

# Etude de l'impact de l'urbanisation sur les oiseaux

## Analyse des Données Multivariées

*Rapport de projet  
Effectué dans le cadre du parcours  
Master 1 Bioinformatique et Biologie des Systèmes  
Année universitaire 2023-2024*



Image générée par IA générative<sup>(1)</sup>

## ABRÉVIATIONS

### Espèces étudiées

- ❖ **Cir** : Trembleur brun (*Cinclocerthia ruficauda*)
- ❖ **Cof** : Sucrier à ventre jaune (*Coereba flaveola*)
- ❖ **Col** : Moucherolle gobemouche (*Contopus latirostris*)
- ❖ **Dep** : Paruline caféiette (*Dendroica plumbea*)
- ❖ **Euh** : Colibri falle-vert (*Eulampis holosericeus*)
- ❖ **Euj** : Colibri madère (*Eulampis jugularis*)
- ❖ **Elm** : Élénie siffleuse (*Elaenia martinica*)
- ❖ **Gem** : Colombe à croissants (*Geotrygon mystacea*)
- ❖ **Lon** : Sporophile rouge-gorge (*Loxigilla noctis*)
- ❖ **Maf** : Moqueur corossol (*Margarops fuscatus*)
- ❖ **Mas** : Moqueur grivotte (*Margarops fuscus*)
- ❖ **Meh** : Pic de la Guadeloupe (*Melanerpes herminieri*)
- ❖ **Myo** : Tyran janeau (*Myiarchus oberi*)
- ❖ **Orc** : Colibri huppe (*Orthorhyncus cristatus*)
- ❖ **Saa** : Saltator gros-bec (*Saltator albicollis*)
- ❖ **Tul** : Grive à pieds jaunes (*Turdus lherminieri*)

### Caractéristiques de l'environnement

- ❖ **IdPE** : identifiant de la station d'écoute
- ❖ **UE** : unité écologique (adapté d'après Rousteau 1996)
- ❖ **agri** : surfaces agricoles totales (densité)
- ❖ **alt** : altitude (en mètres)
- ❖ **arbu** : strates arbustives et herbacées (densité)
- ❖ **bati** : urbanisation, bâtiment (densité)
- ❖ **ecot** : écotone (en mètres par hectare)
- ❖ **fcoeur** : cœur de forêt (densité de forêt située à plus de 250 m de la lisière)
- ❖ **foret** : forêt (densité)
- ❖ **heter** : hétérogénéité de l'habitat (indice de Shannon)
- ❖ **hydro** : réseau hydrographique (en mètres par hectare)
- ❖ **lat** : latitude
- ❖ **lisi** : lisière de forêts (en mètres par hectare)
- ❖ **loc** : localisation géographique (Basse-Terre ou Grande-Terre)
- ❖ **long** : longitude
- ❖ **pente** : pente (en degrés)
- ❖ **prai** : prairies (densité)
- ❖ **prec** : précipitations (en millimètres)
- ❖ **temp** : température (en degrés Celsius)

## RÉSUMÉ

La croissance rapide et incontrôlée des villes empiète sur les territoires anciennement sauvages. En effet, l'urbanisation grignote de plus en plus de terrains sur les anciens habitats de la faune autochtone. Ces changements sont la cause d'extinction et menacent de nombreuses espèces dans le monde.

Nous allons travailler sur des données provenant d'une étude menée en Guadeloupe et récoltées par diverses sources (Parc National de Guadeloupe, ONCFS, bureau d'études et associations). Cette étude s'intéresse aux communautés d'oiseaux endémiques des Antilles sur le territoire guadeloupéen.

Lors de cette étude, nous allons effectuer des analyses de couplage de tableaux entre trois jeux de données. "*habitat*" composé de 19 variables, lui-même composé de 16 variables quantitatives, 2 qualitatives et 1 compteur. Pour notre étude, nous avons exclu les variables qualitatives de ce jeu de données. Nous avons aussi le jeu de données "*espèces*" composé de 8 variables pour décrire les 16 espèces d'oiseaux : 7 qualitatives et 1 quantitative. Enfin, nous avons "*communaute*", ce dernier est un tableau de comptage de la localisation des 16 espèces dans les divers points d'écoutes de "*habitat*".

Lors de cette étude, nous allons essayer de trouver si l'urbanisation a un impact sur les caractéristiques des 16 espèces, en particulier sur leur taille. Pour ce faire, nous allons utiliser des méthodes de couplage de tableaux, des analyses en composantes principales, des analyses en correspondances multiples et des analyses factorielles discriminantes.

Au fil de nos analyses, nous avons pu montrer une tendance à l'impact des divers facteurs liés à l'habitat des oiseaux dont les aménagements humains tels que l'urbanisation et les terres agricoles sur la taille des oiseaux. Néanmoins, il est nécessaire de pousser plus profondément les observations liées à l'impact de l'Homme pour s'en convaincre.

Points clés :

- Identification et classification des espèces d'oiseaux étudiées.
- Analyse des paramètres environnementaux influençant leur répartition.
- Impact des activités humaines et des modifications paysagères sur la biodiversité aviaire.
- Implications pour la conservation et la gestion des écosystèmes.
- Influence modérée de l'urbanisation sur la répartition des oiseaux endémiques.
- Importance du milieu forestier dans la préservation des espèces sensibles.
- Faible structuration des communautés aviaires par les variables environnementales seules.
- Aucune cohabitation préférentielle ou corrélation alimentaire détectée.
- Légère tendance entre taille des espèces et urbanisation, non significative.
- Pistes complémentaires : analyses temporelles, modélisation de niches, prise en compte de pollutions urbaines.

Mots clés : biodiversité, oiseaux, Guadeloupe, écologie, habitat, conservation, environnement, répartition des espèces, statistiques, analyses multivariées

## INTRODUCTION

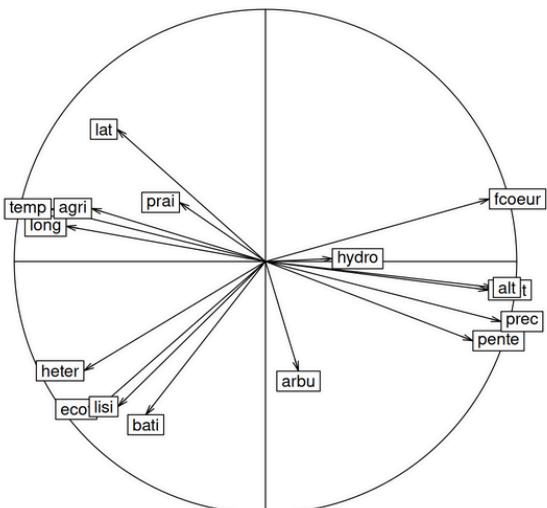
Entre 1999 et 2010, la surface du territoire urbain en France s'est accrue de 19 %, passant de 100 000 à 119 000 km<sup>2</sup>. Désormais, **21,8 % du territoire métropolitain est urbain**<sup>(2)</sup>. Qu'en est-il alors de l'**impact de cette croissance rapide sur la faune sauvage** ?

Nous allons dans cette étude nous concentrer sur la répartition mais aussi les caractéristiques que présentent les espèces d'oiseaux endémiques guadeloupéens (Annexe 1) en fonction du taux d'urbanisation du milieu. Les données ont été collectées par divers organismes (Parc National de Guadeloupe, ONCFS, bureaux d'études, associations). Nous utiliserons trois jeux de données pour cette analyse. Le fichier "*communaute.txt*" recense la présence-absence de **16 espèces** d'oiseaux sur **914 points d'écoute** sur une période de **3 ans**, entre 2009 et 2011. Le fichier "*habitat.txt*" décrit les variables environnementales liées à la topographie, au climat, à l'occupation du sol et à l'urbanisation. Le fichier "*especes.txt*" compile les informations taxonomiques et écologiques des espèces étudiées.

L'étude sera réalisée via diverses analyses à l'aide du langage R qui possèdent de nombreux outils statistiques sous forme de packages. Nous utiliserons les packages *ade4*<sup>(3)</sup>, *vegan*<sup>(4)</sup>, *lmttest*<sup>(5)</sup>, *MASS*<sup>(6)</sup> et *tidyverse*<sup>(7)</sup>. En effet, ces derniers nous permettront d'effectuer des analyses en composantes principales, des analyses factorielles des correspondances, des analyses des correspondances multiples, mais aussi des couplages de tableaux et d'analyses de co-inertie. Les données issues du dataset "*habitat*" possèdent de nombreuses unités différentes (Annexe 2), pour répondre à cette problématique, deux normalisations seront évaluées : la transformation log+1 et la normalisation à la moyenne.

## LES COMPORTEMENTS DIFFÈRENT-ILS SELON LA QUALITÉ DE VIE DU MILIEU ?

Nous effectuerons un test du  $\chi^2$  d'indépendance entre présence de l'espèce et caractéristiques du milieu. Nous nous concentrerons ensuite sur les variables relatives à l'occupation du sol (prairie, champs agricole, ville...). Seront réalisées : une analyse en composantes principales et une analyse factorielle des correspondances respectivement sur le jeu de données "*habitat*" et "*communaute*". Par la suite, un couplage de tableaux, une analyse canonique des correspondances et une co-inertie seront effectués à partir des analyses précédentes.



Le test du  $\chi^2$  d'indépendance nous montre une corrélation entre la présence d'une espèce et la caractéristique ciblée du milieu. Nous avons pu constater que les modalités *hydro*, *foret*, *lisi*, *pente*, *alt* et *prec* n'avaient aucune influence sur la présence des espèces. Cependant, l'unité écologique, la température, la longitude et la latitude semblent avoir une influence sur la répartition des espèces. De même que l'occupation au sol avec la présence de forêt profonde dense, de surfaces arbustives et herbacées, et du taux d'urbanisation (Annexe 3).

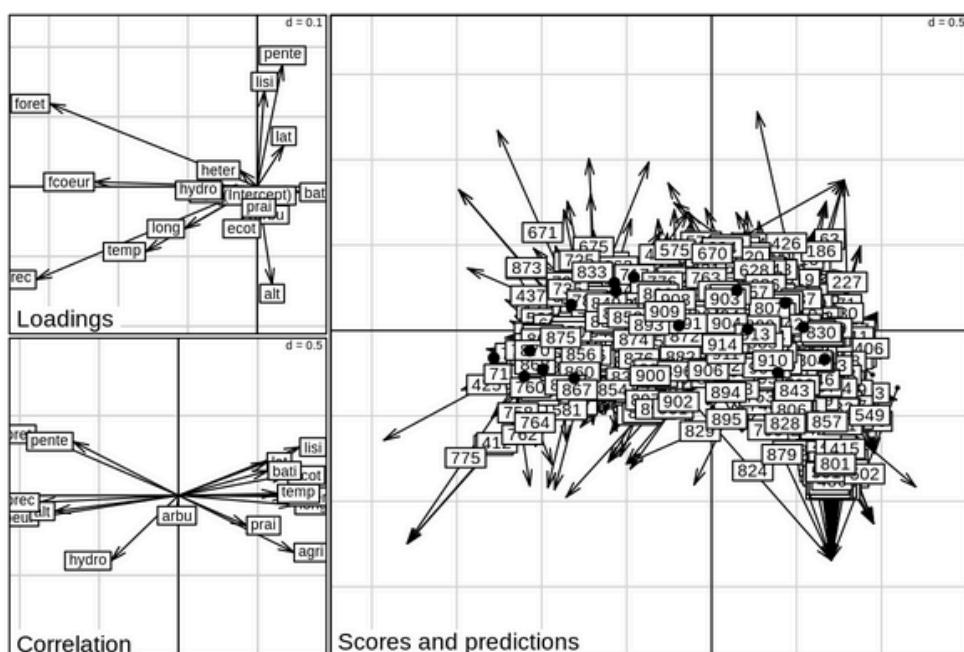
Suite à l'analyse précédente, nous avons pu observer qu'une partie seulement du jeu de données était susceptible de nous intéresser pour la suite. Nous allons donc réaliser une ACP normée,

pour réduire la dimensionnalité des données sur l'environnement et synthétiser l'information. La normalisation est nécessaire puisque chacune des modalités du jeu de données "habitat" n'ont pas les mêmes ordres de grandeur. A l'aide du cercle de corrélation (Fig.1), nous pouvons observer la participation des différentes modalités sur les composantes principales. La majorité des modalités s'expliquent par les deux premières composantes principales. Plus particulièrement, les modalités semblent se regrouper en quatre groupes, on peut donc résumer le nombre des modalités à quatre pour l'analyse. Pour expliquer l'impact de l'occupation du sol sur la répartition des espèces aviaires, nous choisissons les modalités : *agri*, *bati*, *fcoeur*, *foret* qui réfèrent respectivement à la densité agricole, au taux d'urbanisation, à la densité de la forêt et de la forêt profonde.

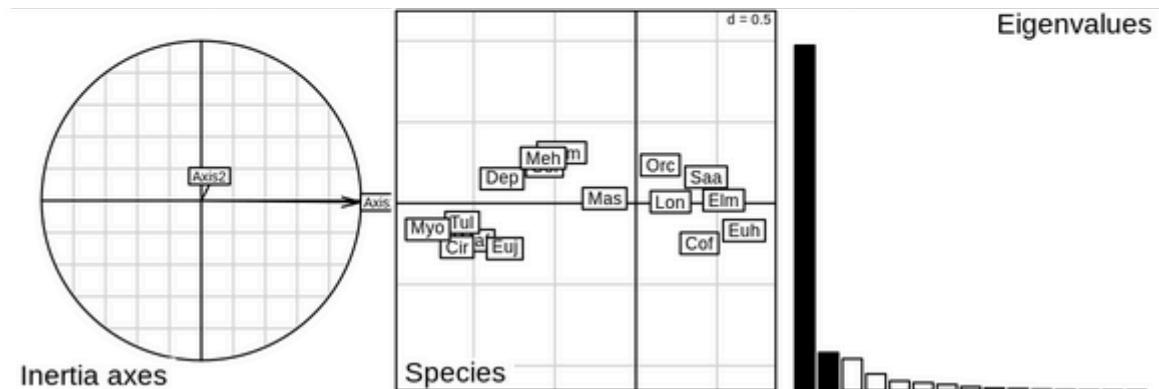
Nous allons effectuer une analyse factorielle des correspondances sur le dataset "communaute" (Annexe 5), ce qui nous permet d'analyser au mieux ce tableau de contingence représentant la présence-absence de chacune des espèces sur les lieux d'écoute. Dans cette dernière, nous observons que les familles d'animaux sont très peu expliquées par les 2 premiers axes à l'exception des Euh (Colibri falle-vert), ces derniers sont majoritairement dans les sites : 401, 801 et 199. Hormis cela, cette analyse nous indique qu'il y a peu de corrélation et/ou peu de variance entre les espèces.

Pour approfondir nos recherches, allons réaliser un couplage de tableaux, une analyse canonique des correspondances et une analyse de co-inertie. Nous avons donc réalisé une analyse canonique des correspondances entre "especes" et "habitat".

Comme nous pouvons l'observer sur la figure 2a, nous obtenons les prévisions grâce au dataset "habitat" normalisé avec la méthode du log+1. Le graphique *Loadings* nous éclaire sur le pouvoir explicatif de chaque modalité sur chaque axe. Tandis que le graphique *Correlation* montre la projection des modalités sur l'axe. Nous avons, par le graphique *Scores and predictions*, les étiquettes des lieux, des flèches représentant les distances de la projection des données provenant du dataset "especes" et de la prédiction des positions par régression sur les variables de milieu.

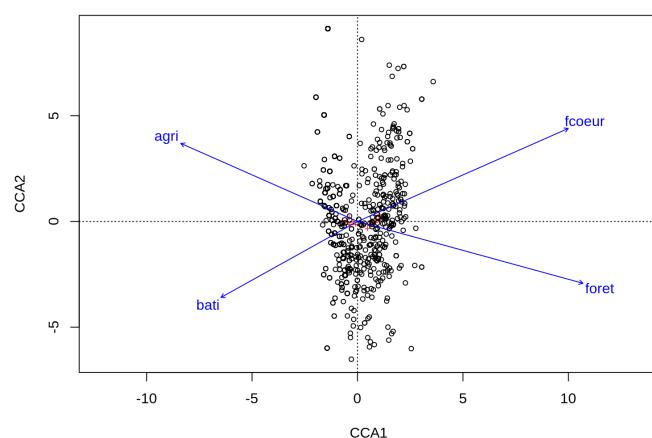


**Figure 2a : Projection vectorielle des modalités.** Les données analysées sont extraites de l'analyse canonique des correspondances avec les résultats de l'analyse en composantes principales du dataset "habitat" (normalisation avec le log+1) et de l'analyse factorielle des correspondances du dataset "communaute".



**Figure 2b : Cercle de corrélation, position des espèces à la moyenne des sites qui les contiennent et éboulis des valeurs propres.** Les données analysées sont extraites de l'analyse canonique des correspondances avec les résultats de l'analyse en composantes principales du dataset “habitat” (normalisation avec le log+1) et de l'analyse factorielle des correspondances du dataset “communaute”.

A l'aide de la figure 2b, nous observons la partie expliquée pour le dataset “especes”, les espèces sont très fortement séparées par l'axe 1 mais très peu par le second axe. Nous avons donc séparé les espèces en fonction de leur habitat.



**Figure 3 : Projection expliquant la répartition des espèces avec les conditions occupations du sol et notamment l'urbanisation.** Les données analysées sont extraites de l'analyse canonique des correspondances avec les résultats de l'analyse en composantes principales du dataset “habitat” (normalisation à la moyenne) et de l'analyse factorielle des correspondances du dataset “communaute”.

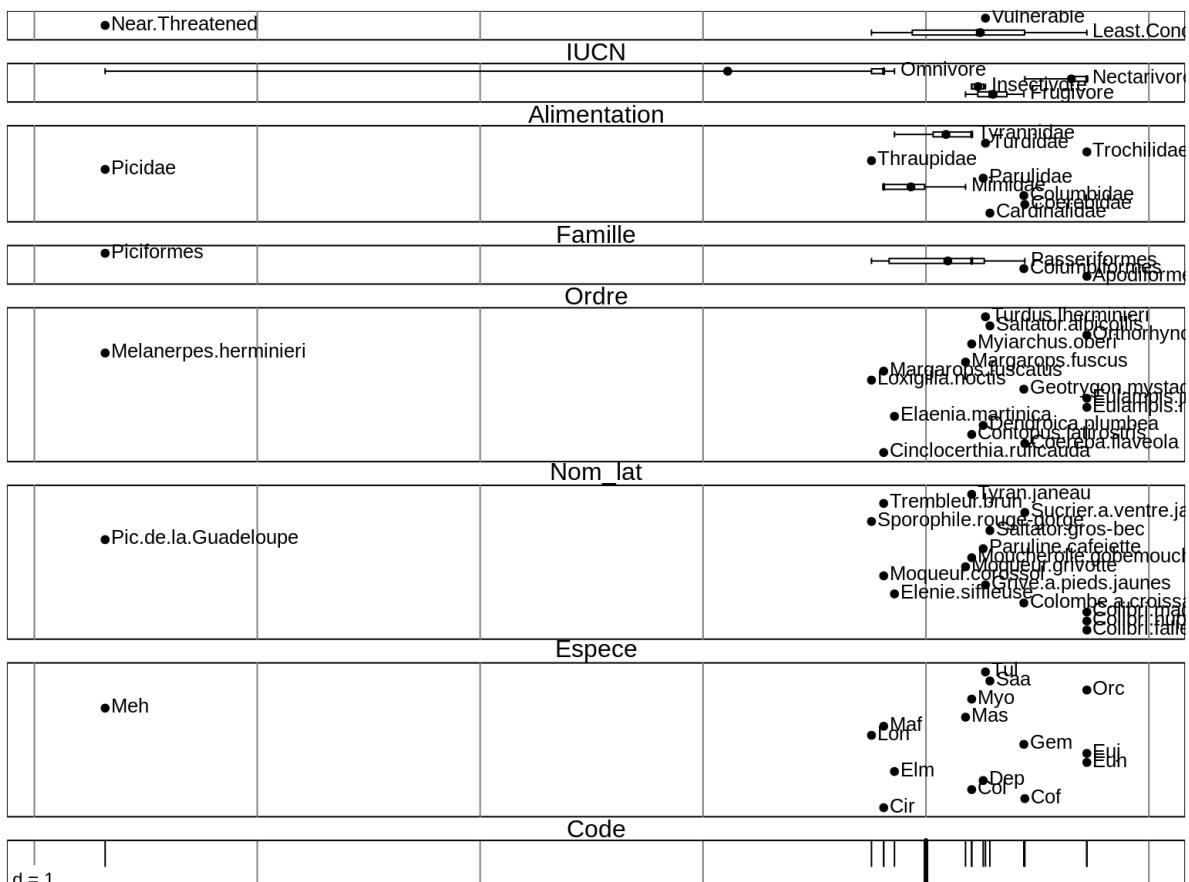
Enfin, nous avons fait une analyse canonique des correspondances entre les données issues de “communaute” et “habitat” avec une normalisation à la moyenne (Annexe 6). Nous avons par la suite, diminué le nombre de modalités de “habitat” pour l'analyse à l'aide de l'analyse en composantes principales précédentes (Fig.3). En effet, elles sont représentatives des groupes et correspondent à l'occupation du sol. Ce graphique révèle une structuration faible des espèces en fonction de l'occupation du sol. Nous avons nos quatre modalités, orthogonales deux à deux, expliquant très fortement les positions des espèces par la projection sur les 2 axes. L'occupation du sol influence faiblement la répartition des espèces aux sites d'écoute, avec un gradient d'opposition, avec d'un côté, l'urbanisation et l'agriculture, et de l'autre, les milieux forestiers. L'urbanisation semble apparaître comme un facteur essentiel mais non déterminant dans la distribution des communautés d'oiseaux.

Une analyse de la répartition des oiseaux en fonction de celles des autres espèces a été réalisée (Annexe 7). On peut ainsi vérifier si certains cohabitent ensemble et même plus, si elles semblent préférer la compagnie d'autres congénères. Après étude de la classification phylogénétique (Annexe 8a, 8b, 8c) mais aussi des comportements alimentaires, nous n'observons pas de rapprochement significatif entre espèces et donc pas de comportements particuliers selon le milieu.

*Plus particulièrement, les oiseaux protégés sont-ils davantage présents dans les milieux urbains ou les évitent-ils ?*

Pour la suite de l'analyse, nous allons étudier la présence des espèces protégées sur les lieux à forte densité urbaine. En effet, il est pertinent de se demander si les espèces dans les zones urbaines sont davantage protégées par l'Homme de par l'abondance de nourriture (issue des détritus humains), l'absence de prédateurs naturels et la présence de nombreux habitats en hauteur et sécuritaires. A contrario, il est également envisageable que les zones urbaines entraînent un plus fort taux de mortalité de par la forte densité d'animaux domestiques, de véhicules, de pollution et de par l'absence d'alimentation et d'habitats naturels.

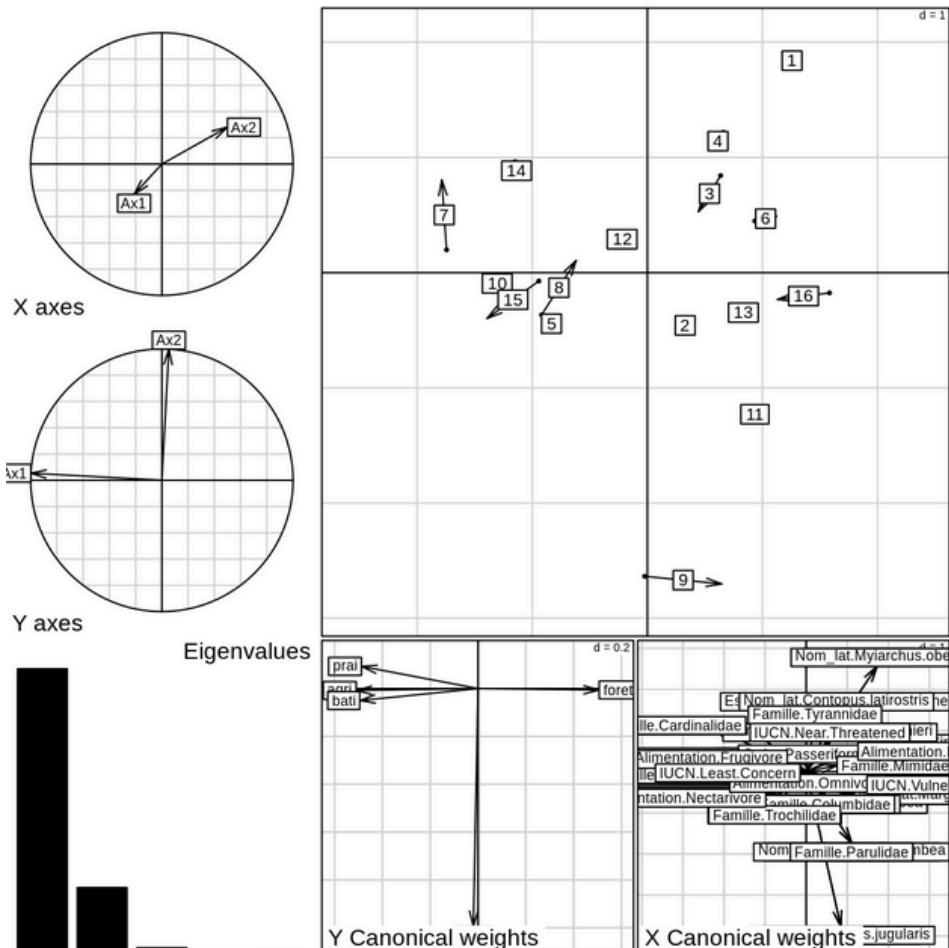
Pour ce faire, nous allons réaliser une analyse des correspondances multiples (Fig.4). Nous pouvons observer que *Melanerpes herminieri* (Meh), espèce endémique menacée, se distingue nettement des autres espèces. Tandis que les espèces insectivores notamment, et plus particulièrement appartenant à des familles comme les Tyrannidae ou Mimidae, sont à l'opposé, représentant sans doute des espèces plus communes et généralistes.



**Figure 4 : Boxplot de l'analyse des correspondances multiples.** Les données analysées sont extraites du jeu de données “especes”.

Puis, nous avons fait une analyse en composantes principales du dataset “habitat”. Nous allons utiliser les résultats précédents afin de réaliser une analyse de co-inertie entre les deux tableaux : “especes” et “habitat”. Nous pouvons remarquer que l'analyse en composantes principales (Fig.5) explique davantage la co-inertie comparée à l'analyse des correspondances multiples malgré le fait

qu'elle explique deux axes. Les deux analyses montrent une bonne corrélation car les projections des deux axes sont proches. Le graphique montrent une superposition des oiseaux en fonction de leurs habitats et de leur espèce. Les vecteurs résultant de l'analyse de co-inertie ont une norme faible pour la majorité. Nous pouvons donc noter que la modalité espèce est corrélée avec les variables de l'environnement.



**Figure 5 : Graphiques regroupant les résultats de l'analyse de la co-inertie.** Les données analysées sont extraites du jeu de données “espèces” et “habitat”. Le X représente l'analyse en correspondances multiples et le Y représente l'analyse en composantes principales. L'histogramme représente l'éboulis des valeurs propres.

```
t = -1.7395, df = 14,
p-value = 0.1039
alternative hypothesis: true
correlation is not equal to 0

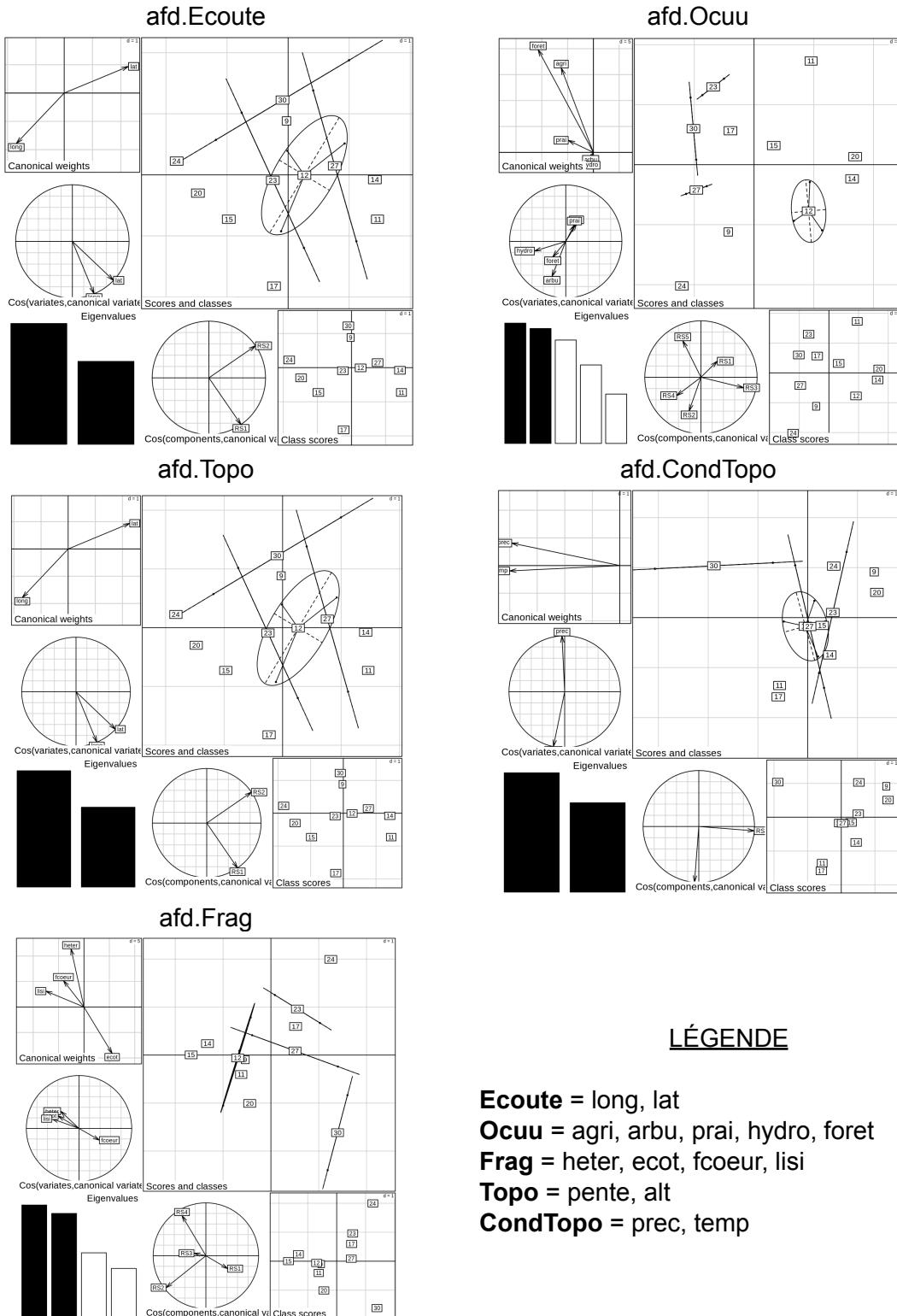
sample estimates:
  cor
-0.4215692
```

## LA PHYSIOLOGIE DIFFÈRE-T-ELLE SELON LES CARACTÉRISTIQUES DU MILIEU ?

Afin de tester si nos données sur les habitats sont corrélées directement aux espèces, nous avons fait les moyennes des modalités par espèces. Puis, nous avons réalisé un test de corrélation (Fig.6) pour déterminer s'il est pertinent de faire des analyses de couplage tel qu'une analyse factorielle discriminante.

Les tests préliminaires (test de normalité, homéostasie et linéarité) ont été réalisés et sont concluants. Cependant, des doutes subsistent concernant le test de linéarité. Nous allons donc effectuer le test de corrélation pour s'assurer de la dépendance des modalités étudiées. L'hypothèse nulle de

**Figure 6 : Résultats du test de corrélation.** Les données analysées sont extraites du jeu de données “espèces” et “habitat”. Le test à été réalisé en utilisant la commande cor.test avec la méthode de Pearson du langage R.



### LÉGENDE

**Ecoute** = long, lat  
**Ocuu** = agri, arbu, prai, hydro, foret  
**Frag** = heter, ecot, fcoeur, lisi  
**Topo** = pente, alt  
**CondTopo** = prec, temp

**Figure 7 : Analyse factorielle discriminante sur différentes catégories de variables environnementales.**  
 Les données analysées sont extraites du jeu de données “especes” et “habitat”.

ce test (cor.test par la méthode de Pearson) est la dépendance des données. Nous pouvons noter une p-value supérieure au risque de première espèce (0,05). Nous ne rejetons pas l'hypothèse nulle, les variables sont donc corrélées. De par cette analyse, nous pouvons en extraire que les variables environnementales (moyenne des occupations du sol -*agri*, *arbu*, *prai*, *foret*, *bati*) par espèces sont négativement corrélées à la taille des oiseaux.

Pour la suite de l'analyse, nous allons nous intéresser à l'impact de l'environnement sur la physiologie des oiseaux. Pour cela, nous avons réalisé des analyses en composantes principales sur chaque sous-groupes de modalités (Annexe 9). Il est à noter que les analyses des groupes à deux variables n'ont qu'un seul axe et sont donc peu pertinentes mais nécessaires pour réaliser l'analyse factorielle discriminante par sous-groupe.

L'analyse factorielle discriminante réalisée sur différentes catégories de variables environnementales (Fig.7) vise à explorer dans quelle mesure la taille des espèces peut être expliquée par leur environnement. Les analyses ont été menées à l'aide des sous-groupes de variables suivants : lieu d'écoute (*Ecoute*), occupation du sol (*Ocuu*), fragmentation (*Frag*), topographie (*Topo*), et conditions topographiques (*CondTopo*).

Pour le sous-groupe sur les lieux d'écoutes (*Ecoute*), les résultats n'ont pas permis de dégager des classes basées sur la taille des espèces. Certaines variables, comme les groupes de taille 12 et 23, sont peu expliquées par les deux premiers axes. En revanche, les tailles 20 et 24 présentent des coordonnées proches de celles des groupes de taille 9 et 30, tandis que les groupes de taille 11 et 14 sont similaires entre eux, mais opposés à 20 et 24, suggérant une certaine structure spatiale sans pour autant clarifier la distribution selon la taille.

Concernant le sous-groupe sur l'occupation du sol (*Ocuu*), aucune structure claire n'émerge non plus, sauf pour l'espèce de taille 12 (*Moqueur grivotte*), qui se distingue partiellement. Les variables *arbu* et *hydro* semblent peu explicatives. L'espèce de taille 23, quant à elle, se retrouve souvent en contexte forestier et en présence de grandes surfaces agricoles.

Pour les variables liées à la fragmentation de l'habitat (*Frag*), le modèle ne révèle aucune classe bien définie. Bien que ces variables soient informatives, leurs vecteurs sont colinéaires, limitant leur pouvoir discriminant.

Les groupes topographiques (*Topo*) ne permettent pas non plus la création de classes claires. Toutefois, ces deux variables sont explicatives avec des vecteurs colinéaires et permettent une séparation relativement cohérente des espèces sur le plan topographique.

De même, les conditions topographiques (*CondTopo*) partagent cette caractéristique de vecteurs colinéaires. Bien qu'elles soient explicatives, elles n'aboutissent pas à des regroupements nets selon la taille des espèces.

## CONCLUSION & DISCUSSION

A travers cette étude, nous avons pu souligner la présence de liens complexes entre les caractéristiques de l'environnement, en particulier l'urbanisation, et la répartition des espèces d'oiseaux en Guadeloupe. L'ensemble des analyses multivariées nous a permis d'explorer la structure des communautés aviaires endémiques de Guadeloupe en lien avec des variables écologiques fines.

Nous pouvons ainsi en déduire que certaines variables environnementales telles que la température, la latitude, la longitude et le type d'unité écologique influencent significativement la présence des espèces. L'urbanisation, quant à elle, apparaît comme un facteur essentiel mais non suffisant pour expliquer la répartition des oiseaux.

En effet, cette influence reste globalement faible, la répartition des espèces n'est pas strictement déterminée par l'occupation du sol. De plus, aucune corrélation significative n'a été observée entre les espèces en termes de cohabitation, de comportement alimentaire ou en réponse à l'urbanisation. L'analyse de la physiologie semble montrer une légère tendance à la corrélation négative entre la taille des espèces et les milieux urbains. Elle ne se révèle cependant pas statistiquement significative.

Ces résultats semblent également suggérer une relative indifférence des communautés aviaires endémiques de Guadeloupe face à une urbanisation modérée. Toutefois, ils soulignent également l'importance des milieux naturels forestiers dans la conservation des espèces endémiques. En effet, l'espèce protégée *Melanerpes herminieri*, en particulier, se distingue par une forte sensibilité à l'environnement, justifiant des mesures de gestion ciblées.

Cette analyse met en lumière la complexité des relations entre la faune et son environnement dans un contexte d'urbanisation croissante. Une gestion fine du territoire sur la préservation de la biodiversité est à privilégier mais ne semble pas contraire au développement de l'urbanisation. D'autres perspectives d'études pourraient aider à mieux caractériser l'impact de l'urbanisation sur les espèces d'oiseaux.

Une analyse temporelle de la répartition des espèces est à envisager pour observer l'évolution des communautés aviaires sur plusieurs années, en lien avec l'évolution de l'urbanisation. Cela permettrait de distinguer des tendances durables d'adaptation ou de déclin des espèces. De même, les résultats d'une telle étude pourraient mettre en lumière l'existence d'un seuil de densité urbaine au-delà duquel certaines espèces disparaissent ou se raréfient et ainsi définir des limites applicables et concrètes au développement urbain. Une identification et une mise en relation des types de pollutions engendrées par l'urbanisation et la répartition des espèces pourrait apporter des informations complémentaires essentielles. En effet, l'urbanisation ne se résume pas à la densité des bâtiments. Analyser l'influence d'autres facteurs, tels que la pollution sonore, lumineuse ou chimique, permettrait de mieux cerner les contraintes qu'impose l'urbanisation sur les espèces.

Une étude menée sur l'ensemble du globe permettrait de diversifier les zones d'écoute. Une confrontation des résultats obtenus en Guadeloupe avec d'autres territoires similaires, ou au contraire particulièrement différents, pourrait aider à identifier des tendances générales ou au contraire des spécificités locales.

De même, nous pourrions inclure d'autres espèces<sup>(8)</sup> dans le corpus de l'étude. Ainsi, nous pourrions éventuellement observer des différences de comportements face à l'urbanisation des territoires, selon s'il s'agit d'oiseaux endémiques ou migrateurs. Nous pourrions ainsi vérifier si ces derniers proviennent de territoires à l'urbanisation différente ou similaire, indiquant que le choix du lieu de migration est fonction de l'occupation du sol.

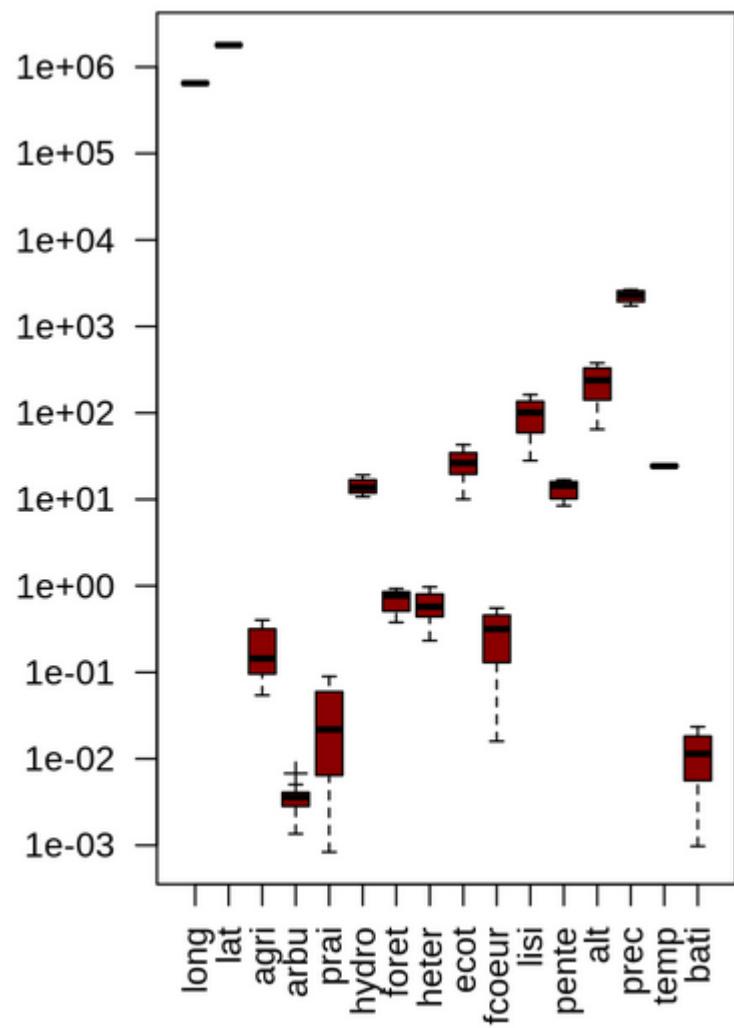
## RÉFÉRENCES

- [1] <https://chatgpt.com/>
- [2] <https://www.insee.fr/fr/statistiques/1280970>
- [3] <https://doi.org/10.32614/CRAN.package.ade4>
- [4] <https://doi.org/10.32614/CRAN.package.vegan>
- [5] <https://doi.org/10.32614/CRAN.package.lmtest>
- [6] <https://doi.org/10.32614/CRAN.package.MASS>
- [7] <https://doi.org/10.32614/CRAN.package.tidyverse>
- [8] <https://www.oiseaux.net/oiseaux/quadeloupe.html>

## ANNEXES



**Annexe 1 : Images des espèces endémiques guadeloupéennes étudiées.** Les données de présence-absence et les attributs des espèces sont respectivement répertoriées dans les dataset “communaute” et “especes”. Les espèces sont nommées selon leur nom scientifique. Pour avoir l'équivalent en abrégé dans les analyses ou encore le nom vernaculaire, se référer au glossaire.



**Annexe 2 : Boxplot représentant les données environnementales mesurées.** Les données analysées sont extraites du jeu de données “habitat”.

	Gem	Maf	Tul	Mas	Orc	Euj	Col	Dep	Meh	Saa	Lon	Cir	Myo	Cof	Euh	Elm
loc																
UE																
long																
lat																
agri																
arbu																
prai																
heter																
ecot																
fcoeur																
temp																
bati																
hydro																
foret																
lisi																
pente																
alt																
prec																

**Annexe 3 : Tableau représentant les résultats du test de  $\chi^2$  d'indépendance sur la répartition d'une espèce d'oiseau en fonction des caractéristiques de l'environnement.** Les données analysées sont extraites des jeux de données "habitat" et "communaute". Une p-value inférieure au risque de première espèce sera représentée par une case verte, tandis que celle dont la p-value est supérieure au risque de première espèce sera représentée par une case rouge.

```

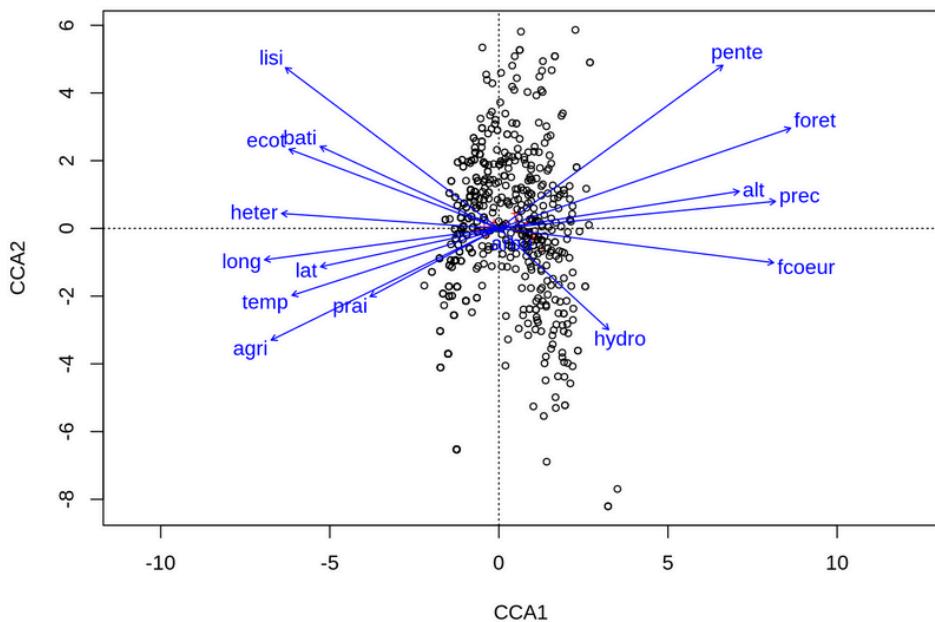
##                  CS1          CS2
## long    -0.28220786  0.096503992
## lat     -0.21011116  0.361947976
## agri   -0.24629728  0.145164973
## arbu    0.04667534 -0.300078275
## prai   -0.12185444  0.161730500
## hydro   0.09465626  0.007728365
## foret   0.31571473 -0.078104816
## heter  -0.25723888 -0.299136122
## ecot   -0.23874791 -0.404577352
## fcoeur  0.31705334  0.171227746
## lisi   -0.20887190 -0.397278555
## pente   0.29376299 -0.217028999
## alt     0.32254790 -0.071134350
## prec    0.33365837 -0.164285853
## temp   -0.30359650  0.144051605
## bati   -0.16950428 -0.419555865

```

**Annexe 4 : Coefficient des modalités sur les deux composantes principales.** Les données analysées sont extraites du jeu de données "habitat".



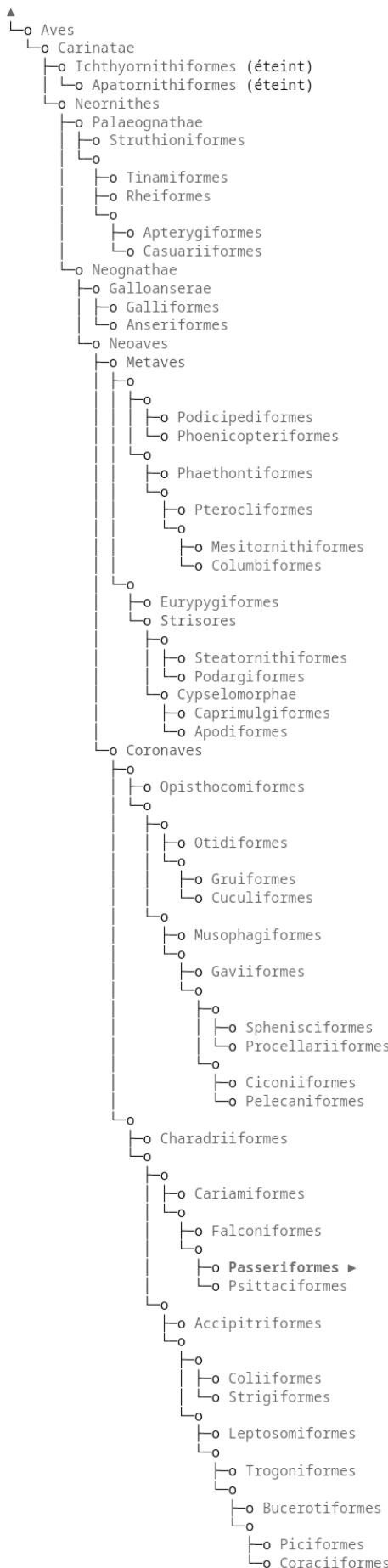
**Annexe 5 : Scatter-plot d'une AFC.** Les données analysées sont extraites du jeu de données “communaute”, les animaux sont représentés par leur code à 3 lettres et les lieux d’écoutes sont représentés par leur identifiant.



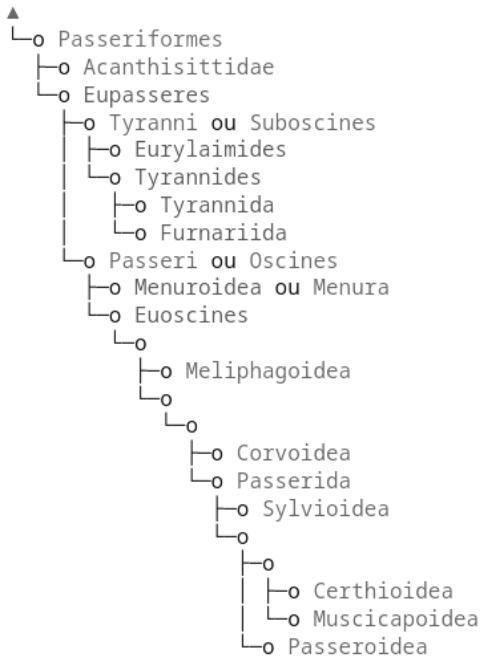
**Annexe 6 : Projection vectorielle des 16 modalités du dataset “habitat” sur celui de “communaute”.** Plus un vecteur est grand, plus il explique la dispersion des données dans son sens.

	Gem	Maf	Tul	Mas	Orc	Euj	Col	Dep	Meh	Saa	Lon	Cir	Myo	Cof	Euh	Elm
Gem	Grey	Green	Red	Green	Green	Green	Red	Red	Red							
Maf			Green	Green	Red					Green	Red			Red	Green	Red
Tul			Grey	Green	Red					Red	Red			Green	Green	Red
Mas			Grey	Green						Red	Green			Green	Red	Red
Orc			Grey	Grey	Red					Green	Red	Red	Red	Red	Green	
Euj										Red	Green					
Col										Green	Red					
Dep										Red	Green			Green		
Meh										Green	Green			Red		
Saa										Green	Green		Red	Green		
Lon										Green	Red		Green	Green		
Cir										Green	Green			Green		
Myo										Green	Green		Red			
Cof										Green			Green		Red	
Euh													Grey		Green	
Elm															Grey	

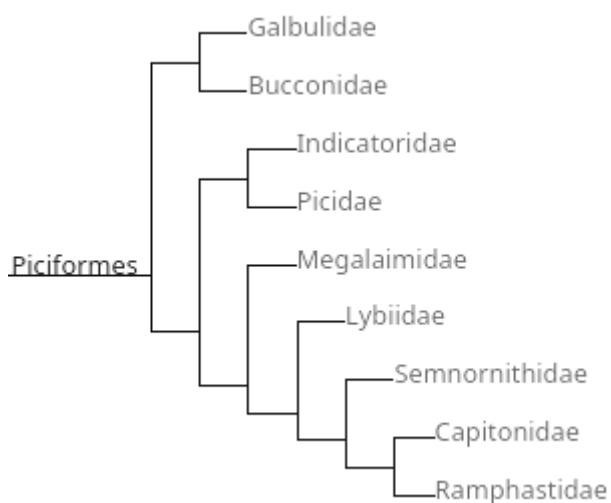
**Annexe 7 : Tableau représentant les résultats du test de  $\chi^2$  d'indépendance sur la répartition d'une espèce d'oiseau en fonction des autres espèces d'oiseaux.** Les données analysées sont extraites du jeu de données “communaute”. Une p-value inférieure au risque de première espèce sera représentée par une case verte, tandis que celle dont la p-value est supérieure au risque de première espèce sera représentée par une case rouge.



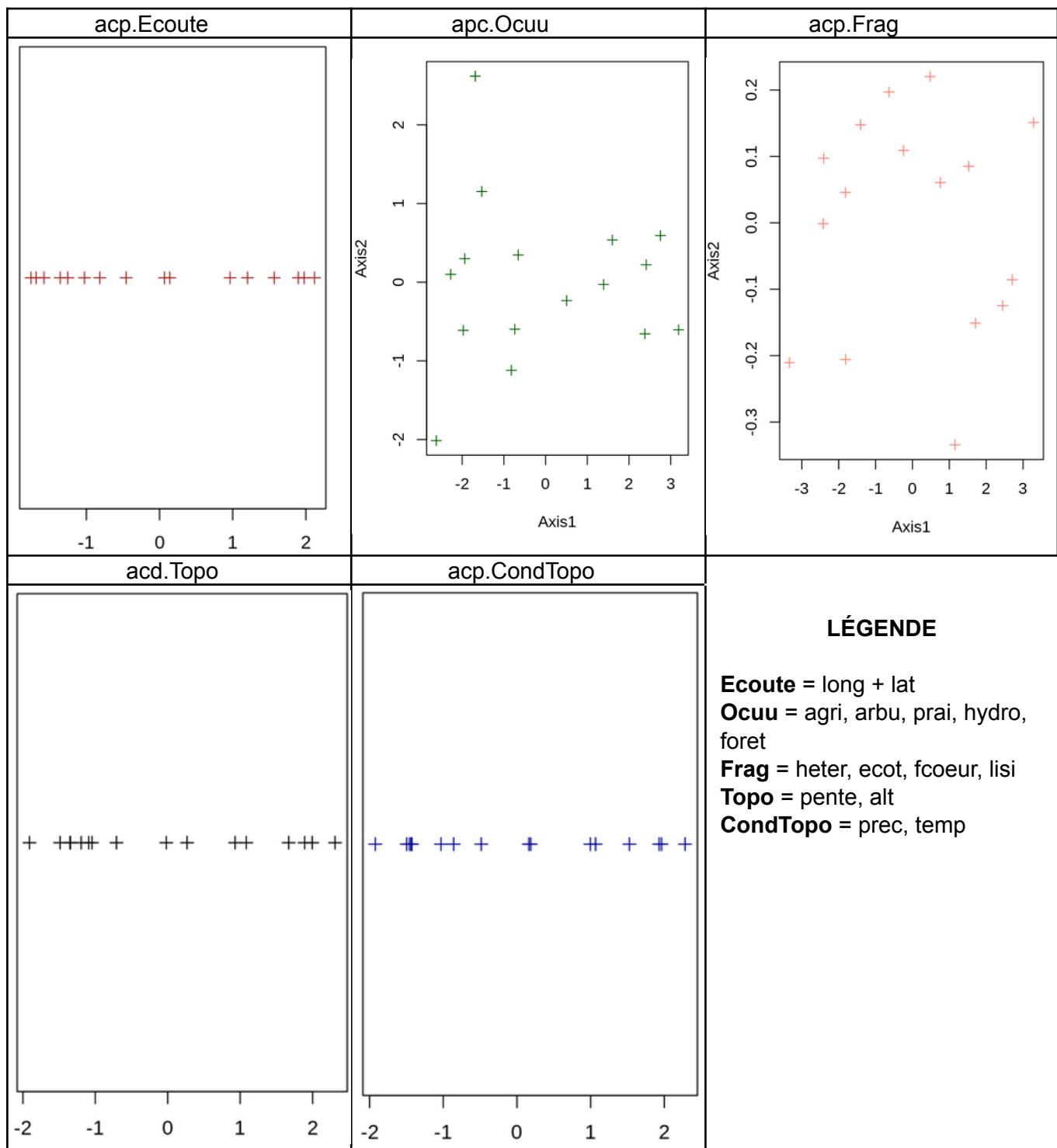
**Annexe 8a : Classification phylogénétique des espèces étudiées.** Les données sont extraites du site wikipedia ([https://fr.wikipedia.org/wiki/Classification\\_phylog%C3%A9nique\\_des\\_oiseaux](https://fr.wikipedia.org/wiki/Classification_phylog%C3%A9nique_des_oiseaux))



**Annexe 8b : Classification phylogénétique des Passeriformes.** Les données sont extraites du site wikipedia ([https://fr.wikipedia.org/wiki/Passeriformes\\_\(classification\\_phylog%C3%A9nique\)](https://fr.wikipedia.org/wiki/Passeriformes_(classification_phylog%C3%A9nique)))



**Annexe 8c : Classification phylogénétique des Piciformes.** Les données sont extraites du site wikipedia (<https://fr.wikipedia.org/wiki/Piciformes>)



**Annexe 9 : scatter-plot des ACP des sous groupes des modalités.** Les données analysées sont extraites du jeu de données “habitat”, les points sont les projections sur les axes principaux.

```
Shapiro-Wilk normality test
```

```
data: meanList  
W = 0.93103, p-value = 0.2532
```

```
> shapiro.test(especies$Taille)
```

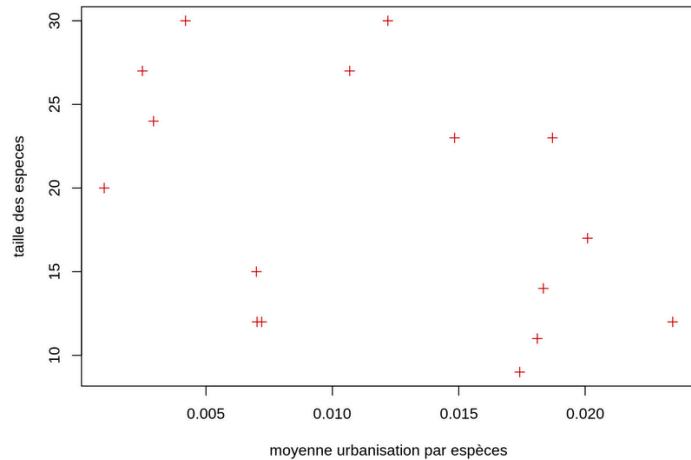
```
Shapiro-Wilk normality test
```

```
data: especes$Taille  
W = 0.91199, p-value = 0.1253
```

```
studentized Breusch-Pagan test
```

```
data: meanList ~ especes$Taille  
BP = 0.34564, df = 1, p-value = 0.5566
```

**Annexe 10 : Tests préliminaires au test de corrélation.** Les tests analysées sont extraites du jeu de données “habitat”.



**Annexe 11 : plot taille des espèces en fonction du taux d'urbanisation.** Les tests analysés sont extraites du jeu de données “habitat” et “especes”.

## DOCUMENTS COMPLÉMENTAIRES

### Document 1 : Script sur le chargement des données.

```
habitat = read.table("../data/habitat.tsv", sep = "\t",
                     stringsAsFactors = T, header = T, na.strings = T)
especies = read.table("../data/especies.tsv", sep = "\t",
                     stringsAsFactors = T, header = T, na.strings = T)
communautes = read.table("../data/communautesnautes.tsv", sep = "\t",
                        stringsAsFactors = T, header = T, na.strings = T)
```

### Document 2 : Script sur le test de $\chi^2$ d'indépendance sur la répartition d'une espèce d'oiseau en fonction de l'environnement.

```
communautes_clean = communautes[, -1]
habitat_clean = habitat[, -1]
esp = colnames(communautes_clean[1,])
hab = colnames(habitat_clean[1,])

write(x = "Test de chi2 sur la répartition d'une espèce d'oiseau en fonction de l'environnement", file = "chi-2_q1_export.txt", append = F)
n=0
for (i in 1:length(communautes_clean[1,])){
  n=n+1
  communautesI = paste("\n-----", esp[n], "-----")
  write(x = communautesI, file = "chi-2_q1_export.txt", append = T)
  m = 0
  for (j in 1:length(habitat_clean[1,])){
    m=m+1
    chi2 = chisq.test(communautes_clean[,i], habitat_clean[,j])
    habitatI = paste(hab[m], ":", chi2$p.value)
    write(x = habitatI, file = "chi-2_q1_export.txt", append = T)
  }
}

write(x = "Résultats d'indépendance entre les données", file = "chi-2_q1_export_indep.txt",
append = F)
write(x = "Résultats de dépendance entre les données", file = "chi-2_q1_export_dep.txt",
append = F)
n=0
for (i in 1:length(communautes_clean[1,])){
  n=n+1
  communautesI = paste("\n-----", esp[n], "-----")
  write(x = communautesI, file = "chi-2_q1_export_indep.txt", append = T)
  write(x = communautesI, file = "chi-2_q1_export_dep.txt", append = T)
  m = 0
  for (j in 1:length(habitat_clean[1,])){
    m=m+1
    chi2 = chisq.test(communautes_clean[,i], habitat_clean[,j])
    habitatI = paste(hab[m], ":", chi2$p.value)

    if (chi2$p.value<0.05){
      write(x = habitatI, file = "chi-2_q1_export_dep.txt", append = T)
    }
    else{
      write(x = habitatI, file = "chi-2_q1_export_indep.txt", append = T)
    }
  }
}
```

**Document 3a** : Sortie du script (Doc.2) sur les résultats de dépendance entre les données du test de la répartition d'une espèce d'oiseau en fonction de l'environnement.

**Document 3b** : Sortie du script (Doc.2) sur les résultats de l'indépendance entre les données du test de la répartition d'une espèce d'oiseau en fonction de l'environnement.

Résultats de dépendance entre les données	Résultats d'indépendance entre les données
----- Gem ----- loc : 3.97849670889657e-11 UE : 3.03482434029665e-18 long : 0.0118774651318067 lat : 0.0233550626079455 arbu : 0.0172428418570418 fcoeur : 8.00808613276792e-18 temp : 0.00645646932239253	----- Gem ----- agri : 0.944655715942457 prai : 0.99999980017232 hydro : 0.477279597304613 foret : 0.719059517019088 heter : 0.886951202471342 ecot : 0.647738008386706 lisi : 0.617019907443181 pente : 0.456503514775866 alt : 0.484447629315317 prec : 0.484447629315317 bati : 0.835736813817381
----- Maf ----- loc : 1.23761285975185e-45 UE : 8.70029216306414e-62 long : 6.01920984072491e-13 lat : 4.02544411299939e-12 fcoeur : 7.03308315077041e-40 temp : 8.01781556673379e-05	----- Maf ----- agri : 0.998051751055625 arbu : 0.404111440773872 prai : 1 hydro : 0.365373976212342 foret : 0.690121759014592 heter : 0.991087005827794 ecot : 0.911490810560114 lisi : 0.708275781927374 pente : 0.456503514775864 alt : 0.484447629315315 prec : 0.484447629315315 bati : 0.989895170252551
----- Tul ----- loc : 3.85814704895311e-36 UE : 9.0189067846949e-49 long : 1.07916402324195e-09 lat : 1.29760756901779e-09 fcoeur : 8.21930866107264e-50 temp : 0.000106314632081956	----- Tul ----- agri : 0.999928600803286 arbu : 0.124007053938423 prai : 1 hydro : 0.311624425559753 foret : 0.720369255703365 heter : 0.985922543480561 ecot : 0.967916959378534 lisi : 0.722321302283637 pente : 0.456503514775865 alt : 0.484447629315316 prec : 0.484447629315316 bati : 0.740100172730744
----- Mas ----- loc : 3.18926991195067e-17 UE : 1.37650598268566e-35 long : 0.0265771995645365 lat : 2.12659845825375e-05 fcoeur : 0.008408252294829	----- Mas ----- agri : 0.461052847570073 arbu : 0.495196374958537 prai : 0.337996022699942 hydro : 0.443321780756528
----- Orc ----- UE : 3.73992779916646e-10 long : 0.00173138195659215	
----- Euj ----- loc : 9.70841877186119e-21 UE : 9.8302364512904e-23 long : 0.000541953802533168 lat : 2.68648776613993e-07 arbu : 0.02742901032858 fcoeur : 7.58548009039081e-37 temp : 0.000126611664787517	

----- Col -----	foret : 0.333964814987431 heter : 0.407617489022434 ecot : 0.306521573198261 lisi : 0.34242318446699 pente : 0.456503514775865 alt : 0.484447629315316 prec : 0.484447629315316 temp : 0.227857857917936 bati : 0.505756264957315
----- Dep -----	----- Orc ----- loc : 0.965928893700934 lat : 0.152934965353574 agri : 0.830793793900777 arbu : 0.109697645647103 prai : 0.615173241147942 hydro : 0.492475072090335 foret : 0.382997374401017 heter : 0.643679967384343 ecot : 0.511458874114799 fcoeur : 0.10839243763107 lisi : 0.424197001131746 pente : 0.456503514775866 alt : 0.484447629315317 prec : 0.484447629315317 temp : 0.44924379403065 bati : 0.394595665615355
----- Meh -----	----- Euj ----- agri : 0.999711433882022 prai : 1 hydro : 0.372116183410579 foret : 0.912454526790457 heter : 0.99759378089184 ecot : 0.961117910605981 lisi : 0.918712894474701 pente : 0.456503514775866 alt : 0.484447629315317 prec : 0.484447629315317 bati : 0.99999413064107
----- Saa -----	----- Col ----- long : 0.452878412067871 lat : 0.298913303985546 agri : 0.999804167528838 arbu : 0.986746346373132 prai : 0.999902539782387 hydro : 0.644828527910327 foret : 0.443754871115629 heter : 0.949343614399626 ecot : 0.673902871082448 lisi : 0.560394967000027 pente : 0.525210198442907
----- Lon -----	
----- Cir -----	
----- Myo -----	
----- Cof -----	

long : 0.000291601186109117  
lat : 4.76382681849988e-06  
fcoeur : 3.04590989072352e-18  
temp : 0.0495207437521548

----- Euh -----

loc : 0.00212964120943696  
UE : 1.5843753595224e-06  
agri : 0.000238369301646243  
prai : 0.00225962829416561  
heter : 0.0136255261118365  
ecot : 0.0179489702307699  
bati : 0.000489735295929632

----- Elm -----

loc : 1.02383214084957e-31  
UE : 9.38409899354379e-46  
long : 1.12882120356029e-10  
lat : 1.14092122885002e-05  
fcoeur : 1.9868657395345e-08  
bati : 0.0268030683302548

alt : 0.484447629315316  
prec : 0.484447629315316  
bati : 0.99222280187718

----- Dep -----

agri : 0.978501599532222  
prai : 0.999999271021437  
hydro : 0.473312812122898  
foret : 0.597410878028803  
heter : 0.857996090474448  
ecot : 0.753901968455594  
lisi : 0.717285362097408  
pente : 0.456503514775865  
alt : 0.484447629315316  
prec : 0.484447629315316  
bati : 0.428244493743504

----- Meh -----

agri : 0.84380037418806  
arbu : 0.0913087253563225  
prai : 0.999011126321269  
hydro : 0.456840278507123  
foret : 0.594209277203768  
heter : 0.763921911002379  
ecot : 0.745025821246824  
lisi : 0.571183382845362  
pente : 0.480458282998902  
alt : 0.484447629315316  
prec : 0.484447629315316  
bati : 0.858012293380698

----- Saa -----

agri : 0.382166990273697  
arbu : 0.340696414906705  
prai : 0.440912897798503  
hydro : 0.486034193218065  
foret : 0.139637346654072  
heter : 0.207584915935045  
ecot : 0.209562318177177  
fcoeur : 0.0635107389099945  
lisi : 0.217945196044388  
pente : 0.494182879377995  
alt : 0.484447629315316  
prec : 0.484447629315316  
temp : 0.330352420237235

----- Lon -----

loc : 0.0575520706054077  
agri : 0.479025013505129  
arbu : 0.68960528599341  
prai : 0.334481228141229  
hydro : 0.470531027241802  
foret : 0.453406068536079

heter : 0.461399149447954  
ecot : 0.40059943102699  
fcoeur : 0.429663096161088  
lisi : 0.523223299144043  
pente : 0.475638361561621  
alt : 0.484447629315315  
prec : 0.484447629315315  
temp : 0.534223666879335  
bati : 0.373289176364727

----- Cir -----

agri : 0.999890325330359  
arbu : 0.374294235218863  
prai : 1  
hydro : 0.366283803458509  
foret : 0.594340371228436  
heter : 0.952621967236661  
ecot : 0.876303045938593  
lisi : 0.714286327953021  
pente : 0.489564097306995  
alt : 0.484447629315316  
prec : 0.484447629315316  
bati : 0.512105109345743

----- Myo -----

long : 0.0606007629709407  
agri : 0.99960316248491  
arbu : 0.900117724126671  
prai : 1  
hydro : 0.28109645412711  
foret : 0.958212864630155  
heter : 0.999874841726895  
ecot : 0.992543463162034  
lisi : 0.649082899958358  
pente : 0.576525471619906  
alt : 0.484447629315317  
prec : 0.484447629315317  
bati : 0.99999999591224

----- Cof -----

agri : 0.999971603641854  
arbu : 0.866419185245612  
prai : 0.999999995702145  
hydro : 0.347182872046765  
foret : 0.940315377577754  
heter : 0.998729387573241  
ecot : 0.991503409145958  
lisi : 0.871780376735711  
pente : 0.456503514775866  
alt : 0.484447629315317  
prec : 0.484447629315317  
bati : 0.66419276491571

```

----- Euh -----
long : 0.313587675475814
lat : 0.845914642901199
arbu : 0.711074088678061
hydro : 0.396364528309974
foret : 0.0820302772532145
fcoeur : 0.999999999879356
lisi : 0.182706578292374
pente : 0.500527405910191
alt : 0.484447629315316
prec : 0.484447629315316
temp : 0.895328802211066

----- Elm -----
agri : 0.662219259652887
arbu : 0.62453525501937
prai : 0.921664751829203
hydro : 0.47152683952313
foret : 0.158728016856255
heter : 0.300078369185385
ecot : 0.204753925656462
lisi : 0.293569909827036
pente : 0.497400315524936
alt : 0.484447629315316
prec : 0.484447629315316
temp : 0.0839550027924124

```

#### Document 4 : ACP normé d'*habitat*.

```

res = chisq.test(habitat[,4:19]) ; res
# Dépendance des données

acp_habitat = dudi.pca(habitat[,4:19], scale=T, nf = 2, scannf = F)
plot(acp_habitat$li,
     pch = 3)
s.corcircle(acp_habitat$co,
            grid = F)
acp_habitat$c1
# 1 variable expliquant peu le modèle : hydro (faible pourcentage
application des axes, <10%)

```

#### Document 5 : AFC d'*especes*.

```

afc1 = dudi.coa(communautes[,-1], nf=2, scannf = F) ; afc1
barplot(afc1$eig)
scatter(afc1)

# Log des données
caract_habitat = habitat[,-(1:3)]
log_habitat = log(caract_habitat +1)

```

**Document 6 : ACC entre l'ACP d'*habitat* et l'AFC d'*espece*.**

```
res = chisq.test(communautes[,2:17]) ; res
# Indépendance des données

cca = pcaiv(afc1, log_habitat, scannf = F, nf = 2)

#projection axe d'inertie
s.corcircle(cca$as)

# Position des relevés par combinaisons de variables environnementales :
# représentation des sites
s.label(cca$li, clabel = 0, cpoint = 1.5)

# Position des espèces à la moyenne des sites qui les contiennent :
# représentation des espèces
s.label(cca$co, add.plot = T)

# Poids canoniques donnant les combinaisons de variance unité
s.arrow(cca$fa)

# Prédiction des positions par régression sur les variables de milieu
s.match(cca$li, cca$ls, clabel = 0.5)

# Ensemble des résultats de l'analyse
plot(cca)

acc1 = cca(communautes[,-1], caract_habitat)
plot(acc1)
#répartition lisière de foret et foret à l'opposé, coeur de la foret
corrélé à foret

acc2 = cca(communautes[,-1] ~ bati + agri + arbu + fcoeur + prai + foret,
caract_habitat)
plot(acc2)
#explication de la répartition avec les conditions occupations du sol et
urbanisation
#agri, bati, hydro, foret à conserver

acc3 = cca(communautes[,-1] ~ bati + agri + fcoeur + foret, caract_habitat)
plot(acc3)
# urbanisation opposé à réseau hydrographique, surface agricole opposé à
foret
```

**Document 7 :** Impression des  $\chi^2$  d'indépendance sur la répartition d'une espèce d'oiseau en fonction des autres.

```
taille = length(communautes_clean[,])-1

write(x = "Test de chi2 sur la répartition d'une espèce d'oiseau en
fonction des autres", file = "chi-2_q1_export2.txt", append = F)
for (i in 1:taille){
  communautesI = paste("\n-----", esp[i], -----")
  write(x = communautesI, file = "chi-2_q1_export2.txt", append = T)
  deb = i+1
  for (j in deb:taille){
    chi2 = chisq.test(communautes_clean[,i], communautes_clean[,j])
    espI = paste(esp[j], ":", chi2$p.value)
    write(x = espI, file = "chi-2_q1_export2.txt", append = T)
  }
}

write(x = "Résultats d'indépendance entre les données", file =
"chi-2_q1_export2_indep.txt", append = F)
write(x = "Résultats de dépendance entre les données", file =
"chi-2_q1_export2_dep.txt", append = F)
for (i in 1:taille){
  communautesI = paste("\n-----", esp[i], -----)
  write(x = communautesI, file = "chi-2_q1_export2_indep.txt", append = T)
  write(x = communautesI, file = "chi-2_q1_export2_dep.txt", append = T)
  deb = i+1
  for (j in deb:taille){
    chi2 = chisq.test(communautes_clean[,i], communautes_clean[,j])
    espI = paste(esp[j], ":", chi2$p.value)

    if (chi2$p.value<0.05){
      write(x = espI, file = "chi-2_q1_export2_dep.txt", append = T)
    }
    else{
      write(x = espI, file = "chi-2_q1_export2_indep.txt", append = T)
    }
  }
}
```

**Document 8 :** Co-inertie entre *habitat* et *communautes*.

```
couplage = cbind(communautes[,-1], habitat[,-1])
habitat_numerique = habitat[,-(1:3)]

#AFC sur tableau de comptage
afc_communautés = dudi.coa(communautes[,-1], nf=2, scannf = F)
scatter(afc_communautés)

#ACP normée sur milieu
acp_habitat = dudi.pca(habitat_numerique, scale=T, nf = 2, scannf = F)
```

```

#set même poids partout
acp_habitat$lw = rep(1, length(acp_habitat$lw))
afc_communautés$lw = rep(1, length(afc_communautés$lw))

coinertie = coinertia(acp_habitat, afc_communautés, scannf = F, nf = 2)
plot(coinertie)
scatter(coinertie)

#AFC sur couplage
afc_couplage = dudi.coa(couplage, scannf = F, nf = 2)

scatter(afc_couplage)

```

**Document 9 :** Création d'une liste répertoriant les lieux où résident chaque espèces et représentation du tableau *habitat*.

```

# Récupération des noms des colonnes correspondant aux espèces (colonnes 2 à 17)
vectNom = names(communautés[2:17])

# Crédation d'une liste vide pour stocker les sous-ensembles
listSubsets = list()
# Parcours de toutes les espèces
for (i in 1:length(vectNom)) {
  especesActu = vectNom[i]
  # Extraction des numéros de lieu (IdPE) où l'espèce a une présence
  foosubset = communautés$IdPE[communautés[[especesActu]] != 0]
  listSubsets[[especesActu]] = foosubset
}; listSubsets$Gem

par(mfrow = c(1,1))
ExGem = colMeans(subset(habitat, habitat$IdPE %in% listSubsets$Gem))
plot(ExGem[-1])

ExMat = t(sapply(names(listSubsets), function(name) {
  colMeans(subset(habitat, habitat$IdPE %in% listSubsets[[name]])))
})

ExMat = ExMat[,-1]
boxplot(ExMat,
        pch = 3, col = "darkred",
        log = "y")

```

**Document 10 :** Boxplot et ACP de chaque groupe de variable.

```
## boxplots par classe de modalité :  
  
# 1. Point d'écoute :  
par(mfrow = c(1,2))  
  
plot(ExMat[,1:2],  
      pch = 3, col = "brown")  
  
acp.Ecoute = dudi.pca(ExMat[,1:2], center = TRUE, scale = TRUE,  
                      ,scannf = FALSE, nf = 1)  
plot(acp.Ecoute$li,  
      pch = 3, col = "brown")  
  
#corrélation +  
  
# 2. Occupation du sol :  
par(mfrow = c(1,3))  
  
boxplot(ExMat[,3:7],  
        pch = 3, col = "darkgreen")  
  
acp.Ocuu = dudi.pca(ExMat[,3:7], center = TRUE, scale = TRUE,  
                     scannf = FALSE, nf = 2)  
plot(acp.Ocuu$li,  
      pch = 3, col = "darkgreen")  
  
s.corcircle(acp.Ocuu$co)  
  
# pas de corrélation  
  
# 3. Fragmentation :  
  
par(mfrow = c(1,3))  
  
boxplot(ExMat[,8:11],  
        pch = 3, col = "salmon")  
  
acp.Frag = dudi.pca(ExMat[,8:11], center = TRUE, scale = TRUE,  
                     scannf = FALSE, nf = 2)  
plot(acp.Frag$li,  
      pch = 3, col = "salmon")  
  
s.corcircle(acp.Frag$co)  
  
# pas de corrélation
```

```

# 4. Topographie
par(mfrow = c(1,2))

plot(ExMat[,12:13],
  pch = 3, col = "black")

acp.Topo = dudi.pca(ExMat[,12:13], center = TRUE, scale = TRUE,
                     scannf = FALSE, nf = 1)
plot(acp.Topo$li,
  pch = 3, col = "black")

# corrélation acp.Topo# corrélation +
# 5. condition topographique
par(mfrow = c(1,2))

plot(ExMat[,14:15],
  pch = 3, col = "darkblue")

acp.CondTopo = dudi.pca(ExMat[,14:15], center = TRUE, scale = TRUE,
                        scannf = FALSE, nf = 1)
plot(acp.CondTopo$li,
  pch = 3, col = "darkblue")

# corrélation -
# 6. Urbanisation
par(mfrow = c(1,2))

par(mfrow = c(1,1))
boxplot(ExMat[,16],
  pch = 3, col = "darkgray")

# c'est une ligne

# 7. Total
par(mfrow = c(1,2))

boxplot(ExMat,
  pch = 3, col = "darkred",
  las = 2, log = "y")
acp.Total = dudi.pca(ExMat[,1:15], center = TRUE, scale = TRUE,
                      scannf = FALSE, nf = 2)

s.corcircle(acp.Total$co)

```

**Document 11 :** AFD par catégorie pour déterminer l'impact des classes d'*habitat* sur la taille des oiseaux.

```
# discrimin = AFD
# les arg : (acp, var qualitative)
afd.Ecoute.Taille = discrimin(acp.Ecoute, as.factor(especes$Taille),
                               , scannf = FALSE, nf = 2)

plot(afd.Ecoute.Taille)

#####
afd.Ocuu.Taille = discrimin(acp.Ocuu, as.factor(especes$Taille),
                             , scannf = FALSE, nf = 2)

plot(afd.Ocuu.Taille)

#####
afd.Frag.Taille = discrimin(acp.Frag, as.factor(especes$Taille),
                            , scannf = FALSE, nf = 2)

plot(afd.Frag.Taille)

#####
afd.Mat.Taille = discrimin(acp.Topo, as.factor(especes$Taille),
                           , scannf = FALSE, nf = 2)

plot(afd.Ecoute.Taille)

#####
afd.CondTopo.Taille = discrimin(acp.CondTopo, as.factor(especes$Taille),
                                 , scannf = FALSE, nf = 2)

plot(afd.CondTopo.Taille)
```

**Document 12 :** Génération des *subsets* des habitats par espèce et association de la taille à l'espèce.

```
for (i in seq_along(vectNom)) {
  subset_name = paste0(vectNom[i], "habitat")
  assign(subset_name, subset(habitat, IdPE %in% listSubsets[[vectNom[i]]]))

  # Ajouter la colonne Taille a chaque habitat en fonction de l'espèce
  temp_df = get(subset_name)
  temp_df[, 18] = especes$Taille[i]
  assign(subset_name, temp_df)
}
```

**Document 13 :** Test de Corrélation entre les tailles des espèces et la moyenne des attributs des habitats par espèce.

```
# test normalité :  
shapiro.test(meanList)  
shapiro.test(especies$Taille)  
  
# test linéarité :  
plot(meanList, especes$Taille,  
      pch = 3, col = "red3",  
      ylab = "taille des especes", xlab = "moyenne urbanisation par especes")  
  
# test homoscédasticité  
library(lmtest)  
bptest(meanList ~ especes$Taille)  
  
cor.test(especies$Taille, meanList,  
         method = "pearson")  
  
cor.test(especies$Taille, meanList,  
         method = "kendall")  
  
cor.test(especies$Taille, meanList,  
         method = "spearman")  
  
# les condition pour spearman étant faible, les 3 méthodes ont été quand même  
testé
```

**Document 14 :** Trie selon les espèces et calcul de la moyenne des attributs des habitats par espèce.

```
# Liste des noms d'habitat  
habitat_names = c("Myohabitat", "Gemhabitat", "Mehhabitat", "Colhabitat",  
"Lonhabitat",  
                  "Cirhabitat", "Euhhabitat", "Orchabitat", "Eujhabitat",  
"Cofhabitat",  
                  "Dephabitat", "Mashabitat", "Mafhabitat", "Saahabitat",  
"Elmhabitat", "Tulhabitat")  
  
# liste des diminutif des noms stocké précédemment dans vectNom  
  
# Liste des colonnes d'intérêt  
vectCol = c("agri", "arbu", "prai", "foret", "bati")  
  
# Initialisation d'une liste pour stocker les données  
meanList = data.frame(matrix(ncol = length(vectCol) + 1, nrow =  
length(habitat_names)))  
  
# Nommer les colonnes du dataframe  
colnames(meanList) = c("habitat", vectCol)
```

```

# Remplir le dataframe
for (i in seq_along(habitat_names)) {
  habitat = habitat_names[i]
  habitat_data = get(habitat) # Récupération du dataframe

  # Calcul des moyennes pour chaque colonne d'intérêt
  mean_values = sapply(vectCol, function(col) mean(habitat_data[[col]], na.rm = TRUE))

  # Stockage des résultats
  meanList[i, ] = c(habitat, mean_values)
}

# Convertir la colonne 'habitat' en facteur ou caractère
meanList$habitat = as.character(vectNom)

# Affichage du résultat
meanList = data.frame(meanList)

# Conversion des colonnes en numérique car bozo est char
meanList[,-1] = lapply(meanList[,-1], as.numeric)

```

**Document 15 :** Analyse de Co-inertie entre les moyennes des caractéristiques des habitats par espèce et les caractéristiques des espèces.

```

# ACP sur les moyennes des caractéristiques des habitats par espèces
acp.meanhabi = dudi.pca(meanList[,-1],
  center = T, scale = T,
  scannf = FALSE, nf = 2)

# AFC des espèces
acm.espèces = dudi.acm(as.data.frame(especes[-6]),
  scannf = FALSE, nf = 2)
boxplot(acm.espèces)

# Analyse de coinertie
coinertie = coinertia(acm.espèces, acp.meanhabi,
  scannf = FALSE, nf = 2); summary(coinertie)
s.corcircle(coinertie$li)
plot(coinertie)

```