

Positioning the R in Open SouRce and Reproducibility: RStudio + GitHub + Renv workflows



March 25, 2025

Dan Kerchner

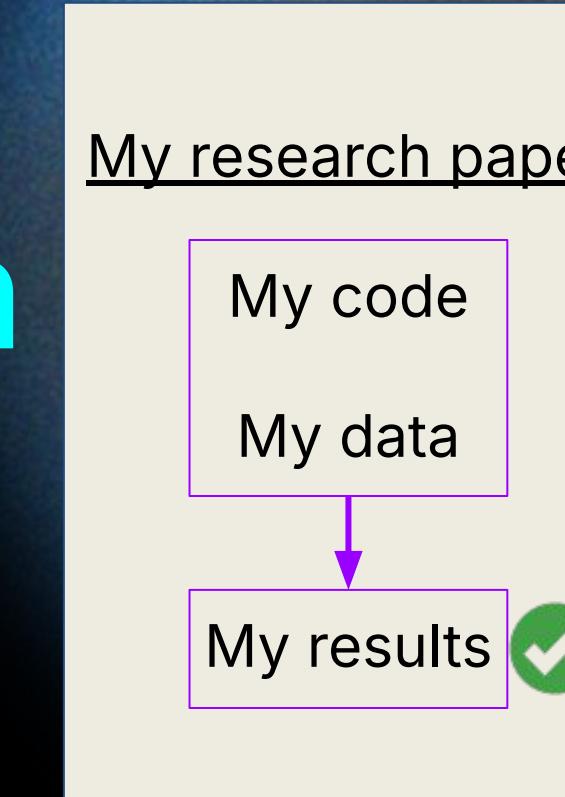
Senior Software Developer | GW Libraries & Academic Innovation
PhD Student | GWSPH Health Data Science, Applied Bioinformatics

Positioning the R in Open SouRce and Reproducibility: RStudio + GitHub + Renv workflows

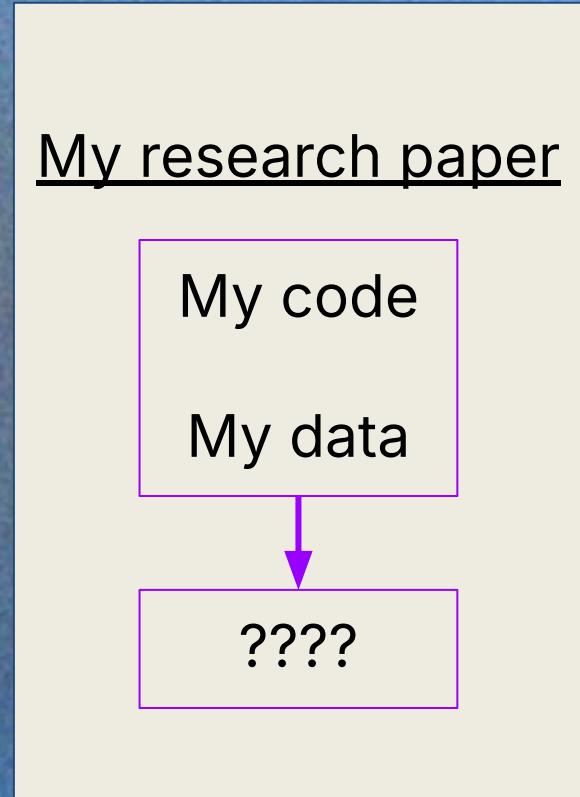


If my research code/data is "open" but not
easily reproducible, is it truly open?

Fresh publication



Reader attempts to reproduce



What could possibly go wrong? 

What could possibly go wrong?

Big problems:

- ✗ Reader tries running code but encounters errors
- ✗ Reader tries running code - it runs, but results are different

Medium problems:

- ✗ Code (as an attachment) isn't set up for collaboration
- ✗ Code (as an attachment) can't be cited independently of the paper
- ✗ Versions of key libraries are cited, but most readers don't know how (or don't want) to successfully install [many] specific versions
- ✗ Version of R/Python is cited, but reader is using another version

An improved R workflow

- ❖ RStudio projects
- ❖ GitHub (via RStudio and/or command line)
- ❖ renv
- ❖ [RSwitch or similar]

An improved R workflow: RStudio projects

Set context to a working directory so that you can always use relative paths like

```
read.csv('data/mydata.csv')
```

and never have to do this

```
read.csv('/Users/Dan/my_stuff/project1/data/mydata.csv')
```

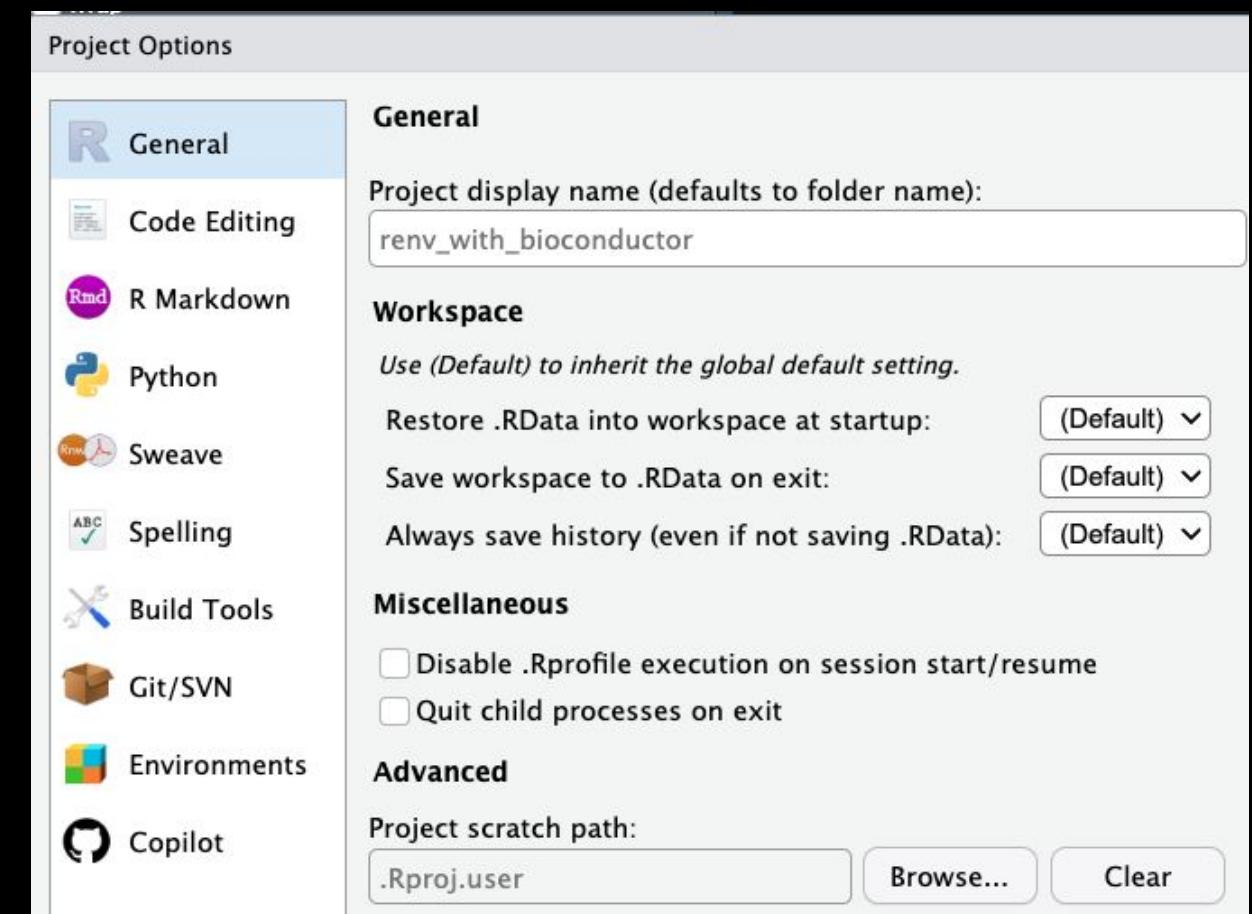
or even this

```
setwd('/Users/Dan/my_stuff/project1/data')
```

```
read.csv('data/mydata.csv')
```

An improved R workflow: RStudio projects

- Provides clean, self-contained project environment
- Isolates project-specific options
- Isolates projects with different library versions (with `renv!`)



```
Version: 1.0
ProjectId: 19132850-b088-48bb-815f-dfce0855230d

RestoreWorkspace: Default
SaveWorkspace: Default
AlwaysSaveHistory: Default

EnableCodeIndexing: Yes
UseSpacesForTab: Yes
NumSpacesForTab: 2
Encoding: UTF-8

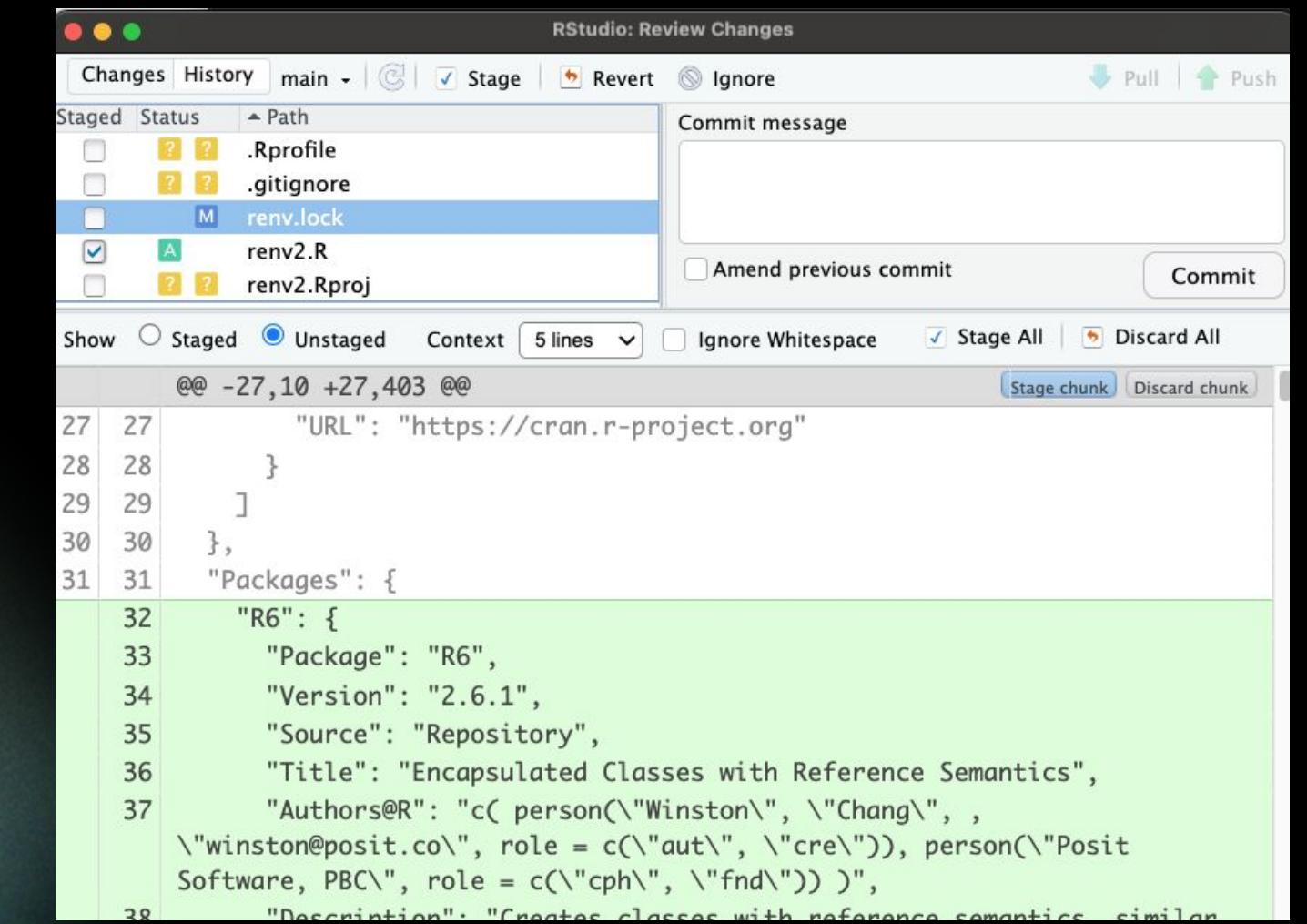
RnwWeave: Sweave
LaTeX: pdfLaTeX
```

An improved R workflow: Git/GitHub

- Readers can not only access code version corresponding to the paper, but also updates (bug fixes...)
- Opens your code to collaboration & engagement
 - issues/discussions
 - contributions as pull requests
- Your code is its own scholarly object and can have its own DOI 
- Additional storage/organization beyond what a journal can host
- Lots of other tools - CI/CD, documentation tools, etc.

An improved R workflow: Git/GitHub

- ❖ Git/GitHub things I think RStudio does **well**
- State of files
- Diff viewer
- History viewer
- Make/switch branches
- Push/pull



The screenshot shows the RStudio interface for managing git changes. The top navigation bar includes 'Changes' (selected), 'History', and 'main'. Below the navigation are buttons for 'Stage', 'Revert', and 'Ignore'. A 'Commit message' field is present with a 'Commit' button. The main area has tabs for 'Staged' (selected) and 'Unstaged'. Under 'Path', several files are listed: '.Rprofile', '.gitignore', 'renv.lock' (highlighted in blue), 'renv2.R', and 'renv2.Rproj'. The 'Show' dropdown is set to 'Unstaged'. The bottom half of the window displays a diff viewer with code snippets. Lines 27-31 show:

```

@@ -27,10 +27,403 @@
27 27     "URL": "https://cran.r-project.org"
28 28   }
29 29   ]
30 30 },
31 31 "Packages": {

```

Line 32 shows the start of the 'R6' package definition.

An improved R workflow: Git/GitHub

- ❖ Git/GitHub things RStudio does **not** do
 - merge/rebase
 - tag
 - reset
 - git rm, git mv

So... where is the dependency management in R???

Ruby+Rails : Bundler

Python : Pip or Conda

JavaScript: npm or yarn

R : ????

12factor.net/dependencies

**"A twelve-factor app
never relies on implicit existence
of system-wide packages."**

An improved R workflow:

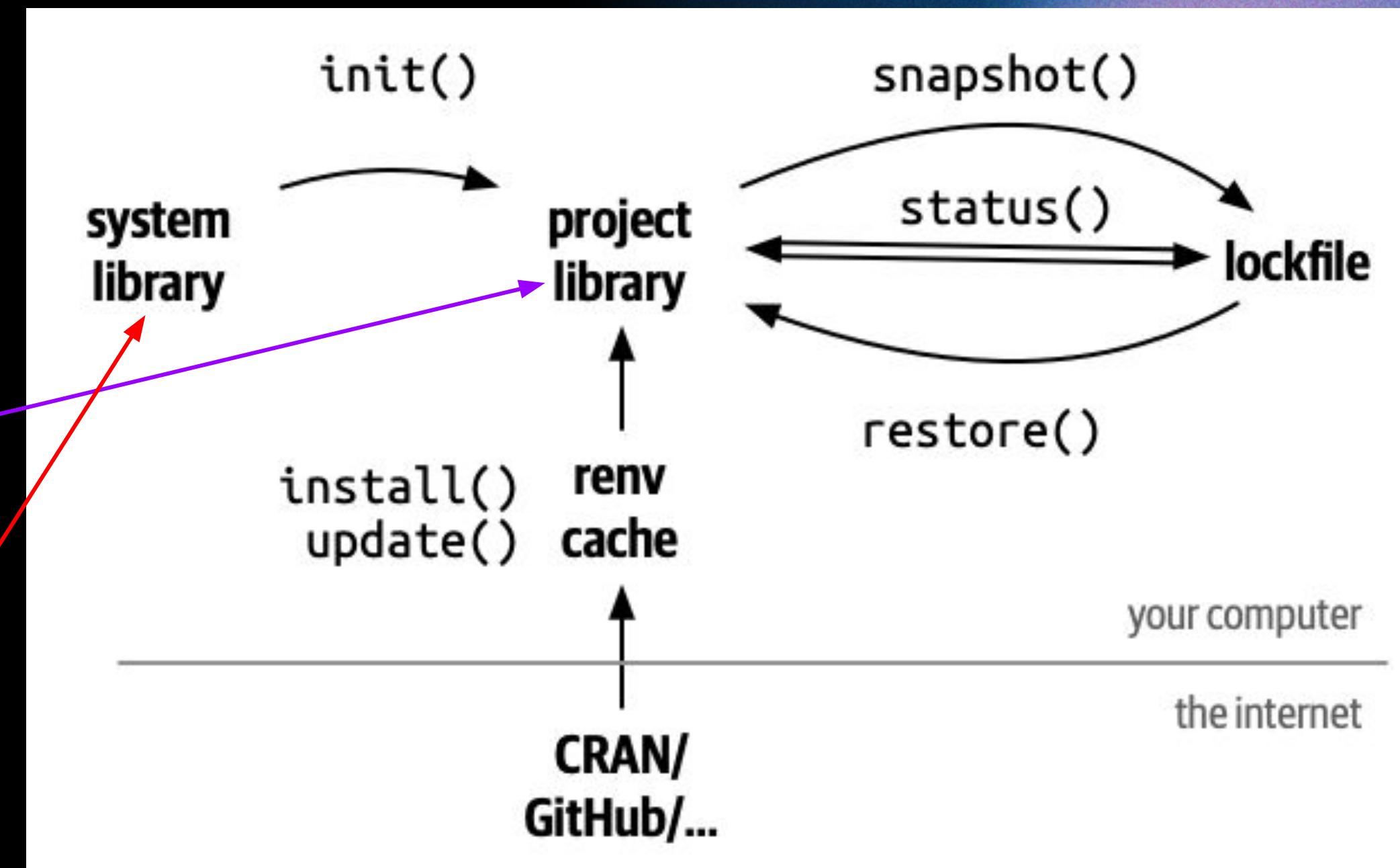
renv

- ❖ is a dependency management toolkit for R. It tracks the packages you use, and their versions
- ❖ creates an `renv.lock` lockfile that records the exact versions of R packages used in a project
- ❖ can restore the package(s) environment for a project by installing the versions recorded in the lockfile

An improved R workflow:

renv

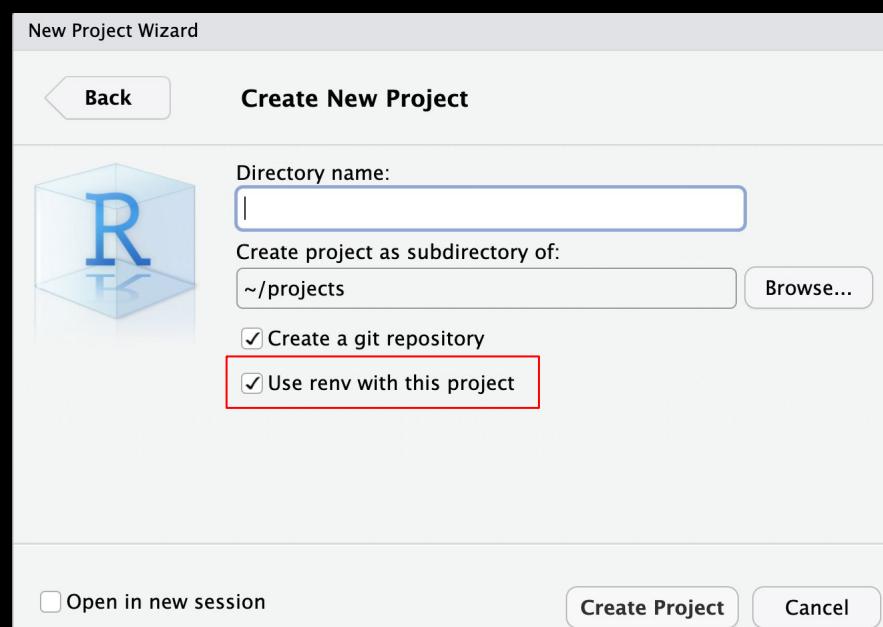
```
> .libPaths()
[1] "/Users/kerchner/projects/renv3/re
nv/library/macos/R-4.4/aarch64-app
le-darwin20"
[2]
"/Users/kerchner/Library/Caches/or
g.R-project.R/R/renv/sandbox/macos
/R-4.4/aarch64-apple-darwin20/f715
6815"
```



From <https://rstudio.github.io/renv/articles/renv.html>

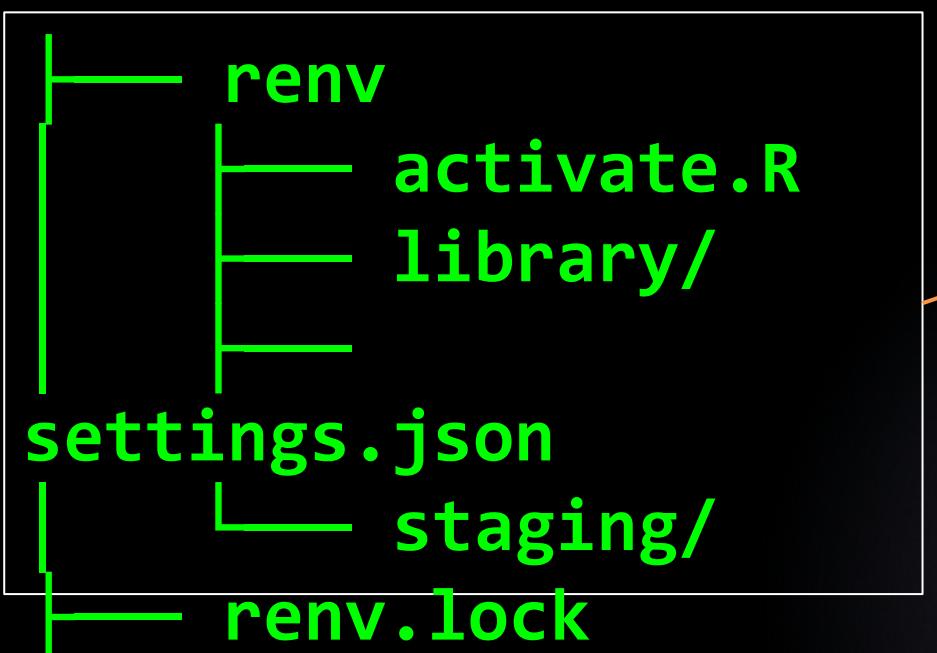
An improved R workflow:

renv



or:

`renv::init()`



```
{
  "R": {
    "Version": "4.4.2",
    "Repositories": [
      {
        "Name": "BioCsoft",
        "URL": "https://bioconductor.org/packages/3.20/bioc"
      },
      {
        "Name": "BioCann",
        "URL": "https://bioconductor.org/packages/3.20/data/annotation"
      },
      {
        "Name": "BioCexp",
        "URL": "https://bioconductor.org/packages/3.20/data/experiment"
      },
      {
        "Name": "BioCworkflows",
        "URL": "https://bioconductor.org/packages/3.20/workFlows"
      },
      {
        "Name": "BioCbooks",
        "URL": "https://bioconductor.org/packages/3.20/books"
      },
      {
        "Name": "CRAN",
        "URL": "https://cran.r-project.org"
      }
    ],
    "Bioconductor": {
      "Version": "3.20"
    },
    "Packages": {
      "Biobase": {
        "Package": "Biobase",
        "Version": "2.66.0",
        "Source": "Bioconductor",
        "Repository": "Bioconductor 3.20",
        "Requirements": [
          "BiocGenerics",
          "R",
          "methods",
          "utils"
        ],
        "Hash": "f6e716bd9ed8acfd2d4137be7d4fa8f9"
      },
      "BiocGenerics": {
        "Package": "BiocGenerics",
        "Version": "0.52.0",
        "Source": "Bioconductor",
        "Repository": "Bioconductor 3.20",
        "Requirements": [
          "R",
          "graphics",
          "methods",
          "stats",
          "utils"
        ],
        "Hash": "3a1a587cfadcfcbf849dfc605cbbb965"
      }
    }
  }
}
```

An improved R workflow:

renv example

```
# install.packages('dplyr')
# install.packages('ggplot2')

library(dplyr)
library(ggplot2)
```

An improved R workflow:

renv example

```
> renv::snapshot()  
The following package(s) will be updated in the lockfile:
```

```
# CRAN  
-----  
- cli [ * -> 3.6.4]  
- colorspace [ * -> 2.1-1]  
- dplyr [ * -> 1.1.4]  
- fansi [ * -> 1.0.6]  
- farver [ * -> 2.1.2]  
- generics [ * -> 0.1.3]  
- ggplot2 [ * -> 3.5.1]  
...  
Do you want to proceed? [Y/n]: y
```

```
- Lockfile written to "~/projects/renv3/renv.lock".  
>
```

An improved R workflow:

`renv::restore()` example

```
R version 4.4.3 (2025-02-28) -- "Trophy Case"  
Copyright (C) 2025 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin20
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.
```

Natural language support but running in an English locale

```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

```
- Project '~/.projects/renv_with_bioconductor' loaded. [renv 1.1.4]  
- The project is out-of-sync -- use `renv::status()` for details.  
[Workspace loaded from ~/.projects/renv_with_bioconductor/.RData]
```

An improved R workflow:

renv example

```
> renv::status()
The following package(s) are in an inconsistent state:

  package      installed recorded used
  dplyr        y            n        y
  generics      y            n        y
  tidyselect    y            n        y

The following package(s) are out of sync [lockfile != library]:

# CRAN
-----
- cluster     [2.1.6 != 2.1.8]
- MASS        [7.3-61 != 7.3-64]
- Matrix       [1.7-1 != 1.7-2]
- nlme        [3.1-166 != 3.1-167]
- survival    [3.7-0 != 3.8-3]

See `?renv::status` for advice on resolving these issues.
```

An improved R workflow:

renv example

```
> renv::restore()
The following package(s) will be updated:

# CRAN
-----
- cluster      [2.1.8 -> 2.1.6]
- MASS         [7.3-64 -> 7.3-61]
- Matrix        [1.7-2 -> 1.7-1]
- nlme          [3.1-167 -> 3.1-166]
- survival     [3.8-3 -> 3.7-0]

Do you want to proceed? [Y/n]: y
```

An improved R workflow: **renv** example

**package installation
may fail if a package
was originally installed
through a binary...
but that binary is no
longer available!**

```
Retrieving 'https://mran.microsoft.com/snapshot/2023-11-07/bin/macosx/contrib/4.2/Matrix_1.6-1.1.tgz' ...
Retrieving 'https://cloud.r-project.org/src/contrib/Archive/Matrix/Matrix_1.6-1.1.tar.gz' ...
  OK [downloaded 2 Mb in 0.5 secs]
Retrieving 'https://mran.microsoft.com/snapshot/2023-12-06/bin/macosx/contrib/4.2/mgcv_1.9-0.tgz' ...
Retrieving 'https://cloud.r-project.org/src/contrib/Archive/mgcv/mgcv_1.9-0.tar.gz' ...
  OK [downloaded 1.1 Mb in 0.3 secs]
Retrieving 'https://mran.microsoft.com/snapshot/2023-11-27/bin/macosx/contrib/4.2/nlme_3.1-163.tgz' ...
Warning: failed to find binary for 'mgcv 1.9-0' in package repositories
Warning: failed to retrieve 'https://mran.microsoft.com/snapshot/2023-12-06/bin/macosx/contrib/4.2/mgcv_1
Warning: failed to find source for 'mgcv 1.9-0' in package repositories
Warning: failed to find binary for 'nlme 3.1-163' in package repositories
Warning: failed to retrieve 'https://mran.microsoft.com/snapshot/2023-11-27/bin/macosx/contrib/4.2/nlme_3
Error: failed to retrieve package 'mgcv@1.9-0'
In addition: Warning messages:
1: curl: (22) The requested URL returned error: 403
2: curl: (6) Could not resolve host: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.co
ost: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.com
3: failed to retrieve 'https://mran.microsoft.com/snapshot/2023-11-28/bin/macosx/contrib/4.2/foreign_0.8-8
4: curl: (6) Could not resolve host: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.co
ost: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.com
5: failed to retrieve 'https://mran.microsoft.com/snapshot/2023-11-07/bin/macosx/contrib/4.2/Matrix_1.6-1
6: curl: (6) Could not resolve host: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.co
ost: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.com
7: failed to retrieve 'https://mran.microsoft.com/snapshot/2023-12-06/bin/macosx/contrib/4.2/mgcv_1.9-0.tg
8: curl: (6) Could not resolve host: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.co
ost: mran.microsoft.comcurl: (6) Could not resolve host: mran.microsoft.com
```

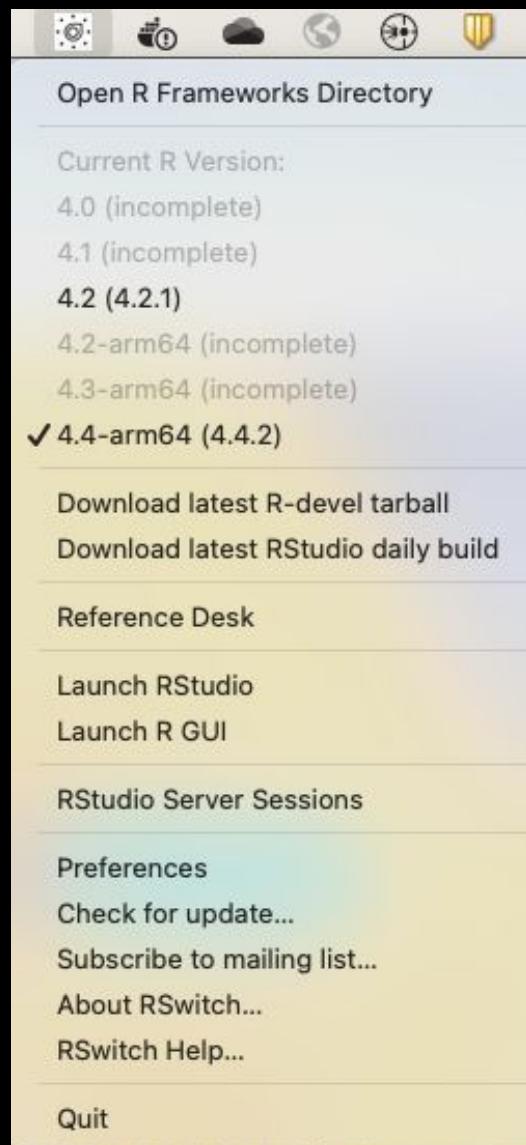
But how do you freeze the operating system, versions of system libraries, compiler versions, etc.?

renv can also be used with Docker!

`vignette("docker", package = "renv")`

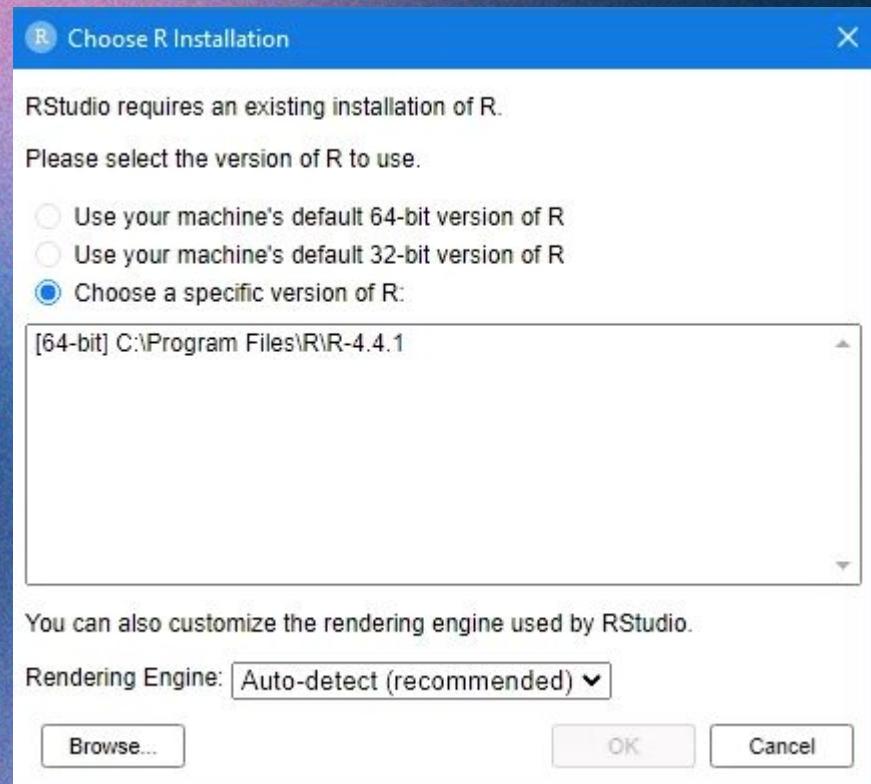
`(https://rstudio.github.io/renv/articles/docker.html)`

What about different versions of R?



**Mac:
Switch more easily using RSwitch**

**Windows:
Tools → General Options**

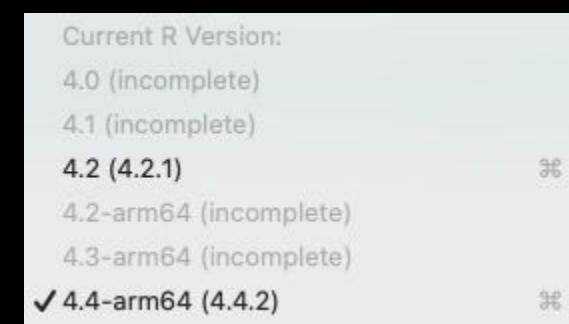
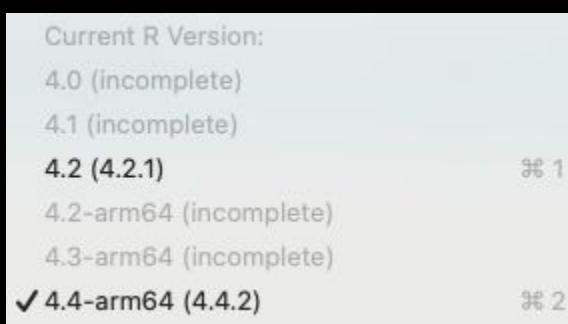


**Linux:
export RSTUDIO_WHICH_R=/usr/local/bin/R**

What about different versions of R?

Mac:

RSwitch only distinguishes between minor (not "point") releases



What about different versions of R?

Also consider Rig

<https://github.com/r-lib/rig>

Summary

- RStudio projects
- Git/GitHub
- `renv` to manage package versions
- `RSwitch` to manage multiple R versions

Thanks!

kerchner@gwu.edu