Discussion following review #1

Problems encountered in your map

You did a good job standardized street type and school types. It seems that you understand the process of auditing, and know ways to correct or standardize the data. Please consider expanding the "Problems encountered in your Map" further. You might want to include examples of other elements (postal codes, Email, phone numbers, URL, etc) that were not properly shaped or formatted and discuss ways in which they can be fixed programmatically. Were you able to find elements that are specific to your chosen area? Different language?

Proposal for further corrections

Postcode:

Postcodes are 5 digits in France and start by 31 in Toulouse area. The tag can be fixed fairly easily by making sure it follows the pattern (‘^31[0-9]{3}$’). Usual problem is that city name is sometimes included after the post code. One can only take the first 5 characters in that case.

Phone number:

Phone numbers are made of country code (+33 in France), followed by area code (5 for Toulouse, 6 or 7 for mobiles and 8 or 9 for special numbers) followed by 8 digits. The tag can be fixed by making sure it follows the pattern ('^\+33 ?[0-9]{1} ?[0-9]{2} ?[0-9]{2} ?[0-9]{2} ?[0-9]{2}$'). Usual problem is the lack of country code and none homogenous number of whitespace characters.

Other ideas about the dataset

Although you provide a lot of meaningful statistics and analysis to the data, there is limited discussion of how the analysis can be further improved. Consider ideas for the data collection process from OpenStreetMaps and other analogs that could benefit data wrangling or problem solving. Broadly discuss how the idea could be implemented and what the potential challenges might be. For instance, is there an implementation that could be made to OpenStreetMaps or your data analysis process that could improve the data? It is OK if your idea is beyond the level of the course, and you are not sure how to implement it entirely, as long as you can discuss what are potential challenges of implementing the idea.

It would be interesting that this work profit to the community. For instance the simple auditing and correction scripts for street names, city names … could be applied directly to OpenStreetMaps data. This could be done either when the map is modified by a user or scheduled on a regular basis to keep the map clean and tidy. If several users like me have the same idea, we would be able to clean all the tags in no time …

Another idea that does not deal with the wrangling but rather with the analysis part of this work is to link the results of my queries with the map from OpenStreetMaps. That way, when I query the restaurants or the schools in SQL I can directly see the results on the map and have a better idea of the geographical location of my results. I’m not sure how to implement such feature, but I’m sure longitude and latitude of the nodes can be used conveniently for that.