

Faculté des Lettres, des Arts et des Humanités de la Manouba

Département de Géographie

1^{er} année Licence de Géomatique

Module cartographie 2

Découper en classes les séries
quantitatives

Kamel Nasraoui

Année universitaire : 2024 - 2025

Le langage statistique

Une série statistique : est un ensemble dans lequel, à chaque individu d'une population ou d'un échantillon, correspond la valeur du caractère.

Une population : est un ensemble fini et théoriquement dénombrable d'êtres animés, d'objets et de concepts. C'est l'ensemble de référence sur lequel porte l'analyse statistique. Il doit être homogène et défini avec la plus grande rigueur. Il est composé d'éléments liés par une ou plusieurs propriétés.

Un individu : l'individu, ou unité statistique, est un élément de l'ensemble de référence ou de la population.

Échantillon : un échantillon est un sous-ensemble prélevé dans une population.

Le langage statistique

Un caractère : un caractère (ou propriété, critère) est un thème commun à tous les individus d'une population : âge d'une population, débit annuel d'un fleuve etc.

Un effectif : est le nombre d'individus d'une modalité ; l'effectif total est l'ensemble des individus d'une population.

Une distribution : une série devient une distribution lorsque les valeurs prises dans les différentes modalités sont ordonnées, classées.

Il existe deux types de séries ou distribution : chronologique et des distributions qui décrivent un moment donné.

Qu'est ce qu'une discrétisation?

Le découpage en classes d'une série statistique est appelé discrétisation dans la mesure où elle consiste à rendre discrète, c'est-à-dire discontinue, une série mesurée d'abord sur une échelle continue de valeur. La discrétisation doit permettre de conserver au mieux l'information géographique continue dans la série statistique, tout en permettant sa transmission par carte avec la meilleure lisibilité possible.

Du point de vue des règles de perception visuelle, la lisibilité de l'image cartographique est plus grande que le nombre de signes distincts, donc de classes est petit. Donc il est donc indispensable de procéder à une réduction de l'information statistique . La simplification de l'information a pour but de la rendre intelligible, c'est-à-dire de la mettre en classes.

L'ordre de grandeur d'une série statistique

Le mode est la valeur la plus fréquente dans une série statistique. Il est en pratique peu utilisé en géographie. Son utilité se limite au cas des séries statistiques discrètes, où le nombre des unités géographiques observées est très grand.

La médiane est la valeur telle que 50% des observations aient une valeur inférieure à elle et 50 % des observations aient une valeur supérieure à elle. Elle partage la série ordonnée en deux classes d'effectifs égaux.

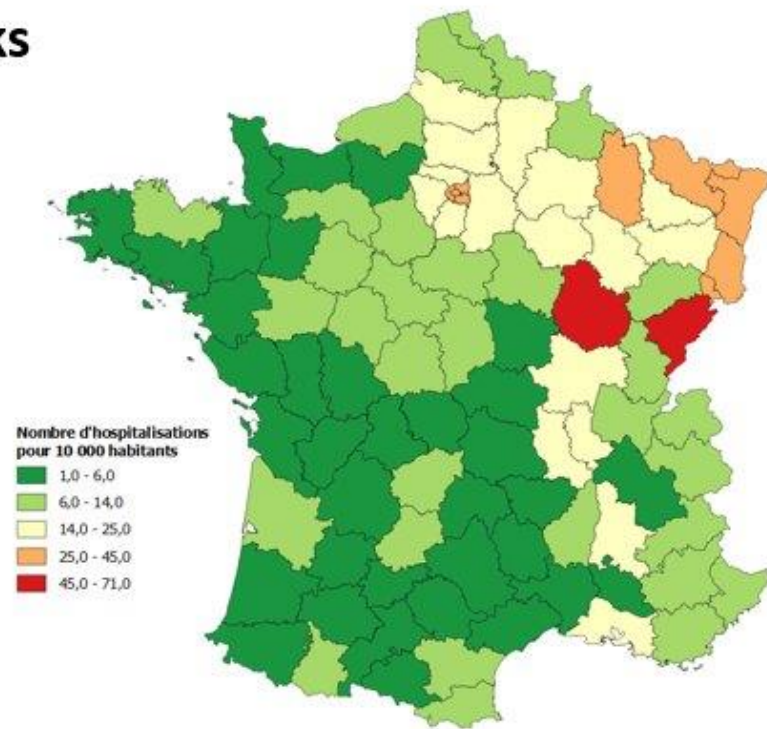
La moyenne est la valeur centrale la plus utilisée pour résumer une distribution. On l'obtient en effectuant la somme de toutes les valeurs de la série statistique, et en divisant ce total par le nombre d'observations.

L'enjeu de discrétisation

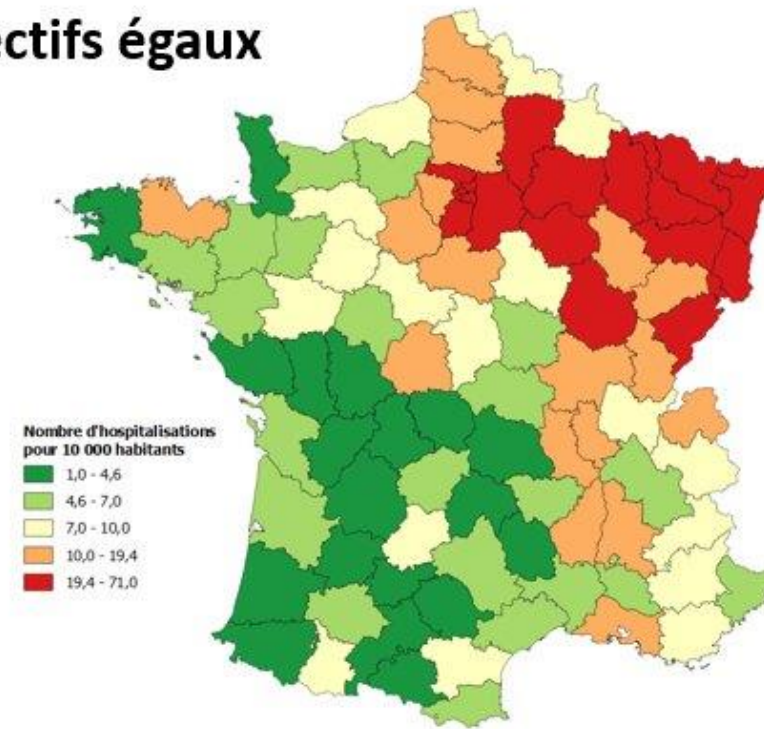
Combiner deux critères contradictoires

- Regrouper les valeurs qui se ressemblent et qui sont différentes des autres.
- Conserver l'ordre de grandeur ; la dispersion et la forme de la distribution.
- Limiter le nombre de classes (6 à 8 classes).
- Donner des seuils qui ont du sens (exemple : moyenne nationale).

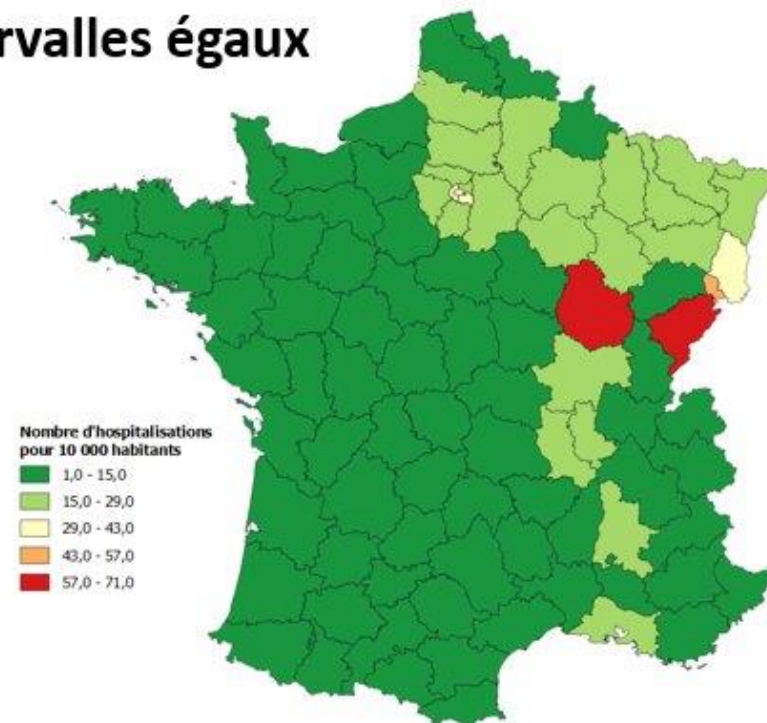
Jenks



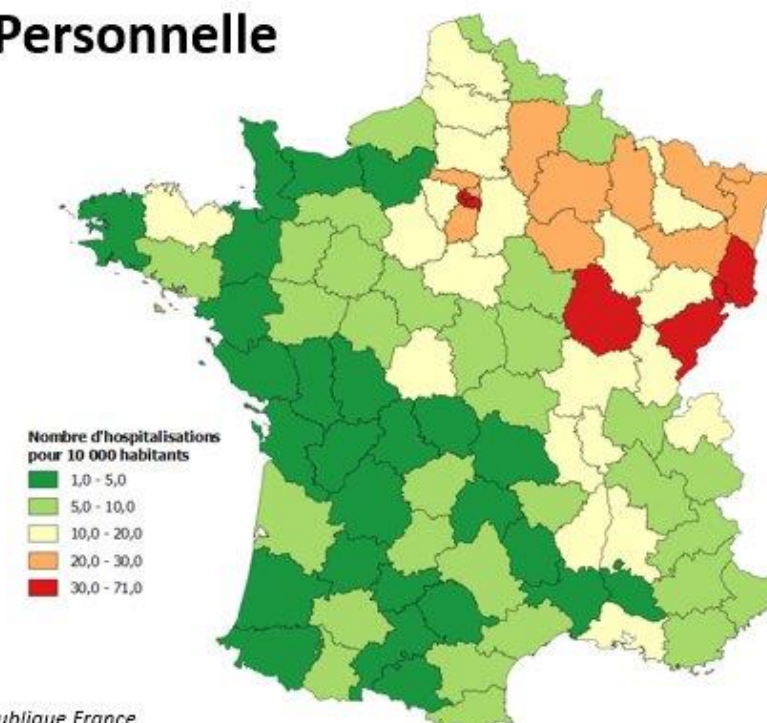
Effectifs égaux

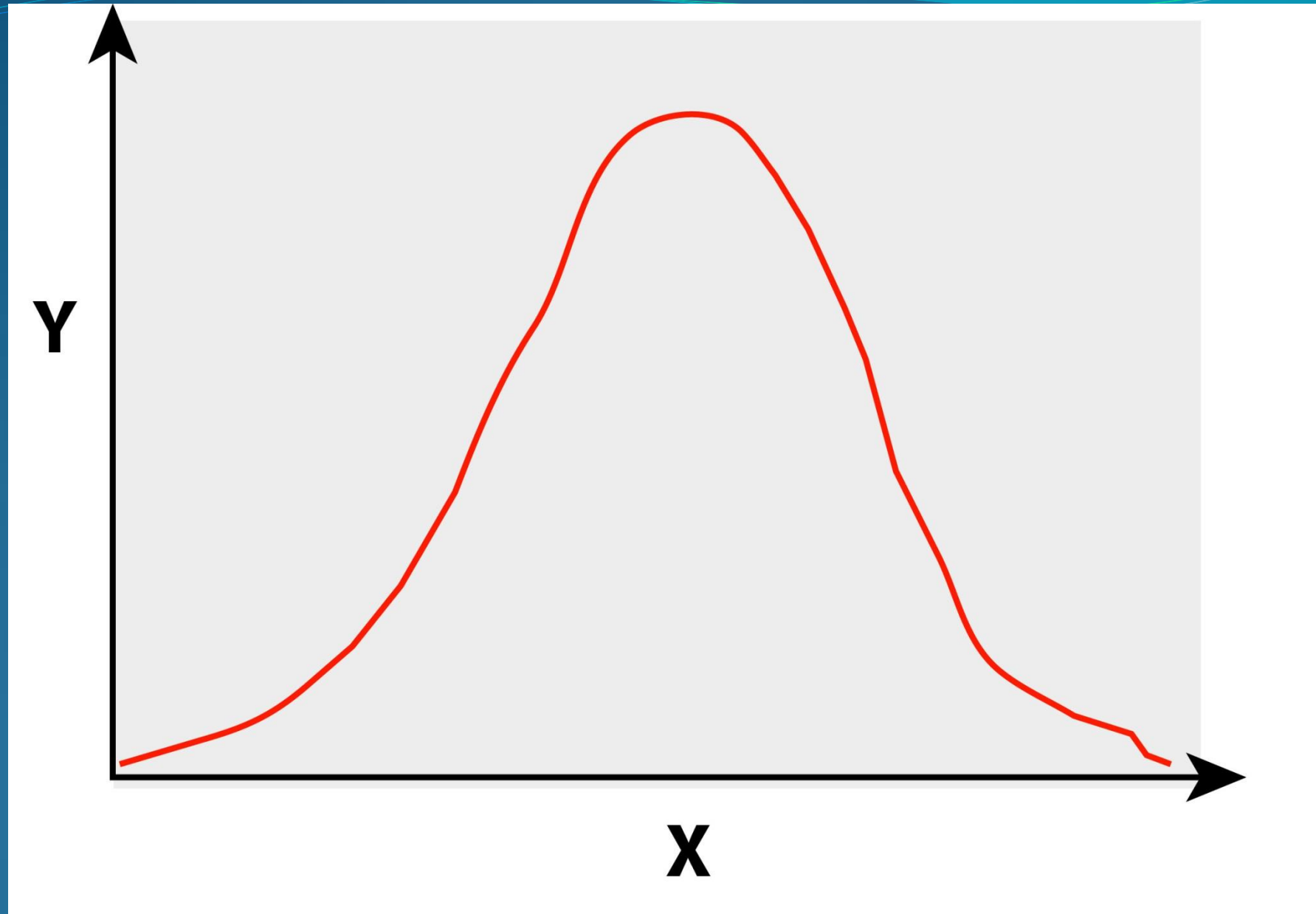


Intervalles égaux

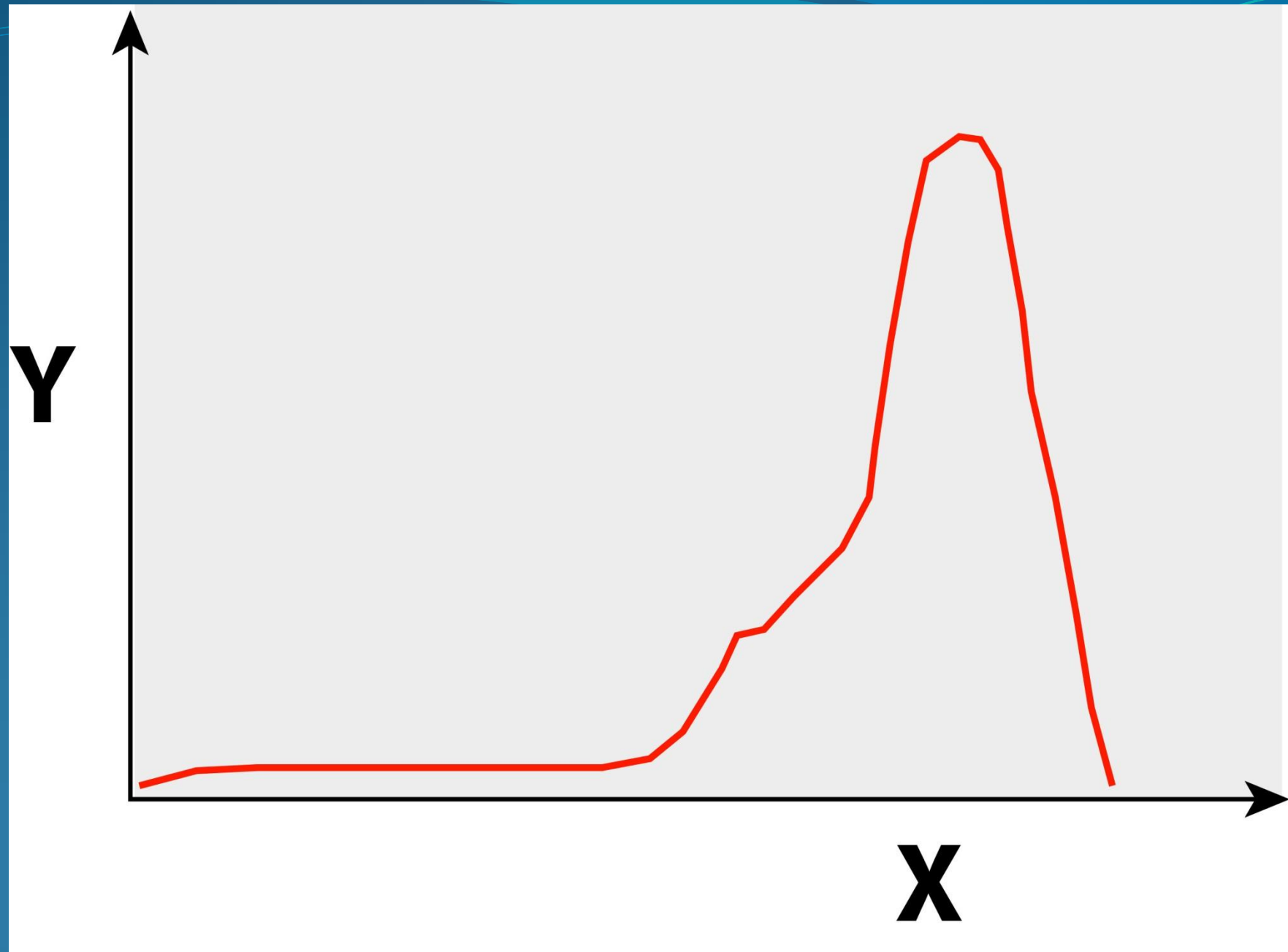


Personnelle

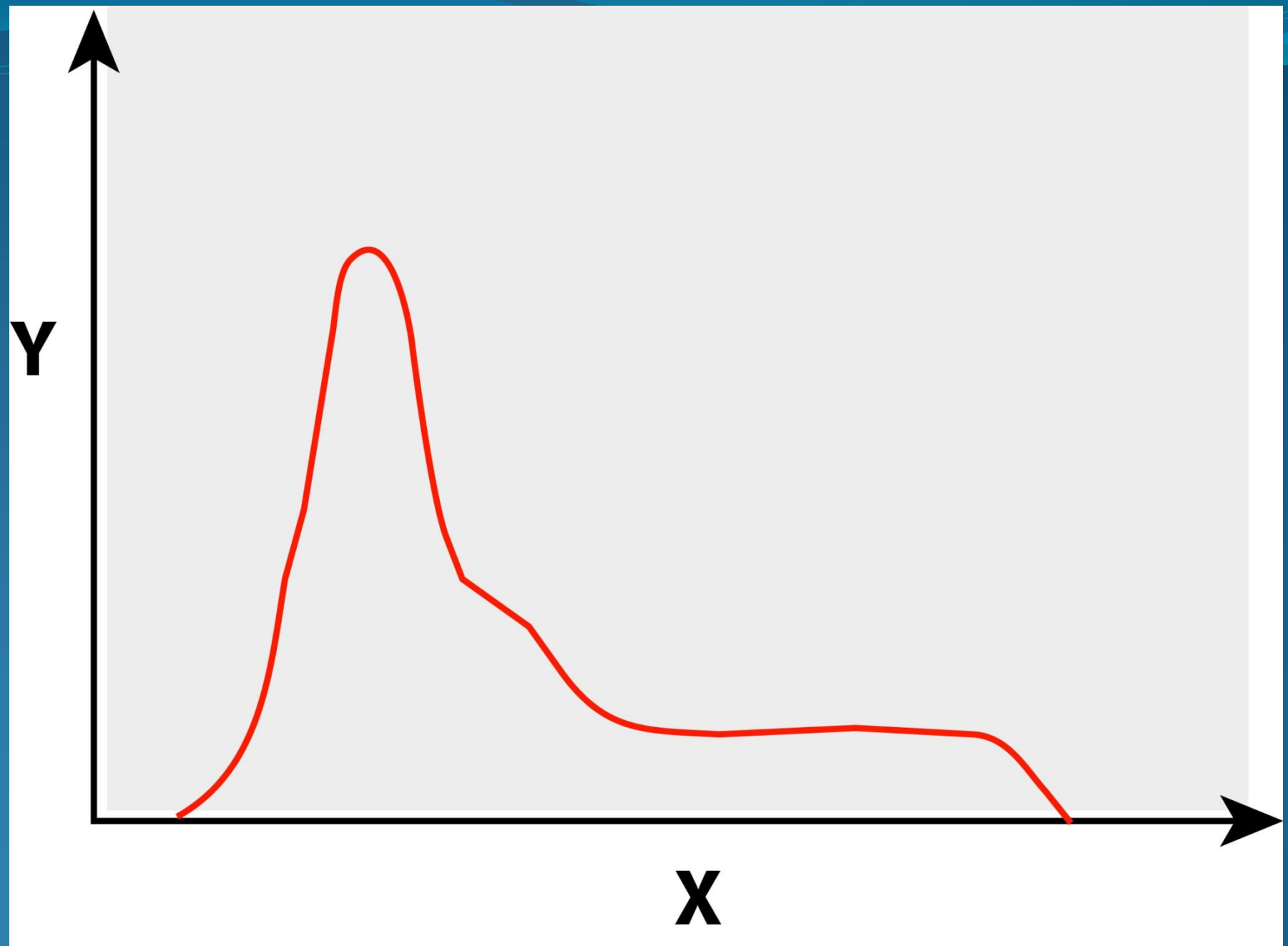




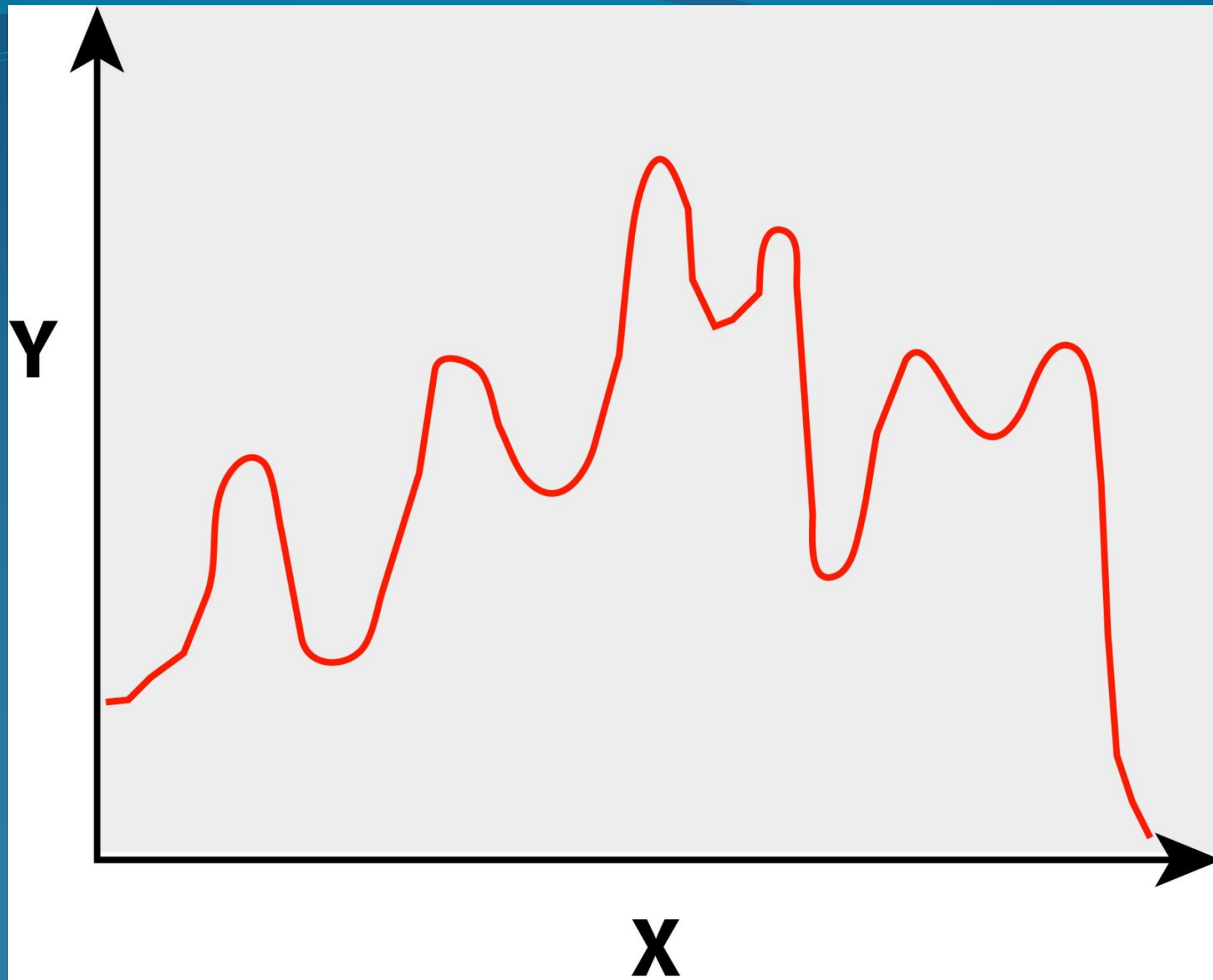
Si la médiane égale à la moyenne = distribution symétrique



Si la médiane est plus forte que la
moyenne (dissymétrie à droite)



Si la médiane est plus faible que la
moyenne (dissymétrique à gauche)



Distribution multimodale

Le choix du nombre de classes

La discrétisation soulève souvent la question du nombre de classes à créer. L'objectif étant de perdre le moins d'information possible tout en retranscrivant au mieux l'organisation statistique des données. De plus, au-delà de l'aspect purement statistique, la discrétisation doit aussi prendre en compte les règles de la perception visuelle afin de transmettre une information géographique efficace et de qualité. Le seuil de la perception visuelle maximale a été défini à huit classes mais ce chiffre peut varier en fonction de la configuration spatiale. L'inconvénient de ces formules repose sur l'hypothèse d'une distribution normale de la série statistique. En effet, si la distribution est asymétrique le nombre de classes ne sera pas optimal.

Diverses formules mathématiques permettent de donner une indication du nombre de classes idéal en fonction de la population de la série statistique :

Formule de Brooks-Carruthers : $k = 5 \times \log (N)$.

Formule de Sturges-Huntsberger : $k = 1 + 3,3 \times \log (N)$.

Le choix du nombre de classes

Population	Brooks-Carruthers	Sturges-Huntsberger	Perception visuelle	Nombre de classe idéal
10	5	4,33	8	4-5
25	6,99	5,66	8	5-6-7
50	8,49	6,66	8	6-7-8
100 (et +)	10 (et +)	7,67 (et +)	8	7-8

Tab : Estimation du nombre de classes

<https://magrit.hypotheses.org>

Les méthodes de discrétisation

Gouvernorat	Taux d'urbanisation (%)
Tunis	100
Ariana	90,8
Ben Arous	90,3
Manouba	74
Nabeul	65,9
Zaghouan	37,5
Bizerte	60,9
Beja	40,4
Jandouba	27,3
Le Kef	49,7
Siliana	36
Kairouan	31,8
Kasserine	39,8
Sidi Bouzid	24,3
Monastir	100
Sousse	79,4
Mahdia	44,4
Sfax	63,7
Gafsa	73,2
Tozeur	70,3
Kebili	53,9
Gabès	67,7
Médenine	77,1
Tataouine	61,1

Tab : Le taux d'urbanisation en Tunisie en 2004

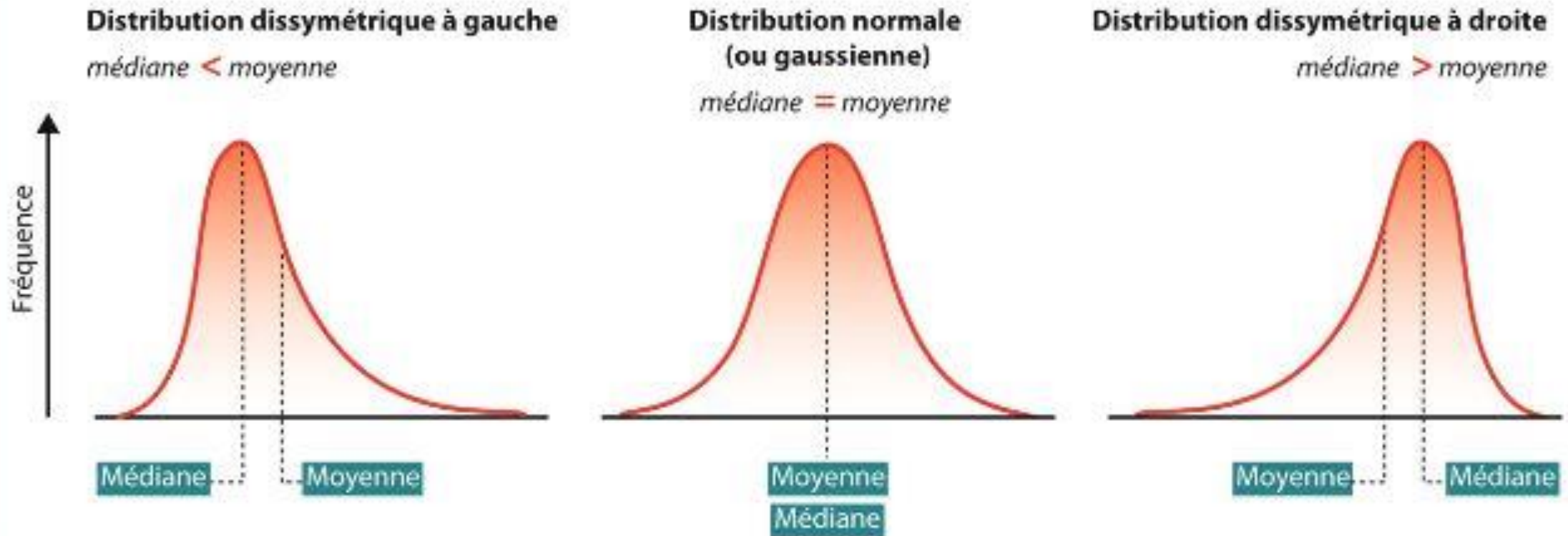
Source : Institut National de la Statistique.

Moyenne de la série statistique = 60,8

Médiane de la série statistique = 62,4

Les méthodes de discrétisation

Déterminer la forme de la distribution avec les paramètres de position

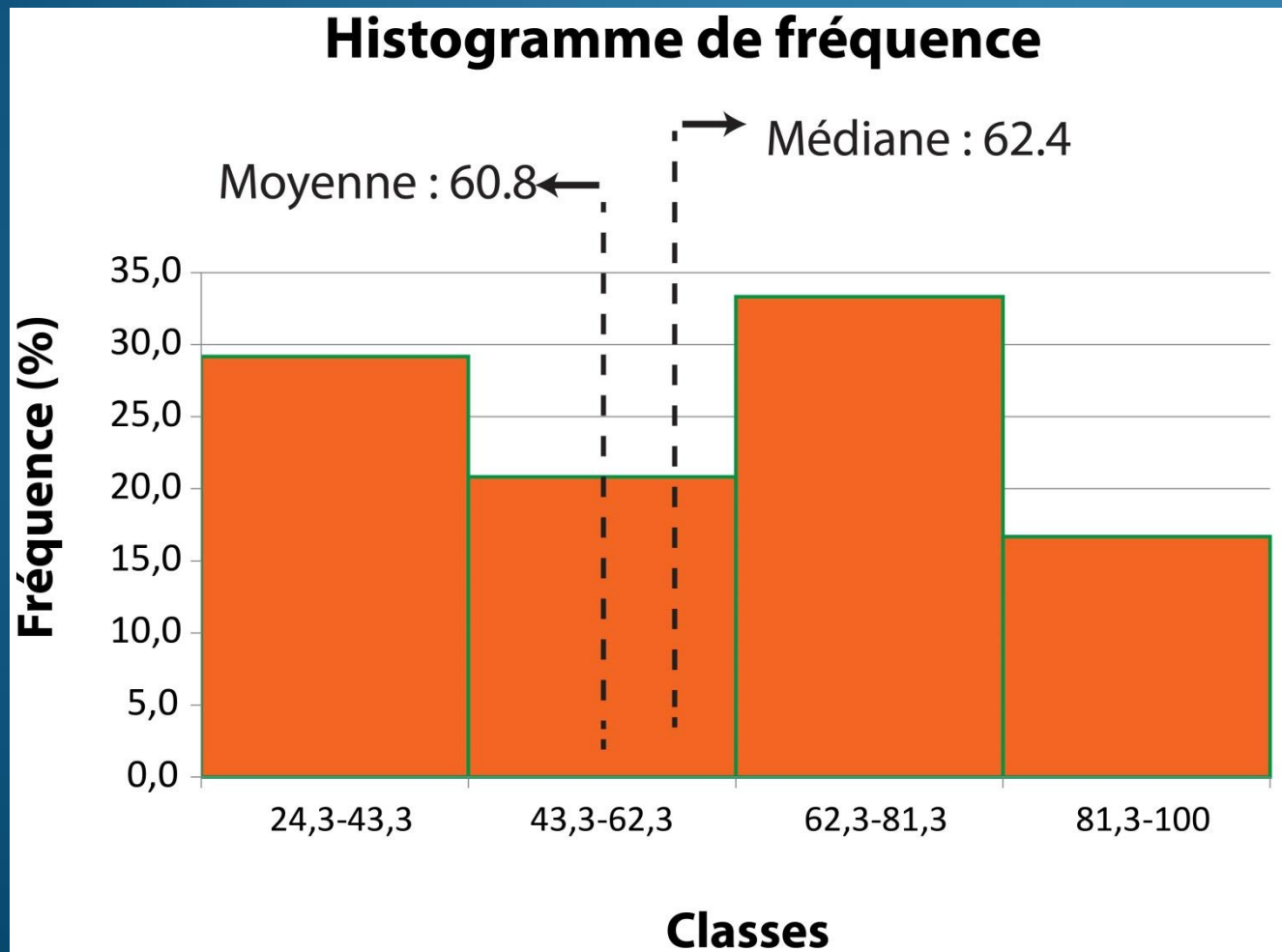


LAMBERT & ZANIN, 2016

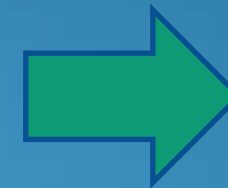
Manuel de cartographie : principes, méthodes et applications, N. Lambert et Ch. Zanin, 2016.

Comment choisir une bonne discrétisation (déterminer la forme de la distribution)?

classes	Effectifs	Fréquence (%)	Fréquence cumulée(%)
24,3-43,3	7	29,2	29,2
43,3-62,3	5	20,8	50,0
62,3-81,3	8	33,3	83,4
81,3-100	4	16,7	100,0



Médiane > Moyenne



Une série
dissymétrique à
droite

Tab : Le taux d'urbanisation en
Tunisie en 2004

Source : Institut National de la
Statistique.

Le nombre de classes

Le nombre de classes :

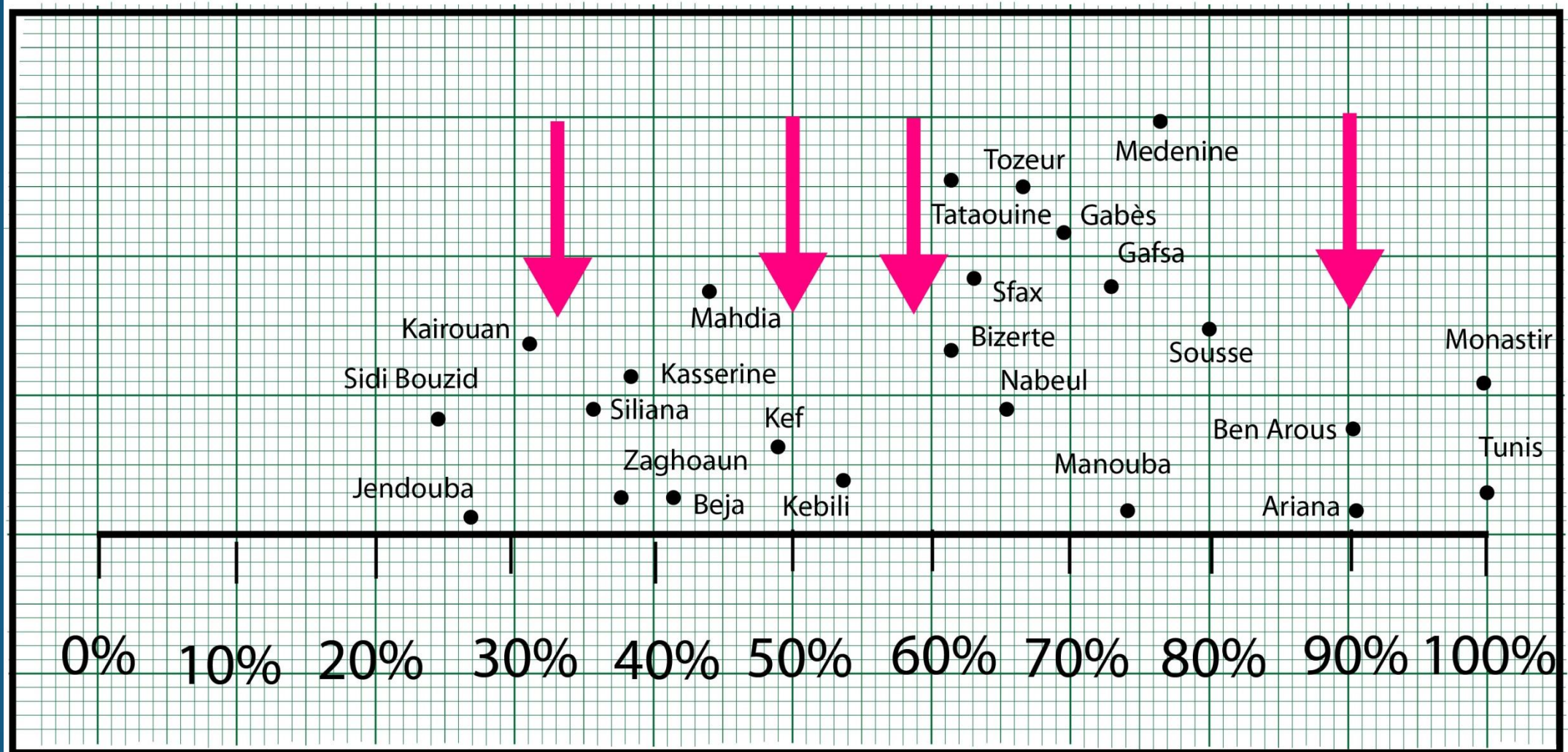
Formule1 : $1 + 3,3 * \log(24) = 5,55$: formule de Sturges Huntsberger

Formule2 : $5 * \log(24) = 6,9$: formule de Brooks-Carruthers

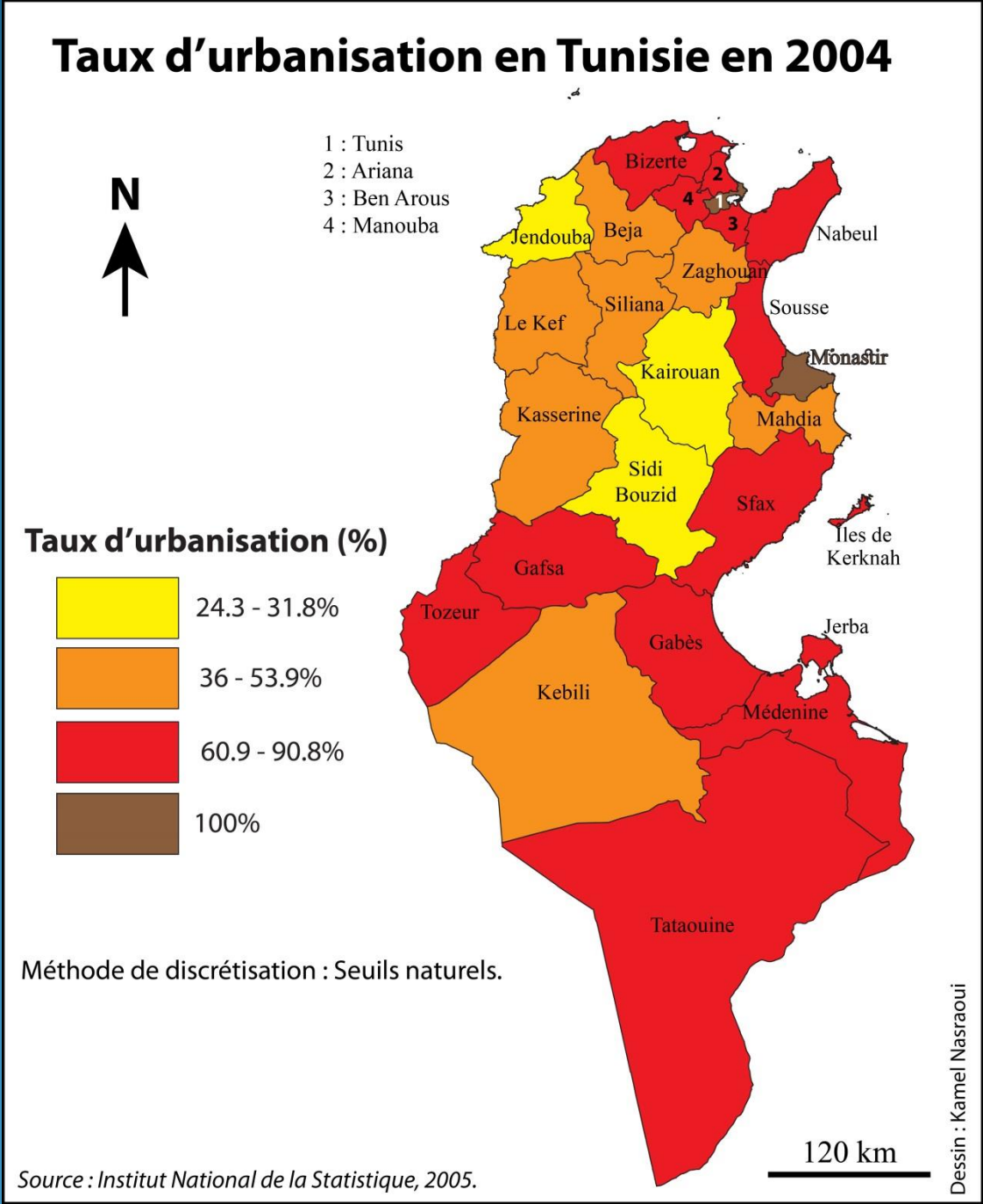
Pour une population de 24 gouvernorats, le nombre de classes varie entre 5 et 7.

La méthode des seuils naturels ou observés

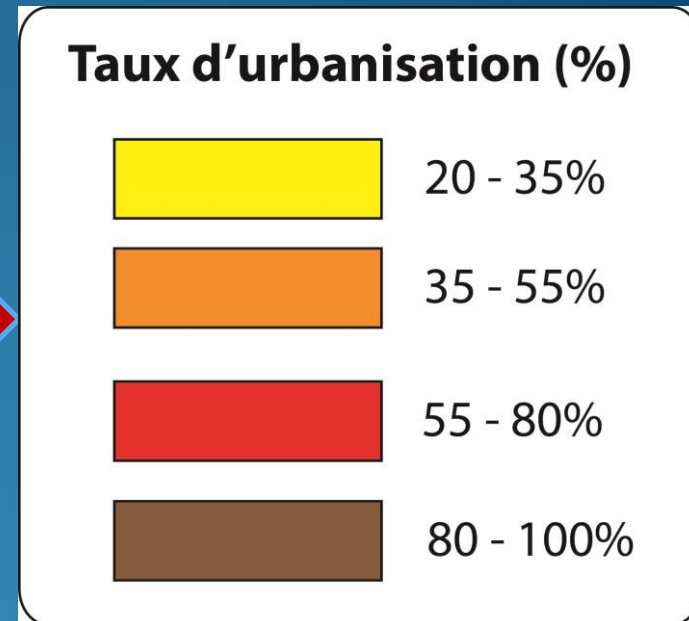
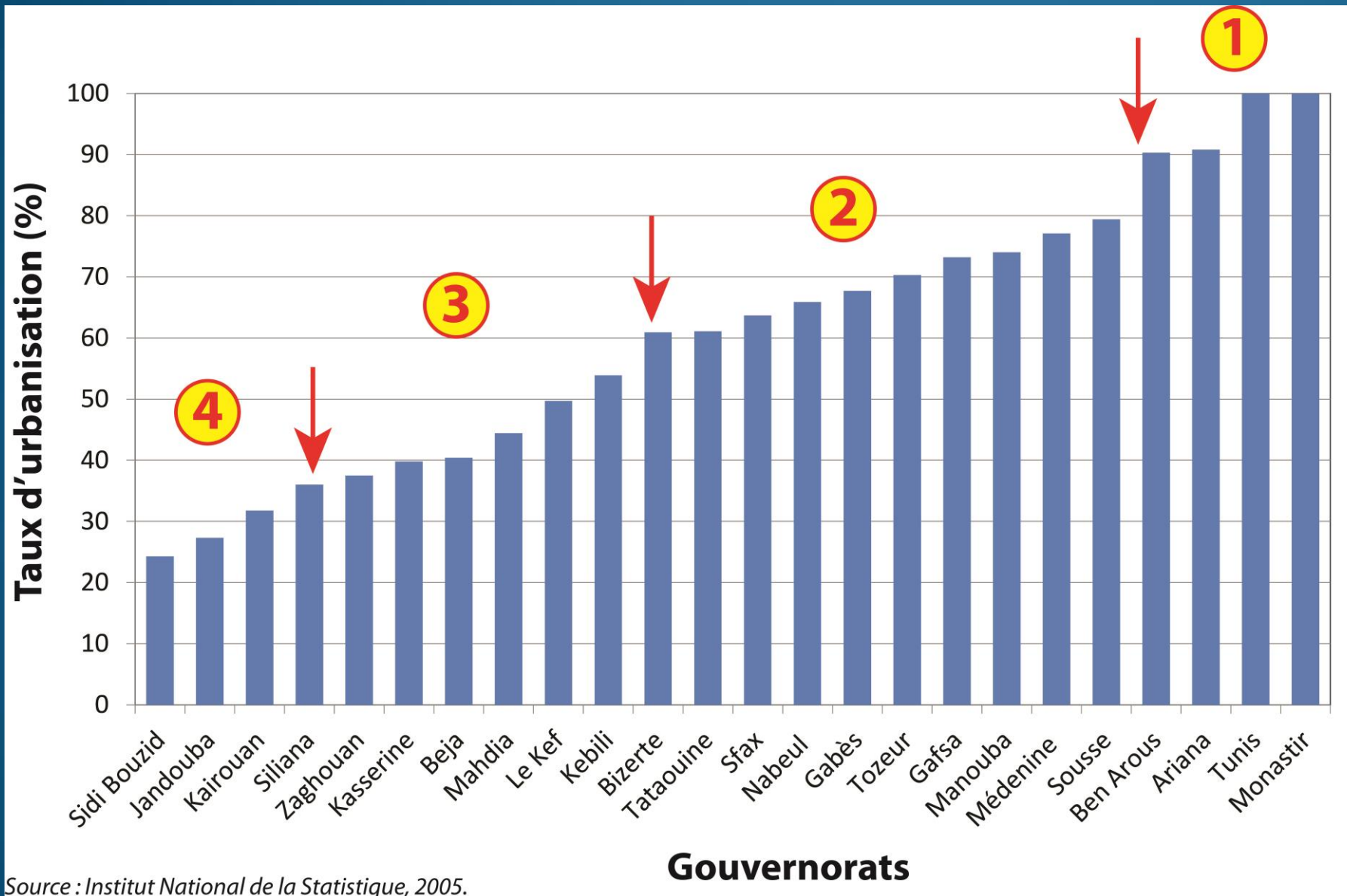
Méthode des seuils naturels



La méthode des seuils naturels ou observés



La méthode d'histogramme ordonné



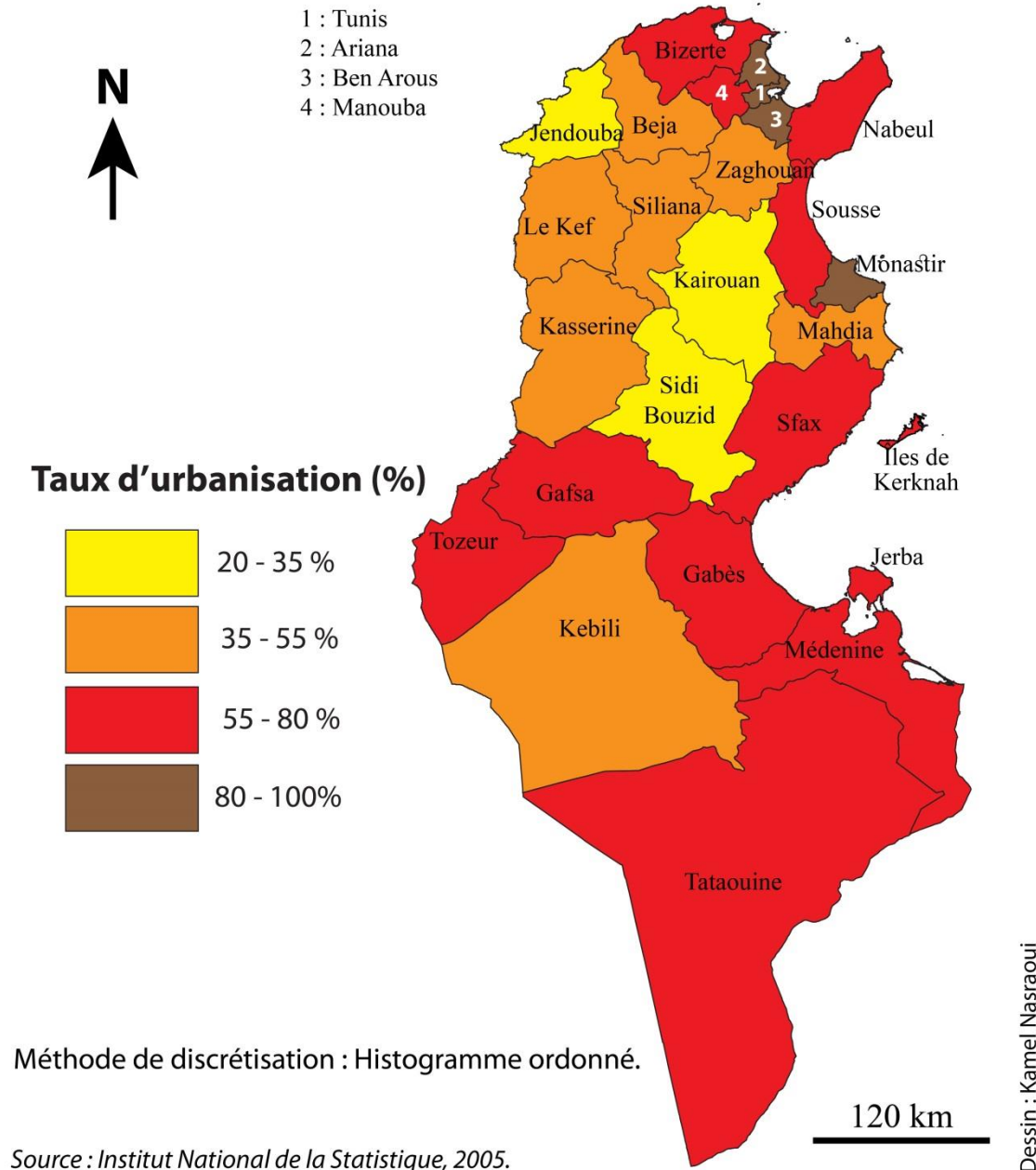
Source : Institut National de la Statistique, 2005.

La méthode des seuils naturels ou observés

- Représenter graphiquement la distribution et repérer les creux et les bosses, chaque creux marquera la limite d'une classe.
- Cette méthode permet de focaliser sur les discontinuités de la série statistique.
- Les effectifs des classes sont inégaux et le découpage est subjectif.

La méthode histogramme ordonné

Taux d'urbanisation en Tunisie en 2004



La méthode à amplitudes égales

Borne minimale = Sidi Bouzid = 24,3

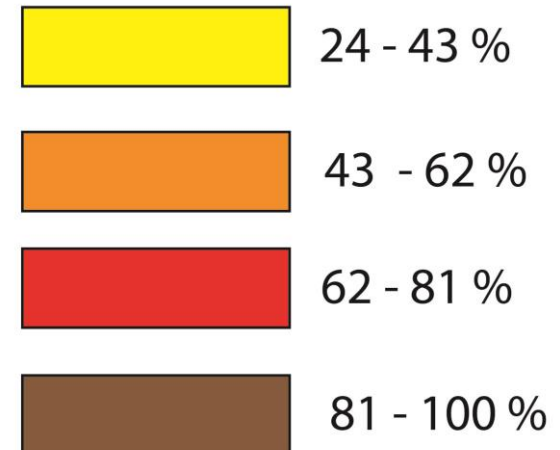
Borne maximale = Tunis et Monastir = 100

L'étendue de la série statistique : $100 - 24,3 = 75,7$

L'amplitude sur 4 classes : $75,7 / 4 = 18,92 \approx 19$

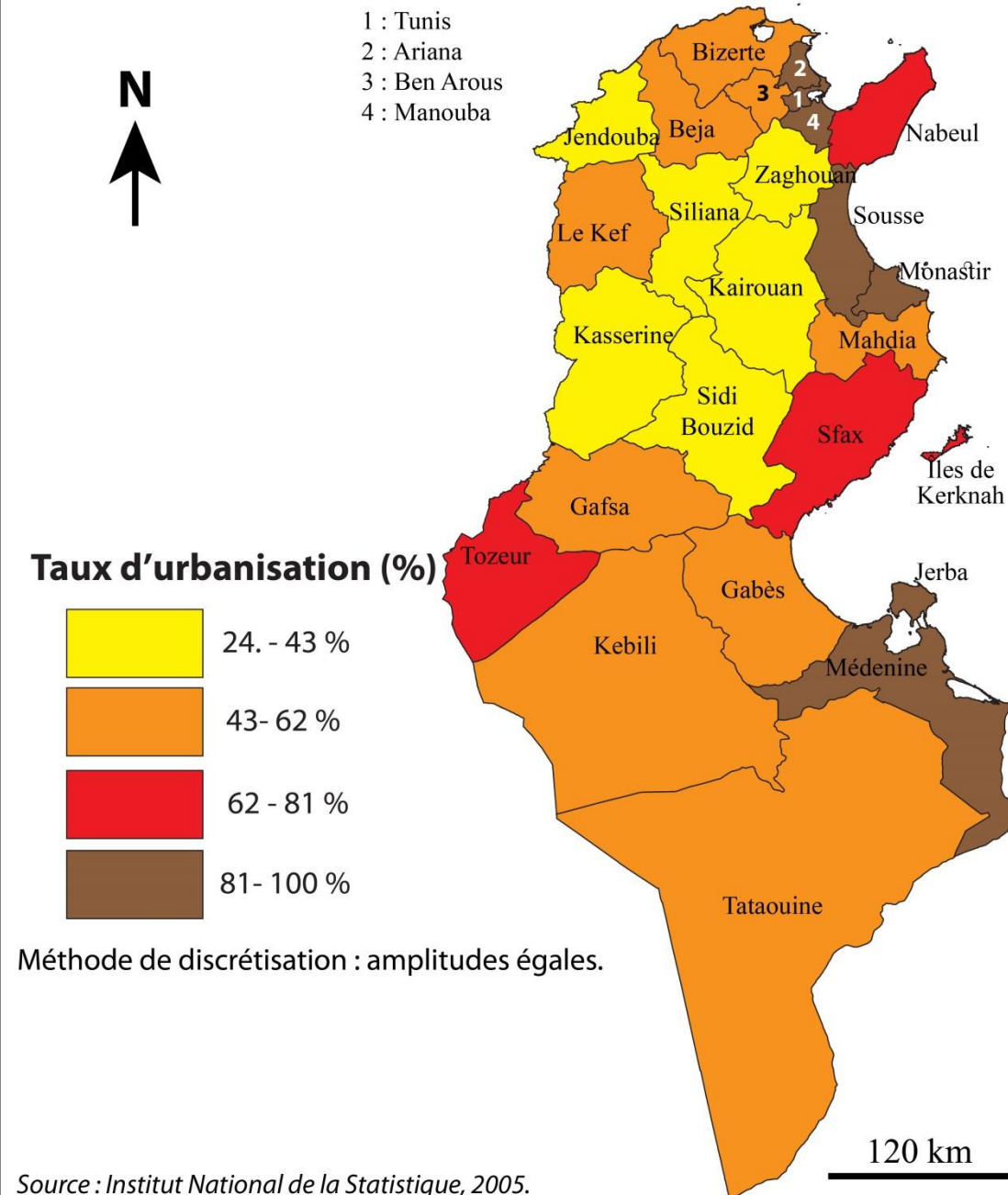
- On l'utilise pour les distributions uniformes et symétriques.
- Cette méthode ne permet pas de comparer plusieurs cartes.
- Il peut y avoir des classes vides.
- À éviter pour les distributions dissymétriques.

Taux d'urbanisation (%)



La méthode à amplitudes égales

Taux d'urbanisation en Tunisie en 2004



Source : Institut National de la Statistique, 2005.

Méthode à effectifs égaux (EF)

On décide : 4 classes donc, $24 / 4 = 6$ gouvernorats

Classement par ordre croissant

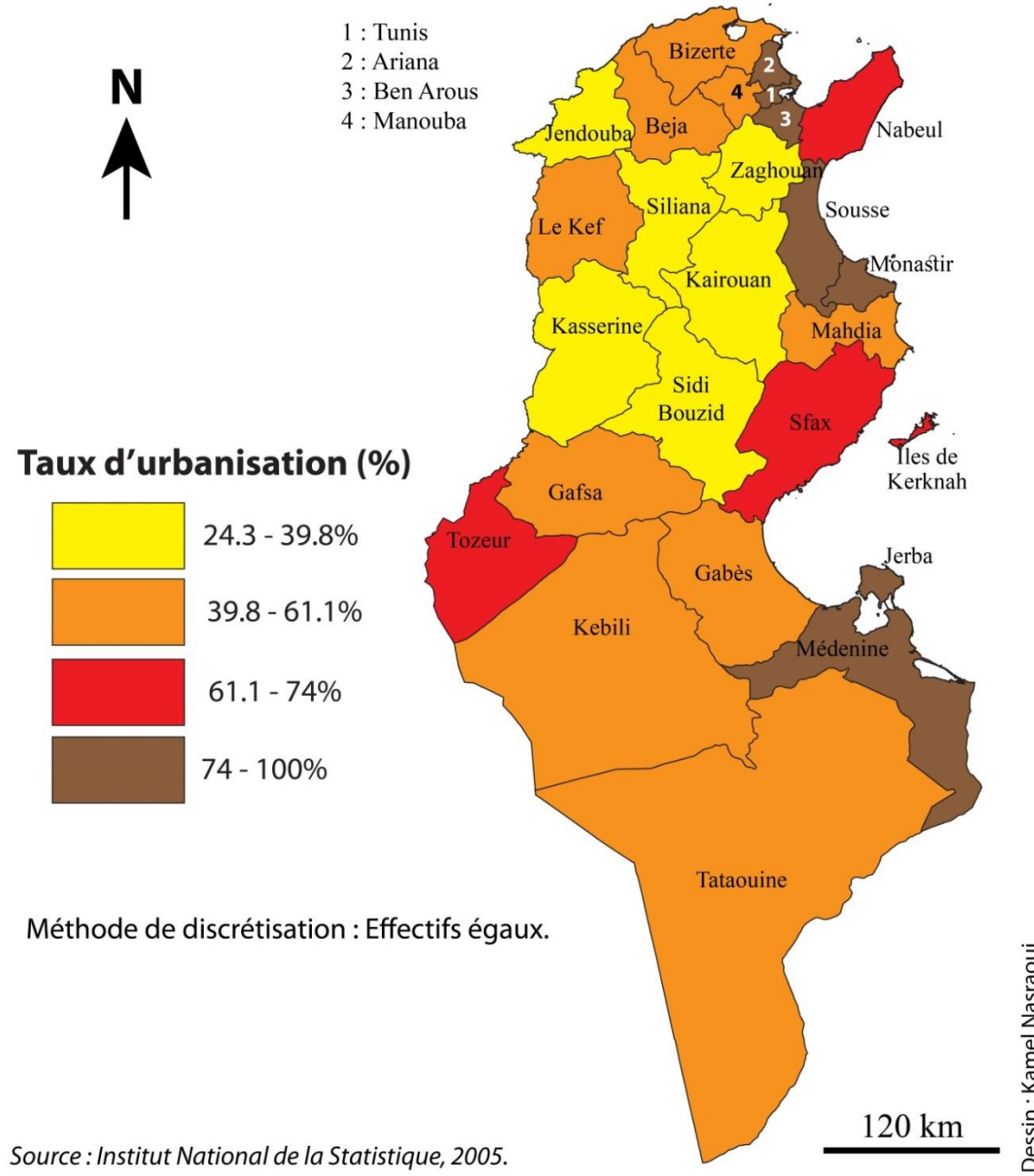
Sidi Bouzid	Jendouba	Kairouan	Zaghouan	Kasserine	Siliana	I
Beja	Mahdia	Kef	Kebili	Bizerte	Tataouine	II
Sfax	Nabeul	Tozeur	Gabès	Gafsa	Manouba	III
Médinine	Sousse	Ben Arous	Ariana	Monastir	Tunis	IV

Méthode de quartiles : puisque on a 4 classes.

Elle peut être utilisée avec n'importe quelle forme de distribution . Elle donne des résultats satisfaisantes au niveau cartographique. Elle est basée sur le classement des individus et non sur les valeurs.

Méthode à effectifs égaux

Taux d'urbanisation en Tunisie en 2004

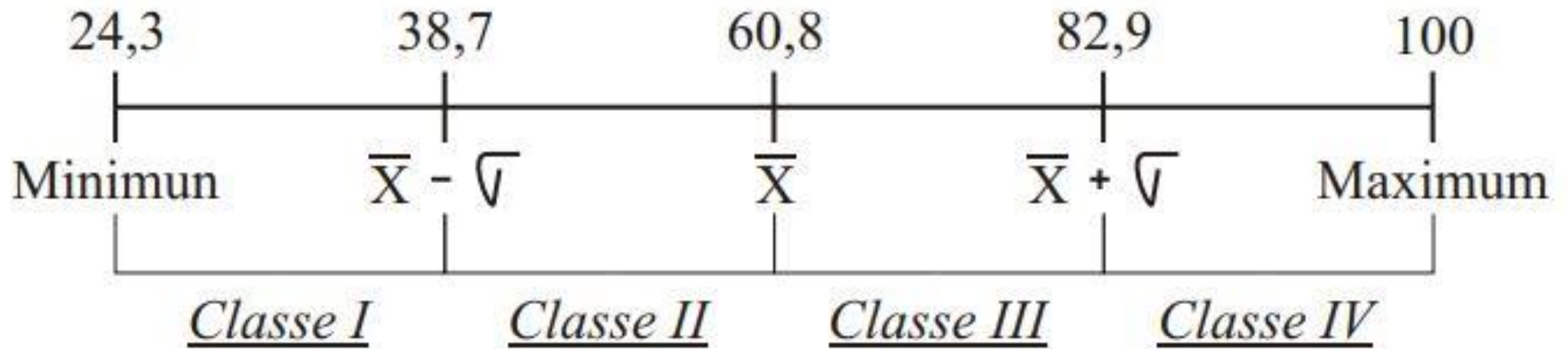


Méthode par l'utilisation de la moyenne arithmétique ($\bar{X}$) et l'écart-type

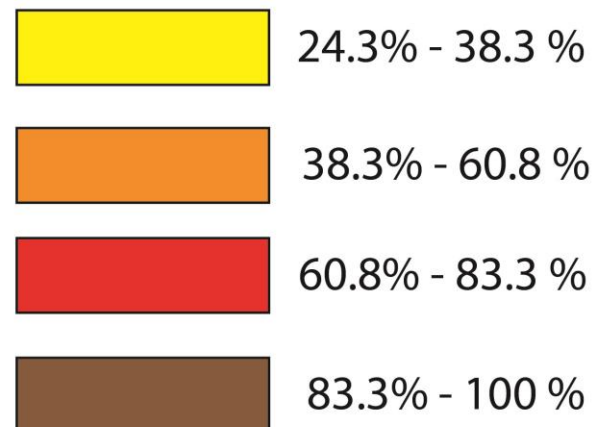
Moyenne arithmétique = $(\bar{X}) \frac{\sum ni}{N} = 60,8$

Écart-type = L'écart -type ($\sqrt{\quad}$) $\sqrt{\frac{\sum (Xi - \bar{X})^2}{N}} = 22,5$

Méthode par l'utilisation de la moyenne arithmétique ($\bar{X}$) et l'écart -type

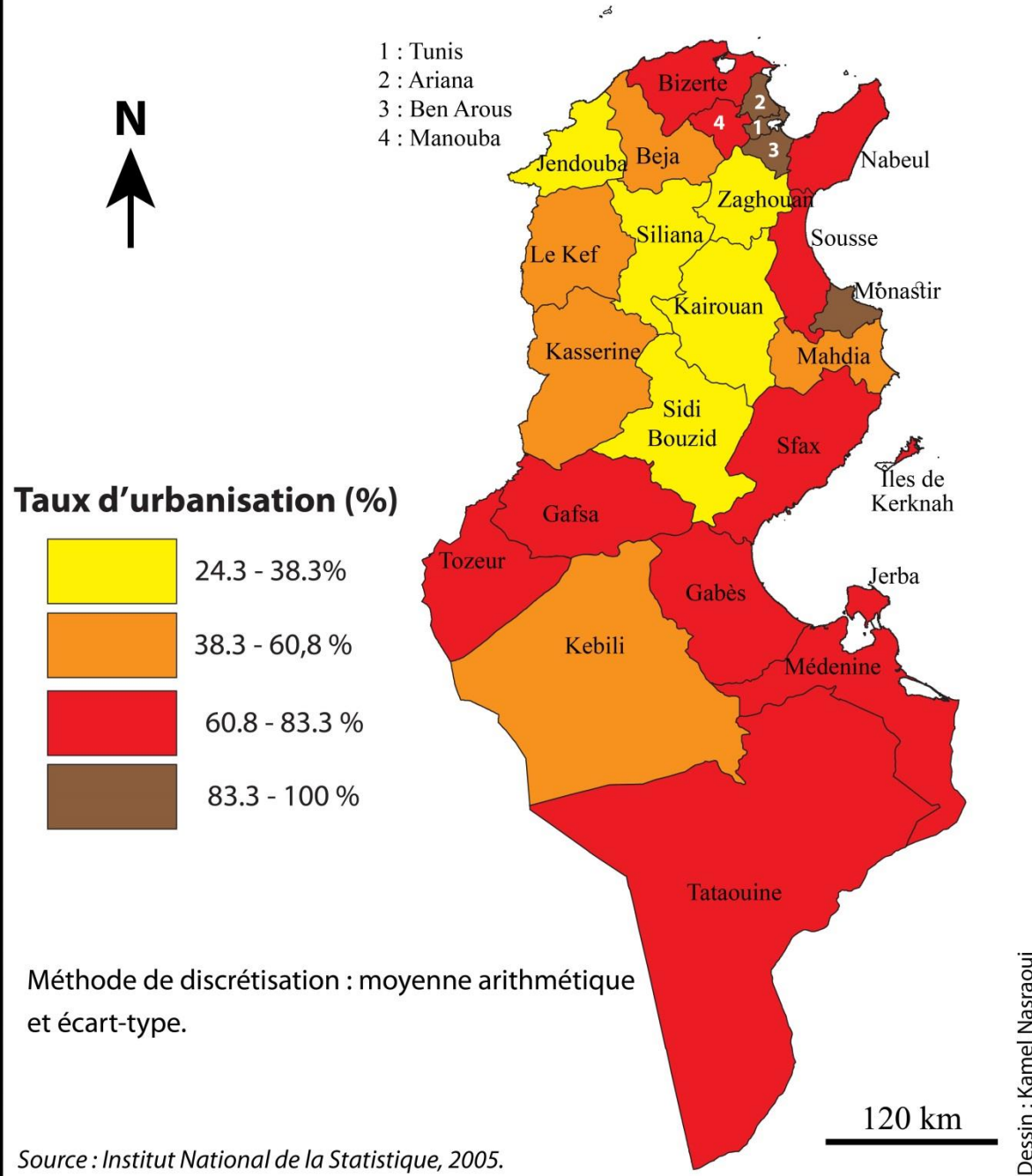


Taux d'urbanisation (%)



Méthode par l'utilisation de la moyenne arithmétique ($\bar{X}$) et l'écart -type

Taux d'urbanisation en Tunisie en 2004



- Cette méthode est la plus employée en cartographie. Elle est idéale pour les distributions normales (forme de cloche).
- Elle permet la comparaison des cartes.

Pour conclure

- Pour une distribution uniforme, on utilise la méthode d'amplitudes égales.
- Pour une distribution symétrique, on utilise la méthode d'amplitudes égales et la discrétisation standardisée basée sur la moyenne et l'écart-type.
- Pour une distribution dissymétrique, on utilise la méthode des effectifs égaux ou des seuils naturels.
- Pour une distribution bimodale, on utilise la méthode des effectifs égaux ou des seuils naturels.

La méthode de discrétisation d'une série statistique

Les classes seront identifiées en légende, à côté des caissons et des symboles qui représentent le figuré qui est leur est affecté sur la carte par leurs limites ou bornes de classes. Trois types de présentations sont possibles :

