Assignment 1

2023-01-27

This data set is from <https://vincentarelbundock.github.io/Rdatasets/articles/data.html>. It is number 61.

Part 2: Import the data set

DF <- read.csv("~/Downloads/NaturalGas.csv")

Part 3: Descriptive Statistics

summary(DF)

## X state statecode year   
## Min. : 1.00 Length:138 Min. : 5 Min. :1967   
## 1st Qu.: 35.25 Class :character 1st Qu.:10 1st Qu.:1972   
## Median : 69.50 Mode :character Median :29 Median :1978   
## Mean : 69.50 Mean :27 Mean :1978   
## 3rd Qu.:103.75 3rd Qu.:44 3rd Qu.:1984   
## Max. :138.00 Max. :45 Max. :1989   
## consumption price eprice oprice   
## Min. : 9430 Min. :0.680 Min. : 1.980 Min. : 5.01   
## 1st Qu.: 49104 1st Qu.:1.380 1st Qu.: 2.433 1st Qu.: 8.24   
## Median :300836 Median :2.775 Median : 4.520 Median :19.86   
## Mean :252902 Mean :3.422 Mean : 5.054 Mean :24.64   
## 3rd Qu.:346429 3rd Qu.:5.310 3rd Qu.: 7.282 3rd Qu.:40.10   
## Max. :637289 Max. :8.060 Max. :10.860 Max. :51.73   
## lprice heating income   
## Min. :0.680 Min. : 481 Min. : 7465   
## 1st Qu.:1.258 1st Qu.:2082 1st Qu.: 9637   
## Median :2.615 Median :4272 Median :11367   
## Mean :3.208 Mean :4155 Mean :11193   
## 3rd Qu.:5.192 3rd Qu.:6498 3rd Qu.:12125   
## Max. :7.870 Max. :7440 Max. :16425

mean(DF$price)

## [1] 3.422319

median(DF$price)

## [1] 2.775

range(DF$price)

## [1] 0.68 8.06

median(DF$year)

## [1] 1978

mean (DF$eprice)

## [1] 5.053551

median(DF$eprice)

## [1] 4.52

range((DF$eprice))

## [1] 1.98 10.86

mean (DF$lprice)

## [1] 3.207681

median(DF$lprice)

## [1] 2.615

range(DF$lprice)

## [1] 0.68 7.87

mean (DF$oprice)

## [1] 24.63558

median(DF$oprice)

## [1] 19.865

range(DF$oprice)

## [1] 5.01 51.73

print(IQR(DF$price))

## [1] 3.93

print(IQR(DF$epice))

## [1] NA

print(IQR(DF$lprice))

## [1] 3.935

print(IQR(DF$oprice))

## [1] 31.8625

list(DF$state)

## [[1]]  
## [1] "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY"  
## [16] "NY" "NY" "NY" "NY" "NY" "NY" "NY" "NY" "FL" "FL" "FL" "FL" "FL" "FL" "FL"  
## [31] "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL" "FL"  
## [46] "FL" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI"  
## [61] "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "MI" "TX" "TX" "TX" "TX" "TX" "TX"  
## [76] "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX" "TX"  
## [91] "TX" "TX" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT"  
## [106] "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "UT" "CA" "CA" "CA" "CA" "CA"  
## [121] "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA" "CA"  
## [136] "CA" "CA" "CA"

Part 4: Transform that data

summary(DF$consumption)

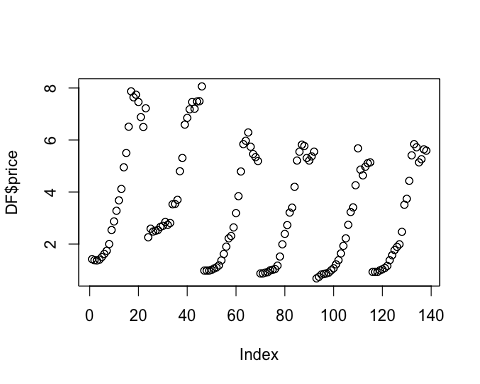
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 9430 49104 300836 252902 346429 637289

summary(log10(DF$consumption))

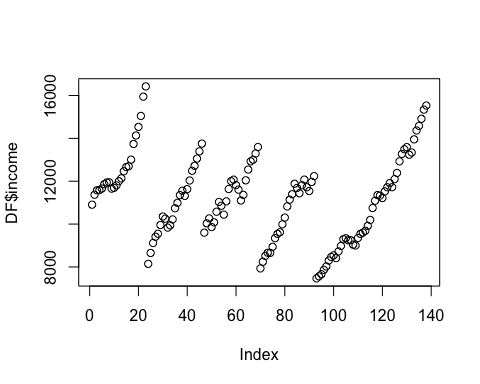
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3.975 4.691 5.478 5.168 5.540 5.804

Part 5: plots

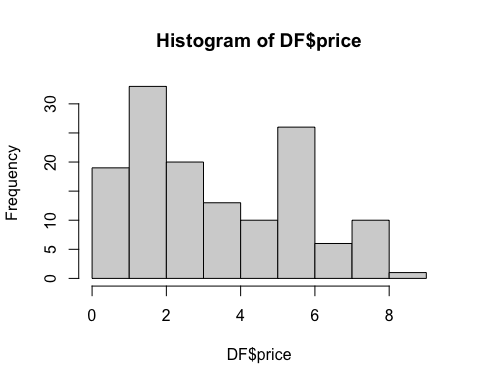
plot(DF$price)



plot(DF$income)



hist(DF$price)



hist(DF$income)

