CS461
Bapa Rao & Russ Abbott
R Overview
What is R

• R Environment
  – Data manipulation, calculation, graphical display
• S Programming language
• Object-oriented
• Interactive
The Language

• Data Types and Data Structures
  – Vectors
  – Factors
    • Like enum, but each element can be a vector
    • Used to specify groupings
      – E.g., students by state
  – Lists
  – Arrays, Matrices
  – Data Frames

• Control Structures
  – If, while, for, repeat, break, next
  – Functions
Useful Functions

• Library()
  – List all loaded functions

• .packages()
  – List all loaded packages

• c()
  – Concatenate objects into a list
Built-in datasets

• Package datasets
• data()
Probability Distributions

**Distribution R name additional arguments**
beta beta shape1, shape2, ncp
binomial binom size, prob
Cauchy cauchy location, scale
chi-squared chisq df, ncp
exponential exp rate
F f df1, df2, ncp
gamma gamma shape, scale
geometric geom prob
hypergeometric hyper m, n, k
log-normal lnorm meanlog, sdlog
logistic logis location, scale
negative binomial nbinom size, prob
normal norm mean, sd
Poisson pois lambda
Student’s t t df, ncp
uniform unif min, max
Weibull weibull shape, scale
Wilcoxon wilcox m, n
Statistical Modeling

• Linear regression modeling is part of basic package
• Other regression models also available
• $Y = X*BETA + e$
  – $Y$ is dependent variable vector (response vector)
  – $X$ is independent variable matrix (model matrix or design matrix—column matrix)
  – $BETA$ is coefficient matrix
  – $e$ is error vector, Normal Independently Distributed with mean 0, all elements have same variance
• $y \sim x$
• `lm()`
• Generalized linear models `glm()`
  – Non-normal
• Nonlinear models
Packages

- `library()`
- `.Library()`
- `(.packages())`
- CRAN [http://cran.r-project.org/index.html](http://cran.r-project.org/index.html)
  - `install.packages("name-of-your-package", lib="~/R/library")`
  - `install.packages(“TTR”) #technical trading rules`
  - `Install.packages(“ttrc) #technical trading rules composite data`
  - `require("TTR")`
  - `require("ttrc")`
  - `X<-c(1.0,8.9,3.2,5.9,9.2,6)`
  - `SMA(x,n=3) #simple moving average`

```r
```
TTR: A Useful financial package

ibm<-getYahooData("IBM",20080117,20080118)
> ibm

    Open    High    Low   Close  Volume Unadj.Close Div Split
2008-01-17  98.2828 99.67996 96.40386 97.41560 12385361    101.1 NA    NA
2008-01-18 102.8308 102.83079 98.76458 99.63178 24774563    103.4 NA    NA

  Adj.Div
2008-01-17    NA
2008-01-18    NA

stockSymbols(exchange=c("AMEX","NASDAQ","NYSE"),sort.by=c("Exchange","Symbol"), quiet=FALSE)

• Full range of functions to support Technical trading rules
• SMA, EMA, MACD, ...
R and Machine Learning

• R is very strong on statistical methods
• There are a number of machine learning packages
  – http://cran.r-project.org/
  – Click task views and select MachineLearning
• Focus on fine-tuning and evaluating machine learning parameters
  – Neural nets, genetic algorithms, support vector machines, bayesian classifiers, and many others