**Project 4**

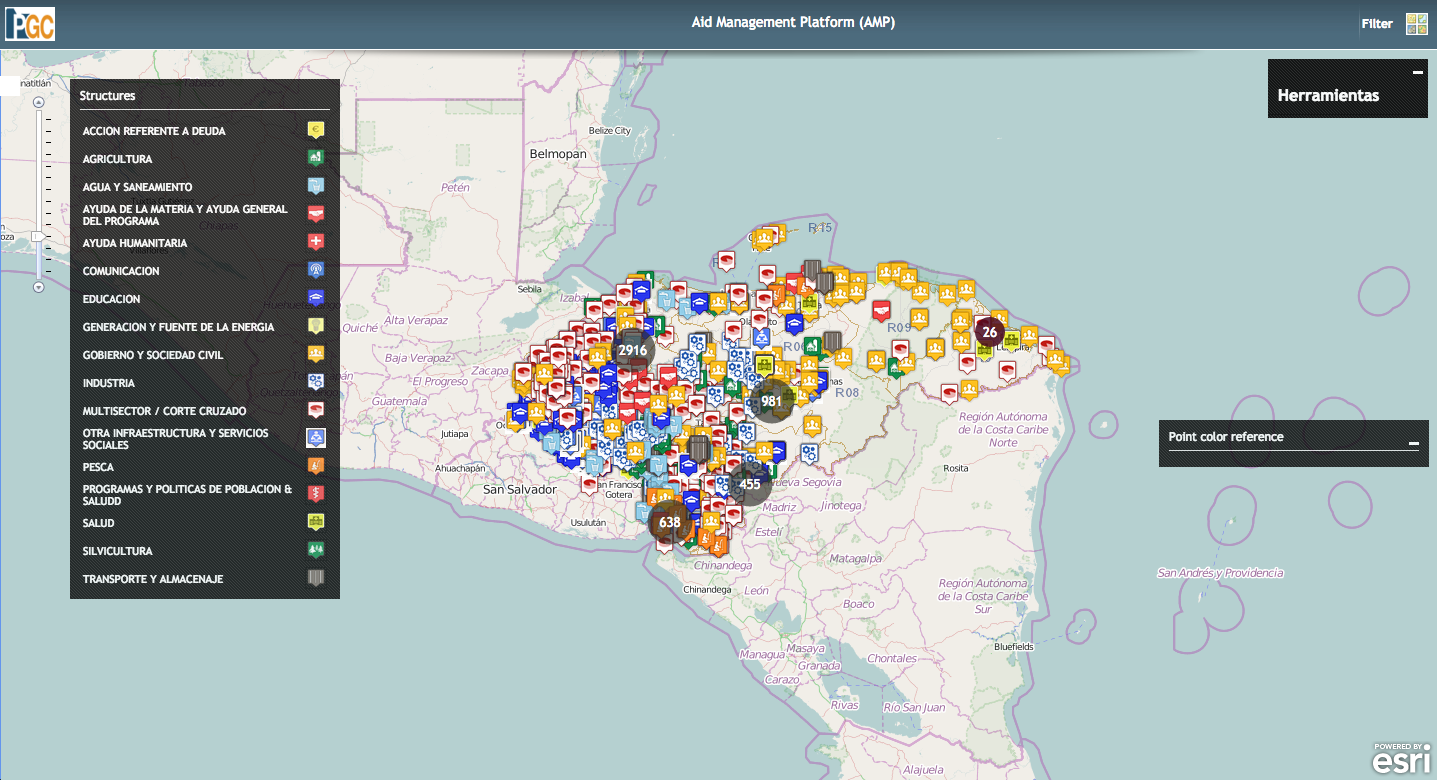
Benjamin Arancibia

*Part 1: Use Case and Obtain the Data*

The data used in this project is a brand new dataset that is only about a week old and was first published on Friday November 7th, 2014. It is a dataset of all foreign aid project locations in Honduras. Some quick stats about the dataset: 420 foreign aid projects geocoded that led to more than 5,000 project locations in Honduras. It is very rare for a foreign aid project to only have 1 location. The dataset that I am using for this project is simpler version of the full dataset. The columns in the dataset for this project are: ID, Title, Latitude, Longitude, Type (sector), and Description. Usually this type of dataset has 20+ columns more, but this is raw dataset that still needs to be cleaned and merged with the data currently present in the Honduras database.

This data is important because the Government of Honduras, Donors in Honduras, and Civil Society groups have no idea where project locations are in Honduras. Creating this geocoded data allows users to merge this project information with other geospatial information to perform better analysis. Ultimately, this better information will allow for better decisions and an improvement in planning by the Government of Honduras.

The data can be found at this website: <http://pgc.sre.gob.hn/> free for download and open to everyone.



*Part 2: Bring the data into R*

I imported the csv into R using the typical read .csv command. The only issue I had is in the .csv the IDs are viewed in scientific notation, this needs to be altered when imported into R. The reason why this needs to be changed is because many IDs start with 8717 and the projects can be merged accidently. This results in around 3000 locations for one project. The structure of the dataframe is as follows.

AMPID = Project ID in Government of Honduras’s system

Title = Title of Project in Government of Honduras’s system

Latitude = latitude of project

Longitude = longitude of project

Type = Sector of Project in Government of Honduras’s system.

Description = Description of project if any

For type (sector), there can be multiple sectors. Right now they are identified as multisector and as part of the cleanup of the dataset data transformations will remove multisector and fill in the appropriate sectors. Also, part of the cleanup of the data will be resolving encoding issues.