Data Visualization: LCELO Process Book

Aslam Cader, Etienne Caquot, Jeanne Chaverot

May 2020

1 Introduction

The goal of this project is to display the impact of the Covid-19 virus on the Swiss population. The data set is a daily time series containing the number of contaminated/deaths/recovered, as well as information about the number of individuals hospitalized in a normal ward, Intensive Care Unit(ICU), or in need of respiratory assistance. As we have access to this data for each Swiss canton, we have decided to use this project in order to enlighten the situation of this health crisis in each of them. The output from this project will help have a better understanding of the situation in each canton, as well as seeing which one(s) are being overwhelmed by this crisis.

2 Path to obtain the final result

2.1 Abandoned goals

The interest in the Covid-19 data-set arose upon realization that as individuals we are surrounded by misconceptions around the reality of this pandemic. As we know, many websites have already used Covid-19 data for general visualizations from a worldwide perspective. We are interested in taking a closer look at Switzerland's situation. First, we wanted to compare the date of the first lock-down (and other security measures) between Switzerland and the rest of the world. We find it interesting to see which Government was the promptest and acted the quickest when faced with this pandemic. Then, we wanted to work on a more human scale. We want to raise awareness on how this virus not only presents rather basic symptoms but is more dangerous than the basic flu to mankind. To do so, we wanted to implement a word-cloud based on symptoms from positively-tested individuals that have presented themselves at the hospital. We also wanted to plot the statistics comparing the Covid-19 to the seasonal flu.
However, these two first ideas were rejected after careful consideration and feedback from our supervisor. It has been suggested that it would be a better idea to focus on a single country in order to extract and display more useful information. Finally, we were told that what was expected were lesser visualizations with deeper meaning and complexity than multiple rather superficial graphs. A visualization such as a dynamic word-cloud, one of our ideas to compare the Covid-19 to the seasonal flu, would be considered too simple and more decorative than an actual source of information.

2.2 Final goals

Therefore, we decided to take a deeper look into Switzerland's reality. Switzerland is divided into 26 cantons, and each of these cantons has its own statistics. In the data visualization protocol, we want to analyze different elements from Swiss hospitals: the number of hospitalizations, number of individuals under respirators, number of individuals in ICU, and number of released individuals. We would like to compare these values with the magnitude of the Canton in question, as well as being able to find similarities between cantons. With an interactive map, we can see the number of infected individuals per Canton.

2.3 Finding the right visualizations

Once we knew that we wanted to work on Switzerland's we needed to define which visualizations to draw. This has been a challenging part as currently, many parties are working on the same subject. We knew that by focusing on Switzerland we had a certain differentiation, however, we needed our ideas to be original. In order to validate our visualizations hypothesis, we conducted a poll to gather some feedback about which information individuals would

like to have access to. This has started an iterative process between brain-storming and finding the appropriate visualization to answer a certain need. We decided to focus on two aspects of our data:

- A first visualization using the number of infected/recovered/deaths per canton
- A second visualization using the data on the situation in hospitals per canton

We chose the two items as visualizations as the first one gives a broad view of the country's situation and the second one, while still focusing on Switzerland, focuses on the hospital's situation. It is common knowledge that many infected individuals end up in intensive care and that many hospitals are overwhelmed. However, we have not seen interactive visualizations depicting this situation.' How many people go in the ICU', 'What fraction of infected individuals need respiratory devices' were some questions we wanted to answer with our work.

3 Challenges and design decisions

In this part, we explain our choices regarding the website design, as well as a detailed explanation of our visualizations.

First of all, we needed to have a dynamic skeleton with a highly responsive website. Working on this subject was interesting and challenging at the same time. Many companies, universities, and research departments have been working on the same data set. Therefore, producing something creative required a lot of brainstorming. That is why we decided to focus on Switzerland: in order to still have a broad view, but with the opportunity to compare the situation between each canton. To help use choosing proper visualization, we used the Data Viz Project [2] website which is pretty well build and displays a large catalog of visualizations. There is a functionality to search by function or by data-set type which really helped us.

3.1 Data Cleaning

The data-set we have is like any real-world data-set dirty and missing data. We used pandas library to counter this problem. There were some cantons were the number we not updated daily but only when the number would change. To bypass this problem, we decided to "fill" the missing entries with the last known number until a new one was seen.

In order to arrange the data such that it fits the required criteria for our design needs, we had to clean and filter it. We selected the data that focused on Switzerland: this gave us the number of cases/deaths/recovered per canton on any given day. However, we believed that this absolute value may not be representative of the virus' spread situation on a canton, as it depends on its density. That is why our first data modification was to transform it in order to change the number of cases/deaths/recovered to a relative value, as a percentage of the canton's population. To do so, we created a dictionary mapping the density to each canton's name. Unfortunately, this mapping is static, we do not fetch the data from an updated data source. We kept the absolute value in our graphs to allow the user to choose the format in which he wants to display the information, see Figure 1 and 2.

3.2 Map overview

Choosing the map was a strategic choice: it's an easy way to compare the evolution of the situation canton by canton. It allows the viewer to understand the data without raw numbers in a more illustrated way. We have chosen orange for cases, red for deaths and green for recovered as orange is usually less alarming that red, while red and green are more attache to negative and positive sentiments. This map is interactive with its slider. The user can see the evolution of the situation in time (starting on February 25th until the date of the deadline). The user can decide which information to display: the number of cases, deaths, or recovered, as well as the format of the number: either absolute or relative. The absolute represents the total number of affected individuals per canton, while the relative value is as a percentage of the canton's population. We believe the relative data is more informative as absolute data from highly populated cantons will always be higher in comparison to smaller cantons. Relative numbers are more representative of a canton's situation.

Below our map, following a discussion with our supervisor, we decided to add a time series displaying the evolution of the number of cases/deaths/recovered in Switzerland. This plot allows the viewer to have a general view of the virus' evolution in the country. As before, this plot adapts in terms of absolute or relative numbers. There

is also a tooltip when you move your mouse over a canton to see all statistics at once.

Designing the map was not easy and we used D3.js along with TopoJson. Having the data for one day and all canton was not that hard. We had problems when we wanted to link the map with the slider, the selection of the right data to display is not easy but we could finally manage it. The most challenging part was with the buttons to change the data-set or the scale for absolute to relative. There were lots of situations that we needed to test to be sure to have the proper update of all the content. For example, we had to track if the slider was moving or not, the tool-tip position...

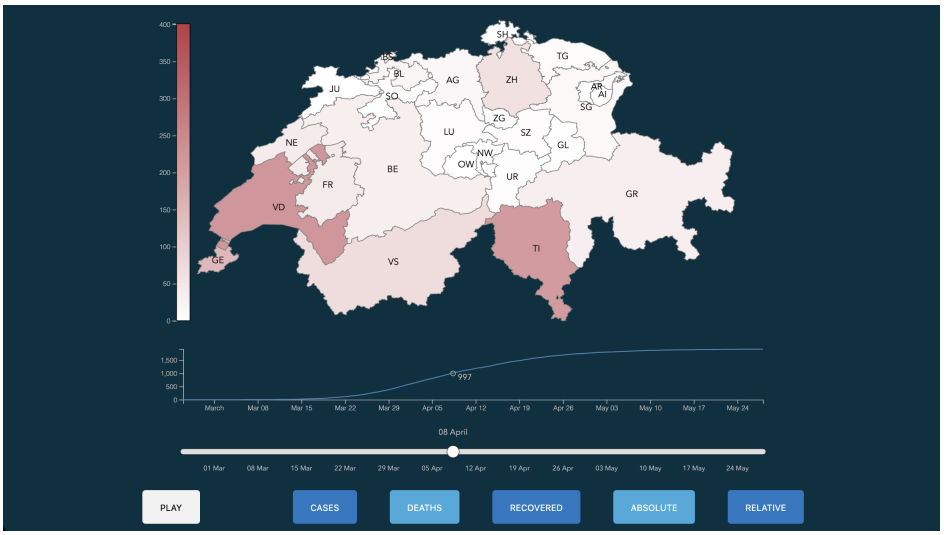

Figure 1: Switzerland map with deaths, absolute count

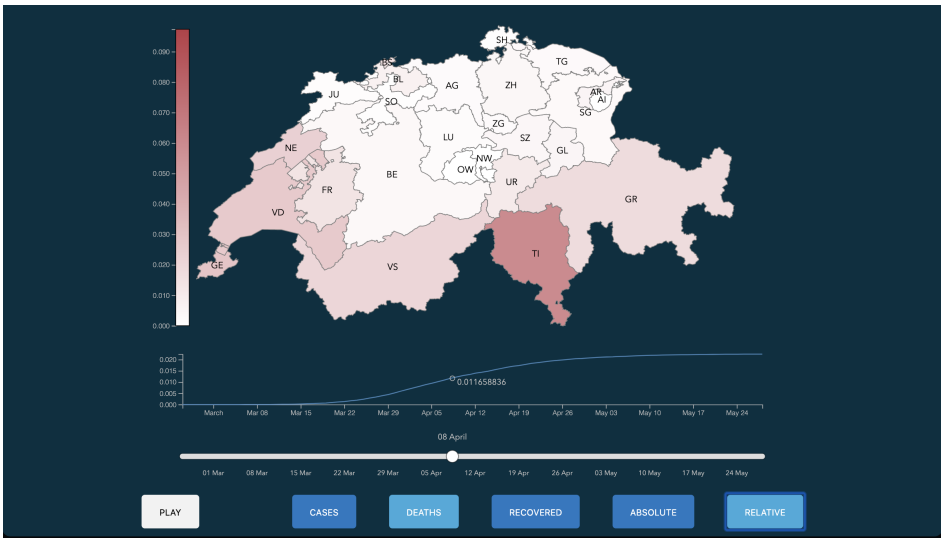

Figure 2: Switzerland map with deaths, relative count

3.3 Graph for hospitalization

This graph shows the distribution of the hospitalized COVID-19 cases: normal ward, ICU ward, or under respiratory assistance. This allows individuals to compare the pressure that each canton's hospital unit has handled so far. We believe that this visualization is a good indicator of a canton's ratio between its density and its health system support.

The viewer can select the cantons he would like to see displayed in the plot and can travel through time with the slider. We decided to let the user decide which cantons he would like to see so that he can compare two or several cantons over time (thanks to the slider). The graph called a Marimekko chart (see Figure 3), allows us to have an easy to interpret illustration, in order to compare the hospitalization situation among cantons. This slice and dice visualization permits a hierarchical view which helps to compare the cases between cantons. An additional feature has been added in case the viewer wishes to see the data displayed for all the cantons (see Figure 4). By clicking a toggle button, you'll be able able to simply display the situation for all cantons. We chose to work with a Marimekko chart as it is more dynamic and responsive, providing a more visual classification of the cantons depending on their hospitalization situations. However, for a more general view, we think it remains interesting to keep a stack bar chart with all cantons.

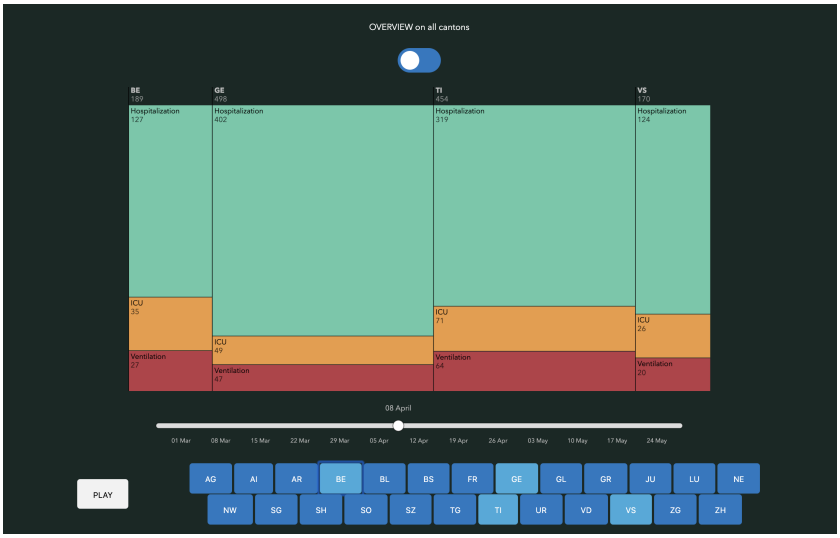

Figure 3: Comparative view for cantons: Marimekko slice and dice view

The major challenges for the Marimekko visualization were the data-format. We had to have a really particular structure of the data, with lists of dictionaries, and then map it to a tree structure to have the cantons on the nodes and on the leafs the data for the canton. It was quite challenging. Again the high number of buttons brought some complexity by handling the insertion of the canton data to the visualization, the slider to select the right date was pretty easy after having done it for the map.

One challenge we faced was to have a nice website design that would help to understand the problem. In order to have coherent color combination, since we are not artists, we relied on an online tool name Colors.co [1] where we could create a palette of colors to have a uniform and nice design of our website.

Another challenge was to find the simplest visualization as possible to show the maximum amount of information. Indeed, making complex visualizations just to have some nice content wasn't enough. In most of the cases, having such content is just pleasant to see but doesn't necessarily bring a lot of information. That's one of the reasons why we stuck to a simple stacked bar chart instead of having a sequence sunburst or something else.

4 Changes since milestone I

In milestone 1, our focus wasn't only around Switzerland, we wanted to compare Switzerland versus the rest of the world, as well as looking into the differences and similarities between the Covid-19 and the flu. However, we were

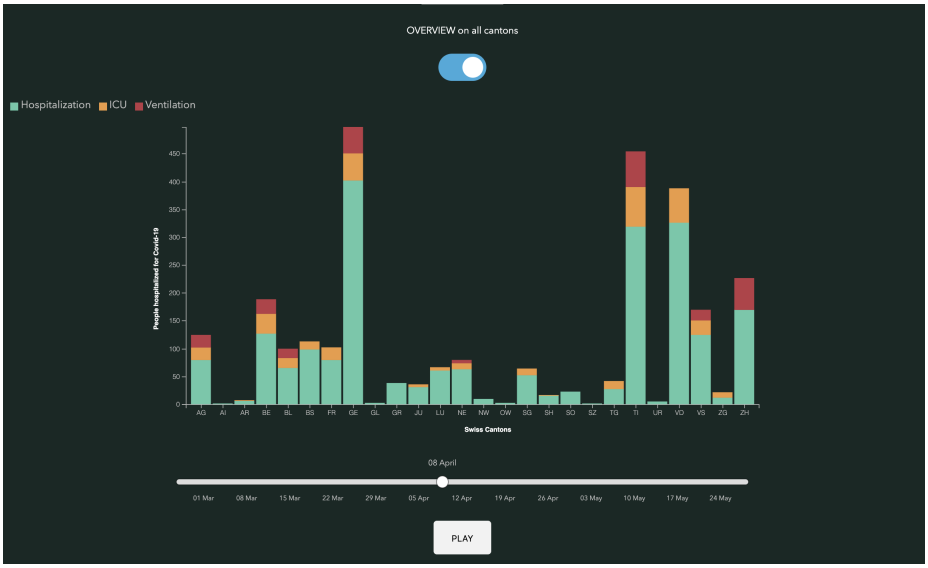

Figure 4: Overview on all cantons: Stack bar chart

correctly indicated that we were going into too many directions at the same time. That is why we decided to focus on Switzerland, in order to dedicate our visualizations around one subject and be more consistent in our work.

5 Peer assessment

In the context of this project, we decided to continuously brainstorm our ideas in order to agree on the storytelling and direction of our visualizations. Etienne focused mainly on the website with the D3.js implementations of our visualizations and design for the website. He worked on the JavaScript code in order to implement the heat map and the Marimekko chart. He also did some data cleaning to fill the missing entries in the data-sets. Jeanne worked in the creation of the visual sketches in order to define a first view of the website. Then, she worked on the data cleaning to give a less biased view of the data. We switched from quantitative data to relative data. She wrote most of the report. Aslam worked on an interactive bubble word-cloud implementation with D3.js in order to detect symptom tendencies in Covid-19 positive patients. However, we later decided not to keep this visualization as we considered it didn't bring enough information and was more of a decorative asset. He worked on the skeleton of the website, did the stacked bar chart. Aslam and Jeanne worked on the screencast scenario and Aslam recorded it.

6 Conclusion

Our team started out this project out of interest for a subject surrounded by questions but with very few answers from the scientific community. As the majority of information treated countries as an overall, we decided to focus on the country we are studying in. We believed Switzerland's interesting territorial disposition would allow us to benefit from the results displayed by our visualizations. Indeed, now that we can explicitly see the situation among cantons, it is easier to grasp a better view of the severity of the health crisis in each of them. Moreover, the Swiss Federal Statistical Office publishes each day new data that was updated in our data set, which allowed us to obtain more content to show day by day[3].

Unfortunately, we have not been able to find a data-set classifying the number of cases/deaths/recovered by age of the individuals. We do believe that such information could be largely beneficial and would allow us to derive more interesting statistics. Finally, we believe the result of our work does answer to the questions we were asking ourselves and hope it will enlighten others users as well on the Covid-19 situation in Switzerland.

References

[1] Fabrizio Bianchi. *Coolors*. https://coolors.co.

[2] ferdio. *Data Viz Project*. https://datavizproject.com.

[3] Office fédérale de la statistique. Page Spéciale COVID-19. https://www.bfs.admin.ch/bfs/fr/home/actualites/covid-19.html.

Stack Exchange Inc. *Stack Overflow.* `https://stackoverflow.com`

Data on canton density: `https://fr.wikipedia.org/wiki/Canton_(Suisse)`