

UN TABLEAU DE DONNEES : POUR QUOI FAIRE ? Révisions concernant les méthodes statistiques uni- et bivariées

**Myriam BARON, Christine ZANIN,
Université Paris 7 – Denis Diderot**

Type : TD

Niveau : 2^e année ou 3^e année de Licence

Durée : 1 séance de 3 heures

Thèmes : Formulation de questions de recherche, de problématiques
Enoncé d'une démarche pour y répondre
Géographie de la population
Etudes de cas
Statistiques univariées, bivariées
Cartographie thématique

Objectifs :

Cette séance de travaux dirigés propose de passer en revue les connaissances susceptibles d'être acquises en statistiques et en cartographie par un étudiant au cours des deux premières années du cursus de géographie (cf. Hors série Feuilles de Géographie intitulé *Itinéraire géographique en pays statistique*). On vérifie que les descriptions de base des tableaux élémentaires de données sont acquises et que la manipulation des outils statistiques et cartographiques est correcte. Enfin, le plus important : qu'avant toute démarche de calcul statistique et de représentation cartographique, les étudiants ont fait l'effort de regarder la nature de l'information dont ils disposaient et de formuler des questions auxquelles ils vont tenter de répondre. Ce travail vise à insister sur la relation qui doit exister entre questionnements, interrogations préalables et traitements statistiques et cartographiques : pour faire bref, les outils ne doivent pas être déconnectés d'une problématique, d'une thématique.

Cette séance peut être placée à la fin d'enseignements concernant l'acquisition des connaissances des statistiques uni- et bivariées et des représentations cartographiques et sera alors considérée comme une synthèse. Elle peut être également une introduction aux méthodes d'analyse des données, dont les apprentissages se font plutôt au cours de la troisième année. Dans ce dernier cas, on insistera davantage sur les articulations entre statistiques descriptives et analyse de données.

Déroulement :

La séance de travail est surtout basée sur la mobilisation des connaissances acquises par les étudiants et repose sur des commentaires de documents distribués au début de la séance. On travaille à partir d'un tableau élémentaire de données qui concerne les 51 Etats des Etats-Unis d'Amérique (la cartographie présente 49 Etats (51 moins Alaska et Hawaï), certains résultats n'étant pas disponibles pour ces deux Etats). On travaille surtout à partir d'un énoncé très détaillé, qui doit permettre aux étudiants de se repérer dans la démarche suivie, de réviser

dans de bonnes conditions mais aussi et surtout de reproduire le même cheminement lors de rencontres ou confrontations ultérieures avec d'autres tableaux de données statistiques.

Cet énoncé détaillé comporte trois temps qui correspondent à **trois démarches** :

- **le premier temps** est celui de la **familiarisation avec l'information communiquée**, de l'appropriation de cette information en formulant des questions plus ou moins générales et en proposant une ou des démarches pour pouvoir apporter des éléments de réponse.

- **le deuxième temps** est consacré **uniquement aux révisions concernant les méthodes de description et d'analyse des tableaux de données** – univariées et bivariées. Il vise surtout à bien vérifier que les étudiants sont capables de relier les résultats statistiques à des contenus thématiques, mais aussi de formuler clairement des conclusions à partir de ces premiers résultats.

- Enfin, **le troisième temps** est consacré aux **révisions concernant la représentation cartographique de données statistiques**. L'accent est mis sur les bonnes relations entre variables statistiques et variables visuelles et sur la mise en page de la carte, le soin extrême qui doit être porté à l'habillage du document cartographique.

UN TABLEAU DE DONNEES : POUR QUOI FAIRE ?

1. Réfléchir peut-être...

Documents :

- Fond de carte de localisation de 49 Etats américains
- Description du tableau élémentaire de données caractérisant les Etats américains

Sources : New-York Times Almanac, 2001, 2002; National Census 1980, 1990, 2000

L'occasion est trop belle de passer en revue le vocabulaire qui permet de décrire et donc de préciser la nature de chacune des variables qui figurent dans le tableau (caractères quantitatif ou qualitatif ; qualitatif nominal ou ordinal ; quantitatif de stock ou de taux, mesurable ou repérable, continu ou discret) et de savoir ce que l'on est en droit de faire avec les calculs statistiques.

On prend soin de préciser que les deux premières variables, Nom de l'état et code, ne sont pas considérées comme des variables mais comme le nom des individus observés du tableau élémentaire.

Il s'agit également de sensibiliser les étudiants aux « familles » de variables au sens où celles-ci regroupent plusieurs caractères qui apportent différents éclairages sur un même phénomène. Le tableau comporte beaucoup de variables qui concernent la répartition des populations dans chacun des Etats américains, comme :

- **des caractéristiques générales** : population totale en 2000 (habitants), part de la population de l'Etat dans la population totale (%), variation de la population de l'Etat (%/1990), densité de population de l'Etat (hab/miles²), population urbaine en 1990 (habitants), population rurale en 1990 (habitants), âge médian de la population de l'Etat (années), nombre de personnes d'origine étrangère en 1980, part des personnes d'origine étrangère dans la population totale en 1980 (%), nombre de personnes d'origine étrangère en 1990, part des personnes d'origine étrangère dans la population totale en 1980 (%).
- **mais aussi une répartition de cette population selon les classes d'âge** : part des moins de 5 ans dans la population totale de l'Etat (%), part des 5-17 ans dans la population totale de l'Etat (%), part des 18-24 ans dans la population totale de l'Etat (%), part des 25-44 ans dans la population totale de l'Etat (%), part des 45-64 ans dans la population totale de l'Etat (%), part des plus de 65 ans dans la population totale de l'Etat (%).
- **sans oublier tout ce qui a trait à l'origine des populations** : part de la population blanche dans l'ensemble de la population de l'Etat (%), part de la population noire dans l'ensemble de la population de l'Etat (%), part de la population amérindienne dans l'ensemble de la population de l'Etat (%), part de la population d'origine asiatique dans l'ensemble de la population de l'Etat (%), part des autres populations

dans l'ensemble de la population de l'Etat (%), part de la population se déclarant appartenir à deux races dans l'ensemble de la population de l'Etat (%), part de la population d'origine hispanique dans l'ensemble de la population de l'Etat (%), nombre de personnes d'origine étrangère en 1980, part des personnes d'origine étrangère dans la population totale en 1980 (%), nombre de personnes d'origine étrangère en 1980, part des personnes d'origine étrangère dans la population totale en 1980 (%)... durant ce regroupement on peut aussi faire un point concernant les renseignements collectés dans le cadre des opérations de recensement américain.

De la même manière, on peut faire recenser tout ce qui a trait au niveau de richesse des populations des Etats, à leurs niveaux d'éducation, enfin aux résultats électoraux enregistrés au cours des trois derniers scrutins présidentiels américains.

Enfin, on réserve un temps plus long au travail de réflexions : démarches préalables indispensables à toute analyse statistique ou à une analyse de données. Ce qui signifie qu'avant de commencer une analyse, il est fondamental d'avoir des connaissances préalables solides sur le thème que l'on souhaite traiter et donc d'avoir effectué une analyse bibliographique sérieuse. On peut alors rappeler que toute étude commence toujours par une définition des objectifs voire d'une problématique. Mais, au fait, qu'est-ce qu'une problématique ?

- C'est une construction cohérente d'un dispositif de recherche.
- Il s'agit ainsi d'une base indispensable à toute recherche.
- Une problématique est composée de problèmes clairement posés (souvent sous forme de questions hiérarchisées).
- La problématique donne le fil conducteur de l'étude qui suit. Toute phase du travail, que vous effectuez, a pour but de vous amener à répondre à votre problématique.

Il s'agit d'amener les étudiants à développer une question ou un ensemble de questions auxquelles on espère répondre soit en totalité, soit en partie à la fin de cette étude. Par exemple, si l'on s'intéresse aux inégalités et aux oppositions à l'intérieur des Etats-Unis d'Amérique, on peut se demander si certains phénomènes démographiques contribuent plus que d'autres à créer, à creuser ces inégalités ou ces oppositions. Si tel est le cas, on peut tenter de mettre en évidence d'éventuels regroupements « régionaux » comme le Sud, le Nord-Est etc. On peut enfin s'interroger sur la longévité temporelle de ces regroupements « régionaux » : autrement dit quelles peuvent être les relations entre évolutions démographiques et structurations spatiales ? On peut enfin se demander s'il y a des relations entre les caractéristiques des populations des Etats américains et les résultats enregistrés lors des dernières élections présidentielles.

Une fois la problématique clairement définie, on réfléchit aux données statistiques qui permettent de répondre aux questions. Il est possible qu'en fonction des données disponibles, on soit obligé de revenir sur la problématique et la modifier, ou alors de supprimer certaines variables et d'en recueillir d'autres. En ce qui concerne nos variables et si on conserve les thématiques proposées à la question précédente, on peut proposer les questions suivantes :

- ① Quelle est la part du territoire américain considéré comme rurale ? Quels sont les Etats américains les plus ruraux ? Où se localisent les concentrations de population ?
- ② Quelles sont les caractéristiques démographiques de la population des Etats américains ? Quels sont les Etats américains les plus jeunes et les plus vieux ? Où se trouve la population urbaine ? Peut-on tracer un isoligne de l'âge médian ?

③ Les Etats du Sud, du Nord, de l'Est et de l'Ouest concentrent-ils la même part de population étrangère ? Quelle est la cohérence territoriale de cette population ? Existe-t-il des différences liées aux origines de cette population étrangère ? Quelle est la répartition de la part de chaque race dans la population étrangère et/ou dans la population totale ?

Si on regroupe les thèmes 1, 2 et 3 : Les populations étrangères se localisent-elles plus particulièrement dans les zones rurales ou urbaines ? On peut émettre l'hypothèse que les populations étrangères se regroupent plutôt dans les zones urbaines.

⑥ Quelle est la couleur politique des Etats américains ? Cette couleur a-t-elle évolué entre 1996 et 2004 ?

UN TABLEAU DE DONNEES : POUR QUOI FAIRE ?

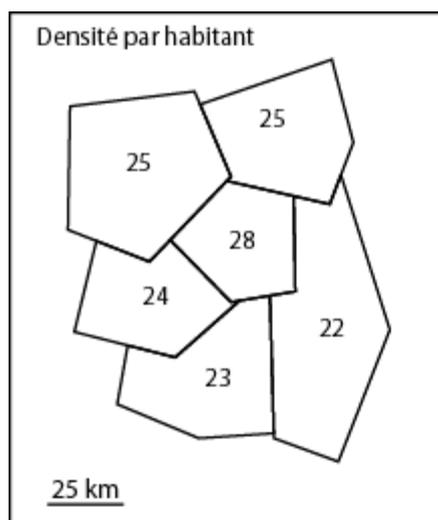
2. Résumer beaucoup d'informations peut-être...

Documents :

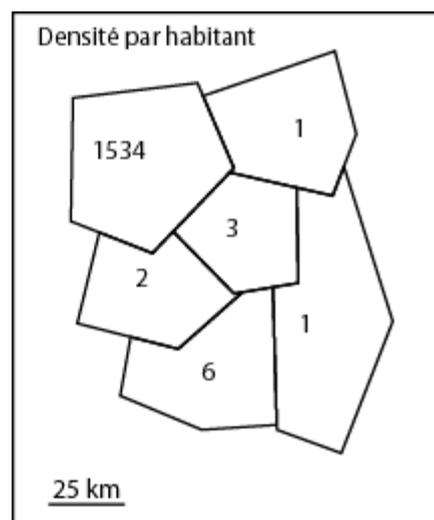
- Fond de carte de localisation de 49 Etats américains
- Tableaux résumant les caractéristiques démographiques des Etats américains
- Matrice de corrélations entre les caractéristiques démographiques des Etats américains

Sources : New-York Times Almanac, 2001, 2002

Les variables retenues pour ce travail sont quantitatives de taux continues : ce qui signifie que la plupart des traitements statistiques uni- ou bivariés sont possibles. Or, il faut bien se rendre compte, que même si le travail proposé ne porte pas sur l'ensemble du tableau de données, il y a encore beaucoup trop d'informations : il faut donc trouver une manière efficace de commencer un résumé de tout ceci. Il est indispensable, dans tout résumé de données chiffrées, d'avoir une idée claire des ordres de grandeur, de l'étalement (ou dispersion) des valeurs, enfin de pouvoir comparer voire classer les variables les unes par rapport aux autres... Dans ce dernier cas on peut chercher à évaluer les différenciations introduites par les variables entre les Etats américains : on utilise alors un paramètre de dispersion. En effet les caractéristiques de dispersion ont pour but d'indiquer globalement, par une seule valeur, l'hétérogénéité d'une distribution. L'interprétation d'un tel paramètre permet de relier dispersion statistique et dispersion géographique mais en sens inverse : la dispersion statistique est d'autant plus grande que le phénomène est concentré en un endroit mais la dispersion géographique est d'autant plus grande que le phénomène est uniformément répandu :



Territoire HOMOGÈNE
Dispersion statistique faible
Dispersion géographique forte



Territoire HÉTÉROGÈNE
Dispersion statistique forte
Dispersion géographique faible
(Concentration spatiale)

Comme on doit comparer la dispersion de plusieurs variables qui ont certes la même unité de mesure mais pas toujours les mêmes ordres de grandeur, on doit choisir un paramètre de dispersion relative qui exprime l'hétérogénéité indépendamment de l'unité de mesure et sans effet de taille. Le paramètre de dispersion relative dont nous disposons ici est le **COEFFICIENT DE VARIATION (en abrégé c.v. défini comme le rapport de l'écart-type sur la moyenne)**. Ce coefficient permet de comparer les variables ayant des unités de mesure différentes et de classer les variables en fonction de leur degré de différenciation introduit entre les Etats américains. La lecture du coefficient de variation dans le tableau 2 nous indique que les variables qui introduisent le moins de différenciations entre les Etats américains sont les variables donnant la répartition de la population selon leur âge (c.v. variant de 0,06 à 0,18). Tandis que les variables qui introduisent le plus de différenciations sont le taux de variation de la population d'origine étrangère entre 1980 et 1990 (c.v. égal à 2,94) et la densité de population en 1990 (c.v. égal à 3,46). On souligne également le fait que les valeurs des coefficients de variation peuvent s'exprimer en %. Par exemple un coefficient égal à 0,06 peut être transformé en % : ce qui signifie que l'écart-type vaut 6% de la moyenne. Il semble important enfin de montrer les limites du coefficient de variation qu'il n'est pas toujours légitime de calculer.

En général, les caractères de stock sont les seuls caractères pour lesquels il soit juste de calculer un paramètre de dispersion relative (et donc le coefficient de variation). En effet, pour que cela ait un sens de parler d'un coefficient de variation de 10 % autour de la moyenne, il faut que la notion de somme ait un sens (10% du total), ce qui n'est vrai que pour les caractères de stock.

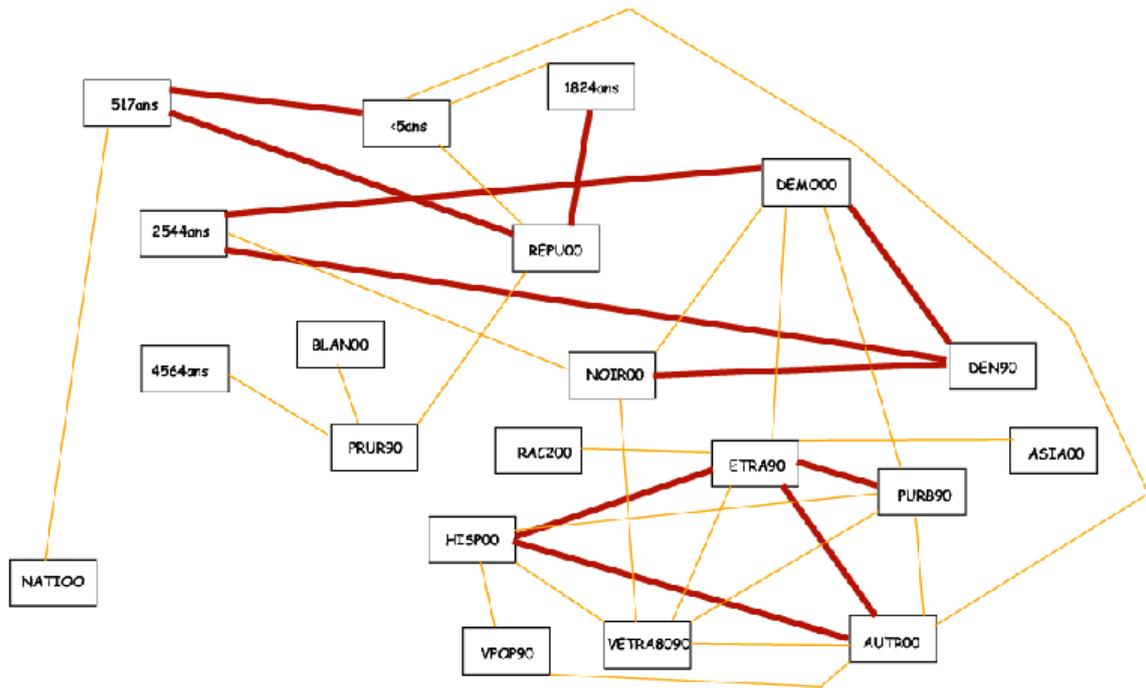
Pour les caractères de rapport, les paramètres de dispersion relative n'ont pas de valeur dans l'absolue et on ne peut les calculer que pour les caractères positifs mesurables, c'est à dire ceux dont toutes les modalités sont positives, dont la moyenne ne peut pas être négative ou nulle et pour lesquels le 0 signifie l'absence du phénomène. On ne le calcule que dans **une perspective comparative** et pour des variables qui utilisent la même unité de mesure, par exemple, comparaison de la dispersion d'une même variable à deux dates différentes (C.V. du taux de chômage des régions françaises en 1980 et en 1990) ; comparaison de la dispersion d'une même variable dans deux ensembles spatiaux différents (C.V. du taux de chômage des régions françaises et des régions italiennes) ou encore, comparaison de la dispersion de deux variables ayant la même unité de mesure

Après avoir rappelé ce qu'est un coefficient de corrélation (à ne pas confondre bien évidemment avec le coefficient de variation vu auparavant), son intervalle de variation et ses différentes significations, on insiste sur ce qu'il nous apprend des différentes combinaisons de variables dans les Etats américains. On peut alors mettre en regard le graphe des associations entre variables (corrélations positives et significatives) et celui des oppositions (corrélations négatives et significatives). Une partie du graphe des corrélations significatives négatives (à savoir toutes celles inférieures à $-0,4$) souligne des évidences sans beaucoup d'intérêt: les Etats « jeunes » s'opposent aux Etats « plus vieux » ! Les Etats où la population noire est forte aux Etats où la population blanche est forte ! Les urbains aux ruraux ! On note cependant que dans les Etats où les densités sont fortes, les votes républicains sont faibles, les Etats ruraux sont ceux où la population étrangère est la plus faible. Au final, on peut proposer de supprimer impérativement une des deux variables population urbaine et population rurale qui sont corrélées à -1 ! La plus forte corrélation suivante est $-0,758$ et concerne les votes lors de l'élection présidentielle de 2000 en faveur du candidat démocrate et les 5 - 17 ans : ce qui signifie que les Etats où les votes en faveur du candidat démocrate à la présidence ont été plus importants qu'en moyenne sont caractérisés par une sous représentation des populations très

jeunes. On met ainsi en évidence une opposition spatiale et non le fait que ce sont les 5-17 ans qui votent plus en faveur du candidat démocrate ! Attention à l'erreur écologique (cf. Grasland, 1994). Cette corrélation reste intéressante et ne peut être supprimée : il n'y a pas de redondance de l'information.

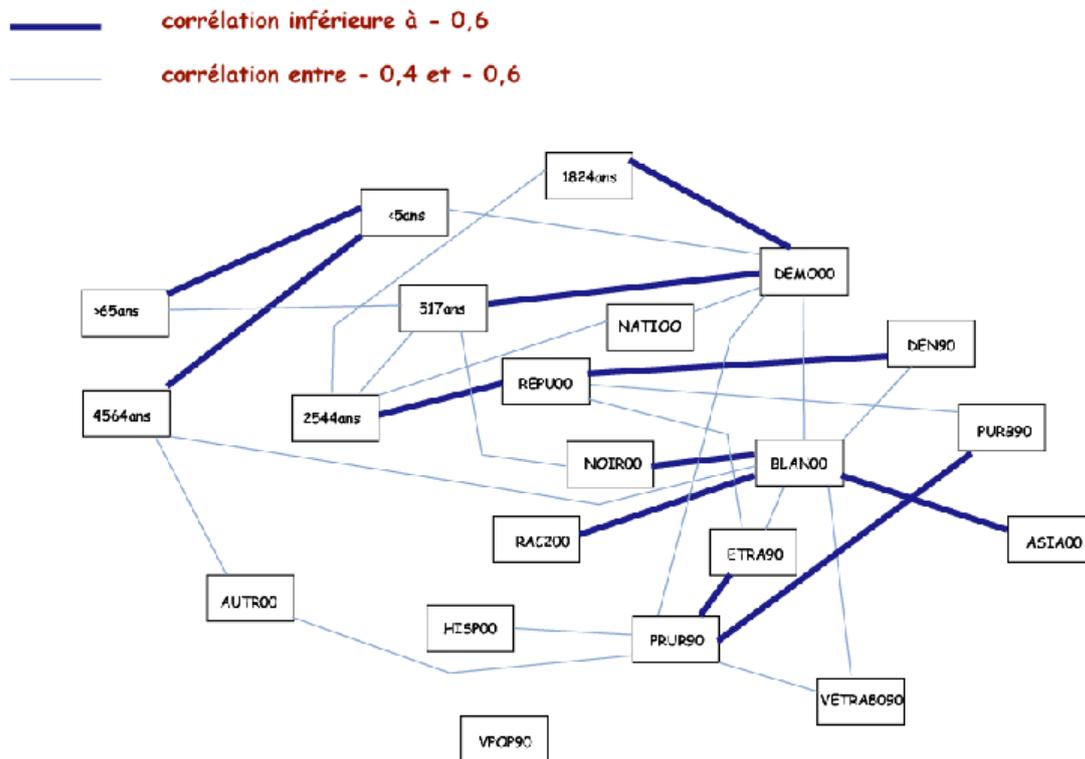
Graphique des corrélations positives entre des caractéristiques des états américains :

- corrélation supérieure à 0,6
- corrélation entre 0,4 et 0,6



corrélations significatives pour 51 états américains (pour 49 degrés de liberté et 5% d'erreur) : $\pm / - 0,2732$

Graphe des corrélations négatives entre les caractéristiques des états américains :



corrélations significatives pour 51 états américains (pour 49 degrés de liberté et 5% d'erreur) : $r = -0,2732$
 Graphiques réalisés à l'aide du logiciel Adobe Illustrator (version 10)

Dans le cadre d'une séance introductive et avant d'entamer des travaux d'interprétations de résultats d'Analyse de Données, on peut insister sur le fait qu'une Analyse en Composantes Principales normée (ACP) peut s'appliquer à cet ensemble de données. Il s'agit en effet de données quantitatives relatives (de taux) dont les unités de mesure ne sont pas identiques : décider de réaliser une ACP normée dans ce cas, c'est privilégier la structure de l'information aux dépens des ordres de grandeur (Sanders, 1989). Une fois ce point précisé, on montre l'intérêt d'avoir sous la main le tableau concernant les ordres de grandeur des variables sur lequel on pourra revenir à tout moment : par exemple, si d'aventure on met en évidence que certains Etats sont caractérisés par une sur représentation des populations amérindiennes, il convient de préciser que cette population ne représente en moyenne que 1,65% de la population totale d'un état. De même, si certains Etats sont caractérisés par une sur représentation des populations blanches dites WASP (White Anglo-Saxon Protestants), il faudra également insister sur le fait qu'en moyenne près de 8 personnes sur 10 sont WASP aux Etats-Unis ! D'après l'analyse des corrélations entre les variables sélectionnées dans le cadre de cette étude, on peut déjà émettre certaines hypothèses quant aux résultats de l'ACP normée :

- Comme la part de population urbaine est associée aux taux élevés de population étrangère, aux différentes origines de cette même population étrangère, à ces variations entre 1980 et 1990, on peut penser que dans la définition et la structuration des premiers facteurs de l'ACP, ces variables se « regroupent » ou se projettent du même côté. De même, comme cette part de population urbaine est opposée à l'importance de la population dite WASP et aux populations plus âgées, on peut penser que dans la définition et la structuration des premiers facteurs de l'ACP, il n'y a aucune chance qu'elles soient regroupées.

- Les Etats les plus densément peuplés, dans lesquels les populations noires sont sur représentées ainsi que les jeunes adultes et/ou les jeunes actifs ne se regrouperont pas, ne se projettent pas du même côté que les Etats dans lesquels les populations très jeunes (à savoir les moins de 24 ans) sont sur représentés.

Bibliographie

Ouvrages indispensables

****Dumolard P., Dubus N., Charleux L., 2003, *Les statistiques en géographie*, Paris : Belin, collection Atouts Géographie, 239 p.

Cet ouvrage est un aperçu très détaillé des principales notions de statistiques uni- et bivariées vues dans le cadre de cette partie. Il présente l'avantage d'être rédigé par des géographes et de comporter de nombreux exercices d'application avec solutions. Sa consultation est indispensable pour être sûr de maîtriser cette partie.

****Grasland C., Saint-Julien T., 1999, « Itinéraires géographiques en pays statistique », *Feuilles de Géographie*, Hors-Série n°1.

Ouvrages de références

**Béguin H., 1979, *Méthodes d'analyse géographique quantitative*, Paris: Litec, collection Droit, 252 p.

Ouvrage très complet, parfois même trop complet. Il présente un large éventail des méthodes et outils statistiques utilisés en géographie. Il comporte des éléments vus en DEUG et, comme l'ouvrage de P.-J. Taylor, il est aussi conseillé dans d'autres enseignements de licence.

***Groupe Chadule, 1994, *Initiation aux pratiques statistiques en Géographie*, Paris: Masson, collection Géographie, 4ème édition, 189 p.

Cet ouvrage fait partie des références qui sont connues depuis le DEUG. Il permet de réviser efficacement les notions supposées acquises en statistique uni- et bivariée. Il permet aussi d'entrevoir les liens possibles entre méthodes d'analyse statistique et méthodes d'analyse de données.

UN TABLEAU DE DONNEES : POUR QUOI FAIRE ?

3. Faire des cartes peut-être...

Documents :

- Fond de carte de localisation de 49 Etats américains
- Tableaux de données servant de base à la réalisation de 2 planches cartographiques

Sources : New-York Times Almanac, 2001, 2002

Quoi de plus satisfaisant que de finir une séance de travail par la réalisation d'une voire de plusieurs cartes ... encore faut-il « reprendre » de bonnes habitudes... Autrement dit identifier clairement la nature des variables statistiques à cartographier.

Pour la première carte, la population totale est une variable de stock et l'évolution de la population entre 1990 et 2000 est une variable de taux mesurable et continue. Il suffit de rappeler qu'une variable de stock est représentée par la combinaison des variables visuelles forme géométrique et taille. La forme peut être un cercle, un carré ou un triangle et sa taille varie en fonction de la quantité à représenter : on rappelle ainsi les principes d'utilisation de l'abaque de César et surtout le fait qu'il convient de prendre en compte le fond de carte sur lequel est représenté le phénomène. Quant à l'évolution de la population de l'état au cours de la dernière période inter-censitaire, il existe plusieurs combinaisons de variables visuelles qui permettent de la représenter cartographiquement. On peut utiliser simplement et uniquement la valeur ; ou une couleur et la valeur ; ou une texture-structure combinée avec le grain. Encore faut-il définir le nombre de classes, qui dépend non seulement du nombre d'unités spatiales à cartographier mais aussi de ce que l'on veut montrer. Dans le cas qui nous occupe si l'on se réfère à l'indice d'Huntsberger :

$$N(\text{cl}) = 1 + (3,3 * \log(N)) \quad \text{où } N(\text{cl}) \text{ est le nombre de classes}$$

N est le nombre d'observations

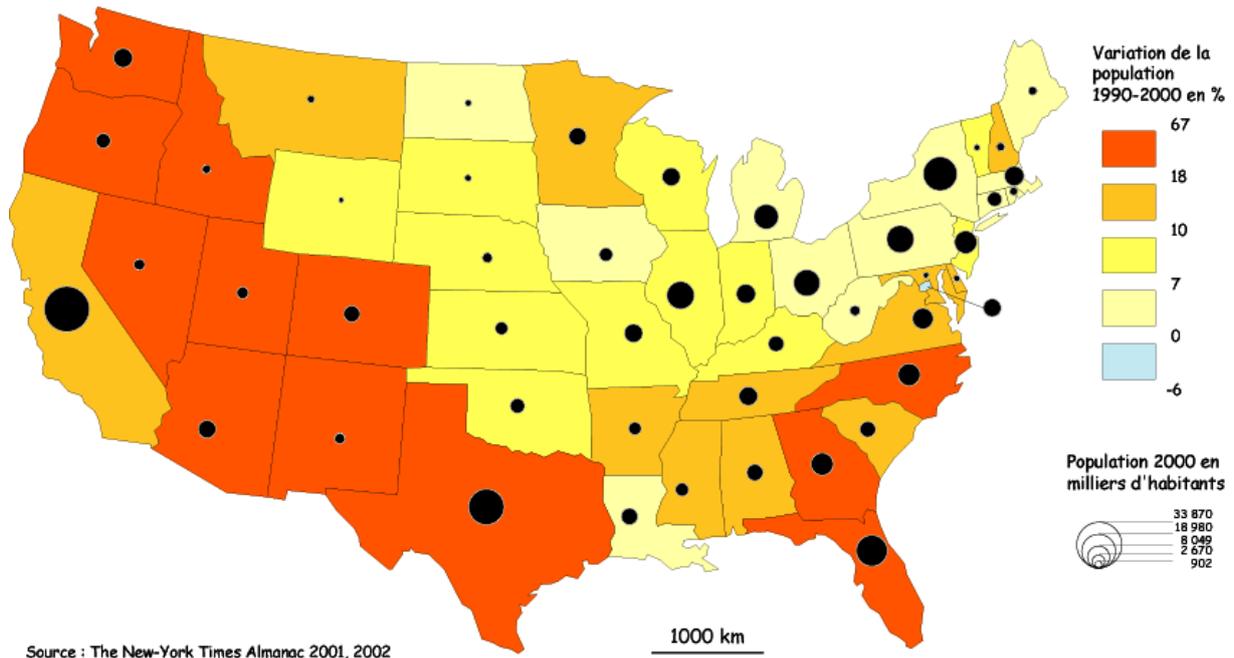
$$N(\text{cl}) = 6,6$$

On peut donc dire que l'on peut faire une partition en au plus 6 classes de l'évolution de la population au cours de la dernière période inter censitaire. Enfin, la méthode de discrétisation adoptée dépendra pour l'essentiel de la forme de la distribution statistique. S'il s'agit d'une distribution qui suit la loi normale, on peut hésiter entre une partition en classes d'effectifs égaux ou une partition en classes d'amplitudes égales. Si la distribution est dissymétrique, le choix de partitions se réduit : seule la partition en classes d'effectifs égaux reste. Enfin, on apporte un soin particulier à la mise en page et à l'habillage de la planche cartographique.

On propose ici deux possibilités de représentation cartographique, selon la même partition en classes. Au vue de ces deux cartes force est de constater que l'information qui ressort n'est pas la même. Dans le premier cas, qui correspond à la dissociation entre les deux variables, le lecteur est immédiatement attiré par les différences de croissance de populations entre 1990 et 2000. En particulier, on retient l'image du croissant des Etats les plus dynamiques qui coure

de l'état de Washington sur la côte pacifique à la Caroline du Nord sur la côte atlantique en passant par l'Arizona, le Texas ou encore la Floride. On retient également le gradient de décroissance des taux de variation de population orienté vers le Nord Nord-Est. Secondairement, on s'intéresse à l'importance de la population des Etats : ce qui permet de nuancer l'effet Sun Belt, Eldorado pour les taux de croissance de population les plus élevés, en soulignant qu'une grande partie des Etats qui comptent parmi les taux de croissance les plus élevés ne sont pas très peuplés (Oregon, Nevada, Arizona, Colorado, Utah, Nouveau Mexique).

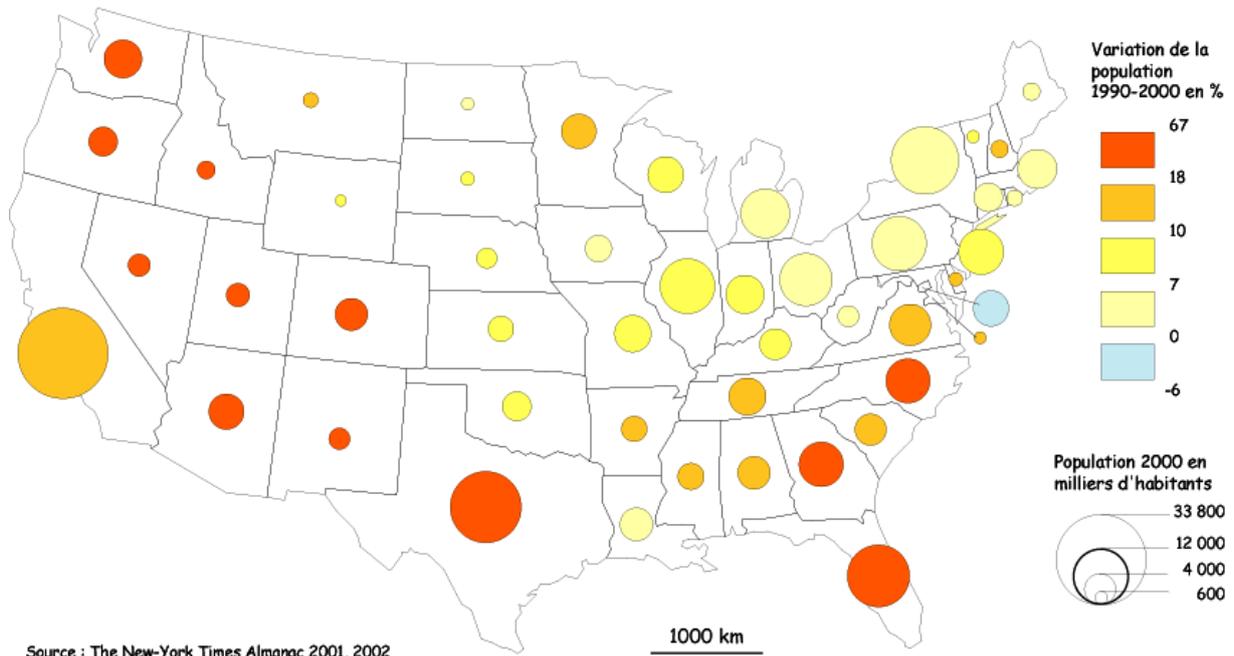
POPULATION TOTALE ET VARIATION DE LA POPULATION 1990-2000



La seconde carte intègre davantage les deux phénomènes et permet directement de faire la relation entre importance de la population de l'état et taux de croissance au cours de la dernière période inter censitaire. Il ressort immédiatement de cette carte que généralement ce ne sont pas les zones les plus peuplées qui connaissent les croissances de population les plus élevées (Etats de New York, du New-Jersey, de l'Ohio, de l'Illinois et dans une moindre mesure la Californie). En revanche, on souligne le cas des Etats dont la population est intermédiaire et qui eux enregistrent des taux de croissance plutôt très élevés surtout quand ils sont situés au Sud, Sud-Ouest et Ouest. Enfin, on remarque la position très particulière de l'état correspondant à la capitale fédérale (Washington D.C.) qui est le seul à perdre de la population au cours des dix dernières années.

Ces deux cartes sont un choix : on aurait pu choisir de comparer deux cartes réalisées avec deux méthodes différentes de partition en classes, ou encore deux cartes réalisées avec des nombres de classes différents. On le voit : les possibilités sont infinies pour sensibiliser les étudiants au fait que les choix faits en matière de production cartographique contribuent à produire des images différentes et qu'il faut dépasser le questionnement qui consiste à penser que les cartes, surtout thématiques, peuvent être objectives.

Population et variation de la population des états américains ne vont pas de pairs



Pour traiter la question sur les résidus, un rappel de leur signification est sans doute nécessaire. En préambule on insistera sur le fait que la détermination d'une équation de régression n'est pas une fin en soi et que seule l'interprétation géographique est intéressante. Les résidus sont le résultat d'une analyse bi-variée entre une variable X et une variable Y. On cherche à expliquer les variations de Y en fonction de celles de X au moyen d'une équation. Attention on rappelle et on insiste sur le fait que la mise en évidence des dépendances entre les variations de X et celles de Y ne signifie pas qu'il y a causalité. Lorsque le coefficient de corrélation entre deux variables est significatif, on est en droit d'essayer de trouver l'équation de la droite de régression linéaire, considérée comme le modèle de la relation, si cette relation était totale. Les résidus correspondent alors à l'écart qui peut exister entre ce modèle et la relation réelle. Plus le coefficient de corrélation est élevé, plus les résidus seront faibles. L'interprétation de ces résidus doit se faire en termes de surestimation et sous-estimation et n'a de signification que par référence au modèle. Ils contiennent deux informations : leur **valeur absolue** (la valeur 0 correspond exactement au modèle, plus la valeur absolue du résidu est élevée, plus l'écart au modèle est important et le comportement de l'individu correspondant est mal pris en compte par le modèle) et leur signe (si le résidu a une valeur positive, cela signifie que le modèle sous estime Y et les valeurs de Y sont plus fortes que ne le laisse supposer le modèle, s'il est négatif, il le surestime et les valeurs de Y sont moins fortes que ne le laisse supposer le modèle).

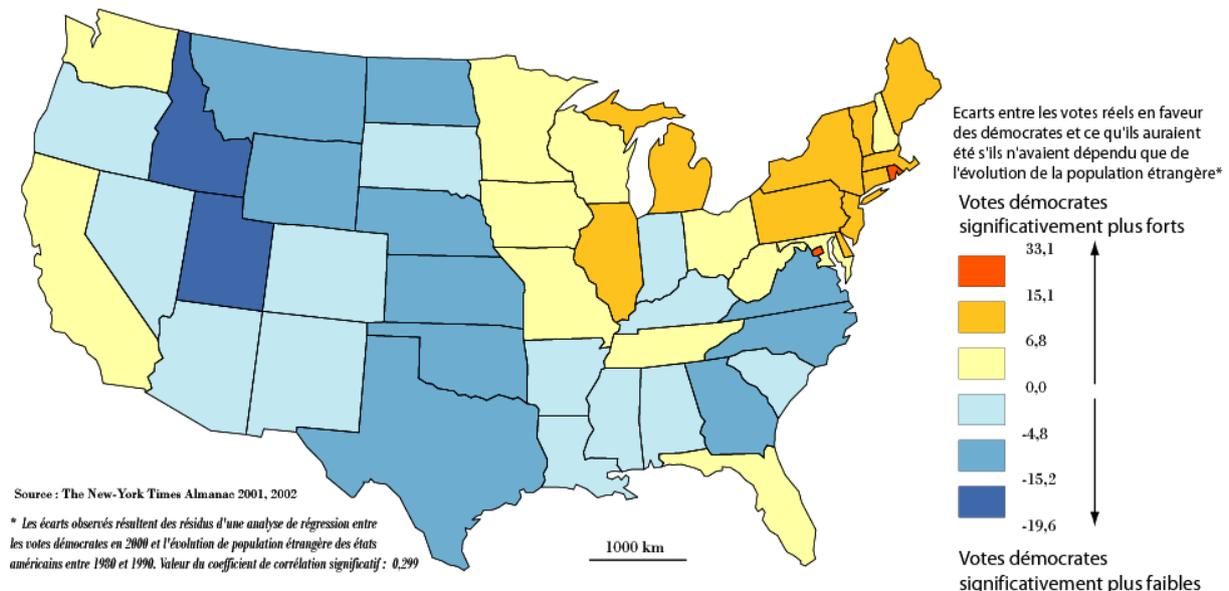
Ici on analyse la relation entre la variation de la population étrangère et le vote démocrate. On cherche à savoir si le vote démocrate (Y) est fonction de la variation de la population étrangère de chaque état américain. Le coefficient de corrélation entre ces deux variables est égal à + 0,3. Ce coefficient est significatif de l'existence d'une relation, elle est positive mais pas très forte (on suppose que plus la variation de la population étrangère est forte, plus le vote démocrate sera fort), on peut donc prévoir que les valeurs absolues des résidus seront plutôt élevées. Pour les résidus positifs, on peut dire que le modèle sous-estime le vote démocrate, le vote démocrate est plus élevé que ne le laisse supposer le modèle. Pour les

résidus négatifs, le modèle surestime les valeurs du vote démocrate : dans ces Etats, le vote démocrate est plus faible que ne le laisse supposer le modèle.

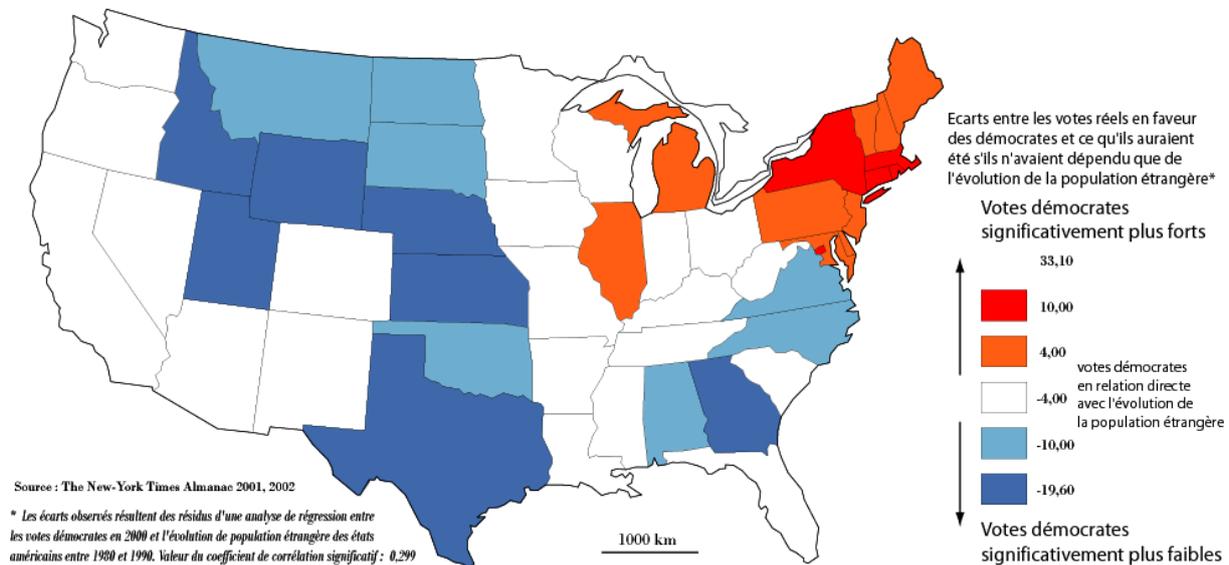
Comme les individus sont ici les unités spatiales, il est possible de réaliser une carte des résidus qui mettra en évidence des phénomènes de différenciation spatiale différents de ceux qu'on représente à l'aide des données « brutes ». Les caractéristiques particulières de la distribution des résidus doivent être traduites correctement :

- La distribution des valeurs de résidus doit être considérée comme une nouvelle variable, ici les deux variables X et Y sont des variables de taux, les résidus présentent la même unité de mesure et donc se représentent comme une variable de taux : on choisit donc la variable visuelle valeur ou texture /structure (en couleur ou noir et blanc).
- la valeur 0 qui correspond au modèle doit être mise en avant : si le nombre de classes est pair, le 0 apparaît comme une limite de classe ; si le nombre est impair (ce qui est souvent préférable) le 0 est inclus dans une classe centrale qui correspond aux individus proche ou égal au modèle.
- On choisit deux types de symboles cartographiques opposés (deux camaïeu ou une double gradation harmonique si on choisit la variable visuelle couleur ; deux texture/structure pour la variable visuelle valeur en noir et blanc ou en couleur): l'un pour les résidus négatifs et l'autre pour les positifs.
- Les valeurs les plus foncées correspondent toujours aux plus fortes valeurs des résidus.

POPULATION ETRANGERE ET VOTES DEMOCRATES



POPULATION ETRANGERE ET VOTES DEMOCRATES



Ces cartes font apparaître des regroupements d'Etats différents de ceux présentées sur les cartes univariées. Le centre des Etats-Unis (le Montana, l'Utah, le Kansas ou le Texas, par exemple) apparaît comme une zone où le vote démocrate est significativement plus faible que ne le laissait prévoir les variations de la population étrangère. À ces classes s'associent quelques Etats du sud est comme l'Alabama, la Géorgie ou la Caroline du nord. Du côté des résidus positifs on trouve des Etats du nord est (le Maine, le Massachusetts, le New Jersey ou la Virginie) où les votes démocrates sont supérieurs à ce que laissent supposer les variations de la population étrangère. Cette carte ne traduit pas des oppositions simples entre les Etats où le vote démocrate est élevé et ceux où il est faible. Le modèle de régression permet d'aller plus loin : il exprime la liaison entre les dynamiques démographiques liées aux populations étrangères et le vote démocrate, ce que l'on peut obtenir aussi en superposant les deux cartes univariées. Mais cela serait peu lisible. La cartographie des résidus offre donc plus de facilités et de précisions.

Bibliographie

Ouvrages indispensables

****Béguin M. et Pumain D., 1994, *La représentation des données géographiques. Statistique et cartographie*, Paris: Armand Colin, Cours, 192 p.

Cet ouvrage a déjà été recommandé dans les enseignements de DEUG relatifs aux méthodes et aux outils de la géographie. Il est surtout centré sur les questions relatives aux représentations cartographiques des résultats statistiques. A regarder de nouveau pour se remémorer les règles de base de la sémiologie graphique.

****Zanin C., Trémelo M.-L., 2003, *Savoir faire une carte. Aide à la conception et à la réalisation d'une carte thématique univariée*, Paris : Belin, collection Sup Géographie, 200 p.

Ouvrages de références

***Brunet R., 1987, *La carte, mode d'emploi*, Paris: Fayard/Reclus, 270 p.

Plus théorique que l'ouvrage de M. Béguin et D. Pumain, *La carte, mode d'emploi* offre une vue plus synthétique de la démarche cartographique et de l'ensemble des questions que l'on peut se poser face à un document cartographique. A consulter pour remettre son travail cartographique en perspective.

**Cauvin C., Reymond H., 1986, *Nouvelles méthodes en cartographie*, Montpellier: GIP Reclus, Reclus Mode d'Emploi, 56 p.

**Cauvin C., Reymond H., Serradj A., 1987, *Discretisation et représentation cartographique*, Montpellier: GIP Reclus, Reclus Mode d'Emploi, 116 p.

Il s'agit d'une collection qui devient actuellement presque introuvable mais qui était conçue très efficacement. Chacun des volumes traitaient d'un type de question et proposait les dernières avancées « géographiques » pour le résoudre. Les volumes rédigés par C. Cauvin et H. Reymond sont remarquables par leurs contenus précis et très pédagogiques ayant trait à des questions relatives à la cartographie.

Révisions concernant les méthodes statistiques uni- et bivariées

Travaux autour d'un tableau élémentaire de données

Documents :- Fond de carte de localisation de 49 Etats américains

- Description du tableau élémentaire caractérisant les Etats américains
- Tableaux résumant les caractéristiques démographiques des Etats américains

Documents à rendre - Une représentation cartographique rendant compte des populations par état et de leurs évolutions entre 1990 et 2000

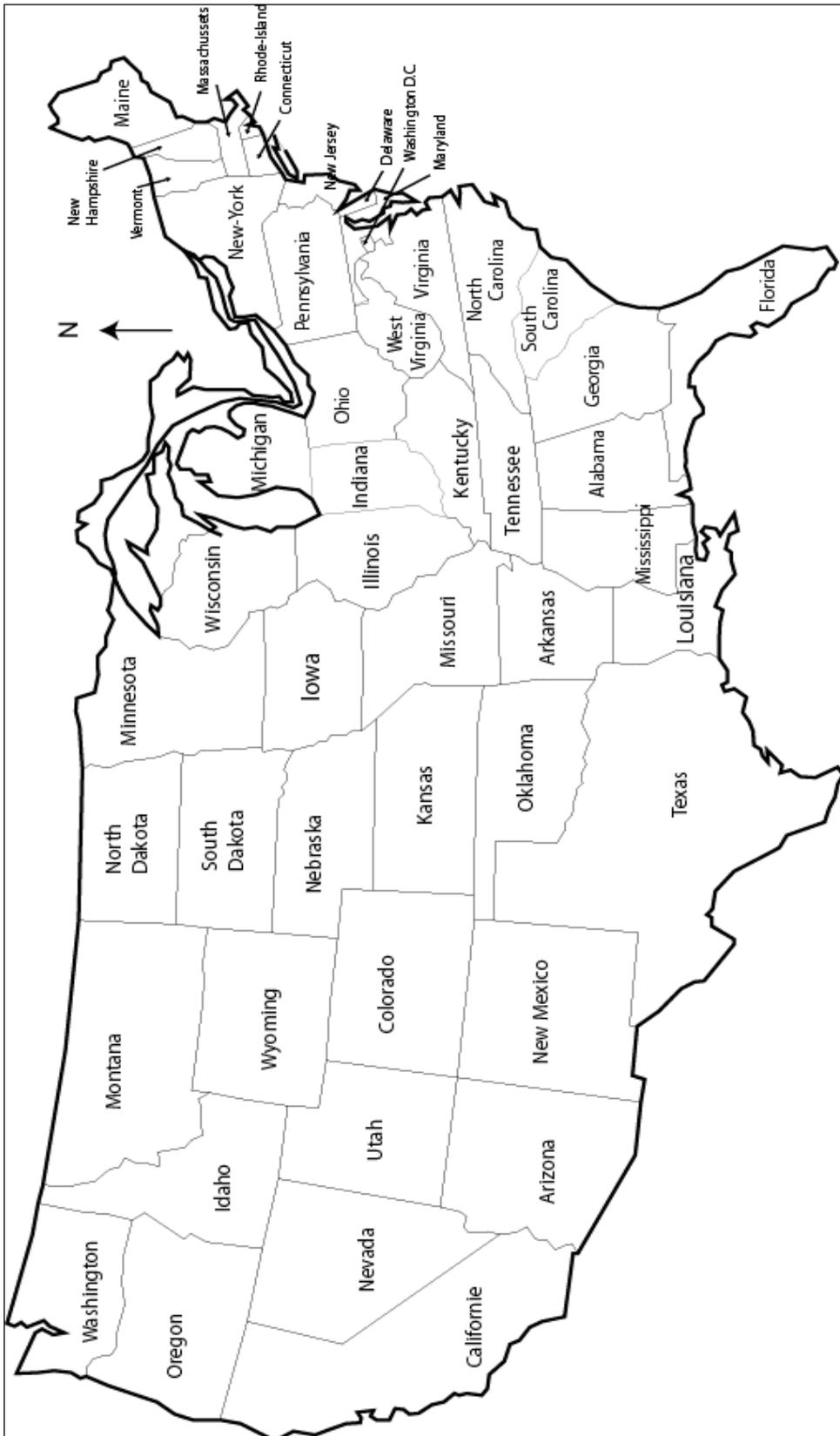
- Une représentation cartographique qui rend compte des écarts entre les votes démocrates - tels que le laissait attendre l'évolution de la population étrangère - et les votes réels

On travaille à partir d'un tableau élémentaire contenant des informations permettant de caractériser les 51 Etats des Etats-Unis d'Amérique.

- ❖ A partir du premier document qui présente l'information contenue dans le tableau élémentaire de données concernant les caractéristiques des 51 Etats américains :
 - Vous indiquerez clairement la nature de chacune des variables qui figurent dans le tableau.
 - Vous indiquerez également les groupes de variables qui caractérisent le même phénomène.
 - Pour chacun de ces groupes de variables, vous tenterez de formuler des questions ou/et des hypothèses de recherche en géographie qui peuvent donner lieu à des analyses statistiques ou/et de données.

- ❖ Dans un deuxième temps, on décide d'étudier les caractéristiques de la répartition des populations dans les 51 Etats des Etats-Unis d'Amérique. Pour ce faire, on dispose d'un tableau qui pour chacun des Etats américains indique la répartition des populations par tranches d'âge, par « races », par origine étrangère, les différentes dynamiques de population. Enfin, pour chacun des Etats, on dispose des résultats aux dernières élections présidentielles. A partir des premiers tableaux qui résument l'information contenue dans le tableau élémentaire de données concernant les caractéristiques démographiques des 51 Etats américains :
 - Vous préciserez la nature des variables retenues.
 - Vous indiquerez les variables qui introduisent le plus de différenciations entre les Etats, et celles qui en introduisent le moins. Vous aurez soin de justifier clairement le choix du paramètre caractéristique qui vous permet de répondre à cette question.
 - Vous construirez le ou les graphes de corrélation, en mettant en évidence les principales associations et oppositions entre les variables. Vous conclurez en émettant des hypothèses sur l'organisation de l'information dans le cas d'une Analyse en Composantes Principales.

49 Etats américains



Description du contenu du tableau concernant les Etats-Unis d'Amérique

Etats Code		Nom complet des Etats Abréviation du nom de l'Etat
Elections présidentielles 1996	Nombre de sièges	Nombre de représentants au Sénat
	Republicains	Nombre de votes en faveur des républicains aux élections de 1996
	Democrates	Nombre de votes en faveur des démocrates aux élections de 1996
	Réforme	Nombre de votes en faveur des réformistes aux élections de 1996
Elections présidentielles 2000	Republicains	Nombre de votes en faveur des républicains aux élections de 2000
	Democrates	Nombre de votes en faveur des démocrates aux élections de 2000
	Verts	Nombre de votes en faveur des verts aux élections de 2000
Population	recensement 2000	Population totale en 2000 (habitants)
	part dans la population totale	Part de la population de l'Etat dans la population totale (%)
	Variation de population 90-00 (%)	Variation de la population de l'Etat (%/1990)
	densité pop99	Densité de population de l'Etat (hab/miles ²)
Rural/Urban	Superficie	Superficie de l'Etat (miles ²)
	Surbain	Superficie de l'espace urbain (miles ²)
	Srurale	Superficie de l'espace rural (miles ²)
	Popurbaine90	Population urbaine en 1990 (habitants)
	Poprurale90	Population rurale en 1990 (habitants)
Population résidente par race et origine hispanique	Blancs	Part de la population blanche dans l'ensemble de la population de l'Etat (%)
	Noirs	Part de la population noire dans l'ensemble de la population de l'Etat (%)
	Amerindiens	Part de la population amérindienne dans l'ensemble de la population de l'Etat (%)
	Asiatiques	Part de la population d'origine asiatique dans l'ensemble de la population de l'Etat (%)
	Autres	Part des autres populations dans l'ensemble de la population de l'Etat (%)
	2 races déclarées ou plus	Part de la population se déclarant appartenir à deux races dans l'ensemble de la population de l'Etat (%)
	Origine hispanique ou latine	Part de la population d'origine hispanique dans l'ensemble de la population de l'Etat (%)
Population	moins de 5 ans	Part des moins de 5 ans dans la population totale de l'Etat (%)
	5-17 ans	Part des 5-17 ans dans la population totale de l'Etat (%)

	18-24 ans 25-44 ans 45-64 ans plus de 65 ans age median	Part des 18-24 ans dans la population totale de l'Etat (%) Part des 25-44 ans dans la population totale de l'Etat (%) Part des 45-64 ans dans la population totale de l'Etat (%) Part des plus de 65 ans dans la population totale de l'Etat (%) Age médian de la population de l'Etat (années)
Population d'origine étrangère	1980 Part dans poptot 1990 Part dans poptot	Nombre de personnes d'origine étrangère en 1980 Part des personnes d'origine étrangère dans la population totale en 1980 (%) Nombre de personnes d'origine étrangère en 1990 Part des personnes d'origine étrangère dans la population totale en 1990 (%)
Revenu par habitant	1980 1990 1999	Revenu par habitant de l'Etat en 1980 (\$/hab) Revenu par habitant de l'Etat en 1990 (\$/hab) Revenu par habitant de l'Etat en 1999 (\$/hab)
% des personnes en deça du seuil de pauvreté	1980 1990 2000	Part des populations de l'Etat vivant en deça du seuil de pauvreté en 1980 Part des populations de l'Etat vivant en deça du seuil de pauvreté en 1990 Part des populations de l'Etat vivant en deça du seuil de pauvreté en 2000
Enseignement public	investissement depenses par élève salaire moyen enseignant	Investissement de l'état pour l'enseignement public Dépense moyenne par élève de l'Etat (\$) Salaire moyen d'un enseignant (\$)
Niveau de diplôme atteint	Personnes de plus de 25 ans 4 années et plus en lycée 4 années et plus d'université	Part des populations de l'Etat ayant plus de 25 ans dans la population totale (%) Part des populations de plus de 25 ans ayant effectué 4 années de scolarité en lycée Part des populations de plus de 25 ans ayant effectué 4 années à l'université
Personnes selon couverture médicale 1999	avec couverture sans couverture %des personnes sans couverture	Nombre d'habitants de l'Etat ayant souscrit une couverture médicale en 1999 Nombre d'habitants de l'Etat n'ayant pas souscrit une couverture médicale en 1999 Part des populations de l'Etat n'ayant pas souscrit de couverture médicale en 1999 (%)

Source : The New-York Times Almanac 2001, 2002

Tableaux 1. Résumés du tableau élémentaire de données

1. Moyennes et écart-types des phénomènes caractérisant les Etats américains

Variables	Codes	Moyennes	Ecart-types	Coefficients de variation
Part des 25-44 ans dans la population totale de l'Etat (%)	2544ans	28,84	1,80	0,06
Part des 45-64 ans dans la population totale de l'Etat (%)	4564ans	21,39	1,35	0,06
Part des 5-17 ans dans la population totale de l'Etat (%)	517ANS	18,28	1,57	0,09
Part des moins de 5 ans dans la population totale de l'Etat (%)	<5ANS	6,57	0,73	0,11
Part des 18-24 ans dans la population totale de l'Etat (%)	1824ans	9,43	1,07	0,11
Part des plus de 65 ans dans la population totale de l'Etat (%)	>65ans	12,20	1,91	0,16
Part de la population blanche dans l'ensemble de la population de l'Etat (%)	BLAN00	78,52	14,39	0,18
Part des votes en faveur des républicains aux élections de 2000	REPU00	49,64	10,30	0,21
Part des votes en faveur des démocrates aux élections de 2000	DEMO00	46,02	10,10	0,22
Part de la population urbaine dans la population de l'Etat en 1990 (%)	PURB90	69,18	15,19	0,22
Part de la population rurale dans la population de l'Etat en 1990 (%)	PRUR90	30,81	15,19	0,49

2. Moyennes et écart-types des phénomènes caractérisant les Etats américains (fin)

Variabiles	Codes	Moyennes	Ecart-types	Coefficients de variation
Part des votes en faveur des verts aux élections de 2000	VERT00	3,10	2,04	0,66
Variation de la population de l'Etat entre 1990 et 2000(%/1990)	VPOP90	13,45	11,38	0,85
Part des personnes d'origine étrangère dans la population totale en 1990 (%)	ETRA90	4,92	4,48	0,91
Part de la population noire dans l'ensemble de la population de l'Etat en 2000(%)	NOIR00	10,88	11,68	1,07
Part de la population d'origine hispanique dans l'ensemble de la population de l'Etat (%)	HISP00	7,79	8,74	1,12
Part des autres populations dans l'ensemble de la population de l'Etat (%)	AUTR00	3,34	3,78	1,13
Part de la population se déclarant appartenir à deux races dans l'ensemble de la population de l'Etat (%)	RAC200	2,40	2,88	1,20
Part de la population amérindienne dans l'ensemble de la population de l'Etat (%)	NATI00	1,65	2,93	1,77
Part de la population d'origine asiatique dans l'ensemble de la population de l'Etat (%)	ASIA00	2,93	5,77	1,97

Taux de variation de la population d'origine étrangère entre 1980 et 1990 (%)	VETRA8090	7,49	22,01	2,94
Densité de population de l'Etat en 1990(hab/miles2)	DEN90	338,19	1171,58	3,46

Tableaux 1. Résumés du tableau élémentaire de données

3. Matrice des corrélations

Codes	VPOP90	DEN90	PURB90	PRUR90	<5ANS	517ANS	1824ans	2544ans	4564ans	>65ans	BLAN00
VPOP90	1	-0,285	0,221	-0,221	0,526	0,200	0,032	-0,277	-0,345	-0,398	-0,014
DEN90	-0,285	1	0,368	-0,368	-0,392	-0,632	-0,290	0,343	-0,035	0,058	-0,486
PURB90	0,221	0,368	1	-1,000	0,295	-0,158	-0,234	0,172	-0,465	-0,170	-0,495
PRUR90	-0,221	-0,368	-1,000	1	-0,295	0,158	0,234	-0,172	0,465	0,170	0,495
<5ANS	0,526	-0,392	0,295	-0,295	1	0,726	0,596	-0,314	-0,626	-0,601	-0,110
517ANS	0,200	-0,632	-0,158	0,158	0,726	1	0,690	-0,428	-0,262	-0,472	0,309
1824ans	0,032	-0,290	-0,234	0,234	0,596	0,690	1	-0,512	-0,145	-0,376	0,081
2544ans	-0,277	0,343	0,172	-0,172	-0,314	-0,428	-0,512	1	-0,041	-0,178	-0,273
4564ans	-0,345	-0,035	-0,465	0,465	-0,626	-0,262	-0,145	-0,041	1	0,273	0,166
>65ans	-0,398	0,058	-0,170	0,170	-0,601	-0,472	-0,376	-0,178	0,273	1	0,181
BLAN00	-0,014	-0,486	-0,495	0,495	-0,110	0,309	0,081	-0,273	0,166	0,181	1
NOIR00	-0,164	0,611	0,160	-0,161	-0,150	-0,492	-0,179	0,405	-0,090	-0,033	-0,635
NATI00	0,097	-0,124	-0,097	0,097	0,290	0,540	0,354	-0,401	0,033	-0,313	0,000
ASIA00	0,001	0,019	0,364	-0,364	0,073	-0,101	-0,014	0,111	0,002	-0,015	-0,616
AUTR00	0,420	0,034	0,551	-0,551	0,411	0,095	-0,081	-0,028	-0,409	-0,271	-0,265
RAC200	0,079	0,007	0,371	-0,371	0,139	-0,008	0,050	-0,037	0,016	-0,068	-0,600
HISP00	0,454	0,013	0,548	-0,548	0,399	0,084	-0,100	-0,090	-0,378	-0,221	-0,296
ETRA90	0,096	0,245	0,710	-0,709	0,068	-0,250	-0,371	0,302	-0,282	-0,033	-0,512
VETRA8090	0,379	0,309	0,535	-0,535	0,290	-0,212	-0,194	0,311	-0,342	-0,360	-0,571
REPU00	0,294	-0,657	-0,451	0,451	0,499	0,684	0,700	-0,616	-0,009	-0,177	0,433
DEMO00	-0,297	0,649	0,434	-0,435	-0,511	-0,758	-0,733	0,625	-0,003	0,270	-0,460
VERT00	-0,079	0,172	0,192	-0,191	-0,071	0,127	-0,055	0,155	0,059	-0,317	-0,010

En gras, valeurs significatives (hors diagonale) au seuil alpha=0.050 (test bilatéral)

Codes	NOIR00	NATI00	ASIA00	AUTR00	RAC200	HISP00	ETRA90	VETRA8090	REPU00	DEMO00	VERT00
VPOP90	-0,164	0,097	0,001	0,420	0,079	0,454	0,096	0,379	0,294	-0,297	-0,07
DEN90	0,611	-0,124	0,019	0,034	0,007	0,013	0,245	0,309	-0,657	0,649	0,17
PURB90	0,160	-0,097	0,364	0,551	0,371	0,548	0,710	0,535	-0,451	0,434	0,19
PRUR90	-0,161	0,097	-0,364	-0,551	-0,371	-0,548	-0,709	-0,535	0,451	-0,435	-0,19
<5ANS	-0,150	0,290	0,073	0,411	0,139	0,399	0,068	0,290	0,499	-0,511	-0,07
517ANS	-0,492	0,540	-0,101	0,095	-0,008	0,084	-0,250	-0,212	0,684	-0,758	0,12
1824ans	-0,179	0,354	-0,014	-0,081	0,050	-0,100	-0,371	-0,194	0,700	-0,733	-0,05
2544ans	0,405	-0,401	0,111	-0,028	-0,037	-0,090	0,302	0,311	-0,616	0,625	0,15

4564ans	-0,090	0,033	0,002	-0,409	0,016	-0,378	-0,282	-0,342	-0,009	-0,003	0,05
>65ans	-0,033	-0,313	-0,015	-0,271	-0,068	-0,221	-0,033	-0,360	-0,177	0,270	-0,31
BLAN00	-0,635	0,000	-0,616	-0,265	-0,600	-0,296	-0,512	-0,571	0,433	-0,460	-0,01
NOIR00	1	-0,295	-0,093	-0,126	-0,171	-0,119	0,032	0,425	-0,349	0,450	-0,33
NATI00	-0,295	1	-0,076	0,164	0,130	0,176	-0,121	-0,053	0,304	-0,403	0,27
ASIA00	-0,093	-0,076	1	0,106	0,960	0,142	0,546	0,145	-0,289	0,248	0,27
AUTR00	-0,126	0,164	0,106	1	0,157	0,972	0,617	0,564	-0,152	0,141	0,11
RAC200	-0,171	0,130	0,960	0,157	1	0,208	0,495	0,148	-0,211	0,150	0,32
HISP00	-0,119	0,176	0,142	0,972	0,208	1	0,615	0,559	-0,122	0,117	0,08
ETRA90	0,032	-0,121	0,546	0,617	0,495	0,615	1	0,477	-0,568	0,531	0,31
VETRA8090	0,425	-0,053	0,145	0,564	0,148	0,559	0,477	1	-0,261	0,299	-0,08
REPU00	-0,349	0,304	-0,289	-0,152	-0,211	-0,122	-0,568	-0,261	1	-0,973	-0,37
DEMO00	0,450	-0,403	0,248	0,141	0,150	0,117	0,531	0,299	-0,973	1	0,15
VERT00	-0,334	0,274	0,270	0,113	0,327	0,085	0,313	-0,083	-0,375	0,157	

En gras, valeurs significatives (hors diagonale) au seuil alpha=0.050 (test bilatéral)

- ❖ Dans un troisième temps, on décide de réaliser soit manuellement, soit à l'aide des logiciels de cartographie et de dessin connus (Philcarto, Adobe Illustrator) :
 - Une carte qui rend compte à la fois des populations par état et de leur évolution entre 1990 et 2000
 - Une carte qui rend compte des écarts entre les votes en faveur des démocrates tels que le laissaient attendre l'évolution de la population étrangère et les votes réels

Pour chacune des cartes, vous accorderez un soin particulier à l'habillage

Tableau 2. Tableau élémentaire décrivant les Etats américains en fonction de la population et des votes démocrates.

Etats	Code	Population POPTOT	part dans	Variation	Variation de	Part des	Résidu de
			la population totale	de population 90-00 (%)	population étrangère 80-90 (%)	votes en faveur des démocrates en 2000	la relation entre variation de la population étrangère et votes démocrates
			PUSA	VPOP90	VETRA8090	DEMO00	DEMORES
Alabama	AL	4447100	1,6	10,1	10,0	41,6	-4,8
Alaska	AK	627000	0,2	14,0	12,5	27,7	-19,0
Arizona	AZ	5131000	1,8	40,0	26,7	44,7	-4,0
Arkansas	AR	2673000	0,9	13,7	10,0	45,9	-0,5
California	CA	33872000	12,0	13,8	43,7	53,5	2,5
Colorado	CO	4301000	1,5	30,6	10,3	42,4	-4,0
Connecticut	CT	3406000	1,2	3,6	-1,2	55,9	11,1
Delaware	DE	784000	0,3	17,6	3,1	55,0	9,6
District of Columbia	DC	572000	0,2	-5,7	51,6	85,2	33,1
Florida	FL	15982000	5,7	23,5	18,3	48,8	1,3

Etats	Code	Population	part dans la population totale	Variation de population 90-00 (%)	Variation de population étrangère 80-90 (%)	Part des votes en faveur des démocrates en 2000	Résidus de la relation entre variation de la population étrangère et votes démocrates
Georgia	GA	8186000	2,9	26,4	58,8	42,9	-10,2
Hawaïi	HI	1212000	0,4	9,3	3,5	55,8	10,3
Idaho	ID	1294000	0,5	28,5	16,0	27,6	-19,6
Illinois	IL	12419000	4,4	8,6	15,3	54,6	7,5
Indiana	IN	6080000	2,2	9,7	-10,5	41,0	-2,5
Iowa	IA	2926000	1,0	5,4	0,0	48,5	3,5
Kansas	KS	2688000	1,0	8,5	25,0	37,2	-11,2
Kentucky	KY	4042000	1,4	9,7	0,0	41,4	-3,6
Louisiana	LA	4469000	1,6	5,9	5,0	44,9	-0,8
Maine	ME	1275000	0,5	3,8	-23,1	49,1	7,3
Maryland	MD	5296000	1,9	10,8	43,5	56,5	5,5
Massachussets	MA	6349000	2,3	5,5	9,2	59,8	13,5
Michigan	MI	9938000	3,5	6,9	-15,6	51,3	8,5
Minnesota	MN	4919000	1,7	12,4	0,0	47,9	2,9
Mississippi	MS	2845000	1,0	10,5	-11,1	40,7	-2,8
Missouri	MO	5595000	2,0	9,3	-5,9	47,1	2,9
Montana	MT	902000	0,3	12,9	-26,1	33,4	-8,0
Nebraska	NE	1711000	0,6	8,4	-10,0	33,3	-10,3
Nevada	NV	1998000	0,7	66,3	29,9	46,0	-3,1
New Hampshire	NH	1236000	0,4	11,4	-15,9	46,8	4,0
New Jersey	NJ	8414000	3,0	8,9	21,4	56,1	8,2
New Mexico	NM	1819000	0,6	20,1	32,5	47,9	-1,6
New York	NY	18976000	6,7	5,5	16,9	60,2	12,9
North Carolina	NC	8049000	2,9	21,4	30,8	43,2	-6,0
North Dakota	ND	642000	0,2	0,5	-34,8	33,1	-7,1
Ohio	OH	11353000	4,0	4,7	-14,3	46,4	3,4
Oklahoma	OK	3451000	1,2	9,7	10,5	38,4	-8,0
Oregon	OR	3421000	1,2	20,4	19,5	47,0	-0,7
Pennsylvania	PA	12281000	4,4	3,4	-8,8	50,6	6,8
Rhode Island	RI	1048000	0,4	4,5	6,7	61,0	15,1
South Carolina	SC	4012000	1,4	15,1	-6,7	40,9	-3,2
South Dakota	SD	755000	0,3	8,5	-21,4	37,6	-4,4
Tennessee	TN	5689000	2,0	16,7	9,1	47,3	1,1
Texas	TX	20852000	7,4	22,8	50,0	38,0	-13,9
Utah	UT	2233000	0,8	29,6	-2,9	26,3	-18,3

Myriam BARON, Christine ZANIN, Révisions concernant les méthodes statistiques uni- et bivariées, Feuilles de Géographie, XII-2004, Feuilles n° ??, ?? p.

Vermont	VT	609000	0,2	8,2	-24,4	50,6	9,0
Virginia	VA	7079000	2,5	14,4	51,5	44,4	-7,7
Washington	WA	5894000	2,1	21,1	13,8	50,2	3,3
West Virginia	WV	1808000	0,6	0,8	-18,2	45,6	3,1
Wisconsin	WI	5364000	1,9	9,6	-7,4	47,8	3,8
Wyoming	WY	494000	0,2	8,9	-15,0	27,7	-15,2

Source : *The New-York Times Almanac, 2002*