R Commander : Petit guide pratique
1. Statistiques de base

Jean-philippe Gaudron

Février 2016

Ce petit guide pratique a été écrit pour faciliter l'utilisation de R Commander. A partir de courts exercices et de questions de recherche simples, il permet d'apprendre à décrire des données univariées et bivariées et à les analyser avec ce logiciel simple, puissant et gratuit. La plupart des chapitres portent sur un test en particulier et peuvent donc être travaillés indépendamment les uns des autres. Mais tous nécessitent de connaître l'interface de R Commander (chapitre Prise en main) et de savoir comment construire ou importer un tableau de données (chapitre Construire un tableau de données…) avec R Commander.

Ce guide a été écrit en utilisant la version 2.2.3 de R Commander pour Windows. Mais compte tenu des faibles différences entre les versions pour Mac, LINUX (Ubuntu) et Windows (le chapitre Prise en main présente l'interface de R Commander sous les trois environnements), il peut être utilisé pour les trois. A ce titre, les annexes présentent les procédures de chargement et d'installation de R et de R Commander sous Windows, Mac et Ubuntu.

Enfin et surtout, ce petit guide pratique n'a pas vocation à se substituer à un manuel d'introduction à la statistique. Au contraire, il part de principe que les lecteurs et lectrices connaissent les statistiques descriptives ainsi que les tests d'inférence statistique usuels, leurs conditions d'utilisation et surtout leur pertinence par rapport aux questions de recherche posées.

Ce document est protégé par les droits d'auteur. Il peut être téléchargé et utilisé gratuitement. Toute modification, même partielle, n'est autorisée qu'avec l'accord écrit de l'auteur.

Pr Jean-Philippe Gaudron
Département des Sciences de l'éducation et de la formation
Université Toulouse Jean Jaurès
F-31058 Toulouse
France
gaudron@univ-tlse2.fr

Table des matières

1. R Commander : découverte et prise en main 5
1.1 Démarrer R Commander 5
1.2 Faire des calculs avec R 8
1.3 Copier, enregistrer les résultats et fermer R Commander 9
2 Construire un tableau de données, spécifier les variables et modifier les valeurs 10
2.1 Construire un tableau de données avec R Commander 10
2.2 Enregistrer un tableau de données 13
2.3 Importer un jeu de données 14
2.4 Ouvrir un fichier de données déjà enregistré sous R 19
2.5 Coder les variables 20
2.6 Calculer de nouvelles variables 22
2.7 Découper une variable numérique en classes 23
2.8 Générer des jeux de données partiels 24
3 Statistiques descriptives 27
3.1 Statistiques descriptives numériques 28
3.1.1 Statistiques descriptives par variable 28
3.1.2 Statistiques descriptives par modalité de variable (sous-groupes) 30
3.2 Construire des graphiques pour décrire des données 33
3.2.1 Dessiner un histogramme 33
3.2.2 Dessiner un graphe Tiges et feuilles 34
3.2.3 Dessiner un graphe quantile-quantile : 36
3.2.4 Dessiner une boîte à moustaches. 37
3.2.5 Dessiner un graphe des moyennes 39
3.2.6 Dessiner un graphe en camembert 41
3.2.7 Dessiner un graphe en barres 42
4 Tableaux croisés et Chi-deux d'indépendance (*Chi-square test of independance*) 44
4.1 Calculer un Chi-deux à partir d'un tableau croisé 44
4.2 Calculer un Chi-deux sur un jeu de données 46
5 La corrélation avec le r de Bravais-Pearson *(correlation with Pearsonôs r)* 48
5.1 Vérifier la nature linéaire de la relation entre les deux observations Avant et Après 49
5.2 Vérifier que les variables Avant et Après se distribuent normalement 50
5.3 Calculer la corrélation entre les deux variables 52
5.4 Calcul de la corrélation pour sous-groupes 52

5.4.1 Vérifier la nature linéaire de la relation entre Avant et Après pour le sous-groupe des filles 53
5.4.2 Vérifier la normalité des deux distributions des scores Avant et Après pour ce sous-groupe des filles 54
5.4.3 Calcul de la corrélation pour un sous-groupe 57
6 La corrélation avec le rho de Spearman (*correlational analysis : Spearmanôs rho*) 60
6.1 Vérifier la nature linéaire de la relation en observant le nuage de points 61
6.2 Vérifier que les variables Temps et Performance se distribuent normalement 62
6.3 Calculer la corrélation entre les deux variables Temps et Performance 63
7 Le test *t* de Student pour un échantillon ou test *t* univarié *(one sample t-test)* 64
8 Le test *t* de Student pour groupes indépendants *(independant sample t-test)* 66
8.1 Vérifier la normalité des groupes 67
8.2 Apprécier l'égalité des variances des 2 groupes 69
8.3 Calcul du *t* de Student pour groupes indépendants 70
9 Le test *U* de Mann-Withney (ou test de la somme des rangs de Wilcoxon) pour groupes indépendants (*Mann-Withney U-test* or *Wilcoxon sum rank test*) 71
10 Le test *t* de Student pour groupes appariés (*paired sample, ou related t-test*) 73
10.1 Vérifier la normalité des 2 groupes 73
10.2 Calcul du test *t* pour groupes appariés 75
10.3 Calcul du test *t* de Student pour sous-groupes appariés 76
10.3.1 Vérifier la normalité d'un sous-groupe 76
10.3.2 Calcul du test *t* pour deux sous-groupes appariés 79
11 Le test du signe de Wilcoxon pour groupes appariés (*Wilcoxon signed ranks test*) 81
12 ANOVA pour groupes indépendants (*One-Way between-groups*) 83
12.1 Tester la normalité des groupes 84
12.2 Tester l'égalité des variances des 3 groupes 86
12.3 Calcul de l'ANOVA pour groupes indépendants 87
12.4 Comparaisons *a posteriori* 88
13 La régression linéaire simple (*Simple Regression*) 89
13.1 Vérifier la nature linéaire de la relation entre *X* et *Y* en observant le nuage de points 90
13.2 Calculer la corrélation entre les deux variables Temps et Performance 91
13.3 La droite de regression 92
13.4 Les tests d'hypothèse : 93
13.4.1 Vérifier les conditions d'homogénéité des variances et de normalité dans les vecteurs en observant les graphiques 94

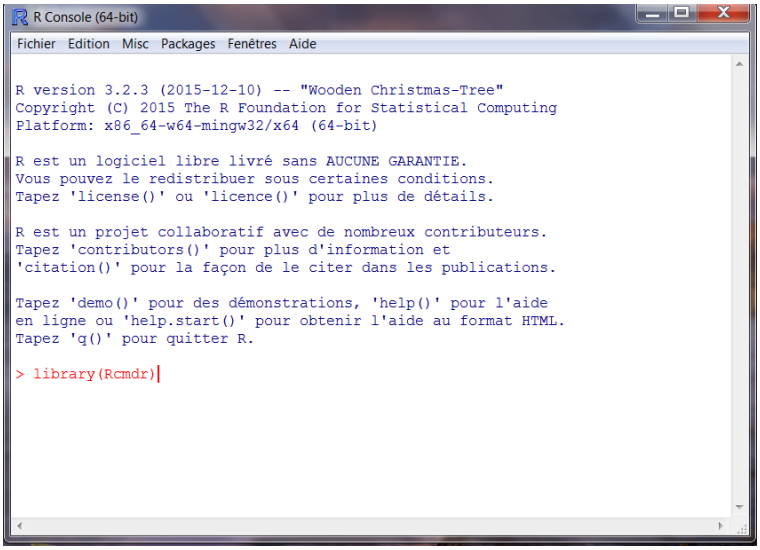
13.4.2 Pour aller plus loin 95
14 Références 98
15 Annexes 99
15.1 Installer R et R Commander sous Windows 100
15.2 Installer R et R Commander sous Mac 108
15.3 Installer R et R Commander sous Ubuntu (Linux) 124

1. R Commander : découverte et prise en main

R Commander (Fox, 2005), ou encore Rcmdr est une interface graphique (*Graphical User Interface*, GUI) du logiciel libre R. R Commander permet de calculer des statistiques de base descriptives et inférentielles à partir de menus déroulants, sans avoir donc à écrire les commandes et maîtriser le langage R. R Commander repose sur la souplesse et la puissance de calcul de R et permet de construire des graphiques, autre point fort de R. Mais R Commander n'est que la partie visible de l'iceberg R, c'est à dire que c'est un paquet (*package*) parmi les centaines de paquets élaborés et offerts par la communauté des chercheurs pour enrichir les fonctionnalités de R, ou pour proposer des méthodes particulières. Pour autant, sa simplicité et son aspect intuitif permettent aux chercheur.e.s de répondre à nombre de questions quantitatives qu'elles et ils se posent ; et petit à petit, de découvrir les potentialités de R.

1.1 Démarrer R Commander

Pour démarrer R Commander, il faut tout d'abord démarrer R en double cliquant sur l'icône R se trouvant sur le bureau. La fenêtre R s'ouvre avec plusieurs paragraphes en bleu donnant quelques informations sur R (voir ci-dessous). Sous le texte, il y a le symbole rouge >, appelé prompt. C'est après ce symbole qu'il faut taper les lignes de commande de R qui s'affichent également en rouge.
Pour lancer R Commander, après le > il faut taper library(Rcmdr) et entrer.



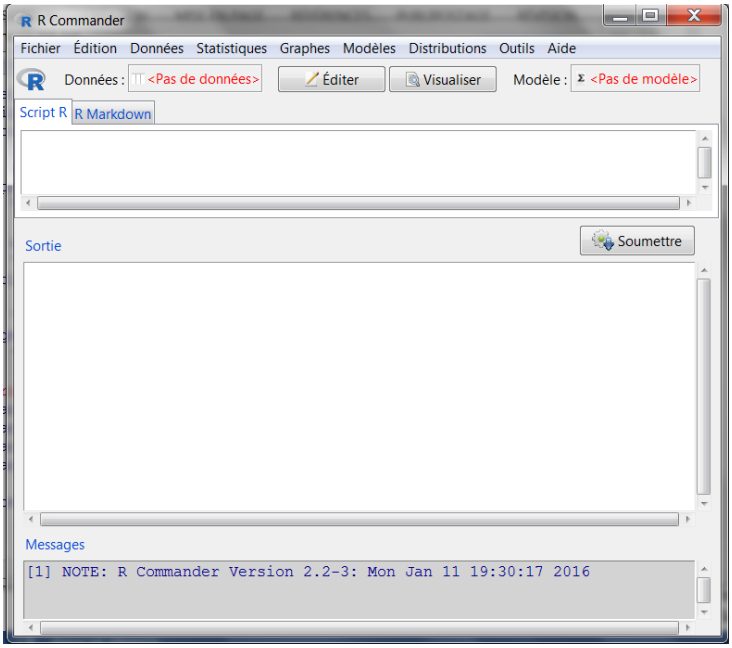
Si la commande est incomplète, le symbole > est remplacé par +
Si la commande est fausse (par exemple, la majuscule de Rcmdr a été oubliée), un message erreur s'affiche. R distingue donc majuscules et minuscules. Ajoutons que les é, è, î,… posent problème : il faut éviter tous les accents, même dans les noms des colonnes des fichiers importés. Dernière précision typographique : les décimales étant en anglais, il faut mettre des points à la place des virgules.

Lorsque R Commander est chargé, la fenêtre s'affiche à l'écran. Comme toute fenêtre, elle reproduit le graphique habituel de l'environnement de l'ordinateur, Windows, Mac ou Ubuntu (Linux). Pour autant, l'arborescence des menus est identique (il y a les mêmes options, les mêmes commandes) et les algorithmes de calculs sont les mêmes. Par la suite, nous utiliserons

le graphique de Windows, mais on aura compris que les procédures, les étapes sont identiques sous Mac ou sous Ubuntu. Ci-dessous sont présentées les fenêtres R Commander sous les différents environnements.
Pour Mac :

Pour Ubuntu :

Pour Windows :



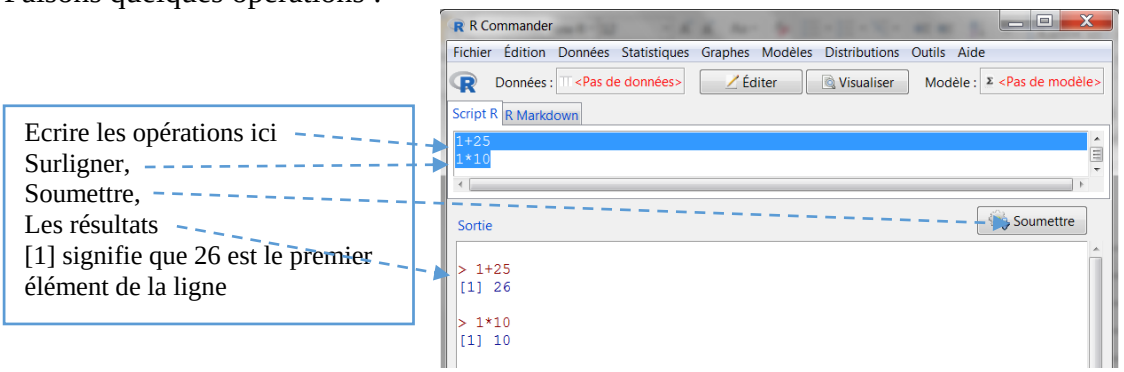
Examinons la fenêtre R Commander
On peut l'agrandir en glissant la souris à partir des bords ou en cliquant le bouton d'agrandissement.
Tout en haut, le nom du programme R Commander et les 3 icônes pour masquer, agrandir/réduire, ou fermer.
En dessous, la barre « menu », avec Fichier, Edition, Données… En cliquant sur un mot, apparaît un menu déroulant à partir duquel on choisit les actions ou les analyses, on le verra plus loin.
Sous la barre des menus, quatre touches : Pas de données (lorsque l'on travaillera avec un jeu de données, son titre apparaîtra là ; Editer ; Visualiser ; Pas de Modèle. Ce sont des touches raccourcis et que l'on retrouve dans les menus déroulants
Une première petite fenêtre blanche sous les deux onglets, script R et Markdown : c'est la fenêtre des commandes dans le langage de programmation de R, appelées scripts. Chaque clic de la souris sur une icône ou une commande d'un menu déroulant correspond à une programmation et y fera apparaître un script. On trouve d'ailleurs le même principe dans certains logiciels de statistiques tel SPSS avec la possibilité de visualiser le langage de programmation dans une fenêtre spécifique appelée Editeur de syntaxe. Simplement, dans R Commander, 'l'éditeur de syntaxe' est sous les yeux en permanence, ce qui est une bonne façon de connaître le langage de R. D'ailleurs, lorsque nous souhaiterons programmer une analyse un peu spécifique, il suffira d'écrire le script dans la fenêtre, le sélectionner (le surligner avec la souris), puis de cliquer sur Soumettre. Et parfois, pour aller plus vite lorsqu'il s'agira de procéder à des traitements répétitifs, il suffira de reprendre le script, de le modifier à sa guise, de le sélectionner et de soumettre.
Notons que la fenêtre des scripts de R Commander est équivalente à la fenêtre de R : une commande lancée sous R Commander est reconnue par R et inversement.
La grande fenêtre en dessous est la fenêtre des sorties : c'est là que nous lirons les statistiques demandées (les commandes sont en rouge), c'est-à-dire les résultats (qui sont en bleu).
Tout en bas, une troisième petite fenêtre pour des messages, qui peuvent apparaître sous la forme d'avis (en vert), de notes (en bleu) ou d'erreurs (en rouge).

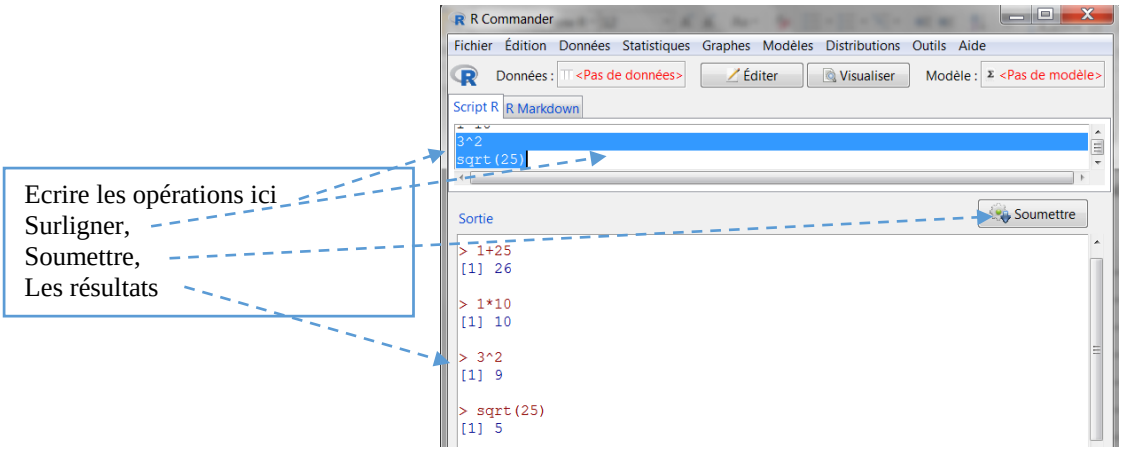
1.2 Faire des calculs avec R

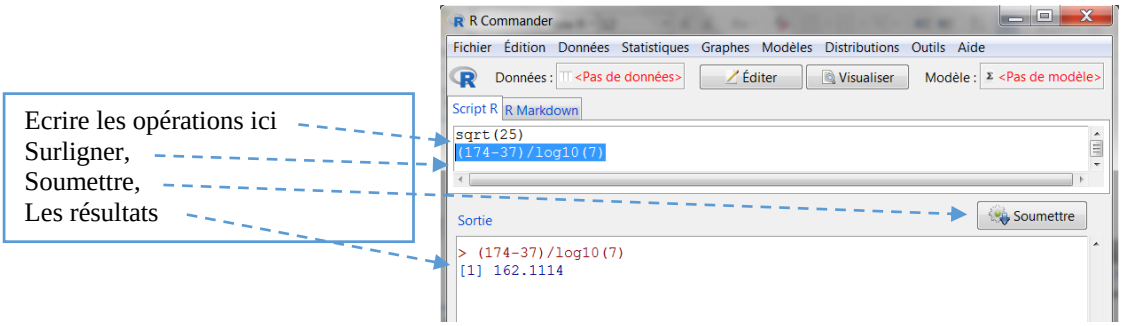
R est avant tout un calculateur, donc R Commander aussi puisque les deux fenêtres sont équivalentes. Le tableau ci-dessous présente quelques opérations élémentaires possibles. Il suffit écrire l'opération dans la fenêtre commande, de surligner et de soumettre (dans R, il suffit d'écrire l'opération après le > et entrer) :

Opération	Signe	Exemple
Addition	x + y	1 + 25
Soustraction	x - y	5 - 1
Multiplication	x * y	1*10
Division	x / y	34/63
Elever à la puissance	x ^ y	3^2
Racine carrée	sqrt(x)	sqrt(25)
Transformation logarithmique	log10(x) ou log(x,10)	Log10(7)

Faisons quelques opérations :





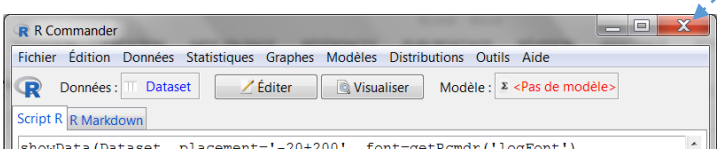


1.3 Copier, enregistrer les résultats et fermer R Commander

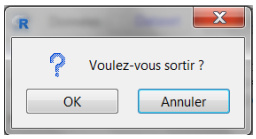
On peut rapidement copier ce qui apparaît dans la fenêtre des sorties, tout simplement en sélectionnant l'extrait que l'on souhaite copier avec la souris, puis Copier (bouton droit). Dans le traitement de texte, cliquer Coller : les résultats sont là.
Pour les graphiques, les images, l'option Copier les images est proposée dans le menu Editer ou en cliquant sur bouton droit de la souris. Il suffit ensuite de coller là où l'on souhaite, ou encore de l'enregistrer comme fichier

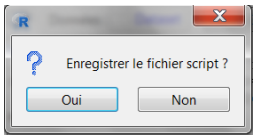
Avant de fermer R Commander, il ne faut pas oublier d'enregistrer le fichier actif au risque de perdre des modifications apportées.

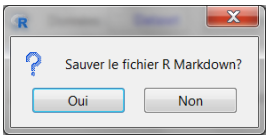
Pour fermer R Commander, à moins de passer par le menu déroulant, Fichier Ą Sortir, il suffit de cliquer sur la croix rouge en haut à droite de la fenêtre de R Commander :

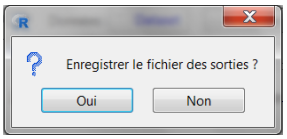


Quatre fenêtres se suivent, demandant la confirmation de la fermeture et l'enregistrement ou non le fichier script (pour réutiliser une autre fois les commandes utilisées), les scripts avec R Mardown, enfin les sorties (pour les conserver).

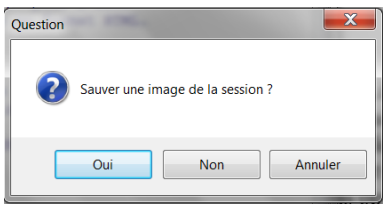


puis




enfin


La fenêtre R Commander se ferme.
Il faut maintenant fermer R en cliquant également sur la croix rouge en haut à droite et répondre à une dernière fenêtre :



Notons que si l'on a fermé R Commander et que l'on souhaite le ré-ouvrir dans la même session de R, il faut écrire en face du prompt, la commande : Commander()

2 Construire un tableau de données, spécifier les variables et modifier les valeurs

Si l'on utilise des statistiques pour résoudre des problèmes, c'est que l'on dispose d'un grand nombre de données numériques. Pour effectuer les statistiques souhaitées, il est obligatoire de présenter ces données sous la forme d'un tableau dans lequel chaque colonne est une variable et chaque ligne un individu (individu au sens statistique du terme, c'est-à-dire une personne, une chose, une famille, un phénomène…, bref, tout ce à quoi s'intéresse la statistique). Chaque case, appelée cellule, est la valeur donnée à un individu (ou sa réponse s'il s'agit par exemple d'un questionnaire) à une variable (par exemple à une question).
Dans ce genre de tableau, la toute première ligne est réservée aux noms des colonnes, c'est à dire au nom des variables (ne pas mettre d'accent et d'espace). La première colonne est généralement réservée aux noms des lignes, c'est à dire les noms des individus qui sont souvent les numéros attribués aux individus, aux protocoles.
Un tel tableau des données peut-être très grand puisqu'il comporte autant de colonnes que de variables (dans un sondage, chaque question est a minima une variable) et autant de lignes que d'individus ayant participé. R Commander permet de construire des tableaux comportant moins de 10 000 valeurs. On peut aussi importer un tableau de données élaboré dans un tableur (Open Office, Excell…) ou dans d'autres logiciels de statistiques tels SPSS, SAS. Ces tableaux peuvent avoir plus de 10 000 valeurs.

Une pratique qui accélère la saisie consiste à remplacer les mots des modalités d'une variable par un chiffre (par exemple 1 pour les filles et 2 pour les garçons, ou encore les chiffres associés aux catégories socio-professionnelles, aux années d'études…). Ces chiffres ne font que nommer ces modalités. Nous verrons ci-dessous comment leur attribuer de nouveau leur nom (coder les variables).
Enfin, pour éviter les erreurs de saisie et pour aller plus vite, on entre les valeurs telles qu'elles apparaissent dans les protocoles. S'il faut faire la somme à plusieurs réponses, nous le ferons plus tard ; s'il faut inverser les réponses à une question négative, nous le ferons pareillement plus tard ; plus tard également s'il faut regrouper des scores en deux ou trois classes. En fait, ce sont des fonctions ou des calculs que permet de faire automatiquement R Commander lorsque que l'on souhaite modifier une variable, ce que nous verrons bientôt (modifier les variables).

2.1 Construire un tableau de données avec R Commander

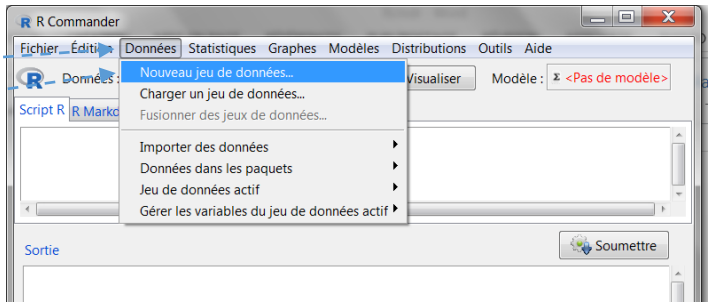
Construire un tableau de données avec R Commander se fait avec l'éditeur de données. Cet éditeur de données présente des fonctions semblables à celles des tableurs : il permet de créer un nouveau tableau de données, de modifier les données d'un tableau existant, d'ajouter ou de supprimer des lignes (*row*) et des colonnes (*column*) au tableau, de copier (*copy*), coller (*paste*) ou couper (*cut*) le contenu des cellules (*cell*). Cependant, il est moins ergonomique, moins rapide pour la saisie des données, il offre moins de possibilités qu'un vrai tableur et il est limité à 10 000 valeurs rappelons-le. Bref, il s'avère pratique pour de petits jeux de données et de petits exercices, sans plus.
Voici à titre d'exercice un petit jeu de données recueillies par une enseignante qui a relevé les notes à une évaluation passée par 6 élèves, trois filles et trois garçons. Comme il a été précisé plus haut, pour effectuer des statistiques, il est nécessaire de présenter les données sous la forme d'un tableau dans lequel chaque colonne est une variable et chaque ligne un individu et le nom des variables en haut des colonnes. Il y ici trois variables, les élèves, leur sexe (codé ici 1 pour

les filles et 2 pour les garçons) et leur note. Et il y a 6 individus, ici 6 élèves : 6 lignes donc, plus la première ligne pour les noms des variables. Voici le tableau que nous allons construire :

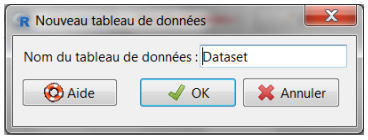
Elève	Sexe	Note
1	1	12,5
2	2	9
3	1	17
4	2	15,5
5	2	12
6	1	10

A partir de la fenêtre R Commander :

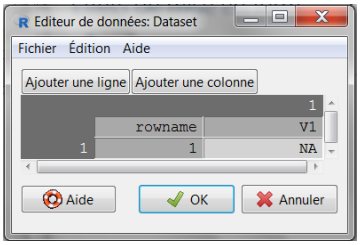
Etape 1
Données
Nouveau jeu de données



La fenêtre Nouveau tableau de données s'ouvre permettant de donner un nom. Laissons le nom par défaut (Dataset) et OK :

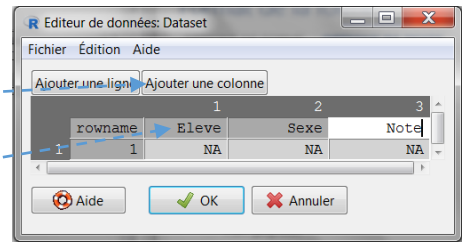


La fenêtre intitulée 'Editeur de données' s'ouvre avec le nom choisi (ici par défaut, Dataset).



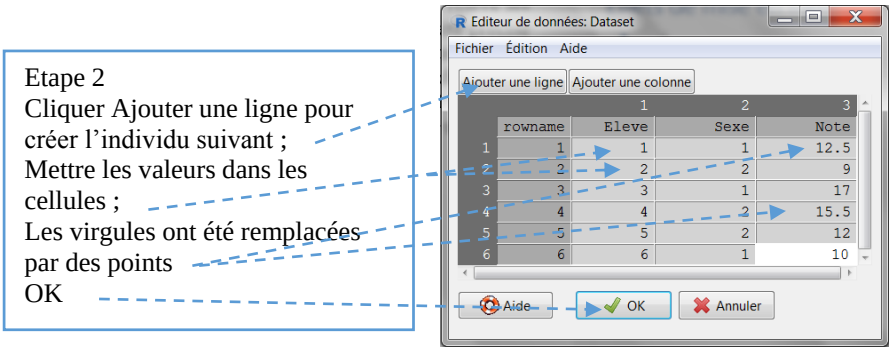
Le plus pratique est de commencer par créer le tableau vierge, c'est-à-dire de créer le nombre de colonnes désirées. On ajoute une colonne en cliquant sur l'onglet 'Ajouter une colonne'. Puis on nomme la colonne.

Etape 2
Cliquer Ajouter une colonne deux fois ;
Nommer les variables en écrivant directement sur les cellules V1, V2 et V3

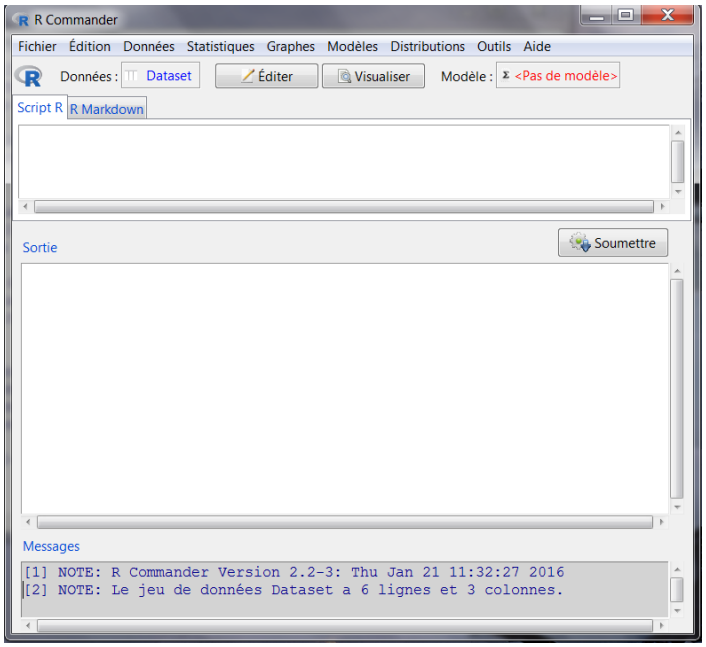


Lorsque les colonnes sont prêtes, on peut passer aux lignes, c'est-à-dire que l'on remplit le tableau individu par individu, c'est-à-dire ligne par ligne. Pour remplir le tableau, il suffit de cliquer dans une cellule et de mettre la valeur correspondante qui s'affiche à la place de NA. NA signifie valeur manquante : s'il y en a, il suffira de sauter la cellule. Comme on le verra au cours des différentes parties, R Commander gère ces valeurs manquantes en proposant par exemple d'en tenir compte, ou pas, dans certains calculs ; ou encore en proposant une option dans laquelle, les lignes comportant des valeurs manquantes sont supprimées lors de la création de nouveaux fichiers avec une partie des données. On passe d'une cellule à l'autre à l'aide des flèches du clavier. Arrivé au bout de la première ligne, on ajoute une seconde ligne en cliquant sur l'onglet Ajouter une ligne etc.
Attention, rappelons que dans la numérotation anglaise, les décimales s'écrivent avec un point à la place de la virgule : par exemple ici, à la place de 12,5, il faut entrer 12.5.

Voici le tableau construit :



Comme il n'y a pas de donnée manquante, le tableau final ne comporte aucune cellule avec NA.
Lorsque la saisie est terminée, appuyez sur OK. La fenêtre 'Editeur de données' disparaît et laisse place à la fenêtre R Commander. En haut, après 'Données', le nom du tableau 'Dataset' apparaît : c'est le jeu de données actif. Tout en bas, une note bleue signale qu'un jeu de données est actif, avec son nom (Dataset) et les nombres de lignes et de colonnes.



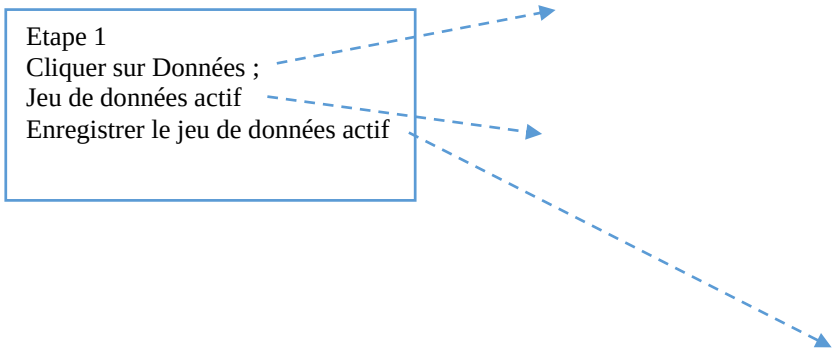
On peut visualiser le tableau sans modification possible, en cliquant sur le bouton Visualiser :

Pour retirer la fenêtre, cliquer sur la croix rouge.

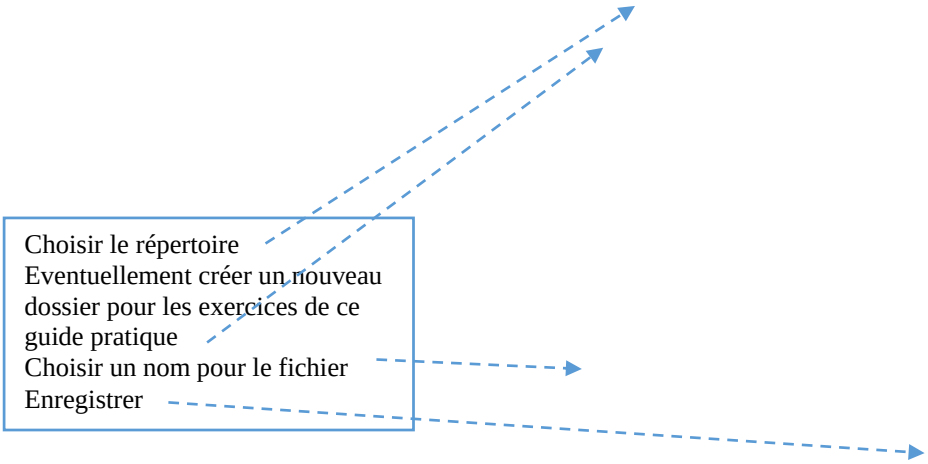
On peut l'éditer pour procéder à des modifications, à l'ajout de nouvelles données ou du nom des variables ou des modalités en cliquant sur le bouton Editer. Et pour retirer cette fenêtre, cliquer sur OK.

2.2 Enregistrer un tableau de données

Il est temps d'enregistrer ce fichier actif.



La fenêtre Enregistrer sous apparaît :.



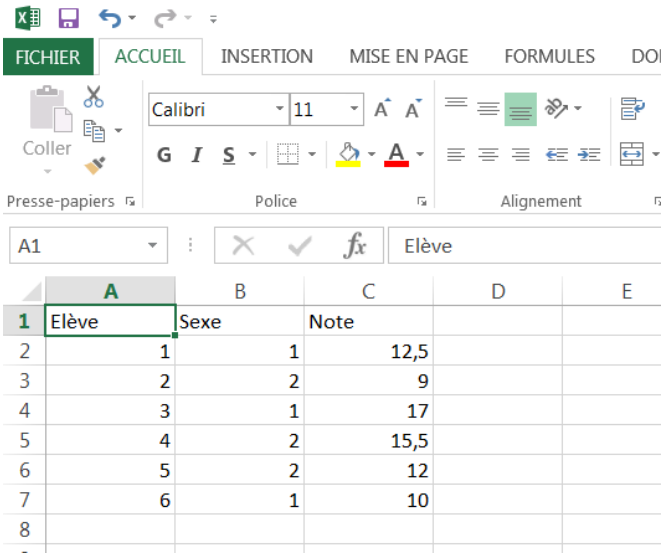
2.3 Importer un jeu de données

Comme beaucoup de logiciels de statistiques, R Commander peut travailler à partir de tableaux construits sur des tableurs (Excel, Open Office…) ou d'autres logiciels (SPSS, SAS, Minitab et STATA…). Il offre à ce titre une très grande souplesse et certains paquets (*packages*) sont dédiés à cela. On peut ainsi choisir le logiciel parmi une liste proposée et on se laisse ensuite guider et répondre aux questions posées quand il y en a. Ainsi, si l'on a un tableau construit dans l'un de ces logiciels, il suffit d'aller sur Données dans la barre de menu, puis Importer un jeu de données et de choisir le logiciel utilisé.
Si l'on a un jeu de données dans une autre logiciel que ceux proposés, ou si la démarche ci-dessus n'aboutit pas (ce qui arrive pour les nouvelles versions de ces logiciels qui ne sont pas encore intégrées dans la version présente de R Commander), la procédure reste néanmoins extrêmement simple, ce que nous allons faire ci-dessous.

Nous allons construire un tableau de données dans le tableur Excel, sachant que la procédure est identique pour le tableur gratuit de la suite Open Office. Nous utiliserons le même petit jeu de données que ci-dessus :

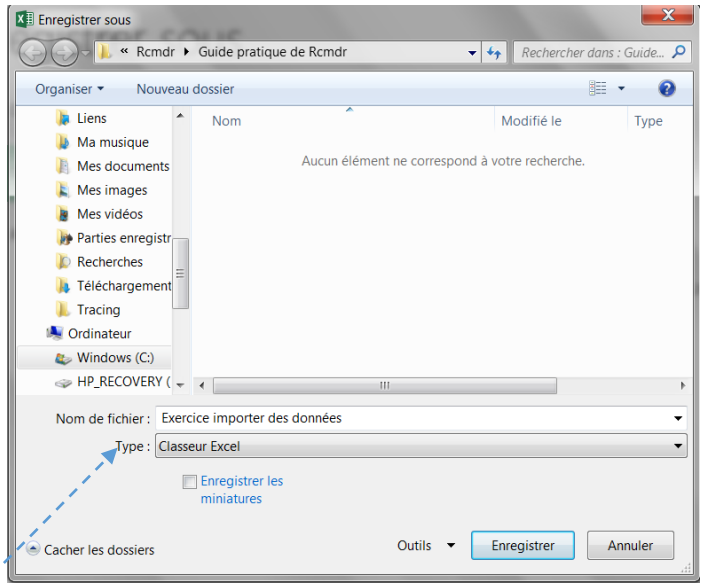
Elève	Sexe	Note
1	1	12,5
2	2	9
3	1	17
4	2	15,5
5	2	12
6	1	10

Voici dans Excel, le tableau crée avec le nom des variables sur la première ligne :



	A	B	C	D	E
1	Elève	Sexe	Note		
2	1	1	12,5		
3	2	2	9		
4	3	1	17		
5	4	2	15,5		
6	5	2	12		
7	6	1	10		
8					

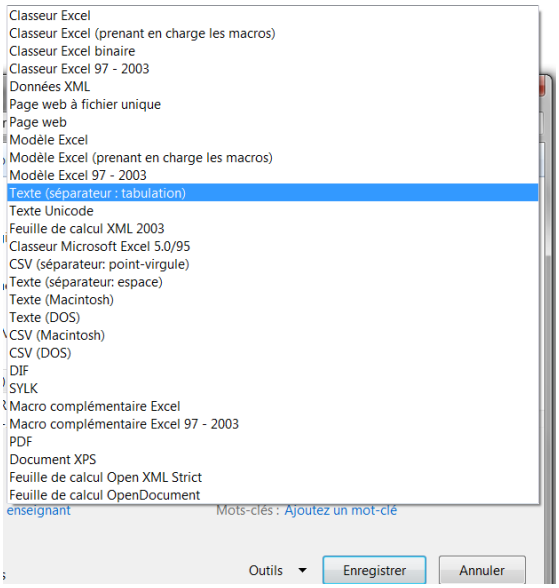
Aller dans Fichier ⇨ Enregistrer sous, pour ouvrir la fenêtre Enregistrer sous :



Cliquer sur Type et la fenêtre des formats d'enregistrement s'ouvre :

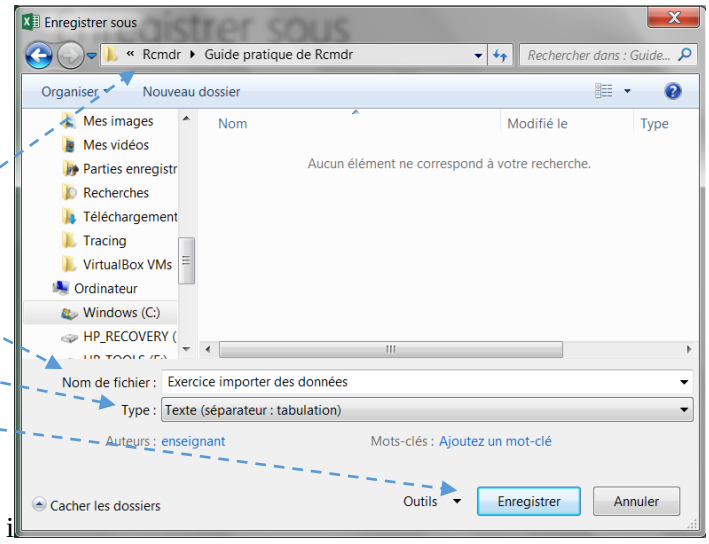
Choisir Texte (séparateur : tabulation):

On pourrait prendre également Texte avec séparateur espace ou point-virgule

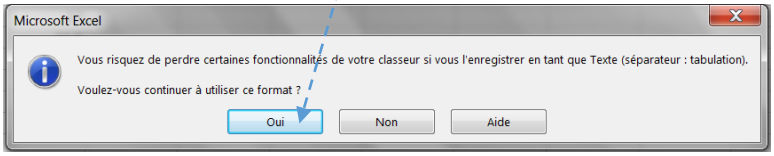


La fenêtre Enregistrer sous revient

Choisir votre répertoire
Choisir un nom pour le fichier
Le Type choisi est indiqué (Texte (séparateur : tabulation)):
Enregistrer



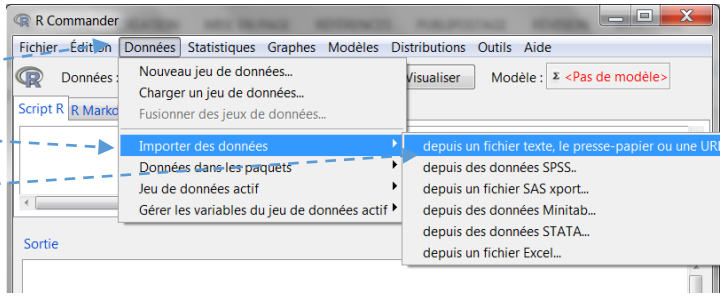
Si la fenêtre suivante apparaît, cliquer Oui



Fermer le fichier Excel.

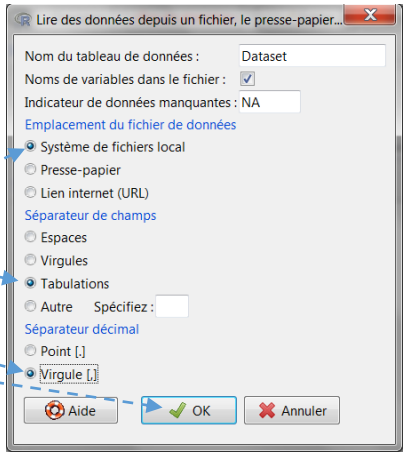
Retournons maintenant dans R Commander :

Cliquer sur :
Données
Importer des données
Depuis un fichier texte, le presse papier ou une URL

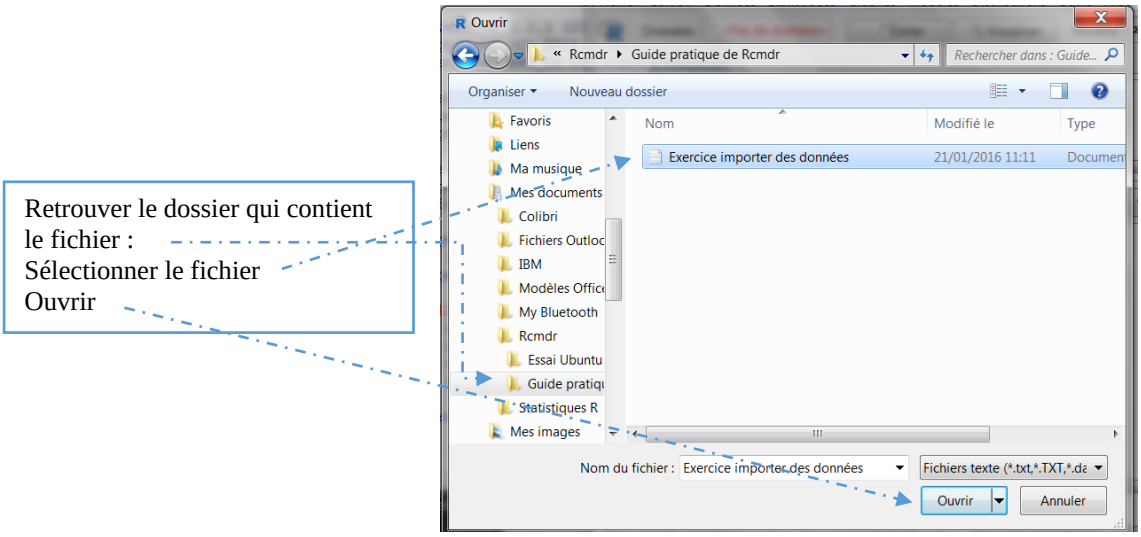


La fenêtre 'Lire depuis un fichier, le presse papier' s'ouvre. Elle offre propose plusieurs options qu'il faudra spécifier : l'emplacement du fichier de données ; le séparateur de champs, le séparateur décimal.

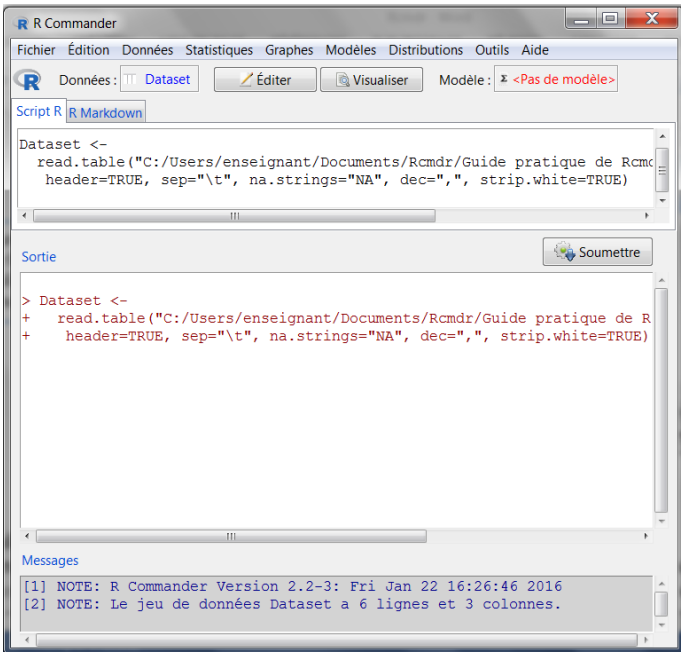
Choisir système de fichiers local :
Le séparateur utilisé dans Excel est Tabulations
Dans le fichier de données, il y a des virgules
OK



La fenêtre Ouvrir apparaît. Aller dans le dossier dans lequel le tableau Excel a été enregistré.



La fenêtre R Commander s'ouvre avec tout en bas une seconde note bleue indiquant, le fichier ouvert, son nom et les nombres de lignes et colonnes.

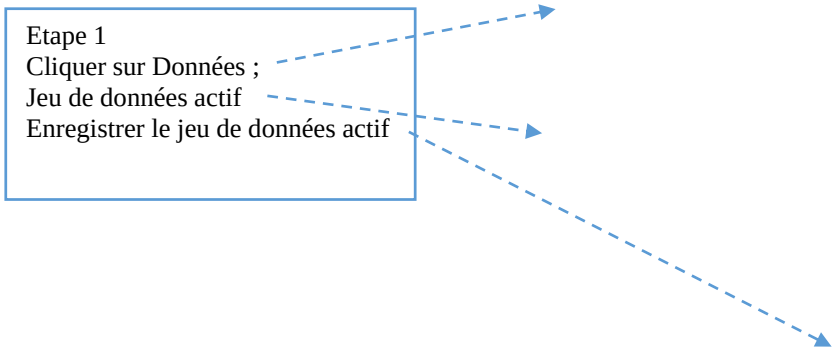


On peut visualiser le tableau sans modification possible, en cliquant sur le bouton Visualiser :

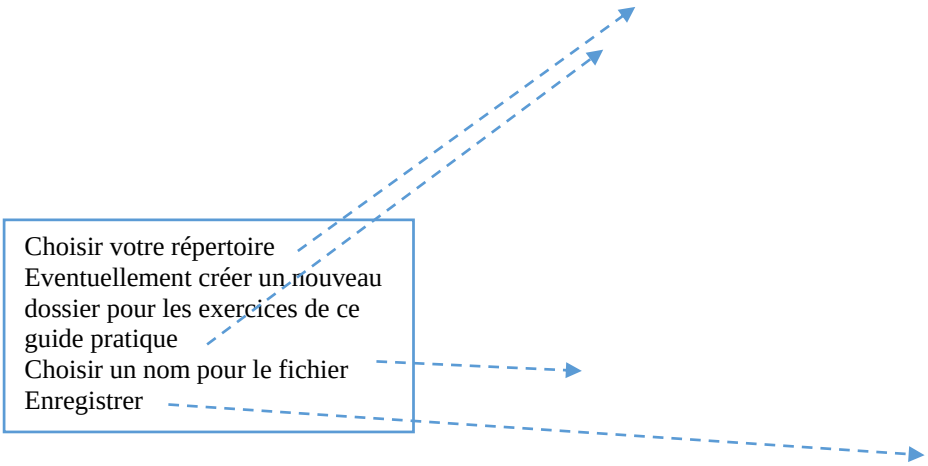
Pour retirer la fenêtre 'Visualiser', cliquer sur la croix en haut à droite.
On peut éditer le fichier pour procéder à des modifications, à l'ajout de nouvelles données ou du nom des variables ou des modalités. Il faut cliquer sur le bouton Editer. Il s'agit exactement du même tableau que si l'on passe par l'éditeur de données comme on a vu plus haut :

Enregistrer un tableau de données

Pour enregistrer ce fichier actif :



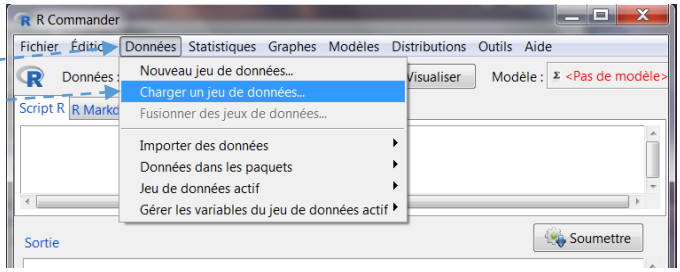
La fenêtre Enregistrer sous apparaît :



2.4 Ouvrir un fichier de données déjà enregistré sous R

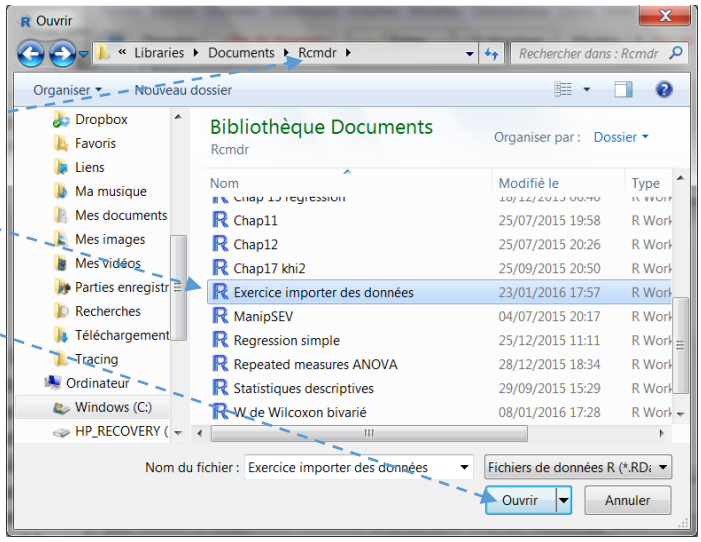
Au cours d'une nouvelle session avec R Commander, ou lorsque l'on souhaite travailler sur un jeu de données déjà enregistré sous R, il faut ouvrir ce jeu, c'est-à-dire « charger le jeu ».

Etape 1
Cliquer sur Données ;
Charger un jeu de données

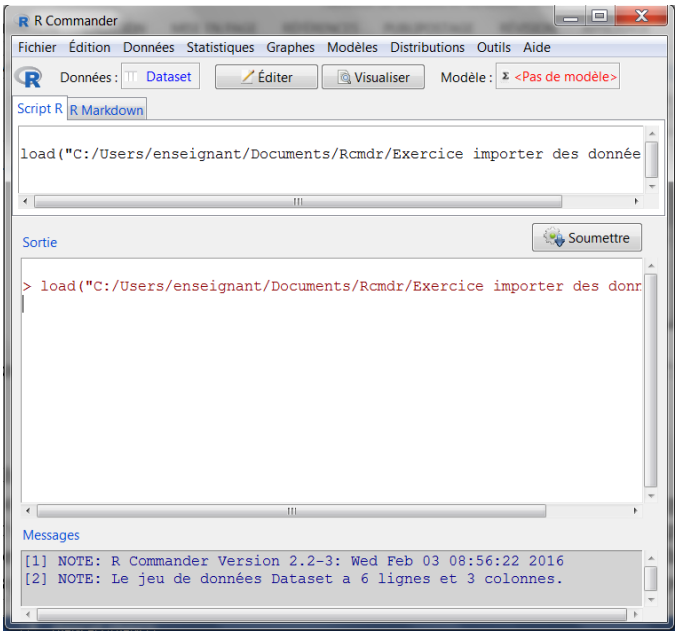


La fenêtre Ouvrir s'ouvre

Aller dans le dossier qui contient le fichier
Choisi le fichier à charger
Ouvrir



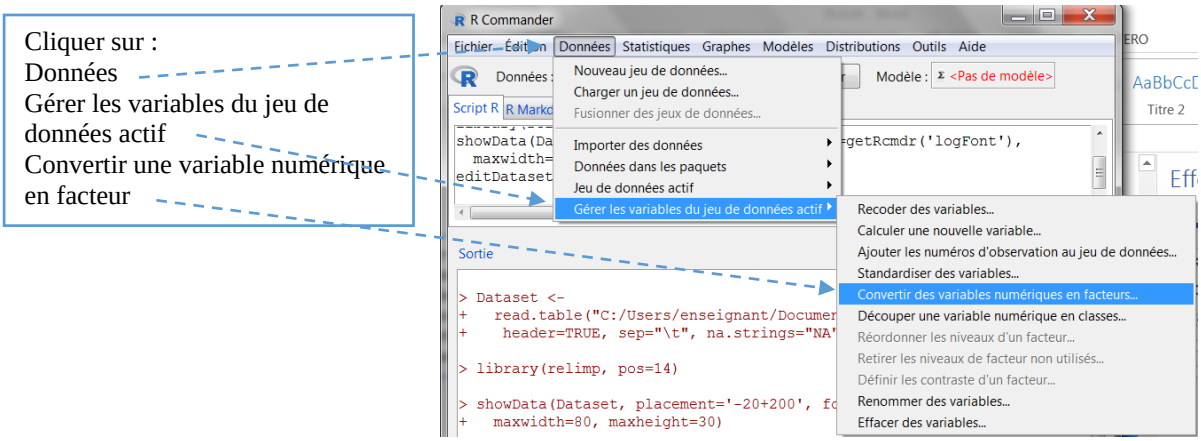
La fenêtre R Commander revient, avec tout en bas le message bleu indiquant le chargement du fichier comme fichier actif les nombres de lignes et de colonnes :



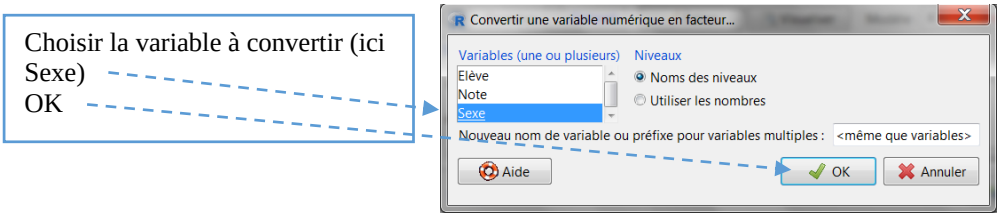
2.5 Coder les variables

Il existe trois sortes de variables dans R Commander : (1) les variables numériques, qui sont des nombres et qui correspondent en fait aux variables dépendantes ou variables à expliquer. Ce sont des valeurs numériques, des scores, des notes, des réponses chiffrées sur une échelle de type Likert, des grandeurs physiques,… ; (2) les facteurs, qui sont des mots et qui correspondent aux variables indépendantes ou variables explicatives. Par souci de simplification, on l'a vu plus haut, on attribue généralement un chiffre aux différentes modalités ou niveaux de ce type de variables, par exemple, 1 pour fille, 2 pour garçon pour la variable sexe. (3) les variables logiques (*true, false*), dont il ne sera pas question dans ce guide.

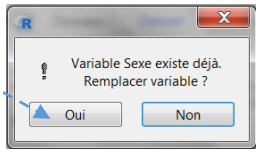
La spécification entre les deux types de variables utilisées dans ce qui suivra, variables numériques et facteurs, se fait automatiquement par R Commander. Autrement dit, en entrant un nombre dans une colonne, R Commander la considère automatiquement comme une variable numérique ; en entrant un mot, R Commander considère la colonne comme un facteur. Si l'on nomme une variable 'Sexe' et que l'on rentre 'fille' ou 'garçon' dans les cellules, R Commander considère la variable 'Sexe' comme un facteur. Mais si l'on rentre 1 pour les filles et 2 pour les garçons, comme dans l'exemple ci-dessus, R Commander considèrera cette même variable 'Sexe' comme une variable numérique. Il faut donc convertir la variable 'Sexe' pour en faire un facteur. La procédure est très simple. Utilisons le même fichier de données que ci-dessus.



La fenêtre 'Convertir une variable numérique en facteur' s'ouvre :

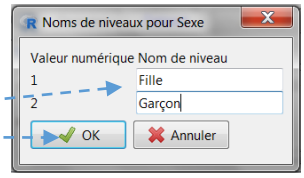


Une fenêtre d'avertissement prévient qu'il existe déjà un nom pour la variable.
Cliquer Oui :

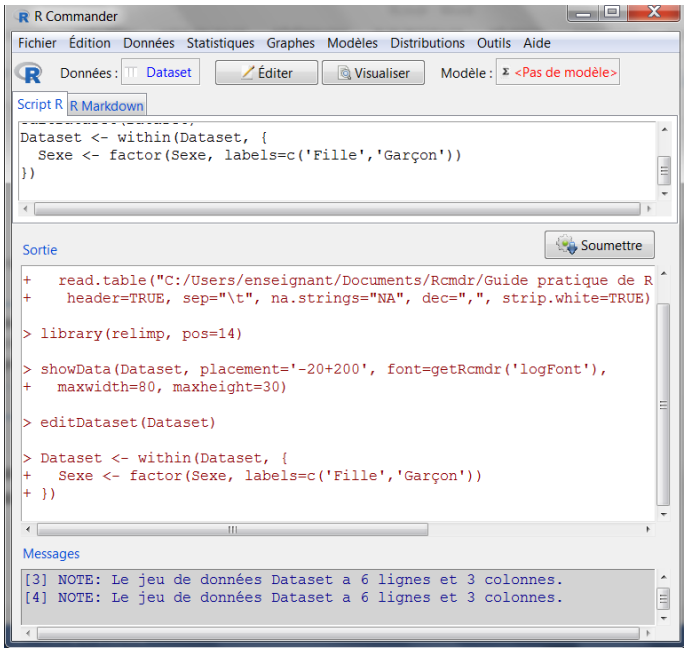


La fenêtre ‘Noms de niveaux pour Sexe’ s’ouvre :

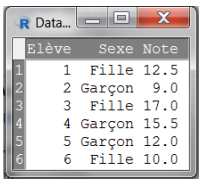
Mettre le nom des modalités pour chacun des niveaux du facteur (ici Fille pour 1 et Garçon pour 2)
OK



La fenêtre R Commander apparaît de nouveau avec tout en bas un nouveau message bleu. On voit également dans la fenêtre des scripts le facteur Sexe et les noms (*labels*) de ses deux modalités, Fille et Garçon.



Cliquons sur Visualiser pour vérifier que les chiffres ont été remplacés par des noms :



Enregistrer le fichier pour conserver les codages effectués.

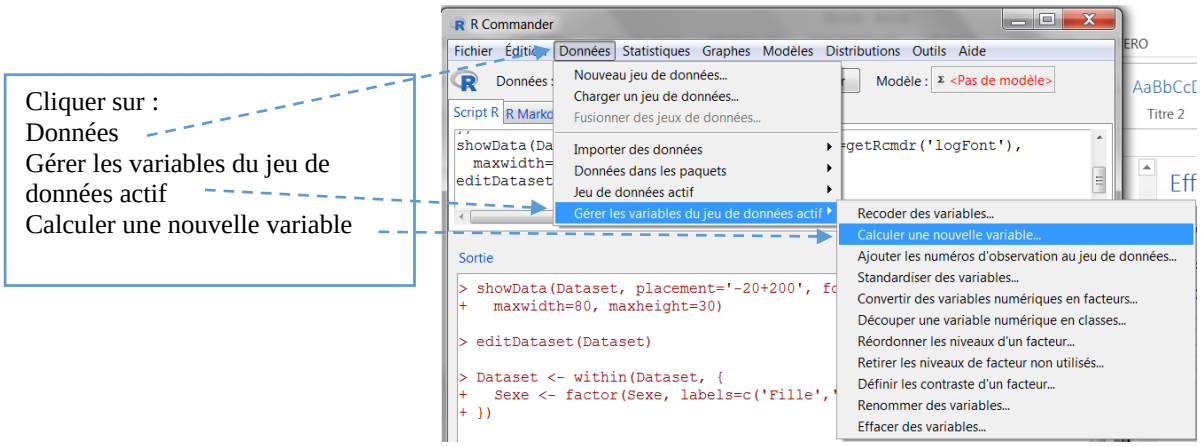
Ce codage des variables est primordial. En effet, R Commander est un logiciel qui intègre dans ses menus et les options offertes, les variables et leur statut. Par exemple, si le jeu de données ne comporte qu’une seule variable quantitative, la commande Corrélation ne peut pas être activée. Ou encore, si le jeu de données ne comporte qu’une variable indépendante (ou facteur) avec 2 modalités, la commande Test pour groupes appariés reste inactive, de même que l’ANOVA. Autrement dit, si cela évitera bien des erreurs, il faut en revanche ne pas se tromper dans le codage des variables ou oublier de coder un facteur.

2.6 Calculer de nouvelles variables

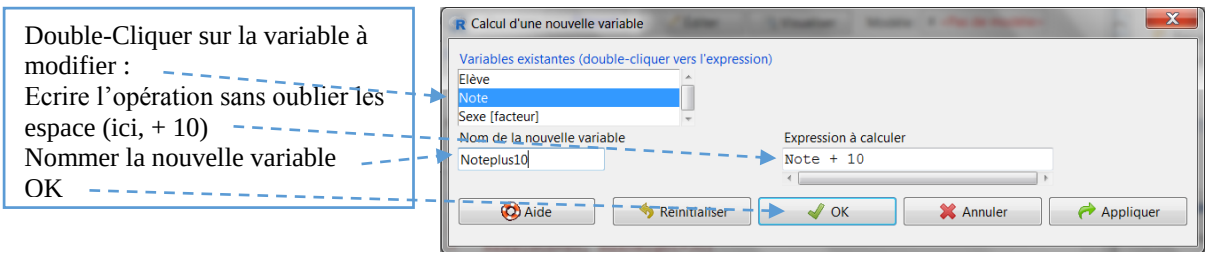
R Commander permet de calculer de nouvelles variables sur un jeu de données. Voici quelques opérations possibles :

Opération	Signe	Exemple
Addition	x + y	Variable 1 + 25
Soustraction	x - y	5 - Variable 1
Multiplication	x * y	Variable 1*10
Division	x / y	Variable 1/63
Elever à la puissance	x ^ y	Variable1^2
Racine carrée	sqrt(x)	sqrt(variable 1)
Transformation logarithmique	log10(x) ou log(x,10)	Log10(variable 1)

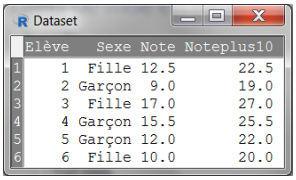
Faisons quelques exercices sur le jeu de données ci-dessus.



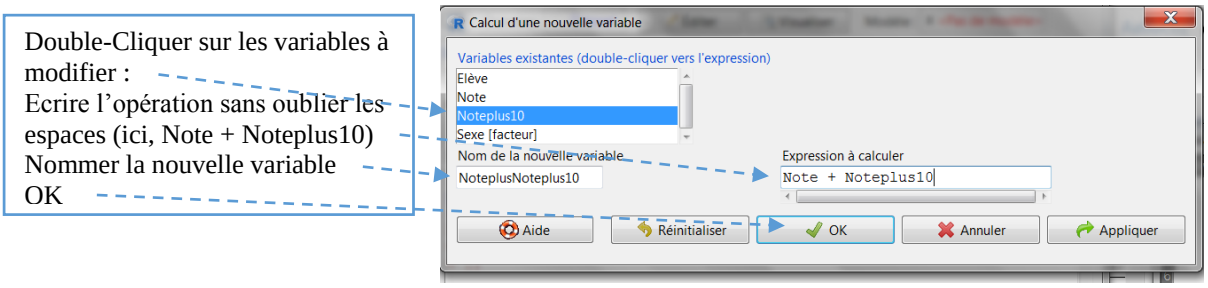
La fenêtre apparaît. Ajoutons 10 aux notes des élèves



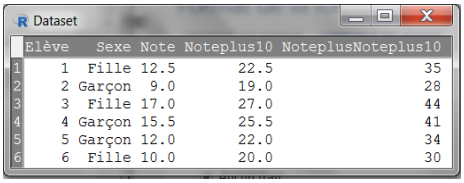
La fenêtre R Commander revient. Visualisons le fichier :



On peut aussi effectuer les opérations avec deux variables. Après avoir Cliqué Données Ą Gérer les variables du jeu actif Ą Calculer de nouvelles variables :



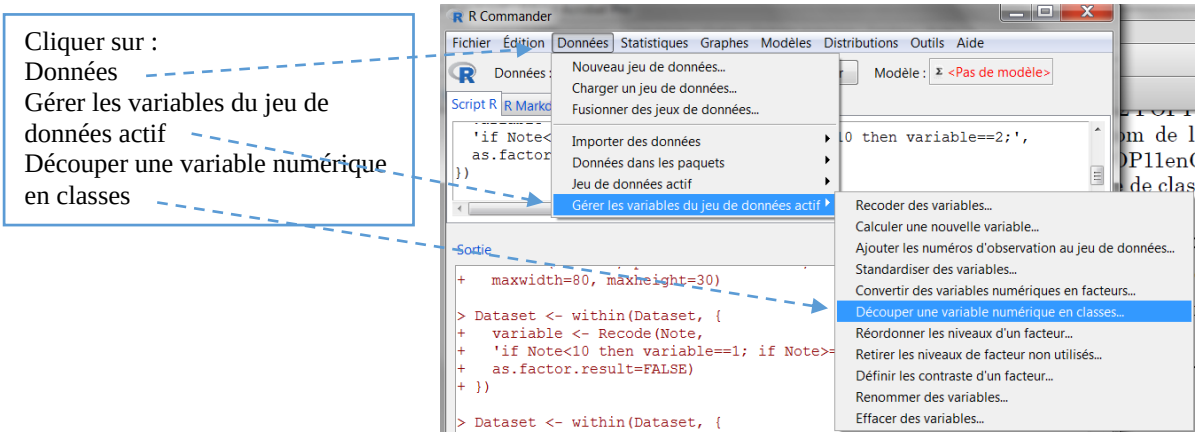
Une fois la fenêtre R Commander revenu, cliquons sur visualiser :



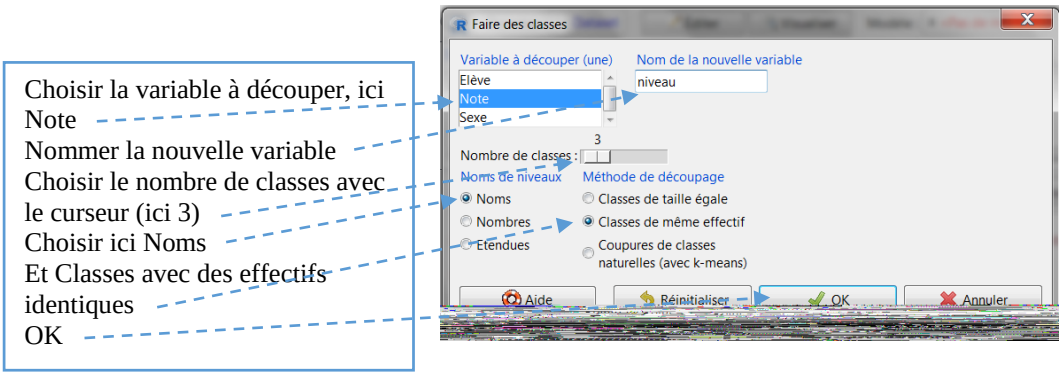
	Elève	Sexe	Note	Noteplus10	NoteplusNoteplus10
1	1	Fille	12.5	22.5	35
2	2	Garçon	9.0	19.0	28
3	3	Fille	17.0	27.0	44
4	4	Garçon	15.5	25.5	41
5	5	Garçon	12.0	22.0	34
6	6	Fille	10.0	20.0	30

2.7 Découper une variable numérique en classes

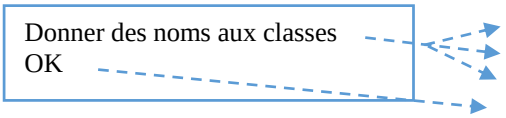
On peut souhaiter découper une variable numérique en classes, par exemple, regrouper les scores à une épreuve en trois classes, scores faibles, moyens et forts, ou encore par déciles… Plus précisément, cette nouvelle variable est un facteur et donc les classes sont autant de modalités. Voici comment procéder :



La fenêtre Faire des classes apparaît :



Comme on a souhaité ici donner des noms aux classes, la fenêtre 'Nom des classes' apparait :



La fenêtre R Commander s'ouvre et cliquons sur Visualiser pour vois les modifications :

A partir de la variable Note, une autre variable, a bien été créée avec 3 classes aux effectifs égaux. Plus précisément, il s'agit d'un facteur à trois modalités dont les effectifs sont égaux.

2.8 Générer des jeux de données partiels

R Commander permet de créer des jeux de données partiels, c'est-à-dire de nouveaux tableaux mais avec une partie des données. On peut ainsi avoir un nouveau jeu de données avec certaines variables seulement, ou encore avec certains individus qui présentent des valeurs particulières sur une variable donnée. Plusieurs symboles ou codes peuvent être utilisés :

Symbole/code	Nom
==	égalité
!=	inégalité
&	et
|	ou
>	Plus grand
<	Plus petit
>=	Plus grand ou égal
<=	Plus petit ou égal
is.na(variable)	Pour inclure les valeurs manquantes d'une variable
!is.na(variable)	Pour exclure les valeurs manquantes d'une variable