

# Analysis of RT distributions with R

## Univariate data

---

**2.6** According to the NY Times, the top-selling MP3 players for August 2003 are as shown in the following table with their market share.

Assume that the total market share is \$22 million.

1. What percent is 'other'.
2. Find the dollar amount (not the percentage amount) for each company.
3. Make a bar chart, dot chart and pie chart of the data, including 'other' using the R commands `barplot`, `dotchart`, `pie`.

<b>Sales of MP3 players.</b>	
<b>Total \$ 22 million.</b>	
Apple	18 %
RCA	15 %
Rio	14.4 %
iRiver	13.5 %
Creative Labs	6.2 %

**2.9** Install the package 'UsingR' (by Verzani) using the following commands:

```
install.packages("UsingR") # this installs the package to the hdd
library("UsingR")         # this loads UsingR into memory
```

Now load the malpractice-award data set 'npdb' using `attach(npdb)`.

In this set the variable `ID` is an identification number unique to a doctor but not traceable back to the doctor. It allows a look at a doctor's malpractice record without sacrificing anonymity. The commands `table(npdb$ID)` create a table of malpractice awards for each of the 6.369 doctors.

What does the command `table(table(ID))` do, and why is this interesting?

## Analysis of RT distributions with R

---

**2.26** A *lag plot* of a data vector plots successive values of the data against each other. By using a *lag plot*, we can tell whether future values depend on previous values: if not, the graph is scattered; if so, there is often a pattern.

Making a lag plot (with lag 1) is quickly done with the indexing notation of a negative number. For example, these commands produce a lag plot of `x`:

```
n <- length(x)
plot(x[-n], x[-1])
```

Look at the lag plots of the following data sets:

```
x <- rnorm(100)
x <- sin(1:100)
```

---

**2.34** The data set `DDT (MASS)` contains independent measurements for the pesticide DDT on kale. Make a histogram and a boxplot of the data. From these estimate the mean and the standard deviation. Check your answers with the appropriate functions.

---

**2.45** Sometimes a data set is so skewed that it can help if we transform the data prior to looking at it. A common transformation for long-tailed data sets is to take the logarithm of the data. For example, the `exec.pay (UsingR)` data set is highly skewed. Look at histograms before and after taking a logarithmic transform. You can transform with the command `log(1+exec.pay, 10)`.

## Analysis of RT distributions with R

Which is better at showing the data and why? Find the median and the mean for the transformed data. How do they correspond to the median and the mean of the untransformed data?

### Bivariate data

---

**3.8** The use of a cell phone while driving is often thought to increase the chance of an accident. The data set `reaction.time` (`UsingR`) is simulated data on the time it takes to react to an external event while driving. Subjects with `control == "C"` are not using a cell phone, and those with `control == "T"` are. Their time to respond to some external event is recorded in seconds. Create side-by-side boxplots of the variable `reaction.time` for the two values of control. Compare the centers and spreads.

---

**3.22** The `galton` (`UsingR`) data set contains data recorded by Galton in 1885 on the heights of children and their parents. The data is discrete, so a simple scatterplot does not show all the data points. In this case, it is useful to jitter the points a little when plotting by adding a bit of noise to each point. The `jitter()` function will do this. An optional argument, `factor =`, allows us to adjust the amount of jitter. Plot the data as below and find a value for `factor=` that shows the data better.

```
attach(galton)
```

```
plot(jitter(parent, factor=1), jitter(child, factor=1))
```

# Analysis of RT distributions with R

## Linear regression

---

**3.26** The data frame `x77` contains data from each of the fifty United States. First coerce the `state.x77` variable into a data frame with

```
x77 <- data.frame(state.x77)
```

For each of the following models, make a scatterplot and add the regression line:

1. The model of illiteracy rate (`Illiteracy`) modeled by high school graduation rate (`HS.Grad`).
  2. The model of life expectancy (`Life.Exp`) modeled by (`Murder()`) the murder rate.
  3. The model of income (`Income`) modeled by the illiteracy rate (`Illiteracy`)
  4. The model of illiteracy (`Illiteracy`) modeled by the life expectancy (`Life.Exp`).
-