# Bridging the Gap between Data versus Technology Producers: An Interactive Visual Interface for Data Exploration

**M. Abounabhan, F. K. Abu Salem, K. Antar, S. Elbassuoni, W. Hojeily, S. Najem**
Computer Science Department, American University of Beirut, National Council for Scientific Research (CNRS)
P. O. Box 11-0236, Riad El Solh, Beirut 1107 2020, Lebanon, Riad al Solh, Beirut 1107 2260, Lebanon
mea32@mail.aub.edu,fa21@aub.edu.lb,kha13@mail.aub.edu,se58@aub.edu.lb,weh01@mail.aub.edu, snajem@cnrs.edu.lb

**Abstract**

With the accumulation of large amounts of vital information surrounding the refugee communities in Lebanon, the current operational apparatus is facing unprecedented challenges in making sense of the deluge in data. In the majority of those instances the governmental sector in Lebanon is still largely lacking in the information systems expertise required to extract knowledge from this data in order to inform policy making. In this paper, we present an open, interactive visual interface to assist non-technical users in the exploratory data analysis of various types of data that are of vital significance. The intended purpose is to reduce the gap between producers of data and users of technology, and encourage evidence-based strategies for the refugee crisis management. In contrast to existing open visual platforms, our tool requires minimal technical expertise and effort, to the extent that not a single download or installation is needed. It also provides a user-friendly frontend and connects the user to data visualisation and analytics tools in the backend developed using Python. Our case studies reveal interesting revelations as a result of temporal and spatial phenomena, following the Syrian war and the influx of refugees into Lebanon.

**Keywords:** Visual Interface, Exploratory Data Analysis, Primary Care Health Data, Refugee Crisis

## 1. Introduction and Background

The status quo surrounding data in the Middle East and North Africa (MENA) region is that there is a general lack of studies and information around the current humanitarian crises following the widespread wars. In light of current events surrounding the Syrian crisis, there has been an uptake in data-driven research in the Middle East specifically as it is an issue of global concern. The MENA region is traditionally known to be a data-desert where not enough data is gathered, or data is not gathered in proper form, or at best, data is gathered in proper form but not efficiently used. The regional instabilities have exposed a large fraction of the population to food and nutrition insecurity in addition to undernutritions comorbidities, and numerous communicable and non-communicable diseases (Taleb et al., 2015; Hwalla et al., 2016). Thus, the internally displaced individuals and refugees have become a burden of the countries respective economies and infrastructures. The need for producing evidence-based strategies to face the emerging sociodemographic changes has become pressing (Breisinger et al., 2012). With the accumulation of large amounts of vital information surrounding the refugee communities in Lebanon, the current operational apparatus is facing unprecedented challenges in making sense of the deluge in data, mostly spread out across various uncentralised sources. In the majority of those instances, the governmental sectors are still largely lacking in the information systems expertise required to extract knowledge from this data to inform policy making, and there seems to be a need to develop scalable and efficient frameworks within reach of users of various levels of technical expertise (Koliopoulos et al., 2015).

Lebanon has still been reeling from the devastation incurred by its own civil war when the war in neighbouring Syria broke out. The operational staff in its ministries are stretched beyond limits and do not seem to be equipped to hop on the data analytics revolution overtaking the developing world. Concomitantly, the lack of a political will to invest in data-intensive research for managing the refugee crisis in Lebanon leads us to think that a minimalist, technically non-invasive platform is more likely to encourage owners of data to begin thinking along data-driven lines. The global burden of disease[1] is a massive initiative that provides a tool to demonstrate health loss from hundreds of diseases, injuries, and risk factors; however, this is not a platform which native owners of data can use at their whim, and personalise the way they see fit. RapidMiner (Fischer et al., 2002), Knime (Berthold et al., 2009), Weka (Hall et al., 2009) and Gephi (Bastian et al., 2009) are some of the leading platforms with the highest potential for scalable big data analytics. These visual tools allow code free mining of big data, but they still require non-trivial installations and a learning curve before one is able to use them. This may be a hindrance for people with extremely limited quantitative skills, who also have no access to technical support within their working environment.

In this paper, we present an open, interactive visual web interface to assist non-technical users in the exploratory data analysis of various types of data that are of vital significance. The intended purpose is to reduce the gap between producers of data and users of technology, and encourage evidence-based strategies for the refugee crisis management. In contrast to existing open visual platforms, our tool requires minimal technical expertise and effort, to the extent that not a single download or installation is needed. It also provides a user-friendly frontend and connects the user to data visualisation and analytics tools in the backend developed using Python. Our case studies reveal interesting revelations as a result of temporal and spatial phenomena, following the Syrian war and the influx of refugees

[1] `http://www.healthdata.org/gbd`

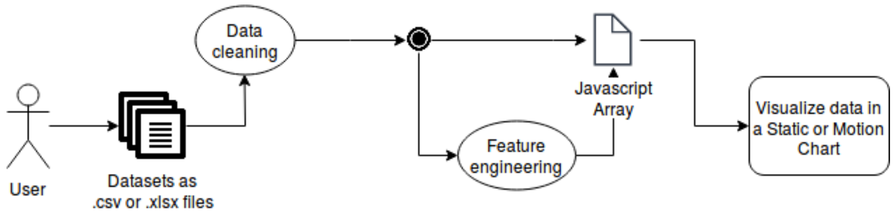



Figure 1: System Architecture

into Lebanon. These revelations become essential if one is to embark on predicting demand on the primary health care system, for example, assuming the crisis will persist. A plethora of work in the literature tackles geospatial and temporal visualizations of social data relevant to the MENA region (see for example, (Ghanem et al., 2014; Jänicke et al., 2013)). They rely on online textual and social media open data. Our work stands out by exploring publically available and private datasets that have been collected surrounding vital population statistics.

The rest of our paper is organised as follows. We first describe the design process and the software prototype information using a system's architecture. We also describe the data formats that our tool is able to process, as well as the anomalies in the data that are permissible and which our tool is able to clean up automatically. We further describe the pivotal datasets showcasing vital information relating to refugees in Lebanon, and some feature engeering we performed to make sense of the time series data representing demand on the primary health sector in Lebanon. Some of the datasets are publicly available – for example, UNHCR's persons of concern[2] and the global terrorism data [3]. One other dataset provided by the Ministry of Public Health in Lebanon cannot be made public. We finally present insights from the data that our interface is able to munge, highlighting the temporal and spatial dependencies of the observations found in the tackled datasets. Particularly, we are interested in the temporal effects of the Syrian refugee influx and its effect on the primary health care centers in Lebanon, as well as the spatial correlations existing between locations of major traumatic events and primary health care centers of high demand. Indeed, these basic manipulations that any novice user can perform using our web interface confirm the authors' viewpoint that any future forecasting of the demand on Lebanese infrastructure should be tackled as a spatiotemporal phenomenon. Readers can interact with our tool from here: `http://104.238.170.191/`.

## 2. The System

Figure 1 depicts the system architecture for our interface. Our system takes as input a dataset to be analysed, which typically consists of one or more comma-separated variable (csv) files or Excel (xls) files. The dataset is then passed through a data cleaning module, which detects anomalies in the data that are automatically eliminated. The cleaned data is then passed through a feature engineering module that extracts temporal and spatial features in the data, which can then be used to perform visual data analytics using various modes such as motion (dynamic) or static charts.

The frontend of our interface uses the Flask microframework to link the frontend to Python. Under the "Graph" option, "Browse" is a frontend option using JavaScript, and "Use Data" is a backend option using Python. Under the "Patient" option, "Browse" and "Data Sample" are both backend options using Python. "Static Visuals", seen in Figure 2, is also backend using Python. More on these options follow below as we describe our case studies.

We explain each module (component) in our system separately next.

### 2.1. Input Files

Our system assumes the input to be one or more csv (or xls) files. The columns in the files correspond to attributes such as dates, geographical locations, gender, age, etc. Each row corresponds to a data instance. Figure 3 shows a dummy snapshot representing what the Ministry of Public Health dataset looks like[4].

### 2.2. Data Cleaning

The data cleaning process accounts for anomalies such as null or missing values in numerical attributes. Currently, we perform two basic data cleaning operations to deal with such anomalies, depending on the type of attribute. For categorical attributes like gender or nationality, for example, we replace a null or missing value with a zero and for continuous attributes, we replace the missing value with the average value of all the categories. In future releases, we plan to utilise third-party data cleaning tools such as OpenRefine[5] or DataWrangler[6] to perform further cleaning operations such as data standardisation and duplicate elimination.

### 2.3. Feature Engineering

The data imported to the "Patients" page goes a step further past cleaning; it undergoes feature engineering process to generate additional features that are needed for spatiotemporal exploration. For instance, we use gazetteers to generate geo-coordinates of locations in the dataset that can be used to perform spatial analysis on the datasets as we explain in our second case study.

[2] Available from `popstats.unhcr.org/en/persons_of_concernpopstats.unhcr.org/en/persons_of_concern.csv`

[3] Available from `http://apps.start.umd.edu/gtd/downloads/dataset/globalterrorismdb_0616dist.xlsx`

[4] For privacy and data sharing agreements, this dataset cannot be revealed

[5] `http://openrefine.org/`

[6] `http://vis.stanford.edu/wrangler/`

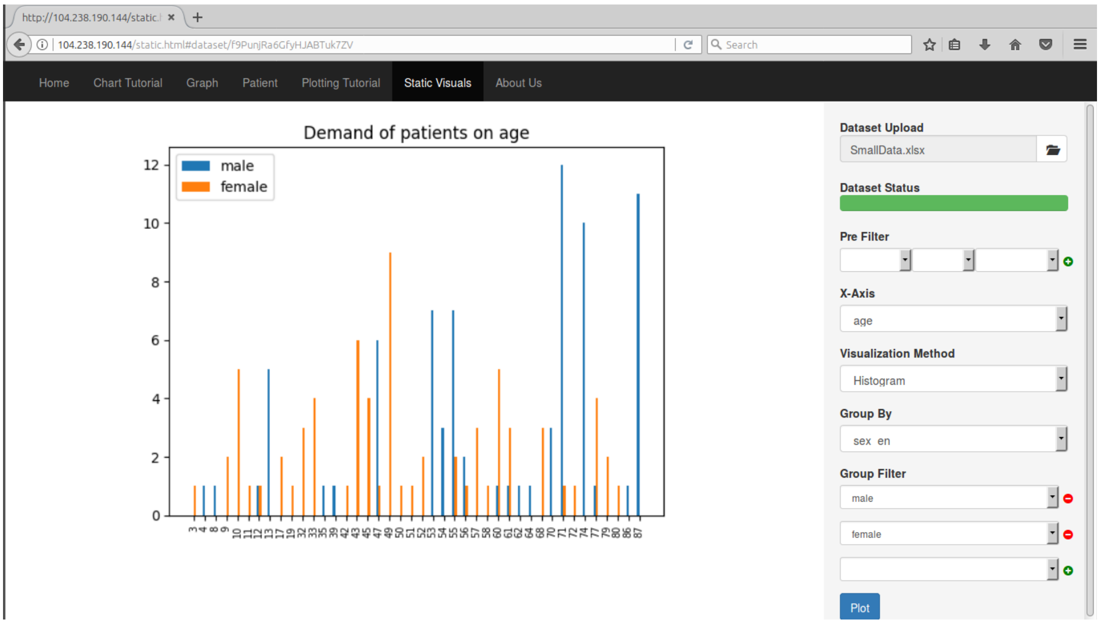
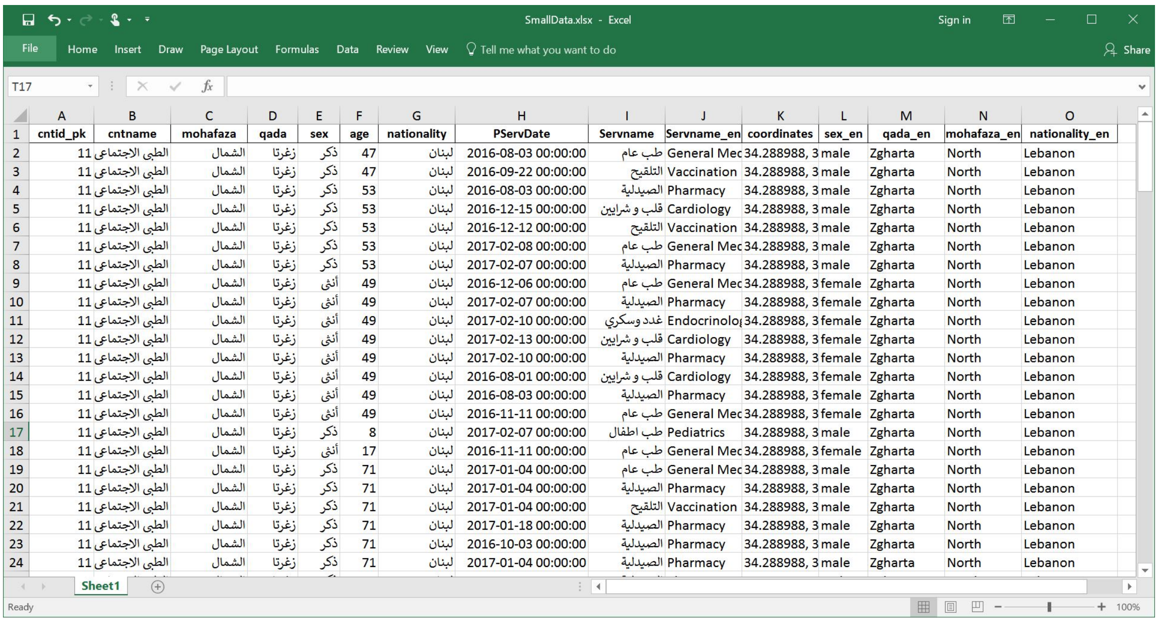



Figure 2: Static Visuals page



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | cntid_pk | cntname | mohafaza | qada | sex | age | nationality | PServDate | Servname | Servname_en | coordinates | sex_en | qada_en | mohafaza_en | nationality_en |
| 2 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 47 | لبنان | 2016-08-03 00:00:00 | طب عام | General Mec | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 3 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 47 | لبنان | 2016-09-22 00:00:00 | التلقيح | Vaccination | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 4 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 53 | لبنان | 2016-08-03 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 5 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 53 | لبنان | 2016-12-15 00:00:00 | قلب و شرايين | Cardiology | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 6 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 53 | لبنان | 2016-12-12 00:00:00 | التلقيح | Vaccination | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 7 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 53 | لبنان | 2017-02-08 00:00:00 | طب عام | General Mec | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 8 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 53 | لبنان | 2017-02-07 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 9 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2016-12-06 00:00:00 | طب عام | General Mec | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 10 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2017-02-07 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 11 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2017-02-10 00:00:00 | غدد وسكري | Endocrinolo | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 12 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2017-02-13 00:00:00 | قلب و شرايين | Cardiology | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 13 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2017-02-10 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 14 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2016-08-01 00:00:00 | قلب و شرايين | Cardiology | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 15 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2016-08-03 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 16 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 49 | لبنان | 2016-11-11 00:00:00 | طب عام | General Mec | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 17 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 8 | لبنان | 2017-02-07 00:00:00 | طب اطفال | Pediatrics | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 18 | 11 | الطبى الاجتماعى | الشمال | زغرتا | أنثى | 17 | لبنان | 2016-11-11 00:00:00 | طب عام | General Mec | 34.288988, 3 | female | Zgharta | North | Lebanon |
| 19 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 71 | لبنان | 2017-01-04 00:00:00 | طب عام | General Mec | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 20 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 71 | لبنان | 2017-01-04 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 21 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 71 | لبنان | 2017-01-04 00:00:00 | التلقيح | Vaccination | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 22 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 71 | لبنان | 2017-01-18 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 23 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 71 | لبنان | 2016-10-03 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | male | Zgharta | North | Lebanon |
| 24 | 11 | الطبى الاجتماعى | الشمال | زغرتا | ذكر | 71 | لبنان | 2017-01-04 00:00:00 | الصيدلية | Pharmacy | 34.288988, 3 | male | Zgharta | North | Lebanon |

Figure 3: A dummy snapshot representing potential MoPH Dataset

### 2.4. Data Analytics and Visualisations

Our data analytics and visualisations are focused on spatiotemporal exploration of the data. Our first step is to allow the user to filter the data using one or more attributes in the file. For example, the user can specify a certain date range or a particular location. Once the dataset is filtered, the user can generate a set of visualisations to get a summary of the data. Our system supports both motion (dynamic) and static charts. Particularly, we support *histograms*, *line plots* and *box plots*, which are all generated using Python's Pyplot library[7] in the "Static Visuals" page seen in Figure 2.

In addition to static visualisations, our system supports motion charts powered by Google Charts[8]. Motion charts are a natural way to explore data in a spatiotemporal fashion. The motion chart is dynamic and thus enables users to understand the change in several indicators as a function of time.

## 3. Case Studies

### 3.1. UNHCR's Persons of Concern

The "Graph" page has a "Browse" component in the top left corner of the page, seen in both Figures 4 and 5, that allows the user to import two csv files. Each of these files should be in such a format that it features an indicator with

[7] `https://matplotlib.org/`

[8] `https://developers.google.com/chart/interactive/docs/gallery/motionchart`

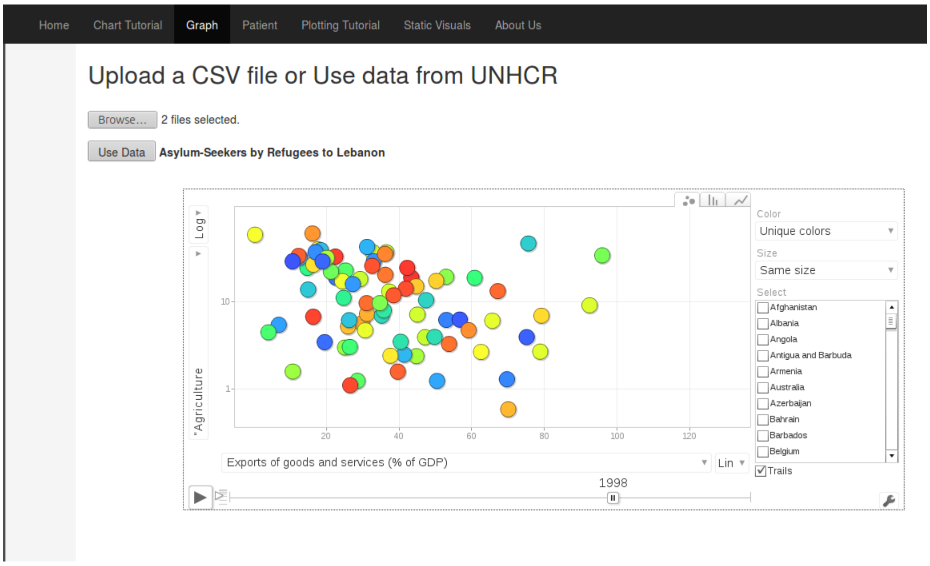



Figure 4: Motion Chart from the imported indicators

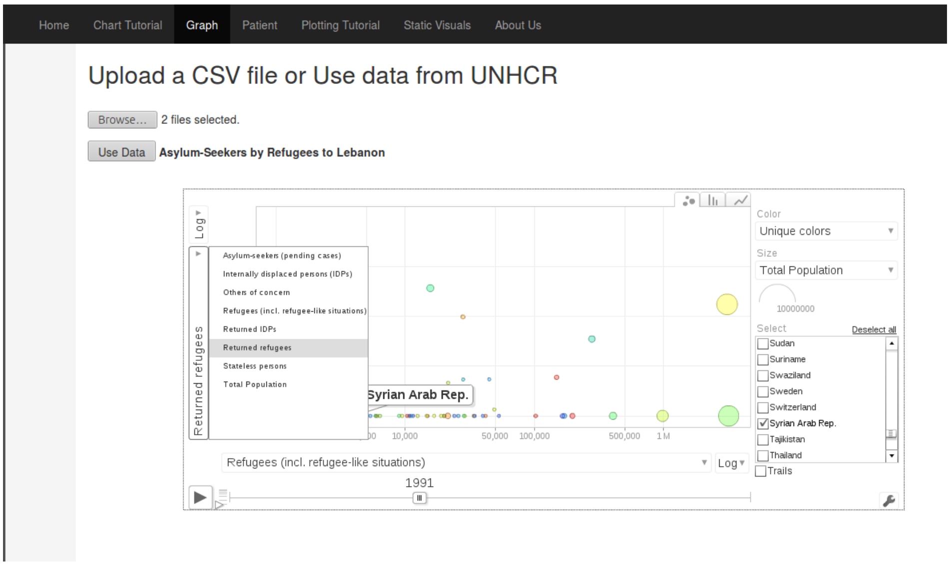



Figure 5: Motion Chart from UNHCR's Persons of Concern dataset: Number of Returned Refugees versus Total Number of Refugees

the columns in years and the rows as locations. In order to retrieve the best results, the years and locations should be common in both indicators. Otherwise, the plots generated and the resulting correlations may not make sense. Once these files are imported, the datasets go through a cleaning process and are displayed in a Motion Chart for the user to view.

The "Use Data" component, located under "Browse", loads a dataset from the United Nations High Commissioner for Refugees (UNHCR) named "Persons of Concern". This dataset is a csv file where the first column denotes the years, second column denotes locations, and the successive columns are the observations which the user would like to represent. Once loaded into the Motion Chart, the user can choose from a collection of columns to represent either axes and the size of the bubble as seen in Figure 5 in the drop-down list of indicators from the UNHCR dataset on the Y-axis. For example, Figure 5 plots the number of Returned refugees against the number of Refugees, across an interval time, versus the total population, indicated by the size of the bubbles. Motion charts like these help shed light on the extent to which refugees are able to integrate, as opposed to return to places of conflict. It also helps shed light on the load borne by the host community in terms of population size.

The format of the import files for the "Browse" component is quite simple to replicate. Users can make quick modifications to their own datasets and replicate the required format in order to visualize their own data on our platform. From Figure 4, the users can observe how their data evolve with time and such, deduce a correlation from their data if one exists. In this figure, one can investigate a correlation between gross GDP and agricultural net output, for example, also across a given span of time. To facilitate studying the

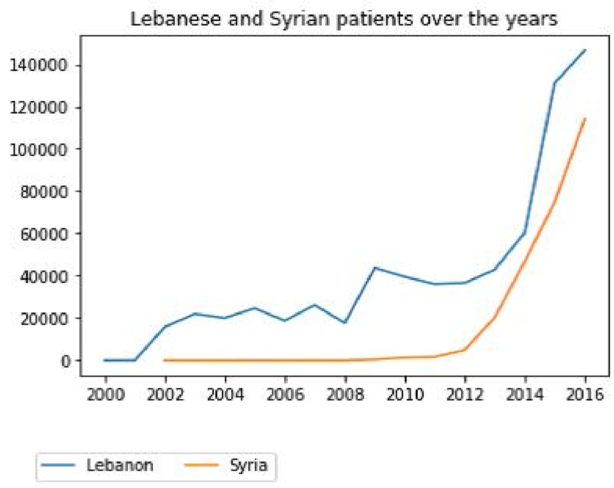



Figure 6: Overall Demand

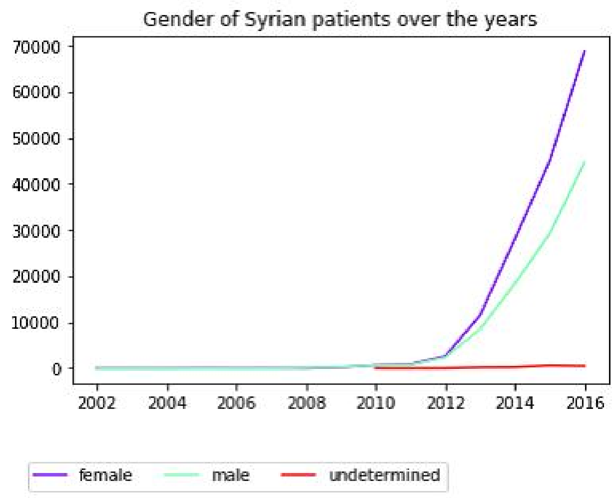



Figure 7: Gender Distribution

dynamic graph, the user is provided with a couple of options found on the left of the graph: the user can select one or more locations to follow their movement with time, and the user can check the "Trail" box found at the lower right corner of the graph to track the development of the selected locations. Such visualizations can be useful for personnel working on the refugee crisis in Lebanon.

### 3.2. Datasets Released by the Ministry of Public Health in Lebanon

In addition to the global challenges common to all, Lebanon is bearing an additional burden associated with its hosting of Syrian, Iraqi, and formerly, Palestinian refugees. It is timely to try and understand the impact that this continuing crisis is bearing on the primary health care system in Lebanon. The pivotal dataset that we have experimented with was received by another party on behalf of the Ministry of Public Health (MoPH) in Lebanon. The data received by the current authors is in aggregated rather than individualised form, and so does not represent human subjects per se. Instead, it constitutes a time series depicting daily demand on the primary health centers in Lebanon since 2009 by any given nationality on Lebanese soil. The locations of the centers are in terms of districts. Using our website, one can manipulate filters to perform exploratory and spatial analysis that help quantify the trends. Particularly, one can choose to observe how trends in demand have been changing by gender, by nationality, by medical departments named in the dataset, by location, by season, by year, and cues into the trends before the Syrian war and afterward.

The "Static Visuals" option of our interface, seen in Figure 2, allows us to make a selection of various exploratory plots. Once someone has produced a csv file similar to the one in Figure 3, they can put in some filters on some of the attributes. For example, one can choose to plot observations for age groups where age is greater than 50, or to plot observations where data is greater than 2011. One can also choose to subset the MoPH dataset by governorate and district. Currently, our tool shows these two terms in Arabic transliteration for ease of use by public servants, respectively shown as "Mohafaza" and "Qada". The x-axis is an independent attribute against which the user can choose to plot the magnitude of demand on given medical centers. The "visualization method" allows the user to choose any of histograms, line plots, or box plots. The "Group by" option allows the user to aggregate demand across, say, more than one district, or more than one governorate.

The initial exploration of demand we conducted reveals that there has been a clear increase in the number of Syrian patients over the past couple of years whereas the number of patients from other nationalities seems to be insignificant. Even though our data spans the years 2000 to recent 2017, we only see a real impact of Syrian patients around 2012 onwards, to the extent that their numbers are equalling that of the Lebanese patients (Figure 6). In Figures 8 and 9, we look at the age distribution of Syrian patients in 2009 and 2016. While in both cases the skewness is positive signaling the highest demand for infant care, the number of Syrian children treated in 2009 was around 100 while it exceeds 30000 in 2016. We now take a look at the geospatial distribution of Syrian patients around the various districts. Choosing the histogram option from our tool and the filtering according to Syrian patients in both 2009 and 2016, we observe the following. While in 2009 Syrian patients were mostly aggregated around the regions of Baabda, Metn and Tripoli, which are mostly industrial and economical hubs that provided jobs for the Syrian population residing in Lebanon at that time. In 2016, the distribution shifted to span multiple districts in the Northern part of Lebanon, which is attributed to the geographical proximity of these regions to hotspots of conflict along the northern borders with Syria.

Filtering at the level of medical services, we observe that the high number of patients going into Pediatrics in 2009 makes sense since the age distribution at that time was skewed towards infants. However, in 2016, while the Pediatrics services are the most visited, there seems to be a high number of people getting into the Pharmacy, Gynecology, General Medicine, and Vaccination services. Filtering at the level of gender, the results in Figure 7 indicate that

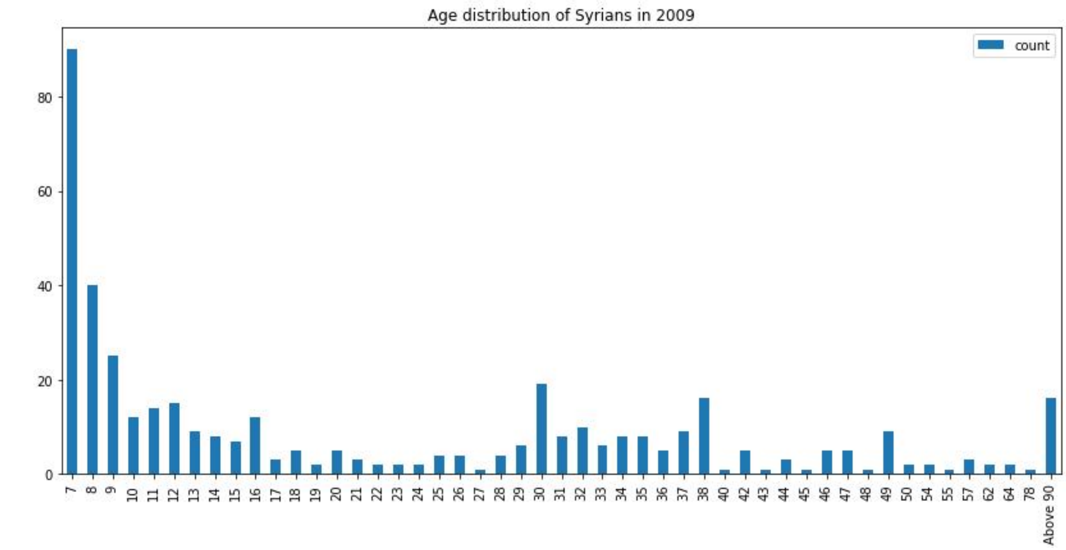



Figure 8: Age distribution 2009

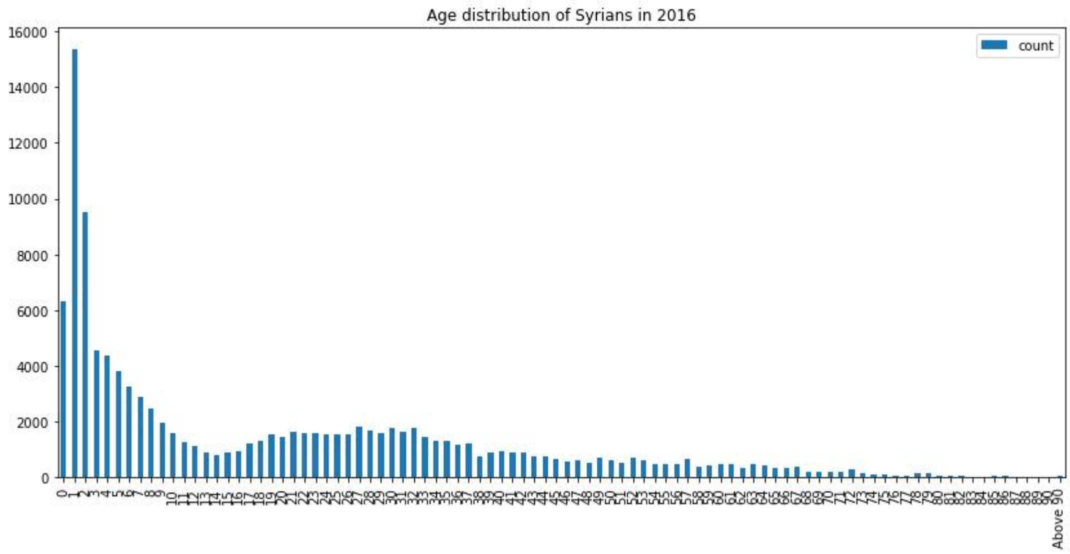



Figure 9: Age distribution 2016

the number of female patients has increased tremendously over the years, significantly faster than that of the males. This seems to confirm earlier official reports that females and children constitute the largest part of Syrian refugee population in Lebanon.

Certain figures can help us pave the way for time series forecasting by studying the trends in the data. For example, Figure 10 shows that the demand on a center in the Akkar district follows a strong increasing trend with "peak-and-trough" patterns and cyclic components since the last three years. The peak-and-trough trends are attributed to the natural fluctuation of demand according to seasons. The cyclic effects can be attributed to the fact that the volatility of the situation causes data to exhibit rises and falls that are not of fixed period. This information is important to gather before one attempts to forecast the expected increase should current influx continues. In contrast, this same center features a different time series for serving the Iraqi refugees. Particularly, Figure 11 shows that the demand has no strong patterns that would help with developing a forecasting model. The mere increases and decreases in secular trend are a result of a less volatile situation with Iraqi refugees whose mobility into Lebanon was not as much of a direct result of traumatic events across shared borders as in the case with the Syrian refugees.

### 3.3. The global Terrorism Dataset: A Spatiotemporal Perspective onto the Primary Health Care Dataset

As mentioned in Section 2., the dataset imported to the "Patient" page underwent feature engineering. Specifically, we used another dataset from the Global Terrorism Database

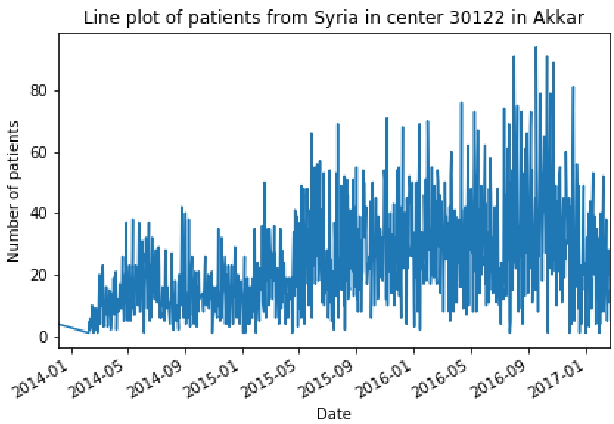



Figure 10: Time series trend – Syrian Patients in Akkar

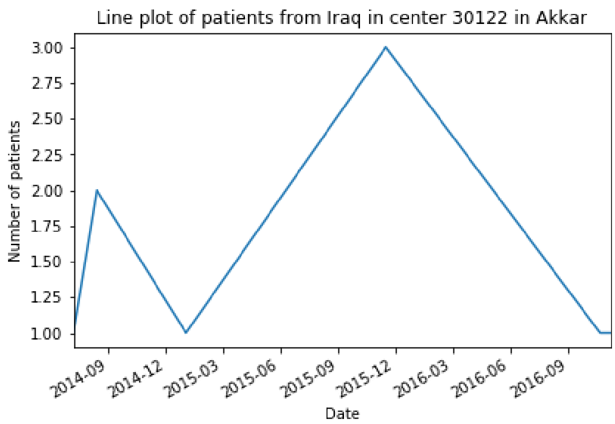



Figure 11: Time series trend – Iraqi Patients in Akkar

(GTD) in order to extract the proximity of the health centers found in the MoPH dataset to the trauma events found in the GTD. The rationale behind this suggested approach is that geographical proximity of violence hotspots to certain geolocations in Lebanon might explain the increase in demand upon centers in the implicated Lebanese districts.

We proceeded by grouping events in the GTD into weeks, then we clustered the events near each other in each week using Python libraries such as Pandas and Scikit-learn (sklearn.cluster). This was done to avoid artificially distinguishing among traumatic events that are near each other or overlapping in space. We created a new DataFrame with the number of clustered events per week, then calculated the number of clustered events nearest to the hospital centers. This information was then appended to the original dataset featuring demand on medical centers. The final product is a dataset accumulating for time observations both the number of patients entering a medical center per week as well as the number of clustered trauma events nearest to a hospital center. Figure 12 shows the effect of high-frequency trauma events to centers around the Beirut and Mount Lebanon area. The centers were colour-coded by the district they are in. The motion chart in Figure 12 shows the demand by time. The x-axis represent the number of trauma clusters within a specified radial distance of about 277 km from the medical center. This choice of distance was arbitrary. The y-axis represents the demand on the particular center. The size of each circle can at this stage either indicate nothing ("same size" option), but it can also be modified by the user to have it correspond to the number of trauma clusters or the actual demand on the center, whichever is more visually appealing to the user. The screen-shot of the motion chart in play showing in Figure 12 reveals an interesting observation: almost all of the centers experiencing highest demand are those which have been detected to be closest to high-frequency trauma centers within about 277 km radial distance from them.

Potential users of this option can have the choice of updating a merged csv file that represents both the demand on medical centers plus their own accumulation of the frequency of traumatic events happening near a specific center. They also have the option of entering their own list of traumatic events provided it is in a format compliant with that of the GTD. For future work and amendment, we plan on giving the user the option of also manipulating the predefined distance now being at 277 km.

## 4. Conclusion and Future Work

The proliferation of data science techniques has generated innovative, timely, and cost efficient ways of capturing actionable intelligence in low-resource, high-risk settings. A data science approach based on machine learning and spatio-temporal analytics can contribute towards more precision for policy making and population interventions at the level of primary health care services offered by the government (Stevens and Pfeiffer, 2011). Particularly, one needs to derive computational insight into the impact and challenges surrounding the primary healthcare system in Lebanon through data provided by the primary health care unit at the Ministry of Public Health. With the help of automated techniques, one can develop real-time forecasting systems of spatiotemporal demand on the primary health sector, as a result of an abnormal rate of population growth in Lebanon, and especially if current trends of refugee influx continue. Our developed platform will help owners of data with no assumed knowledge in code development or software deployment to perform the exploration needed before embarking on any spatiotemporal modeling of the data. Using our tool, we were able to dig deeper into the temporal and spatial features affecting demand on the primary health care centers in Lebanon as a result of the Syrian refugees crisis. In the future, we also plan to allow users to visualize semi-structured and unstructured data by employing Natural Language Processing (NLP) techniques as well.

Our tool is still in its infancy and there are already ample ways it can be improved and made available to the public. Particularly, we are interested in adding a "GIS" component to it that helps users visualise spatial analysis of trends and understand spatial correlations – for example, understanding demand both as a result of seasonality as well as a result of geographical proximity to traumatic events in Syria. We anticipate our tool will be extremely useful for public servants working at the MoPH or staff at the various NGO's in the refugee aid sector with no adequate data analysts or programmers among their ranks.

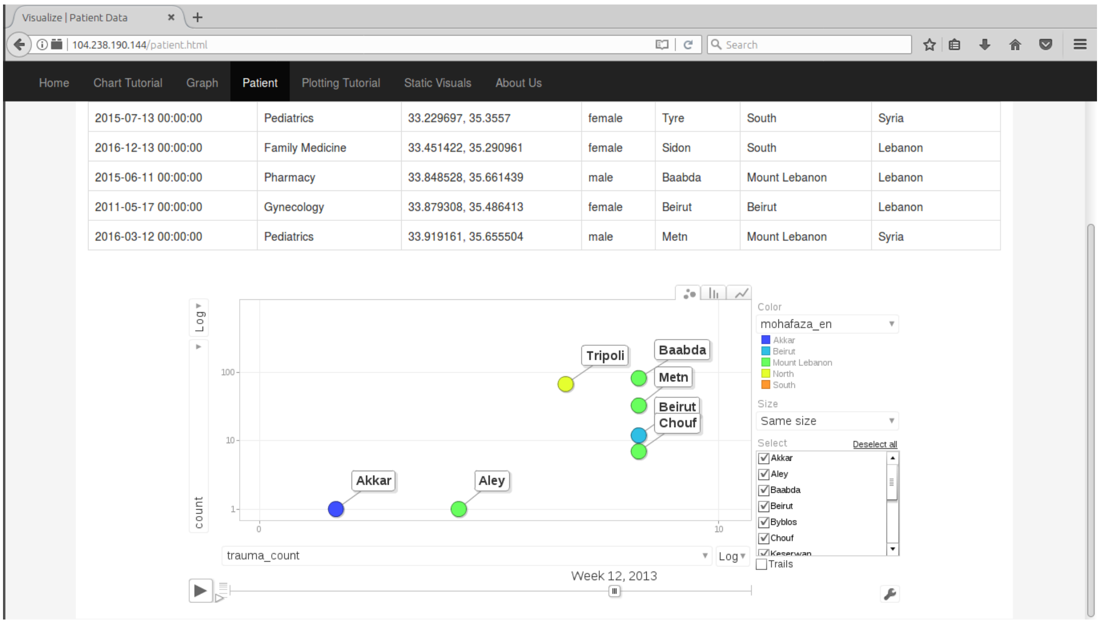



Figure 12: Motion Chart from MoPH dataset

## 5. Acknowledgments

The authors are grateful for the Primary Care Unit at the Ministry of Public Health in Lebanon for proving datasets that record the demand on the primary health care centers before and after the Syrian war.

## 6. Bibliographical References

Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. *Icwsm*, 8:361–362.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., and Wiswedel, B. (2009). Knime-the konstanz information miner: version 2.0 and beyond. *ACM SIGKDD explorations Newsletter*, 11(1):26–31.

Breisinger, C., Ecker, O., Al-Riffai, P., and Yu, B. (2012). *Beyond the Arab awakening: policies and investments for poverty reduction and food security*. Intl Food Policy Res Inst.

Fischer, S., Klinkenberg, R., Mierswa, I., and Ritthoff, O. (2002). Yale: Yet another learning environment–tutorial. *Colloborative Research Center*, 531.

Ghanem, T. M., Magdy, A., Musleh, M., Ghani, S., and Mokbel, M. F. M. (2014). Viscat: Spatio-temporal visualization and aggregation of categorical attributes in twitter data. In *Proc. of ACM SIGSPATIAL*, pages 537–540. ACM Press.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.

Hwalla, N., Weaver, C. M., Mekary, R. A., and El Labban, S. (2016). Public health nutrition in the middle east. *Frontiers in public health*, 4.

Jänicke, S., Christian, H., and Gerik, S. (2013). Geotemco: Comparative visualization of geospatial-temporal data with clutter removal based on dynamic delaunay triangulations. In *Communications in Computer and Information Science*, volume 359. Springer.

Koliopoulos, A.-K., Yiapanis, P., Tekiner, F., Nenadic, G., and Keane, J. (2015). A parallel distributed weka framework for big data mining using spark. In *Big Data (BigData Congress), 2015 IEEE International Congress on*, pages 9–16. IEEE.

Stevens, K. B. and Pfeiffer, D. U. (2011). Spatial modelling of disease using data-and knowledge-driven approaches. *Spatial and spatio-temporal epidemiology*, 2(3):125–133.

Taleb, Z. B., Bahelah, R., Fouad, F. M., Coutts, A., Wilcox, M., and Maziak, W. (2015). Syria: health in a country undergoing tragic transition. *International journal of public health*, 60(1):63–72.