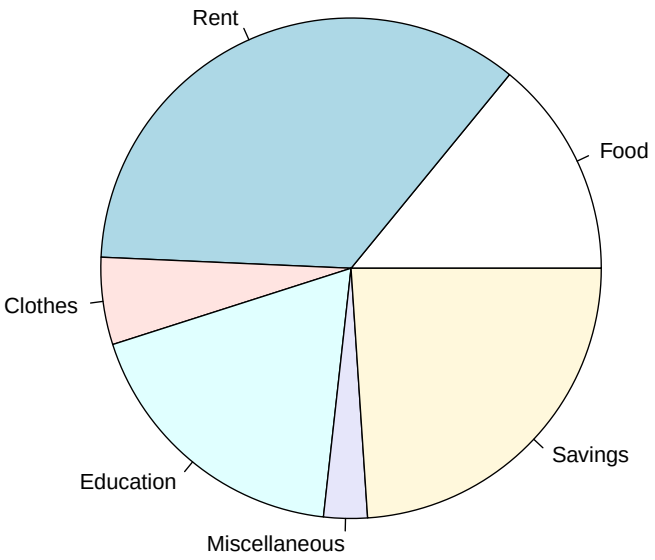
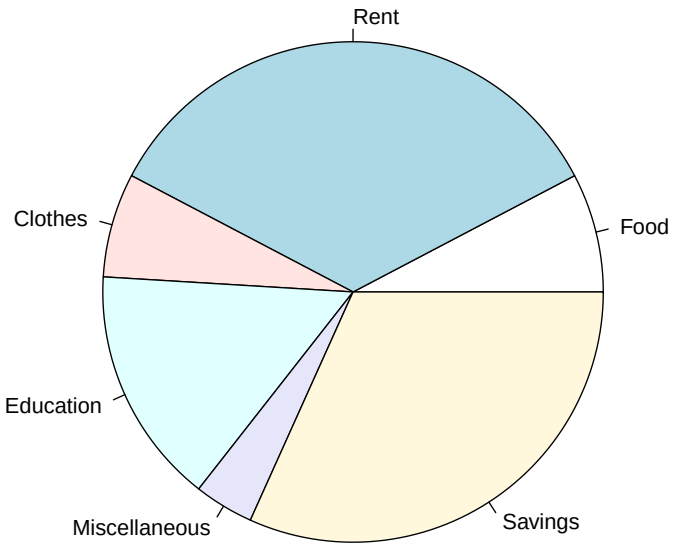
Name: Mahaprasad Mohanty
Registration: 24MDT0061
Course: Statistical Inference
Code: PMDS503P Slot: L33+L34

1 Question 1

```
library(ggplot2)
data <- data.frame(
  Commodity = c("Food", "Rent", "Clothes", "Education", "Miscellaneous", "Savings"),
  FamilyA = c(10, 25, 4, 13, 2, 17),
  FamilyB = c(8, 36, 7, 16, 4, 33)
)
pie(data$FamilyA, labels = data$Commodity, main = "Expenditure of Family A")
```



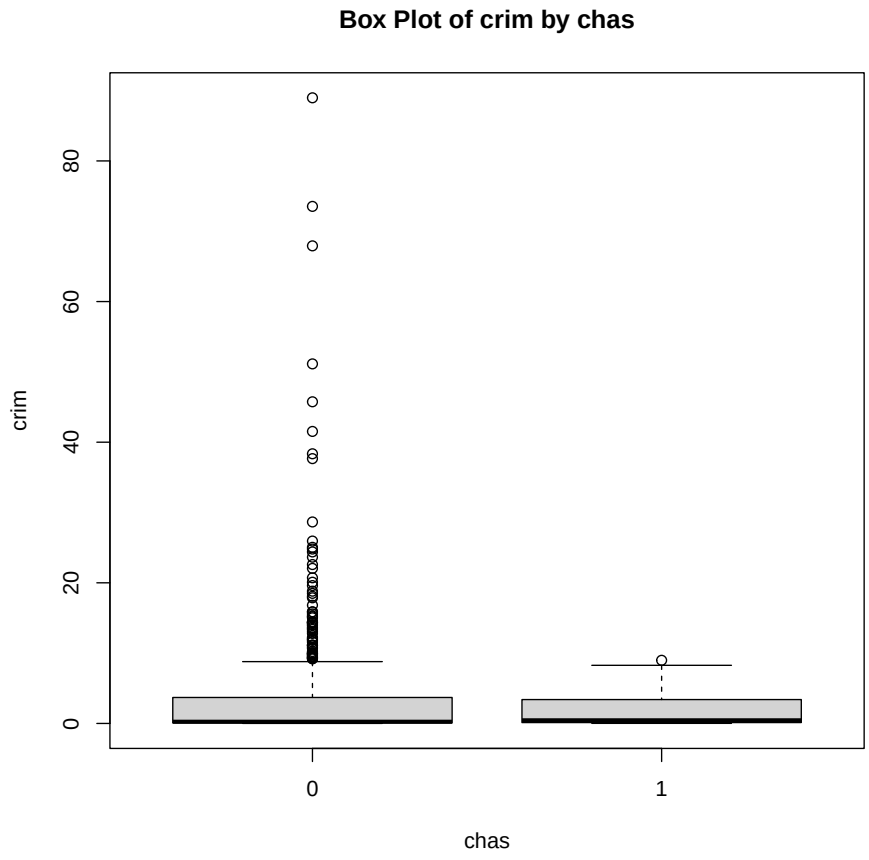
```r
pie(data$FamilyB, labels = data$Commodity, main = "Expenditure of Family B")
```



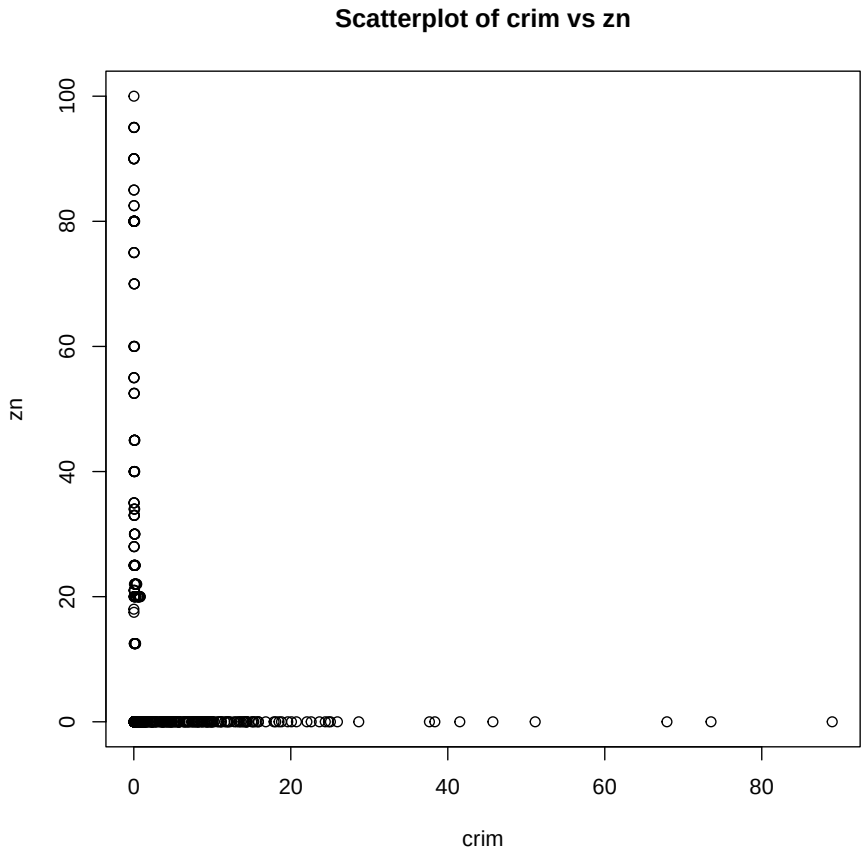
2 Question 2

```r
library(MASS)
library(ggplot2)
data(Boston)

# (i) Display the number of variables in the dataset
num_variables <- ncol(Boston)
cat("Number of variables in the dataset:", num_variables, "\n")

## Number of variables in the dataset: 14
```

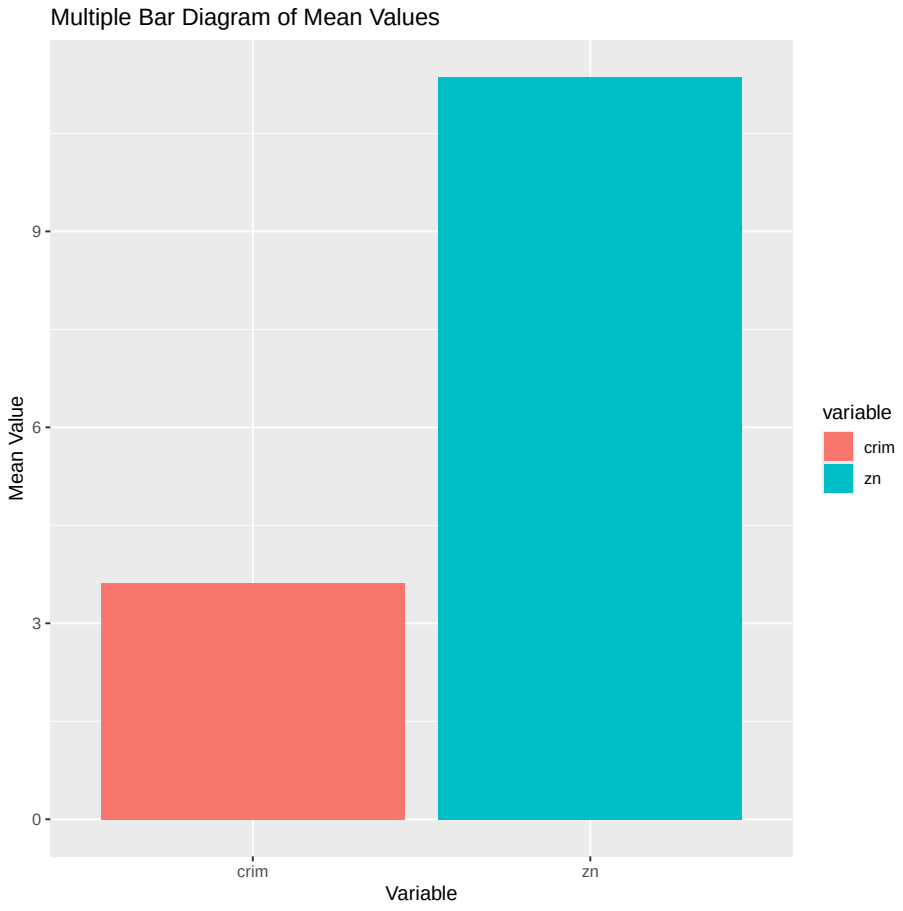
```r
# (ii) Draw a box plot for any two variables
boxplot(Boston$crim ~ Boston$chas,
        main = "Box Plot of crim by chas",
        xlab = "chas", ylab = "crim")
```




```r
# (iii) Scatterplot for any two variables
plot(Boston$crim, Boston$zn,
     main = "Scatterplot of crim vs zn",
     xlab = "crim", ylab = "zn")
```




```
# (iv) Multiple bar diagram (with different colors)
bar_data <- data.frame(
  variable = c("crim", "zn"),
  value = c(mean(Boston$crim), mean(Boston$zn))
)
ggplot(bar_data, aes(x = variable, y = value, fill = variable)) +
  geom_bar(stat = "identity") +
  labs(title = "Multiple Bar Diagram of Mean Values",
       x = "Variable", y = "Mean Value")
```




```
# (v) Observations
# - The Boston dataset has 14 variables.\\
# - The boxplot shows that crime rates are very different for 'crim' and 'chas'. There are s
# - From the scatter plot, there appears to be no linear relationship between 'crim' and 'zn
# - From the bar diagram, we can infer that the mean value of 'zn' is much higher compared t
```

3 Question 3

```
library(dplyr)

##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:MASS':
##
##     select
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(xtable)
data <- read.csv('D:\\Code-stuff\\Stat-Inference\\exam-data.csv')

# Calculating all summary statistics

# Mean
mean_CAT1 <- mean(data$CAT1)
mean_CAT2 <- mean(data$CAT2)
mean_DA <- mean(data$DA)
mean_FAT <- mean(data$FAT)
mean_QUIZ1 <- mean(data$QUIZ1)
mean_QUIZ2 <- mean(data$QUIZ2)

# Median
median_CAT1 <- median(data$CAT1)
median_CAT2 <- median(data$CAT2)
median_DA <- median(data$DA)
median_FAT <- median(data$FAT)
median_QUIZ1 <- median(data$QUIZ1)
median_QUIZ2 <- median(data$QUIZ2)

# Minimum and Maximum
min_CAT1 <- min(data$CAT1)
max_CAT1 <- max(data$CAT1)

min_CAT2 <- min(data$CAT2)
max_CAT2 <- max(data$CAT2)

min_DA <- min(data$DA)
max_DA <- max(data$DA)

min_FAT <- min(data$FAT)
max_FAT <- max(data$FAT)

min_QUIZ1 <- min(data$QUIZ1)
max_QUIZ1 <- max(data$QUIZ1)
```

```r
min_QUIZ2 <- min(data$QUIZ2)
max_QUIZ2 <- max(data$QUIZ2)

# Standard deviation
sd_CAT1 <- sd(data$CAT1)
sd_CAT2 <- sd(data$CAT2)
sd_DA <- sd(data$DA)
sd_FAT <- sd(data$FAT)
sd_QUIZ1 <- sd(data$QUIZ1)
sd_QUIZ2 <- sd(data$QUIZ2)

# Printing all

cat("Summary Statistics:\n")

## Summary Statistics:

summary_stats <- data.frame(
  Variable = c("CAT1", "CAT2", "DA", "FAT", "QUIZ1", "QUIZ2"),
  Mean = c(mean_CAT1, mean_CAT2, mean_DA, mean_FAT, mean_QUIZ1, mean_QUIZ2),
  Median = c(median_CAT1, median_CAT2, median_DA, median_FAT, median_QUIZ1, median_QUIZ2),
  Min = c(min_CAT1, min_CAT2, min_DA, min_FAT, min_QUIZ1, min_QUIZ2),
  Max = c(max_CAT1, max_CAT2, max_DA, max_FAT, max_QUIZ1, max_QUIZ2),
  SD = c(sd_CAT1, sd_CAT2, sd_DA, sd_FAT, sd_QUIZ1, sd_QUIZ2)
)

# printing the table of summary statistics using xtable library
```

Variable	Mean	Median	Min	Max	SD
CAT1	22.35	21.00	5	45	9.44
CAT2	31.53	34.00	0	49	10.67
DA	10.00	10.00	10	10	0.00
FAT	60.44	60.00	0	94	20.90
QUIZ1	12.68	12.00	6	20	2.96
QUIZ2	13.74	14.00	0	20	3.38

Table 1: Summary Statistics