Data Science and Ecological Restoration: A Real-Life Example for Dummies in Either Field.

Introduction

I used to be on a track quite different from data science. I went to a very good school for Forestry and Conservation. And even in one of the best schools in the world for the topic, using computers was a touchy subject. In the handful of courses that were oriented around teaching us excel, geographic information systems, even word processing, there was always a feeling that loomed overhead, with many students giving each other glances across the computer lab: *I picked this major so I could be outside! Not stuck behind a screen!*

Even once I’d graduated, the feeling persisted for some people. My boss would frequently echo frustration about not being able to follow along with us in our field work, and colleagues would frequently struggle to use computers beyond basic office software.

It was never quite that way for me however. I loved computers, and still do, which is the main reason why I started dabbling with exploring data in Python. This blog shows the results of my first foray into data science with actual, real-world data, and the insights that it brought to the project, and hopefully can provide any other environmentalists who struggle with technology a new insight for how this tool can be used.

And if any seasoned data scientists are looking at this, I hope that instead it serves as a little introduction into ecological restoration, and drives at least one more person towards a greener future.

Ecological Restoration: What is it? How’s data science supposed to help?

The goal of any ecological restoration project is to start with heavily modified land – wasteland, old quarries or fields – and to establish a living system within it similar to what existed before human intervention. This benefits both natural populations – Habitat loss is one of the [leading drivers](https://www.nature.com/articles/536143a) of species extinction – and human populations; aside from it being generally pleasant to have forests, they also serve to clean the air and capture useful water, and many other [ecosystem services](https://en.wikipedia.org/wiki/Ecosystem_service).

The data I’m going to examine is from an ecological restoration project in Colombia, run by the environmental NGO Fundación Natura (seriously, [check them out](https://natura.org.co/)). The project is aiming to increase the coverage of [Andean Cloud Forest](https://www.natureandculture.org/ecosystems/andean-cloudforests/) by tree planting in unused agricultural land. The particular plot I will be examining features roughly 2700 trees planted in late 2019. The plantation has seen moderate success but suffers from high mortality and a limited budget. The data collected measures plant biomass and health indicators, taken roughly 3 times a year, with interruptions in the process courtesy of your local global pandemic.

First things first: This is not big data. This won’t feature machine learning. It is, however, real on-the-ground stuff that can improve outcomes and save money, but also highlights the difficulties of applying raw numbers to living organisms.

1: What are your primary metrics for success, and how are they doing? The purpose of a planting operation is primarily to establish itself by surviving and accumulating biomass.

Show survival and growth statistics. Expound a little on significance.

2: Are your actions headed towards the desired outcome? The end goal of ecological restoration is to provide a functional ecosystem, either by the introduced plants seeding themselves, interaction with local flora and fauna, or by increasing biodiversity.

Show species survival statistics, reproductive activity of plants, maybe if one plant growing with another is correlated?

(A cute problem: Guinea pigs love your plants)

3: How does the data inform your next actions: What of the previously identified trends are linked to control variables? What needs to be done now?

Show geographic data, including predation. Suggest either a course of action or a need for additional ground truthing.