

Analyse spatiale et territoriale de données de recensement

Formation Carthageo-Geoprisme 2021 / 1ere journée

C.GRASLAND

15/10/2020

Section 1

2. Données spatiales (sf)

Le format sf (spatial features)

La cartographie et plus généralement les opérations géométriques sur des données spatiales dans R peuvent facilement être effectuées avec le **package sf** (spatial features) qui crée des objets uniques rassemblant à la fois

- un tableau de données (l'équivalent du fichier .dbf)
- une géométrie (l'équivalent du fichier .shp)
- une projection (l'équivalent du fichier .prj)

Lorsqu'on récupère des fonds de carte au format shapefile (.shp) ou dans d'autres formats standards comme GeoJson, la première tâche consiste donc à les convertir au format sf afin de pouvoir les utiliser facilement dans R. L'importation se fait à l'aide de l'instruction `st_read` en indiquant juste le nom du fichier .shp à charger. Les autres fichiers (.dbf ou .proj) seront lus également et intégrés dans l'objet qui hérite de la double classe *data.frame* et *sf*.

Etapes de préparation des données

Dans notre exemple, nous allons suivre les étapes suivantes :

- ➊ Préparer les données statistiques par IRIS dans un *data.frame*
- ➋ Charger un fonds de carte par IRIS au format *sf*
- ➌ Effectuer une jointure entre les deux fichiers par le code IRIS
- ➍ Sauvegarder le résultat
- ➎ Agréger les données statistiques et géométriques par commune
- ➏ Sauvegarder le résultat.

Préparer les données statistiques

On importe le fichier des individus :

programme

```
tab_ind<-readRDS("data2021/94/indiv2017.RDS")
```

résultat

FALSE	CANTVILLE	NUMMI	ACHLR	AEMMR	AGED
FALSE 1	9401	1	3	9	47
FALSE 2	9401	1	3	9	42
FALSE 3	9401	1	3	9	6

Agréger les données

On commence par créer un *tableau long* croisant les deux variables et leur effectif pondéré :

programme

```
tab_long <- tab_ind %>%  
  filter(LPRM == 1) %>%  
  group_by(IRIS, HLML) %>%  
  summarise(NB = sum(IPONDI))
```

résultat

IRIS	HLML	NB
940020101	1	470.39
940020101	2	1054.91
940020102	1	247.79
940020102	2	1651.54
040020102	1	410.87

Pivoter le tableau

Puis on fait “pivoter” le tableau pour l’obtenir en format large :

```
tab_large <- tab_long %>% pivot_wider(id_cols = IRIS,  
                                       names_from = HLML,  
                                       names_prefix = "HLM",  
                                       values_from = NB)
```

résultat

IRIS	HLM1	HLM2
940020101	470.39	1054.91
940020102	247.79	1651.54
940020103	419.87	1030.03
940020104	576.24	967.49
940020105	648.40	1011.51

Ajouter de nouvelles variables

On ajoute de nouvelles variables telles que le nombre total de ménage et le % de ménages en HLM :

```
tab<- tab_large %>% mutate(TOT = HLM1+HLM2,  
                             HLMpct = 100*HLM1/TOT)
```

résultat

IRIS	HLM1	HLM2	TOT	HLMpct
940020101	470.39	1054.91	1525.30	30.84
940020102	247.79	1651.54	1899.34	13.05
940020103	419.87	1030.03	1449.90	28.96
940020104	576.24	967.49	1543.73	37.33
940020105	648.40	1011.51	1659.90	39.06

Examiner la distribution statistique

On examine l'histogramme donnant distribution statistique du % de ménages ordinaires résidant en HLM par IRIS.

programme

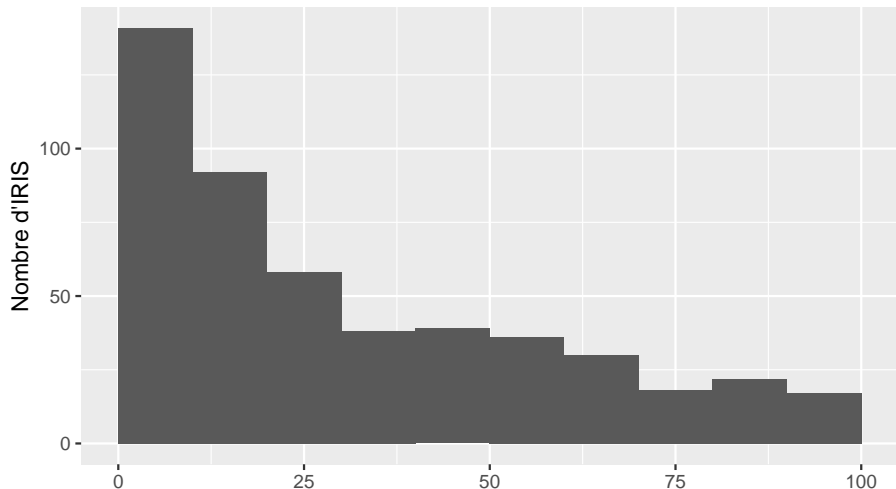
```
p <- ggplot(tab) + aes (x = HLMpct) +  
  geom_histogram(breaks = c(0,10,20,30,40,50,  
                             60,70,80,90, 100))  
  scale_x_continuous("% de ménages en HLM") +  
  scale_y_continuous("Nombre d'IRIS") +  
  ggtitle(label = "Distribution des logements",  
          subtitle = "Source : INSEE, RP 2017")
```

Examiner la distribution statistique

résultat

Distribution des logements sociaux dans le Val de Marne

Source : INSEE, RP 2017



Charger les données géométriques

On importe le fichier des iris du Val-de-Marne qui est au format sf en ne gardant que les colonnes utiles

programme

```
map_iris <- readRDS("data2021/94/map_iris.RDS")
map_iris <- map_iris[,c(4,5,1,2,7)]
names(map_iris) <- c("IRIS", "NOM_IRIS", "COM", "NOM_COM", "geometry")
```

résultat

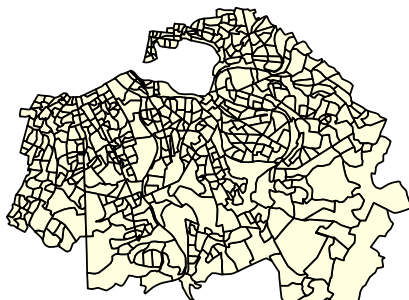
```
FALSE [1] "sf"           "data.frame"
```

IRIS	NOM_IRIS	COM	NOM_COM
940210107	Sorbiers	94021	Chevilly-Larue
940680204	Le Vieux Saint-Maur 4	94068	Saint-Maur-des-Fossés

Visualisation du fonds iris avec sf

On peut facilement produire une carte vierge des iris du Grand Paris en faisant un plot de la colonne *geometry* du fichier *sf*

```
plot(map_iris$geometry,col="lightyellow")
```



Jointure des données IRIS et du fonds de carte

programme

```
map_iris_tab<-merge(map_iris,tab,  
                    by.x="IRIS",by.y="IRIS",  
                    all.x=T,all.y=F)
```

résultat

IRIS	NOM_IRIS	COMNOM_COM	HLM1	HLM2	TOT	HLMp	geometry
940010001	Ablon-sur-Seine	94001Ablon-sur-Seine	NA	NA	NA	NA	MULTIPOLYGON (((658439.7 68...
940020101	Chinagora Berthelot	94002Afortville	470.39	1054.91	1525.30	80.84	MULTIPOLYGON (((656616.9 68...
940020102	Deny Garnier	94002Afortville	247.79	1651.51	1899.31	13.05	MULTIPOLYGON (((656988.2

Sauvegarde du fichier par IRIS

On sauvegarde notre fichier au format .RDS de R

```
saveRDS(map_iris_tab, "data2021/94/map_iris_hlm.RDS")
```

Grâce aux nouveaux packages de R (*dplyr* et *sf*) il est possible d'**agréger simultanément les statistiques et les géométries** après les avoir stockés dans un même objet de type "sf"

Du coup, on peut gagner beaucoup de temps dans les traitements et les analyses cartographiques, en particulier si l'on veut tester différents niveaux d'agrégation.

Agrégation des IRIS en communes

L'agrégation est très facile et elle concerne à la fois les variables (de stock) et les geometries

programme

```
map_com_tab <- map_iris_tab %>%  
  group_by(COM, NOM_COM) %>%  
  summarise(HLM1=sum(HLM1,na.rm=T),  
            HLM2=sum(HLM2,na.rm=T)) %>%  
  st_cast("MULTIPOLYGON")  
  
map_com_tab <- map_com_tab %>% mutate(TOT = HLM1+HLM2,  
                                       HLMpct = 100*HLM1/TOT)
```


Agrégation des iris en communes

résultat statistique

COM	NOM_COM	HLM1	HLM2	TOT	HLMpct
94001	Ablon-sur-Seine	0.00	0.00	0.00	NaN
94002	Alfortville	7593.21	12562.81	20156.02	37.67
94003	Arcueil	4505.06	5357.96	9863.02	45.68

Agrégation des iris en communes

résultat géométrique



Examiner la distribution statistique

On examine l'histogramme donnant distribution statistique du % de ménages ordinaires résidant en HLM par Commune.

programme

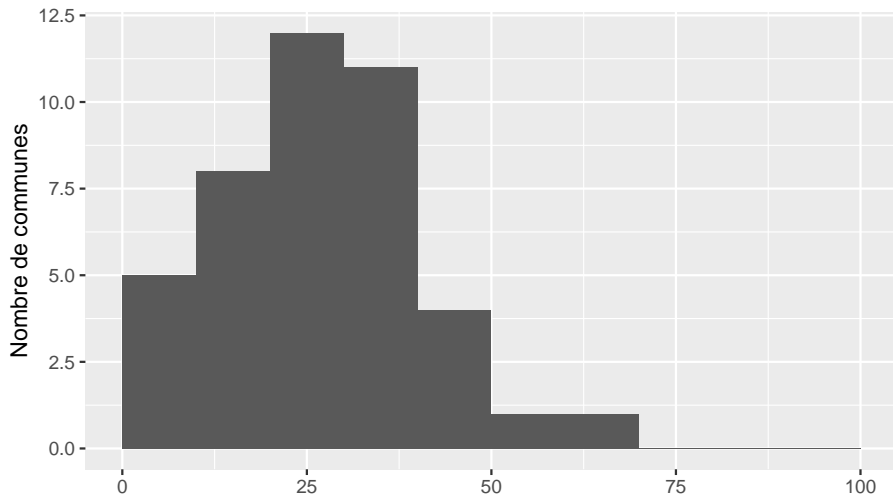
```
p <- ggplot(map_com_tab) + aes (x = HLMpct) +  
  geom_histogram(breaks = c(0,10,20,30,40,50,  
                             60,70,80,90,100)) +  
  scale_x_continuous("% de ménages en HLM") +  
  scale_y_continuous("Nombre de communes") +  
  ggtitle(label = "Distribution des logements",  
          subtitle = "Source : INSEE, RP 2017")
```

Examiner la distribution statistique

résultat

Distribution des logements sociaux dans le Val de Marne

Source : INSEE, RP 2017



Sauvegarde du fichier par commune

On sauvegarde notre fichier au format .RDS de R

```
saveRDS(map_com_tab, "data2021/94/map_com_hlm.RDS")
```