Traffic violations

SQL SCHEMA:

```
create database traffic;

use traffic;

create table violations(dateofstop varchar(100),timeofstop varchar(100),agency varchar(100),subagency
varchar(100),description varchar(100),location varchar(100),lat float,longi float,accident
varchar(100),belts varchar(100),personalinjury varchar(100),propertydamage varchar(100),fatal
varchar(100),commlic varchar(100),hazmat varchar(100),commvechile varchar(100),alcohol
varchar(100),workzone varchar(100),state varchar(100),vechiletype varchar(100),year int,make
varchar(100),model varchar(100),color varchar(100),violationtype varchar(100),charge
varchar(100),article varchar(100),contributetoacc varchar(100),race varchar(100),gender
varchar(100),drivercity varchar(100),driverstate varchar(100),dltype varchar(100),arresttype
varchar(100),geoloc varchar(100));
```

R Script: [1] [2] [3]

```
library(tidyverse)

library(ggplot2)

library(plyr)

data <- read.csv('F:/study/masters/1sem/ait-580/final project/Traffic_Violations.csv')


#remove empty rows and rows with NA's

data <- na.omit(data)


#select required columns

data <- subset(data,select=c(7:19,21,28:30,32))

data <- data.frame(data)


#Handling categorical variables i.e converting Yes ->1, No->0

data$Accident <- ifelse(data$Accident == "Yes",1,0)

data$Belts <- ifelse(data$Belts == "Yes",1,0)

data$Personal.Injury <- ifelse(data$Personal.Injury == "Yes",1,0)

data$Property.Damage <- ifelse(data$Property.Damage == "Yes",1,0)

data$Fatal <- ifelse(data$Fatal == "Yes",1,0)
```

```
data$Commercial.License <- ifelse(data$Commercial.License == "Yes",1,0)
data$HAZMAT <- ifelse(data$HAZMAT == "Yes",1,0)
data$Commercial.Vehicle <- ifelse(data$Commercial.Vehicle == "Yes",1,0)
data$Alcohol <- ifelse(data$Alcohol == "Yes",1,0)
data$Work.Zone <- ifelse(data$Work.Zone == "Yes",1,0)
data$Contributed.To.Accident <- ifelse(data$Contributed.To.Accident == "Yes",1,0)
attach(data)
summary(Accident)
summary(Belts)
summary(Personal.Injury)
summary(Property.Damage)
summary(Fatal)
summary(Commercial.License)
summary(HAZMAT)
summary(Commercial.Vehicle)
summary(Alcohol)
summary(Work.Zone)
summary(Contributed.To.Accident)

#count number of male,female violate dthe traffic rules
gen <- count(Gender)
#Generating visualizations
#ggplot(gen,aes(x=x,y=freq))+geom_point()
plot(Gender,col="Blue")+title(xlab='Gender',ylab = 'Count')

#How many traffic violations are recorded in respective years
yrs <- subset(data,Year>1990 & Year<2018)
filter_yrs <- yrs$Year
plt <- as.data.frame(table(filter_yrs))
```

```
ggplot(plt,aes(x=filter_yrs,y=Freq))+geom_point(col="blue")+labs(title ="Violations recorded in respective years",x="Year",y="Number of violations")


#Box plot representing count of violations in all states compared to drivers violated in their own state

st <- as.data.frame(table(State))

drst <- as.data.frame(table(Driver.State))

st <- st[sample(1:nrow(st),69,replace= FALSE),]

names(drst)[1] <- paste("State")

df <- merge(st,drst,by="State")

df <- na.omit(df)

df <- df[,2:3]

names(df)[1] <- paste("Violations_allstates")

names(df)[2] <- paste("Violations_ownstate")

boxplot(df,ylim=c(0,3000),col=c("blue","brown"))


#correlation test & hypothesis test

cor.test(df$`Violations_allstates`,df$`Violations_ownstate`)

library(ggpubr)

ggscatter(df,x='Violations_allstates',y='Violations_ownstate',add = "reg.line",conf.int = TRUE,cor.coef = TRUE,cor.method = "pearson",color = 'blue')


#Generating logistic reression model & hypothesis test

summary(glm(Accident~Alcohol,family = binomial(link=logit)))
```

References

[1] datascience+, "Mastering R Plot – Part 1: colors, legends and lines," [Online]. Available: https://datascienceplus.com/mastering-r-plot-part-1-colors-legends-and-lines/.

[2] stackoverflow, "R:Select values from data table in range," [Online]. Available: https://stackoverflow.com/questions/5204953/r-select-values-from-data-table-in-range.

[3] R, "Logistic Regression in R," [Online]. Available: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R7_LogisticRegression-Survival/R7_LogisticRegression-Survival3.html.