UK Data Service

Census flow data workshop

Handout instructions #2

Downloading 2021 census flow data and analysing the data programmatically

Vassilis Routsis & Oliver Duke-Williams

January 2025

Getting started

- Start up a browser
- Screenshots of process

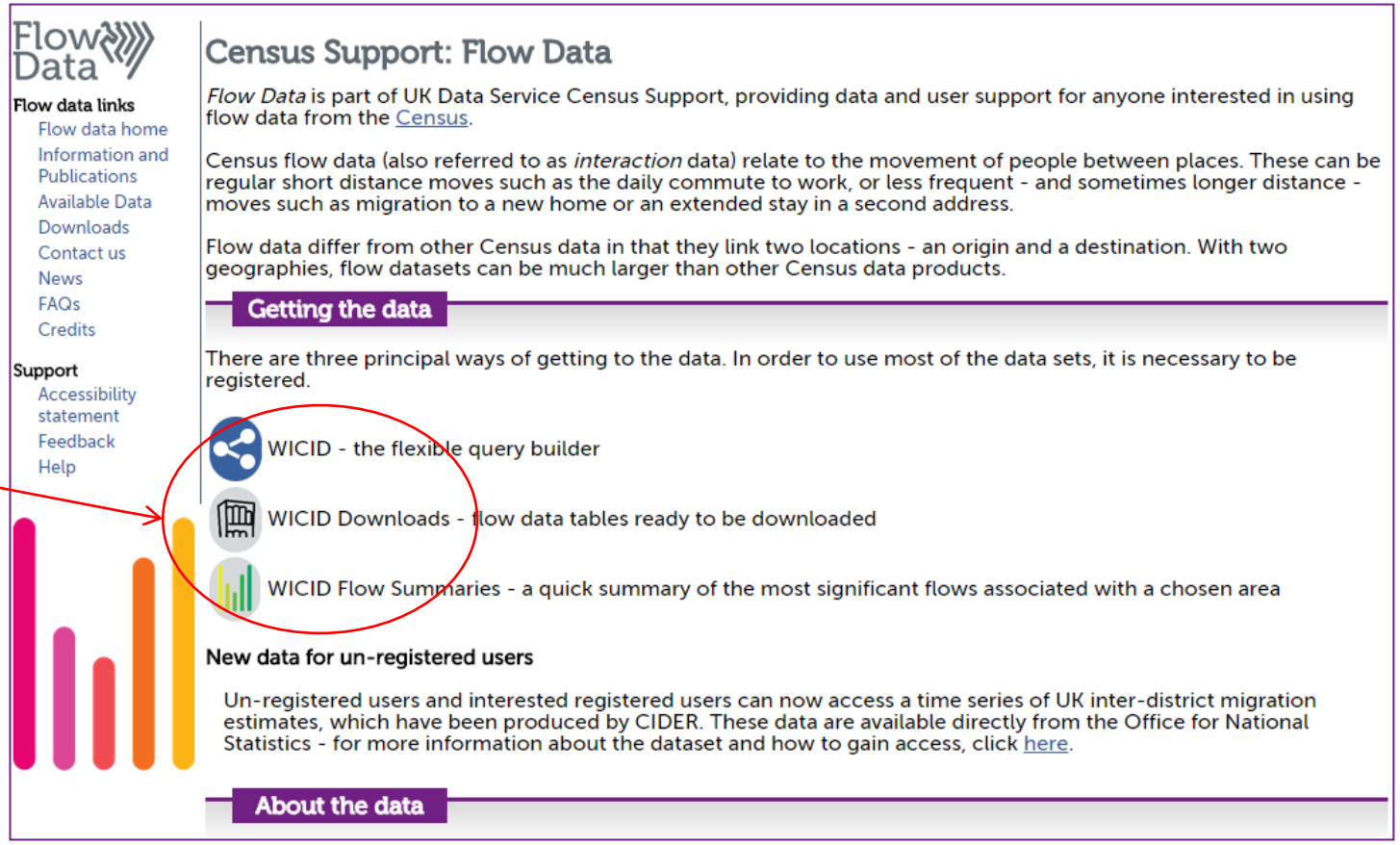
Flow data website

For quick access you can follow the link to

- https://wicid.ukdataservice.ac.uk/

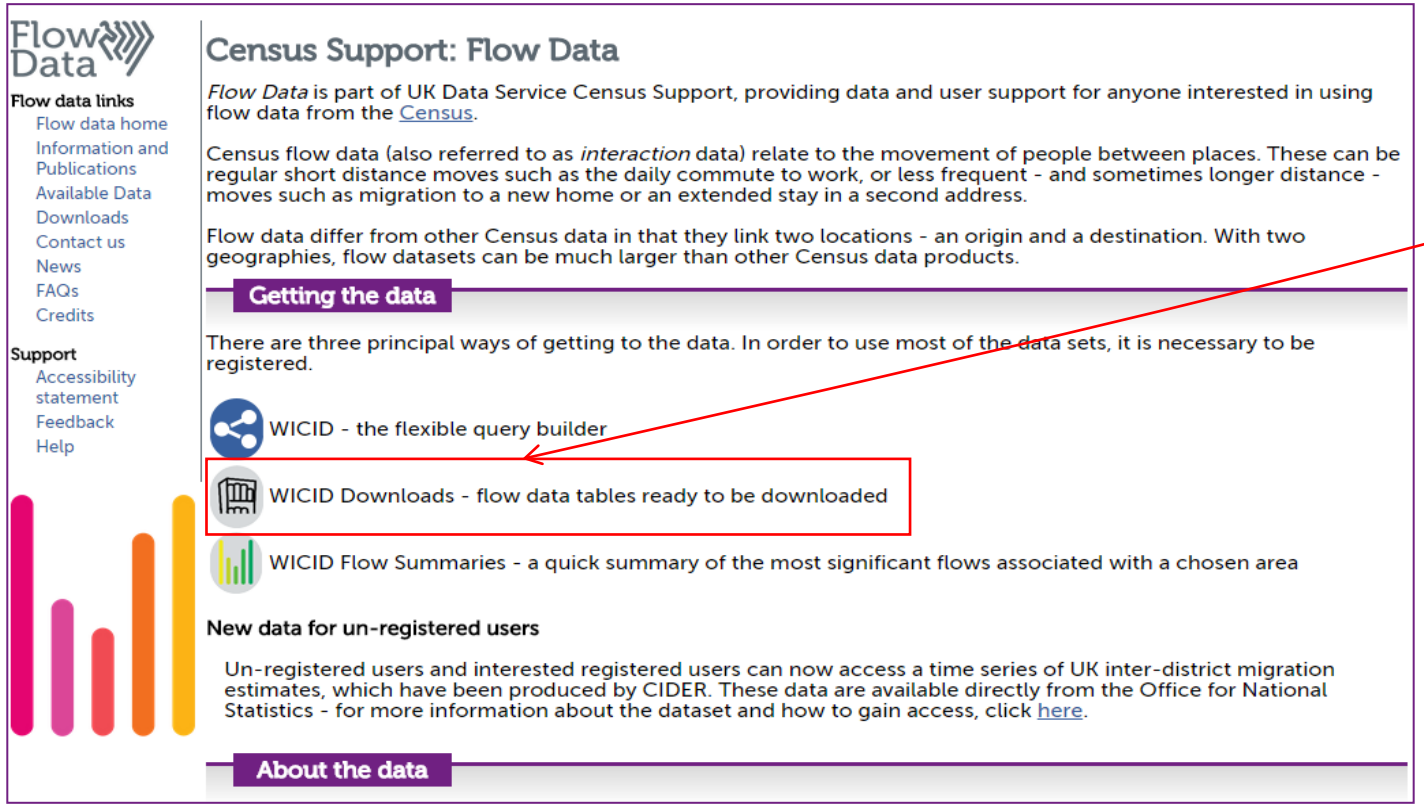
Flow data homepage (i)

- Three main routes to data
  - Flexible queries
  - Downloads
  - Summaries



WICID Query builder

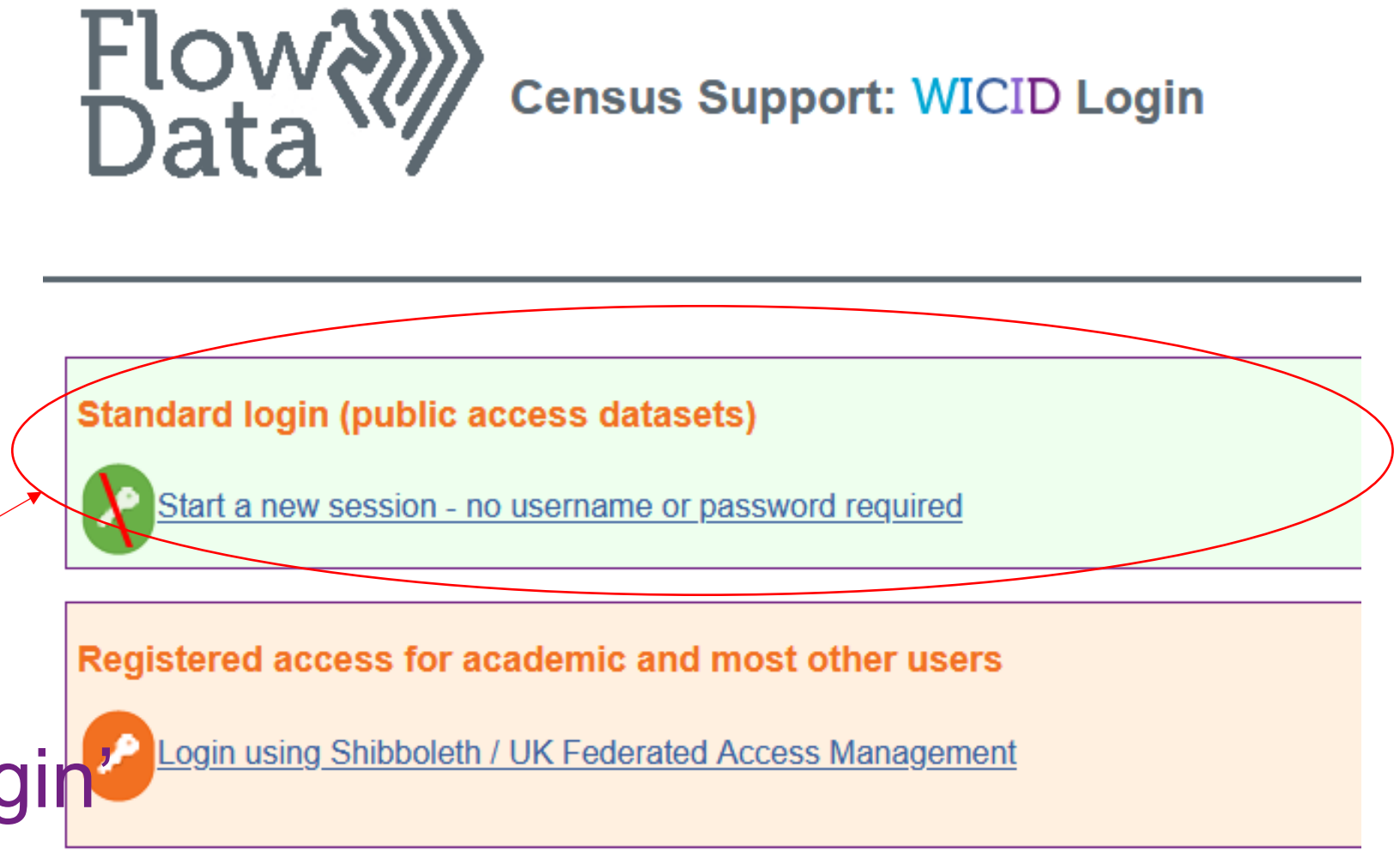
- The WICID general query builder allows users to identify and extract a subset of flows in which they are interested
  - It also provides built-in support for area code labelling
  - Select '**WICID Downloads**' button



Login screen

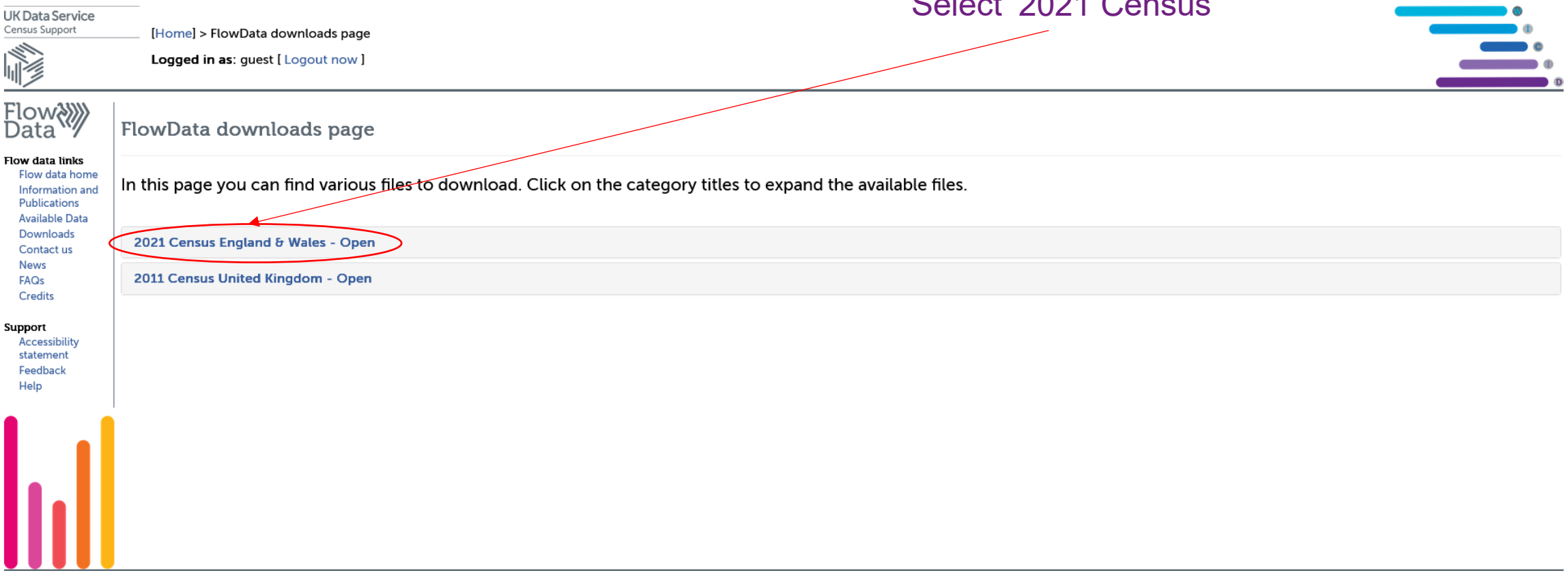
- For use with OGL/Public data, we can use 'Standard Login'
- Registered access offers open data AND safeguarded data

Select ‘Standard login’



WICID downloads page

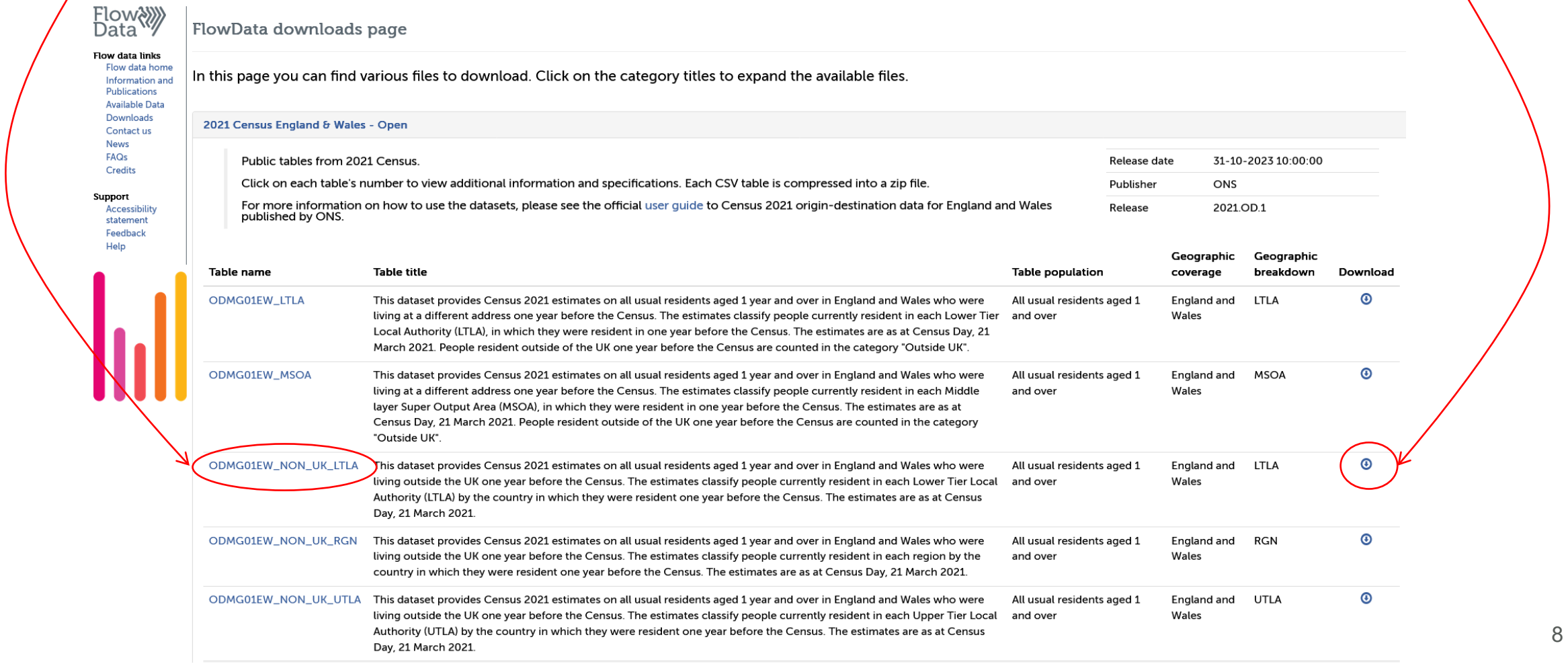
Select ‘2021 Census’



WICID download 2021 tables

We will work with the closest equivalent dataset to its 2011 version, table ODMG01EW_NON_UK_LTLA. Clicking on the table name will display some basic information about the table.

Clicking on the ‘Download’ button will download the dataset



Flow Data

Flow data links
- Flow data home
- Information and Publications
- Available Data
- Downloads
- Contact us
- News
- FAQs
- Credits

Support
- Accessibility statement
- Feedback
- Help

FlowData downloads page

In this page you can find various files to download. Click on the category titles to expand the available files.

2021 Census England & Wales - Open

Public tables from 2021 Census.

Click on each table's number to view additional information and specifications. Each CSV table is compressed into a zip file.

For more information on how to use the datasets, please see the official user guide to Census 2021 origin-destination data for England and Wales published by ONS.

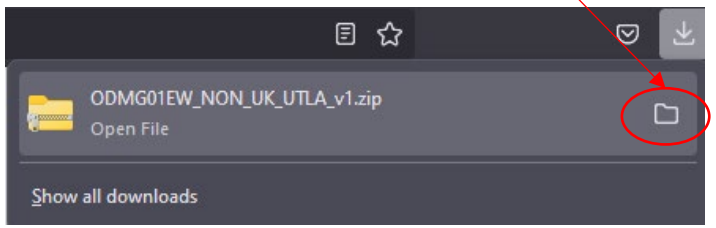
Release date	31-10-2023 10:00:00
Publisher	ONS
Release	2021.OD.1

Table name	Table title	Table population	Geographic coverage	Geographic breakdown	Download
ODMG01EW_LTLA	This dataset provides Census 2021 estimates on all usual residents aged 1 year and over in England and Wales who were living at a different address one year before the Census. The estimates classify people currently resident in each Lower Tier Local Authority (LTLA), in which they were resident in one year before the Census. The estimates are as at Census Day, 21 March 2021. People resident outside of the UK one year before the Census are counted in the category "Outside UK".	All usual residents aged 1 and over	England and Wales	LTLA	
ODMG01EW_MSOA	This dataset provides Census 2021 estimates on all usual residents aged 1 year and over in England and Wales who were living at a different address one year before the Census. The estimates classify people currently resident in each Middle layer Super Output Area (MSOA), in which they were resident in one year before the Census. The estimates are as at Census Day, 21 March 2021. People resident outside of the UK one year before the Census are counted in the category "Outside UK".	All usual residents aged 1 and over	England and Wales	MSOA	
ODMG01EW_NON_UK_LTLA	This dataset provides Census 2021 estimates on all usual residents aged 1 year and over in England and Wales who were living outside the UK one year before the Census. The estimates classify people currently resident in each Lower Tier Local Authority (LTLA) by the country in which they were resident one year before the Census. The estimates are as at Census Day, 21 March 2021.	All usual residents aged 1 and over	England and Wales	LTLA	
ODMG01EW_NON_UK_RGN	This dataset provides Census 2021 estimates on all usual residents aged 1 year and over in England and Wales who were living outside the UK one year before the Census. The estimates classify people currently resident in each region by the country in which they were resident one year before the Census. The estimates are as at Census Day, 21 March 2021.	All usual residents aged 1 and over	England and Wales	RGN	
ODMG01EW_NON_UK_UTLA	This dataset provides Census 2021 estimates on all usual residents aged 1 year and over in England and Wales who were living outside the UK one year before the Census. The estimates classify people currently resident in each Upper Tier Local Authority (UTLA) by the country in which they were resident one year before the Census. The estimates are as at Census Day, 21 March 2021.	All usual residents aged 1 and over	England and Wales	UTLA	

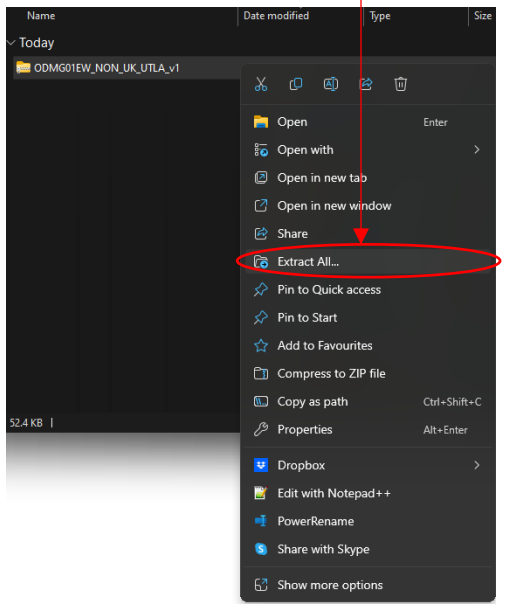
WICID download 2021 tables

The dataset is archived in a ZIP file so you need to extract it.

1. Open the folder where the downloaded ZIP file is



2. Extract the CSV file from the ZIP archive and then Open it with Microsoft Excel.

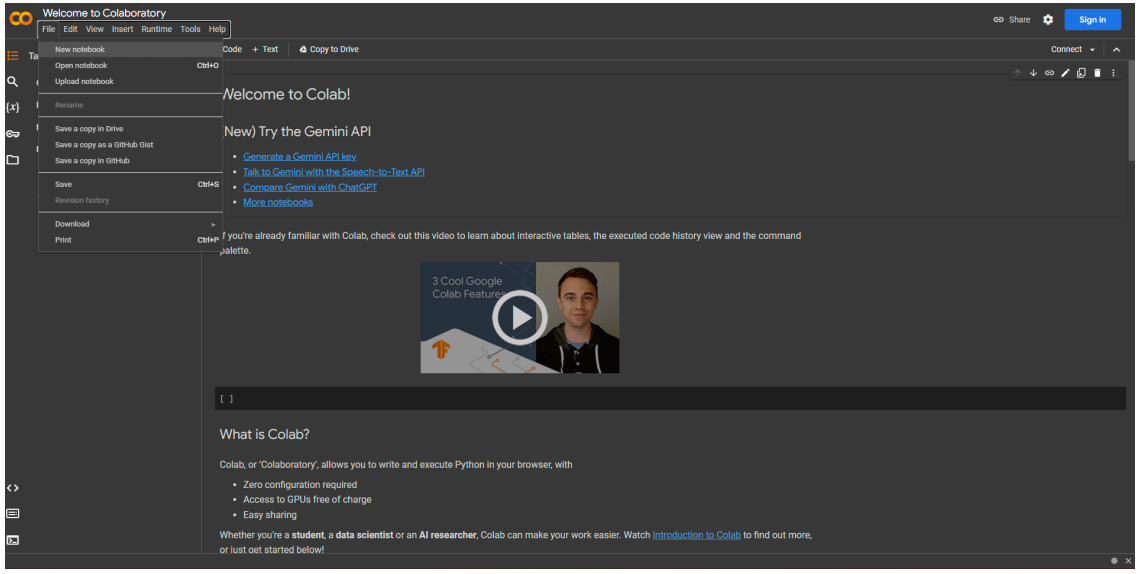


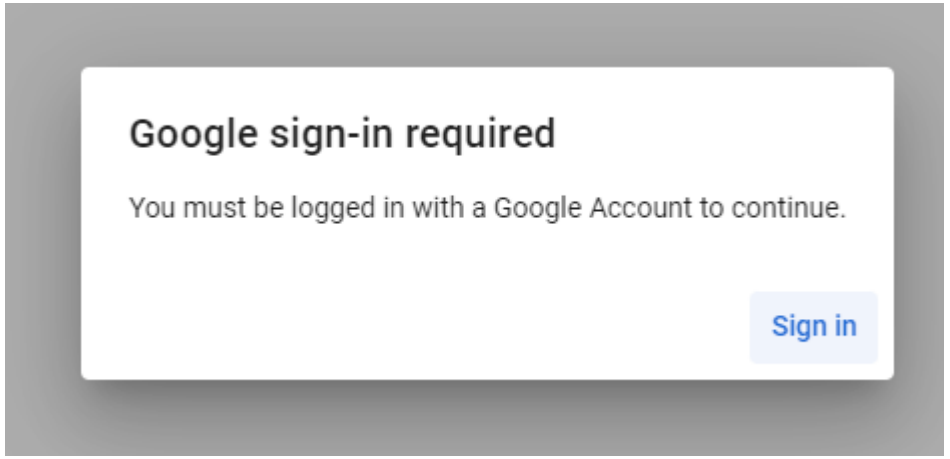
Google Colab

Google Colab is a powerful tool which we can use to run code without the need to install any software in our local machines.

https://colab.research.google.com/

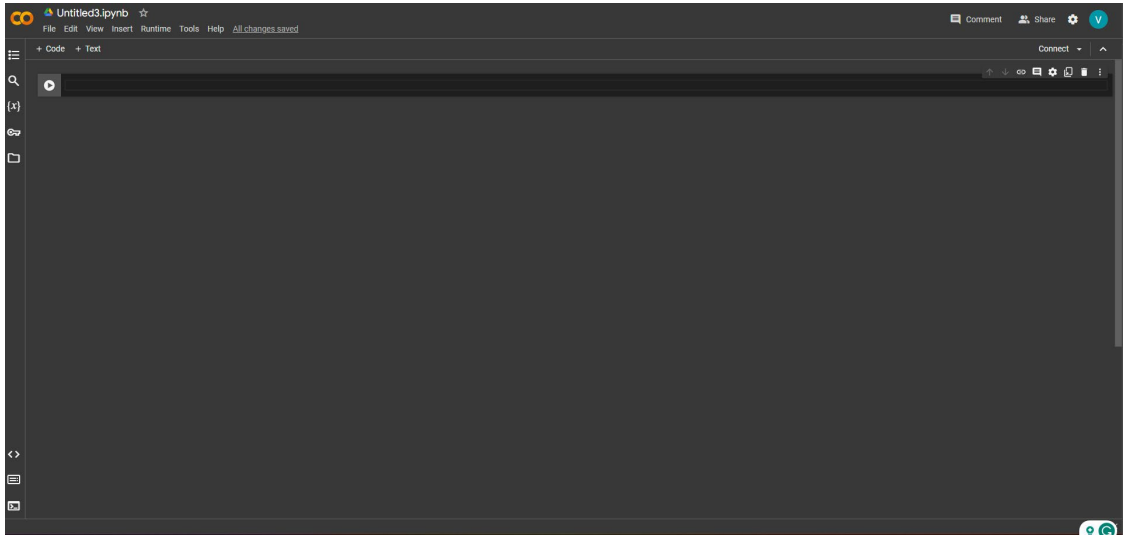
Click File->New notebook and login using your Google account

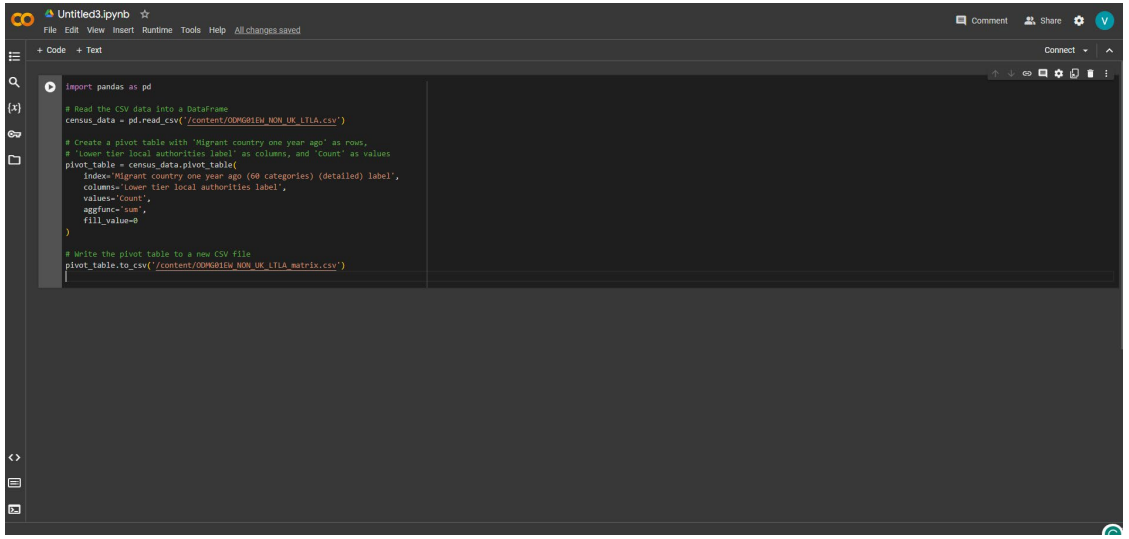




Google Colab

After successful login, we are ready to run our code. Click on the input box next to the play button and paste the code from **exercise2_1_convert_to_matrix_simple.py**.



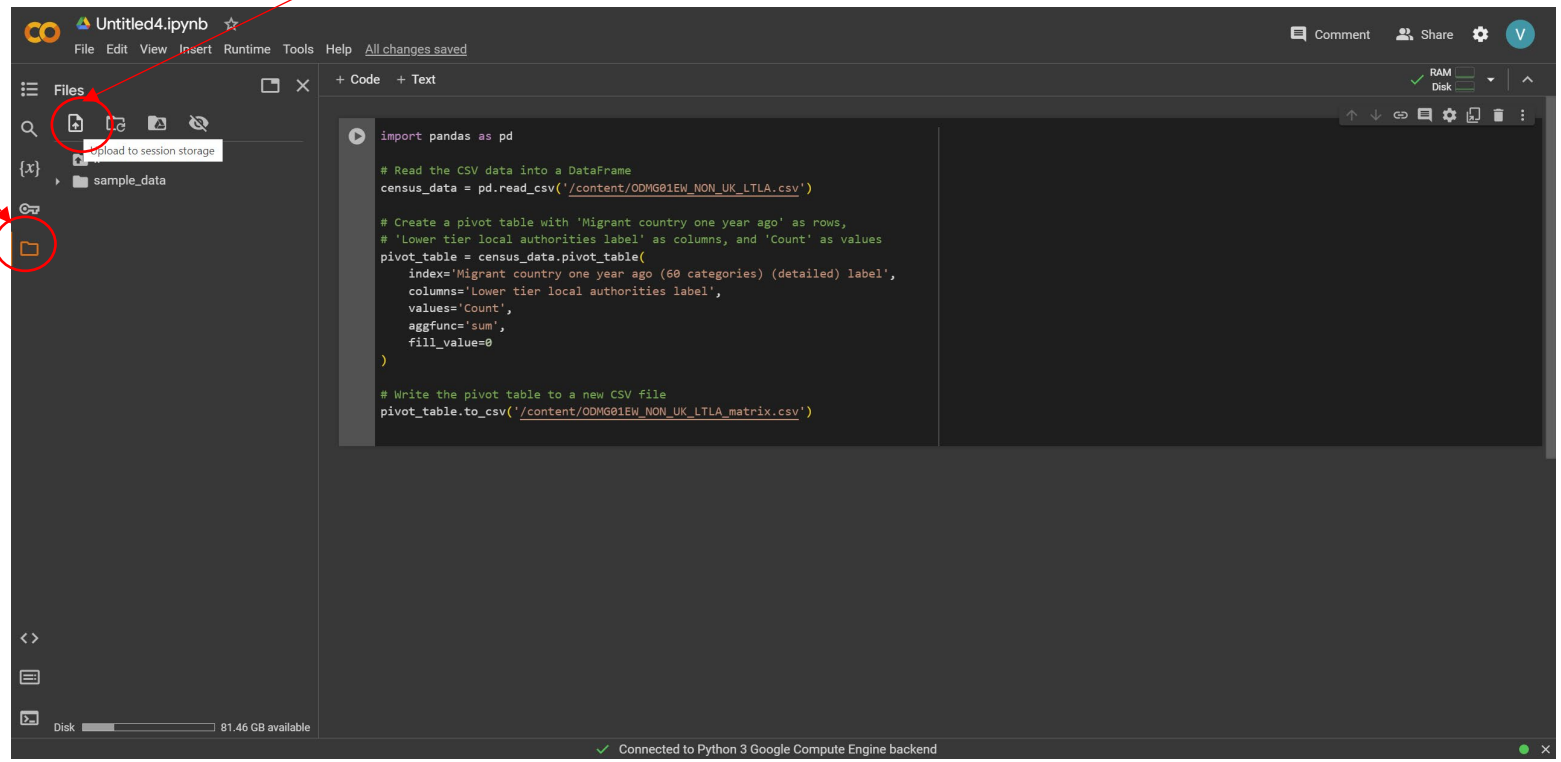


Google Colab

Upload ODMG01EW_NON_UK_LTLA.csv into the Colab environment.

1. Open the files tab

2. Upload the file using the upload dialog box or drag and drop it into storage below the sample_data folder

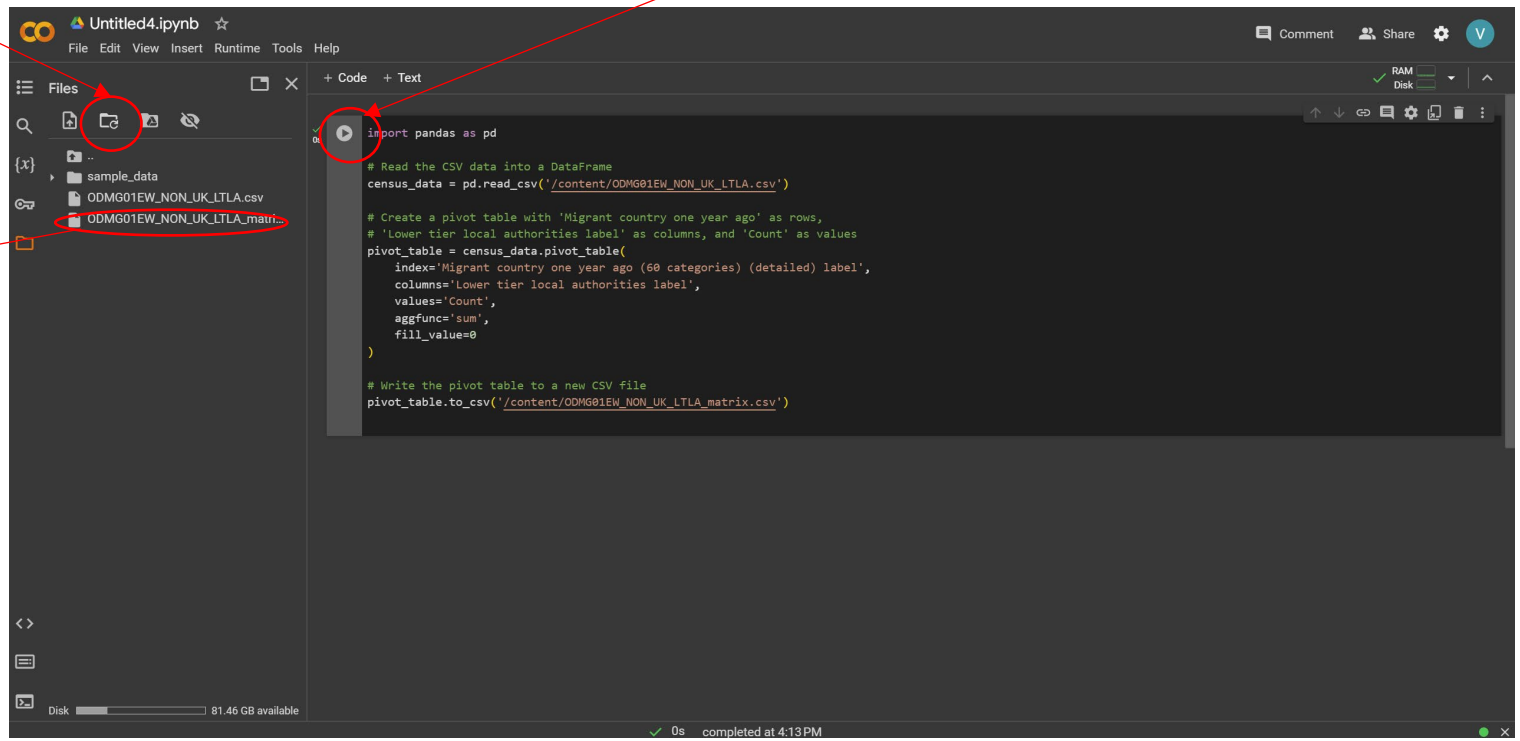


Google Colab

Upload ODMG01EW_NON_UK_LTLA.csv into the Colab environment.

2. Refresh the storage after processing has finished or the generated file might not be visible

1. Execute the script by clicking the play button



3. Hover the mouse over the generated file, click on the three vertical dots that appear and select Download from the context menu

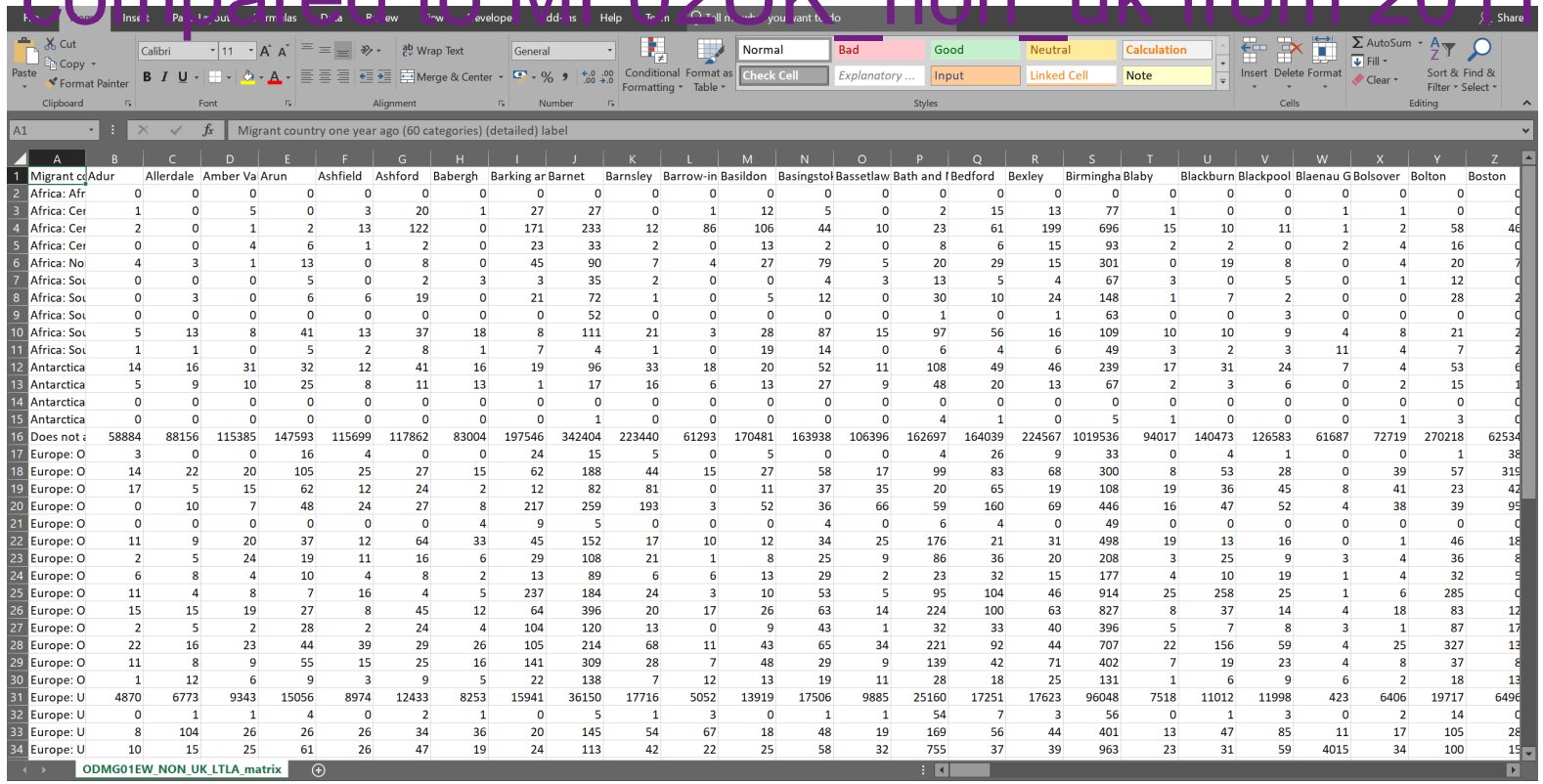
Google Colab

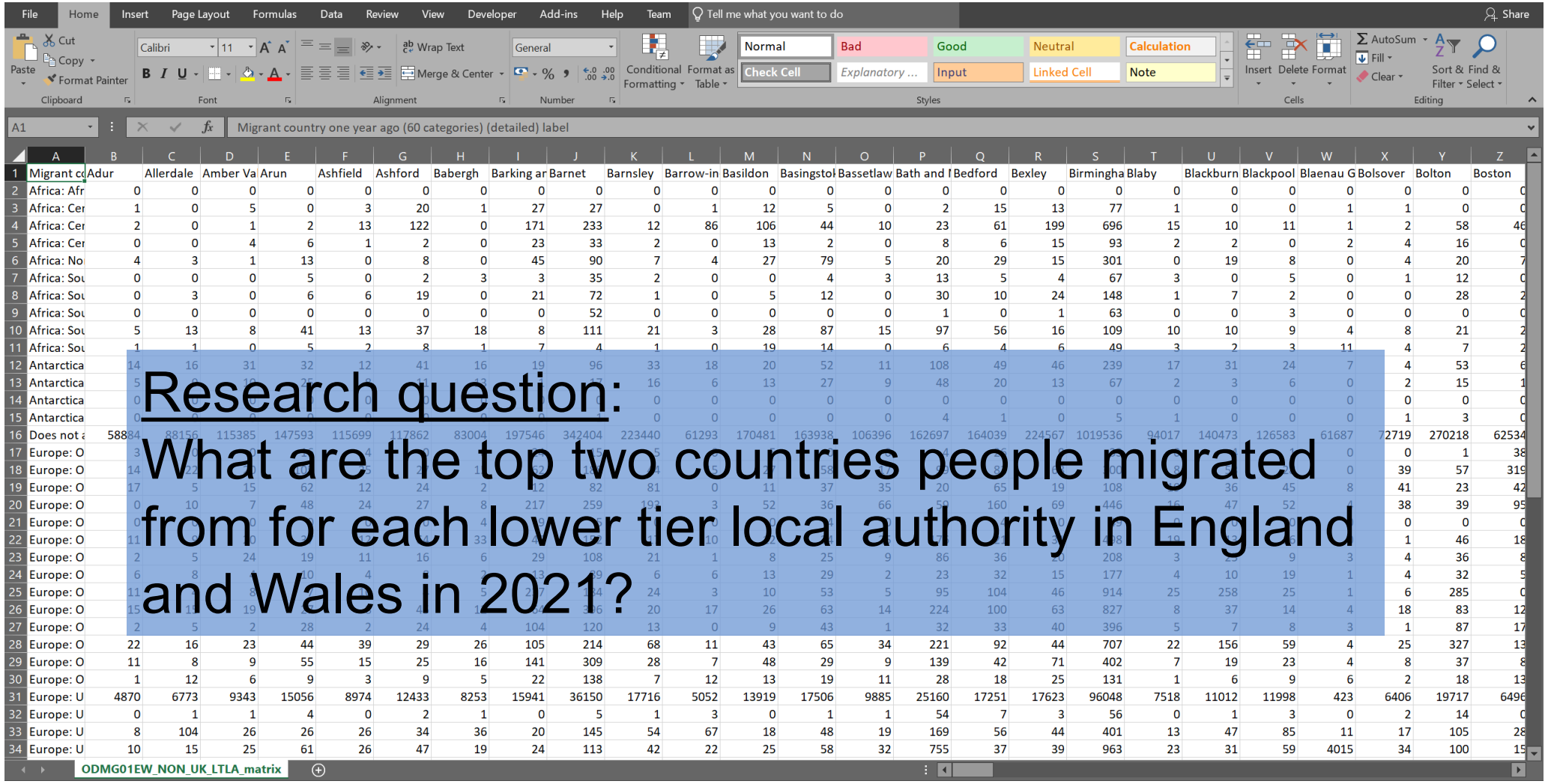
```
import pandas as pd

# Read the CSV data into a DataFrame
census_data = pd.read_csv('/content/ODMG01EW_NON_UK_LTLA.csv')

# Create a pivot table with 'Migrant country one year ago' as rows,
# 'Lower tier local authorities label' as columns, and 'Count' as values
pivot_table = census_data.pivot_table(
    index='Migrant country one year ago (60 categories) (detailed) label',
    columns='Lower tier local authorities label',
    values='Count',
    aggfunc='sum',
    fill_value=0
)

# Write the pivot table to a new CSV file
pivot_table.to_csv('/content/ODMG01EW_NON_UK_LTLA_matrix.csv')
```

Exercises with downloaded data

Open ODMG01EW_NON_UK_LTLA_matrix.csv with Microsoft Excel.

Explore it. What immediate differences you notice in the dataset compared to ME02UK_non_uk from 2011 census?





Research question:

What are the top two countries people migrated from for each lower tier local authority in England and Wales in 2021?

Exercises with downloaded data

An immediate problem is that the 2021 dataset contains the numbers of those who have not migrated (‘does not apply’) as well as the internal migrants from the rest of the UK. So, if we only want to see the numbers of overseas migrants ,we need to clean our data and remove the following rows: ‘Does not apply’, ‘Europe: United Kingdom: England’, ‘Europe: United Kingdom: Northern Ireland’, ‘Europe: United Kingdom: Scotland’, ‘Europe: United Kingdom: Wales’. Select these rows by holding the **CTRL** key and left click each one of these row numbers (16, 31, 32, 33, 34) until all of them are grey. While on top of one of the selected rows, **right click** without holding the CTRL key anymore and select **Delete** from the context menu to remove them. You should be left with 54 rows now including the header.

There is. We can remove the undesired rows programmatically. Let's get back to the drawing board!

Upload **exercise2_2_convert_to_matrix_with_exclusions.ipynb** into Google Colab via the File->Upload a notebook option

```
import pandas as pd

# Read the CSV data into a DataFrame
census_data = pd.read_csv('/content/ODMG01EW_NON_UK_LTLA.csv')

# Define the countries we do not want to include
exclusions = ['Does not apply', 'Europe: United Kingdom: England', 'Europe: United Kingdom: Northern Ireland', 'Europe: United Kingdom: Scotland', 'Europe: United Kingdom: Wales']

# Filter out the rows where 'Migrant country one year ago (60 categories) (detailed) label' matches any of the exclusions
census_data = census_data[~census_data['Migrant country one year ago (60 categories) (detailed) label'].isin(exclusions)]

# Create a pivot table with 'Migrant country one year ago' as rows,
# 'Lower tier local authorities label' as columns, and 'Count' as values
pivot_table = census_data.pivot_table(
    index='Migrant country one year ago (60 categories) (detailed) label',
    columns='Lower tier local authorities label',
    values='Count',
    aggfunc='sum',
    fill_value=0
)

# Write the pivot table to a new CSV file
pivot_table.to_csv('/content/ODMG01EW_NON_UK_LTLA_matrix_excl.csv')
```

Exercises with download data

If you want to repeat and slightly adjust the steps from Exercise 1 you can follow the instructions of the following 8 slides.

Exercises with download data

- While in cell B60, start typing **=LARGE(**
- Using your keyboard arrow keys go up to **B54** which is the last row with data. Press **CTRL+Shift+UP** simultaneously and Excel will select all rows up to the top.
- We don’t want the first row as it’s just the column header. While holding **Shift** key use your **down arrow** until the selection reaches **B2**.
- The formula should now be **=LARGE(B2:B54** . Large function accepts a final argument which allows to select the first largest, second largest etc. Type **,1)** and press **Enter**.
- The complete formula must be **=LARGE(B2:B54,1)** and the result in B60 should be **22.**

Exercises with download data

- We need to find an easy way to repeat this formula for all columns.
- Click on any column in row **54** (the last row containing data). While holding **CTRL** use the **right arrow key**. The active cell should now be the last one **LT54**.
- Move your cursor a bit further down to **LT60** which is the same row with the formula we created earlier. Type anything in, e.g. a number like 1. Press **CTRL+Left key** and the cursor should be move back to our formula cell **B60**.

Exercises with download data

- Use **CTRL+SHIFT+Right key** to select all columns in row 54 up to column LT.
- Press **F2** to enter Edit mode and Excel will automatically move the cursor back to B60 while keeping the cells selection active. Use **CTRL+Enter** to apply the formula. All columns in row 60 should now contain the largest number for each column (local authority).
- Repeat the process in B61 finding the second largest head count. The initial formula should be **=LARGE(B2:B54,2)**. Expand the formula to LT60 as before.

Exercises with download data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
23	Europe: Other Europe: EU countries: Member countries in March 2001: Ireland	6	8	4	10	4	8	2	13	89	6	6	13	29	2	23	32	15
24	Europe: Other Europe: EU countries: Member countries in March 2001: Italy	11	4	8	7	16	4	5	237	184	24	3	10	53	5	95	104	46
25	Europe: Other Europe: EU countries: Member countries in March 2001: Other member countries in March 2001	15	15	19	27	8	45	12	64	396	20	17	26	63	14	224	100	63
26	Europe: Other Europe: EU countries: Member countries in March 2001: Portugal (including Madeira and the Azores)	2	5	2	28	2	24	4	104	120	13	0	9	43	1	32	33	40
27	Europe: Other Europe: EU countries: Member countries in March 2001: Spain (including Canary Islands)	22	16	23	44	39	29	26	105	214	68	11	43	65	34	221	92	44
28	Europe: Other Europe: Rest of Europe: Other Europe	11	8	9	55	15	25	16	141	309	28	7	48	29	9	139	42	71
29	Europe: Other Europe: Rest of Europe: Turkey	1	12	6	9	3	9	5	22	138	7	12	13	19	11	28	18	25
30	Middle East and Asia: Central Asia	0	0	0	0	0	1	0	0	10	0	3	0	9	1	2	1	8
31	Middle East and Asia: Eastern Asia: China	6	3	6	6	3	6	10	15	131	6	4	10	17	7	158	13	23
32	Middle East and Asia: Eastern Asia: Hong Kong (Special Administrative Region of China)	11	2	4	7	4	36	5	17	414	4	3	12	48	7	157	28	32
33	Middle East and Asia: Eastern Asia: Other Eastern Asia	4	9	3	8	4	5	2	9	229	11	7	9	12	8	84	7	13
34	Middle East and Asia: Middle East: Iran	1	0	0	2	0	3	1	2	180	5	0	3	0	1	8	13	4
35	Middle East and Asia: Middle East: Iraq	2	1	0	4	1	0	0	0	5	4	0	2	1	1	6	1	1
36	Middle East and Asia: Middle East: Other Middle East	7	29	24	35	25	72	23	50	586	25	45	48	133	11	158	83	72
37	Middle East and Asia: South-East Asia: Malaysia	4	0	0	3	0	2	1	4	42	6	3	8	3	1	32	4	5
38	Middle East and Asia: South-East Asia: Other South-East Asia	6	7	6	28	13	23	14	9	94	18	7	19	25	10	71	35	38
39	Middle East and Asia: South-East Asia: Philippines	3	4	1	5	4	7	3	14	74	5	20	33	52	0	58	10	21
40	Middle East and Asia: South-East Asia: Singapore	0	8	0	4	0	9	3	2	30	3	0	3	8	0	70	3	17
41	Middle East and Asia: Southern Asia: Afghanistan	1	0	0	0	0	4	3	21	20	2	1	0	0	0	1	2	3
42	Middle East and Asia: Southern Asia: Bangladesh	1	1	2	4	0	3	0	193	28	2	4	8	2	0	1	29	10
43	Middle East and Asia: Southern Asia: India	5	3	3	9	18	52	6	257	372	22	21	108	263	7	153	188	108
44	Middle East and Asia: Southern Asia: Other Southern Asia	1	0	2	0	0	89	0	5	24	0	0	1	35	0	6	3	16
45	Middle East and Asia: Southern Asia: Pakistan	2	1	2	9	1	6	0	112	61	7	1	11	14	1	9	33	13
46	Middle East and Asia: Southern Asia: Sri Lanka	0	0	0	1	4	0	0	10	27	0	5	6	6	3	6	12	12
47	Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
48	The Americas and the Caribbean: Central America: All Central American countries	1	1	2	6	1	2	1	0	38	6	0	0	5	0	7	9	3
49	The Americas and the Caribbean: North America: Canada	4	10	14	8	2	17	11	11	80	11	4	10	22	2	46	18	17
50	The Americas and the Caribbean: North America: Other North America	0	0	1	2	0	0	0	0	1	0	0	0	0	0	4	2	0
51	The Americas and the Caribbean: North America: United States	11	15	31	41	15	33	32	28	274	36	10	25	80	16	155	62	42
52	The Americas and the Caribbean: South America: All South American countries	16	10	4	15	2	9	1	10	135	9	1	22	40	2	40	29	35
53	The Americas and the Caribbean: The Caribbean: Jamaica	0	0	0	0	0	1	0	4	7	0	0	7	2	0	0	7	6
54	The Americas and the Caribbean: The Caribbean: Other Caribbean	1	3	7	6	1	12	6	7	25	4	0	10	10	0	5	9	12
55																		
56																		
57																		
58																		
59																		
60		22	29	31	105	39	122	33	257	586	193	86	108	263	66	224	188	199
61		17	22	31	62	25	89	32	237	414	81	45	106	133	35	221	160	108
62																		

Exercises with download data

- We now need to match the numbers to the countries. For this purpose we will use a slightly more complex Excel formula.
- We will combine two Excel functions into one formula:
  - **INDEX**: returns a value based on the intersection of a row and column position within a given range
  - **MATCH**: returns the relative position of an item within a given range

Exercises with download data

- Go to cell B62 and type in the following formula:
  =INDEX($A$2:$A$54,MATCH(B60,B2:B54,0))

An array of values that contain the labels of the countries.
Notice the dollar sign. It instructs Excel to treat the values as fixed. It can become especially handy in situations when we want to apply auto-fill but want to keep this part of the formula fixed.

B60 is the look-up value which we want to match against the look-up array.
B2:B54 is the look-up array defined as a range of cells.
0 is the match type. Zero tells Excel to find the first value that is exactly equal to look-up value.

Exercises with download data

- Apply the auto-fill for formula in B62 in the exact same way as we did before to find the largest number.
- Go to cell B63 and repeat the process with B62 as our look-up value in order to find the second most popular countries of origin. **=INDEX($A$2:$A$54,MATCH(B61,B2:B54,0))**
- Apply the auto-fill

Exercises with download data

You should now be able to see the top two countries for each lower tier local authority.

	A	B	C	D	E	F
14	Antarctica and Oceania: Australasia: Other Australasia	0	0	0	0	0
15	Antarctica and Oceania: Other Oceania and Antarctica	0	0	0	0	0
16	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Lithuania	3	0	0	16	4
17	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Other EU countries	14	22	20	105	25
18	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Poland	17	5	15	62	12
19	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Romania	0	10	7	48	24
20	Europe: Other Europe: EU countries: Countries that joined the EU between April 2011 and March 2021: Croatia	0	0	0	0	0
21	Europe: Other Europe: EU countries: Member countries in March 2001: France	11	9	20	37	12
22	Europe: Other Europe: EU countries: Member countries in March 2001: Germany	2	5	24	19	11
23	Europe: Other Europe: EU countries: Member countries in March 2001: Ireland	6	8	4	10	4
24	Europe: Other Europe: EU countries: Member countries in March 2001: Italy	11	4	8	7	16
25	Europe: Other Europe: EU countries: Member countries in March 2001: Other member countries in March 2001	15	15	19	27	8
26	Europe: Other Europe: EU countries: Member countries in March 2001: Portugal (including Madeira and the Azores)	2	5	2	28	2
27	Europe: Other Europe: EU countries: Member countries in March 2001: Spain (including Canary Islands)	22	16	23	44	39
28	Europe: Other Europe: Rest of Europe: Other Europe	11	8	9	55	15
29	Europe: Other Europe: Rest of Europe: Turkey	1	12	6	9	3
30	Middle East and Asia: Central Asia	0	0	0	0	0
31	Middle East and Asia: Eastern Asia: China	6	3	6	6	3
32	Middle East and Asia: Eastern Asia: Hong Kong (Special Administrative Region of China)	11	2	4	7	4
33	Middle East and Asia: Eastern Asia: Other Eastern Asia	4	9	3	8	4
34	Middle East and Asia: Middle East: Iran	1	0	0	2	0
35	Middle East and Asia: Middle East: Iraq	2	1	0	4	1
36	Middle East and Asia: Middle East: Other Middle East	7	29	24	35	25
37	Middle East and Asia: South-East Asia: Malaysia	4	0	0	3	0
38	Middle East and Asia: South-East Asia: Other South-East Asia	6	7	6	28	13
39	Middle East and Asia: South-East Asia: Philippines	3	4	1	5	4
40	Middle East and Asia: South-East Asia: Singapore	0	8	0	4	0
41	Middle East and Asia: Southern Asia: Afghanistan	1	0	0	0	0
42	Middle East and Asia: Southern Asia: Bangladesh	1	1	2	4	0
43	Middle East and Asia: Southern Asia: India	5	3	3	9	18
44	Middle East and Asia: Southern Asia: Other Southern Asia	1	0	2	0	0
45	Middle East and Asia: Southern Asia: Pakistan	2	1	2	9	1
46	Middle East and Asia: Southern Asia: Sri Lanka	0	0	0	1	4
47	Other	0	0	0	1	0
48	The Americas and the Caribbean: Central America: All Central American countries	1	1	2	6	1
49	The Americas and the Caribbean: North America: Canada	4	10	14	8	2
50	The Americas and the Caribbean: North America: Other North America	0	0	1	2	0
51	The Americas and the Caribbean: North America: United States	11	15	31	41	15
52	The Americas and the Caribbean: South America: All South American countries	16	10	4	15	2
53	The Americas and the Caribbean: The Caribbean: Jamaica	0	0	0	0	0
54	The Americas and the Caribbean: The Caribbean: Other Caribbean	1	3	7	6	1
55						
56						
57						
58						
59						
60	Largest	22	29	31	105	39
61	Second largest	17	22	31	62	25
62	Largest	Europe: Other Europe: EU countries: Member countries in March 2001: Spain (including Canary Islands)	Middle East and Asia: Middle East: Other Middle East	Antarctica and Oceania: Australasia: Australia	Europe: O	Europe: O
63	Second largest	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Poland	Europe: Other Europe: EU countries: Countries that joined the EU be	Antarctica and Oceania: Australasia: Australia	Europe: O	Europe: O

But isn't there a more efficient way to do this?

Again, there are more efficient methods!

Upload **exercise2_3_top_countries.ipynb** into Google Colab
via the File->Upload a notebook option

```
import pandas as pd

# Load the CSV data into a DataFrame
census_data = pd.read_csv('/content/ODMG01EW_NON_UK_LTLA.csv')

# Define the countries we do not want to include
exclusions = ['Does not apply', 'Europe: United Kingdom: England', 'Europe: United Kingdom: Northern Ireland', 'Europe: United Kingdom: Scotland', 'Europe: United Kingdom: Wales']

# Filter out the rows where 'Migrant country one year ago (60 categories) (detailed) label' matches any of the exclusions
census_data = census_data[~census_data['Migrant country one year ago (60 categories) (detailed) label'].isin(exclusions)]

# Function to find the top N countries for each local authority
def top_n_countries(group, n=2):
    return group.nlargest(n, 'Count')[['Migrant country one year ago (60 categories) (detailed) label', 'Count']]

# Group by 'Lower tier local authorities label' and apply the function
top_countries = census_data.groupby('Lower tier local authorities label', group_keys=True).apply(top_n_countries)

# Reset index to clean up the DataFrame
top_countries = top_countries.reset_index(level=-1, drop=True).reset_index()

# Write the pivot table to a new CSV file
top_countries.to_csv('/content/ODMG01EW_NON_UK_LTLA_top_countries.csv')
```

Different methods. Same results.

```
display(top_countries)
```

1 to 25 of 662 entries Filter

index	Lower tier local authorities label	Migrant country one year ago (60 categories) (detailed) label	Count
0	Adur	Europe: Other Europe: EU countries: Member countries in March 2001: Spain (including Canary Islands)	22
1	Adur	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Poland	17
2	Allerdale	Middle East and Asia: Middle East: Other Middle East	29
3	Allerdale	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Other EU countries	22
4	Amber Valley	The Americas and the Caribbean: North America: United States	31
5	Amber Valley	Antarctica and Oceania: Australasia: Australia	31
6	Arun	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Other EU countries	105
7	Arun	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Poland	62
8	Ashfield	Europe: Other Europe: EU countries: Member countries in March 2001: Spain (including Canary Islands)	39
9	Ashfield	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Other EU countries	25
10	Ashford	Africa: Central and Western Africa: Nigeria	122
11	Ashford	Middle East and Asia: Southern Asia: Other Southern Asia	89
12	Babergh	Europe: Other Europe: EU countries: Member countries in March 2001: France	33
13	Babergh	The Americas and the Caribbean: North America: United States	32
14	Barking and Dagenham	Middle East and Asia: Southern Asia: India	257
15	Barking and Dagenham	Europe: Other Europe: EU countries: Member countries in March 2001: Italy	237
16	Barnet	Middle East and Asia: Middle East: Other Middle East	586
17	Barnet	Middle East and Asia: Eastern Asia: Hong Kong (Special Administrative Region of China)	414
18	Barnsley	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Romania	193
19	Barnsley	Europe: Other Europe: EU countries: Countries that joined the EU between April 2001 and March 2011: Poland	81
20	Barrow-in-Furness	Africa: Central and Western Africa: Nigeria	86
21	Barrow-in-Furness	Middle East and Asia: Middle East: Other Middle East	45
22	Basildon	Middle East and Asia: Southern Asia: India	108
23	Basildon	Africa: Central and Western Africa: Nigeria	106
24	Basingstoke and Deane	Middle East and Asia: Southern Asia: India	263

Show 25 per page

1 2 10 20 27