Exploratory Information Visualization Project
OSM changes visualizer

November 2022

1 Introduction

For this exploratory visualization analysis project, I decided to create an exploratory feel data interactive visualization to portray the change in road infrastructure and amenities in a selected number of developing countries. The topic is inspired by the 9th UN sustainable development goal "Industry, innovation and infrastructure". I will use OpenStreetMaps historical files to allow the user to observe how the road network and amenities have changed in the past 8 years in a developing area of their choosing. The aim is to portray the connection between road infrastructure and amenities with development. From observing the changes in the road network and infrastructure through time for an area of their choosing, the user can get an idea of the development that has occurred in the area during the last 8 years. The amenities shown (hospitals, restaurants and schools) were specifically chosen because, along with other infrastructure, they can be used as indicators for the development of an area. The user can observe whether more restaurants have appeared in the area in the last decade, which is generally an indication of population growth and possibly development. Same reasoning applies to an increased number of schools and hospitals in the area. On the other hand, if there is a noticeable decrease in these types of amenities through time, it could mean that the area is being depopulated. For example, the user can observe both how road networks have expanded in the last decade in the main cities and infer about changes in amenities in suburban areas as people depopulate towards bigger more habitable cities. The visualization can then reveal 1) changes in roads and infrastructures for a given area 2) changes in restaurants, hospitals and schools in the given area. These can be used to determine development and population changes. These observations on the changes occurring in the area can be used to visually concretize the concept of development, which is often expressed only in terms of summary figures such as Gini index, GDP growth, literacy rate. On the hand, this project aims to portray a concrete visual representation of what development looks like in a specific area. A video for the visualization can be found in this Github repository: https://github.com/emimarch/OSM-Changes-Visualizer/tree/master/videos .

2 Application and user interaction

In the picture in Figure 1 you may observe a screenshot of the visualization for an area in Davao City, Philippines. The visualization is deployed as an application, with three main panels. In the first panel the user can select a country. After the country selection, the user will have to insert a pair of coordinates in the dedicated input bars. The coordinates need to belong to the country they have chosen, and they represent the centroid of the area that will be displayed in the map. After selecting a country and the coordinates, the application will remind the user of the country and coordinates selected and start processing the network map. After less than a minute, a graph with time line bar will appear in the bottom panel. The map will be centered at the coordinate pairs provided by the user and will cover approximately a 2 kilometer radius from the centroid coordinates. The timeline will have a play button and stop button. By pressing the play button, the visualization of the road network from year 2014 to year 2022 will begin: each snapshot represents the state of the network in that year, with black lines representing the OSM network (including buildings) and yellow, red, and blue dots representing restaurants, schools and hospitals respectively. The user may stop the visualization at any point as well as drag the indicator on the timeline to manually shift between one year to another. They may also hover above a amenity of their liking, which will trigger a pop-up with the name of the amenity to appear. They can also zoom in the graph and run the timeline visualization for the zoomed area. Next to the map panel there is a legend for the amenities markers and the network lines. Each year showcases the state of the network on the first of January of that year.

Initially, the aim was to provide the user with the choice among all 27 developing countries. Since it is a lot of data to store, it was instead decided to provide a smaller selection of countries: Bolivia, Philippines, and Ghana. The code can be adapted for other countries as well if provided the necessary OSM files.

3 Data

As briefly hinted above, the visualization relies solely on historical OpenStreetMaps (OSMs). OSM is a crowd-sourced project to create a free editable geographic database of the world. It is an open source freely accessible extensive map containing roads, buildings and geographical features. OSM maps may be accessed as .osm files, which contain a list of nodes, ways and relations in the map. Historical OSM, on the other hand, are a merge of the past and current versions of the OSM files. The historical maps contain all the changes made in osm file throughout time, and are often found in osh.osm format. By selecting a specific point in time from an historical map, we can observe the state of the map at that time.

In this project, I used the OSM historical files provided by Geofabrik ([3]). While these files are freely accessible, they are not anonymous and contain

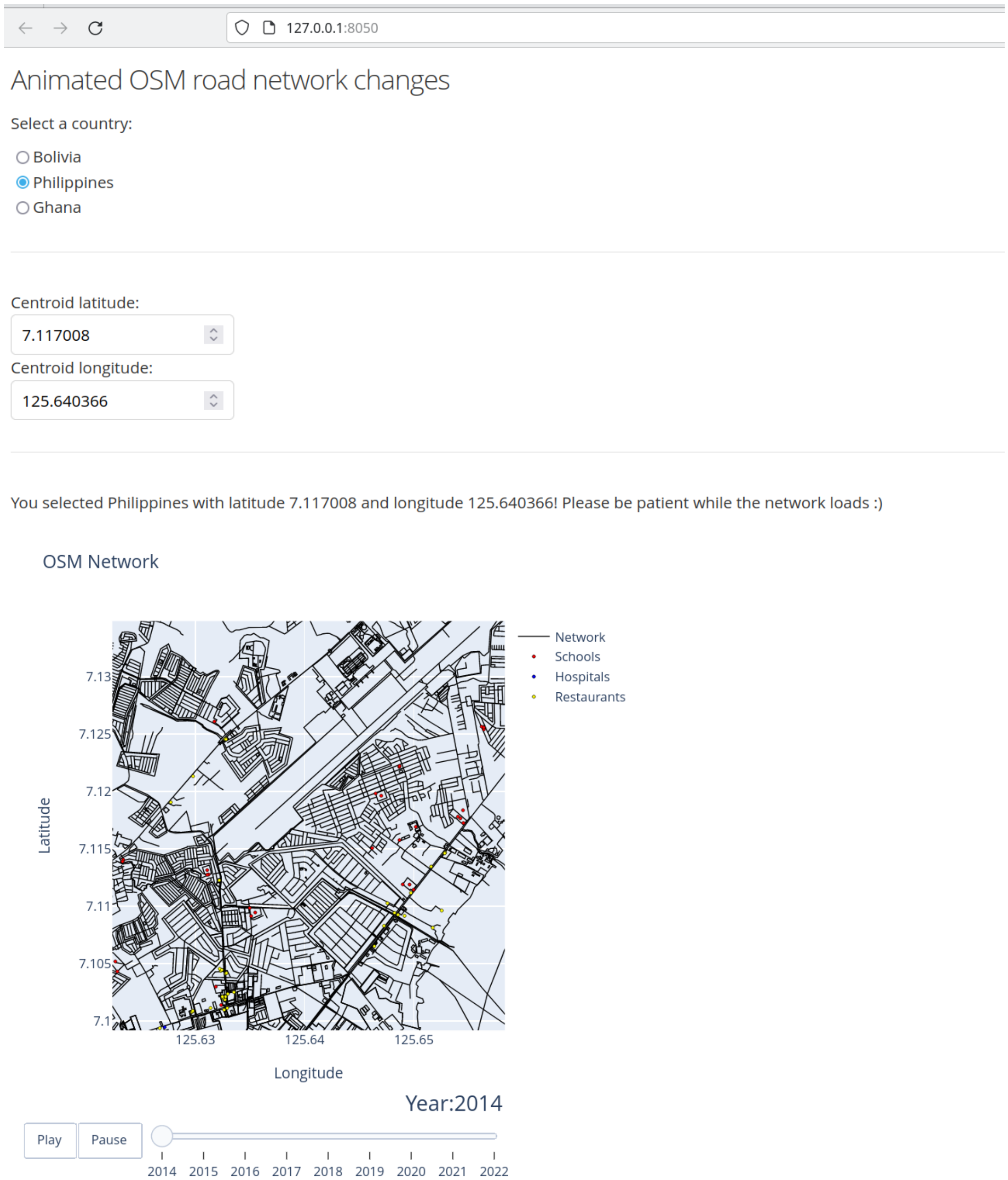

Figure 1: Screenshot of the application for area in Davao

user information. For that reason, it is needed to first create a OSM Geofabrik account to be able to download the historical OSM files. For instance, the historical file for Bolivia can be accessed as `Bolivia-internal.osh.pbf` from Geofabrik as shown in this webpage https://download.geofabrik.de/south-america/bolivia.html with a valid osm account. Creating an account is fairly straightforward and free.

3.1 Data shortcomings

The state of OSM maps today is quite complete, but it does rely on the inputs from users who willingly update the network as it changes through time. Desolated areas might not have been covered as much, and this is a possible issue in my visualization for several reasons. To minimize the effect of having incomplete maps, I selected as a starting year the year 2014, which is about 10 years after the founding of the OSM platform. Nevertheless, some areas, especially in developing countries, have not been as extensively explored and updated by the users. This might result in a perception of false changes throughout time, where for instance the appearance of a new road in the network in the year 2019 compared to the year 2018 might not necessarily mean that the road had been built that year (effectively being a sign of network expansion) but that the road had already been existing before and was only added in the OSM database by an user in 2019 for the first time. This is a shortcoming to keep in mind, and unfortunately impossible to get rid of because OSM does not detain construction year data for most items. This visualization will be most reliable for developed countries and large cities, which are likely the ones that are upkept the most by the osm community.

4 Tools

The osm historical files where downloaded as country wise files, but the visualization only displays the two kilometer radius from the user input coordinates, which can be zoomed in as the user wishes. The reason for not displaying the entire network for a country, if not the globe, is that the OSM files for bigger areas are significantly larger and, since everything is done online and the networks are not prepared beforehand, creating a network update visualization would have taken the application an unreasonable amount of time. Therefore, the sensible approach was to delimit the browsable area to a 2 kilometer radius. In order to do so, it was necessary to use a tool to extract the area of interest from the main historical file and save the state of the network for each year from the historical area extract. Osmium-tool was used to achieve both. Osmium-tool (documentation here https://osmcode.org/osmium-tool/) is a command line tool created to handle both osm and historical osm files. The functions of interest that were used for this project are `extract` to create the historical area extract from a bounding rectangle centered at the user coordinates, `time-filter` to create maps for each year from the extracted areas and `export` to convert my maps into a easily readable format (geojson in this case).

After creating a map for each year for the area of interest selected by the user, the program opens them and merges them as a singular dataframe in python. The python library Plotly express [2] is then used to create the visualization. This is a JS based library that allows to create animated plots. More about Plotly for python here: https://plotly.com/. A lot of data preprocessing is needed before the osm network extracts can be used with plotly, mostly necessary to achieve a faster creation of the plot so to limit the waiting time. The plot is rendered in an app with the three panels described in the previous section using the Dash python library. This is a library for creating web applications with Plotly graphs.

5 Time and space

Time is a fundamental aspect of this exploratory visualization. The aim is to portray changes in road networks and amenities through time to display the shift in infrastructure in developing areas. The article "Timelines Revisited: A Design Space and Considerations for Expressive Storytelling" [1] defines the role of time in visualization based on three variables: its representation, its scale and its layout. For the first, representation, it states that time can be represented as a linear, radial, spiral, grid, or arbitrary. This visualization represents time as linear. This can also be observed by the screenshot in Figure 1, where there is a linear time bar for representing time. I have considered also other representations of time, but this is the one that is the most understandable for the scope of my visualization. I believe that other representations are more sensible when time is the main topic of the the visualization (for instance, the year clock that was displayed in class) but while time is fundamental in this project, it is this not the main character. For the second feature, scale, time can have a chronological, relative, logarithmic, sequential, sequential + interim duration. The visualization will has a chronological scale of one year interval, starting from the year 2014 to the year 2022. A relative time scale could have also been used, where the starting point would have referenced year 2014 and the last point would have referenced year 2022, but this would have not served a good purpose in the visualization, as it is important to refer to the actual year because development factors differ depending on the decade. For instance, 8 years of road changes in the time of the roman empire would look completely different from 8 years of road changes in our modern world. Even if we would compare 8 years of road changes in the 90s with my selected time frame (2014-2022), we would be looking at two completely different contexts of innovation and technology.

The last feature is layout, which can be unified, faceted, segmented, or faceted + segmented. The visualization has a unified timeline, that is a single individual time line that is chronological from the year 2014 to the year 2022. The reason for this is that whilst the aim is to comparing road networks through time, the data is accessible for one specific range of years. In theory, some OSM information can be found from even earlier than 2014, but it was decided that this data would probably too unreliable due to the limitations reported in section 3.1. If reliable older historical data would have been accessible, e.g. for the roman empire, it would have indeed been very nice to have a faceted layout, one portraying the changes in time in the modern age years, and one portraying the changes for the same time span (8 years) in the roman empire.

While time is very important in my visualization, space detains an even greater role. My visualization is mainly based on spatial information and, as such, it comes with all the challenges that are associated with having to handle GIS data. The network is the central character of the project: by observing the changes in the road network and buildings the user is supposed to make inference about the connection between development and infrastructure. A light blue almost white background was chosen on purpose to enhance the dark network that is superposed on it. The network lines are kept very fine to retain definition both when zooming in or out, which is fundamental to observe any changes in very dense networks. The highlighted amenities (restaurants, hospitals, and school) are shown in small colored black-lined markers to pop against the black lines of the network. The two kilometer radius of the area to display was chosen after experimenting with greater or smaller radii. Greater radii were soon dismissed because of several computing constraints (a lot of processing time) but mostly because the definition of the network would have been lost when at a greater scale. Although the user would have been able to zoom in, the purpose of the visualization was not to provide an aggregate view of very large area, because most details would have been lost. Smaller radii have also been considered, but they weren't able to achieve a good enough view for city level changes. The two kilometer radius was a compromise between many factors, both computational and visual, and was deemed reasonable enough to represent changes in small cities or large neighborhoods.

6 Previous ideas

Before I came up with the topic on infrastructure, I was set to do an exploratory feel-data visualization about the connection between deforestation and natural disasters. The format and the visualization would have been in a sense very similar to my current project, with a browsable map displaying deforestation intensity areas and natural disaster zones through time. Unfurtuntaly, gaining access to deforestation GIS data ended up being harder than expected, and storing and converting such data to a more workable format also resulted challenging. Nevertheless, a lot of the research that I did in preparation of this previous topic ended up being useful for my infrastructure visualization. For instance, the usage of the Plotly library for animated visualizations.

References

[1] Brehmer, M., Lee, B., Bach, B., Riche, N. H., and Munzner, T. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE transactions on visualization and computer graphics 23*, 9 (2016), 2151–2164.

[2] Inc., P. T. Collaborative data science. `https://plot.ly`, 2015.

[3] Karlsruhe, G. Geofabrik. `http://www.geofabrik.de/`.