

Example data frame: `bballgames03`. Data on all 2,430 major league baseball games played in the year 2003.

1. Working with dataframes.

```
> names(bballgames03)           # variable names of dataframe
> bballgames03[3,]              # third case, all variables
> bballgames03[,c(1,4)]        # first and fourth variable, all cases
> bballgames03[1:3,1:3]        # first three cases of first three variables
> bballgames03$day             # just the variable named day
> attach(bballgames03)        # makes the variables available as individual vectors
```

Assume in what follows that we have attached the dataframe.

2. One categorical variable. (Example: `bballgames03$home`)

```
> table(home)                  # counts number of cases in each level
> barplot(table(home))         # a barplot of a table
> pie(table(home))             # pie chart from a table, pie charts are evil
```

3. One quantitative variable. (Example: `bballgames03$attendance`)

```
> hist(attendance)            # histogram of variable
> hist(attendance,freq=F)     # density histogram
> stem(attendance)            # stem and leaf plot
> boxplot(attendance)         # boxplot
> summary(attendance)         # some summary statistics of variable
> mean(attendance,na.rm=T)    # mean of variable, remove missing values
> sd(attendance)              # standard deviation of variable
```

4. Two categorical variables. (Example: `bballgames03$day` and `bballgames03$nightday`)

```
> table(day,nightday)         # contingency table, rows day
> barplot(table(day,nightday)) # stacked barplot of a table
> barplot(table(day,nightday),beside=T) # side-by-side barplot of table
> plot(nightday~day)          # an alternate stacked barplot (mosaic plot)
> plot(day,nightday)          # same as previous
> prop.table(table(day,nightday),1) # proportions by row (rows sum to 100%)
> prop.table(table(day,nightday),2) # proportions by column
> prop.table(table(day,nightday)) # proportions over whole table
```

5. One categorical and one quantitative variable. (Example `bballgames03$timemin` and `bballgames$day`)

```
> boxplot(timemin~day)        # side-by-side boxplot (plot for each category of day)
> by(timemin,day,summary)     # apply function summary to timemin by group
```

6. Two quantitative variables. (Example: `bballgames03$vish` and `bballgames03$visscore`)

```
> plot(vish,visscore)         # scatterplot, first variable x
> plot(visscore~vish)         # scatterplot, y~x
> cor(vish,visscore)          # correlation coefficient
> l=lm(visscore~vish)         # computes the least-squares model line
> summary(l)                  # summarizes properties of regression object
> residuals(l)                # residuals of regression model
> abline(l)                   # adds regression line to scatterplot
> predict(l)                   # predicted y given by model
> anova(l)                    # ANOVA table for regression
```