Comparison of the quality of TER flows between the 20 SNCF-regional agreements between 2013 and 2017

HAMIZI Naama MOUSSI Ibtissem MEZIANI Amina

Université Claude Bernard Lyon 1

ABSTRACT

TER means Regional express transport and designates a category of trains allowing intra-regional transport. This appellation excludes the Ile de France, which for its part has what is called the "Transilien". The comparison of TER flows between 2013 and 2017 makes it possible to judge the quality of this type of transport by region and at national level. The definition of criteria such as the number of canceled trains or the number of trains that have circulated versus planned trains provides an objective portrait of the properties of TER networks in France.
The visual analysis of these properties can be useful for the detection of anomalies or improvement tracks of these networks. The article therefore establishes the uses of these visual methods within the existing as a first step and the methodological approach and techniques discussed in order to set up this analysis.

Keywords: TER, public transport quality, visualization, quality indicator, chorplethhe map, dorling cartogram.

1 INTRODUCTION

Our research in the field of the quality of public transport services in France allowed us to discover the creation in 2012 of the "Authority of the quality of service of the transports" (AQST) organization whose role is to analyze and to shed light on public transport in France of the air (planes) or land (trains, trams, buses … etc.) type.
Moreover, we found the data allowing us to consider our study with the SNCF. We will present in this section the two major sources of data for this project.

2 COLLECTION OF THE EXISTING

Our research in the field of the quality of public transport services in France allowed us to discover the creation in 2012 of the "Authority of the quality of service of the transports" (AQST) organization whose role is to analyze and to shed light on public transport in France of the air (planes) or land (trains, trams, buses ... etc.) type. Moreover, we found the data allowing us to consider our study with the SNCF. We will present in this section the two major sources of data for this project.

2.1 Transportation Quality Authority (AQST)

An analysis of the content offered by the AQST via its website [2] shows that the quality of transport was approached by type (bus, aircraft, trains, Transilien, TER). However, the data open to the public is presented by year. The user of this platform can therefore access the quality indicators (regularity, delays) by year and by type of transport. The proposed visualizations make it possible to compare the 20 regional-SNCF conventions as shown in Figure 1 or to apprehend the overall evolution (national level).

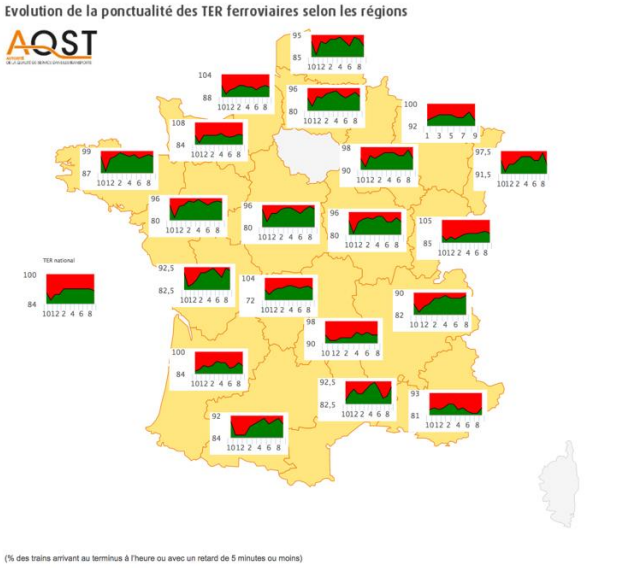

Figure 1: An example of visualization found on the AQST website

2.2 SNCF data

The main source of data for this project will be the SNCF site [3]. The proposed data express the monthly regularity of TERs by region and by month. These data range from 2013 to 2017.

2.3 Community Transportation Innovation Club

In order to define the indicators that we want to implement in this project, the website of the Territories and Cities Technical Department [4] provides insight into the approach to be followed. The role of this site is primarily to provide leads that can lead to innovations in the field of public transport in France.

2.4 Existing visualizations

We also explored the visualizations that can be used be during the project as a base, especially for maps:

- The evolution of influenza from 2014 to 2015 in France [6]. The interest of this presentation is the taking into account of the temporal aspect similar to what we want to do but over 5 years and per year and not per week.

- Analysis of the evolution of various parameters on the French territory over several years. The code is also open and accessible to the public allows the exploitation of the card in d3.js among others.

3 TECHNICAL POINT OF VIEW

The tools we have used are:

1. The JavaScript framework D3.js: this choice was justified by the flexibility that it offers as well as the fact that we present our results via a web interface supposed to be interactive. The version we used in this project is v4.

2. A Bootstrap template centering visualizations vertically allowing us to present our ideas and visualizations incrementally. In addition, the web interface is not the heart of the project we opted for this solution to devote most of our time to optimizing visualizations.

3. Microsoft Excel to cut the data into adapted and different files from one visualization to another.

4 DESIGN TRACKS

The implementation of such a project requires the use of a methodology [4] to decide the transport quality indicator. This first step will allow the implementation of relevant visualizations thereafter.

4.1 Definition of the transport quality indicator

Regarding the choice of the quality indicator of TER flows, there are two ways of looking at things through the data we have collected on the SNCF website. Indeed, the data guide our choices in the sense that the two main variables are the number of planned trains and the number of trains actually circulated. From there two figures emerge:

- A national average that can serve as a reference.
- An average per region that will allow the regions to be compared with each other or with a national reference system.
- These two averages are expressed as follows:

$$mean = \frac{canceled\ trains\ +\ delayed\ trains}{sum\ of\ trains\ per\ unit\ of\ time} \quad (1)$$

4.2 Types of visualizations

The goal of this project is to be able to extract information in a visual way, it is important to imagine the different possible visualizations that can allow this.

4.2.1 Choropleth Cards and Dorling Cartograms

Intuitively and when one is interested in data relating to a geographical area one of the first ideas to be exploited is to project the data on a choropleth map (France in this case here as shown in the figure [1]).
The strong points of this type of representation are:

- The ability to view data in a comprehensive and synthetic way
- The possibility to make comparison of regions with each other and with the national reference system

In our case, the statistic represented in the map is modeled by a gradient of colors and a label obtained on the overflight of the region with the mouse.

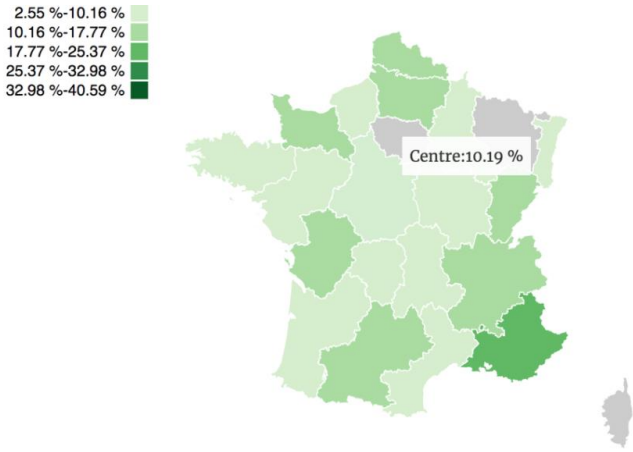

Figure 2: choropleth map allowing the visualization of the indicator of quality by region according to a gradient of colors.

Another way to visualize the variation of the quality indicator between regions is to partially detach from the geographical dimension and to play on the size of the region according to the statistics. This type of visualization is called “dorling cartogram” by the name of their inventor [5]. By avoiding the geographical form, we opt for circles for example as shown the figure below:

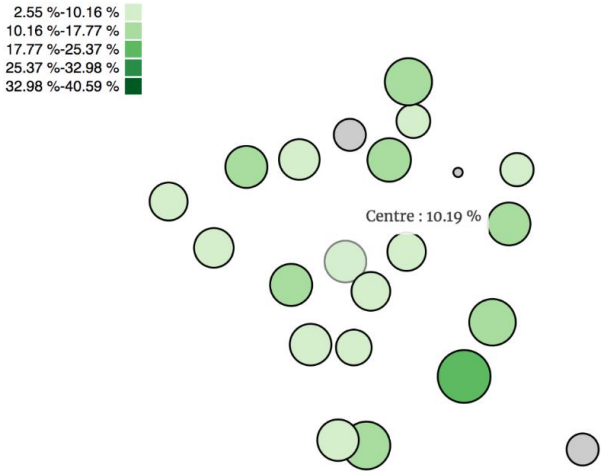

Figure 3: dorling cartogram with circles allowing the visualization of the indicator of quality by region according to a gradient of colors.

The larger a darker is a circle more the number of delays within that region is high. The statistic is always displayed when you move the mouse over the region. The user can therefore within this first approach visualize the evolution as a function of time, and appreciate overall the evolution of the number of delays per year.
The more a map is very green or made up of relatively large circles, the greater the number of delays. The time it takes for the visualization to interpret is therefore relatively limited. After analyzing these visualizations, the user is supposed to have:

- Perceived the evolution of the delays for a given region over the 4 years analyzed
- Compares two given regions over the same period.

In a second phase, it is possible to establish a comparison over two different periods by exploiting the previous choropletes as shown in Figure 4:

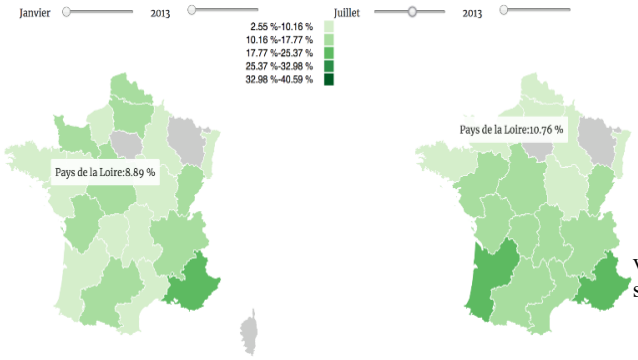

Figure 4: choropleth map allowing the comparison of the indicator of quality by region over two different periods

From there the user is able to compare several regions (the one of his journey for example) over several periods. This second phase makes it possible to give more importance to the temporal dimension.

4.2.2 Histograms and proportion circles

Beyond the inter-regional comparison, we decided to offer the user the possibility of comparing the regions' delays to the national average (of the 20 SNCF regional agreements).This approach makes it possible to detect the regions having a strongly divergent behavior positively or negatively compared to the global average. Thus, a region with a higher than average rate of delay may require further study to discover why this anomaly exists. On the other hand, a region with lower rates of delay than the average may be the subject of the same study for the benefit of other regions, for example in order to improve the national average regarding delays or punctuality of trains.

The figure 5 bellow shows the combination of the two types of visualization used (map, histogram) to highlight or to quantify our indicator of quality in each region according to a reference value, which is the national average of delay:

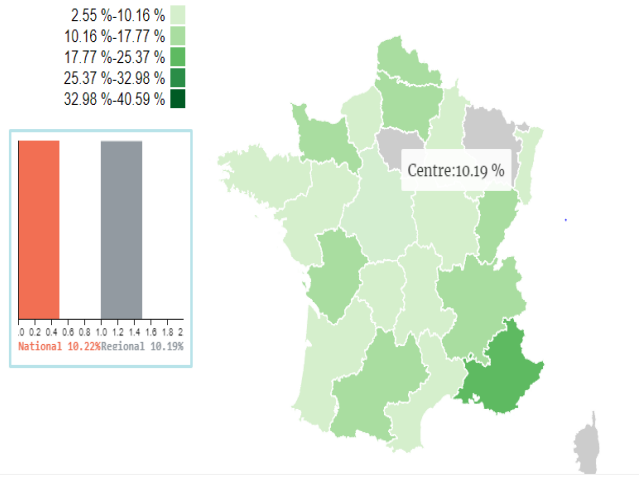

Figure 5: Histogram shown the value of the indicator of quality compared to that of the national average

The figure 6 bellow shows the combination of two other types of visualization used (dorling cartogram, proportion circles) for the same purpose but without considering the geographical view:

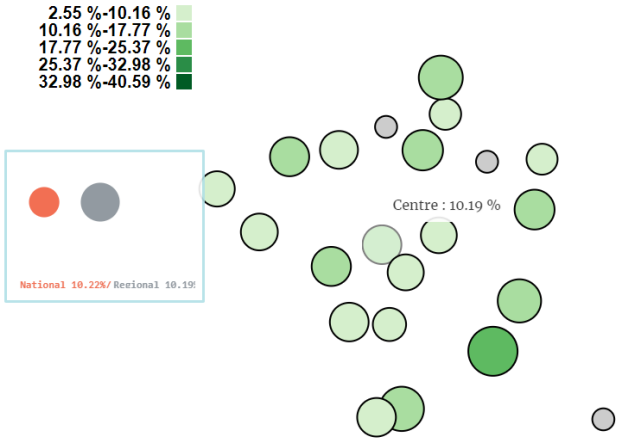

Figure 6: proportion circles shown the value of the indicator of quality compared to that of the national average

4.2.3 Line charts

Contrary to the maps, which allow comparing the regions with each other and expressing in a limited way the temporal dimension, the diagrams make it possible to highlight the evolution in time in a more expressive way. Given that our data range from 2013 to 2017, we can analyze the trend in the quality of TER flows during this time period (by year, or nationally or regionally). At the end of this last stage of the project, the user will have the possibility to:

- Compare all 4-year trends across all regions and the national trend
- Select by successive clicks to display trends related only to the regions that interest him (those of his trip for example).

The figure 5 illustrate what it was described below:

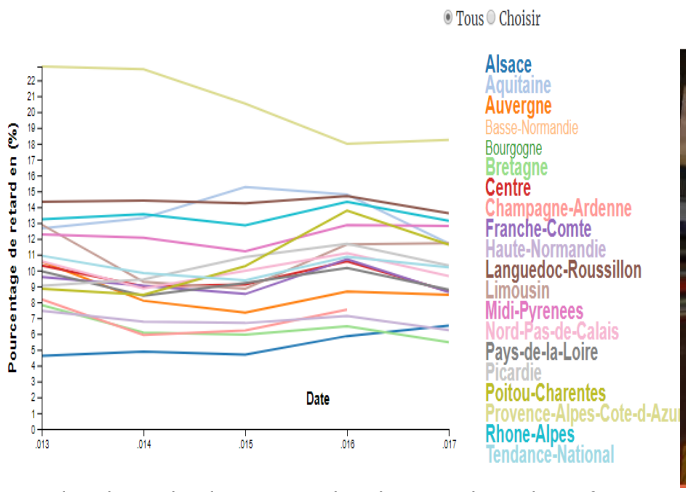

Figure 7: Line charts showing a comparison between the regions of the evolution of the indicator of quality by year

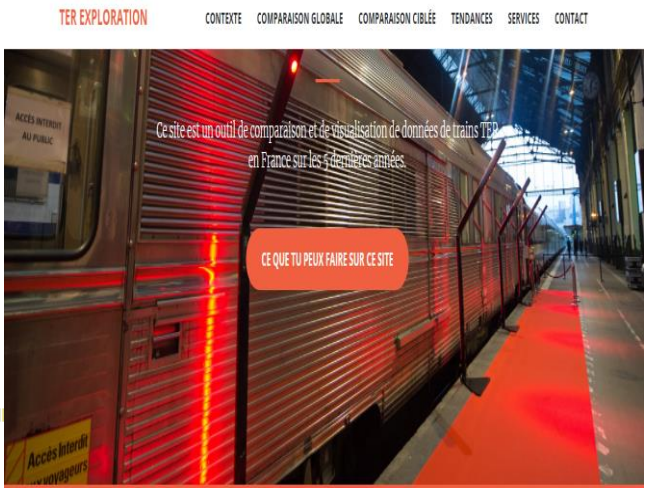

Figure 8: Site template view

5 USER PROFILE

The user targeted by this study can be:

- Any user interested in train flows seeking a resource that can popularize the behavior of TER train flows.
- A person working within the SNCF for example responsible for the analysis and improvement of flows within the TER network.

In the case of the second user, these visualizations and the interpretations they allow can serve as a basis or preliminary work that can lead to further studies involving other datasets (weather, social movements, acts of terrorism ... etc.).

The layout of visualizations within the web interface of the project makes it possible to develop a reasoning that satisfies both users with technical profile or not.

The need may be different but the logic followed to organize the visualizations allows using with a better understanding of the network of TERs in France on several levels:

1. Geographic
2. Temporal
3. Global or localized to a region.

6 OPTIMIZING THE SPEED OF COMPREHENSION OF VISUALIZATIONS

In order to accelerate the process of apprehension of the visualizations we have played on several HMI (human-machine interface) parameters:

✓ The use of a bootstrap template [9] to center the visualizations by scrolling them vertically, which suits the reasoning we want to see the user follow.

✓ The use of graphic components providing immediate visual feedback such as slidebars, buttons setting the display accelerate the display of the information sought by the user.

Figure 9: graphic components used in the project site

✓ Mouseovers and clicks allow the user to hover over information faster or set it for further comparison or analysis.

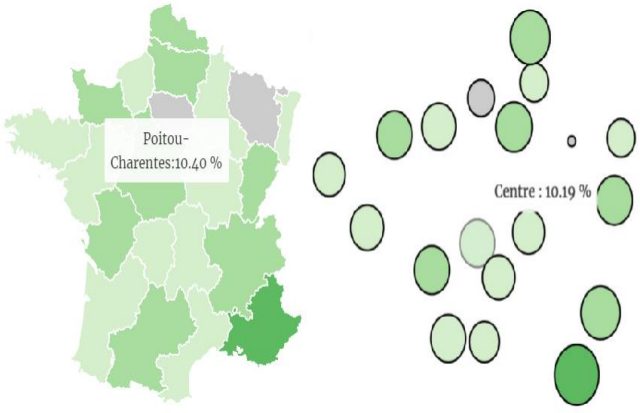

Figure 10: Different information shows using Mouseovers and clicks

7 PROSPECT

To stay in the logic of the project we can imagine a follow-up to this project. Indeed, if the three parts described in this article lead

to the highlighting of national or regional trends, the user could want to push the analysis more in depth.

To do this one could imagine, and still in line with the reasoning followed until then, parties involving the use of additional data (such as weather data, social movements, terrorist attacks, vacation plans, ... etc.). These data can enlighten the analyst about previously detected trends.

In addition, a question could emerge and would aim to question the resilience of the TER network in the face of rare events (natural disasters, terrorist acts, etc.) or foreseeable events, but with significant impact (school holidays, summer period, and return to school / academics).

Moreover, and always with a view to greatly deepen this analysis thus begun, the analyst could imagine an approach involving a level of geographical granularity (by cities for example, or even by stations) and / or temporal (days, hours ... .etc.) as a basis for forecasting work. Predictive visualizations could be imagined to report the state of the network in real time, for example. The idea here is to combine predictive algorithms (machine learnin so) and data visualization.

Thus, a more in-depth work would lead to combining different areas of analysis to generate relevant visualizations and informing the user as much as possible and in more depth.

Finally, a further analysis of the GUI of the presentation interface would be necessary to improve the speed with which a user could apprehend and understand the phenomena explained by the visualizations. The user must finally be part of the objectives of the various works cited here.

8 Conclusion

Using different visualizations can vary the impact of geographical, temporal or global viewpoints or more targeted to a region or a path.

The combination of graphic elements such as slidebars, buttons, mouse movements and clicks with visualizations allow the reduction of the time needed to extract information from them.

Thus, the addition of interactions between the visualization interface and the user thus allows a more flexible interpretation of the results in terms of time and semantics.

Finally, it should be noted that, unlike the existing one, our analysis will allow an analysis over a time interval greater than 12 months (from 2013 to 2017).

References

[1] https://fr.wikipedia.org/wiki/Transport_express_r%C3%A9gional

[2] http://www.qualitetransports.gouv.fr/

[3] https://ressources.data.sncf.com

[4] http://www.innovations-transports.fr

[5] https://www.arcgis.com/home/item.html?id=b686a7679cb747e9825d1d1bb6b26046

[6] http://www.innovations-transports.fr/Mesurer-la-qualite-de-service

[7] http://abenrob.com/Grippe-Narrative/

[8] https://openfrance.lightjs.org/#energie_eolienne/2012/regions

[9] https://startbootstrap.com/template-overviews/creative//

[10] https://lyondataviz.github.io/teaching/lyon1-m2/2017/

[11] https://bl.ocks.org/veltman/33dbad5aa12d92b977fd29128eb5358e

[12] https://bl.ocks.org/mbostock/4055892

[13] http://blockbuilder.org/aurelient/5bb9210591eb86882612a2002faab698

[14] http://blockbuilder.org/romsson/57eb273e01761b829aaa95b28c8193c1

[15] http://blockbuilder.org/romsson/c607174fc6fd5c7bfaa804d46ca70eab

[16] http://bl.ocks.org/JessicaFreaner/8fb0ac6c12aa1dab5f70

[17] http://bl.ocks.org/d3noob/e99a762017060ce81c76

[18] https://mbostock.github.io/protovis/ex/cartogram.html