Assignment1

DEEPAK MACHKURI

2023-09-11

```
knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

##DOWNLOADED THE DATASET
https://www.kaggle.com/datasets/smitisinghal/wholesale-customers-data

```
##imported the dataset
Wholesale_data <- read.csv("C:/Users/durga/OneDrive/Desktop/data.csv", header
= TRUE, sep = "," )

##DESCRIPTIVE STATISTICS
head (Wholesale_data)
```

```
##   Channel Region Fresh Milk Grocery Frozen Detergents_Paper Delicassen
## 1       2      3 12669 9656    7561    214             2674       1338
## 2       2      3  7057 9810    9568   1762             3293       1776
## 3       2      3  6353 8808    7684   2405             3516       7844
## 4       1      3 13265 1196    4221   6404              507       1788
## 5       2      3 22615 5410    7198   3915             1777       5185
## 6       2      3  9413 8259    5126    666             1795       1451
```

```
summary (Wholesale_data)
```

```
##     Channel          Region          Fresh             Milk
##  Min.   :1.000   Min.   :1.000   Min.   :     3   Min.   :   55
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:  3128   1st Qu.: 1533
##  Median :1.000   Median :3.000   Median :  8504   Median : 3627
##  Mean   :1.323   Mean   :2.543   Mean   : 12000   Mean   : 5796
##  3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.: 16934   3rd Qu.: 7190
##  Max.   :2.000   Max.   :3.000   Max.   :112151   Max.   :73498
##     Grocery          Frozen        Detergents_Paper    Delicassen
##  Min.   :    3   Min.   :   25.0   Min.   :    3.0   Min.   :    3.0
##  1st Qu.: 2153   1st Qu.:  742.2   1st Qu.:  256.8   1st Qu.:  408.2
##  Median : 4756   Median : 1526.0   Median :  816.5   Median :  965.5
##  Mean   : 7951   Mean   : 3071.9   Mean   : 2881.5   Mean   : 1524.9
##  3rd Qu.:10656   3rd Qu.: 3554.2   3rd Qu.: 3922.0   3rd Qu.: 1820.2
##  Max.   :92780   Max.   :60869.0   Max.   :40827.0   Max.   :47943.0
```

```
##DESCRIPTIVE STATISTICS FOR QUANTITATIVE VARIABLES
quantitative_variables = c("Milk" , "Grocery")
summary_variables = summary(Wholesale_data[quantitative_variables])
print(summary_variables)
```

```
##       Milk          Grocery
##  Min.   :   55   Min.   :    3
##  1st Qu.: 1533   1st Qu.: 2153
##  Median : 3627   Median : 4756
##  Mean   : 5796   Mean   : 7951
##  3rd Qu.: 7190   3rd Qu.:10656
##  Max.   :73498   Max.   :92780
```

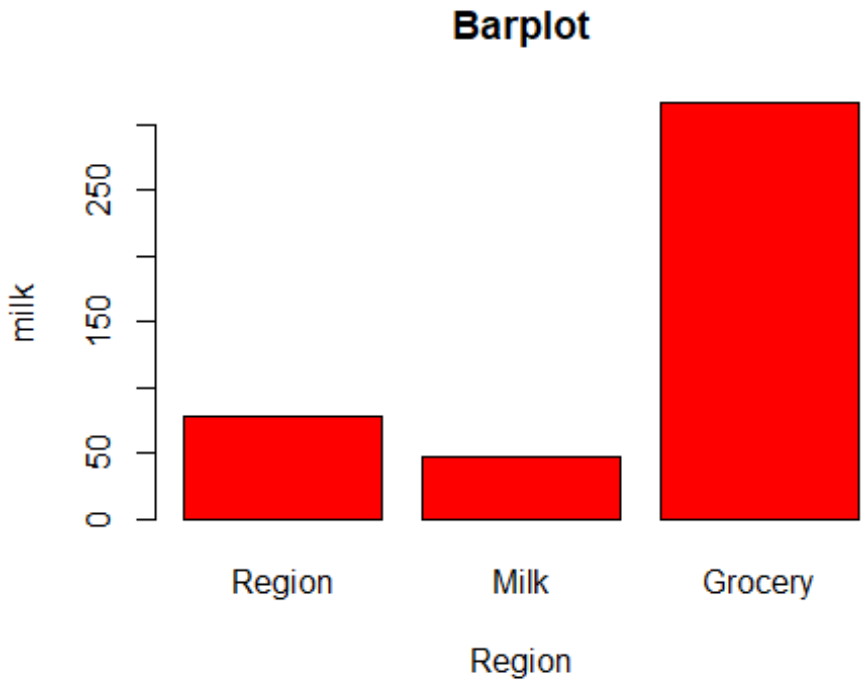
```r
##transformation
Transformation = Wholesale_data$Milk/100
summary(Transformation)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.55   15.33   36.27   57.96   71.90  734.98
```

```r
##Bar Plot
x=table(Wholesale_data$Region)
print(x)
```

```
##
##   1   2   3
##  77  47 316
```

```r
barplot.default(x,names.arg = c("Region", "Milk", "Grocery"),
xlab = "Region", ylab = "milk",
col = "Red", main = "Barplot")
```



```r
##Scatter Plot
##Scatter plot for milk consumption based on Region
plot(x = Wholesale_data$Milk, y = Wholesale_data$Region,
xlab = "Milk",
ylab = "Region",
xlim = c(0,3000),
ylim = c(0, 5000),
main = "Scatterplot"
)
```