The variable \*\*`YEAR`\*\* tells you when the survey was taken, while \*\*`YRIMMIG`\*\* tells you when the person arrived. We use both to isolate the group of \*newly arrived\* immigrants.

The survey you are using is the \*\*ASEC\*\*, which is conducted in \*\*March\*\* of each year.

\* When we analyze the March \*\*2023\*\* survey file (`YEAR == 2023`), we are looking at a snapshot of the population in March 2023.

\* To find the "immigrants in an year," we look for those who arrived in the \*most recent full calendar year\*, which would be \*\*2022\*\*.

\* We can't use 2023 as the arrival year because, by March, only 2-3 months of arrivals for 2023 have even happened. This would give an incomplete count.

Therefore, the logic in the script defines "Recent Immigrants" for any given survey year as those who arrived in the \*previous\* calendar year.

Here is the exact logic from the code:

```r

NATIVITY == 5 & YRIMMIG == (YEAR - 1)

```

-----

### Example

Let's trace the logic for a person in the \*\*March 2023\*\* survey data:

1. The R script reads the data file, and for this person, the `YEAR` variable is \*\*2023\*\*.

2. The `case\_when()` function evaluates the `immigrant\_group`:

\* \*\*Is `NATIVITY == 1`?\*\* No, let's say they are foreign-born (`NATIVITY == 5`).

\* \*\*Is `NATIVITY == 5` AND `YRIMMIG == (YEAR - 1)`?\*\*

\* This translates to: Is `NATIVITY == 5` and `YRIMMIG == (2023 - 1)`?

\* In short: Is the person foreign-born \*\*and\*\* did they arrive in \*\*2022\*\*?

\* If \*\*Yes\*\*, they are classified as a \*\*"Recent Immigrant"\*\*.

\* If \*\*No\*\* (e.g., their `YRIMMIG` is 2019), they are checked by the next line, `YRIMMIG < (YEAR - 1)`, and classified as a \*\*"Prior Immigrant"\*\*.

This method ensures that for every year of the survey, you are consistently capturing the full, preceding calendar year's worth of new arrivals as your "migrant" group.