Colorful Traces

Process Book

Davide Di Dio Tanguy Albrici Fatine Benhsain

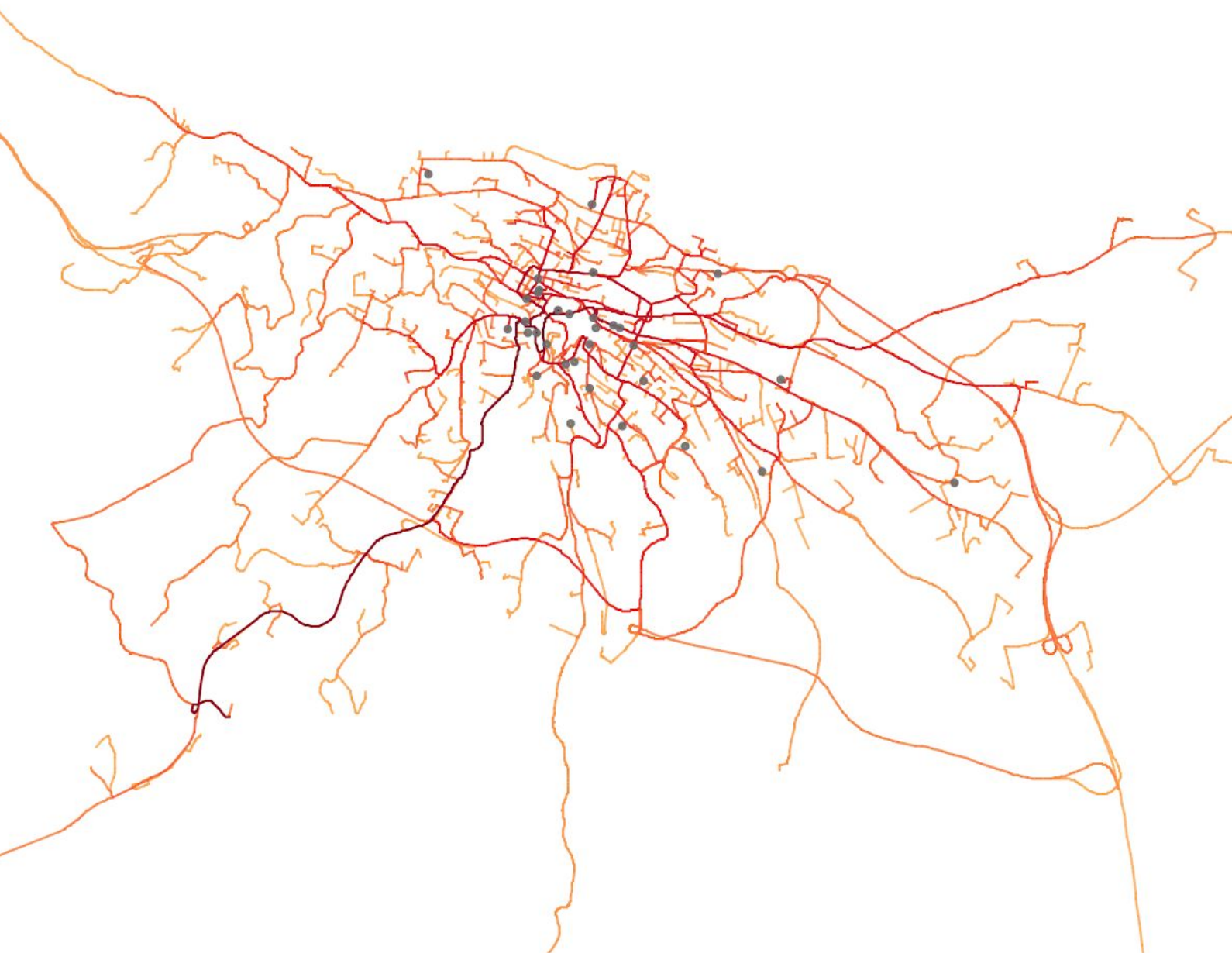


Table of Contents

Introduction

This is the project for the Data Visualization class (COM-480) of Fall 2018 at EPFL, in partnership with the food delivery company Smood. This process book is meant to describe and show the path that led us to our current visualisation.

Overview

This whole project is based on a dataset provided by Smood, which is a company that allows you to order your meal at restaurants near you and get them delivered to your home as fast as possible. It acts as an intermediate between you and restaurants, through the drivers.

Motivation

Startups like Smood are having a big growth and becoming very popular lately, because they combine convenience and technology. It can hence be very interesting to look at the way they work and the way they organise their drivers. This work can also give us insights on the distribution of the demand of food delivery around the city of Lausanne and its suburbs.

Target audience

Our project has a broad range of influence. Since we try to study the patterns and the distributions of the deliveries, there are many parties involved.

The first one is of course the company itself, Smood. Our work can provide a great source of material to exploit for them. By analyzing their drivers' and deliveries' patterns they can enhance their algorithms and provide a better service to their clients.

The users of the Smood service are also a target audience because they are the core of the dataset. The information contained in it summarizes their patterns and the way they use the service. This can also let them see the particularities of their neighborhoods: the existing restaurants/pick-up points, the most popular ones.

Moreover, the other important group of people directly affected by our data are the Smood drivers. Their patterns are the ones analyzed here. From our visualization, they can extract information about the most popular restaurants and/ or neighborhoods in case they can choose the places where they work.

Finally, restaurants owners are also part of the equation. On the one hand, the owners interested in being a Smood partner can use the insights to decide whether it would be profitable for them to

engage with the company. On the other hand, and more generally, all restaurants owners can benefit from knowing the habits of customers when ordering their food.

Inspiration

Our main source of inspiration when designing the heatmap was the Strava Global Heatmap, which is of course not quite on the same scale as our visualization, but it was still a helpful reference to have.

Concept

Exploratory Data Analysis

The dataset given by Smood consists of 2000 samples (i.e. deliveries), that each contain the pick-up and destination locations, the time at which the food should be delivered (as chosen by the customer), as well as a list of OpenStreetMap node ids corresponding to the itinerary taken by the person doing the delivery. After discussing with our contact at Smood, we were given additional metadata that contained the duration and distance of each delivery. Note that the data indicates that delivery are made for one meal at a time (probably intentionally by Smood).

The first step of preprocessing was to retrieve the latitude and longitude of all OSM nodes present in the dataset. To do this, we downloaded OSM data from the Lausanne area, extracted from this the locations of all the nodes present in our dataset.

Then, we wanted to see if we could cluster the pickup locations and identify from which restaurant the food came from. Surprisingly, we observed that there are only 32 unique pickup locations, and none of these seemed to overlap with others, thus clustering was not needed. In order to try to match these locations with a restaurant, we created a map showing these 32 points as well as the locations of the 70 restaurants in Lausanne listed on Smood's website. But this map showed that many pickup points were far from any restaurant, therefore it was impossible to find a reliable match for each pickup location using this method. After reaching out to Smood, we were told that restaurants Smood partners with are changing constantly, and that therefore the ones listed on Smood's website most likely did not match the ones in the our dataset. They were unwilling to share the names of the restaurants with us, but advised us to not restrict our search to restaurants listed on Smood website. However, we decided against doing that, because due to the high uncertainty of such an approach, it would most likely result in an inaccurate outcome.

By doing a similar kind on analysis on, we observed that they are only 528 unique destinations in the dataset.

Designs

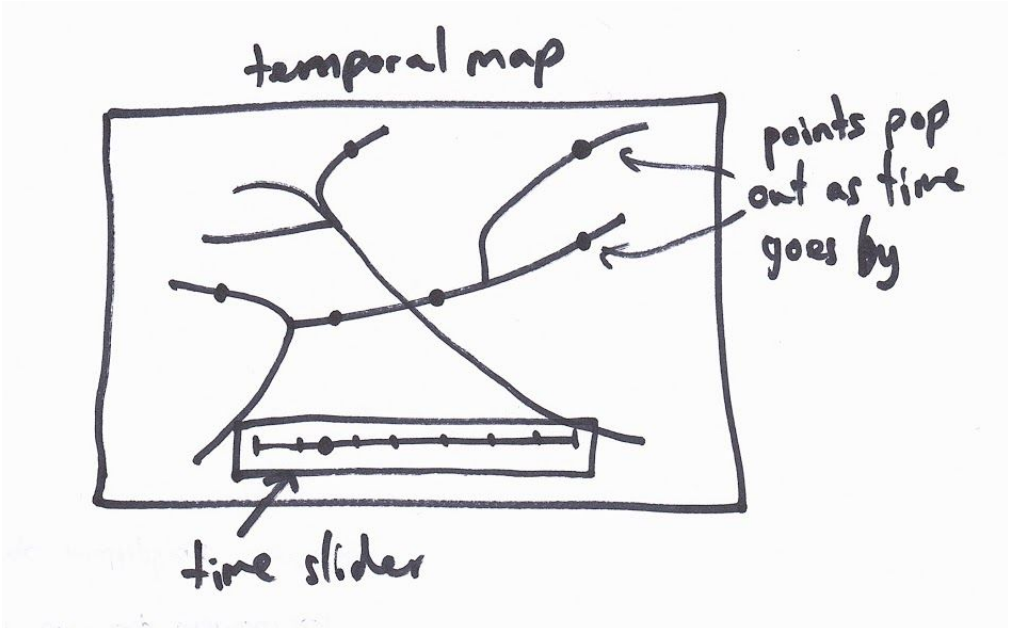
In this section we will go through our design process and the evolution of the different visualizations we considered by including sketches. We will separate it in three parts: first, our very initial concept imagined at the beginning of the project, then our exploration of graph-based visualization, before finally talking specifically about the ideas behind the design of our final visualization.

Initial Concept

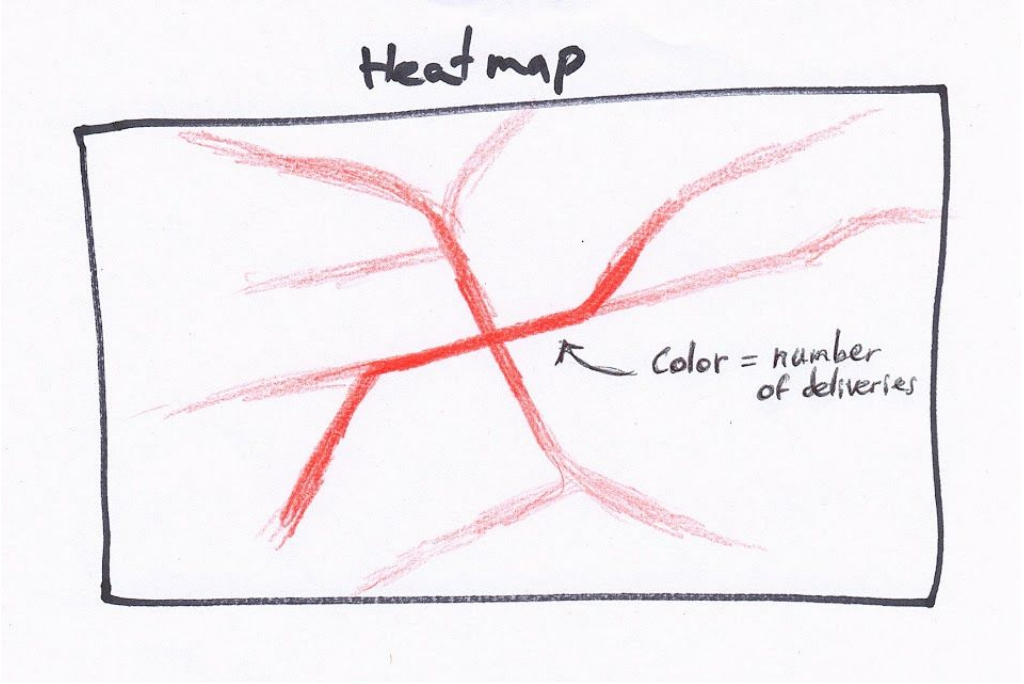
Initially, before even having access to the real dataset, we already had a rough idea of how the general structure of our visualization could look like. Our main idea was to have an interactive dashboard, with on one side a map showing the area of interest, and on the other side a panel that would display different types of insights for the area currently shown on the map. This is illustrated in the sketch below.

As for what would actually be displayed on the map, we considered differents ideas that we could possibly implement in parallel. Some of these ideas were based on data that turned out not to be part of the actual dataset, and therefore had to be dropped.

One idea was to have a temporal map that would show the most frequent delivery locations along the day. We would show an animation where points corresponding to delivery locations would pop up and disappear on the map as we brushed through a typical day for Smood deliveries in Lausanne. This is illustrated in the sketch below.



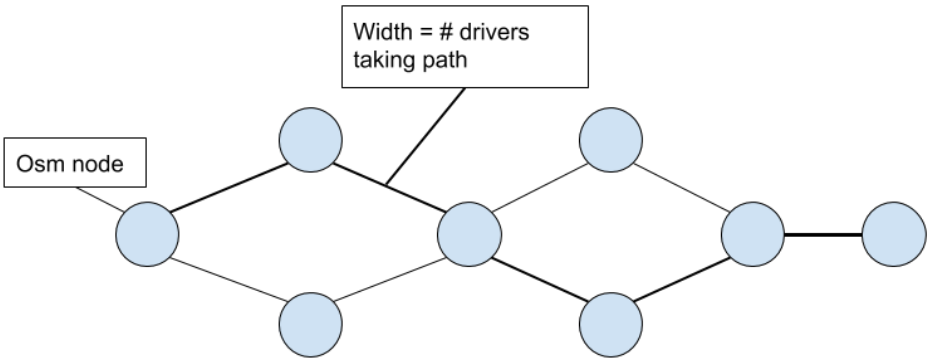
Another idea was to build a heatmap that displays how many times a road or a section of a road was taken by a Smood delivery man, by aggregating all GPS traces. This would allow us to easily identify frequently used roads and driving patterns.



Exploring Graph-Based Visualizations

After receiving the actual dataset, we were advised by our Smood contact to not restrict ourselves to maps, and maybe explore ways to represent the data using graphs. We therefore thought about graph representation based the OSM nodes in the dataset, where we would set the width of any edge relative to the number of drivers that went through this edge, as seen in the sketch below. We

figured that this was a good way to identify the frequently used paths without having the geographic constraint of a map, which would probably make the viz harder to read.



Another-graph based visualization we considered is to have a graph in a style reminiscent of a metro network map, where we would show the paths taken by the deliveries for each restaurant separately, as seen in the sketch below. By selecting a restaurant, we could thus highlight all the delivery made from that restaurant.

Design of Final Visualization

After spending some working in parallel on a heatmap as well as a graph-based visualization similar to the first one described previously, we figured that, in this context, a graph was not very helpful and didn't convey any information that the heatmap couldn't. Therefore, we decided to focus our efforts on the heatmap only.

To add some interactivity to our simple heatmap we decided add the possibility to select only some of the deliveries. We first had the idea to select the deliveries from a single restaurant by clicking on it. Then we took this idea further and added the possibility click on any point along a

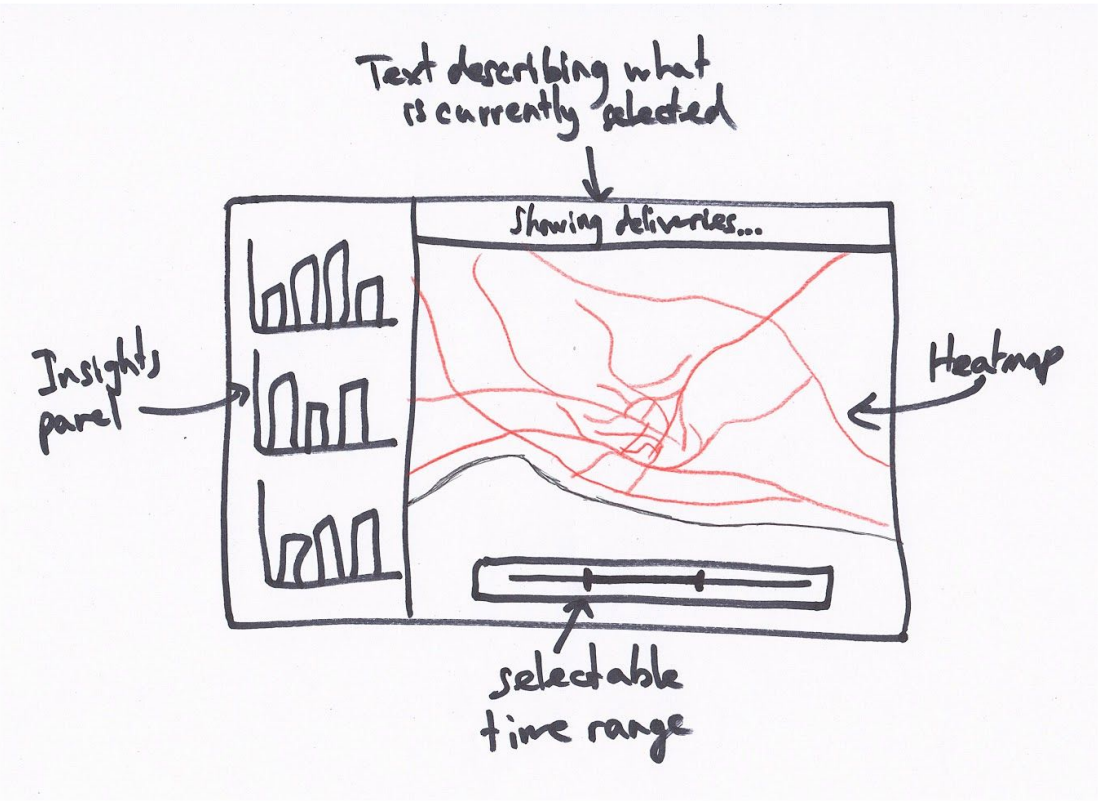
path to display only the deliveries that go through that point. This features adds a lot more unique ways to see the data and is very fun to play with!

Then, to be able to explore the temporal aspect, we had the idea to add a timeline that would allow to select a specific time range.

With all these possible types of selections, we were worried that the user would get lost and not know what he/she is looking at. To solve this, we added a text at the very top of the screen that describes clearly what is currently selected. If something is selected, a reset button next to this text is shown.

We also wanted to find a way to show other aspects of the dataset that couldn't be put on a map, so we decided to add a side panel, very similar to the on the one envisioned at the very beginning of our design process.

The different aspects of the design mentioned above are illustrated in the sketch below.



Differences from the initial proposal

When we first selected this project, we were still unaware of what the real dataset would contain, as the project description was a little vague. We therefore had multiple ideas that relied on different types of information such as the cost of each delivered meal, the routes taken by each individual driver or the time that each driver spends on the road. But after receiving the actual

dataset we realized later that some of these initial ideas were not feasible and thus had to be put aside.

But as it can be seen from our design process, our final visualization still has some ideas that are related to our initial concept, such as of course the heatmap, but also the temporal map. Moreover, even if we didn't end up choosing a graph-based visualization, we still implemented a way to select a single restaurant, an idea that we had initially for the graph visualization as explained in the last section.

Implementation

Technical details:

The main map implementation is built on top of the leaflet API. Most interaction are handled well but drawing multiple object on the map can be a bit unresponsive.
We use Bootstrap, Font Awesome JS and JQuery with plugins to ease the visual implementation.

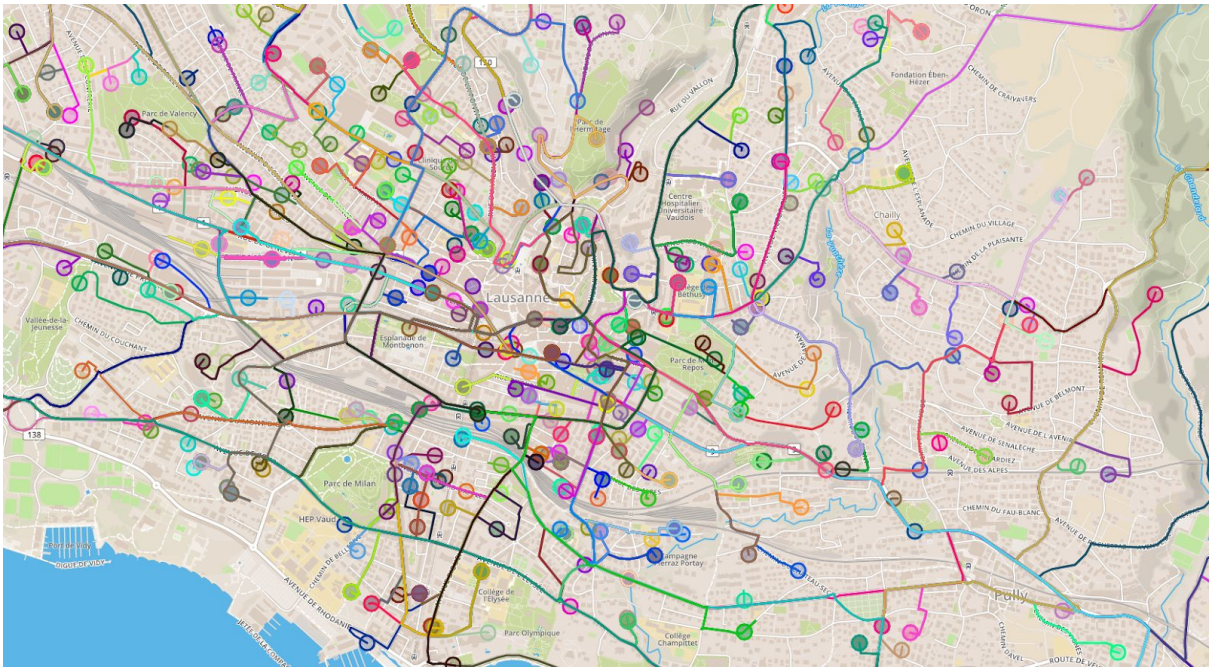
The graph visualization is made using Cytoscape.js (http://js.cytoscape.org).
The time slider is built on top of range-slider script:
(https://github.com/slawomir-zaziablo/range-slider)

The data is processed mostly in python to avoid unnecessary overhead on the user side (and also because we are more comfortable with the language).

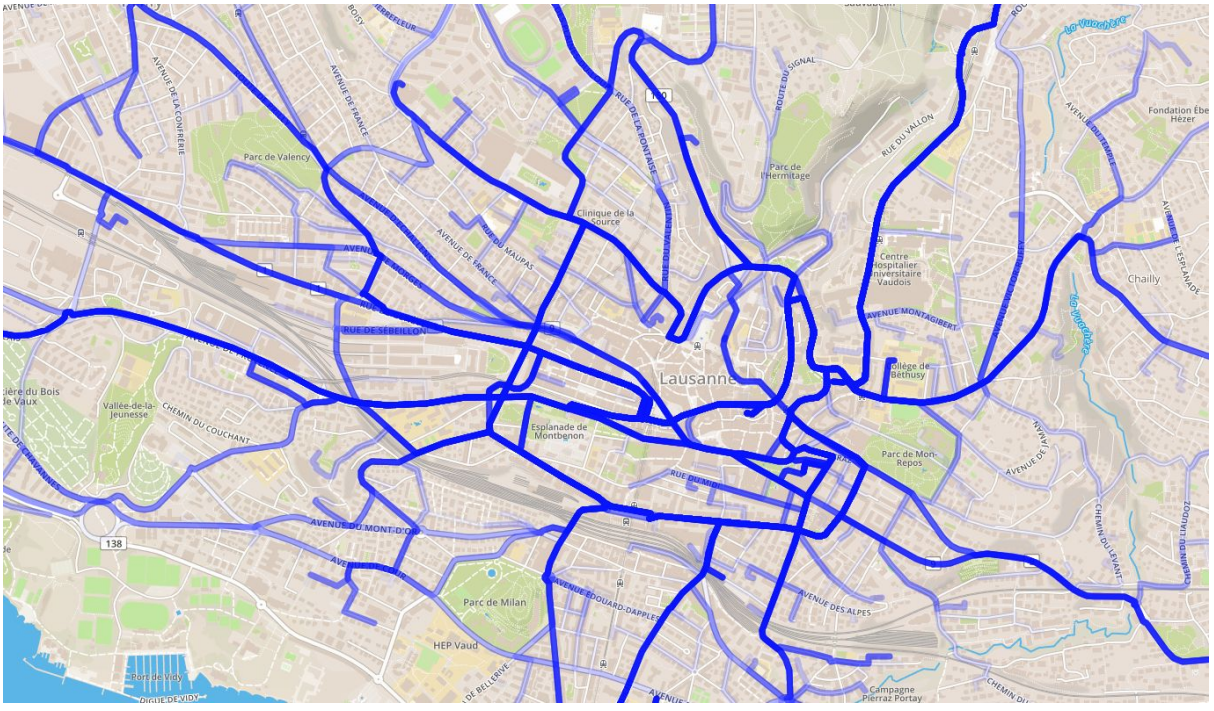
Finally, the website is hosted on github pages using jekyll as backend.

Implementation process:

To reach the final implementation, we followed an iterative process. At first we try two parallel implementations, one with a map and a second one with a graph representation of the data. The first step consists of collecting the position of each node from OpenStreetMap (OSM) (https://www.openstreetmap.org/).

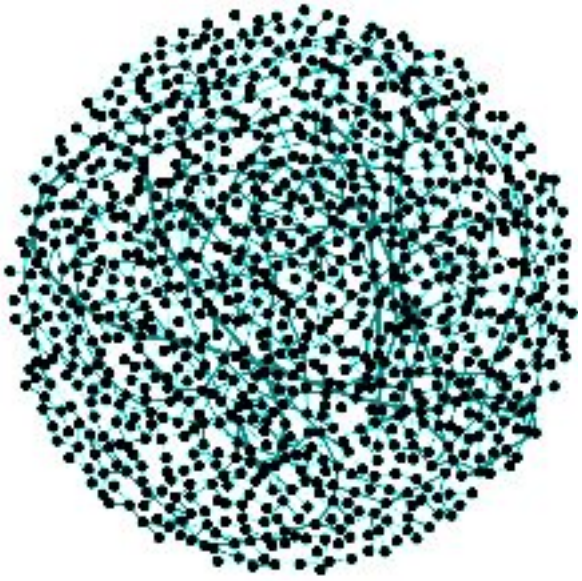


As shown above, the first result is simply a visual representation of every delivery path. The points are restaurants and delivery locations. The technique used is simple: We draw each path with the latitude and longitude of each nodes on top of any existing path. The traces are nice but for now the color of each path is random and the map is completely cluttered. The information that can be extracted from this visualization is minimal, we need to represent the density of paths in some way.

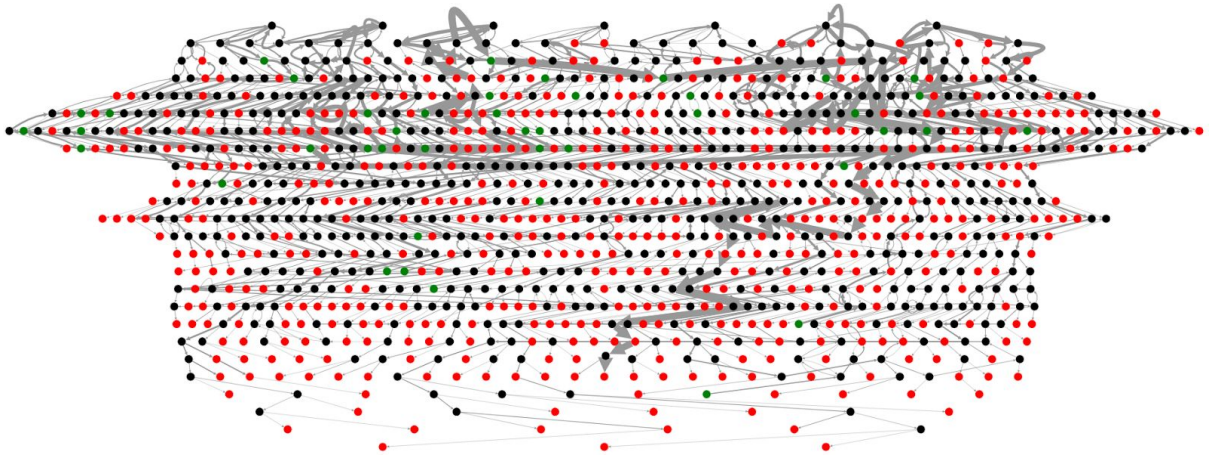


To overcome problems from the last visualization, we try to draw the paths in a single color and

with an opacity of 0.05, as seen above. This method shows the density of each path in a clean way. However this visualization is a dead end. We can not brush our data in any user friendly way with our current data representation and technique. No interactivity can be added to this design.

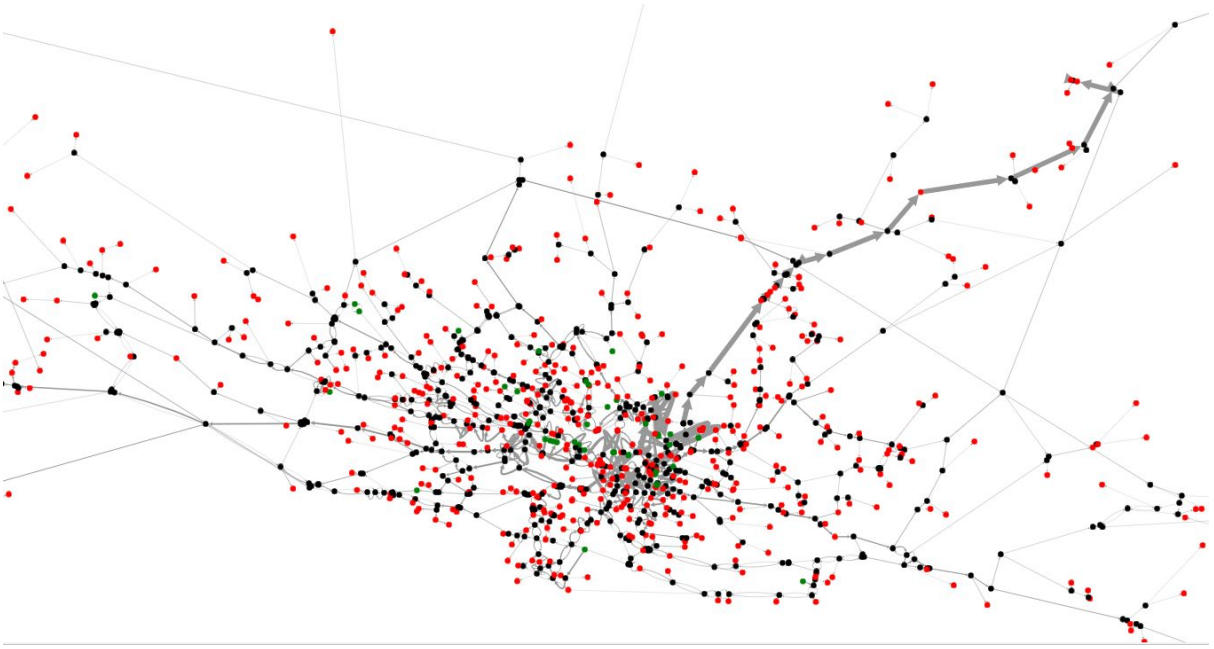


We went on a completely different representation and tried some graph visualizations. We wanted to abstract the position of the OSM nodes but still visualize their interconnectivity. The graph above shows a subset of OSM nodes. The size of the edges between two nodes is proportional to the number of drivers that went through them. Here the visualization is fun but too cluttered to make sense of the data. The graph used D3 interactive force layout which made the visualization feel like a bowl of spaghetti and chickpea soup. Tasty but not the best visually.



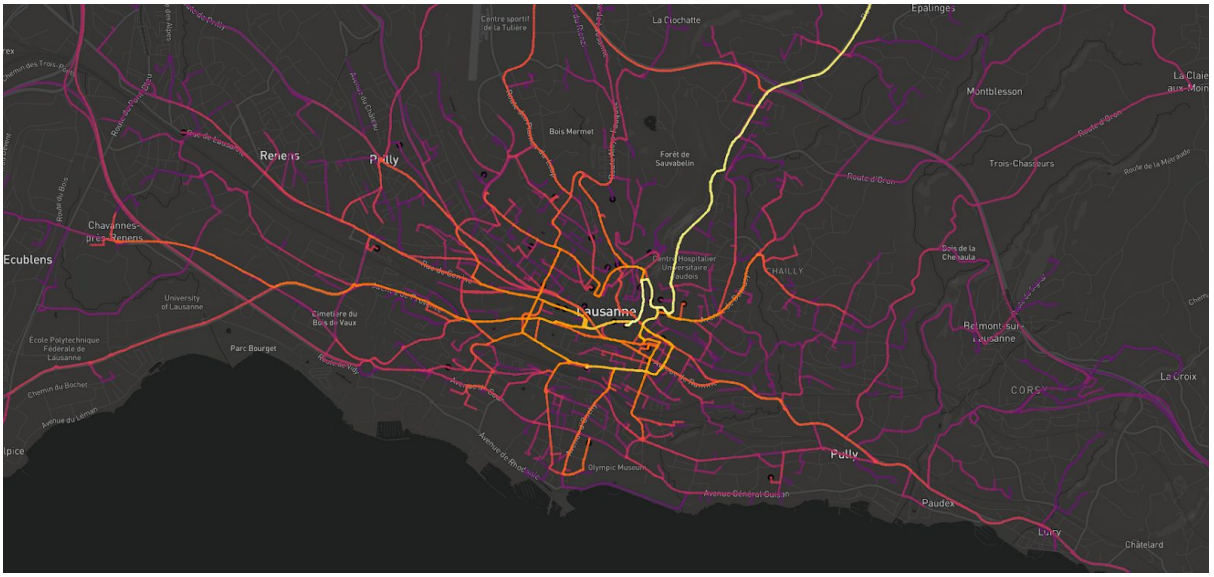
One problem is that too many nodes are on the screen. Most of the nodes in our previous visualization are irrelevant. Since we are interested in the interconnectivity of each OSM Node, we can collapse all the nodes that are the same as their neighbours. The placement of the nodes changed, now we use the Cytoscape breadthfirst layout and display the whole dataset. The result can be seen above. Even if we can see the main paths, the placement of the nodes does not make

any sense. We decided that it would be more intuitive for the user and it would reduce the clutter to place the nodes by their geographic location.



This isn't great, we tried other layout like circle or concentric but the interconnectivity of the nodes did not look good. We tried to add an options to hide restaurants (green), delivery points (red) and intermediate points (black) but none of these options actually provided any useful insight on the data. The most tricky part is to push nodes away from each other but we did not find any satisfying way to implement that.

We then turned full circle and, fortunately, the technique used to build the graph allowed us to improve our map visualization.



Now, we are starting to converge. This time a layer of proper polylines is built on top of the map, each polyline contains the array of ids that represent a delivery. We pick a colormap that respects a linear proportional change in lightness and remains aesthetic with our overall presentation (and that corresponds to Smood's color as a bonus). This representation allows us to build a sophisticated brushing on multiple dimensions of the data. We can create an index pointing to deliveries to display. We have one with every delivery times and one with every edges in the graph.

Final Implementation.

After an iterative process of implementation, we converged to a delivery exploration tool.

When you arrive on the website, a pop-up appears and proposes to explain the functionalities of the visualization step-by-step while showing some interesting insight (the best we could find with this amazing dataset). These pop-ups guide you through the visualization and encourage to try it out for yourself. We decided to go with an intermediate size on the story font since the goal of our visualization is focused on the exploration rather than showing any particular part of the dataset. Hence if a user is more intrigued by the map instead of the story at first, he can enjoy the map and read the story afterward.

The map is interactive, clicking on a path or a restaurant (black circle) filters only on the delivery that go through it. The user can zoom and move the map to focus on different parts of the dataset. We have a double range slider at the bottom that allows the user to select a delivery time. The play button automatically cycles every delivery times until paused or until the end of the day is reached. At the top, a message describes what is currently filtered. Aside this message, a reset button puts the map in its initial state.

The left pane contains three histograms of the deliveries by time, distance and duration. The histograms are updated with the current map filter.

If the user requires a clearer map, he can use the light theme in the option menu. The restaurants can also be hidden from there.

To polish the implementation, we add some eye candy to smooth the transitions. Animations on the polylines and histograms transform the visualization into a finalized product.

Evaluation

In the beginning the dataset seemed a little sparse, not having enough features, leading us to think that it would be hard to extract good visualization from it. However, we managed to exploit the data to its maximum and design the best visualization we could think of.

So, this experience let us learn that we can almost always adapt to the data we have, no matter how small.

The final visualization runs pretty smoothly, it is a strong tool to display any path data as a heatmap with the ability to filter very specific time and space window.

Discovered Insights

From the exploration of the dataset using multiple successive filters, we were able to gather the following insights.

First, most of the deliveries are done in the evening. There is a peak between 18h and 22h, corresponding to dinner time. There is a small peak at lunch time too but it seems very small in comparison with the dinner one. This may indicate that people prefer ordering food from their home.

Another interesting thing is that the most popular path is from Rue du Flon, in the center of Lausanne, to La Vulliette, near the École Hôtelière. In fact it so much more popular than all the other ones that it is strikingly obvious when looking at the visualization.

Additionally, most of the drivers travel less than 10 kilometers in their trip. This can be explained by the fact that restaurants usually only deliver to nearby houses or at least houses in close neighborhoods, and also that restaurants too far from your house are not shown to you anyway.

Furthermore, the trips are not very long. The longest takes 20 minutes, and most of them end in less than 15 minutes. We see two peaks, one around 5 minutes and another around 10 minutes.

Future work

Multiple improvements can be made on this visualization, with our current API it is easy to add new dimensions for the data brushing, we could for instance select on the length of the delivery or on the time required for a delivery.

Another great idea would be to have an additional panel containing the data of the brushed restaurants ordered by popularity. This requires data for each restaurant and we don’t have it but the concept is interesting and the additional insight would be very useful.

To create a stunning effect that also captures the direction of the deliveries, we could create an animation where some wave would flow from the pickup to the delivery point representing the driver at a given time.