

Information Visualization

Arnon Kahani – 305509556

Overview

The inspiration and the data used for the visualization project presented in this report is based on the engineering project, FastLanes¹, which is a part of the Software Engineering final project.

The FastLanes project will provide a framework for querying, processing and visualization public transportation data

This visualization project uses data from the FastLane framework i.e. acquired and saved on file. However, the project doesn't interact directly with the FastLane framework.

What

Domain

The domain of the data used in the following visualization project, is **public transportation data** and more precisely static transportation data i.e. planned trips and stops and not live actual data. The data is defined by the **General Transit Feed Specification** (GTFS) and in the visualization presented in the project only a subset of the data is used.

Data description

This section will describe the raw data and the derived data for the visualization, the GTFS data includes a large amount of information what will be described in this section is the relevant data for this visualization.

Raw data

The raw data used in this project is obtained from the FastLanes project, the data goes through a pipeline from the database to the visualization, the data presented in this section describes the data in the database as it kept in the FastLanes system.

Only the main attributes will be described.

- Calendar - a table describing for each trip what days in the week it operates e.g. For trip id 512456 the days of operation are Sunday, Wednesday, Friday.
- Stops – a table describing the bus stops including geographical information.
- Stop-times – a table describing for each trip what are the stop times for that trip in a specific stop.

The main attributes

- Geographical coordinates of a stop
- Days of the week a trip is being operated - Day
- Stop times (hour : minutes : seconds) of a trip in a specific stop

Derived data

The data derived for this visualization is as follows, for each stop the amount of buses that stop at it per hour for each day (int[7][24]) and the stop's coordinates.

¹ The FastLanes project team members are Einav Ram,Ziv Levy and Arnon Kahani.

JSON representation

```
        "stop_id": "376903",
        "stop_name": "פפ'ירו/ה מזק ביבר לת תיכרומ ווותה",
        "geom": {
            "type": "Point",
            "coordinates": [
                34.79648,
                32.856487
            ]
        },
        "rides": [
            {
                6,
                0,
                0,
                0,
                0,
                11
            }
        ]
    }
}
```

Why

I was unsure how to phrase the **why** of this project I present two options.

- 1) Explore the density of public buses as function of buses per stop at specific times in context of geography and time.
 - 2) Explore the urban landscape as function the geographical density of public buses in specific times in context of geography and time.

The user task is to explore the density of public buses and analyze urban environments.

The goal of this visualization is to provide a view of density of public transportation in a geographical-time context to allow the exploration of the urban environment i.e. which areas are for working, leaving, entertainment, congestion bottlenecks etc.

This visualization has direct connection to the visualizations made by Einav Ram and Ziv Levy (FastLane team), it is the most general and high level visualization of the public transportation density it allows a complete overview map of the density and there for allows the exploration of large geographical and over time.

How

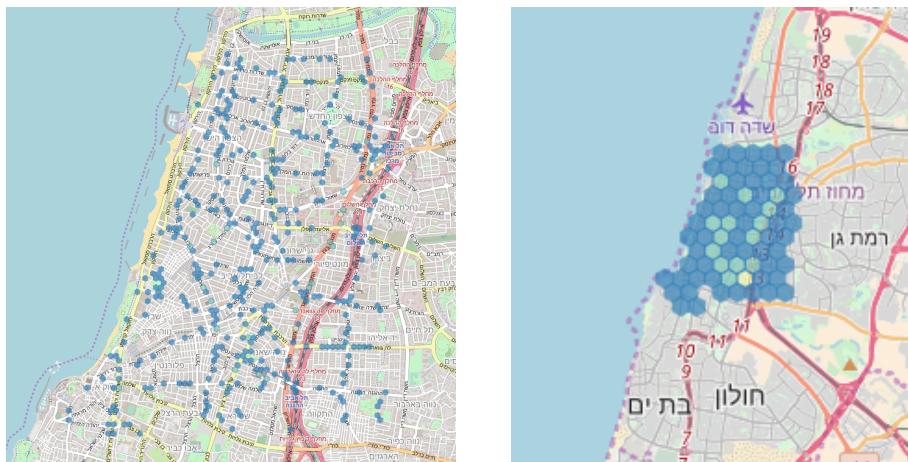
Visual Encoding

1. Hexagon – confines a geographical area and set a bin for the buses in that area.
 2. Color – the relative amount of buses per bin for each view (zoom) the scale is readjusted to match the maximum bin for that view (red high amount of buses and light yellow low amount)
 3. Tooltip – the amount of buses in a hexagon bin

The design process

Stage	Pros	Cons
1	Can infer specific stops density	Only 3 colors hues, hard to understand detailed information. When zooming in the radius is very small and hard to infer and density information. No change of range between zooms
2	Can infer specific stops density and large areas	Color pallet doesn't show the full range so well. Hard to determine what the color means i.e. what is the amount of busses
3	Easier to understand the density (color) and see specify values	No temporal information
4	Able to explore density change of time (specific day)	No user control

Stage 1



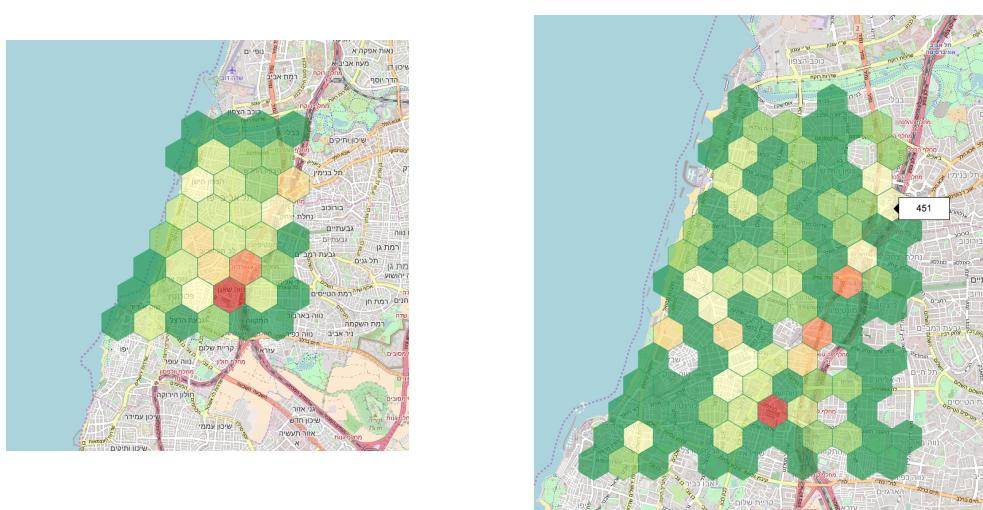
Stage 2

More color class, higher radius and dynamic domain between zooms

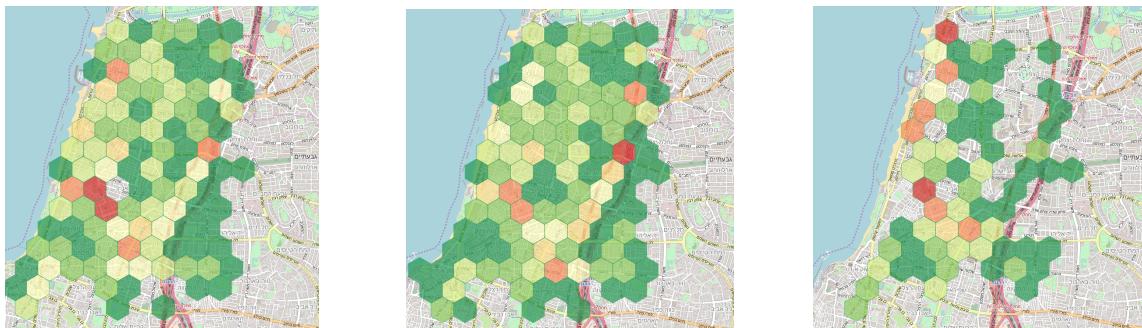


Stage 3

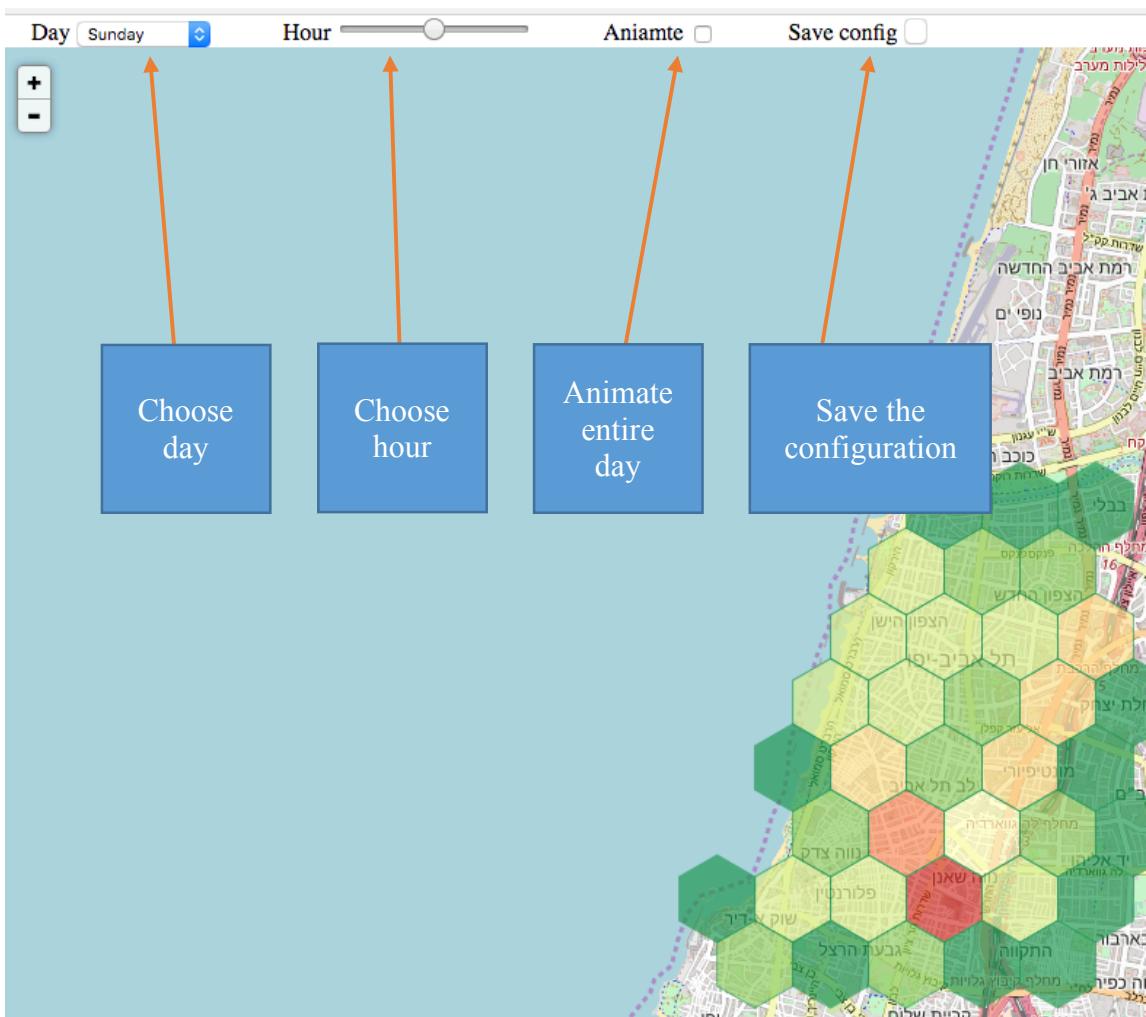
Change color pallet and added tool tip with amount of buses



Stage 4
Temporal change (animation)



Stage 5
User control



General Discussion

The visualization choice

- The visualization chosen in this project is a dynamic (temporal) hexbin approach, there are few reasons for doing so:
 - The hexbin allows to show general themes and in this case density over map, it is not as specific as dot density although if the map is zoomed allow the way it is possible to view specific data points (in this case).
 - It is relatively easy (compared to a continues heatmap) to understand the bounds of the hexbin and density associated with that area.
 - Although it is not as granular as a dot density map it allows using a tool tip to view the information encoded in the bin and understand the color meaning.
- The temporal changes (animation) allows the user to explore that changes over time and understand patterns in the data.
- The color pallet emphasizes the min and max values and changing the domain at each zoom allows the user to analyze different areas with specific context

How does the visualization help the user

There a few reason why the visualization presented in this project my help the users in accomplishing their task.

- 1) The visualization provides the user a view of the density of the public transportation in few granularities and thus providing a view of urban environment in a city cluster view, city wide view and street level view.
- 2) The temporal aspect of the visualization provides the users the ability to explore and learn patterns in density.
- 3) The visualization allows checking specific values and inferring information on areas and not only get a “feel” of the density but rather numbers.
- 4) The user controls allow the users an interactive experience and thus allowing true exploration.
- 5) The colors allow the users to infer the density without seeing the numbers and the data itself.
- 6) The visualization enables the users to view large amount of the data about the density in concentrated matter (e.g. that data in the sample provided is of 600~ stops and around 17K trips).
- 7) The visualization geographical view of the density and thus understand topological attributes.

Value of the visualization

The majority pros of this visualization were presented in the previous section, the value of this visualization is the ability to compress large amount of data in a geographical context with temporal view of the data thus exploring and learning patterns, furthermore the ability to interact with the visualization allows users the freedom to explore as they would like.

How to run the code

The project is available at and therefor there is no need to run it locally if it is need the provided code should be run on localhost server, hosting the files in the **viz** folder the file to open in the browser is the index.html file.

To use the controls, choose the configuration and then click on the save configuration button.

The animation toggle is a checkbox when it is on and after the save configuration button is pressed the animation will begin, to stop it toggle the checkbox off and press the save configuration again.

The animation is for the specific day chosen and it loops the hours of the day.