

Self-Study Exercises

Nishant Gopalakrishnan

Fred Hutchinson Cancer Research Center

17-18 February, 2011

Introduction to NETCDF library
Variables, Dimensions, Attributes.
R packages available.
Advantages of netcdf4

Exercise 1

The goal here is to explore the functions available in R to read text files in an efficient manner reading in smaller chunks at a time.

Question 1

- Make use of the functions `read.table` and `scan` to read in only the row 601 from the file.
- Make use of the `system.time` function to check which function is more efficient.

Solution:

```
> origDat <- system.time(as.numeric(scan("snpData.csv", what = character(0), sep = ",",  
+                                     skip = 600, nlines = 1, quiet = TRUE)))
```

Exercise 2

The experimental data has been provided to you in the form of a NetCDF (version 3) file in the `inst/extdata` folder of your package. Make use of the `ncdump` commandline utility to explore the contents of the NetCDF file. Do not forget to use the `-h` option.

Question 2

Create a function `getCol` to retrieve columns in the range specified by the user from the NetCDF file provided. The function should take two arguments

- file name

- A numeric vector of length 2, indicating the range of columns to be retrieved from the NetCDF file.

Solution:

```
> getCol <- function(file, rng)
+ {
+   rng <- sort(rng)
+   if (length(rng) != 2)
+     stop("rng has to be a vector of length 2")
+   nc <- open.ncdf(file, write = FALSE)
+   get.var.ncdf(nc, varid = "snpData", start= c(1,rng[1]),
+     count = c(1000,rng[2]))
+   close.ncdf(nc)
+ }
```

Exercise 3

The goal here is to read in all the snips for sample 601 from the netcdf file provided and create a small unit test function that checks if the values read from the ncdf file are the same with respect to the raw text files provided.

Question 3

Create a function `checkRow` that

- Reads in all the snps for a sample id provided by the user
- Make use of the `checkEquals` to ensure the values read in from the NetCDF file are correct with respect to the raw snip data obtained from the file "snpData.csv".

Solution:

```
> nc <- open.ncdf(file, write = FALSE)
> dat <- get.var.ncdf(nc, varid = "snpData", start= c(601,1),
+   count = c(1,NCOLS))
> origDat <- as.numeric(scan("snpData.csv", what = character(0), sep = ",",
+   skip = 600, nlines = 1, quiet = TRUE))
> checkEquals(as.numeric(dat), origDat)
> dat <- get.var.ncdf(nc, varid = "snpData", start= c(1000,1),
+   count = c(1,NCOLS))
> origDat <- as.numeric(scan("snpData.csv", what = character(0), sep = ",",
+   skip = 999, nlines = 1, quiet = TRUE))
> checkEquals(as.numeric(dat), origDat)
> close.ncdf(nc)
```