MichaelBasta_Assignment 1

Michael Basta

2022-09-09

```r
# Data used is the build in dataset "Credit" under the library ISLR
library(ISLR)
CreditData <- Credit
summary(Credit)
```

```
##        ID              Income            Limit            Rating
##  Min.   :  1.0   Min.   : 10.35   Min.   :  855   Min.   : 93.0
##  1st Qu.:100.8   1st Qu.: 21.01   1st Qu.: 3088   1st Qu.:247.2
##  Median :200.5   Median : 33.12   Median : 4622   Median :344.0
##  Mean   :200.5   Mean   : 45.22   Mean   : 4736   Mean   :354.9
##  3rd Qu.:300.2   3rd Qu.: 57.47   3rd Qu.: 5873   3rd Qu.:437.2
##  Max.   :400.0   Max.   :186.63   Max.   :13913   Max.   :982.0
##      Cards            Age          Education        Gender      Student
##  Min.   :1.000   Min.   :23.00   Min.   : 5.00   Male :193   No :360
##  1st Qu.:2.000   1st Qu.:41.75   1st Qu.:11.00   Female:207   Yes: 40
##  Median :3.000   Median :56.00   Median :14.00
##  Mean   :2.958   Mean   :55.67   Mean   :13.45
##  3rd Qu.:4.000   3rd Qu.:70.00   3rd Qu.:16.00
##  Max.   :9.000   Max.   :98.00   Max.   :20.00
##  Married              Ethnicity       Balance
##  No :155   African American: 99   Min.   :   0.00
##  Yes:245   Asian           :102   1st Qu.:  68.75
##            Caucasian       :199   Median : 459.50
##                                   Mean   : 520.01
##                                   3rd Qu.: 863.00
##                                   Max.   :1999.00
```

```r
## Descriptive Statistics

# Average Income in Thousands
meanIncome <- mean(CreditData$Income)
print(paste("The mean income is ", meanIncome))
```

```
## [1] "The mean income is  45.218885"
```

```r
# Measure of Dependence - Pearson Correlation
IncomeBalanceCor <- cor(CreditData$Income, CreditData$Balance, method = 'pearson')
print(paste("There's mild positive correlation between the income and the balance", IncomeBalanceCor))
```

```
## [1] "There's mild positive correlation between the income and the balance 0.463656457015757"
```

```r
IncomeCardsCor <- cor(CreditData$Income, CreditData$Cards, method = 'pearson')
print(paste("Here it seems like there is almost no correlation between Income and number of credit cards
```

```
## [1] "Here it seems like there is almost no correlation between Income and number of credit cards -0.0
```

```r
# Measure of Dispersion and Skewness
paste("Standard deviation of Income")
```

```
## [1] "Standard deviation of Income"
```

```r
sd(CreditData$Income)
```

```
## [1] 35.24427
```

```r
paste("Quantile of Credit Card Balance")
```

```
## [1] "Quantile of Credit Card Balance"
```

```r
quantile(CreditData$Balance)
```

```
##      0%     25%     50%     75%    100%
##    0.00   68.75  459.50  863.00 1999.00
```

```r
# Measure of Frequencies for Categorical Variables
table(CreditData$Gender)
```

```
##
##    Male Female
##     193    207
```

```r
table(CreditData$Ethnicity)
```

```
##
## African American            Asian       Caucasian
##               99              102             199
```

```r
# Cross classification counts for gender by Marital status
table(CreditData$Married, CreditData$Gender)
```

```
##
##        Male Female
##   No     76     79
##   Yes   117    128
```

```r
# Data Transformation for numeric variable : Normalization
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
Model_range_normalized <- preProcess(CreditData, method = "range")
Credit_Normalized = predict(Model_range_normalized, CreditData)
paste("Summary of the Income normalized variable")
```

```
## [1] "Summary of the Income normalized variable"
```

```
summary(Credit_Normalized$Income)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.06043 0.12912 0.19778 0.26728 1.00000
```

```
# Data Transformation for Categorical Variable : Dummy Variables
paste("Converting the Ethnicity categorical variable into individual binary variable")
```

```
## [1] "Converting the Ethnicity categorical variable into individual binary variable"
```

```
dummy_model <- dummyVars(~Ethnicity, data = CreditData)
head(predict(dummy_model, CreditData))
```

```
##   Ethnicity.African American Ethnicity.Asian Ethnicity.Caucasian
## 1                          0               0                   1
## 2                          0               1                   0
## 3                          0               1                   0
## 4                          0               1                   0
## 5                          0               0                   1
## 6                          0               0                   1
```
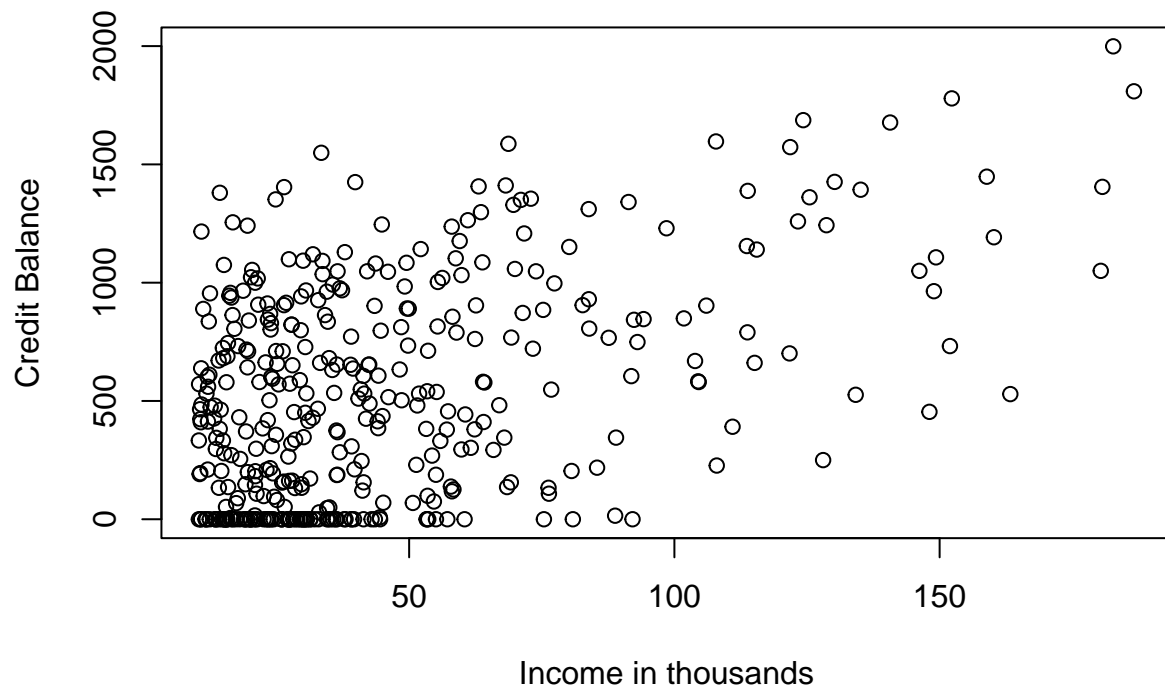
```
# Bar plots for Average Income by years of Education
library(ggplot2)
ggplot(CreditData, aes(x = factor(Education), y = Income)) + geom_bar(stat = "summary", fun = "mean")
```
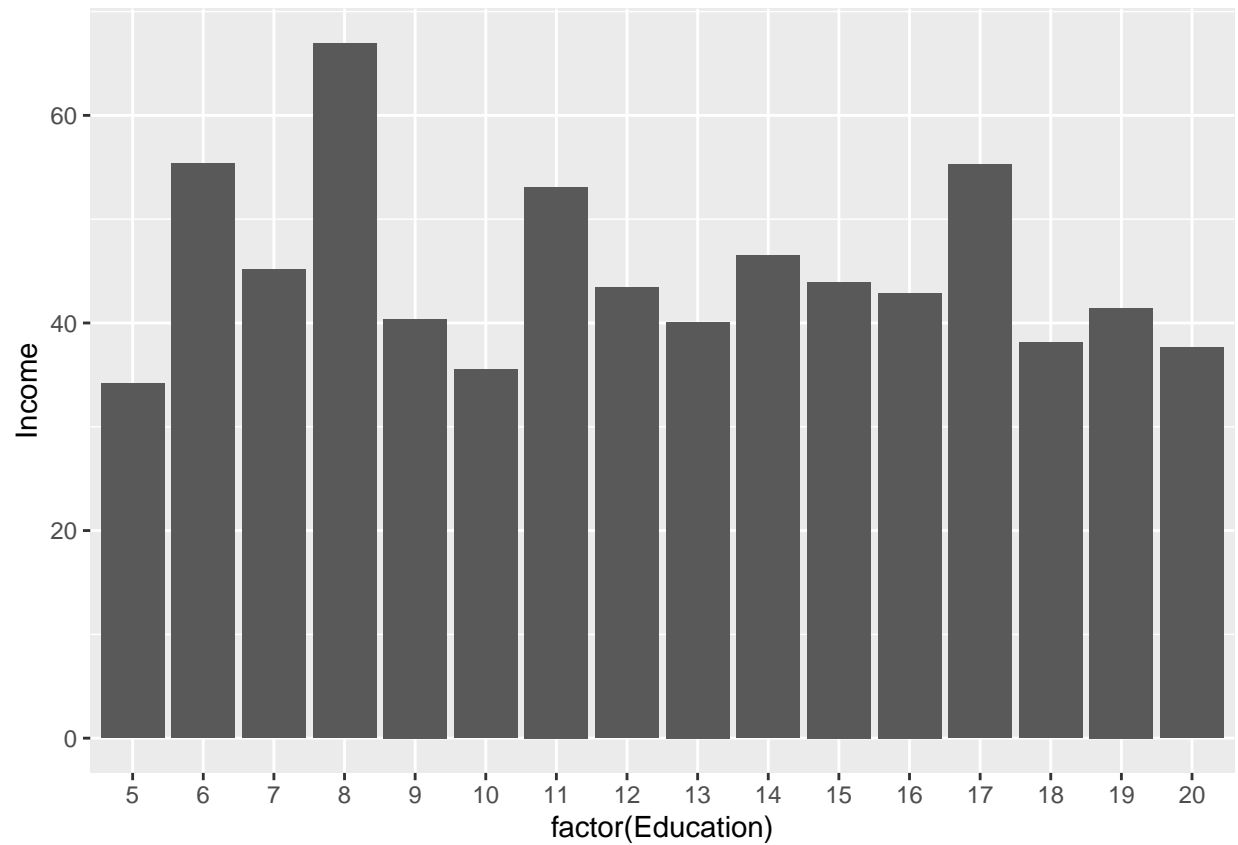
```r
# Scatter plot
plot(CreditData$Income, CreditData$Balance, main = "Scatterplot Income vs Balance", xlab = "Income in t
```

Scatterplot Income vs Balance


```
plot(CreditData$Age, CreditData$Income, main = "Scatterplot Age vs Income", xlab = "Age", ylab = "Income
```

Scatterplot Age vs Income

Income in thousands (y-axis) vs Age (x-axis)