Mosaic Plots

Oliver

Mosaic plots extend the barchart to help us see relationships between multiple categorical variables. We will use the Titanic data to demonstrate the ways in which these plots can be used to visualize how different variables are associated with survival.

First we need to read the Titanic data. Using **read.csv** makes this easy. We also need to code the factor/categorical variables so that they are easier to read.

 titanic = read.csv("http://facweb1.redlands.edu/fac/jim\_bentley/downloads/math111/titanic.csv")
 names(titanic)

## [1] "CLASS" "AGE" "SEX" "SURVIVED"

 titanic$AGE=factor(titanic$AGE,labels=c('Child','Adult'))
 titanic$CLASS=factor(titanic$CLASS,labels=c('0','1','2','3'))
 titanic$SEX=factor(titanic$SEX, labels=c('Female','Male'))
 titanic$SURVIVED=factor(titanic$SURVIVED,labels=c('No','Yes'))

A table of survival by sex is a good place to start.

 attach(titanic)
 table(titanic[,c("SURVIVED","SEX")])

## SEX
## SURVIVED Female Male
## No 126 1364
## Yes 344 367

A mosaic plot that shows this relationship is:

 mosaicplot(table(titanic[,c("SURVIVED","SEX")]),
 main="Survival on the Titanic",
 col=hcl(c(240,120)), off=c(5,5,5,5))



Similarly, we can look at survival by age.

 table(titanic[,c("SURVIVED","AGE")])

## AGE
## SURVIVED Child Adult
## No 52 1438
## Yes 57 654

A mosaic plot for this relationship:

 mosaicplot(table(titanic[,c("SURVIVED","AGE")]),
 main="Survival on the Titanic",
 col=hcl(c(240,120)), off=c(5,5,5,5))



We can also look at age and sex as predictors of survival.

 table(titanic[,c("SURVIVED","AGE","SEX")])

## , , SEX = Female
##
## AGE
## SURVIVED Child Adult
## No 17 109
## Yes 28 316
##
## , , SEX = Male
##
## AGE
## SURVIVED Child Adult
## No 35 1329
## Yes 29 338

A mosaic plot for these variables is:

 mosaicplot(table(titanic[,c("SURVIVED","AGE","SEX")]),
 main="Survival on the Titanic",
 col=hcl(c(240,120)), off=c(5,5,5,5))



Swapping age and sex may make it easier to see relationships.

 mosaicplot(table(titanic[,c("SURVIVED","SEX","AGE")]),
 main="Survival on the Titanic",
 col=hcl(c(240,120)), off=c(5,5,5,5))



Class may also have been associated with survival. We can take a look at age and class as predictors.

 table(titanic[,c("SURVIVED","CLASS","AGE")])

## , , AGE = Child
##
## CLASS
## SURVIVED 0 1 2 3
## No 0 0 0 52
## Yes 0 6 24 27
##
## , , AGE = Adult
##
## CLASS
## SURVIVED 0 1 2 3
## No 673 122 167 476
## Yes 212 197 94 151

A mosaic plot for these variables is:

 mosaicplot(table(titanic[,c("SURVIVED","CLASS","AGE")]),
 main="Survival on the Titanic",
 col=hcl(c(360,240,120,60)), off=c(5,5,5,5))



Again, swapping variables may make it easier to see relationships.

 mosaicplot(table(titanic[,c("SURVIVED","AGE","CLASS")]),
 main="Survival on the Titanic",
 col=hcl(c(360,240,120,60)), off=c(5,5,5,5))

