



Club Géomatique

Automatisation de processus de traitement de données et de production d'outils avec R et SQL

17 novembre 2022

1. – L'automatisation chez Topos



- ✓ Avant : chacun gardait ses données sur son poste en local.
- ✓ Nouveau serveur et construction d'une BDD unique avec PostgreSQL.
- ✓ Création d'un poste pour l'intégration et les traitements de données.
- ✓ Adoption de R pour automatiser ces processus.





2. – Présentation de R



- ☺ Langage de programmation dédié aux statistiques et aux traitements de données, et logiciel libre.
- ☺ Un environnement de développement, RStudio.
- ☺ Lit, transforme et écrit de la donnée.
- ☺ Visualisation : cartes, graphiques, rapports.
- ☺ Les packages apportent une grande polyvalence, selon le besoin :
 - Formats de données et langages
 - Fonctions disponibles => traitements possibles

3. – Intégration de données dans la base de données

🍷 RP INSEE

🍷 Assemblage des documents d'urbanisme

🍷 Registre des preuves de covoiturage

Dans le même environnement, on peut :

- **charger** les packages, les variables, les données (API).

```
response <- GET("https://www.data.gouv.fr/api/1/datasets/5e8ee97c16601da4ee24ffb7/")
json <- content(response, type = 'text', encoding = "UTF-8")
lien <- fromJSON(json)
url <- lien[["resources"]][["url"]]

registre <- list()

for(i in 1:length(url)){
  mois <- url[i] %>%
    str_sub(start = -11) %>%
    str_replace("-", "_")

  download.file(url[i], paste0("",chemin_donnees,"registre_",mois,""), mode="wb")

  csv <- read.csv2(paste0("",chemin_donnees,"registre_",mois,""), encoding = "UTF-8")

  registre[[i]] <- csv
}
registre <- rbindlist(registre, fill=TRUE)
registre <- registre %>%
  as.data.frame()
```



3. – Intégration de données dans la base de données

🍷 Registre des preuves de covoiturage

Dans le même environnement, on peut :

- **charger** les packages, les variables, les données (API),
- **transformer** les données (géolocaliser, géocoder, nettoyer, jointure, traitements spatiaux...).

```
registre_end <- registre %>%  
  select(journey_id, journey_end_lon, journey_end_lat) %>%  
  st_as_sf(coords = c("journey_end_lon", "journey_end_lat"), crs = 4326) %>%  
  st_transform(crs = 2154) %>%  
  rename(geom_end = geometry) %>%  
  as.data.frame()  
  
registre_sf <- registre_start %>%  
  left_join(registre_end, by = c("journey_id" = "journey_id"))
```



3. – Intégration de données dans la base de données

🍷 Registre des preuves de covoiturage

Dans le même environnement, on peut :

- **charger** les packages, les variables, les données (API),
- **transformer** les données (géolocaliser, géocoder, nettoyer, jointure, traitements spatiaux...),
- **intégrer et paramétrer** les données avec des requêtes SQL.

```
dbExecute(conn, paste0("GRANT ALL ON TABLE ",schema,".",table," TO groupe_admin;"))  
dbExecute(conn, paste0("GRANT SELECT ON TABLE ",schema,".",table," TO groupe_utilisateur;"))  
dbExecute(conn, paste0("COMMENT ON TABLE ",schema,".",table," IS 'Registre de preuves de covoiturage - DGITM - ",date,"';"))
```

4. – Création d'outils

Territoires des proximités

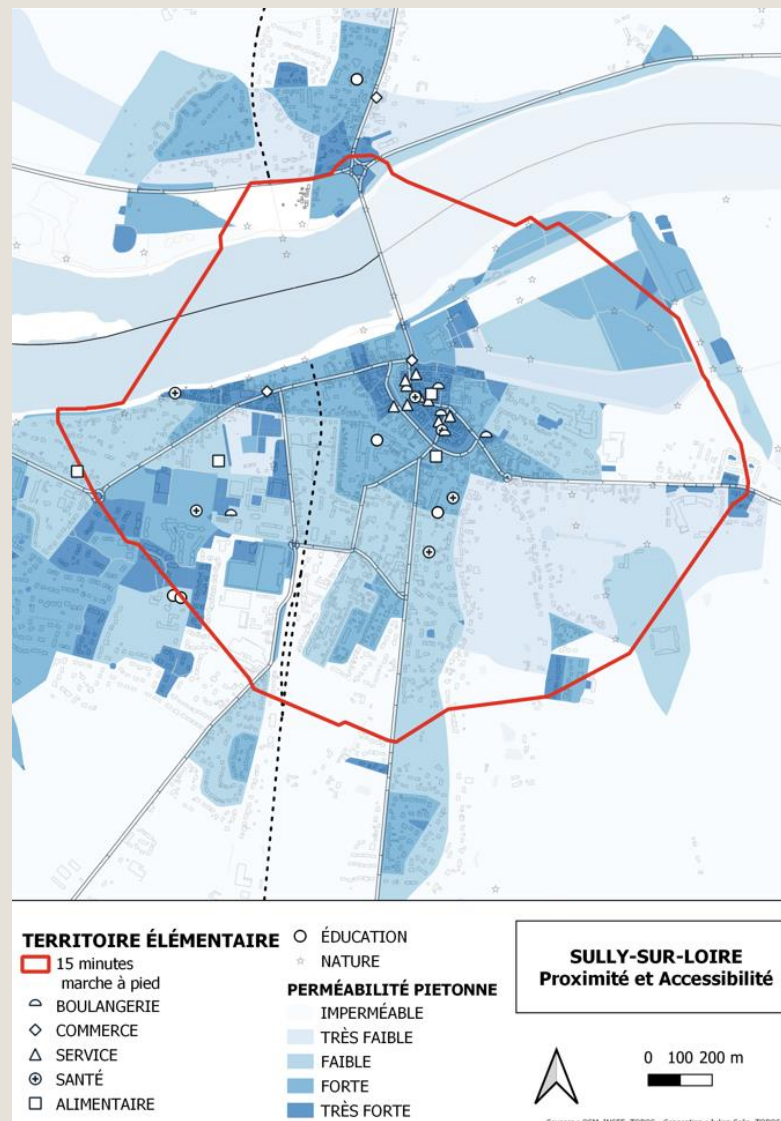
Outil :

- modélisation du territoire sur une thématique particulière.
- croisement de données en vue d'analyse.
- fiabilise la méthodologie, la documente et automatise la mise à jour.

Territoires des proximités : accessibilité des équipements de la vie quotidienne.

Grands principes du script R :

- Combinaison de packages selon les besoins.
- Isochrones (docker openrouteservices).
- Intersection spatiale.
- Intégration en BDD (requête SQL).



4. – Création d'outils



Corrections topologiques

Un sujet compliqué qui revient très souvent (occsol, assemblage des DU...), d'où la nécessité d'avoir une **méthodologie répliquable et automatisée**.



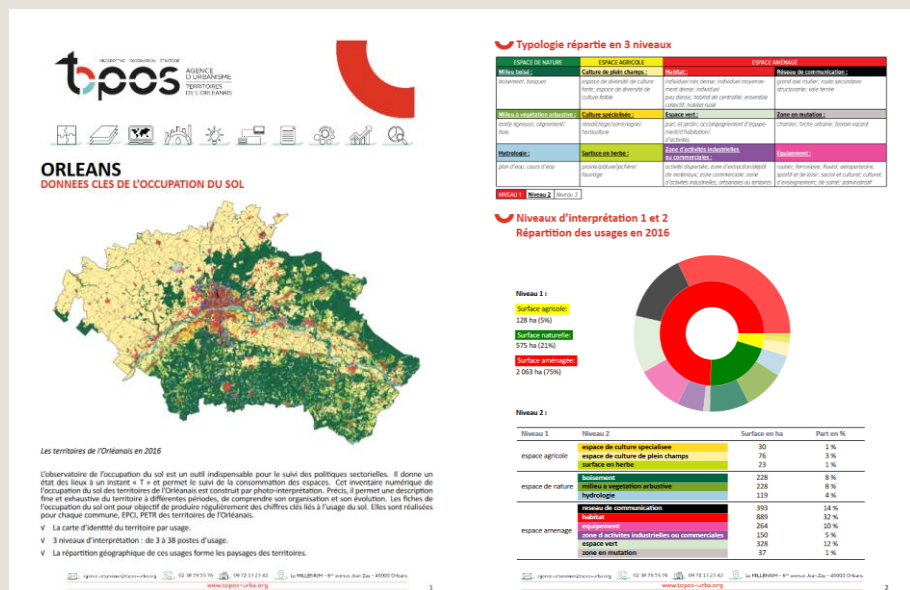
Un script qui mélange **plusieurs langages** :
Grass, PostGIS, R, QGIS.



5. – Développement de rapports automatisés dans markdown

Fiches Occupation du Sol & Parc d'activité économique

- Génération automatique de fiches PDF dataviz (cartes, graphiques, chiffres-clés)
- Mélange des fonctionnalités : R, requêtes SQL, python (Plotly), latex/html.





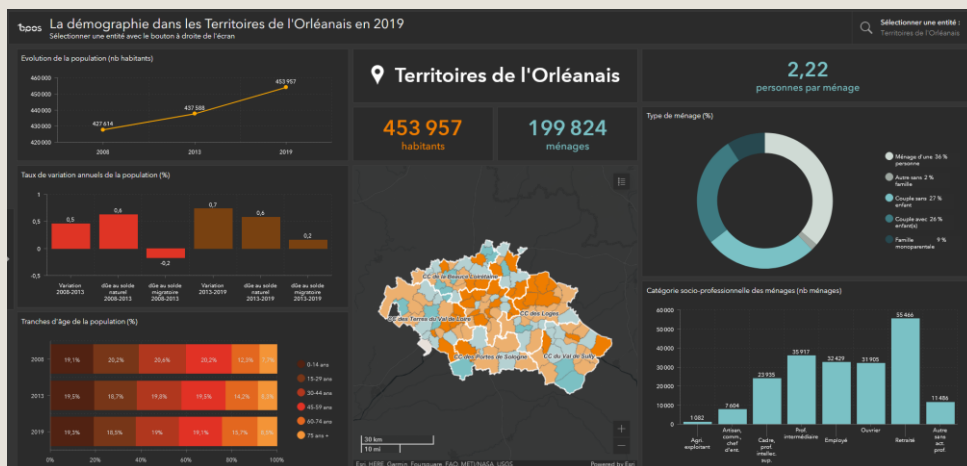
6. – Echanges avec les chargés d'étude

Création d'indicateurs Habitat

- vues à partir du RP INSEE (fiches habiter, études, réutilisations dans d'autres thématiques).

Intégration à ArcGIS Online

- nécessite un **formatage précis** des données pour créer applications, dashboards et story-maps.
- script **automatise** le processus, et permet à chaque mise à jour de créer un tableau de données dans le même format.



data_demo2 (Entités : 134, sélectionnées : 0)

codgeo	libgeo	pn0_pop	pn5_pop	pn10_pop	sntn5n0	sntn10n5	smgn5n0	smgn10
45006	Ardon	1092	1099	1139	1,31363318787914	0,23154629413709	-1,44134752558961	-0,94395
45008	Artenay	1951	1827	1772	0,07462890904839	-0,06689464697173	1,24736890838031	0,68009
45009	Aschères-le-Marché	1136	1166	1166	0,24264746908435	0,42881646655231	-0,76260633140797	-0,42881
45019	Baccon	654	715	681	0,60855023068312	0,97916751080255	-2,37624375790554	0
45024	Baule	2065	2044	2028	0,350811744397991	0,82580038670355	-0,14617156016579	-0,66850
41173	Beauce la Romaine	3488	3449	3194	-0,04618184498622	0,29140065463627	0,27131833929441	1,25666
45028	Beaugency	7339	7519	7715	-0,41360874377624	0,39810395104924	-0,06983004765055	-0,91144
41017	Binas	659	717	717	0,28843020486160	0,44630404463040	-1,96132539305894	-0,44630
45034	Boigny-sur-Bionne	2094	2183	2128	0,06520411822845	0,13023935940605	-0,89422790713304	0,38141
45039	Bonnée	710	673	676	-0,11633967635886	0,08891537479771	1,19248168267855	-0,17783
45043	Bou	1002	904	900	0,44567383236943	0,04436564214505	1,63413738535461	0,04436



7. – Retour d'expérience sur 5 ans

🍷 Un **centre de données et une chaîne de traitements** qui offrent :

- une **facilité de recherche des données** pour les chargés d'étude (manque encore un catalogue de données),
- des **garanties** sur la fiabilité, la pérennité et la mise à jour des données et des outils,
- un **gain de temps** (et donc humain et financier) pour tous grâce à l'automatisation de l'intégration des données, de la création des outils et de leur mise à jour,
- une **agilité** pour répondre à des demandes de partenaires.

MERCI POUR VOTRE ATTENTION !

