

# Project-oriented workflow

Kauê de Sousa

2021

# Content

- Script vs Project
- RStudio project
- Directory structure
- Our workflow

# Lecture

This lecture is available on [Youtube](#)

# Script vs Project

# Script vs Project

## Script

- An R script is simply a text file containing a set of commands and comments. The script can be saved and used later to re-execute the saved commands
- Data analysis in a script-based approach means that you have to set your workplace every time that you need to work on the data
- Sometimes the input files are spread out in different directories and make difficult to others to reproduce the analysis. EVEN FOR YOU IF YOU CHANGE COMPUTER!!

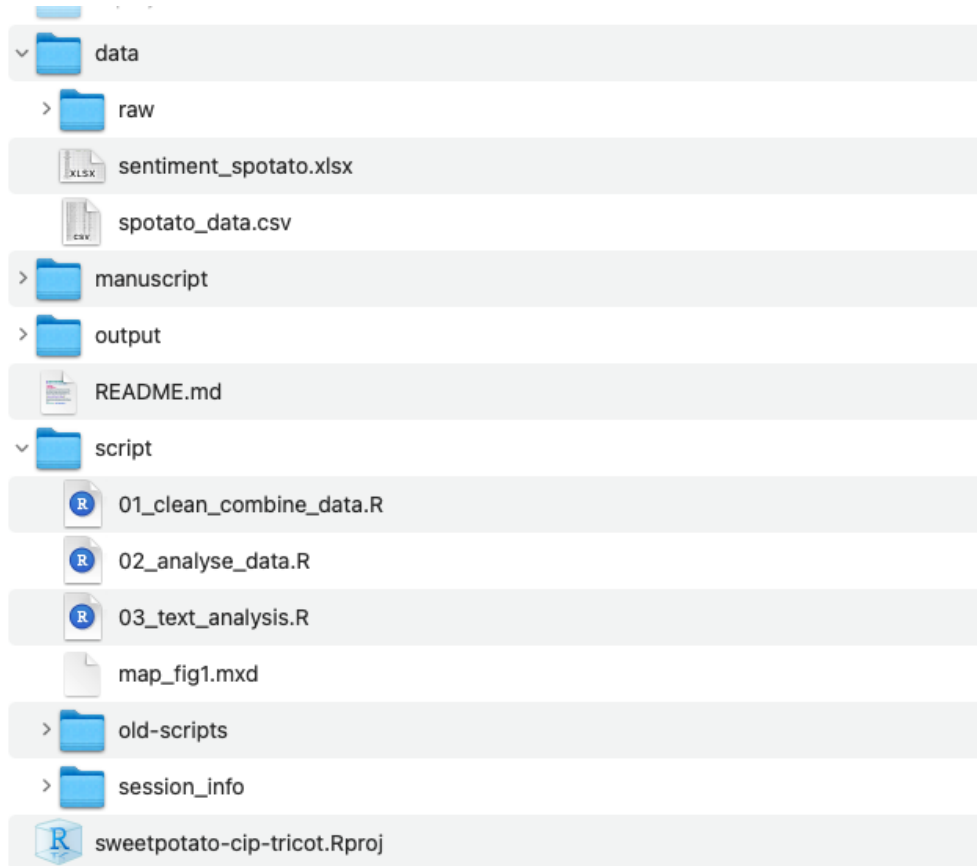
[1] [Read more here](#)

# Script vs Project

## Project

- A folder on your computer that holds all the files relevant to that particular piece of work
- Any resident R script is written assuming that it will be run from a fresh R process with working directory set to the project directory
- This convention guarantees that the project can be moved around on your computer or onto other computers and will still “just work”

[1] [Read more here](#)



(**.Rproj**) a file generated by RStudio for the R project in this directory. This is the file that should be opened to work on the project (e.g. clean data, write scripts, write outputs)

(**data**) contains all the input files that are ready to be analysed. In some cases we use a sub-directory called "raw", which may contain some sensitive data and or the raw data

(**script**) contains all the scripts used for the analysis. They are named in order from 01 to n scripts. This makes easier for others to know where to start. Old scripts (not relevant for the analysis) are maintained in a sub-directory "old-scripts". You may need them in the future

(**output**) is where all the outputs from the analysis are written.

(**manuscript**) an optional sub-directory to store the manuscript or report(s)

[1] [See this folder here](#)

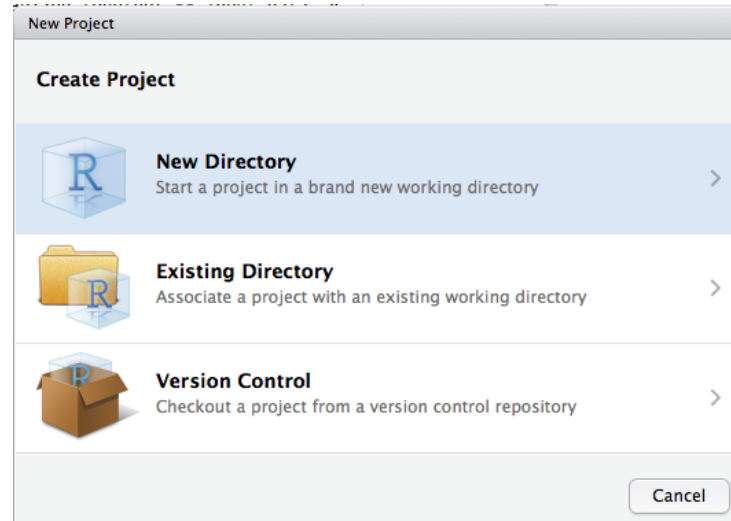
[2] [Read the paper by Moyo et al \(2021\) here](#)

# RStudio project

# RStudio project

Is a context for work on a specific project for data analysis

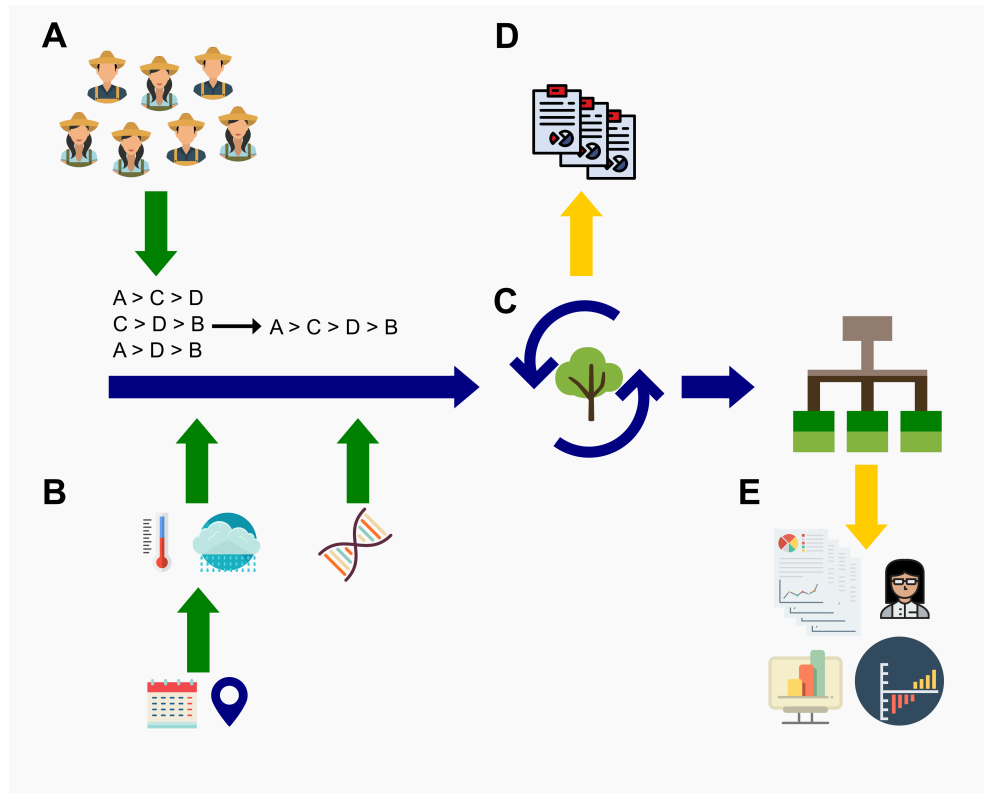
- automatically sets working directory to project folder
- has a separate workspace and command history
- its easy to share and ensure reproducibility



[1] [Read more here](#)



# Our workflow



(A) Several participants contribute with small tasks. Ranking-based approach allows the data to be easily combined.

(B) Explanatory covariates are added (e.g. using geographic coordinates and planting dates, or even DNA markers)

(C) Model selection to find the covariate(s) that best explain the data

(D) Automated reports can be generated and feedback to participants in (A) is given

(E) A stable *rpart* is used for further analysis

# R packages that we developed to support this workflow



**ClimMobTools** API Client for the 'ClimMob' platform in R

**gosset** Toolkit and a workflow to analyse metadata and experimental citizen science data

**PlackettLuce** Functions to prepare rankings data and fit the Plackett-Luce model jointly attributed to Plackett (1975) and Luce (1959)

**chirps** API Client for Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS)

**climatrends** Methods to compute precipitation and temperature indices for climate models in ecology

# Thank you!



Alliance



@desousakaue



k.desousa@cgiar.org

[Back to the main page](#)