

# Data visualization : Milestone 2

AJEGHRIR Mustapha  
mustapha.ajeghrir@epfl.ch  
(Sciper 333806)

Sorin Mircea  
sorin.mircea@epfl.ch  
(Sciper 306618 )

Simon Dayer  
simon.dayer@epfl.ch  
(Sciper 271991)

**Abstract**—The aim of this milestone is to provide to the reader a good understanding of the potential of our project. We will first explain the goal of this project by presenting the main part of this project. Finally, we will go through more challenging ideas that would bring a plus to our implementation

## I. GOAL

Since the last milestone, the goal of this project has been deepened. We started the implementation of the website that will give the user a good understanding of his/her habits through the collected data.

With this thinking in mind, we will first focus on constructing exploratory plots that will give a general summary over all the collected data (this includes both activities and sleep data).

Then, we will focus on finding different correlation and patterns in the activities made by the user.

Finally, we want to provide more interactive visualizations like for example:

- ability to see for each individual day a summary of all the activities done along with health related data and sleep
- ability to plot an interactive geo-heatmap of all the activities location data with the possibility to emphasise on different criterion (different sport to be colored differently, possibility to select a specific workout and emphasise it on the map and so on).

Because these weeks allowed us to put more work into this project, we have aimed to actually implement as much charts as possible (in the detriment of only making sketches).

## II. OVERVIEW OF THE USER DATA

The first challenging idea was to define what a sport oriented person would be interested to see (from a data visualisation point of view). So, after doing the proper research we came to the conclusion that an overview of all the activities is a must have.

### A. Initial geoplot

This is why we decide to first let the user see how its data looks on the world map. In our case, the data source of all visualizations have already been explained in the 1st milestone (it is collected by one of our team members).



Fig. 1: Activities locations

In this graph, we were able to experience what we learned in the Maps lecture.

In order to make this plot, we wrote a Python script that iterates through all the activities and fetches one location point per activity, then we appended these two extra columns (start\_latitude and start\_longitude) to the activities summary CSV that we already had exported from Strava.

This plot does not have any extra interactivity, also, focusing on only one geo point per activity is not really our goal (but a proof of concept) hence the reason to improve it in the next milestone.

### B. Activity types

Other information that will be of interest to the user is comparing different data points based on the type of activity: the time spent doing each type of activity, the consumed calories, relative effort and total distance.

In our case, we have a total of five activities (Ride, Run, Hike, Walk and swim).

Some of these charts are meaningful if the user tries to achieve certain goals, like losing weight, case in which he/her would be interested to finding out by which sport he/she lost the most calories in the shortest amount of time.

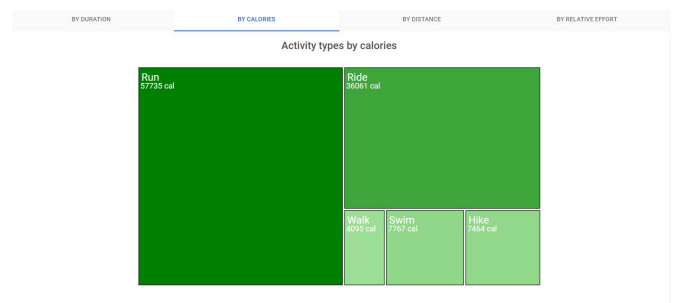


Fig. 2: Activities type

In this graph, with the support of the lecture Mark, channels, we used a tree map graph and the size is adjusted in proportion with the observed quantity.

These are actually a group of charts that have the same backbone code, thus, we are focusing on a modular implementation that allows code reusing.

### C. Sleep data patterns

We also want to present to the user an overview about his/her sleep, hence the next chart in which we have made a histogram of the sleep duration (hrs).

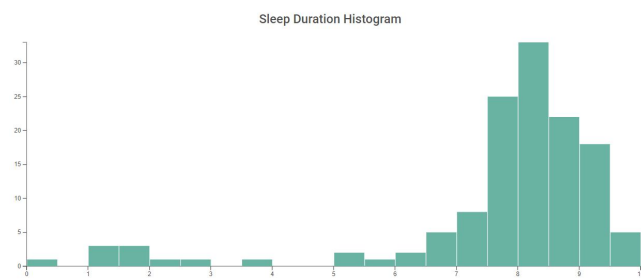


Fig. 3: Sleep duration

### III. FURTHER PLANS

Providing insights, certain correlations and patterns about how some activities can have an impact in a different part of one's life is one of the goals of this project.

As we don't have enough night runs a possible correlation between sleep and late runs isn't possible. Nevertheless, we will study the correlation between the amount of deep sleep / the number of sleep cycles versus the quantity of physical activities done in that day (and other similar metrics).

For plotting correlations between two numerical variables we want to use a 2d heatmap that is going to look like this:

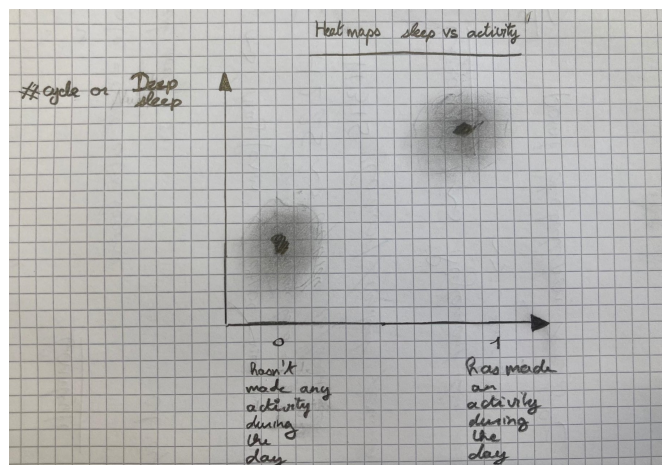


Fig. 4: Sketch heatmap

Lastly, we find interesting if the user could know where are his favorite training tracks. So we plan to generate a geo-heatmap that will look like similar with the following:

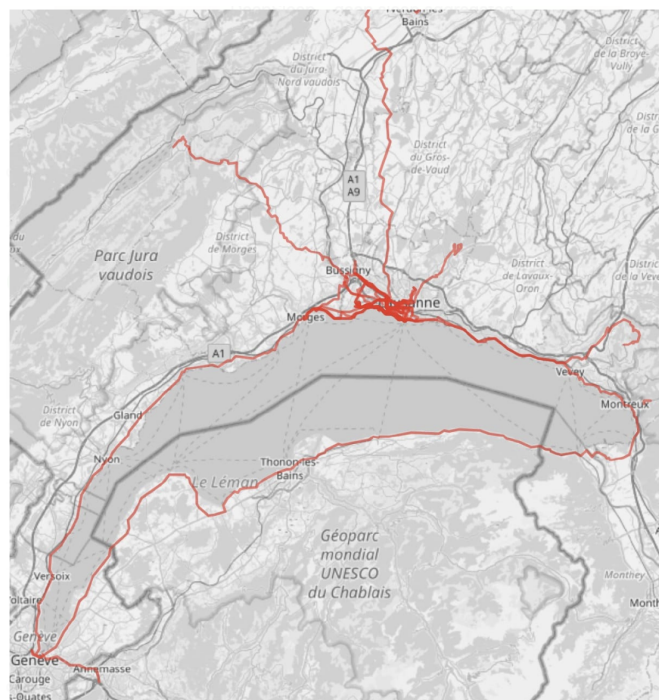


Fig. 5: Geo heat map

The above chart is plotted using the data that we have available, but it resides outside the scope of our project as it is generated by a 3rd party project (but it shows us what we can get to).

### IV. MAIN VIZUALISATION

In this section, we will present our main vizualisation that we want to make more interactive. Instead, we will take as a first support a calendar.

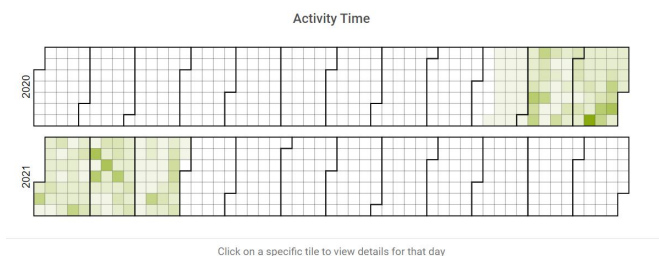


Fig. 6: Activity days

As you can see on the sketch below, when the user will want to see his data for a special day, some cool visualizations will be accessible directly.



Fig. 7: Activity days

## REFERENCES

- [1] Website, 2021. *Personal website of Andre Salyer* [online]. Found at this address: <https://www.andresalyer.com/nike-heat-map/>
- [2] D3Js examples website: <https://observablehq.com/@d3/gallery>
- [3] Activity days chart template <https://github.com/DKirwan/calendar-heatmap>