

Daily Met Surface observations for the NARCCAP region and time period.

This is a convenient compilation of *daily* minimum/maximum temperature, precipitation for the period 1970 - 2011. Monthly average minimum/maximum temperatures and monthly precipitation totals are also included. The basic source is the GHCN data base although this has been cleaned by some by the mica group in processing. Surprisingly the variables *tmin* and *tmax* are the basic measurements for most stations. The daily mean is typically taken to be $(tmin + tmax)/2$.

NOTE Due to different missing value patterns *tmin* and *tmax* have different numbers of stations. For convenience the monthly data has removed the stations that are not common.

Monthly average station values

By "monthly average daily minimum temp" it is meant: find the minimum temperature for each day and then take the average of these for the month. Temperatures are reported in degrees C. These data consist of about 17K stations but typically there are about 9K reporting in a given month. If a station has more than 15 days missing in a month I have recoded the temperatures to missing (NA in R). The R binary data set file can be downloaded from:

www.image.ucar.edu/pub/nychka/mica/monthlyTempR/monthlyTemp.rda

and is about 34Mb.

To access this data set in R set your working directory to the same directory as the **tempMonthly.rda** or in the example below include the full pathname to this file.

```
> load("tempMonthly.rda")
> ls()
[1] "indexMonthly" "mn"
[3] "time"         "tmaxMonthly"
[5] "tminMonthly"  "yr"
```

- indexMonthly is a 6 column data frame that lists the id, lon, lat, elevation and start and end period for each station -- about 17K rows in this data frame.

- tmaxMonthly and tminMonthly are arrays with 3 arguments that contain the monthly means. The dimensions are:

17533, 12, 42, i.e. station, month, year.

stations are in the *same* order as indexMonthly.

Separating the month and year make it easy to find other statistics or subset e.g. just look at June, find annual averages.

So tminMonthly[5222,12, 1] is the value for station #5222 (actually Boulder) for DEC 1970

This array is organized so that if you collapse on 2nd and 3rd dimensions you will get a monthly time series for a station. E.g. c(tminMonthly[5222, ,]) times series of tmin for Boulder

Given this time series the following objects are helpful

- time are the year and fraction of year (e.g. jan 1975 is 1975.042)
- mn indicates the months for the columns
- yr indicates the years

=====

Daily temperatures and Precipitation

The daily data sets for each variable are about 250Mb each.

The R binary files can be downloaded as "tarballs" from:

www.image.ucar.edu/pub/nychka/mica/tmaxR.tar.gz

www.image.ucar.edu/pub/nychka/mica/tminR.tar.gz

www.image.ucar.edu/pub/nychka/mica/tminR.tar.gz

For reference the index files for these data can be obtained separately as

www.image.ucar.edu/pub/nychka/mica/tmaxIndex.rda

www.image.ucar.edu/pub/nychka/mica/tminIndex.rda

www.image.ucar.edu/pub/nychka/mica/tminIndex.rda

so it possible to view the extent of these data without downloading the entire volume.

To expand these files in UNIX:

```
gunzip tmaxR
tar -xvf tmaxR.tar
```

This will create a directory **tmaxR** of about 500 files. Data files have the name tmaxYYYYMM indicating the year and month and includes all the stations in the same order as **tmaxIndex**. So **tmax197503** has the daily values for all stations and for March, 1975.

To keep the index and data together the index file is also included as the last file (in alpha order) in this directory.

Loading each of these data files into R results in an R object, a matrix, with name tmaxYYYYMM where rows index stations and columns index the days of the month. The names of columns are dates as character strings but in the default format to be converted by the **as.Date** function into a date Object .

NOTE: tmin and tmax datasets have different numbers of stations, use the `match` function on the station ids to reconcile these to get common **tmin**, **tmax** values. The precipitation station network is substantially different from the surface temperature network.

////////////////////////////////////

Quick examples for monthly data

```
#Set the working directory to be the one including this file
load( "tempMonthly.rda")
```

these are all the station locations

```
library( fields)
loc<- cbind( indexMonthly$lon, indexMonthly$lat)

ind<- !is.na(tminMonthly[,6,1])
quilt.plot(loc[ind,] , (tminMonthly[ind,6,1] + tmaxMonthly[ind,6,1])/2)
map( "world", add=TRUE)
title("1970 June daily average (tmin+tmax)/2")
```

average for 1975

```
ann1975<- apply( tminMonthly[, ,6], 1, FUN="mean", na.rm=TRUE)
ind2<- !is.na(ann1975)
quilt.plot(loc[ind2,] , ann1975[ind2] )
title( "Tmin june average")
```

number of stations reporting by month and by year

```
good<- apply( !is.na(tminMonthly), c(2,3), FUN="sum")
image.plot( 1:12, 1970:2011, good)

plot( time, c( good), xlab="years", ylab="stations reporting")
```

find the closest station to Boulder,CO coordinates.

```
distBoulder<- rdist.earth( cbind( -105.2, 40), loc)
indexBoulder<- which.min( distBoulder)

t2<- c(tmaxMonthly[indexBoulder , ,] )
t1<- c(tminMonthly[indexBoulder , ,])

plot( time, (t1+t2)/2, type="l" )
title("Boulder, CO monthly mean temperatures")

plot( mn, (t1+t2)/2, xlab="month")
title("Boulder seasonality")
```

Quick examples for daily data

```
#Set the working directory to include the tmaxR directory from the tarball.

load("tmaxR/tmax197503.rda")
dim(tmax197503)
[1] 17596 31

#plot of daily max for March 15, 1975
```

```
load("tmaxR/tmaxIndex.rda") loc<- cbind( tmaxIndexlon, tmaxIndexlat) colnames( tmax197503 ) #double
```

```
check dates are right Y<- tmax197503[,15] ind<- !is.na( Y) library(fields) quilt.plot( loc[ind,], Y[ind]) ``
```

assembling a times series for a station(s)

```
fileNames<- dir("tmaxR")
fileNames<- fileNames[-505]
# omit last one (it is the index file!)
load("tmaxR/tmaxIndex.rda") # load index file
print( tmaxIndex[4500,]) # a station
```

Read the whole daily tmax files into R this will take some memory -- but works on my macbook air(!)

```
for(fName in fileNames){
cat(fName, fill=TRUE)
load(paste0("tmaxR/",fName))
}
# remove the .rda from names
objectNames<- substring(fileNames,1,10)
# loop through all months and accumulate the
# station values
temp<- NULL
for( dName in objectNames){
cat(dName, fill=TRUE)
# select stations 4500 and 4503 this
# of course can be modified to
# grab a more interesting subset of stations
nextMonth<-(get(dName))[c(4500,4503),]
temp<- cbind( temp, nextMonth)
}
time<- as.Date( colnames( temp)) # names of the columns are date strings
matplot( time, t(temp), type="l",lty=1)
title( paste("station 4500 and 4503 daily tmax record") )
```