

## **Atelier *Climate-Smart Solutions*: L'analyse, la présentation et l'interprétation de l'information climatique utile aux agriculteurs**

Présenté par Dr. Roger Stern, Université de Reading/Stats4SD et Dr. Caroline Staub, Université de Floride

**Introduction:** Cette formation de deux jours vise à former des techniciens dont le rôle consiste à récolter, manipuler et analyser des données d'ordre climatique ou environnementale et à présenter les résultats de ces analyses à un public composé de non-techniciens.

**Objectifs:** Les participants devront:

1. Analyser l'information climatique à l'aide du logiciel accès libre R-Instat
2. Développer des graphiques dont l'interprétation est suffisamment simple et pertinente pour guider la discussion des agents de terrain avec les agriculteurs
3. Interpréter les résultats des analyses afin de pouvoir les partager avec les agents de terrain de façon claire et précise.

### **Programme du 5 Juin:**

Debut: 9:00 Fin: 17:00

1. Introduction du programme "climate-smart solutions" du projet AREA.
2. L'intégration de l'information climatique a la prise de décision agricole.
3. L'usage du logiciel R-Instat et du fichier Microsoft Excel pour analyser l'information climatique
4. Les analyses de base
5. La présentation des résultats sous forme de tableaux et graphiques

### **Programme du 6 Juin:**

Debut: 9:00 Fin: 17:00

1. Introduction du menu "climat" du logiciel R-Instat.
2. Les synthèses pratiques: le début et la fin de la saison des pluies.
3. Préparer des graphiques de qualité
4. Analyses additionnelles
5. Evaluation de l'atelier

# Analysing climatic data with R-Instat

## Introduction

R-instat est un logiciel conçu pour l'analyse statistique. Toutes les opérations y sont effectuées à partir du langage statistique R. En plus, R-instat comporte un menu spécial 'climatique'.

Les méthodes actuelles sont conçues pour l'analyse des données quotidiennes. Ce guide utilise des données quotidiennes de deux (02) stations provenant de la Guinée (Conakry) et comportent quatre (04) paramètres chacune. Ces données sont accessibles librement à travers la bibliothèque R-Instat, par conséquent, les exemples présentés dans ce guide peuvent être suivis par n'importe qui. Ce guide fait suite au guide "Preparing climatic data for analysis" et vise à renforcer la maîtrise de l'analyse statistique et de la gestion de l'information climatique.

Nous sommes reconnaissants envers l'agence nationale de météorologie de la Guinée de nous avoir permis d'utiliser ces données à des fins de formation.

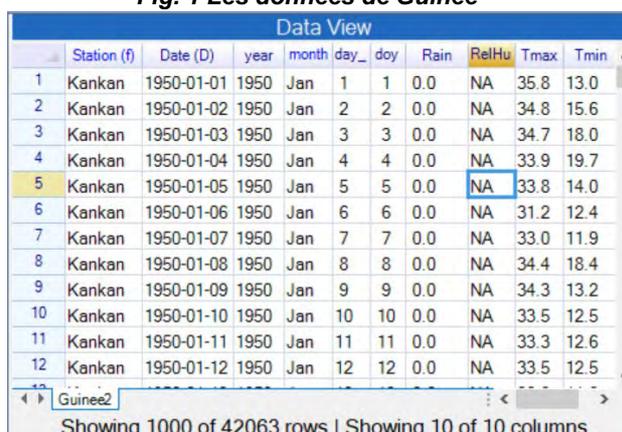
### 1) Produire un résumé pertinent

Si vous avez parcouru le premier guide, vous devriez avoir votre propre copie du fichier intitulé Guinea2.rds.

\*\* **Ouvrez ce fichier**, soit du dossier dans lequel vous l'avez enregistré, soit de la Instat library:

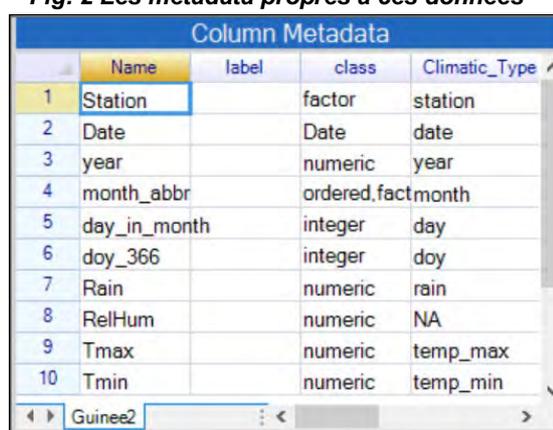
\*\* ('S'il se trouve dans le Instat library, appuyez sur **Open From Library > Load from Instat Collection > Browse > Climatic > Guinea** and ouvrez le fichier intitulé **Guinea2.rds**, Fig. 1

Fig. 1 Les données de Guinee



	Station (f)	Date (D)	year	month	day_	doy	Rain	RelHu	Tmax	Tmin
1	Kankan	1950-01-01	1950	Jan	1	1	0.0	NA	35.8	13.0
2	Kankan	1950-01-02	1950	Jan	2	2	0.0	NA	34.8	15.6
3	Kankan	1950-01-03	1950	Jan	3	3	0.0	NA	34.7	18.0
4	Kankan	1950-01-04	1950	Jan	4	4	0.0	NA	33.9	19.7
5	Kankan	1950-01-05	1950	Jan	5	5	0.0	NA	33.8	14.0
6	Kankan	1950-01-06	1950	Jan	6	6	0.0	NA	31.2	12.4
7	Kankan	1950-01-07	1950	Jan	7	7	0.0	NA	33.0	11.9
8	Kankan	1950-01-08	1950	Jan	8	8	0.0	NA	34.4	18.4
9	Kankan	1950-01-09	1950	Jan	9	9	0.0	NA	34.3	13.2
10	Kankan	1950-01-10	1950	Jan	10	10	0.0	NA	33.5	12.5
11	Kankan	1950-01-11	1950	Jan	11	11	0.0	NA	33.3	12.6
12	Kankan	1950-01-12	1950	Jan	12	12	0.0	NA	33.5	12.5

Fig. 2 Les metadata propres à ces données



	Name	label	class	Climatic_Type
1	Station		factor	station
2	Date		Date	date
3	year		numeric	year
4	month_abbr		ordered, fact	month
5	day_in_month		integer	day
6	doy_366		integer	doy
7	Rain		numeric	rain
8	RelHum		numeric	NA
9	Tmax		numeric	temp_max
10	Tmin		numeric	temp_min

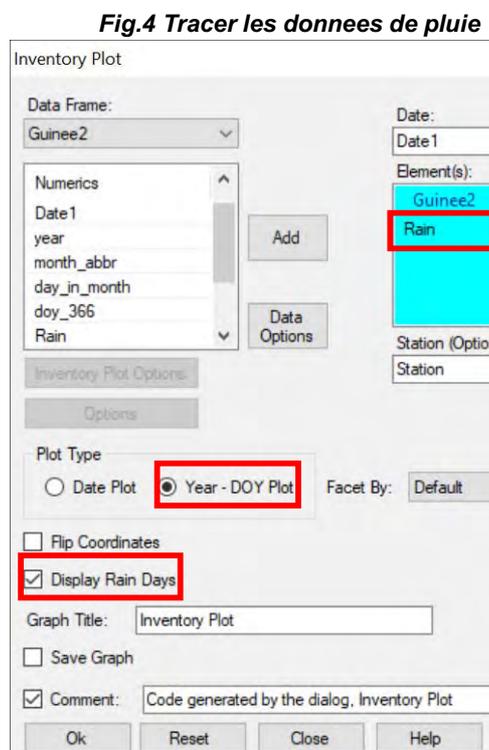
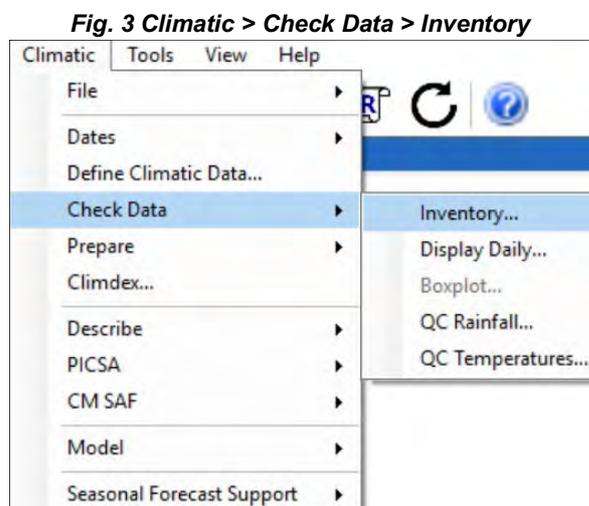
\*\* Sur le **toolbar**, appuyez sur l'icône **i** pour obtenir les metadata (métadonnées) associées à ces données (Fig. 2).

Dans les metadata, chaque ligne fournit des informations sur une variable ou une colonne associées aux données. Vérifiez que l'information inclut le type climatique, c'est-à-dire que ces données ont été préparées pour une analyse climatique.

\*\* Appuyez soit sur le même bouton soit sur la flèche semi-circulaire, pour restaurer les fenêtres selon la disposition par défaut.

Nous parcourons à nouveau un exemple du « **inventory** » du menu **climatic** afin de pouvoir produire des résumés climatique.

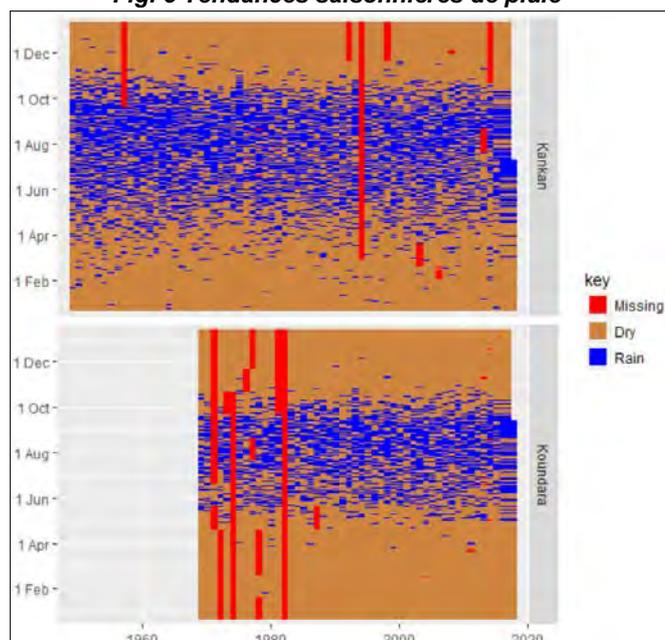
\*\* Sélectionnez **Climatic > Check Data > Inventory**, Fig. 3.



\*\* **Completez le dialogue** comme présenté ci-dessus Fig. 4.

\*\* Appuyez sur **OK** pour obtenir les résultats présentés ci-dessous Fig. 5.

**Fig. 5 Tendances saisonnières de pluie**

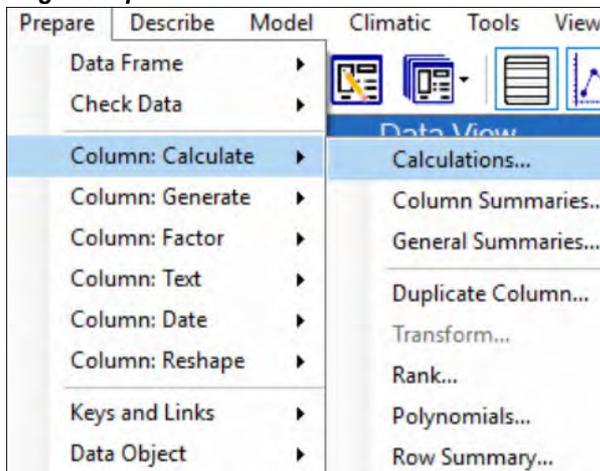


L'une de nos préoccupations concernent les données manquantes, en particulier à Kankon. Il manque des données sur 4 ou 5 ans. Heureusement ces données manquantes se trouvent particulièrement pendant la période sèche de l'année et qui, par conséquent,

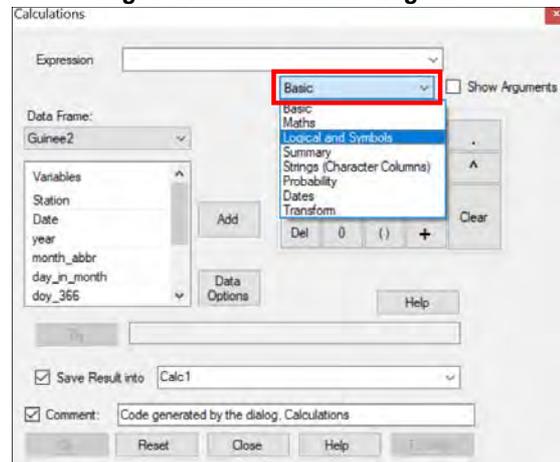
n'affecteront pas beaucoup les totaux annuels. Une autre étape préliminaire consiste à faciliter le calcul du nombre annuel de jours de pluie.

\*\* Appuyez sur **Prepare > Column: Calculate > Calculations**, Fig 6.

**Fig. 6 Prepare > Column: Calculate > Calculations**

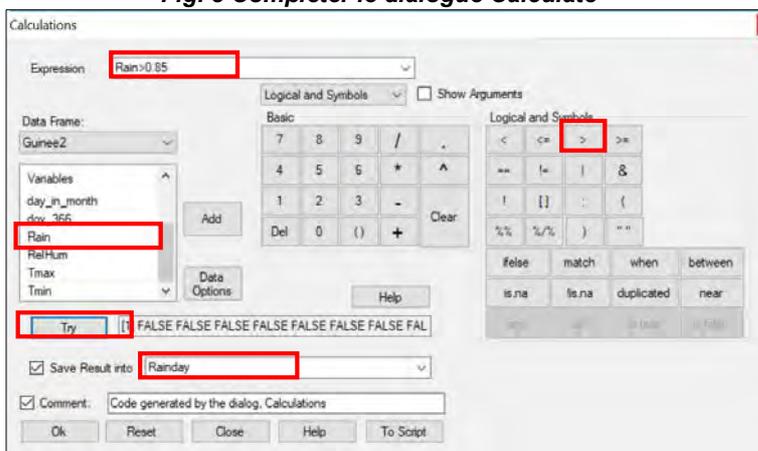


**Fig. 7 Choisir le clavier "logical"**



\*\* Dans la calculatrice, cliquez sur le **menu déroulant** intitulé **Basic** et choisissez le clavier logique (logique), Fig. 7. La calculatrice se développe pour afficher un clavier supplémentaire, Fig. 8

**Fig. 8 Compléter le dialogue Calculate**



**Fig. 9 La nouvelle variable**

	Rain	RelHu	Tmax	Tmin	Rainday (l)
234	0.0	NA	27.3	20.3	FALSE
235	8.2	NA	28.2	21.5	TRUE
236	0.4	NA	27.6	21.1	FALSE
237	0.0	NA	28.2	20.9	FALSE
238	0.0	NA	29.8	20.5	FALSE
239	1.7	NA	29.5	21.2	TRUE
240	2.2	NA	28.7	21.5	TRUE
241	6.0	NA	26.2	21.0	TRUE
242	0.3	NA	28.8	20.9	FALSE
243	17.9	NA	23.6	19.4	TRUE
244	25.3	NA	30.3	19.6	TRUE
245	0.5	NA	27.8	20.7	FALSE

\*\* Comme présenté sur la Fig. 8 sélectionnez la variable intitulée **Rain**. Double-cliquez, ou appuyez sur **Add button** pour l'ajouter à l'expression.

\*\* Appuyez sur le symbole **>** ensuite sur **0.85**, Fig. 8. L'expression devient donc **Rain > 0.85**.

\*\* Appuyez sur **Try**, Fig. 8, pour vous assurer de la validité de votre expression.

Ceci devrait montrer FAUX pour les premières rangées de données, Fig. 8. Ceci est dû au fait que ce sont des jours secs, c'est-à-dire que la quantité de pluie est égale à 0, et est donc inférieure à 0,85 mm. Si vous obtenez: **"Command produced an error or no output to display."** ("La commande a produit une erreur ou aucune sortie à afficher."), vous devrez corriger l'expression.

\*\* Enregistrez les résultats dans une colonne intitulée **Rainday**, Fig, 8 puis appuyez sur **OK**.

Le résultat s'affiche sur Fig. 9 pour un certain nombre de jours.

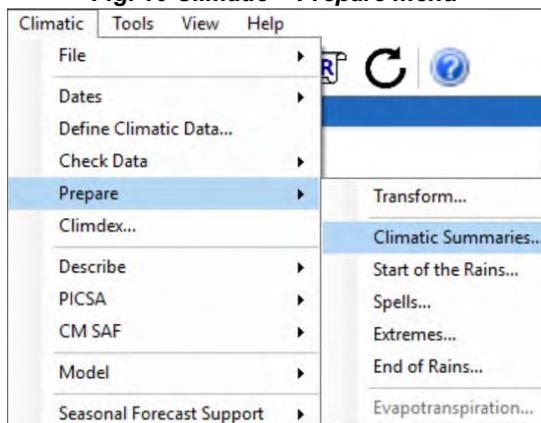
**Faites défiler les données vers le bas** pour confirmer, Fig. 9, que lorsqu'il y a une pluie de plus de 0,85 mm, la nouvelle colonne dit VRAIE. Sinon, elle dit FAUX. C'est un exemple de variable logique de la langue R. Il peut être utilisé dans les calculs, quand TRUE est interprété comme 1, et FALSE = 0.

Ainsi, comme nous le montrons ci-dessous, R-Instat peut totaliser le nombre de VRAIES valeurs chaque année, pour donner le nombre de jours de pluie.

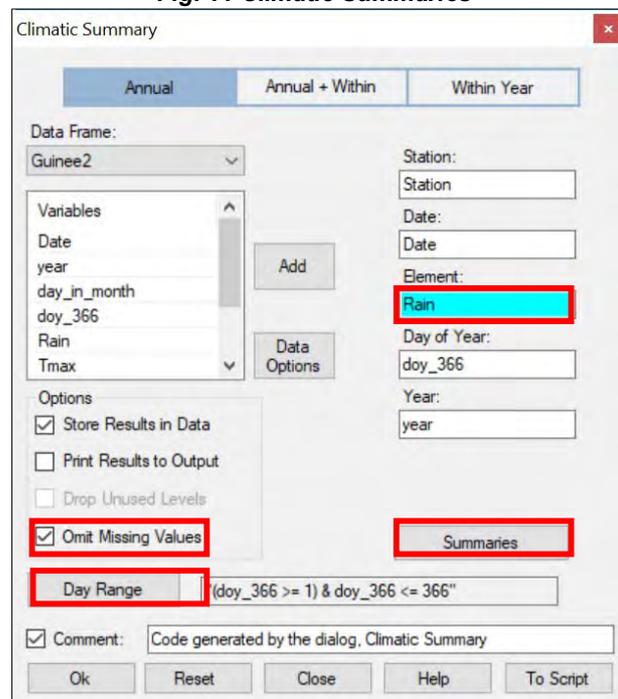
Now move to the Climatic > Prepare menu in R-Instat

\*\* Select **Climatic > Prepare > Climatic Summaries**, Fig. 10.

**Fig. 10 Climatic > Prepare menu**



**Fig. 11 Climatic Summaries**



\*\* Sélectionnez la colonne Rain (Pluie) comme l'élément, Fig. 11.

\*\* Cliquez sur la case à cocher **Omit Missing Values** pour omettre les valeurs manquantes. (Sinon, le processus résumé est annulé chaque fois qu'il y a un jour manquant pendant l'année.)

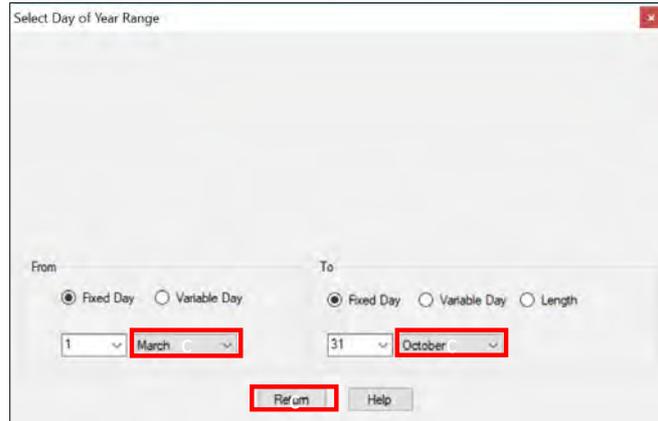
\*\* Appuyez sur le bouton **Summaries** pour produire le sous-dialogue présentée sur la figure 12.

\*\* Décocher **N Non missing and N Total** et cocher **N Missing**, Fig. 12

**Fig. 12** Sous-dialogue de « Summaries »



**Fig. 13** Sous dialogue "Day of Year"



\*\* Appuyez sur **Return** pour revenir au dialogue principal, Fig. 11.

Si l'objectif était d'obtenir les totaux pluviométriques pour toute l'année, il faudrait cliquer sur OK des maintenant. Mais nous cherchons les totaux saisonniers soit de mars à octobre.

\*\* Donc, cliquez sur **Day Range** sur la Fig. 11.

\*\* Changez le **From(De)** mois en **mars** et le mois de **To(A)** jusqu'en **octobre**, Fig. 13.

\*\* Appuyez sur **Return** pour revenir au dialogue principal, Fig. 11; puis appuyez sur **OK**.

Une nouvelle base de données est produite, avec 116 lignes soit l'ensemble des années des deux stations.

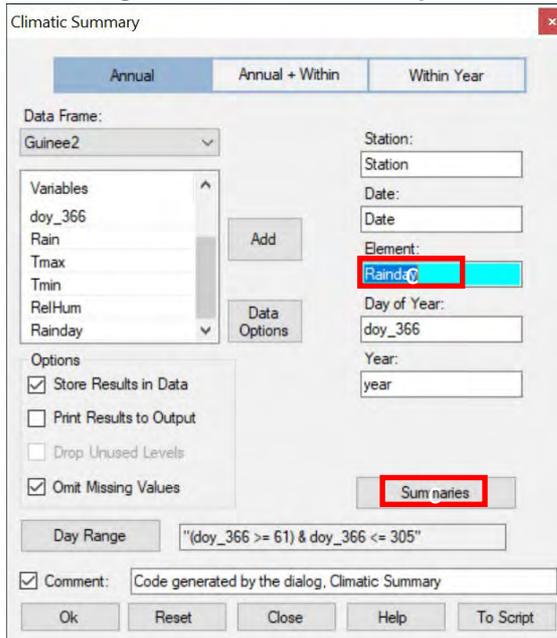
\*\* Retournez au dernier dialogue – appuyez a nouveau sur **Climatic> Préparer> Climatic Summaries**.

\*\* Changez l'élément en **Raindays**, Fig. 14.

\*\* Appuyez sur **Summaries**, Fig. 14 et décochez la case pour produire **N Missing**.

\*\* Appuyez sur **Return** dans le sous-dialogue, puis sur **OK**.

**Fig. 14 Les resumes climatiques**



**Fig. 15 Les resumes produits**

Data View					
	Station (f)	year	count_mis	sum_Rain	sum_Rainday
66	Kankan	2014	0	1313	91
67	Kankan	2016	0	1333	92
68	Kankan	2017	0	494	32
69	Koundara	1970	0	849	61
70	Koundara	1971	0	994	63
71	Koundara	1972	154	186	10
72	Koundara	1973	61	902	63
73	Koundara	1974	31	1264	71
74	Koundara	1975	245	0	0
75	Koundara	1976	0	1238	66
76	Koundara	1977	0	732	54
77	Koundara	1978	31	886	52
78	Koundara	1979	61	917	66
79	Koundara	1980	0	1194	62
80	Koundara	1981	0	1103	85
81	Koundara	1982	32	1142	65

Showing 116 of 116 rows | Showing 5 of 5 columns

Les résultats sont présentés sur la Fig. 15. Par exemple, en 2016, Kankan a enregistré un total de 1333 mm entre mars et octobre après 92 jours de pluie.

Les premières années à Koundara ont aussi beaucoup de valeurs manquantes. Le Climatic Summaries propose une option pour gérer cela. Mais nous vous en présentons une autre ici.

\*\* Utilisez à nouveau le dialogue **Prepare > Column : Calculate > Calculations**. Vous pouvez **recall the last 10 elements** (rappeler les 10 derniers éléments) sur le toolbar.

Complétez le calcul comme indiqué sur la figure 17. Ensuite, toute année comptant plus de 29 jours manquants aura comme résultat le mot « missing » (manquant).

Cet exercice n'est pas si facile, donc les chiffres sont inclus dans la figure 17 pour fournir l'ordre, comme suit:

\*\* Étape 1: Vérifiez que vous êtes sur le bon bloc de données. C'est **Guinee2 by Station\_year**.

\*\* Étape 2: Appuyez sur la fonction **ifelse**.

\*\* Étape 3: Sélectionnez la colonne appelée **count\_missing\_rain**.

\*\* Étape 4: Appuyez sur le signe ">" et ajoutez **29** à la formule.

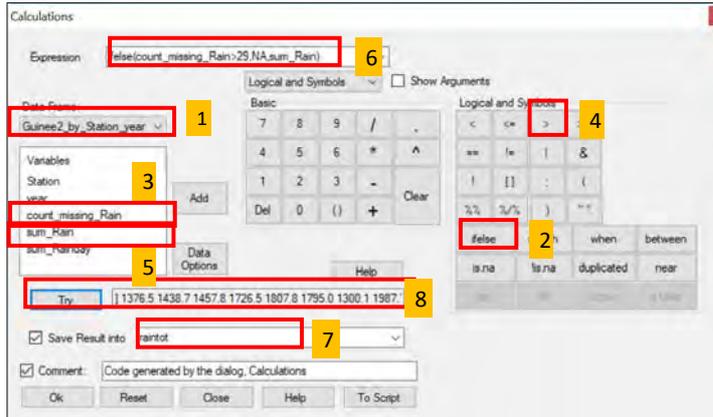
\*\* Tapez **<comma>** puis **NA** puis une autre **<comma>**.

\*\* Étape 5: Sélectionnez la colonne appelée **sum\_Rain**.

\*\* Étape 6: Vérifiez que la formule est maintenant **ifelse (count\_missing\_Rain> 29, NA, sum\_Rain)**.

\*\* Étape 7: Intitulez la colonne résultante « **raintot** ».

**Fig. 16 Setting summary data to missing**



**Fig.17 Resulting data**

Station (f)	year	count_	sum_Rai	sum_Rai	raintot	rainday	meanrai
63 nkan	2012	0	1375	91	1375	91	15.1
64 nkan	2013	0	1233	80	1233	80	15.4
65 nkan	2014	31	878	61	NA	NA	NA
66 nkan	2015	0	1313	91	1313	91	14.4
67 nkan	2016	0	1333	92	1333	92	14.5
68 nkan	2017	0	494	32	494	32	15.4
69 undara	1970	0	849	61	849	61	13.9
70 undara	1971	0	994	63	994	63	15.8
71 undara	1972	154	186	10	NA	NA	NA
72 undara	1973	61	902	63	NA	NA	NA
73 undara	1974	31	1264	71	NA	NA	NA
74 undara	1975	245	0	0	NA	NA	NA
75 undara	1976	0	1238	66	1238	66	18.8
76 undara	1977	0	732	54	732	54	13.6
77 undara	1978	31	886	52	NA	NA	NA
78 undara	1979	61	917	66	NA	NA	NA

\*\* Étape 8: Cliquez sur le bouton **Try** pour vérifier que la commande donne des résultats et non une erreur.

\*\* Appuyez sur **OK** pour produire une nouvelle colonne.

Maintenant, la même chose est faite pour le nombre de jours de pluie.

\*\* Récupérez le dialogue - utilisez la fleche semi-circulaire sur le toolbar.

\*\* Dans la formule, le point 6 de la figure 17 transforme la variable **sum\_Rain** en **sum\_Rainday**. (Habituellement, nous vous suggérons d'éviter de taper dans ce champ, mais il est tentant de taper simplement le mot « day » ici.)

\*\* Renommez le (Étape 7), appelez le « **Rainday** »

\*\* Appuyez sur **OK**.

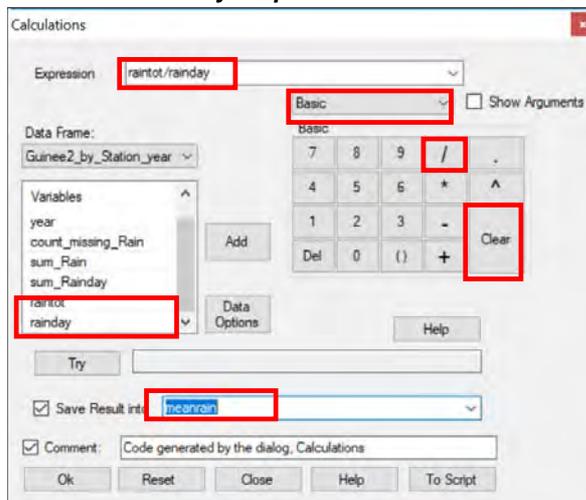
Avec les totaux pluviométriques et le nombre de jours, la pluie moyenne par jour de pluie peut également être calculée facilement.

\*\* Récupérez le dernier dialogue (ou utiliser à nouveau **Prepare>Column>Calculate >Calculations**).

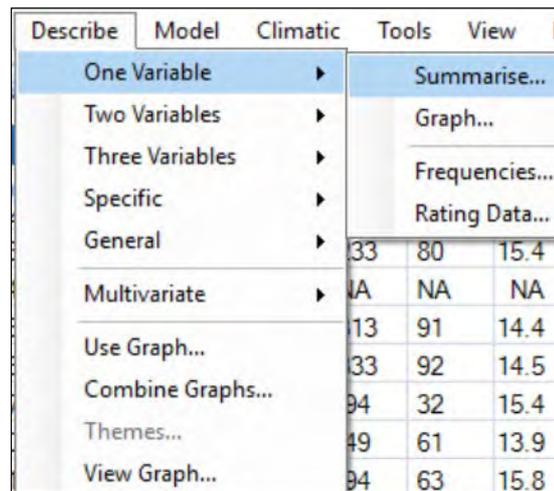
\*\* Remettre la calculatrice sur **Basic**, Fig. 18.

\*\* Appuyez sur le bouton **Clear** (Effacer)

**Fig. 18 Calculer la moyenne quotidienne d'un jour pluvieux**



**Fig. 19 Menu Summarise: Une variable**



\*\* Entrez la formule **raintot / rainday** (figure 18) – pas besoin de taper, il suffit de cliquer.

\*\* Nommez la colonne résultante **meanrain** et appuyez sur **Ok**.

Ces données récapitulatives sont traitées dans les sections suivantes de ce guide.

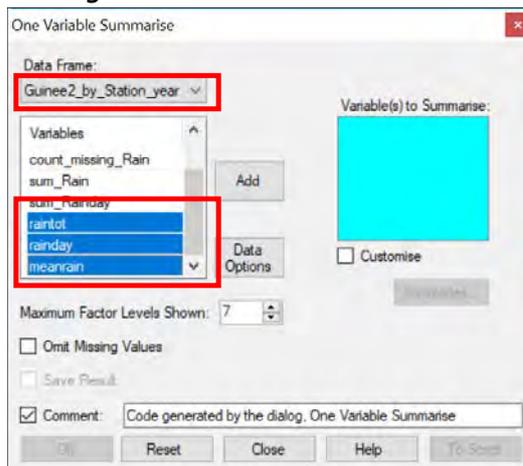
\*\* Pour l'instant, utilisez **Describe > One Variable > Summarize**, Fig. 19, pour avoir une idée initiale des colonnes produites.

\*\* Vérifiez que vous traitez le data frame que nous cherchons, c'est-à-dire les valeurs récapitulatives annuelles.

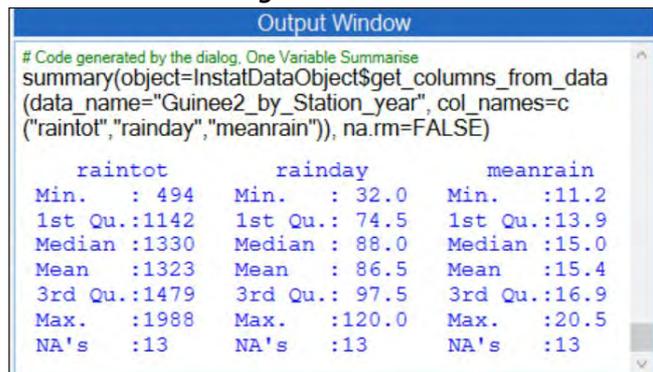
\*\* Sélectionnez les 3 colonnes comme indiqué sur la figure 20.

\*\* Appuyez sur OK.

**Fig. 20 One Variable Summarise**



**Fig. 21 Results**



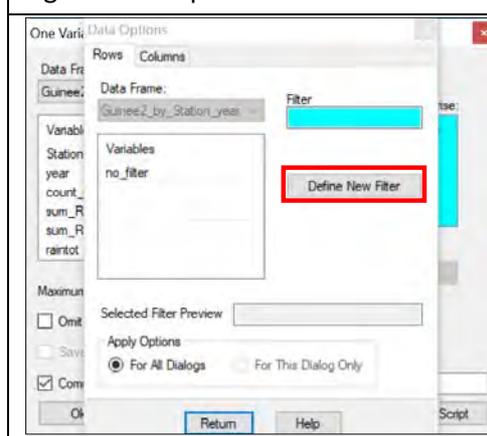
Les résultats sont présentés sur la Fig. 21. Par exemple, 120 était le nombre maximum de jours de pluie dans l'année.

Ces résultats ne sont pas très utiles, car ils traitent les deux stations en même temps.

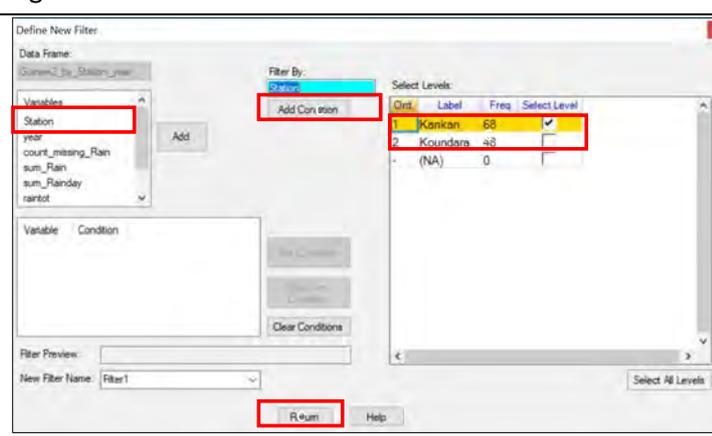
\*\* Retournez au dialogue **Summarise**.

\*\* Appuyez sur **Data Options**, voir Fig. 20.

**Fig. 22 Data Options for a filter**



**Fig. 23 Filter for one station**



\*\* Sélectionnez la colonne **Station**, Fig. 23

\*\* La condition est correcte pour sélectionner uniquement Kankan, donc appuyez sur **Add condition**.

\*\* Appuyez ensuite sur **Return** pour revenir au sous-dialogue de la figure 22.

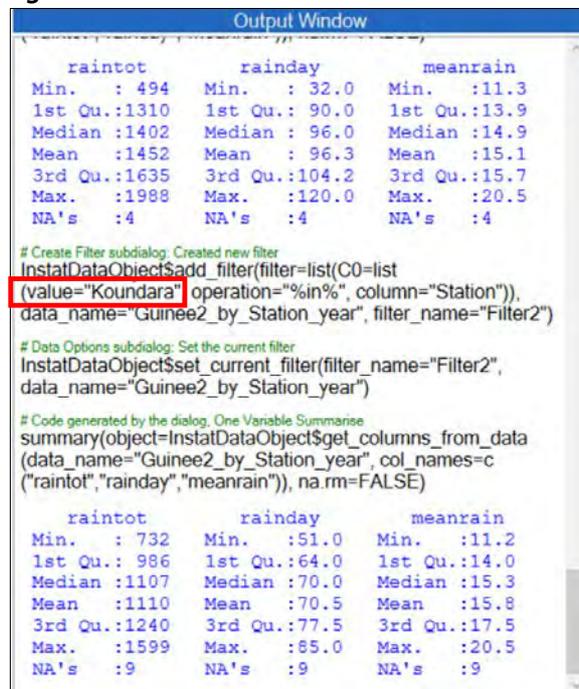
\*\* Appuyez à nouveau sur **Return** pour revenir à la boîte de dialogue Résumer.

\*\* Appuyez sur **OK** pour obtenir les résultats pour Kankan seulement.

\*\* Répétez les Data Options, pour sélectionner Koundara in Fig. 23.

\*\* Retournez au dialogue principal, et appuyez sur **Back** ensuite OK pour obtenir les résultats provenant uniquement de Koundara, see Fig. 24

**Fig. 24 Résultats de stations individuelles**



```
Output Window

  raintot      raintot      raintot
Min.   : 494      Min.   : 32.0      Min.   :11.3
1st Qu.:1310      1st Qu.: 90.0      1st Qu.:13.9
Median :1402      Median : 96.0      Median :14.9
Mean   :1452      Mean   : 96.3      Mean   :15.1
3rd Qu.:1635      3rd Qu.:104.2      3rd Qu.:15.7
Max.   :1988      Max.   :120.0      Max.   :20.5
NA's   : 4         NA's   : 4         NA's   : 4

# Create Filter subdialog: Created new filter
InstatDataObject$add_filter(filter=list(C0=list
(value="Koundara" operation="%in%", column="Station")),
data_name="Guinee2_by_Station_year", filter_name="Filter2")

# Data Options subdialog: Set the current filter
InstatDataObject$$set_current_filter(filter_name="Filter2",
data_name="Guinee2_by_Station_year")

# Code generated by the dialog: One Variable Summarise
summary(object=InstatDataObject$get_columns_from_data
(data_name="Guinee2_by_Station_year", col_names=c
("raintot","raintot","raintot"), na.rm=FALSE)

  raintot      raintot      raintot
Min.   : 732      Min.   : 51.0      Min.   :11.2
1st Qu.: 986      1st Qu.: 64.0      1st Qu.:14.0
Median :1107      Median : 70.0      Median :15.3
Mean   :1110      Mean   : 70.5      Mean   :15.8
3rd Qu.:1240      3rd Qu.: 77.5      3rd Qu.:17.5
Max.   :1599      Max.   : 85.0      Max.   :20.5
NA's   : 9         NA's   : 9         NA's   : 9
```

## Recapitulatif

Dans cette section, vous avez produit des résumés annuels à partir de données quotidiennes. Il s'agissait de totaux pluviométriques, mais cela aurait pu être quelque chose d'autre, telles que les moyennes de température ou les extrêmes. Ces résumés ont été placés dans un 2<sup>e</sup> data frame et ont ensuite été analysés davantage. La calculatrice a été utilisée pour produire d'autres colonnes (variables) également au niveau annuel. Les analyses étaient pour les deux stations combinées. Les installations de filtrage permettent alors d'obtenir des résultats pour des stations individuelles.

## Le début et la fin de la saison des pluies

Si vous avez complété les analyses ci-dessus, un filtre est toujours en cours de fonctionnement, Fig. 1, donc seules les données de Koundara sont visibles. Sur la figure 1, la première ligne est en rouge, ce qui confirme l'utilisation d'un filtre.

\*\* Faites un **clic droit** avec le curseur dans le champ de nom, Fig. 1.

\*\* Prendre la dernière option, **Remove the Last Filter** (pour supprimer le filtre actuel).

Début des pluies

\*\* Choisissez Climatique> Préparer> Début des pluies, Fig. 2.

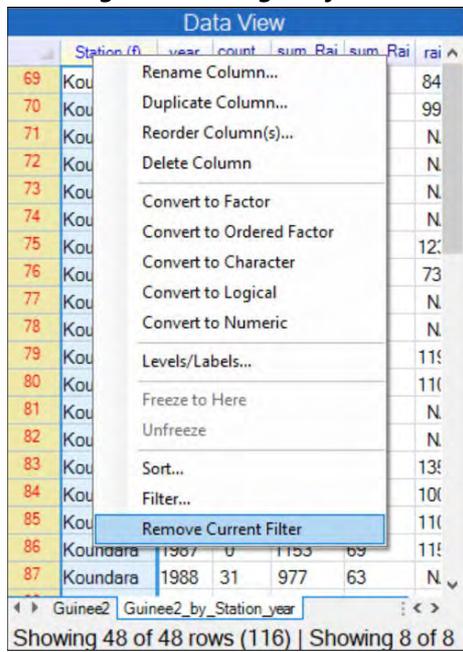
C'est l'élément de menu juste en dessous des résumés climatiques utilisés dans la dernière section

## Start of the rains

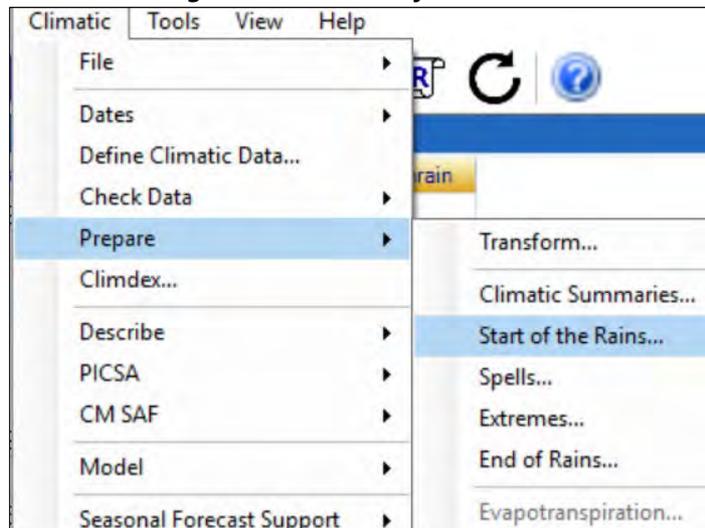
\*\* Sélectionnez **Climatic > Prepare > Start of the Rains**, Fig. 2.

C'est l'élément de menu juste en dessous des Climatic Summaries utilisées récemment.

**Fig. 1 Removing the filter**



**Fig. 2 Choose Start of the rains**



\*\* Dans la Fig. 3, vérifiez que le fichier **Guinea2** de données quotidiennes est utilisé. Notez que les 5 champs de la Station au jour de l'année ont été automatiquement remplis.

\*\* Cliquez sur **Day Range**, Fig. 3 pour afficher le sous-dialogue, Fig. 4.

\*\* Dans la figure 4, modifiez la date **From** : choisissez **15 avril**. Les cultures sont rarement plantées plus tôt que ça.

\*\* Changer la date **To**, choisissez **le 31 juillet**. Nous supposons que la plupart des agriculteurs auront planté d'ici-là.

\*\* Appuyez sur **Return**.

**Fig. 3 Le debut des pluie**

**Fig. 4 Selectionnez une date de semis précoce**

\*\* Dans la Fig. 3, cliquez sur la case **Total Rainfall** et modifiez le nombre de jours **de 2 à 3**.

\*\* Appuyer sur **OK**.

Cela ajoute une autre colonne, intitulée start, dans le data frame avec les données annuelles.

Cela représente une potentielle date de début des pluies chaque année.

Maintenant, préparons une deuxième définition qui représentera également une période de sécheresse.

\*\* Retournez au dialogue précédent.

\*\* Cliquez sur la case à cocher pour ajouter une « période seche », Fig. 5

\*\* Modifier le spell length (longueur de la periode seche) à 7 jours et la durée de cette période à 21 jours.

\*\* Nommez la colonne résultante s « startdry », Fig. 5.

\*\* Appuyer sur **OK**.

**Fig. 5 Ajouter une condition de période sèche**

Start of Rains

Data Frame: Guinee2

Station: Station

Date: Date

Rain Column: Rain

Year: year

Day of Year: doy\_366

Threshold: 0.85

Day Range: "doy\_366 >= 106 & doy\_366 <= 213"

Conditions for Start of Rains

Total Rainfall Over Days: 3 Calculate Rainfall Value by:

Number of Rainy Days

Dry Spell Maximum Dry Days: 7 Overall Interval Length: 21

Dry Period

New Column Name: startdry

**Fig. 6 Difference between start dates**

Calculations

Expression: startdry-start

Data Frame: Guinee2\_by\_Station\_year

Save Result into: startdiff

Comment: Code generated by the dialog, Calculations

Calculez maintenant la différence entre ces deux colonnes

\*\* Utilisez **Prépare > Column: Calculate > Calculations** pour obtenir la calculatrice R-Instat

\*\* Sur la Fig. 6, l'expression est **startdry - start**.

\*\* Appelez cette nouvelle colonne **startdiff**.

\*\* Appuyez sur **OK**.

Cela ajoute une colonne supplémentaire aux données. Nous interprétons le résultat que lorsque la colonne startdiff est nulle, c'est-à-dire que la condition de la période sèche n'a pas d'effet, la première date de début est correcte. Dans le cas contraire, il y a eu une période sèche de plus de 7 jours au cours des 3 premières semaines (21 jours) après la semis, et il a donc fallu replanter.

## La fin de la saison des pluies

\*\* Utilisez **Climatic > Prepare > End of the Rains**, Fig. 7.

\*\* Sur la figure 8, confirmez que le data frame est bien Guinee2, c'est-à-dire les données quotidiennes. Les contrôles sur le côté droit ont été alimentés automatiquement.

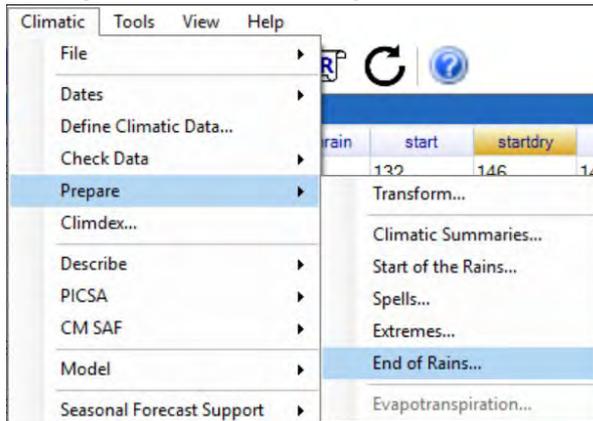
\*\* Appuyez sur le bouton Day Range pour lancer le sous-dialogue.

\*\* Changez la date la plus proche au 1er septembre. Appuyez sur Return.

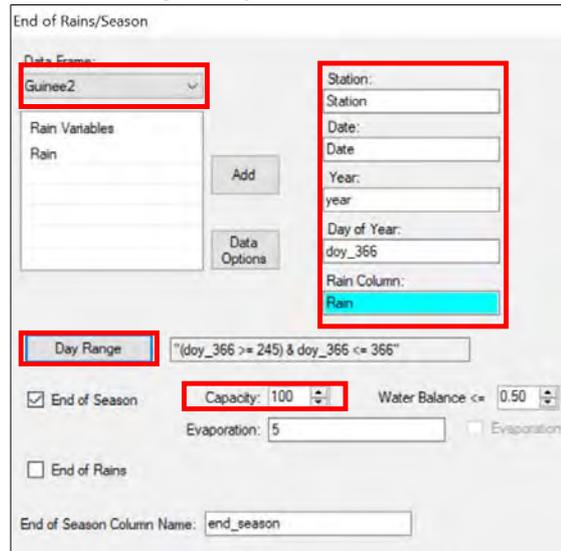
\*\* Cliquez sur la case End of Season. Tapez une capacité de 100mm.

\*\* Appuyez sur **OK**

**Fig. 7 Sélectionner de la fin de la saison**



**Fig. 8 La fin de la saison**



Nous utilisons un modèle d'équilibre hydrique simplifié pour calculer la date de la fin de la saison.

Maintenant il s'agit de soustraire la date de début de la date de fin, pour obtenir la longueur de la saison.

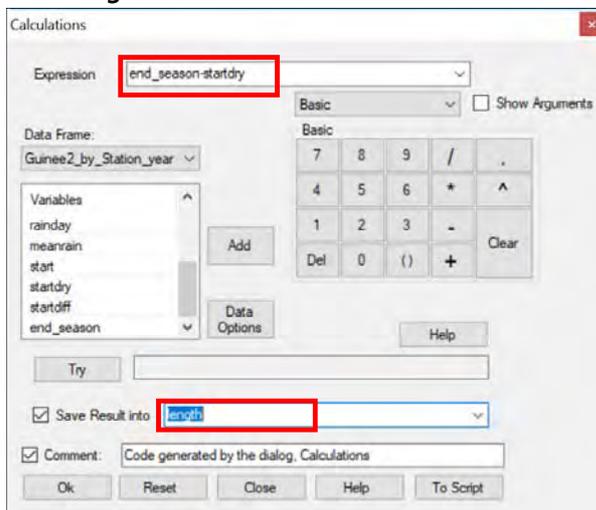
\*\* Utilisez **Prepare > Column: Calculate > Calculations**.

\*\* Sur la figure 9, l'expression est comme suit : **end\_season - startdry**.

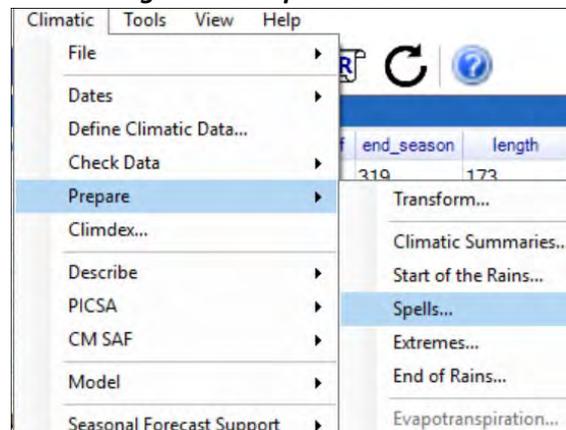
\*\* Appelez la nouvelle colonne « **length** » (duree).

\*\* Appuyer sur **OK**.

**Fig. 9 Calculer la duree de la saison**



**Fig. 10 Menu periode seche**



## Duree des periodes seches

Calculez maintenant la durée du plus long épisode de sécheresse de la saison. Cela pourrait être pour un nombre de mois fixe, par ex. Juillet à septembre. Ici, nous calculons le plus long épisode de sécheresse entre les dates de début et de fin de la saison, qui elles même changent chaque année.

\*\* Sélectionnez **Climatic > Prepare > Spells**, Fig. 10.

**Fig 11 Duree de la periode seche**

**Fig. 12 Choisir la gamme de jour**

- \*\* Dans le dialogue, vérifiez que les données quotidiennes - **Guinée2** sont utilisées.
- \*\* Ajouter la colonne **Rain** comme élément, Fig. 11.
- \*\* Appuyez sur le bouton **Day Range**, Fig. 11.
- \*\* Dans le sous-dialogue, sur **From** cliquez sur **Variable Day**, Fig. 12.
- \*\* Vérifiez que cette trame de données contient les **summary data**, c'est-à-dire Guinea2\_by\_Station\_year, Fig. 12.
- \*\* Choisissez **startdry** comme colonne de départ.
- \*\* Dans la section **To**, utilisez également le paramètre **Variable Day** et sélectionnez la colonne appelée **end\_season**.
- \*\* Appuyez sur **Return**.
- \*\* Appuyez sur **OK**

**Fig. 13 The summary data**

Data View														
	Station (f)	year	count_mis	sum_Rain	sum_Rainday	raintot	rainday	meanrain	start	startdry	startdiff	end_season	length	spells
1	Kankan	1950	0	1376	93	1376	93	14.8	132	146	14	319	173	7
2	Kankan	1951	0	1439	107	1439	107	13.4	109	109	0	335	226	11
3	Kankan	1952	0	1458	110	1458	110	13.3	127	127	0	315	188	7
4	Kankan	1953	0	1726	109	1726	109	15.8	118	159	41	321	162	8
5	Kankan	1954	0	1808	116	1808	116	15.6	128	128	0	342	214	13
6	Kankan	1955	0	1795	120	1795	120	15.0	115	138	23	331	193	11
7	Kankan	1956	0	1300	96	1300	96	13.5	150	150	0	308	158	11
8	Kankan	1957	0	1988	118	1988	118	16.8	108	108	0	330	222	11
9	Kankan	1958	31	1216	85	NA	NA	NA	106	130	24	293	163	6
10	Kankan	1959	0	1762	100	1762	100	17.6	124	138	14	326	188	13
11	Kankan	1960	0	1899	117	1899	117	16.2	118	118	0	324	206	10
12	Kankan	1961	0	1356	97	1356	97	14.0	113	113	0	309	196	12
13	Kankan	1962	0	1870	107	1870	107	17.5	107	107	0	320	213	7
14	Kankan	1963	0	1599	102	1599	102	15.7	116	162	46	319	157	12
15	Kankan	1964	0	1684	111	1684	111	15.2	126	126	0	317	191	10
16	Kankan	1965	0	1403	105	1403	105	13.4	137	137	0	320	183	16
17	Kankan	1966	0	1469	110	1469	110	13.4	137	137	0	322	185	10

Showing 116 of 116 rows | Showing 14 of 14 columns

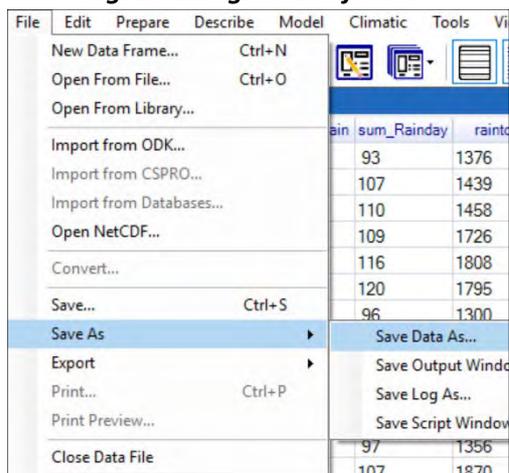
La figure 13 montre qu'il y a maintenant beaucoup de colonnes de données que l'on pourrait examiner davantage. Par exemple, à Kankan en 1950:

- Les précipitations totales, de mars à octobre, ont été de 1376 mm
- Il y a eu 93 jours de pluie et il n'y avait pas de valeurs manquantes dans cette période
- Il y a eu un faux départ qui a eu lieu le jour no. 132, c'est-à-dire le 11 mai.
- Le démarrage réussi a eu lieu 2 semaines plus tard, c'est-à-dire qu'il a fallu replanter.
- La fin de la saison a eu lieu le jour no. 319, c'est-à-dire le 14 novembre
- D'où une durée de la saison de 173 jours, soit près de 6 mois.
- La plus longue période de sécheresse de la saison a été de 7 jours.

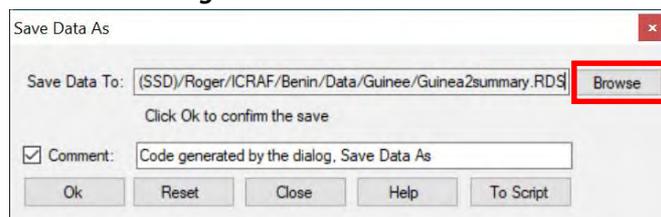
L'analyse complémentaire commence dans la section suivante.

Pour compléter cette section, enregistrez le fichier comme indique ci dessous

**Fig. 14 Enregistrez le fichier**



**Fig. 15 Choisissez un nom**



\*\* Utilisez **File > Save As > Save Data As**, Fig. 14.

\*\* Appuyez sur **Browse**, Fig. 15. Dans la boîte de dialogue suivante, choisissez un répertoire et un nom de fichier appropriés, puis cliquez sur **Save**.

\*\* De retour à la Fig. 15. Cliquez sur **OK**.

Vous pouvez également utiliser **File> Export** si vous le souhaitez. Mais il y a une grande différence entre le fichier exporté et le fichier enregistré. L'exportation est pour un seul data frame, tandis que l'enregistrement (Save As) est pour les données quotidiennes et annuelles en même temps, ce qui permet de poursuivre l'analyse ultérieurement.

# PREPARATION DES DONNEES CLIMATIQUES POUR DES ANALYSES AVEC R-INSTAT

## 1) Introduction

R-instat est un logiciel conçu pour la statistique générale. Toutes les opérations y sont effectuées à par du langage statistique R. En plus, R-instat comporte un menu spécial 'climatique'.

Nous prévoyons à l'avenir, à partir du menu 'climatique' faciliter l'analyse des données climatiques de n'importe quelle échelle, comme par exemple celles d'une station automatique. Les méthodes actuelles sont particulièrement conçues pour l'analyse des données journalières. Ce guide utilise des données journalières de deux (02) stations de la Guinée (Conakry) comportant chacun quatre (04) paramètres. Ces données sont dans la bibliothèque R-Instat, par conséquent les exemples dans ce guide peuvent être suivis par les utilisateurs qui souhaitent le faire.

## 2) Remerciements

Nous exprimons nos remerciements et notre gratitude au Service de la Météorologie Nationale de la Guinée de nous avoir permis d'utiliser leurs données dans la préparation de ce guide, et aussi de permettre que leurs données soient ajoutées à la bibliothèque de R-Instat.

## 3) Mise en forme des données

Nous allons tout d'abord ouvrir les données dans leur format original. Il s'agit de deux (02) fichiers Excel, et donc nous commençons par présenter les données dans Excel, plutôt que dans R-Instat, (voir Fig. 1).

*Fig. 1 Données d'une station sous Excel*

	A	B	C	D	E	F
1		<b>TMIN journalière</b>				
2	Eg gh id	Eg el abbrevi	Year	Month	Day	Value
3	17KKAN1S	TMIN	1950	01	01	13
4	17KKAN1S	TMIN	1950	01	02	15.6
5	17KKAN1S	TMIN	1950	01	03	18
6	17KKAN1S	TMIN	1950	01	04	19.7
7	17KKAN1S	TMIN	1950	01	05	14
8	17KKAN1S	TMIN	1950	01	06	12.4
9	17KKAN1S	TMIN	1950	01	07	11.9
10	17KKAN1S	TMIN	1950	01	08	18.4
11	17KKAN1S	TMIN	1950	01	09	13.2

Les données sont sous le "format" adéquat pour R-Instat, c'est-à-dire une ligne de données pour chaque jour. Les données de cette station commencent en 1950 et se poursuivent jusqu'en 2016 ou 2017.

Les données ne sont pas toujours sous ce "format" adéquat. L'Annexe 1 présente comment transformer les données au format r-instat.

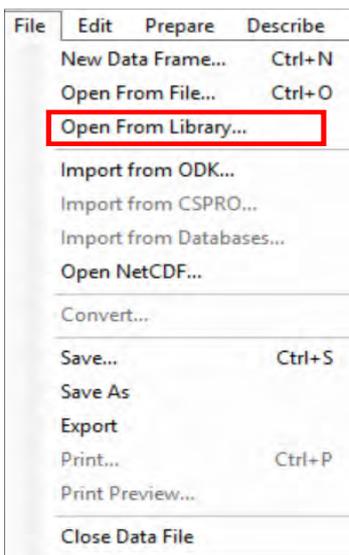
Le fichier Excel de la Figure 1 comporte quatre (04) feuilles, soit un paramètre par feuille. Si l'analyse portait uniquement sur un seul paramètre, et pour cette seule station, ces données pouvaient être importées dans R-Instat comme indiqué ci-dessous, et continuer l'analyse à partir de la section 4. Cependant, fusionner ces 4 feuilles en seul fichier sera bénéfique pour la réalisation de plusieurs analyses. Il est également possible de les combiner avec les données de la 2<sup>e</sup> station. C'est ce que ferons.

Notre objectif est de combiner les quatre paramètres pour les deux stations dans une seule et même feuille de données.

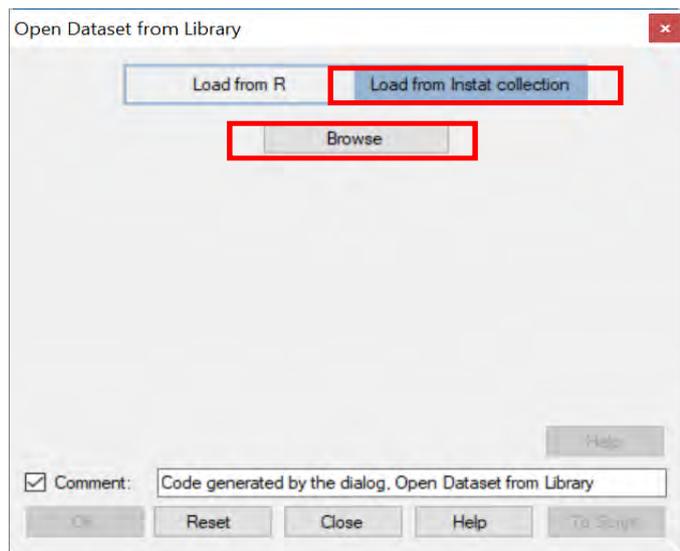
\*\* **Ouvrir** R-instat

\*\* Utiliser le menu **File > Open from library** pour ouvrir le fichier, voir Fig. 2. (Nous utilisons ici les données contenues dans la bibliothèque de r-instat. Pour vos propre données, utilisez plutôt le menu **File > Open**).

**Fig. 2 File > Open from Library**



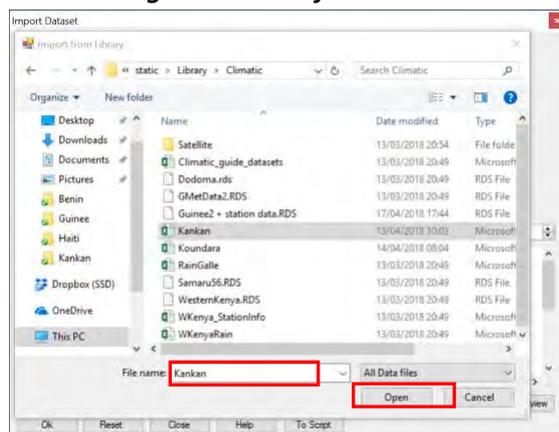
**Fig. 3 Charger les données à partir de collection de Instat**



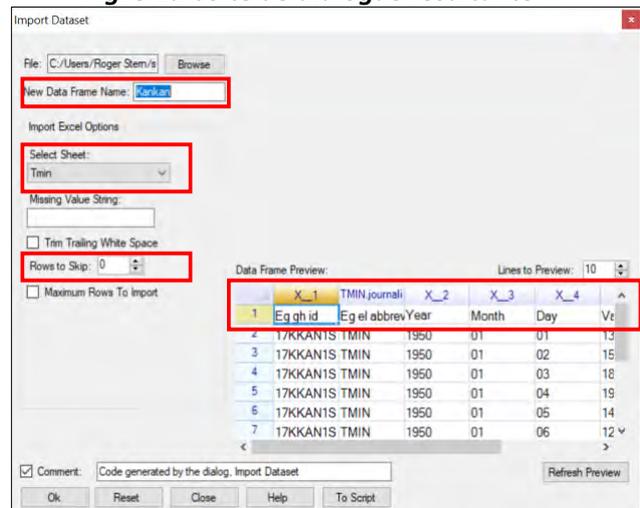
\*\* Cliquez sur **Load from the Instat collection**, Fig. 3

\*\* Cliquez sur **Browse**, ensuite choisissez **Climatic** et **Guinée**

**Fig. 4 Choisir le fichier Excel**



**Fig. 5 La boîte de dialogue résultante**



\*\* Choisissez le fichier **Kankan.xlsx**, et ensuite cliquez sur **Open**, Fig. 4.

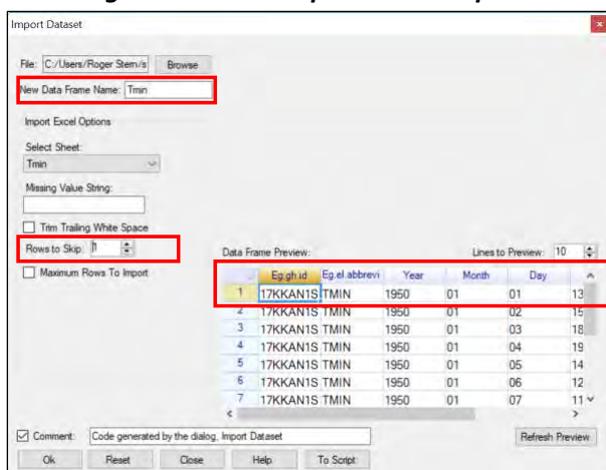
\*\* Examinez la boîte de dialogue Fig. 5. **NE PAS CLIQUER SUR OK** pour l'instant. La feuille Excel n'est pas encore prête pour être importée.

En examinant la feuille Excel de plus près, on constate que la première ligne est un en-tête, et les noms des variables sont dans la ligne 2 de la feuille.

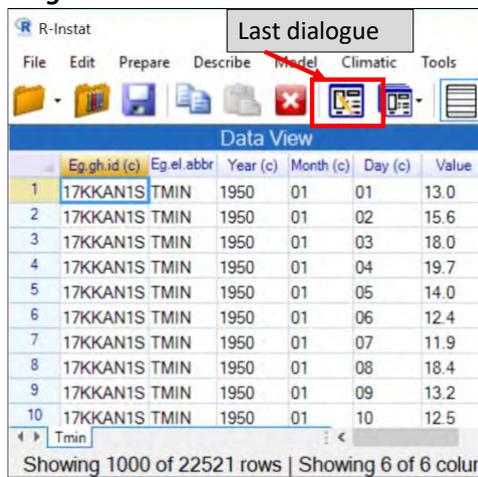
\*\* Modifier **Row to skip** de **0** à **1**, Fig. 5.

\*\* Remplacer dans la section **New Data Frame Name** 'Kankan' par **Tmin**. Le résultat est semblable à la Fig. 6.

**Fig. 6 Les données prêtent à l'import**



**Fig. 7 Les données dans R-Instat**



\*\* Cliquez sur **OK** pour importer les données dans R-instat, Fig.7.

Le bas de la figure 7 indique qu'il y a 22 521 jours (lignes de données), dont seulement 1000 sont représentées. La grille dans R-Instat est juste une fenêtre montrant une partie des données. L'ensemble des données est stocké dans R.

\*\* Cliquez sur le bouton **Last Dialog**, Fig.7. Cela vous ramène à la Figure 6.

\*\* Changez de feuille dans la section **Select Sheet**,

\*\* Choisissez **Tmax** et cliquez sur **OK**.

Vous avez maintenant deux (02) feuilles de données ouvertes dans R-instat. La feuille de Tmax comporte 22215 lignes de données.

\*\* Répétez le même processus 2 fois pour importer les données **Rain** et **RelHum**. Vous avez maintenant 4 feuilles de données dans r-instat, Fig. 8.

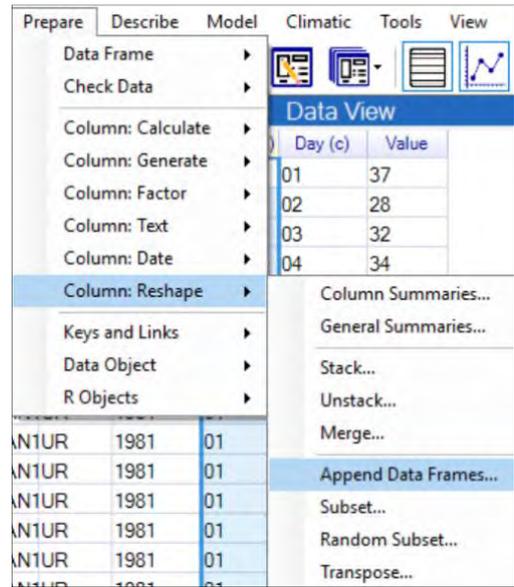
Il y a 24080 lignes de données sur la pluie et 11045 lignes pour l'humidité relative.

**Fig. 8 Les 4 feuilles de données dans R-Instat**

Data View						
	Eg.gh.id	Eg.el.ab	Year (c)	Month (c)	Day (c)	Value
1	17KKAN1UR	1981	01	01	37	
2	17KKAN1UR	1981	01	02	28	
3	17KKAN1UR	1981	01	03	32	
4	17KKAN1UR	1981	01	04	34	
5	17KKAN1UR	1981	01	05	40	
6	17KKAN1UR	1981	01	06	37	
7	17KKAN1UR	1981	01	07	35	
8	17KKAN1UR	1981	01	08	38	
9	17KKAN1UR	1981	01	09	33	
10	17KKAN1UR	1981	01	10	33	

Showing 1000 of 11045 rows | Showing 6 of 6

**Fig. 9 Le menu Prepare**

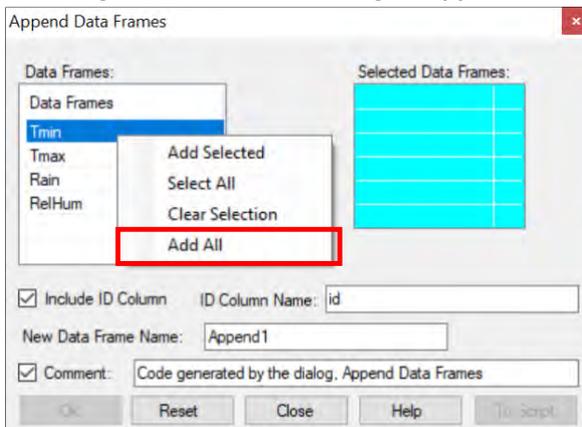


La tâche suivante consiste à fusionner les données des différentes feuilles dans une seule feuille de données. Ceci est fait en 2 étapes.

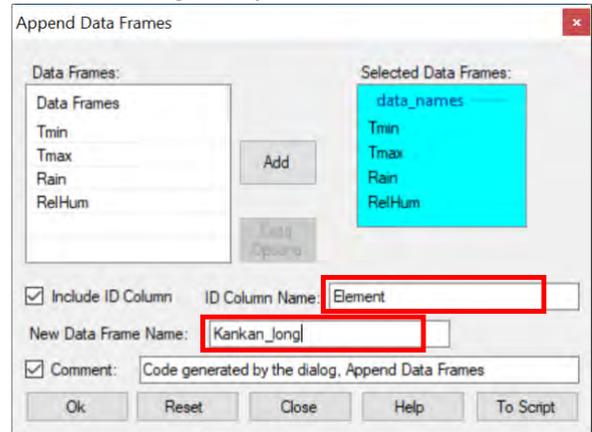
Étant donné que nous sommes toujours dans la préparation des données, nous utiliserons le menu **Prepare**, Fig. 9.

\*\* Dans le menu **Prepare**, choisir **Column: Reshape**, ensuite **Append Data Frame**, Fig. 9.

**Fig. 10 La boîte de dialogue Append**



**Fig. 11 Ajout terminé**



Nous voulons sélectionner tous les paramètres (feuilles de données).

\*\* **Faites un clic droit** dans le sélecteur de données, Fig. 10, et cliquez sur **Add All**.

\*\* Changez **ID Column Name** en **Element**, puis **Data Frame Name** en **Kankan\_long**. Fig. 11.

\*\* **Cliquez** sur **OK**.

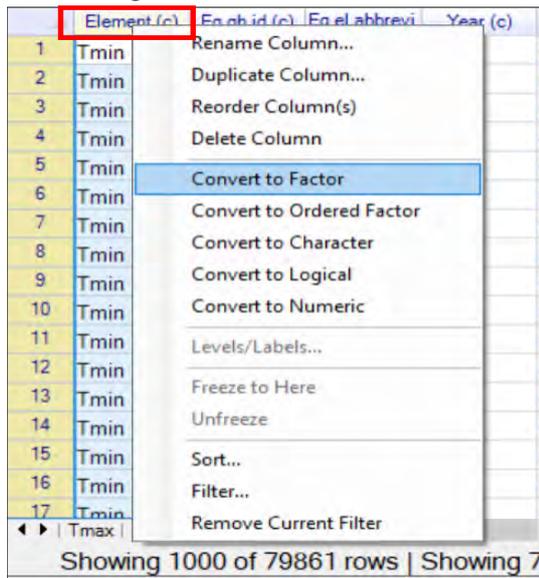
Cela a produit une nouvelle feuille de données contenant 79 861 lignes.

La première colonne, appelée **Elément** est type texte (caractère). Elle doit être convertie en une colonne de type catégorique, ce qui correspond au type Facteur dans R.

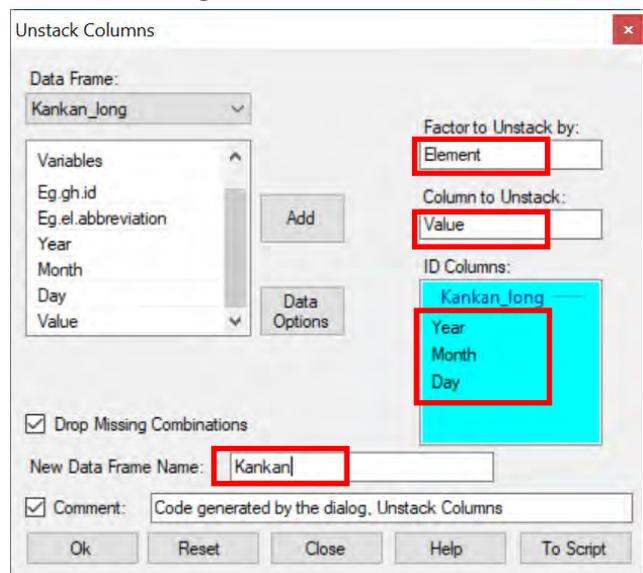
\*\* Cliquez avec le bouton droit sur le nom **Element**. On a le menu déroulant la figure 12.

\*\* Choisir Convert to Factor. Un (f) devrait apparaître après le nom de la colonne.

**Fig 12. Le menu du clic droit**



**Fig. 13 Unstack**



\*\* Allez à **Prepare > Column: Reshape > Unstack**, Fig. 9.

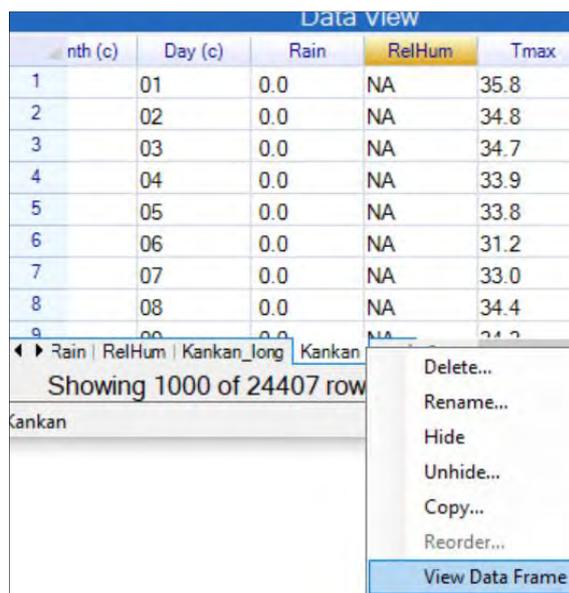
\*\* Terminer le dialogue **Unstack** comme indiqué sur la figure 13.

\*\* La case **Factor to Unstack** contient la colonne **Element**, tandis que **Column to Unstack** contient **Value**.

\*\* Il y a 3 **ID Columns** (Identifiants des colonnes) nommés, **Year**, **Month**, **Day**, et le nom de la nouvelle feuille (**New Data Frame Name**) est **Kankan**.

\*\* Cliquez sur OK.

**Fig. 14 Clic droit sur la feuille de données**



**Fig. 15 Les données de Kankan dans la vue de R**

Year	Month	Day	Rain	RelHum	Tmax	Tmin	
21781	2010	01	20	0.0	36	35.4	15.3
21782	2010	01	21	0.0	37	36.0	14.2
21783	2010	01	22	0.0	22	36.3	15.5
21784	2010	01	23	0.0	31	36.5	15.1
21785	2010	01	24	0.0	34	36.1	15.6
21786	2010	01	25	0.0	31	36.4	16.5
21787	2010	01	26	0.0	31	36.0	19.5
21788	2010	01	27	0.0	33	36.1	16.7
21789	2010	01	28	0.0	38	36.0	21.0
21790	2010	01	29	0.0	34	37.4	20.4
21791	2010	01	30	0.0	33	38.4	16.7
21792	2010	01	31	0.0	33	38.0	15.4
21793	2010	02	01	0.0	30	38.0	16.0
21794	2010	02	02	0.0	26	38.0	11.0
21795	2010	02	03	0.0	30	38.5	15.5
21796	2010	02	04	0.0	30	39.0	16.9
21797	2010	02	05	0.0	44	38.1	18.0
21798	2010	02	06	0.0	24	37.8	17.4
21799	2010	02	07	0.0	27	36.8	17.1

La grille, ou encore feuille de r-instat est une fenêtre montrant juste une partie des données.

\*\* Faites un clic droit sur le nom **Kankan**, ensuite cliquez sur **View Data Frame**. Fig. 14.

\*\* Faites défiler les données dans la vue de R, pour voir les données complètes. Fig. 15.

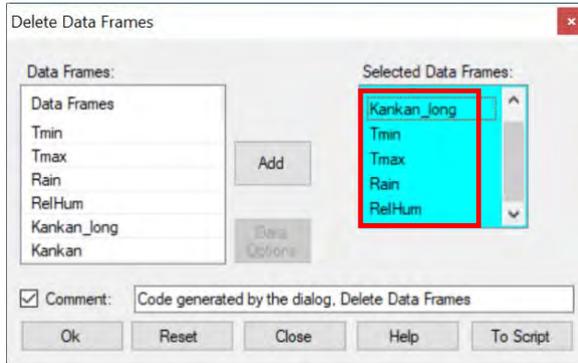
Nous pouvons fermer les autres feuilles de données, à l'exception de celle nommée **Kankan**. Celle elle que l'on utilisera pour la suite de notre analyse.

**\*\* Sélectionnez** la feuille **Kankan\_long**

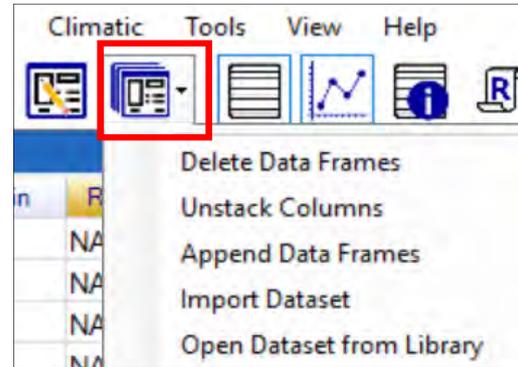
**\*\* Faites un clic droit** au bas de la fenêtre (Fig. 14), et cliquez sur **Delete**, Fig. 16.

**\*\* Ajouter** les autres feuilles à supprimer, sauf bien sûr celle de Kankan, Fig. 16.

**Fig. 16 Suppression des feuilles non désirées**



**Fig. 17 Les dialogues récemment utilisés**



**\*\* Cliquez** sur **OK**.

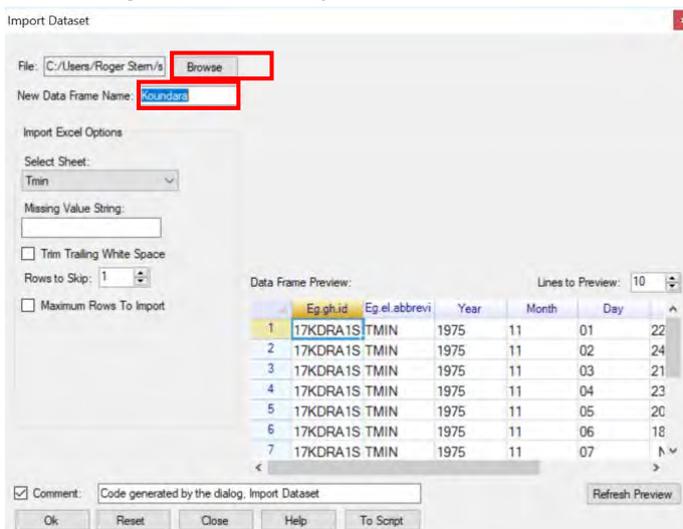
Les dialogues utilisés jusqu'ici sont désormais facilement accessibles via la barre d'outils.

**\*\* Cliquez** sur l'icône (Fig. 17) pour voir les dialogues précédemment utilisés.

**\*\* Choisir Import Dataset** à partir de la liste résultante, Fig.17.

**\*\* Cliquez** sur **Browse** dans la boîte de dialogue résultante, Fig. 18, et choisissez la deuxième station **Koundara**.

**Fig. 18 Un second jeu de données**



**Fig. 19 Les données de 2 stations**

	Year (c)	Month (c)	Day (c)	Rain	RelHu	Tmax	Tr
1	1970	01	01	0.0	NA	NA	NA
2	1970	01	02	0.0	NA	NA	NA
3	1970	01	03	0.0	NA	NA	NA
4	1970	01	04	0.0	NA	NA	NA
5	1970	01	05	0.0	NA	NA	NA
6	1970	01	06	0.0	NA	NA	NA
7	1970	01	07	0.0	NA	NA	NA
8	1970	01	08	0.0	NA	NA	NA
9	1970	01	09	0.0	NA	NA	NA
10	1970	01	10	0.0	NA	NA	NA
11	1970	01	11	0.0	NA	NA	NA
12	1970	01	12	0.0	NA	NA	NA
13	1970	01	13	0.0	NA	NA	NA
14	1970	01	14	0.0	NA	NA	NA

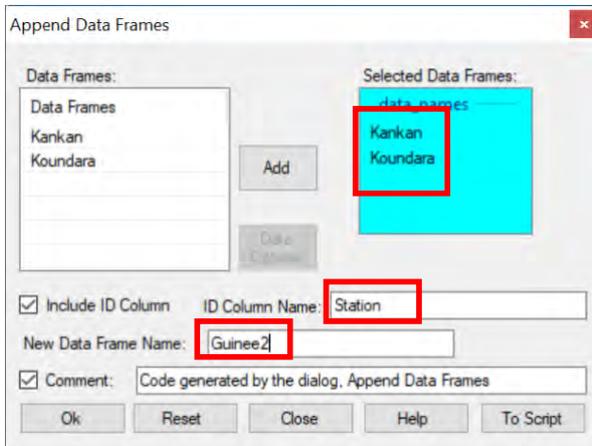
**\*\* Changez le nom** de la feuille en **Tmin**, puis **cliquez** sur **OK**.

**\*\* Répétez les étapes** à partir de la Figure 7 a la page 3 jusqu'à la Figure 16, page 6, pour les données de **Koundara**.

Le résultat avec les 2 feuilles de données se présente tel que sur la Figure 19.

\*\* Utilisez **Prepare > Column: Reshape > Append** pour fusionner les 2 feuilles de données en une seule, Fig. 20.

**Fig. 20 Joindre les données de 2 stations**



**Fig. 21 Les données des 2 stations**

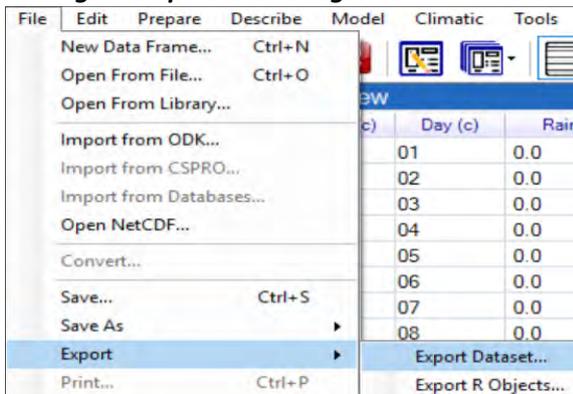
	Station (c)	Year	Month	Day	Rain	RelHum	Tmax	Tmin
1	Kankan	1950	01	01	0.0	NA	35.8	13.0
2	Kankan	1950	01	02	0.0	NA	34.8	15.6
3	Kankan	1950	01	03	0.0	NA	34.7	18.0
4	Kankan	1950	01	04	0.0	NA	33.9	19.7
5	Kankan	1950	01	05	0.0	NA	33.8	14.0
6	Kankan	1950	01	06	0.0	NA	31.2	12.4
7	Kankan	1950	01	07	0.0	NA	33.0	11.9
8	Kankan	1950	01	08	0.0	NA	34.4	18.4
9	Kankan	1950	01	09	0.0	NA	34.3	13.2
10	Kankan	1950	01	10	0.0	NA	33.5	12.5
11	Kankan	1950	01	11	0.0	NA	33.3	12.6
12	Kankan	1950	01	12	0.0	NA	33.5	12.5
13	Kankan	1950	01	13	0.0	NA	32.0	11.0
14	Kankan	1950	01	14	0.0	NA	29.2	12.3

Cette étape est maintenant terminée. Les données des deux stations, et avec les quatre paramètres, sont dans une seule feuille de données, Fig. 21. Celle-ci comporte 40475 lignes (jours) de données.

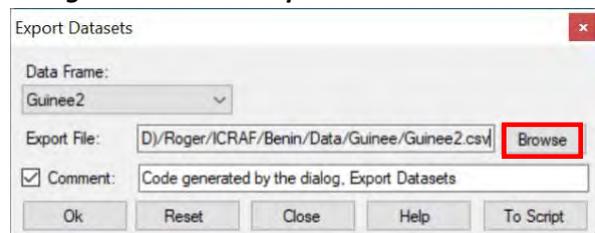
Pour terminer cette première partie, nous allons sauvegarder ces données dans des fichiers.

\*\* Allez à **File > Export > Export Dataset**, Fig.22

**Fig. 22 Export et sauvegarde des données**



**Fig. 23 Choisir ou exporter les données**



\*\* Cliquez sur **Browse**, et choisissez l'emplacement où sauvegarder les données, Fig. 23.

\*\* Après avoir **choisi le nom du fichier**, vous retournerez à la Fig. 22, cliquez sur **OK**. Le fichier sera sauvegardé qu'après avoir cliqué sur **OK**.

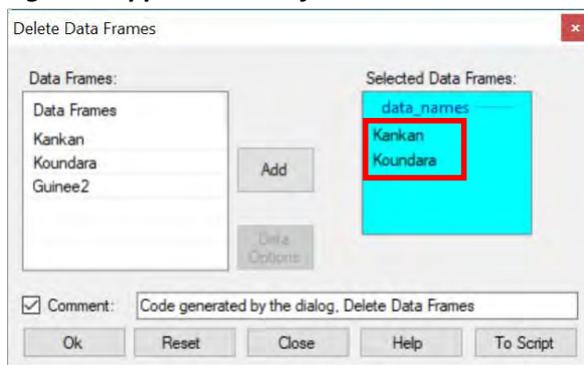
Il est sauvegardé au format CSV par défaut, qui est facilement lisible avec Excel.

\*\* **Faites un clic droit** sur Kankan au bas de la feuille.

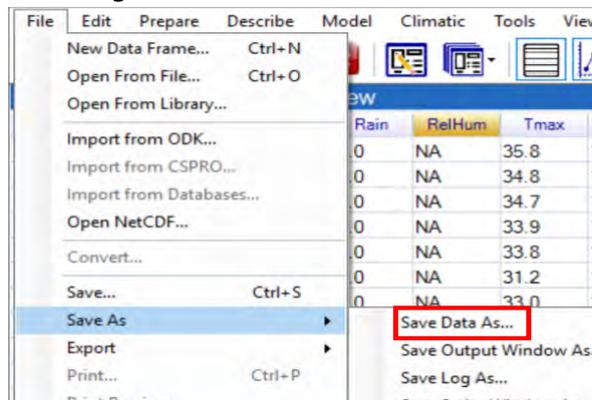
\*\* Cliquez sur **Delete** pour supprimer les feuilles de données individuelles, Fig.23. (I think they can first save the individual datasets before closing them)

\*\* Allez à **File > Save As > Save Data As**, Fig. 24, pour sauvegarder le fichier au format RDS lisible par R-instat.

**Fig. 24. Suppression des feuilles additionnelles**



**Fig. 25 File > Save As**



## Résumé

Cette première partie nous permis d’organiser les données. Elles sont prêtes pour les analyses climatiques avec R-Instat. La tâche principale a été de ‘reformatier’ les données dans une forme utilisable par R-Instat.

Les transformations dépendent de la "forme" initiale des données. D'autres points de départ sont examinés à l'Annexe 1.

Nous avons utilisé les menus **File** et **Prepare** de R-Instat. Dans le menu **Prepare**, nous avons particulièrement utilisé le sous-menu **Prepare > Column: Reshape**, puis les dialogues **Append** et **Unstack**. Pour d’autres formats de données, l’on utilisera des options supplémentaires de ce sous-menu, en particulier **Stack** et **Merge**.

## 4) Ajouter une colonne de Date

\*\* Si vous continuez cette partie à partir de la section précédente, continuez avec les données.

(\*\* Dans le cas contraire, ouvrez le fichier sauvegardé plus haut. Vous pouvez aussi faire **File > Open from Library > Load from Instat Collection > Browse > Climatic > Guinee** and choose the file called **Guinee2.csv**)

**Fig. 1 Convertie la colonne Station en facteur**

	Station (c)	Year	Month	Day	Rain	RelHum	Tmax	Tmin
1	Kankan						35.8	13.0
2	Kankan						34.8	15.6
3	Kankan						34.7	18.0
4	Kankan						33.9	19.7
5	Kankan						33.8	14.0
6	Kankan						31.2	12.4
7	Kankan						33.0	11.9
8	Kankan						34.4	18.4
9	Kankan						34.3	13.2
10	Kankan						33.5	12.5
11	Kankan						33.3	12.6

**Fig. 2 Transformer les colonnes Year, Month, Day en numerique**

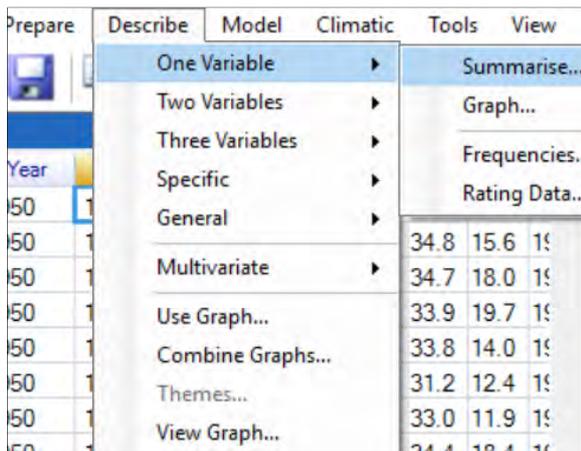
	Station (f)	Year (c)	Month (c)	Day (c)	Rain	RelHu	Tmax	Tmin
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				
	Kankan	1950	01	01				

\*\* **Faites un clic droit** sur le nom **Station** et choisissez **Convert to Factor**, Fig.1.

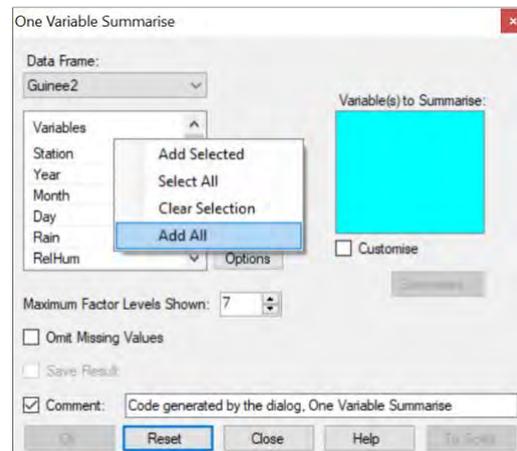
\*\* **Faites un clic droit** sur les colonnes **Year, Month** et **Day** (vous pouvez les sélectionner ensemble) et choisissez **Convert to Numeric**, Fig. 2.

Vérifions maintenant que les données sont à peu près comme nous le souhaitons, nous ne voulons pas de mauvaises surprises!

**Fig. 3 Choisir le dialogue Summarise**



**Fig. 4 Clic droit pour sélectionner toutes les variables**

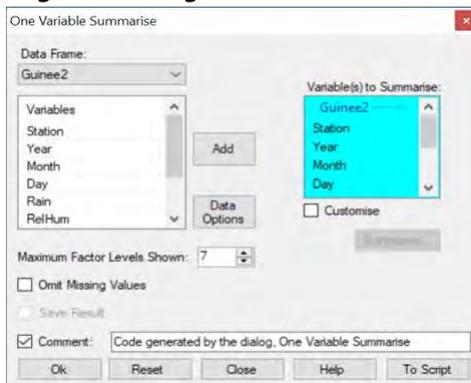


\*\* **Allez a Describe > One Variable > summarize**, Fig. 3.

Remarquez que le bouton OK est grisé, nous devons sélectionner les données à résumer.

\*\* **Faites un clic droit** dans le sélecteur de données, Fig.4, et **cliquez** sur **Add All**. (Ou **sélectionner** toutes les variables et cliquez sur le bouton **Add**.)

**Fig. 5 Le dialogue Summarise terminé**



**Fig. 6 Resultats**

Station	Year	Month	Day	Rain
Kankan :24407	Min. :1950	Min. : 1.00	Min. : 1.0	Min. : 0.0
Roundara:16068	1st Qu.:1974	1st Qu.: 4.00	1st Qu.: 8.0	1st Qu.: 0.0
	Median :1989	Median : 6.00	Median :16.0	Median : 0.0
	Mean :1988	Mean : 6.47	Mean :15.7	Mean : 3.7
	3rd Qu.:2003	3rd Qu.: 9.00	3rd Qu.:23.0	3rd Qu.: 0.6
	Max. :2017	Max. :12.00	Max. :31.0	Max. :162.7
				NA's :366
RelHum	Tmax	Tmin		
Min. : 5	Min. :21	Min. : 5		
1st Qu.: 44	1st Qu.:32	1st Qu.:19		
Median : 64	Median :34	Median :21		
Mean : 61	Mean :34	Mean :20		
3rd Qu.: 78	3rd Qu.:36	3rd Qu.:23		
Max. :100	Max. :45	Max. :28		
NA's :1766	NA's :4737	NA's :6043		

La sélection est maintenant terminée et le bouton Ok est par conséquent activé. **Cliquez** sur **OK**.

Examinons maintenant les résultats. Certains points intéressants sont indiqués en rouge sur la figure 6. Ce sont:

- Le facteur Station a seulement 2 niveaux, ce qui est normal car nous avons 2 stations. Il y a plus de données à **Kankan** par rapport à **Koundara**, et il n'y a pas de valeurs manquantes dans cette variable.

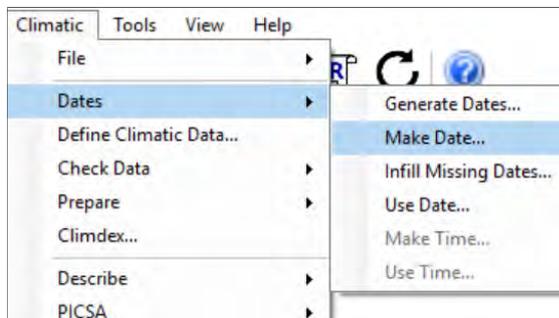
- Pour la colonne Pluie, le minimum est 0 mm (jour sec) et le maximum est 163 mm. Ce sont des valeurs plausibles. Il y a moins de 400 valeurs manquantes. Elles sont désignées par NA dans R.
- Les colonnes **Year, Month, Day** sont également celles attendues. Nous avons par exemple les jours qui sont compris entre 1 et 31, et il n'y a pas de valeurs manquantes.
- Il y a plus de valeurs manquantes dans les 3 autres paramètres climatiques. Les valeurs minimales et maximales sont raisonnables.
- Il n'y a pas de valeurs vraiment bizarres, comme par exemple « -99 » qui auraient dû être transformées en valeurs manquantes.

Tout ceci est donc intéressant pour la suite.

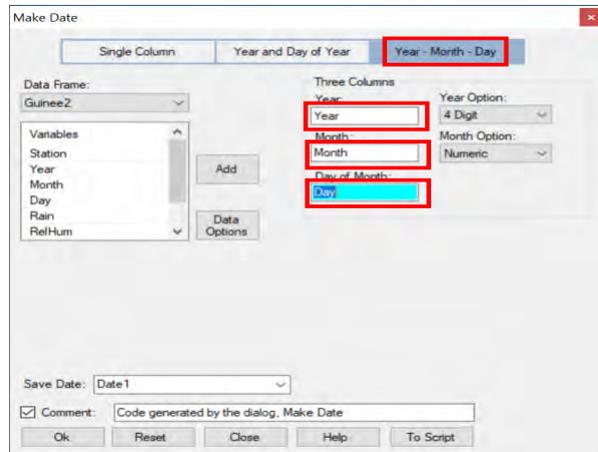
L'étape suivante consiste à créer une seule variable de date.

\*\* Allez à **Climatic > Dates > Make Date**, Fig. 7.

**Fig. 7 Ajouter une variable date**



**Fig.8 Le dialogue Climatic > Date > Make Date**



\*\* Dans la boîte de dialogue Fig. 8, **cliquez** sur le bouton **Year-Month-Day**, parce que ces 3 colonnes sont présentent dans le fichier actuel.

\*\* Complétez le dialogue en **ajoutant** ces **3 colonnes** tel qu'indiqué sur la Fig. 8 puis **cliquez** sur **OK**.

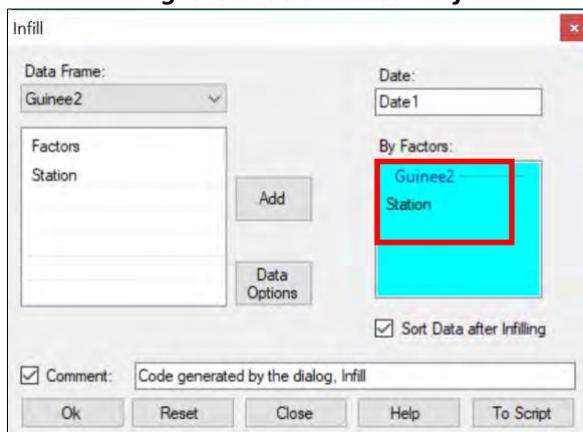
Cela a ajouté une colonne date, de type (D) a la feuille de données.

L'étape suivante consiste à vérifier s'il y a des dates manquantes dans le fichier. Cela est différent des valeurs manquantes dans les données. Il peut arriver que des années entières aient été omises dans le fichier.

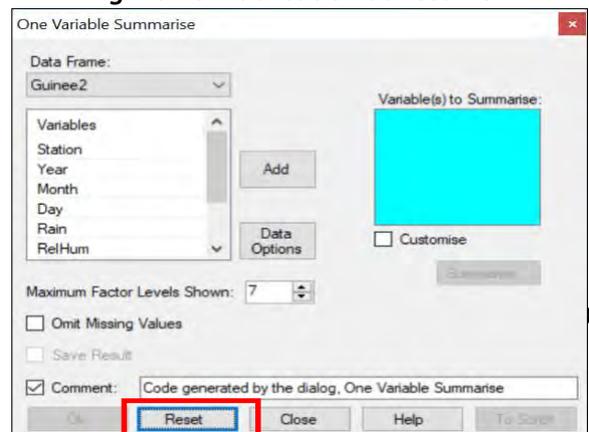
\*\* Notez le nombre de lignes (jours) de données. Il y en a actuellement 40475.

\*\* Allez à **Climatic > Dates > Infill Missing Dates** Fig. 9.

**Fig. 9 Climatic >Dates > Infill**



**Fig. 10 Réinitialisation du résumé**



Vous constatez que la case Date de la Figure 9 est remplie automatiquement.

\*\* Cliquez dans le champ **By Factors** et **Add the Station Column**, Fig. 9.

\*\* Cliquez sur **OK**.

La longueur du fichier est maintenant de 42063 lignes. Donc environ 1600 lignes (jours) ont été ajoutées.

\*\* Utilisez encore le menu **Describe > One Variable > Summarize**. (Souvenez-vous que vous pouvez avoir aussi le bouton de la barre d'outil pour les 10 derniers dialogues utilisés.)

\*\* Cliquez sur le bouton **Reset**, Fig.10.

\*\* **Faites un clic droit** dans le **sélecteur de données** (comme précédemment), et **cliquez** sur **Add All**.

\*\* Cliquez sur **OK**.

**Fig. 11 Résultats du menu Describe > One Variable > Summarise**

Station	Year	Month	Day	Rain
Kankan :24653	Min. :1950	Min. : 1.0	Min. : 1.0	Min. : 0.0
Koundara:17410	1st Qu.:1974	1st Qu.: 4.0	1st Qu.: 8.0	1st Qu.: 0.0
	Median :1989	Median : 6.0	Median :16.0	Median : 0.0
	Mean :1988	Mean : 6.5	Mean :15.7	Mean : 3.7
	3rd Qu.:2003	3rd Qu.: 9.0	3rd Qu.:23.0	3rd Qu.: 0.6
	Max. :2017	Max. :12.0	Max. :31.0	Max. :162.7
	NA's :1588	NA's :1588	NA's :1588	NA's :1954
RelHum	Tmax	Tmin	Datel	
Min. : 5	Min. :21	Min. : 5	Min. :1950-01-01	
1st Qu.: 44	1st Qu.:32	1st Qu.:19	1st Qu.:1974-05-25	
Median : 64	Median :34	Median :21	Median :1988-10-16	
Mean : 61	Mean :34	Mean :20	Mean :1987-12-03	
3rd Qu.: 78	3rd Qu.:36	3rd Qu.:23	3rd Qu.:2003-03-09	
Max. :100	Max. :45	Max. :28	Max. :2017-08-31	
NA's :19354	NA's :6325	NA's :7631		

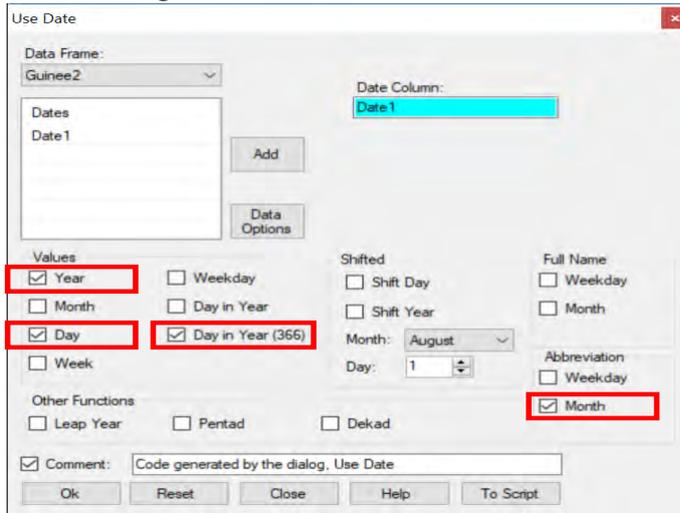
On obtient les résultats de la Fig. 11. Etant donné que des données ont été remplies, il manque des valeurs dans les colonnes Année, Mois et Jour. Nous devons donc y remédier. Heureusement, il n'y a pas de valeurs manquantes dans les colonnes **Date** et **Station**.

\*\* Allez à **Climatic > Dates > Use Date**, voir Fig. 4 pour le menu.

\*\* **Complétez** tel que qu'indiqué sur la Figure 12 et **cliquez** sur **OK**.

Cela a généré 4 nouvelles colonnes, Fig. 13, pour l'année, le mois (avec les étiquettes), le jour du mois et le jour de l'année. L'on peut maintenant remplacer les 3 premières colonnes (avec des dates manquantes) par celles nouvellement générées après le remplissage.

**Fig. 12 Climatic > Dates > Use Date**



**Fig. 13 Resulting columns generated**

	Date1 (D)	month	year	day_in_m	doy
1	1950-01-01	Jan	1950	1	1
2	1950-01-02	Jan	1950	2	2
3	1950-01-03	Jan	1950	3	3
4	1950-01-04	Jan	1950	4	4
5	1950-01-05	Jan	1950	5	5
6	1950-01-06	Jan	1950	6	6
7	1950-01-07	Jan	1950	7	7
8	1950-01-08	Jan	1950	8	8
9	1950-01-09	Jan	1950	9	9
10	1950-01-10	Jan	1950	10	10
11	1950-01-11	Jan	1950	11	11
12	1950-01-12	Jan	1950	12	12
13	1950-01-13	Jan	1950	13	13

Showing 1000 of 42063 rows | Showing 13

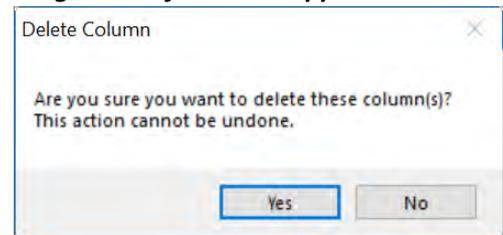
\*\* **Sélectionnez** les 3 colonnes de date originales (non désirées), **faites un clic droit** et **choisissez Delete colons**, Fig. 14.

\*\* Confirmez la suppression, Fig. 15.

**Fig. 14 Supprimer les colonnes non désirées**

Statio	Year	Month	Day	Rain	RelHu	ISU
Kanka	1950	1	1			
Kanka	1950	1	2			
Kanka	1950	1	3			
Kanka	1950	1	4			
Kanka	1950	1	5			
Kanka	1950	1	6			
Kanka	1950	1	7			
Kanka	1950	1	8			
Kanka	1950	1	9			
Kanka	1950	1	10			

**Fig. 15 Confirmer la suppression**



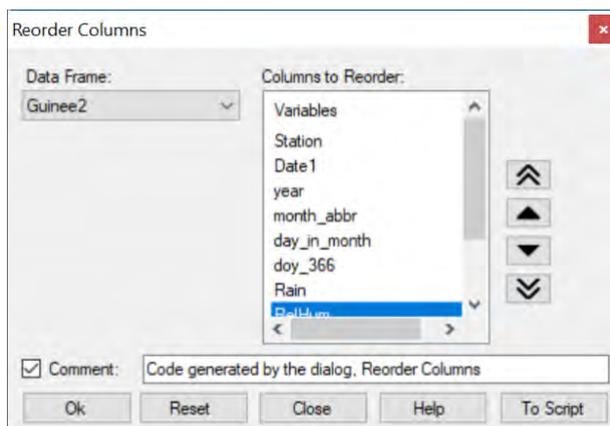
Enfin pour cette section, nous allons réorganiser les colonnes du fichier de manière à avoir d'abord les colonnes des dates avant celles des données.

\*\* **Faites un clic droit** dans le champ des noms, Fig. 14, et sélectionnez **Reorder Columns**.

\*\* **Réordonner** les colonnes comme indiqué sur la Figure 16.

\*\* Cliquez sur OK. Les données se présentent maintenant comme sur la figure 17.

**Fig. 16 Réorganiser les colonnes de la feuille**



**Fig. 17 Les données résultantes**

	Station	Date1 (D)	year	month	day_in	doy_	Rain	RelHu
1	Kankan	1950-01-01	1950	Jan	1	1	0.0	f
2	Kankan	1950-01-02	1950	Jan	2	2	0.0	f
3	Kankan	1950-01-03	1950	Jan	3	3	0.0	f
4	Kankan	1950-01-04	1950	Jan	4	4	0.0	f
5	Kankan	1950-01-05	1950	Jan	5	5	0.0	f
6	Kankan	1950-01-06	1950	Jan	6	6	0.0	f
7	Kankan	1950-01-07	1950	Jan	7	7	0.0	f
8	Kankan	1950-01-08	1950	Jan	8	8	0.0	f
9	Kankan	1950-01-09	1950	Jan	9	9	0.0	f
10	Kankan	1950-01-10	1950	Jan	10	10	0.0	f
11	Kankan	1950-01-11	1950	Jan	11	11	0.0	f
12	Kankan	1950-01-12	1950	Jan	12	12	0.0	f
13	Kankan	1950-01-13	1950	Jan	13	13	0.0	f

Showing 1000 of 42063 rows | Showing 10 of 10

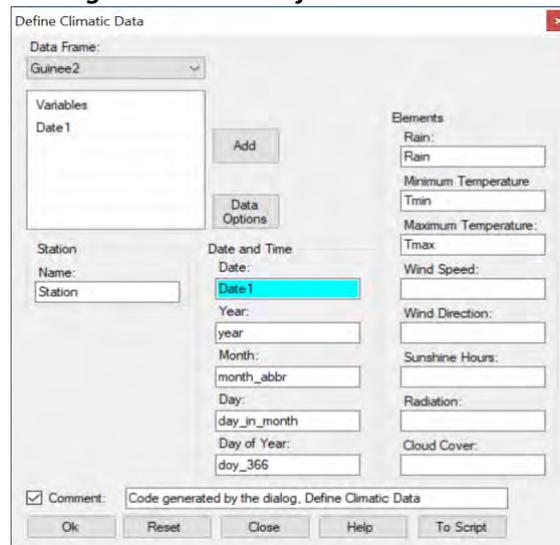
Vous pouvez maintenant enregistrer les données comme décrit précédemment à la section 3. Cependant, la section suivante étant très courte, vous pouvez d'abord le faire avant l'enregistrement.

5) Définir les données comme climatiques

La dernière étape pour la préparation des données consiste à les définir comme données climatiques.

\*\* Utiliser le menu *Climatic > Define Climatic data*, Fig. 1.

**Fig. 1 Climatic > Define Climatic Data**



Pour ce jeu de données, les cases de la boîte de dialogue ont été remplies automatiquement. C'est dû au fait que R-Instat ait identifié les noms des colonnes. Dans le cas contraire, les cases doivent être remplies manuellement.

\*\* Cliquez sur **OK**.

Aucun changement visible n'a été observé. Nous allons donc profiter de cette occasion pour présenter une troisième fenêtre dans R-Instat.

Jusqu'à présent, vous n'avez utilisé que 2 fenêtres : celle pour les données et une autre pour les résultats.

\*\* Dans la barre des tâches, cliquez sur l'icône avec un « i » (i pour informations), Fig. 2.

**Fig. 2 La colonne métadonnées**

	Name	label	class	Climatic_Type	Is_Hidden	Scientific
1	Station		factor	station	FALSE	FALSE
2	Date1		Date	date	FALSE	FALSE
3	year		numeric	year	FALSE	FALSE
4	month_abbr		ordered.factor	month	FALSE	FALSE
5	day_in_month		integer	day	FALSE	FALSE
6	doy_366		integer	doy	FALSE	FALSE
7	Rain		numeric	rain	FALSE	FALSE
8	RelHum		numeric	NA	FALSE	FALSE
9	Tmax		numeric	temp_max	FALSE	FALSE

	Station	Date1 (D)	year	month	day_in	doy_	Rain	RelHu	Tmax	Tmin
1	Kankan	1950-01-01	1950	Jan	1	1	0.0	NA	35.8	13.0
2	Kankan	1950-01-02	1950	Jan	2	2	0.0	NA	34.8	15.6
3	Kankan	1950-01-03	1950	Jan	3	3	0.0	NA	34.7	18.0
4	Kankan	1950-01-04	1950	Jan	4	4	0.0	NA	33.9	19.7
5	Kankan	1950-01-05	1950	Jan	5	5	0.0	NA	33.8	14.0
6	Kankan	1950-01-06	1950	Jan	6	6	0.0	NA	31.2	12.4
7	Kankan	1950-01-07	1950	Jan	7	7	0.0	NA	33.0	11.9
8	Kankan	1950-01-08	1950	Jan	8	8	0.0	NA	34.4	18.4
9	Kankan	1950-01-09	1950	Jan	9	9	0.0	NA	34.3	13.2

\*\* **Tirez** pour agrandir la fenêtre des métadonnées, tel que sur la Figure 2.

Dans cette fenêtre, on a une ligne qui correspond à chaque colonne de la fenêtre de données. L'élément nouveau est que nous avons maintenant une colonne d'informations nommé Climatic\_Type dans les métadonnées.

Cela simplifiera les dialogues pour les analyses climatiques pour les prochaines sections de ce guide. Notez également qu'une étiquette peut être ajoutée pour donner plus de détails sur le contenu de n'importe quelle colonne.

\*\* Cliquez encore sur le bouton «i» de la barre d'outils pour fermer la fenêtre des métadonnées. Vous pouvez également utiliser la flèche bouclée pour réinitialiser les fenêtres à leurs valeurs par défaut.

Enregistrez enfin les données. Nous sommes maintenant prêts à démarrer les analyses.

\*\* Allez à **File > Save As > Save Data As**, on a la boîte de dialogue de la Fig. 3.

\*\* Cliquez sur **Browse**, Fig. 3 et choisissez ou sauvegarder les données.

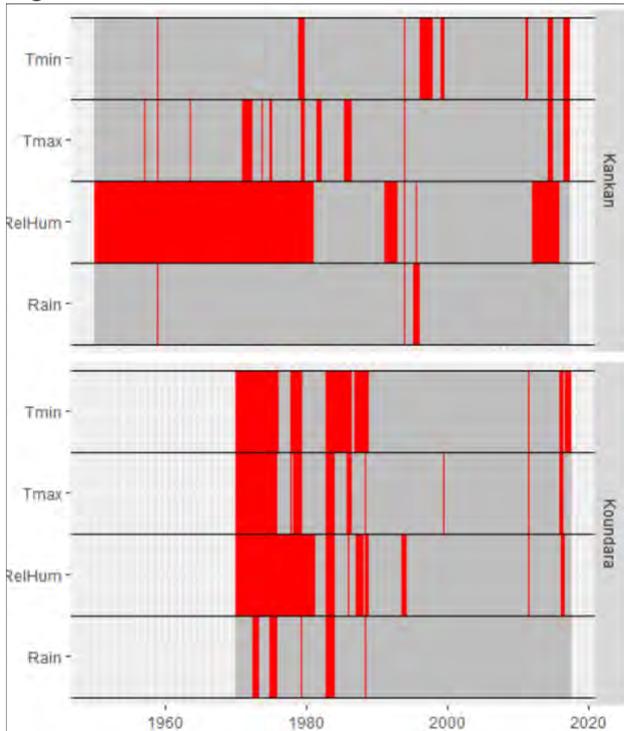
**Fig. 3 File > Save As > Save Data As**

**Fig. 4 Exportation des données**

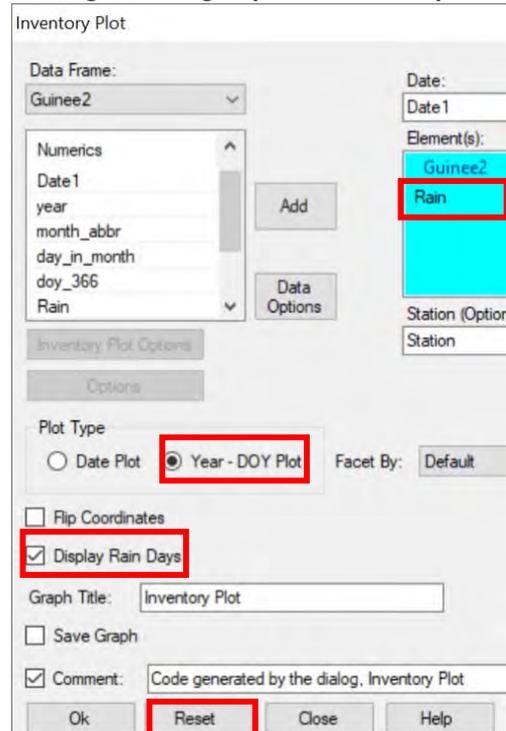
\*\* **Cliquez** sur **Save** dans la boîte de dialogue résultante pour retourner à la Fig. 3.



**Fig. 3 Schema d'inventaire de Kankan et Koundara**



**Fig. 4 Dialogue pour le second plot**



La figure 3 présente l'inventaire pour les 2 stations, le graphique du haut étant celui de Kankan. Les zones rouges indiquent la présence de données manquantes. Les données pluie présentent quelques valeurs manquantes. Les données de température pour Kankan commencent en 1950, et il y a un peu plus de valeurs manquantes pour Tmax.

Les données de précipitations pour Koundara commencent vers 1970, tandis les données de température commencent quelques années plus tard. Il y a des périodes manquantes occasionnelles dans la première partie des enregistrements de pluie, mais presque aucune plus tard.

\*\* Retournez à la boîte de dialogue **Inventory**. (Utilisez le bouton de la barre d'outils.)

\*\* Cliquez sur le bouton **Reset**, Fig. 4.

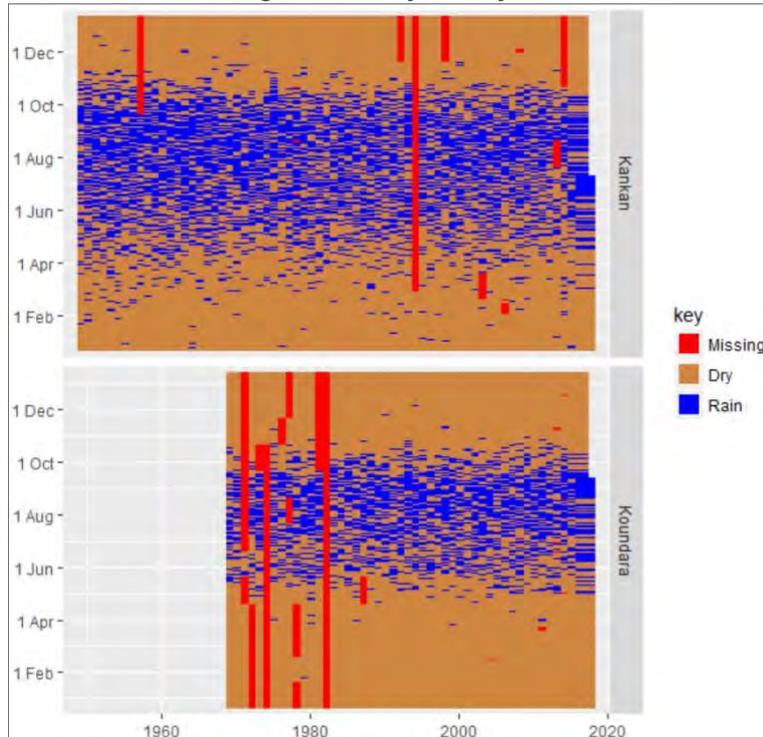
\*\* Juste sélectionner le paramètre **Rain**, Fig. 4.

\*\* Cochez les autres cases tel que sur la Figure 4 pour afficher **display the rain days** (jours de pluie) et **days in the year** (jours de l'année).

\*\* Cliquez sur Ok pour obtenir les résultats de la Fig. 5.

Nous constatons à partir de la figure 5 qu'il y a une seule saison des pluies dans les 2 villes, avec celle de Kankan étant la plus longue. Les données sont d'assez bonne qualité de façon. Par exemple, il n'y a pas de valeurs très étranges.

**Fig. 5 Pattern for rainfall**

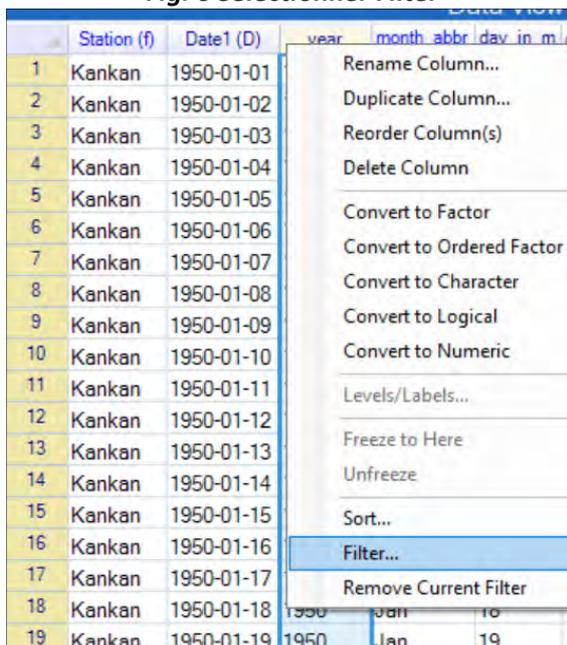


L'option suivante dans **Climatic > Check Data** présente les valeurs journalières de manière plus détaillée.

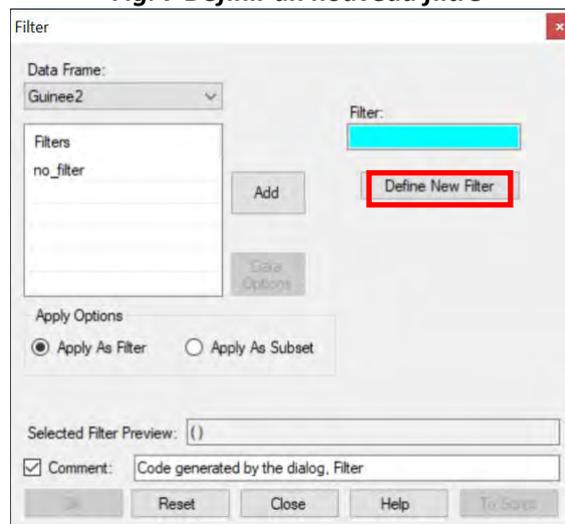
Cette opération prend plus de temps pour une plus grande de quantités de données. Nous allons tout d'abord appliquer un filtre pour analyser les premières années de Koundara. Le filtrage est une option très puissante dans R-Instat.

**\*\*Placez** la souris sur le **nom des colonnes** et **faites un clic droit**, puis cliquez sur **Filter**, Fig. 6.

**Fig. 6 Sélectionner Filter**

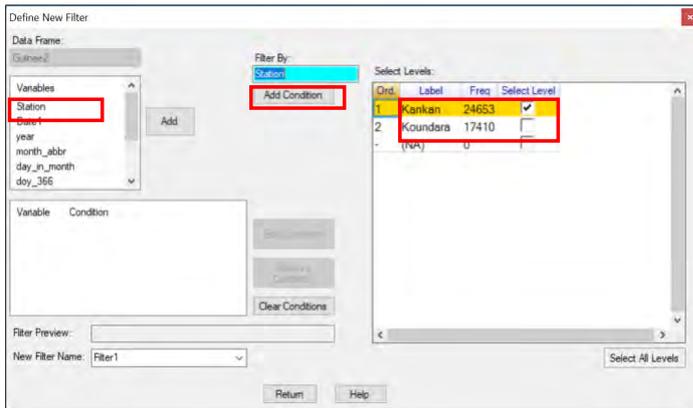


**Fig. 7 Définir un nouveau filtre**

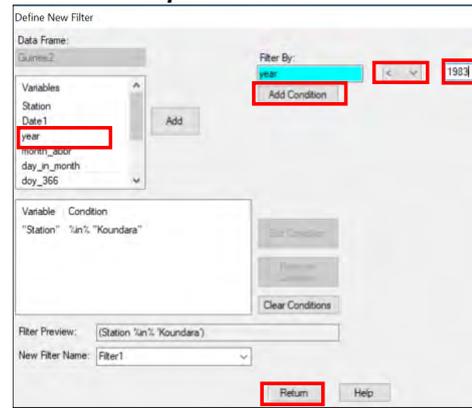


**\*\* Dans** la boîte de dialogue **Filter**, **cliquez** sur **new filter**, Fig. 7.

**Fig. 8 Filtre pour un niveau de facteur**



**Fig. 9 Filtre pour des années particulières**



Sélectionnez d'abord **Station**.

\*\* Nous choisissons de filtrer à partir de la variable **Station**, Fig. 8.

\*\* **Sélectionnez Koundara** plutôt que Kankaran, ensuite **cliquez** sur **Add Condition**.

Sur la Figure 9, nous constatons que la condition a été appliquée par ce message « Station %in% 'Koundara' ». Nous allons maintenant ajouter une autre condition.

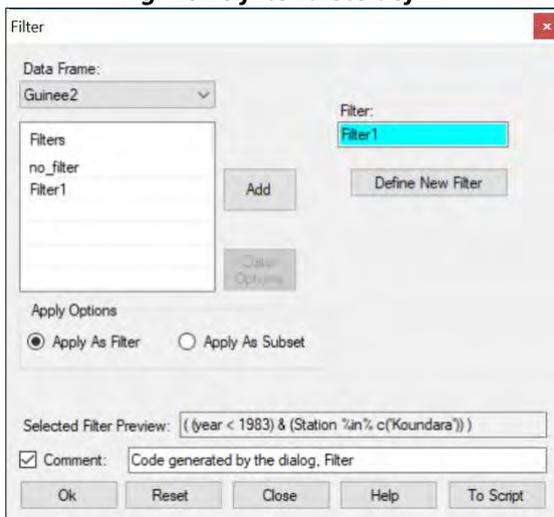
\*\* Choisissez la colonne **year**, Fig. 9.

\*\* **Définissez** la condition '<' (inférieure a), et **saisissez l'année 1983**, Fig. 9.

\*\* **Cliquez** sur **Add Condition**

\*\* Maintenant que nous avons ajouté les 2 conditions, cliquez sur **Return**.

**Fig. 10 Le filtre a été défini**



**Fig. 11 Le filtre est appliqué**

	Station (f)	Date1 (D)	year	month_abbr	day_in_m	doj
24654	Koundara	1970-01-01	1970	Jan	1	1
24655	Koundara	1970-01-02	1970	Jan	2	2
24656	Koundara	1970-01-03	1970	Jan	3	3
24657	Koundara	1970-01-04	1970	Jan	4	4
24658	Koundara	1970-01-05	1970	Jan	5	5
24659	Koundara	1970-01-06	1970	Jan	6	6
24660	Koundara	1970-01-07	1970	Jan	7	7
24661	Koundara	1970-01-08	1970	Jan	8	8
24662	Koundara	1970-01-09	1970	Jan	9	9
24663	Koundara	1970-01-10	1970	Jan	10	10
24664	Koundara	1970-01-11	1970	Jan	11	11
24665	Koundara	1970-01-12	1970	Jan	12	12
24666	Koundara	1970-01-13	1970	Jan	13	13

Showing 1000 of 4748 rows (42063) | Showing 10 of

\*\* De retour sur la boîte de dialogue **Filter**, Fig. 10, **cliquez** sur **Ok**.

Sur la figure 11, on constate que la première colonne est maintenant en rouge. Cela indique qu'un filtre est actif. Les données commencent maintenant à être sélectionnées. Nous constatons aussi que seulement 4748 lignes de données (sur les 42063 lignes) ont été sélectionnées pour cette étape de l'analyse.

# Tutoriel d'introduction : Partie 1: Description des données

## Introduction

Bienvenue dans ce tutoriel d'introduction R-Instat. R-Instat est un logiciel de statistiques gratuit alimenté par le langage de programmation et logiciel R. Il est conçu pour exploiter la puissance du système statistique R, tout en étant aussi facile à utiliser que d'autres paquets de statistiques de configuration pointer-cliquer traditionnels.

R-Instat est le premier produit à avoir été développé dans le cadre de l'African Data Initiative (ADI), un projet collaboratif visant à renforcer la maîtrise de l'analyse de statistiques et de la gestion de l'information en Afrique et au-delà. L'objectif global du projet African Data Initiative va au-delà du développement de ce logiciel, mais R-Instat est une étape préliminaire importante.

L'auditoire d'R-Instat a été établi à travers la campagne de financement participatif qui a lancé son développement. Selon nous, il était important de développer davantage de logiciels statistiques gratuits, open source et faciles à utiliser, qui auraient pour but d'encourager de bonnes pratiques statistiques.

"Instat" dans "R-Instat" se réfère à un simple ensemble de statistiques développé dans les années 1980 avec des objectifs et publics cibles similaires à ceux d'R-Instat, et une grande partie de la philosophie de R-Instat est inspirée par Instat. Instat a inclus un menu spécial pour l'analyse des données climatiques et R-Instat a suivi ce modèle. En plus R-Instat inclut un autre menu spécial pour l'analyse des données sur les marchés publics.

Nous vous recommandons de suivre les instructions d'installation. Dans ce document, nous nous concentrons sur l'introduction du logiciel une fois celui-ci installé.

L'équipe ADI (R-Instat) R-Instat@AfricanMathsInitiative.net

## Lancer R-Instat pour la première fois

Maintenant que vous avez installé R-Instat, il est temps de le faire démarrer.

Lorsque R-Instat s'exécute pour la première fois, si vous n'avez pas utilisé le logiciel R auparavant ou si une version récente de R a été installée sur votre ordinateur, il se peut que ce message apparaisse :

"Would you like to use a personal library instead?" ("Voudriez-vous plutôt utiliser une collection personnelle"?)

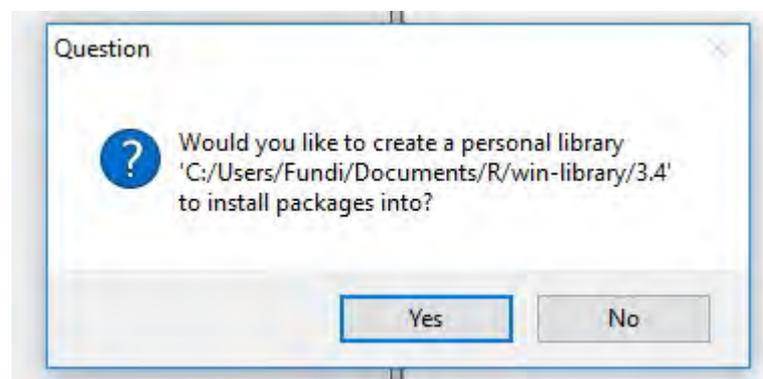


→ Cliquez sur Yes (Oui) pour que le logiciel procède à l'installation des paquets R requis sur votre ordinateur dans un dossier situé dans vos documents.

Après avoir cliqué sur Yes (Oui), vous verrez sans doute le message suivant :

"Would you like to create a personal library" ("Voulez-vous créer une bibliothèque personnelle")

→ Cliquez à nouveau sur Yes (Oui) pour autoriser R à installer les packages dans le dossier spécifié.



Si ce message n'apparaît pas (et qu'aucune erreur n'apparaît), cela voudra dire que les paquets R ont été correctement installés - il est probable que vous disposiez déjà des structures de dossiers nécessaires.

Quand une commande prend plus de temps à s'exécuter (ici à installer des paquets), vous verrez le message "Sorry for the Wait" ("Désolé pour l'attente") (ci-dessus), qui vous indiquera que R-Instat est toujours en cours.

La première fois que vous exécuterez R-Instat, il vous faudra patienter, car de nombreux paquets R seront installés. L'exécution se fera beaucoup plus rapidement par la suite car les paquets seront déjà installés.

Une fois le dialogue d'attente a disparu, vous pourrez commencer à utiliser R-Instat!

## Explorer R-Instat

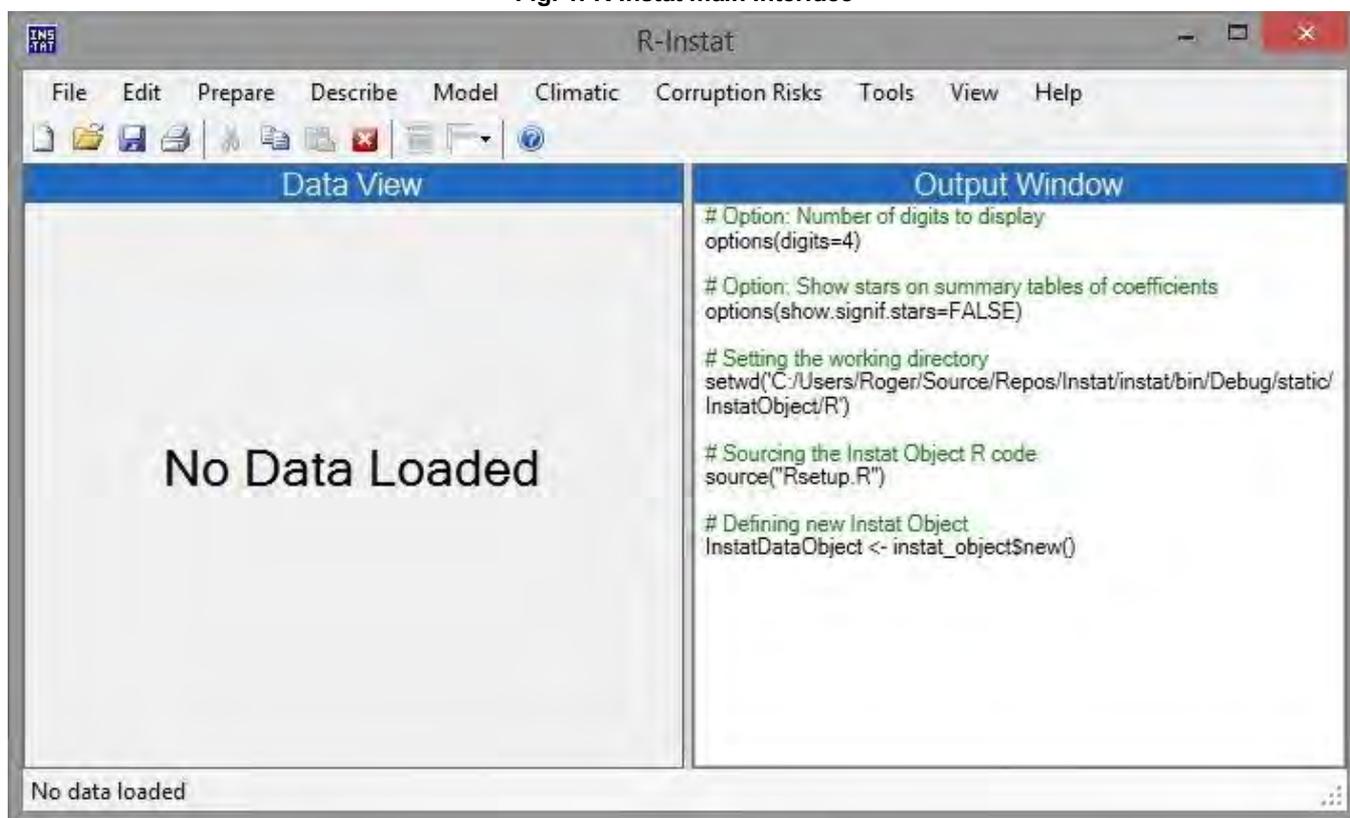
Cette section fournit une première série d'exemples pour vous aider à vous familiariser avec R-Instat et ses caractéristiques de base.

### 1. L'installation.

Actuellement, R-Instat est disponible uniquement pour Windows. Nous conseillons aux utilisateurs de Mac et Linux de l'utiliser à travers une machine Windows virtuelle. Nous prévoyons de développer une version multiplateforme bientôt.

Une fois installé et activé, vous devriez voir l'écran ci-dessous :

Fig. 1: R-Instat main Interface



### 2. Votre première tâche - Importer des données de la bibliothèque

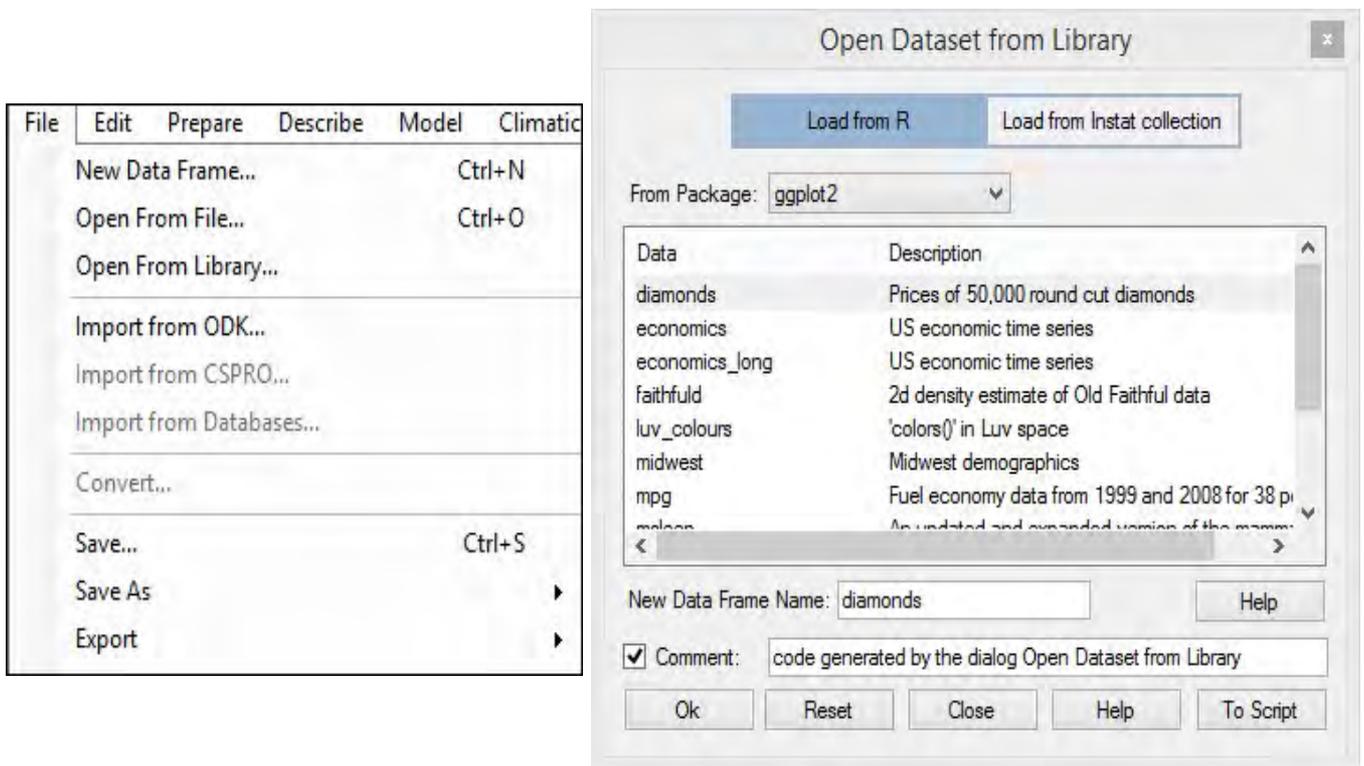
→ Appuyez sur File> Open From Library (Fichier> Ouvrir à partir de la bibliothèque).

→ Appuyez sur la liste déroulante From Package et choisissez ggplot2.

→ Choisissez le premier exemple, « diamonds » comme indiqué dans la Fig. 8. Vous devriez voir qu'un second bouton d'aide est maintenant activé, juste en dessous de la liste de bases de données.

→ Appuyez sur ce bouton pour obtenir plus d'informations sur la base de données soit l'ensemble des données utilisées par Hadley Wickham, l'auteur du paquet ggplot2.

**Fig. 2. Utiliser l'une des bases de données de la bibliothèque**



→ Revenez maintenant à la boîte de dialogue, sélectionnez à nouveau la base de données « diamonds » et appuyez sur OK.

Fig. 3 The diamonds data

Data View										
	carat	cut (o.f)	color (o.f)	clarity (o.f)	depth	table	price	x	y	z
1	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
3	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
6	0.24	Very Good	J	VVS2	62.8	57.0	336	3.94	3.96	2.48
7	0.24	Very Good	I	VVS1	62.3	57.0	336	3.95	3.98	2.47
8	0.26	Very Good	H	SI1	61.9	55.0	337	4.07	4.11	2.53
9	0.22	Fair	E	VS2	65.1	61.0	337	3.87	3.78	2.49
10	0.23	Very Good	H	VS1	59.4	61.0	338	4.00	4.05	2.39
11	0.30	Good	J	SI1	64.0	55.0	339	4.25	4.28	2.73
12	0.23	Ideal	J	VS1	62.8	56.0	340	3.93	3.90	2.46
13	0.22	Premium	F	SI1	60.4	61.0	342	3.88	3.84	2.33
14	0.31	Ideal	J	SI2	62.2	54.0	344	4.35	4.37	2.71
15	0.20	Premium	E	SI2	60.2	62.0	345	3.79	3.75	2.27
16	0.32	Premium	E	I1	60.9	58.0	345	4.38	4.42	2.68
17	0.30	Ideal	I	SI2	62.0	54.0	348	4.31	4.34	2.68
18	0.30	Good	J	SI1	63.4	54.0	351	4.23	4.29	2.70
19	0.30	Good	J	SI1	63.8	56.0	351	4.23	4.26	2.71

Showing 1000 of 53940 rows | Showing 10 of 10 columns

→ Faites défiler les données vers le bas, vous verrez que seulement 1000 lignes apparaissent. C'est juste une fenêtre sur une partie du data frame (la matrice de données) disponibles dans le paquet

→ Faites un **clik droit** sur l'onglet inférieur, Fig. 4.

→ Choisissez la dernière option, **View Data** (Afficher les données). De cette façon, vous pourrez voir l'ensemble des données, illustré ci-dessous (Fig. 4).

Fig. 4. Visualiser la base de données

The image shows a context menu for the 'diamonds' tab with the following options: Insert..., Delete..., Rename..., Reorder..., Copy..., Hide, Unhide, and View Data Frame (highlighted). To the right, a window titled 'diamonds' displays a table of diamond data with columns: carat, cut, color, clarity, depth, table, price, x, y, z. The table contains rows 53922 through 53940.

	carat	cut	color	clarity	depth	table	price	x	y	z
53922	0.70	Very Good	E	VS2	62.8	60.0	2755	5.59	5.65	3.53
53923	0.70	Very Good	D	VS1	63.1	59.0	2755	5.67	5.58	3.55
53924	0.73	Ideal	I	VS2	61.3	56.0	2756	5.80	5.84	3.57
53925	0.73	Ideal	I	VS2	61.6	55.0	2756	5.82	5.84	3.59
53926	0.79	Ideal	I	SI1	61.6	56.0	2756	5.95	5.97	3.67
53927	0.71	Ideal	E	SI1	61.9	56.0	2756	5.71	5.73	3.54
53928	0.79	Good	F	SI1	58.1	59.0	2756	6.06	6.13	3.54
53929	0.79	Premium	E	SI2	61.4	58.0	2756	6.03	5.96	3.68
53930	0.71	Ideal	G	VS1	61.4	56.0	2756	5.76	5.73	3.53
53931	0.71	Premium	E	SI1	60.5	55.0	2756	5.79	5.74	3.49
53932	0.71	Premium	F	SI1	59.8	62.0	2756	5.74	5.73	3.43
53933	0.70	Very Good	E	VS2	60.5	59.0	2757	5.71	5.76	3.47
53934	0.70	Very Good	E	VS2	61.2	59.0	2757	5.69	5.72	3.49
53935	0.72	Premium	D	SI1	62.7	59.0	2757	5.69	5.73	3.58
53936	0.72	Ideal	D	SI1	60.8	57.0	2757	5.75	5.76	3.50
53937	0.72	Good	D	SI1	63.1	55.0	2757	5.69	5.75	3.61
53938	0.70	Very Good	D	SI1	62.8	60.0	2757	5.66	5.68	3.56
53939	0.86	Premium	H	SI2	61.0	58.0	2757	6.15	6.12	3.74
53940	0.75	Ideal	D	SI2	62.2	55.0	2757	5.83	5.87	3.64

Il y a 10 colonnes (variables) de données dans ce fichier, dont 7 sont **numeric** (numériques) et 3 sont **categorical** (catégorielles). Les colonnes catégorielles sont dénommées « **factors** » et sont désignées à l'aide d'un "f" après le nom de la colonne. Ces colonnes catégorielles sont effectivement ordonnées, par exemple la deuxième colonne, à savoir « the cut » (la coupe) des diamants varie de **Fair** à **Ideal**. Les colonnes catégorielles ordonnées sont désignées à l'aide du symbole "(o.f)" à la suite du titre de la colonne dans R-Instat.

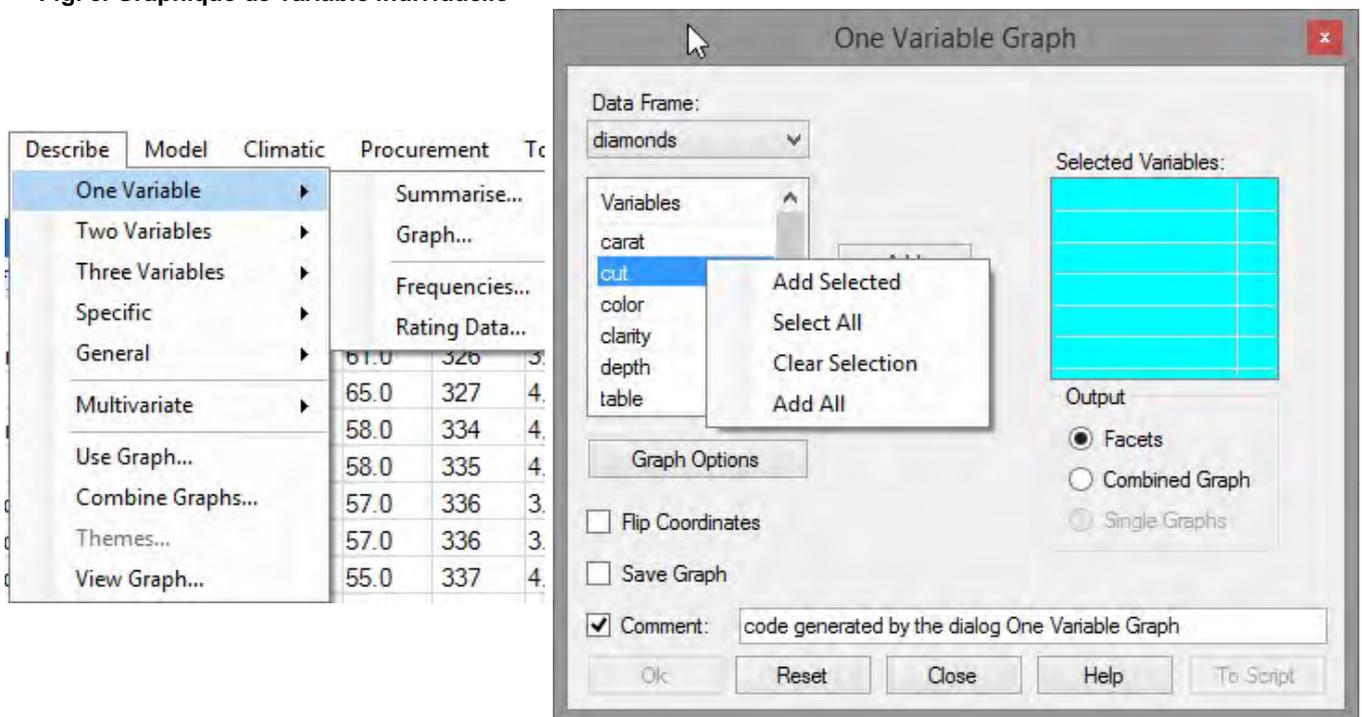
Ces données sont prêtes à être analysées, nous pouvons donc vous montrer quelques graphiques à l'aide du menu « Describe ».

### 3. Quelques graphiques

→ Appuyez sur **Describe > One Variable > Graph**, Fig. 5. (\* Décrire > Une variable > Graphique, Fig. 5.)

→ \*\*\*\* Cliquez avec le bouton droit \*\*\*\* dans le sélecteur de variables et choisissez l'option **Add All** (Ajouter tout). (Vous pouvez également, si vous le souhaitez, sélectionner toutes les colonnes, puis cliquer sur \*\*\*\* Add/Ajouter \*\*\*\*) Fig. 5.

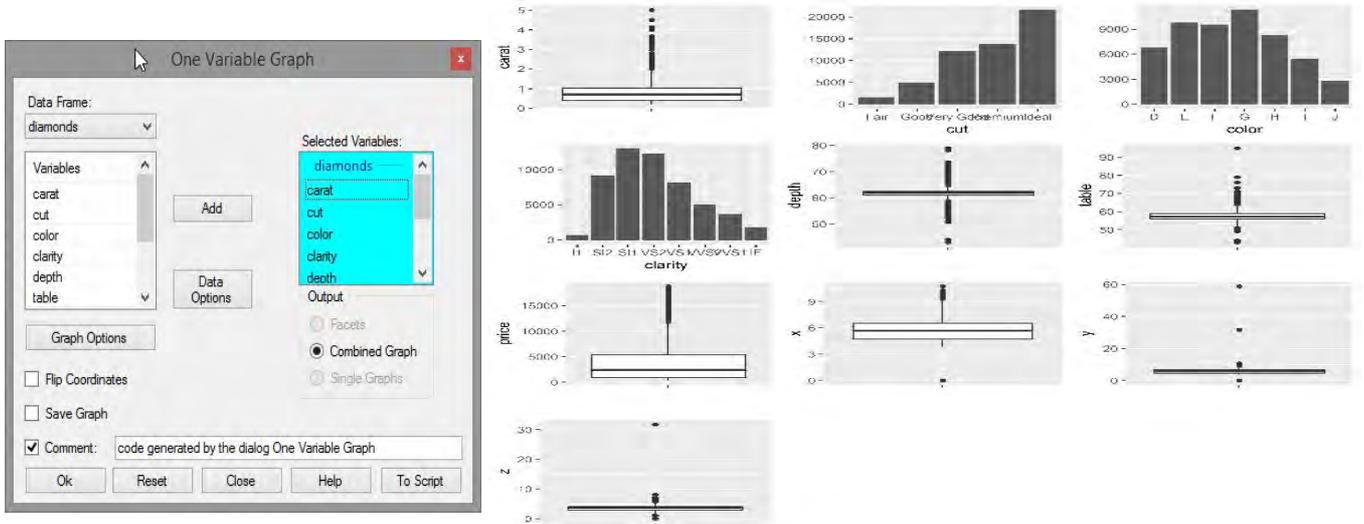
Fig. 5. Graphique de variable individuelle



Dans la boîte de dialogue de la figure 5, le bouton radio n'est plus Facets (Facettes) mais Combine Graph (Graphique combinées), voir Fig. 6. En effet, les variables sélectionnées ont des types de données différents. Certaines colonnes sont catégorielles tandis que d'autres sont numériques.

→ Appuyez sur OK pour afficher les résultats indiqués sur la Fig. 6.

**Fig. 6. Graphique Variable individuelle**

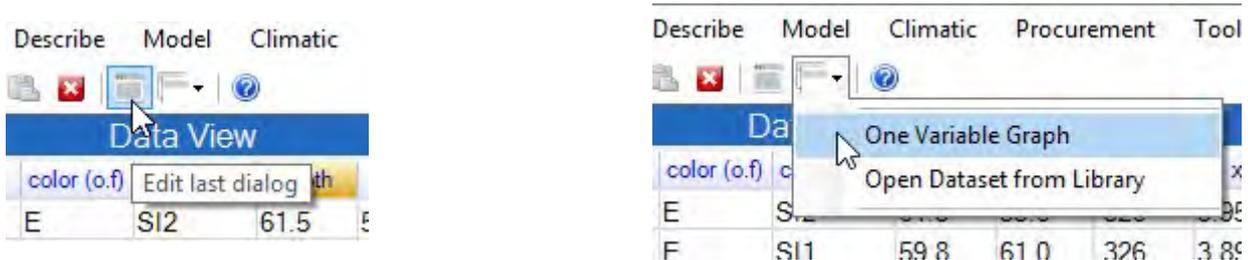


Vous connaissez sûrement déjà les « box plots » Nous vous en parlerons un peu plus tard, même si ce tutoriel vise principalement à montrer comment utiliser R-Instat, plutôt qu'à enseigner les statistiques.

Souvent, les résultats de l'utilisation d'un dialogue peuvent être améliorés, de sorte que vous souhaiterez l'utiliser à nouveau. Vous pouvez utiliser les mêmes options de menu que sur la Fig. 5, mais il existe un moyen plus rapide.

→ Appuyez sur la petite image de dialogue dans la barre d'outils (toolbar), voir Fig. 7, qui vous ramène au dialogue précédent. (Ou l'icône suivante vous permet de revenir à l'un des dialogues précédents.)

**Fig. 7. Utilisez le menu outils (Toolbar) pour retrouver un dialogue précédent**



Vous voyez que le dialogue s'est «souvenu» des paramètres tels que vous les avez laissés, lorsque vous avez appuyé sur OK. C'est souvent préférable ainsi.

→ Mais cette fois, appuyez sur le bouton Reset (Réinitialiser) en bas de la boîte de dialogue pour effacer tous les paramètres.

→ Omettez les 4 premières variables et sélectionnez les 6 dernières.

Comme ce sont toutes des colonnes numériques, les boutons radio sur la droite vous ont permis de créer un graphique « à facettes », de sorte que vous puissiez voir ce que c'est!

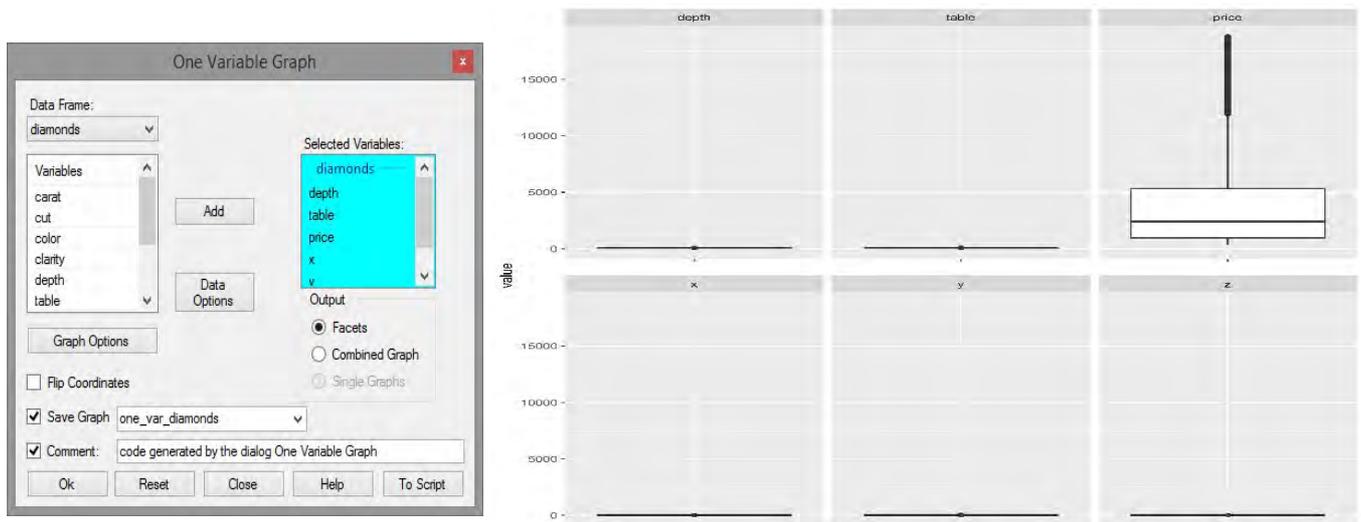
→ Cliquez également sur la case à cocher pour sauvegarder le graphique.

- Nommez-le **one-var diamonds** (Prière d'inclure un «tiret» et un espace.)
- Appuyez maintenant sur **OK**

Le dialogue n'a pas fonctionné. Au lieu de cela, il donne un message que "le nom ne peut contenir un espace" (ou un tiret). C'est le nom d'un objet dans R et ceux-ci ne sont pas autorisés.

- Appuyez sur OK pour effacer la boîte de message.
- Changez le nom de façon à ce qu'il devienne **OneVarDiamonds**, Fig. 8, et cliquez à nouveau sur **OK**.

**Fig. 8. Le dialogue d'une variable individuelle Avec un graphique a facettes**



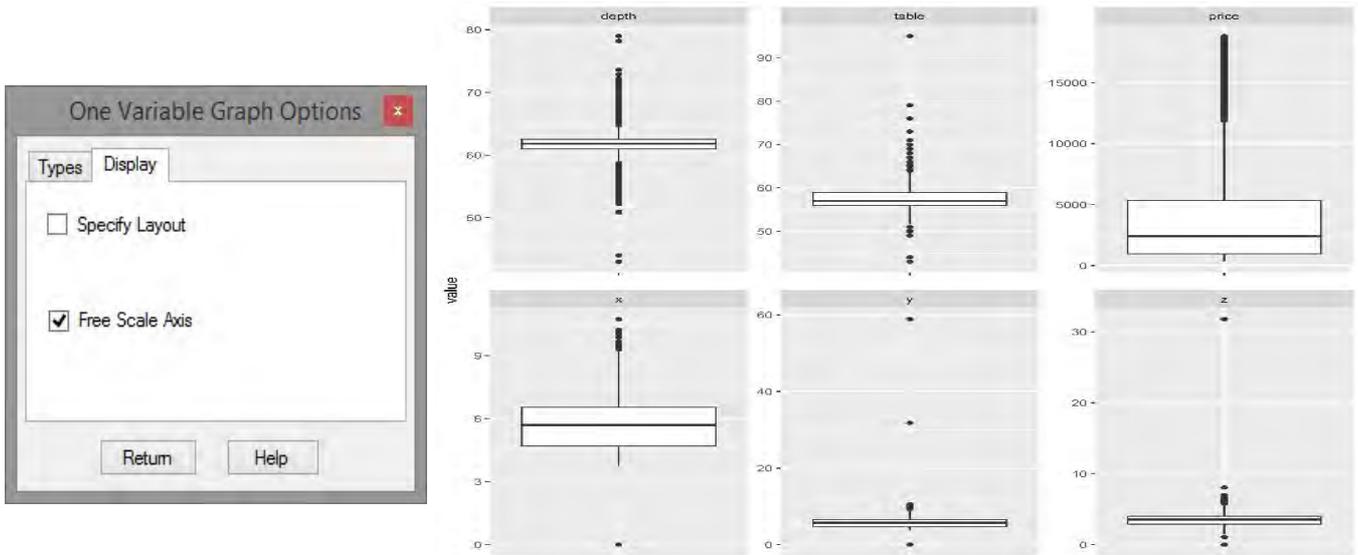
Ceci représente un graphique « à facettes », Fig. 8. Il s'agit de plusieurs graphiques dont l'axe y est le même. L'axe individuel n'est pas toujours nécessaire. Cependant, dans notre cas, les différentes variables ont des échelles très différentes et nous devons en tenir compte dans le graphique.

- Revenez au même dialogue.
- Cliquez sur le bouton **Graph Options** (Options graphiques).

Vous voyez maintenant un sous-dialogue avec seulement 2 onglets, Fig. 9. Un onglet vous permet de changer le type de graphique qui est affiché.

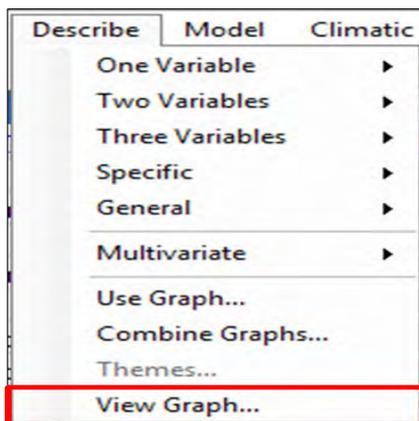
- Appuyez sur l'onglet **Display**, puis sur le **Free Scale Axis** (l'axe Free Scale).
- Appuyez à nouveau sur le bouton **Return**, puis sur **OK**, pour afficher le graphique illustré (Fig. 9).

Fig. 9. Le graphique d'une variable individuelle

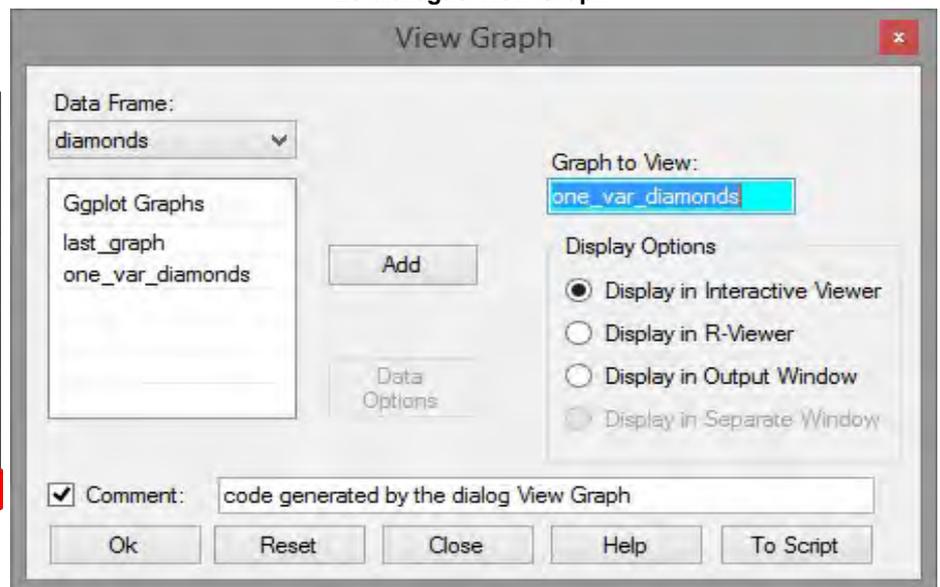


→ Appuyez sur **Describe** > **View Graph** pour examiner d'avantage ce dernier graphique, Fig. 10.

Fig. 10. Le Menu Describe



Le dialogue View Graph

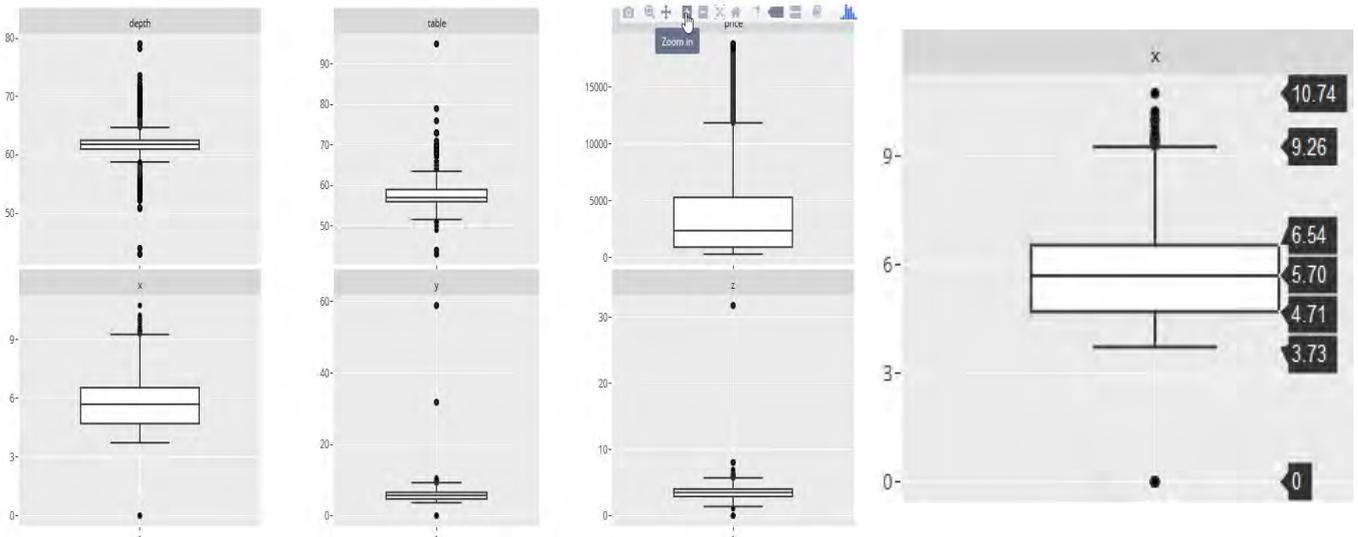


→ Appuyez sur **OK** pour afficher les graphiques dans une fenêtre séparée (interactive), Fig. 11.

→

Survolez un graphique particulier pour ajouter automatiquement des résumés numériques, Fig. 11.

Fig. 11. Les resultats View Graph



#### 4. Quelques résumés

Les analyses impliquent souvent des résumés numériques et graphiques.

→ Appuyez sur Describe > One variable> Summarise (Décrire >Une variable>Résumer).

→ Sélectionnez à nouveau toutes les variables (comme vous l'avez fait lors de la première utilisation de la boîte de dialogue Graphique), Fig. 12.

→ Appuyez sur OK pour afficher les résultats indiqués sur la Fig. 12.

Fig. 12. Resume d'une variable individuelle

Dialogue affichant des resultats

Ce n'est pas tout à fait juste. La variable indiquée sur la figure 12 n'est pas très claire. Certaines catégories ont été combinées.

→ Retournez au dernier dialogue.

→ Dans le dialogue, Fig. 12, modifiez les **Maximum Factor Levels Shown** (niveaux de facteur maximum) affichés de 7 à 10. Appuyez sur **OK**.

Les niveaux sont maintenant tous donnés pour cette colonne de facteur.

→ Examinez la correspondance entre les valeurs données pour la variable x sur la Fig. 12 et celles du box plot pour la même variable sur la Fig. 11. Pour vous faciliter la tâche, elles sont présentées côte à côte sur la Fig. 13.

La médiane des deux graphiques est très proche (Fig. 13). Les autres valeurs sont-elles les mêmes ? La correspondance est-elle utile pour comprendre (ou enseigner) ce que fournit un box plot ?

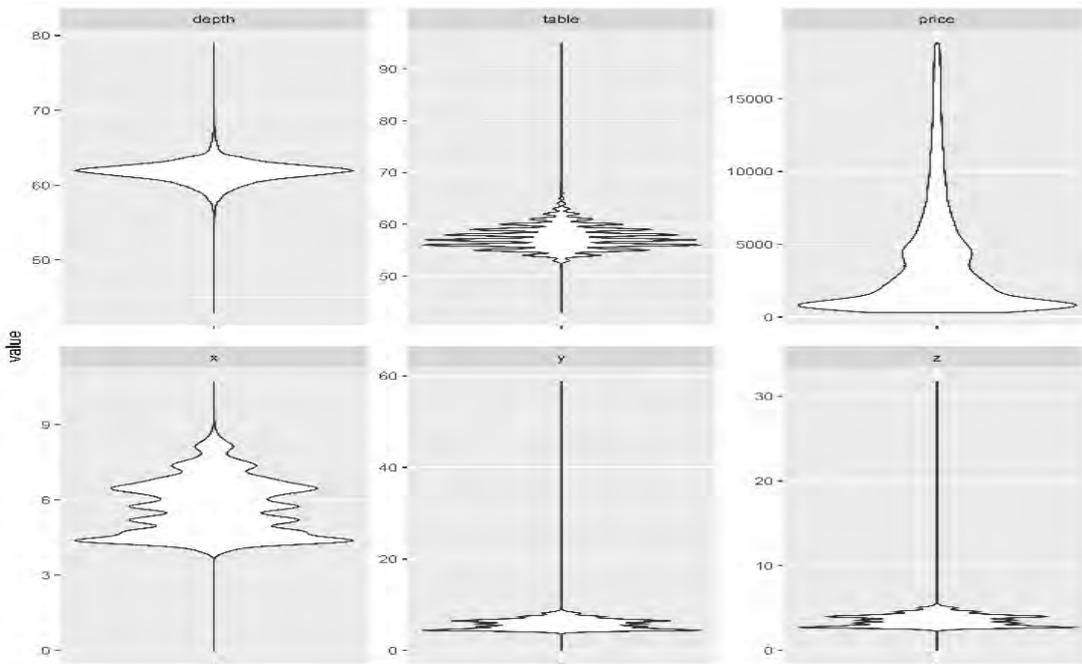
## 5. Un petit défi

→ Retournez à la boîte de dialogue Describe/One variable/Graph (Décrire > Une variable > Graphique).

→ Avec les mêmes 6 variables, de **Depth to z** (Depth à z), passez d'un boxplot à un Violin Plot (Ne vous inquiétez pas si vous ne savez pas encore ce qu'est un Violin Plot).

→ Examinez la forme curieuse (Fig.13) de certaines des variables, en particulier celle dénommée **table**. (Vous devriez remarquer quelque chose qui n'est pas forcément évident à partir d'un boxplot.)

**Fig. 13 Curious results from a violin plot**



→ Examinez ceci davantage (Un indice : Appuyez sur **Describe > One variable > Frequency** (Décrivez > Une variable > Fréquences.)

## 6. Une analyse plus ambitieuse

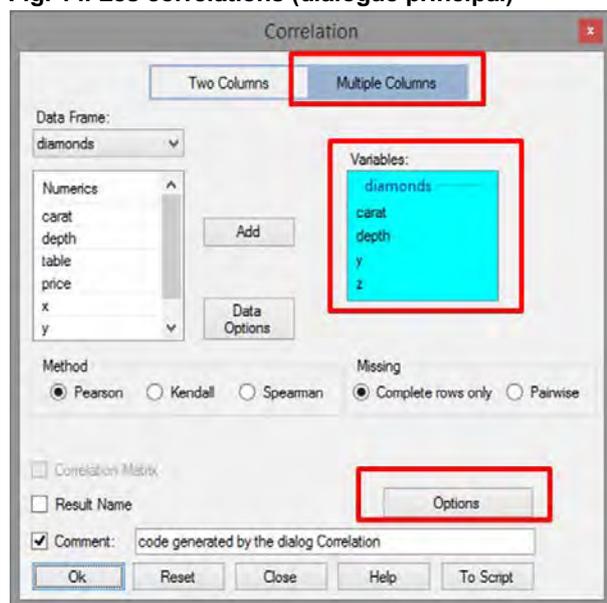
→ Appuyez sur Describe > One variable > Frequency (Décrire > Multivarié > Corrélations). (Notez que seules les colonnes numériques sont visibles pour cette boîte de dialogue.)

→ Sélectionnez le bouton Multiple Columns en haut de la boîte de dialogue, Fig. 14.

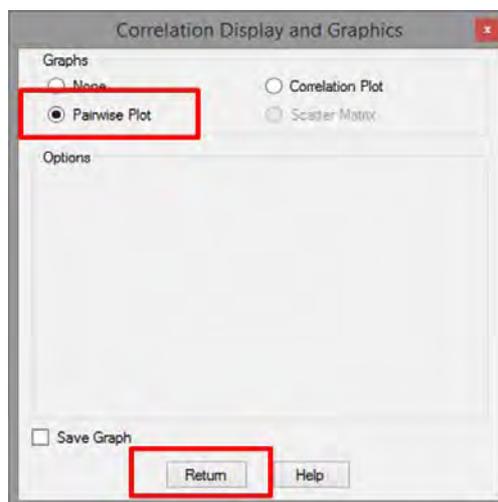
→ Sélectionnez les 2 premières variables (Carat et Depth) et les deux dernières (y et z), Fig. 14.

→ Cliquez sur le bouton Options pour accéder au sous-dialogue, Fig. 14.

**Fig. 14. Les corrélations (dialogue principal)**



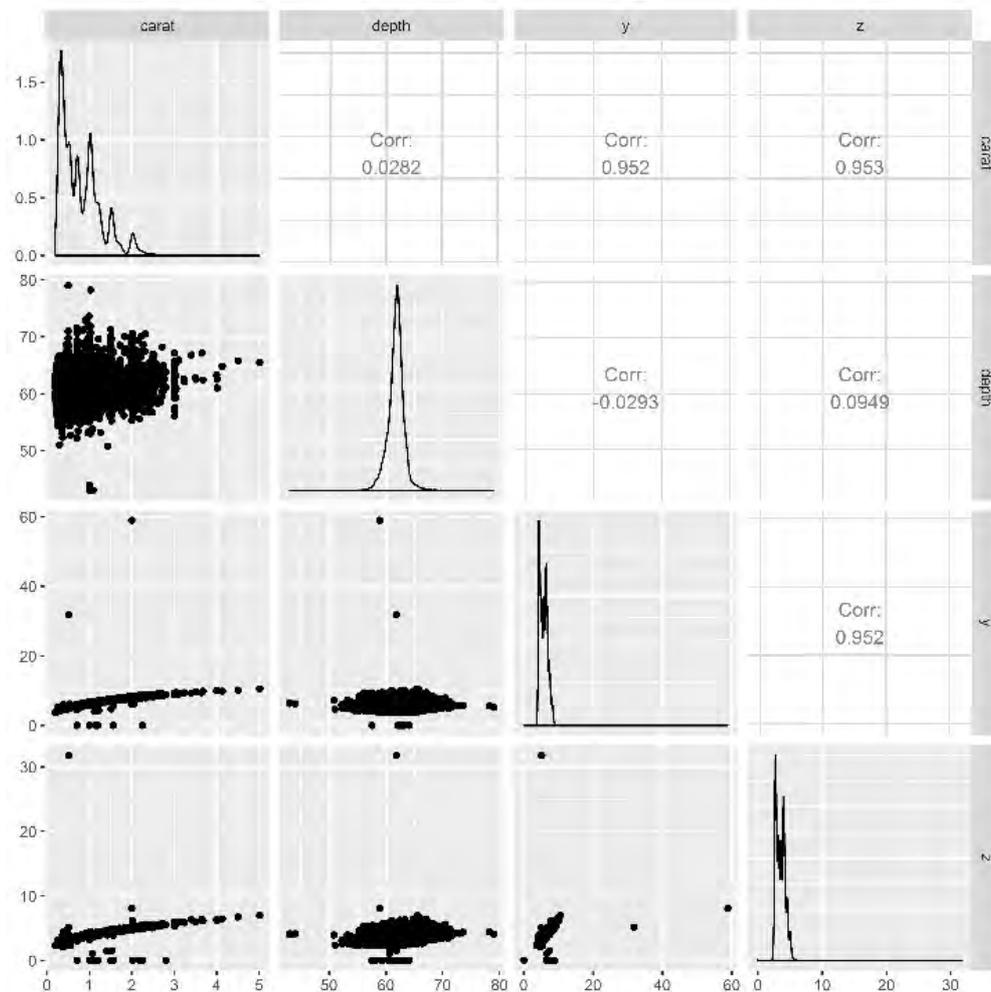
**Sous-dialogue**



→ Sélectionnez le Pairwise Plot. Puis appuyez sur Return (Retour)

→ Appuyez sur OK pour afficher les résultats illustrés ci-dessous (Fig. 15).

Fig. 15 Corrélations



## 7. Réflexions

Il est facile de suivre les instructions sans vraiment comprendre ce qui se passe. Nous énumérons ici quelques-uns des points qui ont été couverts:

- **File > Open Library**(Fichier>Ouvrir à partir de la bibliothèque) a été utilisé pour choisir un ensemble de données pour l'analyse. De même, la boîte de dialogue **File > Open Library** peut être utilisée pour importer vos propres données.
- Les données étaient pre-préparées pour cette analyse. Nous avons donc utilisé le menu **Describe** (Décrire).
- L'exploration initiale des données commence souvent par l'examen des variables une à la fois. Nous avons donc commencé par une description à l'aide de **Describe >One variable>Graph** (Décrire> Une variable> Graphique).
- Dans presque toutes les boîtes de dialogue, la première étape consiste à **Select the Variables** (sélectionner les variables) à analyser.
- Nous avons souvent eu à revenir à une boîte de dialogue pour reviser l'analyse.

- Les dialogues "se souvenaient" des réglages récents, de sorte que les changements se faisaient rapidement.
- Certains dialogues ont des sous-dialogues offrant des options additionnelles.
- Du point de vue statistique, nous avons facilement pu générer différents types de graphiques qui peuvent s'avérer utiles.
- Notez que la figure 15 est une fusion entre un graphique et un tableau, présentant quelques caractéristiques des deux.

## 8. Prochaines étapes

N'hésitez pas continuer à explorer le menu Describe avec cet ensemble de données et produire des tableaux et de graphiques additionnels pour explorer davantage les données. La partie suivante du tutoriel explore le menu **Prepare** (Préparer) à l'aide d'une deuxième base de données de la bibliothèque R-Instat.

## 9. Commentaires et signalement des bugs

R-Instat est encore en mode de développement avec de nombreuses améliorations et de nouvelles fonctionnalités prévues pour les futures versions. Merci de partager vos commentaires pour nous aider à améliorer R-Instat. Il existe plusieurs façons de nous les faire parvenir:

1. Pour des commentaires de type général, vous pouvez nous contacter par courriel à [R-Instat@AfricanMathsInitiative.net](mailto:R-Instat@AfricanMathsInitiative.net).

2. Notre page « problems » sur notre compte GitHub peut être utilisée pour signaler des bugs spécifiques ou des suggestions et c'est la meilleure façon de contacter l'équipe de développement. Notez l'accès à notre page "problems" ci dessous:

<https://github.com/africanmathsinitiative/R-Instat/issues>. Cliquez sur le bouton vert Nouveau problème sur le côté droit pour envoyer votre message.

Lorsque vous signalez un bug ou un problème, donnez-nous un maximum de précisions pour que nous puissions reproduire le bug, en collant le code R du fichier journal et en nous envoyant les données si possibles.

L'équipe R-Instat, Initiative africaine de données

# Tutoriel d'introduction : Partie 2: Une deuxième base de données

## Introduction

Ce guide fait suite à la partie 1 du tutoriel d'introduction à R-Instat. Nous vous recommandons de commencer par la première partie. Notez cependant que les étapes et de données utilisées dans cette partie ne dépendent pas de celles de la partie 1.

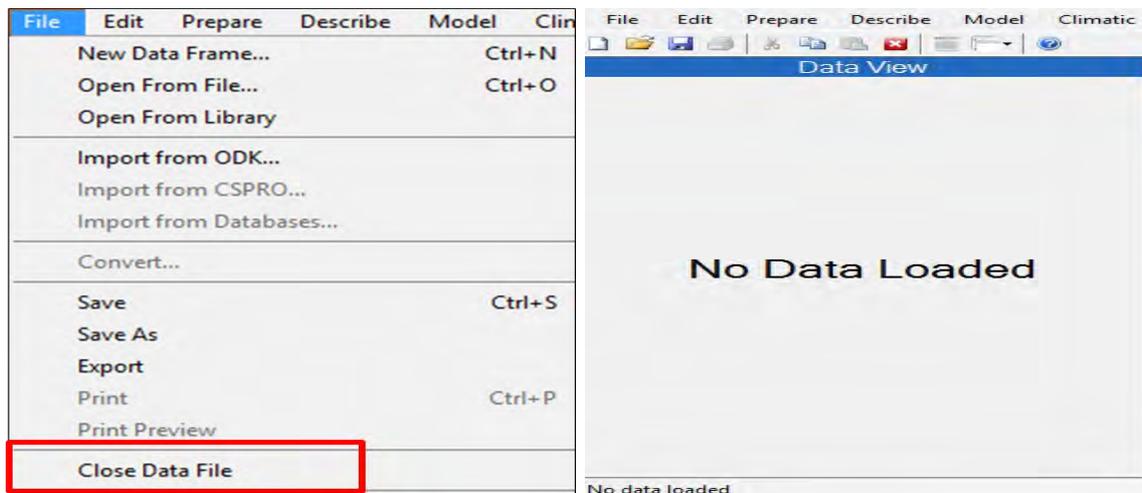
### 1. La base de données Dodoma

Il s'agit des données climatiques quotidiennes de Dodoma en Tanzanie, de 1935 à 2013. Nous sommes reconnaissants envers l'agence nationale de météorologie Tanzanienne de nous avoir permis d'utiliser ces données à des fins de formation.

→ Si les données Diamonds (diamants) sont toujours affichées, utilisez **File>Close Data File** (Fichier> Fermer le fichier de données) Fig. 16.

→ On vous demandera « are you sure » (si vous êtes sûr). Répondez **Yes** (Oui).

Fig. 16. Fermer le fichier precedent

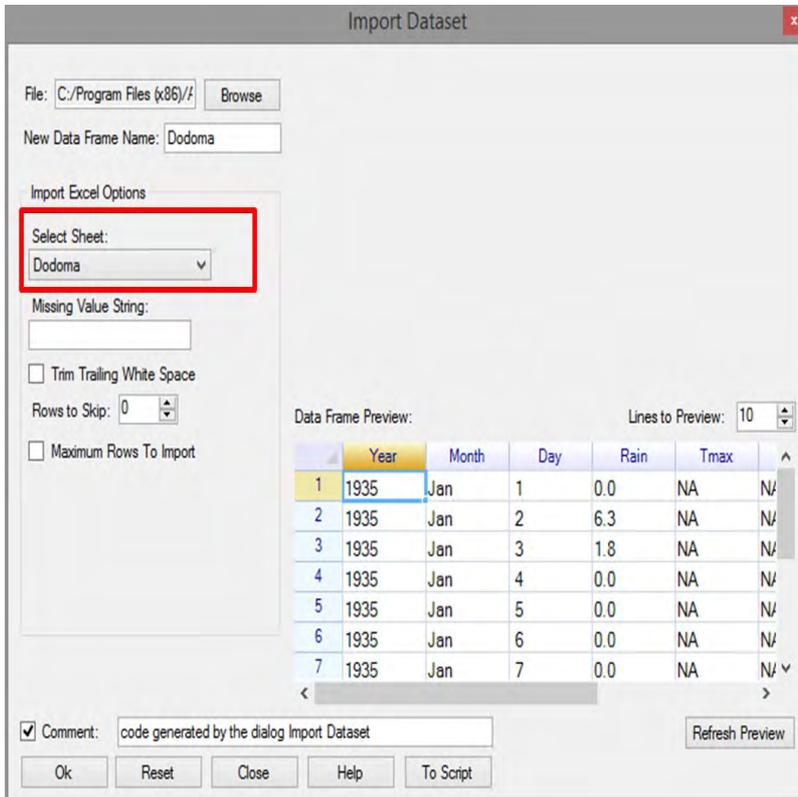


→ Utilisez **File > Open from Library**. (Fichier> Ouvrir depuis la bibliothèque). Prenez l'option Load from Instat Collection, puis appuyez sur Browse.

→ Choisissez **Climatic** et sélectionnez le fichier Excel **Climatic\_guide\_datasets**.

→ Ce fichier Excel a plusieurs feuilles. Choisissez celui appelé **Dodoma**, voir Fig. 17

**Fig. 17 Ouvrir le fichier Dodoma**



Un premier objectif est de fournir des graphiques de moyennes de températures annuelles, à la fois maximales et minimales. Comme ces données sont disponibles à l'échelle journalière, la première étape consiste à calculer la moyenne annuelle. Par conséquent, nous aurons recours au menu Prepare.

Notez sur la figure 1 que la base de données comprend 28 855 observations.

Une différence par rapport à l'exemple des diamants de la partie 1 est que les valeurs manquantes sont visibles immédiatement.

Fig. 18 Résumé des données quotidiennes de Dodoma

The screenshot shows a software interface with two main panels: 'Data View' and 'Output Window'.

**Data View:** A table with 7 columns: Year, Month (c), Day, Rain, Tmax, Tmin, and Sunh. The first 12 rows are shown, all for the year 1935 and month Jan. The 'Rain' column has values 0.0, 6.3, 1.8, 0.0, 0.0, 0.0, 0.0, 0.5, 0.0, 0.0, 0.0, 0.0. The 'Tmax' and 'Tmin' columns are all 'NA'. The 'Sunh' column is also all 'NA'. The status bar indicates 'Showing 1000 of 28855 rows | Showing 7 of 7 columns'.

**Output Window:** Contains R code and a summary of the data. The code is:
 

```
summary(object=InstatDataObject$get_columns_from_data(data_name="Dodoma",
col_names=c("Year","Month","Day","Rain","Tmax","Tmin","Sunh")), na.rm=FALSE)
```

 The summary output is:
 

Year	Month	Day	Rain
Min. :1935	Length:28855	Min. : 1.0	Min. : 0.00
1st Qu.:1954	Class :character	1st Qu.: 8.0	1st Qu.: 0.00
Median :1974	Mode :character	Median :16.0	Median : 0.00
Mean :1974		Mean :15.7	Mean : 1.57
3rd Qu.:1994		3rd Qu.:23.0	3rd Qu.: 0.00
Max. :2013		Max. :31.0	Max. :119.80
			NA's :91

 Below this, there is another summary for Tmax, Tmin, and Sunh:
 

Tmax	Tmin	Sunh
Min. :15	Min. : 8	Min. : 0
1st Qu.:27	1st Qu.:15	1st Qu.: 8
Median :29	Median :17	Median :10
Mean :29	Mean :17	Mean : 9
3rd Qu.:30	3rd Qu.:18	3rd Qu.:11
Max. :36	Max. :26	Max. :14
NA's :8631	NA's :8703	NA's :18451

→ Utilisez la boîte de dialogue **Describe > One variable > Summary** (Décrire> Une variable> Résumer).

→ **Choisissez toutes les colonnes**, puis appuyez sur **OK** pour produire les récapitulatifs illustrés à la Fig. 18.

Les résultats incluent le nombre de valeurs manquantes et il s'avère que plus de 8 000 valeurs de température sont manquantes. (Comme cette caractéristique n'était pas évidente dans la sortie similaire de la partie 1 (figure 12), il s'ensuit que les données sur les diamants n'avaient aucune valeur manquante.)

Les données de précipitations sur la figure 18 sont de 1935. La station a ajouté des enregistrements de température plus tard.

→ Cliquez à l'aide du **bouton droit** sur **l'onglet du bas** et choisissez la dernière option **View Data** (Afficher les données) pour afficher l'ensemble des données.

→ Faites défiler ces données pour confirmer que les températures ont commencé à partir de 1958.

Ceci indique que la plupart des 8 000 données de température manquantes de la figure 18 s'expliquent par le fait que les mesures de ces éléments ont commencé plus tard que les autres.

Souvent, la préparation des données pour l'analyse prend plus de temps que l'analyse elle-même. Nous avons essayé de rendre le menu Prepare (Préparer) aussi simple que possible à utiliser. Il y a 5 étapes à parcourir, même pour des tâches simples comme celles-ci. Nous espérons que vous apprécierez, les étapes ci-dessous. Nous verrons dans la section 4 qu'elles offrent un bon côté.

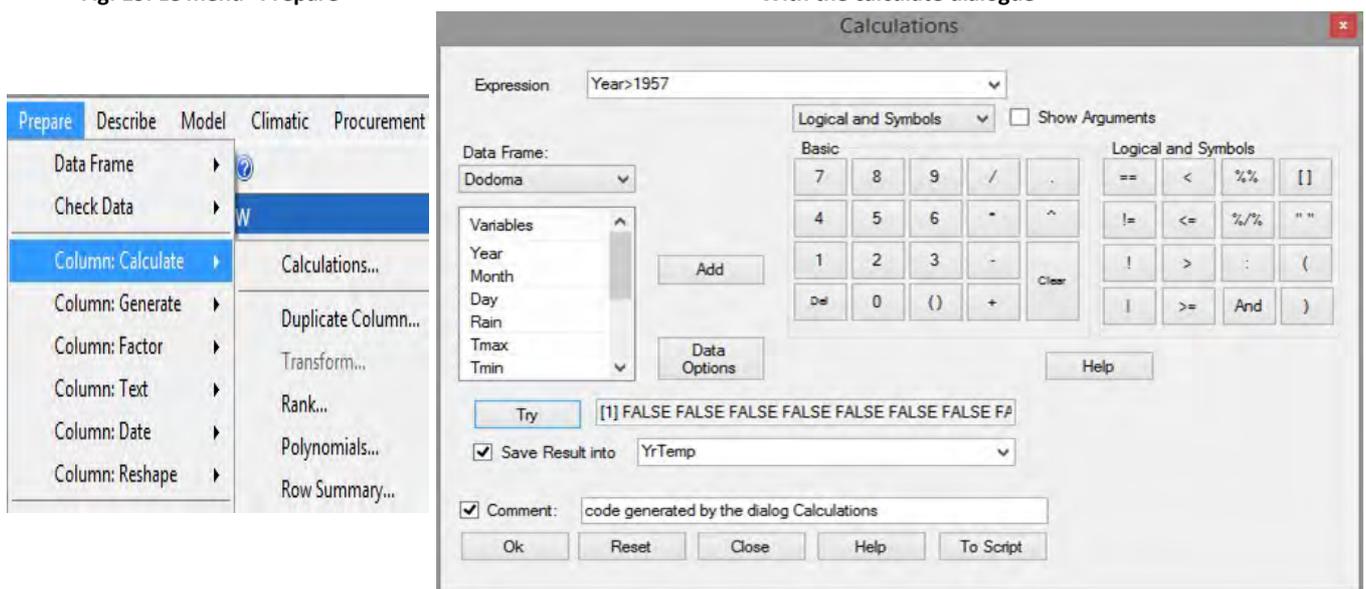
## Préparer les données

Souvent, l'étape de préparation comprend le calcul de colonnes additionnelles.

→ Ouvrez le dialogue **Prepare > Column: Calculate > Calculations** (Préparer> Colonne: Calculer> Calculs) comme indiqué ci-dessous Fig. 19.

Fig. 19. Le menu "Prepare"

With the calculate dialogue



Ce menu fonctionne en tant que calculateur de colonnes. Plusieurs claviers sont disponibles.

→ Cliquez sur le contrôle **Basic** et choisissez **Logical and Symbols**. Un clavier supplémentaire s'ouvre comme indiqué sur la figure 19.

→ **Double-cliquez sur** la colonne **Year** (Année), (ou cliquez et appuyez sur Add (Ajouter) pour la placer dans le champ de la formule, qui se trouve en haut de la boîte de dialogue.

→ Complétez la formule en ajoutant **> 1957**, de façon à lire **Year (Année) > 1957**, voir Fig. 19.

→ Cliquez sur le bouton Try et le résultat devrait être **FALSE, FALSE, FALSE ...** comme sur la figure 19, car les premières lignes de données datent de 1935 - donc pas plus de 1957!

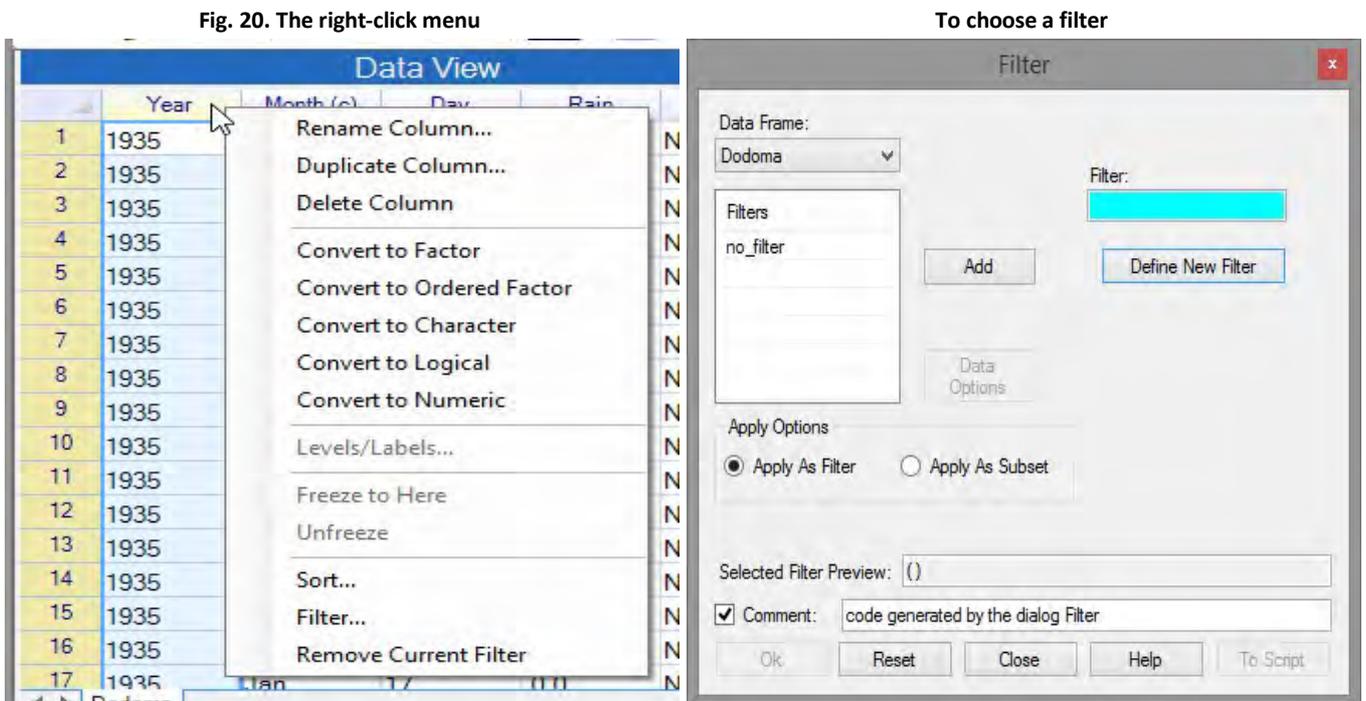
→ Donnez un nom à la nouvelle colonne pour enregistrer les résultats, comme par exemple **YrTemp**. Puis appuyez sur **OK**.

Une nouvelle colonne de données a été créée.

L'étape suivante consiste à appliquer un **filtre**, de sorte que les données à analyser ne commencent qu'en 1958, et que « TRUE » apparaisse dans la nouvelle colonne. De nombreuses tâches courantes du menu Prepare (Préparer) sont rapidement accessibles via un menu spécial clic-droit qui est illustré ci-dessous (Fig. 20.)

→ Placez le curseur sur la ligne du haut (contenant les noms) et faites un **clik droit**, Fig. 20.

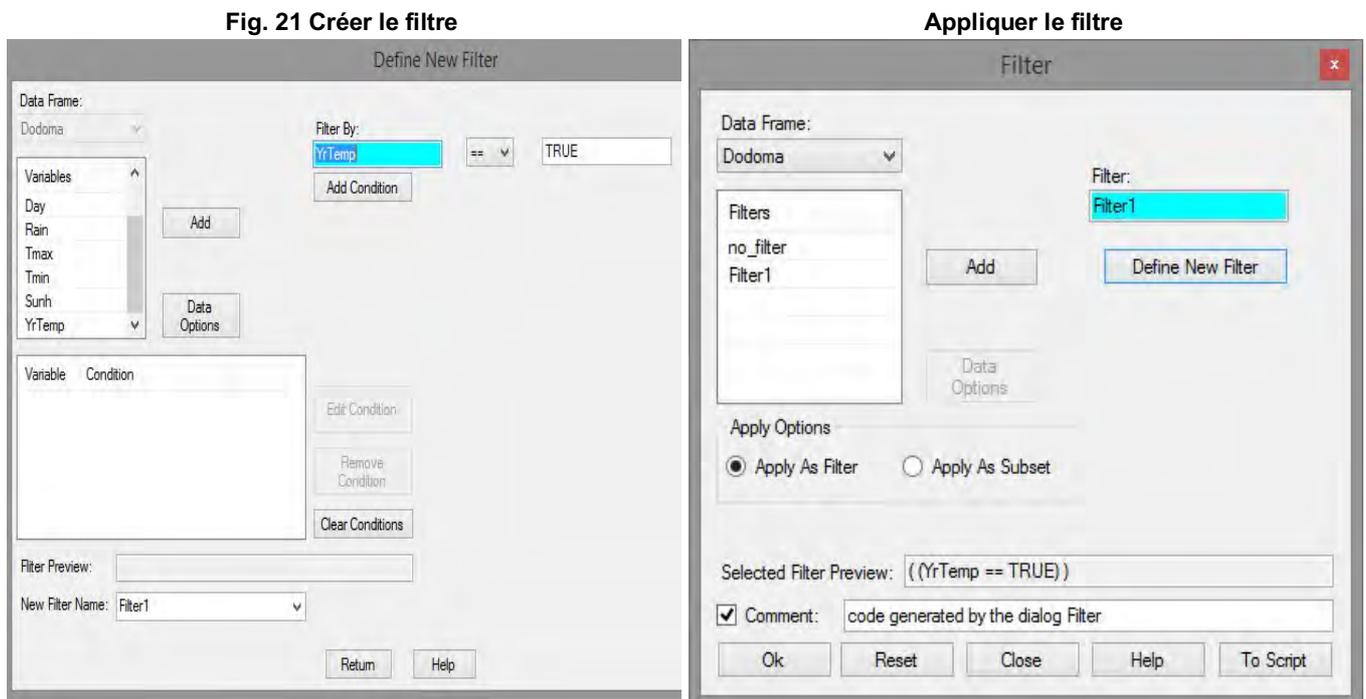
→ Choisissez le dialogue **Filter** (Filtre) dans ce menu, voir Fig 20.



→ Cliquez sur **Define New Filter** (définir un nouveau filtre) Fig. 20.

→ Dans le sous-dialogue, choisissez la colonne YrTemp. Remplissez la condition pour qu'elle indique **YrTemp == TRUE**

(Notez que == n'est pas une erreur, et le mot **TRUE** doit être en majuscules, Fig. 21)



→ Appuyez sur **Add Condition** pour ajouter une condition, Fig. 21, puis appuyez sur **Return** (Retour).

→ Dans le dialogue principal du filtre, appuyez sur **OK** pour appliquer le filtre (fig. 21.). Notez que la première colonne, qui contient les numéros correspondant aux lignes, est maintenant affichée en rouge et la première est la ligne 8402, c'est-à-dire le 1er janvier 1958.

La troisième étape préparatoire consiste à **changer le mode de la colonne Year (Année) de numeric (numérique) à factor (factorielle)**.

→ Placez le curseur dans la colonne **Year** (Année) et dans la rangée supérieure (nom). Faites un **clic droit**, Fig. 22.

→ Cliquez sur **Convert to Ordered Factor** (Convertir en facteur ordonné).

**Fig. 22. Conversion numérique > factorielle de la colonne Année**

**Les données résultantes**

	Year	Month (c)	Day	Rain	Tmax	Tmin
8402	1958				28.6	18.7
8403	1958	Jan	2	0.0	29.7	18.8
8404	1958	Jan	3	0.0	29.7	17.6
8405	1958	Jan	4	7.1	30.5	18.8
8406	1958	Jan	5	8.9	31.2	19.2
8407	1958	Jan	6	2.0	31.1	19.1
8408	1958	Jan	7	0.0	27.2	18.1
8409	1958	Jan	8	0.0	28.9	18.8
8410	1958	Jan	9	0.0	30.0	16.7
8411	1958	Jan	10	0.0	30.1	17.3
8412	1958	Jan	11	0.0	31.2	19.3
8413	1958	Jan	12	0.0	31.2	19.1
8414	1958	Jan	13	0.0	32.1	18.3
8415	1958	Jan	14	0.0	31.8	18.6
8416	1958	Jan	15	0.0	32.9	18.3
8417	1958	Jan	16	0.0	33.6	17.8
8418	1958	Jan	17	0.0	34.1	19.2

Les données quotidiennes sont maintenant prêtes à être résumées pour produire les moyennes annuelles.

→ Ouvrez le dialogue **Prepare > Column: Reshape > Column Summaries** (Préparer> Colonne: Remodeler> Résumés de colonnes) Fig. 23.

Fig. 23. Le résumé de colonnes

Le dialogue résultant

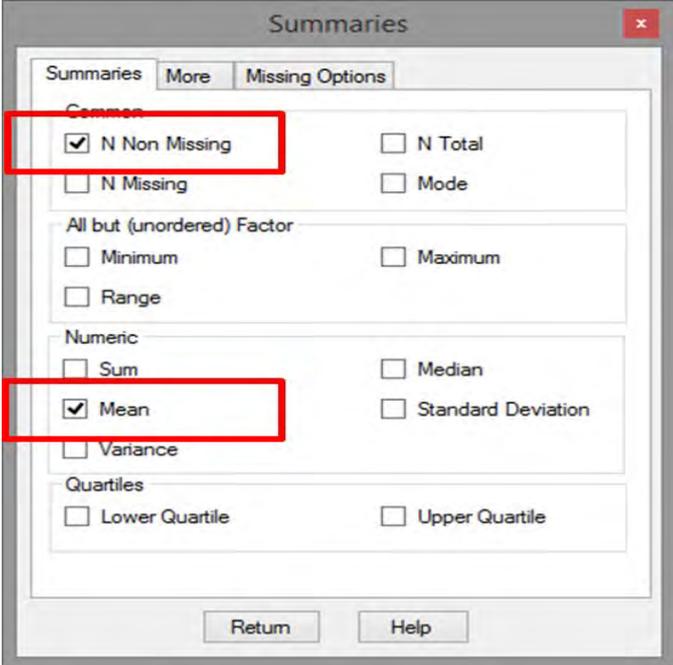
→ Complétez les informations comme indiqué sur la Fig. 23, c'est-à-dire **Tmin et Tmax** dans le dialogue principal, **Année** dans le sous-dialogue, et l'option cochée pour **Omit Missing Values** (Omettre les valeurs manquantes).

→ Appuyez ensuite sur **Summaries** pour passer au sous-dialogue, Fig. 24.

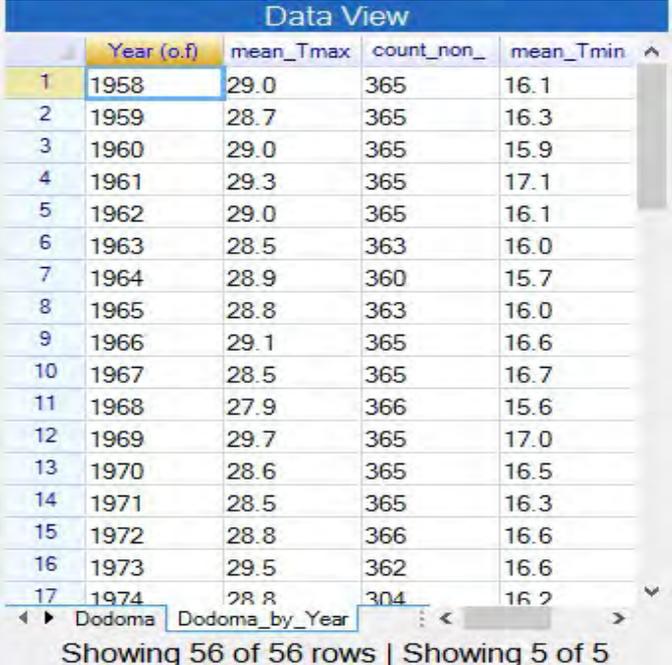
→ Complétez l'info du sous-dialogue comme indiqué sur la figure 24, c'est-à-dire avec seulement deux résumés pour le **N Not Missing** (N Ne manquant pas) et le **Mean** (la moyenne). Puis appuyez sur Return (Retour).

→ Appuyez sur **OK** pour produire les résumés, Fig. 24.

**Fig. 24. Résumé – le sous-dialogue**



**Et le données résultantes**



	Year (o.f)	mean_Tmax	count_non_	mean_Tmin
1	1958	29.0	365	16.1
2	1959	28.7	365	16.3
3	1960	29.0	365	15.9
4	1961	29.3	365	17.1
5	1962	29.0	365	16.1
6	1963	28.5	363	16.0
7	1964	28.9	360	15.7
8	1965	28.8	363	16.0
9	1966	29.1	365	16.6
10	1967	28.5	365	16.7
11	1968	27.9	366	15.6
12	1969	29.7	365	17.0
13	1970	28.6	365	16.5
14	1971	28.5	365	16.3
15	1972	28.8	366	16.6
16	1973	29.5	362	16.6
17	1974	28.8	304	16.2

La figure 24 démontre également que nous avons maintenant **deux bases de données**, l'une à l'échelle quotidienne et l'autre avec les résumés annuels. Cette seconde data frame est celle que nous utiliserons pour produire les graphiques.

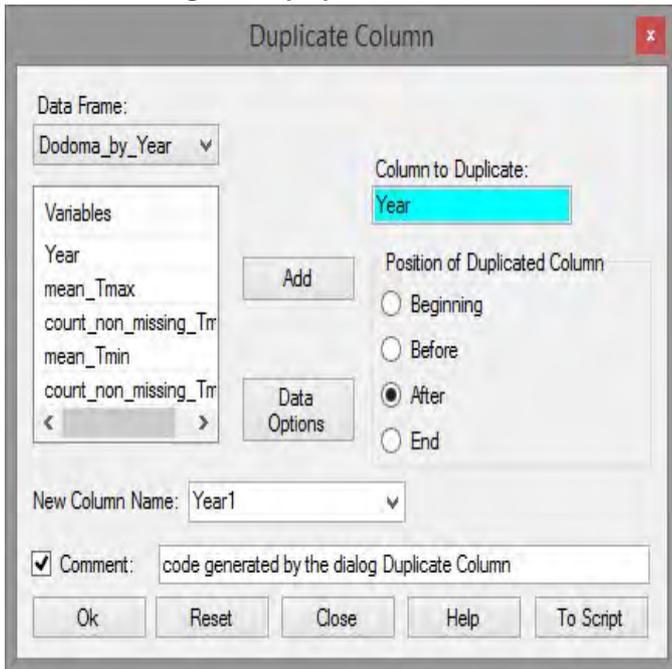
### 3. Produire les graphiques

Il ne reste plus qu'une dernière étape préparatoire. La colonne Year (Année) de la colonne Summary Data (Résumé de données) est en mode factor (factorielle). Pour les graphiques, il faudra la reconvertir au mode numeric (numérique). Il est souvent pratique d'avoir accès aux deux différents modes (numérique et factoriel)

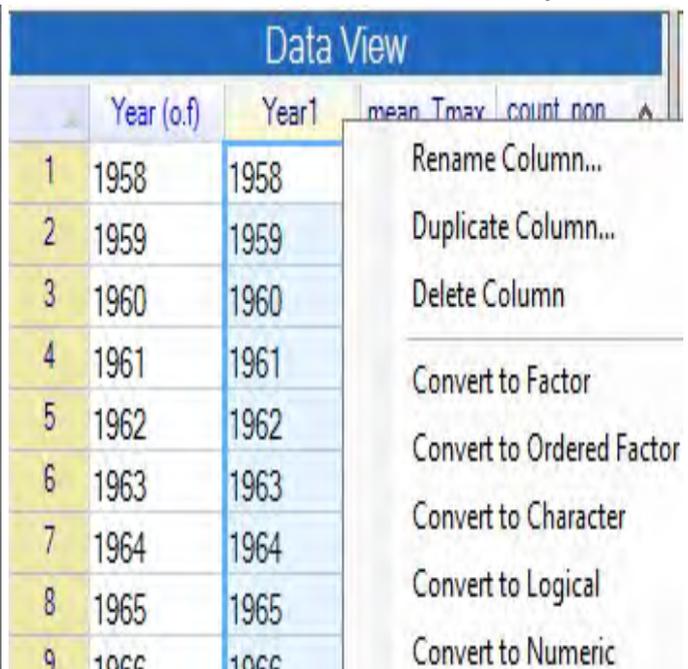
→ Utiliser **Prepare > Calculate > Duplicate Column** (Préparer> Calculer> Colonne en double (ou faites un clic droit et choisissez l'élément approprié.)) → Complétez le dialogue comme indiqué sur la Fig. 25. Appuyez sur **OK** pour produire une autre colonne appelée Année1.

→ **Cliquez** avec le **bouton droit** sur le titre/le nom de la colonne **Year1** et convertissez la colonne en mode **numeric** (numérique) Fig. 25.

Fig. 25. Dupliquer une colonne



Convertir la colonne en mode numérique

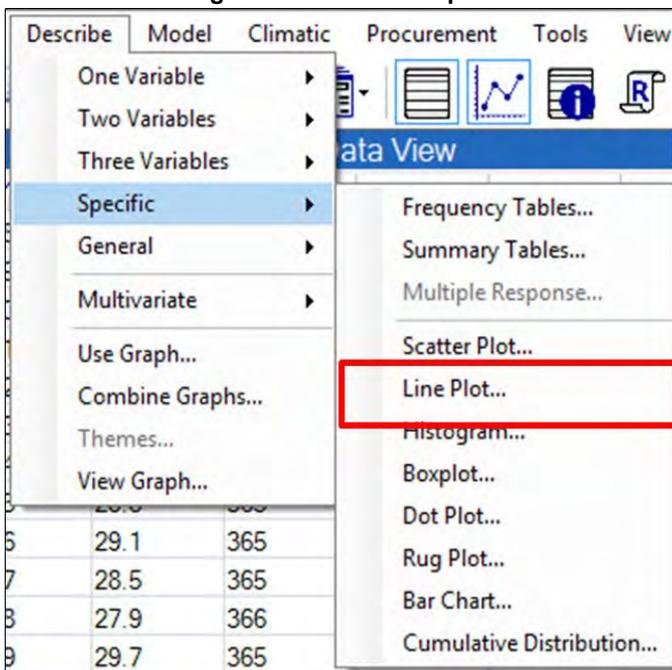


Nous sommes enfin prêts à créer les graphiques.

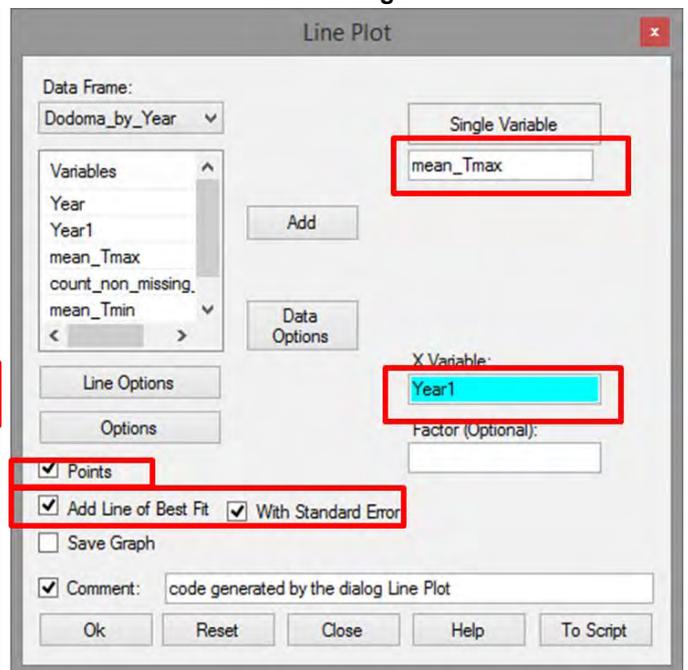
→ Utiliser **Describe > Specific > Line Plot** (Décrire> Spécifique> Tracé de ligne), Fig. 26.

→ Complétez le dialogue comme indiqué sur la Fig. 26 pour le **mean\_Tmax**. Appuyer sur **OK**.

Fig. 26. The menu line plot



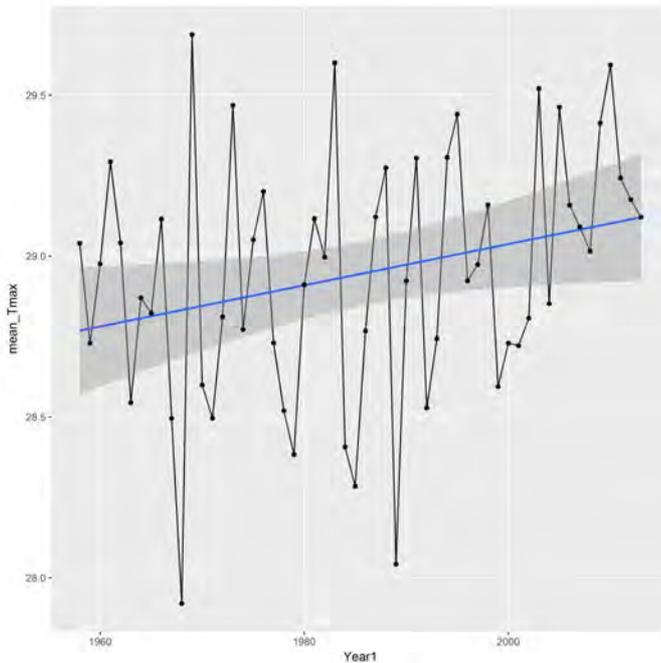
Et le dialogue



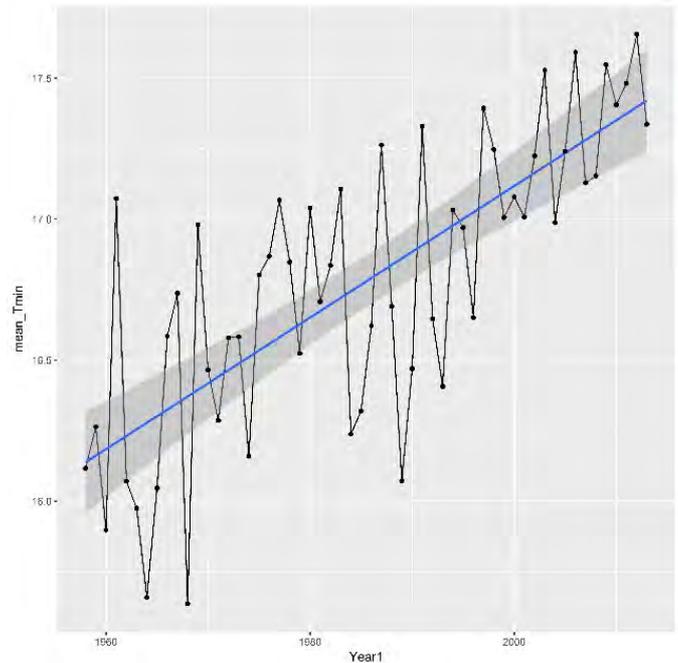
Le graphique produit est représenté sur la figure 27.

→ Retournez à la boîte de dialogue Line Plot et remplacez **mean\_Tmin** par **mean\_Tmax**. Appuyez sur **OK** pour afficher le second graphique également illustré ci-dessous Fig. 27

Fig. 27. Le graphique Tmax



Et Tmin



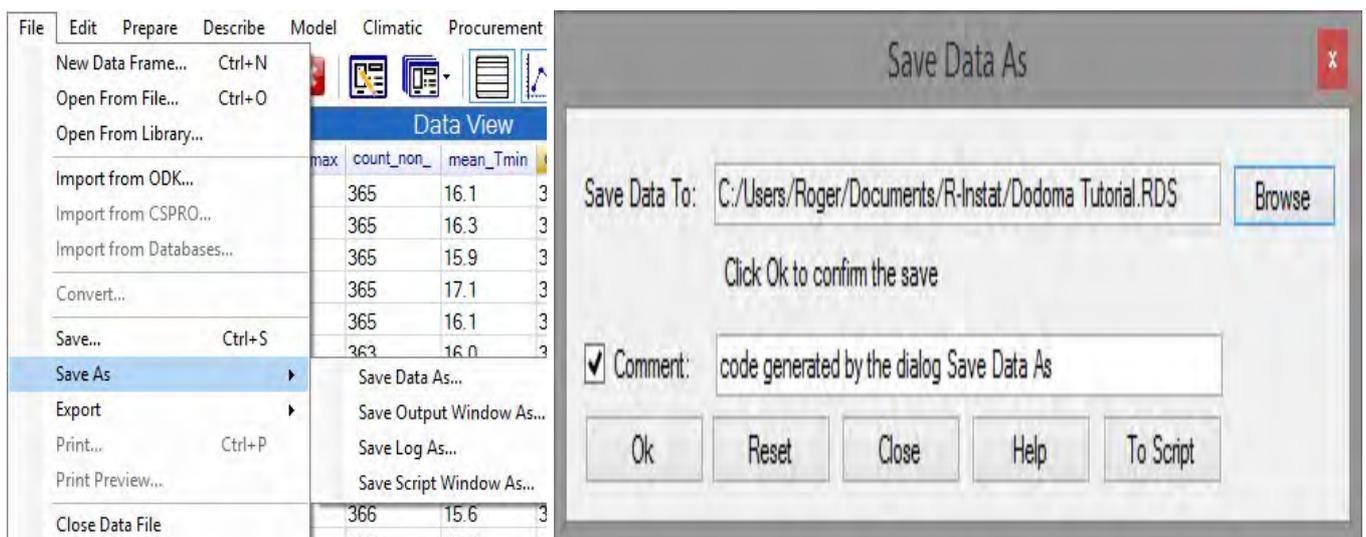
#### 4. Sauvegarder / enregistrer les données

Avant d'utiliser une base de données différente, enregistrez/ sauvegardez ces données afin de pouvoir les retrouver plus tard.

→ Utilisez la boîte de dialogue **File > Save As** (Fichier > Enregistrer sous), Fig. 28. Choisissez l'option **Save Data As** (Enregistrer les données sous).

→ Appuyez sur **Browse** (Parcourir), Fig. 28. Choisissez un répertoire et un nom appropriés. Appuyez sur **OK** dès votre retour à la boîte de dialogue Save data (Enregistrez les données).

Fig. 28. Enregistrer/sauvegarder les données



L'extension RDS est incluse pour indiquer que le fichier est enregistré en tant que fichier de données R. C'est le bon côté auquel nous avons fait allusion dans la section 1. Si elle est bien faite, les données ne

doivent être organisées qu'une seule fois. Le fichier et les data frames, peuvent être réutilisés et l'analyse poursuivie ultérieurement.

### **3. Prochaines étapes**

Il y a plus d'analyses qui peuvent être explorées avec ces données dans R-Instat et nous vous encourageons maintenant à essayer. La partie suivante de ce tutoriel se concentre sur l'utilisation de données étiquetées.

### **4. Commentaires et signalement des bugs**

R-Instat est encore en mode de développement avec de nombreuses améliorations et de nouvelles fonctionnalités prévues pour les futures versions. Merci de partager vos commentaires pour nous aider à améliorer R-Instat. Il existe plusieurs façons de nous les faire parvenir:

1. Pour des commentaires de type général, vous pouvez nous contacter par courriel à [R-Instat@AfricanMathsInitiative.net](mailto:R-Instat@AfricanMathsInitiative.net).

2. Notre page « problems » sur notre compte GitHub peut être utilisée pour signaler des bugs spécifiques ou des suggestions et c'est la meilleure façon de contacter l'équipe de développement. Notez que notre page de problems est accessible à tous:

<https://github.com/africanmathsinitiative/R-Instat/issues>. Cliquez sur le bouton vert New Issue (Nouveau problème) sur le côté droit pour envoyer votre message.

Lorsque vous signalez un bug ou un problème, donnez-nous un maximum de précisions pour que nous puissions reproduire le bug, en collant le code R du fichier journal et en nous envoyant les données si possible.

Équipe R-Instat, Initiative africaine de données