
*Visualization of sentiments towards Syrian Refugees using Twitter data, aimed at UN Officials and journalists for quick comprehension of countries that are more receptive towards refugees on social media*

Jonathan Grundy (jdg526@nyu.edu), Manushi Majumdar (mam1612@nyu.edu),
Maria Ortiz (meo342@nyu.edu), Xia Wang (xw876@nyu.edu)

Project repository: https://github.com/jgrundy/infoviz_refugee_project
Video: https://vimeo.com/167366472
Demo: http://jgrundy.github.io/infoviz_refugee_project/index.html

**What is the problem you want to solve and who has this problem?**

According to the UNHCR (United Nations High Commissioner for Refugees), there are close to 5 million "persons of concern"[1] that have been displaced from Syria to five neighboring countries (Turkey, Lebanon, Jordan, Egypt and Iraq) since 2013. Other sites place the total number of displaced Syrians is over 9 million, and they have fled to countries all over the world. European countries especially have seen a massive influx of migrants from Syria in the past 6 months

**Does the number of refugees in a country influence the local population's attitude towards them?**

Through our analysis, the team is assessing what are the prevailing sentiments towards Syrian refugees compared to population density of refugees per country. We are interested in seeing how people are reacting to this influx and seeing if people's reaction is correlated with the number of refugees arriving. This information could be relevant to international organizations such as UN Refugee Agencies, Amnesty International, countries' embassies, and refugee aid agencies.

[1] UNHCR http://data.unhcr.org/syrianrefugees/regional.php, last accessed May 16, 2016

**What are the driving analytical questions you want to be able to answer with your visualization?**

- Where are people tweeting about this problem? The objective is to locate countries that are tweeting about this issue.
- What are their general attitudes? Tweets will be grouped into three buckets depending on their sentiment score. We had initially thought of these values as hostile (values -1 to above 0), neutral (value equal to 0), welcoming (values above 0). After some analysis this was changed to have a wider scope, assigning welcoming from 0.01 and above, neutral from -0.01 - 0.01 and hostile from -0.01 and below.
- Does the number of refugees in a country substantially influence a country's perception of refugees on Twitter?

**Data used, description and sources**

| Attributes | Type | Description | Value Range/ Category | Derived |
|---|---|---|---|---|
| Country | Categorical | Name | None | No |
| Country Geometry | Categorical | Spatial geometry | None | No |
| Country Population | Quantitative | Number of people in country | 0 - 1.4 billion | No |
| Number of refugees per country | Quantitative | Proportion of refugees to established population | 0%-100% | Scaling |
| Tweet Location | Categorical | Spatial geometry | English-tweeting nations | No |
| Tweet Text | Categorical | String | None | Subsetting |
| Tweet Count | Quantitative | Number of tweets | 0 - inf | Aggregation |
| Sentiment | Ordinal or Divergent | Sentiment score | Color Hue | Binning |

**Sources:**

- Country geometry data: Mapbox[2]
- Country population: United Nations[3]

[2] MapBox https://www.mapbox.com/mapbox-studio/
[3] United Nations Data Tables http://data.worldbank.org/indicator/SP.POP.TOTL

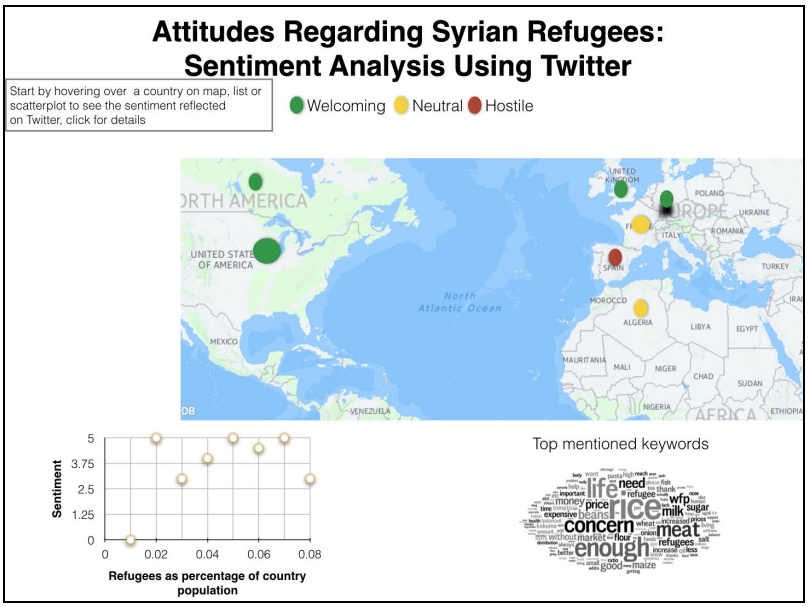
- Number of refugees per country: UNHCR[4], UNHCR Syrian Refugee Regional Response[5] and Tweet Location, Text, Count: Twitter API[6]
- Sentiment: Python TextBlob module[7]

**What have others done to solve this or related problems?**

- The UNHCR (United Nations High Commissioner for Refugees) has created a map displaying by country the number of refugees broken down into six categories: IDP (internally displaced persons), refugees, asylum seekers, returnees, stateless persons and others.[8]
- Research showing how to perform Twitter sentiment analysis in multiple different languages.[9]
- Questions different European countries have about Syrian refugees.[10]
- Attempts to map the locations, flows, and attitudes on refugees leaving Syria.[11]
- Google Trends “European Migrant Crisis”. Ranks countries by most searched places to move to, top questions related to immigration by country, and which countries are doing most searches on immigration.[12]
- Track migration flows around the world by country[13] and in an interactive map.[14]

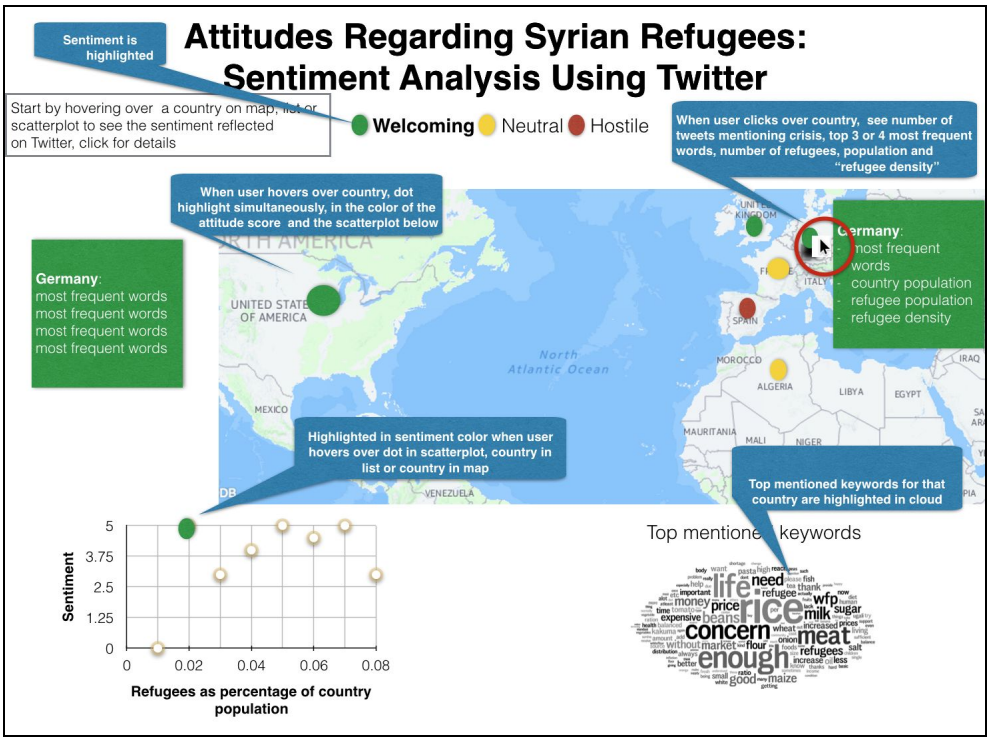
**Initial Mockup**

**Figure 1: Initial Mockup- Static Image**




[4] UNHCR Popstats http://popstats.unhcr.org/en/overview
[5] UNHCR Syria Regional Response Portal http://data.unhcr.org/syrianrefugees/regional.php
[6] Twitter API https://dev.twitter.com/streaming/overview
[7] Python Text Blob http://textblob.readthedocs.io/en/dev/
[8] UN Popstats http://popstats.unhcr.org/en/overview
[9] Technical University Berlin http://www.dai-labor.de/fileadmin/files/publications/narr-twittersentiment-KDML-LWA-2012.pdf
[10] Telegraph.co.uk http://www.telegraph.co.uk/news/uknews/immigration/11863452/Refugee-crisis-How-do-European-countries-attitudes-differ-on-refugees.html
[11] Wired.co http://www.wired.co.uk/news/archive/2015-09/11/europe-syria-refugee-crisis-maps
[12] Google Trends https://www.google.co.uk/trends/story/GB_cu_tru9jk8BAABEtM_en
[13] http://peoplemov.in/
[14] http://www.lucify.com/the-flow-towards-europe/

**Figure 2: Initial Mockup - Interactive Sketch**




**How to read the sketch**

Data is displayed in faceted view, linked highlighting shows where user is in all three sections, detail on demand when user clicks

- Map shows choropleth of “refugee density”
- Scatterplot below
- Aggregate keywords and quantity encoded in word cloud
- Sentiment is encoded in dot color hue

**How the design helps answer the questions**

- The map design makes it easy to locate where countries that are tweeting about this issue
- The map provides a snapshot of sentiments by country at a glance
- Relatable color encoding for Welcoming (green), Neutral (yellow), Hostile (red)
- Scatterplot can show correlation (or lack of) between country sentiment and number of refugees as percentage of population

**Project Update**

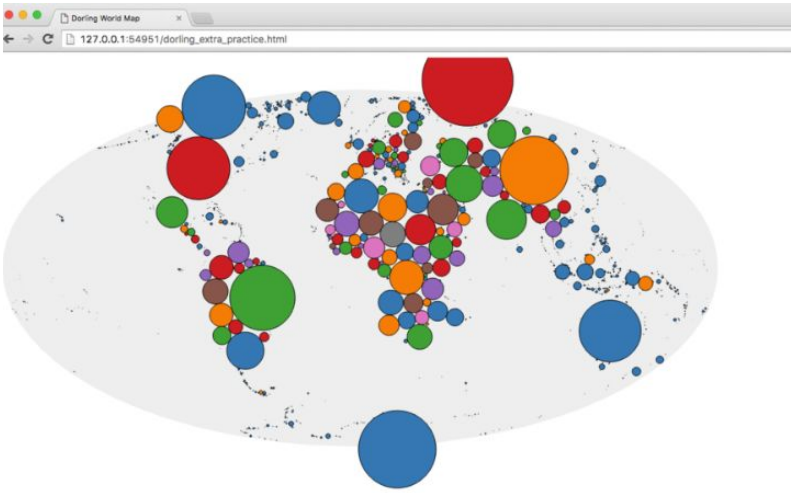
- Upon initial feedback and recommendation from Professor Bertini, we decided to use a Circular Dorling Cartogram. This technique uses an algorithm to draw circles that are scaled to the attribute size and plots them with spatial relation to a world map[15]. The premise is: “ if an algorithm is left to run on a normal map for a few hundred iterations it eventually produces a solution in which no circles overlap and as many as possible are still in close contact with their original geographic neighbours”, thus producing a world map.We had considered doing a

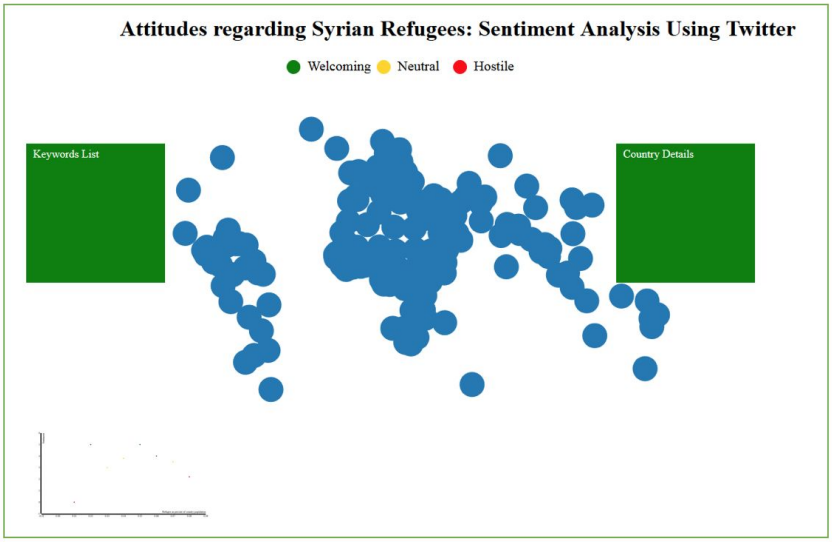
[15] Dorling, D. (1996). *Area Cartograms.* Bristol, England: Department of Geography, University of Bristol, England. Page 32.

choropleth map, however using the Dorling cartogram we can take advantage of both size and color to encode population size and and sentiment

- Eliminating the word cloud and instead using a list of top keywords by country. This was for two reasons- 1) we would not have enough words for a substantial cloud and 2) it did not encode additional relevant information
- We considered using three sentiments: welcome, neutral and hostile and thus three colors, green, yellow and red. This would ultimately give us a scatterplot that would not be particularly illustrative and using only three colors would not give enough nuance. In addition, these colors would not be good for color blindness, so we changed the scale to be blue (for welcoming), yellow and red.

**Figure 3: Screenshot of early work on Dorling Cartogram and Project Update**



## Final Visualization

Data Analysis

### I. Obtaining data on refugees by country

The first step was finding data on refugee flows. Detailed data was found via an inter-agency portal[16] that tracks refugee movements with a two month time delay for 5 countries: Egypt, Iraq, Jordan, Lebanon and Turkey.

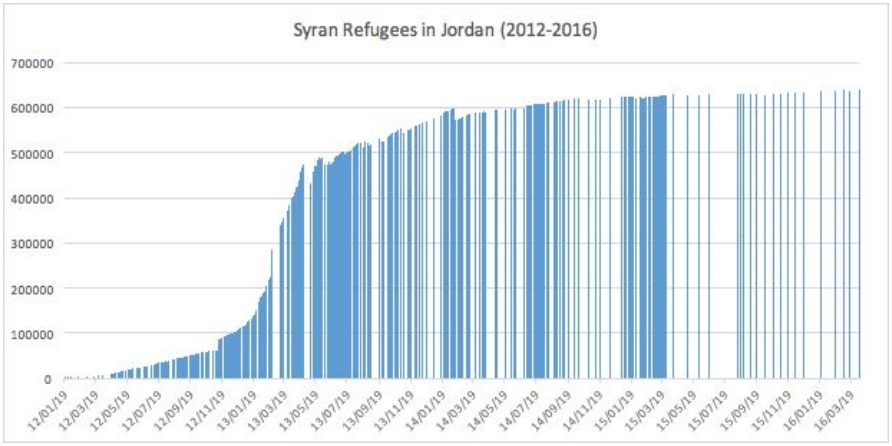
**Figure 4: Screenshot of raw data for countries in Excel**

I974

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | country | year | date | refugees | percent_pop | | |
| 963 | Turkey | 2011 | 11/12/16 | 8,000.00 | 0.00011% | | |
| 964 | Turkey | 2012 | 12/01/15 | 9,500.00 | 0.00013% | | |
| 965 | Turkey | 2012 | 12/02/29 | 10,658.00 | 0.00014% | | |
| 966 | Turkey | 2012 | 12/04/18 | 23,955.00 | 0.00032% | | |
| 967 | Turkey | 2012 | 12/04/23 | 23,702.00 | 0.00031% | | |
| 968 | Turkey | 2012 | 12/04/25 | 29,441.00 | 0.00039% | | |
| 969 | Turkey | 2012 | 12/04/26 | 29,082.00 | 0.00038% | | |
| 970 | Turkey | 2012 | 12/05/02 | 29,037.00 | 0.00038% | | |
| 971 | Turkey | 2012 | 12/05/07 | 28,750.00 | 0.00038% | | |
| 972 | Turkey | 2012 | 12/05/21 | 28,517.00 | 0.00038% | | |
| 973 | Turkey | 2012 | 12/06/01 | 30,172.00 | 0.00040% | | |
| 974 | Turkey | 2012 | 12/06/04 | 31,986.00 | 0.00042% | | |
| 975 | Turkey | 2012 | 12/06/12 | 33,110.00 | 0.00044% | | |
| 976 | Turkey | 2012 | 12/06/22 | 37,839.00 | 0.00050% | | |
| 977 | Turkey | 2012 | 12/06/25 | 38,085.00 | 0.00050% | | |
| 978 | Turkey | 2012 | 12/06/26 | 38,784.00 | 0.00051% | | |
| 979 | Turkey | 2012 | 12/07/04 | 40,804.00 | 0.00054% | | |
| 980 | Turkey | 2012 | 12/07/05 | 41,270.00 | 0.00054% | | |
| 981 | Turkey | 2012 | 12/07/12 | 43,058.00 | 0.00057% | | |
| 982 | Turkey | 2012 | 12/07/20 | 48,387.00 | 0.00064% | | |
| 983 | Turkey | 2012 | 12/07/23 | 49,092.00 | 0.00065% | | |
| 984 | Turkey | 2012 | 12/07/31 | 49,893.00 | 0.00066% | | |
| 985 | Turkey | 2012 | 12/08/01 | 49,743.00 | 0.00066% | | |

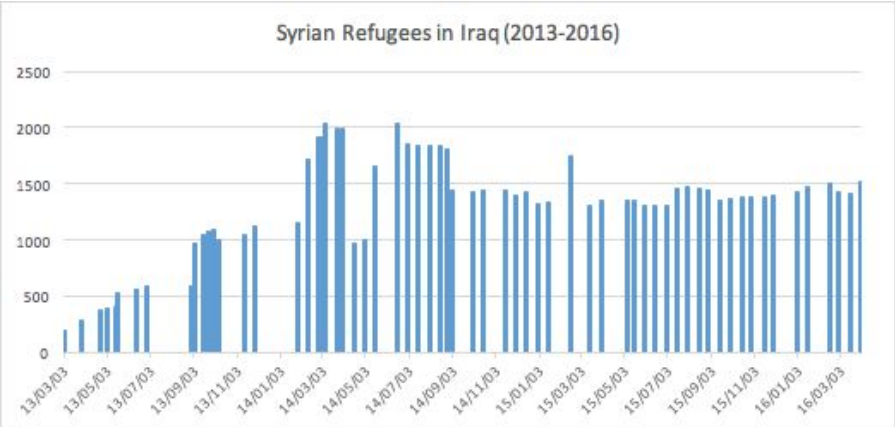
In **Lebanon**, refugee counts reached a peak in September 2014 of almost 300,000, and the latest number for March 2016 is 256,126. For **Iraq**, during 2014 the numbers fluctuated between 2000 and 1000, with 1,508 being the latest number. For **Turkey**, numbers have climbed steadily from 174,598 at the beginning of 2013 to 2,748,367 in March of this year.
**Egypt** has 199,556 refugees as of March of this year, although their count peaked in November of 2014 with 138,543 refugees.
For **Jordan,** the refugee population remained constant as of March of 2015 and with a latest count of 642,862.

**Fig 5: Refugee counts for Jordan**




[16] Syria Regional Refugee Response, Inter-agency Information Sharing Portal

**Fig 6: Refugee counts for Iraq**




The graphs above point out that the numbers are very disparate for each country, which is why we decided to use a proportion of refugees to total population to try to make this more comparable.

The UNHCR tracks different classifications of "persons of concern":

- Refugees, which are individuals that have been officially granted that status
- Asylum seekers which are individuals seeking refugee status but have not been recognized
- Internally displaced persons (IDPs) which are individuals who have left their homes because of situations of generalized violence and violations of human rights
- "Stateless persons" who are "not considered as nationals by any State on under the operation of the law"[17].

The most robust data in terms of refugee flows from Syria to the rest of the world is up to 2013. As of 2013, an inter-agency task force was formed to closely track the regions that were seeing the greatest influx- Jordan, Iraq, Egypt, Lebanon and Turkey. From 2013 onwards, data is only available for people who have applied for refugee status. For purposes of this analysis, we considered asylum seekers as refugees. For 2016, data specifically concerning Syrian asylum seekers is only available for 28 countries.

## II. Data collection from Twitter and preparation

### 1. Twitter Scraping

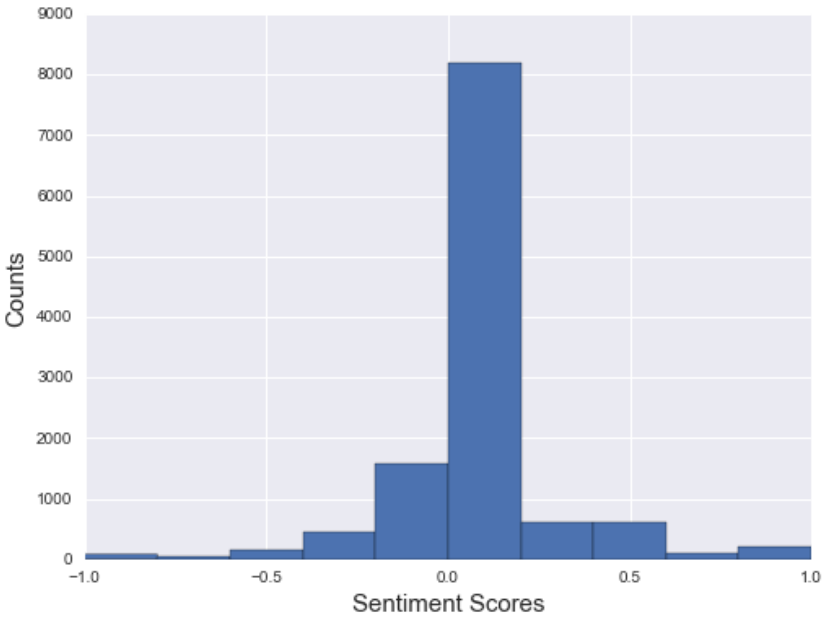
Despite the fact several studies and projects picked up the topic of the tweets about the Syrian refugees, no publicly available data is found. Therefore, the team decided to scrape data from Twitter and explore the Twitter users' attitude towards the Syrian refugees based on this data.

There are two Twitter APIs which are possible choices for the scraping task: the search API and the public streaming API. The difference between the two APIs for this project lies in that the search API has a rate limit for every 15 minutes, while the public streaming API doesn't not. In the end, the public streaming API is chosen to scrape the data since more data can be pulled for the same amount of time the scraper is running. The project has been structured in such as way as to allow for future expansion from using historical data to a real-time feed.

[17] UNHCR Popstats http://popstats.unhcr.org/en/overview

The public streaming API provides a keyword query functionality so that API users can filter the content of the tweets they would like to search. For this project we used 4 keywords: *syria*, *refugee*, *syrian* and *migrant crisis* to select the more relevant tweets.

**Figure 7: Histogram of the tweet Sentiment Analysis Scores**




The scraper created a text file of 5 hours of tweets in json format, encompassing a total of 12062 geotagged tweets.

The tweets in json format are then transformed into a CSV file with columns for the tweet content, the sentiment analysis score and the country where that tweet was published. For the sentiment analysis, we used the Python TextBlob module to parse each tweets and return a score ranging from -1 to 1. A score of, -1 indicates extremely negative sentiments expressed and 1 indicates extremely positive sentiments. As can be observed in Figure 7, more than 8000 tweets lie within the 0.0 to 0.16 bucket among all the sentimental scores. There are 2 caveats here. First, sentiment analysis is a two-step process: the TextBlob module would first decide the subjectivity of the tweet content, and only the subjective sentences are assigned a sentiment score; the objective sentences would be assigned 0 value directly. Second, TextBlob might not have a vocabulary large enough to understand all the tweets. So some tweets would be given 0 value due to the fact that TextBlob cannot parse them. In all, the number of 0 values might be higher than the 'true' distribution, but we assume it would not affect our later analysis.

To acquire the country of the tweet, we used the tweet's latitude and longitude and used google maps API to locate the country, and attached the name of the country to the tweets in the end. A snapshot of the processed twitter data is shown in Figure 7.

**Figure 8: Screenshot of raw Twitter data with sentiment score**

| | | country | senti_score | tweet_content |
|---|---|---|---|---|
| 2 | 0 | Canada | -0.3 | Disturbing footage shows horrific reality for Syrians under attack by their own government. #Daraya #Syria |
| 3 | 1 | Canada | -0.4 | A chess set from cheese cartons. A game of survival. Tonight BBCRadio4 20 BST \"Held Hostage in #Syria\u2026 htt\u20 |
| 4 | 2 | Canada | 0 | May Allah curse Bashar al Asad & his allies! Another massacre in #MaraatAlNouman, #Syria just minutes ag\u2026 |
| 5 | 3 | Spain | -0.3 | Disturbing footage shows horrific reality for Syrians under attack by their own government. #Daraya #Syria |
| 6 | 4 | Denmark | -0.5 | Absurd at h\u00f8re #dkpol \u00e6vle om at IS er den st\u00f8rste trussel, n\u00e5r en kvart mill borgere er myrdet af |
| 7 | 5 | USA | 0 | #Iran:Ex-governor of #yazd injured in #Syria war |
| 8 | 6 | Canada | 0.375 | Rich gulf Arab nations\u2019 refugee response questioned |
| 9 | 7 | Thailand | 0 | #ndtv Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria: Air strikes on a vegetable mark... |
| 10 | 8 | Canada | 0 | \u062f\u062e\u0648\u0644 \u0644\u0633\u064a\u0627\u0631\u0627\u062a \u0627\u0644\u0623\u0645\u0645 \u0 |
| 11 | 9 | Canada | -0.3 | Disturbing footage shows horrific reality for Syrians under attack by their own government. #Daraya #Syria |
| 12 | 10 | India | 0 | Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria: Air strikes on a vegetable ... |
| 13 | 11 | Canada | 0 | @Ali_Husain333 @Nandianehat Ali is made because of my anti radical position in Syria. |
| 14 | 12 | Canada | -0.2 | Four Afghan fighters part of the IRGC Fatimiyoun Brigade recently killed in #Syria |
| 15 | 13 | Canada | 0 | Obama Dismisses 'Some Abstract Notion of Red Lines' in Syria |
| 16 | 14 | India | 0 | Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria |
| 17 | 15 | Canada | 0 | #FollowMe #F2F #FF Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria... |
| 18 | 16 | USA | 0 | #Jordan recalls its #Iran ambassador |
| 19 | 17 | India | 0 | Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria: Air strikes on a vegetable market in opposition |
| 20 | 18 | India | 0 | Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria: Air strikes on a vegetable market in ... |
| 21 | 19 | Israel | -0.1 | Manpads are no use against mod Russian Jets & Helicopters: Counter measures invented longtime ago: Useless in S |
| 22 | 20 | UK | 0.3 | Our own @Zia_Salik discussing the current situation in Syria 5 years on from start of the conflict #ANightInAleppo htt\u2 |
| 23 | 21 | Canada | 0 | LOLOLOL!!! #ReleasetheTranscripts #Haiti #Aristide #Honduras #BertaCaceres #Iraq #Syria #Libya #\u2026 htt\u2026 |
| 24 | 22 | India | 0 | Air Strike On Market Kills Around 40 In Opposition-Held Northwest Syria |
| 25 | 23 | Canada | -0.3 | Disturbing footage shows horrific reality for Syrians under attack by their own government. #Daraya #Syria |
| 26 | 24 | Riyadh Saudi | 0.14444444 | Air strike in Syria against the morning market in Kafranbel \u2013 as you can easily see, a clear military target |
| 27 | 25 | Canada | 0 | Syria\u2019s Saudi-backed opposition says Delegation leaders are leaving Geneva |

A preliminary look at the data obtained from Twitter scraping showed we would not cover all the countries we had refugee data for, most notably Jordan and Lebanon which are two of the five countries tracked by the Syrian Refugee Response site. We decided to go ahead and use the countries we had refugee and Twitter data for, in order to avoid a temporal mismatch. This is a common challenge when looking at Twitter data by location, as only around 1%[18] of all tweets tend to have a geotag that indicated their latitude and longitude. Due to our scraping method, we did not store the non-geotagged tweets. Total number of tweets used was 10,345 and we obtained tweets from 85 countries. Out of these 85, we had 2016 refugee data for 21 countries.

### 2. Assigning sentiment score to Twitter data

The tweets are then assigned a sentimental score by the Python package TextBlob sentiment analyzer. The scores and the tweet texts can be seen in Figure 8. From this we wanted to see how many tweets were emitted by country and group them into positive, neutral, and negative tweets.

**Figure 9: Screenshot of tweet classification**

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | Count of country | Column Labels | | | | |
| 4 | Row Labels | Hostile | Neutral | Welcoming | (blank) | Grand Total |
| 5 | 2000 | 16 | 58 | 16 | | 90 |
| 6 | Afghanistan | 7 | | | | 7 |
| 7 | Africa | 14 | 9 | 2 | | 25 |
| 8 | Argentina | | 1 | | | 1 |
| 9 | Armenia | | 9 | 3 | | 12 |
| 10 | Australia | 24 | 71 | 28 | | 123 |
| 11 | Austria | 5 | 10 | 4 | | 19 |
| 12 | Azerbaijan | 1 | 6 | 2 | | 9 |
| 13 | Bangladesh | 1 | | 1 | | 2 |
| 14 | Belarus | 17 | 50 | 7 | | 74 |
| 15 | Belgium | 12 | 18 | 12 | | 42 |
| 16 | Bermuda | 1 | | | | 1 |
| 17 | Bosnia and Herzegovina | 2 | 5 | 2 | | 9 |
| 18 | Botswana | | 1 | | | 1 |
| 19 | Brazil | 9 | 26 | 14 | | 49 |
| 20 | Bulgaria | 1 | 3 | | | 4 |
| 21 | Burundi | 15 | | | | 15 |
| 22 | Cameroon | 1 | 11 | 1 | | 13 |
| 23 | Canada | 1127 | 3086 | 1171 | | 5384 |
| 24 | China | 14 | 48 | 16 | | 78 |
| 25 | Colombia | 1 | 10 | 2 | | 13 |

Using excel pivot tables we were able to make this grouping, as well as obtaining the average sentiment.

Quora

https://www.quora.com/What-percentage-of-tweets-are-geotagged-What-percentage-of-geotagged-tweets-are-ascribed-to-a-venue

**Figure 10: Final data table**

| | A | B | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Country | Pop2014 | Asylum2016 (data for 21 countries) | Average of senti_sco | hostile_tweets | neutral_tweets | welcoming_tweets | total_tweets | Count of country | % ref to pop asylum 2016 |
| 2 | Germany | 80970732 | 27146 | 0.03593007 | 17 | 46 | 18 | 81 | 81 | 3.3525694E-04 |
| 3 | Austria | 8545908 | 1272 | -0.0241427 | 5 | 10 | 4 | 19 | 19 | 1.4884317E-04 |
| 4 | Sweden | 9696110 | 1065 | 0.07786759 | 5 | 69 | 28 | 102 | 102 | 1.0983786E-04 |
| 5 | Switzerland | 8188102 | 476 | 0.00636498 | 20 | 73 | 20 | 113 | 113 | 5.8133130E-05 |
| 6 | Belgium | 11231213 | 508 | -0.0026 | 12 | 18 | 12 | 42 | 42 | 4.5231089E-05 |
| 7 | Bulgaria | 7223938 | 205 | -0.0375 | 1 | 3 | | 4 | 4 | 2.8377874E-05 |
| 8 | Finland | 5461512 | 66 | -0.0918669 | 4 | 7 | 3 | 14 | 14 | 1.2084566E-05 |
| 9 | Serbia | 7129366 | 79 | -0.0246868 | 20 | 39 | 8 | 67 | 67 | 1.1080929E-05 |
| 10 | France | 66217509 | 397 | 0.02872402 | 43 | 106 | 52 | 201 | 201 | 5.9953932E-06 |
| 11 | Ireland | 4615693 | 20 | 0.04825103 | 17 | 46 | 18 | 81 | 81 | 4.3330438E-06 |
| 12 | Spain | 46476032 | 151 | -0.0038552 | 21 | 101 | 9 | 131 | 131 | 3.2489865E-06 |
| 13 | United Kingd | 64559135 | 177 | 0.11483414 | 134 | 612 | 298 | 1044 | 1044 | 2.7416724E-06 |
| 14 | Canada | 35543658 | 94 | 0.02415157 | 1127 | 3086 | 1171 | 5384 | 5384 | 2.6446349E-06 |
| 15 | Slovenia | 2061980 | 1 | -0.0170918 | 12 | 29 | 8 | 49 | 49 | 4.8497076E-07 |
| 16 | Romania | 19904360 | 6 | 0.02320611 | 11 | 72 | 8 | 91 | 91 | 3.0144149E-07 |
| 17 | Croatia | 4238389 | 1 | 0 | | 1 | | 1 | 1 | 2.3593870E-07 |
| 18 | United State | 318857056 | 70 | 0.00138978 | 148 | 465 | 140 | 753 | 753 | 2.1953411E-07 |
| 19 | Australia | 23470118 | 2 | 0.0208098 | 24 | 71 | 28 | 123 | 123 | 8.5214740E-08 |
| 20 | Poland | 38011735 | 3 | 0.07142857 | 1 | 3 | 3 | 7 | 7 | 7.8922996E-08 |
| 21 | Turkey | 75932348 | 2 | -0.0422951 | 57 | 135 | 15 | 207 | 207 | 2.6339236E-08 |
| 22 | Japan | 127131800 | 1 | 0.00776768 | 8 | 87 | 5 | 100 | 100 | 7.8658526E-09 |

## III. Creating visualization and binding data

- Map created with Mapbox + d3 + leaflet
- Added color encoding to countries with refugee counts using the color purple, with the intensity of the hue denoting a greater population of refugees.
- Using tooltip to create cards for a list of the key words
- Creating scatter plot and pie chart with global synthesis of tweets and sentiment scores

## IV. Visualization results

- **Countries with highest refugee counts in our visualization: Germany and Bulgaria**
- **Overall sentiments on Twitter are Welcoming.** Sentiments**:** of all the tweets, a similar proportion are Neutral and Hostile, around 28%, Welcoming tweets are 43%. Our classification thresholds were widened slightly, assigning Welcoming from 0.01 and above, Neutral from -0.01 - 0.01 and Hostile from -0.01 and below.
- **Sentiments follow a loosely discernible regional pattern.** There is a “welcoming cluster” around Germany and a negative cluster formed by Turkey, Bulgaria and Serbia.
    - **Welcoming: United Kingdom, Sweden, Poland, Germany, Ireland, France, Romania, Canada and Australia.**
    - **Neutral: United States, Japan, Croatia, Spain, Switzerland and Belgium**
    - **Hostile: Slovenia, Serbia, Bulgaria, Turkey, Austria and Finland**
- **Number of refugees is not correlated to sentiment.** Germany that had the highest refugee count also had positive sentiments, and others with low numbers had both positive and negative.

**Figure 11: World Map, main viz**

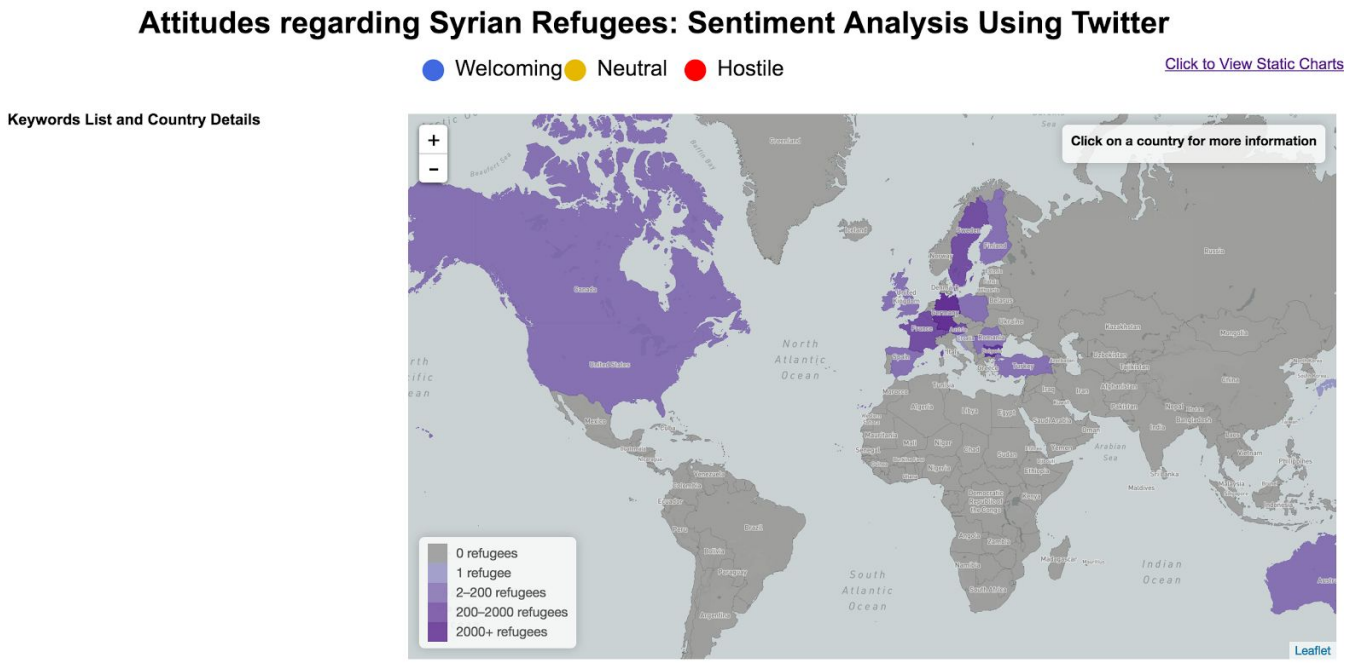



**Figure 12: "Negative cluster"**

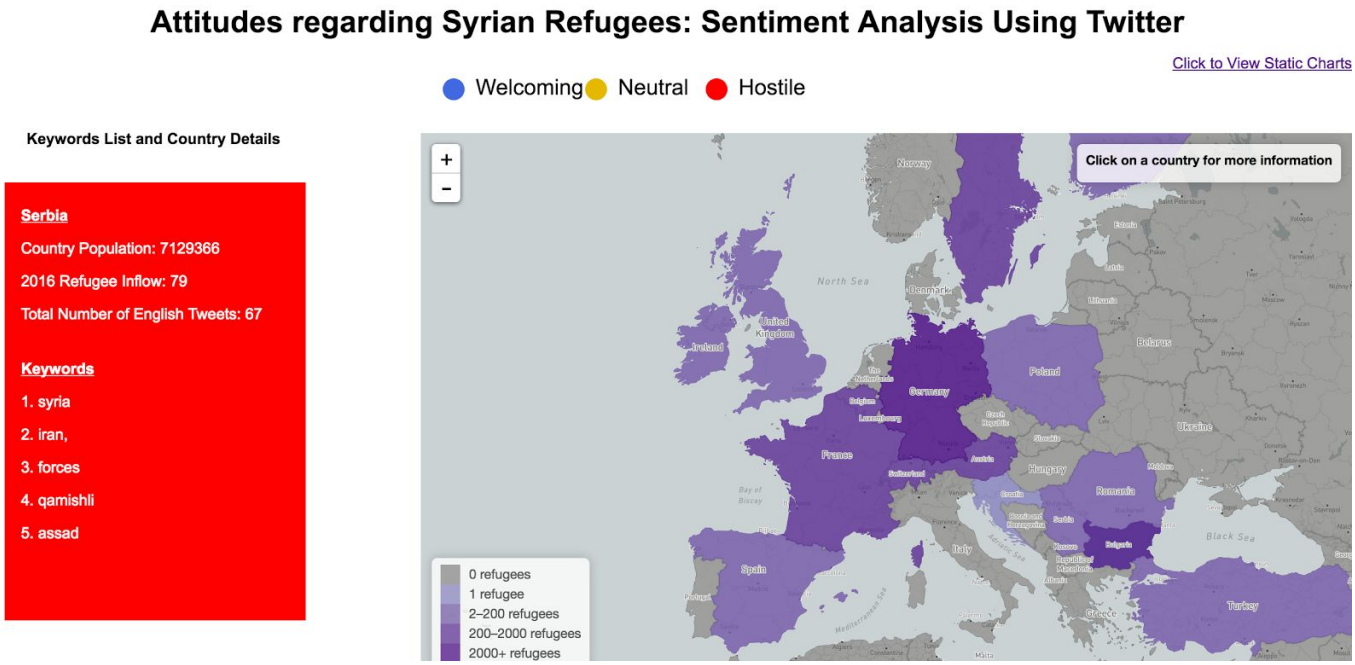
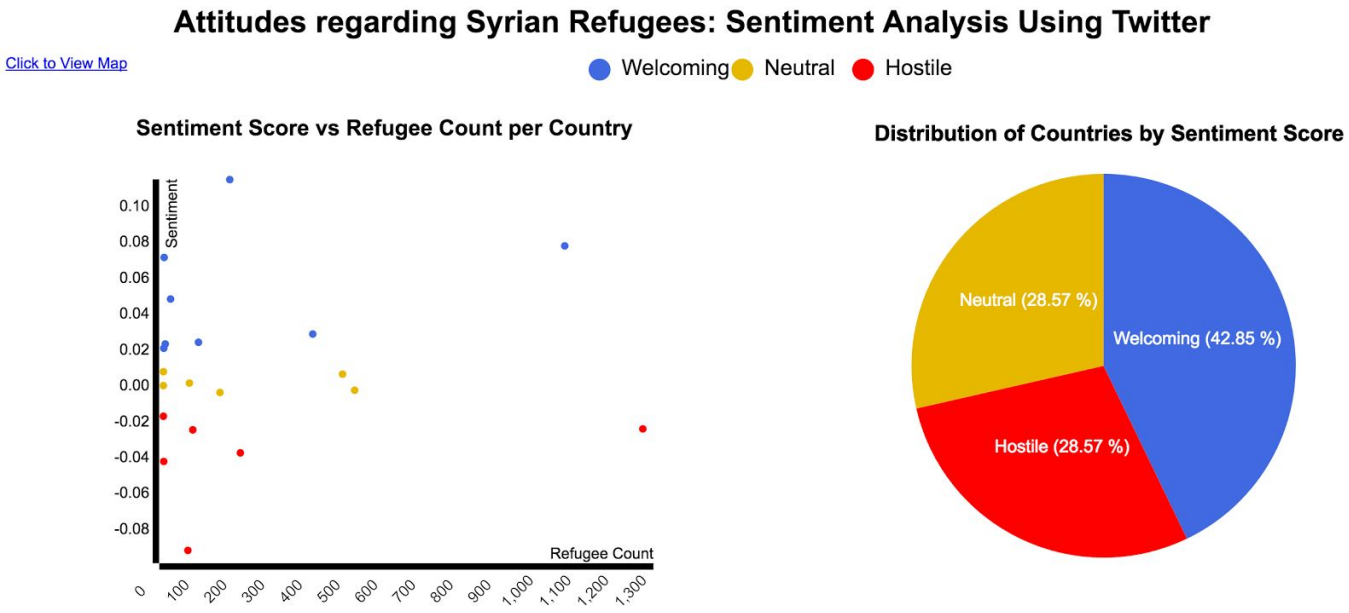


**Figure 13: Scatter plot and pie chart**




NOTE: Germany being an outlier taking 27146 Syrian refugees so far in 2016, has been deliberately excluded from the scatter plot.

## Limitations and Future Works

Visualization:

- We discovered the Dorling Cartogram would have only been useful if data on refugees would have been available for all countries or at least all the countries that we obtained tweets from.
- For future work, we would be like to be able to pull in a live stream of Twitter data and have the choropleth map adjust on a daily basis.
- In addition, the map could be linked to update the countries for which information on Syrian refugees is available as it comes online, pulling data directly from the UNHCR portal.

Sentiment Analysis:

- As was mentioned before, the TextBlob module may mistakenly assign a value of 0 (neutral) to certain tweets it “considers” objective, or because of a lack of vocabulary.
- We were limited to doing the analysis on English tweets. Future work could include doing sentiment analysis on some of the major languages to get a more accurate depiction.
- The project has been set up to allow use of a real-time Twitter feed.

Data:

- As our data for 2016 numbers of refugees was not accurate, a correlation analysis showed the number of refugees explained very little about the sentiment score, and this can also be seen in our visualization. With accurate, updated counts, we would be able to to see if there was a correlation between the sentiments towards refugees expressed on Twitter and the amount of refugees offered official resettlement aid within that country.