The Swiss Transportation System
Process Book

Introduction

This project undertook the challenge of designing and developing an interactive visualization platform with a focus on the Swiss transportation network. The initiative sprang from the need to provide an accessible and comprehensive understanding of the vast array of Swiss transportation operations, encompassing trains, buses, and trams, with an emphasis on offering users enhanced insights into schedules, routes, delays, and passenger traffic.

Our primary objective was to offer a tool that served railway enthusiasts, the general public, travelers, and commuters alike. The platform was designed to shed light on the reliability and punctuality of Swiss trains and other modes of transport, revealing recent delay data in a visually compelling way. Beyond that, we aimed to expose the popularity and utility of various stations and routes, allowing users to discern trends in passenger traffic and make more informed decisions about their commuting options.

Our approach excelled in its innovative combination of multiple modes of transport into a singular platform, offering an integrated perspective on the Swiss transportation system. The uniqueness of the platform lies in the strategic implementation of a network and map visualizations, which beautifully capture the structure and topology of Switzerland. In addition to this, we integrated an interactive table and other data visualizations, providing a powerful toolset for users to compile, examine, and understand the data. These visualization strategies facilitate the exploration of the intricate hierarchy and interactions between different transportation systems, encapsulating the essence of our original approach.

In this process book, we will walk you through the journey of creating this platform, shedding light on the path we took to achieve the final result, the challenges we faced, the design decisions we made, and the roles of each team member. We hope this narrative provides a comprehensive view into our project, highlighting the importance of each step in creating a tool that marries the informative with the innovative.

Data and Preprocessing

In the initial stages of our project, the data preprocessing was a significant component, as it laid the foundation for the subsequent visualization and interpretation of the data. We relied on the comprehensive data provided by Open data platform mobility Switzerland[1], a customer information platform for Swiss public transport and individual mobility that generously offers mobility data free of charge.

Our selected dataset, a considerable collection at 100GB, comprised an individual entry for each journey taken by a transport vehicle from one stop to another, detailing specifics such as the date, delay, and transport type. A decision was made to limit our data to normal weekdays from the year 2022, excluding weekends and holidays, which further streamlined our data focus.

The magnitude of our dataset presented the first significant challenge. Handling 100GB of data required robust computational resources and thoughtful strategies for data management. To this

[1] https://opentransportdata.swiss/en/dataset

end, we utilized Spark and various data compression methods to process the data efficiently.

Our preprocessing pipeline was custom-built to clean each daily file used in our analysis. Establishing this pipeline was a journey in itself, punctuated by various challenges. It was crucial to ensure the accuracy and logical consistency of the entries, such as confirming that vehicles arrived at a given stop before their recorded departure time, and that no vehicle reported an implausible average speed of 2000 km/h. See figures 1 and 2. It was also essential to reconstruct entire train journeys using only the information about stops at specific times, a task that required meticulous attention to detail. These challenges tested our problem-solving skills and capacity to work with large, complex datasets, ultimately strengthening our approach and yielding valuable insights.

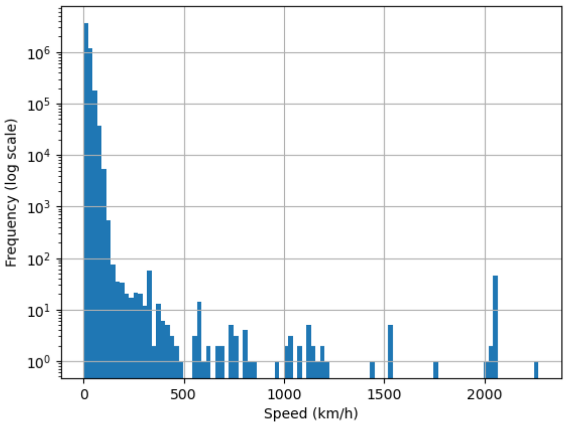

Figure 1: Distribution of speed between stops before preprocessing

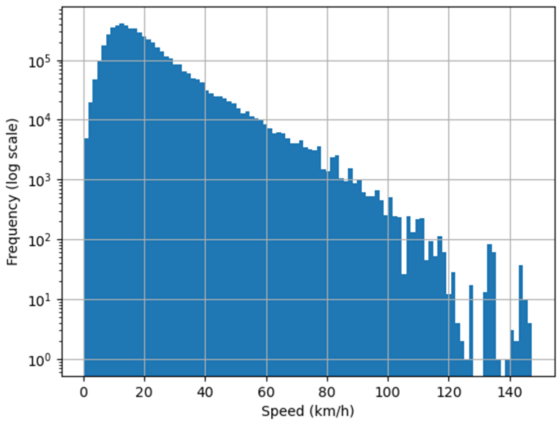

Figure 2: Distribution of speed between stops after removing outliers

Circle Packing

The Circle Packing visualization was a key component in unveiling the hierarchical structure of Switzerland's transportation systems. In our first milestone, we created a bar plot to count the number of entries for each transportation category, which provided a glimpse into the complexity of the system's hierarchy. See figures 3 and 4.

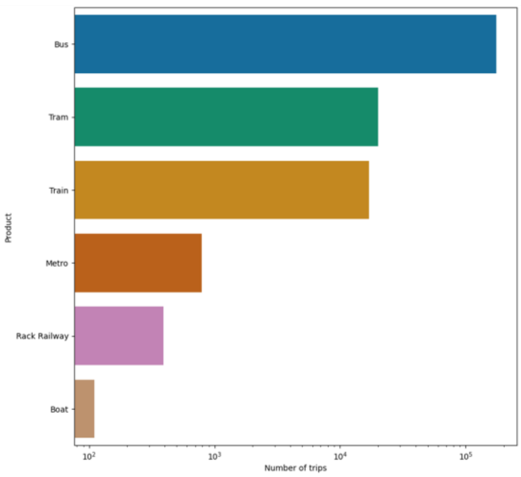

Figure 3: Number of trips per transport

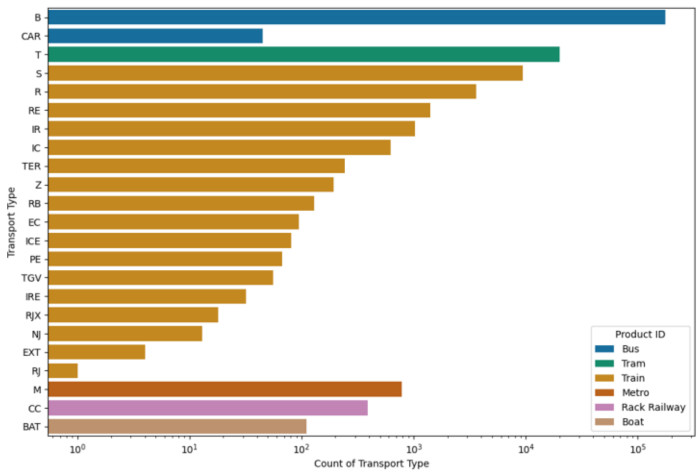

Figure 4: Transport types by products

Taking it a step further, our circle plot illustrated three levels of this hierarchy:

1. Transport products such as trains, buses, and trams.
2. Transport types, including specific tram numbers, intercity trains, and interregional transport.
3. Specific labeled transport types within these categories, like IR15 or IC9.

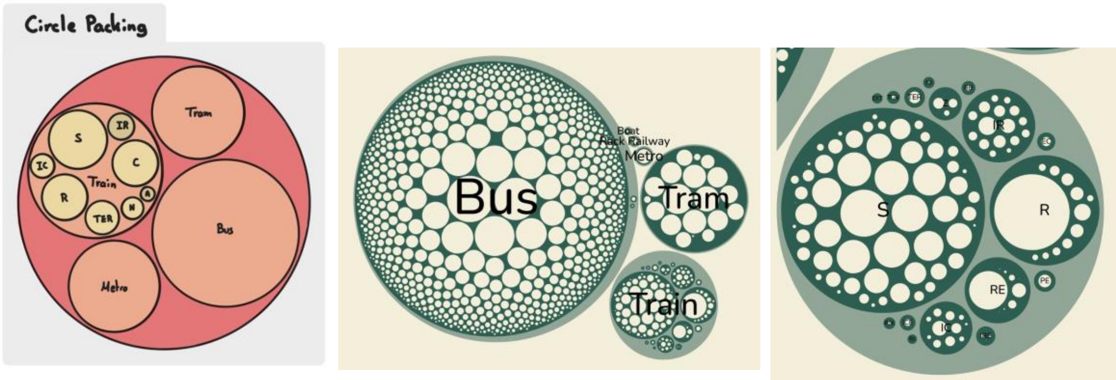

Figure 5: First sketch of the circle packing visualization and final results

This multilevel structure allowed for a nuanced exploration of the transportation system. To support user understanding, we also provided additional information for each transport type. When a user clicks on a specific mode of transport in the visualization, a brief description would appear, offering pertinent information and facilitating a more informed user experience. See figure 5.

Interactive Table

The Interactive Table presents many relevant features of stops. Users can sort the table in descending or increasing order by features such as delays or the number of trains arriving at a specific stop in 2022 by clicking on the header on the desired feature. When users click on a cell, alongside the table, a synchronized histogram displays the distribution of the feature selected cell is under and it highlights where it stands in the plot. Since there are too many stops to be displayed on just one page, pagination is implemented in the table.

Figure 6: First sketch of the table and final results

Hierarchical Edge Bundling

The Hierarchical Edge Bundling visualization represented an innovative approach to elucidating the connections between different transportation stops in Switzerland's cantons. In the development of our final product, we decided to use the Hierarchical Edge Bundling approach, which compactly displays information about connections between different transportation stops. This visualization comprises nodes representing the stops, and edges connecting the nodes symbolizing the trips between these stops. The edge thickness is proportionate to the number of trips that take place between two connected stops.

To further enhance the user experience, we incorporated an interactive minimap, which allows users to select specific cantons. Upon selection, the system dynamically alters the display, adding or smoothly removing corresponding nodes and edges associated with the selected canton. A critical design decision was to ensure that our visualization is accessible to everyone, including those with color vision deficiency. Our color choices were inspired by research on colorblind-friendly palettes and principles of color theory to ensure the best possible contrast and visibility for all users. Figure 8 demonstrates the color differentiation in our final visualization.

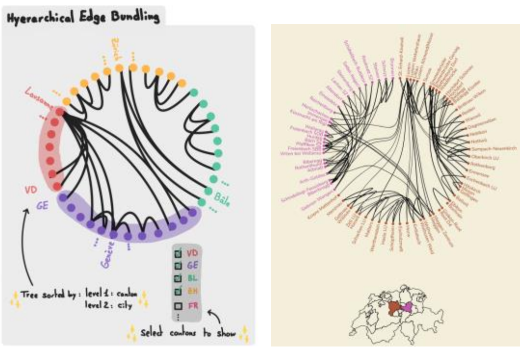

Figure 7: Sketch of the hierarchical edge bundling visualization and final result

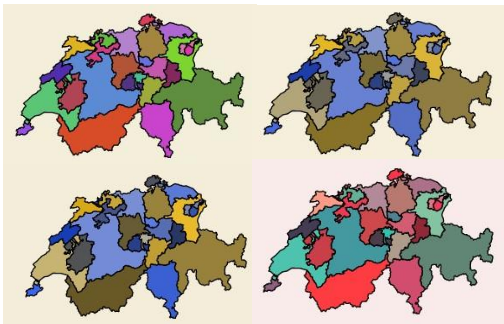

Figure 8: Minimap viewed through various forms of colorblindness

Network

The Network visualization forms a substantial part of our project, with each node representing a train stop and links indicating the number of transport modes connecting one stop to another.

Initially, we used D3.js to create graphs that would automatically run a simulation where interconnected nodes attracted each other, and unconnected nodes repelled each other. While this was interesting to implement, the computational cost of generating a network with approximately 2000 nodes and 3000 links in SVG using an online simulation engine was prohibitive.

Subsequently, we opted for Sigma.js for its computational efficiency. The trade-off was the necessity to perform all smooth animations in native JavaScript.

With the new setup, we computed multiple layouts, enabling smooth transitions between different node organizations. For instance, the geographic topology layout maps nodes to their projected latitude and longitude on the screen. The Force Atlas layout applies a force-directed algorithm that simulates physical forces on the nodes and edges to determine an aesthetically pleasing arrangement.

Additionally, nodes could vary in size and color based on several compelling metrics computed across the entire graph such as degree, page rank or betweenness centrality.

To further enhance usability, we refined the network to allow easy node searches by city or name. Highlighted nodes could be easily found, with the view automatically zoomed in on them. This level of user interactivity, coupled with visually engaging animations and transitions, elevates the overall experience of exploring the Swiss transportation network through our visualization.

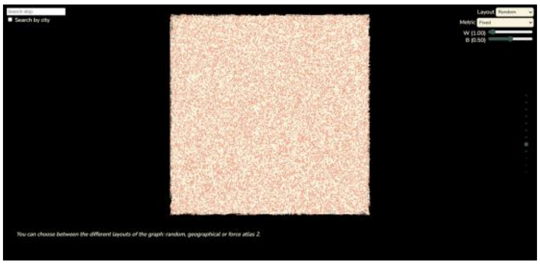

Figure 9: Figure showing the initial random network

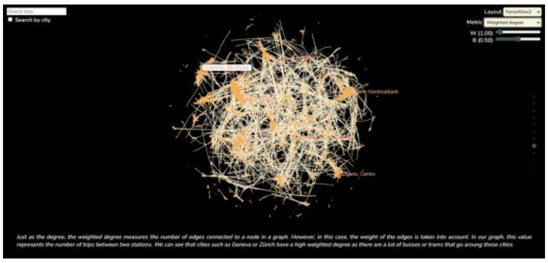

Figure 10: Network with Force Atlas 2 showing the weighted degree

Hourly Histogram

The hourly histogram served as a valuable tool to identify peak transportation times for each transport type. Initially, for milestone 1, we created individual histograms for each transport mode. Recognizing that this method could be further enhanced, we thought of incorporating smooth transitions between each transport type. This improvement would provide a more coherent and comprehensive view of peak times, facilitating users in understanding the ebbs and flows of different modes of Swiss transportation over a 24-hour period.

Map

The interactive map visualization was a cornerstone of our project, aiming to simulate the movement of trains across Switzerland throughout the day. This visualization would enable users to gain an intuitive understanding of how trains traverse the country, clearly illustrating the connectivity between various cities.

The primary challenge we faced in this stage was the need for a platform that could smoothly handle the visualization of hundreds of trains moving concurrently, without compromising on speed and performance. Consequently, we undertook several iterations of optimization and efficiency enhancements to achieve this.

Initially, we utilized D3.js to generate a map of Switzerland. Our aim was to project geojson paths on the website, exploring various possible topologies such as cantons or districts. This stage also incorporated interactive elements, allowing users to zoom in on specific cantons or districts and view their names. However, this approach had its shortcomings; the rendering process used SVGs, and more complex topologies like districts generated a multitude of polygons. This made the map noticeably laggy, even without any trains added. See figure 11.

Figure 11: D3.js interactive map visualization highlighting parts of Switzerland at different granularities

To address this, we transitioned to using THREE.js, creating a scene where we could efficiently add objects like the Swiss borders and trains. Leveraging WebGL for its excellent performance when rendering multiple simple shapes, this method significantly reduced the lag.

To ensure the smooth running of the simulation, several optimizations were implemented:

1. We continued to use D3.js for projecting Swiss coordinates of the borders and stop locations onto the screen. To speed up the process, we introduced caching for the projection results, especially for scenarios where projection was repeatedly called on the same coordinates, such as train locations.
2. Initially, the trains were represented by dots on the map, but we later upgraded these to small rectangles, each with a specific orientation. To efficiently compute these orientations without slowing down each frame, we utilized a caching strategy.
3. Given the high frequency of trains appearing and disappearing from the map, we created an entity pool. This allowed us to set aside objects not in use, ready to be reused for another trip at a different time. This strategy prevented constant memory allocation and deallocation with each frame, contributing to overall efficiency.
4. Another challenge emerged when loading train journey data. Given the size of the data files, loading all the information at once resulted in significant waiting times for users. To mitigate this, we partitioned the data into 15-minute simulation segments, enabling the website to load each file only as needed. This on-demand data loading strategy streamlined the user experience, allowing simulations to commence more swiftly without waiting for the entire dataset to be loaded upfront.

These enhancements significantly improved the efficiency and performance of our interactive map visualization, bringing us closer to our goal of providing a comprehensive, user-friendly visualization platform for understanding Switzerland's transportation network.

With the basic map framework in place, we sought to layer in more informative elements. We employed color-coding to distinguish trains based on their average delays and speed at a given time, enhancing the clarity of the visualization. We also experimented with a heatmap to illustrate the distribution of different types of stops across Switzerland. See figure 12. However, integrating this heatmap alongside the isochronic map posed complexities and was ultimately deemed superfluous, resulting in its removal from the final visualization. This decision underscores our ongoing commitment to maintaining an efficient, user-friendly, and meaningful interactive map.

An isochronic map illustrates travel time from a particular location to various points across a geographic region. In our implementation, rather than color-coding or indicating steps to denote hours, we opted to represent travel times by distorting the spatial structure of Switzerland on the map.

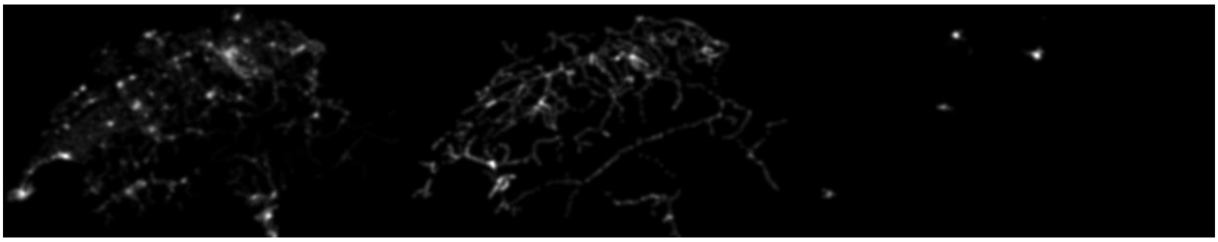

Figure 12: Heatmaps describing the concentration of transport products, respectively bus, train and tram

To incorporate this into our visualization, we initially chose a set of key locations such as Zurich, Basel, Lausanne, and Geneva, and precomputed the travel time in seconds from these cities to all other locations in Switzerland.

Creating an effective transport network for this task was a challenge. We conceptualized each transport stop as a node, with links between them carrying weights corresponding to the travel time between the stops. We made an assumption that nodes within a 1km radius of each other could be reached within a 20-minute timeframe.

This approach yielded accurate travel times for a subset of locations, but the discrete values were less helpful in depicting a complete picture. Therefore, we aimed to find continuous travel times for all locations in Switzerland. Two methods were adopted:

1. Nearest Neighbors: A location takes the travel time value of the nearest stop. However, this method presented the problem of discontinuity between stations with vastly differing travel times. See figure 13.
2. Linear Interpolation: The travel time at a location is interpolated based on surrounding nodes. This method offers more continuity in the data representation. See figure 14 for a comparison with the former method.

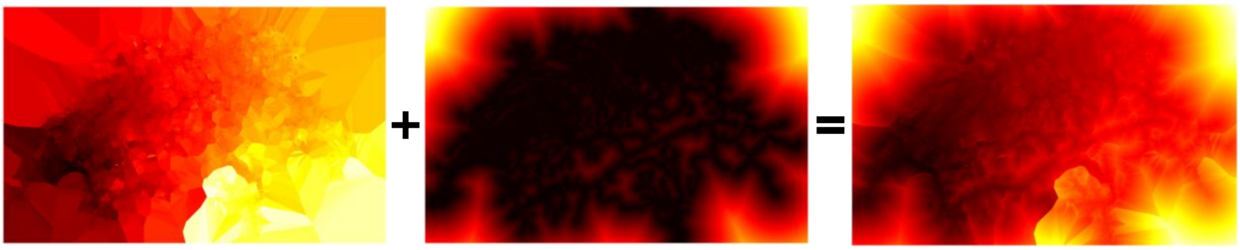

Figure 13: Final result of the nearest neighbour isochronic map centered on Lausanne generated from summing up the time zones to the walking time

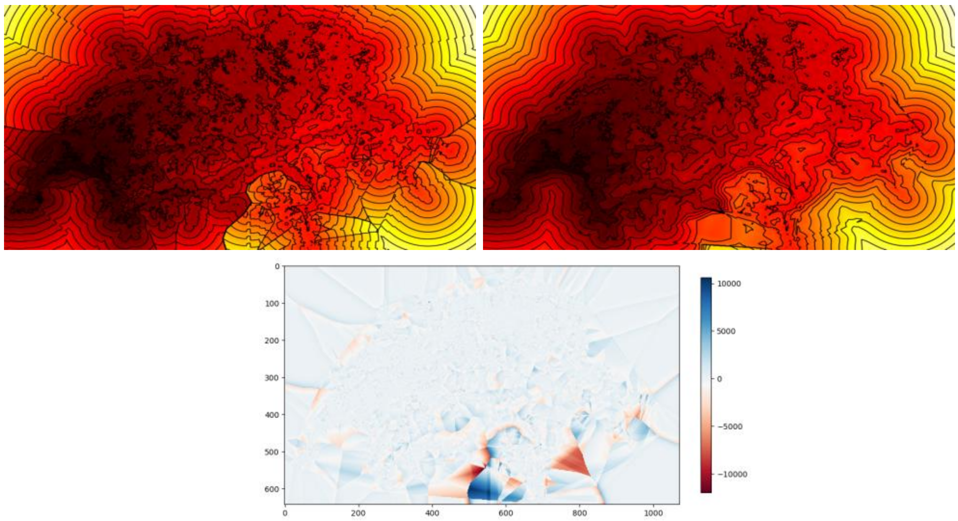

Figure 14: Nearest neighbor vs linear interpolation methods for isochronic map centered on Lausanne with hourly steps on top, the difference in seconds below

Finally, we opted for the second method. We also factored in walking speed for travelers moving away from stations, assuming a straight-line walking speed of 6 km/h. This inclusion rounded off our comprehensive approach to modeling the Swiss transport system using an isochronic map.

Peer assessment

As we reflect on our project journey, it's crucial to acknowledge and celebrate the unique contributions of each team member. Below is a brief peer assessment, summarizing the individual responsibilities and roles fulfilled by various members of our team in creating a successful project.

- **Primary Data Preprocessing Pipeline:** The cleaning and structuring of raw data via a preprocessing pipeline was efficiently performed by **Ozan** and **Arnaud**.
- **Traffic Network:** **Arnaud** handled the processing of the network, effectively illustrating stop nodes with edges that represented the number of transports passing through them.
- **Time Network:** The network showcasing stop nodes with edges defined by the number of seconds between stops was processed meticulously by **Arnaud**.
- **Data analysis:** A concise analysis on basic statistics, transport data hierarchy, and delay histograms was conducted by **Ozan** and **Arnaud**.

- **Aggregated Dataset:** All 2022 entries were gathered and integrated into an aggregated dataset using Spark containing more advanced traffic statistics by **Arnaud**.
- **Circle Packing:** The circle packing visualization for transport system hierarchy was implemented skillfully by **Arnaud**.
- **Stops Table:** The interactive table of stops was devised by **Defne**, **Ozan**, and **Arnaud**, while the histogram of table attributes was created by **Defne**.
- **Hierarchical Edge Bundling:** The hierarchical circle layout visualization for canton interactions via transports was developed by **Ozan** and **Arnaud**.
- **Network Visualization:** The comprehensive and interactive network demonstrating the overall transportation system, complete with multiple layouts and metrics, was formulated by **Ozan** and **Arnaud**.
- **Hourly Plot:** **Defne** and **Ozan** capably implemented a histogram to depict traffic importance throughout an average day.
- **Map:** The simulation of the Swiss transport system on the Swiss map, coupled with the bending isochronic map, was meticulously constructed by **Arnaud**.
- **Sketches:** **Arnaud** meticulously crafted visualization sketches under the insightful direction of **Defne** and **Ozan**.
- **Text Composition and Milestones Documentation:** **Defne**, **Ozan** and **Arnaud** undertook the crucial task of writing and refining the project's milestones and text content for our website, ensuring clarity, conciseness, and overall coherence.
- **Graphics Design:** The task of creating the background and image designs for the website was executed by **Arnaud**.
- **Website Layout:** **Ozan** and **Arnaud** were responsible for formulating the sections and determining the overall structure of the website.