

# Une enquête sur les profils « data »

*Quelle place pour R et ses utilisateurs ?*

**Antoine GIRARD**

*Data analyst freelance*

06 33 72 87 12

[Antoine.girard@datag.fr](mailto:Antoine.girard@datag.fr)

## Une enquête quantitative par sondage

*indépendante et sans parti pris...*

### Cible



Métiers **opérationnels** de la **data**

- *Data Analyst, data Scientists, data Engineer*
- Professionnels en activité

### Méthode



Sondage **empirique**

- 311 réponses complètes récoltées

# Diffusion du questionnaire

---

**Diffusion** du questionnaire exclusivement via **LinkedIn**

*Mars / Avril 2024*



**Biais de sélection**

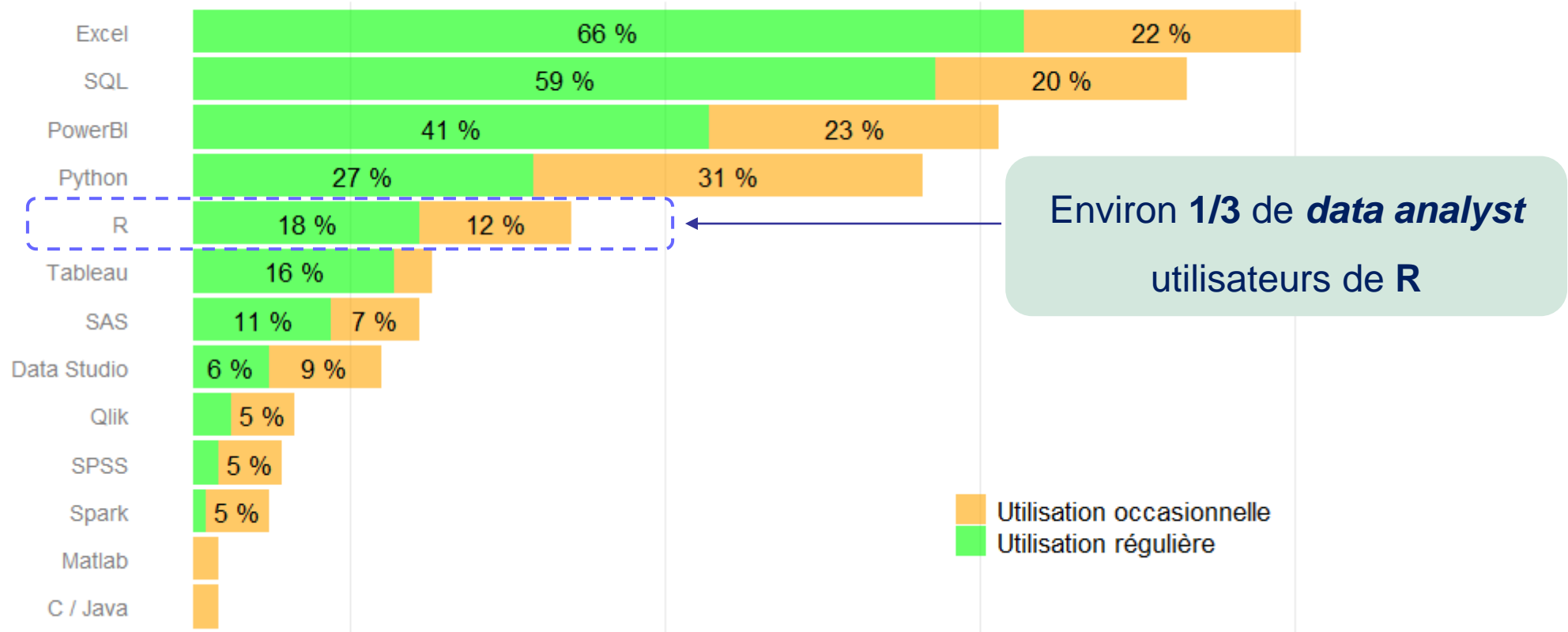
de par l'algorithme de LinkedIn

**Ciblage** « manuel » de **profils data**

**Invitations** « aléatoires »

# Quelle place pour R parmi tous les outils ?

## Fréquence d'utilisation des outils data *Data analyst / chargés d'études uniquement*



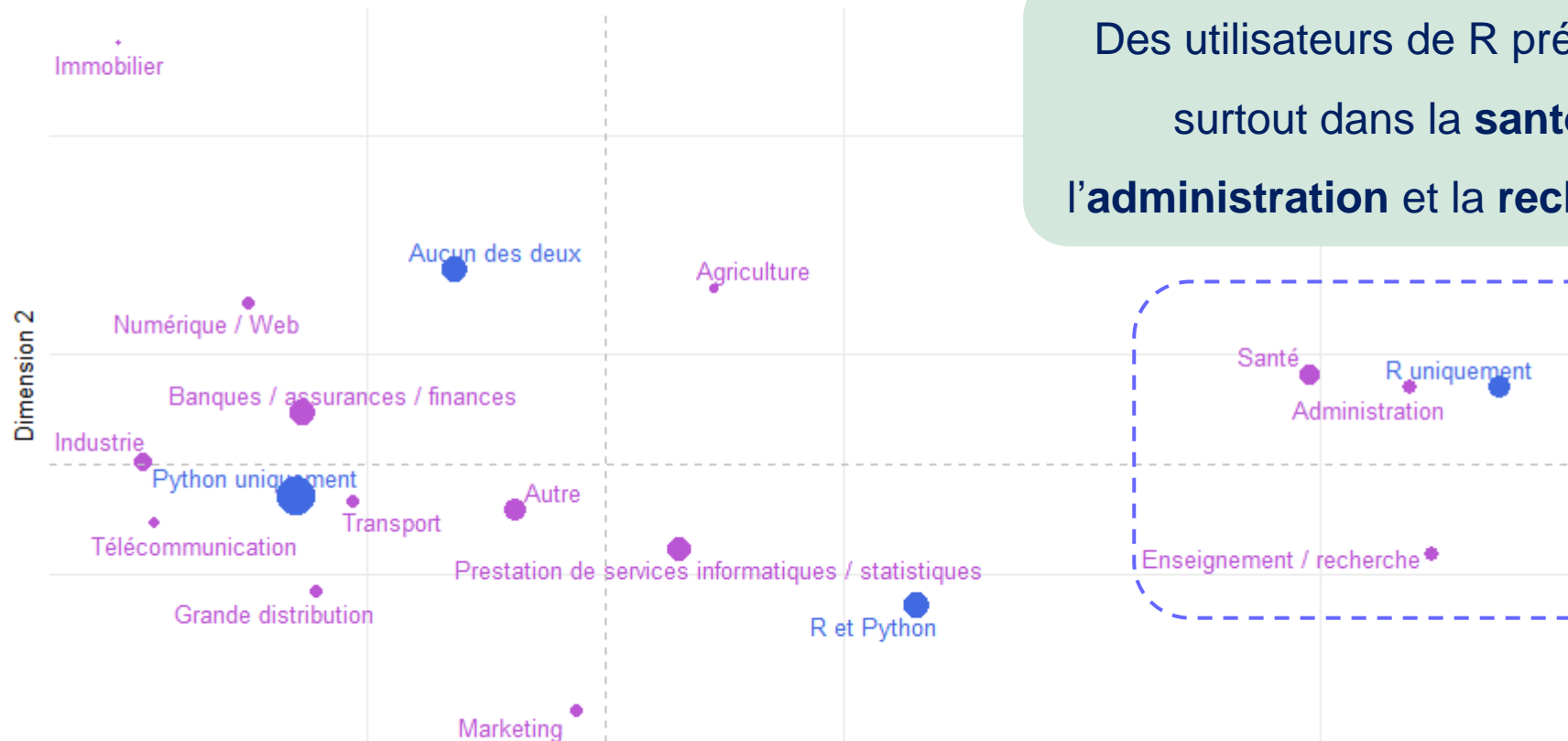
Question posée : pour chacun de ces outils que vous avez déjà utilisés, indiquez désormais si vous l'utilisez régulièrement, occasionnellement ou jamais dans votre poste actuel.

*Data scientists (base faible)* : 40% d'utilisateur, mais tous utilisateurs de Python également

# Qui utilise R ?

## Liens entre le secteur d'activité et l'utilisation de R / Python

*Tous profils*



Des utilisateurs de R présents surtout dans la **santé**, l'**administration** et la **recherche**

Méthode utilisée : analyse des correspondances entre secteur d'activité et l'utilisation conjointe de R / Python

*Pas de différences significatives sur l'âge ou le sexe R vs Python*

# Pourquoi utiliser R ?

## Corrélations missions / outils

*Tous profils*

	Data Studio	Excel	PowerBI	Python	Qlik	R	SAS	Spark	SQL	Tableau
Visualisation de données / dataviz	-0.22	0.17	0.26	0.01	0.01	0.03	-0.13	-0.18	0.1	0.21
Tableaux de bord / reporting	-0.16	0.3	0.35	-0.15	0.08	-0.1	0	-0.2	0.13	0.19
Suivi de modèles / ML Ops	0.03	-0.08	-0.09	0.27	-0.01	-0.01	-0.08	0.2	0.02	-0.04
Statistiques	-0.1	0.05	-0.16	0.01	0.02	0.42	0.16	-0.07	-0.16	0.08
Modélisation / mise en place de modèles de Machine Learning	0.03	-0.12	-0.11	0.35	-0.04	0.25	-0.05	0.2	-0.16	0
Mise en production /déploiement de modèles	0.02	-0.09	-0.02	0.12	0.05	-0.03	-0.12	0.14	0.11	0
Flux de données / data pipelines	0.03	-0.13	0.03	0.28	0.16	-0.26	-0.14	0.24	0.32	0.08
Expérimentation / tests	0.03	-0.09	-0.03	0.22	-0.02	0.1	-0.09	0.18	0.04	0
Construction ou gestion de bases de données	0.09	0.06	-0.05	0.01	0	-0.04	0.09	0.04	0.18	0.12
Analyse métier / Business insights	-0.18	0.14	0.35	0.13	0.1	-0.19	-0.11	-0.13	0.28	0.22

Des utilisateurs de R qui se démarquent des autres outils  
par l'importance des **statistiques**

# Synthèse R

---



R (largement) distancé par Python mais  
utilisé par 1/3 des *data analyst*



Particulièrement présent sur les secteurs de la santé et les  
administrations



Une spécificité fondée avant tout sur les statistiques

# Mise à disposition des résultats

---

- **Données brutes et traitements** :

- Document d'analyse « focus R » plus complet
- Fichiers csv et scripts R de traitements de données
- Lien : [https://gitlab.com/angirard1/enquete\\_data\\_2024/-/tree/main/](https://gitlab.com/angirard1/enquete_data_2024/-/tree/main/)

*Des suggestions ou critiques d'ordre méthodologique / analytique ?*

*Je prends note !*