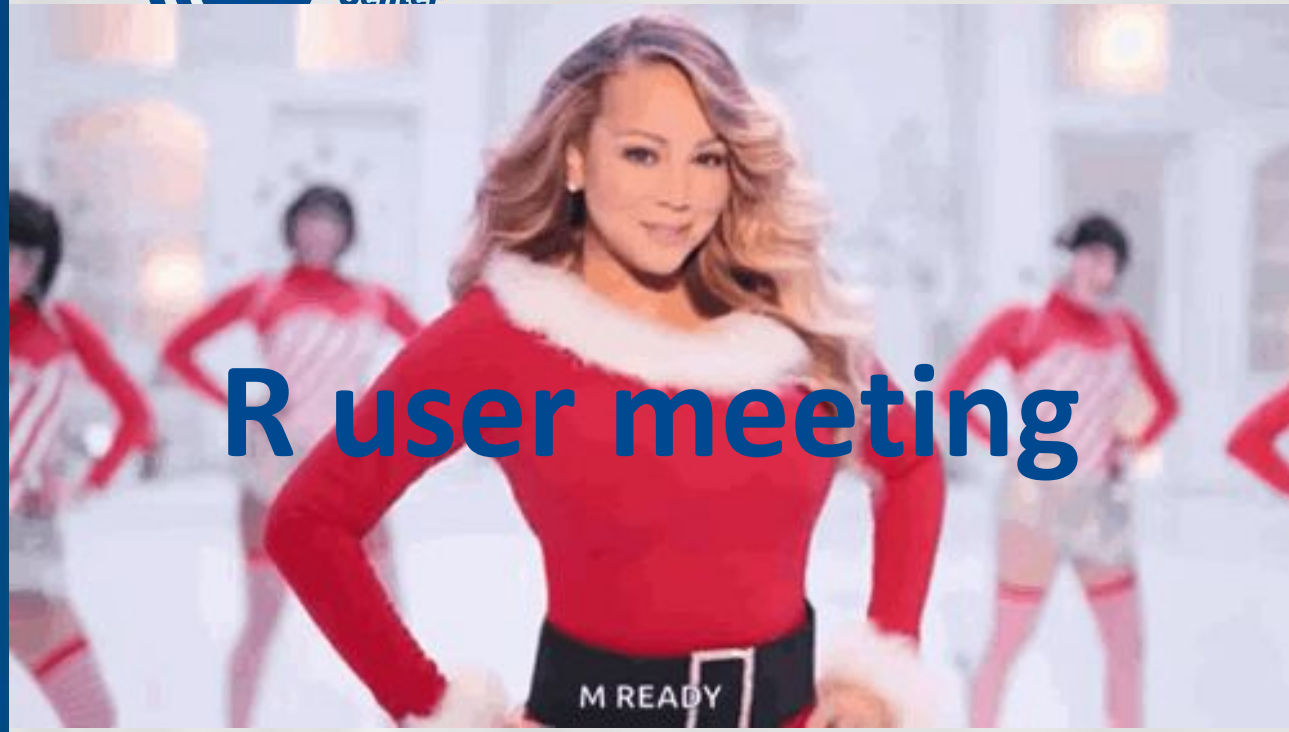
Barcelona Supercomputing Center




# R user meeting

An-Chi Ho, Eva Rifà, Victòria Agudetse

contributor: 🎄

14/12/2023

# Agenda

1. Ice-breaker: Citation
2. News
    - General
    - s2dv
    - CSTools
    - CSIndicators
    - esviz
    - SUNSET
3. Presentation: Christmas activity 🎄
4. Q&A

# Ice-breaker


BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

# Package Citation

If you use any package in your publication, it is recommended putting it in the citation.

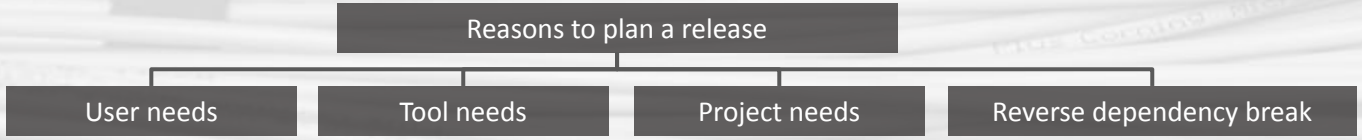
How to do it?

- Search the package's repo and CRAN page to check if they have recommendation. E.g., https://cran.r-project.org/web/packages/ggplot2/citation.html
- Open an R session and type `citation("<package_name>")`

# General


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# R TOOLS ROADMAP 2024




| | General Direction | Q1 2024 | Q2 2024 | Q3 2024 | Long term |
|---|---|---|---|---|---|
| **s2dv** (v2.0.0) | Enhance statistical methods | Include new statistical methods | | | Efficiency improvement |
| **startR** (v2.3.0) | - Maintain the flexibility while improving user experience<br>- Focus on using Autosubmit as workflow manager | Development for decadal loading in SUNSET. | - Multiple steps<br>- Code refactoring | | - Load and interpolate irregular grid<br>- Load GRIB |
| **CSTools** (v5.1.1) | Increase the general methods and enhance the relevance of the s2dv_cube | - Add use cases or tutorials<br>- Complete CST_SaveExp | | Include new methods for s2dv_cube | Convert s2dv_cube to xarray |
| **CSIndicators** (v1.1.1) | Facilitate the calculation of new indicators through existing functionalities | Develop subseasonal case and new features to SPEI functions. | - Health Indicators integration<br>- Add use cases | | Add new vignettes for bioclimatic and drought indicators |
| **esviz** (WIP) | Establish a common function structure for various needs | - Standardize function parameters<br>- Time series and scorecard functions | First release | | - Reduce device dependency<br>- Improve different projections |
| **ClimProjDiags**<br>**multiApply**<br>**easyNCDF** | No planned development. Only maintenance. | | | | |
| **General** | User meeting, support new package development by GHR, shiny server maintenance<br>While developing the packages for SUNSET or any specific project, it's important to keep the packages as general and individual as possible. | | | | |

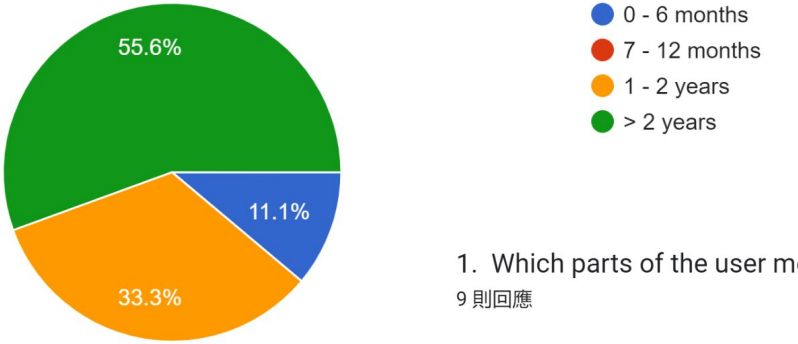
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

*Made by An-Chi Ho, Eva Rifà, Victoria Agudetse (Dec. 2023)*

# R Tool Survey

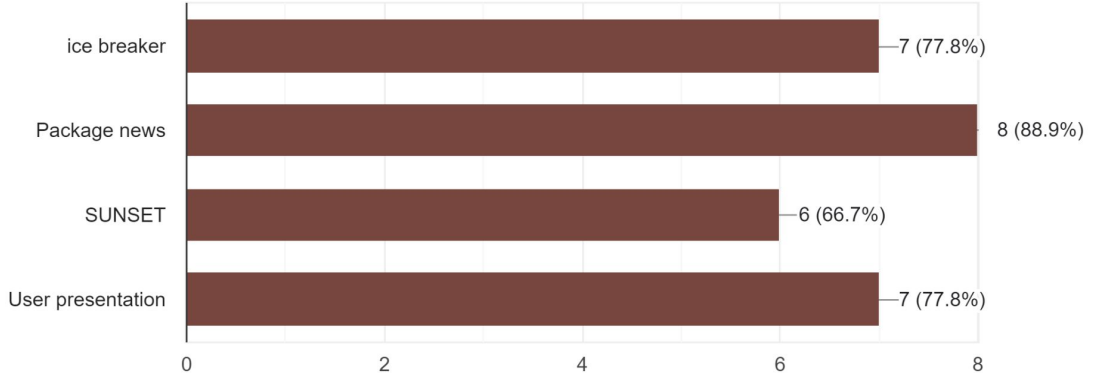
0. How long have you been in BSC R community?

9 則回應




1. Which parts of the user meeting do you find useful?

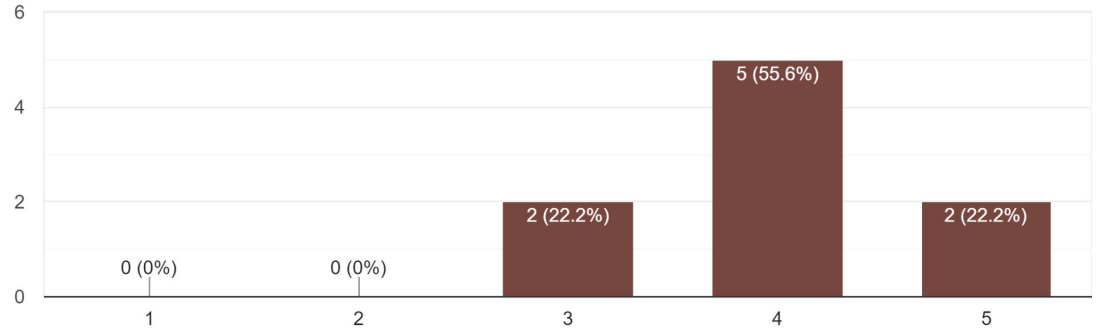
9 則回應



# R Tool Survey

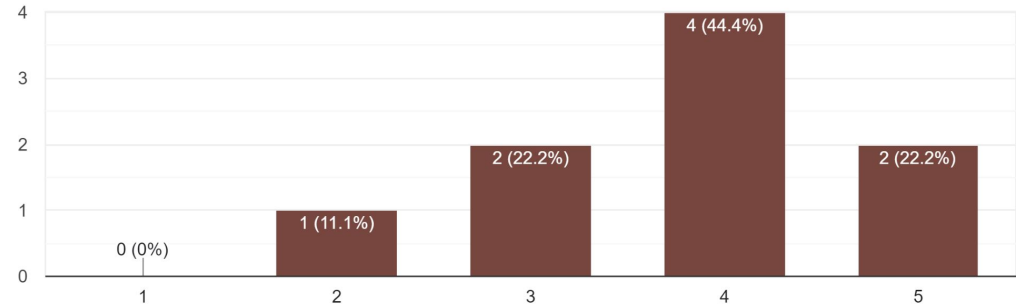
2. What percentage of the PACKAGE NEWS can you follow usually?

9 則回應




3. What percentage of the SUNSET UPDATE can you follow usually?

9 則回應

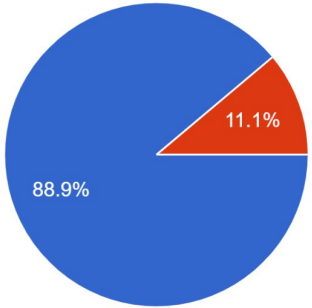



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# R Tool Survey

4. Which frequency of the meeting do you prefer?

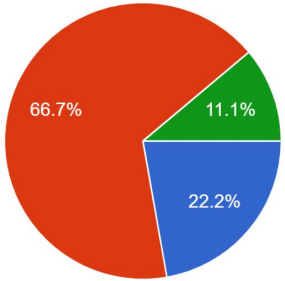
9 則回應




- monthly
- bi-monthly
- seasonal
- bi-annual
- annual
- No meeting required, I prefer other ways of development update (Choose "Others" and specify below, e.g. newsletters, ...)

5. Would you like to have an individual R user meeting? It is for general discussion on how you use the R tools.

9 則回應

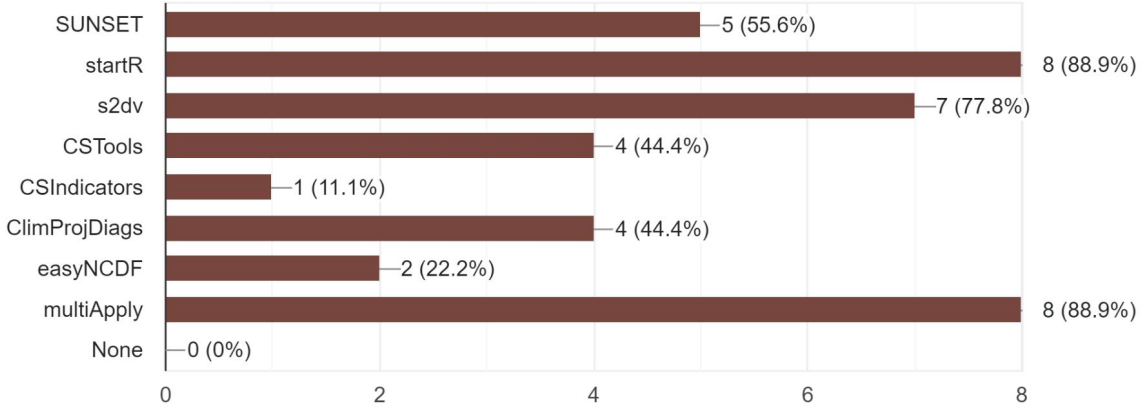



- Yes
- No
- Maybe (Choose "Others" and explain below)
- I like a specific training on SUNSET actually. But before that, I know that I should try by myself on my specific case...

# R Tool Survey

6. Which internal packages/tools do you use? Note that if you only use the package within SUNSET, you should only select SUNSET.

9 則回應




7. Do you use any external R packages that you would like to recommend?

> mapview and leaft to create interactive shiny apps to show maps

# R Tool Survey

8. What things did you find difficult when you started using R tools in our department? Is there any missing information that would be nice to have for prep-up?

Specially at the beginning but now still, there are many functions that I'm not aware that already exist in the department packages; it is useful to have some of them "advertised" in the monthly meetings (like you already do in the format of the icebreaker, maybe you could show more examples of use of internal functions)

Explanations on using array with named dimensions that is the approach of all tools

I've used the R-packages developed in the department during several years, but a few months ago I've started to use SUNSET. I think it would be helpful to keep the documentation updated.

StartR was a black box. I'm still struggling with some plotting functions of s2dv but they are in the process to be updated. And I'd like to use some functions of the CST packages since I know I can easily change a startR output to a s2dv_cube! I'll need to practice.

I found it difficult to know to how/where to find a specific function for what I needed to do.

An overview about what each R package does would have been nice at the beginning.

# Rsenal: A GitLab repo for useful R scripts

https://earth.bsc.es/gitlab/es/rsenal


This is the place to store the general R scripts or functions for the R community in BSC Earth Science Department. You can find, for example, the example scripts for some functions introduced in the R user meeting; the fun Christmas activities; useful helper functions to make your coding life easier; etc.

If you have some useful scripts or functions that may benefit others, don't hesitate to share! You can open an issue or directly a branch, explain the functionality, and we can include it in our arsenal ⚔️

# My last R user meeting… (An-Chi)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

But see you in the office still!

# s2dv


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# CDORemap() error message improvement

New parameter “print_sys_msg” to help debug when they function fails due to cdo command.

```
#'@param print_sys_msg A logical value indicating to print the messages from
system CDO commands. The default is FALSE to keep function using clean.

tas <- CDORemap(tas, lon, lat, 't170grid', 'bil', TRUE, print_sys_msg = T)

cdo    sellonlatbox (Abort): Longitudinal dimension is too small!
terminate called without an active exception
Error in CDORemap(tas, lon, lat, "t170grid", "bil", TRUE, print_sys_msg = T) :
  CDO remap failed. Set 'print_sys_msg' to TRUE to see CDO system message..
In addition: Warning message:
! Warning: CDORemap: Using CDO version 1.9.8.
```

status: In branch master

# Corr() output dimension fix

The output dimensions had some mistakes for certain cases, when dat_dim and memb_dim are NULL. It is fixed now.

The problem was only in dimensions; the output values are correct (or it would return errors).

status: In branch master

issues: https://earth.bsc.es/gitlab/es/s2dv/-/issues/111

# startR


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# Collect and merge chunk results on HPCs

The classic usage of startR is to divide a large piece of data in chunks, run the jobs on HPCs in parallel, reduce the data size significantly, then retrieve the result on local workstation. However, it limits the usage that if the data size is still larger than the workstation memory, you can't collect the data.

Now, you can collect the result by Collect() on HPCs. Check the usage in the issue (link below). Note that the current development is for ecFlow; if you use Autosubmit as workflow manager, the installed version is able to do this already.

A kind reminder: Remember to clean the chunk files (.Rds) when it is not needed to relieve the burden of esarchive.



status: In master

issues: https://earth.bsc.es/gitlab/es/startR/-/issues/189

# CSTools


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# CST_SaveExp new developments

**New features**

- Allow to save arrays with Dates that only have either `'ftime_dim'` or `'sdate_dim'`
- New parameter `units_hours_since` (only used when single_file is TRUE)
- Added parameter to save global attributes (`global_attrs`)
- Added saving time bounds (parameter `'time_bnds'`)

**Code improvements**

- Now the saving is done by `easyNCDF::ArrayToNc`. Before, the library `ncdf4` was used.
- Reorganized the initial checks, improved unit test and documentation.

**NOTE:** You can test the function with some examples here.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

status: In branch develop-SaveCube

Check issue: https://earth.bsc.es/gitlab/external/cstools/-/issues/141

# CSIndicators


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# CSIndicators new release 1.1.0 (2023-11-20)

**Fixes**

- Correct functions in order that the s2dv_cube reflect time aggregation: added `'time_bounds'` attribute
- Correct s2dv_cube output coordinates consistency.
- Change default value of `time_dim` to be `'time'`.

**New features**

- New SPEI functions: PeriodPET, PeriodAccumulation with rolling accumulation and PeriodStandardization.
- New functions to compute bioclimatic indicators: PeriodMax, PeriodMin and PeriodVariance.
- Added `'memb_dim'` parameter to MergeRefToExp.
- Substitute CST_Load by CST_Start in vignettes.

**Other**

- Improved documentation of function MergeRefToExp.
- Include again ClimProjDiags and s2dv dependencies.
- Include new publication in documentation.

# New SPEI indicator functions

- ★ **CST_PeriodPET**: function to Compute the Potential Evapotranspiration.
  - Input (CST_)PeriodPET: A named list of MD arrays (s2dv_cubes) with the needed variables
  - Methods available: 'hargreaves' (tmin and tmax), 'hargreaves_modified' (tmin, tmax and pr) and 'thornthwaite' (tmean).
- ★ **CST_PeriodAccumulation**: Returns the sum (accumulation) of a given variable in a period.
  - 2 methods: accumulation through time dimension and rolling accumulation (rollwidth)
- ★ **CST_PeriodStandardization:**
  - Transform the original values to standardized units that are comparable in space and time and at different SPEI time scales.

**NOTE:** Some development is done here for a wrapper function to compute the indicator if needed.

status: In CSIndicators v1.1.0

Check issue: https://earth.bsc.es/gitlab/es/csindicators/-/issues/27 (closed)

Next developments: https://earth.bsc.es/gitlab/es/csindicators/-/issues/39

# New SPEI indicator functions

**SPEI**

Step 0: variables (e.g. tasmin, tasmax, prlr) are loaded in the same s2dv_cube

Step 1: units transformation

| | original | transformed |
|---|---|---|
| **Temperature** | K | C |
| **Precipitation** | mm/day | mm/month |

Step 2: checks

- data units
- provided data vs evapotranspiration estimation method

**hcst**, **fcst** *and* **obs** *are treated independently until standardization*

Step 3: estimate evapotranspiration

| | |
|---|---|
| hargreaves | **tasmin**, **tasmax**, *lat* |
| hargreaves modified* | **tasmin, tasmax, prlr**, *lat* |
| thornthwaite* | **tas** (mean), *lat* |

`CST_PeriodPET`

* *need to add estimation of solar radiation and subseasonal case*

Step 4: accumulation

`CST_PeriodAccumulation`

Step 5: standardization

`CST_PeriodStandardization`

**CSIndicators functions**

**MODULE structure in SUNSET**

Note: Simplified slide from author Alba Llabrés

# Bugfix in SelectPeriodOnDates

- **Problem:** the dates were not subset correctly when the order of the dates was first the start date and then the forecast time.
  - Only when different start dates started on different days of the month.
  - **Example subset period**: `start = list(10, 3)` and `end = list(20, 3)`

```
> dim(dates)
sdate ftime
    3    31
```

```
# subset method
dates[res]
```

```
> result[,1]    WRONG!
[1] "2011-03-10" "2013-03-10" "2011-03-11"
```

ftime dim

sdate dim

```
     ...      [,9]         [,10]        [,11]        [,12]      ...      [,20]        [,21]        [,22]
[1,] ... "2011-03-09" "2011-03-10" "2011-03-11" "2011-03-12" ... "2011-03-20" "2011-03-21" "2011-03-22"
[2,] ... "2012-03-08" "2012-03-09" "2012-03-10" "2012-03-11" ... "2012-03-19" "2012-03-20" "2012-03-21"
[3,] ... "2013-03-09" "2013-03-10" "2013-03-11" "2013-03-12" ... "2013-03-20" "2013-03-21" "2013-03-22"
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

status: In branch master

Check issue: https://earth.bsc.es/gitlab/es/csindicators/-/issues/41

# esviz

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# ShapeToMask efficiency improvement

- ★ Added unit test
- ★ Added initial checks
- ★ Added atomic function to loop over each region
- ★ Parallelization with ‘foreach’
  - If ncores parameter is not NULL:

```
registerDoParallel(ncores)
mask <- foreach(shp_i = 1:nrow(shp), .combine = 'cfun', .packages='sf') %dopar%
          .shapetomask(shp = shp, n = shp_i, lon = lon, lat = lat,
                       xy.sfc = xy.sfc, find_min_dist = find_min_dist,
                       shp_col_name_ids = shp_col_name_ids,
                       max_dist = max_dist, region = region, ...)
registerDoSEQ()
```

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

status: In main branch

Check issue: https://earth.bsc.es/gitlab/es/esviz/-/issues/4

# Avoid error when all data are NAs

Functions VizEquiMap(), VizLayout(), VizRobinson(), VizStereoMap() have a more secure way to decide var_limits (the data range, which is needed for generating a proper color bar) so they won't return errors at ColorBar function, which usually says "the var_limits cannot be infinite values."

# SUNSET


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# Loading tas-tos blended data

It is now possible to load tas (near-surface air temperature) and tos (sea-surface temperature) blended together as one dataset in SUNSET by simply requesting ‘tas-tos’ as the variable in your recipe.

The parameter sic_threshold also allows you to choose the threshold for sea ice concentration, to determine the concentration of sea ice at which tos should be replaced by tas. The default value is 0.15.

```
Variables:
  name: tas-tos
  freq: monthly_mean
  units: C
  sic_threshold: 0.2
```

status: in master

# Multi-model pooled method

Multi-model approaches are able to build a hindcast and forecast by combining two or more systems in different ways. The first method available is the ‘pooled’ method: a simple ‘pooling’ of the individual ensemble members of all of the systems. More complex methods will be added in the future.

The recipe now accepts the option multimodel: yes to build a multi-model after loading and post-processing each individual model requested by the user.

```
Datasets:
  System:
    - {name: ECMWF-SEAS5.1}
    - {name: CMCC-SPS3.5}
    - {name: DWD-GCFS2.1}
  Multimodel:
    execute: both # Mandatory: Either both, yes/true or no/false
    approach: pooled # Mandatory: pooled, mean, median
    createFrom: Anomalies # Mandatory: Which module output to create the multimodel from
  Reference:
    - {name: ERA5}
```

status: in branch dev-multimodel2

# Use cases

Two more use case tutorials have been developed. They focus on different modules and on the usage of the SUNSET launcher to do multiple verifications (with and without autosubmit):

- Hands-on 1.2: Computation of Scorecards with Autosubmit
- Hands-on 1.3: Computation of El Niño indices for two seasonal models

If you want, you can try them out and provide feedback in the GitLab issue:
https://earth.bsc.es/gitlab/es/sunset/-/issues/88

status: in branch doc-usecases

# Christmas presentation


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# From last year…

Plot your own Christmas tree by **R base functions** and spice up with an **emoji** twist 🎄😉💗

✨Inspired by
https://github.com/R-CoderDotCom/christmas-tree

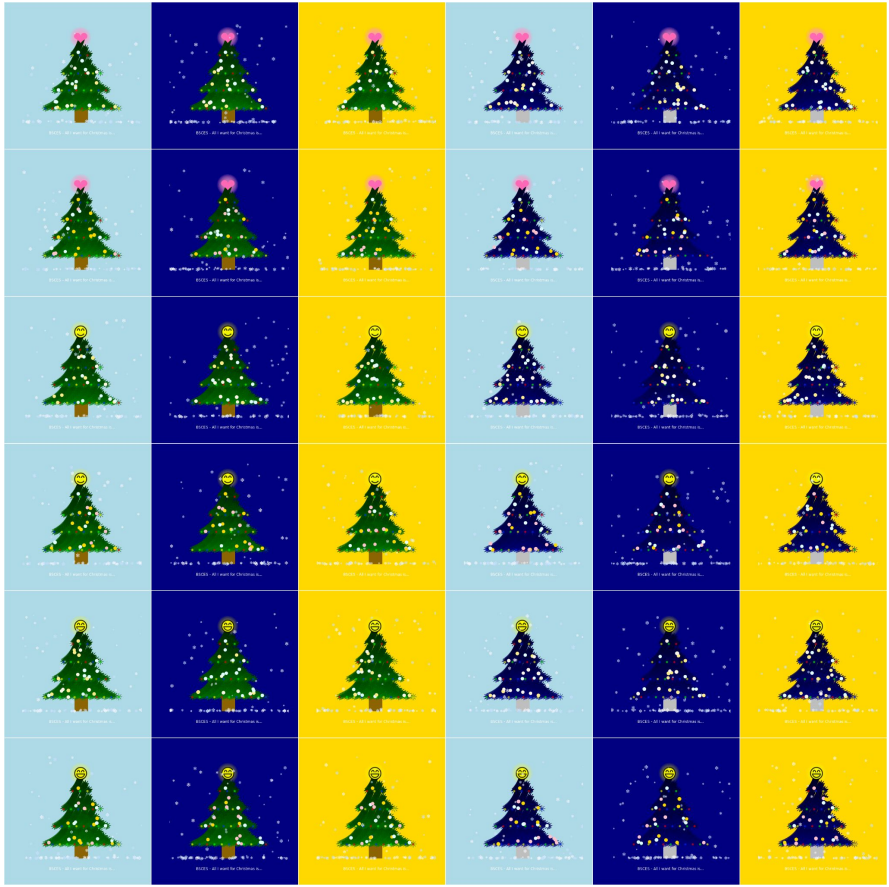
🎭Package “emo”: https://github.com/hadley/emo

🎨Colors reference:
https://www.rapidtables.com/web/color/RGB_Color.html
https://www.datanovia.com/en/blog/awesome-list-of-657-r-color-names/

# This year, we're going to make it bigger!

Mimic SUNSET, we can put multiple options in the recipe this time, and several atomic recipes will be created. Then, submit jobs on Nord3, one job for one atomic recipe. In the end, use montage to combine all the plots together into one Christmas card.

Find everything you need on Rsenal: https://earth.bsc.es/gitlab/es/rsenal/-/tree/main/Christmas/2023



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

# Thanks for joining

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Next meeting: 1st Feb. 12h