Ben-Gurion University of the Negev

Faculty of Engineering Science

Department of Software and Information Systems Engineering

Department of Industrial Engineering and Management

Shadowed Segment Maps and Patterns Detection of Bus lines in Beer Sheva

Thesis Submitted in Partial Fulfillment of the Requirements for the M.Sc. Degree

Submitted By: Liad Nahmias

Supervisors:

Prof. Noam Tractinsky, Department of Software and Information Systems Engineering

Dr. Peter Bak, Department of Industrial Engineering and Management

E-mail: liadn@post.bgu.ac.il

14 October 2019

Ben-Gurion University of the Negev

Faculty of Engineering Science

Department of Software and Information Systems Engineering

Department of Industrial Engineering and Management

Shadowed Segment Maps and Pattern Detection of Bus lines in Beer Sheva

Thesis Submitted in Partial Fulfillment of the Requirements for the M.Sc. Degree

Submitted By: Liad Nahmias

Supervisors:

Prof. Noam Tractinsky, Department of Software and Information Systems Engineering

Dr. Peter Bak, Department of Industrial Engineering and Management

Author:.. Date:......................14.10.2019.........................

Supervisor:.. Date:......................14.10.2019.........................

Supervisor:.. Date:......................14.10.2019.........................

Chairman of Graduate Studies Committee: .... Date:...... 14.10.2019.......

14 October 2019

ABSTRACT

The public transportation system serves a vital role in modern life. It facilitates the movement of masses of people from one place to another, in relatively quick, cheap, and easy ways. Its importance is underscored by the growth in population, the movement to big cities and the deleterious effects of private cars (e.g., traffic jams, air polution, gas prices, etc.).

The field of public transportation consists of planning and building new infrastructures, purchsing new means of transportation for different needs, planning schedules and capacities of transportation lines, and monitoring and tracking the existing system.

This proposal is aimed at providing a means for monitoring the public transportation system and planning schedules by understanding the true picture on the roads (delays, number of passengers between stops, etc.). For this purpose, we present a new visualization called Shadowed Segments Maps (SSM) for monitoring the activity of bus lines. This visualization is constructed based on several big data sources, including metadata (list of stops, list of routes, schedules, etc.) and real-time data (stops at which the bus opened its doors, number of passengers that boarded or alighted, times of departure and arrival, etc.). This visualization compares planned and actual data using time-based, rather than area-based, visualization.

The proposed visualization greatly improves the ability to analyze very large amounts of raw transportation data, and may help to detect delays and define patterns and behaviors of specific bus lines. We envision it as a support system for decision-making for both experts and researchers in the public trasportation field. The results of an expert evaluation study provide support for this outlook.

The context of the current system is the metropolitan area of Beer Sheva , Israel, which currently enjoys the most accurate public transportation data reporting in the country. However, the visualization, the developed system, and the uderlying algorithms, will be scalable to other areas as well, as data from those areas become available.

The contribution of the work is as follows: (a) A unique visual way of detect delays at the public transportation based on comparison the planed and actual data (b) Monitoring existing features from the data as number of passenger, duration of openning and closing doors etc. and unique ones that are calculated from the data as the average speed of the bus (c) Displaying daily/ weekly/monthly of data for detecting routine patterns.

Keywords - Visualization, Public Transportation, Passengers, Openning/closing doors. SSM, Shadowed Segments Maps, Buses

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisors Prof. Noam Tractinsky and Dr. Peter Bak for their helpful advice, insightful ideas, and continuous guidance throughout the research.

I would like to thank the Department of Geography and Environmental Development at Ben Gurion University for helping me define the areas that need improvement in the field of public tranporation and for aiding me in obtaining the real-time data by mediating between me and the Ministry of Transportation of Israel.

Also, I would like to thank all seven experts that dedicated time to evaluate the SSM visualization, and Ayalon Highways for letting me use their facilities.

On a personal note, I thank my parents for providing me with support and continuous encouragement throughout my years of study.

Also, I would like to thank my two best friends - Michael Khavkin and Netanel Chaim for their time and effort in editing and consulting.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Every now and then we witness numerous problems in the public transportation in Israel. These include massive traffic jams, an increasing number of private cars on the roads, delays in train and bus schedules, etc. Consequently, it appears that the majority of citizens in Israel prefer to rely on their own private vehicles rather than on public transportation. This appears to be one of the main public transportation problems in Israel.

Speakers at the 58th Israeli Geographic Association conference estimated that if 10% of private car drivers switched to using public transportation, it would be enough to improve the massive traffic jams on Israel's roads.

From a general prespective, another problem that might be a major factor hindering the improvment of the existing public transportation infrastructures is inadequate monitoring. The raw data that the Israeli Ministry of Transportation collected from the different bus operating companies is not clear, not consistent, and problematic to analyze. Thus, it is difficult to provide accurate information on the state of public transportation in Israel, let alone suggest how it can be improved. This project focuses on improving this state of affairs.

The objective of this project is to visualize traffic data in a manner that will aid transportation experts, researchers, and administrators in improving the traffic flow in a specific metropolitan area.

For this purpose, we have studied public transportation data from the metropolitan area of Beer Sheva. Using visualization methods, we will try to "tell the story," in terms of space and time, of each bus route. The stories will be based on analyses of data gathered from publicly-available sources that track the movement of public buses in the Beer Sheva metropolitan area. The data will include metadata of routes (ID, list of stops, schedule, etc.) and real-time data (stops at which the doors have been opened, number of passengers who boarded and alighted, departure and arrival times, etc.).

In contrast to previous works done in this field, we do not use a geographic analysis as part of the analysis phase, meaning that our visualisation is not based on the traditional way of showing transportation routes on a map. Instead, we abstract the data and visualize it on two parallel straight lines, a method called Shadowed Segment Maps.

The resulting visualization is meant to compare the planned script of each route (based on the route's planned schedule) with the "actual story" (based on real-time data). This comparison should help us detect patterns of deviations from the planned schedule. For example, we can detect the overly busy segments (e.g., when the bus was late on most if its route runs), empty segments (when no passangeers boarded or alighted the bus), long-tail trips (a bus route in which most of the passengers had already alighted by the time the route ends), etc. Such observations should help planners improve bus scheduling, consider route changes, and in general, provide better public transportation services to the metropolitan area.

The rest of this work is structured as follows: The following section provides background on the two fields of public transportation and visualization, Section 3 presents related work and Section 4 discusses the context of this project. Section 5 describes the study’s method and Section 6 describes the novel visualization proposed in this study. Section 7 deals with the experts' reviews and their results, while Section 8 offers conclusions. Section 9 presents future work and suggestions for improving the usability and integrity of the current visualization, as well as raises possible limitations. Appendix A and Appendix B include the full survey given to the seven experts for evaluating the visualization and the experts' notes, respectivly.

2. BACKGROUND

Scholtz [Scho17] first defined the importance of interactive visualizations being both **useful** and **useable**. While Section 5.3.3 will discuss the usability component in depth, making a useful visualization requires a deep understanding of the domain (i.e., public transportation), its analytical needs, and the possible design space of the visualization component.

This section provides important key notes on the two main aspects of this research: public transportaion and visualization, for creating a helpful visualization.

2.1. Public Transportation

[Walk12] defined “Public Transportation” as consisting of "regularly scheduled vehicle trips, open to all paying passengers with the capacity to carry multiple passengers whose trips may have different origins, destinations and purposes” ([Walk12] p. 13-14).

This definition consists of four properties:

- "Regularly scheduled vehicle trips" – Public transportation provides a constant schedule.
- "Open to all paying passengers” – Public transportation provides services to any person who can pay. In most countries, this service is subsidized by the government.
- “That can carry multiple passengers” – The ability to carry many passengers in one vehicle.
- "Whose trips may have different origins, destinations and purposes” – Carpools, taxi services and school buses are excluded.

2.1.1. Lines and Stops

Every public transportation infrastructure consists of two main components: stops and the lines (roads/rails) connecting them.

The geography of the area defines where the lines will pass and what their layout will be. According to [Walk12], there are 3 different types of lines. Figure 1 depicts the three types:

- **Direct** – A line from point A to point B without any turns. This is the most desirable line from a planner's perspective; it runs directly to the target via the shortest path.
- **Circuitous** – Any line that has turns.
- **Deviating** – A line that runs direct except for a single turn. This is the most undesirable line from a planner's perspective; a single part of the line forces the driver who is driving in the original direction to waste time and effort.

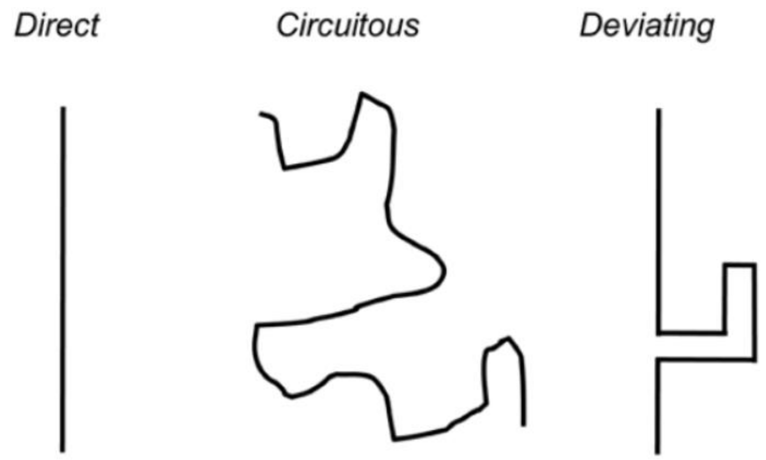


Figure 1: Three possible types of lines. (From [Walk12] P.48)

The decision of where to place bus stops should be based on the number of stops and their locations. Choosing stop locations involves various considerations. For example, private residents may prefer that stops will not be placed near their homes because of the noise and air pollution. In contrast, store owners may prefer that the stops be placed near their businesses, so that passengers can visit them easily when they get off the bus. Figure 2 depicts the manner in which planners define the circles of arrival to the stop by foot, by bike and by car. The walking distance to the stop is measured in quarter mile units (approximately 400 meters).

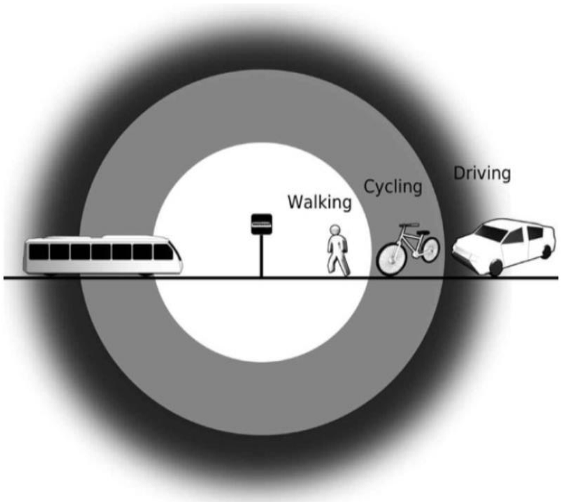


Figure 2: Stop circles (walking, cycling and driving). (From [Walk12] P.60)

2.1.2. Public Transportation in Israel

In Israel, there are two main modes of public transportation: train and road (i.e., bus).

- **Train**: **(Out of the scope of this project)**
  "Israel Railways" is the sole operator of rail transportation in Israel (https://www.rail.co.il/). In this work, we will not address this mode of public transportation.
- **Buses**:

In contrast to the train, buses in Israel are operated by many operators. Figure 3 describes four types of buses used by public transportation operators in Israel and their characteristics (number of passengers, bus length, floor height, and the ability to lower the bus closer to the ground).

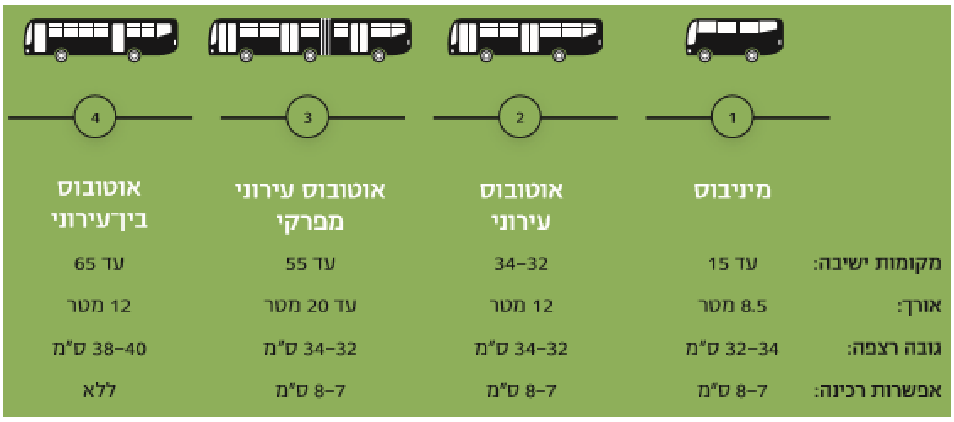


Figure 3: The four types of buses in Israel. (From [IsIA16] P.37)

In general, the Israeli Ministry of Transportation divides the country into four metropolitan areas: Haifa, Jerusalem, Tel-Aviv and Beer Sheva.

2.1.3. Parameters of Public Transportation in Israel

In 2016, the Israeli Ministry of Transportation and the Transportation Authority published a report containing instructions for the planning and operating of bus services in Israel [IsIA16]. First, they described the three dimensions of service:

- **Availability** – The time required for passengers to arrive at the closest stop (this includes the time it takes a passenger to arrive at the first stop on the route on foot).
- **Accessibility** – The number of destinations a passenger can reach (this including switching routes and lines).
- **Suitability to all passengers** – Whether the service is suitable for different neighborhoods and cities and to different types of passengers with different needs and constraints (adults, children, handicapped, etc.)

Secondly, the report describes seven quantitative measures and defines the required quality of the service.

Table 1 lists the notations used throughout the seven equations in Table 2:

Variable	Description
PB_i	The Number of passengers boarding at stop i
i	Stop i in a route that consists of I stops
L	Route length in km

l_a	Part of the route in km, in which the number of passengers exceeds the capacity of the bus
$\mathbf{O}_{i-1,i}$	Number of passengers that continue to consecutive stops
$\mathbf{l}_{i-1,i}$	The distance in km between two consecutive stops
C	The capacity of the bus
T	Trip duration
$\overline{\mathbf{T}}$	Average trip duration
N	Number of trip duration samples
P	The percentage of passengers arriving at the stop on time
H	The time declared in the schedule (estimated time at which the bus was supposed to be at the stop)
m	Delay in minutes

Table 1: Variables and their description. (From [IsIA16] P.82)

Table 2 lists those measures.

Parameter	Description	Equation	Measurement Frequency	Boundary Values
Number of boarding passengers	Total number of passengers on the bus throughout the trip	$\sum_{i=1}^{I} PB_i$	Peak times	Less than 10 boarding passengers
Number of boarding passengers per km	Total number of passengers on the bus divided by line length (km)	$\frac{\sum_{i=1}^{I} PB_i}{L}$	Peak times	Less than 1 passenger per km in the lines inside the cities
Usage of capacity	The sum of continuing passengers multiplied by the distance between two consecutive stops (km), divided by the capacity of the bus multiplied by line length (km)	$\frac{\sum_{i=2}^{I} O_{i-1,i} * l_{i-1,i}}{C * L}$	Peak times	Less than 30%

Trip Crowdedness	Percentage of the line's length in which the number of passengers onboard the bus exceeds the bus capacity (number of passengers sitting and standing)	$\frac{l_a}{L}$	Daily	Above 15%
Velocity per segment	Segment length divided by trip duration	$\frac{\mathrm{L}}{T}$	Daily	Less than 15 km/h inside the city and less than 40% outside the city
Variance in trip durations	Variance in trip durations from origin to destination	$\frac{(T_j - \bar{T})^2}{N}$	Daily	Above 6 minutes inside the city
Frequency of on-time arrival at a stop	If 90% of the time the bus is late to the stop by less than 2 minutes, it gets the highest value	$P(H+m)$	Daily	Depends on the time spaces

Table 2: The seven quantitative measurements. (From [IsIA16] P.81)

2.2. Visualization

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.” (Tamara Munzner [Munz14] P.1).

In other words, visualization is the methodology of graphic presentation of data for performing analyses, finding patterns and sets, and conducting explorations. Thus, it is a powerful tool for supporting decision-making. Nonetheless, at the basis of every visualization we can find raw-data.

[Munz14] suggests dividing the raw data into three primary groups: Ordinal, Nominal/Categorical, and Quantitative. Specifically,

- **Ordinal data** –Data which carries importance of order (e.g., high, medium, low). Usually, ordinal data relates to alphabetic text, but it can relate to numbers too.
- **Nominal/Categorical data** – Data which carries no importance of order (and thus, cannot be numeric). For example, Gender (male or female).
- **Quantitative data –** Numeric data. This group can be further divided into two distinct subgroups:
    - **Interval data –** Numeric data with an arbitrary zero that does not stand for anything. For example, the zero grade in IQ tests is irrelevant.
    - **Ratio data –** Numeric data which is centered around a zero value which has some significance. For example, a salary.

The second element of visualization is **encoding**. Good encoding can be the difference between a concise and an unclear visualization. Encoding can be improved by considering the data's properties and matching these properties using the data classification mentioned above. Each datum we display with our visualization has properties based on its classification (Ordinal, Nominal/Categorical and Quantitative). These properties include **position**, **length**, **color**, **volume**, **density**, **angle, opacity,** etc.

2.2.1. Visualization Properties

Each type of data can be encoded using various properties (shape, position, color, size, tilt, etc.). Each property has an effectiveness and importance for emphasizing the data and the message the specific visualization is meant to convey. In 1986, Jock Mackinlay [Mack86] claimed that it is important to design an effective graphical user interface, and Mackinlay’s extended accuracy rate system for properties was published by Cleveland and McGill for ordinal, categorical and quantitative data. Figure 4 depicts the model that was suggested for rating accuracy.

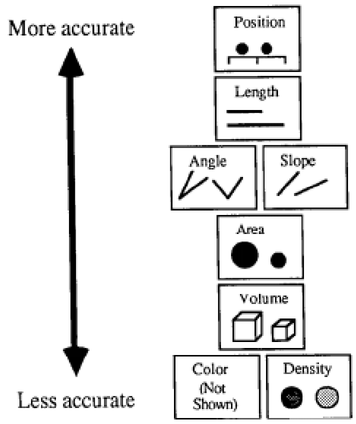


Figure 4: Accuracy rate for quantative data by Cleveland and McGill. (From [Mack86] *P.17)*

In contrast, Mackinlay's model (Figure 5) shows the preferred order of properties for each type of raw-data (ordinal, categorical and quantitative). Both Mackinlay's and Cleveland and McGill's models suggest that the most valuable property for all data types is position. However, there are some disagreements regarding the other data properties. Note that Mackinlay did not provide empirical evidence for his model.

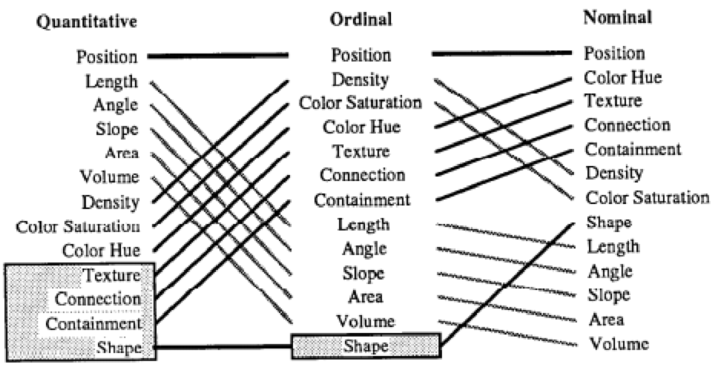


Figure 5: Mackinlay's model for encoding all three type of raw data (Nominal, Ordinal and Quantitative). The gray parts were found as ineffective. (From [Mack86] *P.18)*

2.2.2. Visualization Evaluation

What properties can be used to evaluate a visualization? The phrase "Beauty is in the eye of the beholder" suggests that different people have different perspectives regarding aesthetics. However, when designing a functional visualization, the goal is to deliver a similar message to all viewers by using effective visual encoding.

One criterion for visualization evaluation is based on Tufte's [Tuft08] principal of **data-ink ratio,** which measures the ratio between the amount of ink used in the visualization to the amount of data it represents (Equation 1). Tufte suggests that the greater the ratio, the better the visualization.

$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$

$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

$$= 1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}$$

Equation 1: Tufte's explanation for calculating the data-ink ratio. From ([Tuft08] P.93)

Mackinlay [Mack86] suggested two main criteria a visualization's quality: **Expressiveness** and **Effectiveness**. A visualization is of high expressiveness if it shows all the facts in the data (and only them). A visualization is effective if it is more readable to the viewer and if it delivers the intended message.

2.2.3. The "What-Why-How"Model

Munzner ([Munz14]) suggested a schematic model, according to which, the process of creating a visualization should consider three main aspects - What, Why and How (Figure 6). The consideration of these three aspects is usually done sequentially, in phases. Hence, they are termed "phases". The "What-Why-How" model based on Munzner's Nested model for Visualization Design and Validation, which includes four levels – Domain problem characterization and Data abstraction (What phase), Encoding technique design (Why phase) and Algorithm design (How phase) [Munz09].

Figure 6: Munzner's Model "What-Why-How" and the relationships between the three phases. From ([Munz14] P.17)

The First phase comprises the understanding of **What** data we have and how to classify it to its different types. This phase is crucial for creating an effective and understandable visualization (as mentioned in section 2.2.1). Figure 7 provides a full depiction of the "What" phase. It defines the types of the data, the datasets, and finally, the attributes. Munzner's definitions for data types and datasets derive from their representation: are they a coordinate on a map? a node in a network? data in a table? The attributes are related to the type itself: categorical or quantitative.

Figure 7: The "What" phase. From ([Munz14] P.20)

The second phase in the model represents **Why** a visualization was chosen for this task. This phase is meant to help define more clearly what the actions and goals of the visualization are. Generally speaking, this phase consists of two main parts: actions and targets. Figure 8 shows the possible actions, e.g., the visualization will help analyze the data or show hidden facts within it. The second part is the targets: representation of trends, anomalies or connections between datapoints.

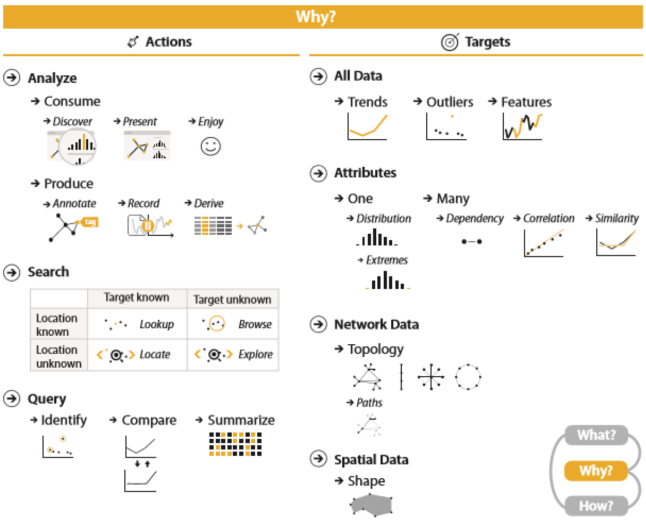


Figure 8: The "Why" phase. From ([Munz14] P.42)

In the third and final phase, the focus is on **How** to create the visualization. Generally speaking, it relates to effectively encoding the different types of data in order to deliver the message in a clear and convincing way. Figure 9 depicts the different ways one can encode the data by using visual properties, manipulating them, and even filtering them and showing only the required data. For example, showing geographic data on a map may help us extract more information from the data. This is the same as displaying data on a graph to see a trend. Another aspect is encoding the data to properties (color, size, shape, etc.).

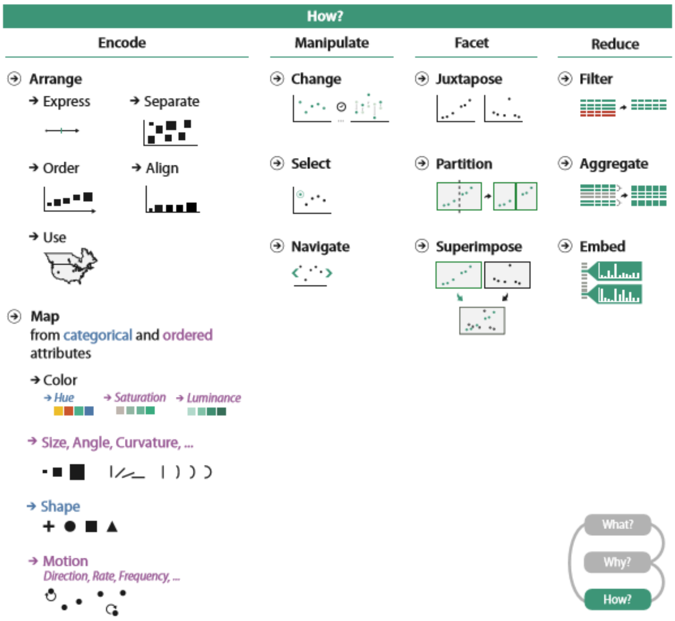


Figure 9: The "How" phase. From ([Munz14] P.58)

2.2.4. Visualization of Movement

Public transportation consists of geographical data and time-series. Thus, it can be classified as a type of visualization of movement. Some basic terms of movement in visualization are reviewed below.

Andrienko + et al. [AABK71] suggested that movement in visualization consists of three basic dimensions: Time, Space, and Object.

- **Time** is defined as a continuous set with linear order and distances between moments or occasions (e.g., hour, minute, etc.)
- **Space** is a set of places with different distances between them (e.g., city, neighborhood, etc.)
- **Object** is any physical or abstract entity (e.g., car, bus, etc.).

3. Related Work

In this section, we will survey related work, focusing on the analysis of public transportation around the globe using various analyses, different visualization methods, and explore how to build and conduct expert reviews.

3.1. *Visualizations of Public Transportation*

Many papers have been published on using a visualization of the public transportation for analyzing different user tasks, defining poor areas, creating a current model of the existing system, etc. Each one of them used different visual data encoding to meet its needs. Surprisingly, very few authors have used a visualization to deal with the delays in public transportation. We will review the most notable of them, list their strongpoints, and explain how we used these strongpoints to create a superior encoding.

3.1.1. Spatially-based Visualizations

Some researchers used the public transportation data geographically, on maps, to present theses and existing issues. There are many advantages in using maps as a visual method to derive more information from geographical data and to understand the spatial dimension. In a work by [Aljo14], a spatial analysis was conducted on the deficiencies and potential of the public transportation in Jeddah, Saudi Arabia. The author showed that factors such as poverty, regions with a high proportion of car ownership, and poor accessibility to public transportation are reasons for the lack of use of public transportation. Figure 10 and Figure 11 display the public transportation map, and Figure 12 depicts the population density. According to the author, around 50% of the population has low or no access to the current public transportation.

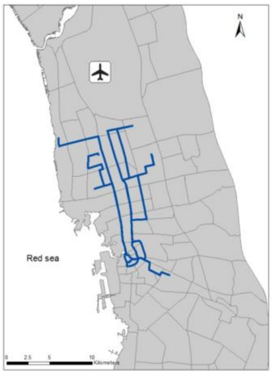


Figure 10: The regulated public transportation system map. (From [Aljo14] P.4)

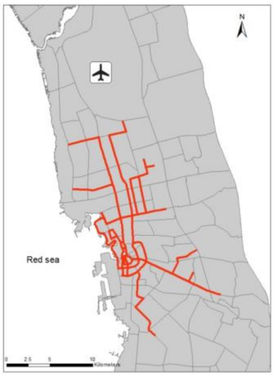


Figure 11: The unregulated buses system map. (From [Aljo14] P.4)

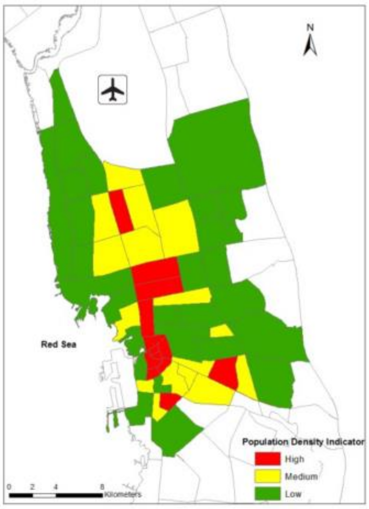


Figure 12: The population density indicator. (From [Aljo14] P.8)

Another way of utilizing a visualization on a map is for navigation needs. Figure 13 depicts the four main modes of transportation in the city of Berlin: S-Bhan (green dots), U-Bhan (blue dots), tram (red dots), bus (purple dots) and the Fähren (light blue dots). By choosing the mode of transportation and the specific route (available at: https://interaktiv.morgenpost.de/berlin-an-deiner-linie/) on the map, the specific geographic location appears, and using animation, the directions of the route are shown.

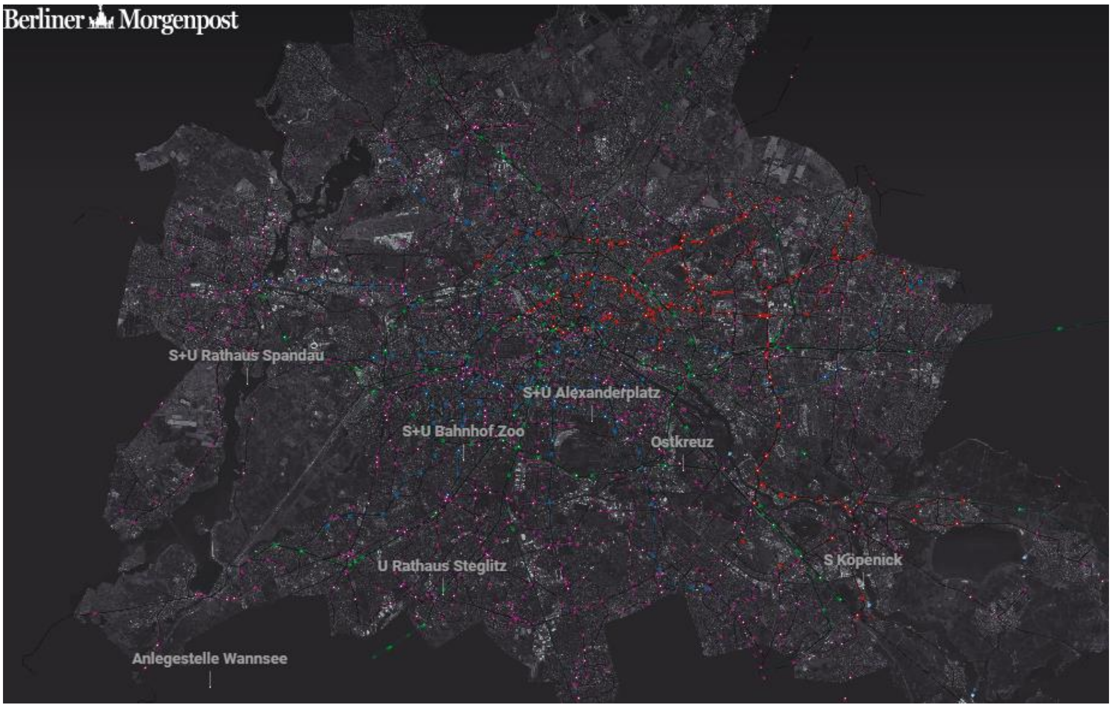


Figure 13: Berlin's Transportation on a map for navigation. URL: https://interaktiv.morgenpost.de/berlin-an-deiner-linie/

Despite the clear advantages of using a map, which presents the true state of the geographic area, in our visualization we decided to put the map aside and "straighten" the data on a horizontal line to allow a clear and more meaningful comparison between the planned and actual (real-time) data. Even though we gave up the advantages of a map, we managed to effectively visualize characters such as distances, traffic lights, roundabouts and turns.

3.1.2. Network-based Visualizations

Other researchers used a different method of visualization, a system of circles and lines, to analyze public transportation as a network or a graph (vertexes – stops, and edges - roads). In [WuGS04], the authors presented a method of looking at public transportation as a scale-free network (a network in which degree distribution is logarithmic scaled). This way, the nodes with the highest degree (the most connections) can serve as hubs, or in public transportation terms, central stations. The analysis was conducted on the Beijing public transportation system. Figure 14 depicts the Beijing public transportation map, while Figure 15 shows the network representing the data on the Beijing public transportation system.

Figure 14: Beijing public transportation map. (From [WuGS04] P.4)

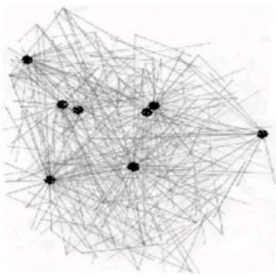


Figure 15. Network from Beijing public transportation data. The black nodes are the hubs (Central stations). (From [WuGS04] P.4)

Another important work which used the network structure of public transportation is by [Syrj16]. This author performed a very thorough analysis of the urban bus network. In the paper, the authors reviewed a large volume of data (bus stops, lines, bus traffic and general traffic on the roads in Finland) to understand the network, monitor traffic patterns, identify bottlenecks, etc.

These articles which converted the public transportation data to a network/graph, are the foundation of our visualization. Each stop in our visualization will be represented as a circle, while the connection between two neighbors stops will be represented as a line.

We adopted the perception of stops as circles and roads as lines in our visualization, but we did not build the entire graph at once. Our visualization focuses on one bus line at a time (origin-destination stops).

3.1.3. Data-encoded

For encoding the data to visual properties, we adopted some ideas from the following three papers.

The first paper, by [SMDS14], presented a unique visualization which helps to define patterns between binary properties. The authors created a new visualization which helps to show binary traits using matrices. The main contribution of their work is the ability to connect entities using the functions "AND" and "OR". Figure 16s depicts the two actions in the top left corner, and the number of entities in the top right corner. This visualization is a useful tool for showing if any relationships exist, and if so, what are the strengths of these relationships (using opacity).

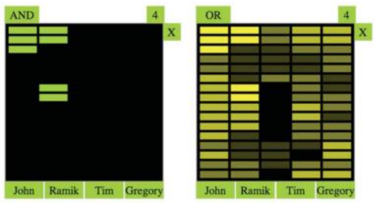
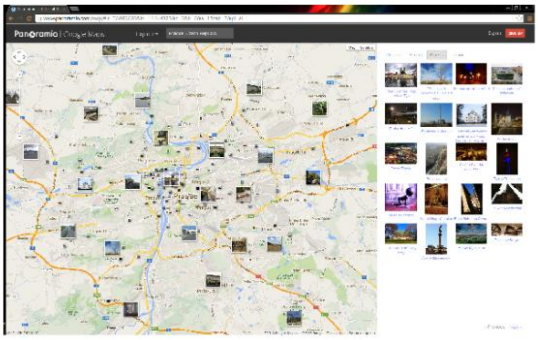


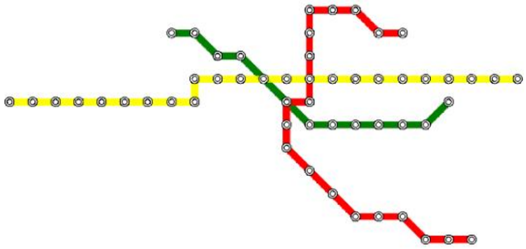
Figure 16: OnSet visualization matrices. (From [SMDS14] P.6)

The second work showed the simplicity of subway/metro maps. [ClYo14] used a visualization of the metro maps to help plan trips in cities by adding points of interest. This visualization can help tourists and other passengers find the way to their desired destinations.

Figure 17 displays the superimposition of the two maps, a and b, on the lower map. The combination creates new visualization that helps the tourists that are new for the city.



(a) Tourist destinations map



(b) Schematic metro map

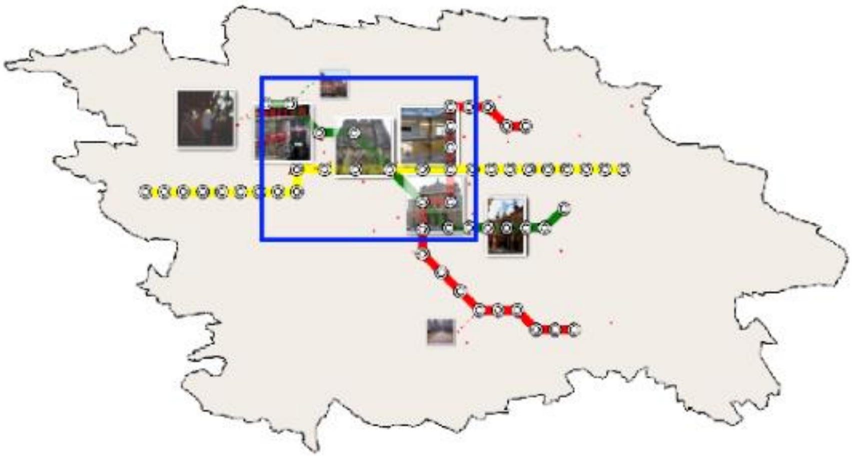


Figure 17: Superimposition of the metro map and points of interests. the lower map is the final product.

The metro map is the fundamental visual data encoded by our visualization, but in contrast to the usual use of preserving the geographical position of the stops, we flatten the lines from the geographic form to a horizontal line (the first stop is the left circle and the last stop is the right circle).

The third work presented a creative visualization of the mobility of the public transportation system. The paper by [ZFAE14] shows a very unique visualization in the form of a combined three-visualization model: an isochrone map view for geographical information, an isotime flow map view for effective comparison and manipulation of temporal information, and an OD-pair journey view for finding the most efficient route to a given destination, considering factors such as trip length, transfer times, etc. Figure 18 depicts the combination of the three visualizations types.

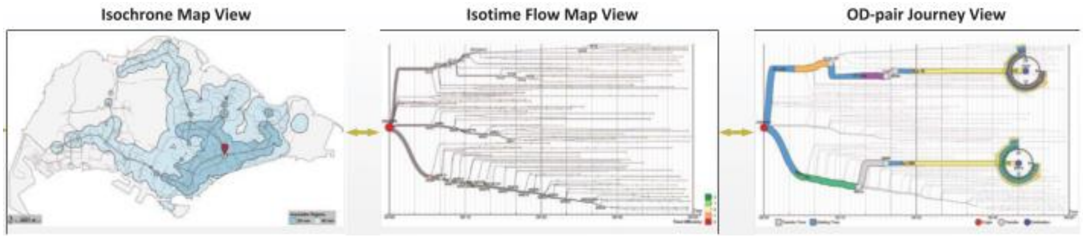


Figure 18: The three visualizations comprising the final one. (From [ZFAE14] P.3)

Figure 19 depicts the OD component more thoroughly. In this figure shows two lines from the origin stop CC9/EW8 to two destination stops (64151 and 03369). The blue line represents the waiting time, hence, the waiting time is longer on the route to stop 64151. The gray line represents the transfer time, hence, the transfer time is shorter on the route to stop 64151. The mobility clock at the end of each route indicates the expected duration of the trip on a given hour.

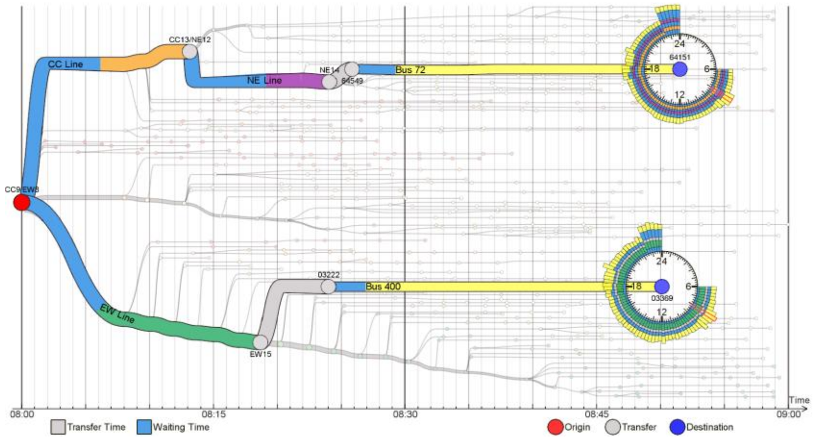


Figure 19: OD visualization options from origin stop CC9/EW8 to two destination stops (64151 and 03369). ((From [ZFAE14] P.3)

This visualization is a good example of using colors to provide the user with quick insight regarding aspects that are important to him when planning a trip (such as transfer times, total trip duration, etc.). Moreover, removing the geographical data presented in a map is logical step when planning a trip, since usually a passenger's physical location on the map is unimportant, and the only thing that matters is the origin and destination stops. Although our main user task is different, since our visualization displays delays in public transportation for monitoring and planning improvement purposes, the two visualizations share the horizontal lines motif, relinquishing the use of a geographical map, and using time-based comparisons.

The most recent visualizations that were found to be similar to our visualization used horizontal\vertical\diagonal parallel lines to depict the movement of the public transportation in different places around the globe. This idea is attributed to E.J. Marey in a visualization from 1885 (From [Tuft08] P. 31) .

Figure 20 depicts E. J. Marey’s visualization of the train schedule from the city of Paris to Lyon in the 1880s. The horizontal axis represents the time in hours, while the vertical axis shows the stops between Paris and Lyon. Every trip is depicted by a diagonal line which starts at the time and place the trip began, and moves upward from left to right until reaching the final stop and the time the train arrived there.

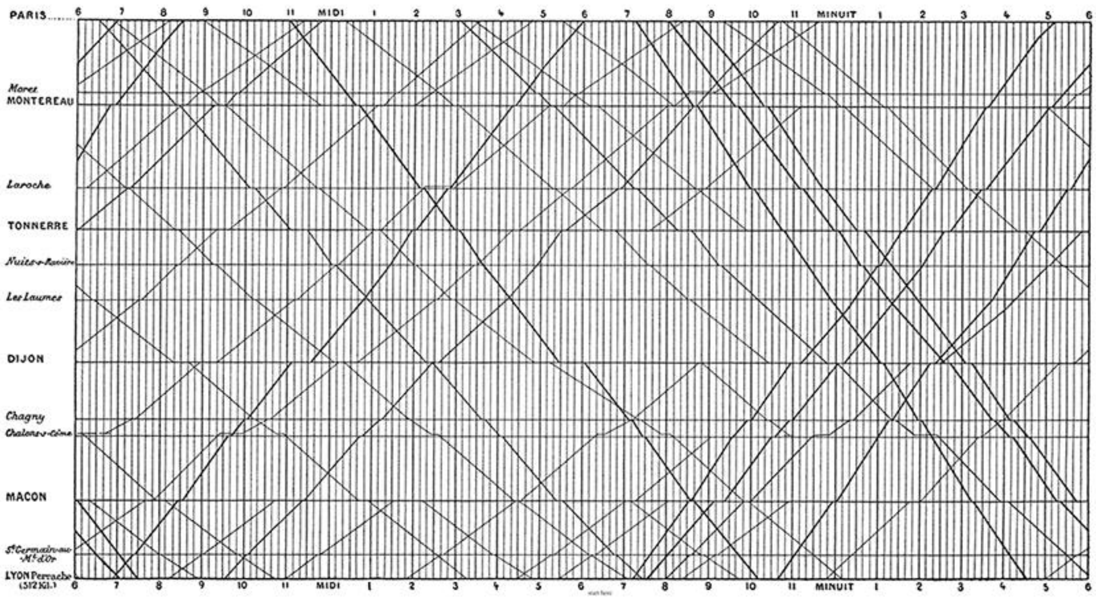


Figure 20: Train from Paris to Lyon by E. J. Marey, 1885. From [Tuft08] P. 31

An example of a visualization that relies on E. J. Marely's work was The BART (Bay Area Rapid Transit) System Schedule of San Francisco. Figure 21 (available at https://public.tableau.com/profile/kiyoshi#!/vizhome/BART/BARTSchedule) shows the schedule of the BART during the hours of the day. The user can choose a specific route or point of interest and filter the data accordingly.

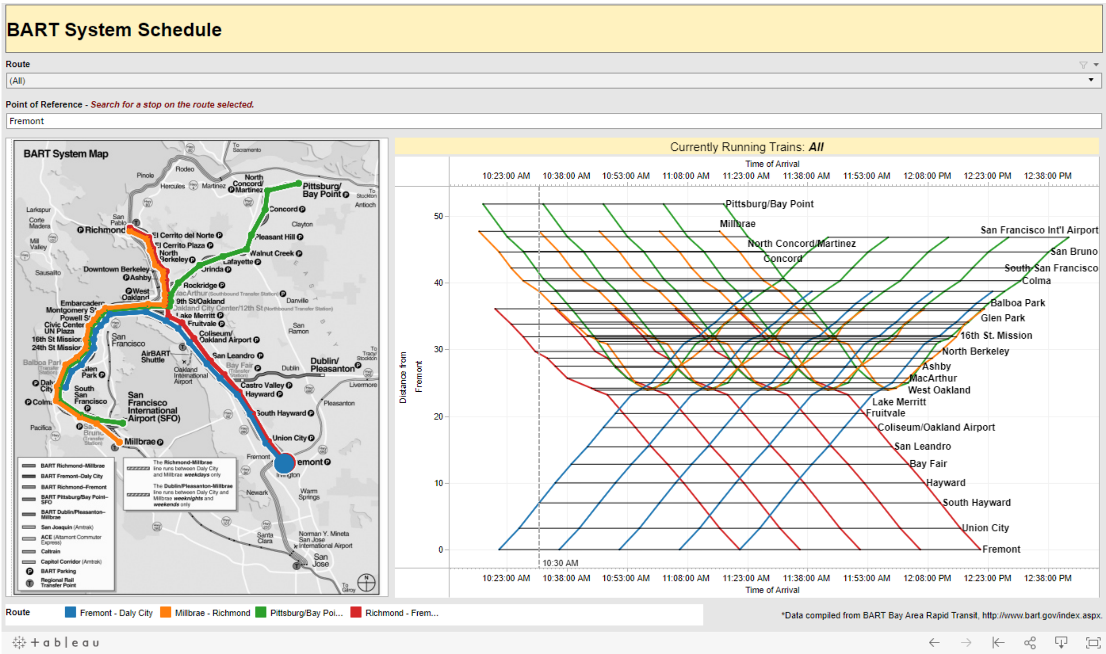


Figure 21: The BART System Schedule by visualization.

Another example is a visualization by Barry and Card, and has not been formally published. Barry and Card took the Boston subway system data from Boston's Massachusetts Bay Transit Authority (MBTA) from February 2014 and displayed it on their GitHub [http://mbtaviz.github.io/]. While their work has not been formally published, it presents several visualizations for analyzing delays. It used an effective combination of a small map and parallel lines for monitoring delays.

Figure 22 displays the three subway lines (red, blue and orange). On the left corner there is the map of the three subway lines, and in the middle, there are parallel lines, each one representing a station, while the vertical lines represent the time of day. The red, blue and orange represent trains riding each line from the origin station to the destination station. The longer the trip, the bigger the slope of the line.

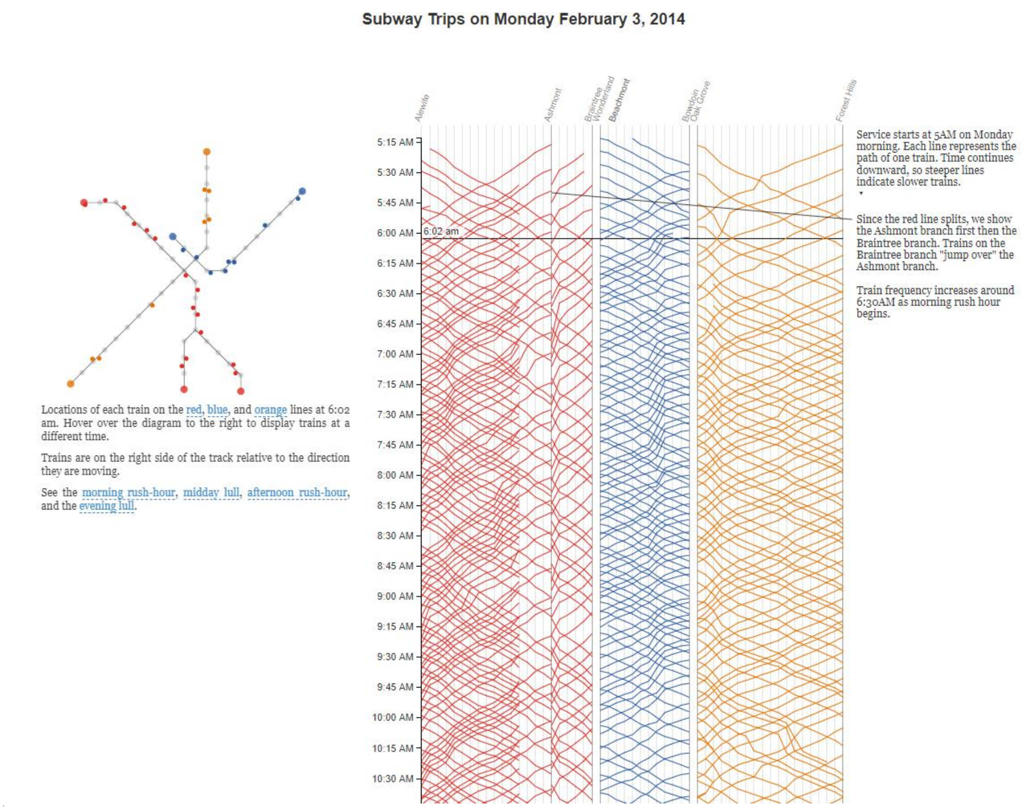


Figure 22: The visualization of the MBTA data

In contrast to our visualization, the passengers' data was displayed using a different visualization. Figure 23 represents the passengers who entered and exited the MBTA system (the total number of passengers in all the three lines). Each day of the week has a separate "entries" and "exits" scale by hours. Using a color palette which ranges from white (low density) to red (high density) to represent the average number of passengers per minute.

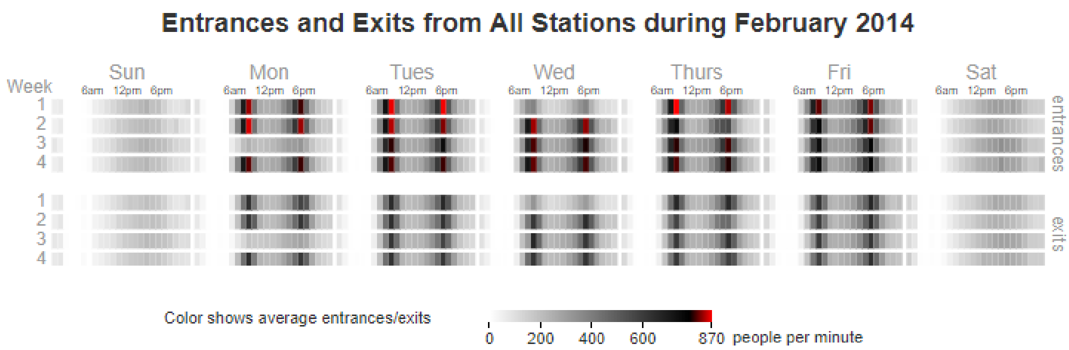


Figure 23: Visualization of passenger density over time in the MBTA system

Unlike Barry and Card's work and the Bart System Schedule, our visualization combines all three visualization in one, without using a small map. This will be thoroughly explained in Section 6.

3.2. Visualiztaion Evaluation

The second phase of this project is evaluating the visualization and its contribution. The evaluation process of a visualization is difficult and complex, as each visualization was built as a visual solution for a specific user task.

Lam et al. [LBIP12] displayed the shocking statistic that between the years 1995-2010 more than half of the previous works from the four main visualization venues (Eurovis, Infovis, VAST and IVS) have not used any sort of evaluation for their visualizations. Therefore, Lam et al. and many other authors tried to create some order in the chaos.

Lam et al. gathered seven scenarios, analyzed them, and suggested an evaluation method. For our type of work (evaluating the relevance of a visualization), they suggested an informal evaluation (i.e., an experts' review), a usability test, field observation, and a laboratory questionnaire.

Scholtz [Scho17] also mentioned the growing field of creating interactive visualization systems, and the resulting need for new evaluation methods. Scholtz published a short book in which he summarized methods for user-centered evaluation. In his book, he emphasizes the importance of understanding whether the visualization is useful to the user in its intended setting, and whether the visualization solves the domain problem in a simple and interpretable way, while remaining faithful to the data (he referred to this as “sensemaking”).

In accordance with these papers, since our visualization is intending to serve researchers and domain experts from the field of public transportation, we decided to evaluate it using a small group of experts (a qualitative evaluation) rather than a large number of students (a quantitative evaluation). The process is called an experts' review, and is based on domain experts who evaluate the visualization using their experience, knowledge, and understanding of the needs of the field. This evaluation process is usually carried out by conducting interviews with the experts and them performing tasks using the visualization. Both of these methods are used interchangeably in the different papers and are matched to the specific needs.

[ElYi15] reviewed in depth some evaluation methods, presenting each of their strengths and shortcomings. The advantage of using experts is the ability to obtain deeper and more meaningful insights. However, the main problems of relying on this method are the lack of availability and compliance of experts who are willing to participate in the evaluation process, and not achieving a statistical significance (small group). To overcome this obstacle, the authors suggest sending the experts a written questionnaire containing simple, open-ended questions, in order to receive deep insights from using the system with respect to some tasks.

[Carp08] reviewed a vast range of evaluation options for visualizations, both quantitative and qualitative. An experts' review, which is a qualitative method, does not require a massive experiment and includes an interview and an observation component. The interview should be open but very coherent, it should be based on asking open-ended questions while avoiding leading questions. In addition, the authors recommend taking notes during the interview. Furthermore, regarding the observation component, the authors recommend taking the notes as close to the observation as possible.

[ToMö05] shared their insights about the points on which the interviewer should focus the experts, what to evaluate, and to what extent he should make the review independent and informal.

As for the number of experts, there is no rule of thumb. The availability of experts in each field and their willingness to participate are the main factors influencing their number. Some examples from previous works include [ElST07], which interviewed only 2 experts, while [ZhWS17] interviewed 13. For this project, we interviewed 7 experts.

4. CONTEXT

Our project focuses on the public transportation in the city of Beer Sheva. More specifically, the project focuses on bus routes operated by the major transportation company in the metropolitan area, Dan Beer Sheva.

4.1. Dan Beer Sheva - Numbers

Dan Beer Sheva operates 33 bus lines in the city of Beer Sheva. Each line has two different routes (a total of 65 unique routes). The first route is from stop A (origin) to stop B (destination), passing through several stops on the way. The second route is from stop B (origin) to stop A (destination), passing through several stops on the way, which are not necessarily the same stops as in the first route.

During a week, the 33 bus lines make 16,929 trips (Real-time data on specific route) and drive through 577 stops.

4.2. The Revolution in Dan Beer Sheva

In 2017, Dan Beer Sheva changed the decades-old star-like route model and adopted a grid model for the city of Beer Sheva (http://www.danbr7.co.il/pages/1060.aspx). In a star-like route model, each route starts and ends at the central station. In contrast, the new model defines six additional locations as "central stations", and the bus lines can also start/end their trips in these locations. Figure 24 shows a schematic depiction of the old "star-like" route model. Figure 25 depicts the new model.

Figure 24: Dan Beer Sheva - The Old Public Transportation Model. All lines start and end at the central station. (From http://www.danbr7.co.il/pages/1060.aspx)

Figure 25: Dan Beer Sheva - The New Public Transportation Model. The lines start and end at different locations and neighborhoods. (From http://www.danbr7.co.il/pages/1060.aspx)

5. DEVELOPMENT METHOD

For this project, we adopted the "What-Why-How" model to define a systematic way of developing our visualization method. Those three phases will be thoroughly discussed in this section.

5.1. "What" – Data

The data for this project will be divided into two parts: the planned data, which was collected from open sources and includes the bus lines, schedules, stop order, distances, etc., and the actual data, which was obtained from the Ministry of Transportation and includes data on specific trips conducted during the first week of December 2017 (the bus lines' trips per hour, the number of passengers who boarded and alighted, the times the bus doors were opened/closed, etc.).

5.1.1. Raw-Data

The data types used and their corresponding notations are as follows:

Planned data:

- **33** unique bus lines, e.g., **2, 3A, 81.**
- **65** unique routes[1], e.g., **'81': ['9269399','9271551']**
- **577** unique stops
- **748** unique segments (a part of the road between two neighboring stops)
- **16,929** trips (Real-time data on a specific route) during a week
- **19.41 km/h** is the average speed of the bus lines (from the Ministry of Transportation data [calculated by dividing the length of the route by the estimated time]. This was also approved by our experts.

Actual data:

- **11,202** trips during the week 01.12.2017-07.12.2017 (**before data cleaning – delete duplicates stops in the same trip, stops with id -1 etc.**)
- **9,478** trips during the week 01.12.2017-07.12.2017 **(after data cleaning)**
    - **a. 4,327** long-tail trips during the week
        - **i. 2,094** started with long-tail trips during the week
        - **ii. 4,253** ended with long-tail trips during the week
        - **iii. 1,010 both** started and ended with long-tail trips during the week

[1] Except for line 24, which has 3 routes, and line 82 and line 23, which have 1 route each. All other bus lines have 2 routes per bus line.

5.1.2. Basic Terms

- **Passengers** – People who uses the public transportation.
- **Stop** – A bus station.
- **Line** – A bus identifier.
- **Boarding** – Getting on the bus
- **Alighting** – Getting off the bus
- **Route** – The course a bus passes. Consists of segments.
- **Segment** – A part of the road between two neighboring stops.
- **Trip** – Real-time data on a specific route. Contains data on trip onset, stops at which the bus opened and closed its doors, number of passengers who boarded and alighted the bus.
- **Long-Tail Trip** – A bus route on which at the beginning of the route or a few segments before the final stop the bus drives empty without passengers.
- **Busiest Segment in Route** – The segment on which the bus was late on most trips of on specific day.

5.1.3. Sources

The data we used in this project was collected from different sources. The list of sources and the relevant data gathered are as follows:

- **Bus.co.il -** A self-coded crawler extracted all the data on lines, routes, schedules, up/down, and estimated times form the website: https://www.bus.co.il/ **[the data was collected on 22.02.18]**
- **Moovit** – For double checking of the lines and routes. In addition, it was used to convert stop IDs to names and routes on map photos (https://moovitapp.com/). **[the data was collected on 22.02.18-01.03.18]**
- **Google Maps (street view) -** All section data such as length, roundabouts, traffic lights, and turns (https://www.google.co.il/maps?source=tldsi&hl=en). **[the data was collected on 22.02.18-01.03.18]**
- **Beer Sheva Municipality –** For double checking of traffic lights data (http://opendata.br7.org.il/dataset/light-traffics) and for roundabouts data (http://opendata.br7.org.il/dataset/fountains). **[the data was collected on 22.02.18]**
- **Ministry of Transportation** – CSV file with the actual data. **[the data was collected on September 2019]**

5.2. "Why" - User Tasks

The user tasks define the goals of the visualization. It is important to describe user tasks as clearly as possible to help build the correct visualization with respect to a task.

5.2.1. User Tasks

Our visualization was developed based on the user tasks presented below. The tasks were determined by consulting experts in the field of public transportation (Prof. Eran Ben-Elia, Mr. Dror Bogin, and Mr. Dan Rader) and in the field of visualization (Dr. Peter Bak).

- To compare a trip's actual data to the planned data and to the average de facto
- To detect delays and early arrivals, by whole trip and by segments
- To identify crowded segments
- To detect the most busy stop on a given trip (boarding/alighting/both)
- To detect long-tail trips
- To detect stops at which the bus did not stop
- To detect the average speed of the bus on each segment
- To display connections between traffic lights, roundabouts, and turns and delays/slow speed
- To create daily, weekly and monthly reports on all trips on the same route

5.3. *"How" – SSM (Shadowed Segment Maps)*

In this section, we will elaborate on the design space, relevant data attributes, and the idioms and marks required for the creation of the Shadowed Segment Maps (SSM) visualization.

5.3.1. Design Space

The design space defines and allows to explore all of the options for creating the visualization by displaying all of the possible connections between the data attributes to the idioms and marks. Each idiom/mark consists of one or more visual channels. As shown Figure 26, the design space of a SSM visualization comprises eight data attributes (left column) which are connected to visual channels (right column) and four marks (middle column).

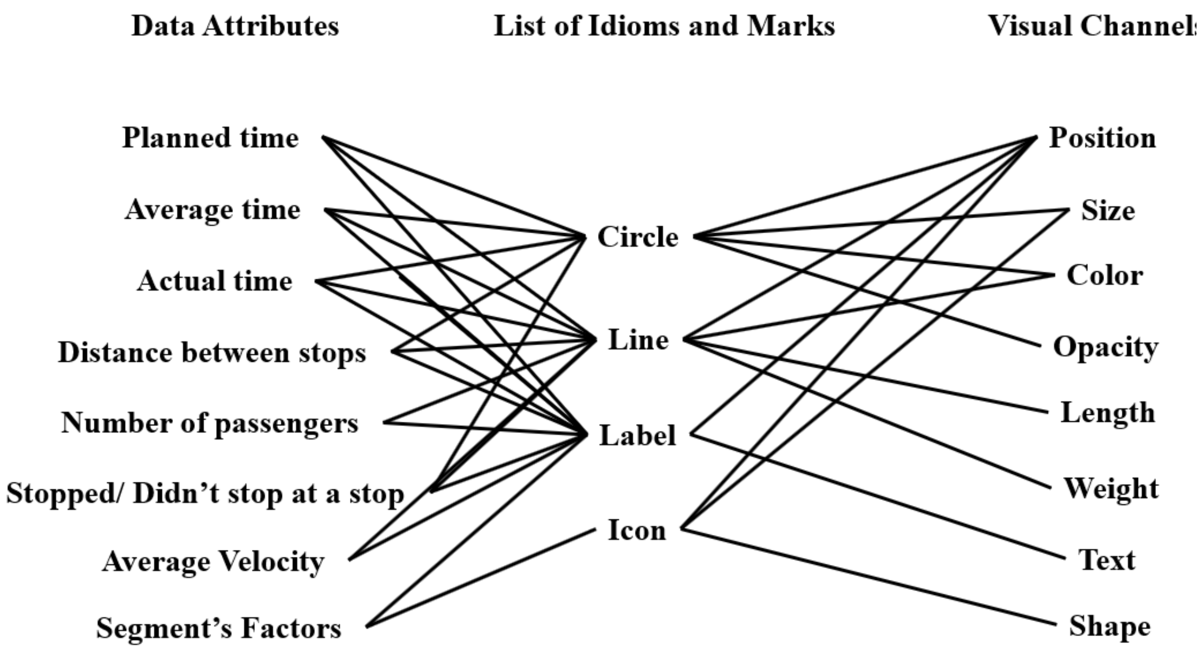


Figure 26: The design space with all the possibilities for creating the SSM visualization

As shown in Figure 26, the relevant data can be specified to these data attributes:

- **Planned time** (Quantitative) - representation of the times according to the schedule.
- **Average time** (Quantitative) - representation of the average times of the selected bus line according to the whole schedule times of the bus line in the data.
- **Actual time** (Quantitative) - representation of the times according to historic data collected by the Ministry of transportation.
- **Distance between stops** (Quantitative) - representation in meters of the distance between two neighboring stops.
- **Number of passengers** (Quantitative) - representation of the number of passengers on the bus on each segment.
- **Stopped/ did not stop at a stop** (Categorical) **-** representation of the stops on which the bus stopped/did not stop and open its doors.
- **Average velocity** (Quantitative) - representation of average speed of the bus.
- **Segment characteristics** (Categorical & Quantitative) **-** representation of roundabouts, traffic lights, and turns without roundabout or traffic light on the segment.

The four marks (circle, line, label, and icon) and their visual channels allow to visualize the data attributes from different points of views:

- **Circle:**

a. Position – planned time/actual time/average time/distance between neighboring stops
b. Radius size **-** number of passengers who boarded and alighted/number passengers who boarded/number of passengers who alighted/a subtraction of the number of passengers who alighted from the number of passengers who boarded at a given stop
c. Color **-** bus stopped/did not stop
d. Opacity **-** bus stopped/ did not stop

- **Line:**
  a. Position – planned time/actual time/average time/distance between neighboring stops
  b. Length – planned time/actual time/average time/distance between neighboring stops
  c. Weight – number of passengers on the bus throughout the segment/time between neighboring stops
  d. Color – average velocity of the bus/number of passengers
- **Label:**
  a. Position – beginning of the line (to the right of it)/end of the line (to the left of it)/above the circle (arrival time)/beneath the circle
  b. Text – planned time/actual time/average time/departure time/distance between neighboring stops/numbers (of passengers, of roundabouts, etc.)
- **Icon**:
  a. Position – above/beneath the line
  b. Size – number of traffic lights/turns/roundabouts
  c. Shape – roundabout sign/traffic light/turn sign

5.3.2. Objective and Rational

The objective of creating this visualization is to monitor trip delays and the density of passengers in a straightforward and clear manner, and to identify patterns such as long-tail trips. As mentioned in the Related Work section, some previous visualizations encoded the geographic data of public transportation on a map. Locating stops and segments on a map helps passengers to navigate, find bus routes near attractions and workplaces, estimate the distance between two locations, and so forth. However, as a visualization designed for monitoring purposes, the main goal of this proposed visualization is to avoid irrelevant data as much as possible. By showing only relevant data, we can help experts find the problems and patterns they are looking for. Thus, the decision to put the map aside was made consciously in order to juxtapose the planned data on the bus line (as they are published in the operator's site) and the real-time data in a simple and straightforward way.

Moreover, to compensate for the absence of a map and the value it holds, displayed on the different segments of the bus line are the relevant road characteristics, such as roundabouts, traffic lights, and road turns. In terms of “ink-data ratio” calculation, we saved unnecessary graphic display this way. The decision to display each bus line and trip as lines was inspired by the design of metro maps. The display is clean, simple (circles representing stops and lines between consecutive stops representing segments), and easy to use for both traveling and analyzing purposes. The distinction between the planned trip and the actual trip is meant for performing comparisons and detecting the segments and bus lines that do not behave as expected.

5.3.3. The Overall System - Usability

The URL of the SSM visualization:

http://proj.ise.bgu.ac.il/liadn/Bus_stops_Weekly.html

Figure 27 displays the main menu of the SSM visualization. The three drop lists on the left let the user choose the bus line and the specific date and time of the trip. To the right of the drop lists, the user can choose the “Alignment to Left” option, which aligns all of the lines to the same left point, and the “Diagonal Lines” option, which creates the curtain view. Each stop is connected to itself through the different trips on the canvas (e.g., Figure 33) and lets the user see the delays at the different stops and identify the stops where the delays become early arrivals. The next buttons are “DISPLAY”, which runs the report, and “CANCEL”, which stops generating the report. The last blue button “…” toggles the advanced options.

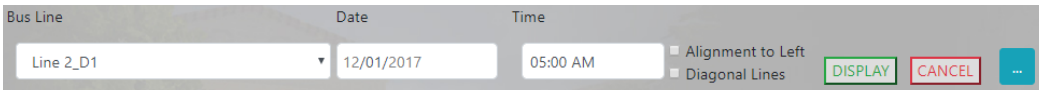


Figure 27: Main menu of the SSM visualization.

Figure 29 displays the advanced options of the SSM visualization. The options are divided into four main categories:

- **Speed and Timeliness Coloring** – allows the user to color the lines according to delays and bus speed.
    a. **Regular** – using black and white lines
    b. **Early-Late Cumulative** – colors the entire trip line red if the bus was delayed more than 1 minute, blue if it arrived more than 1 minute early, and gray if it arrived on time.
    c. **Early-Late Relative** – color each stop on the trip line red if the bus was delayed in reaching the specific stop, blue if it arrived early, and gray if it arrived on time.

d. **Speed** – colors each segment on the trip line red if the bus drove at a speed lower than 15 km/h, blue if it drove at a speed greater than 25 km/h, and gray if it drove at a speed between 15-25 km/h.

Figure 28 Shows the legend of the colors for the bus speed and time. Each category is divided into a seven colors palette.

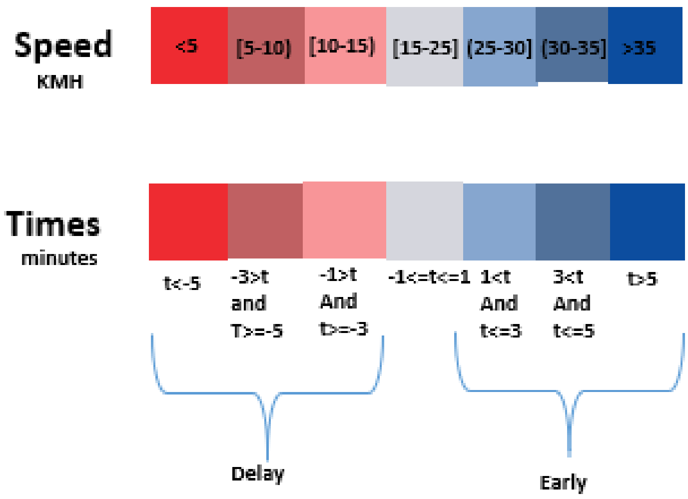


Figure 28: Color legend by bus speed and time.

- **Passenger Movement** – defines the use of the stops, by passengers who alighted, who boarded, or both
  a. **Alight**ed **-** the stops radius on the trip line is normalized only by the number of passengers who alighted
  b. **Board -** the stops radius on the trip line is normalized only by the number of passengers who boarded
  c. **Both clicked** –the stops radius on the trip line is normalized by the number of passengers who alighted and who boarded
- **Time Display Options** – formats the time display
  a. **Cumulative** – displays the minutes on the stops on the actual line it took the bus in total to arrive. For example, stop 1 – 0 minutes, stop 2 – 1 minute, stop 3 – 4 minutes, stop 4 – 6 minutes etc.
  b. **Relative** – displays the minutes of the stops on the actual line it took the bus to arrive between the two neighbor stops. For example (like the numbers in the cumulative above), stop 1 – 0 minutes, stop 2 – 1 minute, stop 3 – 3 minutes, stop 4 – 2 minutes etc.
- **Line View** – varying types of reports
  a. **Single Trip** – displays a report on a specific trip
  b. **Daily** – displays a report on all trips on a chosen date
  c. **Weekly** – displays a report on all trips on a chosen time during the week

d. **Monthly** – displays a report on all trips on a chosen time during the month

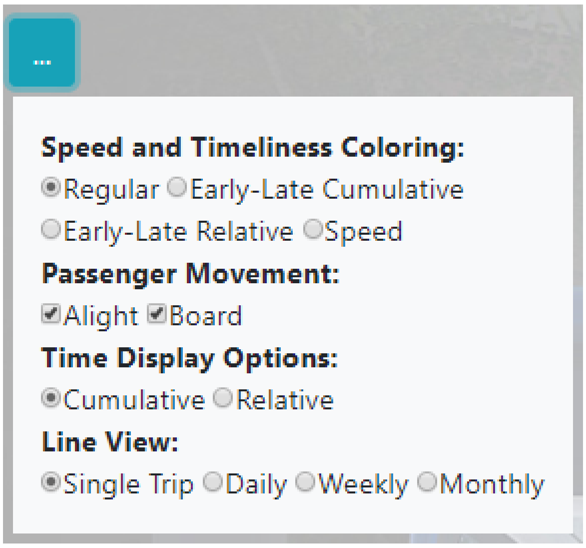


Figure 29: The "Advanced Options" menu.

6. SSM Visualization

6.1. Our Proposed Design

The Shadowed Segment Maps (SSM) visualization was developed by following two main principles: the first is that improvement must come from knowing the situation de facto, in this specific project, knowing the actual data compared to what is expected is crucial. The second principle is related to the visual part. Just as the shadow of the same person can become longer, shorter or deviate to the sides, the estimated data (the planned data for each route) will be represented by a line indicating the stop's position on the route over planned arrival time, while the actual trip data will be represented as the first line's shadow. For example, if the trip started before the designated time, the shadow line representing it will start more to the left of the starting point of the planned data line, and if it ended after the estimated time, the shadow line will be extended to the right of the end point of the planned data line).

Figure 30 depicts the following story: The user is requesting data on the trip performed by Bus Line 2 (in the upper left corner (Line2_D1)) on 01.12.17 at 05:30. The upper line represents the planned data for this bus line (31 circles representing 31 stops along the route from left to right) and the estimated time periods for arrival at each stop from the beginning of the trip are shown above each respective circle. The timer starts at the first stop as 00:00:00 and stops at the last stop with a time of 00:41:00 (which means it took the bus 41 minutes to complete the trip). Each line stretching between two neighboring stops represents the segment. Beneath each segment appear the three icons of the following segment characteristics: roundabouts, traffic lights, and road turns, as well as the number of each of them. Table 3 summarizes the hover options available to the user, depending on the specific line (planned or actual) and stop or segment. For example, if the user hovers with the cursor over a planned segment, a pop-up noting the length of the segment in meters and the time in minutes will appear. Another example, if the user hovers with the cursor over a stop on the actual line, a pop-up noting the stop's ID, the number of passengers who boarded and alighted at this stop, and the time the bus doors were opened and closed will appear.

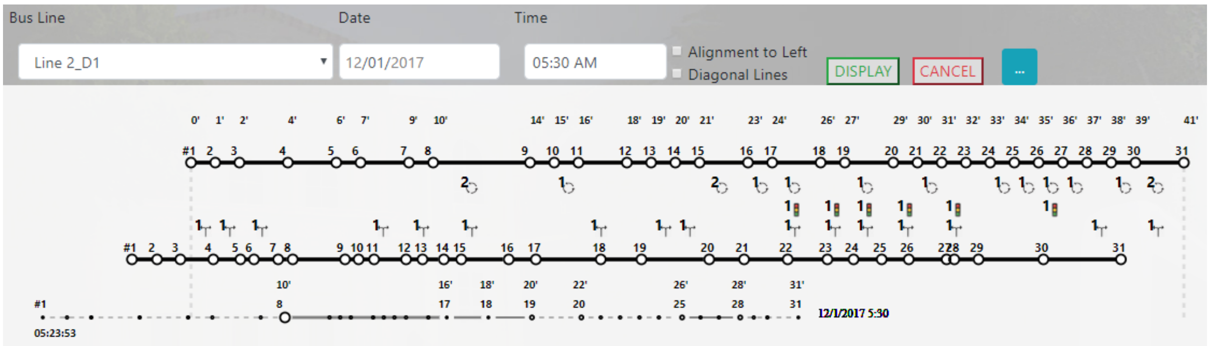


Figure 30: A POC (Prove of Concept) for the SSM visualization. The upper line represents the general data of bus line 2 while the bottom line represents the trip performed by this line on 01.12.17 at 05:30.

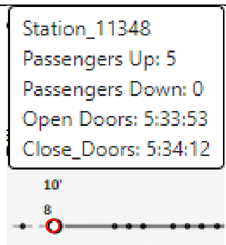
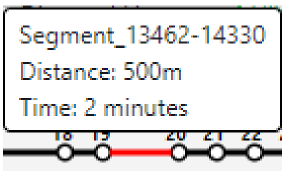
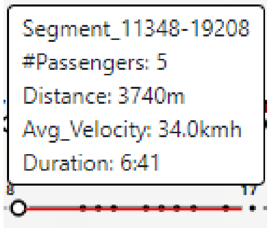
	Planned	**Actual**
Stop		
Segment		

Table 3: SSM hover pop-up options

The middle line (average line) represents the average trip duration of the selected route, e.g., line2_D1, from all of the data in our database from the Israeli Ministry of Transportation.

The bottom line (shadowed line) represents the actual trip performed by Bus Line 2_D1 on 01.12.17 at 05:30. The trip actually started at 05:23:53 (the time the bus closed its doors and started the trip), and because it started earlier than 05:30, the bottom line starts to the left of the upper line. Here too the circles represent the stops and the lines represent the segments, but the size of the circles depends on the number of passengers that the specific stop served, meaning passengers who either boarded the bus or alighted from it at that stop. For example, stop 1 is represented by a dot (a circle with a very small radius) because 0 passengers boarded and 0 passengers alighted from the bus at this point, while at stop 8 (the second circle on the shadowed line) 5 passengers boarded the bus, and this is represented by a bigger circle. All the dots without an ID number are stops at which the bus did not stop (for example, stops 2, 3, 4, 5, 6, 7, 9, 10, etc.). There is no data about the time the bus arrived at these stops and left them, thus, we used a calculation do determine this. For example, the segment between stop 28 and stop 31 takes 4 minutes (according to the plan), but only 3 minutes de facto (actual line), thus, the calculation to determine the position of stop 29 is based on the relative duration on the planned line. The time between stop 28 to 29 is 1 minute, the time between stop 29 to 30 is 1 minute, and the time between stop 30 to 31 is 2 minutes, so the position of stop 29 is a quarter of the segment's time: (0.25*3 minutes = 0.75).

The weight of the line represents the number of passengers on the bus on a given segment, excluding a case of 0 or 1 passengers, which is represented by a dashed line. For example, the line representing the segment between stop 1 and stop 8 is dashed, because no passengers were aboard the bus when it started the trip up to stop 8. The weight of the line between stop 8 and stop 17 is 5 passengers (the weight of the line is equal to the number of passengers on the bus on the segment divided by the maximal number of passengers on the bus during the trip).

Except for the first stop, the times noted above all of the other stops represent the time the bus doors were opened. As previously mentioned, the time noted above the first stop represents the time the bus began its trip. Hovering with the cursor over the stops (circles) on the shadowed line will display information about the stop, including its ID, the time the bus doors were opened and closed, and the number of passengers who boarded and alighted at this stop. Similarly, hovering with the cursor over a segment will display the segment's ID, the number of passengers who were aboard the bus on this segment, the length of the segment, the average velocity of the bus on this segment, and the duration of the trip on this segment.

To sum up all the details presented in Figure 30regarding the actual trip that took place on 01.12.17 at 05:30 (the shadowed line), the bus set out six minutes earlier than planned, had a long tail until stop 8 (no passengers boarded until stop 8), the entire trip took 31 minutes to complete, and the bus served only eight passengers from eight stops (out of a total of 31 stops). After examining the story of the route driven by Bus Line 2 on 01.12.17 at 05:30, Figure 31 depicts the same route driven on 01.12.17 at 07:30. The story displayed in this figure is completely different: this trip took 52 minutes to complete, served 28 out of the 31 stops. Between stop 16 and stop 17, 71 passengers were aboard the bus, almost twice the expected capacity of the bus (the bus is meant to serve a maximum of 34 passengers). Despite the designated starting time of 07:30, the bus started its trip at 07:29:28. Using a label only for the departure time is meant to help the user in cases where the bus started a seconds before after planned time, it’s hard to tell by only looking and the difference between the planned and the actual lines.

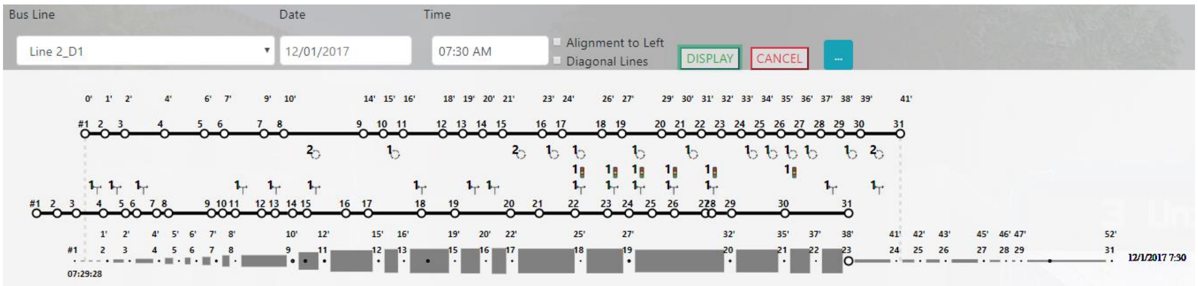


Figure 31: A POC for SSM visualization. The upper line represents the general data of bus line 2 while the bottom line represents the trip this bus line performed on 01.12.17 at 07:30.

Our proposed visualization has two additional features: a weekly report and a daily report. Figure 32 depicts the trips performed by Bus Line 2 at 07:30 over the first week of December 2017 (we received data on only three trips from the Ministry of Transportation). In all three trips, the bus set out from the first stop earlier than expected.

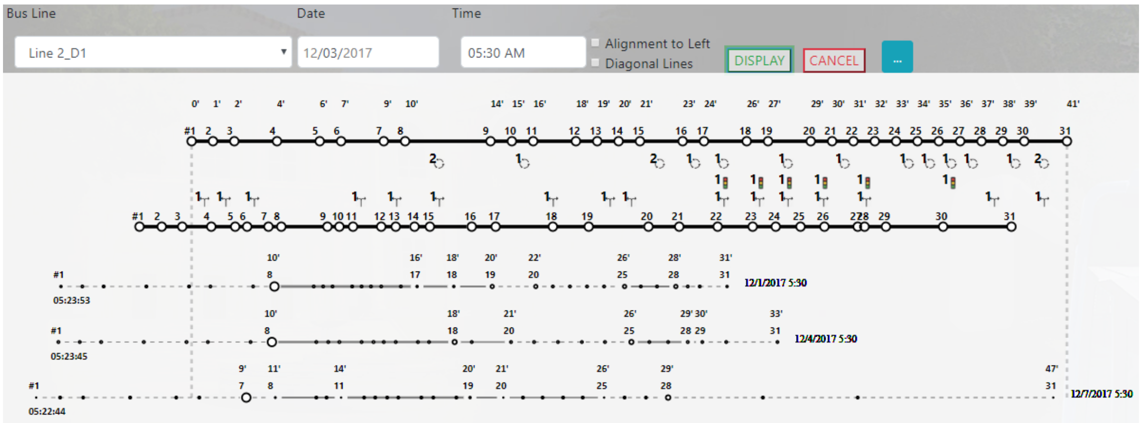


Figure 32: A Snapshot from the SSM visualization. The upper line represents the general data of bus line 2 while the bottom lines represent the trips performed during the week at 05:30.

Figure 33 Depicts a daily report on Bus Line 2 from 01.12.17.

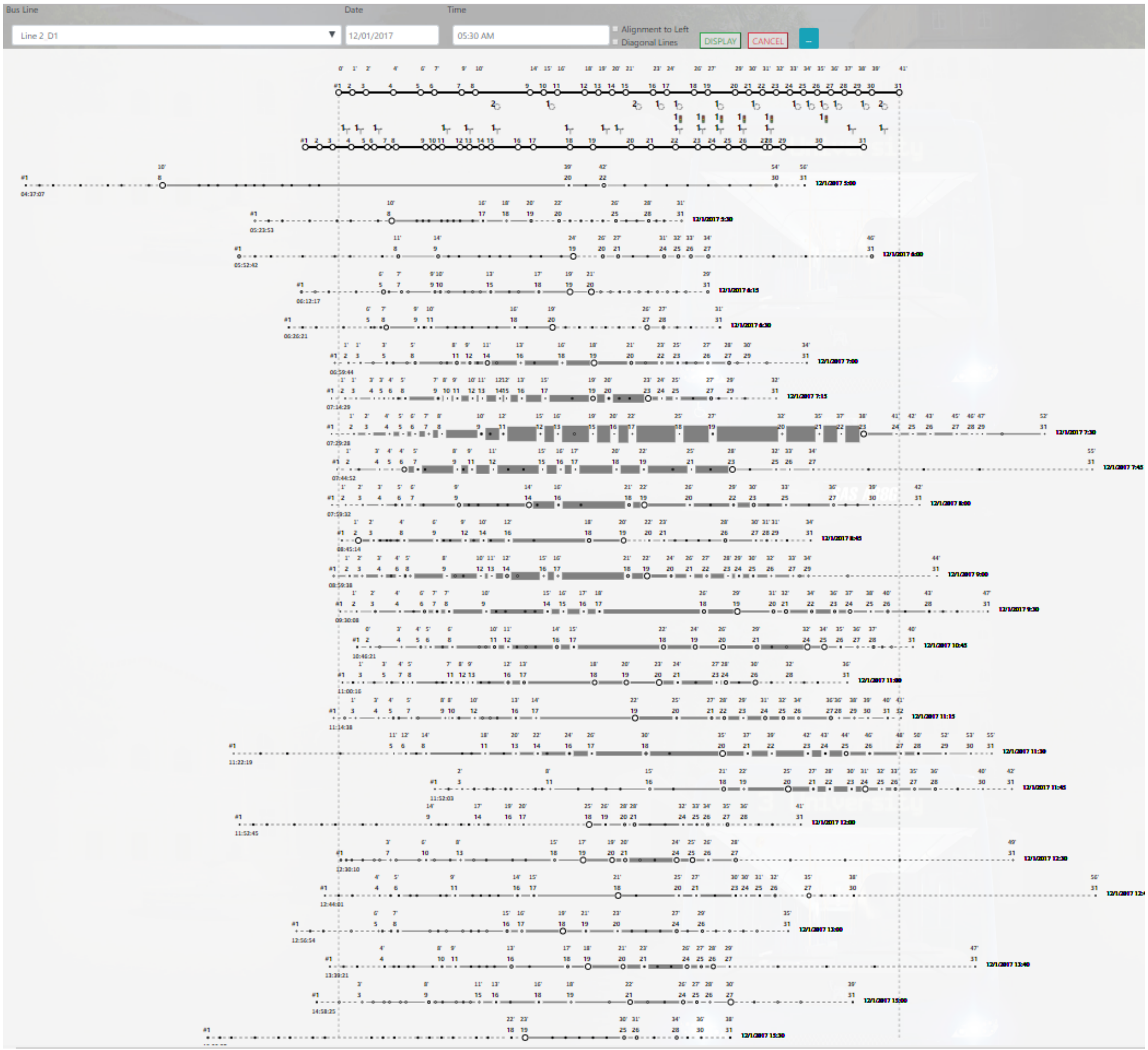


Figure 33: A Snapshot from the SSM visualization. The upper line represents the general data of bus line 2 while the bottom lines represent some of the trips performed on 01.12.17. The background image is from url: https://steamcommunity.com/market/listings/753/324310-SAS%20A38G%20%28Profile%20Background%29?l=danish

To give the visualization more power, we used three visual channels (position, color, and line) to emphasize the user tasks.

1. **Position – Alignment to Left**

   The fundamental principal of the SSM visualization is the time comparisons between the planned line and the actual line based on their starting and end positions on the canvas. It is difficult to compare bus lines that start at different locations (e.g., Figure 33), especially in the daily reports. Thus, we have created the "Alignment to Left" option. By choosing this option, all of the lines will be organized from the same starting point on the left side of the canvas. This aids in finding delays by quickly looking at the "finish line", and also aids in identifying the longest delays.

Figure 34 focuses on single trip performed by Bus Line 2 in direction 1 on 01.12.17 at 05:30. The Alignment to Left option quickly reveals that the actual line is shorter than the planed line, which implies that the bus arrived at the final stop earlier than planned.

Figure 35 provides another a good example of the usefulness of this option. It is easy to see that the first actual line (the third from the top) is the longest trip (56 minutes) on this specific day.

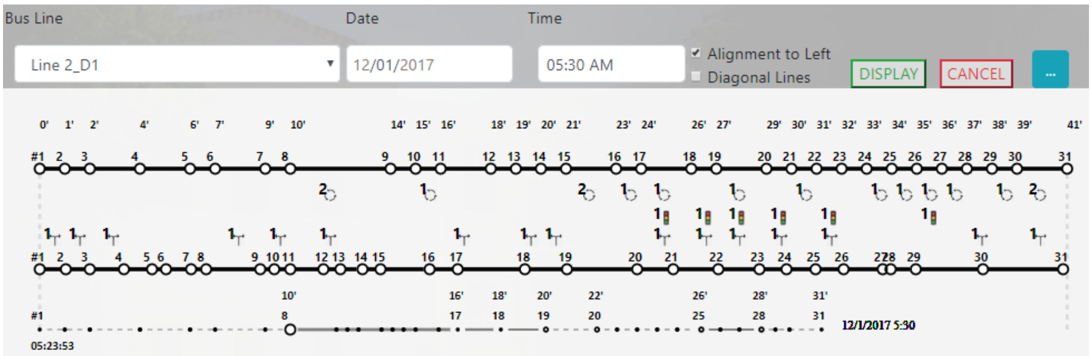


Figure 34: A Snapshot from the SSM visualization. Compared to Figure 30, the same line data and time are displayed, but in this snapshot the "Alignment to Left" option was activated, causing all of the lines to start at the same point.

2. Color – color for speed and time

Adding colors to the SSM lines provides another efficient way of looking at the bus lines' data. The user can quickly detect delays and early arrivals, as well as define patterns of rush hours, find slot times at which it would be beneficial to increase the frequency of bus arrivals, and monitor the bus's speed for finding problematic segments (slow driving speed) and monitoring the drivers. Figure 28 depicts the project's color palette.

Figure 35 displays a daily report on Bus Line 2 direction 1 from 01.12.17. The red lines represent delays and the blue lines represent early arrivals. Thus, it is easy to see that between 06:15-07:15 the roads were empty and the trips were completed ahead of time, but at 07:30 we can see the morning rush hours.

Figure 35: A Snapshot from the SSM visualization. Using the "Early-Late Cumulative" and "Alignment to Left" options, the daily report displays the delays (red lines) and the early arrivals (blue lines).

Figure 36 displays the same daily report as Figure 35, but this time, the segments' colors represent the bus's speed. The segments colored in red can tell where the bus drove too slowly, most likely because of traffic jams or traffic lights. For example, the segment between stop 18 and stop 19 seems to be problematic, perhaps because it is one of the six segments which include traffic lights.

Figure 36: A Snapshot from the SSM visualization. Using the "Speed" and "Alignment to Left" options, the daily report displays the speed of the bus by segment (red for slow speed and blue for high speed).

Using Figure 35 in conjunction with Figure 36 allows the user to analyze delays at the single trip level (Figure 35) and the bus speed at the segment level (Figure 36) in order to find problematic segment(s) that may be the cause of delays.

3. Line – Diagonal Lines ("Curtains" of Stops)

A unique pattern which is incorporated into the SSM visualization is the “curtains of stops”. By choosing the “Diagonal Lines” option, the stops between the different trips will be connected by lines that create a curtain-like pattern. Figure 37 displays a good example of a curtain. Each connecting line that moves left (from the top down) represents an early arrival, while a line moving to the right represents a delay (i.e., the bus arrived at the station later than it should have).

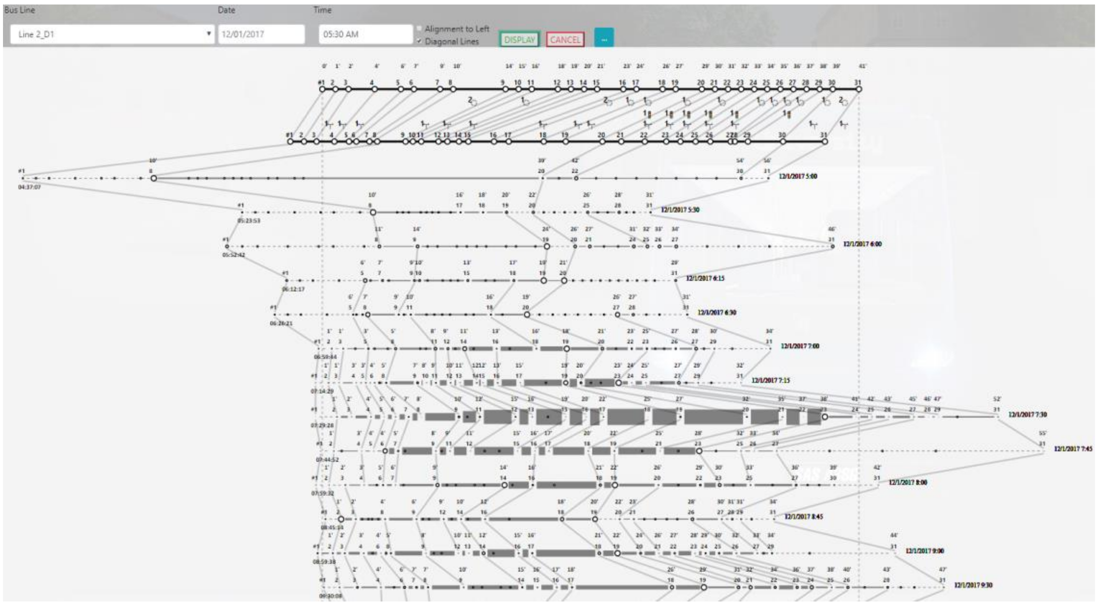


Figure 37: A Snapshot from the SSM visualization. By using a line to connect the same stop across different trips, we can define delays and early arrivals by the angle of the connecting line.

Figure 38 displays the combination of "Diagonal Lines" and “Alignment to Left” options. It is apparent that all the first stops are connected by the same vertical start line.

This combination can improve the user's understanding of patterns of delays and early arrivals compared to the planned trip. Some insights from this combination include:

- In morning trips, the bus did not stop until stop 5.
- Stop 20 seems to be a good stop to analyze. In most trips, the bus stopped at this stop, and if it arrived at this stop ahead of the planned time, the trip was completed either on time or ahead of time (for example, the trips that started at 05:00, 07:00, 07:15, etc.). However, if the bus arrived at this stop after the planned time (for example, the trips that started at 07:30, 09:00, 09:30, etc.), the entire trip was completed after the planned time.

- In most trips, the bus was delayed when arriving at stop 27 (the connecting line moves from left to right).

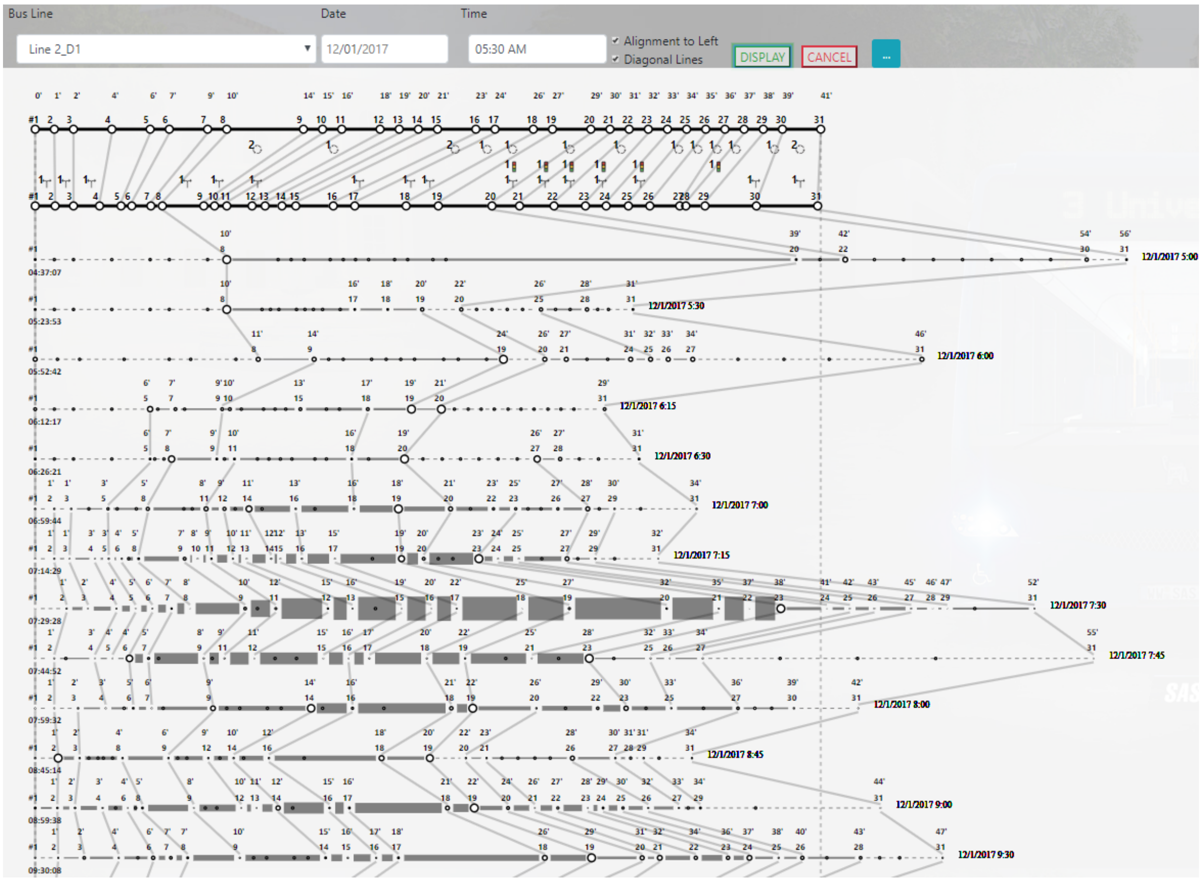


Figure 38: A Snapshot from the SSM visualization. Alignment to Left with Diagonal Lines.

6.2. The SSM Visualization's Specific Design Space

As shown in Section 5.3.1, the design space defines all the possibilities for creating the SSM visualization. Figure 39 displays the specific design space used for the creation of the SSM visualization, as presented in Sections 6.1 and 6.2.

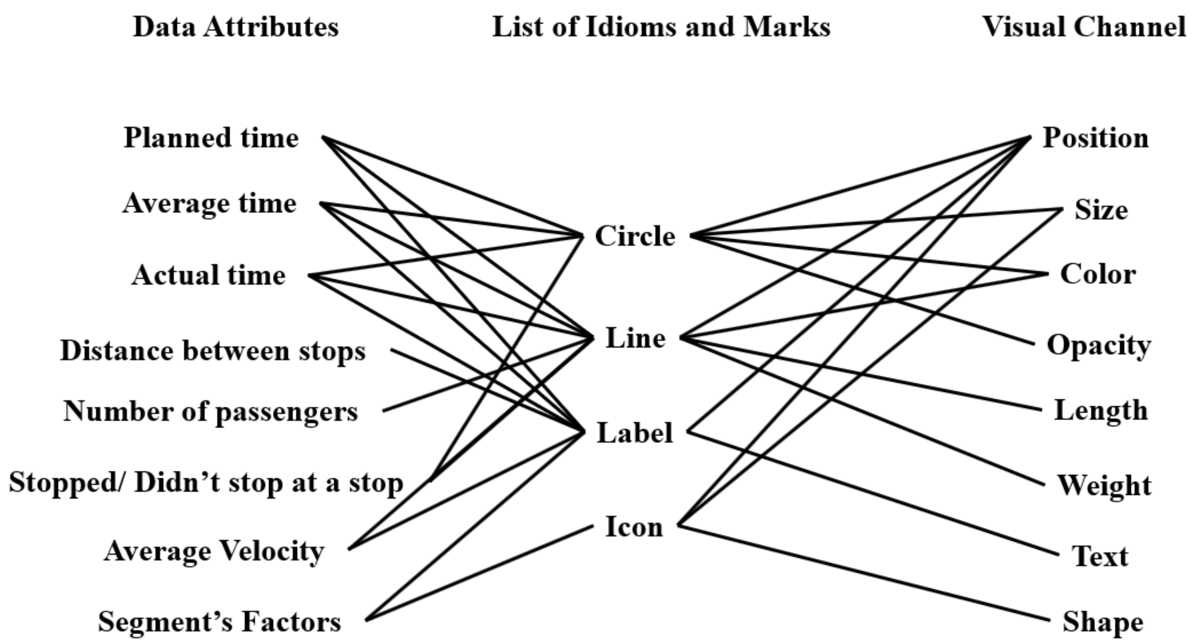


Figure 39: The specific design space for creating the SSM visualization

Figure 39 is the result of deep contemplation regarding the ultimate way of representing the user tasks. The most salient representations and their rationale are listed below.

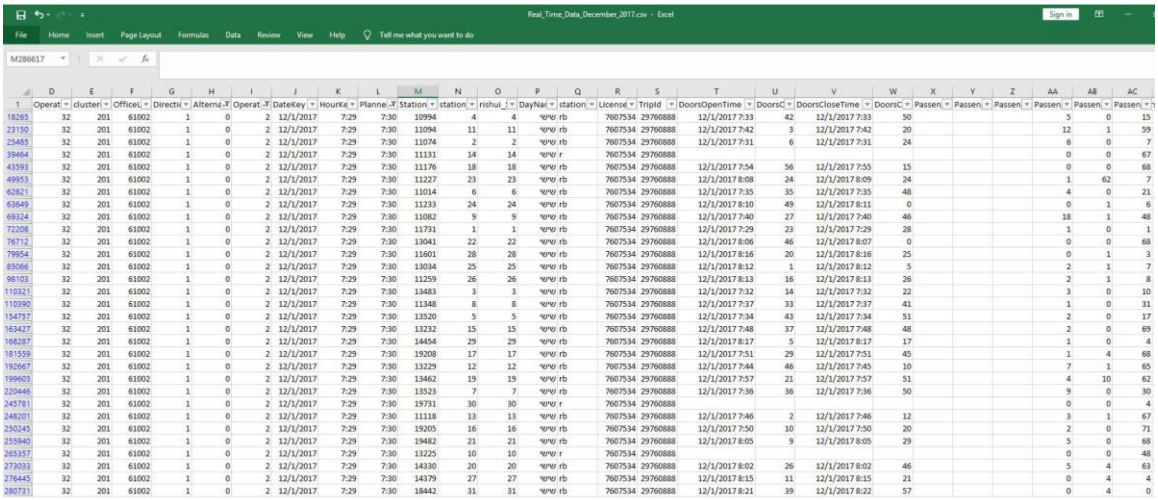
- The unique aspect of the SSM visualization is using a design similar to that of metro maps, but unlike metro maps, the distance between two stops is defined by the trip duration and not the physical distance between them. This way, the SSM visualization allows the user to compare the planned data with the actual data.
- The number of passengers on the bus affects the weight of the line, hence it easy to observe segments with many passengers, as well as long-tail trips. Specifically, for long-tail trips, using a dashed line instead of a line with a weight of 0 (when there are no passengers on the bus) allows long-tails to pop up to the expert.
- The stops that the bus did not use (where it kept driving and did not open the doors for passengers to board or alight) are displayed on a calculated position on the line, and allow the user to gain more information regarding the trip (for example, section 6.1 page 34-35). The decision to display those stops was made to try to tell the expert the "whole story".
- The segment's characteristics (roundabouts, traffic lights, and road turns) are displayed beneath each segment of the planned line (the upper line) with an icon and a label indicating their number, to compensate for the absence of a geographic map.
- The average velocity is an attribute calculated by dividing the distances from google maps by the actual time. Displaying this data attribute using the color channel on actual lines is also an option, and may help determine whether the bus drove too fast or too slowly (most likely because of the segment's characteristics or traffic jams, etc.). Moreover, after the experts'

review, we became aware that the systems used by experts at “Ayalon Highways” do not display this information.

- The last data attribute is the dashed vertical lines at the start and end of the planned time. These lines indicate if the trips started/arrived earlier or later than planned, especially in the daily, weekly and monthly reports.

6.3. Data vs Visualization

To emphasize the advantages of using our visualization, consider the existing source of data for similar analyses, which is a real-time data excel sheet from the Ministry of Transportation of Israel, shown in Figure 40. Figure 40 shows the actual data of the same trip, performed by line 2_D1 on 01.12.17 at 07:30. The SSM visualization (Figure 41) displays the same data much more succinctly. It also includes the planning data and the average data.


Real_Time_Data_December_2017.csv - Excel

M286617

	D	E	F	G	H	I	J	K	L	M	N	O
1	Operat	clusteri	OfficeL	Directi	Alterna	Operat	DateKey	HourKe	Planne	Station	station	rishul_
18265	32	201	61002	1	0	2	12/1/2017	7:29	7:30	10994	4	4
23150	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11094	11	11
25485	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11074	2	2
39464	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11131	14	14
43593	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11176	18	18
49953	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11227	23	23
62821	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11014	6	6
63649	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11233	24	24
69324	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11082	9	9
72208	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11731	1	1
76712	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13041	22	22
79954	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11601	28	28
85066	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13034	25	25
98103	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11259	26	26
110321	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13483	3	3
110390	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11348	8	8
154757	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13520	5	5
163427	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13232	15	15
168287	32	201	61002	1	0	2	12/1/2017	7:29	7:30	14454	29	29
181559	32	201	61002	1	0	2	12/1/2017	7:29	7:30	19208	17	17
192667	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13229	12	12
199603	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13462	19	19
220446	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13523	7	7
245781	32	201	61002	1	0	2	12/1/2017	7:29	7:30	19731	30	30
248201	32	201	61002	1	0	2	12/1/2017	7:29	7:30	11118	13	13
250245	32	201	61002	1	0	2	12/1/2017	7:29	7:30	19205	16	16
255940	32	201	61002	1	0	2	12/1/2017	7:29	7:30	19482	21	21
265357	32	201	61002	1	0	2	12/1/2017	7:29	7:30	13225	10	10
273033	32	201	61002	1	0	2	12/1/2017	7:29	7:30	14330	20	20
276445	32	201	61002	1	0	2	12/1/2017	7:29	7:30	14379	27	27
280731	32	201	61002	1	0	2	12/1/2017	7:29	7:30	18442	31	31

	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	DayNar	station	License	TripId	DoorsOpenTime	DoorsC	DoorsCloseTime	DoorsC	Passen	Passen	Passen	Passen	Passen	Passen
18265	שישי	rb	7607534	29760888	12/1/2017 7:33	42	12/1/2017 7:33	50				5	0	15
23150	שישי	rb	7607534	29760888	12/1/2017 7:42	3	12/1/2017 7:42	20				12	1	59
25485	שישי	rb	7607534	29760888	12/1/2017 7:31	6	12/1/2017 7:31	24				6	0	7
39464	שישי	r	7607534	29760888								0	0	67
43593	שישי	rb	7607534	29760888	12/1/2017 7:54	56	12/1/2017 7:55	15				0	0	68
49953	שישי	rb	7607534	29760888	12/1/2017 8:08	24	12/1/2017 8:09	24				1	62	7
62821	שישי	rb	7607534	29760888	12/1/2017 7:35	35	12/1/2017 7:35	48				4	0	21
63649	שישי	rb	7607534	29760888	12/1/2017 8:10	49	12/1/2017 8:11	0				0	1	6
69324	שישי	rb	7607534	29760888	12/1/2017 7:40	27	12/1/2017 7:40	46				18	1	48
72208	שישי	rb	7607534	29760888	12/1/2017 7:29	23	12/1/2017 7:29	28				1	0	1
76712	שישי	rb	7607534	29760888	12/1/2017 8:06	46	12/1/2017 8:07	0				0	0	68
79954	שישי	rb	7607534	29760888	12/1/2017 8:16	20	12/1/2017 8:16	25				0	1	3
85066	שישי	rb	7607534	29760888	12/1/2017 8:12	1	12/1/2017 8:12	5				2	1	7
98103	שישי	rb	7607534	29760888	12/1/2017 8:13	16	12/1/2017 8:13	26				2	1	8
110321	שישי	rb	7607534	29760888	12/1/2017 7:32	14	12/1/2017 7:32	22				3	0	10
110390	שישי	rb	7607534	29760888	12/1/2017 7:37	33	12/1/2017 7:37	41				1	0	31
154757	שישי	rb	7607534	29760888	12/1/2017 7:34	43	12/1/2017 7:34	51				2	0	17
163427	שישי	rb	7607534	29760888	12/1/2017 7:48	37	12/1/2017 7:48	48				2	0	69
168287	שישי	rb	7607534	29760888	12/1/2017 8:17	5	12/1/2017 8:17	17				1	0	4
181559	שישי	rb	7607534	29760888	12/1/2017 7:51	29	12/1/2017 7:51	45				1	4	68
192667	שישי	rb	7607534	29760888	12/1/2017 7:44	46	12/1/2017 7:45	10				7	1	65
199603	שישי	rb	7607534	29760888	12/1/2017 7:57	21	12/1/2017 7:57	51				4	10	62
220446	שישי	rb	7607534	29760888	12/1/2017 7:36	36	12/1/2017 7:36	50				9	0	30
245781	שישי	r	7607534	29760888								0	0	4
248201	שישי	rb	7607534	29760888	12/1/2017 7:46	2	12/1/2017 7:46	12				3	1	67
250245	שישי	rb	7607534	29760888	12/1/2017 7:50	10	12/1/2017 7:50	20				2	0	71
255940	שישי	rb	7607534	29760888	12/1/2017 8:05	9	12/1/2017 8:05	29				5	0	68
265357	שישי	r	7607534	29760888								0	0	48
273033	שישי	rb	7607534	29760888	12/1/2017 8:02	26	12/1/2017 8:02	46				5	4	63
276445	שישי	rb	7607534	29760888	12/1/2017 8:15	11	12/1/2017 8:15	21				0	4	4
280731	שישי	rb	7607534	29760888	12/1/2017 8:21	39	12/1/2017 8:22	57				0	4	0

Figure 40: The actual data from the ministry of Transportation of Israel - Line 6_D1 01.12.17 07:30

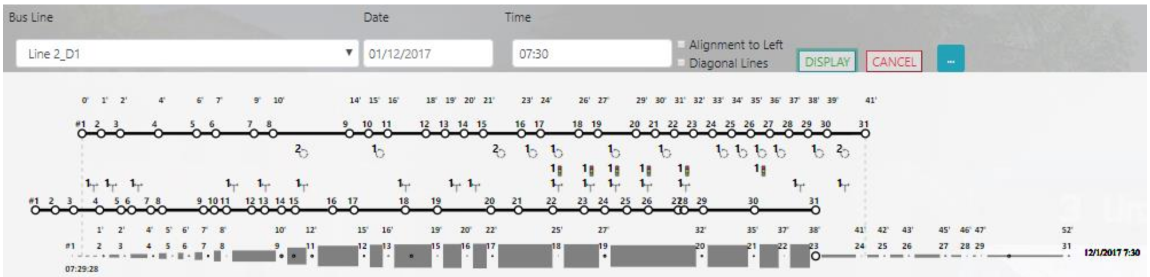


Figure 41: SSM visualization - Line 6_D1 01.12.17 07:30

Moreover, Figure 40 and Figure 41 depict only one trip. When analyzing a daily/weekly/monthly report, the data displayed as one snippet using our visualization provides a big advantage. Our visualization of this data makes drawing conclusions quicker and simpler, as it is much easier to draw such conclusions from a snippet than to start filtering an excel sheet.

Figure 42 and Figure 43 illustrate this big advantage by showing a weekly report of trips performed by line 6_D1 at 07:00 AM. The SSM visualization allows the user to see immediately and clearly long tails, which trip started before time, which stops were the busiest (the circles with the greatest diameter), etc., while using the excel sheet will require considerably more time to get the same answers (23 stops * 4 trips = 92 records).

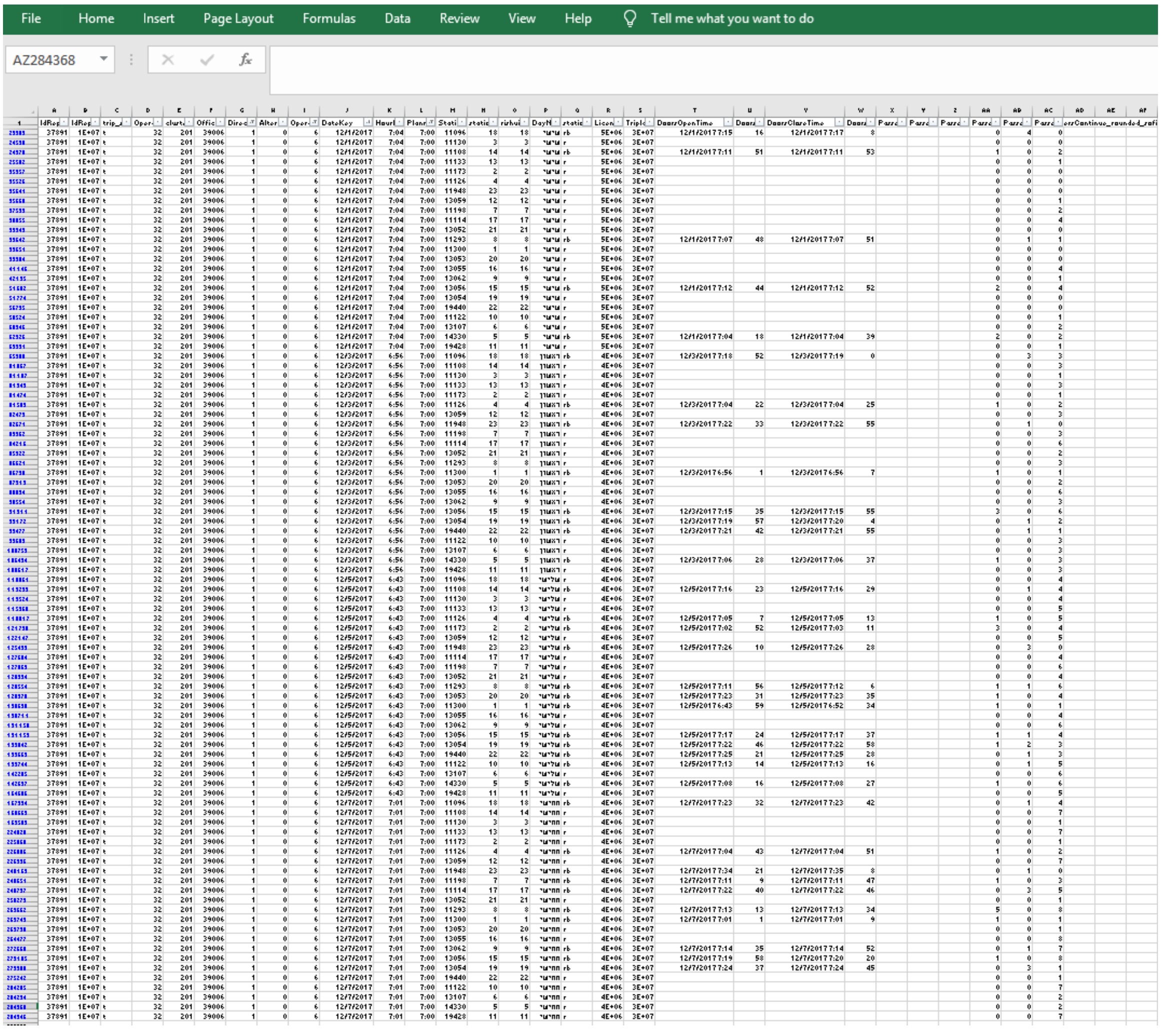


Figure 42: The actual data from the ministry of Transportation of Israel - Line 6_D1 weekly report of 07:00 (4 different days - 92 records)

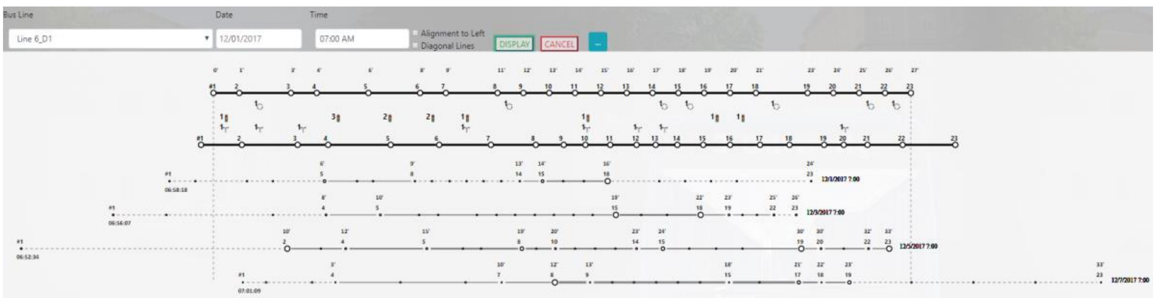


Figure 43: SSM visualization - Line 6_D1 weekly report of 07:00

6.4. Special Examples

6.4.1. Busy Segment

Figure 44 shows that during the trip performed by Line 2_D1 on 01.12.17 at 07:30, between stop 16 and stop 17 (segment 19205-19208) there were 71 passengers on the bus. According to Figure 3, this is more than twice the capacity of a crosstown bus (32-34).

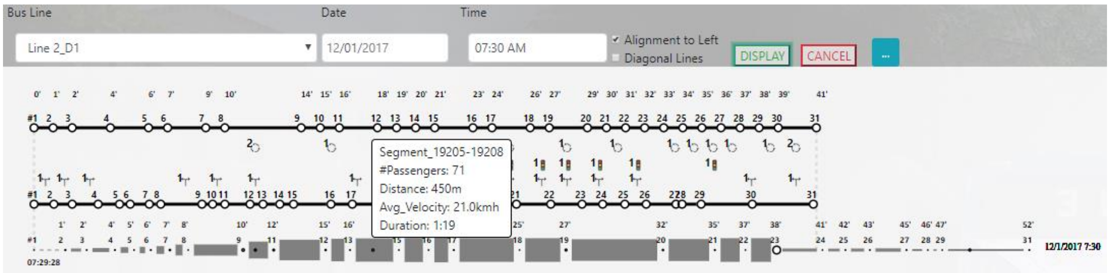


Figure 44: SSM visualization - Line 2_D1 on 01.12.17 07:30. The white pop up is by hover with the mouse on the line between stop 16 to 17

6.4.2. Long-Tails

Figure 45 shows that during the week, beginning at 05:30 and until stop 7 or 8, the bus did not pick up any passengers (or only one). The same thing happened from stop 28 or 29. This information can help planners create a dynamic plan for this line. For example, such a plan may allow the bus to skip several stops early in the morning or change its route altogether.

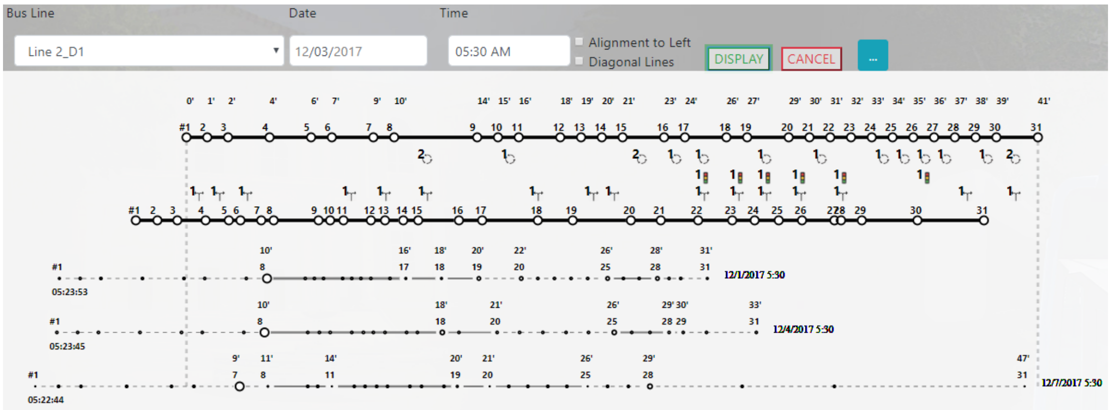


Figure 45: SSM visualization - Line 2_D1 on 05:30 Weekly Report

6.4.3. *Rush hours*

Figure 46 shows the Line 4_D1 rush hours on Sunday (03.12.17). It is clear that between 5:20-6:45 the bus arrived at the last stop earlier than planned, while between 07:00-8:36 it arrived later than planned.

More delays in arrival started at 13:12.

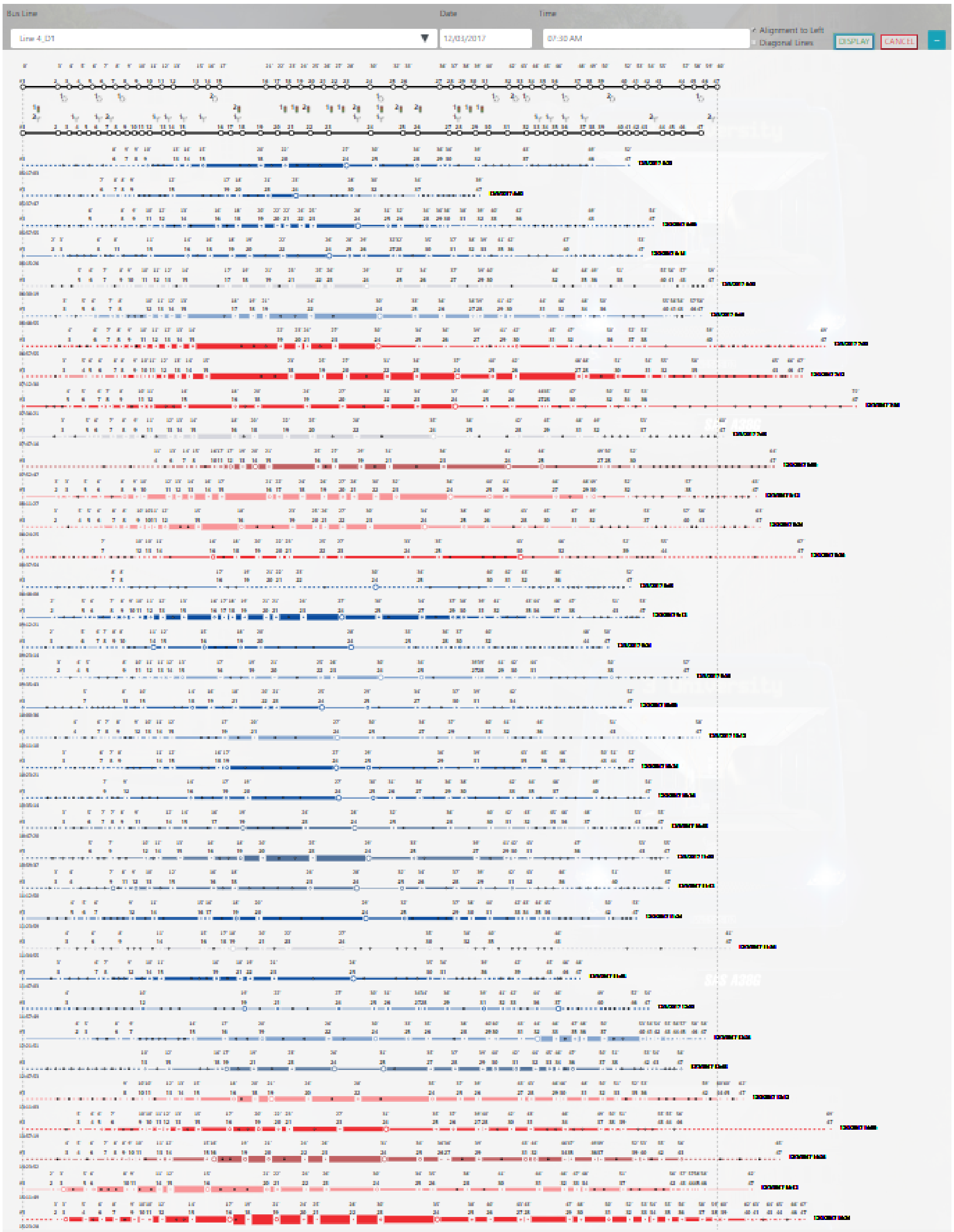


Figure 46: SSM visualization - Line 4_D1 on 03.12.17 Daily report colored by Early-Late Cumulative.

6.4.4. Low Speed

Figure 47 shows the daily trips of line 3_D1. A careful look reveals that between Stop 5 and Stop 7, the bus travels at a low speed in most of the trips.

Figure 47: SSM visualization - Line 3_D1 on 04.12.17 Daily report colored by average speed.

By using a map and a picture from the area, we can see why there is low speed issue between Stop 5 and Stops 7, as showed in Figure 47. Figure 48, Figure 49 and Figure 50 help understand that the bus needs to pass 5 lanes in order to turn to the left. This situation caused the bus to drive slowly, especially in the rush hours.

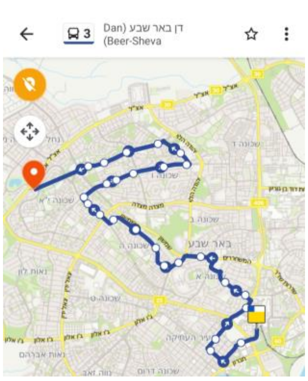


Figure 48: Line 3_D1 total route on a map from Moovit.

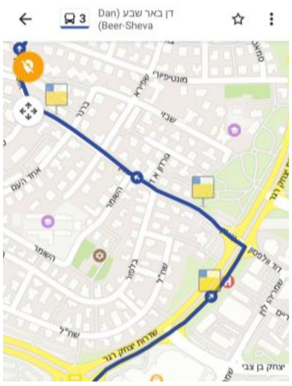


Figure 49: Line 3_D1 stops 5 to stop 7 on a map from Moovit.

Figure 50: Line 3_D1 stop 5 from google view.

6.5. Existing Alternative Design

The SSM visualization presented in our work is not the first one created for monitoring purposes. TransCAD[2] is a Geographic Information System (GIS) designed specifically for use by transportation professionals for the purpose of storing, displaying, managing, and analyzing transportation data. It combines GIS and transportation modeling capabilities in a single integrated platform, thus providing capabilities that are unmatched by any other package. “Ayalon Highway[3]” is an Israeli governmental company specializing in management of projects in the field of infrastructure and transportation. It is responsible for the construction and development of urban transport roads in central Israel. The company uses TransCAD for analyzing and monitoring bus trips. Figure 51 displays a snippet from TransCAD depicting the daily data on bus line 18 in Tel Aviv in 2013. Each stop is visualized as a red bus icon. Above the icon are both a pie chart and a label showing the number of passengers that boarded and alighted during the chosen period. Pop-up windows allow users to change preferences and create excel reports using the data.

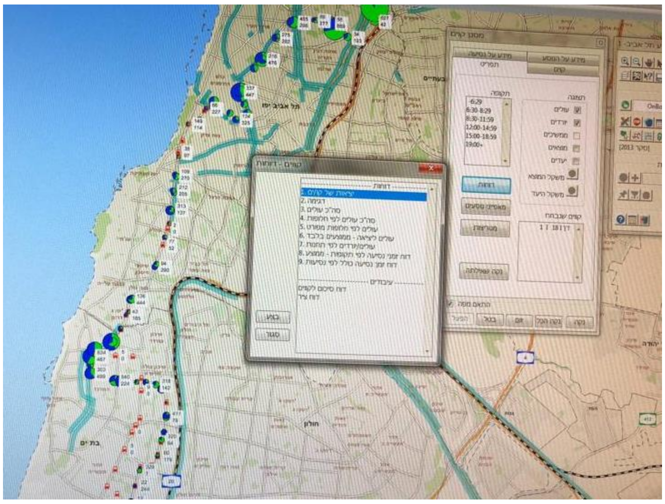


Figure 51: TransCAD snippet of a trip analysis based on a specific bus line's data

The visualization depicted in Figure 51 is used only for showing the total number of passengers during a specific period of a bus line's trip on a map (it does not include timepoints from real-time data). To analyze the data for details, the “Ayalon Highway” personnel need to export them to an excel spreadsheet (an output from TransCAD).

6.6. *Evaluation Process*

Evaluation of the SSM visualization can be based on either novice or expert evaluators. Given the novelty and exploratory nature of the SSM, a more qualitative data collection method is preferred,

---

[2] https://www.caliper.com/tcovu.htm
[3] http://www.ayalonhw.co.il/pages/160.aspx

that would provide rich findings and potentially useful insights by relying on expert users. Hence, we recruited a small group of 7 experts from the field of public transportation. The experts were asked to evaluate the SSM visualization and to compare it to the visualizations they are currently working with. An elaborated experts' review is provided in Section 7.

7. EXPERTS' REVIEW

This section presents in depth the experts' review, the questions it included, background information on each of the experts, and their evaluations. As the evaluation was conducted in Hebrew, Appendix A and Appendix B include the full experts' review interview and the experts' notes in Hebrew to preserve all the details.

7.1. The Experts

Seven experts participated in our experts' review interview. The interview was carried out anonymously; Therefore, we will only expose the experts' professional background, gender and position. We will use an identifying symbol based on the order they evaluated our system (the first expert will be referred to as "expert 1", etc.):

1. **Expert 1 (Male)** – A professor from the Department of Geography and Environmental Development at Ben Gurion University. Main research interests include investigating public transportation traveling habits using technological tools for the purpose of creating a balanced transportation system.
2. **Expert 2 (Male)** – A civil engineer from the Technion - Israel Institute of Technology. At his previous position was involved in planning the future design of public transportation based on surveys. At his current position, involved in creating a model of the Tel-Aviv metropolis in future years (based on new buildings, new roads, etc.). Both of these positions were at the "Ayalon Highways" company.
3. **Expert 3 (Male)** – Head of the Planning Department at the "Ayalon Highways" company. His previous positions included planning of public transportation in the Jerusalem metropolis and serving as the assistant supervisor of the Dan metropolis public transportation. He also replaced the supervisor of the Beer Sheva metropolis public transportation for several months.
4. **Expert 4 (Male)** – Statistician and head of the Data Collection Department at the "Ayalon Highways" company. In charge of developing the information systems for the company.
5. **Expert 5 (Male)** – A member of the Data Collection Department at the "Ayalon Highways" company. In the past months his work focused on the Beer Sheva metropolis, and before then, on the Dan metropolis.

6. **Expert 6 (Male)** – A civil engineer from the Technion institute. Works at the Modeling Department at the “Ayalon Highways” company.
7. **Expert 7 (Female)** – Has an economy B.A Degree from Ben Gurion University. Participated in a public cadet program of the public civil service. Currently works at Transportation Department at the Beer Sheva municipality, works in conjunction with the “Dan Beer Sheva” operator in receiving public complaints regarding public transportation delays and problems. Her job includes placing bus stops and increasing the efficiency of the bus lines in Beer Sheva.

7.2. The Interview

Based on the papers presented in Section 4 under “Experts' Review”, the interview consisted of two main phases: the practical use of the visualization, and the phase including open and closed questions to receive feedback.

The interview for this project included:

1. Personal introduction – personal introduction and a short conversation about the expert's background
2. Short introduction of the visualization – explaining the main objective and providing a short explanation about the visualization and the possible insights that may be derived from working with it.
3. Two pilot tasks – the experts followed the tasks and were asked for their insights after performing them. The two tasks included 16 questions.
4. Free time with the visualization – the expert got some free time (5-7 minutes) to explore the visualization. 4 optional tasks were provided as well. In addition, at the end of the interview, the experts received the visualization's URL to allow them free use of it.
5. Final survey – consisted of 7 specific questions about the pros and cons of the visualization.

The full interview is provided in Appendix A.

7.3. Two Pilot Tasks

Table 4 presents the questions asked as part of the interview and the experts' performance in the pilot tasks.

Task ID	Task Description	Question [correct answer]	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7
1	**Choose "Line3_D2" on 04.12.17 at 07:17 AM and under "…" choose "Single-trip"**	**How many stops did this bus line have?** [31 stops]	V	V	V	V	V	V	V
		How many roundabouts, traffic lights, and road turns are between stop 26 and stop 27 (Segment 13459-11119)? [0 roundabouts, 3 traffic lights, 1 road turn]	V	V	V	V	V	V	V
		How long did the total planned trip take? [40 minutes]	V	V	V	V	V	V	V
		Did the bus stop in every stop? [no, for example, the bus did not stop between stops 1-5]	V	V	V	V	V	V	V
		When did the bus start the trip? [At 16:59:07 PM]	V	V	V	V	V	V	V
		How many passengers were aboard the bus between stop 20 and stop 21? What is the distance between these two stops? [28 passengers, 350 meters]	V	V	V	V	V	V	V
		Did the bus start the trip on time? [Yes]	V	V	V	V	V	V	V
		Was the duration of trip as planned? Was it shorter or longer? [No, the trip duration was shorter than planned]	V	V	V	V	V	V	V

2	**Choose "Line16A_D2" on 03.12.17 at 16:30 PM and under "…" choose "Weekly" and "Cumulative"**	**Did the bus line start its trip ahead of time/late/on time?** [On time]	Did not complete task number 2 because he did not like working with the system	V	V	V	V	V	V
		Were most of the trips conducted during the week completed on time/late/ahead of time? [Late]		V	V	V	V	V	V
	Under "…" choose "Speed"	**On which segments did the bus drive at a speed of less than 10 km/h (deep blue color)?** [Segments 9-10, 11-12, 23-24]		**X**	**X**	**X**	V	V	V
		On which segments did the bus drive at a speed of more than 25 km/h (deep red color)? [Segments 18-20, 25-30]		V	V	**X**	V	V	V
	Choose "Alignment to Left" and under "…" choose "Daily" and "Early-Late Cumulative"	**Was the entire trip completed on time/late/ahead of time?** [Late]		V	V	V	V	V	V
		When was the longest trip performed? [On 15:45 PM, the trip took 56 minutes]		V	V	V	V	V	V

Table 4: The Results of the Two Pilot Tasks

7.4. Final Survay

At the end of the evaluation process, the experts were asked to answer seven identical questions. Table 5**Error! Reference source not found.** elaborates the questions and the answers:

ID	Question	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7
1	Did you find the visualization useful for your job?	No	No (not working on the single line level at my current job)	Yes	Yes	Yes	Yes	Yes
2	Would you want to work with the visualization as a standard system at your work?	No	Yes (the interface is good)	Yes (Obviously)	Yes	Yes (Very much)	Yes (Very much)	Yes
3	Did you find the average line helpful?	No (Confusing)	Yes	No	No	No	No	Yes
4	Was the lack of a map helpful/not helpful/neutral?	Not helpful (The lack of a map does not allow working with the system)	Neutral (Prefer adding a small map)	Neutral (Prefer adding a small map)	Helpful	Helpful	Neutral (Prefer adding a small map)	Neutral (Prefer adding a small map)
5	What parts of the visualization did you like? what are its strengths compared to the systems you are currently using at your job?	Did not like	The dashed line for long tails	The varying types of reports	The lack of a map	The lack of a map	Simple	More advanced than the existing and future systems at my work
6	Which features did you find unhelpful?	Better screen utilization	The roundabouts, traffic lights, and road	Changing the speed color palette	The weight of the line should be changed to 5	Nothing specific	Nothing specific	Nothing specific

			turns. Also, the speed color palette and more vertical space between two lines		levels, and using statistics schema for speed (not range of speeds but mean and standard deviation)			
7	What features would you like to add to this visualization?	Connection to the map, but did not find the visualization useful	Small map on the right corner	Small map on the right corner	Small map on the right corner but did not find it useful	Small map on the right corner	Small map on the right corner	Small map on the right corner and emphasis on **the lights with high priority to public** transportatio n

Table 5: Final Survey

7.5. Results

The experts' review shed light on the strengths and weaknesses of the visualization. Moreover, the experts provided suggestions for improvement that could serve more tasks and users.

As mentioned in this thesis, the main goal of our project was to monitor the public transportation system for the purpose of improving the transportation schedule and understanding the status on the roads, with an emphasis on trip delays. More specifically, our visualization was built based on 9 user tasks (described in Section 5.2.1) which were formulated in conjunction with other experts, which did not participate in our final experts' review.

Our project not only contributes to the understanding of trip delays, but also created a system which is superior to those currently used by experts in the field (at least in Israel) in some aspects. Our visualization provides benefits in terms of monitoring the bus's speed (“**The speed component does not exist in the current systems**”, *Expert 3*), as a tool for investigating trip delays (“**The visualization in its current status can be used as a tool for verifying public complaints regarding delays and other problems in the bus lines**”, *Expert 3 & Expert 4*), and for defining new ways of looking at the public transportation data (“**The comparison by times and not by distances is really good. It is easy to see if there is delay or an early arrival**”, *Expert 6*; “**The "curtains" of stops are very good, but I believe that 90% of the users will find them unclear**”, *Expert 4*). Nevertheless, it is important to emphasize that not all of the experts found the SSM visualization to be helpful or useful (“**The lack of use of a geographic map renders the system unusable. I personally did not like the system at all**”, *Expert 1*).

Based on the experts' evaluation, we identified 4 aspects for improving the SSM in future development efforts:

- **Adding a map -** the full separation from the map is a salient characteristic of the SSM visualization. As we thought, for analysis work, the map did not add any value, but most of the experts still suggested that a map be included (e.g., as a small insert on the upper right corner) to receive the “geographic context” of the bus line they are examining. Moreover, it seems that the segment characteristics we added (traffic lights, roundabouts, and road turns) at every segment, did not provide the expected added value.
- **Modifying the color palette** – both the pilot tasks and the final survey showed that using the “cold-hot” color palette for defining delays and early arrivals was intuitive and easy to understand for all of the experts, but when it came to representation of the bus's speed, it was unreadable and confusing (if a delay is red, then the low speed is supposed to be red, because it caused the delay). It seems that the speed function should receive a different color palette which is not the same as the one used for delays.
- **Customizing the average line** – most of the experts found this line irrelevant and confusing. The line was created based on all the data we gathered on each specific bus

line. Future customization of this line (such as displaying only the average of weekdays/weekends, etc.) will improve its integrity and make it relevant to the SSM visualization.

- **Allowing customization** – as this project dealt with visualization, the system used to run it was designed to be as simple as possible. From the experts' review, it seemed that while they enjoyed this simplicity, they were not against complicating it in order to have more freedom to change features such as the divide of the colors palette, the space between the actual lines (e.g., in the daily/weekly/monthly report), etc.

8. CONCLUSIONS

In this thesis, we addressed the problem of monitoring public transportation delays by proposing a novel interactive visualization system. The SSM visualization harnesses the advantages of familiar visual properties, such as circles, lines, and colors, to create a new way of analyzing geographic data without using a map.

Furthermore, the fundamentals of the SSM include a comparison between planned and actual trips, which provides a strong ability to monitor one specific trip or a daily/weekly/monthly trips, for the purpose of analysis and improvement.

The evaluation of the SSM visualization included 7 experts in the field of public transportation (a transportation researcher at Ben Gurion University, 5 engineers from the “Ayalon Highways” company and an expert from the Public Transportation Department of the Beer Sheva municipality). They performed two pilot tasks and answered a survey regarding the advantages and disadvantages of the SSM visualization. Most of them responded with a positive feedback, and even emphasized the advantages of the SSM technique over their existing tools.

Although some limitations exist regarding the use of our visualization system, it provides a novel and practical solution to the problem of monitoring the performance of public transportation and can be considered a solid foundation for a new type of public transportation monitoring applications.

9. Future Work

9.1. Limitaions

9.1.1. The POC visualization we built works for the Beer Sheva metropolitan area. It can work for other metropolises, but this would require a short data collection session (bus lines, bus stops, distances, etc.) for the planned line display and the actual data for the actual line display.

9.1.2. The future infrastructure of the visualization system must include new methods such as node.js, which enable faster loading of big data (our current infrastructure can hold 1 month of data at good working speed).

9.1.3. The geographic connection to the map is not currently available. This means that working with visualization is reserved almost exclusively to experts in the field. A future connection to geographic data (i.e., a map) could help more users work with the system.

9.1.4. The POC visualization menu we built is as simple as possible to allow quick learning of how to use the visualization, and to display as much data as possible in a single snapshot. There are many possible options the researchers/experts can use, and asked for, such adding options as choosing the boundaries of the driving speed and delays, changing the color palette for their convenience, changing the spacing between the lines, changing the font size, etc., that were previously hidden from them in the first version of the POC.

9.1.5. The current infrastructure does not support real time tracking of bus lines trips but only historic data. For planning needs it is enough but for real time monitoring of the current status of the system, displaying the data in real time it is crucial.

9.1.6. The visualization focused on the individual bus line, its “story” and does not connect it to the whole system in the metropolitan. It does display stops and segments that share couple bus lines and the consequences from that. When we are dealing delays of the public transportation, it is important to remember we are dealing the whole system but for the first step starts from the individual to the system (inside out).

9.1.7. The visualization does not deal with dynamic planning. For now, the ministry of transportation of Israel provides only one planning data for each line. If this situation will be change, it will require another future work to compare the new data and with the actual data.

9.2. Suggestions – The Ministry of Transportation

9.2.1. In order to create a full and correct baseline, opening and closing the bus doors when arriving at a stop and when leaving it, respectively, should be mandatory, even if there are no passengers at the stop. This is important for binding the trip to the correct scale.

9.2.2. The data collected in the Ministry of Transportation's database is very unorganized. For example, there are duplicated stops on the same trip. We recommended checking the raw data received from the different operators.

9.2.3. Asking the different operators provide dynamic planning data. The visualization emphasis the importance of dynamic planning - the planning data of the early morning hours supposed to be different than the rush hours and also different between weekdays and weekends.

9.3. Suggestions – Future Work on the Visualization

9.3.1. Show faded out trips on the schedule we do not have specific data from the ministry of transportation. For example, if we do not have the data on the trip performed by line 2 in direction 1 on 01.12.17 at 05:00, leave a space for it without a line or a line with high opacity.

9.3.2. When looking at the visualization as a system meant for the user, it must include more freedom: an option to change the color boundaries (every 5 km/h, by mean and standard deviation, etc.), choosing a range of hours for a specific bus line, controlling the space between the actual lines, etc.

9.3.3. Finding a free data and adding a note near the traffic lights with system that gives preference to public transportation. Moreover, any other open source data that could shed more light on the segments – as pedestrian crossings, bumpers etc.

9.3.4. The average line should be customizable to the users. For now, it includes all the data we have in our database, but should be changeable to the average of the same day, only weekdays, only weekends, etc.

9.3.5. Using the power of this visualization as a single line which organizes the data from a range of hours, days and by rush hours, weekends, holidays, etc., to display a summary of a specific bus line. For example, displaying all the data on Line2_D1 on 01.12.17 between 13:00-15:00 will display statistical information (sum/average/max/min) on the passengers on the different segments, the average time of opening/closing the bus doors on a specific stop, average speed, etc.

9.3.6. Adequate the visualization to other public transportation vehicle as trains.

10. APPENDIX A – THE EXPERTS' REVIEW INTERVIEW

- בקשה לאפשר להקליט את הפגישה (לציין כי זה ראיון אנונימי)

1. <u>הכרות</u> **(5-7 דקות)**
   1. על עצמי: סטודנט לתואר שני במחלקה להנדסת מערכות מידע באוניברסיטת בן גוריון. בנוסף עובד כ-Data Scientist בחברת סטארט-אפ.
   2. על הפרויקט: מטרת הפרויקט הינה לספק מעקב אחר התנהלות התחבורה הציבורית בהשוואה ללו"ז המתוכנן על מנת לקבלת תמונת מצב אמיתית. אנו מציגים ויזואליזציה חדשה שנקראת "**מפות סגמנטים מוצללות**" לטובת ניטור מערך האוטובוסים באמצעות הצגת מה שהתרחש בזמן אמת והשוואתו לתכנון.
   3. אודות הנבחן:
      1. תיאור קצר על הרקע המקצועי והתפקיד
      2. מערכות מידע איתם עובד
      3. פרויקטים ראשיים/צרכי עבודה מהמערכות

2. <u>הצגה קצרה של הויזואליזציה</u> **(5-7 דקות)**
   1. בחירת קו "Line2_D1" בתאריך ה-01.12.17 בשעה 07:30
      1. הסברים על השוני בין שלושת הקווים
      2. כמות הנוסעים – עובי הקו (להראות שבמקטע מסוים עלו 71 נוסעים – נתון חריג)
      3. כיצד לזהות איחור – המיקום של הקווים הוא לפי זמנים
      4. השימוש בסוגי הדו"חות (יומי/שבועי) ותצוגת הזמנים והצבעים
      5. הצגת ה"וילונות" (החיבורים בין התחנות)
      6. תחנות שלא עצר בהן
      7. הסבר על ה-pop up מהקו המתוכנן והפועל
      8. הצמדה לשמאל
      9. הבדל בין הזמנים של הקו בפועל (זמן תחילת נסיעה משמאל למטה בעוד הדקות של משך הנסיעה עצמה למעלה)
      10. צביעה לפי מהירות וזמנים

3. <u>שתי משימות פיילוט</u> **(10-20 דקות)**
   1. בחר "line 3_D2 " בתאריך ה-04.12.17 בשעה 07:17 ותחת "..." בחר single-trip. לחץ "Display"
      1. **קו מתוכנן**
         1. כמה תחנות יש לקו? **[31 תחנות]**
         2. כמה כיכרות, רמזורים ופניות בין תחנה 26 לתחנה 27 (סגמנט 11119-13459)? **[0 כיכרות, 3 רמזורים, 1 פניות]**
         3. מהו הזמן הכולל של הקו? **[40 דקות]**
      2. **קו בפועל**

1. האם האוטובוס עצר בכל התחנות? **[לא, לדוגמא, בין החנות 1ל-5 הוא לא עצר]**
2. מתי האוטובוס התחיל את נסיעתו? **[בשעה 07:16:59]**
3. כמה נוסעים היו על האוטובוס בין תחנה 20 לתחנה 21? ומהו המרחק בין התחנות הללו? **[28 נוסעים, 350 מטר]**

3. כלל הקווים

1. האם האוטובוס התחיל את נסיעתו בזמן? **[כן]**
2. האם משך הנסיעה היה כפי שתוכנן? קצר או ארוך יותר? **[לא, היה קצר יותר]**

2. בחר קו "line 16A_D2" בתאריך ה-03.12.17 בשעה 16:30 ותחת "..." בחר "Weekly" ו-"Cumulative". לחץ "Display".

1. האם הקו התחיל את נסיעתו מוקדם/מאוחר/בזמן? **[בזמן]**
2. האם רוב הנסיעות לאורך השבוע הסתיימו באיחור/בזמן/לפי הזמן? **[איחרו]**
3. תחת "..." בחר את האופציה של "Speed", ולחץ "Display".
4. באלו מקטעים האוטובוס נסע פחות מ-10 קמ"ש (צבע אדום כהה)? **[מקטעים 9-10, 11-12 ו-23-24]**
5. בתאריך ה-04.12.17, באילו מקטעים האוטובוס נסע ביותר מ-25 קמ"ש (צבע כחול כהה)? **[תחנות 18-20 ותחנות 25-30]**
6. בחר "Alignment to Left" ותחת "..." "Daily", "Early-Late Cumulative"
7. האם כל הנסיעות הסתיימו בזמן/באיחור/הקדימו? **[איחרו]**
8. מתי הייתה נסיעה שארכה הכי הרבה? **[בשעה 15:45, 53 דקות]**

4. <u>זמן חופשי</u> **(5-7 דקות)**

1. אופציונאלי:

1. האם קו "line 112A_D1" בתאריך 06.12.17 בשעה 18:30 איחר ומה משך הטיול בדקות? [**single-trip, date and time, cumulative, alignment to left/diagonal**]
2. מה התחנה בה עלו הכי הרבה נוסעים בקו "line 32_D1" בתאריך ה-01.12.17 בשעה 10:30? [**single-trip, date and time, board**]
3. כל הנסיעות בהן איחר קו "line 4_D1" בתאריך ה-03.12.17? [**daily report, alignment to left/cumulative, early-late cumulative**]
4. לאורך השבוע, הצג את כל האיחורים וההקדמות של קו "line 24_D1" בשעה 10:45? [**weekly, date and time, early-late cumulative, alignment to left**]

5. <u>שאלון מסכם</u> **(5-10 דקות)**

1. האם מצאת את הויזואליזציה שימושית לעבודתך?
2. האם היית רוצה להשתמש בויזואליזציה הנ"ל ככלי בעבודתך?
3. האם מצאת את הקו הממוצע תורם?
4. האם חוסר השימוש במפה טוב/חסר?

5. מה אהבת בויזואליזציה/ מה הן החוזקות שלה בהשוואה לכלים בהם אתה משתמש בעבודתך?
6. על אילו פיצ'רים הייתי מוותר?
7. אלו פיצ'רים נוספים היית רוצה להוסיף בעתיד לויזואליזציה?

11. APPENDIX B – THE EXPERTS' NOTES

במהלך הערכת המומחים, עלו מספר הערות מכל מומחה לגבי השימושיות, היעילות ונקודות לשיפור ושימור לויזואליזציה.

<u>מומחה 1</u>:

1. חוסר המפה לא מאפשר שימוש במערכת
2. הגוונים השחורים והאפורים מקשים ולא גורמים לרצות לעבוד עם המערכת
3. "לחשוב על הקו הממוצע" – למצע בחלוקה לפי שלוש קטגוריות: ימי השבוע, שישי ושבת
4. להוסיף מקרא בצד שמאל של הקווים (P – Planned, A – Average, R – Real Time)
5. תמונת הרקע מושכת את תשומת הלב
6. השוואה לפי זמנים ולא לפי מרחקים לא טובה
7. ניתן לוותר על הדיוק לרמת השניות
8. הוספת מפת הקו בצד ימין
9. להשתמש בשמות התחנות במקום ID תחנה
10. סימון תחנות שמאפשרות מעבר לקווים אחרים אל מול תחנות רגילות
11. במקום מספר הקו וכיוון (לדוגמא Line2_D1) לציין תחנת יעד ומוצא
12. ניצול טוב יותר של המסך בהצגת קו יחיד

<u>מומחה 2</u>:

1. ריווח של הקווים - מרגיש צפוף מדי
2. ייצוג הכיכרות, רמזורים ופניות הוא טוב אך היה מציג אותו באופן ויזואלי אחר (רק כמספר או לשמור אותם ב-Hoover של הסגמנט)
3. הצבעים כחול אדום לאיחורים מעולים, לעומת זאת במהירויות – לא מובן.
4. השוואה לbaseline של דן באר שבע היא בעייתית כי היא לא דינאמית (זמני הגעה לתחנות)
5. אהב את הקו המקווקו לטובת ייצוג הזנבות

<u>מומחה 3</u>:

1. ישנה חשיבות גדולה לוודוא תקינות הנתונים שמתקבלים ממשרד התחבורה. ציין כי כיום הם מקבלים נתונים שעברו פרמטרים מסוימים של תקינות.
2. ציין כי מרחב באר שבע הוא מצוין לעבודה מסוג זה כי 100% מהאוטובוסים מצוידים בחיישניי GPS, ספירת נוסעים וסגירה ופתיחת דלתות.
3. במרחב תל אביב יש חשיבות למספר הקו, הכיוון והאלטרנטיבה, ולכן ייתכן שייצוג לפי Line2_D1 הוא לא מספיק במרחבים אחרים – אבל ציין שבמערכת זו נוח לו ככה
4. ביקש לרווח את הקווים – צפוף לטעמו
5. "המערכת במצבה הנוכחי תתאים מאוד לעבודה של הגורמים שמקבלים את תלונות הציבור בנושא איחורים ובעיות באוטובוס."
6. עבודה מול הנסיעה הבודדת פחות מתאימה למתכננים.
7. דו"ח צביעה לפי איחורים הוא מצוין – "אין לנו כיום דו"ח שמאפשר משהו כזה"

8. היה שמח אם יהיה דו"ח של יום בשבוע, לדוגמא יום ג' לאורך כל החודש – יאפשר לראות בעייתיות ביום בשבוע ואולי לתכנן אחרת את הקווים.
9. כיום הם עובדים המון מול
10. "עם כל האהבה שלי למפות, יותר נוח לי לעבוד ככה". במידה ומוסיפים מפה קטנה בצד, היה שמח לקבל נקודות עניין כמו תחנות רכבת, בתי חולים, בתי עלמין וכו' (יעזור לניתוחים עתידיים).
11. עניין המהירויות הוא מצוין וחסר היום במערכות הקיימות – היה מעדיף את המהירויות כליבלים ולא צבעים.
12. דו"ח הוילונות – "מעניין מאוד, עוזר לראות עד כמה הקו יציב. דו"ח שכזה יהבהב למתכננים שהקו לא יציב. מאפשר להראות כמה בלאגן יש"
13. בהתייחס לוילונות, היה שמח אם ניתן לתת ציון מספרי לפי השונות של הקו.
14. כמערכת, הכלי נותן מענה לצורך של עד כמה הקו מלא, איפה עולים ויורדים רק אולי ברמות אגריגציה גבוהות יותר – לא של הנסיעה הבודדת.
15. סיכם "מקווה שנראה את זה בשימוש, כיום חסרה גישה טובה ומהירה למידע"

__מומחה 4:__

1. חלוקת מספר הנוסעים לסקלה של חמש ולא כנתון מצטבר: עד 10 נוסעים בישיבה, 10 נוסעים ועד שיש עדיין מקומות ישיבה (34-46), מלא בישיבה (34-36), עד 5 אנשים עומדים (נקרא "חצי עמידה"), עד עשרה נוסעים עומדים.
2. הוספה של מידע בסוף כל נסיעה ספציפית: סה"כ נוסעים, מספר ממשיכים מירבי, מקדם תחלופה (היחס בין העולים ליורדים לאורך הנסיעה)
3. "שימוש במהירות ממוצעת של 19.4 קמ"ש היא נתון מאוד הגיוני לעיר באר שבע, מעיד על עבודת הכנה טובה"
4. חלוקת המהירות היה מבצע לא לפי קטגוריות של 10 קמ"ש למעלה ולמטה אלא לפי ממוצע וסטיית תקן
5. דו"ח וילונות – טוב מאוד אבל חושב של-90% מהמשתמשים הוא לא יהיה מובן
6. "המערכת מרשימה, רואים שנעשתה פה הרבה עבודה טובה"
7. "המערכת יותר טובה מהמערכות שיש היום למתכננים. ישנם מספר שיפורים קטנים שציינתי אך המערכת היא כבר עכשיו מוצר מדף טוב מאוד שמתאים גם לבחון תלונות מהציבור"
8. "זה כלי טוב לתחבורה הציבורית. חשבתי שזה יהיה בזבוז זמן אבל אני שמח שהדור החדש כותב כלים ברמה"

__מומחה 5:__

1. הייתי מוותר על ייצוג הכיכרות, רמזורים ופניות – לא משרת אותי בצורה כלשהי
2. הצגת השעה של הקו בצד ימין וזמן הנסיעה המדויק בצד שמאל – טובה ונוחה לעבודה
3. "כמי שעובד בעיקר עם נתונים "יבשים" ו-Raw data חוסר המפה מוצדק לטעמי ואפילו הייתי אומר עדיף"
4. "השימוש במערכת היה פשוט ונוח"
5. מבין שזאת לא מטרה הויזואליזציה, אבל היה שמח לראות ממוצעים יומיים של כמויות עולים ויורדים

מומחה 6:

1. "השימוש במערכת היה פשוט ואינטואיטיבי"
2. הצגת המרחקים בין תחנות לפי זמן ולא לפי מיקום טובה, בקלות מבינים שיש איחור או הקדמה
3. הצגת מספר סידורי עולה לתחנות ושמירת ה-ID של התחנה ב-hoover מעולה.
4. הצבעים האדומים כחולים למהירויות הם בעייתיים
5. "מאוד נהניתי לעבוד על המערכת ביחד לממשקים של התוכנות הקיימות"

מומחה 7:

1. "לדעתי זה יכול להועיל"
2. "בונים לנו מערכת מידע יותר בסיסית ממה שרואים פה – זו באמת מערכת טובה"
3. לחשוב על הפרדה עם הזמן למול דוח יומי – לבחון איך
4. תצוגת הקו בצד שמאל למעלה – לפי שם תחנות ולא מק"ט
5. איפה עצר בפועל – סימון בצבע אחר של תחנה בה האוטובוס לא עצר ולא לפי מספור.
6. לבחון אפשרות של סימון ליד רמזור עם מערכת של העדפה לתחבורה ציבורית
7. השימוש בקו הוא טוב - ניתן להבין הרבה דברים אך החיבור למפה חסר
8. היעדר המפה – יכול להיות שזה עניין של שינוי דיסקט אבל היום עובדים רק עם מפות, קישור למפה של הסגמנט או התחנה
9. עוד פילוח שעתי של ממוצע ביום – אפשרות לקו אחד ממוצע לפי שעה

12. REFERENCES

[AABK71] ANDRIENKO, GENNADY ; ANDRIENKO, NATALIA ; BAK, PETER ; KEIM, DANIEL ; WROBEL, STEFAN: *Visual Analytics of Movement*. Bd. 1, 1871 — ISBN 978-85-232-0700-7

[Aljo14] ALJOUFIE, M.: Spatial analysis of the potential demand for public transport in the city of Jeddah, Saudi Arabia. In: *WIT Transactions on the Built Environment* Bd. 138 (2014), S. 113–123 — ISBN 9781845647780

[Carp08] CARPENDALE, SHEELAGH: Evaluating information visualizations. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* Bd. 4950 LNCS (2008), S. 19–45 — ISBN 354070955X

[ClYo14] CLAUDIO, PIO ; YOON, SUNG EUI: Metro transit-centric visualization for city tour planning. In: *Computer Graphics Forum* Bd. 33 (2014), Nr. 3, S. 271–280 — ISBN 1467-8659

[ElST07] ELMQVIST, NIKLAS ; STASKO, JOHN ; TSIGAS, PHILIPPAS: DataMeadow: A visual canvas for analysis of large-scale multivariate data. In: *VAST IEEE Symposium on Visual Analytics Science and Technology 2007, Proceedings* (2007), S. 187–194 — ISBN 9781424416592

[ElYi15] ELMQVIST, NIKLAS ; YI, JI SOO: Patterns for visualization evaluation. In: *Information Visualization* Bd. 14 (2015), Nr. 3, S. 250–269

[IsIA16] ISRAEL, PUBLIC TRANSPORTATION DEPARTMENT ; ISRAEL, PUBLIC TRANSPORTATION AUTHORITY ; ASHAK, ROBERT: *Instructions for Planning and Operating The buses in Israel*, 2016

[LBIP12] LAM, HEIDI ; BERTINI, ENRICO ; ISENBERG, PETRA ; PLAISANT, CATHERINE ; LAM, HEIDI ; BERTINI, ENRICO ; ISENBERG, PETRA ; PLAISANT, CATHERINE ; U. A.: Empirical Studies in Information Visualization : Seven Scenarios To cite this version : Empirical Studies in Information Visualization : Seven Scenarios. In: *IEEE Transactions on Visualization and Computer Graphics* Bd. 18 (2012), Nr. 9, S. 1520–1536

[Mack86] MACKINLAY, JOCK: Automating the design of graphical presentations of relational information. In: *ACM Transactions on Graphics* Bd. 5 (1986), Nr. 2, S. 110–141 — ISBN 1558605339

[Munz09] MUNZNER, TAMARA: A Nested Model for Visualization Design and Validation (2009)

[Munz14] MUNZNER, TAMARA: *Visualization Analysis and Design*, 2014 — ISBN 9781466508934

[SMDS14] SADANA, RAMIK ; MAJOR, TIMOTHY ; DOVE, ALISTAIR ; STASKO, JOHN: OnSet: A visualization technique for large-scale binary set data. In: *IEEE Transactions on Visualization* *and Computer Graphics* Bd. 20 (2014), Nr. 12, S. 1993–2002 — ISBN 10.1109/TVCG.2009.122

[Syrj16] SYRJÄRINNE, PAULA: *Urban Traffic Analysis with Bus Location Data*, 2016 — ISBN 9789520300685

[ToMö05] TORY, MELANIE ; MÖLLER, TORSTEN: Evaluating Visualizations : Do Expert Reviews Work ? Graphics, Usability and Visualization Lab When formal user studies fail. In: *IEEE computer graphics and applications* Bd. 25 (2005), Nr. 5, S. 8–11

[Tuft08] TUFTE, EDWARD R: The visual display of quantitative Information. In: *The Visual Display of Quantitative Information* (2008) — ISBN 9780961392147

[Walk12] WALKER, JARRETT: *Human Transit*, 2012 — ISBN 9781597269711

[WuGS04] WU, JIANJUN ; GAO, ZIYOU ; SUN, HUIJUN: Urban transit system as a scale--free network. In: *Modern Physics Letters B* Bd. 18 (2004), S. 1043–1049

[ZFAE14] ZENG, WEI ; FU, CHI WING ; ARISONA, STEFAN MULLER ; ERATH, ALEXANDER ; QU, HUAMIN: Visualizing mobility of public transportation system. In: *IEEE Transactions on Visualization and Computer Graphics* Bd. 20 (2014), Nr. 12, S. 1833–1842 — ISBN 1077-2626 VO - PP

[ZhWS17] ZHUGE, XU ; WANG, HAO ; STRAZDINS, GIRTS: Evaluating the data visualization for demanding marine operations. In: *Techno-Ocean 2016: Return to the Oceans* (2017), S. 471–477 — ISBN 9781509024452

תקציר

מערכת התחבורה הציבורית משרתת תפקיד עיקרי בחיים המודרניים. היא מאפשרת את התנועה של כמויות אנשים ממקום אחד לשני באופן מהיר, זול ופשוט. החשיבות עולה ככל שישנו גידול באוכלוסייה, מעבר לערים הגדולות וכן ההשלכות של השימוש ברכב הפרטי (פקקים, זיהום אוויר, מחירי הדלק וכו').

תחום התחבורה הציבורית מורכב מתכנון ובניית תשתיות חדשות, רכישת רכבים ומערכות בהתאם לצרכים, תכנון לוחות זמנים ותכולות קווי התחבורה וניטור ומעקב אחר המערכות הקיימות.

מטרת התזה הינה לספק ניטור של מערכת התחבורה הציבורית ולסייע בתכנון לוחות הזמנים ע"י הבנת תמונת המצב בדרכים (איחורים, כמות הנוסעים בין תחנות וכו'). לטובת מטרה זו, אנו מציגים ויזואליזציה חדשה שנקראת מפות סגמנטים מוצללות לטובת ניטור פעילות קווי האוטובוסים. הויזואליזציה מבוססת על מקורות ביג דאטה, הכוללים נתונים סטאטיים אודות הקווים (רשימת תחנות, רשימת קווים, לוחות זמנים וכו') ונתוני זמן אמת (תחנות בהן פתח האוטובוס את הדלתות, מספר הנוסעים שעלו וירדו, זמני יציאה והגעה וכו'). הויזואליזציה משווה בין התכנון לנתונים בפועל בהתבסס על זמנים, בשונה מהשוואה ויזואלית מבוססת מיקום גיאוגרפי.

הויזואליזציה המוצעת מספקת שיפור ניכר באופן הניתוח של כמויות גדולות של מידע לא מעובד, מסייעת בזיהוי איחורים, תבניות והתנהגות של קווי אוטובוס ספציפיים.

בעייני רוחנו, הויזואליזציה משמשת כמערכת תומכת החלטה למומחים וחוקרי תחבורה ציבורית. תוצאות הערכת המומחים מספקות בסיס לחזון זה.

המערכת נבנתה על מידע ממטרופולין באר שבע, ישראל, שכעת מחזיק את המידע המהימן ביותר אודות דיווחי תחבורה ציבורית מכלל מדינת ישראל. על אף זאת, הויזואליזציה תוכננה באמצעות אלגוריתמים שיכולים לתת מענה גם באזורים אחרים במידה והמידע בהם זמין.

התרומה של עבודה זו היא: (א) דרך ויזואלית ייחודית לזיהוי איחורים בתחבורה הציבורית בהתבסס על השוואה בין המתוכנן לביצוע בפועל (ב) ניטור פיצ'רים קיימים במידע כמו מספר נוסעים, משך פתיחת וסגירת דלתות וכו' ופיצ'רים ייחודים שניתן להסיקם מהמידע כמו מהירות ממוצעת של האוטובוס (ג) הצגה יומית/שבועית/חודשית של המידע לזיהוי תבניות שגרתיות.

מילות מפתח – ויזואליזציה, תחבורה ציבורית, נוסעים, פתיחה וסגירת דלתות, SSM, מפות סגמנטים מוצללות, אוטובוסים

אוניברסיטת בן גוריון בנגב

הפקולטה למדעי ההנדסה

המחלקה להנדסת תוכנה ומערכות מידע

המחלקה להנדסת תעשייה וניהול

מפות סגמנטים מוצללות וזיהוי תבניות של קווי האוטובוס בבאר שבע

חיבור זה מהווה חלק מהדרישות לקבלת תואר מוסמך M.Sc.

מאת: ליעד נחמיאס

מנחים:

פרופסור נעם טרקטינסקי, המחלקה להנדסת תוכנה ומערכות מידע

ד"ר פיטר בק, המחלקה להנדסת תעשייה וניהול

חתימת המחבר:... תאריך:......................14.10.2019.......................

חתימת המנחה:... Noam Tractinsky תאריך:......................14.10.2019.......................

חתימת המנחה:... תאריך:......................14.10.2019.......................

חתימת יו"ר ועדת הוראה:... תאריך:........................14.10.2019.......................

14 באוקטובר 2019

אוניברסיטת בן גוריון בנגב

הפקולטה למדעי ההנדסה

המחלקה להנדסת תוכנה ומערכות מידע

המחלקה להנדת תעשייה וניהול

מפות סגמנטים מוצללות וזיהוי תבניות של קווי האוטובוס בבאר שבע

חיבור זה מהווה חלק מהדרישות לקבלת תואר מוסמך M.Sc.

מאת: ליעד נחמיאס

מנחים:

פרופסור נעם טרקטינסקי, המחלקה להנדסת תוכנה ומערכות מידע

ד"ר פיטר בק, המחלקה להנדסת תעשייה וניהול

כתובת אימייל של המחבר: liadn@post.bgu.ac.il

14 באוקטובר 2019