

## 4. Concatenate, Interleave and Merge

Daniel Alconchel Vázquez

**Exercise 1.** The previous code doesn't work properly. Fix it. Rename variables from table *mylib.studenci* to concatenate it with table *mylib.students* (do it in one data step).

```
LIBNAME mylib "/home/u63324691/soulutions";
```

```
data mylib.students;
  length Name$9. Surname$10. ShoeSize 3.;
  input name$ surname$ shoesize;
  datalines;
  Jan Kowalski 45
  Maja Harda 36
  Kasia Kil 39
  Honorata Mobke 41
  Kajetan Straszliwy 49
  ;
run;
```

```
data mylib.studenci;
  length Imie$15. Nazwisko$10. NrButa 3.;
  input Imie$ Nazwisko$ NrButa;
  datalines;
  Janina Fobke 49
  Kasia Zwinna 38
  Anastazja Zimmer 40
  Tomasz Nowak 43
  ;
run;
```

```
data mylib.concatenated_students2;
  set mylib.students mylib.studenci(rename=(Imie=Name Nazwisko=Surname
NrButa=ShoeSize));
run;
```

Tabla: MYLIB.CONCATENATED\_STUDENTS2 Ver: Nombres de columna

Columnas: Seleccionar todo

Nº total de filas: 9 Nº total de columnas: 3

	Name	Surname	ShoeSize
1	Jan	Kowalski	45
2	Maja	Harda	36
3	Kasia	Kil	39
4	Honorata	Mobke	41
5	Kajetan	Straszliwy	49
6	Janina	Fobke	49
7	Kasia	Zwinna	38
8	Anastazj	Zimmer	40
9	Tomasz	Nowak	43

Exercise 2. Concatenate data sets *mylib.students* and *sashelp.class*.

```
data mylib.exercise2;
    set mylib.students sashelp.class;
run;
```

Tabla: MYLIB.EXERCISE2 Ver: Nombres de columna

Columnas: Seleccionar todo

Nº total de filas: 24 Nº total de columnas: 7

	Name	Surname	ShoeSize	Sex	Age	Height
1	Jan	Kowalski	45		.	.
2	Maja	Harda	36		.	.
3	Kasia	Kil	39		.	.
4	Honorata	Mobke	41		.	.
5	Kajetan	Straszliwy	49		.	.
6	Alfred		.	M	14	69
7	Alice		.	F	13	56.5
8	Barbara		.	F	13	65.3
9	Carol		.	F	14	62.8
10	Henry		.	M	14	63.5
11	James		.	M	12	57.3
12	Jane		.	F	12	59.8

Exercise 3. Interleave the following data sets.

```

data publications;
  length Department Manager $ 10;
  input Manager $ Department $ Project $ StaffCount;
  datalines;
Cook Writing WP057 5
Deakins Writing SL827 7
Franscombe Editing MP971 4
Henry Editing WP057 3
King Production SL827 5
Krysonski Production WP057 3
Lassister Graphics SL827 3
Miedema Editing SL827 5
Morard Writing MP971 6
Posey Production MP971 4
Spackle Graphics WP057 2
;
run;

data research_development;
  length Department Manager $ 10;
  input Project $ Department $ Manager $ StaffCount;
  datalines;
MP971 Designing Daugherty 10
MP971 Coding Newton 8
MP971 Testing Miller 7
SL827 Designing Ramirez 8
SL827 Coding Cho 10
SL827 Testing Baker 7
WP057 Designing Hascal 11
WP057 Coding Constant 13
WP057 Testing Slivko 10
;
run;

```

We are going to interleave them by Department:

```

proc sort data=publications out=out_1;
by StaffCount;
run;
proc sort data=research_development out=out_2;
by StaffCount;
run;
data interleaving_exercise3;
    set out_1 out_2;
    by StaffCount;
run;

```

Tabla: WORK.INTERLEAVING\_EXERCISE3 | Ver: Nombres de columna | | Filtrar: (ninguno)

Columnas 🔍 N° total de filas: 20 N° total de columnas: 4

<input checked="" type="checkbox"/>	Seleccionar todo			
<input checked="" type="checkbox"/>	Department	1	Coding	8
<input checked="" type="checkbox"/>	Manager	2	Coding	10
<input checked="" type="checkbox"/>	Project	3	Coding	13
<input checked="" type="checkbox"/>	StaffCount	4	Designing	10
		5	Designing	8
		6	Designing	11
		7	Editing	4
		8	Editing	3
		9	Editing	5
		10	Graphics	3

**Exercise 4.** Using MERGE create data set *coworkers1* in which each observation contains the name of erasmus student from table *mylib.erasmus\_students*.

```
data mylib.erasmus_students;
  LENGTH Name$9. Surname$10. ShoeSize 3.;
  input Name$ Surname$ ShoeSize;
  datalines;
  Jean Baptiste 49
  Kate Zweerink 37
  Helga Jensen 40
  Yo Tang 49
  ;
run;
```

```
data mylib.studenci;
  Length Imie$9. Nazwisko$10. NrButa 3.;
  Input Imie$ Nazwisko$ NrButa;
  datalines;
  Anna Maria 49
  Kasia Zwinna 38
  Helena Jeszke 40
  Tomasz Tomaszke 43
  Karol Wielki 42
  ;
run;
```

```
data coworkers;
  merge mylib.erasmus_students mylib.studenci;
run;
```

Tabla: WORK.COWORKERS Ver: Nombres de columna

Columnas

Nº total de filas: 5 Nº total de columnas: 6

	Name	Surname	ShoeSize	Imie	Nazwisko	NrButa
1	Jean	Baptiste	49	Anna	Maria	49
2	Kate	Zweerink	37	Kasia	Zwinna	38
3	Helga	Jensen	40	Helena	Jeszke	40
4	Yo	Tang	49	Tomasz	Tomaszke	43
5			.	Karol	Wielki	42

**Exercise 5.** Merge table *toy* with *toy\_gender\_age* by variable Code.

```

data toy;
  length Code$4 Description$12 CompanyCode $4;
  input Code Description CompanyCode;
  datalines;
0567 Princess 1038
0568 Prince 1068
0569 Ringo 1038
0570 House 1068
0545 Train 1068
0529 Ball 1038
;
run;

```

```

data toy_gender_age;
  length Code$4 Gender$1;
  input Code$ Gender$ AgeRangeLow AgeRangeHigh;
  datalines;
0567 F 5 10
0568 M 5 10
0569 M 4 9
0570 F 3 8
0545 M 5 8
0520 F 8 12
0531 F 4 12
;
run;

```

```

proc sort data=toy;
  by Code;
run;
proc sort data=toy_gender_age;
  by Code;
run;
data exercise4;
  merge toy toy_gender_age;
  by Code;
run;

```

Nº total de filas: 7 Nº total de columnas: 6

	Code	Description	CompanyC...	Gender	AgeRangeLow	AgeRangeHigh
1	052	Ball	1038	F	8	12
2	053			F	4	12
3	054	Train	1068	M	5	8
4	056	Princess	1038	F	5	10
5	056	Prince	1068	M	5	10
6	056	Ringo	1038	M	4	9
7	057	House	1068	F	3	8

**Exercise 6.** Merge tables *toy* and *toy\_gender\_age* by Code to obtain

- only observations where Code is from set *toy*.
- only observations where Code is from set *toy\_gender\_age*.
- observations where Code is common for both sets.

```
data together_toy;
  merge toy (in=first) toy_gender_age (in=second);
  by Code;
  if first;
  PUT _ALL_;
run;
```

Tabla: WORK.TOGETHER\_TOY Ver: Nombres de columna

Columnas

Nº total de filas: 6 Nº total de columnas: 6

Filas 1-6

	Code	Description	CompanyCode	Gender	AgeRangeLow	AgeRangeHigh
1	052	Ball	1038	F	8	12
2	054	Train	1068	M	5	8
3	056	Princess	1038	F	5	10
4	056	Prince	1068	M	5	10
5	056	Ringo	1038	M	4	9
6	057	House	1068	F	3	8

```
data together_toy_gender_age;
  merge toy (in=first) toy_gender_age (in=second);
  by Code;
  if second;
  PUT _ALL_;
run;
```

Tabla: WORK.TOGETHER\_TOY\_GENDER\_AGE | Ver: Nombres de columna | Filtrar: (ninguno)

Columnas: ☒ Seleccionar todo

Nº total de filas: 7 Nº total de columnas: 6

	Code	Description	CompanyCode	Gender	AgeRangeLow	AgeRangeHigh
1	052	Ball	1038	F	8	12
2	053			F	4	12
3	054	Train	1068	M	5	8
4	056	Princess	1038	F	5	10
5	056	Prince	1068	M	5	10
6	056	Ringo	1038	M	4	9
7	057	House	1068	F	3	8

```
data together;
  merge toy (in=first) toy_gender_age (in=second);
  by Code;
  if first or second;
  PUT _ALL_;
run;
```

Tabla: WORK.TOGETHER | Ver: Nombres de columna | Filtrar: (ninguno)

Columnas: ☒ Seleccionar todo

Nº total de filas: 7 Nº total de columnas: 6

	Code	Description	CompanyCode	Gender	AgeRangeLow	AgeRangeHigh
1	052	Ball	1038	F	8	12
2	053			F	4	12
3	054	Train	1068	M	5	8
4	056	Princess	1038	F	5	10
5	056	Prince	1068	M	5	10
6	056	Ringo	1038	M	4	9
7	057	House	1068	F	3	8

**Exercise 7.** Merge table *toy* with the following data set by *CompanyCode* to obtain variables *Code*, *Description* and *Factory*.



```

data company;
  length CompanyCode$4 Factory$12;
  input CompanyCode$ Factory$;
  datalines;
1038 SuperToys
1068 HiperToys
;
run;

```

```

proc sort data=toy;
  by CompanyCode;
run;
proc sort data=company;
  by CompanyCode;
run;
data exercise7;
  merge toy company;
  by CompanyCode;
  drop CompanyCode;
run;

```

Tabla: WORK.EXERCISE7 | Ver: Nombres de columna | | Filtrar: (ninguno)

Columnas: ☒ Seleccionar todo

☒ Code ☒ Description ☒ Factory

Nº total de filas: 6 Nº total de columnas: 3

	Code	Description	Factory
1	056	Princess	SuperToys
2	056	Ringo	SuperToys
3	052	Ball	SuperToys
4	056	Prince	HiperToys
5	057	House	HiperToys
6	054	Train	HiperToys

**Exercise 8.** Read section MANY-TO-MANY MERGE in article. Merge tables toy and

```

data factory;
  length CompanyCode$4 FactoryCode$4 FactoryState$2;
  input CompanyCode$ FactoryCode$ FactoryState$;
  datalines;
1038 1234 PI
1038 1235 KU
1038 1236 HG
1038 1237 GR
1086 1238 BT
1068 1239 VR
;
run;

```

```

proc sort data=toy
  by CompanyCode;
run;
proc sort data=factory;
  by CompanyCode;
run;
data exercise8;
  merge toy factory;
  by CompanyCode;
run;

```

(I do not understand the objective of this exercise. DataSet *toy* has Code4 Description12 CompanyCode \$4 and *factory* has CompanyCode FactoryCode FactoryState\$; so the only common variable is CompanyCode). I have read the article but it is using different variables.