Analysis of Customer Data in R by Hemani Panchmatiya

2025-09-07

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

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
customers <- read.csv("customers[1].csv")
str(customers)
```

```
## 'data.frame':    200 obs. of  14 variables:
##  $ CustID       : int  1530016 1531136 1532160 1532307 1532356 1532387 1533017 1533561 1533697 15337
##  $ Sex          : chr  "Female" "Male" "Male" "Male" ...
##  $ Race         : chr  "Black" "White" "Black" "White" ...
##  $ BirthDate    : chr  "12/16/1986" "5/9/1993" "5/22/1966" "9/16/1964" ...
##  $ College      : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ HouseholdSize: int  5 5 2 4 5 2 3 5 3 2 ...
##  $ ZipCode      : int  90047 90026 90027 90029 90017 90028 90063 90050 90038 90057 ...
##  $ Income       : int  53000 94000 64000 60000 47000 67000 84000 76000 42000 71000 ...
##  $ Spending2017 : int  287 1227 523 516 555 631 229 1474 300 925 ...
##  $ Spending2018 : int  241 843 719 582 845 452 153 1079 247 708 ...
##  $ NumOfOrders  : int  3 12 9 13 7 9 2 23 3 4 ...
##  $ DaysSinceLast: int  101 262 122 129 97 150 144 122 238 171 ...
##  $ Satisfaction : chr  "Very Dissatisfied" "Neutral" "Very Satisfied" "Very Dissatisfied" ...
##  $ Channel      : chr  "SM" "TV" "TV" "SM" ...
```

```
summary(customers[, c("HouseholdSize", "Income", "Spending2017", "Spending2018", "NumOfOrders", "DaysSi
```

```
##  HouseholdSize      Income        Spending2017     Spending2018
##  Min.   :1.00   Min.   : 31000   Min.   :  42.0   Min.   :  50.0
##  1st Qu.:2.00   1st Qu.: 54000   1st Qu.: 339.0   1st Qu.: 383.8
##  Median :3.00   Median : 69000   Median : 601.5   Median : 662.0
##  Mean   :3.11   Mean   : 72730   Mean   : 658.8   Mean   : 659.6
##  3rd Qu.:4.00   3rd Qu.: 91000   3rd Qu.: 911.5   3rd Qu.: 962.2
##  Max.   :5.00   Max.   :167000   Max.   :1851.0   Max.   :1250.0
##   NumOfOrders    DaysSinceLast
##  Min.   : 1.00   Min.   :  6.0
##  1st Qu.: 5.00   1st Qu.: 97.0
##  Median : 9.50   Median :180.0
##  Mean   :10.55   Mean   :182.9
##  3rd Qu.:15.00   3rd Qu.:267.0
##  Max.   :31.00   Max.   :360.0
```

```r
table(customers$Sex)
```

```
##
## Female   Male
##     73    127
```

```r
table(customers$Race)
```

```
##
##  American Indian            Asian            Black         Hispanic
##                5               15               57               41
## Pacific Islander            White
##                3               79
```

```r
table(customers$College)
```

```
##
##  No Yes
##  32 168
```

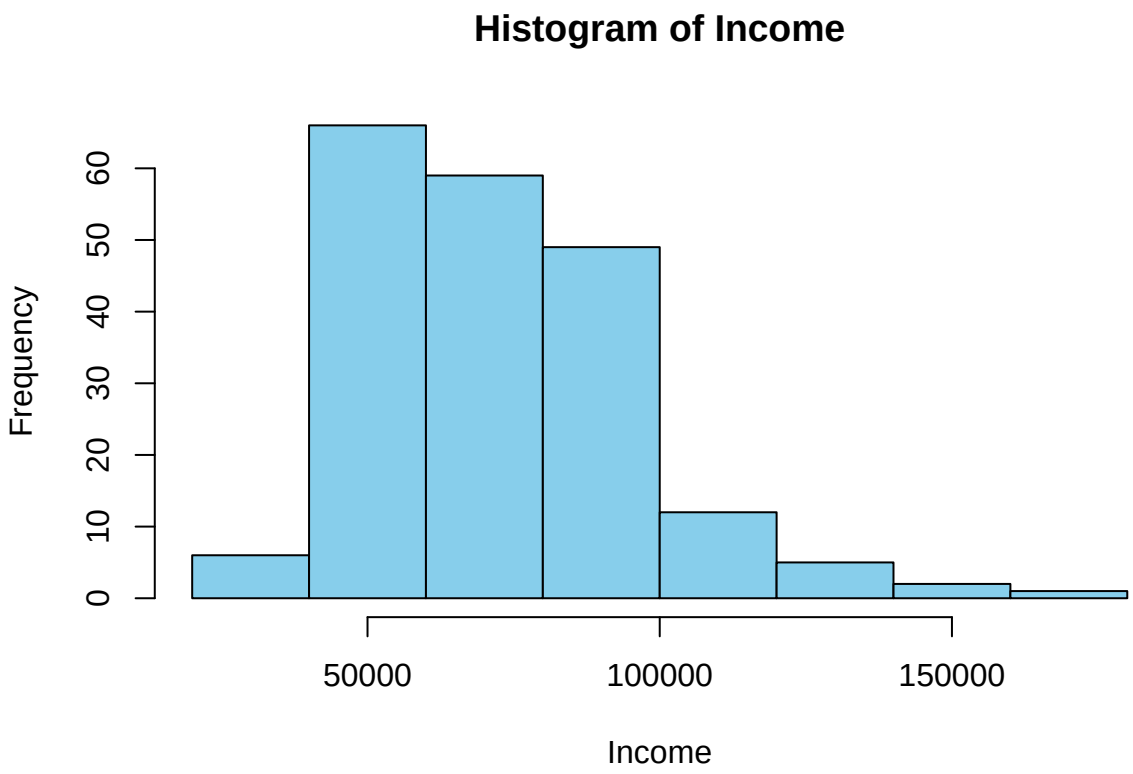
```r
table(customers$Satisfaction)
```

```
##
##               Neutral Somewhat Dissatisfied    Somewhat Satisfied
##                    35                    17                    94
##     Very Dissatisfied        Very Satisfied
##                     9                    45
```

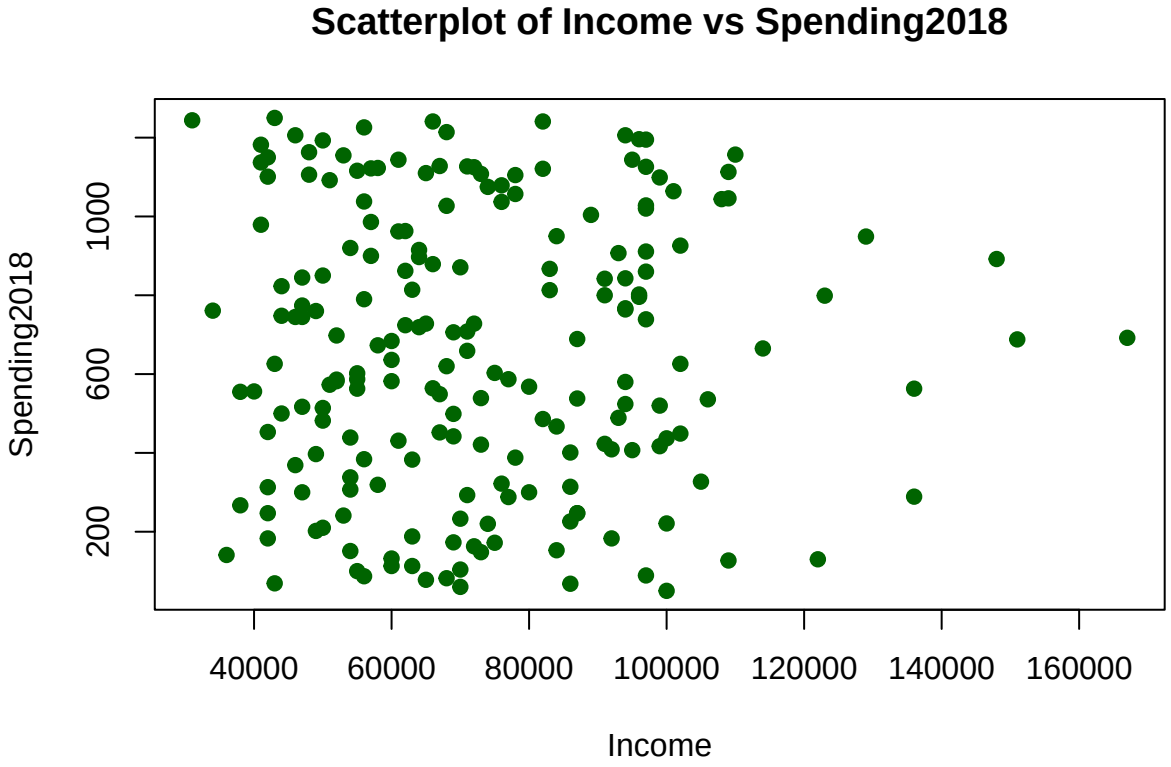
```r
table(customers$Channel)
```

```
##
## Referral       SM       TV      Web
##       38       39       57       66
```

```r
customers$SatBinary <- ifelse(customers$Satisfaction %in% c("Very Satisfied", "Somewhat Satisfied"),
                              "Satisfied", "Not Satisfied")
hist(customers$Income,
     main = "Histogram of Income",
     xlab = "Income", col = "skyblue", border = "black")
```




```
plot(customers$Income, customers$Spending2018,
     main = "Scatterplot of Income vs Spending2018",
     xlab = "Income", ylab = "Spending2018", pch = 19, col = "darkgreen")
```



#The End

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.