R Programming CIA 2

Pranav - 1940223

2020-01-03

setwd("~/rPrograms/CIA 2 Dataset")  
mydata1 = read.csv("NTCA - TIGERNET.csv", header = TRUE)  
  
  
#THE DATA SET  
head(mydata1, 6)

## Date Location Inside...Outside  
## 1 01/03/19 Pench Inside Tiger Reserve  
## 2 01/04/19 Maharashtra ,East Melgaht Outside Tiger Reserve  
## 3 01/05/19 Kanha Inside Tiger Reserve  
## 4 01/11/19 Maharashtra,Amravati Outside Tiger Reserve  
## 5 01/19/2019 Periyar Inside Tiger Reserve  
## 6 01/19/2019 Kanha Inside Tiger Reserve  
## Mortality...Seizure State Sex Age  
## 1 Mortality Maharashtra <NA> NA  
## 2 Mortality Maharashtra <NA> 2.3  
## 3 Mortality Madhya Pradesh <NA> 7.8  
## 4 Mortality Maharashtra <NA> 4.0  
## 5 Mortality Kerala Male 10.4  
## 6 Mortality Madhya Pradesh Male 2.0

#INTRODUCTORY INFORMATION  
#This data set is about the observed mortalities and seizures of tigers  
#in the different stats of India, in the year 2019.  
  
sum(!is.na(unique(mydata1$Location)))

## [1] 59

sum(!is.na(unique(mydata1$State)))

## [1] 16

#The data set takes its data from   
#59 different locations  
#16 different states  
  
  
#AGE RELATED  
total = sum(!is.na(mydata1$Date))  
i = 1  
tenPlusCount = 0  
fivePlusCount = 0  
onePlusCount = 0  
oneMinusCount = 0  
  
while(i <= total)  
{  
 if(!is.na(mydata1$Age[i]))  
 {  
 if(mydata1$Age[i] >= 10 && mydata1$Mortality...Seizure == "Mortality")  
 {  
 tenPlusCount = tenPlusCount + 1  
 } else if(mydata1$Age[i] >= 5 && mydata1$Mortality...Seizure == "Mortality")  
 {  
 fivePlusCount = fivePlusCount + 1  
 } else if(mydata1$Age[i] >= 1 && mydata1$Mortality...Seizure == "Mortality")  
 {  
 onePlusCount = onePlusCount + 1  
 } else if(mydata1$Age[i] < 1 && mydata1$Mortality...Seizure == "Mortality")  
 {  
 oneMinusCount = oneMinusCount + 1  
 }  
 }  
 i = i + 1  
}  
#Number of tigers 10 years or older'  
tenPlusCount

## [1] 4

#Number of tigers 5 years or older but less than 10 years  
fivePlusCount

## [1] 12

#Number of tigers 1 year or older but less than 5 years  
onePlusCount

## [1] 18

#Number of tigers less than 1 year old  
oneMinusCount

## [1] 2

#Number of tigers with age unknown  
sum(is.na(mydata1$Age))

## [1] 54

#With the above results, we see that a majority of the tigers' ages are unknown.  
#However, going with what is known, we see maximum mortality in adult tigers from 1 to 5 years old.  
#The high mortality among adult tigers may suggest...  
  
#1. Higher fatal confrontations between tigers of   
# these ages or between humans and tigers of these ages  
#2. Greater number of tigers in this age group  
  
  
#GRAPHS

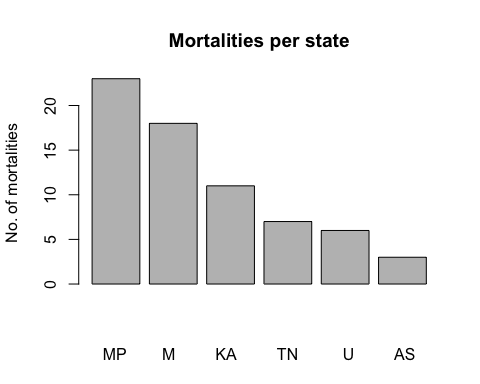
#Number of mortalities per state  
MP\_m = sum(mydata1$Mortality...Seizure == "Mortality" & mydata1$State == "Madhya Pradesh")  
M\_m = sum(mydata1$Mortality...Seizure == "Mortality" & mydata1$State == "Maharashtra")  
KA\_m = sum(mydata1$Mortality...Seizure == "Mortality" & mydata1$State == "Karnataka")  
TN\_m = sum(mydata1$Mortality...Seizure == "Mortality" & mydata1$State == "Tamil Nadu")  
U\_m = sum(mydata1$Mortality...Seizure == "Mortality" & mydata1$State == "Uttarakhand")  
AS\_m = sum(mydata1$Mortality...Seizure == "Mortality" & mydata1$State == "Assam")  
  
data1 = c(MP\_m, M\_m, KA\_m, TN\_m, U\_m, AS\_m)  
data1

## [1] 23 18 11 7 6 3

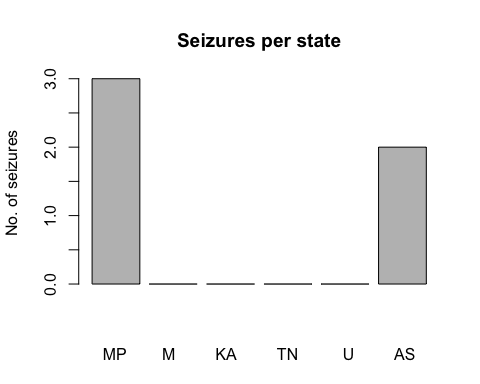
states = c("MP", "Maharashtra", "Karnataka", "Tamil Nadu", "Uttarakhand", "Assam")  
  
MP\_s = sum(mydata1$Mortality...Seizure == "Seizure" & mydata1$State == "Madhya Pradesh")  
M\_s = sum(mydata1$Mortality...Seizure == "Seizure" & mydata1$State == "Maharashtra")  
KA\_s = sum(mydata1$Mortality...Seizure == "Seizure" & mydata1$State == "Karnataka")  
TN\_s = sum(mydata1$Mortality...Seizure == "Seizure" & mydata1$State == "Tamil Nadu")  
U\_s = sum(mydata1$Mortality...Seizure == "Seizure" & mydata1$State == "Uttarakhand")  
AS\_s = sum(mydata1$Mortality...Seizure == "Seizure" & mydata1$State == "Assam")  
  
data2 = c(MP\_s, M\_s, KA\_s, TN\_s, U\_s, AS\_s)  
data2

## [1] 3 0 0 0 0 2

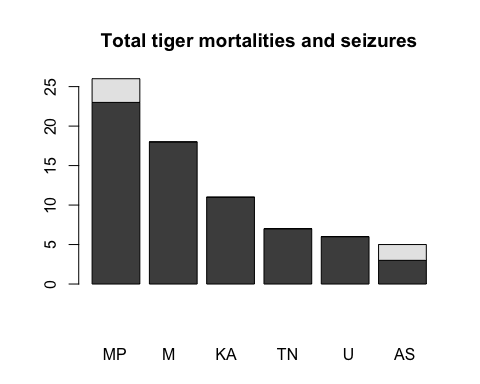
#Mortalities per state  
barplot(main = "Mortalities per state", data1, xlab = "MP M KA TN U AS", ylab = "No. of mortalities")



#Seizures per state  
barplot(main = "Seizures per state", data2, xlab = "MP M KA TN U AS", ylab = "No. of seizures")

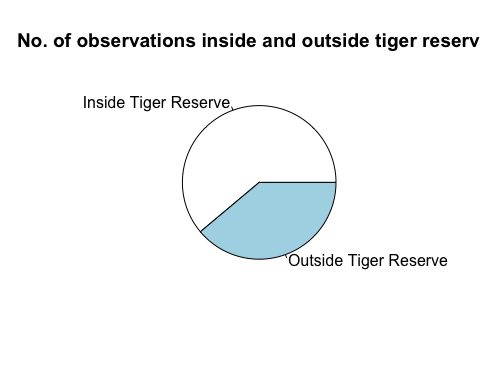


#Seizures are very few overall.  
#They are primarily in MP and Assam  
  
#Seizures and mortalities per state  
X = matrix(c(data1, data2), nrow = 2, byrow = TRUE)  
barplot(main = "Total tiger mortalities and seizures", X, xlab = "MP M KA TN U AS")



#With this graph, we can see that  
#1. Maximum observations, mortalities and seizures are in MP  
#2. Assam has the highest seizure to mortality ratio,   
# meaning a larger portion of observed tigers were   
# seized rather than found dead  
#3. No other state has records of seizures

a = sum(mydata1$Inside...Outside == "Inside Tiger Reserve" | mydata1$Inside...Outside == "Inside Tiger Reserve")  
b = sum(mydata1$Inside...Outside == "Outside Tiger Reserve" | mydata1$Inside...Outside == "Outside tiger Reserve")  
pie(main = "No. of observations inside and outside tiger reserves", c(a, b), labels = c("Inside Tiger Reserve", "Outside Tiger Reserve"))



#A majority of observations were taken inside tiger reserves  
#However, this majority is not significant, as a considerable number of observations are from outside tiger reserves as well  
  
  
#STRUCTURE  
str(mydata1)

## 'data.frame': 90 obs. of 7 variables:  
## $ Date : Factor w/ 75 levels "01/03/19","01/04/19",..: 1 2 3 4 5 5 6 7 8 9 ...  
## $ Location : Factor w/ 59 levels "Andhra Pradesh,Nandyal Division, Compt No. 525",..: 39 29 13 30 42 13 10 23 44 47 ...  
## $ Inside...Outside : Factor w/ 4 levels "Inside Tiger Reserve",..: 2 4 2 4 2 2 2 4 2 2 ...  
## $ Mortality...Seizure: Factor w/ 2 levels "Mortality","Seizure": 1 1 1 1 1 1 1 1 1 1 ...  
## $ State : Factor w/ 16 levels "Andhra Pradesh",..: 9 9 8 9 7 8 15 8 11 12 ...  
## $ Sex : Factor w/ 2 levels "Female","Male": NA NA NA NA 2 2 2 NA NA NA ...  
## $ Age : num NA 2.3 7.8 4 10.4 2 NA NA NA NA ...

#This data set is largely comprised of non-numeric fields.  
#There are large numbers of unknown values in the fields 'Sex' and 'Age'.  
#There is a considerable amount of repetition in the data, since thought the data set is large, the unique values are much lesser  
  
  
#SUMMARY  
summary(mydata1)

## Date Location Inside...Outside  
## 07/08/19 : 3 Bandhavgarh: 6 Inside Tiger Reserve: 1   
## 01/05/19 : 2 Bandipur : 4 Inside Tiger Reserve :54   
## 01/19/2019: 2 Jim Corbett: 4 Outside tiger Reserve: 1   
## 01/24/2019: 2 Kanha : 4 Outside Tiger Reserve:34   
## 02/26/2019: 2 Nagarhole : 4   
## 02/28/2019: 2 Pench : 4   
## (Other) :77 (Other) :64   
## Mortality...Seizure State Sex Age   
## Mortality:82 Madhya Pradesh:26 Female: 7 Min. : 0.300   
## Seizure : 8 Maharashtra :18 Male :38 1st Qu.: 2.000   
## Karnataka :11 NA's :45 Median : 4.400   
## Tamil Nadu : 7 Mean : 4.856   
## Uttarakhand : 6 3rd Qu.: 7.200   
## Assam : 5 Max. :12.000   
## (Other) :17 NA's :54

#There are 90 observations and 7 variables  
#The data set includes a very diverse range of ages  
#Maximum observations are from Madhya Pradesh  
#Most observed tigers are male  
#Out of the 16 states in the data set, only 6 have a significant amount of data  
#There is a considerable amount of unavailable data in the fields 'Sex' and 'Age'