

Team Improviz - Process book

Vincent Cabrini & Martin Beaussart & Luca Joss



The Journey -The brainstorming

Our journey for this project started with choosing what we wanted to show. The first criteria was having enough data to tell something interesting and not just simple statistics. Another important criteria was in choosing data with which we would learn something interesting. We ended up with two possible datasets, a personality test which had interesting questions with the response time and crimes in Chicago from 2001 to now.

Both were great for different reasons, but we chose the crimes in Chicago from 2001 to now. This choice was motivated by the fact that we are used to receive small amount of information about crimes, usually one at a time, from news. We thought it would be more interesting to see criminality with a more global view, on larger time lapses, and try to observe if there are factors with a visible impact. The idea was on displaying general information about it such as the density for a given day or the density of a given type of crime.

Once the main dataset was chosen, we found additional datasets, one which contained the location of police stations in Chicago, one with socio-economic indicators of Chicago population and one which for affordable rental houses in Chicago. These would be used to have additional information which could have an impact on crime density.

-The journey begins

With the dataset chosen we now had to find what and how we wanted to display our data.

The first ideas were mostly based on using a map of Chicago to display the data, but we quickly realized that we would need support from other structures to display our data in a more pleasant way. At this point the idea for more global representations came to us. Firstly a circle with general values as shown in Figure 1, and as well as a stream graph to show information on different types of crimes registered by the Chicago police, which would be displayed as in Figure 2.

In addition to these general displays, we wanted to use a map of Chicago, using Carto and Leaflet, to have a geographical representation. We had an initial idea of showing police stations location and observe if they had an impact on crime density. And rendering this as a small animation, to better see if there was an impact around them.

We thought about using the fact we had different location information, such as the Community Area where the crime happened, to display a sort of grid on the map. This quickly led to having a section of our project containing a map of Chicago with buttons to display different views, this initial sketch looked like Figure 3.

Seeing we had the Latitude and Longitude of crimes, the second step would be to display crime density with a heatmap. This would allow, with a quick glance, to see if there are more dangerous areas, in a smoother way than with line delimiters. And finally, we wanted to be able to observe the evolution of crimes for certain interval of time or at special dates, such as New Year's Eve or Christmas. The interesting point being able to see if some dates had an impact on criminality.

For extra features, we thought about the following ideas, which could improve the user experience when navigating in our project.

A first idea was in adding a "following" mechanism for the animated features, such as the crime density over a time laps. This mechanism would have the camera zoom in specific points, change the rate at which time increments or show a given type of crime for a specific date.

Another addition was to have a modifiable filters so the user could "play"

with the display. The idea being to have the user chose the date and type of crime. This would be great for curious people wanting to see what happened a given date, or where a specific type of crime happens more often.

Lastly if time would permit it, we thought about introducing Machine Learning algorithms to try and predict crimes in certain areas or at given times. As for the filter, this would add a playful side to the project.

Having submitted our ideas, an interesting point was observed, which stated that it would be nice to have the filtering feature in the main features. Which we proceeded in moving to our standard goals and can be found in our final project.

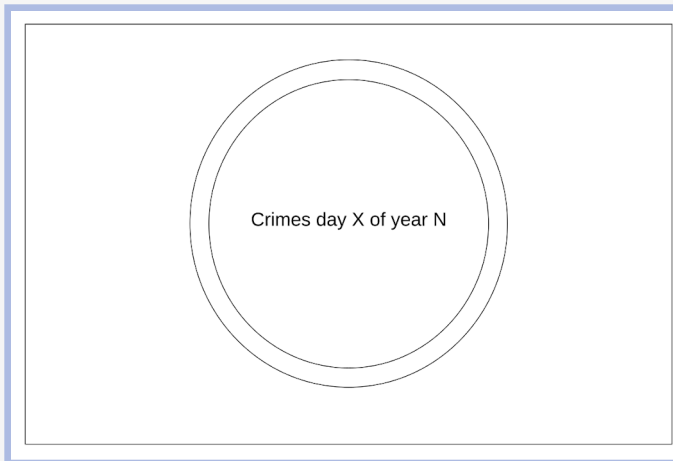


Figure 1: Initial sketch of the circle for general display of information

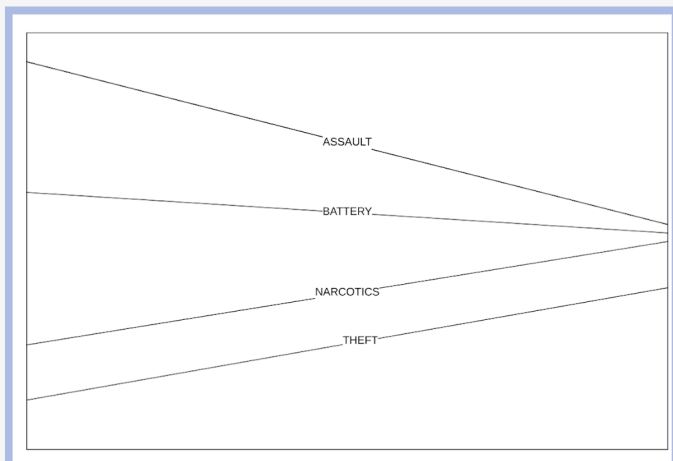


Figure 2: Initial sketch of stream graph for general information on types of crime

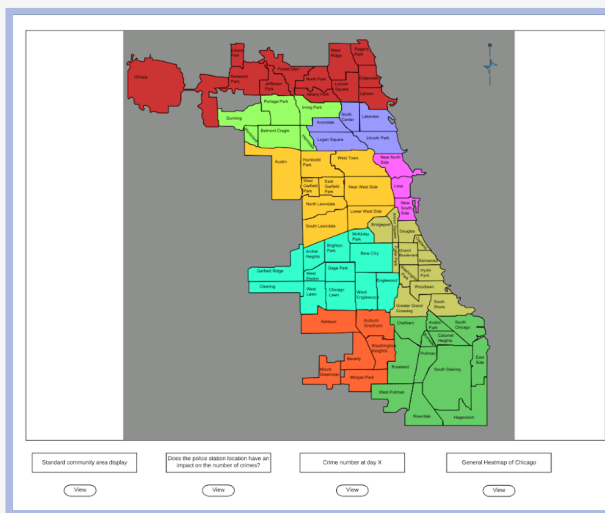


Figure 3: Initial sketch of the map for the specific problematic

-The creation

Disclaimer: The following description is not accurate chronologically speaking since features were developed in parallel by all team members.

We now knew what material we had, our data, and what we wanted to create with. The first step was in implementing the circle which shows the average crimes a given day of the year and the stream graph which shows what type of crime if more frequent. This gives a really great introduction to the user, the first things he sees is a general overview which is easy to understand. In addition this also helps when wanting to play with the filter since it gives a general idea of when and what is more frequent. These were implemented using standard D3 functionalities, with only mouse hover interaction. The idea was not for the user to spend all the time on these features, but use them to better understand the big picture of our data.

-The geographical view

Now that the user has been eased into our world of criminality in Chicago we could start adding more detailed features. The map with different applicable features is placed at this point. Some changes from our initial ideas had to be done once we started implementing them. For starters we realized that Leaflet was enough for what we intended to do, and thus only used its applications for the map. The other modification was in using a heatmap for the animation with the police stations would have a better result since it shows precise locations of the crimes. And since the goal is to observe if the location of the police stations has an impact on the crimes, we thought about showing an evolution day by day of crimes with markers showing the location of police stations.

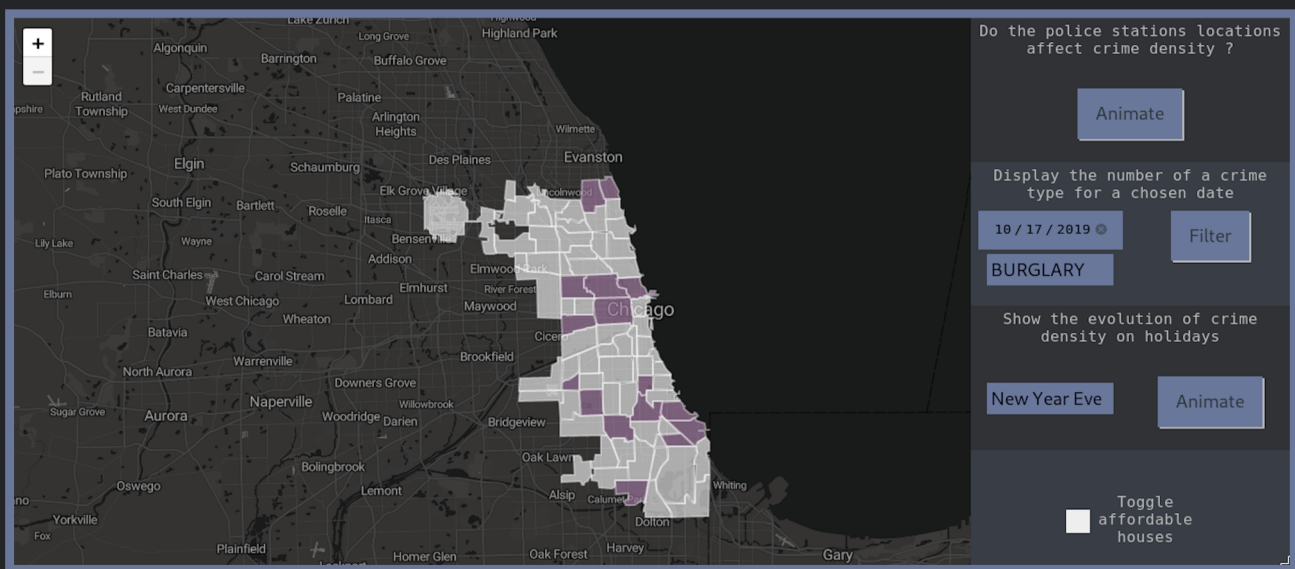


Figure 4: Burglaries in Chicago the day of Luca's birthday

The second feature we added, was the filtering by date and type of crime. As for the location of crimes, instead of using the heatmap once again, we chose to use the initial idea of using a more grid like delimiter. Having the socio-economic data, we chose to use the Community Areas of Chicago. Even if this would render the information less accurate than the heatmap, in addition to giving the user a different perspective, the Community Areas will also be used for the more detailed statistics afterwards, and thus having an interesting combination with the last feature. This feature is more of a sand-box feature, in the meaning it gives the user the power to freely navigate into the data, for example in Figure 4 we could observe how many burglaries there were during the birthday of our team member Luca.

The next addition to our map was the special dates. After seeing the result of the police station feature, the idea of using the heatmap to show a general idea seemed the best choice. But since the data displayed was for one day only, the transition time would need to be slower, giving enough time to the user to observe the evolution in time. We focused on well-known dates such as New Year, Christmas, Easter. A drop menu is used to let the user chose the specific date he wishes to display, the display will show evolution from 2001 to 2019.

Lastly, more as a bonus for the previous features on the map, we introduced a check-box to toggle affordable houses locations, to be able to observe if this has an impact on crime density.

Having the map to give a general idea of where crimes are more dense, and which crime was more frequent, the user should now start to want to have more detailed information on Community Areas of Chicago. We wanted to give him the possibility to explore these detailed statistics in addition to visualise the geographical data. This feature works very well with the filter of the map, since a pop-up showing the name of the Community Area is displayed, as shown in Figure 5.

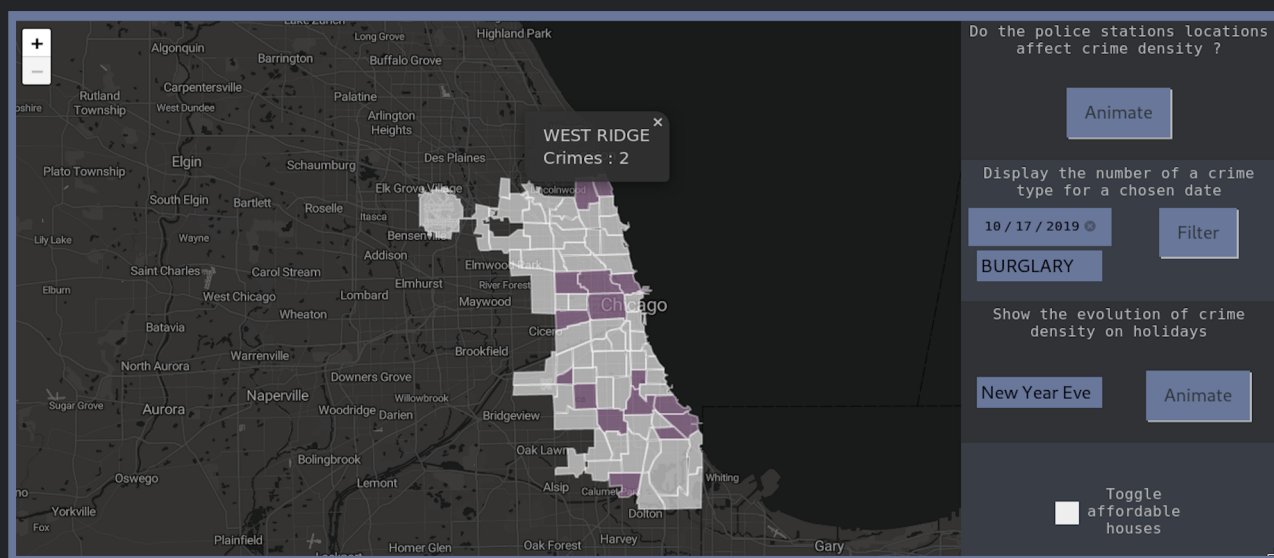


Figure 5: Popup showing Community Area name and number of crimes at the selected date

This last section can be seen as three parts, the central circle, which can be seen as the "controls" of the feature, and above and below it the statistical data. The circle is compromised of one inner circle being the year selection and the outer circle the Community Area selection. This gives the user a quick and responsive system to navigate through the different options.

Above the circle are displayed bars representing the types of crimes and the amount of the given crime for the chosen Community Area and year. As with the initial stream graph, this give a nice view of what types of crimes are more common and which ones are very rare. Underneath the circle are displayed interesting social statistics for the chosen Community Area. This allows the user to observe differences between Community Areas.

Once everything was in place, we attacked the projects visual identity, which colors to use, how to setup our layers and make the navigation pleasant to the user. Mostly focusing on dark colors, since our subject is not joyful, and even upsetting when seeing some results.

-The Challenges

In this section, the difficulties in our choices and implementations will be discussed.

The first issue we happened onto was with our dataset. The first problem was having many empty sections for the year 2001 and very little data for the year 2020. This led us to remove rows containing NaN values and remove the year 2020. For the year 2020 we chose to drop it because we wanted to have a general view of the data, so an incomplete year didn't seem to have its place in our project. As for the dropped values, we wanted to have data we could trust and easy to use, without the need of extrapolations.

Once this was done came another issue with the dataset, its size. Being really big was a storage issue as well as a speed issue. We chose to use our big data set to create smaller ones containing only the columns which were needed for the specific feature.

The map section had a few challenges as well. Firstly, the heatmap were a bit of a challenge, since the intensity of each point and the scales for the colors needed to be correctly chosen to highlight the desired message. As well as chosen a speed adapted to what needs to be shown.

Another issue was in arranging the user part of this section, having it not crowd too much the map, which would contain the display of the data, but at the same time not having it too compressed and unpleasant to use.

In the last part, the "control" circle were not an issue if taken alone since our design focused on non-exotic visuals, to maintain a sombre ambiance. But managing the display of all the data, and having all these elements to be aesthetically pleasing, while showing what needs to be shown, was a bit more challenging.

-Peer Review

Every member had to do some data parsing on python for the features developed by him. For the feature implementations and general workload, the separation was as following:

Martin Beaussart worked on the initial circle, stream graph, the statistical section and the visual part of the process book.

Vincent Cabrini worked on the map display, heatmap features, finalised to initial circle and finalising the website visualisation.

Luca Joss worked on the filtering and special dates as well as writing the content of the process book.

Everything was reviewed by the other team members and the choices were made together. Even if the main workload of a feature was usually done by one person, the other members verified and modified if needed.