

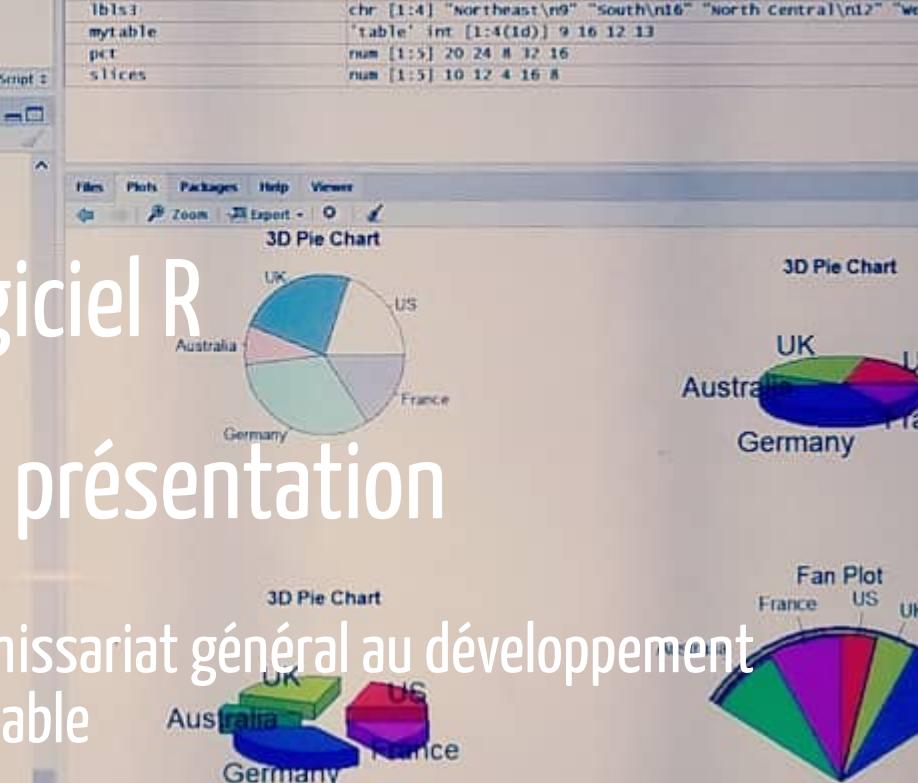
```
unit.angle * pi/180  
x - radius * cos(t2p), y - radius * sin(t2p))  
  
ges * dx[i]))  
[i], x[i + 1], length.out = n),  
(Psy, 0), density = density[i], angle = angle[i],  
[i], col = col[i], lty = lty[i],  
[0:1]))  
labels[i])  
rchar(lab) {  
  * p$y, c(1, 1.05) * p$y)  
  [i] * Psy, labels[i], xpd = TRUE,  
(p$x < 0, 1, 0), ...)  
  
)  
graphics>  
warnings() to see them)  
main="3D Pie chart",  
explode=0.3,  
  
"explode" is not a graphical parameter:  
: "Explode" is not a graphical parameter  
is not a graphical parameter  
explode=0.3,  
  
warnings() to see them)  
main="Fan Plot")  
warnings() to see them)
```

Le logiciel R

Une (rapide) présentation

Journée de la donnée 2019 - Commissariat général au développement durable

18 novembre 2019



R c'est quoi ?

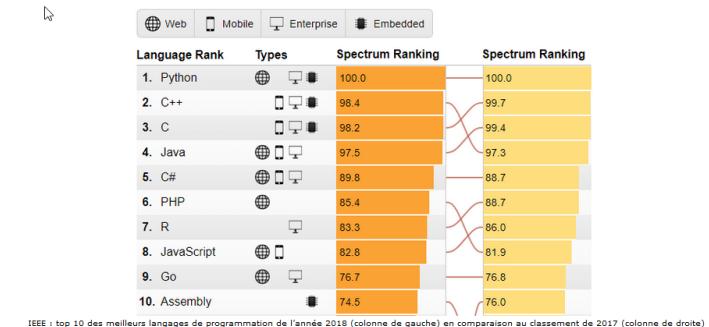
- Un langage de programmation interprété exécutable pas à pas → lignes de code
- Crée au milieu des années 90 mélange de S et du langage Lisp pour le dédier à l'analyse de données et à la création de graphiques
- Un projet né en 1993, une version 1.0.0 publiée le 29 février 2000
- Un logiciel libre (GNU GPLv2)
- Utilisé en recherche, en enseignement, dans le journalisme, l'administration et en entreprise
- Que l'on peut utiliser via un IDE de référence : [Rstudio](#)
- Modulaire : système de packages installables permettant d'enrichir R de fonctionnalités. Le [CRAN](#) rassemble les packages ayant été soumis à un processus de validation mais on peut trouver aussi beaucoup de packages hors du CRAN sur des forges (github, gitlab...).



Un logiciel en pleine explosion

- Développement du logiciel libre
- Développement des sciences de la données

Top 10 des langages de programmation (classement IEEE)



R et le MTES-MCT

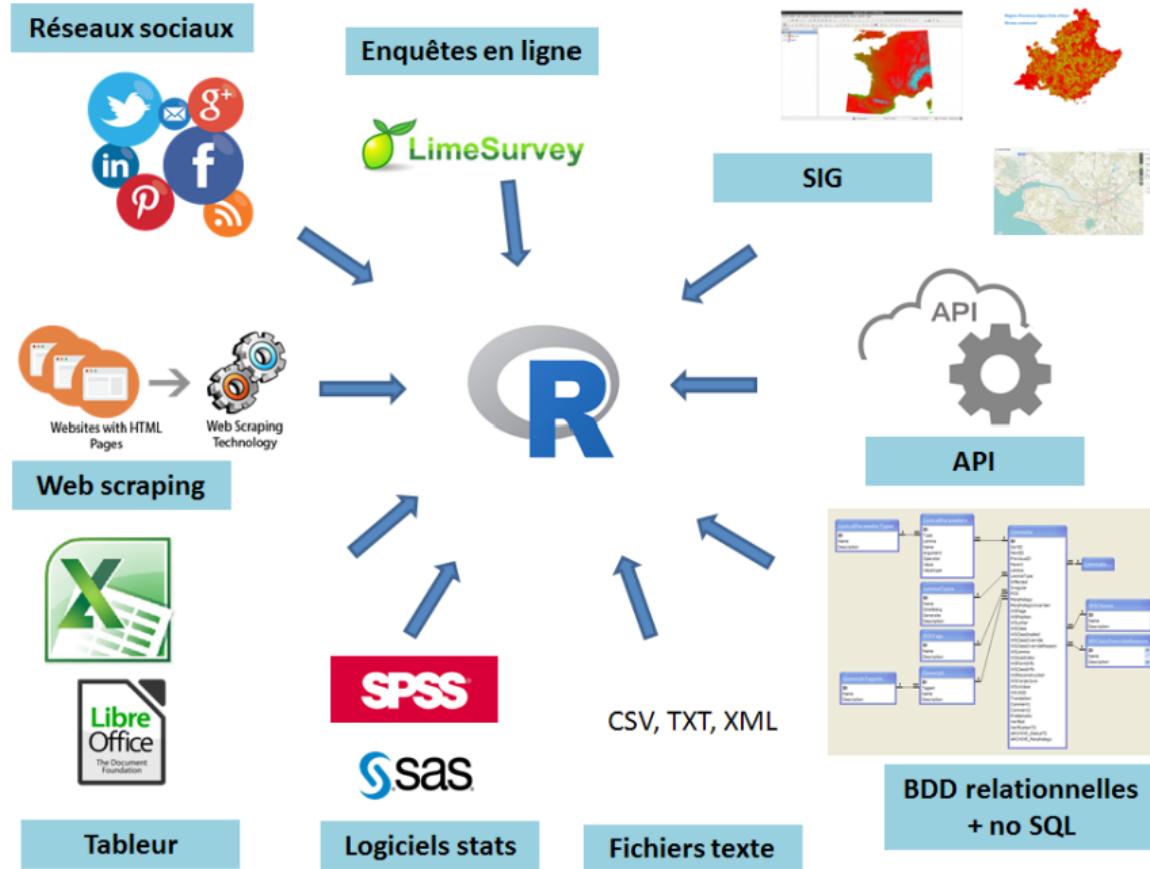
- R répond à des besoins croissant de traitements et d'analyse de données
- R s'inscrit dans la promotion du logiciel libre du ministère (Qgis, PosgreSQL, Libre Office...)
- R est l'avenir dans la statistique publique : adopté comme standard par l'Insee (à l'horizon 2025) et divers Services Statistiques Ministériels (dont le SDES).
- R remplace progressivement SAS au sein de la statistique publique



Que peut on faire avec R ?

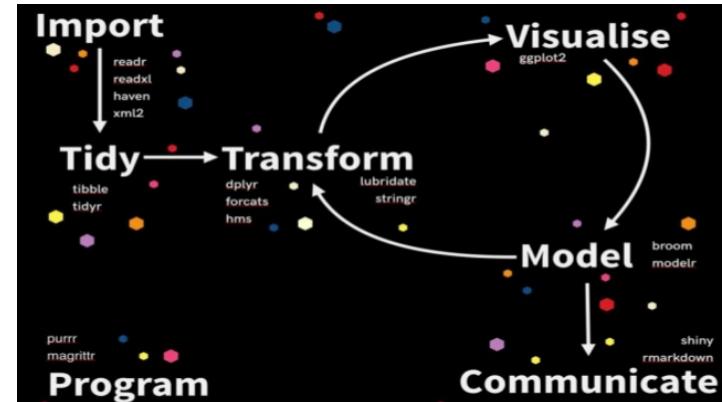


Lire toutes sortes de données...



... et gérer l'intégralité de la chaîne de vie de la donnée

- Mettre en forme
- Analyser
- Modéliser
- Présenter / publier des résultats
 - Datavisualisation / Tableaux / Cartes
 - Publications / Rapports
 - Livres
 - Présentations
 - Applications web



Idées reçues



Idées reçues

Complexe

- Développement rapide hors sphère académique
- Formation de « littéraires »



Idées reçues

Complexe

Peu
performant

- Ok jusqu'à 10 000 000 lignes et quelques dizaines de colonnes
- Version serveur (disponible au SDES), parallélisation
- Requêtage simple de bases de données



Idées reçues

Complexe

- Tidyverse : une syntaxe unifiée pour l'essentiel des travaux data

Peu
performant

Hétérogène



Idées reçues

Complexe

- Journal of statistical software
- CRAN

Peu
performant

Hétérogène

Pas de
certifications



Idées reçues

Complexe

- 2 versions par an

Peu
performant

Hétérogène

Pas de
certifications

Pas stable



Les avantages

Les avantages

Gratuit

- Installable sur tout les postes du ministère
(contrairement à SAS)



Les avantages

Gratuit

- Partage du code, mutualisation

Libre



Les avantages

Gratuit

- Un outil pour vraiment faire du traitement de données complexe
- Permet simplement de faire des choses qui seraient très complexes en bureautique

Libre

Adapté



Les avantages

Gratuit

- Un seul outil pour faire des cartes, graphiques et le rapport.
- Pas besoin de passerelles

Libre

Adapté

Intégré



Les avantages

Gratuit

- Traçabilité
- Reproductibilité
- Capitalisable (programmation fonctionnelle)

Libre

A utiliser avec git et gitlab/github pour tirer parti de toute cette richesse

Adapté

Intégré

En ligne de code



Les avantages

Gratuit

- Développements rapides
- Une documentation riche
- Une communauté utilisateurs sur laquelle s'appuyer

Libre

Adapté

Intégré

En ligne de code

Populaire



Les limites

- Programmation en lignes de code
- Foisonnement de packages complémentaires
 - Choix
 - Maintenance de certains packages
 - Gestion des dépendances entre packages



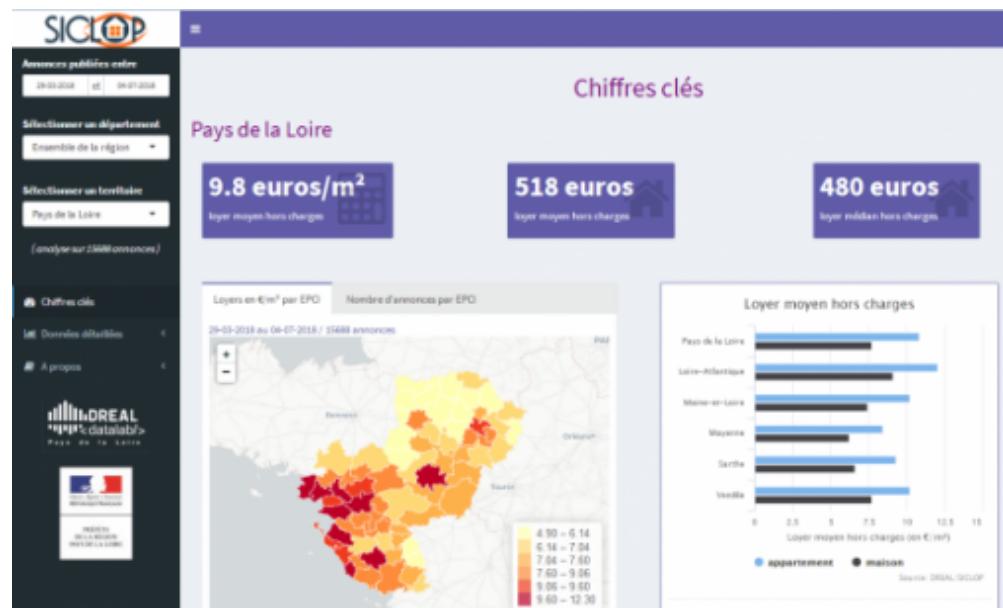
Quelques exemples d'utilisations



SICLOP

- Application web de valorisation de données scrappées sur ouest france immo et nettoyée.

Packages utilisés : Tidyverse, Shiny, ggplot2, highcharter, leaflet



Rapport de conjoncture logement construction

- Rapport reproductible de valorisation des données de conjoncture.

Packages utilisés : Tidyverse,
Rmarkdown,
ggplot2,drealthemes,drealdown

Lien



Observation, études et statistiques

Le suivi de la conjoncture logement construction en Pays de la Loire

2ÈME TRIMESTRE 2019 : LE NOMBRE DE DEMANDEURS D'EMPLOI AU PLUS BAS DEPUIS MI-2011

Maël THEULIERE - Dreal Pays de la Loire - mael.theuliere@developpement-durable.gouv.fr

2019-09-20

TABLE DES MATIERES

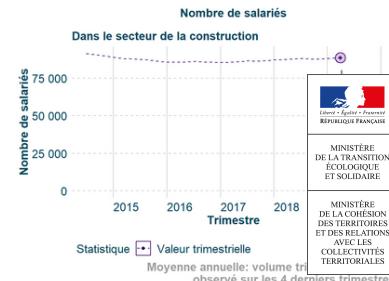
[Emploi](#)
[Intérim](#)
[Demandeurs d'emploi](#)
[Créations d'entreprises](#)
[Construction neuve](#)
[Vente de logements neufs](#)
[Appartements](#)
[Maisons](#)
[Prêts à taux zéro](#)
[Eco prêts à taux zéro](#)
[Financement du logement social](#)
[Aides de l'Anah](#)

Emploi

88 533 salariés travaillent dans la construction fin mars

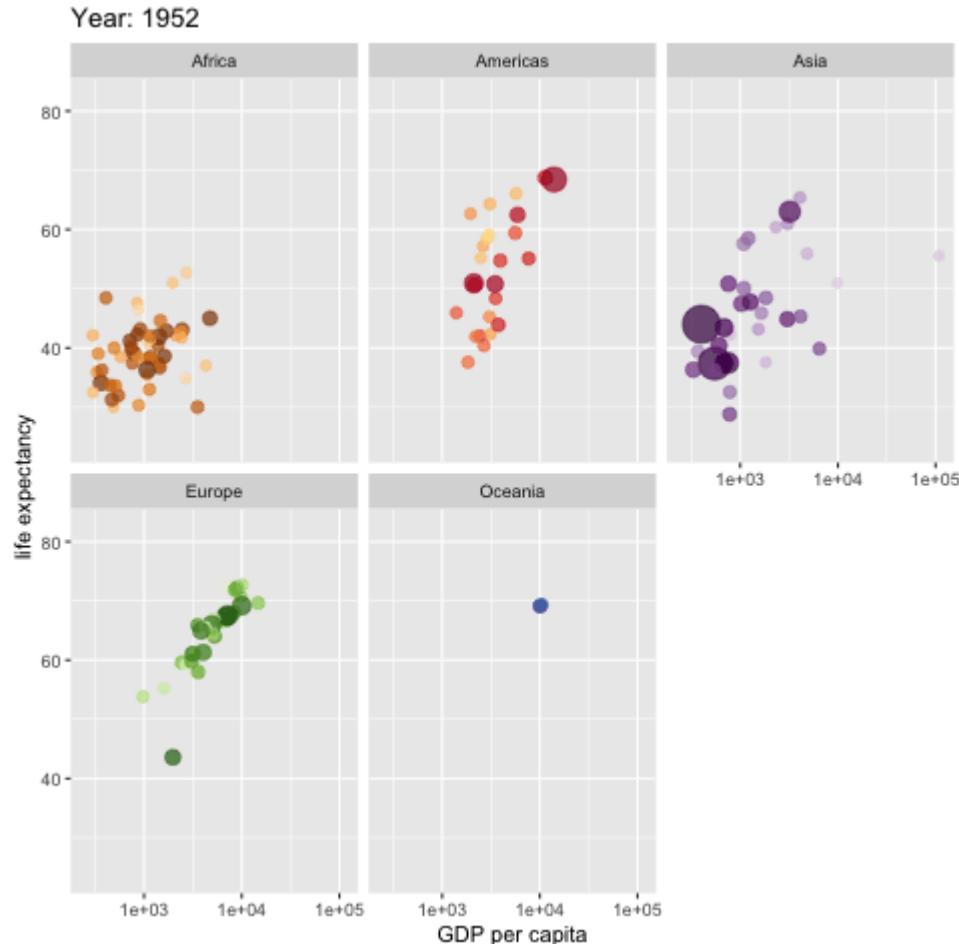
+1,5% sur un an

Fin mars 2019, 88 533 salariés travaillent dans le secteur de la construction, c'est 1,5 % de plus qu'en auparavant, une croissance qui s'accélère par rapport à 2018. Cette hausse de l'emploi salarié est cependant moins forte qu'au niveau national (+ 2,7 %). Sur la période, la hausse de l'emploi dans la construction est portée par la Loire-Atlantique (+ 3,3 %) et la Vendée (+ 3,7 %).



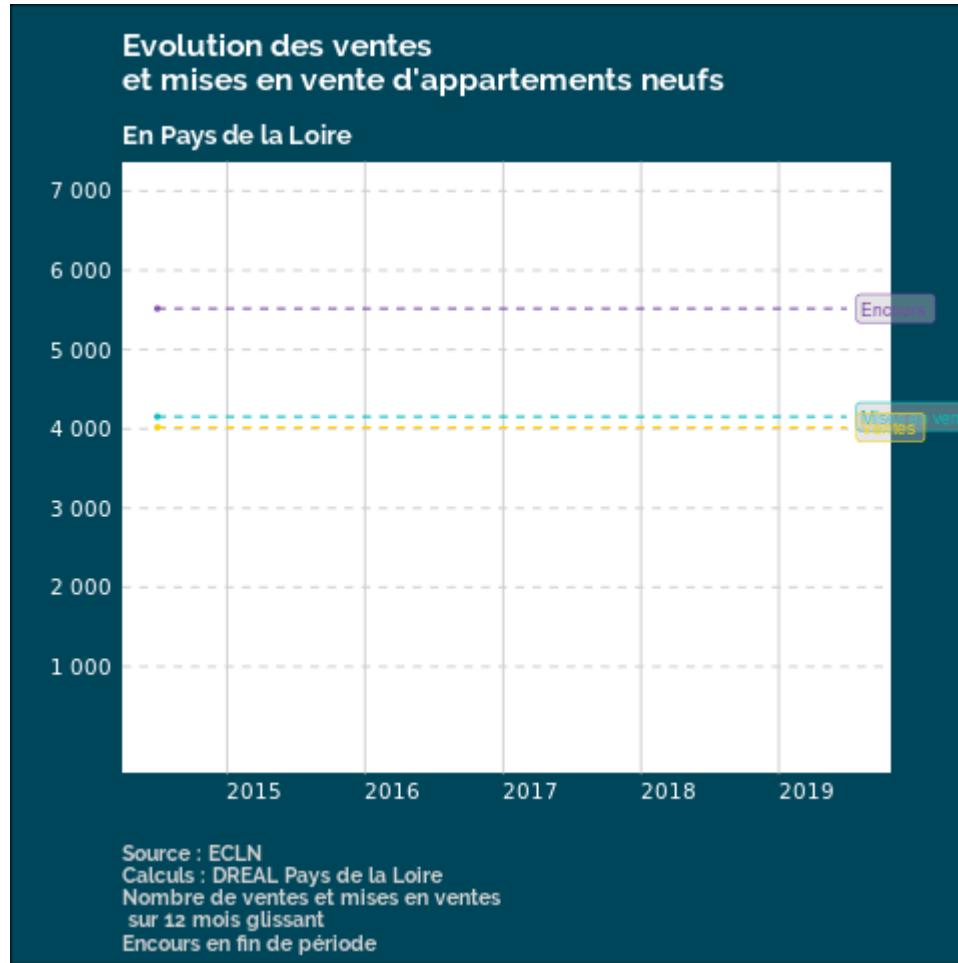
gganimate

Permet la réalisation des graphiques animés



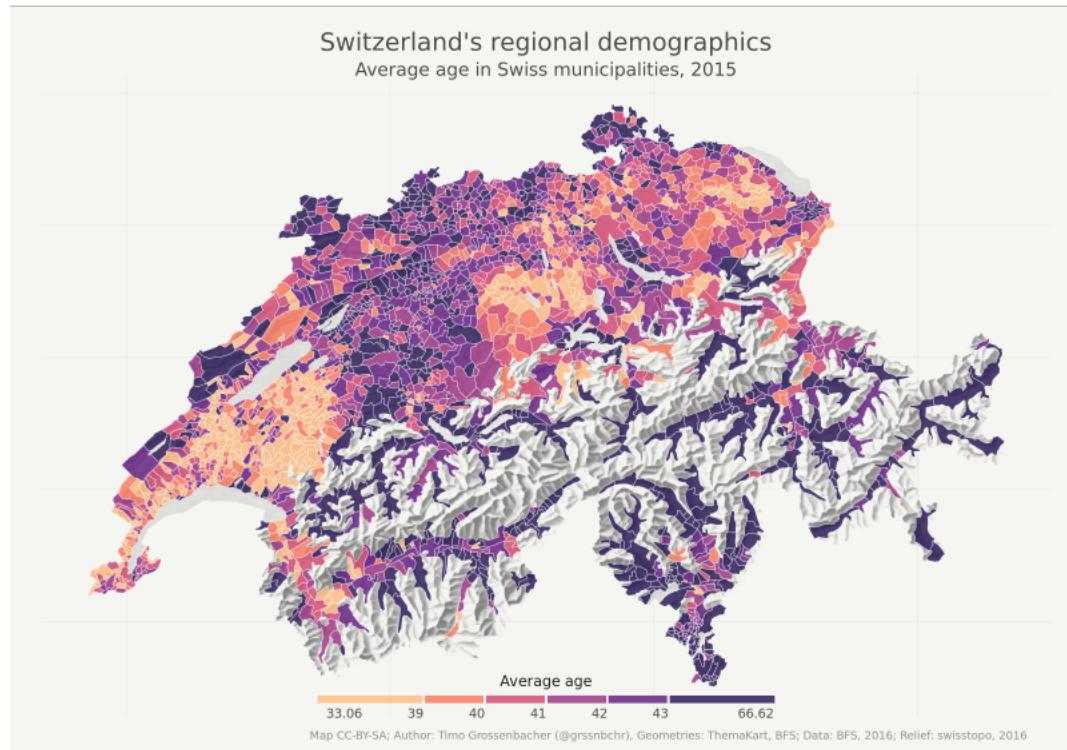
gganimate

Permet la réalisation des graphiques animés



Cartoviz

Exemple d'usage de R pour cartographier l'âge moyen au sein des villes suisses.



R par rapport à d'autres outils



R versus Python



- Dédié au traitement des données
- Programmation fonctionnelle
- Meilleur sur la dataviz



- Généraliste
- Language objet
- Pas les mêmes problèmes de dépendances

R versus Qgis



- Ligne de code
- Permet de traiter tout type de données
- Outil intégré



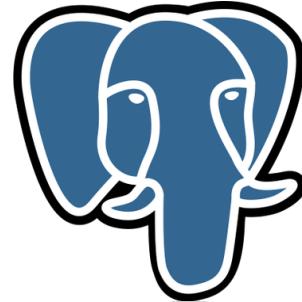
- Interface fenêtrée
- Dédié aux données géomatiques
- Dédié à la cartographie



R versus PostgreSQL



- Language in memory
- Gestion de l'ensemble du cycle de vie de la donnée
- Toutes les fonctionnalités géomatiques ne sont pas (encore) implémentées



- Traitement des données sur disque
- Dédié au traitement de la donnée
- Permet tout type de traitement géomatique avec Postgis

Se mettre à R au MTES-MCT



L'offre de formation du CVRH

Modules déployés

- Premiers programmes en R
- Préparation des données
- Statistiques descriptives
- Analyses des données multidimensionnelles
- Datavisualisation : produire des graphiques, des cartes et des tableaux

Se mettre à R au MTES-MCT

L'offre de formation du CVRH

Modules en cours de finalisation

- Publications reproductibles avec RMarkdown
- Analyse spatiale

Modules à venir

- Applications interactives avec Shiny
- Big data et optimisation du code
- Le travail collaboratif avec R

Mis au fil de l'eau en open source sur [github](#), avec un [site web](#) d'accès aux versions des modules les plus à jour



Se mettre à R au MTES-MCT

Des outils

Un kit d'installation prêt à l'emploi

Un serveur Shiny pour le MTES - MCT

Une faq pour le partage de connaissance



Se mettre à R au MTES-MCT

Des lieux d'échange

Un salon des utilisateurs R sur Ariane

Une liste d'échange entre utilisateurs

Pour s'inscrire, envoyer un message vide avec le titre "subscribe labo.communaute-r" à l'adresse sympa@developpement-durable.gouv.fr ou cliquer directement sur [ce lien](#)



Se mettre à R au MTES-MCT

Des événements de communication autour de R

- Cette journée !
- Des ateliers utilisateurs à Orléans
- Des funcamp en 2020 avec l'Insee autour de R



Quelle complémentarité avec les autres outils du ministère ?

- Qgis et R ne s'adressent pas au même public
- Géokit 3 et Géoidd sont des outils à l'avenir incertain, mais plus à cause du datalake que de R
- R est très complémentaire du datalake : se connecter aux données par API pour produire des valorisations reproductibles.
- Il est fortement conseillé d'utiliser git et gitlab/github avec R afin de tracer son travail au mieux et de travailler à plusieurs : vers une approche **DataOps** de nos travaux "data".

Une vision du futur



Une vision du futur

Aujourd'hui

- des chargés d'études perdant beaucoup de temps dans le traitement des données et l'utilisation d'outils multiples pour valoriser leurs données



Une vision du futur

Demain

- Des outils de traitements de la données intégrant l'ensemble du cycle de vie de la données et mis en oeuvre par des spécialistes du domaine : de l'intégration à la valorisation
- Des chargés d'études recentrés sur l'analyse et le portage des politiques publiques

