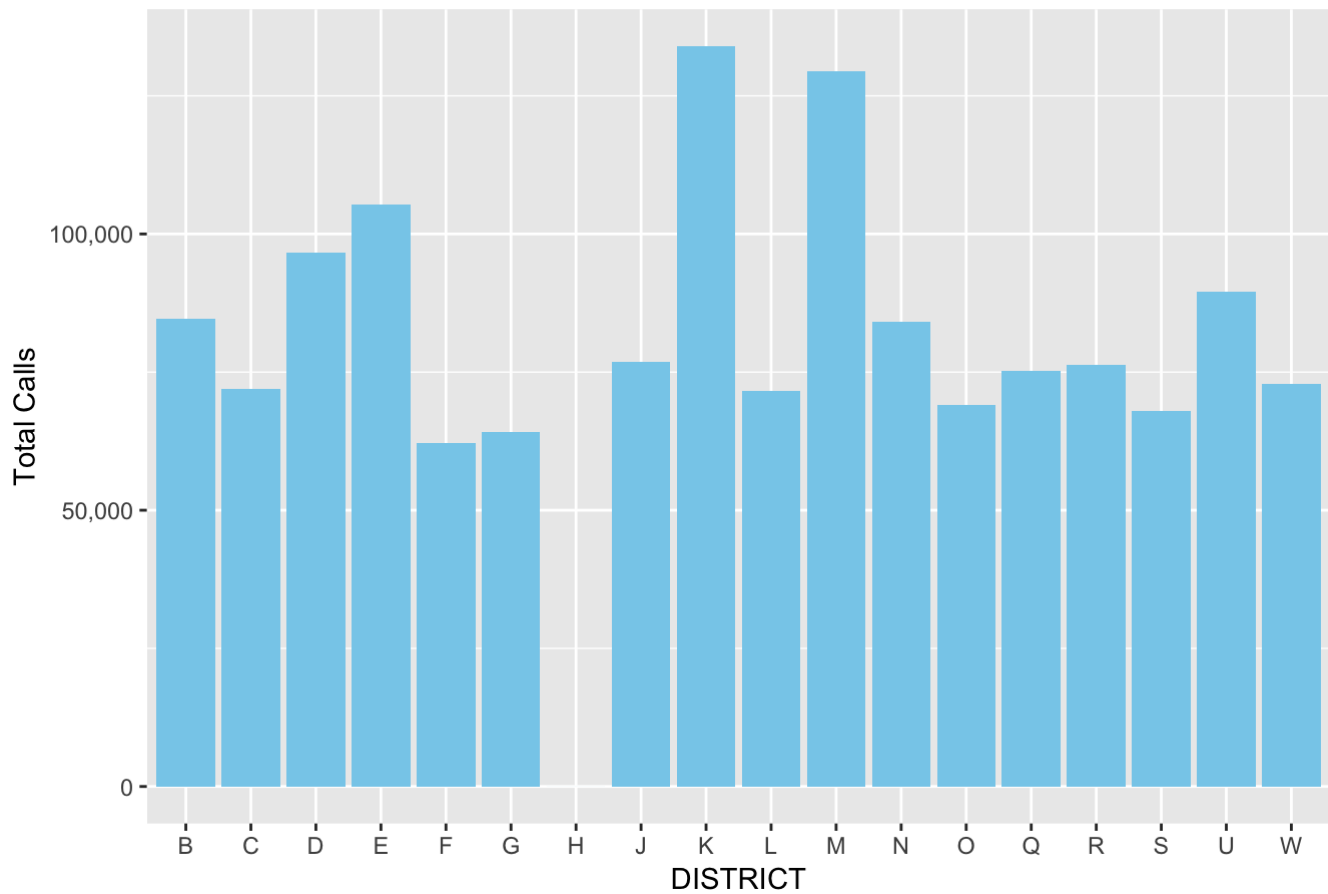
Basic summaries

```
#Numbers  
table(data_v1$HOUR)  
table(data_v1$MONTH)  
table(data_v1$DISTRICT)  
table(data_v1$WEEKDAY)
```

Visualization

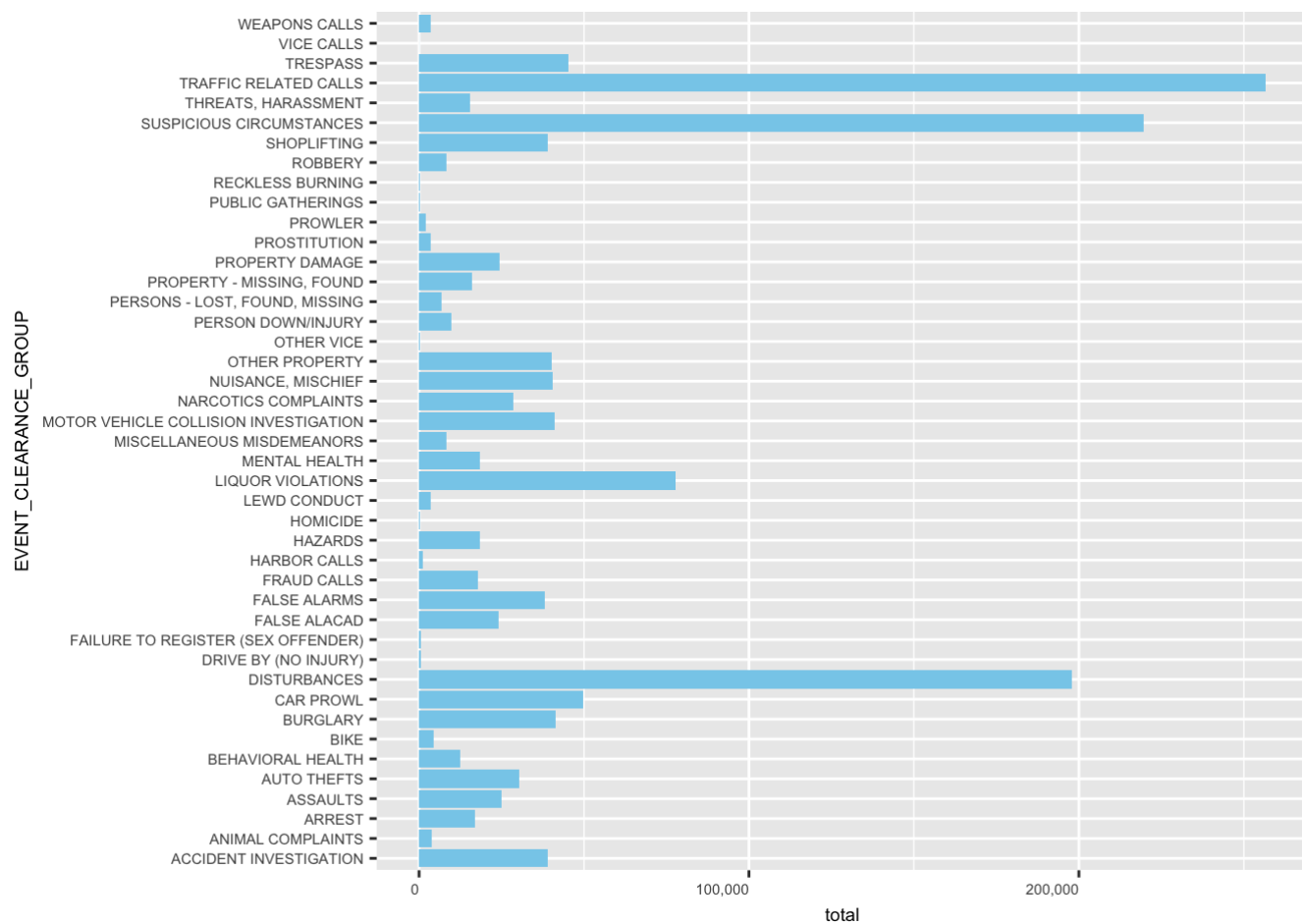
```
#By Districts  
data_v1 %>%  
  group_by(DISTRICT) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_col(mapping = aes(x = DISTRICT, y = total), fill = 'skyblue') + ylab("Total Calls") + scale_y_continuous(labels = scales::comma) + ggtitle("Number of calls by Districts")
```

Number of calls by Districts



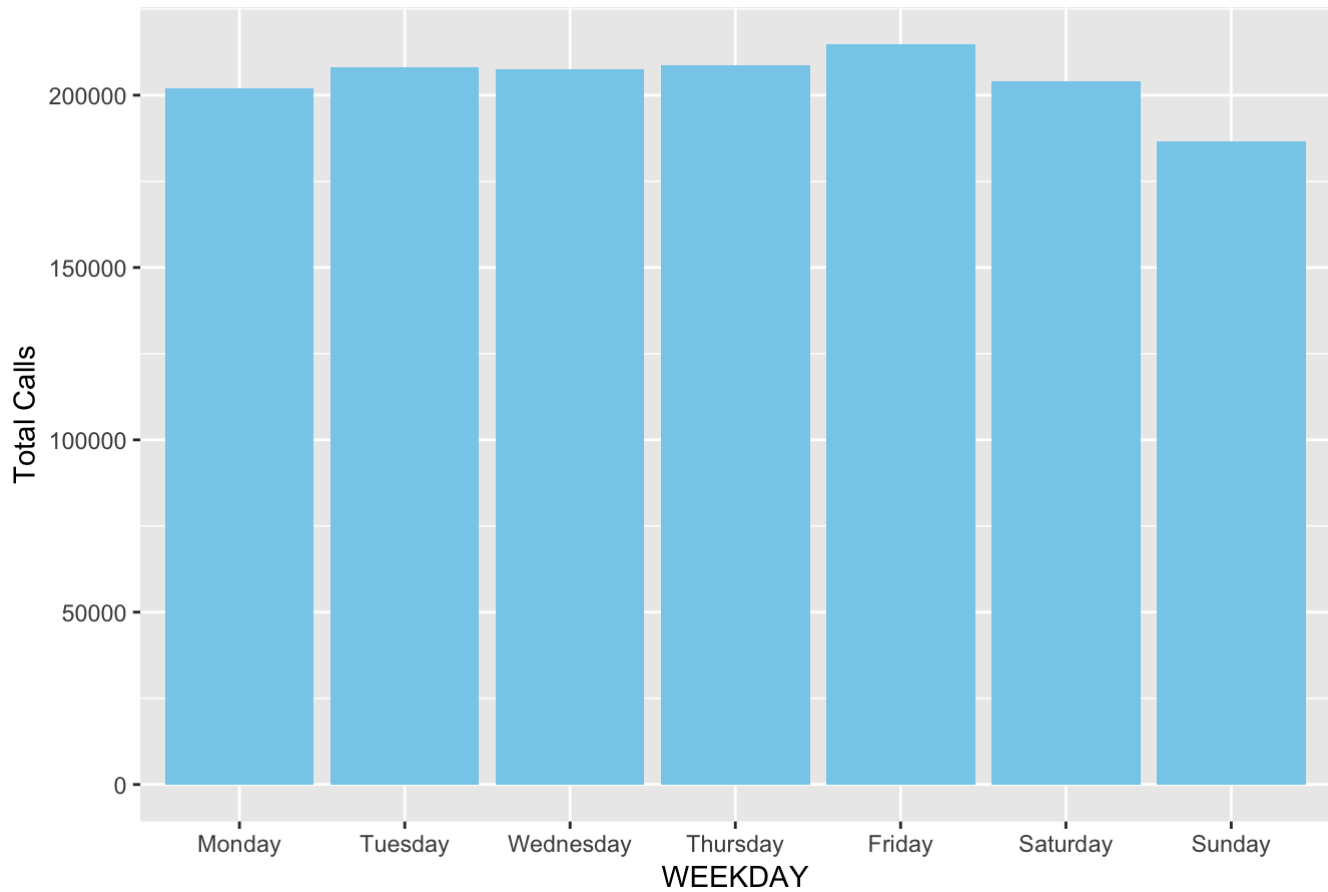
```
#Districts M,K,E,D,U have the most calls  
#District H has very low call
```

```
data_v1 %>%  
  group_by(EVENT_CLEARANCE_GROUP) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_col(mapping = aes(y = total, x=EVENT_CLEARANCE_GROUP), fill = 'skyblue') + coord_flip() + scale_y_continuous(labels = scales::comma) + theme(text = element_text(size=7), axis.text.x = element_text(hjust=1))
```



```
#By weekday
data_v1 %>%
  group_by(WEEKDAY) %>%
  summarise(total = n()) %>%
  ggplot() + geom_col(mapping = aes(x = WEEKDAY, y = total), fill = 'skyblue') + scale_x_
_discrete(labels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday")) + ylab("Total Calls") + ggtitle("Number of calls by Weekday")
```

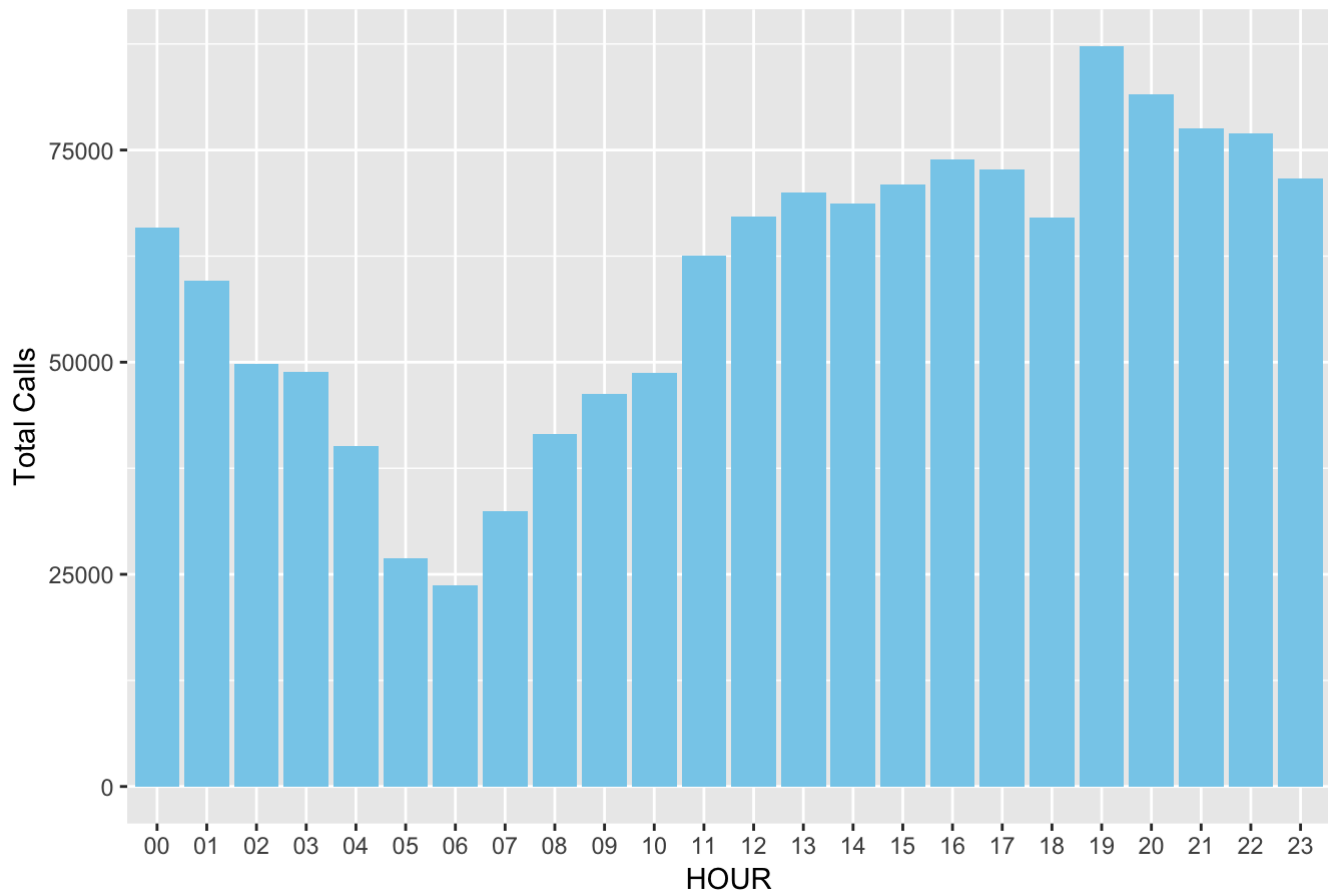
Number of calls by Weekday



#Friday has the most calls

```
#By Time
data_v1 %>%
  group_by(HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_col(mapping = aes(x = HOUR, y = total), fill = 'skyblue') + ylab("Total Calls") + ggtitle("Number of calls by Hours")
```

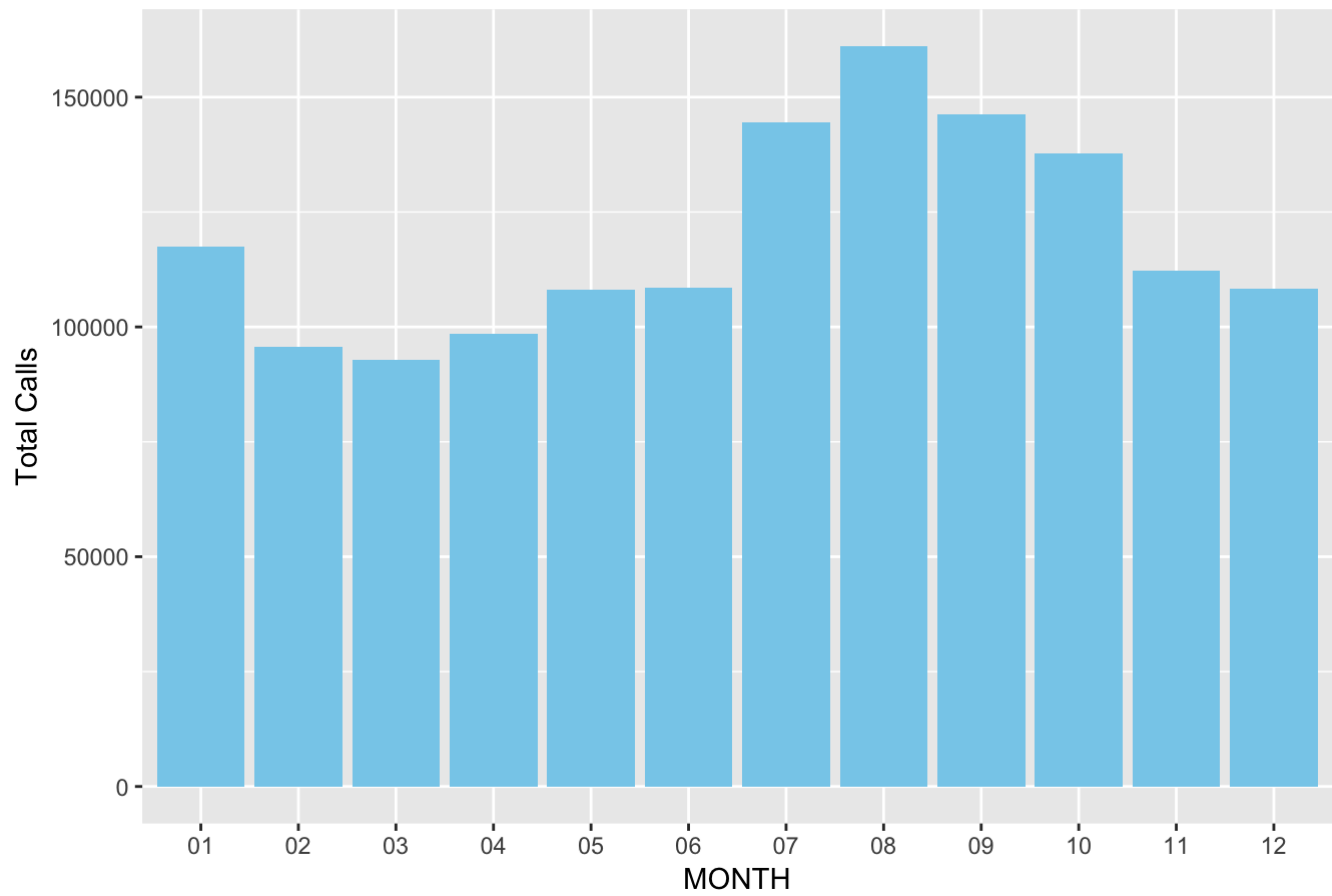
Number of calls by Hours



#7PM is peak

```
#By month
data_v1 %>%
  group_by(MONTH) %>%
  summarise(total = n()) %>%
  ggplot() + geom_col(mapping = aes(x = MONTH, y = total), fill = 'skyblue') + ylab("Total Calls") + ggtitle("Number of calls by Months")
```

Number of calls by Months



```
## Summary of each incident  
table(data_v1$EVENT_CLEARANCE_GROUP)
```



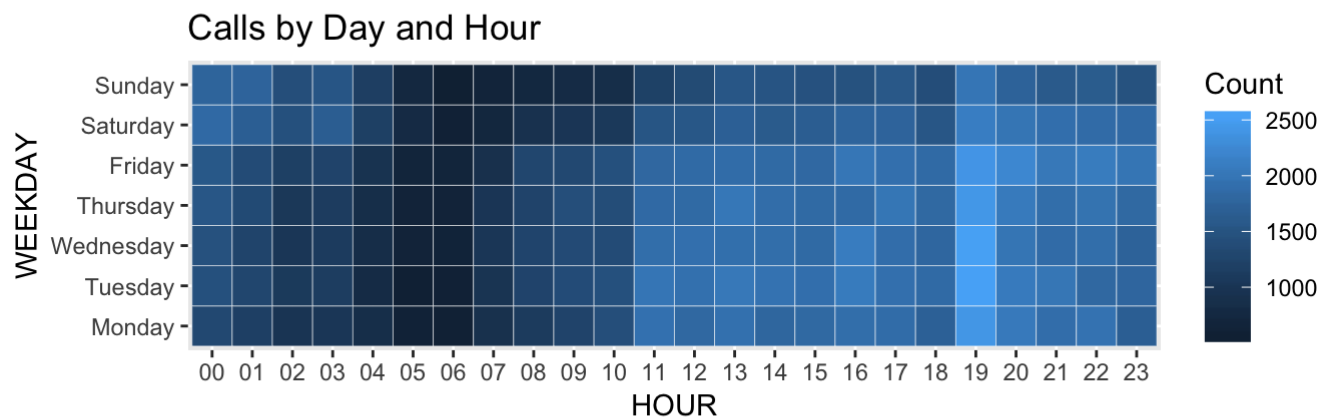
```
##
##          ACCIDENT INVESTIGATION
##                      39005
##          ANIMAL COMPLAINTS
##                      3933
##          ARREST
##                      16938
##          ASSAULTS
##                      25015
##          AUTO THEFTS
##                      30466
##          BEHAVIORAL HEALTH
##                      12448
##          BIKE
##                      4364
##          BURGLARY
##                      41467
##          CAR PROWL
##                      49748
##          DISTURBANCES
##                      198015
##          DRIVE BY (NO INJURY)
##                      632
##  FAILURE TO REGISTER (SEX OFFENDER)
##                      517
##          FALSE ALACAD
##                      23959
##          FALSE ALARMS
##                      38191
##          FRAUD CALLS
##                      17709
##          HARBOR CALLS
##                      1162
##          HAZARDS
##                      18430
##          HOMICIDE
##                      225
##          LEWD CONDUCT
##                      3422
##          LIQUOR VIOLATIONS
##                      77635
##          MENTAL HEALTH
##                      18313
##          MISCELLANEOUS MISDEMEANORS
##                      8332
##  MOTOR VEHICLE COLLISION INVESTIGATION
##                      40968
##          NARCOTICS COMPLAINTS
##                      28528
##          NUISANCE, MISCHIEF
##                      40496
##          OTHER PROPERTY
##                      40183
```

```
##                OTHER VICE
##                394
##                PERSON DOWN/INJURY
##                9850
##                PERSONS - LOST, FOUND, MISSING
##                6789
##                PROPERTY - MISSING, FOUND
##                16141
##                PROPERTY DAMAGE
##                24309
##                PROSTITUTION
##                3491
##                PROWLER
##                2076
##                PUBLIC GATHERINGS
##                374
##                RECKLESS BURNING
##                291
##                ROBBERY
##                8344
##                SHOPLIFTING
##                38867
##                SUSPICIOUS CIRCUMSTANCES
##                219556
##                THREATS, HARASSMENT
##                15375
##                TRAFFIC RELATED CALLS
##                256692
##                TRESPASS
##                45392
##                VICE CALLS
##                29
##                WEAPONS CALLS
##                3567
```

```
# Get the count of the calls by Day and Hour
day_hour <- data_v1[data_v1$YEAR == "2016", c("WEEKDAY", "HOUR")] %>%
  group_by(WEEKDAY, HOUR) %>%
  summarise(Count = n())
day_hour <- as.data.frame(day_hour)

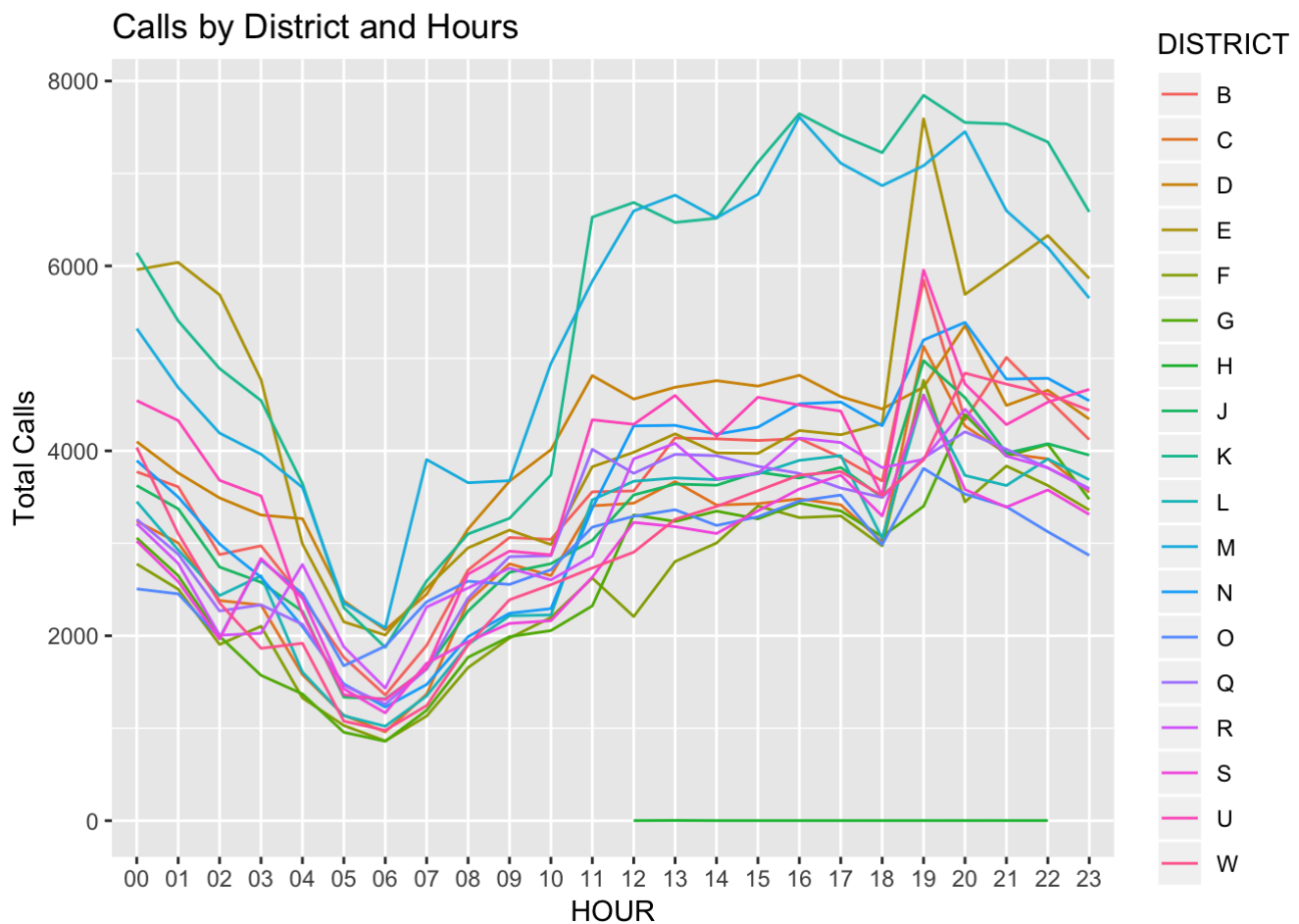
# Change the type of the variables
day_hour$WEEKDAY<- as.factor(day_hour$WEEKDAY)
day_hour$HOUR <- as.factor(day_hour$HOUR)

# Building heatmap using ggplot2
ggplot(day_hour, aes(y=WEEKDAY, x=HOUR, fill = Count)) + geom_tile(color = "white", size
= 0.1) + coord_equal() + labs(title="Calls by Day and Hour") + scale_y_discrete(labels
= c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))
```



- Do traffic => look into certain types of incidents.
- How to make use of it => Where to put the cops.
- How things happen over time. Example: heatmap year + types.

```
data_v1 %>%
  group_by(DISTRICT, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = DISTRICT, color = DISTRICT)) + labs(title="Calls by District and Hours") + ylab("Total Calls")
```



Divide data into different subsets in EVENT_DESCRIPTION_GROUP to focus on

```
# Public Gatherings
public_gatherings <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "PUBLIC GATHERINGS")
table(public_gatherings$MONTH)
```

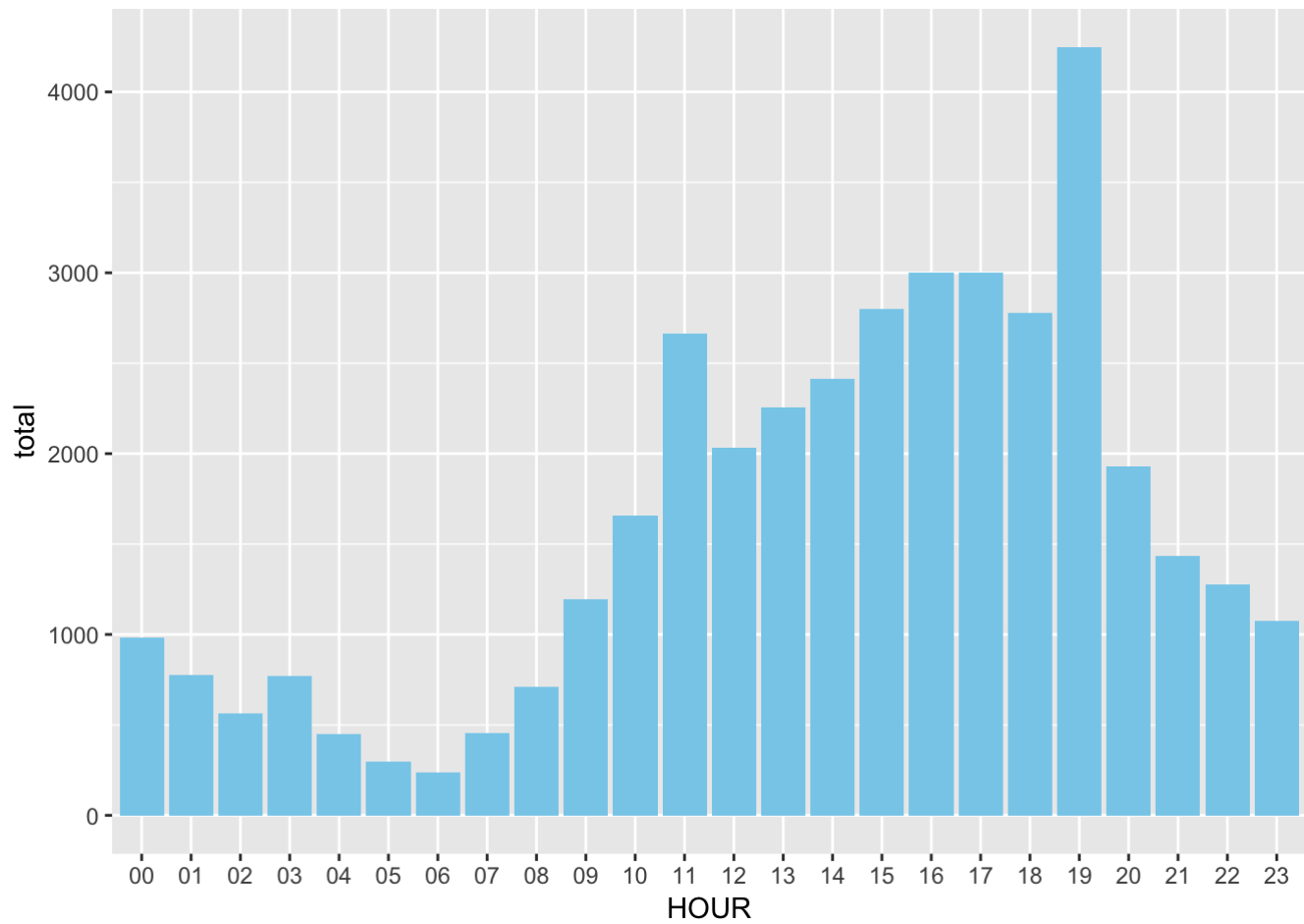
```
##
## 01 02 03 04 05 06 07 08 09 10 11 12
## 22 22 36 28 34 32 34 41 48 30 32 15
```

```
# Accidents
accidents <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "ACCIDENT INVESTIGATION")
table(accidents$HOUR)
```

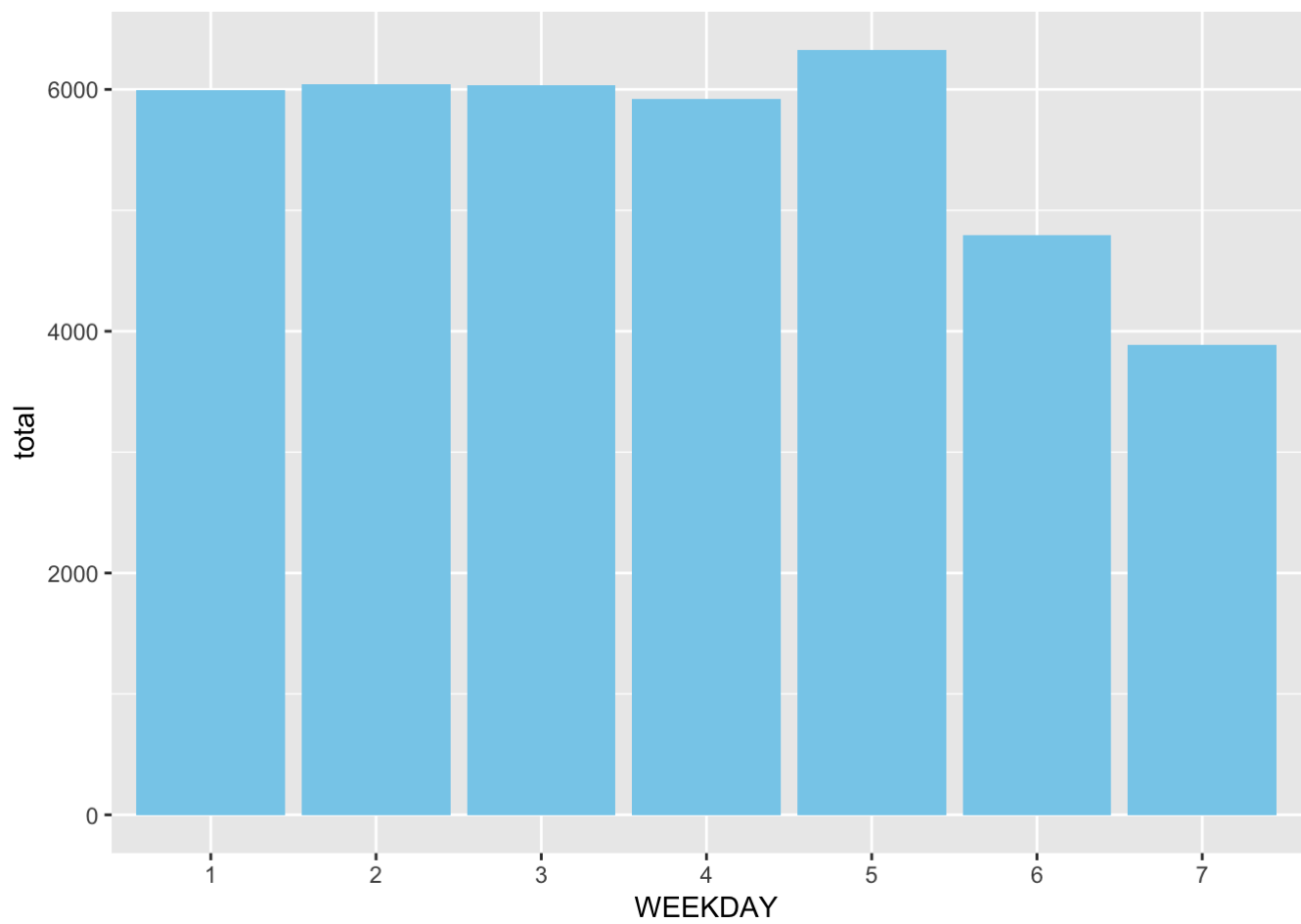
```
##
## 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14
## 984 776 562 769 452 295 238 454 712 1195 1655 2664 2031 2257 2415
## 15 16 17 18 19 20 21 22 23
## 2800 3003 3003 2777 4249 1930 1433 1278 1073
```

=> because of accident => dangerous at 7 for other incidents as well.

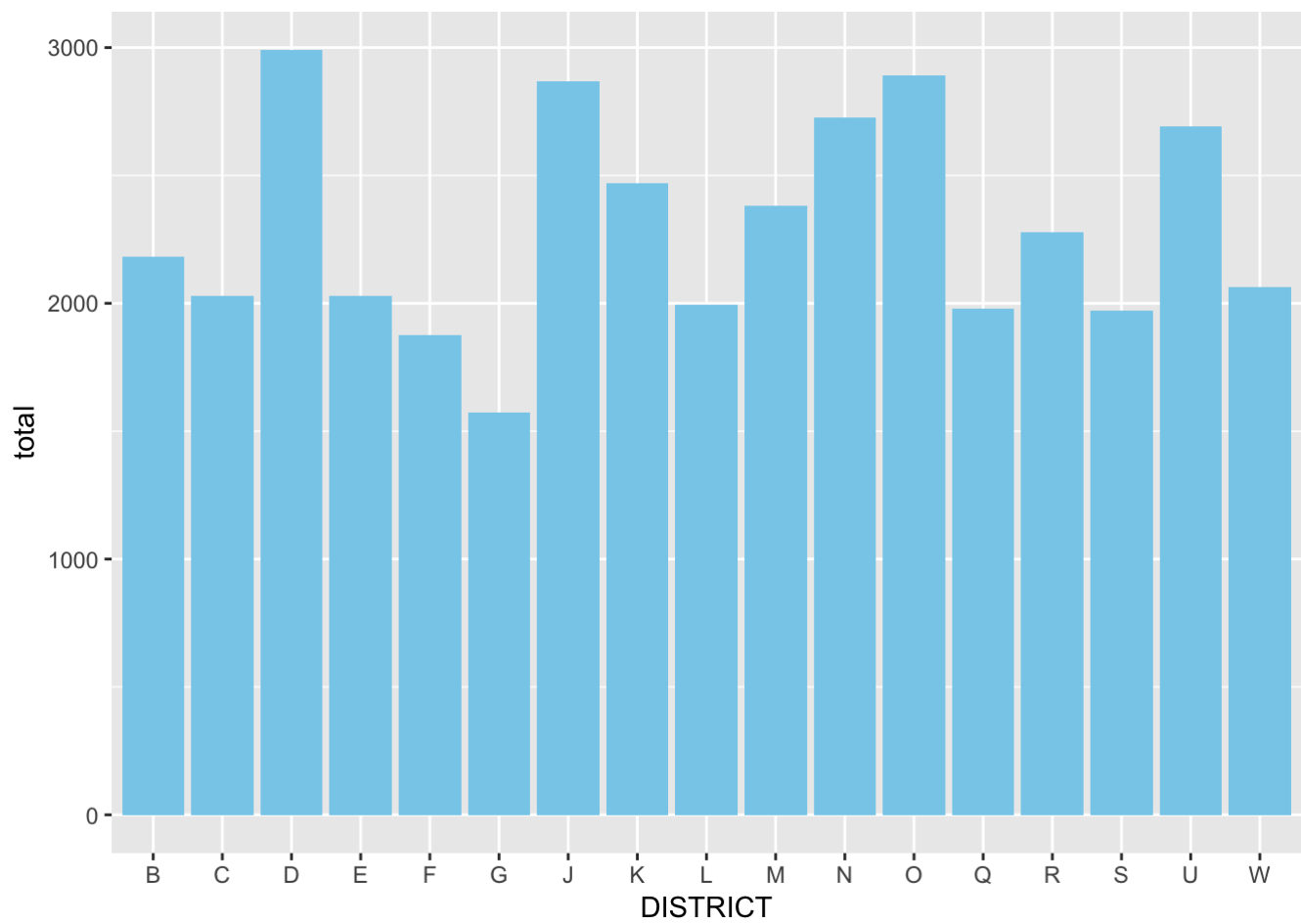
```
accidents %>%  
  group_by(HOUR) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_col(mapping = aes(x = HOUR, y = total), fill = 'skyblue')
```



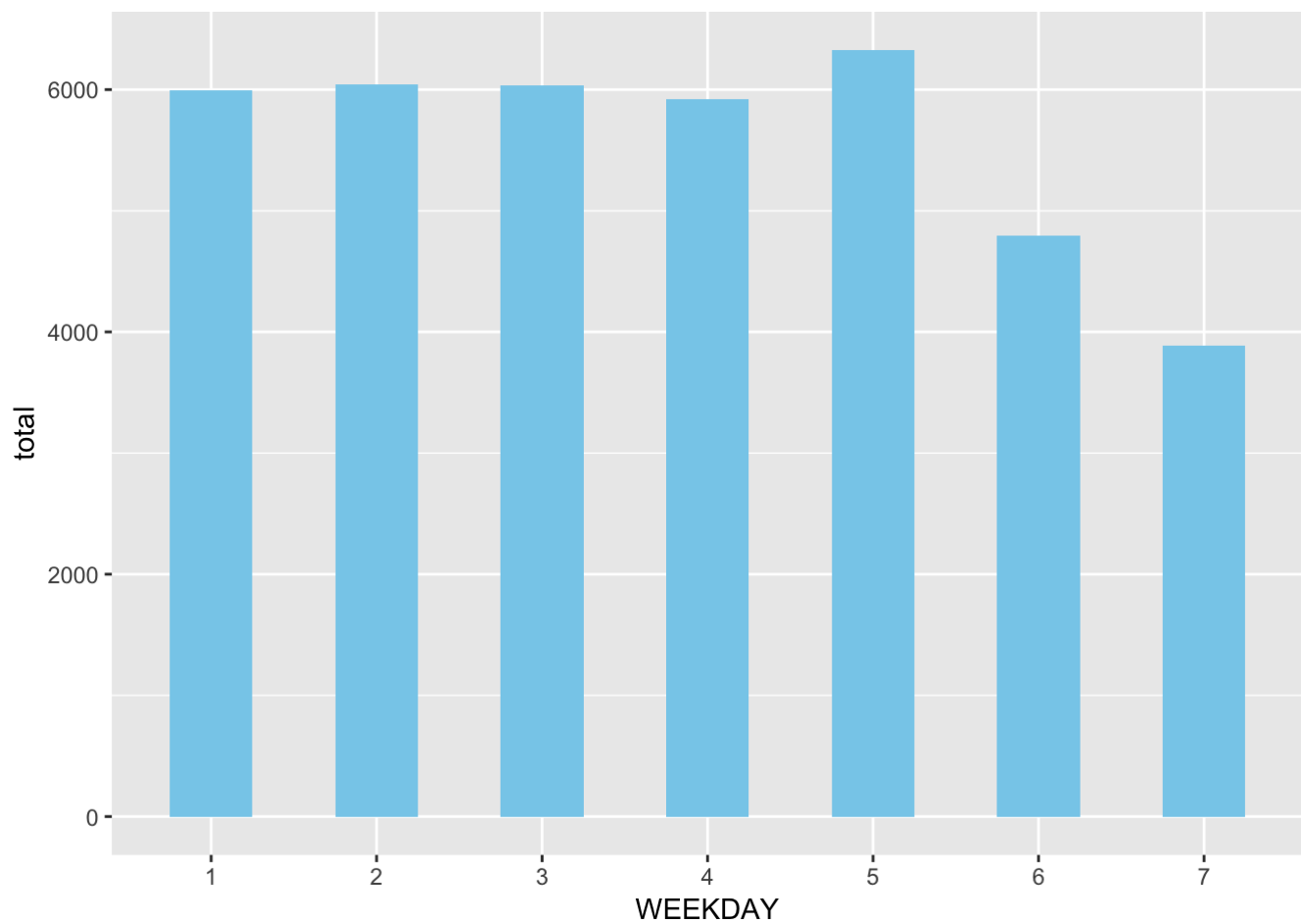
```
accidents %>%  
  group_by(WEEKDAY) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_col(mapping = aes(x = WEEKDAY, y = total), fill = 'skyblue')
```



```
accidents %>%  
  group_by(DISTRICT) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_col(mapping = aes(x = DISTRICT, y = total), fill = 'skyblue')
```



```
accidents %>%
  group_by(WEEKDAY) %>%
  summarise(total = n()) %>%
  ggplot() + geom_col(mapping = aes(x = WEEKDAY, y = total), width = 0.5, fill = 'skyblue')
```



```
# Time stamps analysis
table(accidents$YEAR, accidents$MONTH)
```

```
##
##      01  02  03  04  05  06  07  08  09  10  11  12
## 2010    0    0    0    0    1    2  657 1449 1357 1564 1447  632
## 2011 1376  551  574 1289 1256 1225 1253 1332 1264 1316  846 1228
## 2012 1174 1054 1304 1255 1400 1398 1392 1270 1314 1554 1419 1317
## 2013 1362  966    0    0    0    0    0    6    0    0    8    0
## 2014    0    0    0    0    0    0    13   33   23   25   24   40
## 2015   35    0    0    0    0    0    0    0    0    0    0    0
```

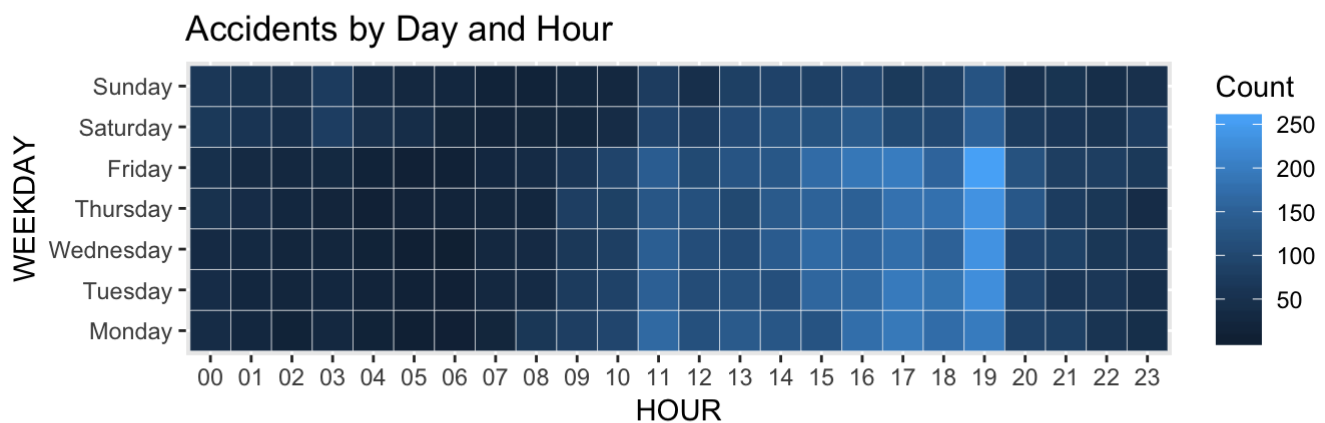
```
# Missing a lot of data in 2010, 2013, 2014, 2015
```



```
## Get the count of the calls by Day and Hour
accidents_plot <- accidents[accidents$YEAR == "2011", c("WEEKDAY", "HOUR")] %>%
  group_by(WEEKDAY, HOUR) %>%
  summarise(Count = n())
accident_plot <- as.data.frame(accidents_plot)

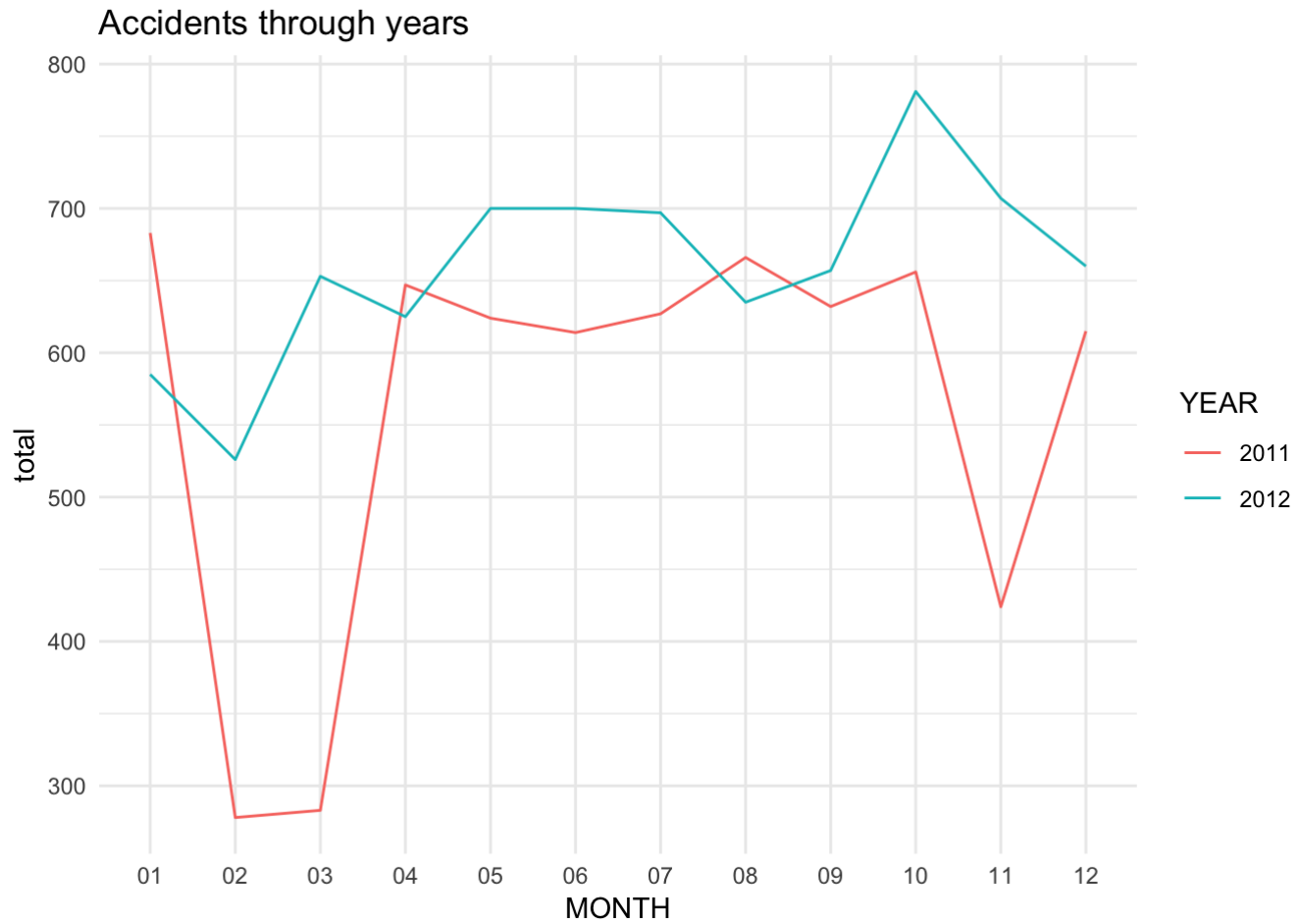
## Change the type of the variables
accident_plot$WEEKDAY<- as.factor(accident_plot$WEEKDAY)
accident_plot$HOUR <- as.factor(accident_plot$HOUR)

## Building heatmap using ggplot2
ggplot(accident_plot, aes(y=WEEKDAY, x=HOUR, fill = Count)) + geom_tile(color = "white",
  size = 0.1) + coord_equal() + labs(title="Accidents by Day and Hour") + scale_y_discret
e(labels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"
))
```



```
# Time series plot
accidents %>%
  filter(YEAR == c("2011", "2012")) %>%
  group_by(YEAR, MONTH) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = MONTH, y = total, group = YEAR, color = YEAR))
+ labs(title="Accidents through years") + theme_minimal()
```

```
## Warning in YEAR == c("2011", "2012"): longer object length is not a
## multiple of shorter object length
```

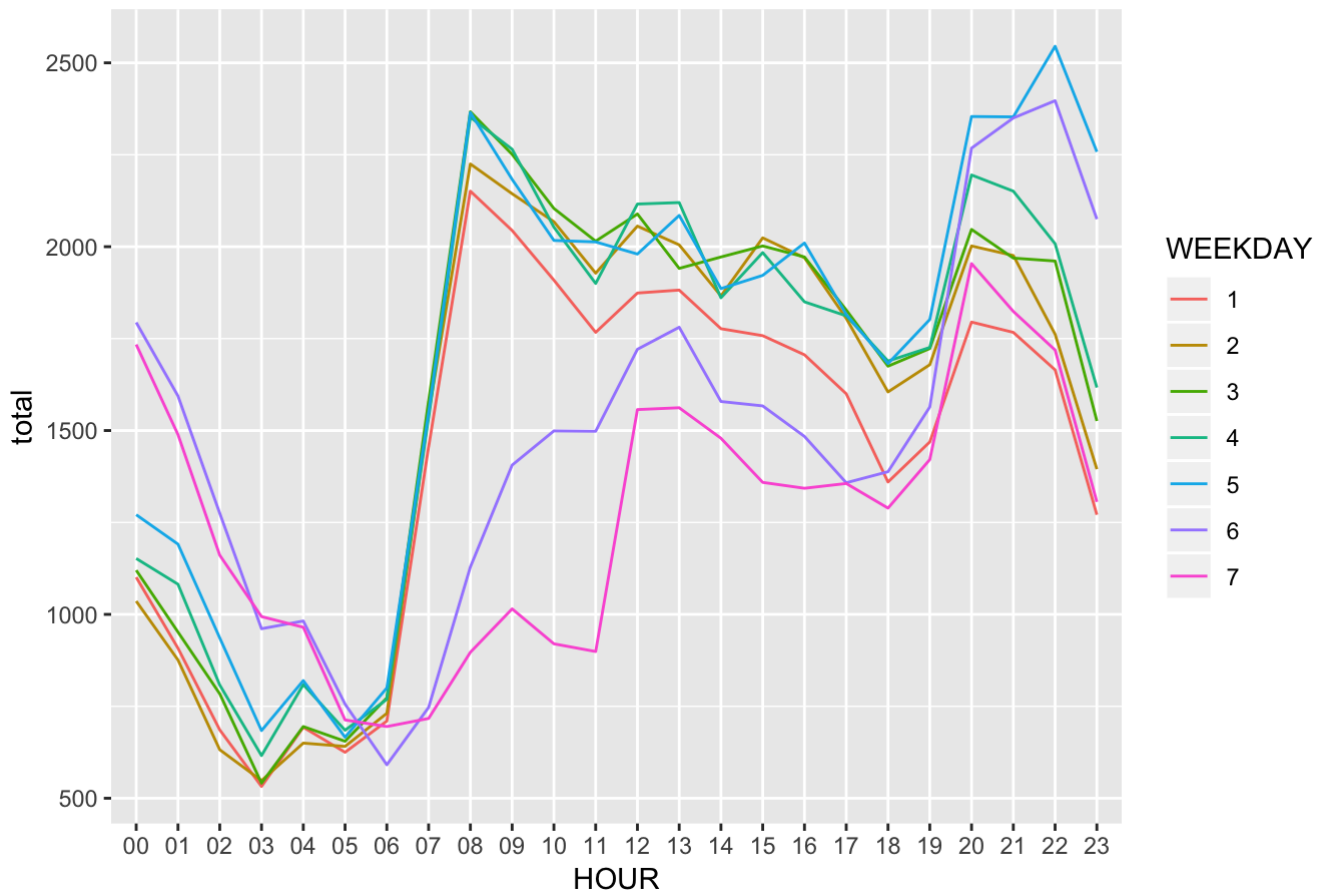


```
# Traffic
traffic <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "TRAFFIC RELATED CALLS")
table(traffic$WEEKDAY, traffic$HOUR)
```

```
##
##      00    01    02    03    04    05    06    07    08    09    10    11    12    13
##  1 1101   908   686   532   693   625   710 1455 2151 2044 1909 1767 1874 1882
##  2 1036   876   632   548   650   641   731 1547 2225 2144 2068 1928 2056 2005
##  3 1120   952   784   541   695   655   773 1580 2367 2251 2104 2015 2089 1941
##  4 1152 1082   808   616   809   685   769 1534 2352 2265 2053 1900 2116 2120
##  5 1271 1191   936   684   820   666   800 1549 2365 2183 2017 2013 1980 2085
##  6 1794 1594 1275   961   982   756   591   747 1129 1406 1499 1498 1721 1781
##  7 1734 1489 1161   994   965   713   695   717   897 1015   920   899 1557 1562
##
##      14    15    16    17    18    19    20    21    22    23
##  1 1777 1758 1706 1600 1360 1469 1795 1767 1665 1271
##  2 1866 2024 1970 1805 1605 1679 2002 1976 1762 1395
##  3 1972 2002 1972 1828 1675 1723 2047 1969 1961 1526
##  4 1861 1984 1850 1812 1689 1726 2195 2151 2008 1617
##  5 1886 1922 2010 1816 1682 1802 2354 2353 2545 2258
##  6 1579 1567 1484 1358 1388 1564 2268 2350 2397 2075
##  7 1479 1359 1343 1356 1289 1421 1954 1824 1719 1306
```

```
traffic %>%
  group_by(WEEKDAY, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = WEEKDAY, color = WEEKDAY)) + labs(title="Traffic Calls")
```

Traffic Calls



```
unique(traffic$EVENT_CLEARANCE_DESCRIPTION)
```

```
## [1] "TRAFFIC (MOVING) VIOLATION"  
## [2] "PARKING VIOLATION (EXCEPT ABANDONED VEHICLES)"  
## [3] "PEDESTRIAN VIOLATION"  
## [4] "DRIVING WHILE UNDER INFLUENCE (DUI)"  
## [5] "MOTORIST ASSIST"  
## [6] "BLOCKING VEHICLE"  
## [7] "ABANDONED VEHICLE"  
## [8] "PURSUIT"  
## [9] "TRAFFIC CONTROL (SPECIAL EVENTS)"  
## [10] "TRAFFIC - COMMUNITY TRAFFIC COMPLAINT (CTC)"  
## [11] "TRAFFIC - BICYCLE VIOLATION"  
## [12] "TRAFFIC - SCHOOL ZONE ENFORCEMENT"
```

```
traffic %>%  
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%  
  summarise(total = n())
```

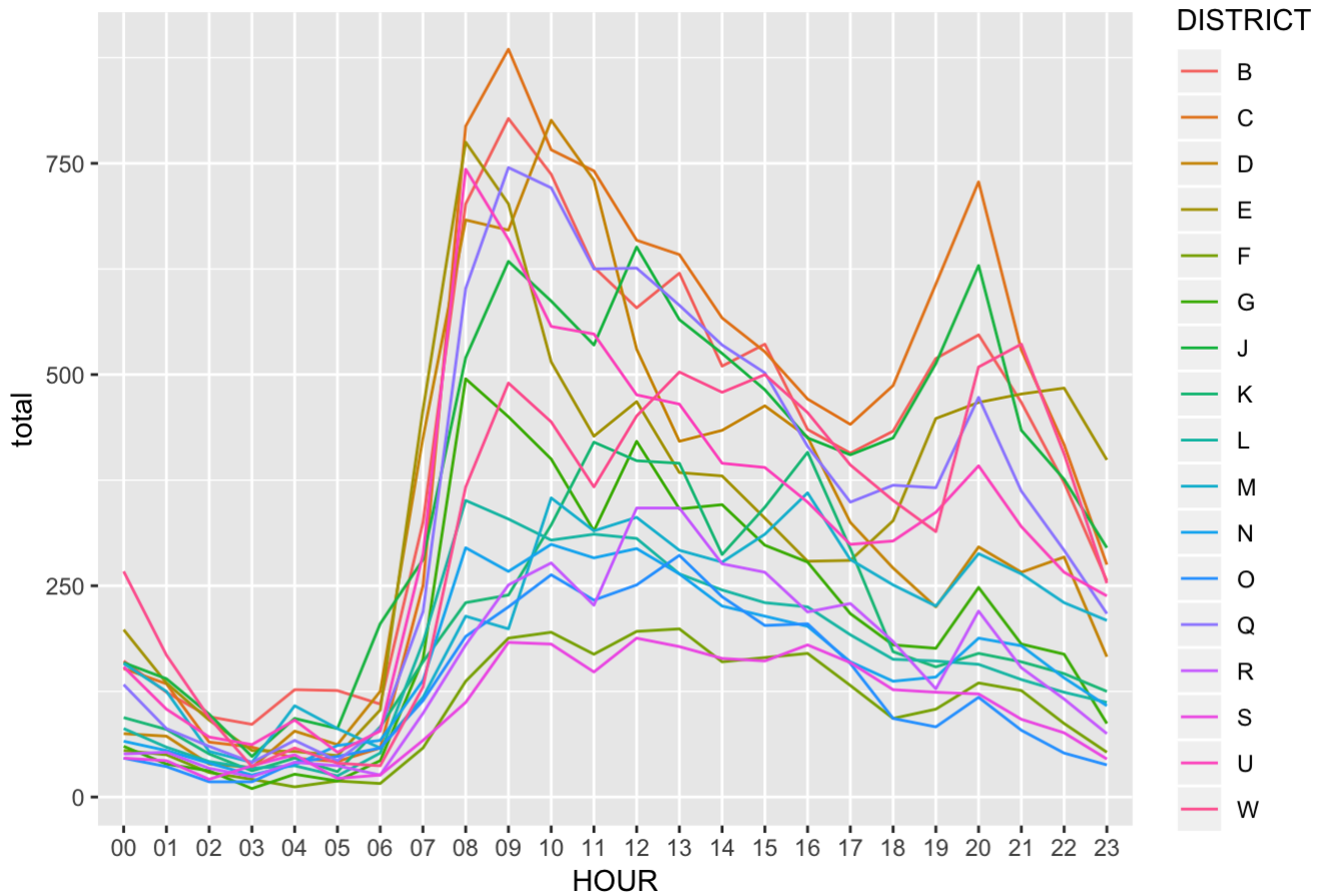
```
## # A tibble: 12 x 2  
##   EVENT_CLEARANCE_DESCRIPTION      total  
##   <chr>                        <int>  
## 1 ABANDONED VEHICLE             6512  
## 2 BLOCKING VEHICLE             11625  
## 3 DRIVING WHILE UNDER INFLUENCE (DUI) 14879  
## 4 MOTORIST ASSIST                7047  
## 5 PARKING VIOLATION (EXCEPT ABANDONED VEHICLES) 104190  
## 6 PEDESTRIAN VIOLATION          5882  
## 7 PURSUIT                       132  
## 8 TRAFFIC - BICYCLE VIOLATION        64  
## 9 TRAFFIC - COMMUNITY TRAFFIC COMPLAINT (CTC)    101  
## 10 TRAFFIC - SCHOOL ZONE ENFORCEMENT           31  
## 11 TRAFFIC (MOVING) VIOLATION        106057  
## 12 TRAFFIC CONTROL (SPECIAL EVENTS)         172
```

```
traffic_parking <- traffic %>%  
  filter(EVENT_CLEARANCE_DESCRIPTION == "PARKING VIOLATION (EXCEPT ABANDONED VEHICLES)")  
  
nrow(traffic_parking)
```

```
## [1] 104190
```

```
traffic_parking %>%  
  group_by(DISTRICT, HOUR) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = DISTRICT, color = DISTRICT)) + labs(title="Parking Violations")
```

Parking Violations



```
# Suspicious circumstances
suspicious <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "SUSPICIOUS CIRCUMSTANCES")
table(suspicious$WEEKDAY, suspicious$HOUR)
```

```
##
##      00    01    02    03    04    05    06    07    08    09    10    11    12    13
## 1 1509 1355 1090 1090 1248  722  536  656  710  734  799  967 1466 1434
## 2 1652 1574 1226 1131 1250  788  552  645  745  846  869 1041 1495 1462
## 3 1684 1568 1212 1160 1297  763  629  655  812  850  879 1038 1571 1496
## 4 1664 1544 1248 1121 1244  780  608  655  758  864  905  999 1510 1507
## 5 1746 1532 1317 1232 1250  782  595  638  732  846  900  986 1467 1490
## 6 1951 1775 1526 1503 1421  881  586  565  646  675  635  836 1279 1260
## 7 1919 1752 1413 1455 1328  856  568  548  521  627  585  723 1110 1087
##
##      14    15    16    17    18    19    20    21    22    23
## 1 1317 1450 1534 1405 1258 1752 2144 1963 1950 1773
## 2 1409 1475 1499 1474 1336 1823 2205 1979 1984 1881
## 3 1428 1385 1549 1447 1301 1742 2186 1969 1932 1906
## 4 1344 1409 1480 1492 1280 1786 2120 1941 1984 1881
## 5 1357 1427 1525 1448 1273 1723 2254 2059 2077 2061
## 6 1128 1215 1298 1256 1083 1560 1944 1903 1934 1935
## 7 1069 1079 1120 1226 1002 1379 1825 1665 1740 1661
```

```
unique(suspicious$EVENT_CLEARANCE_DESCRIPTION)
```

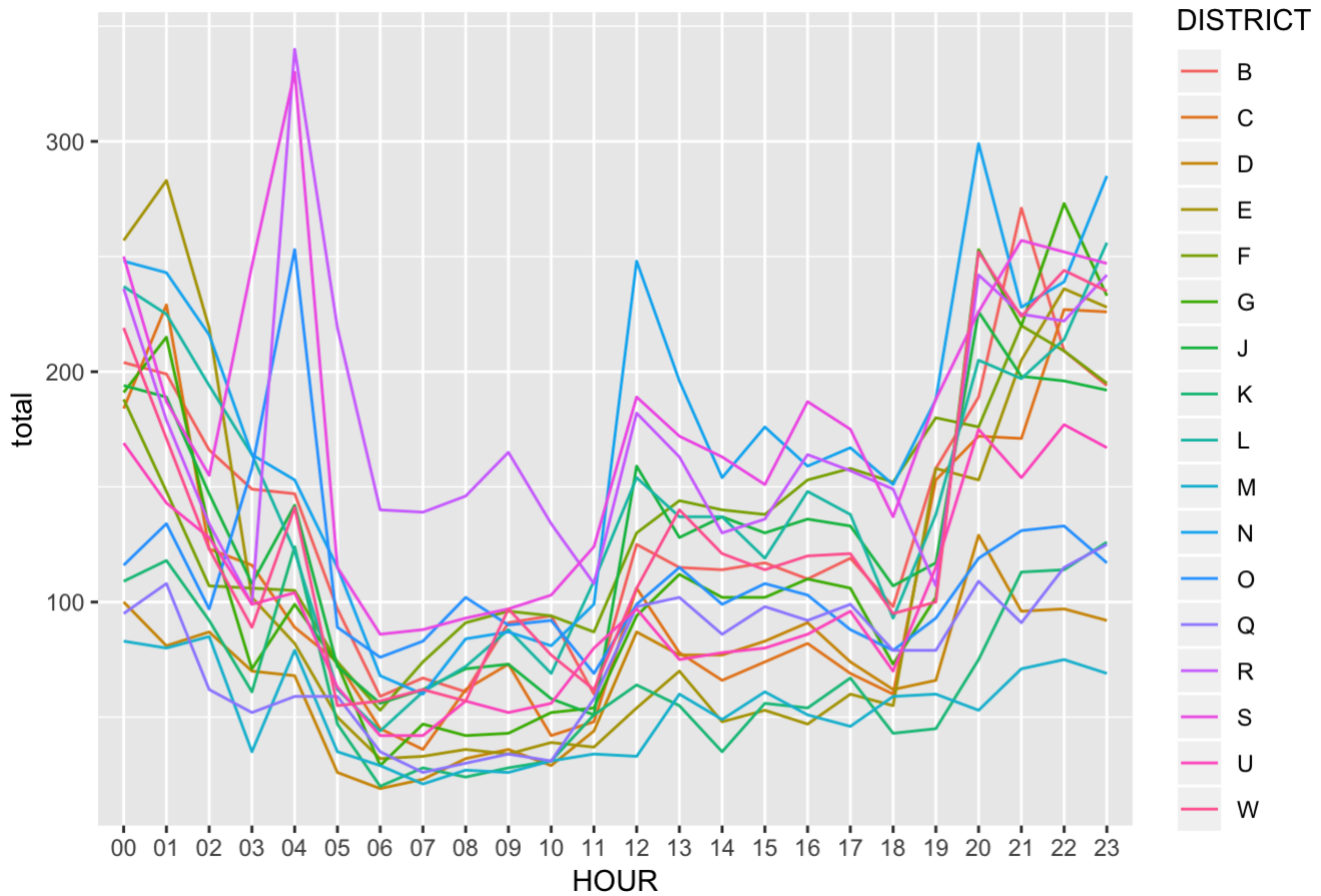
```
## [1] "SUSPICIOUS VEHICLE"  
## [2] "SUSPICIOUS PERSON"  
## [3] "SUSPICIOUS CIRCUMSTANCES - BUILDING (OPEN DOOR, ETC.)"
```

```
suspicious %>%  
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%  
  summarise(total = n())
```

```
## # A tibble: 3 x 2  
##   EVENT_CLEARANCE_DESCRIPTION      total  
##   <chr>                      <int>  
## 1 SUSPICIOUS CIRCUMSTANCES - BUILDING (OPEN DOOR, ETC.)    9130  
## 2 SUSPICIOUS PERSON                                162612  
## 3 SUSPICIOUS VEHICLE                                    47814
```

```
suspicious_vehicle <- suspicious %>%  
  filter(EVENT_CLEARANCE_DESCRIPTION == "SUSPICIOUS VEHICLE")  
  
suspicious_vehicle %>%  
  group_by(DISTRICT, HOUR) %>%  
  summarise(total = n()) %>%  
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = DISTRICT, color = DISTRICT)) + labs(title="Suspicious Vehicles in Districts")
```

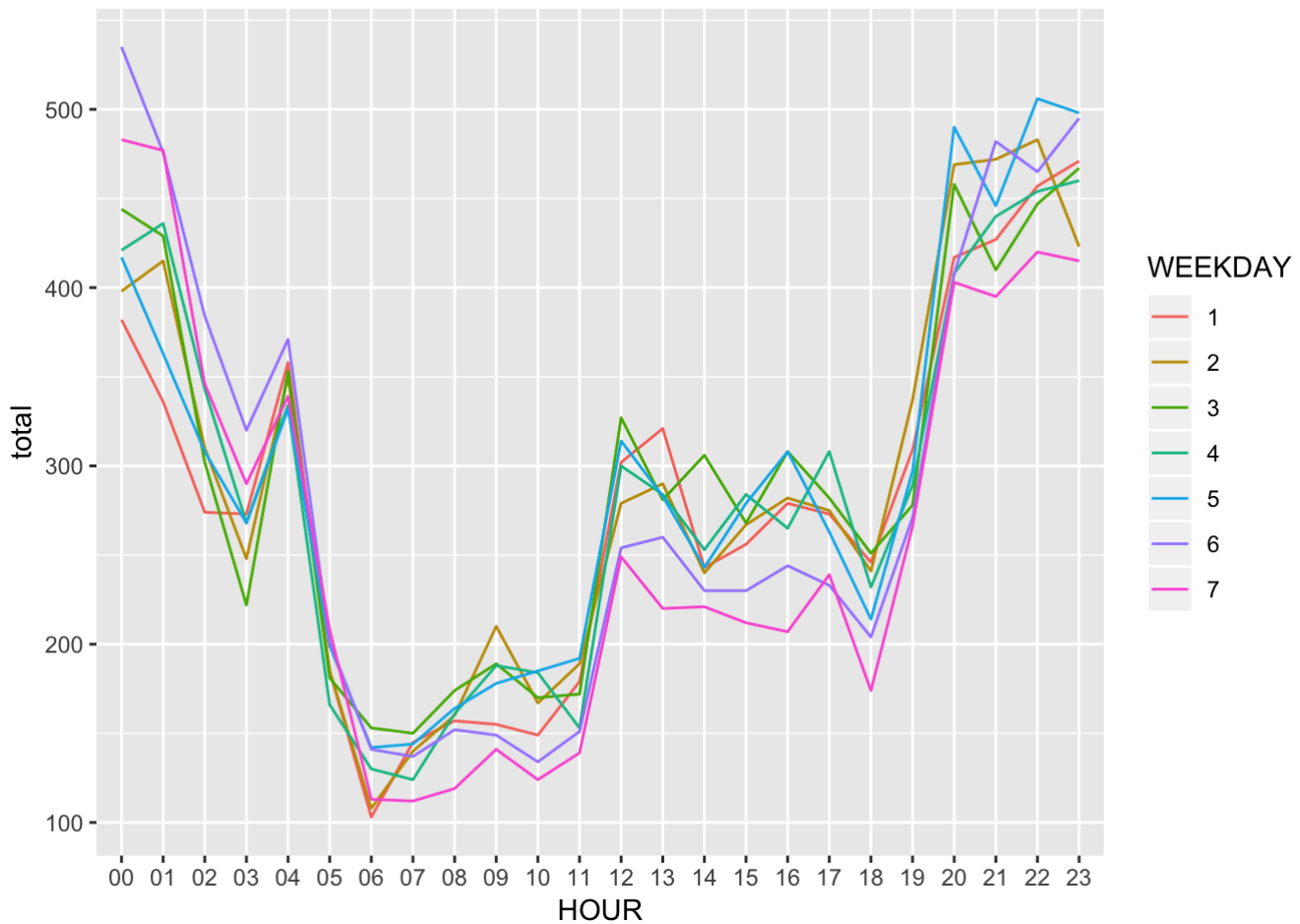
Suspicious Vehicles in Districts



```
suspicious_vehicle %>%
  group_by(DISTRICT) %>%
  summarise(total = n())
```

```
## # A tibble: 17 x 2
##   DISTRICT total
##   <fct>      <int>
## 1 B          3213
## 2 C          2605
## 3 D          1646
## 4 E          2571
## 5 F          3218
## 6 G          2828
## 7 J          3082
## 8 K          1580
## 9 L          3383
## 10 M         1252
## 11 N         4108
## 12 O         2643
## 13 Q         1822
## 14 R         4158
## 15 S         4218
## 16 U         2405
## 17 W         3082
```

```
suspicious_vehicle %>%
  group_by(WEEKDAY, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = WEEKDAY, color = WEEKDAY))
```



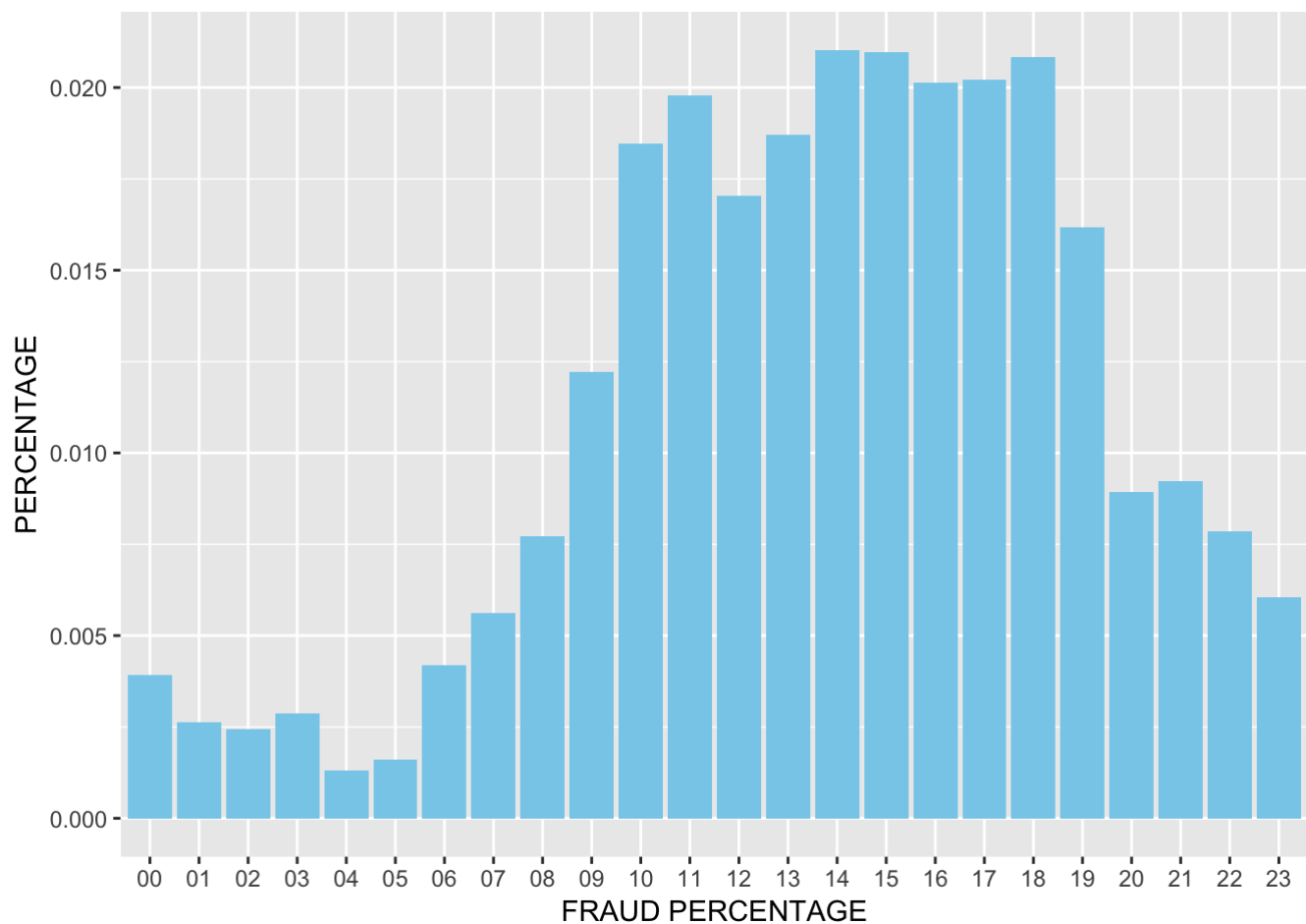
```
# Fraud

fraud <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "FRAUD CALLS")

fraud_percentage <- as.data.frame(table(fraud$HOUR)/table(data_v1$HOUR))

names(fraud_percentage)[1] <- "HOUR"
names(fraud_percentage)[2] <- "PERCENTAGE"

fraud_percentage %>%
  ggplot() + geom_col(mapping = aes(x=HOUR, y=PERCENTAGE), fill = 'skyblue') + xlab("FRAUD PERCENTAGE")
```

```
# Assaults
assaults <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "ASSAULTS")
assaults %>%
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(total = n())
```

```
## # A tibble: 3 x 2
##   EVENT_CLEARANCE_DESCRIPTION total
##   <chr>                  <int>
## 1 ASSAULTS, FIREARM INVOLVED    755
## 2 ASSAULTS, GANG RELATED        99
## 3 ASSAULTS, OTHER             24161
```

```
#Disturbances
disturbances <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "DISTURBANCES")
disturbances %>%
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(total = n())
```

```
## # A tibble: 6 x 2
##   EVENT_CLEARANCE_DESCRIPTION    total
##   <chr>                        <int>
## 1 DISTURBANCE, GANG RELATED      325
## 2 DISTURBANCE, OTHER           140973
## 3 FIGHT DISTURBANCE             10248
## 4 JUVENILE DISTURBANCE           947
## 5 NOISE DISTURBANCE             36999
## 6 NOISE DISTURBANCE, RESIDENTIAL 8523
```

```
#Weapons
weapons <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "WEAPONS CALLS")
weapons %>%
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(total = n())
```

```
## # A tibble: 2 x 2
##   EVENT_CLEARANCE_DESCRIPTION    total
##   <chr>                        <int>
## 1 PERSON WITH A GUN             2141
## 2 PERSON WITH A WEAPON (NOT GUN) 1426
```

```
#Liquor
liquor <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "LIQUOR VIOLATIONS")
liquor %>%
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(total = n())
```

```
## # A tibble: 3 x 2
##   EVENT_CLEARANCE_DESCRIPTION    total
##   <chr>                        <int>
## 1 LIQUOR VIOLATION - ADULT      17591
## 2 LIQUOR VIOLATION - INTOXICATED PERSON 59280
## 3 LIQUOR VIOLATION - MINOR      764
```

```
#Narcotics
narcotics <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "NARCOTICS COMPLAINTS")
narcotics %>%
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(total = n())
```

```
## # A tibble: 6 x 2
##   EVENT_CLEARANCE_DESCRIPTION      total
##   <chr>                        <int>
## 1 MARIJUANA PUBLIC USE (NOT DISPENSARY) 1626
## 2 NARCOTICS ACTIVITY REPORT            2209
## 3 NARCOTICS FOUND, RECOVERED           1211
## 4 NARCOTICS WARRANT SERVICE             107
## 5 NARCOTICS, DRUG TRAFFIC LOITERING     3465
## 6 NARCOTICS, OTHER                     19910
```

```
#Summary
table(assaults$DISTRICT)
```

```
##
##      B      C      D      E      F      G      H      J      K      L      M      N      O      Q      R
## 1087  811 1986 2578 1126 1353      0  947 3646 1015 2924 1396 1040  852  997
##      S      U      W
## 1086 1439  732
```

```
table(weapons$DISTRICT)
```

```
##
##      B      C      D      E      F      G      H      J      K      L      M      N      O      Q      R      S      U      W
## 123 110 214 283 212 235      0 116 328 112 383 218 199 115 288 322 201 108
```

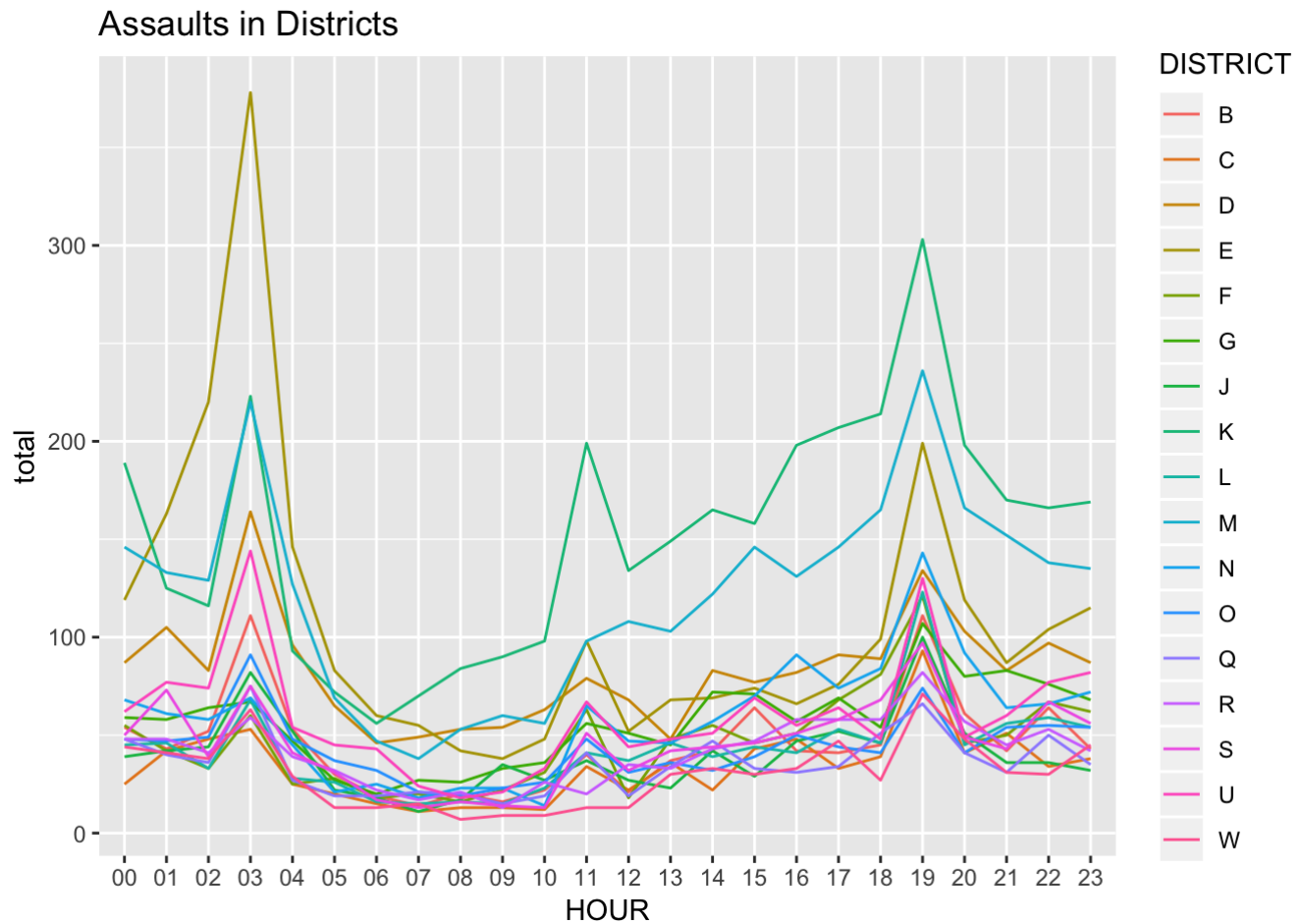
```
table(assaults$HOUR)
```

```
##
##      00      01      02      03      04      05      06      07      08      09      10      11      12      13      14
## 1186 1186 1157 2002  975  640  486  439  464  494  556 1050  759  872 1018
##      15      16      17      18      19      20      21      22      23
## 1086 1132 1215 1255 2190 1287 1138 1239 1189
```

```
table(weapons$HOUR)
```

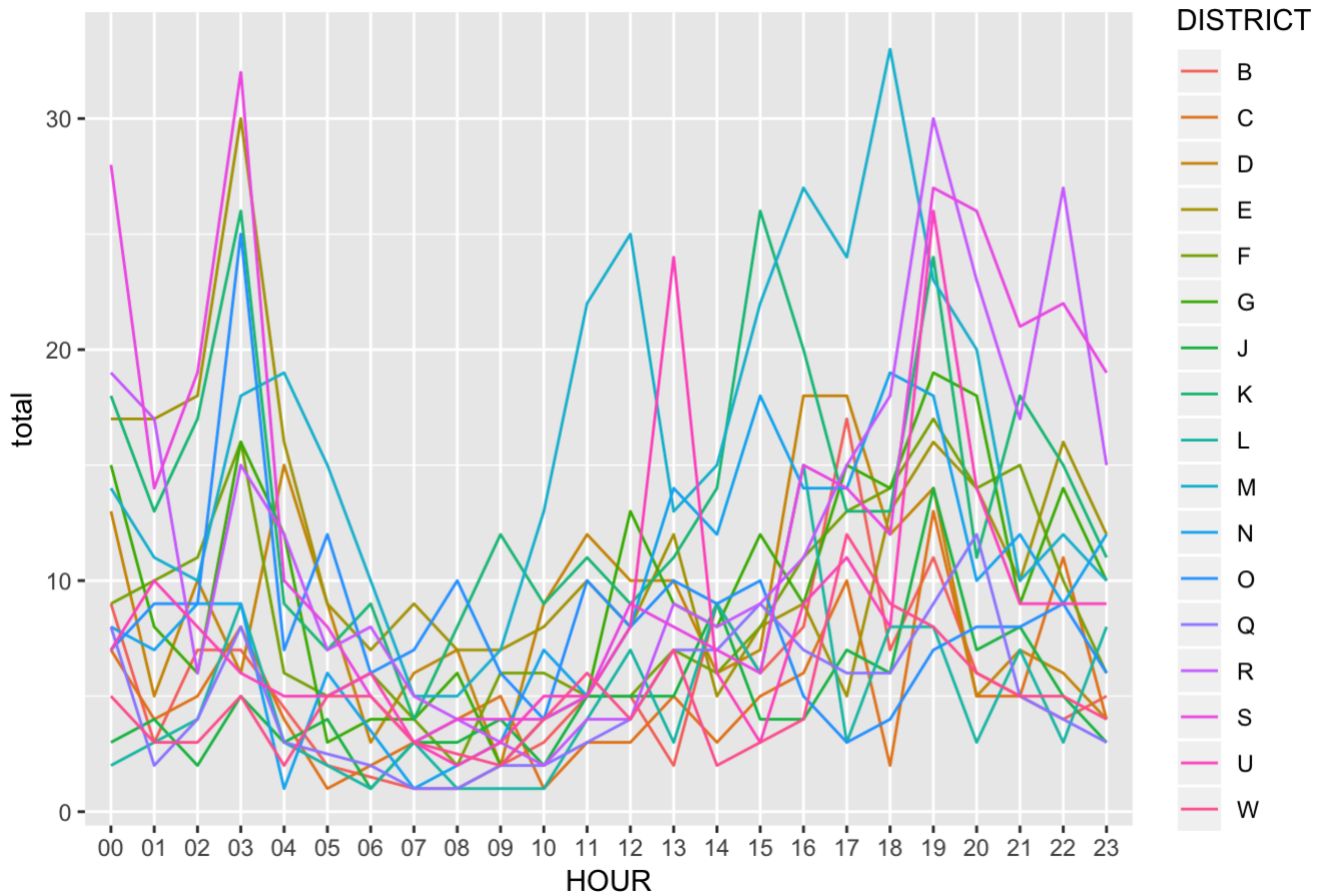
```
##
##      00      01      02      03      04      05      06      07      08      09      10      11      12      13      14      15      16      17
## 189 140 148 241 127  98  73  62  63  69  80 120 135 156 135 162 192 200
##      18      19      20      21      22      23
## 198 284 202 171 181 141
```

```
#Plot
assaults %>%
  group_by(DISTRICT, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = DISTRICT, color = DISTRICT)) + labs(title="Assaults in Districts")
```



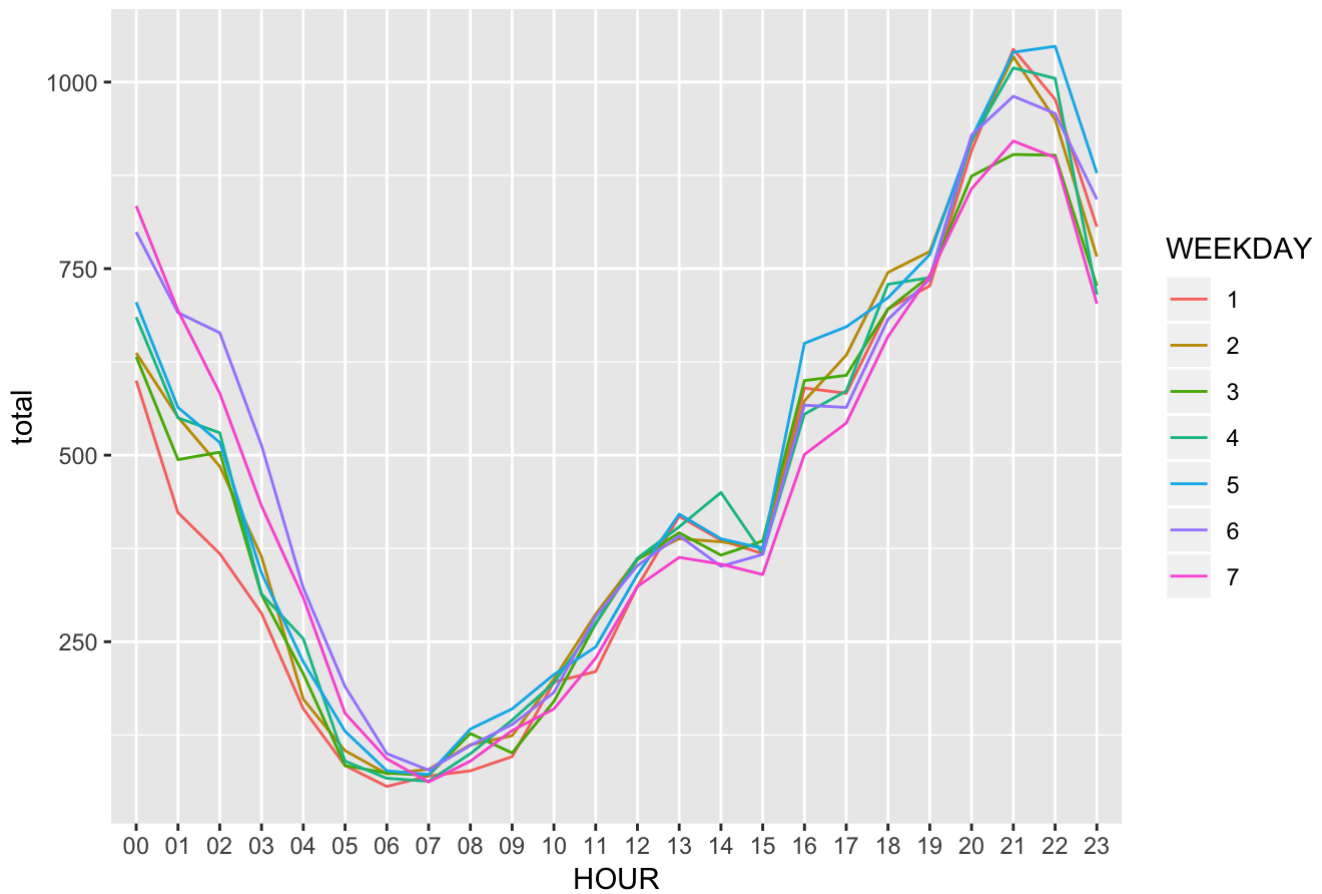
```
weapons %>%
  group_by(DISTRICT, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = DISTRICT, color = DISTRICT)) + labs(title="Weapons in Districts")
```

Weapons in Districts



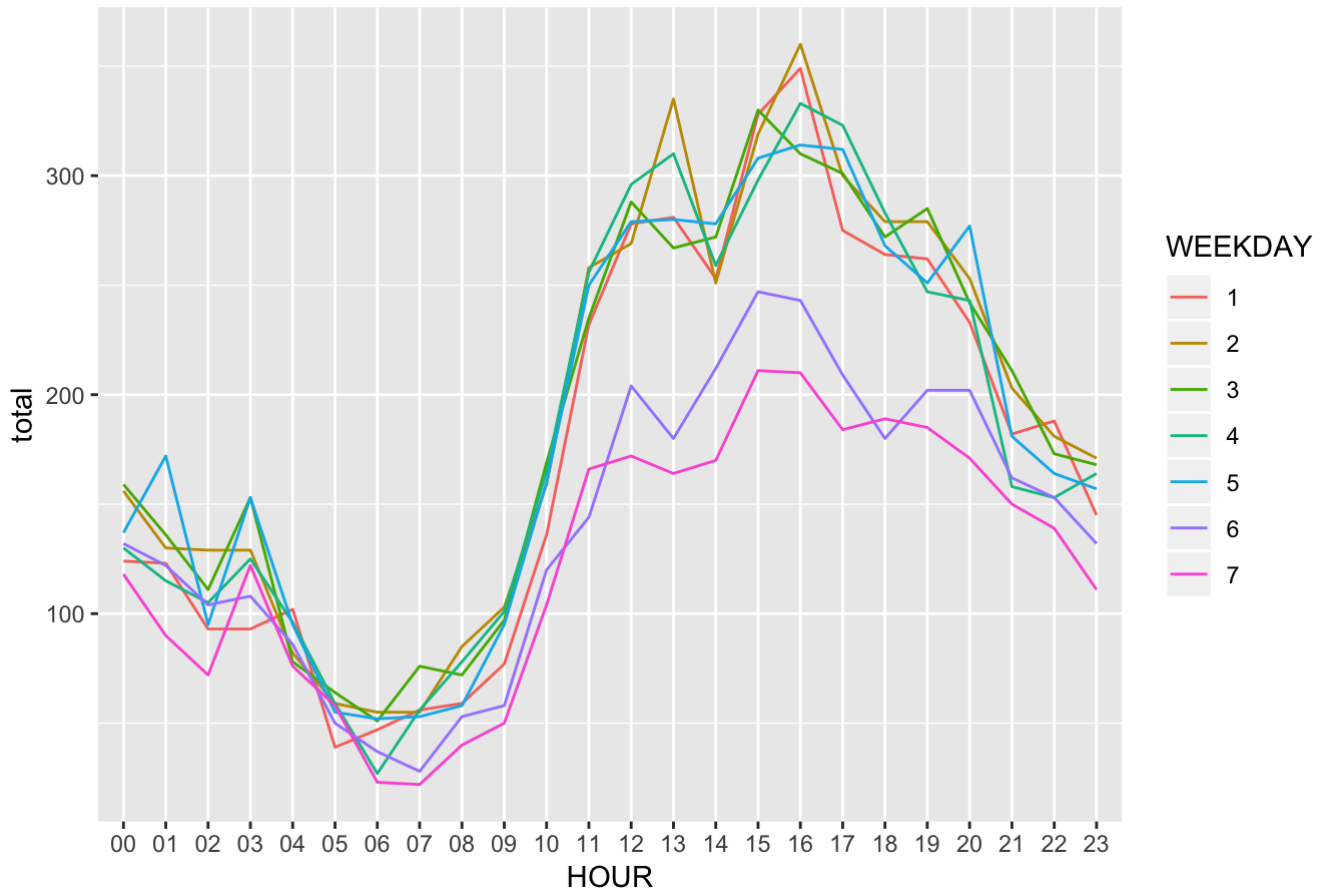
```
#Plots
liquor %>%
  group_by(WEEKDAY, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = WEEKDAY, color = WEEKDAY)) + labs(title="Liquor in Weekdays")
```

Liquor in Weekdays

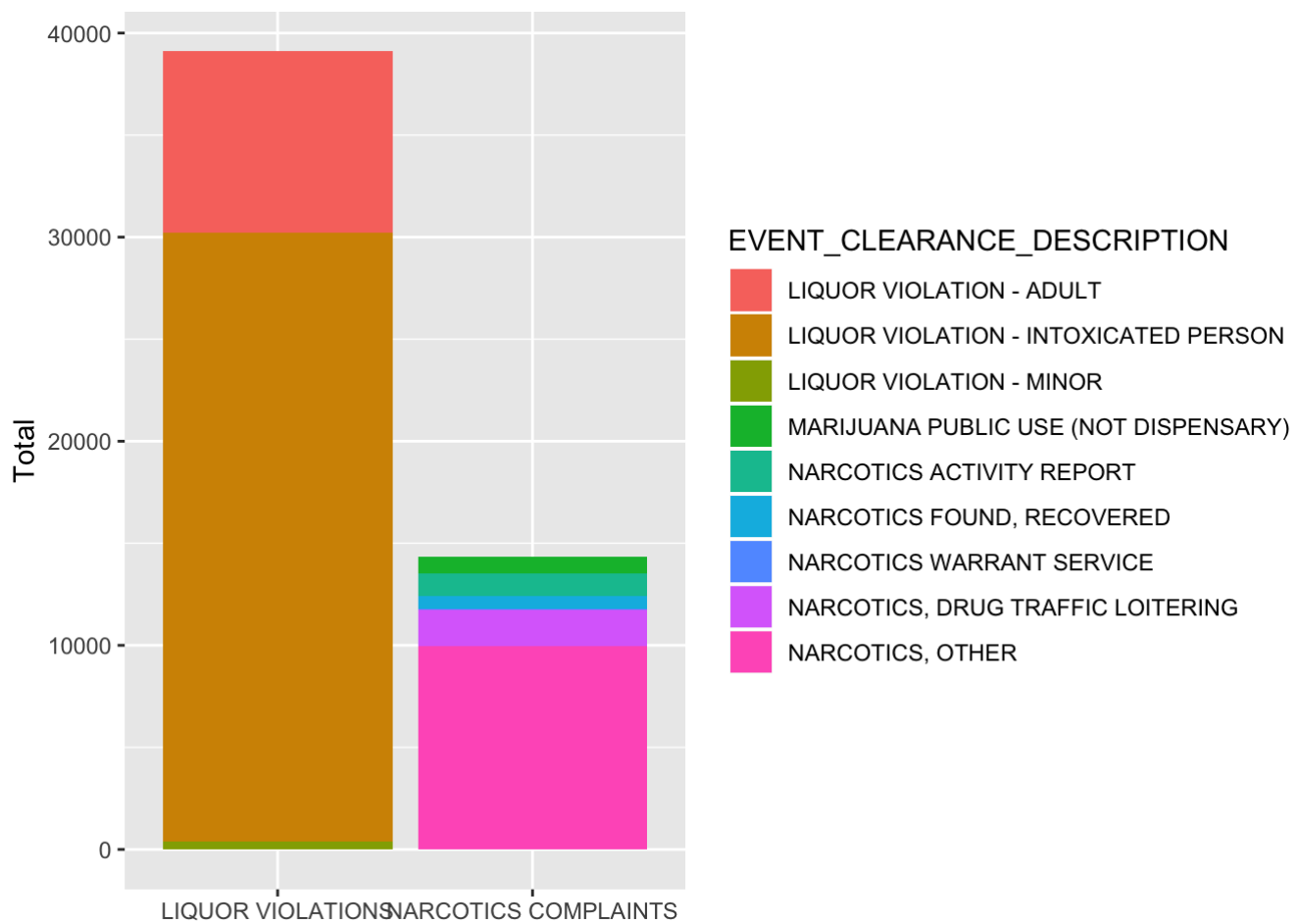


```
narcotics %>%
  group_by(WEEKDAY, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = WEEKDAY, color = WEEKDAY)) + labs(title="Narcotics Complaints in Weekdays")
```

Narcotics Complaints in Weekdays



```
#Substance as a whole
substance <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == c("NARCOTICS COMPLAINTS", "LIQUOR VIOLATIONS"))
substance %>%
  group_by(EVENT_CLEARANCE_GROUP, EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(Total = n()) %>%
  ggplot() + geom_col(mapping = aes(x=EVENT_CLEARANCE_GROUP, y=Total, fill=EVENT_CLEARAN
CE_DESCRIPTION)) + xlab(NULL)
```

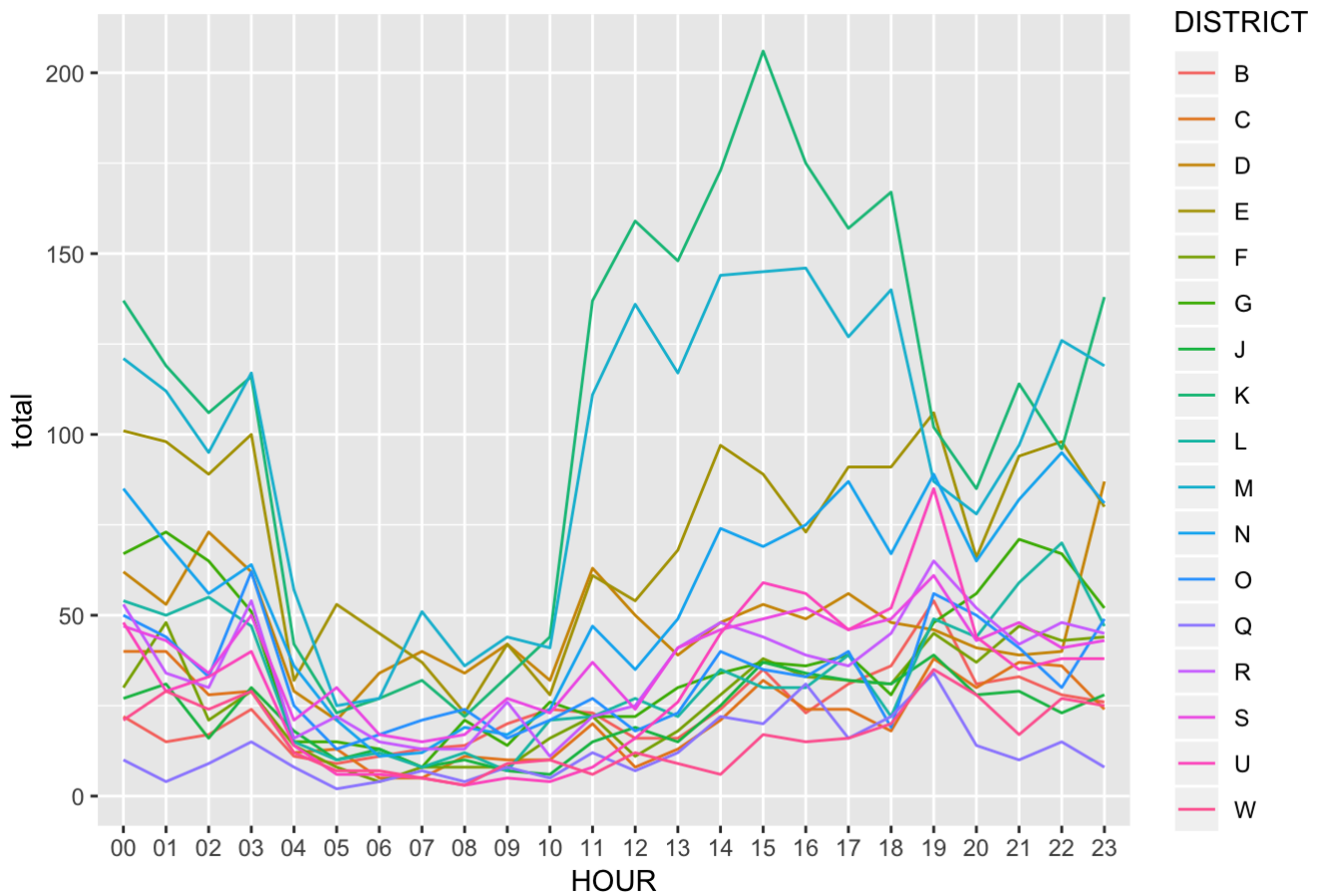


```
# Arrests
arrest <- data_v1 %>%
  filter(EVENT_CLEARANCE_GROUP == "ARREST")
arrest %>%
  group_by(EVENT_CLEARANCE_DESCRIPTION) %>%
  summarise(total = n())
```

```
## # A tibble: 2 x 2
##   EVENT_CLEARANCE_DESCRIPTION total
##   <chr>                      <int>
## 1 FELONY WARRANT SERVICE      8371
## 2 MISDEMEANOR WARRANT SERVICE 8567
```

```
arrest %>%
  group_by(DISTRICT, HOUR) %>%
  summarise(total = n()) %>%
  ggplot() + geom_line(mapping = aes(x = HOUR, y = total, group = DISTRICT, color = DISTRICT)) + labs(title="Arrests in Districts")
```

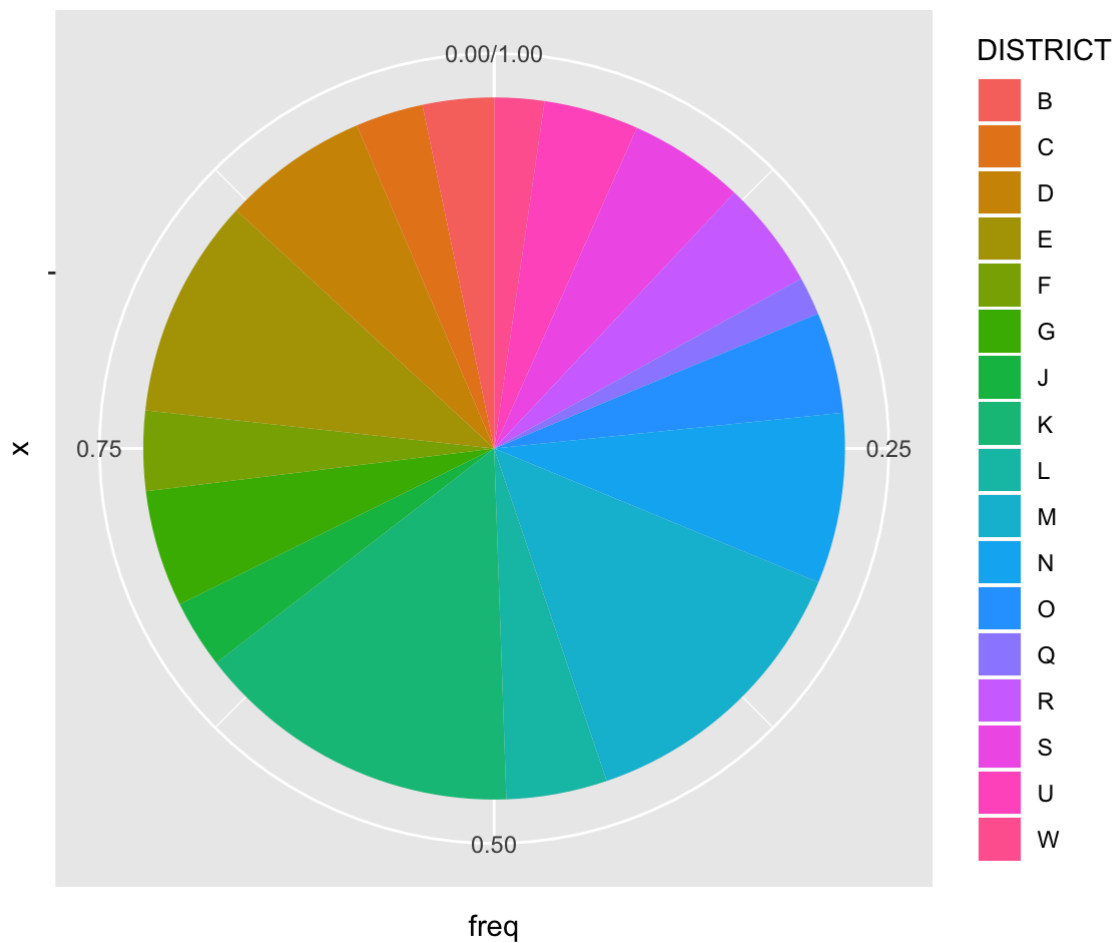

Arrests in Districts



```
table(arrest$DISTRICT)
```

```
##
##      B      C      D      E      F      G      H      J      K      L      M      N      O      Q      R
##  556  528  1141  1716  623  910      0  531  2558  787  2299  1330  787  299  839
##      S      U      W
##  904  741  389
```

```
#Pie chart
arrest %>%
  group_by(DISTRICT) %>%
  summarise (n = n()) %>%
  mutate(freq = n / sum(n)) %>%
  ggplot(mapping = aes(x="", y=freq, fill=DISTRICT)) + geom_bar(width = 1, stat = "identity") + coord_polar("y", start=0)
```



```
# Try modelling
#set.seed(123)
#data_v1$EVENT_CLEARANCE_GROUP = as.factor(data_v1$EVENT_CLEARANCE_GROUP)
#train_index <- sample(1:nrow(data_v1), 0.75 * nrow(data_v1))
#test_index <- setdiff(1:nrow(data_v1), train_index)

#train <- data_v1[train_index,]
#test <- data_v1[test_index,]
```

```
# library(caret)
# library(e1071)

# modell <- train(EVENT_CLEARANCE_GROUP~DISTRICT+LONGITUDE+LATITUDE+MONTH+HOUR+WEEKDAY,
# data = train, method = "knn")

# Too may outputs to classify (Almost 40). Not focus on this!
```