Awesome R

A curated list of awesome R frameworks, packages and software. Inspired by awesome-machine-learning.

Integrated Development Environment

Integrated Development Environment

- RStudio - A powerful and productive user interface for R. Works great on Windows, Mac, and Linux.
- JGR - JGR (speak 'Jaguar') is a Java Gui for R.
- Emacs + ESS - Emacs Speaks Statistics is an add-on package for emacs text editors.
- StatET - An Eclipse based IDE (integrated development environment) for R.
- Revolution R Enterprise - Revolution R would be offered free to academic users and commercial software would focus on big data, large scale multiprocessor functionality.
- R Commander - A package that provides a basic graphical user interface.
- IPython - An interactive Python interpreter,and it supports execution of R code while capturing both output and figures.
- Deducer - A Menu driven data analysis GUI with a spreadsheet like data editor.
- Radiant - A platform-independent browser-based interface for business analytics in R, based on the Shiny package.

Syntax

Packages change the way you use R.

- magrittr - Let's pipe it.
- pipeR - Multi-paradigm Pipeline Implementation.
- lambda.r - Functional programming and simple pattern matching in R.

Data Manipulation

Packages for cooking data.

- dplyr - Blazing fast data frames manipulation and database query.

- data.table - Fast data manipulation in a short and flexible syntax.
- reshape2 - Flexible rearrange, reshape and aggregate data.
- readr - A fast and friendly way to read tabular data into R.
- tidyr - Easily tidy data with spread and gather functions.
- broom - Convert statistical analysis objects into tidy data frames.
- rlist - A toolbox for non-tabular data manipulation with lists.
- ff - Data structures designed to store large datasets.
- lubridate - A set of functions to work with dates and times.
- stringi - ICU based string processing package.
- stringr - Consistent API for string processing.

Graphic Displays

Packages for showing data.

- ggplot2 - An implementation of the Grammar of Graphics.
- ggvis - Interactive grammar of graphics for R.
- rCharts - Interactive JS Charts from R.
- lattice - A powerful and elegant high-level data visualization system.
- rgl - 3D visualization device system for R.
- Cairo - R graphics device using cairo graphics library for creating high-quality display output.
- extrafont - Tools for using fonts in R graphics.
- showtext - Enable R graphics device to show text using system fonts.
- dygraphs - Charting time-series data in R.
- rbokeh - R Interface to Bokeh.

Reproducible Research

Packages for literate programming.

- knitr - Easy dynamic report generation in R.
- xtable - Export tables to LaTeX or HTML.
- rapport - An R templating system.
- rmarkdown - Dynamic documents for R.
- slidify - Generate reproducible html5 slides from R markdown.
- Sweave - A package designed to write LaTeX reports using R

Web Technologies and Services

Packages to surf the web.

- shiny - Easy interactive web applications with R.
- RCurl - General network (HTTP/FTP/...) client interface for R.
- httpuv - HTTP and WebSocket server library.
- XML - Tools for parsing and generating XML within R.
- rvest - Simple web scraping for R.
- OpenCPU - HTTP API for R.

Parallel Computing

Packages for parallel computing.

- parallel - R started with release 2.14.0 which includes a new package parallel incorporating (slightly revised) copies of packages multicore and snow.
- Rmpi - Rmpi provides an interface (wrapper) to MPI APIs. It also provides interactive R slave environment.
- foreach - Executing the loop in parallel.
- SparkR - R frontend for Spark.

High Performance

Packages for making R faster.

- Rcpp - Rcpp provides a powerful API on top of R, make function in R extremely faster.
- Rcpp11 - Rcpp11 is a complete redesign of Rcpp, targetting C++11.
- compiler - speeding up your R code using the JIT

Language API

Packages for other languages.

- rJava - Low-level R to Java interface.
- jvmr - Integration of R, Java, and Scala.
- rJython - R interface to Python via Jython.
- rPython - Package allowing R to call Python.
- runr - Run Julia and Bash from R.
- RJulia - R package Call Julia.
- RinRuby - a Ruby library that integrates the R interpreter in Ruby.
- R.matlab - Read and write of MAT files together with R-to-MATLAB connectivity.
- RcppOctave - Seamless Interface to Octave and Matlab.
- RSPerl - A bidirectional interface for calling R from Perl and Perl from R.
- V8 - Embedded JavaScript Engine.
- htmlwidgets - Bring the best of JavaScript data visualization to R.
- rpy2 - Python interface for R.

Database Management

Packages for managing data.

- RODBC - ODBC database access for R.
- DBI - Defines a common interface between the R and database management systems.
- RMySQL - R interface to the MySQL database.
- ROracle - OCI based Oracle database interface for R.
- RPostgreSQL - R interface to the PostgreSQL database system.
- RSQLite - SQLite interface for R
- RJDBC - Provides access to databases through the JDBC interface.
- rmongodb - R driver for MongoDB.
- rredis - Redis client for R.
- RCassandra - Direct interface (not Java) to the most basic functionality of Apache Cassanda.
- RHive - R extension facilitating distributed computing via Apache Hive.
- RNeo4j - Neo4j graph database driver.

Machine Learning

Packages for making R cleverer.

- AnomalyDetection - AnomalyDetection R package from Twitter.
- h2o - Deeplearning, Random forests, GBM, KMeans, PCA, GLM
- Clever Algorithms For Machine Learning
- Machine Learning For Hackers
- rpart - Recursive Partitioning and Regression Trees
- randomForest - Breiman and Cutler's random forests for classification and regression
- lasso2 - L1 constrained estimation aka ‘lasso’
- gbm - Generalized Boosted Regression Models
- e1071 - Misc Functions of the Department of Statistics (e1071), TU Wien
- tgp - Bayesian treed Gaussian process models

- rgp - R genetic programming framework
- arules - Mining Association Rules and Frequent Itemsets
- frbs - Fuzzy Rule-based Systems for Classification and Regression Tasks
- rattle - Graphical user interface for data mining in R
- ahaz - Regularization for semiparametric additive hazards regression
- arules - Mining Association Rules and Frequent Itemsets
- bigrf - Big Random Forests: Classification and Regression Forests for Large Data Sets
- bigRR - Generalized Ridge Regression (with special advantage for p >> n cases)
- bmrm - Bundle Methods for Regularized Risk Minimization Package
- Boruta - A wrapper algorithm for all-relevant feature selection
- bst - Gradient Boosting
- C50 - C5.0 Decision Trees and Rule-Based Models
- caret - Classification and Regression Training
- CORElearn - Classification, regression, feature evaluation and ordinal evaluation
- CoxBoost - Cox models by likelihood based boosting for a single survival endpoint or competing risks
- Cubist - Rule- and Instance-Based Regression Modeling
- earth - Multivariate Adaptive Regression Spline Models
- elasticnet - Elastic-Net for Sparse Estimation and Sparse PCA
- ElemStatLearn - Data sets, functions and examples from the book: "The Elements of Statistical Learning, Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani and Jerome Friedman
- evtree - Evolutionary Learning of Globally Optimal Trees
- frbs - Fuzzy Rule-based Systems for Classification and Regression Tasks
- GAMBoost - Generalized linear and additive models by likelihood based boosting
- gamboostLSS - Boosting Methods for GAMLSS
- gbm - Generalized Boosted Regression Models
- glmnet - Lasso and elastic-net regularized generalized linear models
- glmpath - L1 Regularization Path for Generalized Linear Models and Cox Proportional Hazards Model
- GMMBoost - Likelihood-based Boosting for Generalized mixed models
- grplasso - Fitting user specified models with Group Lasso penalty
- grpreg - Regularization paths for regression models with grouped covariates
- hda - Heteroscedastic Discriminant Analysis
- ipred - Improved Predictors
- kernlab - kernlab: Kernel-based Machine Learning Lab
- klaR - Classification and visualization
- lars - Least Angle Regression, Lasso and Forward Stagewise
- lasso2 - L1 constrained estimation aka ‘lasso’
- LiblineaR - Linear Predictive Models Based On The Liblinear C/C++ Library
- LogicReg - Logic Regression
- maptree - Mapping, pruning, and graphing tree models
- mboost - Model-Based Boosting
- mvpart - Multivariate partitioning
- ncvreg - Regularization paths for SCAD- and MCP-penalized regression models
- nnet - eed-forward Neural Networks and Multinomial Log-Linear Models
- oblique.tree - Oblique Trees for Classification Data
- pamr - Pam: prediction analysis for microarrays
- party - A Laboratory for Recursive Partytioning
- partykit - A Toolkit for Recursive Partytioning
- penalized - L1 (lasso and fused lasso) and L2 (ridge) penalized estimation in GLMs and in the Cox model
- penalizedLDA - Penalized classification using Fisher's linear discriminant
- penalizedSVM - Feature Selection SVM using penalty functions
- quantregForest - quantregForest: Quantile Regression Forests
- randomForest - randomForest: Breiman and Cutler's random forests for classification and regression
- randomForestSRC - randomForestSRC: Random Forests for Survival, Regression and Classification (RF-SRC)
- rda - Shrunken Centroids Regularized Discriminant Analysis
- rdetools - Relevant Dimension Estimation (RDE) in Feature Spaces
- REEMtree - Regression Trees with Random Effects for Longitudinal (Panel) Data
- relaxo - Relaxed Lasso
- rgenoud - R version of GENetic Optimization Using Derivatives

- rgp - R genetic programming framework
- Rmalschains - Continuous Optimization using Memetic Algorithms with Local Search Chains (MA-LS-Chains) in R
- rminer - Simpler use of data mining methods (e.g. NN and SVM) in classification and regression
- ROCR - Visualizing the performance of scoring classifiers
- RoughSets - Data Analysis Using Rough Set and Fuzzy Rough Set Theories
- rpart - Recursive Partitioning and Regression Trees
- RPMM - Recursively Partitioned Mixture Model
- RSNNS - Neural Networks in R using the Stuttgart Neural Network Simulator (SNNS)
- RWeka - R/Weka interface
- RXshrink - RXshrink: Maximum Likelihood Shrinkage via Generalized Ridge or Least Angle Regression
- sda - Shrinkage Discriminant Analysis and CAT Score Variable Selection
- SDDA - Stepwise Diagonal Discriminant Analysis
- svmpath - svmpath: the SVM Path algorithm
- tgp - Bayesian treed Gaussian process models
- tree - Classification and regression trees
- varSelRF - Variable selection using random forests
- xgboost - eXtreme Gradient Boosting Tree model, well known for its speed and performance.
- SuperLearner and subsemble - Multi-algorithm ensemble learning packages.
- Introduction to Statistical Learning
- BreakoutDetection - Breakout Detection via Robust E-Statistics from Twitter.
- igraph - A collection of network analysis tools.

Natural Language Processing

Packages for Natural Language Processing.

- tm - A comprehensive text mining framework for R.
- openNLP - Apache OpenNLP Tools Interface.
- koRpus - An R Package for Text Analysis.
- zipfR - Statistical models for word frequency distributions.
- tmcn - A Text mining toolkit for international characters especially for Chinese.
- rmmseg4j - R interface to the Java Chinese word segmentation system of mmseg4j.
- Rwordseg - Chinese word segmentation.

Bayesian

Packages for Bayesian Inference.

- coda - Output analysis and diagnostics for MCMC.
- mcmc - Markov Chain Monte Carlo.
- MCMCpack - Markov chain Monte Carlo (MCMC) Package.
- R2WinBUGS - Running WinBUGS and OpenBUGS from R / S-PLUS.
- BRugs - R interface to the OpenBUGS MCMC software.
- rjags - R interface to the JAGS MCMC library.
- rstan - R interface to the Stan MCMC software.

Finance

Packages for dealing with money.

- quantmod - Quantitative Financial Modelling & Trading Framework for R.
- TTR - Functions and data to construct technical trading rules with R.
- PerformanceAnalytics - Econometric tools for performance and risk analysis.
- zoo - S3 Infrastructure for Regular and Irregular Time Series.
- xts - eXtensible Time Series.
- tseries - Time series analysis and computational finance.
- fAssets - Analysing and Modelling Financial Assets.

Bioinformatics

Packages for processing biological datasets.

- Bioconductor - Tools for the analysis and comprehension of high-throughput genomic data.
- genetics - Classes and methods for handling genetic data.
- gap - An integrated package for genetic data analysis of both population and family data.
- ape - Analyses of Phylogenetics and Evolution.
- pheatmap - Pretty heatmaps made easy.

R Development

Packages for packages.

- devtools - Tools to make an R developer's life easier.
- testthat - An R package to make testing fun.
- R6 - simpler, faster, lighter-weight alternative to R's built-in classes.
- pryr - Make it easier to understand what's going on in R.
- roxygen - Describe your functions in comments next to their definitions.
- lineprof - Visualise line profiling results in R.
- packrat - Make your R projects more isolated, portable, and reproducible.
- installr - Functions for installing softwares from within R.
- Rocker - R configurations for Docker.

Other Interpreter

Alternative R engines.

- renjin - a JVM-based interpreter for R.
- pqR - a "pretty quick" implementation of R
- fastR - FastR is an implementation of the R Language in Java atop Truffle and Graal.
- riposte - a fast interpreter and JIT for R.
- TERR - TIBCO Enterprise Runtime for R.
- RRE - Revolution R Enterprise.
- CXXR - Refactorising R into C++.

Resources

Where to discover new R-esources.

Websites

- R-project - The R Project for Statistical Computing.
- R Bloggers - There are people scattered across the Web who blog about R. This is simply an aggregator of many of those feeds.
- DataCamp - Learn R data analytics online.
- Quick-R - An excellent quick reference.
- Advanced R - An in-progress book site for Advanced R.
- CRAN Task Views - Task Views for CRAN packages.
- The R Programming Wikibook - A collaborative handbook for R.

Books

- The Art of R Programming - It's a good resource for systematically learning fundamentals such as types of objects, control statements, variable scope, classes and debugging in R.

- R in Action - This book aims at all levels of users, with sections for beginning, intermediate and advanced R ranging from "Exploring R data structures" to running regressions and conducting factor analyses.
- Use R! - This series of inexpensive and focused books from Springer publish shorter books aimed at practitioners. Books can discuss the use of R in a particular subject area, such as bayesian networks, ggplot2 and Rcpp.

Reference Card

- R Reference Card 2.0 - Material from R for Beginners by permission of Emmanuel Paradis (Version 2 by Matt Baggott).
- Data Mining Refcard - R Reference Card for Data Mining.
- Regression Analysis Refcard - R Reference Card for Regression Analysis.
- Reference Card for ESS - Reference Card for ESS.
- R Markdown Cheat sheet - Quick reference guide for writing reports with R Markdown.
- Shiny Cheat sheet - Quick reference guide for building Shiny apps.

Other Awesome Lists

- awesome-awesomeness
- lists

Contributing

Your contributions are always welcome!