Chose Your Own Project - Machine Learning Submission

HarvardX Data Science Capstone - PH125.9x

Simon Gibson

2024-05-26

Contents

Introduction

For the 9th Course in the HarvardX Data Science course we have been asked to create two recommendation systems. The first was a Movie Recommendation System using the MovieLens dataset. The second is a "Choose your Own Project." For this a we have chosen a Workforce Recommendation System - mixing weather forecasts with Police 911 call information to see if it is possible to predict Police staffing requirements based on weather based trends.

We are using the Seattle Police Department 911 Incident Response data set found here : https://www.kaggle.com/datasets/sohier/seattle-police-department-911-incident-response

For Weather data we will use National Oceanic and Atmospheric Administration (NOAA) data. Michael Minns' tutorial is inciteful for weather analysis. It can be found here: https://michaelminn.net/tutorials/r-weather/index.html This weather data does not appear to be available via an api call or similar and is quite a manual download process. Due to download constraints we will be using a locally sourced dataset covering the years 2001 to 2002.

In order to test the results of the recommendation system we are using the root-mean-square error (RMSE) to measure the difference between the values predicted by the model and the observed values.

Method

The first step is to clear any set variables so we do not introduce anything unexpected into the data we are working with.

Then we install the packages required to manipulate the data.

Following that, the data is downloaded and then divided into 2 sets. The first set is used to train the algorithm and the second set is used to validate the algorithm. By dividing the data the problem of over-training and thus producing skewed results can be avoided.

The creation of the 2 sets involves the following steps. Initially required packages are installed if not installed and then loaded. Next the data is downloaded if the zip files are not found. Column names are set and the data is converted into forms more easily processed. Then the data is joined. Finally the joined data is split into 2 sets - the edx set used to train the algorithm and the final_holdout_test set that will be used to validate the algorithm and calculate the final RMSE score.

Data Investigation

Weather Dataset

Looking at the first 5 rows of data we can see the following:

STATION	NAME	DATE	PRCP	SNOW	TAVG	TMAX	TMIN	TSUN	WT01	WT02	WT03	WT04	WT05	WT06	WT07	WT08	WT09	WT11	WT13	WT14	WT15	WT16	WT17	WT18	WT19	WT21	WT22	WV01	WV03
USC00450872	BREMERTON, WA US	2000-01-01	0.23	0	NA	44	38	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
USC00450872	BREMERTON, WA US	2000-01-02	0.00	0	NA	44	31	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
USC00450872	BREMERTON, WA US	2000-01-03	0.10	0	NA	45	32	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
USC00450872	BREMERTON, WA US	2000-01-04	1.38	0	NA	47	35	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
USC00450872	BREMERTON, WA US	2000-01-05	0.02	0	NA	51	30	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
USC00450872	BREMERTON, WA US	2000-01-06	0.01	0	NA	44	34	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

The weather data starts from 2000-01-01 and ends 2002-12-31. It has the following column headers: STATION, NAME, DATE, PRCP, SNOW, TAVG, TMAX, TMIN, TSUN, WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT11, WT13, WT14, WT15, WT16, WT17, WT18, WT19, WT21, WT22, WV01, WV03. If we look at a single row, row 100 in this case, we can see the following: USC00450872, BREMERTON, WA US, 2000-04-09, 0, 0, NA, 66, 41, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA

For our investigation the following columns may be of interest - STATION, NAME, DATE, PRCP, SNOW, TMAX TMIN. The PRCP (Precipitation) and SNOW (Snowfall) columns are in inches and the TMAX (Maximum Temperature) TMIN (Minimum Temperature) are in Fahrenheit. These will be converted to their metric equivalents using the measurements package. We are interested in determining if there is any relationship between crime reports and weather so we will determine the closest weather station using latitude and longitude bearings returned by the ggmap package.

We have data from 20 stations:

	NAME	STATION
X2000.01.01	BREMERTON, WA US	USC00450872
X2000.01.01.1	EVERETT, WA US	USC00452675
X2000.01.01.2	MONROE, WA US	USC00455525
X2000.01.01.3	TOLT SOUTH FORK RESERVOIR, WA US	USC00458508
X2000.01.01.4	RENTON MUNICIPAL AIRPORT, WA US	USW00094248
X2000.01.01.5	KENT, WA US	USC00454169
X2000.01.01.6	TACOMA NUMBER 1, WA US	USC00458278
X2000.01.01.7	LANDSBURG, WA US	USC00454486
X2000.01.01.8	CEDAR LAKE, WA US	USC00451233
X2000.01.01.9	SNOQUALMIE FALLS, WA US	USC00457773
X2000.01.01.10	WAUNA 3 W, WA US	USC00459021
X2000.01.01.11	PALMER 3 ESE, WA US	USC00456295
X2000.01.01.12	TACOMA NARROWS AIRPORT, WA US	USW00094274
X2000.01.01.13	EVERETT SNOHOMISH CO AIRPORT, WA US	USW00024222
X2000.01.01.14	SEATTLE TACOMA AIRPORT, WA US	USW00024233
X2000.01.01.15	SEATTLE SAND POINT WEATHER FORECAST OFFICE, WA US	USW00094290
X2000.01.01.16	SEATTLE BOEING FIELD, WA US	USW00024234
X2000.11.22.4	GIG HARBOR 3.4 NW, WA US	US1WAPR0075
X2001.08.11.3	OLALLA 1.4 WNW, WA US	US1WAKP0013
X2001.12.02.10	WOODINVILLE 0.9 ENE, WA US	US1WAKG0078

Of 17773 rainfall measurements, 7869 recorded rainfall, and 9794 recorded no rainfall. The maximum rainfall during this period was 109.22mm which fell on 2002-01-07 at BREMERTON, WA US. Heavy rainfall is defined by NIWA as rainfall of over 100mm in 24 hours[1] and this occurred 3 times during the period we have data for.

[1] https://niwa.co.nz/natural-hazards/extreme-weather-heavy-rainfall

Of 17773 snowfall entries, 158 recorded snowfall, and 10382 recorded no snowfall, with 7233 not recording data. The maximum snowfall during this period was 609.6mm which fell on 2002-02-01 at TOLT SOUTH FORK RESERVOIR, WA US.

Over the period we have data for we have a maximum temperature of 37.22 at LANDSBURG, WA US and a minimum of -26.67 degrees Celsius at LANDSBURG, WA US. The mean maximum temperature was 15.08 while the mean minimum temperature was 6.23 degrees Celsius.

```
#We have data from `r n_distinct(Weather$STATION)` stations:

#r kable(Seattle_Stations, format = "markdown")`
#Seattle_Weather %>% group_by(Seattle_Weather$STATION)
```

```
#convert dates to posix dates from character class. ChatGPT used [OpenAI. ChatGPT (GPT-4). 2024. OpenAI
Seattle_911 <- Seattle_911 %>%
   mutate(`Event.Clearance.Date` = as.POSIXct(`Event.Clearance.Date`, format = "%m/%d/%Y %I:%M:%S %p", 
#Stop R converting Offense Number etc into Scientific Notation
options(scipen = 999)
summary(Seattle_911)
```

```
##   CAD.CDW.ID        CAD.Event.Number      General.Offense.Number
##  Length:1433853     Min.   : 9000209585   Min.   :      20114   
##  Class :character   1st Qu.:12000003919   1st Qu.: 2010347066   
##  Mode  :character   Median :14000330696   Median : 2012287942   
##                     Mean   :13660906898   Mean   : 1641393178   
##                     3rd Qu.:16000132431   3rd Qu.: 2015388964   
##                     Max.   :17000320040   Max.   :20122212881   
##                                                                 
##  Event.Clearance.Code Event.Clearance.Description Event.Clearance.SubGroup
##  Length:1433853       Length:1433853              Length:1433853          
##  Class :character     Class :character            Class :character        
##  Mode  :character     Mode  :character            Mode  :character        
##                                                                           
##                                                                           
##                                                                           
##                                                                           
##  Event.Clearance.Group Event.Clearance.Date            Hundred.Block.Location
##  Length:1433853        Min.   :2009-06-17 16:14:00.0   Length:1433853        
##  Class :character      1st Qu.:2011-12-31 06:44:00.0   Class :character      
##  Mode  :character      Median :2014-10-01 21:58:00.0   Mode  :character      
##                        Mean   :2014-03-05 07:13:45.2                         
##                        3rd Qu.:2016-04-21 13:45:28.0                         
##                        Max.   :2017-08-29 11:44:01.0                         
##                        NA's   :11584                                         
##  District.Sector     Zone.Beat         Census.Tract         Longitude    
##  Length:1433853     Length:1433853     Length:1433853     Min.   :-122  
##  Class :character   Class :character   Class :character   1st Qu.:-122  
##  Mode  :character   Mode  :character   Mode  :character   Median :-122  
##                                                           Mean   :-122  
##                                                           3rd Qu.:-122  
##                                                           Max.   :-122  
##                                                           NA's   :1     
##     Latitude  Incident.Location  Initial.Type.Description Initial.Type.Subgroup
##  Min.   :47   Length:1433853     Length:1433853           Length:1433853       
##  1st Qu.:48   Class :character   Class :character         Class :character     
##  Median :48   Mode  :character   Mode  :character         Mode  :character
##  Mean   :48
##  3rd Qu.:48
##  Max.   :48
##  NA's   :1
##  Initial.Type.Group At.Scene.Time
##  Length:1433853     Length:1433853
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

```
(unique(Seattle_911$`Event.Clearance.Description`))
```

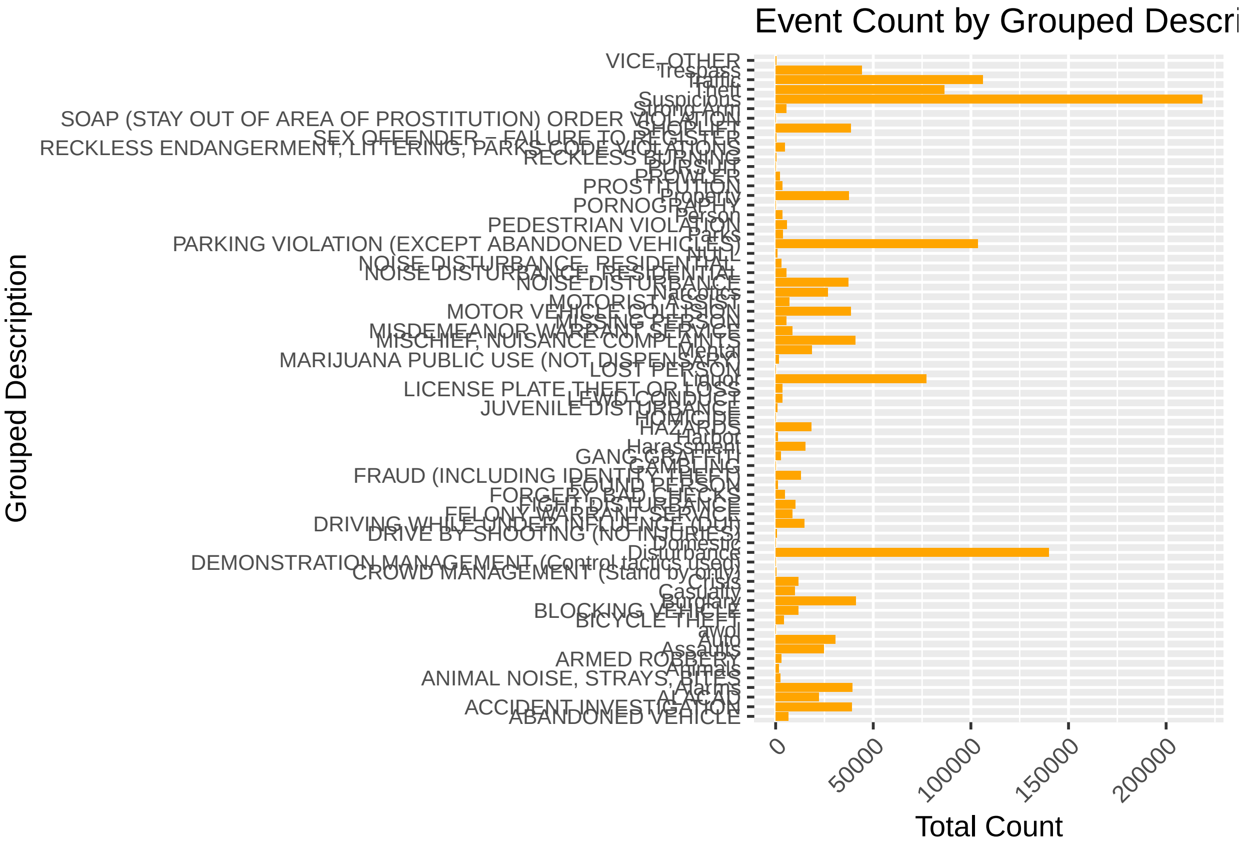
```
##   [1] "FIGHT DISTURBANCE"
##   [2] "THEFT - MISCELLANEOUS"
##   [3] "MISCHIEF, NUISANCE COMPLAINTS"
##   [4] "TRAFFIC (MOVING) VIOLATION"
##   [5] "SUSPICIOUS VEHICLE"
##   [6] "MENTAL COMPLAINT"
##   [7] "LIQUOR VIOLATION - INTOXICATED PERSON"
##   [8] "DISTURBANCE, OTHER"
##   [9] "TRESPASS "
##  [10] "ASSAULTS, OTHER"
##  [11] "SUSPICIOUS PERSON"
##  [12] "NARCOTICS, OTHER"
##  [13] "ACCIDENT INVESTIGATION"
##  [14] "NOISE DISTURBANCE"
##  [15] "PARKING VIOLATION (EXCEPT ABANDONED VEHICLES)"
##  [16] "SHOPLIFT"
##  [17] "PROWLER"
##  [18] "BURGLARY - RESIDENTIAL, OCCUPIED"
##  [19] "HAZARDS"
##  [20] "PROPERTY DESTRUCTION"
##  [21] "SUSPICIOUS CIRCUMSTANCES - BUILDING (OPEN DOOR, ETC.)"
##  [22] "THEFT - CAR PROWL"
##  [23] "BICYCLE THEFT"
##  [24] "PROSTITUTION"
##  [25] "RECKLESS ENDANGERMENT, LITTERING, PARKS CODE VIOLATIONS"
##  [26] "PEDESTRIAN VIOLATION"
##  [27] "BURGLARY - RESIDENTIAL, UNOCCUPIED"
##  [28] "NOISE DISTURBANCE, RESIDENTIAL "
##  [29] "ANIMAL NOISE, STRAYS, BITES"
##  [30] "ANIMALS - INJURED, DEAD, DANGEROUS"
##  [31] "HARASSMENT, THREATS"
##  [32] "CASUALTY (NON CRIMINAL/TRAFFIC) - MAN DOWN, SICK PERSONS, INJURED, DOA)"
##  [33] "AUTO RECOVERY"
##  [34] "MISSING PERSON"
##  [35] "ALARMS - RESIDENTIAL BURGLARY (FALSE)"
##  [36] "MENTAL PERSON PICK-UP OR TRANSPORT"
##  [37] "STRONG ARM ROBBERY "
##  [38] "ALARMS - VEHICLE (FALSE)"
##  [39] "ALARMS - COMMERCIAL BURGLARY (FALSE)"
##  [40] "MISDEMEANOR WARRANT SERVICE"
##  [41] "LIQUOR VIOLATION - MINOR"
##  [42] "DRIVING WHILE UNDER INFLUENCE (DUI)"
##  [43] "LIQUOR VIOLATION - ADULT"
##  [44] "ALACAD - COMMERCIAL BURGLARY (FALSE)"
##  [45] "MOTORIST ASSIST"
##  [46] "TRESPASS"
##  [47] "PROPERTY - FOUND (NON SPD GO#)"
##  [48] "BLOCKING VEHICLE"
##  [49] "PROPERTY - FOUND (FOLLOW UP TO SPD CASE)"
##  [50] "RECKLESS BURNING"
##  [51] "FELONY WARRANT SERVICE"
##  [52] "PARKS EXCLUSION "
##  [53] "CRISIS COMPLAINT - GENERAL"
##  [54] "PERSON WITH A GUN"
##  [55] "BURGLARY - COMMERCIAL "
##  [56] "AUTO THEFT"
##  [57] "ALARMS - RESIDENTIAL PANIC (FALSE)"
##  [58] "THEFT - AUTO ACCESSORIES"
##  [59] "TRESPASS - PARKS EXCLUSION"
##  [60] "HARASSMENT, THREATS - BY TELEPHONE, WRITING"
##  [61] "AUTO THEFT AND RECOVERY"
##  [62] "ALARMS - COMMERCIAL PANIC (FALSE)"
##  [63] "CASUALTY - DRUG RELATED (OVERDOSE, OTHER)"
##  [64] "PERSON WITH A WEAPON (NOT GUN)"
##  [65] "ABANDONED VEHICLE"
##  [66] "NARCOTICS, DRUG TRAFFIC LOITERING "
##  [67] "LEWD CONDUCT"
##  [68] "NARCOTICS ACTIVITY REPORT"
##  [69] "ARMED ROBBERY"
##  [70] "BURGLARY - UNOCCUPIED STRUCTURE ON RESIDENTIAL PROPERTY"
##  [71] "ALACAD - RESIDENTIAL BURGLARY (FALSE)"
##  [72] "ASSAULTS, FIREARM INVOLVED"
##  [73] "FRAUD (INCLUDING IDENTITY THEFT)"
##  [74] "GANG GRAFFITI"
##  [75] "LICENSE PLATE THEFT OR LOSS"
##  [76] "HARBOR - CODE VIOLATION"
##  [77] "PURSUIT"
##  [78] "FOUND PERSON"
##  [79] "FORGERY, BAD CHECKS"
##  [80] "NARCOTICS FOUND, RECOVERED"
##  [81] "JUVENILE DISTURBANCE"
##  [82] "MOTOR VEHICLE COLLISION"
##  [83] "DRIVE BY SHOOTING (NO INJURIES)"
##  [84] "ASSAULTS, GANG RELATED "
##  [85] "HARBOR - WATER EMERGENCIES"
##  [86] "PROPERTY - MISSING"
##  [87] "NARCOTICS, DRUG TRAFFIC LOITERING"
##  [88] "DISTURBANCE, GANG RELATED"
##  [89] "VICE, OTHER"
##  [90] "HARBOR - ASSIST BOATER (NON EMERGENCY)"
##  [91] "SEX OFFENDER - FAILURE TO REGISTER"
##  [92] "SOAP (STAY OUT OF AREA OF PROSTITUTION) ORDER VIOLATION"
##  [93] "HOMICIDE"
##  [94] "MARIJUANA PUBLIC USE (NOT DISPENSARY)"
##  [95] "BURGLARY - COMMERCIAL"
##  [96] "HARBOR - DEBRIS, NAVIGATIONAL HAZARDS"
##  [97] "NARCOTICS WARRANT SERVICE"
##  [98] "NOISE DISTURBANCE, RESIDENTIAL"
##  [99] "TRAFFIC CONTROL (SPECIAL EVENTS)"
## [100] "LOST PERSON"
## [101] "HARBOR - BOAT ACCIDENT"
## [102] "AWOL "
## [103] "HARBOR - BOATING UNDER THE INFLUENCE"
## [104] "HARBOR - VESSEL ABANDONED"
## [105] "GAMBLING"
## [106] "HARBOR - VESSEL THEFT AND RECOVERY"
## [107] "STRONG ARM ROBBERY"
## [108] "PARKS EXCLUSION"
## [109] "PORNOGRAPHY"
## [110] "HARBOR - MARINE FIRE"
## [111] "HARBOR - VESSEL THEFT"
## [112] "ALACAD - VEHICLE (FALSE)"
## [113] "ALACAD - RESIDENTIAL PANIC (FALSE)"
## [114] "LIQUOR VIOLATIONS (BUSINESS)"
## [115] "CROWD MANAGEMENT (Stand by only)"
## [116] "ALACAD - COMMERCIAL PANIC (FALSE)"
## [117] "HARBOR - VESSEL RECOVERY"
## [118] "CRISIS COMPLAINT - PICK-UP OR TRANSPORT"
## [119] ""
## [120] "DEMONSTRATION MANAGEMENT (Control tactics used)"
## [121] "ASSAULTS, GANG RELATED"
## [122] "TRAFFIC - COMMUNITY TRAFFIC COMPLAINT (CTC)"
## [123] "AWOL"
## [124] "TRAFFIC - BICYCLE VIOLATION"
## [125] "TRAFFIC - SCHOOL ZONE ENFORCEMENT"
## [126] "NULL"
## [127] "DOMESTIC SEX TRAFFICKING, ADULT"
## [128] "DOMESTIC HUMAN TRAFFICKING, ADULT"
```

The Seattle 911 dataset contains data from 2009-06-17 16:14:00 through to 2017-08-29 11:44:01. During this period, 1433853 CAD events were recorded.

```r
# Aggregate the data to get counts for each 'Event Clearance Desc'
event_counts <- Seattle_911 %>%
  count(`Event.Clearance.Description`, name = "count") %>%
  arrange(desc(count))

# Convert the count column to numeric if it isn't already
event_counts$count <- as.numeric(event_counts$count)

#Group Categories to provide simplified view (120 categories, some duplicated)
event_counts <- event_counts %>%
  mutate(GroupedDescription = case_when(
    grepl("^ALACAD", `Event.Clearance.Description`, ignore.case = TRUE) ~ "ALACAD",
    grepl("^ALARMS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Alarms",
    grepl("^ANIMALS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Animals",
    grepl("^ASSAULTS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Assaults",
    grepl("^AUTO", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Auto",
    grepl("^AWOL", `Event.Clearance.Description`, ignore.case = TRUE) ~ "awol",
    grepl("^BURGLARY", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Burglary",
    grepl("^CASUALTY", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Casualty",
    grepl("^CRISIS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Crisis",
    grepl("^DISTURBANCE", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Disturbance",
    grepl("^DOMESTIC", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Domestic",
    grepl("^TRAFFIC", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Traffic",
    grepl("^HARASSMENT", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Harassment",
    grepl("^HARBOR", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Harbor",
    grepl("^LIQUOR", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Liquor",
    grepl("^MENTAL", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Mental",
    grepl("^NARCOTICS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Narcotics",
    grepl("^PARKS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Parks",
    grepl("^PERSON", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Person",
    grepl("^PROPERTY", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Property",
    grepl("^STRONG", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Strong Arm",
    grepl("^SUSPICIOUS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Suspicious",
    grepl("^THEFT", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Theft",
    grepl("^TRESPASS", `Event.Clearance.Description`, ignore.case = TRUE) ~ "Trespass",
    TRUE ~ `Event.Clearance.Description`
  )) %>%
  # Filter out rows with NA or empty GroupedDescription
  filter(!is.na(GroupedDescription) & GroupedDescription != "") %>%
  # Group by GroupedDescription and Summarise
  group_by(GroupedDescription) %>%
  summarise(total_count = sum(count , na.rm = TRUE))


# View the result
print(event_counts)
```

```
## # A tibble: 69 x 2
##    GroupedDescription          total_count
##    <chr>                             <dbl>
##  1 ABANDONED VEHICLE                  6550
##  2 ACCIDENT INVESTIGATION            39070
##  3 ALACAD                            22184
##  4 ANIMAL NOISE, STRAYS, BITES        2356
##  5 ARMED ROBBERY                      2901
##  6 Alarms                            39343
##  7 Animals                            1581
##  8 Assaults                          24834
##  9 Auto                              30559
## 10 BICYCLE THEFT                      4340
## # i 59 more rows
```

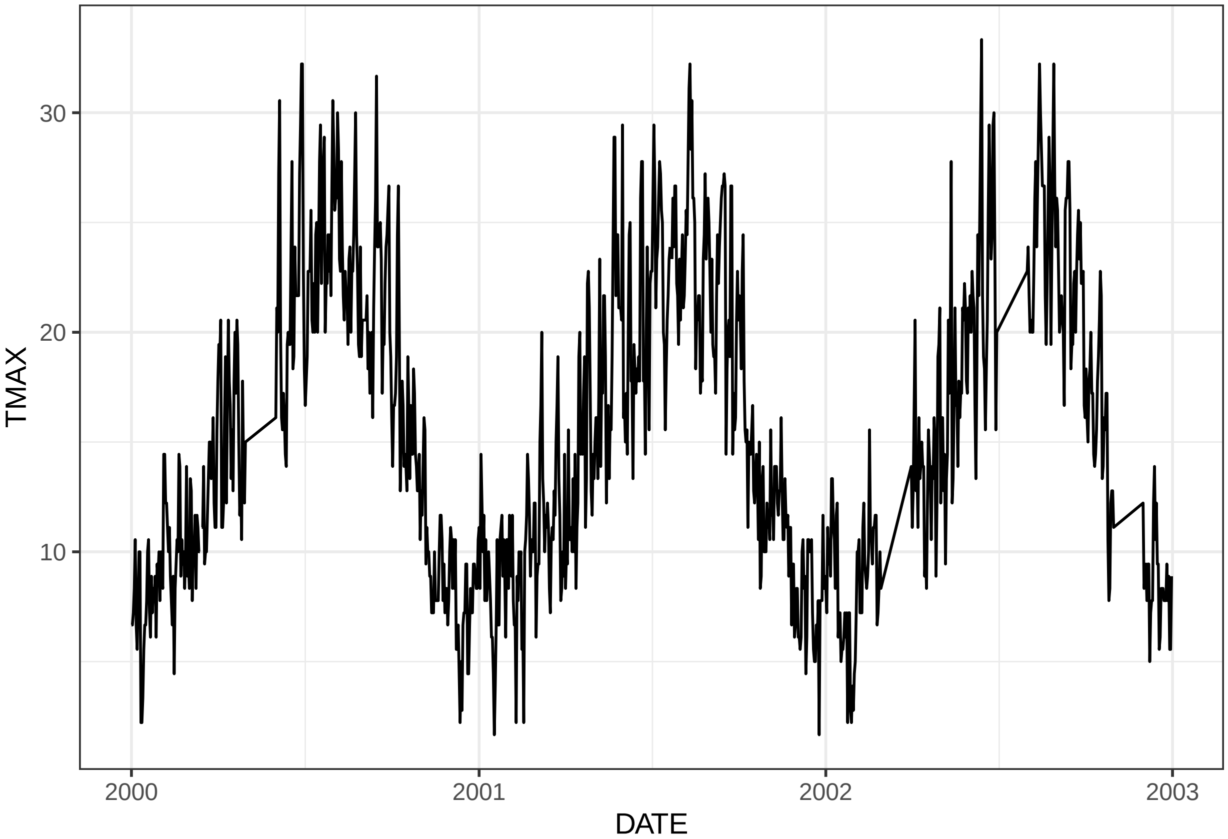
```
#event_count_graph <- ggplot(event_counts,aes(GroupedDescription)) +
 # geom_bar(stat = identity, fill = "orange") +
#  coord_flip()
#Sevent_count_graph

event_count_graph <- ggplot(event_counts, aes(x = GroupedDescription, y = total_count)) +
  geom_col(fill = "orange") +
  coord_flip() +
  labs(title = "Event Count by Grouped Description",
       x = "Grouped Description",
       y = "Total Count") +
  theme(axis.text.y = element_text(size = 8), # Reduce text size
        axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x-axis labels

# Print the graph
print(event_count_graph)
```



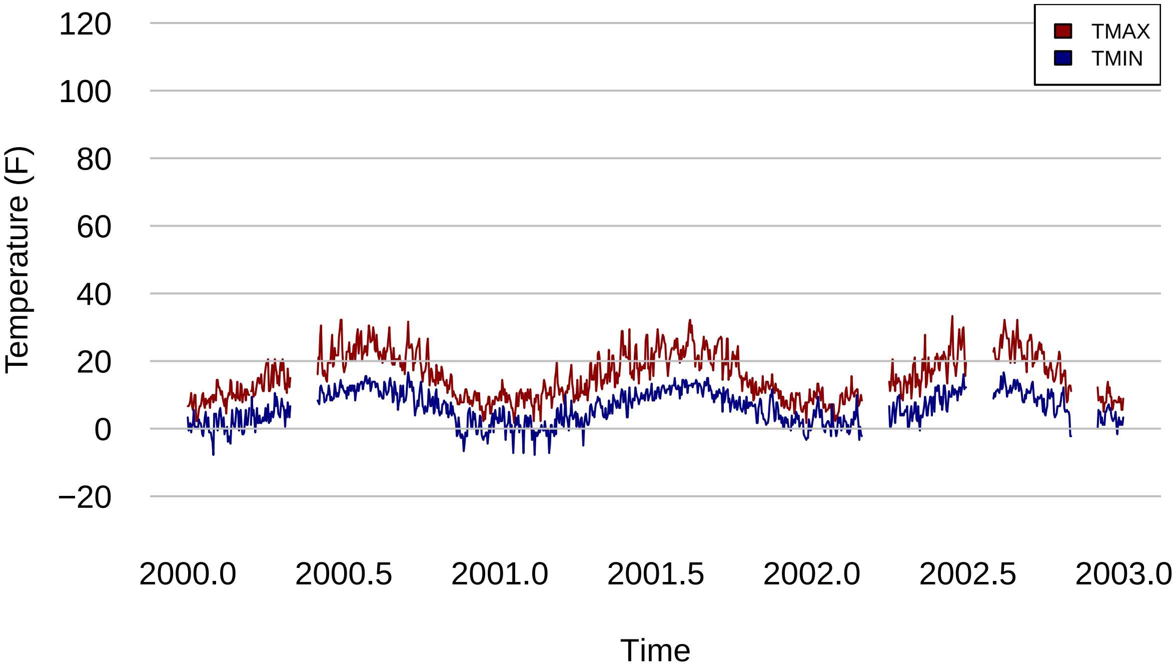
```
#event_counts_summary <- event_counts %>%
#  group_by(GroupedDescription) %>%
#  summarise(TotalCount = sum(count))


# Count occurrences of each description
#top_descriptions <- event_counts %>%
#  count(`Event.Clearance.Description`, name = "Count") %>%
#  arrange(desc(Count)) %>%
#  slice_head(n = 10)

# Print the top 10 descriptions
#print(top_descriptions)

# Plotting if necessary
#ggplot(top_descriptions, aes(x = reorder(`Event.Clearance.Description`, -Count), y = Count)) +
#  geom_bar(stat = "identity") +
#  coord_flip() +  # Flip coordinates for better readability
#  theme_minimal() +
#  labs(title = "Top 10 Most Common Event Clearance Descriptions",
#       x = "Event Clearance Description",
#       y = "Count")
```

To do - investigation of police data map weather station locations correlate weather station locations with police data

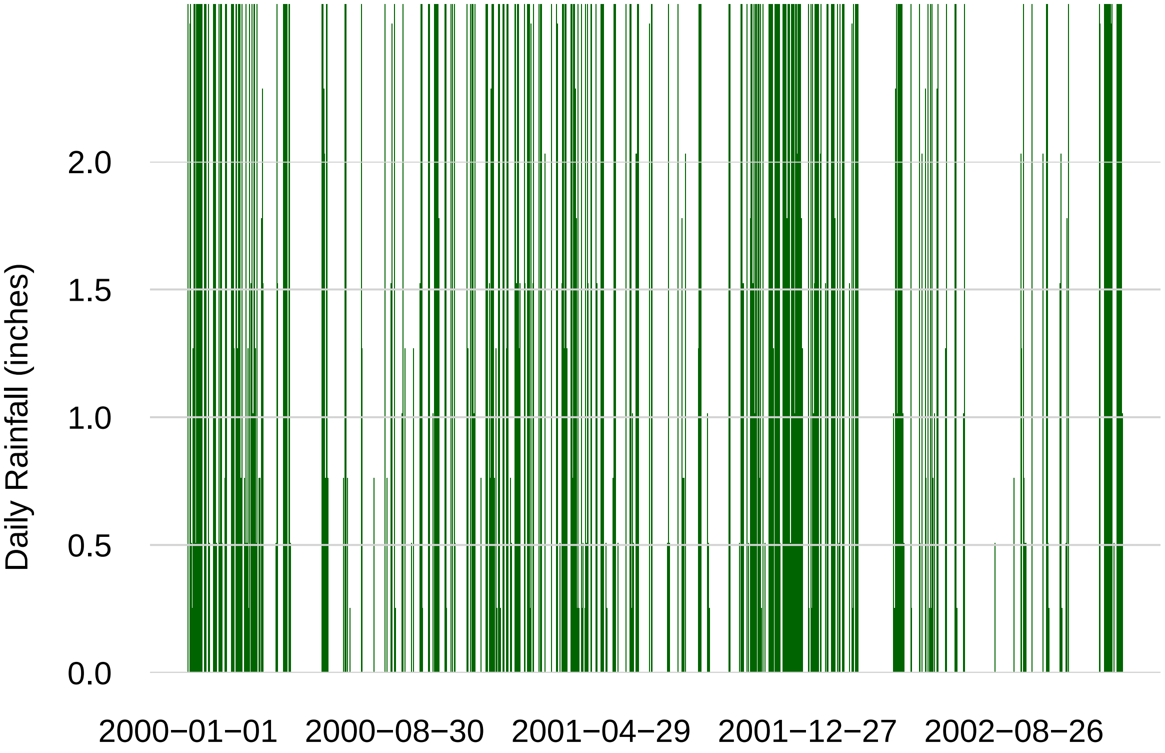
```
# Group Data by weather station
weather_data_grouped <- Seattle_Weather %>%
  group_by(STATION)

# find average maximum temperature
average_max_temp <- weather_data_grouped %>%
  summarise(avg_max_temp = mean(TMAX, na.rm = TRUE))

# Get unique station codes
station_codes <- unique(Seattle_Weather$STATION)

# Create a list to store data frames for each station
station_data_list <- list()

# Loop through each station code and filter data for that station
for (station_code in station_codes) {
  station_data <- filter(Seattle_Weather, STATION == station_code)
  station_data_list[[station_code]] <- station_data
}
ggplot(station_data_list[["USC00450872"]], aes(x=DATE, y=TMAX)) +
 geom_line() +
 theme_bw()
```



```
 USC00450872 <- station_data_list[["USC00450872"]]

historical = xts(USC00450872[,c("TMAX","TMIN","PRCP")], order.by=as.Date(USC00450872$DATE))

historical = ts_regular(historical)

historical = suppressWarnings(na.fill(historical, "extend"))

historical = window(historical, start=as.Date("2000-01-01"), end=as.Date("2020-12-31"))



plot(ts_ts(historical$TMAX), col="darkred", bty="n", las=1, fg=NA,
    ylim=c(-20, 120), ylab="Temperature (F)")

lines(ts_ts(historical$TMIN), col="navy")

grid(nx=NA, ny=NULL, lty=1, col="gray")

legend("topright", fill=c("darkred", "navy"), cex=0.7,
    legend=c("TMAX", "TMIN"), bg="white")
```



```r
barplot(historical$PRCP, border=NA, col="darkgreen", ylim=c(0, 2),
    space=0, bty="n", las=1, fg=NA, ylab="Daily Rainfall (inches)")

grid(nx=NA, ny=NULL, lty=1)
```



References

1.
2.

3.
4. https://www.neonscience.org/resources/learning-hub/tutorials/da-viz-coop-precip-data-r