

# A reproducible analysis of CRAN Task Views

to understand the state of an R package ecosystem

## Introduction

CRAN Task Views are a collection of R packages related to a specific topic.

But are these packages reliable, updated, and in line with current community best practices?

We perform a reproducible, and reusable, analysis of the Epidemiology CRAN Task View, aiming to:

- Evaluate the ecosystem health and resilience
- Identify training and support needs

## Methods

1. Create a list of R packages relevant to a specific topic via systematic search and surveys
2. Fetch package DESCRIPTION, and version history via the pkgsearch R package
3. Clean author names (Author field is not standardized)
4. Extract or manually identify GitHub source repository
5. Check for the presence of specific metadata or files via the GitHub API

Epiverse  
powered by data.org

Epidemiology Task View analysis

Analysis

About



Number of packages

89

Package authors

400

GitHub stars

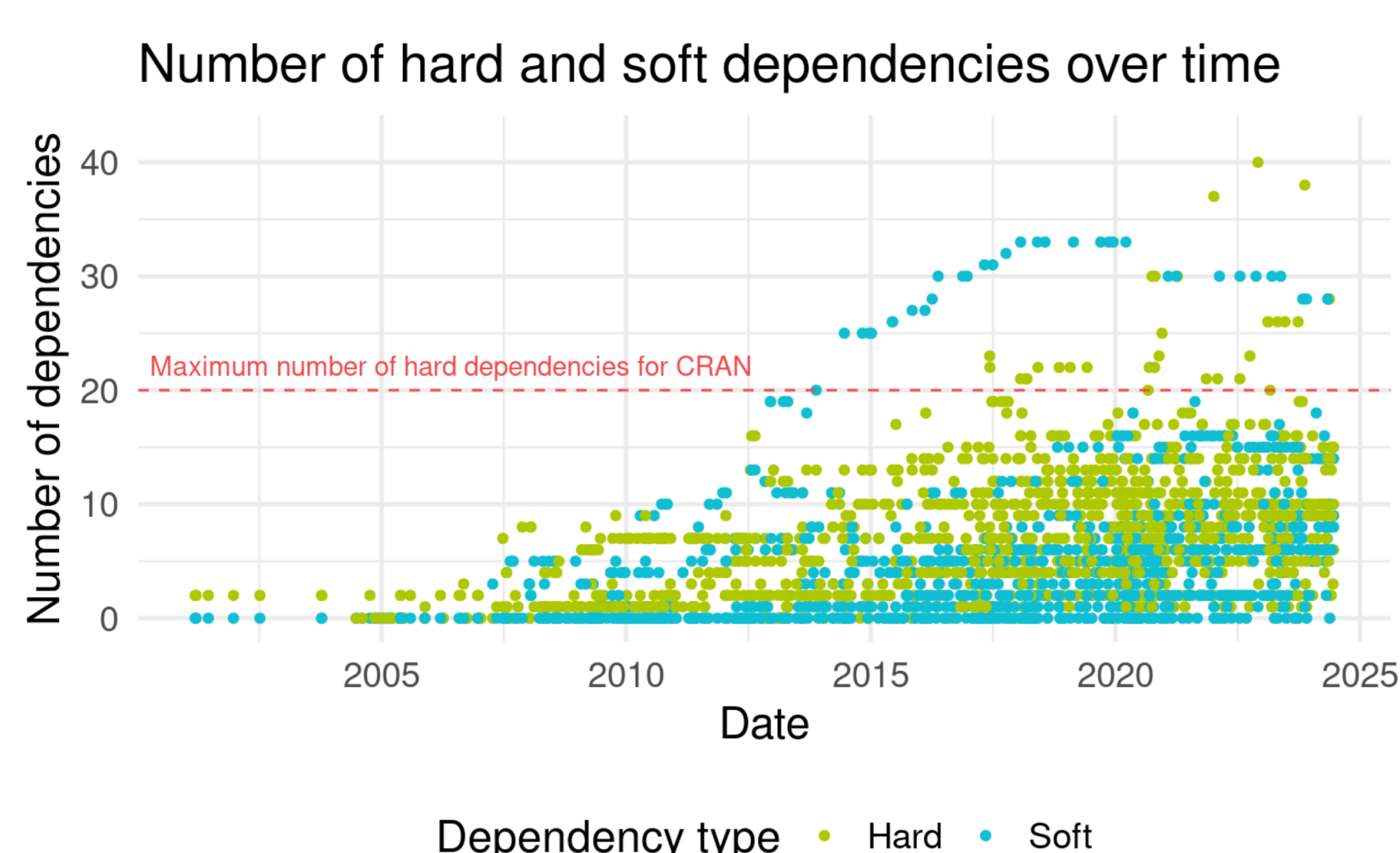
1,171

CRAN downloads

16M

Data points

Trend line



### Adherence to good practices

Good practices as defined by [rOpenSci dev guide](#), and [Epiverse-TRACE blueprints](#).

#### Sustainability

Does not depend on deprecated packages  
*XML, RCurl, RUnit, plyr, or reshape2 packages* 85%

#### Metadata

Has Authors@R field 85%

Has a GitHub URL  
*In 'BugReports' or 'URL' fields in 'DESCRIPTION'* 71%

Has ORCID in Author field 45%

Has LICENSE.md 33%

Provides a DOI for citation 31%

#### Documentation

Uses Roxygen  
*As indicated by the presence of the 'RoxygenNote' field in 'DESCRIPTION'* 82%

Has a knitr vignette 57%

Has NEWS.md 55%

Has README.Rmd 43%

Uses pkgdown 26%

#### Testing & CI

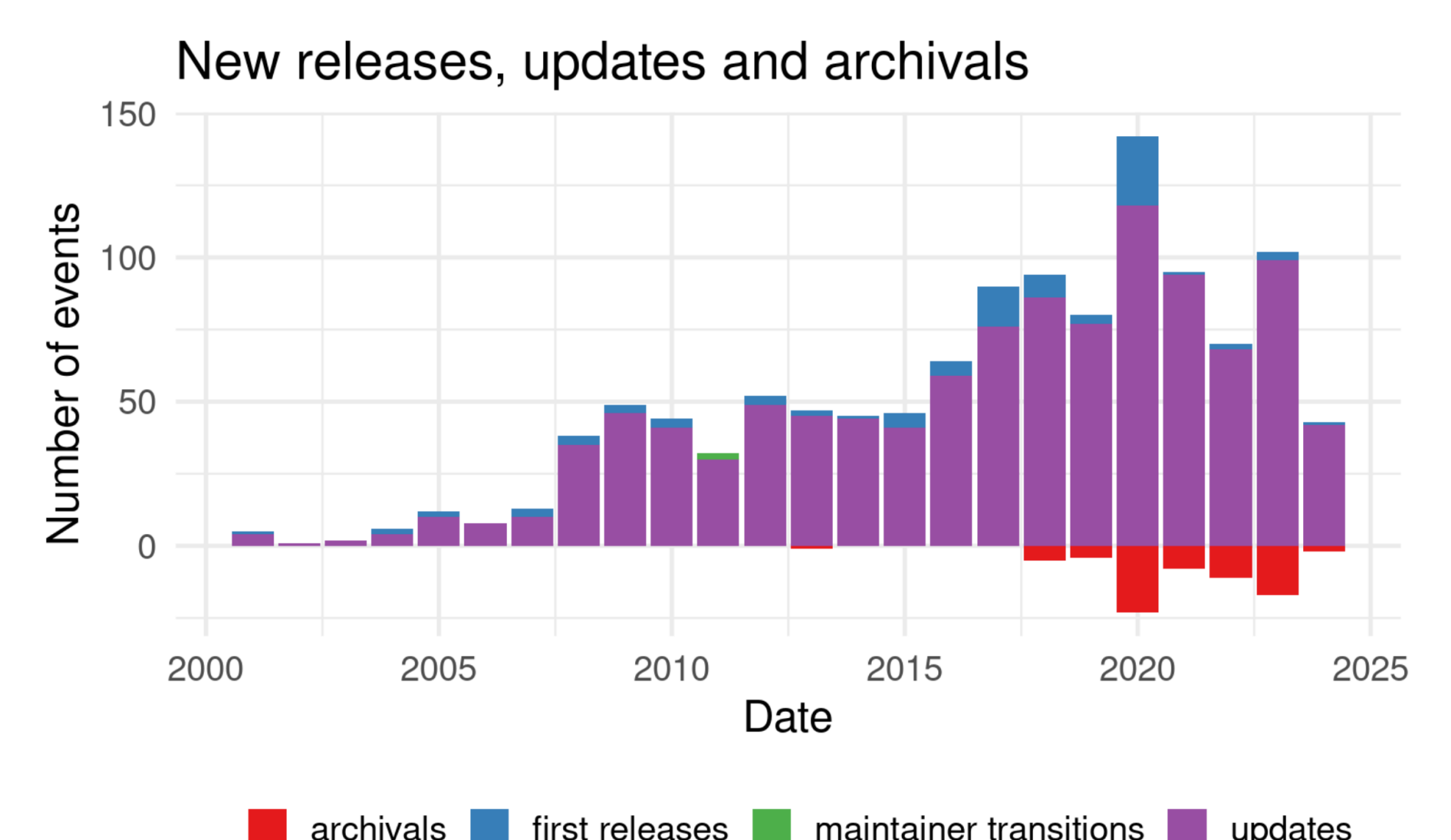
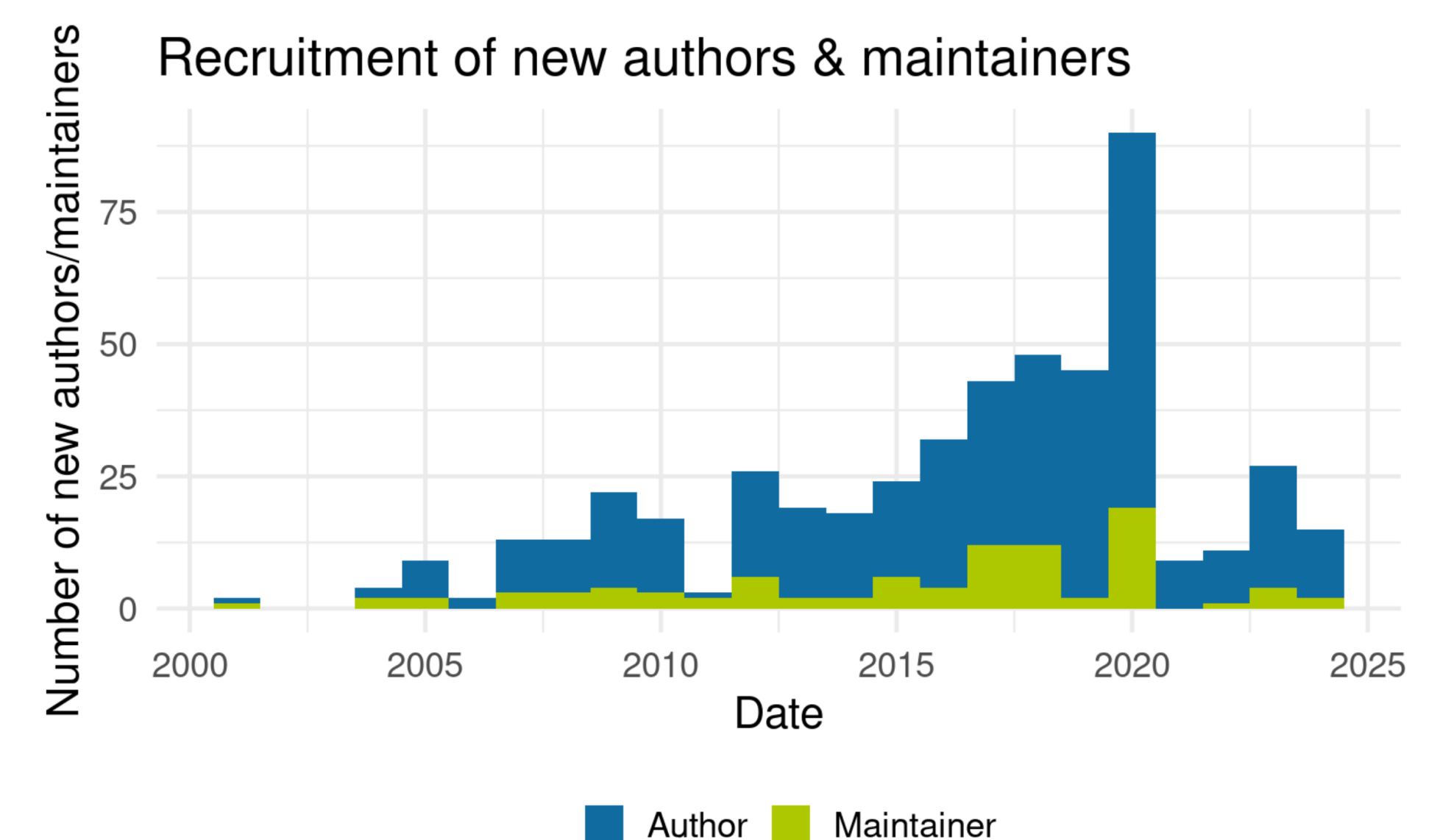
Uses a testing framework  
*As indicated by the presence of testthat, testit, unitizer, RUnit, tinytest in 'Suggests'* 55%

Uses GitHub Actions 48%

#### Community

Has a contributing guide 7%

Has a code of conduct 7%



## Discussion

The Epidemiology R package ecosystem is strong, with reasonable activity, especially when needed (e.g., 2020 release surge). But:

- Struggle to recruit new authors & maintainers over the past couple of years
- Inflation of the number of dependencies, exceeding CRAN soft limit
- Relatively low adherence to good practices

## Future directions

- Extract cleaning functionality to separate package
- Compare results to CRAN in its entirety
- Submit pull requests to packages to:
  - Improve their metadata, and consequently the quality of this analysis
  - Propose modern alternative to deprecated packages
- Organize trainings on testing and pkgdown

Live analysis available at <https://epiverse-connect.github.io/ctv-analysis>



Epiverse  
powered by data.org