# Assignment_3

```
library(readxl)
library(dplyr)
library(knitr)
accident <- read_excel("Road Traffic Accident (Batch 3B).xlsx")
accident <- data.frame(accident)
str(accident)
```

```
## 'data.frame':    50 obs. of  11 variables:
##  $ Serial.number      : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Age..yrs.          : num  24 63 23 41 25 61 35 27 38 48 ...
##  $ Gender             : chr  "Male" "Male" "Female" "Female" ...
##  $ Occupation         : chr  "Bussinesman" "Retired Person" "Student" "Student" ...
##  $ Site.of.death      : chr  "On the way" "Hospital" "Spot" "Hospital" ...
##  $ Time.of.death      : chr  "After 1 hour" "After 120  hours" "Immediately" "After 144  hours" ...
##  $ Alco.Drug.etc      : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Use.of.helmet      : chr  "No" "No" "No" "Na" ...
##  $ Type.of.vehicle    : chr  "Scooter" "Scooty" "Scooter" "Kinetic Honda" ...
##  $ Period.of.coma.hrs.: num  1 120 NA 144 192 NA NA NA NA 2 ...
##  $ Person.died        : chr  "Rider" "Rider" "Pillion" "Rider" ...
```

```
View(accident)
```

# 1. How many children (<= 18 years) have died in the accident and how many of them are males and females?

```
died_of_accident <-  accident %>% filter(`Age..yrs.` <= 18) %>% group_by(Gender) %>% summarise(Total_accidents = n())

  died_of_accident <- rbind(died_of_accident,c("TOTAL",sum(died_of_accident$Total_accidents)))

  kable(died_of_accident)
```

| Gender | Total_accidents |
|---|---|
| Female | 2 |
| Male | 6 |
| TOTAL | 8 |

# 2. How many children had taken alcohol? Were they riders or pillion when they died?

```
kable(accident %>% select(Age..yrs.,Alco.Drug.etc,Person.died) %>% filter(Age..yrs.<=18,Alco.Drug.etc=="Yes"))
```

| Age..yrs. | Alco.Drug.etc | Person.died |
|---|---|---|
| 17 | Yes | Pillion |
| 12 | Yes | Pillion |
| 18 | Yes | Rider |

# 3. How many children have died immediately in the accident as pillion riders who had consumed alcolhol?

```
kable(accident %>% filter(Age..yrs.<=18,Alco.Drug.etc=="Yes",Time.of.death=="Immediately") %>% group_by(Person.died) %>% sum
marise(Total_died=n()))
```

| Person.died | Total_died |
|---|---|
| Pillion | 2 |
| Rider | 1 |

# 4. How many have died under the influence of alcohol/drug and without wearing helmet?

```
no.helmet_alcohol<- accident %>% filter(Alco.Drug.etc=="Yes",Use.of.helmet=="No")

kable(no.helmet_alcohol)
```

| Serial.number | Age..yrs. | Gender | Occupation | Site.of.death | Time.of.death | Alco.Drug.etc | Use.of.helmet | Type.of.vehicle | Period.of.coma.hrs. | P |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | Male | Bussinesman | On the way | After 1 hour | Yes | No | Scooter | 1 | R |
| 2 | 63 | Male | Retired Person | Hospital | After 120 hours | Yes | No | Scooty | 120 | R |
| 3 | 23 | Female | Student | Spot | Immediately | Yes | No | Scooter | NA | P |
| 8 | 27 | Male | Bussinesman | Hospital | Immediately | Yes | No | Scooter | NA | R |
| 15 | 22 | Male | Student | Hospital | After 284 hours | Yes | No | TVS 50 | 384 | P |
| 17 | 58 | Male | Polic constable | Spot | Immediately | Yes | No | Scooter | NA | R |
| 20 | 19 | Female | Watchman | Hospital | After 5 hours | Yes | No | Kinetic Honda | 5 | R |
| 26 | 21 | Male | Bussinesman | Hospital | After 72 hours | Yes | No | Kinetic Honda | 72 | R |
| 27 | 17 | Female | Carpenter | Spot | Immediately | Yes | No | TVS 50 | NA | P |
| 30 | 34 | Female | Student | Spot | Immediately | Yes | No | TVS 50 | NA | R P |
| 36 | 12 | Male | Rag picker | On the way | Immediately | Yes | No | Kinetic Honda | NA | P |
| 37 | 18 | Male | Bussinesman | Spot | Immediately | Yes | No | Hero Honda | NA | R |
| 39 | 72 | Male | Retired Person | Hospital | After 48 hours | Yes | No | Hero Honda | 48 | R |
| 43 | 19 | Female | Employee | Spot | Immediately | Yes | No | Scooter | NA | R |
| 44 | 32 | Male | Employee | On the way | Immediately | Yes | No | Scooter | NA | P |
| 46 | 32 | Male | Student | Spot | Immediately | Yes | No | RX 100 | NA | R |
| 47 | 42 | Male | Employee | Spot | Immediately | Yes | No | Scooty | NA | R |
| 48 | 23 | Male | Mechanic | Spot | Immediately | Yes | No | Scooter | NA | R |
| 50 | 25 | Male | Student | On the way | Immediately | Yes | No | RX 100 | NA | R |

```
paste("People died under the influence of alcohol/drug and without wearing helmet:",nrow(no.helmet_alcohol))
```

```
## [1] "People died under the influence of alcohol/drug and without wearing helmet: 19"
```

# 5. How many of them have died between the age group of 31 - 40 years inspite of wearing helmet?

```
age_31_40 <- accident %>% filter(Age..yrs. %in% 31:40,Use.of.helmet=="Yes")

paste("Number of people died between the age group of 31 - 40 years inspite of wearing helmet:",nrow(age_31_40))
```

```
## [1] "Number of people died between the age group of 31 - 40 years inspite of wearing helmet: 2"
```

# 6. What is the ratio of male and female who died on the spot?

```
accident_spot_Gender <- accident %>% filter(Site.of.death=="Spot") %>%  group_by(Gender) %>% summarise(Total_died_on_spot=n
()) %>% arrange(-Total_died_on_spot)

kable(accident_spot_Gender)
```

| Gender | Total_died_on_spot |
|---|---|
| Male | 14 |
| Female | 5 |

```
paste("Ratio between male and female who died on the spot ->",accident_spot_Gender$Total_died_on_spot[1],":",accident_spot_G
ender$Total_died_on_spot[2])
```

```
## [1] "Ratio between male and female who died on the spot -> 14 : 5"
```

```
paste("OR",accident_spot_Gender$Total_died_on_spot[1]/accident_spot_Gender$Total_died_on_spot[2])
```

```
## [1] "OR 2.8"
```

# 7. Whai is the medianal survival time and IQR of RTA victims who died in the hospital?

```
accident_died <- accident %>% filter(Site.of.death=="Hospital")
accident_died$Time.of.death <- as.numeric(trimws(gsub("[a-z|A-Z]*","",accident_died$Time.of.death)))

summ_died <- summary(accident_died$Time.of.death)

paste("For RTA victims who died in the hospital ")
```

```
## [1] "For RTA victims who died in the hospital "
```

```
paste("Median:",summ_died[3],"IQR:",summ_died[4]-summ_died[2])
```

```
## [1] "Median: 72 IQR: 95.4583333333333"
```

# 8. How many of them have died in the hospital for whom wearing of helmet details not available? Do you think these people could have wore helmet at the time of accident?

```
kable(accident %>% filter(Site.of.death == "Hospital" & Use.of.helmet!= "Yes" &  Use.of.helmet  != "No") %>% group_by( Time.of.death ) %>% summarise(Death_Count = n()))
```

| Time.of.death | Death_Count |
|---|---|
| After 144 hours | 1 |
| After 260 hours | 1 |
| After 36 hours | 1 |
| After 5 hours | 2 |
| After 72 hours | 1 |

```
## Since the death of people without helmet was on spot, it is said that these passengers could have worn helmet during accident.
```

# 9. Whai is the medianal period of coma (hrs) and IQR of RTA victims who died in the hospital?

```
accident_died <- accident %>% filter(Site.of.death=="Hospital")

summ_died <- summary(accident_died$Period.of.coma.hrs)

paste("For RTA victims who died in the hospital ")
```

```
## [1] "For RTA victims who died in the hospital "
```

```
paste("Median:",summ_died[3],"IQR:",summ_died[4]-summ_died[2])
```

```
## [1] "Median: 72 IQR: 103.791666666667"
```

# 10. What is the ratio of students to bussinessmen who died in RTA under the influence of alcohol?

```
student_bussinesman <- accident %>% filter(Occupation %in% c("Student","Bussinesman"),Alco.Drug.etc=="Yes") %>% group_by(Occupation) %>% summarise(Total_died=n())


paste("Ratio between Student and Bussinesman who died on the spot ->",student_bussinesman$Total_died[2],":",student_bussinesman$Total_died[1])
```

```
## [1] "Ratio between Student and Bussinesman who died on the spot -> 6 : 5"
```

```
paste("OR",student_bussinesman$Total_died[2]/student_bussinesman$Total_died[1])
```

```
## [1] "OR 1.2"
```

# 11. State some of the hypothesis which may be formulated based on this data?

HYPOTHESIS POSSIBLE:

1.

H0<- Alcohol Consumption and Time of Death do not have a significant effect on each other

H1<- Alcohol Consumption and Time of Death do not have a significant effect on each other

2.

H0<-Type of vehicle and Period of Death (Coma Period) do not have a Significant effect on each other

H1<-Type of vehicle and Period of Death (Coma Period) have a Significant effect on each other

3.

H0<- Usage of Helment does not affect Time of Death

H1<- Usage of Helmet affects Time of Death

Testing Above Hypothesis

# 12. Is there sufficient data to test these hypothesis? If so use the available and test those hypothesis.

```
ch1<- chisq.test(accident$Time.of.death ,accident$Alco.Drug.etc)
ch1
```

```
##
##  Pearson's Chi-squared test
##
## data:  accident$Time.of.death and accident$Alco.Drug.etc
## X-squared = 17.928, df = 16, p-value = 0.3281
```

```
## INTERPRETATION:
#                  As the P value (0.3281) > P critical (0.05) Null Hypothesis is accepted and it is concluded that the cons
umption of Alcohol during Road Accident does affect the TIme of Death.
```

```
ch2<- chisq.test(accident$Type.of.vehicle,accident$Time.of.death )
ch2
```

```
##
##  Pearson's Chi-squared test
##
## data:  accident$Type.of.vehicle and accident$Time.of.death
## X-squared = 104.02, df = 112, p-value = 0.6921
```

```
## INTERPRETATION:
#                 As the P Value (0.6921) > P critical (0.05), Null Hypothesis is accepted and we can say thaat there is no
 significant relation in type of vehicle and time of death
```

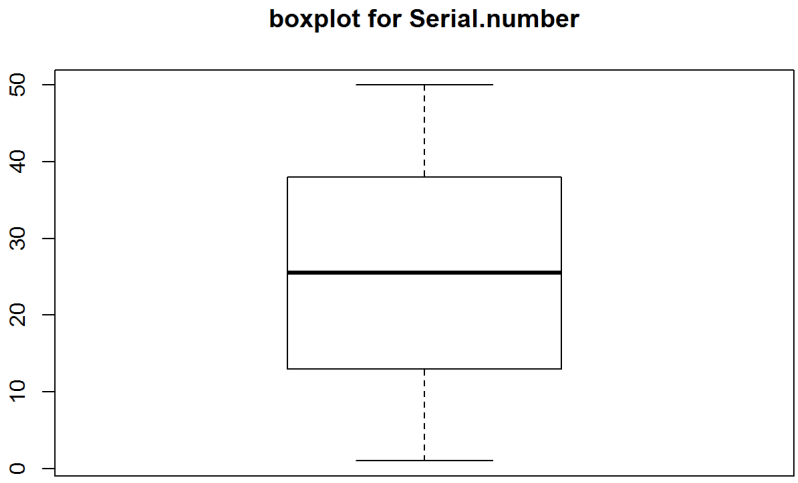
```
ch3<- chisq.test(accident$Use.of.helmet,accident$Time.of.death)
ch3
```

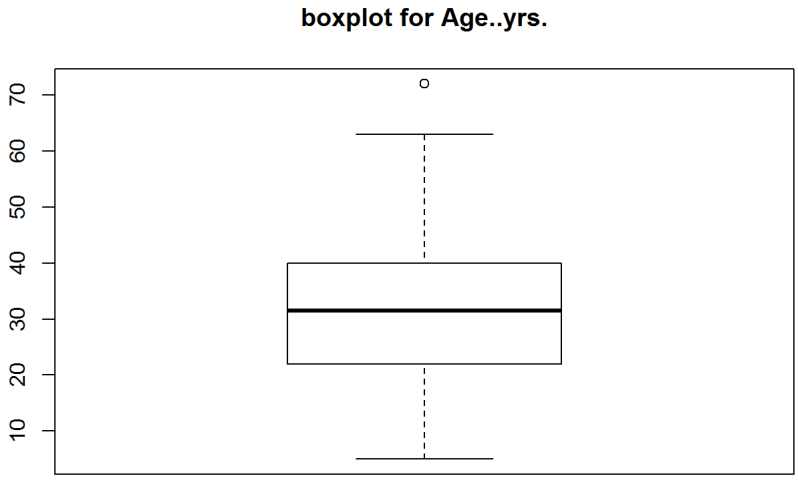
```
##
##  Pearson's Chi-squared test
##
## data:  accident$Use.of.helmet and accident$Time.of.death
## X-squared = 40.295, df = 32, p-value = 0.149
```

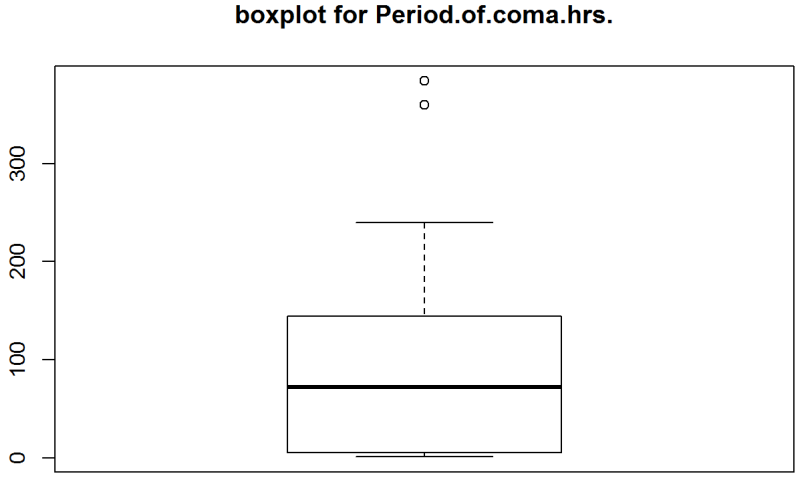
```
## INTERPRETATION:
#                As the P value (0.149) > P critical (0.05) Null Hypothesis is accepted and it is concluded that the Use of
 Helmet during Road Accident does affect the TIme of Death for the given sample.
```

# 13. Use appropriaate graphs to present the data

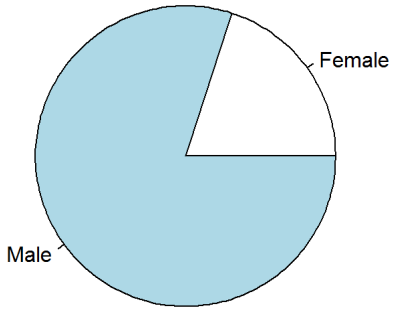
```
library(ggplot2)

#Converting all character columns into factor
accident <- as.data.frame(lapply(accident, function(x){if(length(unique(x))<=12){return(as.factor(x))}else {return(x)}}))


plot_graph <- function(database)
{
  num_col <- sapply(database, is.numeric)
  fac_col <- sapply(database,is.factor)

  for (i in names(database)[num_col]) {

     boxplot(database[,i],names = i)
     title(paste("boxplot for",i))

  }
  for (j in names(database)[fac_col]) {
  slices <- table(database[,j])
  lbls <- names(slices)
  Main <- paste("Pie Chart for",j)
  pie(slices, labels = lbls, main=Main)
  }

}
plot_graph(accident)
```


boxplot for Serial.number
0
10
20
30
40
50

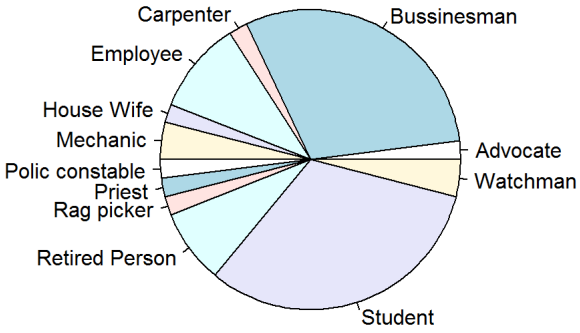

boxplot for Age..yrs.
10
20
30
40
50
60
70


boxplot for Period.of.coma.hrs.
0
100
200
300

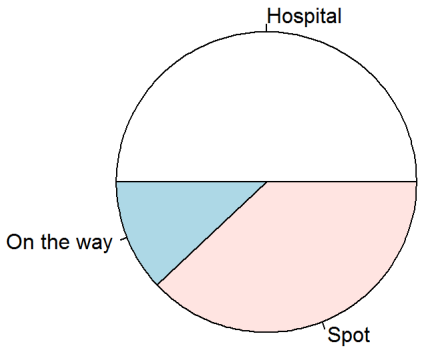
**Pie Chart for Gender**



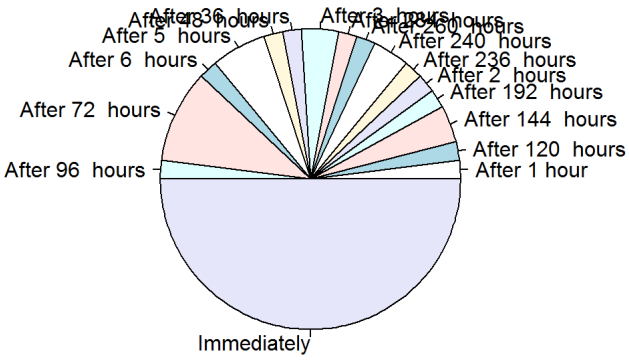

**Pie Chart for Occupation**



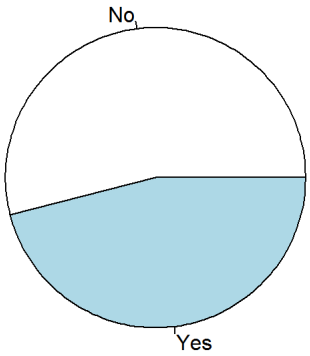

**Pie Chart for Site.of.death**



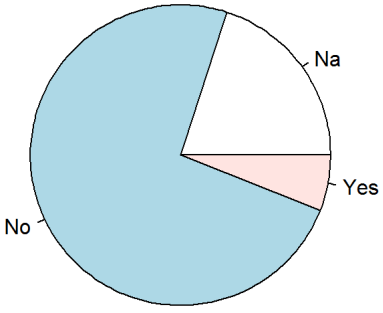
**Pie Chart for Time.of.death**



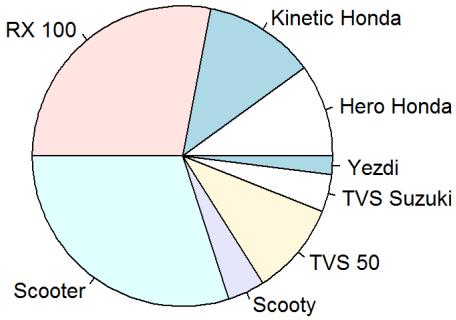

**Pie Chart for Alco.Drug.etc**




**Pie Chart for Use.of.helmet**



**Pie Chart for Type.of.vehicle**




**Pie Chart for Person.died**