

# Head coach dismissal effect on football team performance

Mathis Derenne<sup>1</sup>, Ewann Boismartel<sup>1</sup>, Scott Thomas<sup>1</sup>, and Romain Biguet<sup>1</sup>

<sup>1</sup>University of Rennes

---

## Abstract

This paper aims to explore the relationship between the team's performance and sports coach. To do that, we use traditional statistical methods applied to Men's Football First Divisions of major leagues between 2015 and 2023. We investigate this relation on three different aspects: the effect of head coach tenure, the effect of regular renewal of sports coach by club, and the effect of changing clubs on the coach's performance. Our results show that there is a positive correlation between head coach tenure and team's performance, while the regular renewal of a sports coach by a club has a negative relation on the team's performance. We also find that head coach who regularly change club is negatively correlated with team's performance. These results suggest that stability and continuity are important factors for the success of a football team.

---

**Keywords** coach dismissal, team performance

## 1. INTRODUCTION

### 1.a. Cadre de la problématique

### 1.b. Source des données

Les données utilisées au cours de cette analyse sont extraites de deux sites spécialisés dans les statistiques de football : [Fbref](#) et [Transfermarkt](#).

- Fbref offre une gamme complète de données statistiques sur les joueurs, les équipes, les ligues et les compétitions de football du monde entier. Il propose des informations détaillées telles que les buts marqués, les passes décisives, les tirs au but, les interceptions et bien d'autres statistiques.
- Transfermarkt est une ressource en ligne majeure pour tout ce qui concerne les transferts de joueurs de football, les rumeurs de transferts, les valeurs marchandes des joueurs ainsi que les informations sur les contrats. Il offre une base de données exhaustive des joueurs, des clubs et des agents, ainsi que des détails sur les transferts passés et actuels.

Ces sites sont utilisés par les amateurs de football, les journalistes et les professionnels pour rester informés sur les évolutions au cours de la saison ou pendant les trêves/mercatos.

### 1.c. Fiabilité des données

Ces sites sont très utilisés et considérés comme fiables. Fbref est entretenu par l'entreprise Sport Reference qui gère également d'autres sites spécialisés dans les statistiques sportives, comme Baseball-Reference et Basketball Reference. Les données sur Fbref sont souvent vérifiées et mises à jour régulièrement, ce qui contribue à leur fiabilité. Pour Transfermarkt, c'est aussi un site très utilisé pour les rumeurs de transferts et les transferts en général, il a une réputation de site fiable. Le site recueille des données sur les transferts, les valeurs

**Published** Apr 13, 2024

## Data Availability

L'ensemble des fichiers et données relatif à ce travail sont disponible en accès libre sur le [dépot GitHub](#) sous licence MIT.

marchandes des joueurs et d'autres détails liés aux contrats à partir de diverses sources, y compris les médias et les communiqués officiels des clubs. Cependant, c'est un site reliant des rumeurs de transferts, donc il peut y avoir des inexactitudes ou des spéculations qui ne se concrétisent pas toujours. Il est donc conseillé de vérifier les informations avec d'autres sources fiables, notamment lorsqu'il s'agit de transferts non confirmés.

#### 1.d. Références et outils utilisés

L'intégralité du travail de récupération, de pré-traitement, d'analyse et visualisation des données a été réalisé au sein de Notebook Jupyter.

La récupération des données footballistique a été effectuée à l'aide du package R [WorldFootballR](#). Ce package est régulièrement mis à jour et implémente des outils de web scraping afin d'extraire les données des principaux sites footballistiques.

Le pré-traitement, l'analyse et la visualisation des données a été effectué sous Python à l'aide de bibliothèques standards : Pandas, Numpy, Matplotlib, Seaborn, Scipy.

L'écriture de ce papier a été réalisé dans un fichier Markdown.

[MyST](#) fait partie d'un écosystème d'outils qui cherchent à améliorer le travail de communication scientifique en favorisant le développement d'une science reproductible et indexable. Cet outil a été utilisé pour permettre la diffusion de ce papier de recherche au format d'un [site statique](#) et d'un [PDF](#) répondant aux exigences de qualité scientifique.

MyST permet de réutiliser les entrées et les sorties des Notebooks Jupyter. Ainsi l'ensemble des figures, tableaux et une majorité des variables présentes dans ce papier sont directement issus des Notebooks Jupyter. À titre d'exemple, il est possible de renouveler l'intégralité de l'étude à d'autres ligues ou d'autres périodes en modifiant simplement les valeurs des variables utilisées dans les Notebooks Jupyter :

```
country <- c("ENG", "ESP", "ITA", "GER", "FRA")
year <- c(2015:2022)
```

Enfin, la lecture de *Fundamentals of Data Visualization* (Wilke, 2019) a permis d'améliorer la qualité des graphiques et de la présentation des données en les rendant plus clairs et informatifs.

## 2. EXTRACTION DES DONNÉES

L'extraction se fait aisément à l'aide de [WorldFootballR](#).

Un premier jeu de données concernant les matchs est récupéré à partir de [Fbref](#). Il contient des informations sur les matchs de football, notamment les équipes qui ont joué, le score final, le lieu du match et la date du match. Un second jeu de données concernant les entraîneurs sportifs est récupéré à partir de [Transfermarkt](#). Il contient des informations sur les entraîneurs de football, notamment leur nom, leur date de naissance, leur nationalité, les dates de début et de fin de leur mandat, ainsi que des statistiques sur les matchs qu'ils ont dirigés.

Table 1: Extrait du jeu de données sur les mandats des entraîneurs sportifs

Table 2: Extrait du jeu de données sur les matchs

## 3. TRAITEMENT DES DONNÉES

Le traitement des données est une étape essentielle dans le processus d'analyse des données. Il permet de vérifier l'intégrité des données, de les adapter aux besoins de l'étude et de les préparer pour l'analyse statistique.

Davantage de détails sur le traitement des données sont disponibles dans les Notebooks Jupyter associés à ce papier.

### 3.a. Données d'intérêts

On exclut des données de résultat de matchs, les matchs ne concernant pas des équipes de première division. De plus, on exclut des données les entraîneurs sportifs n'ayant pas été actifs entre le début de la saison 2015 et la fin de la saison 2023 (14/02/2024).

### 3.b. Correspondance des noms d'équipes

Les résultats de matchs contiennent 31 équipes et les mandats de coaches contiennent 13 équipes. Cependant, le nom de certaines équipes est différent entre les deux jeux de données. Par exemple, « Liverpool » dans les résultats de match devient « Liverpool FC » dans les mandats des entraîneurs.

Il existe 59 équipes présentes dans les mandats de coaches qui n'ont pas de correspondance dans les résultats de match.

Team in head coach records	Team in match results
Newcastle United	Newcastle Utd
Crystal Palace	Crystal Palace
Chelsea FC	Chelsea
Arsenal FC	Arsenal
Liverpool FC	Liverpool

Table 3: Exemple de correspondance des noms d'équipes

Nous avons utilisé l'algorithme de la distance Levenshtein (Levenshtein, 1965) afin de faire correspondre les équipes des mandats d'entraîneurs aux équipes des résultats de matchs. Nous restreignons par ailleurs la recherche d'une équipe correspondante à la liste des équipes du pays correspondant afin de limiter le nombre de correspondances possibles. L'ajustement du score de Levenshtein a permis de rapidement vérifier la validité des correspondances de plus faible certitude.

```
from thefuzz import process

team_name_mapping = {}

# For each country
for country in coach_teams_by_country.index:
    # Get teams for this country
    coach_teams = coach_teams_by_country[country]
    match_teams = match_teams_by_country.get(country, [])

    # For each team in coach_teams
    for coach_team in coach_teams:
        # Find the best match in match_teams
        matching_scores = process.extract(coach_team, match_teams, limit=1)

        if len(matching_scores) != 0 and matching_scores[0][1] >= 60:
```

```

team_name_mapping[coach_team] = matching_scores[0][0]
else:
    team_name_mapping[coach_team] = None
    print(f"No match found for {coach_team} among {match_teams} in
{country}")

```

Code 1: Utilisation de l'algorithme de la distance Levenshtein

### 3.c. Intégrité des données

La vérification de la qualité des données est une étape essentielle dans le processus d'analyse des données. Elle permet de s'assurer que les données sont complètes, cohérentes et fiables. Ces vérifications ont été effectués par le biais d'hypothèses simples sur les données :

- toutes les équipes ont une league et un pays associé
- un seul coach par équipe à une date donnée
- le nombre total de match est égale à la somme du nombre de victoires, de matchs nuls et de défaites
- etc.

Nous avons trouvé 11 enregistrements de mandats d'entraîneurs sportifs qui avaient lieu alors qu'un autre chef-entraîneur assurait l'entraînement de l'équipe. Ces enregistrements concernent les équipes suivante : 'FC Empoli, Hellas Verona, LOSC Lille, Newcastle United, Rayo Vallecano, SSC Napoli, Stade Reims, Stade Rennais FC, Torino FC, West Ham United'. Ces enregistrements sont exclus du jeu de données.

Table 4: Example of inconsistency in the head coach data

### 3.d. Modification des données des matchs

Les [Table 2](#) sont modifiés de manière à ce que chaque match soit divisé en deux lignes : une pour chaque équipe. Ainsi, chaque ligne représente le résultat d'une équipe lors d'un match : la date du match, le résultat final, la présence ou non de l'équipe à domicile, le score final de l'équipe ainsi que le nom de l'équipe.

De plus, nous ajoutons aux données de match l'ancienneté du coach au sein de l'équipe lorsque ce match a été joué. Cela nous permettra de voir si l'ancienneté de l'entraîneur au sein d'une équipe a un impact sur sa performance. Le tableau suivant est ainsi obtenu :

Table 5: Extrait du jeu de donnée sur les matchs après pré-traitement

## 4. PRÉSENTATION DES DONNÉES

Nous avons collecté les résultats des matchs et les mandats des entraîneurs sportif chef de la première division masculine de football pour les saisons 2015 - 2023 et pour les ligues suivantes : 'Premier League, La Liga, Ligue 1, Bundesliga, Serie A'.

Cela correspond à un total de 14452 matchs joués par 161 équipes. Parmi ces 161 équipes, 95 équipes n'ont pas d'information sur les mandats des entraîneurs chefs pour tout ou partie des saisons 2015 à 2023. En proportion de match c'est '46' % des matchs qui ne comportent pas d'information sur l'entraîneur chef.

Les données des mandats des entraîneurs sportifs contiennent 381 mandat d'entraîneurs chefs parmi 249 entraîneurs uniques.

Table 6: Summary of the match data

### 4.a. L'avantage de jouer à domicile

En moyenne, l'équipe qui joue à domicile marque '1.50' but contre '1.18' but pour l'équipe jouant à l'extérieur ('22'% moins de buts).

En moyenne, les équipes jouant à domicile remportent '43' % des matchs contre seulement '29' des matchs pour les équipes jouant à l'extérieur. Les matchs nuls représentent '26' % des matchs.

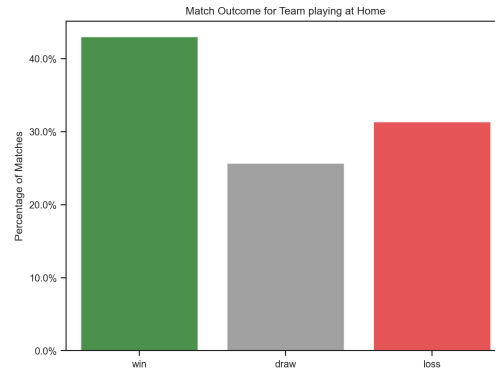


Figure 1: Venue effect on team's performance (2015 - 2023)

#### 4.b. Les dynamiques temporelles

Les saisons de football sont divisées en deux périodes : la saison régulière et la saison hors-saison. La saison régulière est la période pendant laquelle les équipes jouent des matchs de championnat et de coupe, tandis que la saison hors-saison marque une période de pause permettant à l'équipe de s'entraîner et de se préparer pour la saison suivante.

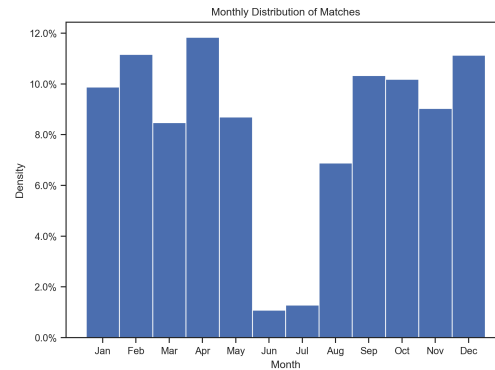


Figure 2: Monthly Distribution of Matches (2015 - 2023)

Une proportion importante des nominations et licenciement des entraîneurs sportifs a lieu en période de hors-saison ((Figure 3) et Figure 4). En effet, les clubs profitent de cette période pour revoir leur stratégie et renouveler certains membres de l'équipe. Le changement d'entraîneur au cours de la saison régulière est plus rare et risqué, car l'équipe est déjà engagée dans des compétitions et le changement d'entraîneur peut perturber la dynamique de l'équipe. Le licenciement d'un entraîneur en cours de saison pourrait indiquer une situation critique pour le club, comme une série de mauvais résultats ou des conflits internes.

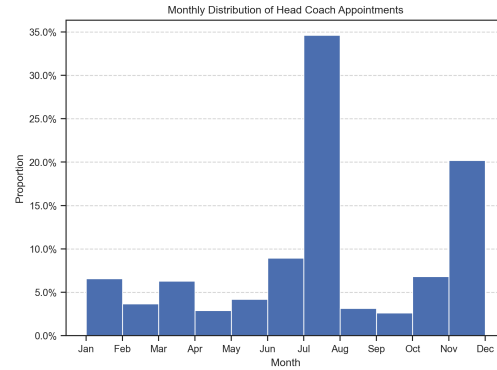


Figure 3: Monthly Distribution of Head Coaches Appointments

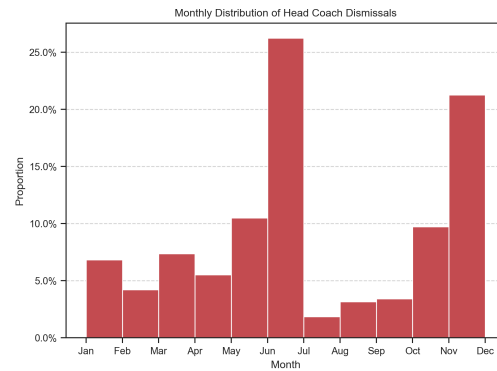


Figure 4: Monthly Distribution of Head Coaches Dismissals

Les périodes de nominations et de licenciement des entraîneurs sportifs sont relativement stables au cours des saisons, mais aussi entre les clubs (Figure 5).

<Figure size 1920x1440 with 0 Axes>

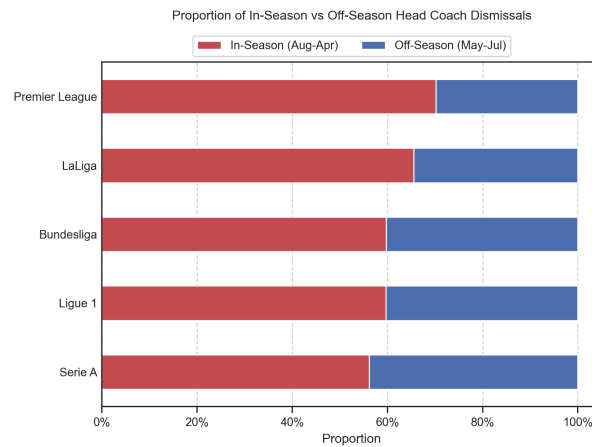


Figure 5: Proportion of in-season vs off-season head coach dismissal per league

## 5. ANALYSE STATISTIQUE

### 5.a. Définitions

Le coefficient de corrélation de Pearson (valeur de  $r$ ) est une mesure statistique qui évalue la force et la direction de la relation linéaire entre deux variables continues. Il est souvent utilisé pour quantifier la relation entre deux variables.

- Lorsque  $r = 1$ , cela indique une corrélation linéaire parfaite positive.
- Lorsque  $r = -1$ , cela indique une corrélation linéaire parfaite négative.
- De plus, lorsque  $r = 0$ , cela signifie qu'il n'y a aucune corrélation linéaire entre les deux variables.

La p-value (valeur de  $p$ ) est une mesure statistique utilisée pour déterminer la signification statistique d'un résultat. Dans le contexte de la corrélation de Pearson, la p-value est utilisée pour évaluer si la corrélation observée entre deux variables est statistiquement significative ou non. Si la p-value est inférieure à seuil (ici  $p < 0.05$ ), on rejette l'hypothèse selon laquelle il n'y a pas de corrélation dans la population, et on conclut qu'il y a une corrélation significative entre les deux variables.

#### 5.b. Fréquence de renouvellement du coach sur la performance du club

Dans cette partie, nous allons étudier les effets de la fréquence de renouvellement des coaches sur la performance d'une équipe. L'objectif est de voir les effets sur les résultats obtenus lorsque qu'une équipe change régulièrement d'entraîneur.

La [Figure 6](#) nous informe quant à la distribution du nombre d'entraîneurs employés par les clubs durant la période 2015 - 2023. On observe que plus de 85 % des clubs ont employés au moins 3 entraîneurs différents suggérant qu'il y a un renouvellement régulier des entraîneurs dans les clubs de football.

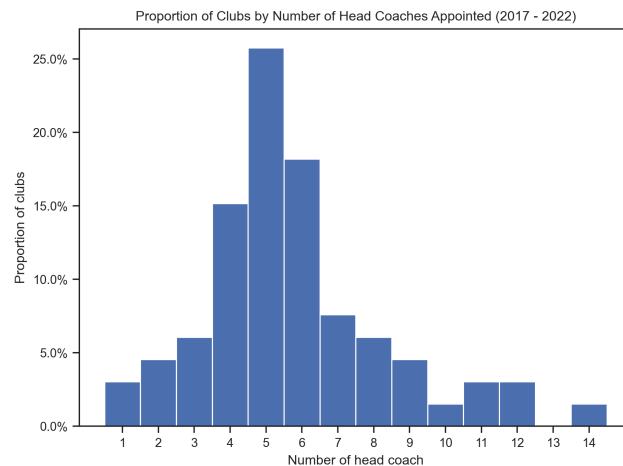


Figure 6: Proportion of Clubs by Number of Head Coaches Appointed (2015 - 2023)

La [Figure 7](#) montre que les entraîneurs de la Premier League restent en poste plus longtemps que ceux des autres ligues. De plus, les équipes de la Premier League changent moins souvent d'entraîneur que celles des autres ligues. À l'inverse, la LaLiga renouvelle fréquemment ses entraîneurs.

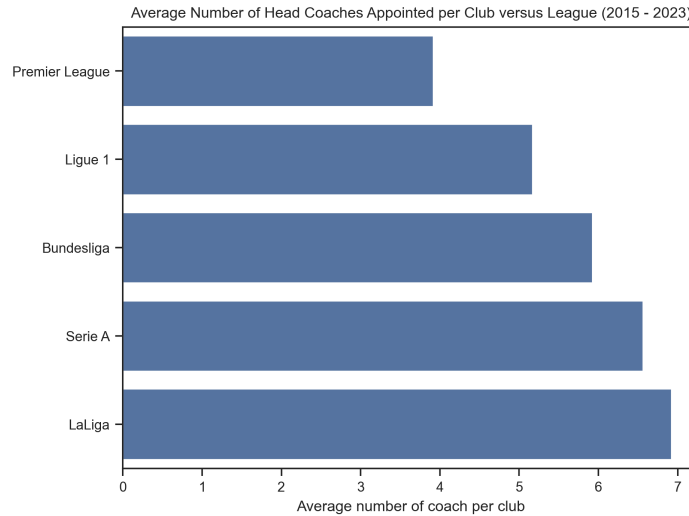


Figure 7: Average Number of Head Coaches Appointed per Club versus League (2015 - 2023)

Les [Figure 8](#), [Figure 9](#) et [Figure 10](#) s'intéressent aux ratios de victoires, de matchs nuls et de défaites des clubs en fonction du nombre d'entraîneurs nommés à la tête de l'équipe durant la période 2015 - 2023. Ces relations pourraient montrer l'effet de la fréquence de remplacement d'un coach sur les performances de l'équipe.

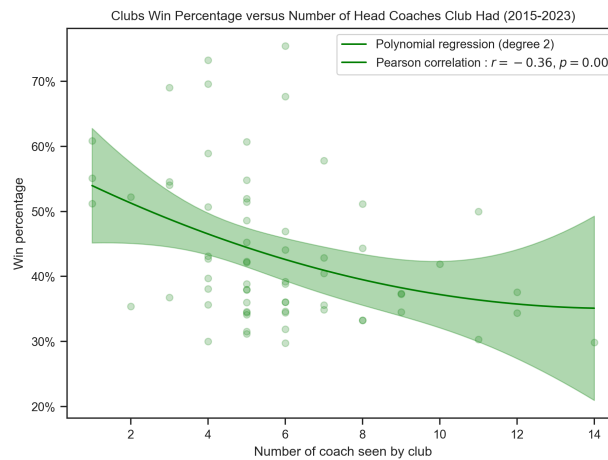


Figure 8: Win Ratio of Clubs versus Number of Head Coaches Appointed by Club



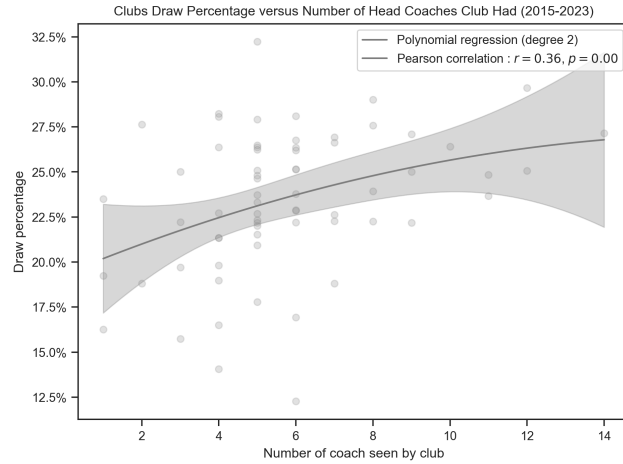


Figure 9: Draw Ratio of Clubs versus Number of Head Coaches Appointed by Club

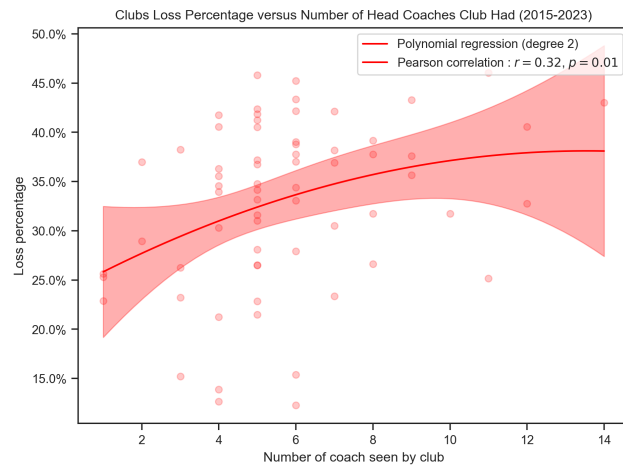


Figure 10: Loss Ratio of Clubs versus Number of Head Coaches Appointed by Club

Ainsi, nous observons :

- [Figure 8](#) une corrélation négative modérée ( $r = -0.33$ ) statistiquement significative ( $p = 0.01$ ) entre le nombre d'entraîneurs nommés par le club et son ratio de victoires.
- [Figure 9](#) une corrélation positive modérée ( $r = 0.35$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre d'entraîneurs nommés par le club et son ratio de matchs nuls.
- [Figure 10](#) une corrélation positive faible ( $r = 0.28$ ) statistiquement significative ( $p = 0.03$ ) entre le nombre d'entraîneurs nommés par le club et son ratio de défaites.

Ces résultats suggèrent que le renouvellement régulier des entraîneurs a un impact négatif sur les performances des équipes. Cela pourrait s'expliquer par le fait que les entraîneurs ont besoin de temps pour s'adapter à leur nouvel environnement et pour mettre en place leur stratégie de jeu. De plus, les entraîneurs qui restent plus longtemps à la tête de l'équipe ont tendance à mieux connaître les joueurs et à mieux comprendre les forces et les faiblesses de l'équipe, ce qui peut contribuer à améliorer les performances de l'équipe.

### 5.c. Mobilité des coachs sur la performance des équipes

Ici, nous allons observer la relation entre les performances des entraîneurs et le nombre de clubs par lesquels ils sont passés. Cette relation nous permettra de voir si le nombre de clubs qu'un entraîneur a faits a un impact sur les résultats de l'équipe.

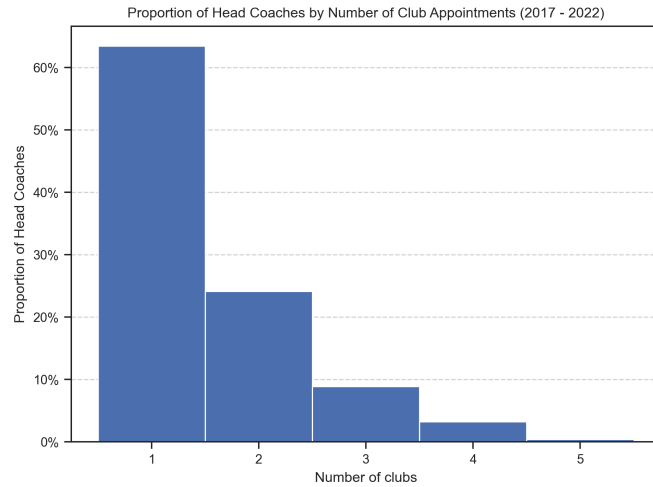


Figure 11: Proportion of Head Coaches by Number of Club Appointments (2015 - 2023)

Au cours de la période 2017 - 2022, plus de 60 % des entraîneurs sportifs n'ont entraîné qu'un seul club. Environ 20 % des entraîneurs ont entraîné 2 clubs et seulement 10 % des entraîneurs ont entraîné plus de 3 clubs au cours de cette période (voir [Figure 11](#)).

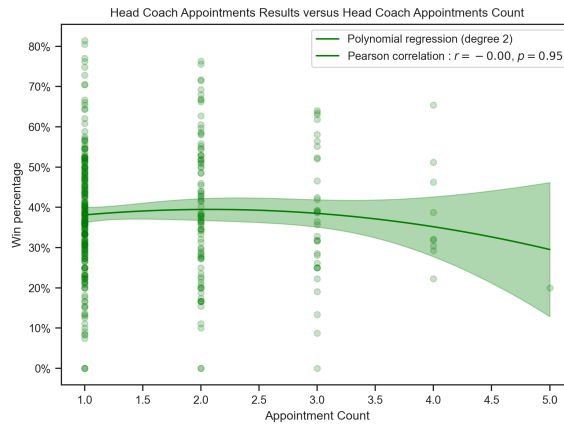


Figure 13: Head Coach Appointment Win Ratio versus Appointment Count

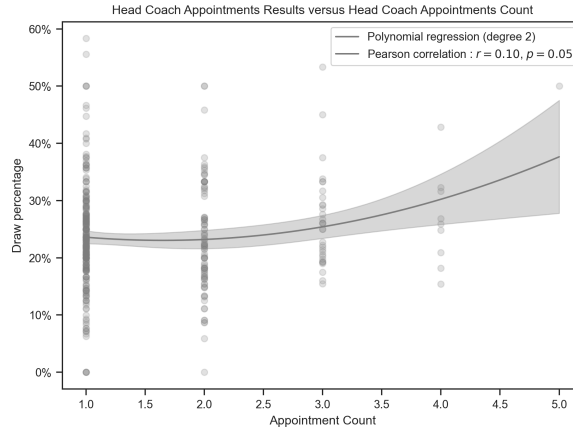


Figure 14: Head Coach Appointment Draw Ratio versus Appointment Count

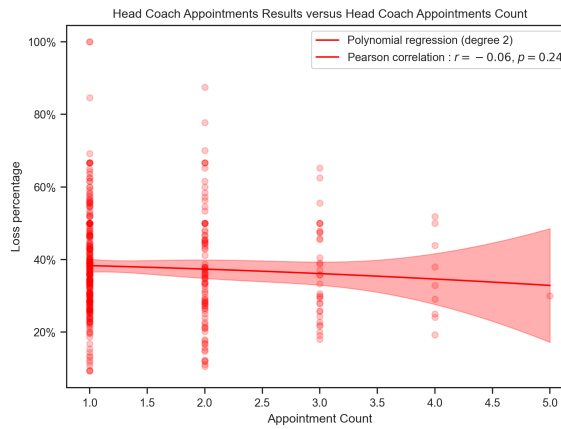


Figure 15: Head Coach Appointment Loss Ratio versus Appointment Count

Figure 12: Head Coach Appointment Performance versus Appointment Count

La [Figure 12](#) s'intéresse aux pourcentages de victoires, de matchs nuls et de défaites des mandats individuel des entraîneurs par rapport au nombre de clubs pour lesquels ils ont travaillés durant la saison 2015 - 2023. Ces graphiques permettent de visualiser la relation entre l'expérience acquise par l'entraîneur et l'impact sur la performance de leurs équipes.

Nous observons qu'il n'existe pas de corrélation entre le nombre de clubs pour lesquels un entraîneur a travaillé et son pourcentage de victoires ( $p = 0.48$ ), de matchs nuls ( $p = 0.09$ ) ou de défaites ( $p = 0.07$ ). Cela suggère que l'expérience acquise par l'entraîneur en travaillant pour plusieurs clubs n'a pas d'impact statistiquement observable sur la performance de l'équipe.

La [Figure 16](#) suivante diffère quant à elle car elle s'intéresse à la performance agrégée des entraîneurs par rapport au nombre total de club pour lesquels ils ont travaillés durant la saison 2015 - 2023.

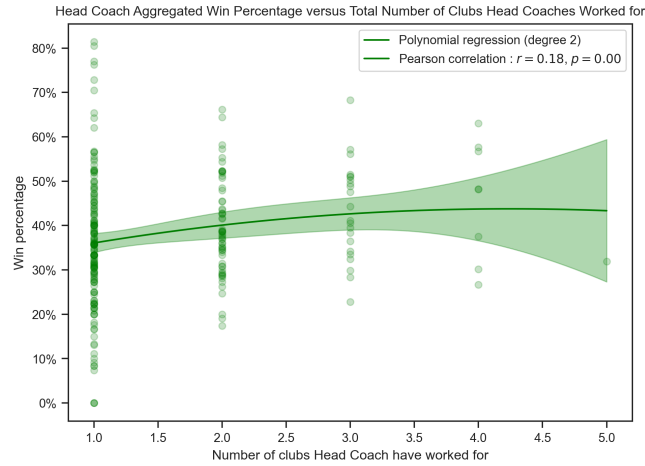


Figure 17: Head Coach Aggregated Win Ratio versus Total Number of Appointments

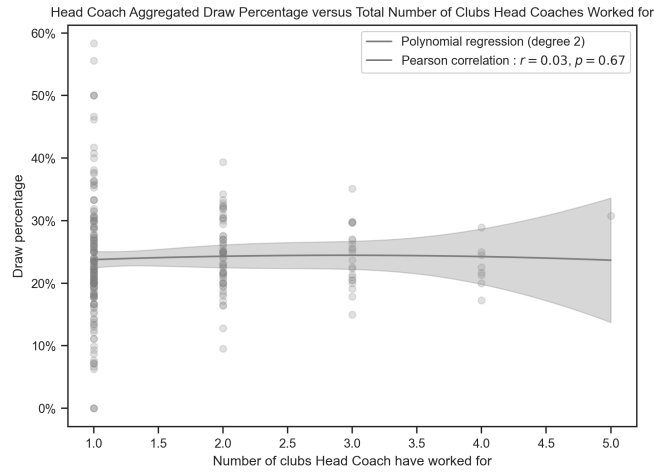


Figure 18: Head Coach Aggregated Draw Ratio versus Total Number of Appointments

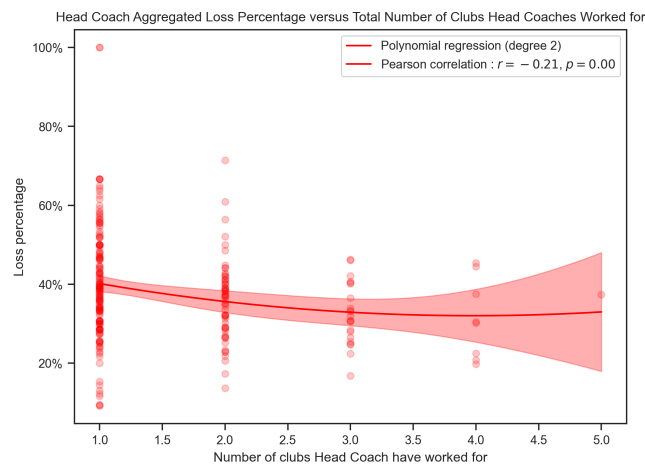


Figure 19: Head Coach Aggregated Loss Ratio versus Total Number of Appointments

Figure 16: Performance of Head Coaches versus Number of Clubs Appointments

Ces graphiques permettent de visualiser la relation entre la mobilité des entraîneurs et la performance de leurs équipes. Ainsi, nous observons :

- [Figure 17](#) une corrélation positive faible ( $r = 0.23$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre de clubs entraînés par le coach et son ratio de victoires.
- [Figure 18](#) une corrélation positive très faible ( $r = 0.01$ ) mais statistiquement non significative ( $p = 0.89$ ) entre le nombre de clubs entraînés par le coach et son ratio de matches nuls.
- [Figure 19](#) une corrélation négative faible ( $r = -0.25$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre de clubs entraînés par le coach et son ratio de défaites.

Ces graphiques ne suggèrent pas que le nombre de clubs pour lesquels un entraîneur a travaillé a un impact positif sur la performance des équipes qu'il entraîne, car l'on a précédemment montré avec la [Figure 12](#) qu'il n'existait pas de corrélation statistiquement significative. Cependant, ce graphique suggère que les entraîneurs qui ont montré une performance supérieure à celle de leurs pairs ont tendance à travailler pour un plus grand nombre de clubs.

Cela pourrait s'expliquer par le fait que les entraîneurs performants soient davantage reconnus et donc sollicités par d'autres clubs, favorisant ainsi leur mobilité entre les clubs.

#### 5.d. Ancienneté du coach sur la performance de l'équipe

Dans cette partie, nous cherchons une éventuelle relation entre l'expérience qu'un entraîneur a pu accumuler avec son ancienneté et les performances de l'équipe lorsqu'il est à sa tête. L'ancienneté pourrait bien être un facteur important au sein d'une équipe, car l'entraîneur gagne en expérience au fil des clubs par lesquels il est passé. Mais cette expérience a-t-elle une réelle incidence sur les résultats obtenus ?

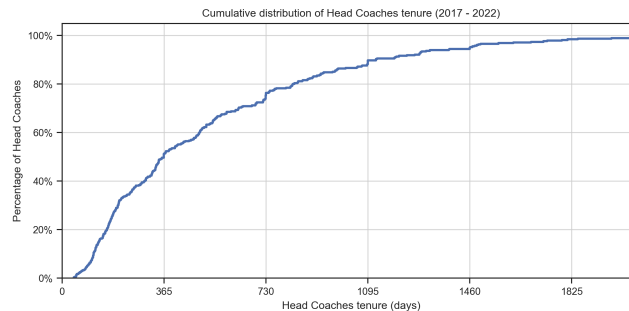


Figure 20: Empirical Cumulative Distribution Function of Head Coaches Tenure For Completed Appointments

La [Figure 20](#) montre la distribution de l'ancienneté des entraîneurs sportifs au sein de leur club. On observe que plus de 50 % des entraîneurs sportifs sont renouvelés après 1 an de mandat. Ce pourcentage augmente à 80 % après 2 ans de mandat et à 90 % après 3 ans de mandat.

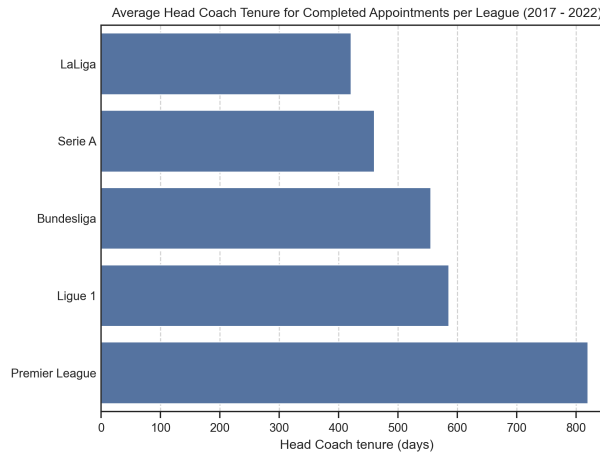


Figure 21: Average Head Coach Tenure for Completed Appointments per League

La Figure 21 s'intéresse à l'ancienneté des entraîneurs sportif et au renouvellement des entraîneurs sportifs par rapport aux ligues d'intérêt.

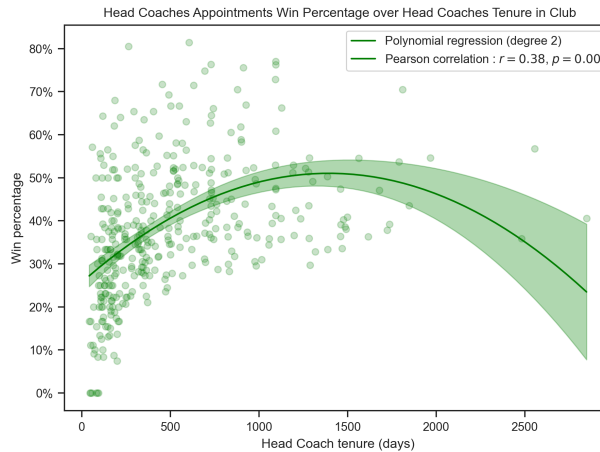


Figure 22: Win Ratio of Head Coaches Appointments versus Head Coach Tenure

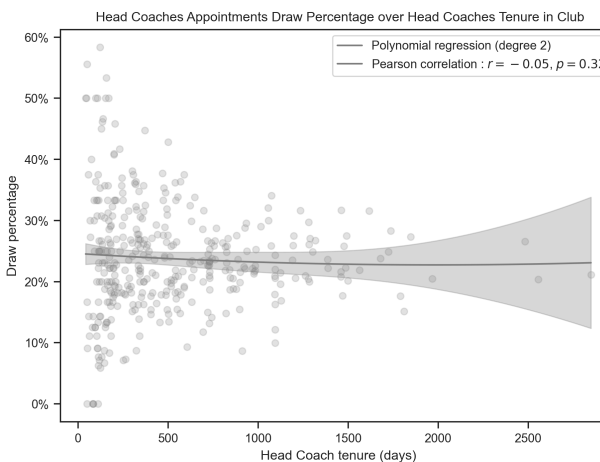


Figure 23: Draw Ratio of Head Coaches Appointments versus Head Coach Tenure

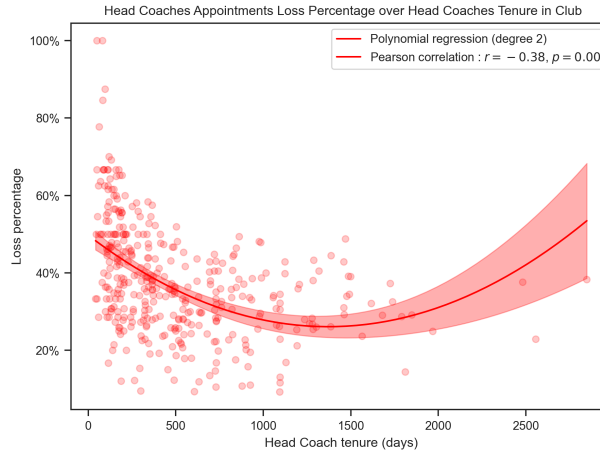


Figure 24: Loss Ratio of Head Coaches Appointments versus Head Coach Tenure

Les Figure 22 et Figure 23 et Figure 24 s'intéressent aux différents ratios de victoires, de matchs nuls et de défaites en fonction de la durée du coach au sein du club. Cette analyse permet de mettre en évidence le lien existant entre les résultats directs du coach et la durée de son mandat à la tête de l'équipe.

- Figure 22 une corrélation positive modérée ( $r = 0.38$ ) statistiquement significative ( $p = 0.00$ ) entre la durée du coach et le ratio de victoires.
- Figure 23 une corrélation négative faible ( $r = -0.06$ ) mais statistiquement-non significative ( $p = 0.25$ ) entre la durée du coach et le ratio de matchs nuls.
- Figure 24 une corrélation négative modérée ( $r = -0.37$ ) statistiquement significative ( $p = 0.00$ ) entre la durée du coach et le ratio de défaites.

Il est à noter que deux des trois corrélations sont statistiquement significatives, étant donné que les valeurs de  $p$  sont inférieures à 0.05, qui est notre seuil d'acceptation. Ainsi, selon le coefficient de corrélation de Pearson  $r$ , nous observons que lorsque qu'un entraîneur gagne ses matchs, il augmente en même temps sa durée à la tête de l'équipe. En revanche, pour les défaites, plus il y en a, moins longtemps il reste coach de l'équipe. Les résultats neutres, représentés par les matchs nuls, non pas de signification statistique au vu du coefficient  $p > 0.05$ .

#### 5.d.i. Lien plus fin entre ancienneté du coach et performance de l'équipe:

Cette partie est différente de la précédente, car ici, on associe les résultats de l'entraîneur à son ancienneté lors du match. Dans ce cas, nous avons un lien plus précis et statistiquement significatif entre les deux variables que lorsque l'on regarde la performance moyenne de l'entraîneur par rapport à son ancienneté dans un club.

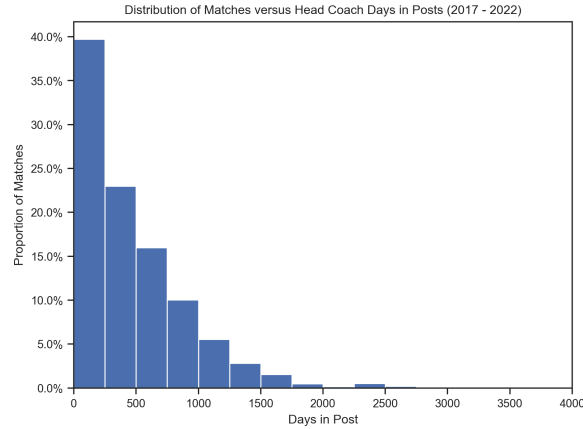


Figure 25: Distribution of Matches versus Head Coach Tenure on Match Day

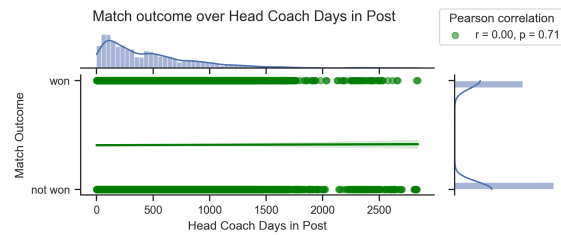


Figure 26: Match Win Outcome versus Head Coach Tenure on Match Day

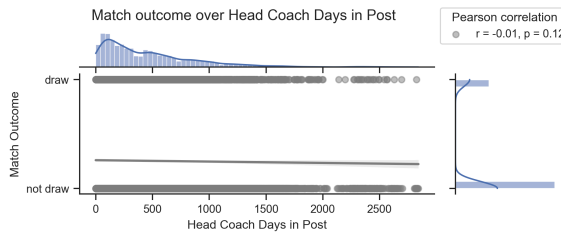


Figure 27: Match Draw Outcome versus Head Coach Tenure on Match Day

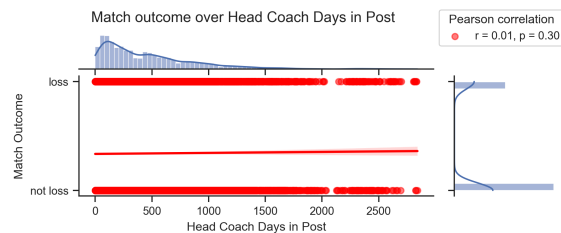


Figure 28: Match Loss Outcome versus Head Coach Tenure on Match Day

L'ancienneté, tout entraîneurs confondus a un effet positif sur la performance de l'équipe. Cela peut s'expliquer par le fait que les entraîneurs ont besoin de temps pour s'adapter à leur nouvel environnement et pour mettre en place leur stratégie de jeu. De plus, les entraîneurs qui restent plus longtemps à la tête de l'équipe ont tendance à mieux connaître les joueurs et à mieux comprendre les forces et les faiblesses de l'équipe, ce qui peut contribuer à améliorer les performances de l'équipe. Néanmoins, il est aussi probable que les équipes qui ont de bons résultats ont tendance à garder leurs entraîneurs plus longtemps, ce qui peut expliquer en partie la corrélation positive entre l'ancienneté du coach et la performance de l'équipe.



### 5.d.ii. Une visualisation graphique de l'effet de l'ancienneté du coach sur la performance de l'équipe:

Parmi l'ensemble des matchs où l'on possède des informations sur l'entraîneur sportif et où l'entraîneur sportif avait moins de 1500 jours d'ancienneté lors du match :

- le pourcentage de match gagné est de '40.92' %.
- le pourcentage de match nul est de '25.42' %.
- le pourcentage de match perdu est de '33.67' %.

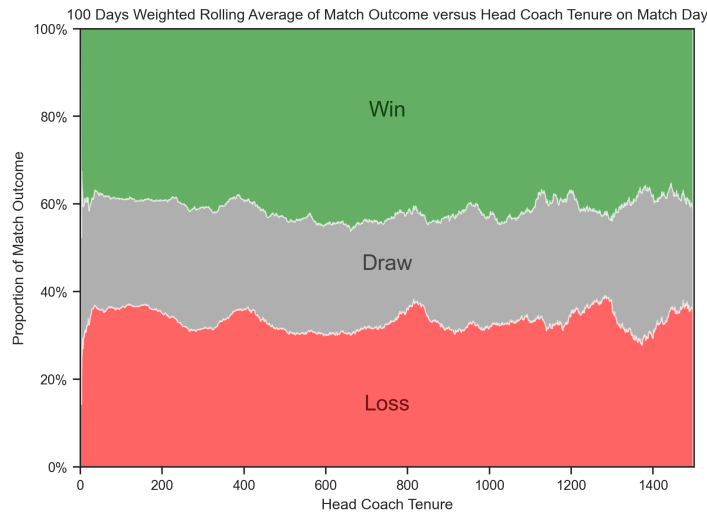


Figure 29: Weighted Rolling Average of Match Outcome versus Head Coach Tenure on Match Day

Le Figure 29 illustre la proportion des résultats des matchs en fonction de l'ancienneté de l'entraîneur principal lors du match. Cette proportion est calculée à l'aide d'une moyenne pondérée sur une période de 100 jours. La zone verte représente les victoires, la zone grise représente les matchs nuls et la zone rouge représente les défaites. On peut observer que les victoires tendent à augmenter avec l'ancienneté de l'entraîneur, tandis que les défaites ont tendance à diminuer.

```
import numpy as np

def weighted_rolling_mean(data, weights, window_size=30):
    def weighted_mean(x):
        return np.average(data.loc[x.index], weights=weights.loc[x.index])

    return data.rolling(window_size, min_periods=1).apply(weighted_mean,
raw=False)
```

Code 2: Calcul des moyennes mobiles pondérées

## 6. CONCLUSION

L'analyse des données a permis de mettre en évidence trois corrélations entre la performance d'une équipe et l'entraîneur sportif. Les résultats montrent que l'ancienneté du coach au sein de l'équipe est corrélée positivement avec la performance de l'équipe. En d'autres termes, plus un coach reste longtemps à la tête de l'équipe, meilleure sont les performances de l'équipe. De plus, le renouvellement régulier d'un coach sportif est corrélé négativement avec la

performance de l'équipe. Enfin, les entraîneurs qui changent régulièrement de club ont tendance à voir une amélioration de la performance de l'équipe.

Ces résultats suggèrent que la stabilité et la continuité sont des facteurs importants de la performance sportive d'une équipe. Les entraîneurs qui restent longtemps à la tête de l'équipe ont tendance à mieux connaître les joueurs et à mieux comprendre les forces et les faiblesses de l'équipe, ce qui peut contribuer à améliorer les performances de l'équipe. En revanche, les entraîneurs qui changent régulièrement de club ont tendance à voir une amélioration de la performance de l'équipe. Cela semble indiquer qu'un renouvellement régulier des entraîneurs peut être bénéfique pour l'équipe, mais que la stabilité et la continuité d'un coach peuvent également avoir un impact significatif sur les performances de l'équipe, en particulier à court terme.

Cependant, il est difficile de tirer des conclusions définitives sur la causalité de ces relations, car il existe de nombreux autres facteurs qui peuvent influencer la performance d'une équipe de football. Par exemple, la qualité des joueurs, la stratégie de jeu, la gestion du club et d'autres facteurs peuvent également jouer un rôle important dans la performance de l'équipe. Il est donc important de prendre en compte ces facteurs lors de l'analyse des données et de ne pas tirer de conclusions hâtives sur la relation entre le licenciement d'un coach et la performance de l'équipe.

## 7. APPRONFONDIR ET AMÉLIORER CETTE ÉTUDE

- Observer les causes et conséquences d'un licenciement en cours de saison sur la performance de l'équipe le reste de la saison
- Aggrandir la tailles des données d'intérêts (expérience passé de l'entraîneur pas représentative car limité à certains clubs de 1ère division)
- Ajouter les données manquantes des entraîneurs sportifs (proportion importante de matchs sans entraîneur)

## REFERENCES

- Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics. Doklady*, 10, 707–710. <https://api.semanticscholar.org/CorpusID:60827152>
- Wilke, C. O. (2019). *Fundamentals of Data Visualization*. O'Reilly Media. <https://clauswilke.com/dataviz/>