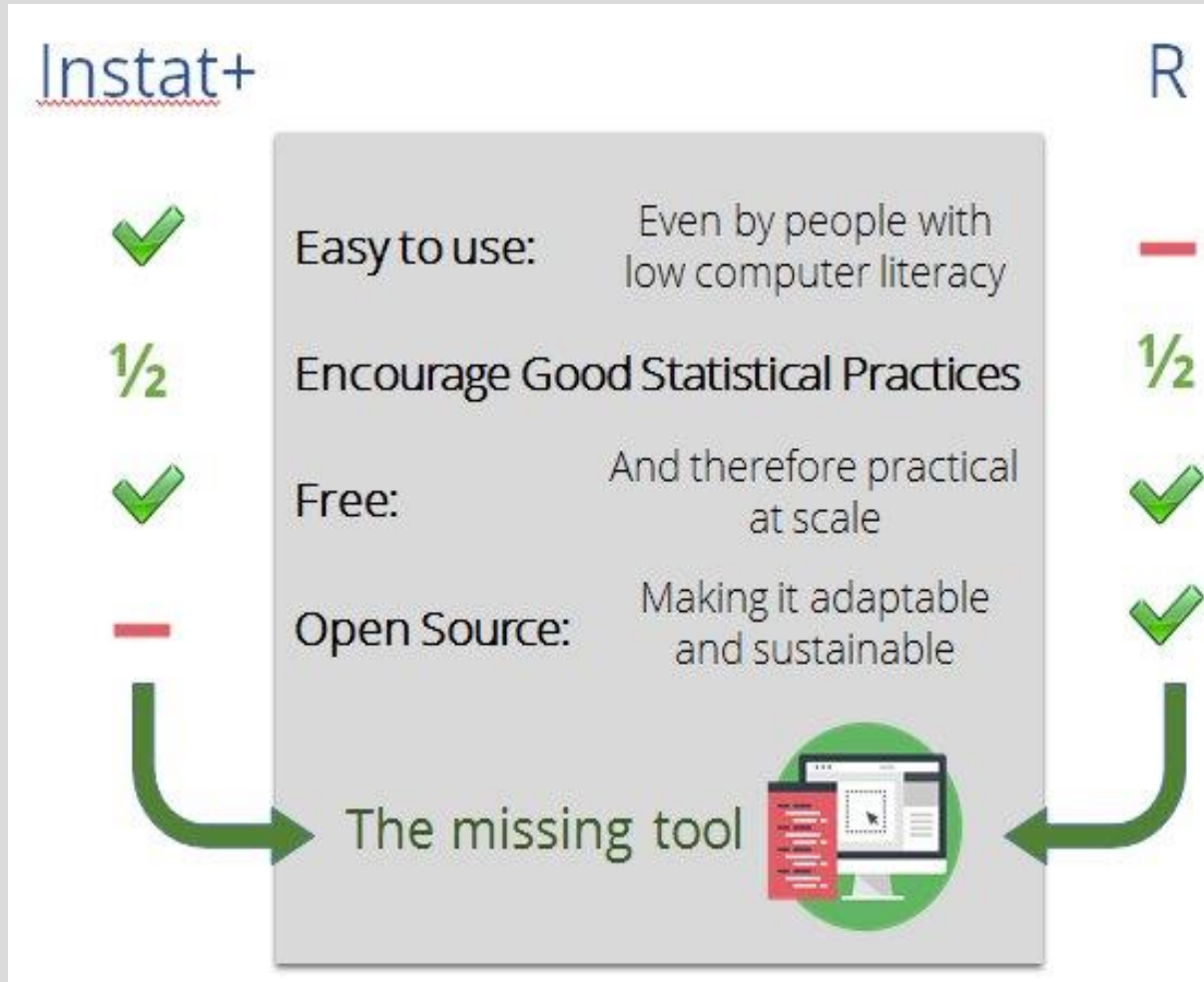


Applications of R-Instat to analyse and validate  
satellite-based climate time series

# **Introduction to R-Instat**

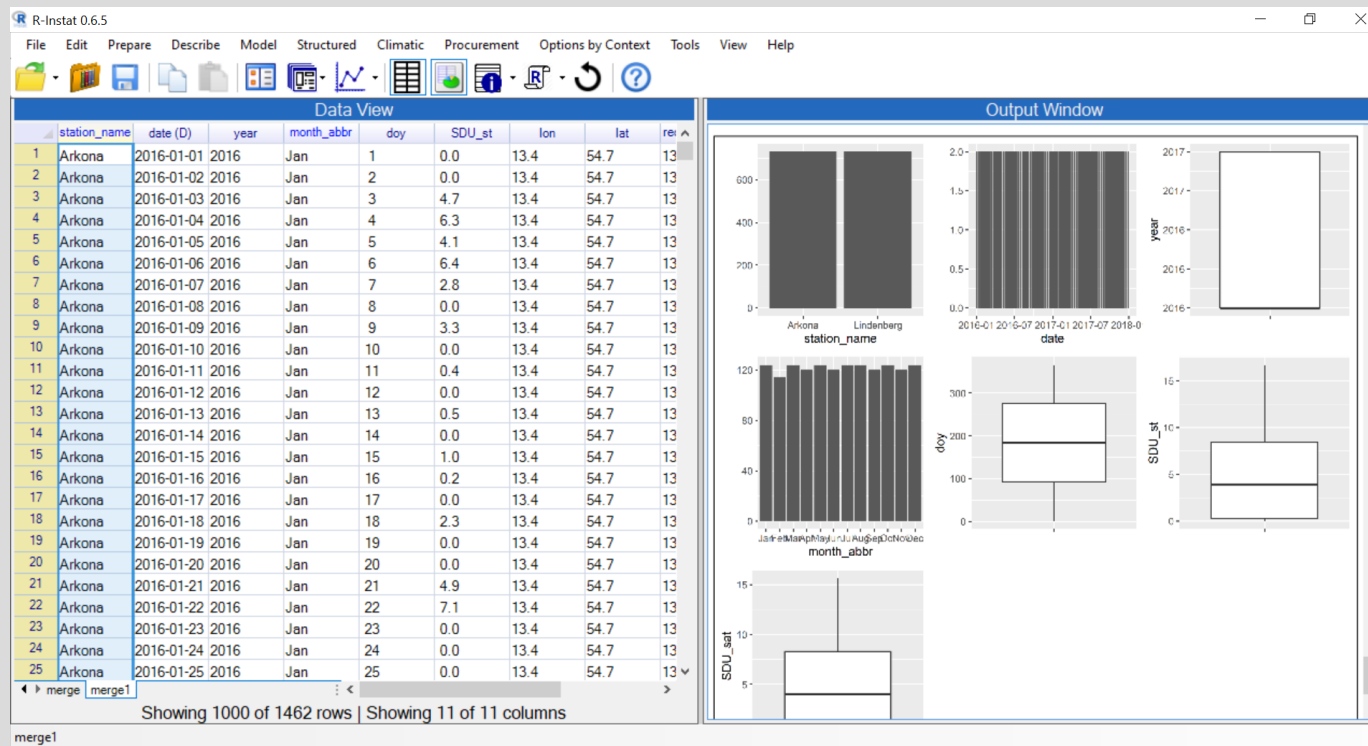
Danny Parsons  
IDEMS International  
3 February 2021

# Background to R-Instat



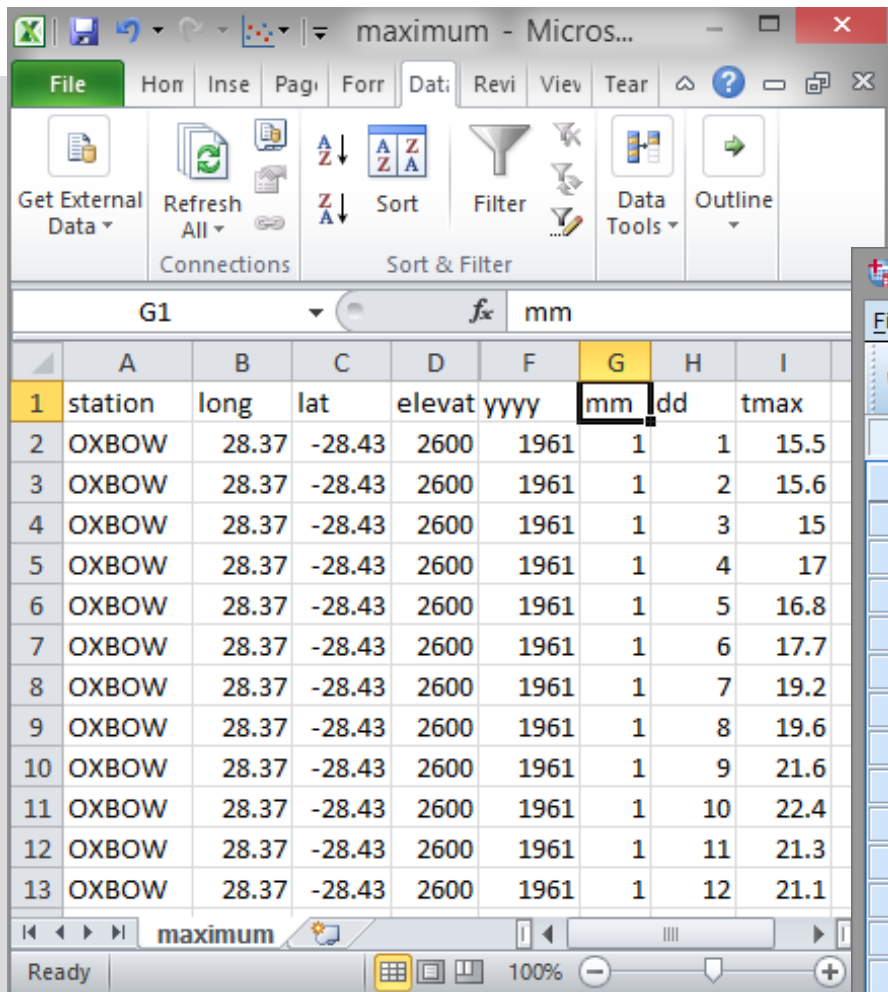
- Crowd Sourcing campaign 2015
- Identified a gap in statistical software
- Work in Africa and UK
- Support from CCAFS, EUMETSAT, WMO and UK Met for climate features

# R-Instat



- Menu-driven front end to R
- Designed to make R easy to use
- Particularly for those who already use a spreadsheet.
- A general statistics package
- With a special Climatic menu
- Free to download and open source

# R-Instat easy to use?



maximum - Micros...

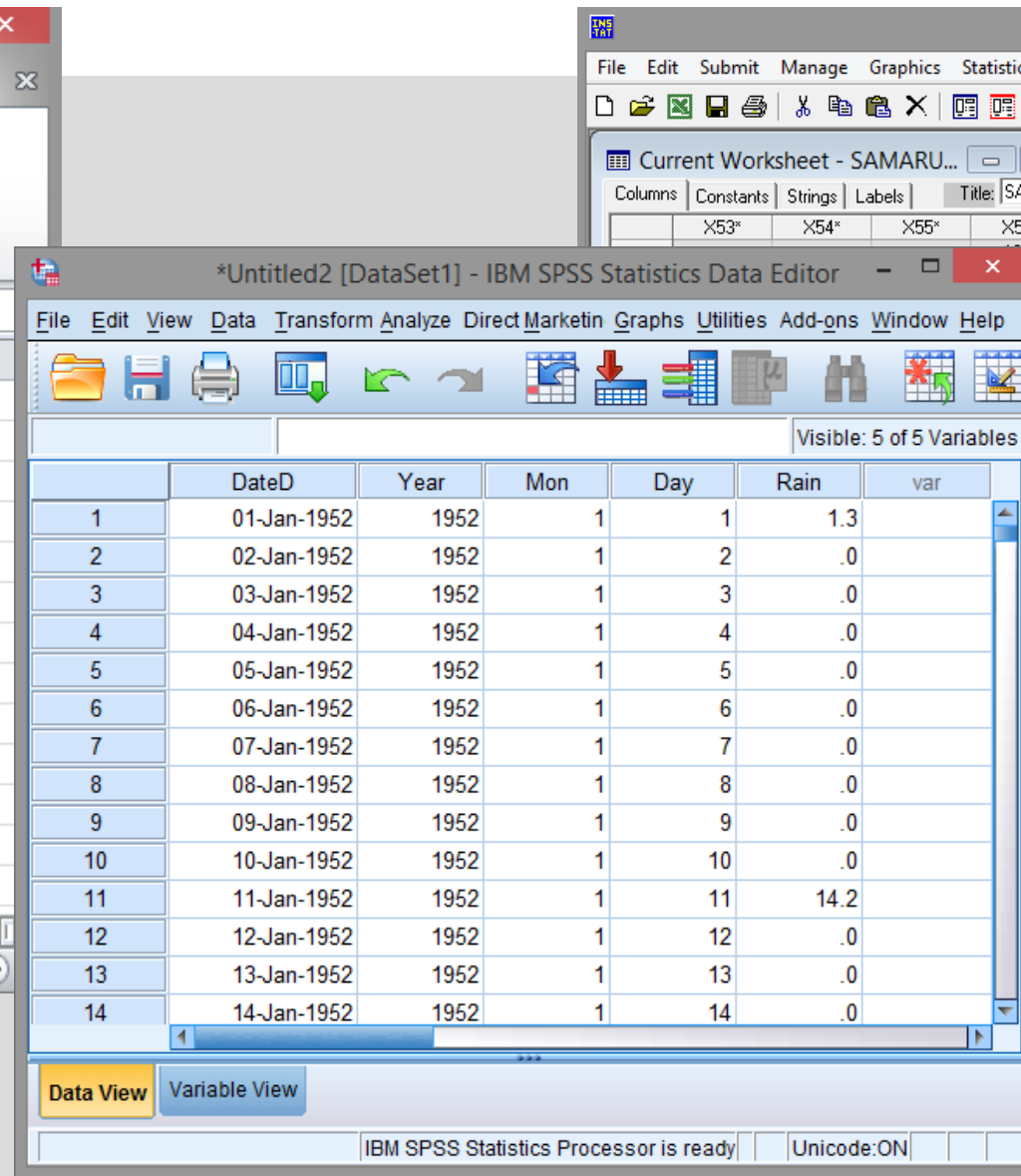
File Home Insert Page Layout Formulas Data Review View Tools Ribbon

Get External Data Refresh All Connections Sort & Filter

G1 fx mm

	A	B	C	D	F	G	H	I
1	station	long	lat	elevat	yyyy	mm	dd	tmax
2	OXBOW	28.37	-28.43	2600	1961	1	1	15.5
3	OXBOW	28.37	-28.43	2600	1961	1	2	15.6
4	OXBOW	28.37	-28.43	2600	1961	1	3	15
5	OXBOW	28.37	-28.43	2600	1961	1	4	17
6	OXBOW	28.37	-28.43	2600	1961	1	5	16.8
7	OXBOW	28.37	-28.43	2600	1961	1	6	17.7
8	OXBOW	28.37	-28.43	2600	1961	1	7	19.2
9	OXBOW	28.37	-28.43	2600	1961	1	8	19.6
10	OXBOW	28.37	-28.43	2600	1961	1	9	21.6
11	OXBOW	28.37	-28.43	2600	1961	1	10	22.4
12	OXBOW	28.37	-28.43	2600	1961	1	11	21.3
13	OXBOW	28.37	-28.43	2600	1961	1	12	21.1

maximum 100%



\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

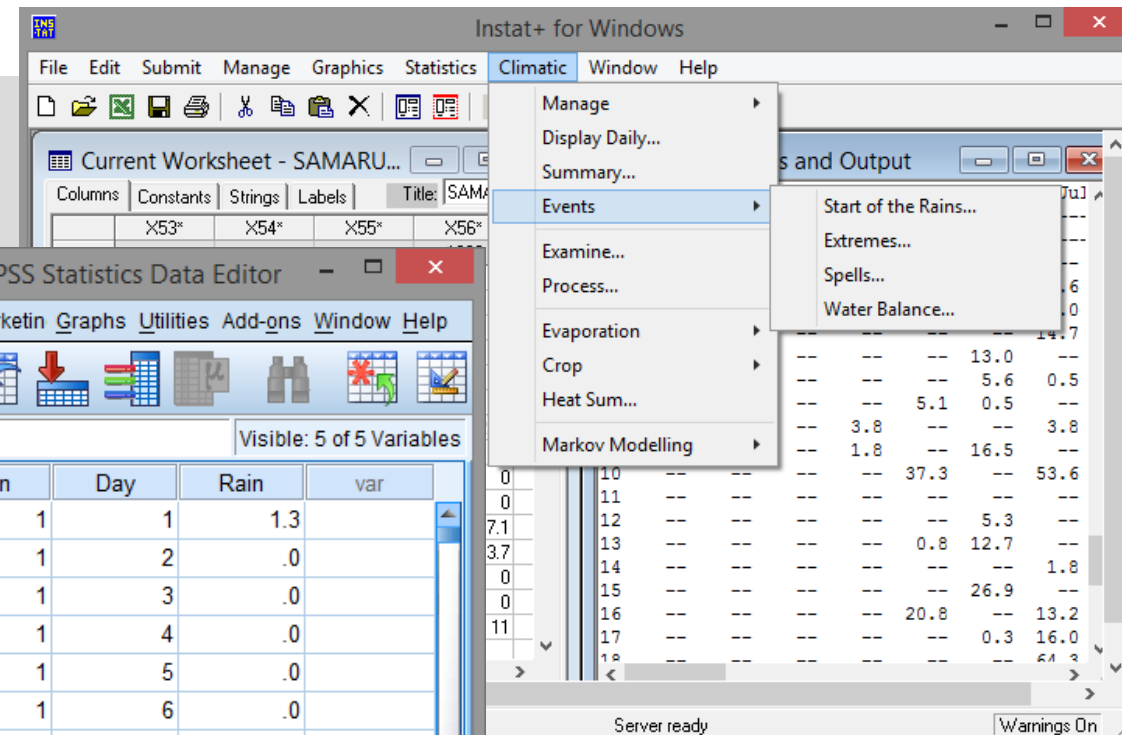
File Edit View Data Transform Analyze Direct Marketin Graphs Utilities Add-ons Window Help

Visible: 5 of 5 Variables

	DateD	Year	Mon	Day	Rain	var
1	01-Jan-1952	1952	1	1	1.3	
2	02-Jan-1952	1952	1	2	.0	
3	03-Jan-1952	1952	1	3	.0	
4	04-Jan-1952	1952	1	4	.0	
5	05-Jan-1952	1952	1	5	.0	
6	06-Jan-1952	1952	1	6	.0	
7	07-Jan-1952	1952	1	7	.0	
8	08-Jan-1952	1952	1	8	.0	
9	09-Jan-1952	1952	1	9	.0	
10	10-Jan-1952	1952	1	10	.0	
11	11-Jan-1952	1952	1	11	14.2	
12	12-Jan-1952	1952	1	12	.0	
13	13-Jan-1952	1952	1	13	.0	
14	14-Jan-1952	1952	1	14	.0	

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON



Instat+ for Windows

File Edit Submit Manage Graphics Statistics Climatic Window Help

Current Worksheet - SAMARU...

Columns Constants Strings Labels Title: SAMARU...

Events

- Start of the Rains...
- Extremes...
- Spells...
- Water Balance...

	--	--	--	13.0	--
	--	--	--	5.6	0.5
	--	--	5.1	0.5	--
	--	3.8	--	--	3.8
	--	1.8	--	16.5	--
			37.3	--	53.6
			--	--	--
			--	5.3	--
			--	0.8	12.7
			--	--	1.8
			--	--	26.9
			--	20.8	13.2
			--	--	16.0
			--	--	61.3

Server ready Warnings On

# Similar to a spreadsheet for data analysis?

✓ Name at the top of column

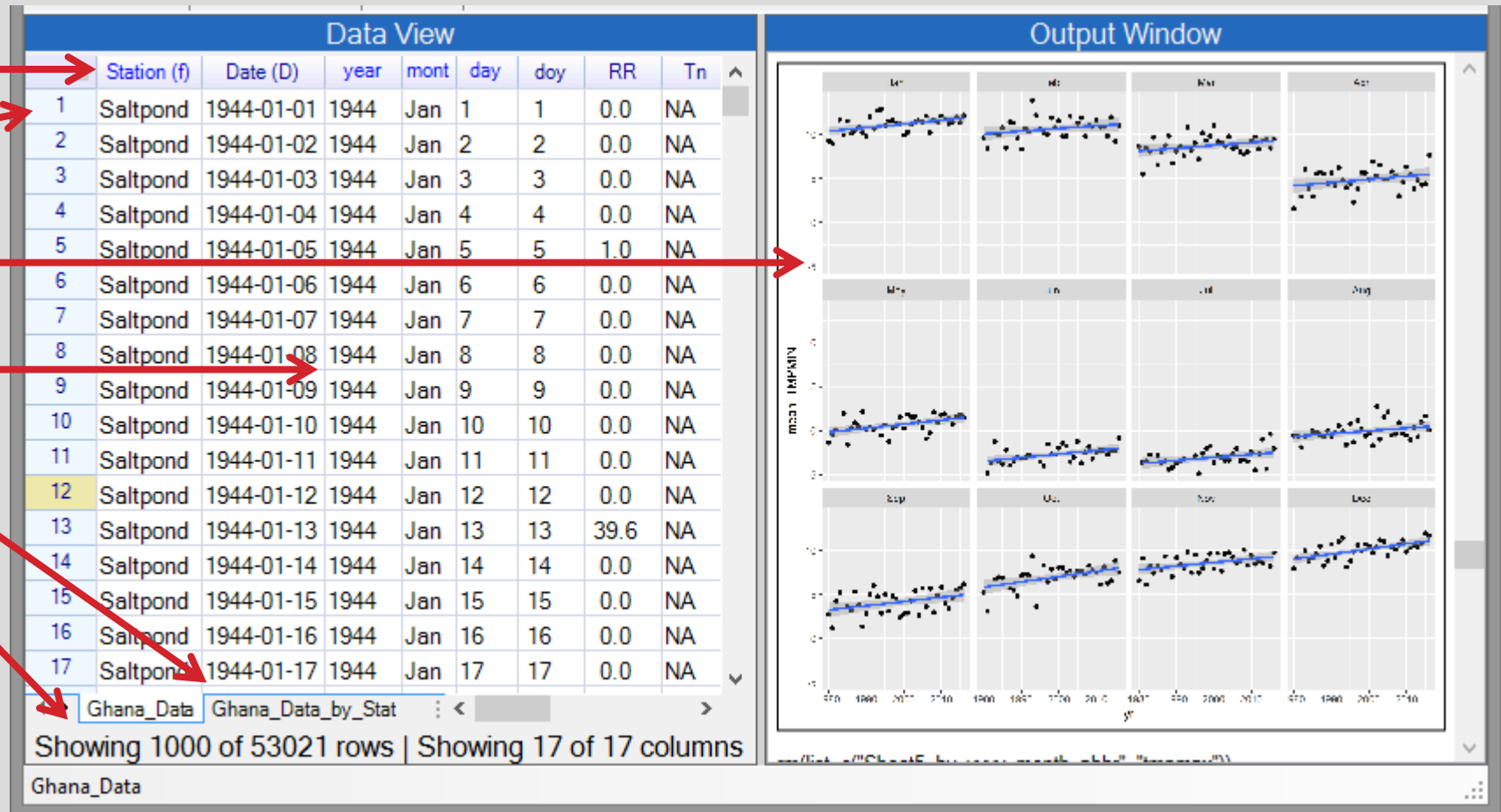
✗ Data start at row 1

✗ Results in a different window

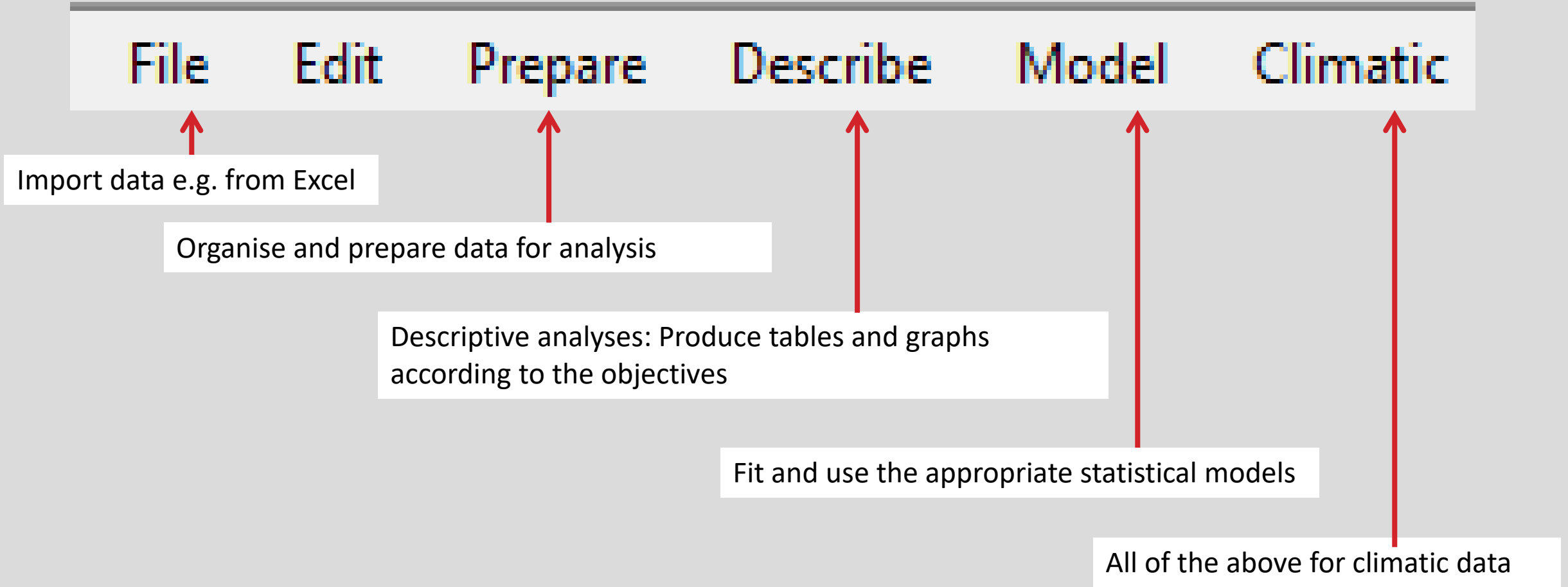
✓ Data in “columns” – one type of data only per column

✓ Multiple sheets (data frames)

✗ Just a window onto the data



# The menus in R-Instat



# Download and installing R-Instat

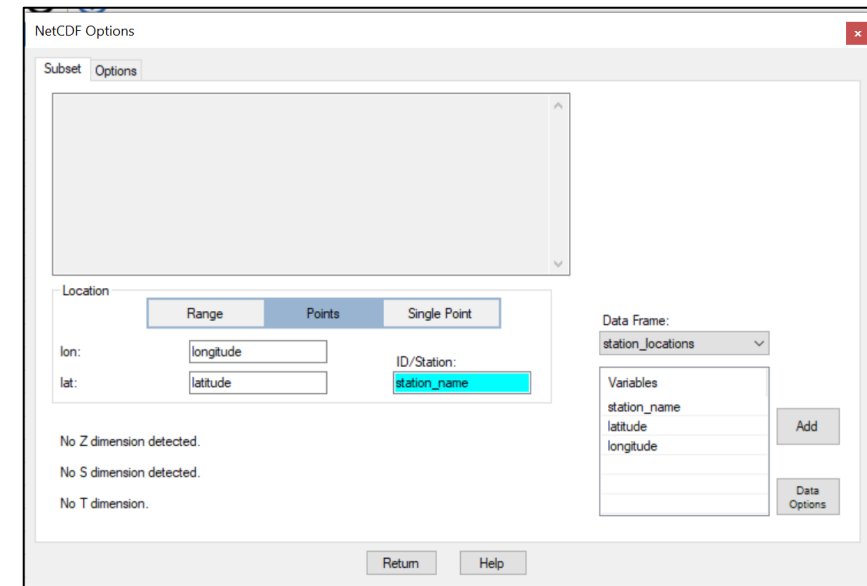
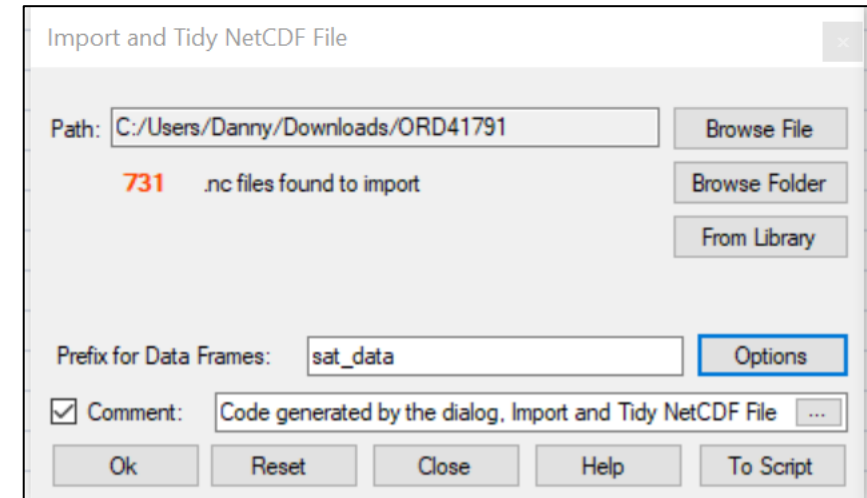
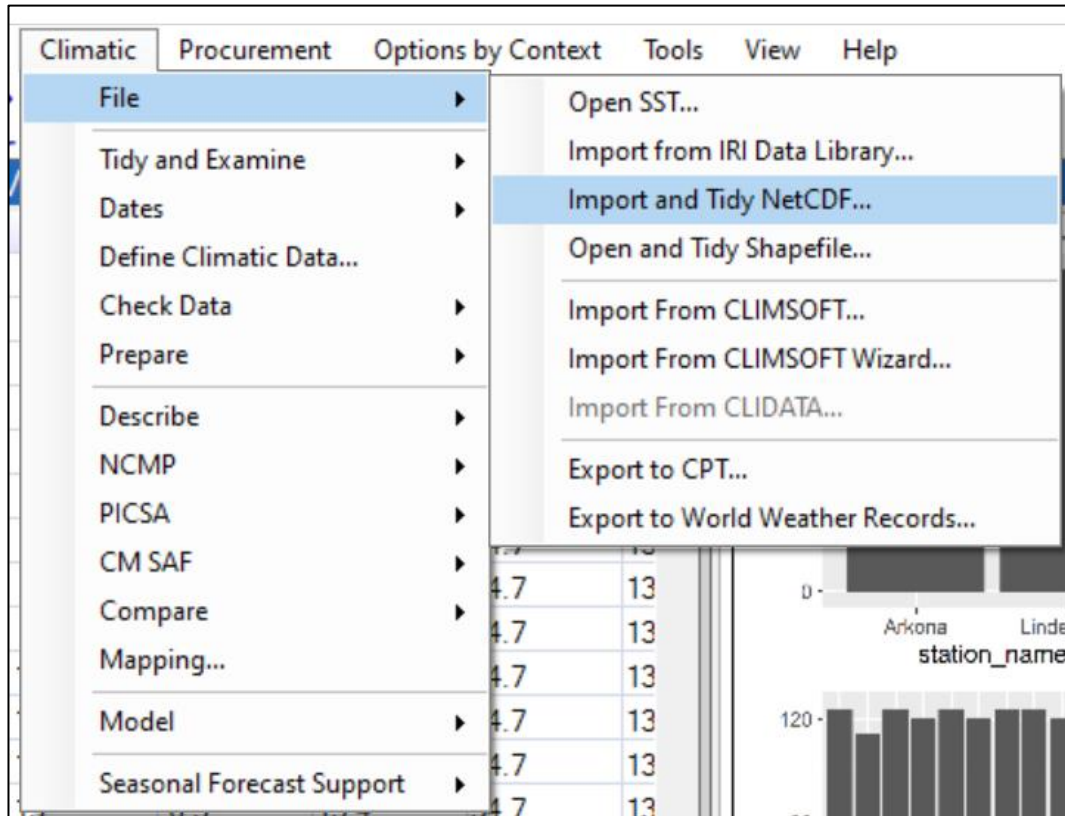
- Download from <http://r-instat.org/Download.html> (550MB)
- Currently a **Windows** software
- Install on Ubuntu 18.04 using Wine [with this script](#)
- Other Linux and Mac users
  - can be installed on a virtual Windows machine
  - cross-platform version coming soon (mid-2021)
- See Moodle course for more details  
<https://training.eumetsat.int/course/view.php?id=381>

# Case study – comparing sunshine duration (SDU) at two stations in Germany

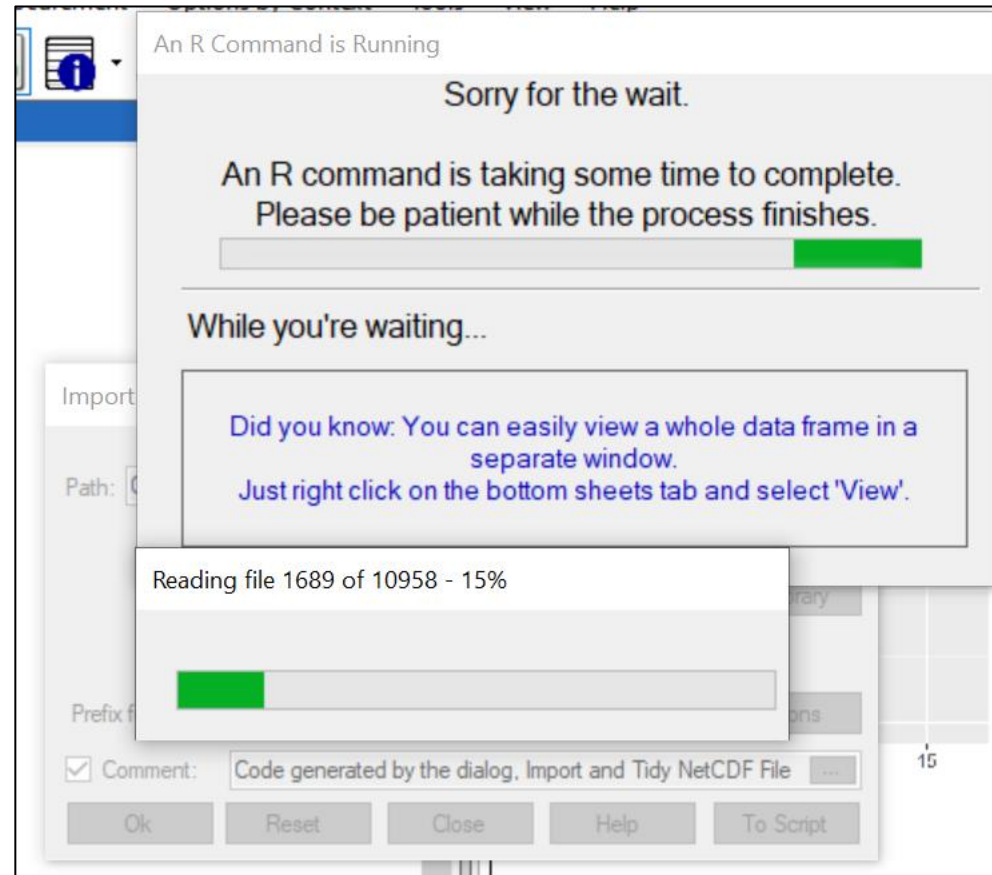
- Import and tidy station data with daily sunshine duration (SDU)
- Extract and import satellite estimates of SDU for two points
  - Import station locations
  - Extract two points from satellite data (NetCDF files) corresponding to stations
- Merge station and satellite data into one sheet (data frame)
- Prepare and tidy the merged data
- Graphical comparisons: correlations, scatter plots, seasonality plots
- Numerical comparisons: summary metrics



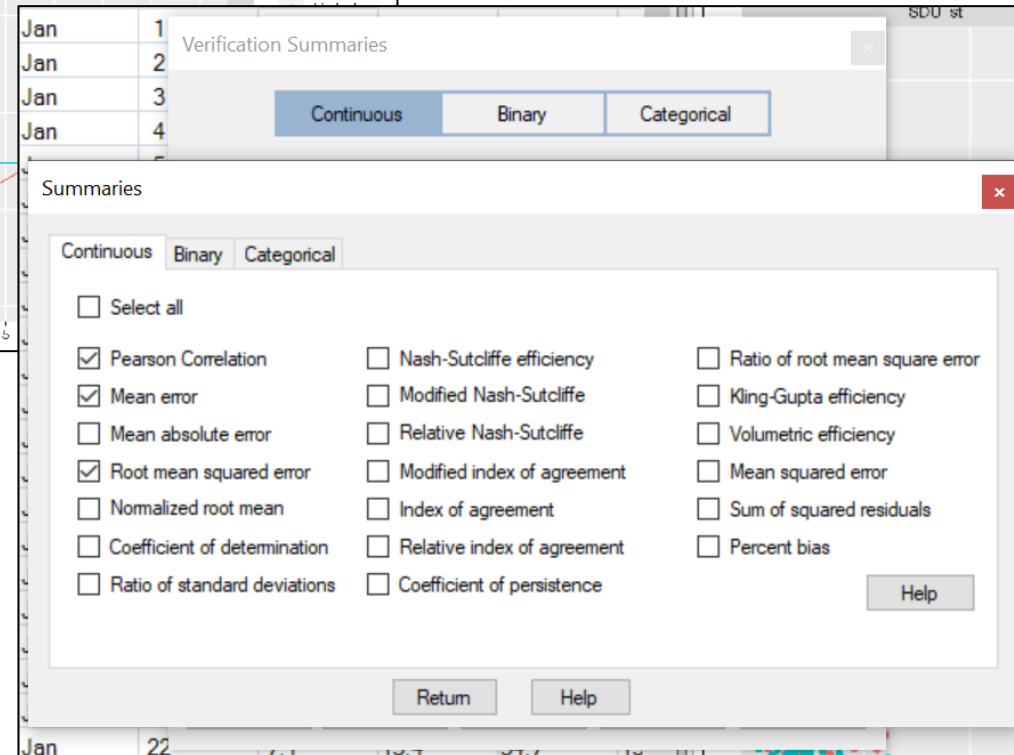
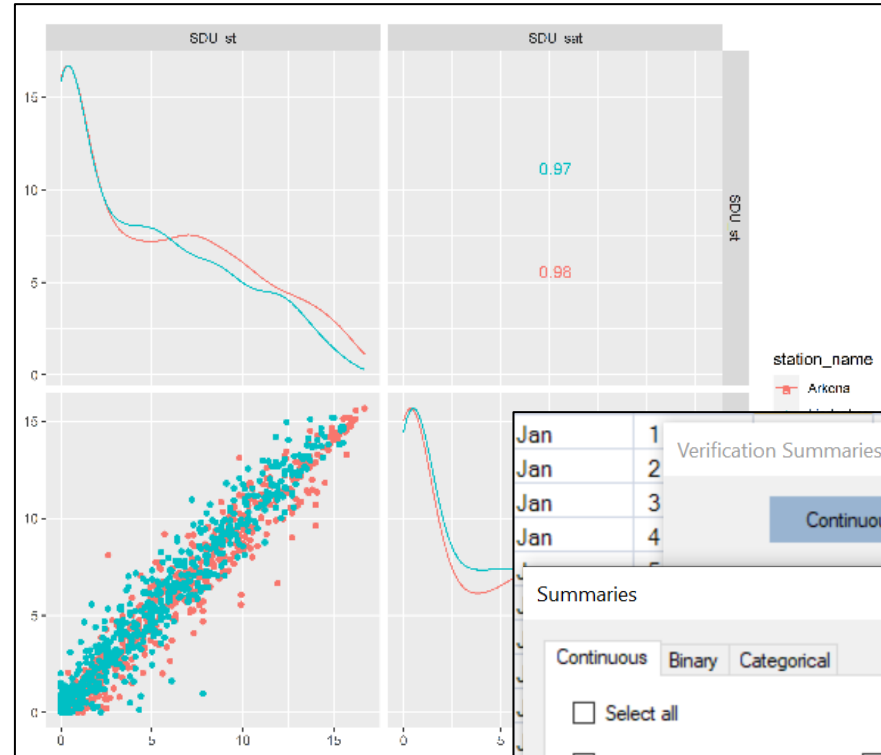
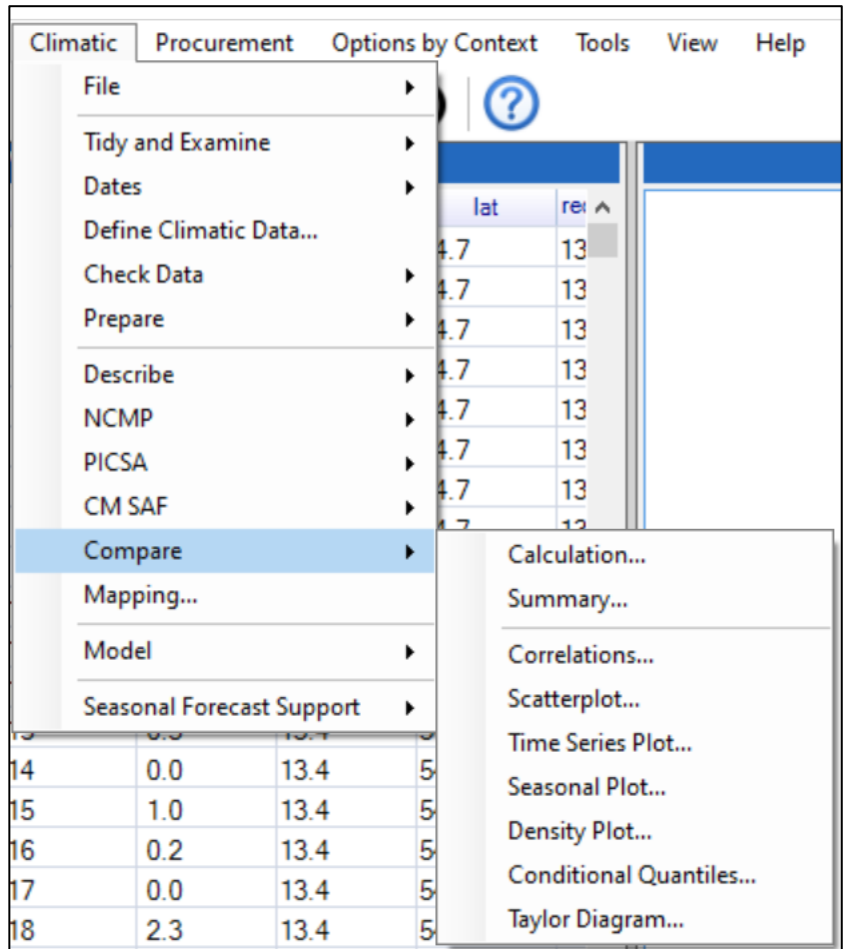
# Importing gridded data - NetCDF files



# Importing gridded data - NetCDF files



# Comparing satellite and station data



# Prepare – Totals from daily data

Prepare Describe Model Climatic Procurement

Data Frame

Check Data

Column: Calculate

Column: Generate

Column: Factor

Column: Text

Column: Date

Column: Reshape

Keys and Links

Data Object

R Objects

Column Summaries...

General Summaries...

Stack...

Unstack...

Merge...

Append Data Frames...

Subset...

Random Subset...

Transpose...

	month_abbrev	day_in_month	count
	Jan	1	1
	Jan	2	2
	Jan	3	3
	Jan	4	4
	Jan	5	5

Column Statistics

Data Frame: Ghana\_Data

Variable(s) to Summarise: Ghana\_Data, RR

By Factor(s): Ghana\_Data, Station, year

Options

☒ Store Results in Data

☐ Print Results to Output

☐ Drop Unused Levels

☒ Omit Missing Values

☒ Comment: code generated by the dialog Column Statistics

Ok Reset Close Help To Script

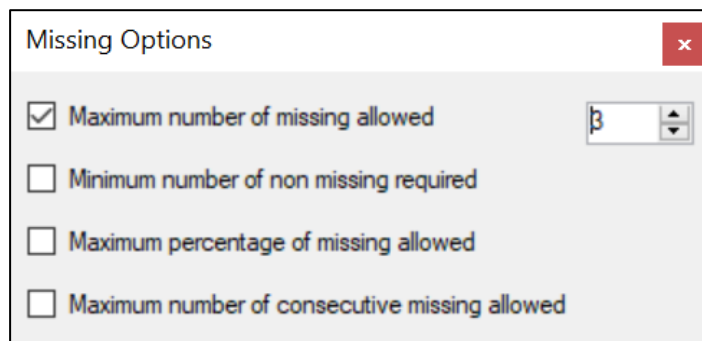
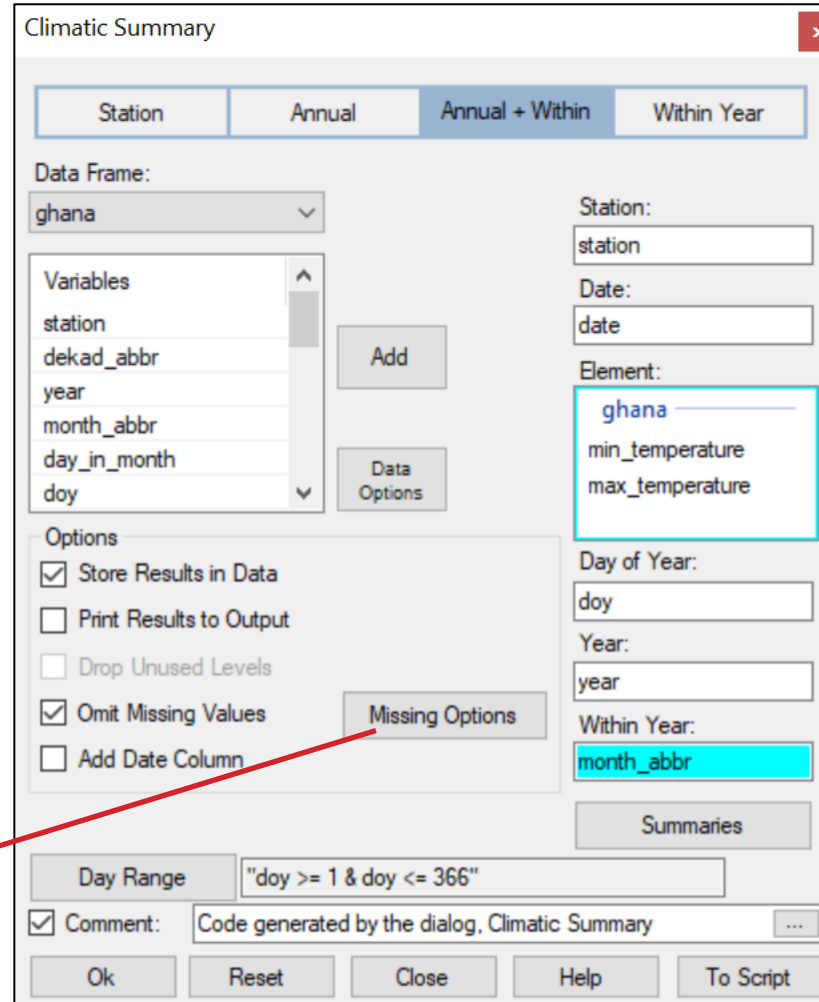
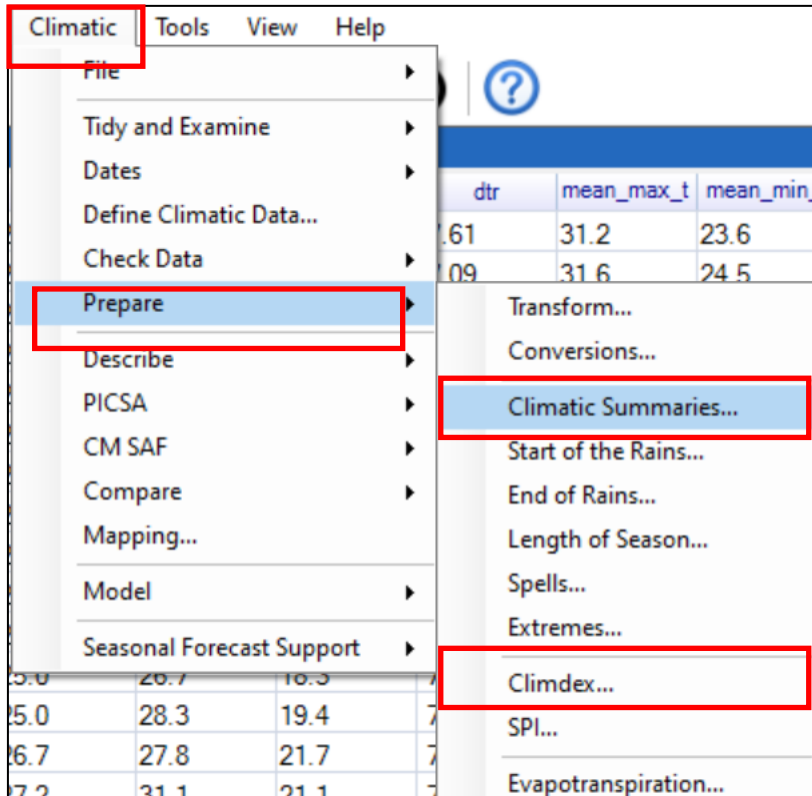
Data View					
	Station (f)	year (f)	count_n	count_RR	sum_RR
1	Saltpond	1944	366	366	724
2	Saltpond	1945	365	365	644
3	Saltpond	1946	365	365	1001
4	Saltpond	1947	365	365	1017
5	Saltpond	1948	366	366	1089
6	Saltpond	1949	273	273	772
7	Saltpond	1950	365	365	776
8	Saltpond	1951	365	365	1428
9	Saltpond	1952	366	366	1376
10	Saltpond	1953	365	365	699
11	Saltpond	1954	365	365	986
12	Saltpond	1955	365	365	1485
13	Saltpond	1956	365	366	1151
14	Saltpond	1957	365	365	702
15	Saltpond	1958	365	365	749
16	Saltpond	1959	365	365	789
17	Saltpond	1960	366	366	1004

Ghana\_Data Ghana\_Data\_by\_Station\_year

Showing 146 of 146 rows | Showing 5 of 5

Ghana\_Data\_by\_Station\_year

# Climatic – Totals from daily data

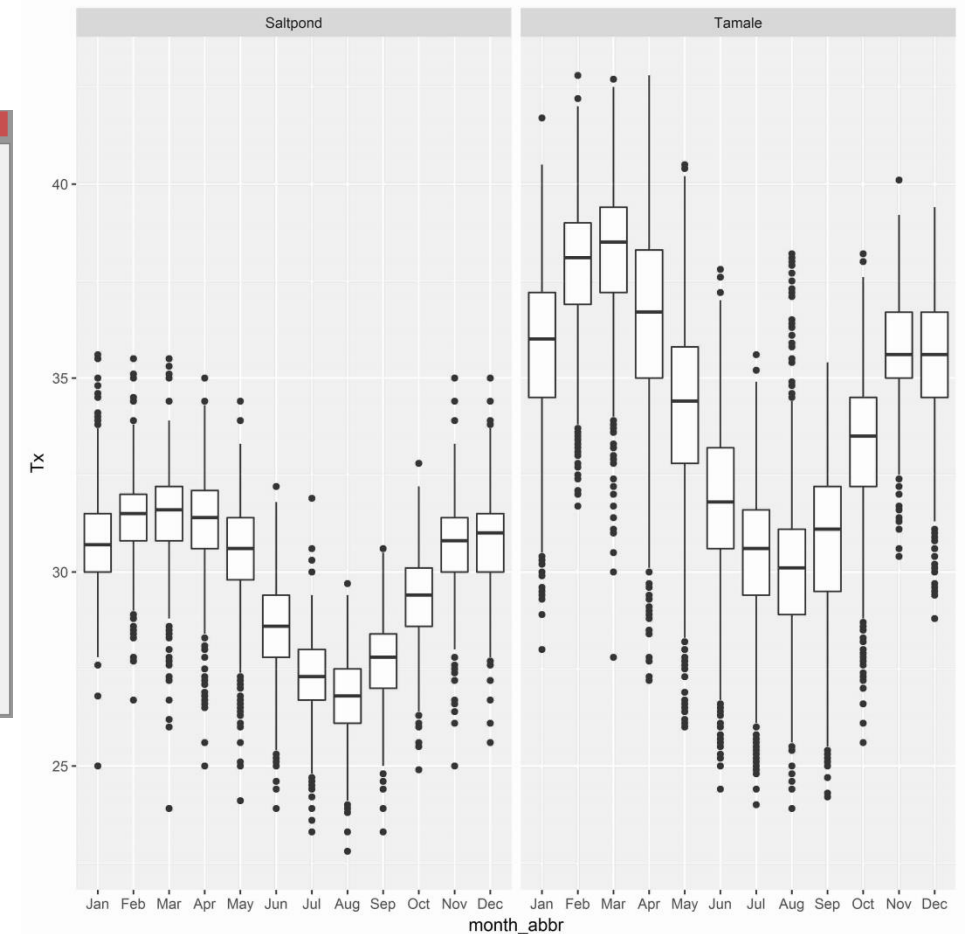
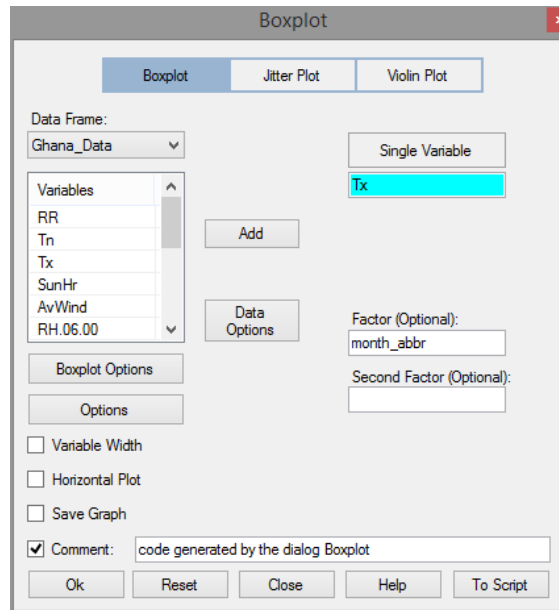
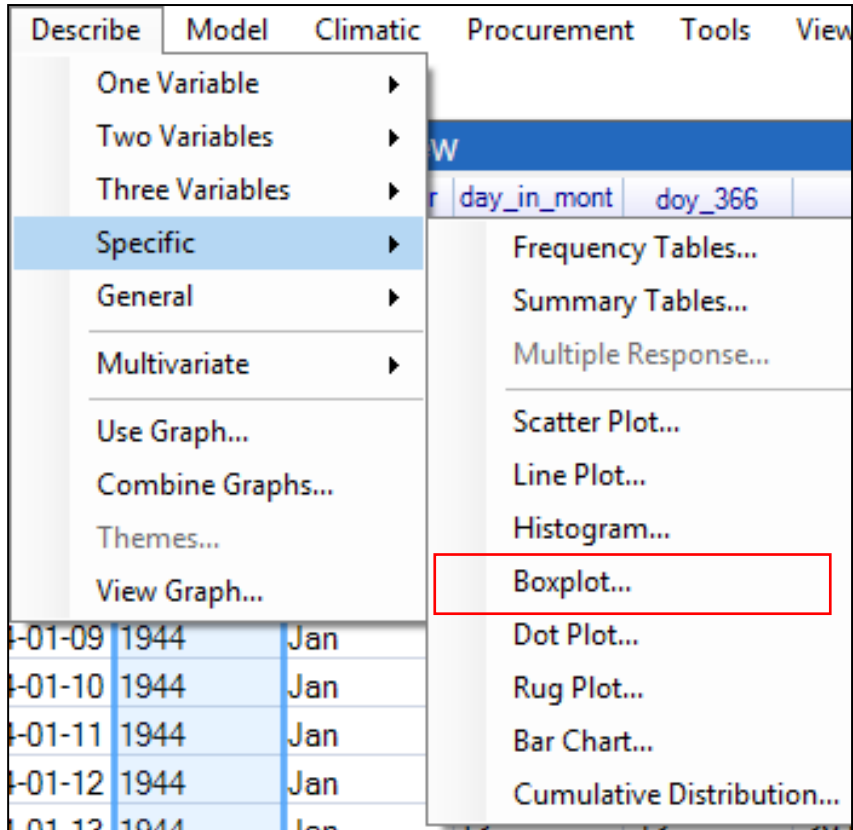


Data View						
	station (f)	year	month_	mean_max	mean_min	dtr
1	Saltpond	1960	Jan	31.2	23.6	7.61
2	Saltpond	1960	Feb	31.6	24.5	7.09
3	Saltpond	1960	Mar	31.5	23.8	7.71
4	Saltpond	1960	Apr	31.2	23.8	7.33
5	Saltpond	1960	May	30.7	23.6	7.05
6	Saltpond	1960	Jun	28.4	23.1	5.36
7	Saltpond	1960	Jul	27.8	21.7	6.10
8	Saltpond	1960	Aug	27.2	22.0	5.18
9	Saltpond	1960	Sep	28.1	22.5	5.55
10	Saltpond	1960	Oct	29.9	23.1	6.76
11	Saltpond	1960	Nov	31.6	23.2	8.34
12	Saltpond	1960	Dec	31.0	23.4	7.58
13	Saltpond	1961	Jan	31.0	23.3	7.73
14	Saltpond	1961	Feb	31.7	23.9	7.77
15	Saltpond	1961	Mar	32.2	24.6	7.53
16	Saltpond	1961	Apr	32.0	23.9	8.16
17	Saltpond	1961	May	31.4	24.5	6.95

ghana | ghana\_by\_station\_year | ghana\_by\_station\_year\_month

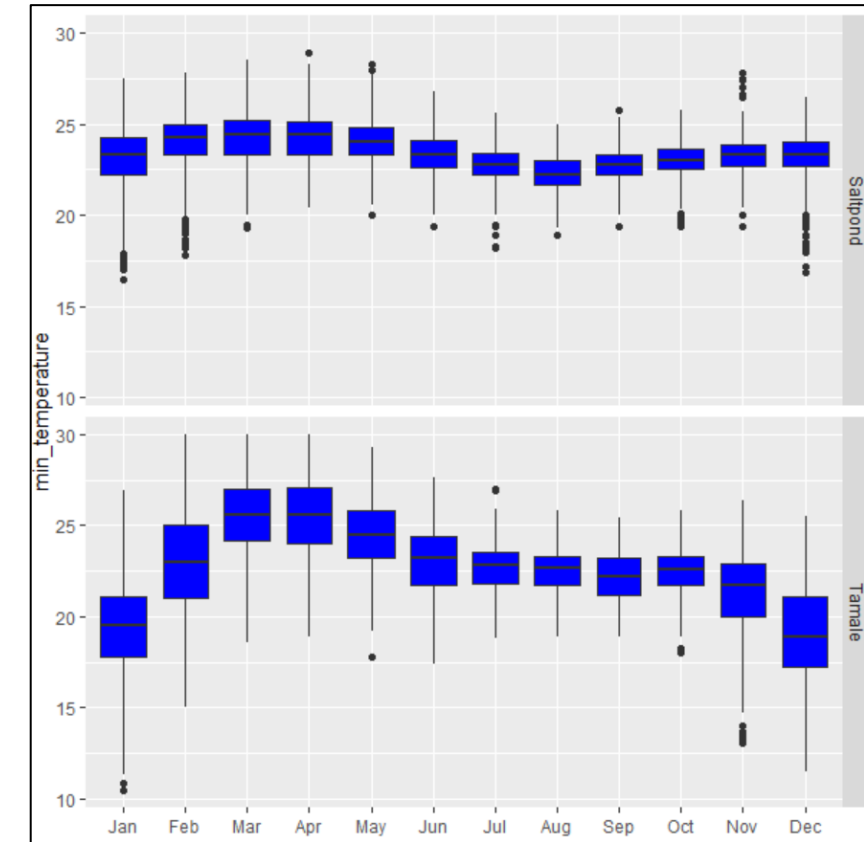
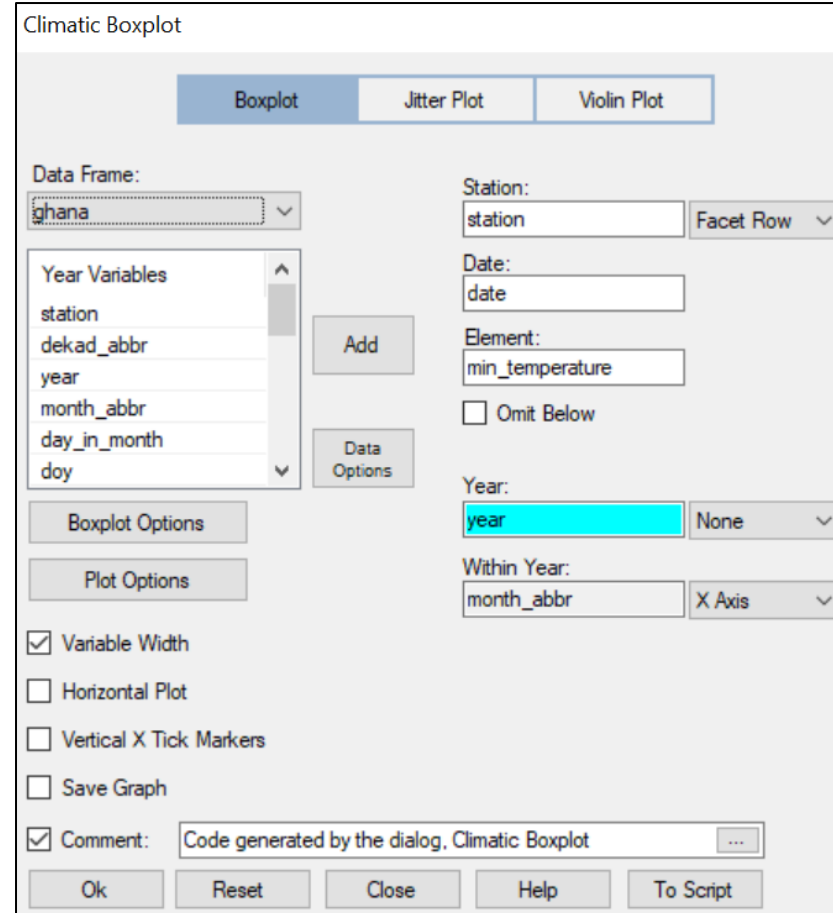
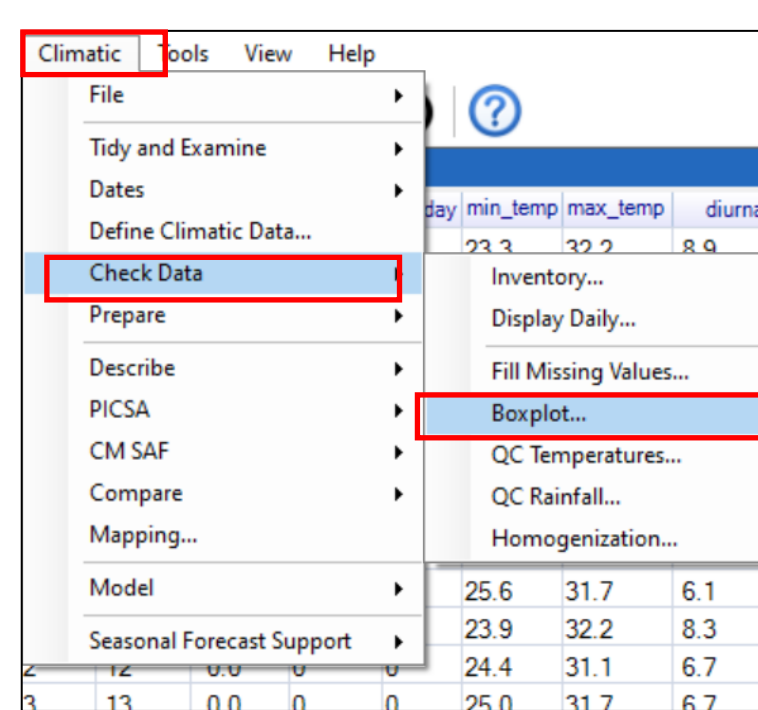
Showing 1000 of 1368 rows | Showing 11 of 11 columns

# Describe: Boxplots

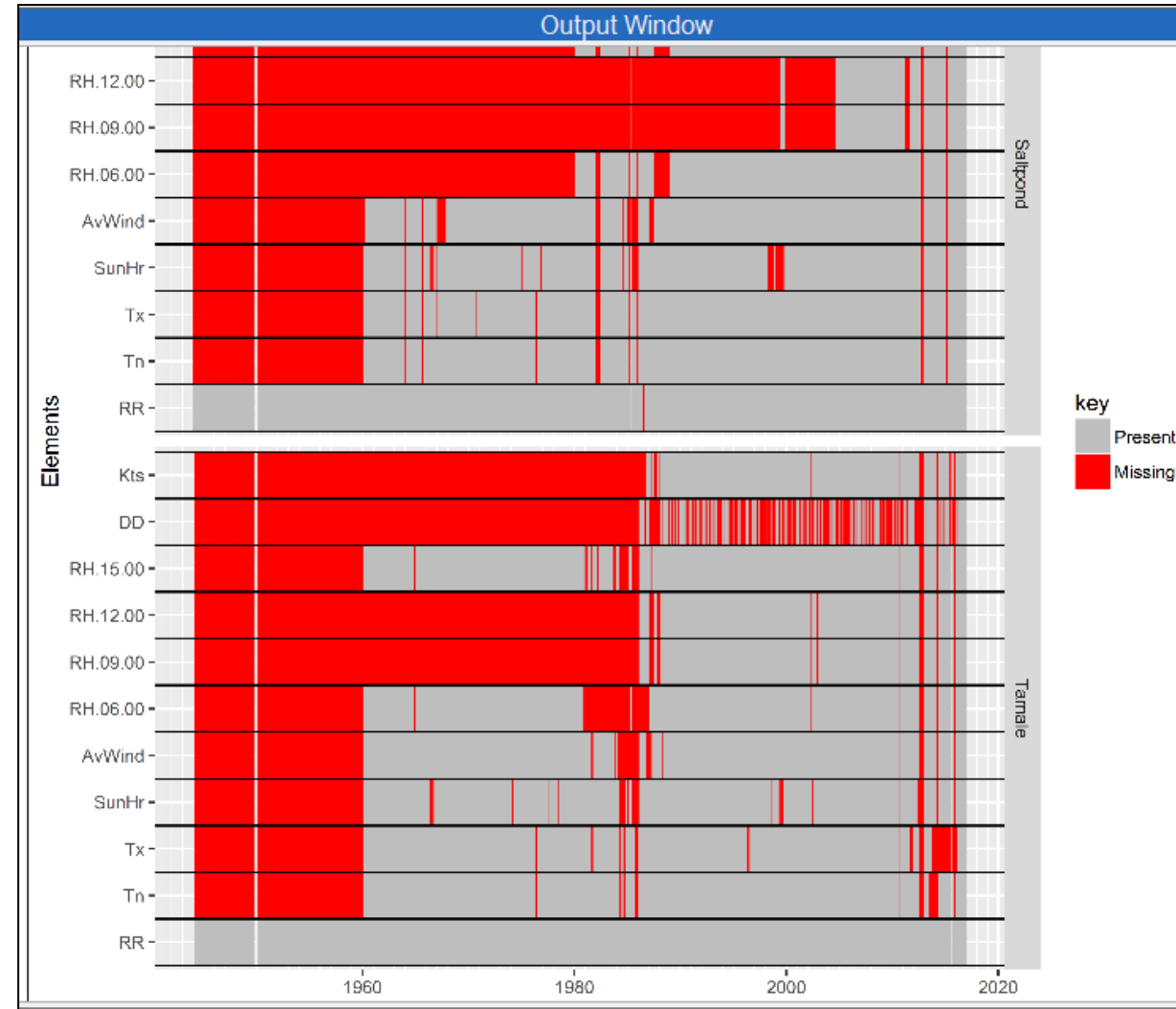
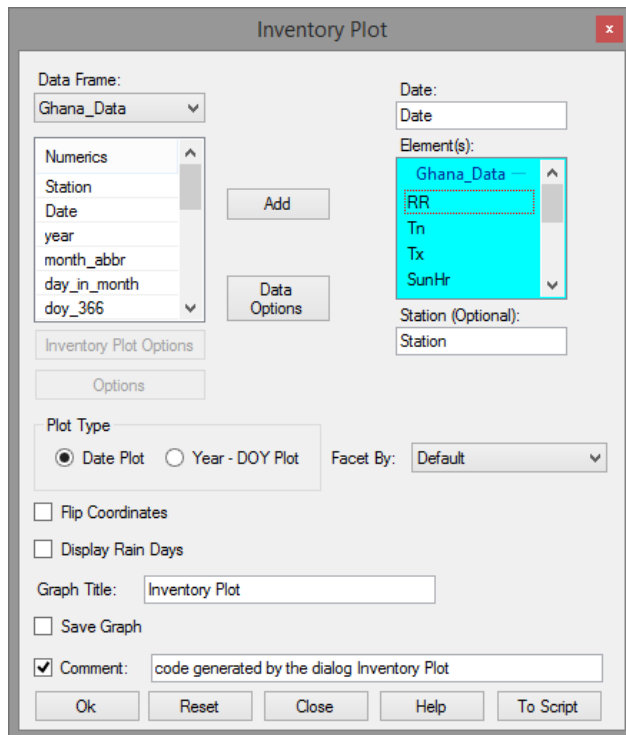
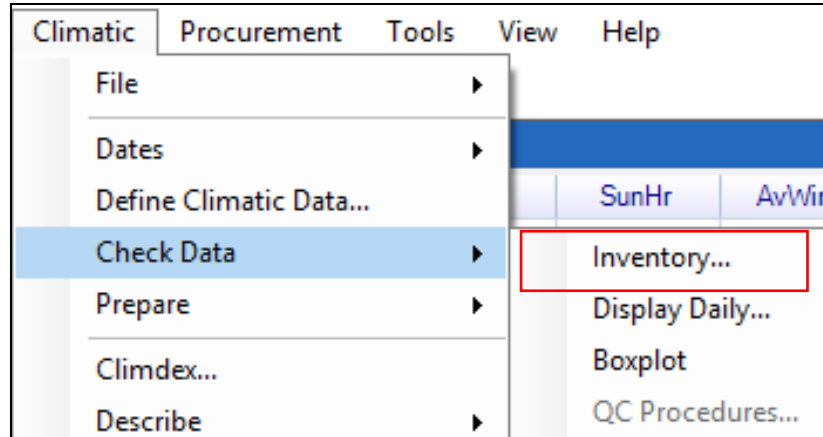




# Climatic: Boxplots



# Climatic: Inventory





# Exercises and practice

1. Practice the steps from the demo with the Germany data
    - practice data preparation and analyses
  2. Analyse the pre-prepared example data from Germany
    - focus on the statistical analyses and comparisons in more detail
    - Use pre-processed data files
  3. Prepare and Analyse your own data
    - If you have your own station data and can download CM SAF data
    - Use the climatic guide to help prepare your data
- Use the documents and videos on Moodle as guidance
  - Ask questions on Slido if you need any help: <https://slido.com> code #EUMSC10

# Getting help and giving feedback

We want to hear from you!

## **During this course**

- Post questions on <https://slido.com> code **#EUMSC10**
- Post results on Padlet [https://padlet.com/CMSAF/cmsaf\\_padlet](https://padlet.com/CMSAF/cmsaf_padlet)

## **After the course**

- Feature requests and bug reports, post an issue on GitHub <https://github.com/africanmathsinitiative/R-Instat/issues>
- General email [rinstat@africanmathsinitiative.net](mailto:rinstat@africanmathsinitiative.net)
- or Danny [danny@idems.international](mailto:danny@idems.international)