Hw #2

Lisa Chiobi

9/17/2022

Creating a Data Frame

```r
library(readxl)
setwd("~/Desktop")
# Create Data Frame

#Name <- c("Ronak","Sam","Joe","Mary","Jane","Sally")
#Age <- c(21,24,20,15,19,22)
#Gender <- c("M","M","M","F","F","F")
#GPA <- c(3.7,2.9,3.1,4.0,3.4,3.5)
#myDataFrame <- data.frame(Name,Age,Gender,GPA)
#myDataFrame
myDataFrame<-read_excel("HwTwo.xlsx")

# Explore Data Frame
names(myDataFrame)
```

```
##  [1] "Emp_Age"                "Emp_Gender"
##  [3] "Emp_JobSatisfaction"    "Emp_ImpJobCharacteristic"
##  [5] "Emp_Tenure"             "Emp_PromotionTime"
##  [7] "Emp_DecisionPower"      "Emp_BudgDecision"
##  [9] "Emp_Proud"              "Emp_HigherOffer"
## [11] "Emp_ColleagueRelations"
```

```r
ncol(myDataFrame)
```

```
## [1] 11
```

```r
nrow(myDataFrame)
```

```
## [1] 122
```

```r
dim(myDataFrame)
```

```
## [1] 122  11
```

```
head(myDataFrame)
```

```
## # A tibble: 6 x 11
##   Emp_Age Emp_Gender Emp_JobSa~1 Emp_I~2 Emp_T~3 Emp_P~4 Emp_D~5 Emp_B~6 Emp_P~7
##     <dbl>      <dbl>       <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1      35          1           2       4       3       1       2       1       2
## 2      33          1           2       3       9       5       2       1       2
## 3      23          1           1       1     1.5       1       2       2       1
## 4      60          1           1       1      20       3       2       2       1
## 5      35          1           2       1       3       3       2       1       2
## 6      34          2           2       1       6       1       2       2       2
## # ... with 2 more variables: Emp_HigherOffer <dbl>,
## #   Emp_ColleagueRelations <dbl>, and abbreviated variable names
## #   1: Emp_JobSatisfaction, 2: Emp_ImpJobCharacteristic, 3: Emp_Tenure,
## #   4: Emp_PromotionTime, 5: Emp_DecisionPower, 6: Emp_BudgDecision,
## #   7: Emp_Proud
```

```
tail(myDataFrame)
```

```
## # A tibble: 6 x 11
##   Emp_Age Emp_Gender Emp_JobSa~1 Emp_I~2 Emp_T~3 Emp_P~4 Emp_D~5 Emp_B~6 Emp_P~7
##     <dbl>      <dbl>       <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1      49          1           2       4     1.5       5       4       1       1
## 2      35          2           2       2      10       4       3       2       2
## 3      22          1           1       5       1       1       1       2       2
## 4      33          2           1       5      11       5       2       1       2
## 5      29          2           1       5       2       2       2       1       1
## 6      22          1           2       2    1.25       1       2       2       2
## # ... with 2 more variables: Emp_HigherOffer <dbl>,
## #   Emp_ColleagueRelations <dbl>, and abbreviated variable names
## #   1: Emp_JobSatisfaction, 2: Emp_ImpJobCharacteristic, 3: Emp_Tenure,
## #   4: Emp_PromotionTime, 5: Emp_DecisionPower, 6: Emp_BudgDecision,
## #   7: Emp_Proud
```

```
str(myDataFrame)
```

```
## tibble [122 x 11] (S3: tbl_df/tbl/data.frame)
##  $ Emp_Age               : num [1:122] 35 33 23 60 35 34 61 59 37 30 ...
##  $ Emp_Gender            : num [1:122] 1 1 1 1 1 2 2 1 2 1 ...
##  $ Emp_JobSatisfaction   : num [1:122] 2 2 1 1 2 2 1 2 1 1 ...
##  $ Emp_ImpJobCharacteristic: num [1:122] 4 3 1 1 1 1 1 5 5 5 ...
##  $ Emp_Tenure            : num [1:122] 3 9 1.5 20 3 6 0.75 1.5 3 5 ...
##  $ Emp_PromotionTime     : num [1:122] 1 5 1 3 3 1 5 2 4 2 ...
##  $ Emp_DecisionPower     : num [1:122] 2 2 2 2 2 2 2 3 1 3 2 ...
##  $ Emp_BudgDecision      : num [1:122] 1 1 2 2 1 2 2 2 1 1 ...
##  $ Emp_Proud             : num [1:122] 2 2 1 1 2 2 2 2 2 2 1 ...
##  $ Emp_HigherOffer       : num [1:122] 5 2 5 1 4 4 4 4 2 2 ...
##  $ Emp_ColleagueRelations : num [1:122] 2 1 2 1 2 4 1 1 3 2 ...
```

```
myDataFrame$Gender <- myDataFrame$Emp_Gender
myDataFrame$Age <- myDataFrame$Emp_Age
```

```
class(myDataFrame$Gender)
```

```
## [1] "numeric"
```

```
myDataFrame$Gender <- as.factor(myDataFrame$Gender)
class(myDataFrame$Gender)
```

```
## [1] "factor"
```

```
# Column
myDataFrame$Age
```

```
##    [1] 35 33 23 60 35 34 61 59 37 30 34 34 27 38 41 58 34 48 26 39 29 36 25 39 40
##   [26] 53 39 27 35 25 29 23 40 36 64 43 28 48 52 32 23 44 36 33 52 38 34 62 36 37
##   [51] 39 61 20 22 36 48 58 50 24 44 30 32 32 44 34 42 40 37 32 31 44 39 30 41 39
##   [76] 33 25 31 41 42 33 36 39 39 62 62 34 52 40 43 41 64 26 45 33 36 45 51 38 57
##  [101] 45 43 37 33 51 43 42 25 40 57 38 41 32 39 43 50 49 35 22 33 29 22
```

```
myDataFrame[2]
```

```
## # A tibble: 122 x 1
##    Emp_Gender
##         <dbl>
##  1          1
##  2          1
##  3          1
##  4          1
##  5          1
##  6          2
##  7          2
##  8          1
##  9          2
## 10          1
## # ... with 112 more rows
```

```
myDataFrame[c("Age")]
```

```
## # A tibble: 122 x 1
##      Age
##    <dbl>
##  1    35
##  2    33
##  3    23
##  4    60
##  5    35
##  6    34
##  7    61
##  8    59
##  9    37
## 10    30
## # ... with 112 more rows
```

```r
myDataFrame[2:3]
```

```
## # A tibble: 122 x 2
##    Emp_Gender Emp_JobSatisfaction
##         <dbl>               <dbl>
## 1           1                   2
## 2           1                   2
## 3           1                   1
## 4           1                   1
## 5           1                   2
## 6           2                   2
## 7           2                   1
## 8           1                   2
## 9           2                   1
## 10          1                   1
## # ... with 112 more rows
```

```r
myDataFrame[c("Age","Gender")]
```

```
## # A tibble: 122 x 2
##       Age Gender
##     <dbl> <fct>
## 1      35 1
## 2      33 1
## 3      23 1
## 4      60 1
## 5      35 1
## 6      34 2
## 7      61 2
## 8      59 1
## 9      37 2
## 10     30 1
## # ... with 112 more rows
```

```r
# Rows
myDataFrame[2,]
```

```
## # A tibble: 1 x 13
##   Emp_Age Emp_Gender Emp_JobSa~1 Emp_I~2 Emp_T~3 Emp_P~4 Emp_D~5 Emp_B~6 Emp_P~7
##     <dbl>      <dbl>       <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1      33          1           2       3       9       5       2       1       2
## # ... with 4 more variables: Emp_HigherOffer <dbl>,
## #   Emp_ColleagueRelations <dbl>, Gender <fct>, Age <dbl>, and abbreviated
## #   variable names 1: Emp_JobSatisfaction, 2: Emp_ImpJobCharacteristic,
## #   3: Emp_Tenure, 4: Emp_PromotionTime, 5: Emp_DecisionPower,
## #   6: Emp_BudgDecision, 7: Emp_Proud
```

```r
# CombinedmyDataFrame[2:3,2:3]

#Levels

myDataFrame$Gender
```

```
##   [1] 1 1 1 1 1 2 2 1 2 1 2 2 2 1 1 1 2 2 2 1 2 1 1 2 1 1 2 2 1 1 1 2 1 2 2 2 2
##  [38] 2 1 1 1 2 2 2 1 1 1 1 2 1 2 2 1 1 1 2 2 1 1 2 2 1 1 1 1 2 1 1 2 2 1 1 1 1
##  [75] 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 2 2 1 2 1 1 2 2 2 2 1 1 2 1
## [112] 2 1 1 1 2 1 2 1 2 2 1
## Levels: 1 2
```

```
table(myDataFrame$Gender)
```

```
## 
##  1  2 
## 70 52
```

```
table(myDataFrame$Age)
```

```
## 
## 20 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 48 
##  1  3  3  1  4  2  2  1  3  3  2  5  7  7  4  7  4  4  9  5  5  3  5  4  3  3 
## 49 50 51 52 53 57 58 59 60 61 62 64 
##  1  2  2  3  1  2  2  1  1  2  3  2
```

```
table(myDataFrame$Age, myDataFrame$Gender)
```

```
## 
##       1 2
##   20 1 0
##   22 3 0
##   23 2 1
##   24 1 0
##   25 4 0
##   26 1 1
##   27 0 2
##   28 0 1
##   29 1 2
##   30 2 1
##   31 0 2
##   32 4 1
##   33 3 4
##   34 3 4
##   35 3 1
##   36 3 4
##   37 3 1
##   38 3 1
##   39 6 3
##   40 5 0
##   41 2 3
##   42 1 2
##   43 2 3
##   44 2 2
##   45 1 2
##   48 0 3
##   49 1 0
##   50 1 1
```

```
##    51 0 2
##    52 3 0
##    53 1 0
##    57 1 1
##    58 1 1
##    59 1 0
##    60 1 0
##    61 0 2
##    62 3 0
##    64 1 1
```

```r
# Filtering

myDataFrame[myDataFrame$Gender == "1",]
```

```
## # A tibble: 70 x 13
##    Emp_Age Emp_Gender Emp_JobS~1 Emp_I~2 Emp_T~3 Emp_P~4 Emp_D~5 Emp_B~6 Emp_P~7
##      <dbl>      <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1       35          1          2       4       3       1       2       1       2
## 2       33          1          2       3       9       5       2       1       2
## 3       23          1          1       1     1.5       1       2       2       1
## 4       60          1          1       1      20       3       2       2       1
## 5       35          1          2       1       3       3       2       1       2
## 6       59          1          2       5     1.5       2       1       2       2
## 7       30          1          1       5       5       2       2       1       1
## 8       38          1          1       4      15       1       2       1       2
## 9       41          1          3       1       1       4       2       2       4
## 10      58          1          2       4      36       5       2       1       2
## # ... with 60 more rows, 4 more variables: Emp_HigherOffer <dbl>,
## #   Emp_ColleagueRelations <dbl>, Gender <fct>, Age <dbl>, and abbreviated
## #   variable names 1: Emp_JobSatisfaction, 2: Emp_ImpJobCharacteristic,
## #   3: Emp_Tenure, 4: Emp_PromotionTime, 5: Emp_DecisionPower,
## #   6: Emp_BudgDecision, 7: Emp_Proud
```

```r
# Ordering

myDataFrame[order(myDataFrame$Age),]
```

```
## # A tibble: 122 x 13
##    Emp_Age Emp_Gender Emp_JobS~1 Emp_I~2 Emp_T~3 Emp_P~4 Emp_D~5 Emp_B~6 Emp_P~7
##      <dbl>      <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1       20          1          2       4       3       2       3       1       3
## 2       22          1          2       5       1       3       3       1       2
## 3       22          1          1       5       1       1       1       2       2
## 4       22          1          2       2    1.25       1       2       2       2
## 5       23          1          1       1     1.5       1       2       2       1
## 6       23          2          1       2    0.75       1       3       1       1
## 7       23          1          2       1       1       1       2       2       3
## 8       24          1          2       5       1       2       2       2       2
## 9       25          1          1       4       2       1       3       2       2
## 10      25          1          1       5     2.5       5       2       1       1
## # ... with 112 more rows, 4 more variables: Emp_HigherOffer <dbl>,
## #   Emp_ColleagueRelations <dbl>, Gender <fct>, Age <dbl>, and abbreviated
```

```
## #    variable names 1: Emp_JobSatisfaction, 2: Emp_ImpJobCharacteristic,
## #    3: Emp_Tenure, 4: Emp_PromotionTime, 5: Emp_DecisionPower,
## #    6: Emp_BudgDecision, 7: Emp_Proud
```

```r
myDataFrame[order(-myDataFrame$Age),]
```

```
## # A tibble: 122 x 13
##    Emp_Age Emp_Gender Emp_JobS~1 Emp_I~2 Emp_T~3 Emp_P~4 Emp_D~5 Emp_B~6 Emp_P~7
##      <dbl>      <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1       64          2          1       4    7           5       2       1       1
## 2       64          1          1       5   29           5       2       2       2
## 3       62          1          1       1   52.2         4       2       1       2
## 4       62          1          1       2   27           5       2       1       2
## 5       62          1          1       1    1.91        2       1       2       2
## 6       61          2          1       1    0.75        5       3       2       2
## 7       61          2          1       5   13           4       2       2       1
## 8       60          1          1       1   20           3       2       2       1
## 9       59          1          2       5    1.5         2       1       2       2
## 10      58          1          2       4   36           5       2       1       2
## # ... with 112 more rows, 4 more variables: Emp_HigherOffer <dbl>,
## #   Emp_ColleagueRelations <dbl>, Gender <fct>, Age <dbl>, and abbreviated
## #   variable names 1: Emp_JobSatisfaction, 2: Emp_ImpJobCharacteristic,
## #   3: Emp_Tenure, 4: Emp_PromotionTime, 5: Emp_DecisionPower,
## #   6: Emp_BudgDecision, 7: Emp_Proud
```

```r
# Change Column Name
# names(myDataFrame)[4] <- "Grade"
# head(myDataFrame)
# names(myDataFrame)[4] <- "GPA"
# head(myDataFrame)

# Changing Age for Ronak
# myDataFrame [1,2] <- 31
# head(myDataFrame)
# myDataFrame$Age[1] <- 33
# head(myDataFrame)

# Graphing
barplot(myDataFrame$Age)
```
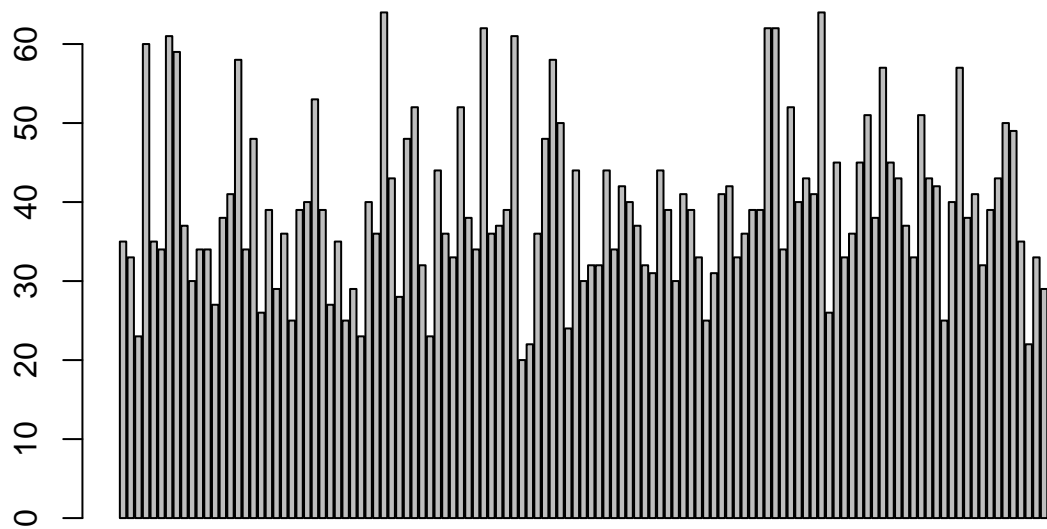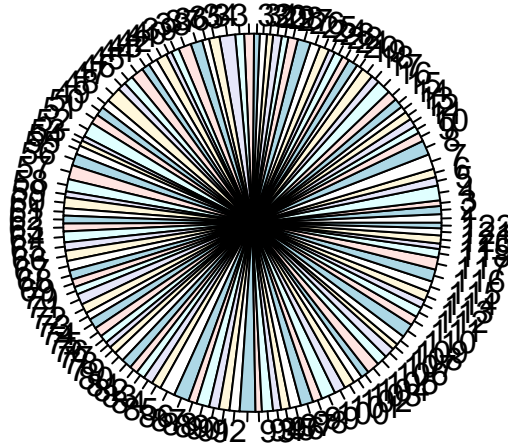
```
barplot(myDataFrame$Age, main = "Age of People", xlab = "People", ylab = "Age", col = "Blue")
```
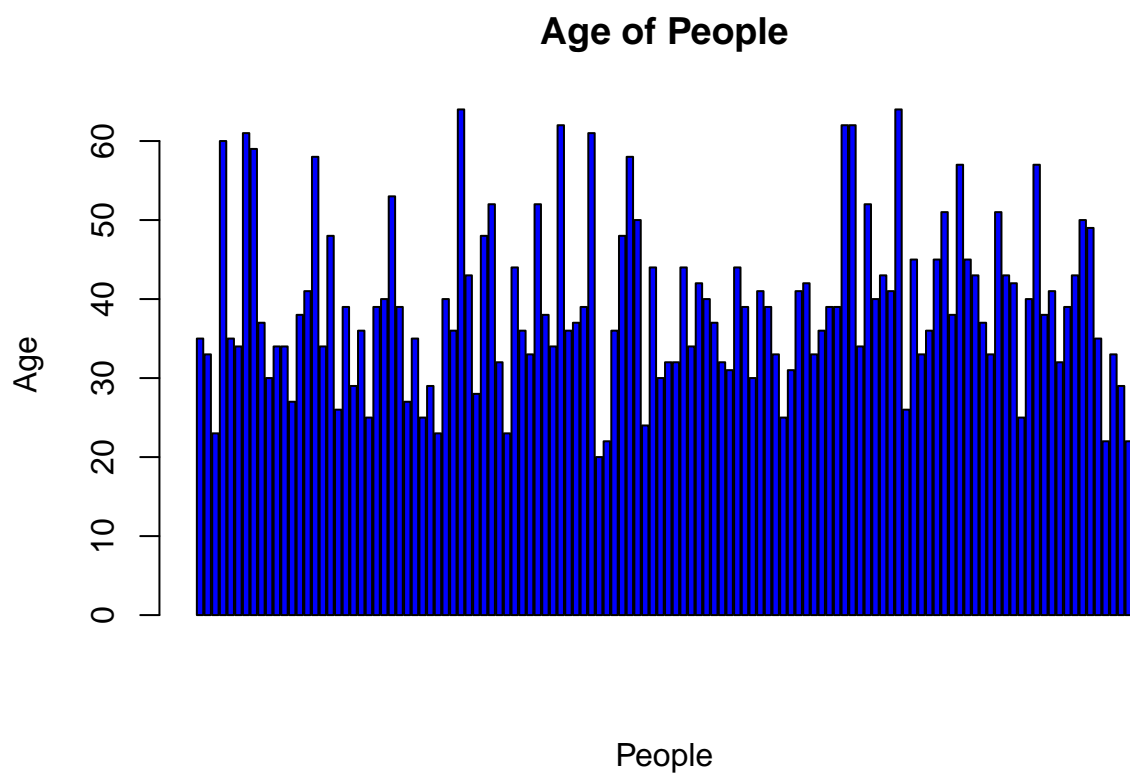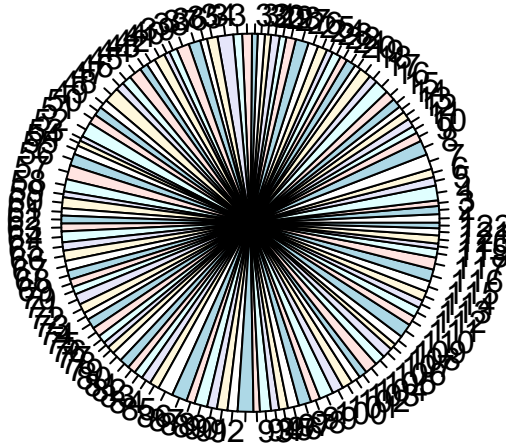
Age of People


```
pie(myDataFrame$Age)
```

```
pie(myDataFrame$Age, main = "Age of People")
```

Age of People


```r
#stem(myDataFrame$GPA)
stem(myDataFrame$Age)
```
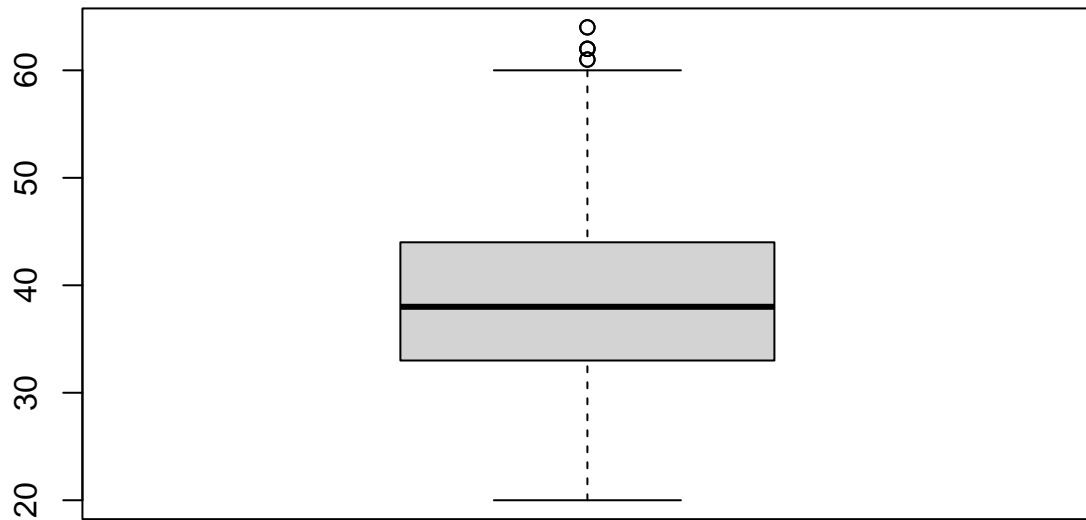
```
## 
##   The decimal point is at the |
## 
##   20 | 0
##   22 | 000000
##   24 | 00000
##   26 | 0000
##   28 | 0000
##   30 | 00000
##   32 | 000000000000
##   34 | 00000000000
##   36 | 00000000000
##   38 | 0000000000000
##   40 | 0000000000
##   42 | 00000000
##   44 | 0000000
##   46 |
##   48 | 0000
##   50 | 0000
##   52 | 0000
##   54 |
##   56 | 00
```

```
##   58 | 000
##   60 | 000
##   62 | 000
##   64 | 00
```

```
hist(myDataFrame$Age)
```

Histogram of myDataFrame$Age


```
boxplot(myDataFrame$Age)
```
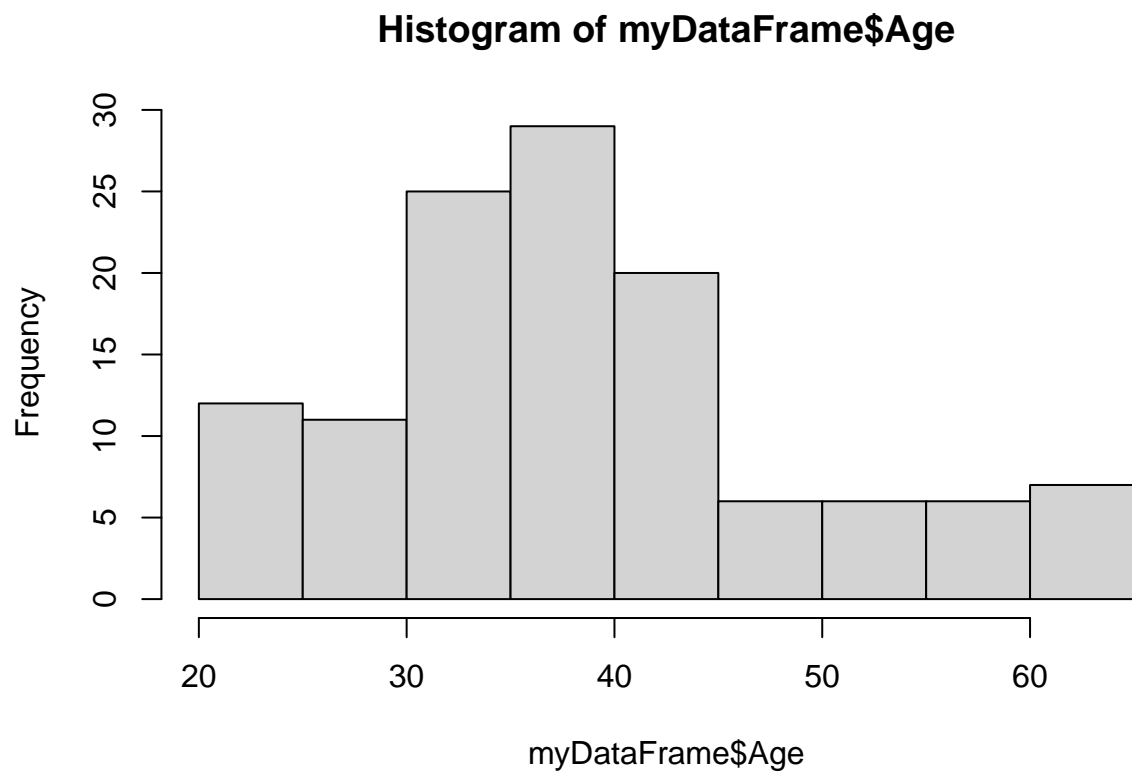
```
boxplot(myDataFrame$Age,myDataFrame$Gender)
```
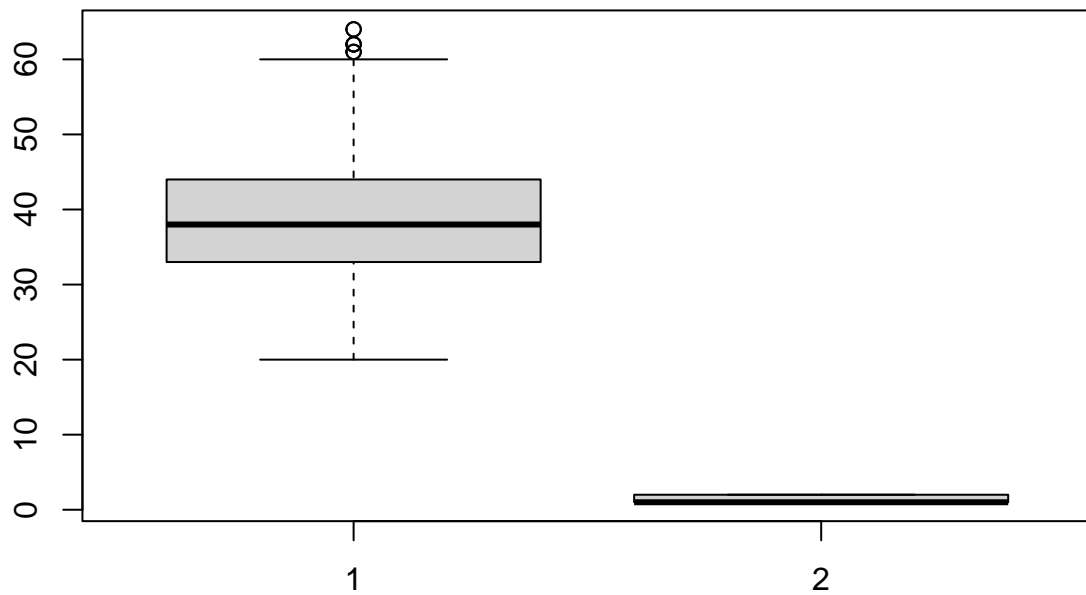
Summary Stats

```
min(myDataFrame$Age)
```

```
## [1] 20
```

```
max(myDataFrame$Age)
```

```
## [1] 64
```

```
range(myDataFrame$Age)
```

```
## [1] 20 64
```

```
StatRange = max(myDataFrame$Age) - min(myDataFrame$Age)
StatRange
```

```
## [1] 44
```

```
mean(myDataFrame$Age)
```

```
## [1] 39.09836
```

```
sd(myDataFrame$Age)
```

```
## [1] 10.4339
```

```
var(myDataFrame$Age)
```

```
## [1] 108.8663
```

```
sqrt(var(myDataFrame$Age))
```

```
## [1] 10.4339
```

```
fivenum(myDataFrame$Age)
```

```
## [1] 20 33 38 44 64
```

```
IQR(myDataFrame$Age)
```

```
## [1] 11
```

```
quantile(myDataFrame$Age)
```

```
##   0%  25%  50%  75% 100%
##   20   33   38   44   64
```

```
summary(myDataFrame$Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    20.0    33.0    38.0    39.1    44.0    64.0
```

```
boxplot.stats(myDataFrame$Age)
```

```
## $stats
## [1] 20 33 38 44 60
##
## $n
## [1] 122
##
## $conf
## [1] 36.42649 39.57351
##
## $out
## [1] 61 64 62 61 62 62 64
```

```
boxplot.stats(myDataFrame$Age)$out
```

```
## [1] 61 64 62 61 62 62 64
```

```
# Data Frame Summary
```

```
summary(myDataFrame)
```

```
##      Emp_Age        Emp_Gender      Emp_JobSatisfaction Emp_ImpJobCharacteristic
## Min.   :20.0   Min.   :1.000   Min.   :1.000       Min.   :1.000
## 1st Qu.:33.0   1st Qu.:1.000   1st Qu.:1.000       1st Qu.:2.000
## Median :38.0   Median :1.000   Median :1.000       Median :5.000
## Mean   :39.1   Mean   :1.426   Mean   :1.631       Mean   :3.705
## 3rd Qu.:44.0   3rd Qu.:2.000   3rd Qu.:2.000       3rd Qu.:5.000
## Max.   :64.0   Max.   :2.000   Max.   :4.000       Max.   :5.000
##    Emp_Tenure      Emp_PromotionTime Emp_DecisionPower Emp_BudgDecision
## Min.   : 0.080   Min.   :1.000     Min.   :1.000     Min.   :1.000
## 1st Qu.: 2.000   1st Qu.:2.000     1st Qu.:2.000     1st Qu.:1.000
## Median : 5.000   Median :4.000     Median :2.000     Median :1.000
## Mean   : 8.224   Mean   :3.287     Mean   :2.279     Mean   :1.451
## 3rd Qu.:11.000   3rd Qu.:5.000     3rd Qu.:3.000     3rd Qu.:2.000
## Max.   :52.250   Max.   :5.000     Max.   :4.000     Max.   :2.000
##    Emp_Proud      Emp_HigherOffer Emp_ColleagueRelations Gender     Age
## Min.   :1.000   Min.   :1.000   Min.   :1.000          1:70   Min.   :20.0
## 1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000          2:52   1st Qu.:33.0
## Median :2.000   Median :4.000   Median :2.000                 Median :38.0
## Mean   :1.787   Mean   :3.254   Mean   :1.787                 Mean   :39.1
## 3rd Qu.:2.000   3rd Qu.:4.000   3rd Qu.:2.000                 3rd Qu.:44.0
## Max.   :4.000   Max.   :5.000   Max.   :4.000                 Max.   :64.0
```

```
by(myDataFrame$Age,myDataFrame$Gender,mean)
```

```
## myDataFrame$Gender: 1
## [1] 38.67143
## ----------------------------------------------------------------
## myDataFrame$Gender: 2
## [1] 39.67308
```

```
by(myDataFrame$Age,myDataFrame$Gender,sd)
```

```
## myDataFrame$Gender: 1
## [1] 11.06204
## ----------------------------------------------------------------
## myDataFrame$Gender: 2
## [1] 9.599267
```

```
by(myDataFrame$Age,myDataFrame$Gender,summary)
```

```
## myDataFrame$Gender: 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   20.00   32.00   38.00   38.67   43.00   64.00
## ----------------------------------------------------------------
## myDataFrame$Gender: 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.00   33.00   38.50   39.67   44.25   64.00
```

```r
aggregate(myDataFrame$Age,list("Type" = myDataFrame$Gender),median)
```
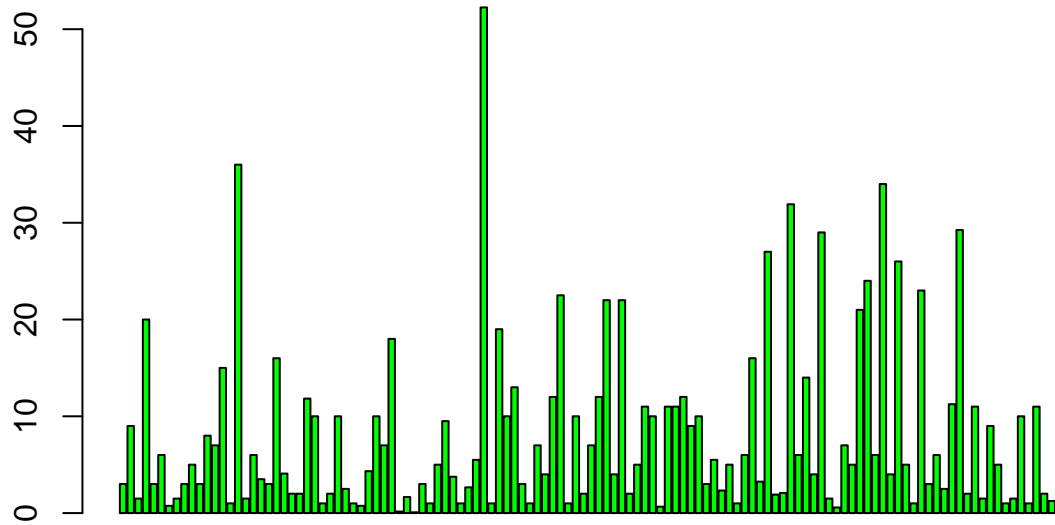
```
##   Type    x
## 1    1 38.0
## 2    2 38.5
```

```r
aggregate(myDataFrame$Age,list("Type" = myDataFrame$Gender),summary)
```

```
##   Type   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
## 1    1 20.00000  32.00000 38.00000 38.67143  43.00000 64.00000
## 2    2 23.00000  33.00000 38.50000 39.67308  44.25000 64.00000
```

```r
# EMPLOYEE TENURE INSIGHTS
```

```r
barplot(myDataFrame$Emp_Tenure, col="Green")
```



```r
min(myDataFrame$Emp_Tenure)
```

```
## [1] 0.08
```

```r
max(myDataFrame$Emp_Tenure)
```

```
## [1] 52.25
```

```
mean(myDataFrame$Emp_Tenure)
```

```
## [1] 8.223525
```

```
plot(myDataFrame$Age,myDataFrame$Emp_Tenure)
```
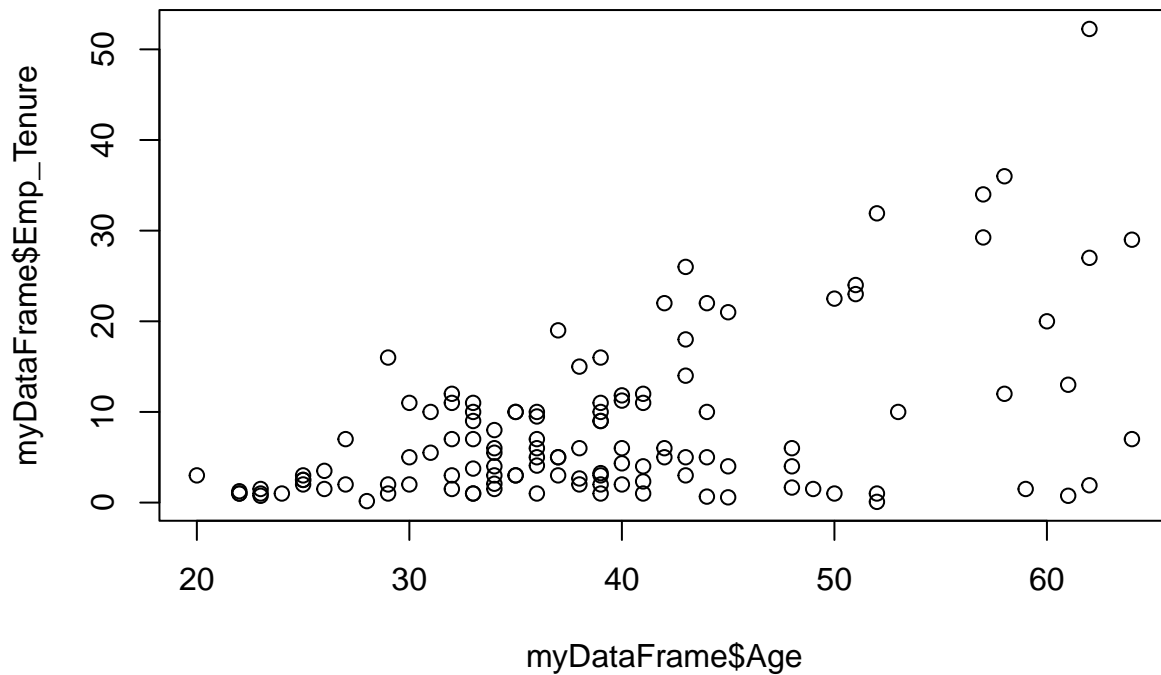


```
summary(myDataFrame$Emp_Tenure)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.080   2.000   5.000   8.224  11.000  52.250
```
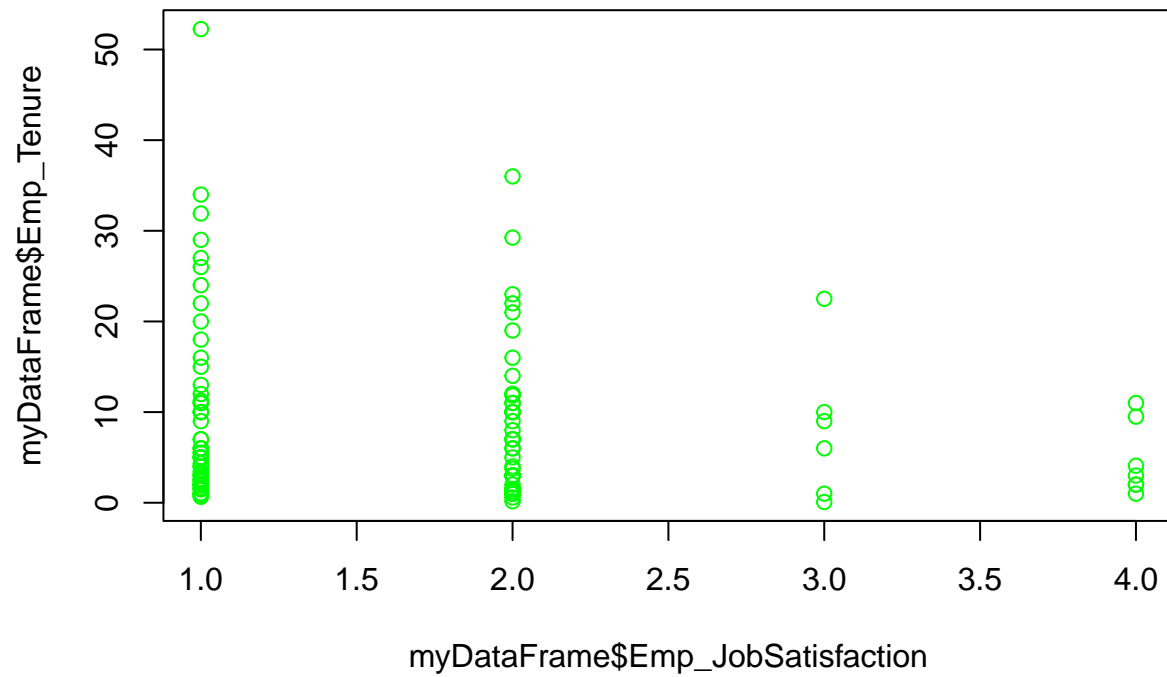
```
# EMPLOYEE JOB SATISFACTION INSIGHTS
```

```
mean(myDataFrame$Emp_JobSatisfaction)
```
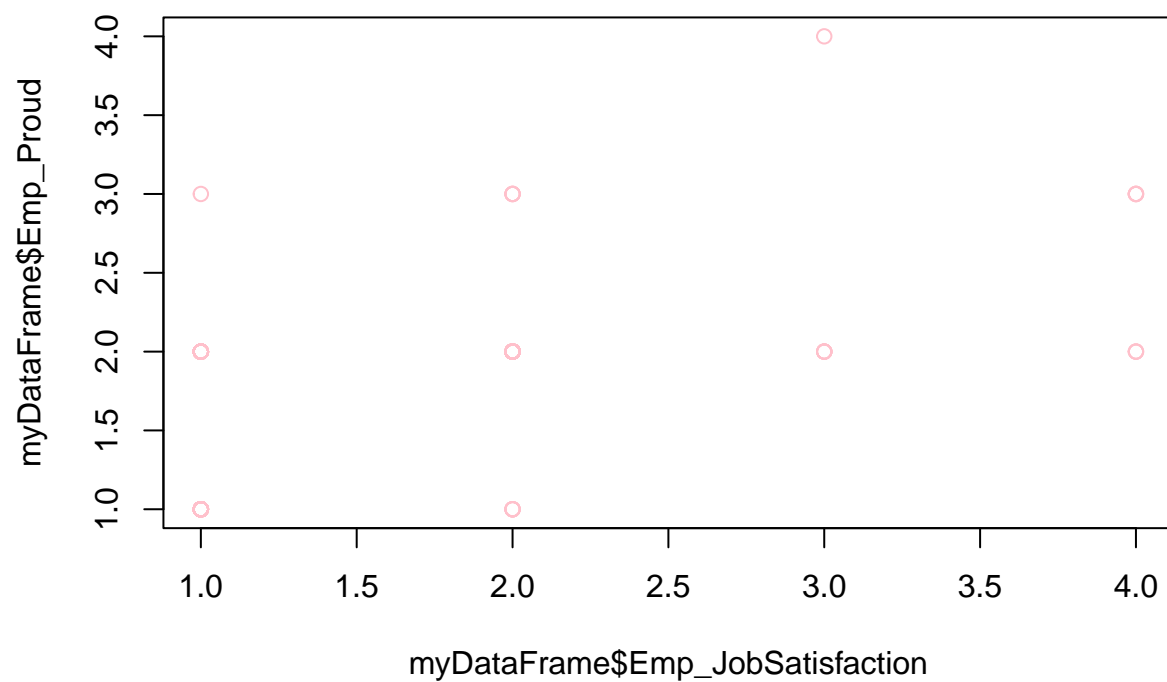
```
## [1] 1.631148
```

```
summary(myDataFrame$Emp_JobSatisfaction)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.631   2.000   4.000
```

```
plot(myDataFrame$Emp_JobSatisfaction, myDataFrame$Emp_Tenure, col="Green")
```



```
plot(myDataFrame$Emp_JobSatisfaction, myDataFrame$Emp_Proud, col="Pink")
```

```
# EMPLOYEE COWORKER RELATIONS INSIGHTS
```
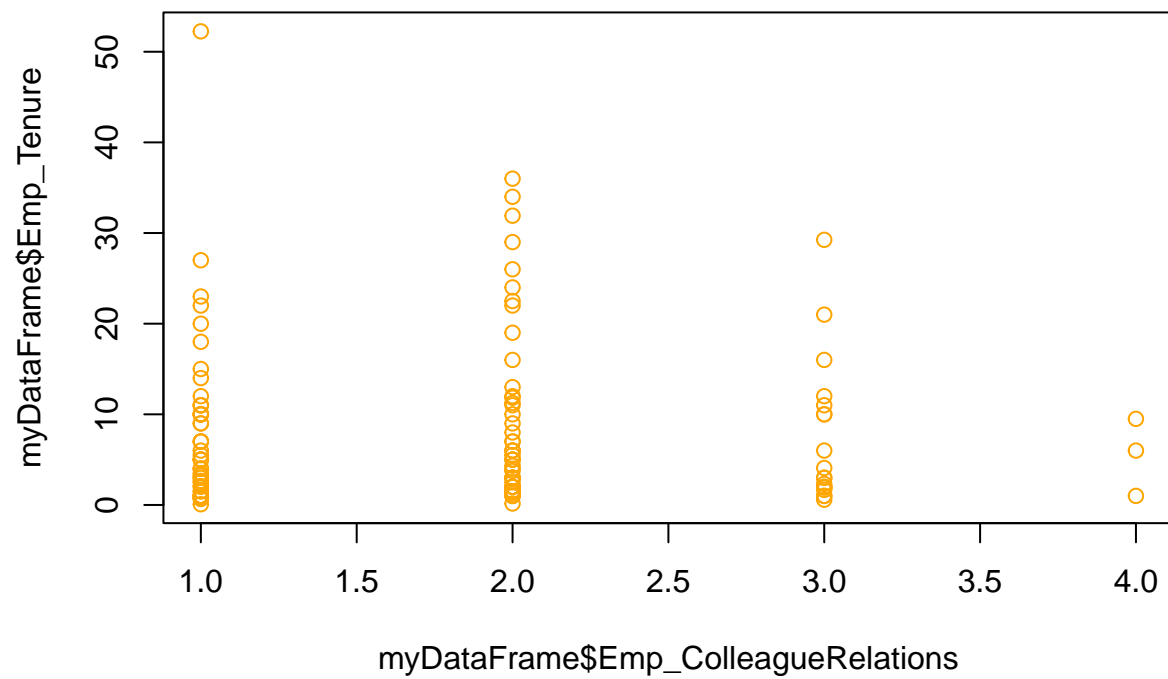
```
mean(myDataFrame$Emp_ColleagueRelations)
```
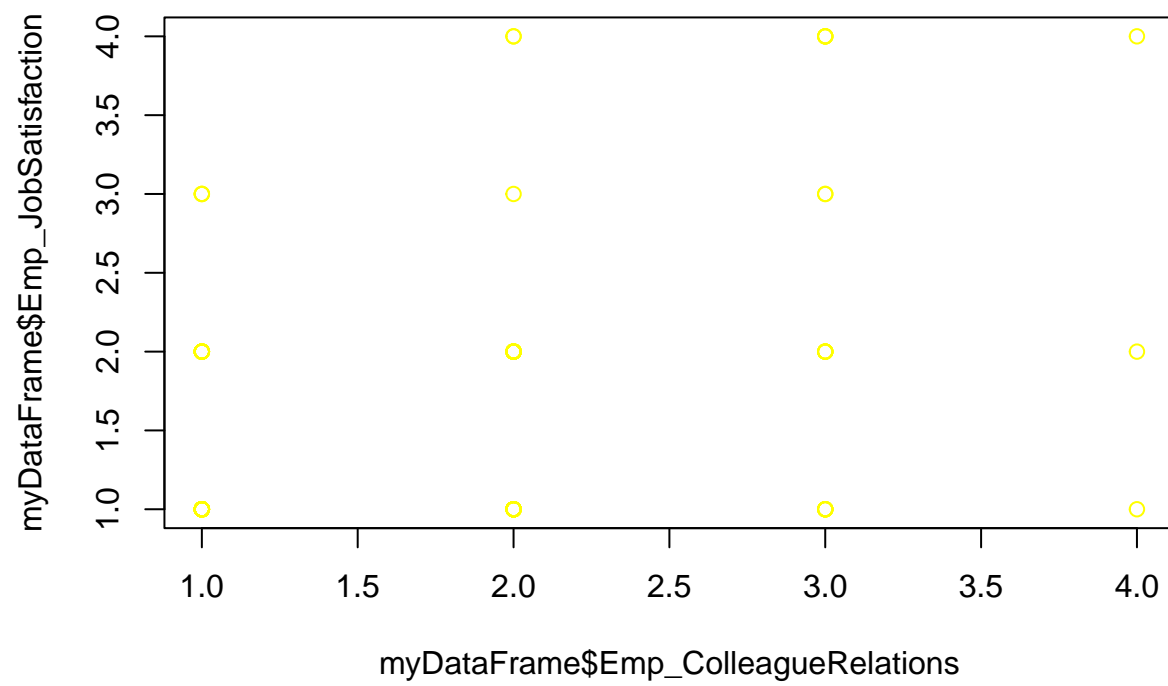
```
## [1] 1.786885
```

```
summary(myDataFrame$Emp_ColleagueRelations)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   1.787   2.000   4.000
```
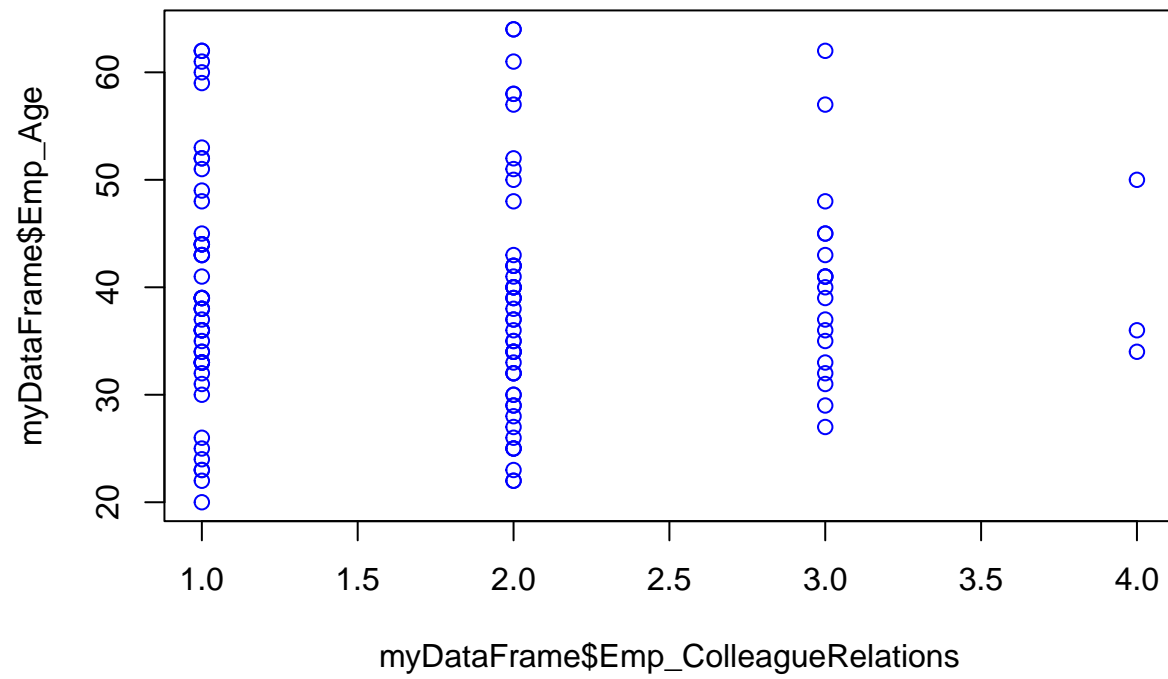
```
plot(myDataFrame$Emp_ColleagueRelations, myDataFrame$Emp_Tenure, col="Orange")
```

```
plot(myDataFrame$Emp_ColleagueRelations, myDataFrame$Emp_JobSatisfaction, col="Yellow")
```
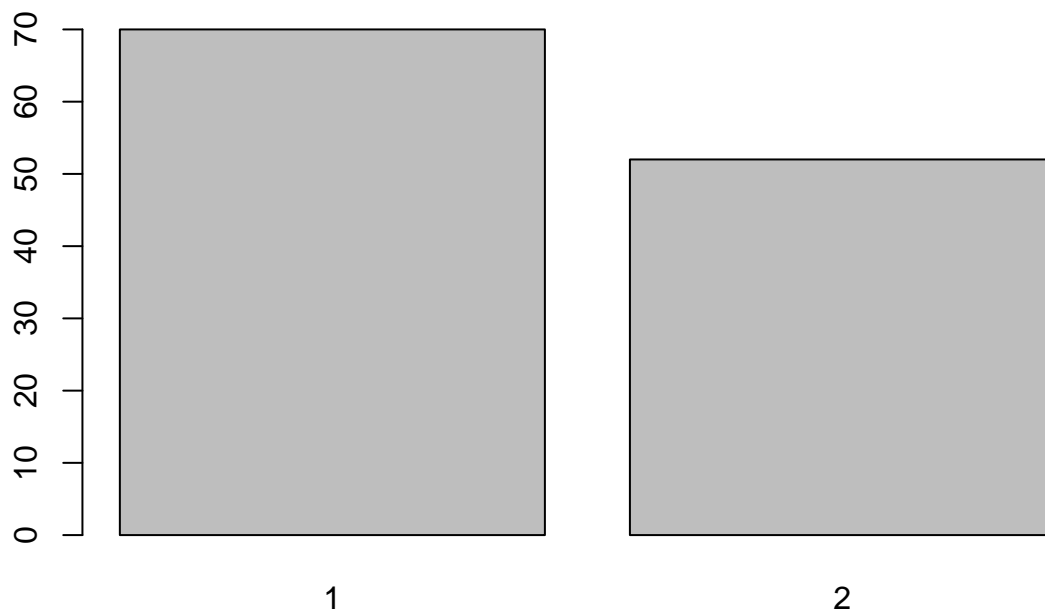
```
plot(myDataFrame$Emp_ColleagueRelations, myDataFrame$Emp_Age, col="bLUE")
```

```
# GENDER INSIGHTS

barplot(table(myDataFrame$Gender))
```

```
pie(table(myDataFrame$Gender), main = "Gender Proportions: 1=Male,2=Female")
```

Gender Proportions: 1=Male,2=Female