DS\_05.R

anany

2024-06-15

#load necessary libraries  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

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

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

library(lubridate)

## Warning: package 'lubridate' was built under R version 4.3.3

##   
## Attaching package: 'lubridate'

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

library(leaflet)

## Warning: package 'leaflet' was built under R version 4.3.3

library(sp)

## Warning: package 'sp' was built under R version 4.3.3

library(readr)

## Warning: package 'readr' was built under R version 4.3.3

#loading the dataset  
Road\_accident <- read\_csv("~/Road\_accident.csv")

## Rows: 12316 Columns: 32

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (27): Day\_of\_week, Age\_band\_of\_driver, Sex\_of\_driver, Educational\_level...  
## dbl (4): Driving\_experience, Service\_year\_of\_vehicle, Number\_of\_vehicles\_i...  
## time (1): Time  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

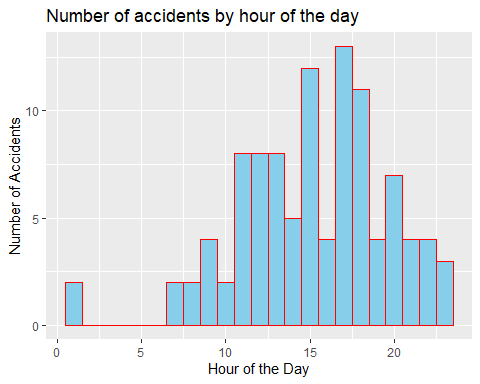
print(Road\_accident)

## # A tibble: 12,316 × 32  
## Time Day\_of\_week Age\_band\_of\_driver Sex\_of\_driver Educational\_level   
## <time> <chr> <chr> <chr> <chr>   
## 1 17:02 Monday 18-30 Male Above high school   
## 2 17:02 Monday 31-50 Male Junior high school  
## 3 17:02 Monday 18-30 Male Junior high school  
## 4 01:06 Sunday 18-30 Male Junior high school  
## 5 01:06 Sunday 18-30 Male Junior high school  
## 6 14:15 Friday 31-50 Male Junior high school  
## 7 17:30 Wednesday 18-30 Male Junior high school  
## 8 17:20 Friday 18-30 Male Junior high school  
## 9 17:20 Friday 18-30 Male Junior high school  
## 10 17:20 Friday 18-30 Male Junior high school  
## # ℹ 12,306 more rows  
## # ℹ 27 more variables: Vehicle\_driver\_relation <chr>, Driving\_experience <dbl>,  
## # Type\_of\_vehicle <chr>, Owner\_of\_vehicle <chr>,  
## # Service\_year\_of\_vehicle <dbl>, Defect\_of\_vehicle <chr>,  
## # Area\_accident\_occured <chr>, Lanes\_or\_Medians <chr>, Road\_allignment <chr>,  
## # Types\_of\_Junction <chr>, Road\_surface\_type <chr>,  
## # Road\_surface\_conditions <chr>, Light\_conditions <chr>, …

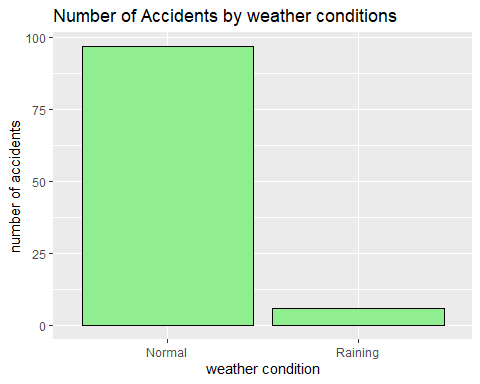
#checking for missing values   
Road\_accident<-na.omit(Road\_accident)  
  
#converting time column to appropriate data type  
Road\_accident$Time<-hms::as\_hms(Road\_accident$Time)  
  
#summary statistics for the dataset  
summary(Road\_accident)

## Time Day\_of\_week Age\_band\_of\_driver Sex\_of\_driver   
## Length:103 Length:103 Length:103 Length:103   
## Class1:hms Class :character Class :character Class :character   
## Class2:difftime Mode :character Mode :character Mode :character   
## Mode :numeric   
##   
##   
## Educational\_level Vehicle\_driver\_relation Driving\_experience  
## Length:103 Length:103 Min. : 0.000   
## Class :character Class :character 1st Qu.: 1.000   
## Mode :character Mode :character Median : 4.000   
## Mean : 4.971   
## 3rd Qu.: 8.000   
## Max. :16.000   
## Type\_of\_vehicle Owner\_of\_vehicle Service\_year\_of\_vehicle  
## Length:103 Length:103 Min. :1.000   
## Class :character Class :character 1st Qu.:3.000   
## Mode :character Mode :character Median :4.000   
## Mean :4.757   
## 3rd Qu.:6.500   
## Max. :9.000   
## Defect\_of\_vehicle Area\_accident\_occured Lanes\_or\_Medians Road\_allignment   
## Length:103 Length:103 Length:103 Length:103   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Types\_of\_Junction Road\_surface\_type Road\_surface\_conditions  
## Length:103 Length:103 Length:103   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Light\_conditions Weather\_conditions Type\_of\_collision   
## Length:103 Length:103 Length:103   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Number\_of\_vehicles\_involved Number\_of\_casualties Vehicle\_movement   
## Min. :1.000 Min. :1.000 Length:103   
## 1st Qu.:2.000 1st Qu.:1.000 Class :character   
## Median :2.000 Median :1.000 Mode :character   
## Mean :1.883 Mean :1.534   
## 3rd Qu.:2.000 3rd Qu.:2.000   
## Max. :3.000 Max. :4.000   
## Casualty\_class Sex\_of\_casualty Age\_band\_of\_casualty Casualty\_severity   
## Length:103 Length:103 Length:103 Length:103   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Work\_of\_casuality Fitness\_of\_casuality Pedestrian\_movement Cause\_of\_accident   
## Length:103 Length:103 Length:103 Length:103   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Accident\_severity   
## Length:103   
## Class :character   
## Mode :character   
##   
##   
##

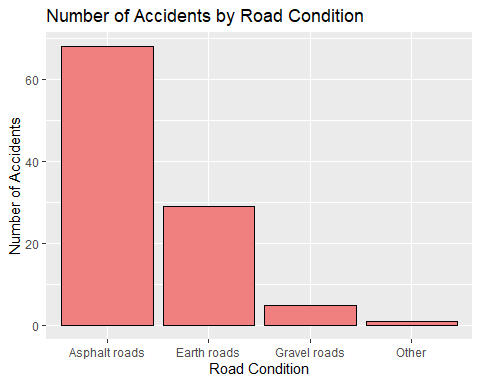
#extracting hours from the time column  
Road\_accident$Hour<-hour(Road\_accident$Time)  
  
#ploting the number of accidents by hour  
ggplot(Road\_accident,aes(x=Hour))+geom\_histogram(binwidth = 1,fill="skyblue",color="red")+labs(title = "Number of accidents by hour of the day",x="Hour of the Day",y="Number of Accidents")



#accidents by weather condition  
ggplot(Road\_accident,aes(x=Weather\_conditions))+geom\_bar(fill="lightgreen",color="black")+labs(title = "Number of Accidents by weather conditions",x="weather condition",y="number of accidents")



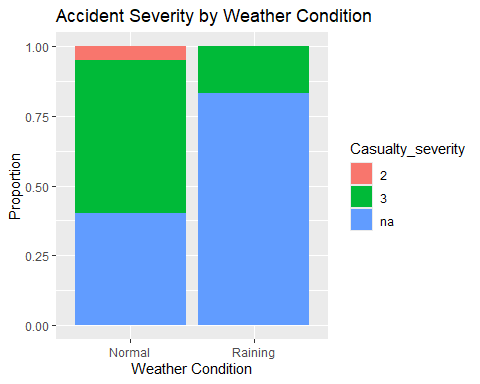
#accidents by road condition  
ggplot(Road\_accident,aes(x=Road\_surface\_type))+geom\_bar(fill="lightcoral",color="black")+labs(title = "Number of Accidents by Road Condition",x="Road Condition",y="Number of Accidents")



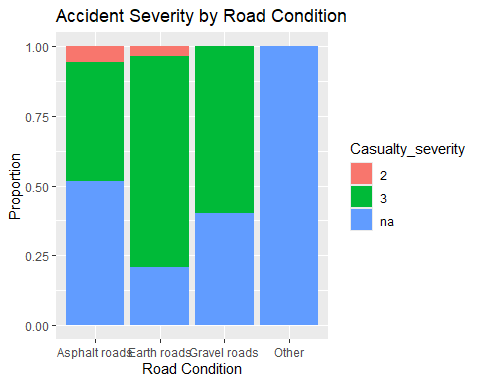
# Example columns: Latitude and Longitude  
if ("Latitude" %in% colnames(Road\_accident) & "Longitude" %in% colnames(Road\_accident)) {  
 leaflet(data = Road\_accident) %>%  
 addTiles() %>%  
 addCircleMarkers(~Longitude, ~Latitude, radius = 3, color = ~Severity,  
 popup = ~paste("Date:", Date, "<br>", "Time:", Time, "<br>", "Severity:", Severity),  
 clusterOptions = markerClusterOptions()) %>%  
 addLegend("bottomright", pal = colorFactor(c("blue", "orange", "red"), domain = Road\_accident$Severity),  
 values = ~Severity, title = "Severity")  
} else {  
 print("Latitude and Longitude columns are not available in the dataset.")  
}

## [1] "Latitude and Longitude columns are not available in the dataset."

# Analyze the impact of weather, road condition, and time of day on accident severity  
ggplot(Road\_accident, aes(x=Weather\_conditions, fill=Casualty\_severity)) +  
 geom\_bar(position="fill") +  
 labs(title="Accident Severity by Weather Condition", x="Weather Condition", y="Proportion")



ggplot(Road\_accident, aes(x=Road\_surface\_type, fill=Casualty\_severity)) +  
 geom\_bar(position="fill") +  
 labs(title="Accident Severity by Road Condition", x="Road Condition", y="Proportion")



ggplot(Road\_accident, aes(x=Hour, fill=Casualty\_severity)) +  
 geom\_histogram(binwidth=1, position="fill") +  
 labs(title="Accident Severity by Hour of the Day", x="Hour of the Day", y="Proportion")

## Warning: Removed 15 rows containing missing values or values outside the scale range  
## (`geom\_bar()`).

