

Transformer un fichier individu en fichier famille

Camille Bouchard-Coulombe

Département de Démographie, Université de Montréal

Objectifs de recherche

Objectifs de recherche

Couples linguistiquement exogames

Objectifs de recherche

Couples linguistiquement exogames



Langue maternelle des enfants  Langue d'usage

Objectifs de recherche

Couples linguistiquement exogames



Langue maternelle des enfants \longleftrightarrow ^{VS} Langue d'usage

Familles à l'étude

- Résidant au Québec
- Enfants âgés entre 2 et 17 ans

Objectifs de recherche

Couples linguistiquement exogames



Langue maternelle des enfants \longleftrightarrow ^{VS} Langue d'usage

Familles à l'étude

- Résidant au Québec
- Enfants âgés entre 2 et 17 ans

Données

- Échantillon 20% Recensement canadien de 2006
 - Volet détaillé sur la connaissance des langues

Fichier de données

Fichier de données

Fichier brut Recensement 2006 : 1 ligne = 1 individu

Ligne	Rôle	Âge	Langue maternelle	Langue d'usage
1	Père	28	Français	Français
2	Mère	29	Anglais	Français
3	Enfant	4	Français	Français

Fichier de données

Fichier brut Recensement 2006 : 1 ligne = 1 individu

Ligne	Rôle	Âge	Langue maternelle	Langue d'usage
1	Père	28	Français	Français
2	Mère	29	Anglais	Français
3	Enfant	4	Français	Français

Unité d'analyse=FAMILLE

Fichier de données

Fichier brut Recensement 2006 : 1 ligne = 1 individu

Ligne	Rôle	Âge	Langue maternelle	Langue d'usage
1	Père	28	Français	Français
2	Mère	29	Anglais	Français
3	Enfant	4	Français	Français

Unité d'analyse=FAMILLE

Analyses descriptives

Régressions logistiques

Fichier de données

Fichier brut Recensement 2006 : 1 ligne = 1 individu

Ligne	Rôle	Âge	Langue maternelle	Langue d'usage
1	Père	28	Français	Français
2	Mère	29	Anglais	Français
3	Enfant	4	Français	Français

Unité d'analyse=FAMILLE

Analyses descriptives

Régressions logistiques

But: Fichier famille où 1 ligne = 1 famille

Reconfiguration du fichier de données

Reconfiguration du fichier de données

1. Exclusion d'individus

Reconfiguration du fichier de données

1. Exclusion d'individus



2. Transformation du fichier

1. Exclusion d'individus

2. Transformation du fichier

Base de données: 6 470 000 observations

Base de données: 6 470 000 observations



- \neq Québec
- \neq structure familiale
- Famille \neq enfants
- Famille conjoints même sexe
- Famille monoparentale avec enfants
- Enfants <2 ans et >17 ans

Base de données: 6 470 000 observations



- \neq Québec
- \neq structure familiale
- Famille \neq enfants
- Famille conjoints même sexe
- Famille monoparentale avec enfants
- Enfants < 2 ans et > 17 ans



Individus restants:

- Pères, Mères, Enfants ≥ 2 ans et ≤ 17 ans

1. Exclusion d'individus



2. Transformation du fichier

Fichier brut Recensement 2006 : 1 ligne = 1 individu

Fichier brut Recensement 2006 : 1 ligne = 1 individu

ID	E_fampp	C_fampp	ID_PERSONO	Rôle	Âge
184	0	1	3	Père	28
184	0	1	4	Mère	29
184	0	1	5	Enfant	4
186	1	1	1	Père	37
186	1	1	2	Enfant	6
186	1	1	3	Enfant	10
186	1	1	4	Mère	36
187	1	2	1	Mère	50
187	1	2	2	Père	55
187	1	2	3	Enfant	13
187	1	2	4	Enfant	15
187	1	2	5	Enfant	17

Fichier brut Recensement 2006 : 1 ligne = 1 individu

ID	E_fampp	C_fampp	ID_PERSONO	Rôle	Âge
184	0	1	3	Père	28
184	0	1	4	Mère	29
184	0	1	5	Enfant	4
186	1	1	1	Père	37
186	1	1	2	Enfant	6
186	1	1	3	Enfant	10
186	1	1	4	Mère	36
187	1	2	1	Mère	50
187	1	2	2	Père	55
187	1	2	3	Enfant	13
187	1	2	4	Enfant	15
187	1	2	5	Enfant	17

ID :
Ménages

Fichier brut Recensement 2006 : 1 ligne = 1 individu

ID	E_fampp	C_fampp	ID_PERSONO	Rôle	Âge
184	0	1	3	Père	28
184	0	1	4	Mère	29
184	0	1	5	Enfant	4
186	1	1	1	Père	37
186	1	1	2	Enfant	6
186	1	1	3	Enfant	10
186	1	1	4	Mère	36
187	1	2	1	Mère	50
187	1	2	2	Père	55
187	1	2	3	Enfant	13
187	1	2	4	Enfant	15
187	1	2	5	Enfant	17

ID :
Ménages

E_Fampp :
Famille économique

Fichier brut Recensement 2006 : 1 ligne = 1 individu

ID	E_fampp	C_fampp	ID_PERSONO	Rôle	Âge
184	0	1	3	Père	28
184	0	1	4	Mère	29
184	0	1	5	Enfant	4
186	1	1	1	Père	37
186	1	1	2	Enfant	6
186	1	1	3	Enfant	10
186	1	1	4	Mère	36
187	1	2	1	Mère	50
187	1	2	2	Père	55
187	1	2	3	Enfant	13
187	1	2	4	Enfant	15
187	1	2	5	Enfant	17

ID :
Ménages

E_Fampp :
Famille économique

C_Fampp :
Famille de recensement

Fichier brut Recensement 2006 : 1 ligne = 1 individu

ID	E_fampp	C_fampp	ID_PERSONO	Rôle	Âge
184	0	1	3	Père	28
184	0	1	4	Mère	29
184	0	1	5	Enfant	4
186	1	1	1	Père	37
186	1	1	2	Enfant	6
186	1	1	3	Enfant	10
186	1	1	4	Mère	36
187	1	2	1	Mère	50
187	1	2	2	Père	55
187	1	2	3	Enfant	13
187	1	2	4	Enfant	15
187	1	2	5	Enfant	17

ID :
Ménages

E_Fampp :
Famille économique

C_Fampp :
Famille de recensement

ID_PERSONO:
Individus du ménage

1. Numéro d'identification → ID_fam

1. Numéro d'identification → ID_fam

ID (ménage)	E_fampp	C_fampp	ID_fam
184	0	1	18401
184	0	1	18401
184	0	1	18401
186	1	1	18611
186	1	1	18611
186	1	1	18611
186	1	1	18611
187	1	2	18712
187	1	2	18712
187	1	2	18712
187	1	2	18712
187	1	2	18712

Obtenu en concaténant
ID et
E_fampp et
C_fampp

1. Numéro d'identification → ID_fam

ID (ménage)	E_fampp	C_fampp	ID_fam
184	0	1	18401
184	0	1	18401
184	0	1	18401
186	1	1	18611
186	1	1	18611
186	1	1	18611
186	1	1	18611
187	1	2	18712
187	1	2	18712
187	1	2	18712
187	1	2	18712
187	1	2	18712

Obtenu en concaténant
ID et
E_fampp et
C_fampp

Clé d'appariement

2. Numéro d'identification → ID1

2. Numéro d'identification → ID1

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	Âge
184	0	1	3	18401	1	Père	28
184	0	1	4	18401	2	Mère	29
184	0	1	5	18401	3	Enfant	4
186	1	1	1	18611	1	Père	37
186	1	1	2	18611	2	Enfant	6
186	1	1	3	18611	3	Enfant	10
186	1	1	4	18611	4	Mère	36
187	1	2	1	18712	1	Mère	50
187	1	2	2	18712	2	Père	55
187	1	2	3	18712	3	Enfant	13
187	1	2	4	18712	4	Enfant	15
187	1	2	5	18712	5	Enfant	17

3. Ordonner les individus d'une famille

3. Ordonner les individus d'une famille

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
184	0	1	3	18401	1	Père → 1	1	28
184	0	1	4	18401	2	Mère → 2	2	29
184	0	1	5	18401	3	Enfant → 3	3	4
186	1	1	1	18611	1	Père	1	37
186	1	1	2	18611	2	Enfant	3	6
186	1	1	3	18611	3	Enfant	3	10
186	1	1	4	18611	4	Mère	2	36
187	1	2	1	18712	1	Mère	2	50
187	1	2	2	18712	2	Père	1	55
187	1	2	3	18712	3	Enfant	3	13
187	1	2	4	18712	4	Enfant	3	15
187	1	2	5	18712	5	Enfant	3	17

1. En fonction du rôle



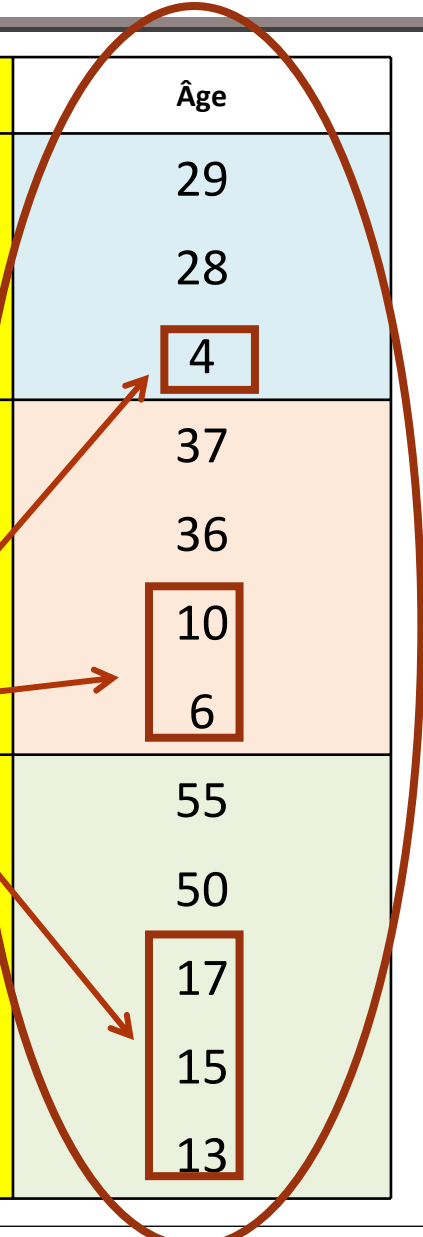
3. Ordonner les individus d'une famille

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
184	0	1	4	18401	2	Mère	2	29
184	0	1	5	18401	1	Père	1	28
184	0	1	6	18401	3	Enfant	3	4
186	1	1	4	18611	4	Père	1	37
186	1	1	4	18611	2	Mère	2	36
186	1	1	3	18611	3	Enfant	3	10
186	1	1	3	18611	3	Enfant	3	6
187	1	2	1	18712	1	Mère	2	55
187	1	2	3	18712	3	Enfant	3	17
187	1	2	4	18712	4	Enfant	3	15
187	1	2	5	18712	5	Enfant	3	13

1. En fonction du Rôle

2. Ordre décroissant de l'âge

Âge des enfants en ordre décroissant

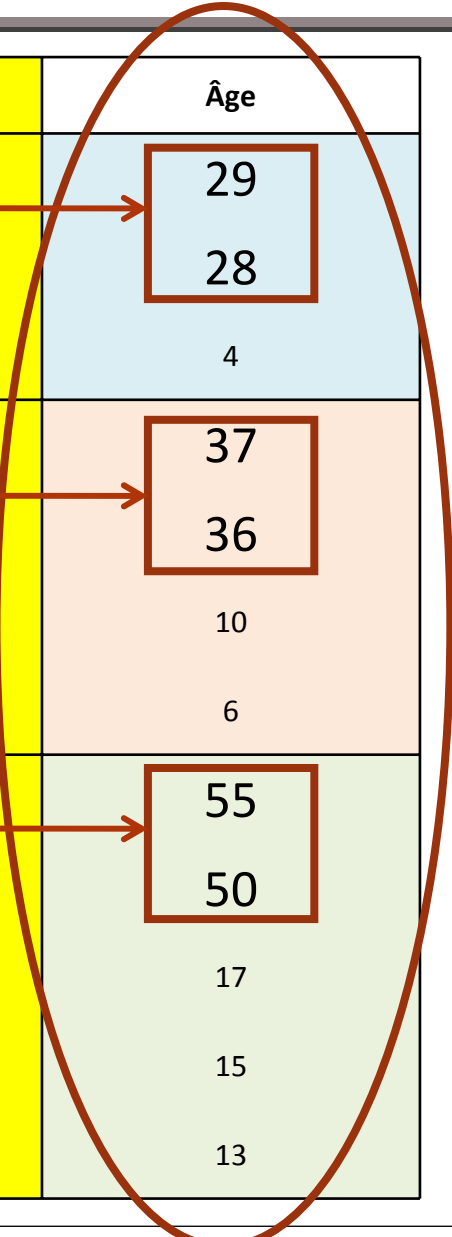


3. Ordonner les individus d'une famille

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
184	0	1	4	18401	2	Mère	2	29
184	0	1	5	18401	1	Père	1	28
184	0	1	5	18401	3	Enfant	3	4
186	1	1	4	18611	4	Père	1	37
186	1	1	3	18611	2	Mère	2	36
186	1	1	3	18611	3	Enfant	3	10
186	1	1	2	18611	2	Enfant	3	6
187	1	2	2	18712	2	Père	1	55
187	1	2	1	18712	1	Mère	2	50
187	1	2	3	18712	3	Enfant	3	17
187	1	2	4	18712	4	Enfant	3	15
187	1	2	5	18712	5	Enfant	3	13

1. En fonction du Rôle

2. Ordre décroissant de l'âge



3. Ordonner les individus d'une famille

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
184	0	1	3	18401	1	Père	1	28
184	0	1	3	18401	2	Mère	2	29
184	0	1	3	18401	3	Enfant	3	4
186	1	1	4	18611	4	Père	1	37
186	1	1	4	18611	4	Mère	2	36
186	1	1	2	18611	2	Enfant	3	10
186	1	1	2	18611	2	Enfant	3	6
187	1	2	2	18712	2	Père	1	55
187	1	2	1	18712	1	Mère	2	50
187	1	2	3	18712	3	Enfant	3	17
187	1	2	4	18712	4	Enfant	3	15
187	1	2	5	18712	5	Enfant	3	13

1. En fonction du Rôle

2. Ordre décroissant de l'âge

3. En fonction de ID2

4. Recréer la variable ID1

4. Recréer la variable ID1

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
184	0	1	3	18401	1	Père	1	28
184	0	1	4	18401	2	Mère	2	29
184	0	1	5	18401	3	Enfant	3	4
186	1	1	1	18611	1	Père	1	37
186	1	1	4	18611	4	Mère	2	36
186	1	1	3	18611	3	Enfant	3	10
186	1	1	2	18611	2	Enfant	3	6
187	1	2	2	18712	2	Père	1	55
187	1	2	1	18712	1	Mère	2	50
187	1	2	3	18712	3	Enfant	3	17
187	1	2	4	18712	4	Enfant	3	15
187	1	2	5	18712	5	Enfant	3	13

4. Recréer la variable ID1



ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
184	0	1	3	18401	1	Père	1	28
184	0	1	4	18401	2	Mère	2	29
184	0	1	5	18401	3	Enfant	3	4
186	1	1	1	18611	1	Père	1	37
186	1	1	4	18611	2	Mère	2	36
186	1	1	3	18611	3	Enfant	3	10
186	1	1	2	18611	4	Enfant	3	6
187	1	2	2	18712	1	Père	1	55
187	1	2	1	18712	2	Mère	2	50
187	1	2	3	18712	3	Enfant	3	17
187	1	2	4	18712	4	Enfant	3	15
187	1	2	5	18712	5	Enfant	3	13

5. Création de fichiers distincts en fonction ID1

5. Création de fichiers distincts en fonction ID1

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Père					1	Père		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Mère					2	Mère		

5. Création de fichiers distincts en fonction ID1

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Père					1	Père		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Mère					2	Mère		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Enfant rang 1					3	E1		

5. Création de fichiers distincts en fonction ID1

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Père					1	Père		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Mère					2	Mère		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Enfant rang 1					3	E1		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Enfant rang 2					4	E2		

5. Création de fichiers distincts en fonction ID1

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Père					1	Père		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Mère					2	Mère		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Enfant rang 1					3	E1		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Enfant rang 2					4	E2		

ID (ménage)	E_fampp	C_fampp	ID_PERSONO	ID_fam	ID1	Rôle	ID2	Âge
Fichier Enfant rang 3					5	E3		

6. Conserver les variables utiles

7. Renommer les variables restantes

7. Renommer les variables restantes

	ID_fam	ID1_p	Rôle_p	ID2_p	Âge_p
Fichier Père (_p)					

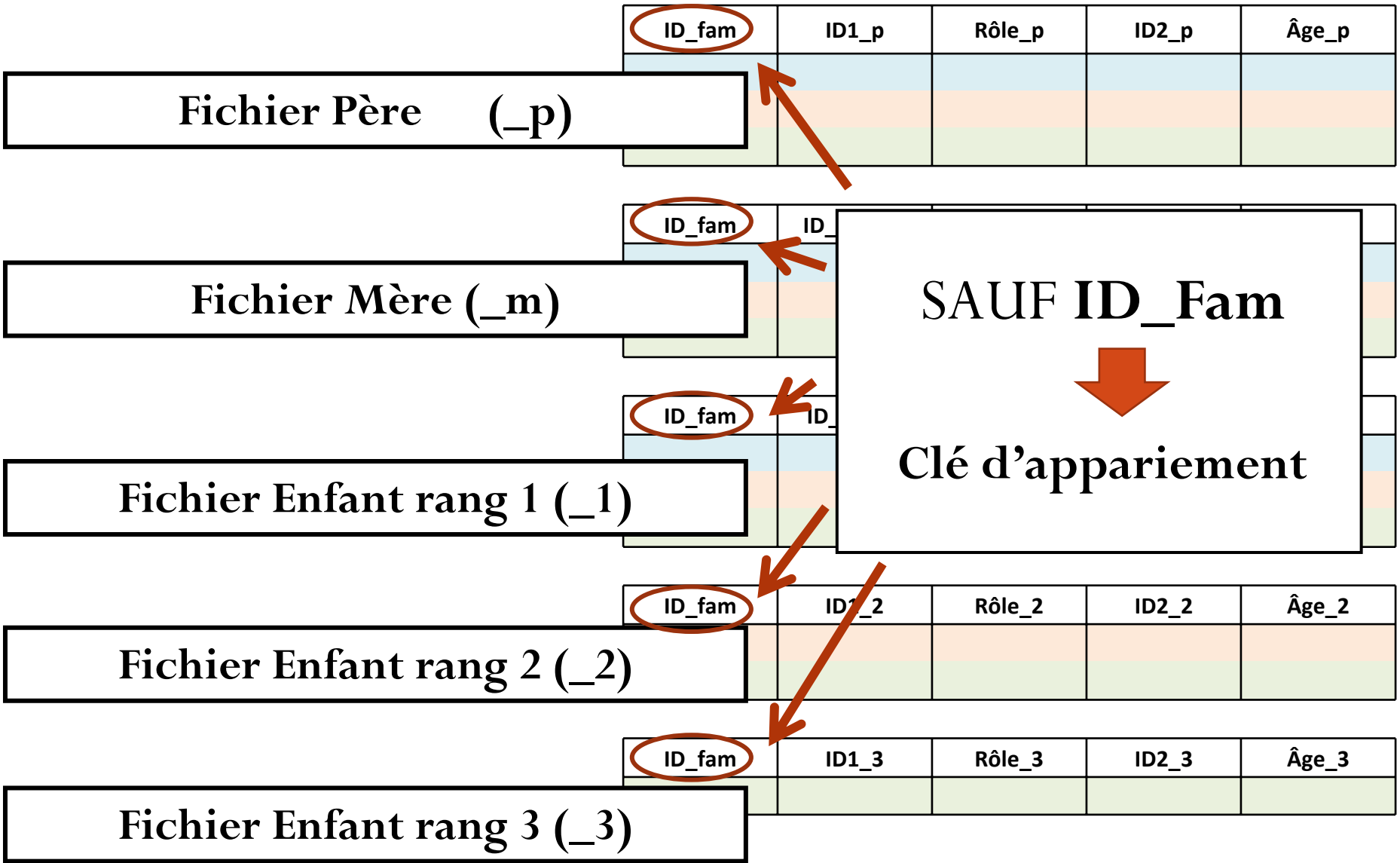
	ID_fam	ID1_m	Rôle_m	ID2_m	Âge_m
Fichier Mère (_m)					

	ID_fam	ID1_1	Rôle_1	ID2_1	Âge_1
Fichier Enfant rang 1 (_1)					

	ID_fam	ID1_2	Rôle_2	ID2_2	Âge_2
Fichier Enfant rang 2 (_2)					

	ID_fam	ID1_3	Rôle_3	ID2_3	Âge_3
Fichier Enfant rang 3 (_3)					

7. Renommer les variables restantes



8. Fusionner les fichiers en fonction de la clé d'appariement

8. Fusionner les fichiers en fonction de la clé d'appariement

ID_fam	ID1_p	Rôle_p	ID2_p	Âge_p	ID1_m	Rôle_m	ID2_m	Âge_m
18401	1	Père	1	28	2	Mère	2	29
18611	1	Père	1	37	2	Mère	2	36
18712	1	Père	1	55	2	Mère	2	50

ID1_1	Rôle_1	ID2_1	Âge_1	ID1_2	Rôle_2	ID2_2	Âge_2
3	Enfant	3	4
3	Enfant	3	10	4	Enfant	3	6
3	Enfant	3	17	4	Enfant	3	15

ID1_3	Rôle_3	ID2_3	Âge_3
.	.	.	.
.	.	.	.
5	Enfant	3	13

Création valeurs
manquantes chez les familles
qui n'ont pas d'enfants de
rang 2 et +

9. Supprimer certaines familles


9. Supprimer certaines familles

- Familles dont TOUS les enfants:
 - -2 ans
 - + 17 ans

9. Supprimer certaines familles

- Familles dont TOUS les enfants:
 - -2 ans
 - + 17 ans
- Valeurs manquantes

9. Supprimer certaines familles

- Familles dont TOUS les enfants:
 - -2 ans
 - + 17 ans Valeurs manquantes
- Familles dont les grands-parents = parents

Conclusion

Conclusion

Fichier famille complet → Différentes applications possibles

Conclusion

Fichier famille complet → Différentes applications possibles

Sélection familles linguistiquement exogames

Conclusion

Fichier famille complet → Différentes applications possibles

Sélection familles linguistiquement exogames

 Unité d'analyse = FAMILLE 

Conclusion

Fichier famille complet → Différentes applications possibles

Sélection familles linguistiquement exogames

 Unité d'analyse = FAMILLE 

- Limites:
- Difficile d'analyser individuellement chaque enfant
 - Impossible d'identifier les familles recomposées

Conclusion

Fichier famille complet → Différentes applications possibles

Sélection familles linguistiquement exogames

 Unité d'analyse = FAMILLE 

- Limites:
- Difficile d'analyser individuellement chaque enfant
 - Impossible d'identifier les familles recomposées

Avantage: Conserve information pour chaque enfant

Remerciements

- Centre Interuniversitaire québécois de statistiques sociales (CIQSS)
- Office québécois de la langue française

Pour information : Camille Bouchard-Coulombe
camille.bouchard-coulombe@umontreal.ca



Office québécois
de la langue
française

