DFG Colombia Land-Use Project Report

Fall 2019 (3/2019 – 05/2019)
Sponsor: Ana María González Forero (FEM – Fundación por la Educación Multidimensional)
Team: Vivek Kantamani (Project Lead), Rea Radhika Rustagi, Amy Zhang, Xinyue (Sally) You, Ankita Agrawal

Objectives:
1. Data Cleaning / Exploratory Data Analysis
2. Visualization / Modeling

Hurdles:
1. Validity of data.
2. Laying out concrete objectives and endpoints.

After establishing the Colombia Land-Use Project team in late March, our initial meetings in April consisted primarily of Data Cleaning and Exploratory Data Analysis (EDA) per the OSEMN framework. It was brought to our attention via Professor Dorie that part of the collaboration with FEM centered around understanding the data Ana had, and how it could be used to meaningful ends. This led to our first hurdle – the validity of the data.

The data took four forms: (a) Sociological Data, (b) Population Data, (c) Geographic Township Data, and (d) FEM Data.

Dataset (c) was easily visualized and facilitated our understanding of the size and populations of various townships. Dataset (d) allowed us to visualize a map of Colombia townships that FEM had either (1) contacted, (2) not contacted, or (3) filed a land titling claim for. Datasets (a) and (b) were essentially random assortments of socioeconomic outcomes (malnutrition, poverty, access to potable water, unemployment, etc.) for ethnic vs nonethnic groups that Ana had amalgamated from various sources. We had many discussions about how to reliably make use of this data, and we ultimately had to scrape the primary sources Ana had drawn from to ensure that the data was reliable and sound in structure.

These datasets were essentially a miscellany of information unsystematically and unreliably organized. They touched on a variety of topics and were not linked by any particular parameters. Though we were able to produce three visualization that proved fruitful (see Results below), they ultimately stem from basic exploratory data analysis.

The datasets were consolidated from academic and government reports – they each had their own unique scope and method of data organization / consolidation. Thus, many of the broader visualizations that we hoped to elicit from this portion of the project proved unrealistic – the data could not support such conclusions.

In summary, the data that we had access to was an assortment of geographic and sociological data pulled from various government and research reports that could not be cohesively tied together to more fruitful ends. However, productive visualizations that illuminated the extent of FEM's work and influence in certain regions, as well as visualizations of the plight of ethnic communities were beneficial.

Results:
1. Three visualizations – a map of FEM activity in Colombia (produced by Amy), a visualization of sociological data (produced by Rea), and a visualization of township density (produced by Vivek).

Summer 2019 (05/2019 – 08/2019)
Sponsor: Ana María González Forero (FEM – Fundación por la Educación Multidimensional)
Team: Vivek Kantamani (Project Lead), Rea Radhika Rustagi

Objectives:	Hurdles:
1. FEM Survey Support – Statistical Consulting	1. Project Scope
2. Land Mapping (*See Future Directions)	

The visualization aspect of this project was completed in mid to late May, and the visualizations that we had produced were sent to Ana pending her input. Originally, we had discussed two objectives for the summer aspect of this project: ongoing support for FEM's ethnographic surveys, and the pursuit of a land mapping project.

As Ana had returned to Columbia in late May following the conclusion of her Obama Foundation Scholars program, we began to communicate primarily regarding FEM's survey work. Across the meetings, we pursued the objective of statistical consulting – handling errant data points, amalgamating data in a responsible way, survey design, responsible use of the data collection tool Kobotoolbox, etc. I feel that we made good progress on establishing data handling protocols for FEM over the course of the summer.

In mid-July, we began to discuss the prospect of a land mapping project. Much of the public land data to be mapped is in analog formats in Colombia government repositories. The information is occasionally a series of geographic descriptions rather than coordinates. This would have to be digitized and mapped using natural language processing and GIS expertise. Due to the robust funding, established multifaceted teams, and time required for this project to come to fruition – we drafted a proposal to facilitate collaboration among future partners.

Due in large part to our ongoing statistical consulting of FEM's data and in playing with the data we had at our disposal, both teams developed a better understanding of the others' capabilities and limitations – especially with regard to operating practices and future goals.

FEM's data collection surveys are constituted primarily by fundamental survey information of individual households – number of family members, ethnic affiliation, income, location, etc. (primarily categorical factors). Though this is a robust resource of data on ethnic populations previously unavailable – there are limitations in applying this to long-term predictive models (i.e. logistic regression for multiclass classification, linear regression, random forest etc. seemed irrelevant for application to these categorical factors).

Ana is interested in collecting data that is more useful in these predictive applications. In particular, a well-planned socioeconomic study on indigenous populations could be used to draw more meaningful conclusions about indigenous populations. This is a future direction for the project.

The land mapping project is another future direction of the project.

Results:
1. Established data handling methodology for FEM through statistical consulting. Miscellaneous support.
2. Understanding the need for more robust studies (and data) to draw meaningful conclusions from.

Future Directions
Sponsor: Ana María González Forero (FEM – Fundación por la Educación Multidimensional)

Ana is very interested in collaborating with DSI in some capacity on the following projects.

1. Colombia Public Land Mapping

There has been no concerted effort at any point in time to map Colombia's public land. Doing so would support the defense of the rights of underprivileged and indigenous communities by facilitating litigation in support of public land that has been conscripted for illegal purposes.

Much of the public land data to be mapped is in analog formats in Colombia government repositories. The information is occasionally a series of geographic descriptions rather than coordinates. This would have to be digitized and mapped using natural language processing and GIS expertise. Some of the more recent data is already digital and can be mapped more easily, however generating a temporal map (the end goal) will require concerted effort.

Ana has established contact with GSAPP (Columbia School of Architecture), the Columbia School of Journalism, and with Juan Francisco Saldarriaga Chaux, a researcher in the Center for Spatial Research at Columbia University, who are interested in collaborating on this project.

Attached please find the drafted project proposal.

2. Socioeconomic Studies of Indigenous Populations

The data that FEM currently has performed are basic surveys of households – income, number of family members, ethnic and demographic information, etc. Much of the data that they have sought out are academic or government reports on the outcomes of various ethnic groups – unemployment, income disparities, access to water, etc.

Moving forward, FEM would like to design robust socioeconomic studies of underprivileged ethnic and indigenous groups to draw conclusions about the disparities faced by these populations. Ana has established talks with the Political Science and Economics departments of the University of Los Andes and the Universidad Tecnologica de Bolivar to establish these studies. FEM's surveys have laid a groundwork for future data collection, and the universities will offer study design and implantation.

Ana was hoping to partner with DSI to facilitate these socioeconomic studies.

Future Involvement of the Team
Ana has just gotten back to us about finalizing the visualizations and tying up loose ends regarding drafting project proposals. We believe that this is a natural endpoint for Rea and my involvement. The future directions and projects require substantial expertise and are more suited to subject experts.

<div align="center">

Colombia Public Land Mapping Project Proposal

</div>

Executive Summary

In accordance with the goal of Fundación por la Educación Multidimensional (FEM) to reduce the social and economic inequality faced by marginalized ethnic groups in Colombia, FEM would like to propose a spatial and temporal mapping of titled public land in Colombia. In order to connect local communities with the technical and legal resources to protect their cultural lands and territories, it is necessary to develop a more concrete understanding of public land titling in Colombia.

Background and Problem Statement

Since 1991, recognition of collective land tenure rights in Colombia has benefited from constitutional backing. However, though the Constitution recognizes the collective rights of indigenous and afro-descendent communities in the Pacific and Caribbean regions of the country (Caribe no Insular), indigenous reservations (resguardos) remain vulnerable to developmental initiatives (e.g. mining, tourism) that facilitate large-scale violations of collective rights. In addition to natural resource extraction and conflicts with local farmers, ethnic groups face impending issues as Colombia changes its regulations on land access without effectively considering collective land rights.

Colombia is one of the few countries in Latin America without a detailed map of public lands. Many of the areas not yet mapped have traditionally been inhabited by ethnic groups, and though Colombia has officially recognized large tracts of land in forested regions, understanding the distribution and legitimacy of pending land titling requests requires further research and study.

In order to establish a productive discourse between ethnic communities and the government, it is necessary to rectify the contradictory regulations on land use and natural resources, and to determine the locations and extent of pending land claims initiated by ethnic groups. Establishing a comprehensive temporal map of Colombian public land would help to codify a framework for understanding collective land rights in Colombia and facilitate these goals.

The need to construct a territorial memory is also a key aspect of Colombian peace keeping efforts given that a key issue in the armed conflict has been the control of large portions of land, and the consequent dispossession of its legitimate owners. Public land, being legally "stable," offers a great measure to analyze land grabbing practices and the contemporary use of those lands. Creating a GIS base for historical analysis will empower historians, researchers and journalists of new perspectives to explain the armed conflict.

Proposed Work Plan

As there has been no detailed mapping of public lands in Colombia to date, the project will require collaboration from varied partners to establish a concrete timeline, funding, and action plan.

1. Establishing the Project
   a. Identify key partners to collaborate on the academic research and technical mapping of public land data.

2. Exploratory Phase
    a. Document the state of Colombian public land data and establish parameters for encoding the information into preferred formats.
    b. Identify key deliverables.
3. Acquire Funding
4. Establish GIS Team
    a. Establish and facilitate the work of a GIS team capable of mapping the data.
5. Ongoing Project Support