



The National Center for Atmospheric Research is sponsored by the National Science Foundation. This work is supported by NSF EASM grant 737208

Manipulating and Visualising Data

Cindy Bruyère and Mari Jones



I have my model runs... Now what?

- What tools are available for data analysis?
- What science questions are you assessing?
- How large is your data set?
- How do you want to present your results?
 - *Visually e.g. maps of differences in variables*
 - *Tabulated e.g. statistical significance analyses*
 - *Correlations and dependencies*

Visualization Packages

	NCL	RIP4	ARWpost (GrADS)	UPP	VAPOR	IDV
Directly ingest WRF data	Y	N <i>converter</i>	N / (Y) <i>converter</i>	N <i>converter</i>	N / (Y) <i>converter</i>	N / (Y)
Intermediate files	N	lots	large file	Y	large file	Y
WPS DATA	Y	Y	Y	N	N	Y
wrfinput data	Y	Y	Y	N	N	Y
Input format	<i>netCDF</i>	<i>netCDF</i>	<i>netCDF</i>	<i>netCDF / binary</i>	<i>netCDF</i>	<i>netCDF</i>
Vertical Output Coordinate	eta pressure height	eta pressure height	eta pressure height	pressure	eta	if pre- processed
Software required (All binaries are free)	<i>NCL</i>	<i>NCARG</i>	<i>GrADS</i>	<i>GrADS / GEMPAK</i>	<i>VAPOR</i>	<i>JAVE</i>
Diagnostics	<i>some</i>	<i>> 100</i>	<i>some</i>	<i>some</i>	<i>limited</i>	<i>limited</i>

MatLab / IDL / **R** / ferret / python

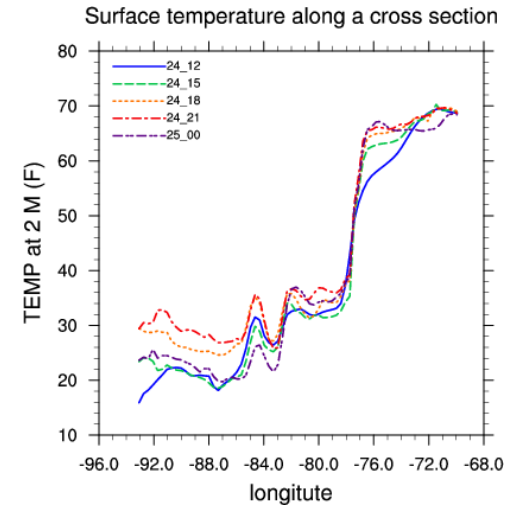
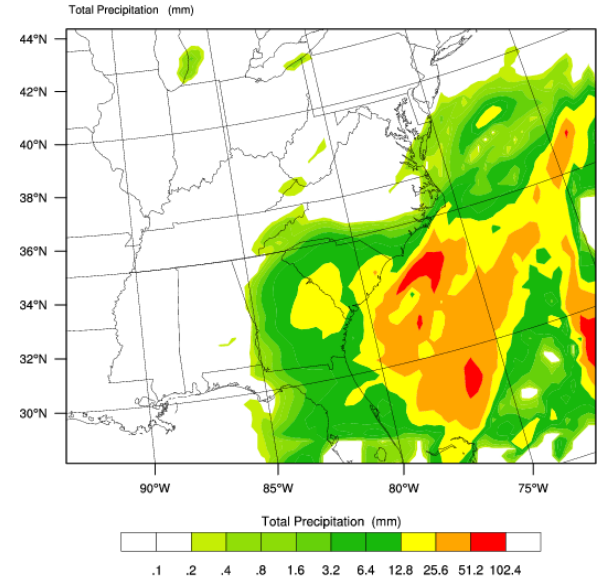
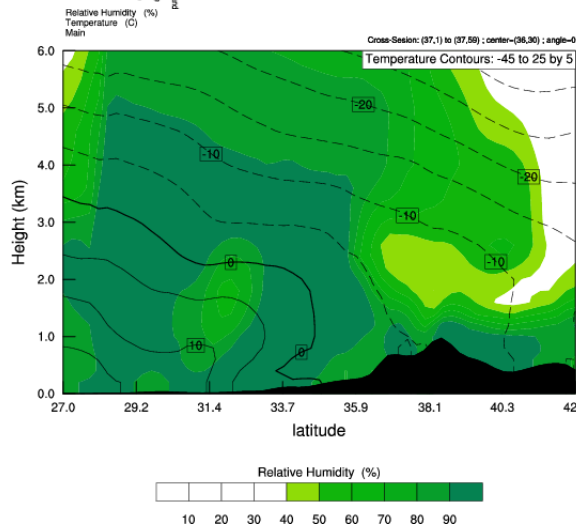
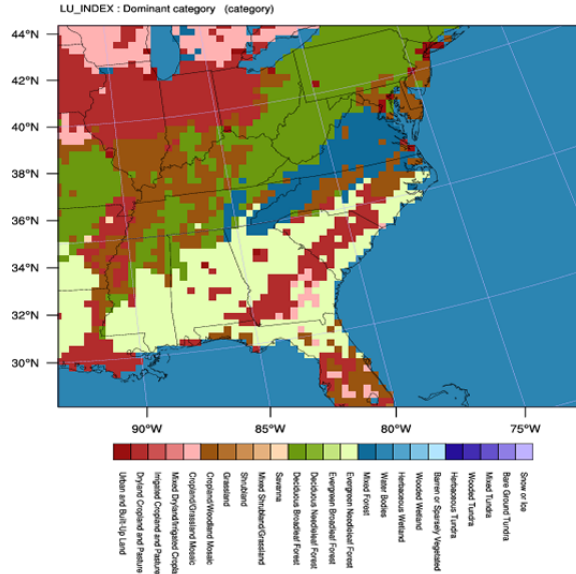
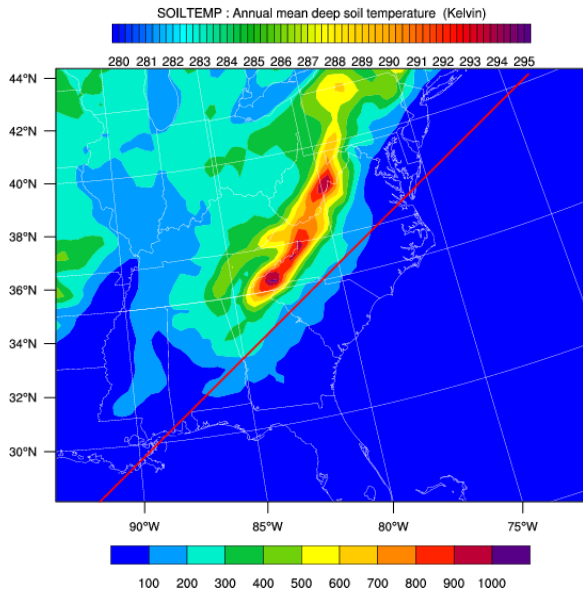
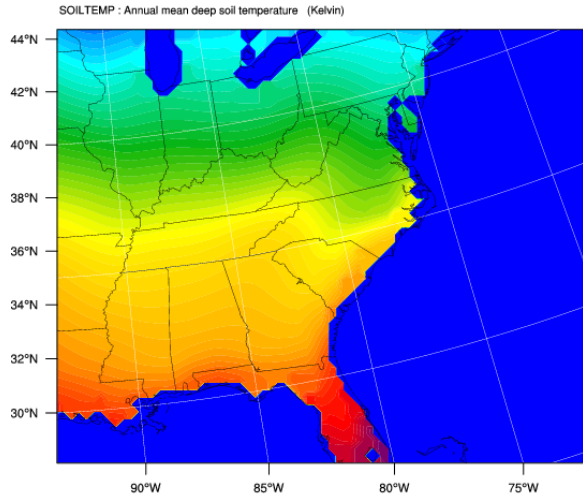
NCL

- **N**CAR **C**ommand **L**anguage
- <http://www.ncl.ucar.edu>
- Read WRF_ARW data directly
- Generate a number of graphical plots
 - *Horizontal, cross-section, skewT, meteogram, panel*
- Download
 - <http://www.ncl.ucar.edu/Download>
 - *Fill out short registration form (there is a short waiting period)*
 - *Read and agree to OSI-based license*
 - *Get version 6 or **later** (current v6.1.2 – highly recommended)*
- NCARG_ROOT environment variable
 - *setenv NCARG_ROOT /usr/local/ncl*

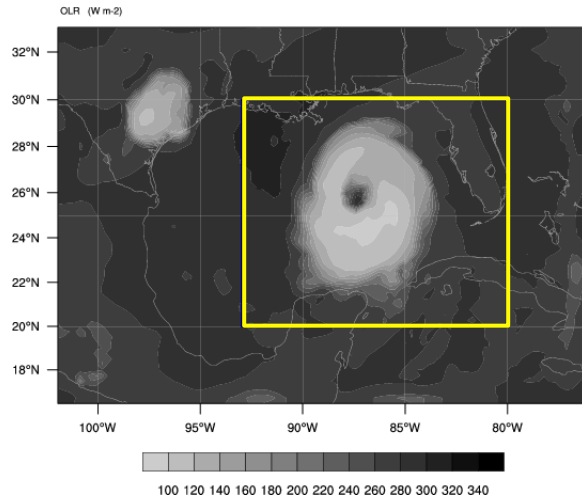
NCL

- What it's good for:
 - *Spatial visualization of climate data*
 - *Data manipulation*
 - *Large data files*
 - *Link Fortran / C programs directly to NCL scripts*
- Limitation
 - *Statistical tools (some are available, but tools like R are better)*

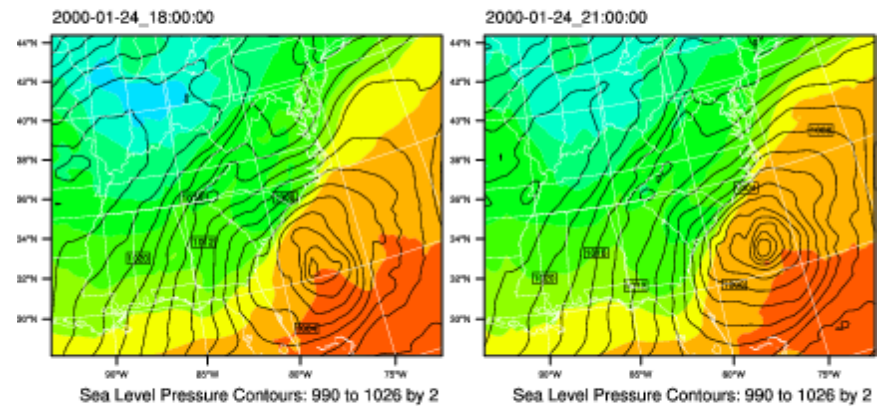
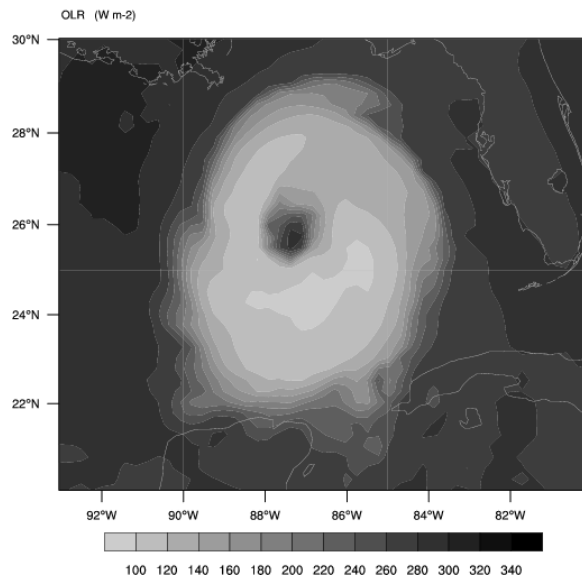
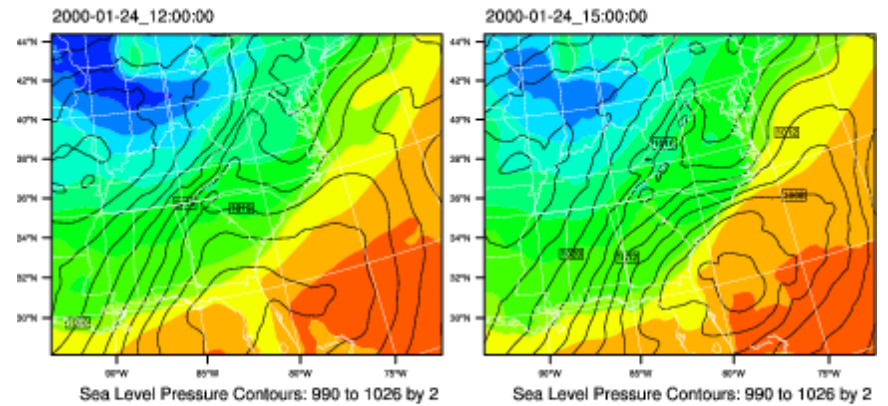
Examples



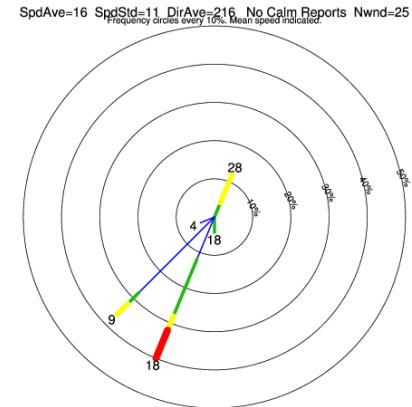
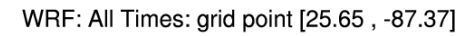
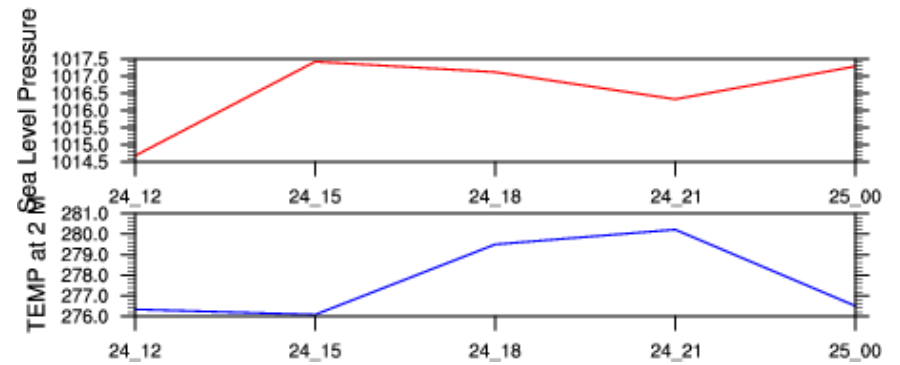
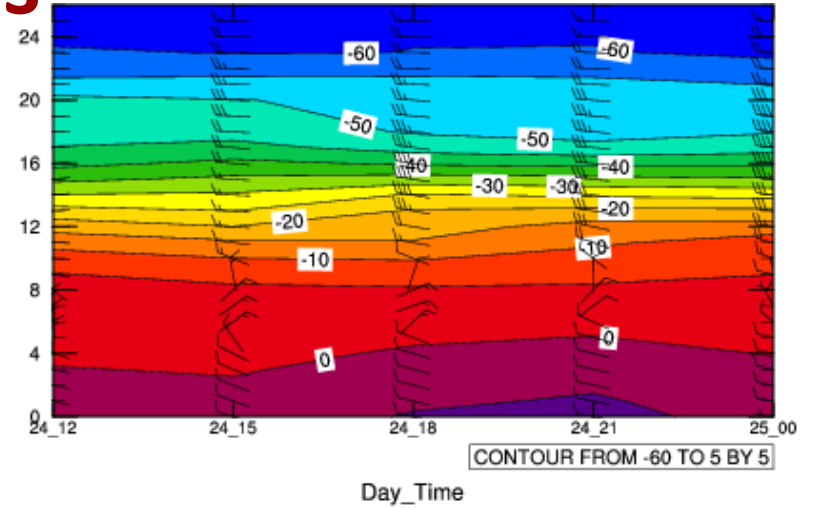
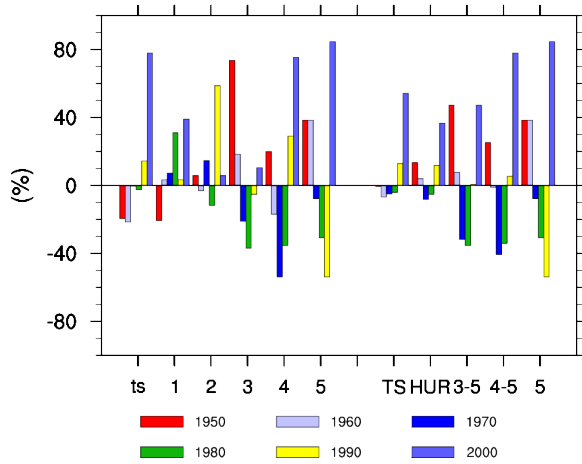
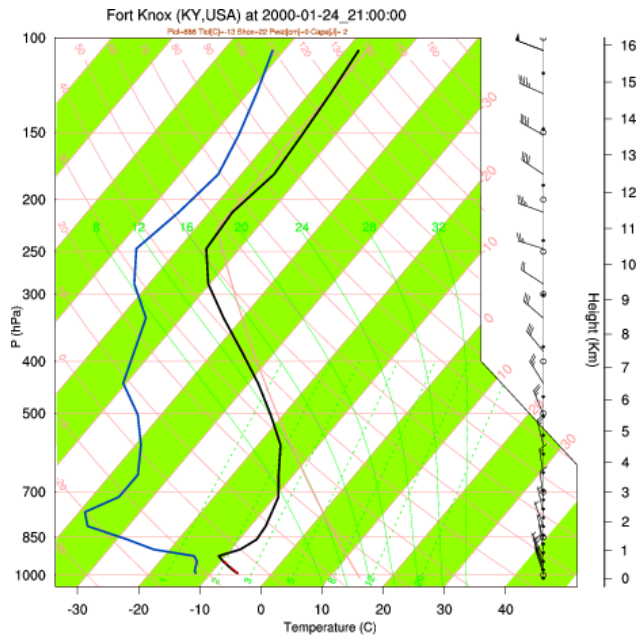
Examples



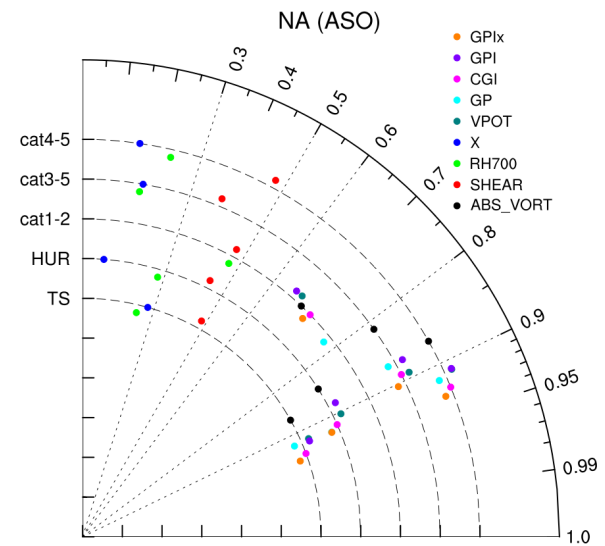
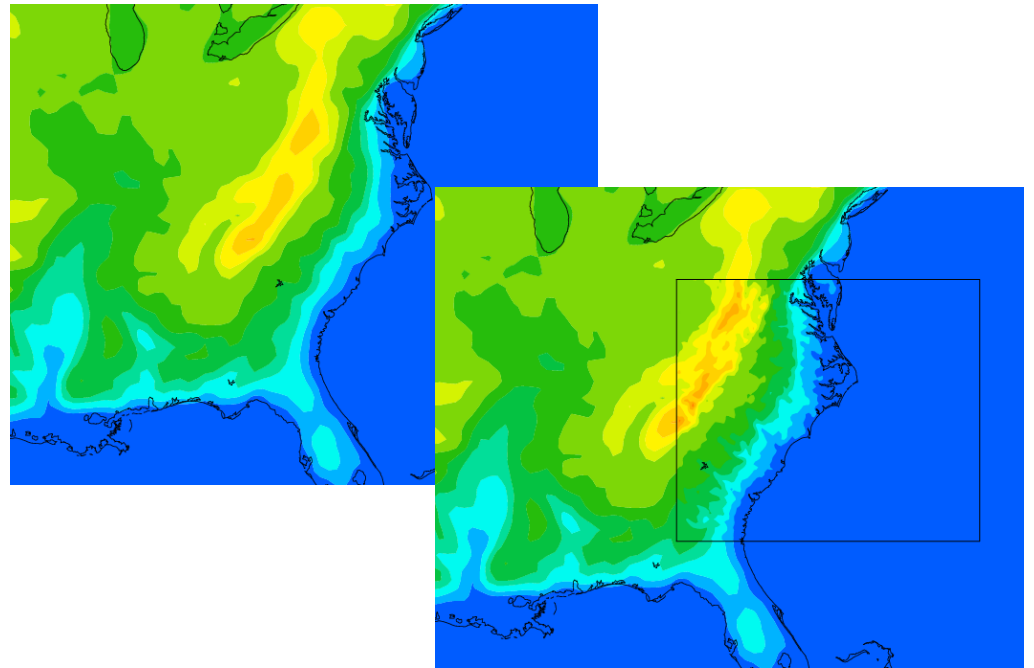
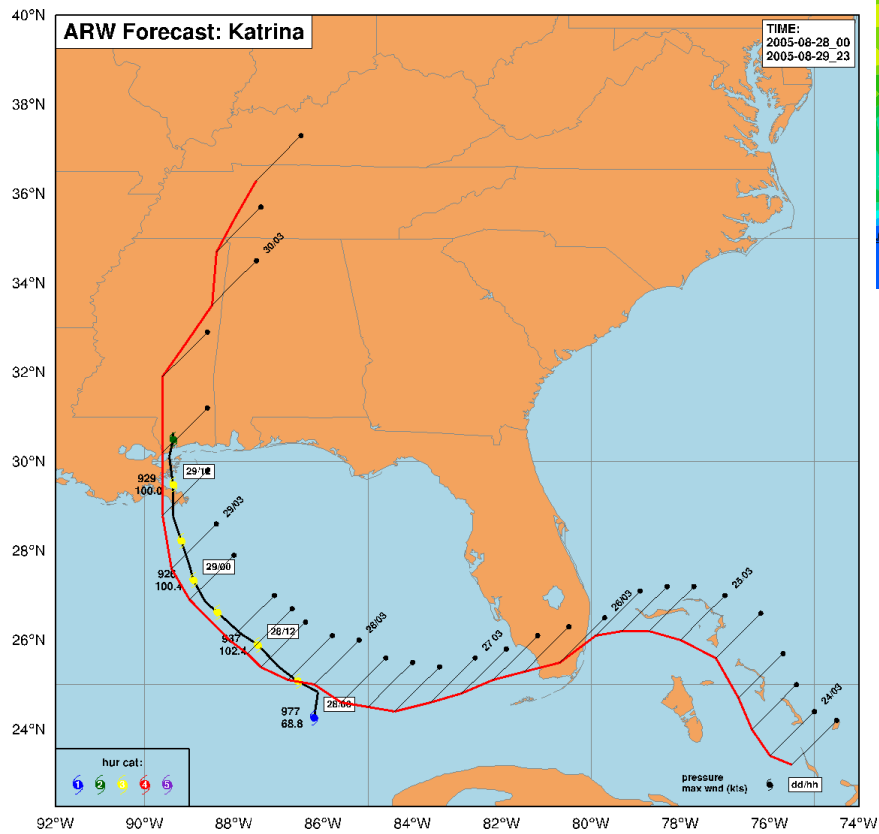
TEMP at 2 M (K)



Examples



Examples



Change fields in netCDF file

```
load "$NCARG_ROOT/lib/ncarg/nclscripts/csm/gsn_code.ncl"
```

```
Begin
```

```
  a = addfile("./met_em.d01.2000-01-24_12:00:00.nc","w")
```

```
  sst = a->SST          ; read a field
```

```
  sst = sst + 10        ; change the entire field
```

```
  a->SST = sst           ; write the field
```

```
  var= a->LANDUSE
```

```
  var(0,62,84) = 7      ; change a single point
```

```
  a->LANDUSE = var
```

```
end
```

Output ASCII Data

```
t2_point = a->T2(:,locY,locX)
q2_point = a->Q2(:,locY,locX)
```

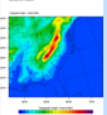
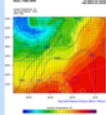
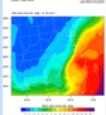
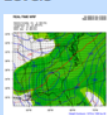
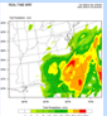
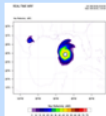
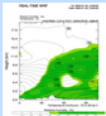
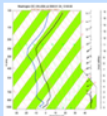
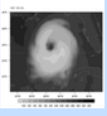
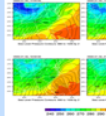
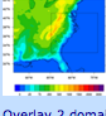
`write_matrix`

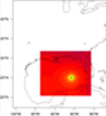
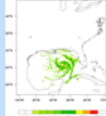
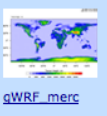
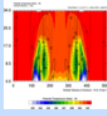
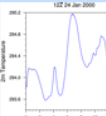
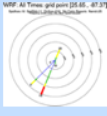
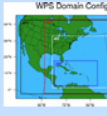
```
asciwrite ("t2.txt" , t2_point)
asciwrite ("q2.txt" , sprintf("%9.3f", q2_point))
```

```
npts = dimsizes(t2_point)
data = new( npts, "string")
  do npt=0,npts-1
    data(npt) = sprintf("%7.1f ", t2_point(npt))
    data(npt) = data(npt) + sprintf("%7.1f ", q2_point(npt))
  end do
asciwrite ("t2_q2.txt", data)
```

Generate Plots: *A good start - OnLine Tutorial*

<http://www.mmm.ucar.edu/wrf/OnLineTutorial/Graphics/NCL/index.html>

Basic Plots  Basic Plot Setup <i>(This series of examples takes users through same basic steps in generating plotting scripts.)</i> Get and plot a single field Multiple input files Plot all fields in a file	Basic Surface Plots  Surface 1 Surface 3 Surface 2 Surface with multiple input files	Plots on Model Levels  Clouds Levels from wrfout files Levels from metgrid files	Plots on Interpolated Levels  Height Levels Pressure Levels
Plotting Precipitation  Precipitation	Diagnostics  CAPE dBZ PW Vorticity <i>(More diagnostics are available, shown are only some newer/special diagnostics)</i>	Cross-section Plots  Height - Through a Pivot Point Height - Point A to Point B Pressure Limited Vertical Extent For 2D fields Smooth terrain	Skew_T Plots  Skew_T
Overlay and Zoom  Overlay Zoom Overlay & Zoom	Panel Plots  Panel 1 Panel 2	Overlay Domains  Overlay 2 domains	

Moving Nest  Moving Nest <i>Keeping a fixed background for moving nest domain</i>	Moving Nest  Precip Tendencies for a Moving Nest <i>Keeping a fixed background for moving nest domain, and plotting rainfall tendencies in the overlapping regions</i> Domain 1 and 2 Precip Tendencies on a single plot <i>(domain 2 is a moving nest)</i>	Global WRF  qWRF merc	Idealized cases  wrf_Grav2x wrf_Hill2d wrf_Squall_2d_x wrf_Squall_2d_y wrf_Seabreeze2x wrf_BWave wrf_QSS
Time Slice Plots  Meteoqrgrams WRF Time Series data	Track Cyclone  Plot Cyclone Vortex	Wind Roses  Create a wind rose	Preview Domain  Preview <i>Note: This example script makes use to the WPS namelist directly</i>

Resources

- **The special WRF functions have unique resources:**

http://www.mmm.ucar.edu/wrf/OnLineTutorial/Graphics/NCL/NCL_functions.htm

- **All general NCL resources can also be used to control the plot:**

<http://www.ncl.ucar.edu/Document/Graphics/Resources>

- | | |
|-----------------------------------|---------------------------------|
| • am (annotation manager) | • sf (scalar field) |
| • app (app) | • st (streamline) |
| • ca (coordinate array) | • tf (transformation) |
| • cn (contour) | • ti (title) |
| • ct (coordinate array table) | • tm (tickmark) |
| • dc (data comm) | • tr (irregular transformation) |
| • err (error) | • tx (text) |
| • gs (graphic style) | • vc (vectors) |
| • gsn (gsn high-level interfaces) | • vf (vector field) |
| • lb (label bar) | • vp (view port) |
| • lg (legends) | • wk (workstation) |
| • mp (maps) | • ws (workspace) |
| • pm (plot manager) | • xy (xy plots) |
| • pr (primitives) | |

NCO tools

<http://nco.sourceforge.net/>

- **ncdiff**
 - *Difference two file*
`ncdiff input1.nc input2.nc -o output.nc`
- **ncrcat (nc cat)**
 - *Write specified variables / times to a new file*
`ncrcat -v RAINNC wrfout* -o RAINNC.nc`
`ncrcat -d Time,0,231 -v RAINNC wrfout* -o RAINNC.nc`
- **ncra (nc average)**
 - *Average variables and write to a new file*
`ncra -v OLR wrfout* -o OLR.nc`
- **ncks (nc kitchen sink)**
 - *Combination of NCO tools all in one (handy: one tool for multiple operations)*
Specifically handy to split files
`ncks -d Time,1,1 wrfout -o wrfout1.nc`

ncview

http://meteora.ucsd.edu/~pierce/ncview_home_page.html

no variable selected

Ncview 1.93a David W. Pierce 1 Feb 2006

*** SELECT A VARIABLE TO START ***

Quit

-->|

◀

▶

⏏

▶▶

Edit

?

Delay:

Opts

3gauss

Inv P

Inv C

Mag X1

Linear

Axes

Range

blowup

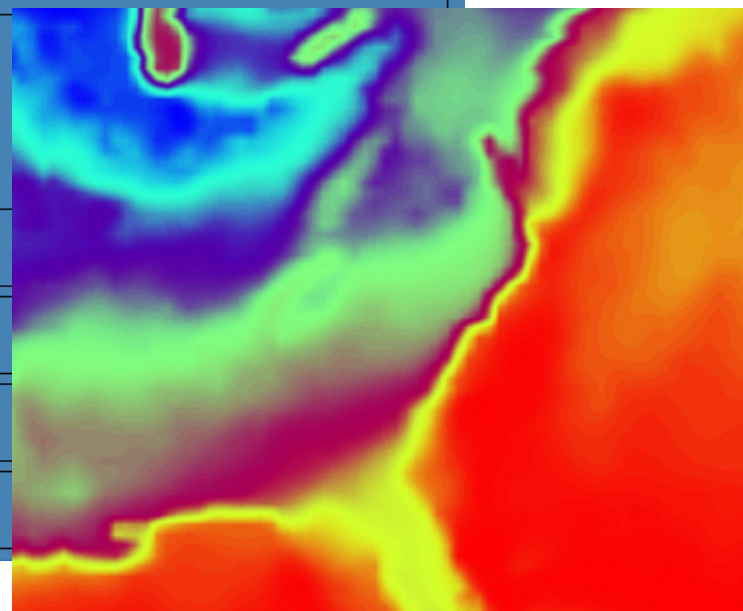
Print

(18) 1d vars

(48) 2d vars

(13) 3d vars

Dim:	Name:	Min:	Current:	Max:	Units:
	Time	Min:		Max:	Units:
	bottom_top	Min:		Max:	Units:
	south_north	Min:		Max:	Units:
	west_east_st	Min:		Max:	Units:



R

- A freely available “language and environment for statistical computing and graphics”
- Widely used, with a global support community
- Fully extensible using different libraries/packages
 - *“package” is a plugin with additional functions*
 - *Usually comes with online help*
 - *Sometimes has a “vignette” of example uses*
- Wealth of tutorials, books, blogs, example code
- Not so great with large data sets (use NCL!)

Framing the Question

“To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.”

- R.A. Fisher

- Explore your data
- Design a realistic and sensible hypothesis
- Build statistical models to test it
- Choose an appropriate significance test



Exploratory Analysis

- Read in the data... NetCDF, .csv, .dat,
- Look at it:

```
> datafile
```

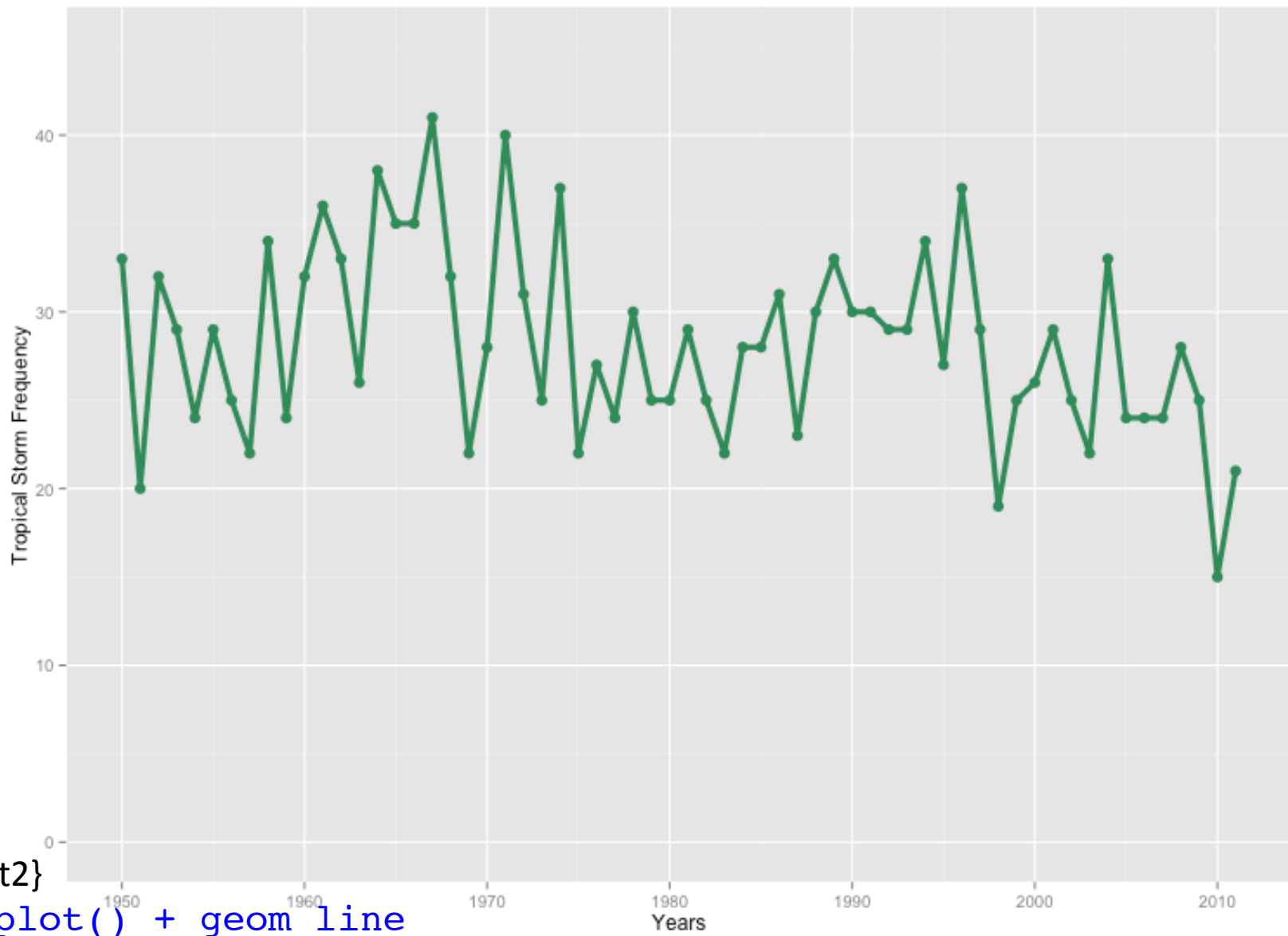
	YEAR	ts	c1	c2	c3	c4	SST	Rel_SST	ABS_VORT	SHEAR
1	1950	33	14	8	6	5	26.47666	1.423286	4.47e-05	0.1155347
2	1951	20	17	6	11	7	26.74715	1.397581	4.61e-05	0.1141776
3	1952	32	21	9	12	8	26.43839	1.134402	4.58e-05	0.1085700
4	1953	29	17	2	15	9	26.58802	1.293598	4.55e-05	0.1015015
5	1954	24	15	4	11	8	26.64362	1.603314	4.55e-05	0.1207132
6	1955	29	20	10	10	7	26.09393	1.058517	4.60e-05	0.1091396

- Summarise it:

```
> summary(datafile)
```

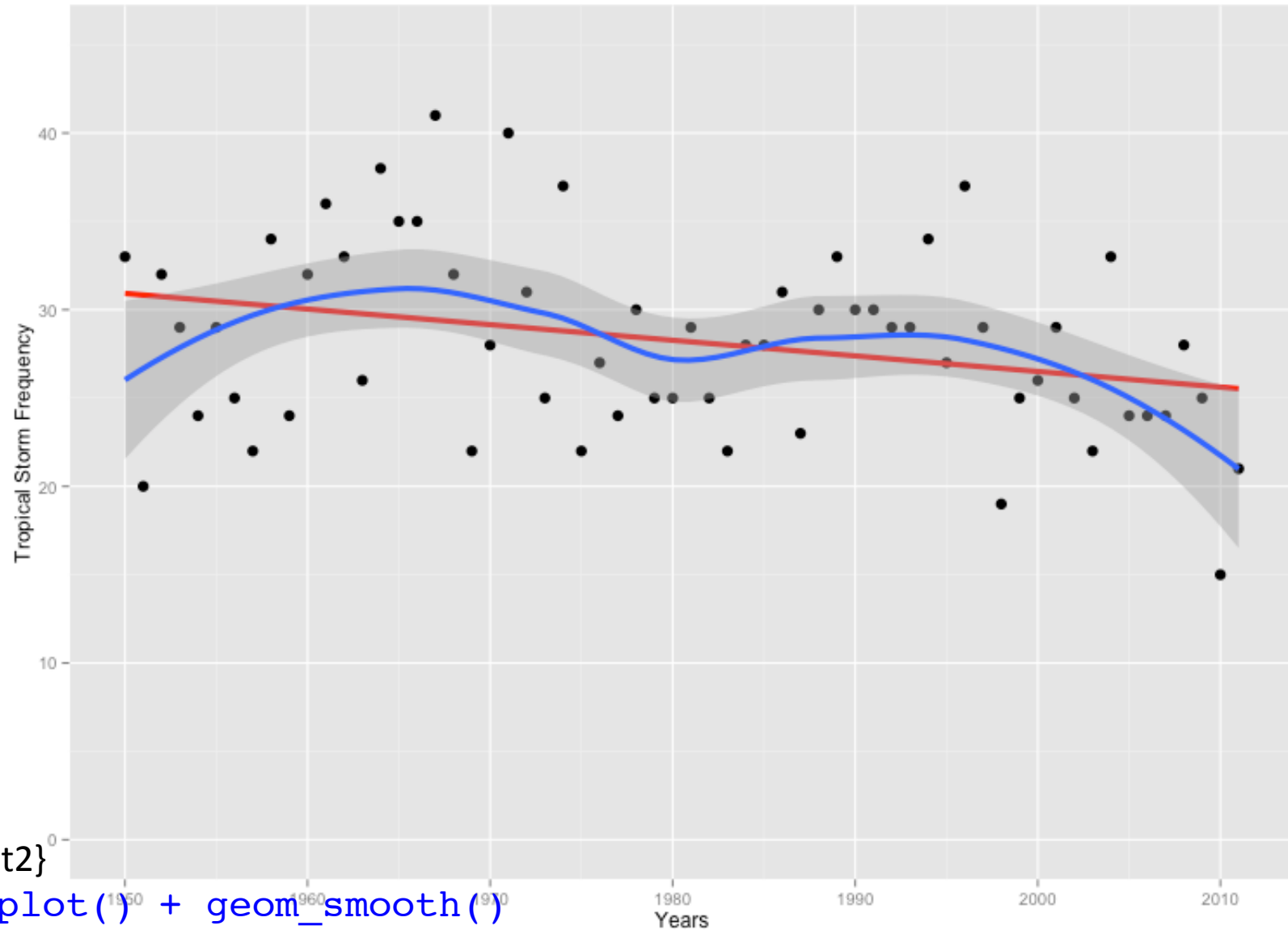
YEAR	ts	c1	c2
Min. :1950	Min. :15.00	Min. : 9.00	Min. : 1.000
1st Qu.:1965	1st Qu.:24.25	1st Qu.:15.00	1st Qu.: 6.000
Median :1980	Median :28.00	Median :17.00	Median : 8.000
Mean :1980	Mean :28.23	Mean :17.55	Mean : 7.823
3rd Qu.:1996	3rd Qu.:32.00	3rd Qu.:20.75	3rd Qu.:10.000
Max. :2011	Max. :41.00	Max. :26.00	Max. :14.000

Plot the time series



```
{ggplot2}  
> ggplot() + geom_line
```

... add some simple trend lines

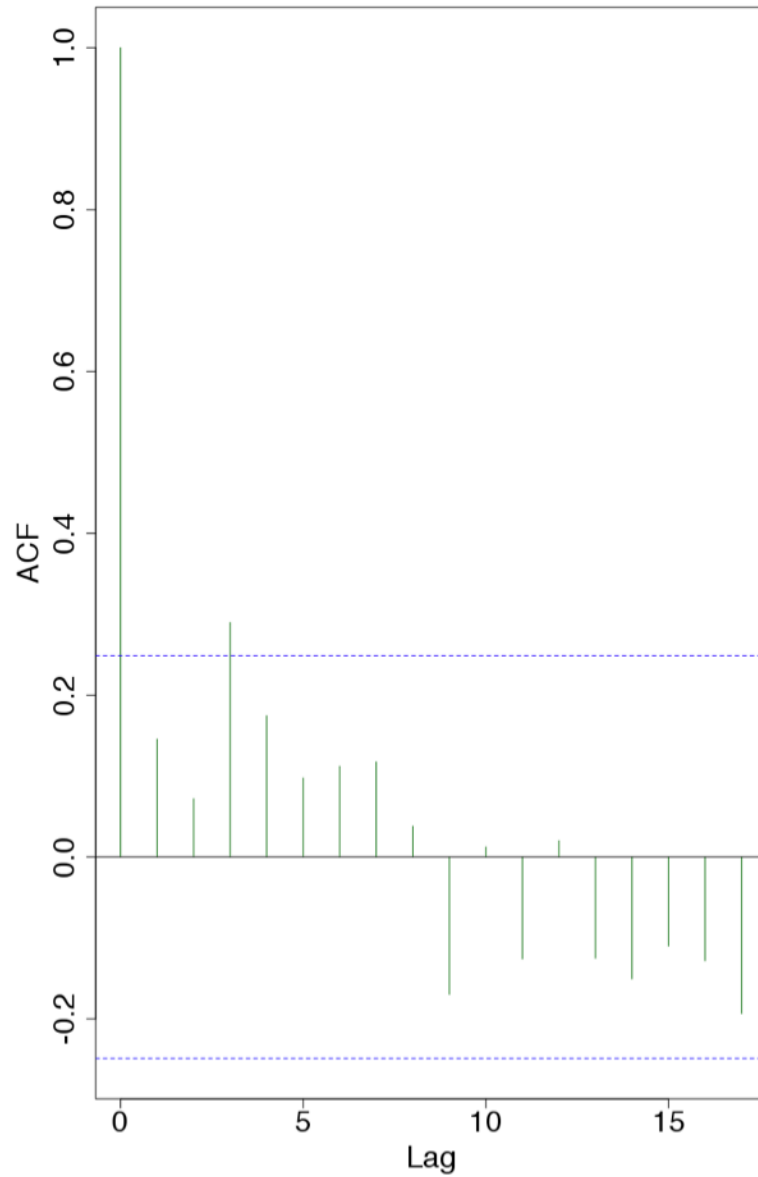
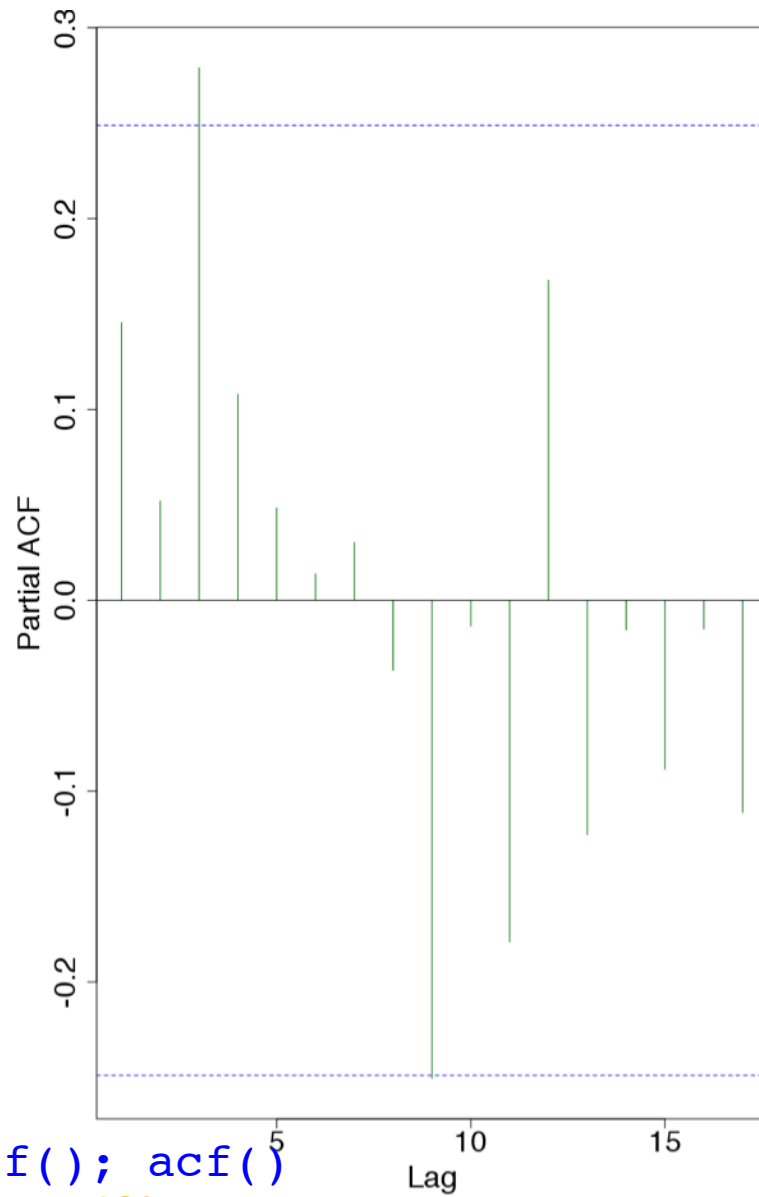


```
{ggplot2}
```

```
> ggplot() + geom_smooth()
```

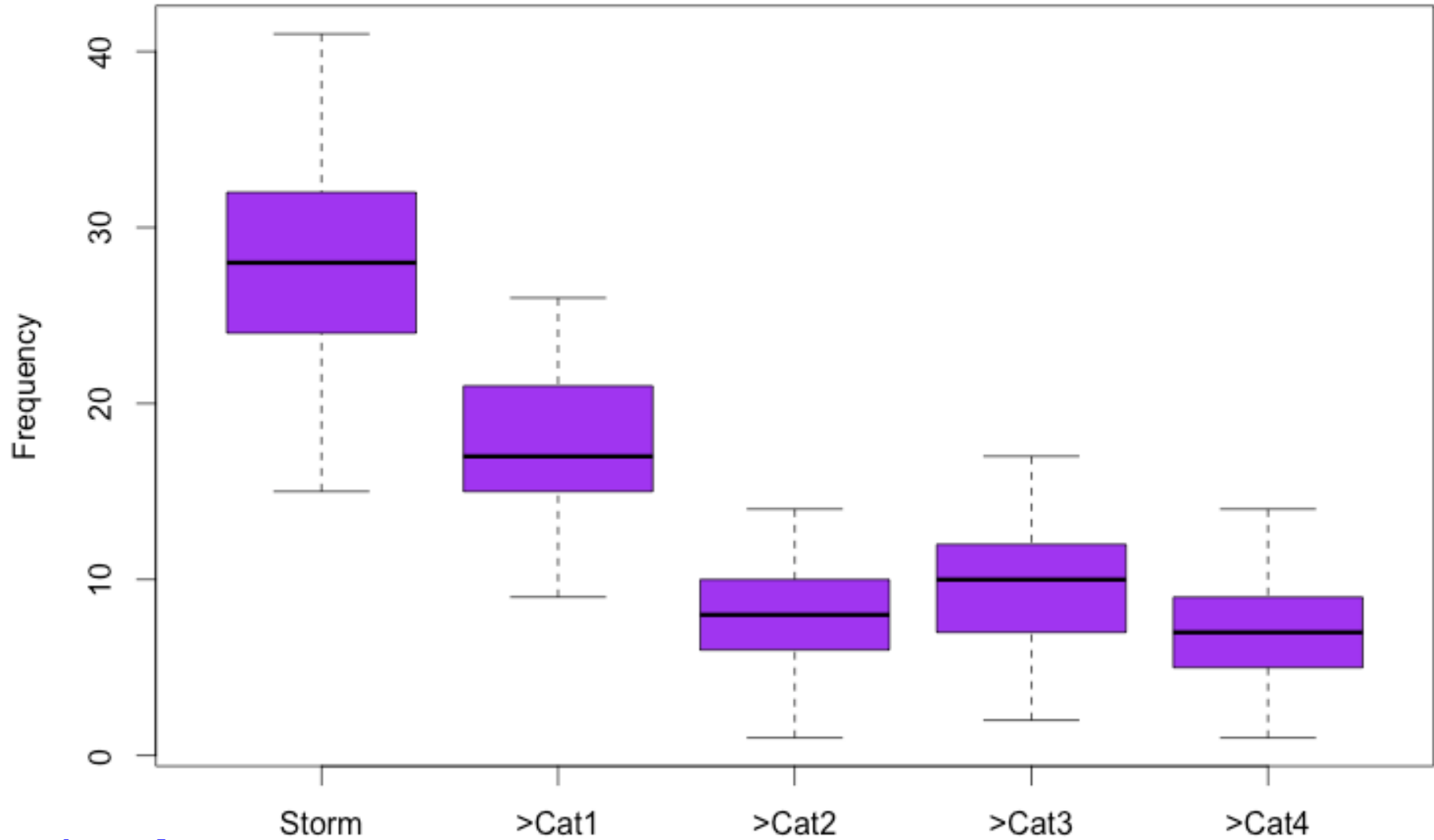
```
> lm()
```

Check Autocorrelation



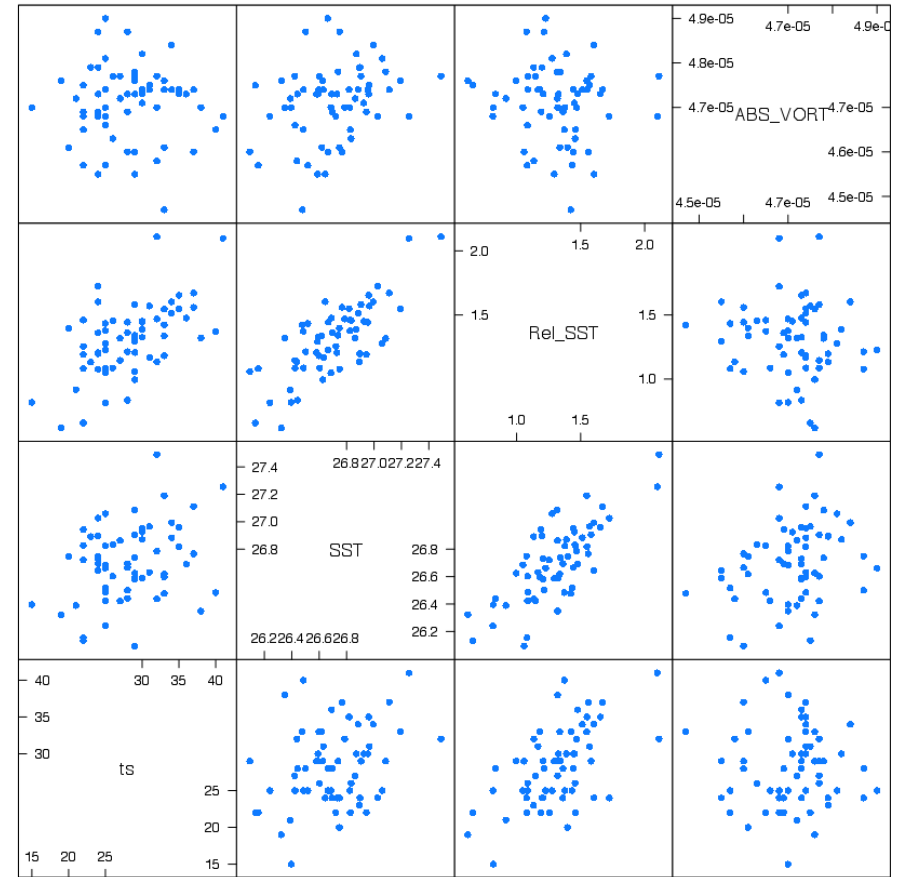
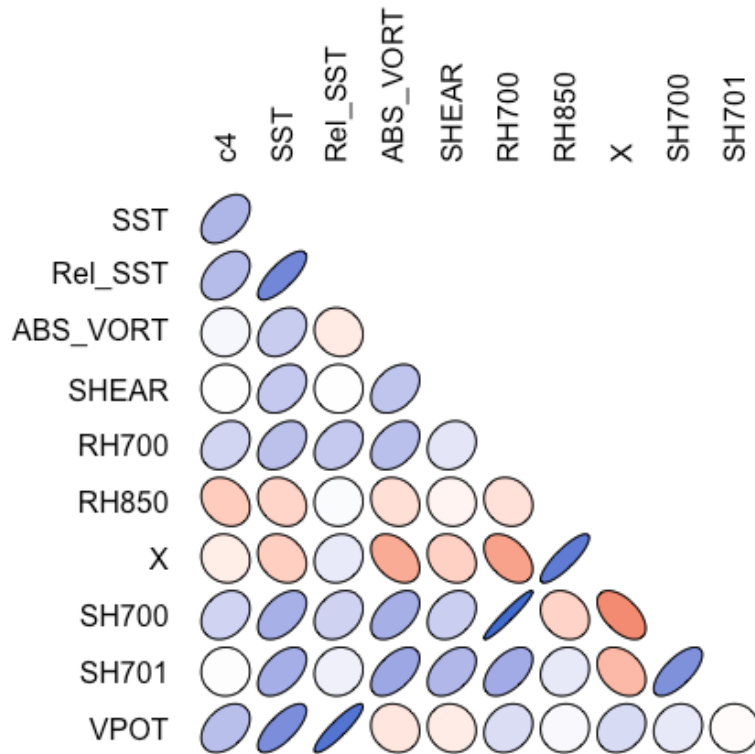
```
> pacf(); acf()
```

Check the variability in the data



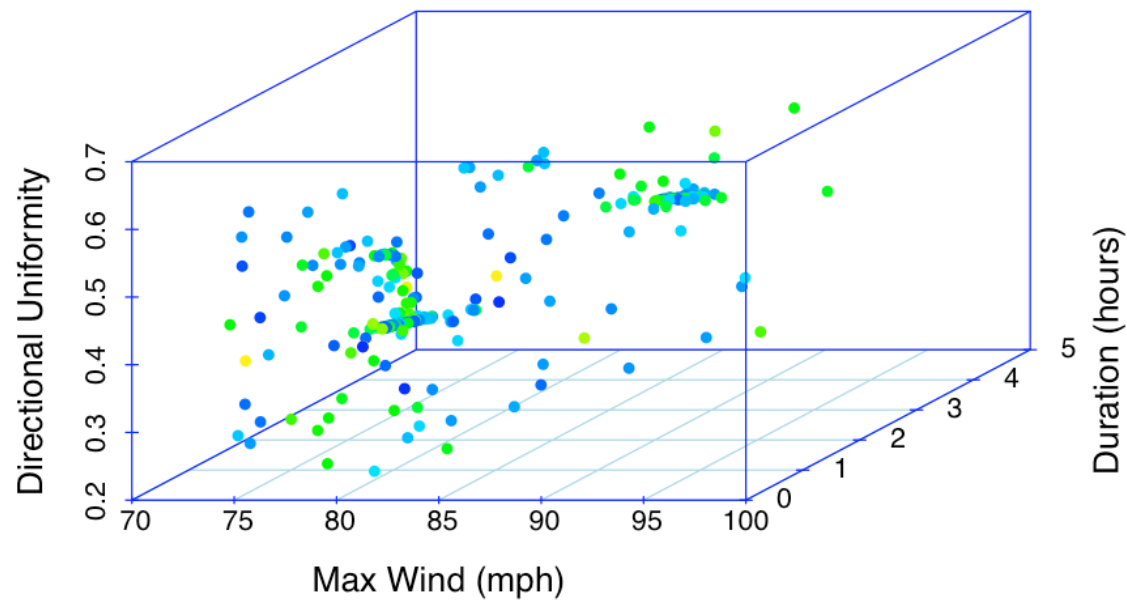
```
> boxplot()  
> var(); quantile(); mean()
```

Examine correlations



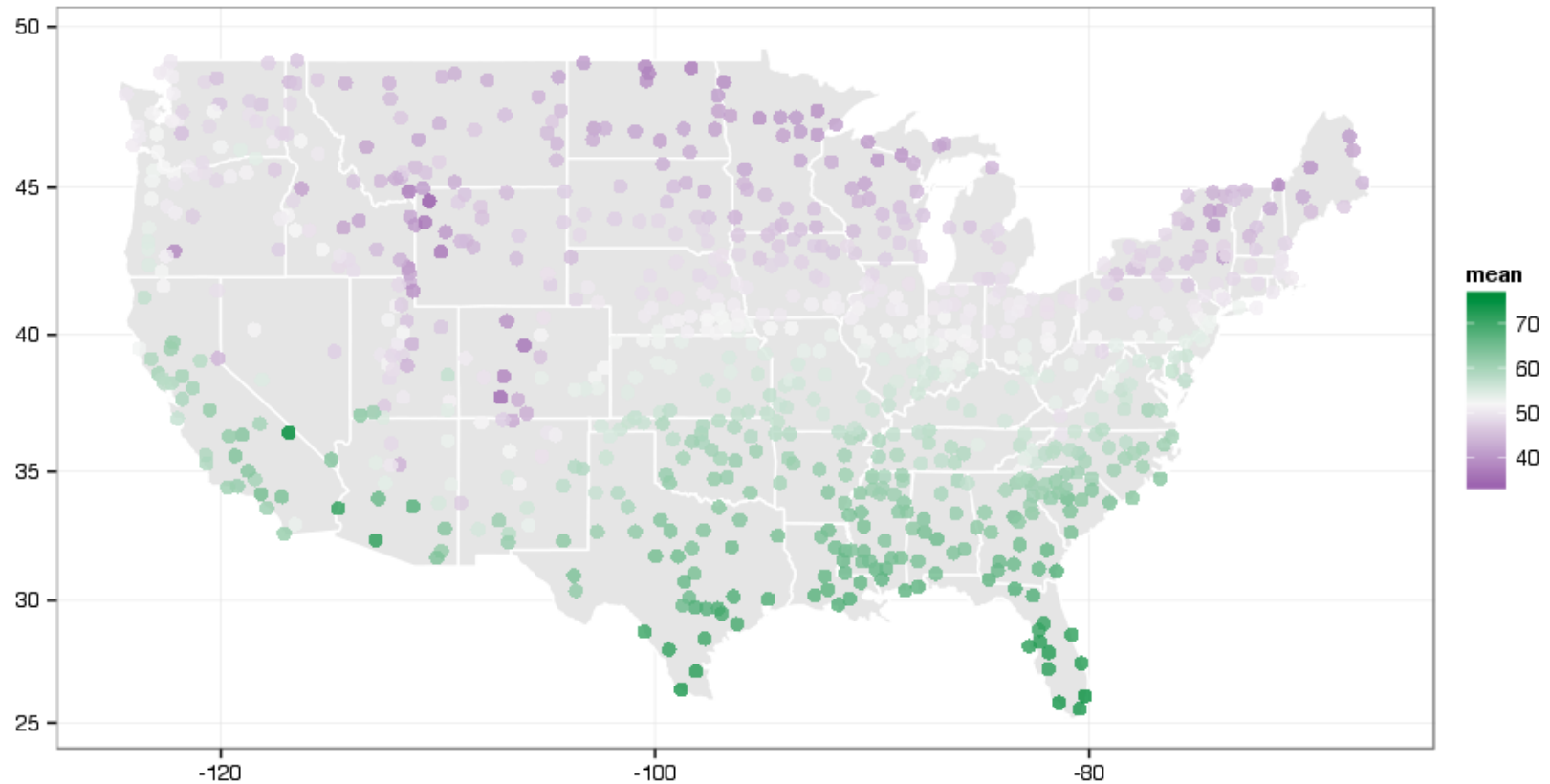
```
{ellipse}
> plotcorr()
{lattice}
> splom()
> aov(); acov; corr()
```

What's important in the data?



```
{scatterplot3d}  
> scatterplot3d()
```

Examine spatial patterns

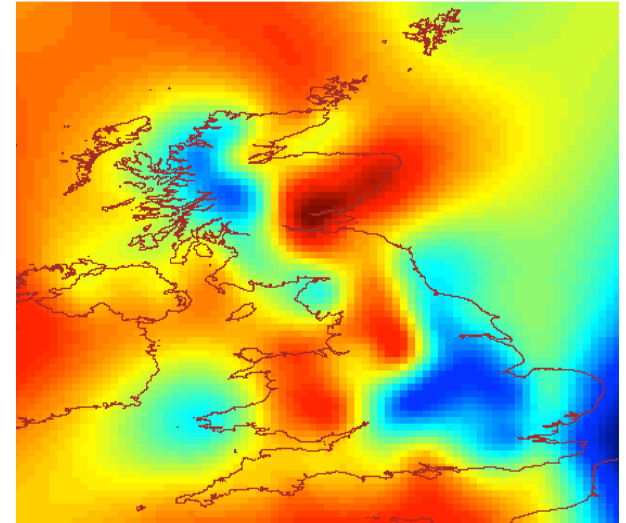
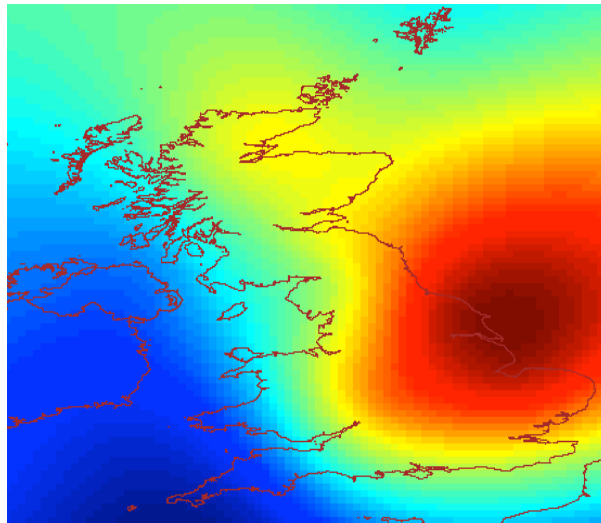
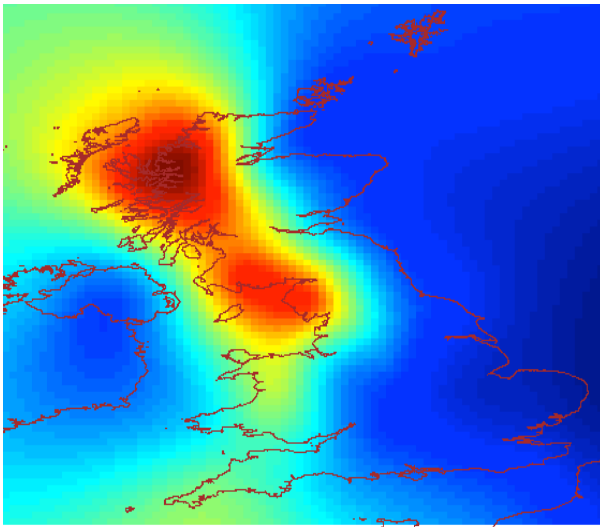


```
{ggplot2}  
> ggplot() + geom_polygon()  
{lubridate}  
> ymd(); month(); year()
```

Recap

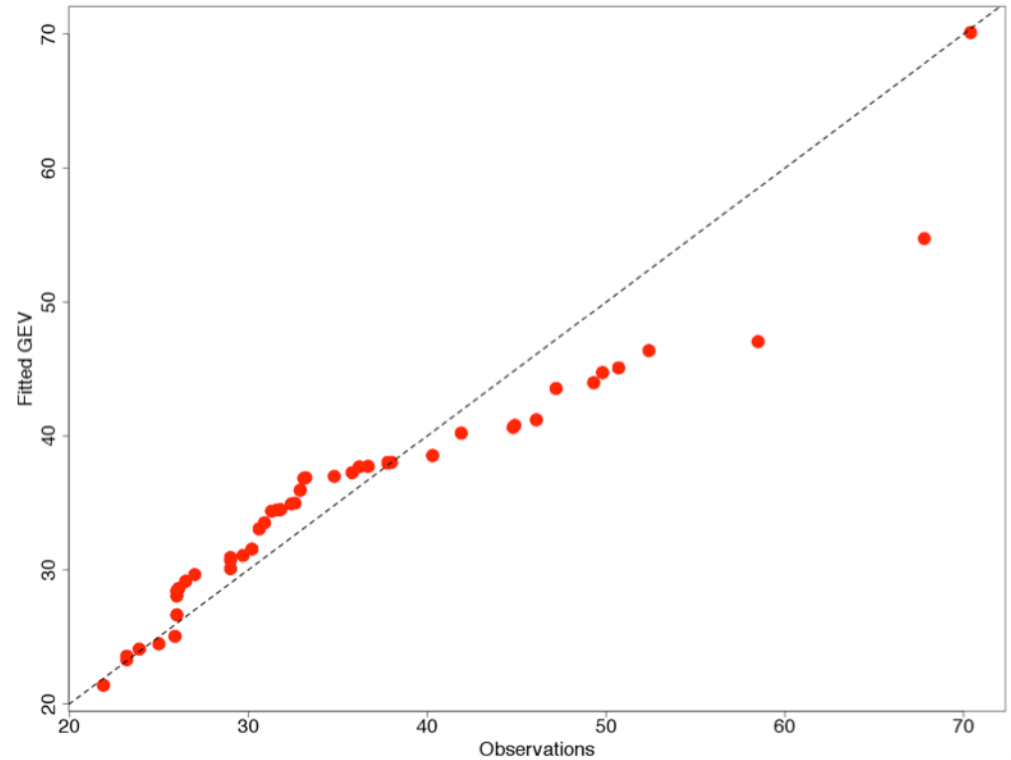
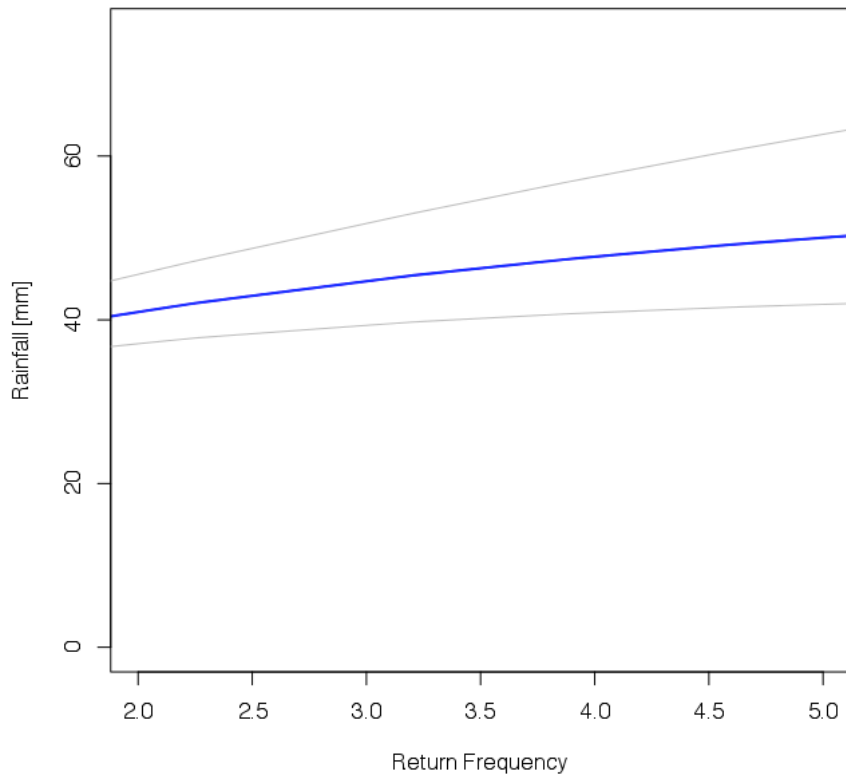
- Explore your data
- Design a realistic and sensible hypothesis
- **Build statistical models to test it**
- **Choose an appropriate significance test**
 - *Base packages include t-test, F-test, chi-squared*
 - *Fitted model outputs include standard error estimates*
 - *Most model outputs give relative statistics e.g. AIC*
 - *Visual tests (Q-Q plots) are just as important as statistical tests!*

Identify Principal Components



```
> princomp()  
{fields}  
> Tps(); surface()
```

Fit Extreme Value Distributions

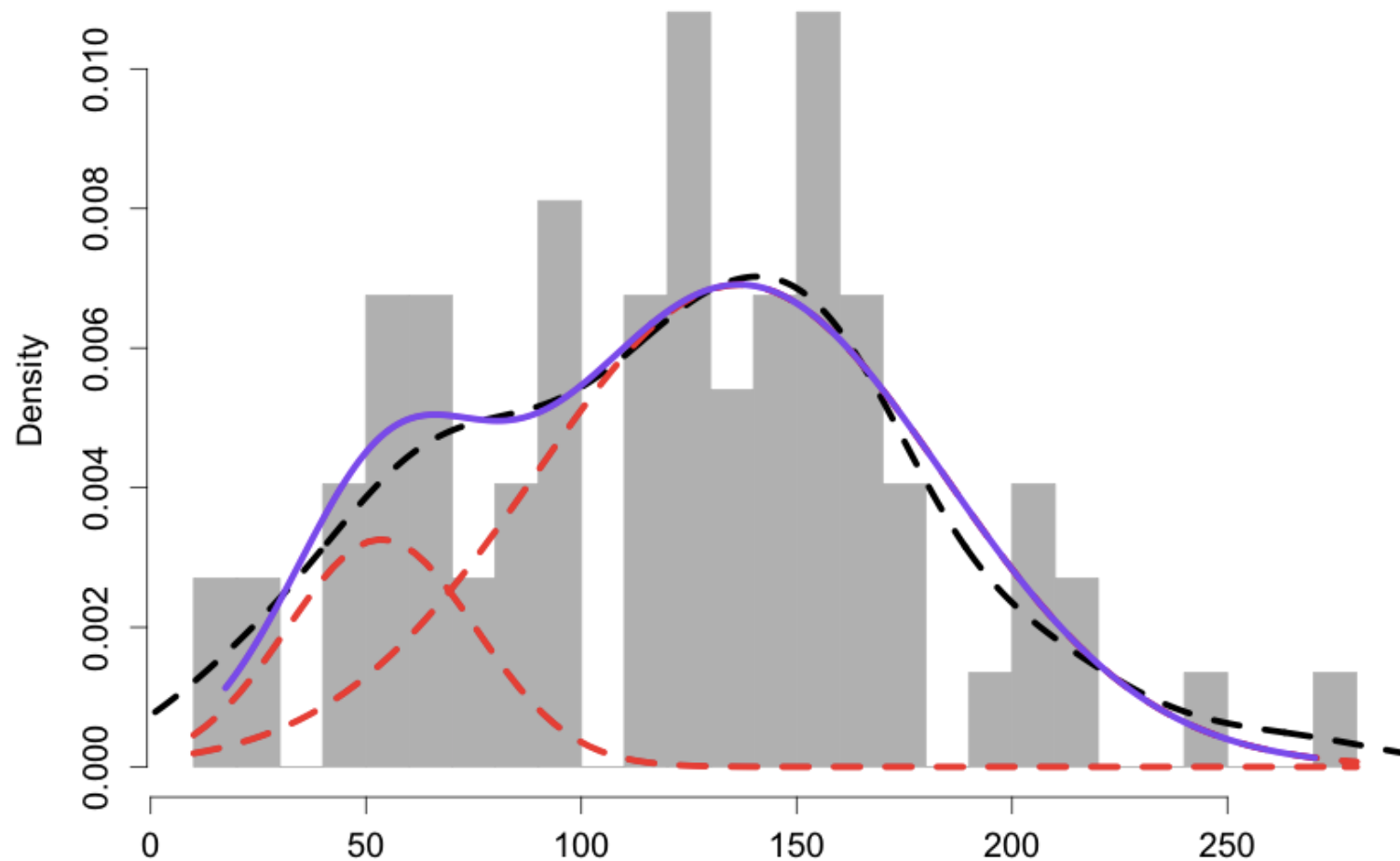


```
{extRemes}
```

```
> gev.fit()
```

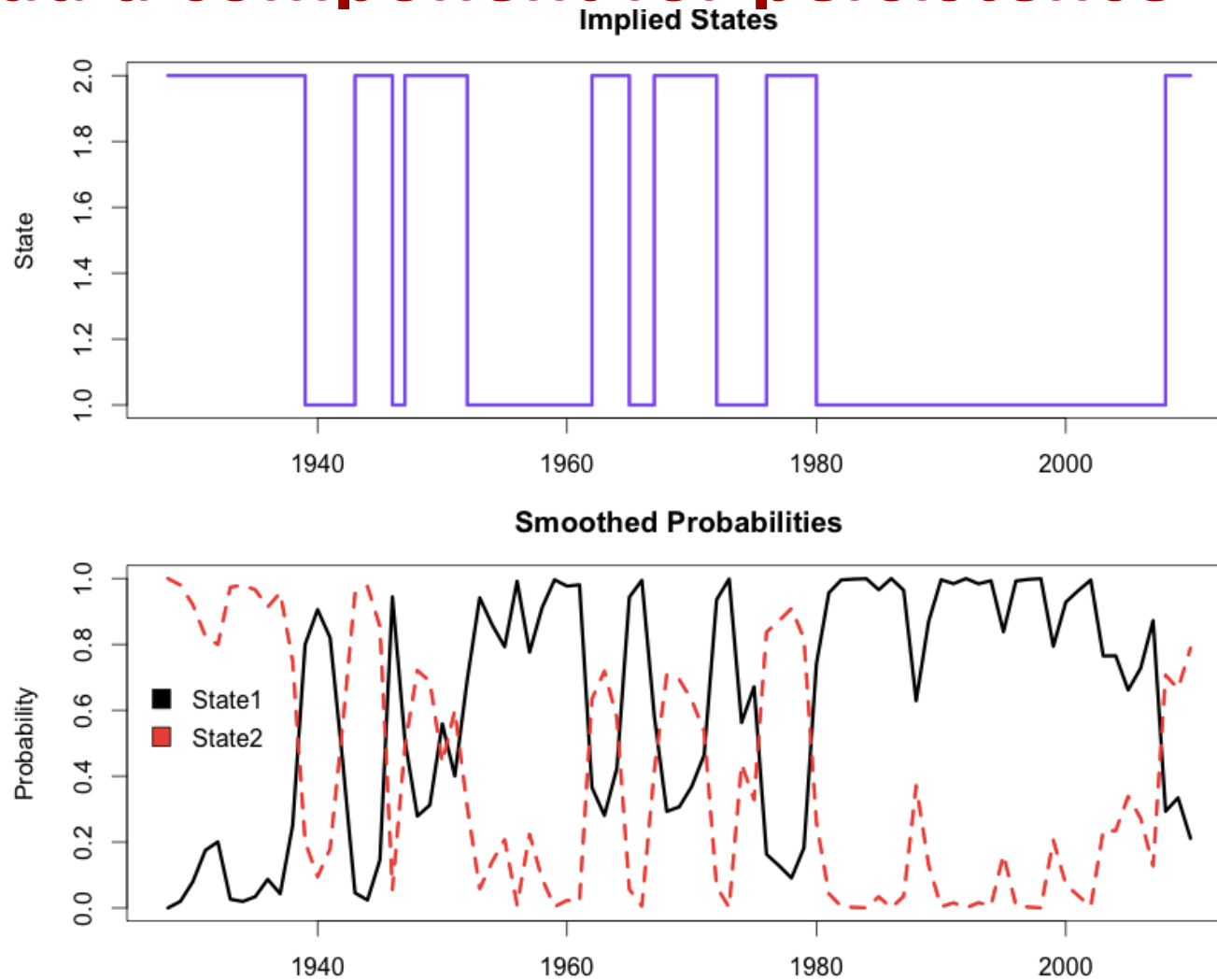
```
> plot(); qqplot()
```

Fit Mixture Models



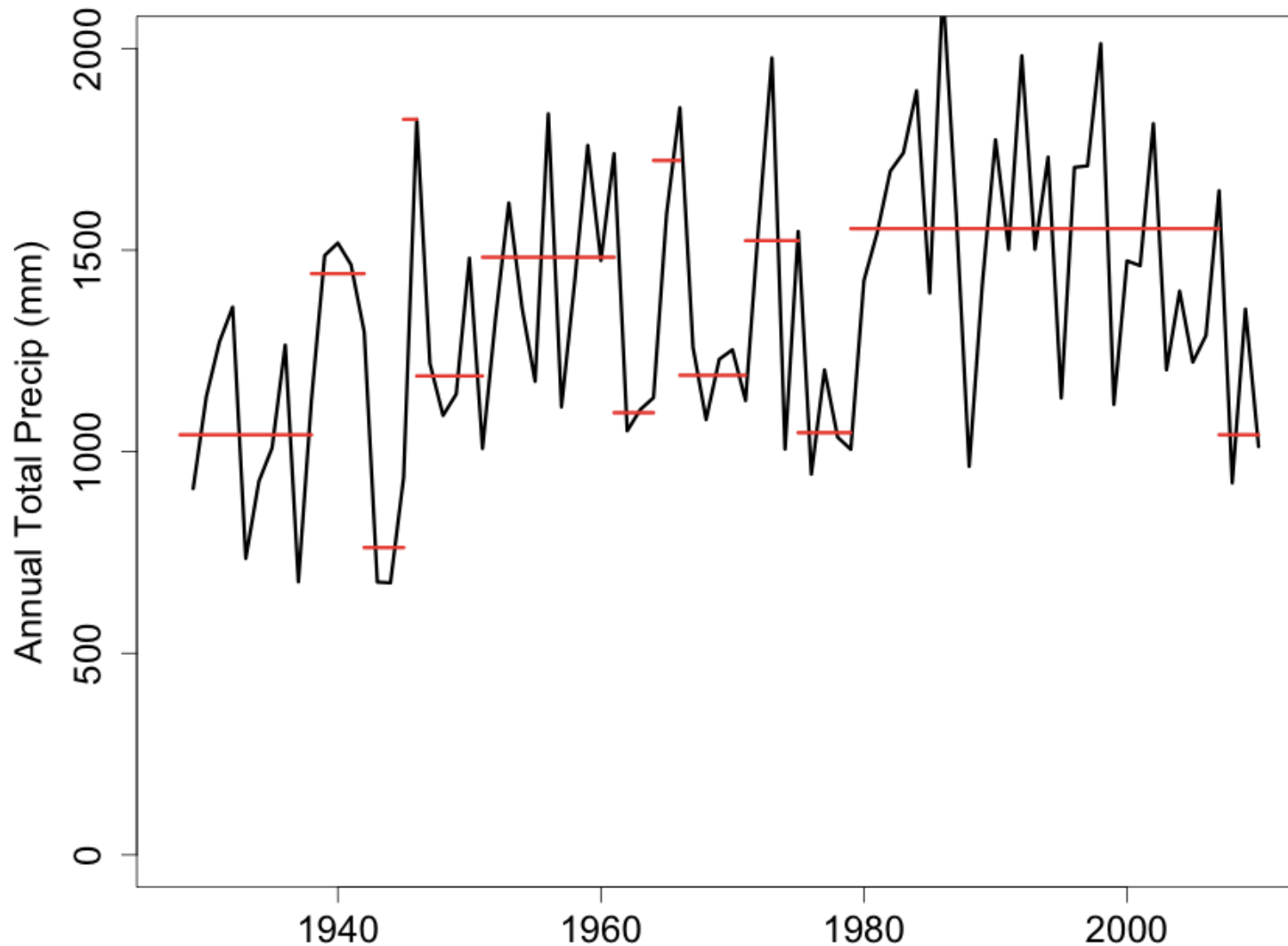
```
{mixtools}  
> normalmixEM()  
> AIC(); BIC(); loglikn()  
> pchisq()
```

add a component for persistence



```
{depmixS4}  
> depmix(); posterior(); forwardbackward()  
> matplot()
```

... to examine multi-annual cycles



```
> diff()  
> plot() + segments()
```

[illegible]

Getting started with R

<http://cran.r-project.org> Download and install R

<http://www.rstudio.com> a nice interface for scripting

<http://cran.r-project.org/web/packages/> the packages

<http://www.statmethods.net> An introduction

<http://www.cookbook-r.com> Another introduction

<http://a-little-book-of-r-for-time-series.readthedocs.org> And another one for time series. Also for multivariate analysis

<http://cran.r-project.org/doc/contrib/Short-refcard.pdf> A handy listing of commonly used commands. Lots of other useful references in “contrib”

<http://www.r-project.org/conferences.html> useR! Annual conferences with video presentations

<http://pairach.com/2012/02/26/r-tutorials-from-universities-around-the-world/> various tutorials

For those @#\$*&! moments

<http://rseek.org> An R-centric search engine

<http://stats.stackexchange.com> Q&A site for stats and R

www.stackoverflow.com Q&A site for many different computer languages

<http://rwiki.sciviews.org/doku.php> the R Wiki

<http://www.r-bloggers.com> a compilation of blog links

http://www.burns-stat.com/pages/Tutor/R_inferno.pdf

Because others share your pain!

Useful R Packages

cluster	extended cluster analysis
extRemes	extreme value toolkit
ggplot2	graphics package for beautiful charts
lubridate	for manipulating time series
maps/maptools	includes interface for ESRI shapefiles
mgcv	generalized linear and additive modeling
ncdf	read and write NetCDF files in R
plyr	for splitting or merging data
RColorBrewer	different colour palettes
reshape2	rearrange data tables
sp	interface for spatial data
sqldf	perform SQL queries in R
stringr	string related manipulation
wordcloud	create wordclouds for abstracts