
ASSIGNMENT: DATA ANALYSIS AND TRANSFORMATION WITH ALTERYX

STUDENT: ANASTASIA HASWANI

COURSE: CEBD 1150

INSTRUCTOR: ADRIÀN GONSÀLEZ SÀNCHEZ

PART 1

ANALYZING AND TRANSFORMING DATA WITH ALTERYX

CONTENTS

Load and prepare the data	3
What are the first 10 countries that received the most refugees in 2015.	8
Provide the top 10 countries which are the highest “generators” of refugees in 2010.....	11
Number of people that came to Canada in 2018	12
Number of people for each population type for Canada in 2018.....	13
Top10 countries that generated newcomers for Canada in 2018	14
Total number of people received yearly.....	14
Countries that provided the highest number of people for Canada each year.....	15
Yearly distribution of the number of people from each country-Origin	17
Syrian Refugee Crisis.....	19
Top10 countries that hosted the most refugees from Syria in 2015.....	19
A comparison between the number of received refugees and total population	20

LOAD AND PREPARE THE DATA

1. The data was loaded as .csv file following given instructions. The data is taken from the UN Refugee Agency¹.
2. In Alteryx we used the **Input Data** tool and loaded the file. We see that there are few lines in the file which are not part of the dataset.

Input Data (9) - Configuration

Connect a File or Database

C:\Users\nasta\BIG_DATA\CEBD-1150\Alteryx Assignment\unhcr_popstats_export_time_series_all_data.csv

Options

Name	Value
1 Record Limit	
2 File Format	Comma Separated Value (*.csv)
3 Search SubDirs	<input type="checkbox"/>
4 Output File Name as Field	No
5 Delimiters	.
6 First Row Contains Field Names	<input checked="" type="checkbox"/>

Preview (first 100 records)

Record	Date extracted: 2018-12-13 04:37:24 +02:00	Year	Country / territory of asylum/residence	Origin
1	[Null]	1951	Australia	Various/Unknown
2	[Null]	1951	Austria	Various/Unknown
3	[Null]	1951	Belgium	Various/Unknown
4	[Null]	1951	Canada	Various/Unknown
5	[Null]	1951	China, Hong Kong SAR	Various/Unknown
6	[Null]	1951	Denmark	Various/Unknown

To resolve this issue, we choose the option **Start Data Import On Line** in Configuration window, and set it to 4. The **First Row Contains Field Names** option is checked already. Each time we make changes we press Run button and then we see the results using **Browse** tool:

Browse (10) - Configuration

Profile

192,227 records displayed, 6 fields, 3.9 MB

A Year

Showing a sample of your whole dataset.

A Country / territory of asylum/residence

Showing a sample of your whole dataset.

A Origin

Showing a sample of your whole dataset.

Results - Browse (10) - Input

Record	Year	Country / territory of asylum/residence	Origin
1	1951	Australia	Various/Unknown
2	1951	Austria	Various/Unknown
3	1951	Belgium	Various/Unknown

¹ http://popstats.unhcr.org/en/time_series

In Configuration window of the **Browse** tool we can see all the basic information about each field.

3. Using **Browse** tool, we saw that *Year* and *Value* fields are String type.

Summary

Type	Records	Data Type Size
V_String	192227	254

Length Statistics

Min	4
Max	4
Average	4.0
Shortest Value	1951
Longest Value	1951
First Alphanumeric Value	1951

Results - Browse (6) - Input

Record	Year	Country / territory of asylum/residence	Origin	Population typ
1	1951	Australia	Various/Unknown	Refugees (incl. i
2	1951	Austria	Various/Unknown	Refugees (incl. i
3	1951	Belgium	Various/Unknown	Refugees (incl. i
4	1951	Canada	Various/Unknown	Refugees (incl. i
5	1951	China, Hong Kong SAR	Various/Unknown	Refugees (incl. i
6	1951	Denmark	Various/Unknown	Refugees (incl. i
7	1951	France	Various/Unknown	Refugees (incl. i
8	1951	Germany	Various/Unknown	Refugees (incl. i
9	1951	Greece	Various/Unknown	Refugees (incl. i
10	1951	Italy	Various/Unknown	Refugees (incl. i

4. To automatically transform types, we used **Auto Field** tool. The results in **Browse** tool showed that *Year* has been successfully transformed. But the *Value* field was stayed String.

Summary

Type	Records	Data Type Size
Int16	192227	2

Value Statistics

Value by count

Summary

Type	Records	Data Type Size
V_String	192227	7

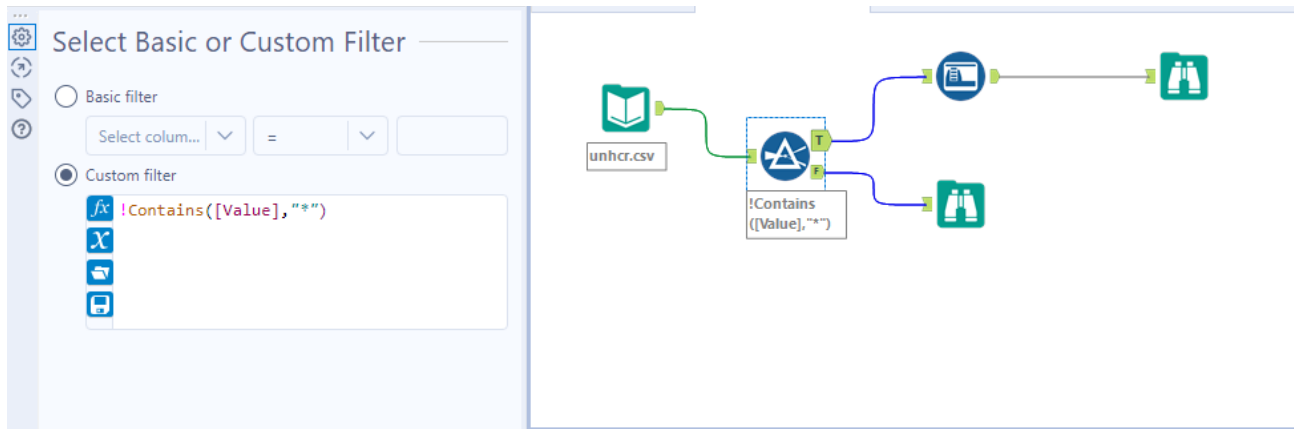
Length Statistics

Min	1
Max	7
Average	1.9
Shortest Value	2
Longest Value	1120000
First Alphanumeric Value	*
Last Alphanumeric Value	99983
Blanks	0
Values with Leading Whitespace	0
Values with Trailing Whitespace	0

For the *Year* we see that the records are made from 1951 till 2018.

For the *Value* field we see that the first alphanumeric value is *. That is not a number. Probably that's why the **Auto Field** tool couldn't transform the type to numeric.

5. Lets try to filter out the records containing this symbol *. We will use the **Filter** tool.

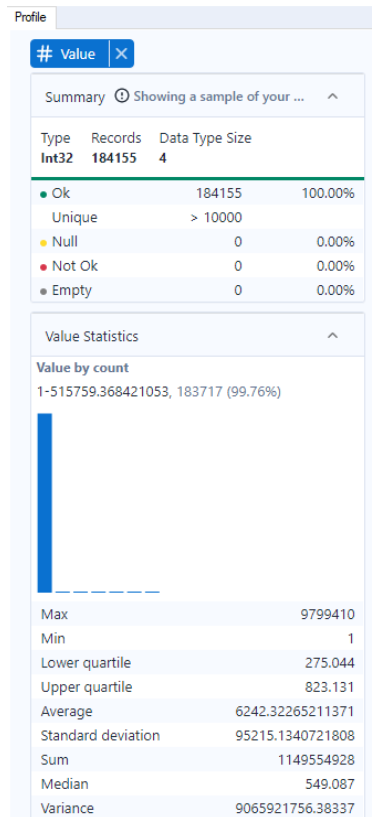
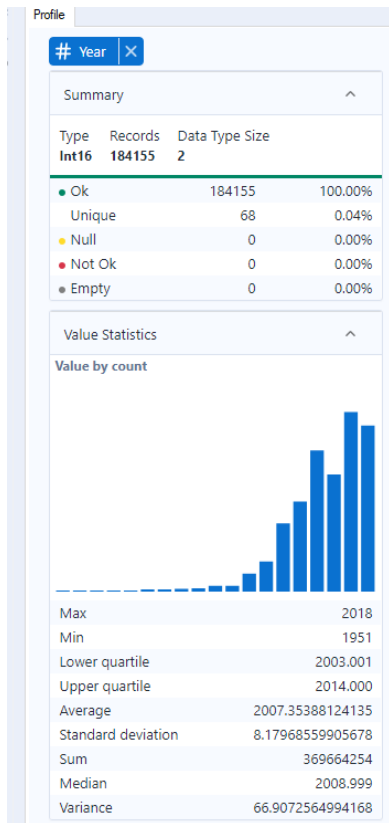


Using the **Browse** tool, we can see that there were around 8000 records with * instead of the number. It doesn't make sense to replace * by 0 so we will not use them.

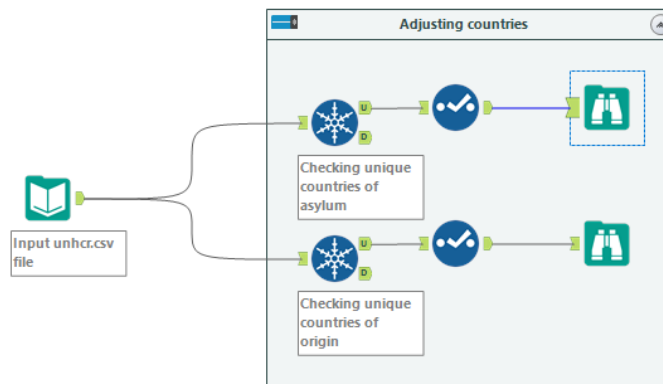
The screenshot shows the Alteryx interface for the Browse tool. On the left, the 'Browse (12) - Configuration' panel shows the 'Profile' tab with 8,072 records displayed. On the right, the 'Results - Browse (12) - Input' tab shows a table of 8,072 records with 5 fields: Record, Year, Country / territory of asylum/residence, Origin, and Population type. The Value field is highlighted in the first row.

Record	Year	Country / territory of asylum/residence	Origin	Population type	Value
1	2017	Afghanistan	Iraq	Refugees (incl. refugee-like situations)	*
2	2017	Afghanistan	Uzbekistan	Asylum-seekers	*
3	2017	Albania	Afghanistan	Asylum-seekers	*
4	2017	Albania	Algeria	Asylum-seekers	*
5	2017	Albania	Dem. Rep. of the Congo	Refugees (incl. refugee-like situations)	*
6	2017	Albania	Egypt	Refugees (incl. refugee-like situations)	*
7	2017	Albania	Iran	Defensor (incl. refugee-like situations)	*

After applying Auto Field tool to the filtered data, the Value field was transformed to Numeric type. Value type is *Int32* and Year type is *Int16* as its values don't get higher than 4-digit numbers.



6. Let us get closer look at the countries. Using **Unique** and **Select** tool we will select *unique countries of asylum* and *countries of origin* and check the results:

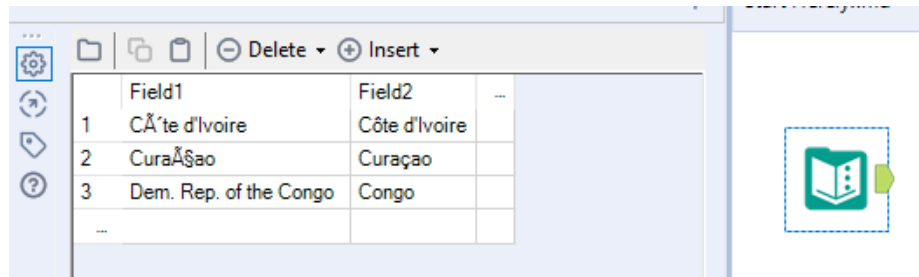


Results - Browse (11) - Input

1 of 1 Fields | Cell Viewer | 203 records displayed, 3618 bytes

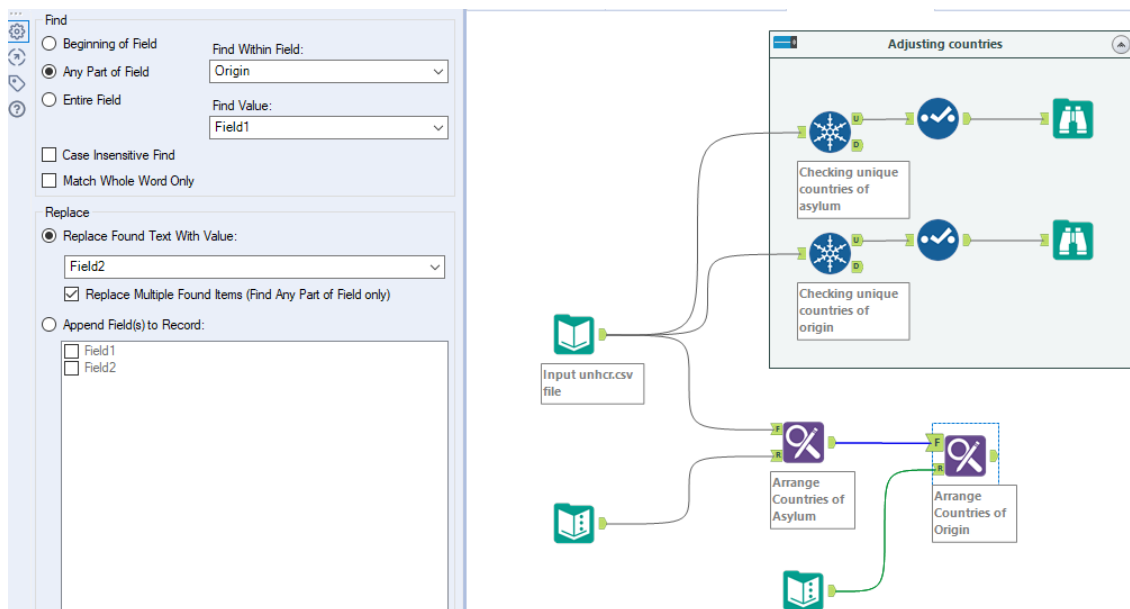
Record	Country / territory of asylum/residence
44	Comoros
45	Congo
46	Costa Rica
47	Croatia
48	Cuba
49	Curaçao
50	Curaçao
51	Cyprus
52	Czech Rep.
53	Côte d'Ivoire
54	Côte d'Ivoire
55	Dem. People's Rep. of Korea
56	Dem. Rep. of the Congo
57	Denmark
58	Djibouti

We found three countries that have two versions of its names. Let's create a small table with the names of those countries and their duplicate names, using the **Text Input** tool:

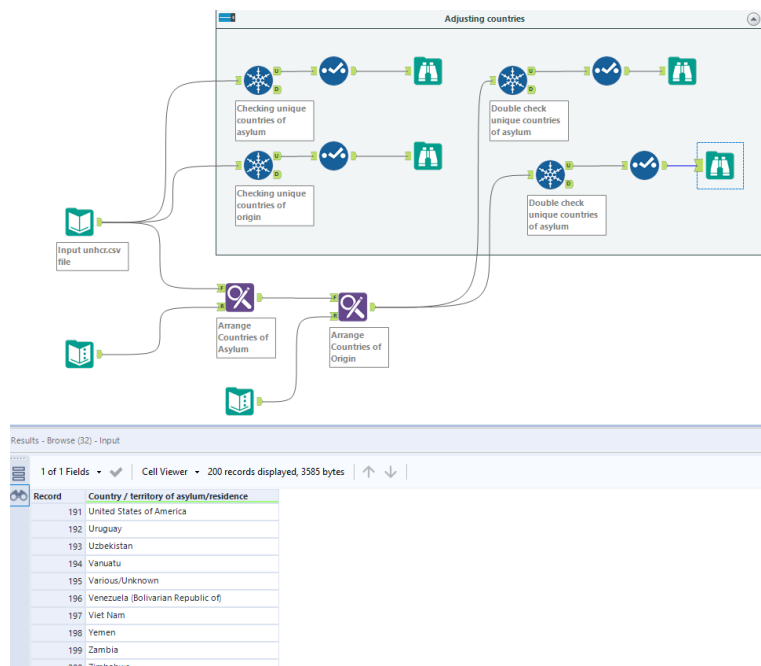


	Field1	Field2	...
1	Côte d'Ivoire	Côte d'Ivoire	
2	Curaçao	Curaçao	
3	Dem. Rep. of the Congo	Congo	
...			

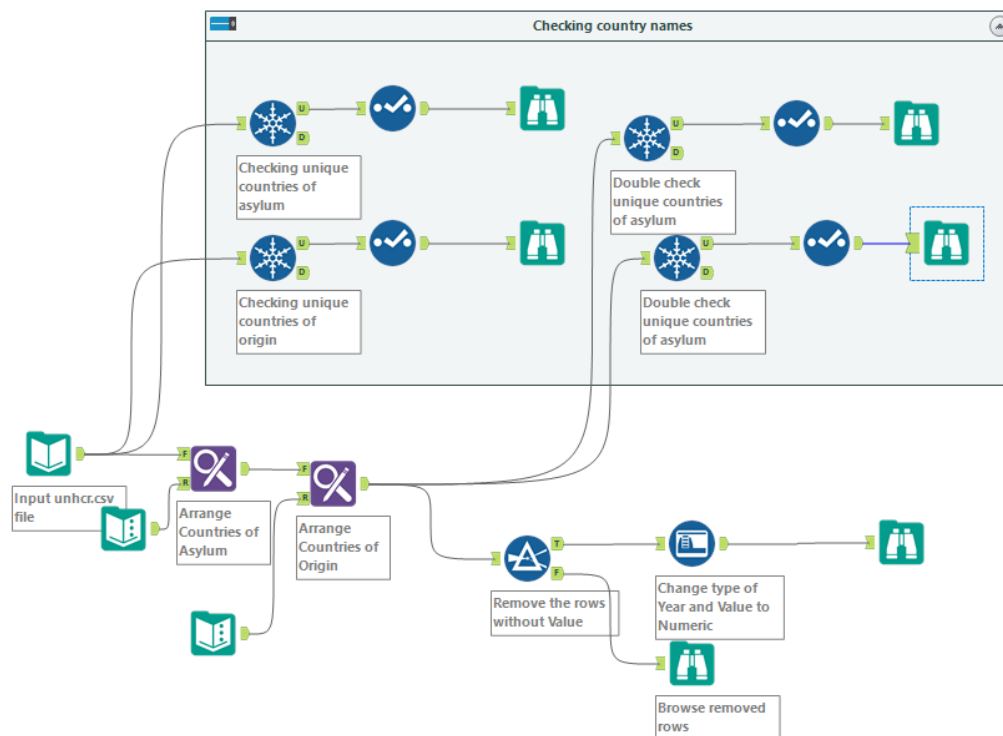
7. Now using **Find Replace** tool we will replace values from the left field of our Text Input table by right fields for *Country of Asylum*. We will repeat it for *Country of origin* too.



Next, we will repeat step 6 for the output and check the results. Before we had 203 distinct countries. Now we have 200.



8. We started our data transformations by removing the records with the empty *Values*. But if we kept this step at the beginning it would be hard to compare the results of replacing the name of the countries (there was a country which disappears from the dataset after removing the records containing *). That is why we shifted them. Nevertheless, the order of these two steps does not change the output.



We will consider now that the transformation part is completed. Next step is to find the answer to this question:

WHAT ARE THE FIRST 10 COUNTRIES THAT RECEIVED THE MOST REFUGEES IN 2015.

1. Using the Filter tool we will keep the records from 2015.

Select Basic or Custom Filter

☒ Basic filter

Year = 2015

☐ Custom filter

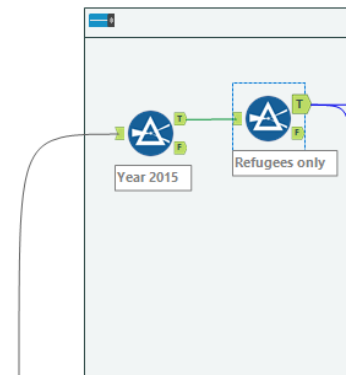
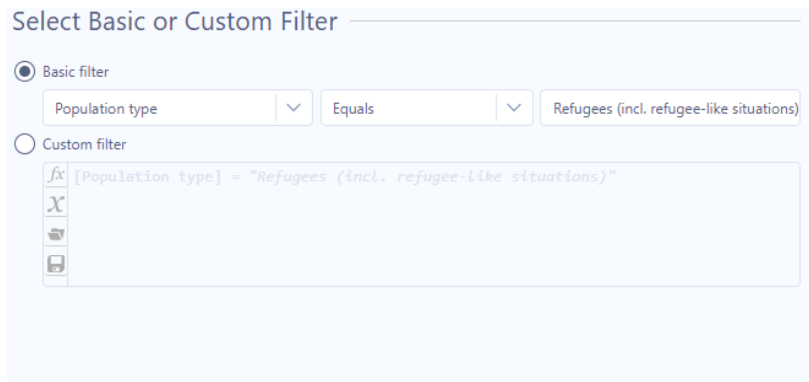
$[Year] = 2015$

Results - Filter (15) - Out - True

5 of 5 Fields | Cell Viewer | 11,051 records displayed

Record	Year	Country / territory of asylum/residence	Origin	Population type
1	2015	Afghanistan	Afghanistan	Internally displaced persons
2	2015	Afghanistan	Afghanistan	Returned IDPs
3	2015	Afghanistan	Afghanistan	Others of concern
4	2015	Afghanistan	Iran (Islamic Rep. of)	Refugees (incl. refugee-like situation)
5	2015	Afghanistan	Iran (Islamic Rep. of)	Asylum-seekers
6	2015	Afghanistan	Iraq	Refugees (incl. refugee-like situation)

2. Next step is to filter the data by keeping refugees only. We will use the same **Filter** tool.



3. Now we want to group our data by *Country of asylum* field and calculate the summary of *Value* for each country. For this we will use **Summarize** tool. In the lower part of configuration window, we click on *Add* and choose *Group By* option for *Country of asylum* field and *Sum* option for the *Value* field.

Record	Country / territory of asylum/residence	Sum_Value
1	Alghanistan	257554
2	Albania	105
3	Algeria	95041
4	Angola	15557
5	Anguilla	1
6	Antigua and Barbuda	15
7	Argentina	3216

In the *Results* window we see that now we have a table of countries with the total number of refugees as *Sum_Value* field.

4. Next step is to sort the data by Value. We will use **Sort** tool and choose *Sum_Value* field to be sorted in *descending* order.

The screenshot displays an Alteryx Designer workflow. On the left, the 'Fields' pane shows 'Sum_Value' selected with a descending sort order. The main workflow area contains several tools: 'Input unhcr.csv file', 'Arrange Countries of Asylum', 'Arrange Countries of Origin', 'Remove the rows without Value', 'Change type of Year and Value to Numeric', and 'Browse removed rows'. A tool container titled 'Top 10 countries which received the refugees in 2015' contains tools for filtering by 'Year 2015', 'Refugees only', calculating the 'Total number of refugees for each country', and sorting by 'number of refugees'. The 'Results - Sort (18) - Output' pane shows a table with 173 records displayed.

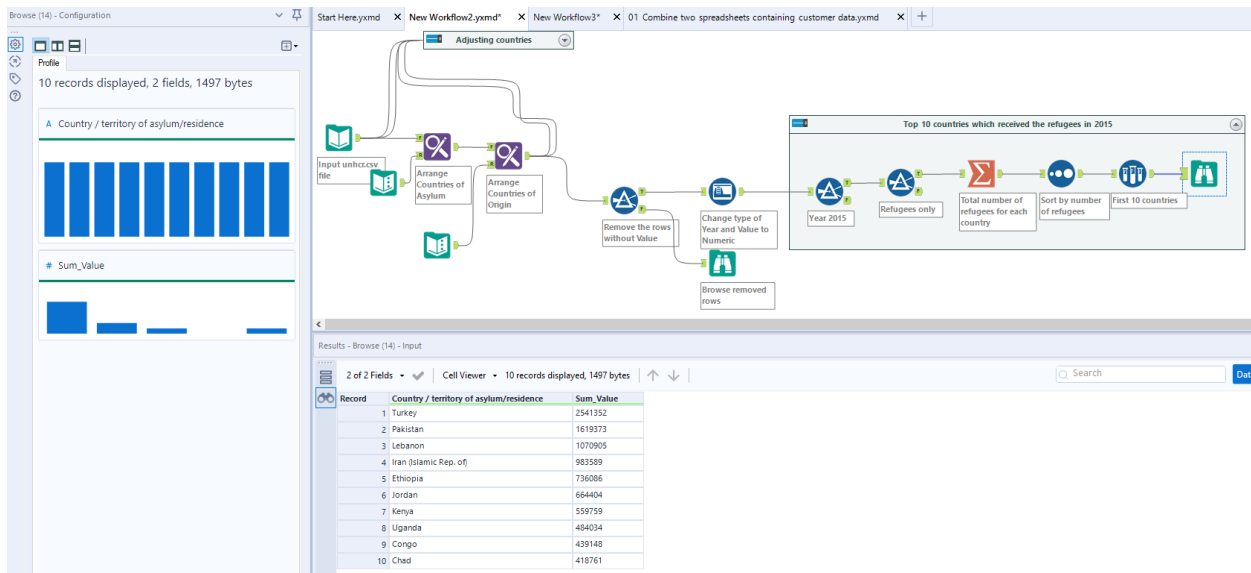
Record	Country / territory of asylum/residence	Sum_Value
1	Turkey	2541352
2	Pakistan	1619373
3	Lebanon	1070905
4	Iran (Islamic Rep. of)	983589
5	Ethiopia	736086
6	Jordan	664404
7	Kenya	559759
8	Uganda	484034

5. To get the first 10 countries we use **Sample** tool and choose *First N rows* option with $N=10$.

This screenshot shows the same Alteryx workflow, but with an additional 'Sample' tool added to the 'Top 10 countries which received the refugees in 2015' container. The 'Select Sample Type' pane on the left is visible, showing 'First N rows' selected with $N = 10$. The 'Results - Sample (21) - Output' pane now displays only the first 10 records of the sorted data.

Record	Country / territory of asylum/residence	Sum_Value
1	Turkey	2541352
2	Pakistan	1619373
3	Lebanon	1070905
4	Iran (Islamic Rep. of)	983589
5	Ethiopia	736086
6	Jordan	664404
7	Kenya	559759
8	Uganda	484034
9	Congo	439148
10	Chad	418761

6. The last step is to view the results with the **Browse** tool. For better organize the tools we created 2 **Tool Containers**. We named them *Checking country names* and *Top 10 countries which received refugees in 2015* and placed corresponding groups of tools inside each one of them.



Below are the results:

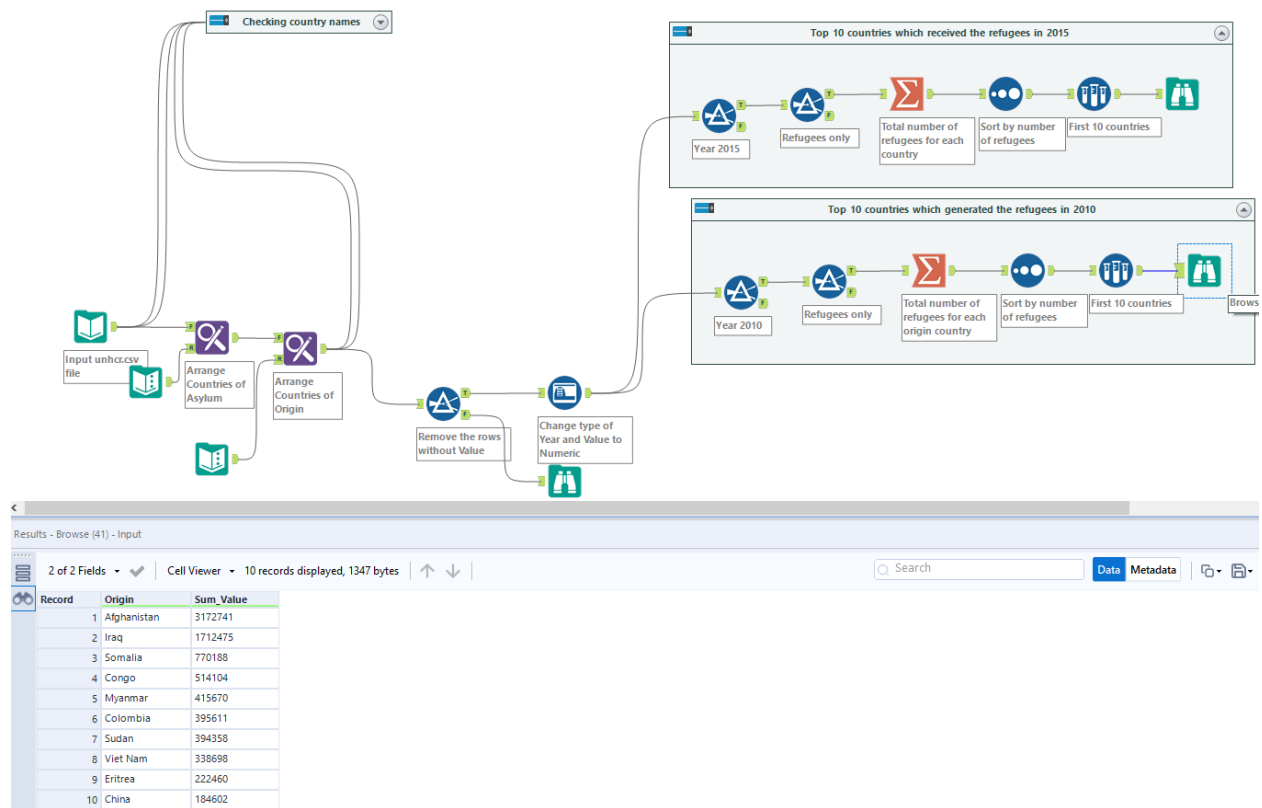
Country / territory of asylum/residence	Sum_Value
Turkey	2541352
Pakistan	1561162
Lebanon	1070854
Iran (Islamic Rep. of)	979437
Ethiopia	736086
Jordan	664118
Kenya	553912
Uganda	477187
Congo	428050
Chad	369540

PROVIDE THE TOP 10 COUNTRIES WHICH ARE THE HIGHEST “GENERATORS” OF REFUGEES IN 2010

For this task we click on the Tool Container *Top 10 countries which received refugees in 2015*, copy and paste it. This sequence will not differ much from the top10 that we calculated previously.

We will do changes in three places:

1. Inside **Filter** tool we replace the year 2015 with 2010.
2. Inside **Summarize** tool we replace *Country of asylum* with *Origin*.
3. We will change the name of the **Tool Container** with *Top 10 countries which generated refugees in 2010*.



Below are the results:

Origin	Sum_Value
Afghanistan	3054709
Iraq	1683579
Somalia	770154
Congo	497372
Myanmar	415670
Colombia	395577
Sudan	387288
Viet Nam	338698
Eritrea	222460
China	184602

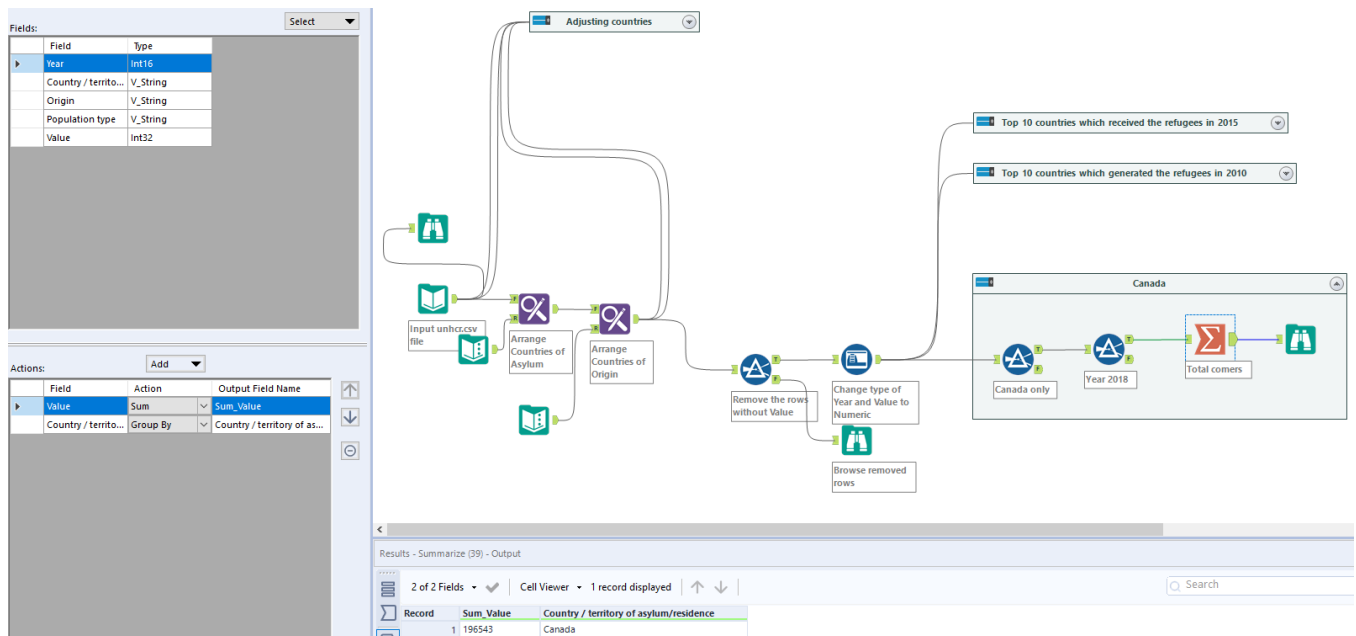
NUMBER OF PEOPLE THAT CAME TO CANADA IN 2018

To find out we will create another Tool Container and call it *Canada*. Then we will follow these steps:

- Use **Filter** tool and set it this way:
Contains([Country / territory of asylum/residence],"Canada")
AND
NOT
Contains([Population type],"Returned refugees")
AND
NOT

Contains([Population type], "Stateless persons")

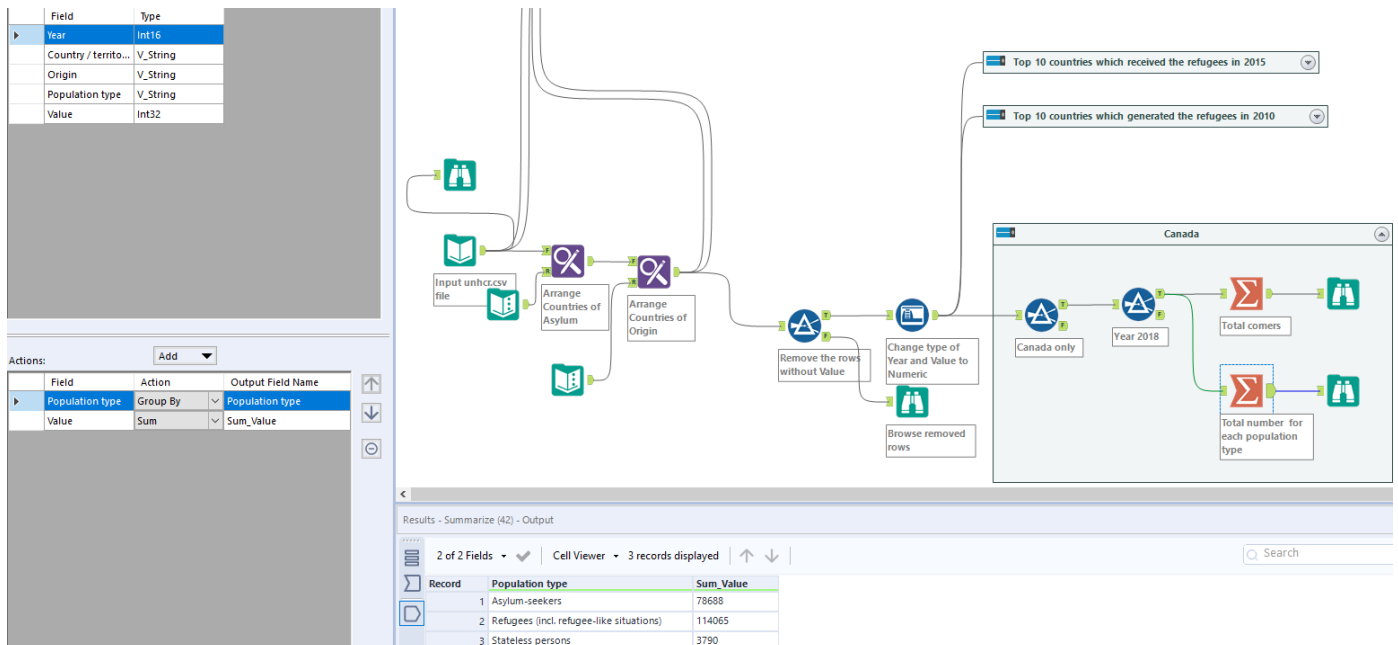
2. Use another **Filter** tool to set the *Year* to 2018 like we did before.
3. Use **Summarize** tool and set it this way: *Group By Country of asylum*, and *Sum of Value*.



Result obtained: in 2018 Canada received 196,543 persons.

NUMBER OF PEOPLE FOR EACH POPULATION TYPE FOR CANADA IN 2018

1. Use **Filter** tool and set it [this way](#):
2. Use another **Filter** tool to set the *Year* to 2018 like we did before (from last task too).
3. Use **Summarize** tool and set it this way: *Group By Population type*, and *Sum of Value*.

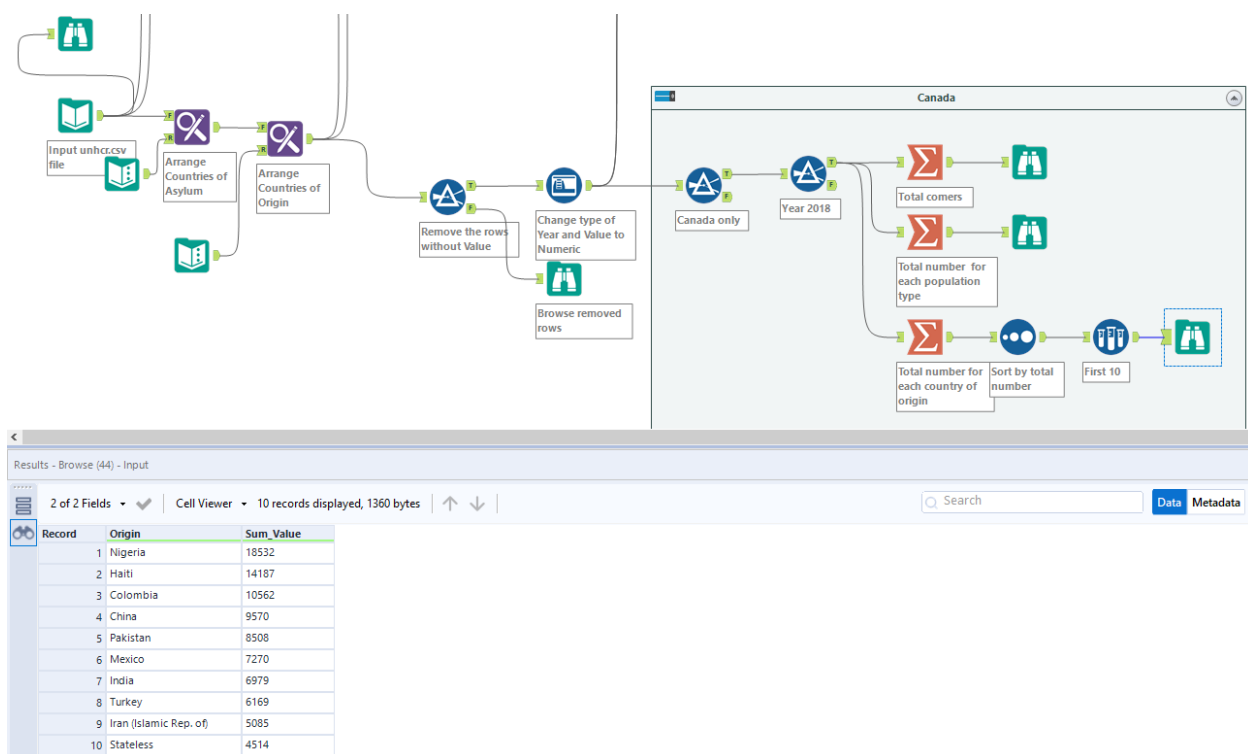


Results obtained:

Population type	Sum_Value
Asylum-seekers	78688
Refugees (incl. refugee-like situations)	114065
Stateless persons	3790

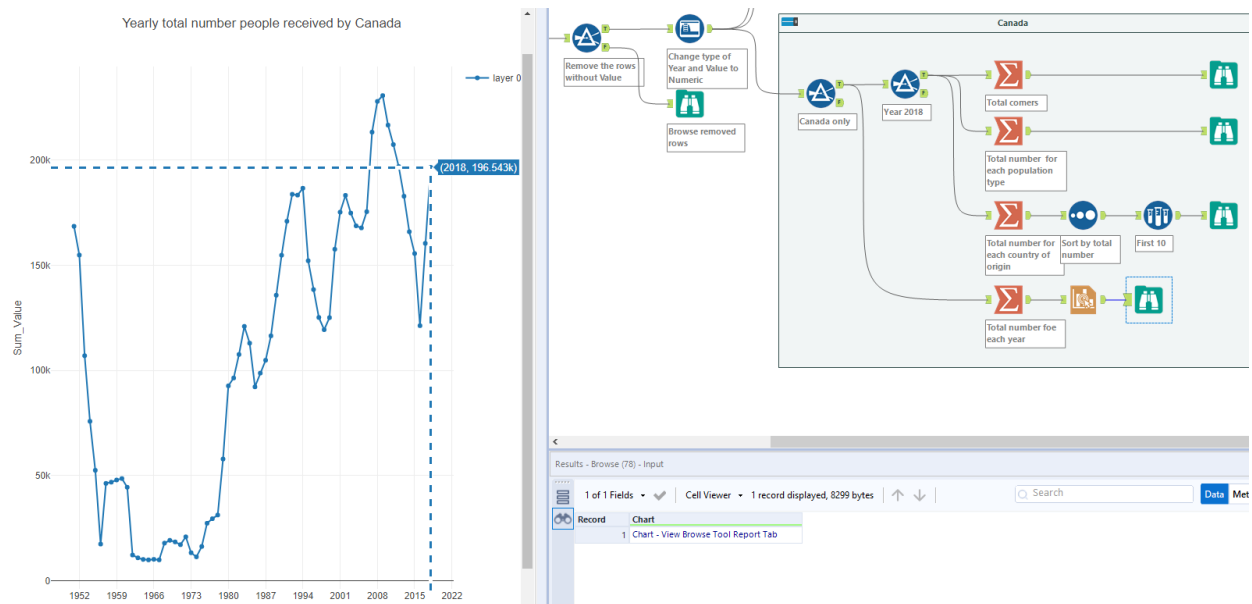
TOP10 COUNTRIES THAT GENERATED NEWCOMERS FOR CANADA IN 2018

1. Repeat first two steps from previous task.
2. In **Summarize** set *Group By* to *Origin*, and *Sum of Value*.
3. Use **Sort** tool and sort by *Sum_Value* in *descending* order.
4. Using **Sample** tool, get first 10 rows for the dataset.



TOTAL NUMBER OF PEOPLE RECEIVED YEARLY

1. Use **Filter** tool and set it [this way](#).
2. In **Summarize** set *Group By* to *Year*, and *Sum of Value*.
3. Use **Interactive Chart** tool and configure it this way:
 - a. In *Layer* section set *Type* to *Line*,
 - b. Set *X axis* to *Year*
 - c. Set *Y axis* to *Sum_Value*.
4. Use **Browse** tool to see the results (choose *Report* tab in *configuration window*).



Results: This graph is easy to read and gives us a lot of information:

- The highest number of people was received in 2009.
- In 2018 Canada received 196,543 persons. We can now compare this number to the one we got previously and see that they are equal.
- This irregular graph shows a frequency of approximately 10 years. After the 80s, it becomes more obvious.

COUNTRIES THAT PROVIDED THE HIGHEST NUMBER OF PEOPLE FOR CANADA EACH YEAR

1. Use **Filter** tool and set it [this way](#).
 2. In **Summarize** set *Group By Year*, *Group By Origin*, and *Sum of Value*.
- Here is except of our resulting data:

Year	Origin	Sum_Value
1993	Chile	65
1993	Iran (Islamic Rep. of)	1
1993	South Africa	9
1993	Various/Unknown	183241

3. Now we will use another **Summarize** tool in order to get the Max value of each year. We will have to *Group by Year* and *Max of Sum_Value*. The result will be only one row for each year but without Origin Country:

Year	Max_Sum_Value
1993	183241

4. To have origin country name for each max value we will join the results of both **Summarize** tools using **Join** tool on two fields: *Year from both tables* and *Sum_Value with Max_Sum_Value*. As output, we need *Year, Origin* and *Max_Sum*.

The screenshot shows the Alteryx Join tool configuration. The 'Join by Specific Fields' tab is selected. The left table has fields 'Year' and 'Sum_Value'. The right table has fields 'Year' and 'Max_Sum_Value'. The join is configured to join on 'Year' from both tables and 'Sum_Value' from the left table with 'Max_Sum_Value' from the right table. Below the configuration is a table showing the input fields and their properties.

Input	Field	Type	Size	Rename	Description
<input checked="" type="checkbox"/> Left	Year	int16	2		
<input checked="" type="checkbox"/> Left	Origin	V_String	254		
<input checked="" type="checkbox"/> Left	Sum_Value	int64	8		
<input type="checkbox"/> Right	Year	int16	2	Right_Year	
<input checked="" type="checkbox"/> Right	Max_Sum_...	int64	8	Right_Max_Sum_Value	
<input type="checkbox"/>	*Unknown	Unknown	0		Dynamic or Unknown Fields

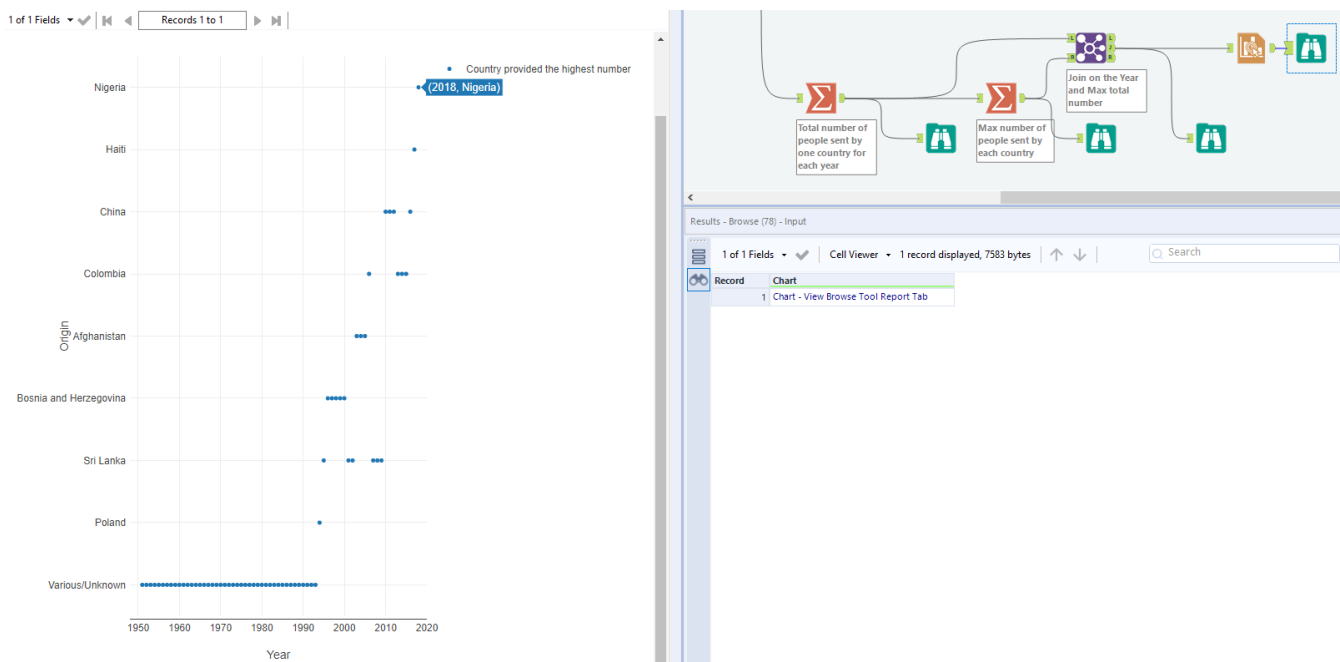
The workflow diagram on the right shows the process: a Summarize tool calculates the total number of people sent by one country for each year. Another Summarize tool calculates the maximum number of people sent by each country. These two results are joined on the Year and Max total number fields. The final output is a table with columns Year, Origin, and Max_Sum.

The except of output for 1993:

Year	Origin	Right_Max_Sum_Value
1993	Various/Unknown	183241

We finally have what we want: Countries that provided the highest number of people for Canada each year.

5. Using **Interactive Chart**, we will visualize it. In *Layer* section:
 - a. *Type: Scatter*
 - b. *X: Year*
 - c. *Y: Origin*



Below are some of the conclusions drawn from the graph:

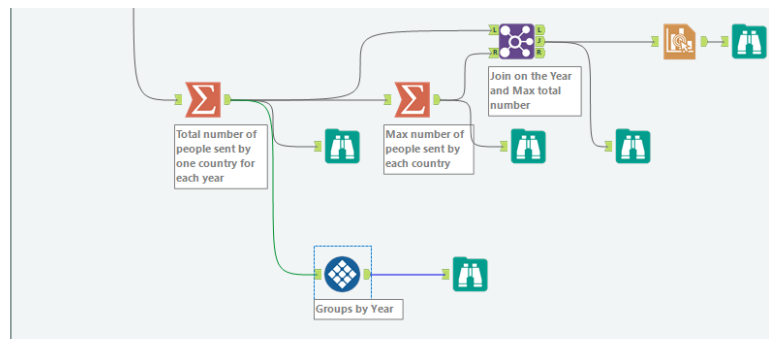
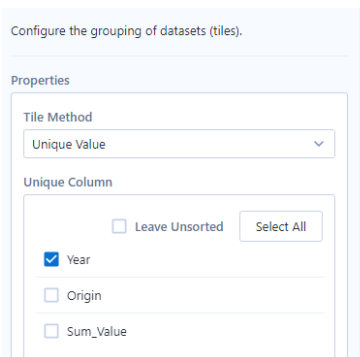
- We can verify one of the previous tasks (getting the country that provided the most comers in 2018). The results match: in 2018 it was Nigeria.
- We see that for the four consecutive years (1996-2000) the leader was Bosnia and Herzegovina.
- Sri Lanka has become the leader the most times: in 1995, 2001, 2002, 2007, 2008 and 2009.

We see that from 1951 to 1992 the leader is *Various/Unknown*. We can assume that these during these years the agency wasn't specifying the origin countries for many or all the people.

To have a better look we will build another chart.

YEARLY DISTRIBUTION OF THE NUMBER OF PEOPLE FROM EACH COUNTRY-ORIGIN

1. Use **Filter** tool and set it [this way](#).
2. In **Summarize** set *Group By Year*, *Group By Origin*, and *Sum of Value*.
3. Use **Tile** tool to break all our records into groups by *Year*.

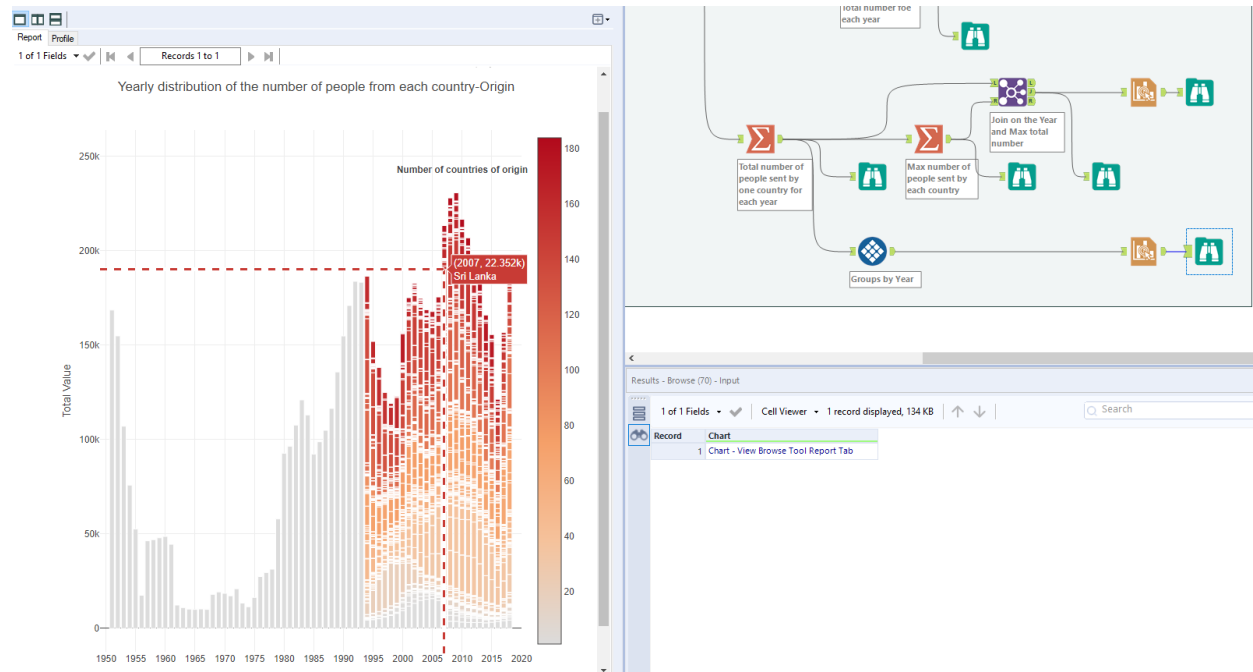


Here we see that from 1951 to 1992 the only value in *Origin* field is *Various/Unknown* (we saw it before when we summarized by *Origin*). But here we have two new fields: *Tile_num* and *Tile_sequence_num* that gives sequence number to each record in one groupe. Below is the exept of the output:

Year	Origin	Sum_Value	Tile_Num	Tile_SequenceNum
1994	Uzbekistan	26	44	157
1994	Various/Unknown	3255	44	158
1994	Venezuela (Bolivarian Republic of)	363	44	159
1994	Viet Nam	14007	44	160
1994	Yemen	79	44	161
1994	Zambia	15	44	162
1994	Zimbabwe	25	44	163
1995	Afghanistan	4575	45	1
1995	Albania	239	45	2
1995	Algeria	905	45	3
1995	Andorra	1	45	4

We see that the group of the year 1994 has tile number 44 and consists of 163 Origin countries.

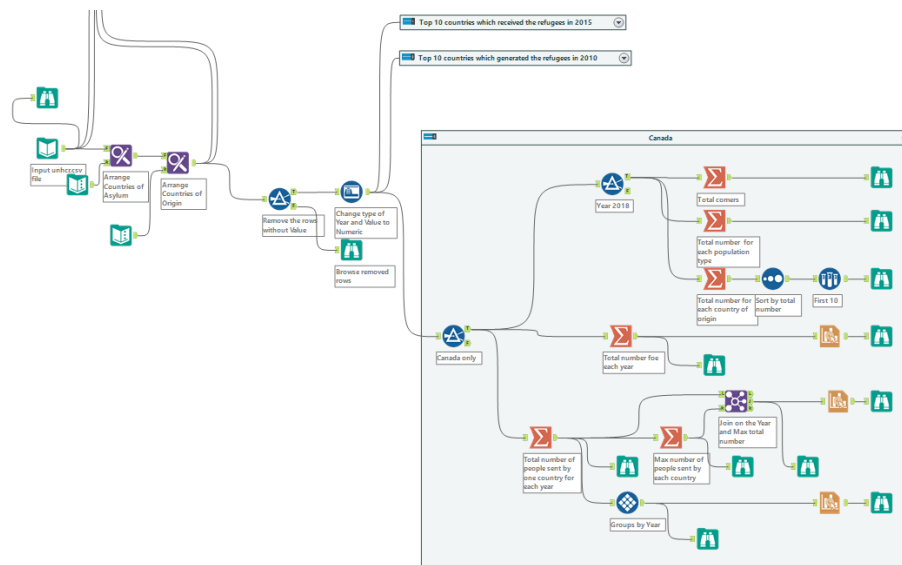
4. Next, we will build a chart using **Interactive Chart** tool and set it up this way:
 - a. In the *Layer* section: *Type to Bar*; *X to Year*; *Y to Sum_Value*
 - b. In the *Style Layer* section: *Each point in a trace is colored according to data to Tile_SequenceNum*
 - c. Adjust some other details according to our liking.



Now we can visually verify all our assumptions and get a lot of information all from one chart:

- The specification of the countries of origin started in 1994.
- 2009 was the year when Canada received the highest number of people.
- The mid-sixties were the years when Canada received the least number of people.
- Every time we point to certain bar section, we will see the country of origin and the total number of people it sent to Canada

Now our “Canada Tool Container” looks like this:



SYRIAN REFUGEE CRISIS.

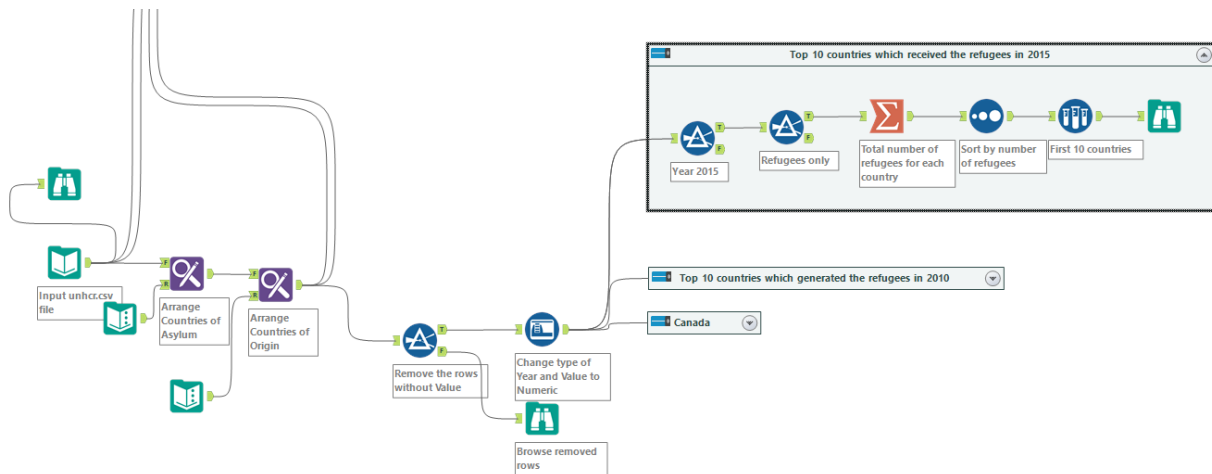
Total population of Syria (<https://www.google.com/publicdata/>):

2010: 21,362,500

2015: 17,996,400

TOP10 COUNTRIES THAT HOSTED THE MOST REFUGEES FROM SYRIA IN 2015

1. Let us go back to our 2015-year tool container:



2. We need *Year 2015* filter and *Refugees only* filter too. To its output we will attach a new **Summarize tool** and set *Group By Country of asylum*, *Group By Origin*, and *Sum of Value*.
3. Then with the help of the **Filter tool** we will keep records where country of origin is Syria.
4. Using **Sort tool**, we will sort our data by *Sum_Value* in *descending* order.
5. With the **Sample tool** we will keep only 10 first records.
6. With the **Select tool** we will keep only two fields: *Country of asylum* and *Sum_Value*.

Record	Country / territory of asylum/residence	Sum_Value
1	Turkey	2503549
2	Lebanon	1062690
3	Jordan	628223
4	Iraq	244830
5	Egypt	117635
6	Germany	115604
7	Sweden	52707
8	Armenia	16611
9	Netherlands	16184
10	Austria	14894

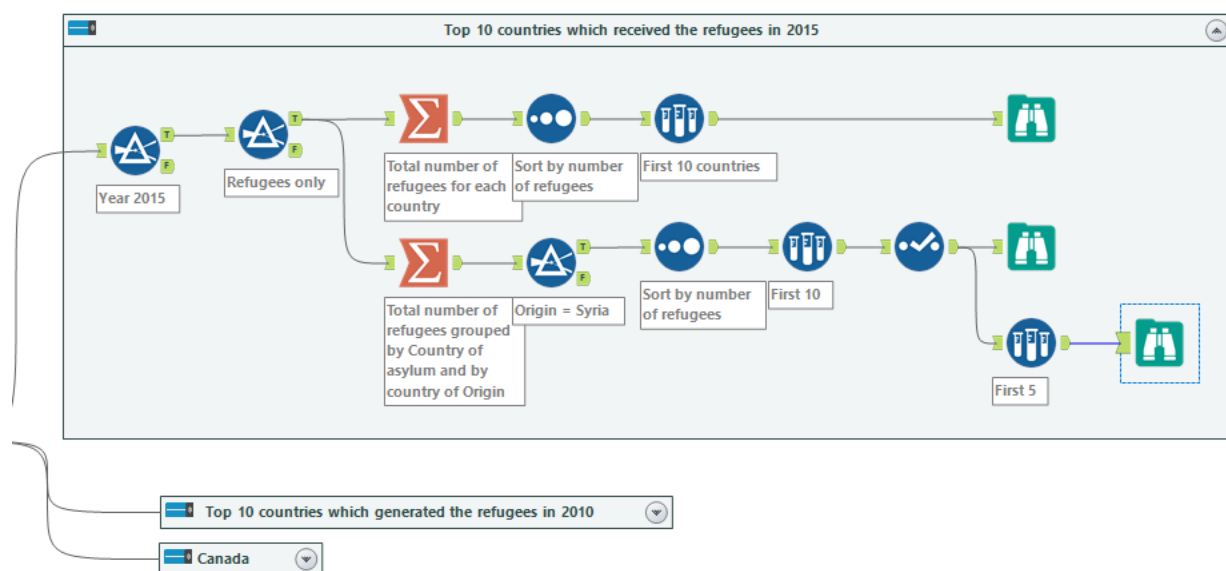
Results:

Country / territory of asylum/residence	Sum_Value
Turkey	2503549
Lebanon	1062690
Jordan	628223
Iraq	244830
Egypt	117635
Germany	115604
Sweden	52707
Armenia	16611
Netherlands	16184
Austria	14894

A COMPARISON BETWEEN THE NUMBER OF RECEIVED REFUGEES AND TOTAL POPULATION

We will use the same tool container.

- Let's get top5 countries that received the refugees from Syria by attaching a **Sample tool** to the previous "branch" and choosing the 5 first records.

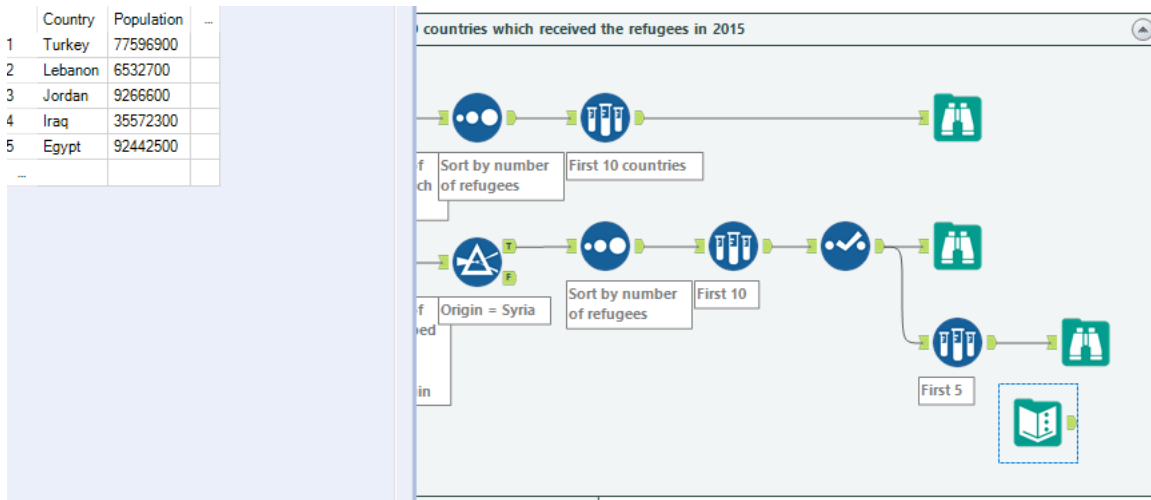


Results - Browse (92) - Input

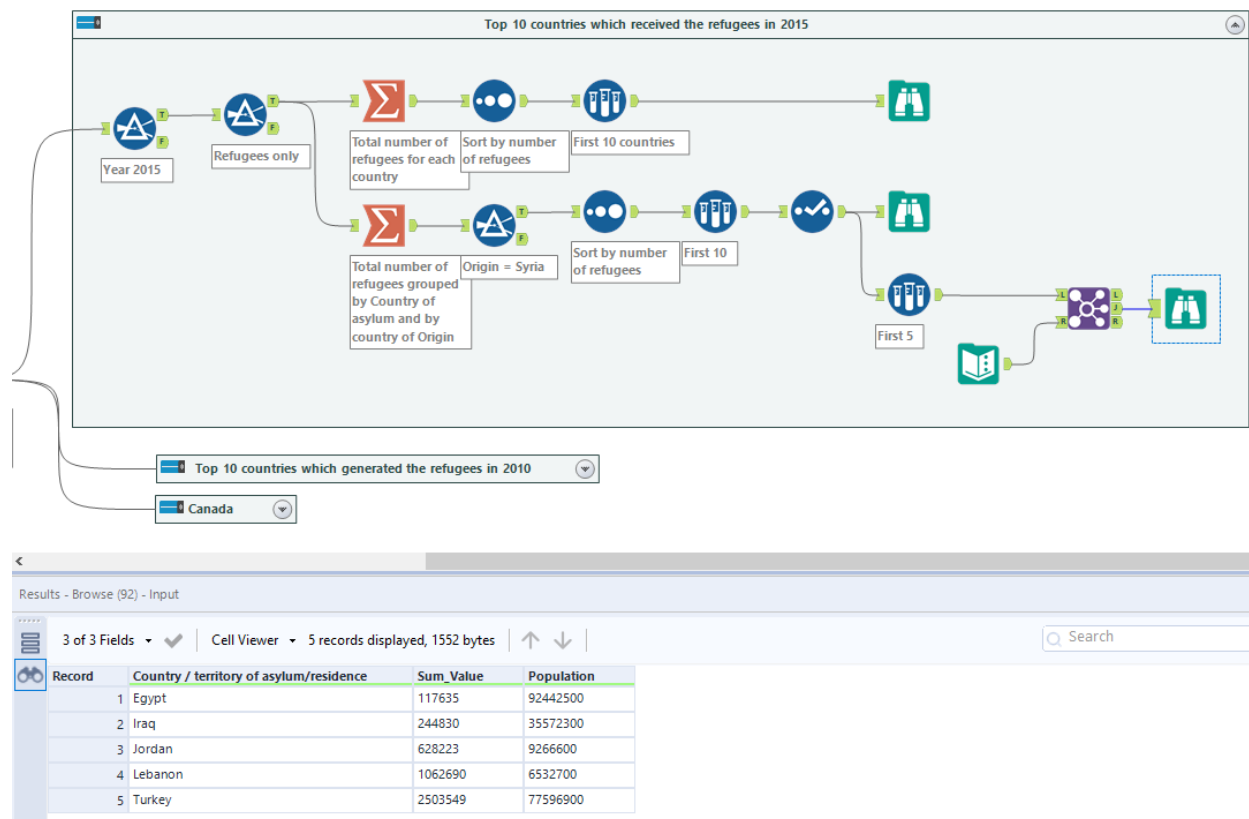
2 of 2 Fields | Cell Viewer | 5 records displayed, 1367 bytes

Record	Country / territory of asylum/residence	Sum_Value
1	Turkey	2503549
2	Lebanon	1062690
3	Jordan	628223
4	Iraq	244830
5	Egypt	117635

- Next step – we will create a table with a list of those countries and their population in 2015 taken from <https://www.google.com/publicdata>, using **Text Input Tool**:



3. With **Join tool**, we will connect this table with the output from step 1 (*Join by Country of asylum and Country*).



4. Using **Multi-Field Formula Tool**, we will create a new field by calculating the percentage of number of Syrian refugees accepted by each of the countries from the total population of these countries. For that we will select *Sum_Value Field* and add this expression: $[_{CurrentField}] * 100 / [Population]$. (See left image below)

Then we will add another **Multi-Field Formula Tool** and create a field with the percentage of number of refugees accepted by each of the countries from the population of Syria in 2015 (17,996,400). The expression will change to $[_{CurrentField}] * 100 / 17996400$. (See right image below).

Select: Numeric Fields:

☒ Sum_Value
☐ Population
☐ Dynamic or Unknown Fields

☒ Copy Output Fields and Add
's % from Population as a Suffix

☒ Change Output Type to FixedDecima Size: 4,1

Variables Functions Saved Expressions

☒ Current Field
☒ Original Fields
☒ Constants

Expression:
[CurrentField_]*100/[Population]

Multi-Field Formula (101) - Configuration

Select: Numeric Fields:

☒ Sum_Value
☐ Population
☐ Sum_Value's % from Population
☐ Dynamic or Unknown Fields

☒ Copy Output Fields and Add
's % from Syria Population as a Suffix

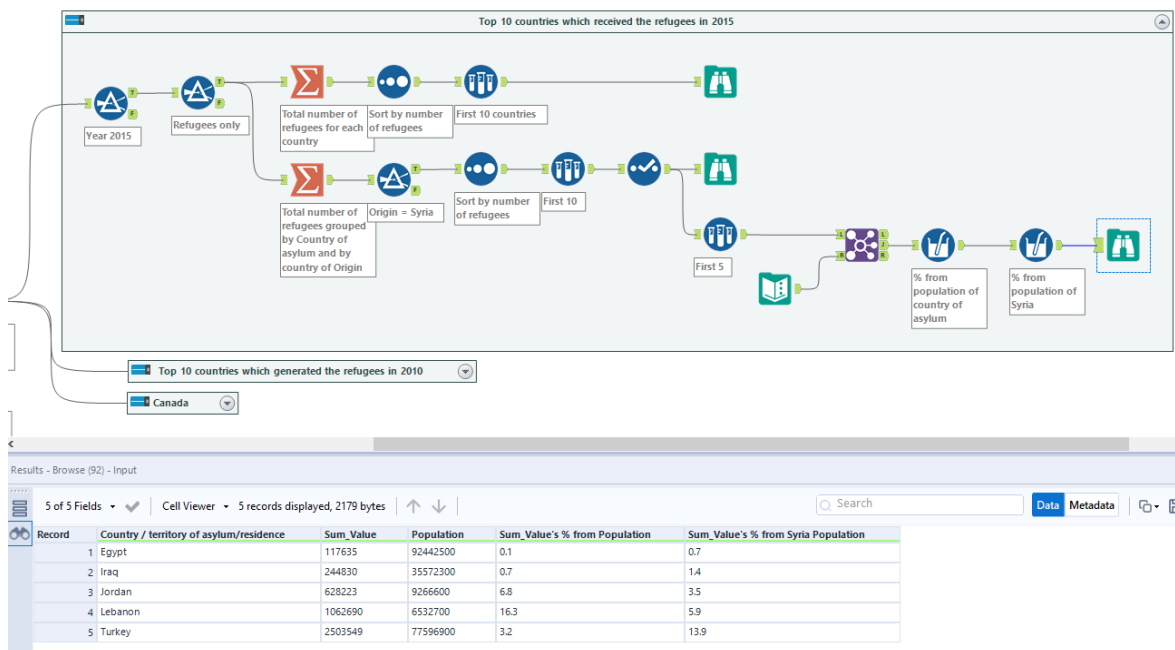
☒ Change Output Type to FixedDecima Size: 4,1

Variables Functions Saved Expressions

☒ Current Field
☒ Original Fields
☒ Constants

Expression:
[CurrentField_]*100/17996400

Here is how it will look:



Results obtained:

Country of asylum	Sum_Value	Population	Sum_Value's % from Population	Sum_Value's % from Population of Syria
Egypt	117635	92442500	0.1	0.7
Iraq	244830	35572300	0.7	1.4
Jordan	628223	9266600	6.8	3.5
Lebanon	1062690	6532700	16.3	5.9
Turkey	2503549	77596900	3.2	13.9

From the numbers we see that Lebanon has the highest percentage of refugees from the total population. It proves that the high flow of refugees greatly influenced the economy and life of the country in general.

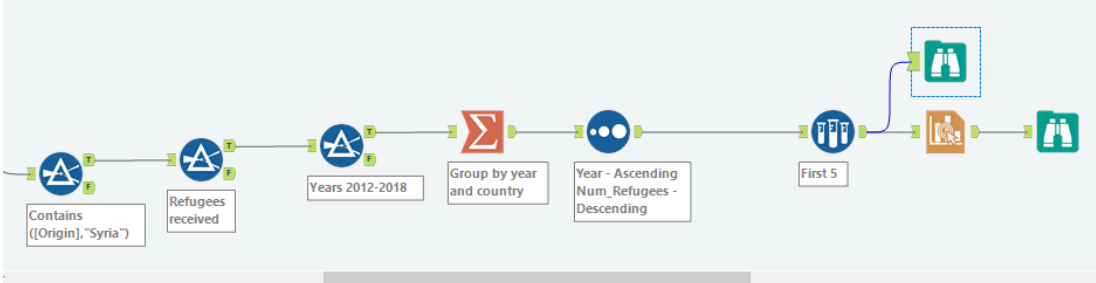
On another side, the highest percentage of refugees from the population of Syria left to Turkey. Lebanon is still on the second place.

YEARLY SYRIAN REFUGEES' TRENDS COMPARISON FOR THE TOP5 COUNTRIES FOR THE PERIOD 2012-2018

Let us start from discovering the top5 countries which received Syrian refugees from 2012 to 2018.

1. **Filter Tool:** *Origin Contains Syria*
2. **Filter Tool:** *Population type Equals Refugees (incl. refugee-like situations)*
3. **Filter Tool:** *[Year] in (2012, 2013, 2014, 2015, 2016, 2017, 2018)*
4. **Summarize Tool:** *Group By Year, Group By Country, Sum of Value (renamed to Num_Refugees)*
5. **Sort Tool:** *Year: ascending; Num_Refugees: descending*
6. **Sample Tool:** *first 5 rows, grouped by Year*
7. **Interactive Chart Tool:**
 - a. **Layer:** *Type: Bar, X: Year, Y: Num_Refugees.*
 - b. **Transforms:** *Split: Enabled, Group by Country*
 - c. **Chart:** *Title: Yearly top5 countries that received the highest number of Syrian refugees*

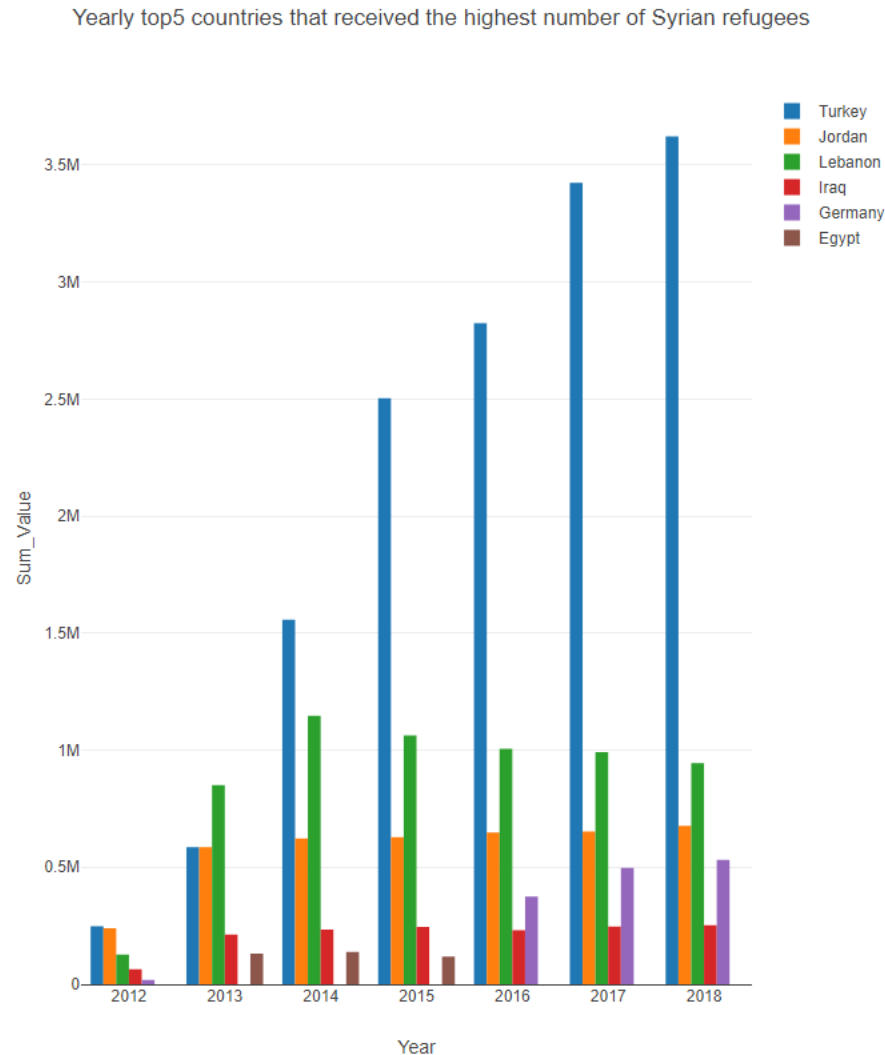
Here is how our branch looks like now (the output is from the Sample Tool):



Results - Browse (252) - Input

Record	Year	Country / territory of asylum/residence	Num_Refugees
1	2012	Turkey	248466
2	2012	Jordan	238798
3	2012	Lebanon	126939
4	2012	Iraq	63586
5	2012	Germany	18165
6	2013	Lebanon	851284
7	2013	Turkey	585601
8	2013	Jordan	585304
9	2013	Iraq	212809
10	2013	Egypt	131659
11	2014	Turkey	1557899
12	2014	Lebanon	1147494
13	2014	Jordan	623112
14	2014	Iraq	234196
15	2014	Egypt	138381
16	2015	Turkey	2503549

Let us look at the chart that we've got:

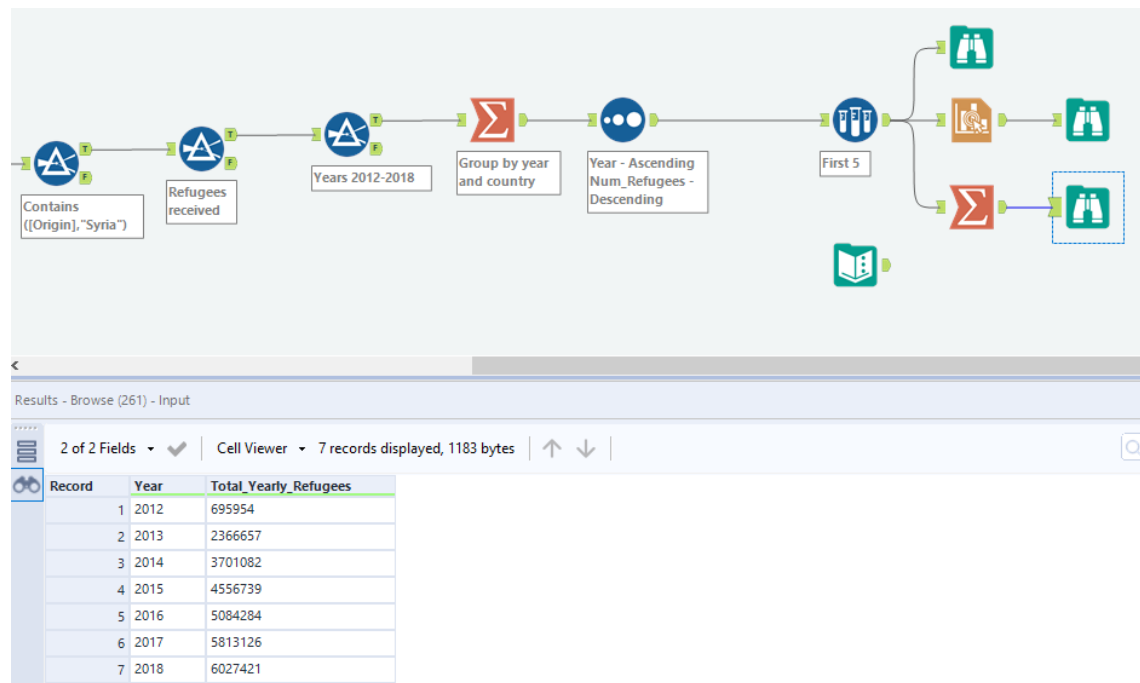


For the yearly top5 countries, during the period 2012-2018, we've got an intersection of six countries: Turkey, Jordan, Lebanon, Iraq, Germany, and Egypt. We will continue our analysis with the first 5 countries: **Turkey, Jordan, Lebanon, Iraq, and Germany**. We chose Germany over Egypt out of personal interest and equally because in 2016, 2017, and 2018 the number of refugees in Germany increased.

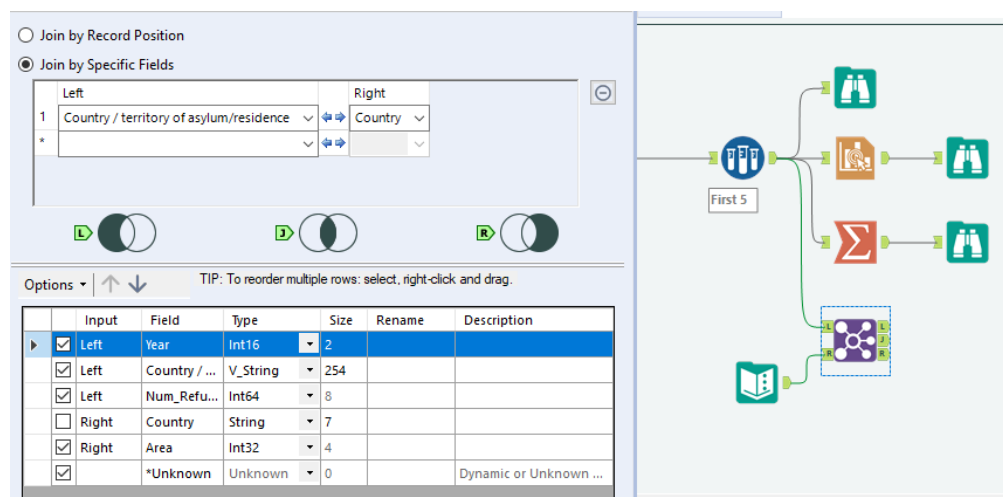
8. We will now create a **Text Input** with the list of the countries and their Area in square kilometers for future calculations:

Country	Area
Turkey	783562
Lebanon	10452
Jordan	89342
Germany	357368
Iraq	437072

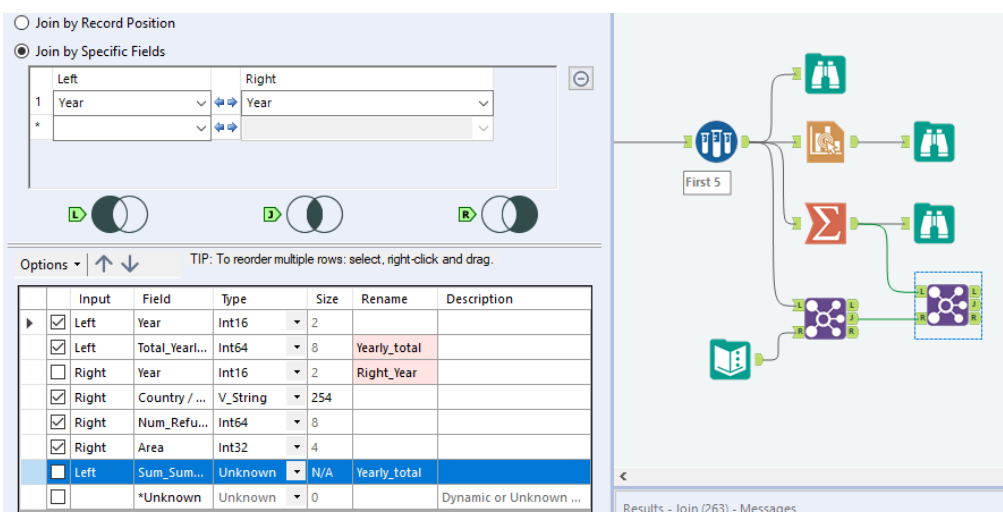
9. We will connect **Summarize Tool** to the last Sample tool and calculate the Yearly Total amount of Syrian refugees for the future calculations: *Group By Year, Sum of Num_Refugees renamed to Total_Yearly_Refugees*:



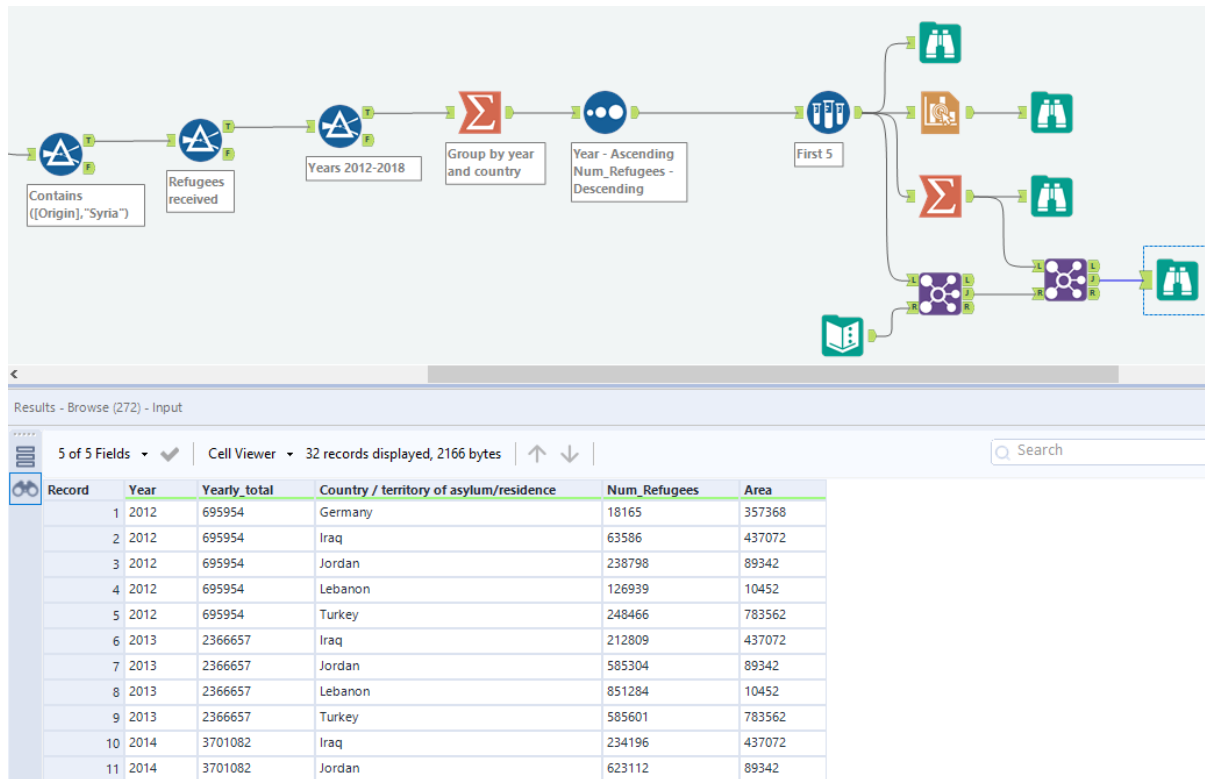
10. Now, we will connect the output of Sample Tool and Text Input Tool using **Join Tool**:



11. Using **Join Tool** we will now connect the output we've got with the *Yearly Total* amounts from step 9:



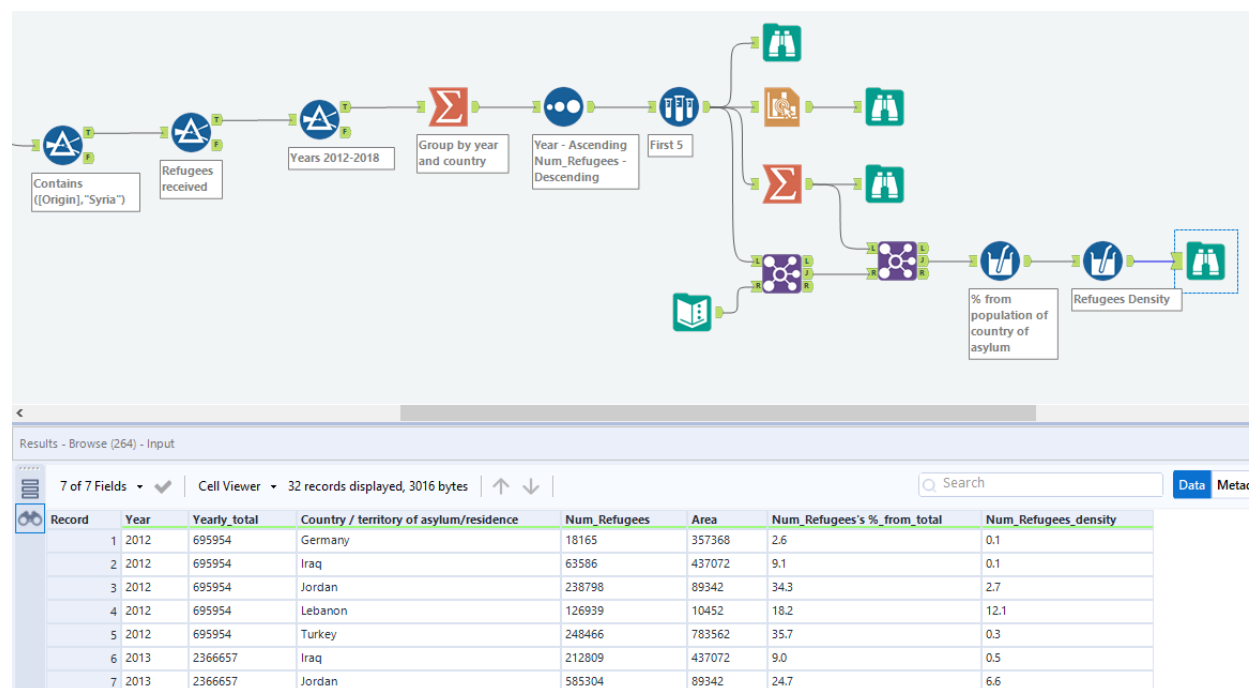
Output:



12. We will not calculate the yearly percentage of refugees for each country from the total yearly number of refugees. We will use **Multi-Field Formula Tool**: For the *Num_Refugees* field checked we use this *Expression*: $[_CurrentField_] * 100 / [Yearly_total]$ and change output type to *FixedDecima, Size 4.1*.

13. With another **Multi-Field Formula Tool** we will calculate the refugees density based on the country area: for the *Num_Refugees* field checked we use this *Expression*: $[_CurrentField_] / [Area]$ and change output type to *FixedDecima, Size 5.1*.

Now we finally have all the necessary fields:



The last step is to create few charts that will hopefully shows us some interesting trends.

14. **Interactive Chart Tool:**

- a. **Layer:** Type: Line, X: Year, Y: Num_Refugees
- b. **Transforms:** Enabled, Group By Country
- c. **Chart:** Title: *Total Syrian refugees 2015-2018*

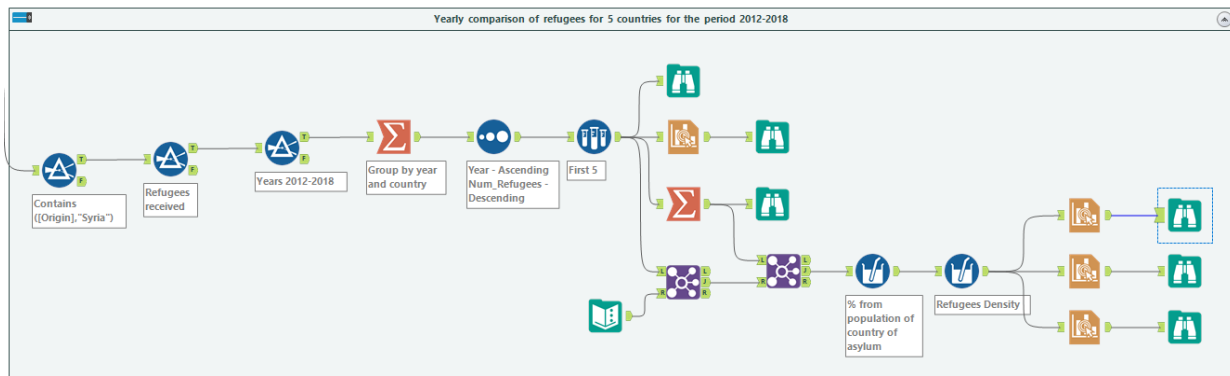
15. **Interactive Chart Tool:**

- a. **Layer:** Type: Line, X: Year, Y: Num_Refugees_%_from_total
- b. **Transforms:** Enabled, Group By Country
- c. **Chart:** Title: *Percentage from total yearly Syrian refugees 2015-2018*

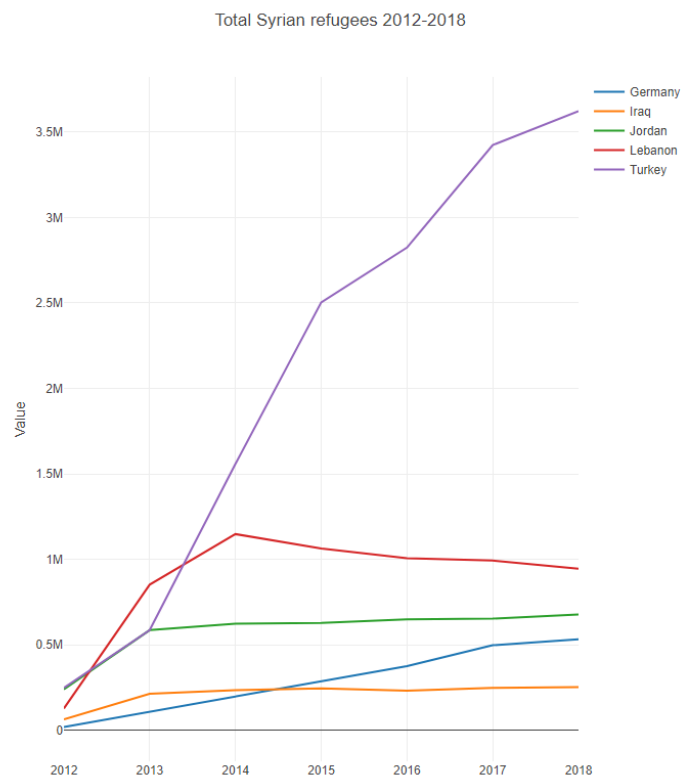
16. **Interactive Chart Tool:**

- a. **Layer:** Type: Line, X: Year, Y: Num_Refugees_density
- b. **Transforms:** Enabled, Group By Country
- c. **Chart:** Title: *Syrian refugees' density 2015-2018 (per square kilometer)*

This is how our **Tool Container** looks:



Let us check our graphical results:

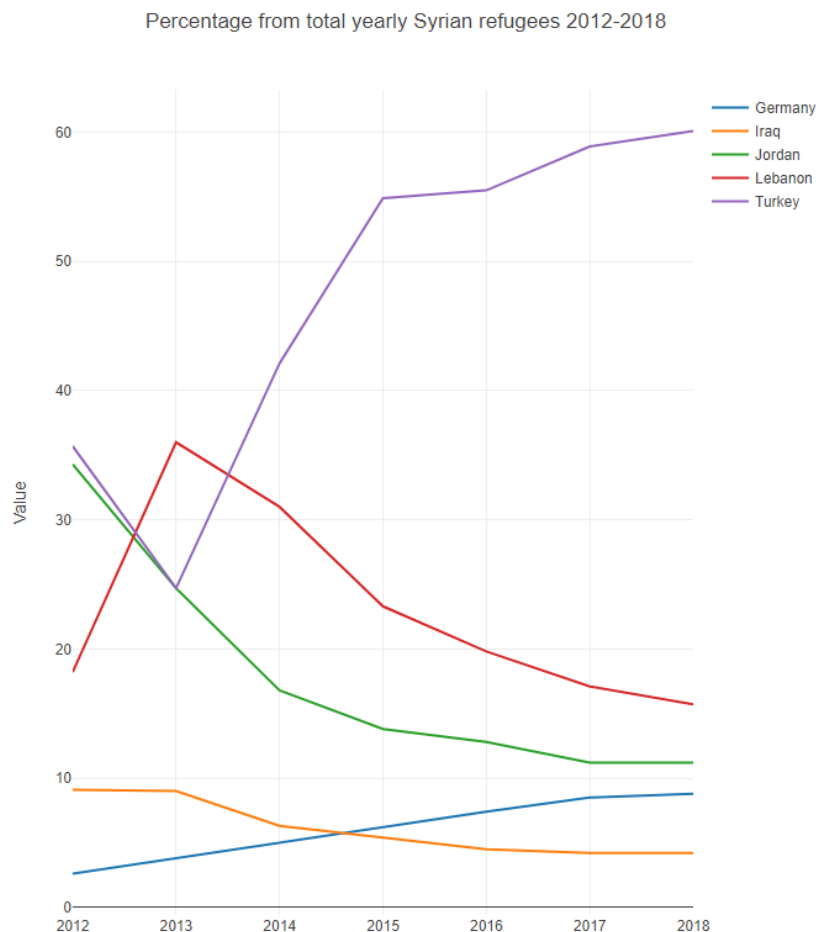


Turkey is home to the highest number of Syrian. Financial aid from EU was promised in November 2015 but the promise is still not fulfilled. The number continues to raise attending the marge of 3.6 million refugees.

Lebanon attempted the pick in 2014, after the total number slightly decreased. Nevertheless, according to Government of Lebanon estimates, the country hosted 1.5 million Syrian refugees as of October 2016. This number shows that there are around 500k nonregistered refugees.

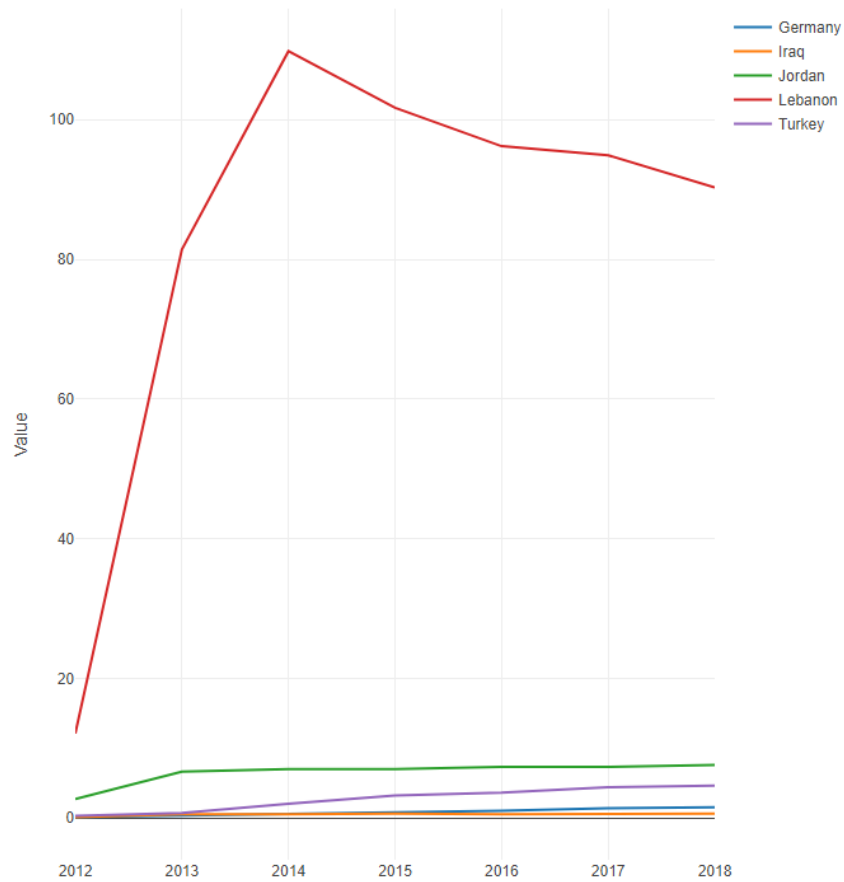
Iraq and Jordan have a stable number of refugees after 2013, reaching in 2018 676k in Jordan and 252k in Iraq.

Germany's trend is slowly going up, reaching 532k refugees in 2018.



In 2013 Lebanon acquired the highest percentage of the total number of Syrian refugees (36%), for the other years Turkey is leading, receiving up to 60%.

Syrian refugees' density 2012-2018 (per square kilometer)



This trend shows that among all the countries, Lebanon as an exceedingly small country has the highest refugees' density per square kilometer. This crisis has increased population drastically and had added a large strain on the country's infrastructure. The Lebanese government chose not to establish camps, and thus the refugees have settled throughout country.