Barcelona Supercomputing Center
Centro Nacional de Supercomputación

R user meeting

An-Chi Ho and Eva Rifà

contributor: Martin Lotto

06/10/2022

Agenda

1. Ice-breaker: R tips
2. News
   - General R
   - s2dv
   - startR
   - ClimProjDiags
   - CSTools
3. User presentation: [Martin]
4. Q&A
   -

Ice-breaker

R tips

RStudio shortcut

- (y <- seq(1,10, 2.5)) # print directly
- alt + “-” # <-
- ctrl + shift + c # comment lines
- ctrl + i # indent

Make the screen clean: suppress printing

- suppressPackageStartupMessages(library(CSTools)) # don’t print the package loading info
- suppressWarnings() & suppressMessages()
- option(warn = -1) # any negative value; global option changed
- system(paste0("cdo -s ", sellonlatbox, "remap", method, ",", grid, " ", tmp_file, " ", tmp_file2), ignore.stdout = T, ignore.stderr = T) #ignore system message
- invisible(1:4); invisible(lapply(list('CSTools', 'abind'), library, character.only = TRUE))

R tips

Keep results

- summary <- capture.output(summary(as.vector(data)))
- "logger" package
- sink() #Send R Output to a File

Debug

- browser(); browser(skipCalls = 1) # create break point
- trace(PlotEquiMap, edit = T) # Change the loaded function temporarily

Find an existing function to save your work

https://twitter.com/hadleywickham/status/1571603361350164486?s=20&t=CRUX3gCMyF8sWe61KNasyQ

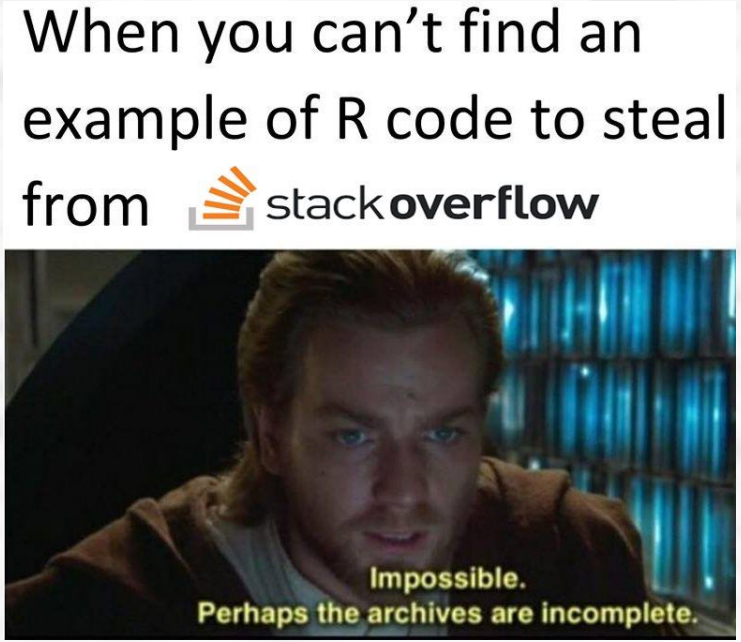
→ **What tips do you have?**

General R

Individual R user meeting

Slack channel “r-tools”

- Welcome to raise all the R-related questions



… Go to r-tools Slack channel.

Instability of RStudio server on Nord3v2

- Random error
- Endless queuing of the job
- Webpage cannot be open
- Can only load a small amount of data

→ Please report the error message (open an issue on request) and provide the scripts to help us solve the problem.

→ Try to use RStudio server on Worksation (should be more stable)
https://earth.bsc.es/wiki/doku.php?id=computing:workstations#using_rstudio-server_in_ws

* Same problem with Jupyter Notebook.

Plotting function house-cleaning

The in-house plotting functions:

- s2dv: **PlotEquiMap**, PlotSection, **PlotStereoMap**, **PlotLayout**, PlotBoxWhisker, PlotClim, PlotAno, PlotACC, Plot2VarsVsLTime, PlotMatrix, PlotVsLTime, AnimateMap
- CSTools: **PlotMostLikelyQuantileMap**, **PlotCombinedMap**, PlotForecastPDF, PlotPDFsOLE, PlotTriangles4Categories

Problems

- PlotLayout doesn't work well with some of the functions. IN THEORY it should be.
- Parameter naming is a mess and not comprehensive.
- Several small bugs

Desired features

- Different projection
- Fixed grid size (keep it as square)
- Find more (about PlotEquiMap) in https://earth.bsc.es/gitlab/es/s2dv/-/issues/63

Plan to start in the late October. Will be a major version change (compatibility break.)

License change

We need to change all the packages’ license to GPL-3. We will do it in the next release of each package.

s2dv

PlotLayout

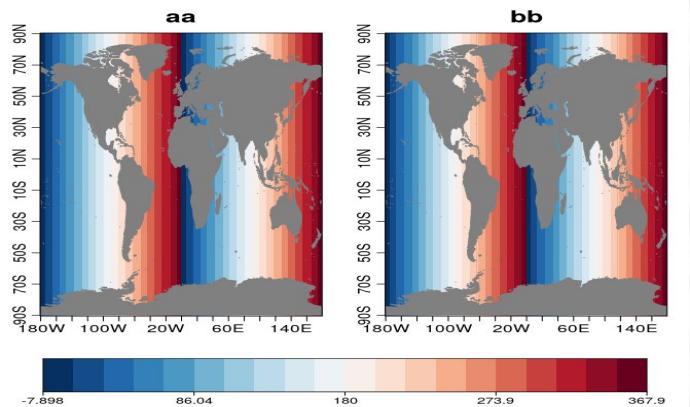
- New parameter *subplot_titles_scale*: Scale factor for the subplots top titles.

Status: in master

- **Before:**

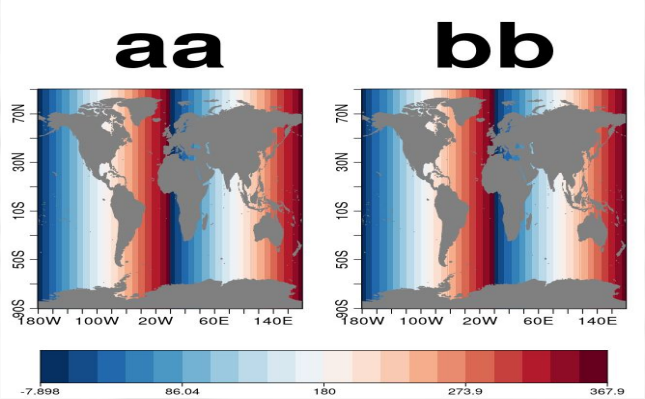
```
asd <- list(list(filled.continents = F, title_scale = 0.5),
            list(filled.continents = T, title_scale = 1))

PlotLayout(PlotEquiMap, c('lat', 'lon'), var = asplit(data, 1),
           lon = lon, lat = lat, special_args = asd,
           titles = c('aa', 'bb'))
```



- **Now:**

```
PlotLayout(PlotEquiMap, c('lat', 'lon'), var = asplit(data, 1),
           lon = lon, lat = lat, subplot_titles_scale = 4,
           titles = c('aa', 'bb'))
```

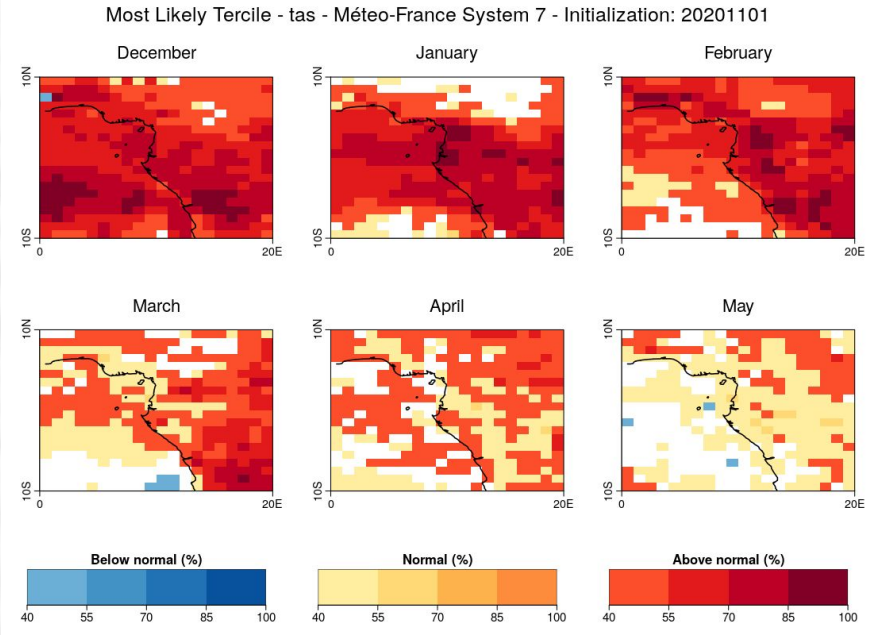


PlotLayout: work with PlotMostLikelyQuantileMap

In theory, PlotLayout() should work with all the in-house functions. But it doesn't now.

Status: Fixed but need more checks; in [develop-PlotLayout_PlotMostLikelyQuantileMap]

Follow the development on https://earth.bsc.es/gitlab/es/s2dv/-/issues/74



(Created by Victòria)

PlotEquiMap

- Allow **dots**, **varv**, **varu** and **contours** parameters coordinates dimensions to be unordered.

PlotEquiMap(var, lon, lat, varu, varv, dots, contours)

Status: in branch develop-PlotEquiMap_dots

- **Accepted**

var dims	dots, varv, varu, contour dims
(lon = 2, lat = 3)	(lon = 2, lat = 3)
(lat = 3, lon = 2)	(lon = 2, lat = 3)
(lat = 2, lon = 2)	(lat = 2, lon = 2)
(2, 3)	(2, 3)
(3, 2)	(3, 2)
(2 , 2)	(2, 2)

* ex. length(lon) = 2, length(lat) = 3

- **Not accepted:**
  - Different coordinates dimensions as lon and lat
  - Different dimension names as var

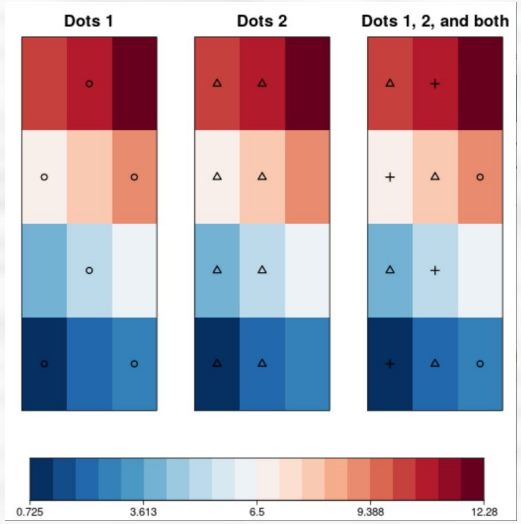
PlotEquiMap: plot multiple dots

This functionality already exists!

Definition of `dots`:

Array of same dimensions as 'var' or with dimensions c(n, dim(var)), where n is the number of dot/symbol layers to add to the plot. A value of TRUE at a grid cell will draw a dot/symbol on the corresponding square of the plot. By default all layers provided in 'dots' are plotted with dots, but a symbol can be specified for each of the layers via the parameter 'dot_symbol'.

Discussion on gitlab:
https://earth.bsc.es/gitlab/es/s2dv/-/issues/76



Corr output

Status: to be started

- **To discuss:**

  Corr(): change “conf.lev” to “alpha”, and return “sign”, “pval”, (and “conf”)

  - **Now:**

    a) conf: ACC(), Corr(), RMS(), Regression(), Spread(), Trend()

    b) alpha: DiffCorr(), ResidualCorr()

- **Future development:** dat_dim = NULL by default

  - **Now:**

    a) dat_dim = ‘dataset’: ACC(), Corr(), RatioSDRMS(), RMS(), RMSSS() and UltimateBrier().

    b) dat_dim = NULL: BrierScore(), RPS(), RPSS(), CRPS() and CRPSS().

    c) dat_dim = c(‘dataset’, ‘member’): Clim(), Ano_CrossValid() → REMAIN THIS DEFAULT

DiffCorr significance test

Status: Partially done in master but need further discussion

Follow the discussion: https://earth.bsc.es/gitlab/es/s2dv/-/merge_requests/110

new parameter test.type:

A character string indicating the type of significance test. It can be "**two-sided**" (to assess whether the skill of "exp" and "ref" are significantly different with a z-test on Fisher z-transformed correlation coefficients, Hinkel et al, 1988) or "**one-sided**" (to assess whether the skill of "exp" is significantly higher than that of "ref" with the test described in Steiger, 1980). The default value is "two-sided".

startR

extra_queue_params

In the documentation, `extra_queue_params = list('#SBATCH --mem-per-cpu=3000')` was used a lot but this line doesn't work in Nord3v2. Practical_guide.md & use cases are corrected.

Definition: list of character strings with additional queue headers for the jobs to be submitted to the HPC.

Check nord3v2 manual: https://www.bsc.es/support/Nord3v2-ug.pdf

```
res <- Compute(wf, chunks = list(latitude = 2),
               threads_load = 2, threads_compute = 4,
               cluster = list(
                 queue_host = 'nord4',
                 queue_type = 'slurm',
                 temp_dir = '/gpfs/scratch/bsc32/bsc32734/startR_hpc/',
                 cores_per_job = 4, max_jobs = 4, job_wallclock = '00:10:00',
                 extra_queue_params = list('#SBATCH --constraint=medmem'),
                 bidirectional = FALSE, polling_period = 10),
               ecflow_suite_dir = '/home/Earth/aho/startR_local/',
               wait = TRUE)
```

Find the error message from job submission

If the job fails during running, it turns red on ecFlow_ui. But If the job is not submitted successfully at very first, ecFlow_ui stays blue (queue) forever.

- ssh to Nord3v2
- Go to `temp_dir` (*e.g., /gpfs/scratch/bsc32/bsc32734/startR_hpc/*)
- Go to the innest level (*e.g., STARTR_CHUNKING_7398543523/STARTR_CHUNKING_7398543523/computation/longitude_CHUNK_1/latitude_CHUNK_1/sdate_CHUNK_1/var_CHUNK_1/dat_CHUNK_1/*)

```
[bsc32734@login4 dat_CHUNK_1]$ ls
Chunk.1  Chunk.job1  Chunk.submit_time
```

- vi Chunk.1

```
sbatch: error: You cannot submit a job requesting memory
parameters, memory is automatically set for each asked cpu
(2G/core by default, 4G/core for medmem, 8G/core for highmem)
sbatch: error: Batch job submission failed: Memory required
by task is not available
```

Start() bugfixes

Status: in branch develop-error_longitude_dims

- **Error when the whole range of longitudes was requested:**

```
lons.min <- 0
lons.max <- 359
circular.sort <- CircularSort(0, 360)
Start([...], transform_params = list(grid = 'r360x180', method = 'bilinear'), [...])
```

- **Output:**

```
* Progress: 0%Error in easyNCDF::ArrayToNc(setNames(list(data_array, lons, lats), c("var",
: The dimension 'longitude' is defined or used more than once in the provided data but the
dimension specifications do not match.
```

ClimProjDiags

WeightCells

Used to compute the square-root of the cosine of the latitude weighting on the given array.

Status: in branch develop-WeightCells

- **To discuss:**
  - Function name
  - Weighting method:

a) square root b) cos(lat) c) areacello

  - Potential development

https://earth.bsc.es/gitlab/es/ClimProjDiags/-/issues/10

```
WeightCells <- function(data, lat, lat_dim = 'lat', ncores =
NULL) {

  [...]

  wt <- sqrt(cos(lat * pi/180))

  res <- Apply(data = data,
               target_dims = c(lat_dim),
               fun = .WeightCells,
               wt = wt,
               ncores = ncores)$output1
  [...]

  return(res)
}

.WeightCells <- function(data, wt) {
  data <- wt*data
  return(data)
}
```

CSTools

s2dv_cube object

- **Current structure**

```
$ data          : int [1:2, 1:10, 1:5] 1 2 3 …
$ lon           : NULL
$ lat           : NULL
$ Variable      : NULL
$ Datasets      : NULL
$ Dates         : NULL
$ when          : NULL
$ source_files: NULL
- attr(*, "class")= chr "s2dv_cube"
```

- **New structure**

```
$data: an array
$dims
$coords
    #(names are changable, according to names(dim(data))
    $sdate:
    $time: _____
    $lon:
$attrs
    $Variable
    $Datasets
    $source_files
    $when
    $Dates: _____  (x)
    $load_parameters: (x)
```

- **Main changes:**

i) The coordinates level

ii) The names of the coordinates

iii) More similar to xarray

iv) $Dates and $load_parameters deprecated

Status: to be started

raster / Malaria vignette

https://earth.bsc.es/gitlab/mlotto/templates_to_share/-/tree/main/spatial_stuff

Q & A

Thanks for joining

Next meeting: 3rd Nov. 2022 (4 pm)