

SocCR

Anatole GAYANT, Jules AVIGNON, Clément PEREON

2023-04-28

<!DOCTYPE html>



Pour notre projet de R avancé, nous avons choisi de récupérer les données de tous les matchs de l'édition 2020-2021 de la Ligue des champions à partir de l'API football-data.org. Pour cela, nous avons utilisé les librairies httr et dplyr.

Nous avons d'abord envoyé une requête GET à l'API en spécifiant l'URL et le token d'authentification dans les en-têtes de la requête. Nous avons ensuite converti la réponse en texte brut et l'avons parsée en utilisant la fonction fromJSON de jsonlite.

Finalement, on extrait les données des matchs en accédant à l'élément "matches" du JSON renvoyé par l'API.

Voici le code utilisé pour récupérer les données :

```
library(httr)
library(jsonlite)

#Récupération des données de tous les matchs de la Champions League
url <- "https://api.football-data.org/v4/competitions/CL/matches?season=2020"
response <- GET(url, add_headers("x-auth-token" = "000055533034025bac788669973df2"))
data <- fromJSON(rawToChar(response$content))
matches <- data$matches
```

Liste des matchs

On récupère les données de tous les matchs de la Ligue des Champions 2020-2021 depuis une API, et les stocke dans un objet "matches". Ensuite, le code utilise les packages "dplyr", "DT" et "ggplot2" pour transformer ces données en un tableau affichant les scores des matchs avec les noms des équipes, les dates, et triés par ordre chronologique décroissant. Enfin, la fonction "datatable" de DT est utilisée pour afficher ce tableau dans le R Markdown, avec des options pour la pagination, la recherche et le renommage des colonnes.

library(dplyr)

Warning: le package 'dplyr' a été compilé avec la version R 4.1.3

library(DT)

Warning: le package 'DT' a été compilé avec la version R 4.1.3

library(ggplot2)

Warning: le package 'ggplot2' a été compilé avec la version R 4.1.3

matches_tbl <- as_tibble(matches)

matches_scores <- select(matches_tbl, homeTeam, awayTeam, score, utcdate) |> mutate(homeTeam_name = pull(homeTeam, "name"), awayTeam_name = pull(awayTeam, "name"), home_score = pull(scores\$fulltime, "home"), away_score = pull(scores\$fulltime, "away"), total_score = paste(home_score, "-", away_score), date = as.Date(utcdate)) |> select(homeTeam_name, awayTeam_name, total_score, date) |> arrange(desc(matches_tbl\$utcdate))

datatable(matches_scores, options = list(pageLength = 10, lengthMenu = c(10, 20, 50)), filter = "top", rownames = FALSE, colnames = c("Home Team", "Away Team", "Total Score", "Date"), caption = "Tableau des scores Champions League 2021")

Showing 10 of 178 entries

Tableau des scores Champions League 2021

| Home Team | Away Team | Total Score | Date |
|------------------------|------------------------|-------------|------------|
| All | All | All | All |
| Manchester City FC | Chelsea FC | 0 - 1 | 2021-05-29 |
| Chelsea FC | Real Madrid CF | 2 - 0 | 2021-05-06 |
| Manchester City FC | Paris Saint-Germain FC | 2 - 0 | 2021-05-04 |
| Paris Saint-Germain FC | Manchester City FC | 1 - 2 | 2021-04-28 |
| Real Madrid CF | Chelsea FC | 1 - 1 | 2021-04-27 |
| Borussia Dortmund | Manchester City FC | 1 - 2 | 2021-04-14 |
| Liverpool FC | Real Madrid CF | 0 - 0 | 2021-04-14 |
| Chelsea FC | FC Porto | 0 - 1 | 2021-04-13 |
| Paris Saint-Germain FC | FC Bayern München | 0 - 1 | 2021-04-13 |
| FC Porto | Chelsea FC | 0 - 2 | 2021-04-07 |

Showing 1 to 10 of 178 entries

Association du tour

```
matches_scores$tour[1] <- "finale"
```

```
## Warning: unknown or uninitialised column: 'tour'.
```

```
matches_scores$tour[2:16] <- "demi1"
matches_scores$tour[16:31] <- "quarts"
matches_scores$tour[31:46] <- "huitiemes"
matches_scores$tour[46:nrow(matches_scores)] <- NA
matches_scores <- na.omit(matches_scores)
```

On utilise le fait que les matchs soient classés par ordre croissant pour leur assigner un tour. (On s'entend uniquement au match à partir des 8 ème de finale)

Représentation graphique

```
# Transformation du format des scores
matches_scores$home_score <- as.numeric(sub("(\\d-\\d)+$", "(\\d-\\d)+$", matches_scores$total_score))
matches_scores$away_score <- as.numeric(sub("(\\d-\\d)+$", "(\\d-\\d)+$", matches_scores$total_score))

# Création de la variable 'winner'
matches_scores$winner <- with(matches_scores, ifelse(home_score > away_score, homeTeam_name, ifelse(home_score < away_score, awayTeam_name, "draw")))

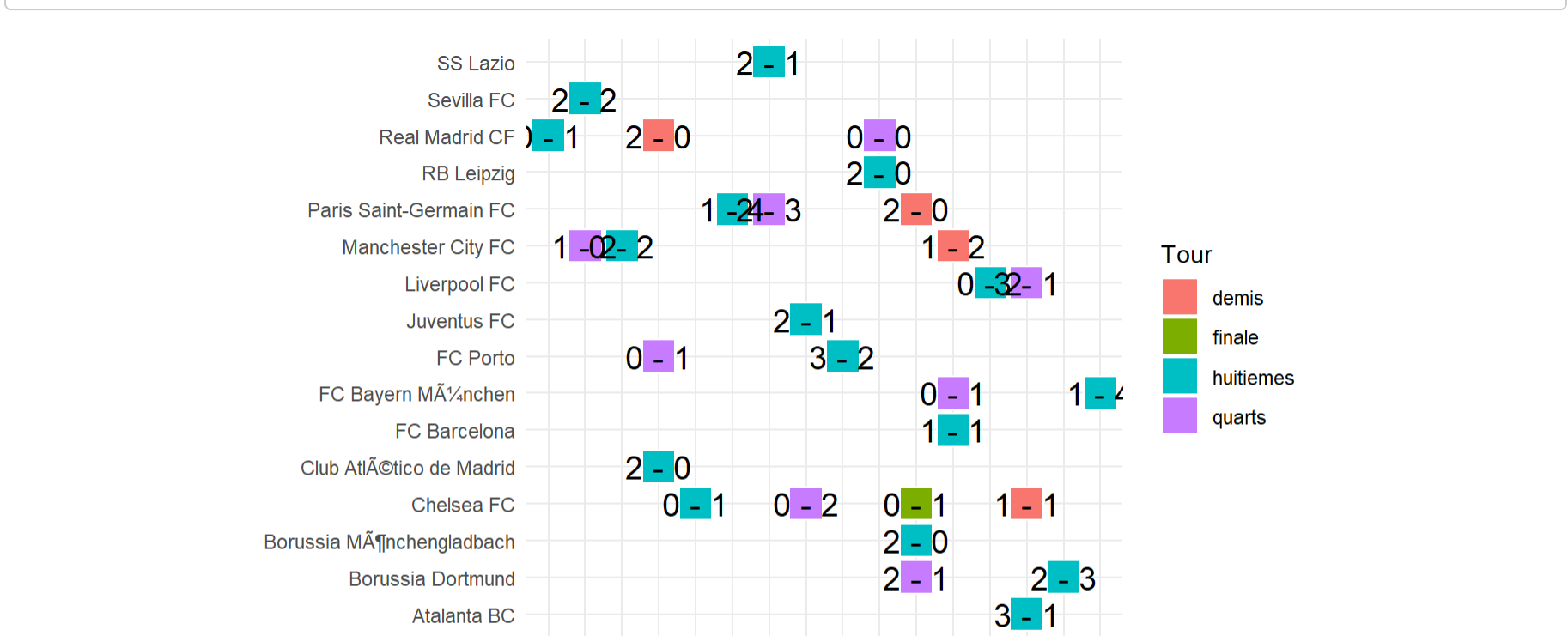
# Création du graphique avec ggplot2
ggplot(matches_scores, aes(x = homeTeam_name, y = awayTeam_name, fill = tour)) +
  geom_tile(aes(width = 0.5, height = 0.5), color = "white", size = 0.5) +
  geom_text(aes(label = total_score), size = 5) +
  scale_fill_discrete(name = "Tour") +
  theme_minimal() +
  labs(x = NULL, y = NULL) +
  coord_fixed() +
  theme(axis.text.x = element_text(angle = 45, hjust=1))
```

Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.

I please use 'linewidth' instead.

This warning is displayed once every 8 hours.

Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.



Nous nous intéressons maintenant à la page wikipédia de l'édition 2020-2021 de la Ligue des champions :

https://wikipedia.org/wiki/Ligue_des_champions_de_l'UEFA_2020-2021

```
library(rvest)
```

```
## Warning: le package 'rvest' a été compilé avec la version R 4.1.3
```

```
url <- "https://fr.wikipedia.org/wiki/Ligue_des_champions_de_l'UEFA_2020-2021"
data_html <- read_html(url)
```

Voici, pour chaque pays, le nombre d'équipes qualifiées à l'issue de la phase de groupe :

```
# Tableau des équipes qualifiées à l'issue de la phase de groupe
css_selector <- "new-content-text > div > table:nth-of-type(25)"
df_qualif <- data_html |> html_nodes(css_selector) |> html_table()
names(df_qualif[[1]])[2] <- "premiers"
names(df_qualif[[1]])[3] <- "deuxiemes"
```

Showing 10 of 8 entries

| Groupe | premiers | deuxiemes |
|--------|---------------------|--------------------------|
| 1 A | Bayern Munich | Atletico de Madrid |
| 2 B | Real Madrid | Borussia Mönchengladbach |
| 3 C | Manchester City | FC Porto |
| 4 D | Liverpool FC | Atalanta Bergame |
| 5 E | Chelsea FC | Séville FC |
| 6 F | Borussia Dortmund | SS Lazio |
| 7 G | Juventus FC | FC Barcelone |
| 8 H | Paris Saint-Germain | RB Leipzig |

Showing 1 to 8 of 8 entries

Voici le détail de la phase à élimination directe avec, pour chaque tour, les scores aux matchs aller, aux matchs retours et totaux.

Huitièmes de finale

```
css_selector <- "new-content-text > div > table:nth-of-type(26)"
df_huit <- data_html |> html_nodes(css_selector) |> html_table()
datatable(df_huit[[1]])
```

Showing 10 of 8 entries

| Équipe | Aller | Total | Retour | Équipe |
|----------------------------|-------|--------|----------|---------------------|
| 1 Borussia Mönchengladbach | 0 - 2 | 0 - 4 | 0 - 2 | Manchester City |
| 2 SS Lazio | 1 - 4 | 2 - 6 | 1 - 2 | Bayern Munich |
| 3 Atlético de Madrid | 0 - 1 | 0 - 3 | 0 - 2 | Chelsea FC |
| 4 RB Leipzig | 0 - 2 | 0 - 4 | 0 - 2 | Liverpool FC |
| 5 FC Porto | 2 - 1 | 4e - 4 | 2 - 3 ap | Juventus FC |
| 6 FC Barcelone | 1 - 4 | 2 - 5 | 1 - 1 | Paris Saint-Germain |
| 7 Séville FC | 2 - 3 | 4 - 5 | 2 - 2 | Borussia Dortmund |
| 8 Atalanta Bergame | 0 - 1 | 1 - 4 | 1 - 3 | Real Madrid |

Showing 1 to 8 of 8 entries

Quarts de finale

```
# Tableau des quarts de finale
css_selector <- "new-content-text > div > table:nth-of-type(27)"
df_quart <- data_html |> html_nodes(css_selector) |> html_table()
datatable(df_quart[[1]])
```

Showing 10 of 4 entries

| Équipe | Aller | Total | Retour | Équipe |
|-------------------|-------|--------|--------|---------------------|
| 1 Manchester City | 2 - 1 | 4 - 2 | 2 - 1 | Borussia Dortmund |
| 2 FC Porto | 0 - 2 | 1 - 2 | 1 - 0 | Chelsea FC |
| 3 Bayern Munich | 2 - 3 | 3 - 3e | 1 - 0 | Paris Saint-Germain |
| 4 Real Madrid | 3 - 1 | 3 - 1 | 0 - 0 | Liverpool FC |

Showing 1 to 4 of 4 entries

Demi-finales

```
# Tableau des demi-finales
css_selector <- "new-content-text > div > table:nth-of-type(28)"
df_demi <- data_html |> html_nodes(css_selector) |> html_table()
datatable(df_demi[[1]])
```

Showing 10 of 2 entries

| Équipe | Aller | Total | Retour | Équipe |
|-----------------------|-------|-------|--------|-----------------|
| 1 Paris Saint-Germain | 1 - 2 | 1 - 4 | 0 - 2 | Manchester City |
| 2 Real Madrid | 1 - 1 | 1 - 3 | 0 - 2 | Chelsea FC |

Showing 1 to 2 of 2 entries

Finale

```
# Tableau de la finale
css_selector <- "new-content-text > div > table:nth-of-type(29)"
df_finale <- data_html |> html_nodes(css_selector) |> html_table()
datatable(df_finale[[1]])
```

Showing 10 of 1 entries

| Équipe 1 | Score | Équipe 2 |
|-------------------|-------|------------|
| 1 Manchester City | 0 - 1 | Chelsea FC |

Showing 1 to 1 of 1 entries

Meilleures attaques

```
css_selector <- "new-content-text > div.new-parser-output > table:nth-child(195) > tbody > tr > td:nth-child(1) > table > tbody"
df_buteurs <- data_html |> html_nodes(css_selector) |> html_table()
datatable(df_buteurs[[1]])
```

Showing 10 of 32 entries

| Rang | Club | Buts[25] | Buts/match | Matches joués |
|------|----------------------------|----------|------------|---------------|
| 1 | 1 Bayern Munich | 27 | 2.7 | 10 |
| 2 | 2 Manchester City | 25 | 1.92 | 13 |
| 3 | 3 Chelsea FC | 23 | 1.77 | 13 |
| 4 | 4 Paris Saint-Germain | 22 | 1.83 | 12 |
| 5 | 5 Real Madrid | 19 | 1.58 | 12 |
| 6 | 6 Borussia Dortmund | 19 | 1.9 | 10 |
| 7 | 7 Juventus FC | 18 | 2.25 | 8 |
| 8 | 7 FC Barcelone | 18 | 2.25 | 8 |
| 9 | 9 Borussia Mönchengladbach | 16 | 2 | 8 |
| 10 | 10 Liverpool FC | 15 | 1.5 | 10 |

Showing 1 to 10 of 32 entries

Graphique des meilleurs attaques

Liste de correspondance entre clubs et couleurs de maillot

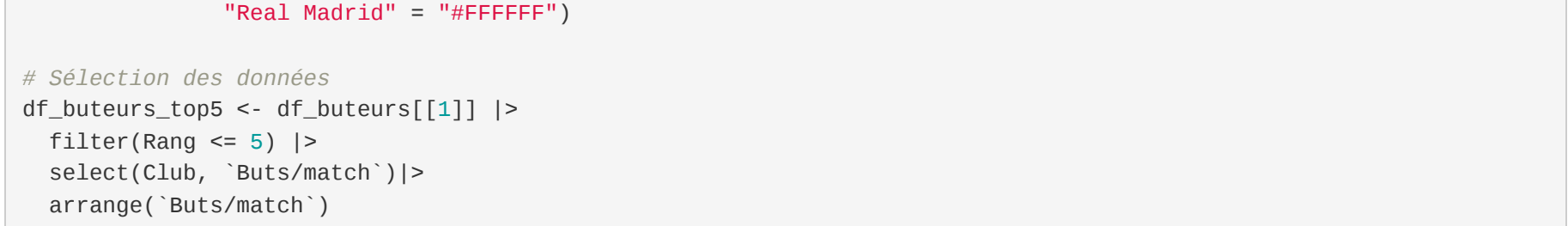
colors <- list("Bayern Munich" = "#005599", "Manchester City" = "#75A086", "Chelsea FC" = "#034A99", "Paris Saint-Germain" = "#004444", "Real Madrid" = "#FFFFFF")

Sélection des données

df_buteurs_top <- df_buteurs[[1]] |> filter(rank < 5) |> select(Club, Buts/match) |> arrange(Buts/match)

Histogramme

ggplot(df_buteurs_top, aes(x = Club, y = Buts/match, fill = Club)) + geom_bar(stat = "identity") + scale_fill_manual(values = colors) + labs(title = "Histogramme des buts par match des 5 équipes qui ont marqué le plus", x = "Club", y = "Buts par match")



Meilleurs buteurs

```
css_selector <- "new-content-text > div.new-parser-output > table:nth-child(196) > tbody > tr > td:nth-child(1) > table > tbody"
df_buteurs <- data_html |> html_nodes(css_selector) |> html_table()
datatable(df_buteurs[[1]])
```

Showing 10 of 32 entries

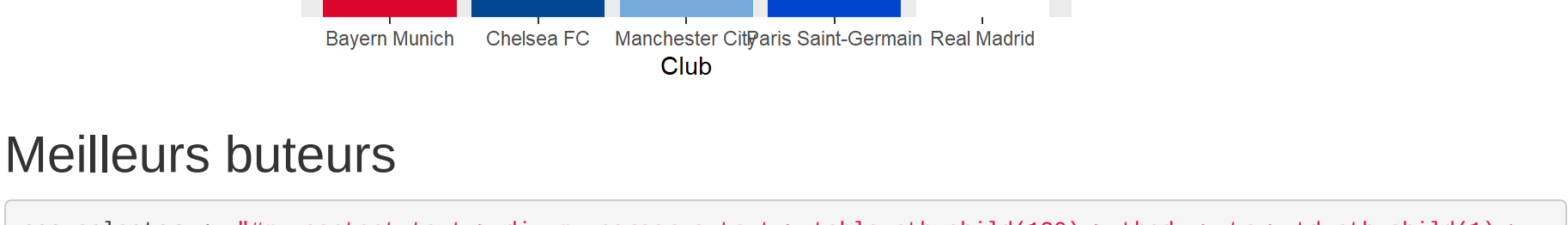
| Rang | Buteur | Club | Buts[22] | Passes | Minutes jouées |
|------|------------------------|--------------------------|----------|--------|----------------|
| 1 | 1 Erling Braut Haaland | Borussia Dortmund | 10 | 2 | 705 |
| 2 | 2 Kylian Mbappé | Paris SG | 8 | 3 | 900 |
| 3 | 3 Neymar | Paris SG | 6 | 3 | 746 |
| 4 | 4 Álvaro Morata | Juventus FC | 6 | 1 | 597 |
| 5 | 5 Mohamed Salah | Liverpool FC | 6 | 1 | 781 |
| 6 | 6 Olivier Giroud | Chelsea FC | 6 | 0 | 257 |
| 7 | 7 Youssef En-Nesyri | FC Séville | 6 | 0 | 386 |
| 8 | 8 Marcus Rashford | Manchester United | 6 | 0 | 416 |
| 9 | 9 Karim Benzema | Real Madrid | 6 | 0 | 842 |
| 10 | 10 Alassane Plea | Borussia Mönchengladbach | 5 | 3 | 540 |

Showing 1 to 10 of 32 entries

df_buteurs_top <- df_buteurs[[1]] |> slice(1:5) |> arrange(desc(Buts[22]))

ggplot(df_buteurs_top, aes(x = reorder(Buteur, Buts[22]), y = Buts[22])) + geom_bar(stat = "identity", fill = "steelblue") + coord_flip()

theme(axis.text.x = element_text(angle = 45, justify = "left", hjust = 1)) + labs(title = "Histogramme du nombre de buts des 5 meilleurs buteurs", x = "Buteur", y = "Nombre de buts")



Meilleurs passeurs

css_selector <- "www-content-text > div.rw-parser-output > table:nth-child(100) > tbody > tr > td:nth-child(2) > table > tbody"
df_passeurs <- data_html %> html_nodes(css_selector) %> html_table() %> data_table(df_passeurs[[1]])

Show10entries

Search:

| | Rang | Passeur | Club | Passes(23) | Buts | Minutes jouées |
|----|------|-----------------|--------------------------|------------|------|----------------|
| 1 | 1 | Juan Cuadrado | Juventus FC | 6 | 0 | 551 |
| 2 | 2 | Kevin De Bruyne | Manchester City | 4 | 3 | 609 |
| 3 | 3 | Joshua Kimmich | Bayern Munich | 4 | 1 | 617 |
| 4 | 4 | Ángel Di María | Paris SG | 4 | 1 | 697 |
| 5 | 5 | Kylian Mbappé | Paris SG | 3 | 8 | 900 |
| 6 | 6 | Neymar | Paris SG | 3 | 6 | 746 |
| 7 | 7 | Alessane Pléa | Borussia Mönchengladbach | 3 | 5 | 540 |
| 8 | 8 | Kingsley Coman | Bayern Munich | 3 | 3 | 459 |
| 9 | 9 | Angelino | RB Leipzig | 3 | 3 | 627 |
| 10 | 10 | Phil Foden | Manchester City | 3 | 2 | 386 |

Showing 1 to 10 of 10 entries

Previous1Next

Agregation

En agrégeant, on obtient le nombre de passes décisives par équipe.

passes_equipe <- df_passeurs[[1]] %> group_by(Club) %> summarize(nb_passes = sum('Passes(23)'))
passes_equipe

```
## # A tibble: 6 x 2
##   Club          nb_passes
##   <chr>          <int>
## 1 Bayern Munich          7
## 2 Borussia Mönchengladbach  3
## 3 Juventus FC            6
## 4 Manchester City          7
## 5 Paris SG             10
## 6 RB Leipzig             3
```