The R environment for statistical computing and graphics.

Douglas Nychka,
Emaj

- Some History
- The langauge
- Some examples
- Contributed packages and help.
- Connections to other environments.

NCAR

Supported by the National Science Foundation DMS

Some History

The **S** language was developed in the Bell labs statistics group from 1979 through 1992. It is a C-like language (hence the single letter), command line driven, but focuses on many functions for statistics and graphics.

S was largely an academic curiosity until it was commercialized in 1993 as S-plus. One difficulty of using S-plus was creating and distributing new packages to add more functionality.

More History

 was initially designed by Gentleman and Ihacka at the University of Auckland. Over time it has grown as a statistics community based *GNU project* and so is free.

Currently R is the primary tool for most research in statistics and is also used extensively for data analysis and teaching.

Some Attributes

R is an interpretative language that has many similarities with Matlab or IDL.

It is not as efficient as Matlab for linear algebra but can handle more complicated data structures and provides more statistical functions.

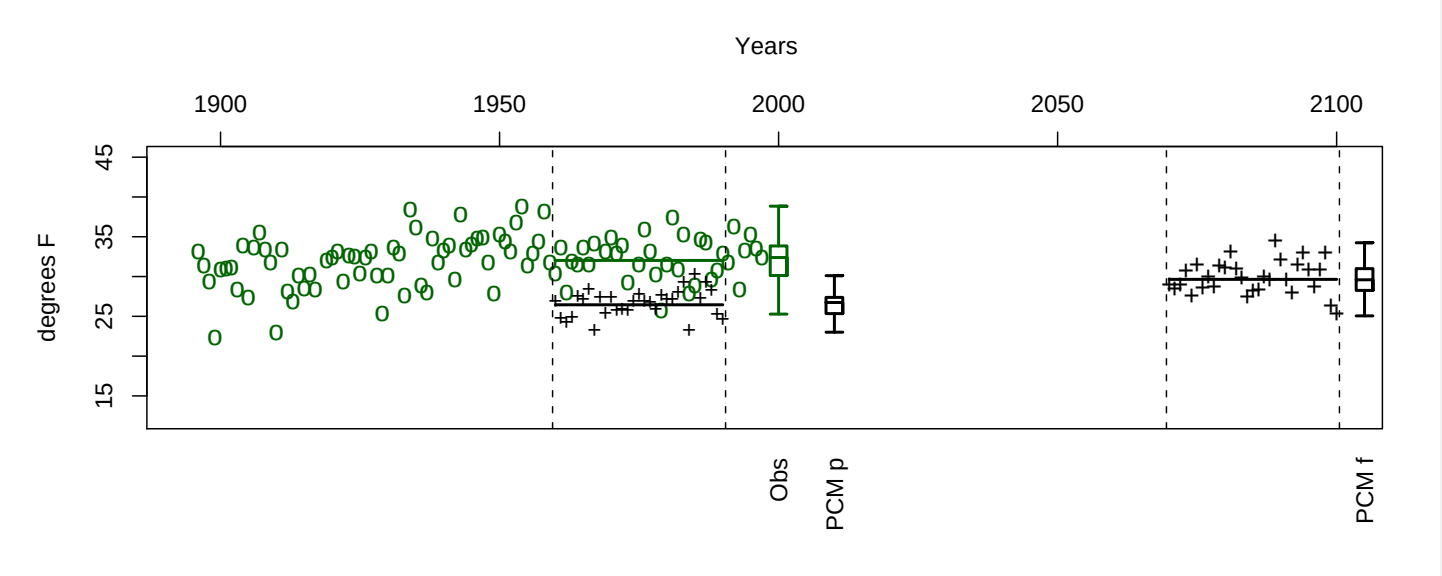
One strength of R is the ease of creating high quality contributed packages.

R runs on Windows, Mac OSX and many varieties of Unix.

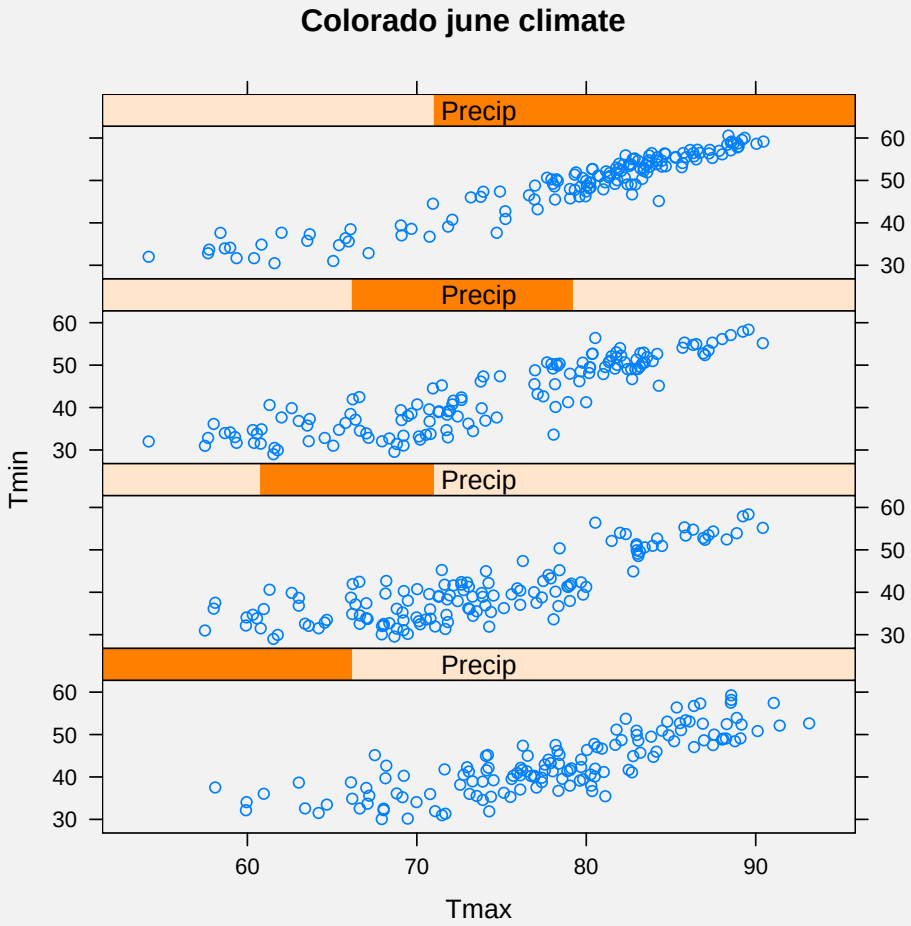
Some Examples

Graphics:

R provides flexibility in drawing many kinds of plots



, but maps and perspective plots are not as extensive as NCL or Matlab.



Lattice library provides useful graphical summaries involving repeating simple plots.

Note that statistical graphics is designed to be spare but with high information content.

STAT 101: Linear Regression

Regressing on June T max on T min

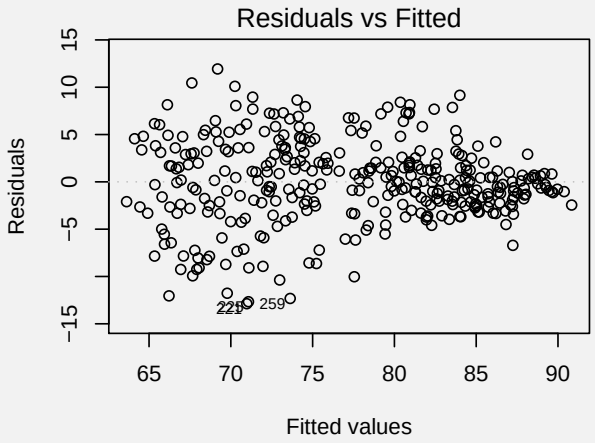
`out= lm( Tmax~Tmin)` Fitting linear model.

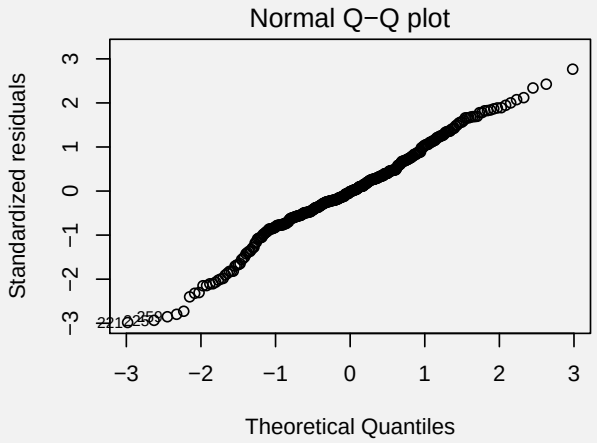
`summary( out)` Overloading of the summary function.

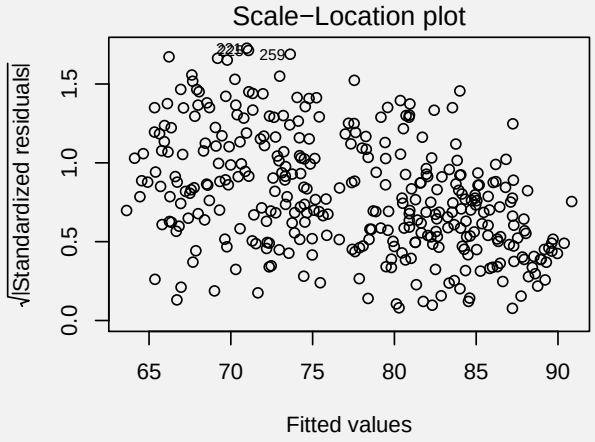
```
Call:
lm(formula = Tmax ~ Tmin)

Residuals:
      Min        1Q    Median        3Q       Max
-12.87607  -2.38583  -0.06851   2.64686  11.93170

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.64072    1.27195   30.38   <2e-16 ***
Tmin         0.86162    0.02779   31.00   <2e-16 ***

Residual standard error: 4.328 on 348 degrees of freedom
Multiple R-Squared: 0.7342,     Adjusted R-squared: 0.7334
F-statistic: 961.3 on 1 and 348 DF,  p-value: < 2.2e-16
```

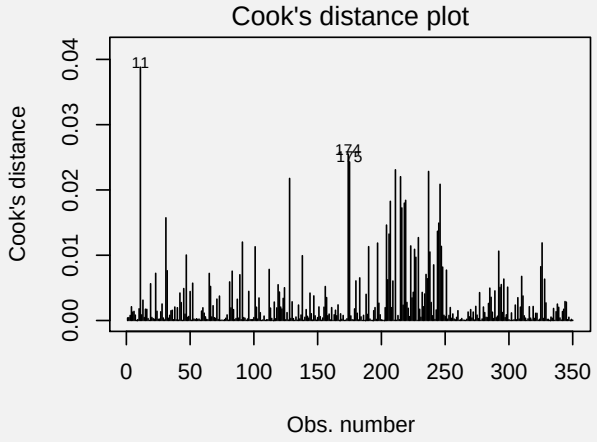
Regression diagnostics

`plot( out)`









Packages

There are hundreds of contributed R packages that provide more specific functions. e.g.

- *ncdf* Reads and writes netcdf format.
- *fields*, *geoR*, *KRiSP* Analysis of spatial data
- *Rawindsonde* skew-t plots
- *SparseM* Sparse matrix methods
- *extRemes* GUI based analysis of extremes

Packages posted to the central archive have passed a series of checks for completeness and have Windows binaries and UNIX source.

Package install is

In UNIX (with sudo)

`R CMD install fields` a single line !

In Windows

selecting an item from a pull down menu

or

self extracting exectuable

R provides utilities for creating documentation and many tutorials

- Markup language for help file
- Automatic testing of documentation syntax and examples.
- Conversion to html, pdf and indexing.
- Tutorials and manuals English (15), French(2), German, Italian (6), Japanese(1), Spanish(5),Polish, Portuguese, Chinese
- Searching

Connections

R provides usual capability of executing shell commands.

- *tcltk* Interface to tcltk scripting and GUI language
- *ggobi* Dynamic graphics

The Omega Project

`www.omegahat.org` Careful reading of the preamble to the project web page reveals the developers (computational statisticians) will do basically whatever they want
but most activities seem to have some anchor in R/S.

- *Rmatlab* Bidirectional connection to Matlab
- *RSPython*
- *RSPerl*
- ... Many projects I don't understand!