Day 2 - EU-SILC
Part III Practical Training Session: Panel data structure & analyses

Exercise 2 – Cumulating panel data files across years

Merge panel D-File and R-File from year blocks 2017-2020, 2016-2019, 2015-2018. Append these files. Identify rotational groups in each year block for Switzerland and Poland. Identify, which years for which rotational groups for which year-blocks have to used and appended to the year block 2017-2020 to avoid any duplicates. Calculate within person sex changes across years to estimate wrongfully matched persons.

Data sets needed
- Saved panel data 2020 from first exercise
  - l20d.dta
  - l20r.dta
- Panel data 2019
  - UDB_l19D_ver_2022_release2.dta
  - UDB_l19R_ver_2022_release2.dta
- Panel data 2018
  - UDB_l18D_ver_2022_release2.dta
  - UDB_l18R_ver_2022_release2.dta

Variables needed
- keep all variables, important here:
- D-File
  - DB010 Year of the survey
  - DB020 Country alphanumeric
  - DB030 Household ID
  - DB075 Rotational group
- R-File
  - RB010 Year of the survey
  - RB020 Country alphanumeric
  - RB030 Personal ID
  - RB040 Household ID
  - RB090 Sex

Solution steps
- Merge panel D file with panel R file from the year block 2017-2020
  - use l20d.dta and l20r.dta from above
- Merge panel D file with panel R file from the year block 2016-2019 and save
- Merge panel D file with panel R file from the year block 2015-2018 and save

- Keep only one case of respective level, go back to full data if you have change the data
- Inspect data to identify rotational group pattern across the three year blocks 2017-2020, 2016-2019, and 2015-2018
  - For sake of simplicity, only look at CH and PL
  - Follow procedure shown in presentation
  - Create a variable that contains the span of the respective rotational group, i.e. difference between the latest and the earliest year within countries and rotational groups
  - Tabulate span variable by country and rotational group for all three year blocks
  - Identify if country has only four-year rotational groups or if and when the country has six-years rotational groups
- For countries, which have only four-year rotational groups, keep the rotational group that spans all for years 2016 to 2019 in the year block 2016-2019 and the rotational group that spans all for years 2015 to 2018 in the year block 2015-2018, and append those to the year block 2017-2020
- For countries, which have also have six-years rotational groups, keep the additional year(s) which are contained in the earlier annual blocks, but not in the later annual blocks, and append those to the year block 2017-2020

Hints
- Use such a code like this to create a span variable

```
* Create a span variable that contains the difference between
* the first year and the last year of the rotational group
bysort country DB075: egen minyear=min(year)
bysort country DB075: egen maxyear=max(year)
gen span=maxyear-minyear+1
```

- Tabulate span variable for each country, e.g. here Austria

```
tab span DB075 if country=="AT"
           |          Rotation Group
      span |        1          2          3 |     Total
-----------+---------------------------------+----------
         2 |        0          0      6,815 |     6,815
         3 |        0      9,775          0 |     9,775
         4 |   12,526          0          0 |    12,526
-----------+---------------------------------+----------
     Total |   12,526      9,775      6,815 |    29,116
```

- Write down for each year block, separately by year, which rotational group span which years

AT	Rotational group	2015	2016	2017	2018	2019	2020
	3					X	X
	2				X	X	X
	1			X	X	X	X

- Do this for all year blocks