

MAFE Methodological Note 9

Manipulation des bases biographiques

Application à la transnationalité des unions

SAS

Marc Thevenin - Daniele Rossi

Janvier 2015



*Funded under the
Socio-economic
Sciences & Humanities
Theme*



TABLE DE MATIERES

INTRODUCTION	3
PARTIE I : Création de la base d'analyse en format individus/années	7
1) Base listant toutes les années de vie du répondant jusqu' à l'année de l'enquête	7
2) Base individus-années listant les pays de résidence du répondant.....	9
3) Base individus-années listant les pays de résidence des conjoints	14
a) Pays de résidence des partenaires	14
b) Les périodes de séparation du répondant	15
c) Les périodes d'union du répondant	16
c) Appariement des bases répondants et conjoints.....	21
PARTIE II. Exemples de création de variables d'analyse	24
1) Création des variables de statut matrimonial.....	24
Les années de mariages.....	24
b) Variables dichotomiques de statut matrimonial	25
d) Variable synthétique de statut matrimonial	27
2) Variable de comparaison des pays de résidence des répondants et des partenaires.....	28
a) Comparaison des codes pays entre le répondant et le(les) partenaires.....	28
b) Création des variables synthétiques de localisation	29
ANNEXES	41
Création de la base d'analyse	42
1) Création de la base individus années <i>year.dta</i> à partir d'un module biographique	42
a) Modules dont les trajectoires ne sont pas strictement continues	42
b) A partir d'un module strictement continu, <i>house.dta</i> ou <i>activity.dta</i>	42
2) Pays de résidence du répondant à partir de la base « house »	43

INTRODUCTION

L'objectif de ce tutoriel est de donner des outils de manipulation des modules biographiques de l'enquête MAFE avec SAS dans le cadre d'une œuvre une analyse biographique à temps discret. En termes de prérequis avec le logiciel, seules les opérations comme la création de variables et la fusion de bases sont nécessaires, les autres instructions étant brièvement présentées dans le tutoriel avec si nécessaire des liens vers le manuel de référence. Du fait de la complexité des données, il convient **impérativement de consulter** la documentation méthodologique présente sur le site mafeproject.com, en particulier le document décrivant toutes les bases mises à disposition (« [Introduction to the MAFE datasets](#) »).

Objectifs du tutoriel

- I. **Changement de niveau d'observation** : les modules biographiques de MAFE sont livrés en format individu-événement (ex : dans le fichier des unions, pour chaque individu, une ligne par union). Si ce type de format permet déjà de réaliser des analyses, la nécessité de combiner plusieurs modules ou l'utilisation de méthodes en temps discret demande de modifier le niveau d'observation. On va donc créer une base d'analyse qui donne *pour chaque année* des informations sur le pays de résidence du répondant et de son ou ses conjoints, ainsi que sa situation maritale. Le niveau d'observation passe donc de l'événement vécu à l'année vécue.
- II. **Création de « variables du moment »**: On cherchera à créer des variables synthétiques qui permettent de renseigner pour chaque instant d'observation les caractéristiques d'état d'un individu. Ces variables sont souvent appelée TVC pour « Time Varying Covariate » dans la littérature technique relative à ce type de problématique. Par exemple, on cherchera à renseigner une variable donnant pour chaque année le statut matrimonial annuel du répondant et une information sur la transnationalité des unions qui tient compte des situations passées.

D'une base individus-événements à une base individus-périodes

En entrée : on dispose d'un fichier *individus-événements*. A l'exception de la base *general.dta*, tous les modules biographiques de MAFE sont livrés sous ce format avec une variable indiquant le rang lié à l'événement (**num_event**), l'année d'entrée dans l'événement (qxxxd) et l'année de sortie (qxxxf), ainsi que diverses caractéristiques propres à l'événement.

Dans la base *union.dta* par exemple, le rang de l'union est donné par la variable **num_uni**, l'année de début de l'union par **q102d**, l'année de fin par **q102f**, le motif de rupture de l'union par **q102** et l'année de mariage par **q103**.

Si l'année de fin du dernier épisode est une valeur manquante, cela indique que l'événement était toujours en cours au moment de l'enquête.

Base individu-événements

ident	num_event	qxxxd	qxxxf	VarX
SN000000	1	1970	1974	a
SN000000	2	2005	.	b

En sortie : on souhaite disposer d'un fichier *individus-périodes* (ou plus précisément, *individus-années*). Après avoir calculé la durée de l'évènement, on va transformer la base pour à avoir une ligne par année.

Base individu-années

ident	year	num_event	Qxxxd	qxxxf	VarX
SN000000	1970	1	1970	1972	a
SN000000	1971	1	1970	1972	a
SN000000	1972	1	1970	1972	a
SN000000	2005	2	2005	2008	b
SN000000	2006	2	2005	2008	b
SN000000	2007	2	2005	2008	b
SN000000	2008	2	2005	2008	b

Ceci implique :

1. Une transformation des bases individus-événements en bases individus-périodes en tenant compte des trajectoires continues et des fin d'évènement l'année de l'enquête.
2. Une transformation, appelée souvent reshaping dans la littérature technique, des bases lorsque plusieurs évènements de même type ont lieu en même temps (cas de « temporal overlapping »). Ici, le « temporal overlapping » est engendré par la présence d'unions polygames. Dans ce tutoriel, la polygamie est seulement traitée du point de vue d'un mari (le répondant) avec plusieurs épouses. Lorsque le répondant est une femme, bien qu'on dispose dans MAFE d'informations sur les épisodes de migration des autres co-épouses, on ne connaît pas leurs années d'entrée et de sortie de l'union.

Ressources mises à disposition

- **sas_tuto.sas**: la totalité du programme décrit dans ce document (sauf les compléments qui figurent en annexe).
- Extraits des bases de données de MAFE: **general**, **migration**, **house**, **union** et **network**, pour 10 répondants et une sélection de variables. Les bases ainsi que les formats sont disponibles dans le fichier mafe.tutorial.files :

http://mafeproject.site.ined.fr/fichier/s_rubrique/20386/mafe.tutorial.files.janvier.2015.fr.zip

Documentation sur les bases

Avant toute manipulation des bases, il est fortement conseillé de consulter la documentation de l'enquête, surtout celle dédiée aux bases de données du document « **presentation of the datasets** », et plus particulièrement pour ce tutoriel, la partie consacrée aux modules biographiques (à partir de la page :

- Base **general** : page 32-33.
- Base **house** : page 34.
- Base **migration** : page 36.
- Base **union** : pages 34.
- Base **network** pages 42-46.

Différences avec les tutoriels STATA et R¹

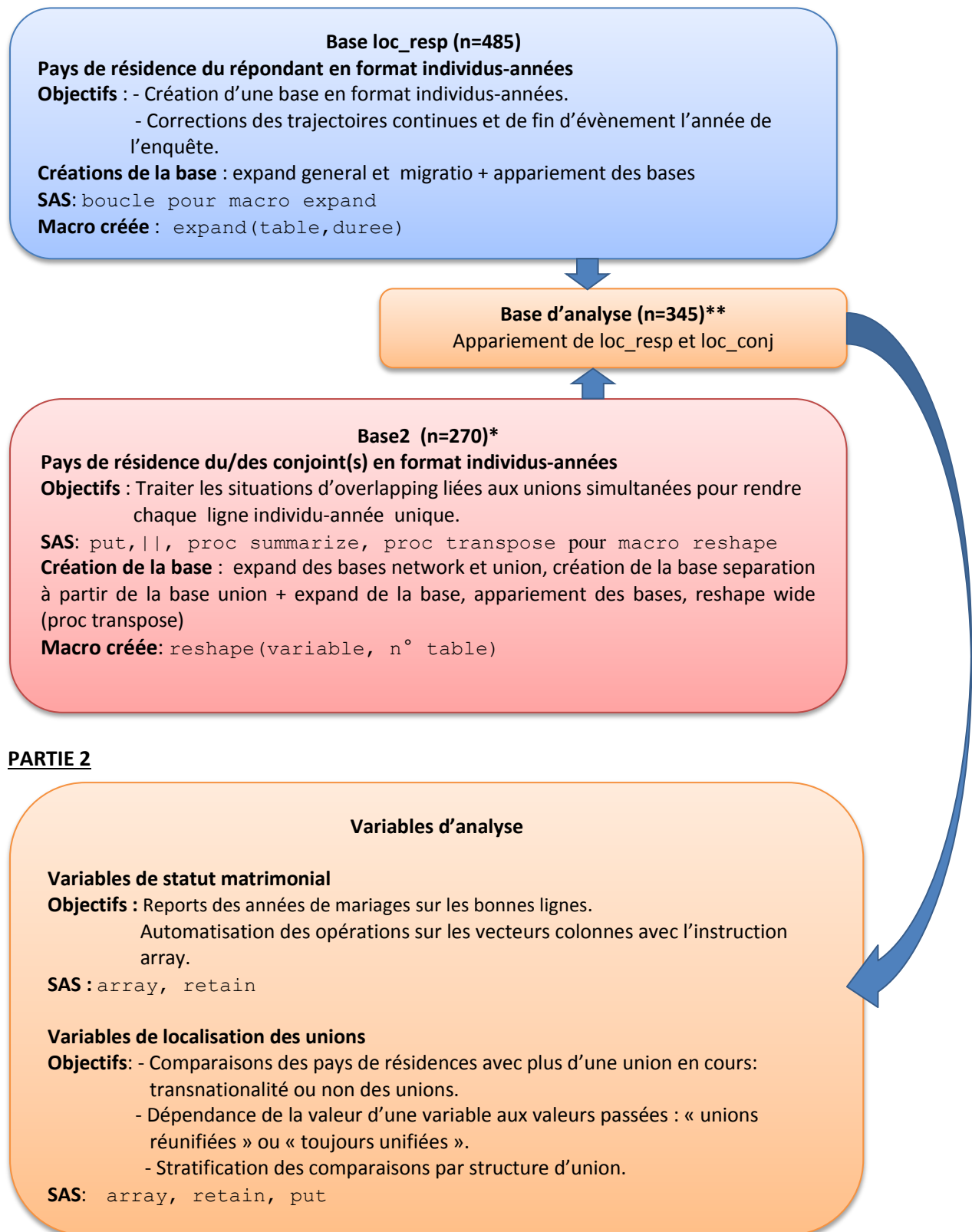
La logique des programmes **STATA** et **SAS** est quasi identique, la seule différence réside dans le report des informations pour la variable q103 (motifs de séparation) qui est ici résolu dans la première partie.

Celle avec **R** est différente. La majorité des variables créées dans la seconde partie sont générée sous R dès la constitution de la base d'analyse avant la correction de l'overlapping. Contrairement au tutoriel avec Stata, ce programme n'a pas été traduit en SAS.

¹ Diffusion prévue pour le début de l'année 2015

Les principales étapes et procédures SAS introduites

PARTIE 1



PARTIE 2

* L'effectif de la base est supérieur à celui obtenu avec Stata (n=255) en raison de la création de la base separation. Cela n'a aucune conséquence sur la taille de la base finale d'analyse.

** Pour un âge supérieur ou égal à 14 ans.

PARTIE I : Création de la base d'analyse en format individus/années

1) Base listant toutes les années de vie du répondant jusqu' à l'année de l'enquête

Objectif :

Créer une ligne d'observation par année de vie

1. Création d'une variable durée (entre l'année de naissance et 2008, année de l'enquête).
2. « Allongement » de la base pour chaque individu : les observations vont être dupliquées avec l'instruction `expand`.

On utilisera ici la base *general*, dont la structure est classique, avec autant de lignes que de répondants. D'autres options sont possibles en utilisant directement les modules biographiques².

Base « general »

variable name	storage type	display format	value label	variable label
ident	str7	%9s		Biographic individual identifier
q1a	int	%17.0g	Q1A	Year of birth

Dans un premier temps, on crée une variable de durée entre 2008 (année de l'enquête) et l'année de naissance (variable **q1a**). On doit ajouter 1 à la différence entre les deux années pour obtenir le nombre exact d'années de la naissance à 2008.

On verra plus tard que cet ajout génère des doublons lorsque la transition entre deux évènements a lieu la même année (trajectoire continue). Ces doublons devront faire l'objet d'une correction.

```
data year; set mafe.general;
a2008=2008;
dur=a2008-q1a+1;
run;
```

```
+-----+
|  ident  num_un~n  q1a  a2008  dur |
+-----+
1. |    E1          1  1972    2008   37 |
2. |   E103         1  1978    2008   31 |
3. |   E147         1  1954    2008   55 |
4. |   F112         1  1951    2008   58 |
5. |   F121         1  1972    2008   37 |
+-----+
6. |   F127         1  1964    2008   45 |
7. |   I172         1  1971    2008   38 |
8. | S011202        1  1975    2008   34 |
9. | S079601        1  1936    2008   73 |
+-----+
```

² Se reporter à l'annexe du tutoriel Stata. On peut soit utiliser la variable **q1a** présente dans n'importe quel module biographique en sélectionnant une seule observation par individu, ou prendre une base dont les trajectoires sont strictement continues (*house* ou *activity*) qui couvrent toutes les années de vie du répondant.

Pour transformer la base en individu-années (une ligne par année) on va effectuer une boucle qui réplique la ligne un nombre de fois égal à la valeur de **duree**, soit ici le nombre d'années d'existence du répondant jusqu'en 2008 compris. Comme cette étape sera répétée sur plusieurs bases, on va générer une macro (expand) avec deux arguments : la table utilisée et la variable de durée. Cette instruction génère la variable **j** qui compte la durée depuis le début de l'évènement.

```
%macro expand(table, duree);
data &table;
set &table;
do j=1 to &duree;
output;
end;
run;
%mend expand;
```

Avant expand

id	duree
1	1
2	6

Après expand

id	duree	j
1	1	1
2	6	1
2	6	2
2	6	3
2	6	4
2	6	5
2	6	6

On applique la macro expand à la base *year* et on crée les variables **year** et **age** à partir de la variable **j**. Pour initialiser les années à partir de l'année de naissance de l'individu on remonte d'un rang avec **j-1**, qui donne l'âge de l'individu au 31 décembre.

```
%expand (year, duree)
data year; set year;
year=q1a+j-1;
age= j-1;
run;
```

Appariée à d'autres modules en format individus-années, la base *year* permet de récupérer toutes les années «sans évènement».

```
data year(drop= num_union q1a j dur a2008); set year;
    label year= "year";
    label age= "age";
run;
```


	ident	year	age
38.	E103	1978	0
39.	E103	1979	1
40.	E103	1980	2
41.	E103	1981	3
42.	E103	1982	4
43.	E103	1983	5
44.	E103	1984	6
45.	E103	1985	7
46.	E103	1986	8
47.	E103	1987	9
48.	E103	1988	10
49.	E103	1989	11
50.	E103	1990	12
51.	E103	1991	13
52.	E103	1992	14
53.	E103	1993	15
54.	E103	1994	16
55.	E103	1995	17
56.	E103	1996	18
57.	E103	1997	19
58.	E103	1998	20
59.	E103	1999	21
60.	E103	2000	22
61.	E103	2001	23
62.	E103	2002	24
63.	E103	2003	25
64.	E103	2004	26
65.	E103	2005	27
66.	E103	2006	28
67.	E103	2007	29
68.	E103	2008	30

2) Base individus-années listant les pays de résidence du répondant

La démarche présentée ici, qui utilise la base *migration*, n'est pas la plus rapide, mais elle exploite *year*. Une alternative avec la base *house* est présentée en annexe.

Objectifs :

1. Rectifier la dernière année d'un évènement qui se termine l'année de l'enquête (p1)
2. Supprimer les doublons générés par les trajectoires continues (p2)

Base « *migration* »

variable name	storage type	display format	value label	variable label
ident	str7	%9s		Biographic individual identifier
num_mig	byte	%8.0g		Migration number
q601d	int	%17.0g	Q601D	Start year of the stay outside Senegal
q601f	int	%17.0g	Q601F	End year of the stay outside Senegal
q602	float	%41.0g	COUNTRY	Country of stay

	ident	num_mig	q601d	q601f	q602	
1.	E1	1	1991	1992	NIGERIA	
2.	E1	2	2001	.	SPAIN	
3.	E103	1	1992	1995	SPAIN	
4.	E103	2	1997	.	SPAIN	
5.	E147	1	1994	1995	PORTUGAL	
6.	E147	2	1995	.	SPAIN	
7.	F112	1	1968	.	FRANCE	
8.	F121	1	1993	1995	FRANCE	
9.	F121	2	1998	2008	ITALY	(p2)
10.	F121	3	2008	.	FRANCE	
11.	F127	1	1989	1996	FRANCE	(p2)
12.	F127	2	1996	1997	UK	(p2)
13.	F127	3	1997	1997	PORTUGAL	(p2)
14.	F127	4	1997	1998	SERBIA & MTN	(p2)
15.	F127	5	1998	2000	HUNGARY	(p2)
16.	F127	6	2000	.	FRANCE	
17.	I172	1	1995	2004	ITALY	
18.	I172	2	2007	.	ITALY	
19.	S011202	1	2005	2008	SPAIN	(p1)
20.	S079601	1	1954	1956	INDONESIA	(p2)
21.	S079601	2	1956	1958	ALGERIA	
22.	S079601	3	1958	1961	MADAGASCAR	
23.	S079601	4	1961	1963	FRANCE	

Cas (p1): Fin du dernier évènement l'année de l'enquête

Le temps étant discret avec comme unité l'année, le dernier évènement peut se terminer en 2008 avant la passation de l'entretien.

Si le dernier épisode de migration a eu lieu en 2008, on doit légèrement tronquer l'information pour créer la variable de durée en ramenant la fin de l'évènement à sa dernière année de survie au 31 décembre, soit l'année 2007. On génère tout d'abord la variable end_08 qui permet d'identifier ces situations et de ne pas tronquer une seconde fois ces observations lorsqu'on corrige p2.

Pour les épisodes non achevés au moment de l'entretien, on remplace la valeur manquante de **q601f** par 2008.

```
data migration;
  set mafe.migration;
  if q601f=2008 then end_08=1;
  if q601f=2008 then q601f=2007;
  if q601f=. then q601f=2008;
run;
```

	ident	num_mig	q601d	q601f	q602	end_08
1.	E1	1	1991	1992	NIGERIA	.
2.	E1	2	2001	.	SPAIN	.
3.	E103	1	1992	1995	SPAIN	.
4.	E103	2	1997	.	SPAIN	.
5.	E147	1	1994	1995	PORTUGAL	.
6.	E147	2	1995	.	SPAIN	.
7.	F112	1	1968	.	FRANCE	.
8.	F121	1	1993	1995	FRANCE	.
9.	F121	2	1998	2007	ITALY	1
10.	F121	3	2008	.	FRANCE	.
11.	F127	1	1989	1996	FRANCE	.
12.	F127	2	1996	1997	UNITED KINGDOM	.
13.	F127	3	1997	1997	PORTUGAL	.
14.	F127	4	1997	1998	SERBIA AND MONTENEGRO	.
15.	F127	5	1998	2000	HUNGARY	.
16.	F127	6	2000	.	FRANCE	.
17.	I172	1	1995	2004	ITALY	.
18.	I172	2	2007	.	ITALY	.
19.	S011202	1	2005	2007	SPAIN	1
20.	S079601	1	1954	1956	INDONESIA	.
21.	S079601	2	1956	1958	ALGERIA	.
22.	S079601	3	1958	1961	MADAGASCAR	.
23.	S079601	4	1961	1963	FRANCE	.

On applique de nouveau la macro `expand` après avoir créé la variable de durée sur chaque évènement (épisode de migration), et on génère une variable **year**.

```
data migration; set migration;
dur=q601f-q601d+1;
if dur=. then delete;
run;

%expand(migration, dur);

proc sort; by ident num_mig; run;
data migration;
set migration;
year=q601d+j-1;
by ident num_mig;
run;
```

Cas (p2) : Suppression des doublons générés par les trajectoires continues

Une trajectoire est considérée comme continue si le début d'un évènement a lieu la même année que la fin de l'évènement précédent³. Une même année se trouve donc inscrite sur deux lignes différentes : une fois dans la variable **q601f** en (t-1) et une fois dans la variable **q601d** en (t). L'allongement de la base provoque alors un doublon au niveau de l'année.

³ Ce type de transition est systématique dans les bases *activity*, *house* et *network*, l'individu ayant toujours un statut d'activité et un pays de résidence.

Exemple pour ident= S079601

	ident	num_mig	q601d	q601f	q602	year
113.	S079601	1	1954	1956	INDONESIA	1954
114.	S079601	1	1954	1956	INDONESIA	1955
115.	S079601	1	1954	1956	INDONESIA	1956
116.	S079601	2	1956	1958	ALGERIA	1956
117.	S079601	2	1956	1958	ALGERIA	1957
118.	S079601	2	1956	1958	ALGERIA	1958
119.	S079601	3	1958	1961	MADAGASCAR	1958
120.	S079601	3	1958	1961	MADAGASCAR	1959
121.	S079601	3	1958	1961	MADAGASCAR	1960
122.	S079601	3	1958	1961	MADAGASCAR	1961
123.	S079601	4	1961	1963	FRANCE	1961
124.	S079601	4	1961	1963	FRANCE	1962
125.	S079601	4	1961	1963	FRANCE	1963

Ici on va de nouveau légèrement tronquer l'information en privilégiant l'année du nouvel évènement. On supprime donc le doublon généré par **q601f**. On considère donc que tous les évènements se sont terminés au 31 décembre de l'année précédente. Les deux restrictions (**year** différent de 2008 et **end_08** différent de 1) évitent de tronquer une seconde fois l'information lorsque le dernier évènement s'est achevé en 2008 (p1).

```
data migration(drop=j dur); set migration;
if (q601f=year & year ne 2008 & end_08 ne 1 ) then delete;
run;
```

	ident	num_mig	q601d	q601f	q602	year
106.	S079601	1	1954	1956	INDONESIA	1954
107.	S079601	1	1954	1956	INDONESIA	1955
108.	S079601	2	1956	1958	ALGERIA	1956
109.	S079601	2	1956	1958	ALGERIA	1957
110.	S079601	3	1958	1961	MADAGASCAR	1958
111.	S079601	3	1958	1961	MADAGASCAR	1959
112.	S079601	3	1958	1961	MADAGASCAR	1960
113.	S079601	4	1961	1963	FRANCE	1961
114.	S079601	4	1961	1963	FRANCE	1962

On remarque que lorsque la trajectoire continue a lieu en 2008 (F121), le problème est réglé en amont lorsqu'on a remplacé la valeur de **q601f** par 2007.

Il ne reste plus alors qu'à apparier la base *migration* avec la base *year* créée plus haut, en triant préalablement les observations par l'identifiant individuel **ident** et la variable **year**.

```
proc sort out= migration;
by ident year; run;

proc sort data=year;
by ident year; run;

data loc_resp (rename=(q602= loc_resp) drop=end_08);
merge year migration;
by ident year;
run;
```

La variable renommée **loc_resp** (précédemment **q602**) a maintenant des valeurs manquantes pour les années n'appartenant pas à un épisode migratoire, c'est à dire les années où le répondant résidait au Sénégal. Il suffit de recoder la variable en remplaçant ces valeurs manquantes par le code pays du Sénégal (99341).

```
data loc_resp (keep = ident year age loc_resp);
  set loc_resp;
  if loc_resp=. then loc_resp=99341;
run;
```

```
+-----+
|  ident  year  age  loc_resp |
+-----+
291. | S079601  1936   0  SENEGAL |
292. | S079601  1937   1  SENEGAL |
293. | S079601  1938   2  SENEGAL |
294. | S079601  1939   3  SENEGAL |
295. | S079601  1940   4  SENEGAL |
296. | S079601  1941   5  SENEGAL |
297. | S079601  1942   6  SENEGAL |
298. | S079601  1943   7  SENEGAL |
299. | S079601  1944   8  SENEGAL |
300. | S079601  1945   9  SENEGAL |
+-----+
301. | S079601  1946  10  SENEGAL |
302. | S079601  1947  11  SENEGAL |
303. | S079601  1948  12  SENEGAL |
304. | S079601  1949  13  SENEGAL |
305. | S079601  1950  14  SENEGAL |
306. | S079601  1951  15  SENEGAL |
307. | S079601  1952  16  SENEGAL |
308. | S079601  1953  17  SENEGAL |
309. | S079601  1954  18  INDONESIA |
310. | S079601  1955  19  INDONESIA |
311. | S079601  1956  20  ALGERIA |
312. | S079601  1957  21  ALGERIA |
313. | S079601  1958  22  MADAGASCAR |
314. | S079601  1959  23  MADAGASCAR |
315. | S079601  1960  24  MADAGASCAR |
316. | S079601  1961  25  FRANCE |
317. | S079601  1962  26  FRANCE |
318. | S079601  1963  27  SENEGAL |
319. | S079601  1964  28  SENEGAL |
320. | S079601  1965  29  SENEGAL |
321. | S079601  1966  30  SENEGAL |
|      .      .      .      . |
|      .      .      .      . |
|  Idem      .      .      . |
|      .      .      .      . |
|      .      .      .      . |
363. | S079601  2008  72  SENEGAL |
+-----+
```

3) Base individus-années listant les pays de résidence des conjoints

Objectif

Transformation d'une base de données pour traiter les situations d'overlapping (polygamie) avec l'instruction `proc transpose` (création d'une macro pour simplifier le programme)

a) Pays de résidence des partenaires

Dans un premier temps on va récupérer les informations sur les pays de résidence des conjoints dans la base « network ». On fait comme dans la section précédente après avoir sélectionné les partenaires dans les personnes du réseau (q901=1).

Base *network.dta*

variable name	storage type	display format	value label	variable label
ident	str7	%9s		Biographic individual identifier
num_netmig	float	%9.0g		Number of migrant's migration
q904d	int	%17.0g	Q904	Beginning year of the migration
q904f	int	%17.0g	Q904	Ending year of the migration
country	float	%15.0g	COUNTRY	Country
q901	byte	%10.0g	Q901	Link with Ego
id_uni	str10	%10s		Union ID

	ident	num_netmig	q904d	q904f	country	q901	id_uni
1.	E103	1	1979	.	SPAIN	Partner	U_E1032
2.	E103	1	1989	.	SPAIN	Partner	U_E1031
3.	F112	1	1965	.	FRANCE	Partner	U_F1121
4.	I172	1	1986	1990	GUINEA	Partner	U_I1722
5.	I172	2	1990	1991	SENEGAL	Partner	U_I1722
6.	I172	3	1991	.	ITALY	Partner	U_I1722
7.	S011202	1	2000	2002	ITALY	Partner	U_S0112021
8.	S011202	2	2002	.	SPAIN	Partner	U_S0112021

Note : dans le tableau ci-dessus on a seulement sélectionné les informations relatives aux partenaires (q901=1)

```
data net_conj; set mafe.network;
if q901=1;
if q904f=2008 then end_08=1;
if q904f=2008 then q904f=2007;
if q904f=. then q904f=2008;
run;

data net_conj; set net_conj;
dur=q904f-q904d+1;
if dur=. then delete;
run;

%expand(net_conj, dur);

proc sort;
by ident id_uni num_netmig; run;

data net_conj; set net_conj;
year=q904d+j-1;
```

```

by ident id_uni num_netmig;
run;

data net_conj(keep= ident id_uni country q904d q904f year);
  set net_conj;
if (q904f=year & year ne 2008 & end_08 ne 1) then delete;
run;
proc sort; by id_uni year; run;

```

b) Les périodes de séparation du répondant

variable name	storage type	display format	value label	variable label
ident	str7	%9s		Biographic individual identifier
num_union	byte	%9.0g		Number of the union
id_uni	str10	%10s		Union ID
q102d	int	%7.0g	Q102D	Start years of the relationship
q102f	int	%7.0g	Q102F	End year of the relationship
q103	byte	%10.0g	Q103	Type of dissolution
q105	int	%8.0g	Q105	In which year did you get married
q107	float	%15.0g	COUNTRY	In which country was she/he borned

Dans les informations concernant les unions, on dispose des causes d’une éventuelle rupture. Avec l’allongement de la base, cette information n’a été répliquée que sur les années où l’union était en cours, mais pas après. Pour résoudre ce problème, on va créer une base *separation* qui donne la cause de rupture à partir de l’année de la séparation et jusqu’en 2008.

Après l’allongement de la base *union*

	ident	year	q1031	num_union
155.	F127	1984	Separation or divorce	1
156.	F127	1985	Separation or divorce	1
157.	F127	1986	Separation or divorce	1
158.	F127	1987	Separation or divorce	1
159.	F127	1988	Separation or divorce	1

Après la création de la base *separation*, expand et appariement avec la base *union*

	ident	year	q1031	num_union
155.	F127	1984	.	1
156.	F127	1985	.	1
157.	F127	1986	.	1
158.	F127	1987	.	1
159.	F127	1988	.	1
160.	F127	1989	Separation or divorce	.
161.	F127	1990	Separation or divorce	.
162.	F127	1991	Separation or divorce	.
163.	F127	1992	Separation or divorce	.
164.	F127	1993	Separation or divorce	.
165.	F127	1994	Separation or divorce	.
166.	F127	1995	Separation or divorce	.
167.	F127	1996	Separation or divorce	.
168.	F127	1997	Separation or divorce	.
169.	F127	1998	Separation or divorce	.
170.	F127	1999	Separation or divorce	.
171.	F127	2000	Separation or divorce	.
172.	F127	2001	Separation or divorce	.
173.	F127	2002	Separation or divorce	.

174.	F127	2003	Separation or divorce	.
175.	F127	2004	Separation or divorce	.
176.	F127	2005	Separation or divorce	.
177.	F127	2006	Separation or divorce	.
178.	F127	2007	Separation or divorce	.
179.	F127	2008	Separation or divorce	.

Pour générer la durée on va prendre **q102f** comme variable de début de l'évènement et 2008 pour marquer la fin⁴.

```
data separation (keep= ident num_union q102d q102f q103 id_uni);
set mafe.union;
if q103=. then delete;
run;
data separation; set separation;
dur=2008-q102f+1;
run;
%expand(separation, dur);
data separation; set separation;
year=q102f+j-1;
by id_uni;
run;
data separation (keep=ident num_union id_uni q103 year);
set separation;
run;
proc sort; by id_uni year; run;
```

c) Les périodes d'union du répondant

Avant de traiter le problème de l'overlapping, on transforme la base *union* en format individus-années, et on l'apparie aux bases *séparation* et *network*.

```
data union(drop=q103); set mafe.union;
if q102f=2008 then end_08=1;
if q102f=2008 then q102f=2007;
if q102f=. then q102f=2008;
run;

data union;
set union;
dur=q102f-q102d+1;
if dur=. then delete;
run;

%expand(union, dur);

data union; set union;
year=q102d+j-1;
run;

data union(keep =ident num_union id_uni q102d q102f q103 q105 q107 year);
```

⁴ Si on souhaitait créer des épisodes de séparation se terminant l'année de début de l'union suivante, il suffirait de remonter cette information d'une ligne avec la fonction `lag` appliquée à la variable `q102d`. Avec cette démarche on fait disparaître la cause de la rupture lorsque la nouvelle mise en union a lieu la même année que la séparation (trajectoire continue) où lorsque la séparation se fait dans une union polygame (évènement simultané/overlapping).


```

set union;

if (q102f=year & year ne 2008 & end_08 ne 1) then delete;
run;

proc sort; by id_uni year;
run;

/* merge union net_conj separation */
data loc_conj(drop=q904d q904f q102d q102f end_08); merge net_conj union
separation ; by id_uni year;
if num_union=. then delete;

```

A ce stade, le pays de résidence du partenaire (country) avant la première migration se trouve en valeur manquante si le conjoint n'est pas né dans un des pays européens où l'enquête a été réalisée⁵. On utilise la variable **q107** (pays de naissance du conjoint) de la base *union.dta* pour récupérer cette information.

```

if country=. & q103=. then country=q107;
if q103 ne . then country=.;
run;

data loc_conj; set loc_conj;
rename country=loc_conj;
label loc_conj="loc_conj";
run;

```

Le problème de l'overlapping

L'**overlapping** s'observe lorsqu'au moins deux événements de même type se déroulent simultanément. Une situation classique d'overlapping est, par exemple, lorsqu'un individu a plusieurs enfants vivants en même temps. Ici, il s'agit des unions polygames.

	ident	num_un~n	id_uni	q102d	q102f	q103	q105	q107	
1.	E1	1	U_E11	1990	.	.	1993	SENEGAL	
2.	E103	1	U_E1031	1992	1996	Separation	1992	SENEGAL	
3.	E103	2	U_E1032	2001	2008	Separation	2001	SPAIN	
4.	E147	1	U_E1471	1979	.	.	1979	SENEGAL	
5.	E147	2	U_E1472	1990	.	.	1990	SENEGAL	Overlapping à partir
6.	E147	3	U_E1473	2004	.	.	.	SENEGAL	de 1990
7.	F112	1	U_F1121	1967	.	.	1968	SENEGAL	
8.	F121	1	U_F1211	1998	.	.	1998	SENEGAL	
9.	I172	1	U_I1721	1985	1993	Deceased	1987	SENEGAL	
10.	I172	2	U_I1722	2002	.	.	2004	SENEGAL	
11.	S011202	1	U_S0112021	1999	.	.	1999	SENEGAL	
12.	S079601	1	U_S0796011	1965	.	.	1965	SENEGAL	
13.	S079601	2	U_S0796012	1977	.	.	1977	SENEGAL	Overlapping à partir
14.	S079601	3	U_S0796013	1984	1988	Deceased	1984	SENEGAL	de 1977

⁵ Par exemple, dans MAFE Sénégal, ceux qui ne sont pas nés en Espagne, France ou Italie (se reporter à la description du module network dans le document « presentation of the dataset »).

Dans l'exemple ci-dessous, après l'allongement de la base, les années renseignées par la variable **year** sont uniques jusqu'en 1989 (une union), doublées de 1990 à 2003 (deux unions simultanées), puis triplées de 2004 à l'année de l'enquête (trois unions simultanées).

	ident	num_un~n	q103	q105	year	loc_conj
30.	E147	1	.	1979	1979	SENEGAL
31.	E147	1	.	1979	1980	SENEGAL
32.	E147	1	.	1979	1981	SENEGAL
33.	E147	1	.	1979	1982	SENEGAL
34.	E147	1	.	1979	1983	SENEGAL
35.	E147	1	.	1979	1984	SENEGAL
36.	E147	1	.	1979	1985	SENEGAL
37.	E147	1	.	1979	1986	SENEGAL
38.	E147	1	.	1979	1987	SENEGAL
39.	E147	1	.	1979	1988	SENEGAL
40.	E147	1	.	1979	1989	SENEGAL
41.	E147	1	.	1979	1990	SENEGAL
42.	E147	2	.	1990	1990	SENEGAL
43.	E147	1	.	1979	1991	SENEGAL
44.	E147	2	.	1990	1991	SENEGAL
45.	E147	1	.	1979	1992	SENEGAL
46.	E147	2	.	1990	1992	SENEGAL
47.	E147	1	.	1979	1993	SENEGAL
48.	E147	2	.	1990	1993	SENEGAL
49.	E147	1	.	1979	1994	SENEGAL
50.	E147	2	.	1990	1994	SENEGAL
51.	E147	1	.	1979	1995	SENEGAL
52.	E147	2	.	1990	1995	SENEGAL
53.	E147	1	.	1979	1996	SENEGAL
54.	E147	2	.	1990	1996	SENEGAL
55.	E147	1	.	1979	1997	SENEGAL
56.	E147	2	.	1990	1997	SENEGAL
57.	E147	1	.	1979	1998	SENEGAL
58.	E147	2	.	1990	1998	SENEGAL
	.					
	.					
	Idem de 1999 à 2003					
	.					
69.	E147	1	.	1979	2004	SENEGAL
70.	E147	2	.	1990	2004	SENEGAL
71.	E147	3	.	.	2004	SENEGAL
72.	E147	1	.	1979	2005	SENEGAL
73.	E147	2	.	1990	2005	SENEGAL
74.	E147	3	.	.	2005	SENEGAL
75.	E147	1	.	1979	2006	SENEGAL
76.	E147	2	.	1990	2006	SENEGAL
77.	E147	3	.	.	2006	SENEGAL
78.	E147	1	.	1979	2007	SENEGAL
79.	E147	2	.	1990	2007	SENEGAL
80.	E147	3	.	.	2007	SENEGAL
81.	E147	1	.	1979	2008	SENEGAL
82.	E147	2	.	1990	2008	SENEGAL
83.	E147	3	.	.	2008	SENEGAL

Les tableaux ci-dessous illustrent la démarche avec une variable propre à chaque évènement (X) qui ne reste constante au cours du temps (exemple : le sexe d'un individu).

De (t1 à (t4) l'évènement est unique ; de (t5) à (t9), il s'observe 2 fois simultanément ; et en (t10) suite à la fin de l'évènement 1, il est à nouveau unique.

Avant reshape

t	num_event	X
t1	1	v1
t2	1	v1
t3	1	v1
t3	1	v1
t4	1	v1
t5	1	v1
t5	2	v2
t6	1	v1
t6	2	v2
t7	1	v1
t7	2	v2
t8	1	v1
t8	2	v2
t9	1	v1
t9	2	v2
t10	2	v2

Après reshape

t	num_event_1	X_1	num_event_2	X_2
t1	1	v1	.	.
t2	1	v1	.	.
t3	1	v1	.	.
t3	1	v1	.	.
t4	1	v1	.	.
t5	1	v1	2	v2
t6	1	v1	2	v2
t7	1	v1	2	v2
t8	1	v1	2	v2
t9	1	v1	2	v2
t10	.	.	2	v2

Pour réaliser cette transformation de la base, on utilise l'instruction **proc transpose** (http://www.stata.com/bookstore/stata12/pdf/d_reshape.pdf) afin d'obtenir un bloc de variables pour chaque union. Comme cette procédure SAS ne tolère qu'une seule variable non constante d'une union à l'autre, on va alléger la syntaxe avec une macro commande.

Etape 1 : on va créer un identifiant (**id**) unique combinant l'identifiant du répondant (**ident**) et l'année de l'observation (**year**). Il s'agit de la variable d'identification de l'instruction **proc transpose**, soit une ligne par individu-année. On transforme tout d'abord la variable **year** en format caractère en créant la variable **y** avec l'instruction **put**, puis on concatène les deux variables avec **||**.

```
data loc_conj(drop= y); set loc_conj;
y=put(year, $4.);
id=ident || y;
run ;
```

Etape2:

Pour corriger ces situations d'overlapping on va privilégier ici la solution la plus simple, soit l'élargissement de la base sur chaque séquence d'union (**num_union**).

Cela va avoir pour conséquence de créer aussi des blocs de variables pour les individus dont les unions se sont succédées sans se chevaucher (période union1=> rupture union1 +période d'au moins un an sans union => période union2.....). C'est avec cette option que la suite du tutoriel sera développée⁶.

⁶ Une variante est d'effectuer le reshape de la base uniquement pour les personnes pour lesquelles on observe de l'overlapping. Avec cette option, le programme de création des variables de la seconde partie devra être modifiée.

Au préalable, deux variables vont être créées pour la seconde partie du tutoriel: la variable **nbu** qui renvoie pour chaque année le nombre total d'union en cours, et **num_uni** qui donne comme **num_union**⁷ le rang de l'union, mais seulement les années où elle n'a pas été rompue.

La variable **nbu** est créée à partir de l'output de proc summary. Par id, on va conserver le nombre total d'observations (N), qui donne le nombre d'unions en cours chaque année. La base nbu sera appariée plus tard avec la base d'analyse.

```
/* variable num_uni = num_union*/
data loc_conj; set loc_conj;
if q103=. then num_uni=num_union;
run;
proc sort data= loc_conj; by id; run;

/* variable nbu => base nbu*/
proc summary data=loc_conj;
class id; var num_uni;
where num_uni ne .;
output out= nbu(drop= _type_ _freq_) N= nbu;
run;
data nbu; set nbu;
if id="" then delete;
proc sort data= nbu; by id; run;
```

L'ensemble des variables propres aux unions et donc susceptibles de varier d'une union à l'autre doivent entrer dans l'instruction proc transpose ([Aide SAS 9.4](#)), il s'agit ici de q103, q105, loc_conj et num_uni. Pour chacune de ces 4 variables on va générer une base (bi : b1,...,b4).

```
%macro reshape(var,i);
proc transpose data= loc_conj
out= b&i prefix=&var;
by id; id num_union;
var &var; run;
data b&i (drop=_NAME_ _LABEL_); set b&i; run;
%mend reshape;
```

```
%reshape(q103,1);
%reshape(q105,2);
%reshape(loc_conj,3);
%reshape(num_uni,4);
```

Les 4 quatre bases sont finalement appariées pour créer la base loc_conj.

```
data c2; merge b1 b2; by id; run;
data c3; merge c2 b3; by id; run;
data c4; merge c3 b4; by id; run;
data loc_conj1; merge c4 nbu; by id; run;
proc sort data= loc_conj1; by id; run;

proc datasets;
delete c2-c4 b1-b4;
```

⁷ **num_union** disparaît avec l'exécution de proc transpose.

```
run;
```

En reprenant l'illustration précédente, le résultat obtenu après le `reshape` est le suivant (où **nbe** représente le nombre d'événements) :

ident	T	id	num_event_1	var_1	num_event_2	var_2	nbe
1	t1	1t1	1	v1	.	.	1
1	t2	1t2	1	v1	.	.	1
1	t3	1t3	1	v1	.	.	1
1	t3	1t4	1	v1	.	.	1
1	t4	1t4	1	v1	.	.	1
1	t5	1t5	1	v1	2	v2	2
1	t6	1t6	1	v1	2	v2	2
1	t7	1t7	1	v1	2	v2	2
1	t8	1t8	1	v1	2	v2	2
1	t9	1t9	1	v1	2	v2	2
1	t10	1t10	.	.	2	v2	1

Dans le cas de MAFE, Le résultat du `reshape` avec les variables de localisation des partenaires donne:

	ident	year	loc_conj1	loc_conj2	loc_conj3	nbu
680.	E147	1979	SENEGAL	.	.	1
681.	E147	1980	SENEGAL	.	.	1
682.	E147	1981	SENEGAL	.	.	1
683.	E147	1982	SENEGAL	.	.	1
684.	E147	1983	SENEGAL	.	.	1
685.	E147	1984	SENEGAL	.	.	1
686.	E147	1985	SENEGAL	.	.	1
687.	E147	1986	SENEGAL	.	.	1
688.	E147	1987	SENEGAL	.	.	1
689.	E147	1988	SENEGAL	.	.	1
690.	E147	1989	SENEGAL	.	.	1
691.	E147	1990	SENEGAL	SENEGAL	.	2
692.	E147	1991	SENEGAL	SENEGAL	.	2
693.	E147	1992	SENEGAL	SENEGAL	.	2
694.	E147	1993	SENEGAL	SENEGAL	.	2
695.	E147	1994	SENEGAL	SENEGAL	.	2
696.	E147	1995	SENEGAL	SENEGAL	.	2
697.	E147	1996	SENEGAL	SENEGAL	.	2
698.	E147	1997	SENEGAL	SENEGAL	.	2
699.	E147	1998	SENEGAL	SENEGAL	.	2
700.	E147	1999	SENEGAL	SENEGAL	.	2
701.	E147	2000	SENEGAL	SENEGAL	.	2
702.	E147	2001	SENEGAL	SENEGAL	.	2
703.	E147	2002	SENEGAL	SENEGAL	.	2
704.	E147	2003	SENEGAL	SENEGAL	.	2
705.	E147	2004	SENEGAL	SENEGAL	SENEGAL	3
706.	E147	2005	SENEGAL	SENEGAL	SENEGAL	3
707.	E147	2006	SENEGAL	SENEGAL	SENEGAL	3
708.	E147	2007	SENEGAL	SENEGAL	SENEGAL	3
709.	E147	2008	SENEGAL	SENEGAL	SENEGAL	3

c) Appariement des bases répondants et conjoints

On peut finalement fusionner les informations sur le(s) partenaire(s) à celles du répondant, sans oublier la base *nbu* qui donne le nombre d'unions en cours. Au préalable, on génère la variable *id* dans la base *loc_resp* afin de pouvoir l'apparier avec la base *nbu*. Dans le cadre de ce tutoriel, on va limiter l'âge de l'individu à 14 ans qui correspond dans les observations sélectionnées à l'âge le plus jeune pour la première union déclarée.

```

data loc_resp (drop= y); set loc_resp;
y=put(year, $4.);
id=ident || y;
run ;
proc sort; by id; run;

data analys; merge loc_resp loc_conj1 nbu; by id;
if age<14 then delete;
if nbu=. then nbu=0;
run;

```

On réordonne enfin les variables avec l'instruction `retain` pour réunir toutes les informations de chaque union dans un même bloc.

```

data analys(drop=id);
retain ident year age loc_resp
      num_uni1 q1031 q1051 loc_conj1
      num_uni2 q1032 q1052 loc_conj2
      num_uni3 q1033 q1053 loc_conj3
      num_uni4 q1034 q1054 loc_conj4
      num_uni5 q1035 q1055 loc_conj5
      nbu;
set analys;
run;

proc datasets;
delete loc_conj loc_conj1 loc_resp1 migration nbu net_conj separation
union year ;
run;

```

	ident	year	age	loc_resp	loc_conj1	num_uni1	loc_conj2	num_uni2
41.	E147	1968	14	SENEGAL
42.	E147	1969	15	SENEGAL
43.	E147	1970	16	SENEGAL
44.	E147	1971	17	SENEGAL
45.	E147	1972	18	SENEGAL
46.	E147	1973	19	SENEGAL
47.	E147	1974	20	SENEGAL
48.	E147	1975	21	SENEGAL
49.	E147	1976	22	SENEGAL
50.	E147	1977	23	SENEGAL
51.	E147	1978	24	SENEGAL
52.	E147	1979	25	SENEGAL	SENEGAL	1	.	.
53.	E147	1980	26	SENEGAL	SENEGAL	1	.	.
54.	E147	1981	27	SENEGAL	SENEGAL	1	.	.
55.	E147	1982	28	SENEGAL	SENEGAL	1	.	.
56.	E147	1983	29	SENEGAL	SENEGAL	1	.	.
57.	E147	1984	30	SENEGAL	SENEGAL	1	.	.
58.	E147	1985	31	SENEGAL	SENEGAL	1	.	.
59.	E147	1986	32	SENEGAL	SENEGAL	1	.	.
60.	E147	1987	33	SENEGAL	SENEGAL	1	.	.
61.	E147	1988	34	SENEGAL	SENEGAL	1	.	.
62.	E147	1989	35	SENEGAL	SENEGAL	1	.	.
63.	E147	1990	36	SENEGAL	SENEGAL	1	SENEGAL	2
64.	E147	1991	37	SENEGAL	SENEGAL	1	SENEGAL	2
65.	E147	1992	38	SENEGAL	SENEGAL	1	SENEGAL	2
66.	E147	1993	39	SENEGAL	SENEGAL	1	SENEGAL	2
67.	E147	1994	40	PORTUGAL	SENEGAL	1	SENEGAL	2

68.		E147	1995	41	SPAIN	SENEGAL	1	SENEGAL	2	
69.		E147	1996	42	SPAIN	SENEGAL	1	SENEGAL	2	
70.		E147	1997	43	SPAIN	SENEGAL	1	SENEGAL	2	

71.		E147	1998	44	SPAIN	SENEGAL	1	SENEGAL	2	
72.		E147	1999	45	SPAIN	SENEGAL	1	SENEGAL	2	
73.		E147	2000	46	SPAIN	SENEGAL	1	SENEGAL	2	
74.		E147	2001	47	SPAIN	SENEGAL	1	SENEGAL	2	
75.		E147	2002	48	SPAIN	SENEGAL	1	SENEGAL	2	

76.		E147	2003	49	SPAIN	SENEGAL	1	SENEGAL	2	
77.		E147	2004	50	SPAIN	SENEGAL	1	SENEGAL	2	
78.		E147	2005	51	SPAIN	SENEGAL	1	SENEGAL	2	
79.		E147	2006	52	SPAIN	SENEGAL	1	SENEGAL	2	
80.		E147	2007	53	SPAIN	SENEGAL	1	SENEGAL	2	

81.		E147	2008	54	SPAIN	SENEGAL	1	SENEGAL	2	

Note : seulement pour les deux premières unions du répondant. Les variable q1031, q1032, q1051 et q1052 ne sont pas reportées dans la capture de la base.

PARTIE II. Exemples de création de variables d'analyse

L'objectif est de générer des variables de type TVC (Time Varying Covariate) sur le statut matrimonial du répondant et sur les localisations (pays de résidence) des différents partenaires en distinguant les unions « transnationale », « réunifiée » et « toujours unifiée ». L'intérêt des deux dernières modalités est que les valeurs dépendent de celles prises dans le passé.

1) Création des variables de statut matrimonial

Objectif

Reports des informations sur les années de mariages sur les bonnes lignes individu-année

Les années de mariages

Ce type de problème se posait avec les motifs de rupture, mais a été résolu en créant la base separation. Pour les années de mariage (**q105**), l'information est aussi présente sur les lignes où l'union n'était encore qu'informelle. On choisit de la supprimer les années où l'union n'était qu'informelle, et pour la forme on remplace l'année par la modalité 1 associée au label « mariage ». Il suffit de comparer l'année du mariage (q105X) à l'année de l'observation. Si la valeur de **q105** est inférieure à celle de **year**, cela indique que l'union ne correspondait pas à un mariage.

+-----+ ident year num_uni1 q1051 +-----+				
5.	E1	1990	1	1993
6.	E1	1991	1	1993
7.	E1	1992	1	1993
8.	E1	1993	1	1993
9.	E1	1994	1	1993
10.	E1	1995	1	1993
	.			
	.			
	.	idem		
	.			
20.	E1	2008	1	1993
+-----+				

Comme il peut y avoir plusieurs mariages par individu, on va utiliser l'instruction `array` ([Aide Sas 9.4](#)) dans une étape `data` pour effectuer la correction sur toutes les unions plus rapidement. Les deux premières lignes de l'instruction donnent les valeurs prises par la variable **i** (numéro d'occurrence => rang de l'union) pour **q105i** et **num_unii**, soit ici **i=1,2,3,4,5**. La seconde initialise la boucle, et la dernière exécute la correction pour chaque valeur de **i**, donc sur chaque sur chaque union.


```

data analys(drop=i); set analys;
/*Valeurs prises par i*/
array q105{5} q1051-q1055;
array num_uni{5} num_uni1-num_uni5;
/*Initialisation de la boucle*/
do i=1 to 5;
/*Instructions*/
if q105{i}>year & num_uni{i} ne . then q105{i}=.;
if q105{i} ne . then q105{i}=1;
end;
run;

```

	ident	year	num_uni1	q1051
5.	E1	1990	1	.
6.	E1	1991	1	.
7.	E1	1992	1	.
8.	E1	1993	1	Marriage
9.	E1	1994	1	Marriage
10.	E1	1995	1	Marriage
	.			
	.			
	.	idem		
	.			
20.	E1	2008	1	Marriage

b) Variables dichotomiques de statut matrimonial

Les variables **C**, **V**, **D**, **U**, **M** indiquent si l'individu est –pour chaque année donnée– célibataire, veuf, séparé ou divorcé, en union informelle, marié. Si C ne peut être égale à 1 qu'avant la première union, les autres situations peuvent se chevaucher : un individu peut être marié (M=1) et divorcé ou veuf d'une union précédente (D=1 ou V=1).

Sauf pour C qui est créée directement car elle ne nécessite d'utiliser que num_uni1 et q1031, on récupère dans un premier temps l'information des différents statuts pour chaque union (1 si oui). Ensuite, on additionne les valeurs obtenues avec la fonction sum pour générer les variables **v_s**, **d_s**, **u_s** et **m_s**. Si le résultat est supérieur à 0 alors l'individu est dans le statut en question.

Comme pour la correction des années de mariage, on automatise les instructions avec array.

```

/* Dummy for being alone and never in partnership */
data analys; set analys;
if num_uni1=. & q1031=. then C=1;
if C=. then C=0;
label C= " Alone, never in partnership =1" ;
run;

/* Dummy for widowhood */
data analys(drop= i V_:); set analys;
array q103[5] q1031-q1035;
array V_[5] V_1-V_5;
do i=1 to 5;
    if q103[i]=2 then V_[i]=1;
end;
v_s = sum(of V_[*]);
if v_s>0 then V=1;
if V=. then V=0;
label V= " Widowed =1" ;

```

```

run;

/** Dummy for separation **/
data analys(drop= i D_:); set analys;
array q103[5] q1031-q1035;
array D_[5] D_1-D_5;
do i=1 to 5;
    if q103[i]=1 then D_[i]=1;
end;
d_s = sum(of D_[*]);
if d_s>0 then D=1;
if D=. then D=0;
label D= " Divorced or separated =1" ;
run;

/** Dummy for free union **/
data analys(drop= i U_:); set analys;
array q105[5] q1051-q1055;
array num_uni[5] num_uni1-num_uni5;
array U_[5] U_1-U_5;
do i=1 to 5;
    if num_uni[i] ne . & q105[i]=. then U_[i]=1;
end;
u_s = sum(of U_[*]);
if u_s>0 then U=1;
if U=. then U=0;
label U= " Partnership, not married=1" ;
run;

/** Dummy for marriage**/
data analys(drop= i M_:); set analys;
array q105[5] q1051-q1055;
array num_uni[5] num_uni1-num_uni5;
array M_[5] M_1-M_5;
do i=1 to 5;
    if num_uni[i] ne . & q105[i] ne . then M_[i]=1;
end;
m_s = sum(of M_[*]);
if m_s>0 then M=1;
if M=. then M=0;
label M= " Married=1" ;
run;

```

	+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+														
	ident	year		q1031	q1051	num_uni1	q1032	q1052	num_uni2	C	V	D	U	M	
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----															
180.	I172	1985		.	.	1	.	.	.	0	0	0	1	0	
181.	I172	1986		.	.	1	.	.	.	0	0	0	1	0	
182.	I172	1987		.	Marriage	1	.	.	.	0	0	0	0	1	
183.	I172	1988		.	Marriage	1	.	.	.	0	0	0	0	1	
184.	I172	1989		.	Marriage	1	.	.	.	0	0	0	0	1	
185.	I172	1990		.	Marriage	1	.	.	.	0	0	0	0	1	
186.	I172	1991		.	Marriage	1	.	.	.	0	0	0	0	1	
187.	I172	1992		.	Marriage	1	.	.	.	0	0	0	0	1	
188.	I172	1993	Partner deceased	0	1	0	0	0	
189.	I172	1994	Partner deceased	0	1	0	0	0	
190.	I172	1995	Partner deceased	0	1	0	0	0	
191.	I172	1996	Partner deceased	0	1	0	0	0	
192.	I172	1997	Partner deceased	0	1	0	0	0	
193.	I172	1998	Partner deceased	0	1	0	0	0	
194.	I172	1999	Partner deceased	0	1	0	0	0	
195.	I172	2000	Partner deceased	0	1	0	0	0	
196.	I172	2001	Partner deceased	0	1	0	0	0	
197.	I172	2002	Partner deceased	2	0	1	0	1	0	
198.	I172	2003	Partner deceased	2	0	1	0	1	0	
199.	I172	2004	Partner deceased	Marriage	2	0	1	0	0	1	
200.	I172	2005	Partner deceased	Marriage	2	0	1	0	0	1	

201.		I172	2006	Partner deceased	.	.	.	Marriage	2	0	1	0	0	1	
202.		I172	2007	Partner deceased	.	.	.	Marriage	2	0	1	0	0	1	
203.		I172	2008	Partner deceased	.	.	.	Marriage	2	0	1	0	0	1	

d) Variable synthétique de statut matrimonial

A partir de ces 5 nouvelles variables, on peut synthétiser l'information en créant la variable **marital** variable qui chaque année donne une information sur le statut matrimonial du répondant. L'ensemble des modalités créées ici n'est qu'une possibilité parmi d'autres.

```
data analys; set analys;
    if C=1 then marital=1;
else if D=1 & nbu=0 then marital=2;
else if V=1 & nbu=0 then marital=3;
else if U=1 & nbu=1 then marital=4;
else if M=1 & nbu=1 then marital=5;
else if U=1 & M=0 & nbu>1 then marital=6;
else if M=1 & nbu>1 then marital=7;
run;

proc formats;
value marital
    1='Alone, never in union'
    2='Alone, separated'
    3='Alone, widowed'
    4='Monogamy, informal union'
    5='Monogamy, married'
    6='Polygamy, only informal unions'
    7='Polygamy, at least one marriage';

proc freq data=analys;
table marital;
format marital marital.;
run;
```

	ident	year	num_unil	q1031	q1051	nbu	C	V	D	U	M	marital
82.		F112	1965	.	.	.	1	0	0	0	0	Alone, never in union
83.		F112	1966	.	.	.	1	0	0	0	0	Alone, never in union
84.		F112	1967	1	.	1	0	0	0	1	0	Monogamy, informal union
85.		F112	1968	1	.	Marriage	1	0	0	0	1	Monogamy, married
86.		F112	1969	1	.	Marriage	1	0	0	0	1	Monogamy, married
87.		F112	1970	1	.	Marriage	1	0	0	0	1	Monogamy, married
88.		F112	1971	1	.	Marriage	1	0	0	0	1	Monogamy, married
89.		F112	1972	1	.	Marriage	1	0	0	0	1	Monogamy, married
90.		F112	1973	1	.	Marriage	1	0	0	0	1	Monogamy, married
91.		F112	1974	1	.	Marriage	1	0	0	0	1	Monogamy, married
92.		F112	1975	1	.	Marriage	1	0	0	0	1	Monogamy, married
93.		F112	1976	1	.	Marriage	1	0	0	0	1	Monogamy, married
94.		F112	1977	1	.	Marriage	1	0	0	0	1	Monogamy, married
95.		F112	1978	1	.	Marriage	1	0	0	0	1	Monogamy, married
96.		F112	1979	1	.	Marriage	1	0	0	0	1	Monogamy, married
		.										
		etc										
		.										
125.		F112	2008	1	.	Marriage	1	0	0	0	1	Monogamy, married

marital	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
Alone, never in union	75	21.74	75	21.74
Alone, separated	6	1.74	81	23.48
Alone, widowed	9	2.61	90	26.09
Monogamy, informal union	24	6.96	114	33.04
Monogamy, married	132	38.26	246	71.30
Polygamy, only informal unions	9	2.61	255	73.91
Polygamy, at least one marriage	90	26.09	345	100.00

2) Variable de comparaison des pays de résidence des répondants et des partenaires

Objectifs

- Comparaisons des pays de résidence avec unions simultanées: transnationalité ou non des unions.
- Dépendance de la valeur d'une variable aux valeurs passées : « unions réunifiées » ou « toujours unifiées ».
- Stratification des comparaisons par structure d'union.

a) Comparaison des codes pays entre le répondant et le(les) partenaires

La nouvelle variable **uni_loc** prend la valeur 1 lorsque le répondant et se(s) partenaire(s) vivent dans le même pays, 0 sinon. On utilise de nouveau l'instruction **array** pour effectuer cette comparaison avec des unions simultanées.

Dans un première temps on cherche à savoir pour chaque union si le répondant et le/la partenaire vivent une année donnée dans le même pays ($L_{[i]}=1$ ou 0). On somme ces variables dichotomiques en créant la variable (**l_s**) que l'on compare à la variable **nbu** générée dans la première partie du tutoriel. Si **l_s** est inférieur à **nbu**, alors le répondant ne vit pas dans le même pays que le/la partenaire ou au moins dans un pays différent de l'un des partenaires en cas de polygamie.

```
data analys(drop= i L_.); set analys;
array loc_conj[5] loc_conj1-loc_conj5;
array num_uni[5] num_uni1-num_uni5;
array L_[5] L_1-L_5;
do i=1 to 5;
    if loc_resp = loc_conj[i] & num_uni[i] ne . then L_[i]=1;
else if loc_resp ne loc_conj[i] & num_uni[i] ne . then L_[i]=0;
end;
l_s = sum(of L_[*]);
if l_s=nbu & nbu>0 then uni_loc=1;
if l_s<nbu & nbu>0 then uni_loc=0;
run;
```

b) Création des variables synthétiques de localisation

b1. Exemple avec un répondant avec une seule union

Pour montrer la démarche sur un cas simple sans union simultanée, on va sélectionner un seul répondant (E1).

```
data E1(keep=ident year num_union loc_resp loc_conj); set analys;
if ident="E1";
if num_uni1 ne .;
run;
```

Comptage des années d'union et des années dans le même pays

Pour créer les modalités de localisation «always unified » et « reunified », qui dépendent des valeurs prises dans le passé par **uni_loc**, on va compter le nombre d'années passées dans l'union (**cuni**) et le nombre d'années où les partenaires vivaient dans le même pays (**cuni_loc**), et comparer les valeurs de ces deux variables. Lorsque la valeur de **cuni_loc** est inférieure à celle de **cuni**, cela indique que les partenaires n'ont pas toujours vécu dans le même pays si **uni_loc** est égal à 1.

SAS n'ayant pas de fonction pour calculer directement des cumuls, on va utiliser l'instruction **retain** associée à la fonction **first** pour effectuer ce comptage. Pour compter le nombre d'années depuis le début de l'union par exemple, L'instruction **retain** indique que pour chaque année la valeur à prendre en compte pour **cuni** est la valeur de l'année précédente, implicitement dans la ligne de code: **else cuni[en t]=uni[en t]+cuni[en t-1]**. **First.ident** permet d'initialiser la première valeur de **cuni** à **ident** fixé. Dans cet exemple avec le répondant E1, **uni** prend toujours la valeur 1.

```
data E1; set E1;
uni=1;
by ident;
retain cuni;
if first.ident then cuni=uni;
                    else cuni=uni+cuni;
run;
```

On fait la même opération avec la variable **uni_loc** qui peut prendre ici la valeur 1 ou 0, selon que le répondant et le partenaire vivent ou non dans le même pays.

```
data E1; set E1;
by ident;
retain cuni_loc;
if first.ident then cuni_loc=uni_loc;
                    else cuni_loc=uni_loc+cuni_loc;
run;
```

Ici, les deux partenaires ont vécu dans le même pays en 1990 et de 1992 à 2000. Les sommes cumulées de **cuni** et **cuni_loc** ne sont donc égales que pour la première année de l'union

(**cuni=cuni_loc**). Il s'en est suivi une année de transnationalité (**uni_loc=0**) puis 9 années de réunification (**uni_loc=1** et **cuni>cuni_loc**), le couple étant de nouveau transnational de 2001 jusqu'au moment de l'enquête en 2008.

+-----+ ident year age uni_loc uni cuni cuni_loc +-----+								
1.		E1	1990	18	1	1	1	1 Always unified
2.		E1	1991	19	0	1	2	1 Transnational
3.		E1	1992	20	1	1	3	2 Reunified
4.		E1	1993	21	1	1	4	3 Reunified
5.		E1	1994	22	1	1	5	4 Reunified
6.		E1	1995	23	1	1	6	5 Reunified
7.		E1	1996	24	1	1	7	6 Reunified
8.		E1	1997	25	1	1	8	7 Reunified
9.		E1	1998	26	1	1	9	8 Reunified
10.		E1	1999	27	1	1	10	9 Reunified
11.		E1	2000	28	1	1	11	10 Reunified
12.		E1	2001	29	0	1	12	10 Transnational
13.		E1	2002	30	0	1	13	10 Transnational
14.		E1	2003	31	0	1	14	10 Transnational
15.		E1	2004	32	0	1	15	10 Transnational
16.		E1	2005	33	0	1	16	10 Transnational
17.		E1	2006	34	0	1	17	10 Transnational
18.		E1	2007	35	0	1	18	10 Transnational
19.		E1	2008	36	0	1	19	10 Transnational
+-----+								

Création de la TVC

Pour chaque année, les différentes situations sont donc les suivantes:

- Si **uni = 0**, le répondant n'est pas en couple.
- Si **uni_loc = 0**, les partenaires de l'union ne vivent pas dans le même pays. Il y a transnationalité du couple.
- Si **uni_loc =1** et **cuni>cuni_loc**, les deux partenaires vivent tous les deux dans le même pays l'année t, mais cela n'a pas été toujours le cas (nombre d'années de l'union>nombre d'années où les deux personnes vivaient dans le même pays). Il y a réunification du couple.
- Si **uni_loc =1** et **cuni=cuni_loc**, les partenaires de l'union vivent dans le même pays en t et ont toujours vécu dans le même pays jusque là.

```
data E1; set E1;
    if uni_loc=0 then trans1=1;
else if uni_loc=1 & cuni>cuni_loc then trans1=2;
else if uni_loc=1 & cuni=cuni_loc then trans1=3;
run;

proc formats;
value trans_last
1='Transnational'
2='Reunified'
3='Always unified';

proc freq data=E1;
table trans1 ident*trans1;
format trans1 trans_last.;
run ;
```

b2. Prise en compte de toutes les unions et du statut matrimonial

On va étendre le raisonnement à toutes les unions, possiblement simultanées, et aux situations où le répondant n'était pas en couple. La présence d'unions multiples oblige à établir des règles de priorité dans la codification. Par exemple, si un homme est en même temps dans une union transnationale avec une épouse et réunifiée avec une autre épouse, on considérera que son statut est « transnational ». D'autres règles de priorités pourraient être établies selon la recherche menée.

On va raisonner ici en termes de structure d'unions en créant la variable **sunid** qui identifie pour chaque répondant et chaque année la combinaison des unions en cours : par exemple, si l'union 1 et l'union 2 ont lieu en même temps, **sunid=12** ; si à un autre moment ce sont les unions 2 et 3 qui sont simultanées, alors **sunid=23**). Les comparaisons des pays de résidence se feront alors sur ces structures. De nouveau cette règle n'est pas absolue, d'autres choix peuvent être opérés.

Sans union simultanée

Si un individu n'a jamais connu de situation d'overlapping, la variable **sunid** est toujours égale à **num_union**.

ident	t	num_uni_1	num_uni_2	num_uni_3	sunid
2	t1
2	t2	1	.	.	1
2	t3	1	.	.	1
2	t3	1	.	.	1
2	t4	.	2	.	2
2	t5	.	2	.	2
2	t6	.	2	.	2
2	t7	.	2	.	2
2	t8	.	2	.	2
2	t9	.	2	.	2
2	t10
2	t11	.	.	3	3
2	t12	.	.	3	3
2	t13	.	.	3	3

Avec unions simultanées

Dans l'exemple illustré par le tableau ci-dessous on observe de l'overlapping. Les 3 unions permettent d'identifier 4 structures d'unions différentes: la structure [1] entre t2 et t3, la structure [12] entre t4 et t9, la structure [23] en t11 et t12 et la structure [2] en t10 et en t13. Si le répondant et la partenaire 2 (**num_union=2**) ont toujours vécu dans le même pays, la transnationalité de la structure [23] en t11 et t12 vient du fait que la partenaire 3 résidait dans un autre pays que le répondant. Après la rupture de cette dernière union, on retrouve la structure [2] qui est ici de nouveau considérée comme « always unified ».

ident	t	num_uni_1	num_uni_2	num_uni_3	sun_i_id	cuni	cuni_loc	trans
1	t1	alone
1	t2	1	.	.	1	1	1	always unified
1	t3	1	.	.	1	2	1	transnational
1	t3	1	.	.	1	3	2	reunified
1	t4	1	2	.	12	1	1	always unified
1	t5	1	2	.	12	2	2	always unified
1	t6	1	2	.	12	3	3	always unified
1	t7	1	2	.	12	4	4	always unified
1	t8	1	2	.	12	5	5	always unified
1	t9	1	2	.	12	6	6	always unified
1	t10	.	2	.	2	1	1	always unified
1	t11	.	2	3	23	1	0	transnational
1	t12	.	2	3	23	2	0	transnational
1	t13	.	2	.	2	2	2	always unified

Le format de la variable d'identification de la structure d'union doit être de type caractère sans quoi l'on risquerait d'avoir la même valeur de **sun_i_id** pour deux structures différentes. Tout simplement, avec les unions simultanées 1 et 2 on obtiendrait le même id que l'union unique 3 avec un format numérique. Pour créer cette variable on utilise l'instruction de concatenation || après avoir modifié le format des variables avec la fonction put.

Identifiant de la structure d'union (sun_i_id)

```
data analys(drop=x1 x2 x3 x4 x5) ; set analys;
x1=put(num_uni1, $1.);
x2=put(num_uni2, $1.);
x3=put(num_uni3, $1.);
x4=put(num_uni4, $1.);
x5=put(num_uni5, $1.);
id_uni= x1 || x2 ||x3 || x4 || x5;
run;

data analys; set analys;
sun_i_id= ident || id_uni;
run;
```

Ensuite, la TVC est construite comme en b1), à ceci près que les variables **cuni** et **cuni_loc** sont maintenant stratifiées par l'identifiant de la structure d'union **sun_i_id**. On ajoute également une modalité à la variable **trans**, celle où l'individu n'est pas en union (nbu=0). Par rapport à l'exemple avec E1, **uni** ne prend la valeur 1 seulement si le répondant est en union.

```
/** cuni **/
data analys; set analys;
if nbu>0 then uni=1;
run;

proc sort data=analys; by sun_i_id;
run;

data analys1; set analys1;
by sun_i_id;
retain cuni;
if first.sun_i_id & nbu>0 then cuni=uni;
                                else cuni=uni+cuni;
run;

/** cuni_loc **/
```



```

data analys; set analys;
by suni_id;
retain cuni_loc;
if first.suni_id then cuni_loc=uni_loc;
                        else cuni_loc=uni_loc+cuni_loc;
run;

/* variable trans */

proc format;
value transa
0= " No partner/spouse "
1= " Transnational "
2= " Reunified "
3= " Always unified " ;
run;

data analys; set analys;
if uni_loc=. then trans1=0;
else if uni_loc=0 then trans1=1;
else if uni_loc=1 & cuni>cuni_loc then trans1=2;
else if uni_loc=1 & cuni=cuni_loc then trans1=3;
format trans1 transa.;
run;

proc freq data=analys;
table trans1;
format trans1 transa.;
run;

```

trans1	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
No partner/spouse	90	26.09	90	26.09
Transnational	54	15.65	144	41.74
Reunified	57	16.52	201	58.26
Always unified	144	41.74	345	100.00

Si l'on souhaite plus de détails au niveau du statut matrimonial, on peut utiliser la variable **marital**. Ici on a juste distingué les épisodes « jamais en union », des « séparations et veuvage ».

```

proc format;
value transb
0= " Never in union "
1= " Separated or widowed "
2= " Transnational "
3= " Reunified "
4= " Always unified " ;
run;

data analys; set analys;
if trans1=0 & C=1 then trans2=0;
else if trans1=0 & (D=1 | V=1) then trans2=1;

```

```

else if trans1=1          then trans2=2;
else if trans1=2          then trans2=3;
else if trans1=3          then trans2=4;
format trans2 transb. ;
run

proc freq data=analys;
table trans2;
format trans2 transb.;
run;

```

trans2	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
Never in union	75	21.74	75	21.74
Separated or widowed	15	4.35	90	26.09
Transnational	54	15.65	144	41.74
Reunified	57	16.52	201	58.26
Always unified	144	41.74	345	100.00

Extrait de la base trans entre 1950 et 2008 (tous les 5 ans)

	ident	year	age	id_hh	uni	uni_loc	cuni	cuni_loc	trans2
5.	E1	1990	18	1....	1	1	1	1	Always unified
10.	E1	1995	23	1....	1	1	6	5	Reunified
15.	E1	2000	28	1....	1	1	11	10	Reunified
20.	E1	2005	33	1....	1	0	16	10	Transnational
23.	E1	2008	36	1....	1	0	19	10	Transnational
27.	E103	1995	17	1....	1	0	4	3	Transnational
32.	E103	2000	22	Separated or widowed
37.	E103	2005	27	.2...	1	0	5	0	Transnational
40.	E103	2008	30	Separated or widowed
43.	E147	1970	16	Never in union
48.	E147	1975	21	Never in union
53.	E147	1980	26	1....	1	1	2	2	Always unified
58.	E147	1985	31	1....	1	1	7	7	Always unified
63.	E147	1990	36	12...	1	1	1	1	Always unified
68.	E147	1995	41	12...	1	0	6	4	Transnational
73.	E147	2000	46	12...	1	0	11	4	Transnational
78.	E147	2005	51	123..	1	0	2	0	Transnational
81.	E147	2008	54	123..	1	0	5	0	Transnational
82.	F112	1965	14	Never in union
87.	F112	1970	19	1....	1	1	4	3	Reunified
92.	F112	1975	24	1....	1	1	9	8	Reunified
97.	F112	1980	29	1....	1	1	14	13	Reunified
102.	F112	1985	34	1....	1	1	19	18	Reunified
107.	F112	1990	39	1....	1	1	24	23	Reunified
112.	F112	1995	44	1....	1	1	29	28	Reunified
117.	F112	2000	49	1....	1	1	34	33	Reunified
122.	F112	2005	54	1....	1	1	39	38	Reunified
125.	F112	2008	57	1....	1	1	42	41	Reunified
130.	F121	1990	18	Never in union
135.	F121	1995	23	Never in union
140.	F121	2000	28	1....	1	0	3	0	Transnational
145.	F121	2005	33	1....	1	0	8	0	Transnational
148.	F121	2008	36	1....	1	0	11	1	Transnational
151.	F127	1980	16	Never in union
156.	F127	1985	21	1....	1	1	2	2	Always unified
161.	F127	1990	26	.2...	1	0	2	0	Transnational
166.	F127	1995	31	..3..	1	0	2	0	Transnational
171.	F127	2000	36	..3..	1	0	7	0	Transnational
176.	F127	2005	41	..3.5	1	1	4	4	Always unified
179.	F127	2008	44	..3.5	1	1	7	7	Always unified
180.	I172	1985	14	1....	1	1	1	1	Always unified
185.	I172	1990	19	1....	1	1	6	6	Always unified
190.	I172	1995	24	Separated or widowed
195.	I172	2000	29	Separated or widowed
200.	I172	2005	34	.2...	1	0	4	0	Transnational
203.	I172	2008	37	.2...	1	0	7	0	Transnational
205.	S011202	1990	15	Never in union
210.	S011202	1995	20	Never in union
215.	S011202	2000	25	1....	1	0	2	1	Transnational
220.	S011202	2005	30	1....	1	1	7	2	Reunified
223.	S011202	2008	33	1....	1	0	10	3	Transnational
224.	S079601	1950	14	Never in union
229.	S079601	1955	19	Never in union
234.	S079601	1960	24	Never in union
239.	S079601	1965	29	1....	1	1	1	1	Always unified
244.	S079601	1970	34	1....	1	1	6	6	Always unified
249.	S079601	1975	39	1....	1	1	11	11	Always unified
254.	S079601	1980	44	12...	1	1	4	4	Always unified
259.	S079601	1985	49	123..	1	1	2	2	Always unified
264.	S079601	1990	54	12...	1	1	10	10	Always unified
269.	S079601	1995	59	12...	1	1	15	15	Always unified
274.	S079601	2000	64	12...	1	1	20	20	Always unified
279.	S079601	2005	69	12...	1	1	25	25	Always unified
282.	S079601	2008	72	12...	1	1	28	28	Always unified
287.	S093701	1950	18	Never in union
292.	S093701	1955	23	Never in union
297.	S093701	1960	28	Never in union
302.	S093701	1965	33	1....	1	1	5	5	Always unified
307.	S093701	1970	38	.2...	1	0	1	0	Transnational
312.	S093701	1975	43	.234.	1	1	1	1	Always unified
317.	S093701	1980	48	.2345	1	1	3	3	Always unified

322.		S093701	1985	53	.2345	1	1	8	8	Always unified	
327.		S093701	1990	58	.2345	1	1	13	13	Always unified	
332.		S093701	1995	63	.2345	1	1	18	18	Always unified	
337.		S093701	2000	68	.2345	1	1	23	23	Always unified	
342.		S093701	2005	73	.2345	1	0	28	27	Transnational	
345.		S093701	2008	76	.2345	1	0	31	27	Transnational	
+-----+											

Extrait de la base trans entre 1985 et 1995

	ident	year	age	id_hh	uni	uni_loc	cuni	cuni_loc	trans2
1.	E1	1986	14	Never in union
2.	E1	1987	15	Never in union
3.	E1	1988	16	Never in union
4.	E1	1989	17	Never in union
5.	E1	1990	18	1....	1	1	1	1	Always unified
6.	E1	1991	19	1....	1	0	2	1	Transnational
7.	E1	1992	20	1....	1	1	3	2	Reunified
8.	E1	1993	21	1....	1	1	4	3	Reunified
9.	E1	1994	22	1....	1	1	5	4	Reunified
10.	E1	1995	23	1....	1	1	6	5	Reunified
24.	E103	1992	14	1....	1	1	1	1	Always unified
25.	E103	1993	15	1....	1	1	2	2	Always unified
26.	E103	1994	16	1....	1	1	3	3	Always unified
27.	E103	1995	17	1....	1	0	4	3	Transnational
58.	E147	1985	31	1....	1	1	7	7	Always unified
59.	E147	1986	32	1....	1	1	8	8	Always unified
60.	E147	1987	33	1....	1	1	9	9	Always unified
61.	E147	1988	34	1....	1	1	10	10	Always unified
62.	E147	1989	35	1....	1	1	11	11	Always unified
63.	E147	1990	36	12...	1	1	1	1	Always unified
64.	E147	1991	37	12...	1	1	2	2	Always unified
65.	E147	1992	38	12...	1	1	3	3	Always unified
66.	E147	1993	39	12...	1	1	4	4	Always unified
67.	E147	1994	40	12...	1	0	5	4	Transnational
68.	E147	1995	41	12...	1	0	6	4	Transnational
102.	F112	1985	34	1....	1	1	19	18	Reunified
103.	F112	1986	35	1....	1	1	20	19	Reunified
104.	F112	1987	36	1....	1	1	21	20	Reunified
105.	F112	1988	37	1....	1	1	22	21	Reunified
106.	F112	1989	38	1....	1	1	23	22	Reunified
107.	F112	1990	39	1....	1	1	24	23	Reunified
108.	F112	1991	40	1....	1	1	25	24	Reunified
109.	F112	1992	41	1....	1	1	26	25	Reunified
110.	F112	1993	42	1....	1	1	27	26	Reunified
111.	F112	1994	43	1....	1	1	28	27	Reunified
112.	F112	1995	44	1....	1	1	29	28	Reunified
126.	F121	1986	14	Never in union
127.	F121	1987	15	Never in union
128.	F121	1988	16	Never in union
129.	F121	1989	17	Never in union
130.	F121	1990	18	Never in union
131.	F121	1991	19	Never in union
132.	F121	1992	20	Never in union
133.	F121	1993	21	Never in union
134.	F121	1994	22	Never in union
135.	F121	1995	23	Never in union
156.	F127	1985	21	1....	1	1	2	2	Always unified
157.	F127	1986	22	1....	1	1	3	3	Always unified
158.	F127	1987	23	1....	1	1	4	4	Always unified
159.	F127	1988	24	1....	1	1	5	5	Always unified
160.	F127	1989	25	.2...	1	0	1	0	Transnational
161.	F127	1990	26	.2...	1	0	2	0	Transnational
162.	F127	1991	27	.2...	1	0	3	0	Transnational
163.	F127	1992	28	..34.	1	1	1	1	Always unified
164.	F127	1993	29	..34.	1	1	2	2	Always unified
165.	F127	1994	30	..3..	1	0	1	0	Transnational
166.	F127	1995	31	..3..	1	0	2	0	Transnational
180.	I172	1985	14	1....	1	1	1	1	Always unified
181.	I172	1986	15	1....	1	1	2	2	Always unified
182.	I172	1987	16	1....	1	1	3	3	Always unified
183.	I172	1988	17	1....	1	1	4	4	Always unified
184.	I172	1989	18	1....	1	1	5	5	Always unified
185.	I172	1990	19	1....	1	1	6	6	Always unified
186.	I172	1991	20	1....	1	1	7	7	Always unified
187.	I172	1992	21	1....	1	1	8	8	Always unified
188.	I172	1993	22	Separated or widowed
189.	I172	1994	23	Separated or widowed
190.	I172	1995	24	Separated or widowed
204.	S011202	1989	14	Never in union
205.	S011202	1990	15	Never in union

206.		S011202	1991	16	Never in union		
207.		S011202	1992	17	Never in union		
208.		S011202	1993	18	Never in union		
209.		S011202	1994	19	Never in union		
210.		S011202	1995	20	Never in union		
259.		S079601	1985	49	123..	1	1	2	2	Always unified	
260.		S079601	1986	50	123..	1	1	3	3	Always unified	
261.		S079601	1987	51	123..	1	1	4	4	Always unified	
262.		S079601	1988	52	12...	1	1	8	8	Always unified	
263.		S079601	1989	53	12...	1	1	9	9	Always unified	
264.		S079601	1990	54	12...	1	1	10	10	Always unified	
265.		S079601	1991	55	12...	1	1	11	11	Always unified	
266.		S079601	1992	56	12...	1	1	12	12	Always unified	
267.		S079601	1993	57	12...	1	1	13	13	Always unified	
268.		S079601	1994	58	12...	1	1	14	14	Always unified	
269.		S079601	1995	59	12...	1	1	15	15	Always unified	
322.		S093701	1985	53	.2345	1	1	8	8	Always unified	
323.		S093701	1986	54	.2345	1	1	9	9	Always unified	
324.		S093701	1987	55	.2345	1	1	10	10	Always unified	
325.		S093701	1988	56	.2345	1	1	11	11	Always unified	

326.		S093701	1989	57	.2345	1	1	12	12	Always unified	
327.		S093701	1990	58	.2345	1	1	13	13	Always unified	
328.		S093701	1991	59	.2345	1	1	14	14	Always unified	
329.		S093701	1992	60	.2345	1	1	15	15	Always unified	
330.		S093701	1993	61	.2345	1	1	16	16	Always unified	

331.		S093701	1994	62	.2345	1	1	17	17	Always unified	
332.		S093701	1995	63	.2345	1	1	18	18	Always unified	

Extrait de la base trans entre 2000 et 2008

	ident	year	id_hh	uni	uni_loc	cuni	cuni_loc	trans2
15.	E1	2000	1....	1	1	11	10	Reunified
16.	E1	2001	1....	1	0	12	10	Transnational
17.	E1	2002	1....	1	0	13	10	Transnational
18.	E1	2003	1....	1	0	14	10	Transnational
19.	E1	2004	1....	1	0	15	10	Transnational
20.	E1	2005	1....	1	0	16	10	Transnational
21.	E1	2006	1....	1	0	17	10	Transnational
22.	E1	2007	1....	1	0	18	10	Transnational
23.	E1	2008	1....	1	0	19	10	Transnational
32.	E103	2000	Separated or widowed
33.	E103	2001	.2...	1	0	1	0	Transnational
34.	E103	2002	.2...	1	0	2	0	Transnational
35.	E103	2003	.2...	1	0	3	0	Transnational
36.	E103	2004	.2...	1	0	4	0	Transnational
37.	E103	2005	.2...	1	0	5	0	Transnational
38.	E103	2006	.2...	1	0	6	0	Transnational
39.	E103	2007	Separated or widowed
40.	E103	2008	Separated or widowed
73.	E147	2000	12...	1	0	11	4	Transnational
74.	E147	2001	12...	1	0	12	4	Transnational
75.	E147	2002	12...	1	0	13	4	Transnational
76.	E147	2003	12...	1	0	14	4	Transnational
77.	E147	2004	123..	1	0	1	0	Transnational
78.	E147	2005	123..	1	0	2	0	Transnational
79.	E147	2006	123..	1	0	3	0	Transnational
80.	E147	2007	123..	1	0	4	0	Transnational
81.	E147	2008	123..	1	0	5	0	Transnational
117.	F112	2000	1....	1	1	34	33	Reunified
118.	F112	2001	1....	1	1	35	34	Reunified
119.	F112	2002	1....	1	1	36	35	Reunified
120.	F112	2003	1....	1	1	37	36	Reunified
121.	F112	2004	1....	1	1	38	37	Reunified
122.	F112	2005	1....	1	1	39	38	Reunified
123.	F112	2006	1....	1	1	40	39	Reunified
124.	F112	2007	1....	1	1	41	40	Reunified
125.	F112	2008	1....	1	1	42	41	Reunified
140.	F121	2000	1....	1	0	3	0	Transnational
141.	F121	2001	1....	1	0	4	0	Transnational
142.	F121	2002	1....	1	0	5	0	Transnational
143.	F121	2003	1....	1	0	6	0	Transnational
144.	F121	2004	1....	1	0	7	0	Transnational
145.	F121	2005	1....	1	0	8	0	Transnational
146.	F121	2006	1....	1	0	9	0	Transnational
147.	F121	2007	1....	1	1	10	1	Reunified
148.	F121	2008	1....	1	0	11	1	Transnational
171.	F127	2000	..3..	1	0	7	0	Transnational
172.	F127	2001	..3..	1	0	8	0	Transnational
173.	F127	2002	..3.5	1	1	1	1	Always unified
174.	F127	2003	..3.5	1	1	2	2	Always unified
175.	F127	2004	..3.5	1	1	3	3	Always unified
176.	F127	2005	..3.5	1	1	4	4	Always unified
177.	F127	2006	..3.5	1	1	5	5	Always unified
178.	F127	2007	..3.5	1	1	6	6	Always unified
179.	F127	2008	..3.5	1	1	7	7	Always unified
195.	I172	2000	Separated or widowed
196.	I172	2001	Separated or widowed
197.	I172	2002	.2...	1	0	1	0	Transnational
198.	I172	2003	.2...	1	0	2	0	Transnational
199.	I172	2004	.2...	1	0	3	0	Transnational
201.	I172	2006	.2...	1	0	5	0	Transnational
202.	I172	2007	.2...	1	0	6	0	Transnational
203.	I172	2008	.2...	1	0	7	0	Transnational
215.	S011202	2000	1....	1	0	2	1	Transnational
216.	S011202	2001	1....	1	0	3	1	Transnational
217.	S011202	2002	1....	1	0	4	1	Transnational
218.	S011202	2003	1....	1	0	5	1	Transnational
219.	S011202	2004	1....	1	0	6	1	Transnational
220.	S011202	2005	1....	1	1	7	2	Reunified
221.	S011202	2006	1....	1	1	8	3	Reunified
222.	S011202	2007	1....	1	0	9	3	Transnational
223.	S011202	2008	1....	1	0	10	3	Transnational

274.		S079601	2000	12...	1	1	20	20	Always unified	
275.		S079601	2001	12...	1	1	21	21	Always unified	
276.		S079601	2002	12...	1	1	22	22	Always unified	
277.		S079601	2003	12...	1	1	23	23	Always unified	
278.		S079601	2004	12...	1	1	24	24	Always unified	
279.		S079601	2005	12...	1	1	25	25	Always unified	
280.		S079601	2006	12...	1	1	26	26	Always unified	
281.		S079601	2007	12...	1	1	27	27	Always unified	
282.		S079601	2008	12...	1	1	28	28	Always unified	
337.		S093701	2000	.2345	1	1	23	23	Always unified	
338.		S093701	2001	.2345	1	1	24	24	Always unified	
339.		S093701	2002	.2345	1	1	25	25	Always unified	
340.		S093701	2003	.2345	1	1	26	26	Always unified	
341.		S093701	2004	.2345	1	1	27	27	Always unified	
342.		S093701	2005	.2345	1	0	28	27	Transnational	
343.		S093701	2006	.2345	1	0	29	27	Transnational	
344.		S093701	2007	.2345	1	0	30	27	Transnational	
345.		S093701	2008	.2345	1	0	31	27	Transnational	
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+										

ANNEXES

Création de la base d'analyse

1) Création de la base individus années *year* à partir d'un module biographique

a) Modules dont les trajectoires ne sont pas strictement continues

Chaque module biographique contient des variables communes (le sexe, l'année de naissance, niveau d'éducation....) issues de la base *general*. On va utiliser de nouveau la variable **q1a**.

Par exemple dans la base « union ».

variable name	storage type	display format	value label	variable label
ident	str7	%9s		Biographic individual identifier
num_union	byte	%9.0g		Number of the union
q1a	int	%17.0g	Q1A	Year of birth

La seule différence avec la première solution est qu'on doit avoir une seule observation (ligne) par individu. Pour les personnes qui ont déclarés plus d'une union, on ne retient que la première (**num_union=1**), en n'oubliant pas de conserver également ceux qui n'en n'ont déclaré aucune (**num_union=.**).

```
data year; set mafe.union;
if num_union=1 | num_union=.;
dur=2008-q1a+1;
run;
```

Les instructions sont ensuite strictement identiques à celles du début du tutoriel.

```
%expand(year, dur);

data year; set year;
year=q1a+j-1;
age= j-1;
run;

data year(drop= q1a j dur a2008); set year;
    label year= "year";
    label age= "age";
run;
```

b) A partir d'un module strictement continu, *house.dta* ou *activity.dta*

On va prendre la base *house*, sans utiliser la variable **q1a** (année de naissance) présente dans chaque module de l'enquête.

Toutes les années de vie du répondant jusqu'à l'année de l'enquête sont couvertes, et la valeur de **q301d** pour **num_log=1** correspond à l'année de naissance de l'individu. On procède comme dans les autres modules (*migration*, *union* et *network*) en n'oubliant pas de supprimer les doublons générés par la macro `expand`, les trajectoires étant strictement continues (**q301d(en t+1)=q301f(en t)**).

```
data year; set mafe.house;
if q301f=2008 then end_08=1;
if q301f=2008 then q601f=2007;
if q301f=. then q301f=2008;
run;

data year; set year;
dur=q301f-q301d+1;
run;

%expand(year, dur);

proc sort; by ident num_log; run;
data year; set year;
year=q301d+j-1;
age=j-1;
by ident num_log;

data year(keep=ident year age); set year;
if (q301f=year & year ne 2008 & end_08 ne 1 ) then delete;
run;
```

2) Pays de résidence du répondant à partir de la base « house »

La démarche est strictement identique à la précédente, sachant qu'on dispose du pays de résidence avec la variable **q302**. C'est en fait la solution la plus rapide car on génère directement la base *loc_resp*.

```
data loc_resp; set mafe.house;
if q301f=2008 then end_08=1;
if q301f=2008 then q601f=2007;
if q301f=. then q301f=2008;
run;

data year; set year;
dur=q301f-q301d+1;
run;

%expand(loc_resp, dur);

proc sort; by ident num_log; run;
data year; set year;
year=q301d+j-1;
by ident num_log;
run;

data loc_resp(drop=j dur end_08); set loc_resp;
if (q301f=year & year ne 2008 & end_08 ne 1 ) then delete;
run;
```