



Enterprise R Package Development

Rob Carnell, Huntington,
Enterprise Analytics Director

4 October 2019

Outline

A tutorial on how to build packages in R that can be deployed to the Enterprise or to the Comprehensive R Archive Network (CRAN) including continuous integration, testing, and code coverage analysis.

- Quick Introduction to R
- Introduction to R packages
- Open Source Licenses in the Corporate Ecosystem
- Package Development Tech Stack





What does the R in Rob stand for?

A Little About Me

- Diverse Education
 - BS Physics
 - MS Statistics
 - MEM in Engineering Management
 - Nuclear Engineering in US Navy
- Data Science Experience
 - Research - Probabilistic Risk Assessment, National Security, Bioinformatics, Forensics, Transportation
 - Marketing
 - Banking
- Languages Spoken
 - R, c/C++, python, SAS, Java, C#, javascript, VBA, ASP.NET

Quick Introduction to R

-  - "R is a free software environment for statistical computing and graphics."
 - **Free** like "free beer", and **Free** like "freedom to modify, extend, and distribute" the source
 - Technically, a functional scripting language with object oriented programming options. C and Fortran under the hood.
-  Studio
 - An Integrated Development Environment (IDE) for programming in R

Introduction to R Packages

Packages in R "provide a mechanism for loading optional code, data and documentation as needed. The R distribution itself includes about 30 packages."

CRAN Packages:

- **CRAN** (The Comprehensive R Archive Network): A world-wide mirrored repository of curated extensions (packages) for the R environment.
- **Testing:** CRAN packages are tested on entry and periodically thereafter and are scanned by the CRAN team for use with designated R versions and designated dependent packages.
- **Security:** Code is open source and is usually inspected. Closed source 3rd party binary libraries are discouraged or not allowed. Packages are manually reviewed for a user's first package or if something unusual is requested.

Introduction to R Packages (cont)

Github Packages + other:

- More of the wild west
- Corporate users need to do due diligence before using a package from github if access is even allowed in the Enterprise
- Most often used as a "faster" way to distribute package bug fixes than the CRAN process
- Also allows package users to see the level of rigor around the package development: continuous integration, testing, code documentation, etc.

Use Case for Enterprise Package Development

- Great way to standardize, document, and distribute a package
- Can use internal git repository as the package repository
- Much easier for users to re-use validated code
- Automated unit testing and validation against new versions of R and dependent packages
- Fits better within a corporate governance structure as a self-contained tool, rather than a collection of scripts

Open Source Licenses in the Corporate Ecosystem

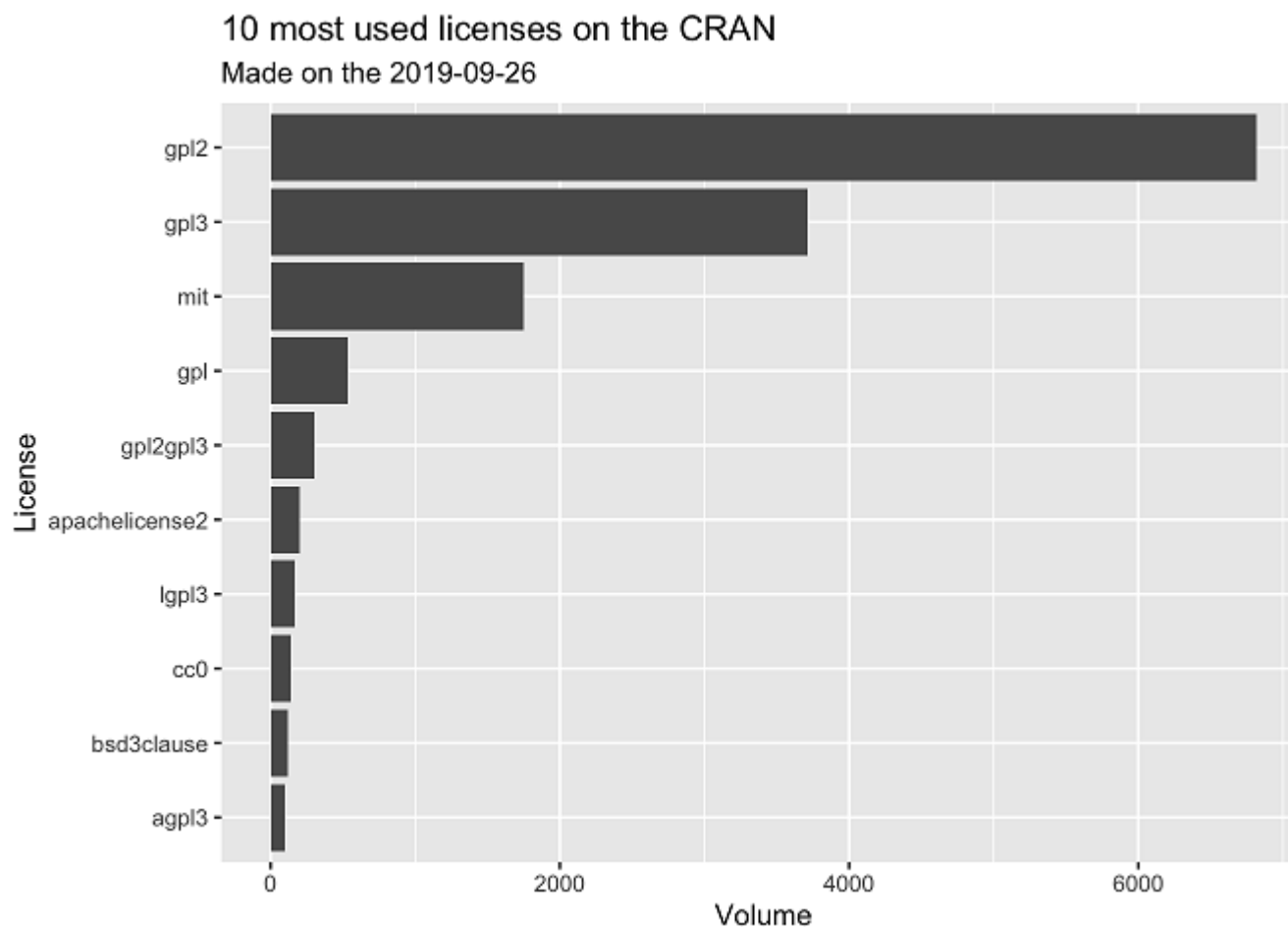
Open Source Licenses

- [R](#): GNU General Public License (GPL) version 2 or 3
- Most R packages are governed under:
 - MIT
 - GPL, LGPL, AGPL
 - BSD
 - CCBY-4
 - Apache 2.0

Please discuss with your company's legal department about specific policies on these licenses.

Read more at the [Free Software Foundation](#)

Most Popular CRAN Licenses



From [Here](#)

Package Development Tech Stack

Open Source Tech Stack:

- [R](#)
- Integrated Development Environment: [RStudio](#)
- Continuous Integration: [Travis CI](#)
- Test Coverage: [CodeCov](#)
- Source Control: [git](#)
- Documentation, Project Management, Collaboration: [github](#)

Analogs for each element can be implemented internally in the corporate environment or enterprise licenses can be bought



Demonstration of Package Environment

Learn More + R References

The gold standard reference from the R Project:

- [Writing R Extensions - R](#)

R packages:



- [roxygen2 - Inline R Documentation](#)
- [covr - Test Coverage for Packages](#)
- [devtools - Package development tools](#)

Learn More + Getting Started References

Others:

- [STAT 545 - University of British Columbia, Jenny Bryan](#)
- [The Political Methodologist](#)
- [devtools cheatsheet - RStudio](#)
- [Developing R Packages - jtleek](#)
- [R Package primer](#)
- [R Packages - Hadley Wickham](#)

Thank You

®; Huntington® and  Welcome® are federally registered service marks of Huntington Bancshares Incorporated. © 2019 Huntington Bancshares Incorporated.