

**Using Free and Open Source Software (FOSS) mapping and charting tools
to visualize refugee and immigrant integration data**

Wing-Yan Wong

Pennsylvania State University

Master of Geographic Information System

Capstone Project

Table of Contents

- Abstract 1
- 1 Introduction 1**
- 2 Background 1**
- 3 Goal and Objectives 3**
- 4 Methods 3**
 - 4.1 Data Sources 3
 - 4.2 Technologies 4
- 5 Usability Evaluation 5**
- 6 Results 6**
 - 6.1 Refugees and Asylum Seekers 6
 - 6.2 Demographics 8
 - 6.3 Education 8
 - Highest Educational Attainment 9
 - Early School Leavers 11
 - 6.4 Employment 11
 - Employment Rate 11
 - 6.5 Social Inclusion 13
- 7 Next Steps 13**
- 8 Acknowledgements 14**
- References 14
- Links 15

Abstract

In 2015, 1.3 million asylum seekers surged in Europe, and the continuous flow of refugees and asylum seekers has been a challenge for humanitarian agencies and local governments to provide assistance and protection. Driven by European Union (EU) policy documents such as the Common Basic Principles (CBP) for Immigrant Integration Policy, Europe 2020, and Indicators of Immigrant Integration, the need for a data visualization tool that enables researchers and policy makers to assess the situations has emerged. This paper discusses an approach using JavaScript libraries (leaflet, d3, crossfilter, dc and d3.parcoords) to visualize the refugee demographic data and immigrant integration indicators in EU countries. By combining these JavaScript libraries and open web mapping techniques, displaying high dimensional data online is a breeze; the interactive graphs such as crossfilters and parallel coordinates with brushing capabilities allow this complex dataset to be queried and viewed quickly using a web browser. Several use case scenarios will demonstrate how the web-based tool could be used to answer questions.

Keywords: Common Basic Principles, Europe 2020, European Union, Indicators of Immigrant Integration, Free and Open Source Software, D3.js, Dc.js, Crossfilter.js, Leaflet.js, Web Mapping, Refugee, Asylum Seeker.

1 Introduction

In 2015, 1.3 million asylum seekers surged in Europe. Under the United Nations High Commissioner for Refugees (UNHCR), Convention of 1951 and Protocol of 1967 relating to the status of refugees, a person with refugee status can receive legal protection, other assistance and social rights from the countries who have signed the document. Different from the refugee status, asylum seekers are people who have applied for the refugee status, however, the claims have not yet been determined. On average, it takes 2 years for asylum seekers to be registered as refugees in the host countries.

UNHCR works to assist and protect refugees globally. Budgets are funded for purposes including shelter, water and sanitation, food, medical assistance, and education. However, with Europe's recent sharp increase in refugees and asylum seekers, not only are resources insufficient to meet their needs, the actual funded amount is far less than the requested (Dobbs & Clayton, 2015). As a result, the needs are dramatically outpacing resources. Providing basic needs is the immediate response to the crisis; in the long-term, we need to develop the sustainable solutions that can help the refugees to rebuild their lives.

2 Background

The European Union (EU) is a politico-economic union of 28 member states. In November 2004, the EU's Justice and Home Affairs Council, agreed upon and adopted the *Common Basic Principles for Immigrant Integration Policy* (European Economic and Social Committee, 2004). The CBPs became the driving force behind the integration of the migrants in the EU. Among 11 CBPs, 4 of them are closely related to this project:

- CBP 1 “Integration is a dynamic, two-way process of mutual accommodation by all immigrants and residents of Member State”.
- CBP 3 “Employment is a key part of the integration process and is central to the participation of immigrants, to the contributions immigrants make to the host society, and to making such contributions”.
- CBP 4 “Basic knowledge of the host society’s language, history, and institutions is indispensable to integration; enabling immigrants to acquire this basic knowledge is essential to successful integration”.
- CBP 5 “Efforts in education are critical to preparing immigrants, and particularly their descendants, to be more successful and more active participants in society”.

Europe 2020 is the EU’s 10-year strategy for social and economic development (European Commission, 2016); it addresses how the policy makers should confront key long-term challenges such as demographic change and globalization. In 2010, the European Commission set 5 headline targets for EU countries to achieve in 2020. Topics include employment, education, fighting poverty and social exclusion, research and development, and climate/energy. The three headline targets related to this project are:

- The employment rate of the population aged 20-64 should increase from the current 69% to at least 75%, including through the greater involvement of women, older workers and the better integration of migrants in the work force.
- The early school leavers should be reduced to less than 10% from the current 15%, while the share of the population aged 30-34 having completed university education should increase from 31% to at least 40% in 2020.
- The number of Europeans living below the national poverty lines should be reduced by 25%, lifting over 20 million people out of poverty (using 2008 as baseline year).

Driven by the Common Basic Principles for Immigrant Integration Policy and the EU 2020 Strategy, EU countries are encouraged to grow their societies smartly, sustainably and inclusively. As reported in the *Using EU Indicators of Immigrant Integration* (European Commission, 2013), researchers and policy makers are advised to use the indicators for:

- Understanding the national contexts for integration in a comparative way,
- Evaluating the results of policies, and
- Using targets to mainstream and improve integration

Suggested in the *Using EU Indicators of Immigrant Integration*, there are three different types of factors influencing integration outcome: first, migration population, second, general policies and factors such as labor market and education system, and third migration and integration policies and factors. The characteristics of the migrant population that have strong influence on the integration outcome are: age, gender, education, job quality, and country of origin.

With that, this study focuses on understanding the migration population and the national contexts for integration in a comparative way.

3 Goal and Objectives

The project goal is to develop the web-based visual analytics tools for researchers and policy makers to interactively explore the refugee and immigrant integration data in EU countries. The project objectives are:

- To view how many asylum seekers and refugees have entered in EU countries over the years,
- To understand who the refugees are demographically,
- To identify which country and group of people need assistance the most, and
- To compare how well the host countries integrated the migrants in the past.

4 Methods

In order to visualize the refugee and immigrant integration indicator data, a web map with various graphic representations are implemented for two reasons: displaying data online can reach a large audience without time and location constraints, and using interactive charts and graphs can show the highly dimensional datasets more effectively.

Only Free and Open Source Software (FOSS) is adopted in this research because of the low cost and high flexibility. Sections 4.1 and 4.2 will describe the data sources and technologies used, and will explain the benefits to the project.

4.1 Data Sources

For the spatial data, 33 country boundaries in GeoJSON format were downloaded from GitHub [1]. Considering the country boundaries are rarely changed and not likely to be used for spatial analysis in this project, GeoJSON files are used to represent the country geometry on the web map for the simplicity and fast loading time. 33 countries consist of the existing 28 EU countries and 5 non-EU countries (Finland, Iceland, Macedonia, Switzerland, and Turkey). The 5 non-EU countries are added because they are included in the immigrant integration indicator datasets.

For the non-spatial data, the statistical data in .csv format were downloaded from two websites, the United Nations High Commissioner for Refugees (UNHCR) [2] and Eurostat [3]. UNHCR works to assist and protect refugees globally; Eurostat provides statistics for European countries only.

The refugee and asylum seeker data were downloaded from the UNHCR population statistics page. The csv file has the following columns:

- Year
- Host country
- Origin country
- Number of asylum seekers
- Number of refugees

The “persons of concern” demographic data were also downloaded from the UNHCR population statistics page. The csv file has the following columns:

- Year
- Host country
- Age group
- Gender

According to UNHCR, persons of concern include the following groups of people: refugees, asylum-seekers, internally displaced persons (IDPs), returned refugees, returned IDPs, stateless persons, and others of concern. Among the 33 host countries, some of them do not have data available; therefore, the total persons of concern are less than the total of refugee and asylum seeker in this project. Nevertheless, the persons of concern dataset is included because understanding the demographic characteristics is important for both refugee crisis response and immigrant integration planning.

The following EU Migration Integration Indicators were downloaded from the Eurostat website:

- Education: highest educational attainment, early school leavers, participation in lifelong learning, young people not in education, employment or training (NEET)
- Employment: employment rate
- Social Inclusion: people at risk of poverty and social exclusion (AROPE)

4.2 Technologies

FOSS JavaScript (.js) libraries are used in this study. In order to display the highly dimensional statistical data, d3.js, dc.js, and crossfilter.js are utilized.

The D3.js (Data-Driven Documents) JavaScript library [4] manipulates documents based on data. Combining HTML, Scalable Vector Graphics (SVG) and Cascading Style Sheets (CSS), data visualizations such as bar charts, row charts, pie charts can be created and accessed by anyone with a web browser. The crossfilter.js library [5] supports extremely fast multidimensional data manipulation and filtering. The dc.js (Dimensional Charting) library [6] is designed for coordinated visualization utilizing both d3.js and crossfilter.js together.

In order to show the population of refugees and asylum seekers on a choropleth map and on graphs and charts that provide brushing capability, leaflet.js [7] and dc-leaflet.js [8] are implemented. Leaflet.js is an open source JavaScript library for interactive mapping. The base map data source is contributed and maintained by OpenStreetMap [9]. Leaflet.js has a well-documented API and a large user community that has contributed plentiful plugin libraries. For example, dc-leaflet.js is an addon that enables the brushing capability between the map and charts.

For the web front-end framework, bootstrap.js [10] was chosen. Its downloadable templates allow faster web development with professional appearance. jQuery.js [11] was also adopted because it simplifies the implementation of JavaScript on web design. It supports HTML/DOM manipulation, CSS manipulation, HTML event methods, and much more. Features such as collapsible panels and animated buttons on the web page can be created with only a few lines of code.

The Penn State Access Account Storage Space (PASS) is used for web hosting. It offers 10 GB free online storage for students, faculty, and staff for file backup and personal web space. It minimizes maintenance cost for this project.



Name	Size	Type
..		Folder
css		Folder
data		Folder
font-awesome		Folder
fonts		Folder
geo		Folder
img		Folder
js		Folder
About.html	12.9 KB	HTML Document
demographic.html	17.8 KB	HTML Document

Figure 1. The project file architecture in PASS.

Figure 1 above depicts the directory structure on the web server. When users open the web browser, it will be pointed to the html files in the webservice which is the base directory. The folders store the reference files in different formats: css folder stores the cascading style sheets; the data folder stores all CSV format data; the font-awesome and font folders store the special designed fonts; the geo folder stores the GeoJSON file; the img folder stores the graphics in PNG formats; and the js folder stores all the JavaScript libraries.

5 Usability Evaluation

In order to test if the web interface design is intuitive enough for users to navigate around with no prior training, a group of five people were invited to participate in a usability evaluation. User 1 and 2 are Master of GIS students with Penn State University. They were enrolled in the graduate course Challenges in Global Geospatial Analytics that is related to humanitarian crisis and data visualization; user 3 is a software engineer who is experienced in web interface design; user 4 is a software architect who has both technical and management skills; user 5 is a project manager. Table 1 below summarizes the users’

skills and knowledge in 4 areas: GIS, humanitarian, programming, and management. Thus, there are two users in each area.

Table 1. Skills and knowledge matrix of the participants in the usability evaluation.

User	GIS	Humanitarian	Programming	Management
1	x	x		
2	x	x		
3			x	
4			x	x
5				x

The five users answered the same eleven questions posted on the web page [12]; responses and feedback were collected by email. Suggestions such as adding tool tips and ordering the year ascendingly in the select combo box were adopted. An unknown programming bug in the paginated table was also reported by one of the users, even though it was not in the questionnaire.

6 Results

The Refugee and Immigrant Integration Dashboard (R2i) [13] was created, and the source codes were uploaded to GitHub [14]. R2i is a web-based mapping and charting tool to explore the refugee and immigrant integration data. The portal page provides access to three different datasets: Refugees and Asylum Seekers, Demographics, and the Immigrant Integration Indicators.

6.1 Refugees and Asylum Seekers

The Refugees and Asylum Seekers page [15] has two main components: a choropleth map on the left; select combo boxes, bar charts, row charts and table view on the right. As shown in Figure 2, the map shows the 33 host countries which are colored by the total of refugees and asylum seekers in thousands (k). On the right, there are collapsible panels containing various interactive items for user to filter the data by host country, by origin country, and by year. At the bottom, there is also a paginated table for user to view and download the dataset.

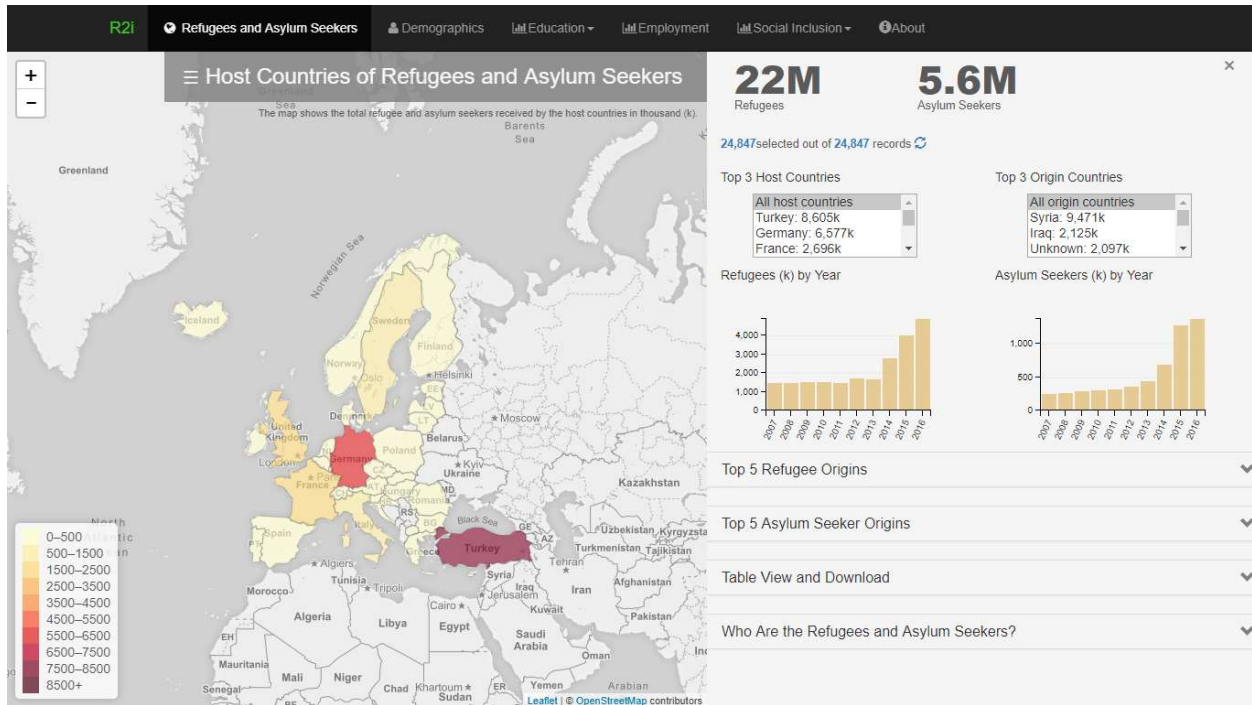


Figure 2. The Refugees and Asylum Seekers page displays the entire dataset between year 2007 and 2016 by default.

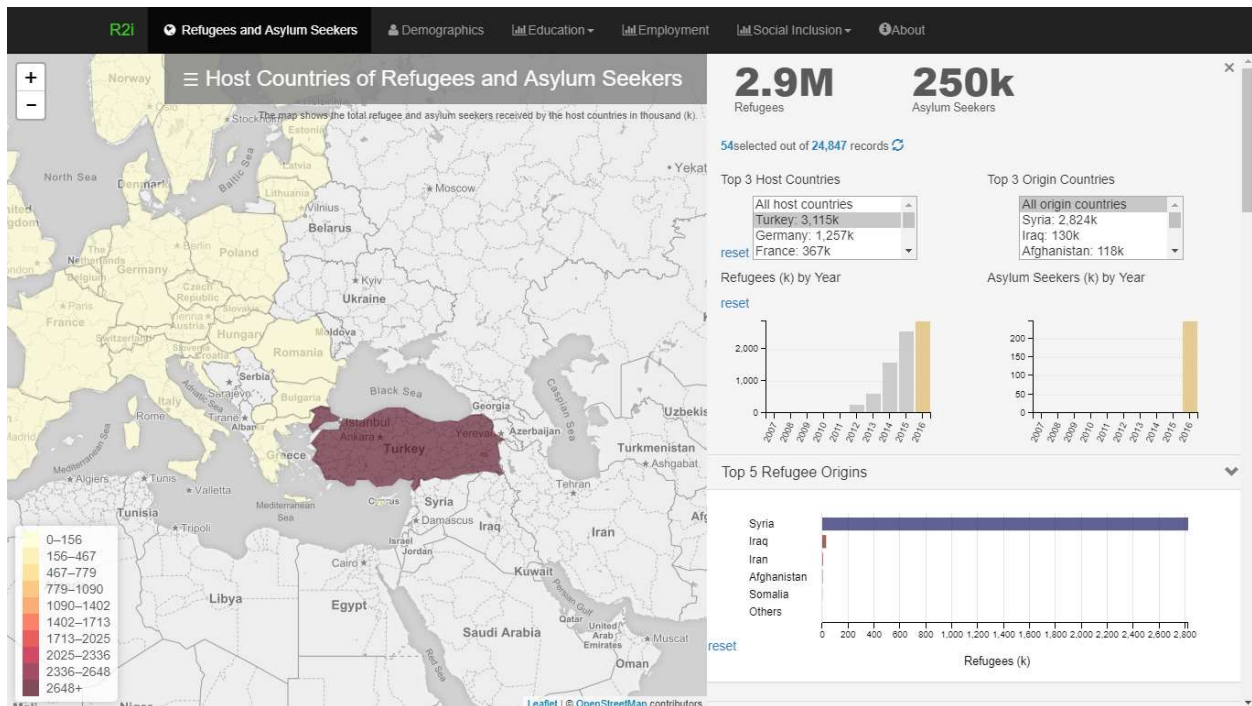


Figure 3. The dataset is filtered by year 2016 and top host country Turkey; the resulting top refugee origin country is Syria.

The advantage of brushing the map and charts together is the multi-dimensional dataset can be queried and visualized on one page. For example, illustrated in Figure 3, by choosing the year 2016 and top host country Turkey, we learn that the top refugee origin is Syria. The choropleth map and legend is also updated accordingly. Since all the graphics and styles are CSS and SVG drawn by the web browser, the data filtering, manipulation and visualization is extremely fast.

6.2 Demographics

The demographics page [16] shows the persons of concern characteristics by year, age group, gender, and host country; the user can filter the data by choosing one or more of these attributes. In Figure 4 below, the stacked bar chart on the left shows the year on the x-axis, displays the total persons of concern on the y-axis, and stacks the bar by age group. On the right, there are 3 select combo boxes that allow the user to filter the data by year, age group and host country. At the bottom is a row chart showing the top 5 host countries, and a pie chart displaying the gender by percentage.

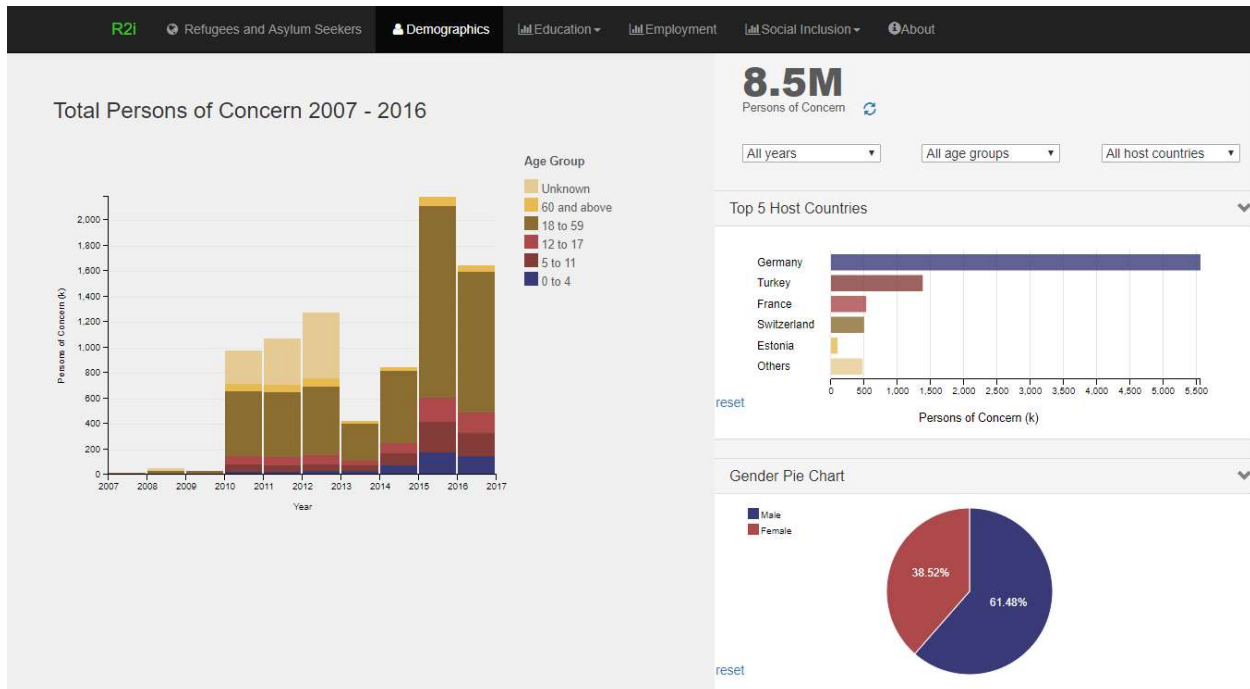


Figure 4. The Demographics page shows the Persons of Concern characteristics by year, age group, gender, and host country.

6.3 Education

A parallel coordinate plot is a set of graphs that have multiple y axes showing a high-dimensional dataset. Each y-axis represents one column in the .csv file. Figure 5 below shows one of the immigrant integration indicators, the highest educational attainment [17] in the EU countries. From the left, the plot shows the host country, year, age group, primary education (%), secondary education (%), tertiary education (%), and the country of birth

(COB)¹. The user left-clicks and drags along any of the y axes to filter the data. The y axes re-orderable. The panel on the right is the table view of the dataset. The corresponding line on the plot is highlighted when user places the mouse over each row in the table.

In addition to the highest educational attainment, 3 other education indicators are included in this project: early school leavers [18], lifelong learning [19], and NEET [20]. The following paragraphs will illustrate how the parallel coordinate plots can visualize the educational gap between the population born outside and within the host countries.

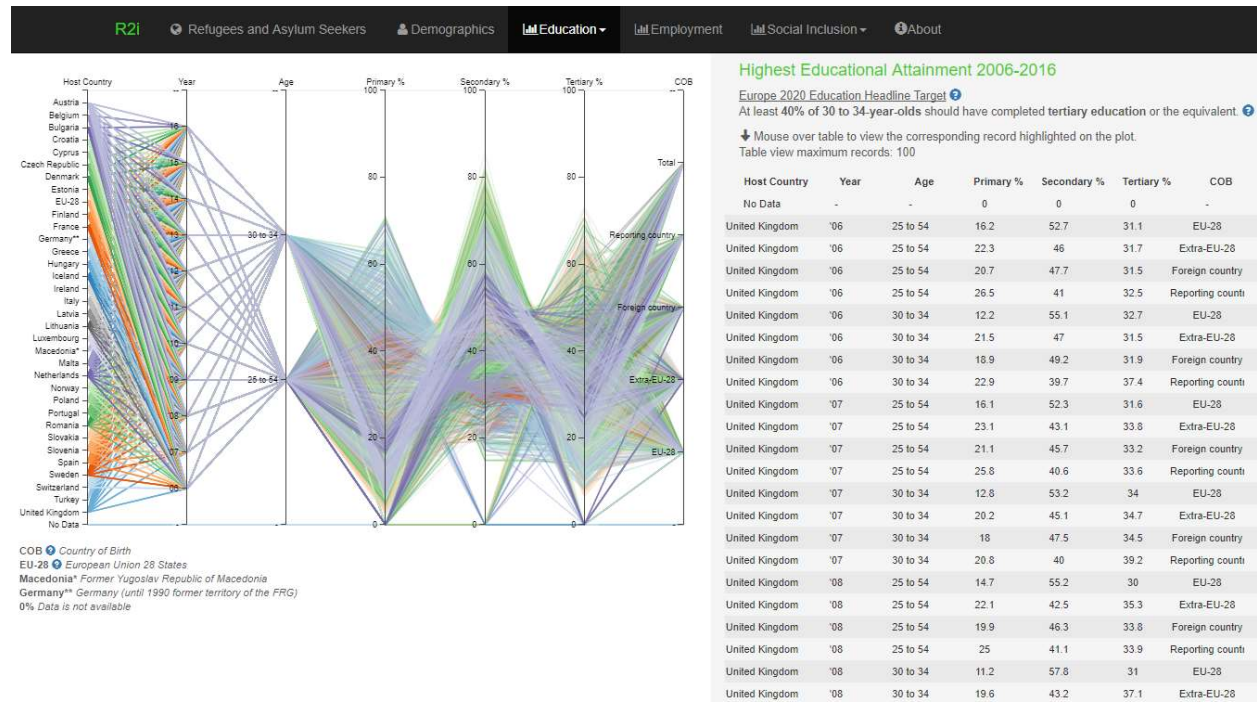


Figure 5. The Highest Educational Attainment page shows the percentage in primary, secondary, and tertiary education in the host countries by country of birth.

Highest Educational Attainment

One of the Europe 2020 education headline targets is at least 40% of 30- to 34-year-olds should have completed tertiary education or the equivalent. In order to find out which countries have already achieved the goal, the user can filter the data by choosing the year 2016, age group 30-40, tertiary education 40% and above, and the COB total population. The top 5 countries shown in Figure 6 are: Lithuania (58.7%), Luxembourg (54.6%), Switzerland (53.5%), Cyprus (53.4%), and Ireland (52.9%).

In some countries, the foreign-born populations have higher percentage in tertiary education attainment than those born in the reporting country. For example, as shown in Figure 7, in 2016, the population born outside United Kingdom (54.7%) has higher tertiary

¹ COB: Total population, foreign born population, reporting country born population, population born in the EU-28 countries and population born outside the EU-28 countries.

education attainment than those born within the United Kingdom (45.6%); the comparison between population born outside and within Turkey is similar (36.8% vs 26.2%). Thus, the

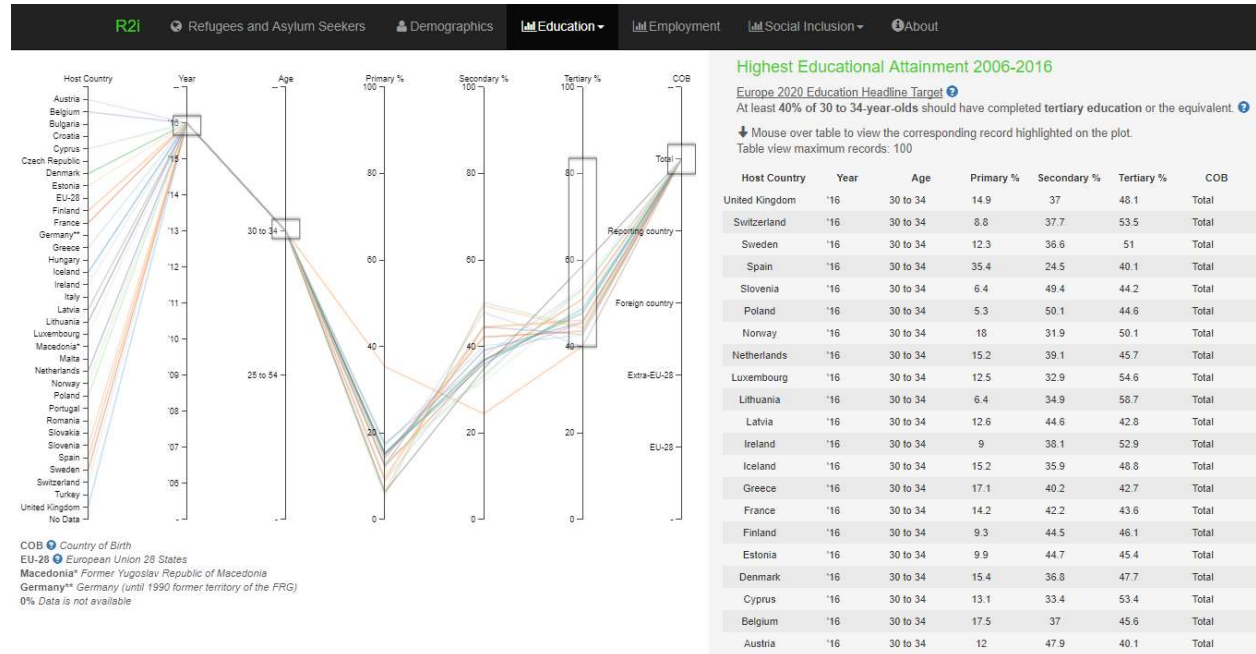


Figure 6. The list of host countries in which at least 40% of the 30- to 40-year-old population has completed tertiary education in 2016.

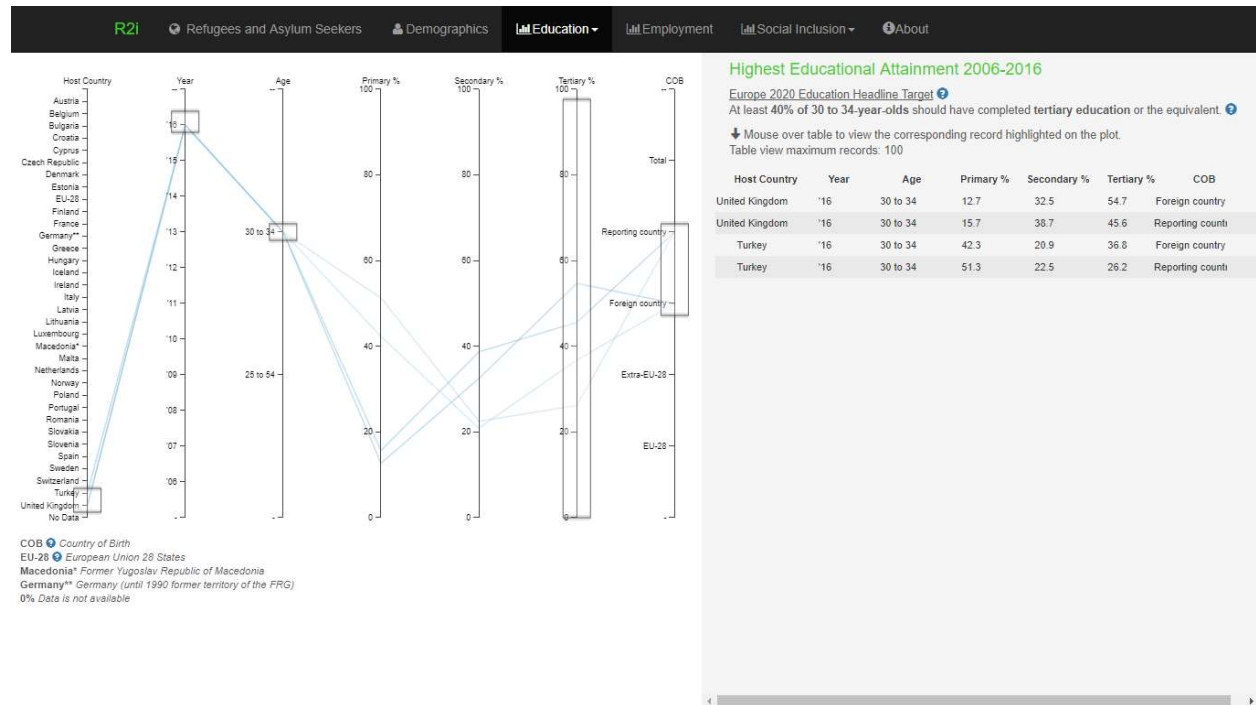


Figure 7. The comparison of highest educational attainment between the population born outside and within the host countries.

data suggests that in some cases the immigrants help the host countries to achieve the Europe 2020 education goal.

Early School Leavers

Another Europe 2020 education headline target is the early school leavers should be reduced to less than 10%. The young foreign-born persons are generally at greater risk of leaving education without having completed upper secondary education level than the people born in the host countries. For examples, Turkey (31.7% vs 18%); Switzerland (5.9% vs 1.4%); Sweden (7.9% vs 2.4%) as shown in Figure 8. The United Kingdom is the exception in that the young foreign-born persons have a lower early school leaver rate (3.4% vs 5.6%) than the persons born in the host country.



Figure 8. The young foreign-born persons have higher percentages in the Early School Leaver indicator than people born in the host countries.

6.4 Employment

Employment rate is one of the headline targets in Europe 2020. The plot [21] in Figure 9 shows the host country, year, gender, employment (%) and country of citizenship (COC). COC is similar to COB; instead of referring to the origin of birth, it is made reference to the country of citizenship.

Employment Rate

The employment headline target is to increase the employment rate of the population aged 20-64 to at least 75%, including through the greater involvement of women, older workers and the better integration of migrants in the work force. In 2016, many countries already have 75% of employment rate in the aged 20-64 population as illustrated in Figure 9.

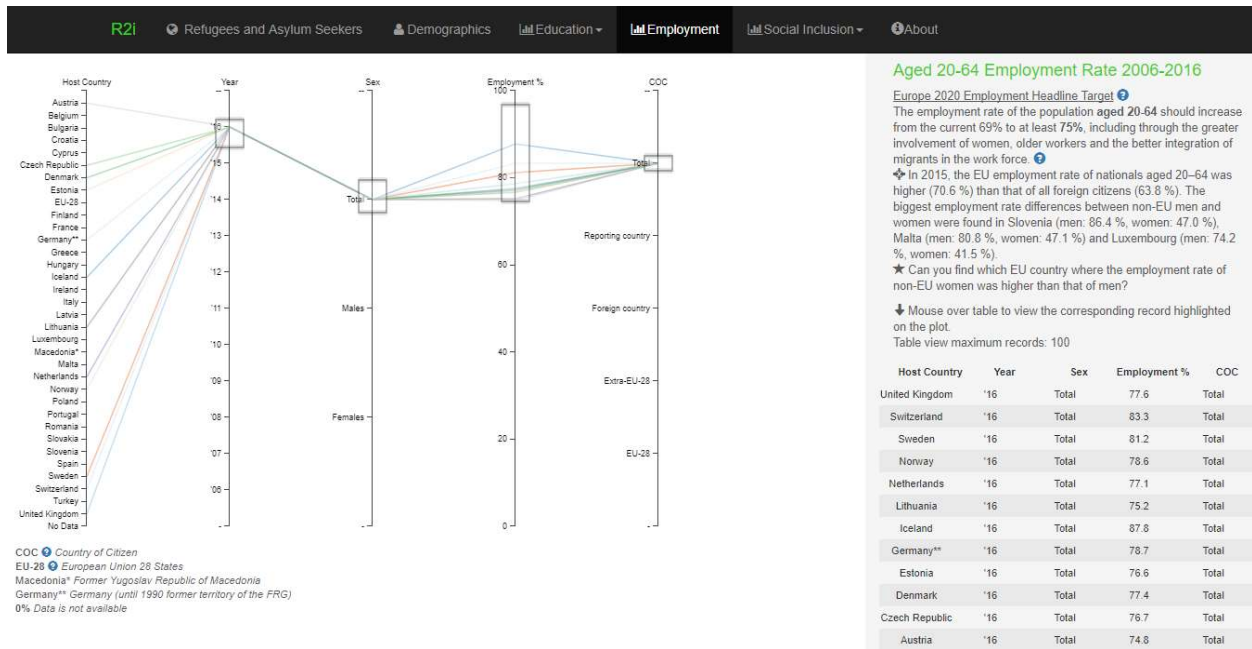


Figure 9. The list of countries where the total population has at least 75% employment rate in 2016.

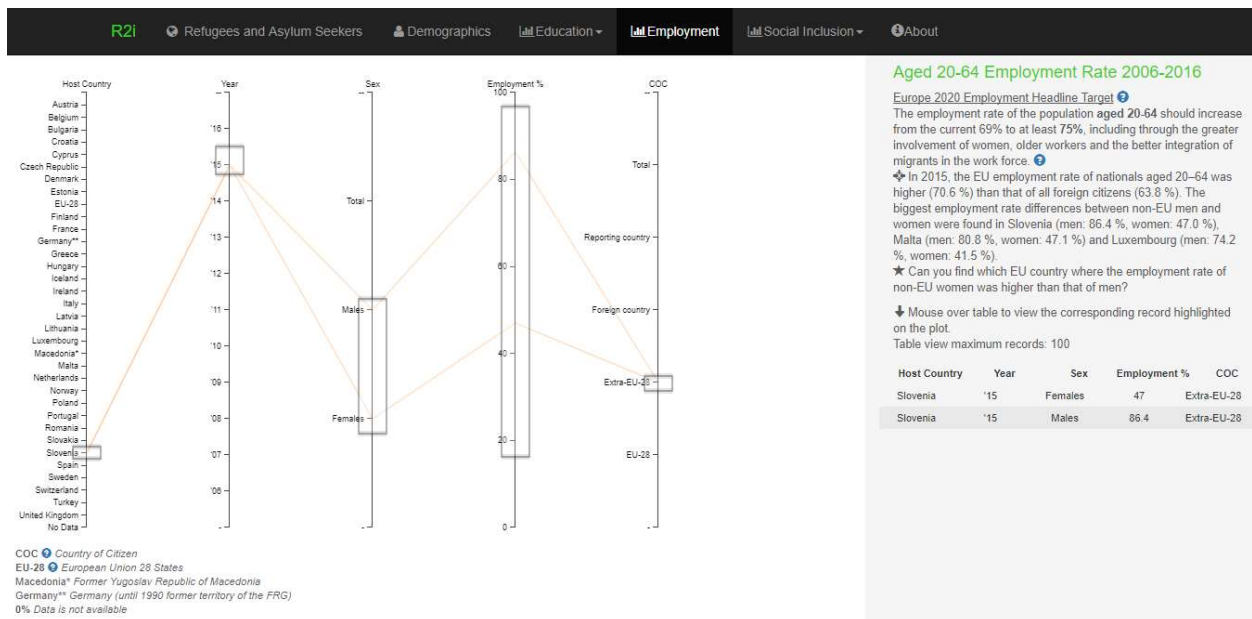


Figure 10. The gap in employment rate between non-EU men and women in Slovenia 2015 (men: 86.4 %, women: 47.0 %).

Comparing the EU employment rate of nationals aged 20-64 with all foreign citizens, in 2015 the former was higher (70.6%) than the latter (63.8%). The biggest employment rate difference between non-EU men and women was found in Slovenia (men: 86.4 %, women: 47.0 %) as shown in Figure 10.

6.5 Social Inclusion

Social inclusion, or AROPE, is one of the Europe 2020 headline targets. Both COB [22] and by COC [23] datasets are included. In Figure 11 below, the plot shows that in EU-28 countries, the non-EU-born population had a much higher percentage in AROPE than the native-born population.

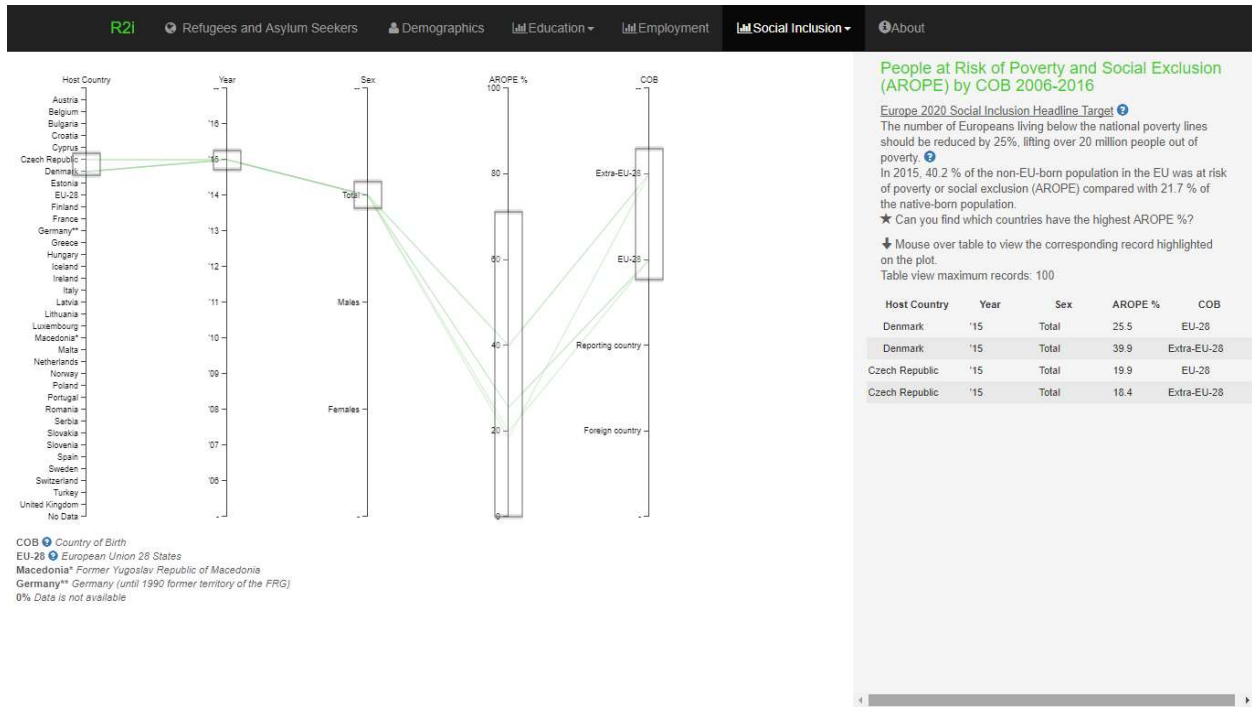


Figure 11. Comparing the population of Denmark in 2015, the people born outside EU-28 are at greater risk of poverty and social exclusion (39.9%) than people born in EU-28 countries (25.5%).

7 Next Steps

Developing a web-based tool is more than creating a button that will trigger the expected result; the success of the tool should also take the usability into consideration. With that, further enhancements could be done by adding more interactive items for the user to filter the data and manipulate the visualization. For example, adding a check box to sort the table view by column or a radio button to switch between refugee and asylum seeker datasets.

In addition, the data could be accessed by using the application programming interface (API) provided by UNHCR [24] instead of downloading the csv files from the website. The advantage of using the API is that whenever the data is updated by UNHCR, R2i will reflect the latest data automatically.

8 Acknowledgements

I would like to give special thanks to Jim Detwiler at Pennsylvania State University for his advice and mentorship. I would also like to thank the users who participated in the usability evaluation and give credit to the FOSS communities for all the amazing tutorials and examples of using the JavaScript libraries.

References

- Dobbs, L., & Clayton, J. (2015). Funding shortage leaves Syrian refugees in danger of missing vital support. *United Nations High Commissioner for Refugees*. Retrieved on November 16, 2016 from <http://www.unhcr.org/en-us/news/latest/2015/6/558acbbc6/funding-shortage-leaves-syrian-refugees-danger-missing-vital-support.html>
- European Commission (2013). Using EU Indicators of Immigrant Integration. *European Commission*. Retrieved on November 16, 2016 from [http://www.integration.ie/website/omi/omiwebv6.nsf/page/AXBN-A4LPWN18405825-en/\\$File/using%20eu%20indicators%20of%20immigrant%20integration.pdf](http://www.integration.ie/website/omi/omiwebv6.nsf/page/AXBN-A4LPWN18405825-en/$File/using%20eu%20indicators%20of%20immigrant%20integration.pdf)
- European Commission (2016). Europe 2020 in a nutshell. *European Commission*. Retrieved on November 16, 2016 from http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/index_en.htm
- European Economic and Social Committee (2004). Common Basic Principles for Immigrant Integration Policy. *European Economic and Social Committee*. Retrieved October 22, 2016 from http://www.eesc.europa.eu/resources/docs/common-basic-principles_en.pdf

Links

- [1] <https://github.com/johan/world.geo.json/tree/master/countries>
- [2] <http://popstats.unhcr.org/en/overview>
- [3] <http://ec.europa.eu/eurostat/web/migrant-integration/data/database>
- [4] <https://d3js.org/>
- [5] <https://github.com/square/crossfilter>
- [6] <https://dc-js.github.io/dc.js/>
- [7] <http://leafletjs.com/index.html>
- [8] <https://github.com/Intellipharm/dc-addons>
- [9] <http://www.openstreetmap.org/copyright>
- [10] <http://getbootstrap.com/>
- [11] <https://jquery.com/>
- [12] <http://personal.psu.edu/wxw5193/webserver/Evaluation.html>
- [13] <http://personal.psu.edu/wxw5193/webserver/R2iDashboard.html>
- [14] <https://github.com/wingyw/r2i>
- [15] <http://personal.psu.edu/wxw5193/webserver/Index.html>
- [16] <http://personal.psu.edu/wxw5193/webserver/demographic.html>
- [17] http://personal.psu.edu/wxw5193/webserver/edu_attainment.html
- [18] http://personal.psu.edu/wxw5193/webserver/edu_early_leaver.html
- [19] http://personal.psu.edu/wxw5193/webserver/edu_lifelong_learning.html
- [20] http://personal.psu.edu/wxw5193/webserver/edu_NEET.html
- [21] <http://personal.psu.edu/wxw5193/webserver/employment.html>
- [22] http://personal.psu.edu/wxw5193/webserver/social_AROPE_COB.html
- [23] http://personal.psu.edu/wxw5193/webserver/social_AROPE_COC.html
- [24] http://data.unhcr.org/wiki/index.php/API_Documentation