

Assignment 1

Noorah

```
library()

df<- read.csv("~/Desktop/spring 2021/ML/norah/cereal.csv")

small.df = df [c(1:10),c(1:8)]
print (small.df)

##           name mfr type calories protein fat sodium fiber
## 1         100% Bran   N   C         70      4   1    130  10.0
## 2       100% Natural Bran   Q   C        120     3   5     15   2.0
## 3             All-Bran   K   C         70      4   1    260   9.0
## 4 All-Bran with Extra Fiber   K   C         50      4   0    140  14.0
## 5         Almond Delight   R   C        110     2   2    200   1.0
## 6   Apple Cinnamon Cheerios   G   C        110     2   2    180  1.5
## 7         Apple Jacks     K   C        110     2   0    125   1.0
## 8             Basic 4     G   C        130     3   2    210   2.0
## 9         Bran Chex     R   C         90     2   1    200   4.0
## 10        Bran Flakes    P   C         90     3   0    210   5.0

str(df)

## 'data.frame':   77 obs. of  16 variables:
## $ name      : chr  "100% Bran" "100% Natural Bran" "All-Bran" "All-Bran with Extra Fiber" ...
## $ mfr       : chr  "N" "Q" "K" "K" ...
## $ type      : chr  "C" "C" "C" "C" ...
## $ calories: int   70 120 70 50 110 110 110 130 90 90 ...
## $ protein:  int    4  3  4  4  2  2  3  2  3 ...
## $ fat       : int    1  5  1  0  2  2  0  2  1  0 ...
## $ sodium:   int   130 15 260 140 200 180 125 210 200 210 ...
## $ fiber     : num   10  2  9 14 1 1.5 1 2 4 5 ...
## $ carbo     : num    5  8  7  8 14 10.5 11 18 15 13 ...
## $ sugars    : int    6  8  5  0  8 10 14  8  6  5 ...
## $ potass    : int   280 135 320 330 -1 70 30 100 125 190 ...
## $ vitamins: int    25  0 25 25 25 25 25 25 25 25 ...
## $ shelf     : int    3  3  3  3 1 2 3 1 3 ...
## $ weight    : num    1  1  1  1 1 1 1.33 1 1 ...
## $ cups      : num   0.33 1 0.33 0.5 0.75 0.75 1 0.75 0.67 0.67 ...
## $ rating    : num   68.4 34 59.4 93.7 34.4 ...
```

#Print out descriptive statistics for a selection of quantitative and categorical variables ##Quantitative variables **cereal calories**

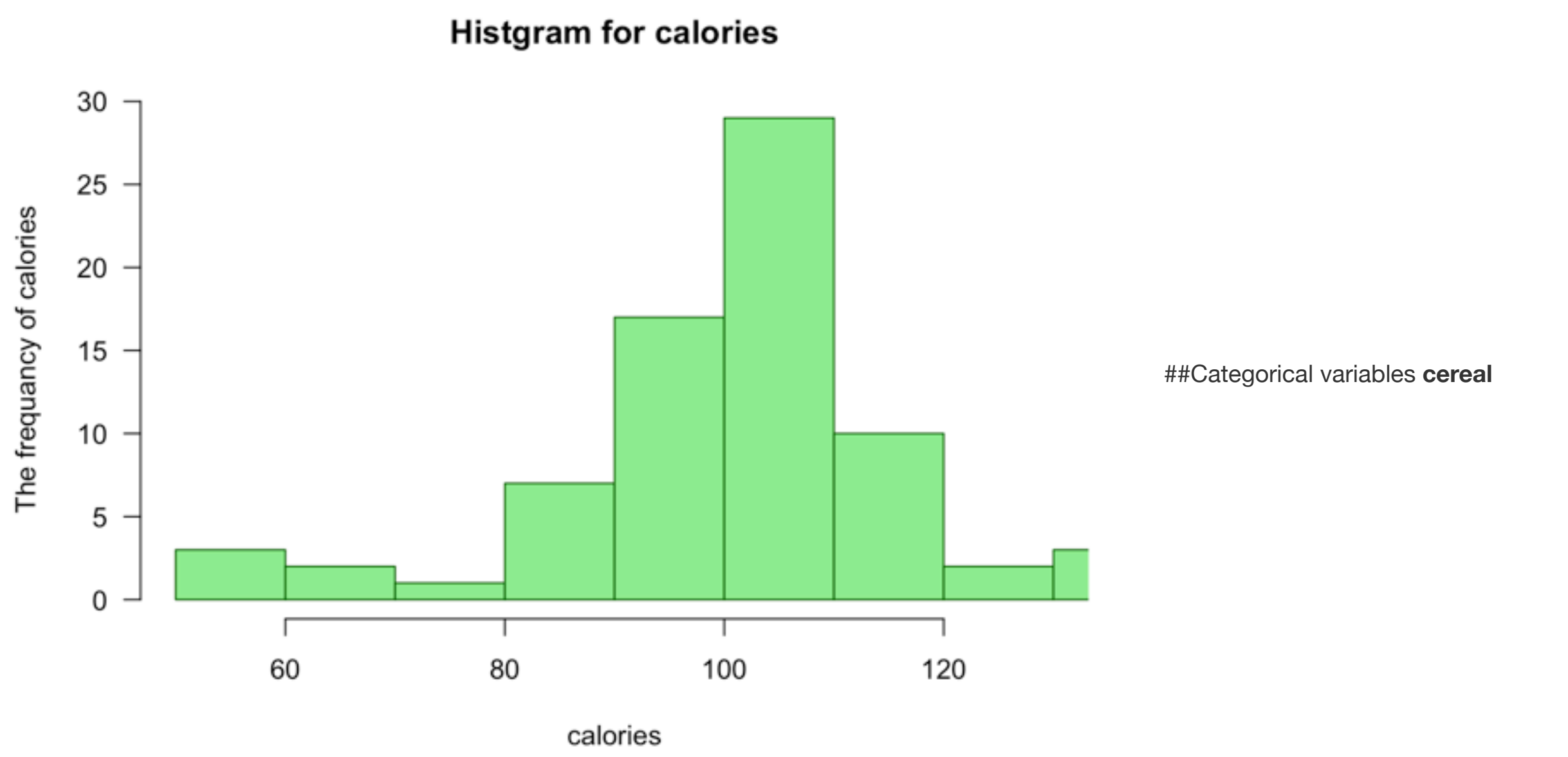
```
class(df$calories)

## [1] "integer"

summary(df$calories)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      50.0   100.0   110.0   106.9   110.0   160.0

hist(df$calories,
     main="Histogram for calories",
     xlab="calories",
     border = "darkgreen",
     col = "lightgreen",
     xlim = c(50,130),
     ylab = "The frequency of calories",
     las=1,
     breaks = 10)
```



manufacture

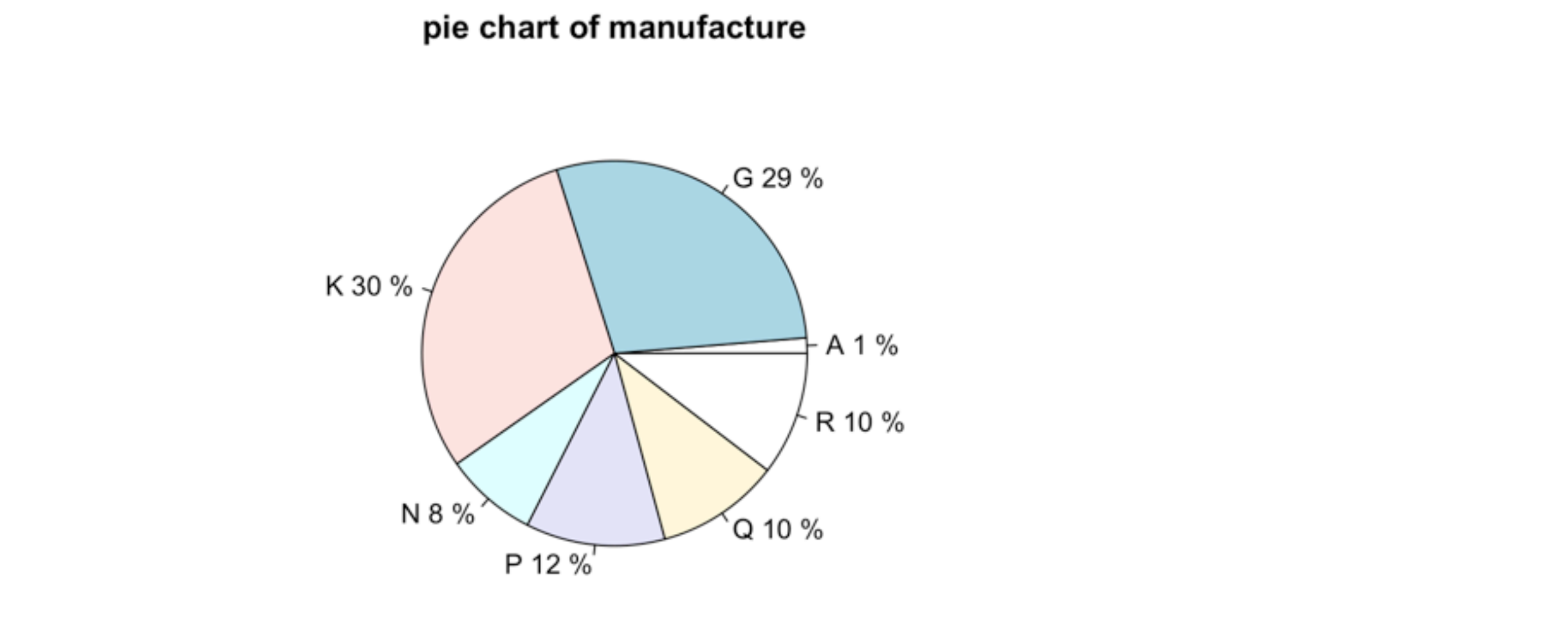
```
class(df$mfr)

## [1] "character"

table(df$mfr)

##
##  A  G  K  N  P  Q  R
##  1 22 23  6  9  8  8

mypct= round((table(df$mfr))/(sum(table(df$mfr)))*100)
lbls=paste(names(table(df$mfr)),mypct,"%")
pie(table(df$mfr),lbls,main = "pie chart of manufacture")
```



#Transform at least one variable

```
class(df$mfr)

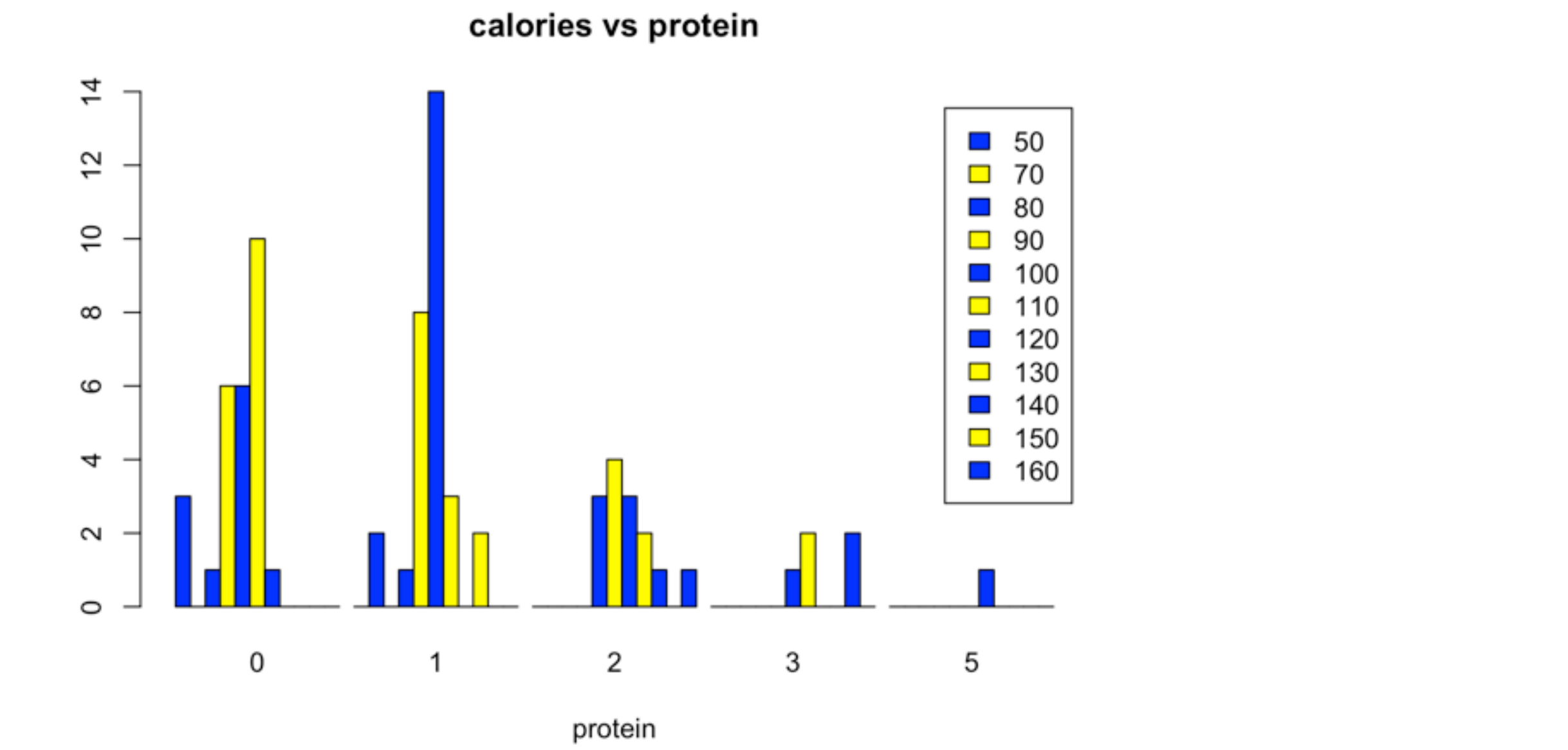
## [1] "character"

class(as.factor(df$type))

## [1] "factor"
```

#Plot at least one quantitative variable, and one scatterplot

```
counts <- table(df$calories,df$fat)
barplot(counts, main= "calories vs protein",
        xlab="protein",
        col=c("blue","yellow"),
        legend=rownames(counts), beside=TRUE)
```



```
plot(df$calories)
```

