**Methods**

This project focuses on data reuse. There is no raw data generated for this project, only pieces entered and then analyzed and presented in different ways. For data reuse to be meaningful, it has to be original work based on existing materials. Data reuse must be guided by best practices to standardize and preserve these materials. This includes both controls on data entry as well as files chosen for longevity.

**Introduction**

This project is a landscape type of survey. The focus is on comparisons between sites across a variety of localities to show commonality and variation. To this end, there are four primary data sources representing four published documents on the Recuay sites in the Callejón de Huaylás. Though none of this data is originally produced, it is in the analysis, replication, and interpretation originality may be achieved.

The theses by Victor Ponte (UW-M) and Rebecca Bria (Vanderbilt) present analysis and descriptions of several highland sites along the Río Santa west of the Cordillera Blanca in the Callejon de Huaylás. Victor Ponte preformed a series of skeletal analyses from a variety of sites in the southern headwaters of Río Santa, near the modern town of Huaraz at the southern end of the Callejon. Rebecca Bria worked at the site of Hualcayán and in quite a few of the other smaller sites associated with that locality. While the site is nearer the central core, it is at the peripheral edge of it. The Río Santa datasets will then be compared with the Río Puenka on the eastern side of the Cordillera Blanca. The Río Puenka is documented by Bebel Ibarra in a variety of documents including his documents and the reports from the project (?How do you cite government records, not the final report just a pile of formal documents per site?). The final dataset is from Valverde Barbosa who produced additional documentation of the Recuay tradition.

There has been some work with the Moche incursion into Recuay territory, with a focus on the Recuay and Moche interactions with the Wari. By around AD 800, both groups were overrun and archaeologically near indistinguishable from the Wari themselves. These cultural exchanges are the focus of this project. The datasets assembled here provide meaningful insight into these interactions.

Using already published data permits access to a far greater volume of data than any individual project could have reasonably collected directly. It opens up the opportunity of viewing and studying the Recuay at the regional level through the lens of the already known sites. This method does limit the project to only data which has already been published. It also opens up the possibility for error. Any error made in the original documentation is compounded in further reanalysis and discussion of the data. Any error made internally through miscomprehension of the data similarly creates possibility of error in the final product. Nonetheless, regional landscape studies are essential to the study of culture.

**Discussion of the Sites**

Bria’s work focused on the ritual economy at the Hualcayán site north of Caraz by Santa Cruz. She wrote about feasting episodes and the archaeological marks they can leave. There are several sites adjacent to the larger village which are documented in her work. Her project involved a lounge durée approach to study long term trends in the Hualcayán site and locality from early Kotosh and Chavín through the Recuay and into the Wari. Her analysis posed a bit of a challenge as she focused in depth at Hualcayán and only did surface collection at the other seventeen sites. This ends up creating a sort of lop-sided dataset. Most of the sites contain geographic information, the cultural affiliation, and description of the mortuary or residential buildings.

Barbosa’s 2008 Master’s thesis contains in depth analysis of three sites in the headwaters of the Río Ancash. These points are just south of Bria’s points. These sites each contain many mortuary structures. They ran a clustering analysis and structure volume analysis on each of the three sites. This project focused on the Chullpas themselves, their dimensions and directions of opening. The analysis was not focused on the human remains themselves, but the mortuary monuments.

Victor Ponte’s 2015 Master’s thesis is one source of data for this project. He discusses a variety of hilltop sites near the modern city of Huaraz. These sites demonstrate the tremendous variability in the Recuay hinterland. These sites include small villages and some terraced agricultural mounds. In all, a total of forty-one sites were observed and catalogued. Twenty-four of these sites included burials. Ponte’s work includes bioarchaeological assessments for all remains possible. He assessed mortuary materials and their contexts. These sites contained a wide variety of mortuary goods including Kaolin, orangewares, and effigy vessels. Copper Tupus, disks, and spatulas dominated the metal assemblage. Through the sample he was able to show there was a meaningful social stratification present. Wealthier sites with more restrictive monumental architecture and large quantities of mortuary goods stood in contrast to smaller monuments with less mortuary goods. He included his own date ranges which were correlated to Lau 2002.

Bebel Ibarra has served as the PI for the Proyecto Arqueológico Huari-Ancash since 2006. The project focused on the archaeology in the region immediately surrounding the modern village of Huari. Though there is evidence of occupation from pre-Chavín through Spanish colonial eras, the area had never been intensively studied prior to this project. The project identified many sites, mortuary complexes, and isolated manifestations. Notably the Marcajirca and Llamacorral sites from the Recuay era occur in this dataset.

**General Description of the Datasets**

The datasets are standardized and constructed deliberately to facilitate comparison. Critically the values are all coded in the same way and the type of variable kept. In this way, I can take efforts to minimize error, inaccuracy, and inconsistency in the data reuse. When engaging in data reuse these errors are expected. It is up to the re-user to minimize their affects on the final project.

The focus of this project is to compare a set of cultural features across hinterland sites from the Recuay era of the Early Intermediate Period. These sites are far from the central core of the Callejón de Huaylas yet resemble strongly their more monumental counterparts. The literature review of Recuay sites from across the twentieth century, there are a set of traits used to divide the Recuay from their neighbours. During the Cultural Historian era the Recuay were defined heavily by the seriated styles of material culture. Processual research focused on exchange and trade, militarism and conquest across the highlands. Material culture became manifested as these large cultural processes. Contemporary research too identifies the importance of understanding the material remains so to better reconstruct the Recuay as an archaeological culture.

Projects utilizing GIS require some basic spatial data in order to be useful. Fortunately Peru uses the UTM zone 18 south for official purposes. One projection for the whole nation eliminates the need to project the data once it has been entered. Latitude and Longitude are given in meters from the datum, that is meters from the origin of the zone 18 south. This does pose a question as significant figures are not common across researchers. With respect to GPS data, the significant figures rounded to meter versus centimeter or decimeter may not be all that essential given the orders of magnitude involved. The points are variable enough to make sub-meter accuracy less significant than if the points were all in closer proximity. Nonetheless these discrepancies in data entry represent a choice to not standardize, but take the values provided by other researchers. Every CSV file uses the standard measurement for location on the surface of the earth, regardless of how specific the measurements were. The additional spatial variable is altitude, discussed always in meters above sea level, or in metros sobre el nivel del mar. As with the other spatial variables, values published were taken directly and not rounded.

Most of the sites used in this project have a seriated period associated with them, but only a few have any grounded dates. For the purposes of this project, all the seriated estimates are placed within the Lau 2002 (b). Throughout the twentieth century many attempts were made to assess chronology. Early Cultural Historians focused on the phases and estimated the time between them. Processual research began to use grounded dating, but was nonetheless limited in scope by sample size. The Lau 2002(b) article is heavily grounded by comparison and forms the basis for the dataset. This variability was represented was first as a general era, a direct transcription of the way the author or researcher described the era. From here the era is placed into Lau’s grounded dating manually by assigning the grounded date ranges to the sites. In some instances the seriated dates are more specific than the estimated dates. In these instances the more specific published date is used.

Additionally, the authors used context and material remains at the site to interpret and estimate cultural affinity. Their cultural assessments are generally stated outright or interpreted by their discussion. The Recuay are an umbrella term for a variety of material and ceramic expressions over the first millennia AD. How specifically the researchers decided to break down the Recuay era is remarkably local in nature. For the purposes of this project, general phases are used to describe the Recuay, Huarás, Post-Recuay, Moche, and Wari. Specific local styles such as the Warmi style of Wari-influenced Recuay ceramics found in the Chinchawas site are coded as post-Recuay because they represent a phase of cultural transition brought on by the Wari.

The next set of variables are the attempt to represent the observed mortuary materials and their contexts. These variables are deliberately general and selected with analysis in mind. First are qualitative descriptions of the mortuary and non-mortuary architecture. Mortuary architecture is usually the graves themselves and given values such as Chullpa or Chamber to match with the general names assigned to the Recuay mortuary contexts. Non-grave structures typically are mounds or walls, structures adjacent the mortuary structures. These walls and platforms would play roles in the feasting and ancestor veneration throughout the Recuay world.

Mortuary treatment varied over the entire Recuay era. Early Recuay tombs were inaccessible in caves or buried under boulders. Early Recuay tombs also include mausoleums, underground tombs accessible through labyrinthine galleries. These early treatments appear to be holdovers from Chavín. In time mortuary treatment gave way to platforms and more accessible cist tombs. Mortuary treatment became an opportunity for community wide feasting events. The Recuay pioneered a highlands ancestor veneration tradition. The evidence is in the tombs.

Chullpas as mortuary monuments made their first appearance in the Recuay world from the north in the sixth century AD. Between the seventh and eighth centuries Chullpas would proliferate eventually becoming a common mortuary treatment. Though Chullpas are associated with the expanding Wari empire/bundle, Recuay peoples adopted these long before the Wari were established. Chullpas represent a newer more accessible means of burial. The highlands have a long tradition of accessible *Mallqui* who could be interacted with directly for the benefit of the living. Chullpas fit into this adaptive strategy well.

This project includes examples of early platforms and boulder burials. There are galleries and Chullpas from a variety of contexts. In isolation each represents a separate mortuary context. Some are monumental and suggest a significant investment of time and materials. Others are reuses of naturally occurring features in the landscape. When viewed all together, the hinterlands have a diverse set of mortuary contexts. The dataset demonstrates that the same type of mortuary customs built in remarkable monumentality in the Callejón are present in the southern and eastern peripheral sites as well.

In the interest of calculation, there are variables with simple binary presence or absence of burials and ceramic materials. These are verbal Yes or No type of variables, yet they can be converted into zeroes and ones should the need arise. There are also counts, allowing each type of burial to be represented independently of one another. Again the tally is taken directly from the researchers and coded numerically deliberately to facilitate statistical analysis and synthesis.

Ceramic made from highland Kaolinite rich clays are a hallmark of the Recuay. During the early Recuay era, they were valued as ceremonial and mortuary goods. Recuay finewares assemblages contain variable amounts across time, but the presence of Kaolin is an indicator of Recuay ceramic across researchers. Kaolin ceramic was also valued as a trade good, as evidenced by its relative prevalence throughout the Moche, Gallinazo, and neighbouring cultures. There also is a prevalence of Moche materials in Recuay contexts. The variables allow styles or types of ceramic to be displayed with a special focus on foreign materials.

In terms of cultural affiliation, most of the ceramics can be identified as Recuay, Moche, or Post-Recuay. Other diagnostic examples are Kaolin or Akillpo. Diagnostic ceramic often determines how the cultural affiliation is determined or approximated. Additionally highly valued materials such as Kaolin ceramic or foreign Moche materials indicate a more wealthy or elite type of burial. Ceramic style is more a qualitative variable which can be used in the larger discussion of changes in mortuary treatment over time and interaction with the wider group.

Within the dataset the type of burial and status goods are the beginning. In an attempt to better understand and represent the dead, there is a breakdown whenever possible for each of the mortuary contexts. These could aid in identification as well. For instance the Moche tended to segregate infants from adults and adolescents in their mortuary contexts. Infants were frequently found intermixed with the adults and adolescents in the highlands and Recuay mortuary contexts. We also find the Recuay tended to mix adult male and female. The dataset deliberately codes each count, whenever available, to demonstrate these trends.

Each of these variables are numeric, a tallied count of the total number of remains. Additionally the binary presence or absence of burial maximizing the opportunities to run statistical tests. The mortuary breakdown begins with a simple MNI, a total tallied count of the estimated minimum number of individuals each assemblage represents. These numbers vary from a minimum of one to a maximum of 129 at the Hualcayan site. There are quite a few null values, which are 0, throughout the dataset.

In the interest of gaining a more complete understanding of the dataset, the mortuary treatments are further broken down. Each file contains separate columns for adult males, females, and indeterminate. There is also a column for a combined total adult. Some sites would not have a full profile or breakdown, so in these instances they might only have a total adult column. The data may not be written standard, but it is coded into the files as standard. In keeping with the standards of human remainsno attempt was made to categorize adult or subadult sex as these tend to be inaccurate and tenuous at best (Ubelaker and Buikstra 1994: 16). A total tally of subadult and infant is the only valid way to represent these subadult dead.

**Testable Mortuary Patterns**

The central heartland of the Callejón de Huaylás represent a far wealthier set of villages and burials. Sites such as such as Pashash and Yayno represent some of the richest finds in the region. Stratification here is evident from the mortuary treatment. Key trading localities such as Chinchawas also demonstrate the variability in prestige goods such as ceramics from the coast.

Elites and powerful individuals can manifest their wealth in mortuary treatment. Presence and volume of foreign ceramic is one key variable to assess the relative wealth. Common foreign ceramics in the hinterlands are from Moche, Gallinazo, and Cajamarca. Later in the Recuay era, the Wari style ceramic becomes valued and there is a local emulating of the Wari ceramic style as well as indirect exchange for foreign pieces.

Recuay people inhabited a world where deceased personal, familial, group ancestors, or progenitors were accessible to their descendants. Burials in caves or in Chullpas were accessible to the living through defined entrances in the stone and evidence of access and reuse over centuries. Even when the entrance was sealed, the ceramics were placed near the entrance. These could be fine ware foreign and domestic, common ware much akin the general use ceramic, or highland micro-ceramics with rim diameters less than ten centimeters decorated very similar their larger counterparts.

Metalwork is another common demonstration of wealth or status through the Chullpas. Though there are occasional artistic or representative metal pieces found in mortuary contexts, the most common are Tupu pins. These pins served as functional means of holding ponchos and shawls to the wearer, as well as decorative expressions of wealth. There is immense variation in the style and type of Tupu from leaf shaped pins to very ornate and decorative pieces.

Additionally, the Recuay tended to divide their settlements between residential, sacred, and mortuary sectors. In some of the larger settlements, the mortuary sector rose to the scale of a necropolis. Some of the larger settlements such as Chinchawas included walls and causeways to direct the population to and through these sectors. Regardless of scope, the Recuay divided space between the living and the dead. Though the sectors are far less monumental in the hinterlands, a division of space may be observed.

Finally the architectural elements involved in early mortuary treatment were single cists, and cave burials. There is even reuse of Chavín temple stones to make Recuay mortuary monuments across the river from the older temple. Early treatment also includes platforms, great stones where significant mummy burials were stored and displayed. Direct interaction between the living and the dead was intended as evidenced by the reuse of tombs over time, the display of mummies, and the presence of anachronistic ceramics. After AD 600 Chullpa burials became prevalent in the heartlands and hinterlands as the arrival of the Wari cultural bundle. Though there is a change in the architectural elements found within the Recuay world, the mortuary customs were absorbed into a new style. Chullpa burials from the later Recuay world show similar stratification of grave good, presence of foreign material, metalwork, and continued reuse and reentry of the tombs.

**Four Testable Variables**

We would expect mortuary treatment in the hinterland to follow similar patterns to those in the heartland. Four primary variables can be created to identify and test based on published data to assess two overarching questions: Social Differentiation, Archaeological heterogeneity, and Cultural Change.

Social differentiation- The Recuay was a society that left meaningful and measurable marks of social differentiation in mortuary contexts. I propose to measure social differentiation by assessing status materials found at certain sites. It is expected that some of the sites will have significant access to Kaolinite, Moche, or other foreign finewares, while other sites are somewhat limited. Metal such as Tupu pins were used to hold *Mallqui* bundles together. These markers do not necessarily indicate stratification but certainly indicate a differentiation in mortuary treatment.

Archaeological heterogeneity- The Recuay world was one of significant division. There are many architectural styles, artistic motifs, and cultural attributes which connect these people together socially. But it seems the Recuay world is one where the power manifested locally as part of that wider culture. The heartland is a diverse tapestry of Recuay cities, each of which could be called a local core with a strong pull for other local cities and sites. It is expected that in the hinterland there will be some cities or sites such as possibly Hualcayán or Marcajirca which act as local cores.

Changes in culture- there are two periods of immense change with respect to the Recuay world. First comes after around AD 500 when the Recuay and coastal groups’ trade expands noticeably. These expanded trade networks manifested archaeologically as mixing of cultural materials. The second big change comes after AD 700 when the Wari expanded throughout the Central and Northern Andes. There is intensive Chullpa building in this era, as well as a high prevalence of Wari ceramic and cultural material. The expectation is that these changes in culture will be visible, and that the arrival of the Wari in the hinterland should again manifest as it did in the heartland at a different scale.

1) concentration of foreign goods in tombs

2) presence of metal work

3) concentration in type of tombs

4) presence of Chullpas/ segregated mortuary sectors

These trends are only meaningful when assigned an appropriate date range. Different authors used slightly different systems. Some dates are absolute, some seriated, some only labeled by the Early Intermediate Period. For this reason Lau’s chronology is used as a standard dating system for the Recuay culture. Whatever the author’s respective chronological system, the dates are converted to standardized Lau’s chronology.

The basic type of analysis is a common spatial test for clustering. First Moran’s I is a test on the variance of spatially local points. Getti’s G is a test on values, weighted nearest neighbor type of P-test. Both tests have both a global test on all the points across the entire dataset, and a local test which can generate the specific clusters and outliers. Together these four tests can show spatial autocorrelation and clustering within the dataset.

1) Concentration of ceramics can be assessed first as a test on the presence or absence. Concentrations of ceramic should be visible and show the use of ceramics in mortuary contexts are meaningfully similar across the Recuay hinterland. Then a test on the types, weighted so that Kaolin and foreign made fineware are heavier than the usual Recuay decorated plainwares. This test could show that there are meaningful and similar clusters of status goods across the Recuay hinterland.

2) Concentration of metal work can be assessed in the same way, but there is only a binary presence or absence variable. This is mostly because of how the original researchers documented the sites. Some had detail on the types of metalworking used while others only expressed a simple presence of metalwork. In any case presence of metalwork alone does not indicate prestige burials, but the relative abundance can.

Together the tests on ceramic and metalwork show commonality throughout the hinterlands. If these commonalities are observed, it follows they are from the same sample. More than anything these two represent an test on variance and whether the samples are consistent throughout the hinterlands. Concentrations of prestige goods should be stable throughout the hinterland. This would imply the social stratification is comparable throughout the dataset at different sites.

3) Concentration in Tomb types over time, such as platforms, subterranean burials, or Chullpas, is really more of a question of change in mortuary custom over time. This could be representative of cultural changes associated with the Wari Cultural Bundle. The changes in the heartland are well documented, with whole cities taking on a distinctly Wari flavor in the seventh and eighth centuries. The change in Recuay mortuary custom could demonstrate or approximate the highly variable rate the cultural bundle moves through the highlands.

Together with variable one, these trends could also be though of as representing the changes in Recuay material culture over time. These trends in tomb construction and in mortuary treatment provide a powerful view of the shift in culture. For instance presence of Moche, Gallinazo, or Cajamarca ceramic in a cist burial would show the tomb dates from likely before AD 550, while Wari inspired vessels in a Chullpa tomb would indicate after 550 AD. The dates are relative at most of the sites. Chinchawas produced the general seriated sequence and grounded dating of Recuay ceramics. None of these datasets have absolute dates associated with them.

4) Chullpas are present only in the Late Recuay era, a time when the Wari cultural bundle’s cultural influence was spreading dramatically throughout the highlands. Their presence alone indicates a later site, but it also can indicate a higher status individual. Chullpas were built so the Recuay could deposit and access significant ancestors. Though it is difficult to assess the specific practices of ancestor veneration and gift giving, the presence of accessible mummies in a significant stone burial monument indicates a person who was significant in life.

The assessment of Chullpas in the hinterland allows assessment of the Wari’s influence spreading and also the practices of ancestor veneration. In the Callejón heartland the Chullpa arrived long before other manifestations of the Wari bundle. The Chullpa fit well into their pre-existing practices of ancestor veneration in platform burials by making the ancestors even more accessible to their descendants. As it was adopted, the transformation of the highlands from the Recuay sequence to the pan-Andean Wari sequence began.

**General Discussion of the Variables**

When building a dataset for reuse, the key is standardization of every piece used to build the dataset. As I was personally responsible for all the decisions made throughout these processes, it falls to myself to provide detail on the decisions made. Not all the authors wrote in English. There were some discrepancies when assessing relative antiquity of the sites. Not all of the variables are represented or discussed by all the authors. The ideal dataset for this project would be filled, all the cells with values and any ability to ground truth or otherwise verify the data.

Each CSV file represents a distinct data source. Linking data together in a GIS does not require the data to all have the same source file. Regardless I compiled all four into a single Recuay Dataset so that each could be

**General Spatial Variables**

|  |  |
| --- | --- |
| **ID** | This is a simple counting number. Each site is a unique and nonrecurring value. The variable itself is included not only as a placeholder but also as a numeric variable for geoprocessing on. |
| **Site\_Name** | In the reports, an official designation of the site was nearly always given and this designation was documented. This variable represents the most precise means of identifying a site. |
| **Special** | Some sites would have a special name used to describe them. There are enough sites with only an official designation to render this variable less useful than the Site\_Name. All spellings were taken directly from the respective document or report. |
| **Longitude\_E** | Meters East of the origin of UTM zone 18 South. Taken directly from the authors without rounding or significant figures. |
| **Latitude\_N** | Meters North of the origin of UTM zone 18 South. Taken directly from the authors without rounding or significant figures. |
| **Location** | Combination variable of Longitude\_E and Latitude\_N into one column. This allows both projection of XY data into a GIS environment, but also have the two together for other geoprocessing techniques. The values from the columns are simply copied with out edit into Location column. |
| **Altitude\_MASL** | Altitude of the site, reported by the researchers. Copied directly into the column without rounding or significant figures. |

**Cultural Variables**

|  |  |
| --- | --- |
| **Era** | Period assigned based on seriation or dating scheme, correlated to Lau 2002 (b) |
| **Culture** | Author’s interpretation of the cultural affinity of the site based on material remains and context of the site. |
| **Date\_Start** | Start date assigned from Lau 2002 (b) which allows estimation for the date range of a site to be in a standard, comparable, and grounded format. |
| **Date\_End** | End date assigned from Lau 2002 (b) which allows estimation for the date range of a site to be in a standard, comparable, and grounded format. |

**General Mortuary Variables**

|  |  |
| --- | --- |
| **NonGrave\_Structure** | Architectural elements from the site that are not distinctly mortuary in nature. Walls, corrals, and mounds are common examples. |
| **Grave\_Type** | Architectural style of mortuary monument. Chullpa, Platform, Cave, and subterranean are common. |
| **Number\_Chullpa** | Tallied count of the number of Chullpas present. |
| **Number\_Cave** | Tallied count of the number of caves identified. |
| **Number\_**  **Subterranean** | Tallied count of the number of subterranean identified. |
| **Number\_Platform** | Tallied count of the number of Platforms present. |
| **Burials** | Binary variable identifying the presence or absence of mortuary materials. Coded by Yes or No. |
| **Ceramic** | Binary variable identifying the presence or absence of ceramic materials. Coded by Yes or No. |
| **Ceramic\_Type** | Diagnostic ceramics present at the site. Only when cultural affiliation of the ceramic is recorded or documented by the authors. |
| **Metal** | Binary Variable identifying presence or absence of metal materials of any type. Coded Yes or No |

**Mortuary Breakdown**

|  |  |
| --- | --- |
| **MNI** | Total Minimum Number of individuals |
| **Adult\_Male** | Count of the total number of adults estimated as male. |
| **Adult\_Female** | Count of the total number of adults estimated as female. |
| **Adult\_**  **Indeterminate** | Count of the total number of adults of indeterminate sex |
| **Total\_Adult** | Tallied total count of all adults recorded at the site. For sites without a full biological profile, total adult is all that is available and represents the highest level of specificity possible. |
| **Total\_Subadult** | Count of the total number of individuals estimated as subadults. |
| **Total\_Infant** | Count of the total number of individuals estimated as infants. |

**Coded Variables for the first round Analysis**

|  |  |
| --- | --- |
| **Burials\_Num** | Binary for presence or absence of human material. 0= No 1=Yes |
| **Ceramic\_Num** | Binary for presence or absence of ceramic material. 0=No 1=Yes |
| **Ceramic\_Type\_**  **Num** | Coded value for the type of ceramic present.  0=Absent 1=Recuay Plainware 2=Recuay Fineware or Kaolin 3=Foreign |
| **Metal\_Num** | Binary for presence or absence of metal material. 0=No 1=Yes |

**Principles of Data Reuse**

The above discussion goes into detail on the data entry and decisions made to create the files. The prerequisite for this type of analysis to produce meaningful results is valid data. As it stands the data was collected from government documents, reports, master’s theses, and dissertations. Therefore a presumption of validity is implicit in the reuse of the data. Were the research they present inadequate, reasonably a degree would not have been granted or the government would not have received the form.

Once the validity of the data is present, the next step is to standardize the data. Variables and entry define the type of analysis which can be done. Several decisions were made as to variable definition and quantification. Decisions were in the interest of creating a dataset from across published sources which could be analyzed as a whole or as individual pieces. Familiarity with the programs, namely ArcGIS, allows some anticipation of potential pitfalls caused by different types of variables.

Deciding what types of variables directly informs the coding and data entry. Here is the first source of error directly introduced during this project. Though there could be error in the course of the field research, such inconsistencies would be mostly undetectable from the outside once the reports are written. Any error created by entering and coding data into datasheets is my own. Efforts are taken throughout the processes to minimize these potential for errors.

The spreadsheets were all saved as Comma Separated Value or CSV files. CSV files represent a very simple tabular way to store the data. The file itself is a series of values separated by commas which can be reconstructed into a spreadsheet in any proprietary or open sourced tabulation software. It is a simple file format and thus a very small file size. Owing to the simplicity of the file format, the CSV has become the preferred file format for long-term data storage.

Once the data was entered, the data was cleaned. Cleaning data is one way to assess and minimize potential error from entry inconsistencies in the data. Cleaning includes searching for errors in the entry and discrepancies in the data. These can include anything from misspelling to extra letters in one cell to superfluous spaces. Once the sheets were completed and in their usable form, they were updated to the Open Refine environment. Each column was faceted by text or number and analyzed. There were occasional minor errors, though in general the data was fairly cleaned. One source of inconsistency were in entries of the binary variables such as “NO” versus “No” or “no” text entries. Others included minor spelling or spacing errors. As all data entry was done personally, there was limited ambiguity in the meanings of such errors. Once the data was prepared it became the **DatasetRecuay\_Cleaned** which was used for the first round of analysis.

In the interest of good data reuse and management technique, there is a Git Hub repository set up for the data. Git Hub is one way to store and curate raw data and files. Git Hub is a software for hosting and storing repositories online. Though it was originally designed to facilitate remote software development, scholars can use Git Hub as a space to store and back up any materials necessary for the research process.

**I’ll put it all into one simple repository that I host once I’m done. It doesn’t exist quite yet**