

Travail de session

Sadou Ba

Introduction

Depuis plus d'un siècle, il y a un engouement international pour les événements sportifs qui sont diffusés à travers divers canaux tel que la radio, la télévision et maintenant internet (Muller et *al*, 2023). Dû à l'attrait encore plus croissant des événements sportifs, les gains monétaires sont en constante croissance. Ce qui pousse plusieurs personnes fortunées à acheter des équipes sportives pour profiter de la manne actuelle (Gomez, 2022). Toutefois, il y a un autre intérêt qui pousse certains grands fortunés de ce monde de s'investir dans le milieu sportif et c'est de pouvoir avoir une plus grande influence. Ainsi, dans une optique d'acquérir plus d'influence par une opération de soft power (Verschuuren, 2013). C'est notamment le cas de riche pétromonarchie du golf tel que le Qatar qui tente d'asseoir son statut de puissance stable dans un Moyen-Orient instable en espérant de cette manière préserver l'intégrité de son territoire contre les velléités de leurs adversaires (Verschuuren, 2013). Un des sports qui soulèvent le plus les passions et qui peut toucher un public à l'échelle mondiale est le football (Power & *al*, 2020). Ce que nous observons est que de nombreux supporters soutiennent le propriétaire du club ce qui en fait un nombre considérable de relai pour les idées promut par les propriétaires en question. Ces constats nous ont poussés à nous poser comme question s'il serait possible de voir l'influence des propriétaires fortunés et étrangers de club de sport sur les orientations du club. Ayant constaté une hausse des joueurs internationaux dans le football depuis les années 90, nous aimerions voir si cela pourrait être lié à la plus grande part de propriétaire d'équipes étrangers. Ainsi, nous regarderons s'il y a une corrélation entre le fait que le propriétaire d'un club soit étranger et le pourcentage de joueurs internationaux dans l'équipe. Pour ce faire, nous observerons les 20 meilleures équipes de la ligue anglaise de football la premier League. Dans ce travail, il sera question de la façon dont nous avons obtenu nos données, de comment nous avons traités celles-ci, de nos résultats obtenus et des conclusions que nous pouvons en retirer.

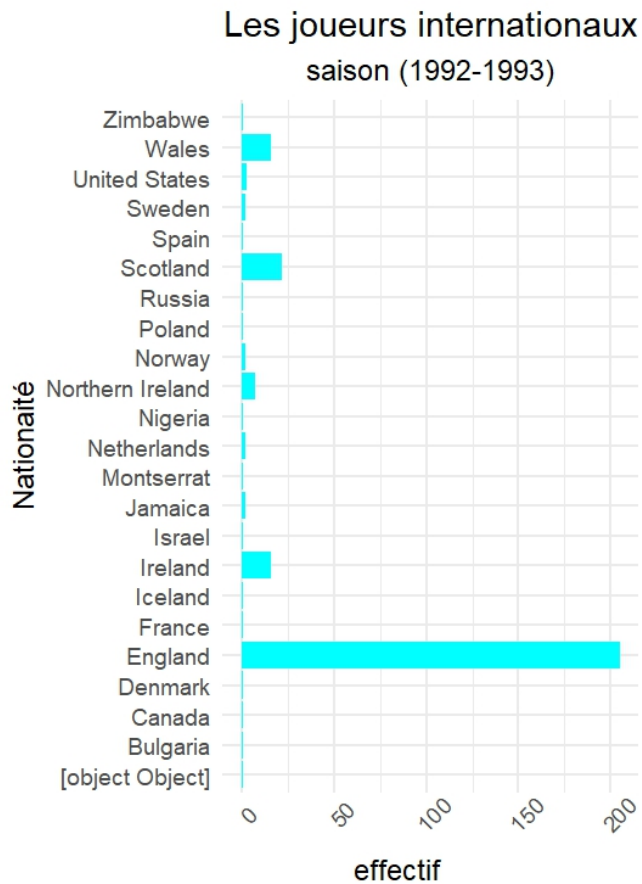
Données et méthodes

Pour effectuer nos analyses, les données que nous avons utilisées nous proviennent du site web de la Premier League. Ils ont une section dédiée aux statistiques sur les joueurs de l'ensemble des équipes de la Premier League. Ces statistiques comportent des données sur le nom du joueur, de son club, et de sa nationalité et du nombre de buts effectués durant la saison. Ce qui nous intéressent dans ce cas-ci, sont les données sur la nationalité, l'équipe et la saison de jeu. Ainsi, nous voulons obtenir les données pour 2 saisons particulières soit celle de 1992-1993 et celle de 2023-2024. Ces deux saisons nous serviront pour constater l'évolution des joueurs internationaux au sein de la Premier League. Pour avoir accès à ces données, il faut procéder à du moissonnage du web (Web-scraping). Pour ce faire nous avons utilisés une série de paquetages inclus dans le logiciel de programmation R. Le paquetage qui nous est le plus nécessaire, est Rselenium. Rselenium est un paquetage qui a pour but de permettre de l'automatisation de l'utilisation d'un navigateur web à distance. Dans cette automatisation, il est question de pouvoir ouvrir une page web, de cliquer sur des éléments ainsi que les sélectionner. Un autre paquetage qui serait utile est le rvest, qui permet d'extraire le code html des pages web pour importer les données du site web dans R sous divers formats, dans notre cas, il s'agira d'un format table qui aura en sortie un 2 dataframe pour les 2 saisons observés. Les étapes pour procéder à notre collecte de données consiste à en premier définir un webdriver pour Rselenium soit quel navigateur web nous utiliserons pour ouvrir notre page web. Par la suite, il faudra indiquer les éléments css présent sur le site web pour pouvoir cliquer dessus, notamment les publicités, les cookies, pour sélectionner la saison et changer de pages. Nous avons aussi mis en place des boucles itératives pour sortir chacune des pages de chacune des saisons pour les inclure dans des listes. Avec l'usage de rvest nous indiquerons finalement la table qu'il faut extraire.

Une fois nos données importées sous forme de dataframe. Nous n'avons plus qu'à procéder à nos analyses. En ce faisant, nous allons d'abord effectuer des graphiques. Les deux premiers graphiques porteront sur la le nombre de joueurs par nationalité au sein de la Premier League pour les 2 saisons. Le troisième et dernier graphique quant à lui portera sur la proportion de joueurs internationaux par équipe dans la saison 2023-2024. Et à la suite de ses visualisations, l'analyse statistique qui nous intéresse le plus sur la corrélation entre la proportion de joueurs internationaux et le fait que le propriétaire soit étranger ou local. Pour ce faire, nous ferons une régression linéaire avec aucun contrôle. Nos deux variables sont qualitatives, un étant nominal soit la nationalité qui est la variable dépendante et la variable indépendante qui est une variable binaire soit local ou étranger. À noter que les données sur les propriétaires de club sont tirées de Wikipédia et ont été mis en format xlsx sur Excel, car le processus aurait été trop long sur R,(Wikipédia, 2024).

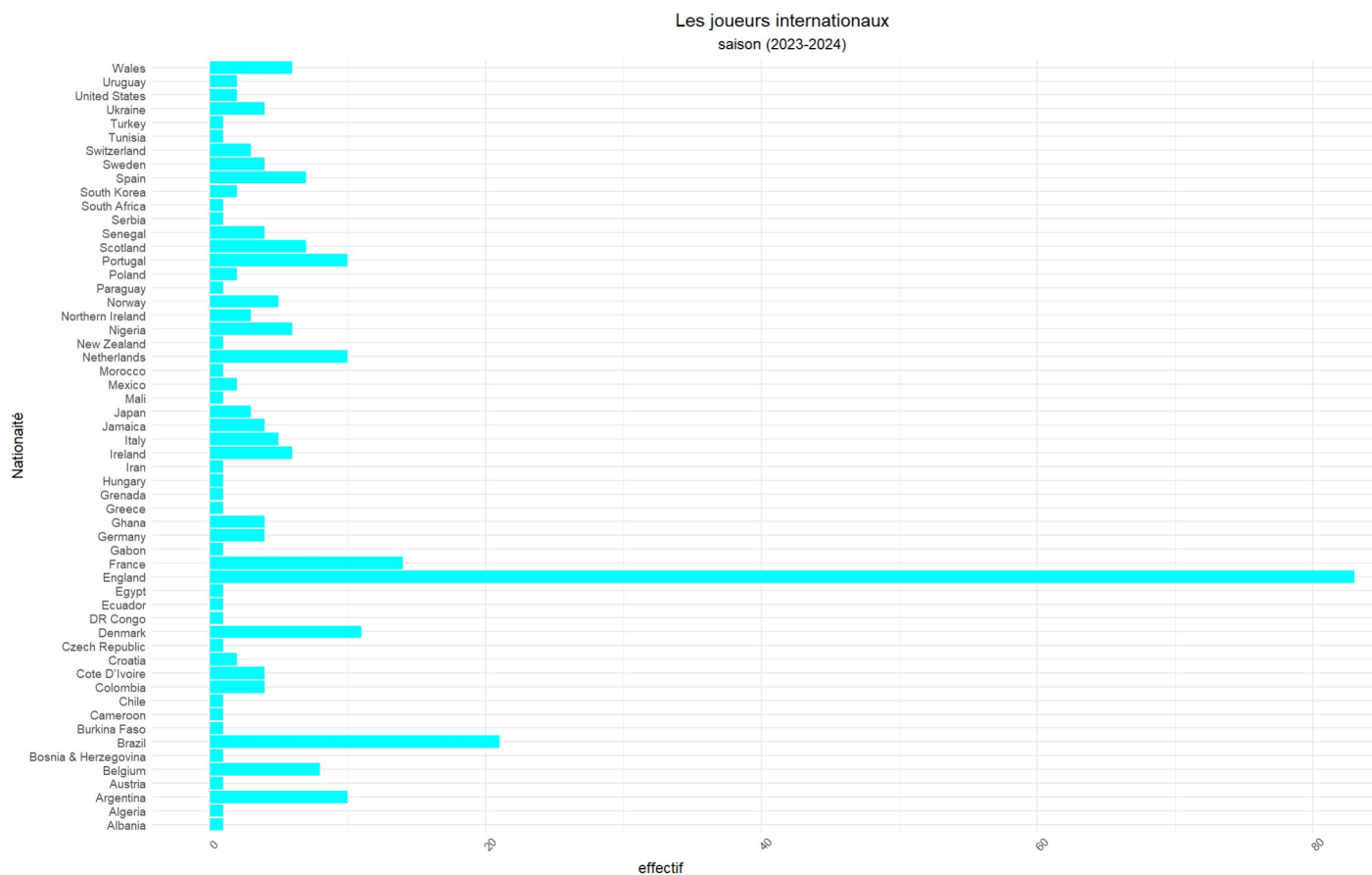
Résultats

Figure 1



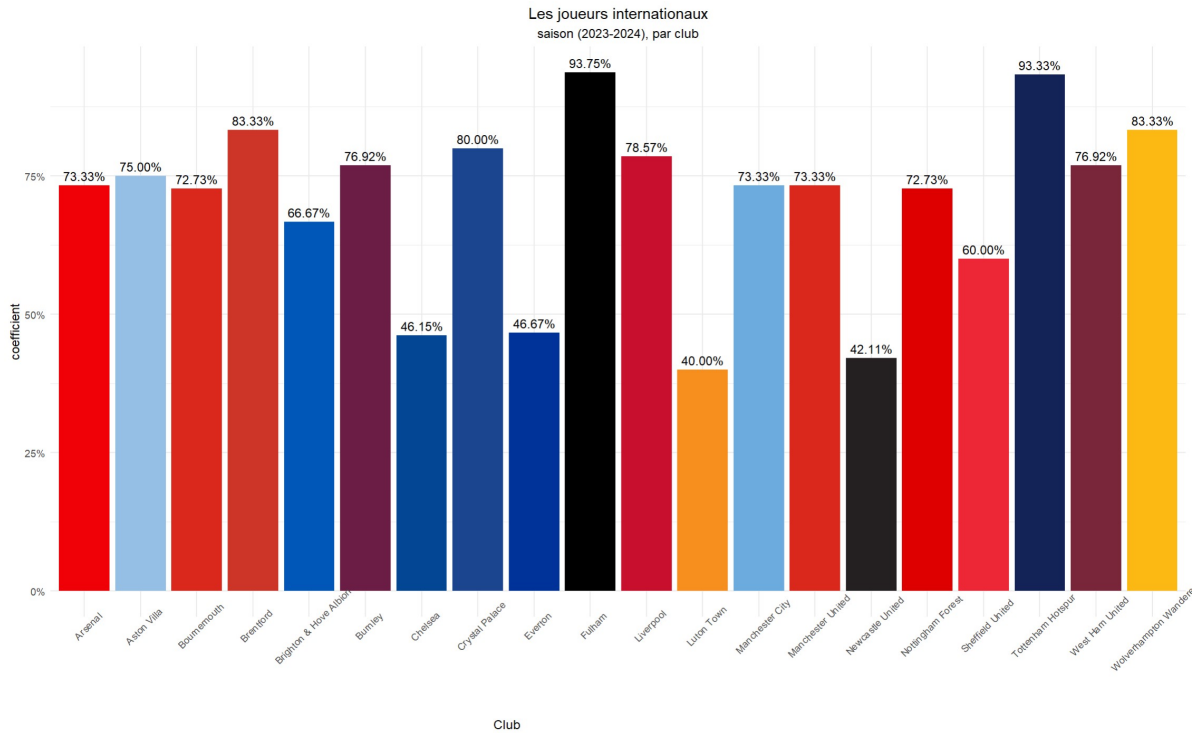
Dans la figure 1, nous pouvons voir que la nationalité la plus représentée est celle de l'Angleterre avec plus de 200 joueurs anglais et que même après cela, les autres nationalités qui suivent font toutes partie de la Grande-Bretagne soit le pays de Galles, l'Écosse ou encore l'Irlande du Nord. Ainsi, nous observons que la majorité des joueurs avaient des origines anglo-saxonnes.

Figure 2



Dans la figure 2, nous observons dans ce cas-ci, une tout autre dynamique, pour à peu près le même effectif de joueurs, les joueurs anglais sont toujours les plus nombreux, mais leur proportion a nettement diminué étant maintenant au nombre d'un peu plus de 80 joueurs. Le deuxième pays est maintenant le Brésil et nous observons une plus grande diversité de nationalité et une plus grande répartition des joueurs. Ainsi, nous constatons une nette progression des joueurs internationaux au sein des équipes.

Figure 3



Dans la figure 3, nous observons que la majorité des équipes ont en haut de 60% de joueurs internationaux. L'équipe ayant le plus de joueurs internationaux est Fulham avec 93,75% de joueurs internationaux et l'équipe avec la plus faible proportion de joueurs internationaux est Luton Town, avec 40% de joueurs internationaux. À savoir que Luton Town est la propriété des supporters du club ce qui rend notre question de recherche encore plus valide, étant donné que le fait que les supporters locaux du club pourraient peut-être influencer la sélection des joueurs pour que ceux-ci représentent la communauté qui les supportent.

Figure 4

	(1)
(Intercept)	0.708
	(0.081)
ownerforeign	-0.005
	(0.090)
Num.Obs.	20
R2	0.000
R2 Adj.	-0.055
AIC	-12.2
BIC	-9.3
Log.Lik.	9.119

Dans la figure 4, nous observons qu'il n'y a pas de corrélation entre la nationalité du propriétaire d'un club et de la proportion de joueurs internationaux dans l'équipe. Ainsi, Le R^2 est le coefficient qui explique la proportion du modèle qui cause nos données observées et ici, nous observons que le modèle de régression linéaire que nous avons fourni n'explique pas (0%) les données observées.

Conclusion

Pour conclure, après analyse des données des joueurs de la Premier League nous constatons que même s'il y a eu un essor des joueurs internationaux dans la période étudiée entre 1992 et 2024, ceci n'est pas expliqué par le fait que les propriétaires des clubs soient étrangers. Toutefois, il y a plusieurs autres facteurs qui auraient pu être pris en compte tel que le poids financier du club, la zone géographique qu'il représente ainsi que le classement du club. Le fait qu'il n'y ait pas de corrélation pour ce modèle, ne veut toutefois pas dire que les propriétaires de club n'influencent pas le monde du sport, et il faudrait peut-être envisager d'autres types d'analyses pour quantifier cela.

Bibliographie

Gomez, C. (2022). *LA PUISSANCE ET L'UNION EUROPÉENNE: LE SPORT COMME OUTIL DE SOFT POWER ?* https://www.iris-france.org/wp-content/uploads/2022/07/06-Puissance_UE_SportLab.pdf,

Muller, M. (2023). *Peak event: The rise, crisis and potential decline of the Olympic Games and the World Cup*. <https://doi.org/10.1016/j.tourman.2022.104657>.

Power, M. J., Widdop, P., Parnell, D., Carr, J., & Millar, S. R. (2020). Football and politics: The politics of football. *Managing Sport and Leisure*, 25(1-2), 1-5. <https://doi.org/10.1080/23750472.2020.1723437>, consulté le 19 avril 2024.

Verschuuren, P. (2013). Les multiples visages du « sport power ». *Revue internationale et stratégique*, 89, 131-136. <https://doi.org/10.3917/ris.089.0131>, consulté le 19 avril 2024.

Wikipédia, List of owners of football club, https://en.wikipedia.org/wiki/List_of_owners_of_English_football_clubs, consulté le 19 octobre 2024.

Annexe

```
rm(list=ls())
library(lubridate)
# install.packages("rvest")
library(readxl)
library(rvest)
library(tidyverse)
library(RSelenium)
library(ggthemes)
library(netstat)
library(wdman)
library(pingr)
library(modelsummary)
# site web

chromeCommand <- chrome(retcommand = T, verbose = F, check = F )
chromeCommand

rD <- RSelenium::rsDriver(browser = "chrome",
                           chromeversion = "123.0.6312.106",
                           verbose = FALSE)

remDr <- rD[["client"]]

remDr$navigate("https://www.premierleague.com/stats/top/players/goals")

Sys.sleep(4)
# enlever cookies et publicités
acceptCookies <- remDr$findElement(using = "css selector",
                                     value = "#onetrust-accept-btn-handler")
acceptCookies$clickElement()

Sys.sleep(2)

closeAdd <- remDr$findElement(using = "css selector",
                              value = "#advertClose")
closeAdd$clickElement()

source <- remDr$getPageSource()[[1]]
# Obtenir les données
```

```

list_seasons <- read_html(source) %>%
  html_nodes("ul[data-dropdown-list=FOOTBALL_COMPSEASON] > li") %>%
  html_attr("data-option-name")

season93 <- which(list_seasons == "1992/93")
season24 <- which(list_seasons == "2023/24")
list_seasons <- list_seasons[c(season93, season24)]

data_seasons <- vector("list", length(list_seasons))

for (j in seq_along(list_seasons)){

  season <- remDr$findElement(using = "css selector",
                              value = ".current[data-dropdown-current=FOOTBALL_COMPSEAS

  season$clickElement()
  Sys.sleep(2)

  ELEMseason <- remDr$findElement(using = "css selector", value = str_c("ul[data-dropdown-
  ELEMseason$clickElement()
  Sys.sleep(2)

  data_seasons[[j]] <- remDr$getPageSource()[[1]] %>%
    read_html() %>%
    html_table() %>%
    .[[1]] %>%
    as_tibble(.name_repair = "unique") %>%
    select(-ncol(.))

  # Avoir les informations pour chaque page
  btnNextExists <- remDr$getPageSource()[[1]] %>%
    read_html() %>%
    html_node(".paginationNextContainer.inactive") %>%
    html_text() %>%
    is.na()

  while (btnNextExists){

```

```

    btnNext <- remDr$findElement(using = "css selector",
                                  value = ".paginationNextContainer")
    btnNext$clickElement()
    Sys.sleep(2)

    table_n <- remDr$getPageSource()[[1]] %>%
      read_html() %>%
      html_table() %>%
      .[[1]] %>%
      as_tibble(.name_repair = "unique") %>%
      select(-ncol(.))

    data_seasons[[j]] <- bind_rows(data_seasons[[j]], table_n)

    btnNextExists <- remDr$getPageSource()[[1]] %>%
      read_html() %>%
      html_node(".paginationNextContainer.inactive") %>%
      html_text() %>%
      is.na()

    Sys.sleep(1)
  }
}

view(data_seasons[[1]])

view(data_seasons[[2]])

merged_data <- bind_rows(data_seasons, .id = "Seasons")

team_data_2023 <- data_seasons[[2]] %>%
  group_by(Club) %>%
  summarise(
    TotalPlayers = n(),

```

```

    TotalNationalities = n(),
    LocalNationality = sum(Nationality == "England")
  ) %>%

  mutate(Coefficient = (TotalNationalities - LocalNationality) / TotalPlayers) %>%
  slice(-1)

data_couleurs <- data_frame(Club = unique(team_data_2023$Club),
                             couleurs = c("#EF0107", "#95BFE5", "#DA291C", "#cd3529", "#005
                             "#C8102E", "#F78F1E", "#6CABDD", "#DA291C", "#241

team_data_2023 <- merge(team_data_2023, data_couleurs, by.x = "Club", by.y = "Club", all.x

ggplot(data = team_data_2023, mapping = aes(x = Club, y = Coefficient, fill = couleurs)) +

  geom_bar(stat = "identity", position = "dodge") +
  geom_text(aes(label = scales::percent(Coefficient)),
            position = position_dodge(width = 1), vjust = -0.5) +
  scale_fill_identity() +
  scale_y_continuous(labels = scales::percent) +

  labs(title = "Les joueurs internationaux",
        subtitle = "saison (2023-2024), par club",
        x = "Club",
        y = "coefficient") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 0.5),
    axis.title.x = element_text(hjust = 0.4)
  )

ggplot(data = data_seasons[[1]], mapping = aes(x = Nationality)) +
  geom_bar(fill = "cyan") +
  labs(title = "Les joueurs internationaux",
        subtitle = "saison (1992-1993)",
        x = "Nationalité",
        y = "effectif") +
  coord_flip() +

```

```

theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5),
  plot.subtitle = element_text(hjust = 0.5),
  axis.text.x = element_text(angle = 45, hjust = 0.5),
  axis.title.x = element_text(hjust = 0.4)
)

ggplot(data = data_seasons[[2]], mapping = aes(x = Nationality)) +
  geom_bar(fill = "cyan", position = position_dodge(width = 2)) +
  labs(title = "Les joueurs internationaux",
        subtitle = "saison (2023-2024)",
        x = "Nationalité",
        y = "effectif") +
  coord_flip() +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 0.5),
    axis.title.x = element_text(hjust = 0.4)
  )

data_ownership <- read_excel("travail_session/data_ownership.xlsx")

data_2023_merged <- merge(team_data_2023, data_ownership, by.x = "Club", by.y = "Club", all = TRUE,
  select(c(Club, Coefficient, owner))

data_2023_merged <- data_2023_merged %>% mutate(owner = replace(owner, is.na(owner), "Britannique"))

regression_owner <- lm(Coefficient ~ owner, data = data_2023_merged)

modelsummary(regression_owner)

summary(regression_owner)

#system("taskkill /im java.exe /f", intern=FALSE, ignore.stdout=FALSE)
#pingr::ping_port("localhost", 4567)
# Bibliographie pour le code

```

```
# (https://www.appsilon.com/post/webscraping-dynamic-websites-with-r,  
# https://joshuamccrain.com/tutorials/web_scraping_R_selenium.html,)
```