

Méthodes de classification et cartographie statistique

Guillaume Daudin

Sciences Po, OFCE

Atelier ATHIS IV, 29 septembre 2007

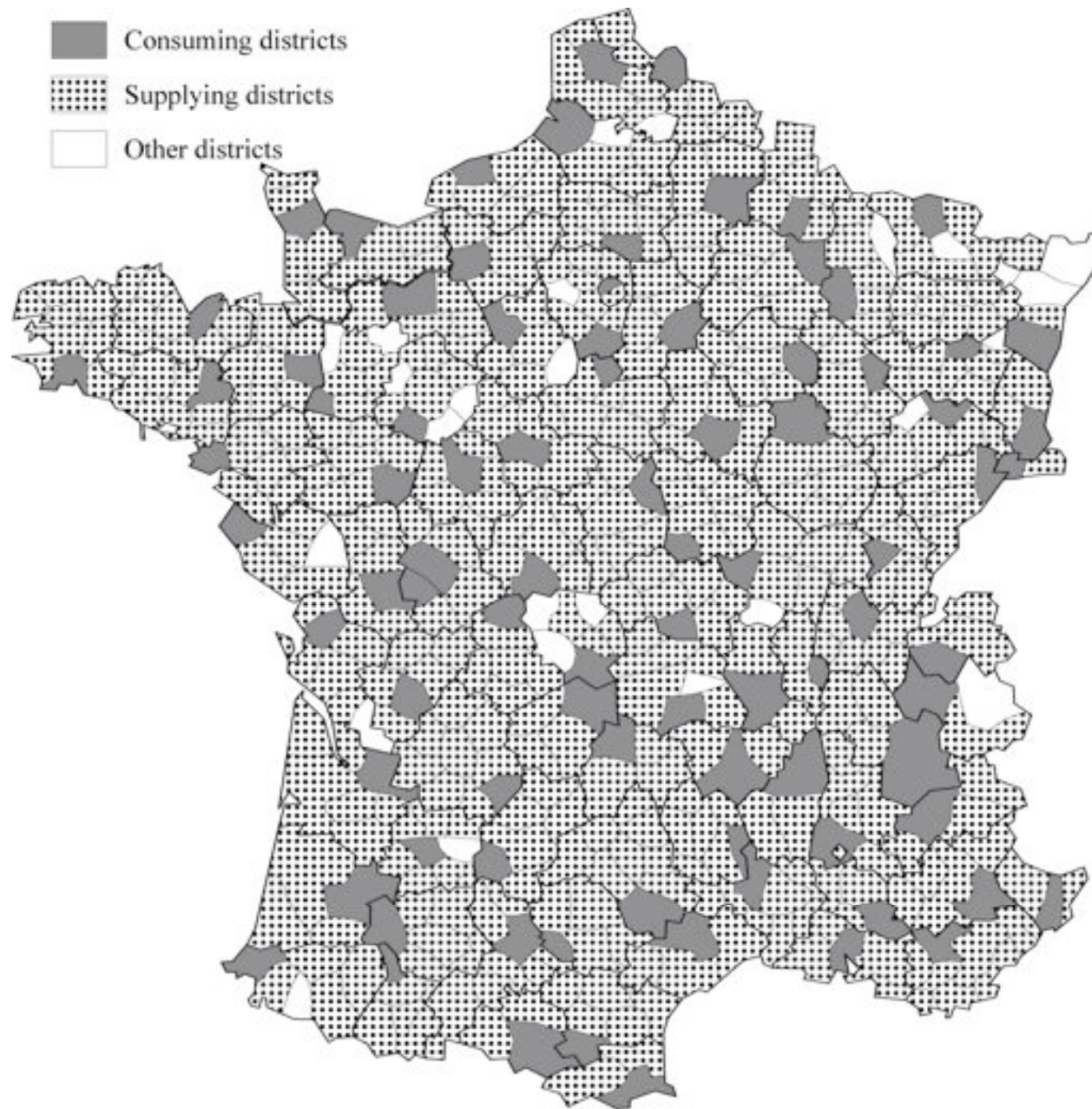
ENS-LSH Lyon

Classification et cartographie?

- D'où est-ce que je parle ?
 - Je viens de l'économie (minorité parmi les sciences auxiliaires de l'histoire)
 - Le but : diffusion des pratiques ?
 - Je devrais sans doute vous parler de méthodes que les économistes maîtrisent : donc pas les méthodes de classification (et moi-même...)
- Ce que je vais faire
 - Vous présenter le prétexte à ma présentation
 - Approche par des régressions logistiques
 - Approche par classification
 - Et toujours un peu de cartographie statistique

Prétexte de la présentation

- Sources
 - Données du « Maximum », 1794 (sur 1791)
 - Pour contrôler les prix, copier le marché
 - Récolte des liens commerciaux entre districts (552) par types de biens (14)
- Questions
 - Marchés régionaux français (mais pourquoi est-ce intéressant ?)
 - Taille du marché français % au marché anglais (lien RI)
- Problèmes de l'analyse
 - (La critique de la source est faite ailleurs)
 - Données binaires, pas des flux commerciaux
 - Rôle pour le caprice des administrateurs



Outil
cartographique:
Philcarto

(Waniez)

[http://perso.club-
internet.fr/philgeo](http://perso.club-internet.fr/philgeo)

Nombre
d'observations:

$88 \times 552 \times 14 = 728\ 640$

Base de données
relationnelle

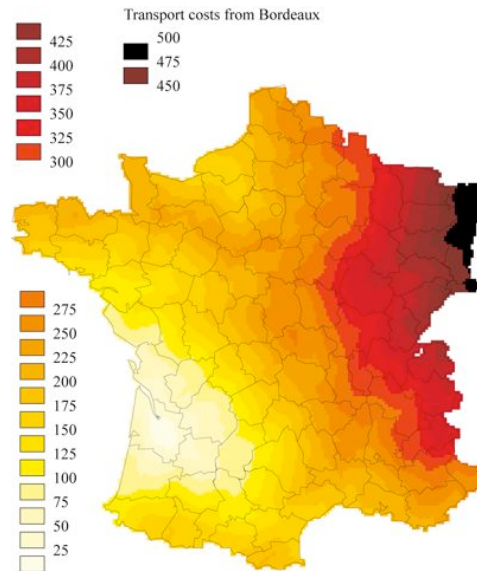
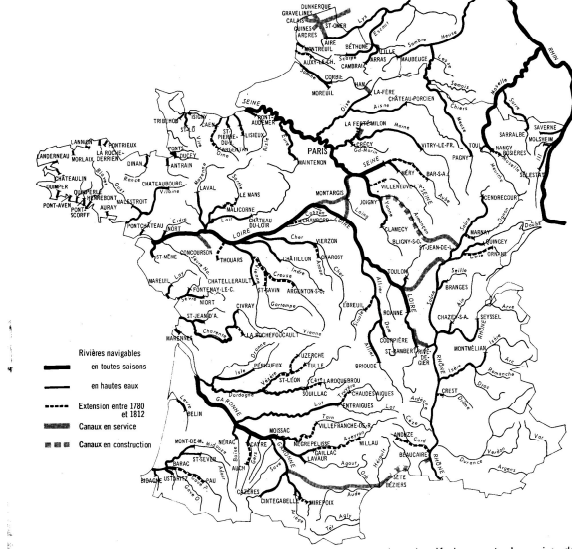
Gérée avec
FileMaker 9

Première approche: équation de gravité

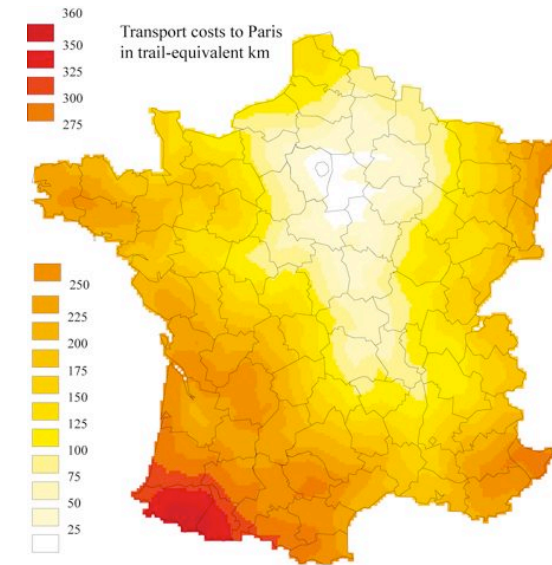
- Quel objectif ?
 - Comparer la taille du marché français et britannique
- Quelle méthode ?
 - Existence d'un lien déterminé par :
 - La nature du bien (rapport poids/valeur)
 - Les coûts de transport, les droits de douane
 - Caractéristiques des districts consommateurs et producteurs
 - On peut étendre les relations observées à celles des districts consommateurs sans information
- Quels outils ?
 - Calcul des coûts de transport : UCINET (sur les réseaux)
 - Calcul des relations (régressions logistiques) et extrapolation: STATA

Calculer les coûts de transport

LE RÉSEAU GÉNÉRAL DES VOIES NAVIGABLES, 1780-1812

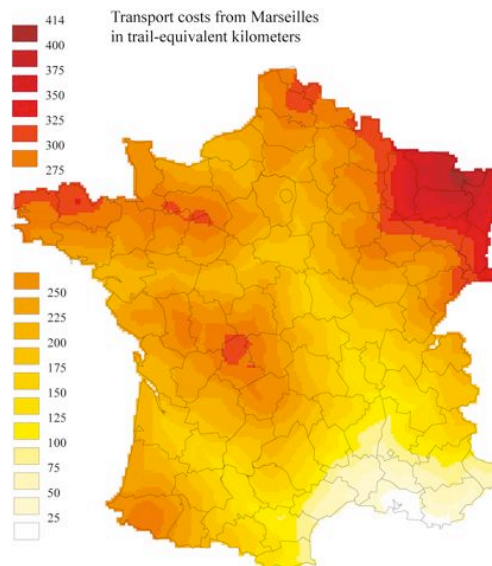


Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin

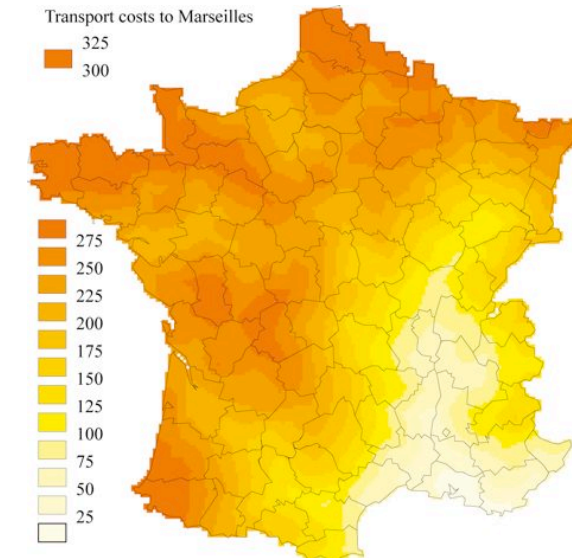


Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin

LES GRANDES VOIES DE COMMUNICATION, 1789



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin

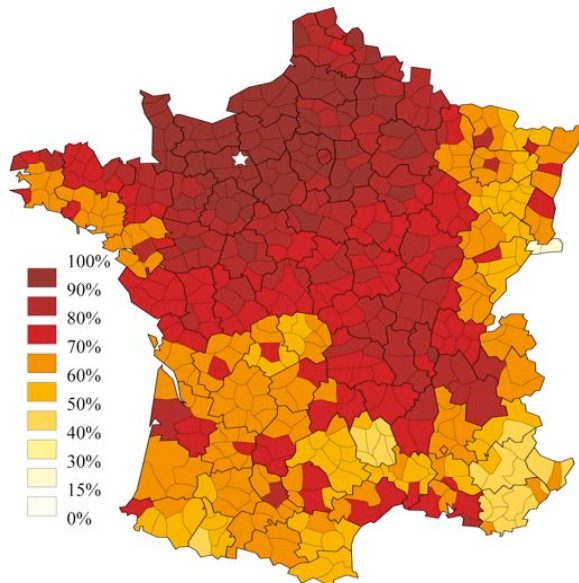
Résultats (1)

- Un des problèmes de l'interprétation des régressions
 - (Ou: espérons que les mauvaises habitudes des économistes ne vont pas se répandre!)
 - Significativité statistique examinée, mais pas la significativité économique

Chg des coûts de transport: +100%	Papier
Modification du rapport des probabilités	-89%
Nouvelle probabilité si proba. initiale = 50%	9,7%
Nouvelle probabilité si proba. initiale = 10%	1,2%

- Puis extrapolation
 - Marchés = population pondérées par les chances qu'un lien existe
 - L'intérêt des régressions: mesure de l'incertitude
 - Outil scientifique (et rhétorique !) important
- Ce qui n'est pas là : l'économétrie spatiale

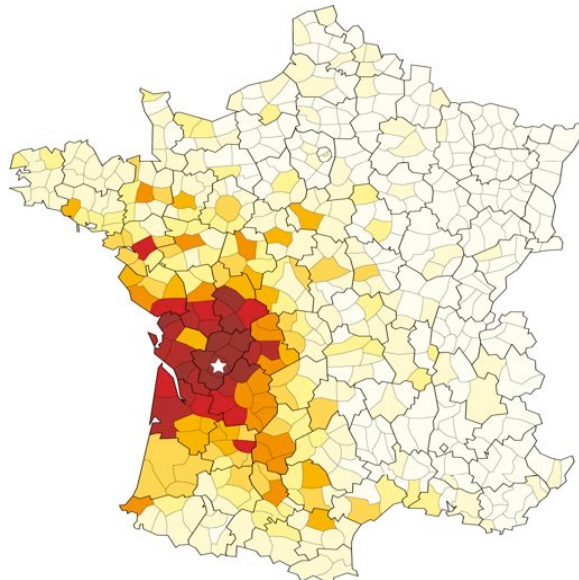
Résultats (2)



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin

Marchés pour la quincaillerie

Saint-Étienne (Loire)	22.4/ 24.9 /26.5
L'Aigle (Orne)	19.3/ 21.9 /23.9
Paris (Seine)	17.0/ 19.9 /22.4
Thiers (Puy-de-Dôme)	16.4/ 19.4 /22.0
Rouen (Seine-Inférieure)	14.4/ 17.4 /20.1



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin

Marchés pour le papier

Angoulême (Charente)	5.4/ 7.7 /10.6
Tournon (Ardèche)	2.2/ 3.7 /6.0
Rouen (Seine-Inférieure)	1.6/ 3.1 /5.7
Thiers (Puy-de-Dôme)	1.5/ 2.7 /5.0
Montargis (Loiret)	1.0/ 2.3 /4.6

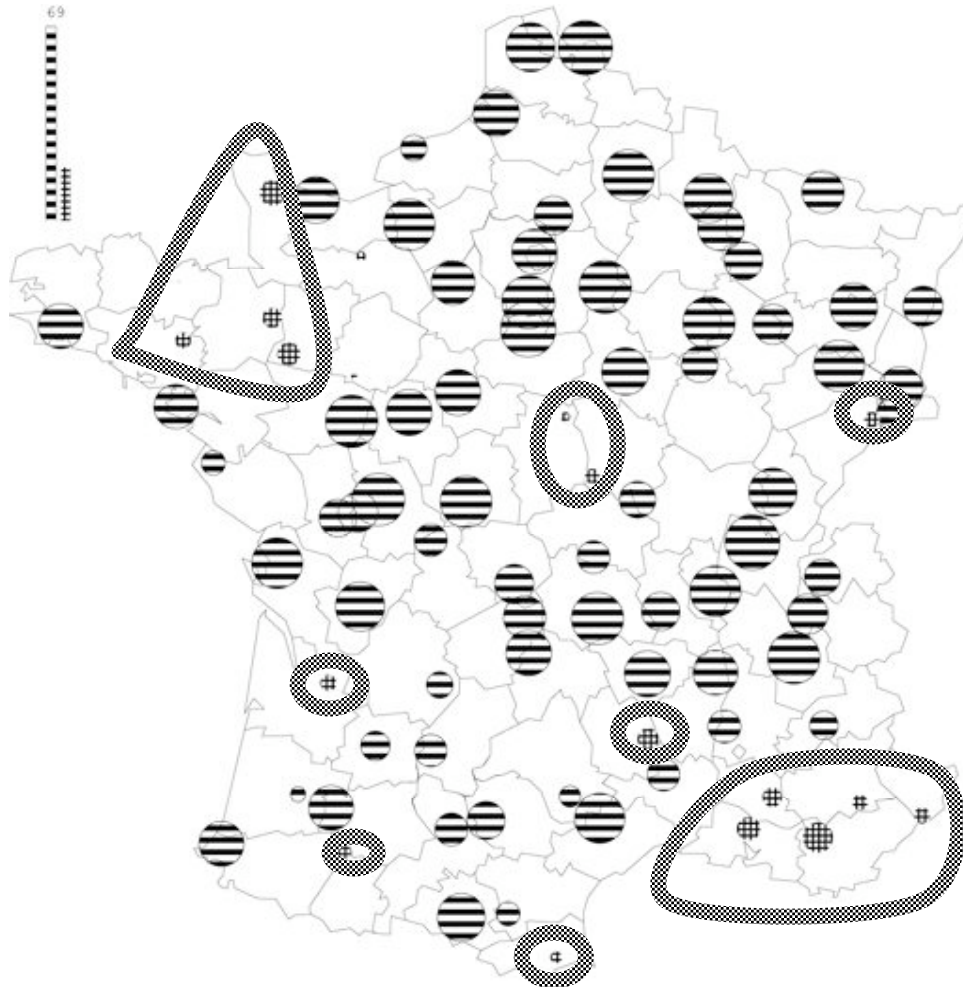
Seconde approche: méthodes de classification

- Quel objectif ?
 - Identifier la géographie des marchés régionaux
 - À partir des réflexions de Christaller et Lösch
 - Les membres d'un marché ont le même type de consommations...
 - ... et font partie d'un ensemble spatialement continu
- Quels outils ?
 - Donc on peut identifier un marché par la similarité des zones de ravitaillement
 - Ce que fait la classification : identifie des groupes d'individus similaires
 - Mais il faut vérifier la continuité... Utilité de la cartographie

Mes angoisses

- Quelle méthode?
 - Nombreuses mesures de distance (ici Jaccard)
 - Nombreuses méthodes pour déterminer la partition pour un nombre donné de groupes (méthode hiérarchique / de partition ; déterministes ou probabilistes)
 - En plus, j'ai appris hier l'existence de méthodes que je ne connaissais pas... (ici k-mean déterministe)
 - Nombreuses méthodes pour déterminer la partition optimum (ici silhouettes)
 - Nombreuses méthodes pour rapprocher les partitions (ici, Rand)
- Circonstances aggravantes
 - Données binaires plutôt que quantitatives
 - Pas de mesure de l'incertitude autour des classifications
- Comment faire ?
 - Raisonement a priori ? (Peut-on le faire sans base théorique ?)
 - Test de robustesse par la multiplication des méthodes ?
 - Critère de choix lié à la question (il faut en avoir une...)
 - La « pire » méthode pour son hypothèse ?

Biens divers (légers) + quincaillerie



Légende:

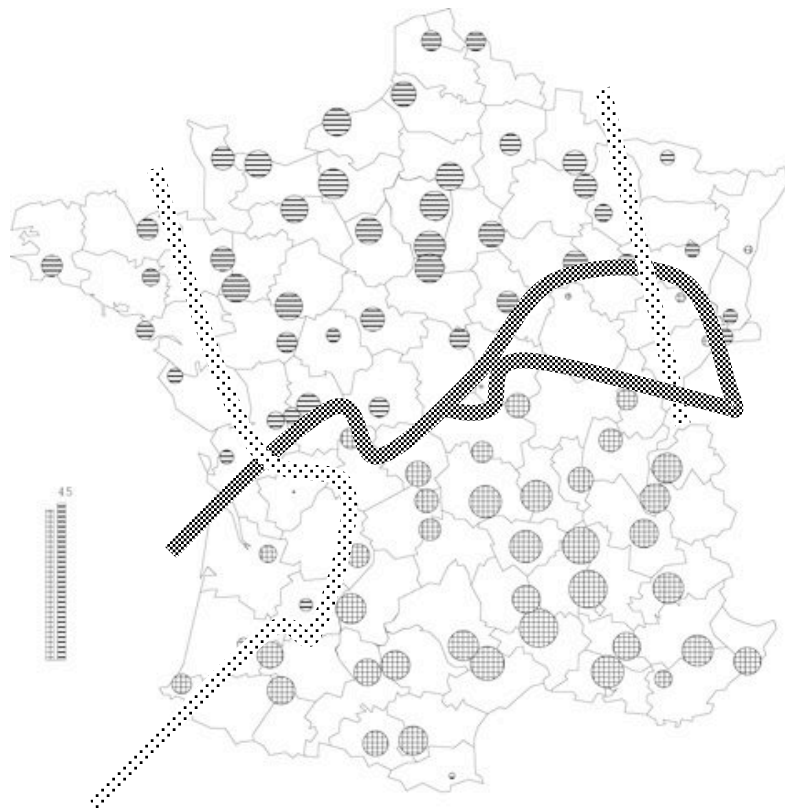
Un disque par consommateur

Couleur: groupe

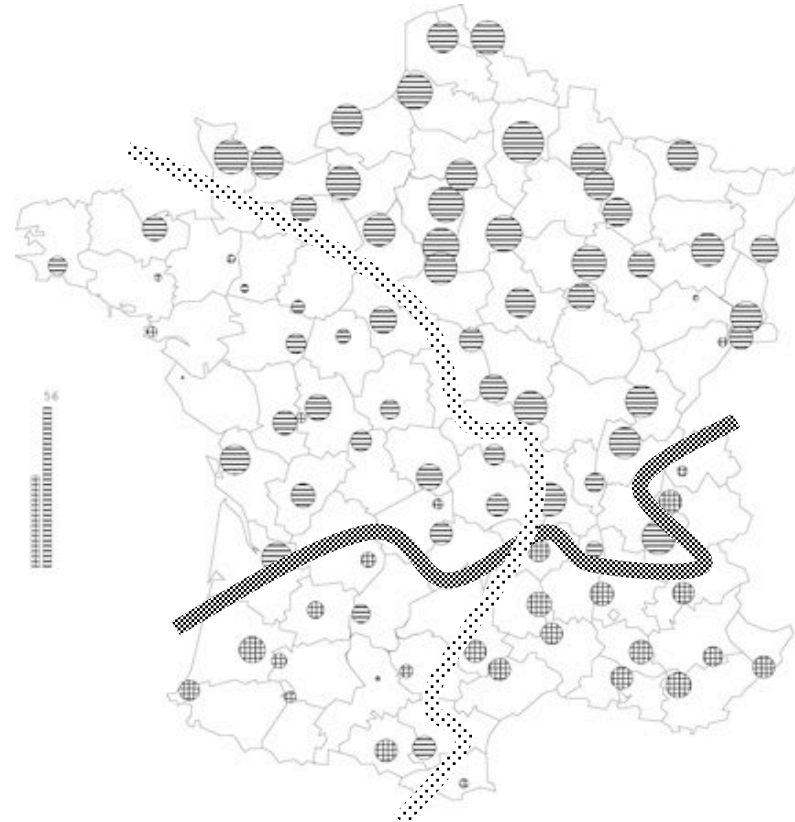
Taille: qualité de la
classification

Interprétation:

Marché unifié : le “deuxième
marché” est mal défini et sans
cohérence



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin

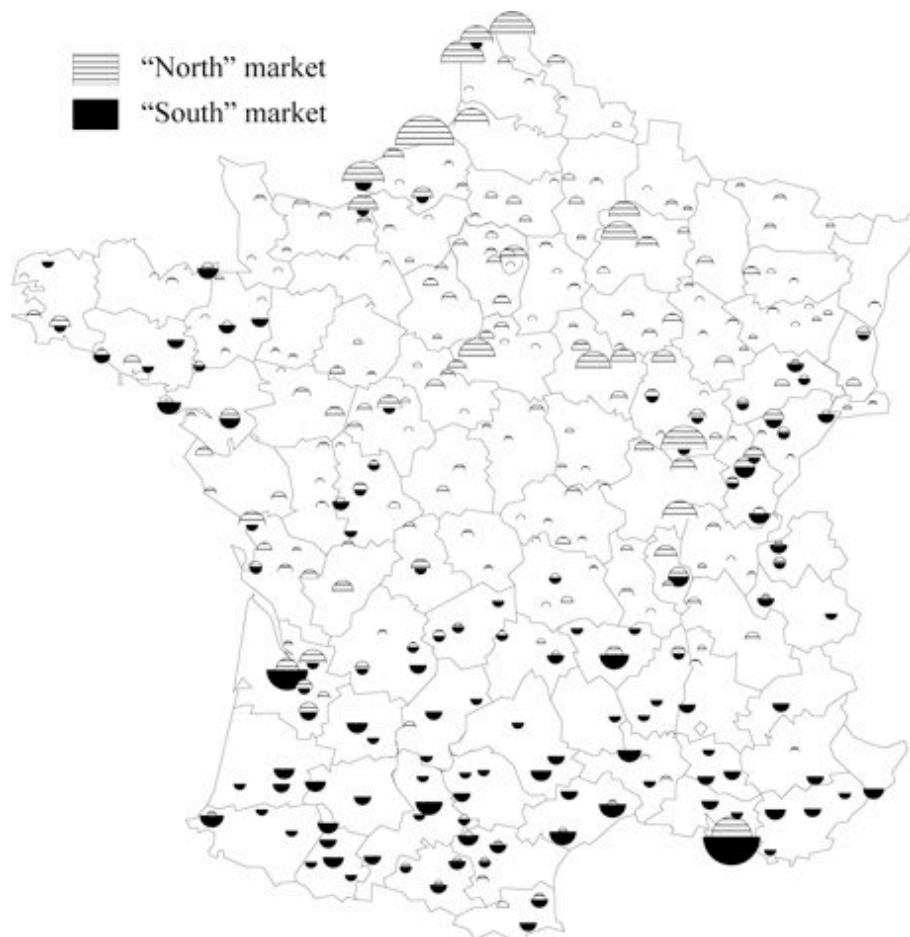
- Gauche : textiles; Droite : nourriture et boissons
- Opposition Nord / Sud
 - Nord organisé autour de la Seine
 - Sud centré sur le bas Rhône
- L'Atlantique, la Loire et la Garonne sont mal liés à l'un et à l'autre
 - Groupe suivant: la Garonne

Conclusion

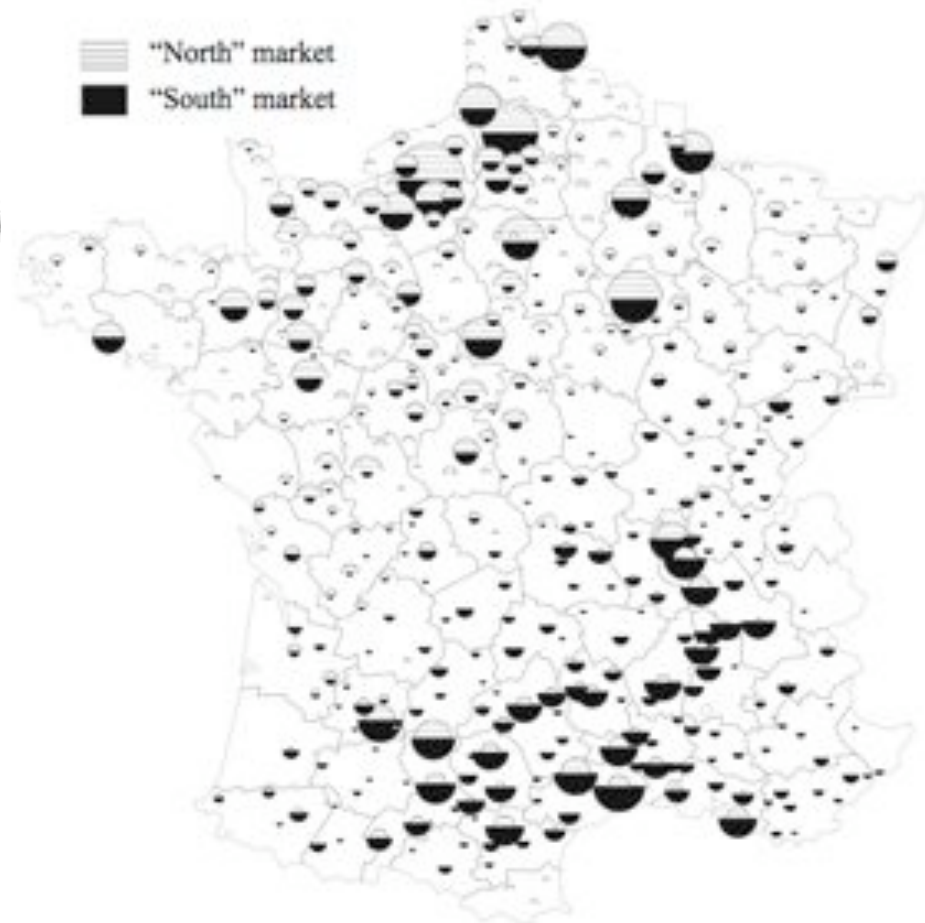
- Ce que j'ai essayé de faire
 - À partir d'une source exceptionnelle
 - Certaines méthodes qui me (nous) sont familières (régressions logistiques)
 - Et à d'autres qui le sont moins (classification)
 - Aidé par des outils simples de cartographie statistique
- Et alors ?
 - La significativité non-statistique, c'est important
 - Mes angoisses sur le choix des méthodes de classification
 - Le problème de la mesure du doute
- Et à propos...
 - Oui, les marchés français étaient plus étendus que les britanniques à la fin du XVIIIe siècle

Différencier les marchés

- Carte de la consommation de chaque marché
- Gauche: nourriture ; droite : textiles



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin



Fait avec Philcarto - <http://perso.club-internet.fr/philgeo>
Guillaume Daudin