

# Analyse spatiale et territoriale du logement social

## Formation Carthageo-Geoprisme 2022

Claude Grasland, Université de Paris (Diderot)

## Section 1

### **Localisation territoriale**

# 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("data/menag2018.RDS")
```

## résultat

FALSE	COMMUNE	ARM	IRIS	ACHL	AEMM
FALSE 1	75056	75101	751010101	C114	2014
FALSE 2	75056	75101	751010101	A11	2008
FALSE 3	75056	75101	751010101	A11	2012

## Agréger les données

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

### programme

```
tab_long <- tab_ind %>%  
  filter(HLML != "Y") %>%  
  group_by(IRIS, HLML) %>%  
  summarise(NB = sum(IPONDL))
```

### résultat

IRIS	HLML	NB
751010101	1	179.08
751010101	2	329.51
751010102	1	3.16
751010102	2	96.72
751010102	1	10.07

# 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,  
                                     values_fill = 0)
```

## résultat

IRIS	HLM_1	HLM_2
751010101	179.08	329.51
751010102	3.16	96.72
751010103	10.97	129.56
751010201	148.72	1182.91
751010202	16.34	923.22

## 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 = HLM_1+HLM_2,  
                           HLM_pct = 100*HLM_1/TOT)
```

### résultat

IRIS	HLM_1	HLM_2	TOT	HLM_pct
751010101	179.08	329.51	508.59	35.21
751010102	3.16	96.72	99.89	3.17
751010103	10.97	129.56	140.53	7.81
751010201	148.72	1182.91	1331.63	11.17
751010202	16.34	923.22	939.56	1.74



# Examiner la distribution statistique

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

## programme

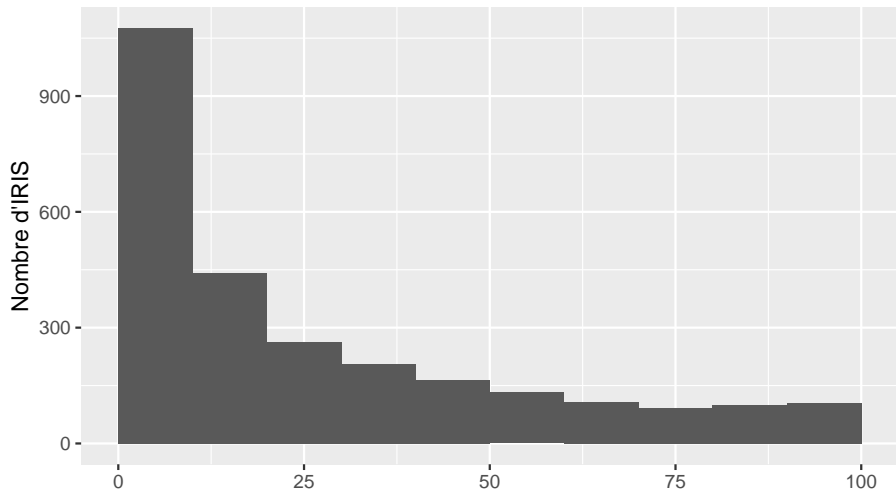
```
p <- ggplot(tab) + aes (x = HLM_pct) +  
  geom_histogram(breaks = c(0,10,20,30,40,50,  
                             60,70,80,90, 100))  
  scale_x_continuous("% de ménages en HLM") +  
  scale_y_continuous("Nombre d'IRIS") +  
  ggtitle(label = "Distribution des logements",  
          subtitle = "Source : INSEE, RP 2018")
```

# Examiner la distribution statistique

## résultat

Distribution des logements sociaux dans Paris & PC

Source : INSEE, RP 2018



# 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("data/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

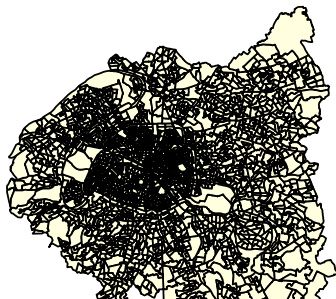
```
FALSE [1] "sf"           "data.frame"
```

IRIS	NOM_IRIS	COM	NOM_COM
751176511	Ternes 11	75117	Paris 17e Arrondissement
920240403	Klock	92024	Clichy

# 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	HLM_HLM_TOT	HLM_geo	geometry
751010101	Int-Germain l'Auxerrois 1	75101Paris 1er Ar-rondissement	179.0829.51	508.595.21	MULTIPOLYGON(((651771.668...
751010102	Int-Germain l'Auxerrois 2	75101Paris 1er Ar-rondissement	3.16	96.7299.893.17	MULTIPOLYGON(((651668.768...

# Sauvegarde du fichier par IRIS

On sauvegarde notre fichier au format .RDS de R

```
saveRDS(map_iris_tab, "data/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(HLM_1=sum(HLM_1,na.rm=T),  
            HLM_2=sum(HLM_2,na.rm=T)) %>%  
  st_cast("MULTIPOLYGON")  
  
map_com_tab <- map_com_tab %>% mutate(TOT = HLM_1+HLM_2,  
                                       HLM_pct = 100*HLM_1/TOT)
```



# Agrégation des iris en communes

## résultat statistique

COM	NOM_COM	HLM_1	HLM_2	TOT	HLM_P
75101	Paris 1er Arrondissement	952.20	8321.07	9273.27	10.
75102	Paris 2e Arrondissement	562.85	11696.77	12259.63	4.
75103	Paris 3e Arrondissement	1517.28	18267.52	19784.80	7.

# 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

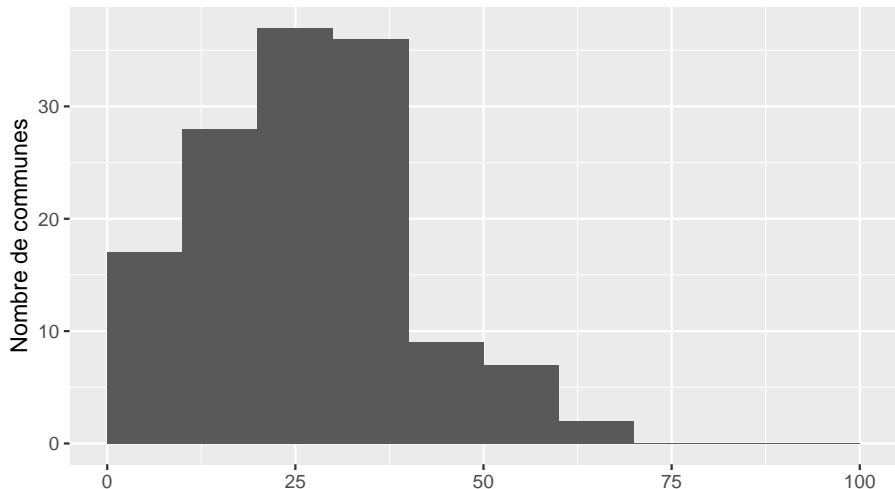
```
p <- ggplot(map_com_tab) + aes (x = HLM_pct) +  
  geom_histogram(breaks = c(0,10,20,30,40,50,  
                             60,70,80,90,100)) +  
  scale_x_continuous("% de ménages en HLM") +  
  scale_y_continuous("Nombre de communes") +  
  ggtitle(label = "Distribution des logements",  
          subtitle = "Source : INSEE, RP 2018")
```

# Examiner la distribution statistique

## résultat

### Distribution des logements sociaux dans Paris et PC

Source : INSEE, RP 2018



# Sauvegarde du fichier par commune

On sauvegarde notre fichier au format .RDS de R

```
saveRDS(map_com_tab, "data/map_com_hlm.RDS")
```