



École nationale
de la statistique
et de l'analyse
de l'information



RAPPORT DE STAGE 1A STATISTIQUE PUBLIQUE

Établissement d'accueil : Direction Régionale de l'Insee de Nouvelle-Aquitaine, établissement de Bordeaux

Développer le caractère explicatif de la note trimestrielle de conjoncture

Emmanuel FRANCO

Maître de stage :

Rémi DUMAS

Préambule et Remerciements

Le stage s'est déroulé du **2 juin au 24 juillet 2020**.

En raison du contexte de crise sanitaire, le stage s'est entièrement déroulé en distanciel.

Malgré cette contrainte, je tiens à remercier tout particulièrement Rémi Dumas qui m'a accompagné tout au long de cette découverte de la statistique publique au sein du Service Études et Diffusion (SED), mais également à Stéphane Levasseur, chef de la division Développement économique et emploi, qui était présent à tout moment en cas de besoin.

Je tiens également à remercier l'ensemble de l'équipe pour l'ambiance studieuse et détendue qui régne à leurs côtés et qui rend fort agréable le travail en leur compagnie, *bien que virtuelle*.

Table des matières

I Description de l'environnement de travail et observations	1
1 Description de l'environnement de travail	1
2 À propos du management	2
3 Analyse de deux indicateurs qualité	3
3.1 Le secret statistique	3
3.2 L'impartialité et l'objectivité	3
II Travaux confiés et description des missions	4
1 L'emplois dans les différents secteurs d'activité au sein des départements	4
1.1 Une périodicité de l'emploi	4
1.2 Une dépendance spatiale dans l'évolution de l'emploi	5
1.3 Une dépendance de l'emploi entre secteurs d'activités	5
2 Analyse de secteurs d'activité à l'échelle régionale	6
2.1 Emploi dans la construction	6
2.2 Un lien entre l'emploi dans l'hébergement et la restauration et le nombre de nuitées dans l'hôtellerie .	7
III Ressenti de l'expérience et prise de recul	8
1 Ressenti de l'expérience	8
2 Prise de recul / analyse d'expérience	8
Annexe	1
1 Informations complémentaires	1
1.1 Départements de Nouvelle-Aquitaine	1
1.2 Secteurs d'activité	2
2 Grille d'entretiens	2
3 Code	2
3.1 Analyse exploratoire-v1	2
3.2 Analyse exploratoire-v2	6
3.2.1 Importations et traitement de données	6
3.2.2 Fonctions	16
3.3 Analyse exploratoire-v3	17
4 Sorties	20

Première partie

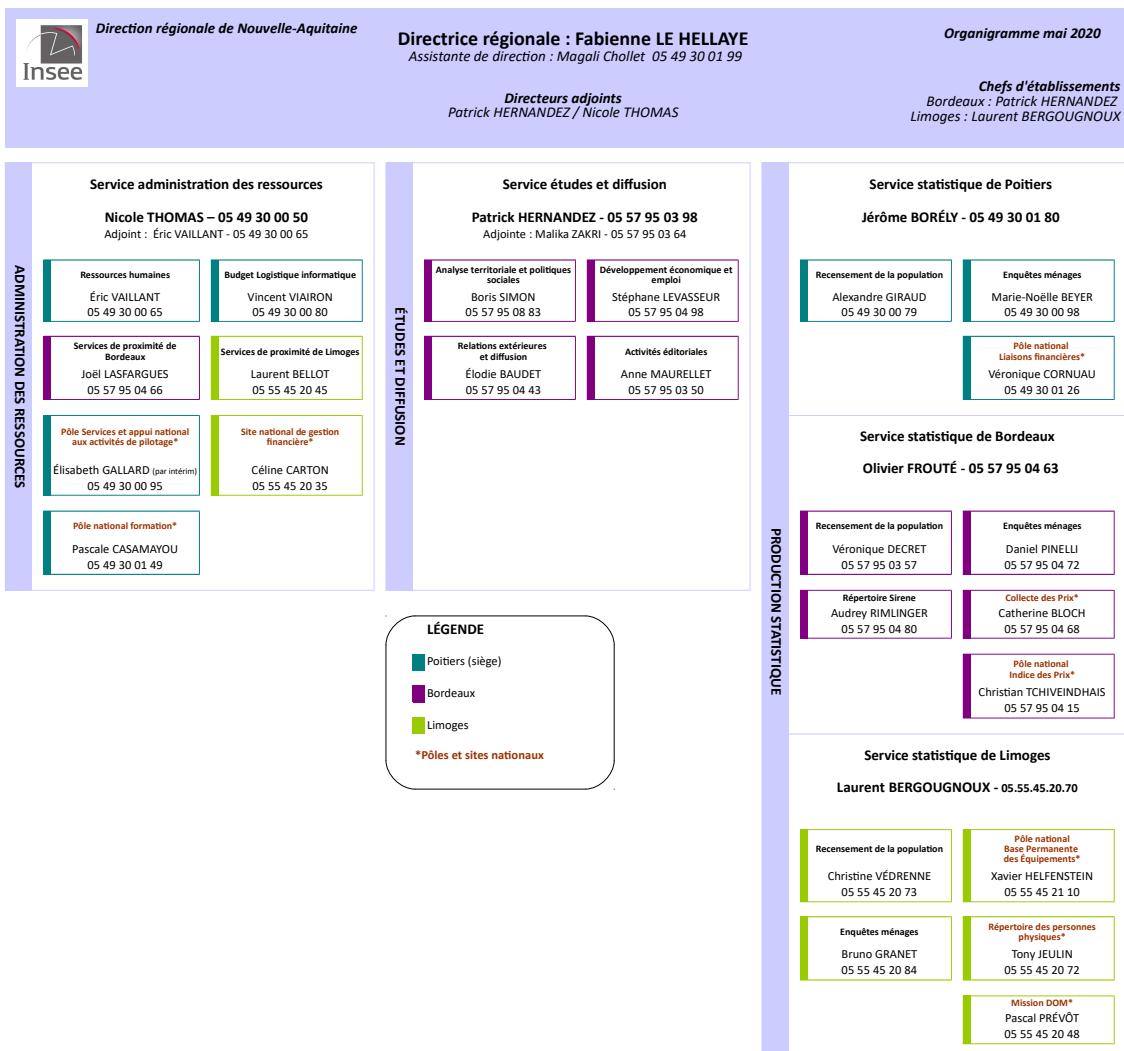
Description de l'environnement de travail et observations

1 Description de l'environnement de travail

Lors de la réforme territoriale de 2015, les régions Poitou-Charentes, Limousin et Aquitaine ont fusionné pour devenir la Nouvelle-Aquitaine. Lors de ce changement, les trois directions régionales se sont vues regroupées en un siège unique à Poitiers, tout en conservant les services de production statistique ainsi que des pôles et missions à compétences nationales à Bordeaux et Limoges. De plus, pour des questions de commodité, les Services Études et Diffusion de chaque nouvelle région sont désormais aux chefs-lieux des régions. C'est le cas du Service Études et Diffusion dans lequel je réalise mon stage, qui se trouve à Bordeaux.

Voici comment s'articule l'ensemble de la direction régionale de Nouvelle-Aquitaine :

FIGURE 1 – Organigramme simplifié de la Direction Régionale de l'Insee de Nouvelle-Aquitaine au 1^{er} mai 2020



SOURCE : Insee, Direction Régionale de Nouvelle-Aquitaine

TABLE 1 – Caractéristiques des agents du SED de Nouvelle-Aquitaine

	Hommes	Femmes	Total
A	11	10	21
B & C	7	9	16
Total	18	19	37

SOURCE : Insee, Sirhius

L’Insee a pour habitude de redéfinir périodiquement ses orientations. En 2015, il a regroupé un ensemble d’orientations sous le nom d’*Insee Horizon 2025* [1].

Les orientations de l’Insee se déclinent en 4 principaux objectifs :

- Innover et être à la pointe des nouvelles sources de données
- Faire preuve d’agilité et de sobriété
- Coordonner la statistique européenne
- Aller au devant de tous les publics

Ces objectifs sont mis en place selon les modèles définis de la gouvernance de la statistique publique et en lien avec le Conseil national de l’information statistique (Cnis).

"Aller au devant de tous les publics" signifie qu'il faut s'adresser à tous et sur tous les sujets. C'est une des principales critiques adressées à l'Insee et constitue un paradoxe. La marque Insee est reconnue de tous mais il serait reproché de ne s'adresser qu'aux experts et pas suffisamment à l'ensemble de la population. Pour s'adresser à tous, il s'agit de s'exprimer de façon claire et accessible mais aussi de ne pas laisser de côté une partie des territoires français.

Par exemple, en 2018 la Direction Régionale de Nouvelle-Aquitaine s'est vu dotée d'une nouvelle mission [2] : celle de consolider le relevé et le traitement de l'information pour les départements et régions d'outre-mer. C'est chose faite puisque l'Insee présente désormais des indicateurs annuels pour les 5 départements ultramarins en plus de ceux pour la métropole [3].

2 À propos du management

Le management de l’Insee repose principalement sur la coordination interne des publications de données, d’indicateurs et de documents, ainsi qu’avec les différents Services Statistiques Ministériels. Le suivi de la qualité est également un élément important du management à l’Insee [4]. C’est l’Unité Qualité à la Direction de la Méthodologie et de la Coordination Statistique et Internationale (DMCSI) qui assure le respect des bonnes pratiques instanciées à l’échelle européenne.

Au sein du service dans lequel j’ai réalisé mon stage, la communication par mail était privilégiée. Dans le contexte du travail à distance, il y a également eu de nombreuses visioconférences via l’outil Zoom, notamment pour réaliser des réunions mais aussi des présentations sur divers sujets. L’intranet, nommé Agora, est également un outil de communication très puissant qui permet la diffusion de l’information à l’échelle de l’institut. Malheureusement, dans le contexte imposé, je n’ai pas eu accès à cette ressource.

3 Analyse de deux indicateurs qualité

3.1 Le secret statistique

Le secret statistique est un ensemble de règles, c'est une forme de secret professionnel auquel les membres de la statistique publique sont tenus lors de la diffusion des exploitations de données [5].

Les règles de diffusion des indicateurs statistiques et des publications font parti du secret statistique. Ces règles de publications reposent sur deux principes : l'annonce préalable du moment exact d'une diffusion et l'égalité de traitement des utilisateurs. Le droit d'accès à une information en avance est restreinte et si un utilisateur extérieur en bénéficie, son identité doit être rendue publique [6].

En outre, certaines informations ne doivent pas être diffusées. Dans l'exemple d'un recensement de logements, des informations propres à une adresse ne doivent pas apparaître au travers des données publiées et ne doivent pas non plus pouvoir être reconstruites par recouplement de différentes sources d'information.

Le secret statistique est bien présent concrètement dans la vie du statisticien public. En effet, en ce qui me concerne, j'ai été amené durant mon stage à manipuler des données dites "sous embargo", c'est-à-dire non diffusables au moment où je les ai exploitées.

3.2 L'impartialité et l'objectivité

En particulier, la façon de recueillir les données et la façon de les traiter ne doivent pas être biaisés par des avis personnels. Pour garantir cela, un contrôle de qualité des indicateurs et procédures utilisés est réalisé par le comité du label [7].

Le comité du label a pour mission de contrôler l'information statistique, produite entre autre par l'Insee et recueillis par le Comité national de l'information statistique (Cnis), qui se charge de mettre en relation les producteurs et consommateurs de statistiques.

Il est ainsi très important que la méthodologie mise en oeuvre lors d'une étude soit communiquée. En ce qui me concerne, la discussion de la méthodologie pour extraire l'information des données était au cœur des discussions avec Rémi Dumas mon maître de stage.

Deuxième partie

Travaux confiés et description des missions

L'ensemble de mes travaux ont porté sur la région Nouvelle-Aquitaine.

Mes travaux ont eu pour vocation de participer à la refonte de la note trimestrielle de conjoncture de la région Nouvelle-Aquitaine.

J'ai principalement étudié l'évolution dans le temps du nombre d'emplois en fonction des secteurs d'activité selon la nomenclature agrégée A17 [8] et en fonction de l'emplacement géographique de ces emplois (parmi les départements de la région Nouvelle-Aquitaine).

Pour cela, j'ai utilisé les données brutes des estimations trimestrielles d'emploi (ETE) par départements qui sont coproduits trimestriellement par l'ACOSS¹, l'Insee, la MSA² et la DARES³.

L'objectif ultime de mon stage fut de trouver des liens entre l'évolution du nombre d'emplois au sein d'un département dans un secteur d'activité et l'évolution du nombre d'emplois au sein d'un autre département et/ou un autre secteur d'activité.

Le découpage géographique utilisé au sein de la région Nouvelle-Aquitaine fut celui des départements.

Dans la nomenclature utilisée, les secteurs d'activité sont au nombre de 17 et possèdent chacun un code Insee.

*** *Nomenclature A17 des secteurs d'activité en annexe* ***

1 L'emploi dans les différents secteurs d'activité au sein des départements

1.1 Une périodicité de l'emploi

La périodicité de l'emploi varie en fonction des secteurs d'activité et des départements observés.

On remarque que, lorsqu'il y a périodicité de l'emploi, elle est souvent annuelle. Toutefois l'emploi concernant certains secteurs d'activité dans certains départements semblent suivre un cycle d'une période de plusieurs années.

Par exemple, en Charente, dans le secteur KZ (des activités financières et d'assurance), on observe ce qui semble être une périodicité de 7 ans. Malheureusement, la durée d'observation dont je disposais était trop courte pour pouvoir l'affirmer.

Les périodicités annuelles sont en revanche plus fréquentes et bien observables. C'est la raison pour laquelle on traite le plus souvent des variations "nettes" de l'emploi, c'est-à-dire en ayant retiré les effets des variations saisonnières.

1. Agence Centrale des Organismes de Sécurité Sociale

2. Mutualité Sociale Agricole

3. Direction de l'Animation, de la Recherche, des Études et des Statistiques

1.2 Une dépendance spatiale dans l'évolution de l'emploi

 Deux évolutions d'emplois dans des départements et/ou secteurs d'activités différents peuvent bien être liées l'une à l'autre mais avec un **décalage temporel**. Par exemple, on peut imaginer que si l'emploi augmente dans l'industrie automobile, l'emploi dans le domaine de la métallurgie augmentera aussi mais avec du retard avec pour cause les stocks de pièces métalliques dont disposent les usines de production automobile.
Il s'agit donc d'en tenir compte pour ne pas passer à côté d'éventuels liens intéressants.

Suite à des observations préalables de l'évolution de l'emploi dans les différents secteurs, j'ai pu constater une singularité de cette évolution dans le secteur d'activité MN0 (activités scientifiques et techniques ; services administratifs et de soutien ; hors intérim). J'ai donc décidé de m'y intéresser plus particulièrement.

J'ai constaté que l'évolution de l'emploi était lié, dans une certaine mesure, au positionnement géographique après avoir créé un indicateur adapté.

En effet, plus des départements sont proches, plus on retrouve des tendances similaires dans l'évolution de l'emploi mais parfois décalées dans le temps.

J'ai tenté d'expliquer ce décalage temporel dans les évolutions d'emploi entre départements mais le lien entre ces décalages temporels et la distance géographique n'était pas significatif.

1.3 Une dépendance de l'emploi entre secteurs d'activités

On peut distinguer certains groupes de secteurs d'activité au sein desquels l'évolution de l'emploi suit globalement les mêmes tendances aux mêmes périodes. De même, il existe des paires de secteurs d'activité dont les tendances s'opposent.

 Parmi deux secteurs d'activité dont l'emploi suit les mêmes tendances aux mêmes périodes, l'un des deux n'est pas forcément à l'origine des tendances de l'autre. En effet, il peut exister une cause commune aux évolutions concomitantes de l'emploi dans ces deux secteurs sans pour autant qu'un lien direct n'existe.

Prenons l'exemple d'une usine d'emballage de lait et d'une usine de production d'une boisson lactée. Si une catastrophe sanitaire bovine survient, la production de lait sera touchée et la prix de la matière première de ces deux usines augmentera. Le prix de vente de ces deux produits va augmenter et ainsi la consommation globale de ces produits diminuera. Par suite, une diminution du nombre d'emplois sera visible dans ces deux usines. La tendance observée sur l'emploi suite à ce choc dans les deux usines sera la même sans que la cause ne vienne d'une de ces deux usines. Il peut en être de même à l'échelle des secteurs d'activité.

On peut distinguer deux groupes de secteur d'activité dont les évolutions sont comparables au sein de chacun des groupes.

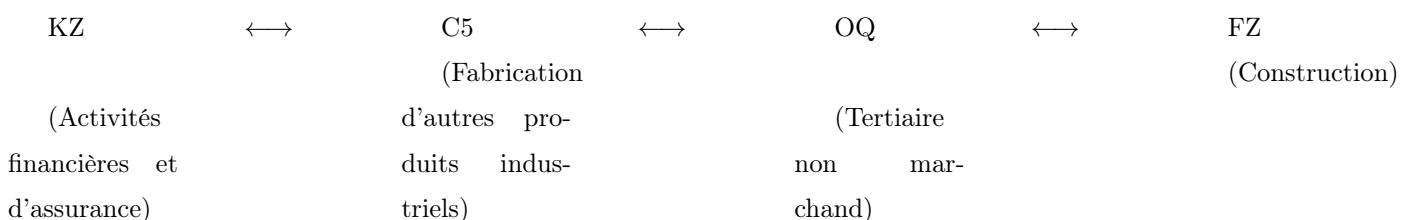
GROUPE 1 :

- C3 (Fabrication d'équipements électriques, électroniques, informatiques ; fabrication de machines)
- C5 (Fabrication d'autres produits industriels)
- FZ (Construction)

GROUPE 2 :

- KZ (Activités financières et d'assurance)
- LZ (Activités immobilières)
- GZ (Commerce ; réparation d'automobiles et de motocycles)
- JZ (Information et communication)
- MN0 (Activités scientifiques et techniques ; services administratifs et de soutien ; hors intérim)

On peut également distinguer des paires de secteurs d'activité dont les évolutions d'emplois s'opposent suivant le schéma suivant : (\longleftrightarrow pour une opposition entre deux secteurs)



Prolongement possible : Analyser ces évolutions au regard du taux directeur de l'économie.

2 Analyse de secteurs d'activité à l'échelle régionale

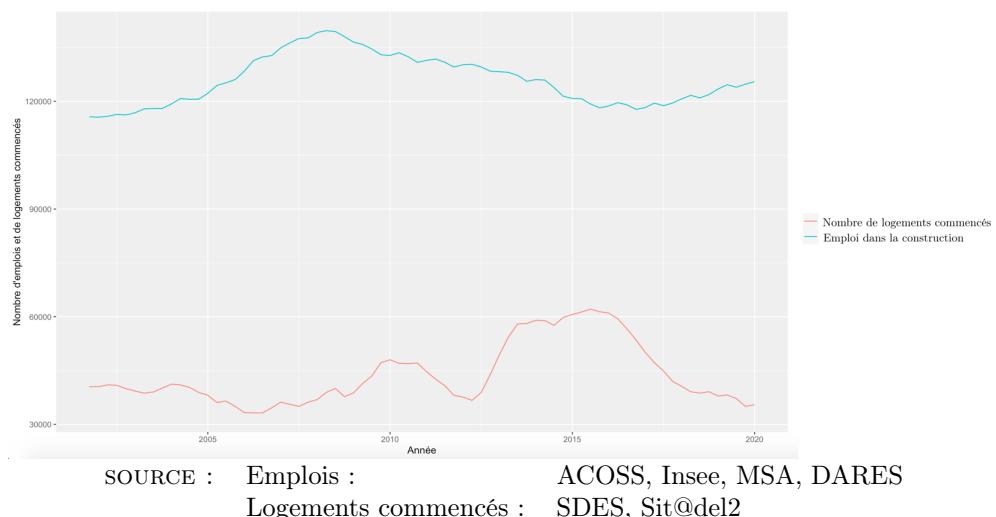
L'emploi est désormais considéré à l'échelle régionale et non plus départementale.

2.1 Emploi dans la construction

Toujours en cherchant à expliquer les évolutions d'emplois, je me suis intéressé au secteur FZ (Construction).

Le nombre de logements dont la construction est commencée semble à priori fortement lié au nombre d'emplois dans ce secteur.

FIGURE 2 – Évolution de l'emploi dans la construction et nombre de logements commencés



On pourrait penser que le nombre de logements commencés suivrait globalement les mêmes tendances que l'emploi dans le secteur de la construction avec un retard des logements commencés sur l'emploi.

On décale horizontalement les deux courbes l'une par rapport à l'autre afin de trouver la position dans laquelle ces deux courbes présenteraient le lien le plus fort.

Après analyse, il n'en est rien. Je n'ai donc pu tirer aucune conclusion.

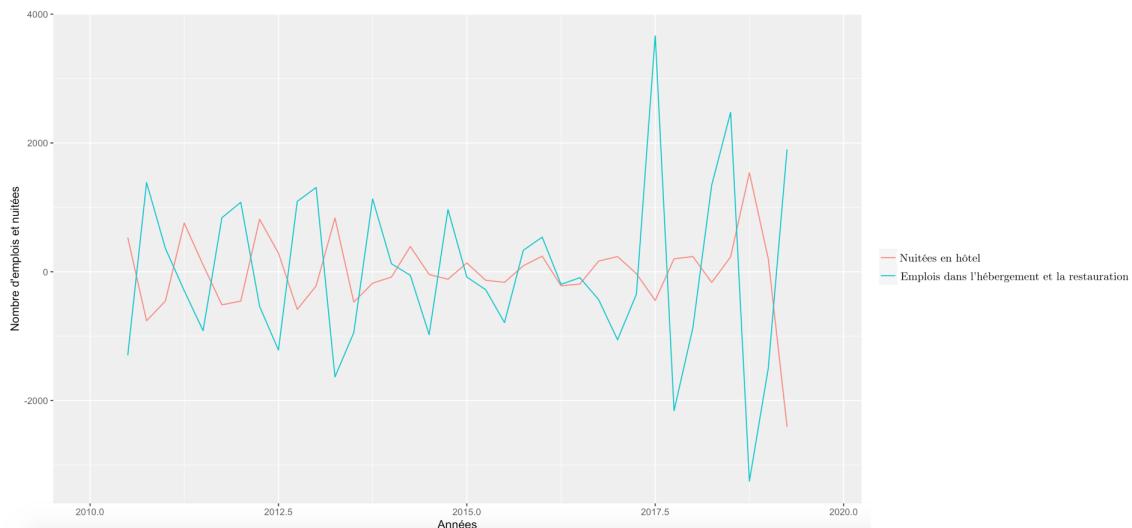
2.2 Un lien entre l'emploi dans l'hébergement et la restauration et le nombre de nuitées dans l'hôtellerie

De même que précédemment, on peut penser que l'évolution de certains paramètres serait liée à l'évolution de l'emploi dans le secteur IZ (Hébergement et restauration). En particulier, le nombre de nuitées dans l'hôtellerie.

On remarque premièrement une forte saisonnalité de l'emploi dans ce secteur et dans le nombre de nuitées.

Après avoir retiré les effets de saisonnalités et la tendance générale de ceux deux courbes, on obtient par superposition la figure suivante :

FIGURE 3 – Évolution de l'emploi dans l'hébergement/restauration et nombre de nuitées dans l'hotellerie (une fois retirés la tendance et les effets de saisonnalité)



SOURCE : Emplois : ACROSS, Insee, MSA, DARES
Logements commencés : SDES, Sit@del2

Les deux courbes semblent suivre des tendances similaires mais avec un retard d'environ 6 mois des nuitées en hôtel sur l'emploi dans ce secteur. Ce résultat se confirme grâce à une analyse ad-hoc.

 Sur ce graphique, le nombre d'emplois et le nombre de nuitées sont parfois négatifs. Cela est dû au fait qu'il ne s'agit pas du nombre total d'emplois (*resp.* nombre total de nuitées) mais seulement de la part que l'on n'explique ni par une tendance générale ni par un phénomène de saisonnalité.

En effet, si à une date donnée le nombre d'emplois attendu au vu de la tendance générale et du phénomène de saisonnalité est supérieur au nombre d'emplois réellement observé, il apparaîtra une valeur négative pour l'emploi à cette date sur la courbe ci-dessus. Il en est de même pour le nombre de nuitées.

NB : La part de l'emploi et du nombre de nuitées que l'on représente sur ces courbes est appellé *résidus*.

Troisième partie

Ressenti de l'expérience et prise de recul

1 Ressenti de l'expérience

Lorsque j'ai commencé ce stage, j'ai tout d'abord été surpris par la liberté qui m'a été accordée : il m'a été laissé le choix des données exploitées, des méthodes employées, des pistes de recherches. On m'a demandé d'innover, de surprendre même ! Je ne m'attendais pas à cela, mais lorsque la réussite est au rendez-vous, la satisfaction est d'autant plus grande si nous sommes l'instigateur de nos propres recherches.

L'Insee attache une importance toute particulière à sa ressource humaine car la taille réduite, compte tenu du maillage complet du territoire, ne permet pas de déperdition de compétences. Ainsi, une grande attention est portée aux agents lors des mobilités fonctionnelles afin que l'adéquation entre compétences et missions soit optimale.

L'Insee s'affranchit clairement du cliché de l'entreprise froide où le climat tendu entre les employés est maintenu par un directeur tyranique. L'ambiance est détendue mais pour autant chacun connaît son travail et sa place au sein de l'institut. Lorsqu'il m'a été donné de participer à un "tour de table" (c'est une réunion de concertation, pour faire le point sur les travaux en cours et ceux à venir), j'ai pu constater la redoutable efficacité des prises de paroles successives, traitant largement d'un grand nombre de sujets en un temps record.

Je n'ai aucun regret hormis le fait que j'ai dû laisser des pistes d'explorations de côté par manque de temps.

Je n'ai rencontré aucune difficulté extérieure au stage. Pour des raisons personnelles j'ai décidé de rester chez moi à Bruz pendant la période de confinement et la période du stage. J'ai donc continué de vivre normalement, rien n'a changé pour moi durant cette période. Certes, j'ai été confronté à la solitude puisque tous mes amis sont rentrés dans leurs familles respectives, loin de l'Ensaï, mais cela ne m'a pas gêné outre mesure.

2 Prise de recul / analyse d'expérience

Ce stage a parfaitement répondu à mes attentes initiales et même plus encore. Je souhaitais découvrir l'environnement de travail du statisticien public. Il m'a été donné de participer à des réunions, de partager le quotidien des agents de l'Insee. Le travail en distanciel ne m'a donc pas empêché de bénéficier d'une immersion totale.

Je ressors de cette expérience avec une maîtrise du logiciel R nettement supérieure à ce qu'elle était auparavant puisque ce fut mon outil de travail pendant toute la période du stage. Grâce à mon maître de stage, j'ai découvert de nombreuses fonctionnalités et syntaxe particulièrement adaptées au traitement de données dont je ne soupçonnais même pas l'existence. Ce surplus d'expérience me sera fort utile à l'avenir car R est un logiciel d'usage permanent dans le traitement de données.

J'ai pu mettre en oeuvre mon savoir-faire statistique à de nombreuses reprises. Notamment lorsqu'il s'agissait d'étudier la relation entre un éloignement spatial entre départements et une similitude des séries temporelles d'emplois à décalage temporel près. J'ai pu effectuer de nombreux tests statistiques pour étudier la significativité de différents résultats obtenus, en terme de corrélation par exemple.

Dans le cadre de relations professionnelles, le savoir-être passe bien-sûr par le respect d'autrui et notamment une communication polie et courtoise. En outre, chacun travaille sur un sujet qui peut être très différent de celui sur lequel nous sommes à l'oeuvre. Il est donc nécessaire de ne pas solliciter de manière intempestive ses collègues de manière à préserver un climat serein et studieux. Dans mon cas, lorsque j'avais des questions ou une remarque à soumettre

à mon maître de stage ou à tout autre membre du personnel, je veillais à attendre la fin de journée pour regrouper mes interventions et ainsi ne pas interférer négativement avec leur travail.

Un élève intéressé par ce stage l'année prochaine devrait savoir qu'il peut être intéressant de s'avancer sur le programme de deuxième année pour aborder en particulier l'analyse des séries temporelles. En dehors de ce conseil pratique, rien de particulier à mes yeux si ce n'est que le stage est fort intéressant et exaltant car il traite de sujets d'intérêt pratique comme les dépendances spacio-temporelles et sectorielles de l'emploi.

Annexe

1 Informations complémentaires

1.1 Départements de Nouvelle-Aquitaine



SOURCE : www.ccomptes.fr

16 Charente

17 Charente-Maritime

19 Corrèze

23 Creuse

24 Dordogne

33 Gironde

40 Landes

47 Lot-et-Garonne

64 Pyrénées-Atlantiques

79 Deux-Sèvres

86 Vienne

87 Haute-Vienne

1.2 Secteurs d'activité

TABLE 2 – Nomenclature A17 des secteurs d'activité

Code	Secteur d'activité
AZ	Agriculture
C1	Fabrication de denrées alimentaires, de boissons et de produits à base de tabac
C2DE	Cokéfaction et raffinage ; industries extractives, énergie, eau, gestion des déchets et dépollution
C3	Fabrication d'équipements électriques, électroniques, informatiques ; fabrication de machines
C4	Fabrication de matériels de transport
C5	Fabrication d'autres produits industriels
FZ	Construction
GZ	Commerce ; réparation d'automobiles et de motocycles
HZ	Transports et entreposage
IZ	Hébergement et restauration
JZ	Information et communication
KZ	Activités financières et d'assurance
LZ	Activités immobilières
MN0	Activités scientifiques et techniques ; services administratifs et de soutien ; hors intérim
MN78	Intérim
RU	Autres activités de services
OQ	Tertiaire non marchand

2 Grille d'entretiens

Voici les points principaux qui ont été abordés durant les différents entretiens :

Présentation de la division, des personnes et des moyens
Cadre institutionnel et légal et environnement professionnel
Missions du service/de la division
Insertion dans l'institut et évolutions futures
Spécificités et points divers
Ressenti personnel
Avenir professionnel

3 Code

3.1 Analyse exploratoire-v1

```
## Emploi salarié en fin de trimestre - Industries manufacturières, industries extractives et autres (A5-E)

X1 <- read.csv("./données/Emplois et chômage/emplois_industrie_manufacturière.csv", header=TRUE, sep=";")

X1 <- as.data.frame(X1)
for (i in 1:length(X1)){
  for (j in 1:length(X1[i])){
    X1[i,j] <- as.numeric(X1[i,j])
  }
}
X1 <- X1[4:length(X1[,1]),]
```

```

colnames(X1)[2] <- "donnee"

## Emploi salarié en fin de trimestre - Fabrication de denrées alimentaires, de boissons et de produits à

X2 <- read.csv("./données/Emplois et chômage/emplois_denrees_alimentaires.csv", header=TRUE, sep=";")
X2 <- as.data.frame(X2)

for (i in 1:length(X2)){
  for (j in 1:length(X2[i,]))
    X2[i,j] <- as.numeric(X2[i,j])
}

X2 <- X2[4:length(X2[,1]),]
colnames(X2)[2] <- "donnee"

## Emploi salarié en fin de trimestre - Cokéfaction et raffinage ; industries extractives, énergie, eau, g

X3 <- read.csv("./données/Emplois et chômage/emplois_energie.csv", header=TRUE, sep=";")
X3 <- as.data.frame(X3)
for (i in 1:length(X3)){
  for (j in 1:length(X3[i,]))
    X3[i,j] <- as.numeric(X3[i,j])
}

X3 <- X3[4:length(X3[,1]),]
colnames(X3)[2] <- "donnee"

## Emploi salarié en fin de trimestre - Fabrication de matériels de transport (A17-C4) - France hors Mayot

X4 <- read.csv("./données/Emplois et chômage/emplois_materiel_de_transport.csv", header=TRUE, sep=";")
X4 <- as.data.frame(X4)
for (i in 1:length(X4)){
  for (j in 1:length(X4[i,]))
    X4[i,j] <- as.numeric(X4[i,j])
}

```

```

X4 <- X4[4:length(X4[,1]),]
colnames(X4)[2] <- "donnee"

plot(X1$donnee, type="l")
plot(X2$donnee, type="l")
plot(X3$donnee, type="l")
plot(X4$donnee, type="l")

ks.test(X1$donnee, "pnorm", mean(X1$donnee), sd(X1$donnee))

## Distribution empirique de la loi de X1$donnee

x <- X1$donnee
?rep
n <- 30
m <- min(x)
M <- max(x)
f <- c()
for (i in 1:(n-1)){
  f <- append(f,length(x[which(x >= m+i*(M-m)/n & x < m+(i+1)*(M-m)/n])))
}
plot(f, type="l")

install.packages("stringr", dependencies=TRUE)
library(stringr)
install.packages("FactoMineR")
library(FactoMineR)
library(missMDA)

## Emploi salarié en fin de trimestre - Hébergement et restauration (A17-I2) - Nouvelle-Aquitaine

X1 <- read.csv("./données/Emplois et chômage/emplois_hebergement_et_restauration.csv", header=TRUE, sep=",")
X1 <- as.data.frame(X1)
for (i in 1:length(X1)){
  for (j in 1:length(X1[i,]))
```

```

X1[i,j] <- as.numeric(X1[i,j])
}
X1 <- X1[4:length(X1[,1]),]
colnames(X1)[2] <- "donnee"

## Frequentation touristique (nuitees, arrivees)

X2 <- read.csv("./donnees/Tourisme/frequentation_touristique.csv", header=TRUE, sep=";")
X2$Libellé <- as.character(X2$Libellé)
X2 <- X2[which(str_detect(X2$Libellé, "Nouvelle-Aquitaine")),]
X <- X2[,4:length(X2[1])]

X <- X2[,4:length(X2[1])]
X <- as.data.frame(X)
tmp <- data.frame(matrix(rep(0,length(X[1]))*length(X[,1])), nrow = length(X[,1]))
for (i in 1:length(X[,1])){
  for (j in 1:length(X[1])){
    tmp[i,j] <- as.numeric(as.character(X[i,j]))
  }
}
X <- tmp

## Taux de chomage localise

X3 <- read.csv("./donnees/Emplois et chômage/emplois_industrie_manufacturiere.csv", header=TRUE, sep=";")
X3 <- X3[which(str_detect(X3$Libellé, "Nouvelle-Aquitaine")),]
X3 <- X3[,84:155]
tmp <- c()
for (i in 1:length(X3[1])){
  tmp <- append(tmp,as.numeric(as.character(X3[1,i])))
}
X3 <- tmp

## Taux de chômage en Nouvelle-Aquitaine et Emplois dans le secteur de l'Hébergement-Restauration en Nouve

plot.new()
par(mar=c(4,4,3,5))
plot(X3,col="blue",axes=F,xlab="",ylab="")
axis(2, ylim=c(0,10),col="blue",col.axis="blue") #,at=seq(0, 10, by=2))
mtext("Taux de chômage en Nouvelle-Aquitaine",side=2,line=2.5,col="blue")
par(new = T)

```

```

plot(X1$donnee,col="red",axes=F,xlab="",ylab "") #,ylim=c(20,40))
axis( 4 ,col="red",col.axis="red") #,at=seq(20, 40, by=5))
mtext("Emplois dans le secteur de l'Hébergement-Restauration en Nouvelle-Aquitaine",side=4,line=2.5,col="red")
axis( 1 , ylim=c(20,40),col="black",col.axis="black") #,at=seq(0, 40, by=1))
mtext("Trimestres",side=1,line=2.5,col="black"))

L1 <- c()
L2 <- c()
for (i in 2:length(X3)){
  L1 <- append(L1,X3[i]-X3[i-1])
  L2 <- append(L2,X1$donnee[i]-X1$donnee[i-1])
}

plot.new()
par(mar=c(4,4,3,5))
plot(-L1,col="blue",axes=F,xlab="",ylab="", type="l")
axis(2, ylim=c(0,10),col="blue",col.axis="blue") #,at=seq(0, 10, by=2))
mtext("Opposé de l'évolution du taux de chômage en Nouvelle-Aquitaine",side=2,line=2.5,col="blue")
par(new = T)
plot(L2,col="red",axes=F,xlab="",ylab="", type="l") #,ylim=c(20,40))
axis( 4 ,col="red",col.axis="red") #,at=seq(20, 40, by=5))
mtext("Evolution de l'emplois dans le secteur de l'Hébergement-Restauration en Nouvelle-Aquitaine",side=4,col="red")
axis( 1 , ylim=c(20,40),col="black",col.axis="black") #,at=seq(0, 40, by=1))
mtext("Trimestres",side=1,line=2.5,col="black"))

## Test de corrélation entre le taux de chômage et l'emploi dans l'hôtellerie-restauration

rho <- sum((X3-mean(X3))*(X1$donnee-mean(X1$donnee)))/sqrt(sum((X3-mean(X3))^2)*sum((X1$donnee-mean(X1$donnee))^2))

T <- sqrt(length(X3)-2)*rho/sqrt(1-rho^2)

# On trouve une p-valeur du test à 10^-4, ON PEUT CONSIDERER LES ECHANTILLONS DECORRELES

install.packages("esquisse")
library(esquisse)

```

3.2 Analyse exploratoire-v2

3.2.1 Importations et traitement de données

```
rm(list=ls())
```

```

library(dplyr)
source("programmes/fonctions.r", encoding = "UTF-8")
library(stringr)
library(tidyverse)
install.packages("corrplot")
library(corrplot)

# tables de passage
table_dep <- readr::read_csv("tables_passage/departement2019.csv")
table_reg <- readr::read_csv("tables_passage/region2019.csv")
table_secteur <- readr::read_csv2("tables_passage/table_a17.csv")

# chargement des ETE

data <- readxl::read_xls("donnees/brut_dep_a17_2020t1.xls") %>%
  gather("trimestre", "nb_emp", starts_with("eff")) %>%
  select(-REG) %>%
  rename(REG = REG2016) %>%
  mutate(
    periode = str_sub(trimestre,4,9),
    annee = as.numeric(str_sub(periode,1,4)),
    trimestre = str_to_upper(str_sub(periode,5,6))
  )

data75 <- data %>% filter(REG == "75", annee > 2010) %>% group_by(periode) %>%
  summarise(nb_emp = sum(nb_emp, na.rm = T))
plot(data75$nb_emp)

result_acf <- acf(data75$nb_emp)

```

```

# chargement des communes (ancienne version)
communes <- read.csv("donnees/communes/communes_france.csv", header=FALSE)
colnames(communes) <- c("id", "dept", "nom_ville_minusc_tirets", "nom_ville_majusc_tirets", "nom_ville_minusc_")
communes <- communes %>% mutate(code_postal = as.character(code_postal)) %>% mutate(code_postal = as.numer
communes <- communes %>% group_by(dept) %>% mutate(popul_relative = population_2010/sum(population_2010))

# Chargement de la base ville
villes <- readxl::read_xls("donnees/ville.xls")

# chargement des communes (nouvelle version)
communes_v2 <- read.csv("donnees/communes/communes_v2.csv", sep = ";", header=TRUE)
communes_v2 <- communes_v2 %>% mutate(code_insee = DEPCOM, nom = COM, population = PTOT) %>% select(-PMUN,
communes_v2 <- communes_v2 %>% mutate(code_insee = code_insee %>% as.character() %>% as.numeric())
communes_v2 <- communes_v2 %>% mutate(dept = as.character(floor(as.numeric(communes_v2$code_insee)/1000)))
communes_v2 <- communes_v2 %>% group_by(dept) %>% mutate(popul_relative = population/sum(population))
communes_v2 <- communes_v2 %>% mutate(longitude = villes$Longitude, latitude = villes$Latitude)

#####
df <- tibble(a = c("3", "4", "5"))
b <- 2

data2 <- data %>%
  group_by(Région) %>%
  mutate(Max2019 = max(eff2019t4)) %>%
  select(Région, Département, eff2019t4, Max2019)

#####
ete <- data %>%
  gather(key = "periode", value = "nb_emplois", -REG, -DEP, -A5, -A17) %>%
  mutate(
    Année = str_sub(periode, 4, 7),
    Trimestre = periode %>% str_sub(8,9) %>% str_to_upper
  ) %>%
  select(-periode) %>% mutate(Periode = paste(Année, Trimestre, sep = "-"))

# On restreint le champ car l'emploi public n'est pris en compte qu'à partir de 2010

ete_restreint <- ete %>% filter(A17!="OQ") %>% mutate(Année = as.numeric(Année))

ete_restreint_ap_2010 <- ete_restreint %>% filter(Année > 2010, REG == "75") %>% mutate(Année = as.character(Année))

# secteur MNO dans chaque département

```

```

ete_restraint_ap_2010_MN0_16 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='16')
ete_restraint_ap_2010_MN0_17 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='17')
ete_restraint_ap_2010_MN0_19 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='19')
ete_restraint_ap_2010_MN0_23 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='23')
ete_restraint_ap_2010_MN0_24 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='24')
ete_restraint_ap_2010_MN0_33 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='33')
ete_restraint_ap_2010_MN0_40 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='40')
ete_restraint_ap_2010_MN0_47 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='47')
ete_restraint_ap_2010_MN0_64 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='64')
ete_restraint_ap_2010_MN0_79 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='79')
ete_restraint_ap_2010_MN0_86 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='86')
ete_restraint_ap_2010_MN0_87 <- ete_restraint_ap_2010 %>% filter(A17=='MN0', DEP=='87')

```

```

emp_MN0_16 <- as.list(ete_restraint_ap_2010_MN0_16)$nb_emplois
emp_MN0_17 <- as.list(ete_restraint_ap_2010_MN0_17)$nb_emplois
emp_MN0_19 <- as.list(ete_restraint_ap_2010_MN0_19)$nb_emplois
emp_MN0_23 <- as.list(ete_restraint_ap_2010_MN0_23)$nb_emplois
emp_MN0_24 <- as.list(ete_restraint_ap_2010_MN0_24)$nb_emplois
emp_MN0_33 <- as.list(ete_restraint_ap_2010_MN0_33)$nb_emplois
emp_MN0_40 <- as.list(ete_restraint_ap_2010_MN0_40)$nb_emplois
emp_MN0_47 <- as.list(ete_restraint_ap_2010_MN0_47)$nb_emplois
emp_MN0_64 <- as.list(ete_restraint_ap_2010_MN0_64)$nb_emplois
emp_MN0_79 <- as.list(ete_restraint_ap_2010_MN0_79)$nb_emplois
emp_MN0_86 <- as.list(ete_restraint_ap_2010_MN0_86)$nb_emplois
emp_MN0_87 <- as.list(ete_restraint_ap_2010_MN0_87)$nb_emplois

```

```

communes_du_16 <- communes[which(floor(communes$code_postal/1000)==16),]
communes_du_17 <- communes[which(floor(communes$code_postal/1000)==17),]
communes_du_19 <- communes[which(floor(communes$code_postal/1000)==19),]
communes_du_23 <- communes[which(floor(communes$code_postal/1000)==23),]
communes_du_24 <- communes[which(floor(communes$code_postal/1000)==24),]
communes_du_33 <- communes[which(floor(communes$code_postal/1000)==33),]
communes_du_40 <- communes[which(floor(communes$code_postal/1000)==40),]
communes_du_47 <- communes[which(floor(communes$code_postal/1000)==47),]
communes_du_64 <- communes[which(floor(communes$code_postal/1000)==64),]
communes_du_79 <- communes[which(floor(communes$code_postal/1000)==79),]
communes_du_86 <- communes[which(floor(communes$code_postal/1000)==86),]
communes_du_87 <- communes[which(floor(communes$code_postal/1000)==87),]

```

```

ete_75_C1_ann <- ete %>% filter(REG == "75", A17 == "C1", Trimestre == "T4") %>%

```

```

group_by(DEP) %>%
mutate(
  evol_nb = nb_emplois - lag(nb_emplois),
  evol_tx = round(100 * evol_nb / lag(nb_emplois) , 1)
)

ete_75_C1_ann %>%
select(DEP, Annee, evol_tx) %>%
spread>Annee, evol_tx)

ete2 <- ete %>%
mutate(Periode = paste>Annee, Trimestre, sep = "-"))

#####
# TRACES GRAPHIQUES #####
#####

# Emploi en Nouvelle-Aquitaine entre 2011 et 2019
# Une répartition hétérogène de l'emploi par secteur d'activité
ggplot(ete_restreint_ap_2010) +
  aes(x = A17, y = nb_emplois) +
  geom_boxplot(fill = "#0c4c8a") +
  labs(x = "Secteur", y = "Nombre d'emplois en fin de chaque trimestre") +
  theme_minimal()

#####
# EXPLOITATION DES DONNEES #####
#####

### Recherche de retards et d'avance entre 2 courbes

# Ecart quadratique moyen entre 2 séries de données

d <- function(x,y){
  if (length(x) != length(y)){warning("Longueurs différentes entre 2 séries dont on calcule l'écart quadratique")}
  return(mean((x-y)^2))
}

# Décalage des séries l'une par rapport à l'autre d'au maximum une demi longueur dans les 2 sens

ajust_decal_series <- function(x,y){


```

```

if (length(x) != length(y)){warning("Longueurs différentes entre 2 séries qu'on cherche à comparer")}
n <- length(x)-1
tab <- (-n):n
vect <- c()
for (k in n:0){
  vect <- append(vect, d(y[(k+1):length(y)], x[1:(length(x)-k)])))
}
for (k in 1:n){
  vect <- append(vect, d(y[1:(length(y)-k)], x[(k+1): length(x)])))
}
cbind(tab,vect)
m <- min(vect)
return(c(tab[which(vect == m)],log(m/10000)))
}

d_communes <- function(nom1,nom2){
  com1 <- communes [which(communes$nom_ville_minusc_tirets == nom1),]
  com2 <- communes [which(communes$nom_ville_minusc_tirets == nom2),]
  tmp <- (com1$longitude_deg*cos(pi*com1$latitude_deg/180)-com2$longitude_deg*cos(pi*com2$latitude_deg/180))
  return(pi*6371.009*sqrt(tmp)/180)
}

d_communes_v2 <- function(nom1,nom2){
  com1 <- communes_v2 [which(communes_v2$nom_ville_minusc_tirets == nom1),]
  com2 <- communes_v2 [which(communes_v2$nom_ville_minusc_tirets == nom2),]
  tmp <- (com1$longitude_deg*cos(pi*com1$latitude_deg/180)-com2$longitude_deg*cos(pi*com2$latitude_deg/180))
  return(pi*6371.009*sqrt(tmp)/180)
}

d_dep <- function(num1, num2){
  dep1 <- communes %>% filter(dept == num1)
  dep2 <- communes %>% filter(dept == num2)
  longitude1 <- sum(dep1$longitude_deg*dep1$popul_relative)
  latitude1 <- sum(dep1$latitude_deg*dep1$popul_relative)
  longitude2 <- sum(dep2$longitude_deg*dep2$popul_relative)
  latitude2 <- sum(dep2$latitude_deg*dep2$popul_relative)
  tmp <- (longitude1*cos(pi*latitude1/180)-longitude2*cos(pi*latitude2/180))^2+(latitude1-latitude2)^2
  return(pi*6371.009*sqrt(tmp)/180)
}

d_dep_v2 <- function(num1, num2){
  dep1 <- communes_v2 %>% filter(dept == num1)
  dep2 <- communes_v2 %>% filter(dept == num2)

```

```

longitude1 <- sum(dep1$longitude*dep1$popul_relative)
latitude1 <- sum(dep1$latitude*dep1$popul_relative)
longitude2 <- sum(dep2$longitude*dep2$popul_relative)
latitude2 <- sum(dep2$latitude*dep2$popul_relative)
tmp <- (longitude1*cos(pi*latitude1/180)-longitude2*cos(pi*latitude2/180))^2+(latitude1-latitude2)^2
return(pi*6371.009*sqrt(tmp)/180)

}

dist <- c()
vect <- c("16", "17", "19", "23", "24", "33", "40", "47", "64", "79", "86", "87")
for (i in 1:11){
  for (j in (i+1):12){
    dist <- dist %>% append(d_dep(vect[i], vect[j]))
  }
}
}

dist_v2 <- c()
for (i in 1:11){
  for (j in (i+1):12){
    dist_v2 <- dist_v2 %>% append(d_dep_v2(vect[i], vect[j]))
  }
}
}

prox <- c(8.32,2.27,7.3,2.44,12.39,1.17,2.38,9.62,5.78,7.98,4.48,8.47,9.57,8.52,11.98,8.57,8.21,6.98,6.87

table_emplois_distance <-
  data.frame(dep=c("16_17", "16_19","16_23", "16_24", "16_33", "16_40", "16_47", "16_64", "16_79", "16_86"))

rho_ecart <- cor(table_emplois_distance$dist, table_emplois_distance$ecart_temp)
T_ecart <- 8*rho_ecart/sqrt(1-rho_ecart^2)
p_val_ecart <- 2*(1-pt(abs(T_ecart), 65))

rho_prox <- cor(table_emplois_distance$dist, table_emplois_distance$prox)
T_prox <- 8*rho_prox/sqrt(1-rho_prox^2)
p_val_prox <- 2*(1-pt(abs(T_prox), 65))

rho_ecart_v2 <- cor(table_emplois_distance$dist_v2, table_emplois_distance$ecart_temp)
T_ecart_v2 <- 8*rho_ecart/sqrt(1-rho_ecart_v2^2)
p_val_ecart_v2 <- 2*(1-pt(abs(T_ecart_v2), 65))

```

```

rho_prox_v2 <- cor(table_emplois_distance$dist_v2, table_emplois_distance$prox)
T_prox_v2 <- 8*rho_prox_v2/sqrt(1-rho_prox_v2^2)
p_val_prox_v2 <- 2*(1-pt(abs(T_prox_v2), 65))

# Corrélogramme par secteur

emplois_secteur <- readxl::read_xls("donnees/brut_dep_a17_2020t1.xls") %>%
  select(A17, REG2016, eff2011t1, eff2011t2, eff2011t3, eff2011t4, eff2012t1, eff2012t2, eff2012t3, eff2012t4) %>%
  rename(REG=REG2016) %>%
  filter(REG == "75") %>%
  select(-REG) %>%
  group_by(A17) %>%
  mutate(eff2011t1 = sum(eff2011t1), eff2011t2 = sum(eff2011t2), eff2011t3 = sum(eff2011t3), eff2011t4 = sum(eff2011t4))

emplois_secteur <- emplois_secteur[1:17,]

col <- colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))

M <- t(emplois_secteur[,2:length(emplois_secteur[1,])])
secteurs <- c()
for (i in 1:17){
  secteurs <- secteurs %>% append(as.character(emplois_secteur[i,1]))
}
colnames(M) <- secteurs

corrplot(cor(M), method="color", col=col(200),
         type="upper", order="hclust",
         addCoef.col = "black", # Ajout du coefficient de corrélation
         tl.col="black", tl.srt=45, #Rotation des étiquettes de textes
         # Combiner avec le niveau de significativité
         p.mat = NULL, sig.level = 0.01, insig = "blank",
         # Cacher les coefficients de corrélation sur la diagonale
         diag=FALSE
)

# Corrélogramme de décalage temporel par secteur

tab <- c()

ACF <- function(L){


```

```

tmp <- c()
for (i in 1:(length(L)-3)){tmp <- append(tmp, cor(L[1:(length(L)-i)], L[(i+1):length(L)]))}
return(tmp)
}

for (i in 1:length(M[,])){tab <- rbind(tab, ACF(M[,i]))}

row.names(tab) <- secteurs
colnames(tab) <- dec_temp

corrplot(tab, tl.srt=45)

# Corrélogramme de décalage temporel par département

emplois_dep <- readxl::read_xls("donnees/brut_dep_a17_2020t1.xls") %>%
  select(A17, DEP, REG2016, eff2011t1, eff2011t2, eff2011t3, eff2011t4, eff2012t1, eff2012t2, eff2012t3, e
  rename(REG=REG2016) %>%
  filter(REG == "75") %>%
  select(-REG) %>%
  group_by(DEP) %>%
  mutate(eff2011t1 = sum(eff2011t1), eff2011t2 = sum(eff2011t2), eff2011t3 = sum(eff2011t3), eff2011t4 = s

tmp <- c()
dep <- c("16", "17", "19", "23", "24", "33", "40", "47", "64", "79", "86", "87")
for (x in dep){
  tmp_dep <- emplois_dep[which(emplois_dep$DEP == x),]
  tmp <- tmp %>% rbind(tmp_dep[1,] %>% select(-A17))
}

tmp_2 <- tmp[,2:length(tmp[1,])]
rownames(tmp_2) <- as.character(tmp[,1])

tmp_3 <- c()
for (i in 1:12){
  tmp_3 <- tmp_3 %>% append(tmp[i,1]$DEP)
}

rownames(tmp_2) <- tmp_3

emplois_dep <- t(tmp_2)

tab <- c()

for (i in 1:length(emplois_dep[,])){tab <- rbind(tab, ACF(emplois_dep[,i]))}

```

```

dec_temp <- c()
for (i in 1:34){
  dec_temp <- dec_temp %>% append(str_c("T+",as.character(i)))
}

row.names(tab) <- dep
colnames(tab) <- dec_temp

corrplot(tab, tl.srt=45)

# Corrélogramme de décalage temporel par département*secteur

emplois_dep_sec <- readxl::read_xls("donnees/brut_dep_a17_2020t1.xls") %>%
  select(A17, DEP, REG2016, eff2011t1, eff2011t2, eff2011t3, eff2011t4, eff2012t1, eff2012t2, eff2012t3, e
  rename(REG=REG2016) %>%
  filter(REG == "75") %>%
  select(-REG)

tmp_2 <- c()
for (i in 1:nrow(emplois_dep_sec)){
  tmp_2 <- tmp_2 %>% rbind(emplois_dep_sec[i,] %>% select(-A17, -DEP))
}
tmp_2 <- as.data.frame(tmp_2)

tmp_3 <- c()
for (i in 1:nrow(tmp_2)){
  tmp_3 <- tmp_3 %>% rbind(as.numeric(tmp_2[i,]))
}

tab <- c()

for (i in 1:nrow(tmp_3)){tab <- rbind(tab, ACF(tmp_3[i,]))}

dep_sec <- c()
for (x in dep){
  for (y in secteurs){
    dep_sec <- dep_sec %>% append(str_c("dep", x, ", ", y))
  }
}

```

```

row.names(tab) <- dep_sec
colnames(tab) <- dec_temp

ind <- 17*(0:11)+1

rch_ind <- function(L, x){
  i <- 1
  while (x != L[i] & i <= length(L)){
    i <- i+1
  }
  if (i == length(L)+1){return(NA)}
  return(i)
}

departement <- readline(prompt="Département : ")

i <- rch_ind(dep, departement)
corrplot(tab[ind[i]]:(ind[i+1]-1), , t1.srt=45)

```

3.2.2 Fonctions

```

# Installation et chargement de packages
install_and_load <- function(P) {
  Pi <- P[!(P %in% installed.packages() [, "Package"])]
  if (length(Pi)>0) install.packages(Pi)
  for(i in P) library(i, character.only = TRUE)
}

```

```

packages <-
  c("downloader",
    "glue",
    "hms",
    "httr",
    "lubridate" ,
    "readxl",
    "rio" ,
    "rlang" ,
    "tidyverse"
  )

```

```
install_and_load(packages)
```

3.3 Analyse exploratoire-v3

```
# chargement des packages
library(dplyr)
if (!("tidyverse" %in% installed.packages() [, "Package"])) install.packages("tidyverse")
library(tidyverse)
if (!("lubridate" %in% installed.packages() [, "Package"])) install.packages("lubridate")
library(lubridate)
if (!("forecast" %in% installed.packages() [, "Package"])) install.packages("forecast")
library(forecast)
if (!("ggfortify" %in% installed.packages() [, "Package"])) install.packages("ggfortify")
library(ggfortify)
if (!("gtools" %in% installed.packages() [, "Package"])) install.packages("gtools")
library(gtools)

# fonction notin
"%notin%" <- Negate("%in%")

table_A17 <- read_csv2("donnees/logements/tables_passage/table_A17.csv")

### emploi dans la construction ###
emploi_FZ <- readxl::read_xls("donnees/logements/brut_dep_a17_2020t1.xls") %>%
gather(-REG2016,-REG,-DEP,-A5,-A17, key = "trimestre", value = "nb_emp") %>%
mutate(
  periode = str_sub(trimestre, 4,9),
  annee = str_sub(periode,1,4),
  trimestre = str_sub(trimestre, 9,9)
) %>%
select(-REG) %>% rename(REG = REG2016) %>%
arrange(desc(annee), REG, DEP) %>%
filter(A17 == "FZ") %>%
filter(REG == "75") %>%
group_by(periode) %>%
summarise(nb_emp = as.numeric(sum(nb_emp, na.rm = T)))

### nbre de logements commencés dans la région ###
log_com <- read_csv2("donnees/logements/log_com.csv") %>% #https://www.insee.fr/fr/statistiques/serie/0017
select(-Codes, periode = Libellé, logcom = 'Nombre de logements commencés - Cumul sur douze mois - Total'
slice(-c(1:2)) %>%
```

```

separate periode, c("annee", "mois"), sep = "-") %>%
filter(mois %in% c("03", "06", "09", "12")) %>%
mutate(
  periode = case_when(
    mois == "03" ~ paste0(annee, "t1"),
    mois == "06" ~ paste0(annee, "t2"),
    mois == "09" ~ paste0(annee, "t3"),
    mois == "12" ~ paste0(annee, "t4"),
    TRUE ~ "Erreur !"
  ),
  logcom = as.numeric(logcom)
) %>% select(periode, logcom)

# periodes communes #
per_list <- intersect(log_com$periode, emploi_FZ$periode)
per_start <- c(per_list %>% last() %>% str_sub(., 1, 4), per_list %>% last() %>% str_sub(., 6, 6)) %>% as.numeric
# series temporelles, ça commence au T4 2001 et c'est par trimestres #
ts_logcom <- log_com %>%
  filter(periode %in% per_list) %>%
  select(logcom) %>%
  ts(frequency = 4, start = per_start)
ts_emploi_FZ <- emploi_FZ %>%
  filter(periode %in% per_list) %>%
  select(nb_emp) %>%
  ts(frequency = 4, start = per_start)

# visualisation des deux séries #
forecast::autoplot(cbind(ts_logcom, ts_emploi_FZ))

# Cross correlation entre ts_logcom et ts_emploi_FZ #
forecast::Ccf(as.numeric(ts_logcom), as.numeric(ts_emploi_FZ))

# Correlation glissante, fenetre de 4 trimestres #
corr_FZ <- function(width){
  gtools::running(ts_logcom, ts_emploi_FZ, fun=cor, width = width)
}

corr_FZ(4) %>%
  enframe() %>% summarise(moyenne = mean(value, na.rm = T))

autoplot(cbind(ts_logcom, stats::lag(ts_emploi_FZ, -18)))

# Cross correlation entre ts_logcom et ts_emploi_FZ #
forecast::Ccf(as.numeric(ts_logcom), as.numeric(ts_emploi_FZ), type = "correlation", xlab = "toto", main =

```

```

ts_emploi_FZ_decal <- stats::lag(ts_emploi_FZ, -18)
# periodes communes #
per_list <- intersect(log_com$periode ,ts_emploi_FZ_decal )

ts_emploi_FZ_decal <- ts(ts_emploi_FZ_decal,frequency = 4, start = c(2006,2) )

ts_emploi_FZ_decal$periode

#### emploi dans la'hébergement-restauration ####
emploi_IZ <- readxl::read_xls("donnees/logements;brut_dep_a17_2020t1.xls") %>%
  gather(-REG2016,-REG,-DEP,-A5,-A17, key = "trimestre", value = "nb_emp") %>%
  mutate(
    periode = str_sub(trimestre, 4,9),
    annee = str_sub(periode,1,4),
    trimestre = str_sub(trimestre, 9,9)
  ) %>%
  select(-REG) %>% rename(REG = REG2016) %>%
  arrange(desc(annee), REG, DEP) %>%
  filter(A17 == "IZ") %>%
  filter(REG == "75") %>%
  group_by(periode) %>%
  summarise(nb_emp = as.numeric(sum(nb_emp, na.rm = T)))

nuitees <- read_csv2("donnees/logements/nuitees.csv") %>%
  select(1,hotels = 2, campings = 4) %>%
  slice(-c(1:2)) %>%
  separate(Libellé, c("annee", "mois"), "-") %>% mutate_all(as.numeric) %>%
  rowwise %>%
  mutate(nuitees = sum(hotels , campings, na.rm = T)) %>%
  mutate(trimestre = case_when(
    mois < 4 ~ "1",
    mois < 7 ~ "2",
    mois < 10 ~ "3",
    TRUE ~ "4",
  )) %>% unite("periode", c(annee,trimestre), sep = "t") %>%
  group_by(periode) %>% summarise(nuitees = sum(nuitees, na.rm = T)) %>% arrange(desc(periode))

# periodes communes #
per_list <- intersect(nuitees$periode ,emploi_IZ$periode)
per_start <- c(per_list %>% last() %>% str_sub(.,1,4),per_list %>% last() %>% str_sub(.,6,6)) %>% as.numer

```

```

# series temporelles, ça commence au T4 2001 et c'est par trimestres #
ts_nuitees <- nuitees %>%
  filter(periode %in% per_list) %>%
  select(nuitees) %>%
  ts(frequency = 4, start = per_start)
ts_emploi_IZ <- emploi_IZ %>%
  filter(periode %in% per_list) %>%
  select(nb_emp) %>%
  ts(frequency = 4, start = per_start)

autoplot(cbind(ts_nuitees, ts_emploi_IZ))

## on decompose les series pour travailler sur le trend et sur les residus
ts_nuitees.dec <- decompose(ts_nuitees, type="add") # decompose the TS
autoplot(ts_nuitees.dec) # view composantes

ts_emploi_IZ.dec <- decompose(ts_emploi_IZ, type="add") # decompose the TS
autoplot(ts_emploi_IZ.dec) # view composantes

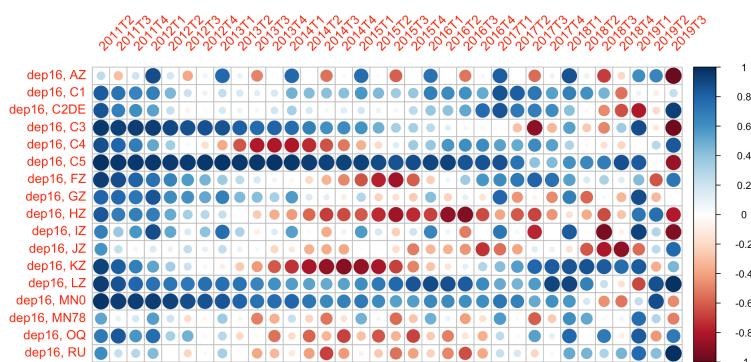
# les deux series
autoplot(list(ts_nuitees.dec, ts_emploi_IZ.dec))

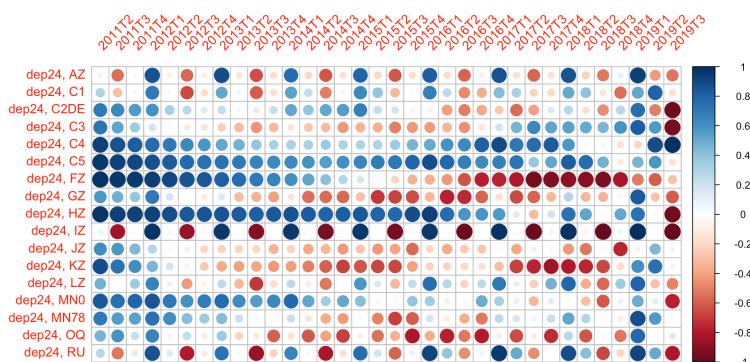
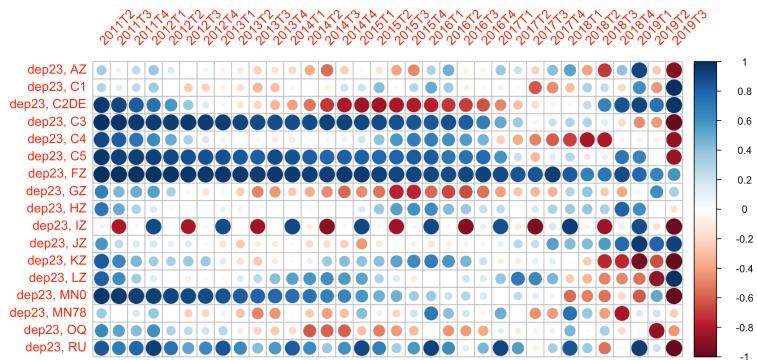
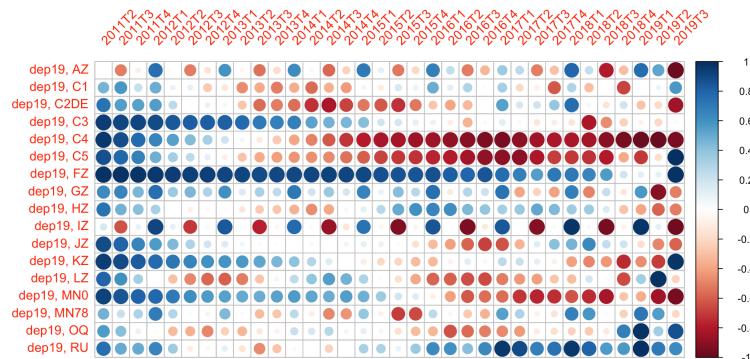
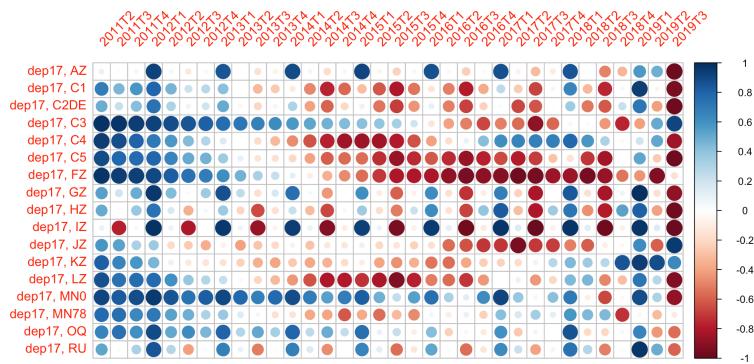
# visualisation des residus
autoplot(cbind(ts_nuitees.dec$random, ts_emploi_IZ.dec$random))
plot(ts_nuitees.dec$random)
plot(ts_emploi_IZ.dec$random)

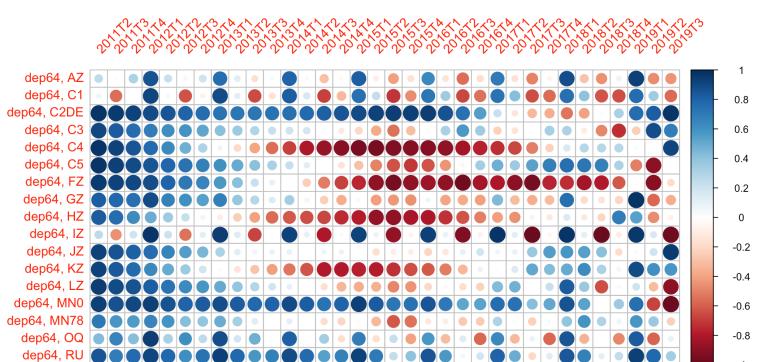
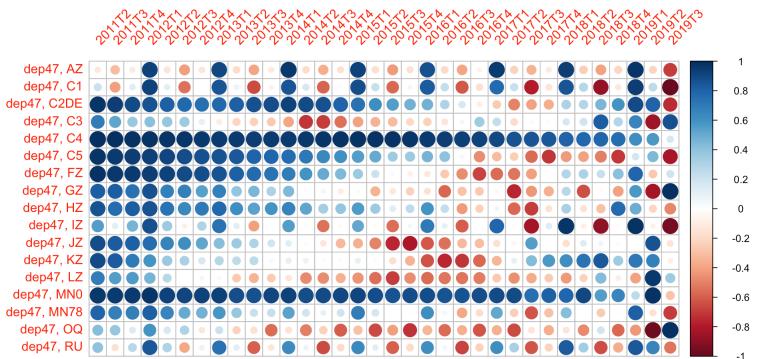
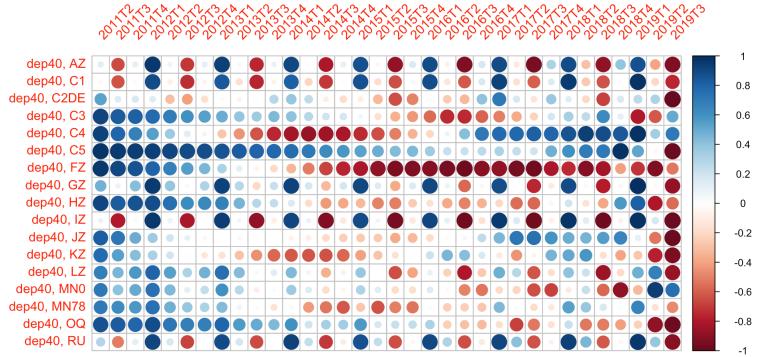
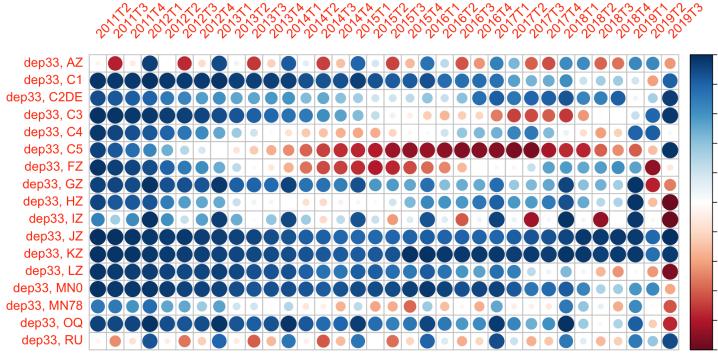
forecast::Ccf(as.numeric(ts_nuitees.dec$random), as.numeric(ts_emploi_IZ.dec$random),
             type = "correlation", xlab = "Déphasage trimestriel", main = "Diagramme de correlation crois-
             sub = "Nuitées hôtel+campings et l'emploi dans l'hébergement-restauration")

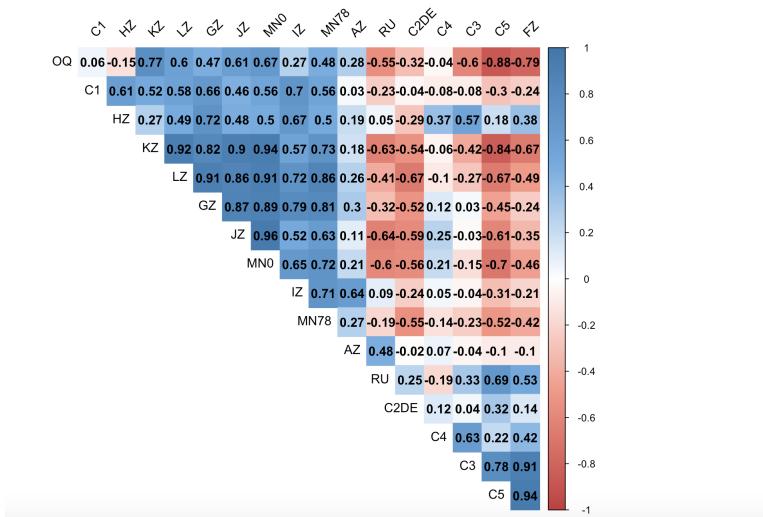
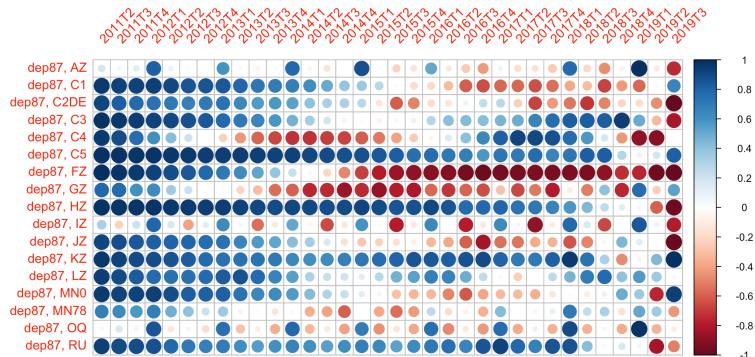
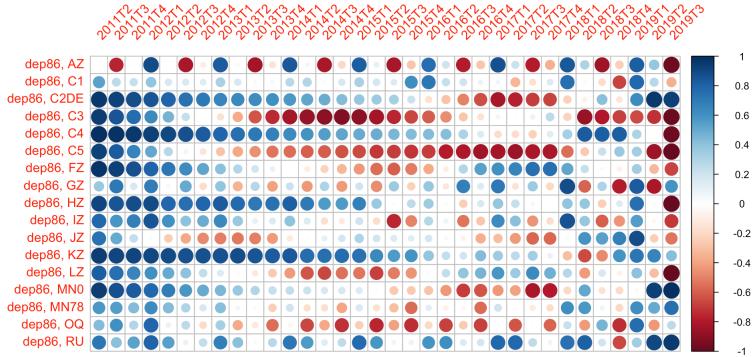
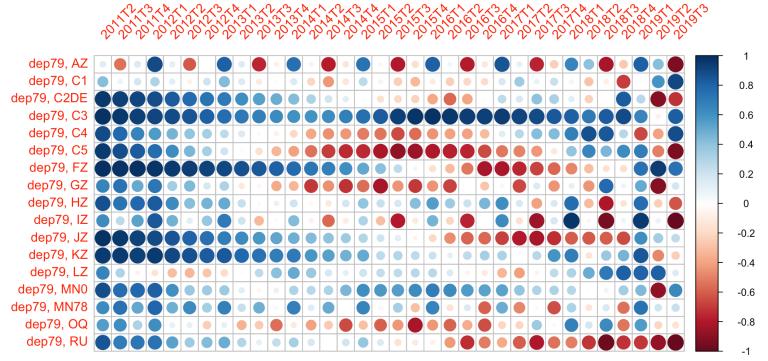
```

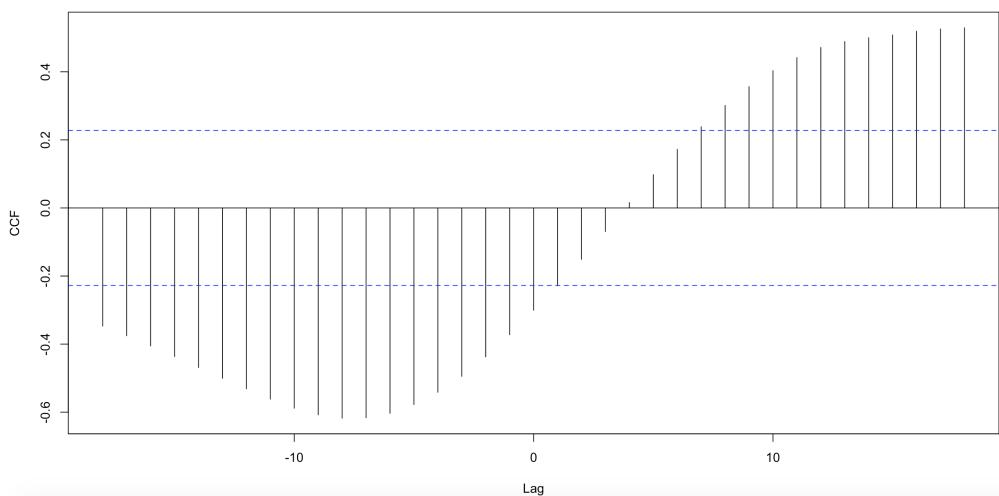
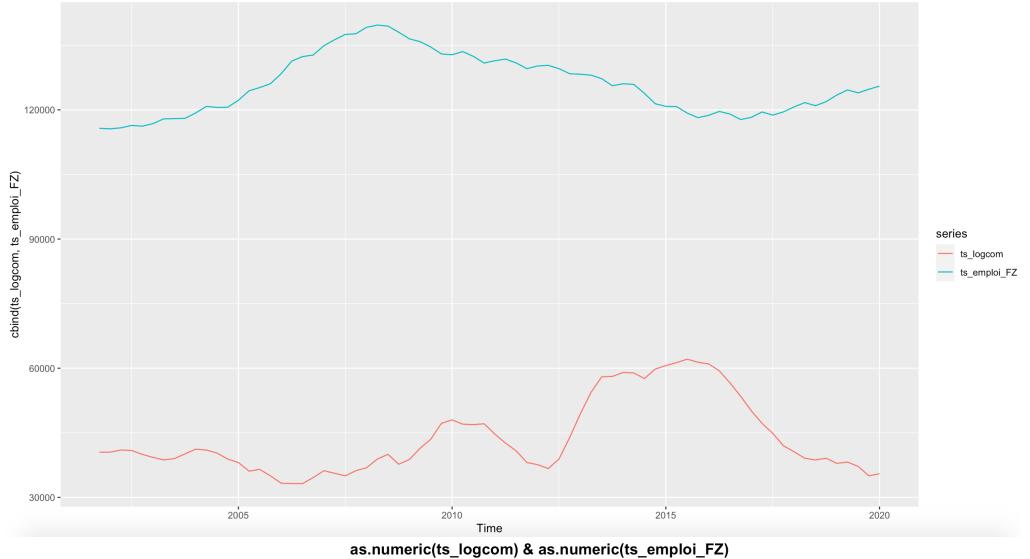
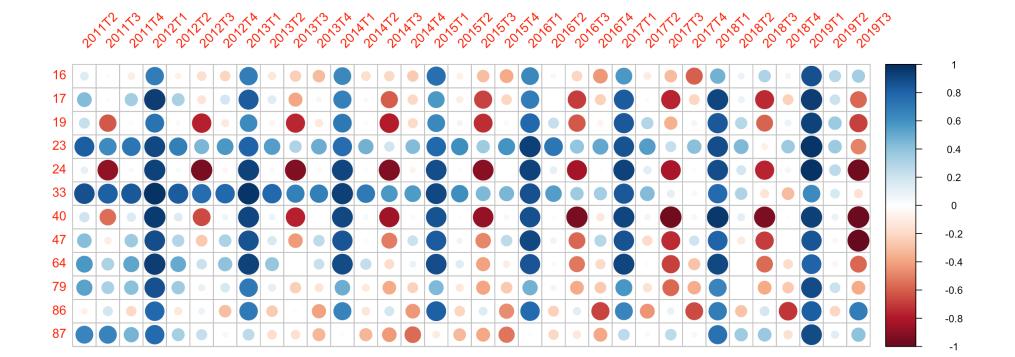
4 Sorties

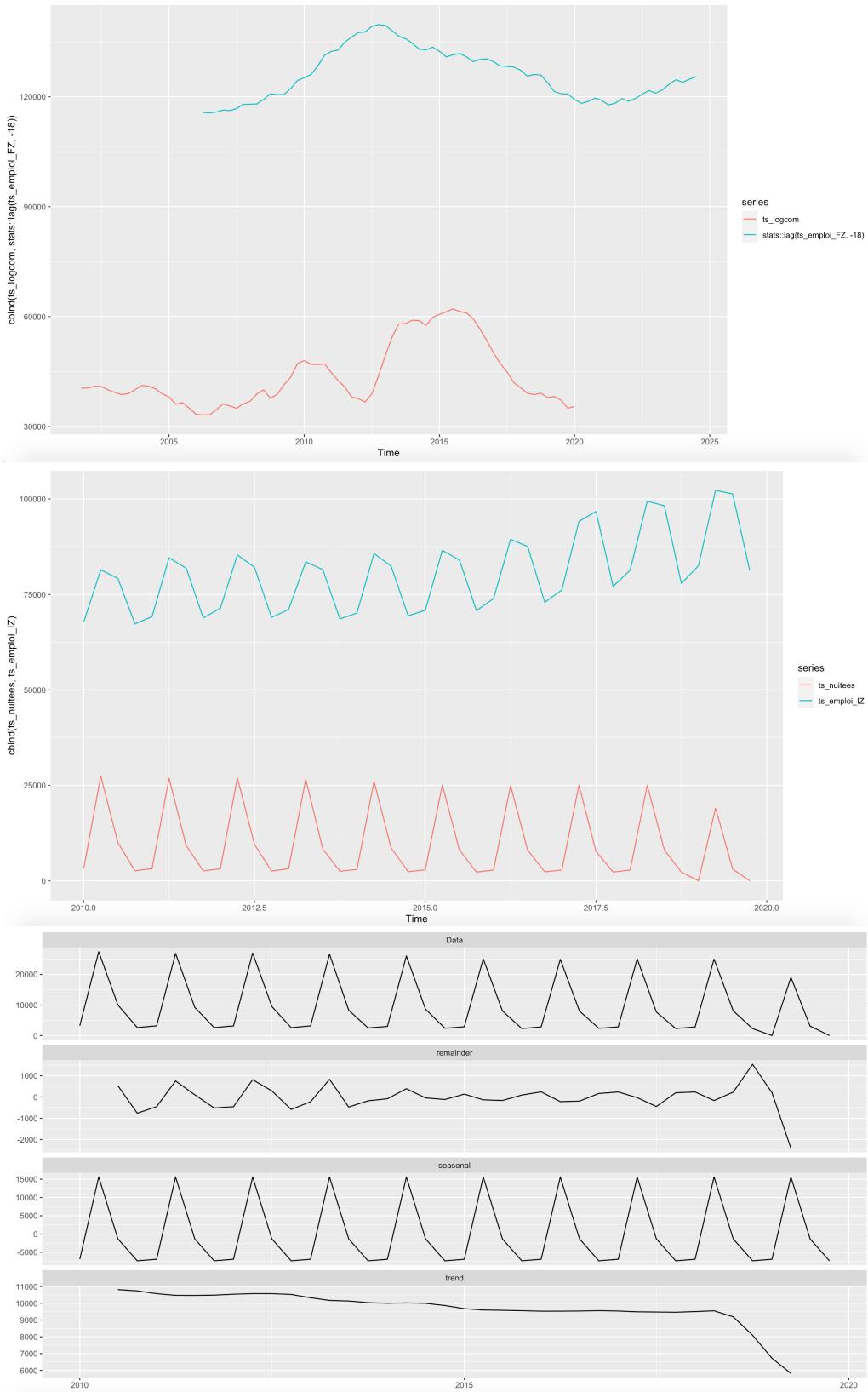


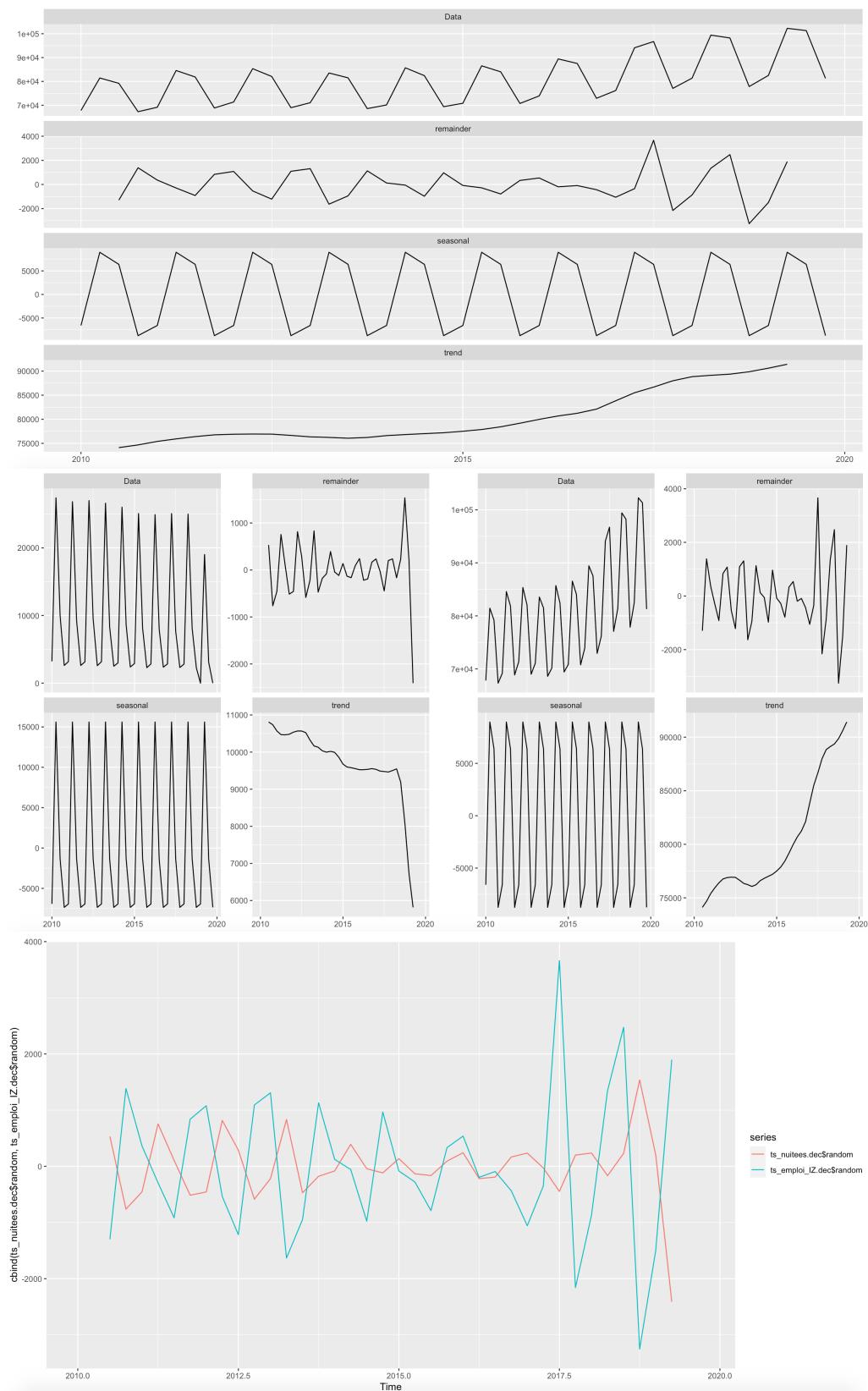












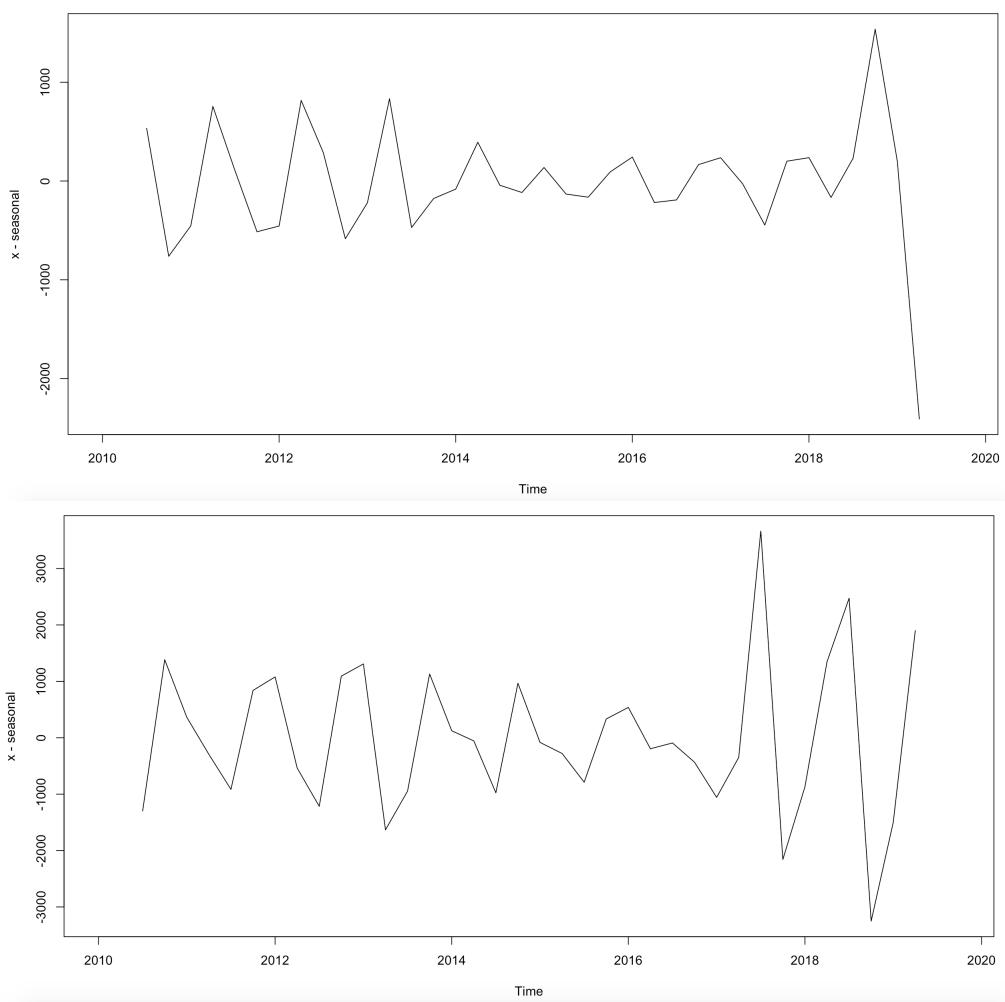
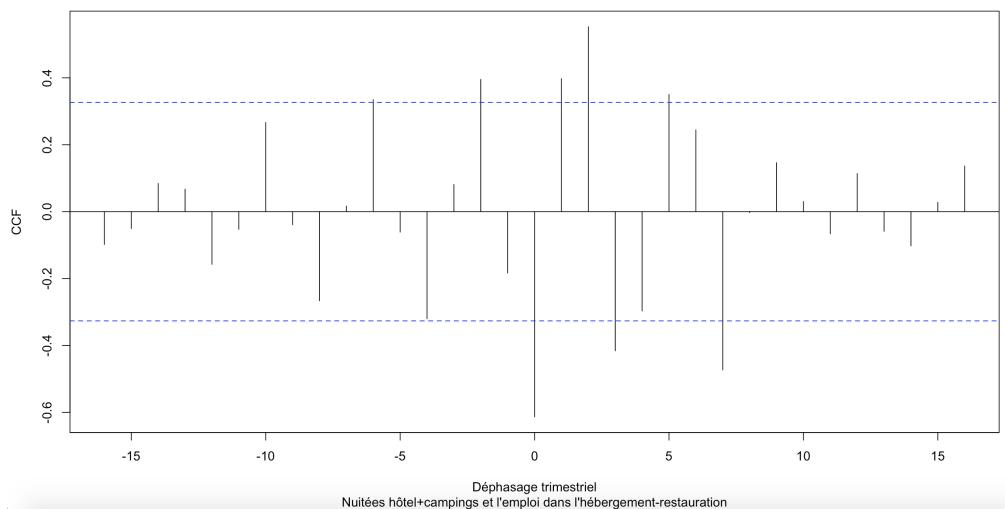


Diagramme de corrélation croisée



Bibliographie

- [1] Insee Horizon 2025 (<https://www.insee.fr/fr/information/2382227>)
- [2] Connaître l'Insee (<https://www.insee.fr/fr/information/1300620>)
- [3] L'insertion dans le marché du travail reste moins aisée dans les DOM en 2017 (<https://www.insee.fr/fr/statistiques/3568843>)
- [4] Les instances de gouvernance (<https://www.insee.fr/fr/information/4137751>)
- [5] Secret Statistique (<https://www.insee.fr/fr/information/1300624>)
- [6] Règles de diffusion des indicateurs statistiques et des publications (<https://www.insee.fr/fr/information/3324236>)
- [7] Impartialité et objectivité (<https://www.insee.fr/fr/information/4175122>)
- [8] La nomenclature agrégée - NA, 2008 (<https://www.insee.fr/fr/information/2028155>)