Attitudes Regarding Syrian Refugees: Sentiment Analysis Using Twitter

Project Update

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

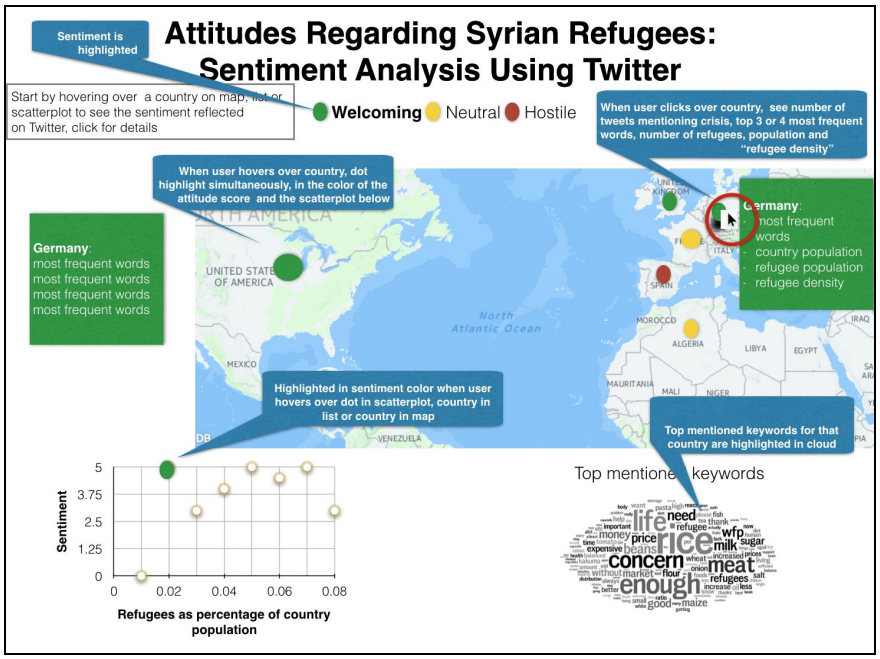
Describe how you addressed the main issues raised in the feedback you received in your project proposal:

- Map was judged as adequate tool, however the Dorling cartogram was seconded as a better option
- Twitter data: we had considered scraping tweets with a general scraper, but then decided to use the Twitter streaming API. The scraper ceases to function every two hours, but we are working on this issue.

If you changed your analytical questions provide the new list of questions here:

- Our questions have not changed, however, we are thinking more in depth about who our target audience is and how could they effectively use this visualization.
- Given that the data we have found on Syrian refugees is disparate, we are considering how to best integrate this. For example, there is detailed information on how the number of Syrian refugees has changed over time for 5 countries (Egypt, Iraq, Jordan, Lebanon, Turkey) and this is updated on bi-weekly or monthly basis, and we have data as current as April of this year. For other countries, we only have one aggregate figure from 2014, which creates a mismatch in the temporality of our twitter data versus the actual refugee population.
- We are encountering that the geotagged tweets (1-3% is the cited proportion of geotagged tweets compared to total volume) do not correspond to our countries of interest, so the historic data may not be useful in this case

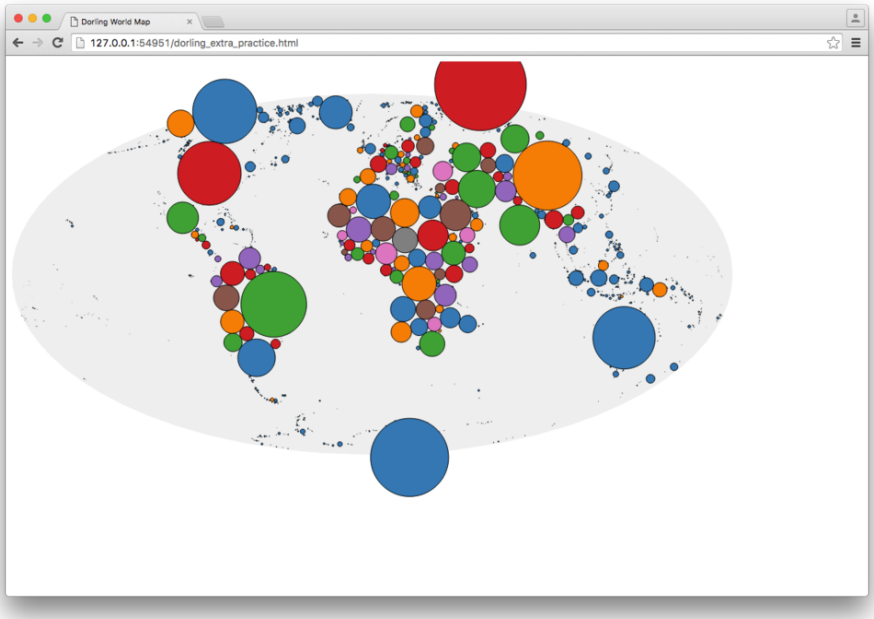
If you changed your visualization, describe what you changed and how:
First iteration:



Changes agreed upon:

- 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 had considered doing a choropleth map, however using the Dorling cartogram we can take advantage of both size and color to encode population size and and sentiment.
- We had considered using three sentiments: welcome, neutral and hostile and thus three colors. This would ultimately give us a scatterplot that would not be particularly illustrative and using only three colors would not give enough nuance.

Add one or more images of the current implementation of your project:



Explain how to read them:

Size represents population of the country, color represents sentiment. We plan to use at least five color to represent the sentiment, and still have to decide on a color pallet.

Describe what else needs to be done until the end of the project (e.g., what is missing to complete the project):

- Fix our twitter scraper
- Filter data for countries for which we have geotagged tweets
- Finish preparing dorling cartogram
- Add other visual elements to the final page and link everything together

Github link: https://github.com/NYU-CS6313-SPRING2016/Group-1-INET-Twitter-Human-Rights