
Sommaire

I - Contexte et description du DataFrame

II - Question et méthodes utilisées

III - Réponse à la question et conclusion

I-

Nous avons travaillé sur un DataFrame créé par Hugo Mathien et regroupant plus de 25 000 Matches de Football de 2008 à 2016 récupéré sur le site Kaggle regroupant des milliers de jeux de données. Le DataFrame est un fichier SQLite que nous avons parsé afin d'obtenir des fichiers .csv pour RStudio. Il contient des tables comme Matches, Joueurs, Equipe, Ligues etc. Nous avons choisi de travailler sur la table la plus volumineuse, la table Matches et notamment sur le nombre de buts inscrits.

II -

De nos jours le foot prend une part importante dans les médias, chaque semaine on peut remarquer en première page que des grands joueurs ont battu leur propre record de buts sur une saison. On a donc choisi de vérifier si le nombre moyen de buts a augmenté depuis 2008.

Nous avons décidé d'utiliser des histogrammes afin de voir s'il y avait une différence significative au niveau des buts marqués à domicile et à l'extérieur. On remarque que les buts à l'extérieur sont principalement compris entre 0 et 1. En revanche pour le nombre de buts inscrits à domicile on peut remarquer que trois valeurs se détachent (1,2,3) ce qui montre bien que le nombre de buts marqués par matchs est plus important à domicile.

```
library(dplyr);
```

```
##  
## Attaching package: 'dplyr'
```

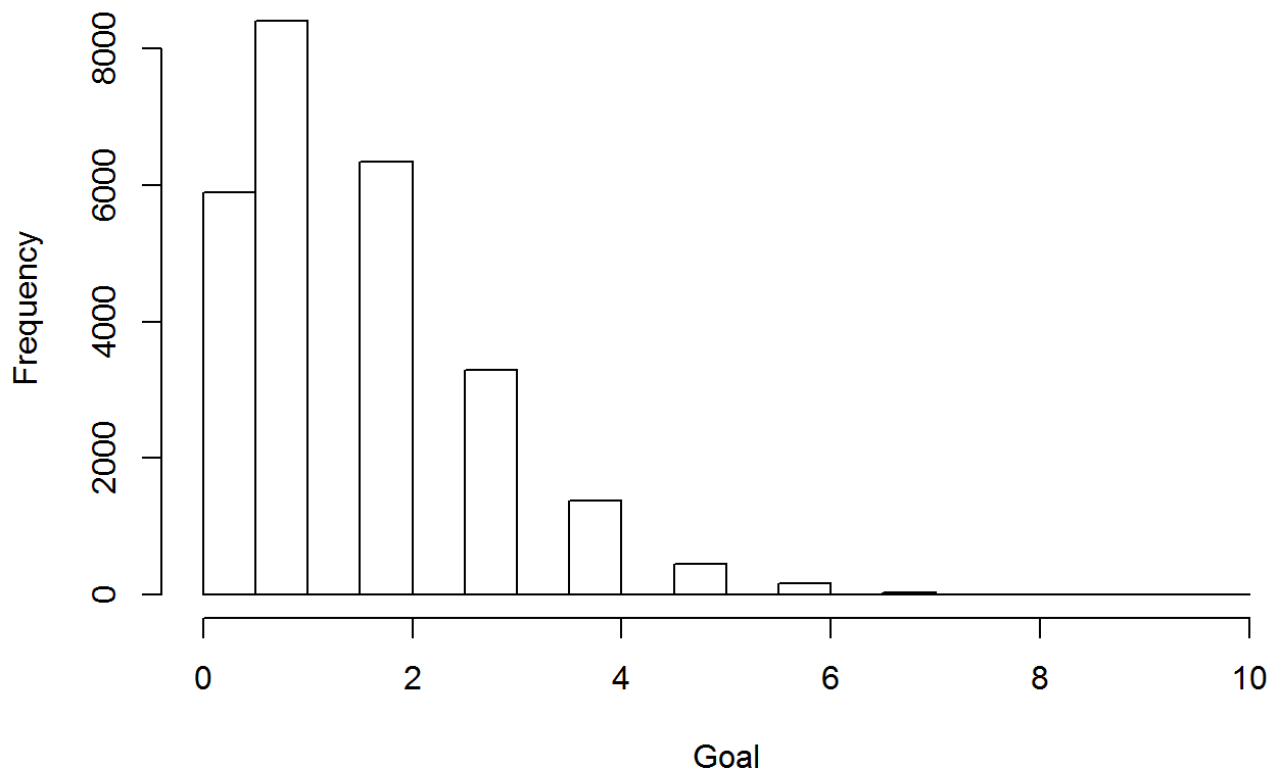
```
## The following objects are masked from 'package:stats':  
##  
##     filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

```
df <- read.csv("C:/Users/Asus G75/Desktop/RStudio/Pepite/Match.csv", header=TRUE, sep=";");
```

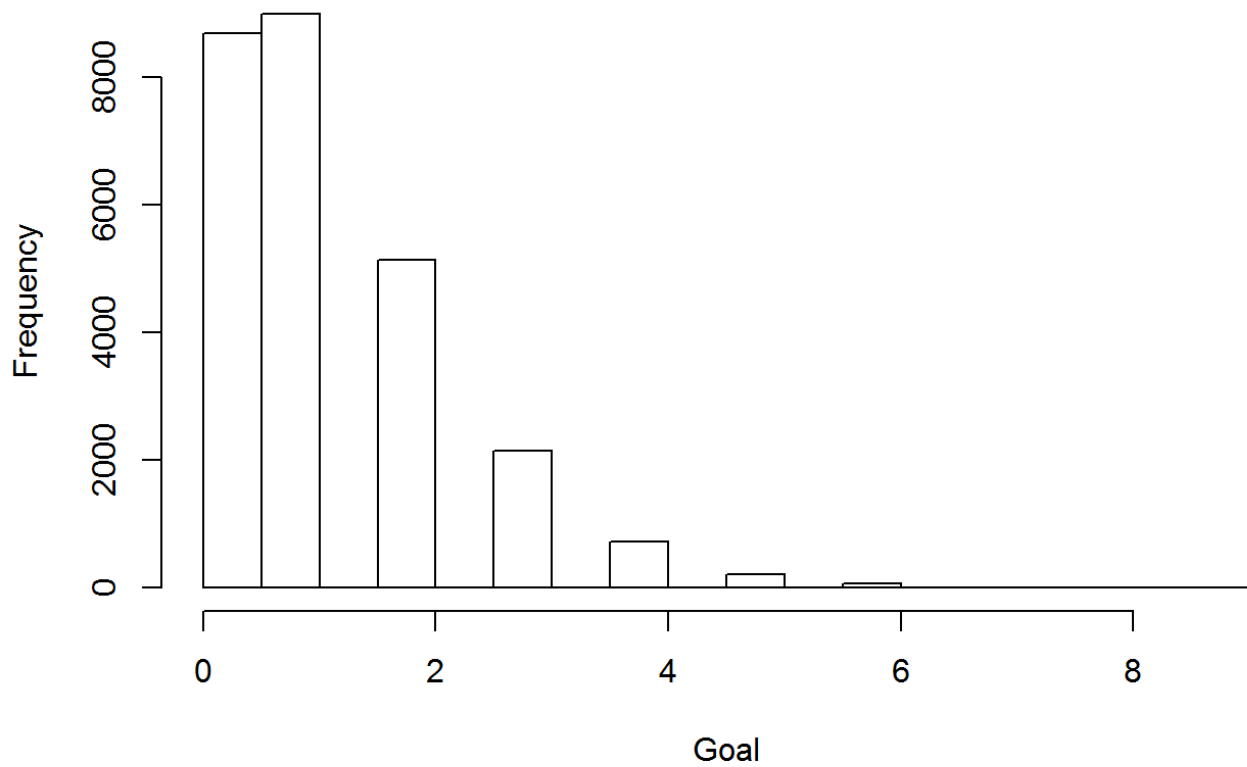
```
hist(df$home_team_goal, xlab="Goal", main="Home Team Goal")
```

Home Team Goal



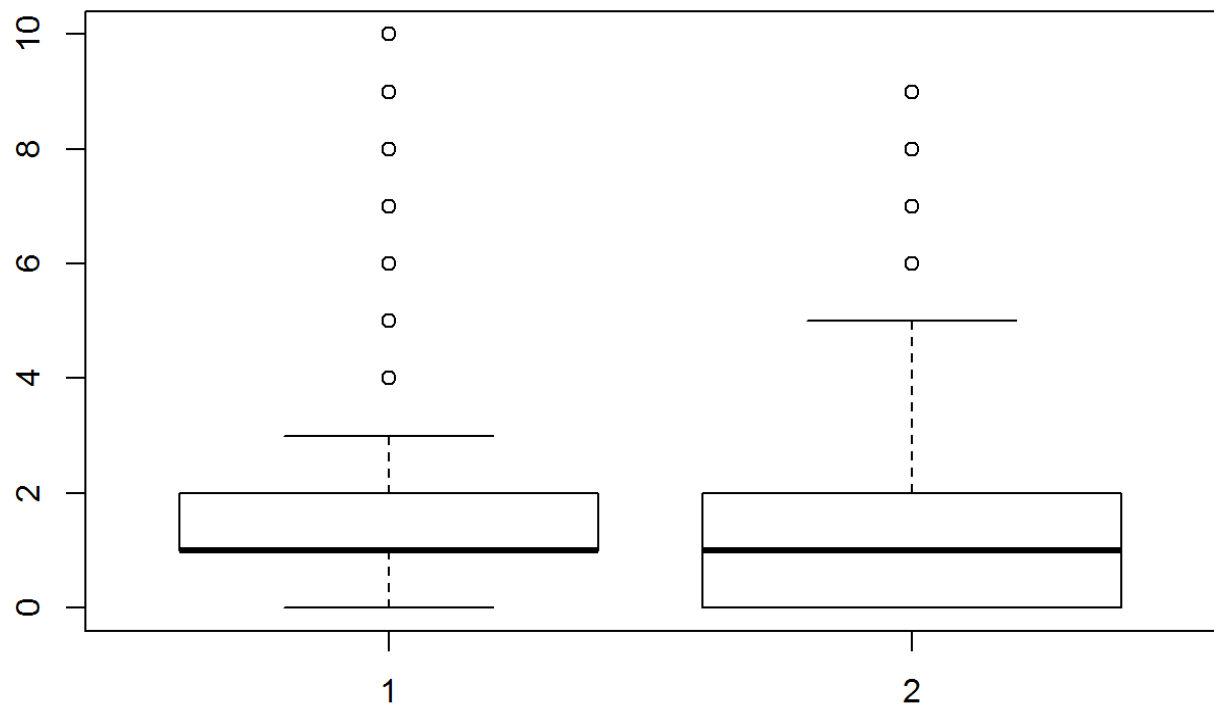
```
hist(df$away_team_goal, xlab="Goal", main="Away Team Goal")
```

Away Team Goal



On a cherché à comprendre plus profondément comment l'équipe à domicile marquait plus souvent une ou plusieurs fois alors que l'équipe qui joue à l'extérieur ne marque souvent pas. Pour approfondir la question, nous avons utilisé un boxplot afin de se rendre compte de la répartition des buts. On remarque que l'équipe qui joue à domicile a une répartition bien plus régulière de ses résultats.

```
boxplot(df$home_team_goal, df$away_team_goal);
```



On a alors cherché à comprendre quel impact avait le fait d'être à domicile et on a donc utilisé Summary afin d'avoir la moyenne des buts à domicile et à l'extérieur. On remarque bien ici que le nombre de buts moyens pour l'équipe à domicile est bien supérieur au nombre de but que marque l'équipe qui joue à l'extérieur.

```
summary(df$home_team_goal);
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   1.000   1.000   1.545   2.000   10.000
```

```
summary(df$away_team_goal);
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   0.000   1.000   1.161   2.000    9.000
```

Concernant la question de base, on a tout d'abord cherché le nombre de but moyen total d'un match, le but étant de se faire une idée de sa valeur avant de se lancer dans une analyse saison par saison.

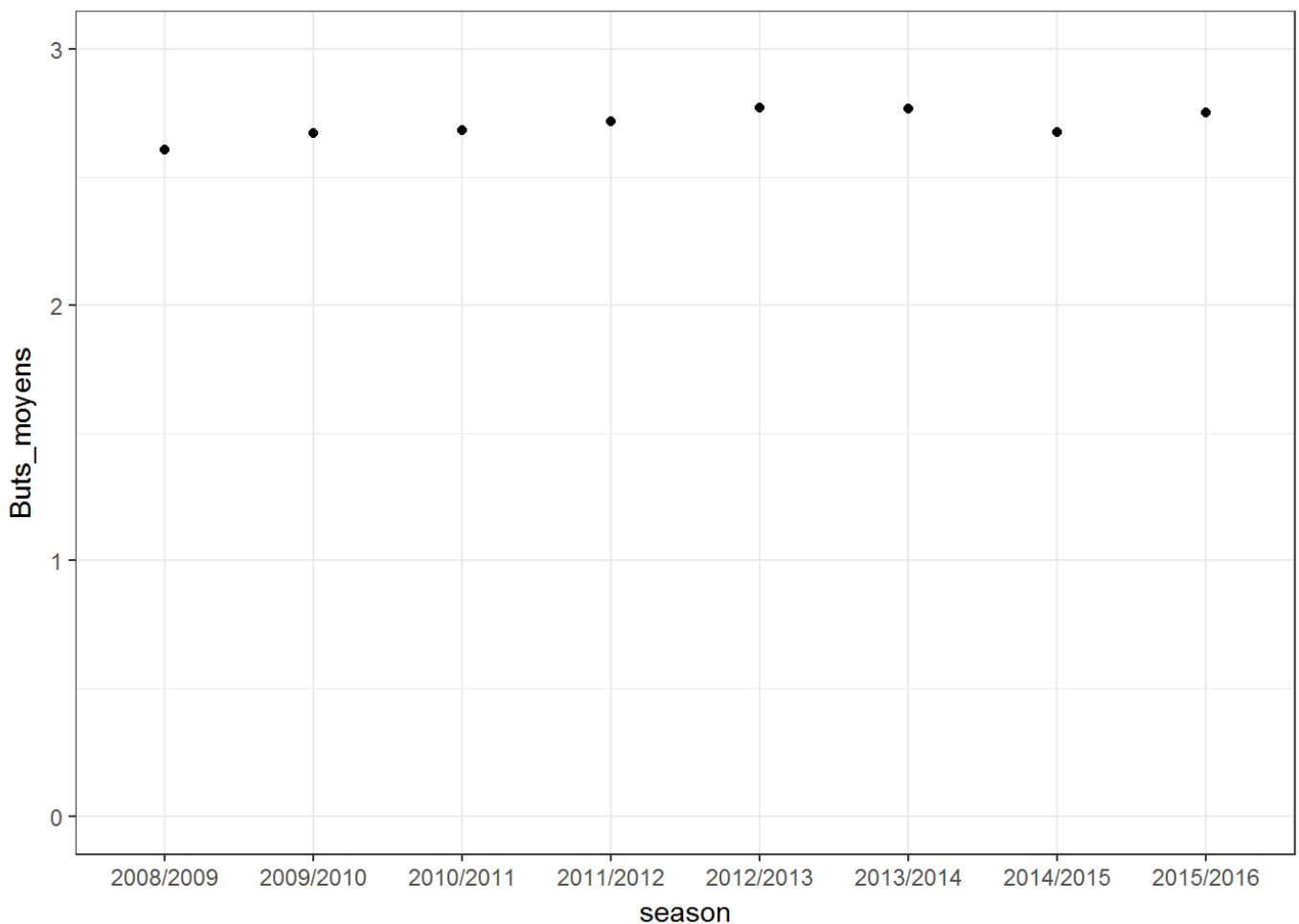
```
totalbutHome <- mean(df$home_team_goal);
totalbutAway <- mean(df$away_team_goal);
MoyButTotal <- (totalbutHome + totalbutAway)
head(MoyButTotal)
```

```
## [1] 2.705531
```

On a finalement cherché à savoir si le nombre de buts augmentait en fonction des saisons et on a comparé le nombre de buts moyens par saison allant de 2008 à 2016. Globalement si on répond à la question de base étant : “Le nombre de but a t-il augmenté de 2008 à 2016”, la réponse est oui. Grâce a un histogramme et à la librairie dplyr, nous avons pu récupérer la moyenne de buts marqués par saison et constater une augmentation, certe légère d’environ 0.3 buts moyen supplémentaire par saison. D’une moyenne tendant vers 2.5 en 2008, on se rapproche aujourd’hui de plus en plus des 3 buts par match.

```
library(ggplot2);

df %>%
  group_by(season) %>%
  summarise(Buts_moyens = mean(home_team_goal+away_team_goal)) %>%
  ggplot(aes(x=season, y=Buts_moyens)) +
    geom_point() + ylim(0,3) + theme_bw();
```



III -

On remarque d’après nos analyses que le nombre de buts moyens par saisons à augmenté. Cependant le nombre moyens de buts n’a pas fait que d’augmenter depuis 2008 il n’est donc pas croissant. Pour terminer cette analyse au vu de l’évolution du nombre de buts depuis 2008, nous pouvons nous poser la question :

Loading [Contrib]/a11y/accessibility-menu.js rd’hui ?