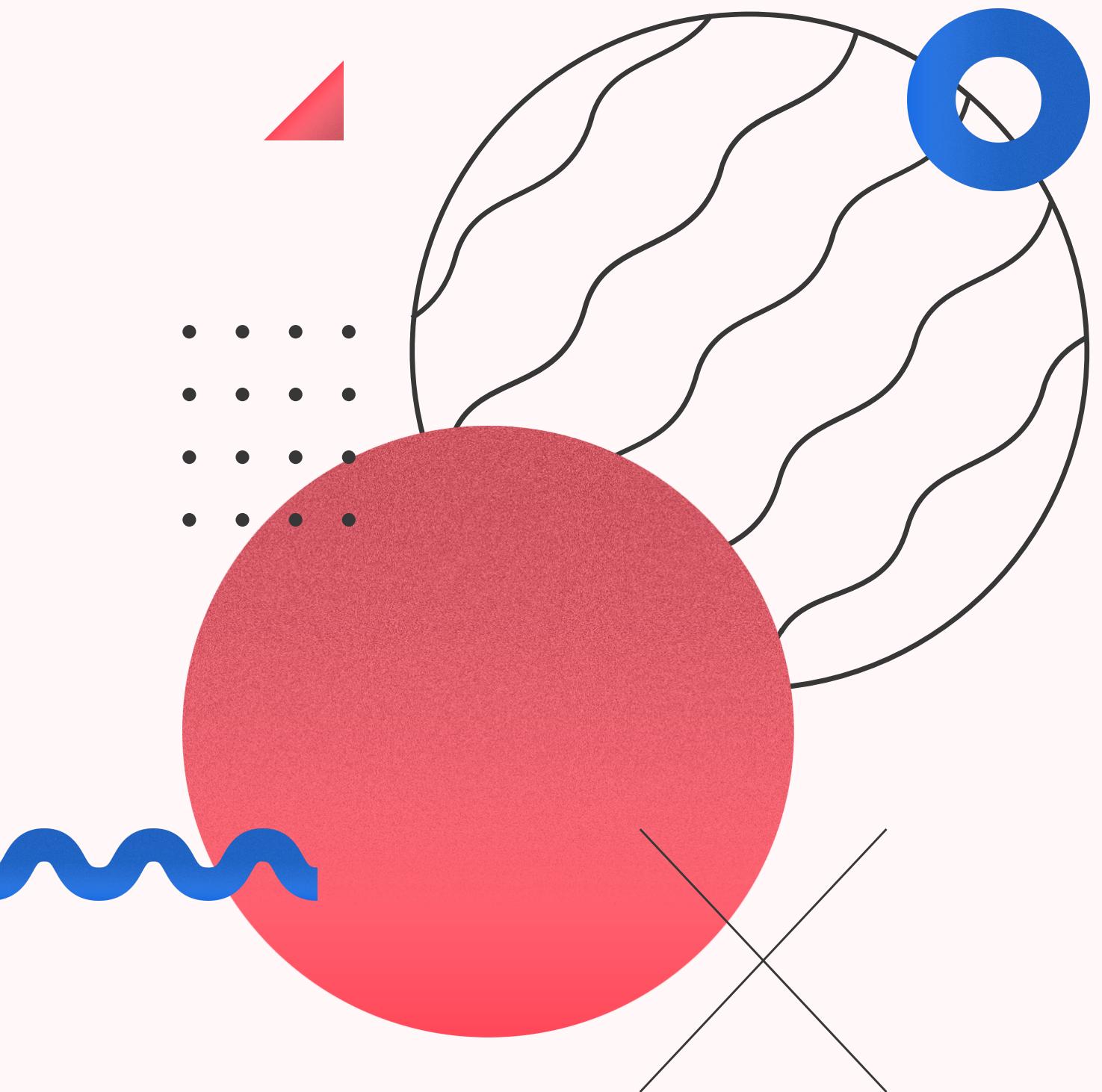


# Analyse de données

Bornes Wi-Fi de Paris

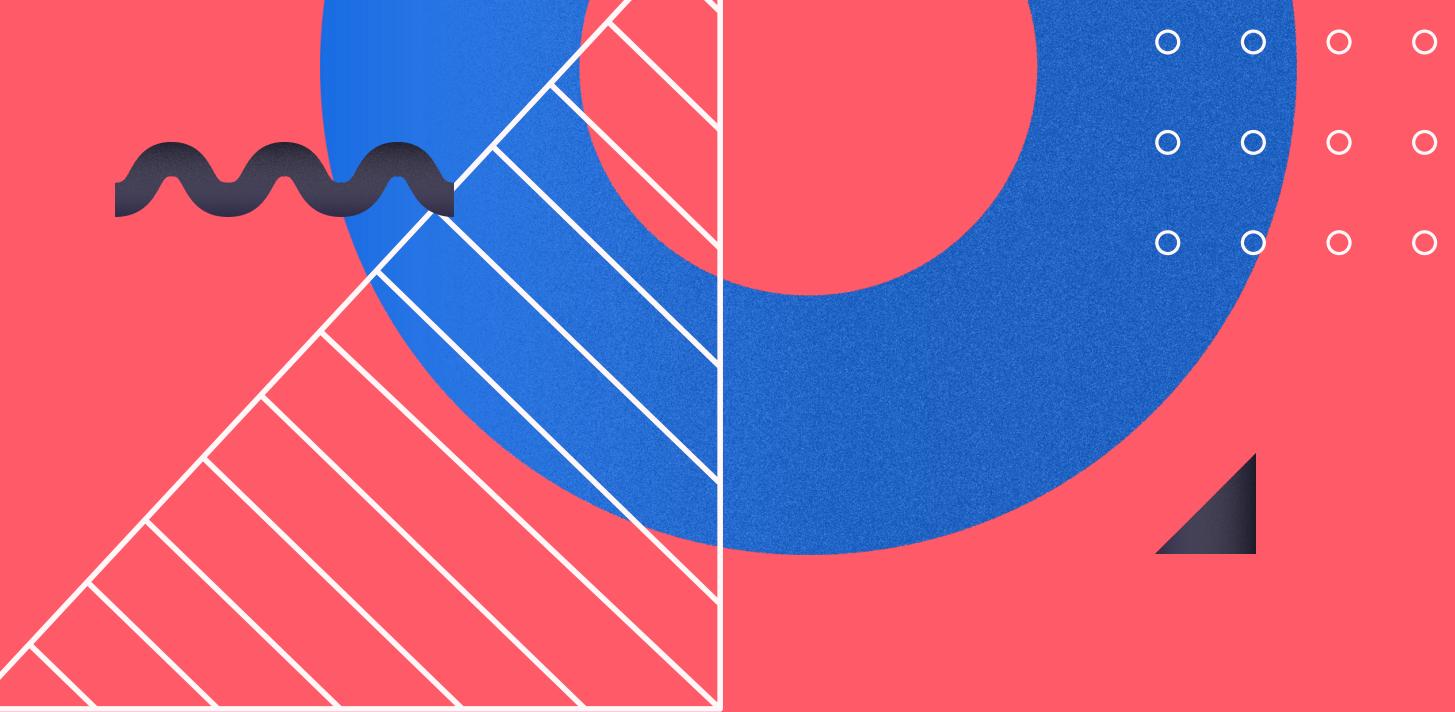


# Utilisation des bornes

## Wi-Fi de la ville de Paris

Données contiennent : identifiant du hotspot, arrondissement, langue utilisateur, marque de l'appareil, heure et durée de connexion

	<b>Code Site</b>	<b>Date heure début</b>	<b>Code postal</b>	<b>Constructeur appareil</b>	<b>Langue utilisateur</b>	<b>Temps de sessions en minutes</b>
0	PW0055	2019-09-01T15:35:00+02:00	75020	Apple	Français	52.600000
1	PW0119	2019-09-01T15:51:00+02:00	75004	Apple	Français	20.333333
2	PW0194	2019-09-01T15:51:00+02:00	75010	autre	Français	6.516667
3	PW0058-2	2019-09-01T15:52:00+02:00	75012	Apple	Français	52.716667
4	PW0161	2019-09-01T16:01:00+02:00	75007	Apple	Russe	20.166667
...	...	...	...	...	...	...
1627733	PW0043	2021-09-02T18:16:38+02:00	75004	autre	Français	10.416667
1627734	PW0134	2021-09-03T05:59:43+02:00	75002	autre	Français	5.383333
1627735	PW0133	2021-09-03T06:14:21+02:00	75017	autre	Français	0.483333



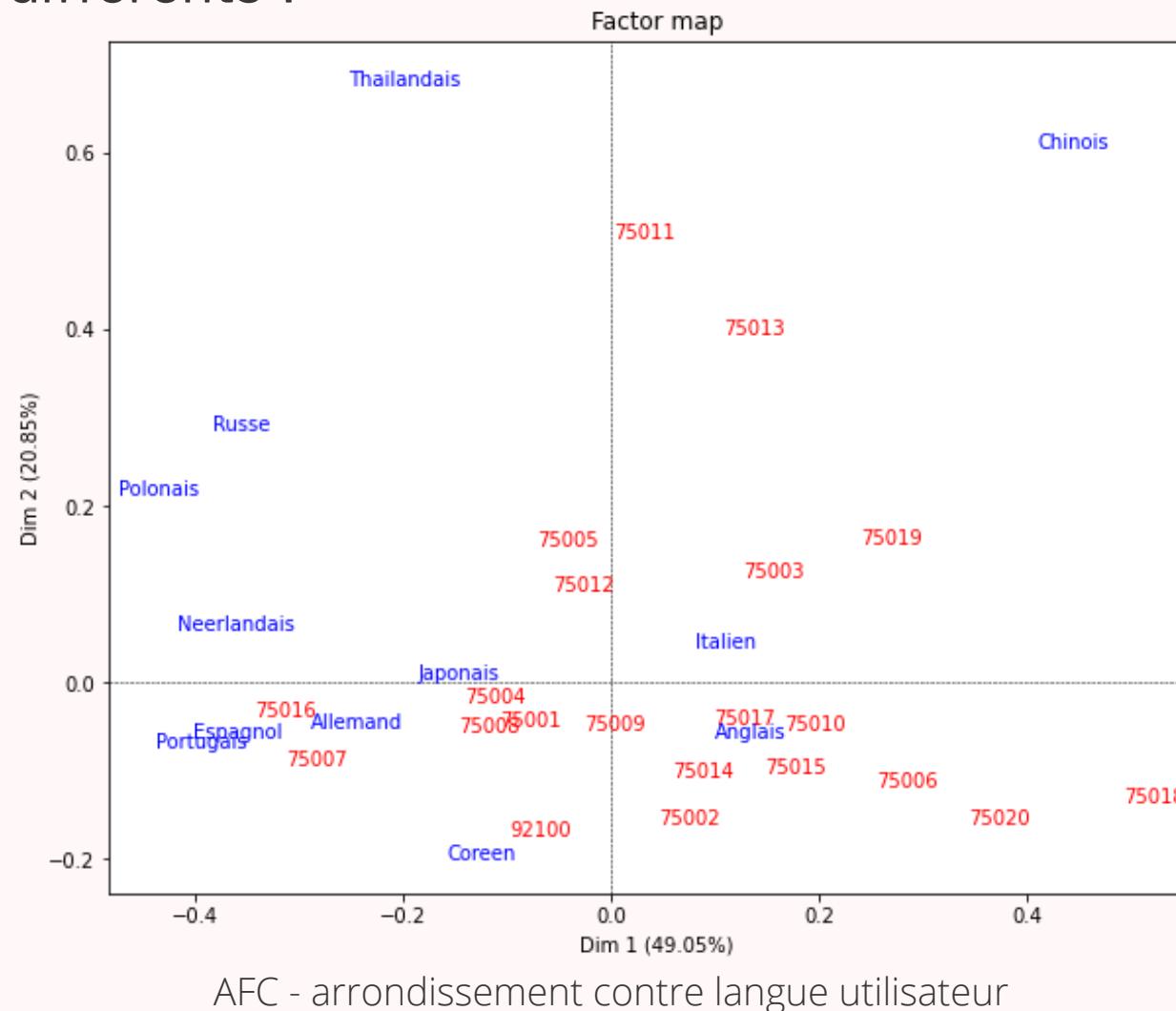
### Question

Qui utilise les bornes Wi-Fi, à quelle fréquence, et où ?

Données mises à disposition par la Direction des Systèmes d'Information et du Numérique - Ville de Paris, de octobre 2019 à octobre 2021

# Analyse factorielle des correspondances : langue utilisateur / arrondissement

est-ce qu'il y a des comportements par nationalités différents ?

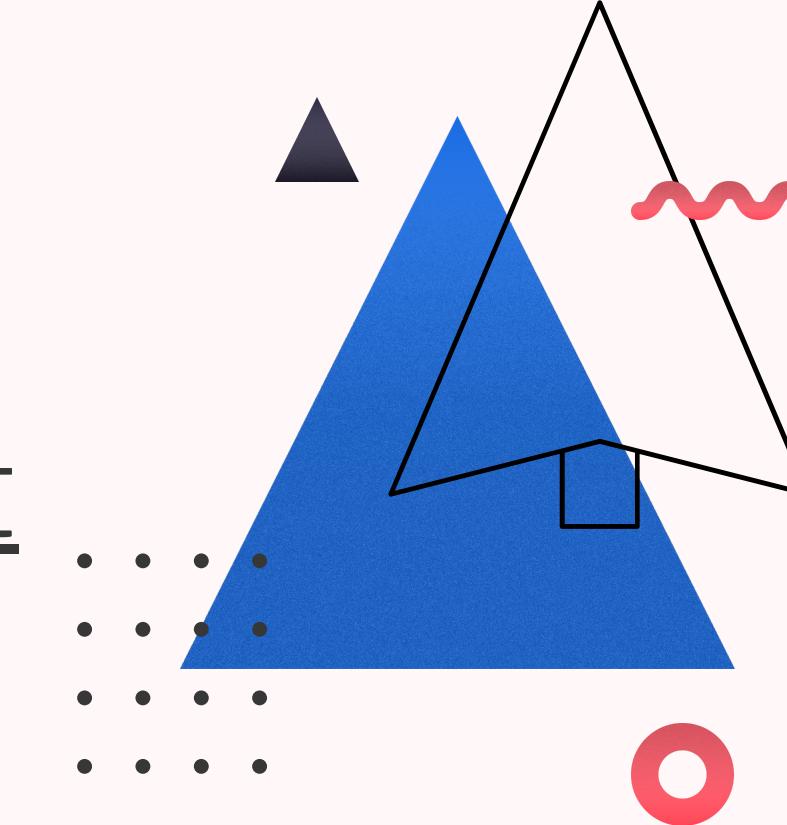


**Analysis :** un axe vraiment explicable

16e (arc de triomphe) et 7e (tour Eiffel) > touristique

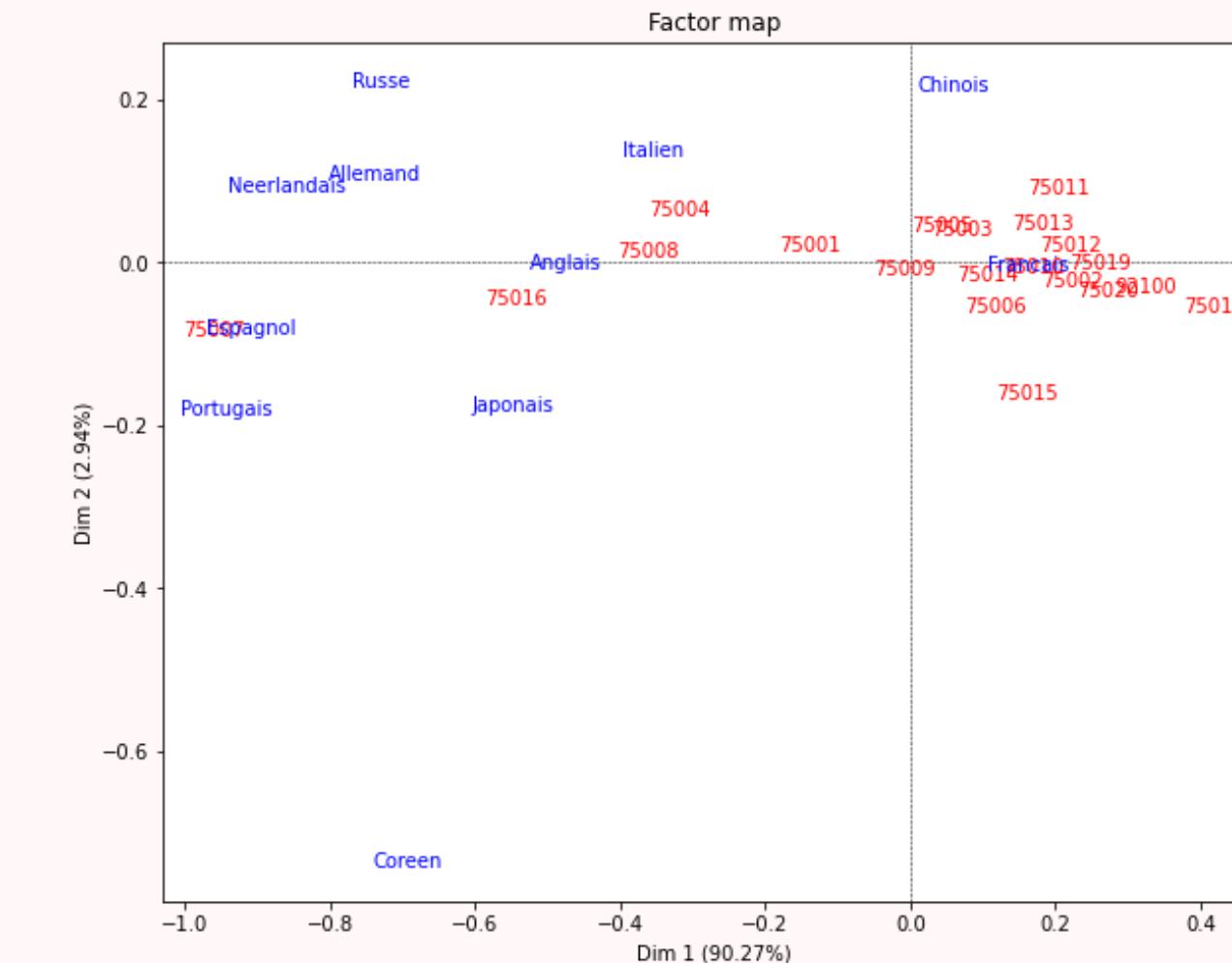
13e > communauté chinoise importante

reste > travailleurs/habitants plus permanents



Ce qui n'a pas marché :

- essayer d'expliquer le deuxième axe (vertical)
- continuer d'enlever les langues sous arrondissement qui étirent les axes > cercle infini



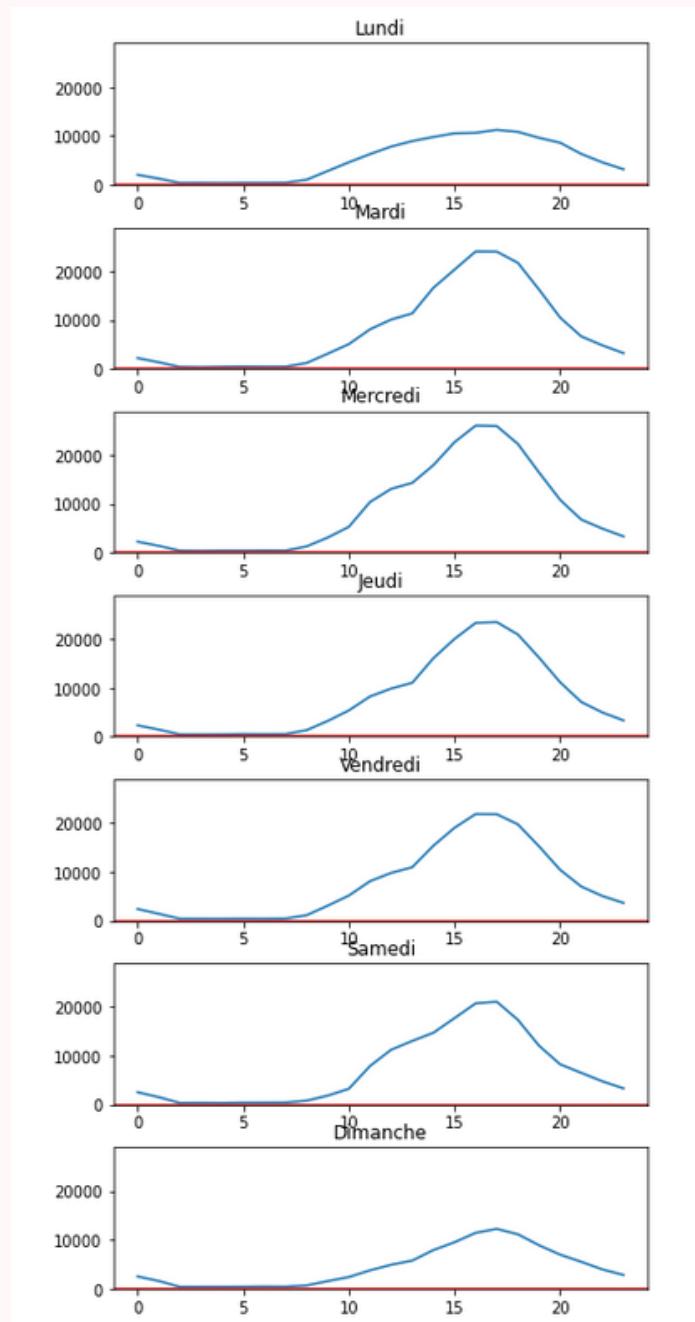
**Autres essais :**

AFC sans : Arabe, Indonésien, Polonais, Thaïlandais, et 18e ; comme le 2e axe n'est pas très explicatif, c'est difficile d'arriver à quelque chose d'homogène.

Avec les français même graphe mais un peu plus tassé sur la droite (habitants).

# AFC: jour de la semaine/ heure de connexion

-> heures de connexions segmentée en 24 variables

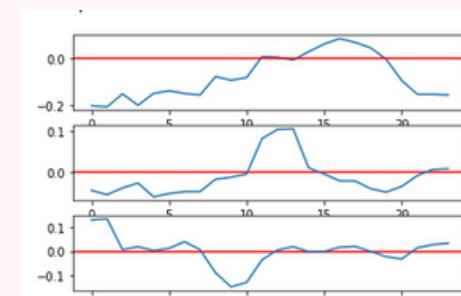


## Interprétation des axes:

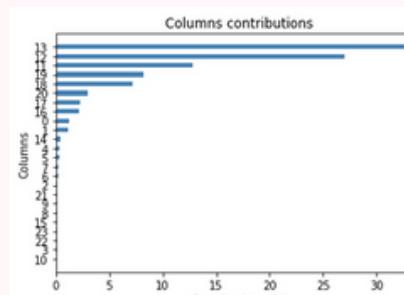
**axe 1:** travail dans l'aprem

**axe 2:** midi

**axe 3:** travail le soir plutôt que le matin



Coordonnées des 3 axes

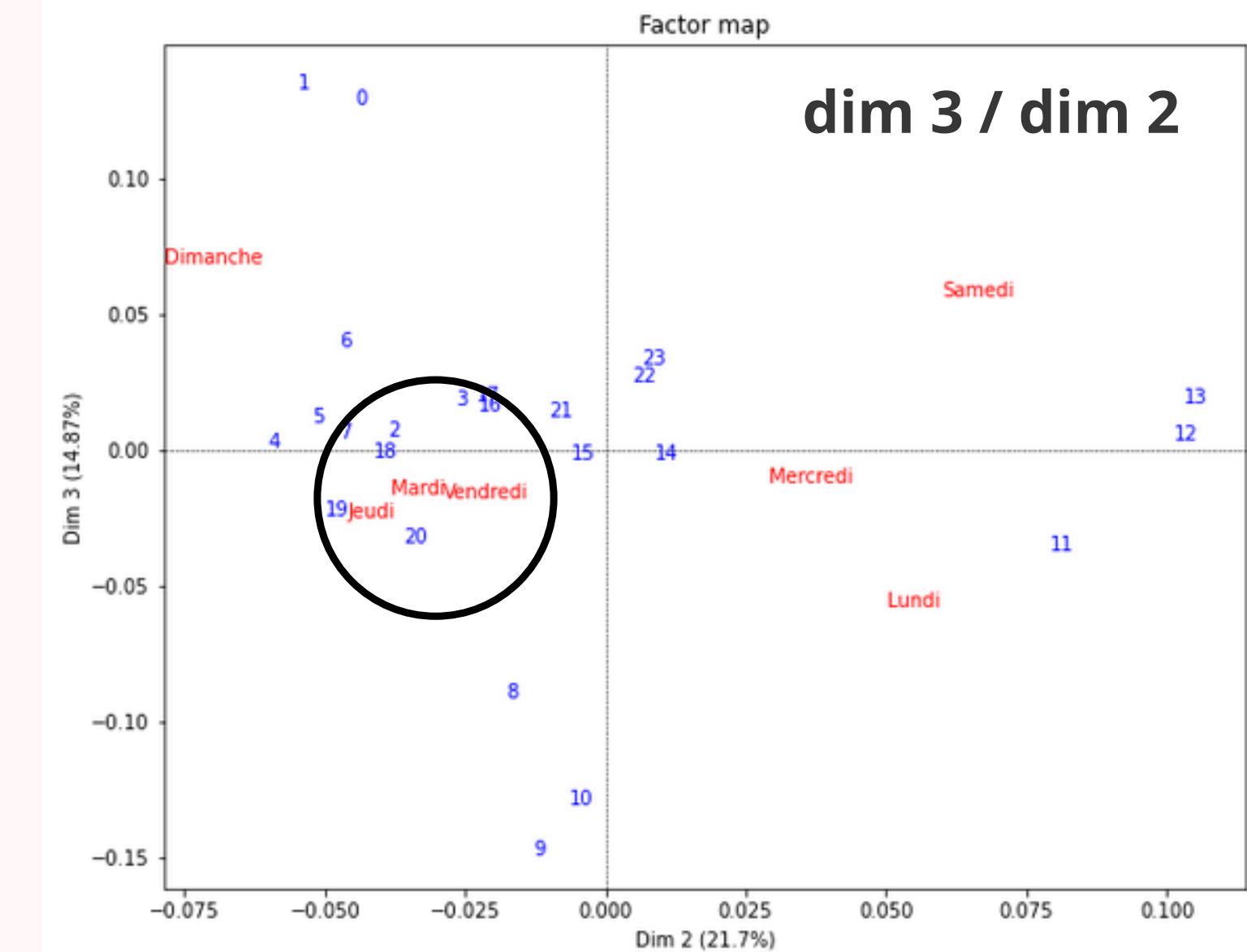


Contribution axe 2

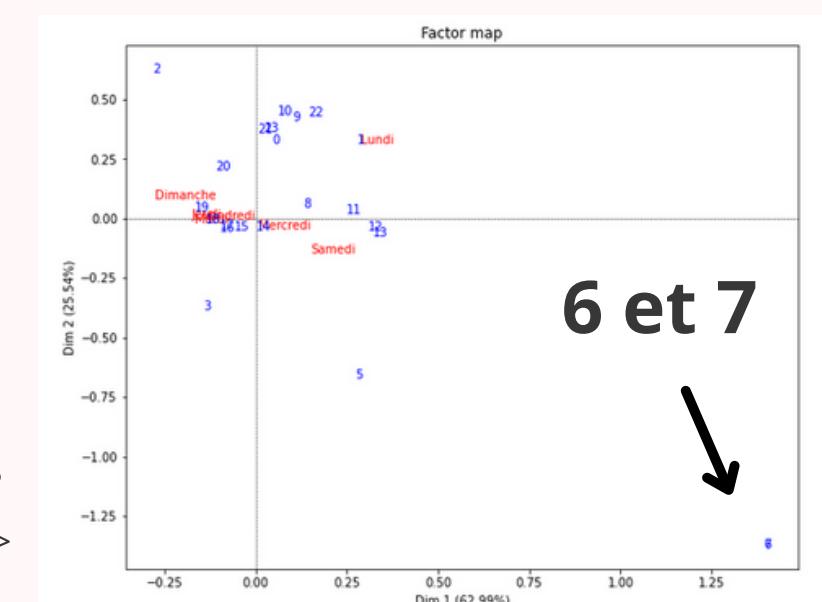
## Difficultés:

- heures de la nuit en supplémentaire
- découpage des jours

Scree plot final

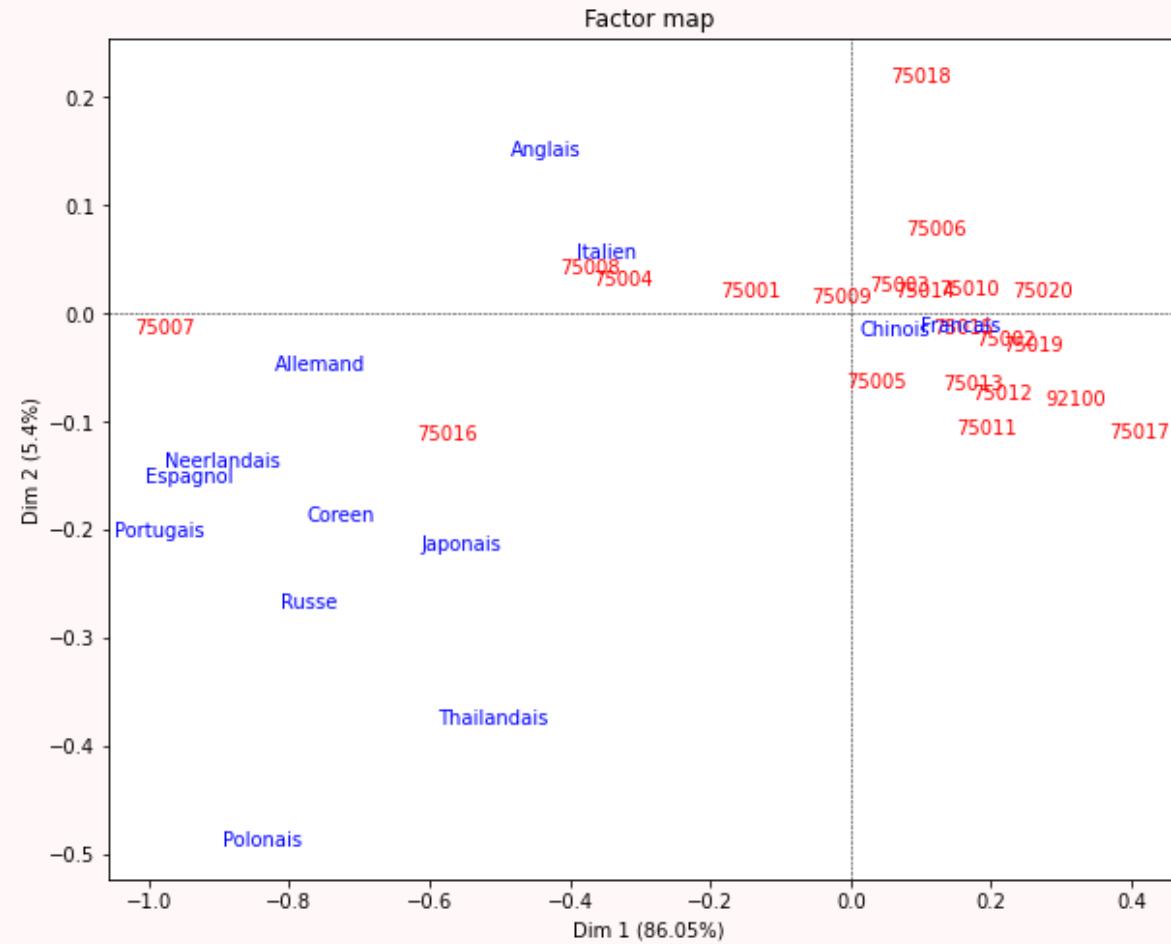


plot final axe 3 et 2

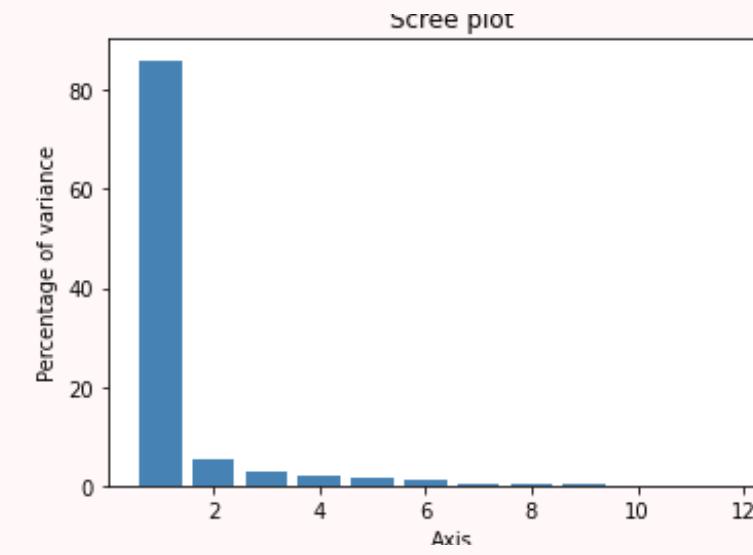


avant d'enlever les  
heures de la nuit >

# Annexe 1: AFC - arrondissement VS langue utilisateur

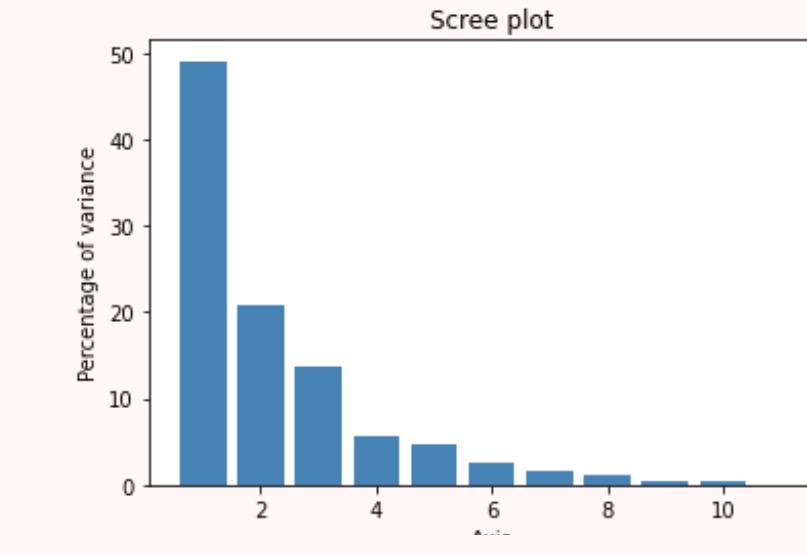


Après deux itérations, sans les langues peu représentées ou qui étirent les axes (Arabe et Indonésien), en gardant les Français  
Axe touristique <-> habitant très renforcé



Pourcentage d'explication de la variance (1 itération)

Les valeurs propres sont claires : un seul axe est représentatif

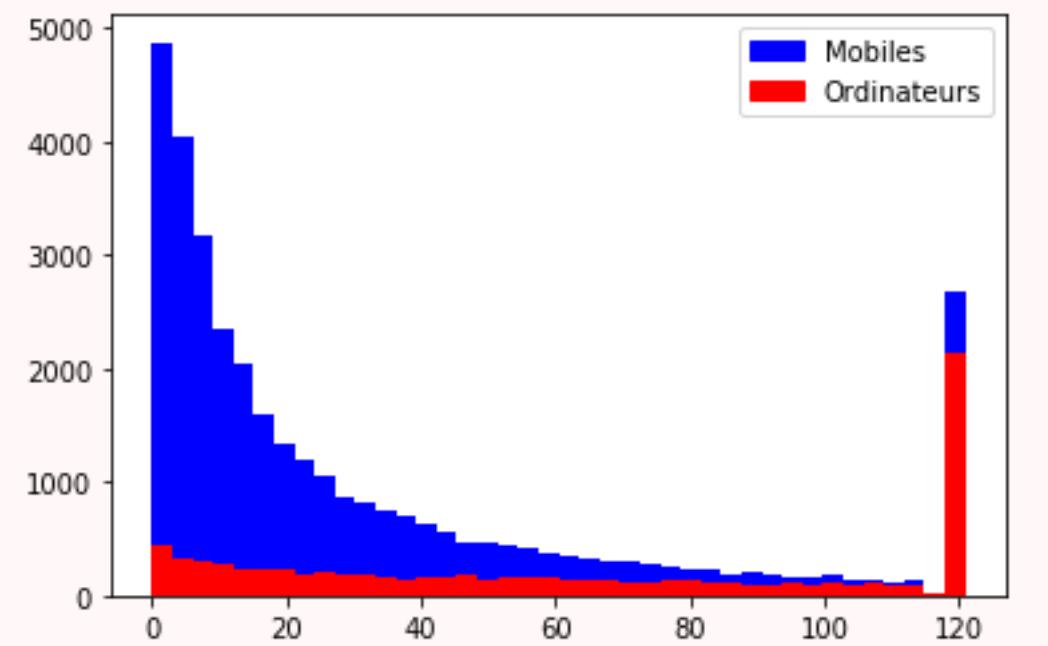


Pourcentage d'explication de la variance (sans les français)

L'axe touristique <-> habitant est moins renforcé. Il y a peut être plus une partie communauté.

## à faire en plus:

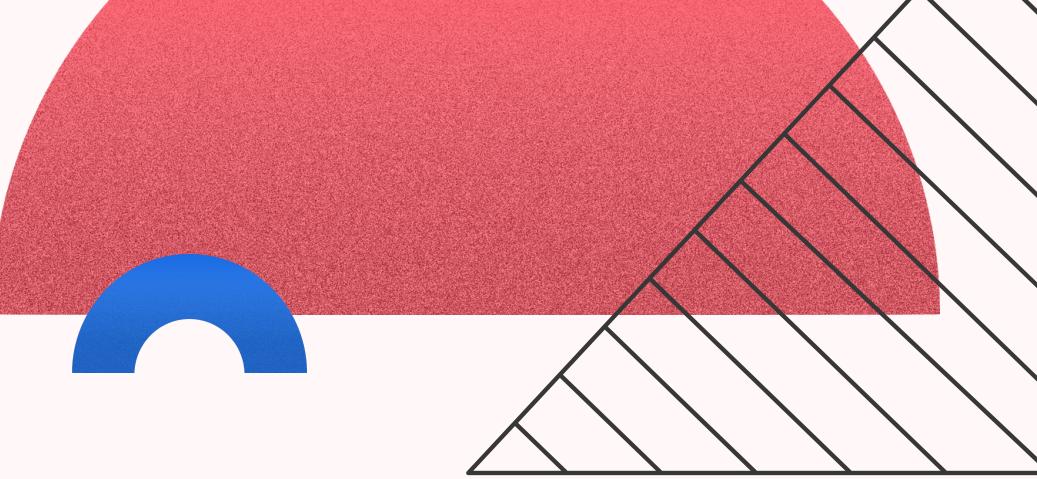
- regrouper les langues par continents (Afrique du Nord, Asie, Europe de l'Est, ...)
- comparer les marques d'appareils utilisés à la langue



Nombre de connexions (en ordonnées)  
par temps de connexion (en abscisses)

Analyse efficace : les gens utilisant des ordinateurs restent connectés bien plus longtemps (au dessus de 2h) que ceux utilisant des mobiles.

Cependant, le nombre de connexion par mobile reste 10 fois plus élevée.



## Annexe 2 : temps de connexion par type d'appareil

