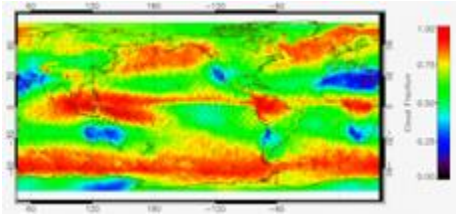
HNE Eberswalde
Hochschule für nachhaltige Entwicklung (FH)

PIK

# Visualization of Climate Data in R



By: Mehrdad Nezamdoost
Mehrdad@Nezamdoost.info
April 2013

# Content

- Why visualization is important?
- How climate data store ?
- Why R?
- Packages which used in R
- Some Result in R
- Combining Script in SimEnvVis
- Future work and opportunities
- Over View

# What is Visualization-Terminology

**What is Visualization?**

computer-supported generation of graphical representations from data

**What can Visualization do?**

intuitive, compressed representation of complex data

# Use of Visualization/Visualization Task

**explorative analysis**

- starting point: unknown data set without any hypothesis about the data - interactive, often non-directional search

**confirmative analysis**

- starting point: hypothesis about the data - validation of the hypothesis

**presentation**

- starting point: already known facts – presentation / communication of the results

# How climate data store ?(NetCDF Format)

Founded by NASA and developed by Unidata (Natinal Science Foundation )

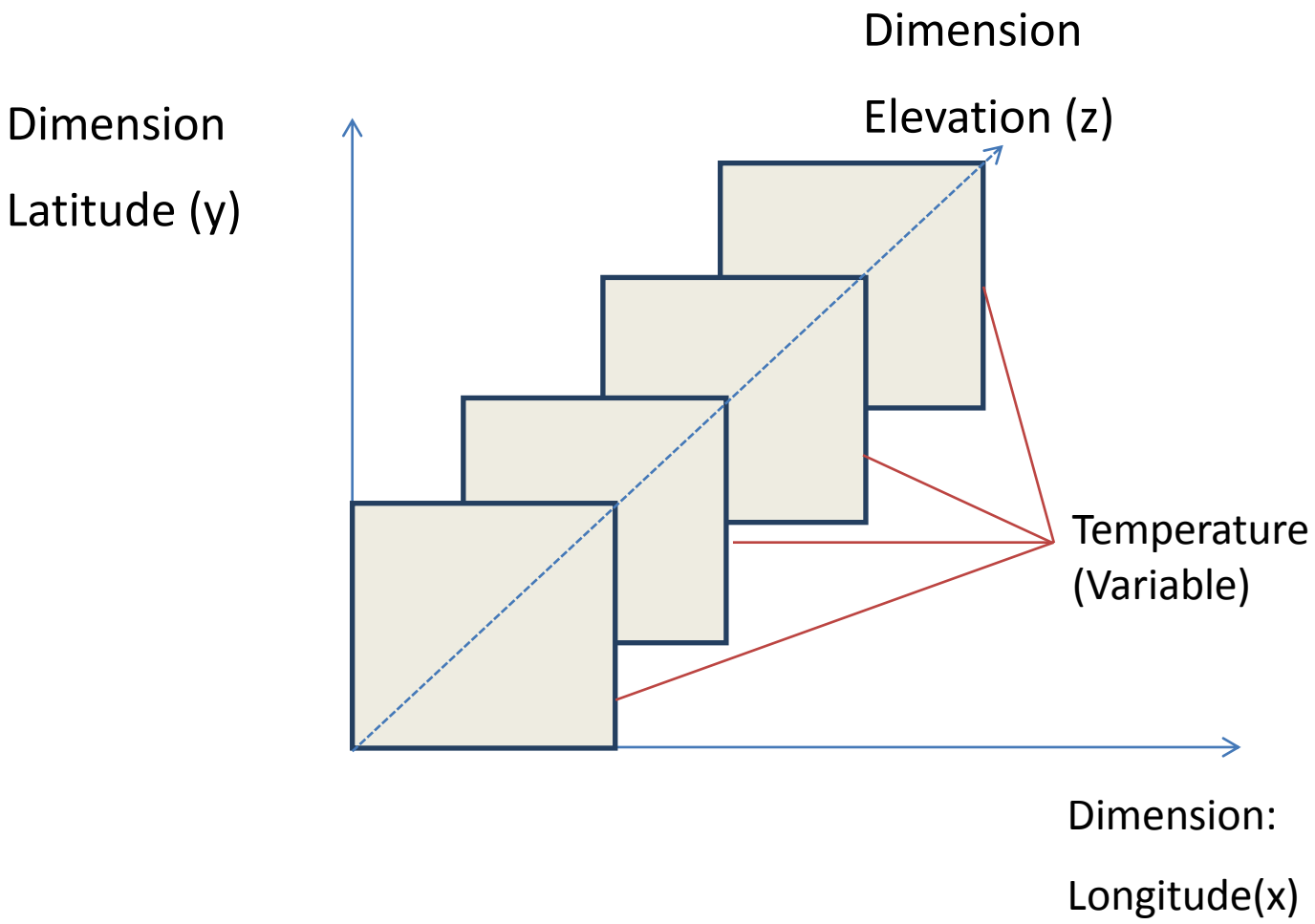
machine-independent format for representing scientific data-*.nc

- Self Describing
- Portable
- Shareable
- C , C++ , Java , R , python , Ruby , ArcGIS and…
- Current version 4.1 (2010)

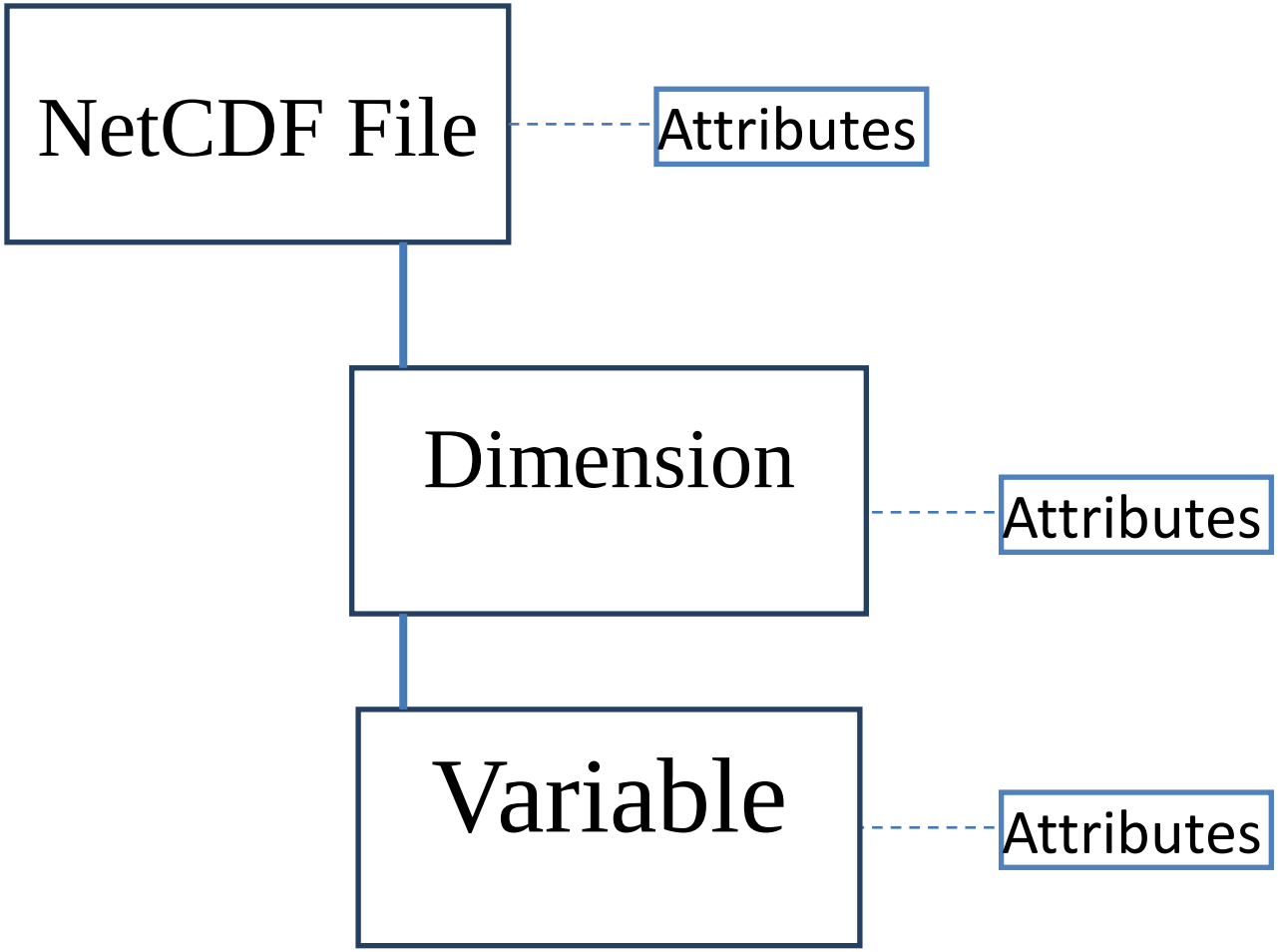
---

http://www.unidata.ucar.edu/software/netcdf/credits.html (access at 22.4.2013)

# How its look like?



# How its look like?



# Look inside NetCDF File

- List of 10
- $ id : int 4
- $ ndims : int 3
- $ natts : int 7
- $ unlimdimid : num 3
- $ filename : chr "C:\\PIK\\Data\\n4 177f9060-a351-11e1-b55e-e1515196aa7„
- $ varid2Rindex: num [1:7] 0 0 0 1 2 3 4
- $ writable : logi FALSE
- $ dim :List of 3
- ..$ longitude:List of 8
- .. ..$ name : chr "longitude"
- .. ..$ len : int 720
- .. ..$ unlim : logi FALSE
- .. ..$ id : int 1
- .. ..$ dimvarid : num 1
- .. ..$ units : chr "degrees_east"
- .. ..$ vals : num [1:720(1d)] 0.25 0.75 1.25 1.75 2.25 2.75 3.25 3.75 4.25 4.75 ...
- .. ..$ create_dimvar: logi TRUE
- .. ..- attr(*, "class")= chr "dim.ncdf„
- ……
- $ nvars : num 4
- $ var :List of 4
- ..$ biomass_carbon_burning_nonCF:List of 16
- .. ..$ id : int 4
- .. ..$ name : chr "biomass_carbon_burning_nonCF"
- .. ..$ ndims : int 3
- .. ..$ natts : int 4
- .. ..$ size : int [1:3] 720 279 4
- .. ..$ prec : chr "float"
- .. ..$ dimids : num [1:3] 1 2 3
- .. ..$ units : chr "kg m-2"
- .. ..$ longname : chr "biomass carbon burning"
- .. ..$ dims : list()
- .. ..$ dim :List of 3

File Description

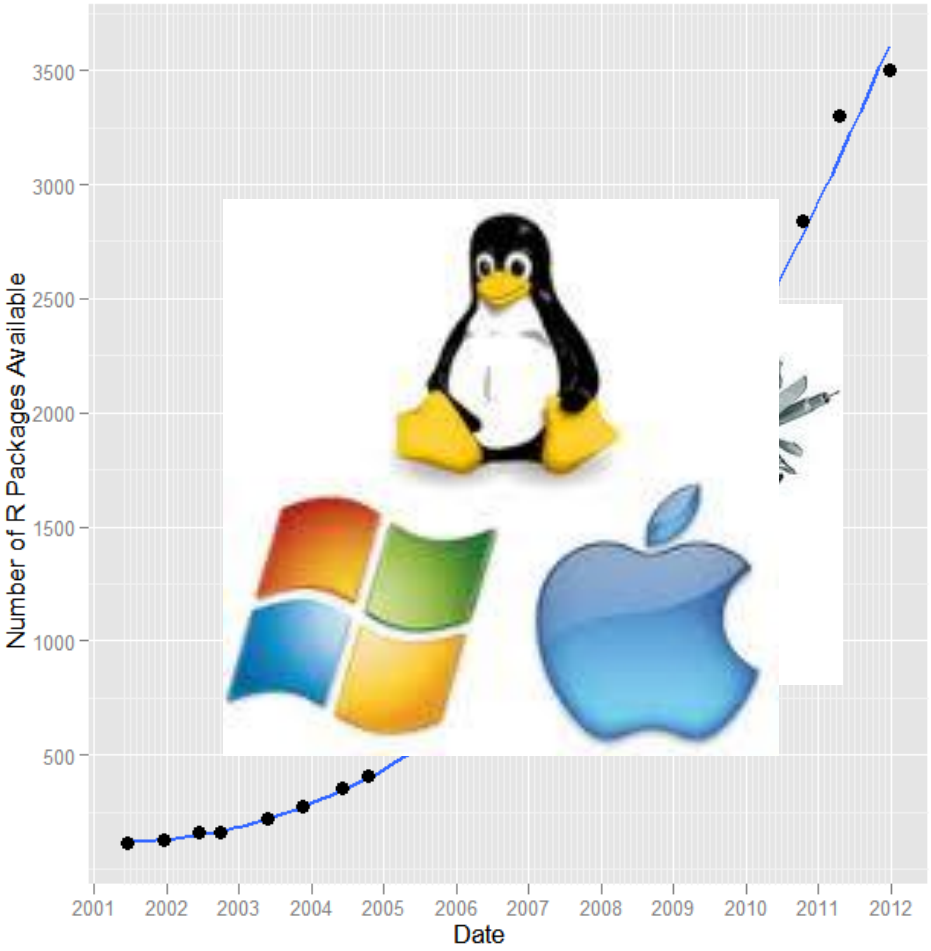
1st Dimension Description

1st Variable Description

# How to read *.nc file

- Linux: ncdump command
- Widows : ncBrowse
- R: ncdf , RNetCDF  ( read.nc(file.choose())

# Why R?




Advantage:

-Open Source

-huge support from user

-user working for user

-fast growing

-easy to understand

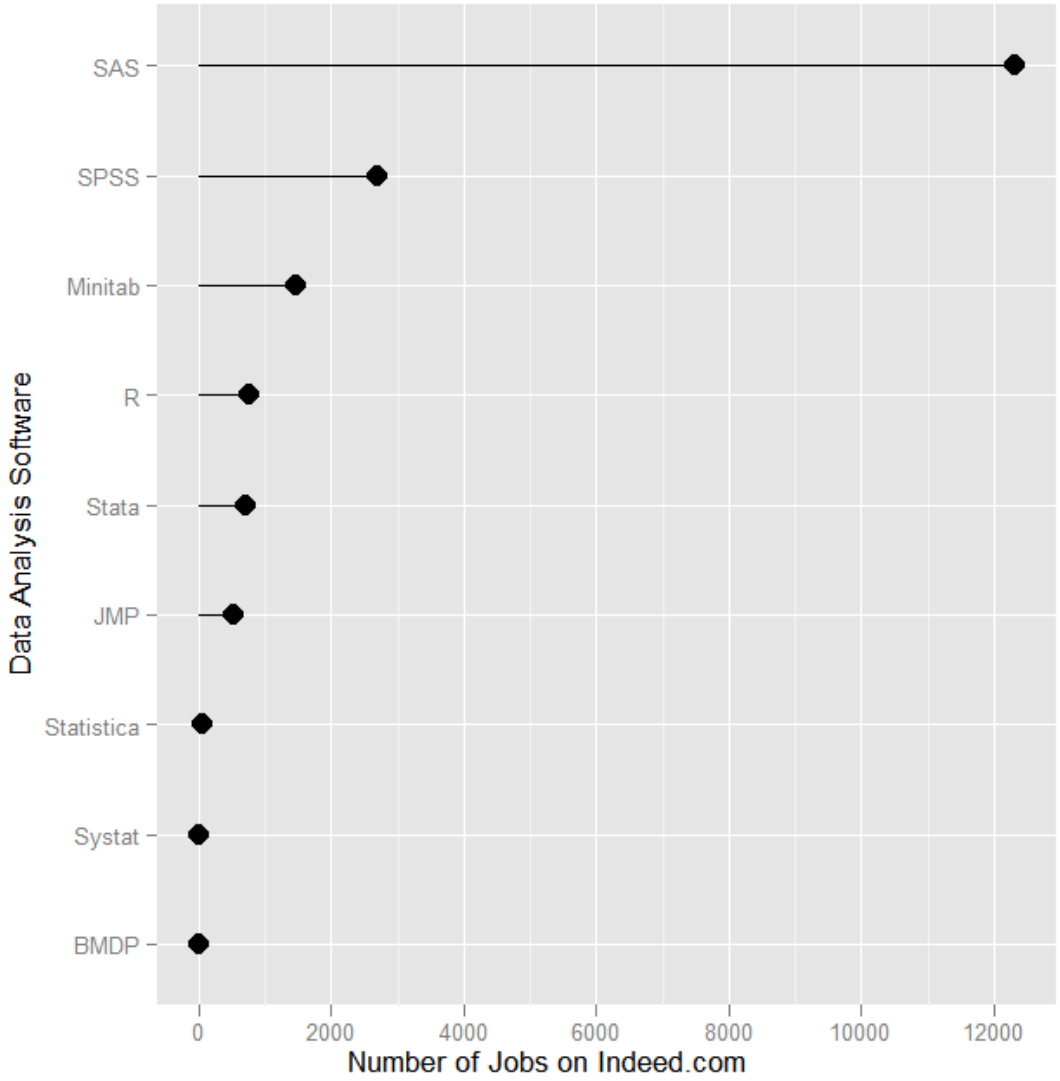
-Change the shape of software as you wish depend on what packages you load

-Cross platform

- Transparent

https://sites.google.com/site/r4statistics/popularity by by Robert A. Muenchen (access at 22.4.2013)

# Why not R ?




Disadvantage:

-changing in version caused some package is not working properly

-The Transparency is good for scientific work and in job world other software may use more often

# Packages which is used

- Visualization :
  - ggplot2
  - Raster(+plot)
- Data preparation :
  - Plyrpackage
  - RNetCDF
  - Arrayhelper

  (Choosing right package)

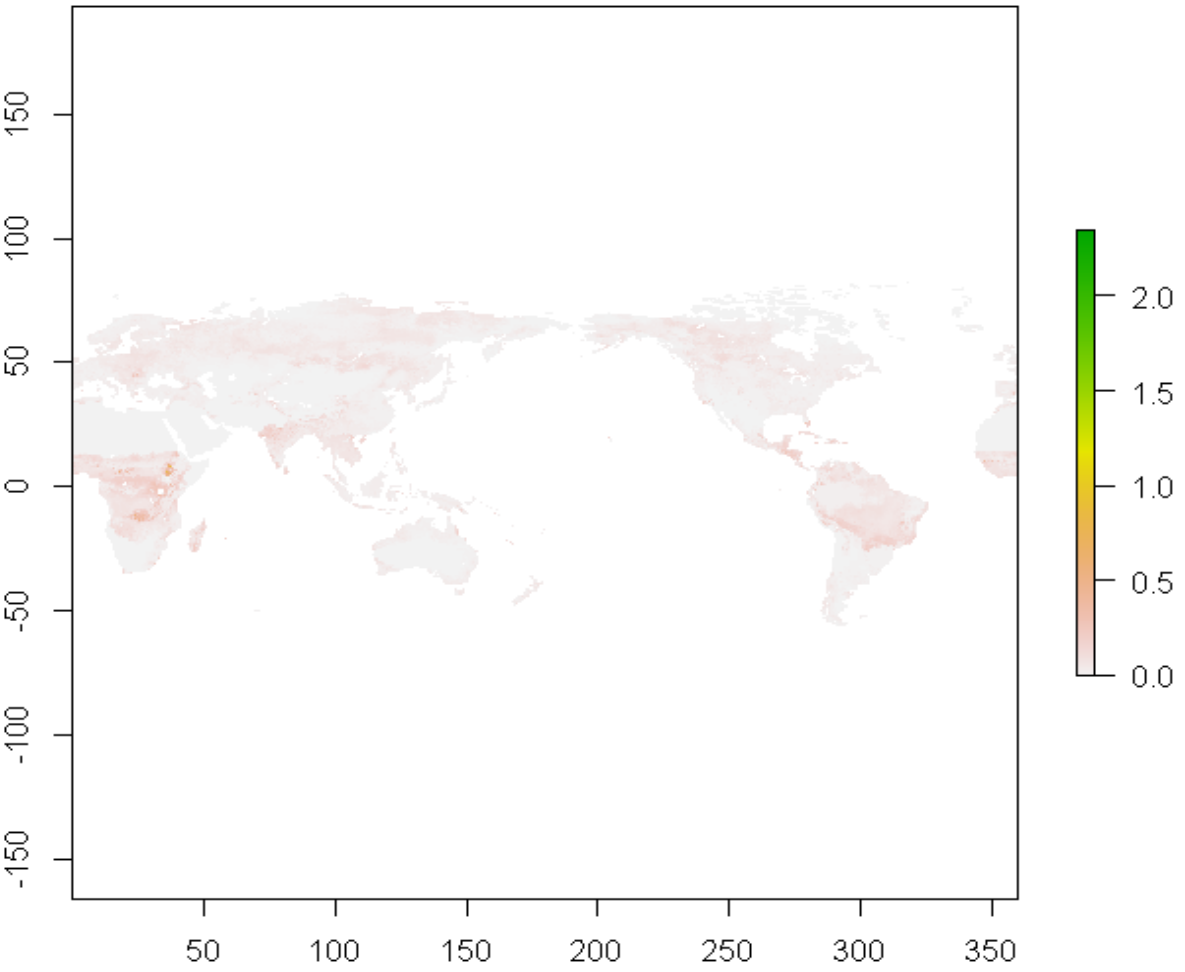
# Some Result(script)

- Script Netcdf to Data frame

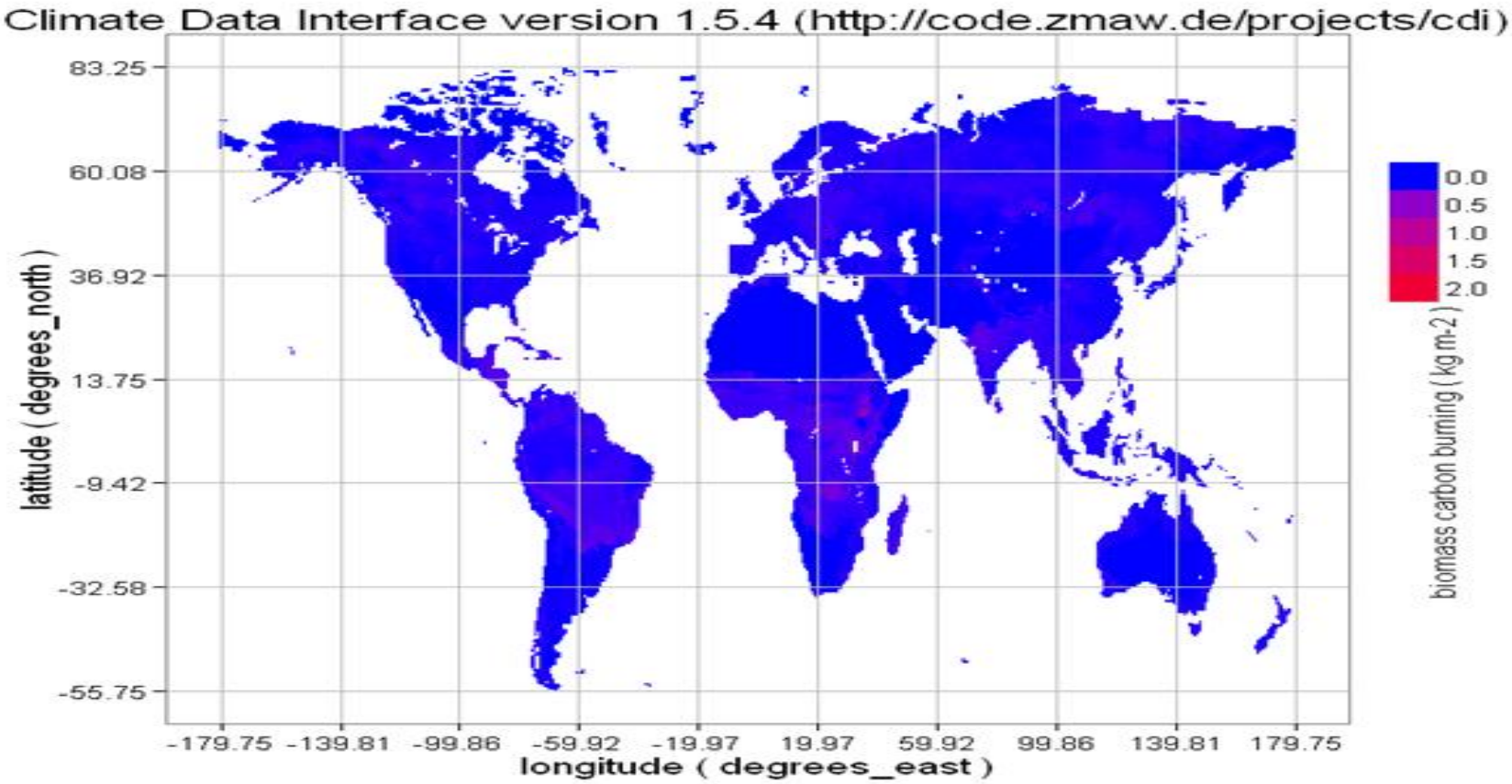
N:\Study\FIT\3 HNEE\Research Project\PIK\PIK\R scripts\NCtoDataFrame_ver1_5_RNetCDF_array.

```
# the thing that I could do with 3 line I spend 2 weeks to find out

Ncdf2df=function(nc=1)
{
      library(RNetCDF)
      library(arrayhelpers)

      if (nc==1)nc=open.nc(file.choose())
      nd=file.inq.nc(nc)[[1]]# number of dimension
      nv=file.inq.nc(nc)[[2]]# number of variable start from 0
#_____________________________________Extract Dimension__________

for(i in 0:(nd-1))
      {
                  Dname=(var.inq.nc(nc,i)[[2]])#name of variable
                  Dvalue=var.get.nc(nc,i)
                  if (i==0)
                  {
                  ll=list(V=Dvalue)
                  names(ll)[[i+1]]<-Dname
                  }
```

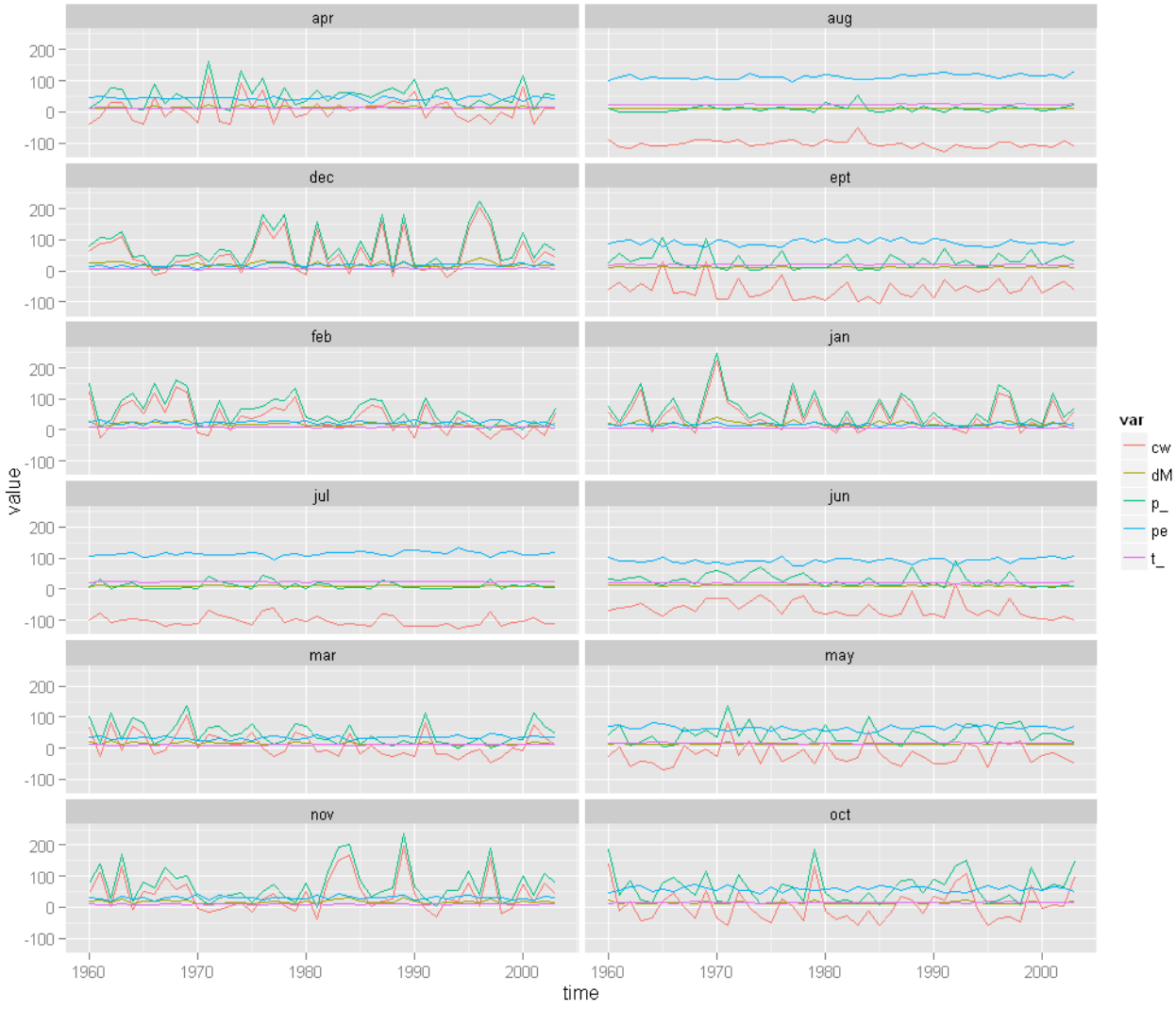
# Some Result: raster+plot **presentation**



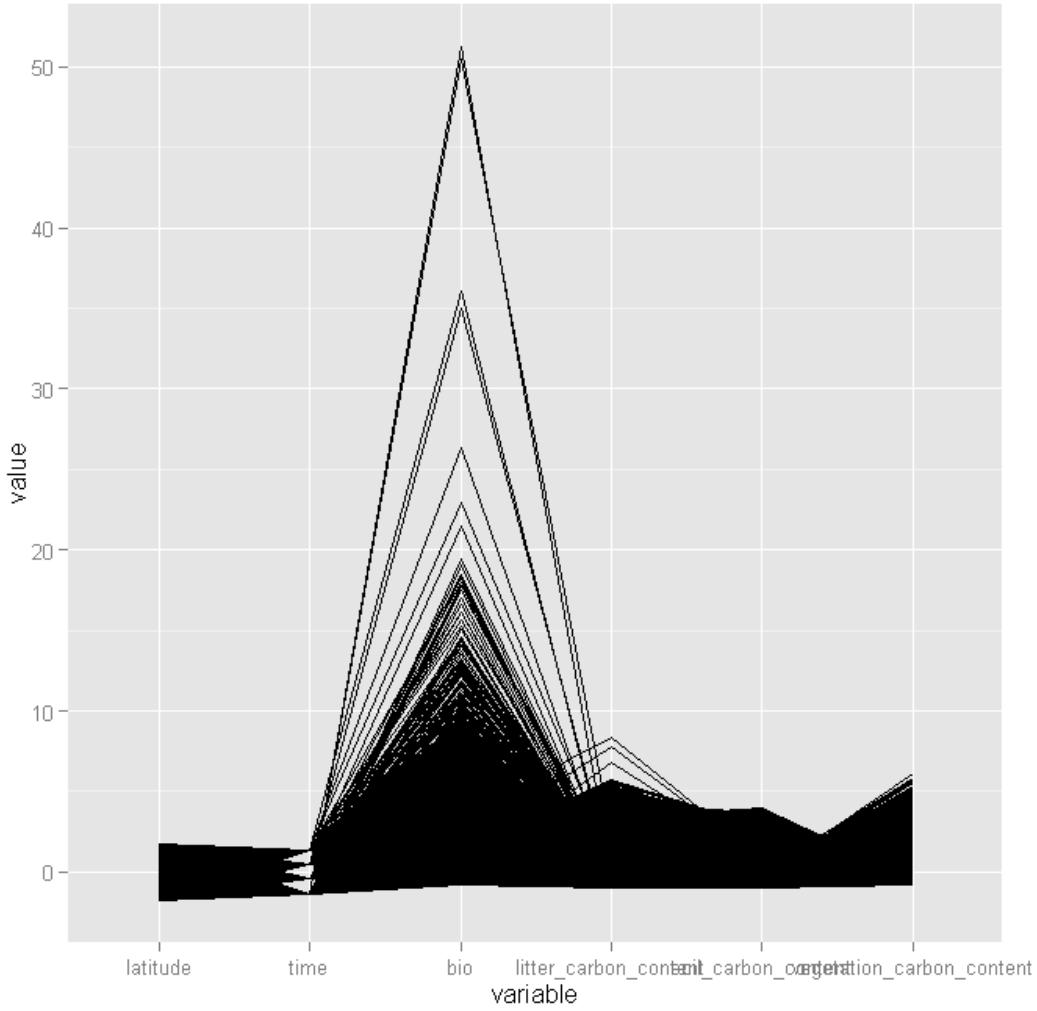
# Some Result: heat map,ggplot2 **presentation**



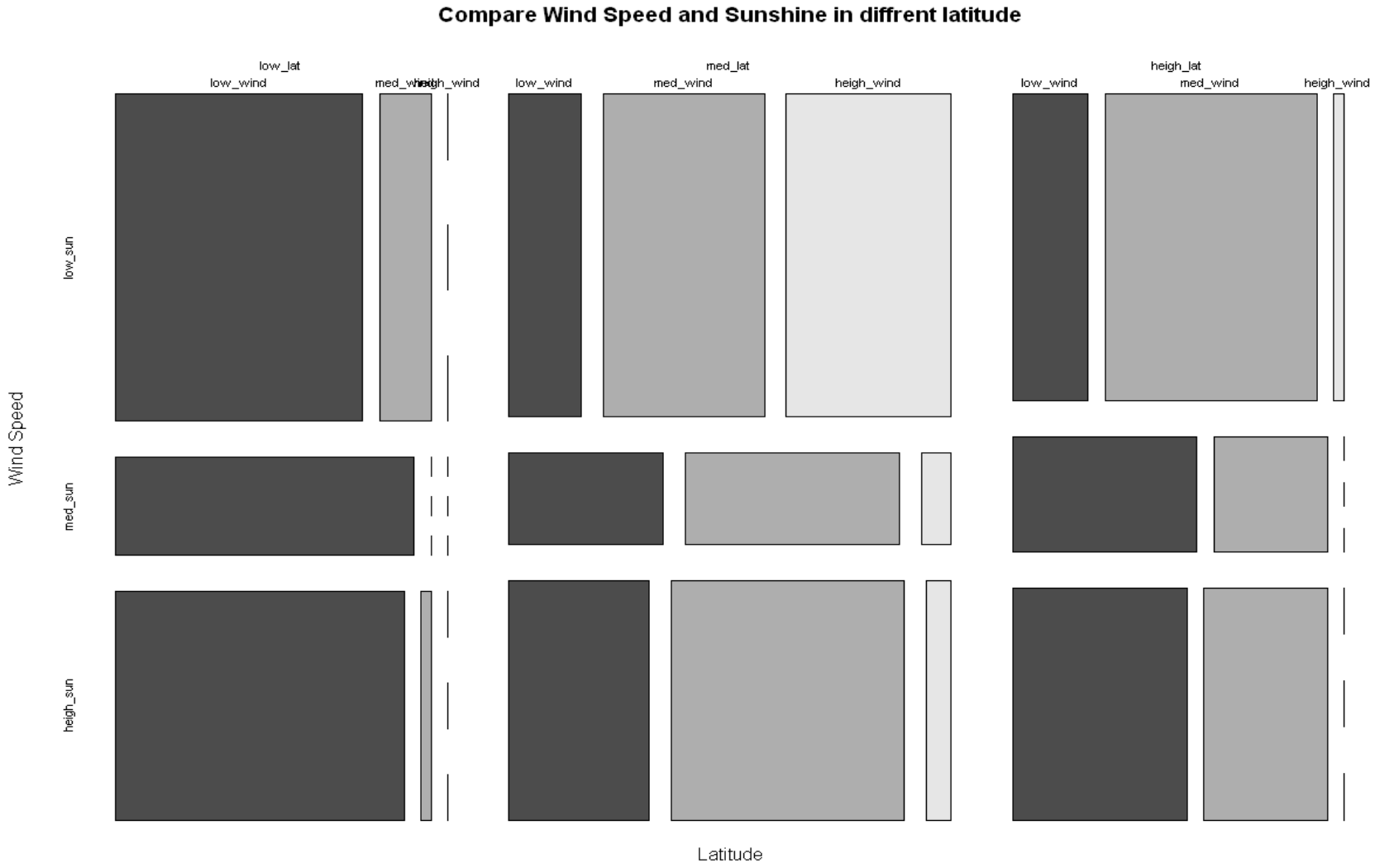
# Some Result: Time Seriese **presentation**



# Some Result: parallel coordinate **explorative analysis**

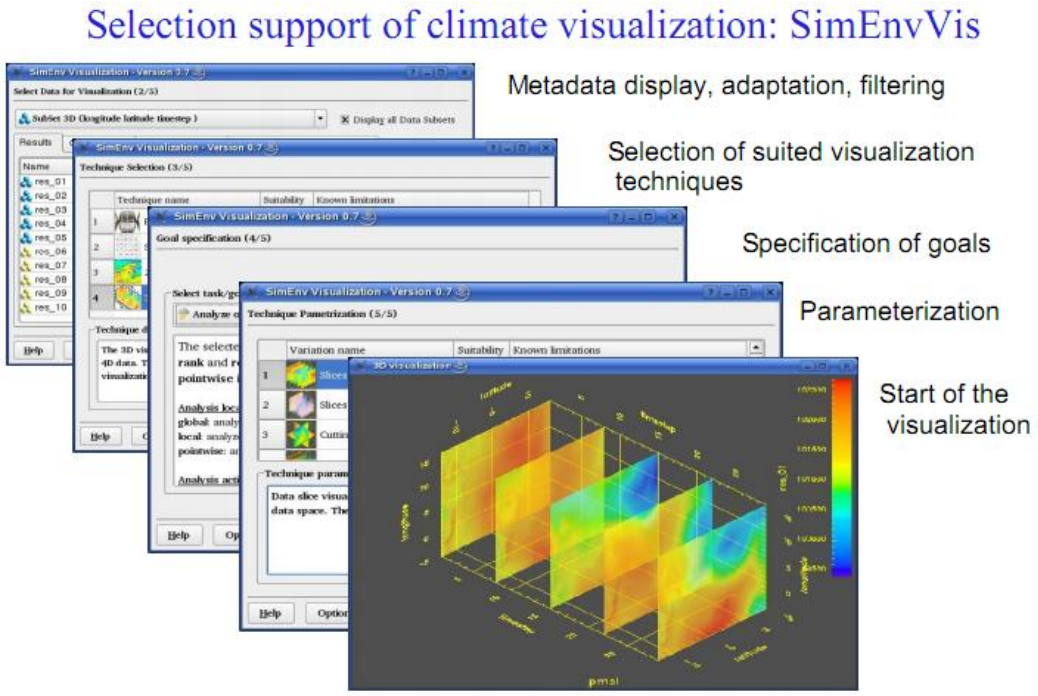


# Mosaic Plot
# **explorative analysis**



# Combining Script in SimEnvVis

- SimEnvVis is the open source software to Visualize the climate data , PIK .



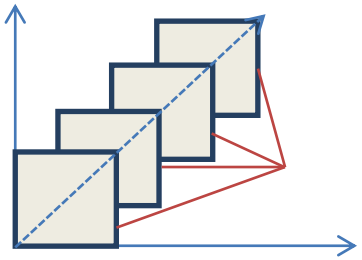

- Language which used for integration :
  C++ and Qt for UI ( R Studio)

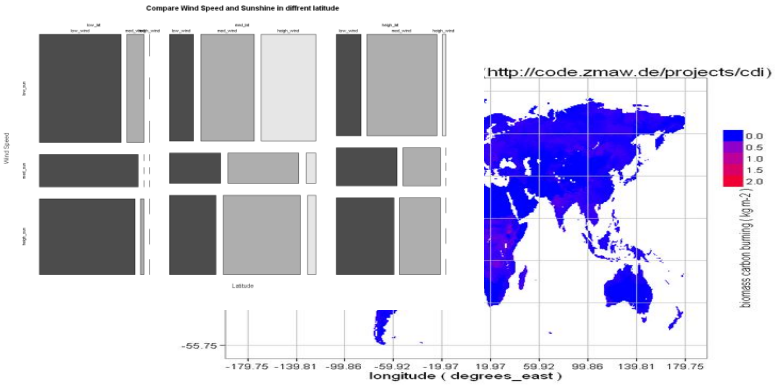
http://www.pik-potsdam.de/research/transdisciplinary-concepts-and-methods/projects/modsimenv/simenv

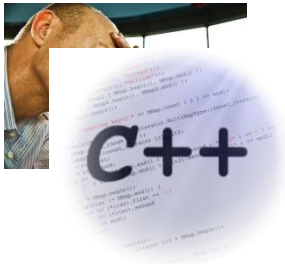
# And More Result …

- Getting to know new part of science Visualization
- Getting to know Open Active Open Source community (Qt, R , Linux , Python , QGIS, CMS, JQuery )→ C++
- Knowing valuable and free source bank ( Github , Source Forge)
- Object Oriented Concept in C++
- Data Management in R
- Be patient (4H each day)

# Over View




Compare Wind Speed and Sunshine in diffrent latitude
http://code.zmaw.de/projects/cdi)
biomass carbon burning ( kg/m-2 )
0.0
0.5
1.0
1.5
2.0
-55.75
-179.75 -139.81 -99.86 -59.92 -19.97 19.97 59.92 99.86 139.81 179.75
longitude ( degrees_east )


c++

# Conclusion and Future research

- Working more in Technical part of Visualization like hypothesis about the data OR validation of the hypothesis
- Entering to more in data mining issues like sensitivity analysis
- Or developing model , testing models and finding proper tuning of existing model for achieving certain goal