

STID 1 - Programmation Statistique

TP5

Visualisation et Agrégation

Anthony SARDELLITTI

2023-01-01

Contents

1	Ressources documentaires	1
2	Exercices	1
2.1	Exercice 1 : Importer les données	2
2.2	Exercice 2 : Statistiques descriptives	2
2.3	Exercice 3 : Visualisation	3
2.4	Exercice 4 : Agrégation	3

1 Ressources documentaires

Pour réaliser ce TP, vous aurez besoin des ressources suivantes :

- [Fonctions de tests et comparaisons](#)
- [Indexation](#)
- [Filtres et sélection](#)
- [Les fonctions de tests et opérateurs de comparaison](#)
- [Trier](#)
- [Agréger](#)

2 Exercices

Pour rendre ce TP, vous pouvez envoyer un mail à anthony.sardellitti@hotmail.fr OU compléter votre repository GitHub.

Pensez à commenter votre code.

Ce TP reprend les notions évoquées dans les chapitres 2 à 7. On utilise le fichier `NBA.csv` qui décrit les tirs effectués au cours de la saison 2014-2015 de NBA. Le fichier est issu du site [Kaggle](#). Il a été adapté et modifié pour ce cours. Les données sont accessibles en [cliquant ici](#).

Voici la présentation du jeu de données :



Figure 1: NBA

- `GAME_ID` : ID du match
- `LOCATION` : Lieu du match (*Home / Away*)
- `GAME_RESULT` : Résultat du match (*Won / Lost*)
- `PERIOD` : Numéro de quart-temps et prolongations éventuelles
- `SHOT_CLOCK` : Durée de la possession de l'équipe au moment du tir
- `DRIBBLES` : Nombre de dribbles avant le tir
- `TOUCH_TIME` : Durée de la possession du joueur avant le tir
- `SHOT_DIST` : Distance de tir en *foot* ($1 \text{ foot} = 0,30 \text{ mètre}$)
- `PTS_TYPE` : Tentative à 2 ou 3 points (les lancers francs à 1 point ne sont pas répertoriés)
- `SHOT_RESULT` : Résultat du tir (*made / missed*)
- `CLOSE_DEF_DIST` : Distance entre le tireur et le défenseur le plus proche en *foot* ($1 \text{ foot} = 0,30 \text{ mètre}$)
- `SHOOTER` : Nom du tireur

2.1 Exercice 1 : Importer les données

- Importer le jeu de données `NBA.csv` avec la fonction `read.csv()`.
- Combien de lignes, colonnes sont présentes dans ce dataset (utilisez les fonctions adaptées) ?
- Afficher un résumé des données avec la fonction adaptée.
- On souhaite analyser les variables `PERIOD` et `PTS_TYPE`. Modifier le type de ces variables pour les transformer en type factor. Vérifiez si cela a fonctionné avec un résumé des données.

2.2 Exercice 2 : Statistiques descriptives

- Calculer les **déciles** de la variable présentant la distance entre le tireur et le défenseur le plus proche (`CLOSE_DEF_DIST`).

- b. Combien de matchs se sont déroulés lors de cette saison de basket ?
- c. Combien y-a-t-il de **SHOOTER** différents ?
- d. Créer une variable **SHOT_DIST_METRE** qui affiche la variable **SHOT_DIST** convertit en mètre ($1\ foot = 0,30\ mètre$).

2.3 Exercice 3 : Visualisation

- a. Afficher dans un diagramme en barre la répartition des tirs par période.
- b. Afficher dans un diagramme circulaire la proportion de tirs réussis et ratés en pourcentage avec les étiquettes.
- c. Afficher l'histogramme de la distribution de la distance de tirs
- d. Afficher un diagramme en barre représentant la répartition en pourcentage de tirs à 2 et 3 points pour chaque période.
- e. Construire une boîte à moustache de la distribution des la distances de tirs.
- f. Construire les boîtes à moustache de la distribution de la distance de tirs en fonction des périodes.

2.4 Exercice 4 : Agrégation

Avec la fonction `aggregate` :

- a. On souhaite construire un data frame avec les 10 **SHOOTER** qui tentent des tirs le plus loin.
- b. On souhaite construire un data frame avec les 5 **SHOOTER** qui ont inscrit le plus de points.
- c. On souhaite construire un data frame avec le nombre de points marqués pour chaque période