

# DataViz project: Lost and found at SNCF

ShangNong Hu\*

Sebastien Delolme-Sabatier†

Valentin Reymond‡

## 1 INTRODUCTION

This project was realized during the classes of Data Visualisation from University Claude Bernard Lyon1s Bio-Informatics Masters Degree.

The objective of this project was to visualize data obtained from SNCFs public data about. With this project, we wanted to answer some questions such as What are the train stations where we report the most lost objects ? What are the train stations where we report the most found objects ? Are lost objects and found objects the same or is there a difference between them ? In which train station we have the highest chances to find the objects that we lost ? Is there a relation between the number of lost objects and the influx inside a train station . We will focus on the train stations in France.

In this document, we will first introduce the datasets that we are going to use to realize this visualization, and next, we will present our design ideas.

## 2 RELATED WORKS

SNCF published among of data in open-access, such as train frequency. In our study, we focus on few datasets:

- **Lost statements** (661843 entries) [3]: this dataset contains informations about the type and the nature of the lost objects, the date of the report and the train station in which the object has been reported (with its name or its id).
- **Founded items** (396 941 entries) [4]: Such as the preceding dataset, this dataset contains informations about the type and the nature of the item that has been found, the date and the place it was found and an indication if the object has been returned or not
- **Train stations in France** (5032 entries) [2]: this dataset contains multiple informations about train stations in France. In this dataset, we can find things such as the station id, the city in which it is located, its geographical coordinates, the type of trains they receive, etc...
- **Influx by train station** (3040 entries) [1]: this dataset contains the informations about the train station (id, place) and the number of travellers from 2016 to 2014. Note that travellers and non travellers are distinguished.

These datasets are well-formated and exist in CSV or JSON format for export. They are free-access but we should notice SNCF via a form for each one if we want to use them. We will combine them into a multi-dimentional display powered by D3.js. First, each dataset will be filtered and parsed. Then we display their representation in a HTML page.

\*e-mail: shangnong.hu@etu.univ-lyon1.fr

†e-mail: sebastien.delolme-sabatier@etu.univ-lyon1.fr

‡e-mail: valentin.reymond@etu.univ-lyon1.fr

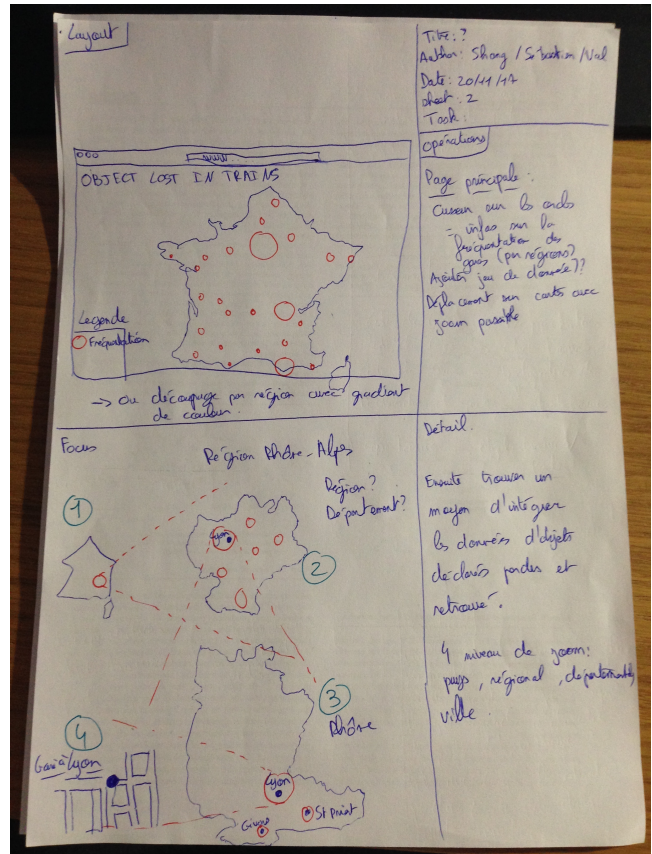


Figure 1: Design sheet number 2.

## 3 CONCEPTION

The visualization work will be done thanks to D3.js, the JavaScript library for producing dynamic and interactive data visualizations in web browsers.

Multiple ideas of designs have been explored. We decided to choose the one that, in our point of view, will allow to best describe the data. The window will be represented the way shown in fig 1. The user will first have a map of the France with train stations represented on it. He then will have the opportunity to click on one of the regions of the France which will allow him to have a detailed point of view as presented in the focus part. Finally, a click on a specific train stations will provide a tool-type object which will give more informations to the user.

Even If, we chose this representation, we considered an other type of representation such as described in fig 2. In this representation, the data are represented through the time axis. The user will be able to click on a specific train in a specific year and he will be able to visualize the data from a more precise point of view (the lost and found items by years, by weeks or by item type for example). As for the first representation, tool tips would be displayed when the user hover his mouse on a part of the diagram such as details on the train stations in which objects were found.

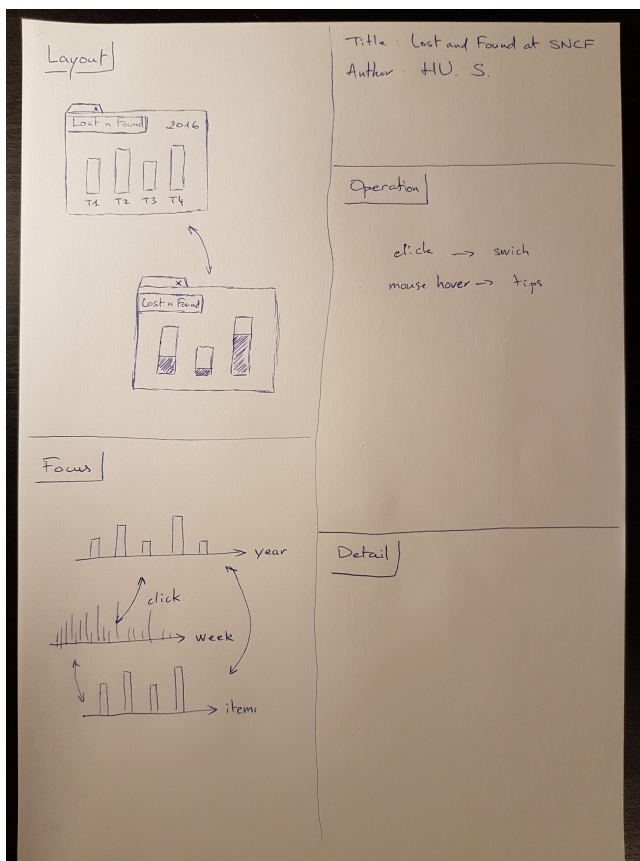


Figure 2: Design sheet number 3.

## REFERENCES

- [1] <https://data.sncf.com/explore/dataset/frequentation-gares/>.
- [2] <https://data.sncf.com/explore/dataset/liste-des-gares/>.
- [3] <https://data.sncf.com/explore/dataset/objets-trouves-gares/>.
- [4] <https://data.sncf.com/explore/dataset/objets-trouves-restitution/>.