

**Εργαστήριο Ανώτερης Γεωδαισίας  
Μεταπτυχιακό Πρόγραμμα ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ  
«Αναλυτικές Μέθοδοι στη Γεωπληροφορική»  
(Ακαδ. Έτος 2022-23)**

ΟΝΟΜΑΤΕΠΩΝΥΜΟ .....

ΕΞΑΜΗΝΟ .....

Ημερομηνία Παράδοσης : **10/1/2023**

**ΘΕΜΑΤΙΚΗ ΕΡΓΑΣΙΑ #4**

**Σκοπός:** Η παρούσα Θεματική Εργασία αποσκοπεί στην εξοικείωση σας με την εισαγωγή στο λογισμικό R αρχείων δεδομένων της μορφής netCDF που συνήθως χρησιμοποιούνται για την ανταλλαγή δεδομένων γεωεπιστημονικού ενδιαφέροντος.

**(1)** Για αυτό το μέρος της Θεματικής Εργασίας θα χρειαστεί να φέρετε για ανάλυση στο R ένα αρχείο δεδομένων θερμοκρασίας του αέρα (σε υψόμετρο 2 m) σε παγκόσμια κάλυψη από τον ιστότοπο <https://crudata.uea.ac.uk/cru/data/hrg/> της Μονάδας Κλιματικών Ερευνών (Climatic Research Unit, CRU) του Πανεπιστημίου East Anglia. Αποκτήστε πρόσβαση στα δεδομένα ακολουθώντας το σύνδεσμο **CRU TS v. 4.06 → Local copy** ([https://crudata.uea.ac.uk/cru/data/hrg/cru\\_ts\\_4.06/](https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.06/)) και στη συνέχεια ακολουθώντας τους συνδέσμους [cruts.2205201912.v4.06](https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.06/cruts.2205201912.v4.06) ([https://crudata.uea.ac.uk/cru/data/hrg/cru\\_ts\\_4.06/cruts.2205201912.v4.06](https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.06/cruts.2205201912.v4.06)) → **tmp** ([https://crudata.uea.ac.uk/cru/data/hrg/cru\\_ts\\_4.06/cruts.2205201912.v4.06/tmp/](https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.06/cruts.2205201912.v4.06/tmp/)).

Σημειώστε ότι, στον ίδιο ιστότοπο, υπάρχουν παρόμοια σύνολα δεδομένων για άλλες κλιματικές παραμέτρους όπως, ημερήσιο εύρος θερμοκρασιών (*diurnal temperature range, **dtr***), βροχόπτωση (*precipitation, **pre***), συχνότητα υγρών ημερών (*wet - day frequency, **wet***), πίεση ατμών (*vapour pressure, **vap***) και κάλυψη νεφών (*cloud cover, **cld***), κ.ά.

Από τα εμφανιζόμενα διαθέσιμα αρχεία, για τους σκοπούς της Θεματικής Εργασίας, ενδιαφέρει να επιλέξετε ένα από τα αρχεία **tmp** με επέκταση **.tmp.dat.nc.gz** (προσοχή, όχι τα αρχεία **.tmp.dat.gz** ή **.tmp.stn.gz**).

Τα εν λόγω αρχεία ενδιαφέροντος είναι στο μορφότυπο **NetCDF** (ή ακριβέστερα σε μορφή συμπιεσμένων αρχείων **NetCDF**) και περιέχουν δεδομένα μέσης μηνιαίας θερμοκρασίας σε σημεία κανάβου 0.5° x 0.5° , τα οποία είναι οργανωμένα ανά δεκαετίες, από το 1901 έως και το 2022. Παρατηρήστε ότι το αρχείο [cru\\_ts4.05.1901.2021.tmp.dat.nc.gz](https://crudata.uea.ac.uk/cru/data/hrg/cru_ts4.05.1901.2021.tmp.dat.nc.gz) περιλαμβάνει δεδομένα από ολόκληρο τον 20<sup>ο</sup> αιώνα έως και το 2021 και για το λόγο αυτό έχει περίπου 10πλάσιο όγκο σε σχέση με τα αντίστοιχα αρχεία δεκαετίας.

Για τους σκοπούς της Θεματικής Εργασίας επιλέξτε είτε το αρχείο της περιόδου 1901-2021 (εάν επιθυμείτε να δοκιμάσετε τις δυνατότητες του R για Μεγάλα Δεδομένα), είτε ένα από τα αρχεία **.tmp.dat.nc.gz** για μια δεκαετία της δικής σας προτίμησης, π.χ. για την 1η δεκαετία του 21<sup>ου</sup> αι., το αντίστοιχο αρχείο θα είναι το **cru\_ts4.06.2001.2010.tmp.dat.nc.gz**.

1. Καθορίστε στον Η/Υ σας ένα νέο χώρο εργασίας του R, π.χ. ένα νέο (άδειο) ντοσιέ με την ονομασία **askisi4**, κάνοντας χρήση της συνάρτησης **setwd(...path/askisi4)**. Δημιουργήστε εκεί ένα κενό αρχείο κειμένου, π.χ. με την ονομασία **.Rhistory**. Από την κονσόλα του R (ή του RStudio) ανοίξτε το εν λόγω αρχείο (→ **Load history** → **.Rhistory**). Επιθεωρήστε το περιεχόμενο του χώρου εργασίας με την εντολή **ls()**, προκειμένου να βεβαιωθείτε ότι ο χώρος εργασίας είναι όντως άδειος (η εντολή **ls()** θα πρέπει να σας δείξει την απάντηση **character(0)**)

Από την κονσόλα του R (ή του RStudio) αποθηκεύστε το άδειο **.Rhistory** αρχείο της συνεδρίας σας με την ονομασία **askisi4\_1.Rhistory** (→ **Save history** → **askisi4\_1.Rhistory**). Με τον τρόπο αυτό θα έχετε διασφαλίσει ότι με την προσθήκη νέων εντολών στη συνέχεια της συνεδρίας σας θα καταγραφούν τελικά στο «φρέσκο» **askisi4\_1.Rhistory** αρχείο που μόλις δημιουργήσατε. Απλά στο τέλος της συνεδρίας σας, πριν την έξοδό σας από το R, από την κονσόλα του R αρκεί εκτελέσετε πάλι το βήμα → **Save history** → **askisi4\_1.Rhistory**.

2. Σημειώστε ότι για τους σκοπούς της Θεματικής Εργασίας θα χρειαστεί να έχετε εγκαταστήσει και φορτώσει στο χώρο εργασίας σας του R τα πακέτα **R.utils**, **ncdf4**, **sf**, **raster**, **RColorBrewer**, **lattice**, **rasterVis**, **chron**, ... Φροντίστε να τα έχετε εγκαταστήσει και φορτώσει στο χώρο εργασίας του R πριν προχωρήσετε στα επόμενα ερωτήματα της Θεματικής Εργασίας.
3. Χρησιμοποιώντας **μέσα από το R** τις κατάλληλες εντολές, με τη χρήση των συναρτήσεων **download.file()**, **gunzip()** και **list.files()**, κατεβάστε και αποσυμπίεστε στο χώρο εργασίας του R το αντίστοιχο NetCDF αρχείο **...tmp.dat.nc** που έχετε επιλέξει και το οποίο θα χρειαστείτε για τα επόμενα ερωτήματα. Βεβαιωθείτε ότι κατά τη χρήση της συνάρτησης αποσυμπίεσης **gunzip()** θα χρησιμοποιήσετε τις κατάλληλες παραμέτρους εισόδου/εξόδου ώστε το αρχικό συμπιεσμένο αρχείο **...tmp.dat.nc.gz** να διατηρηθεί στο χώρο εργασίας σας (και όχι να διαγραφεί μετά την αποσυμπίεση). Στο αποσυμπιεσμένο αρχείο μπορείτε να δώσετε μια ονομασία της δικής σας επιλογής (π.χ., **cru\_year1.year2**), αλλά αυτό θα πρέπει να έχει απαραίτητα την επέκταση (extension) **.nc** (δηλ., **cru\_year1.year2.nc**). Εκτυπώστε μια λίστα των αρχείων στο χώρο εργασίας του R, προκειμένου να βεβαιωθείτε ότι το συμπιεσμένο και το αποσυμπιεσμένο NetCDF αρχείο βρίσκονται εκεί.

Στο R διατίθενται τα πακέτα **ncdf4** και **raster** τα οποία παρέχουν την υποστήριξη που απαιτείται για την ανάγνωση και τη σύνταξη αρχείων NetCDF. Το πακέτο **ncdf4** υποστηρίζει αρχεία NetCDF έκδοσης 4 και παλαιότερες. Με αυτό είναι εύκολο να δημιουργήσετε νέες διαστάσεις, μεταβλητές και αρχεία netcdf σε μορφή 3 ή 4 και να επεξεργαστείτε υπάρχοντα αρχεία NetCDF.

Αντίστοιχα, το πακέτο **raster** έχει επίσης τη δυνατότητα ανάγνωσης και γραφής αρχείων NetCDF. Υπάρχουν κάποια ζητήματα που χρήζουν ιδιαίτερης προσοχής, με τη χρήση του πακέτου **raster**, καθώς σε διαδικασίες μετατροπής δεδομένων από τη μορφή NetCDF στη μορφή **raster**, μπορούν να προκύψουν διάφορες διαχειριστικές δυσκολίες όπως π.χ. από το γεγονός ότι οι συντεταγμένες NetCDF αναφέρονται στο κέντρο των κελιών ενός πλέγματος σημείων, ενώ σε **raster** δεδομένα οι συντεταγμένες αναφέρονται στις γωνίες των κελιών.

Αν και δεν είναι απαραίτητο για τους σκοπούς της Θεματικής Εργασίας, για την ενημέρωσή σας, πριν προχωρήσετε στα επόμενα ερωτήματα, μεταβείτε στον ιστοχώρο <http://netcdf4excel.github.io/> όπου μπορείτε να βρείτε μια χρήσιμη εφαρμογή (**NetCDF4Excel\_3\_3\_setup**) στο Excel που επιτρέπει να διαβαστούν πολύ γρήγορα NetCDF αρχεία προκειμένου να πάρει κανείς μια άμεση εικόνα του περιεχομένου τους. Χρησιμοποιήστε την εφαρμογή, από το Excel, με το αποσυμπιεσμένο netCDF αρχείο **.tmp.dat.nc** που σας ενδιαφέρει, ώστε να πάρετε μια πρώτη εικόνα των περιεχομένων του συγκεκριμένου αρχείου.

Διαβάστε/ανοίξτε το αποσυμπιεσμένο αρχείο NetCDF που έχετε κατεβάσει στο χώρο εργασίας σας, χρησιμοποιώντας τη συνάρτηση **nc\_open()** του **ncdf4** και εκχωρήστε τα δεδομένα του σε ένα αντικείμενο του R με την ονομασία **ncfdata**. Εκτυπώστε στην οθόνη σας ορισμένες βασικές πληροφορίες για τα δεδομένα του αρχείου, οι οποίες μπορούν να ληφθούν χρησιμοποιώντας τη συνάρτηση **print()** ή τη συνάρτηση **str()**.

```
{  
  # Save the print(nc) dump to a text file `*_metadata.txt`  
  sink(' *_metadata.txt')  
  print(ncfdata_file)  
  sink()  
}
```

Εναλλακτικά, χρησιμοποιώντας την παραπάνω αναφερόμενη απλή συνάρτηση που κάνει χρήση της συνάρτησης **sink()**, οδηγήστε τα αποτελέσματα της εκτύπωσης των πληροφοριών για το περιεχόμενο του αποσυμπίεσμένου NetCDF αρχείου σε ένα αρχείο κειμένου με την ονομασία **'ncfdata.txt'**.

Στο χώρο εργασίας σας θα πρέπει να έχει δημιουργηθεί το αρχείο **ncfdata.txt** με όλες τις προηγούμενες βασικές πληροφορίες για τα δεδομένα του NetCDF αρχείου. Ανοίξτε το και επιθεωρήστε το περιεχόμενό του προκειμένου να κατανοήσετε ακριβώς ποιο ήταν το αποτέλεσμα των προηγούμενων εντολών.

4. Επιθεωρήστε τα NetCDF δεδομένα του αρχείου προκειμένου να διαπιστώσετε πως μεταβάλλονται οι μεταβλητές **lon** και **lat** των συντεταγμένων γεωγραφικού μήκους και πλάτους (**longitude** και **latitude**), χρησιμοποιώντας τη συνάρτηση **ncvar\_get()** του **ncdf4**, και εκτυπώστε τις τιμές για το πλήθος τους (δηλ. τις διαστάσεις των διανυσμάτων τιμών **lon** και **lat**). Ακολουθώντας εκτυπώστε τις πρώτες και τις τελευταίες μερικές τιμές τους χρησιμοποιώντας τις συναρτήσεις **head()** και **tail()**. Επιβεβαιώστε τον αριθμό των διαθέσιμων τιμών γεωγραφικού μήκους και γεωγραφικού πλάτους χρησιμοποιώντας τη συνάρτηση **dim()** ή **length()**.
5. Διαβάστε τη μεταβλητή του χρόνου, δηλ. εκχωρήστε τα δεδομένα χρόνου σε μια μεταβλητή **time** και εξακριβώστε τις ιδιότητες των διαστάσεων των μεταβλητών (*τα attributes*) -εν προκειμένω της μεταβλητής του χρόνου- χρησιμοποιώντας τις συναρτήσεις **ncvar\_get()** και **ncatt\_get()**, καθώς και το πλήθος των χρονικών τιμών χρησιμοποιώντας τη συνάρτηση **dim()**. Σημειώστε, ότι ο χρόνος εδώ εκφράζεται στη μορφή "χρόνος από ..." (π.χ. *days since ...*) και θα μετατραπεί παρακάτω σε μια πιο αναγνωρίσιμη μορφή. Σημειώστε ότι τα χαρακτηριστικά των μεταβλητών που αφορούν στα δεδομένα έχουν δύο συνιστώσες, **hasatt** (μια λογική μεταβλητή), και **value** π.χ., τον πραγματικό "χρόνο", που προκύπτει από την συμβολοσειρά.
6. Δεδομένου ότι μεταβλητή του χρόνου, δίνεται σε μονάδες "από τη χρονική στιγμή ...", είναι απαραίτητο να μπορεί να μετατραπούν οι τιμές που διαβάστηκαν σε "πραγματικές" (ή πιο γνώριμες) τιμές χρόνου διαχωρίζοντας τη χρονική συμβολοσειρά χαρακτηριστών **tunits\$value** στα συστατικά μέρη της, με τη συνάρτηση **strsplit()** και τη συνάρτηση **unlist()** - π.χ., μετατρέποντας "days since 1900-1-1" → στα επιμέρους 3 συστατικά της, **1900 1 1** (δηλ. σε έτος, μήνα, ημέρα). Στη συνέχεια χρησιμοποιώντας τη συνάρτηση **chron()** να καθορίσετε τις πραγματικές ημερομηνίες, σε ένα αντικείμενο με την ονομασία **nowdates** π.χ. στη μορφή 01/16/19 -Μήνας / Ημέρα / Δύο τελευταία ψηφία για το Έτος, οι οποίες ανάγονται κάθε φορά από την αφετηρία μέτρησης του χρόνου.

Εναλλακτικά για μια πιο εύκολη μετατροπή χρησιμοποιήστε τη συνάρτηση **as.Date()** σε μια εντολή της μορφής **as.Date(x, origin=...)**.

Παρατηρήστε τη διαφορετική μορφοποίηση των ημερομηνιών που δίνει η συνάρτηση **as.Date()**.

7. Εκχωρήστε σε μια συστοιχία δεδομένων, με την ονομασία **tmp.array**, τα δεδομένα των θερμοκρασιών (δηλ. τις τιμές της μεταβλητής **'tmp'**, όπου προσοχή **'tmp'** εδώ υποδηλώνει **temperature**, και όχι τη συντομογραφία για τον όρο **'temporary'**), καθώς επίσης τα χαρακτηριστικά τους (attributes) **long\_name**, **units** και **\_FillValue**.

Επιβεβαιώστε, χρησιμοποιώντας τη συνάρτηση **dim()**, τις διαστάσεις της εν λόγω συστοιχίας **tmp.array** που περιέχει τις τιμές για τα γεωγραφικά μήκη και πλάτη, και τις χρονικές εποχές που διαβάστηκαν. Βεβαιωθείτε ότι διαστάσεις είναι συμβατές με τις διαστάσεις των αντικειμένων (μεταβλητών) **latitude**, **longitude** και **time** από τα προηγούμενα ερωτήματα 4 και 5.

Στη συνέχεια, διαβάστε και εκτυπώστε μερικά από τα "συνολικά/καθολικά χαρακτηριστικά" (ή τα μεταδεδομένα) του αρχείου των δεδομένων (**global attributes**) του netCDF αρχείου, όπως τα: **title**, **institution**, **datasource**, **references**, **history** και **Conventions**.

8. Εκτυπώστε ποια είναι η τιμή **FillValue** που χρησιμοποιείται στα τυχόν κελιά όπου δεν υπάρχουν δεδομένα θερμοκρασίας. Εάν χρησιμοποιήσετε τη συνάρτηση **head()** επί της συστοιχίας **tmp.array** των δεδομένων, θα διαπιστώσετε ότι υπάρχουν κενά κελιά τα οποία υποδηλώνονται με τον συμβολισμό NA που καταλαβαίνει το R, ώστε να αγνοεί τα εν λόγω κελιά κατά την επεξεργασία των δεδομένων της συστοιχίας. Χρησιμοποιώντας κατάλληλα τη συνάρτηση **na.omit()** ελέγξτε και εκτυπώστε πόσα κελιά της συστοιχίας **tmp.array** περιέχουν έγκυρες μετρήσεις θερμοκρασίας (α) στο πρώτο επίπεδο (layer) των δεδομένων, και (β) συγκεντρωτικά σε όλα τα επίπεδα (layers) των δεδομένων.
9. Εξάγετε το τμήμα δεδομένων της συστοιχίας **tmp.array** που αντιστοιχούν στον 3<sup>ο</sup> μήνα του 4<sup>ου</sup> έτους της δεκαετίας στην οποία έχετε εστιάσει την ανάλυσή σας. Αφού προηγουμένως φορτώσετε τη βιβλιοθήκη **RColorBrewer** που χρειάζεται για την επιλογή κατάλληλης παλέτας χρωμάτων, εκχωρήστε τις αντίστοιχες τιμές σε ένα αντικείμενο με την ονομασία **tmp.month.layer** και εκτυπώστε τον αντίστοιχο χάρτη που απεικονίζει τις εν λόγω τιμές, χρησιμοποιώντας τη συνάρτηση **image()** του R. Στις παραμέτρους εισόδου της συνάρτησης **image()** χρησιμοποιήστε την παράμετρο **col=rev(brewer.pal(10,"RdBu"))** για την επιλογή παλέτας γήινων χρωμάτων, και καθορίστε ως επικεφαλίδα του γραφήματος την ετικέτα **"raster image of a data layer"**.
10. Προκειμένου να δημιουργήσετε ένα χάρτη καλύτερης αισθητικής ή/και χαρτογραφικής πληρότητας χρησιμοποιήστε τη συνάρτηση **levelplot()** από το πακέτο **lattice**. Η συνάρτηση **expand.grid()** χρησιμοποιείται για τη δημιουργία ενός συνόλου από 720 x 360 ζεύγη τιμών γεωγραφικού μήκους και πλάτους (με τις αποχρώσεις των χρωμάτων κατά τα γεωγραφικά πλάτη να κυμαίνονται πιο γρήγορα), ένα για κάθε στοιχείο του πίνακα τιμών του αντικειμένου **tmp.month.layer**. Η συνάρτηση **levelplot()** απαιτεί ως παράμετρο εισόδου τον αριθμό **cut** των χρωματικών αλλαγών που εκφράζουν τα επίπεδα διακυμάνσεων στο εύρος (εν προκειμένω) των θερμοκρασιών και, τις τιμές **at** των σημείων αλλαγής των εν λόγω κατηγοριών θερμοκρασίας. Εν προκειμένω, συνιστάται να χρησιμοποιήσετε 11 επίπεδα διακυμάνσεων (δηλ., **cut = 11**) και ως αντίστοιχα σημεία χρωματικής αλλαγής των θερμοκρασιών τις τιμές (-50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 50).

Μέχρι σε αυτό το στάδιο της Θεματικής Εργασίας εξετάστηκαν τρόποι διαχείρισης των μεταβλητών του NetCDF αρχείου ενδιαφέροντος ως μονοδιάστατα διανύσματα (π.χ. γεωγραφικά μήκη ή πλάτη) και δισδιάστατοι πίνακες («φέτες (slices)/επίπεδα» δεδομένων ή εικόνες raster). Επιπρόσθετα, τα δεδομένα μπορείτε να τα διαχειριστείτε ως πολυδιάστατους πίνακες ή συστοιχίες/arrays («τούβλα/bricks» εικόνων raster). Σε τέτοιες δομές δεδομένων, όπως στο παράδειγμα του συγκεκριμένου συνόλου δεδομένων θερμοκρασίας που διερευνάτε εδώ, οι τιμές συντεταγμένων για κάθε σημείο πλέγματος είναι, γεωγραφικά μήκη, γεωγραφικά πλάτη και χρονικές εποχές.

Αυτός ο τύπος δομής ενός συνόλου δεδομένων μπορεί να απεικονιστεί αποτελεσματικά χρησιμοποιώντας το πακέτο **raster** του R, το οποίο περιέχει ένα σημαντικό αριθμό συναρτήσεων για τη διαχείριση και επεξεργασία, όπως υποδηλώνει το όνομά του, συνόλων δεδομένων στη μορφή εικόνων raster και πολλές άλλες γεωχωρικές μορφές, συμπεριλαμβανομένων αρχείων δεδομένων σε μορφότυπους NetCDF.

11. Από την πολυδιάστατη μορφή των δεδομένων θερμοκρασίας του NetCDF αρχείου σας, εκχωρήστε το σύνολο των δεδομένων θερμοκρασίας του αντικειμένου **ncfdata** που έχετε ήδη δημιουργήσει στο ερώτημα (4) σε μια μεταβλητή με την ονομασία **tmp\_raster**, χρησιμοποιώντας τη συνάρτηση **brick()** του πακέτου **raster**, η οποία λειτουργεί στο αρχείο που περιέχει τα δεδομένα.

Εκτυπώστε μια σύντομη περίληψη των περιεχομένων του αντικειμένου **tmp\_raster** και των επιμέρους μεμονωμένων χαρακτηριστικών των δεδομένων του (π.χ., τις διαστάσεις του), καθώς επίσης εξακριβώστε την κλάση του, προκειμένου να βεβαιωθείτε ότι αυτό έχει αποθηκευθεί ως εικόνα raster και όχι ως συστοιχία (array).

Απεικονίστε τα **δεδομένα του 1<sup>ου</sup> επιπέδου** του τούβλου των δεδομένων **tmp\_raster**, χρησιμοποιώντας την έκδοση της συνάρτησης **levelplot()** του πακέτου **rasterVis**, η οποία είναι παρόμοια με την αντίστοιχη συνάρτηση του πακέτου **lattice**. Επιπλέον, θα χρειαστείτε να

χρησιμοποιήστε ως όρισμα στην κλήση της `levelplot()` τη συνάρτηση `subset()` του πακέτου `raster`, στη μορφή `subset(brick.dataset.object, specific layer)`, η οποία εξάγει ένα μόνο στρώμα από το τούβλο `raster` που ενδιαφέρει. Η συνάρτηση `levelplot()` απαιτεί τον καθορισμό του ορίσματος `par.settings = mapTheme`, όπου η μεταβλητή μπορεί να καθορίζει κάποια κατάλληλη παλέτα χρωμάτων. Μπορείτε να πειραματιστείτε με 2-3 από τις ακόλουθες επιλογές:

```
# RdBuTheme()
# rasterTheme()
# YlOrRdTheme(region = brewer.pal(9, 'YlOrRd'), ...)
# RdBuTheme(region = brewer.pal(9, 'RdBu'), ...)
# BuRdTheme(region = rev(brewer.pal(9, 'RdBu')), ...)
# PuOrTheme(region = brewer.pal(9, 'PuOr'), ...)
# GrTheme(region = rev(brewer.pal(9, 'Greys')), ...)
# BTCTheme(region = BTC(n=9), ...)
```

Χρησιμοποιήστε ως τελική επιλογή χρωματικής παλέτας την παραμετροποίηση

```
# rasterTheme(region = rev(brewer.pal(10, "RdBu")))
```

που καθορίζει μια από τις λεγόμενες αποκλίνουσες (diverging) παλέτες του πακέτου `RColorBrewer`.

Δοκιμάστε μια απλούστερη κλήση της συνάρτησης `levelplot()` στη μορφή `levelplot(subset(...))`. Παρατηρήστε ότι το διαφορετικό γράφημα που δημιουργείται είναι της μορφής πλέγματος (lattice graph).

12. Στο ντοσιέ `myrdata` στις ιστοσελίδες του μαθήματος υπάρχει το συμπιεσμένο αρχείο `ne_110m_admin_0_countries.zip` (υπενθυμίζεται να συμβουλευτείτε τις οδηγίες στην Θεματική Εργασία #3, ως προς το ποια είναι η διαδρομή πρόσβασης, URL, για το συγκεκριμένο ντοσιέ). Χρησιμοποιήστε τη συνάρτηση `download.file()` για να το μεταφέρετε στο χώρο εργασίας σας του R. Ακολούθως, κάνοντας χρήση της συνάρτησης `unzip()` του R, αποσυμπιέστε το για να ανακτήσετε τα περιεχόμενα του – επίσης στο χώρο εργασίας του R – μεταξύ των οποίων θα πρέπει να περιλαμβάνεται το αρχείο `ne_110m_admin_0_countries.shp` με τις ακτογραμμές και τα όρια των κρατών σε παγκόσμια κλίμακα.

Διαβάστε το συγκεκριμένο αρχείο κάνοντας χρήση της συνάρτησης `read_sf()` του πακέτου `sf` και δημιουργήστε ένα αντικείμενο με την ονομασία `world_outline` κάνοντας χρήση των συναρτήσεων `as()` του R και `st_geometry()` του πακέτου `sf`. Απεικονίστε τον παγκόσμιο χάρτη των ακτογραμμών και ορίων των χωρών με μια απλή εντολή της μορφής `plot(world_outline, ...)`.

13. Απεικονίστε ξανά τα **δεδομένα του 1<sup>ου</sup> επιπέδου** του τούβλου των δεδομένων `tmp_raster`, χρησιμοποιώντας την συνάρτηση `levelplot()` του πακέτου `rasterVis`, καθορίζοντας την χρωματική κλίμακα με την επιλογή `rasterTheme(region = rev(brewer.pal(10, "RdBu")))`, όπως ακριβώς κάνατε στο προηγούμενο ερώτημα και προσθέστε στον παραγόμενο χάρτη τις ακτογραμμές και τα όρια των χωρών κάνοντας χρήση μιας επιπλέον εντολής της μορφής

```
plt + latticeExtra::layer(sp.lines(world_outline, col="gray30", lwd=0.5))
```

όπου `plt` υποδηλώνει το αντικείμενο στο οποίο έχετε προηγουμένως εκχωρήσει το αποτέλεσμα της κλήσης της συνάρτησης `levelplot()`, στη μορφή `levelplot(subset(brick.dataset.object, specific layer), ...)` χωρίς τις ακτογραμμές. Εν προκειμένω, η κλήση της συνάρτησης `layer()` είναι απαραίτητο να γίνεται στη μορφή `latticeExtra::layer()` για να γίνει διάκριση μεταξύ των συνώνυμων συναρτήσεων `layer()` στα πακέτα `latticeExtra` και `ggplot2`.

14. Από την πολυδιάστατη μορφή των δεδομένων θερμοκρασίας του NetCDF αρχείου σας, εκχωρήστε το σύνολο των δεδομένων θερμοκρασίας του αντικειμένου `ncfdata` που έχετε ήδη δημιουργήσει στο ερώτημα (4) σε μια μεταβλητή με την ονομασία `tmp_stack`, χρησιμοποιώντας τη συνάρτηση `stack()`, η οποία λειτουργεί σε ένα όνομα αρχείου. Αυτό θα δημιουργήσει μια στοίβα (stack εικόνων raster) του συνόλου των επιπέδων των δεδομένων του NetCDF αρχείου.

Με χρήση της συνάρτησης **subset()** εκχωρήστε σε ένα αντικείμενο με την ονομασία **tmp\_stack\_year** το υποσύνολο των 12 επιπέδων ενός συγκεκριμένου έτους (π.χ. το 1<sup>ο</sup> έτος) των στοιβαγμένων δεδομένων (δηλ. τα δεδομένα θερμοκρασιών για ένα ολόκληρο έτος). Δώστε ετικέτες για τους μήνες με μια εντολή της μορφής

```
names(tmp_stack_year) <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul",
"Aug", "Sep", "Oct", "Nov", "Dec").
```

Δημιουργήστε ένα πολλαπλό γράφημα των επιμέρους εικόνων raster (μία για κάθε μήνα του έτους) που συνιστούν το αντικείμενο **tmp\_stack\_year** . Αυτό μπορεί να γίνει κάνοντας χρήση της συνάρτησης **levelplot(tmp\_stack\_year, ...)**, όπως και στο προηγούμενο ερώτημα, και δημιουργώντας γραφήματα με ή χωρίς τις ακτογραμμές και τα όρια των χωρών.

15. Υποθέστε ότι θέλετε να εξάγετε από τα δεδομένα θερμοκρασιών του NetCDF αρχείου την χρονοσειρά στο σημείο του κανάβου που είναι πλησιέστερα στην τοποθεσία της Αθήνας με γεωγραφικό μήκος 23.73530°E και γεωγραφικό πλάτος 37.98420°N. Υπολογίστε τους δείκτες **j** και **k** για τους οποίους το σημείο με συντεταγμένες **lon(j)** και **lat(k)** είναι το σημείο του κανάβου των δεδομένων που βρίσκεται πλησιέστερα στην προαναφερόμενη τοποθεσία. Τα διανύσματα τιμών **lon** και **lat** που τα έχετε ήδη δημιουργήσει στο ερώτημα 4, περιέχουν τα γεωγραφικά μήκη και πλάτη των σημείων του κανάβου των δεδομένων θερμοκρασίας του NetCDF αρχείου. Ακολουθώντας εξάγετε τη ζητούμενη χρονοσειρά θερμοκρασιών **tmp.array[j, k, ]** από τη συστοιχία τιμών του αντικειμένου **tmp.array** που ήδη έχετε δημιουργήσει στο ερώτημα 8. Για διευκόλυνση εκχωρήστε τις τιμές της εν λόγω χρονοσειράς σε ένα διάνυσμα τιμών με ονομασία **tmp\_ts**.

Από το διάνυσμα **nowdates** των ημερομηνιών που έχετε ήδη δημιουργήσει στο ερώτημα 6 προσδιορίστε την αρχική και την τελική ημερομηνία των χρονικών δεδομένων, π.χ. τις μεταβλητές **beg\_date** και **end\_date** (σε μορφή συμβολοσειράς χαρακτήρων) κάνοντας χρήση των συναρτήσεων **strsplit()** και **as.character()**. Ακολουθώντας μετατρέψτε τις εν λόγω τιμές σε αριθμητικού τύπου, π.χ. τις μεταβλητές **beg\_date\_num** και **end\_date\_num**. Δημιουργήστε ένα γράφημα της χρονοσειράς **tmp\_ts**, χρησιμοποιώντας ως άξονα x του γραφήματος τα έτη αναφοράς (σε δεκαδική μορφή) των παρεμβαλλόμενων τιμών θερμοκρασίας. Καθορίστε κατάλληλες λεζάντες, της επιλογής σας, για τους άξονες και τον τίτλο του γραφήματος.

Σε αυτό το σημείο, θα πρέπει να έχετε ολοκληρώσει την εισαγωγή και επιθεώρηση των δεδομένων του αρχείου **...tmp.dat.nc** που σας ενδιαφέρει, και συνεπώς μπορείτε να το κλείσετε, με την κατάλληλη εντολή. Αποθηκεύστε με την ονομασία **askisi4\_1.Rhistory** το **.Rhistory** αρχείο της μέχρι τώρα συνεδρίας σας στο R.

- (2)** Θεωρήστε ότι χρειάζεστε να αναλύσετε χωρικές χρονοσειρές μετεωρολογικών ή περιβαλλοντικών δεδομένων από ένα αρχείο NetCDF χρησιμοποιώντας το πακέτο **raster** του R. Για τους σκοπούς της Θεματικής Εργασίας θεωρήστε ότι πρόκειται να εξετάσετε δεδομένα υγρασίας του εδάφους, τα οποία παρέχονται σε μια τρισδιάστατη συστοιχία χωρικών δεδομένων ως συνάρτηση του χρόνου. Προηγουμένως δημιουργήστε ένα νέο (άδειο) **.Rhistory** αρχείο ακολουθώντας τις οδηγίες στο ερώτημα (1) της προηγούμενης ενότητας της Θεματικής Εργασίας.

### Multiple map grid series

Continental Gridded dataset in [NetCDF4 format](#) for selected variable over multiple time-steps.  
 NetCDF files can be viewed in some GIS software, including [QGIS](#) and [ArcGIS](#) (using the [Multidimension Toolbox](#)).  
 Complete model outputs (1911-current) are available on request. [Enquire](#)

Time Period	Actual	Relative
Days for 2021	<a href="#">Download NetCDF</a>	<a href="#">Download NetCDF</a>
Months	<a href="#">Download NetCDF</a>	<a href="#">Download NetCDF</a>
Years	<a href="#">Download NetCDF</a>	<a href="#">Download NetCDF</a>

Θα χρειαστεί να πραγματοποιήσετε μη αυτόματη λήψη αυτού του συνόλου δεδομένων από τον ιστότοπο του Γραφείου Μετεωρολογίας της Αυστραλίας ακολουθώντας τη διαδικτυακή διεύθυνση: <http://www.bom.gov.au/water/landscape/> → Soil moisture → Deep soil moisture → Download data

- (i) Κάνοντας χρήση της συνάρτησης **download.file()** με κατάλληλες παραμέτρους εισόδου για τη διεύθυνση

URL και την ονομασία του αρχείου εξόδου στον χώρο εργασίας του Η/Υ σας κατεβάστε το netCDF αρχείο που αντιστοιχεί στην χρονική περίοδο **'Years -- Actual'**, όπως φαίνεται στην παραπάνω εικόνα. Σημειώστε ότι ο σύνδεσμος του συγκεκριμένου αρχείου είναι:

[http://www.bom.gov.au/jsp/awra/thredds/fileServer/AWRACMS/values/year/sd\\_pct.nc](http://www.bom.gov.au/jsp/awra/thredds/fileServer/AWRACMS/values/year/sd_pct.nc)

Για την ανάγνωση, εξαγωγή και διαχείριση των δεδομένων του συγκεκριμένου αρχείου netCDF θα χρησιμοποιήσετε τα πακέτα **ncdf4** και **raster** του R.

1. Ως πρώτο βήμα φορτώστε τα δεδομένα (από το αρχείο που κατεβάσατε) τακτοποιημένα στη δομή ενός 'τούβλου' (δηλ. τρισδιάστατα δεδομένα) χρησιμοποιώντας τη συνάρτηση **brick()** του πακέτου **raster** και εκχωρώντας τα αντίστοιχα στοιχεία σε ένα αντικείμενο (μεταβλητή) με την ονομασία **soil\_moisture.b**.
2. Εξακριβώστε εάν το αντικείμενο **soil\_moisture.b** έχει αποθηκευθεί ή όχι με τη δομή ενός διανύσματος τιμών. Εκτυπώστε στην οθόνη σας την κλάση και το πλήθος των στοιχείων (τιμών), καθώς και τα στοιχεία περιγραφής της δομής του αντικειμένου **soil\_moisture.b**. Ακολουθώντας, κάνοντας χρήση της συνάρτησης **sink()**, διοχετεύστε την εκτύπωση των στοιχείων περιγραφής της δομής του εν λόγω αντικειμένου σε ένα .txt αρχείο, στο χώρο εργασίας του R, με την ονομασία **dsoilm\_data.str.txt**.

Ανοίξτε σε ένα επεξεργαστή κειμένου (π.χ. Wordpad, Notepad) το αρχείο **dsoilm\_data.str.txt** και επιθεωρήστε στα περιεχόμενα του, όπως: τον αριθμό στηλών και γραμμών (**ncols και nrows**) στις οποίες είναι τακτοποιημένα τα δεδομένα και σημειώστε στο πεδίο εντολών του R τις παρατηρούμενες τιμές υπό την μορφή απλών γραμμών παρατήρησης, π.χ. με την ένδειξη

```
# area extend (longitude) = ??? , ???          area extend (latitude) = ???, ???
```

Τέλος επιθεωρήστε τα περιεχόμενα του αντικειμένου **soil\_moisture.b** εκτελώντας απλά την εντολή **soil\_moisture.b**. Εστιάστε την προσοχή σας στις τιμές για τις διαστάσεις (**nrow, ncol**), το πλήθος των κελιών (**ncell**) και των επιπέδων (**nlayers**) της δομής των δεδομένων netCDF, καθώς επίσης την ανάλυση (**resolution κατά x και y**) και την περιοχή που καλύπτουν τα δεδομένα (extend κατά γεωγραφικά μήκη και πλάτη). Σημειώστε τις παρατηρούμενες τιμές τους με την ένδειξη

```
# ncol = ???, nrow = ??? , ncell = ??? , nlayers = ???
```

```
# resolution x = ???, resolution y = ???
```

όπου ??? είναι οι αντίστοιχες παρατηρούμενες τιμές που έχουν προκύψει από τις προαναφερόμενες επιθεωρήσεις σας των στοιχείων του αρχείων netCDF.

Χρησιμοποιήστε τη συνάρτηση **plot()** στο αντικείμενο **soil\_moisture.b** προκειμένου να λάβετε χάρτες απεικόνισης των δεδομένων της υγρασίας του εδάφους για τις διάφορες χρονικές περιόδους των δεδομένων του netCDF αρχείου.

Δημιουργήστε, με μια παρόμοια εντολή, έναν αντίστοιχο χάρτη μόνο για έναν από τα έτη που περιέχονται στα δεδομένα, π.χ. δοκιμάστε το 4<sup>ο</sup> κατά σειρά έτος.

Αντίστοιχα, δημιουργήστε ένα γράφημα που να περιέχει πέντε αντίστοιχους χάρτες από το 4<sup>ο</sup> κατά σειρά έτος έως και το 7<sup>ο</sup> κατά σειρά έτος.

Εξάγετε τις τιμές ενός από τα επίπεδα των παρατηρήσεων που εμπεριέχονται στο αρχείο netCDF (και στο αντικείμενο **soil\_moisture.b**). Π.χ., δοκιμάστε το 1<sup>ο</sup> το 6<sup>ο</sup> και το τελευταίο επίπεδο (layers) των δεδομένων και δημιουργήστε αντίστοιχα αντικείμενα με τις ονομασίες **layer\_1**, **layer\_2**, και **layer\_last**.

Ως επόμενο βήμα, από το επίπεδο των δεδομένων **layer\_1**, εστιάστε στα δεδομένα ενός τμήματος της περιοχής. Εν προκειμένω εστιάστε στο κεντρικό τμήμα της Αυστραλίας, με την ονομασία **subregion** στο αντίστοιχο αντικείμενο του R που θα πρέπει να δημιουργήσετε. Για τον καθορισμό του αντικειμένου **subregion** χρησιμοποιήστε τη συνάρτηση **extend()**, ώστε η περιοχή να περιορίζεται

από τα γεωγραφικά όρια longitude (140° E, 150° E) και latitude (44° S , 30° S). Ακολούθως να εξαγάγετε τις σχετικές πληροφορίες για την εν λόγω υπο-περιοχή και να μετατρέψετε τα αντίστοιχα raster δεδομένα για οπτικοποίηση και ανάλυση. Χρησιμοποιήστε τη συνάρτηση **crop()** στο επίπεδο **layer-1 των δεδομένων** και στο αντικείμενο **subregion** που έχετε ήδη καθορίσει, για να εκχωρήσετε τα αντίστοιχα στοιχεία σε μια μεταβλητή με την ονομασία **r.crop**.

- (i) Για την περιοχή **subregion** εξάγετε τις raster τιμές δεδομένων που αντιστοιχούν σε ένα προφίλ σημείων κατά μήκος του μεσημβρινού  $\varphi=37^\circ$  S, και τα γεωγραφικά μήκη από  $\lambda=140^\circ$  έως  $\lambda=150^\circ$  με βήμα  $0.25^\circ$ . Υπόδειξη, θυμηθείτε τη χρήση των συναρτήσεων **seq()** και **rep()**. Εκχωρήστε τις αντίστοιχες τιμές στα αντικείμενα **lon.pts** και **lat.pts**, και ακολούθως χρησιμοποιήστε τις συναρτήσεις **plot()** επί της μεταβλητής **r.crop** και της συνάρτησης **points()** με κατάλληλη χρήση των **lon.pts** και **lat.pts** για το σχεδιασμό του προφίλ των σημείων. Με τον τρόπο αυτό θα πρέπει να πάρετε ένα γράφημα που θα απεικονίζει τα σημεία του προφίλ ενδιαφέροντος πάνω στο raster **r.crop** της υπο-περιοχής **subregion::layer\_1**.
- (ii) Εκχωρήστε με την έννοια της στήλης τιμών, τις γεωγραφικές συντεταγμένες των σημείων **lon.pts** και **lat.pts**, σε μια μεταβλητή με την ονομασία **extract.pts**. Ακολούθως χρησιμοποιώντας τη συνάρτηση **extract()** επί των αντικειμένων **r.crop** και **extract.pts** εξάγετε με δι-γραμμική παρεμβολή (**method="bilinear"**) τις raster τιμές δεδομένων που αντιστοιχούν στα σημεία του συγκεκριμένου προφίλ. Εκχωρήστε τις εν λόγω τιμές στα σημεία του προφίλ σε ένα αντικείμενο με την ονομασία **ext**, του οποίου εκτυπώστε τα στοιχεία στην οθόνη σας. Ακολούθως δημιουργήστε ένα γράφημα του προφίλ, χρησιμοποιώντας τη συνάρτηση **plot()** επί των αντικειμένων **lon.pts** και **ext** και με ετικέτες **"Longitude"**, και **"Soil Moisture"**, αντίστοιχα στους άξονες x και y.
- (iii) Η συνάρτηση **extent()** δουλεύει κατά παρόμοιο τρόπο και με πολλαπλά layers δεδομένων. Για παράδειγμα, εκχωρήστε σε ένα αντικείμενο με την ονομασία **multi\_layers**, τα δεδομένα από το αντικείμενο **soil\_moisture.b** για τους τελευταίους κατά σειρά 12 μήνες (δηλ. τα τελευταία 12 επίπεδα (layers) δεδομένων του **soil\_moisture.b**). Με τη χρήση της συνάρτησης **extent()** επί των αντικειμένων **multi\_layers** και **extract.pts**, και τη μέθοδο δι-γραμμικής παρεμβολής, εκχωρήστε τα αντίστοιχα raster δεδομένα σε ένα νέο αντικείμενο με την ονομασία **new.extract**, του οποίου εκτυπώστε τις τιμές του στην οθόνη σας. Επίσης δημιουργήστε ένα γράφημα των προφίλ ενδιαφέροντος για τα πολλαπλά επίπεδα (12 μήνες) δεδομένων, χρησιμοποιώντας τη συνάρτηση **plot()** επί των αντικειμένων **lon.pts** και **new.extract** και με ετικέτες **"Longitude"**, και **"Soil Moisture"**, αντίστοιχα στους άξονες x και y.

Επαναλάβετε τα ερωτήματα των προηγούμενων ενότητων (i), (ii) και (iii), αλλά αυτή τη φορά ξεκινώντας από την εξαγωγή raster τιμών δεδομένων που αντιστοιχούν σε προφίλ σημείων κατά μήκους του γεωγραφικού παραλλήλου  $\lambda=147^\circ$  E, και για τα γεωγραφικά πλάτη από  $\varphi=44^\circ$  S έως  $\varphi=32^\circ$  S, με βήμα  $0.25^\circ$ .

Σε αυτό το σημείο, θα πρέπει να έχετε ολοκληρώσει την εισαγωγή και επιθεώρηση των δεδομένων του αρχείου **... sd\_pct.nc** που σας ενδιαφέρει, και συνεπώς μπορείτε να το κλείσετε, με την κατάλληλη εντολή. Αποθηκεύστε με την ονομασία **askisi4\_2.Rhistory** το .Rhistory αρχείο της μέχρι τώρα συνεδρίας σας στο R.

**ΣΗΜΕΙΩΣΗ** – Με τα παραδοτέα σας θα πρέπει να συμπεριληφθούν τα εκάστοτε .Rhistory αρχεία που θα προκύψουν από τις ξεχωριστές συνεδρίες του R που διεκπεραιώσατε για την υλοποίηση κάθε μέρους της Θεματικής Εργασίας.