EPA1351. Advanced Discrete Simulation
Assignment 3. Data Visualization

Introduction

In this exercise we want you to visualize a larger dataset in the best possible way, avoiding the visualization issues that were presented in class. In this particular case, we want to focus on an analysis and presentation of the road / bridge data in terms of vulnerability and criticality of the road system in Bangladesh. It is important to understand how busy it is on different roads, and which part of the roads are most heavily used for goods transportation, i.e. which roads or part of roads have a high percentage of truck traffic (relating to criticality). Of course we are interested to present this data in relation to the properties of the road itself, i.e. the location and quality of bridges and the number of lanes on each stretch of road (relating to vulnerability).

The RMMS dataset has a file called 'traffic.htm' for each road, for different types of traffic and for different chainages. This file contains a table with traffic data for Heavy Truck, Medium Truck, Small Truck, Large Bus, Medium Bus, Micro Bus, Utility vehicle, Car, Auto Rickshaw, Motor Cycle, Bi-Cycle, Cycle Rickshaw, and Cart. Some data are actually measured, others are estimated. The traffic is measured as AADT, which is the total volume of vehicle traffic in both directions on a section of road for a year divided by 365 days. Actual traffic in an hour can be higher – e.g., weekdays versus weekends, or day versus night. Still, the AADT gives a very good indication about the importance of a road, and about the number of truck passages per day or per year, which can be translated to the number of tonnes transported via that road per year.

Furthermore, we have of course the location of bridges on the road and their quality, which can give insight into the vulnerability of that particular stretch of road in case the bridge would be damaged. You can use your own cleaned dataset to get insight into the bridge locations, or alternatively the cleaned dataset that was provided as input to Assignment 2.

Also, the number of lanes on the road plays an important role to get insight into the importance of a road, and into the capacity of the road for cargo transportation. For this purpose, the RMMS dataset has a file called ROAD.widths.processed.txt, which is a tab-delimited text file with information about the number of lanes on each stretch of road. Dividing the AADT by the number of lanes provides insight into how busy each lane is, and where we can expect most congestion.

Assignments

1. Study literature and define **'vulnerability'** and **'criticality'** of a road, where vulnerability relates to the probability that roads get impassable after natural disasters (cyclones, heavy rain, flooding, mudslides, earthquakes, etc.), and criticality relates to the amount of goods transported over the road, i.e. the economic importance of a road. Provide two definitions with corresponding literature references in your report, and use an operationalization of your definition in the subsequent assignments.

2. Analyze the **full N1 road** for traffic density per mode of transport, and relate that visually to the criticality and vulnerability. Build your visualization in such a way that it provides good insight into the vulnerabilities / criticality relation of road segments and of bridges with different quality.

As a **bonus** exercise (max 1 point extra for your mark), generalize the approach from the first assignment to provide visual insight into any road from the road dataset. It should be possible to run your analysis for a given road. Make it as easy as possible for the user to do so.

3. Compare all roads with respect to criticality / vulnerability and provide (visually supported) insight into the **top-10** of most vulnerable as well as most critical roads and/or road segments for cargo transport.

We expect that you will work with Python, Excel, R, and possibly other programs to do the visualizations. It is not necessary to create visualizations using a Web browser, as this might take too much time. Jupyter / iPython notebook, and R scripts are typically the results we would expect.

Note 1: Exercise 3 is really different from exercise 2: for the #2 we ask you to provide **full** visual insight into **any** road from the total dataset of N, R, and Z roads in an interactive manner, whereas in #3, you are tasked to find the **most 'important'** road segments across all roads (note that you have to define that 'importance'), without the need to have a user interface to choose a road. Therefore, you can work on both exercises in parallel.

Note 2: For the visualizations think carefully whether you want to use a geographical or schematic view of the criticality / vulnerability of a road.

Note 3: Provide sufficient instructions so that the teachers and TAs are able to run your programs and replicate your results.

Hand in a ZIP-file with your **definitions and literature** references, the analysis **programs and scripts**, the **generated visualizations**, and a **short report** of your **main findings** about the criticality and vulnerability of these roads, and how you visualizations support the communication of these findings best. The report should have a length of about 5-10 pages (including appendices), but it is no problem if the report is longer. Upload the Zip-file to Brightspace with a clear name (preferably, "**GroupXX_Lab3.zip**"), so our TAs can find it.

Time to spend and Deadline

Like the previous exercises, we will use 2 weeks to complete it. So, there are 8 lab hours dedicated to complete the lab assignment. In addition, you are each expected to spend another **8 hours maximally** per person on carrying out the exercise. Don't overspend your hours, and see how far you can get with the analysis and visualization in 16 hours total. You could already get a pass mark if you define criticality and vulnerability, analyze the N1 road in detail and provide a first pass at analyzing and comparing the total road set. Of course this also depends on the quality of your visualizations. Divide the work well within your group, and make sure you use the available hours of all team members combined well. Because this will be a combination of using Python (to prepare the data sets), Excel (to analyze some of the data on beforehand), R (to make scripts to analyze data in more detail) and possibly statistical packages such as SPSS (to make calculations that are not available in R), we expect all team members to be able to contribute equally.

Deadline for handing in the ZIP-file of Lab session 3 on Brightspace is Monday in week 7, 09:00.