

Récupérer automatiquement des coordonnées en R à partir de nom de lieux

Boîte à outils

- projet personnel à partir de besoins dans le cadre de mon post-doctorat
- thématiques de l'atelier concernées : 2 - synthèses archéologiques urbaines et territoriales ; 4 - exploration multidimensionnelle de sources croisées pour l'archéologie ;

compte rendu de présentation

Gravier Julie

Que ce soit pour une présentation dans un colloque, la réalisation d'un cours ou plus largement dans un projet de reprise de données existantes, nous avons souvent le souhait de cartographier la répartition des objets archéologiques – au sens large – mis au jour et analysés. Toutefois, il est assez fréquent que la documentation à partir de laquelle nous puissions créer ces cartes soit présentée dans le cadre de publications sous forme de tableaux ou de listes énumératives, sans précision des latitudes et longitudes des objets étudiés. Leurs géolocalisations est généralement faite manuellement, point par point, lieu par lieu, dans un SIG. Le problème principal de cette procédure tient au fait que sa probabilité de réalisation est quasi inversement proportionnelle au volume de données à traiter et au temps que l'on détient...

On présente dans cette boîte à outils comment récupérer des coordonnées de manière automatique afin de réduire le temps et l'ingratitude de ce travail. Plus précisément, nous expliquons ci-dessous à travers un code commenté comment aboutir à une carte en environ 2/3 minutes de temps de calcul – ici, pour 88 coordonnées (x;y). L'automatisation du géocodage s'appuie sur l'API Photon, qui se fonde sur les données OpenStreetMap (<http://photon.komoot.de/>), et le package éponyme « photon » d'interfaçage de l'API avec R (<http://github.com/rCarto/photon>). Cette automatisation n'est toutefois pas un outil magique et la précision des informations sémantiques en entrée (adresse, nom de lieu, contrainte de recherche dans l'espace, etc.) a des implications directes sur la précision des coordonnées en sortie.

Importer et adapter les données

À partir du bilan sur la Protohistoire récente en Picardie publié par François Malrain, Stéphane Godefroy et Frédéric Gransar et du tableau recensant les études carpologiques réalisées (fig. 1), nous souhaitons récupérer les coordonnées des lieux d'où proviennent les échantillonnages carpologiques. Plus précisément, le tableau de la publication recense les auteur.e.s ayant réalisé l'étude, le département d'appartenance du site archéologique, la commune, le nom du site, le.s responsable.s de l'opération archéologique, la datation associée à l'échantillon carpologique. Hélas, compte tenu de la manière dont a été encodé le .pdf, il est relativement peu aisé de l'exploiter directement dans R. Il est de ce fait plus rapide de copier-coller préalablement le tableau dans un .txt (ici intitulé *RAP.txt*).

Commune	numéro de site	Commune	Recherche	Fonction	Intervention
Loeuilly	site 207	80485	3	habitat	linéaire
Long	site 208	80486	4	indéterminé	linéaire
Marsuil-Caubert	site 211	80512	1	habitat	autre
Marcelcave	site 209	80507	4	habitat	linéaire
Marcelcave	site 210	80507	4	funéraire	linéaire
Martainneville	site 212	80518	3	habitat	linéaire
Martainneville	site 213	80518	3	funéraire	linéaire
Méricourt-sur-Somme	site 214	80532	1	habitat	autre
Morvillers-Saint-Saturnin	site 215	80573	3	funéraire	autre
Noyelles-sur-Mer	site 216	80600	4	funéraire	linéaire
Plachy-Buyon	site 217	80627	3	habitat	linéaire
Pont-de-Metz	site 218	80632	4	habitat	autre
Pont-de-Metz	site 219	80632	3	habitat	linéaire
Pont-Rémy	site 220	80635	4	habitat	linéaire
Pont-Rémy	site 221	80635	4	funéraire	linéaire
Quend	site 222	80649	4	habitat	carrière
Revelles	site 223	80670	4	habitat	linéaire
Ribemont-sur-Ancre	site 224	80672	2	funéraire	autre
Roye	site 225	80685	4	habitat	autre
Rue	site 226	80688	4	habitat	carrière
Rue	site 227	80688	4	funéraire	carrière
Sailly-Flibeaucourt	site 228	80692	4	habitat	linéaire
Sailly-Flibeaucourt	site 229	80692	4	funéraire	linéaire
Sains-en-Amiénois	site 230	80696	4	funéraire	autre
Saint-Sauveur	site 231	80718	4	habitat	linéaire
Saint-Sauveur	site 232	80718	4	habitat	linéaire
Saint-Sauveur	site 233	80718	4	funéraire	linéaire
Saleux	site 234	80724	4	habitat	linéaire
Saleux	site 235	80724	4	habitat	linéaire
Saleux	site 236	80724	4	funéraire	linéaire
Soyescourt	site 237	80741	4	habitat	linéaire
Soyescourt	site 238	80741	4	habitat	linéaire
Soyescourt	site 239	80741	4	habitat	linéaire
Soyescourt	site 240	80741	4	funéraire	linéaire
Toutencourt	site 241	80766	4	funéraire	linéaire
Toutencourt	site 242	80766	4	funéraire	linéaire
Vignacourt	site 243	80793	4	habitat	linéaire
Vignacourt	site 244	80793	4	funéraire	linéaire
Ville -Le -Marclat	site 245	80795	4	habitat	linéaire
Ville-le-Marclat	site 246	80795	4	habitat	linéaire
Villers-Bretorneux	site 247	80799	3	funéraire	autre
Villers-les-Royes	site 248	80803	4	funéraire	autre
Vismes-au-Val	site 249	80809	3	habitat	linéaire
Vismes-au-Val	site 250	80809	3	funéraire	linéaire

DOCUMENT ANNEXE II

LISTE DES ÉTUDES CARPOLOGIQUES RÉALISÉES EN PICARDIE

Étude par	Dépt.	Commune	Site	Responsable	Datation
Matterne	Aisne	Bazoches-sur-Vesle	La Foulerie	S. Dessenne	La Tène D
Matterne	Aisne	Bazoches-sur-Vesle	Les Chantraines	Cl. Pommepuy	La Tène
Matterne	Aisne	Beaurieux	Les Grèves	C. Hazelgrove	toutes périodes
Bakels, 1995	Aisne	Berry-au-Bac	Le Chemin de la pêche		Néolithique
Matterne	Aisne	Bohain	Le Moulin Mayeux	Ph. Lemaire	I ^{er} siècle
Matterne	Aisne	Bucy-Le-Long	Le Grand Marais	Cl. Pommepuy	Hallstatt / La Tène A
Matterne	Aisne	Bucy-Le-Long	Le Grand Marais	Br. Robert	Hallstatt / La Tène A
Hopf, 1969	Aisne	Chassey		R. Rowlet	La Tène A
Matterne	Aisne	Ciry-Salsogne	Le Bruy	R. Cottiaux	La Tène C
Matterne	Aisne	Coussancourt		C. Hosdez	Gallo-romain
Pals, inédit	Aisne	Juvincourt-et-Damary	Le Gué de Mauchamp	D. Bayard	Hallstatt / La Tène
Matterne	Aisne	Limé	La Prairie	L. Hachem	La Tène A
Matterne	Aisne	Limé	Les Fussis	B. Hénon	I ^{er} -III ^e siècles

Étude par	Dépt.	Commune	Site	Responsable	Datation
Bakels, 1984	Aisne	Menneville	Derrière le Village	Br. Robert	La Tène A
Matterne	Aisne	Ploisy		Fr. Gransar	âges du Fer et Gallo-romain précoce
Matterne	Aisne	Saint-Quentin	Bois de Cambronne	Ph. Lemaire	I ^{er} siècle
Matterne	Aisne	Sermoise	Prés du Bout de la Ville	Fr. Gransar	La Tène
Matterne	Aisne	Vermand	Champ des Lavoirs	Ph. Lemaire	La Tène
Auxiette, 2003	Aisne	Villeneuve-Saint-Germain	Les Étommelles	G. Auxiette	Hallstatt
Marinval, 1991	Oise	Beauvais	Hôtel de Ville	J.-M. Fémolant	Moyen Âge
Matterne	Oise	Beauvais	Palais des Sports	J.-M. Fémolant	Haut-Empire
Heim, 1973	Oise	Béthisy-Saint-Martin	Le Barillet	M. Jouve	La Tène
Marinval, inédit	Oise	Chambly	La Marnière	J.-M. Fémolant	La Tène D
Bakels, 1984, 1986	Oise	Orouy	Les Ouis	G.-P. Woimant	IV ^e siècle
Matterne 1997	Oise	Compiègne	Le Confluent	Fr. Prodeo	Néolithique
Bakels, 1984	Oise	Compiègne	Le Fond Pernant	B. Lambot	La Tène A
Matterne 1997	Oise	Compiègne	Place des Hallettes	M. Petitjean	Moyen Âge
Quérel, 2002	Oise	Estrées-Saint-Denis	Les Sablons	P. Quérel	I ^{er} -II ^e siècles
Malrain et al. 1996	Oise	Jaux	Le Camp du Roy	Fr. Malrain	La Tène D
Matterne	Oise	Lacroix-Saint-Ouen	Les Longues Rayes	M. Talon	La Tène A
Matterne	Oise	Le Plessis-Brion	Le Fond	Fr. Malrain	Haut-Empire
Marinval, 1992, de Hingh, 1992/93 & Matterne, inédit	Oise	Longueil-Sainte-Marie	La Queue de Rivecourt et Le Bois Harlé	D. Maréchal	II ^e -III ^e siècles
Matterne	Oise	Longueil-Sainte-Marie	Le Vivier des Grès	Fr. Malrain	La Tène C
Matterne	Oise	Longpoint	Les Prés Véry	Fr. Malrain & E. Pinard	La Tène A
Matterne	Oise	Pont-Sainte-Maxence	Le Jonquoire	Fr. Malrain	Augustéen, I ^{er} siècle
Reynolds dans Talon, 1982	Oise	Rouvillers	Terre d'Élogettes	M. Talon	IV ^e siècle
Matterne	Oise	Précý-sur-Oise	2 rue du Martray	M. Derbois	Moyen Âge
Matterne	Oise	Verberie	Les Gâts	E. Pinard	La Tène A
Matterne	Somme	Ablaincourt-Pressoir	Le Chemin Blanc de Bovent	F. Lemaire	La Tène D2
Matterne	Somme	Amiens	Boulevard de Belfort	E. Binet	I ^{er} siècle
Matterne, 1998	Somme	Amiens	ZAC Cathédrale	D. Gernéhl	II ^e siècle
Bakels	Somme	Athies	Le Chemin de Croix	V. Harnay	Moyen Âge
Matterne et al.	Somme	Atilly	La Pâture des Grands Valuzeaux	A. Dubois	Gallo-romain précoce et Haut-Empire
Pals, inédit	Somme	Behen	Les Grands Riots	D. Bayard	Gallo-romain
Matterne	Somme	Bettencourt-Saint-Ouen	La Jocour	Ph. Lemaire	Haut-Empire
Pals, inédit	Somme	Boullancourt-en-Séry		O. Gonnet	Gallo-romain
Matterne	Somme	Cagny	Ferme de l'Épinette	P. Barbet	IV ^e siècle
Matterne	Somme	Cizancourt-Licourt	La Sole des Galets	Ph. Lefèvre	La Tène A
Matterne	Somme	Croixvaux	L'Aérodrome	St. Gaudéfroy	La Tène C2-D2
Matterne	Somme	Croixvaux	La Dériole		Gallo-romain
Bakels, 1999	Somme	Dury	Le Moulin	V. Harnay	Moyen Âge
Quéré, 2000	Somme	Dury	Le Camp Rolland	P. Quérel	III ^e siècle
Matterne	Somme	Eaucourt	Les Monts Bergeron	R. Haurillon	II ^e -III ^e siècles
Matterne	Somme	Ennemain	L'Orme	H. Petitot	I ^{er} et III ^e -IV ^e siècles
Matterne	Somme	Ennemain	Notre Dame de Joie	L. Wosny	Bronze final-Hallstatt
Matterne	Somme	Estrées-Deniécourt	Derrière le Jardin du Berger	D. Bayard / G. Prilaux	I ^{er} -IV ^e siècles
Matterne	Somme	Etreillers	La Grande Pâtur	N. Soupard	Gallo-romain
Matterne	Somme	Forest-Monthiers	Le Fond de Bernay	P. Quérel	La Tène D
Matterne	Somme	Framerville-Rainécourt	Le Fond d'Herleville	R. Rougier	Bronze final-Hallstatt
Matterne	Somme	Fresnes-Mazancourt	La Sole du Bois d'en Haut	R. Rougier	Hallstatt
Matterne	Somme	Fresnes-Mazancourt	La Sole du Moulin	R. Rougier	La Tène C-D
Matterne	Somme	Fricamps	Les Epassés	N. Buchez	La Tène C2-D2
Matterne	Somme	Gauville	Le Bout de la Rue du Bois	N. Soupard	La Tène D1
Matterne	Somme	Glizy	Les Terres de Ville	St. Gaudéfroy	La Tène C1
Matterne	Somme	Ham	Le Bois aux Cailloux	Ph. Barbet	La Tène B
Matterne	Somme	Harbonnières	Le Fond de Warcourt	Ph. Lefèvre	La Tène A-D
Matterne	Somme	Herleville	La Fosse Mangeart	G. Prilaux	La Tène A
Matterne	Somme	Hornoy-le-Bourg	Ancien Grand Bois		Gallo-romain
Pals, inédit	Somme	Le Translay	Le Chemin de Morival	Ph. Barbet-D. Bayard	La Tène finale
Matterne	Somme	Marcelcave	Le Chemin d'Ignaucourt	L. Blondiaux	I ^{er} -II ^e siècles
Matterne	Somme	Monchy-Lagache	Le Fond de Guizancourt	L. Blondiaux	I ^{er} siècle
Matterne	Somme	Monchy-Lagache	La Mare de Flez	N. Soupard	Gallo-romain et Haut Moyen Âge

figure 1 : extrait du tableau étudié (Malrain, Godefroy, Gransar 2005 : 165-166)

Packages nécessaires

```
library(tidyverse)
library(photom) # devtools::install_github(repo = 'rCarto/photom') : package
développé pour faire le lien entre R et l'API Photon
library(sf) # package spatial
library(tmap) # package de visualisation spatiale dynamique
```

```
rap_picardie <- read_table("data/RAP.txt") # Lire le tableau de données
head(rap_picardie)
```

```
## # A tibble: 6 x 1
##   `Étude par Dépt. Commune Site Responsable Datation`
##   <chr>
## 1 Matteredne Aisne Bazoches-sur-Vesle La Foulerie S. Dessenne La Tène D
## 2 Matteredne Aisne Bazoches-sur-Vesle Les Chantraines Cl. Pommeputy La Tène
## 3 Matteredne Aisne Beaurieux Les Grèves C. Haselgrove toutes périodes
## 4 Bakels Aisne Berry-au-Bac Le Chemin de la pêcheurie Néolithique
## 5 Matteredne Aisne Bohain Le Moulin Mayeux Ph. Lemaire Ier siècle
## 6 Matteredne Aisne Bucy-Le-Long Le Grand Marais Cl. Pommeputy Hallstatt
```

Le tableau contient une seule colonne de type *character*. On va devoir séparer les éléments en différentes colonnes sachant que l'on souhaite récupérer les informations sur la commune (et le département) dans laquelle a eu lieu l'opération archéologique.

```
rap_picardie_df <- rap_picardie %>%
  rename(variable = `Étude par Dépt. Commune Site Responsable Datation`) %>%
  separate(col = variable, c("Etude", "Departement", "Commune"), sep = " ", remove =
FALSE) %>% # séparation des premiers éléments en 3 colonnes afin de récupérer le
nom des communes et des départements
  mutate(hdf = "Hauts-de-France, France", # création d'une colonne avec des
informations spatiales complémentaires
         lieux = str_c(Commune, Departement, hdf, sep = ", ")) # concaténation des
colonnes liées aux noms des communes, dép., région et pays
```

Deux choses sont importantes à noter dans le code ci-dessus :

1. la fonction *rename()* n'est pas nécessaire en soi. Elle permet seulement d'avoir un nom de variable court dans la 2e ligne de code pour que ce dernier soit plus lisible lorsque l'on travaille sur petit écran ;
2. nous avons créé une colonne **lieux** contenant le nom de la commune, le département, la région actuelle (Hauts-de-France) et le pays (France). C'est cette colonne qui nous servira à faire le géocodage. L'ajout d'information (région & pays) et la concaténation avec celles déjà existantes (commune & département) sert en tant que contrainte spatiale dans la recherche des coordonnées lors de l'utilisation de l'API Photon. En effet, par exemple, plusieurs communes ont le même nom en France. Sans contrainte de recherche dans la région Hauts-de-France, il est possible que le résultat de la recherche automatique soit relatif à une commune située en dehors de la région.

Géocodage et vérification de sa validité

```
rap_geoloca <- geocode(location = rap_picardie_df$lieux, key = "place", limit = 1) #  
récupération des coordonnées avec photon  
rap_geoloca_sf <- st_as_sf(x = rap_geoloca, coords = c("lon", "lat")) %>% #  
construction d'un objet spatial de type sf  
  st_set_crs(crs = 4326) %>% # Les coordonnées récupérées sont en Long/Lat, WGS 84  
  st_transform(crs = 2154) # transformation du système de coordonnées en Lambert 93
```

Une fois que l'on a récupéré les coordonnées et créé un objet de type *sf*, soit un objet spatial dans R (Pebesma 2018), on va pouvoir vérifier la validité des coordonnées de manière dynamique avec le package *tmap*

```
tmap_mode("view") # actionner le mode dynamique  
tm_shape(rap_geoloca_sf) + # le "shape" à visualiser (objet sf)  
  tm_bubbles(size = 0.4, col = "darkorange") # type de visualisation
```

Si l'on est satisfait de la validité des points récupérés automatiquement, on peut exporter l'objet *sf* en un *.shp*

```
st_write(rap_geoloca_sf, "data/geoloca_etudes_carpo_RAP25.shp")  
## Writing layer `geoloca_etudes_carpo_RAP25' to data source  
`data/geoloca_etudes_carpo_RAP25.shp' using driver `ESRI Shapefile'  
## features:      88  
## fields:       13  
## geometry type: Point
```

Cartographie

Enfin, on peut cartographier les résultats pour notre future conférence, cours, publication, etc. (fig. 2).

Imports des fonds de carte

```
sf_fond_europe <- read_sf("data/geom_eu34_nuts13.shp") %>% # fond Europe  
  st_transform(crs = 2154) # transformation en Lambert 93  
france <- read_sf("data/REGION.shp") %>% # fond des régions françaises actuelles  
  st_transform(crs = 2154) # transformation en Lambert 93  
  
## zoom sur la zone d'étude (la région Hauts-de-France)  
bbox_haut_de_france <- st_bbox(france %>% filter(INSEE_REG == "32")) # 32 étant le  
code des Hauts-de-France
```

Carte

```
ggplot() +  
  geom_sf(data = sf_fond_europe, fill = "grey98", color = "grey98") +  
  geom_sf(data = france, fill = "grey98", color = "grey40") +  
  geom_sf(data = rap_geoloca_sf, color = "grey30", fill = "#e66101", shape = 21,  
size = 2) +  
  coord_sf(xlim = bbox_haut_de_france[c(1,3)], ylim = bbox_haut_de_france[c(2,4)]) +  
  ggthemes::theme_igray() +  
  labs(subtitle = "Les études carpologiques réalisées en Picardie en 2005",  
caption = "J. Gravier 2020 | LabEx DynamiTe, UMR Géographie-cités.\nSources :  
Malrain, Gaudefroy, Gransar 2005") +  
  theme(axis.text.x = element_blank(),  
axis.text.y = element_blank(),  
axis.ticks = element_blank()) +  
  ggspatial::annotation_scale(location = "tr", width_hint = 0.5)
```

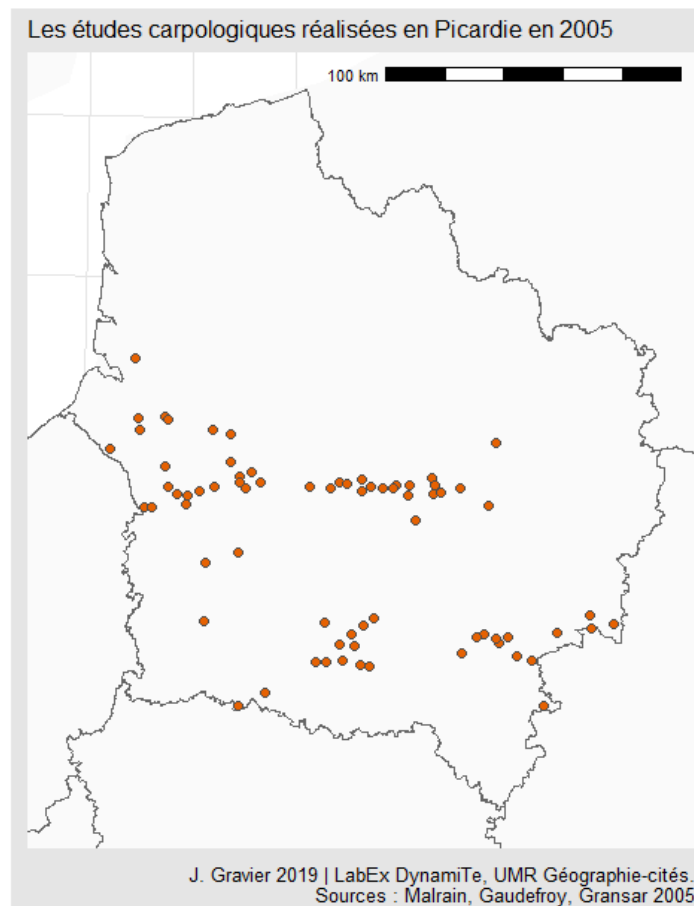


figure 2 : coordonnées récupérées à partir de l'API Photon

références citées

Malrain, F., Godefroy, S., Gransar, F., « La Protohistoire récente : IIIe siècle - Ière moitié du premier siècle avant notre ère », *Revue archéologique de Picardie*, La recherche archéologique en Picardie : Bilans et Perspectives, 2005, 3/4, pp. 127-167

Pebesma, E., *sf: Simple Features for R. R package version 0.7-3.*, 2018