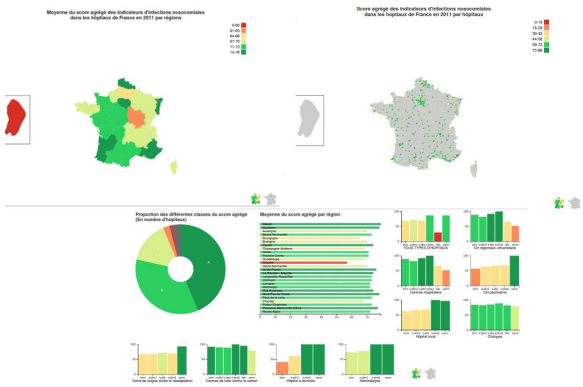
D3js : Nosocomial Disease in France (2011)

Billal Chellali, Jeremy Besson

TEASER IMAGE



ABSTRACT

As part of the first semester of Master 2 Data Science, we carried out a project in Data Visualization which Mr. Tabar, Mr. Vuillemot and Mr. Bonneel gave us. This project aims to become familiar with D3js while applying a concrete subject of everyday life. They left us the choice about the dataset. Indeed, we chose to study nosocomial diseases with the dataset from France in 2011 [1]. This dataset contains all hospital and clinical in 2011 [2] and a set of indicators which directly concern nosocomial infections in France.

The objective was to represent the data spatially and to be able to decipher them in depth. We wanted to create a pertinent dynamic visualization monitoring for the High Authority of Health in France and every organizations involved in fight against nosocomial diseases.

1. INTRODUCTION

Nosocomial infection, is an infection that is acquired in a hospital or other health care facility. To emphasize both hospital and non hospital settings, it is sometimes instead called a health care–associated infection (HAI or HCAI). Such an infection can be acquired in the hospital, nursing home, rehabilitation facility, outpatient clinic, or other clinical settings. Infection is spread to the susceptible patient in the clinical setting by various means. Health care staff can spread infection, in addition to contaminated equipment, bed linens, or air droplets. The infection can originate from the outside environment, another infected patient, staff that may be infected, or in some cases, the source of the infection cannot be determined. In some cases the microorganism originates from the patient's own skin microbiota, becoming opportunistic after surgery or other procedures that compromise the protective skin barrier. Though the patient may have contracted the infection from their own skin, the infection is still considered nosocomial since it develops in the healthcare setting. We have thus worked on it to show the organizations of hospitals the factors influencing this disease.

One in 20 hospitalized patients has a nosocomial infection in France. The Inspectorate of Health Surveillance (InVS) presented the results of a large survey carried out between May and June 2012 in collaboration with the centers of coordination of the fight against nosocomial infections. 1938 French health establishments, more than 90% of hospital beds, were consulted to measure the number of people infected on a

given day. Of the 300,330 hospitalized patients, 15,180 (5.1%) had one or more nosocomial infections. Figures similar to those observed in 2006, the date of the last survey of this scale, where 4.97% of patients hospitalized suffered nosocomial diseases. As a result, there has been no overall improvement in the number of people affected. Nevertheless, InVS observes a sharp change in psychiatric services, where infections have declined by 21% between 2006 and 2012.

2. RELATED WORK

Many works has been done with geographical representation using d3.js. D3 offer many type of projection make data visualisation with geographical [9]. Indeed, with the geographic and projection functions under d3, we can do different projects with maps [8].
Nevertheless, we have seen few projects concerning map and statistical representations in France.

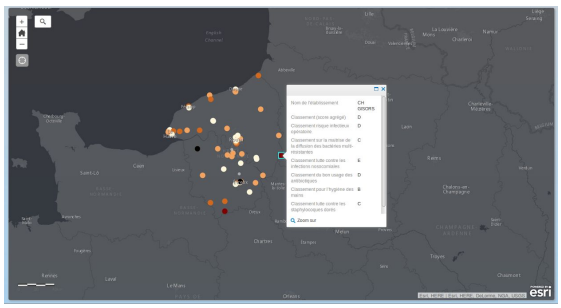


figure 1: Lutte contre les infections nosocomiales dans les établissements haut-normands from [7]

Has you can see on figure 1 a part of the dataset we focus on was already used for a specific France region. This visualisation only include specific geolocated data. And no aggregation has been done to see interesting patterns or something else than the original data with additional geolocation. This visualisation was done with ArcGIS an easy way to create map.
In our case we wanted to do more than just a map representation. We wanted to create an application where we can see interesting pattern in the data and dig into those interesting pattern. That why D3.js is a good technology to achieve this. We can combine geographical data with statistical data.

3.PROJECT DESCRIPTION

As we said in the abstract our goal was to construct a monitoring interface for health professional base on 2011 hospital data.
We choose to offer a web interface divided in three main part. On a first part you will find a France map where every 2011 regions are represented with differents colors. Each color will represent how establishments in this region are fighting against nosocomial infections.
On a second part you will find a France map again but with every establishment geolocated on the map. We also use the same coloring system to see where data come from and easily detect outlier and understand other aggregation. On the last visualization part we built a dynamic monitoring system where you can see every indicators by France region, by type of hospital and by category explains in the next parts.

3.1 Search Data & Clean Data

We started the project with a dataset [1] that was correctly cleaned. Every establishment was record with there name, region, type and every indicators with two scales. A numeric scale 0 to 100 generally and an alphabetical scale (A,B,C,D,E) where number 100 and category A are the best scores. Every establishment were also record with a Finess number which is a unique identifier for each Health establishment in France.

However to construct a map who represent the global repartition of health establishment we needed their geolocation. This information was missing in the original dataset. To solve that we found a dataset where every Finess number where record with the current establishment postal address [3]. This document contained more establishment than in [1] because it also included other health establishment like pharmacy, laboratory not concerned by nosocomial infections. To get only addresses we needed we used QGIS 2.18 software to perform a join on Finess number we had in the original dataset. After an extraction step to get every addresses separately we use this website [4] to transform every addresses into geolocation. The final dataset was built by gathering the original dataset with postal addresses and geolocation of establishments.

3.2 Score By Region

In a first step, we visualized the mean of the scores by regions. The score is a variable that is calculated on the basis of different indicators per establishment, the higher it is and the better the establishment is ranked relative to others by considering the various indicators. Indeed, we performed an average of his scores by regions. Concerning the colors, we decided to use a contrast between red and green. The lower the score, the more the color will be red, and the higher the score, the more the color will be green. This choice of color is applied on all the visualization of our project for the user understanding.

Before performing the visualization on d3, we began to think about it through the Public Table software. We had first noticed regions that were distinguished by their score. Notably, Guyana which had a relatively low score. On this fact we were obliged to add it in our visualization d3. To add the guyana, we first searched his coordinates, because in our basic json file, we did not have it. Once the Guyana was added, it was necessary to be able to change the coordinates while remaining coherent. Indeed, the coordinates found were not in adequation with our representation of France, therefore we had decided to modify them to be able to represent France and Guyana well centered and zoomed.
At the end of our first visualization on d3, we had the same as we had already done on Tableau Public, especially with the Guyana.

3.3 Establishments in France

The second visualization represents the aggregate score of indicators of nosocomial infections in the hospitals of France in 2011 by hospitals. In fact, we decided to display the different hospitals, pointing the mouse on a hospital you can see the data on such as its aggregate score, his town and scores of different indicators. We can show the distribution of the different establishments in France and we have a data visualization more concrete. With this visualization an application user can access with a better view at each row of the dataset so it is easier to understand where data of others visualisation are from.

3.4 Bar Chart with Indicators by Regions

After building a map visualisation to show the global repartition of nosocomial infections in France we wanted to find a way to go deeply in the data we focus on. That why we choose to implement a dynamic crossfilter to show in a web browser the impact of every indicator on the global nosocomial indicator by regions and by hospital type. There were six specifics indicators for health facilities in 2011.

ICALIN2. Is a general indicator on fight against nosocomial infections in institutions.
It represent the organization, resources and actions perform against nosocomial infections.

ICQHA2. A specific indicator on hand

hygiene. A key measure to prevent many nosocomial infections.

ICA-LISO. A specific indicator on operative infectious risk. It represent the institution's commitment to assessing and improving practices and controlling infectious risk in surgery.

ICATB. Indicator on good use of antibiotics. Reflects the level of commitment of the healthcare institution, in a strategy to optimize the effectiveness of antibiotic treatments.

ICA-BMR. Indicator that makes visible the level of commitment of the health establishment, in an approach aimed at controlling the diffusion of multi-resistant bacteria.

SARM. Methicillin-resistant Staphylococcus aureus in establishment. This indicator reflect its ability to control it through measures to prevent patient-to-patient transmission and a policy to control Prescriptions for antibiotics. This multi-drug-resistant bacterium is frequently the cause of nosocomial infections.

We normalized all these indicators to show them in a range of 0 to 100. The more a hospital has a good rating the more is associated value will be close to 100. After that we build a crossfilter with dc.js technology. It permit us to show dynamically all of those indicators aggregate by region and type of establishment at the same time. You'll find nine bar chart these correspond to each type of hospital you can find in France and another bar chart for indicator aggregate on all hospital.
On the application you first see information of the entire France and you can directly click on any regions names to see all indicators aggregate for this or those specifics regions.
We also add a pie chart to represent the repartition of the global class on nosocomial infections in any specific regions or the entire country. The information was indicated in the dataset for every establishment. Each of them have a class A, B, C, D, or E and it represent generally "good" or "bad" hospital against nosocomial infections. We choose to let the opportunity to the user to select and see information of any specific class. Interesting to rapidly see information on best or worst practice.

4. DISCUSSION

We have introduced our Visualization on 2011 Nosocomial data in France. We have also shown how it can be used by professional to understand deeply the nosocomial problem in France. We now briefly discuss some general advantages and limitations of this approach, as well as some extensions to our current system.

4.1 Advantages

Homogeneous coloring visualization. We build three visualization with the same coloring scale to help the application user to understand more the data. The objective was to accentuate correlations between locations and indicators. We also used a strong color separation [5] to accentuate distinction between bad and good notation in these data. Theses colors were chosen for a good human perception.

Animated platform to dig data. Animated representation of data allow users to ask the question they want to those data [6]. That why we develop this application with animation and relation between those animation. Users can select the part of data who concern him or study a particular distribution of those data with some selection with only few click.

4.2 Limitations

Noisy geolocation data. Many factor influences construction of a geolocated

dataset. Geolocation change when time progress for example an establishment can close since the data were collected. Moreover we face some difficulties to get correct postal address because some street name change over the time and it's harder to get geolocation in this case.

Overseas departments representation. We have as mentioned above added Guyana. Nevertheless, we could not add more regions of DOM-TOM. Indeed, we found an interactive web application with d3 which displayed France with its DOM-TOM but by entering the code, it was very specific to the application and we could not adapt it to our application . We also found data in geojson on DOM-TOM. First we had to find its data but with the geometry type "MultiPolygon" we tried to adapt its data so that they were consistent with our but we did not have the good results. It was later that we found the right data with a geometry type "Polygon". By adding these data to our geojson file, the data did not appear to be correct. So to conclude, we decided not to waste time with the others, because we already had Guyana and we thought that if we had time to finish, we would try to find a way to solve the problem of its regions.

4.3 System extensions

Enrich the dataset. We wanted to gather more data into the visualisation to explore more correlation. For example we should be able to see if nosocomial infections is linked with hospitals funds.

Switch between years. Another interesting improvement can be done if we add a dataset of every years nosocomiale investigation. If we allow the user to select the year he want to see he will be able to see how nosocomial infections decrease over years and where it increase if it the case. Unfortunately those data wasn't already open after 2011.

5. CONCLUSION

To conclude, our project has been really interesting. We were able to have a real project, concrete, mounted from the beginning (data search), at the end (interactive web visualization) in Data Visualization. Which is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns. With interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed. This project allowed us to discover d3.js, it helps us attach our data to DOM (Document Object Model) elements. Then we could use CSS, HTML, and SVG showcase this data. Finally, we could make the data interactive through the use of D3.js data-driven transformations and transitions. We used D3.js because it lets us build the data visualization framework that we wanted. Graphic / Data Visualization frameworks make a great deal of decisions to make the framework easy to use. D3.js focuses on binding data to DOM elements. Therefore, being students of Data Science, we are now glad to have been able to realize this project which allowed us to have an experience in Data Visualization. That's why this experience in the world of Data Science enabled us to strongly confirm our future profession.

6. VISUALIZATION

url=https://jeremybesson.github.io/

7. REFERENCES

[1] Les indicateurs relatifs aux infections nosocomiales dans les établissements de santé - 2011
URL=https://www.data.gouv.fr/fr/datasets/les-indicateurs-relatifs-aux-infections-nosocomiales-dans-les-etablissements-de-sante-2011-576189/

[2] Les Hôpitaux et Cliniques en chiffres
URL=http://www.sanitaire-social.com/hopitaux-et-cliniques/definition_organisation/m2/6

[3] Extraction du Fichier National des Etablissements Sanitaires et Sociaux (FINESS) par établissements géolocalisés
URL=https://www.data.gouv.fr/fr/datasets/extraction-du-fichier-national-des-etablissements-sanitaires-et-sociaux-finess-par-etablissements/

[4] Géocoder transform postal addresses into geolocation
URL=http://www.1001maps.fr/index.php?page=043

[5] Simone Bianco, Francesca Gasparini, Raimondo Schettini, January 2014. Color Coding for Data Visualization.
URL=https://www.researchgate.net/publication/268813843_Color_Coding_for_Data_Visualization

[6] Heer J, Robertson G. 2007– Animated Transitions in Statistical Data Graphics.

[7] Lutte contre les infections nosocomiales dans les établissements haut-normands
URL=http://www.arcgis.com/apps/Panels/index.html?appid=6f33692335c349f0b8b5ff98ffd26c3a

[8]DataMaps
URL=https://datamaps.github.io

[9] d3-geo-projection
URL=https://github.com/d3/d3-geo-projection