

Statistique

Licence de Géographie
Aménagement

1^{er} année

2017-2018

Organisation du cours

- 11 séances de 2 heures :
 - 9 séances de cours
 - 2 séances d'évaluation (contrôle continu)
- Organisation des séances :
 - 50 minutes de présentation
 - 1 heure de mise en application (exercices)

Objectifs du cours

- **Acquérir les notions essentielles de la statistique descriptive**, c'est-à-dire apprendre comment décrire de façon claire et concise l'information apportée par des observations nombreuses et variées sur un phénomène donné.
- Plus précisément, il s'agira de **trier** des données, de les **décrire**, de les **résumer** sous forme de **tableaux**, de **graphiques**, et sous forme d'un petit nombre de **paramètres clés**.

Plan du cours

- **Séance 1** - définition de la statistique et vocabulaire usuel
- **Séance 2** - représentation de l'information : tableaux et graphiques
- **Séance 3** - mesures des inégalités
- **Séance 4** - statistique descriptive univariée : les paramètres statistiques
- **Séance 5** - échantillonnage
- **Séance 6, 7, 8, 9** - statistique descriptive bivariée : relation entre deux caractères

Séance 1

**Définition de la statistique et
vocabulaire usuel**

- **Ensemble d'outils** permettant de **décrire** et **d'analyser**, de façon quantifiée, **des phénomènes** repérés par des **observations nombreuses**, de même nature, susceptibles d'être dénombrées et classées.
- Elle permet :
 - **De créer une information systématique** (observations de même nature sur un ensemble homogène d'objets), ce qui autorise les comparaisons ;
 - **De traiter l'information** ainsi créée (résumés graphiques, numériques, étude de relation entre phénomènes, etc.) ;
 - **De connaître la fiabilité de l'information**, notamment en cas de sondage ;
 - **De progresser vers des applications opérationnelles.**
- Elle ne permet pas :
 - **De remplacer le raisonnement de type explicatif ;**
 - **De remplacer la culture et les connaissances** du phénomène étudié.

I. Qu'est-ce que la statistique ? 2. Confusion statistique et statistiques

- On n'entend parfois que les statistiques sont « fausses » ou « truquées ». Cette affirmation vient de la **confusion entre LA statistique et LES statistiques**.
- **La statistique** forme un ensemble, un corpus, d'outils destinés à produire et analyser de l'information. Bien utilisée, elle est rigoureuse.
- **Les statistiques ou données statistiques** ne sont que les résultats numériques auxquels conduit l'application de ces outils. Elles peuvent donc être fausses, mal interprétées ou même détournées pour défendre un intérêt particulier.

On distingue traditionnellement **deux branches** dans la statistique :

- **La statistique descriptive** dont l'objectif est de représenter l'information d'une manière compréhensible et utilisable;
- **La statistique inférentielle** qui a pour but de fournir des résultats relatifs à une population à partir de mesures statistiques réalisées sur des échantillons.

- Le terme statistique est issu du vocable latin ***status* (Etat)**. La statistique est au départ **liée à l'idée de dénombrement, d'inventaire et même de recensement** réalisé par les États pour des besoins économiques et militaires.
- Les débuts des recensements se situent à l'époque des grandes civilisations antiques. Parmi les plus anciennes on peut citer celle de Sumer (entre 3000 et 5000 Av. JC).
- L'utilisation par les États des statistiques à des fins plus élaborées (**arithmétique politique**) **date du milieu du XVIIème siècle** et constitue la **naissance de la statistique moderne**.
- Les méthodes et les lois statistiques connurent leur **grand développement au XIXème** sous l'impulsion principalement de mathématiciens (Laplace, Poisson, Pearson, etc.)
- C'est seulement au **XXème siècle** que la **statistique devient une science à part entière**, autonome.

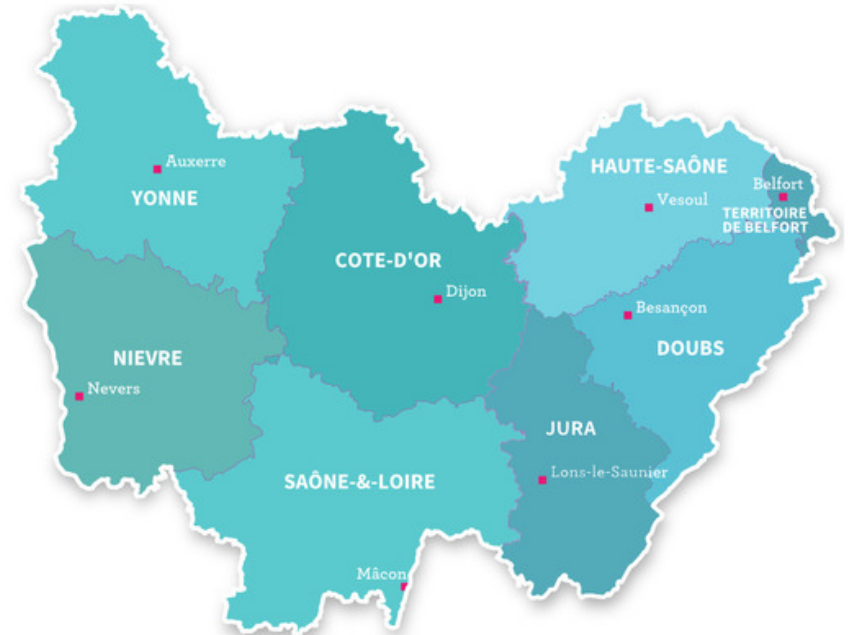
- La statistique est **très largement utilisée en géographie** surtout **depuis les années 60** avec le développement de l'informatique :
- Les données que traite le géographe présentent certaines particularités :
 - **Les individus statistiques sont fréquemment des unités spatiales**, elles sont donc géo-localisables et cartographiables.
 - **Ces unités spatiales sont souvent des agrégats** : ensemble d'habitants, d'entreprises etc.
- Les outils statistiques permettent au géographe de répondre à ses questions de prédilection : quels sont les principes de l'organisation de l'espace ? Existe-t-il des régularités, des gradients, des spécificités locales ?

Exemple: étude d'une région

➤ POPULATION STATISTIQUE

Définition: ensemble d'éléments, de personnes, d'objets, de pays observés

Notation: Ω (dire oméga majuscule)



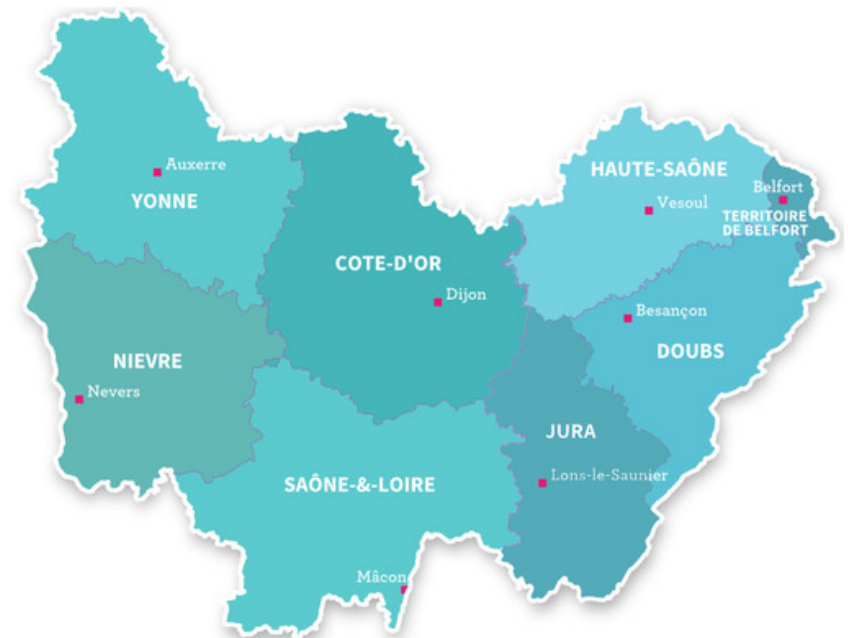
Exemple: étude d'une région

➤ POPULATION STATISTIQUE

Définition: ensemble d'éléments, de personnes, d'objets, de pays observés

Notation: Ω (dire oméga majuscule)

⇒ La région BFC

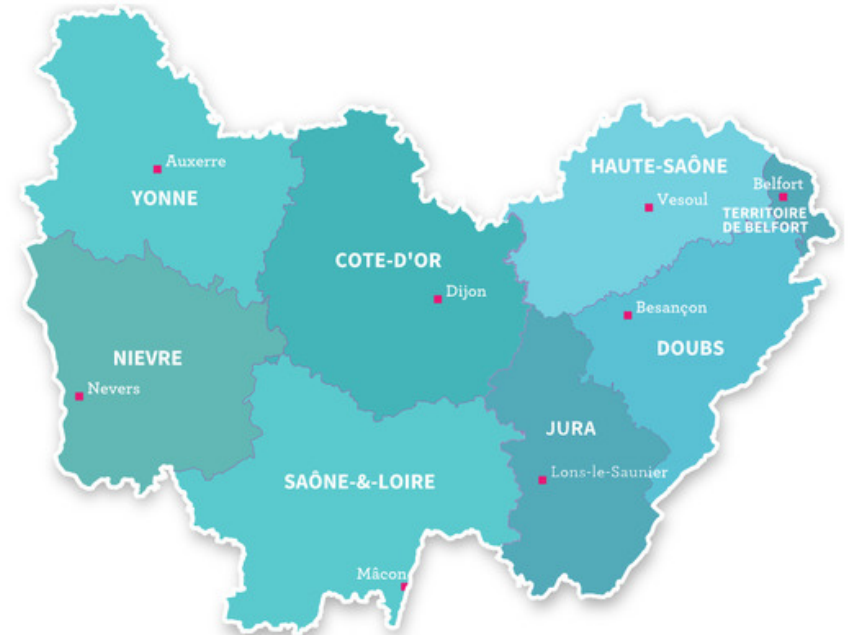


Exemple: étude d'une région

➤ INDIVIDUS OU UNITÉS STATISTIQUES

Définition: éléments, objets, personnes composant la population

Notation: ω (dire oméga minuscule)



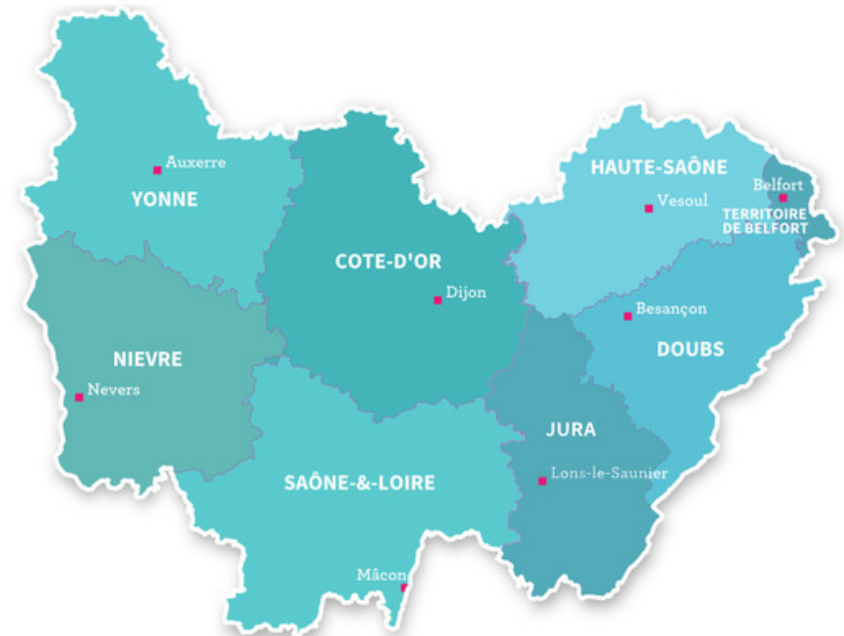
Exemple: étude d'une région

➤ INDIVIDUS OU UNITÉS STATISTIQUES

Définition: éléments, objets, personnes composant la population

Notation: ω (dire oméga minuscule)

⇒ les départements

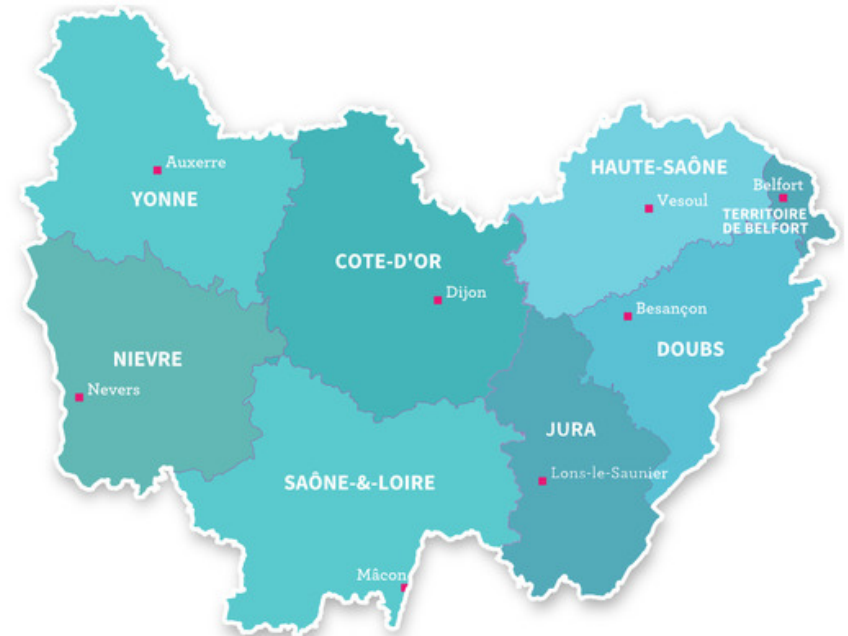


Exemple: étude d'une région

➤ TAILLE DE LA POPULATION

Définition: nombre d'individus que compte la population

Notation: N



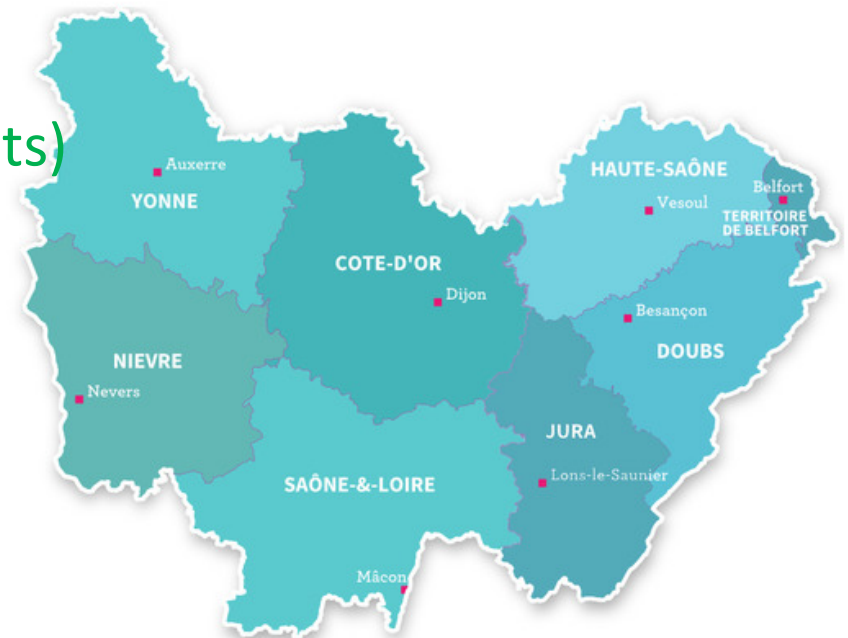
Exemple: étude d'une région

➤ TAILLE DE LA POPULATION

Définition: nombre d'individus que compte la population

Notation: N

⇒ $N = 8$ individus (les départements)

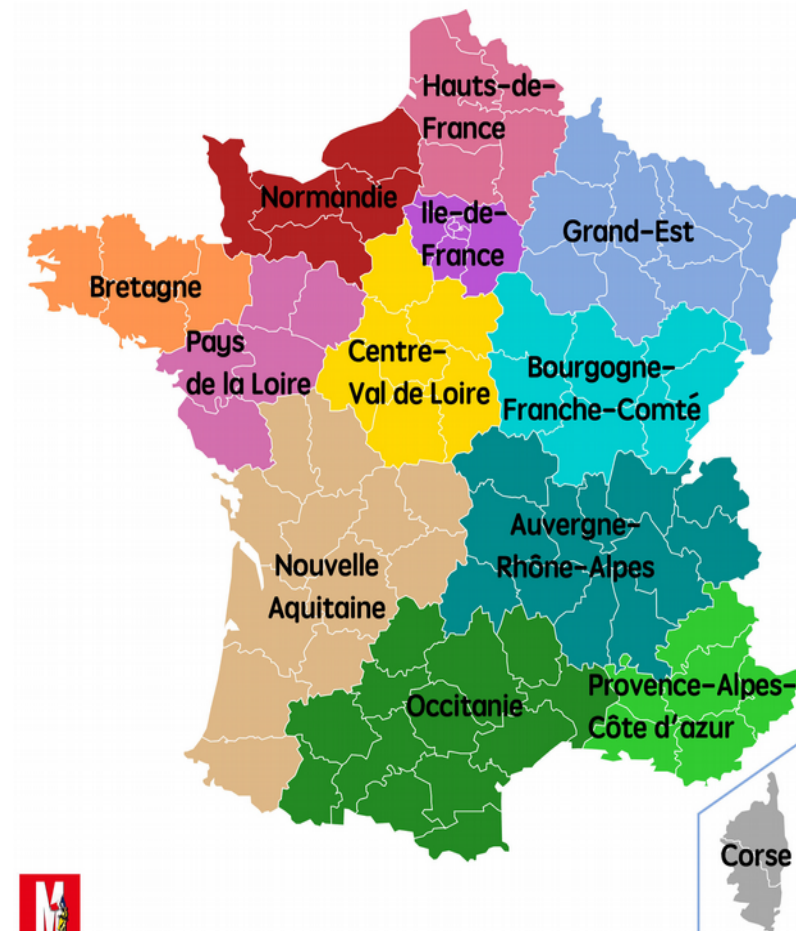


Exemple: étude de quelques départements

➤ ÉCHANTILLON

Définition: Ensemble d'individus extraits d'une population étudiée

On les utilise quand la population est trop importante ou le budget limité

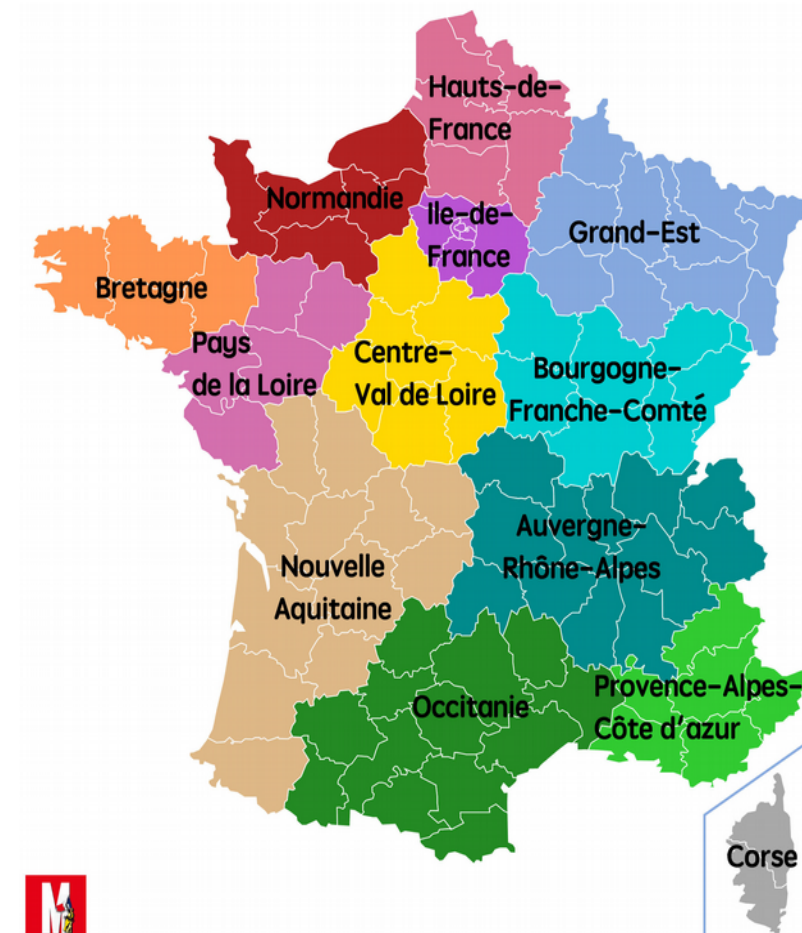


Exemple: étude de quelques départements

➤ TAILLE DE L'ÉCHANTILLON

Définition: Nombre d'individus composant l'échantillon

Notation : n



Exemple: étude de quelques départements

➤ TAILLE DE L'ÉCHANTILLON

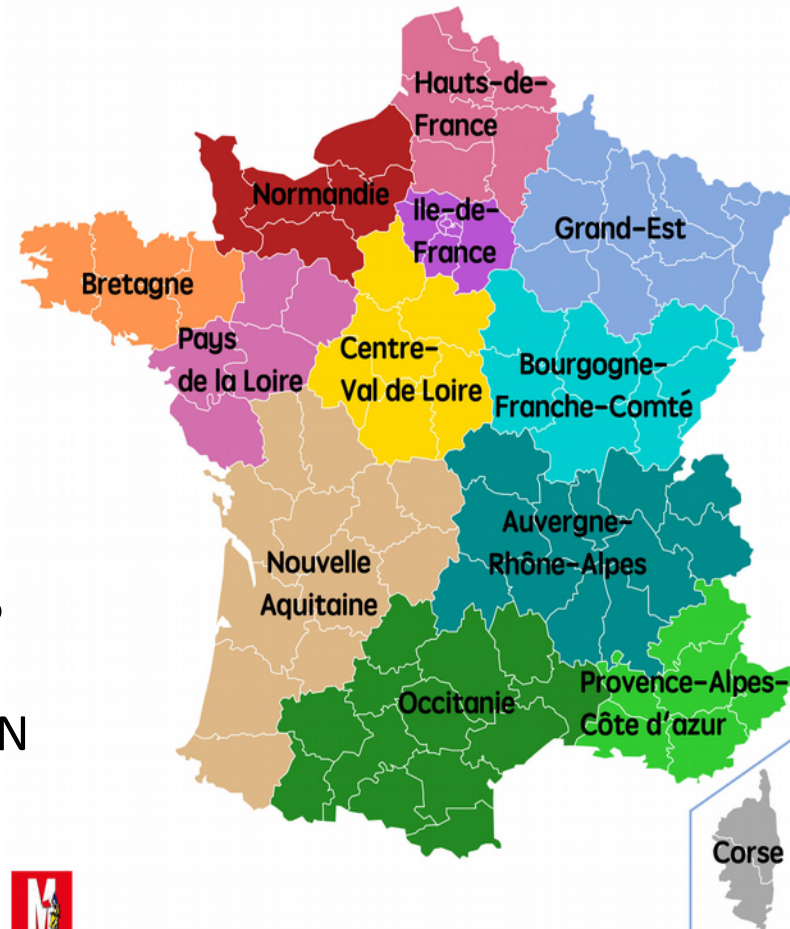
Définition: Nombre d'individus composant l'échantillon

Notation : n

Si je tire au hasard 5 départements français : $n = 5$

Le **TAUX DE SONDAGE** correspond au rapport : n/N

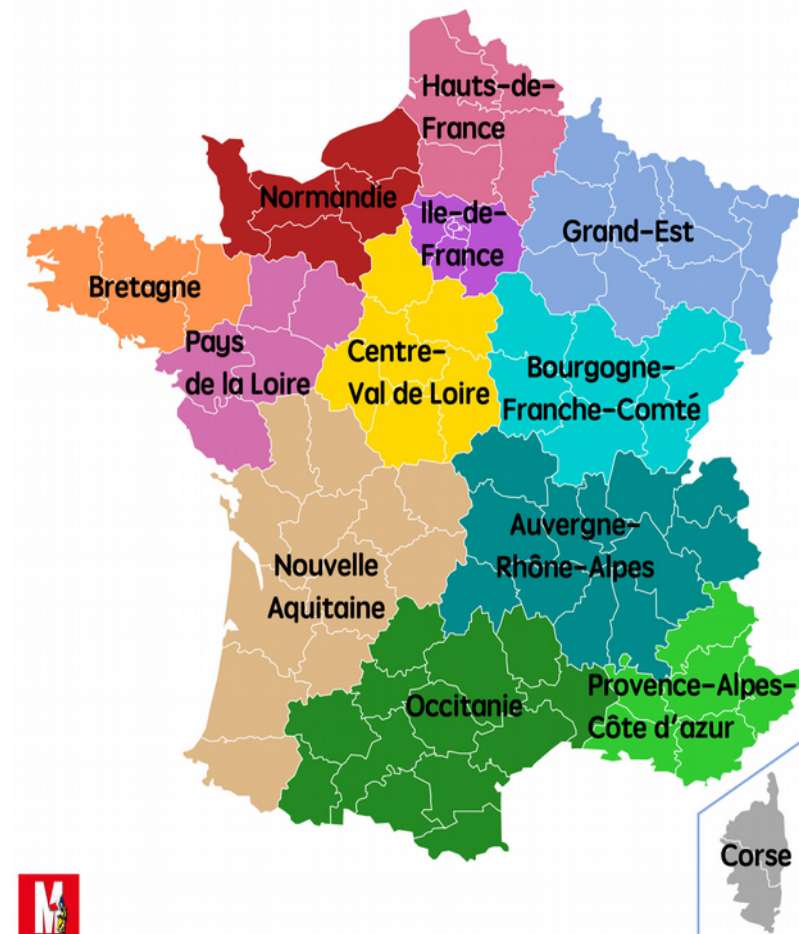
⇒ $5/96=0,052$ (5,2%)



➤ CARACTÈRE ou VARIABLE

Définition: Données caractérisant la population

Notation : on note la première **X**, la seconde **Y**, etc.



Exemple: étude d'une région

➤ CARACTÈRE ou VARIABLE

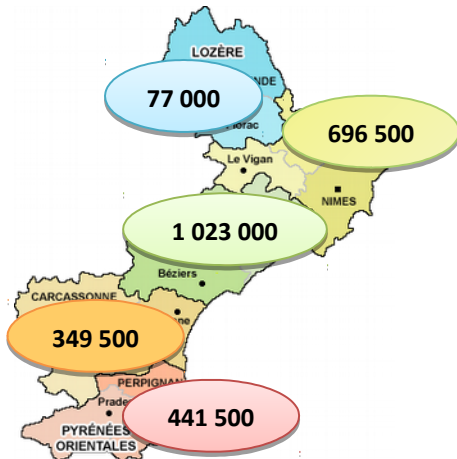


Fig. 1 : Nombre d'habitants par département en 2008, INSEE

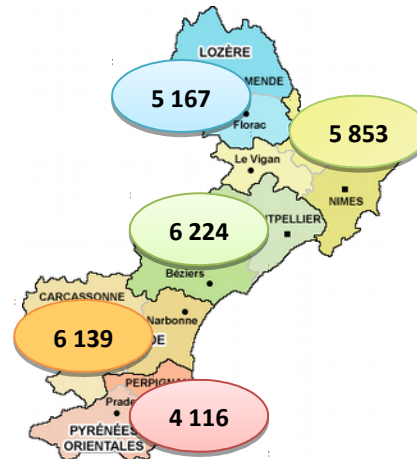


Fig. 2 : Superficie en km², INSEE

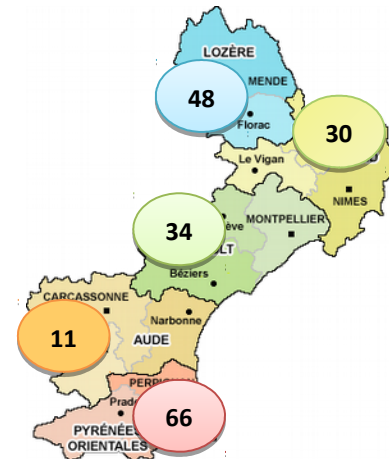


Fig. 3 : Code géographique

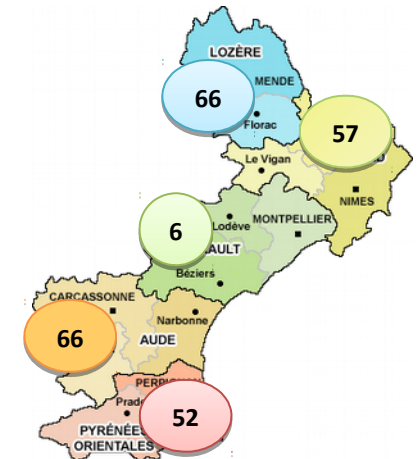


Fig. 4: Classement général des départements où il fait bon vivre en 2008, L'Express

Exemple: étude d'une région

➤ CARACTÈRE ou VARIABLE

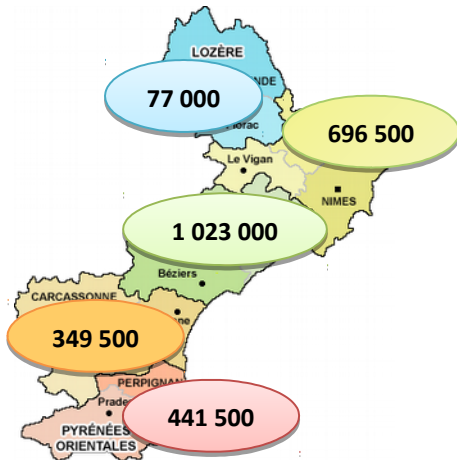


Fig. 1 : Nombre d'habitants par départements en 2008, INSEE

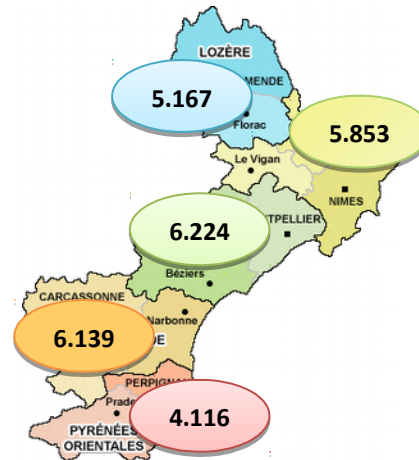


Fig. 2 : Superficie en km², INSEE

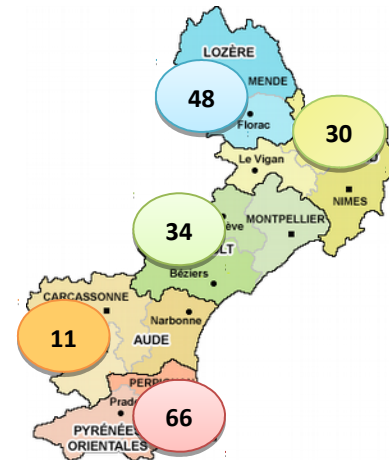


Fig. 3 : Code géographique

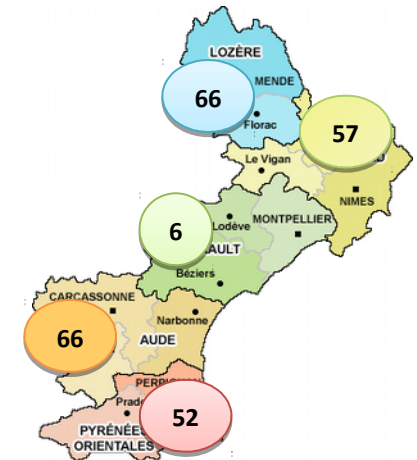


Fig. 4: Classement général des départements où il fait bon vivre en 2008, L'Express

⇒ 4 variables

Exemple: étude d'une région

➤ MODALITÉ

Définition: Valeurs possibles prises par le caractère

Caractéristiques : un individu = 1 et 1 seule modalité du caractère

Notation : le nombre de modalités est généralement noté **k**

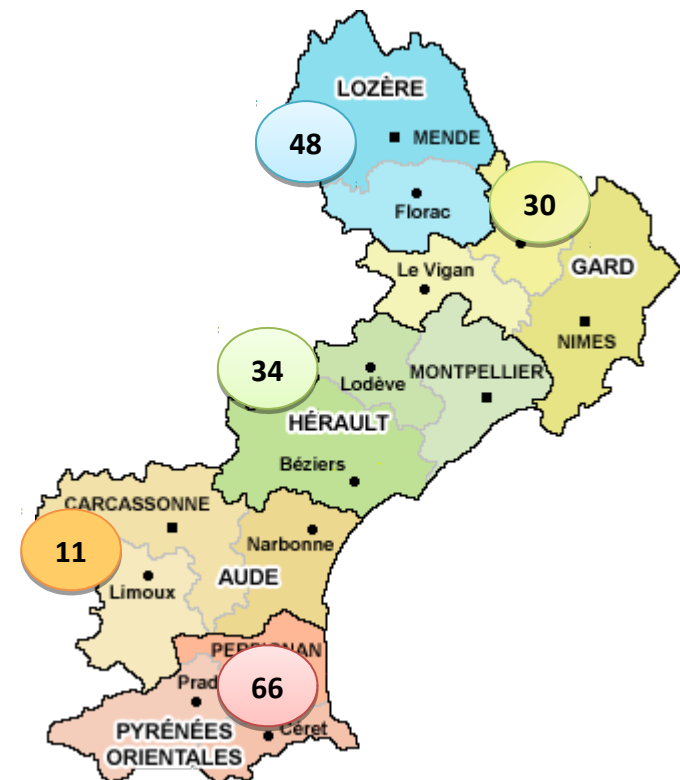


Fig. Code géographique

Exemple: étude d'une région

➤ MODALITÉ

Définition: Valeurs possibles prises par le caractère

Caractéristiques : un individu = 1 et 1 seule modalité du caractère

Notation : le nombre de modalités est généralement noté **k**

⇒ **k = 5 modalités (5 valeurs possibles)**

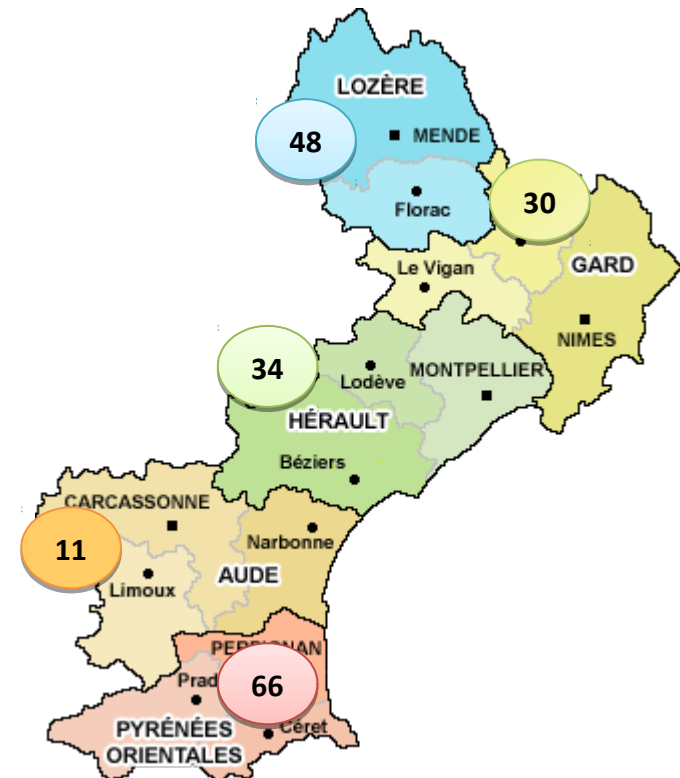


Fig. Code géographique

Exemple: étude d'une région

➤ DOMAINE DE LA VARIABLE

Définition: Ensemble des modalités

Notation : { }

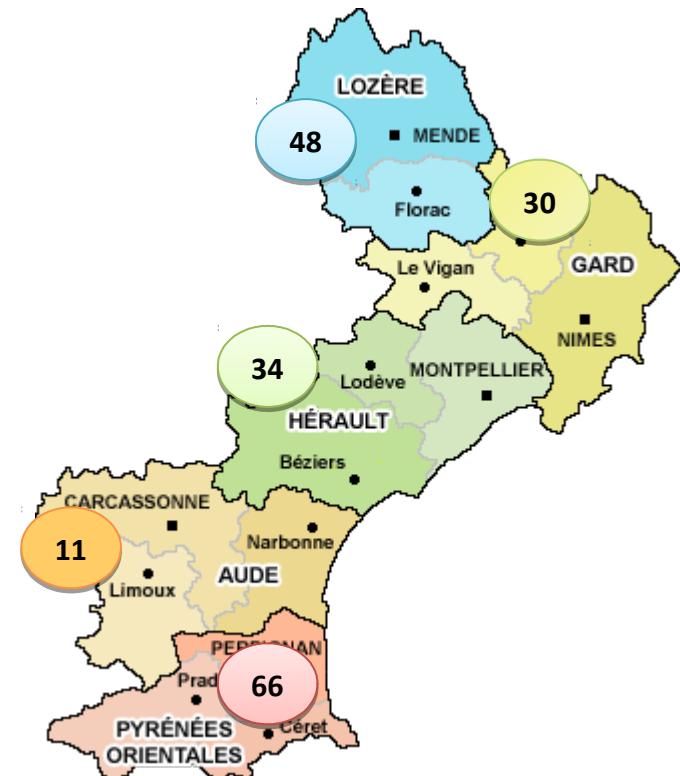


Fig. Code géographique

Exemple: étude d'une région

➤ **DOMAINE DE LA VARIABLE**

Définition: Ensemble des modalités

Notation : { }

⇒ Le domaine de la variable code géographique est {11 ; 66 ; 34 ; 48 ; 30}

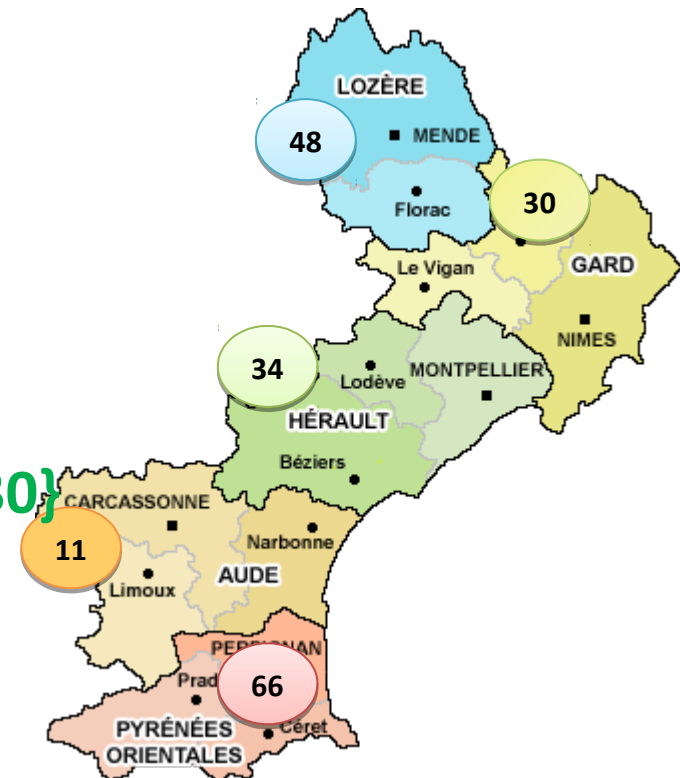


Fig. Code géographique

Exemple: étude d'une région

➤ SÉRIE STATISTIQUE

Définition: Suite des valeurs prises par une variable X sur les individus statistiques

Notation : le nombre d'individus est noté n et les valeurs (modalités) de la variable X sont notées : $x_1, \dots, x_i, \dots, x_k$

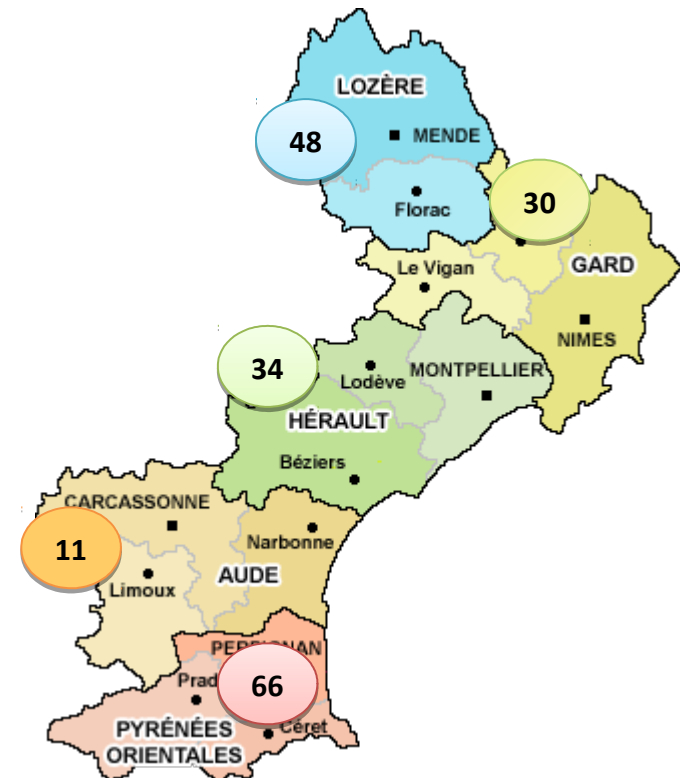


Fig. Code géographique

Exemple: étude d'une région

➤ SERIE STATISTIQUE

Définition: Suite des valeurs prises par une variable X sur les individus statistiques

Notation : le nombre d'individus est noté n et les valeurs de la variable X sont notés : $x_1, \dots, x_i, \dots, x_n$

⇒ Considérons la série statistique du code géographique : 11 66 34 48 30. On peut noter : $x_1=11, x_2=66, \dots, x_5=30$

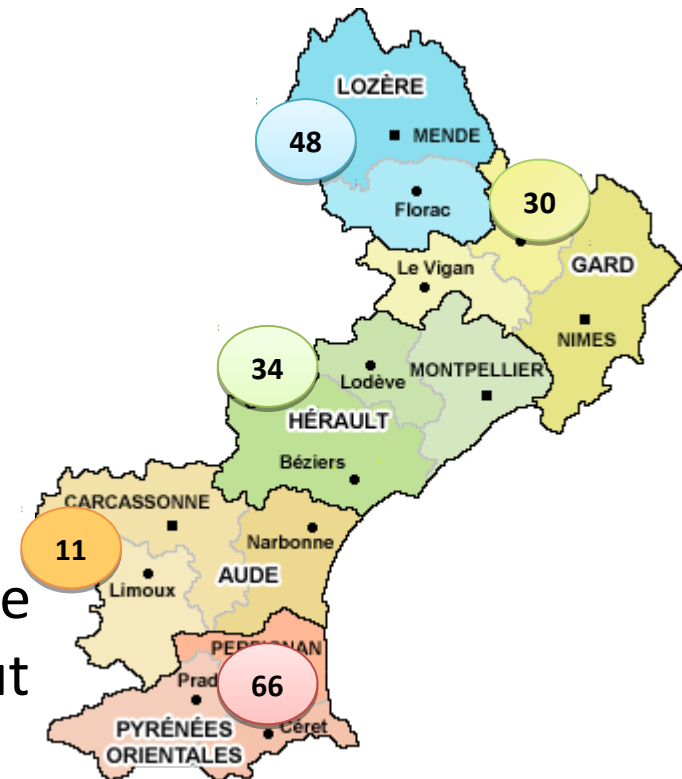
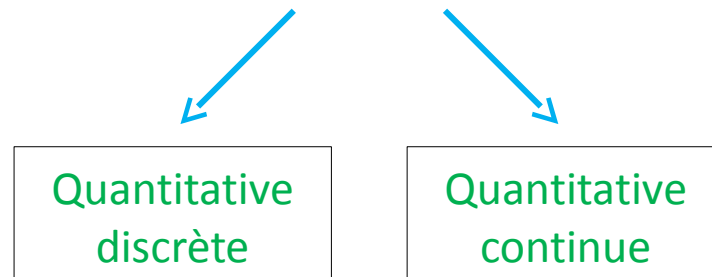


Fig. Code géographique

Exemple: étude d'une région

➤ TYPES DE VARIABLE

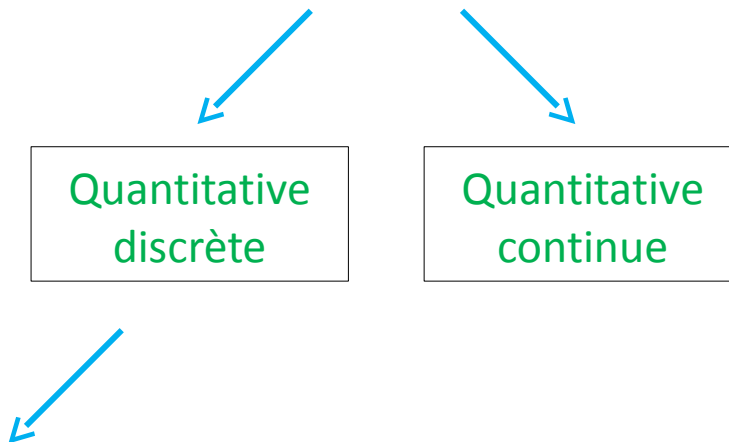
- ✓ Variable **quantitative** : traduit une quantité. Issue de mesure ou dénombrement



Exemple: étude d'une région

➤ TYPES DE VARIABLE

- ✓ Variable **quantitative** : traduit une quantité. Issue de mesure ou dénombrement

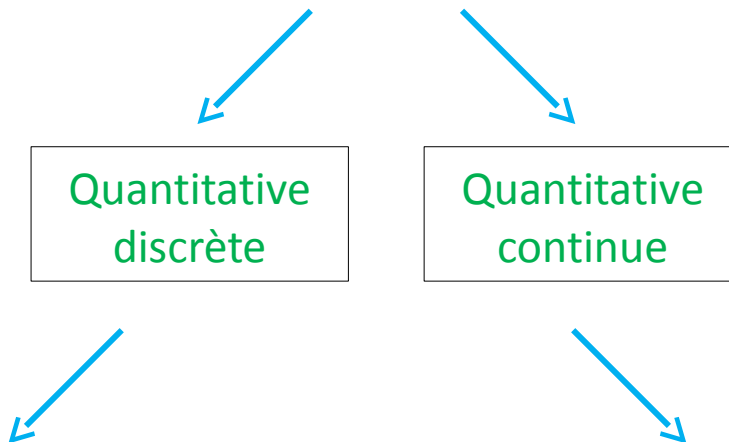


Valeurs isolées et entières dans l'intervalle de variation (ex : âge d'un individu, nombre d'enfants d'un ménage)

Exemple: étude d'une région

➤ TYPES DE VARIABLE

- ✓ Variable **quantitative** : traduit une quantité. Issue de mesure ou dénombrement



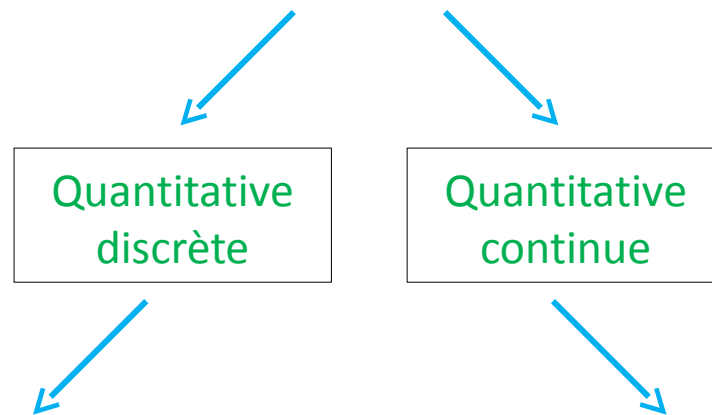
Valeurs isolées et entières dans l'intervalle de variation (ex : âge d'un individu, nombre d'enfants d'un ménage)

Toutes valeurs possibles dans intervalle de variation (ex : taille d'un individu)

Exemple: étude d'une région

➤ TYPES DE VARIABLE

- ✓ Variable **quantitative** : traduit une quantité. Issue de mesure ou dénombrement



Valeurs isolées et entières dans l'intervalle de variation (ex : âge d'un individu, nombre d'enfants d'un ménage)

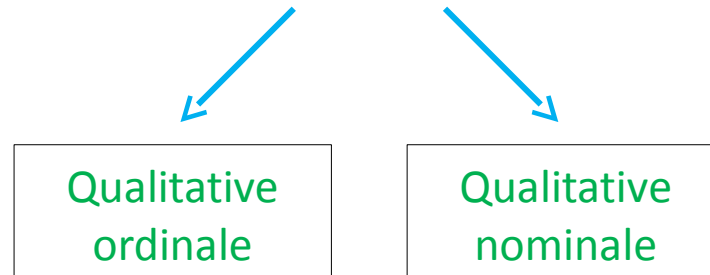
Toutes valeurs possibles dans intervalle de variation (ex : taille d'un individu)

⇒ Cette distinction est à relativiser car toute mesure est limitée en précision. On mesure généralement la taille d'un individu au mieux en centimètres, le nombre de valeurs possibles est donc limité.

Exemple: étude d'une région

➤ TYPES DE VARIABLE

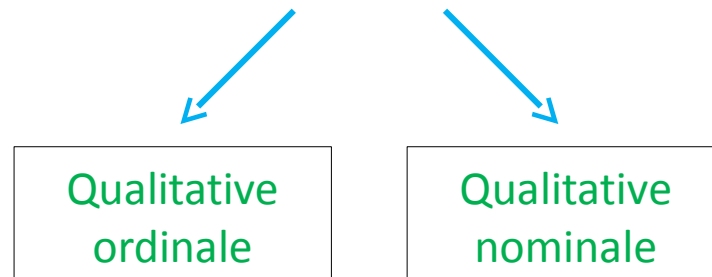
- ✓ Variable **qualitative** : traduit une qualité. Opération de catégorisation



Exemple: étude d'une région

➤ TYPES DE VARIABLE

- ✓ Variable **qualitative** : traduit une qualité. Opération de catégorisation

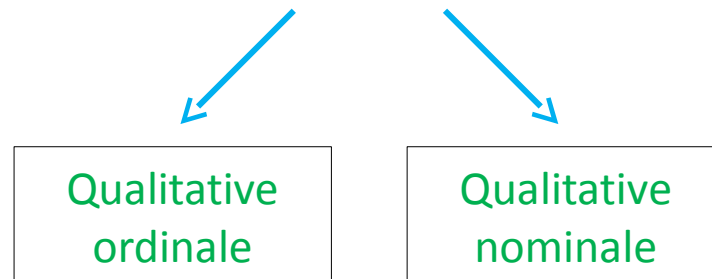


Valeurs ordonnées dans un classement hiérarchique (ex : le rang des élèves dans une classe)

Exemple: étude d'une région

➤ TYPES DE VARIABLE

- ✓ Variable **qualitative** : traduit une qualité. Opération de catégorisation



Valeurs ordonnées dans un classement hiérarchique (ex : le rang des élèves dans une classe)

Valeurs sans aucune hiérarchie (ex: la couleur des yeux d'un individu)

Exemple: étude d'une région

➤ TYPES DE VARIABLE

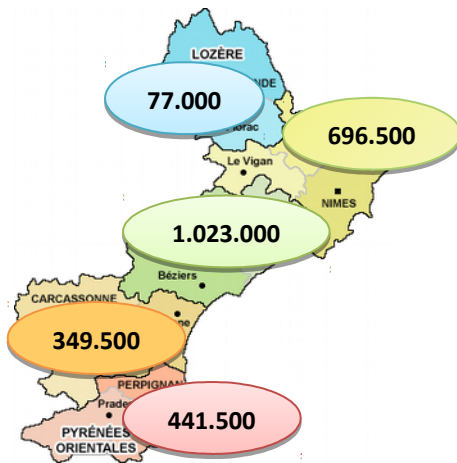


Fig. 1 : Nombre d'habitants par départements en 2008, INSEE

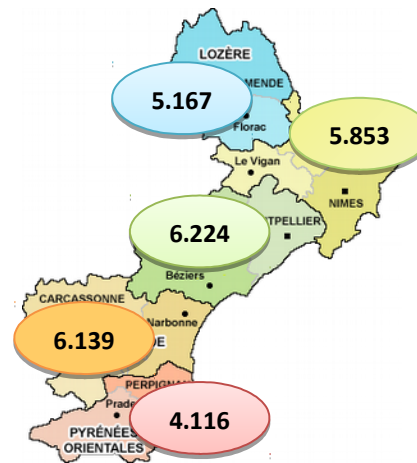


Fig. 2 : Superficie en km², INSEE

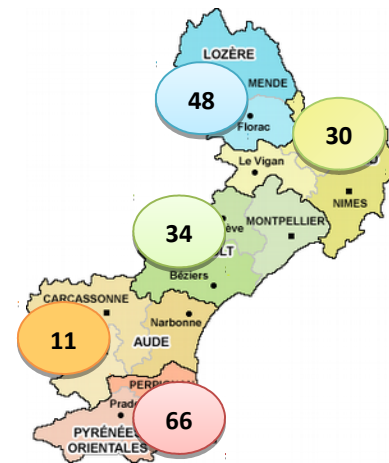


Fig. 3 : Code géographique

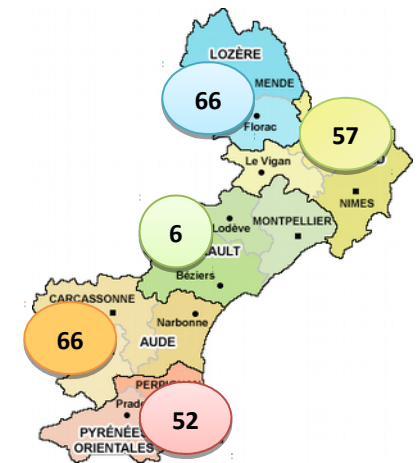


Fig. 4: Classement général des départements où il fait bon vivre en 2008, L'Express

Exemple: étude d'une région

➤ TYPES DE VARIABLE

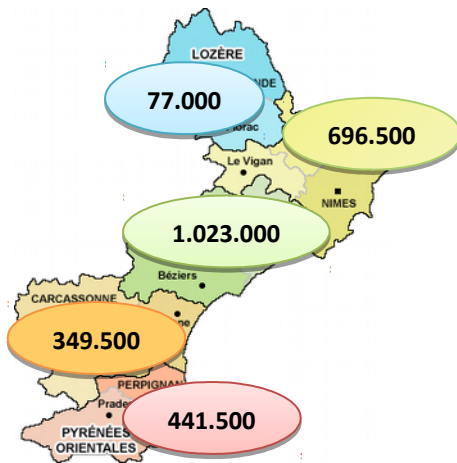


Fig. 1 : Nombre d'habitants par départements en 2008, INSEE

QUANTITATIVE
DISCRÈTE

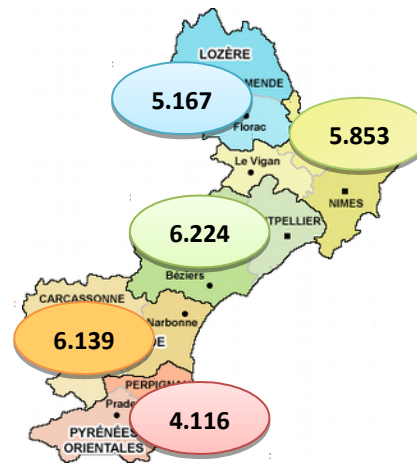


Fig. 2 : Superficie en km², INSEE

QUANTITATIVE
CONTINUE

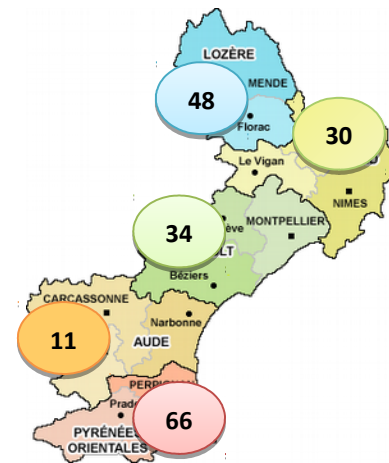


Fig. 3 : Code géographique

QUALITATIVE
NOMINALE

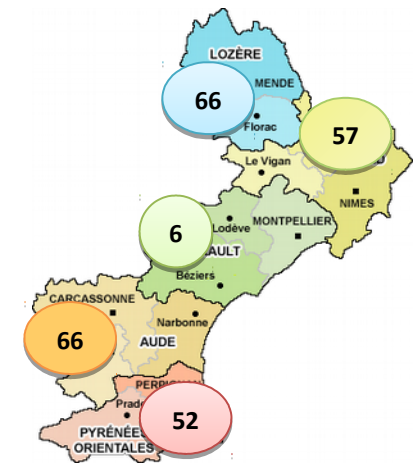
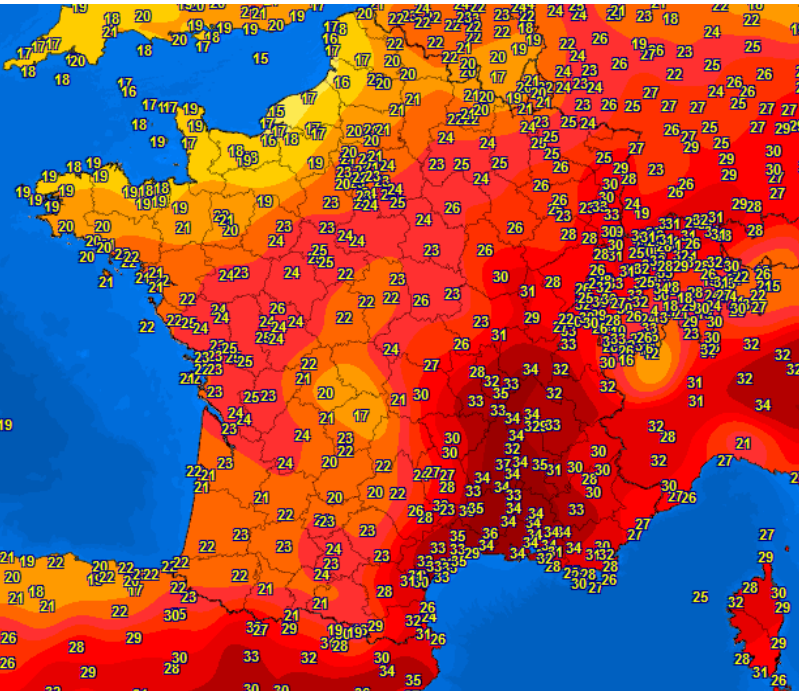


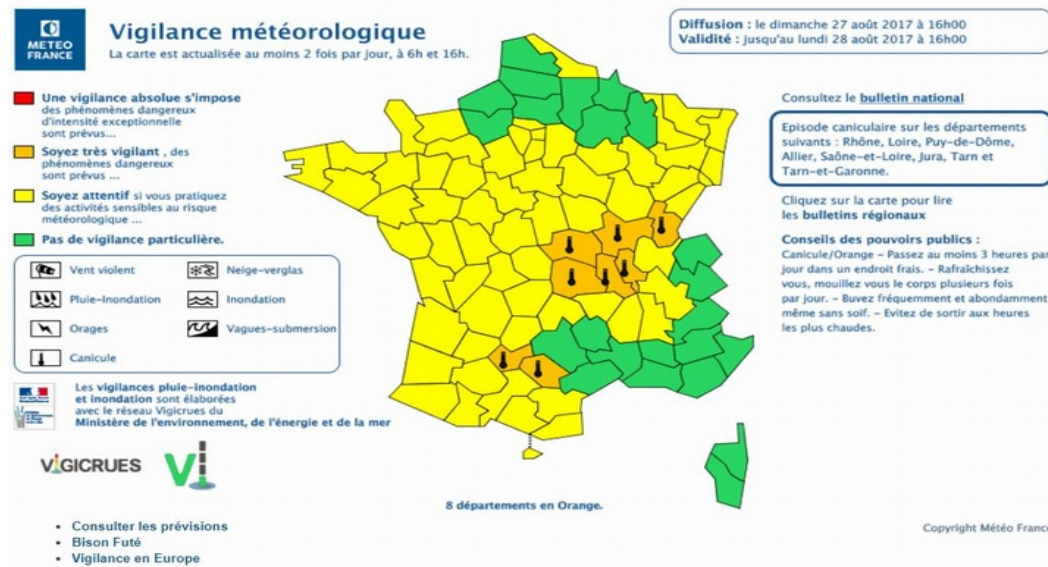
Fig. 4: Classement général des départements où il fait bon vivre en 2008, L'Express

QUALITATIVE
ORDINALE

➤ TYPES DE VARIABLE



PSEUDO
QUANTITATIVE
CONTINUE



QUALITATIVE
ORDINALE

L'ensemble des départements de la région Languedoc-Roussillon
Population

