

Rapport de projet géostatistique

Sujet : “Chômage en Île-de-France”

Groupe : Clélia Ségouin, Joséphine Bocquet, Julie Morel

Date : 22/12/2023

Sommaire :

1. Introduction	2
a. Contexte	2
b. Définition	2
c. Problématique	2
d. Limites géographiques	3
2. Méthodologie	3
a. Présentation des données	3
b. Choix des indicateurs	4
3. Analyse des résultats	5
4. Conclusion	12

1. Introduction

a. Contexte

En 1968, dans la théorie du “spatial mismatch”, J. F. Kain affirme que la distance kilométrique et temporelle entre les lieux de vie des classes populaires et les bassins d’emploi est à l’origine d’un accroissement des inégalités sociales, notamment en termes d’accès à l’emploi (géoconfluences). Il pourrait donc y avoir un lien entre chômage et inégalités d’accès aux moyens de transports.

Par ailleurs, la revue économique « Sortir du chômage en Île-de-France. Disparités territoriales, spatial mismatch et ségrégation résidentielle » (vol. 60, no. 4, 2009) affirme que certaines localités en Île-de-France semblent disposer d’une mauvaise connexion physique aux offres d’emploi. L’enclavement des communes rend également la recherche d’emploi plus complexe en raison d’une prospection longue et coûteuse. Le chômage reposerait donc potentiellement sur des problèmes d’isolement physique et de ségrégation résidentielle.

b. Définition

Selon l’INSEE, le chômage est “l’ensemble des personnes de 15 ans et plus, privées d’emploi et en recherchant un”. Plus précisément, pour mesurer le nombre de chômeurs en France, l’INSEE s’appuie sur le BIT qui définit le chômeur selon caractéristiques suivantes :

- Avoir plus de 15 ans
- Être sans emploi à une semaine donnée
- Avoir trouvé un emploi qui commence dans moins de 3 mois
- Avoir recherché activement un emploi dans les 4 dernières semaines

Le chômage est à l’origine de multiples inégalités socio-spatiales et son origine est multifactorielle. Avec la perspective de la théorie de *spatial mismatch*, nous souhaitons nous intéresser plus particulièrement au lien entre isolement spatial et chômage. L’isolement spatial d’une population vis-à-vis de l’emploi peut se définir selon plusieurs critères :

- Distance kilométrique par rapport aux bassins d’emplois les plus attractifs
- Distance temporelle par rapport aux bassins d’emplois les plus attractifs : ce critère peut être décliné selon le mode de transport utilisé
- Distance coût par rapport aux bassins d’emplois les plus attractifs

La distance aux bassins d’emplois introduit la notion de migration pendulaire, définie comme “les navettes quotidiennes domicile-travail” (géoconfluences). En effet, nous nous sommes intéressés aux flux domiciles-travail afin d’étudier l’accessibilité à l’emploi.

L’objectif de l’étude est donc de qualifier les relations entre inégalités d’accès à la mobilité et taux de chômage dans les communes.

c. Problématique

Le chômage semble pouvoir être lié à l’éloignement géographique et à la non proximité avec l’offre de transport. Nous nous sommes alors demandés : Dans quelle mesure existe-t-il une relation entre chômage et isolement spatial ?

d. Limites géographiques

Nous avons choisi d’orienter et de limiter notre étude à la région Île-de-France. En effet, la région Île-de-France connaît une dynamique particulière due aux divers pôles d’activités présents et à l’attractivité de Paris. Par ailleurs, les données relatives aux transports en France sont souvent hétérogènes et proviennent de sources et d’acteurs différents. Les données de chômage sont regroupées par région par l’INSEE. Il était donc plus facile dans le cadre de cette étude de se pencher sur une région en particulier.

2. Méthodologie

Pour répondre au problème posé en introduction, nous nous sommes concentrés sur le prisme de l’isolement en termes d’inégalités de mobilité. Nous étudions ces inégalités selon plusieurs axes : temps et distances kilométriques des migrations pendulaires selon les différents modes de transports mais également selon l’accès aux transports en commun dans chaque commune. Enfin, l’isolement spatial des populations en Île-de-France est également étudié selon le prisme de l’éloignement par rapport au centre (Paris).

a. Présentation des données

Nous avons choisi des données relatives au chômage en Île-de-France, aux transports, aux revenus et aux flux de personnes entre les communes pour les déplacements domicile-travail.

Les données utilisées sont présentées ci-dessous :

Source	Nom de la donnée	Attributs utilisés	Année
https://www.insee.fr/fr/statistiques/7632867?sommaire=7632977	base-cc-emploi-pop-active-2020_v2	P20_POP1564 : population de 15 à 64 ans en IdF P20_CHOM1564 : nombre de chômeurs en IdF P20_ACT1564 : population active en IdF	2020
Horaires prévus sur les lignes de transport en commun d'Ile-de-France (GTFS Datahub) - data.gouv.fr	Arrêts et horaires des transports en commun en IDF	stop_id : identifiant de chaque arrêt de transport en commun arrival_time : heure d’arrivée pour une ligne à un arrêt	2023
https://analytics.humanum.fr/geographie-cites/aperitif/	listtimes.Rds : temps de trajets entre communes de départ	ORI : code insee de la commune d’origine DES : code insee de la	2018

https://gitlab.huma-n.um.fr/hcommenges/aperitif	et d'arrivée => 1 matrice par mode de transport	commune d'arrivée VAL : valeur du temps de trajet selon le mode de transport	
Principaux indicateurs sur les revenus et la pauvreté aux niveaux national et local en 2020 – Principaux résultats sur les revenus et la pauvreté des ménages en 2020 Insee	Filosofi	MED20 : niveau de vie médian par commune	2020
https://www.insee.fr/fr/statistiques/6454112#consulter	Flux domicile-travail	CODGEO : code insee de la commune de départ DCLT : code insee de la commune d'arrivée NBFLUX_C19_ACTOCC15	2019
https://wxs.ign.fr/859x8t863h6a09o9o6fy4v60/telechargement/prepack/BDOPOV3-TOUSTHEMES-REGION-PACK_233\$BDTOPO_3-3_TOUSTHEMES_SHP_LAMB93_R11_2023-09-15/file/BDOPO_3-3_TOUSTHEMES_SHP_LAMB93_R11_2023-09-15_7z	Communes d'Ile-de-France	POPULATION : population totale par commune	2023
https://opendata.paris.fr/explore/dataset/arrondissements/export/?disjunctive.c_ar&disjunctive.c_arinsee&disjunctive.l_ar	Arrondissements de Paris		2016

b. Choix des indicateurs

A partir des données brutes, nous avons extrait et/ou calculé un certain nombre d'indicateurs afin de pouvoir étudier leur impact sur le chômage dans les communes d'Ile-de-France.

D'une part, nous n'avons pas accès directement à l'information du taux de chômage par commune d'Ile-de-France dans les données disponibles et nous l'avons donc calculé à partir des données de l'INSEE concernant l'emploi en Ile-de-France (base-cc-emploi-pop-active-2020_v2) de la manière suivante :

$$\text{Taux de chômage} = \text{Nombre de chômeurs} / \text{Population active}$$

D'autre part, nous avons choisi de mettre en avant plusieurs indicateurs de mobilité / d'accès aux mobilités à partir des données, afin de comprendre leur potentielle relation avec le chômage en Île-de-France. Ces indicateurs sont les suivants :

- Indicateurs extraits directement des données :
 - Temps de trajet en voiture entre les communes de départ et d'arrivée des flux domicile-travail
 - Temps de trajet en transports en commun entre les communes de départ et d'arrivée des flux domicile-travail
 - Distance entre les communes de départ et d'arrivée des flux domicile-travail
 - Niveau de vie médian par commune
 - Nombre d'emplois par commune : nombre de flux d'arrivée par commune
- Indicateurs calculés à partir des données :
 - Nombre de passages de transports en commun par jour par commune
 - Nombre d'arrêts de lignes de transports en commun par commune
 - Distance à vol d'oiseau entre les communes d'Île-de-France et le centre de Paris (en mètres)

3. Analyse des résultats

Une fois nos données traitées, nous avons cherché à étudier les différents facteurs explicatifs du taux de chômage par commune. Pour cela, nous avons tout d'abord cartographié le taux de chômage par commune en Île-de-France (Figure 1).

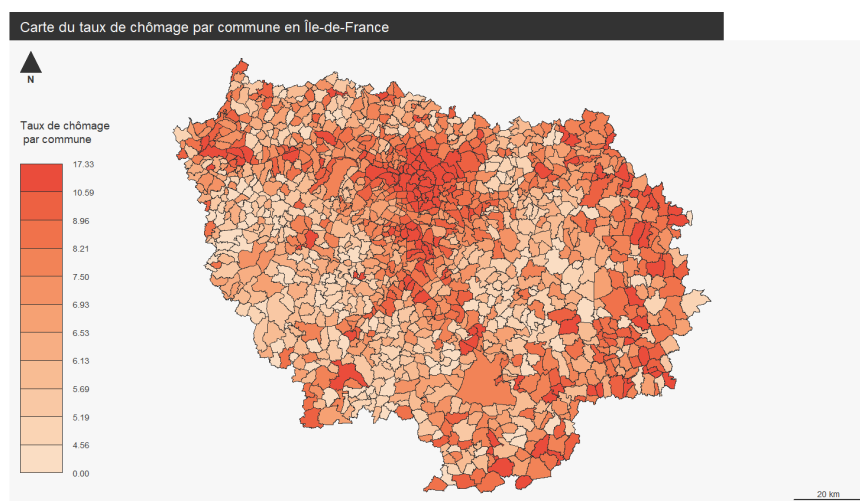


Figure 1 : Carte du taux de chômage par commune en Île-de-France

Comme nous pouvons l'observer sur la carte ci-dessus, la répartition du taux de chômage semble concentrée proche du centre nord-est de Paris ainsi qu'au niveau des frontières est de la région. C'est

donc cette répartition spatiale que nous allons tenter d'expliquer par l'analyse des indicateurs précédemment détaillés.

Tout d'abord, d'après les travaux de Blanchflower et Oswald en 1994, il y a une relation très forte entre le taux de chômage de son bassin d'emploi et le salaire d'un individu. En effet, comme nous pouvons le voir sur la figure ci-dessous, représentant la répartition du revenu par commune d'Île-de-France (Figure 2), la répartition des revenus semble distribuée similairement au taux de chômage.

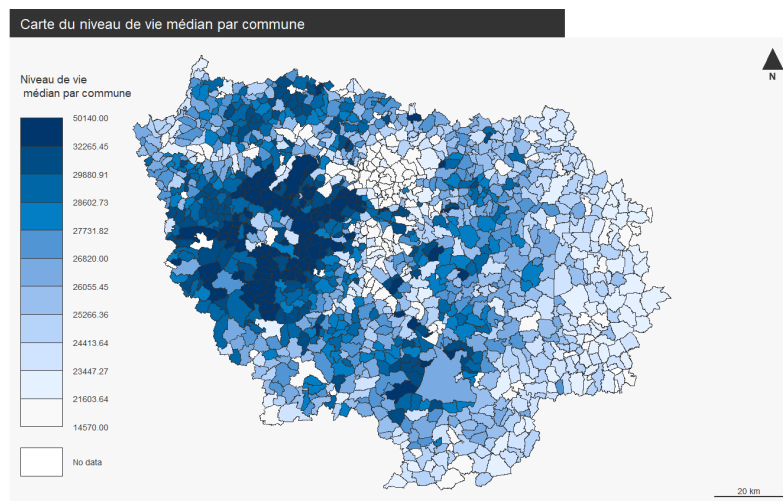


Figure 2 : carte de la répartition du revenu en Ile-de-France

Afin de vérifier la relation précédemment introduite, nous avons cherché à mettre en évidence la relation négative entre la répartition du taux de chômage et le niveau de vie médian par commune en Île-de-France. Pour quantifier cette relation, nous avons effectué une régression linéaire des variables que nous pouvons observer sur la figure 3 ci-dessous. La figure 3 présente la corrélation linéaire du taux de chômage en fonction du revenu par commune étudiée.

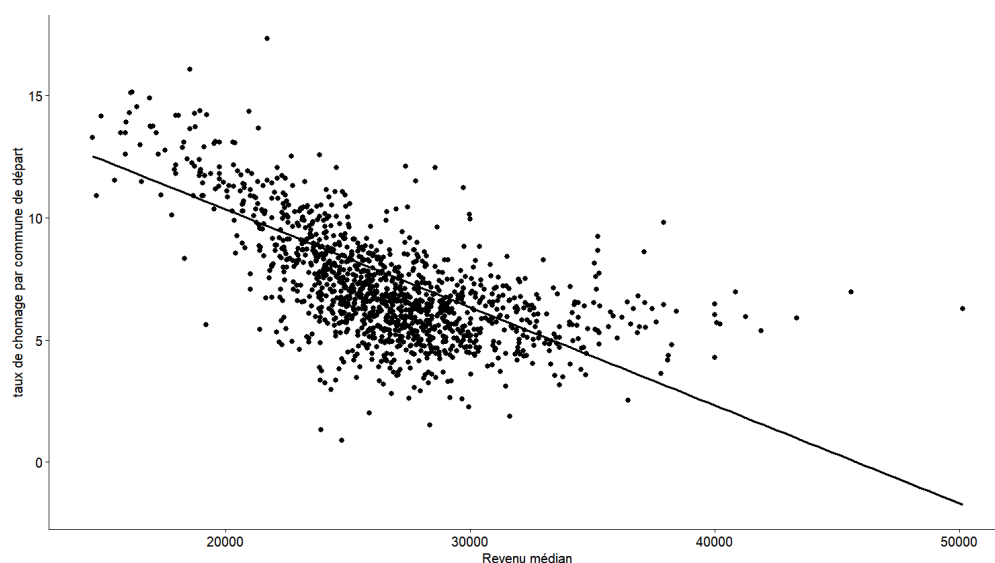


Figure 3 : corrélation du taux de chômage en fonction du revenu

La figure 3 nous confirme l'hypothèse de la corrélation négative de la répartition spatiale du taux de chômage en fonction du revenu en Île-de-France. En effet, le coefficient de corrélation est de -0.62 (cf. figure 4) ce qui implique une relation négative très forte : plus le revenu médian est important plus le taux de chômage est faible. De plus, la p-value ou la valeur de probabilité est très faible ($<2,2 \times 10^{-16}$) ce qui indique une corrélation significative, en effet, sous un seuil de signification l'hypothèse nulle est rejetée. L'hypothèse nulle étant l'hypothèse sous laquelle il n'y a pas de corrélation.

Notre étude étant portée sur l'isolement géographique comme facteur explicatif du taux de chômage, nous allons à présent étudier les différents indicateurs précédemment détaillés. Afin d'avoir une idée générale de la relation entre nos indicateurs nous avons réalisé une matrice de corrélation. La figure 4 ci-dessous représente la matrice de corrélation du taux de chômage et de nos variables.

	Taux chômage	Nombre de passage de TC par jour par habitant	Nombre d'arrêts par commune	Niveau de vie médian	Temps moyen en TC	Temps moyen en VP	Distance au centre de la commune d'arrivée
Taux chômage	1	0,19	0,40	-0,62	-0,28	-0,30	-0,05
Nombre de passage de TC par jour par habitant	0,19	1	0,43	0,06	-0,28	-0,43	-0,36
Nombre d'arrêts par commune	0,39	0,43	1	-0,15	-0,50	-0,65	-0,36
Niveau de vie médian	-0,62	0,06	-0,15	1	0,07	0,01	-0,22
Temps moyen en TC	-0,28	-0,28	-0,50	0,07	1	0,62	0,31
Temps moyen en VP	-0,30	-0,43	-0,65	0,01	0,62	1	0,69
Distance au centre de la commune d'arrivée	-0,05	-0,36	-0,36	-0,22	0,31	0,68	1

Figure 4 : Matrice de corrélation du taux de chômage et de nos indicateurs d'accessibilités

Malgré des coefficients de corrélation linéaire assez faible nous avons cherché à approfondir l'analyse de notre hypothèse de départ soit l'isolement comme facteur explicatif du taux de chômage. D'autre part, nous supposons que les variables peuvent avoir une relation forte mais non linéaire et il est donc important d'explorer chaque relation individuellement.

Tout d'abord, afin d'avoir une approche globale de notre étude, nous nous sommes intéressées à l'isolement des communes selon une distance euclidienne à vol d'oiseaux. Nous avons donc étudié deux distances euclidiennes différentes permettant de quantifier l'isolement : la distance entre la commune de départ des migrations pendulaires et le centre de Paris puis la distance entre les deux communes des migrations pendulaires. La figure 5 représente le taux de chômage en fonction de ces deux distances.

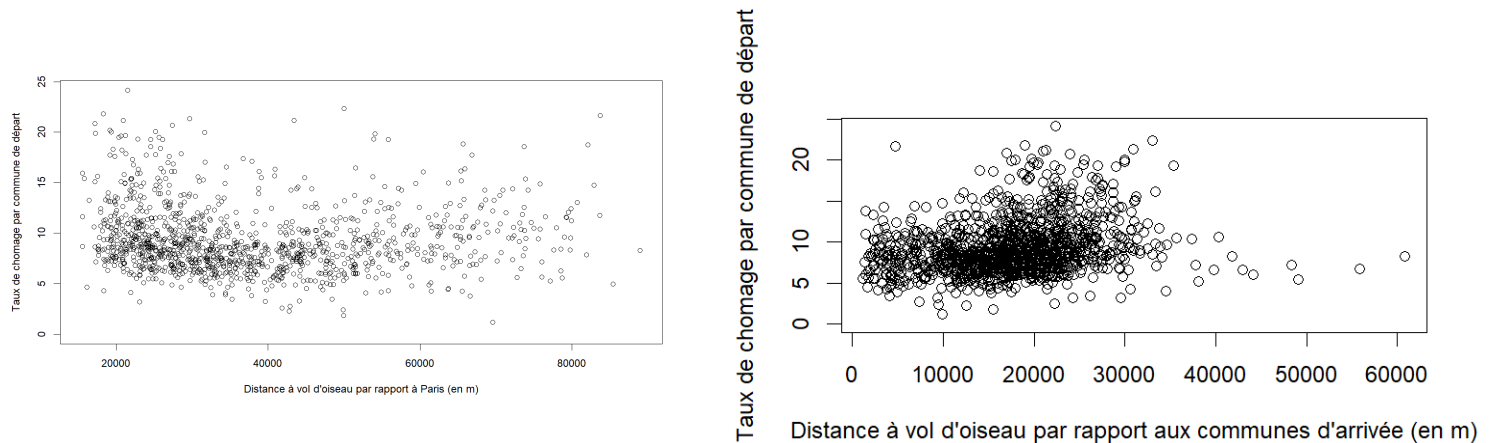


Figure 5 : Analyse du taux de chômage en fonction de la distance entre la commune de départ et le centre de Paris à gauche et en fonction de la distance entre les deux communes à droite.

Nous remarquons sur la figure 5 à droite qu'il semble y avoir une relation linéaire entre les deux variables. En effet, la valeur du coefficient de corrélation est de 0.26, ce qui confirme une légère relation linéaire entre les deux variables. Cependant le coefficient est biaisé par les fortes valeurs outliers. Par ailleurs, nous pouvons observer sur la figure 5 à gauche que la relation entre le taux de chômage et la distance au centre de Paris ne peut être expliquée par une relation linéaire bien qu'il y ait une forte dépendance entre les deux. En effet, on pourrait approximer cette tendance par une courbe polynomiale ayant un coefficient négatif pour une distance inférieure à 40 km puis un coefficient positif ensuite comme nous pouvons l'observer sur la figure 6. Celle-ci permet de mieux rendre compte de la relation polynomiale entre les deux variables. En effet, la figure 6 nous informe de la relation négative entre le taux de chômage et la distance au centre de Paris pour une distance inférieure à 40km, cela se justifie notamment par la concentration de taux de chômage élevé dans le centre nord-est de la région Île-de-France. Cette tendance semble légèrement s'inverser pour une distance supérieure à 40km, avec un taux de chômage élevé pour une distance importante ce qui pourrait justifier les taux de chômage élevé en périphérie est de la région. Cependant, bien que le modèle semble meilleur qu'une régression linéaire, celui-ci a une valeur de R^2 de 0.14 indiquant un mauvais ajustement de nos données par notre modèle et donc une faible part de variance expliquée.

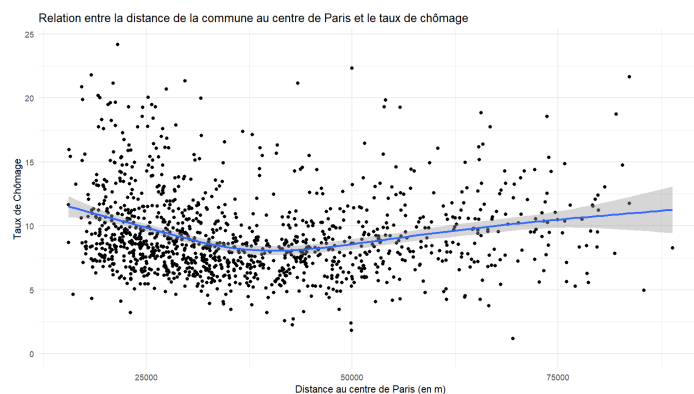


Figure 6 : Approximation polynomiale de la relation du taux de chômage en fonction de la distance au centre de Paris

Nous nous sommes ensuite intéressées plus localement à l'isolement des communes le temps de trajet moyen (en voiture ou en transport en commun) des flux domicile-travail afin de quantifier l'isolement d'une commune. Nous avons ainsi réalisé les cartes ci-dessous (Figure 7) représentant à gauche la répartition des temps de trajet en transport en commun (TC) et à droite la répartition des temps de trajet en véhicule privé (VP) par commune de destination des flux domicile-travail.

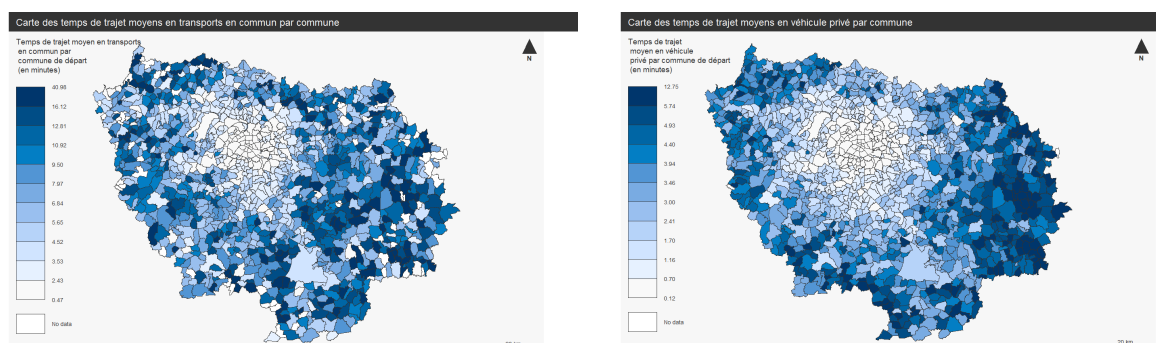


Figure 7 : Répartition des temps de trajet en TC à gauche et en VP à droite par commune de départ des flux domicile-travail.

Comme nous pouvons l'observer sur les cartes ci-dessous, la répartition spatiale des temps de trajet est très dispersée et comporte des valeurs élevées à l'est de Paris et aux frontières de la région. Comme attendu, le centre de Paris est très accessible en transport en commun contrairement aux communes plus éloignées du centre en raison de l'inégalité d'accès aux transports en commun mais aussi en raison de la formation en étoile du réseau de transport.

Malgré une répartition spatiale très différente, nous avons cherché à quantifier la relation entre le temps de trajet par commune de départ des migrations pendulaires et le taux de chômage. La figure 8 représente le taux de chômage en fonction du temps de trajet en transport en commun (TC) et en voiture privée par commune de départ des migrations pendulaires.

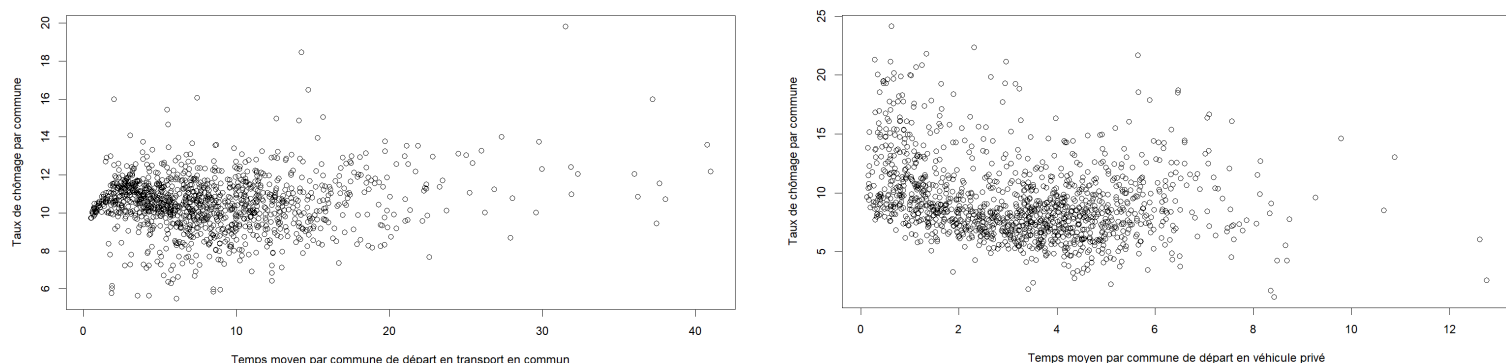


Figure 8 : Analyse du taux de chômage en fonction du temps de trajet en TC à gauche et en VP à droite par commune de départ des migrations pendulaires.

Nous pouvons observer sur la figure 8 deux relations entre les temps de trajet et le taux de chômage qui ne semblent pas linéaires ni significatives ce qui confirme les valeurs de la matrice de corrélation. Malgré une relation non linéaire, les variables semblent tout de même avoir une forte dépendance, notamment pour la figure 8 à gauche où les valeurs semblent suivre une courbe logarithmique. Nous avons donc tenté de linéariser nos données afin de quantifier les relations entre ces variables en passant par le logarithme. La figure 9 ci-dessous présente le logarithme du taux de chômage en fonction des valeurs logarithmiques des temps de trajets par commune de départ des migrations pendulaires.

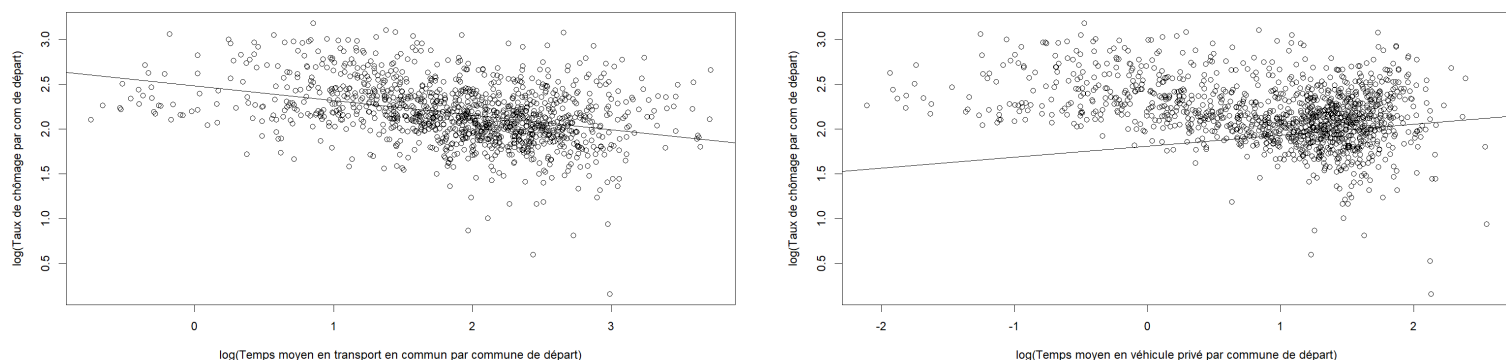


Figure 9 : Analyse du logarithme du taux de chômage en fonction des valeurs logarithmiques des temps de trajets par commune de départ des migrations pendulaires

La figure 9 à gauche nous permet d'observer une relation linéaire significative entre les valeurs logarithmiques du temps de trajet moyen en transport en commun et le taux de chômage. En effet, après la linéarisation des données, nous obtenons un coefficient de corrélation d'une valeur de -0.37 ce qui nous confirme la tendance linéaire entre les deux variables. Cependant, ce résultat est à nuancer avec la valeur de R^2 , soit la part de variance expliquée qui est seulement de 0.12 pour une p-value inférieur à $< 2 \cdot 10^{-16}$. Cela suggère que le modèle est suffisamment significatif mais pas bien adapté aux données. A l'inverse, nous observons une relation non linéaire mais significative entre le temps de trajet moyen en voiture privée et le taux de chômage par commune en Île-de-France. Il est donc intéressant d'approcher cette relation par une régression polynomiale comme nous pouvons l'observer

sur la figure 10 ci-dessous. Ce modèle de régression semble plus pertinent pour nos données malgré une valeur de R^2 très faible, 0.19, mais avec une p-value significative (inférieur à 2×10^{-16}).

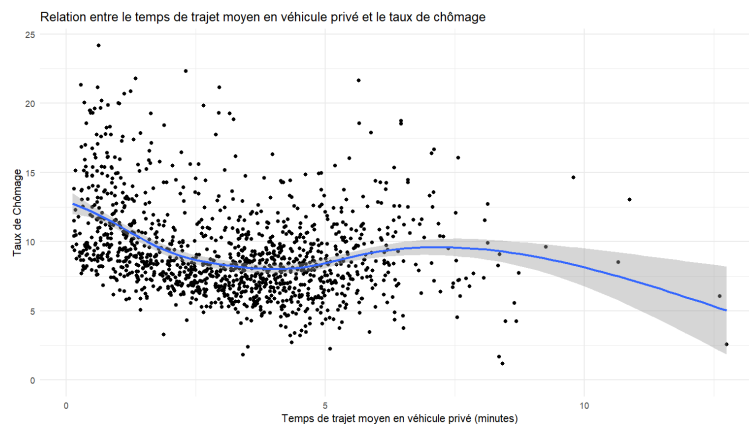


Figure 10 : Approximation polynomiale du taux de chômage en fonction du temps de trajet moyen en voiture privée par commune de départ des migrations pendulaires.

Cependant, comme nous l'avons vu ci-dessus, la région Île-de-France est très inégalitaire spatialement en termes d'accessibilité au transport en commun ce qui peut donc influencer le temps de trajet précédemment étudié. C'est pourquoi nous avons donc cherché à analyser l'inégalité d'accès au transport en commun comme facteur explicatif du taux de chômage, pour cela nous nous sommes concentrées sur les données de fréquence de passage et de nombre d'arrêts par commune. La figure 11 ci-dessous représente la répartition du nombre de passages de transport en commun par jour par habitant par commune en Île-de-France. Cette figure nous informe de la concentration spatiale du nombre de passage par habitant en Île-de-France et confirme les inégalités d'accès au transport en commun.

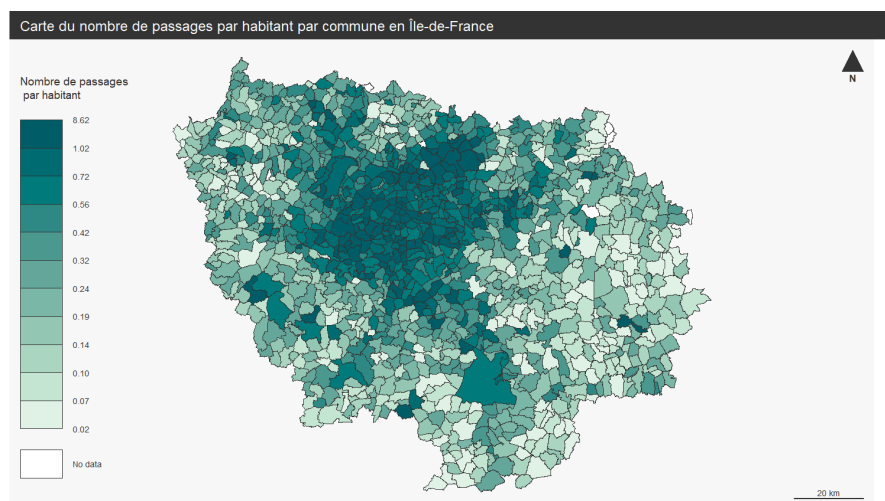


Figure 11 : Carte du nombre de passage par habitant par commune en Île-de-France

La matrice de la figure 4 nous informe d'un coefficient de corrélation assez faible, 0.19, entre le taux de chômage et le nombre de passages par habitant. Nous souhaitons cependant explorer la dispersion par une étude de la relation entre ces deux variables. Après étude des variables, nous n'avons pas trouvé de relation linéaire entre les deux variables ce qui permet de mettre en lumière le biais

important du coefficient selon les valeurs extrêmes, outliers. De plus, nous n'avons pas trouvé de modèle d'approximation pertinent pour l'analyse du taux de chômage en fonction du nombre de passages par habitant par commune.

D'autre part, nous nous sommes intéressés au nombre d'arrêts par habitant par commune en Île-de-France comme indicateur d'isolement. D'après la figure 4, la matrice de corrélation nous informe d'un coefficient de corrélation linéaire de 0.40, nous avons donc cherché à vérifier cette hypothèse. Après étude de la dispersion du taux de chômage en fonction du nombre d'arrêt par habitant, nous n'avons pas pu conclure à un modèle pertinent pour l'étude. Nous nous sommes donc intéressés à la répartition spatiale de cet indicateur, représenté par la figure 12 ci-dessous. Nous pouvons remarquer la répartition inverse à la fréquence de passage par commune, en effet, les communes comprenant un grand nombre d'arrêts par habitant se situent en périphérie. Cela peut notamment être expliqué par le faible nombre d'habitants en périphérie par rapport au centre. Par ailleurs, les arrêts n'ont pas été distingués selon le type de ligne de transport durant notre étude. Nous pouvons ainsi affirmer que le nombre d'arrêts par habitant ne semble pas être un facteur explicatif du taux de chômage.

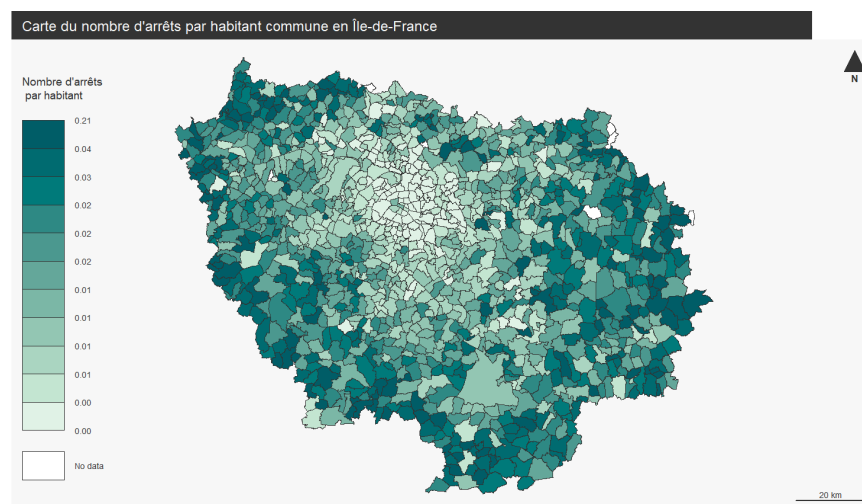


Figure 12 : Carte du nombre d'arrêts par habitant par commune en Île-de-France

4. Conclusion

Pour conclure, nous avons réalisé une étude du taux de chômage selon nos indicateurs définissant l'isolement géographique d'une commune. Nous avons construit donc définis différents indicateurs tel que l'accessibilité à l'emploi mesuré à partir de la distance euclidienne entre chaque flux domicile travail. Ces mesures nous ont permis d'étudier l'impact de ces indicateurs dans l'explication de la répartition du taux de chômage en Île-de-France.

Nous avons généralement observé une faible dépendance linéaire entre la plupart de nos indicateurs et le taux de chômage, contrairement à l'hypothèse émise en introduction d'un lien fort entre chômage et isolement spatial. Cependant, ces dépendances peuvent être modélisées par des approximations polynomiales, ce qui pourrait s'expliquer par une répartition multicentrique des bassins d'emplois dans la région Île-de-France. Il serait intéressant d'approfondir l'étude par une analyse de la régression géographique pondérée (Geographical Weighted Regression) afin de mettre en évidence l'existence d'explications spatialisées au chômage en Île-de-France. De plus, il est important de

nuancer les résultats de notre étude dans la mesure où nous n'avons pas pris en compte l'existence d'autres pôles attractifs pour les emplois des Franciliens.

Bibliographie :

- Migration pendulaire (2023)
<http://geoconfluences.ens-lyon.fr/glossaire/mobilite-pendulaire>
- Chômeur - Bureau International du Travail (07/06/2023)
<https://www.insee.fr/fr/metadonnees/definition/c1129>
- Duguet, Emmanuel, Yannick L'Horty, et Florent Sari. *Sortir du chômage en Île-de-France. Disparités territoriales, spatial mismatch et ségrégation résidentielle*, *Revue économique*, vol. 60, no. 4, 2009, pp. 979-1010.
<https://www.cairn.info/revue-economique-2009-4-page-979.html>
- Blanchflower, David and Oswald, Andrew, (1989), *The Wage Curve*, No 3181, NBER Working Papers, National Bureau of Economic Research, Inc