Retour sur la formation

Bonnes pratiques pour une recherche reproductible en écologie numérique

Les principaux outils - Data toolbox

FRB
FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ

CESAB
CENTRE FOR THE SYNTHESIS AND ANALYSIS OF BIODIVERSITY

GdR EcoStat
GROUPEMENT DE RECHERCHE EN ECOLOGIE STATISTIQUE

Reproductible research

Data + Meta-data + Materials/Methods

Data Management: keep raw data, script the changes

Software: Conventions when writing your code, scripts

Software environment

Workflow

Tracking Changes: versionning

Collaboration: sharing the code with others

Project Organization: Research compendium

Manuscripts: automate the manuscript compilation process as much as possible

Reproductible research

Data + Meta-data + Materials/Methods

Data Management: keep raw data, script the changes

Software: Conventions when writing your code, scripts

Software environment → **Renv**

Workflow → **Targets**

Tracking Changes: versionning → **Git**

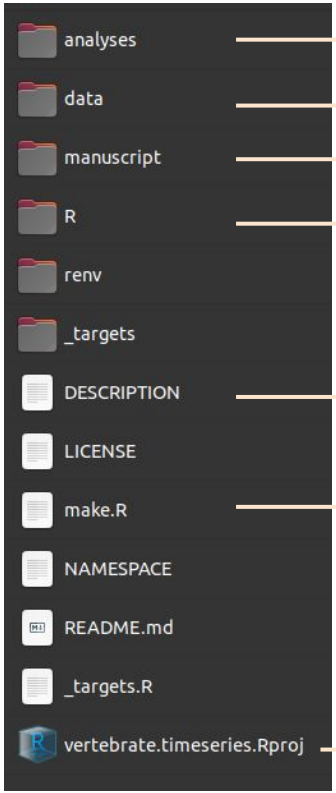
Collaboration: sharing the code with others → **GitHub**

Project Organization: **Research compendium**

Manuscripts: automate the manuscript compilation process as much as possible → **RMarkdown**

Research compendium

Organizing your project



Methods (Scripts R)

Data

Results (Outputs = manuscripts, plots…)

Only R files with function definition

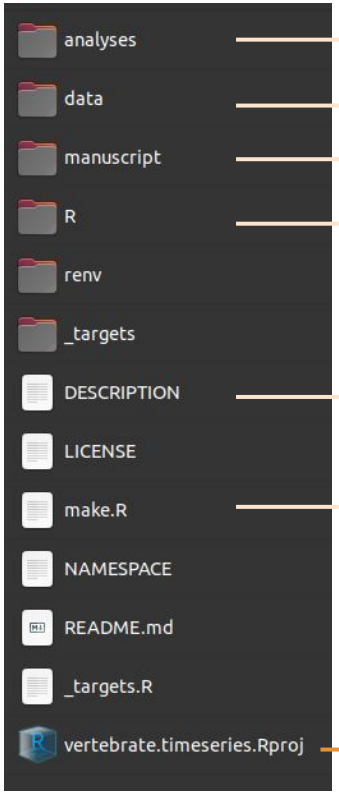
Listing of packages used (`usethis`)

at the root, installation/update of packages (`devtools`), **call to the different analyses**

Rproj at the root

Research compendium

Organizing your project



Methods (Scripts R)

Data

Results (Outputs = manuscripts, plots…)

Only R files with function definition

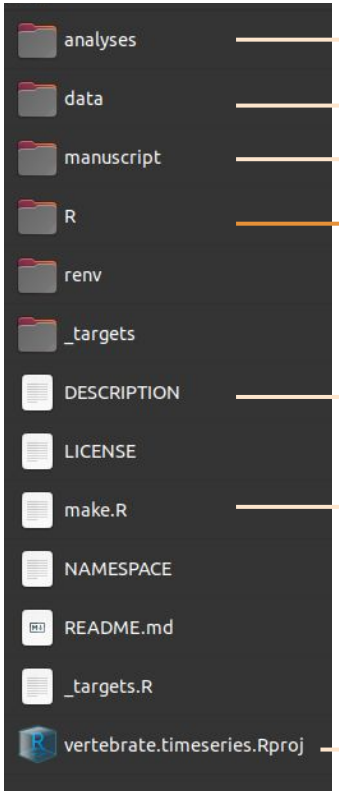
Listing of packages used (`usethis`)

at the root, installation/update of packages (`devtools`), **call to the different analyses**

Rproj at the root (1) *Relative paths will work for everyone*

Research compendium

Organizing your project



Methods (Scripts R)

Data

Results (Outputs = manuscripts, plots…)

Only R files with function definition

2

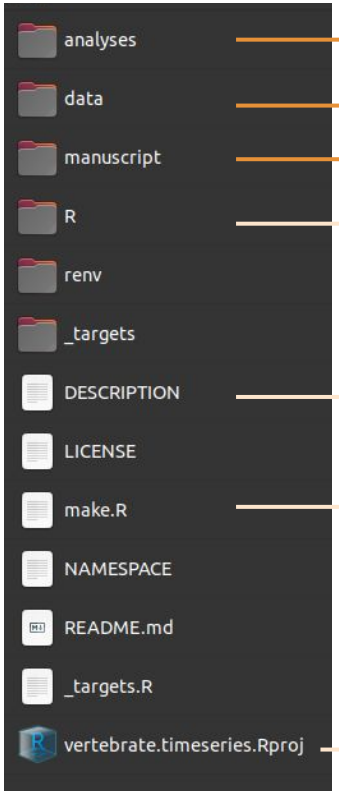
Listing of packages used (`usethis`)

at the root, installation/update of packages (`devtools`), **call to the different analyses**

Rproj at the root

Research compendium

Organizing your project



Methods (Scripts R)

Data

Results (Outputs = manuscripts, plots…)

Only R files with function definition

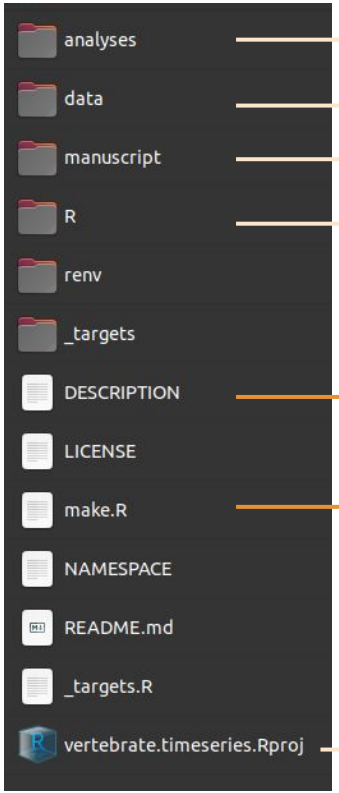
Listing of packages used (`usethis`)

at the root, installation/update of packages (`devtools`), **call to the different analyses**

Rproj at the root

Research compendium

Organizing your project



Methods (Scripts R)

Data

Results (Outputs = manuscripts, plots…)

Only R files with function definition

Listing of packages used (`usethis`)

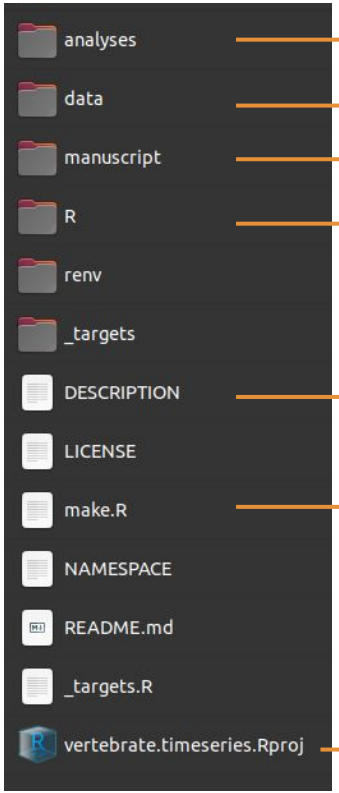
at the root, installation/update of packages (`devtools`), **call to the different analyses**

```
devtools::load_all()
devtools::install_deps()
usethis::use_package("name")
```

Rproj at the root

Research compendium

Organizing your project



Methods (Scripts R)

Data

Results (Outputs = manuscripts, plots…)

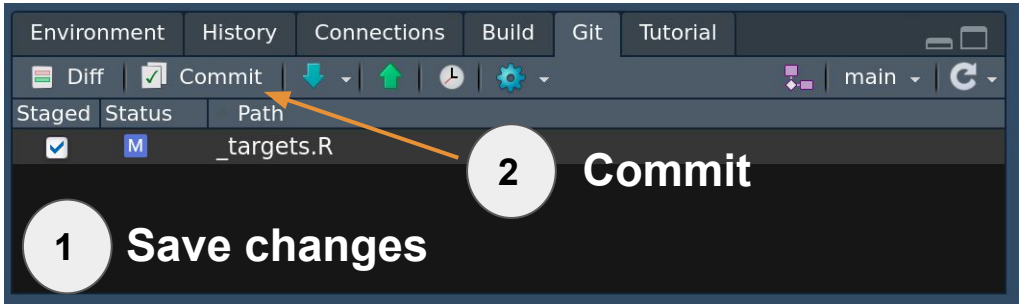
Only R files with function definition

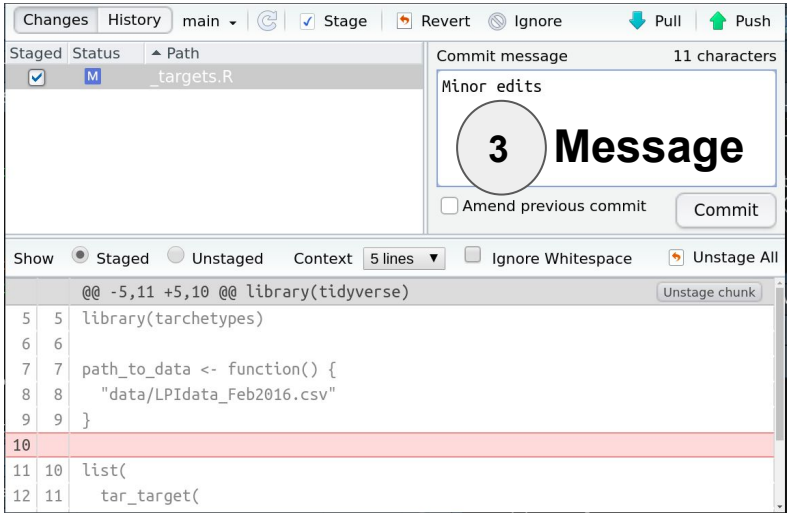
Listing of packages used (`usethis`)

at the root, installation/update of packages (`devtools`), **call to the different analyses**

Rproj at the root

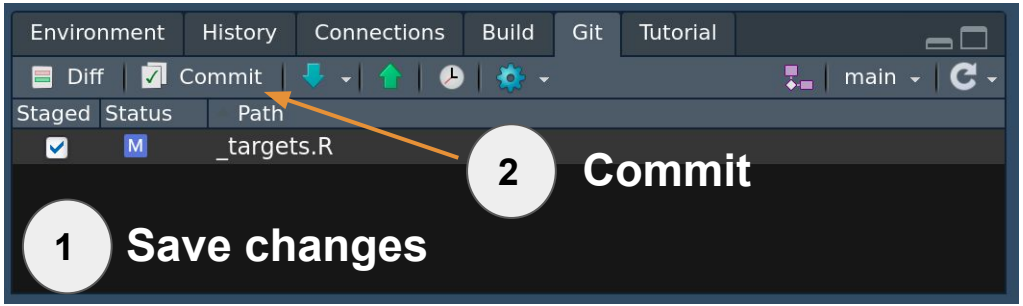
Git / GitHub

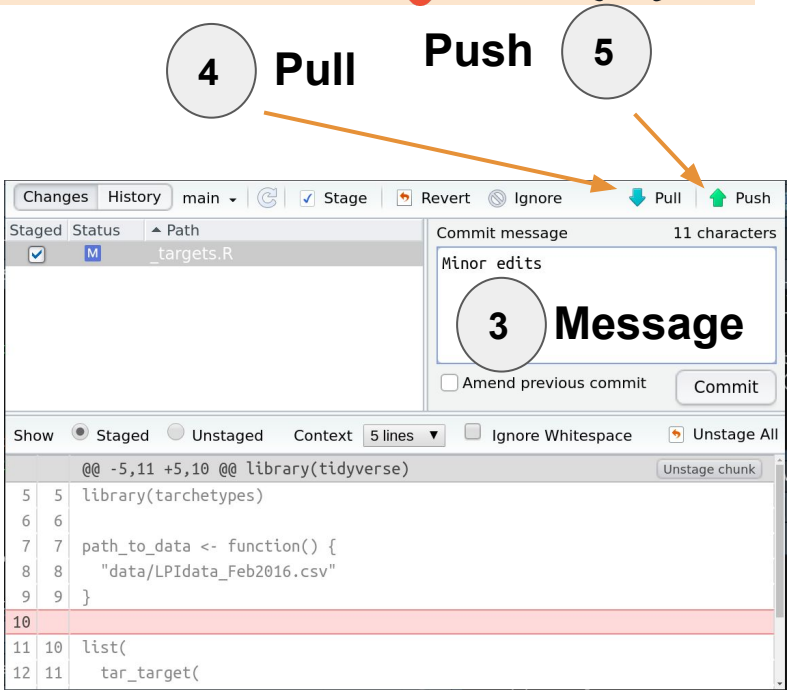




Keep versions of your code

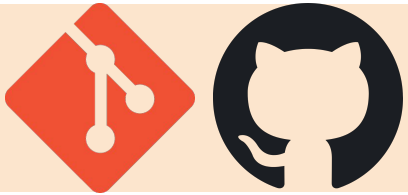
Git / **GitHub**





Keep versions of your code

Git / **GitHub**



6 **Share it with collaborators**

matpelissie / **vertebrate.timeseries** Public

Unwatch 1

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

main | 1 branch | 0 tags | Go to file | Add file | Code

matpelissie Minor edits — 2c04fb5 — 11 seconds ago — 170 commits

R	Population vs. temperature change models	5 days ago
_targets	target maj	last month
analyses	Population vs. temperature change models	5 days ago
data	Join duplication raw problem SOLVED!	5 days ago
manuscript	Minor edits	13 minutes ago
renv/library/R-4.1/x86_64-pc-linux-gnu	Various changes	last month
.Rbuildignore	renv update	last month
.gitignore	Add data LPI	last month
DESCRIPTION	resolution render	last month
LICENSE	Initial commit	last month
NAMESPACE	use_compendium function	last month
README.md	Rationale of the project added	19 hours ago
_targets.R	Minor edits	11 seconds ago
make.R	resolution render	last month
vertebrate.timeseries.Rproj	Data viz	last month

README.md

vertebrate.timeseries

About

No description, website, or topics provided.

Readme
CC0-1.0 License
2 stars
1 watching
1 fork

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 2

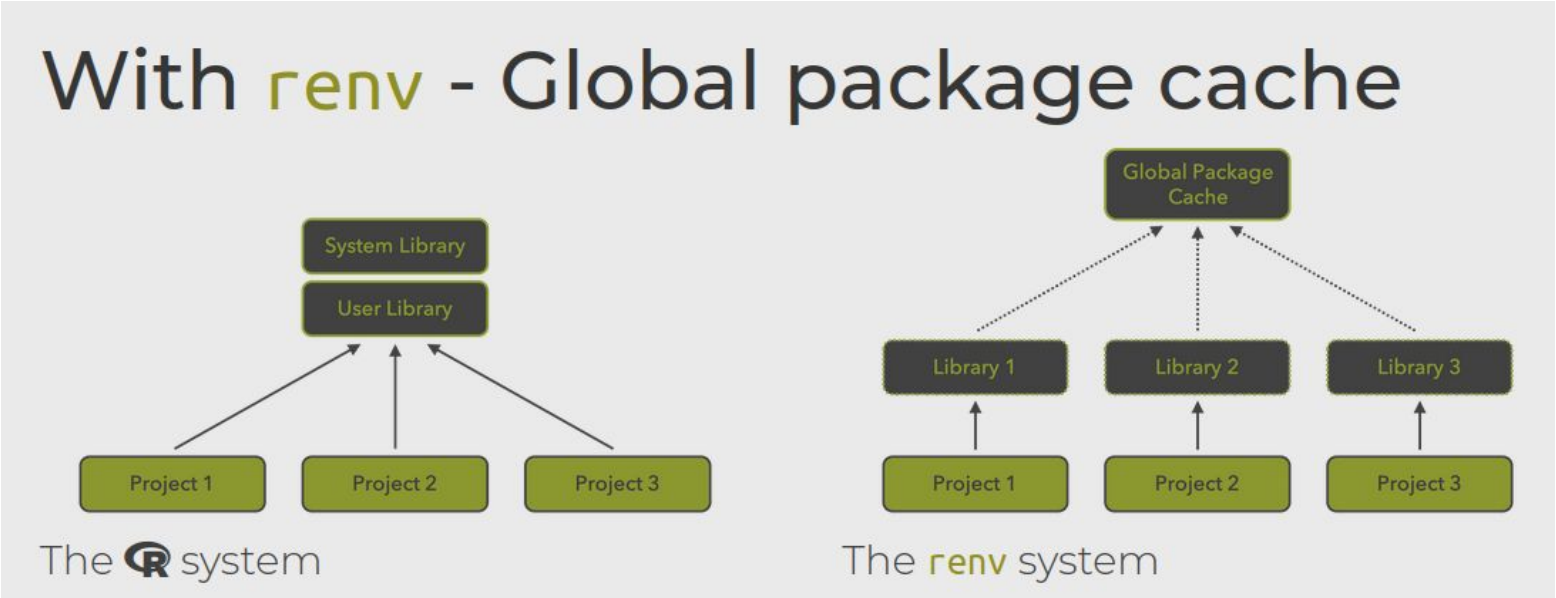
matpelissie
MaelysBoennec

Languages

HTML 75.2% JavaScript 10.0%

Renv

Keep **package version**, run on **other machines**



renv.lock file to share (on GitHub)

Renv


```r
## Initiate renv for the project ----
renv::init()

## Install < pkg_name > for the project ----
renv::install("pkg_name")
renv::install("pkg_name@version")
renv::install("github/pkg_name")

## Install packages listed in DESCRIPTION (and/or R and Rmd files) ----
renv::install()

## Check renv status ----
renv::status()

## Update lockfile (save local environment) ----
renv::snapshot()

## Uninstall unused packages ----
renv::clean()

## Restore local environment ----
renv::restore()
```

Targets

Visualize **links** between **data**, **scripts**, and **results** & keep them **up-to-date**

DEFINITION WORKFLOW

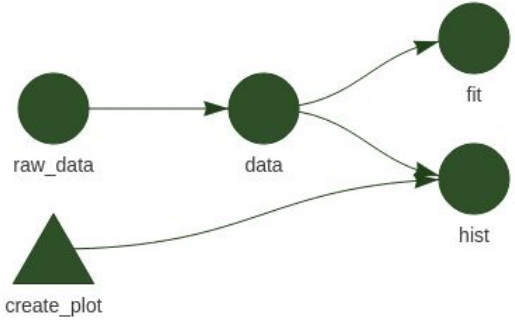
1. Write a function → dans fichier **`make.R`**
2. Add a target to the pipeline → dans fichier **`_targets.R`**
3. Visualize pipeline → dans la console (**`tar_visnetwork() ou tar_glimpse()`**)
4. Make the pipeline → dans la console (**`tar_make()`**)
5. Check the results → dans la console (**`tar_read() ou tar_load()`**)
6. Correct

Targets

Visualize **links** between **data**, **scripts**, and **results** & keep them **up-to-date**

DEFINITION WORKFLOW

1. Write a function → dans fichier **`make.R`**
2. Add a target to the pipeline → dans fichier **`_targets.R`**
3. Visualize pipeline → dans la console (**`tar__visnetwork() ou tar_glimpse()`**)
4. Make the pipeline → dans la console (**`tar_make()`**)
5. Check the results → dans la console (**`tar_read() ou tar_load()`**)
6. Correct




```r
library(targets)

create_plot <- function(data) {
  ggplot(data) +
    geom_histogram(aes(x = Ozone), bins = 12) +
    theme_gray(24)
}

list(
  tar_target(raw_data, airquality),

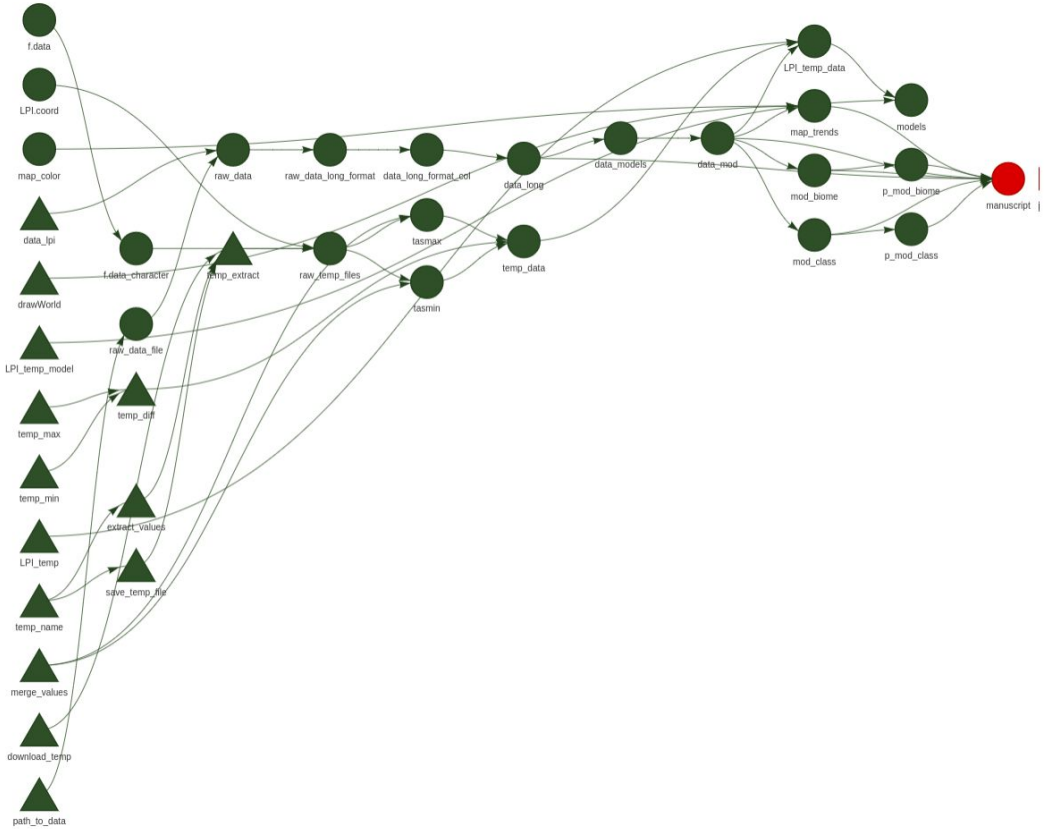
  tar_target(data, raw_data %>% filter(!is.na(Ozone))),

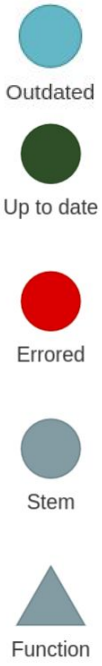
  tar_target(hist, create_plot(data)),

  tar_target(fit, {

    biglm(Ozone ~ Wind + Temp, data)
  })
```

Targets


f.data
LPI.coord
map_color
data_lpi
drawWorld
LPI_temp_model
temp_max
temp_min
LPI_temp
temp_name
merge_values
download_temp
path_to_data
raw_data_file
f.data_character
temp_dir
extract_values
save_temp_file
temp_extract
raw_data
raw_data_long_format
data_long_format_eof
raw_temp_files
tasmax
tasmin
data_long
temp_data
data_models
data_mod
LPI_temp_data
map_trends
mod_biome
mod_class
models
p_mod_biome
p_mod_class
manuscript


Outdated
Up to date
Errored
Stem
Function

Rmarkdown

Integrate **analyses** and **text** in the same document

In practice

Header / Content / Code chunks

Biblio

Export (Knit)

Rmarkdown

.Rmd

```
## Results

The following map illustrates the distribution of the time series. Each
point is colored according to its trend and sized according to the magnitude of
its trend.

```{r, echo=FALSE}
map_trends
```

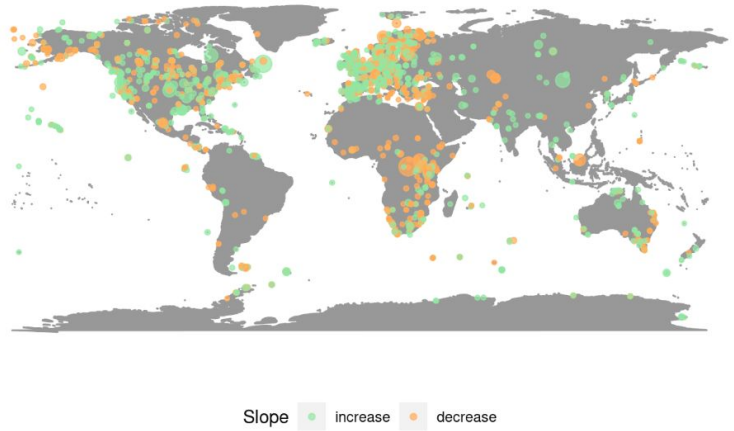
Among the `r length(unique(LPI.long$id))`, `r sum(LPI.mod$slope_p<0.05 & LPI.mod$slope>0)` are
increasing, `r sum(LPI.mod$slope_p<0.05 & LPI.mod$slope<0)` are decreasing and
`r sum(LPI.mod$slope_p>=0.05)` are showing constant trends (p>0.05).
```

.html

Results

The following map illustrates the distribution of the time series. Each point is colored according to its trend and sized according to the magnitude of its trend.

Figure 1: Terrestrial vertebrates population declines and increases worldwide.

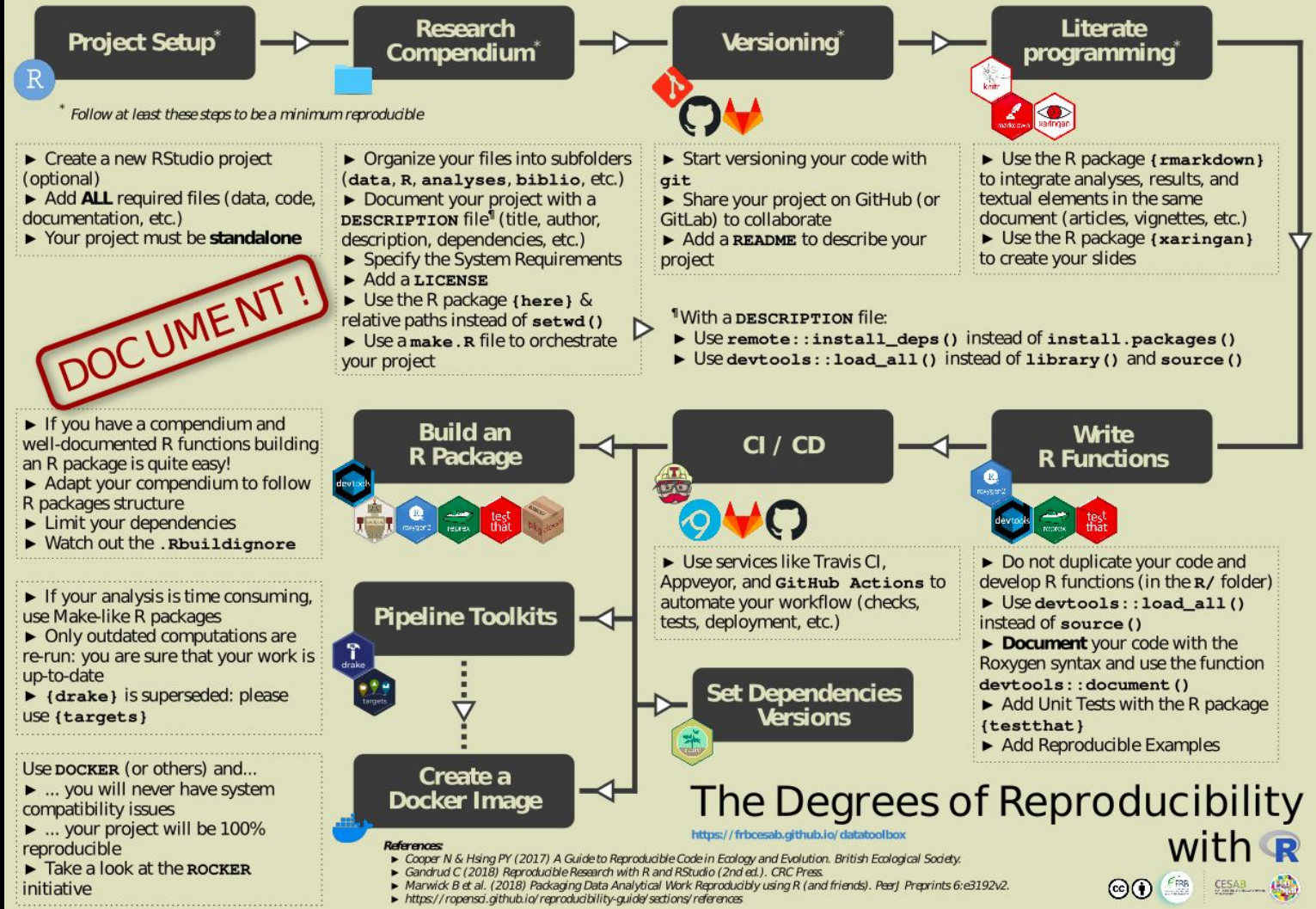


Among the 5529, 1920 are increasing, 1742 are decreasing and 1573 are showing constant trends (p>0.05).

Take home message

- ➔ Tidy your project directories 
- ➔ Keep track of changes and versions   
- ➔ Automate your workflow  

Small effort to adopt => **Save much time!**



https://rdatatoolbox.github.io/img/the_degrees_of_reproductibility.pd