

Rapport de stage

Version 2.0

Comportements d'Achats alimentaires et Performances de l'Agriculture Biologique

Mise en forme des données des fichiers Kantar



Maitres de stage :

Sylvette MONIER-DILHAN

Elise MAIGNE

Marion DESQUILBET

L3 SID – Année universitaire 2014/2015

Mylène MILLES

Responsable universitaire :

Wahiba BAHOUN



Historique du document

Date	N° version	Chapitres concernés	Cause de modification
26/05/2015	V 1.0	Tous	Création du rapport
02/06/2015	V 1.1	Chapitre 4	Ajout de la partie concernant la préparation des données
06/07/2015	V1.2	Chapitre 5	Ajout de la partie concernant les résultats
22/07/2015	V 2	Tous	Finalisation du rapport après lecture par les responsables de stage

Remerciements

J'adresse, en tout premier lieu, mes sincères remerciements à **Sylvette MONIER -DILHAN, Elise MAIGNE** et **Marion DESQUILBET** mes tutrices de stage pour leur accueil au sein de l'INRA, pour leurs partages des connaissances et pour m'avoir assuré les conditions optimales au bon déroulement de mon stage.

Elles ont su être disponible à mes côtés, trouver le moyen d'être à mon écoute tout au long de mon stage et répondre à mes différentes interrogations et sollicitations. Elles ont su m'orienter et m'apprendre à être autonome et à prendre des initiatives, dans un cadre prescrit, en respectant les consignes mais en laissant une place à mon expression personnelle.

Je souhaite aussi remercier **Eric CAHUZAC**, directeur de l'unité au sein de laquelle j'ai effectué mon stage, de m'avoir accueilli pour ces trois mois.

Je remercie également la formation SID, de l'université Paul Sabatier de Toulouse pour leurs enseignements sans lesquels réaliser ce stage n'aurait pas été possible. Je souhaite plus précisément remercier **Wahiba BAHSOUN**.

Je remercie encore, bien évidemment toute l'équipe de l'ODR pour leur patience, leur disponibilité et leur partage de connaissances et d'expériences avec moi. Ils ont tous eu à cœur de me faire découvrir leurs métiers, leurs missions et qui ont toujours été très ouverts et accueillants.

Je tiens aussi à remercier Matthieu CARLI-BASSET avec qui j'ai travaillé tout au long de ce stage.

Résumé

Dans le cadre de ma troisième année de licence SID, Statistique et Informatique Décisionnelle de Toulouse, j'ai effectué mon stage de fin d'études à l'INRA (Institut National de la Recherche Agronomique). J'ai travaillé au sein de l'unité de service Observatoire du Développement Rural.

Ce stage s'inscrivait dans un projet ayant pour but d'analyser dans quelle mesure les compositions des paniers alimentaires des ménages ont des impacts différents en termes de durabilité sur les plans nutritionnel, économique et environnemental selon l'importance des achats de produits issus de l'agriculture biologique (produits Bio) par les ménages.

L'objectif du stage était d'extraire, de mettre en forme et d'analyser des données sur les achats en grandes et moyennes surfaces de produits issus ou non de l'agriculture biologique.

J'ai donc dans un premier temps fait du traitement de données. Tout d'abord, j'ai associé une catégorie et de façon plus large un groupe alimentaire à chaque produit, reclasser certains produits et enfin supprimé les valeurs aberrantes.

J'ai ensuite travaillé sur des critères des ménages pour choisir les panélistes pertinents pour l'étude.

Enfin, j'ai réalisé des études statistiques comme l'étude de la composition des paniers alimentaires des ménages ou la répartition des achats en fonction de la consommation en bio des ménages.

Tous ces traitements et analyses de données ont été réalisés grâce au logiciel de statistique R.

Au terme de ce stage, les données sont mises en forme et pourront être traitées dans la suite du projet. De plus, les études statistiques réalisées nous ont permis de voir que le groupe fruits et légumes, le groupe viande poisson œuf et le lait sont les trois groupes les plus achetés en bio. Dans le groupe viande poisson œuf, les œufs se distinguent des autres catégories avec plus 50 % de la part des dépenses en bio dans ce groupe.

Abstract

In the context of my third year of license SID, statistics and business intelligence, I did an internship at the INRA (...). I worked in the service unit ODR (Observatory Rural Development).

This internship was part of a project which aim was to analyze how the compositions of food basket of households have different impacts in terms of durability on terms of nutritional, economic and environmental, depending on the importance of AB products purchased (organic products) by households.

The aim of the internship was to extract, get in shape and analyze data on purchases in supermarkets and hypermarkets, from organic agriculture or not.

So, in the first time, I did data process. To start with, I associated a food category and more broadly a food group for each product, then I separated files, finally, I deleted outliers. Then, I worked on criteria to select panelists that I would keep in the study.

Finally, I realized statistic analysis like the study of the composition of food basket or distribution of household purchases based on the organic consumption of households.

All these treatments and data analysis were realized using the statistical software R.

At the end of the internship, the data are formatted and can be analyzed. Moreover, statistics studies have shown that the fruit and vegetable group, the meat, fish and egg group and the milk are the three most bought in organic agriculture. In meat fish egg group, eggs are distinguished from other categories with more than 50% on the expenditure part in organic agriculture in this group.

Table des matières

Historique du document	3
Remerciements	4
Résumé	5
Abstract	6
Table des figures.....	9
Liste des tableaux.....	9
Introduction.....	10
1. L'INRA	11
Présentation générale	11
Le centre Toulouse Midi-Pyrénées.....	12
Présentation de l'Observatoire du Développement Rural (l'ODR)	13
2. L'étude.....	15
Présentation	15
Kantar	16
3. Les données.....	17
4. Préparation des données	19
Catégorisation des produits	20
Création de variables.....	21
Découpage de tables	22
Suppression des valeurs aberrantes.....	22
Sélection des ménages	26
Type de consommation	26
Fréquence de déclaration.....	27
5. Résultats de l'étude.....	30
Analyse des parts de marché	30
Dépense Bio.....	31
Par type d'enseignes.....	31
Par groupe	32
Par groupes en fonction des enseignes.....	32
Composition des paniers	33
Répartition des dépenses par groupes.....	33
Répartition des dépenses par catégories	35
Saisonnalité	36

Pourcentage de bio	37
Bilan	39
Conclusion	39
Annexes 1 : Explication du découpage des tables.....	40
Annexe 2 : Schéma Graphviz	42
Annexe 3 : Part des dépenses en fonction de la dépense en produits bio	43

Table des figures

Figure 1 : Organigramme de l'Unité de Service Observatoire du Développement Rural	14
Figure 2 : Répartition des achats bio par types de réseau de distribution Source : Agence Bio	15
Figure 3 : Extrait du tableau de donnée d'un fichier produit	18
Figure 4 : Extrait code R	19
Figure 5: Extrait code R, fonction Grubbs	23
Figure 6 : Description d'une boîte à moustache	24
Figure 7 : Répartition des prix du fichier tomate	25
Figure 8 : Extrait code R	28
Figure 9 : Pourcentage de dépense bio par type d'enseignes	31
Figure 10 : Pourcentage de dépense bio par groupes de produit	32
Figure 11 : Répartition des dépenses bio par groupes en fonction des enseignes	32
Figure 12 : Répartition des dépenses totales par groupes	33
Figure 13 : Répartition des dépenses bio par groupes	33
Figure 14 : Répartition des dépenses totale et bio au sein du groupe viande poisson œuf	34
Figure 15 : Répartition par mois des dépenses totales des tomates	36
Figure 16 : Répartition par mois des dépenses totales des tomates	36
Figure 17 : Répartition des dépenses pour les ménages dépensant « un peu » de produit bio	38
Figure 18 : Répartition des dépenses pour les ménages dépensant « beaucoup » de produit bio	38

Liste des tableaux

Tableau 1 : Tableau des groupes et catégories	20
Tableau 2 : Tableau des unités	22
Tableau 3 : Statistiques descriptives sur la variable semaine d'activité	27
Tableau 4 : Statistiques sur les ménages en fonction des semaines d'activité	27
Tableau 5 : Extrait tableau des variables Semaine d'activité.	28
Tableau 6 : Statistiques part de marché	30
Tableau 7 : Part de marché en valeur et en quantité bio par produits	30
Tableau 8 : Répartition des dépenses totale et bio par catégories	35
Tableau 9 : Tableau de statistiques par intervalles de pourcentage de bio	37

Introduction

Mon stage de fin de troisième année de licence Statistique et Informatique Décisionnelle (SID) de l'université Paul Sabatier, à Toulouse s'est déroulé au sein de l'Institut National de la Recherche Agronomique de Toulouse, sur le site d'Auzeville Tolosane, dans l'unité de service Observatoire du développement rural (ODR). La durée de ce stage a été de 3 mois, du 27 Avril 2015 au 24 Juillet 2015.

La tendance de ces dernières années, c'est le bio. Il fait partie intégrante de notre quotidien. Mais quels produits sont achetés ? Par qui ? A quel prix ? C'est à ces questions que le projet dans lequel mon stage s'inscrit cherche à répondre.

En effet, l'objectif du projet est d'analyser dans quelle mesure les compositions des paniers alimentaires des ménages ont des impacts différents en termes de durabilité sur les plans nutritionnel, économique et environnemental selon l'importance des achats de produits de l'AB (produits Bio) par les ménages.

L'objectif de ce stage était de préparer des données pour par la suite, faire des études statistiques. C'est ce que je vais vous présenter dans ce document.

Dans un premier temps, je présenterai l'INRA en général puis le site de Toulouse et l'unité dans laquelle j'ai travaillé ainsi que le sujet du stage. Dans un deuxième temps, j'exposerai les données et les traitements réalisés. Enfin, avant de conclure, j'analyserai les résultats statistiques obtenus.

Tous les traitements et analyses de données ont été réalisés grâce au logiciel de statistique R.

1. L'INRA

Présentation générale

L'Institut National de la Recherche Agronomique (Inra) est un organisme de recherche scientifique publique, dépendant du ministère délégué à l'Enseignement supérieur et à la Recherche et du ministère de l'Agriculture et de la Pêche

Ses recherches concernent les questions liées à l'agriculture, à l'alimentation et à la sécurité des aliments, à l'environnement et à la gestion des territoires, avec un accent tout particulier en faveur du développement durable

L'Inra a été fondé en 1946, il est aujourd'hui le premier institut européen de recherche agronomique et le deuxième dans le monde en nombre de publications en sciences agricoles et en sciences de la plante et de l'animal

Ses missions sont les suivantes :

- Œuvrer au service de l'intérêt public tout en maintenant l'équilibre entre les exigences de la recherche et les demandes de la société ;
- Produire et diffuser des connaissances scientifiques et des innovations, principalement dans les domaines de l'agriculture, de l'alimentation et de l'environnement ;
- Contribuer à l'expertise, à la formation, à la promotion de la culture scientifique et technique, au débat science/société.



Les recherches de l'INRA s'articulent autour de différentes thématiques réparties au sein de 13 départements scientifiques :

- Environnement et agronomie
- Écologie des forêts, prairies et milieux aquatiques
- Alimentation humaine
- Biologie et amélioration des plantes
- Caractérisation et élaboration des produits issus de l'agriculture
- Génétique animale
- Mathématiques et informatique appliquées
- Microbiologie et chaîne alimentaire
- Physiologie animale et systèmes d'élevage
- Santé animale
- Santé des plantes et environnement
- Sciences pour l'action et le développement
- Sciences sociales, agriculture et alimentation, espace et environnement

Le centre Toulouse Midi-Pyrénées

Le centre Toulouse Midi-Pyrénées fait partie d'un des 17 sites INRA en France.

Son activité est structurée par cinq domaines de recherches tous en partenariat étroit avec d'autres établissements d'enseignement supérieur et de recherche :

1. Génome et amélioration des productions
2. Sécurité sanitaire des aliments
3. Transformation des produits
4. Environnement, territoire et société
5. Economie de l'environnement et des Marchés

L'INRA est organisé en « unités » Le centre de Toulouse Midi-Pyrénées comprend :

- 20 Unités de recherche dont 14 Unités Mixtes
- 5 Unités expérimentales
- 1 Unité d'appui à la recherche

Ces unités sont réparties sur 5 sites géographiques autour de Toulouse.

Avec plus de 850 chercheurs, ingénieurs et techniciens INRA, dont 600 titulaires, le centre de Toulouse Midi-Pyrénées représente environ 10% des publications et près de 12% des brevets de l'INRA.

Présentation de l'Observatoire du Développement Rural (l'ODR)

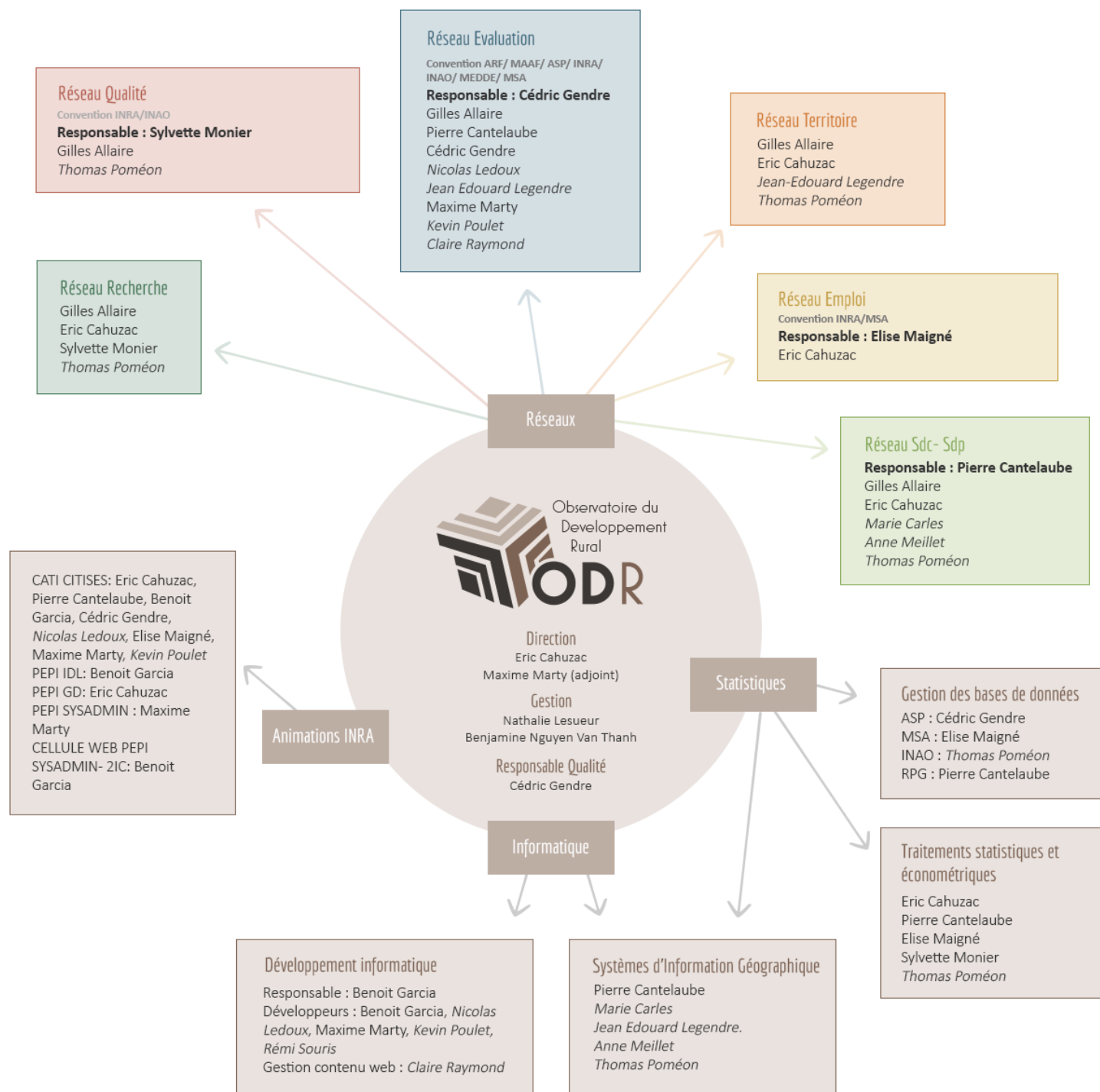
L'Unité de service ODR, située sur le site d'Auzeville et créée le 01/01/2009, est un centre de ressources pour l'évaluation des politiques de développement rural et pour la recherche agronomique et en sciences du développement. Cette unité est composée de chercheurs et d'ingénieurs.

Elle a pour mission principale de maintenir, développer et gérer une plateforme de données. Cette plateforme assure des services pour des unités de recherches et des partenaires publics, gestionnaires de la politique de développement rural.

L'ODR donne accès pour des utilisateurs agréés à des informations et des indicateurs sur l'évolution de l'agriculture (emploi, structures, productions) et de l'économie des territoires ruraux, sur les politiques agricoles et rurales, sur l'agroenvironnement, sur les signes de qualité, etc.



La figure qui suit représente l'organisation de l'Unité de Service Observatoire du Développement Rural.



* Les CDD sont en italiques dans le schéma.

Figure 1 : Organigramme de l'Unité de Service Observatoire du Développement Rural

2. L'étude

Présentation

Ce stage est la première étape d'un projet de l'équipe de recherche qui se déroulera sur plusieurs années.

L'objectif de ce projet est de décrire et quantifier la composition des paniers d'achat des ménages en tenant compte des produits issus de l'agriculture biologique et de l'agriculture conventionnelle. Mais aussi d'étudier dans quelle mesure les compositions de ces paniers ont des impacts différents en termes de durabilité sur les plans nutritionnel, économique et environnemental selon l'importance des achats de produits Bio par les ménages.

L'analyse est réalisée dans deux réseaux de distribution : les grandes ou moyennes surfaces (GMS) et les magasins coopératifs spécialisés dans la vente de produits Bio

Ces données sont issues de deux sources différentes :

- Worldpanel Kantar pour les achats en GMS
- Une base de données de magasins spécialisés pour les achats en réseaux spécialisés bio

Ces bases permettent de couvrir les principaux réseaux de distribution des produits Bio.

En effet, 46% du total des achats bio effectués en France sont effectués en Grandes et Moyennes Surfaces (GMS) et 28 % en distribution spécialisée Bio¹. Parmi les achats en réseaux spécialisés, 50% sont effectués par le réseau étudié dans l'autre stage.

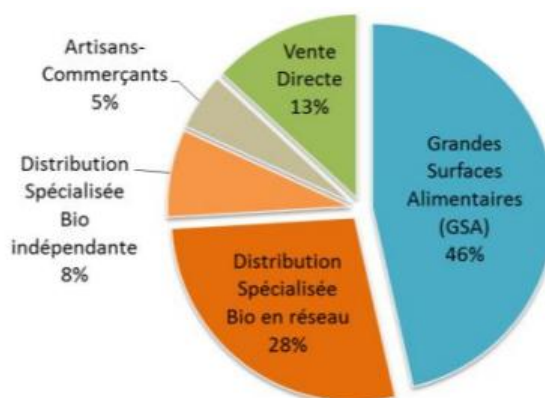


Figure 2 : Répartition des achats bio par types de réseau de distribution Source : Agence Bio

¹

Source : http://www.agencebio.org/sites/default/files/upload/documents/4_Chiffres/BrochureCC/CC2014_chap4_France.pdf

Afin d'étudier ces deux bases de données, deux stages ont été mis en place, d'une part le stage de Matthieu Carli-Basset qui consistait à traiter les données de magasins spécialisés, d'autre part le mien qui avait pour but de traiter les données d'achats des ménages en grandes et moyennes surfaces.

L'objectif du stage est de préparer les données et de mener les premières études statistiques.

De façon plus précise, l'objectif de mon stage a été de préparer les fichiers Kantar et d'étudier la composition des paniers de consommation.

Ces préparations vont nous permettre d'étudier la composition des paniers, en grandes familles de produits (viande, fruits et légumes, céréales, produits laitiers, ...). Une analyse du coût des paniers sera menée pour déterminer dans quelle mesure la composition des paniers intégrant du Bio, et notamment leurs teneurs en viandes et en produits céréaliers, est en mesure de compenser (ou non) le surcoût lié au Bio.

Kantar

Kantar Worldpanel est une entreprise de compréhension et d'analyse du comportement des consommateurs grâce à des panels de consommateurs.

Kantar Worldpanel sélectionne plusieurs milliers de personnes en France selon des critères, pour représenter des groupes de population. Notre panel est composé de personnes ayant des fréquences d'achats différentes : des clients très réguliers ou des clients occasionnels.

Chaque individu participant fournit des informations sur ses achats de produits alimentaires, de toilette, d'hygiène-beauté, ses vêtements, chaussures, accessoires de mode, sa téléphonie mobile, ... Pour notre étude, seuls les produits alimentaires sont traités.



3. Les données

Les données Kantar comprennent :

- Des fichiers produits :

Il y a au total 342 fichiers. Un fichier correspond à tous les actes d'achats effectués pour un type de produit (...). Les types de produits sont plus ou moins homogènes : par exemple le fichier banane est composé uniquement d'achats de banane alors que le fichier boisson sucrée est composé de jus de fruits, de sodas, ... Un fichier produit peut contenir entre 500 et 600 000 lignes.

Pour chaque fichier produit :

- Une observation correspond à un achat,
- Les variables nous donnent des caractéristiques :
 - D'achat :
 - L'identifiant du ménage (nopnlNF)
 - Le type de circuit de vente dans lequel l'achat a été effectué (cvwp)
 - La date de d'achat (dtwa)
 - Le prix de d'achat en euros (Ptwa)
 - Prix unitaire en euros (Pu)
 - Le nombre d'achat d'un même produit lors d'un même achat (Qawa)
 - Le nom du produit (Sa1)
 - ...
 - Du produit :
 - Permet de savoir si le produit est bio ou non (Spe90001)
 - Le type de conditionnement
 - L'origine du produit

Ces variables sont dites « Spécifiques ». Elles sont propres à certains types de produits. Par exemple, la variable qui nous renseigne sur le type de lait ne va être présente que pour les produits de type laitier.

Il y a environ 200 variables par fichier, mais ce nombre n'est pas fixe (variable « Spécifiques »).

- Un fichier pour les ménages :

Le fichier est composé de 22000 lignes. Chaque ligne correspond à un ménage faisant partie du panel de consommateurs.

Les variables sont les caractéristiques du ménage. Elles nous renseignent sur l'âge, la catégorie socioprofessionnelle, le nombre d'enfant, le revenu, année d'étude, type d'habitation ...

J'ai pu faire des jointures entre ces différentes tables de données grâce à l'identifiant du ménage (nopnltnf), en utilisant la fonction 'merge' dans R.

J'avais à disposition les données Kantar pour les années 2011 et 2012. Le but étant de comparer les données Kantar et celles issus de magasins spécialisés, nous avons travaillé sur la période commune aux deux sources de données : l'année 2012.

Voici un extrait d'un fichier produit.

nopnltnf	dtwa	cvwp	...	ptwa	qorig	tuwa	sa1	spe90001
201	30/06/2012	hypermarches chariot	...	2,958	488	poids total en grs	417_viandh ac preb se	non
450	24/04/2012	hypermarches chariot	...	5,222	867	poids total en grs	417_viandh ac preb se	non
320	14/12/2012	hypermarches chariot	...	1,714	125	poids total en grs	417_viandh ac preb se	non
511	01/03/2012	hypermarches chariot	...	1,295	300	poids total en grs	417_viandh ac preb se	non
484	24/10/2012	hypermarches chariot	...	3,939	350	poids total en grs	417_viandh ac preb se	non
654	07/04/2012	hypermarches chariot	...	31,24 7	2025	poids total en grs	417_viandh ac preb se	non

Figure 3 : Extrait du tableau de donnée d'un fichier produit

4. Préparation des données

Certaines modifications ont donc été nécessaires pour pouvoir exploiter les données et avoir des fichiers le plus homogènes possibles en terme de produit.

En effet, compte tenu du nombre de fichiers, on souhaite faire les analyses en faisant des boucles les plus automatiques possibles, et ainsi pouvoir exécuter les programmes et faire des statistiques sans se soucier des spécificités des produits.

J'ai dans un premier temps catégorisé les produits, puis créé de nouvelles variables.

J'ai ensuite séparé le contenu de certaines tables et supprimé des valeurs aberrantes.

Et enfin, j'ai fait une sélection sur nos données de façon à ne conserver que les ménages pertinents pour l'étude.

Comme on vient de le voir dans la présentation des données, je disposais de 342 fichiers produits. Tous ces fichiers sont répertoriés dans un même dossier.

Pour éviter de lancer les programmes de créations de variables, de statistiques 342 fois, j'ai créé une boucle qui parcourt et applique les traitements souhaités à chaque fichier présent dans le dossier.

Voici l'extrait d'un programme qui permet d'afficher le nom et le nombre de ligne pour chaque fichier.

```
1 setwd("C:/Users/milles/Desktop/Mylène/Stage/DONNEES/NV_DONNEE/Rdata")
2 fic = dir(".")
3
4 for (i in 1:length(fic)) {
5   if (substr(fic[i],1,1) == "p")
6   {
7     tab = readRDS(fic[i])
8     print (fic[i])
9     print (nrow(fic[i]))
10  }
11 }
```

Figure 4 : Extrait code R

Les deux premières lignes de ce programme permettent de récupérer les fichiers présents dans le dossier « Rdata » et de les mettre dans une liste.

A la ligne 4, le « for » permet de parcourir chaque élément de la liste créée auparavant.

Le « if » à la ligne suivante permet de vérifier une condition. Ici, le fichier devait commencer par la lettre "p".

Les lignes 7, 8 et 9 correspondent aux instructions appliquées à chaque fichier. Tout d'abord on ouvre « readRDS », puis on affiche son nom et enfin son nombre de lignes.

Catégorisation des produits

Pour les études futures (la répartition des achats par type de produit par exemple), il a été choisi, avec les autres membres du projet, de regrouper les produits en catégorie de produits alimentaires puis, de façon plus large, en groupes.

Il y a 35 catégories de produits alimentaires et 14 groupes présentés dans le tableau suivant.

Groupe alimentaire	Catégorie de produit	Groupe alimentaire	Catégorie de produit
Alcool	Alcool	Produit laitier végétal	Produit laitier végétal
Aliment bébé	Aliment bébé	Snack salé	Condiment
Boissons non sucrées	Café		Fruits salés
	Eau		Sauce
	The		Snack salé
Boissons sucrées	Boissons sucrées	Snack sucré	Autre snack sucré
Féculent	Féculent		Compote
Fruit et légume	Fruit		Confiture
	Fruits secs		Pâtisserie
	Jus de fruit		Sucre et miel
	Légume	Viande Poisson Œuf	Crustacé
Gras	Gras animal		Œuf
	Gras végétal		Poisson
Lait	Lait		Porc
Plat mixte	Plat mixte		Ruminant
Produit laitier	Autre produit laitier		Viande NR (Non renseignée)
	Fromage		Volaille
	Yaourt		

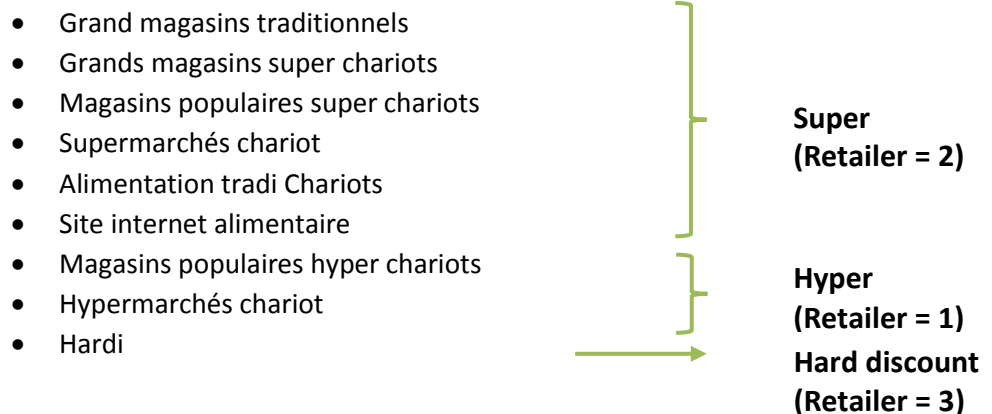
Tableau 1 : Tableau des groupes et catégories

A chaque produit a donc été associée une catégorie. Cette catégorie était évidente pour certain produit (par exemple la catégorie fruit pour le produit banane), mais moins pour d'autres (par exemple le produit « pâte rilette ». Nous ne savions pas s'il s'agissait de viande de porc, de volaille, de ruminant, ... Il a donc fallu chercher des informations plus précises sur le produit grâce aux variables « spécifiques » pour lui associer une catégorie.

Création de variables

Toujours dans un but d'optimisation, certaines variables ont été ajoutées aux tables existantes . Les variables ajoutées correspondent au :

- Numéro du produit (Num_Produit). Par exemple, le numéro 323 pour la tomate.
- Numéro du groupe de produit (NumGroupe). Par exemple, le numéro 6 pour fruit et légume.
- Numéro de la catégorie du produit (NumCategorie). Par exemple, le numéro 64 pour légume.
- Retailer : Regroupement de la variable circuit de magasin (cvwp). Je l'ai regroupé par type de Grandes et Moyennes Surfaces(GMS). Cette variable a 37 modalités. Nous n'en avons gardé que 8 que nous regroupons en 3 types de GMS :



La variable "Retailer" prend la valeur 1 pour Hyper, 2 pour Super et 3 pour Hard Discount et permet de faire des études en fonction du type de magasin dans lequel a eu lieu l'achat

Les autres circuits de ventes alimentaires, comme les marchés ou les magasins spécialisés, ne sont pas gardés pour l'étude. En effet, ces circuits de vente sont mal renseignés dans Kantar (ils ne comptent que pour 10 % des achats alimentaires dans Kantar alors qu'ils représentent en réalité près du tiers).

Découpage de tables

Certains fichiers de produit correspondent à plusieurs catégories de produits tels que nous les avons définis d'après des résultats d'articles scientifiques écrits par des nutritionnistes. Par exemple, le fichier des boissons (63_BRSA) est composé de sodas et de jus de fruit.

La solution a été de créer des nouveaux fichiers à partir du premier, de façon à avoir un nouveau fichier pour chaque catégorie.

Dans l'exemple, le fichier des boissons a été partagé en deux pour donner lieu à deux nouveaux fichiers « produit » : un fichier allant dans la catégorie boisson sucrées et un autre avec fruits et légumes.

Une difficulté rencontrée est que dans certains fichiers l'unité de mesure de la quantité n'est pas unique ; la quantité pour un même produit peut être exprimée en grammes et en kilogrammes, en litres et en millilitres, voire en milites et en grammes (par exemple, le potage est vendu sous différentes formes : en poudre (déshydraté) ou liquide). De plus certains produits sont vendus à l'unité ou au poids, c'est le cas notamment de certains fruits et légumes (par exemple, les avocats ou l'ananas). Or, pour pouvoir comparer ou analyser ses produits, une unité unique est nécessaire pour pouvoir analyser la consommation des produits en quantité.

La solution a été de créer une nouvelle variable, pour nos produits ayant différentes unités, représentant une unité uniformisée : ml ou g ou de séparer le fichier en deux quand les unités étaient g et ml.

Différents cas se sont posés :

Kg	g	L	mL	Unité	Solution
X	X				Conversion des Kg en g
		X	X		Conversion des L en ml
X		X			<u>Séparation en 2 fichiers</u>
X				X	Conversion des unités en g grâce à une variable spéciale
		X		X	Conversion des unités en ml grâce à une variable spéciale

Tableau 2 : Tableau des unités

Suppression des valeurs aberrantes

Après quelques analyses sur les données, nous nous sommes rendu compte que certaines valeurs de prix étaient extrêmes (par exemple : des tomates vendues à 25 € le kilo).

Dans un premier temps, j'ai appliqué la fonction `Grubbs` de R qui permet de repérer les valeurs aberrantes et de les supprimer (les valeurs en question sont remplacées par « NA »).

Voici la fonction `Grubbs.test` et la fonction créée, `Grubbs.outl` (qui exécute la fonction `grubbs.test` jusqu'à ce qu'il n'y ait plus ces valeurs).

```

2 library(outliers)
3
4 grubbs.outl <- function(VAR) {
5   outliers <- NULL
6   test <- VAR
7   pv <- grubbs.result$p.value
8   while (pv < 0.0000000000000001) {
9     outliers <- c(outliers, as.numeric(strsplit(grubbs.result$alternative, " ")[[1]][3]))
10    test <- VAR[!VAR %in% outliers]
11    if (length(test[!is.na(test)]) < 3 ) {
12      VAR_noutl <- ifelse((VAR %in% outliers)==F, VAR, NA)
13      return(VAR_noutl)
14    }
15    if (mad(test, na.rm=T)==0) {
16      VAR_noutl <- ifelse((VAR %in% outliers)==F, VAR, NA)
17      return(VAR_noutl)
18    }
19    else {grubbs.result <- grubbs.test(test)
20          pv <- grubbs.result$p.value }
21  }
22 }
23 VAR_noutl <- ifelse((VAR %in% outliers)==F, VAR, NA)
24 return(VAR_noutl)
25 }
26

```

Figure 5: Extrait code R, fonction `Grubbs`

Le principe mathématique de cette fonction est le suivant.

Principe du test de Grubbs :

Ce test permet de détecter les valeurs aberrantes en termes de dispersion de moyennes. Le test de Grubbs est un test statistique qui vérifie la variabilité d'un ensemble de valeurs. Il permet sur cet ensemble de valeurs d'identifier l'existence de valeurs aberrantes ou isolées à une valeur critique fixée à α % (α étant souvent fixé à 5 % ou 1 %).

Autrement dit :

$$G = \frac{\max |xi - \bar{x}|}{Sx}$$

\bar{x} : Moyenne des valeurs

xi : Valeur la plus éloignées de la moyenne

Sx : Ecart-type des valeurs

Cette fonction a été appliquée pour produits homogènes. Pour cela, j'ai regroupé à l'intérieur d'un fichier produit, les produits de fonction à ce qu'ils soient uniques en concaténant toutes les variables spécifiques du produit (ex : une « petite » tomate, « rouge », « en vrac », « venant de France »). Ces valeurs aberrantes sont ensuite supprimées.

Sur les graphiques suivants, nous pouvons voir la répartition des prix grâce à des boîtes à moustaches avant (à gauche) et après (à droite) l'application de la fonction Grubbs. Une boîte à moustaches est réalisée pour chaque produit ayant les variables spécifiques identiques.

Une boîte à moustache permet de visualiser la répartition des valeurs de la façon suivante :

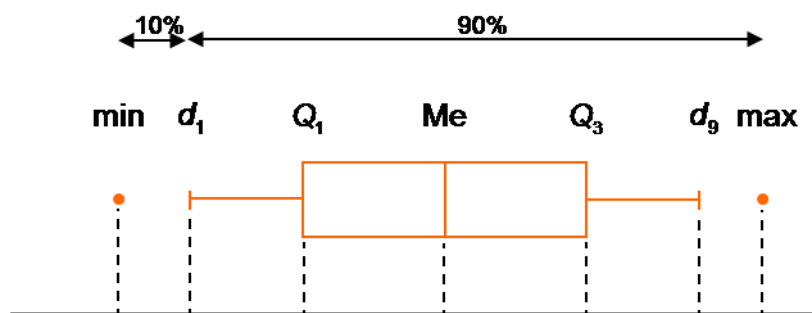
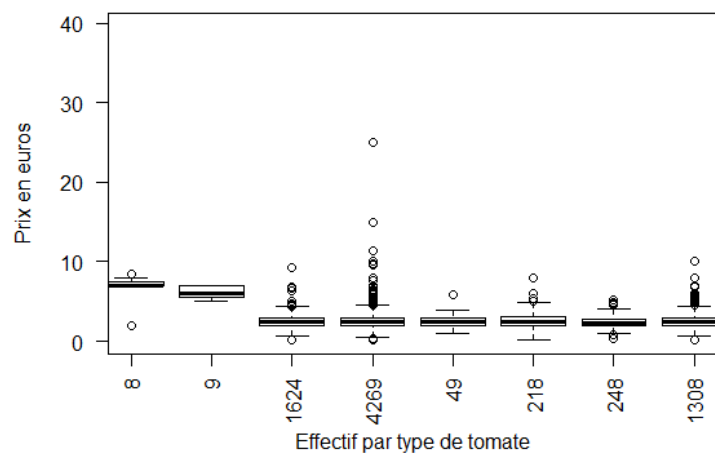


Figure 6 : Description d'une boîte à moustache

En ordonnée est représenté le prix et en abscisse l'effectif. Nous pouvons voir sur la 4ème boîte à moustache en partant de la gauche que les valeurs les plus extrêmes ont été supprimées. L'effectif est passé de 4269 à 4262.

Répartition des prix au kilo par type de tomate
(avant suppression des valeurs aberrantes)



Répartition des prix au kilo par type de tomate
(après suppression des valeurs aberrantes)

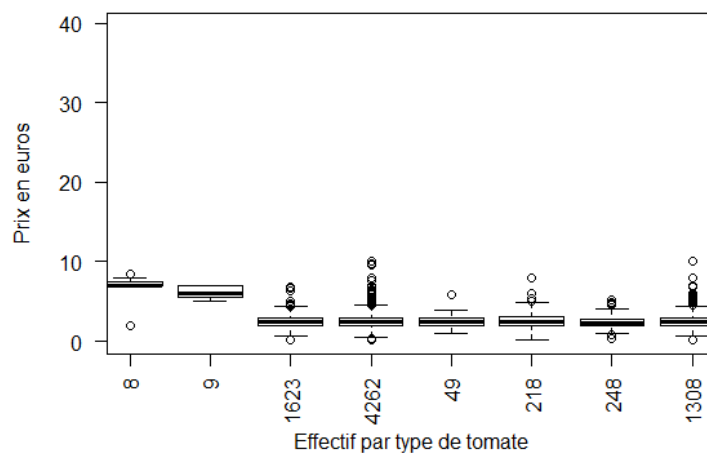


Figure 7 : Répartition des prix du fichier tomate

Le pourcentage minimum d'actes d'achat pour un fichier est de 0 % et le maximum de 0,74 %. La moyenne du pourcentage d'actes d'achat supprimés est de 0,033 %.

Sélection des ménages

Une fois les valeurs aberrantes supprimées, nous avons travaillé sur des critères pour choisir les panélistes gardés dans l'étude. En effet certains panélistes ne déclarent pas que des Produits Frais ou de Grandes Consommations ou encore certains n'ont pas déclaré tous leurs achats. Deux critères de sélection ont donc été appliqués :

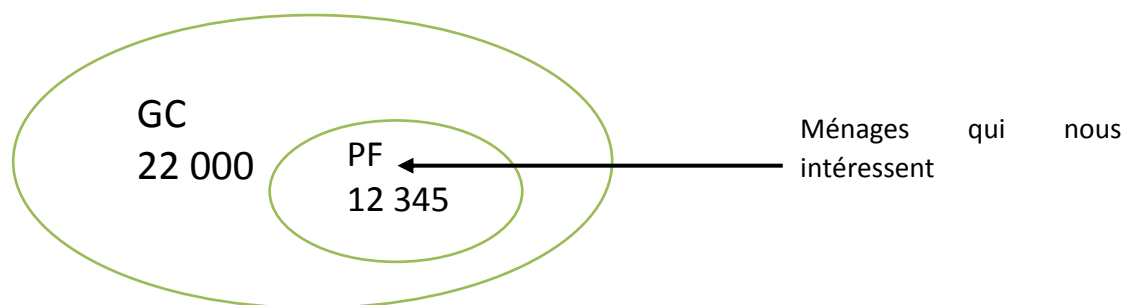
Type de consommation

Notre panel de consommateurs est composé de deux groupes de personnes :

- Les personnes ne devant déclarer que les produits de Grandes Consommation (GC) qu'ils achètent.
- Et les personnes devant déclarer les produits de Grandes Consommation et les Produits Frais (PF) qu'ils achètent.

En gardant ces deux panels, nos résultats seraient biaisés par les consommateurs de produits de Grandes Consommation uniquement. En effet, le poids des produits GC serait potentiellement plus important que le poids des produits PF, alors que notre objectif est d'avoir des consommateurs susceptible d'acheter les deux types de produits.

Etant donné l'objet de l'étude nous avons le panel des consommateurs de produits GC et de produits PF.



Notre jeu de données contient maintenant 12345 ménages.

Fréquence de déclaration

Pour les études statistiques, nous souhaitons travailler sur les ménages qui déclarent systématiquement sinon fréquemment leurs achats. Dans cette partie, nous allons voir comment ces ménages ont été sélectionnés.

Dans les renseignements du ménage, nous disposons d'une variable qui renseigne sur le nombre de semaines d'activités du ménage, c'est-à-dire, le nombre de semaines pendant lesquelles le consommateur a déclaré ses achats.

Pour commencer, j'ai fait des statistiques sur cette variable, qui sont décrites dans les tableaux ci-dessous.

Min	1 ^{er} Quartile	Médiane	3 ^{ème} quartile	Max	Moyenne
1	24	37	45	52	33,52

Tableau 3 : Statistiques descriptives sur la variable semaine d'activité

Nombre de semaines activités	Effectif (nombre de ménages)	Part du nombre ménages (%)	Part des dépenses (%)
Entre 1 et 5	415	3,36	0,37
Entre 6 et 10	738	5,98	1,33
Entre 11 et 15	677	5,48	2,05
Entre 16 et 20	750	6,08	3,34
Entre 21 et 25	866	7,01	4,73
Entre 26 et 30	1 057	8,56	7,05
Entre 31 et 35	1 306	10,58	10,01
Entre 36 et 40	1 587	12,86	13,93
Entre 41 et 45	1 932	15,65	19,82
Entre 46 et 50	2 229	18,06	26,64
Plus de 50	788	6,38	10,72
Total	12 345	100	100

Tableau 4 : Statistiques sur les ménages en fonction des semaines d'activité

Grâce à ces statistiques descriptives, nous pouvons voir que les $\frac{3}{4}$ des ménages ont déclaré leurs achats au moins 24 fois dans l'année, ce qui représente pratiquement une fois toutes les 2 semaines.

En supposant que les consommateurs font leurs courses tous les 15 jours, ces ménages ont une fréquence d'achat satisfaisante. Néanmoins, ces résultats ne nous donnent pas assez d'informations pour conclure.

En effet, un ménage peut avoir participé pendant 24 semaines consécutives. Dans ce cas, il manquerait 6 mois consécutifs d'achat sur une année, et dans le cas d'étude sur la saisonnalité par exemple, les résultats seraient biaisés.

Nous avons donc cherché à voir comment les semaines d'activités des ménages se répartissaient sur l'année.

Pour étudier cette répartition, j'ai créé une variable correspondant à la somme des achats déclarés par un ménage pour chaque semaine de l'année. Il y avait donc 52 variables. J'ai ensuite codifiés ces 52 variables dichotomiques prenant la valeur 1 si la somme des achats déclarée est positive, 0 sinon.

Elles ont ensuite été concaténées de façon à avoir une chaîne de caractères comme dans le tableau 5 ci-dessous.

[illegible]

Tableau 5 : Extrait tableau des variables Semaine d'activité.

Grâce à cette chaîne de caractères, j'ai pu calculer de nouvelles variables comme le nombre de semaines d'inactivité consécutives ou recalculer le nombre de semaines d'activité

Par exemple, le code (Figure 8) suivant permet de recalculer le nombre de semaine d'activité. Son principe est de repérer grâce à une expression régulière tous les "1" présents dans la chaîne de caractère "concaténation" (correspondant à la concaténation des variables dichotomiques présentée au dessus) et de les compter.

```
1 for (i in 1: nrow(tableau)){
2   chaine <- as.character(tableau$concat[i])
3   caractere <- "1"
4   total$actif[i] <- length( unlist( strsplit(chaine,caractere) ) )
5 }
6 }
```

Figure 8 : Extrait code R

Ces nouvelles variables nous permettent de visualiser à quelle période de l'année le consommateur a participé ou non à la déclaration de ces achats.

Par exemple, nous pouvons voir que l'individu 3 a été présent 25 semaines, mais qu'il a 27 semaines d'inactivités consécutives. Il n'a donc pas participé aux déclarations de ses achats pendant une longue période.

Grâce à ces observations, l'équipe du projet a choisi de conserver les ménages qui sont présents au moins 12 semaines d'activité et qui ont au plus 6 semaines d'inactivités consécutives. Ce qui correspond à une déclaration d'achat par mois au moins, et moins d'un mois et demi d'absence. Le mois et demi d'absence est justifié par le fait qu'on souhaite ensuite étudier la saisonnalité d'achat des produits et qu'on ne veut pas biaiser les résultats en enlevant des gens qui ne seraient pas présents tout l'été par exemple (mois de vente de nombreux fruits/légumes de saison).

Après ces critères appliqués, nous disposons d'un échantillon 7900 ménages.

5. Résultats de l'étude

Les statistiques portant sur les produits et sur les Grandes et Moyennes surfaces sont effectués à partir du panel de 12345 ménages : les ménages consommant des produits de grandes consommation et des produits frais.

Les statistiques portant sur la composition des paniers sont effectués à partir du panel de consommateurs retenu dans la partie précédente : les 7 900 ménages consommant des produits de grandes consommation et des produits frais, déclarant leurs achats pendant au moins 12 semaines et ayant au maximum 6 semaines d'inactivités consécutives.

Pour commencer, nous allons voir quelques statistiques descriptives.

Le but de ces études est dans un premier temps d'analyser la composition des paniers.

Puis nous verrons une analyse des paniers selon le pourcentage de bio des ménages.

Analyse des parts de marché

Moyenne part de marché en valeur (produit bio)	Nombre produit sans bio	Nombre de produit total
1.9%	150	372

Tableau 6 : Statistiques part de marché

Sur les 12345 ménages, la moyenne du pourcentage de bio dépensée dans les GMS est de 1,9 %. Cette moyenne est effectuée pour les 372 produits sachant qu'il n'y a aucun achat en bio pour 150 produits.

Produit	Nombre d'achat	part de marché en valeur (produits bio)	part de marché en quantité (produits bio)
136_SUBSTITUT REPAS	1177	60,14%	76,56%
312_ENDIVES FRAICHES	64778	34,32%	35,54%
299_BETTERAVRG FRAIC	4636	21,35%	20,30%
111_LEGUM LYOPHILISE	1623	19,46%	17,15%
314_MAIS FRAIS	458	16,74%	12,33%
395_OEUF CONDITIONNE	244831	12,15%	6,79%
...
30_LAIT CONSOMMATION	331166	8,15%	5,38%
323_TOMATE FRAICHE	154134	5,51%	2,40%

Tableau 7 : Part de marché en valeur et en quantité bio par produits

Le tableau précédent est un extrait du tableau des parts de marché. Pour chaque produit, nous avons le nombre d'achat effectué, la part de marché en valeur bio (=somme des dépenses bio / somme des dépenses total) et la part de marché en quantité bio (=somme des quantités bio / somme des quantités totale).

Le tableau est classé dans l'ordre décroissant. Le produit ayant la plus grande part de marché en bio est donc le produit « substitut de repas », néanmoins il y a que 1000 achats, ce résultat est donc à nuancer.

A l'inverse, le produit « œuf » a 12.15% de part de marché en valeur pour près de 245 000 achats. Il fait donc partie des produits les plus achetés en bio.

Dépense Bio

Les diagrammes en bâtons qui suivent représentent le pourcentage des achats bio réalisé par enseigne et par groupes de produits. Cette étude nous permet de savoir dans quel type d'enseigne les ménages et quels types de produits sont les plus achetés en bio.

Par type d'enseignes

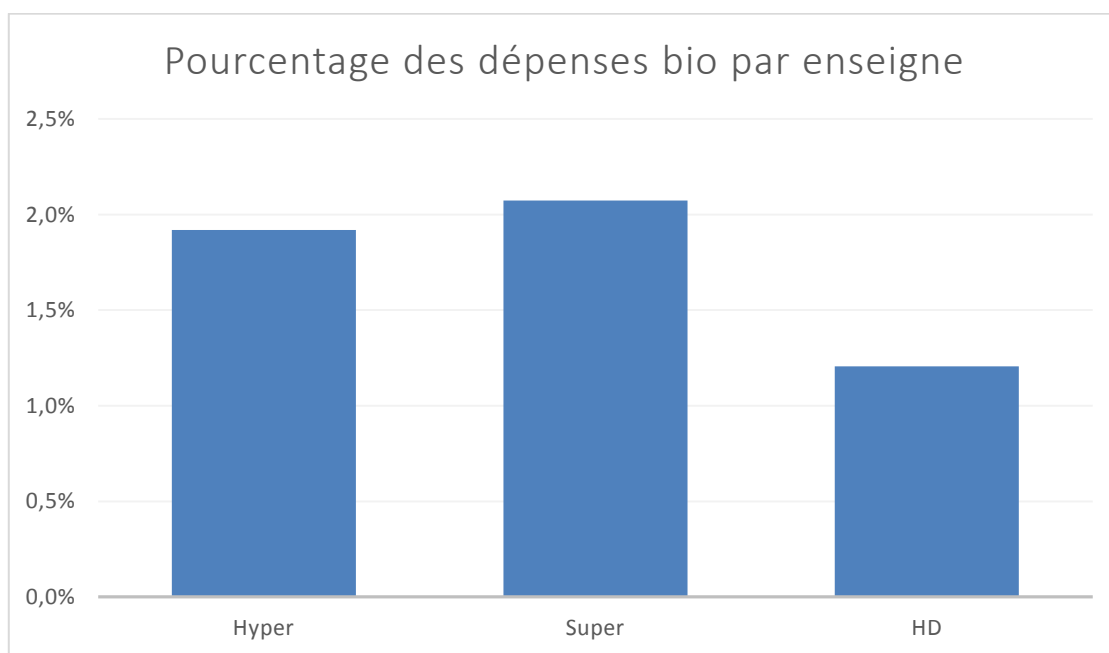


Figure 9 : Pourcentage de dépense bio par type d'enseignes

Nous pouvons voir que la part des achats bio est proche en hyper et supermarché, de 2 % environ, alors qu'elle est de 1,2 % environ en hard discount.

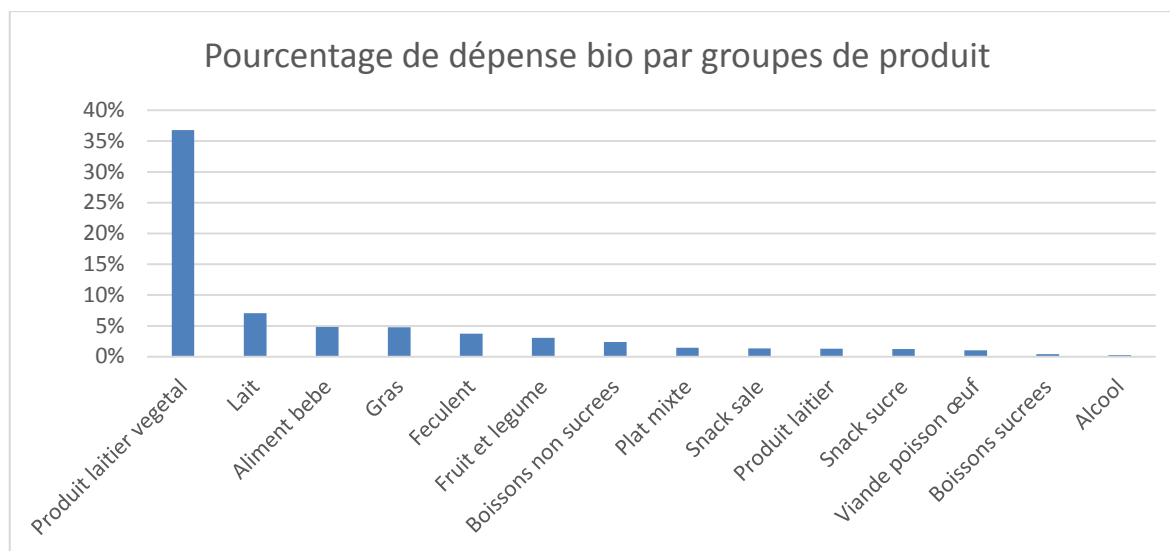
Par groupe

Figure 10 : Pourcentage de dépense bio par groupes de produit

Sur ce diagramme, nous pouvons voir le pourcentage de bio achetée en GMS par groupe de produits. Les principaux produits bio achetés sont les produits laitiers végétaux avec plus de 35 % des achats en bio.

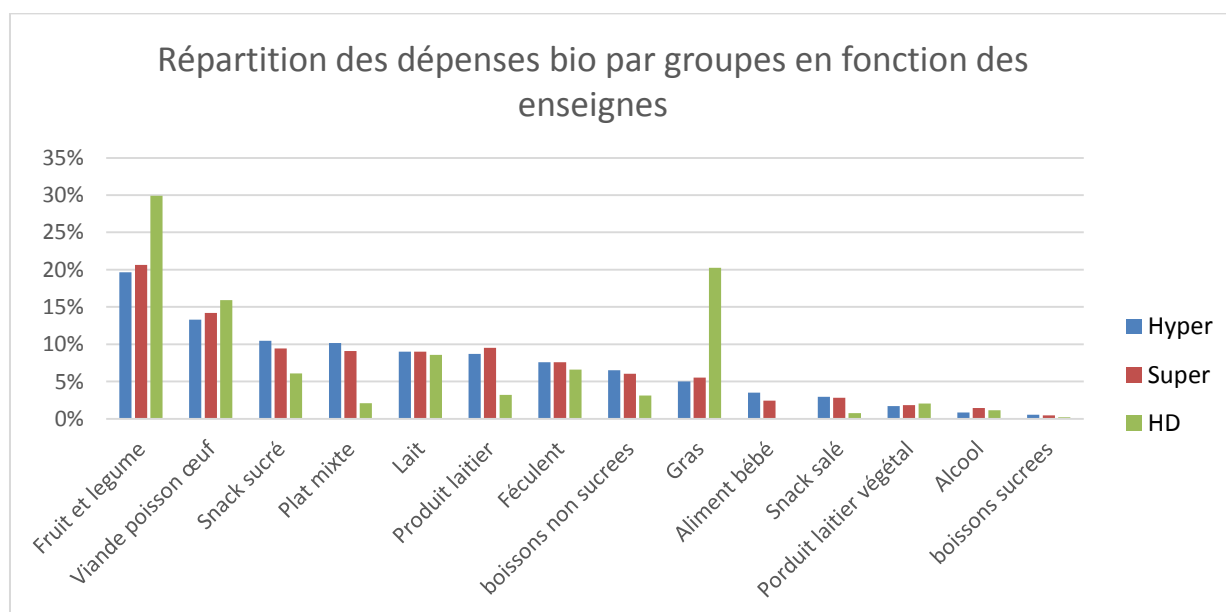
Par groupes en fonction des enseignes

Figure 11 : Répartition des dépenses bio par groupes en fonction des enseignes

Ce diagramme nous permet de voir que la plupart des groupes de produits sont autant achetés en bio dans les 3 types de GMS à l'exception des fruits et légumes et du gras qui sont plus représentés en Hard discount.

Composition des paniers

Répartition des dépenses par groupes

Les deux diagrammes suivants nous permettent de comparer la répartition des achats totaux et des achats bio par groupes de produits. Nous pourrions ainsi voir quel groupe de produits est le plus acheté dans les achats totaux des GMS et dans les achats en bio.

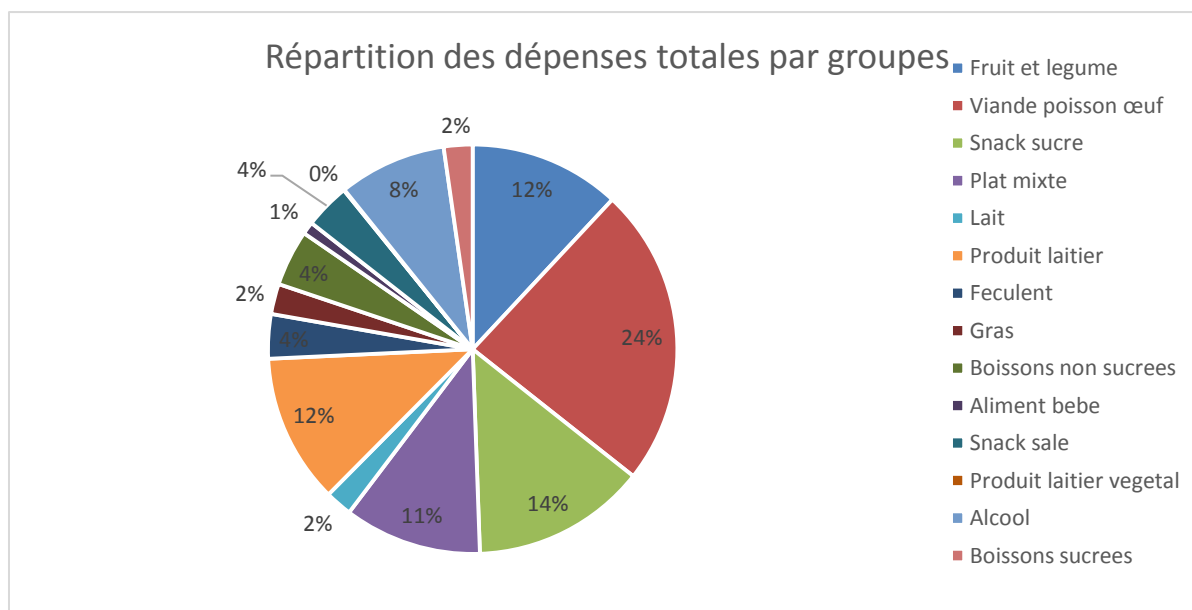


Figure 12 : Répartition des dépenses totales par groupes

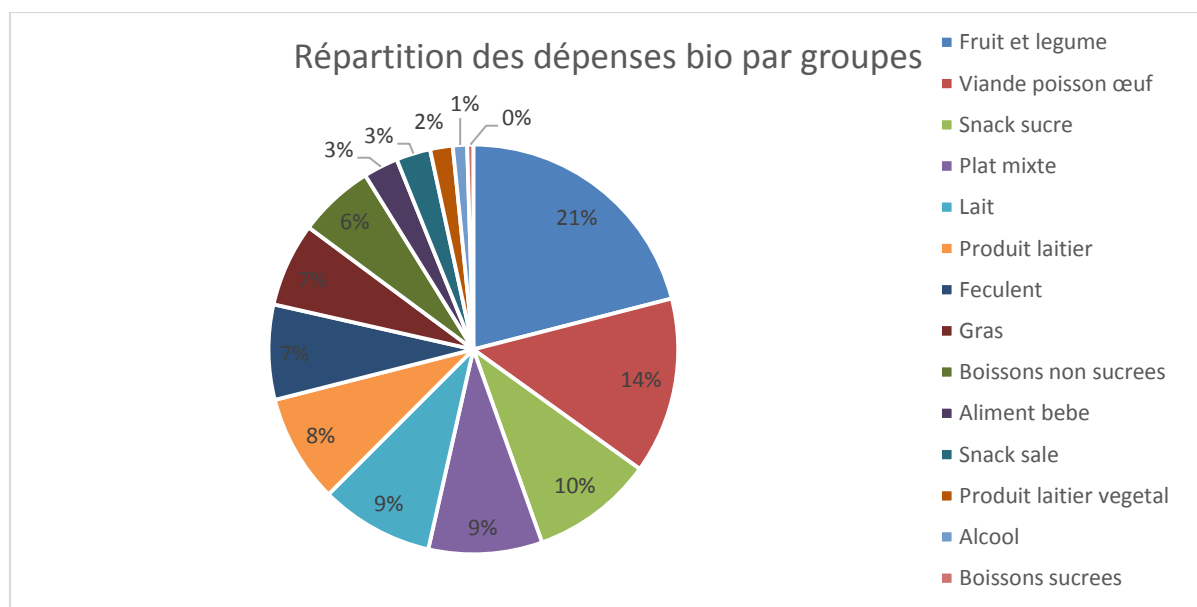


Figure 13 : Répartition des dépenses bio par groupes

Nous constatons que la part des fruits et légumes et celle du lait sont plus importantes dans les achats bios que dans les achats totaux. A l'inverse de la catégorie viande poisson œuf qui représente 24% dans les achats totaux et 14% dans les achats bio.

Pour essayer de voir comment se répartit cette différence ou si une catégorie de produit influence la différence au sein du groupe, nous pouvons faire la même étude par catégorie de produit.

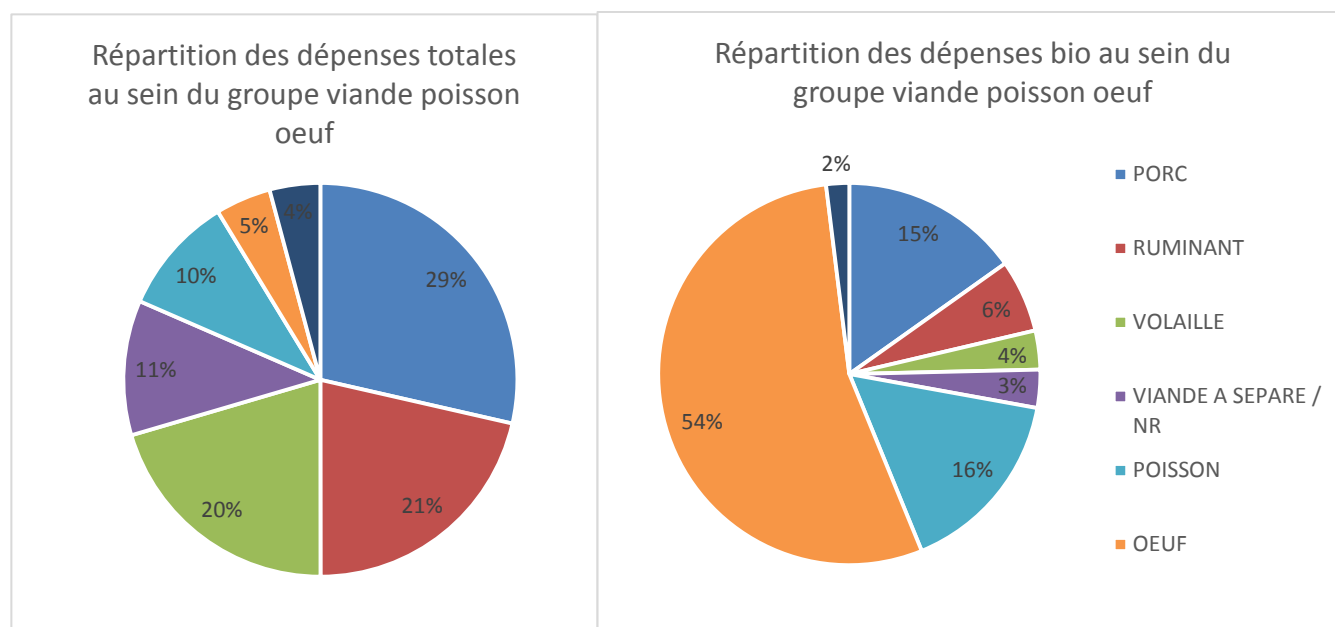


Figure 14 : Répartition des dépenses totale et bio au sein du groupe viande poisson œuf

Au sein du groupe viande poisson œuf, ces diagrammes soulignent une très grande différence dans la répartition des catégories.

En effet, la catégorie de dépense la plus importante en produit non bio est la viande de porc (29%) alors que dans les achats bios elle ne représente que 15% des achats. De même pour la viande de ruminants et de volaille qui représentent environ 20 % en produit non bio et aux alentours de 5% en produit bio.

A l'inverse, la plus grosse catégorie de dépense en produit bio est celle des œufs qui représente plus de la moitié des achats (54%) alors qu'elle en représente que 5% dans les achats non bio.

Répartition des dépenses par catégories

Catégorie	Pourcentage dépense Total	Pourcentage dépense Bio
PLAT MIXTE	10,8%	9,0%
ALCOOL	8,4%	1,1%
AUTRE SNACK SUCRE	7,7%	4,9%
FROMAGE	7,5%	2,2%
PORC	6,7%	2,1%
LEGUME	5,8%	11,9%
RUMINANT	5,0%	0,9%
VOLAILLE	4,8%	0,5%
PATISserie	4,3%	1,6%
FRUIT	4,0%	3,2%
FECULENT	3,6%	7,5%
VIANDE A SEPARER / NR	2,6%	0,4%
YAOURT	2,3%	4,3%
CAFE	2,3%	3,6%
POISSON	2,3%	2,2%
LAIT	2,3%	9,0%
BOISSONS SUCREES	2,2%	0,5%
JUS DE FRUIT	2,0%	5,6%
AUTRE PRODUIT LAITIER	1,9%	2,1%
EAU	1,7%	0,0%
SNACK SALE	1,6%	0,5%
SAUCE	1,3%	1,7%
GRAS ANIMAL	1,2%	2,5%
GRAS VEGETAL	1,2%	4,1%
ŒUF	1,1%	7,5%
ALIMENT BEBE	1,0%	2,7%
CRUSTACE	1,0%	0,3%
COMPOTE	0,9%	1,1%
SUCRE ET MIEL	0,5%	1,3%
THE	0,4%	2,4%
CONDIMENT	0,3%	0,4%
FRUITS SALES	0,3%	0,1%
CONFITURE	0,3%	0,7%
FRUITS SECS	0,3%	0,2%
PRODUIT LAITIER VEGETAL	0,1%	1,8%

Tableau 8 : Répartition des dépenses totale et bio par catégories

Les catégories des légumes, du lait et des œufs sont plus achetées en bio qu'en consommation totale.

Les catégories les plus achetées en total comparé au bio sont le fromage, l'alcool et les viandes porc, volaille et ruminant.

Saisonnalité

Pour étudier les habitudes d'achat des ménages, nous avons analysé la saisonnalité d'achats des fruits et légumes. J'ai donc fait deux diagrammes représentant les dépenses totales et les dépenses bio en fonction des mois des tomates.

Sachant que la saison des tomates s'étend du mois de mai au mois d'août, nous constatons qu'il y a un effet de contre-saisonnalité sur les achats bio. En effet, moins de 15% des dépenses bio sont effectués au cours des mois Décembre, Janvier et Février. Cet effet est moins marqué pour les achats totaux.

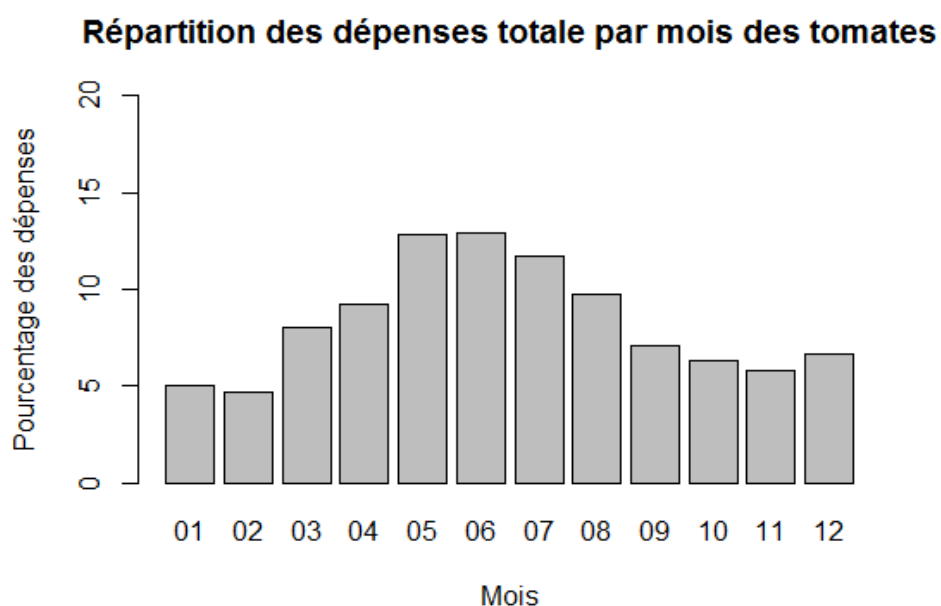


Figure 15 : Répartition par mois des dépenses totales des tomates

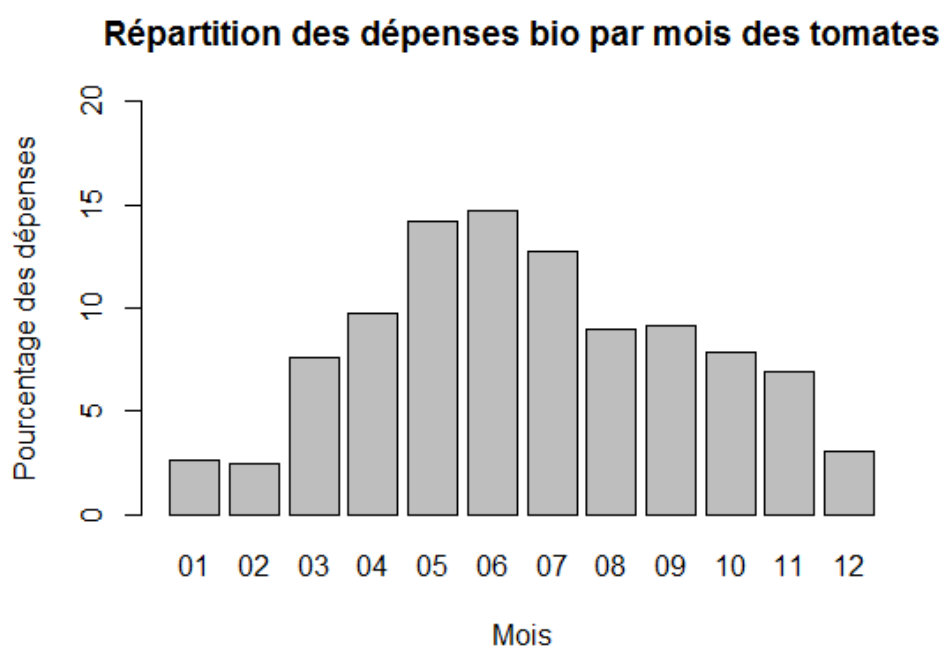


Figure 16 : Répartition par mois des dépenses totales des tomates

Pourcentage de bio

Afin d'étudier la répartition des achats par groupes et catégories en fonction de la consommation en produits issus d'agriculture biologique des ménages, nous avons séparé le fichier en 7 classes (en fonction du pourcentage d'achats bio, voir tableau 9).

La première classe correspond aux ménages n'achetant pas de produit bio et la dernière correspond aux gros acheteurs de produit bio.

Voici quelques statistiques de ces 7 intervalles.

Intervalle (en %)	Effectif	Moyenne (écart-type) Pct bio (en %)	Moyenne des dépenses par ménage		Pourcentage des dépenses par intervalle	
			Totale	Bio	Total	Bio
<= 0.1	992	0,03 (0,03)	2 919	1,2	11,50%	0,30%
] 0.1 ; 1]	4092	0,46 (0,3)	3 383	15,3	54,70%	13,90%
] 1 ; 1.9]	1192	1,37 (0,3)	3 144	43,2	14,80%	11,40%
] 1.9 ; 4]	780	2,7 (0,6)	3 085	82,8	9,50%	14,40%
] 4 ; 9]	492	5,98 (1,4)	2 914	175,5	5,70%	19,20%
] 9 ; 20]	243	13,17 (3,1)	2 708	356,8	2,60%	19,30%
> 20	109	32,42	2 785	889,8	1,20%	21,60%
Total	7900				100%	100%

Tableau 9 : Tableau de statistiques par intervalles de pourcentage de bio

J'ai représenté sur les diagrammes qui suivent (Figure 17 et Figure 18) la répartition des dépenses totale et des dépenses en bio par groupes de produits pour les panelistes consommant entre 0.1 % et 1 % de produits bio et ceux consommant plus de 20 % de produits bio. ²

² Le part des dépenses en produit bio par groupe pour les 7 intervalles se trouve en Annexe 3.

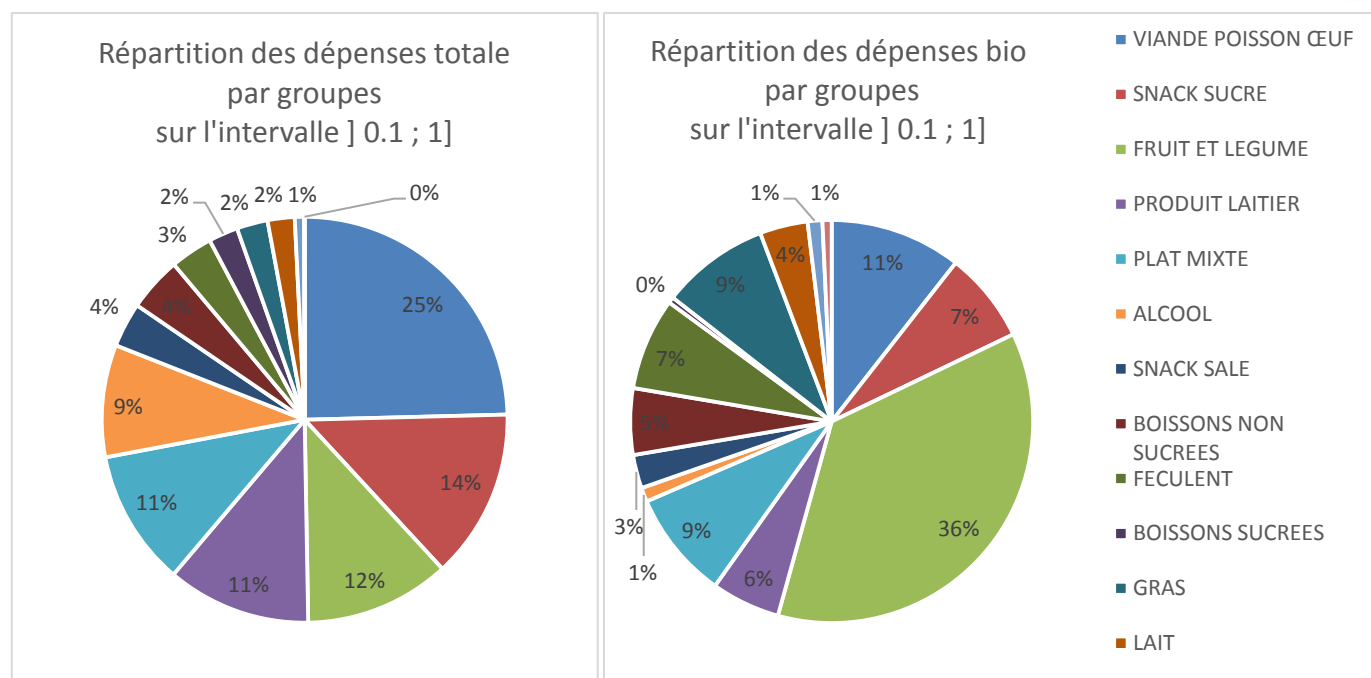


Figure 17 : Répartition des dépenses pour les ménages dépensant « un peu » de produit bio

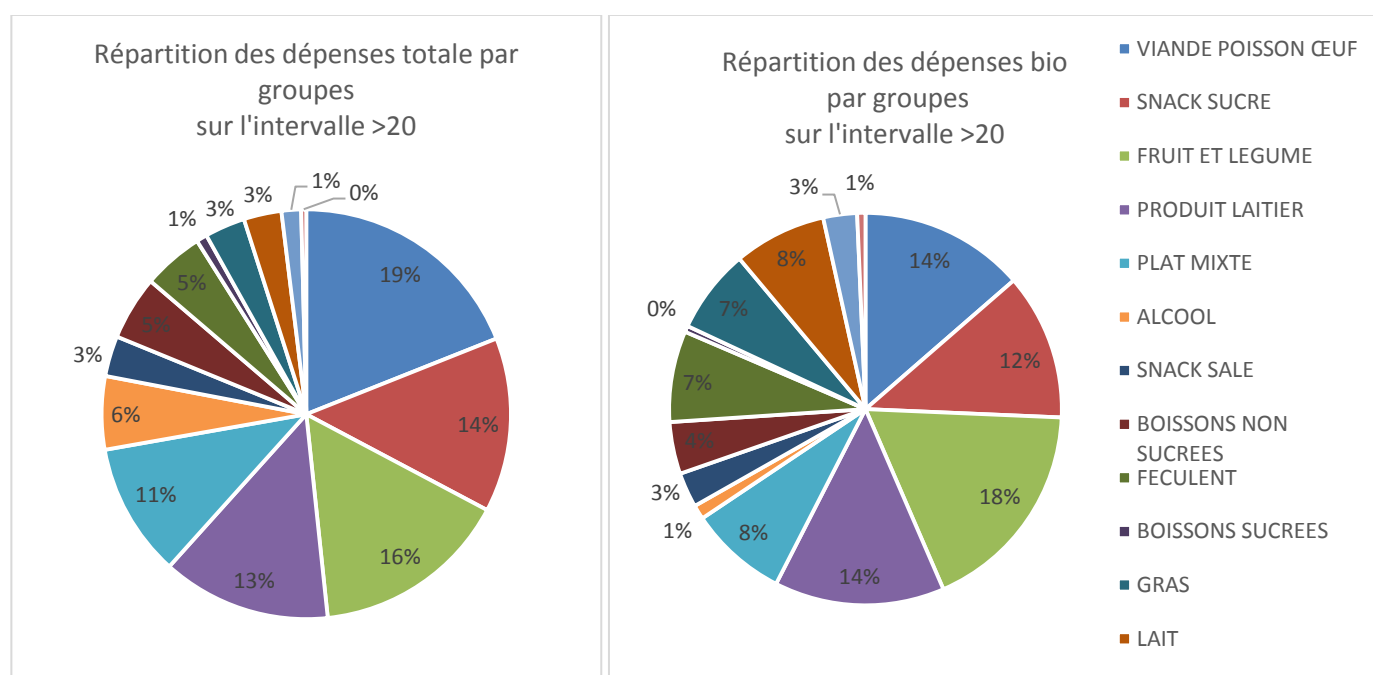


Figure 18 : Répartition des dépenses pour les ménages dépensant « beaucoup » de produit bio

On peut constater que les panelistes ne consommant que très peu de produits bio achètent en majorité des fruits et légumes bio (36 % des dépenses bio contre 12 % des dépenses totales).

Cette différence ne se retrouve pas chez les gros consommateur de bio. En effet, la répartition des dépenses totales et des dépenses bio chez les panéliste consommant plus de 20 % de produits est similaire.

Bilan

Sur le plan personnel ce stage de fin de troisième année de Licence effectué à l'INRA a été pour moi une réussite et m'a apporté de grandes satisfactions, tant sur le plan personnel que professionnel.

J'ai acquis énormément de connaissances sur le logiciel statistique R et en écriture d'algorithme et fonction qui ont été indispensables dans mon travail de traitement et analyse de données, cela m'a permis de mettre en application les compétences que j'ai pu acquérir durant tout mon cursus post-bac.

J'ai également eu l'opportunité de découvrir l'environnement du domaine de la recherche.

Sur le plan professionnel, ce stage a permis de réaliser la première étape du projet de recherche qui était d'extraire et mettre en forme les données sur les achats en grandes et moyennes surfaces de produits issus ou non de l'agriculture biologique.

En effet, à la fin de ce stage je laisse à l'équipe de recherche les données mise en forme et les premières analyses statistiques qui ont été réalisées. De plus, un tableau récapitulatif des comportements d'achats des ménages. Ce tableau permettra de continuer les études statistiques.

Ce stage n'étant que la première étape du projet, mes fichiers et programmes vont être réutilisés par les autres membres du projet. Dans cet optique, j'ai laissé un document expliquant mes fichiers et programmes ainsi qu'un schéma détaillant les liens entre eux (Annexe 2). Ce schéma a été réalisé grâce au logiciel Graphviz.

Conclusion

Ce stage s'inscrivait dans un projet ayant pour but d'analyser dans quelles mesures les compositions des paniers alimentaires des ménages ont des impacts différents en termes de durabilité sur les plans nutritionnels, économiques et environnementaux selon l'importance des achats de produits de l'AB (produits Bio) par les ménages.

Les travaux effectués au cours de ce stage ont permis de laisser des données préparées et quelques résultats statistiques.

En effet, j'ai associé une catégorie et de façon plus large un groupe pour chaque produit pour permettre l'étude des paniers. Les données sont donc prêtes à être traitées.

Un fichier reprenant les variables principales a également été réalisé dans le but de continuer les études statistiques. C'était l'objectif du stage.

Les études statistiques réalisées nous ont permis de voir que le groupe fruits et légumes, le groupe viande poisson œuf et le lait sont les trois groupes les plus achetés en bio. Dans le groupe viande poisson œuf, les œufs se distinguent des autres catégories avec plus 50 % de la part des dépenses en bio dans ce groupe.

Annexes 1 : Explication du découpage des tables

Les tables découpées sont repérables par leur chiffre après l'Under score dans le nom du fichier. S'il est différent de 0, c'est que le fichier a été découpé : le fichier Kantar 11 (Potages) a été découpé en deux :

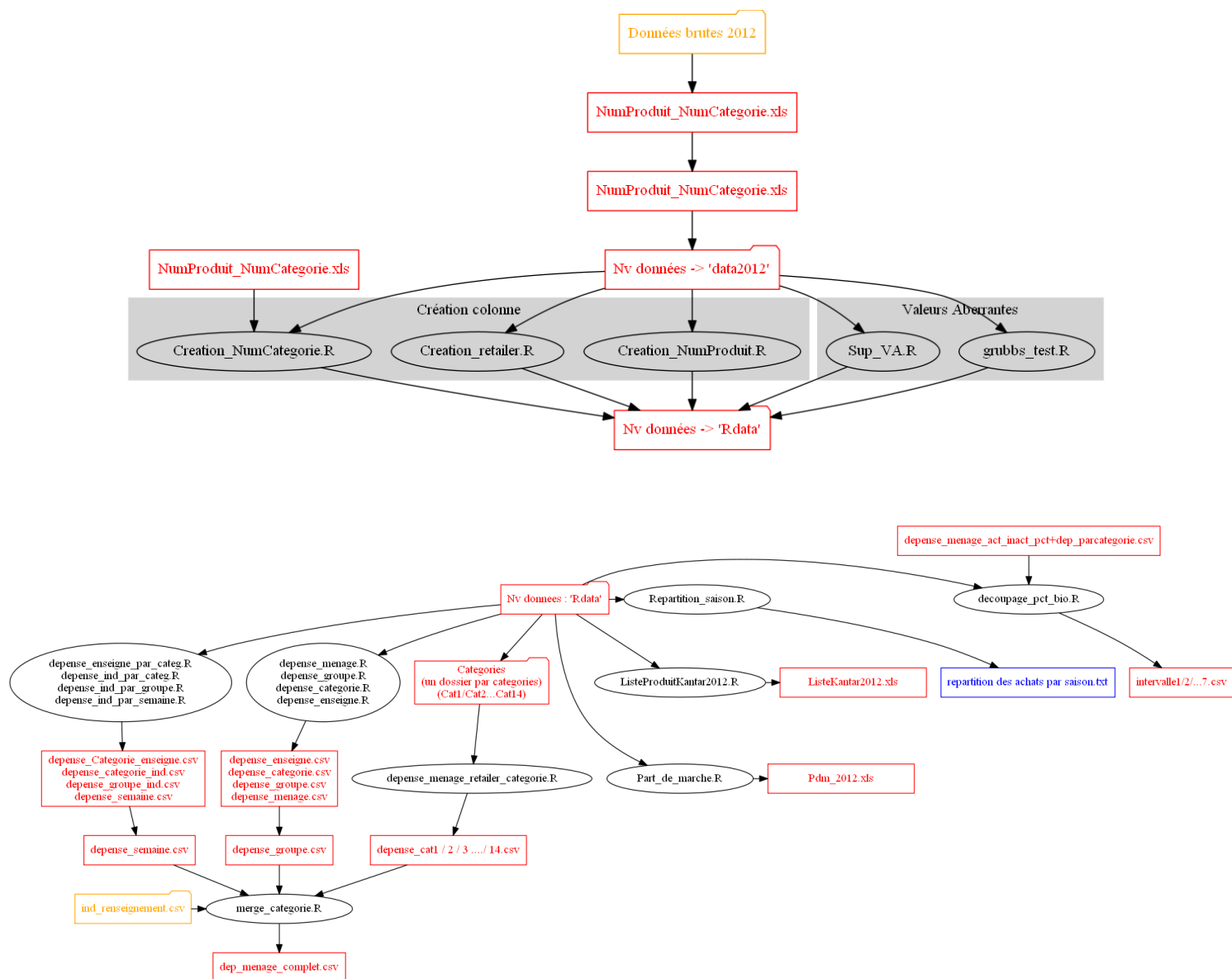
p0011_1NF_Edta, p0011_2NF_Edta

Fichier d'origine	Séparé en ...	Par ...
52	52_1	Spe44 == "13 Abricot ", "208 Figue ", "291 Maïs ", "408 Raisin "
	52_2	Spe32 == "325 Nature "
	52_3	Spe32 == "890 Sucré "
	52_4	Spe32 == "38 Autre ", "959 Salé "
63	63_1	Spe48 == "329 Nectar-Jus de Fruits ", "3109 Concentré Pulpé "
	63_2	le reste
83		Spe224 == "841 Autre Volaille ", "844 Canard ", "850 Poulet ", "1105 Oie ", "2705 Caille ", "2822 Faisan "
	83_1	
	83_2	Spe224 == "1005 Porc "
	83_3	le reste
357		Spe224 == "844 Canard ", "850 Poulet ", "1023 Volaille ", "841 Autre Volaille ", "1105 Oie "
	357_1	
	357_2	Spe224 == "1005 Porc "
	357_3	le reste
416		Spe224 == "35 Assortiment ", "2713 Veau ", "2826 Agneau Mouton ", "842 Boeuf ", "2739 Cheval "
	416_1	
	416_2	Spe224 == "1023 Volaille "
	416_3	Spe224 == "1005 Porc "
417		Spe224 == "35 Assortiment ", "2713 Veau ", "2826 Agneau Mouton ", "842 Boeuf ", "2739 Cheval "
	417_1	
	417_2	Spe224 == "1023 Volaille "
	417_3	Spe224 == "1005 Porc "
430	430_1	Spe224 == "35 Assortiment ", "2713 Veau ", "842 Boeuf ", "2739 Cheval "
	430_2	Spe224 == "1005 Porc "
	430_3	Spe224 == "1023 Volaille ", "848 Lapin "
454	454_1	Spe224 == "35 Assortiment ", "837 Agneau ", "2713 Veau ", "842 Boeuf ", "2739 Cheval "
	454_2	Spe224 == "1005 Porc "
	454_3	Spe224 == "38 Autre ", "1130 Gibier "
91	91_1	Spe39 == "333 Noix de Coco ", "885 Noisette ", "913 Amande "
	91_2	Spe80 == "160 Coulis ", "2415 Fruit Confit ", "2548 Pépité ", "2896 Pâte d'Amande ", "4297 Sucre Aromatisé "
53	53_1	Spe29 == "2473 Végétal "
	53_2	Spe29 == "2477 Animal "

15	15_1	Spe172 == "344 Oui "
	15_2	Spe172 == "334 Non "
362	362_1	Spe44 == "3267 Boudin " & Spe14 == "82 Blanc "
	362_2	le reste
139	139_1	Spe370 == "448 Soja "
	139_2	le reste
140	140_1	Spe77 == "448 Soja "
	140_2	le reste
30	30_1	Spe80 == "448 Soja "
	30_2	le reste
6	6_0	tuwa == "100002 Poids Total en Grs "
18	18_0	tuwa == "100004 Volume Total en ML "
11	11_1	tuwa == "100004 Volume Total en ML "
	11_2	tuwa == "100006 Nombre Total de Parts "
41	41_1	tuwa == "100002 Poids Total en Grs "
	41_2	tuwa == "100004 Volume Total en ML "
81	81_1	tuwa == "100001 Nombre Total "
	81_2	tuwa == "100002 Poids Total en Grs "
110	110_1	tuwa == "100002 Poids Total en Grs " & Spe15 == "334 Non "
	110_2	tuwa == "100002 Poids Total en Grs " & le reste
	110_3	tuwa == "100004 Volume Total en ML " & Spe15 == "334 Non "
	110_4	tuwa == "100004 Volume Total en ML " & le reste
22	22_1	tuwa == "100002 Poids Total en Grs "
	22_2	tuwa == "100004 Volume Total en ML "
33	33_1	tuwa == "100002 Poids Total en Grs "
	33_2	tuwa == "100004 Volume Total en ML "
43	43_1	tuwa == "100002 Poids Total en Grs "
	43_2	tuwa == "100004 Volume Total en ML "
136	136_1	tuwa == "100002 Poids Total en Grs "
	136_2	tuwa == "100004 Volume Total en ML "
155	155_1	Spe60 == " 1004 Blanc de Volaille "
	155_2	le reste

Annexe 2 : Schéma Graphviz

Voici les schémas expliquant les liens entre les codes R (forme arrondie noire) et les fichiers de données (rectangle rouge). Le premier représente les traitements de données et le deuxième les analyses statistiques.



Annexe 3 : Part des dépenses en fonction de la dépense en produits bio

Le tableau ci-dessous représente la part des achats par groupes et en fonction des 7 intervalles de la consommation en produits issus d'agriculture biologique des ménages vu dans le rapport.

Groupe	total_Groupe	pct_bio interv2	pct_bio interv3	pct_bio interv4	pct_bio interv5	pct_bio interv6	pct_bio interv7
VIANDE POISSON ŒUF	23,5%	10,6%	13,5%	16,8%	16,8%	13,4%	13,6%
VIANDE POISSON ŒUF : œuf	1,07%	5,1%	8,4%	11,8%	9,6%	7,4%	4,4%
SNACK SUCRE	13,7%	7,3%	8,4%	9,2%	8,9%	10,1%	12,1%
SNACK SUCRE : autre snack sucré	7,73%	3,5%	4,1%	4,3%	4,9%	5,2%	6,8%
FRUIT ET LEGUME	12,1%	36,4%	24,8%	18,4%	16,0%	16,7%	17,8%
FRUIT ET LEGUME : légume	5,84%	28,9%	17,2%	10,9%	3,3%	6,6%	7,7%
PRODUIT LAITIER	11,8%	5,5%	6,1%	6,8%	8,0%	9,0%	14,0%
PLAT MIXTE	10,8%	8,7%	10,5%	9,7%	8,9%	8,4%	8,1%
ALCOOL	8,4%	1,1%	1,5%	0,9%	1,0%	0,9%	1,2%
SNACK SALE	3,6%	2,7%	2,3%	2,7%	2,2%	3,3%	2,9%
BOISSONS NON SUCREES	4,5%	5,3%	7,1%	9,1%	7,2%	5,1%	4,3%

FECULENT	3,6%	7,4%	7,6%	7,4%	7,9%	7,1%	7,5%
BOISSONS SUCREES	2,2%	0,5%	0,6%	0,7%	0,5%	0,4%	0,5%
GRAS	2,5%	8,7%	8,2%	5,4%	5,6%	6,4%	6,9%
LAIT	2,3%	3,9%	6,3%	8,8%	12,0%	12,2%	7,6%
ALIMENT BEBE	1,0%	1,2%	1,7%	2,1%	2,3%	3,8%	2,8%
PRODUIT LAITIER VEGETAL	0,1%	0,7%	1,6%	2,1%	2,8%	3,1%	0,7%