

# Statistique descriptive avec R

## TP 2

Téléchargez les données au format Excel disponibles à l'adresse suivante :

<https://www.insee.fr/fr/statistiques/4268033>. Dans l'onglet *Figure 2*, copiez le tableau (à partir de la cellule A4) dans un fichier texte et enregistrez-le sous "**inflation.csv**".

1. Identifiez le séparateur de données et le séparateur décimal. Comment spécifier ce dernier dans R ?
2. Importez les données avec la fonction **read.csv**.
3. Vous pourriez rencontrer des problèmes avec les accents et autres caractères spéciaux, dus à un encodage de fichier différent de celui attendu par la fonction d'importation. L'encodage est généralement "**UTF-8**". Précisez-le.
4. Ajoutez un argument **colClasses** pour préciser le type des colonnes dès l'importation. Vous pouvez tester les erreurs générées si vous spécifiez mal le séparateur décimal.
5. Importez à nouveau les données mais avec la fonction **fread** du paquet **data.table**, en précisant l'encodage.
6. Passez vos données en **data.frame** et revenez en **data.table** sans copie.
7. Transformez la première colonne au format **Date**, en fixant les dates au 1er janvier..
8. Vos données sont de la plus récente à la plus ancienne. Triez-les de la plus ancienne à la plus récente à l'aide de la fonction **order**.
9. Avec **matplot**, en utilisant comme premier argument le vecteur des dates en abscisse, et en second argument la data.table sans la colonne des dates, affichez tous les indices d'inflation.
10. Ajoutez la légende en haut à gauche.

11. Les indices sont en base 100 pour l'année 2015. Mettez-les en base 100 par rapport à la première date des données. Rappel : avec des **data.frame** on peut accéder à la colonne **i** avec **inflation[, i]** mais en **data.table**, par défaut, il interprète **i** comme un nom de colonne sans l'évaluer, comportement similaire à **inflation\$i**. Pour l'évaluer, il faut soit utiliser la notation des listes : **inflation[[i]]**, soit forcer l'évaluation de **i** avec **inflation[, i, with =FALSE]**.

**Remarque.** Pour des données de petite taille, le paquet **data.table** est superflu en termes d'efficacité; c'est davantage un choix pour sa syntaxe flexible. Cependant, pour des données importantes ou des manipulations complexes, il est plus efficace et rapide.