

# Penvins Cup 2016 : le cahier des charges

-----  
Pour toute question : denis.poinsot@univ-rennes1.fr  
**Il est interdit** de poser des questions aux membres du jury, merci.  
-----

## Comment votre futur script va être utilisé

Ne placez aucun menu interactif en tête de votre script. En effet nous considérerons comme acquis que l'utilisateur a *déjà* chargé un tableau de données à vérifier dans R (via la fonction `read.table()`), et l'a placé dans un objet de classe `data.frame`. Il va ensuite charger le code de votre script soit "à la souris" (fichier / sourcer du code R) soit en plaçant votre script dans le répertoire de travail et en saisissant la commande : `source("<nom de votre script>.R")`. **Il saisira alors une seule ligne de commande :**

`checkPenvins (<nom de l'objet à vérifier>)`

Toute la suite s'effectuera automatiquement sans autre intervention de sa part. Il lira simplement les messages d'information ou d'erreur produits par le script.

J'attire votre attention sur le fait que l'utilisateur fournit à la fonction `checkPenvins` le **nom** de l'objet de structure `data.frame` à vérifier : votre fonction peut donc utiliser cette information pour accéder aux colonnes de l'objet.

Bien sûr, le jury lira ensuite le code du script lui-même, pour juger de sa qualité et de la clarté des commentaires qu'il contient à l'attention des utilisateurs qui souhaiteraient comprendre son fonctionnement, par exemple dans l'optique de le modifier.

## Liste des tâches que votre fonction `CheckPenvins` doit accomplir

- 1) Vérifier que l'objet est bien de la classe `data.frame`. Si oui, afficher un **message d'information** donnant le nom et la classe de l'objet. Si non, afficher un **message d'erreur** signalant le problème et donnant le nom et la classe de l'objet fautif puis **stopper la fonction**.
- 2) Compter les colonnes de l'objet à examiner. S'il y a 27 ou 31 colonnes, afficher un **message d'information** signalant le nombre de colonnes détecté et la nature supposée du fichier (données de quadrats ou données biométriques). Un nombre différent de colonnes doit provoquer un **message d'erreur** signalant que le nombre de colonnes attendu était 27 ou 31, donner le nombre de colonnes détecté ici puis **stopper la fonction**.
- 3) Vérifier que les intitulés de colonnes correspondent à l'attendu (selon le type de fichier supposé d'après sa taille). Si oui, afficher un **message d'information** indiquant que les intitulés de colonnes sont corrects. Si non, afficher un **message d'erreur** signalant les noms des colonnes dont les intitulés posent problème en citant à chaque fois le nom correct attendu, puis **stopper la fonction**.
- 4) Vérifier que les colonnes sont bien de la classe attendue (factor, ou integer ou numeric, selon le cas). Si oui, afficher un **message d'information** donnant le nom et la classe de chaque colonne. Si non, afficher un **message d'erreur** signalant le nom et la classe des colonnes qui posent problème en citant à chaque fois la classe détectée et la classe attendue, puis **stopper la fonction**.

Si la procédure n'a pas été stoppée à une des étapes 1 à 4 :

5) pour toutes les colonnes de classe "factor", vérifier que la ou les modalités du facteur correspondent à l'attendu. Si oui, afficher un **message d'information** indiquant la/les modalité(s) détectée(s). En cas d'anomalie, afficher un **message d'erreur** donnant la colonne et les numéros et contenus des lignes qui

posent problème ainsi que le contenu qui était attendu. Voir ci-dessous le tableau d'attendu pour les colonnes de classe factor :

colonne	attendu
group	même nom pour toutes les lignes
cible	même cible pour toutes les lignes : "Litlit" ou "Thalap"
resp	même nom pour toutes les lignes
month	même mois pour toutes les lignes : « sept »
mode	même mode pour toutes les lignes : "a" ou "b"
capture	deux types selon les lignes : "quad", "vue"

6) pour toutes les colonnes de chiffres (de classe "integer" ou "numeric"), vérifier que le contenu ne sort pas de l'intervalle mini--maxi attendu. Si tout va bien, afficher un **message d'information** indiquant les valeurs mini et maxi détectées. En cas d'anomalie, afficher un **message d'erreur** donnant les numéros et le contenu des lignes qui posent problème ainsi que l'intervalle mini--maxi qui était attendu. voir ci-dessous le tableau d'attendu pour les colonnes numériques

colonne	attendu
year	2016
coef	108 (=lundi 19 sept) ou 111 (=mardi 20 sept)
dist	même nombre entier pour toutes les lignes, compris entre 10 et 100 (m)
station	nombres entiers divers, entre 0 et 100 (m)
alt	nombres décimaux entre 0 et 5 (m)
surf	nombres décimaux entre 0,16 et 0,48 (m <sup>2</sup> ) [1]
p.roc etc.	nombres entiers entre 0 et 100 (%)
Litlit etc.	nombres entiers entre 0 et 200 (individus) [1]
haut etc.	nombres décimaux entre 1 et respectivement 35, 35 et 20 (mm)

[1] NA systématiquement si capture = "vue"

7) vérifier que les ratios **larg/haut** et **peri/larg** du fichier biométrique sont situés respectivement dans l'intervalle [0,5—1,5] et [0,1—0,75]. Si tout va bien, afficher les ratios mini et maxi détectés et les normes attendues. Dans le cas contraire, afficher autant de **messages d'avertissement** que nécessaire donnant les numéros de ligne, les ratios suspects et les normes attendues.

Bien entendu les vérifications concernant les mesures biométriques ne doivent être effectuées que si un fichier contenant les colonnes en question a été détecté.

La fin de la procédure **CheckPenvins** doit afficher un message de conclusion à votre goût (à supposer que tout n'ait pas été stoppé bien en amont...) signalant **soit** qu'aucune anomalie n'a été détectée **soit** au contraire que le fichier à vérifier ne satisfait pas le cahier des charges.

Il va sans dire (mais ça va mieux en le disant) qu'avant de nous le remettre, vous devez **tester** votre script face à un fichier contenant volontairement des fautes. Méfiez-vous particulièrement des modifications anodines de dernière minute, aussi minuscules soient elles : si vous en faites une, **testez à nouveau** votre script.

-- fin du cahier des charges ---