

Collaborating at Scale: Managing an Enterprise Analytical Computing Ecosystem

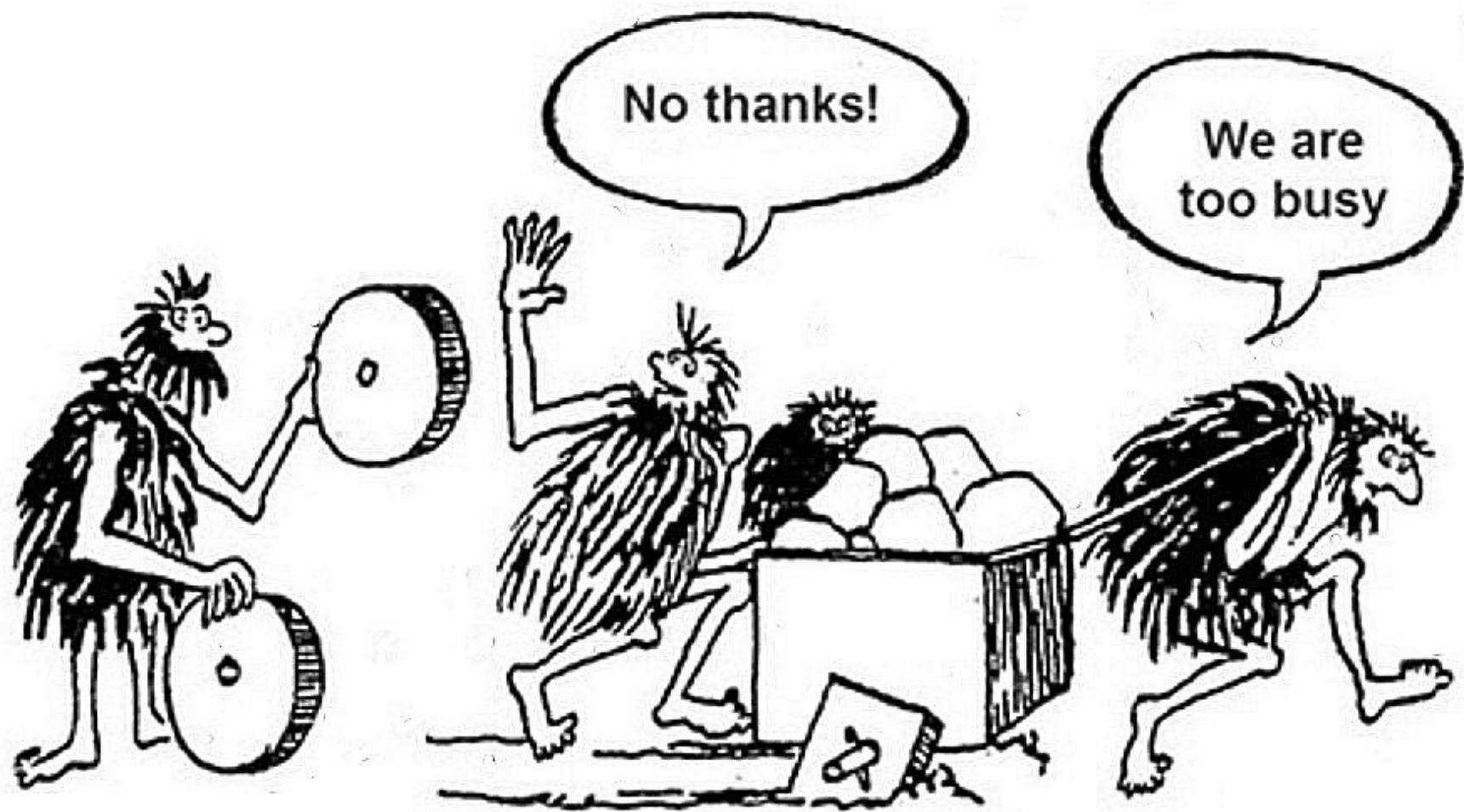
R/Pharma

August 22, 2019

Rena Yang

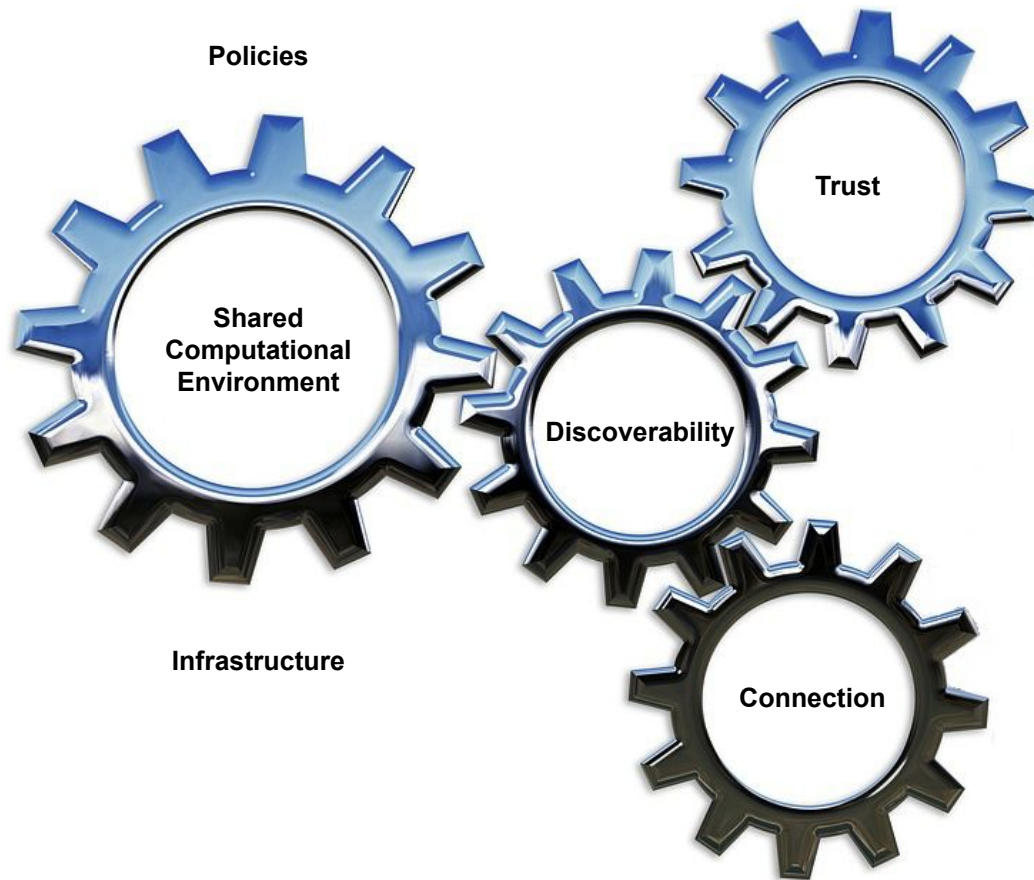


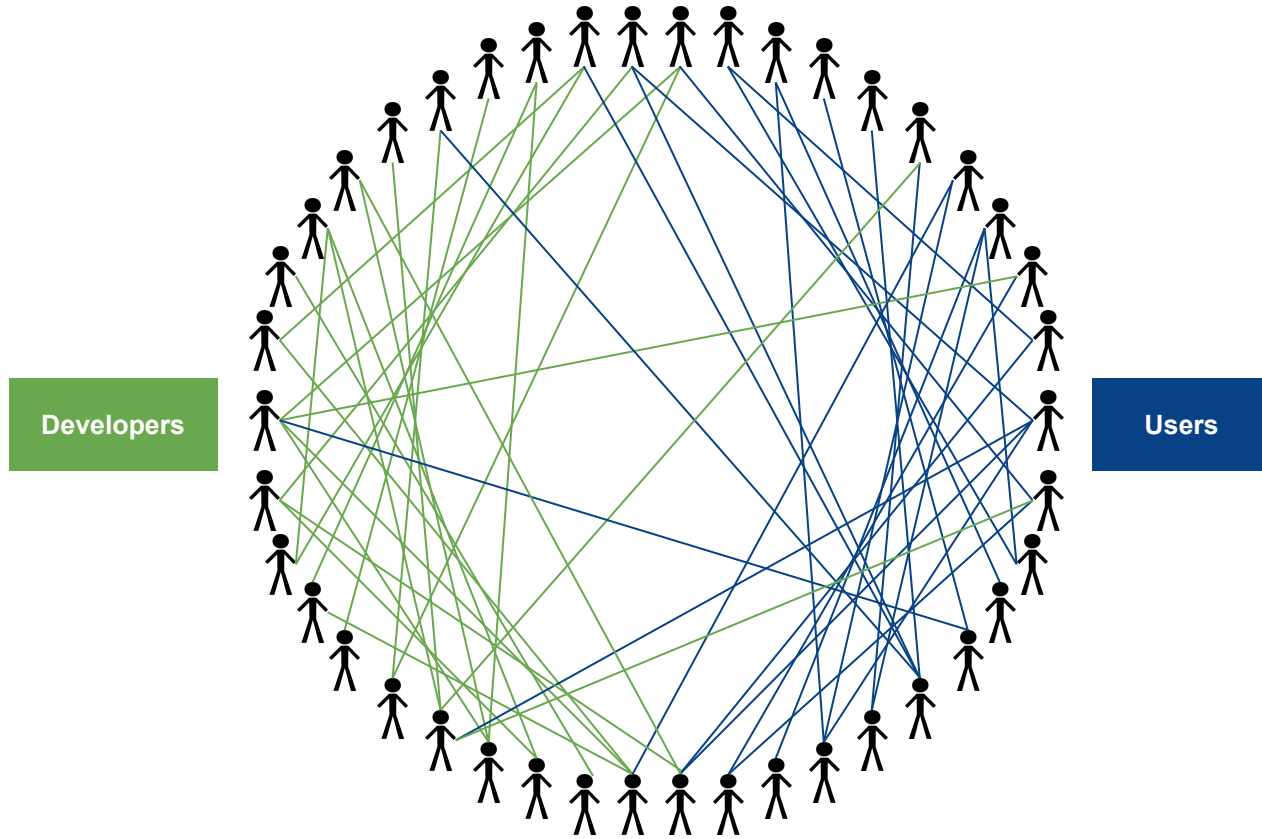
Computational Environments for
Data Analysis and Research

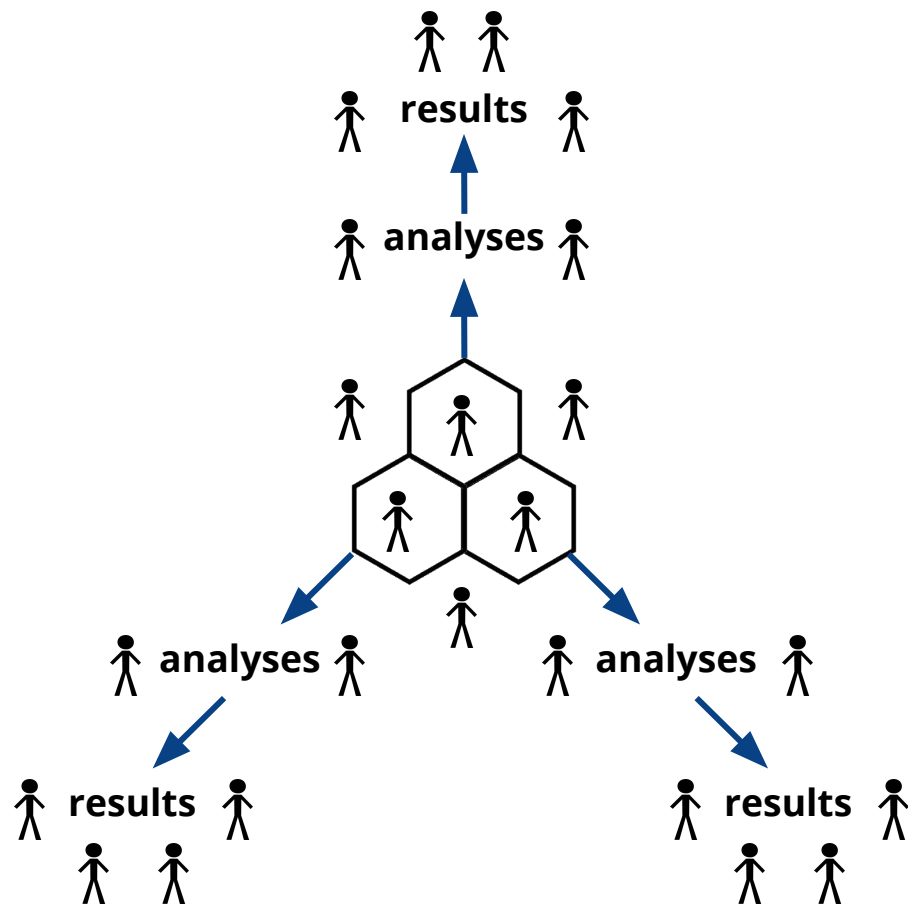












Building Blocks

- Package Publishing Portal
 - Helps users and developers to discover functionality
- Development Hub
 - Provides transparency to build trust
- Shared, Unified Analytical Computing Environments
 - Increases trust that functionality will work together
 - Facilitates collaboration around results
- User Analytics Data
 - Connects package developers with users
 - Connects users who share common tools and interests

Publication Portal

Publication Portal

Package
Landing Page

Package
Documentation












Connection

Publication Portal

Build details

Show 10 entries

Search:

Package Name	Last Attempt Version	Last Attempt Status	Last Attempt Date	Last Built Version	Last Built Status	Last Built Date	Maintainer	Coverage	Build History
ExpressionPlotCorePages	0.33.7	up-to-date	2019-08-20 19:46:35	0.33.4	check warning(s)	2019-05-25 20:22:11		89.7 %	Build log
ExpressionPlotGenomicPlots	0.2.10	up-to-date	2019-08-20 19:46:35	0.2.10	check warning(s)	2019-04-19 12:22:59		81.95 %	Build log
ExpressionPlotGenotype	0.1.32	up-to-date	2019-08-20 19:46:35	0.1.32	check warning(s)	2019-04-19 12:22:59		68.73 %	Build log
ExpressionPlotHttpRequest	0.1.7	up-to-date	2019-08-20 19:46:35	0.1.7	check warning(s)	2019-04-19 12:22:59		85.91 %	Build log
ExpressionPlotImportHTqPCR	0.6.43	up-to-date	2019-08-20 19:46:35	0.6.43	check warning(s)	2019-04-19 12:22:59		96.45 %	Build log
ExpressionPlotImportMicroarray	1.2.77	up-to-date	2019-08-20 19:46:35	1.2.77	check warning(s)	2019-04-22 14:57:59		92.89 %	Build log
ExpressionPlotImportRNASeq	0.99.120	up-to-date	2019-08-20 19:46:35	0.99.120	check warning(s)	2019-05-02 19:12:37		95.43 %	Build log
ExpressionPlotInstall	0.2.30	up-to-date	2019-08-20 19:46:35	0.2.30	check warning(s)	2019-05-16 21:22:38		44.84 %	Build log
ExpressionPlotLoadAffy	0.1.2	up-to-date	2019-08-20 19:46:35	0.1.2	check warning(s)	2019-04-19 12:22:59		81.54 %	Build log
ExpressionPlotLoadAgilent	0.3.2	up-to-date	2019-08-20 19:46:35	0.3.2	check note(s)	2019-04-19 12:22:59		84.31 %	Build log

Package Landing Page

gp.sa.diff: Testing for Differential Events over Lots of Features

Provides methods to test for differential events between samples in 'omics datasets containing many features. This includes differential expression for RNA-seq or microarray data, differential binding for ChIP-seq data, differential abundance for mass cytometry, differential interactions for Hi-C data, etc.

GRAN Release: [GRANtst](#)

Build status: [check note\(s\)](#)

Test Coverage: [93.65 %](#)

Authors: Aaron Lun [cre, aut]

Maintainer: Aaron Lun luna@gene.com



Installation

To install this package, start R and enter:

```
source("http://restst.gene.com/gran/getGRAN-tst.R")
library(GRANtst)
install_packages("gp.sa.diff", type="source")
```

Details

Package	gp.sa.diff
Version	0.99.27
Date	2019-08-13
Title	Testing for Differential Events over Lots of Features
Description	Provides methods to test for differential events between samples in 'omics datasets containing many features. This includes differential expression for RNA-seq or microarray data, differential binding for ChIP-seq data, differential abundance for mass cytometry, differential interactions for Hi-C data, etc.

Package Documentation

Package ‘gp.sa.diff’

August 13, 2019

Version 0.99.27

Date 2019-08-13

Title Testing for Differential Events over Lots of Features

Description Provides methods to test for differential events between samples in 'omics datasets containing many features. This includes differential expression for RNA-seq or microarray data, differential binding for ChIP-seq data, differential abundance for mass cytometry, differential interactions for Hi-C data, etc.

License Genentech, Inc.

Depends SummarizedExperiment,

Imports methods, gp.sa.core, S4Vectors, limma, edgeR

Suggests stats, BiocStyle, BiocGenerics, airway, affy, arrays,
testthat, knitr, rmarkdown, DESeq2

VignetteBuilder knitr

RoxygenNote 6.1.1

NeedsCompilation no

Author Aaron Lun [cre, aut]

Maintainer Aaron Lun <luna@gene.com>

Development Hub

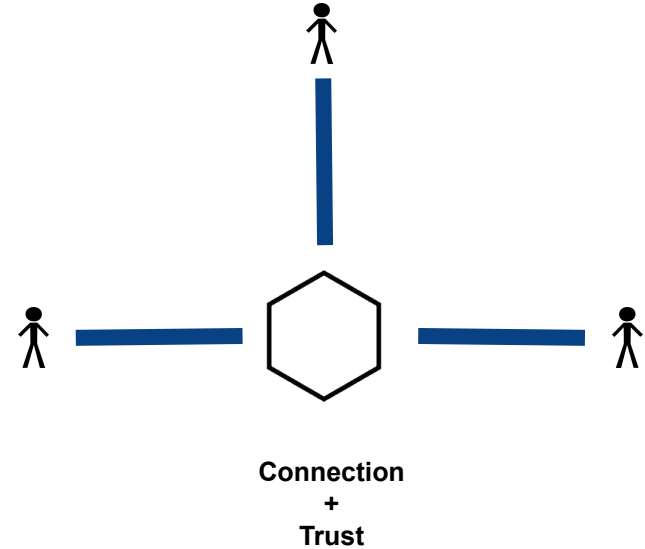
Development Hub

Automated Build,
Test, Install

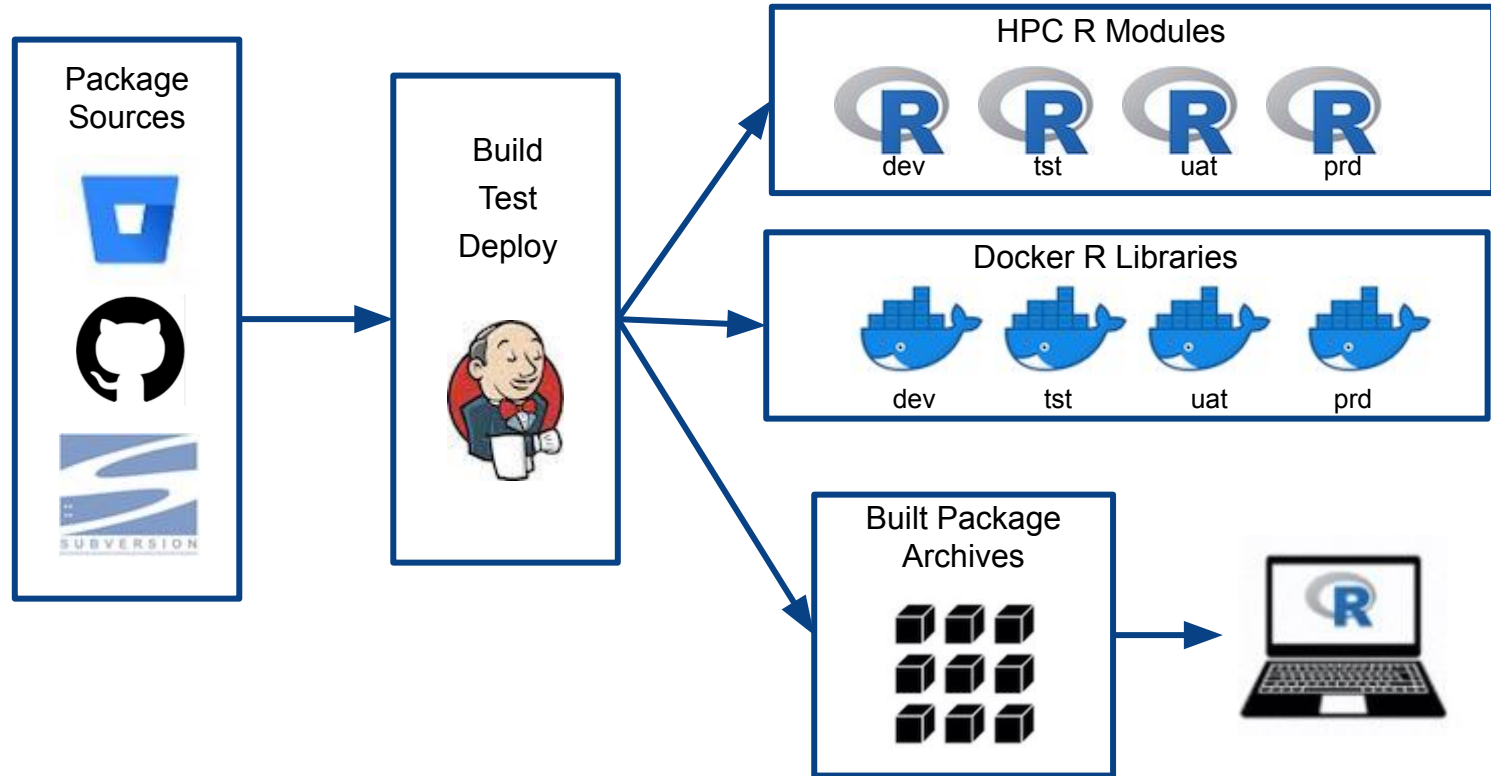
Build, Install,
Test Logs

Package
Source Control
Pointers

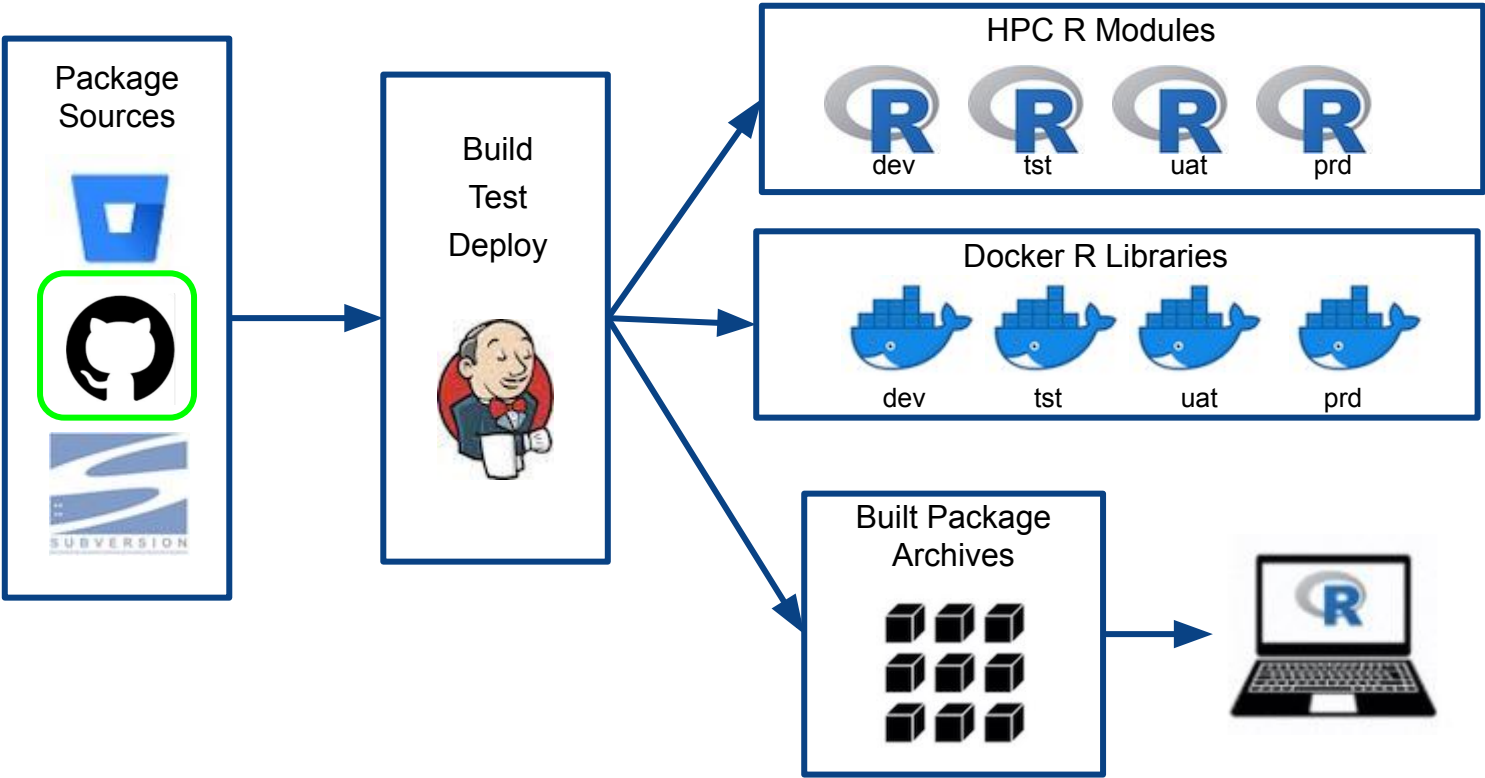
Code Coverage
Tools



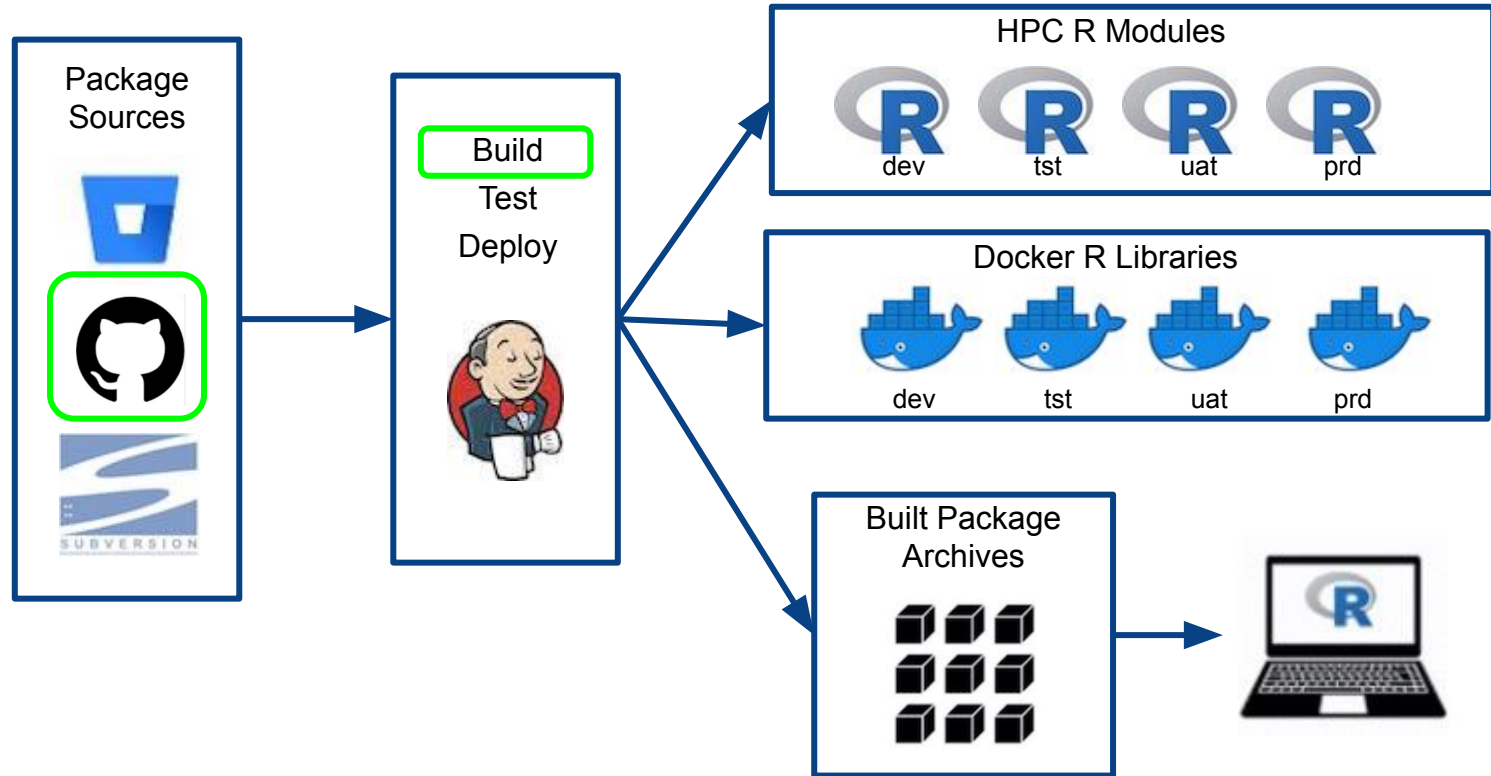
Build, Test, Deploy



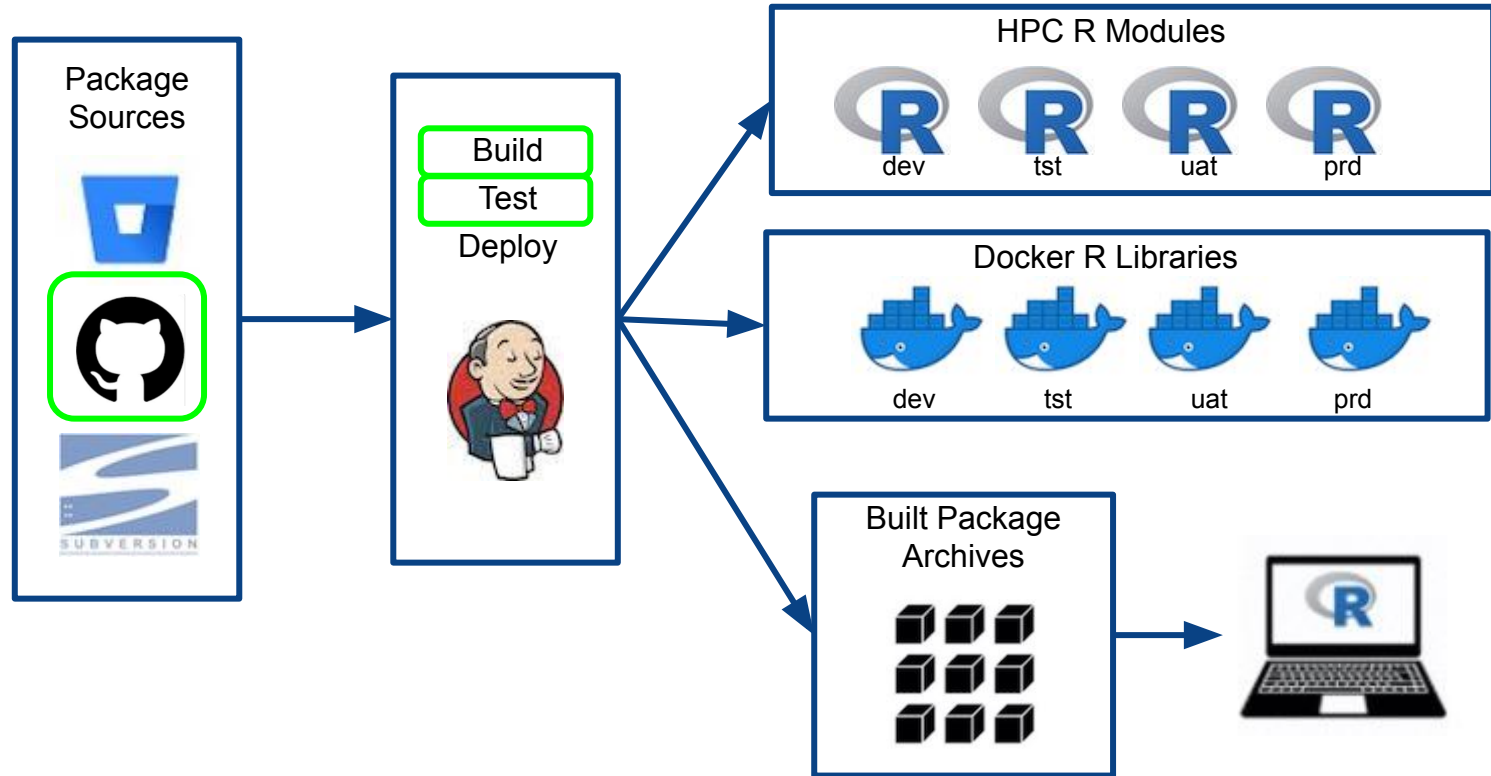
Build, Test, Deploy



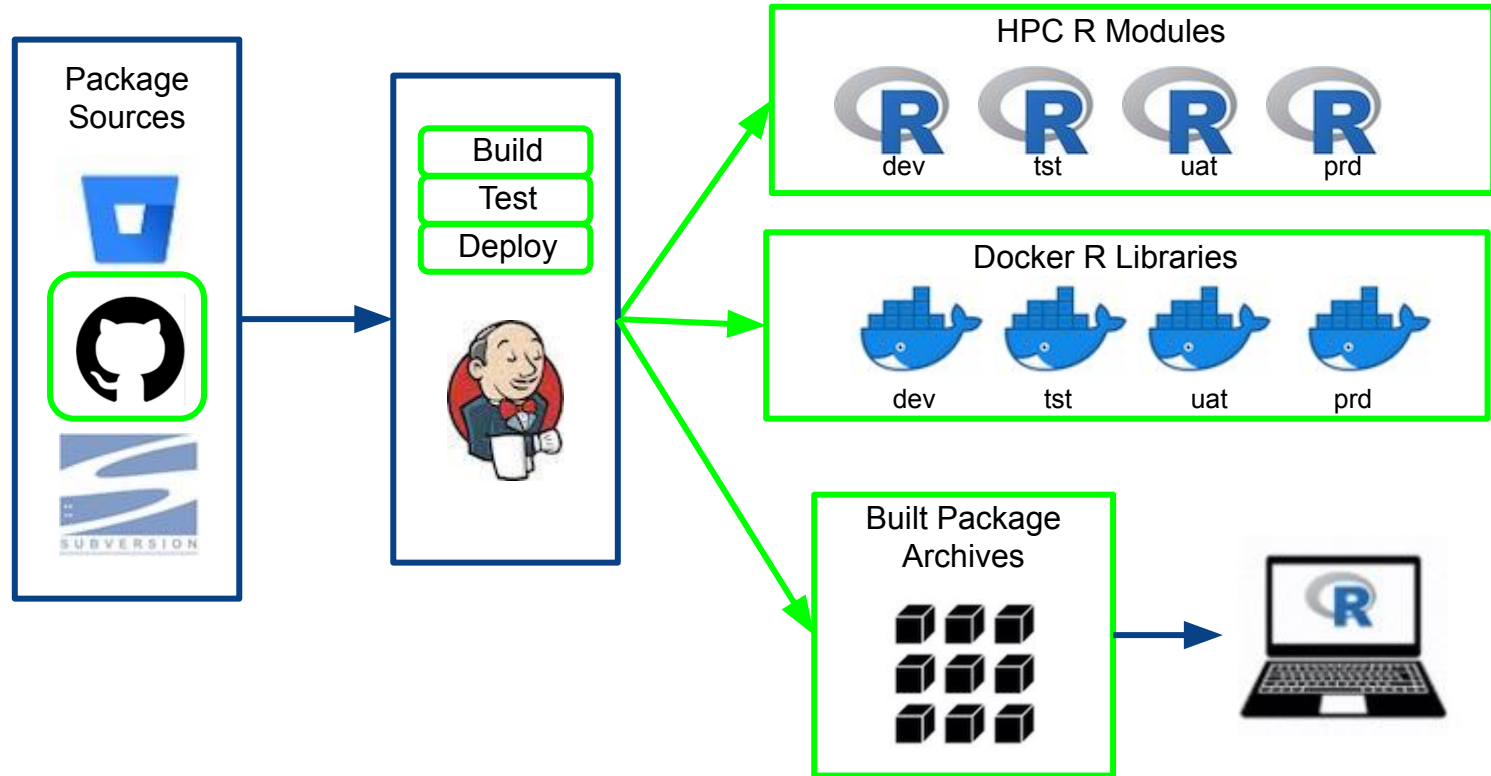
Build, Test, Deploy



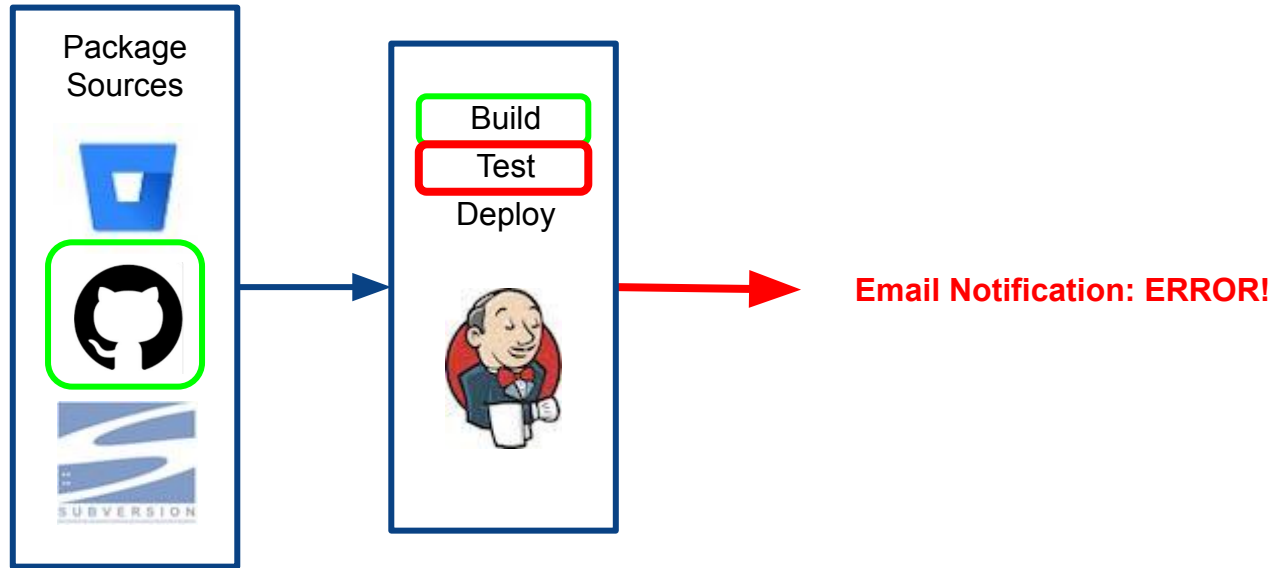
Build, Test, Deploy



Build, Test, Deploy



Build, Test, Deploy



Build, Test, Install Logs

```
* checking installed package size ... OK
* checking package directory ... OK
* checking 'build' directory ... OK
* checking DESCRIPTION meta-information ... WARNING
Non-standard license specification:
  Genentech, Inc.
Standardizable: FALSE
* checking top-level files ... OK
* checking for left-over files ... OK
* checking index information ... OK
* checking package subdirectories ... OK
* checking R files for non-ASCII characters ... OK
* checking R files for syntax errors ... OK
* checking whether the package can be loaded ... OK
* checking whether the package can be loaded with stated dependencies ... OK
* checking whether the package can be unloaded cleanly ... OK
* checking whether the namespace can be loaded with stated dependencies ... OK
* checking whether the namespace can be unloaded cleanly ... OK
* checking loading without being on the library search path ... OK
* checking dependencies in R code ... OK
* checking S3 generic/method consistency ... OK
* checking replacement functions ... OK
* checking foreign function calls ... OK
* checking R code for possible problems ... OK
* checking Rd files ... OK
* checking Rd metadata ... OK
* checking Rd cross-references ... OK
* checking for missing documentation entries ... OK
* checking for code/documentation mismatches ... OK
* checking Rd \usage sections ... OK
* checking Rd contents ... OK
* checking for unstated dependencies in examples ... OK
* checking installed files from 'inst/doc' ... OK
* checking files in 'vignettes' ... OK
* checking examples ... OK
* checking for unstated dependencies in 'tests' ... OK
* checking tests ...
  OK
* checking for unstated dependencies in vignettes ... OK
* checking package vignettes in 'inst/doc' ... OK
* checking running R code from vignettes ...
  'userguide.Rmd' using 'UTF-8' ... OK
  NONE
* checking re-building of vignette outputs ... SKIPPED
* checking PDF version of manual ... OK
* DONE
Status: 1 WARNING
```


Package Source Control Pointers

Build manifest for GRANTst

Show 10 entries

Search:

Package Name	SCM Repo	SCM Type	Branch	Package Subdirectory in SCM Repo
DataSetDB	ssh://git@bncb_r_packages/datasetdb.git	git	master	datasetdb
drusilla	ssh://git@divos/drusilla.git	git	master	.
DT.iTCGA	http://projects/ITCGA/trunk/plumbing/DT.iTCGA/	svn	trunk	.
ecg	ssh://git@bncb_r_packages/germlineeeg.git	git	devel	ecg
EnsemblRest	https://gmbecker/EnsemblRest	git	master	.
EPLog	ssh://git@ep/r.git	git	devel	./packages/EPLog/
EventPredGnr	https://Rpackages/EventPredGnr.git	git	master	.
eventTrack	https://rufiback/eventTrack.git	git	master	.
ExpressionCore	ssh://git@ep/r.git	git	devel	packages/ExpressionCore
ExpressionCoreDB	ssh://git@ep/r.git	git	devel	packages/ExpressionCoreDB

Test Code Coverage

ExpressionPlotImportRNASeq coverage - 95.43%

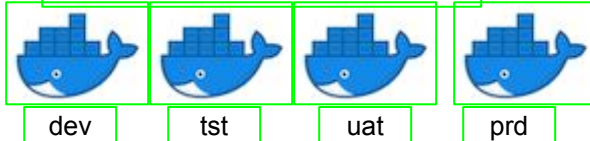
Files	Source						
File	Lines	Relevant	Covered	Missed	Hits / Line	Coverage	
R/ep.renormalize.rnaseq.R	41	15	13	2	1	86.67%	
R/add.data.from.DESeqDataSet.R	283	71	64	7	909	90.14%	
R/estimate.dispersions.R	89	32	29	3	75	90.62%	
R/calculate.mannwhitney.changes.R	201	78	71	7	138	91.03%	
R/calculate.rpkms.R	53	14	13	1	397	92.86%	
R/normalize.R	108	48	45	3	168	93.75%	
R/calculate.voomlimma.changes.R	321	111	106	5	190	95.50%	
R/calculate.changes.for.sumexp.R	449	144	138	6	479	95.83%	
R/ExpressionPlotProject.pileup.R	341	86	84	2	839	97.67%	
R/add.comps.for.rnaseq.R	173	63	63	0	387	100.00%	
R/add.data.from.cds.R	123	32	32	0	388	100.00%	
R/analyze.cds.R	142	30	30	0	251	100.00%	

Shared Unified Computing Environments

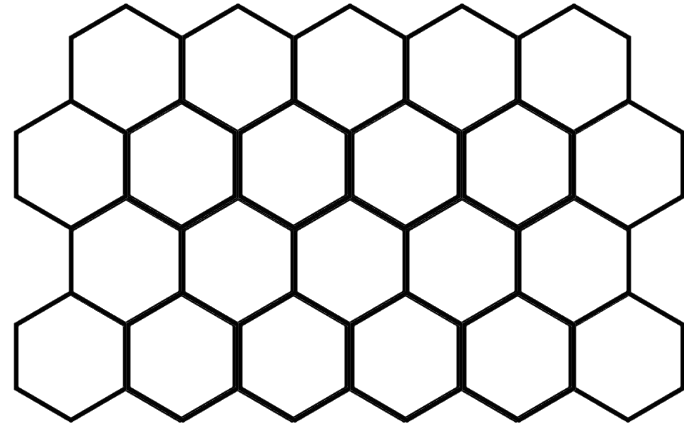
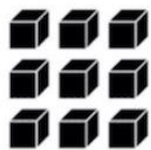
HPC R Modules



Docker R Libraries



Built Package Archives

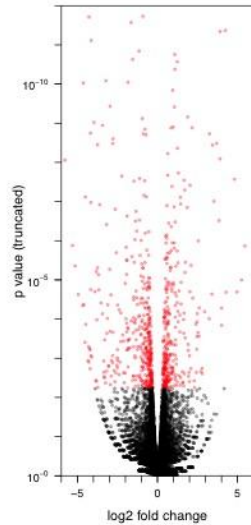


Package Ecosystem



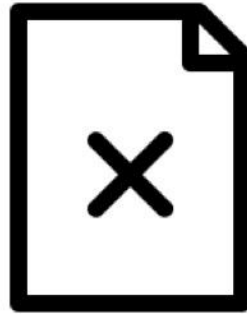
Comparability

Old DESeq



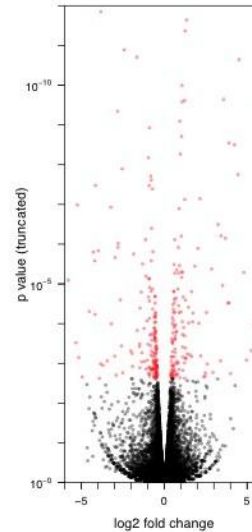
864 genes

Modern DESeq



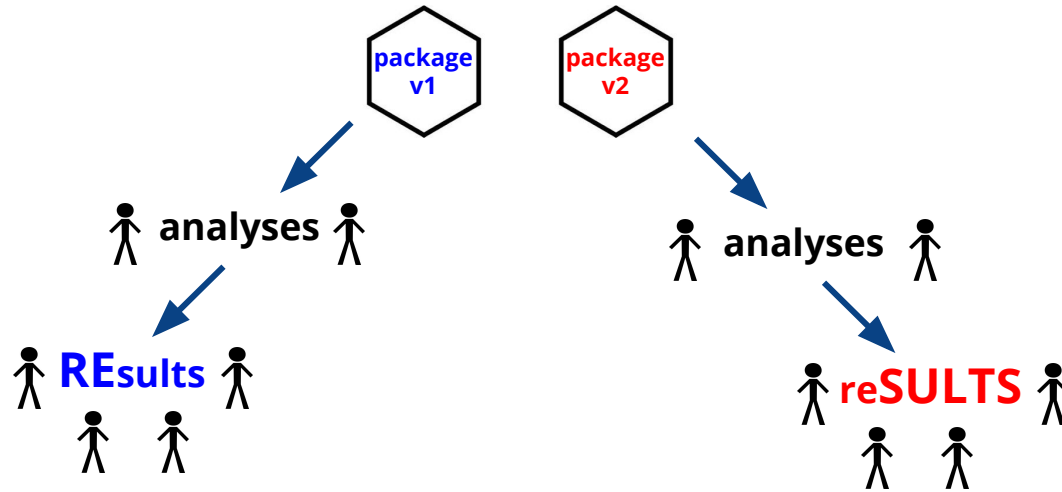
ERROR

Updated code,
Modern DESeq

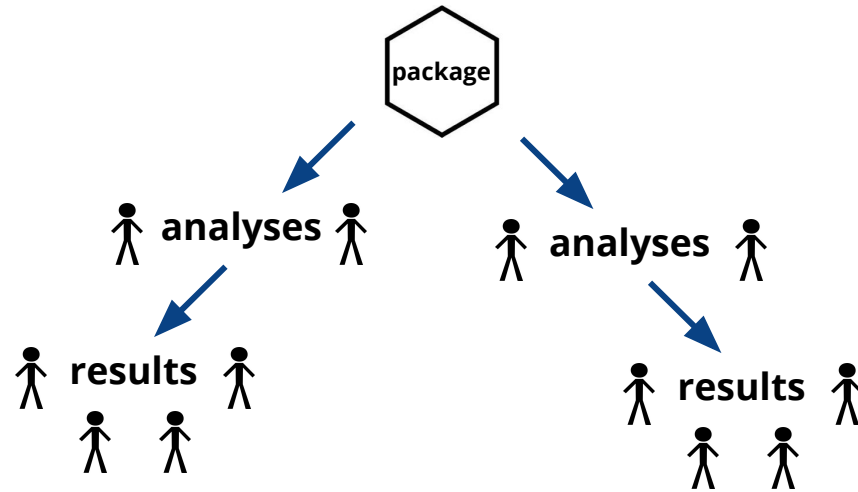


359 genes

Collaborating With Results



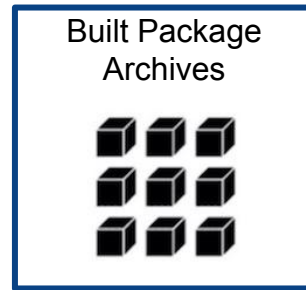
Collaborating With Results



Collaborating With Results Through Time

sessionInfo()

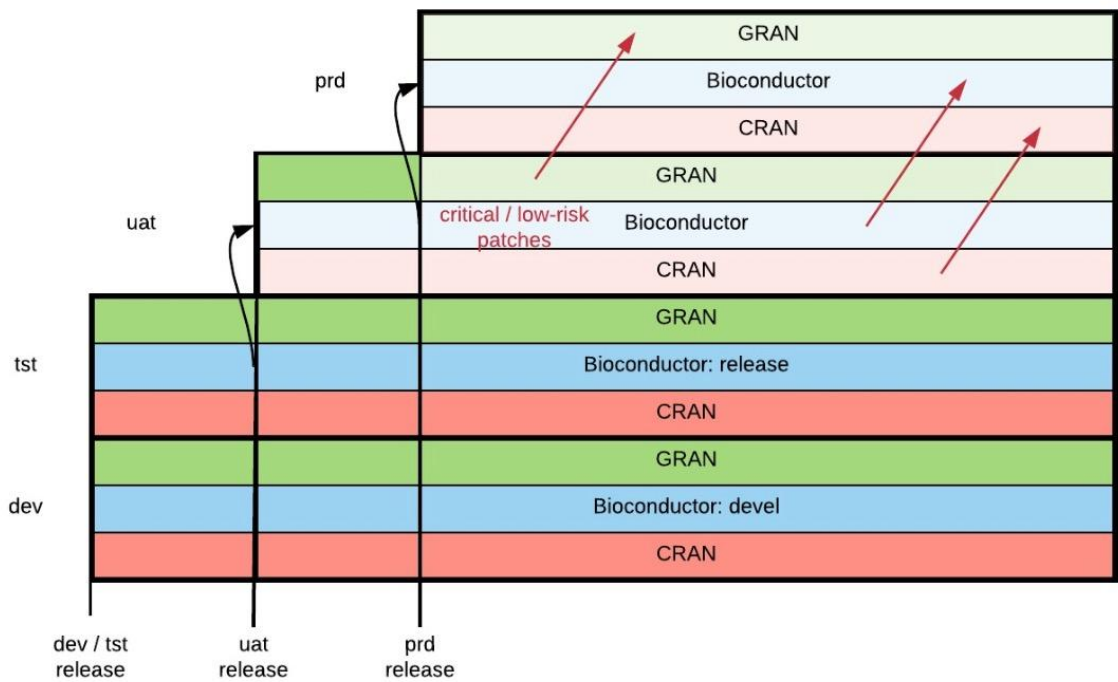
+



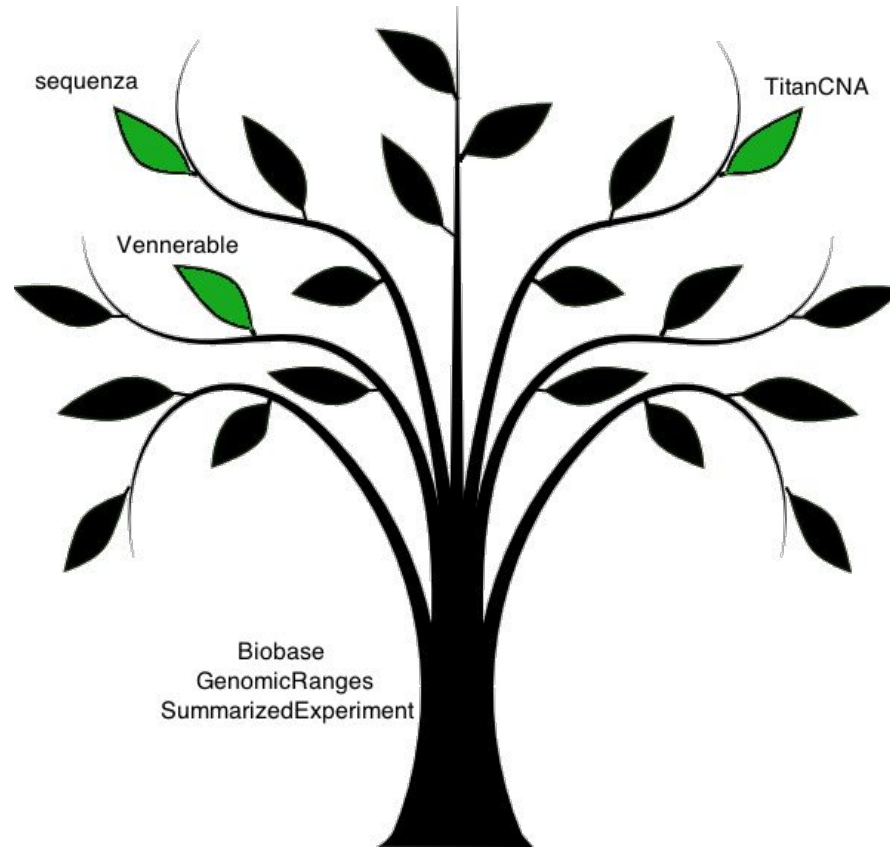
Balance



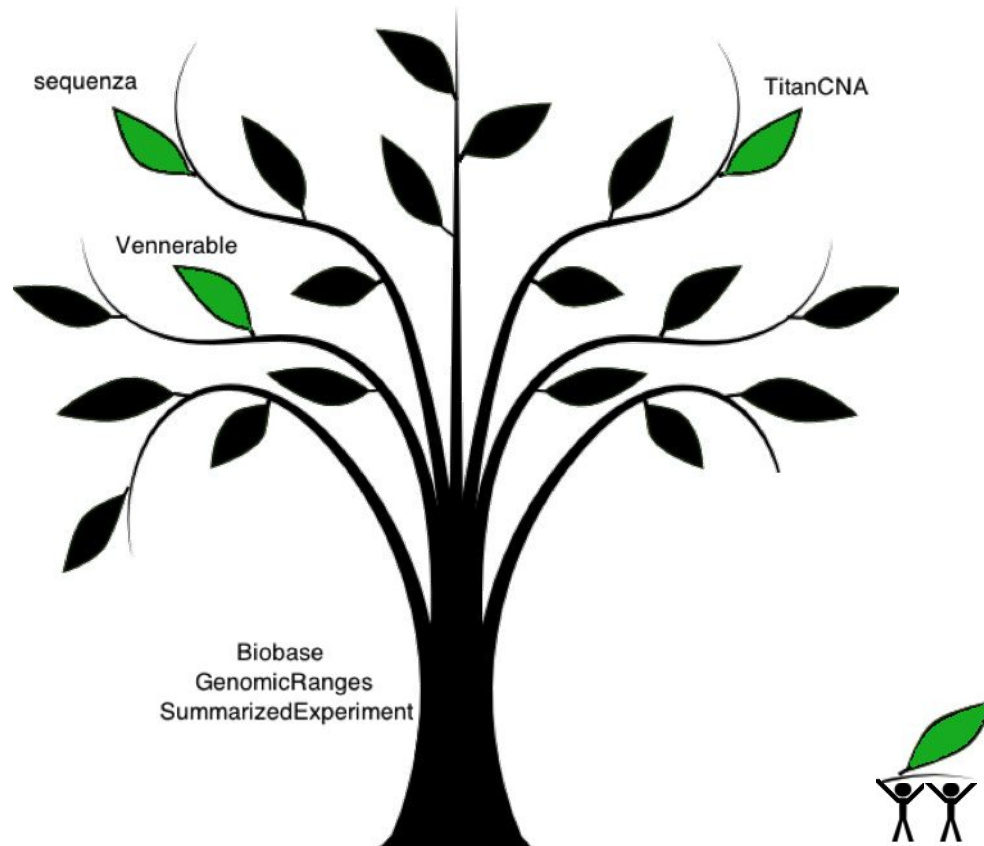
Bounding Risk



Leafy Update Policy for Production



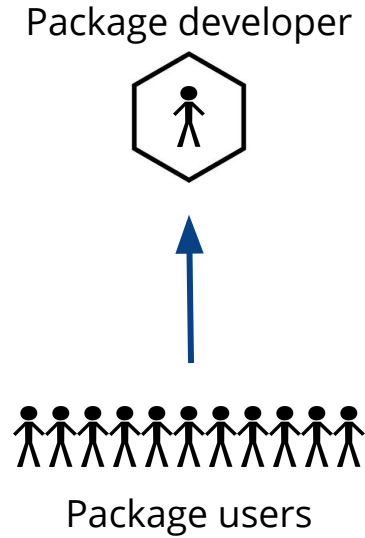
Leafy Update Policy for Production



User Analytics

datetime	user	host	R version	package	package version
2019-08-08 16:14:58	user1	host1	R-3.6.0-Bioc-3.9-prd	package1	1.1.11
2019-08-08 16:12:01	user	host1	R-3.6.0-Bioc-3.9-prd	package2	0.18.5
2019-08-06 07:29:33	user2	host2	R-3.6.0-Bioc-3.10-dev	package3	0.2.4

Making New Connections



Making New Connections

Package developer



?



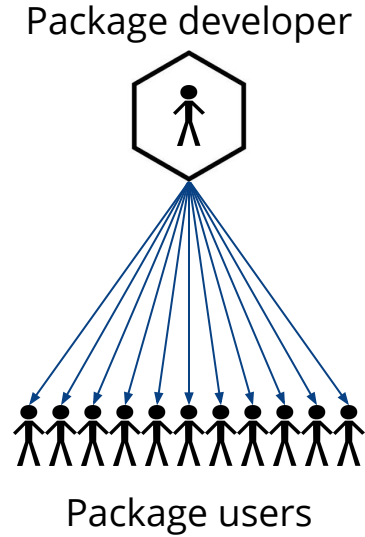
?



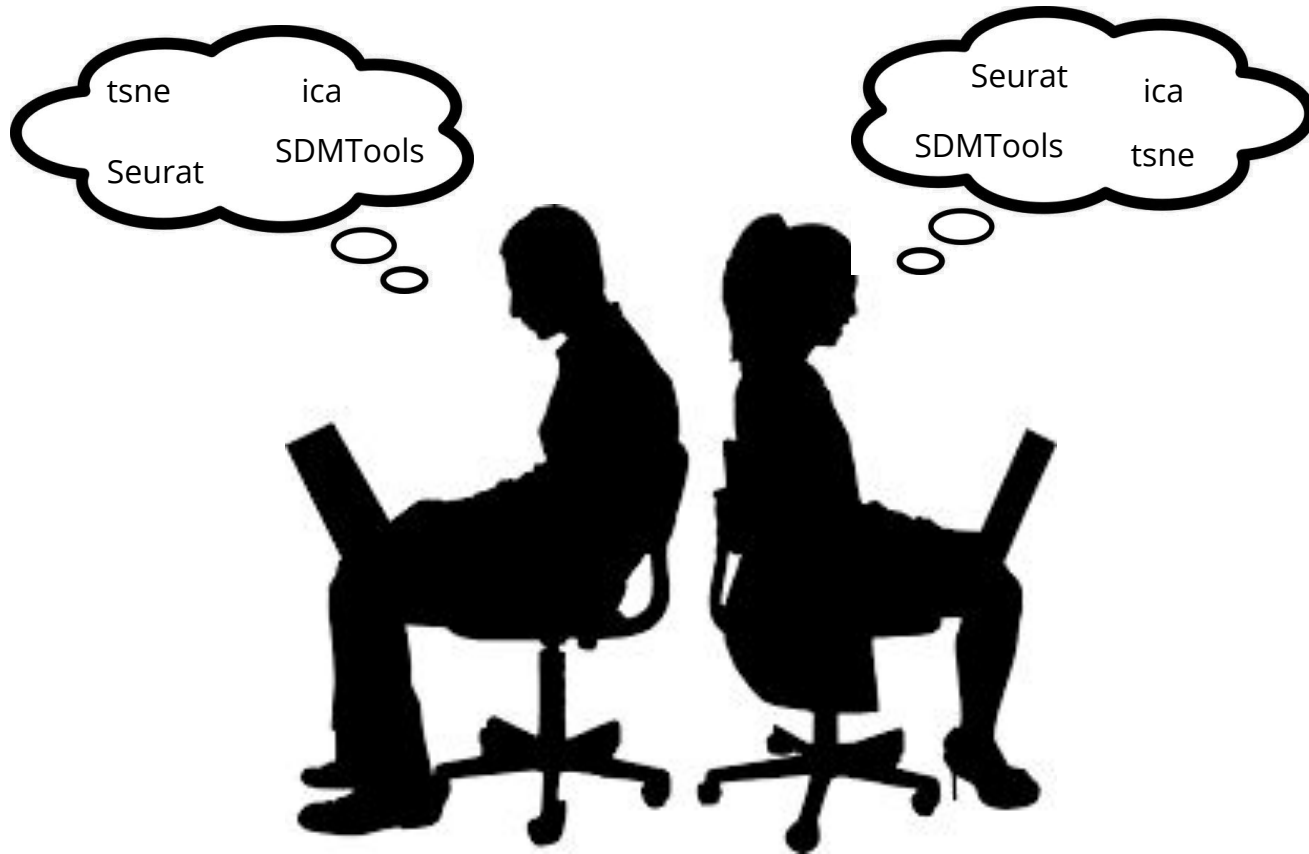
Package users

Making New Connections

user	package
user1	mypackage
user2	mypackage
user3	mypackage



Making New Connections



Making New Connections



Making New Connections

Frequently bought together



+



+



Total price: **\$17.93**

Add all three to Cart

Add all three to List

Making New Connections

Frequently loaded together

Seurat

tsne

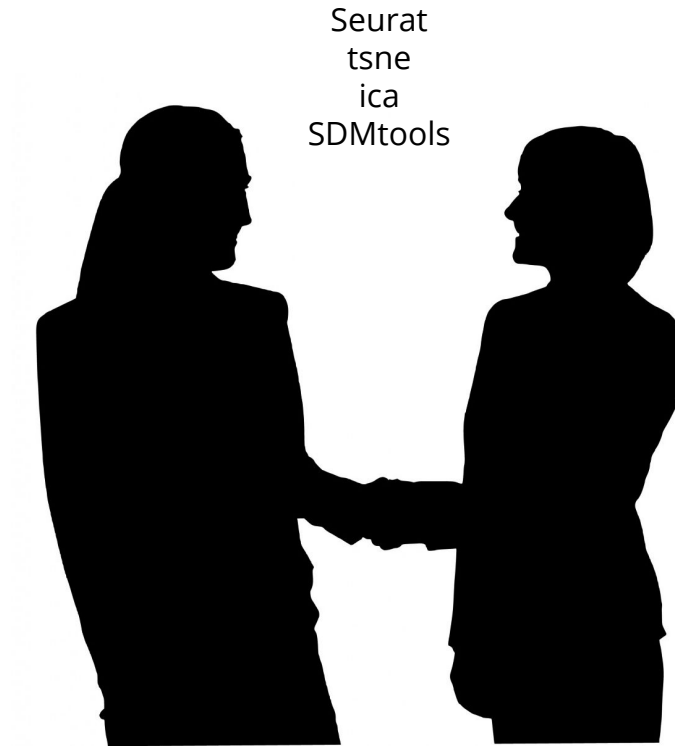
ica

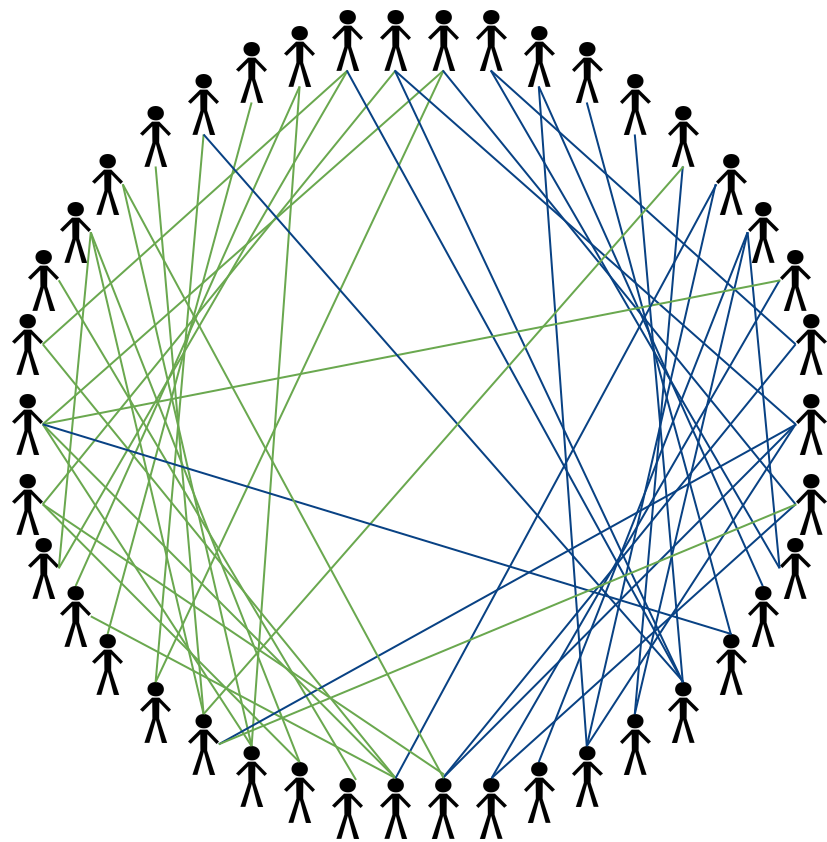
SDMTools

Making New Connections



Making New Connections





Acknowledgements

- Michael Lawrence
- Gabe Becker
- Dinaker Kulkarni
- Henryk Poplawski
- Valerie Obenchain

- Betty Chou
- Jim Fitzgerald
- Allison Vuong
- Kiran Mukhyala
- Pete Haverty

Thank You!