

# Evolution du prix de l'immobilier en France



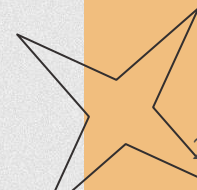
Tasnême-Jenna Louartani  
Maxime Bonnet  
Chaimaa Beladraoui



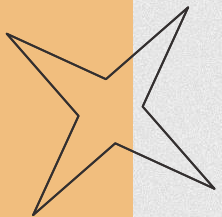


# SOMMAIRE

1. Introduction
2. Analyse des données
3. Etude des données
4. Etudes comparatives
5. Apprentissage
6. Conclusion

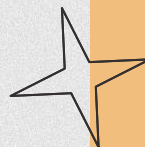






# 01.

## Introduction



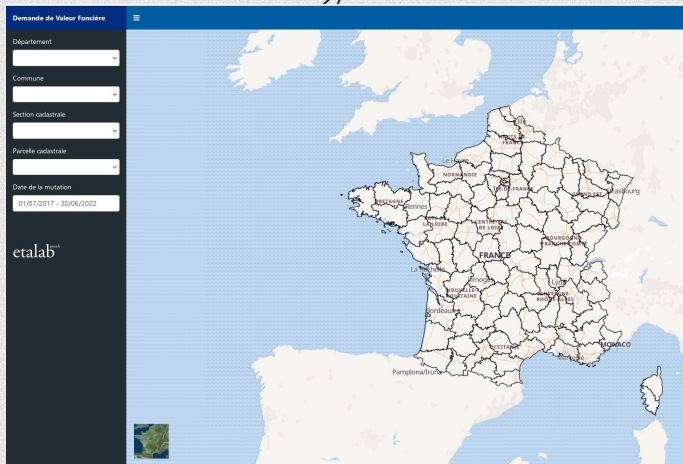


# Base de données DVF (années 2017-2022)

Informations sur :

- La **localisation** des biens et leurs **valeurs foncières**
- La mutation (la **Nature** et la **Date**)
- Le **type** de bien (Local, Maison, Dépendance...)

id_mutation	date_mutation	numero_disposition	nature_mutation	valeur_fonciere	adresse_numero	adresse_suffixe
2017-524	26/08/2017	1	Vente	1750		
2017-572	07/09/2017	1	Vente	144000		
2017-572	07/09/2017	1	Vente	144000		
2017-572	07/09/2017	1	Vente	144000	5079	
2017-572	07/09/2017	1	Vente	144000		
2017-663	14/09/2017	1	Vente	38106		
2017-663	14/09/2017	1	Vente	38106		
2017-663	14/09/2017	1	Vente	38106		
2017-666	15/09/2017	1	Vente	4000		
2017-1173	27/10/2017	1	Vente	211300	5248	
2017-1173	27/10/2017	1	Vente	211300	5248	
2017-1173	27/10/2017	1	Vente	211300		
2017-1291	31/10/2017	1	Vente	850		
2017-1348	20/11/2017	2	Vente	13105.4		
2017-1348	20/11/2017	2	Vente	13105.4		
2017-1348	20/11/2017	2	Vente	13105.4		
2017-1348	20/11/2017	2	Vente	13105.4		
2017-1348	20/11/2017	2	Vente	13105.4		



data.gouv.fr

Recherche

Données Réutilisations Organisations Commencer sur data.gouv.fr v Actualités Nous contacter

Accueil > Jeux de données > DVF

## Demandes de valeurs foncières DVF

105 favoris

Qualité des métadonnées : Mis à jour le 17 octobre 2022 — Licence Ouverte / Open Licence version 2.0

Producteur Métadonnées Actions

**Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique**

Le ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique est chargé de mettre en œuvre la politique du Gouvernement en matière économique, financière, budgétaire et fiscale. Ses missions recouvrent des champs d'action variés tels que l'industrie, le commerce, les...

451 jeux de données 2 réutilisations

**SOMMAIRE**

1. Description
2. Fichiers <sup>(70)</sup>
3. Ressources communautaires <sup>(4)</sup>
4. Réutilisations <sup>(67)</sup>
5. Discussions <sup>(86)</sup>

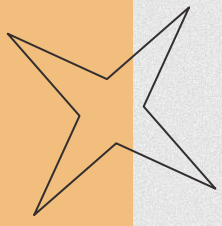




# Problématique

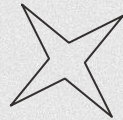
- Evolution de la valeur foncière en France
- Critères sociaux et économiques



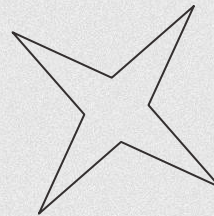


# 02.

## Analyse des données



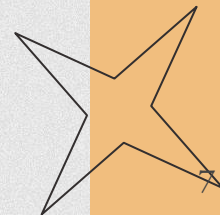


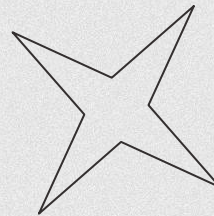


# Problèmes dans la qualité des données

- Champs non renseignés et valeurs NA sur les 40 au total
- L'Alsace et la Moselle ne sont pas soumises au remplissage de la base DVF

nombre_pieces_principales	code_nature_culture	nature_culture	code_nature_culture_speciale	nature_culture_speciale	surface_terrain	longitude	latitude
4	S	sols			391	5.204386	46.19368
0					NA	5.219932	46.19610
1					NA	5.219932	46.19610
NA	AB	terrains a bâtir			605	5.304674	46.04062
NA	AG	terrains d'agrément			259	NA	NA
2	S	sols			471	NA	NA
3	S	sols			165	5.225402	46.19304
6	S	sols			731	5.061935	46.39645
NA	T	terres			1577	4.951945	46.41788
NA	T	terres			929	4.963314	46.30787
5	S	sols			956	4.963243	46.30811
4	S	sols			580	4.910920	46.40844
3	S	sols			948	4.938636	46.42100
NA	P	prés			218	4.938478	46.42136
0	S	sols			1327	5.266108	46.14591
4	S	sols			1327	5.266108	46.14591
NA	P	prés			24175	5.202222	46.40641
NA	S	sols			2693	5.202779	46.40681
NA	S	sols			608	5.202696	46.40711

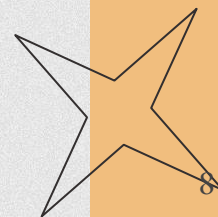




## Problèmes dans la qualité des données

- Champs non renseignés (NA)
- Valeurs aberrantes (Maisons à 1€, appartements à 270M d'euros ou encore surfaces terrain à 1m2)

id_mutation	date_mutation	numero_disposition	nature_mutation	valeur_fonciere
2017-298774	20/07/2017	1	Vente	1
2017-298774	20/07/2017	1	Vente	1
2017-299026	24/07/2017	1	Vente	1
2017-299026	24/07/2017	1	Vente	1
2017-299026	24/07/2017	1	Vente	1

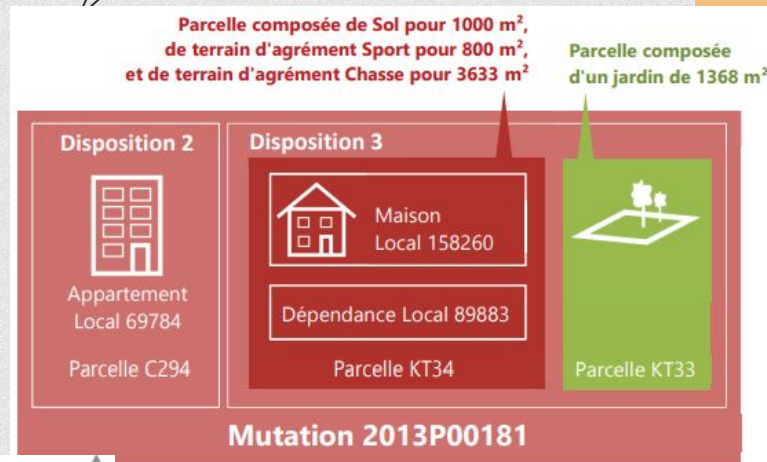
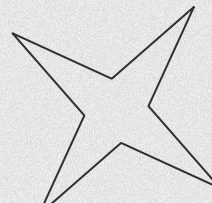




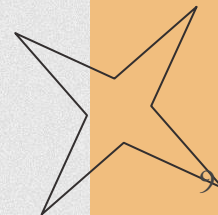


# Problèmes dans la qualité des données

- Champs non renseignés (NA)
- Valeurs aberrantes (Maisons à 1€)
- Doublons
- Erreurs sémantiques (mauvaise localisation des parcelles...)



DVF\_GuideAccompagnement

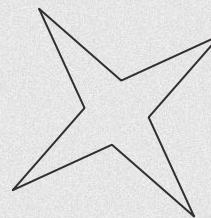
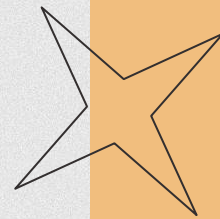


Reference document	No disposition	Valeur foncière	Section	No parcelle	Type local	Identifiant local	Nature culture	Nature culture spéciale	Surface terrain
2013P00181	2	180000	C	294	Appartement	69784			
2013P00181	3	317000	KT	33			Jardin		1368
2013P00181	3	317000	KT	34	Dépendance	89883	Sol		1000
2013P00181	3	317000	KT	34	Dépendance	89883	Agrément	Sport	800
2013P00181	3	317000	KT	34	Dépendance	89883	Agrément	Chasse	3633
2013P00181	3	317000	KT	34	Maison	158260	Sol		1000
2013P00181	3	317000	KT	34	Maison	158260	Agrément	Sport	800
2013P00181	3	317000	KT	34	Maison	158260	Agrément	Chasse	3633

# 03.

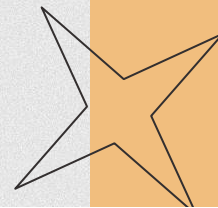


## Etude des données





# Visualisation des données

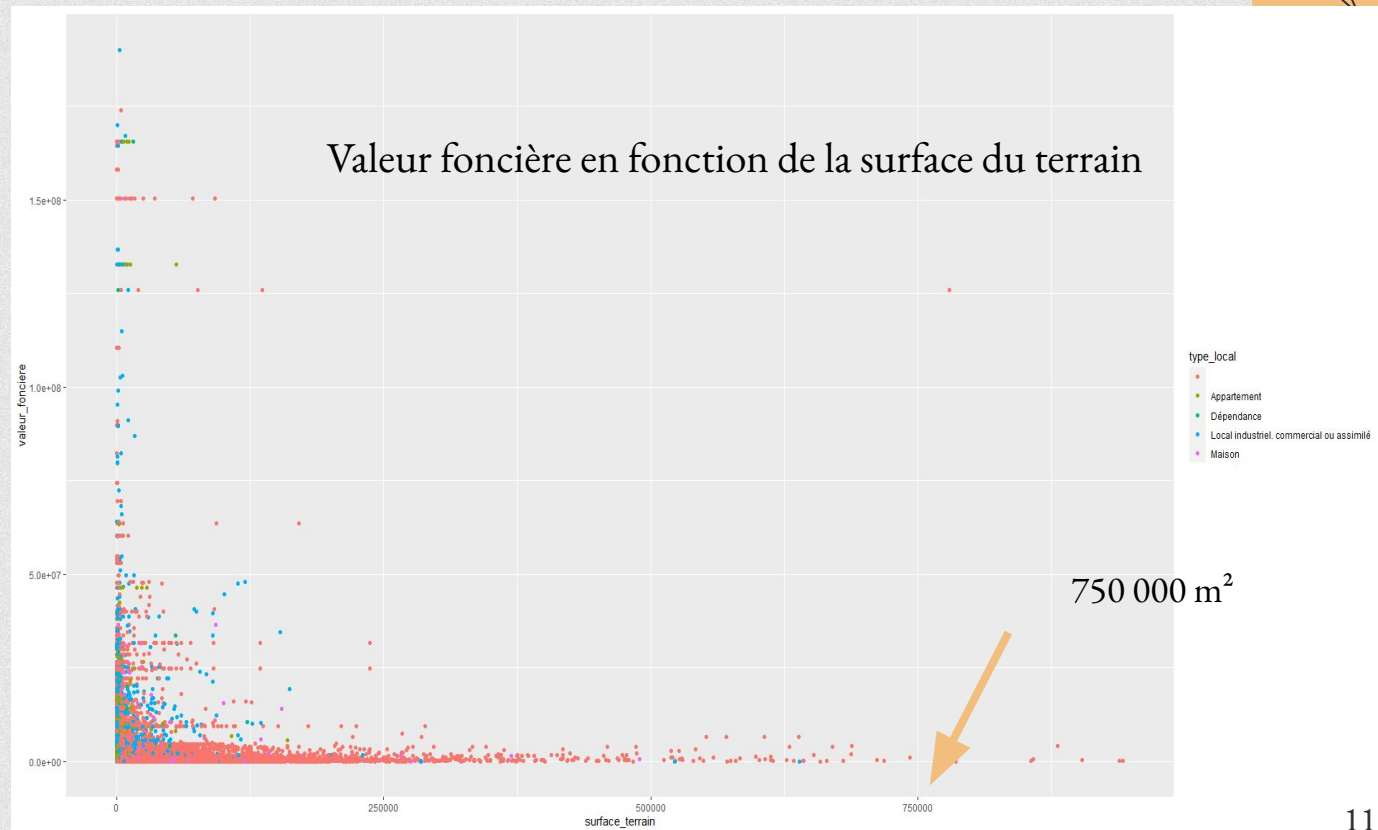


150 000 000  
millions d'euros

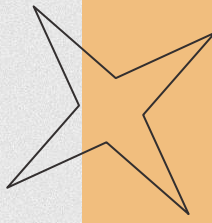


Affichage rapide pour voir  
que des données avec des  
valeurs extrêmement  
élevées tire le graphique  
vers la droite et le haut

Ce sont véritablement  
les seules valeurs  
quantitatives de la base  
DVF

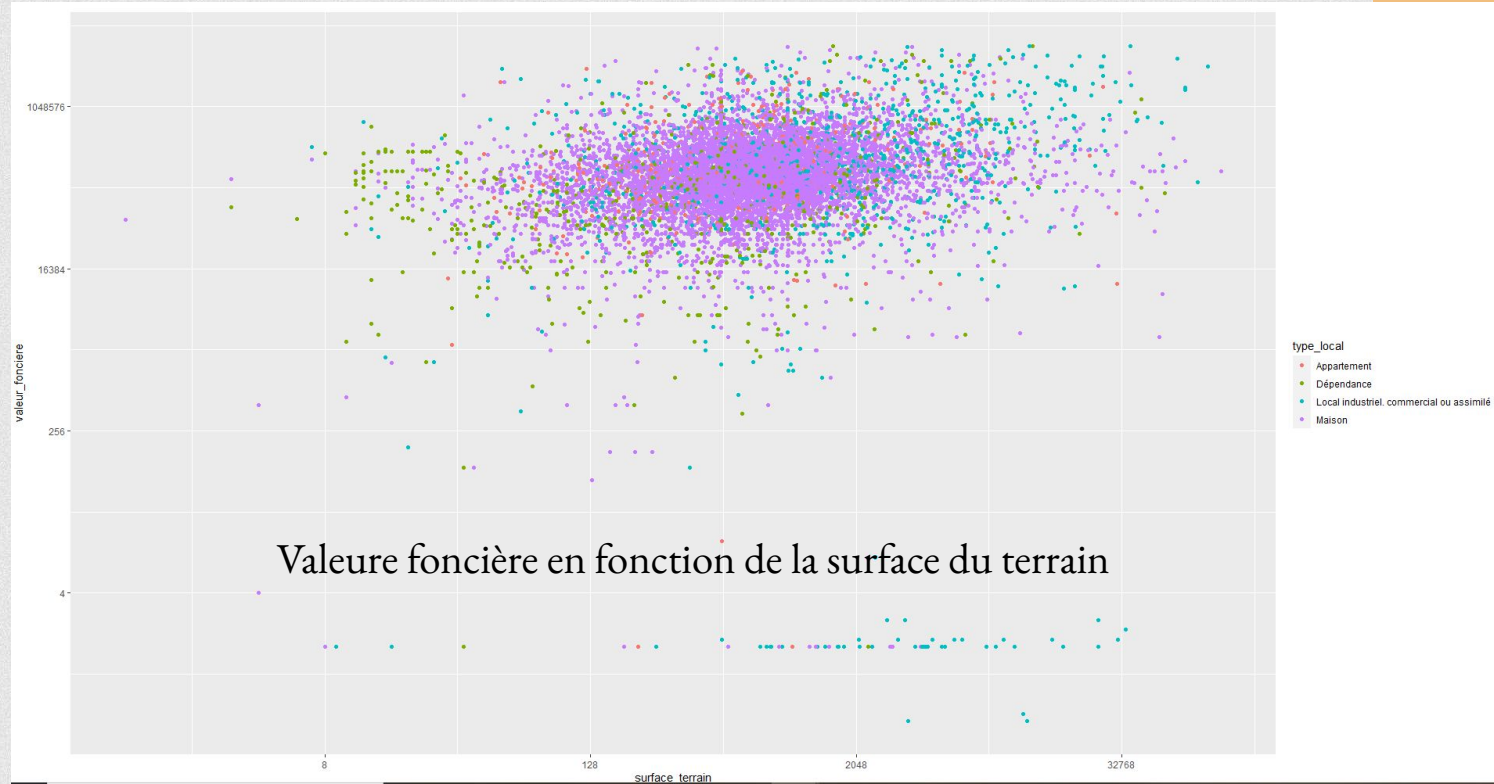


# En échelle logarithmique



Passage à une échelle  $\log_2$   
et en supprimant les biens  
où le type du local n'était  
pas renseigné.

Les données sont  
globalement  
regroupées dans la  
même zone





# Où sont les maisons vendues le plus de fois?

Une maison vendue 6 fois est identifiée par :

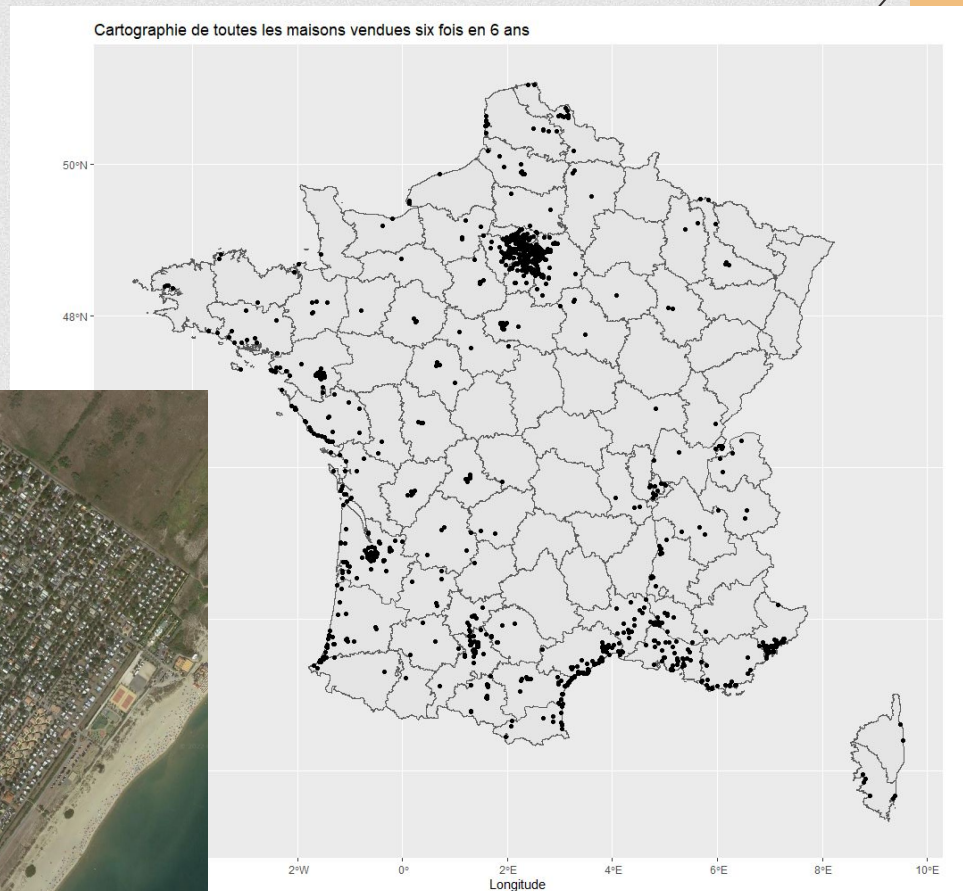
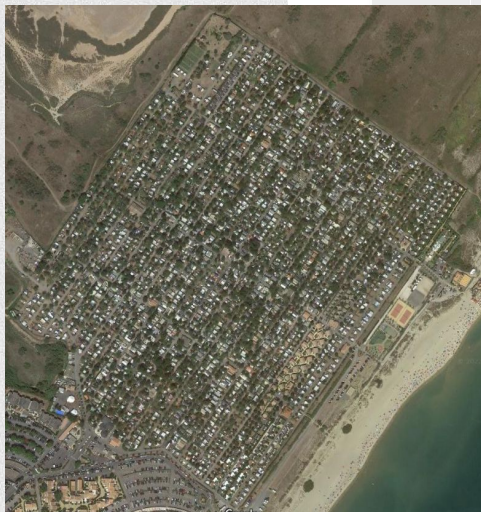
Le type de local -> Maison

La nature de la mutation -> Vente

Id\_parcelle dans chacune des années

- Traduit la présence de camping et de résidence
- Montre également les villes les plus peuplées

Une maison dans une parcelle unique a été vendue plus de 700 fois!

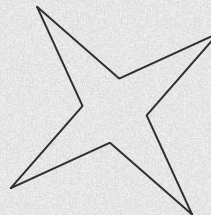
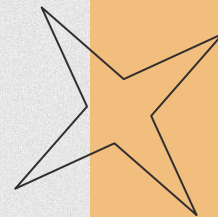
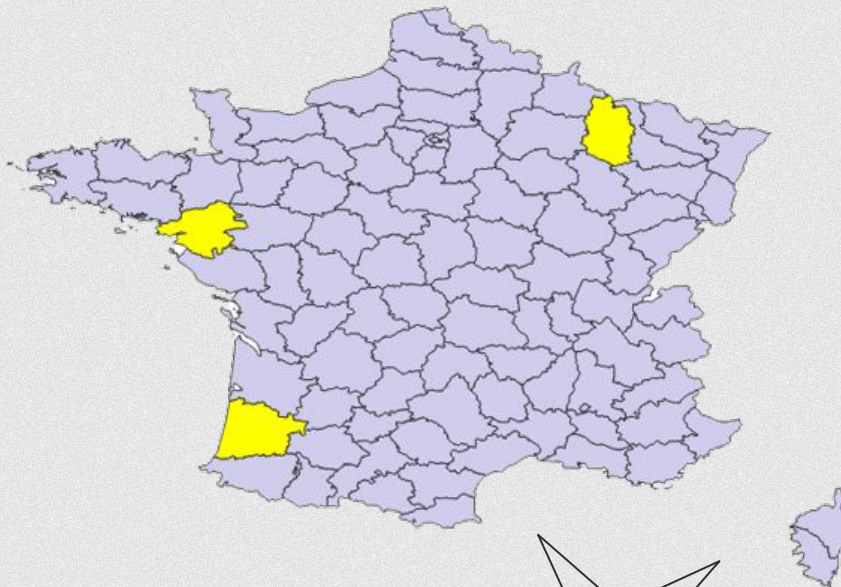




## Présentation de la zone d'étude

Notre étude a été menée dans trois départements différents :

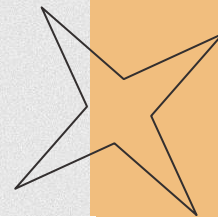
- Les Landes
- La Meuse
- La Loire Atlantique



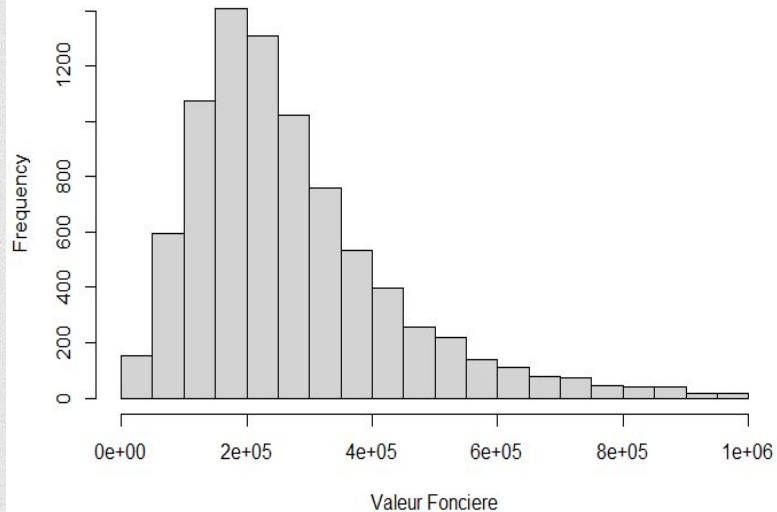




# Evolution de la valeur foncière

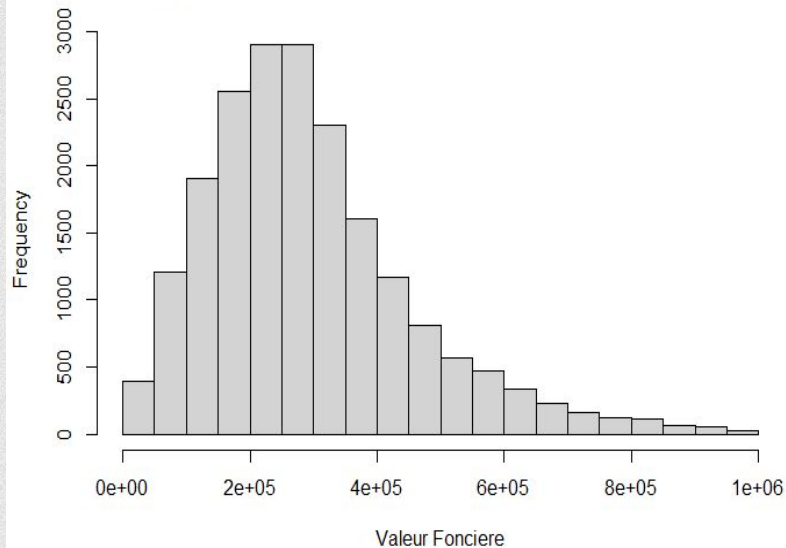


Histogramme de valeurs foncieres à Landes en 2021



Pop : 409 325 INSEE

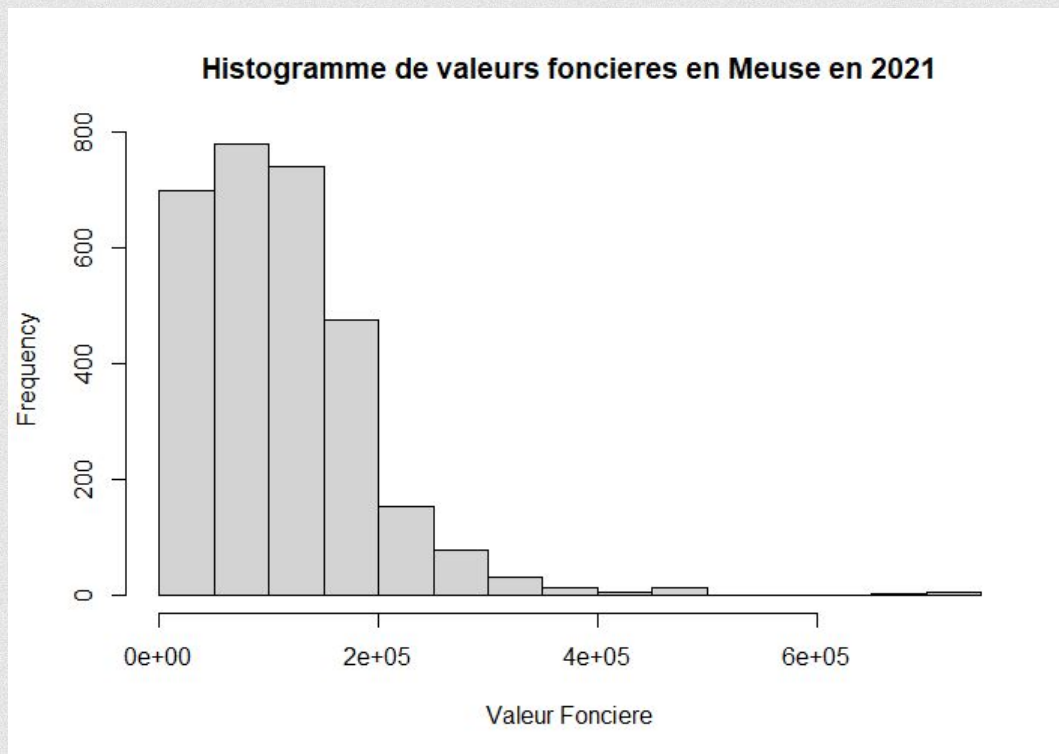
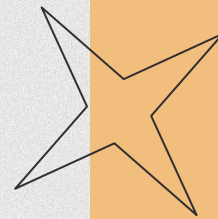
Histogramme de valeurs foncieres en Loire Atlantique en 2021



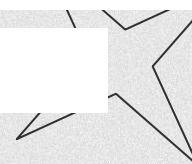
Pop : 1,426 million INSEE



## Evolution de la valeur foncière



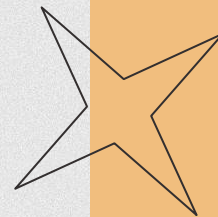
Pop : 184 474 INSEE



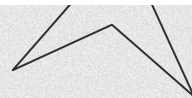
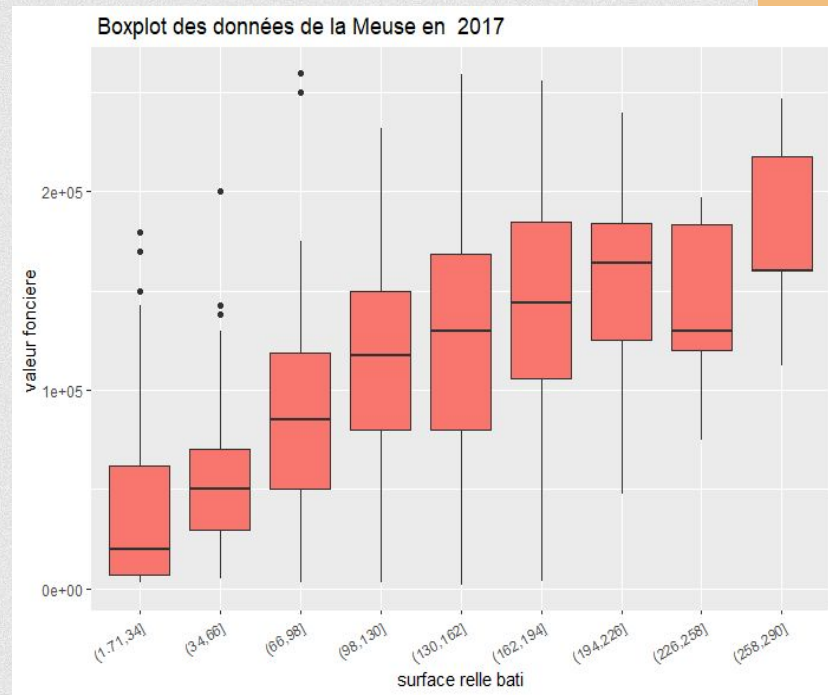
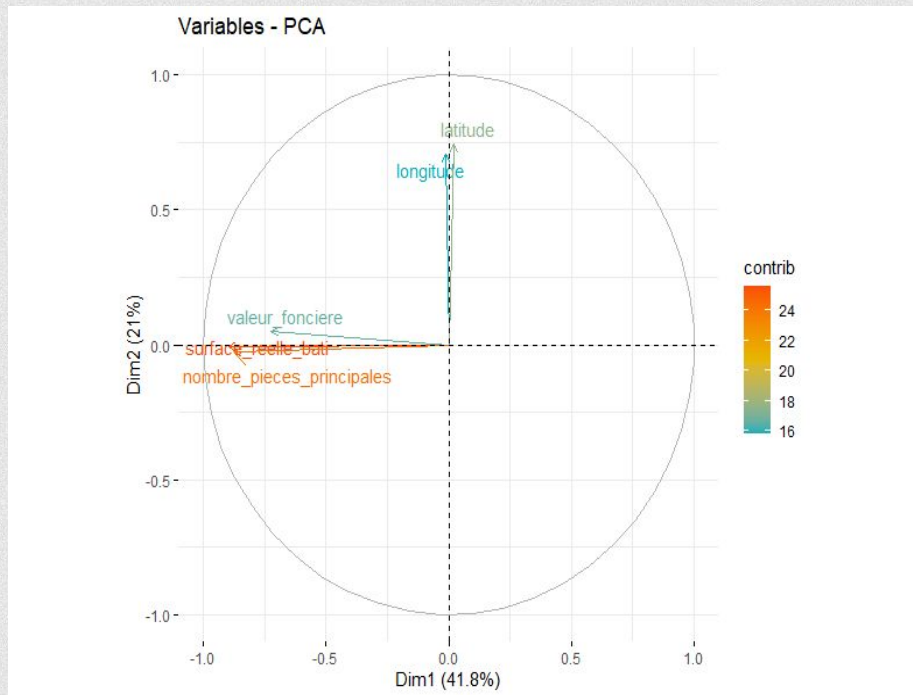




# Interprétation des résultats de l'ACP sur les données DVF 2017

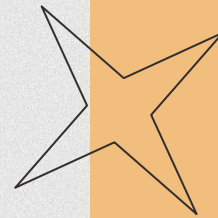


## a) Département de la Meuse

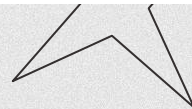
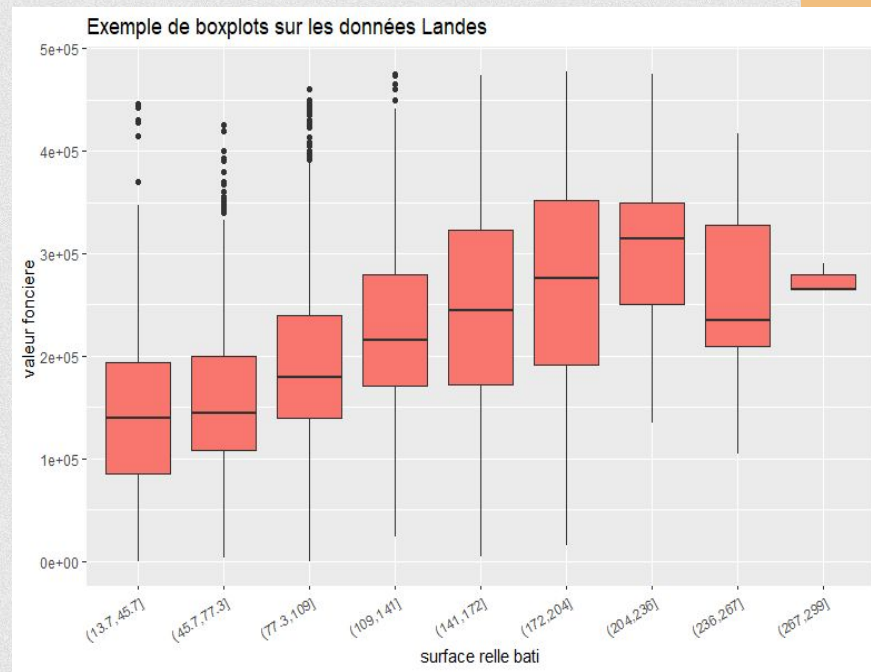
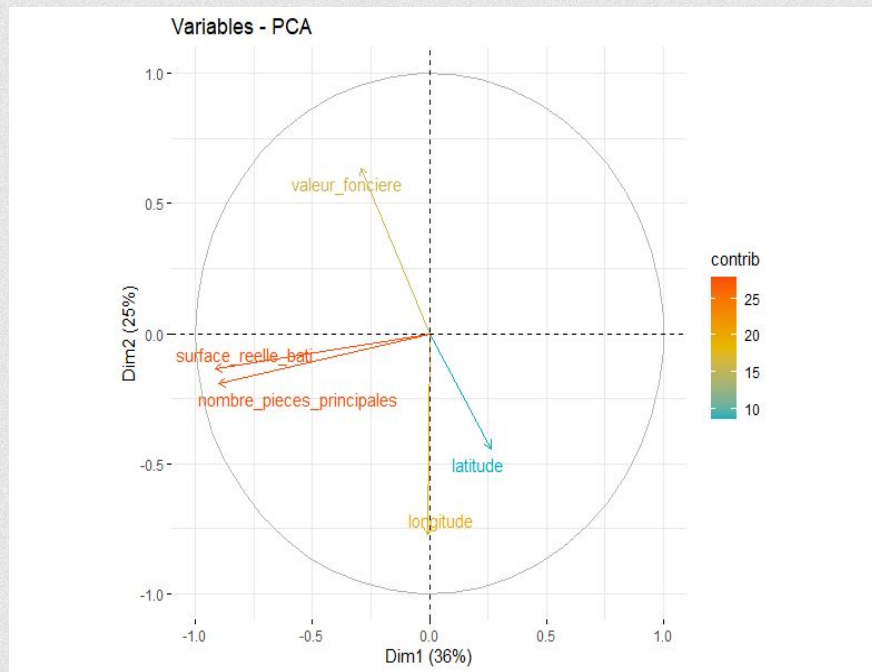




# Interprétation des résultats de l'ACP sur les données DVF 2017



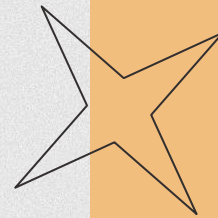
## a) Département des Landes



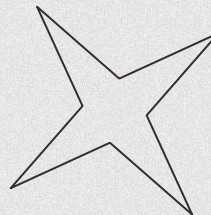
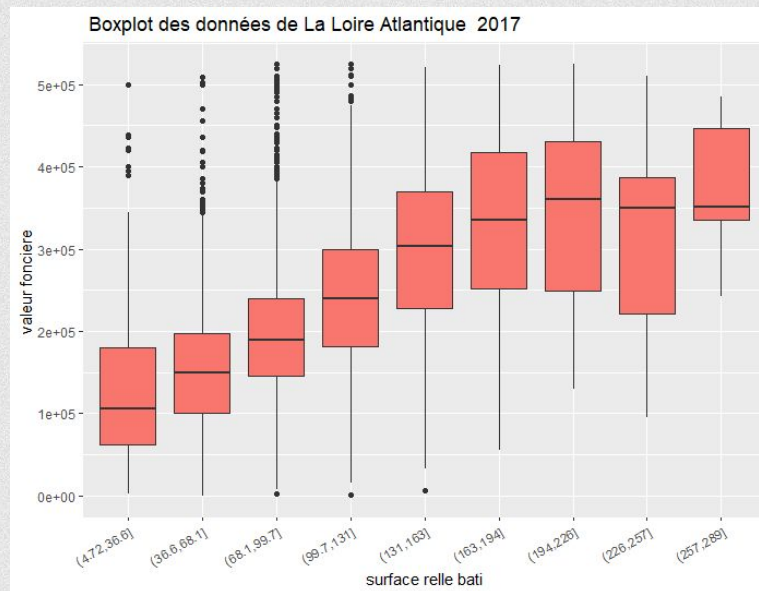
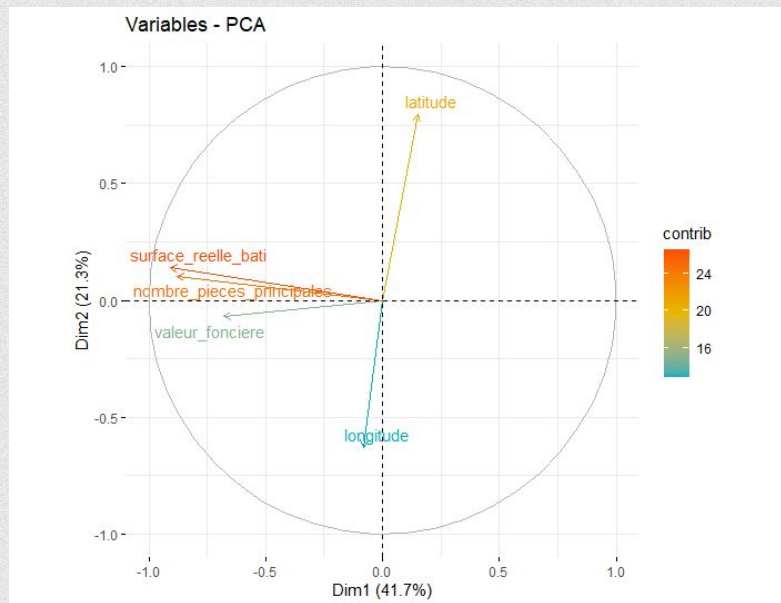




# Interprétation des résultats de l'ACP sur les données DVF 2017



## a) Département de la Loire Atlantique





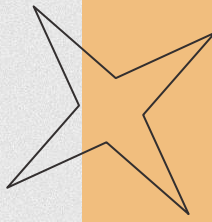
# **04.** **Etudes** **comparatives**



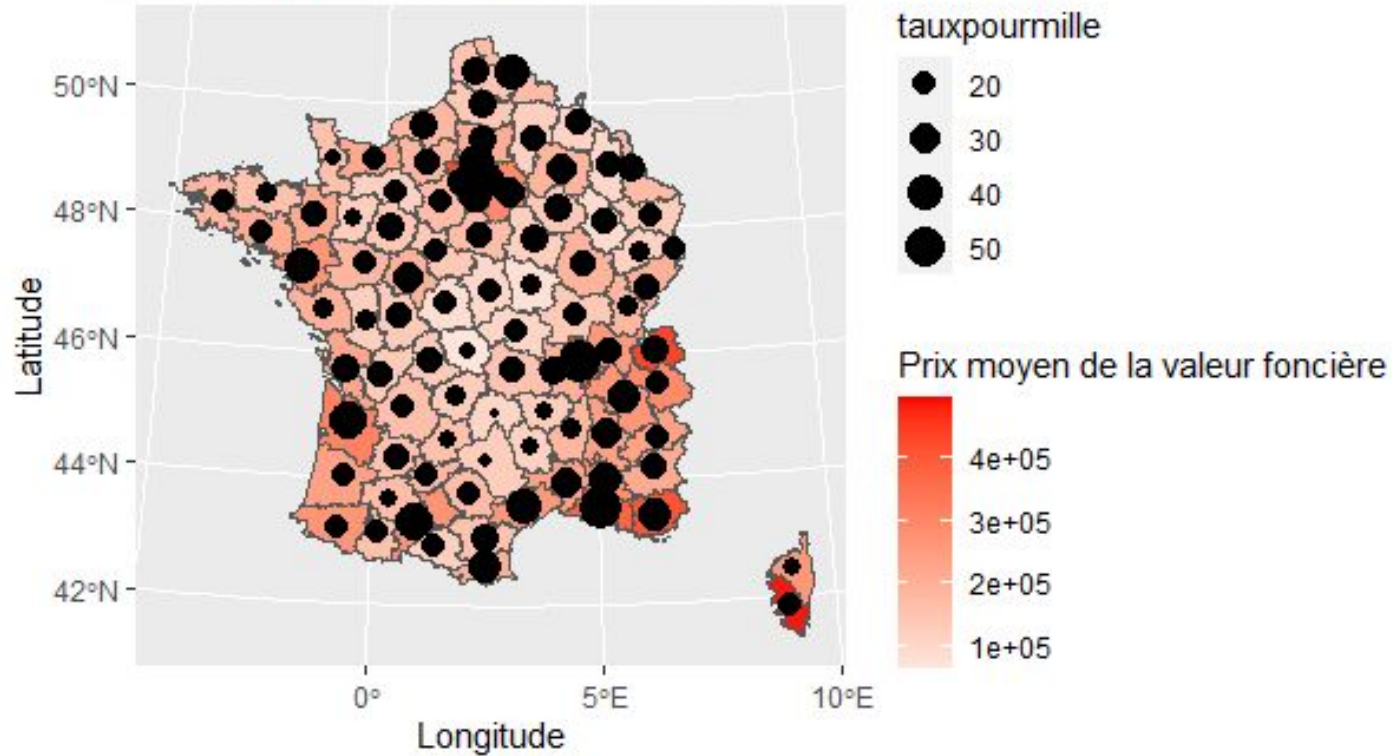




# Criminalité en France

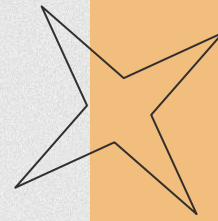


Le taux pour mille de criminalité en France en 2021



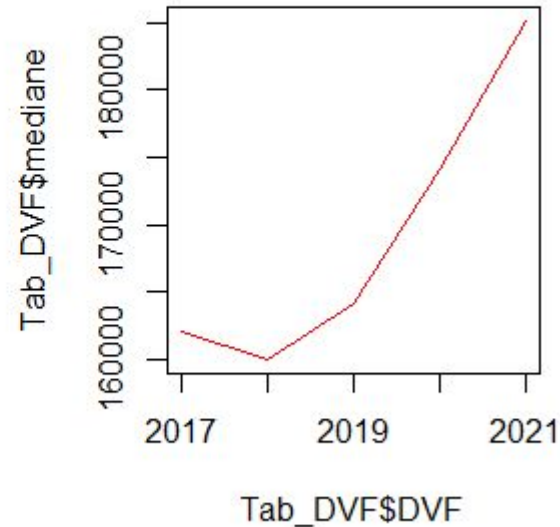
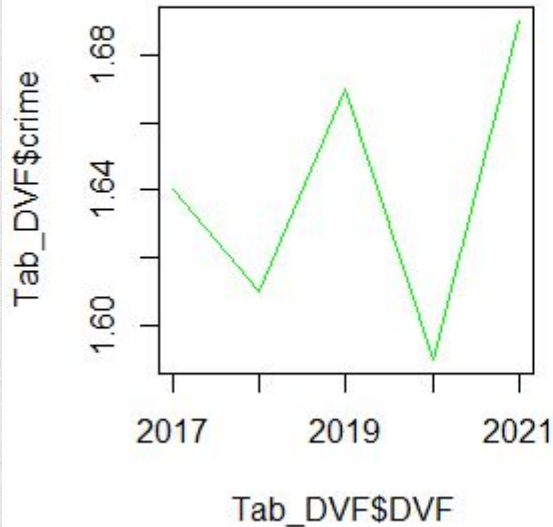


# Criminalité en France



Corrélation entre le taux de criminalité et le prix de la valeur foncière ?

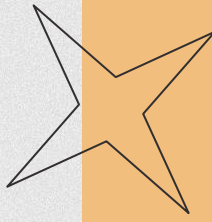
Le taux de criminalité (gauche) et le prix médian de l'immobilier (droite) de 2017 à 2021





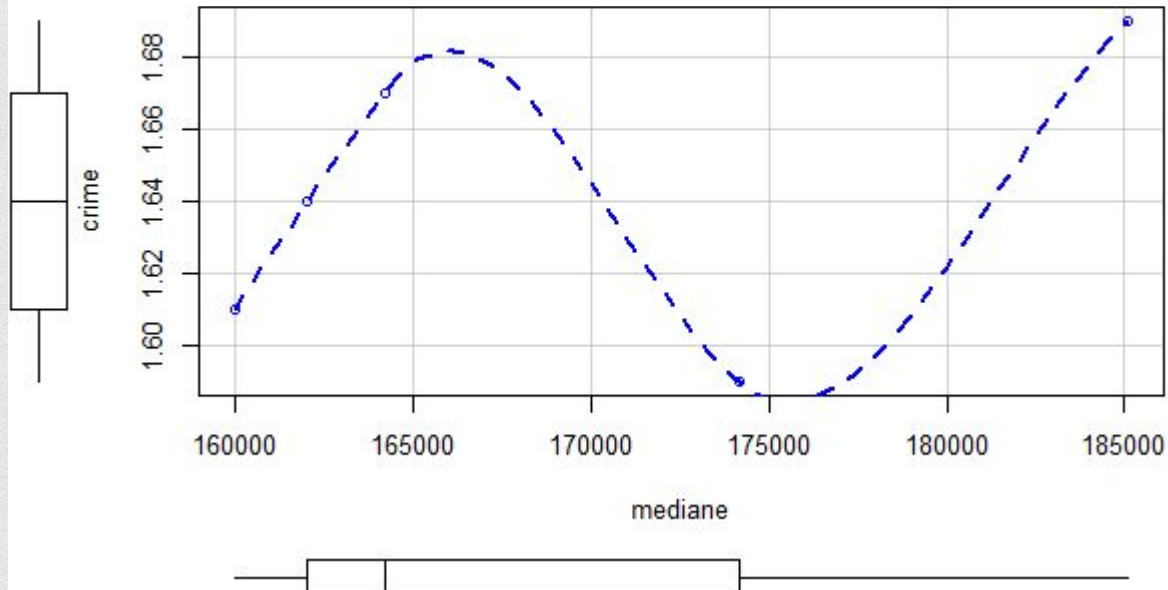


# Criminalité en France



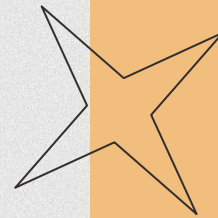
Corrélation entre le taux de criminalité et le prix de la valeur foncière ?

Scatter Plot appliqué pour voir le lien entre le taux de criminalité et le prix médian de l'immobilier

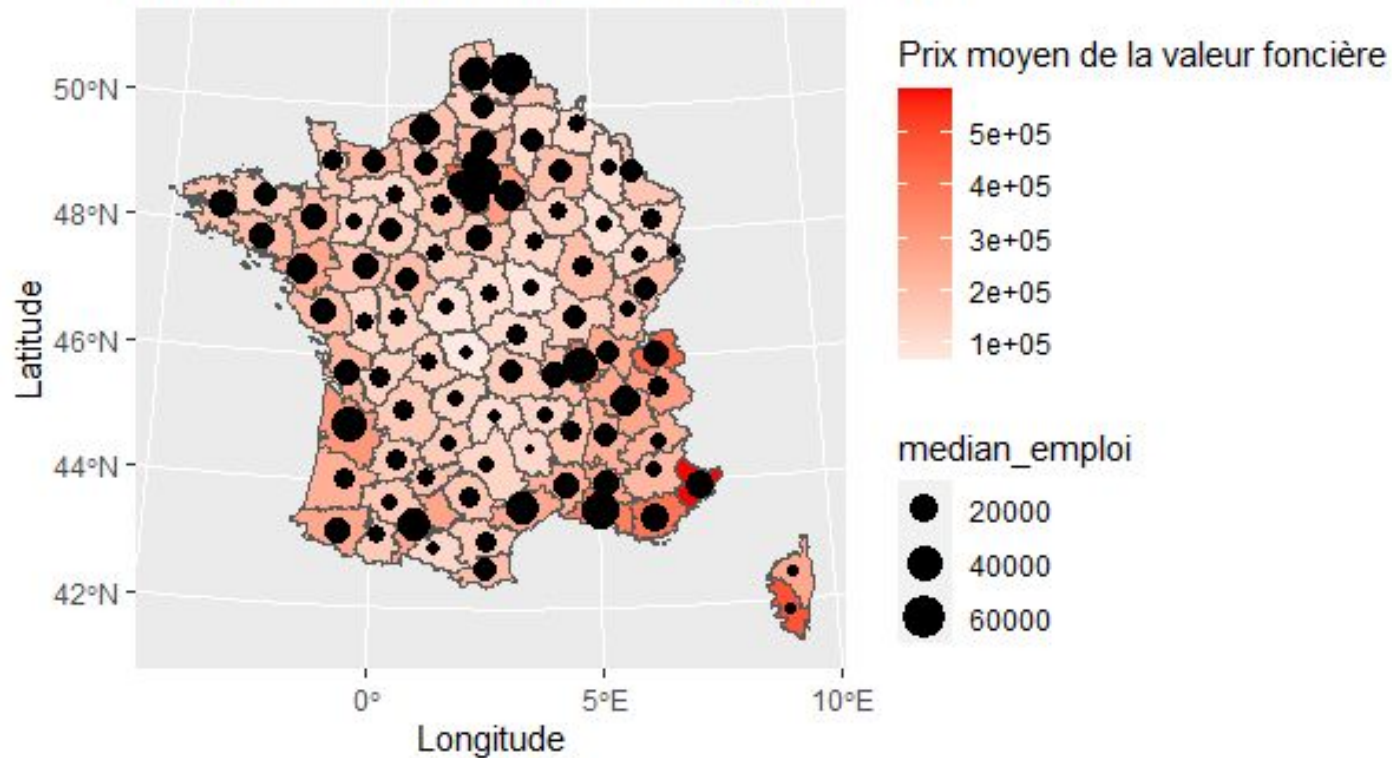




## Nombre de demandeurs d'emploi



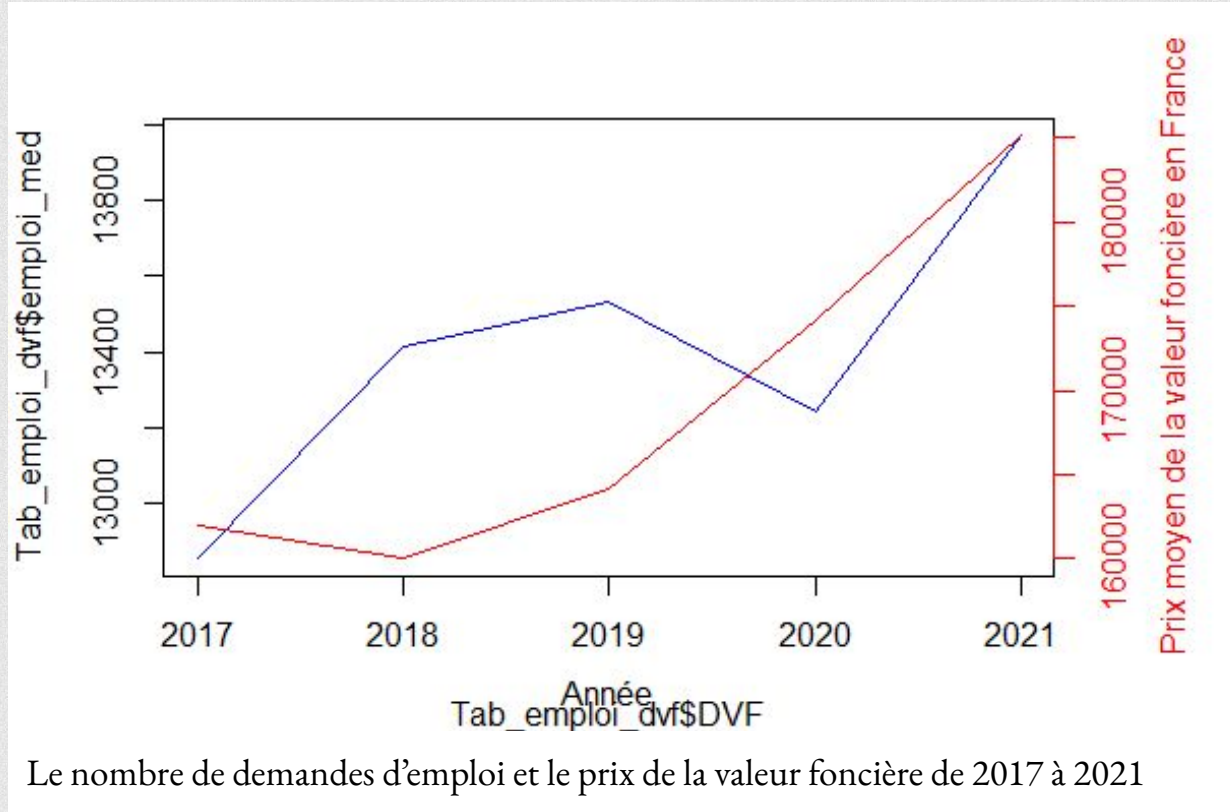
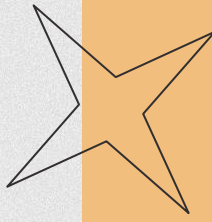
Le nombre de demandeurs d'emploi en 2021

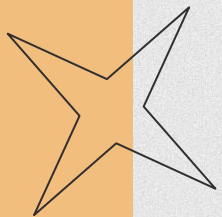






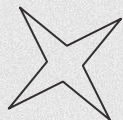
## Nombre de demandeurs d'emploi





# 05.

## Apprentissage





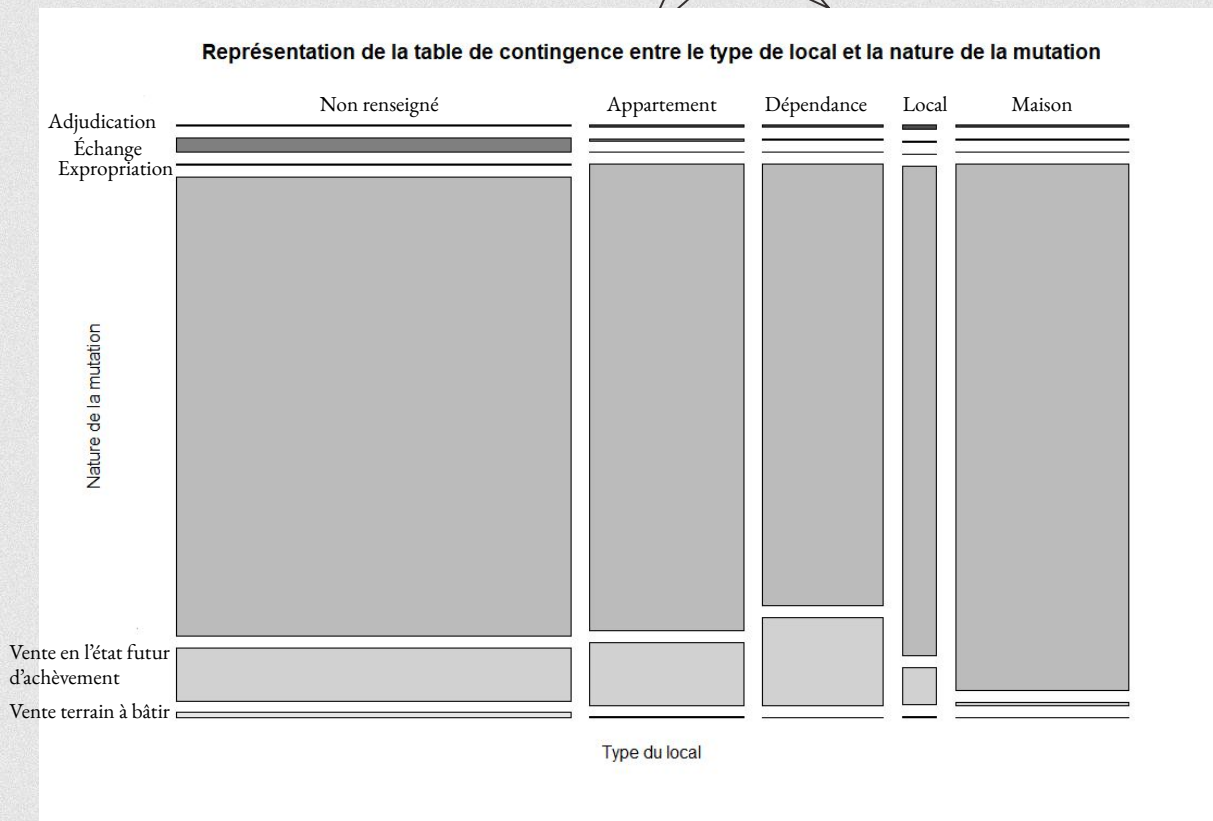


# Représentation des proportions des natures de la mutation avec le type de bien

-Énormément de types de locaux non renseignés

-Quasiment que des ventes et un petit peu de vente en l'état d'achèvement

Ce sont les deux seules variables qualitatives intéressantes. Le champ nature culture était composé de beaucoup trop de statuts.





# Création du jeu de données d'entraînement

Le but est de pouvoir trouver le type du local d'une mutation non étiquetée

Sur les 40 champs nous avons retenu un dizaine d'entre eux:

- Le prix
- La nature de mutation
- La nature culture
- Surface terrain
- Surface habitable (carrez)
- ...

Ajout d'un champs qui résume lorsqu'un bien tient sur plusieurs lignes:

N: nombre de lignes par bien

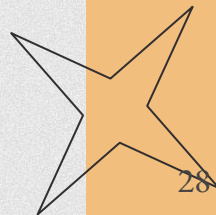
Surface\_total: somme sur toutes les surfaces terrains

Création de variable muette car certains champs sont quasiment jamais remplis mais peuvent déterminer entièrement la classe d'un individu

Suppression des individus qui n'avaient pas les champs remplis (prix, surface...)



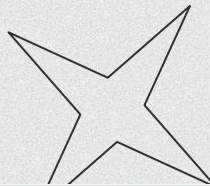
Données d'apprentissage de 150 000 individus.







# Données d'apprentissage



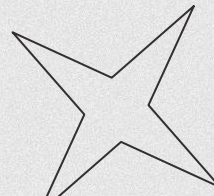
▲	valeur_fonciere ↕	type_local ↕	nature_culture ↕	surface_terrain ↕	n ↕	surface_1_carrez ↕	surface_2_carrez ↕	surface_3_carrez ↕	surface_4_carrez ↕	surface_5_carrez ↕
105	374000.00	Maison	sols	891	1	0	0	0	0	0
106	378390.00	Maison	terrains d'agrément	618	2	0	0	0	0	0
108	264800.00	Maison	sols	806	1	0	0	0	0	0
109	184700.00	Appartement	sols	82	1	0	0	0	0	0
110	350700.00	Dépendance	sols	592	11	0	0	0	0	0
121	120000.00	Maison	sols	320	1	0	0	0	0	0
122	96000.00	Maison	sols	270	1	0	0	0	0	0
123	175000.00	Appartement	sols	341	4	0	0	0	0	0
127	174000.00	Maison	sols	113	1	0	0	0	0	0
128	130000.00	Dépendance	sols	317	4	0	0	0	0	0
132	270000.00	Maison	sols	849	1	0	0	0	0	0
133	176500.00	Maison	sols	746	1	0	0	0	0	0
134	30000.00	Dépendance	sols	257	1	0	0	0	0	0
135	210000.00	Maison	sols	655	1	0	0	0	0	0
136	175000.00	Dépendance	sols	120	1	0	0	0	0	0
137	212282.00	Maison	sols	614	1	0	0	0	0	0
138	175000.00	Maison	sols	149	1	0	0	0	0	0
139	40000.00	Maison	sols	609	1	0	0	0	0	0
140	290700.00	Maison	sols	1963	1	0	0	0	0	0
141	100000.00	Maison	sols	230	1	0	0	0	0	0
142	203040.00	Maison	sols	1000	1	0	0	0	0	0
143	105000.00	Dépendance	sols	1830	2	0	0	0	0	0
145	268910.00	Maison	sols	1440	1	0	0	0	0	0
146	219550.00	Maison	sols	581	1	0	0	0	0	0
147	69500.00	Maison	sols	121	1	0	0	0	0	0
148	95000.00	Maison	sols	235	1	0	0	0	0	0
149	137810.00	Maison	sols	677	1	0	0	0	0	0
150	201550.00	Maison	sols	845	1	0	0	0	0	0

Showing 87 to 115 of 191,446 entries, 10 total columns





## Apprentissage par la méthode random forest de la bibliothèque randomForest



```
> model

Call:
  randomForest(formula = type_local ~ ., data = new_donnees_sans_doublon,      ntree = 400)
    Type of random forest: classification
    Number of trees: 400
No. of variables tried at each split: 3

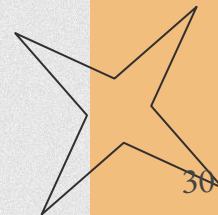
    OOB estimate of  error rate: 12.38%
Confusion matrix:

                Appartement  Dépendance  Local industriel. commercial ou assimilé  Maison  class.error
Appartement              3332             79                                24   3070 0.487778632
Dépendance                1057            2879                                9   8050 0.759983326
Local industriel. commercial ou assimilé              661             236          1098   9086 0.900911470
Maison                    646             641          134 160444 0.008778921
```

Résultat de l'apprentissage sur l'année 2017

Globalement les résultats ne sont pas satisfaisants mais ce qui nous intéresse le plus c'est la classification des maisons. Ce classifieur a un excellent rappel de 99,12% et une précision de 90% pour la classification des maisons.

Avec 100 ou 400 arbres les résultats sont similaires







# Pistes d'amélioration pour l'apprentissage



Apprendre avec le champ nature culture spécial et garder uniquement les attributs qui reviennent le plus souvent. Car limitation de la fonction randomForest sur les valeurs qualitatives: impossibilité d'en avoir plus de 53. Or le champ nature culture spéciale en compte 85.

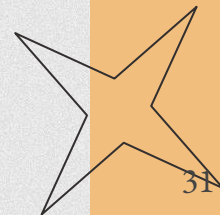
Étudier un autre algorithme d'apprentissage comme SVM.

Trouver plus d'outlier car le jeu de données de base est très grand.

Agréger avec les autres années pour avoir un modèle encore plus robuste.

Faire la phase de test, pas eu les ressources temporelles pour le faire.

Comparer les résultats à un classifieur naïf.



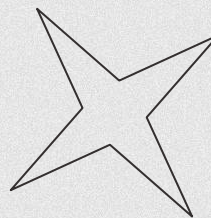
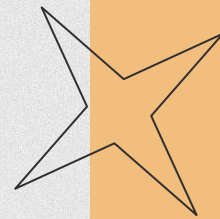


# 06.



## Conclusion

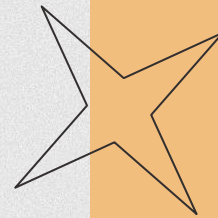
+Perspectives



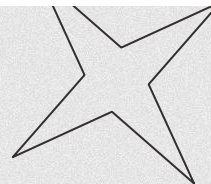
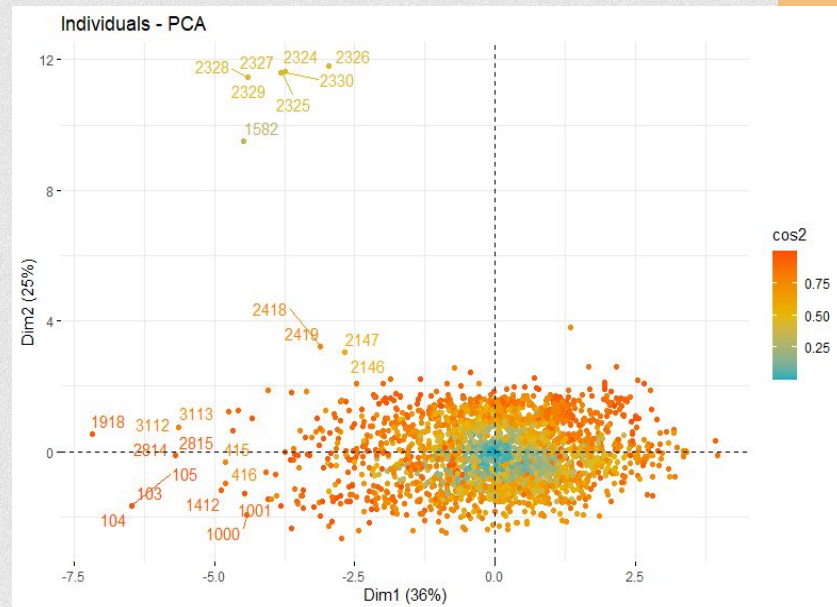
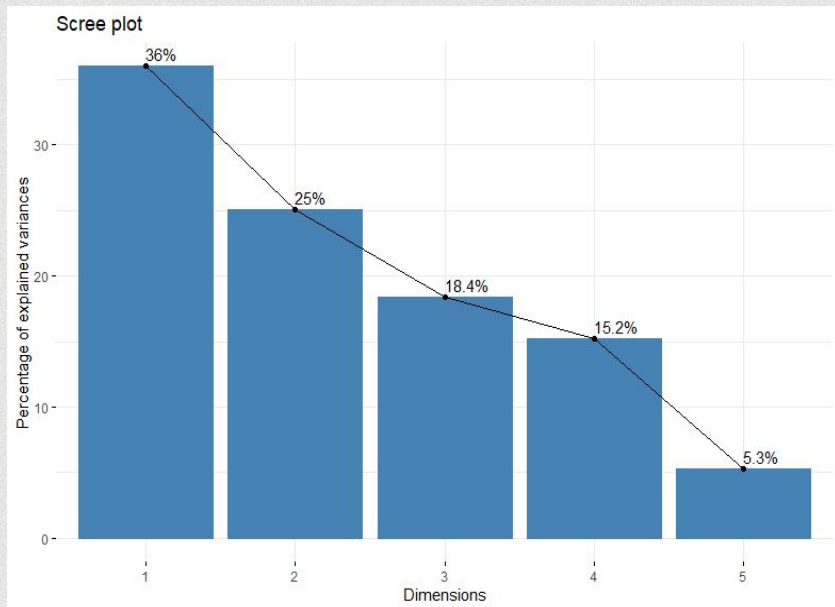




# Interprétation des résultats de l'ACP sur les données DVF 2017

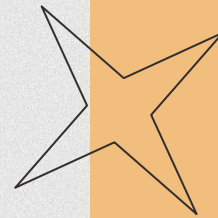


## a) Département des Landes

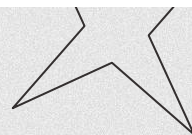
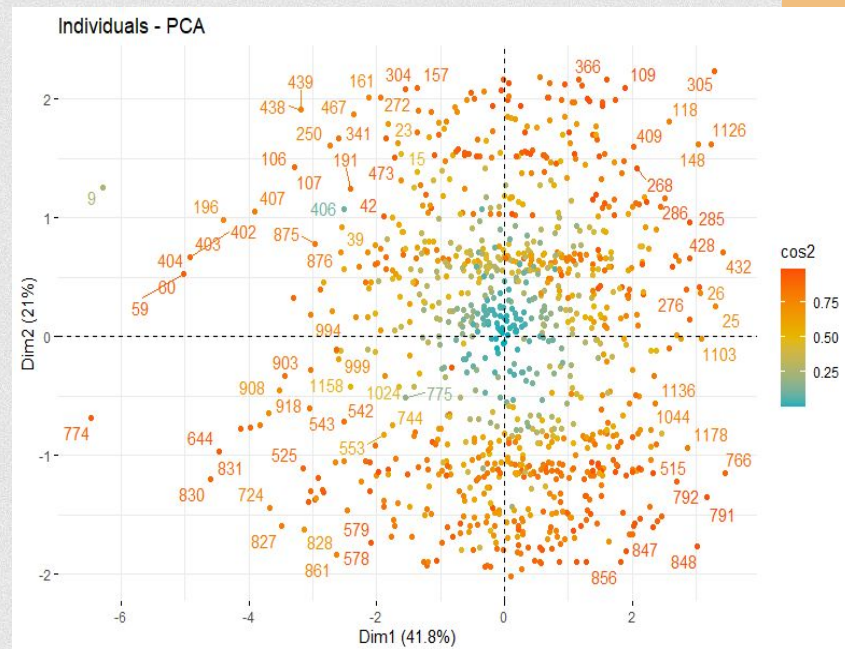
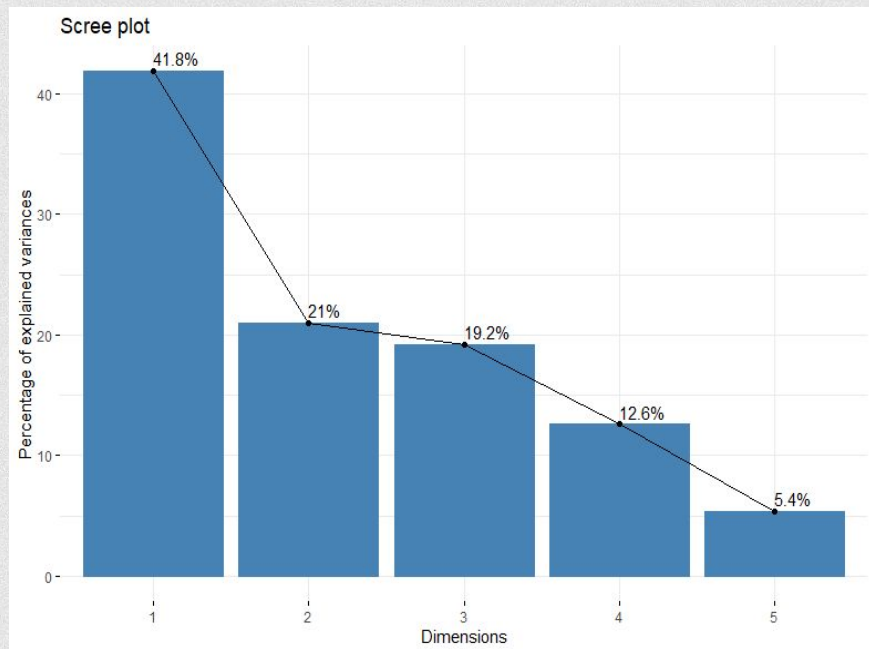




# Interprétation des résultats de l'ACP sur les données DVF 2017



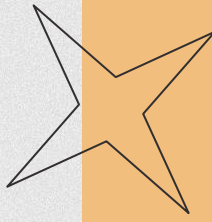
a) Département de la Meuse







# Interprétation des résultats de l'ACP sur les données DVF 2017



a) Département de la Loire Atlantique

