Barcharts

JuJu

## Factor Variables

R calls categorical or qualitative variables, factor variables. When reading CSV files, R makes character variables into factor variables by default. However, when a factor variable has been coded with numbers, R assumes that the variable is quantitative. The ***HTWT*** data shows how this can be a problem.

 htwt = read.csv("http://facweb1.redlands.edu/fac/jim\_bentley/Data/FYS04/HtWt.csv")
 summary(htwt)

## Height Weight Group
## Min. :51.0 Min. : 82.0 Min. :1.00
## 1st Qu.:56.0 1st Qu.:108.2 1st Qu.:1.00
## Median :59.5 Median :123.5 Median :2.00
## Mean :62.1 Mean :139.6 Mean :1.55
## 3rd Qu.:68.0 3rd Qu.:166.8 3rd Qu.:2.00
## Max. :79.0 Max. :228.0 Max. :2.00

Note that the variable ***Group*** has been treated as numeric. It turns out that this variable actually represents the sex of the individual and that males were coded as 1 and females as 2. We convert the numeric variable to a factor variable.

 is.numeric(htwt$Group)

## [1] TRUE

 is.factor(htwt$Group)

## [1] FALSE

 table(htwt$Group)

##
## 1 2
## 9 11

 htwt$Group = factor(htwt$Group, labels=c("Male","Female"))
 is.numeric(htwt$Group)

## [1] FALSE

 is.factor(htwt$Group)

## [1] TRUE

 summary(htwt$Group)

## Male Female
## 9 11

 table(htwt$Group)

##
## Male Female
## 9 11

R uses factor variables to keep track of ordinal data. The ***ordered*** argument should be set to ***TRUE***. We will use data on phone service satisfaction to show how this works.

 phone = c(rep("Poor",840),rep("Fair",1649),rep("Good",4787),rep("Excellent",3208))
 # At this point phone is a list of strings and not a factor
 is.factor(phone)

## [1] FALSE

 # Use the function factor to convert the variable
 phone.u = factor(phone)
 is.factor(phone.u)

## [1] TRUE

 table(phone.u)

## phone.u
## Excellent Fair Good Poor
## 3208 1649 4787 840

 # Note that the output is alphabetical and not properly ordered
 # Recreate phone as an ordered factor variable
 phone.o = factor(phone, levels = c("Poor","Fair","Good","Excellent"), ordered=TRUE)
 table(phone.o)

## phone.o
## Poor Fair Good Excellent
## 840 1649 4787 3208

 # The values in the table are now ordered

## Barcharts

We now create plots to go with the tables.

 # Use base graphics
 barplot(table(htwt$Group))



 barplot(table(phone.u), main="Unordered Factor", col="red")



 barplot(table(phone.o), main="Ordered Factor", col="lightblue")
 # Use lattice plots
 p\_load(lattice)



 histogram(~phone.u)



 histogram(~phone.o)



 barchart(phone.o)



For those who just will not get rid of those stupid pie charts, R will make them. Why anyone would want to is a mystery.

 # Use base graphics
 pie(table(htwt$Group))



 pie(table(phone.u))



 pie(table(phone.o))



 # Can't use lattice since it won't make pie charts