**Geospatial Data and Archaeology**

As diverse as history is, it makes a lot of sense that the methods historians use vary greatly from sub-discipline to sub-discipline. Some historians will spend their days reading old letters or studying ancient art, but it some fields it’s most apt to take a birds-eye view of the situation. For scholars of South American Andean cultures, **geospatial analysis** is a vital tool in their archaeological toolbox. Today we’ll be taking a look at the stories that are told by the humble relics of *estancias*.

**Historical Background**

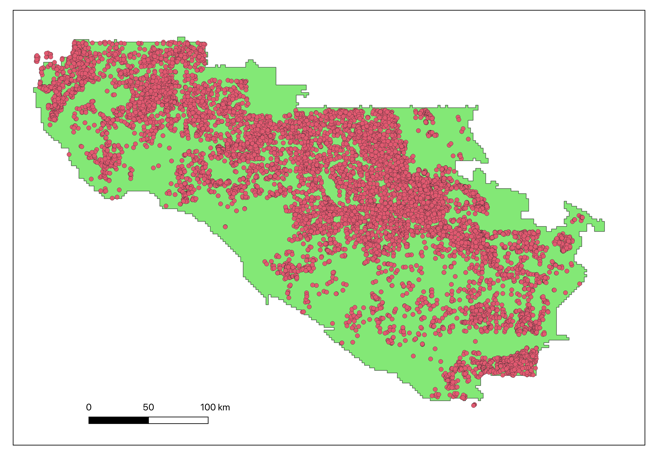
The Andes mountains, which span the west side of the South American continent, have a millennia-old history of human settlement. Most famously, the Inca empire, which existed from around 1200-1533 AD, called the Andes their home. The already-declining Inca empire was ended in the 16th century by the Spanish who conquered the region. This blog post will be focused on the following 300 years of colonial history before local independence. Spanish colonial practices such as the Indian Reductions, which forcibly moved people in the area into more “governable” townships, drastically changed the ways of life of many Andean natives. During this time period, Andean culture adopted many new characteristics introduced by Spanish colonizers, while continuing to maintain a distinctly Andean lifestyle with traditional practices needed to survive and thrive in the Andes.

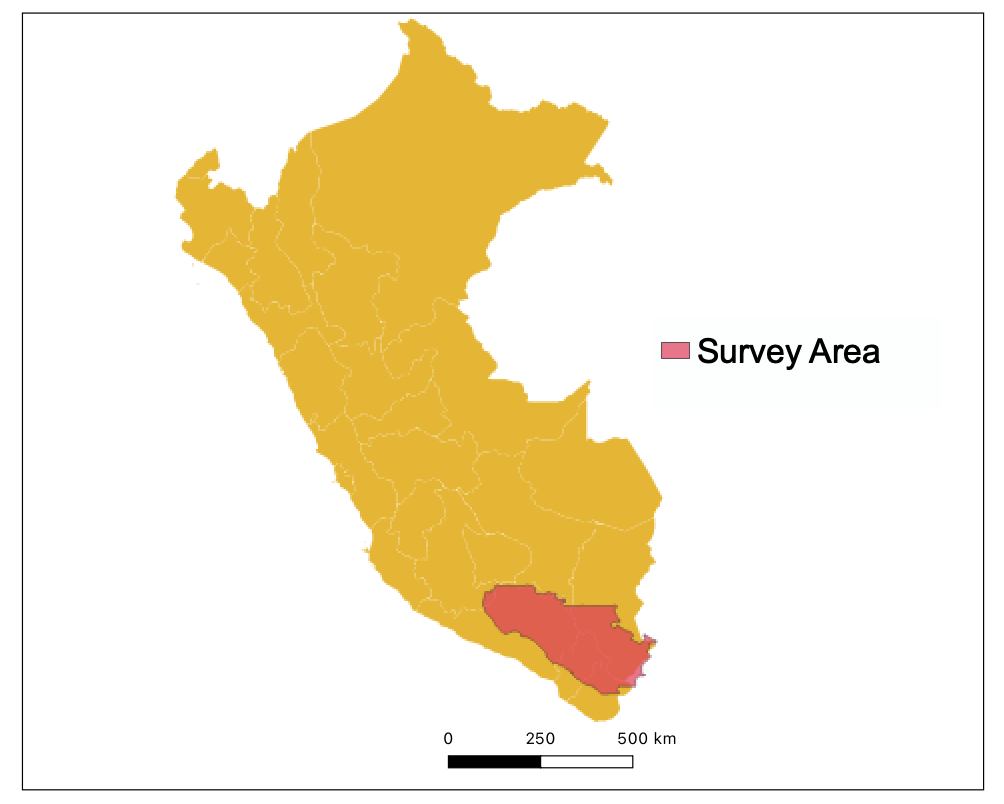
One such practice that offers a glimpse into this time period is the operation of *estancias*. While “estancia” is often used as the Spanish word for ranch, here we have a more specific meaning in mind. Estancias in the colonial Andes were highland pastoralist establishments that contrasted against the lowland (well, as low as they could get) agricultural societies. Estancias maintained various types of livestock but primarily herds of llamas and alpacas, which were raised as work animals and for fur, respectively. Historians know that these estancias played a vital role in Andean society, but compared to the agriculturalists, a lot of information is missing. One of the biggest questions in the study of Andean history is the mechanics of their decentralized economy, and many scholars suspect that estancias are the critical piece we are missing.

**The Remains of Estancias**

One of the few things we do know about estancias is where a lot of them were – because we can still see their remains. If you spend 30 minutes or so scouring over the Andean landscape in Google Earth you may be able to find something that looks like this:



This is an Estancia! All that is left of this long-defunct ranch is the footprint of a building and its associated corral, but it is surely there. We can’t know exactly when it was operational, and we’d be lucky to guess it within a century, but even lacking that knowledge there is a lot we can say if we can just get enough of them. Luckily, we have *the Geospatial Platform for Andean Culture, History, and Archaeology (GeoPACHA),* a federated campaign to document the location of structures such as estancias. A team of student workers has put in many hours poring over satellite imagery giving us a comprehensive survey of 9,500+ estancias over a region of south Peru:



Estancias within survey area

Extent of survey area within Peru

**Geospatial Data Science**

This type of data, which is attached to a physical location on the Earth’s surface, is called *geospatial data*. Effectively gaining insights from geospatial data requires a special set of tools extending the methods that we use for other types of data, like tabular or textual. *Geographic information systems (GIS)* are a type of software specially designed for geospatial analysis and have long been the one-stop-shop for specialists in this field. The two main players in the GIS space are ESRI ArcGIS and QGIS, which both offer a comprehensive suite of capabilities for managing and organizing datasets, using advanced algorithms for deep insights, and visualizing and interacting with data. Any geospatial researcher should know how to use a GIS system for quick, powerful analyses.

But there is a new trend in geospatial analysis that is stepping outside of special-purpose GIS software. More and more we are seeing researchers utilizing general purpose analysis tools to get even more hands-on with their data. In this space there is a clear leader: Python. There is a special corner of the Python package ecosystem dedicated to geospatial analysis, hosting libraries such as PyProj, XArray, GeoPandas, and Shapely. These packages give Python every capability GIS has and more. Using Python for geospatial data lets analysts develop their own tools and implement novel algorithms, while leaning on all the other capabilities that Python offers as well.

But the biggest challenge in geospatial data analysis has always been the data. While disjoint coordinate reference systems and unwieldy filetypes are annoying, the main issue is and always has been the scale. For example, one common source of geospatial data comes from remote sensing devices such as imagery satellites, but these assets contain enormous pictures of Earth that easily occupy terabytes of data. Retrieving and storing these assets has been a long-time bottleneck to geospatial data science.

For this project, I used Microsoft Planetary Computer to tackle this challenge. The Planetary Computer is a remote server mounted onto a data center containing many petabytes of geospatial data. All this data is made available through a convenient API and accessible within a Jupyter Notebook environment. While by no means did this platform eliminate all the challenges of big data, it did take care of the biggest problems, and let me dedicate more time to analysis without worrying excessively about being economical about the data I wanted to investigate.

**Methodology**

The type of analysis I used to investigate these Estancias is called a **suitability analysis.** Given the locations of these 9,500 estancias, can I find any patterns to explain their distributions? I looked at many different variables ranging from solar radiation, to amount of precipitation – but the two with the most compelling results, and the ones I want to talk about today, are vegetation and terrain.

Vegetation data was derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellites, a class of sensors designed to collect regular and high-quality climate measurements on a global scale. To proxy the amount of vegetation a region has, we can use the Normalized Difference Vegetation Index (NDVI).

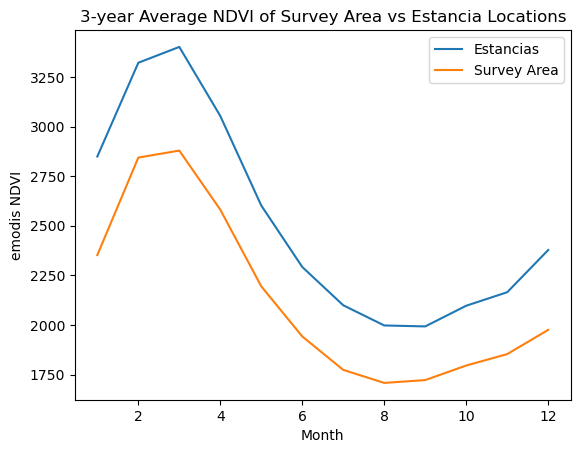
NDVI is based on the reflectance behavior of vegetation, absorbing red light and emitting near-infrared (NIR) light. Because the quantities of red and NIR light are otherwise very closely coupled, a difference in these two different bands of light strongly indicates vegetation.

Terrain data was derived from the Copernicus digital elevation map (DEM). The DEM provides elevation mapping at a 30-meter resolution (1 elevation value per 30 x 30 m square) within an accuracy of 2 meters of altitude. I used the differences in elevation between neighboring cells was used to calculate the grade and aspect of the surface at the locations of each estancia.

To contextualize the values we observe for both of these variables, I also extracted values for the same variables at a set of randomly generated points. The distributions at these random points represent the *null hypothesis*, what we would expect if vegetation and terrain played no role in the location of estancias. Every difference we see between the null hypothesis and the actual estancia locations is a potential insight into what drive’s the location of estancias.

**Vegetation Findings**

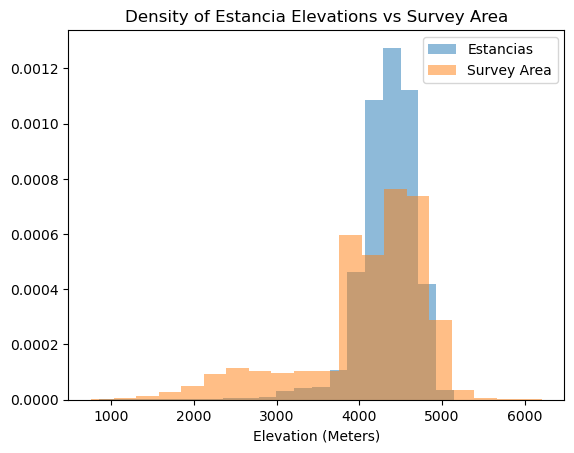
Here’s the graph we get when we look at NDVI values over the course of the year – sampled monthly and averaged over the years 2018-2020 :



The blue line represents the real estancias and the orange line represents the null hypothesis. Remember we are in the Southern hemisphere so the seasonality is opposite what you would find in the Northern hemisphere. What we can gather from this graph is that year-round there is more vegetation at estancia locations than average. This of course makes a lot of sense when we consider that ranchers would be looking for lush areas for their herds to graze. In fact, a known behavior that probably contributes to this graph is the placement of estancias around *bofedales*, highland springs/marshes that are a kind of mini-oasis up in the mountains. What is less intuitive is the lack of seasonality present in the increased vegetation. Estancias do not have to sacrifice above-average vegetation in one season for just average vegetation at other times – they can have the best of both worlds!

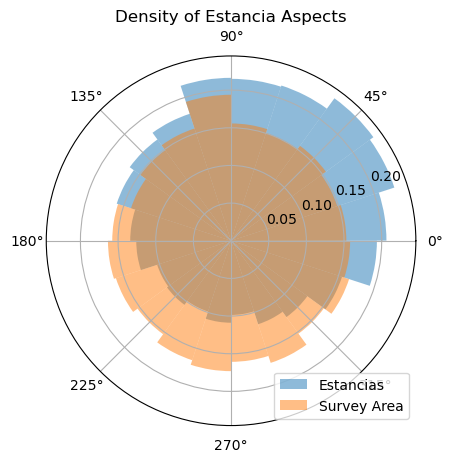
**Terrain Findings**

The first aspect of terrain I looked at was elevation, which illustrates very clearly, we are in the highlands!

****

The average altitude of these estancias was ~4300 meters! If we look more deeply, we see that estancias avoid both extremes. They aren’t built too high, which makes sense because it does get pretty inhospitable at 5,000+ meters. More surprisingly, they won’t go too low either, which warrants a bit further explanation. The most likely cause is that those lower lands are being used for agricultural which needs the low, flat terrain. Estancias, which are more flexible with where they can be located, are thus pushed out from those territories. This trend really highlights the importance of understanding the estancias together with the agricultural settlements they satellite, because otherwise they wouldn’t even be highland in the first place.

The other interesting variable related to terrain I examined was aspect. Aspect is the topography term for the direction of the slope. A northeast aspect indicates that as you walk northeast, you will be walking downhill. This might seem at first like this couldn’t have anything to do with where an estancia would be located, but it actually seems to matter a lot!



The graph above shows that Estancias are significantly more likely to be facing northeast! Why could this possibly be the case? The relevant distinction here is the differing climate between the *windward* and the *leeward* side of a mountain. The windward side is the side which faces the prevailing winds, which in the Andes is the northeast side. These winds bring moist air and cause the landscape to generally be much lusher and more vegetative. In contrast the leeward (in the central Andes, the southwest) side faces away from the wind and hosts a drier more barren climate. What this means is that we are again looking at the estancias’ preference for areas with more vegetation for their herds to graze on.

**Conclusion**

So after all our work, how does this analysis contribute to our knowledge of estancias as a part of colonial Andean society? First off, we can point to vegetation as being the dominant factor motivating the location of estancias. This is so important, in fact, that ranchers would move to the other side of a mountain to build their establishments. This may seem obvious at first, but it demonstrates a flexibility in prospecting that we wouldn’t expect if estancias were strongly tied to **one particular** agricultural settlement. How we do see estancias interact with agricultural settlements however, is that they consistently push the estancias out of the lowland, indicating a general collocation of estancias with agricultural settlements. There isn’t enough here to make definitive statements, but I think this analysis definitely supports the idea that estancias – geographically and thus presumably socially - are around agricultural settlements, but not attached to them.