

Understanding Data Mining

Basics of R

What is R?

R is:

- a language and environment for statistical computing and graphics
- a provider of a wide variety of statistical and graphical techniques
- *free* software
- an environment designed around a true computer language

Basic R Commands with Brief Explanations of their Uses

Command	Use
set.seed()	To ensure results are always the same
help()	To get details about an R command or a data set in R
data()	To load a data set already contained in R
plot()	To generate a scatter plot
title ()	To give a scatter plot a main title
lm()	To create a linear model for a data set
lines ()	To add the graph of a linear model to a data set
cor ()	To find the value of the correlation coefficient of a data set
name<-	To assign a particular name to something, so we can easily recall it later in our code
read.table ()	To load a set of data not already contained in R
anova ()	To create an analysis of variance
step ()	To use forward, backward, or stepwise selection