Project 5 Solutions

Kevin Choe

Collaborators: N/A

TA help: Hilda Ibriga and Katie Brinkers guided with problems 1-5, explained new concepts

Online resources used: Stat 190 Example Book (All problems)

Question 1

```
#Combines accident files into one big dataset
accidents <- rbind(read.csv("/class/datamine/data/fars/1975/ACCIDENT.CSV"),
                   read.csv("/class/datamine/data/fars/1976/ACCIDENT.CSV"),
                   read.csv("/class/datamine/data/fars/1977/ACCIDENT.CSV"),
                   read.csv("/class/datamine/data/fars/1978/ACCIDENT.CSV"),
                   read.csv("/class/datamine/data/fars/1979/ACCIDENT.CSV"),
                   read.csv("/class/datamine/data/fars/1980/ACCIDENT.CSV"),
                   read.csv("/class/datamine/data/fars/1981/ACCIDENT.CSV"))
#Pasting 19 to every value in the year column
accidents$YEAR <- paste0("19", accidents$YEAR)
#Tells all unique values
unique(accidents$YEAR)
```

```
[1] "1975" "1999" "1976" "1977" "1978" "1979" "1980" "1981"
```

Question 2

```
#Printing values where statement is true, false and N/A, specifically looking for true values
table(accidents$DRUNK_DR >= 1&accidents$SCH_BUS >= 1, useNA= "always")
```

```
 FALSE   TRUE   <NA>
284228    101  15730
```

```
#The true value is 101
```

Question 3

```
#Print number of accidents involving more than 1 drunk driver
table(accidents$YEAR[(accidents$DRUNK_DR >= 1)&(accidents$SCH_BUS >= 1)])
```

```
1977 1978 1979 1980 1981
  12   37   17   12   23
```

1978 had the largest number of accidents

Question 4

```
#Finds the mean number of motorists in an accident and drunk drivers
for (i in 0:6) {
  print(paste0("The mean number of motorists in an accident with ", i, " drunk drivers is ", mean(accid
}
```

```
[1] "The mean number of motorists in an accident with 0 drunk drivers is 2.61554022666993"
[1] "The mean number of motorists in an accident with 1 drunk drivers is 2.47407925704472"
[1] "The mean number of motorists in an accident with 2 drunk drivers is 3.66071099650028"
[1] "The mean number of motorists in an accident with 3 drunk drivers is 5.19791666666667"
[1] "The mean number of motorists in an accident with 4 drunk drivers is 5.25"
[1] "The mean number of motorists in an accident with 5 drunk drivers is NaN"
[1] "The mean number of motorists in an accident with 6 drunk drivers is 6"
```

```
#Other way to get the answer
tapply(accidents$PERSONS, accidents$DRUNK_DR, mean)
```

```
       0        1        2        3        4        6
2.615540 2.474079 3.660711 5.197917 5.250000 6.000000
```

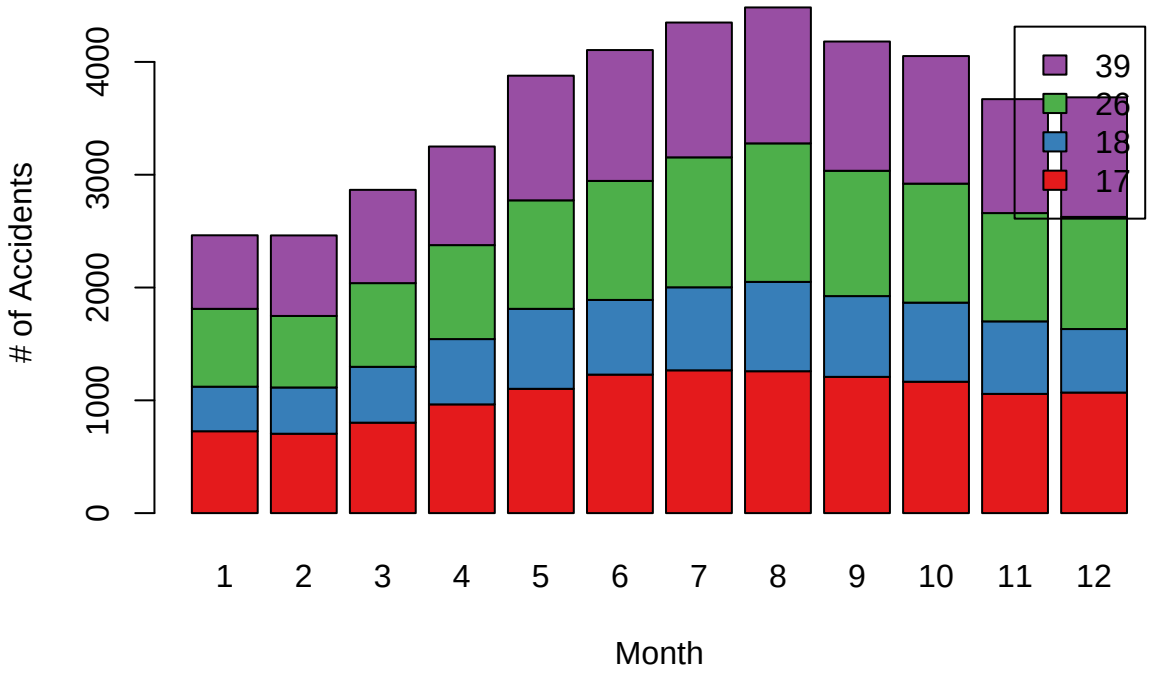
Question 5

```
library(RColorBrewer)
#Creating a vector of states.
sN <- c(17, 18, 26, 39)

# Extract data from accidents only where the state code is either 17, 18, 26, or 39
mwDF <- accidents[accidents$STATE %in% sN,]

t <- table(mwDF$STATE, mwDF$MONTH)

# Using a table comparing the number of accident rows in each month for each state as data for a barplo
barplot(table(mwDF$STATE, mwDF$MONTH),
        legend=rownames(t),
        xlab="Month",
        ylab="# of Accidents",
        main="Accidents for Each State, by Month",
        col=brewer.pal(4, "Set1")
)
```



The most accidents happen in month 8 followed by month 7 and month 9. This does not suprise me because this is when people are most free.

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - We are Purdue.