Weka is Java based software used for data analysis and predictive modeling, R is a programming language...

R program: [http://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/iris_plots/](http://www2.warwick.ac.uk/fac/sci/moac/people/students/peter_cock/r/iris_plots/)

R program: [http://www.statmethods.net/advgraphs/index.html](http://www.statmethods.net/advgraphs/index.html)
Growing Popularity

• Noticed by major publications
  – InfoWorld; “R a Programming Language ‘on the rise’” (October 27, 2010)
  – Forbes; “Power in the Numbers” (May 6, 2010)
  – New York Times; “Data Analysts Captivated by R’s Power” (January 6, 2009)

• Industry polls show increasing use as a big data / data mining tool
  – KD Nuggets (2012) +32% to 31% used
  – Rexer Analytics (2011) close to half of all data miners (47%)
R Programming About

• R is a statistics package created by statisticians that can be compared to a spreadsheet without visual representation

• Similar MATLAB, SAS, SPSS, but favored by academia because it is open source

• Pros:
  – Programmable data analysis
  – Data visualization
  – Pre-processing data from different sources
  – Interactive: interpreted not complied

• Cons:
  – Graphical User Interface
  – Low-level programming
Using R on Gordon

SSH into Gordon using your XSEDE Portal Username and Password

Request and interactive node (for 30 minutes)

Type in the commands “module load R” and then “R” to start the R program
Sample shell script for running R in batch mode

```
#!/bin/sh
#PBS -q normal
#PBS -N run2
#PBS -l nodes=1:ppn=16:native
#PBS -l walltime=01:00:00
#PBS -o run1.out
#PBS -e run1.err
#PBS -V
#PBS -M your email address
#PBS -m abe
#PBS -A your-account-code

cd /home(userid)/your-R-code-directory

Type the following at the command prompt to start:

$ module load R
$ R CMD BATCH --no-save --no-restore myRscript.R
```
Basics

# arithmetic
# "#" is the comment character
# + * / - ^
# < <= > >= == != these all do what you think they do
# Boolean operators are !, & et |
# but you can write && or || instead of |
# and &: the result is then a scalar,
# even if the arguments are vectors
# Assignment operator "<-" or "=" stores the value (object)
x<- log(2.843432) *pi
x

Basic objects
• Vector
• List
• NULL

Vector types
• Numeric
• Logical
• Character
**Environment / Installing Packages; R(Weka) Package**

An R interface to Weka (Version 3.7.7) - RWeka contains the interface code to Weka, a collection of machine learning algorithms for data mining tasks written in Java, containing tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

```r
> w <- read.arff(system.file("arff","weather.nominal.arff",package = "RWeka"))
> m <- J48(play~., data = w)
> m
J48 pruned tree
```

```
outlook = sunny
| humidity = high: no (3.0)
| humidity = normal: yes (2.0)
outlook = overcast: yes (4.0)
outlook = rainy
| windy = FALSE: yes (3.0)
| windy = TRUE: no (2.0)
```

Number of Leaves : 5
Size of the tree : 8
History

Created by Statisticians*

• 1976: John Chambers releases 1st version of S Language for statistics, stochastic simulation and data visualization
• 1995: Ross Ihaka and Robert Gentleman release R as GPL
• 1998: Comprehensive R Archive Network (CRAN) founded
• 2001: R News published for 1st time
• 2004: 1st useR! conference
• 2009: More than 1000 packages available on CRAN

*Daniel Mullensiefen Goldsmiths, University of London; “Using R for data analysis”
Other Intros / Tutorials

The R Manuals
http://www.stat.berkeley.edu/~spector/R.pdf

An Introduction to R
http://cran.r-project.org/doc/manuals/R-intro.html

The R Inferno
http://www.burns-stat.com/

Christopher J Fariss

http://r4stats.com/downloads/
Other Resources

The R Journal

R-bloggers
R news and tutorials contributed by (407) R bloggers

One R Tip A Day

Things I wish I'd known before I started using R
March 13, 2011
By e-bw

Web Application Development with R and Apache

Quick-R
accessing the power of R

statMethods blog

PACE
Predictive Analytics Center of Excellence