Big Data Analytics with R

March 22, 2005

Introduction to R

The goal of the first part of this book is to get you up to speed with the basics of **R** as quickly as possible.

Installation

Follow the procedures according to your operating system.

- Linux: You need to have `blas` and `gfortran` installed on your Linux, for installing the `coin` package.
- *Rgraphviz* requires installation from `source("http://bioconductor.org/biocLite.R")`, then `biocLite("Rgraphviz")`.
- Uncomment the following lines for installing all missing packages (this will take some time):

R and RStudio

- R is a programming language for statistical computing and data analysis that supports a variety of programming styles. See R in Wikipedia
- R has multiple online resources and books.
- R coding style
- R-Bloggers
- Getting help in R
  - RStudio cheat sheet
  - Base R cheat sheet
  - Advanced R cheat sheet
  - Data Visualization cheat sheet
  - R Markdown cheatsheet
  - [R Markdown Basics] (http://rmarkdown.rstudio.com/authoring_basics.html)
  - `help("    ")` command
- R as a calculator. Console: It uses the command-line interface.

Examples

```r
x <- c(1,2,3,4,5,6)    # Create ordered collection (vector)
y <- x^2               # Square the elements of x
print(y)               # print (vector) y
```

```
## [1]  1  4  9 16 25 36
```

```r
mean(y)                # Calculate average (arithmetic mean,
```

```
## [1] 15.16667
```

```r
var(y)                 # Calculate sample variance
```

```
## [1] 178.9667
```

```r
lm_1 <- lm(y ~ x)      # Fit a linear regression model "y =
                       # store the results as lm_1
print(lm_1)            # Print the model from the (linear m
```

```
##
## Call:
## lm(formula = y ~ x)
```