



Jumping Rivers



Ltd
@jumpingrivers@fosstodon.or

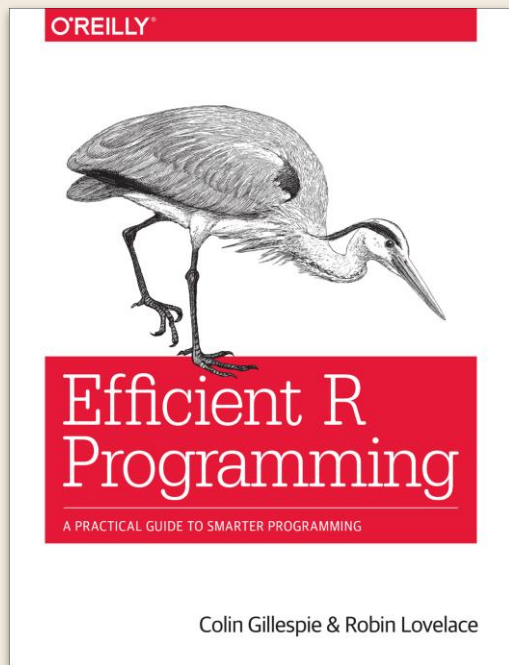
g

**Jumping
Rivers**

R packages in the Pharmaverse

Colin Gillespie

About Me: Colin Gillespie



AI in Production 2027

Almost ex-academic
Co-founder of Jumping Rivers

About: Jumping Rivers

Posit Support & Maintenance

Basic to Enterprise

Kubes, HA, GxP



About: Jumping Rivers

Posit Support & Maintenance

Basic to Enterprise

Kubes, HA, GxP



Validated & Assured R packages

Pharma, insurance, government



About: Jumping Rivers

Posit Support & Maintenance

Basic to Enterprise

Kubes, HA, GxP



Validated & Assured R packages

Pharma, insurance, government



databricks

Data Science Consultancy

R, Python, Shiny... Everything



About: Jumping Rivers

Posit Support & Maintenance

Basic to Enterprise

Kubes, HA, GxP



Validated & Assured R packages

Pharma, insurance, government



databricks

Data Science Consultancy

R, Python, Shiny... Everything



Training

Over 50 courses!

Package choice
comes up all the
time

**Jumping Rivers
does lots of
training**

We always get asked

Can we trust this package?

Confession time

JR

Confession time

- Wave hands
- Waffle about CRAN

Confession time

- Wave hands
- Waffle about CRAN
- Common sense
- Change topic

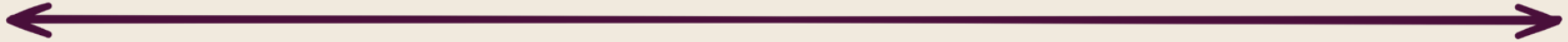
Can we trust “this” R package?

JR

Can we trust “this” R package?

It depends on your **risk appetite**

Risk Appetite



Risk Appetite



Risk Appetite



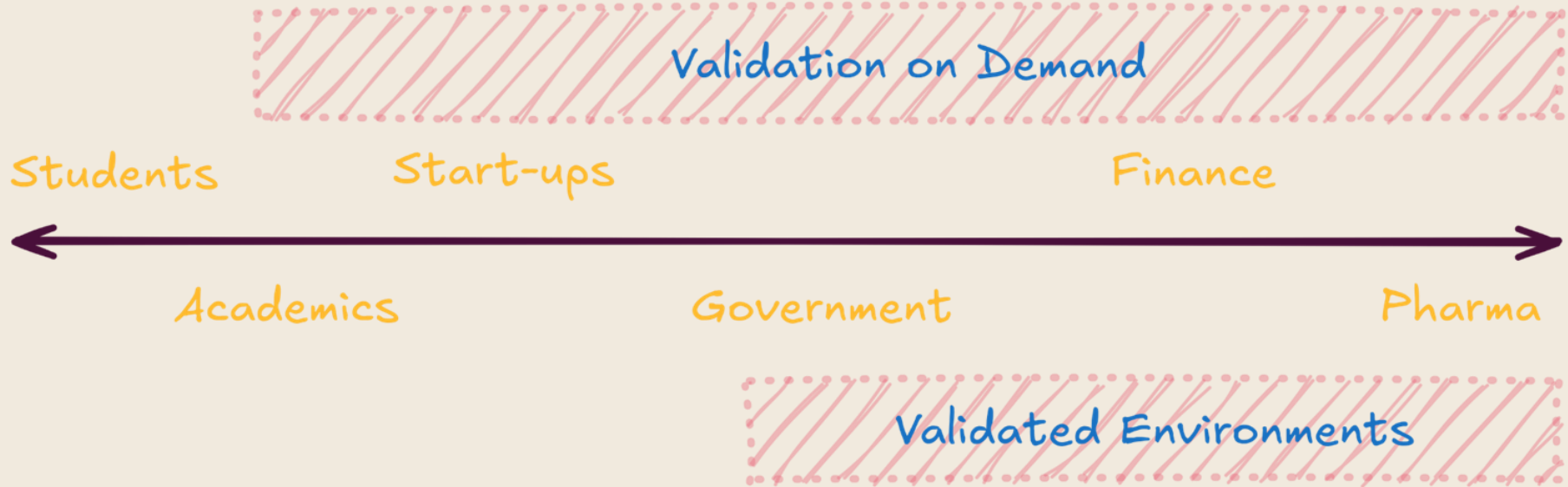
Risk Appetite



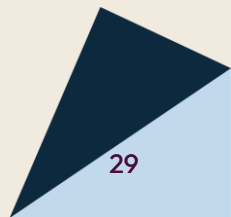
Risk Appetite



Risk Appetite



Many organisations are on both sides



Package Risk

How would you assess a package?



How would you assess a package?



Author “fame”

Code coverage

How would you assess a package?

Author “fame”

Code coverage

Vignettes

Source control

Issue closure rate

How would you assess a package?

Author “fame”

Code coverage

Vignettes

Source control

Issue closure rate

License

NEWS

How would you assess a package?

Author “fame”

Code coverage

Vignettes

Source control

Issue closure rate

License

NEWS

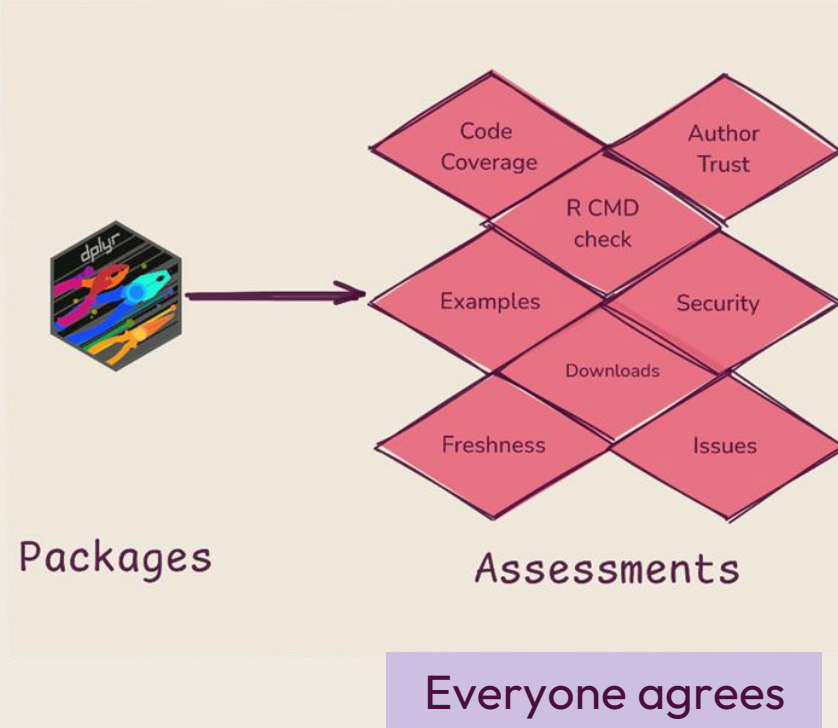
Security

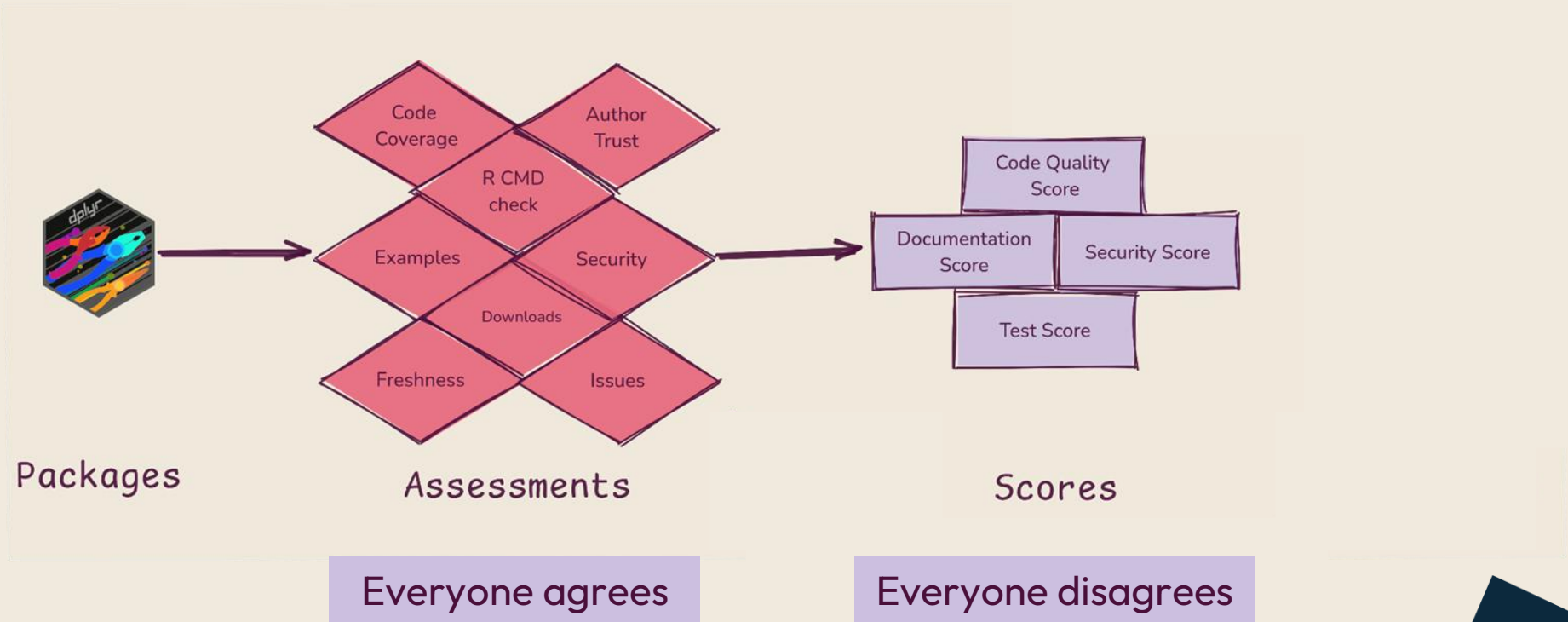
Downloads

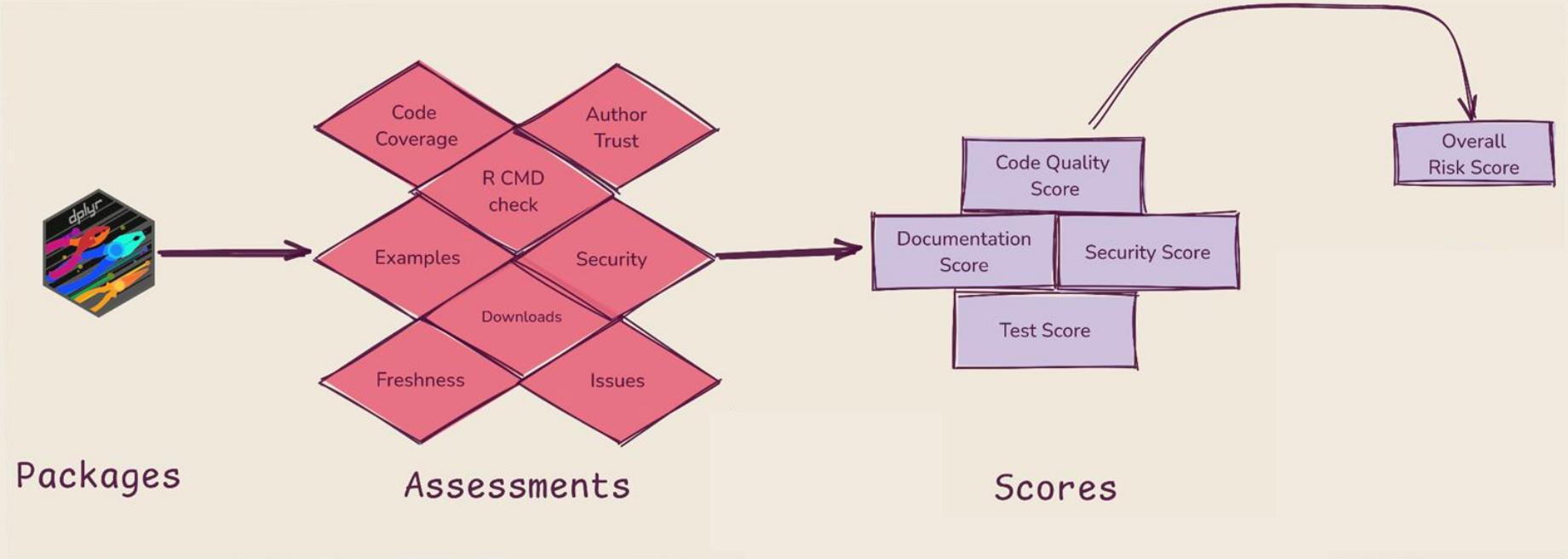
Rev dependencies

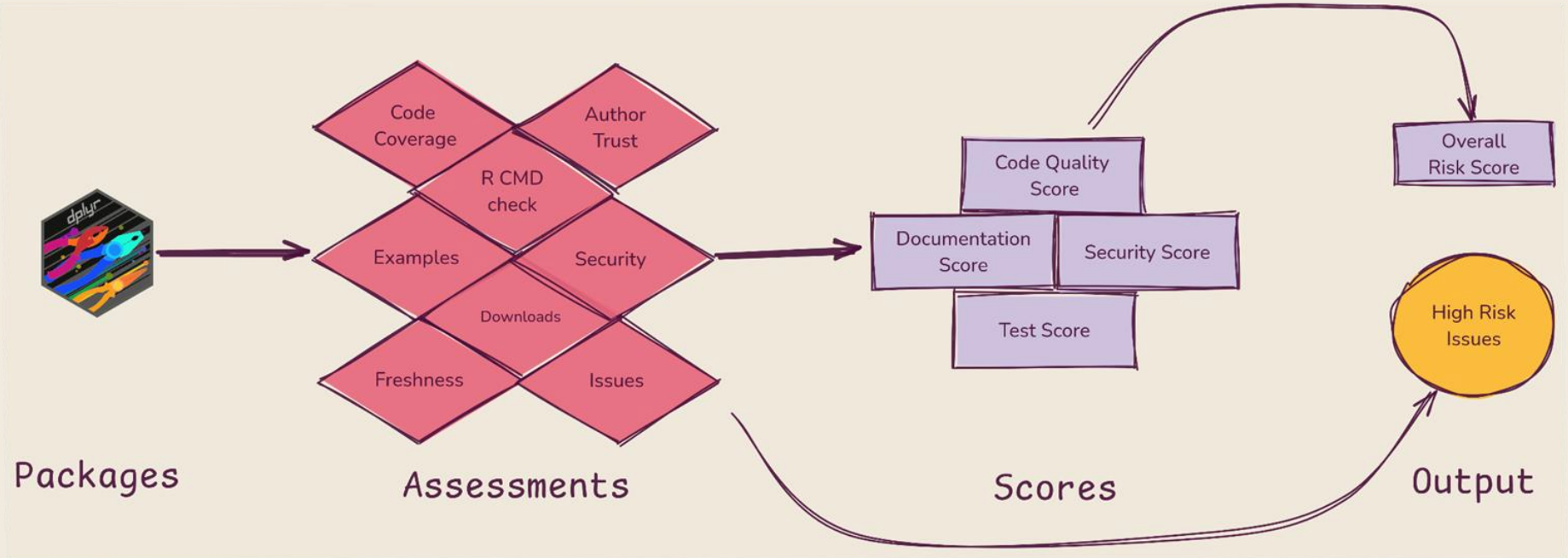


Packages









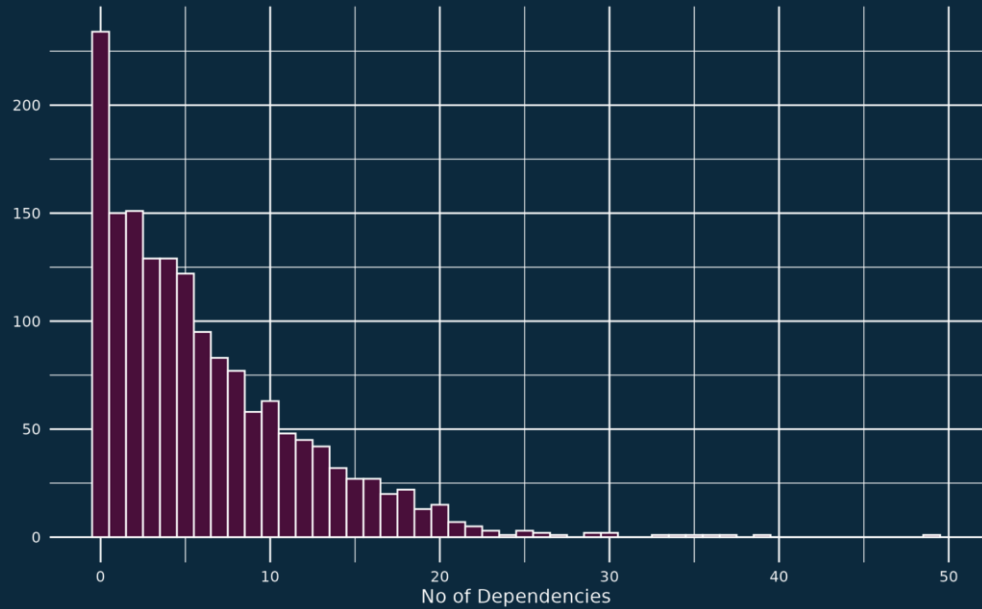
Study

- Last ~1200 packages submitted to CRAN
 - Biased sample - “better quality”
- pharmaverse.org packages

Dependencies

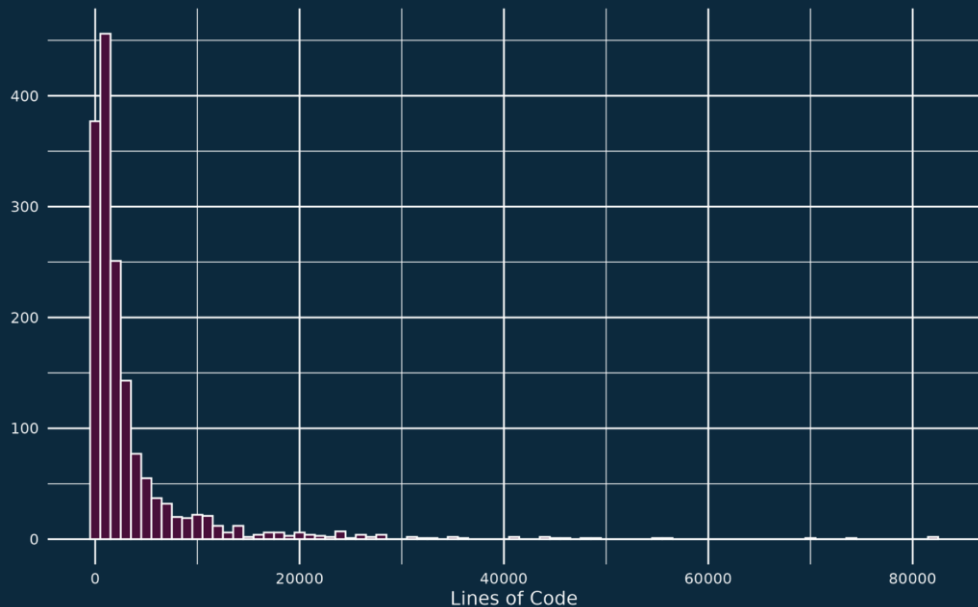
- Q1: 2 dep'
- Q2: 4 dep' vs **9**
- Q3: 9 dep' vs **14**

CRAN vs **pharmaverse**



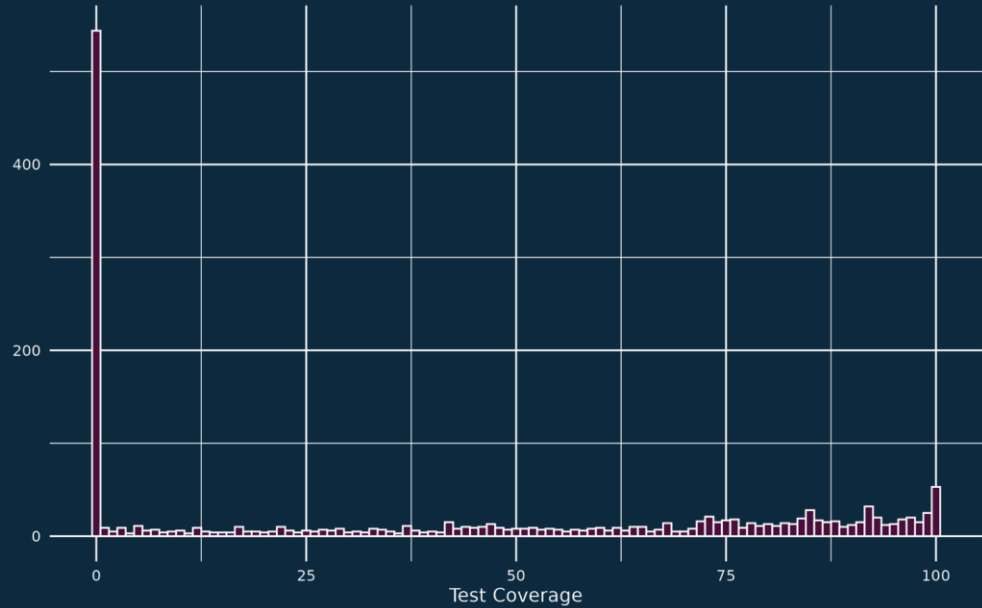
Codebase Size

- Q1: 600 Lines vs **300**
- Q2: 1400 Lines
- Q3: 3300 Lines
- Max: 80,000+



Test Coverage

- Q1: 0 vs **10%**
- Q2: 35% vs **88%**
- Q3: 75% vs **96%**



Contributors

- Q2: 2 vs **7**
- Q3: 4 vs **11**

Licences (SPDX conversion)

- CRAN
 - Mainly MIT & GPL
- pharmaverse.org
 - Apache-2.0 & MIT

How do you score a package?



Vignettes (extra documentation)

- Binary score: 0 or 1

Vignettes (extra documentation)

- Binary score: 0 or 1

What about multiple vignettes?

Do they get a higher score?

Vignettes (extra documentation)

- Binary score: 0 or 1

What about multiple vignettes?

Do they get a higher score?

Told you we would disagree on scores

Dependencies

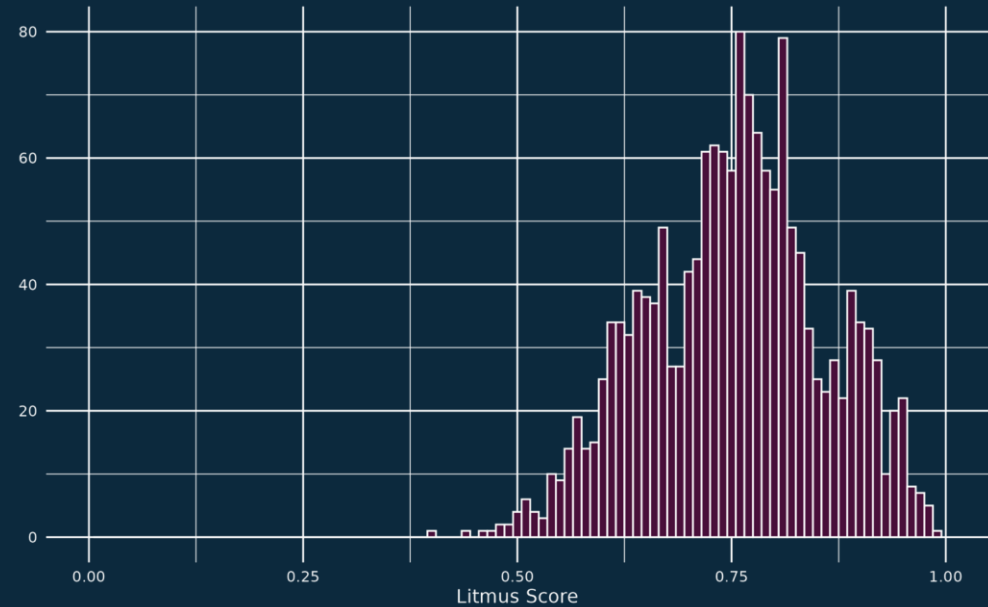
- How big is the Imports list

Dependencies

- How big is the Imports list
- Everyone has an opinion
 - So we cheated!
- Analysed all CRAN packages

Litmus Score

- Q1: 0.7
- Q2: 0.75
- Q3: 0.82



Moving to R

- Be proactive
- What's your procedure for assessing packages
 - Needs to be dynamic

<https://litmus-dashboard.jmpr.io/>

Dashboard: Overview of all packages

918

2 (0%)

918

Package	Score	Count	Percentage
Package 1	918	918	100%

Package	Score	Count	Percentage
Package 2	0	2	0%

Package	Score	Count	Percentage
Package 3	0	0	0%

Package Scores

918
Packages Collected

Manage Score By Category

1
Packages Requiring Attention

2 (0%)
High Risk Packages

0.76
Average Score

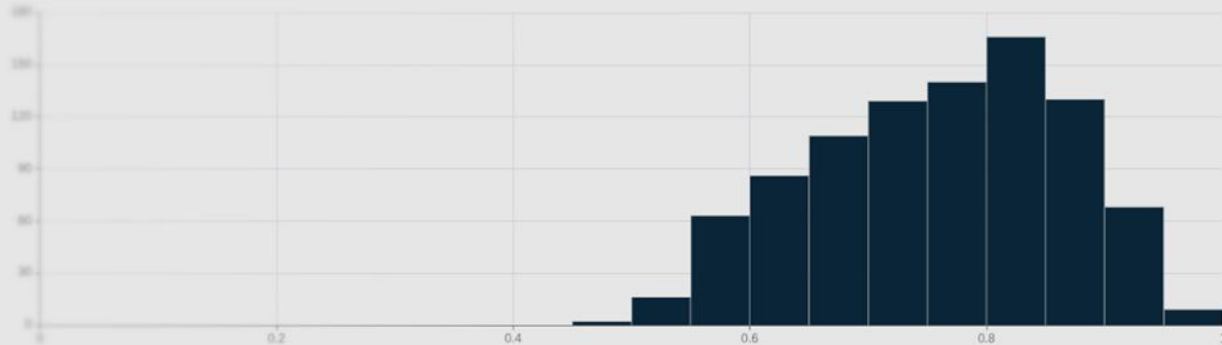
Security 100% Performance 100% Usage 100%

1 Package Found With Dependency Vulnerabilities

Security	100%
Performance	100%
Usage	100%

Collection Scores

> Lowest scoring packages



Explore All Packages

642 (70%)
Low Risk Packages


	datasauRus	1
	warp	0.98

> Highest scoring packages


Collection Summary

 918
Packages in Collection

 0.76
Average Litmus Score

 Average Score Per Category

Documentation	Code	Maintenance	Package Popularity
0.72	0.71	0.81	0.98

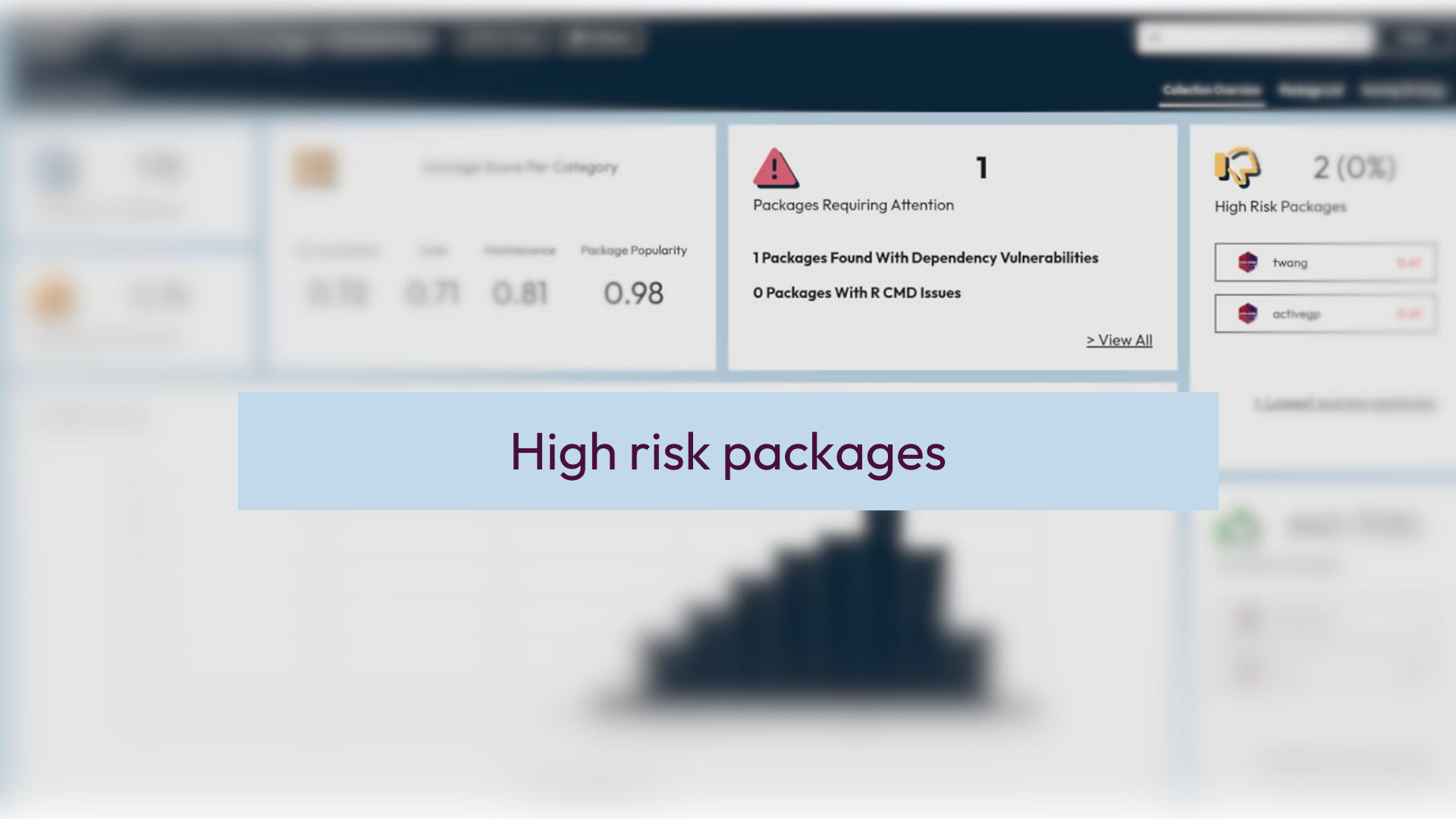
 Packages Requiring Attention

1 Packages Found With Incomplete Information

0 Packages With In-Valid Data


Score Breakdown

High risk packages



Average Score Per Category


Score	Reference	Package Popularity
0.72	0.71	0.81
		0.98



 **1**
Packages Requiring Attention

1 Packages Found With Dependency Vulnerabilities

0 Packages With R CMD Issues

[> View All](#)

 **2 (0%)**
High Risk Packages

-  twang 0.00
-  activegp 0.00



[-- Return to Package List](#)

Package: dotwhisker

Version: 0.8.3

Litmus Score: 0.72

Maintainer: Yue Hu <yuehu@tsinghua.edu.cn>

Website: <https://fsalt.org/dotwhisker/>

Diffify: <https://diffify.com/R/dotwhisker/>

Score Breakdown

Decision Records

JR Decision Support Data

Search

Category

Score

Package Drill down

Litmus Score: 0.72



Scores per metric



Category	Metric	Score	Category Contribution	Score Contribution
▼ documentation (6)				
<input type="radio"/>	Examples: Proportion of Help Pages	0.86	36%	5%
<input type="radio"/>	Has News	1.00	14%	2%
<input type="radio"/>	Exported Objects Documentation	1.00	14%	2%
<input type="radio"/>	NEWS Updated to Current Version	1.00	14%	2%
<input type="radio"/>	Has Vignettes Files	1.00	14%	2%
<input type="radio"/>	Has Website URL	1.00	7%	1%
▼ code (7)				
<input type="radio"/>	Unit Test Coverage	0.00	31%	16%
<input type="radio"/>	Vulnerabilities	1.00	31%	16%
<input type="radio"/>	Package Dependencies	0.16	16%	8%
<input type="radio"/>	Exported Functions Coverage	0.00	6%	3%
<input type="radio"/>	Codebase Size	0.58	6%	3%
<input type="radio"/>	Check Results Score	1.00	6%	3%
<input type="radio"/>	Exported Objects	0.71	3%	2%

Saved Records

Record decisions

may reverse package so
of its popularity. I thin
package. What do otl



@Stella Smith:
Completely agree.

@Alex Naylor:
Yep! Agree, also.

@Andrew Booth:
Record text

Great!

New Record

Record text:

+ Add Record

Summary

- A little bit of thought goes a long way
- Sensible infrastructure
- Rapid validation environment
- **diffify.com**



AI in Production 2027

Questions: colin@jumpingrivers.com