Projet Analyse de données

DU Jean-louis, RACHEDI Yakine, TAHER Wassim, BERKANE Sarah

Dans le cadre du projet de l'UE Analyse de données, nous allons réaliser une analyse statistique de la base de donnée *Baseball.csv* sous le langage **R**.

Table des matières

1 Analyse descriptive des données **2**
1.1 Présentation générale de la base de donnée . . . 2
1.2 Statistique descriptive générale de la base de donnée . . . 3
1.3 Analyse statistique . . . 3
1.3.1 Variable Salary . . . 3
1.3.2 Test des moyennes des salaires pour chaque groupe de League_1986 et League_1987 4
1.3.3 Variables League_1986 et League_1987 . . . 5
1.3.4 Analyse des salaires en fonction de League_1986 et League_1987 avec comparaison 6
1.3.5 Analyse des performances en 1986 . . . 7

2 Analyse en Composantes Principales (ACP) et questions soulevées **10**
2.1 Prétraitement des données . . . 10
2.2 Visualisation des corrélations . . . 10
2.3 Observations . . . 10
2.4 Questions soulevées . . . 10
2.4.1 Les performances de 1986 sont-elles particulières par rapport aux performances sur l'ensemble de la carrière ? . . . 11
2.4.2 Est-ce que la longévité de la carrière et le salaire perçu est lié ? . . . 12

1 Analyse descriptive des données

1.1 Présentation générale de la base de donnée

La base contient 322 lignes (joueurs) et 24 colonnes (variables). Voici la description des variables principales :

Variable	**Type de variables**	**Description**
Name	chr	Le nom du joueur.
Bat_times_86	int	Nombre de fois où le joueur est allé au bâton en 1986.
Hits_86	int	Nombre de coups sûrs réalisés par le joueur en 1986.
Home_runs_1986	int	Nombre de home runs (coups de circuit) réalisés en 1986.
Runs_1986	int	Nombre de points (runs) marqués en 1986.
Runs_batted_1986	int	Nombre de runs batted in (RBI) réalisés en 1986.
Walks_1986	int	Nombre de fois où le joueur a obtenu un walk en 1986.
Longevity	int	Nombre d'années dans les grandes ligues (carrière).
Bat_times_career	int	Nombre total de fois où le joueur est allé au bâton durant sa carrière.
Hits_career	int	Nombre total de coups sûrs réalisés durant la carrière.
Home_runs_career	int	Nombre total de home runs réalisés durant la carrière.
Runs_career	int	Nombre total de runs marqués durant la carrière.
Runs_batted_career	int	Nombre total de runs batted in (RBI) durant la carrière.
Walks_career	int	Nombre total de walks obtenus durant la carrière.
League_1986	chr	Ligue du joueur en 1986 (A : American League, N : National League).
Division_1986	chr	Division du joueur en 1986 (E : Est, W : Ouest).
Team_1986	chr	Équipe du joueur en 1986.
Position_1986	chr	Position occupée par le joueur en 1986.
Put_outs_1986	int	Nombre de put outs réalisés en 1986.
Assists_1986	int	Nombre d'assists réalisés en 1986.
Errors_1986	int	Nombre d'erreurs commises en 1986.
Salary_1987	num	Salaire du joueur pour l'année 1987 (en milliers de dollars).
League_1987	chr	Ligue du joueur en 1987 (A : American League, N : National League).
Team_1987	chr	Équipe du joueur en 1987.

Il y a ainsi 7 variables de type *chr* (chaînes de caractères),16 variables de type *int* (entier) et une variable de type *num* (numérique).

1.2 Statistique descriptive générale de la base de donnée

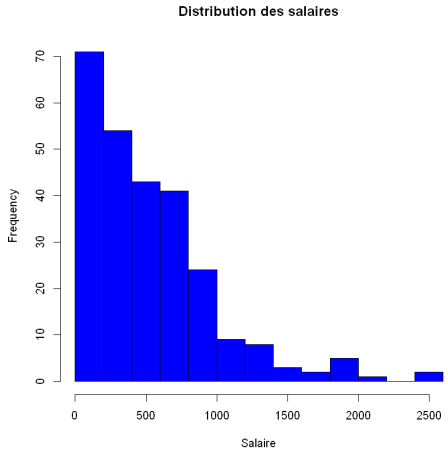
Nous avons réaliser une analyse descriptive générale de la base de donnée à l'aide de la fonction **summary** de **R** :

Variable	Min.	1er Quartile	Médiane	Moyenne	3e Quartile	Max.
Bat_times_86	127.0	272.0	390.5	390.1	512.8	687.0
Hits_86	31.0	68.0	98.5	103.4	137.8	238.0
Home_runs_1986	0.0	4.0	8.5	11.51	16.75	141.0
Runs_1986	12.0	32.0	48.0	52.22	70.0	130.0
Runs_batted_1986	8.0	29.0	45.0	49.37	65.0	121.0
Walks_1986	3.0	22.0	35.5	39.86	54.0	105.0
Longevity	1.0	4.0	6.0	7.67	11.0	24.0
Bat_times_career	166.0	911.2	2065.0	2763.1	4068.5	14053.0
Hits_career	34.0	227.2	552.0	747.7	1095.8	4256.0
Home_runs_career	0.0	16.0	40.0	74.09	95.5	548.0
Runs_career	18.0	106.2	266.0	374.3	556.5	2165.0
Runs_batted_career	9.0	98.25	250.0	347.61	461.0	1659.0
Walks_career	8.0	76.0	178.5	273.4	370.8	1566.0
Put_outs_1986	0.0	109.2	212.0	288.9	325.0	1378.0
Assists_1986	0.0	7.0	39.5	106.9	166.0	492.0
Errors_1986	0.0	3.0	6.0	8.04	11.0	32.0
Salary_1987	67.5	193.0	430.0	542.2	750.0	2460.0

Nous avons observé que certaines variables présentent des valeurs manquantes, notamment **Salary_1987**, avec 59 valeurs absentes (**NA**). Ces valeurs ont été exclues dans nos analyses statistiques à l'aide de la fonction **na.omit()**.

1.3 Analyse statistique

1.3.1 Variable Salary



Observations 1.1.

La distribution est asymétrique et très à droite. Cela indique qu'une majorité des salaires sont concentrés dans les valeurs basses, tandis que peu de joueurs ont des salaires très élevés.

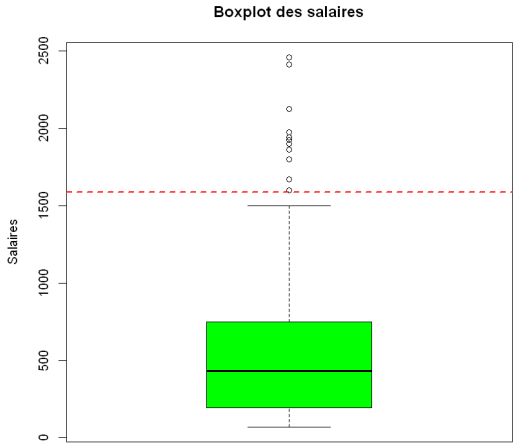
La majorité des joueurs ont des salaires compris entre 0 et 500 (on observe un pic dans cette plage).

Les fréquences diminuent progressivement lorsque les salaires augmentent.

Les barres isolées dans les salaires très élevés (> 2000) suggèrent qu'il y a des joueurs avec des salaires exceptionnellement hauts. Ces valeurs peuvent être des outliers ou refléter des différences dans les niveaux de performance ou de notoriété.

Une telle distribution asymétrique pourrait biaiser les statistiques comme la moyenne, qui sera tirée vers le haut par les valeurs extrêmes. La médiane serait un meilleur indicateur de la tendance centrale dans ce cas.

La médiane est la valeur centrale qui divise les données en deux parties égales, et dans une distribution asymétrique comme celle-ci, les valeurs très élevées (outliers) tirent la moyenne vers le haut, la rendant moins représentative de la tendance générale. La médiane, en revanche, n'est pas influencée par ces valeurs extrêmes. Elle reflète mieux la réalité de la majorité des données.



Boxplot des salaires

Statistique	Valeur
Min.	67.5
1st Qu.	193.0
Median	430.0
Mean	542.2
3rd Qu.	750.0
Max.	2460.0
Intervalle interquartile	557.0

Résumé des statistiques descriptives : Salary

Année	Ligue	Moyenne	Médiane
1886	A	543.44	420.0
	N	540.81	462.5
1987	A	538.53	400.0
	N	546.44	495.0

Moyenne des salaires :

- En 1986, la moyenne des salaires était légèrement plus élevée dans la ligue A (543.44) par rapport à la ligue N (540.81).
- En 1987, la tendance s'inverse légèrement, avec la ligue N ayant une moyenne plus élevée (546.44 contre 538.53 pour la ligue A).

Médiane des salaires :

- En 1986, la médiane des salaires dans la ligue A était de 420, tandis que dans la ligue N, elle était légèrement plus élevée à 462.5.
- En 1987, la ligue N conserve une médiane plus élevée (495 contre 400 pour la ligue A).

1.3.2 Test des moyennes des salaires pour chaque groupe de League_1986 et League_1987

Afin de confirmer nos hypothèses, nous allons efféctuer un test de Student des moyennes de la variable **Salary_1987** en fonction de **League_1987** et **League_1986** :
Pour chaque test de Student, nous définissons les hypothèses suivantes :

- **Hypothèse nulle** H_0 : Il n'y a **pas de différence significative** entre les moyennes des salaires des groupes de la variable League_1986 / 1987.

$$H_0 : \mu_A = \mu_N$$

où μ_A est la moyenne des salaires des joueurs appartenant à la ligue **A** et μ_N la moyenne des salaires des joueurs de la ligue **N**.

— **Hypothèse alternative** H_1 : Il existe une **différence significative** entre les moyennes des salaires des groupes.

$$H_1 : \mu_A \neq \mu_N$$

Nous appliquons un **test de Student avec variances inégales** (car **var.test** nous permet de rejeter, au risque d'erreur 5%, l'hypothèse d'égalité) afin de comparer les deux moyennes de salaires.
Nous obtenons ainisi :

— **Statistique t** : $t = -0.1423$ pour Salary_1987 et League_1987 et $t = 0.047354$ pour Salary_1987 et League_1986

— **p-valeur** : $p = 0.887$ pour Salary_1987 et League_1987 et $p = 0.9623$ pour Salary_1987 et League_1986

La **p-valeur** (dans les deux tests) est largement supérieure au seuil $\alpha = 0.05$. Par conséquent, nous **ne rejetons pas l'hypothèse nulle** H_0.
En d'autres termes, il n'y a **pas de différence significative** entre les salaires moyens des joueurs des ligues **A** et **N** en 1987, ce qui rejoint nos hypothèses de la section précédente.

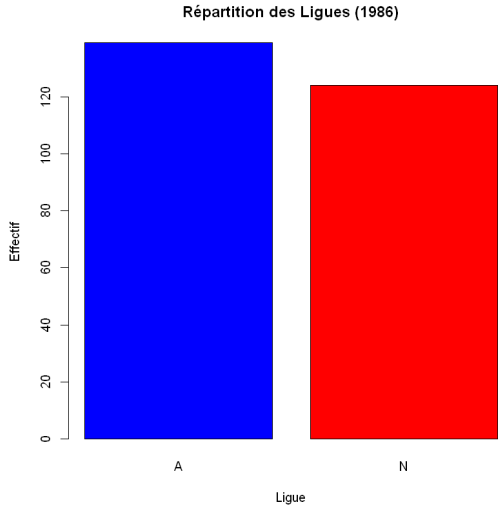
Évolution des salaires d'une année à l'autre :

Ligue A : La moyenne et la médiane des salaires en 1987 sont légèrement plus faibles qu'en 1986. Cela peut indiquer une diminution générale des salaires ou un ajustement dans les salaires des joueurs au sein de cette ligue.

Ligue N : Bien que la moyenne des salaires ait augmenté légèrement (de 540.81 à 546.44), la médiane a également augmenté de manière significative (de 462.5 à 495), ce qui suggère que l'augmentation des salaires dans la ligue N a touché une plus grande partie des joueurs, plutôt que seulement les joueurs les mieux rémunérés.

1.3.3 Variables League_1986 et League_1987

Ces variables sont catégorielles, car elles prennent les valeurs "A" pour représenter la ligue amateur et "N" pour représenter la ligue nationale.



Ligue	A	N
Comptage	139	124
Proportion	52.85 %	47.14%

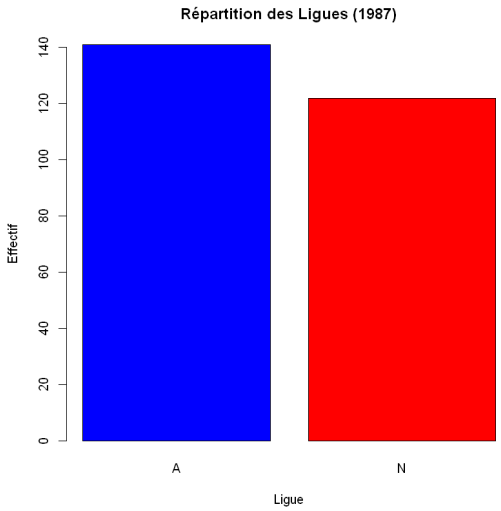
Effectifs :

— Ligue A : 139 joueurs (52.85%).

— Ligue N : 124 joueurs (47.15%).

Analyse : La différence est d'environ 5.7% (52.85% − 47.15%), ce qui reste relativement faible. Cela signifie qu'il y a une répartition presque égale entre les deux ligues, ce qui suggère un échantillon équilibré, même si la ligue A est légèrement plus représentée.

Cette répartition permet de comparer les ligues sans risque majeur de biais lié à une sous-représentation significative d'un groupe.



Ligue	A	N
Comptage	141	122
Proportion	53.61 %	46.38%

Pour la variable `League_1987`, on observe une situation similaire, avec une légère surreprésentation de la ligue A.

Effectifs :

— Ligue A : 141 joueurs (53.61%).

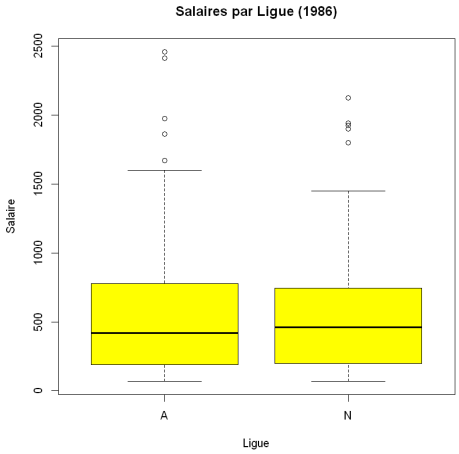
— Ligue N : 122 joueurs (46.39%).

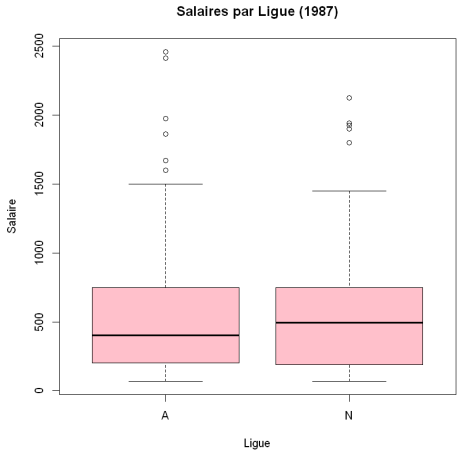
Analyse : La différence entre les deux ligues est d'environ 7.22% (53.61% − 46.39%), légèrement plus marquée qu'en 1986.

La ligue A reste plus représentée, mais l'écart est toujours modéré.

Les données restent relativement équilibrées, bien qu'un petit biais en faveur de la ligue A soit présent.

1.3.4 Analyse des salaires en fonction de League_1986 et League_1987 avec comparaison

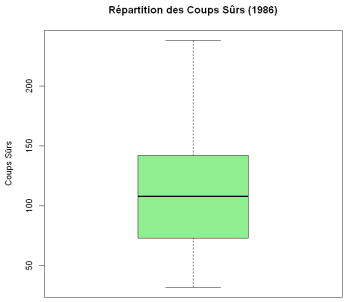




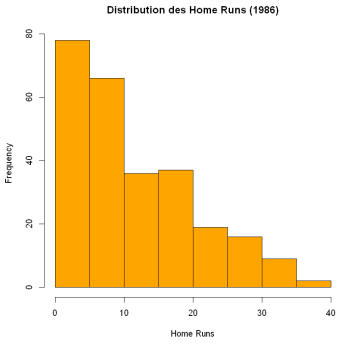
Les ligues de 1986 et 1987 présentent des distributions de salaires globalement similaires, que ce soit en termes de médianes ou de dispersion.

La médiane des salaires pour les deux ligues est d'environ 500 unités monétaires, reflétant une relative homogénéité entre les groupes. Cependant, des outliers sont visibles, avec quelques joueurs ayant des salaires exceptionnellement élevés (au-delà de 1500 et jusqu'à 2500 unités), probablement des stars ou des joueurs très expérimentés.

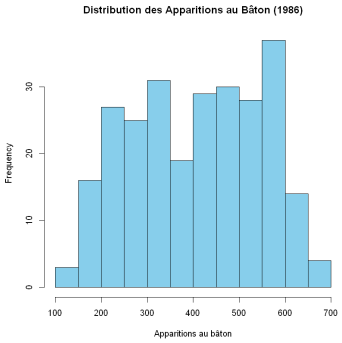
1.3.5 Analyse des performances en 1986



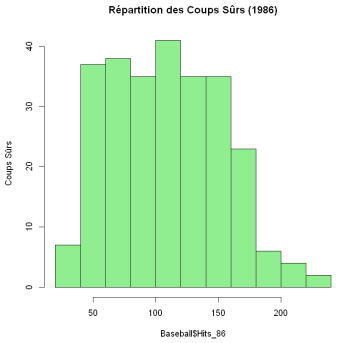
Boîte à moustaches des coups sûrs (Hits).



Distribution des Home Runs.



Histogramme des apparitions au bâton (Bat_times_86).



Histogramme des coups sûrs (Hits).

Analyse graphique des principales variables de performance en 1986.

Variable `Bat_times_86` (apparitions au bâton) L'histogramme (figure **??**) montre qu'une majorité de joueurs a enregistré entre **300 et 600 apparitions** au bâton, ce qui correspond aux joueurs ayant joué régulièrement durant la saison 1986. Les joueurs ayant moins de **200 apparitions** sont probablement des remplaçants ou des joueurs ayant bénéficié de moins d'opportunités de jeu.

Boîte à moustaches des coups sûrs La boîte à moustaches (figure **??**) illustre la répartition des coups sûrs pour l'année 1986. L'axe des ordonnées représente le **nombre de coups sûrs**. La boîte verte correspond à **l'intervalle interquartile (IQR)**, où le premier quartile (Q_1) est égal à **80** et le troisième quartile (Q_3) est égal à **150**. Ainsi, **50% des données** se situent dans cet intervalle. Cette représentation permet de visualiser la dispersion et la centralité des coups sûrs sur cette période, avec quelques valeurs aberrantes situées au-delà des extrémités des moustaches.

Distribution des Home Runs L'histogramme des Home Runs (figure **??**) montre la distribution des performances des joueurs en 1986. L'axe des abscisses représente le **nombre de Home Runs**, tandis que l'axe des ordonnées indique la **fréquence des observations**. On observe une concentration élevée des Home Runs dans les valeurs basses, suivie d'une diminution progressive pour les valeurs plus élevées. Cette distribution **asymétrique** indique que peu de joueurs ont réalisé un nombre important de Home Runs durant cette saison.

Histogramme des coups sûrs L'histogramme (figure **??**) représente la distribution des **coups sûrs (Hits)** pour l'année 1986. L'axe des abscisses montre le nombre de coups sûrs (`Baseball$Hits_86`), et l'axe des ordonnées indique la fréquence des joueurs. La distribution présente un **pic notable** autour de la plage **100 à 150 coups sûrs**, ce qui reflète une concentration des joueurs dans cette catégorie. Cette

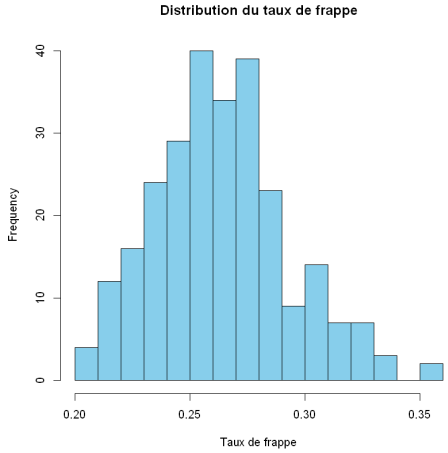
représentation met en évidence une répartition relativement uniforme des coups sûrs, avec une légère concentration dans les intervalles intermédiaires.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2000	0.2452	0.2632	0.2638	0.2798	0.3569

Statistiques descriptives de l'efficacité du joueur à frapper la balle.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
22.0	66.0	103.0	107.8	143.0	230.0

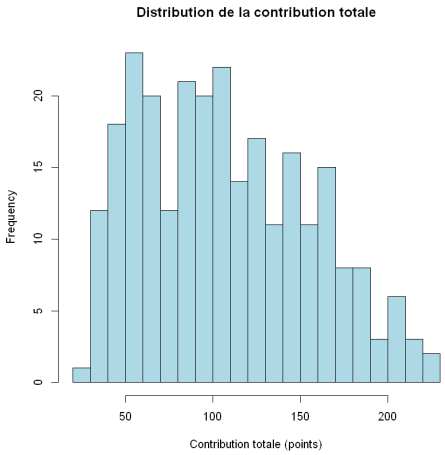
Statistiques descriptives des points contribués par le joueur (directement ou indirectement).

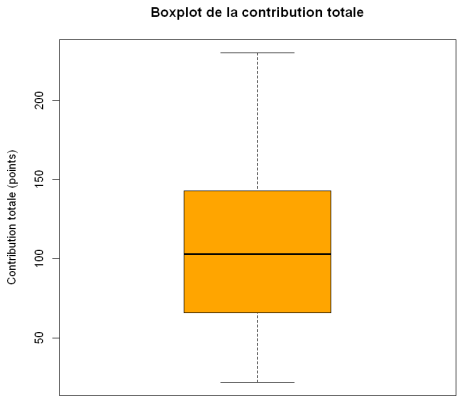


La courbe présente une forme proche de celle d'une distribution normale, symétrique autour d'une valeur centrale. Cette valeur centrale, qui correspond au taux de frappe le plus fréquent parmi les joueurs, se situe autour de 0.26 ou 0.27, représentant un taux de réussite moyen.

La plupart des joueurs ont un taux de frappe compris entre 0.24 et 0.29.

Très peu de joueurs affichent des taux de frappe extrêmement faibles (proches de 0.20) ou très élevés (proches de 0.35).





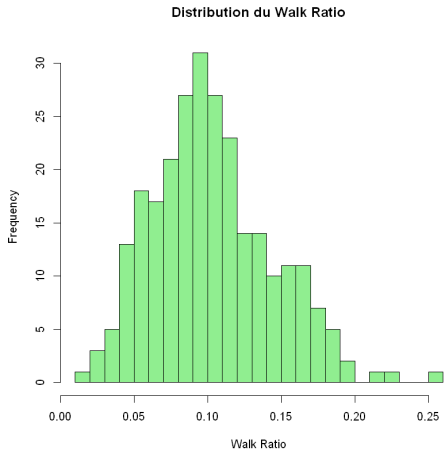
La plupart des joueurs ont contribué, de manière directe ou indirecte, à l'accumulation totale des points.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.01408	0.07564	0.09896	0.10300	0.12730	0.25581

Statistiques descriptives

$$\text{WalkRatio86} = \frac{\text{Walks86}}{\text{BatTimes86}} \tag{1}$$

Ce ratio représente la proportion de fois où un joueur a obtenu une base sur balles par rapport à ses apparitions au bâton.



La distribution des ratios semble légèrement asymétrique vers la droite, indiquant qu'un petit groupe de joueurs a des ratios plus élevés (au-dessus de 0.15), mais cela reste rare. La plupart des joueurs ont des ratios entre 0.05 et 0.15, ce qui est courant.

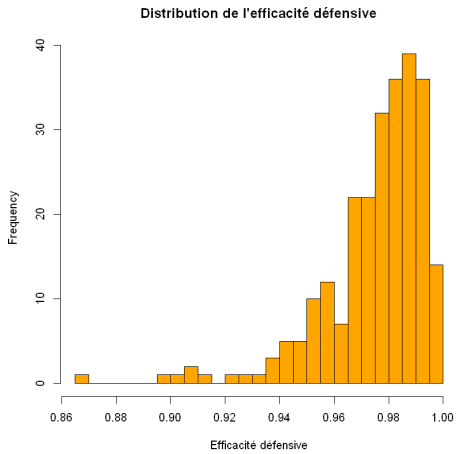
En résumé :

— Les joueurs avec un faible ratio (0 à 0.05) montrent une faible discipline au bâton ou un style agressif.

— Ceux avec un ratio moyen (0.05 à 0.15) ont une discipline moyenne et obtiennent des bases sur balles régulièrement.

— Les joueurs avec un ratio élevé (> 0.15) sont rares et font preuve d'une excellente discipline au bâton.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.8653	0.9682	0.9798	0.9752	0.9888	1.0000	11



La majorité des joueurs de la base de données présentent une efficacité défensive très élevée. Cela suggère soit un niveau défensif globalement élevé pour l'année 1986, soit que les joueurs analysés font partie d'une élite défensive.

2 Analyse en Composantes Principales (ACP) et questions soulevées

L'une des approches utiles face à une large base de données comme celle-ci est de réaliser une Analyse en Composantes Principales (ACP). L'ACP permet de projeter l'information sur 2 ou 3 composantes principales, ce qui facilite l'identification des corrélations possibles entre les variables.

2.1 Prétraitement des données

Avant de réaliser l'ACP, les données sont nettoyées en supprimant les joueurs avec des données manquantes. Une fois les données nettoyées, nous déterminons le nombre optimal de composantes principales à conserver en calculant le pourcentage d'inertie cumulée.

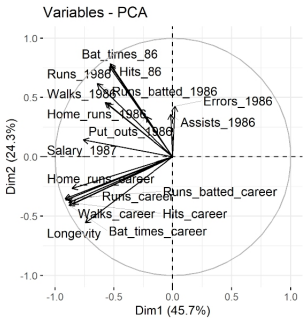
Ax1	**Ax1 :2**	**Ax1 :3**	**Ax1 :4**	**Ax1 :5**
45.67	69.99	80.14	85.43	89.59

Pourcentage cumulée d'inertie expliqué par les composantes principales

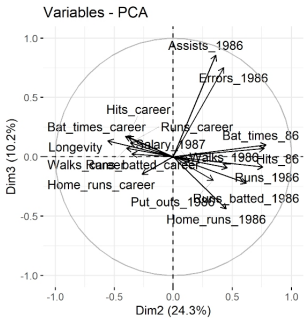
En se basant sur la règle des 80%, nous limitons l'analyse à 3 composantes principales.

2.2 Visualisation des corrélations

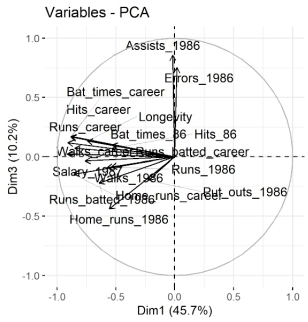
Une fois l'ACP réalisée à l'aide de l'outil disponible (`dudi.pca`), nous traçons les cercles de corrélations pour observer les relations entre les variables.



Corrélation dims 1 et 2



Corrélation dims 2 et 3



Corrélation dims 1 et 3

2.3 Observations

À partir des cercles de corrélations, plusieurs conclusions peuvent être tirées :
— Le **salaire** semble lié aux **performances des joueurs**.
— Les **erreurs** (*errors*) et les **assists** en 1986 ne semblent pas corrélés aux autres statistiques de performances.
— Les **performances globales** et les **performances spécifiques en 1986** sont **négativement corrélées**.
— La **longévité de la carrière** est liée aux **performances au cours de la carrière**.
— La **longévité** et les **performances de carrière** montrent une corrélation modérée.

2.4 Questions soulevées

Ces observations amènent à poser plusieurs questions pour une analyse approfondie :
— L'année 1986 est-elle particulière en termes de performance ?
— Le salaire est-il lié à la performance ?

— La longévité et le salaire d'un joueur sont-ils indépendants ?
— Comment les *assists* et *errors* sont-ils liés aux autres variables ?

2.4.1 Les performances de 1986 sont-elles particulières par rapport aux performances sur l'ensemble de la carrière ?

L'ACP nous avait révélés une corrélation négative entre les performances en 1986 et globales, ce qui nous a beaucoup surpris de prime abord. Après réflexion, comparer les variables globales et les variables de l'année 1986 n'est pas une bonne idée. En effet, les données calculés sur 1986 sont comprises dans celles globales, ce qui fait qu'elles sont forcément plus grandes. C'est peut-être cette relation d'ordre qui a été décelé par l'ACP. Pour analyser la relation entre les performances en 1986 et sur la carrière globale, on va donc modifier les performances globales en les moyennisant sur la longévité de chaque joueur. Malgré le fait que pour certains dont la carrière n'est pas très longue, cela reste une bonne approche pour analyser cela. En rassemblant les mêmes caractéristiques de performance globales et en 1986, on peut faire des tableaux de corrélation. On observe ces tableaux là.

	Bat_times_86	Bat_times_career
Bat_times_86	1	0.19
Bat_times_career	0.19	1

Corrélation du Bat_times

	Home_runs_86	Home_runs_career
Home_runs_86	1	0.51
Home_runs_career	0.51	1

Corrélation des Home_runs

Les corrélations qu'on obtient sont toutes comprises entre 0.19 et 0.51, contrairement à ce que nous révèle l'ACP, ce n'est pas négatif. Mais la corrélation reste limité, les valeurs sont relativement peu élevées. On ne peut pas dire que les performances de 1986 sont représentatifs du reste des années.

Par contre en faisant un tableau de corrélation sur les différentes caractéristiques de performances sur une même année, on voit qu'elles sont toutes corrélées positivement.

	Bat_times_86	Hits_86	Home_runs_86	Runs_86	Runs_batted_86	Walks_86
Bat_times_86	1	0.96	0.55	0.89	0.78	0.61
Hits_86	0.96	1	0.52	0.90	0.78	0.57
Home_runs_86	0.55	0.52	1	0.63	0.85	0.44
Runs_86	0.89	0.90	0.63	1	0.77	0.69
Runs_batted_86	0.78	0.78	0.85	0.77	1	0.56
Walks_86	0.61	0.57	0.44	0.69	0.56	1

Tableau de corrélation pour l'année 86

On observe la même chose sur les variables sur l'ensemble de la carrière. On voit que Hits et Bat_times sont et que Hits et Runs sont très corrélées par exemple. (S'explique par les règles du baseball peut-être)

Observations 2.1.

Graphique (Home_runs_1986 vs Salary_1987) :

La relation entre les "Home Runs" réalisés en 1986 et le salaire en 1987 semble être positive mais faible. La ligne de régression montre une légère tendance croissante. Cependant, il y a beaucoup de dispersion autour de la ligne, ce qui signifie que le nombre de home runs n'explique pas complètement le salaire. Tableau de corrélation :

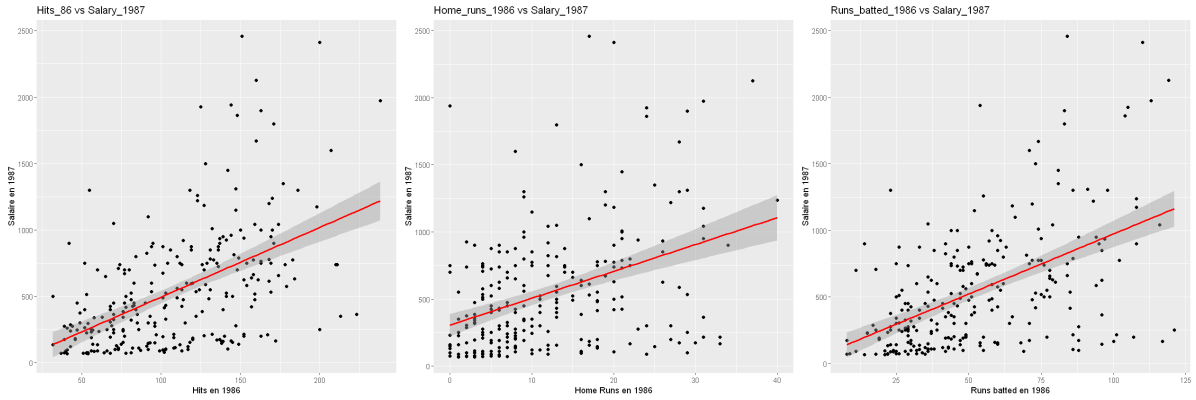
La corrélation entre Home_runs_1986 et Salary_1987 est de 0.39 (positive mais relativement faible).

En revanche, Hits_86 et Runs_batted_1986 montrent des corrélations plus fortes avec $\text{Salary}_1987(0.51et0.52respective$

Pourquoi autant de points éloignés de la ligne de régression ? Cela peut s'expliquer par :

	Bat_times_86	Hits_86	Home_runs_86	Runs_86	Runs_batted_86	Walks_86
Salary_1987	0.47	0.51	0.39	0.49	0.52	0.50

Corrélation performances/salaire



Corrélations entre les performances des joueurs en 1986 et leurs salaires en 1987

Autres facteurs : Le salaire dépend probablement d'autres variables importantes (popularité, ancienneté, position, etc.). Variabilité naturelle : Certains joueurs peuvent avoir des salaires élevés pour des raisons indépendantes de leurs statistiques sportives. Relation faible : La corrélation de 0.39 montre qu'il y a une certaine relation, mais pas assez forte pour expliquer la majorité des salaires.

2.4.2 Est-ce que la longévité de la carrière et le salaire perçu est lié ?

Une autre question soulevé par l'ACP est celle de la relation entre le salaire et la longévité. Intuitivement, on se dirait que plus la carrière est longue, plus le salaire devrait être conséquent. De plus, si la carrière d'un joueur dure, c'est qu'il est performant, cela motive encore plus notre intuition. Pour commencer, on peut afficher la distribution des salaires et des longévités.

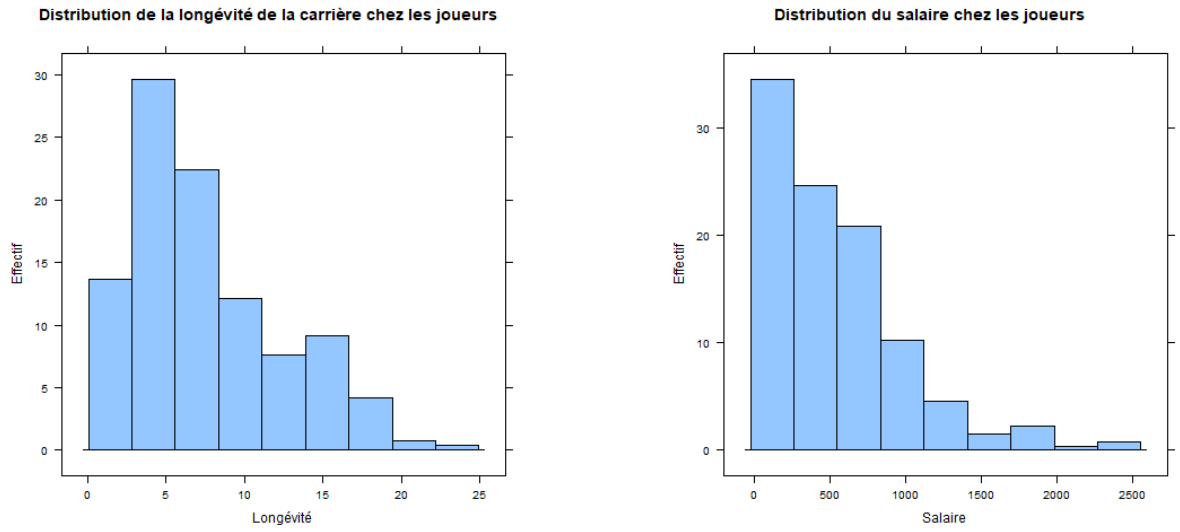
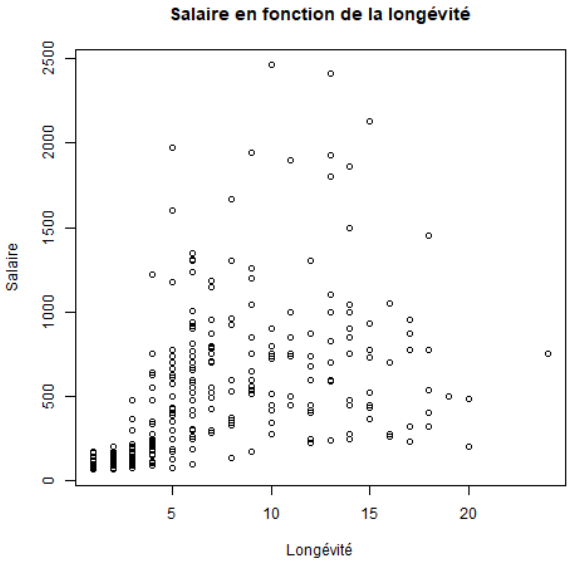
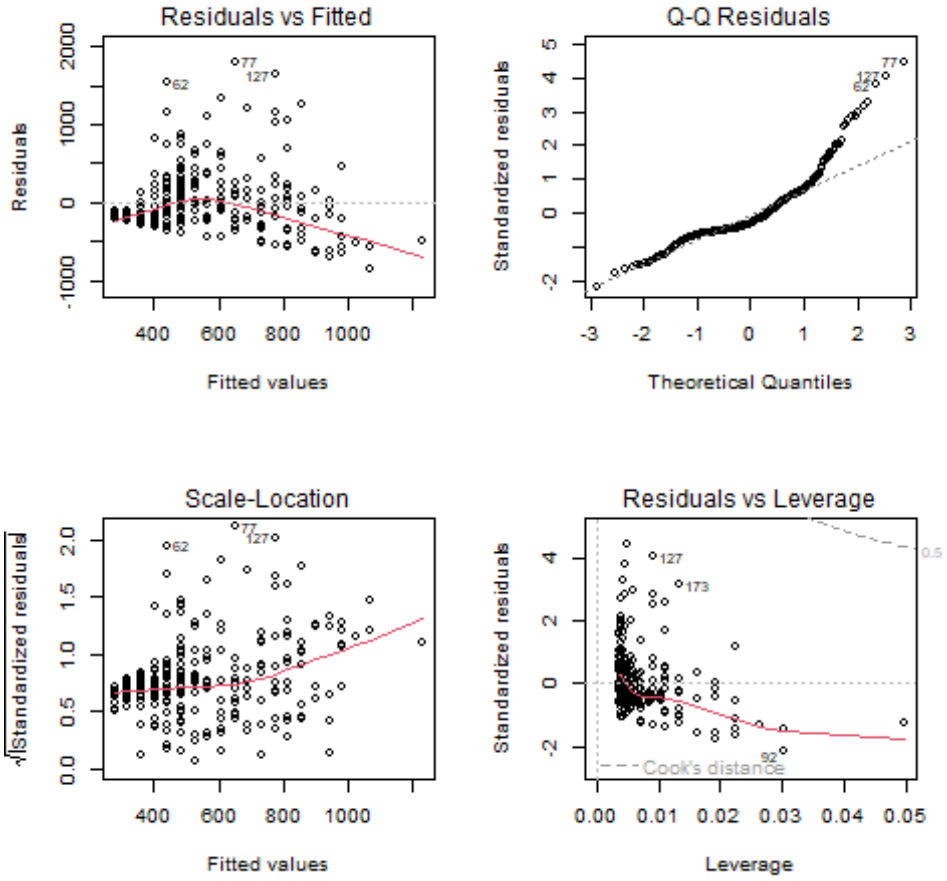


FIGURE 2 – Distribution du salaire et de la longévité

Pour analyser cela, on peut commencer par tracer le salaire en fonction de la longévité.



On remarque que cela n'a pas l'air linéaire du tout. De plus, on a des effectifs très différents selon la longévité. Faire une régression linéaire ne semble pas être une bonne idée. En utilisant le module de régression linéaire, on trouve que les hypothèses ne sont pas vérifié.



Vérification des hypothèses de régression linéaire

On remarque que l'espérance de la distribution n'est pas 0 sur le premier graphique, cela nous confirme

que les espérances conditionnelles ne suivent pas une relation linéaire. Par le deuxième graphe on voit que la normalité résiduelle n'est pas vérifié. Enfin par le troisième graphe, on voit que l'homosédasticité n'est pas vérifié, les écarts de part et d'autre de la ligne rouge ne semble pas du tout régulie.Pour vérifier la relation entre le salaire et la longévité, on peut tenter une ANOVA, avec pour hypothèse H0 que pour toute longévité on obtient le même salaire en moyenne. On a que les effectifs de chaque groupe (nombres d'années de carrière) sont inférieures à 30. Ainsi on ne peut pas faire une ANOVA classique, mais on peut tenter une Anova de Welsh.

Pour cela on a dû rassembler certaines modalités ensemble (16-17, 18-19, 20+).

L'anova nous affiche les résultats suivants :

$$F = 19.613,\ numdf = 17.000,\ denomdf = 55.493,\ p - value < 2.2e - 16$$

Finalement on trouve une p-value de 2.2e-16, l'hypothèse H0 disant que la moyenne de salaire ne dépend pas de la longévité n'est pas vérifié.