STUDENT PERFORMANCE DATASET

Abhiram

2023-02-23

```
library(readr)
library(ggplot2)
library(lattice)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

#import data
df<- read_csv("StudentsPerformance.csv")

## Rows: 1000 Columns: 8

## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (5): gender, race/ethnicity, parental level of education, lunch, test
pr...
## dbl (3): math score, reading score, writing score
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

# to view the first few rows
head(df)

## # A tibble: 6 × 8
##   gender `race/ethnicity` parental level…¹ lunch test …² math …³ readi…⁴
writi…⁵
##   <chr>  <chr>            <chr>            <chr> <chr>     <dbl>   <dbl>
<dbl>
## 1 female group B          bachelor's degr… stan… none          72      72
74
## 2 female group C          some college     stan… comple…      69      90
```

```
88
## 3 female group B          master's degree  stan… none         90      95
93
## 4 male   group A          associate's deg… free… none         47      57
44
## 5 male   group C          some college     stan… none         76      78
75
## 6 female group B          associate's deg… stan… none         71      83
78
## # … with abbreviated variable names ¹`parental level of education`,
## #   ²`test preparation course`, ³`math score`, ⁴`reading score`,
## #   ⁵`writing score`
```

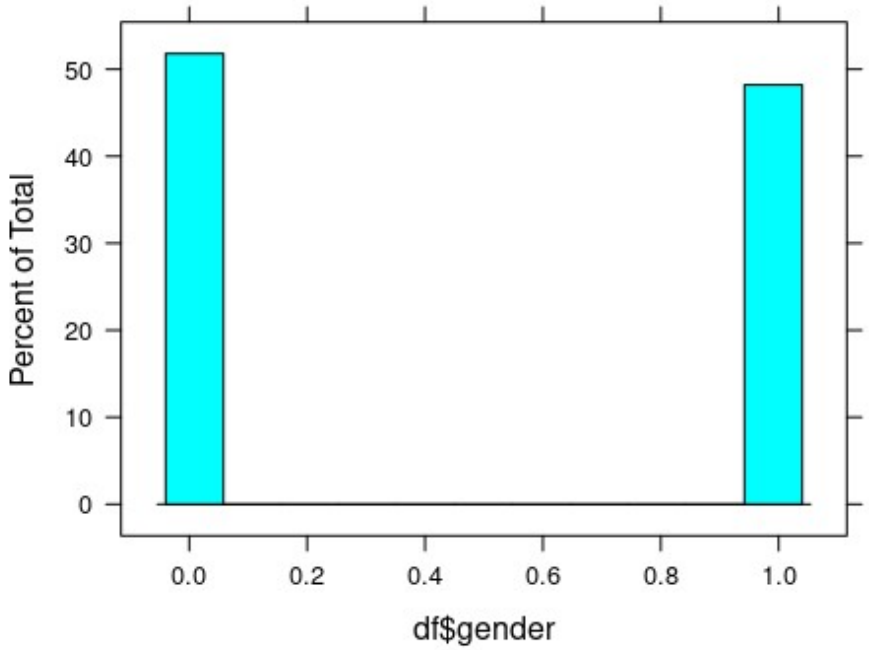
```
# to view summary statistics
summary(df)
```

```
##     gender          race/ethnicity     parental level of education
##  Length:1000        Length:1000        Length:1000
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##     lunch           test preparation course   math score     reading score
##  Length:1000        Length:1000             Min.   :  0.00   Min.   :
17.00
##  Class :character   Class :character        1st Qu.: 57.00   1st Qu.:
59.00
##  Mode  :character   Mode  :character        Median : 66.00   Median :
70.00
##                                             Mean   : 66.09   Mean   :
69.17
##                                             3rd Qu.: 77.00   3rd Qu.:
79.00
##                                             Max.   :100.00   Max.
:100.00
##  writing score
##  Min.   : 10.00
##  1st Qu.: 57.75
##  Median : 69.00
##  Mean   : 68.05
##  3rd Qu.: 79.00
##  Max.   :100.00
```

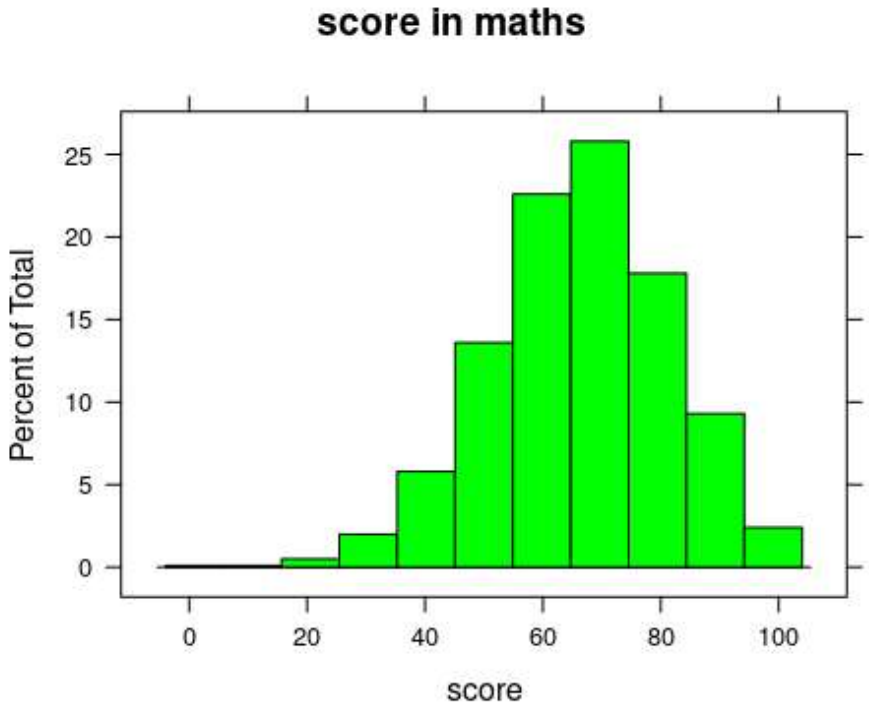
```
df$gender[df$gender == 'male']=1
df$gender[df$gender== 'female']=0
df$gender <- as.integer(df$gender)
count(df,'gender')
```

```
## # A tibble: 1 × 2
##   `"gender"`     n
```

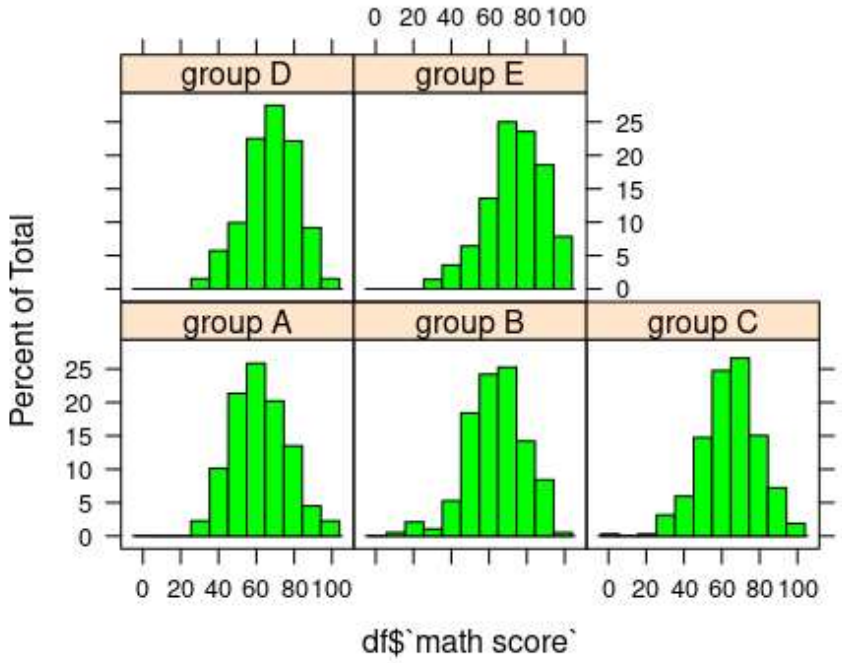
```
##   <chr>      <int>
## 1 gender      1000

print(df)

## # A tibble: 1,000 × 8
##    gender `race/ethnicity` parental leve…¹ lunch test …² math …³ readi…⁴
writi…⁵
##     <int> <chr>            <chr>           <chr> <chr>     <dbl>   <dbl>
<dbl>
##  1      0 group B          bachelor's deg… stan… none         72      72
74
##  2      0 group C          some college    stan… comple…      69      90
88
##  3      0 group B          master's degree stan… none         90      95
93
##  4      1 group A          associate's de… free… none         47      57
44
##  5      1 group C          some college    stan… none         76      78
75
##  6      0 group B          associate's de… stan… none         71      83
78
##  7      0 group B          some college    stan… comple…      88      95
92
##  8      1 group B          some college    free… none         40      43
39
##  9      1 group D          high school     free… comple…      64      64
67
## 10      0 group B          high school     free… none         38      60
50
## # … with 990 more rows, and abbreviated variable names
## #   ¹`parental level of education`, ²`test preparation course`, ³`math
score`,
## #   ⁴`reading score`, ⁵`writing score`

histogram(df$gender)
```



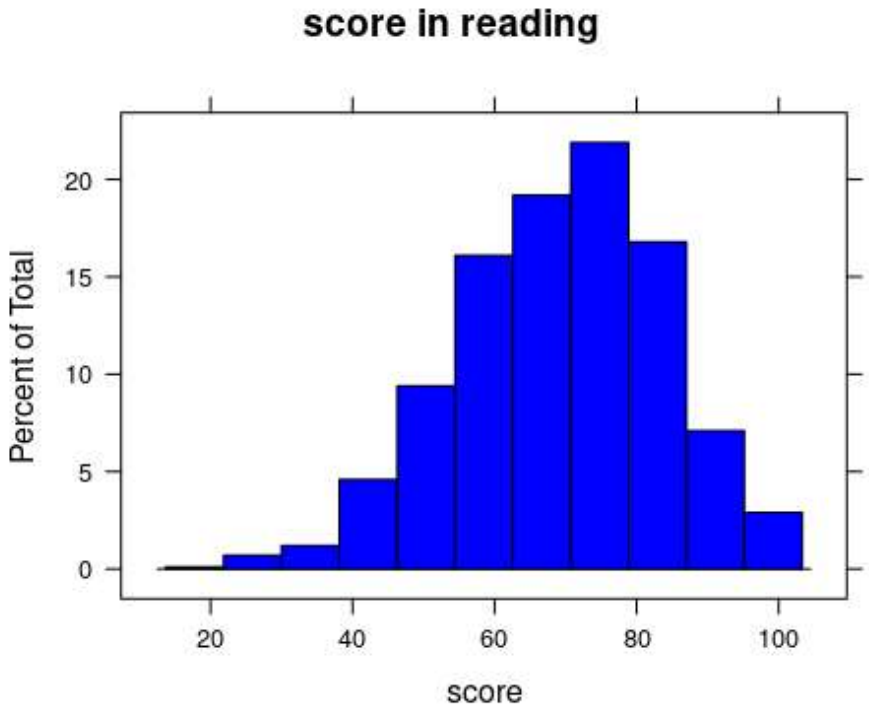

```
histogram(df$`math score`,col='green',main='score in maths',xlab = 'score')
```



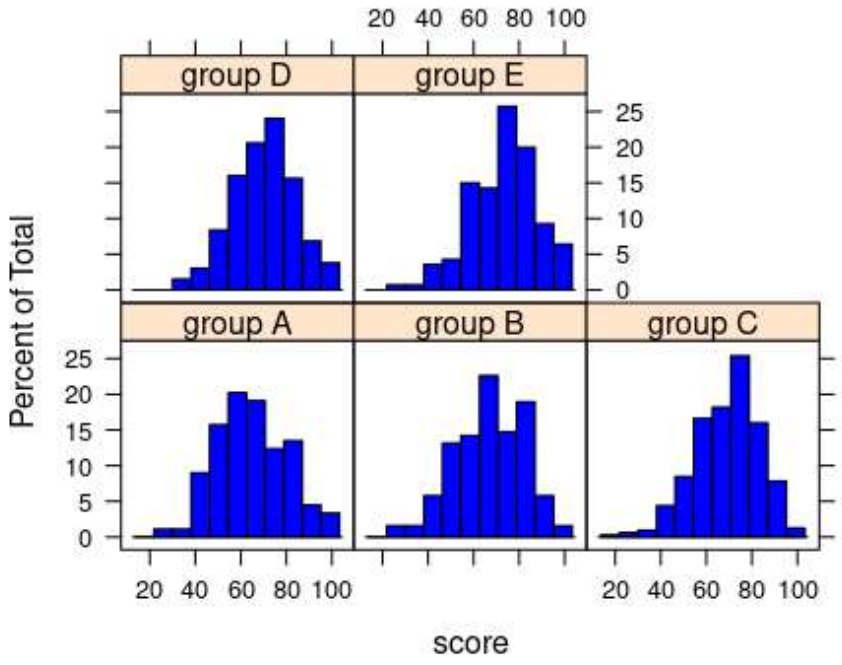

```
histogram(~df$`math score`|df$`race/ethnicity`,data=df,col='green')
```



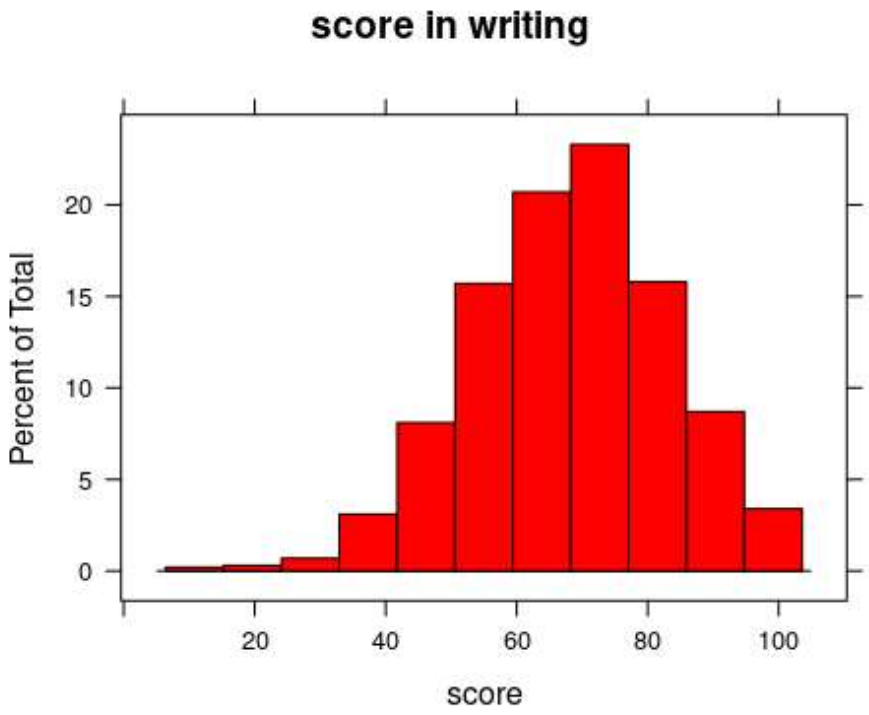

```
histogram(df$`reading score`,col='blue',main='score in reading' ,xlab =
'score')
```



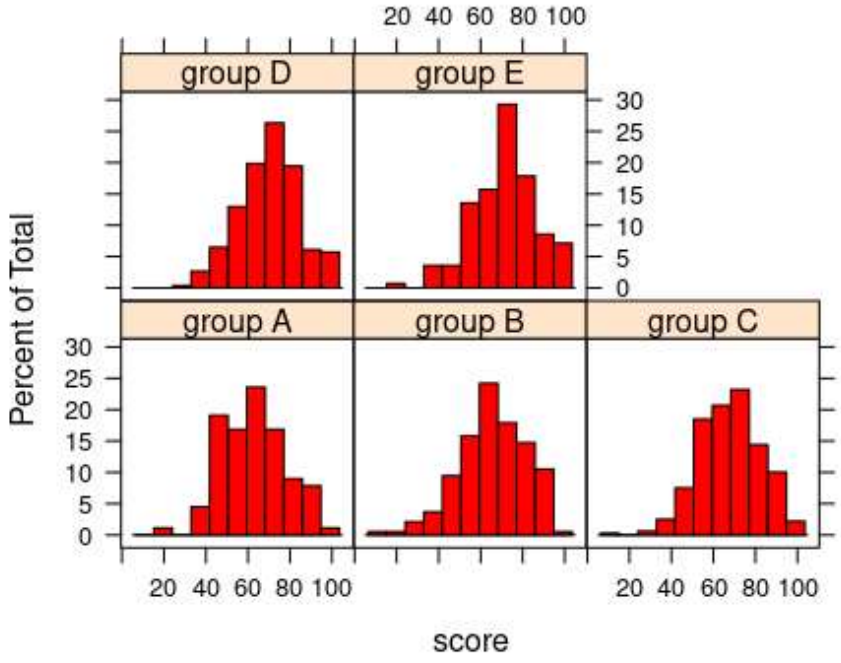
```
histogram(~df$`reading score`|df$`race/ethnicity`,data=df,col='blue',xlab =
'score')
```



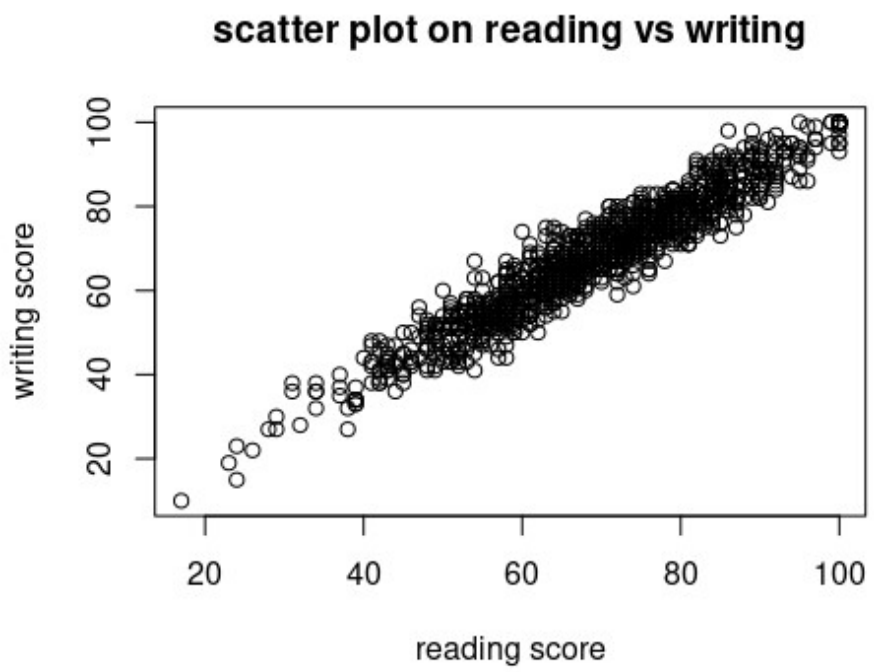

```
histogram(df$`writing score`,col='red',main='score in writing' ,xlab =
'score')
```



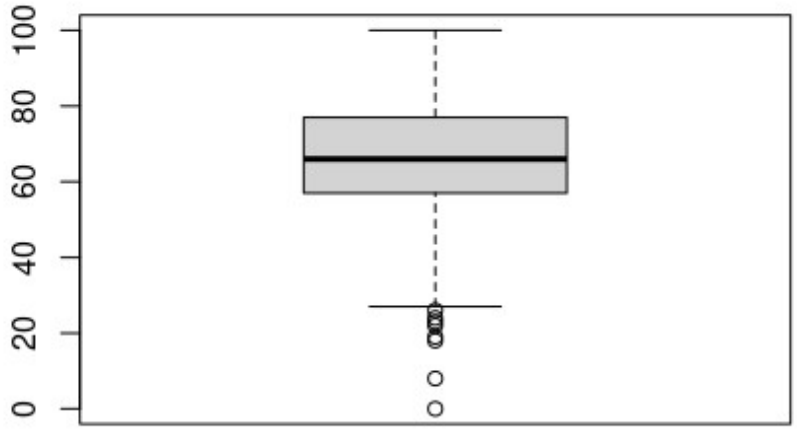

```
histogram(~df$`writing score`|df$`race/ethnicity`,data=df,col='red',xlab =
'score')
```



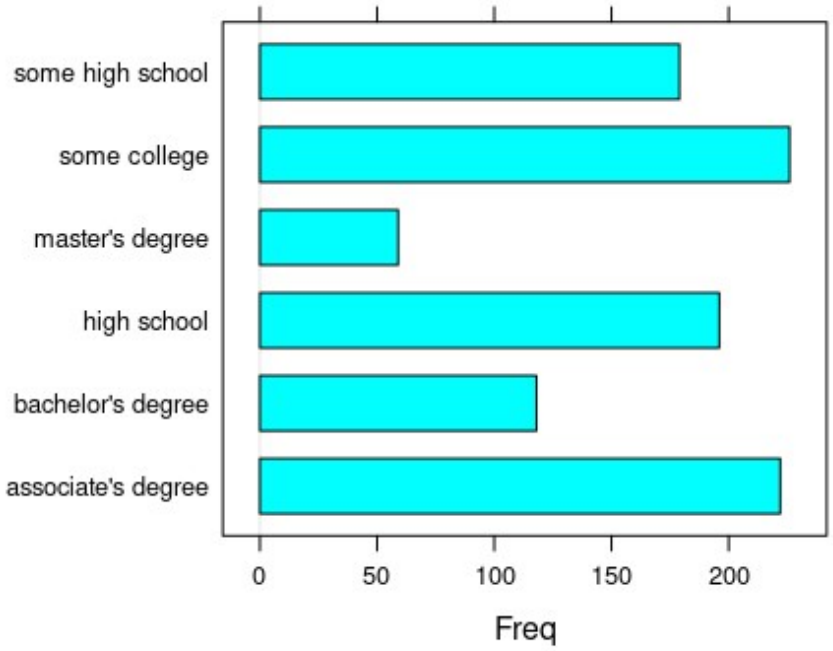
```r
# scatterplot of reading vs writing
plot(df$`reading score`,df$`writing score`,
xlab ='reading score',
ylab= 'writing score',
main='scatter plot on reading vs writing')
```



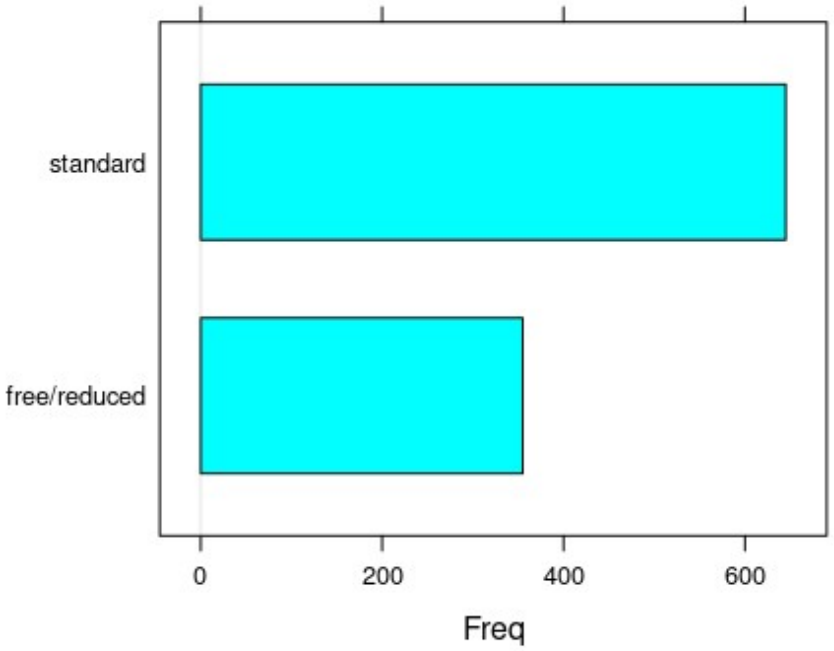

```r
#boxplot
boxplot(df$`math score`)
```



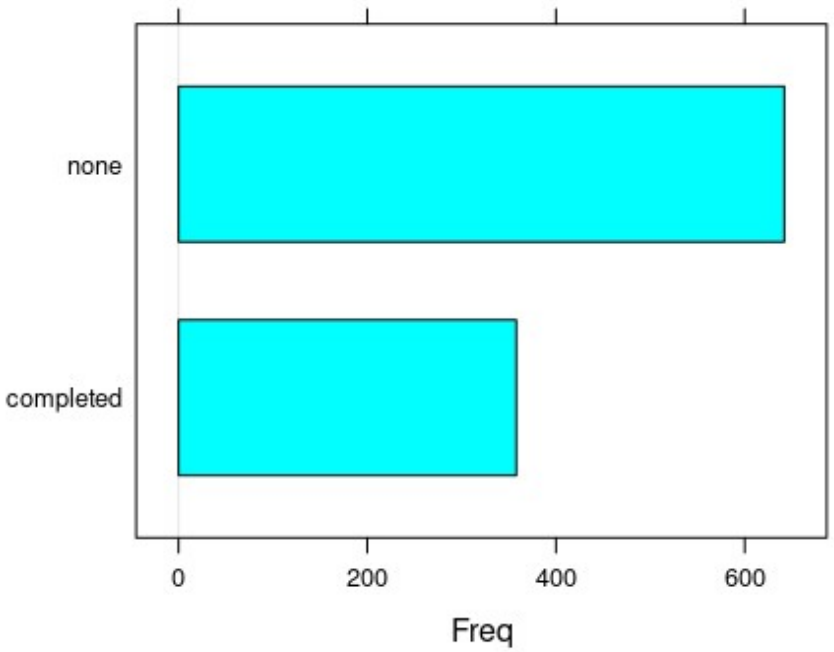

```r
#barchar
barchart(df$`parental level of education`)
```



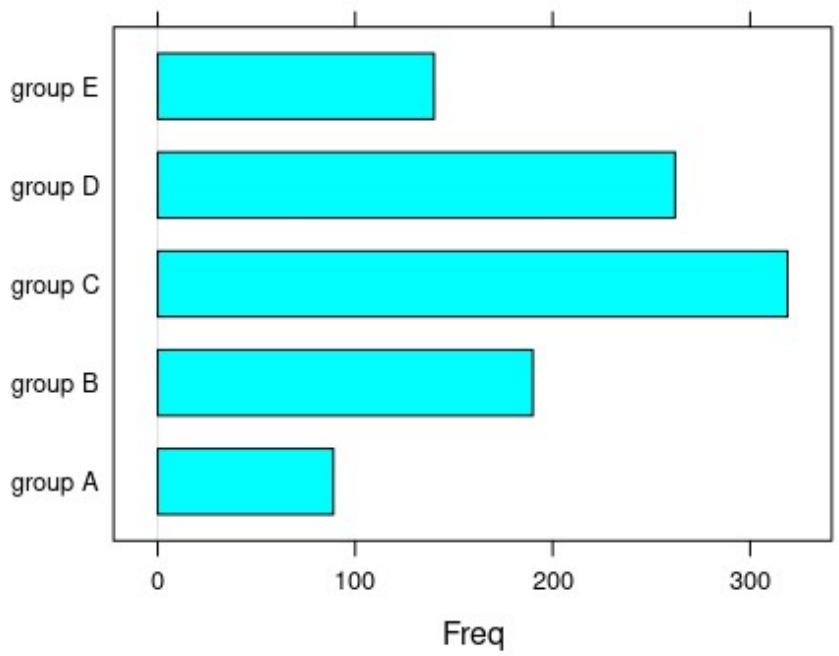
```
barchart(df$lunch)
```



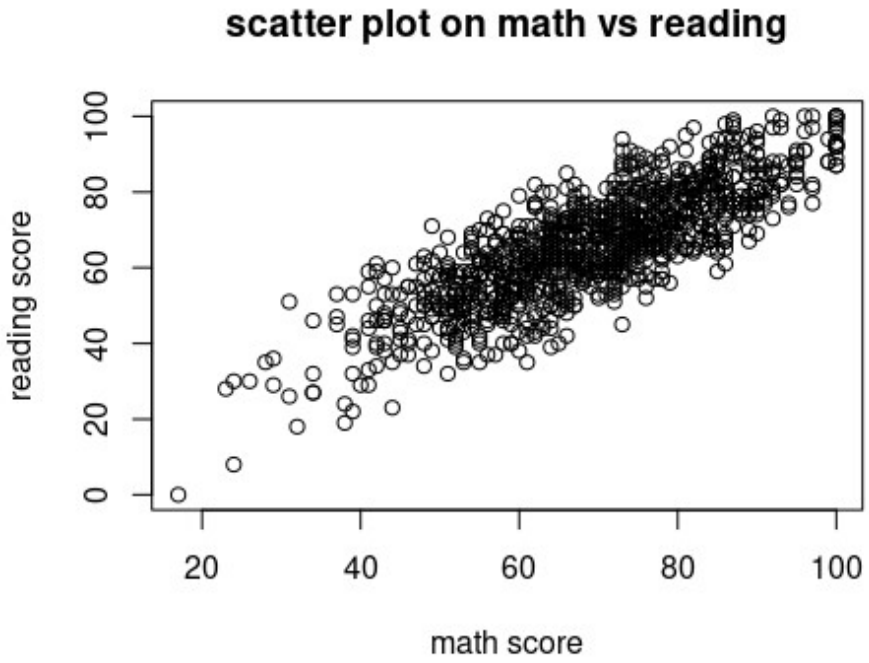

```
barchart(df$`test preparation course`)
```



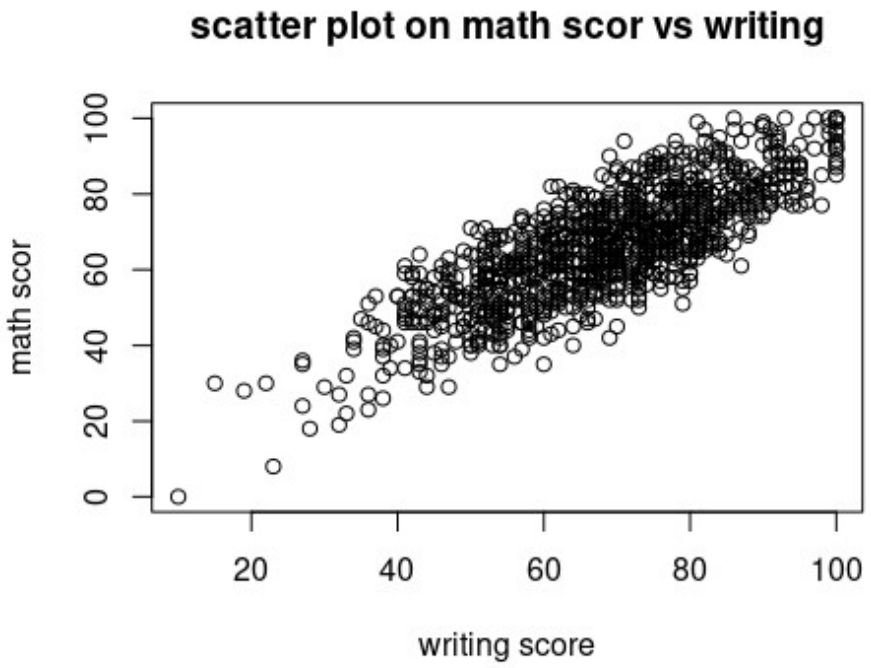
```
barchart(df$`race/ethnicity`)
```




```
# scatterplot of math vs. reading scores
plot(y=df$`math score`, x=df$`reading score`,
xlab ='math score',
ylab= 'reading score',
main='scatter plot on math vs reading')
```




```r
#scatterplot of math scores vs writing
plot(y=df$`math score`,x=df$`writing score`,
xlab ='writing score',
ylab= 'math scor',
main='scatter plot on math scor vs writing')
```




```
library(dplyr)
# group the data by race/ethnicity
grouped_df <- df %>% group_by(df$`race/ethnicity`)
# compute the mean and standard deviation of math score for each group
library(dplyr)
grouped_df <- group_by(df, `parental level of education`)
summary_df <- summarise(grouped_df,
mean.math.score = mean(`math score`, na.rm = TRUE),
sd.math.score = sd(`math score`, na.rm = TRUE))
summary_df
```

```
## # A tibble: 6 × 3
##   `parental level of education` mean.math.score sd.math.score
##   <chr>                                  <dbl>          <dbl>
## 1 associate's degree                      67.9           15.1
## 2 bachelor's degree                       69.4           14.9
## 3 high school                             62.1           14.5
## 4 master's degree                         69.7           15.2
## 5 some college                            67.1           14.3
## 6 some high school                        63.5           15.9
```