BA64060_Assignment1

Ruthvick Bulagakula

2023-09-08

Dataset download

I have downloaded my dataset from Kaggle. This dataset explains about medical details. https://www.kaggle.com/datasets/mirichoi0218/insurance?resource=download

Import

I have imported dataset using read.csv

```
medical_read = read.csv("insurance.csv", header = TRUE, sep = ",")
```

Checking whether data has been read or not

```
head(medical_read)
```

```
##   age    sex    bmi children smoker    region   charges
## 1  19 female 27.900        0    yes southwest 16884.924
## 2  18   male 33.770        1     no southeast  1725.552
## 3  28   male 33.000        3     no southeast  4449.462
## 4  33   male 22.705        0     no northwest 21984.471
## 5  32   male 28.880        0     no northwest  3866.855
## 6  31 female 25.740        0     no southeast  3756.622
```

```
colnames(medical_read)
```

```
## [1] "age"      "sex"      "bmi"      "children" "smoker"   "region"   "charges"
```

Descriptive Statistics

Descriptive Statistics for quantitative variables

```
quant_abc = c("age", "bmi", "charges")
summary_quant = summary(medical_read[quant_abc])
print(summary_quant)
```

```
##       age             bmi           charges
##  Min.   :18.00   Min.   :15.96   Min.   : 1122
##  1st Qu.:27.00   1st Qu.:26.30   1st Qu.: 4740
##  Median :39.00   Median :30.40   Median : 9382
##  Mean   :39.21   Mean   :30.66   Mean   :13270
##  3rd Qu.:51.00   3rd Qu.:34.69   3rd Qu.:16640
##  Max.   :64.00   Max.   :53.13   Max.   :63770
```

Descriptive Statistics for categorical variables

```
cat_sr = c("smoker", "region")
summary_cat = sapply(medical_read[cat_sr], table)
print(summary_cat)
```

```
## $smoker
##
##   no  yes
## 1064  274
##
## $region
##
## northeast northwest southeast southwest
##       324       325       364       325
```

Transformation

```
transformation = medical_read$charges/1000
```

New dataset after transformation

```
sliced_data = medical_read[, c("age","bmi", "charges")]
sliced_data$tran_charges = transformation
head(sliced_data)
```

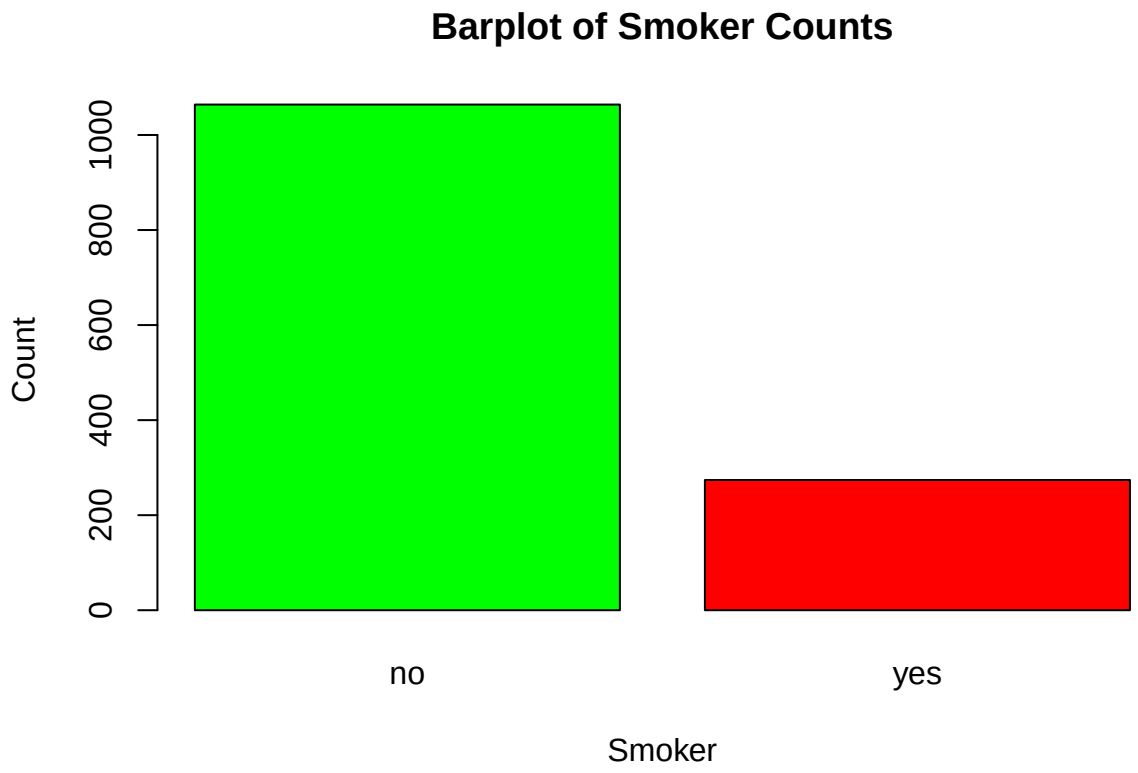
```
##   age    bmi   charges tran_charges
## 1  19 27.900 16884.924    16.884924
## 2  18 33.770  1725.552     1.725552
## 3  28 33.000  4449.462     4.449462
## 4  33 22.705 21984.471    21.984471
## 5  32 28.880  3866.855     3.866855
## 6  31 25.740  3756.622     3.756622
```

Plot

Barplot for smoker and non-smokers

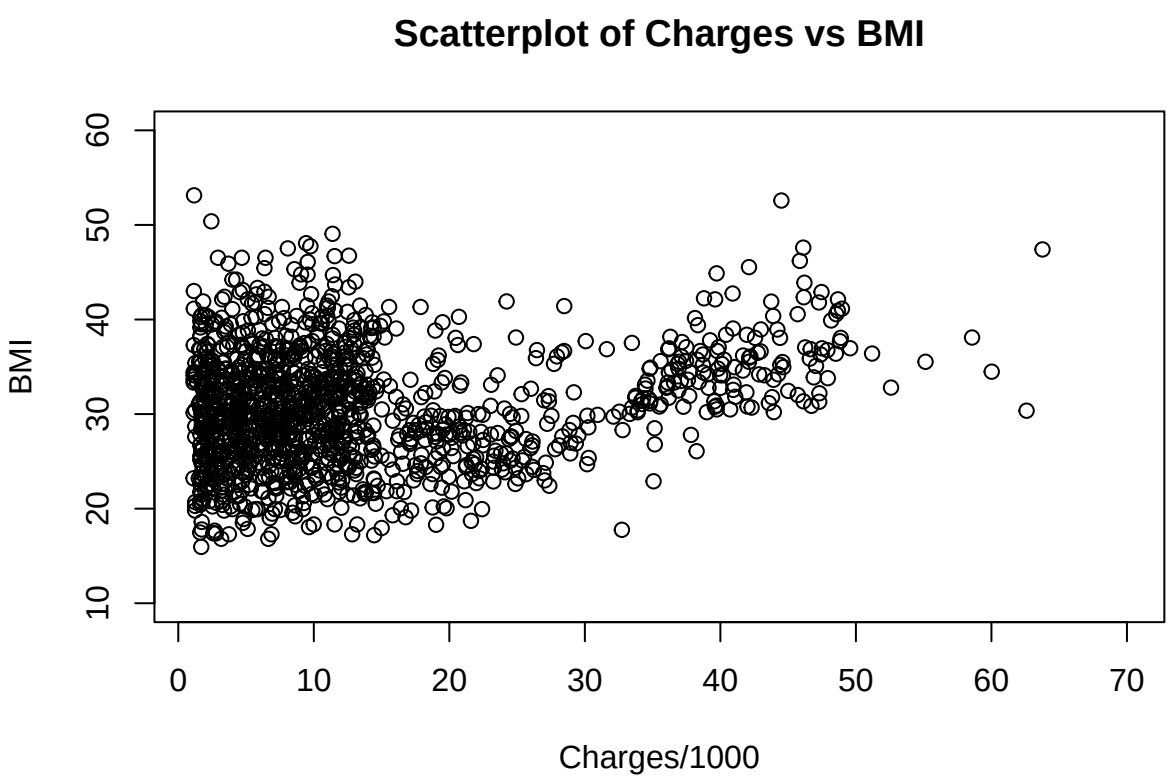
```r
smoker_counts = table(medical_read$smoker)

barplot(smoker_counts,
        xlab = "Smoker",
        ylab = "Count",
        col = c("green", "red"),
        main = "Barplot of Smoker Counts",
        names.arg = names(smoker_counts)
)
```



Scatter plot for Charges and BMI

```r
plot(x = sliced_data$tran_charges, y = sliced_data$bmi,
    xlab = "Charges/1000",
    ylab = "BMI",
    xlim = c(1, 70),
    ylim = c(10, 60),
    main = "Scatterplot of Charges vs BMI"
)
```


Scatterplot of Charges vs BMI
BMI
10
20
30
40
50
60
0
10
20
30
40
50
60
70
Charges/1000