

Transforming scattered analyses into a documented, reproducible and shareable workflow

FOSDEM Brussels, 2020

Sébastien Rochette



This presentation on Github: [statnmap/prez](https://github.com/statnmap/prez)

Sébastien



Data Scientist,  expert,  trainer

- <https://statnmap.com>
- <https://rtask.thinkr.fr>
- <https://github.com/ThinkR-open>
- https://twitter.com/thinkr_fr



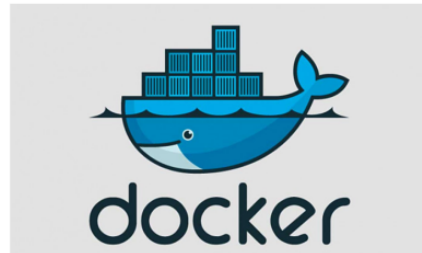
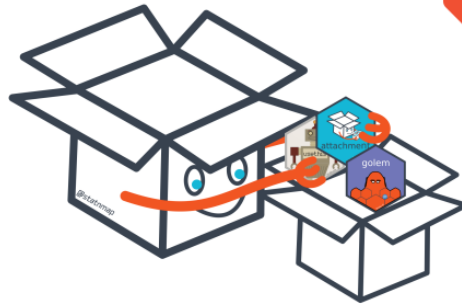
Collaboration fest about R-software



Station Marine de Concarneau - MNHN

Collaboration fest about R-software

- Good practices
- Code sharing



Efficiently sharing data

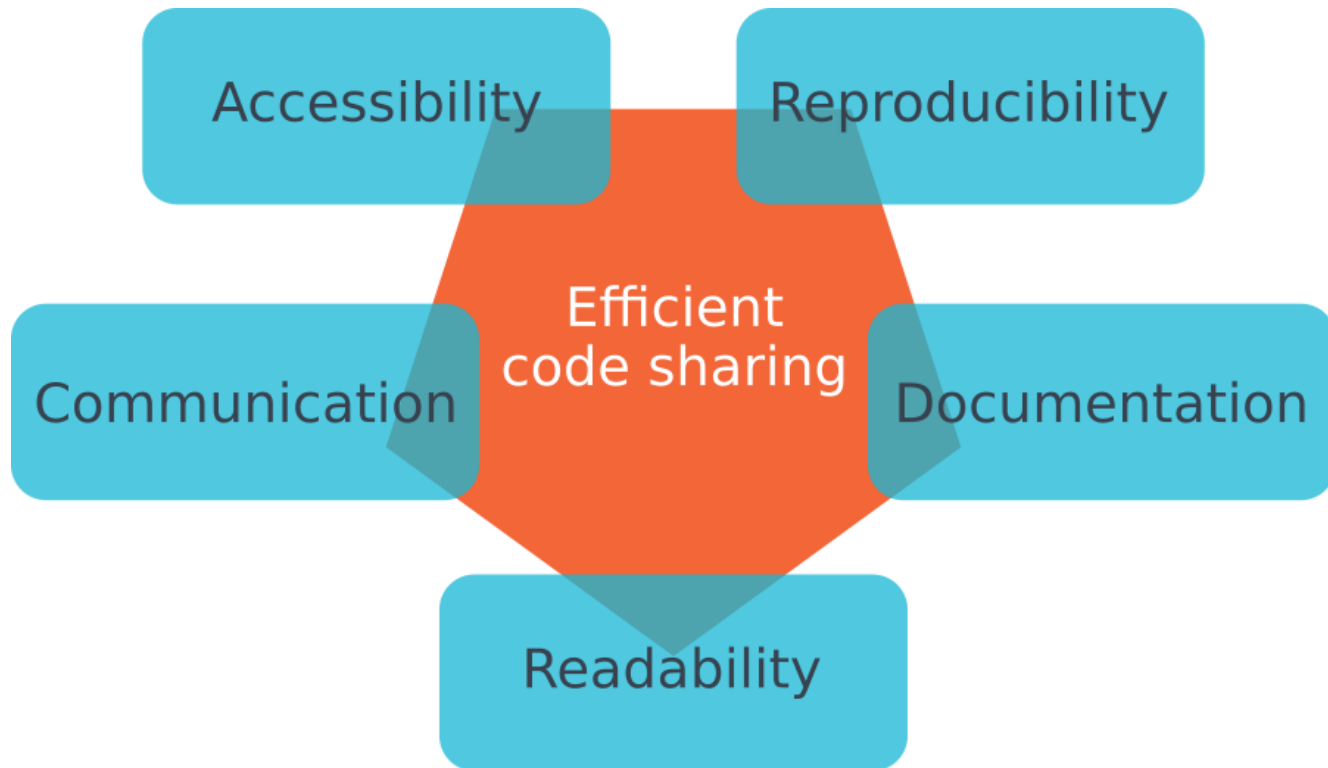
FAIR principles

- Findable
- Accessible
- Interoperable
- Re-usable



www.go-fair.org

Efficiently sharing code



A big step

From coding for yourself...



...to sharing with a community

Do not stay alone



A common project

Let's practice together !



www.vigienature.fr/fr/chaues-souris

Accessibility



cesco-lab / **Vigie-Chiro_scripts** Watch 3 Star 1 Fork 7

Code Issues 6 Pull requests 1 Projects 0 Wiki Security Insights

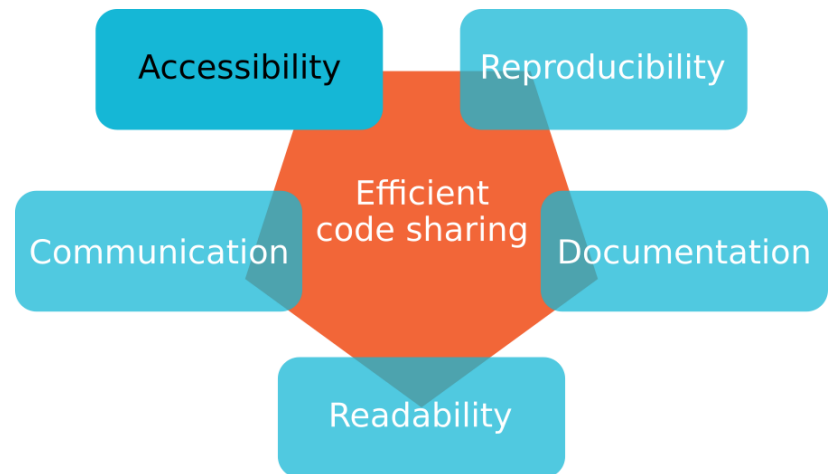
scripts to manage and analyse Vigie-Chiro data

190 commits 1 branch 0 releases 5 contributors GPL-3.0

Tree: 89348fcd19 New pull request Create new file Upload files Find file Clone or download

YvesBas generic biodiv/GI randomforest modelisation Latest commit 89348fc yesterday

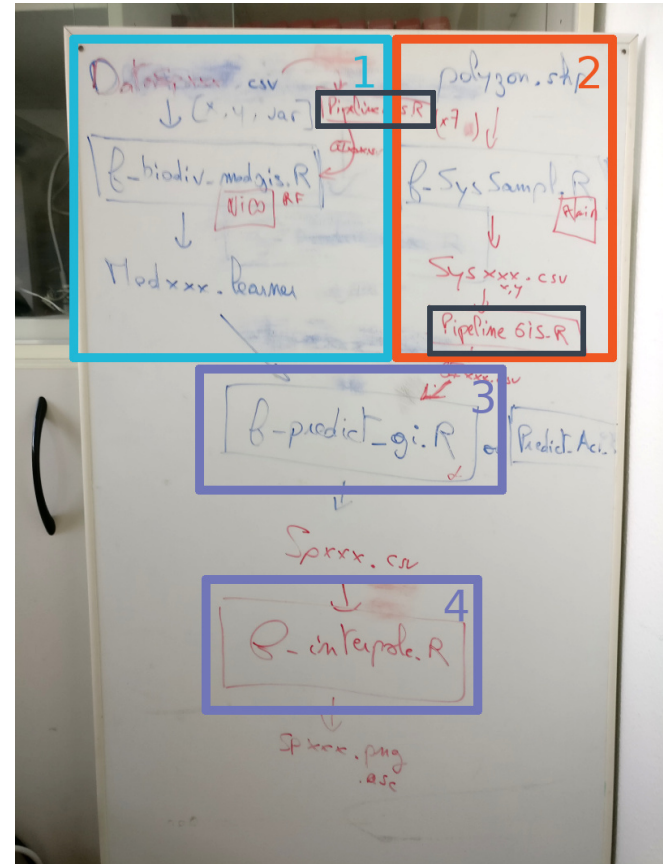
AppShiny	adapted shiny app for vigie-chiro portal outputs	2 months ago
IO	add "carre stoc" sample for test	11 months ago
functions	generic biodiv/GI randomforest modelisation	yesterday
.gitattributes	Initial commit	last year
.gitignore	Update .gitignore	7 months ago
ACombiner.R	Create ACombiner.R	10 months ago
Alnscire.r	tidy	7 months ago
ActFilter.r	extracting activity data on a limited time window within night	2 months ago
ActNew.r	change path name parameter	2 days ago
ActTot.r	separate update and combine list	4 months ago
Act_SplitSp.R	include a filter on channel	4 months ago
AggTC.R	add scripts	last year
AnoPartPoint_pem.R	add scripts	last year
Anomalie_boxplots.R	compute temperatures anomalies	11 months ago
Anomalies_noms_fichiers.R	add scripts	last year
Autocorr.R	compute spatial semi-variograms	7 months ago
Bilan_Annuel_Participation_Vigi...	add filters by region	11 months ago



Management

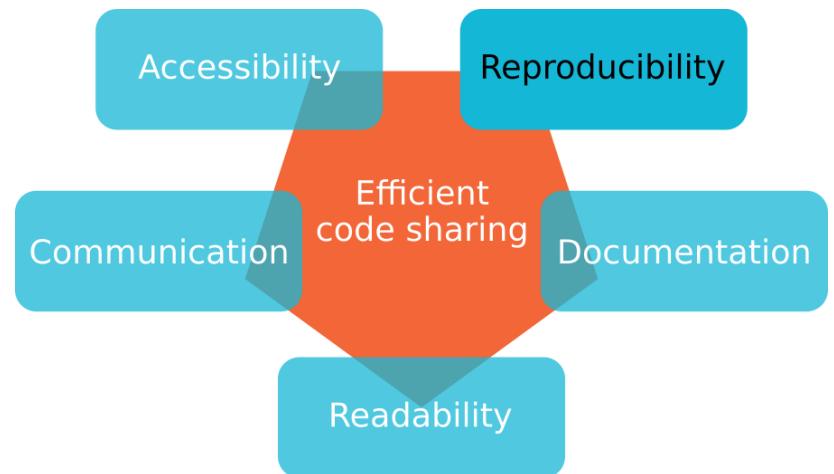
- Dissection
- Separate work into small pieces
- Document issues
- Dispatch
- Manage repository

I felt naked while presenting my code and my project.



Carefully peel back the code

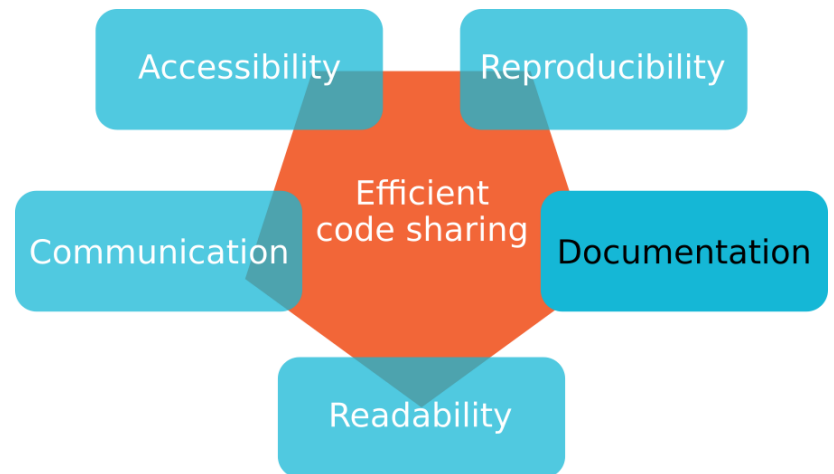
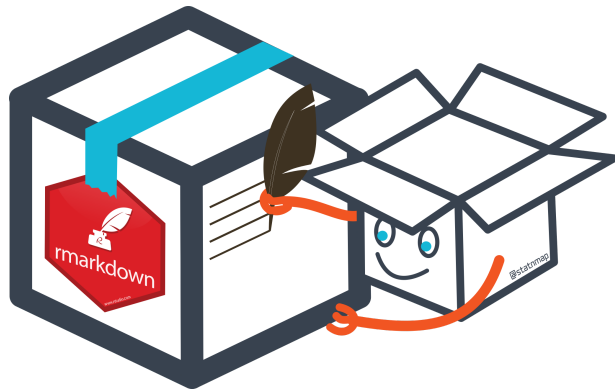
- Identify user-specific code
- Parametrize
- Cut code into pieces
- Functions
- *Reprex*
- Unit tests



Document generously

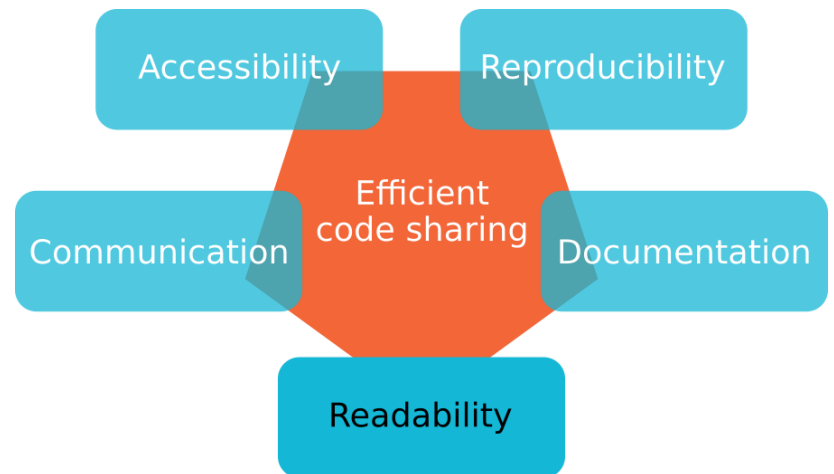
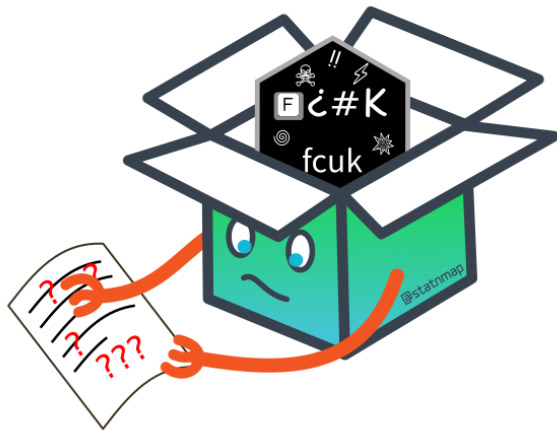
- Vignettes (Notebook)
- Use *reprex*
- Document functions

Transform as package



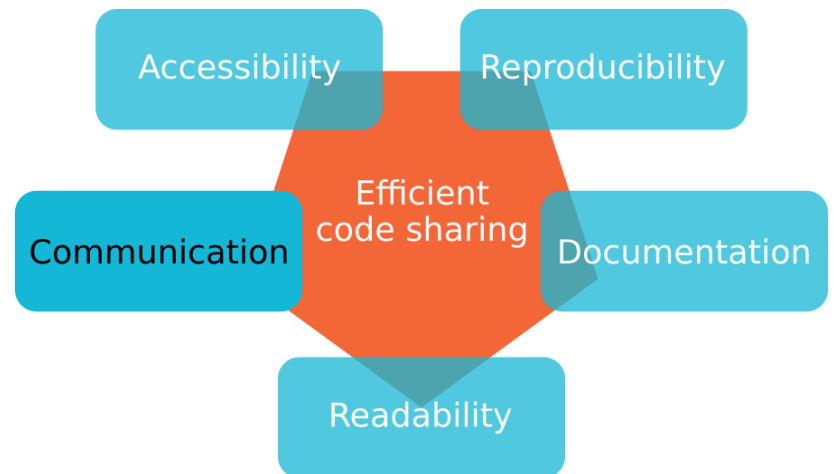
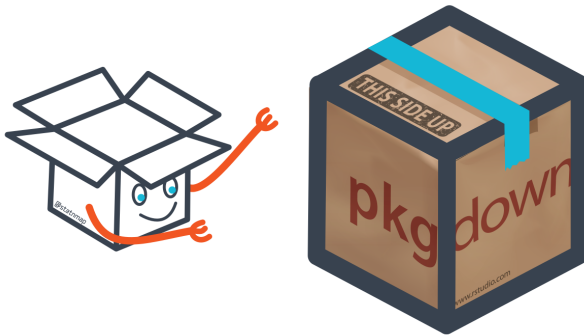
Add the right amount of readability

- Air out the code
- Future you
- Other developers



Communicate abundantly

- Social media
- Blog
- Website



<https://thinkr.fr/transformer-plusieurs-scripts-eparpilles-en-beau-package-r/>

Communicate abundantly



https://cesco-lab.github.io/Vigie-Chiro_scripts/index.html

chirovigie 0.0.1

Vigie-Chiro_scripts

Website with {pkgdown} for documentation: https://cesco-lab.github.io/Vigie-Chiro_scripts/

Summary

The workflows are summarized [here](#) (slides 6 to 20) R scripts are colored blue, php scripts in yellow, manually edited inputs in green, and script outputs in white. Scripts and files that are consuming a lot of memory are shown in a red box.

This repository should be useful mostly for Vigie-Chiro team and participants but might also prove useful to other team interested in managing acoustic monitoring data. We'll try to update and improve documentation as much as we can, but do not hesitate to send an e-mail to yves.bas@mnhn.fr for any issue, question, bug report and so on.

Index

Please find below a description of every objects (scripts, inputs and outputs) in alphabetical order

ActTot.r

This script computes acoustic activity data (number of species occurrence among files) of every species for every full night of stationary recordings.

Links

Browse source code at https://github.com/cesco-lab/Vigie-Chiro_scripts

Report a bug at https://github.com/cesco-lab/Vigie-Chiro_scripts/issues

License

[Full license](#)

GPL-3

Developers

Yves Bas
Author, maintainer

Dev status

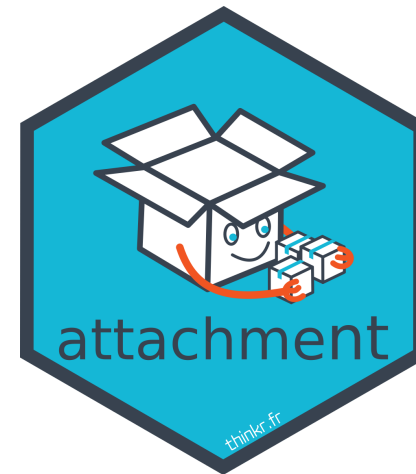
build passing

build passing

https://cesco-lab.github.io/Vigie-Chiro_scripts/index.html

Feedback on the experience

- Mentoring to start
- Accept the exposure
- Welcoming community
- Change in practice for good
- Follow up required



Conclusion

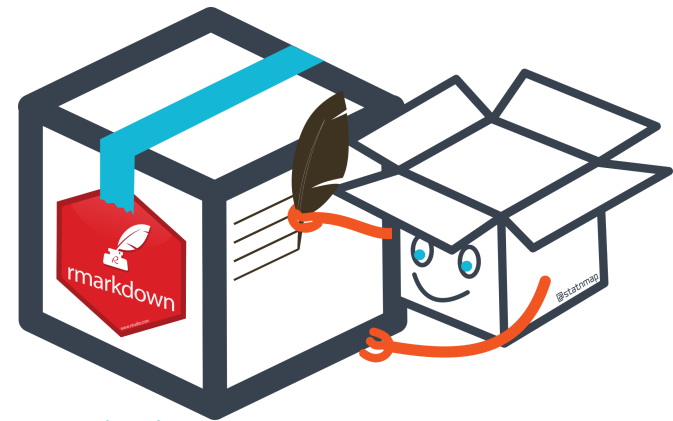
Transforming scattered analyses into a documented, reproducible and shareable workflow is accessible

- Start small
- Get used to documented, reproducible workflow
- Building a package is accessible

Start with documentation

- "Rmd first" / "Notebook first"

<https://rtask.thinkr.fr/when-development-starts-with-documentation/>



Thank you



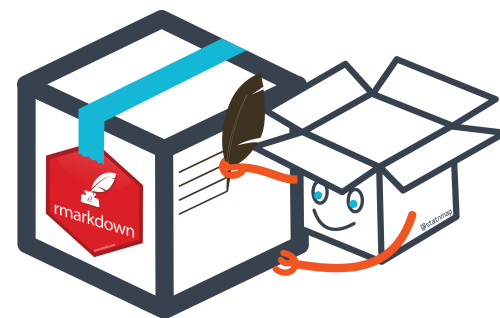
Transforming scattered analyses into a documented, reproducible and shareable workflow is accessible

- Start small
- Get used to documented, reproducible workflow
- Building a package is accessible

Start with documentation

- "Rmd first" / "Notebook first"

<https://rtask.thinkr.fr/when-development-starts-with-documentation/>



🦇 Vigie chiro

<https://rtask.thinkr.fr>
<https://thinkr.fr>

@statnmap
@thinkR_FR