FML_Assignment_1

Sai Goud PVN

2023-09-10

Soucre of my Dataset:

https://www.kaggle.com/datasets/speedoheck/inpatient-hospital-charges?resource

importing Data set into R

```
library(readxl)
fml_inpatientCharges <- read_excel("C:/Users/saigoud/Downloads/fml_inpatientCharges.xlsx")
View(fml_inpatientCharges)
```

The descriptive statistics for the Dataset

```
summary(fml_inpatientCharges)
```

```
##  DRG Definition      Provider Id     Provider Name      Provider Street Address
##  Length:163065      Min.   : 10001   Length:163065      Length:163065
##  Class :character   1st Qu.:110092   Class :character   Class :character
##  Mode  :character   Median :250007   Mode  :character   Mode  :character
##                     Mean   :255570
##                     3rd Qu.:380075
##                     Max.   :670077
##  Provider City      Provider State     Provider Zip Code
##  Length:163065      Length:163065      Min.   : 1040
##  Class :character   Class :character   1st Qu.:27261
##  Mode  :character   Mode  :character   Median :44309
##                                        Mean   :47938
##                                        3rd Qu.:72901
##                                        Max.   :99835
##  Hospital Referral Region Description Total Discharges  Average Covered Charges
##  Length:163065                        Min.   :  11.00   Length:163065
##  Class :character                     1st Qu.:  17.00   Class :character
##  Mode  :character                     Median :  27.00   Mode  :character
##                                       Mean   :  42.78
##                                       3rd Qu.:  49.00
##                                       Max.   :3383.00
```

```
##  Average Total Payments Average Medicare Payments
##  Length:163065          Length:163065
##  Class :character       Class :character
##  Mode  :character       Mode  :character
##
##
##
```

#The Qualitative variables are:

```
summary(fml_inpatientCharges$`Provider Name`)
```

```
##    Length     Class      Mode
##    163065 character character
```

```
summary(fml_inpatientCharges$`Provider Street Address`)
```

```
##    Length     Class      Mode
##    163065 character character
```

```
summary(fml_inpatientCharges$`Provider City`)
```

```
##    Length     Class      Mode
##    163065 character character
```

```
summary(fml_inpatientCharges$`Provider State`)
```

```
##    Length     Class      Mode
##    163065 character character
```

```
summary(fml_inpatientCharges$`Hospital Referral Region Description`)
```

```
##    Length     Class      Mode
##    163065 character character
```

```
summary(fml_inpatientCharges$`Hospital Referral Region Description`)
```

```
##    Length     Class      Mode
##    163065 character character
```

#The Quantitaive Variables are:

```
summary(fml_inpatientCharges$`Provider Zip Code`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1040   27261   44309   47938   72901   99835
```

```r
summary(fml_inpatientCharges$`Total Discharges`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.00   17.00   27.00   42.78   49.00 3383.00
```

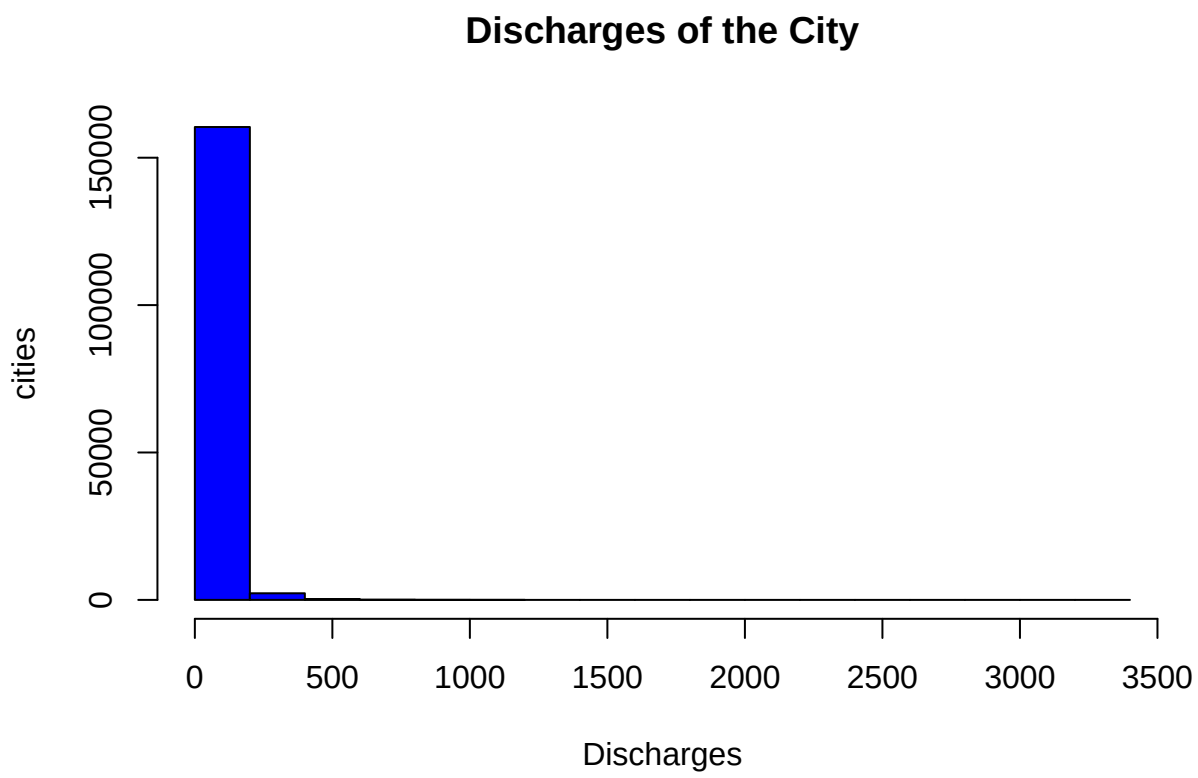
#Transforming of Provider City Data in Quantitaive

```r
fml_inpatientCharges$`Provider City` <-log(fml_inpatientCharges$`Total Discharges`)
summary(fml_inpatientCharges$`Provider City`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.398   2.833   3.296   3.419   3.892   8.127
```

#Plotting one quantitative variable using Histogram

```r
hist(fml_inpatientCharges$`Total Discharges`,
     main="Discharges of the City",
     xlab="Discharges",
     ylab = "cities",
     col ="blue",
     border = "black")
```



Scatterplot of Total Discharges in the cities.

```r
plot(fml_inpatientCharges$`Provider City`,fml_inpatientCharges$`Total Discharges`,
     main = "Scatterplot of City vs Discharges",
     xlab = "Provider City",
     ylab = "Discharges",
     col="green")
```