Best-ReMaP

Healthy Food for a Healthy Future

WP 5 – REFORMULATION AND PROCESSED FOOD MONITORING

T+1 Statistics methodology (task 5.5.1)

Part 1

Anses

20.10.2022

Training for statistics on T+1 data

1st part

20th October 2022

Introduction to R software + verifications to perform on the data collected before being able to produce indicators

Today

2nd part

24th November 2022

Creation of indicators on T+1 data and comparisons with pre-existing data + production of the report

Co-funded by the European Union's Health Programme (2014-2020)

Co-funded by the European Union's Health Programme (2014-2020)

1) Introduction of task 5.5.1

Co-funded by the European Union's Health Programme (2014-2020)

Reminder of the activities objectives of WP5

Activities

Task 5.4.1

- **Data collection** : pictures of products taken in shops
- **Data entry** : entry of all product information in a common template
- **Data codification :**
  - Classification of products of the template into categories and subcategories that allow homogeneous grouping of products according to recipe, ingredients, …
  - The subcategories are defined to enable to identify best formulation and room for reformulation

Task 5.5.1

- **Production of indicators :**
  - Indicators that are defined for monitoring (analysis of the food offer, nutritional values, portion size, …)
  - These indicators are produced at the subcategory level because analysis at the subcategory level have been identified as the “gold standard” during JANPA

Objectives

- **Data available to characterize the food offer and the nutritional quality of processed food at a given time** (number of products by subcategory, nutrient values, list of ingredients…)
- **Monitoring**: knowledge of the evolution over time (reformulation, addition of new products…)
- **Assessment of the impact of nutrition policy measures** both on the food offer and the composition of processed food (commitments with industry, thresholds, taxes)

Co-funded by the European Union's Health Programme (2014-2020)

Task 5.5.1 : Production of statistics for the 2nd snapshot (T+1)

Partners involved in batch 1
Austria
Belgium
Estonia
Germany
Hungary
Romania

+ Ireland
(involved in the first European snapshot T0)

→ Partners involved in the Batch 1 of the second European Snapshot (T+1) will have **to produce statistics/indicators** by Best-ReMaP categories and subcategories of products collected during the T+1

→ **Comparisons with pre-existing data** will be made in order to establish trend assessments of the nutritional quality of the processed food

→ Each partners will have **to produce a report** that will be included in the WP5 deliverable (D5.3)

Co-funded by the European Union's Health Programme (2014-2020)

Timeline of the task 5.5.1

2022												2023					
j	f	m	a	m	j	j	a	s	o	n	d	j	f	m	a	m	j
M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33

Task 5.4.1 : Data collection and data treatment (data entry and encoding)
Mar 2022 - Feb 2023

Collection of pictures **Mar 2022**

Data treatment (data entry and encoding)
Apr 2022 - Jan 2023
(February 2023 at the latest)

Soft drinks | Breakfast cereals | Bread products | Fresh dairy products and desserts | Delicatessen meats and similar

Task 5.5.1 : Production of statistics
Nov 2022 - Apr 2023 (6 months)

Verifications steps **Nov - Jan 2022** *(3 months)*

Production of indicators **Feb 2023** *(1 month)*

Production of the report **Mar - Apr 2023** *(2 months)*

1st Nov 2022

Please send us feedback to confirm that the R and Rstudio software are installed
If you have any problems before this date, please let us know so that we can help you

These deadlines are indications so that you can organize yourselves as well as possible

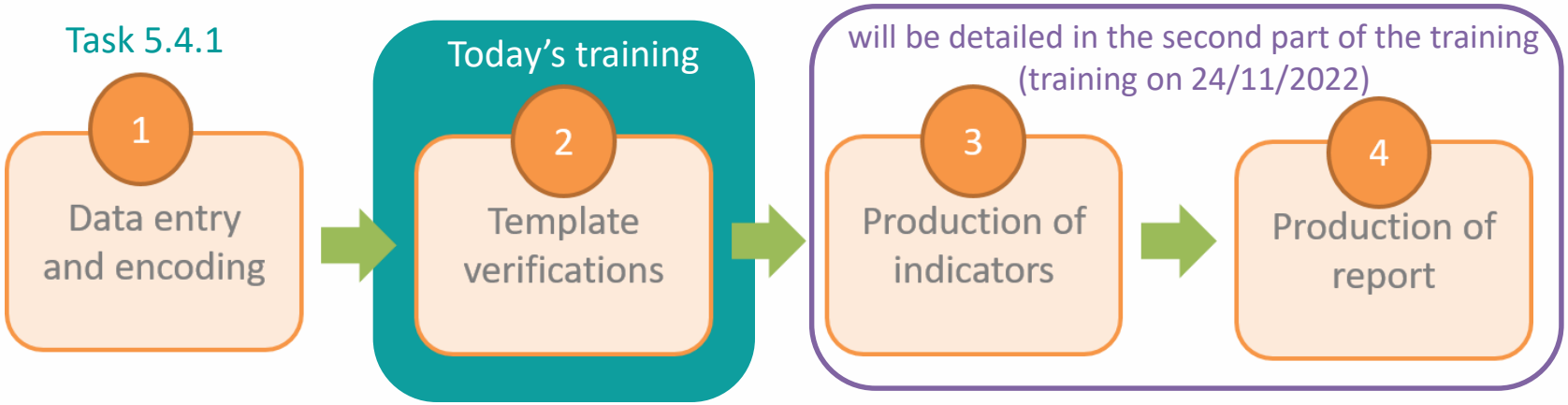
End of verifications and corrections to be made

You should have all figures and tables ready and stored in the "indicators" file

Report ready to be sent to WP5

Co-funded by the European Union's Health Programme (2014-2020)

Summary of the steps in task 5.5.1



It is important to finish each step before moving on to the next

- **Steps 2 and 3** will be done by data processing programs on the Rstudio software which is a free software and therefore accessible to all. The programs have already been created and written in order to harmonise the work and ensure that all partners work in the same way. You will just need to run the programs on your data (you will not have to create any programs).

- **Step 3 and 4** will not be detailed in this training but in the 2nd part of the training on **24th November 2022**

- **Step 4** : you will receive all the elements so that everyone can produce a report on the analysis of his data. A sample report will be provided to ensure that the results for each country are presented in a harmonised way. You will have more information about the report during the 2nd part of the training.

Co-funded by the European Union's Health Programme (2014-2020)

Equipment needed

Tools you already have

- **T+1 collection template** : template that you filled in according to the WP5 methodology during your T+1 data collection and with data for the 5 priority food categories.
- **Pictures of the products** that you have collected for your T+1 data collection
- **Pre-existing data template** : template with your pre-existing national data that have been classified in the Best-ReMaP nomenclature

Tools you will have to download (explanation pages 40-47)

- **R**, **Rstudio** : Free statistical software that you will use to check and correct the data entered in your template and to perform indicators/statistics on your data

- **WP5 R programs** : R programs created by Anses that you will just have to run on the Rstudio software (you will be guided at each step, you will not have to create programs or develop code).
- **Support files for programs** (*wp5_verification_file*; *Best-ReMaP_nomenclature*; *Subcategory_order*; *Years of interest*) **:** files that you will have to download and as it will be used in the R program. These files do **not have to be filled in or modified** (except the file 'Years of interest').

The **WP5 R programs** and **the support files** are available in a **zip folder** on the project intranet by following this link:
https://portal.nijz.si/ssf/a/c/p_name/ss_forum/p_action/1/binderId/21932/entityType/folder/action/view_permalink/novl_url/1

You will need to copy this **folder** as it is on the desktop of your computer (this action is detailed in the slide 32-33)

2) Installation of the necessary equipment and presentation of the Rstudio software

Co-funded by the European Union's Health Programme (2014-2020)

2) Installation of the necessary equipment and presentation of the Rstudio software

A. Preliminary steps

B. Installation of software

C. Introduction to R studio

D. Cleaning of the Rstudio interface

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

1st preliminary step : preparation of the T+1 collection template

→ You must ensure that your T+1 collection template is a **single file** with a **single tab** for all data collected during T+1 (the 5 food categories in the same tab).

→ As your T+1 collection template contains many rows and drop-down menus, you will have to copy it into a **new .xlsx excel file** to keep only the filled rows and remove the drop-down menus.

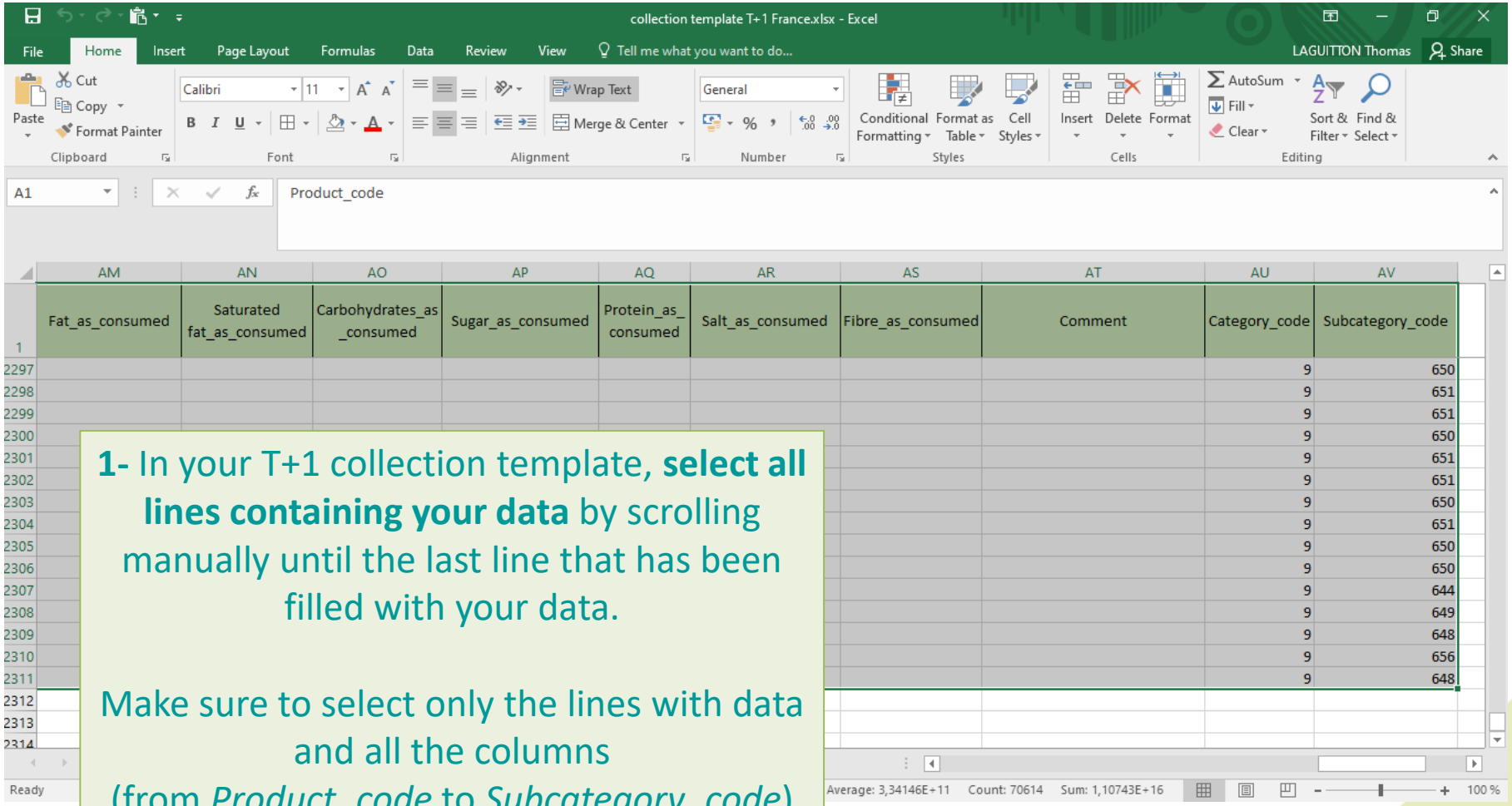
see the following slides for a step-by-step explanation of this procedure

It is also possible for you to perform the verification steps category by category by working on 5 different files (1 per Best-ReMaP category).

Please note that this will result in more file handling and that you will need to gather all your T+1 data in one file at the end of your verification steps. The program to create the indicators will have to be run on the compiled file of all verified categories (see training of 24 November).

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

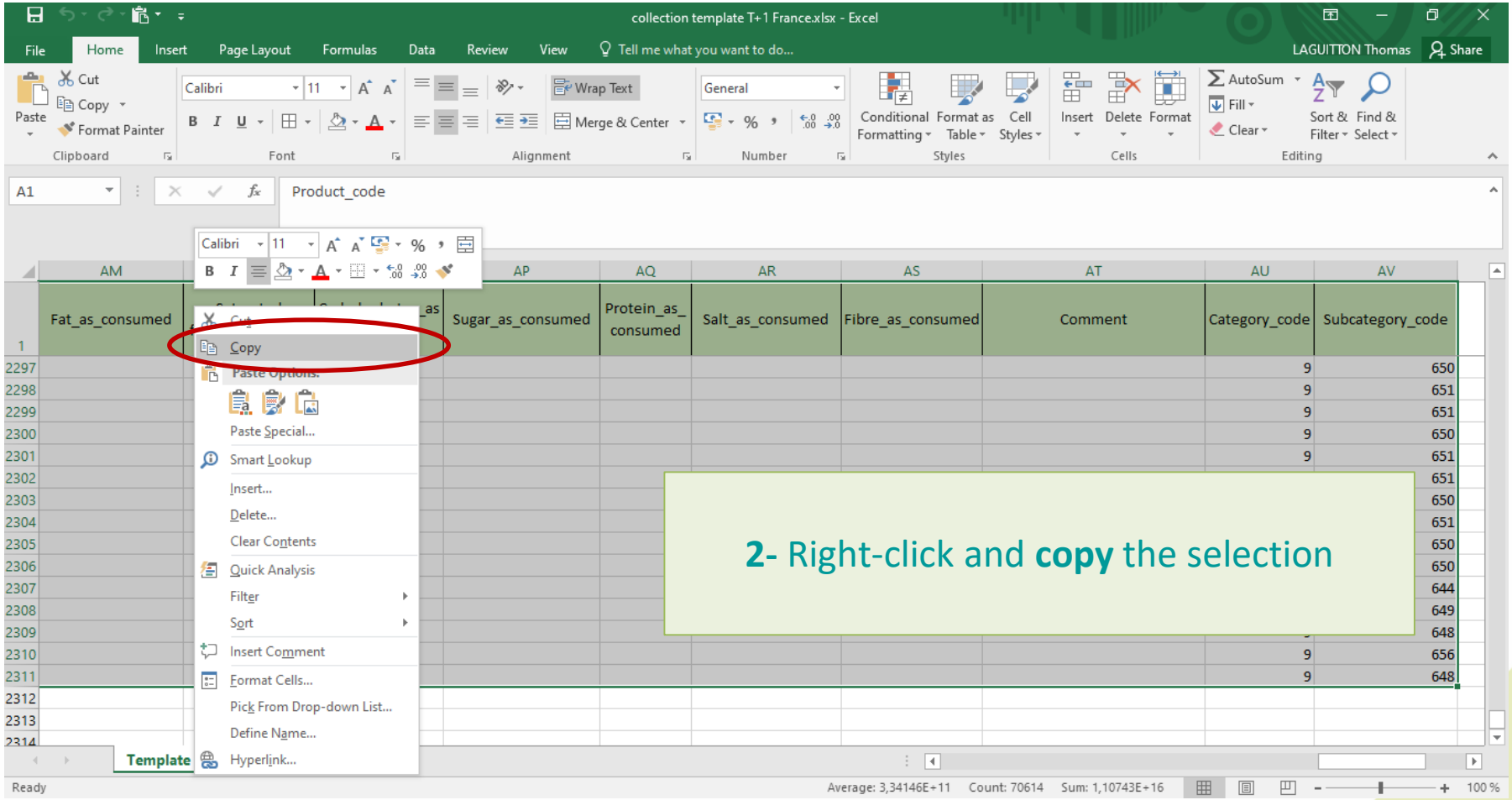


1- In your T+1 collection template, **select all lines containing your data** by scrolling manually until the last line that has been filled with your data.

Make sure to select only the lines with data and all the columns
(from *Product_code* to *Subcategory_code*)

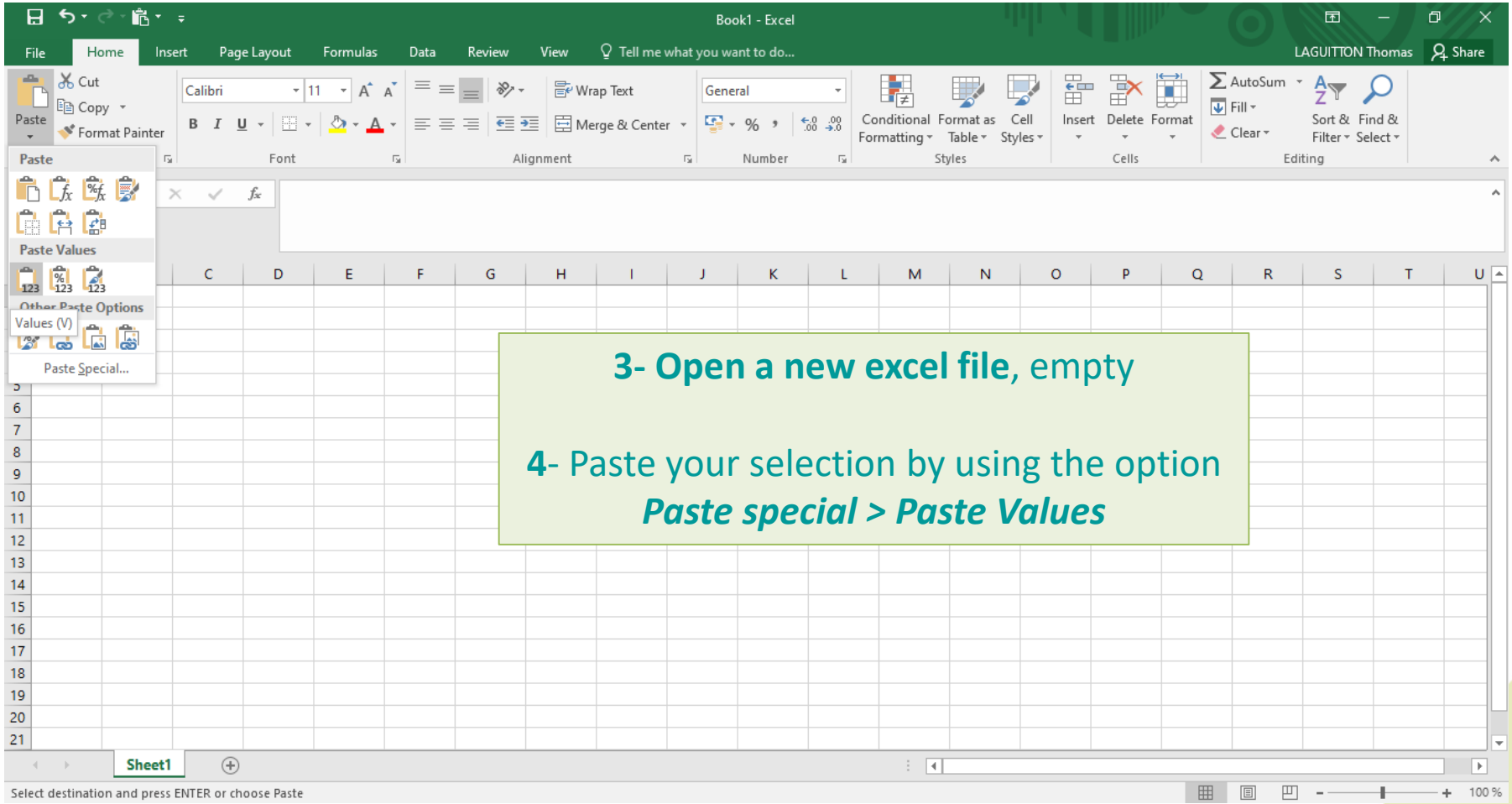
Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps



Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps



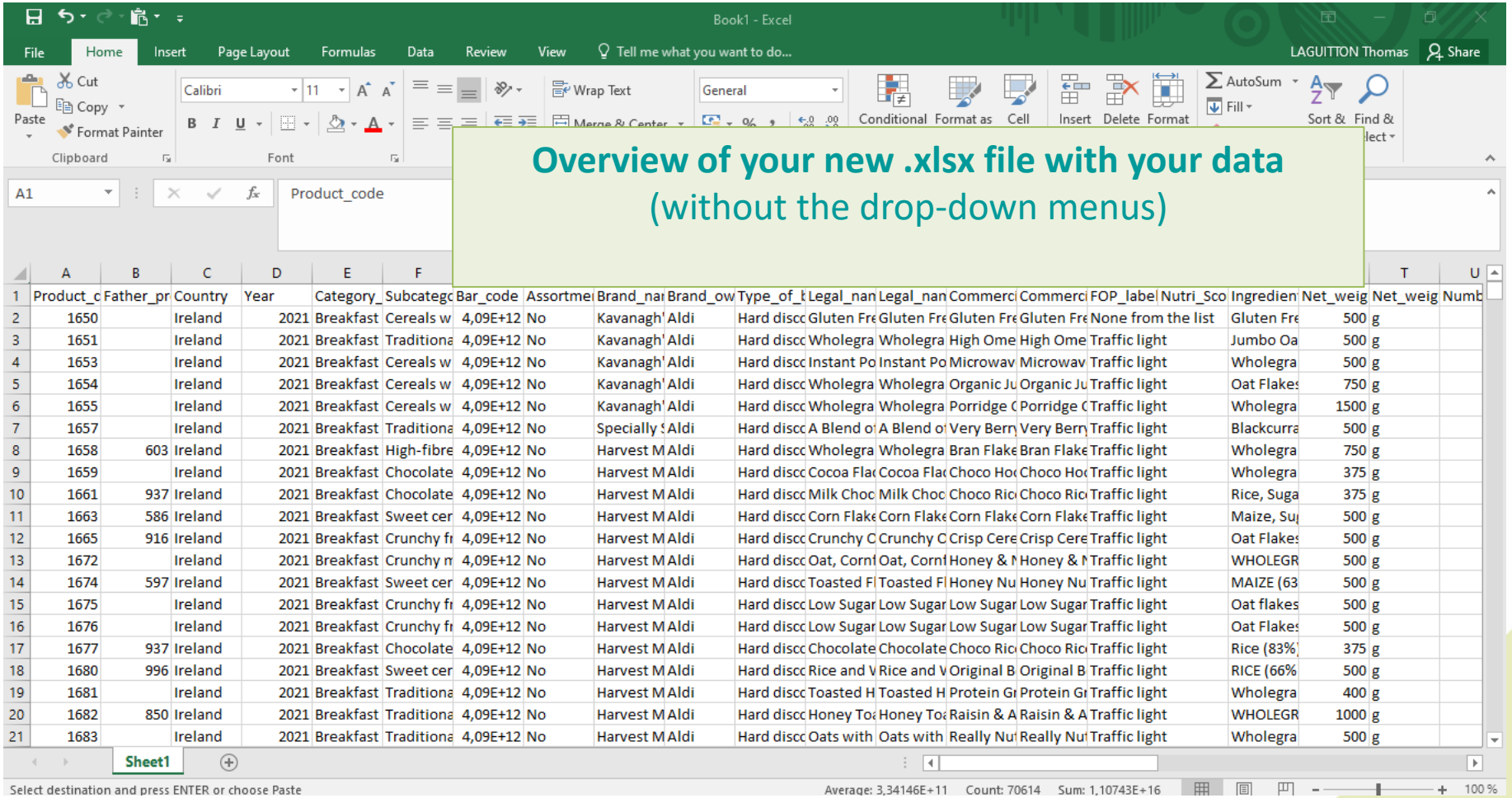
3- Open a new excel file, empty

4- Paste your selection by using the option ***Paste special > Paste Values***

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

Overview of your new .xlsx file with your data
(without the drop-down menus)



	A	B	C	D	E	F	T	U
1	Product_c	Father_pr	Country	Year	Category_	Subcategc	Net_weig	Net_weig	Numb
2	1650		Ireland	2021	Breakfast	Cereals w	500	g	
3	1651		Ireland	2021	Breakfast	Traditiona	500	g	
4	1653		Ireland	2021	Breakfast	Cereals w	500	g	
5	1654		Ireland	2021	Breakfast	Cereals w	750	g	
6	1655		Ireland	2021	Breakfast	Cereals w	1500	g	
7	1657		Ireland	2021	Breakfast	Traditiona	500	g	
8	1658	603	Ireland	2021	Breakfast	High-fibre	750	g	
9	1659		Ireland	2021	Breakfast	Chocolate	375	g	
10	1661	937	Ireland	2021	Breakfast	Chocolate	375	g	
11	1663	586	Ireland	2021	Breakfast	Sweet cer	500	g	
12	1665	916	Ireland	2021	Breakfast	Crunchy fr	500	g	
13	1672		Ireland	2021	Breakfast	Crunchy m	500	g	
14	1674	597	Ireland	2021	Breakfast	Sweet cer	500	g	
15	1675		Ireland	2021	Breakfast	Crunchy fr	500	g	
16	1676		Ireland	2021	Breakfast	Crunchy fr	500	g	
17	1677	937	Ireland	2021	Breakfast	Chocolate	375	g	
18	1680	996	Ireland	2021	Breakfast	Sweet cer	500	g	
19	1681		Ireland	2021	Breakfast	Traditiona	400	g	
20	1682	850	Ireland	2021	Breakfast	Traditiona	1000	g	
21	1683		Ireland	2021	Breakfast	Traditiona	500	g	

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

1st preliminary step : preparation of the T+1 collection template

→ Now you have an **.xlsx file** containing only the lines with your data and without the drop-down menus.
You can rename it **T+1 data collection *country*.xlsx** (with the name of your own country) for example.

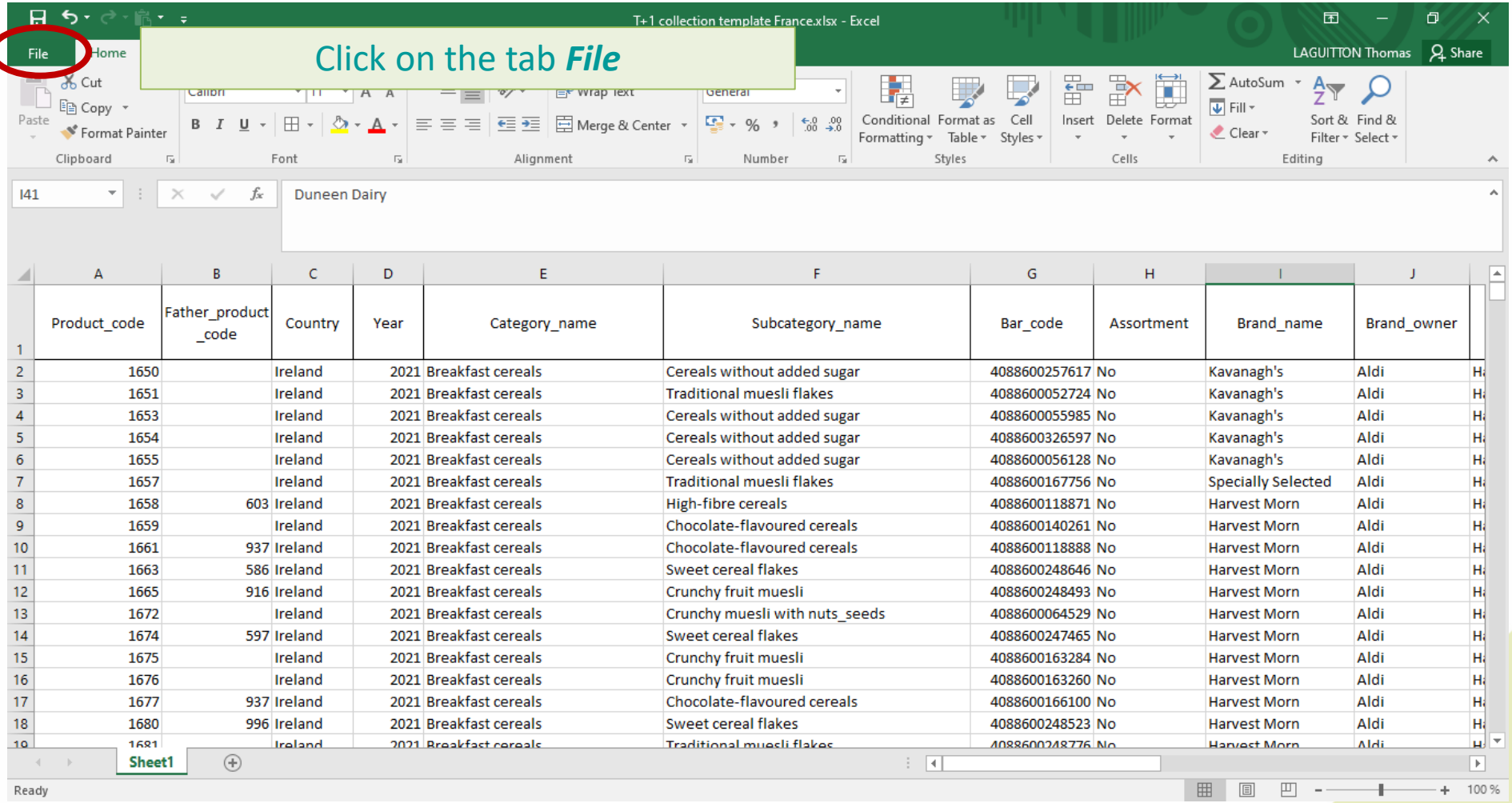
→ You must **make a copy** of this file (T+1 collection template *country*.xlsx) and save it in **.csv (comma separator) format** under the name T+1_data_collection_***country***.csv (with the name of your own country)

The creation of this file in .csv format is very important because it is this file that will be used in the R software for the verification and indicator creation stages.

see the following slides for a step-by-step explanation of this procedure

Co-funded by the European Union's Health Programme (2014-2020)

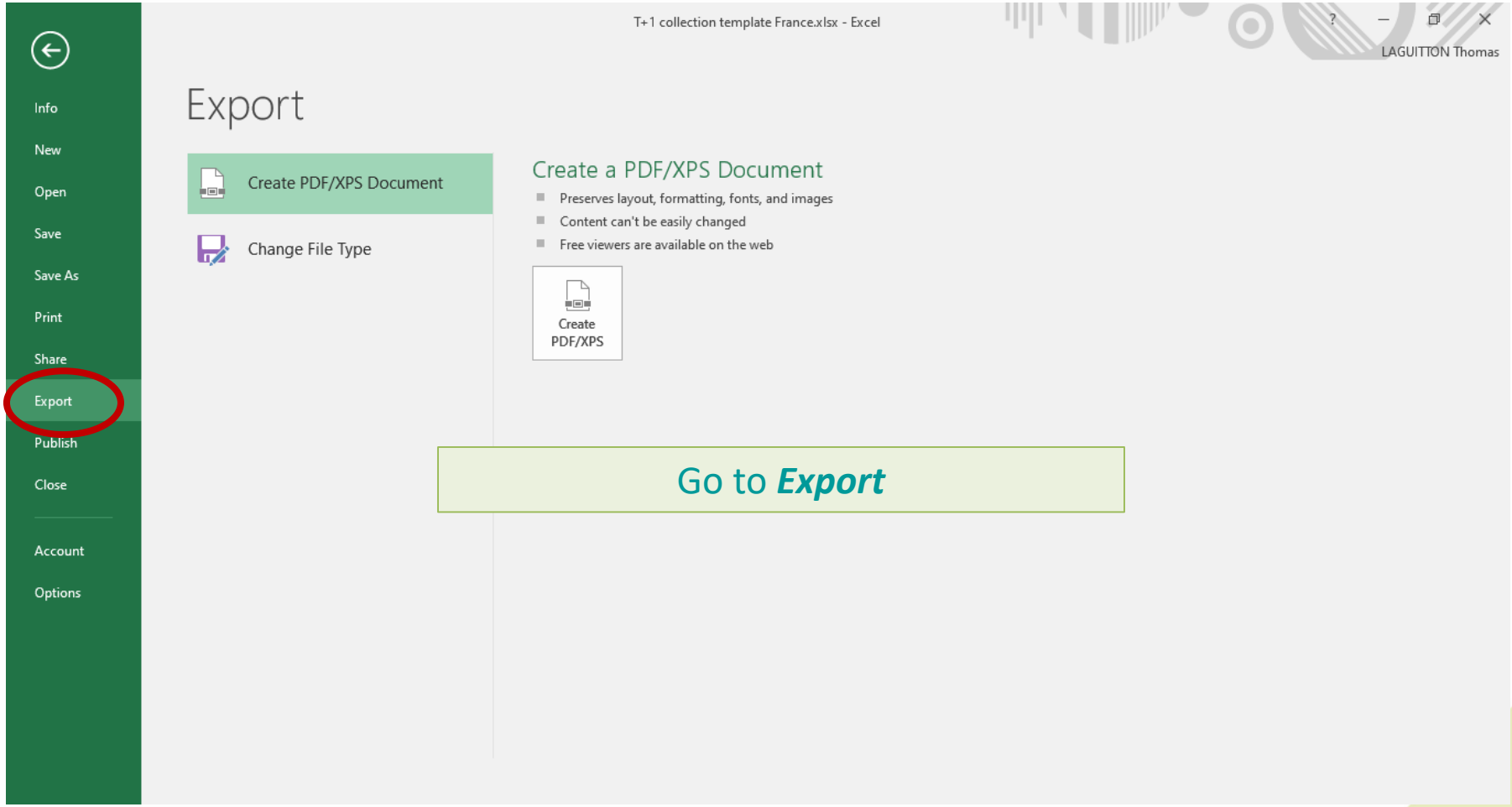
Preliminary steps



	A	B	C	D	E	F	G	H	I	J
1	Product_code	Father_product_code	Country	Year	Category_name	Subcategory_name	Bar_code	Assortment	Brand_name	Brand_owner
2	1650		Ireland	2021	Breakfast cereals	Cereals without added sugar	4088600257617	No	Kavanagh's	Aldi
3	1651		Ireland	2021	Breakfast cereals	Traditional muesli flakes	4088600052724	No	Kavanagh's	Aldi
4	1653		Ireland	2021	Breakfast cereals	Cereals without added sugar	4088600055985	No	Kavanagh's	Aldi
5	1654		Ireland	2021	Breakfast cereals	Cereals without added sugar	4088600326597	No	Kavanagh's	Aldi
6	1655		Ireland	2021	Breakfast cereals	Cereals without added sugar	4088600056128	No	Kavanagh's	Aldi
7	1657		Ireland	2021	Breakfast cereals	Traditional muesli flakes	4088600167756	No	Specially Selected	Aldi
8	1658	603	Ireland	2021	Breakfast cereals	High-fibre cereals	4088600118871	No	Harvest Morn	Aldi
9	1659		Ireland	2021	Breakfast cereals	Chocolate-flavoured cereals	4088600140261	No	Harvest Morn	Aldi
10	1661	937	Ireland	2021	Breakfast cereals	Chocolate-flavoured cereals	4088600118888	No	Harvest Morn	Aldi
11	1663	586	Ireland	2021	Breakfast cereals	Sweet cereal flakes	4088600248646	No	Harvest Morn	Aldi
12	1665	916	Ireland	2021	Breakfast cereals	Crunchy fruit muesli	4088600248493	No	Harvest Morn	Aldi
13	1672		Ireland	2021	Breakfast cereals	Crunchy muesli with nuts_seeds	4088600064529	No	Harvest Morn	Aldi
14	1674	597	Ireland	2021	Breakfast cereals	Sweet cereal flakes	4088600247465	No	Harvest Morn	Aldi
15	1675		Ireland	2021	Breakfast cereals	Crunchy fruit muesli	4088600163284	No	Harvest Morn	Aldi
16	1676		Ireland	2021	Breakfast cereals	Crunchy fruit muesli	4088600163260	No	Harvest Morn	Aldi
17	1677	937	Ireland	2021	Breakfast cereals	Chocolate-flavoured cereals	4088600166100	No	Harvest Morn	Aldi
18	1680	996	Ireland	2021	Breakfast cereals	Sweet cereal flakes	4088600248523	No	Harvest Morn	Aldi
19	1681		Ireland	2021	Breakfast cereals	Traditional muesli flakes	4088600248776	No	Harvest Morn	Aldi

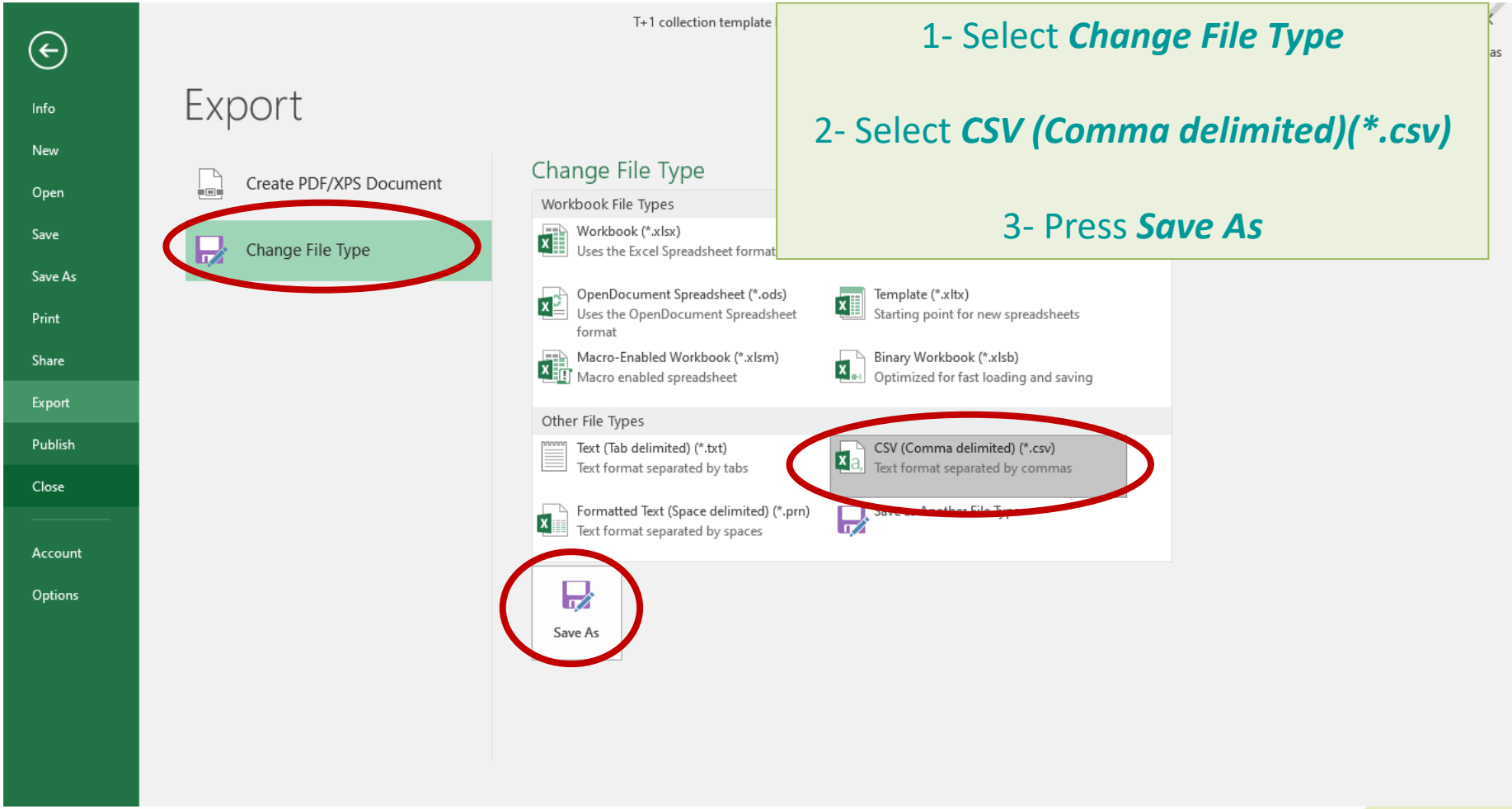
Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps



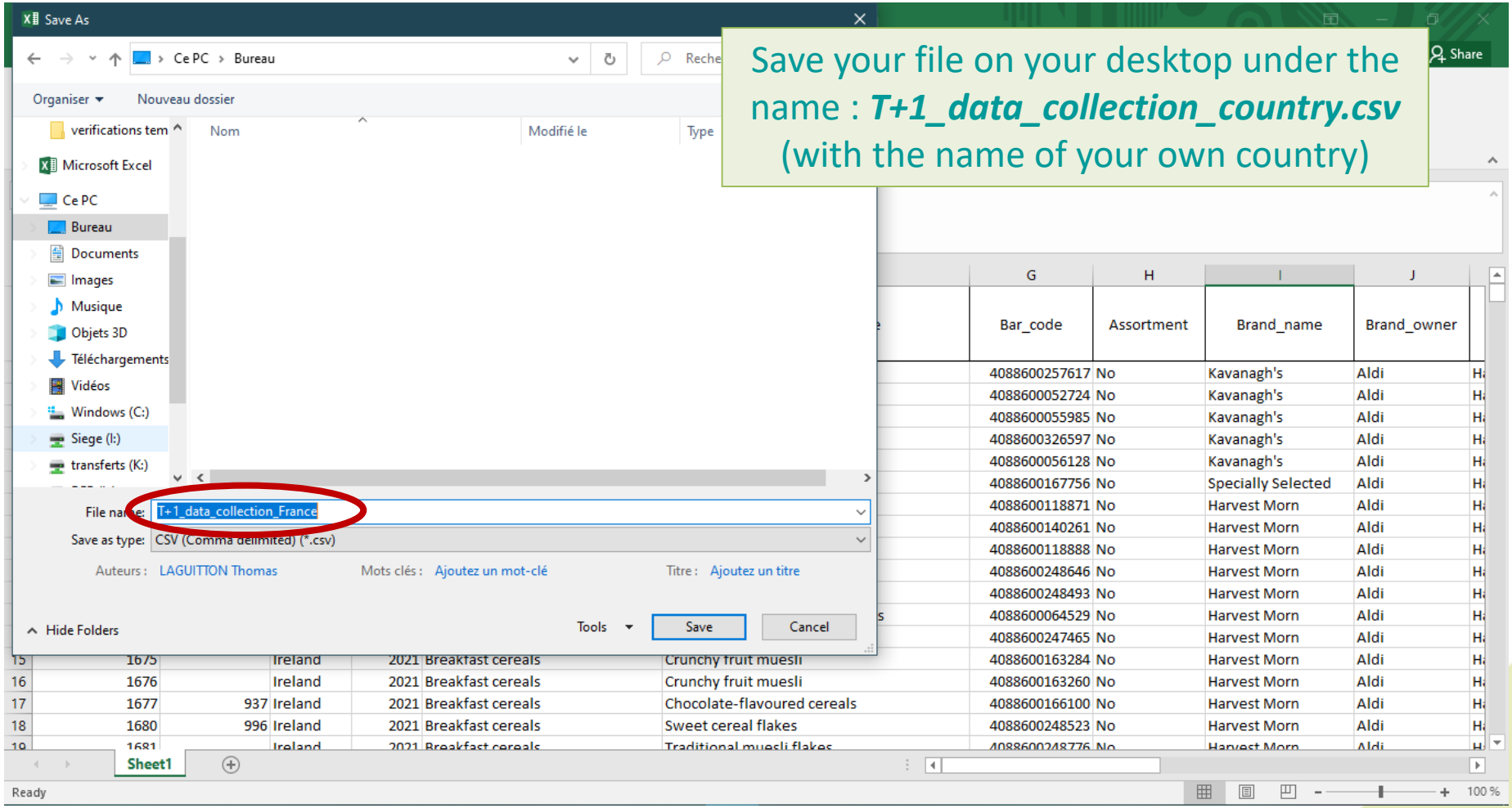
Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps



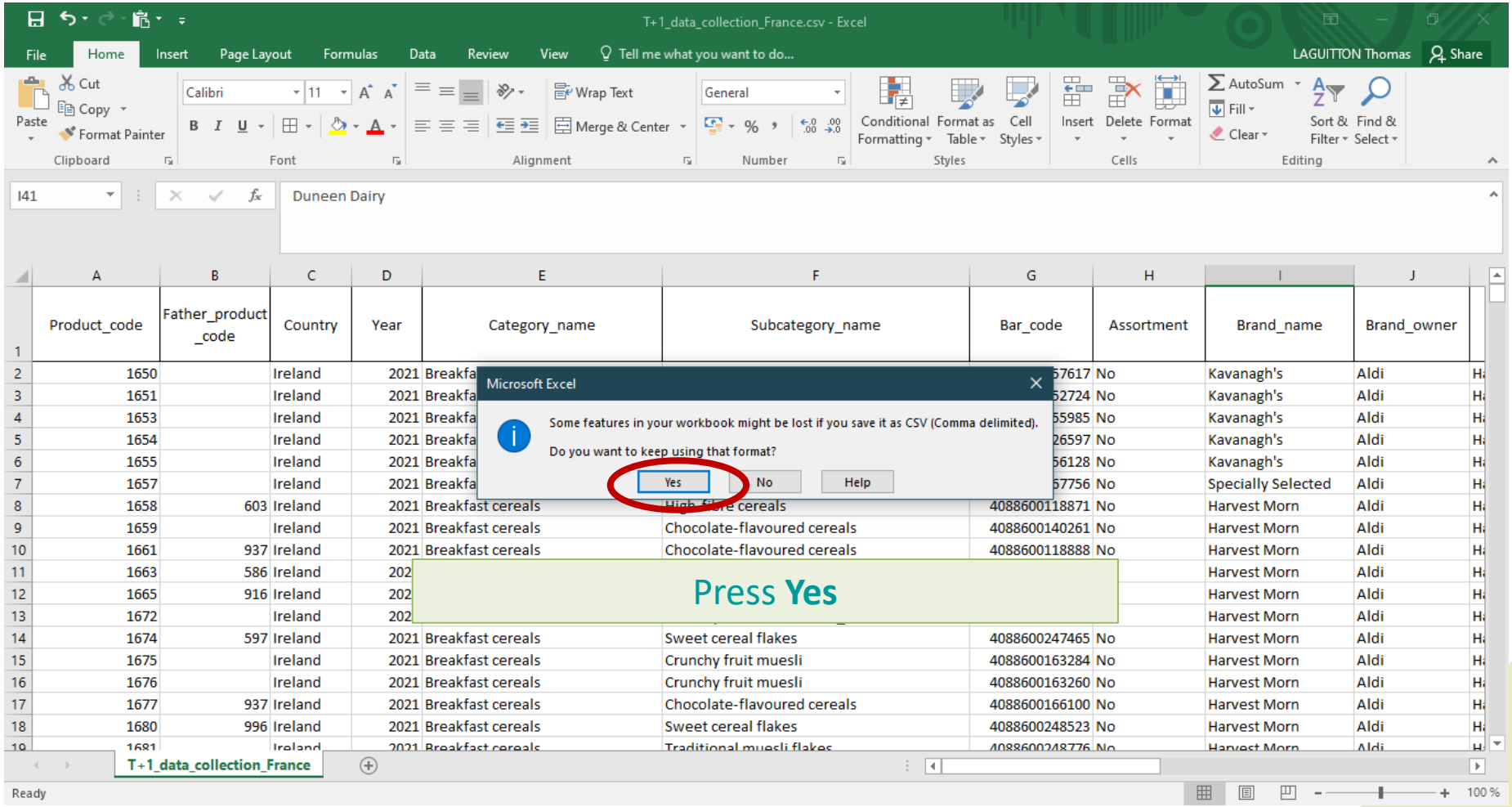
Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps



Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps



Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

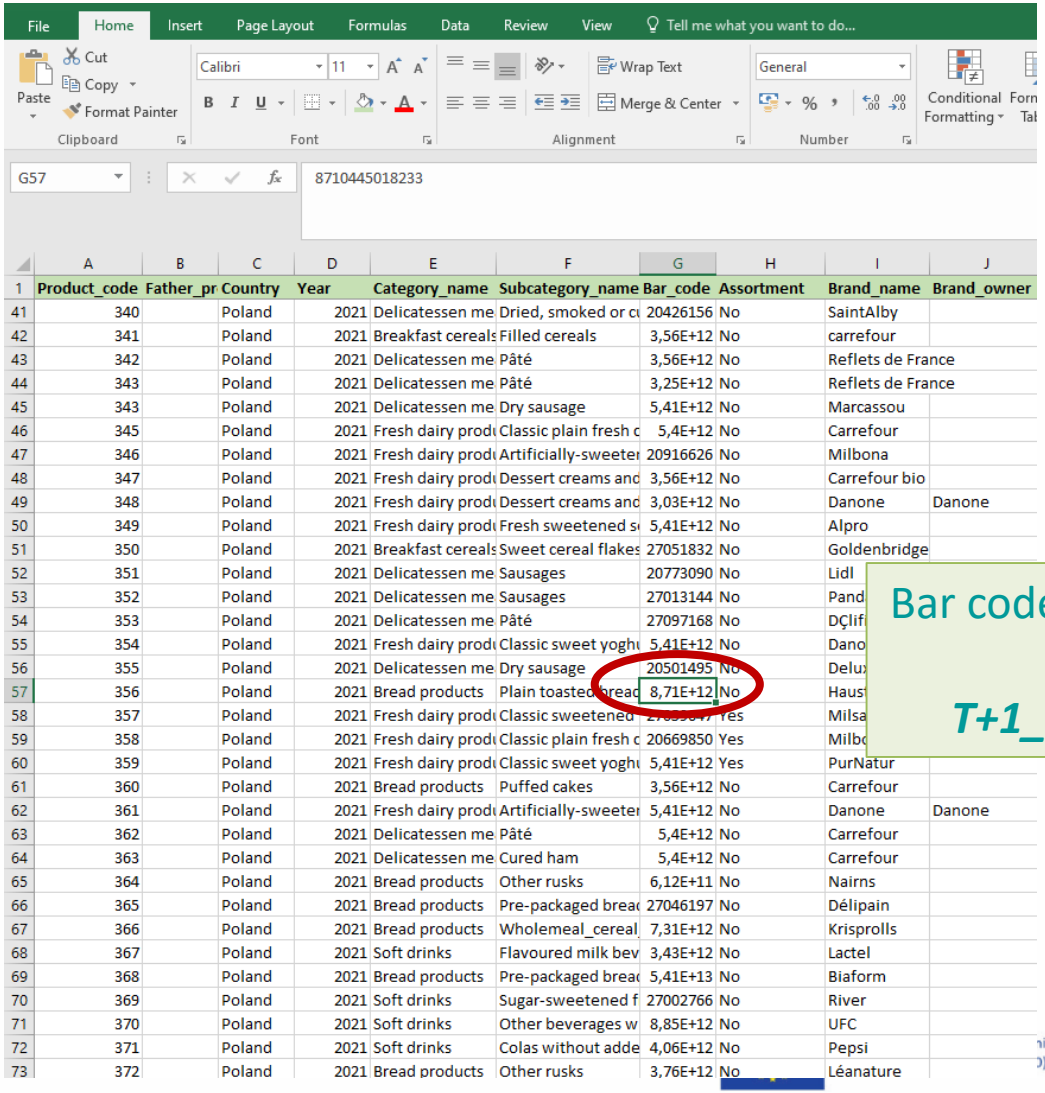
When you save your ***.xlsx*** file in **.csv format**, the barcodes in the **.csv file** appear in scientific writing (e.g. 1.89E+12).
It is necessary to select the column 'bar_codes' and change the column format to **'Number'** with **0 decimal digit**.
The bar codes will appear in full and you will not lose any information. You can then save this new version of your template with the full barcodes and overwrite the old one.

Be careful, as soon as you reopen this new **.csv file**, the barcodes will be written scientifically again and you will have to repeat this procedure.
We therefore invite you to do this procedure only once when you save the **.csv file** and not to reopen the saved file.
(To be read in R, the file must not be opened so you do not need to re-open it)

see the following slides for a step-by-step explanation of this procedure

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

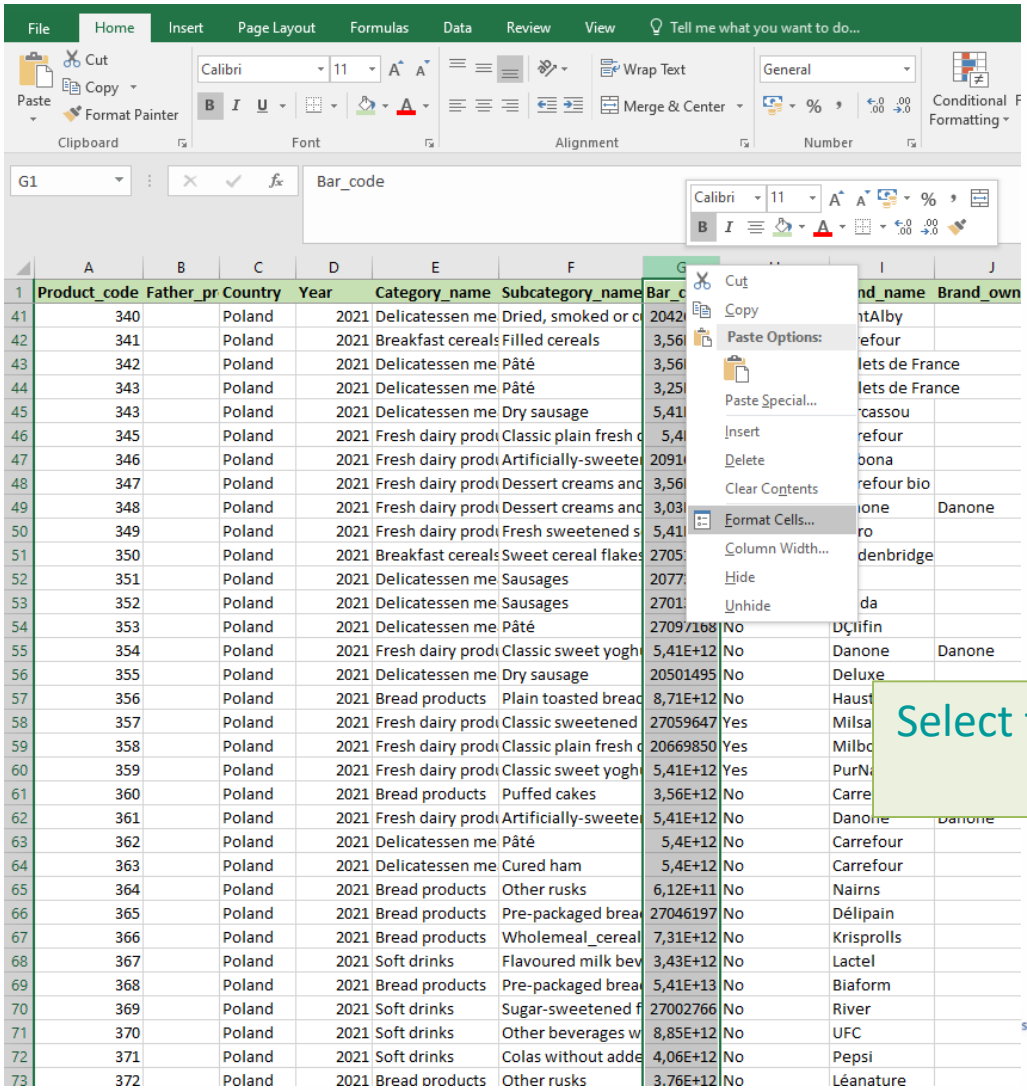


	A	B	C	D	E	F	G	H	I	J
1	Product_code	Father_pr	Country	Year	Category_name	Subcategory_name	Bar_code	Assortment	Brand_name	Brand_owner
41	340		Poland	2021	Delicatessen me	Dried, smoked or c	20426156	No	SaintAlby	
42	341		Poland	2021	Breakfast cereals	Filled cereals	3,56E+12	No	carrefour	
43	342		Poland	2021	Delicatessen me	Pâté	3,56E+12	No	Reflets de France	
44	343		Poland	2021	Delicatessen me	Pâté	3,25E+12	No	Reflets de France	
45	343		Poland	2021	Delicatessen me	Dry sausage	5,41E+12	No	Marcassou	
46	345		Poland	2021	Fresh dairy prod	Classic plain fresh c	5,4E+12	No	Carrefour	
47	346		Poland	2021	Fresh dairy prod	Artificially-sweeter	20916626	No	Milbona	
48	347		Poland	2021	Fresh dairy prod	Dessert creams and	3,56E+12	No	Carrefour bio	
49	348		Poland	2021	Fresh dairy prod	Dessert creams and	3,03E+12	No	Danone	Danone
50	349		Poland	2021	Fresh dairy prod	Fresh sweetened s	5,41E+12	No	Alpro	
51	350		Poland	2021	Breakfast cereals	Sweet cereal flakes	27051832	No	Goldenbridge	
52	351		Poland	2021	Delicatessen me	Sausages	20773090	No	Lidl	
53	352		Poland	2021	Delicatessen me	Sausages	27013144	No	Pand	
54	353		Poland	2021	Delicatessen me	Pâté	27097168	No	DÇlif	
55	354		Poland	2021	Fresh dairy prod	Classic sweet yoghu	5,41E+12	No	Dano	
56	355		Poland	2021	Delicatessen me	Dry sausage	20501495	No	Delu	
57	356		Poland	2021	Bread products	Plain toasted bread	8,71E+12	No	Haus	
58	357		Poland	2021	Fresh dairy prod	Classic sweetened	[illegible]	Yes	Milsa	
59	358		Poland	2021	Fresh dairy prod	Classic plain fresh c	20669850	Yes	Milbo	
60	359		Poland	2021	Fresh dairy prod	Classic sweet yoghu	5,41E+12	Yes	PurNatur	
61	360		Poland	2021	Bread products	Puffed cakes	3,56E+12	No	Carrefour	
62	361		Poland	2021	Fresh dairy prod	Artificially-sweeter	5,41E+12	No	Danone	Danone
63	362		Poland	2021	Delicatessen me	Pâté	5,4E+12	No	Carrefour	
64	363		Poland	2021	Delicatessen me	Cured ham	5,4E+12	No	Carrefour	
65	364		Poland	2021	Bread products	Other rusks	6,12E+11	No	Nairns	
66	365		Poland	2021	Bread products	Pre-packaged brea	27046197	No	Délipain	
67	366		Poland	2021	Bread products	Wholemeal_cereal	7,31E+12	No	Krisprolls	
68	367		Poland	2021	Soft drinks	Flavoured milk bev	3,43E+12	No	Lactel	
69	368		Poland	2021	Bread products	Pre-packaged brea	5,41E+13	No	Biaform	
70	369		Poland	2021	Soft drinks	Sugar-sweetened f	27002766	No	River	
71	370		Poland	2021	Soft drinks	Other beverages w	8,85E+12	No	UFC	
72	371		Poland	2021	Soft drinks	Colas without adde	4,06E+12	No	Pepsi	
73	372		Poland	2021	Bread products	Other rusks	3,76E+12	No	Léanature	

Bar codes appearing in scientific format in your file

T+1_data_collection_country.csv

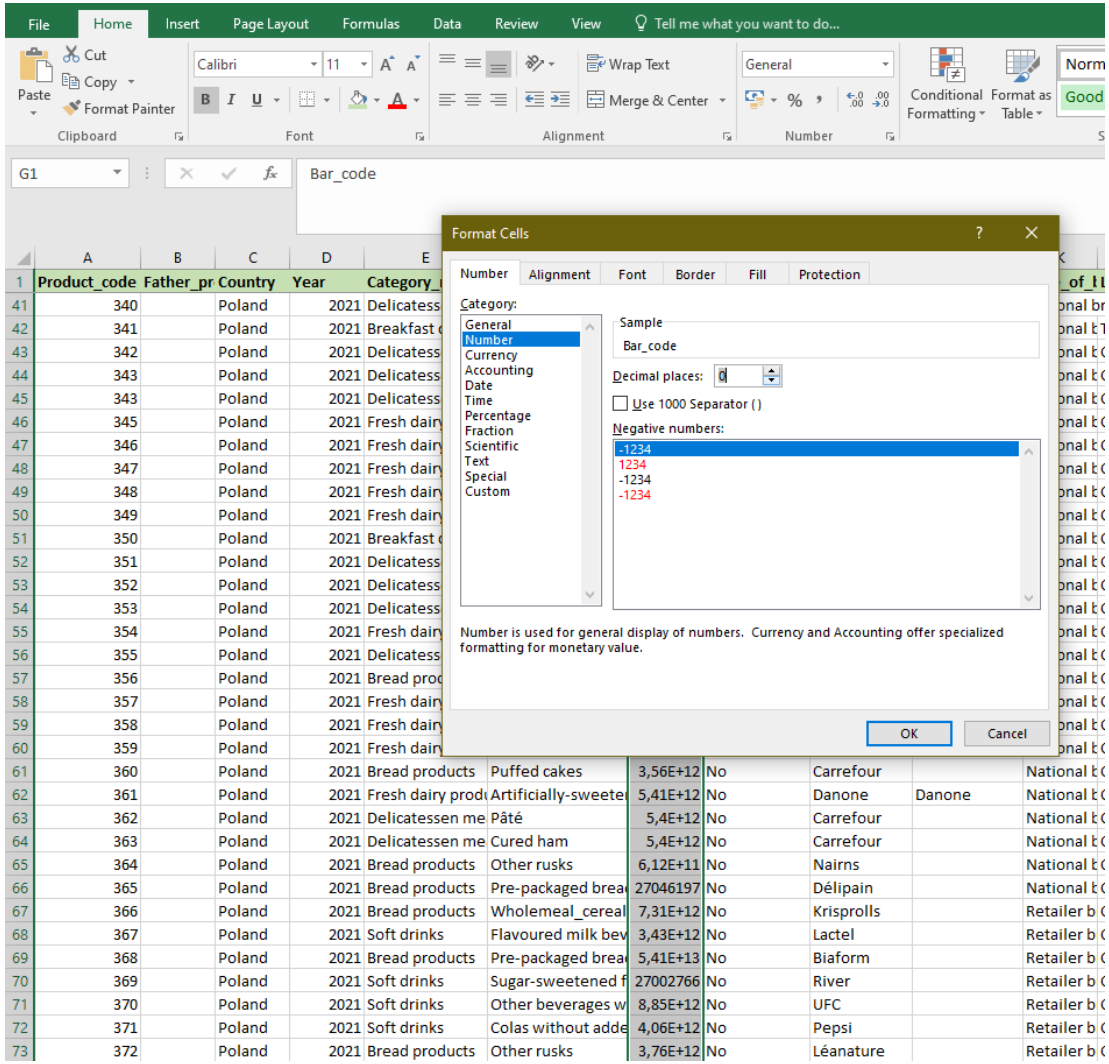
Preliminary steps



	A	B	C	D	E	F	G	H	I	J
1	Product_code	Father_pr	Country	Year	Category_name	Subcategory_name	Bar_c		nd_name	Brand_own
41	340		Poland	2021	Delicatessen me	Dried, smoked or c	2042		ntAlby	
42	341		Poland	2021	Breakfast cereals	Filled cereals	3,56		efour	
43	342		Poland	2021	Delicatessen me	Pâté	3,56		lets de France	
44	343		Poland	2021	Delicatessen me	Pâté	3,25		lets de France	
45	343		Poland	2021	Delicatessen me	Dry sausage	5,41		cassou	
46	345		Poland	2021	Fresh dairy prod	Classic plain fresh c	5,4		efour	
47	346		Poland	2021	Fresh dairy prod	Artificially-sweete	2091		bona	
48	347		Poland	2021	Fresh dairy prod	Dessert creams and	3,56		efour bio	
49	348		Poland	2021	Fresh dairy prod	Dessert creams and	3,03		one	Danone
50	349		Poland	2021	Fresh dairy prod	Fresh sweetened s	5,41		ro	
51	350		Poland	2021	Breakfast cereals	Sweet cereal flakes	2705		denbridge	
52	351		Poland	2021	Delicatessen me	Sausages	2077			
53	352		Poland	2021	Delicatessen me	Sausages	2701		da	
54	353		Poland	2021	Delicatessen me	Pâté	27097168	No	DÇlifin	
55	354		Poland	2021	Fresh dairy prod	Classic sweet yogh	5,41E+12	No	Danone	Danone
56	355		Poland	2021	Delicatessen me	Dry sausage	20501495	No	Deluxe	
57	356		Poland	2021	Bread products	Plain toasted breac	8,71E+12	No	Haust	
58	357		Poland	2021	Fresh dairy prod	Classic sweetened	27059647	Yes	Milsa	
59	358		Poland	2021	Fresh dairy prod	Classic plain fresh c	20669850	Yes	Milbc	
60	359		Poland	2021	Fresh dairy prod	Classic sweet yogh	5,41E+12	Yes	PurN	
61	360		Poland	2021	Bread products	Puffed cakes	3,56E+12	No	Carre	
62	361		Poland	2021	Fresh dairy prod	Artificially-sweete	5,41E+12	No	Danone	Danone
63	362		Poland	2021	Delicatessen me	Pâté	5,4E+12	No	Carrefour	
64	363		Poland	2021	Delicatessen me	Cured ham	5,4E+12	No	Carrefour	
65	364		Poland	2021	Bread products	Other rusks	6,12E+11	No	Nairns	
66	365		Poland	2021	Bread products	Pre-packaged brea	27046197	No	Délipain	
67	366		Poland	2021	Bread products	Wholemeal_cereal	7,31E+12	No	Krisprolls	
68	367		Poland	2021	Soft drinks	Flavoured milk bev	3,43E+12	No	Lactel	
69	368		Poland	2021	Bread products	Pre-packaged brea	5,41E+13	No	Biaform	
70	369		Poland	2021	Soft drinks	Sugar-sweetened f	27002766	No	River	
71	370		Poland	2021	Soft drinks	Other beverages w	8,85E+12	No	UFC	
72	371		Poland	2021	Soft drinks	Colas without adde	4,06E+12	No	Pepsi	
73	372		Poland	2021	Bread products	Other rusks	3,76E+12	No	Léanature	

Select the *Bar_codes* column, right click and go to ***Format Cells***

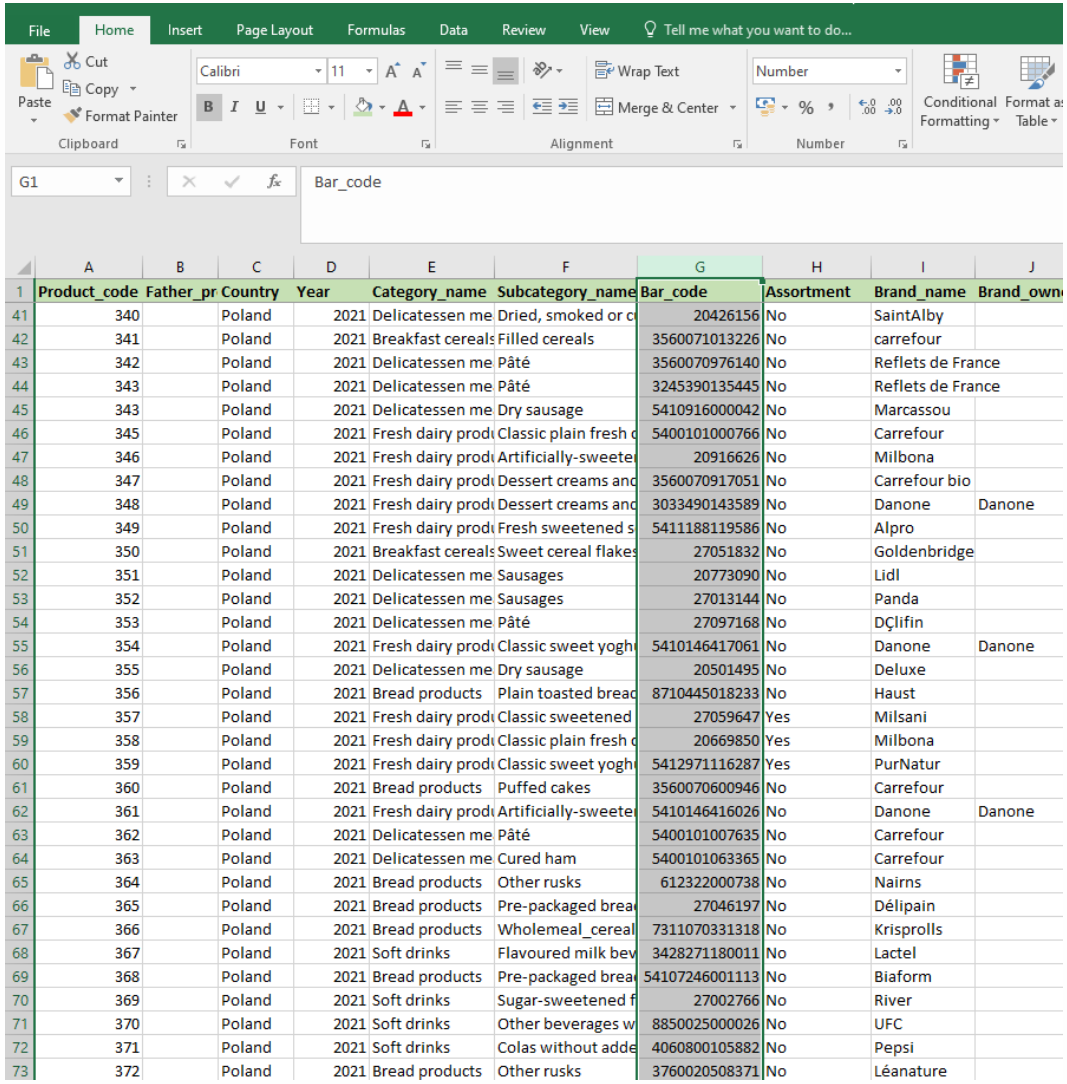
Preliminary steps



	A	B	C	D	E	F	G	H	I	J	K
1	Product_code	Father_pr	Country	Year	Category_						_of_l
41	340		Poland	2021	Delicatess						onal br
42	341		Poland	2021	Breakfast						onal b
43	342		Poland	2021	Delicatess						onal b
44	343		Poland	2021	Delicatess						onal b
45	343		Poland	2021	Delicatess						onal b
46	345		Poland	2021	Fresh dairy						onal b
47	346		Poland	2021	Fresh dairy						onal b
48	347		Poland	2021	Fresh dairy						onal b
49	348		Poland	2021	Fresh dairy						onal b
50	349		Poland	2021	Fresh dairy						onal b
51	350		Poland	2021	Breakfast						onal b
52	351		Poland	2021	Delicatess						onal b
53	352		Poland	2021	Delicatess						onal b
54	353		Poland	2021	Delicatess						onal b
55	354		Poland	2021	Fresh dairy						onal b
56	355		Poland	2021	Delicatess						onal b
57	356		Poland	2021	Bread prod						onal b
58	357		Poland	2021	Fresh dairy						onal b
59	358		Poland	2021	Fresh dairy						onal b
60	359		Poland	2021	Fresh dairy						onal b
61	360		Poland	2021	Bread products	Puffed cakes	3,56E+12	No	Carrefour		National b
62	361		Poland	2021	Fresh dairy prod	Artificially-sweete	5,41E+12	No	Danone	Danone	National b
63	362		Poland	2021	Delicatessen me	Pâté	5,4E+12	No	Carrefour		National b
64	363		Poland	2021	Delicatessen me	Cured ham	5,4E+12	No	Carrefour		National b
65	364		Poland	2021	Bread products	Other rusks	6,12E+11	No	Nairns		National b
66	365		Poland	2021	Bread products	Pre-packaged brea	27046197	No	Délipain		National b
67	366		Poland	2021	Bread products	Wholemeal_cereal	7,31E+12	No	Krisprolls		Retailer b
68	367		Poland	2021	Soft drinks	Flavoured milk bev	3,43E+12	No	Lactel		Retailer b
69	368		Poland	2021	Bread products	Pre-packaged brea	5,41E+13	No	Biaform		Retailer b
70	369		Poland	2021	Soft drinks	Sugar-sweetened f	27002766	No	River		Retailer b
71	370		Poland	2021	Soft drinks	Other beverages w	8,85E+12	No	UFC		Retailer b
72	371		Poland	2021	Soft drinks	Colas without adde	4,06E+12	No	Pepsi		Retailer b
73	372		Poland	2021	Bread products	Other rusks	3,76E+12	No	Léanature		Retailer b

In the ***Number*** tab, choose the '**number**' category, indicate '**0**' for decimal places and click OK

Preliminary steps



	A	B	C	D	E	F	G	H	I	J
1	Product_code	Father_pr	Country	Year	Category_name	Subcategory_name	Bar_code	Assortment	Brand_name	Brand_own
41	340		Poland	2021	Delicatessen me	Dried, smoked or c	20426156	No	SaintAlby	
42	341		Poland	2021	Breakfast cereals	Filled cereals	3560071013226	No	carrefour	
43	342		Poland	2021	Delicatessen me	Pâté	3560070976140	No	Reflets de France	
44	343		Poland	2021	Delicatessen me	Pâté	3245390135445	No	Reflets de France	
45	343		Poland	2021	Delicatessen me	Dry sausage	5410916000042	No	Marcassou	
46	345		Poland	2021	Fresh dairy prod	Classic plain fresh c	5400101000766	No	Carrefour	
47	346		Poland	2021	Fresh dairy prod	Artificially-sweete	20916626	No	Milbona	
48	347		Poland	2021	Fresh dairy prod	Dessert creams and	3560070917051	No	Carrefour bio	
49	348		Poland	2021	Fresh dairy prod	Dessert creams and	3033490143589	No	Danone	Danone
50	349		Poland	2021	Fresh dairy prod	Fresh sweetened s	5411188119586	No	Alpro	
51	350		Poland	2021	Breakfast cereals	Sweet cereal flakes	27051832	No	Goldenbridge	
52	351		Poland	2021	Delicatessen me	Sausages	20773090	No	Lidl	
53	352		Poland	2021	Delicatessen me	Sausages	27013144	No	Panda	
54	353		Poland	2021	Delicatessen me	Pâté	27097168	No	DÇlifin	
55	354		Poland	2021	Fresh dairy prod	Classic sweet yogh	5410146417061	No	Danone	Danone
56	355		Poland	2021	Delicatessen me	Dry sausage	20501495	No	Deluxe	
57	356		Poland	2021	Bread products	Plain toasted breac	8710445018233	No	Haust	
58	357		Poland	2021	Fresh dairy prod	Classic sweetened	27059647	Yes	Milsani	
59	358		Poland	2021	Fresh dairy prod	Classic plain fresh c	20669850	Yes	Milbona	
60	359		Poland	2021	Fresh dairy prod	Classic sweet yogh	5412971116287	Yes	PurNatur	
61	360		Poland	2021	Bread products	Puffed cakes	3560070600946	No	Carrefour	
62	361		Poland	2021	Fresh dairy prod	Artificially-sweete	5410146416026	No	Danone	Danone
63	362		Poland	2021	Delicatessen me	Pâté	5400101007635	No	Carrefour	
64	363		Poland	2021	Delicatessen me	Cured ham	5400101063365	No	Carrefour	
65	364		Poland	2021	Bread products	Other rusks	612322000738	No	Nairns	
66	365		Poland	2021	Bread products	Pre-packaged brea	27046197	No	Délipain	
67	366		Poland	2021	Bread products	Wholemeal_cereal	7311070331318	No	Krisprolls	
68	367		Poland	2021	Soft drinks	Flavoured milk bev	3428271180011	No	Lactel	
69	368		Poland	2021	Bread products	Pre-packaged brea	54107246001113	No	Biaform	
70	369		Poland	2021	Soft drinks	Sugar-sweetened f	27002766	No	River	
71	370		Poland	2021	Soft drinks	Other beverages w	8850025000026	No	UFC	
72	371		Poland	2021	Soft drinks	Colas without adde	4060800105882	No	Pepsi	
73	372		Poland	2021	Bread products	Other rusks	3760020508371	No	Léanature	

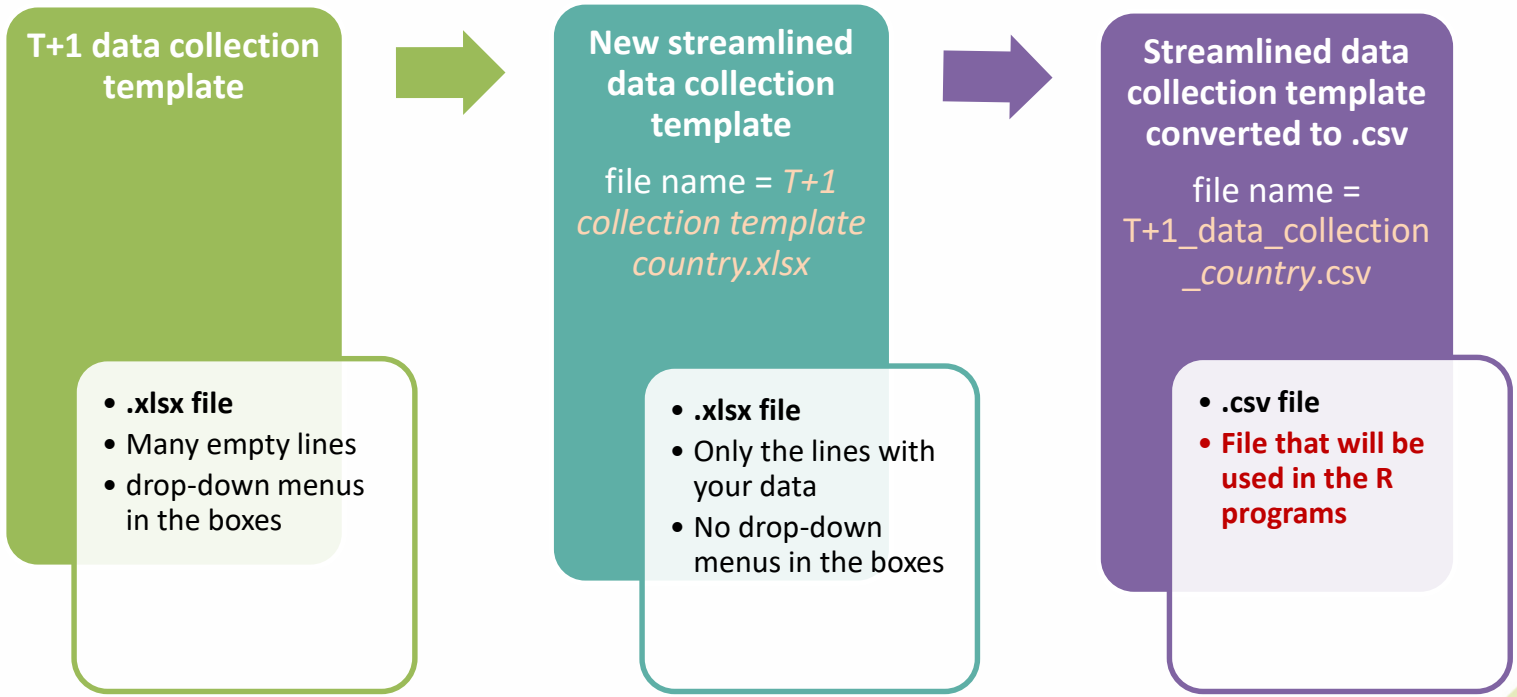
Your barcodes appear in full, you can save this table by overwriting the previous version and close it.

Please note! If you open this file again, you will have to do the same operation again. The numbers are automatically converted to scientific format when opening a .csv file.

Preliminary steps

1st preliminary step : preparation of the T+1 collection template

Summary of the manipulations in the 1st preliminary step



Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

2nd preliminary step : preparation of the pre-existing data template

→ You must **make a copy** of your file containing your pre-existing data that has been reclassified into the Best-ReMaP nomenclature. You must save this copy in **.csv format** under the name pre_existing_data_***country***.csv (with the name of your own country)

→ You can follow the procedure on **pages** **19 to 24** for creating the copy in .csv format.

→ You should also follow the procedure on **pages** **24 to 28** each time you open the file "pre_existing_data_*country*.csv" so that you do not lose any barcode information

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

2nd preliminary step : preparation of the pre-existing data template

Summary of the manipulations in the 2nd preliminary step

Your file containing your pre-existing data that has been reclassified into the Best-ReMaP nomenclature

- **.xlsx file**

File containing your pre-existing data converted to .csv

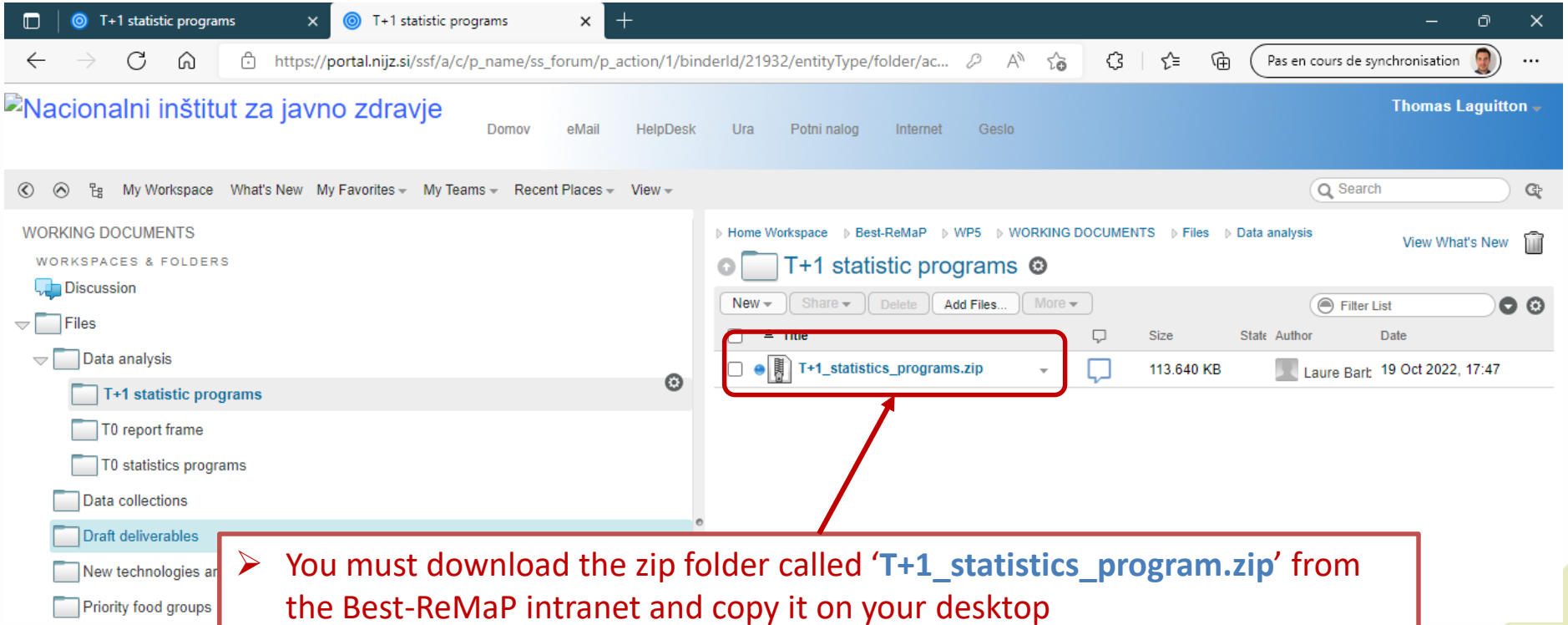
file name = pre_existing_data_*country*.csv

- **.csv file**
- **File that will be used in the R programs**
- **This file will only be used in the 'Verifications template_step_4' program to check the matching of the father products**

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

3rd preliminary step : creation of the working folder



- You must download the zip folder called '**T+1_statistics_program.zip**' from the Best-ReMaP intranet and copy it on your desktop

https://portal.nijz.si/ssf/a/c/p_name/ss_forum/p_action/1/binderId/21932/entityType/folder/action/view_permalink/novl_url/1

- You must unzip this folder before using it

Preliminary steps

3rd preliminary step : creation of the working folder

Please copy this file directly on your desktop

T+1_statistics_programs

T+1_statistics_programs.Rproj

= This is the project you need to open for the task. All the work will be done here.

Files

- wp5_verification_file.csv
- Best-ReMaP_nomenclature.csv
- Subcategory_order.csv
- Years of interest.csv
  → *file to be filled in*

Programs

- R_setup.R
- Verifications template_step_1.R
- Verifications template_step_2.R
- Verifications template_step_3.R
- Verifications template_step_4.R

renv

Folder not to be opened/modified. It contains the working environment of the project.

Indicators

This file is not used for the verification steps.
It will be explained at the 2nd part of T+1 statistics training (24 nov.)

Copy here :

- your file with your collected and encoded data :
→ T+1_data_collection_country.csv

- your file with your preexisting data encoded :
→ pre_existing_data_country.csv

Files must be saved in .csv format (comma delimited)

It is very important that you don't change any names of the files that are listed here (except the file with your collected and encoded data)

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

4th preliminary step : preparation of the file "Years of interest.csv"

→ In the working folder **T+1_statistics_programs** that you copied to your desktop in the 3rd preliminary step, there is a file called "**Years of interest.csv**" in the folder *'Files'.*

→ Before starting the task 5.4.1, we asked you to select which **years of each category** of your pre-existing data **will be used for pairing + comparisons/creation of indicators** with the data collected during Best-ReMaP. The selected years are the ones we called "years of interest".

→ You will therefore need to edit the file "**Years of interest.csv**" manually by entering the years of interest (those that you have selected) of your pre-existing data for each Best-ReMaP category.

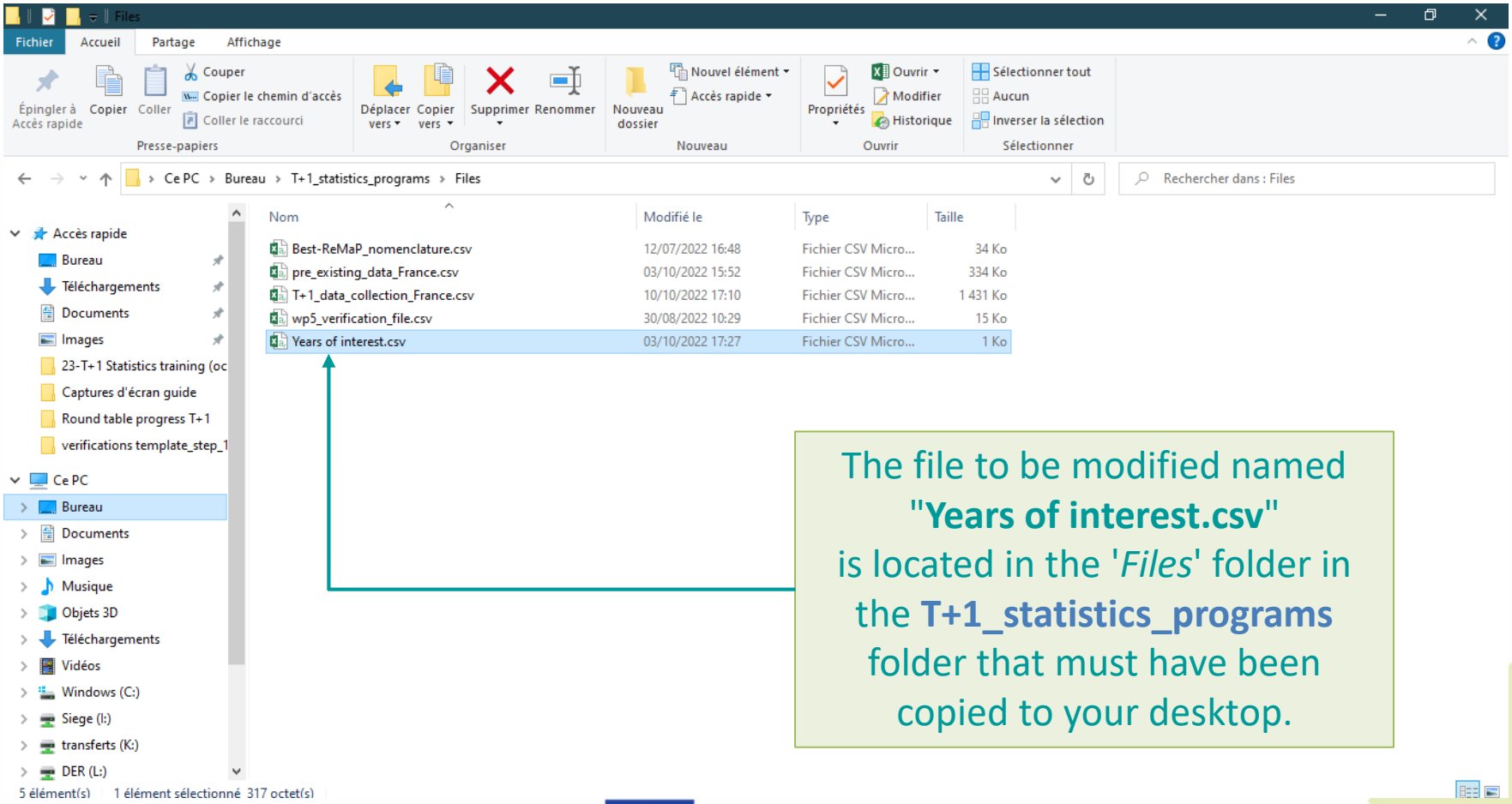
→ Once you have modified this file, you should save the changes.
This file will be used in the 4th data verification step (page 166)

see the following slide for an example

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

4th preliminary step : preparation of the file "Years of interest.csv"

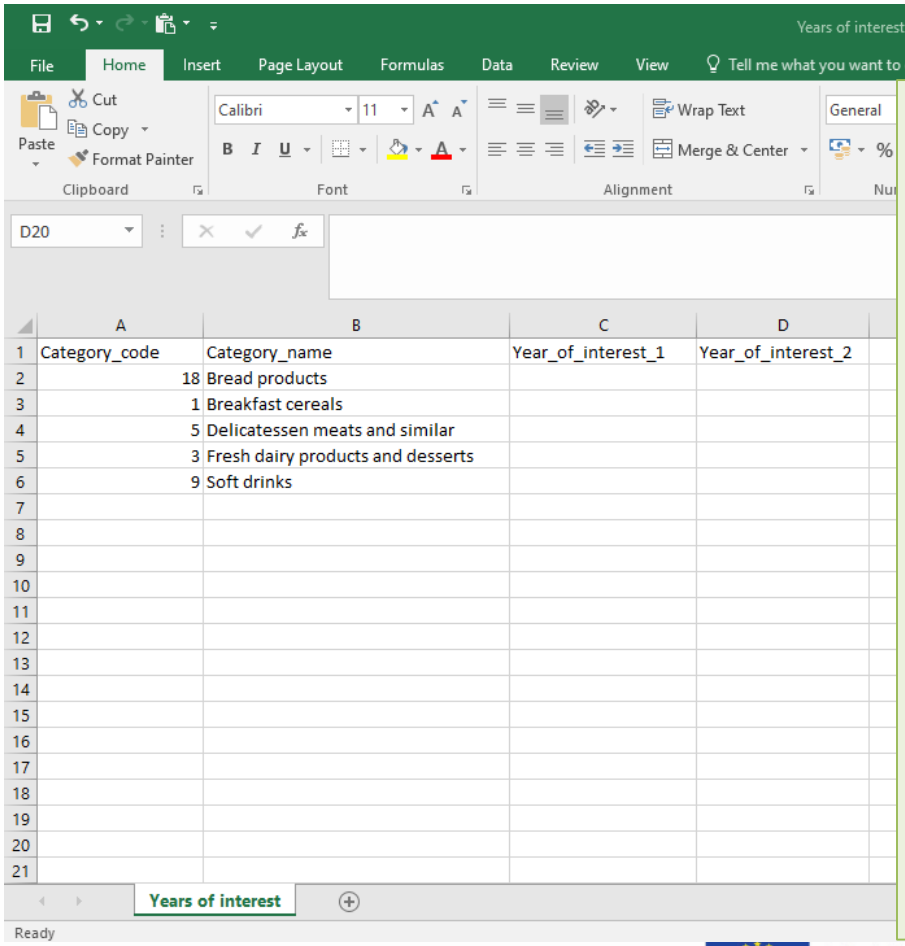


The file to be modified named "**Years of interest.csv**" is located in the '*Files*' folder in the **T+1_statistics_programs** folder that must have been copied to your desktop.

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

4th preliminary step : preparation of the file "Years of interest.csv"



	A	B	C	D
1	Category_code	Category_name	Year_of_interest_1	Year_of_interest_2
2	18	Bread products		
3	1	Breakfast cereals		
4	5	Delicatessen meats and similar		
5	3	Fresh dairy products and desserts		
6	9	Soft drinks		

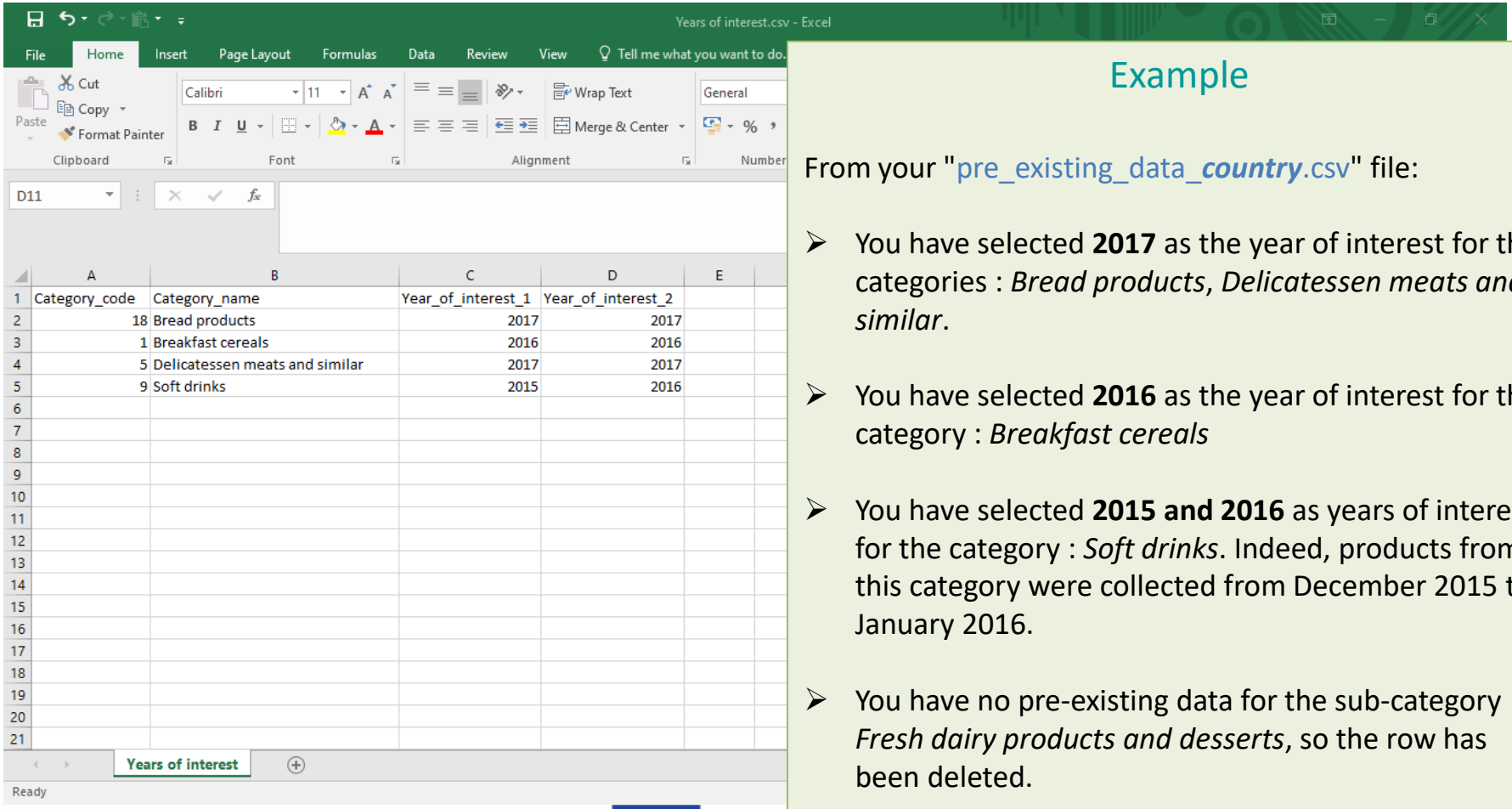
Overview of the file "Years of interest.csv"

- Column **'Year_of_interest_1'** : you must enter for each category the selected year of interest from your "pre_existing_data_***country***.csv" file.
  (other years of collection for the category will not be taking into account during the creation of indicators)
- Column **'Year_of_interest_2':** you must enter the other year of interest of the category if there is a second year of interest for the category (a collection may have occurred on two consecutive years). If there is only one year of interest for the category, it should be repeated in this column.
- If you do not have pre-existing data for one or more of the categories monitored in Best-ReMaP, you must delete the row(s) for that/those category(ies).

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

4th preliminary step : preparation of the file "Years of interest.csv"



	A	B	C	D	E
1	Category_code	Category_name	Year_of_interest_1	Year_of_interest_2	
2	18	Bread products	2017	2017	
3	1	Breakfast cereals	2016	2016	
4	5	Delicatessen meats and similar	2017	2017	
5	9	Soft drinks	2015	2016	

Example

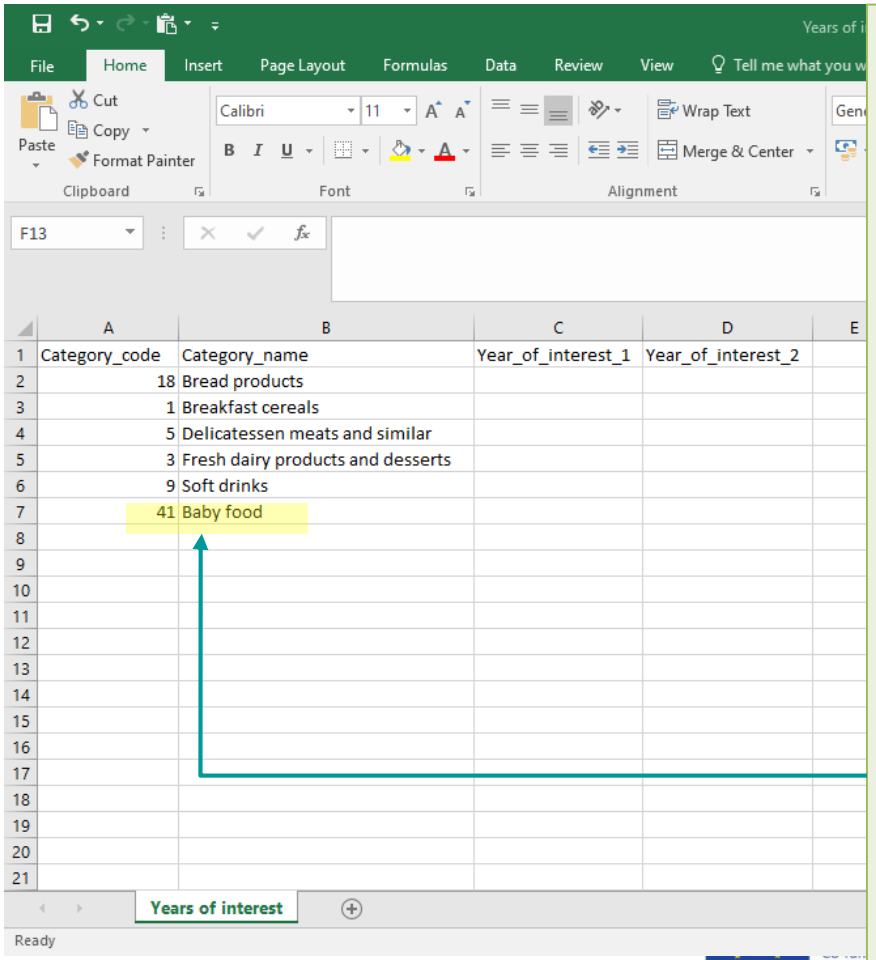
From your "pre_existing_data_***country***.csv" file:

- You have selected **2017** as the year of interest for the categories : *Bread products*, *Delicatessen meats and similar*.
- You have selected **2016** as the year of interest for the category : *Breakfast cereals*
- You have selected **2015 and 2016** as years of interest for the category : *Soft drinks*. Indeed, products from this category were collected from December 2015 to January 2016.
- You have no pre-existing data for the sub-category *Fresh dairy products and desserts*, so the row has been deleted.

Co-funded by the European Union's Health Programme (2014-2020)

Preliminary steps

4th preliminary step : preparation of the file "Years of interest.csv"



	A	B	C	D
1	Category_code	Category_name	Year_of_interest_1	Year_of_interest_2
2	18	Bread products		
3	1	Breakfast cereals		
4	5	Delicatessen meats and similar		
5	3	Fresh dairy products and desserts		
6	9	Soft drinks		
7	41	Baby food		

File "Years of interest.csv"

- This file is a **key file** for the rest of the programs as it allows the identification of the categories for which indicators will be generated. If a category for which you have pre-existing data is forgotten in this file, then it will not be taken into account for the creation of indicators.
  (see next training on 24 November)

- If you have other categories that are not monitored in Best-ReMaP but for which you wish to generate indicators, you must enter this category in the file. But these indicators will not appear in the Best-ReMaP reports.

Example:
You wish to generate indicators for another category which is *Baby food (41)*, you must enter the category_code and category_name + the years of interest of your pre-existing data for this category

2) Installation of the necessary equipment and presentation of the Rstudio software

A. Preliminary steps

B. Installation of software

C. Introduction to R studio

D. Cleaning of the Rstudio interface

Co-funded by the European Union's Health Programme (2014-2020)

Installation of R software

- **Download the R software:**
  Follow this link and select **version 4.1.2** (which is not the latest version but the version on which the programs were created):
  https://cran.r-project.org/bin/windows/base/old/

Version of R to download (click on the link)

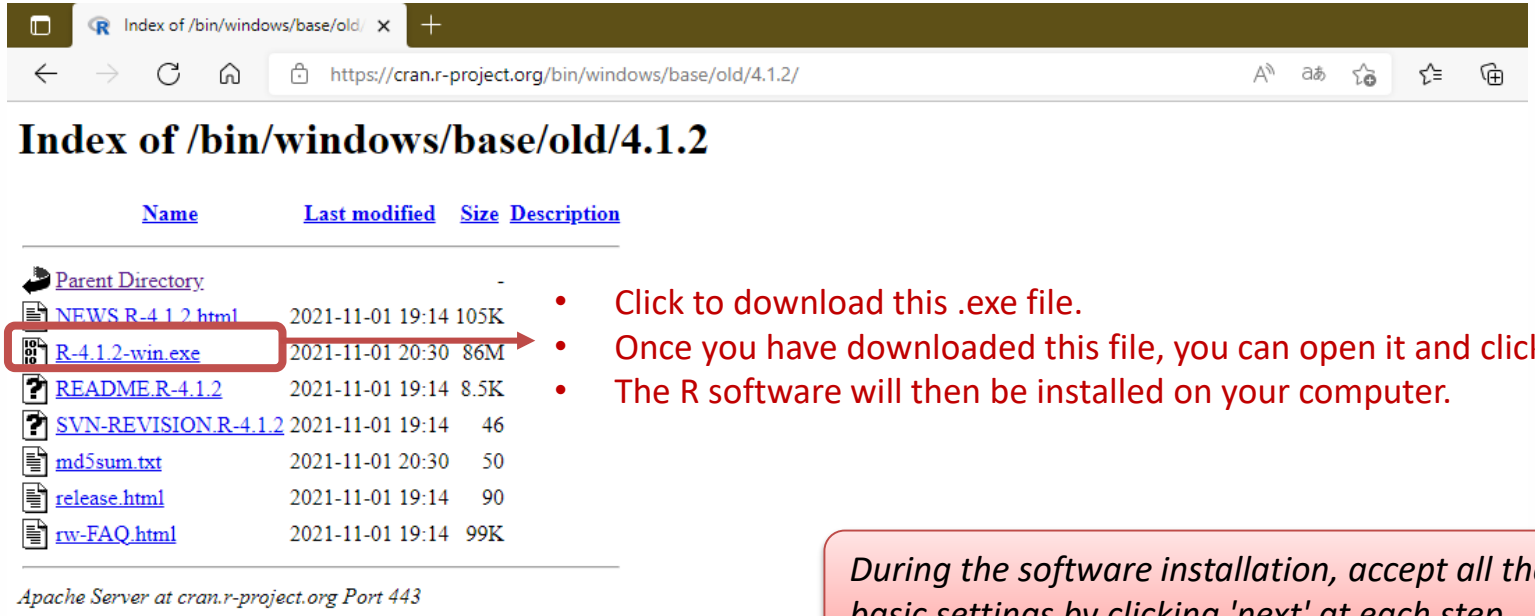
If you already have R on your computer, check which version of the software you have.
If it is a version **other than 4.1.2**, then you need to download version **4.1.2** as shown.
→ You will have **2 versions of R** on your computer.

see the following slide for next step

Co-funded by the European Union's Health Programme (2014-2020)

Installation of R software

- **Download the R software:**
  Follow this link and select **version 4.1.2** (which is not the latest version but the version on which the programs were created):
  https://cran.r-project.org/bin/windows/base/old/



- Click to download this .exe file.
- Once you have downloaded this file, you can open it and click on 'Run'.
- The R software will then be installed on your computer.

During the software installation, accept all the basic settings by clicking 'next' at each step

Co-funded by the European Union's Health Programme (2014-2020)

Installation of R software

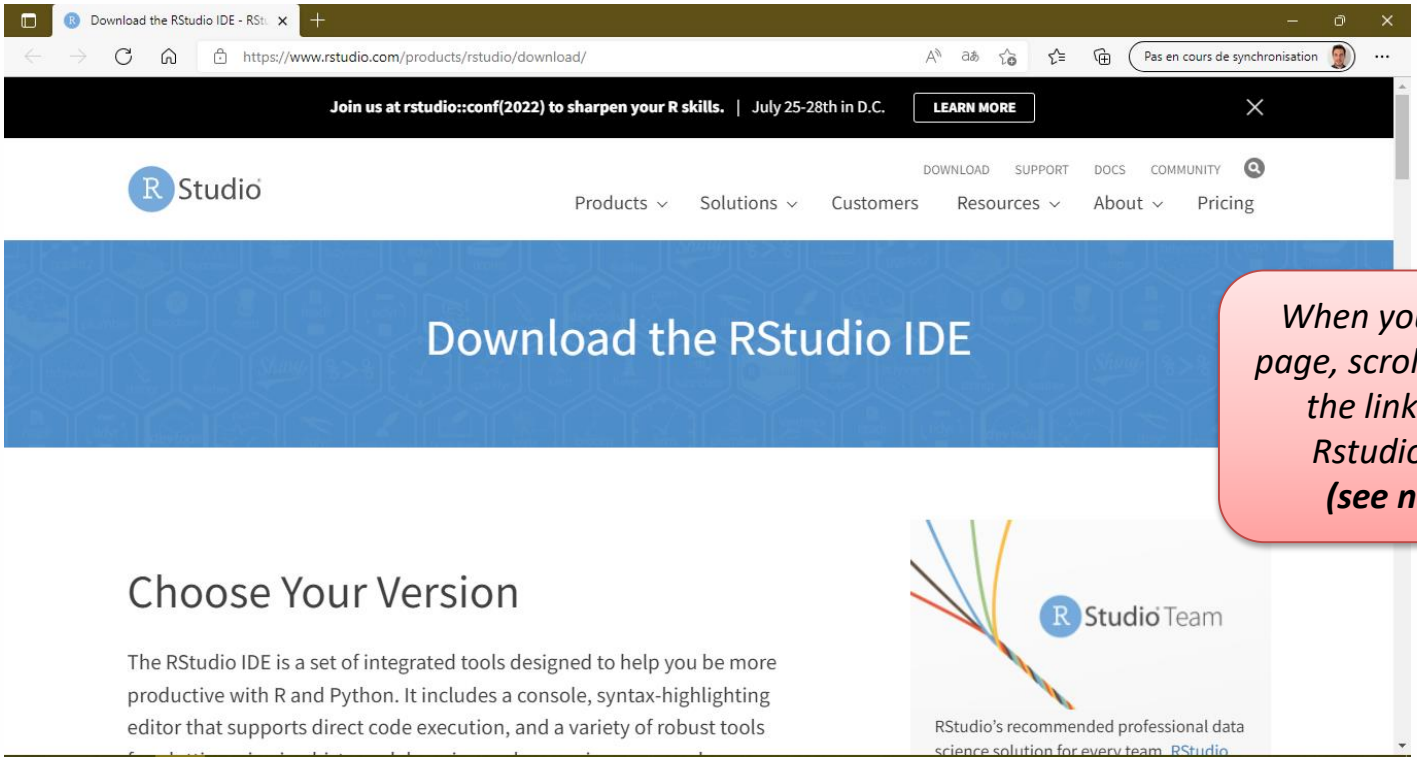
Tutorial video to download and install the R software

→ *This video is available at any time on the Best ReMaP intranet in the WP5 section*
https://portal.nijz.si/ssf/a/c/p_name/ss_forum/p_action/1/entityType/folderEntry/action/view_permalink/entryId/77697/novl_url/1

Co-funded by the European Union's Health Programme (2014-2020)

Installation of Rstudio software

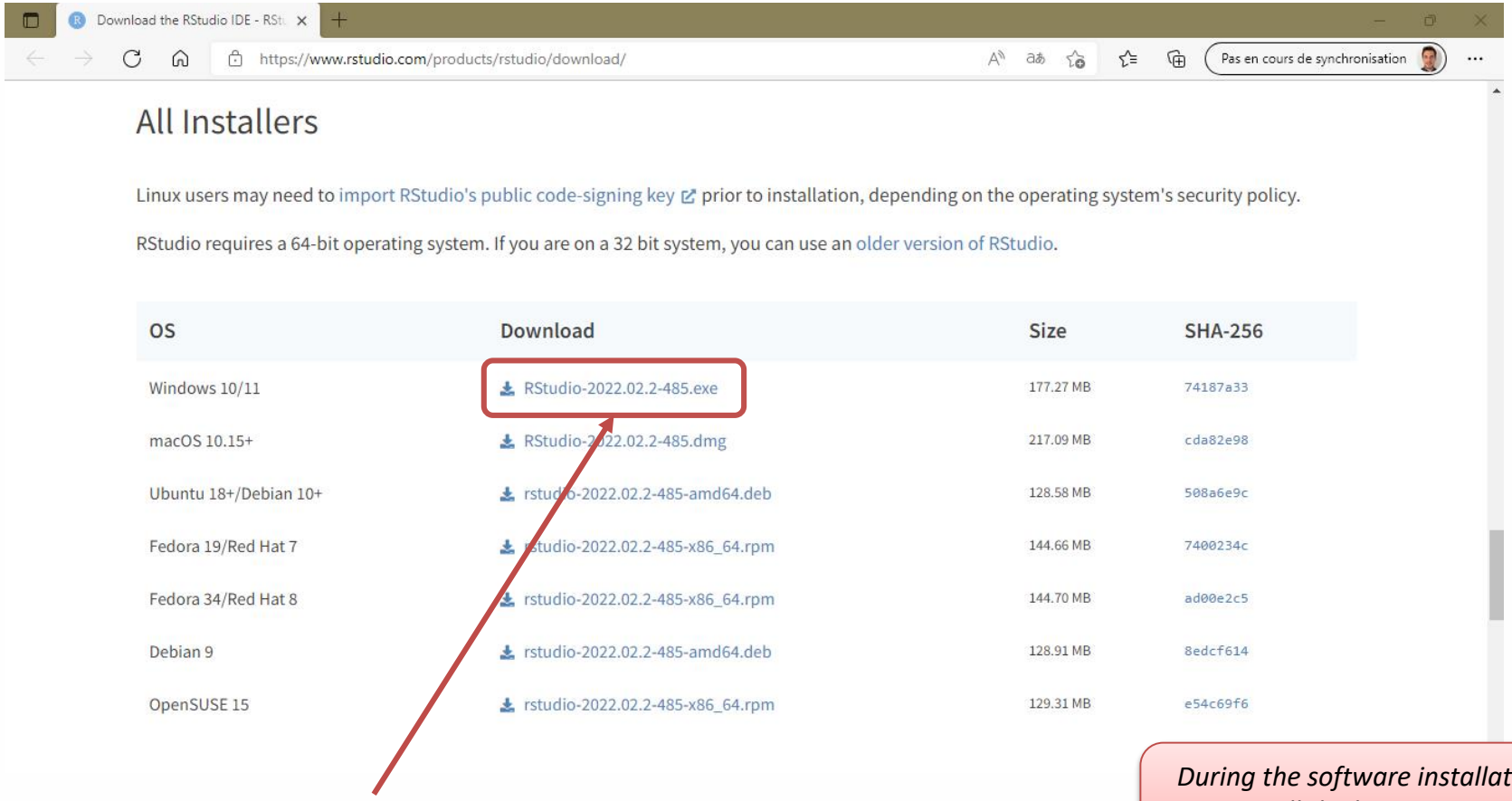
- **Download the Rstudio software:**
  Follow this link :
  https://www.rstudio.com/products/rstudio/download/



When you are on this page, scroll down to find the links to all the Rstudio installers ***(see next slide)***

Co-funded by the European Union's Health Programme (2014-2020)

Installation of Rstudio software



- Click to download this .exe file.
- Once you have downloaded this file, you can open it and click on 'Run'.
- The Rstudio software will then be installed on your computer.

During the software installation, accept all the basic settings by clicking 'next' at each step

Co-funded by the European Union's Health Programme (2014-2020)

Installation of Rstudio software

Tutorial video to download and install the Rstudio software

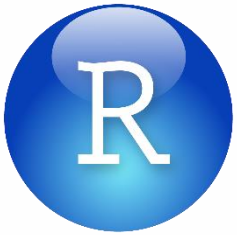
→ *This video is available at any time on the Best ReMaP intranet in the WP5 section*
https://portal.nijz.si/ssf/a/c/p_name/ss_forum/p_action/1/entityType/folderEntry/action/view_permalink/entryId/77698/novl_url/1

Co-funded by the European Union's Health Programme (2014-2020)

Overview of R et Rstudio interfaces

'R'

→ No processing will be done on this interface
→ Software needed to be able to work on Rstudio

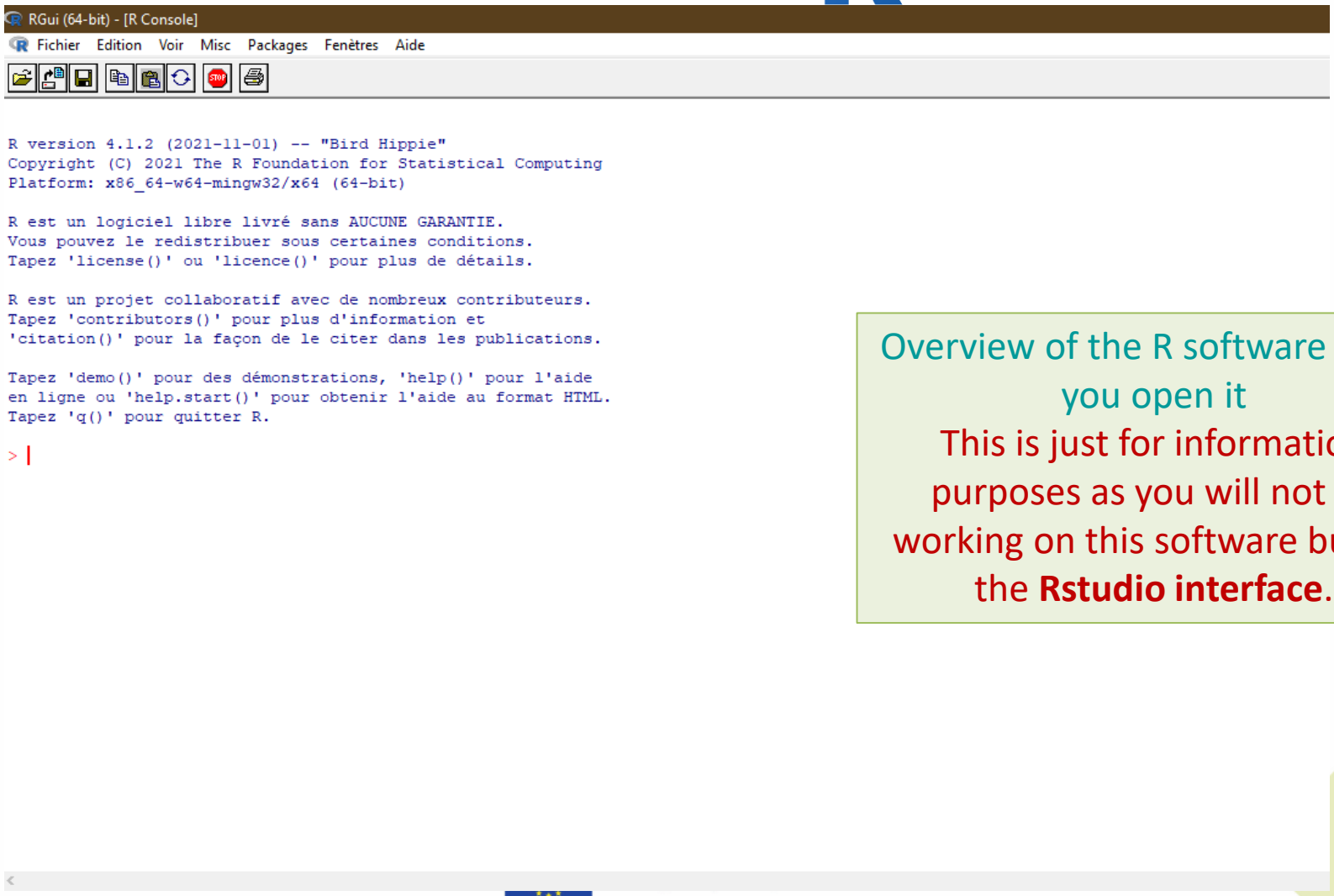


'Rstudio'

→ Interface that allows the software to be used = environment that facilitates input, code execution and visualisation of results
→ Programs will be running through RStudio

Co-funded by the European Union's Health Programme (2014-2020)

Overview of R



Overview of the R software when you open it

This is just for information purposes as you will not be working on this software but on the **Rstudio interface**.

2) Installation of the necessary equipment and presentation of the Rstudio software

A. Preliminary steps

B. Installation of software

C. Introduction to R studio

D. Cleaning of the Rstudio interface

Co-funded by the European Union's Health Programme (2014-2020)

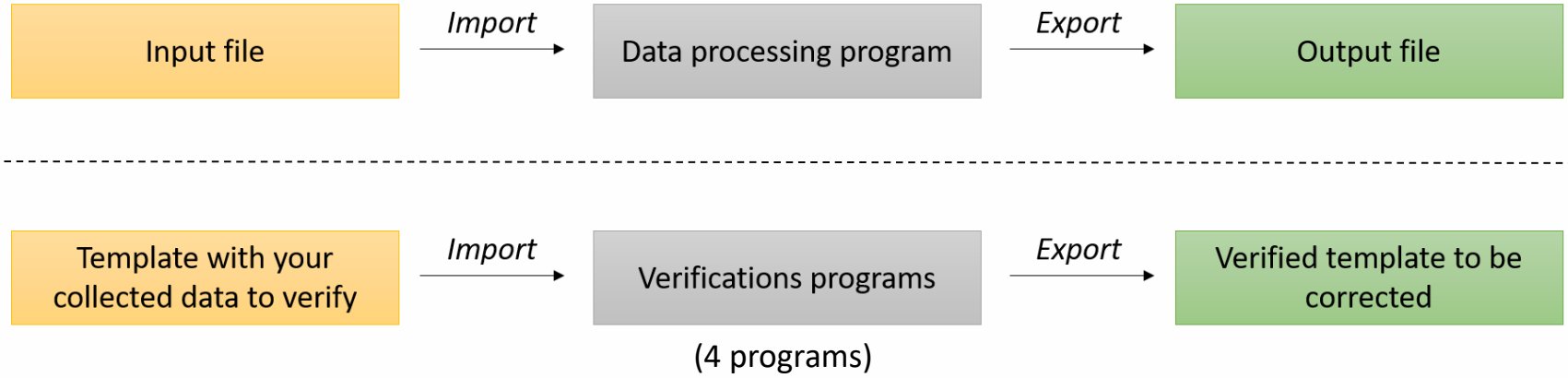
Introduction to Rstudio software

How does data processing software like Rstudio work?

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software

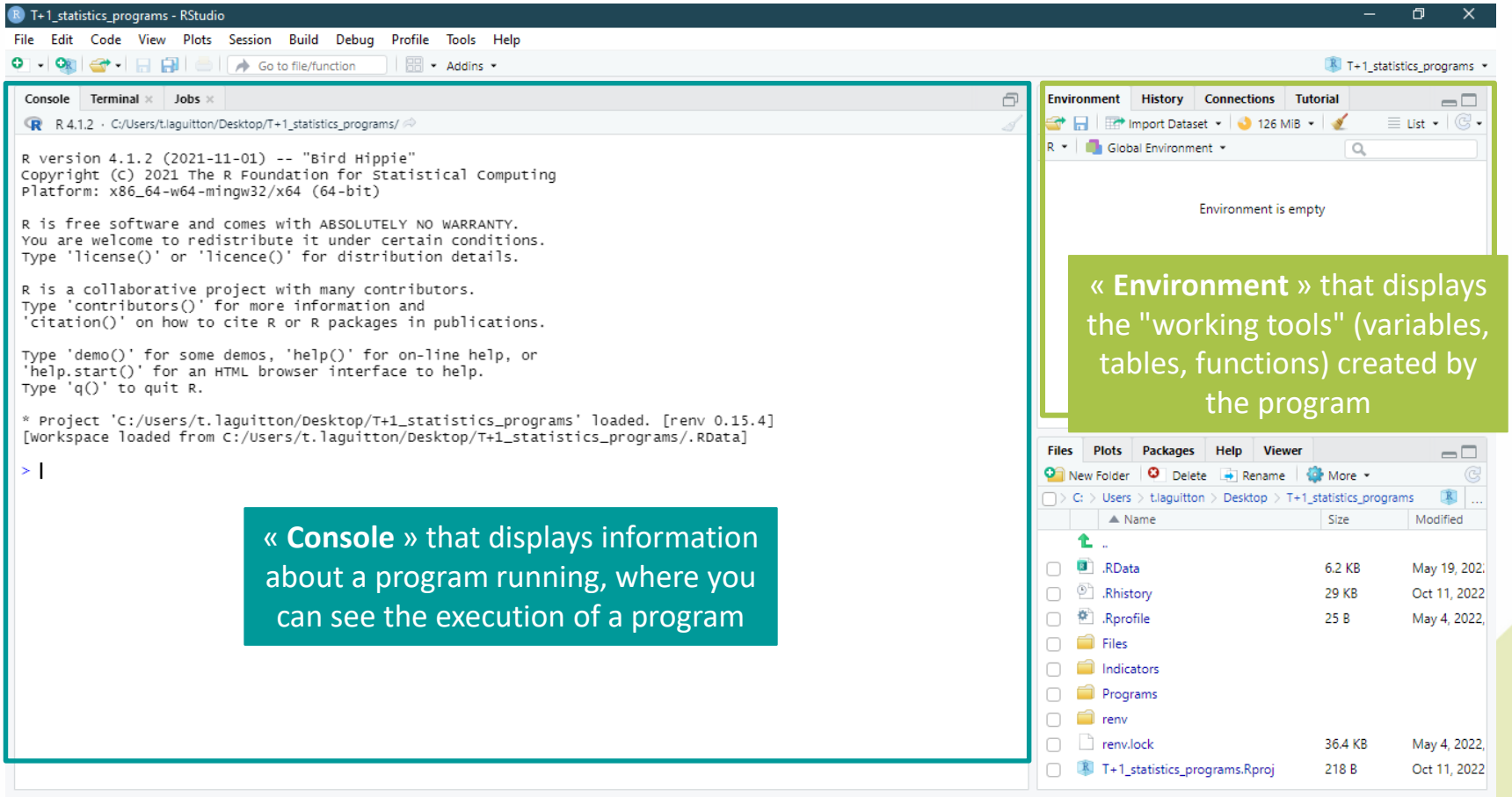
How will the Rstudio data processing software work in Task 5.3.2?



Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software

Overview of the Rstudio software when you open it for the first time.



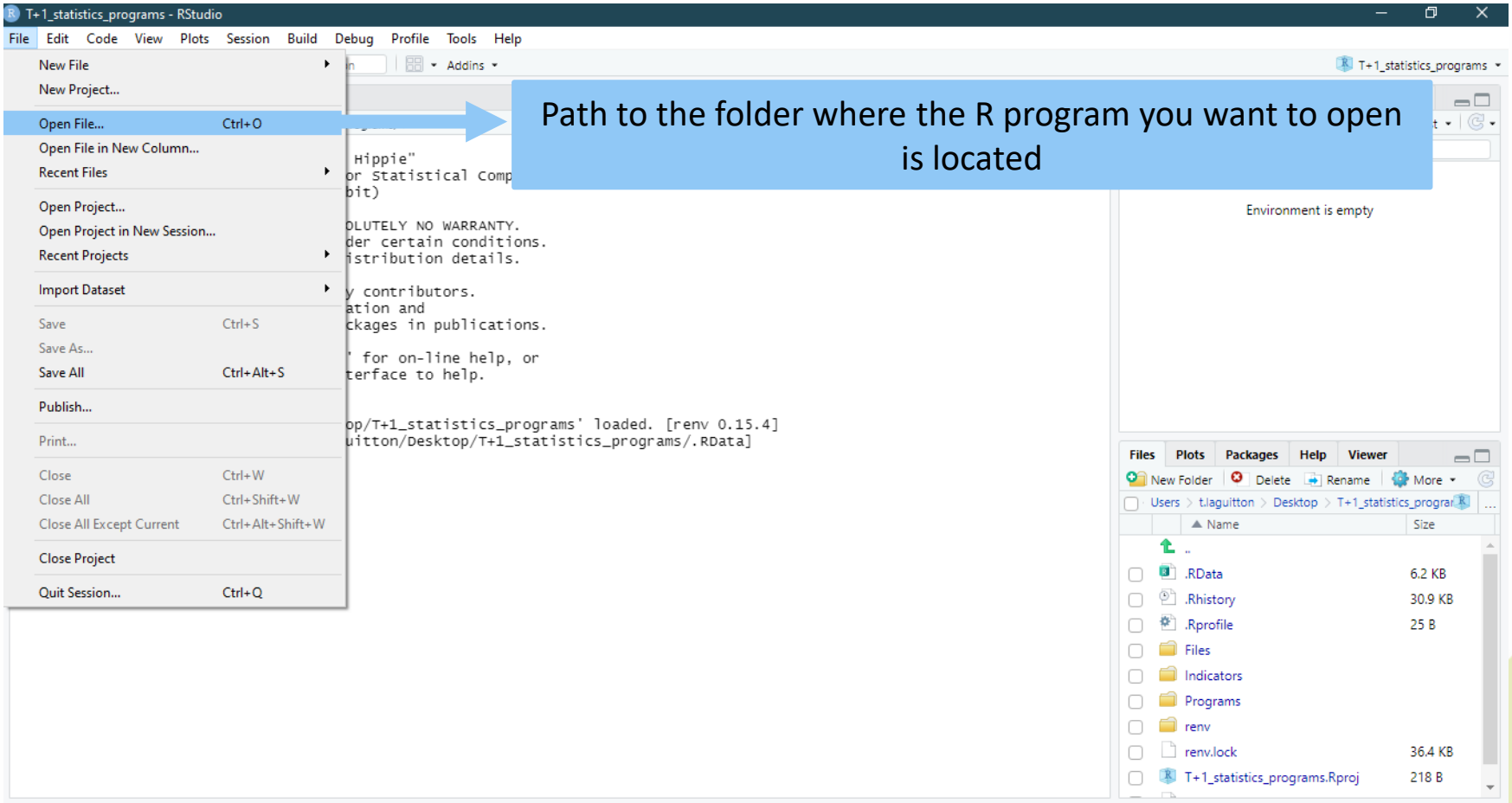
« **Console** » that displays information about a program running, where you can see the execution of a program

« **Environment** » that displays the "working tools" (variables, tables, functions) created by the program

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software

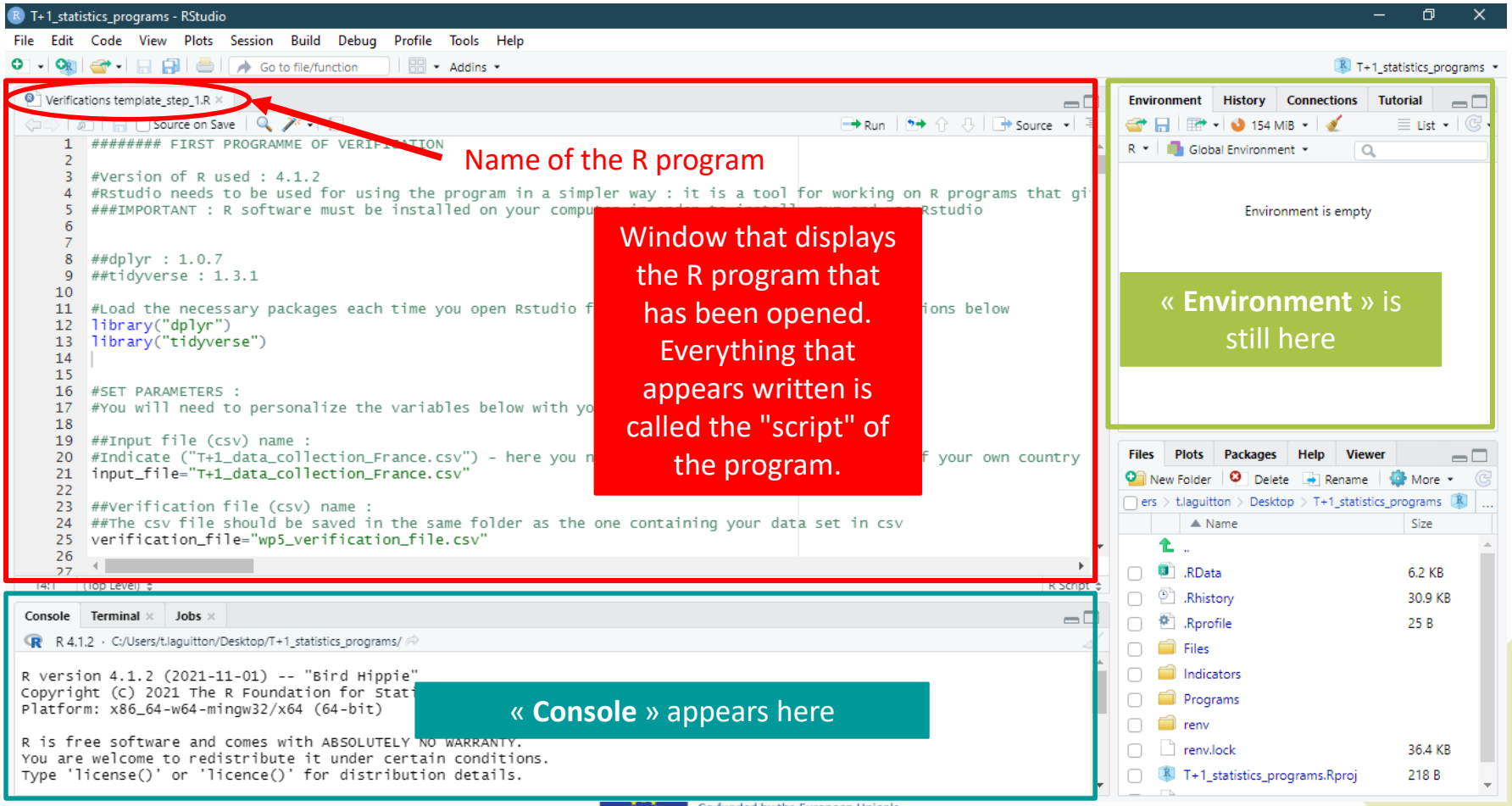
How to open a program in Rstudio



Co-funded by the European Union's Health Programme (2014-2020)

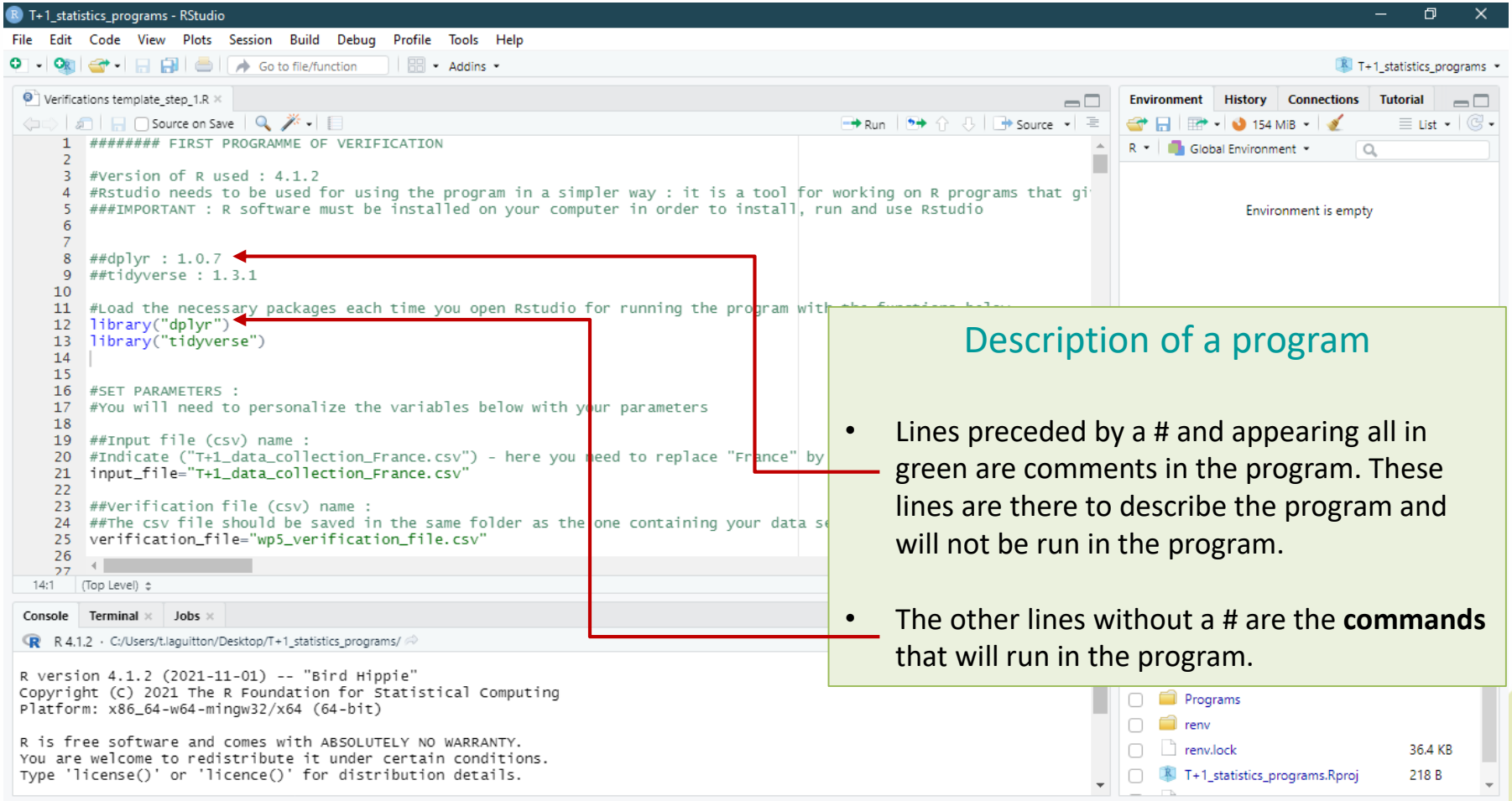
Introduction to Rstudio software

Rstudio when a program is opened



Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software

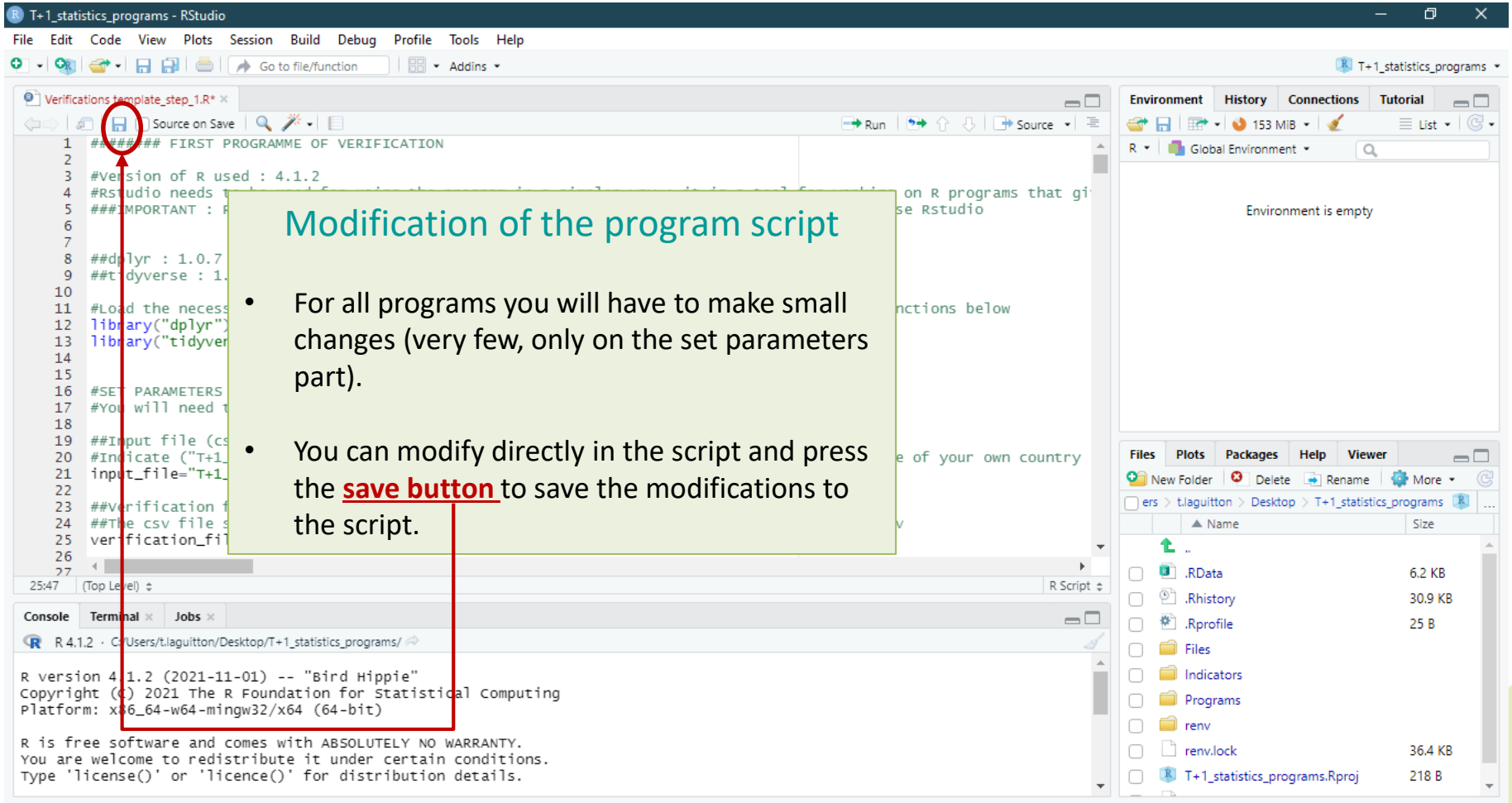


Description of a program

- Lines preceded by a # and appearing all in green are comments in the program. These lines are there to describe the program and will not be run in the program.
- The other lines without a # are the **commands** that will run in the program.

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software

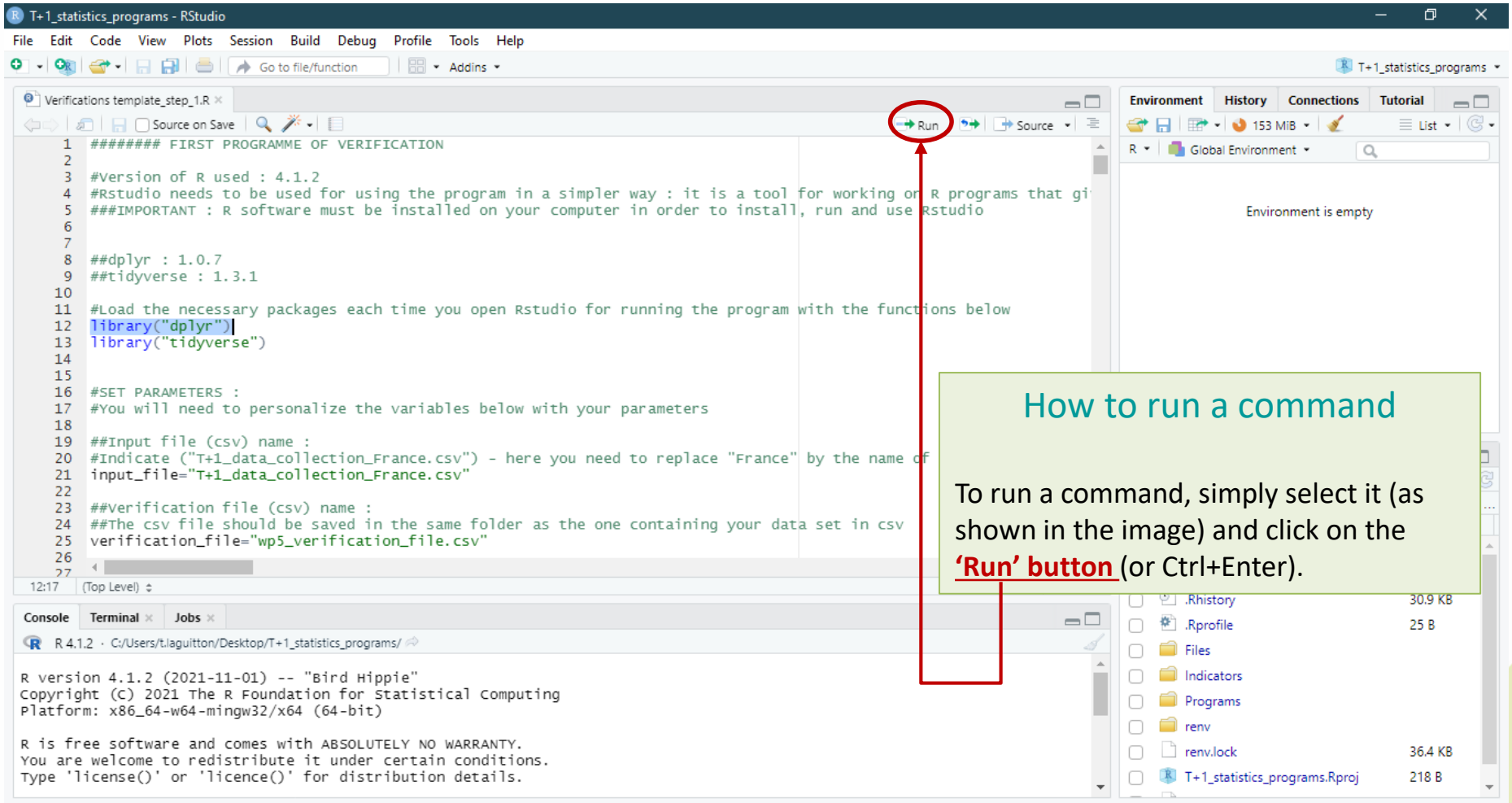


Modification of the program script

- For all programs you will have to make small changes (very few, only on the set parameters part).
- You can modify directly in the script and press the **save button** to save the modifications to the script.

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software

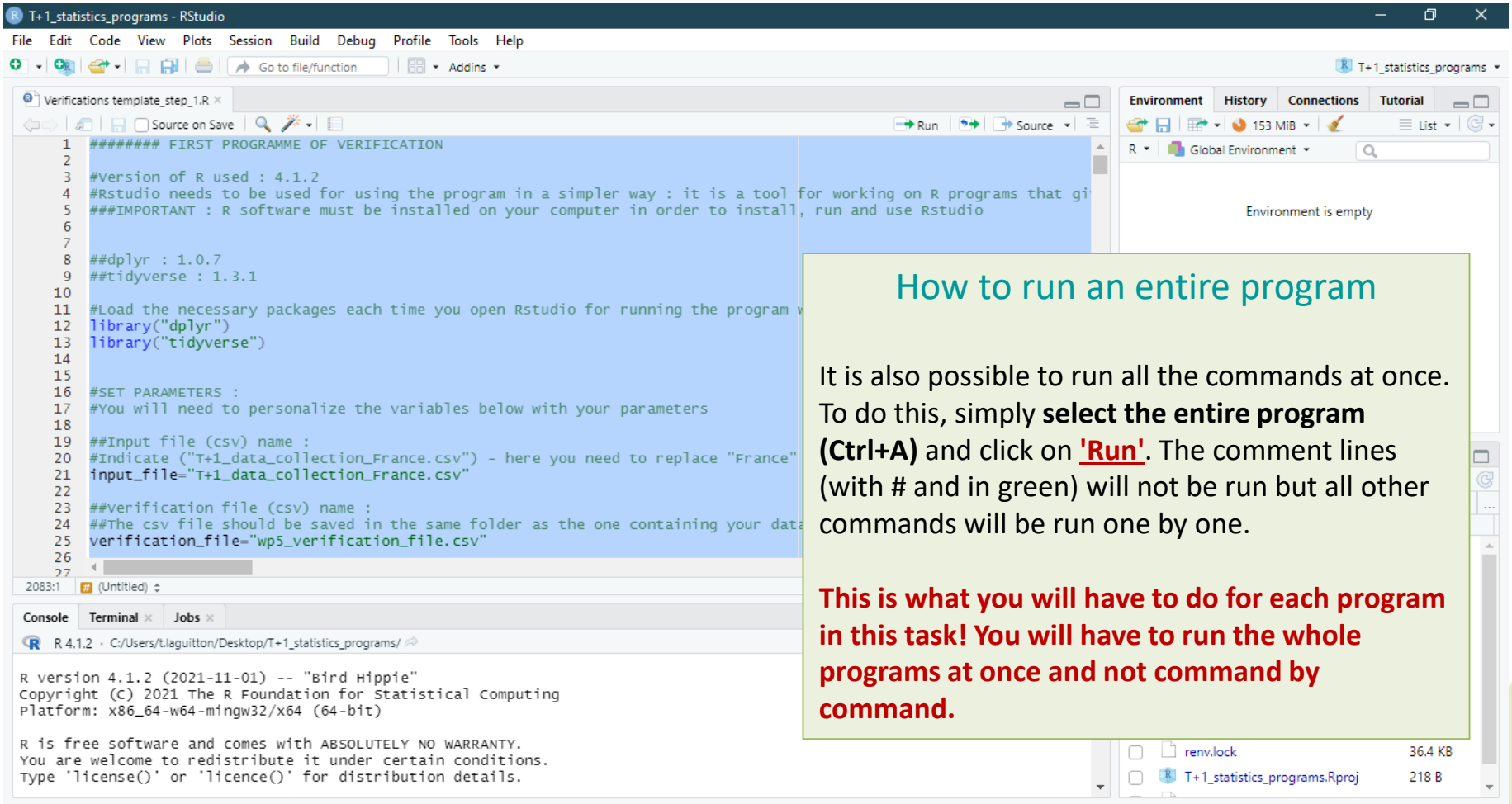


How to run a command

To run a command, simply select it (as shown in the image) and click on the **'Run' button** (or Ctrl+Enter).

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software



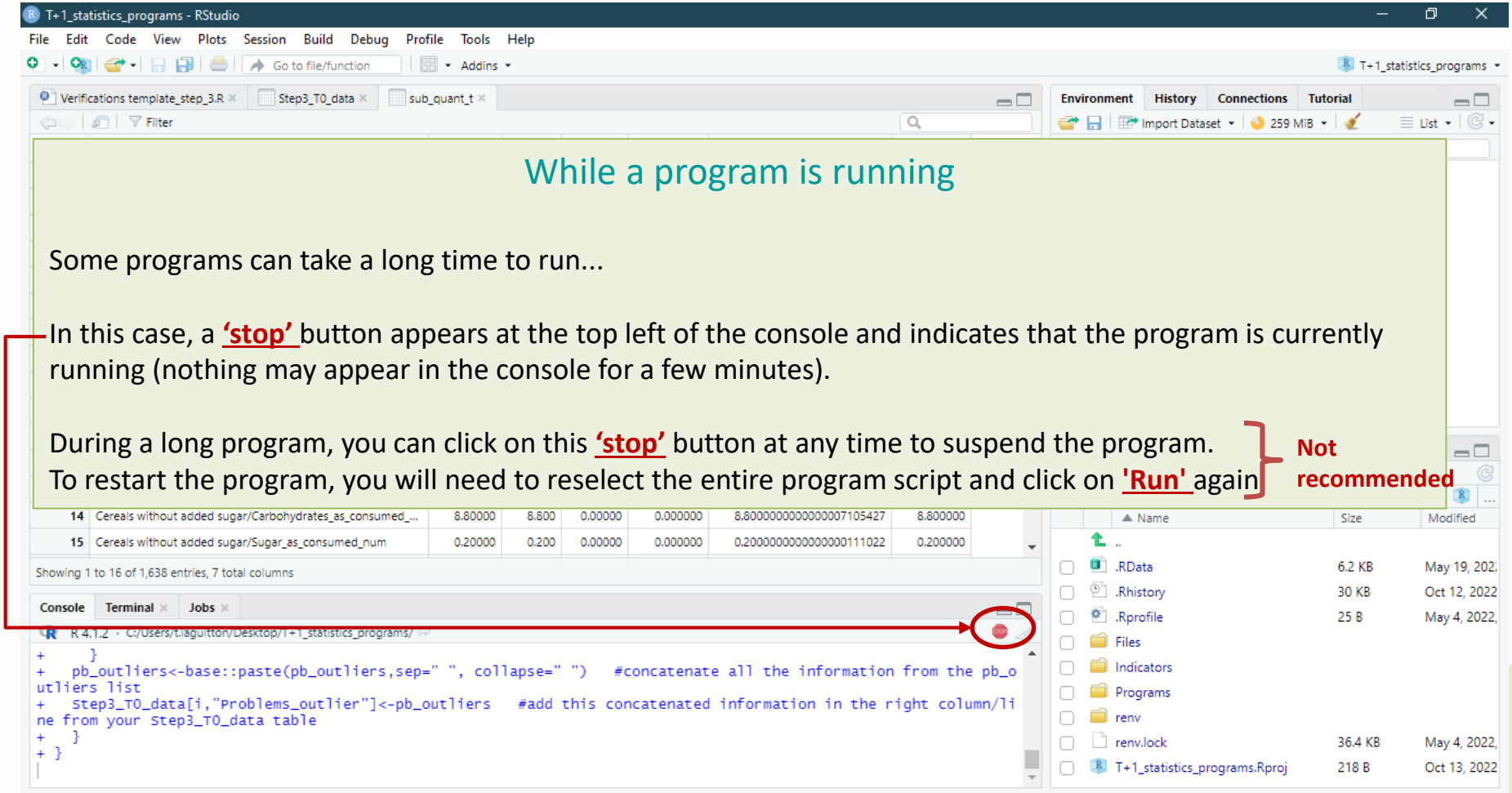
How to run an entire program

It is also possible to run all the commands at once. To do this, simply **select the entire program (Ctrl+A)** and click on **'Run'**. The comment lines (with # and in green) will not be run but all other commands will be run one by one.

This is what you will have to do for each program in this task! You will have to run the whole programs at once and not command by command.

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software



While a program is running

Some programs can take a long time to run...

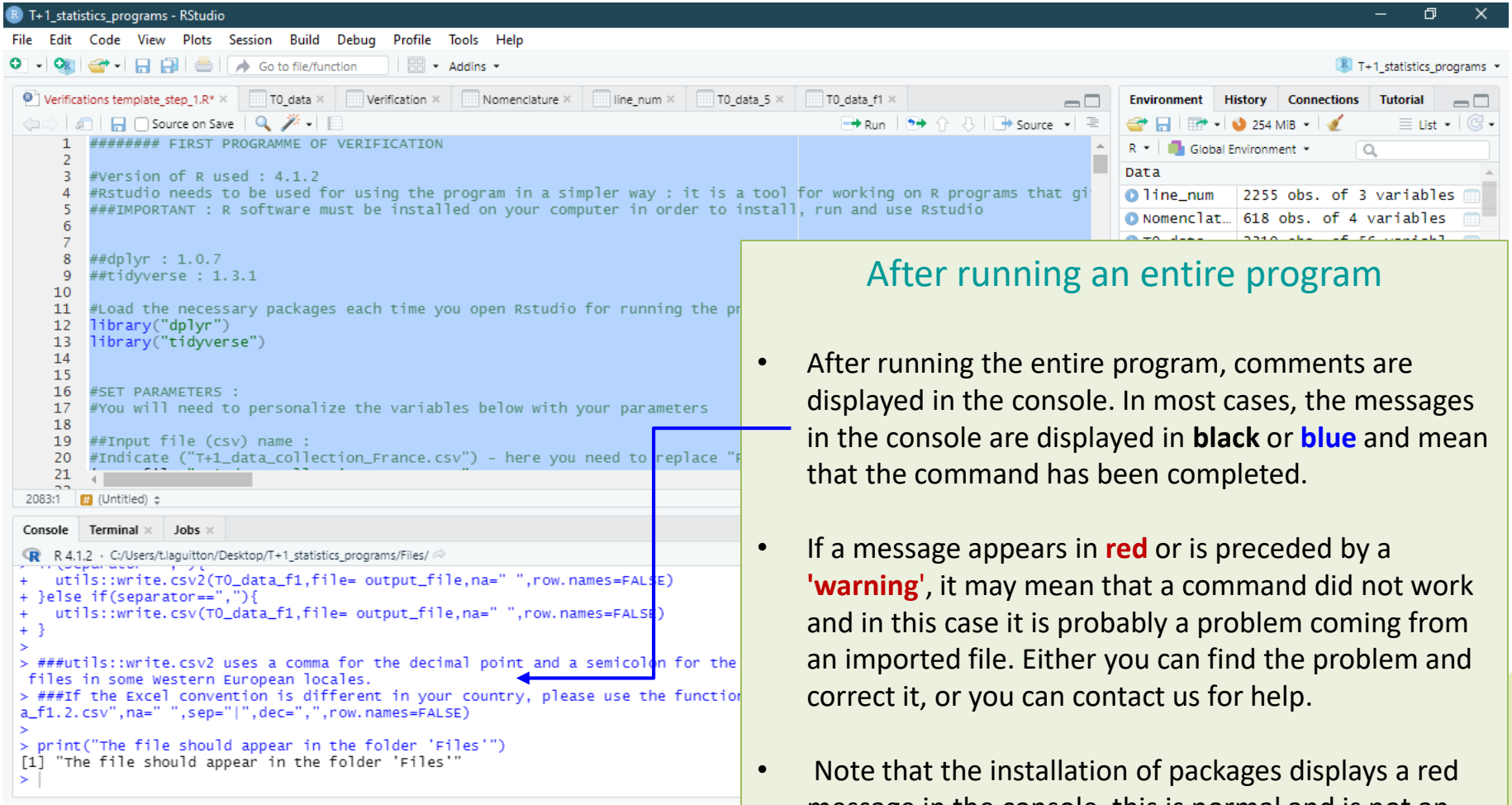
In this case, a **'stop'** button appears at the top left of the console and indicates that the program is currently running (nothing may appear in the console for a few minutes).

During a long program, you can click on this **'stop'** button at any time to suspend the program.
To restart the program, you will need to reselect the entire program script and click on **'Run'** again

Not recommended

Co-funded by the European Union's Health Programme (2014-2020)

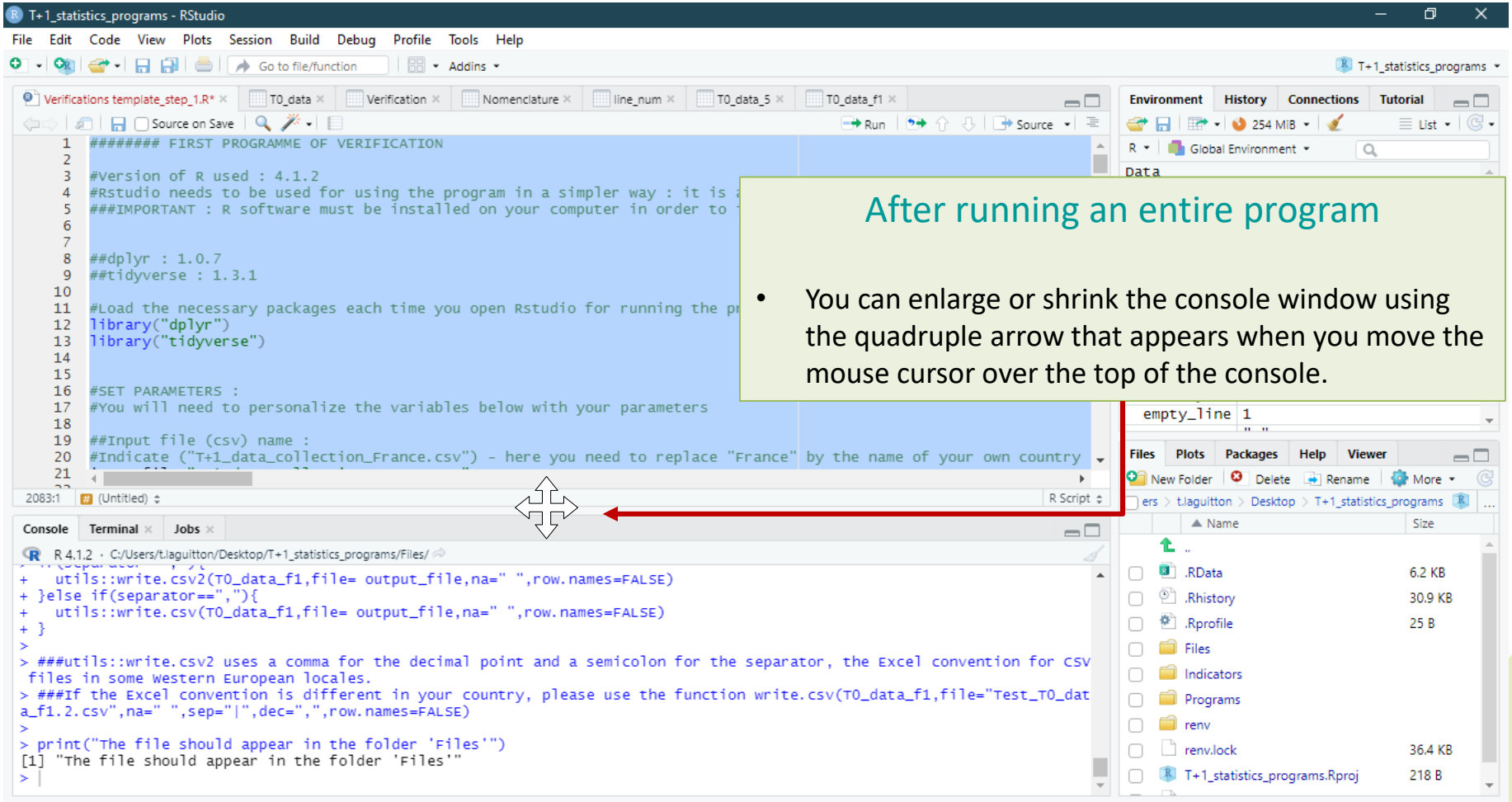
Introduction to Rstudio software



After running an entire program

- After running the entire program, comments are displayed in the console. In most cases, the messages in the console are displayed in **black** or **blue** and mean that the command has been completed.
- If a message appears in **red** or is preceded by a **'warning'**, it may mean that a command did not work and in this case it is probably a problem coming from an imported file. Either you can find the problem and correct it, or you can contact us for help.
- Note that the installation of packages displays a red message in the console, this is normal and is not an error.

Introduction to Rstudio software

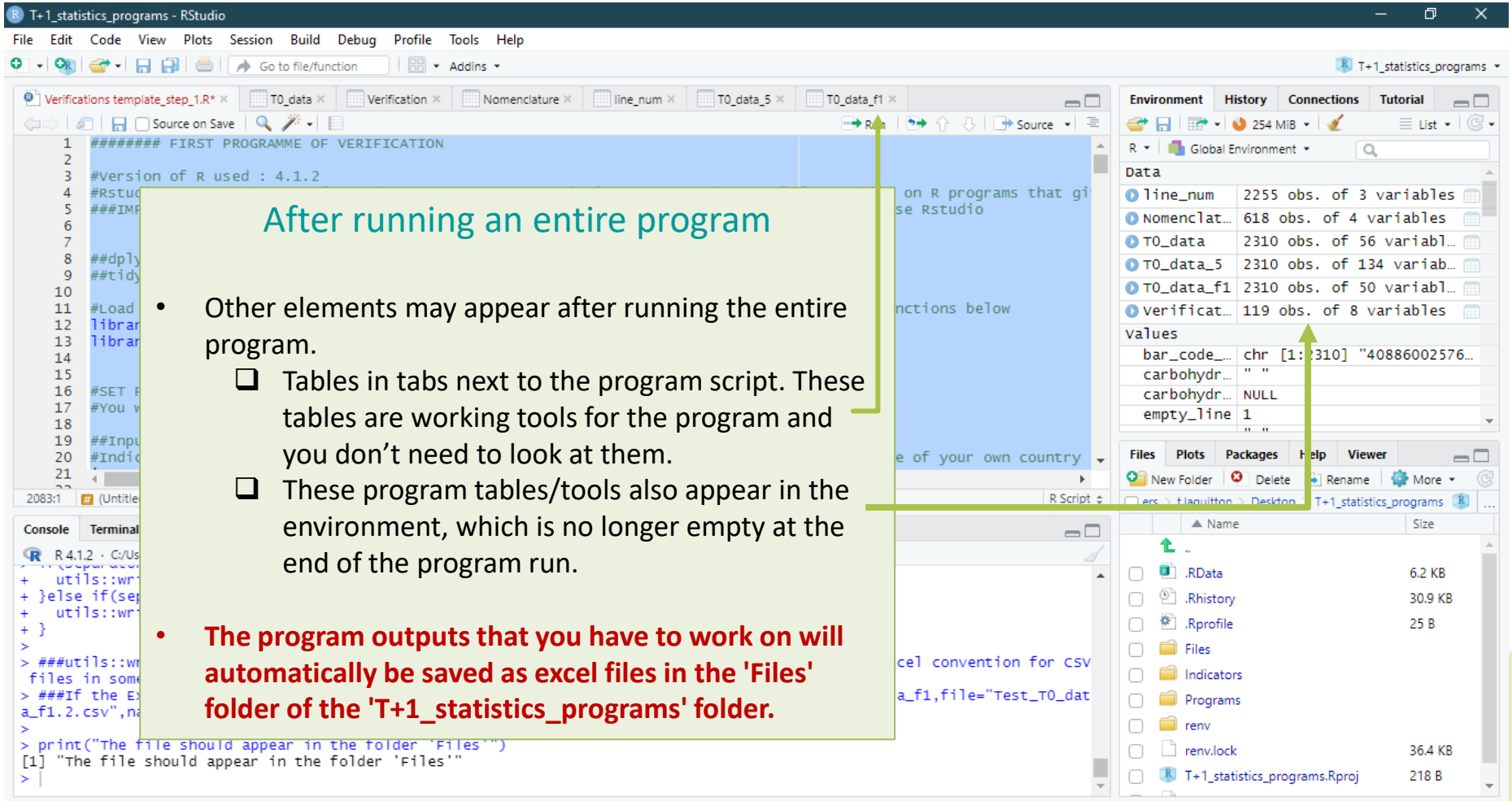


After running an entire program

- You can enlarge or shrink the console window using the quadruple arrow that appears when you move the mouse cursor over the top of the console.

Co-funded by the European Union's Health Programme (2014-2020)

Introduction to Rstudio software



After running an entire program

- Other elements may appear after running the entire program.
  - ❑ Tables in tabs next to the program script. These tables are working tools for the program and you don't need to look at them.
  - ❑ These program tables/tools also appear in the environment, which is no longer empty at the end of the program run.

- **The program outputs that you have to work on will automatically be saved as excel files in the 'Files' folder of the 'T+1_statistics_programs' folder.**

Co-funded by the European Union's Health Programme (2014-2020)

2) Installation of the necessary equipment and presentation of the Rstudio software

A. Preliminary steps

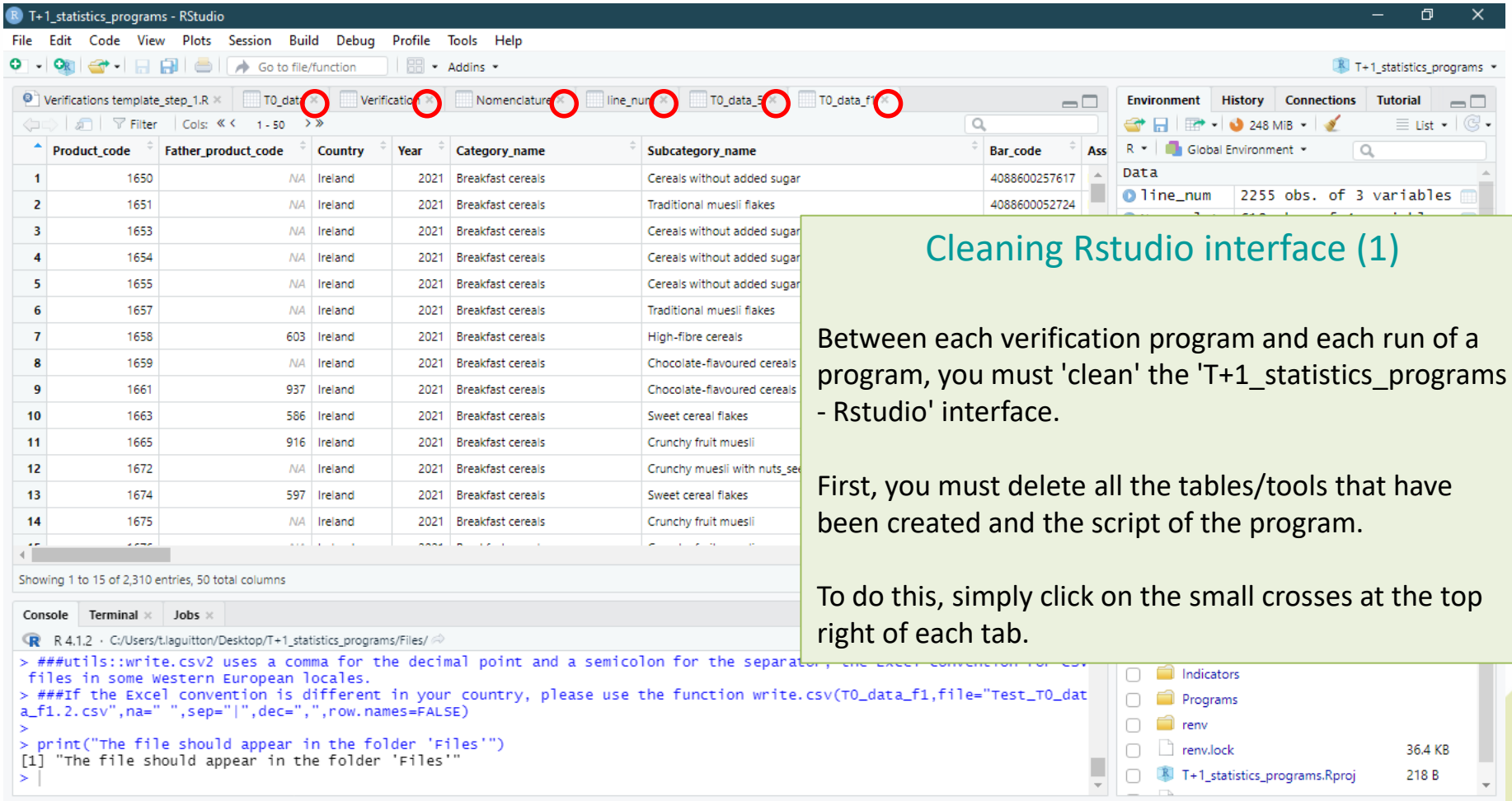
B. Installation of software

C. Introduction to R studio

D. Cleaning of the Rstudio interface

Co-funded by the European Union's Health Programme (2014-2020)

‘Cleaning’ of Rstudio between each program



Cleaning Rstudio interface (1)

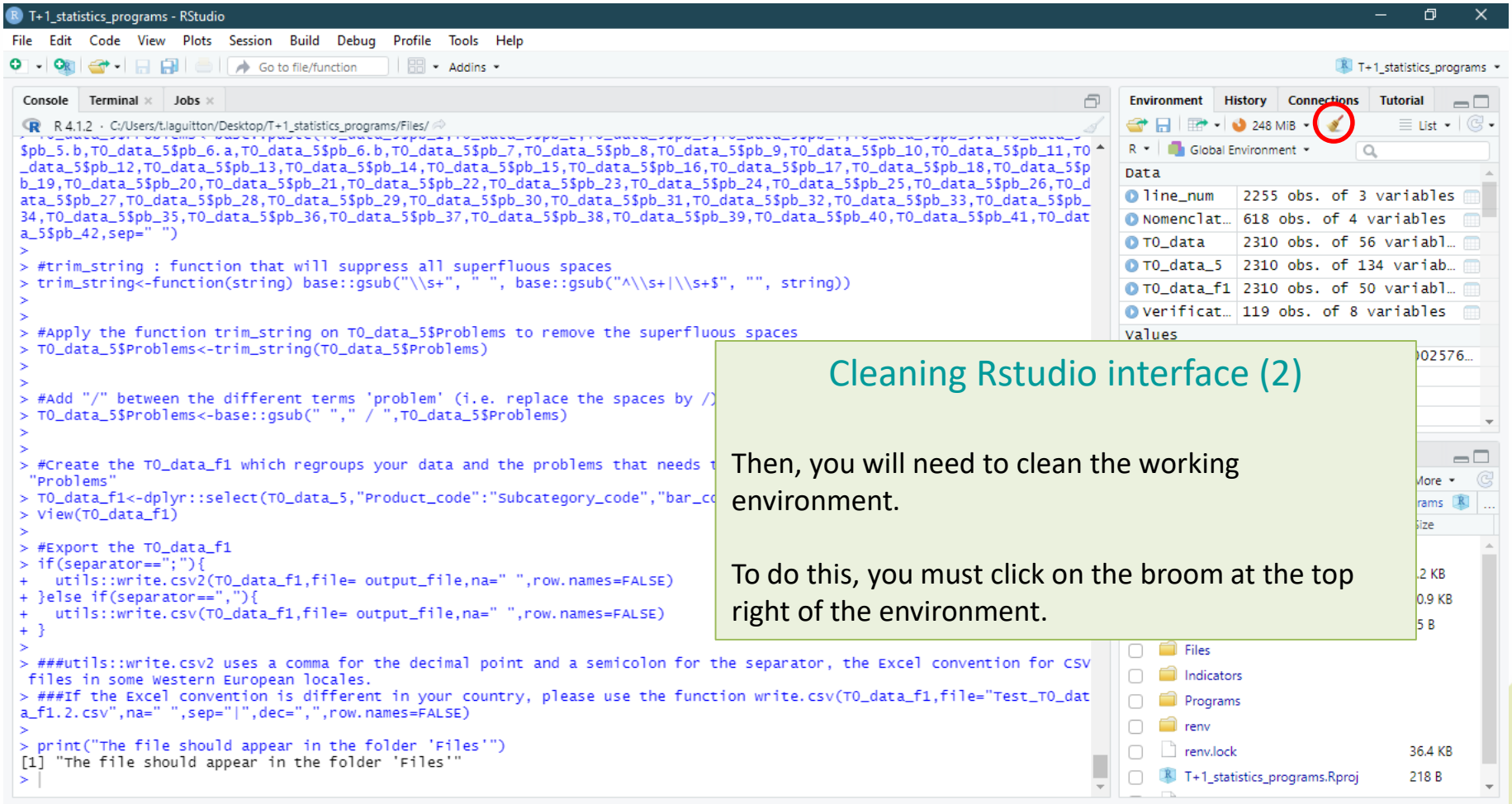
Between each verification program and each run of a program, you must 'clean' the 'T+1_statistics_programs - Rstudio' interface.

First, you must delete all the tables/tools that have been created and the script of the program.

To do this, simply click on the small crosses at the top right of each tab.

Co-funded by the European Union's Health Programme (2014-2020)

‘Cleaning’ of Rstudio between each program



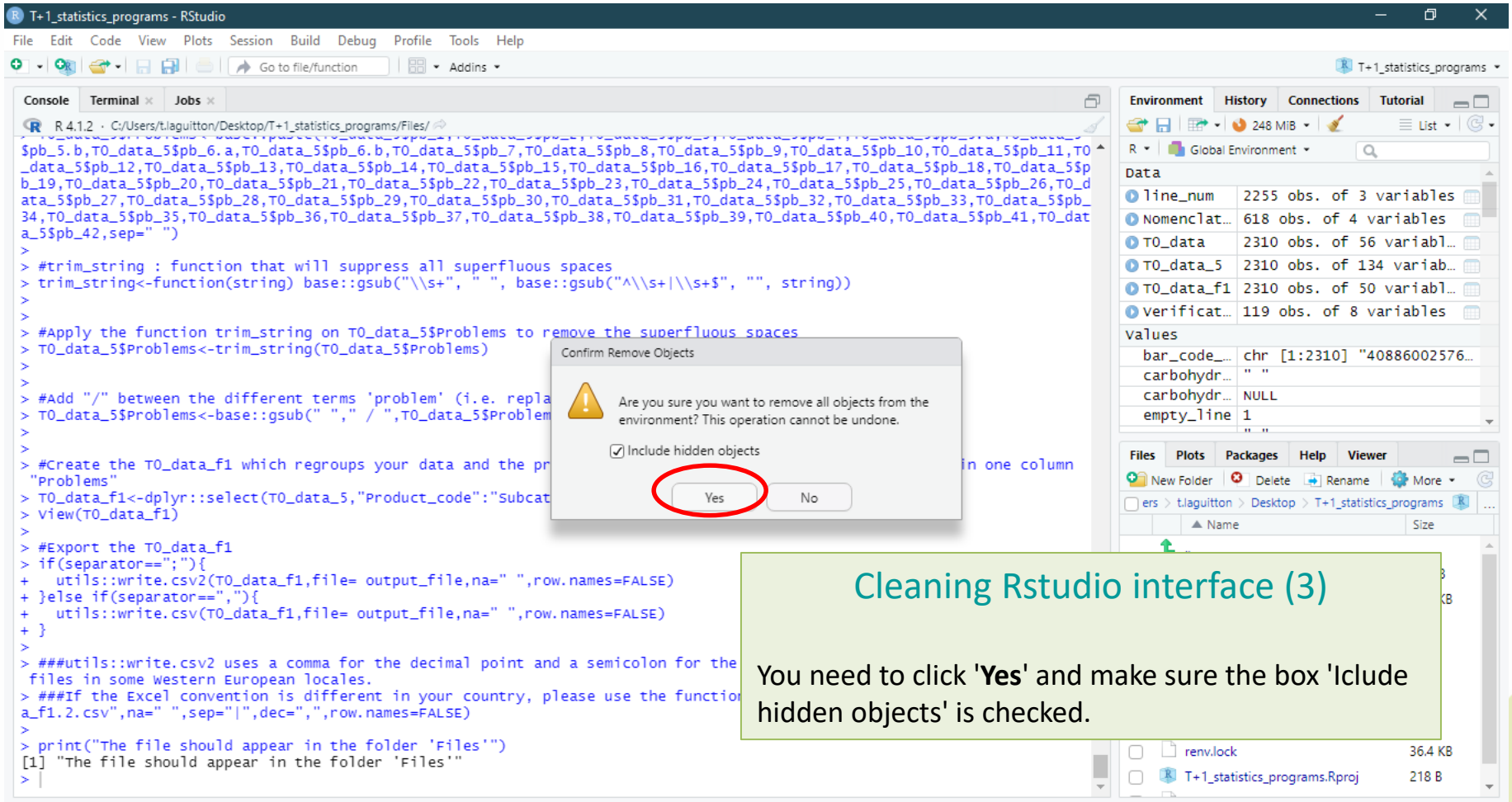
Cleaning Rstudio interface (2)

Then, you will need to clean the working environment.

To do this, you must click on the broom at the top right of the environment.

Co-funded by the European Union's Health Programme (2014-2020)

‘Cleaning’ of Rstudio between each program

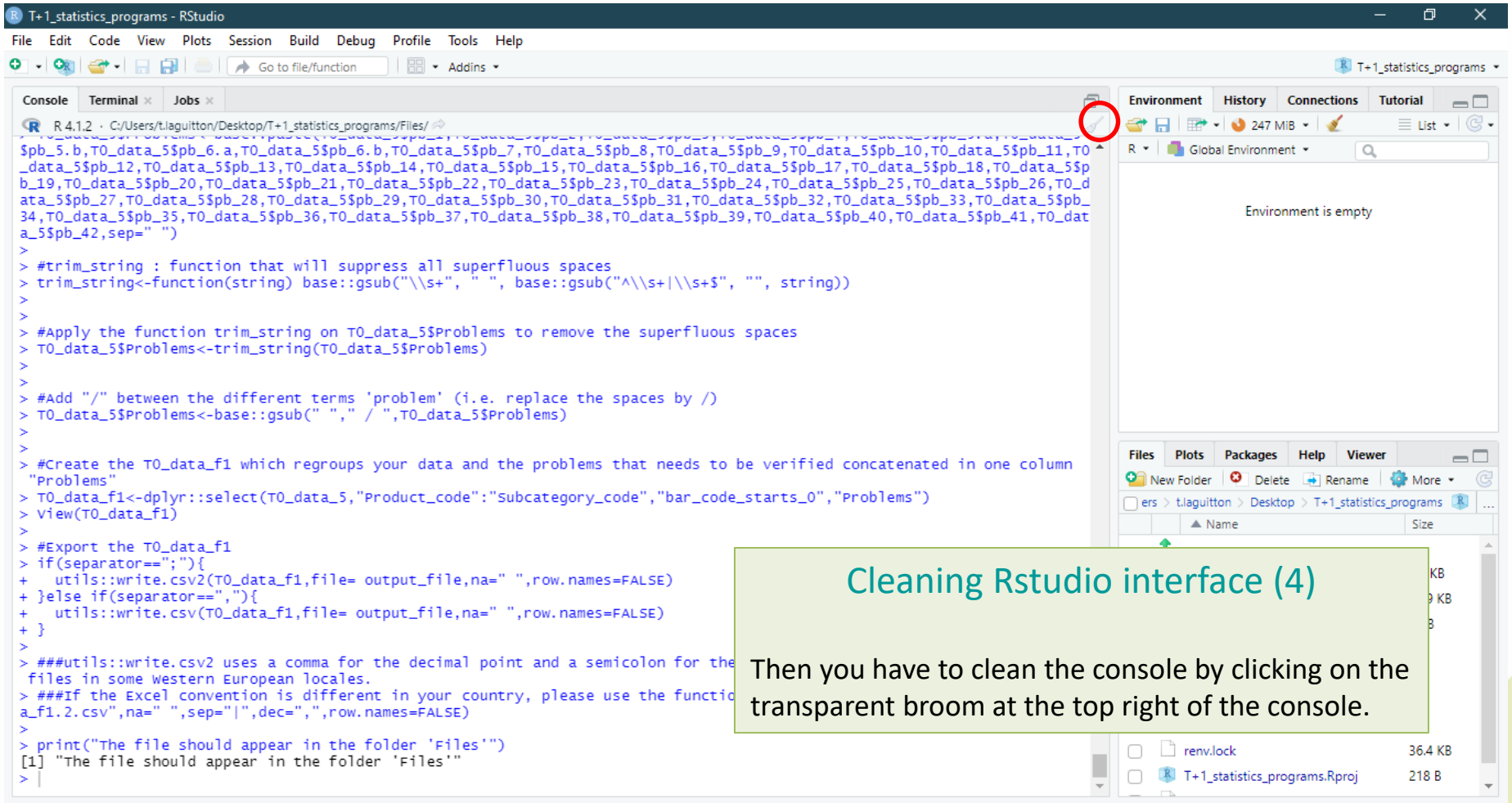


Cleaning Rstudio interface (3)

You need to click '**Yes**' and make sure the box 'Iclude hidden objects' is checked.

Co-funded by the European Union's Health Programme (2014-2020)

‘Cleaning’ of Rstudio between each program

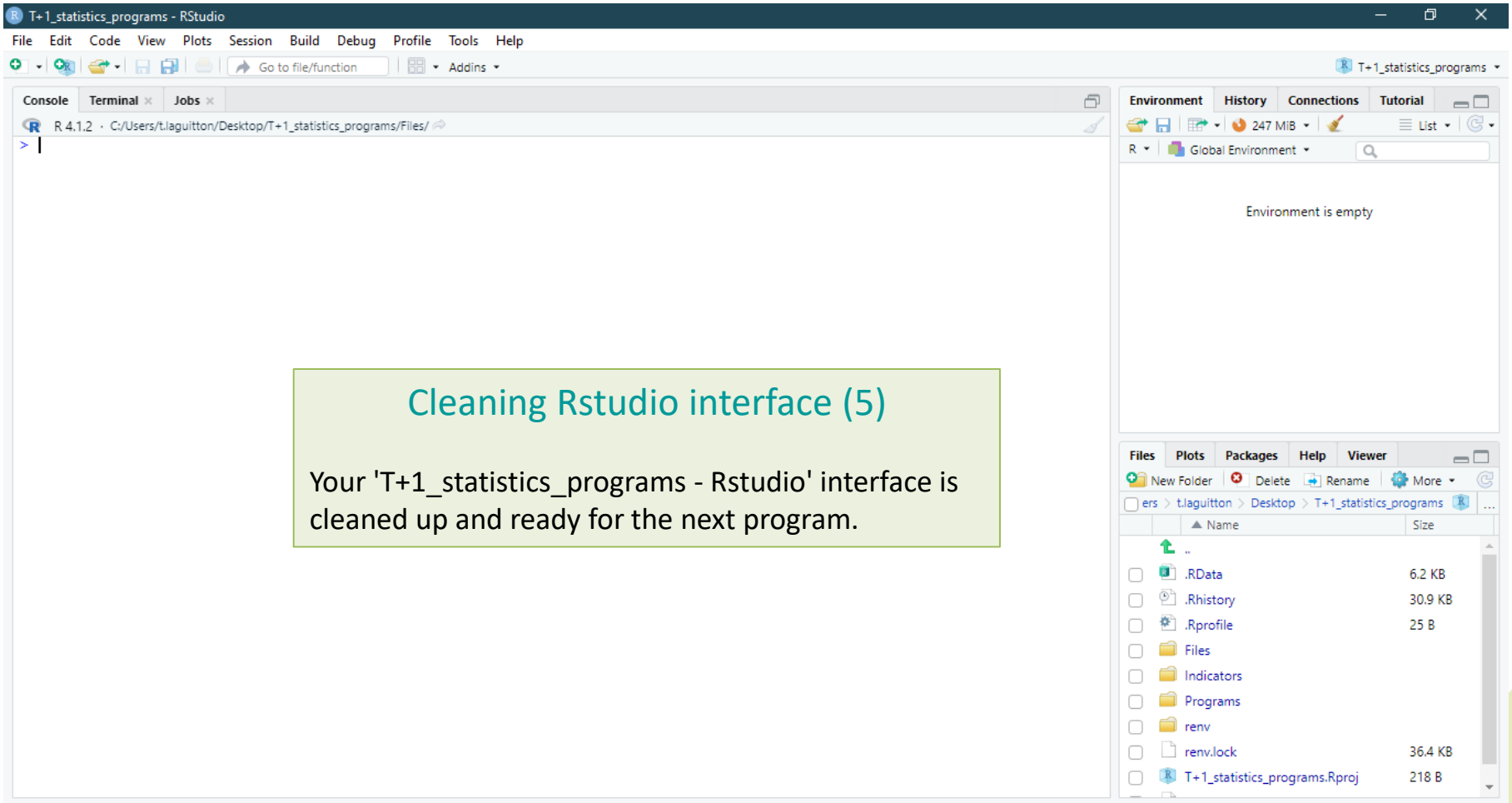


Cleaning Rstudio interface (4)

Then you have to clean the console by clicking on the transparent broom at the top right of the console.

Co-funded by the European Union's Health Programme (2014-2020)

‘Cleaning’ of Rstudio between each program

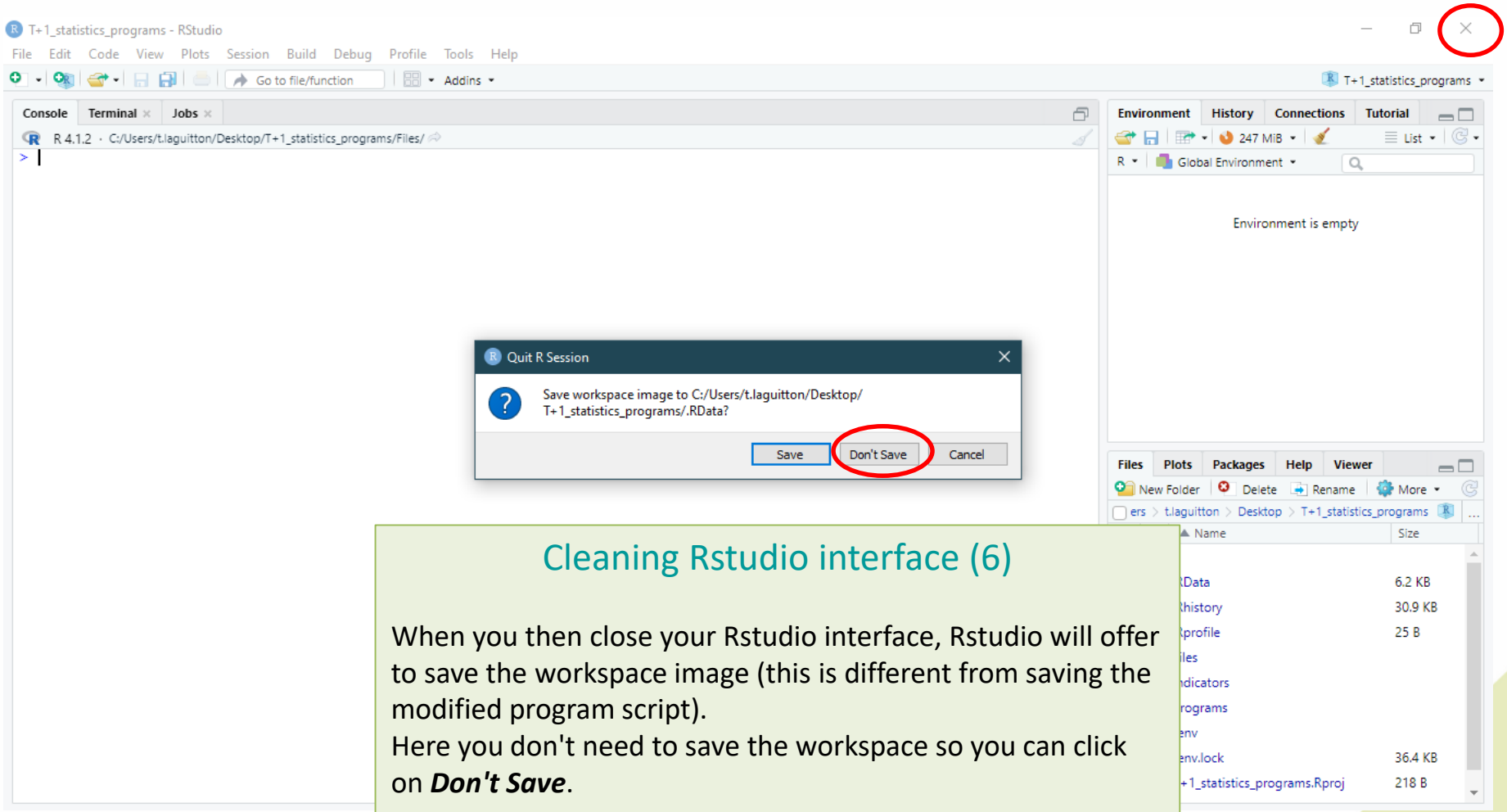


Cleaning Rstudio interface (5)

Your 'T+1_statistics_programs - Rstudio' interface is cleaned up and ready for the next program.

Co-funded by the European Union's Health Programme (2014-2020)

‘Cleaning’ of Rstudio between each program



Cleaning Rstudio interface (6)

When you then close your Rstudio interface, Rstudio will offer to save the workspace image (this is different from saving the modified program script).
Here you don't need to save the workspace so you can click on ***Don't Save***.

Co-funded by the European Union's Health Programme (2014-2020)

3) Running of the programs

A. Part 1 : R setup program (page 78)

B. Part 2 : Verification programs and template cleaning/standardization (page 88)

i. 1st verification program : 'Verifications template_step_1' (page 90)

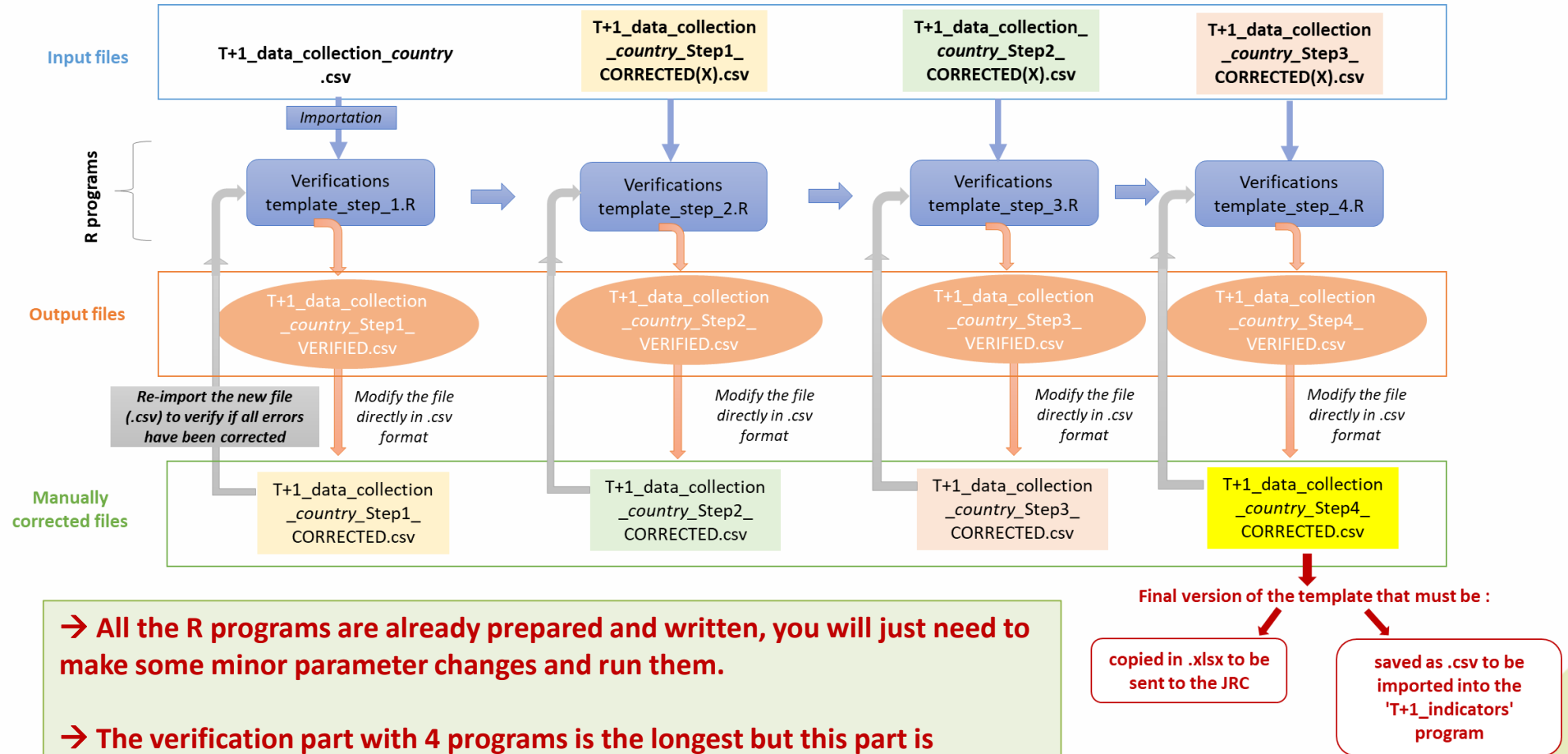
ii. 2nd verification program : 'Verifications template_step_2' (page 119)

iii. 3rd verification program : 'Verifications template_step_3' (page 144)

iv. 4th verification program : 'Verifications template_step_4' (page 166)

Co-funded by the European Union's Health Programme (2014-2020)

Description of the different 'R' programs



Final version of the template that must be :

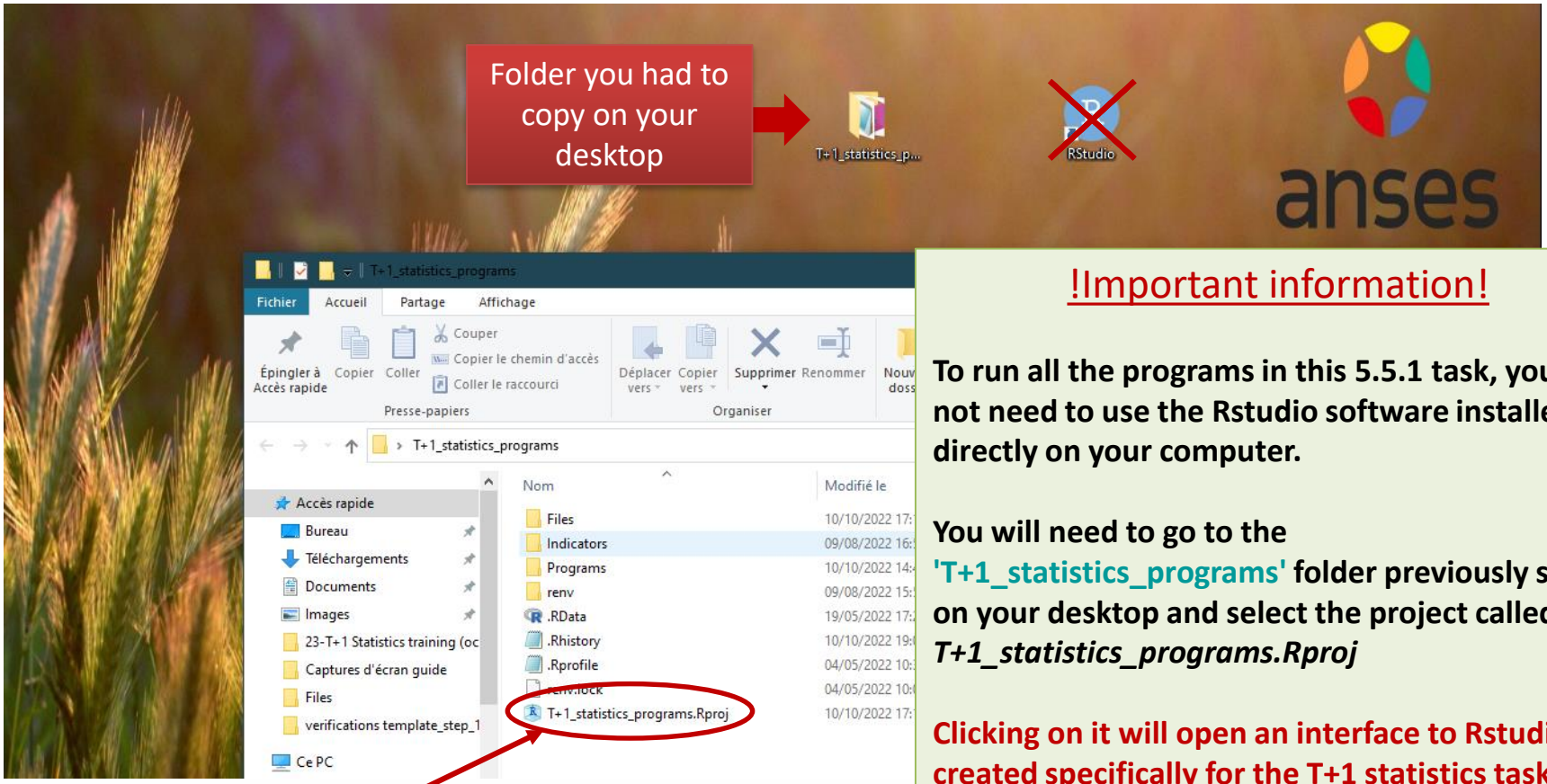
- copied in .xlsx to be sent to the JRC
- saved as .csv to be imported into the 'T+1_indicators' program

→ All the R programs are already prepared and written, you will just need to make some minor parameter changes and run them.

→ The verification part with 4 programs is the longest but this part is essential to ensure the reliability of the indicators that will come out of the last program 'T+1_indicators' (next part of the training).

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use



Rstudio interface created specifically for the project that you always have to work with

!Important information!

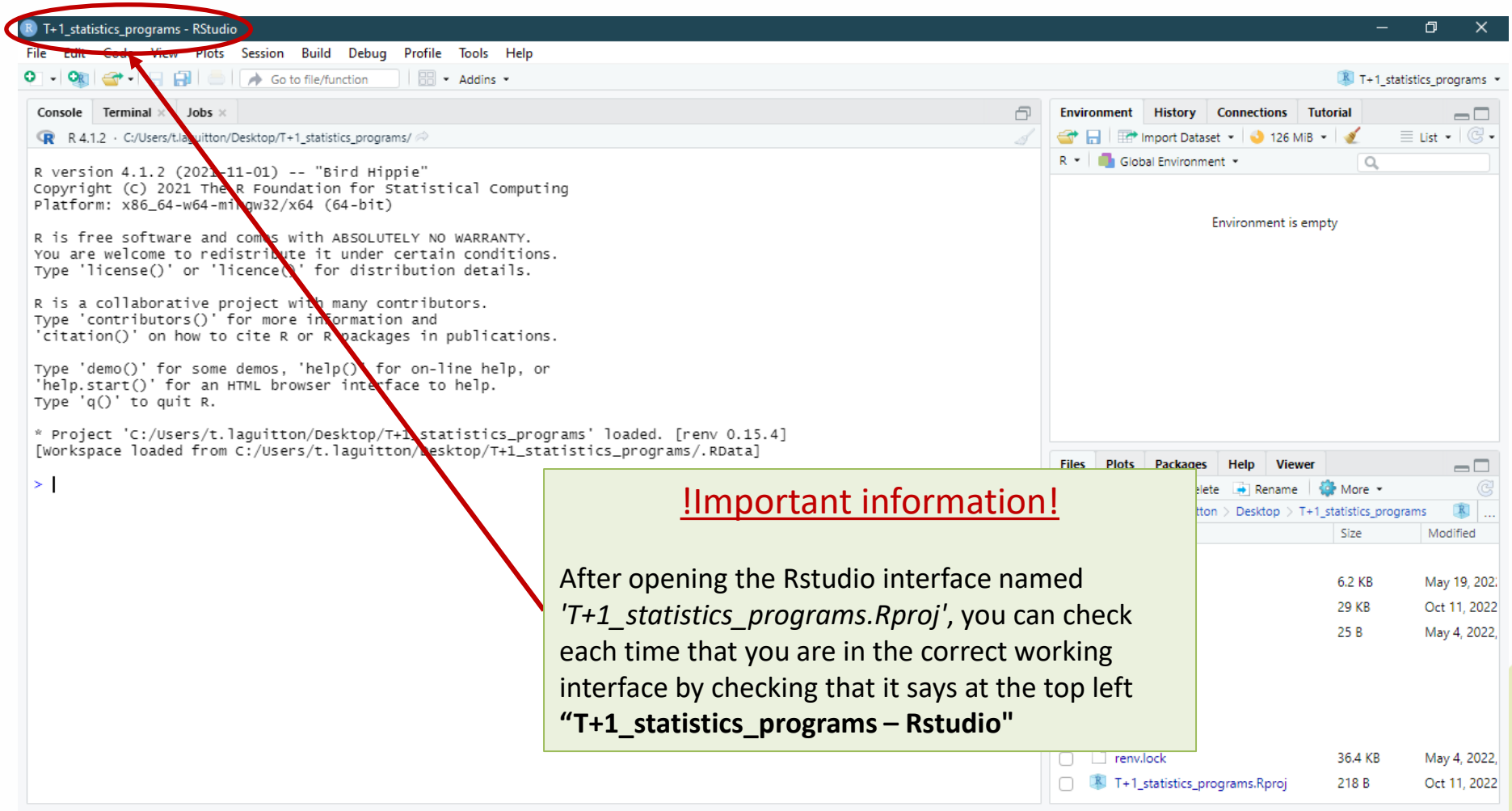
To run all the programs in this 5.5.1 task, you will not need to use the Rstudio software installed directly on your computer.

You will need to go to the 'T+1_statistics_programs' folder previously saved on your desktop and select the project called *T+1_statistics_programs.Rproj*

Clicking on it will open an interface to Rstudio created specifically for the T+1 statistics task. It is very important that you run all your programs on this Rstudio interface.

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use

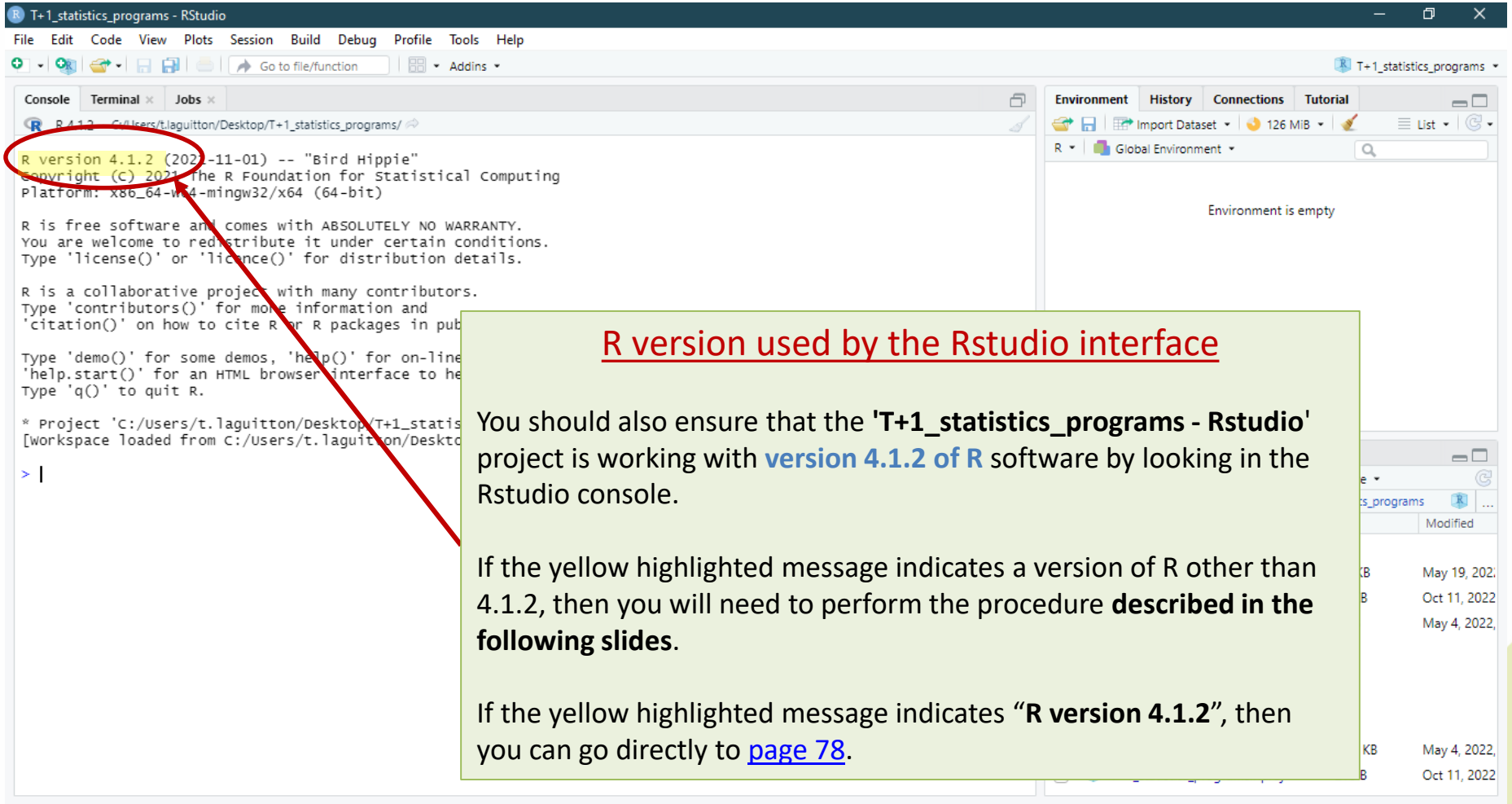


!Important information!

After opening the Rstudio interface named *'T+1_statistics_programs.Rproj'*, you can check each time that you are in the correct working interface by checking that it says at the top left **"T+1_statistics_programs – Rstudio"**

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use



R version used by the Rstudio interface

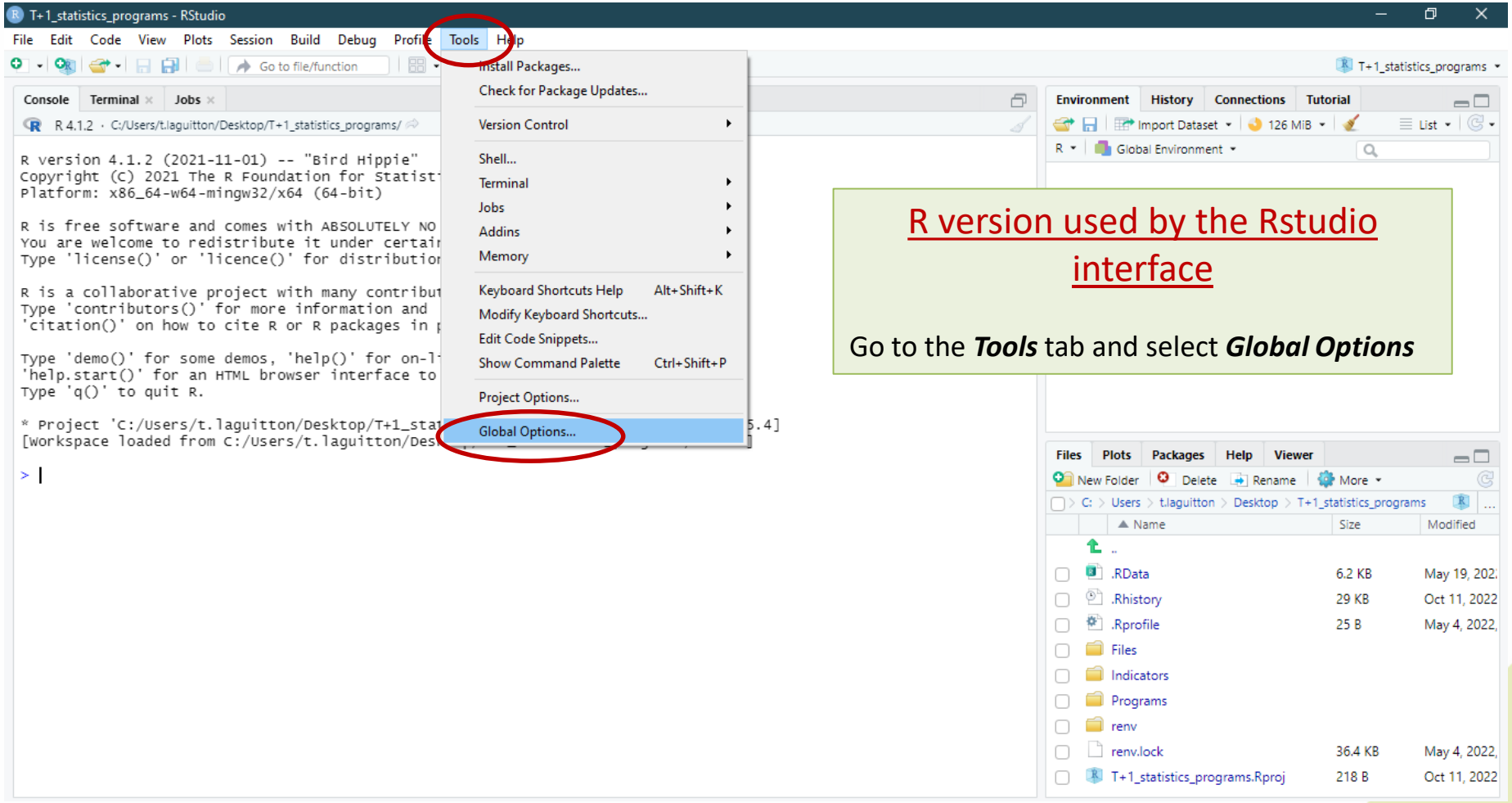
You should also ensure that the **'T+1_statistics_programs - Rstudio'** project is working with **version 4.1.2 of R** software by looking in the Rstudio console.

If the yellow highlighted message indicates a version of R other than 4.1.2, then you will need to perform the procedure **described in the following slides**.

If the yellow highlighted message indicates "**R version 4.1.2**", then you can go directly to page 78.

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use

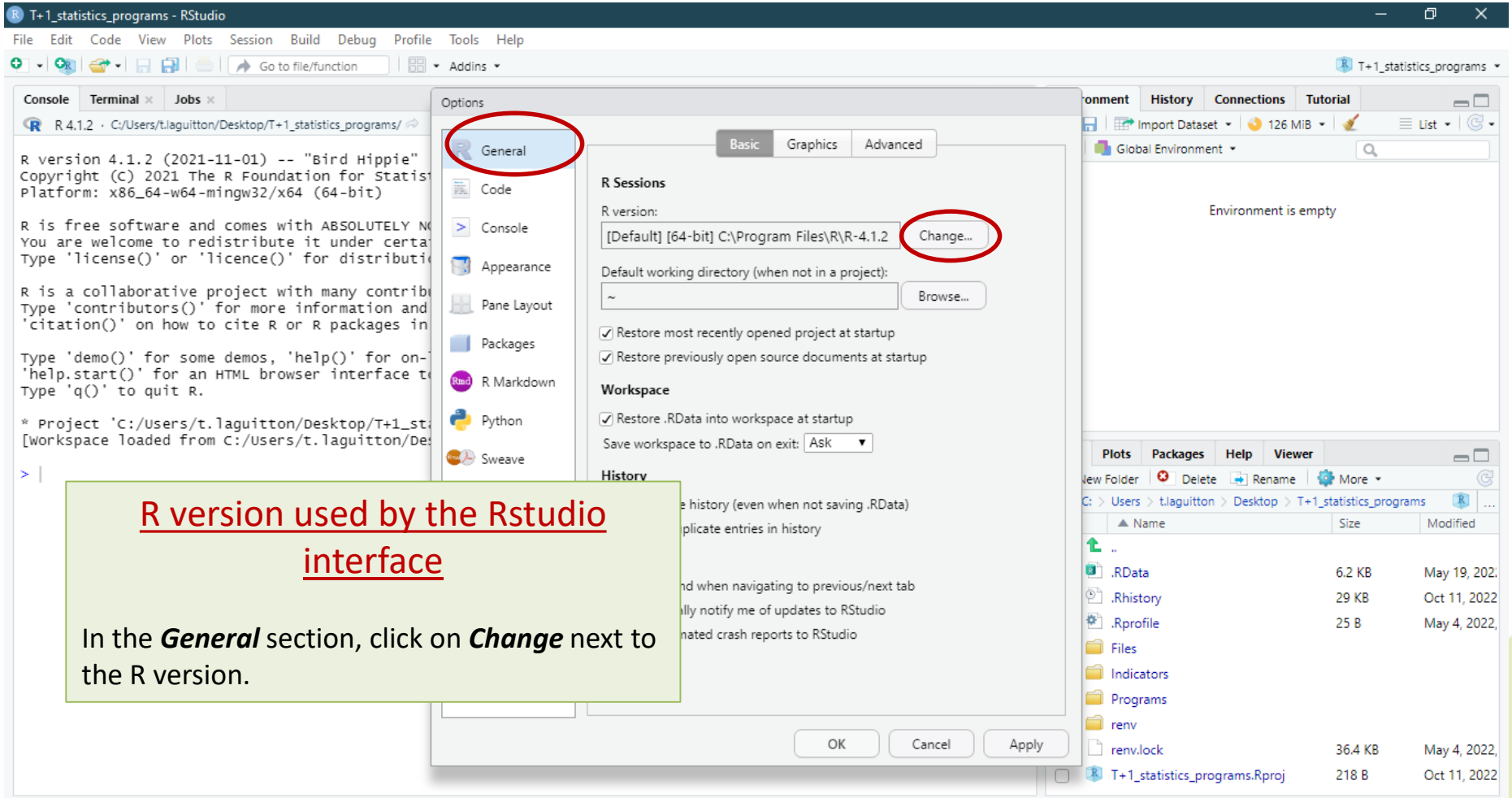


R version used by the Rstudio interface

Go to the ***Tools*** tab and select ***Global Options***

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use

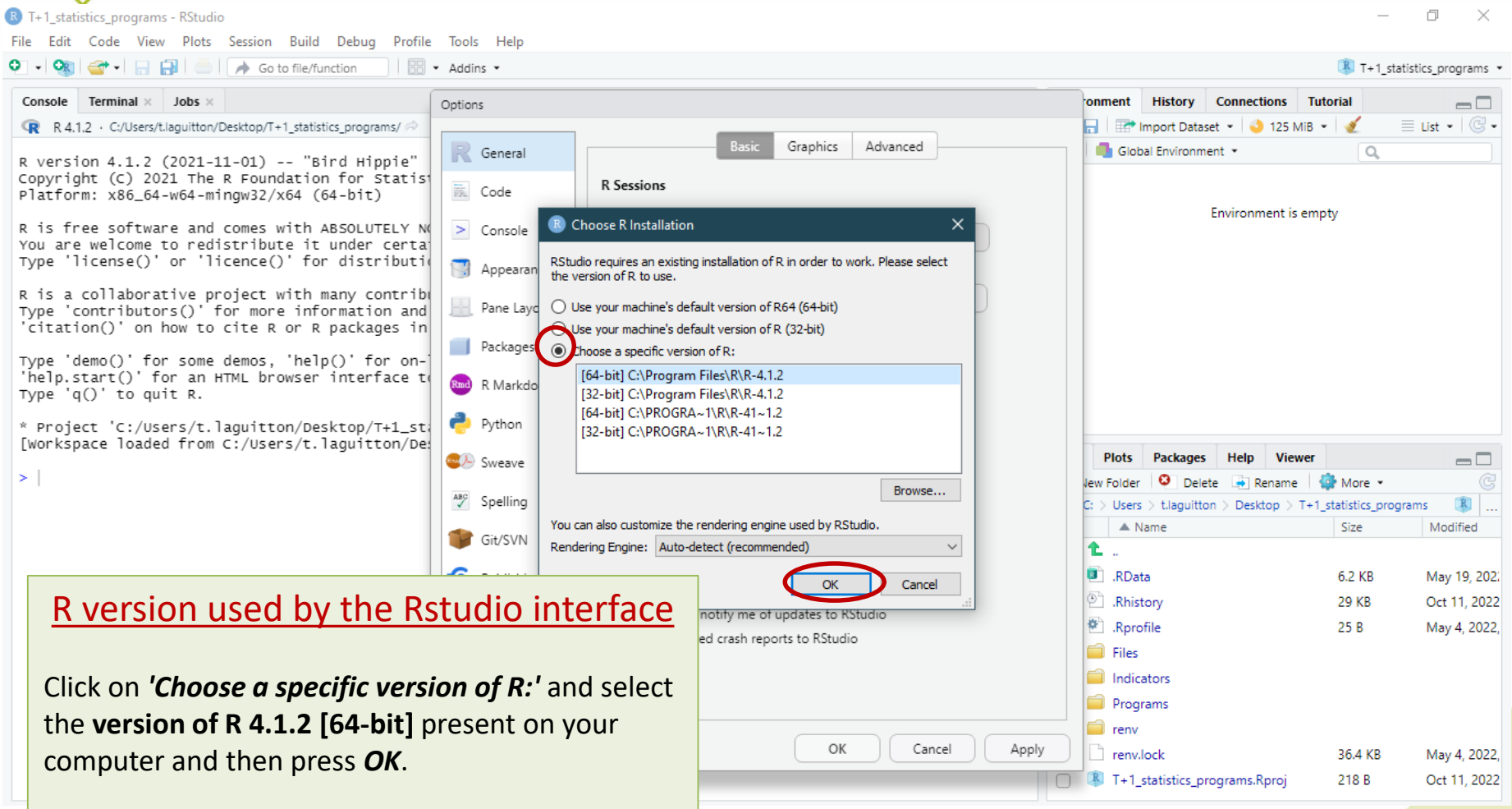


R version used by the Rstudio interface

In the ***General*** section, click on ***Change*** next to the R version.

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use



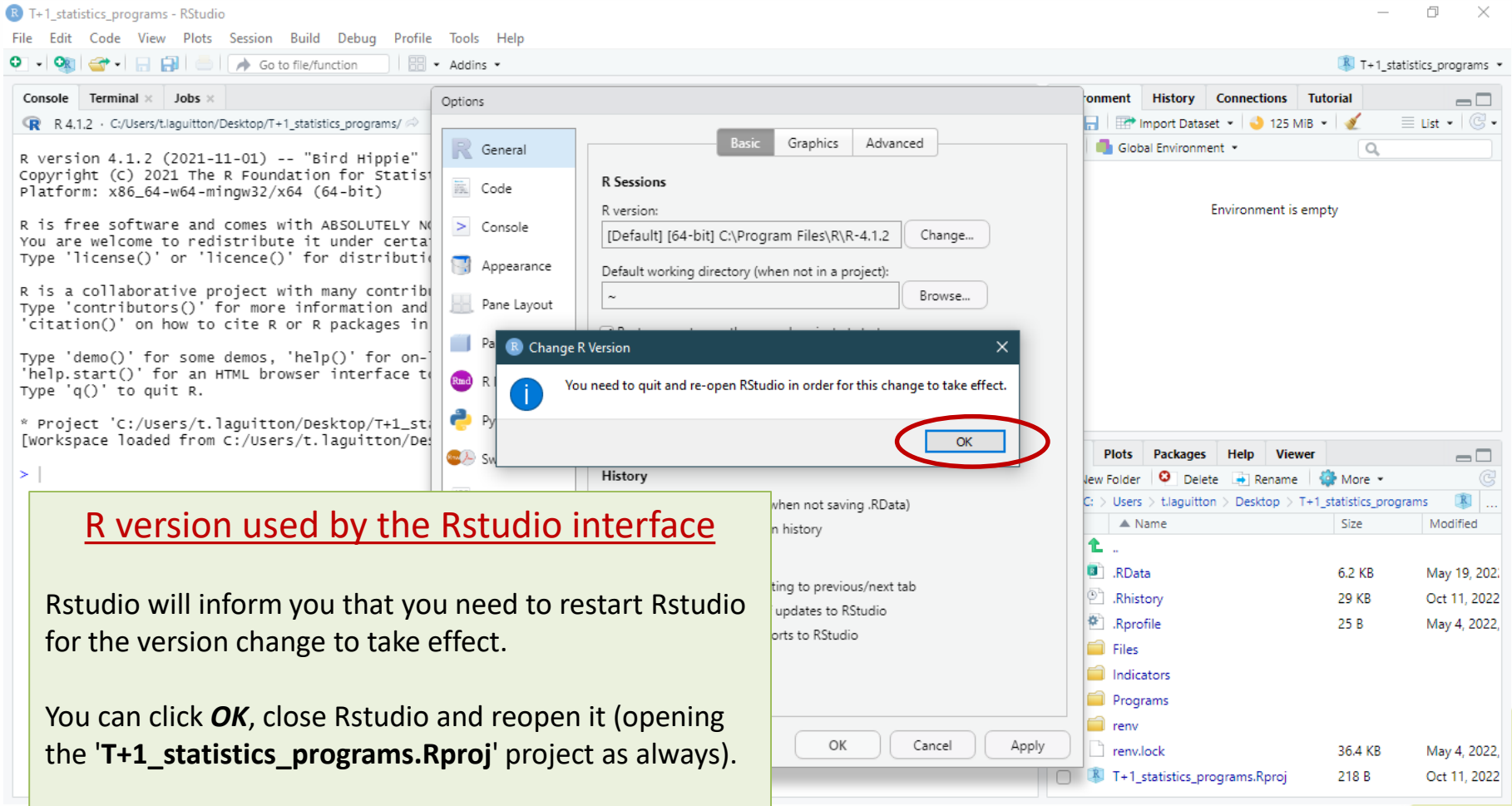
R version used by the Rstudio interface

Click on ***'Choose a specific version of R:'*** and select the **version of R 4.1.2 [64-bit]** present on your computer and then press ***OK***.

(If you have any doubts about which version of R to choose at this point as there are several 4.1.2 versions on offer, please feel free to send us a screenshot of the proposals made and we will help you choose)

Co-funded by the European Union's Health Programme (2014-2020)

Rstudio interface to use



R version used by the Rstudio interface

Rstudio will inform you that you need to restart Rstudio for the version change to take effect.

You can click ***OK***, close Rstudio and reopen it (opening the '**T+1_statistics_programs.Rproj**' project as always).

Version 4.1.2 will be now displayed in the console and you will be able to proceed to the next steps described in the following slides.

3) Running of the programs

A. Part 1 : R setup program

B. Part 2 : Verification programs and template cleaning/standardization

i. 1st verification program : 'Verifications template_step_1'

ii. 2nd verification program : 'Verifications template_step_2'

iii. 3rd verification program : 'Verifications template_step_3'

C. Part 3 : Indicators and statistics production program

Co-funded by the European Union's Health Programme (2014-2020)

‘R_setup’ program

Presentation of the ‘R_setup' program :

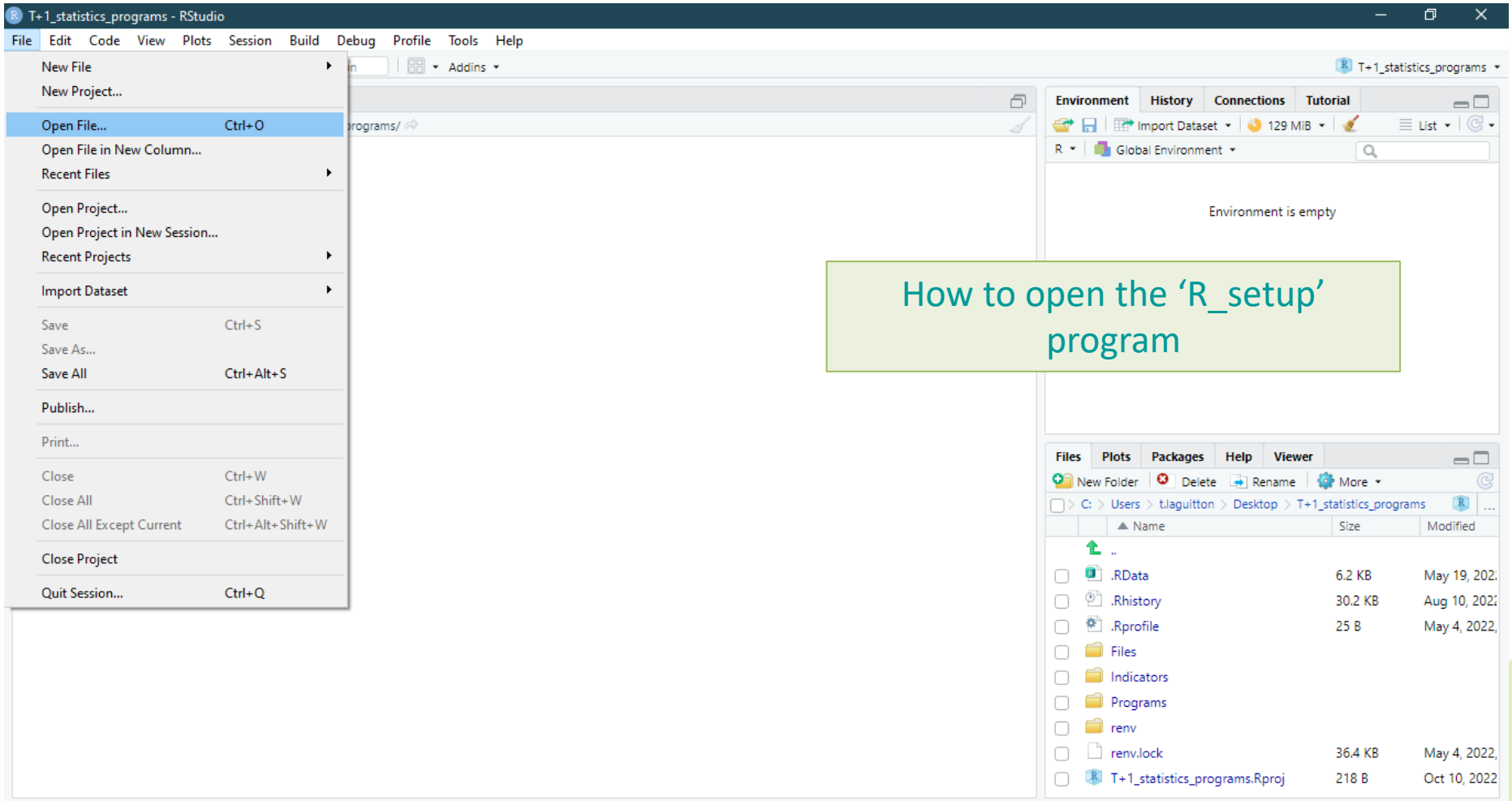
Just after installing the R and R studio software and before starting the verification programs, you will need to run the 'R_setup' program just once. This will allow the installation of packages that contain functions that will be needed for data verification and the creation of indicators.

Requirements before starting the program ‘R_setup’ :

- Before running the ‘R_setup’ program, you must ensure that you have **downloaded** the **R** and **Rstudio** software.
- You must also ensure that you are working on the Rstudio interface called ***T+1_statistics_programs.Rproj*** located in the 'T+1_statistitics_programs' folder on your desktop (see 2nd preliminary step pages 32-33)

Co-funded by the European Union's Health Programme (2014-2020)

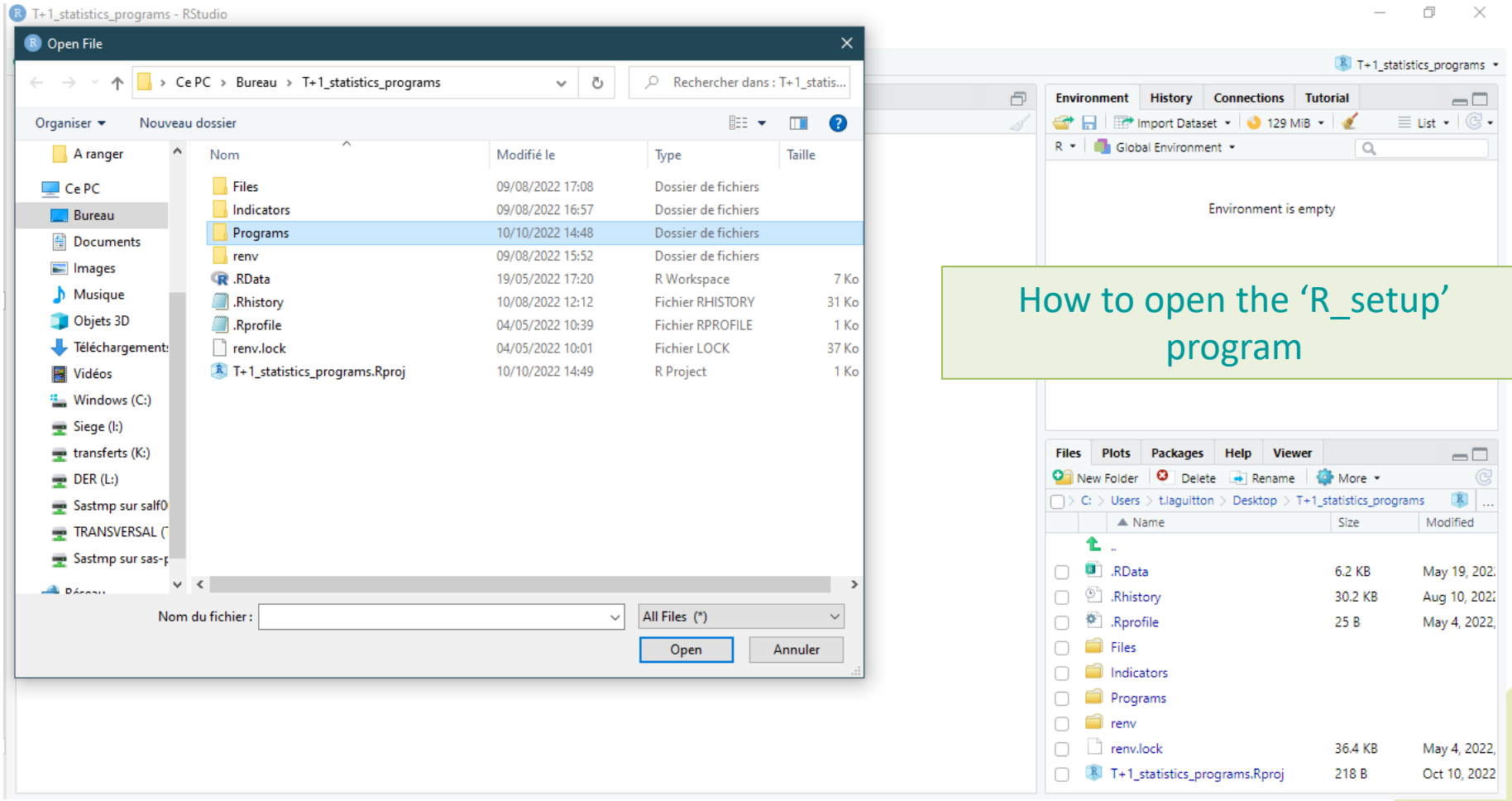
Running the ‘R_setup’ program



How to open the ‘R_setup’ program

Co-funded by the European Union's Health Programme (2014-2020)

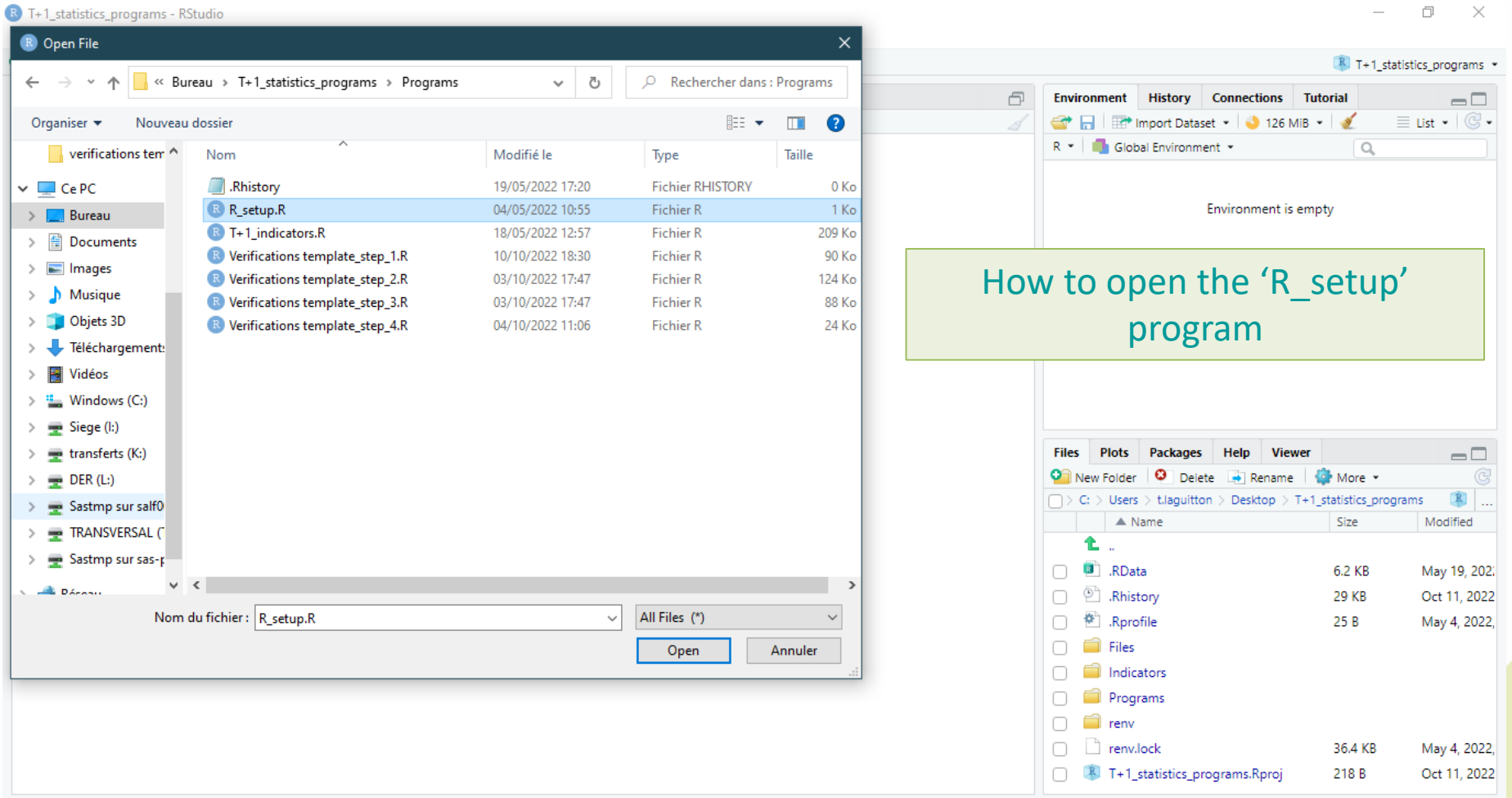
Running the ‘R_setup’ program



How to open the ‘R_setup’ program

Co-funded by the European Union's Health Programme (2014-2020)

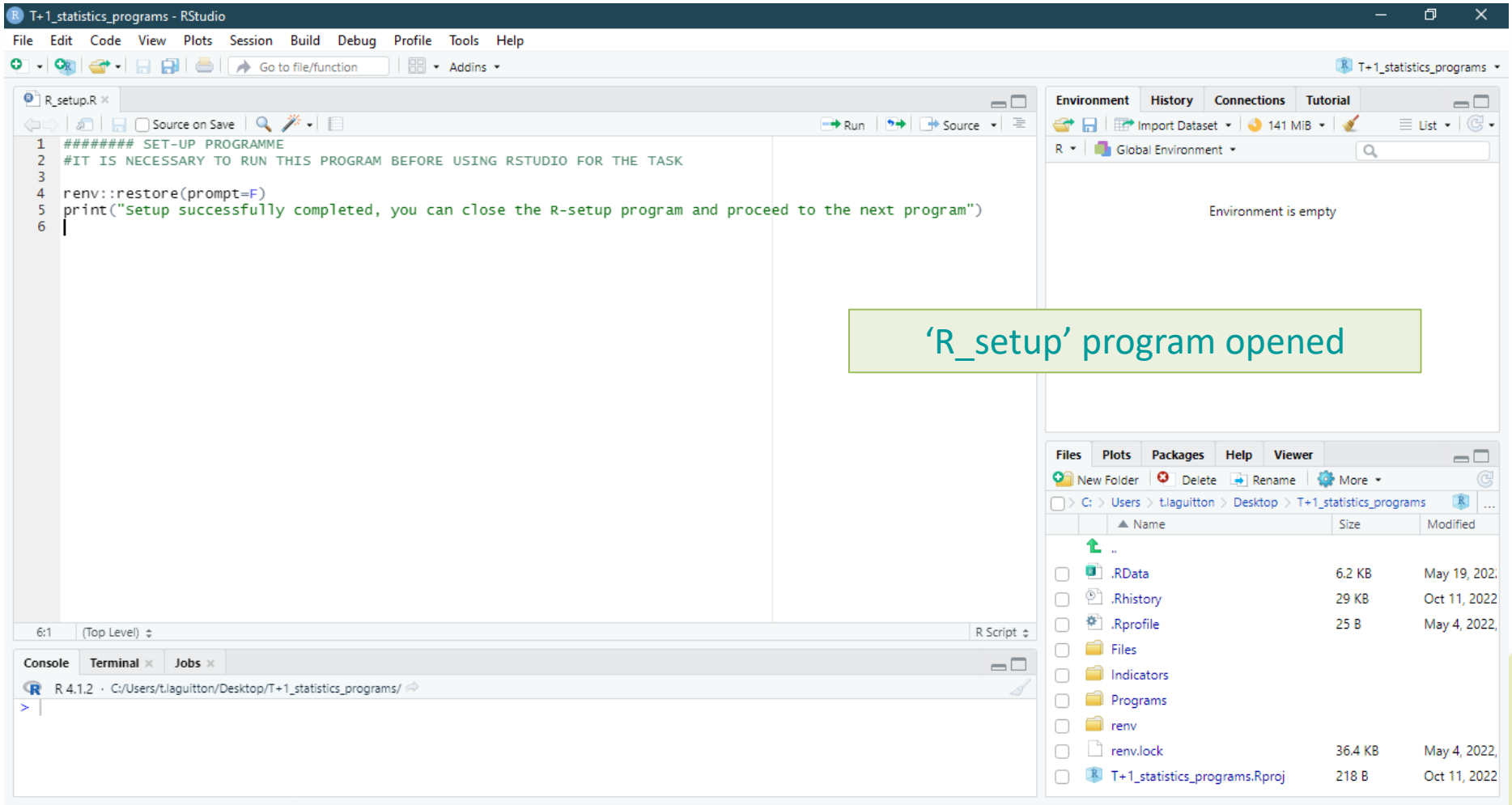
Running the ‘R_setup’ program



How to open the ‘R_setup’ program

Co-funded by the European Union's Health Programme (2014-2020)

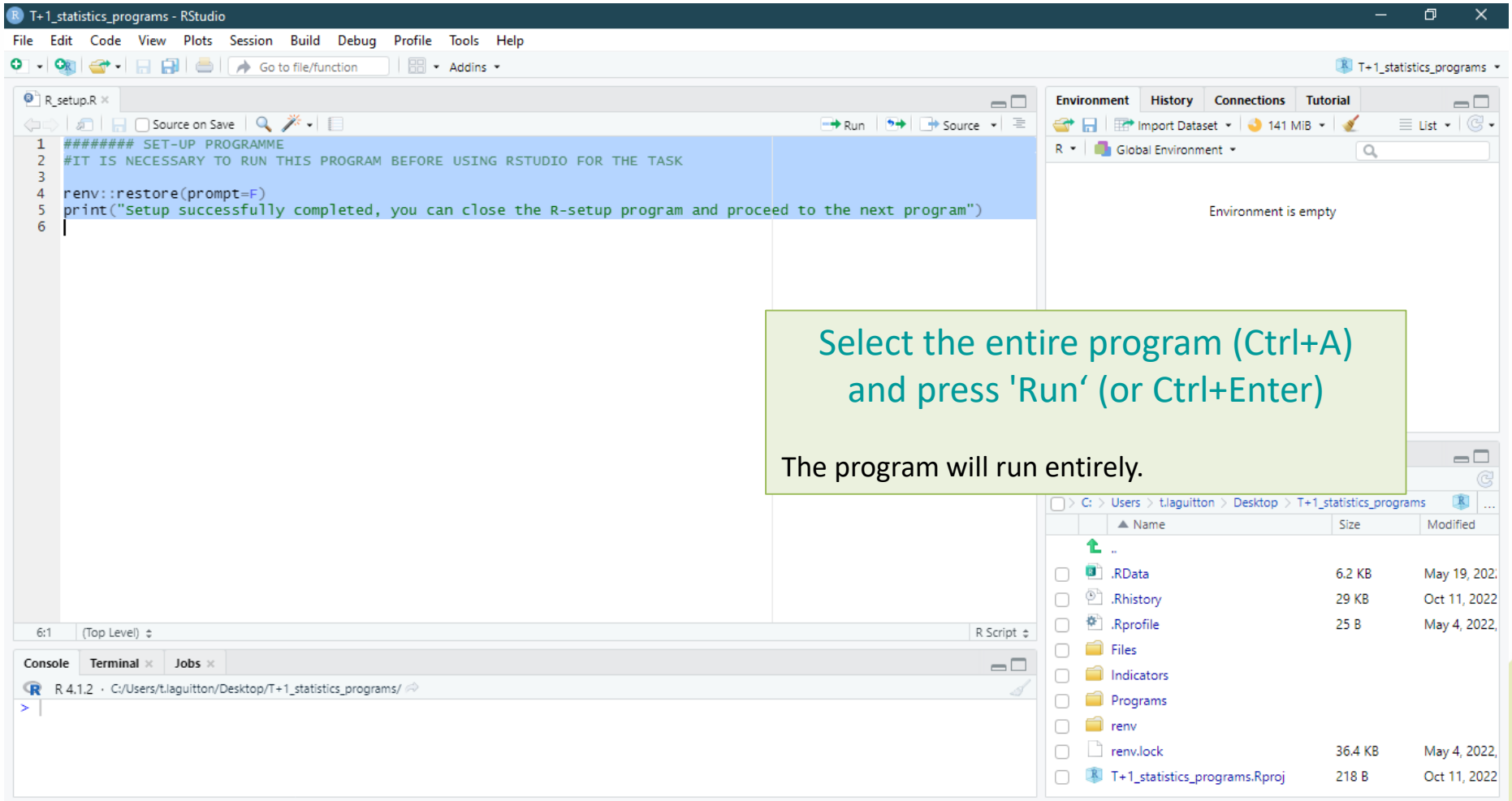
Running the ‘R_setup’ program



‘R_setup’ program opened

Co-funded by the European Union's Health Programme (2014-2020)

Running the 'R_setup' program

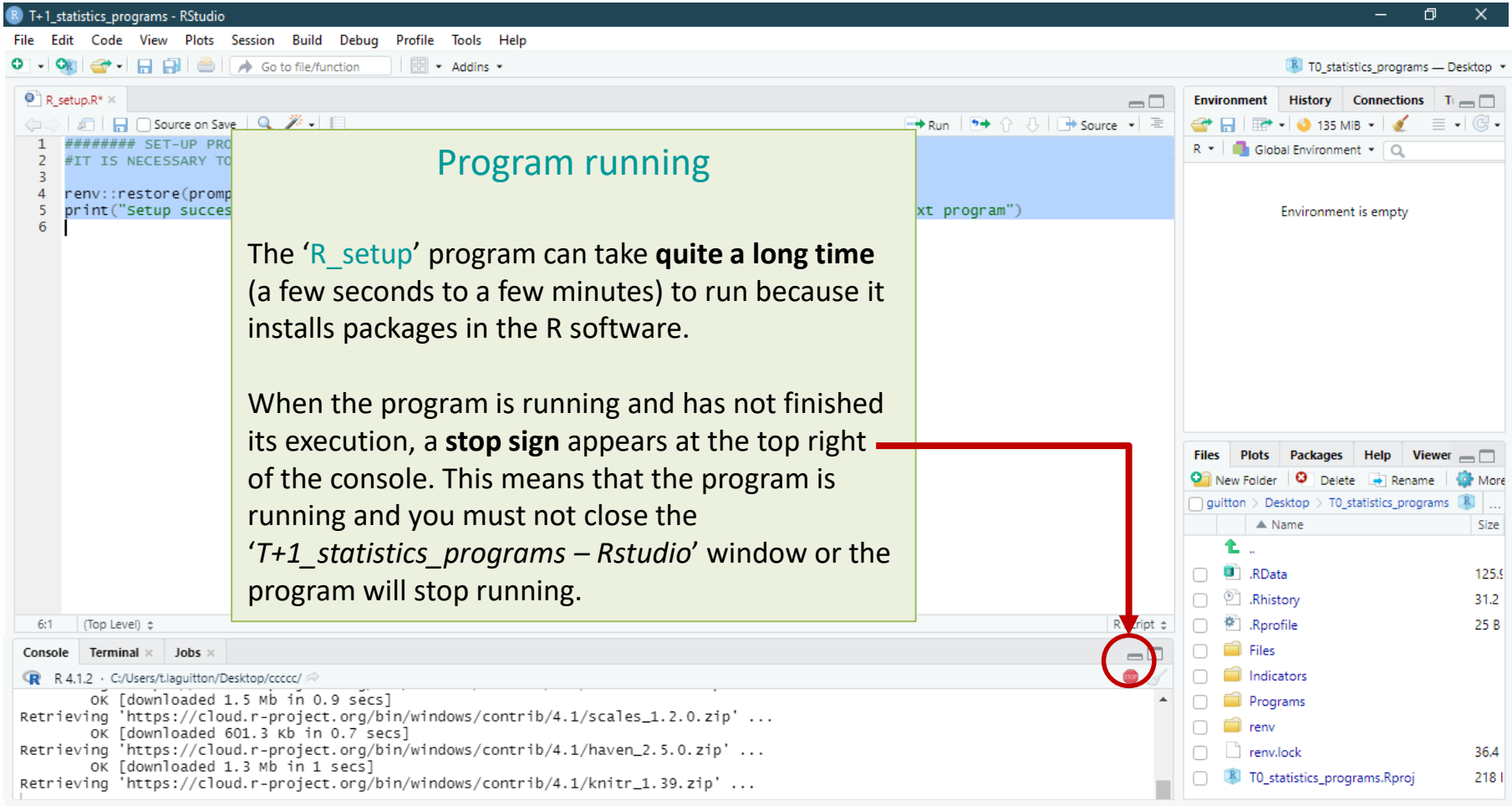


Select the entire program (Ctrl+A) and press 'Run' (or Ctrl+Enter)

The program will run entirely.

Co-funded by the European Union's Health Programme (2014-2020)

Running the 'R_setup' program



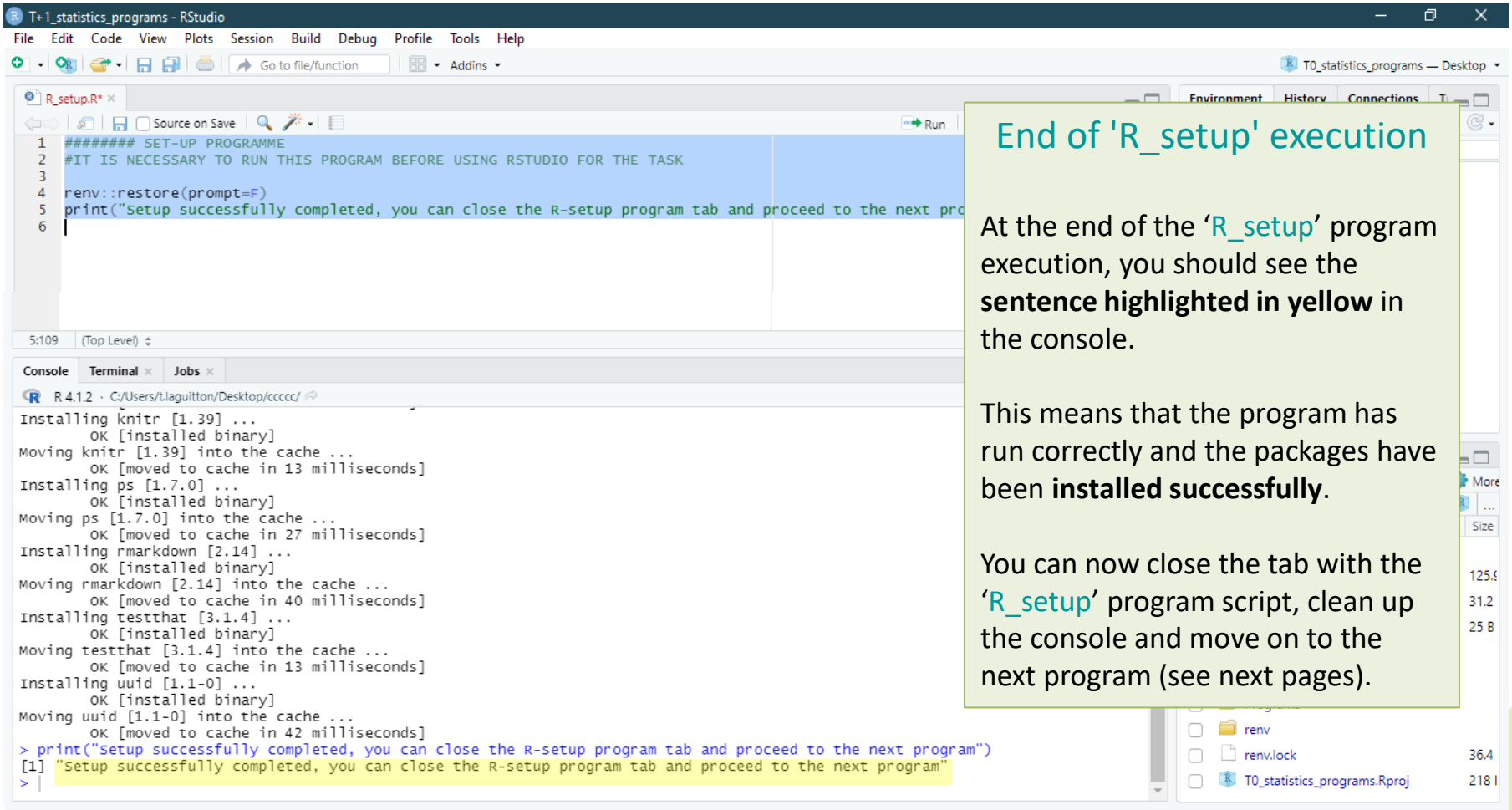
Program running

The 'R_setup' program can take **quite a long time** (a few seconds to a few minutes) to run because it installs packages in the R software.

When the program is running and has not finished its execution, a **stop sign** appears at the top right of the console. This means that the program is running and you must not close the '*T+1_statistics_programs – Rstudio*' window or the program will stop running.

Co-funded by the European Union's Health Programme (2014-2020)

Running the 'R_setup' program



End of 'R_setup' execution

At the end of the 'R_setup' program execution, you should see the **sentence highlighted in yellow** in the console.

This means that the program has run correctly and the packages have been **installed successfully**.

You can now close the tab with the 'R_setup' program script, clean up the console and move on to the next program (see next pages).

Co-funded by the European Union's Health Programme (2014-2020)

‘R_setup’ program

Tutorial video to run the Rsetup program

→ *This video is available at any time on the Best ReMaP intranet in the WP5 section* *https://portal.nijz.si/ssf/a/c/p_name/ss_forum/p_action/1/entityType/folderEntry/action/view_permalink/entryId/77696/novl_url/1*

Co-funded by the European Union's Health Programme (2014-2020)

3) Running of the programs

A. Part 1 : R setup program

B. Part 2 : Verification programs and template cleaning/standardization

i. 1st verification program : 'Verifications template_step_1' (page 90)

ii. 2nd verification program : 'Verifications template_step_2' (page 119)

iii. 3rd verification program : 'Verifications template_step_3' (page 144)

iv. 4th verification program : 'Verifications template_step_4' (page 166)

Co-funded by the European Union's Health Programme (2014-2020)

Verification programs

- In this part, you will run 4 verification programs on your file : T+1_data_collection_***country***.csv

- This part is very important because it will check that there are no input errors in the template and that the data can be used for creating indicators.

  You will need to run these programs in the following order:

  1) Verifications template_step_1.R
  2) Verifications template_step_2.R
  3) Verifications template_step_3.R
  4) Verifications template_step_4.R

- The only information that verification programs cannot check is the **accuracy** of the **Best-ReMaP category and subcategory classification** of the collected products. We therefore suggest that you carefully check the classification of your products collected before starting the verification programs.

Co-funded by the European Union's Health Programme (2014-2020)

3) Running of the programs

A. Part 1 : R setup program

B. Part 2 : Verification programs and template cleaning/standardization

i. 1st verification program : 'Verifications template_step_1'

ii. 2nd verification program : 'Verifications template_step_2'

iii. 3rd verification program : 'Verifications template_step_3'

iv. 4th verification program : 'Verifications template_step_4'

Co-funded by the European Union's Health Programme (2014-2020)

1st verification program : 'Verifications template_step_1'

Presentation of the 'Verifications template step_1' program :

- This is the first out of the 4 programs of verification of the data entered in your template.
- In this program, data entry problems are checked (misspelling of information, missing mandatory fields, ...)

Requirements before starting the program 'Verifications template_step_1' :

- Before running the program, you need to make sure that **a copy** of your T+1 collection template is saved in .csv format in the folder '**Files**' in the '**T+1_statistics_programs**' folder that you have saved on your desktop.

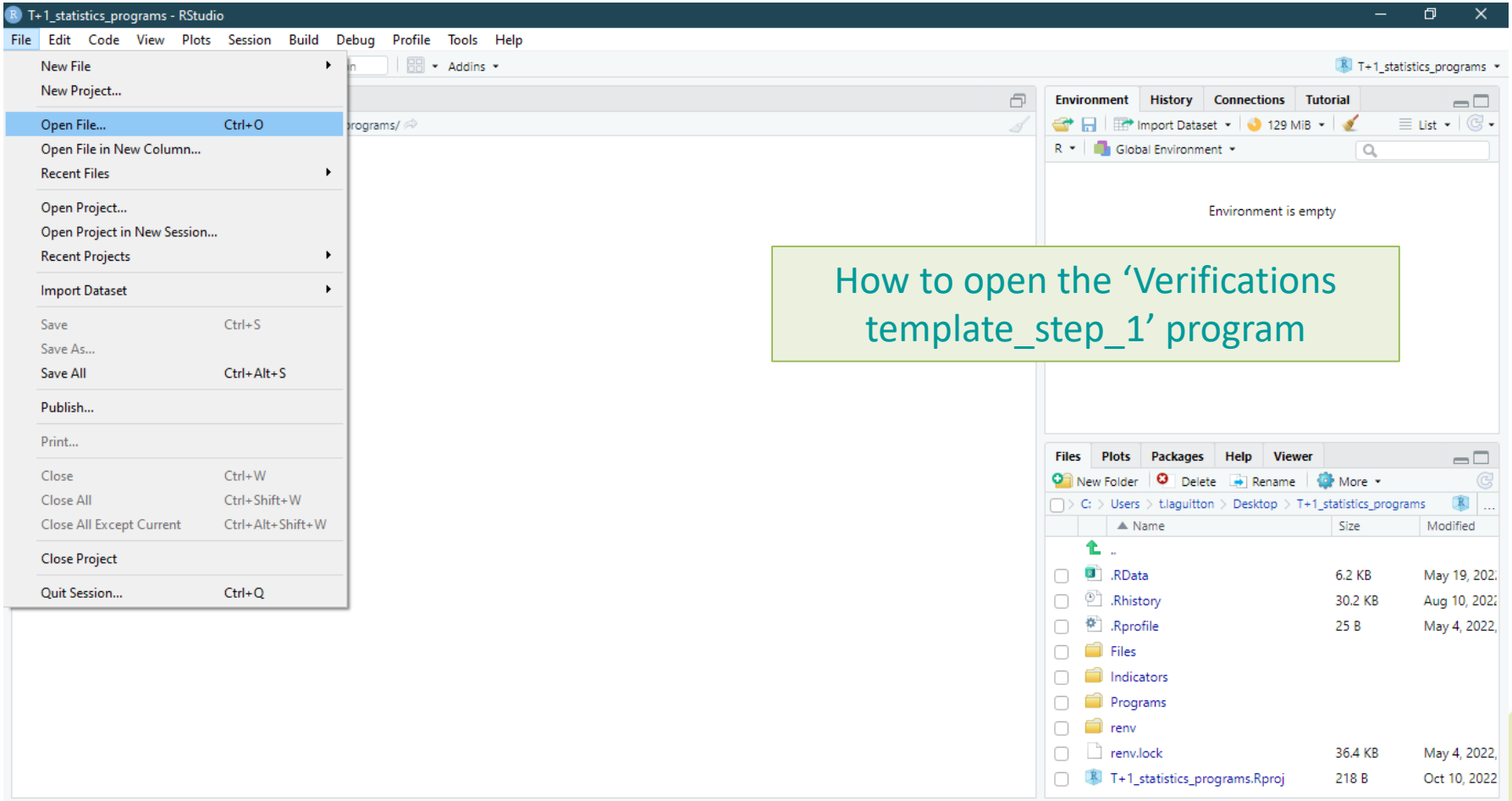
- You need to make sure that your template have been renamed : T+1_data_collection_***country***.csv (with the name of your own country)

- You need to make sure that the barcodes in your file T+1_data_collection_country.csv appear in full and not in scientific format (see procedure pages 24→28)

Your Rstudio interface must have been cleaned up before running the program.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

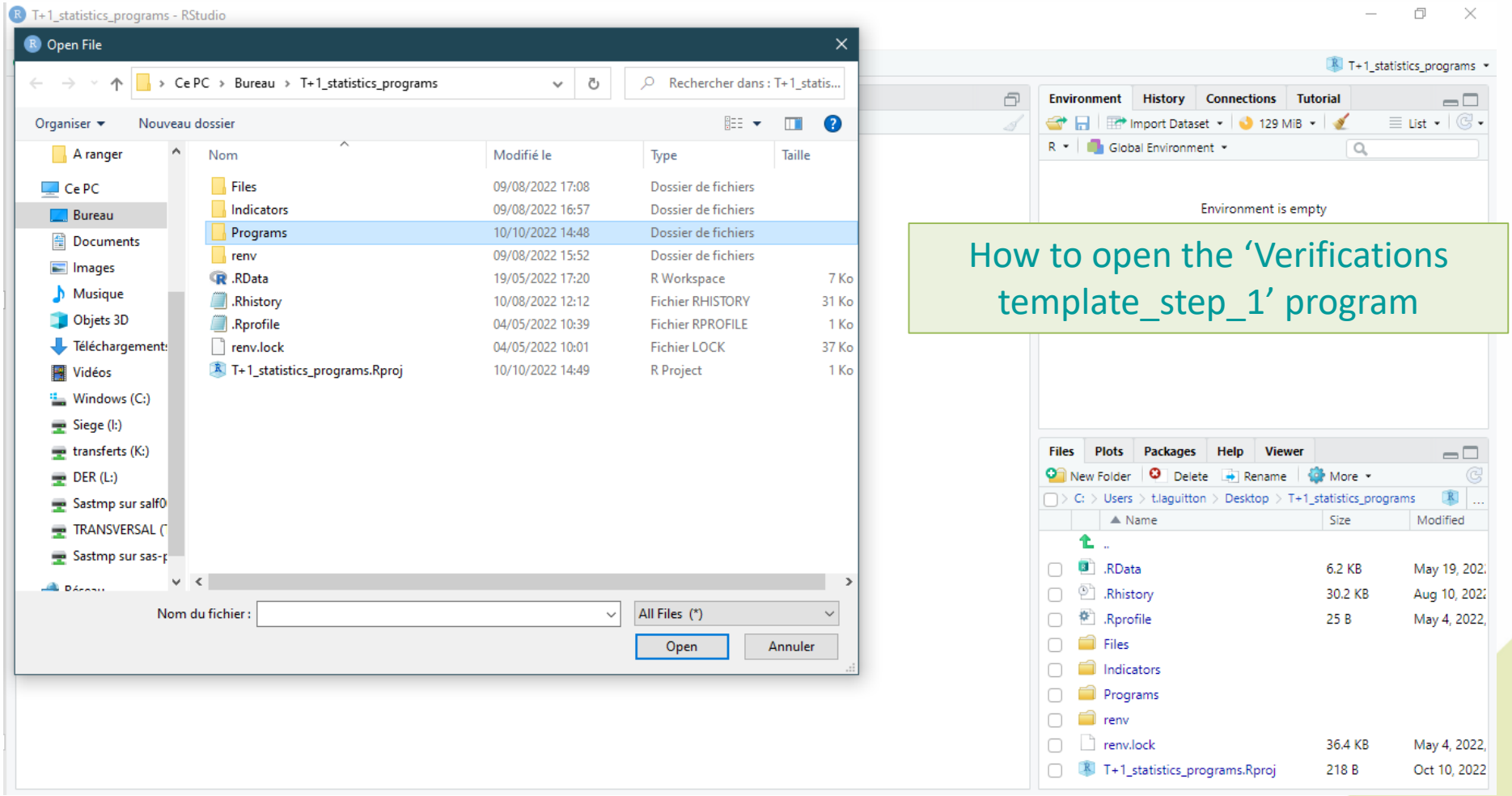
Running of ‘Verifications template_step_1’ program



How to open the ‘Verifications template_step_1’ program

Co-funded by the European Union's Health Programme (2014-2020)

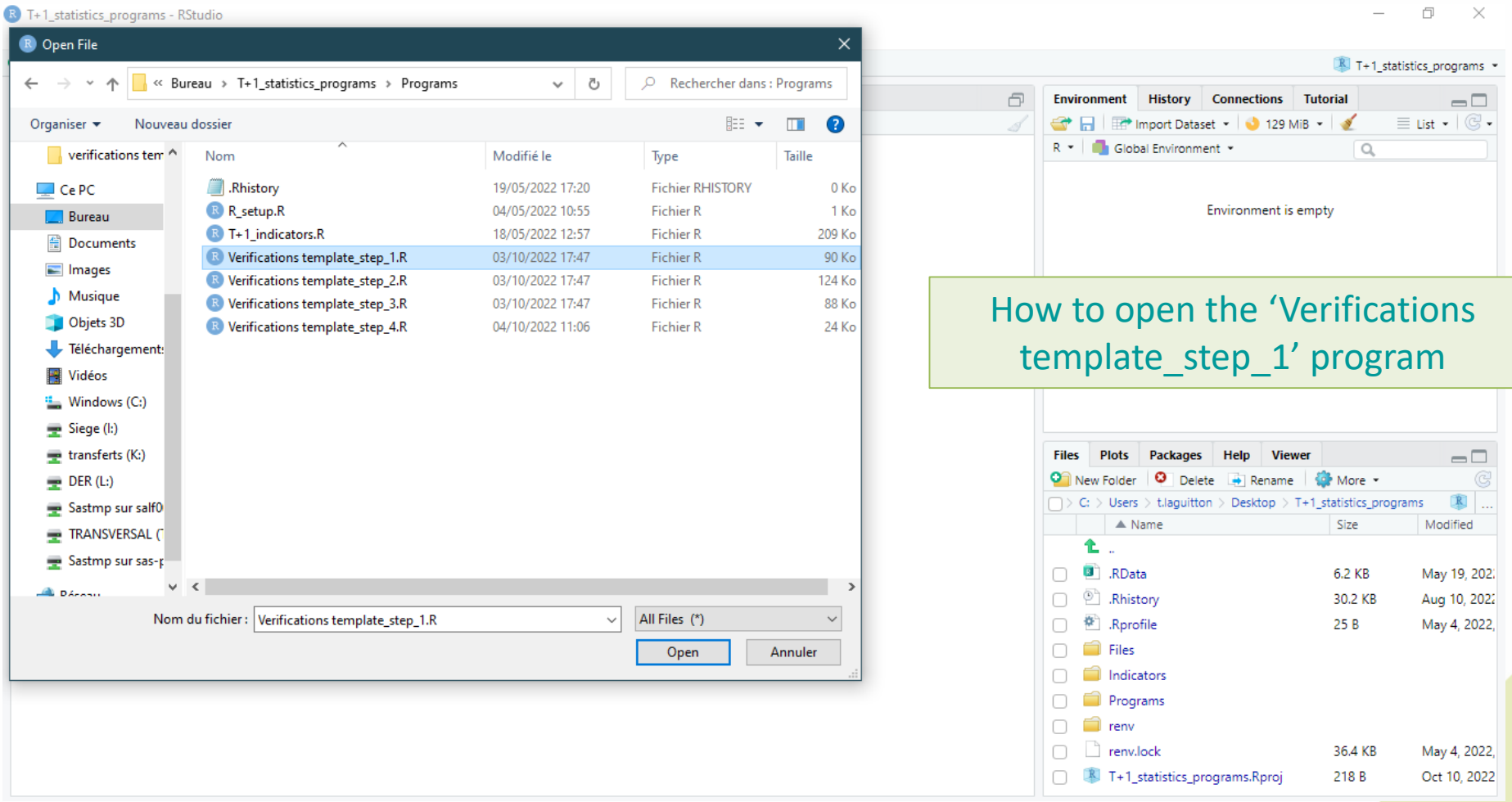
Running of ‘Verifications template_step_1’ program



How to open the ‘Verifications template_step_1’ program

Co-funded by the European Union's Health Programme (2014-2020)

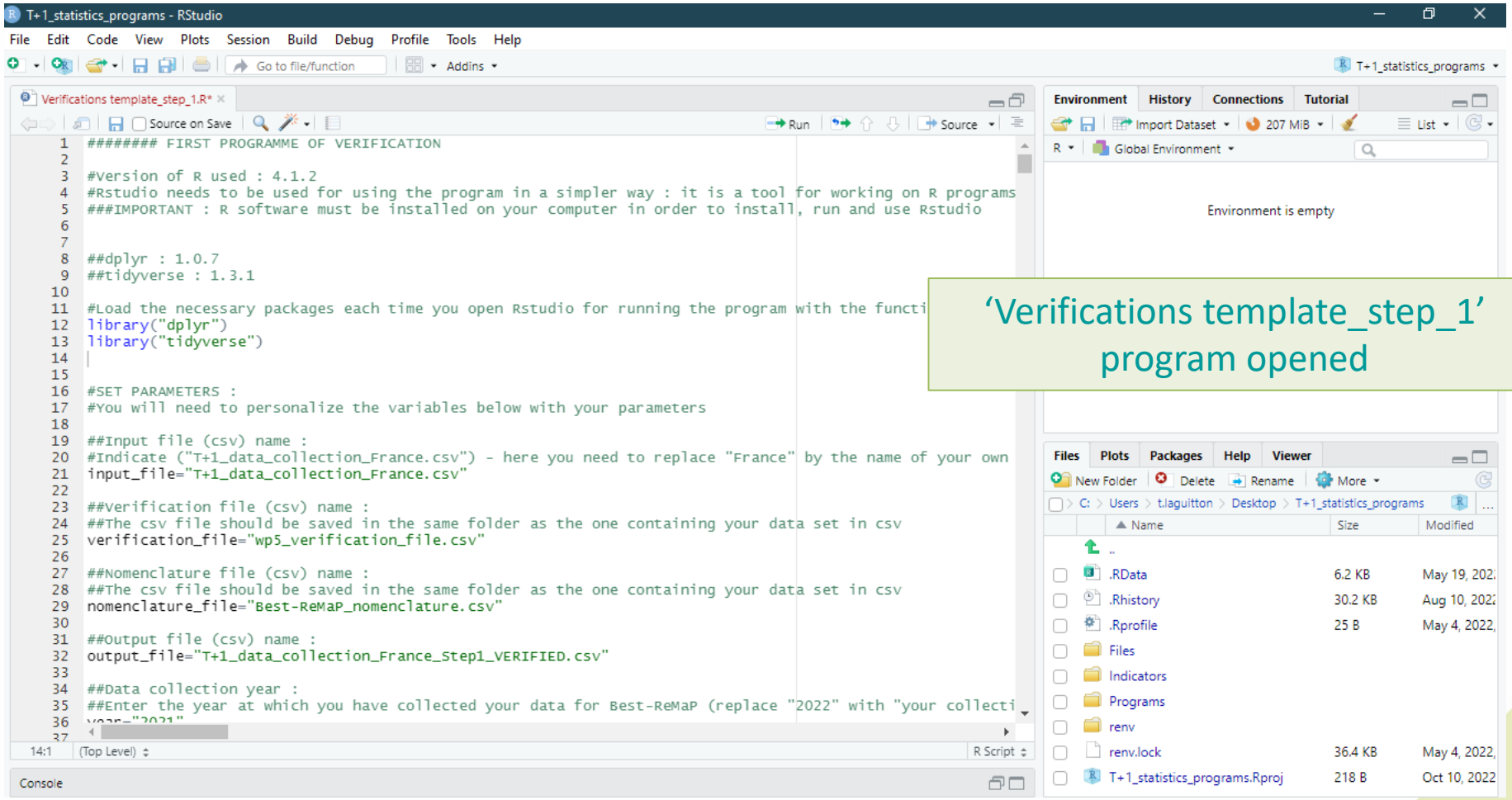
Running of ‘Verifications template_step_1’ program



How to open the ‘Verifications template_step_1’ program

Co-funded by the European Union's Health Programme (2014-2020)

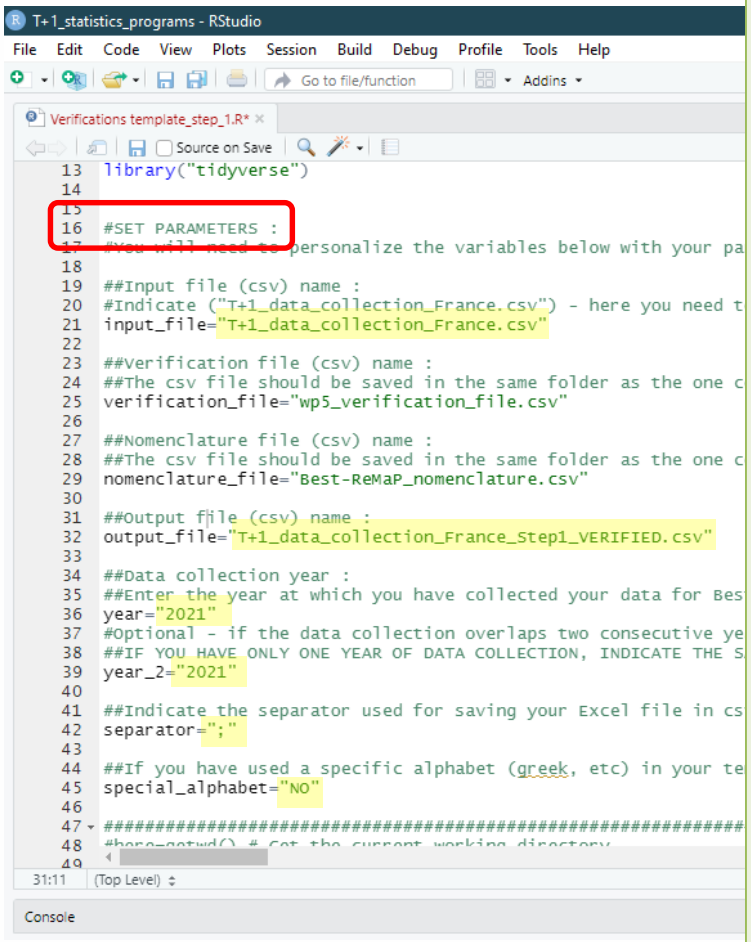
Running of ‘Verifications template_step_1’ program



‘Verifications template_step_1’ program opened

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_1' program



Beginning of each verification program

At the beginning of each verification program, there is a set parameters part (line 16). It is only in this part that you should change information.

In the first run of the verification program 1, you need to change the name of the country with your own country name in the input file (line 21) and the output file (line 32) of the R script.

You must indicate the year in which your T+1 data was collected. If the collection was carried out in the same year, you must enter this year twice (line 36 and 39). If your collection took place in two different years, you must enter both years.

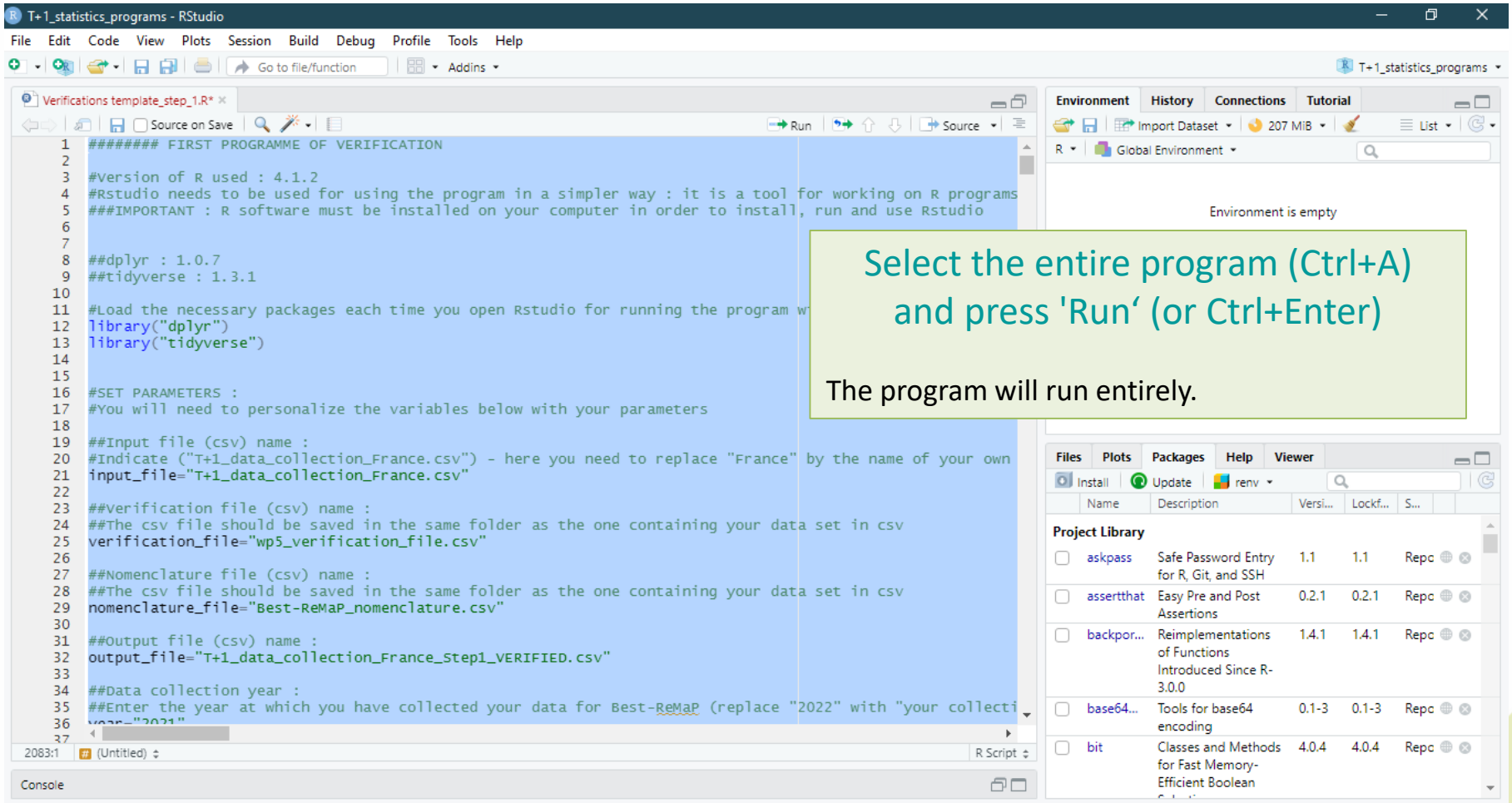
You also need to indicate the separator used in your csv file (line 42). You have the choice only between "**;**" and "**,**". In Europe, the most commonly used separator in the csv format is the "**;**" (pre-filled in the program).

You also need to indicate if you have used a specific alphabet in your template (line 45). This field only concerns countries with a specific alphabet and is therefore pre-filled as "NO".

Example :

- **input file =** "T+1_data_collection_Ireland.csv"
- **output file =** "T+1_data_collection_Ireland_Step1_VERIFIED.csv"
- **year =** "2021"
- **year_2 =** "2022"
- **Separator** = ";"
- **Special_alphabet** = "NO"

Running of ‘Verifications template_step_1’ program

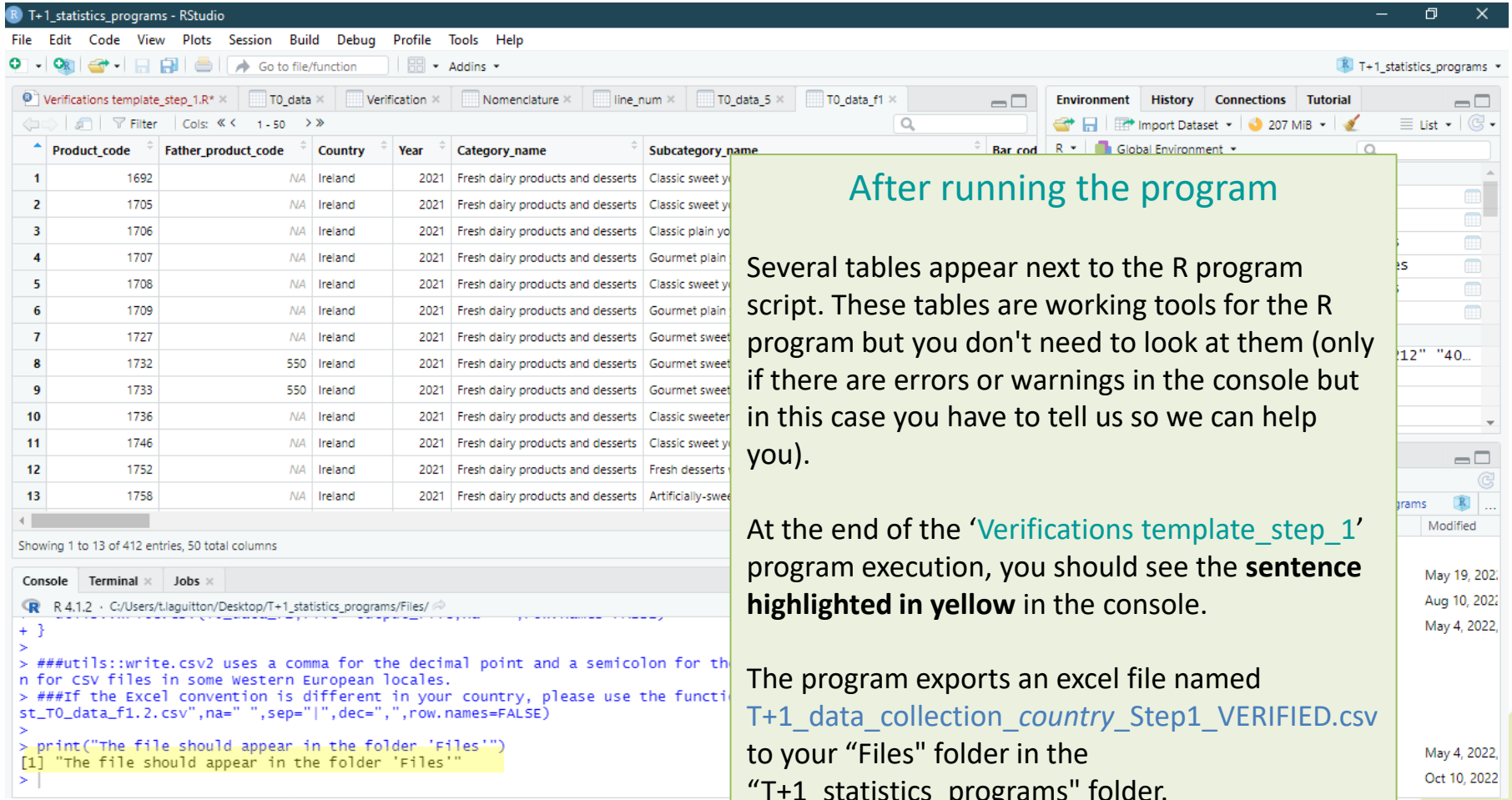


Select the entire program (Ctrl+A) and press 'Run‘ (or Ctrl+Enter)

The program will run entirely.

Co-funded by the European Union's Health Programme (2014-2020)

Running of ‘Verifications template_step_1’ program



After running the program

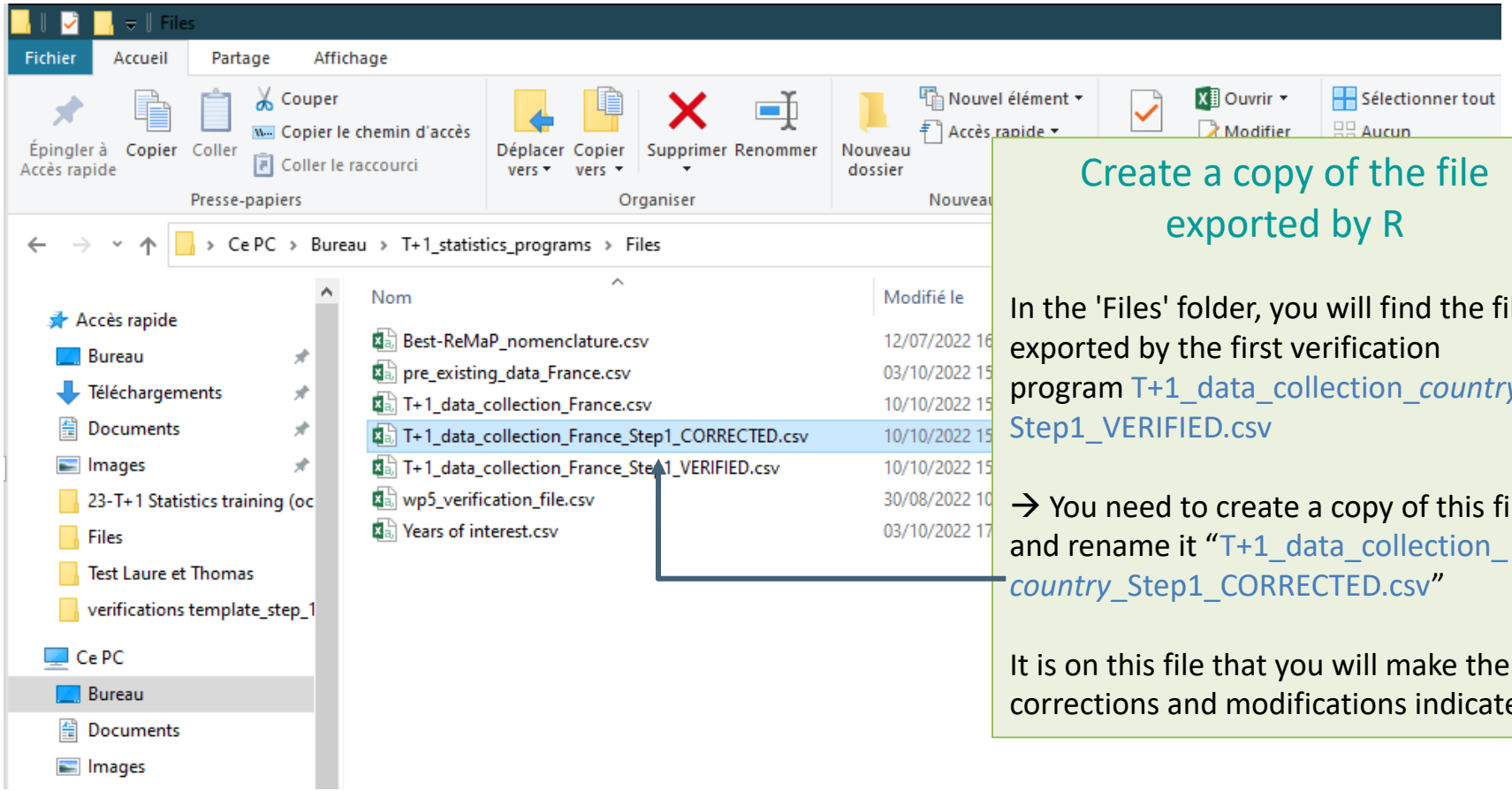
Several tables appear next to the R program script. These tables are working tools for the R program but you don't need to look at them (only if there are errors or warnings in the console but in this case you have to tell us so we can help you).

At the end of the ‘Verifications template_step_1’ program execution, you should see the **sentence highlighted in yellow** in the console.

The program exports an excel file named T+1_data_collection_*country*_Step1_VERIFIED.csv to your “Files" folder in the “T+1_statistics_programs" folder.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_1' program



Create a copy of the file exported by R

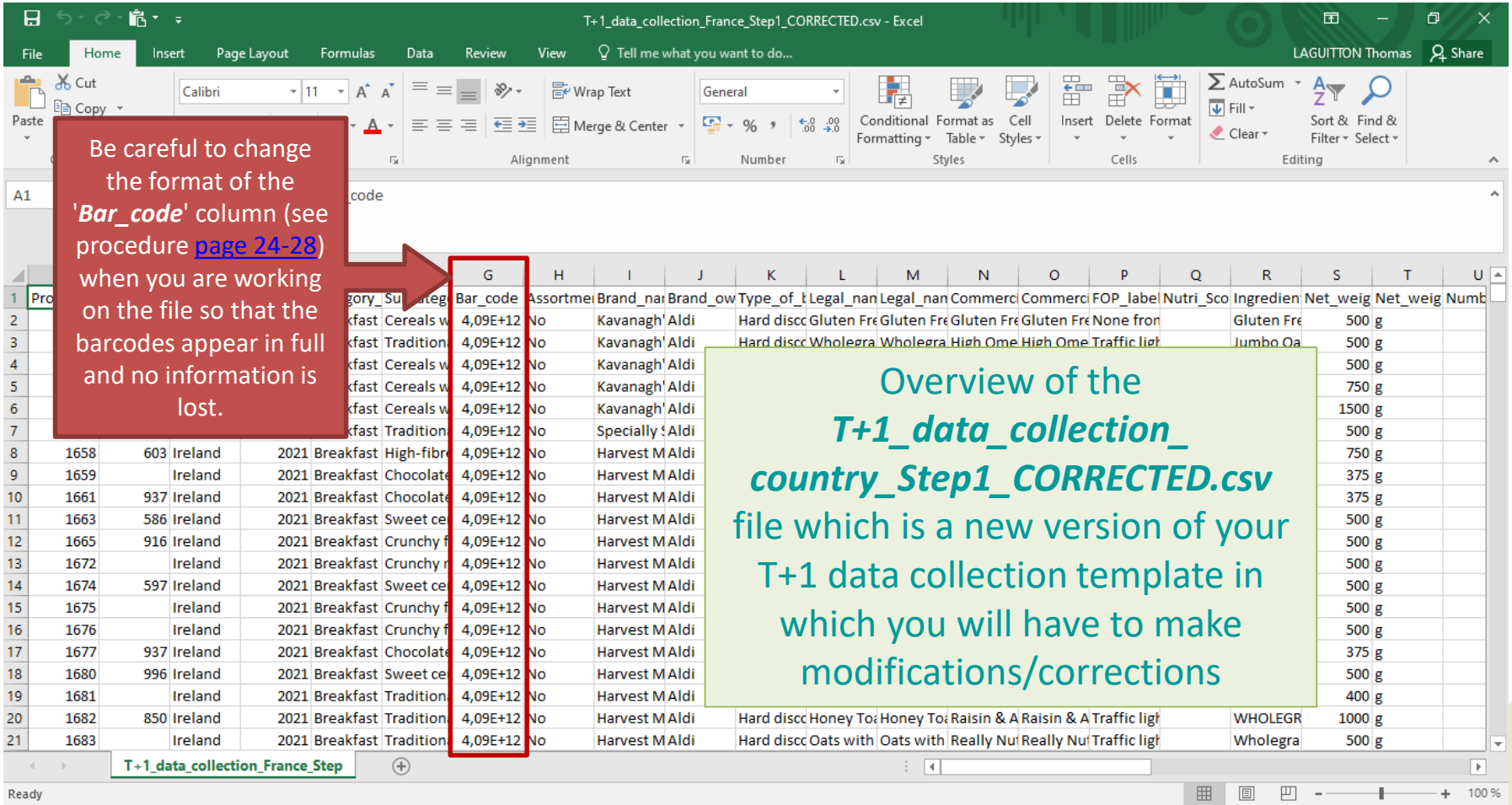
In the 'Files' folder, you will find the file exported by the first verification program T+1_data_collection_*country*_Step1_VERIFIED.csv

→ You need to create a copy of this file and rename it "T+1_data_collection_*country*_Step1_CORRECTED.csv"

It is on this file that you will make the corrections and modifications indicated

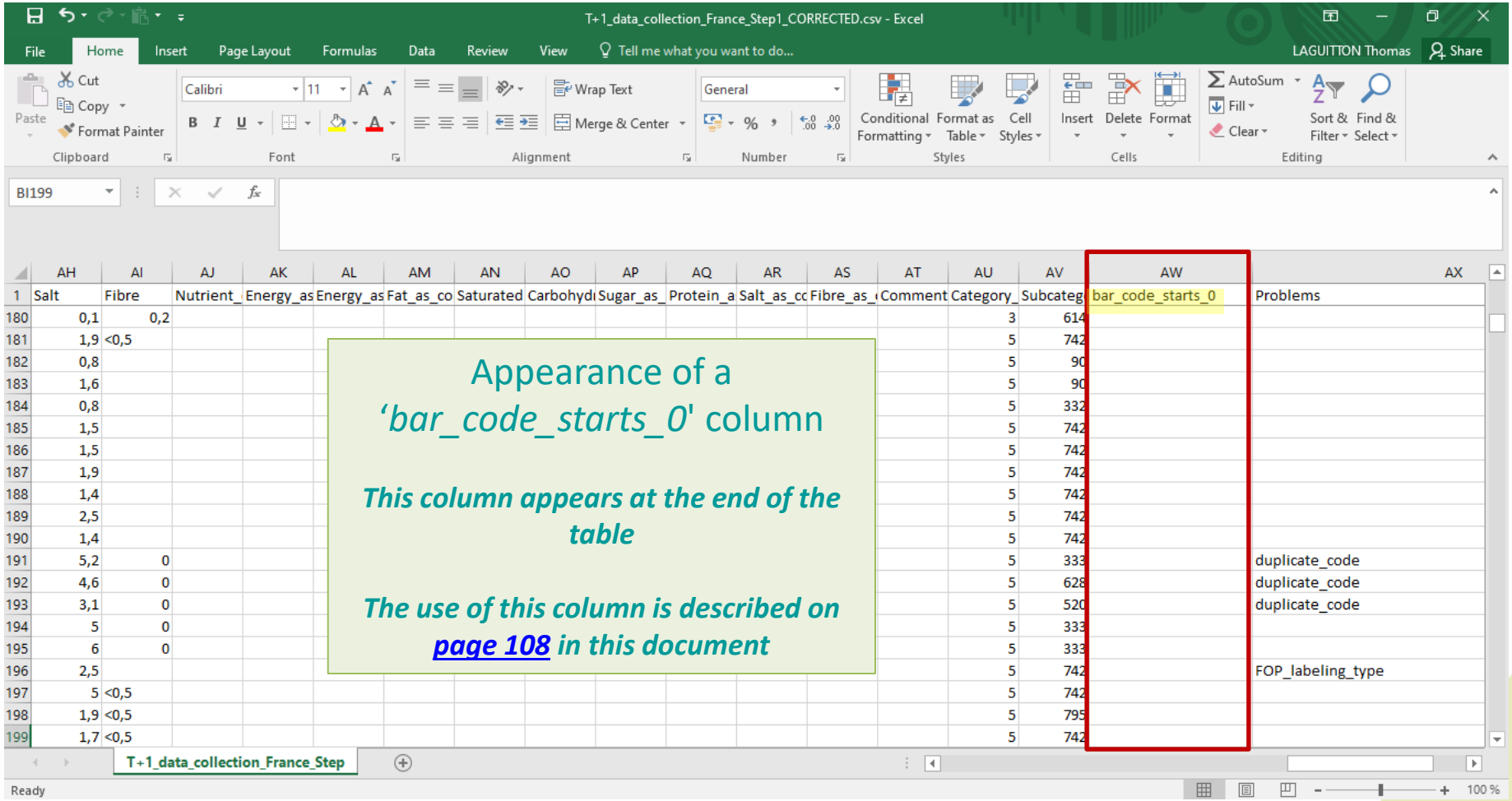
Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program



Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program



	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX
1	Salt	Fibre	Nutrient_	Energy_as	Energy_as	Fat_as_co	Saturated	Carbohydi	Sugar_as_	Protein_a	Salt_as_cc	Fibre_as_(	Comment	Category_	Subcateg	bar_code_starts_0	Problems
180	0,1	0,2												3	614		
181	1,9	<0,5												5	742		
182	0,8													5	90		
183	1,6													5	90		
184	0,8													5	332		
185	1,5													5	742		
186	1,5													5	742		
187	1,9													5	742		
188	1,4													5	742		
189	2,5													5	742		
190	1,4													5	742		
191	5,2	0												5	333		duplicate_code
192	4,6	0												5	628		duplicate_code
193	3,1	0												5	520		duplicate_code
194	5	0												5	333		
195	6	0												5	333		
196	2,5													5	742		FOP_labeling_type
197	5	<0,5												5	742		
198	1,9	<0,5												5	795		
199	1,7	<0,5												5	742		

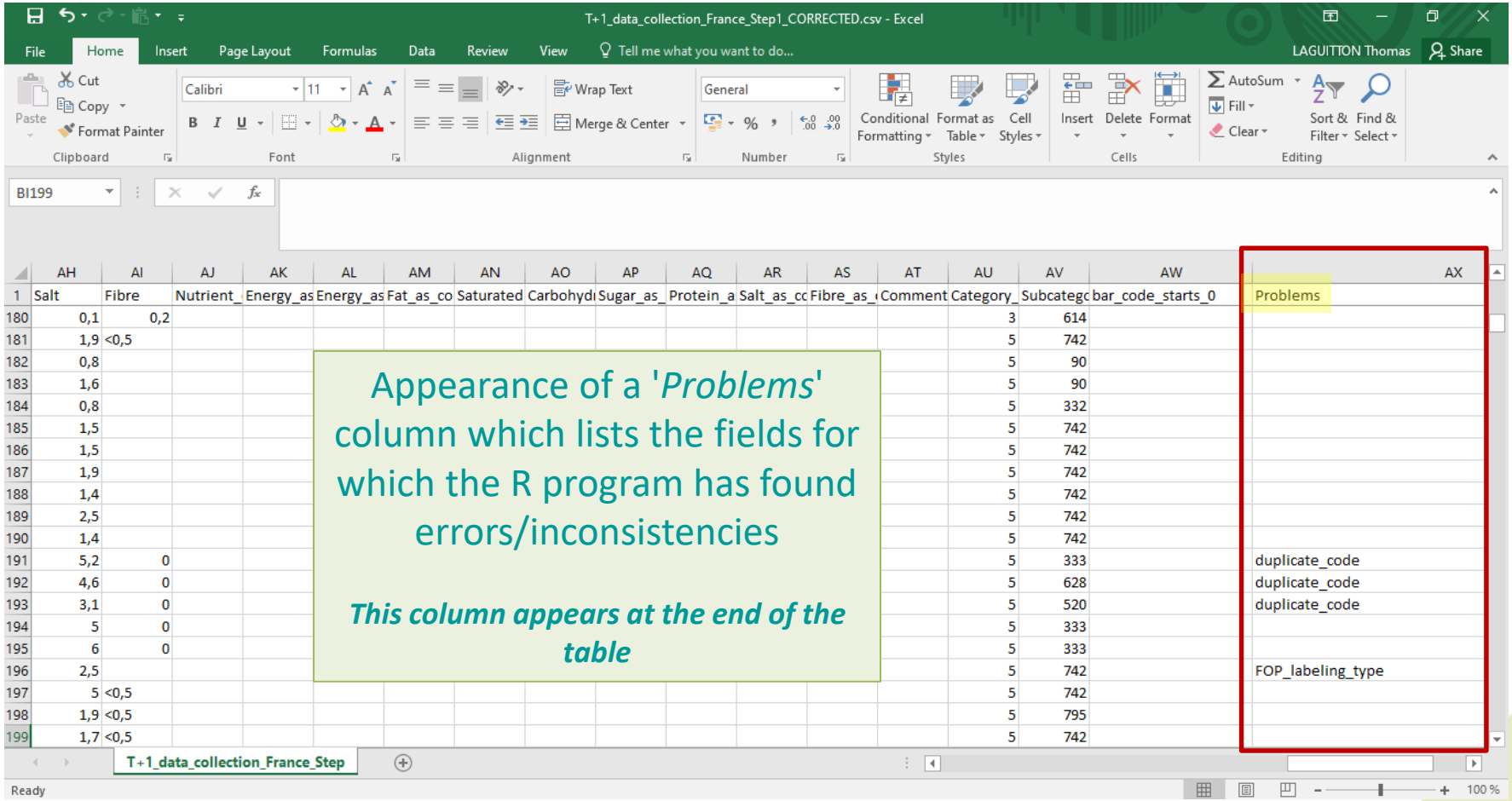
Appearance of a '*bar_code_starts_0*' column

This column appears at the end of the table

The use of this column is described on page 108 in this document

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program



T+1_data_collection_France_Step1_CORRECTED.csv - Excel

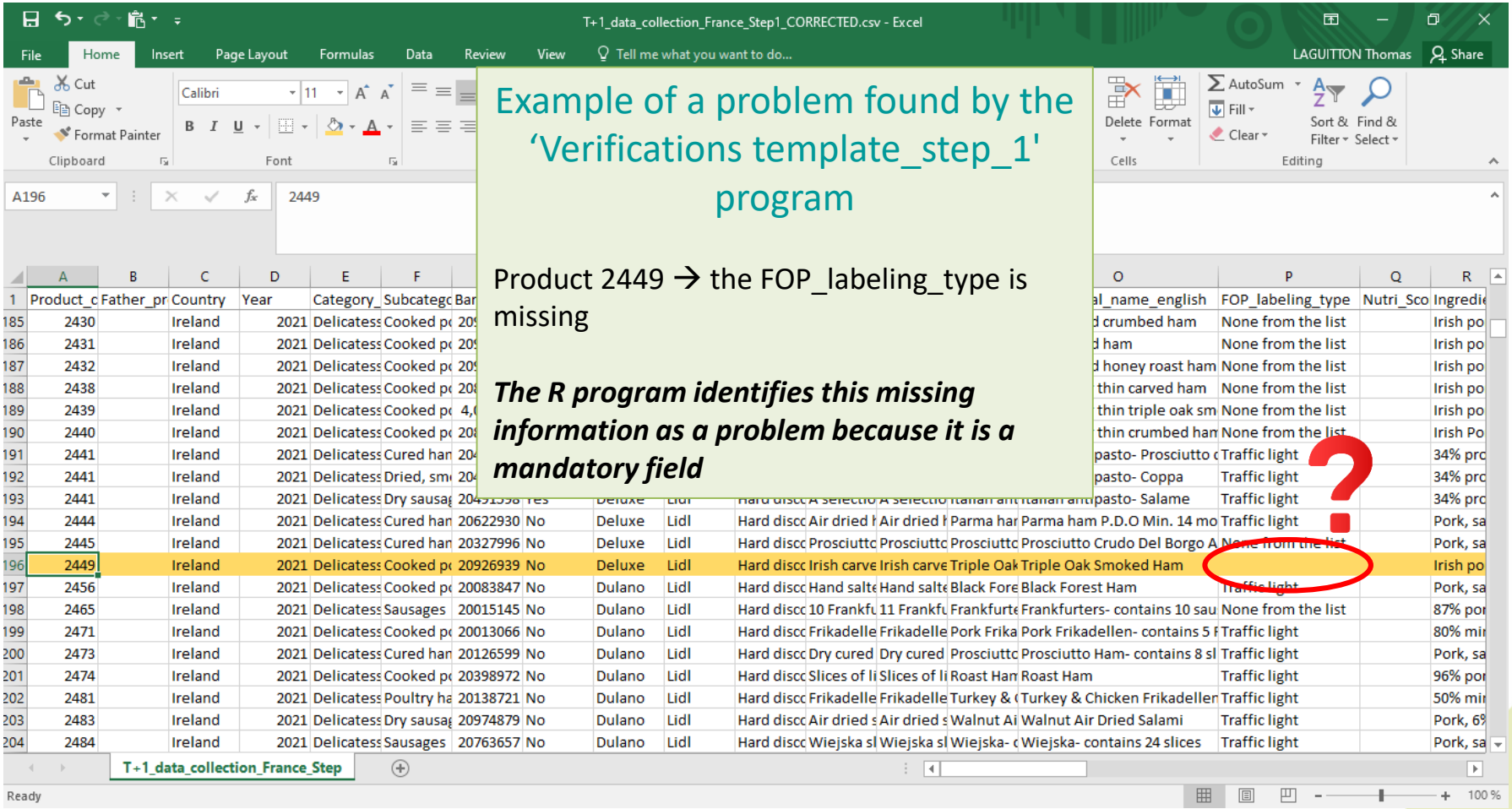
	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX
1	Salt	Fibre	Nutrient_	Energy_as	Energy_as	Fat_as_co	Saturated	Carbohydi	Sugar_as_	Protein_a	Salt_as_cc	Fibre_as_	Comment	Category_	Subcateg	bar_code_starts_0	Problems
180	0,1	0,2												3	614		
181	1,9	<0,5												5	742		
182	0,8													5	90		
183	1,6													5	90		
184	0,8													5	332		
185	1,5													5	742		
186	1,5													5	742		
187	1,9													5	742		
188	1,4													5	742		
189	2,5													5	742		
190	1,4													5	742		
191	5,2	0												5	333		duplicate_code
192	4,6	0												5	628		duplicate_code
193	3,1	0												5	520		duplicate_code
194	5	0												5	333		
195	6	0												5	333		
196	2,5													5	742		FOP_labeling_type
197	5	<0,5												5	742		
198	1,9	<0,5												5	795		
199	1,7	<0,5												5	742		

Appearance of a '*Problems*' column which lists the fields for which the R program has found errors/inconsistencies

This column appears at the end of the table

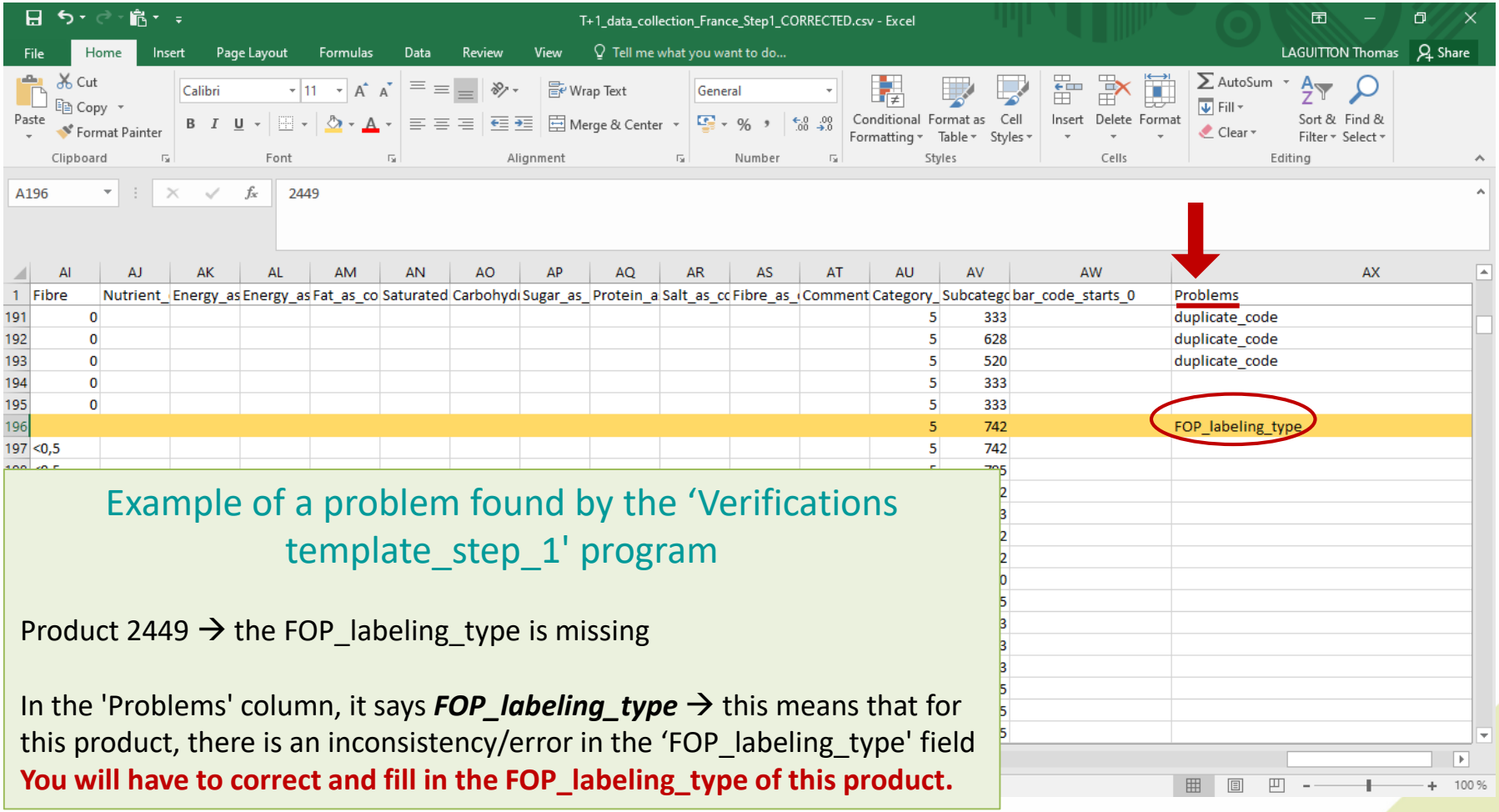
Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program



Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after ‘Verifications template_step_1’ program



Example of a problem found by the ‘Verifications template_step_1' program

Product 2449 → the FOP_labeling_type is missing

In the 'Problems' column, it says ***FOP_labeling_type*** → this means that for this product, there is an inconsistency/error in the ‘FOP_labeling_type' field
You will have to correct and fill in the FOP_labeling_type of this product.

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after ‘Verifications template_step_1’ program

Terms that may appear in the 'Problems' field following the 'R_verifications template_step_1' program, their meaning and what to do

- In the following slides all the terms that can appear in the problem column and how to make corrections will be presented.
- Please note that the verification program 'Verifications template_step_1' is **case sensitive** (upper and lower case letters are important). Any information entered that does not match the **spelling** and **case** of the elements in the drop-down lists of the initial data entry template will appear as an error.
- If a line is completely empty (no fields filled) but errors for all fields appear after running the program, then this line should be deleted.
  Empty rows are normally deleted by the program, but some may still be imported.
- The program 'verifications template_step_1' and all other programs take into account if you have used the template containing only one ‘FOP_labeling_type’ field or if you have used the latest version of the template containing several ‘FOP_labeling_type’ fields

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'R verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action	
Duplicate_code	Different products have the same product code	→ You must change the product code so that all products have a unique code	'Product_code' is a mandatory field
Empty_product_code	The product does not have a unique product code	→ You must create a unique product code that does not already exist for the product	
Country	Incorrect country name (i.e. not contained in the closed list of the input template) or missing country name	→ You must check the spelling of the country by comparing it with the **closed list of the input template** or add the country name if it is missing	'Country' is a mandatory field
Year	year different from that/those indicated in the 'set parameters' part of the program	→ You must correct the collection year that does not match your collection year(s) or add it if it is missing	'Year' is a mandatory field
Category_name	Incorrect category name (i.e. not contained in the closed list of the input template) or missing category name	→ You must check the spelling of the category name by comparing it with the **closed list of the input template** or add the category name if it is missing	'Category_name' is a mandatory field
Category_code	Category code that does not exists or missing category code	→ You must check the category code exists by comparing it with the classification guides of the 5 food categories or add the category code if it is missing	'Category_code' is a mandatory field

Co-funded by the European Union's Health Programme (2014-2020)

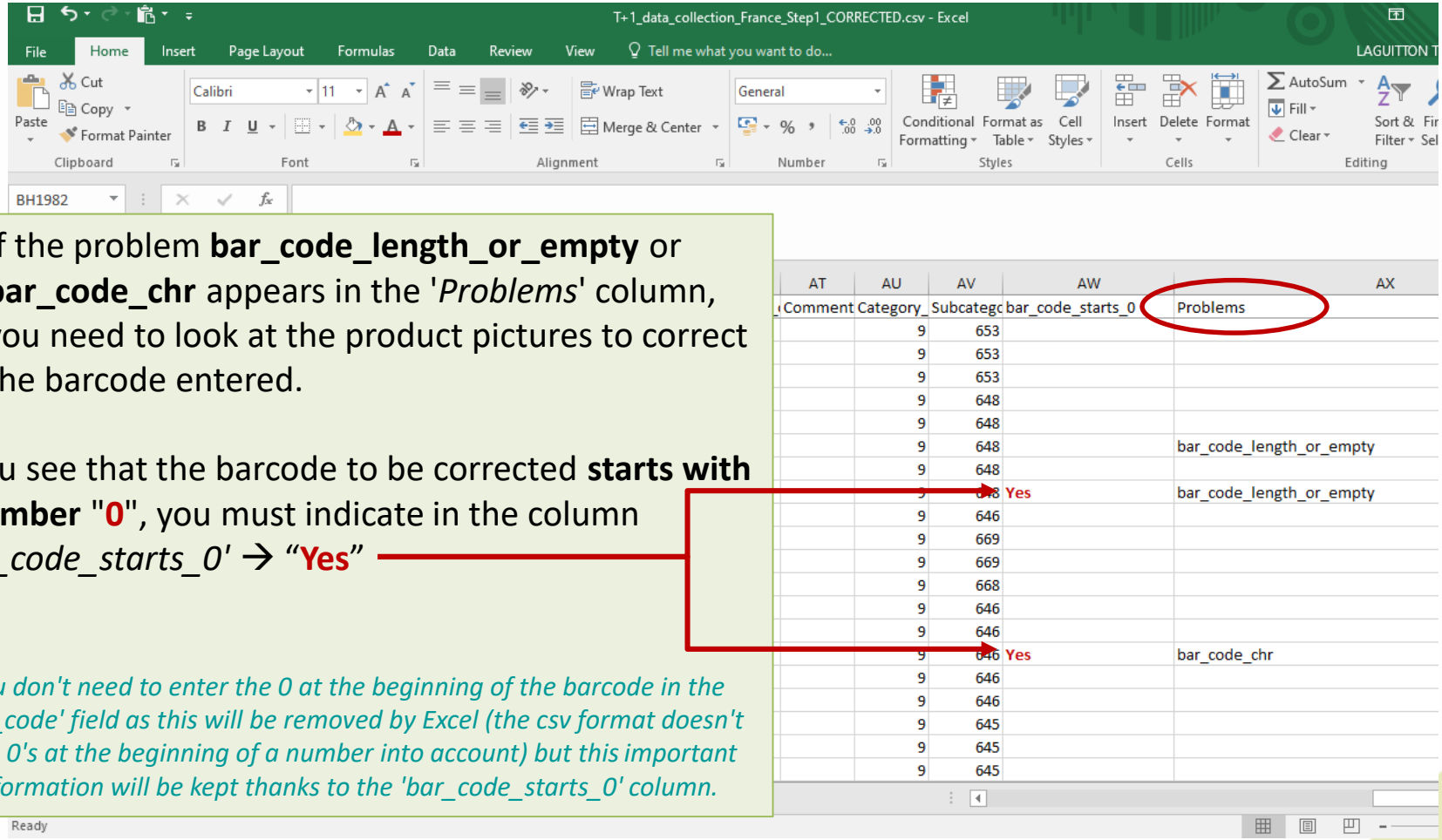
Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'R verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action	
Subcategory_ name	Incorrect subcategory name (i.e. not contained in the closed list of the input template) or missing subcategory name	→ You must check the spelling of the subcategory name by comparing it with the closed list of the input template or add the subcategory name if it is missing	'Subcategory_ name' is a mandatory field
Subcategory_ code	Category code that does not exists or missing category code	→ You must check the subcategory code exists by comparing it with the classification guides of the 5 food categories or add the subcategory code if it is missing	'Subcategory_ code' is a mandatory field
Bar_code_ length_or_ empty	The barcode does not have 8, 12, 13, 14, or 15 digits or is missing	→ You must go back to the product pictures and correct the barcode or add the barcode if you have forgotten it. If the barcode does not appear on the product pictures or is the same as in the picture, you must indicate in the 'Comments' field: "*barcode checked*".	
Bar_code_chr	The barcode contains characters other than numbers that are unwanted	→ You must go back to the product pictures and correct the barcode as there can be no characters other than numbers in a barcode	

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program



- If the problem **bar_code_length_or_empty** or **bar_code_chr** appears in the '*Problems*' column, you need to look at the product pictures to correct the barcode entered.

If you see that the barcode to be corrected **starts with a number** "**0**", you must indicate in the column *'bar_code_starts_0'* → "**Yes**"

You don't need to enter the 0 at the beginning of the barcode in the 'Bar_code' field as this will be removed by Excel (the csv format doesn't take 0's at the beginning of a number into account) but this important information will be kept thanks to the 'bar_code_starts_0' column.

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'R verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action	
Brand_name	Brand name is missing	→ You must go back to the product pictures and add the brand name of the product if you have forgotten it. If the brand name does not appear on the product pictures or is unreadable, you must indicate in the 'Comments' field: "*brand name checked and unreadable/not existing*".	
Type_of_brand	Incorrect type of brand (i.e. not contained in the closed list of the input template) or missing type of brand	→ You must check the spelling of the type of brand by comparing it with the closed list of the input template or add the type of brand if it is missing by looking at the product's brand name	'Type_of_brand' is a mandatory field
• **Legal_name** • **Legal_name _english**	Legal name is missing Legal name in english is missing	→ You must go back to the product pictures and add the legal name of the product if you have forgotten it. If the legal name does not appear on the product pictures or is unreadable, you must indicate in the 'Comments' field: "*legal name checked and unreadable/not existing*". → If the legal name in English is missing, you must translate the legal name and add it to the template	

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'R verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action	
• **Commercial_name** • **Commercial_name_english**	Commercial name is missing Commercial name in english is missing	→ You must go back to the product pictures and add the commercial name of the product if you have forgotten it. If the commercial name does not appear on the product pictures or is unreadable, you must indicate in the 'Comments' field: "*commercial name checked and unreadable/not existing*". → If the commercial name in English is missing, you must translate the commercial name and add it to the template	
• **FOP_labeling_type** • FOP_labeling_type_2 • FOP_labeling_type_3 • FOP_labeling_type_4 ↓ can appear if you use the latest version of the template	Incorrect FOP labeling type (i.e. not contained in the closed list of the input template) or missing FOP labeling type	→ You must check the spelling of the FOP labeling type by comparing it with the closed list of the input template or add the FOP labeling type if it is missing by looking at the product's pictures. If there is no FOP labeling type of interest on the pictures of the product, you must enter 'None from the list'. → When this problem occurs, it can affect the 'FOP_labeling_type' field and also the 'FOP_labeling_type_2/_3 /_4' fields for those using the latest version of the template. In this case, it is necessary to check the spelling of the field concerned.	'FOP_labeling_type' is a mandatory field
Nutri_score	Incorrect nutri-score (not a letter between A and E)	→ You must go back to the product pictures and find the correct nutri-score of the product and enter it in the template	

Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'R verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action
Ingredient_list	Ingredient list is missing	→ You must go back to the product pictures and add the ingredient list of the product if you have forgotten it. If the ingredient list does not appear on the product pictures or is unreadable, you must indicate in the 'Comments' field: *"ingredient list checked and unreadable/not existing*".
Net_weight	The net weight contains characters other than numbers that are unwanted	→ You must go back to the product pictures and correct the net weight as there can be no characters other than numbers in the 'net_weight' field.
Net_weight_unit	The net weight unit is different from « **g** » or « **mL** » (i.e. not contained in the closed list of the input template)	→ You must check the spelling of the net weight unit by paying attention to upper and lower case. It should be entered as "**g**" or "**mL**".
Number_of_units	The number of units contains characters other than numbers that are unwanted	→ You must correct the number of units as there can be no characters other than numbers in the 'number_of_units' field
Portion_size	The portion size contains characters other than numbers that are unwanted	→ You must go back to the product pictures and correct the portion size as there can be no characters other than numbers in the 'portion_size' field.

Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action
Portion_size_unit	The portion size unit is different from « **g** » or « **mL** »	→ You must check the spelling of the net weight unit by paying attention to upper and lower case. It must be entered "**g**" or "**mL**" and not something else.
Nutrient_content_expression_unit	The nutrient content expression unit is different from « **100g** » or « **100mL** »	→ You must check the spelling of the nutrient content expression unit. It must be entered « **100g** » or « **100mL** » and not something else.
• **Energy_kCal** • **Energy_kJ** • **Fat** • **Saturated_fat** • **Carbohydrates** • **Sugar** • **Protein** • **Salt** • **Fibre**	The fields contain characters other than numbers (except "<" and "traces") that are unwanted.	→ You must correct so that only numbers remain and no other characters → If you have any doubt about the values when correcting, go back to the product photos
Nutrient_content_expression_unit_as_consumed	The nutrient content expression unit for products to be reconstituted is different from « **100g** » or « **100mL** »	→ You must check the spelling of the nutrient content expression unit as consumed. It must be entered « **100g** » or « **100mL** » and not something else

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_1' program

Terms that may appear in the 'Problems' field following the 'R verifications template_step_1' program, their meaning and what to do

Problem	Meaning	Action
• **Energy_as_consumed_kCal** • **Energy_as_consumed_kJ** • **Fat_as_consumed** • **Saturated_fat_as_consumed** • **Carbohydrates_as_consumed** • **Sugar_as_consumed** • **Protein_as_consumed** • **Salt_as_consumed** • **Fibre_as_consumed**	The fields contain characters other than numbers (except "<" and "traces") that are unwanted	→ You must correct so that only numbers remain and no other characters → If you have any doubt about the values when correcting, go back to the product photos

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after ‘Verifications template_step_1’ program

- Fields that are indicated as mandatory in the previous table and that show a problem when information is missing

  → **You must correct or enter information for these fields.**

- For nutritional values containing the word “traces”, it is also necessary to **check by hand** and **standardize** the spelling of "**traces**" (so you don't have "Traces", "trace", ...)

- For products where a problem appears but the information entered is correct or missing, you must indicate in the '*Comments*' field that the problem has been checked (see the ‘*Action*' column in the previous tables).

  → **This way, when you will run the verification program again, you will be able to identify problems that appear without being a problem (and have already been verified).**

Co-funded by the European Union's Health Programme (2014-2020)

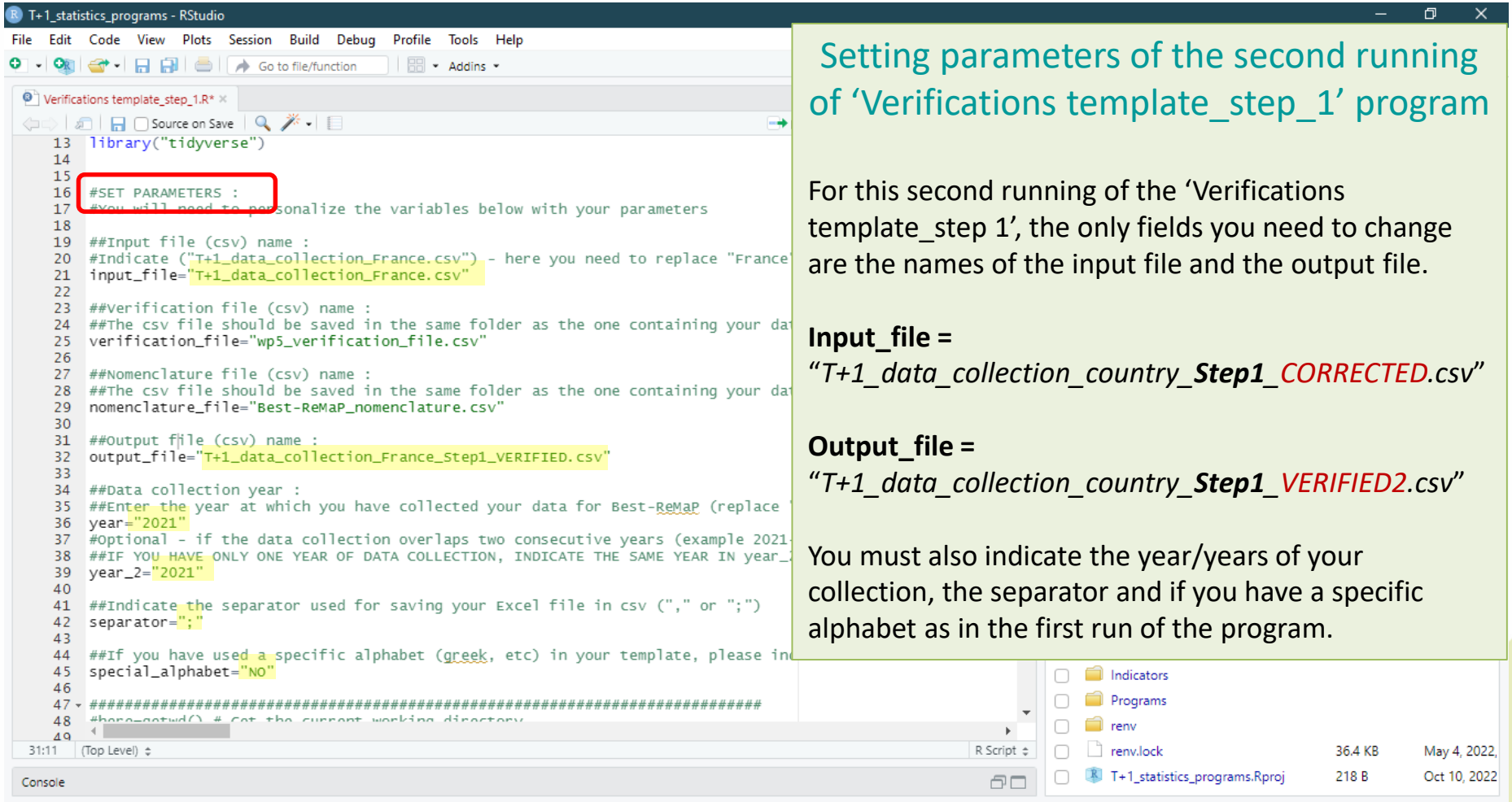
Excel file to modify after ‘Verifications template_step_1’ program

- **Be careful! When saving the file, you must be sure that the barcodes appear in full and not in scientific format. If this is the case, you must repeat the procedure on pages 24→28.**
- Once the corrections have been made and the barcodes are in the correct format, you can save the file *T+1_data_collection_country_Step1_CORRECTED.csv* and close it.
- You will have to run this corrected file again in the ‘Verifications template_step_1' program to make sure you haven't missed a check.

Your Rstudio interface must be cleaned up before running the ‘Verifications template_step_1’ program again.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_1' program



Setting parameters of the second running of 'Verifications template_step_1' program

For this second running of the 'Verifications template_step 1', the only fields you need to change are the names of the input file and the output file.

Input_file =
"*T+1_data_collection_country_**Step1**_CORRECTED.csv*"

Output_file =
"*T+1_data_collection_country_**Step1**_VERIFIED2.csv*"

You must also indicate the year/years of your collection, the separator and if you have a specific alphabet as in the first run of the program.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_1' program

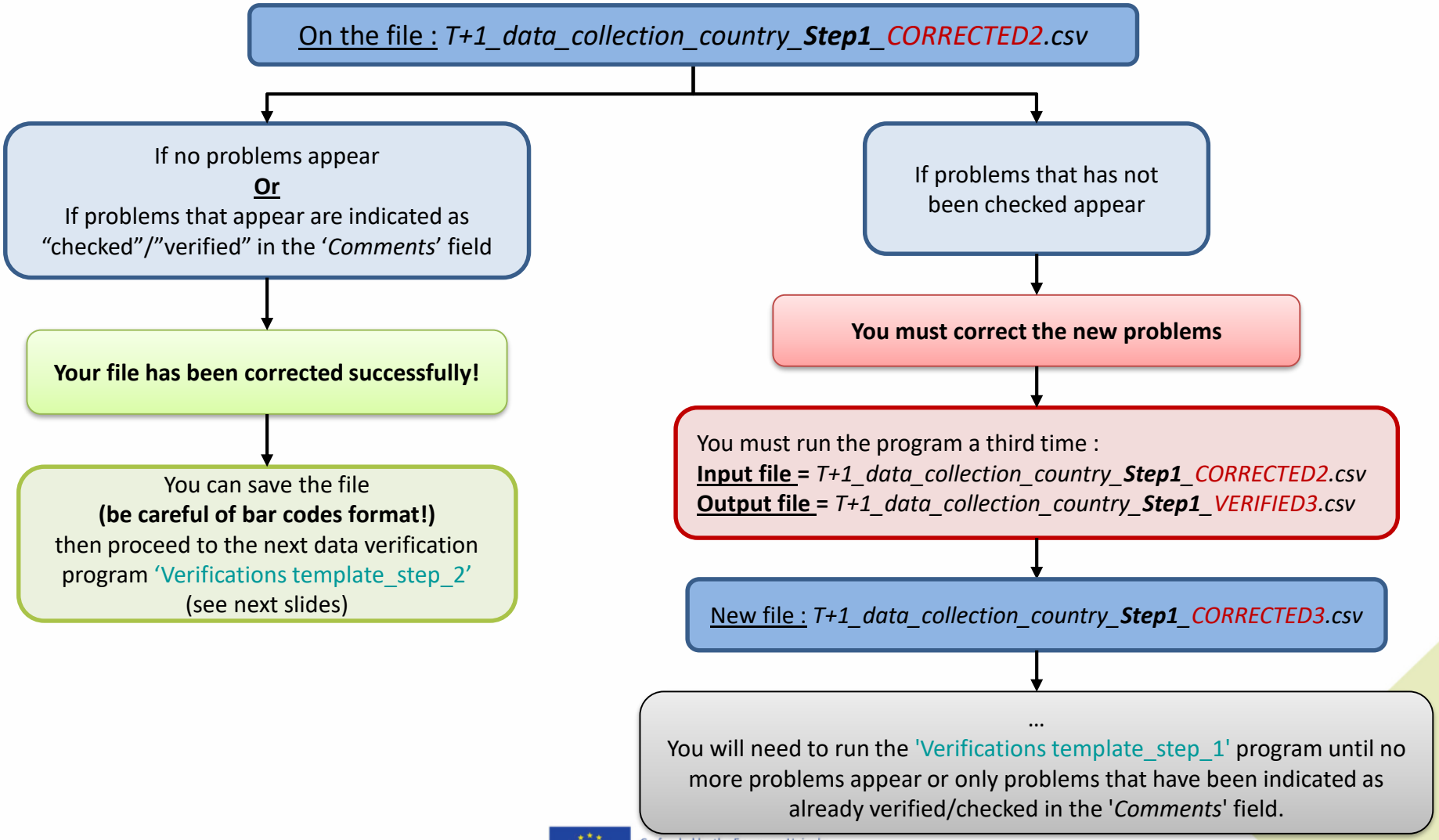
- At the end of this second run, you get in the "***Files***" folder a file called: "*T+1_data_collection_country_**Step1**_VERIFIED2.csv*"

- You must create a copy of this file and call it : *"T+1_data_collection_country_**Step1**_CORRECTED2.csv"*

  → It is on this file that you will make the modifications following the checks

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_1' program



Co-funded by the European Union's Health Programme (2014-2020)

3) Running of the programs

A. Part 1 : R setup program

B. Part 2 : Verification programs and template cleaning/standardization

i. 1st verification program : 'Verifications template_step_1'

ii. 2nd verification program : 'Verifications template_step_2'

iii. 3rd verification program : 'Verifications template_step_3'

iv. 4th verification program : 'Verifications template_step_4'

Co-funded by the European Union's Health Programme (2014-2020)

2nd verification program : 'Verifications template_step_2'

Presentation of the 'Verifications template step_2' program :

- In this second verification program, consistency problems will be highlighted: consistency of categories and their codes, sub-categories and their codes, values and their units, etc.

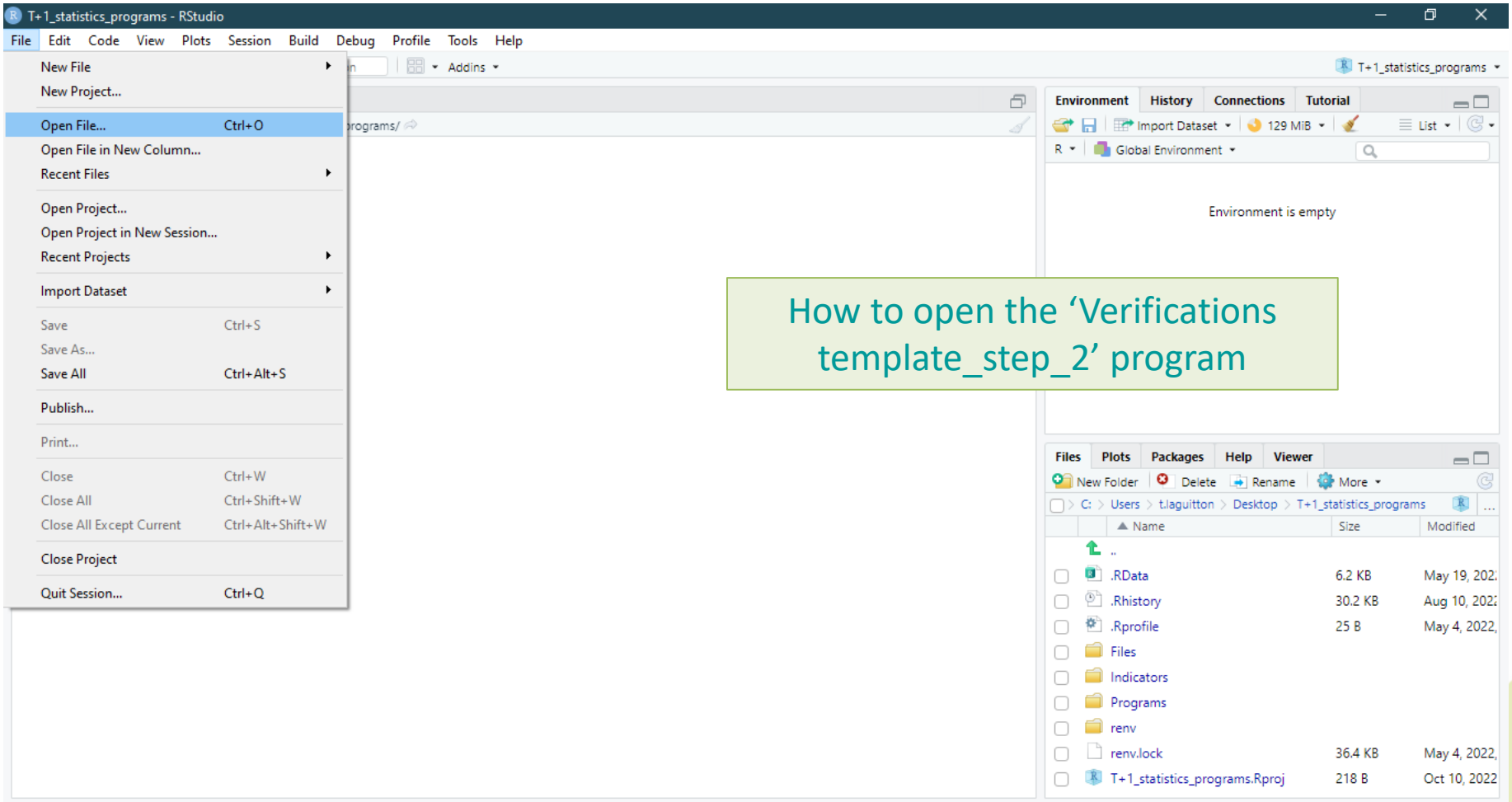
Requirements before starting the program 'Verifications template_step_2' :

- The program 'Verifications template_step_1' should have been run on your data
- You should no longer have any problems appearing or only problems that have been notified as verified after running the program 'Verifications template_step_1'
- You must have your template in your possession and it must now be called:
  T+1_data_collection_***country***_Step1_CORRECTED(X).csv (with the name of your own country)
  ((X) is the number of the last file exported and corrected after the last run of the first verification program)
- You need to make sure that the barcodes in your file T+1_data_collection_***country***_Step1_CORRECTED(X).csv appear in full and not in scientific format (see procedure pages 24→28)

Your Rstudio interface must have been cleaned up before running the program.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

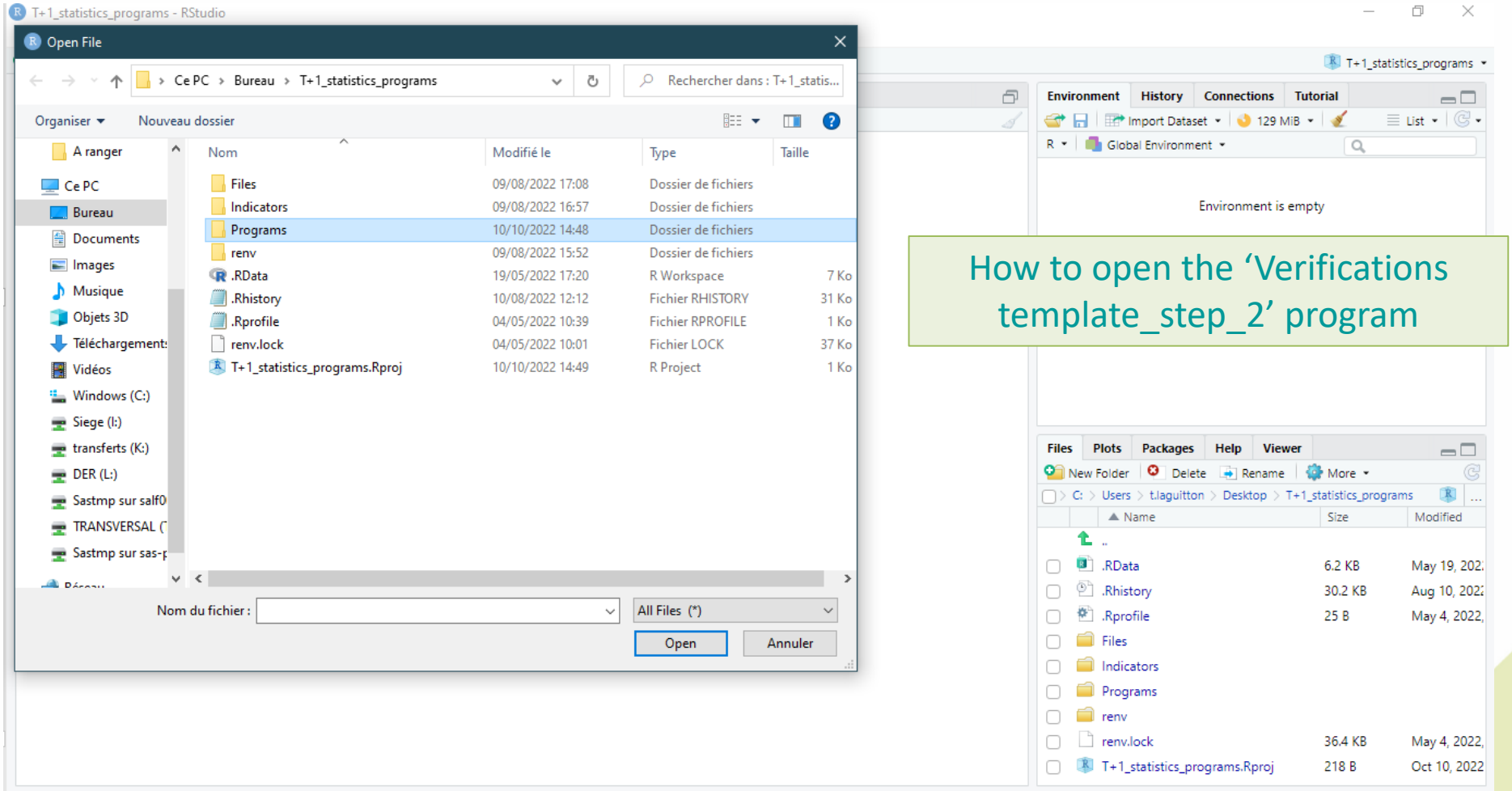
Running of ‘Verifications template_step_2’ program



How to open the ‘Verifications template_step_2’ program

Co-funded by the European Union's Health Programme (2014-2020)

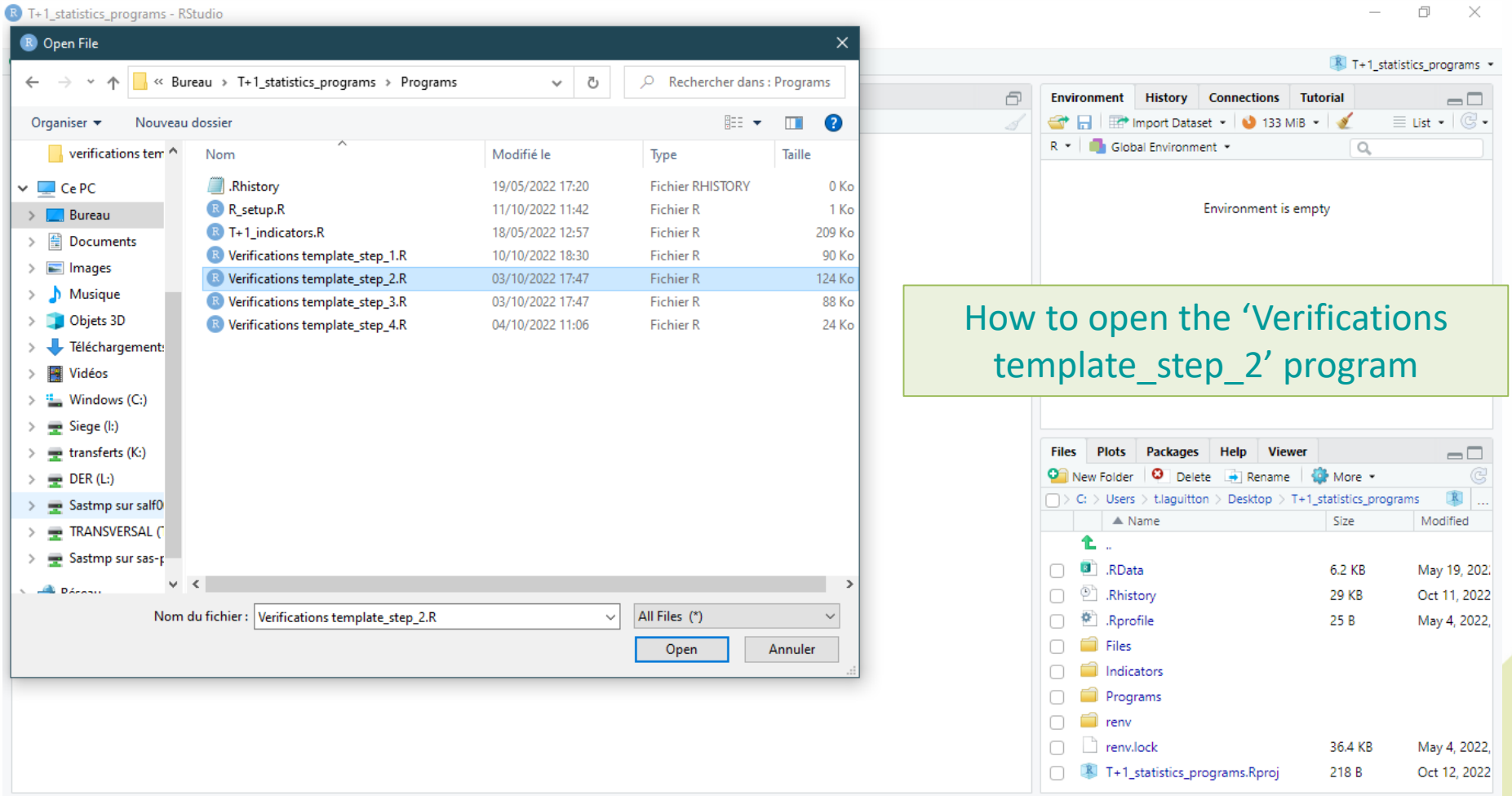
Running of ‘Verifications template_step_2’ program



How to open the ‘Verifications template_step_2’ program

Co-funded by the European Union's Health Programme (2014-2020)

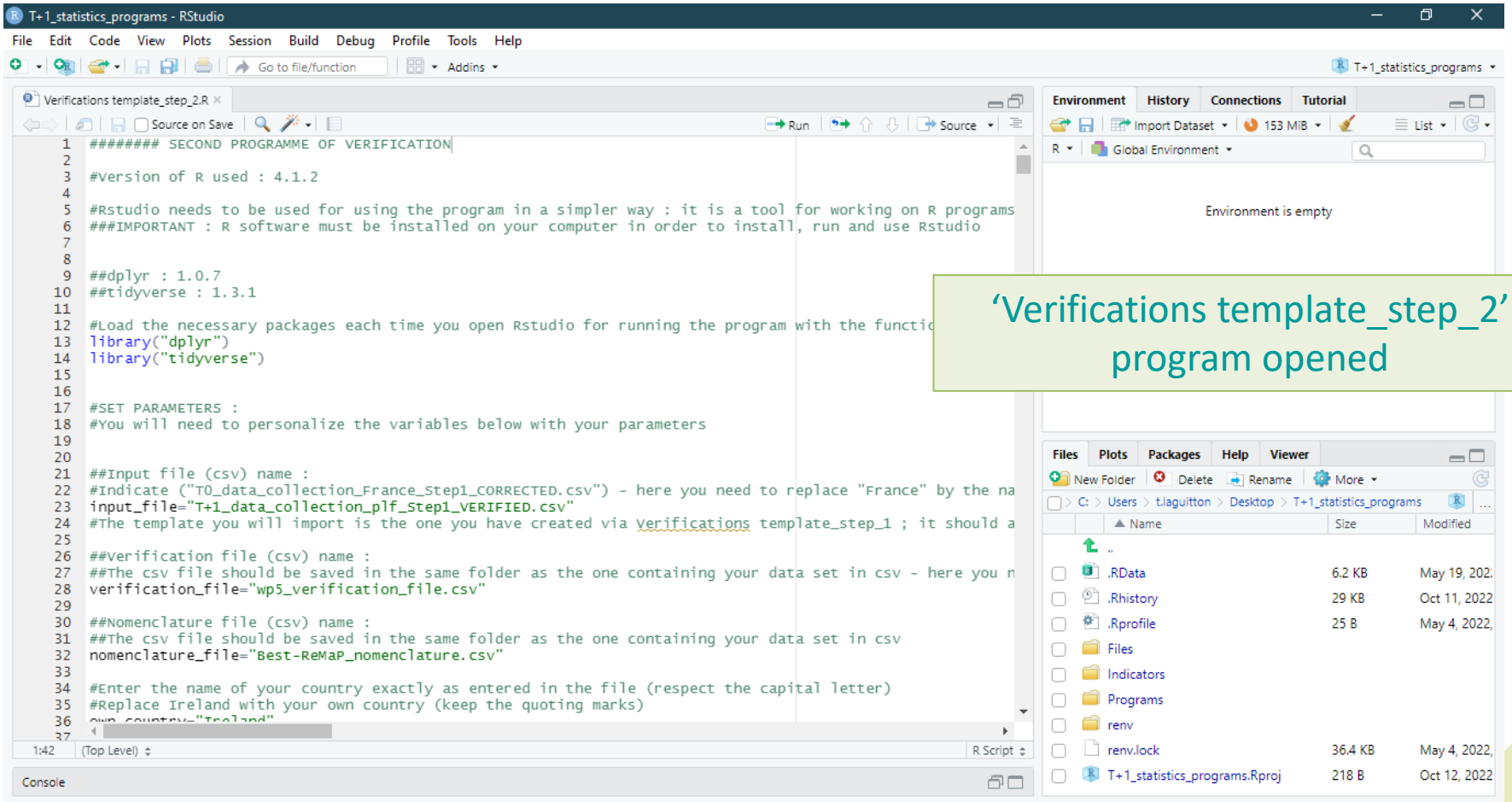
Running of ‘Verifications template_step_2’ program



How to open the ‘Verifications template_step_2’ program

Co-funded by the European Union's Health Programme (2014-2020)

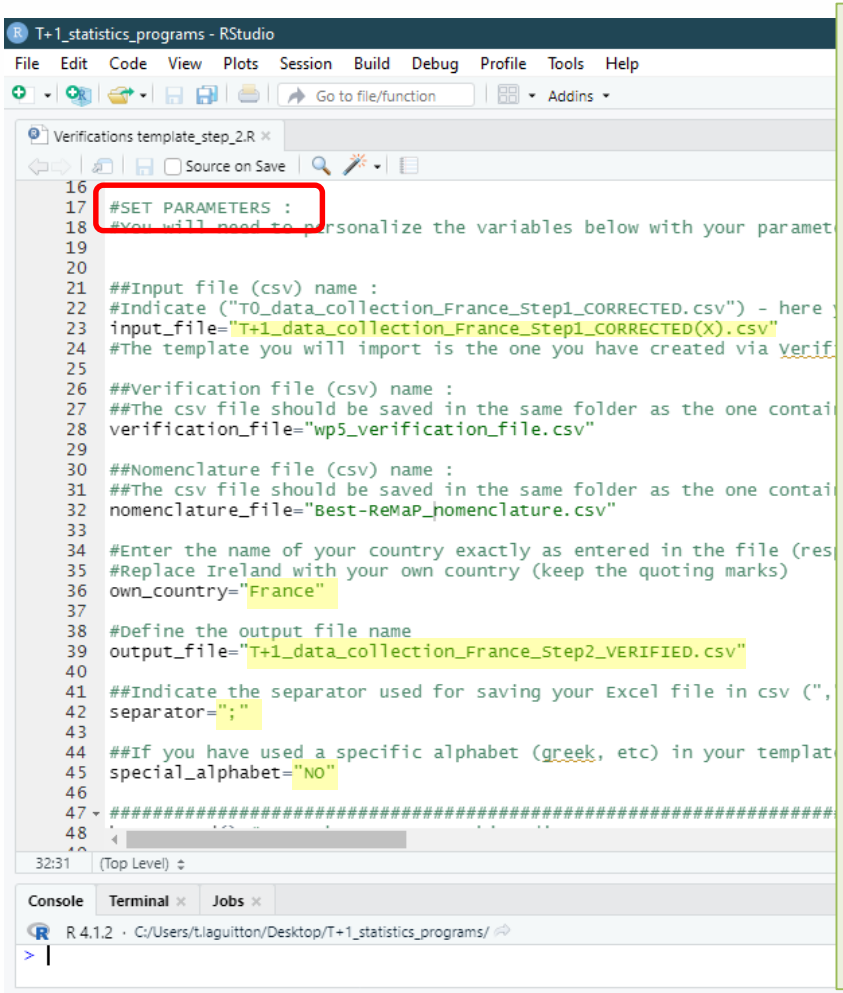
Running of ‘Verifications template_step_2’ program



‘Verifications template_step_2’ program opened

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_2' program



Setting parameters of the 2nd verification program

In the first run of the 2nd verification program, you need to change the name of the country with your own country name in the input file (line 23) and the output file (line 39) of the R script **+ line 36 ('own_country=')**

You must also indicate the separator and if you have a specific alphabet as in the program 'Verifications template_step_1'.

Example :

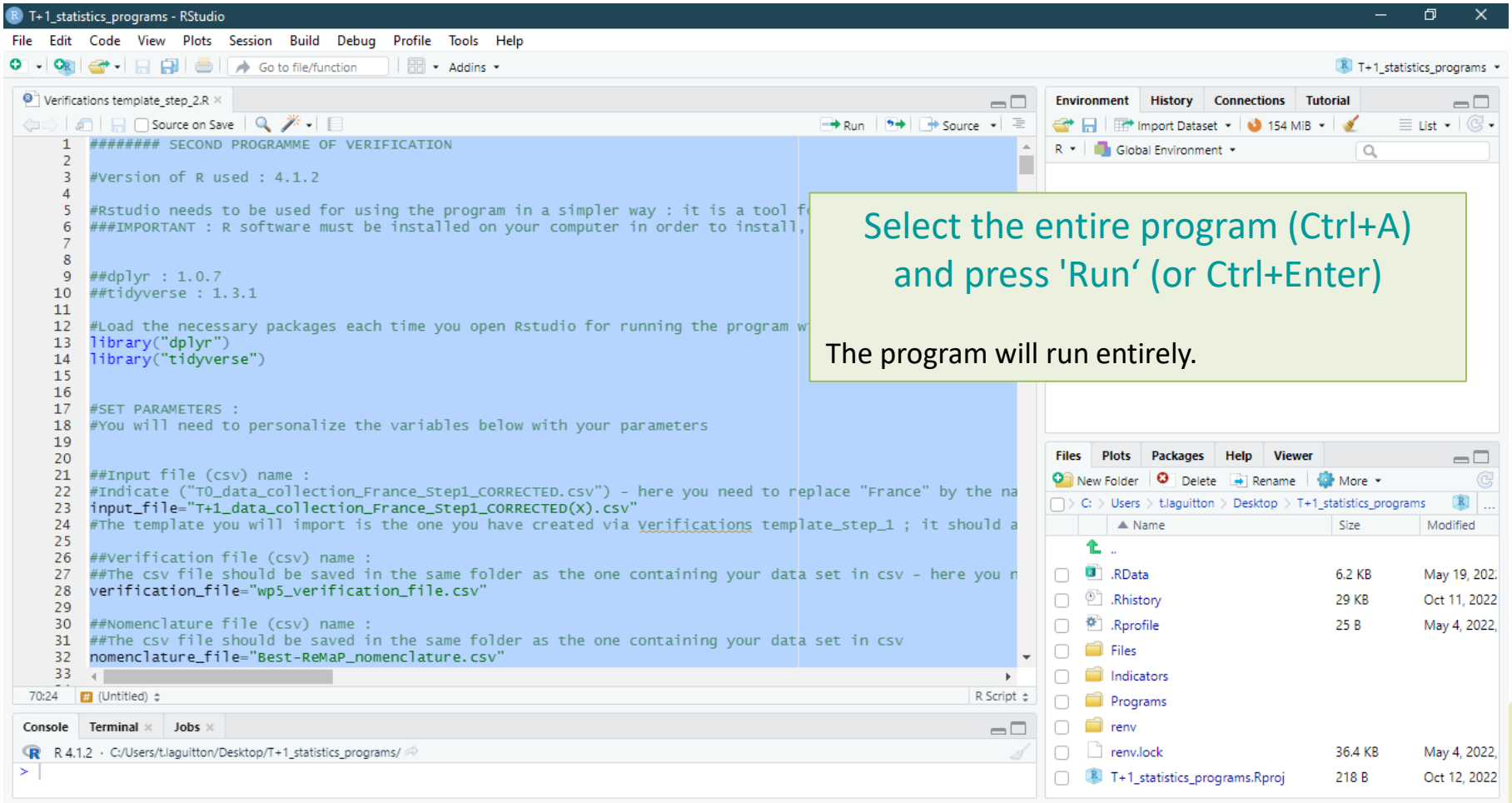
- **Input_file =** "T+1_data_collection_Ireland_**Step1**_CORRECTED(X).csv"

((X) is the number of the last file exported and corrected after the last run of the first verification program)

- **Own_country** = "Ireland" (same spelling as in the country name in the template)
- **Output_file =** "T+1_data_collection_Ireland_**Step2**_VERIFIED.csv"
- **Separator** = ";"
- **Special_alphabet** = "NO"

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_2' program

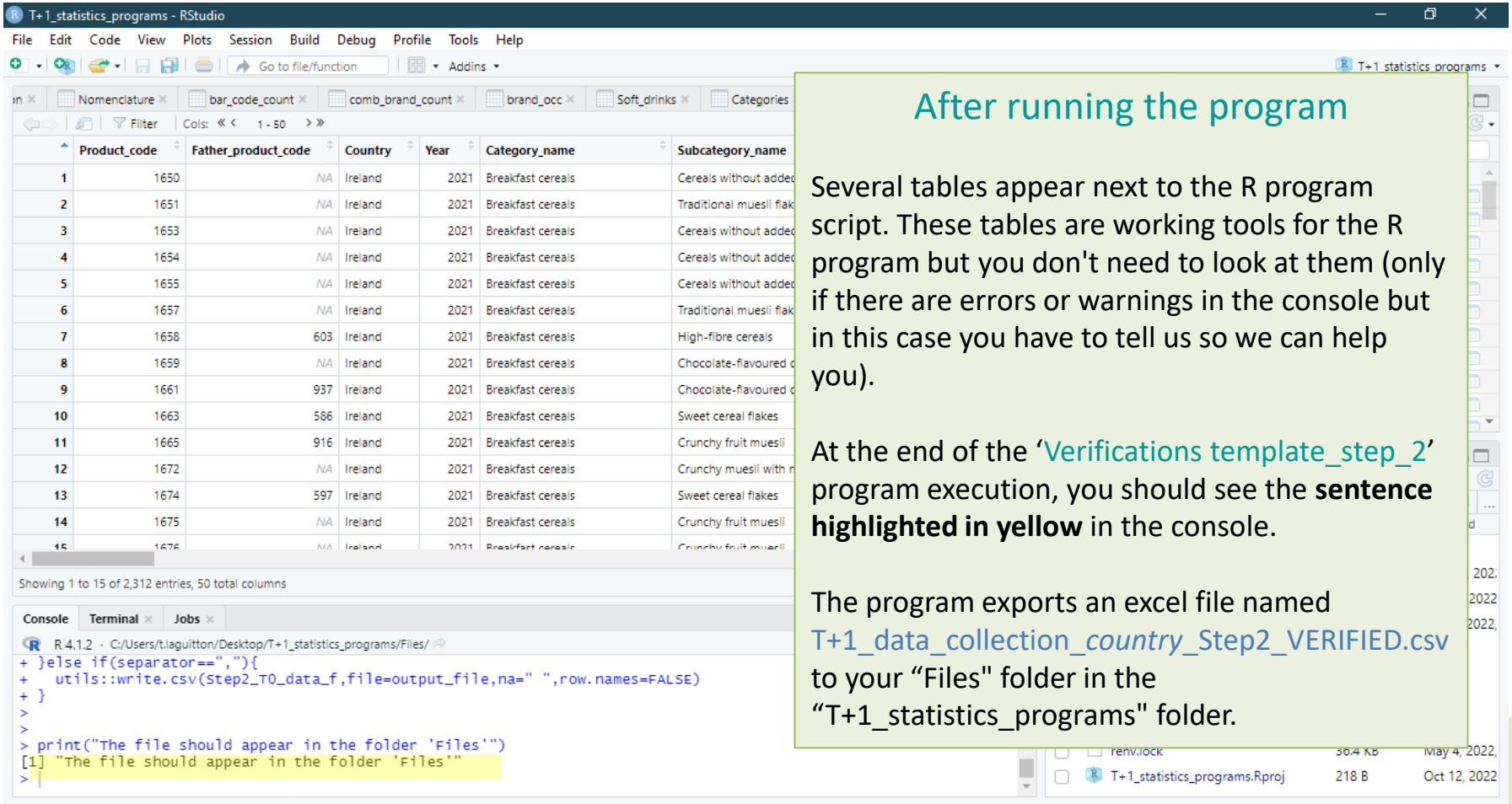


Select the entire program (Ctrl+A) and press 'Run' (or Ctrl+Enter)

The program will run entirely.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_2' program



After running the program

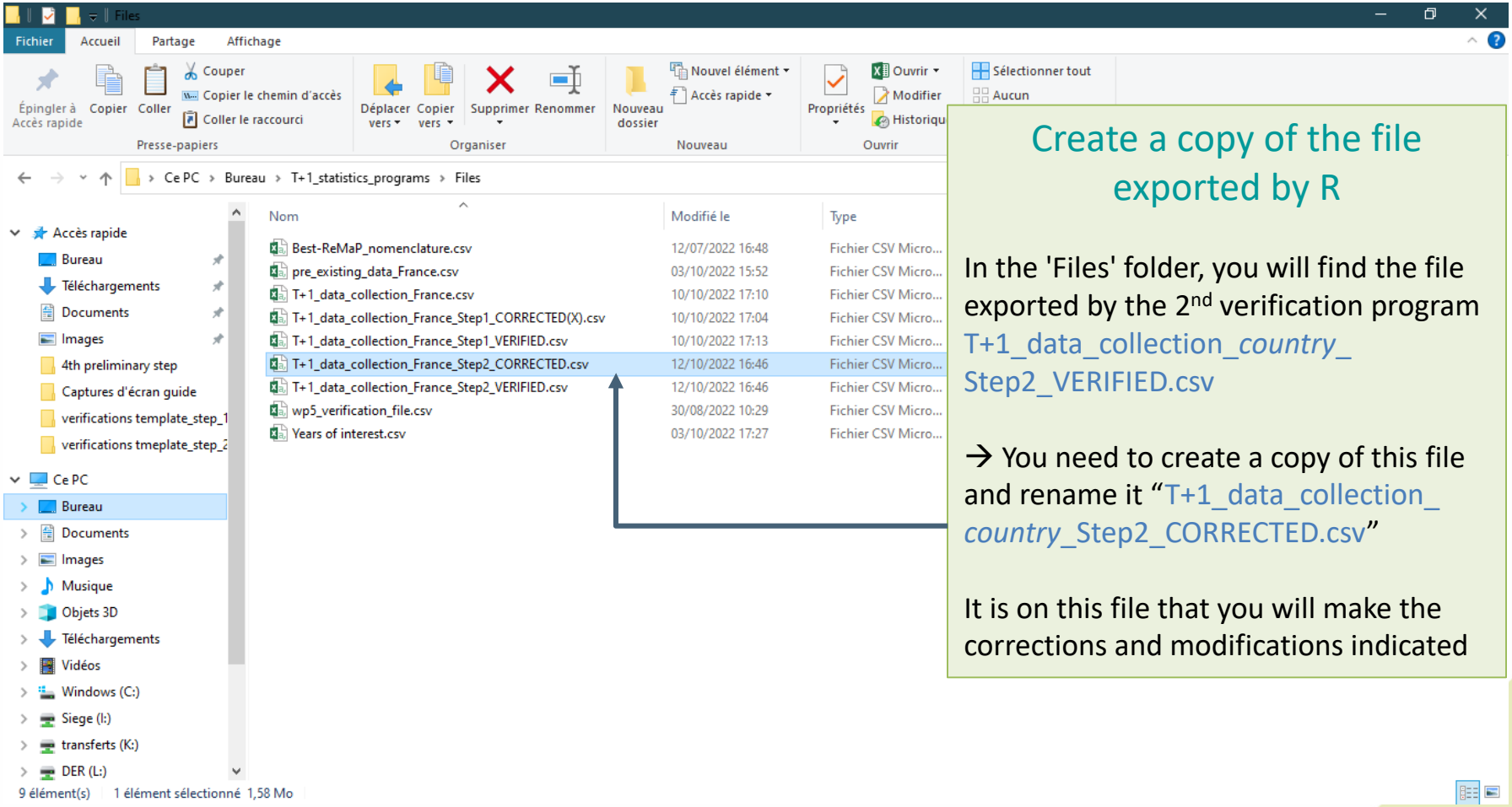
Several tables appear next to the R program script. These tables are working tools for the R program but you don't need to look at them (only if there are errors or warnings in the console but in this case you have to tell us so we can help you).

At the end of the 'Verifications template_step_2' program execution, you should see the **sentence highlighted in yellow** in the console.

The program exports an excel file named T+1_data_collection_*country*_Step2_VERIFIED.csv to your "Files" folder in the "T+1_statistics_programs" folder.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_2' program



Create a copy of the file exported by R

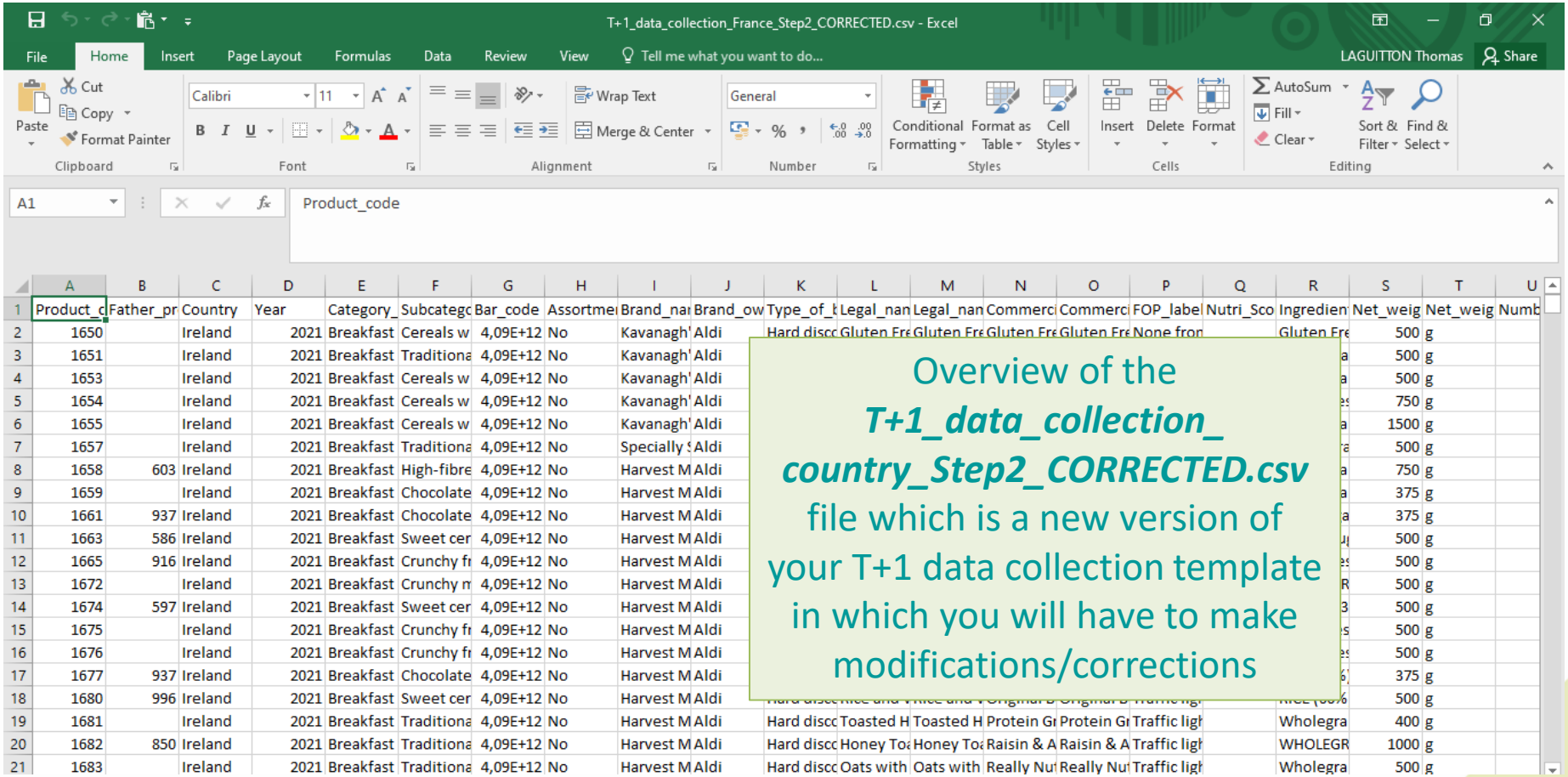
In the 'Files' folder, you will find the file exported by the 2nd verification program T+1_data_collection_*country*_Step2_VERIFIED.csv

→ You need to create a copy of this file and rename it "T+1_data_collection_*country*_Step2_CORRECTED.csv"

It is on this file that you will make the corrections and modifications indicated

Co-funded by the European Union's Health Programme (2014-2020)

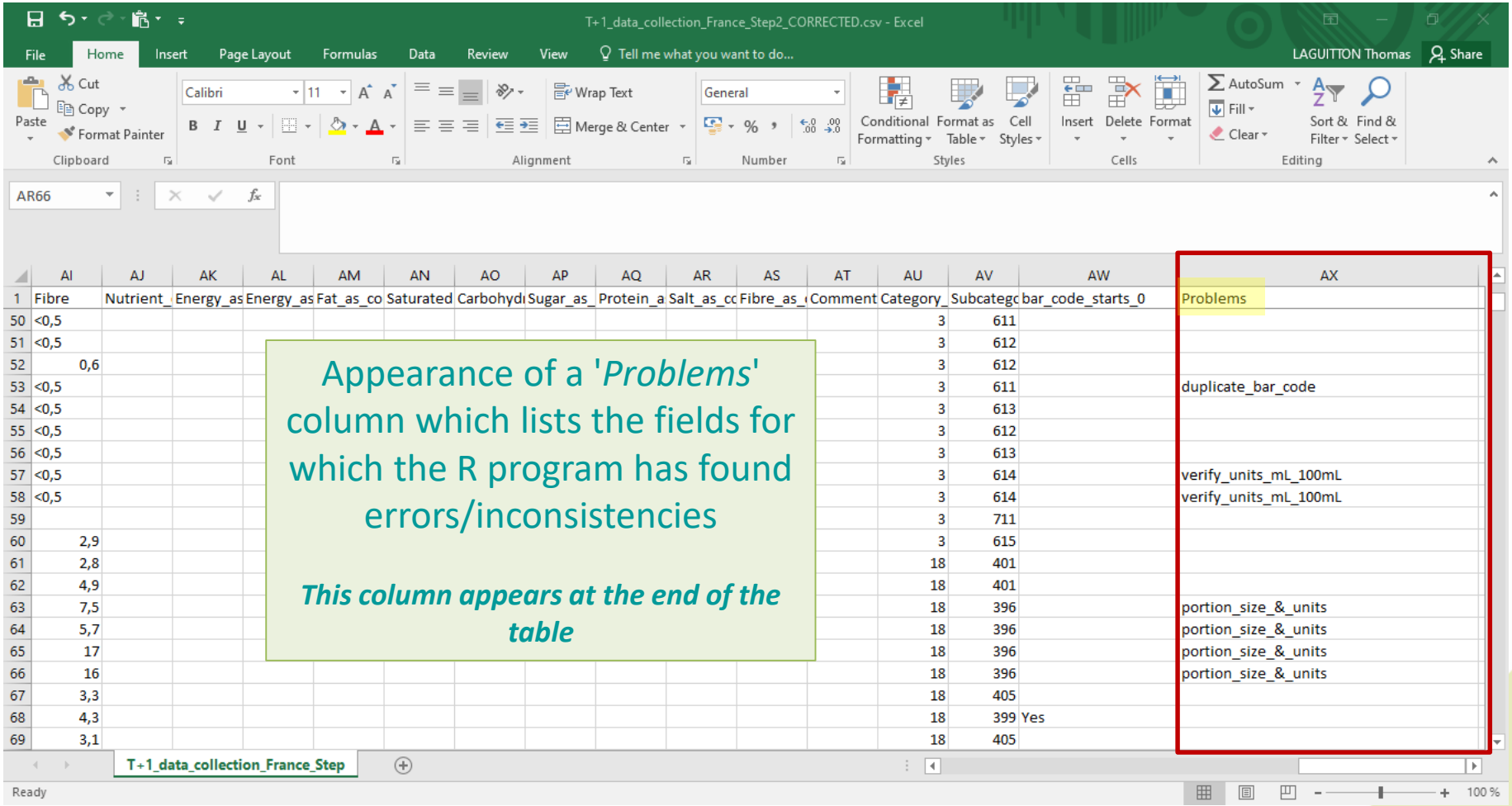
Excel file to modify after 'Verifications template_step_2' program



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Product_c	Father_pr	Country	Year	Category_	Subcategc	Bar_code	Assortmei	Brand_nai	Brand_ow	Type_of_k	Legal_nan	Legal_nan	Commerci	Commerci	FOP_labe	Nutri_Sco	Ingredien	Net_weig	Net_weig	Numb
2	1650		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi	Hard disc	Gluten Fre	Gluten Fre	Gluten Fre	Gluten Fre	None fron		Gluten Fre	500	g	
3	1651		Ireland	2021	Breakfast	Traditiona	4,09E+12	No	Kavanagh'	Aldi									500	g	
4	1653		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi									500	g	
5	1654		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi									750	g	
6	1655		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi									1500	g	
7	1657		Ireland	2021	Breakfast	Traditiona	4,09E+12	No	Specially S	Aldi									500	g	
8	1658	603	Ireland	2021	Breakfast	High-fibre	4,09E+12	No	Harvest M	Aldi									750	g	
9	1659		Ireland	2021	Breakfast	Chocolate	4,09E+12	No	Harvest M	Aldi									375	g	
10	1661	937	Ireland	2021	Breakfast	Chocolate	4,09E+12	No	Harvest M	Aldi									375	g	
11	1663	586	Ireland	2021	Breakfast	Sweet cer	4,09E+12	No	Harvest M	Aldi									500	g	
12	1665	916	Ireland	2021	Breakfast	Crunchy fr	4,09E+12	No	Harvest M	Aldi									500	g	
13	1672		Ireland	2021	Breakfast	Crunchy m	4,09E+12	No	Harvest M	Aldi									500	g	
14	1674	597	Ireland	2021	Breakfast	Sweet cer	4,09E+12	No	Harvest M	Aldi									500	g	
15	1675		Ireland	2021	Breakfast	Crunchy fr	4,09E+12	No	Harvest M	Aldi									500	g	
16	1676		Ireland	2021	Breakfast	Crunchy fr	4,09E+12	No	Harvest M	Aldi									500	g	
17	1677	937	Ireland	2021	Breakfast	Chocolate	4,09E+12	No	Harvest M	Aldi									375	g	
18	1680	996	Ireland	2021	Breakfast	Sweet cer	4,09E+12	No	Harvest M	Aldi									500	g	
19	1681		Ireland	2021	Breakfast	Traditiona	4,09E+12	No	Harvest M	Aldi	Hard disc	Toasted H	Toasted H	Protein Gr	Protein Gr	Traffic ligh		Wholegra	400	g	
20	1682	850	Ireland	2021	Breakfast	Traditiona	4,09E+12	No	Harvest M	Aldi	Hard disc	Honey Toa	Honey Toa	Raisin & A	Raisin & A	Traffic ligh		WHOLEGR	1000	g	
21	1683		Ireland	2021	Breakfast	Traditiona	4,09E+12	No	Harvest M	Aldi	Hard disc	Oats with	Oats with	Really Nut	Really Nut	Traffic ligh		Wholegra	500	g	

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program

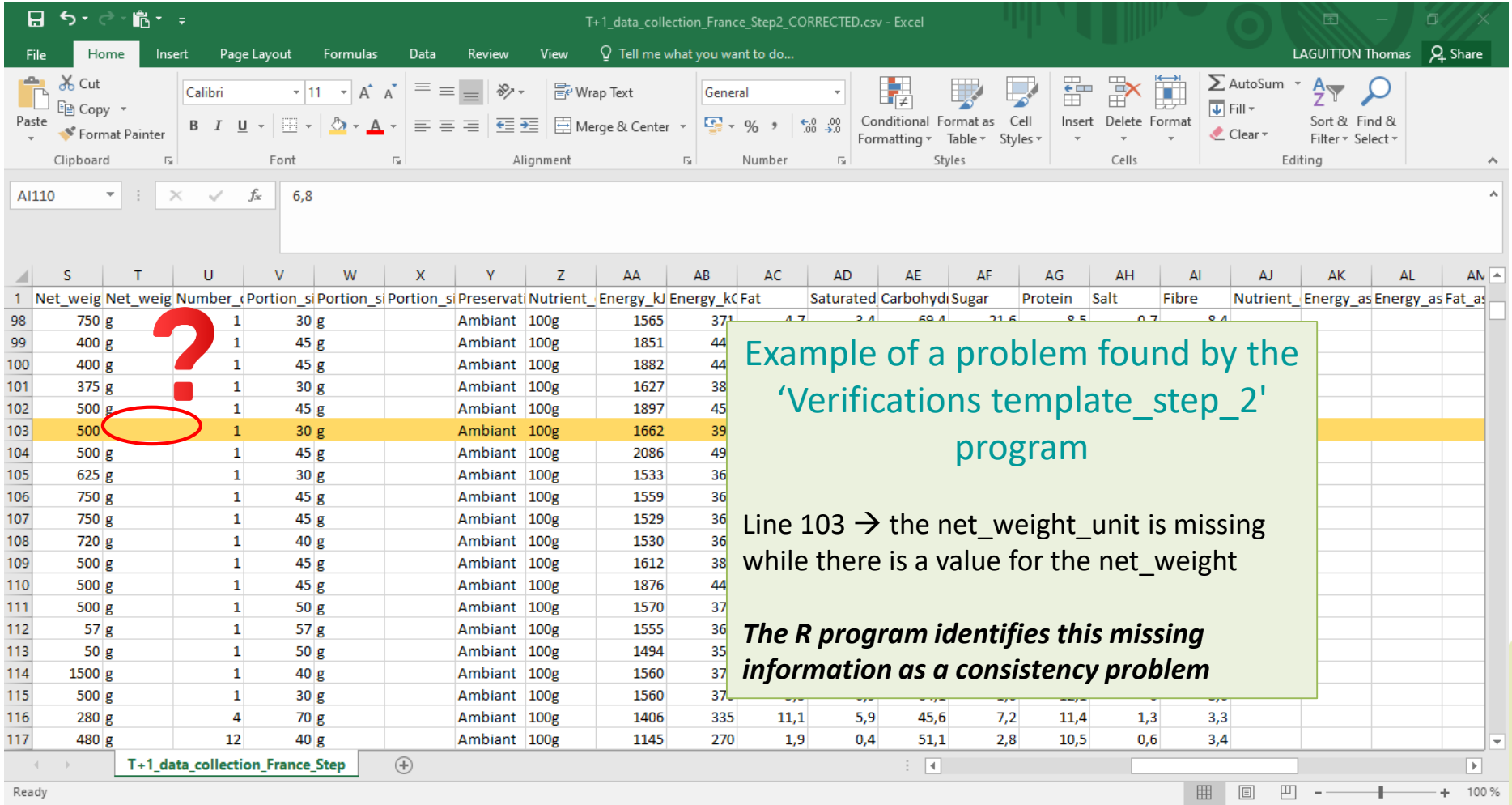


Appearance of a '*Problems*' column which lists the fields for which the R program has found errors/inconsistencies

This column appears at the end of the table

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program



	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1	Net_weig	Net_weig	Number_(	Portion_si	Portion_si	Portion_si	Preservati	Nutrient_(	Energy_kJ	Energy_kC	Fat	Saturated	Carbohydi	Sugar	Protein	Salt	Fibre	Nutrient_	Energy_as	Energy_as	Fat_as
98	750	g	1	30	g		Ambiant	100g	1565	371											
99	400	g	1	45	g		Ambiant	100g	1851	44											
100	400	g	1	45	g		Ambiant	100g	1882	44											
101	375	g	1	30	g		Ambiant	100g	1627	38											
102	500	g	1	45	g		Ambiant	100g	1897	45											
103	500		1	30	g		Ambiant	100g	1662	39											
104	500	g	1	45	g		Ambiant	100g	2086	49											
105	625	g	1	30	g		Ambiant	100g	1533	36											
106	750	g	1	45	g		Ambiant	100g	1559	36											
107	750	g	1	45	g		Ambiant	100g	1529	36											
108	720	g	1	40	g		Ambiant	100g	1530	36											
109	500	g	1	45	g		Ambiant	100g	1612	38											
110	500	g	1	45	g		Ambiant	100g	1876	44											
111	500	g	1	50	g		Ambiant	100g	1570	37											
112	57	g	1	57	g		Ambiant	100g	1555	36											
113	50	g	1	50	g		Ambiant	100g	1494	35											
114	1500	g	1	40	g		Ambiant	100g	1560	37											
115	500	g	1	30	g		Ambiant	100g	1560	37											
116	280	g	4	70	g		Ambiant	100g	1406	335	11,1	5,9	45,6	7,2	11,4	1,3	3,3				
117	480	g	12	40	g		Ambiant	100g	1145	270	1,9	0,4	51,1	2,8	10,5	0,6	3,4				

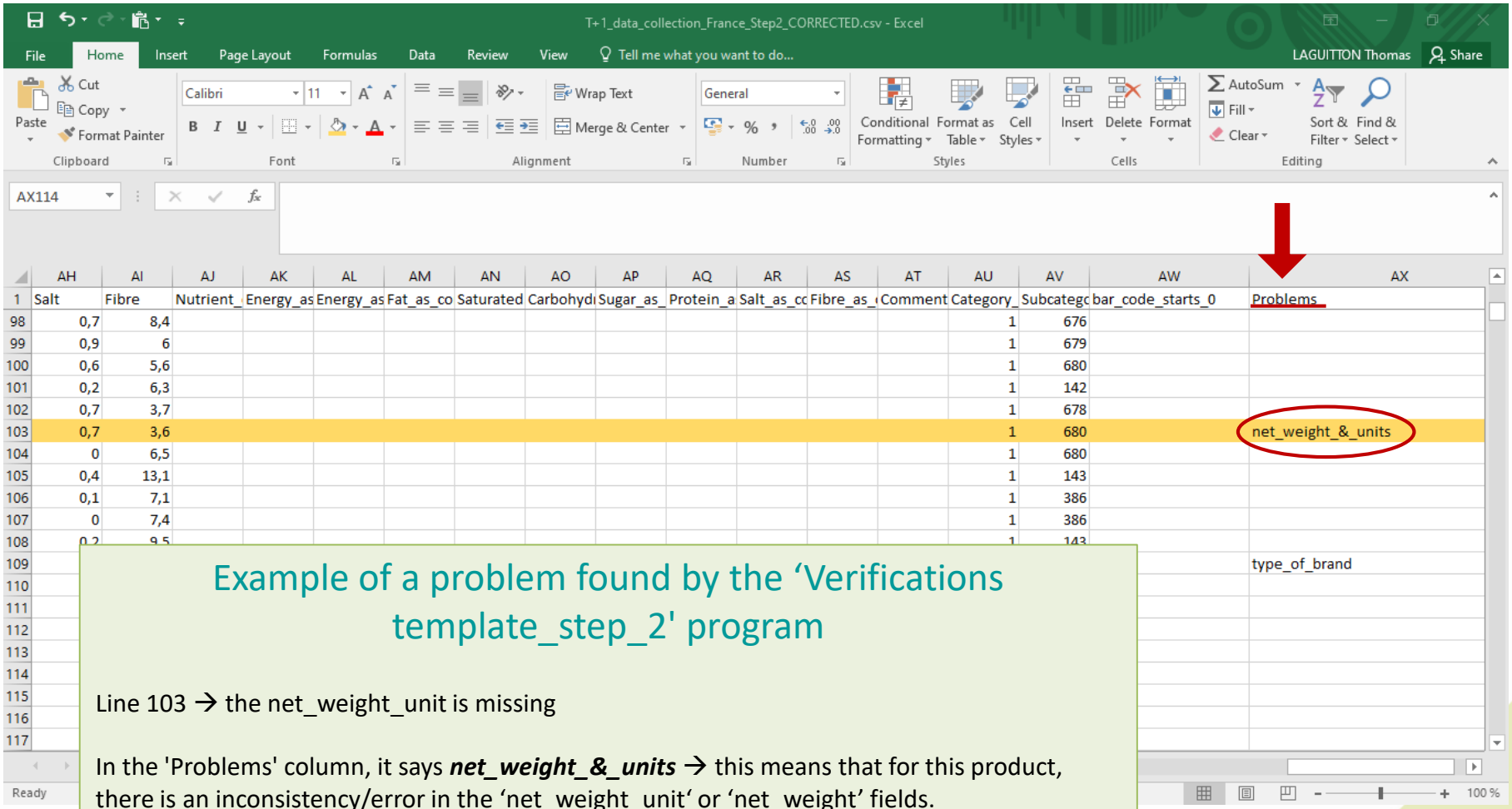
Example of a problem found by the 'Verifications template_step_2' program

Line 103 → the net_weight_unit is missing while there is a value for the net_weight

The R program identifies this missing information as a consistency problem

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program



	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX
1	Salt	Fibre	Nutrient_	Energy_as	Energy_as	Fat_as_co	Saturated	Carbohydi	Sugar_as_	Protein_a	Salt_as_cc	Fibre_as_(	Comment	Category_	Subcategc	bar_code_starts_0	Problems
98	0,7	8,4												1	676		
99	0,9	6												1	679		
100	0,6	5,6												1	680		
101	0,2	6,3												1	142		
102	0,7	3,7												1	678		
103	0,7	3,6												1	680		net_weight_&_units
104	0	6,5												1	680		
105	0,4	13,1												1	143		
106	0,1	7,1												1	386		
107	0	7,4												1	386		
108	0,2	9,5												1	143		
109																	type_of_brand
110																	
111																	
112																	
113																	
114																	
115																	
116																	
117																	

Example of a problem found by the 'Verifications template_step_2' program

Line 103 → the net_weight_unit is missing

In the 'Problems' column, it says ***net_weight_&_units*** → this means that for this product, there is an inconsistency/error in the 'net_weight_unit' or 'net_weight' fields.

You will have to fill in the net_weight_unit of this product

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_2' program, their meaning and what to do

Problem	Meaning	Action
nomenclature	Wrong association between 'Category_name', 'Category_code' , 'Subcategory_name' and 'Subcategory_code'	→ Check the 4 fields and correct those (or the one) that are not correctly associated
Net_weight_ &_units	• The **net weight** is filled but there is no associated **net weight unit** **or** • The **net weight unit** is filled but there is no associated **net weight**	→ If there is a value in the 'net_weight' field, you must add the unit « **g** » or « **mL** » in the 'net_weight_unit' field → If there is a unit in the field 'net_weight_unit', you must go back to the pictures of the product and add the value of the portion size. If there is no net weight for this product, it must be an input error and you can delete the unit. → When you have any doubts, you must go back to the pictures of the product to be sure of what to enter
Portion_size_ &_units	• The **portion size** is filled but there is no associated **portion size unit** **or** • The **portion size unit** is filled but there is no associated **portion size**	→ If there is a value in the 'portion_size' field, you must add the unit « **g** » or « **mL** » in the 'portion_size_unit' field → If there is a unit in the field 'portion_size_unit', you must go back to the pictures of the product and add the value of the portion size. If there is no portion size for this product, it must be an input error and you can delete the unit. → When you have any doubts, you must go back to the pictures of the product to be sure of what to enter

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_2' program, their meaning and what to do

Problem	Meaning	Action
Nutritional_values_&_units	• The **nutrient content expression unit** is filled but there are no associated **nutritional values** for the nutrients **or** • There are **nutritional values** for the nutrients but there is no associated **nutrient content expression unit**	→ If there are nutritional values for the nutrients, you must add the unit « **100g** » or « **100mL** » in the 'nutrient_content_expression_unit' field → If there is a unit in the field 'nutrient_content_expression_unit', you must go back to the pictures of the product and add the nutritional values of each nutrient. If there is nutritional values for this product, it must be an input error and you can delete the unit. → When you have any doubts, you must go back to the pictures of the product to be sure of what to enter
Nutritional_values_as_consumed_&_units	• The **nutrient content expression unit for products to be reconstituted** is filled but there is no associated **nutritional values for the nutrients as consumed** **or** • There are **nutritional values for the nutrients as consumed** but there is no associated **nutrient content expression unit for products to be reconstituted**	→ If there are nutritional values as consumed for the nutrients, you must add the unit « **100g** » or « **100mL** » in the 'nutrient_content_expression_unit_as_consumed' field → If there is a unit in the field 'nutrient_content_expression_unit_as_consumed', you must go back to the pictures of the product and add the nutritional values as consumed of each nutrient. If there is no nutritional values as consumed for this product, it must be an input error and you can delete the unit. → When you have any doubts, you must go back to the pictures of the product to be sure of what to enter

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_2' program, their meaning and what to do

Problem	Meaning	Action
Wrong_country	This is not the name of your country	→ You must enter the name of your own country, making sure that this is the same spelling as the closed list of the input template
Duplicate_bar_code	Same bar code has been found for 2 or more products	→ If the products have the same bar code and are **similar** (= duplicates = same bar code + same information for all the fields), you must delete one of the products to keep only one. → If the products have the same bar code but are **different**, you must check if it is an input error by going back to the pictures of the products. • If it is an input error, you must enter the correct bar code. • If the products really have the same bar code, you must keep them in the template and indicate in the 'Comments' field : « bar code checked and same for several different products »
Type_of_brand	The same brand has been associated with several types of brand. (This problem appears for all products of a same brand if they have been associated with different types of brand)	→ You must filter in the Excel file on a brand name that shows the problem « *Type_of_brand »,* then you must select the correct type of brand and apply it to all the products of the same brand name.

Excel file to modify after 'Verifications template_step_2' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_2' program, their meaning and what to do

Problem	Meaning	Action
Carbohydrates_or_sugar_content	The **sugar** content is greater than the **carbohydrates** content	→ You must go back to the pictures of the product and look at the sugar and carbohydrates content to correct this error. → If the error is on the product label, do not change anything and leave what is written on the product
Carbohydrates_or_sugar_as_consumed_content	The **sugar as consumed** content is greater than the **carbohydrates as consumed** content for products to be reconstituted	→ You must go back to the pictures of the product and look at the sugar and carbohydrates content as consumed to correct this error. → If the error is on the product label, do not change anything and leave what is written on the product
Fat_or_saturated_fat_content	The **saturated fat** content is greater than the **fat** content	→ You must go back to the pictures of the product and look at the fat and saturated fat content to correct this error. → If the error is on the product label, do not change anything and leave what is written on the product
Fat_or_saturated_fat_as_consumed_content	The **saturated fat as consumed** content is greater than the **fat as consumed** content for products to be reconstituted	→ You must go back to the pictures of the product and look at the fat and saturated fat content as consumed to correct this error. → If the error is on the product label, do not change anything and leave what is written on the product

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_2' program, their meaning and what to do

Problem	Meaning	Action
Verify_units_g_100g	a unit in **"g"** appears for a product belonging to the **'Soft drinks'** category	→ You must look at all the fields of the product that have units and find the unit "**g**". (Net_weight_unit, Portion_size_unit, Nutrient_content_expression_unit, Nutrient_content_expression_unit_as_consumed) → You must compare with the pictures of the product to check if this is an input error and correct it if necessary. → **It is not necessarily an input error as some milk beverages can have units in g.**
Verify_units_mL_100mL	a unit in **"mL"** appears for a product belonging to a **category other** than the 'Soft drinks' category	→ You must look at all the fields of the product that have units and find the unit "**mL**". (Net_weight_unit, Portion_size_unit, Nutrient_content_expression_unit, Nutrient_content_expression_unit_as_consumed) → You must compare with the pictures of the product to check if this is an input error and correct it if necessary. → **It is not necessarily an input error as some yoghourts can have units in mL.**

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_2' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_2' program, their meaning and what to do

Problem	Meaning	Action
Enter_Nutri_score	The 'FOP labeling type' field indicates *Nutriscore* but there is no associated nutri-score in the 'Nutriscore' field	→ You must go back to the pictures of the product and enter the letter of the nutri-score that appears on the package. → If there is no nutri-score on the picture, you must correct the entry in the field 'FOP_labeling_type' by choosing another FOP labeling type or *none of the list* (mandatory field)
Remove_Nutri_score	• A nutri score is filled in the 'Nutriscore' field but the 'FOP labeling type' does not indicate *Nutriscore*	→ You must go back to the pictures of the product and check if there is a nutri-score on the package • If there is a nutri-score on the package, you must indicate *Nutriscore* in the 'FOP_labeling_type' field and check that the letter of the nutri-score entered is the right one • If there is no nutri-score on the picture, you must delete the letter in the 'Nutriscore' field and and choose a FOP labeling type or *none of the list* in the field 'FOP_labeling_type' (mandatory field)

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after ‘Verifications template_step_2’ program

Terms that may appear in the 'Problems' field following the ‘Verifications template_step_2' program, their meaning and what to do

This problem can only occur if you are working on the latest version of the template which includes several fields for FOP labeling types (FOP_labeling_type2, FOP_labeling_type3, FOP_labeling_type4)

Problem	Meaning	Action
Incorrect_FOPs	The first field ‘FOP_labeling_type’ indicates *None from the list* but not the other fields ‘FOP_labeling_type2/3/4’. When the first field ‘FOP_labeling_type’ indicates *None from the list*, the other fields ‘FOP_labeling_type2/3/4’ must also indicate *None from the list.*	→ If one or more of the fields ‘FOP_labeling_2/3/4’ is empty, then you must enter *None from the list* in those fields. → If another FOP label (nutriscore, reference intake, ...) is indicated in the field ‘FOP_labeling_type2/3 or 4’, you have to check on the product pictures that this is not an error and if there is indeed an FOP label on the product, it must be indicated in the first field 'FOP_labeling_type' and the other 2/3/4 must indicate 'none from the list'.

Co-funded by the European Union's Health Programme (2014-2020)

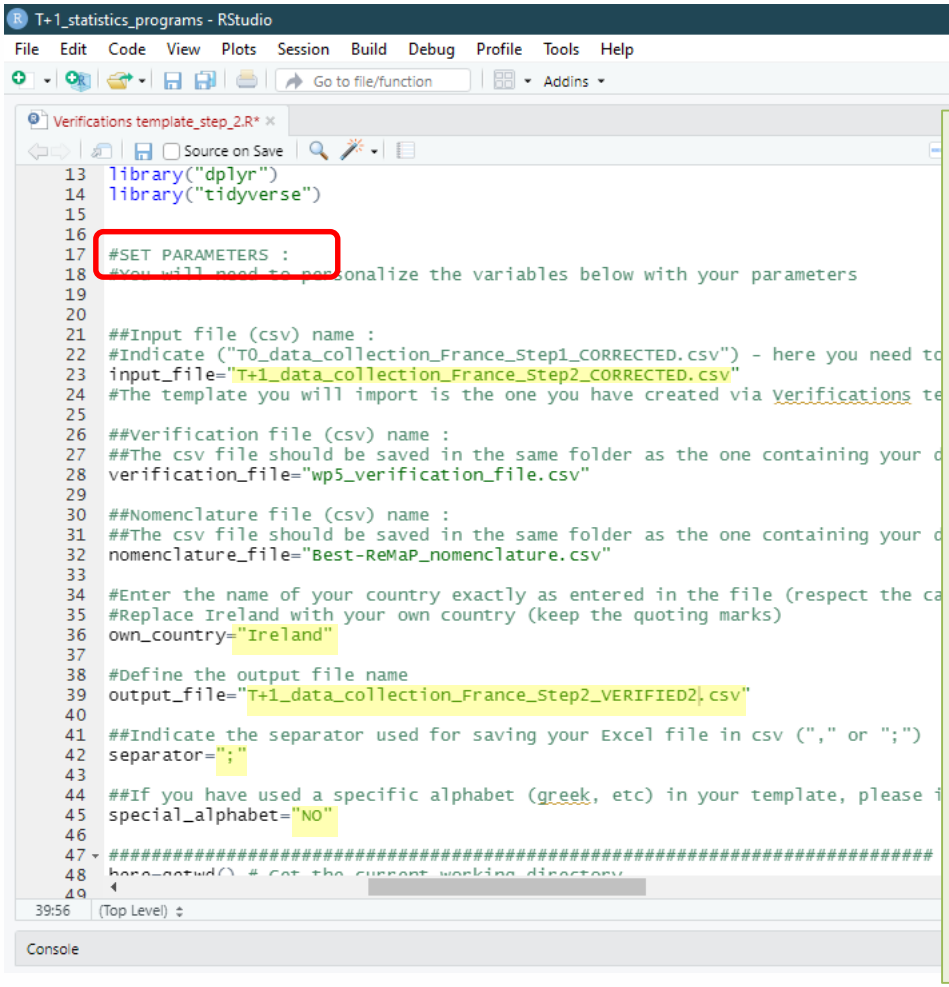
Excel file to modify after ‘Verifications template_step_2’ program

- **Be careful! When saving the file, you must be sure that the barcodes appear in full and not in scientific format. If this is the case, you must repeat the procedure on pages 24→28.**
- Once the corrections have been made and the barcodes are in the correct format, you can save the file *T+1_data_collection_country_Step2_CORRECTED.csv* and close it.
- You will have to run this corrected file again in the ‘Verifications template_step_2' program to make sure you haven't missed a check.

Your Rstudio interface must be cleaned up before running the ‘Verifications template_step_2’ program again.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_2' program



Setting parameters of the second running of 'Verifications template_step_2' program

For this second running of the 'Verifications template_step 2', the only fields you need to change are the names of the input file and the output file.

Input_file =
"*T+1_data_collection_country_**Step2**_CORRECTED.csv*"

Output_file =
"*T+1_data_collection_country_**Step2**_VERIFIED2.csv*"

Make sure the country name on line 36 is yours and if not, enter it.

You must also indicate the separator and if you have a specific alphabet as in the first run of the program.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_2' program

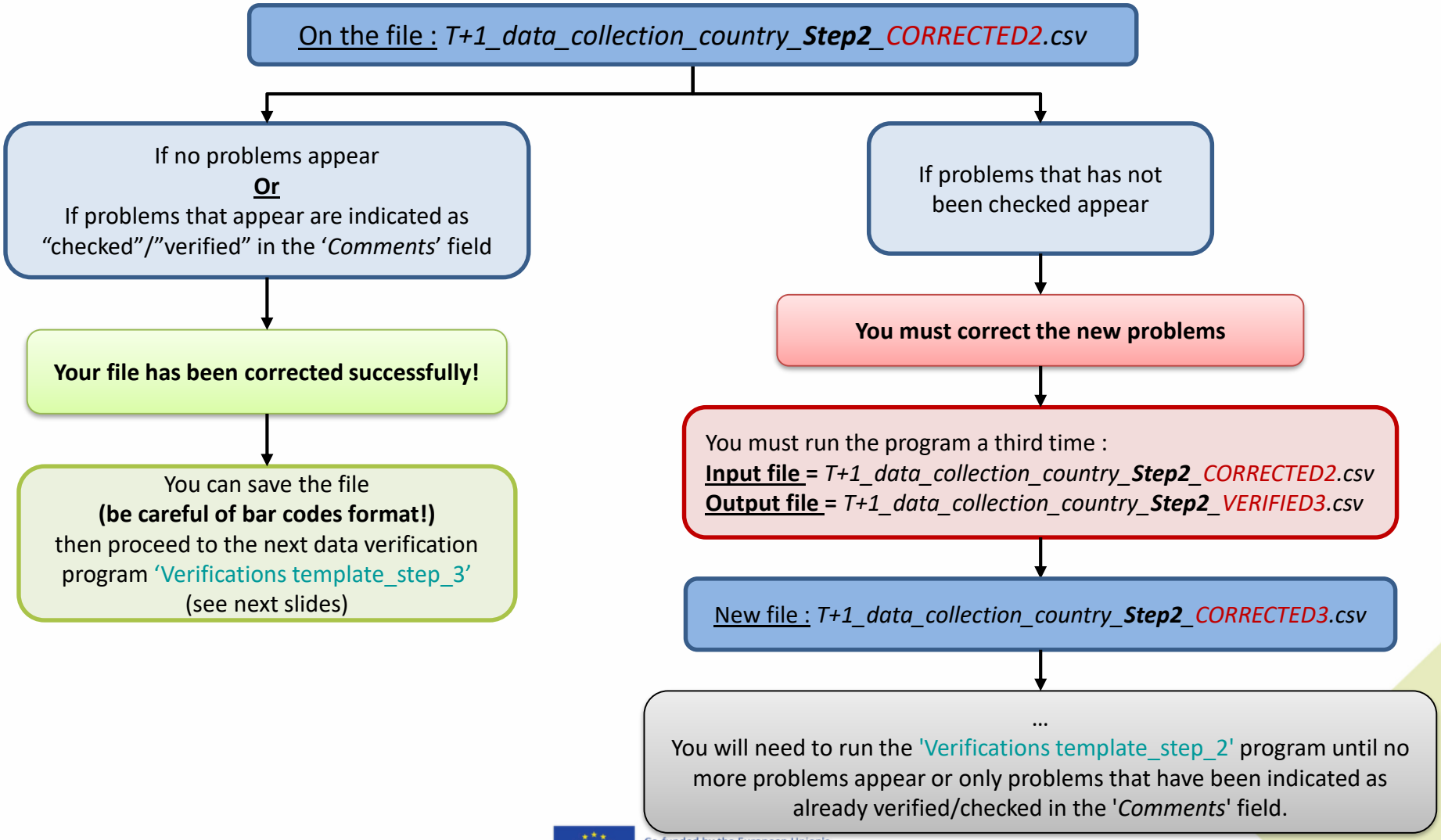
- At the end of this second run, you get in your "files" folder a file called: "*T+1_data_collection_country_**Step2**_VERIFIED2.csv*"

- You must create a copy of this file and call it : "*T+1_data_collection_country_**Step2**_CORRECTED2.csv*"

  → It is on this file that you will make the modifications following the checks

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_2' program



Co-funded by the European Union's Health Programme (2014-2020)

3) Running of the programs

A. Part 1 : R setup program

B. Part 2 : Verification programs and template cleaning/standardization

i. 1st verification program : 'Verifications template_step_1'

ii. 2nd verification program : 'Verifications template_step_2'

iii. 3rd verification program : 'Verifications template_step_3'

iv. 4th verification program : 'Verifications template_step_4'

Co-funded by the European Union's Health Programme (2014-2020)

3rd verification program : 'Verifications template_step_3'

Presentation of the 'Verifications template step_3' program :

- Third verification program : verification of outliers in your nutritional values

- For each nutrient in each subcategory, the following position indicators will be calculated :
  - ❑ 1st quartile (Q1)
  - ❑ 3rd quartile (Q3)
  - ❑ Interquartile range (IQR=Q3-Q1)

  → Nutrient values will be considered outliers if they are below Q1-(IQR*1.5) and above Q3+(IQR*1.5).

Co-funded by the European Union's Health Programme (2014-2020)

3rd verification program : 'Verifications template_step_3'

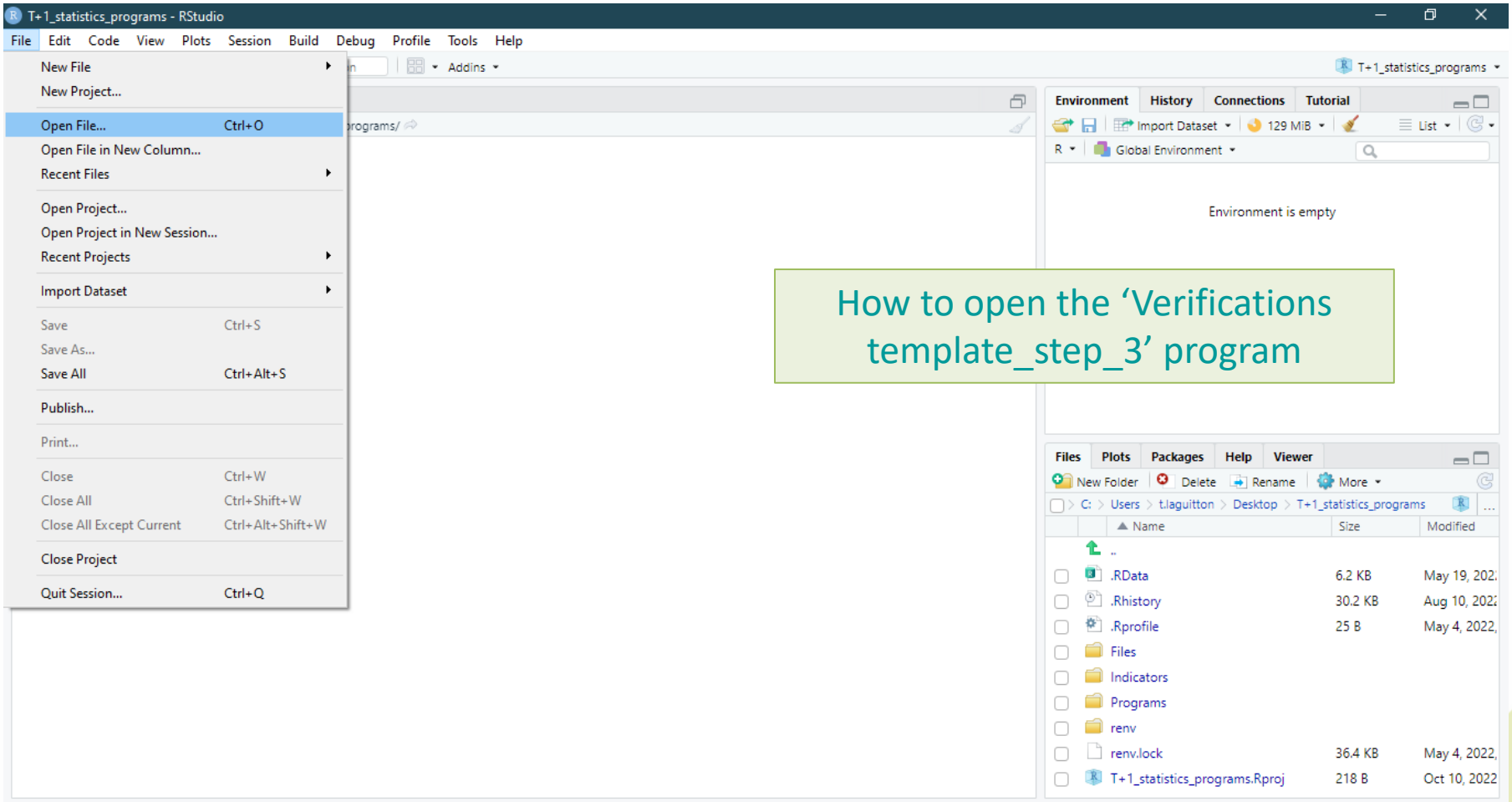
Requirements before starting the program 'Verifications template_step_3' :

- The programs 'Verifications template_step_1' and 'Verifications template_step_2' should have been run on your data
- You should no longer have any problems appearing or only problems that have been notified as verified after running the program 'Verifications template_step_2'
- You must have your template in your possession and it must now be called:
  T+1_data_collection_***country***_Step2_CORRECTED(X).csv (with the name of your own country)
  ((X) is the number of the last file exported and corrected after the last run of the first verification program)
- You need to make sure that the barcodes in your file T+1_data_collection_***country***_Step2_CORRECTED(X).csv appear in full and not in scientific format (see procedure pages 24→28)

Your Rstudio interface must have been cleaned up before running the program.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

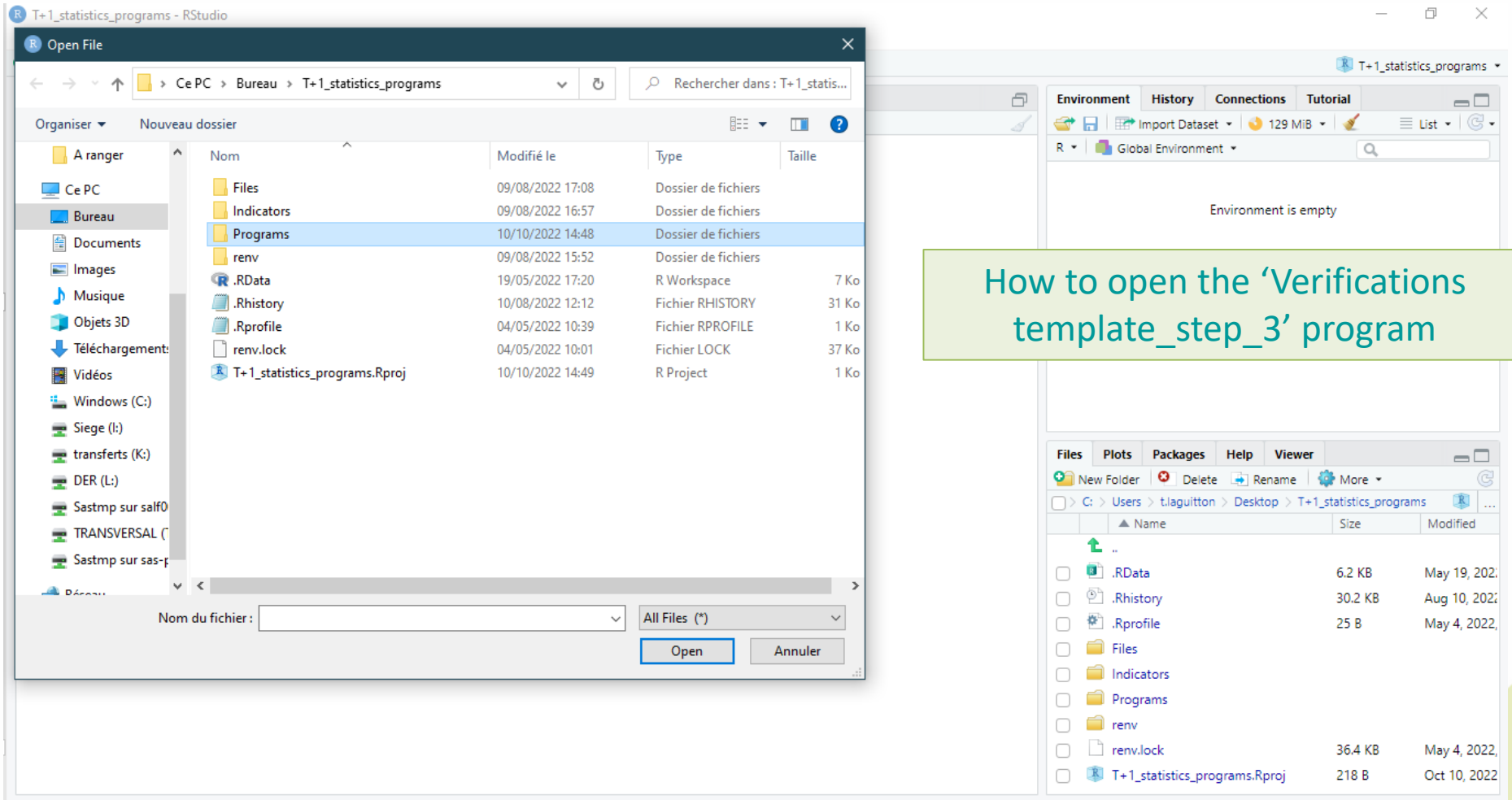
Running of ‘Verifications template_step_3’ program



How to open the ‘Verifications template_step_3’ program

Co-funded by the European Union's Health Programme (2014-2020)

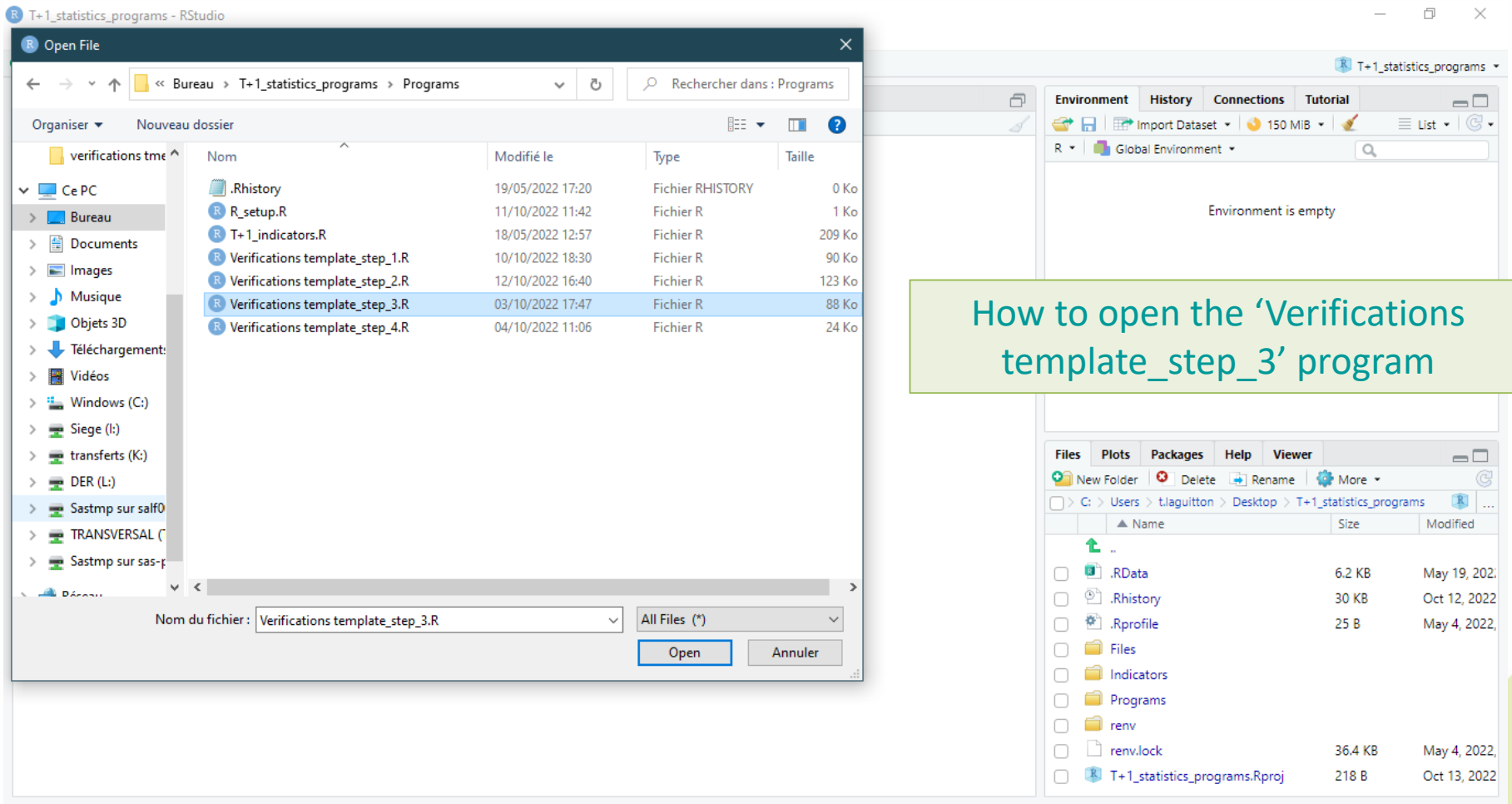
Running of ‘Verifications template_step_3’ program



How to open the ‘Verifications template_step_3’ program

Co-funded by the European Union's Health Programme (2014-2020)

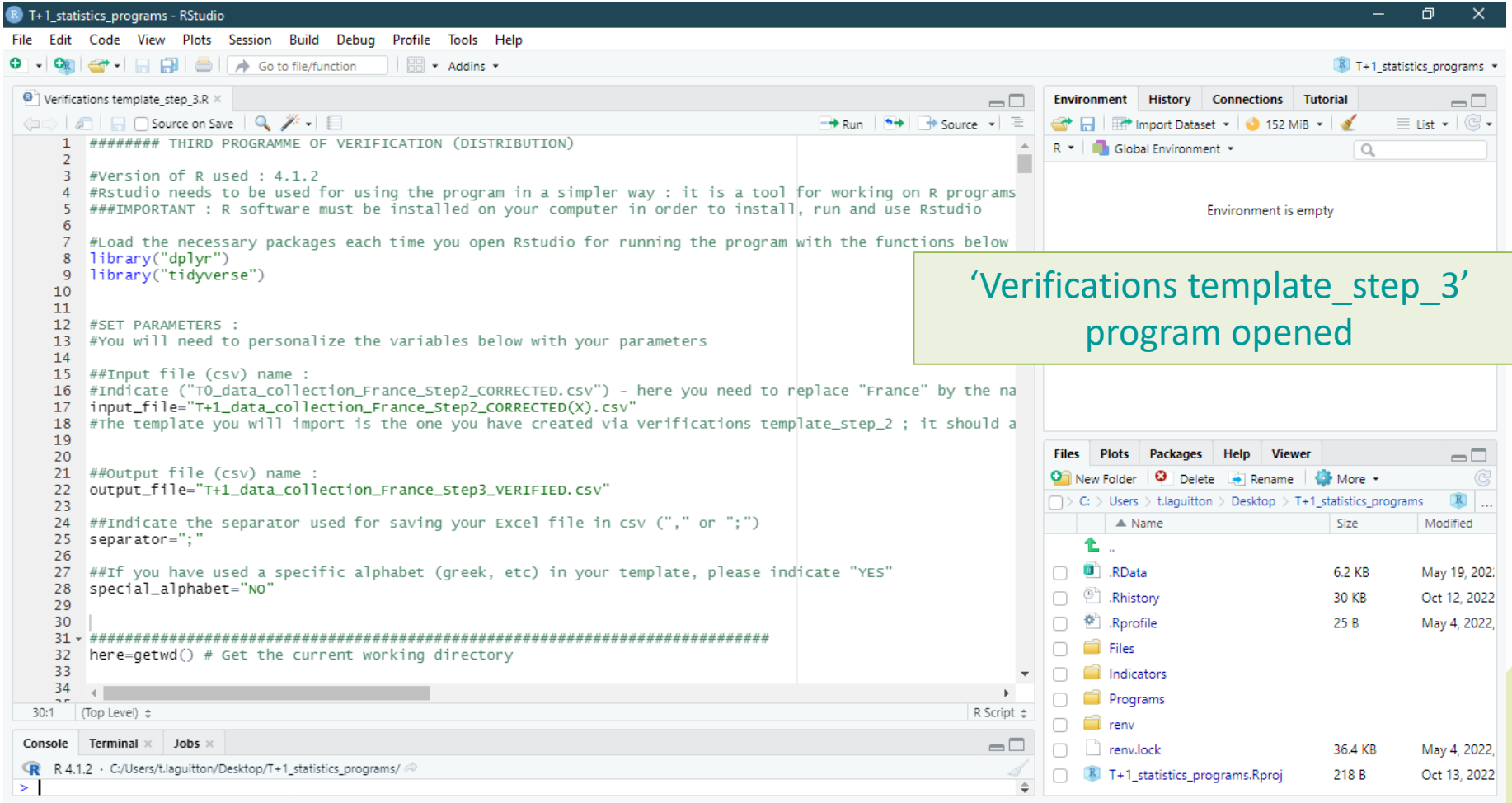
Running of ‘Verifications template_step_3’ program



How to open the ‘Verifications template_step_3’ program

Co-funded by the European Union's Health Programme (2014-2020)

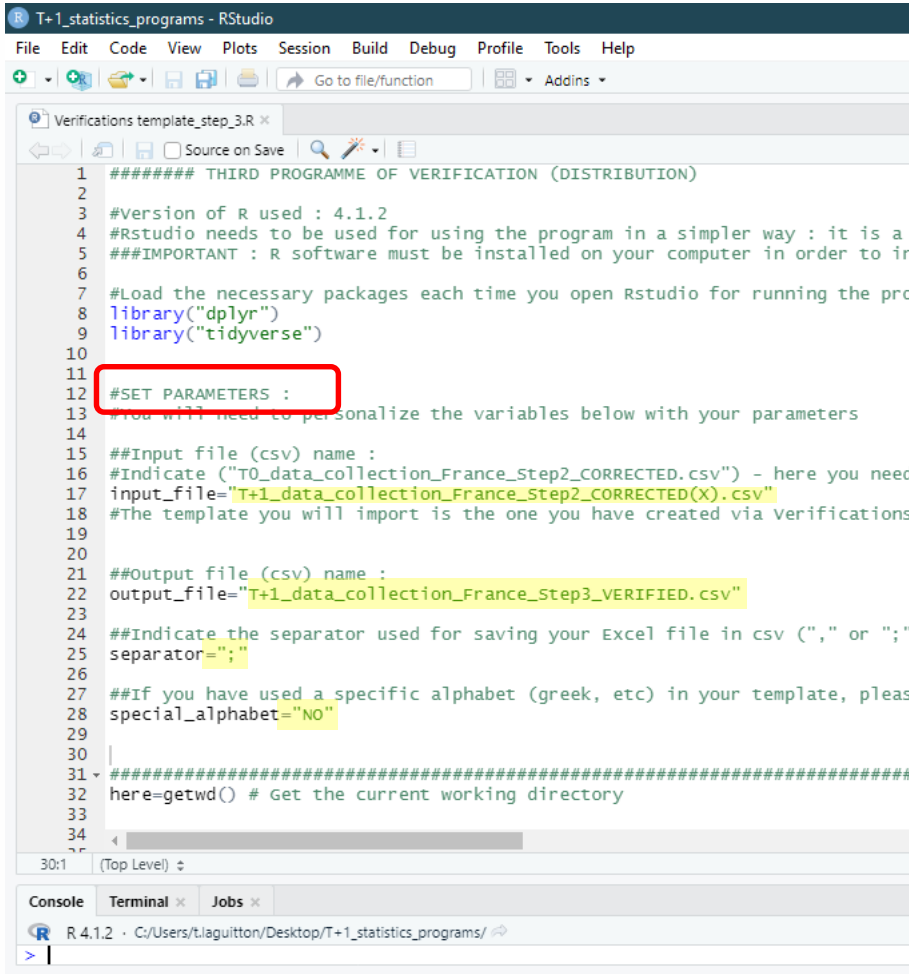
Running of 'Verifications template_step_3' program



'Verifications template_step_3' program opened

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_3' program



Setting parameters of the 3rd verification program

In the first run of the 3rd verification program, you need to change the name of the country with your own country name in the input file (line 17) and the output file (line 22)

You must also indicate the separator and if you have a specific alphabet as in the program 'Verifications template_step_1' and 'Verifications template_step_2'.

Example :

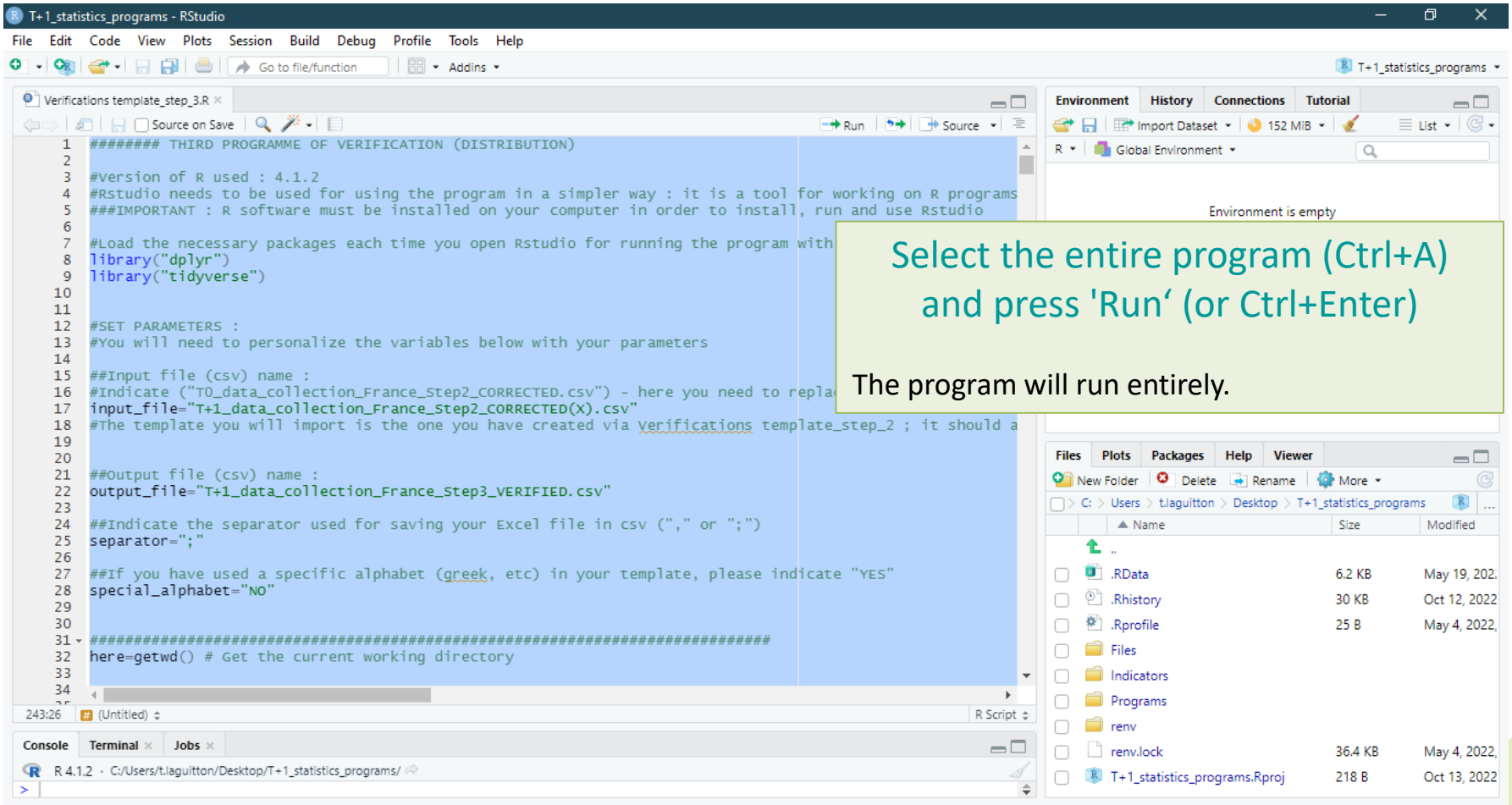
- **Input_file =** "T+1_data_collection_Ireland_Step2_CORRECTED(X).csv"

((X) is the number of the last file exported and corrected after the last run of the second verification program)

- **Output_file =** "T+1_data_collection_Ireland_Step3_VERIFIED.csv"
- **Separator** = ";"
- **Special_alphabet** = "NO"

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_3' program

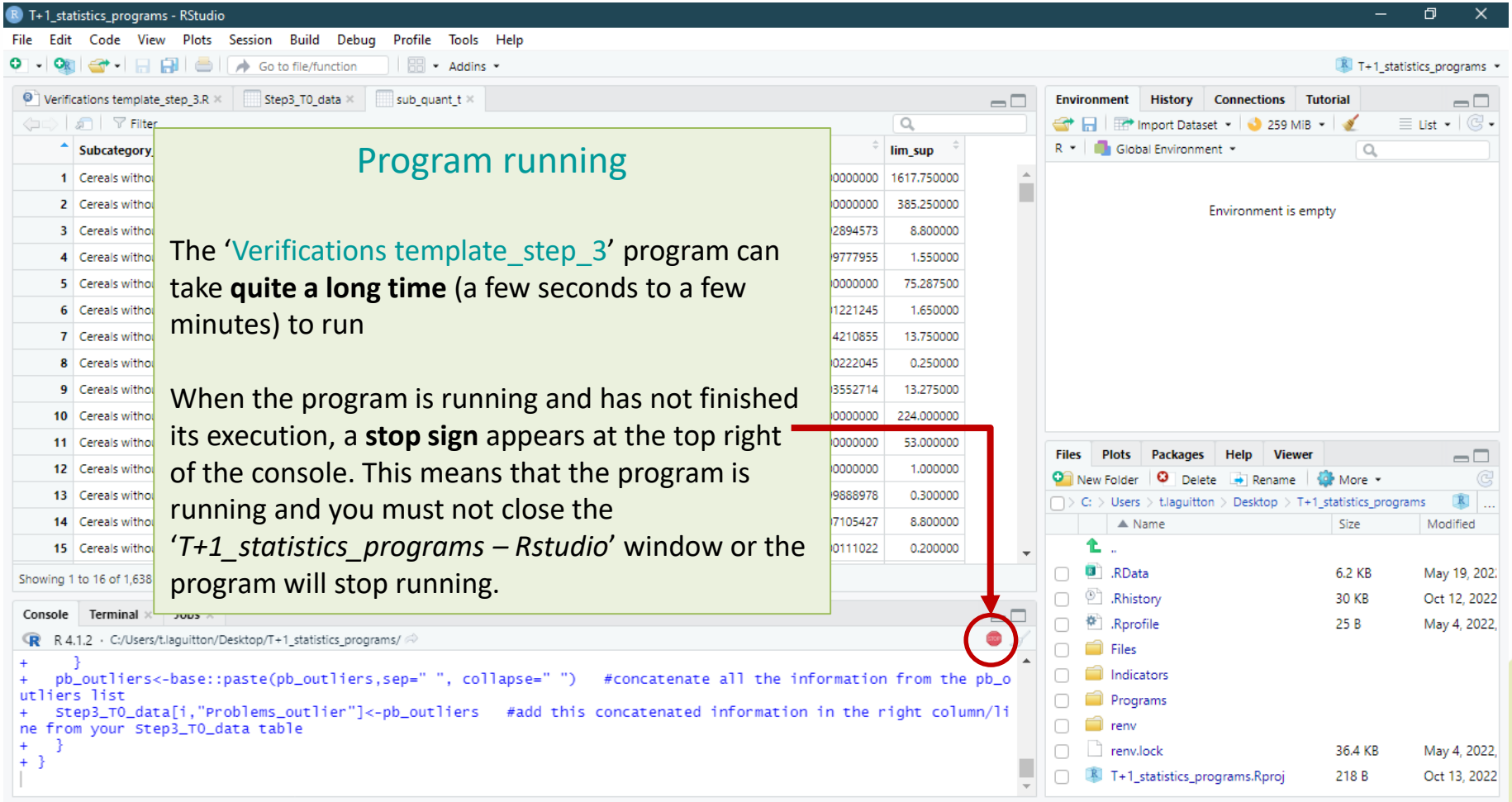


Select the entire program (Ctrl+A) and press 'Run' (or Ctrl+Enter)

The program will run entirely.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_3' program



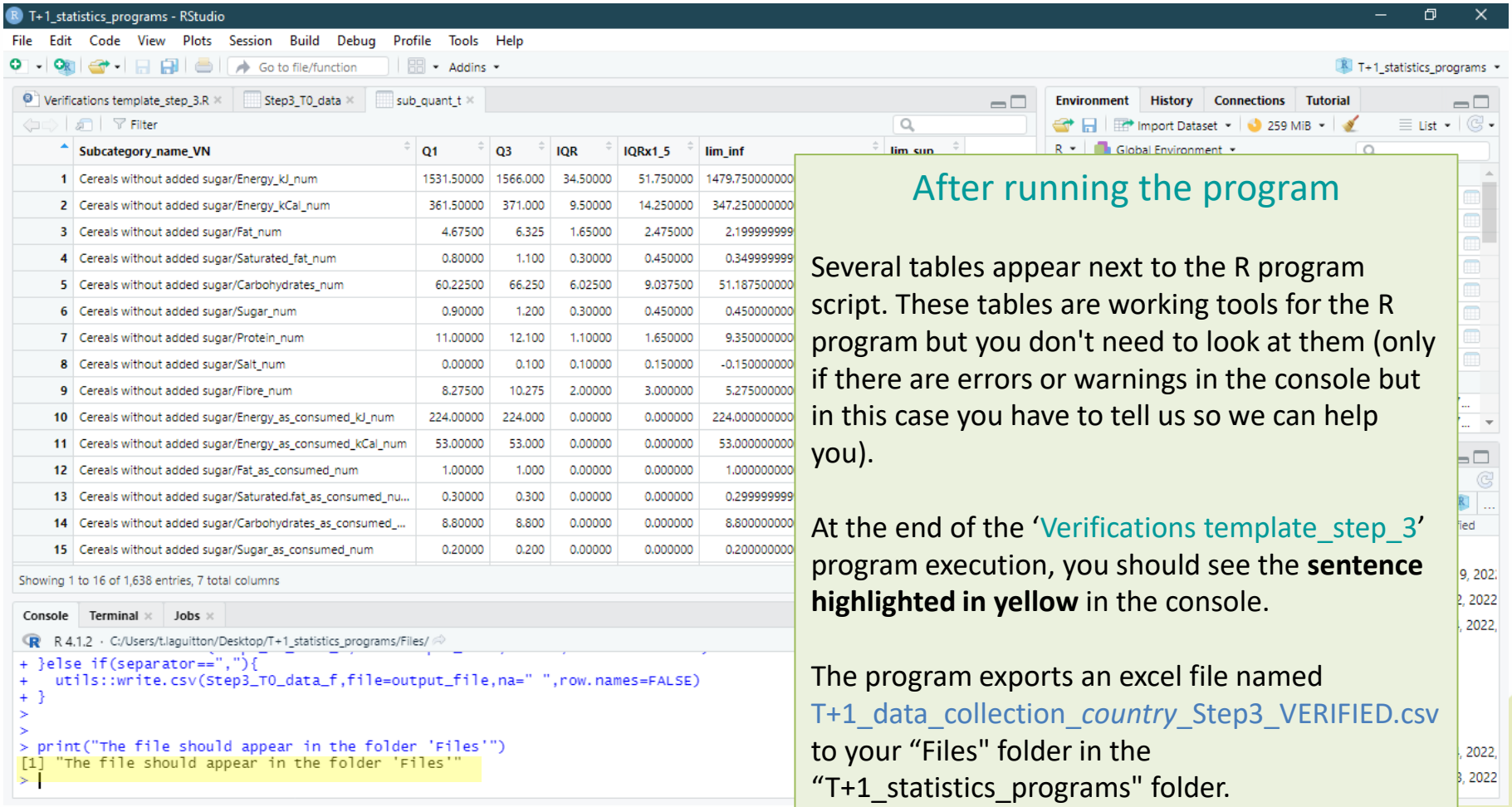
Program running

The 'Verifications template_step_3' program can take **quite a long time** (a few seconds to a few minutes) to run

When the program is running and has not finished its execution, a **stop sign** appears at the top right of the console. This means that the program is running and you must not close the '*T+1_statistics_programs – Rstudio*' window or the program will stop running.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_3' program



After running the program

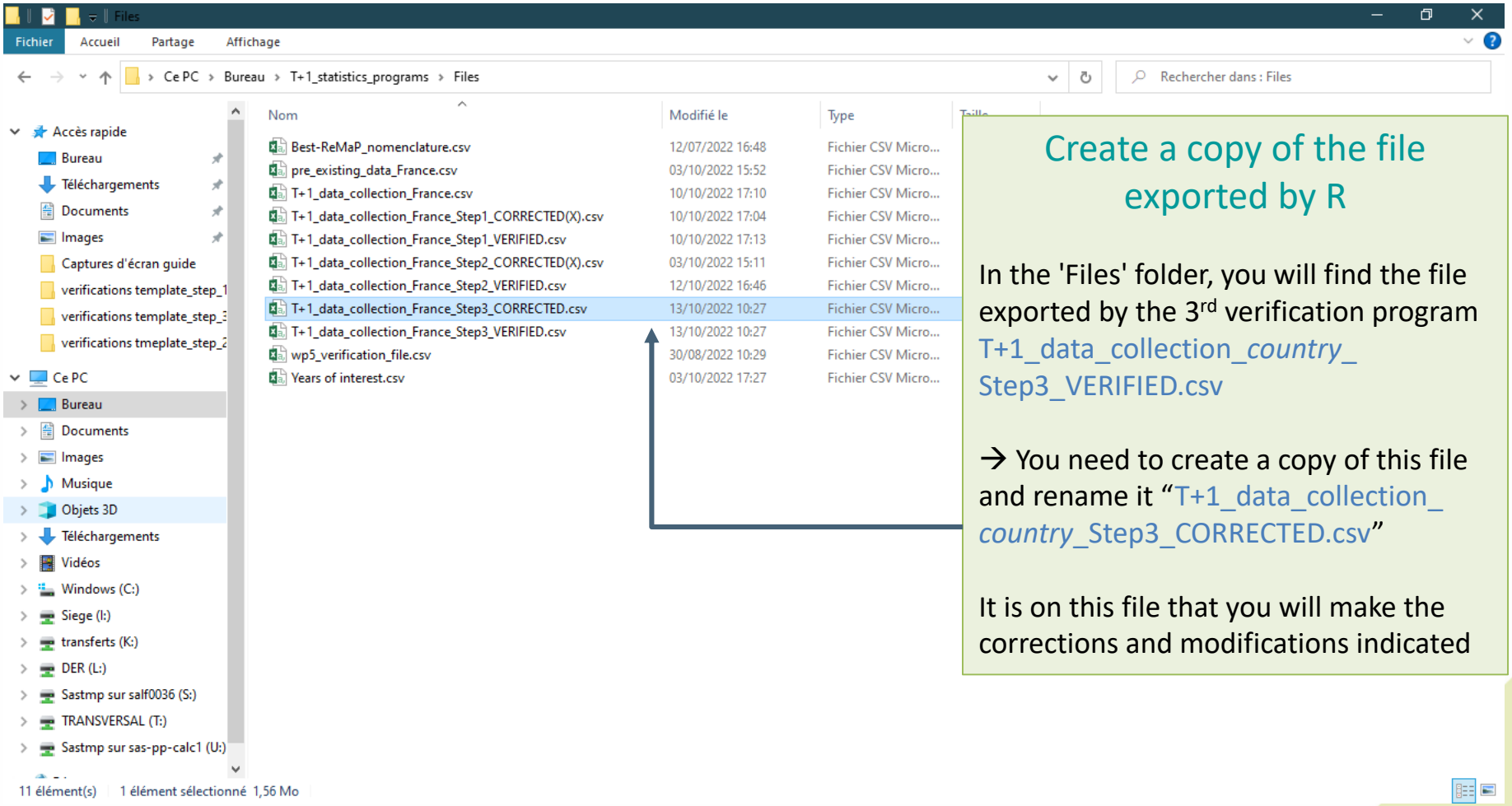
Several tables appear next to the R program script. These tables are working tools for the R program but you don't need to look at them (only if there are errors or warnings in the console but in this case you have to tell us so we can help you).

At the end of the 'Verifications template_step_3' program execution, you should see the **sentence highlighted in yellow** in the console.

The program exports an excel file named T+1_data_collection_*country*_Step3_VERIFIED.csv to your "Files" folder in the "T+1_statistics_programs" folder.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_3' program



Create a copy of the file exported by R

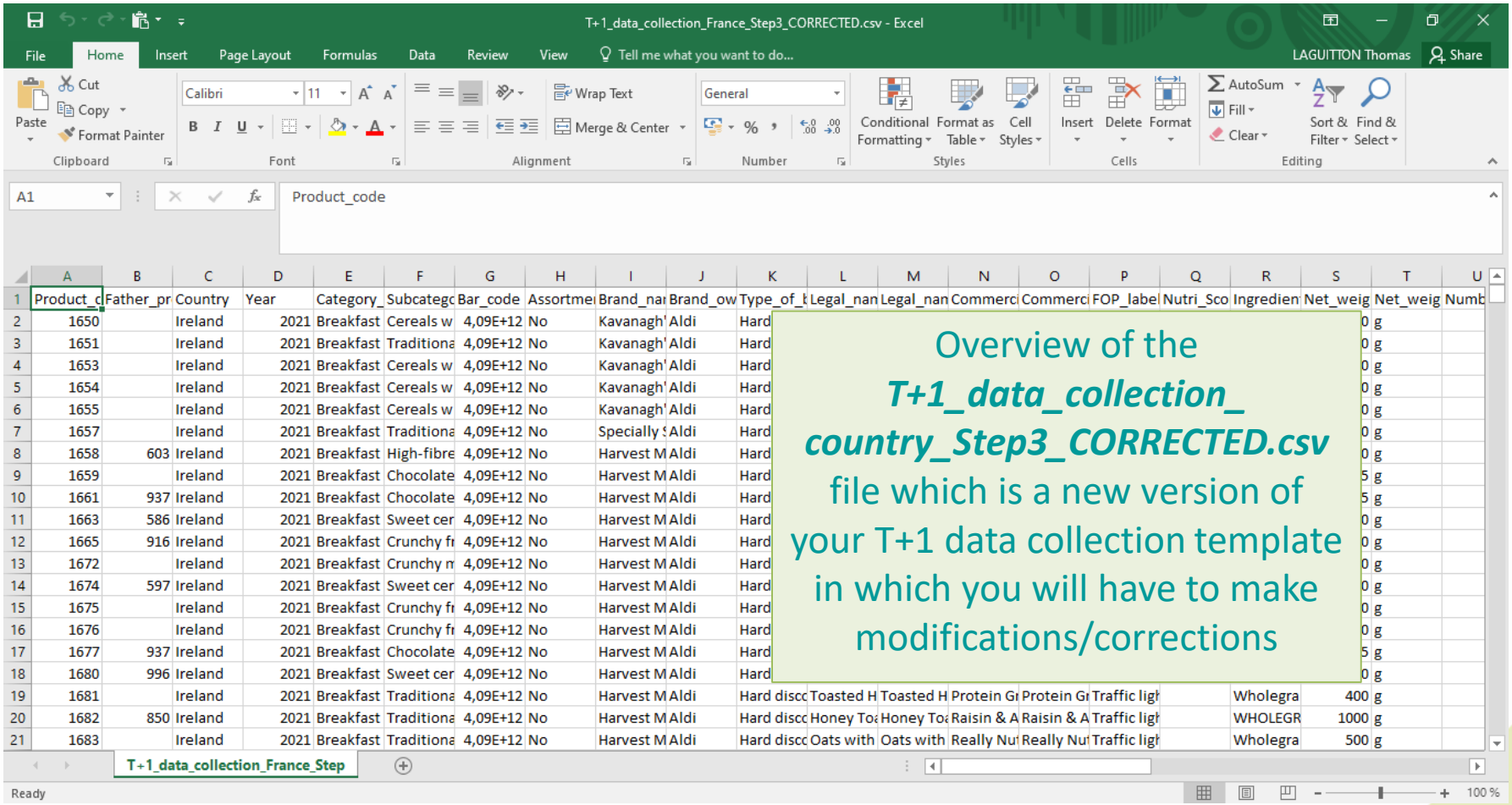
In the 'Files' folder, you will find the file exported by the 3rd verification program T+1_data_collection_*country*_Step3_VERIFIED.csv

→ You need to create a copy of this file and rename it "T+1_data_collection_*country*_Step3_CORRECTED.csv"

It is on this file that you will make the corrections and modifications indicated

Co-funded by the European Union's Health Programme (2014-2020)

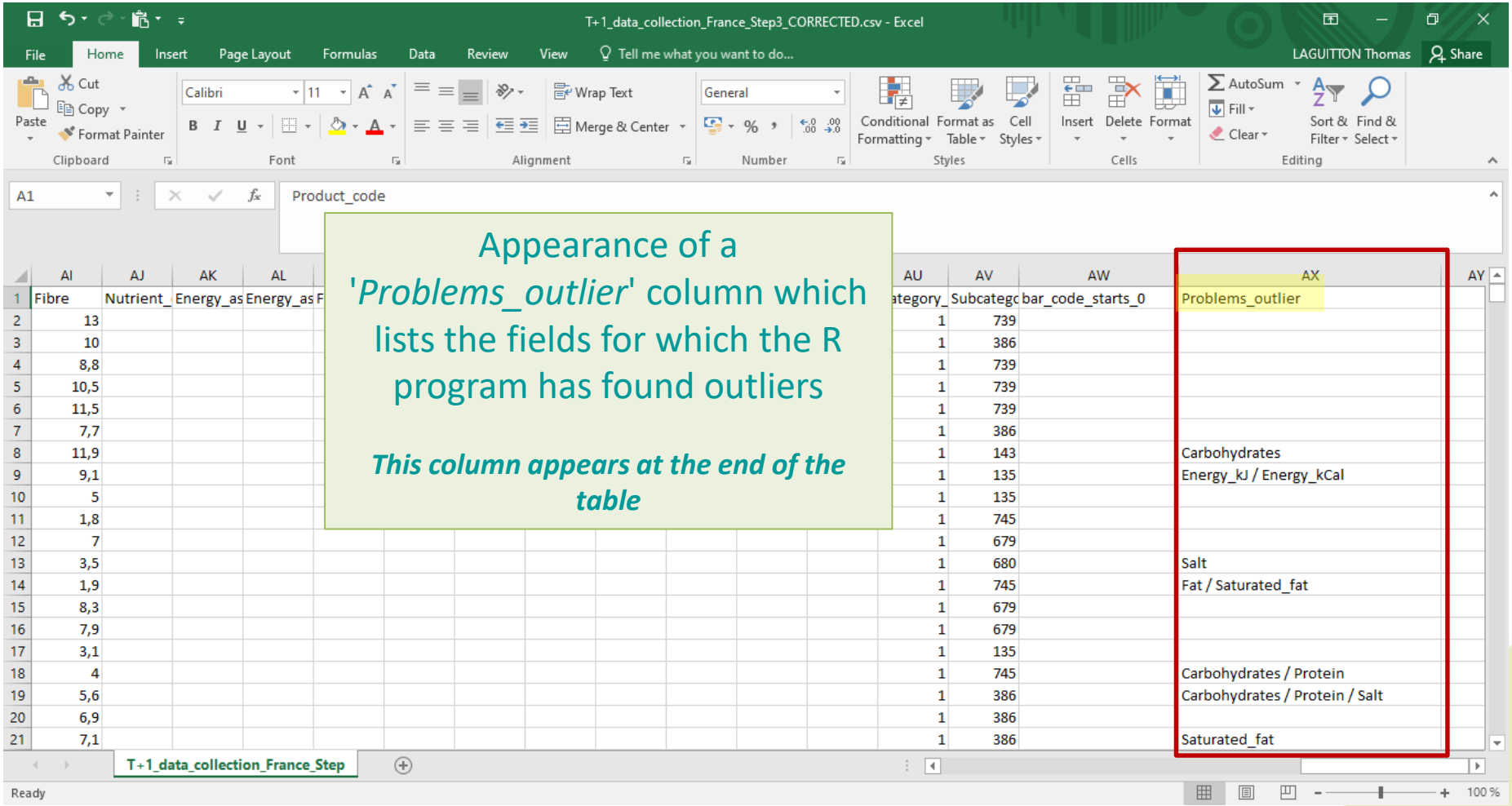
Excel file to modify after 'Verifications template_step_3' program



Overview of the ***T+1_data_collection_country_Step3_CORRECTED.csv*** file which is a new version of your T+1 data collection template in which you will have to make modifications/corrections

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_3' program

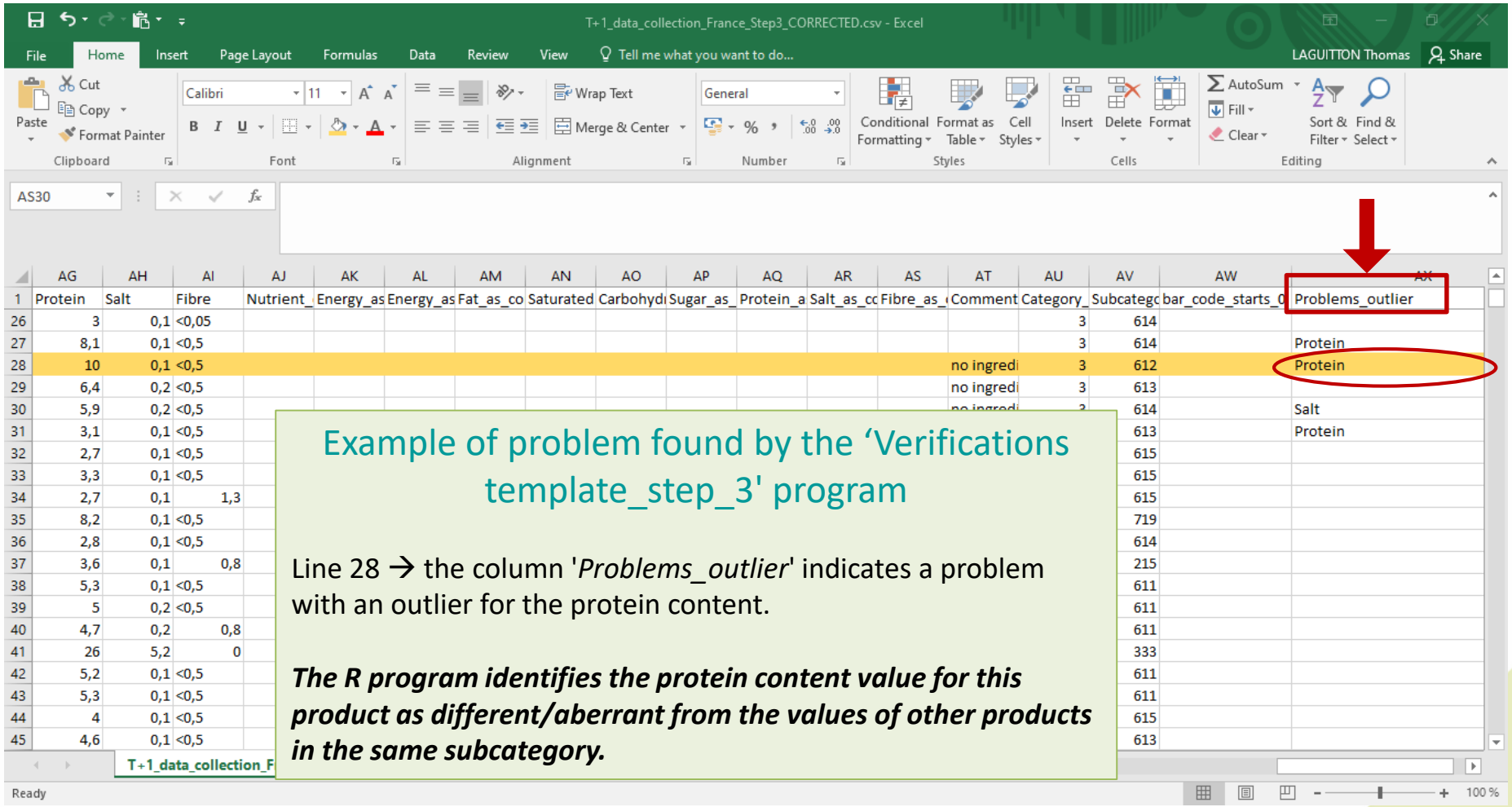


Appearance of a '*Problems_outlier*' column which lists the fields for which the R program has found outliers

This column appears at the end of the table

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_3' program



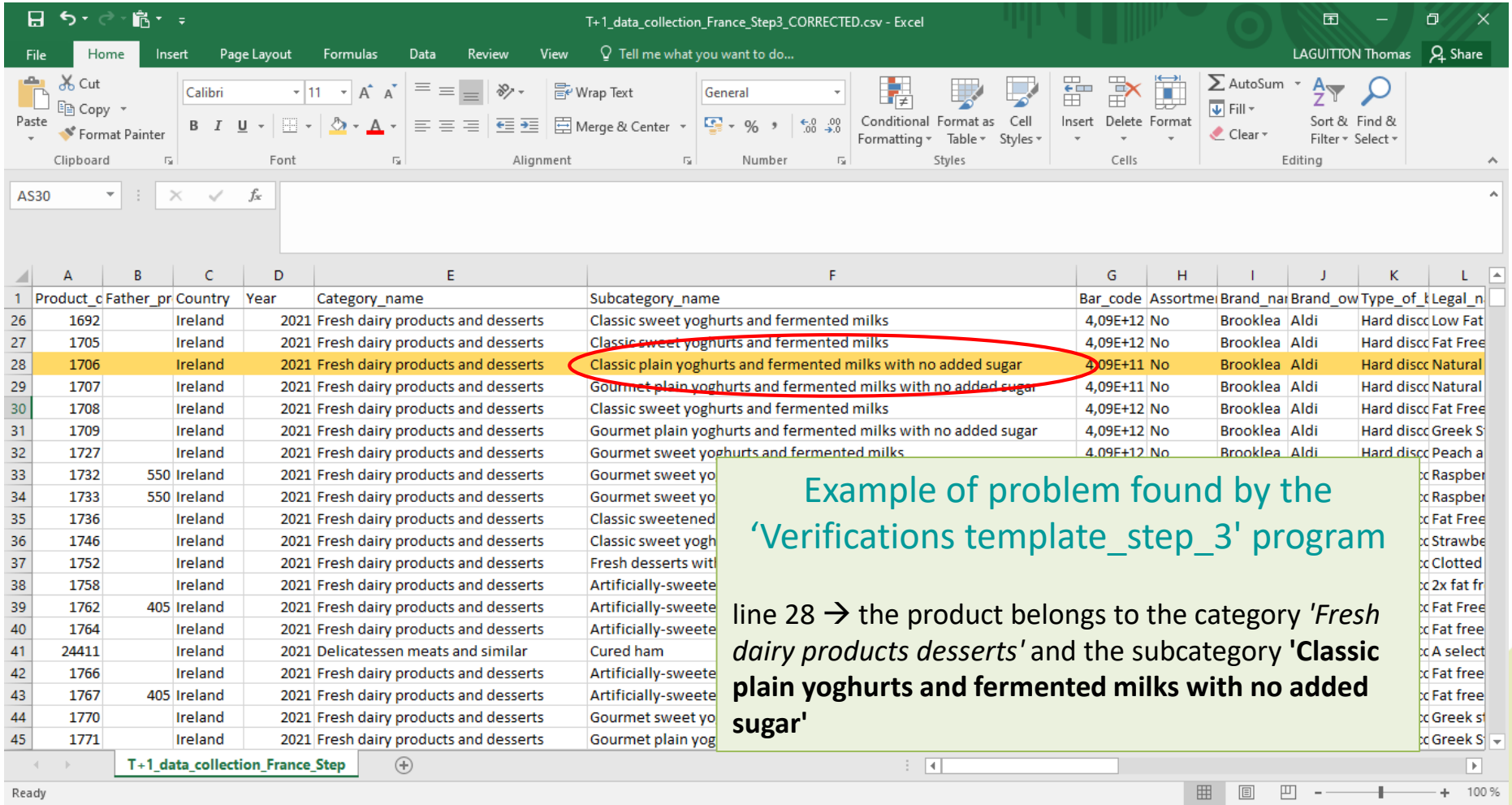
Example of problem found by the 'Verifications template_step_3' program

Line 28 → the column '*Problems_outlier*' indicates a problem with an outlier for the protein content.

The R program identifies the protein content value for this product as different/aberrant from the values of other products in the same subcategory.

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_3' program



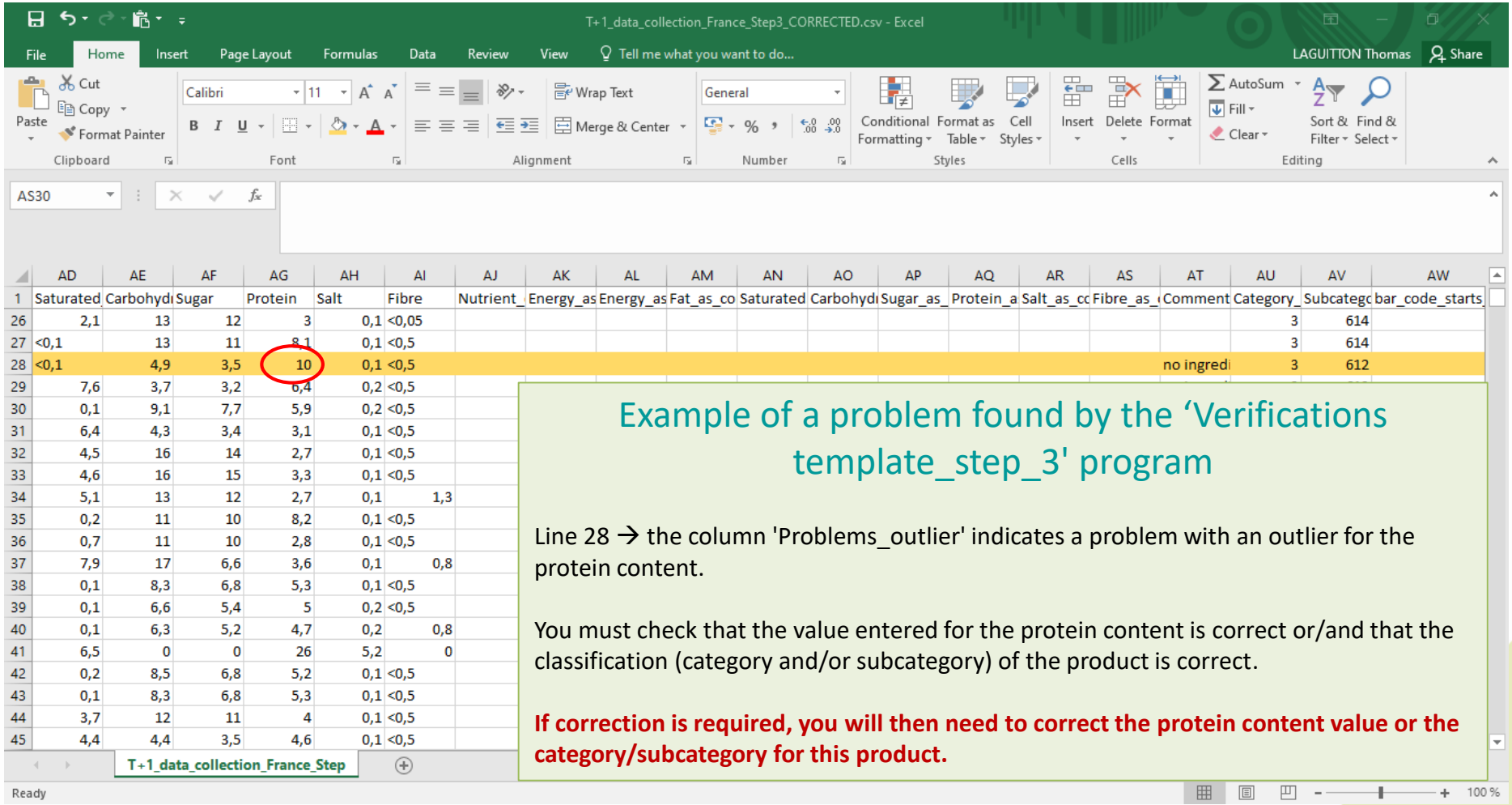
	A	B	C	D	E	F	G	H	I	J	K	L
1	Product_c	Father_pr	Country	Year	Category_name	Subcategory_name	Bar_code	Assortmei	Brand_nai	Brand_ow	Type_of_k	Legal_n
26	1692		Ireland	2021	Fresh dairy products and desserts	Classic sweet yoghurts and fermented milks	4,09E+12	No	Brooklea	Aldi	Hard disc	Low Fat
27	1705		Ireland	2021	Fresh dairy products and desserts	Classic sweet yoghurts and fermented milks	4,09E+12	No	Brooklea	Aldi	Hard disc	Fat Free
28	1706		Ireland	2021	Fresh dairy products and desserts	Classic plain yoghurts and fermented milks with no added sugar	4,09E+11	No	Brooklea	Aldi	Hard disc	Natural
29	1707		Ireland	2021	Fresh dairy products and desserts	Gourmet plain yoghurts and fermented milks with no added sugar	4,09E+11	No	Brooklea	Aldi	Hard disc	Natural
30	1708		Ireland	2021	Fresh dairy products and desserts	Classic sweet yoghurts and fermented milks	4,09E+12	No	Brooklea	Aldi	Hard disc	Fat Free
31	1709		Ireland	2021	Fresh dairy products and desserts	Gourmet plain yoghurts and fermented milks with no added sugar	4,09E+12	No	Brooklea	Aldi	Hard disc	Greek S
32	1727		Ireland	2021	Fresh dairy products and desserts	Gourmet sweet yoghurts and fermented milks	4,09E+12	No	Brooklea	Aldi	Hard disc	Peach a
33	1732	550	Ireland	2021	Fresh dairy products and desserts	Gourmet sweet yo						Raspber
34	1733	550	Ireland	2021	Fresh dairy products and desserts	Gourmet sweet yo						Raspber
35	1736		Ireland	2021	Fresh dairy products and desserts	Classic sweetened						Fat Free
36	1746		Ireland	2021	Fresh dairy products and desserts	Classic sweet yogh						Strawbe
37	1752		Ireland	2021	Fresh dairy products and desserts	Fresh desserts with						Clotted
38	1758		Ireland	2021	Fresh dairy products and desserts	Artificially-sweete						2x fat fr
39	1762	405	Ireland	2021	Fresh dairy products and desserts	Artificially-sweete						Fat Free
40	1764		Ireland	2021	Fresh dairy products and desserts	Artificially-sweete						Fat free
41	24411		Ireland	2021	Delicatessen meats and similar	Cured ham						A select
42	1766		Ireland	2021	Fresh dairy products and desserts	Artificially-sweete						Fat free
43	1767	405	Ireland	2021	Fresh dairy products and desserts	Artificially-sweete						Fat free
44	1770		Ireland	2021	Fresh dairy products and desserts	Gourmet sweet yo						Greek st
45	1771		Ireland	2021	Fresh dairy products and desserts	Gourmet plain yog						Greek S

Example of problem found by the 'Verifications template_step_3' program

line 28 → the product belongs to the category *'Fresh dairy products desserts'* and the subcategory **'Classic plain yoghurts and fermented milks with no added sugar'**

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_3' program



	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
1	Saturated	Carbohydı	Sugar	Protein	Salt	Fibre	Nutrient_	Energy_as	Energy_as	Fat_as_co	Saturated	Carbohydı	Sugar_as_	Protein_a	Salt_as_cc	Fibre_as_	Comment	Category_	Subcategc	bar_code_starts_
26	2,1	13	12	3	0,1	<0,05												3	614	
27	<0,1	13	11	8,1	0,1	<0,5												3	614	
28	<0,1	4,9	3,5	10	0,1	<0,5											no ingredi	3	612	
29	7,6	3,7	3,2	6,4	0,2	<0,5														
30	0,1	9,1	7,7	5,9	0,2	<0,5														
31	6,4	4,3	3,4	3,1	0,1	<0,5														
32	4,5	16	14	2,7	0,1	<0,5														
33	4,6	16	15	3,3	0,1	<0,5														
34	5,1	13	12	2,7	0,1	1,3														
35	0,2	11	10	8,2	0,1	<0,5														
36	0,7	11	10	2,8	0,1	<0,5														
37	7,9	17	6,6	3,6	0,1	0,8														
38	0,1	8,3	6,8	5,3	0,1	<0,5														
39	0,1	6,6	5,4	5	0,2	<0,5														
40	0,1	6,3	5,2	4,7	0,2	0,8														
41	6,5	0	0	26	5,2	0														
42	0,2	8,5	6,8	5,2	0,1	<0,5														
43	0,1	8,3	6,8	5,3	0,1	<0,5														
44	3,7	12	11	4	0,1	<0,5														
45	4,4	4,4	3,5	4,6	0,1	<0,5														

Example of a problem found by the 'Verifications template_step_3' program

Line 28 → the column 'Problems_outlier' indicates a problem with an outlier for the protein content.

You must check that the value entered for the protein content is correct or/and that the classification (category and/or subcategory) of the product is correct.

If correction is required, you will then need to correct the protein content value or the category/subcategory for this product.

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_3' program

Terms that may appear in the 'Problems_outlier' field following the 'Verifications template_step_3' program, their meaning and what to do

Problem	Meaning	Solution
• **Energy_kCal** • **Energy_kJ** • **Fat** • **Saturated_fat** • **Carbohydrates** • **Sugar** • **Protein** • **Salt** • **Fibre** • **Energy_as_consumed_kCal** • **Energy_as_consumed_kJ** • **Fat_as_consumed** • **Saturated_fat_as_consumed** • **Carbohydrates_as_consumed** • **Sugar_as_consumed** • **Protein_as_consumed** • **Salt_as_consumed** • **Fibre_as_consumed**	• The nutritional value of the product for this nutrient appears to be an outlier compared to the nutritional value for this nutrient of other products in the same subcategory.	→ Check the pictures of the product, to be sure that the value entered in the template is the correct one. If not, you must enter the correct value directly in the Excel file in .csv format. **And/or** → Check that the category and subcategory associated to the product are the correct ones (you can help you with the WP5 classification guidelines that have been created for each food category). If not, you must enter the correct information (category name + code and/or subcategory name + code) directly in the Excel file in .csv format. **Else** →Nutritional value and subcategory entered for this product are the correct ones, no correction is needed. You must indicate in the 'Comments' field: "*outliers checked*". **Be careful! A product can have wrong values + wrong classification, it is important to check both for the product.**

Co-funded by the European Union's Health Programme (2014-2020)

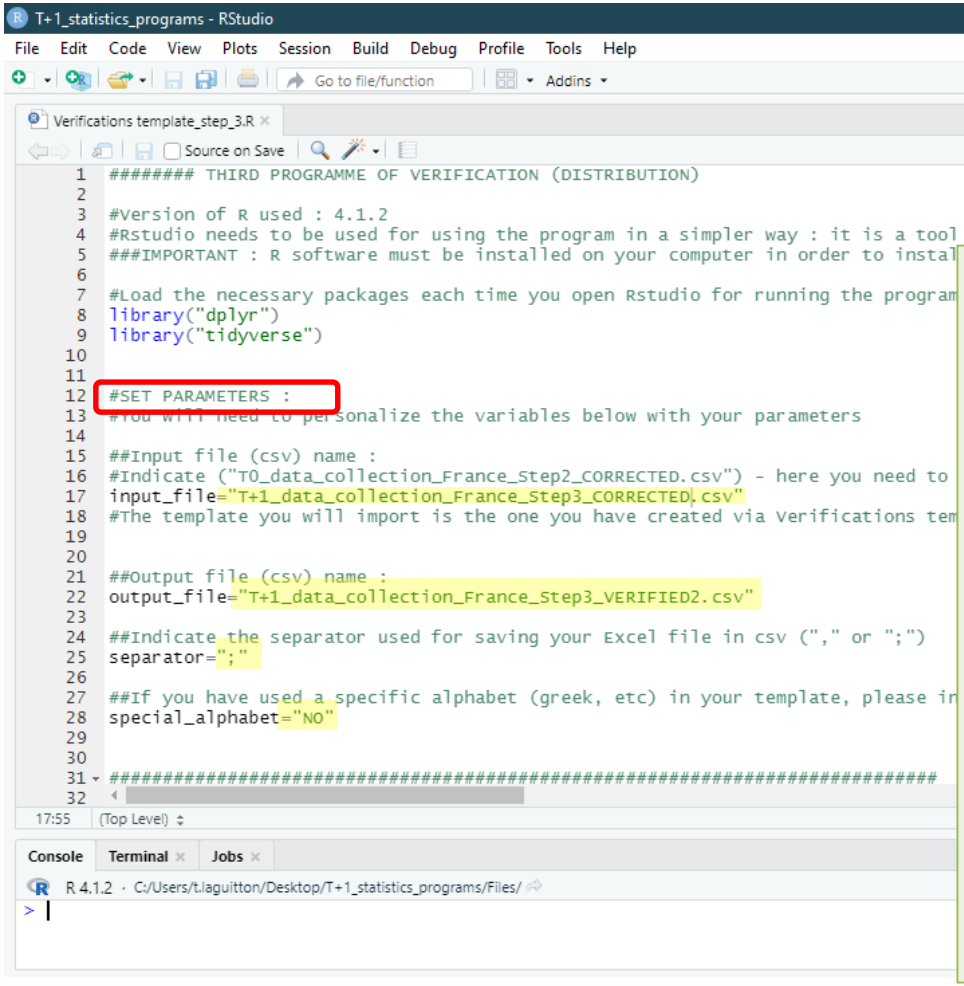
Excel file to modify after ‘Verifications template_step_3’ program

- **Be careful! When saving the file, you must be sure that the barcodes appear in full and not in scientific format. If this is the case, you must repeat the procedure on pages 24→28.**
- Once the corrections have been made and the barcodes are in the correct format, you can save the file *T+1_data_collection_country_Step3_CORRECTED.csv* and close it.
- You will have to run this corrected file again in the ‘Verifications template_step_3' program to make sure you haven't missed a check.

Your Rstudio interface must be cleaned up before running the ‘Verifications template_step_3’ program again.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_3' program



Setting parameters of the second running of 'Verifications template_step_3' program

For this second running of the 'Verifications template_step 3', the only fields you need to change are the names of the input file and the output file

Input_file =
"*T+1_data_collection_country_**Step3**_CORRECTED.csv*"

Output_file =
"*T+1_data_collection_country_**Step3**_VERIFIED2.csv*"

You must also indicate the separator and if you have a specific alphabet as in the first run of the program.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_3' program

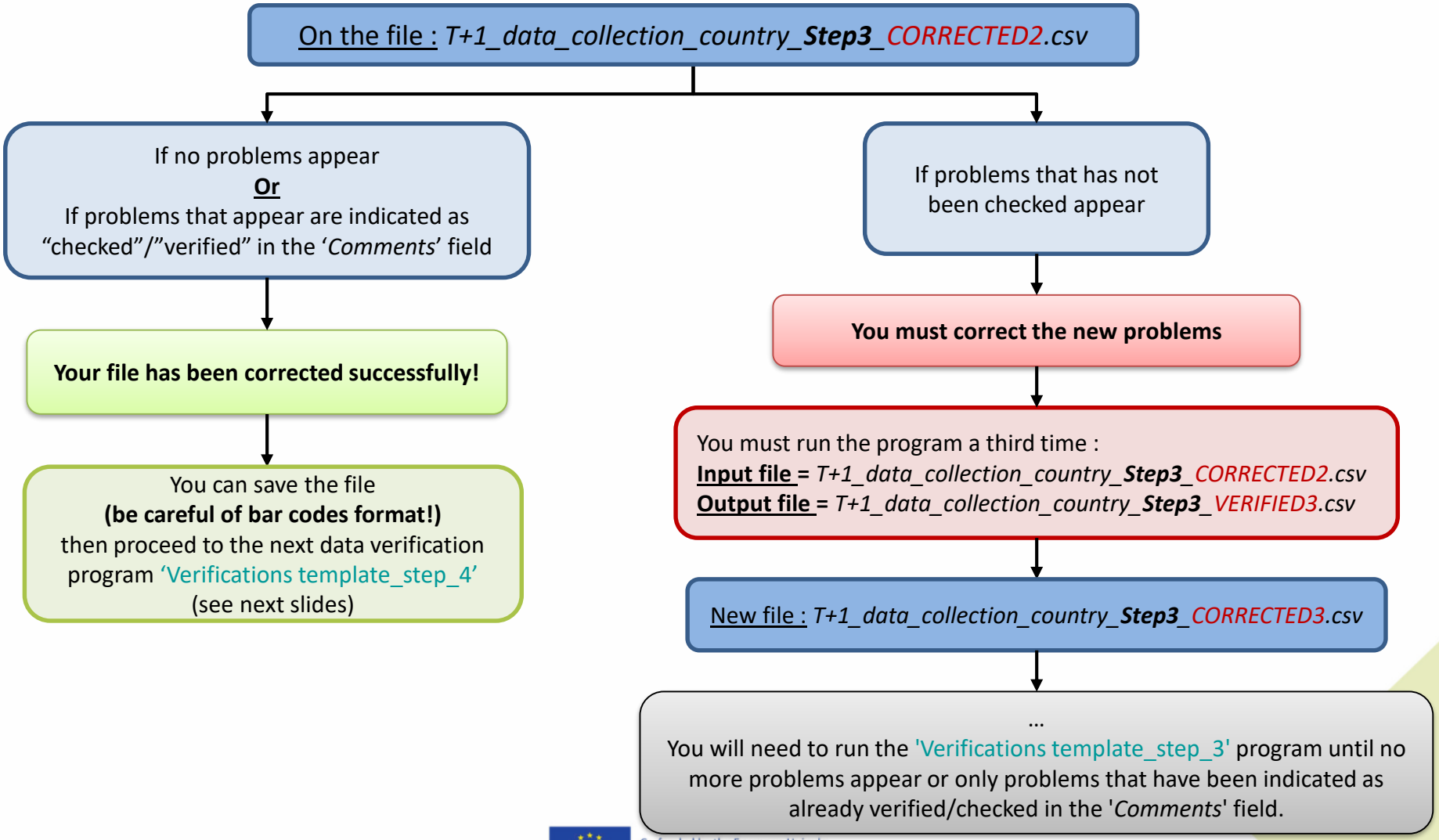
- At the end of this second run, you get in your "files" folder a file called: "*T+1_data_collection_country_**Step3**_VERIFIED2.csv*"

- You must create a copy of this file and call it : "*T+1_data_collection_country_**Step3**_CORRECTED2.csv*"

  → It is on this file that you will make the modifications following the checks

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_3' program



Co-funded by the European Union's Health Programme (2014-2020)

3) Running of the programs

A. Part 1 : R setup program

B. Part 2 : Verification programs and template cleaning/standardization

i. 1st verification program : 'Verifications template_step_1'

ii. 2nd verification program : 'Verifications template_step_2'

iii. 3rd verification program : 'Verifications template_step_3'

iv. 4th verification program : 'Verifications template_step_4'

Co-funded by the European Union's Health Programme (2014-2020)

4th verification program : ‘Verifications template_step_4’

Presentation of the ‘Verifications template step_4' program :

- The 4th verification program allows you to highlight problems in the pairing of your T+1 data with your pre-existing data.

- This progam will allow, for example, to :

  - Check that the information entered matches between 2 paired products (Brand, category name and code, subcategory name and code)

  - Check that the father_product_codes indicated in the T+1 data do exist in the pre-existing data file.

  - Check that unique products_code given in your T+1 data doesn't 'exist in your pre-existing data

  - …

Co-funded by the European Union's Health Programme (2014-2020)

4th verification program : 'Verifications template_step_4'

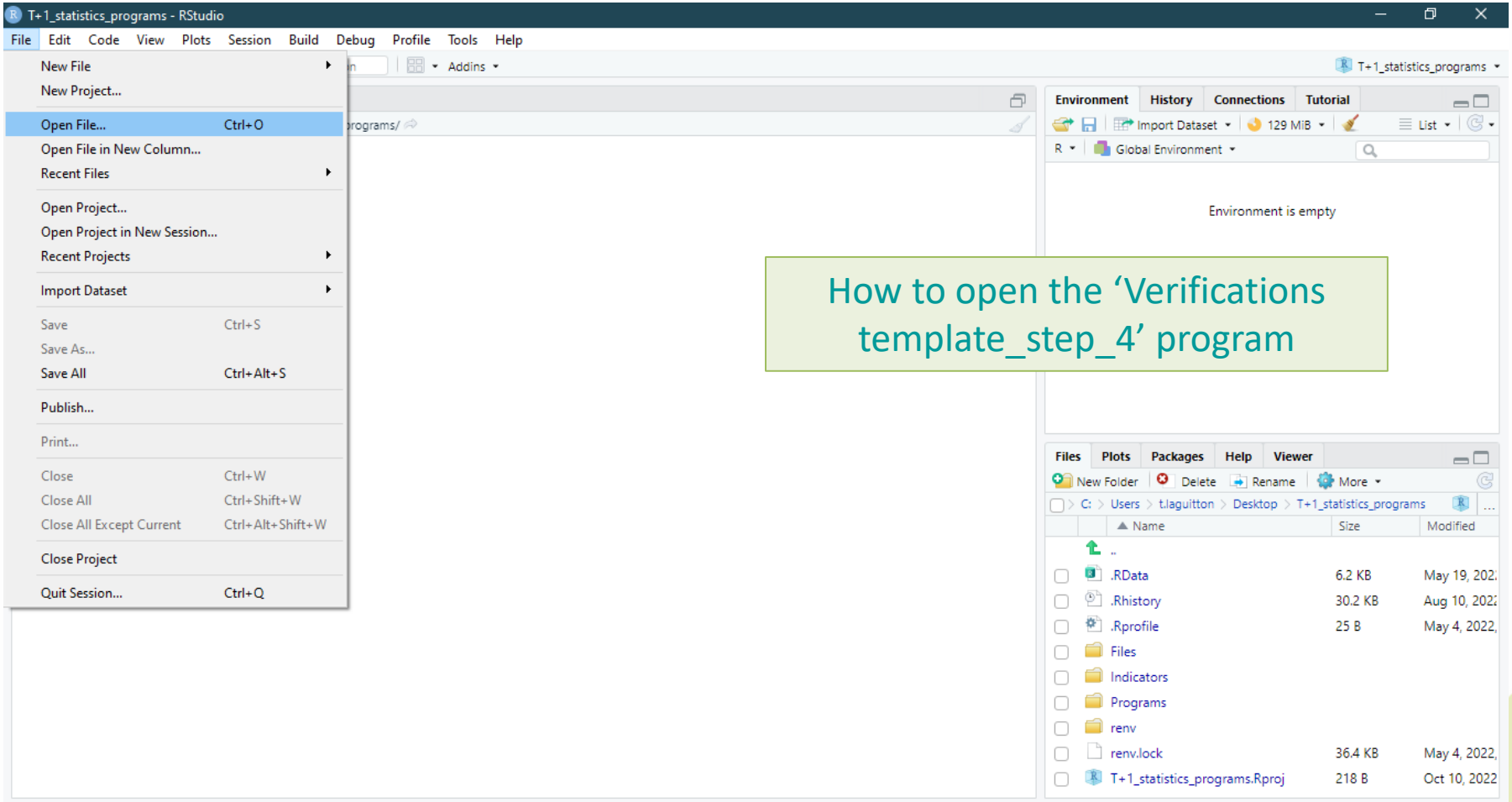
Requirements before starting the program 'Verifications template_step_4' :

- The programs 'Verifications template_step_1', 'Verifications template_step_2' and 'Verifications template_step_3' should have been run on your data
- You should no longer have any problems appearing or only problems that have been notified as verified after running the program 'Verifications template_step_3'
- You must have your template in your possession and it must now be called:
  T+1_data_collection_***country***_Step3_CORRECTED(X).csv (with the name of your own country)
  ((X) is the number of the last file exported and corrected after the last run of the first verification program)

  You need to make sure that the barcodes in this file appear in full and not in scientific format (see procedure pages 24→28)
- You must also have in your possession your pre-existing data file in **.csv format** called pre_existing_data_***country***.csv and your file Years_of_interest.csv that must have been filled in indicating the years chosen in your pre-existing data for pairing with the T+1 data.

Your Rstudio interface must have been cleaned up before running the program.
All cleaning steps are described on pages 62→68.

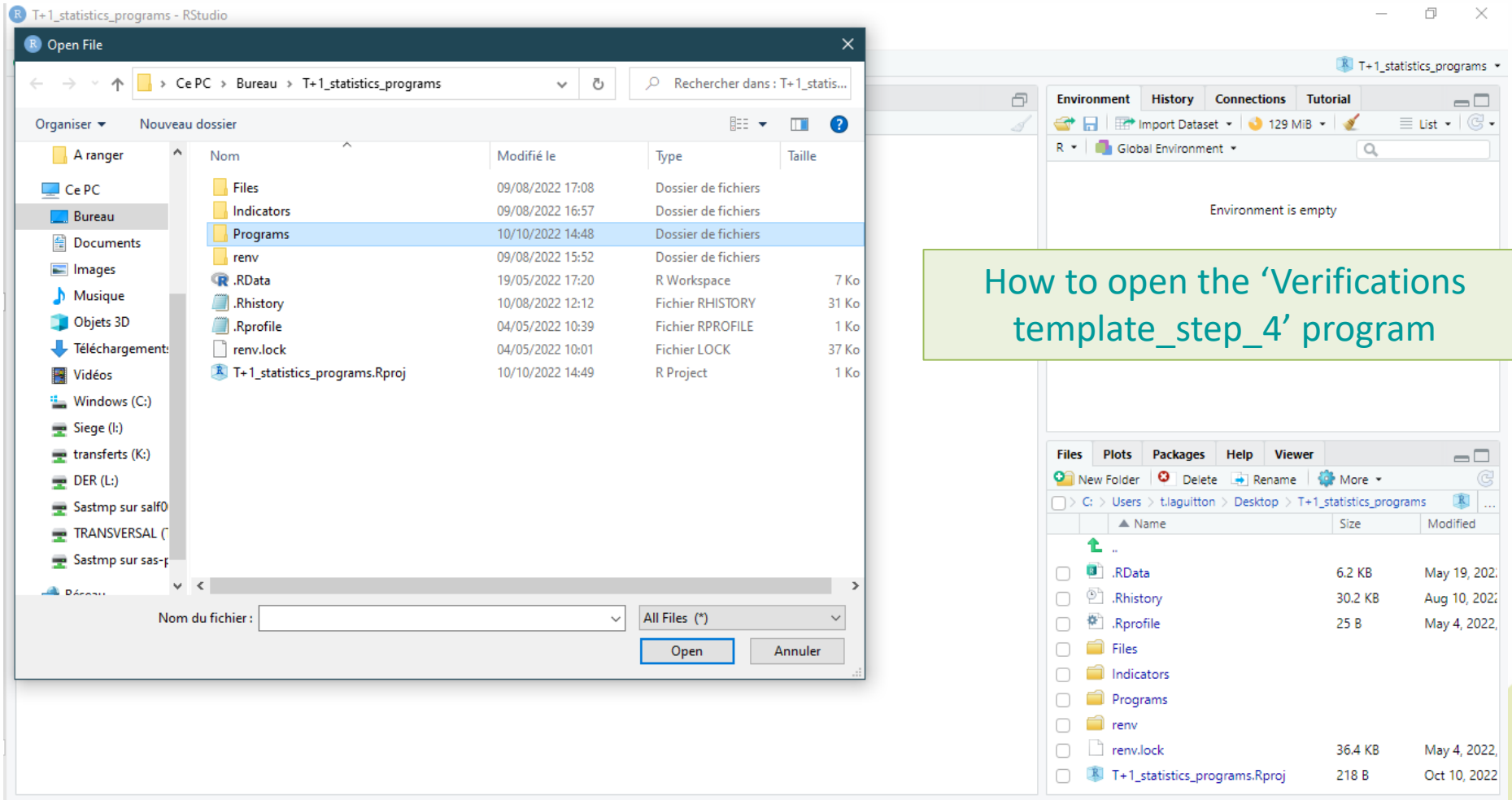
Running of ‘Verifications template_step_4’ program



How to open the ‘Verifications template_step_4’ program

Co-funded by the European Union's Health Programme (2014-2020)

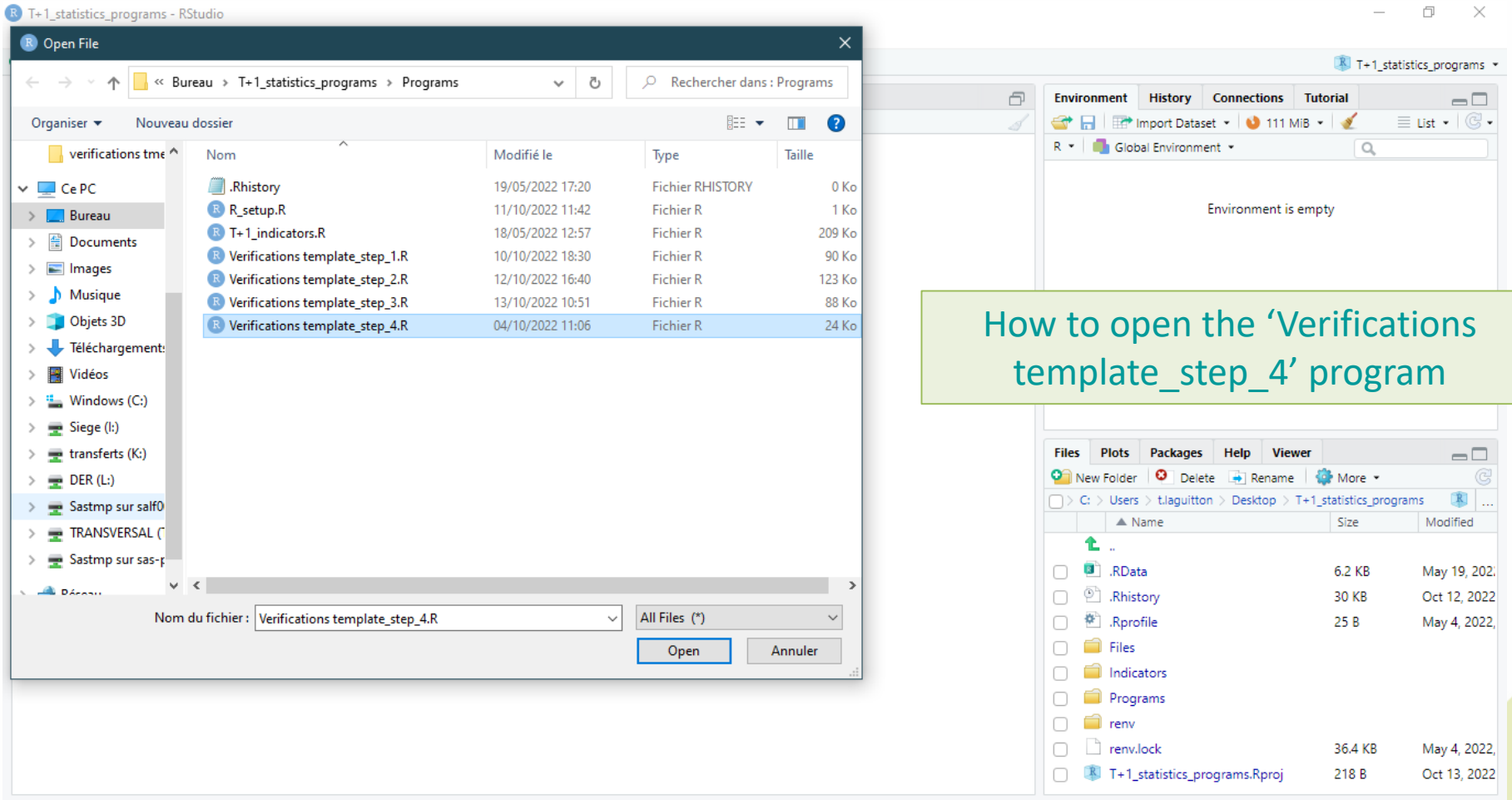
Running of ‘Verifications template_step_4’ program



How to open the ‘Verifications template_step_4’ program

Co-funded by the European Union's Health Programme (2014-2020)

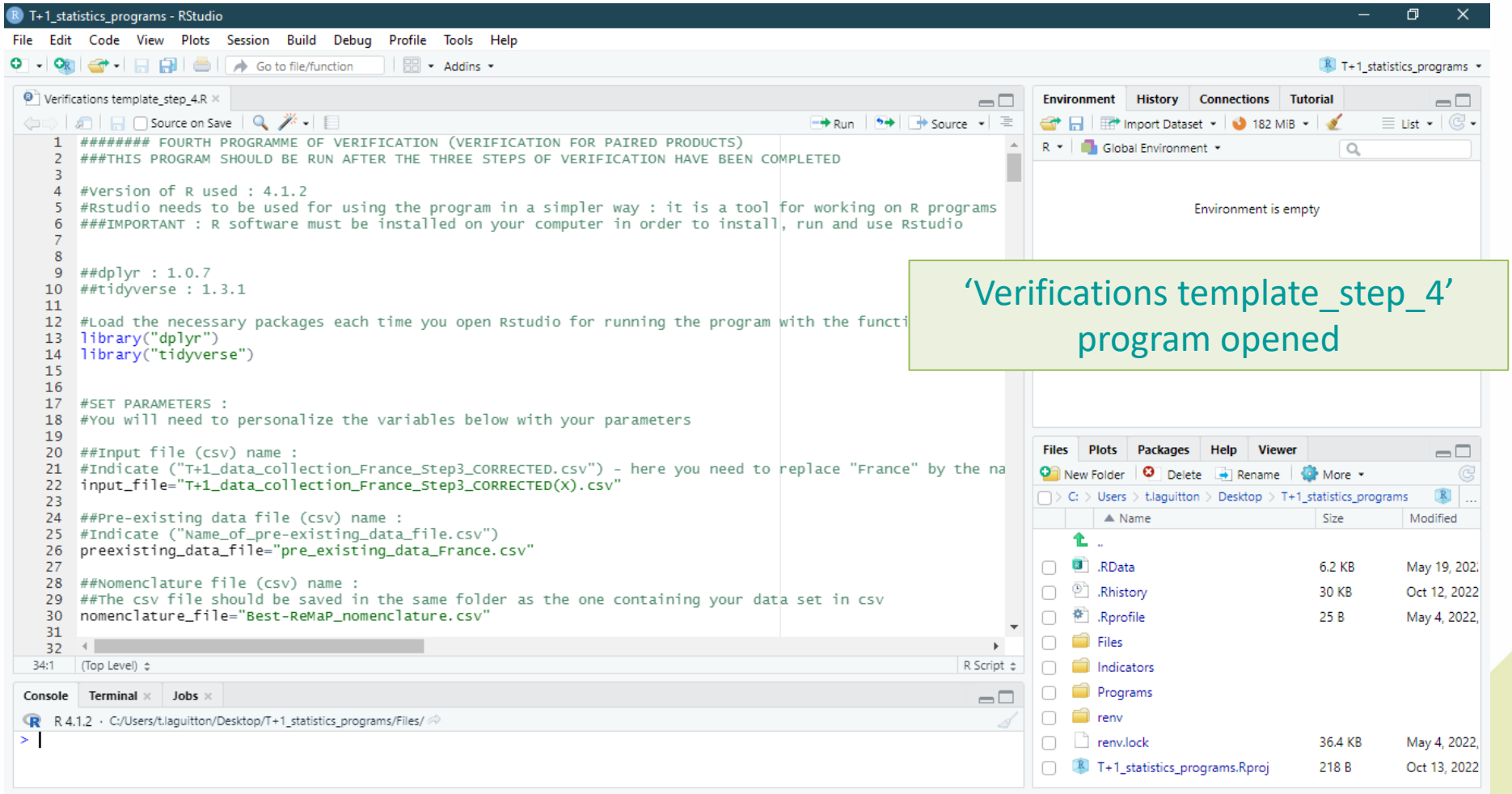
Running of ‘Verifications template_step_4’ program



How to open the ‘Verifications template_step_4’ program

Co-funded by the European Union's Health Programme (2014-2020)

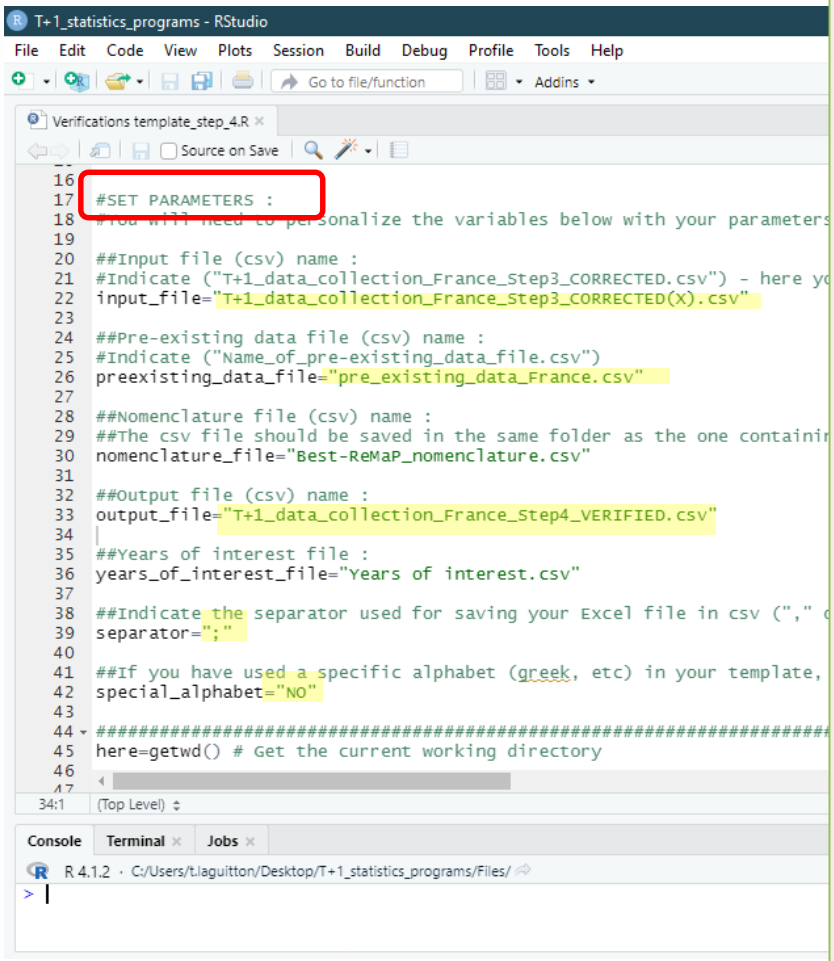
Running of ‘Verifications template_step_4’ program



‘Verifications template_step_4’ program opened

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_4' program



Setting parameters of the 4th verification program

In the first run of the 4th verification program, you need to change the name of the country with your own country name in the input file (line 17) and the output file (line 33)

In this program, you will also need to import your pre-existing data file. To do this, you need to enter the name of your own country in the file name line 26.

You must also indicate the separator and if you have a specific alphabet as in the program 'Verifications template_step_1', 'Verifications template_step_2' and 'Verifications template_step_3'.

Example :

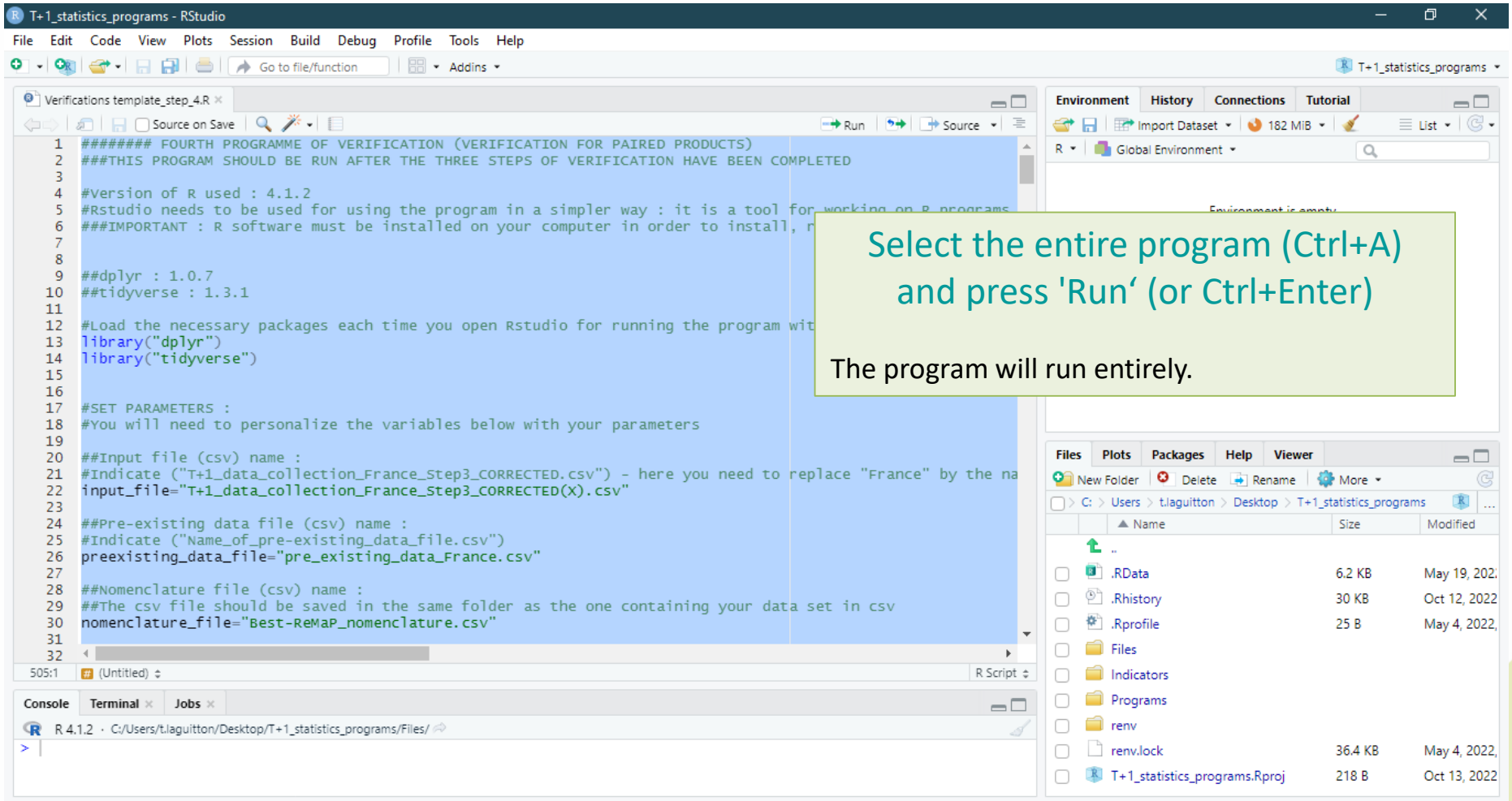
- **Input_file =** "T+1_data_collection_Ireland_Step2_CORRECTED(X).csv"

((X) is the number of the last file exported and corrected after the last run of the second verification program)

- **Preexisting_data_file =** "pre_existing_data_Ireland.csv"
- **Output_file =** "T+1_data_collection_Ireland_Step3_VERIFIED.csv"
- **Separator** = ";"
- **Special_alphabet** = "NO"

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_4' program

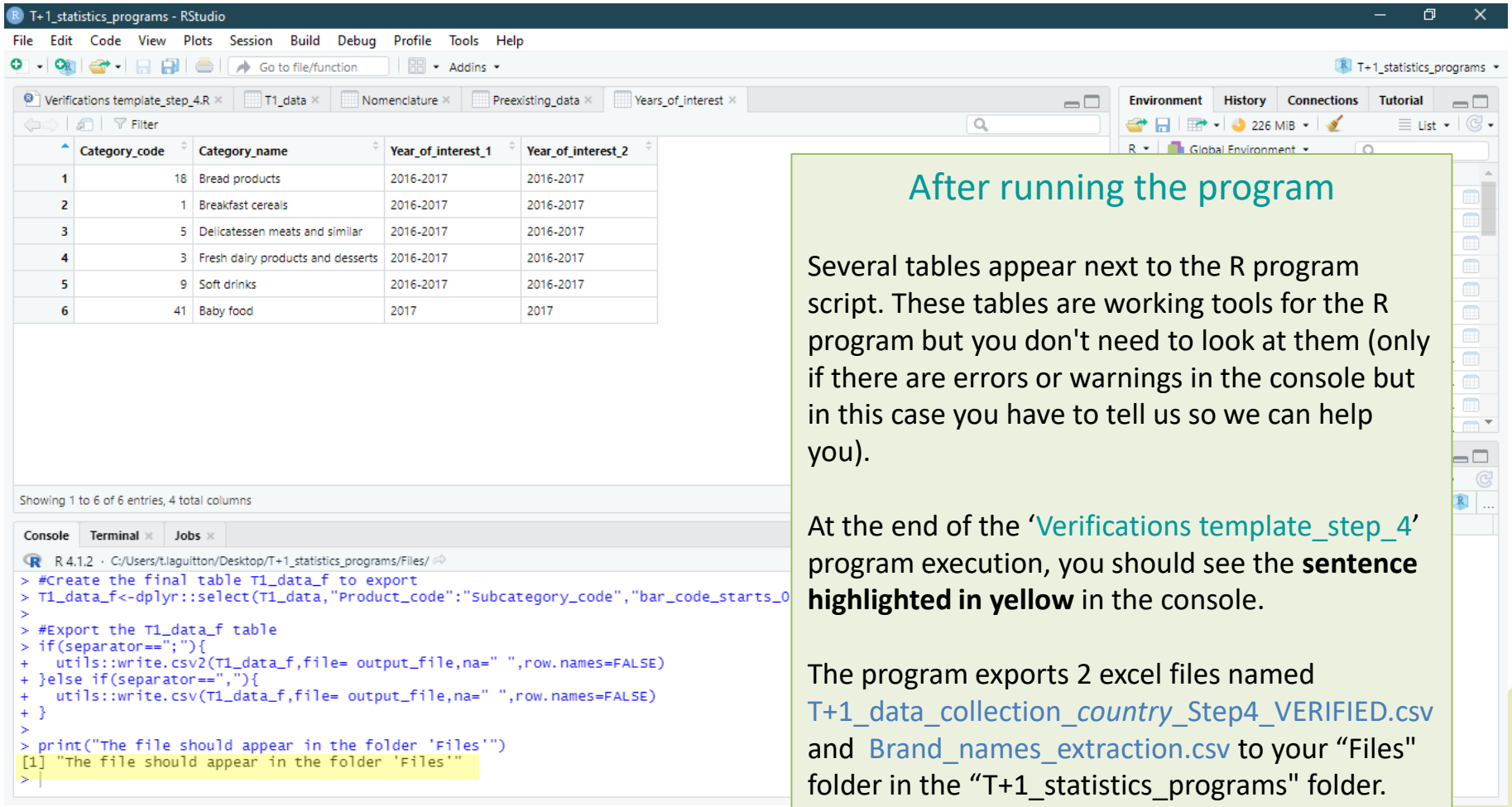


Select the entire program (Ctrl+A) and press 'Run' (or Ctrl+Enter)

The program will run entirely.

Co-funded by the European Union's Health Programme (2014-2020)

Running of ‘Verifications template_step_3’ program



After running the program

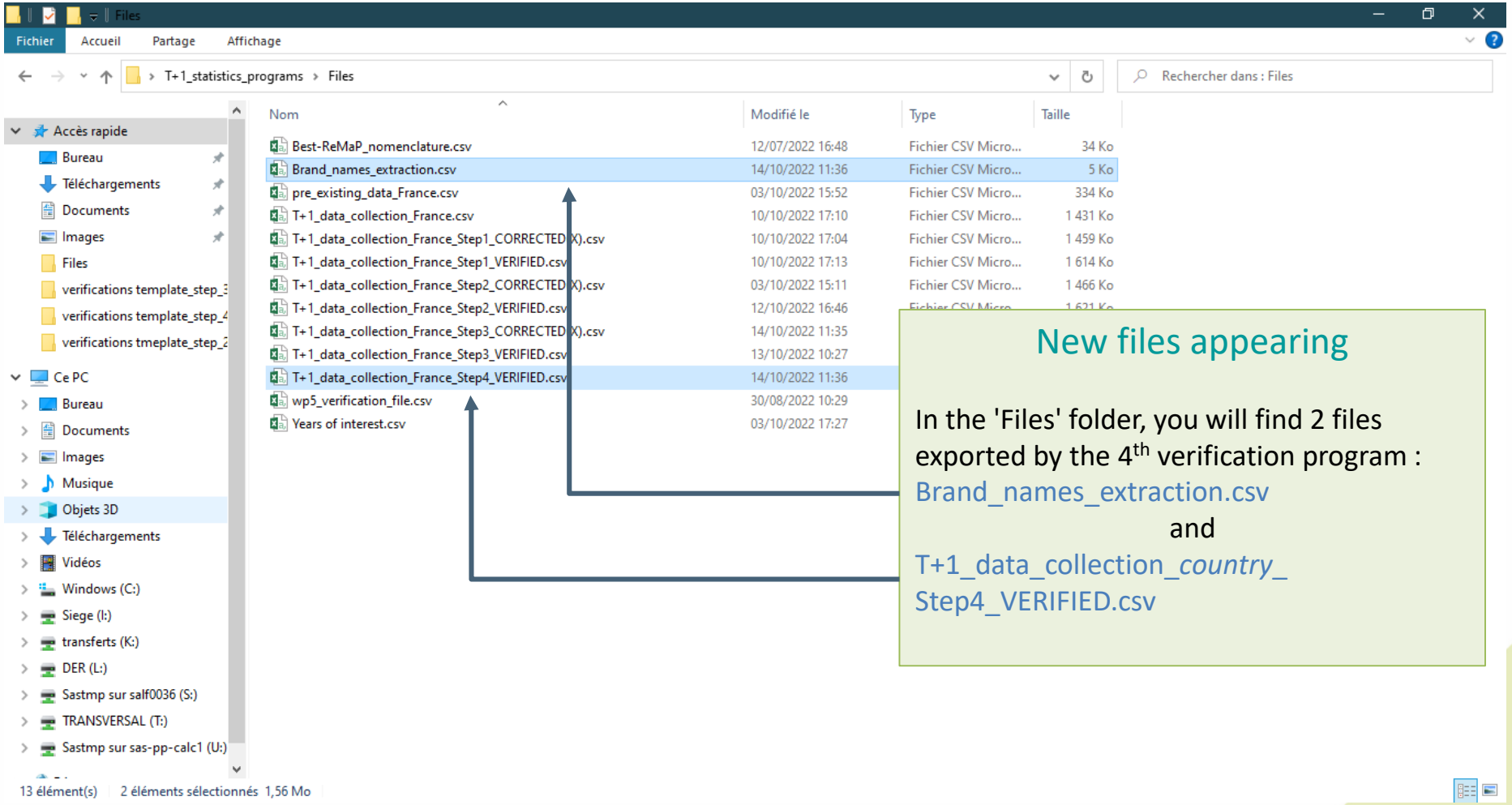
Several tables appear next to the R program script. These tables are working tools for the R program but you don't need to look at them (only if there are errors or warnings in the console but in this case you have to tell us so we can help you).

At the end of the ‘Verifications template_step_4’ program execution, you should see the **sentence highlighted in yellow** in the console.

The program exports 2 excel files named T+1_data_collection_*country*_Step4_VERIFIED.csv and Brand_names_extraction.csv to your “Files" folder in the “T+1_statistics_programs" folder.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_4' program



New files appearing

In the 'Files' folder, you will find 2 files exported by the 4th verification program :

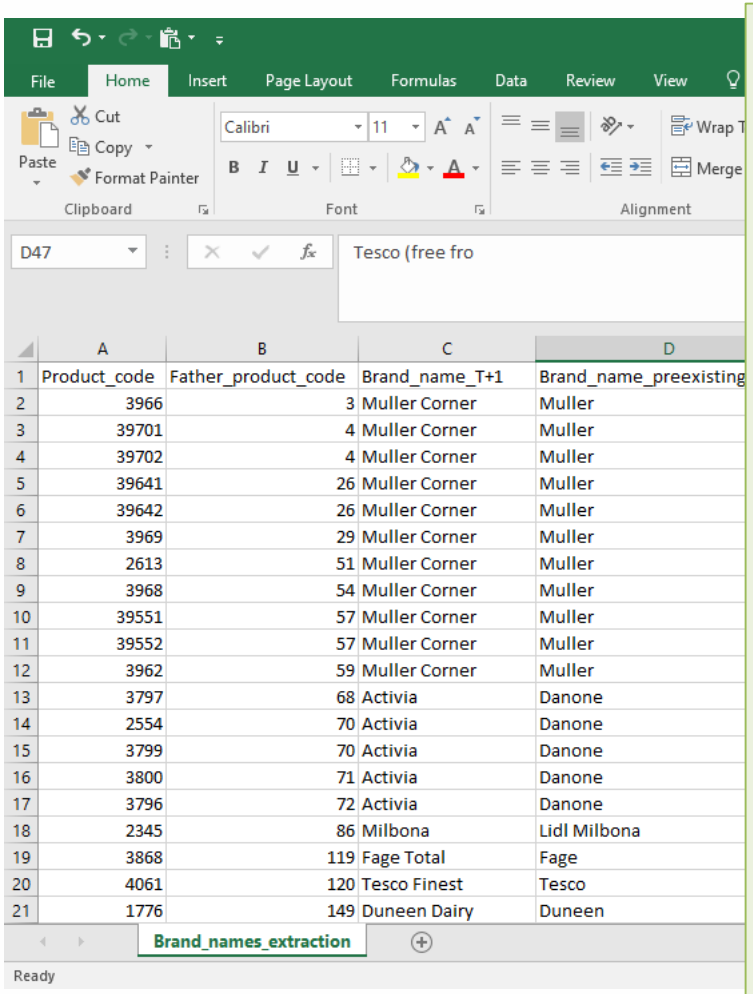
Brand_names_extraction.csv

and

T+1_data_collection_*country*_Step4_VERIFIED.csv

Co-funded by the European Union's Health Programme (2014-2020)

Running of ‘Verifications template_step_4’ program



	A	B	C	D
1	Product_code	Father_product_code	Brand_name_T+1	Brand_name_preexisting
2	3966	3	Muller Corner	Muller
3	39701	4	Muller Corner	Muller
4	39702	4	Muller Corner	Muller
5	39641	26	Muller Corner	Muller
6	39642	26	Muller Corner	Muller
7	3969	29	Muller Corner	Muller
8	2613	51	Muller Corner	Muller
9	3968	54	Muller Corner	Muller
10	39551	57	Muller Corner	Muller
11	39552	57	Muller Corner	Muller
12	3962	59	Muller Corner	Muller
13	3797	68	Activia	Danone
14	2554	70	Activia	Danone
15	3799	70	Activia	Danone
16	3800	71	Activia	Danone
17	3796	72	Activia	Danone
18	2345	86	Milbona	Lidl Milbona
19	3868	119	Fage Total	Fage
20	4061	120	Tesco Finest	Tesco
21	1776	149	Duneen Dairy	Duneen

“Brand_names_extraction.csv” file

This file shows the products paired between your pre-existing data and your T+1 data that have ‘**brand_names’ which are not identical.**

→ This file is only used to check if there are **no major pairing errors**.

Example :

brand_name_preexisting_data = *Kellogg‘s*

brand_name_T+1 = *Carrefour*

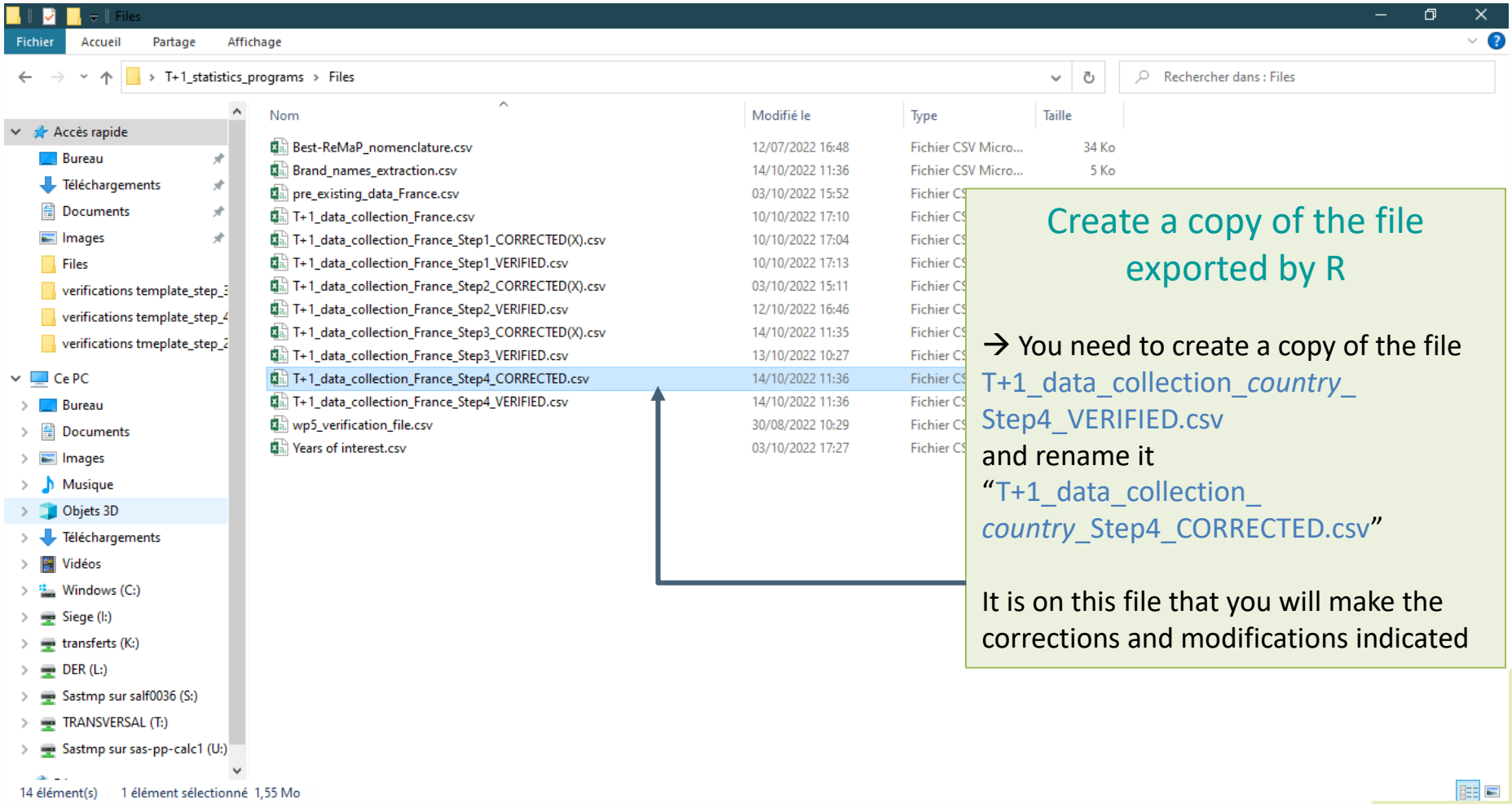
In this case, you need to check the matching of the two products: maybe the product has been matched to the wrong product in the pre-existing data and you need to find the right product in the pre-existing data, maybe there is a typing error in the father_product_code, etc.

→ If two brand names are different but it is a **slight difference** (“Lidl Milbona" vs. “Milbona", "Muller" vs. "Muller Corner", “Kellogg’s” vs. “Kelloggs”, etc.), then you do **not need to modify** the information.

We advise you not to change the 'brand_name' in your pre-existing data as you may no longer have the product pictures in your possession to be sure of what you are changing.

Co-funded by the European Union's Health Programme (2014-2020)

Running of 'Verifications template_step_4' program



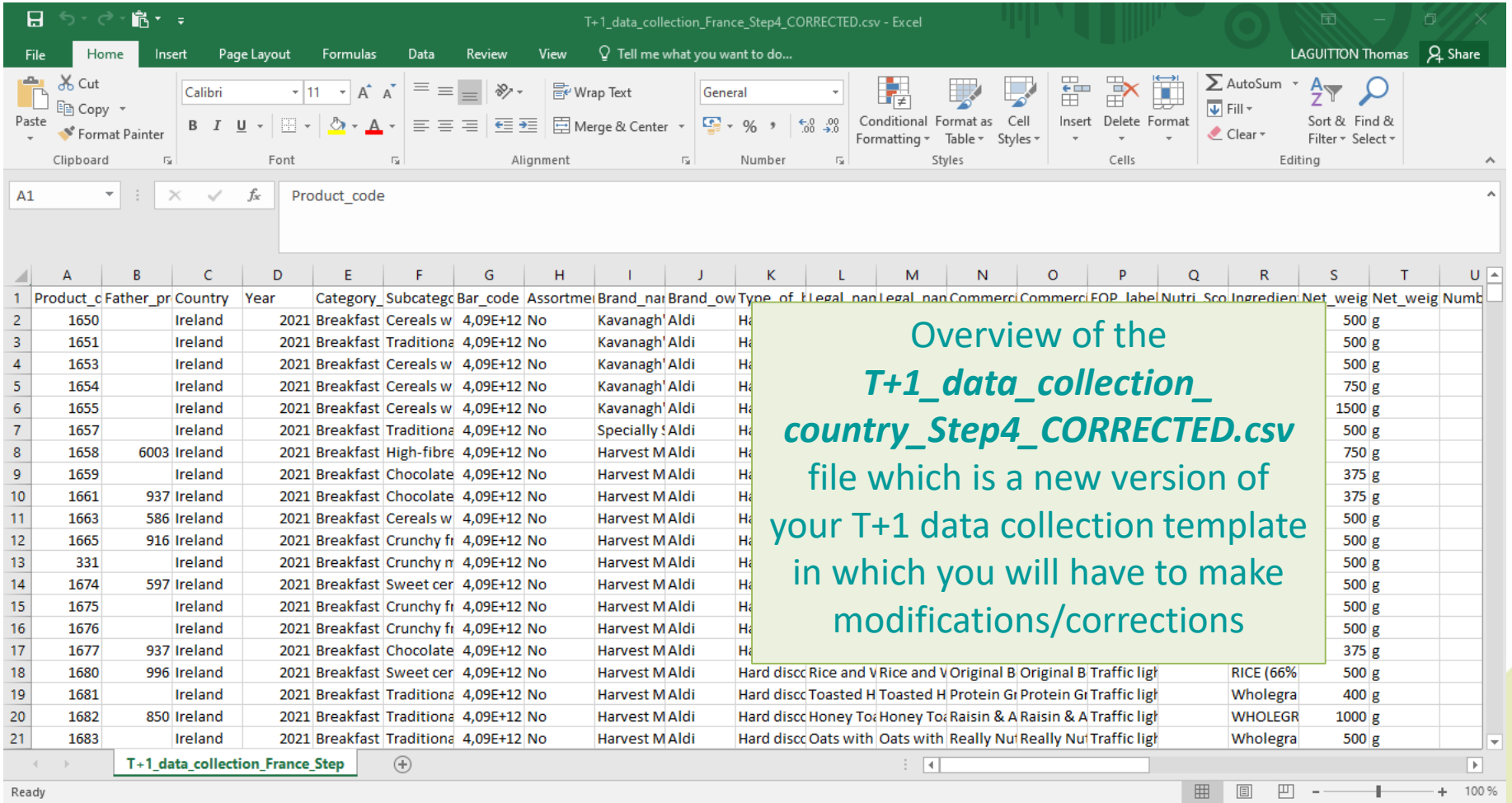
Create a copy of the file exported by R

→ You need to create a copy of the file T+1_data_collection_*country*_Step4_VERIFIED.csv and rename it "T+1_data_collection_*country*_Step4_CORRECTED.csv"

It is on this file that you will make the corrections and modifications indicated

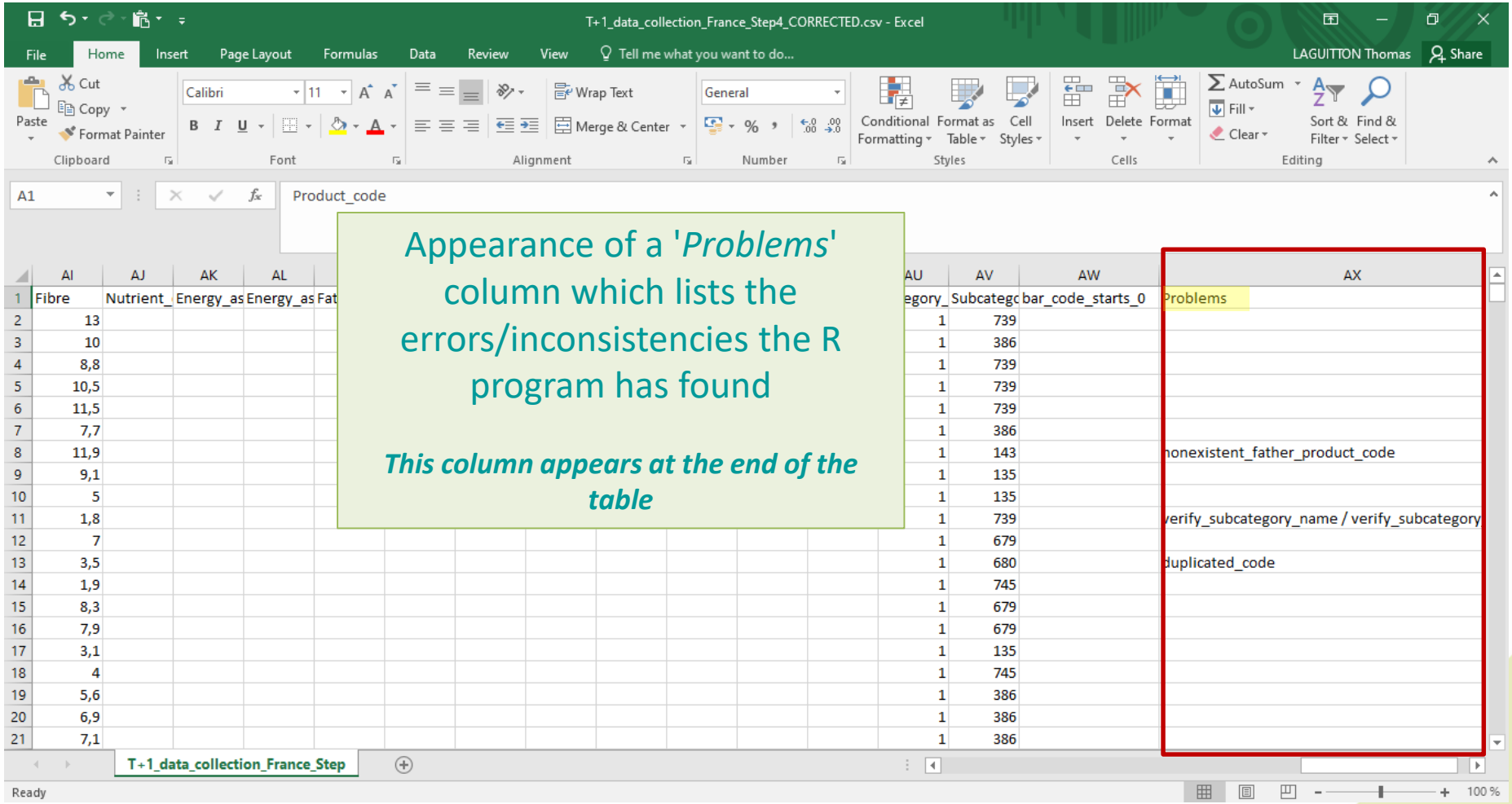
Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_4' program



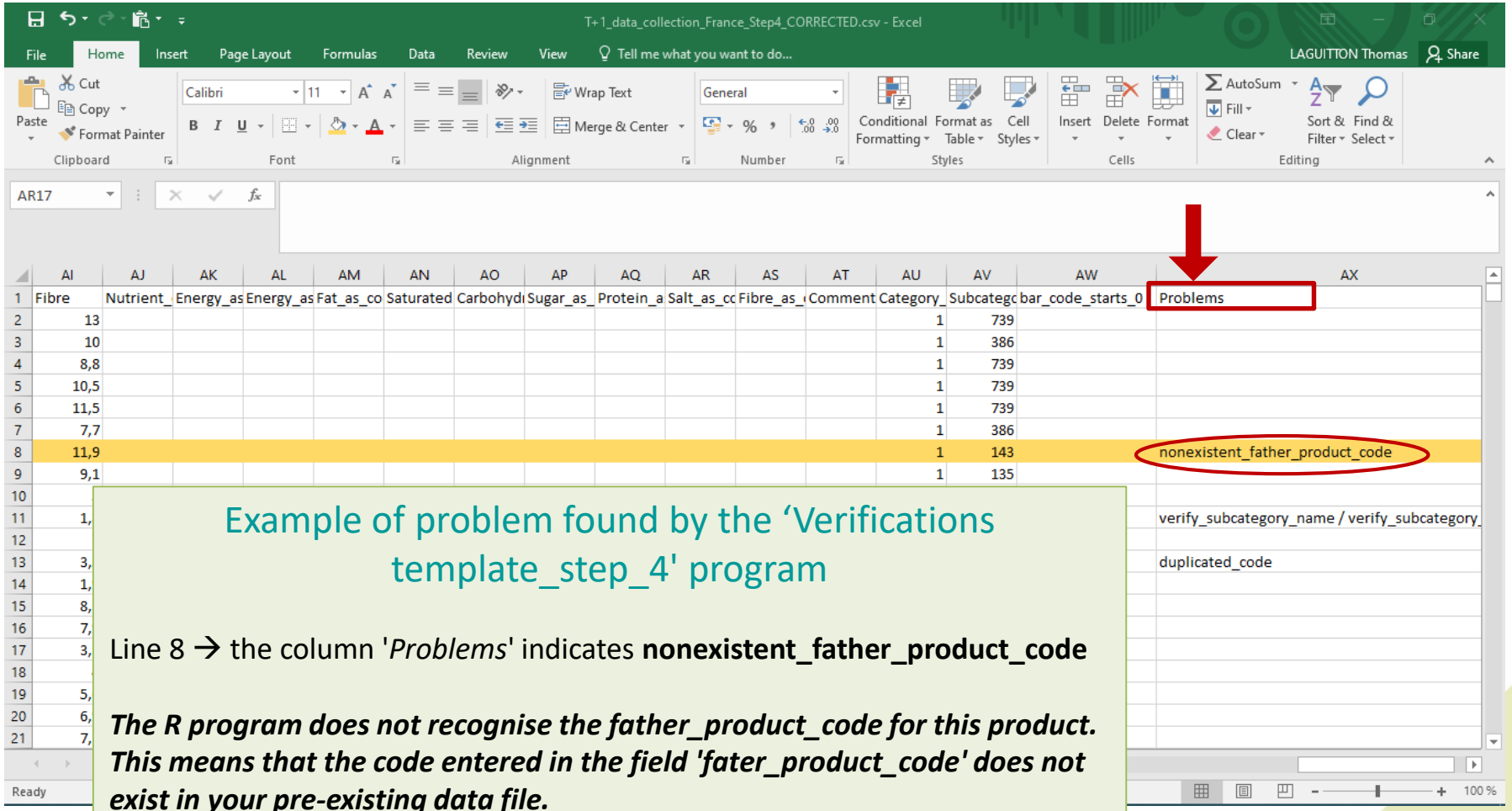
Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_4' program



Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_4' program



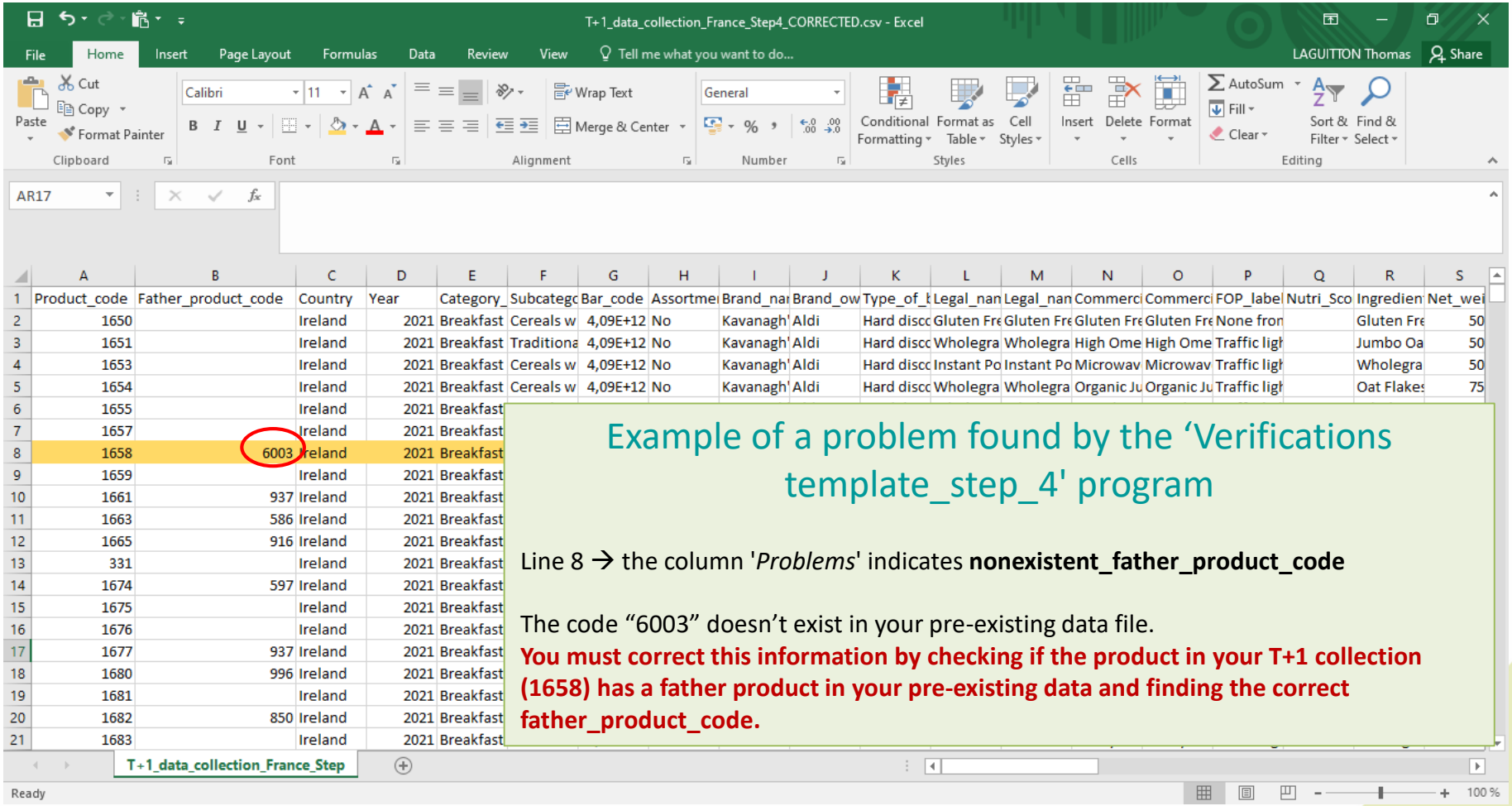
Example of problem found by the 'Verifications template_step_4' program

Line 8 → the column '*Problems*' indicates **nonexistent_father_product_code**

The R program does not recognise the father_product_code for this product. This means that the code entered in the field 'fater_product_code' does not exist in your pre-existing data file.

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_4' program



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Product_code	Father_product_code	Country	Year	Category_	Subcategc	Bar_code	Assortmer	Brand_nar	Brand_ow	Type_of_t	Legal_nan	Legal_nan	Commerci	Commerci	FOP_label	Nutri_Sco	Ingredien	Net_wei
2	1650		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi	Hard disc	Gluten Fre	Gluten Fre	Gluten Fre	Gluten Fre	None fron		Gluten Fre	50
3	1651		Ireland	2021	Breakfast	Traditiona	4,09E+12	No	Kavanagh'	Aldi	Hard disc	Wholegra	Wholegra	High Ome	High Ome	Traffic ligh		Jumbo Oa	50
4	1653		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi	Hard disc	Instant Po	Instant Po	Microwav	Microwav	Traffic ligh		Wholegra	50
5	1654		Ireland	2021	Breakfast	Cereals w	4,09E+12	No	Kavanagh'	Aldi	Hard disc	Wholegra	Wholegra	Organic Ju	Organic Ju	Traffic ligh		Oat Flakes	75
6	1655		Ireland	2021	Breakfast														
7	1657		Ireland	2021	Breakfast														
8	1658	6003	Ireland	2021	Breakfast														
9	1659		Ireland	2021	Breakfast														
10	1661	937	Ireland	2021	Breakfast														
11	1663	586	Ireland	2021	Breakfast														
12	1665	916	Ireland	2021	Breakfast														
13	331		Ireland	2021	Breakfast														
14	1674	597	Ireland	2021	Breakfast														
15	1675		Ireland	2021	Breakfast														
16	1676		Ireland	2021	Breakfast														
17	1677	937	Ireland	2021	Breakfast														
18	1680	996	Ireland	2021	Breakfast														
19	1681		Ireland	2021	Breakfast														
20	1682	850	Ireland	2021	Breakfast														
21	1683		Ireland	2021	Breakfast														

Example of a problem found by the 'Verifications template_step_4' program

Line 8 → the column '*Problems*' indicates **nonexistent_father_product_code**

The code "6003" doesn't exist in your pre-existing data file.

You must correct this information by checking if the product in your T+1 collection (1658) has a father product in your pre-existing data and finding the correct father_product_code.

Co-funded by the European Union's Health Programme (2014-2020)

Excel file to modify after 'Verifications template_step_4' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_4' program, their meaning and what to do

Problem	Meaning	Action
Nonexistent_father_product_code	The **father_product_code** associated with this product does not exist in your pre-existing data.	→ Check if the product has a father product in the pre-existing data file using information such as the product's subcategory, brand name, commercial_name elements, etc. → If the product has a father product, enter the correct product_code from the pre-existing data in the '*Father_product_code*' field. → If the product does not have a father product, delete the wrong *father_product_code* that has been entered.
Duplicated_code	The **product_code** of this product is already assigned to a product in your pre-existing data.	→ You must assign a new product code to this product, making sure that it does not exist in your 'T+1_data_collection_***country***_step_4_CORRECTED.csv' file or in your pre-existing data file.
Not_paired_with_year_of_interest	• The **father product** does not belong to the pre-existing data collection year of interest chosen to make comparisons/indicators with T+1 data. This means that this pairing will not be taken into account for making the indicators.	→ You must check if the T+1 product cannot be paired with another product in your preexisting data file which has been collected on the year of interest for the category. *Please note that the products not paired with the year of interest will not be considered as paired products for the computing of Indicators (the pairing will not be taken into account for this work)*

Excel file to modify after 'Verifications template_step_4' program

Terms that may appear in the 'Problems' field following the 'Verifications template_step_4' program, their meaning and what to do

Problem	Meaning	Action
• **Verify_category_name** • **Verify_category_code**	The product and its father product have **different category names and codes**	→ You must check that you have classified the product of your T+1 Best-ReMaP collection in the **right category**. → If the problem is with a product in your pre-existing data that is classified in the wrong category, you must correct the error in your pre-existing data and return the file to us.
• **Verify_subcategory_name** • **Verify_subcategory_code**	The product and its father product have **different subcategory names and codes**	→ You must check that you have classified the product of your T+1 Best-ReMaP collection in the **right subcategory**. → If the problem is with a product in your pre-existing data that is classified in the wrong category, you must correct the error in your pre-existing data and return the file to us.

WARNING! This is not necessarily an error, the product of the T+1 Best-ReMaP collection may have changed subcategory compared to the pre-existing data.

Example: the product has been reformulated and the sugar has been removed, etc

You must still check that you have not made a classification mistake.

Co-funded by the European Union's Health Programme (2014-2020)

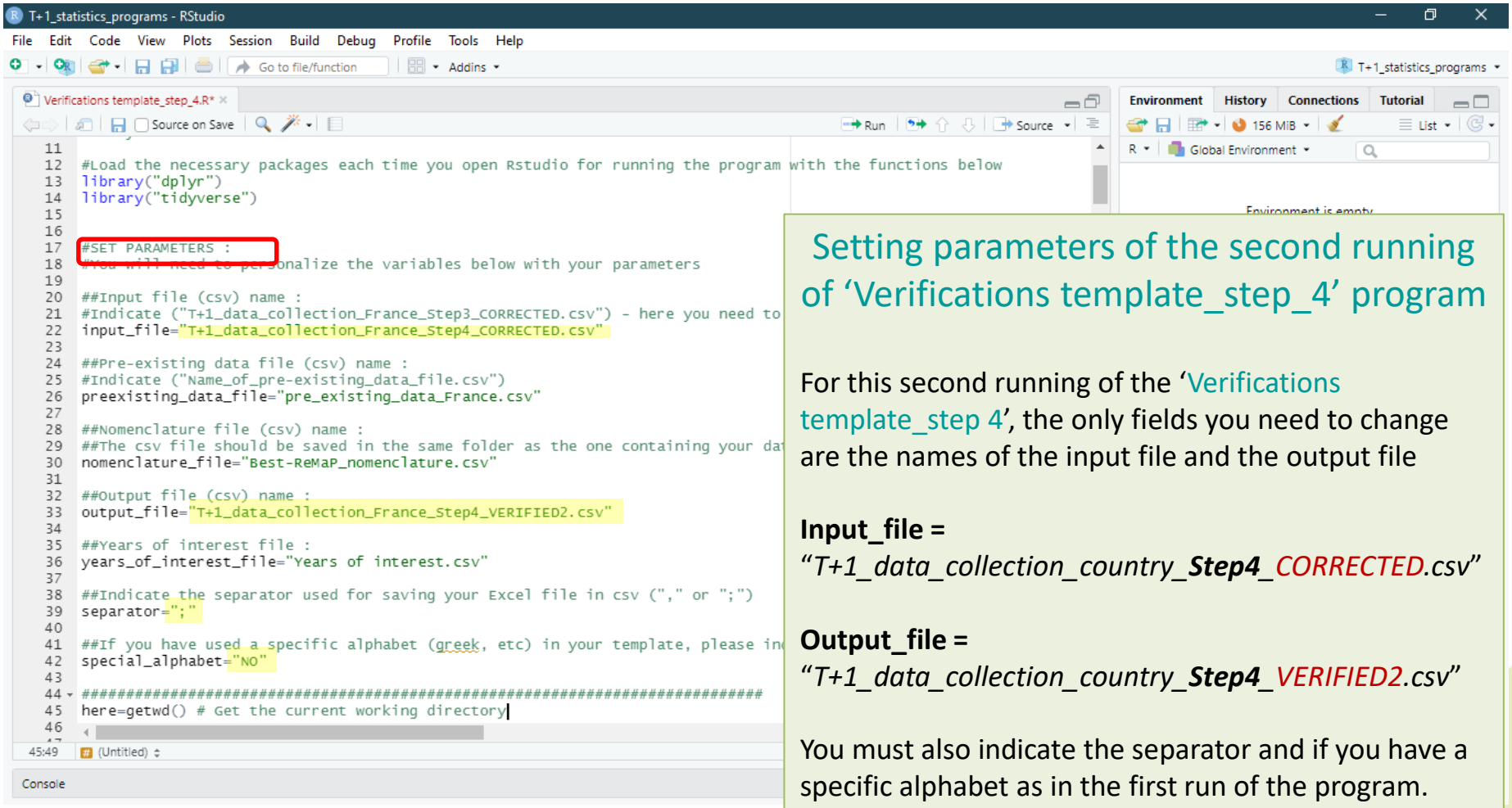
Excel file to modify after ‘Verifications template_step_4’ program

- **Be careful! When saving the file, you must be sure that the barcodes appear in full and not in scientific format. If this is the case, you must repeat the procedure on pages 24→28.**
- Once the corrections have been made and the barcodes are in the correct format, you can save the file *T+1_data_collection_country_Step3_CORRECTED.csv* and close it.
- You will have to run this corrected file again in the ‘Verifications template_step_4' program to make sure you haven't missed a check.

Your Rstudio interface must be cleaned up before running the ‘Verifications template_step_4’ program again.
All cleaning steps are described on pages 62→68.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_4' program



Setting parameters of the second running of 'Verifications template_step_4' program

For this second running of the 'Verifications template_step 4', the only fields you need to change are the names of the input file and the output file

Input_file =
"*T+1_data_collection_country_**Step4**_CORRECTED.csv*"

Output_file =
"*T+1_data_collection_country_**Step4**_VERIFIED2.csv*"

You must also indicate the separator and if you have a specific alphabet as in the first run of the program.

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_4' program

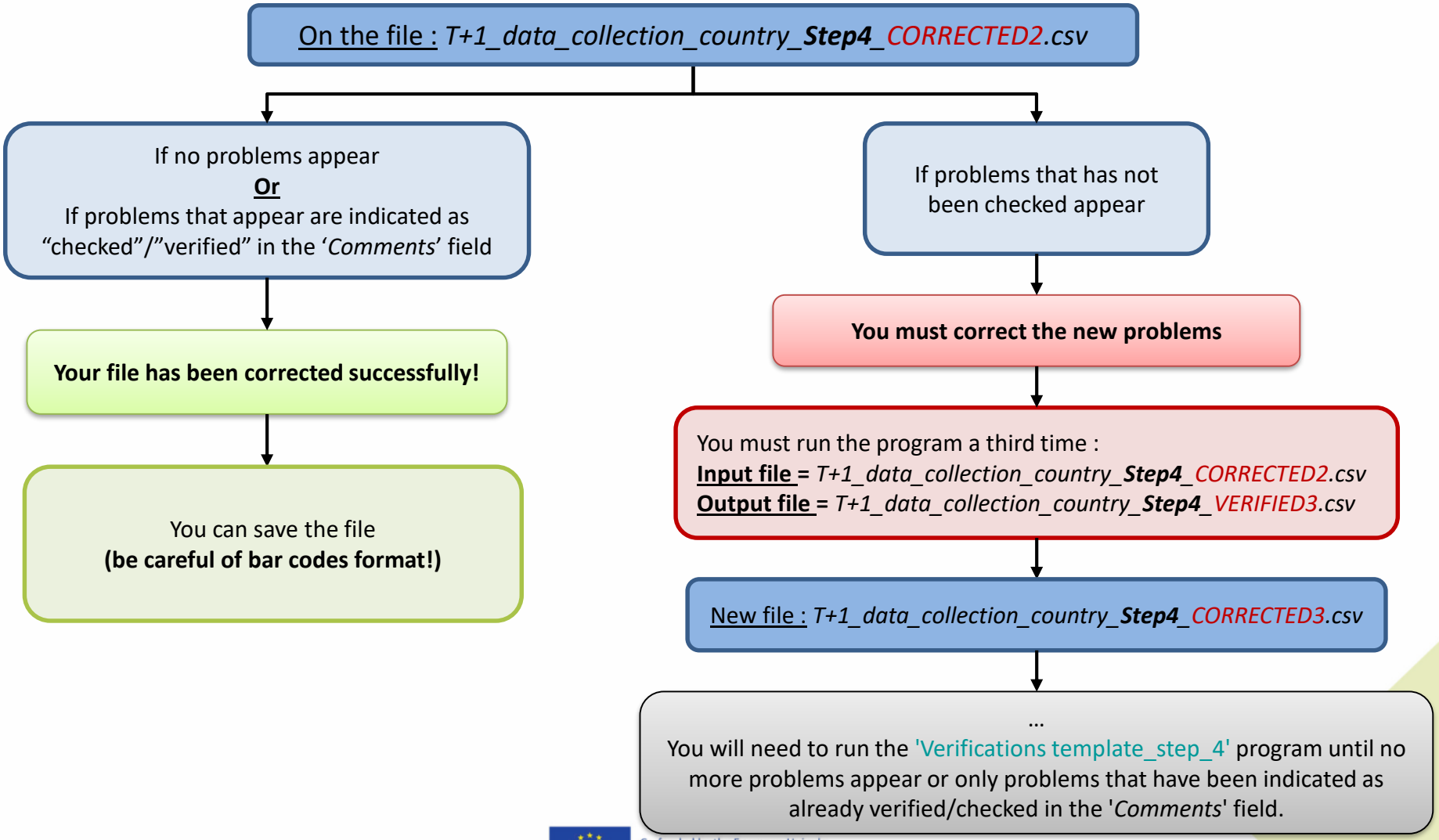
- At the end of this second run, you get in your "files" folder a file called: "*T+1_data_collection_country_**Step4**_VERIFIED2.csv*"

- You must create a copy of this file and call it : "*T+1_data_collection_country_**Step4**_CORRECTED2.csv*"

  → It is on this file that you will make the modifications following the checks

Co-funded by the European Union's Health Programme (2014-2020)

2nd running of 'Verifications template_step_4' program



Co-funded by the European Union's Health Programme (2014-2020)

End of the 4 verification programs

→ You should now have a file called : *T+1_data_collection_country_Step4_CORRECTED(X).csv*
((X) is the number of the last file exported and corrected after the last run of the second verification program)

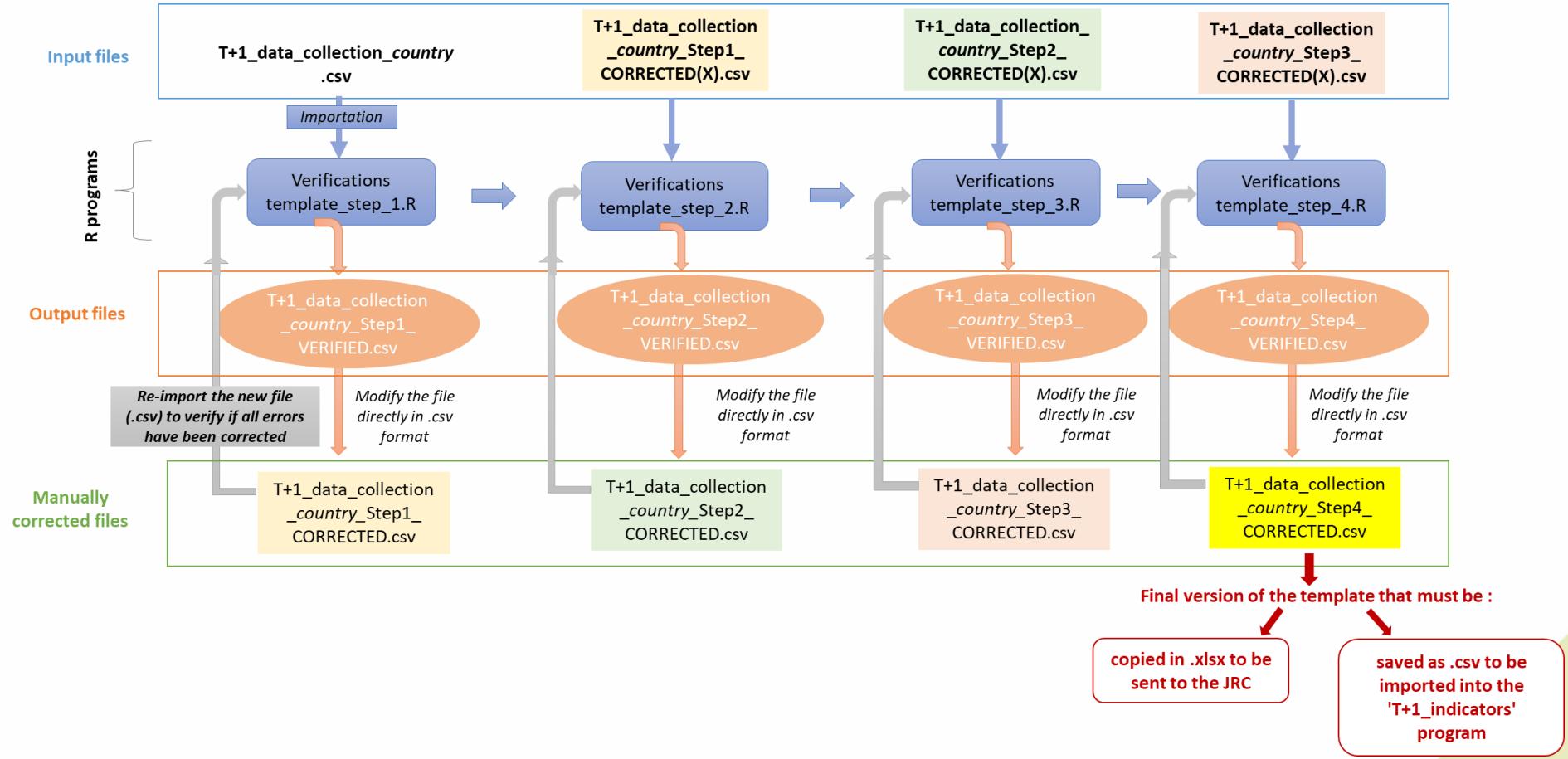
This file is the final version of your data collection template after you have done all the checks and corrected all the errors.

→ You must create a copy of this file and save it in **.xlsx format**
(You can call this file : *T+1_data_collection_country_final.xlsx* for example)

This copy in **.xlsx format** will be the final version of your data collection template that will be transmitted to the **JRC**.

Co-funded by the European Union's Health Programme (2014-2020)

Summary of the 4 verification programs

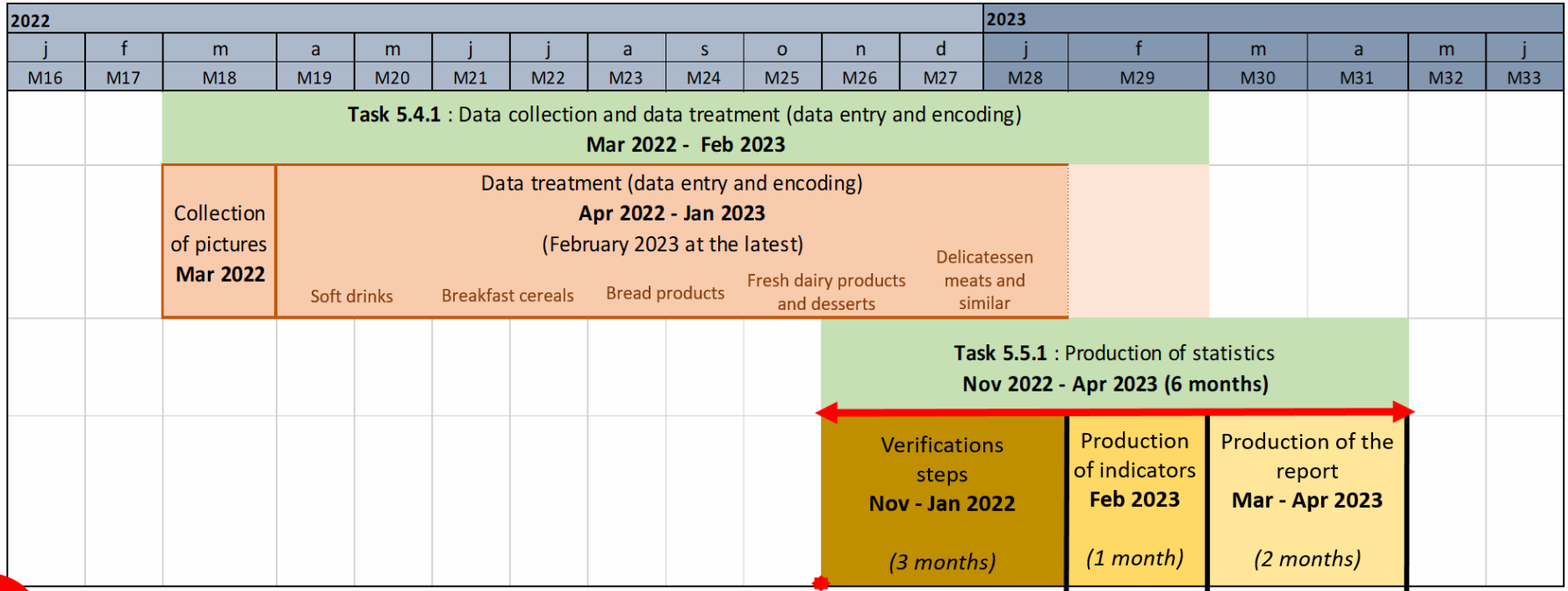


Co-funded by the European Union's Health Programme (2014-2020)

4) Next steps

Co-funded by the European Union's Health Programme (2014-2020)

Timeline of the task 5.5.1



1st Nov 2022

Please send us feedback to confirm that the R and Rstudio software are installed
If you have any problems before this date, please let us know so that we can help you

These deadlines are indications so that you can organize yourselves as well as possible

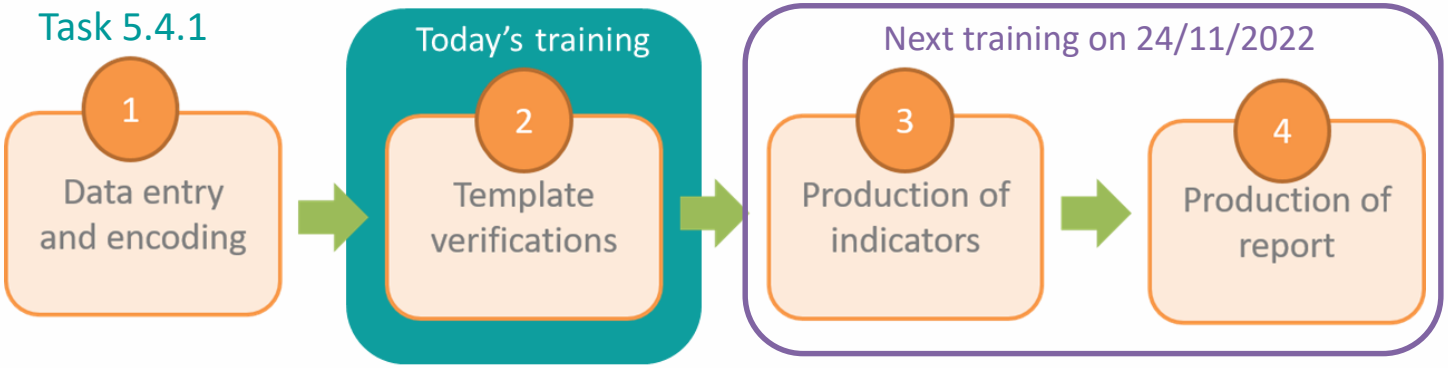
End of verifications and corrections to be made

You should have all figures and tables ready and stored in the "indicators" file

Report ready to be sent to WP5

Co-funded by the European Union's Health Programme (2014-2020)

Next steps



It is important to finish each step before moving on to the next

- Downloading of the R and Rstudio software and the documents needed for task 5.5.1 (working folder on the intranet) **before 1st November 2022**.
  Please give us feedback so that we can help you as much as possible if you have difficulties downloading.

- A next training is planned for **24th November 2022** to present steps 3 and 4 (production of indicators and report)

- If you have any questions or need help at any time during the completion of the steps in Task 5.3.2, you can contact us at: wp5_bestremap@anses.fr
  We will also be able to hold individual meetings to best assist you during this task.

Co-funded by the European Union's Health Programme (2014-2020)

Best-ReMaP

Healthy Food for a Healthy Future

Thank you for your attention!

ANSES

wp5_bestremap@anses.fr

The Joint Action focusing on the implementation of validated best practices in nutrition – Best-ReMaP

This presentation arises from the Joint Action Best-ReMaP. This JA is addressing the adaption, replication and implementation of effective health interventions, based on practices that have been proven to work in the areas of food reformulation, framing of food marketing and public procurement of healthy food in public settings. This presentation was funded by the European Union's Health Programme (2014-2020). The content of this presentation represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency (CHAFEA) or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains.