Analysis and Discussion Chapter

**Date Queries**

The first piece was to take the **DatasetRecuay\_Cleaned** and generate layers to show the different eras. Each of the layers contains the same dataset, just queried to show an era better. Some sites in the dataset were single use burial plots, while others were residential areas occupied for the entire first millennium AD. Some sites are represented multiple times in the various layers below, implying they were in use in multiple eras. All of the sites are represented at least once in the subset layers below.

**Existent Sites in AD 1** The first layer was queried so that the Date Start is less than or equal to 0, implying the site has a start date from the before the first century. This layer is primarily used as a means of showing the dataset before the Recuay. This subset contained twenty-one records.

**Huaras Recuay** This layer was queried so the Date Start equals -200 (200 BC). Only the sites with a Huarás occupation were coded with a start date at -200. There are some sites which eventually had evidence of a Recuay presence and a Huarás presence, these are all in this dataset. This subset contained a total of ten records.

**Classic Recuay** This layer was queried so the Date Start was greater than or equal to zero and the Date Start was less than 600. Together these select sites between AD 1 and 600: sites occupied during the Classic Recuay era. This subset contained a total of thirty-three records.

**Late Recuay** This layer was queried so the Date End was greater than or equal to 650. The Late Recuay Era starts in AD 650. There is a bit of a gap between Classic’s end in 600 and Late’s beginning in 650. That distinction was built into the data deliberately as a distinction to divide Classic and Late sites. This subset contained a total of forty-six records.

**Post Recuay** This layer was queried so the Date End is greater than 700. This would produce only sites whose ending was after the Recuay and in the Wari era. Some sites were used or occupied through Inka and even to Spaniard times, the exact end date for each site is grouped together if it is after AD 700. There were twelve records in this subset.

This dataset contains all the sites present for this project. Naturally most of the sites fall into the Classic and Late Recuay. Any site present in the dataset should have at least a portion of its date range within the Recuay era.

These divisions represent the fundamental units of analysis for the first round.

**Maybe test for oversampling?**

**Description of the test Sequence**

There are a total of six trials in the first round, each of them running a set of analysis on the numerical variables. There are two tests used for this first round of analysis. Moran’s I and Getis G are the primary means of assessing spatial autocorrelation and hotspot analysis. Though both tests discuss clustering, the results are distinct from one another.

Moran’s I is the classic statistic for assessing spatial autocorrelation. It was pioneered in the late 1940s for use in assessing points for spatial ‘joins’ in the data. The test assumes either that the points are independent from one another, or that they were randomly arranged (Moran 1948). For the purposes of this project, all points are assumed independent from one another. Though they are not randomly arranged, there was no indication that any site was assigned values identical to another. There is independence, if not randomness implied.

Moran’s I Global statistic generates a co-efficient between zero and one. The I value of positive one indicates a strong spatial clustering in that area. The points tend to contain high values and low values adjacent or nearby one another. An I value of negative one indicates a strong negative autocorrelation. This implies the points are truly dispersed with high and low values around one another. An I value of zero indicates an absence of a pattern. This indicates there is something akin to randomness as the prominent feature (Rogerson 2015: 302).

Patterns in the data are autocorrelated when observed values geographically near one another possess similar values. The I statistic can demonstrate that values near one another are not random, but that an extraneous factor of geography is influencing the data. This is good for social sciences because many features created by man are placed deliberately. Showing that a given variable is more clustered than could be random is another way of saying that these features were consciously placed on the landscape.

The Global statistic is merely a descriptive statistic, a co-efficient implying relatedness or lack thereof in the overall dataset. The GIS program runs the weighted test, and then outputs the I co-efficient together with a Z-Scored value and a P-Test for significance. The P-Value of .05 was used as a general conventional cutoff. The output also displays a null curve demonstrating the probability the spatial patterns could be random based off the tests for significance. Though the global test can never display where the autocorrelation and dispersal occurs, it can demonstrate the presence of such correlation within the data.

Global statistics are calculated only for the Recuay Dataset, the Classic Recuay, and the Late Recuay. This was a deliberate decision as Moran’s I requires N=30 to produce results. The Huaras and Post Recuay layers had ten and twelve records respectively. Had I calculated Moran’s statistic the value would have outlandish degrees of freedom pushing the test statistic’s value towards insignificance.

Local tests are used to demonstrate the Global clusters spatially. Local Moran’s I runs a related analysis, though the output demonstrates the location of the autocorrelated or dispersed points. Local statistics can detect clusters which may or may not be significant across the entire dataset, but hold local significance (Rogerson 2015: 308-9).

The GIS output for Local (Anselin) Moran’s I is a map layer displaying the points with symbology to match. Where clusters occur they are placed into categories of high-high and low-low clusters. Outlier points occur whenever there is a point with values significantly different from the other points around it. These outliers are labeled based on whether they are high values surrounded by low values or low values surrounded by high. In either case there is an outlier occurring. It also can display in the output nonsignificant values. This case would imply any point is either random or near random with respect to points around them. These points are not autocorrelated with the points around them.

Third test is Getis G General, the global stat calculating hot and cold clusters in the data.

The Fourth test is the Getis Ord Gi which calculates local stats and produces a layer showing the clustering of high and low values.

Describe high-low ect clustering what does it mean?

Each Trial run produced all four tests on the variables.

For the Global stats to be valid, the N value has to be at least thirty. For this reason the Global stats are only calculated on the Classic Recuay, the Late Recuay, and a control with the general Dataset Recuay Cleaned layer. This should indicate whether the data is autocorrelated because of the values per era or because of an overwhelming autocorrelation in the underlying dataset.

**Trial 1**

The first trial is demonstrative. Global and local tests were ran on it to show there is significant autocorrelation and clustering on a variable otherwise insignificant to analysis. The dummy variable selected was the basic ID number, a counting number starting at one and continuing consecutively through the entire dataset. Values with lower ID numbers are simply the first entered into the file, higher ID numbers are the last. Given the numbers were assigned first by author and then combined into one sequence, the data is necessarily autocorrelated.

The Bria points were entered first and the Ibarra points were entered last, and their ID counting number reflects this decision. To put it another way: the datasets are clustered deliberately by ID number. Trial 1 was designed to example that clustering.

The first test was the Global Moran’s I test for autocorrelation. The Recuay Dataset produced an I value of 0.957871 which produced a Z-Score of 9.643685. Together these produce a P-value of 0.00, implying the Recuay Dataset is extraordinarily autocorrelated. The Local Moran’s I produced Global Getis Ord G produced a G value of 0.027608 and a Z-score of 7.352501. Together these produced a P-value of 0.00 implying the hot and cold spots are not randomly clustered.

The Classic Recuay Global Moran’s I produced an M value of 0.938408 and a Z-Score of 7.668169. Together these produce a P-value of 0.00 implying the Classic Recuay dataset is extraordinarily autocorrelated. The Global Getis G produced a G-value of 0.041988 and a Z-score of 5.394184. Together they produced a P-value of 0.00 indicating the hot and cold spots are not randomly clustered.

Local Moran’s I statistics for the Classic era produced Low-Low clustering around the points associated with Bria in the north, and a High-High clustering in the points near Huari from the Ibarra dataset. Barbosa and the sites from Ponte near Huarás are were insignificant. Local Getis G showed hotspots in Huari and coldspots near Santa Cruz. One point in the Barbosa dataset was a lesser cold point. These clusters and hotspots in the Classic era were as expected.

The Late Recuay Global Moran’s I test produced an I value of 0.949636 and a Z-Score of 9.541934. Together these produce a P-value of 0.00 implying the Late Recuay dataset is extraordinarily autocorrelated. Global Getis G produced a G-value of 0.029629 and a Z-score of 6.786307. Together these produce a P-value of 0.00 implying the hot and cold spots are not randomly clustered.

The Local Moran’s I statistics for the Late Recuay show the same pattern as during the Classic: High-High clustering in Huari and Low-Low in the Bria points around Santa Cruz. The Late Recuay dataset has more points than the Classical, but the clusters remained the same. The Local Getis Gi produced hotspots in Huari and coldspots in Santa Cruz. One point in the Barbosa set is a coldspot, while nothing around Huarás is significant.

Trial 1 was designed to establish principle and demonstrate that the dataset is autocorrelated beyond reproach. All three Global Moran’s I stats produced a P-value of 0.00. The dataset is extraordinarily autocorrelated with respect to the entire dataset, and the dated subsets. There is statistically no chance that the deliberately sequential ID number is random. All three Getis Gi stats were 0.00 showing no chance the hot and cold spots were random.

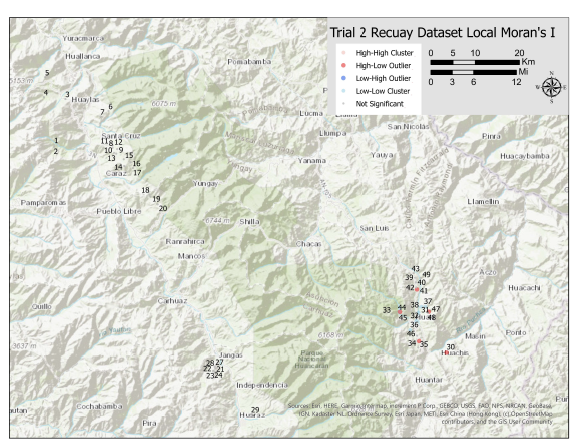
Therefore Trial 1 was a success. Beyond showing deliberate and artificial clustering in the dataset, the values do not technically show or demonstrate anything about the datasets. Trial 1 demonstrated that the tests were sensitive enough to note clustering and autocorrelation in the datasets.

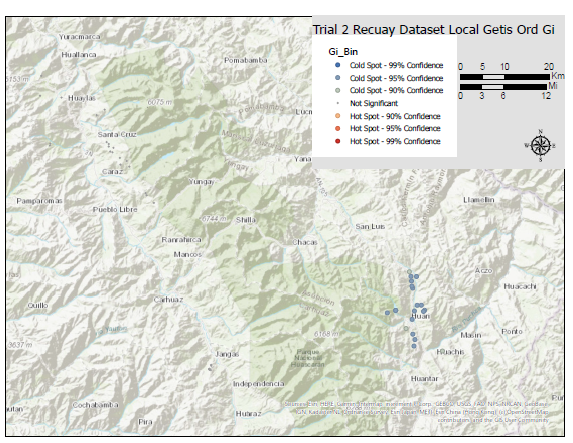
**Trial 2**

The second trial ran the same tests on the Burial\_Num variable, presence or absence of any human remains. The values were all zeroes and ones. Global Statistics are calculated only for the whole Recuay Dataset, Classic Recuay, and Late Recuay eras.

The Recuay Dataset produced a Global Moran’s I of 0.006463 and a Z-Score of 0.266675. The P-Value is 0.789720 indicating negligible clustering not significantly distinct from random. The global Getis Gi statistic was 0.020457 and a Z-Score of -0.172047. The P-value of 0.863401 indicates the general hot and cold clustering is not significantly distinct from random.

Local statistics for the Recuay Dataset showed low-low clustering around the Huari locality with a few outliers at sites 41,47,34,30, and at 44 and 45. Site number 44 is the Llamacorral and number 45 is Ishla Ranra, which are outliers of many tests as they are adjacent the Laguna Purhuay. This low-low clustering around Huari was also a statistical cold-spot from the local Getis Gi test, a **complimentary** result.





Existent Sites in AD1 Local Moran’s I produced only outliers. High-low outliers were present at sites 6 near Santa Cruz, site 47 near Huari, and site 44-45 Llamacorral Ishla Ranra cluster. There were low-high outlier at site 10 near Santa Cruz. There was nothing significant from the local Getis Gi statistic. The Huarás-Recuay set Local Moran’s I produced only outliers as well. Again the 44-45 produced a high-low outlier. Local Getis Gi statistic produced nothing significant.

The Classic Recuay layer produced a Global I value of -0.062948 and a Z-Score of -0.247280. The P-value is 0.804692 indicating slight negative clustering not significantly distinct from random. The Global Getis G value was 0.029525 with a Z-score of -0.578446. The P-value of 0.562963 indicates the hot and cold clustering is not distinct from random. Local Moran’s I analysis for the Classic Recuay layer produced only High-Low outlier at site number 34 Mashuanco, the local Getis Gi produced no significant clustering.

The Late Recuay layer produced a Global Moran’s I of 0.014488 and a Z-Score of 0.075282. The P-Value of 0.939991 indicates the slight clustering is not significantly distinct from random. The global Getis G statistic was 0.020244 with a Z-score of -0.733104. The P-value of 0.463495 indicates the hot and cold clustering is not significantly distinct from random. **Both insignificant**

Local I statistics for the Late Recuay showed a high-high cluster along the Río Ancash and a low-low clustering in Huari. There were several high-low outliers including 44-45, as well as sites 34, 30, and 41. Getis Gi local statistics **complimented** these clusters: there was a hotspot along the Río Ancash and a cold spot in Huari at the 95% significance level.

|  |
| --- |
|  |

The Post Recuay layer contained only outliers spread throughout the dataset. High-low outliers were visible in Huari at sites 30 and 12, Huaraz at site 28, and site 26 near Santa Cruz. Low-high outliers were also present in Huaraz site 24 and Huari site 35. Local Getis Gi produced no significant hot or cold spots in the data.

Given the looting, tropical climate, survival, and recovery biases, it is reasonable that presence or absence of human material might follow a seemingly random pattern in the highlands. All three layers tested for global statistic produced insignificantly distinct from random patterning among the Recuay. Outliers were present throughout the dataset, with no single locality dominating the assemblage of sites.

**Trial 3**

The third trial used the Ceramic\_Num variable, presence or absence of ceramic materials. This variable contains only zeroes and ones.

The Recuay Dataset produced a Global Moran’s I value of 0.243520 and a Z-score of 2.597638. The P-value of 0.009387 indicates the dataset is clustered beyond random chance. The global Getis Gi value is 0.022625 with a Z-score of 1.382472. The p-value of 0.166827 implies the hot and cold clusters are not significantly different from random.

Local statistics for the Recuay dataset showed a high-high clustering in Huari and a low-low clustering in Santa Cruz. High-low outliers were north of the Santa Cruz cluster at sites 5, 7, and 8. There was also a low-high outlier at Huaraz. Local Getis Gi **complimented** the I values showing a hotspot in the Huari locality and the coldspot near Santa Cruz.

Similarly the Existent Sties in AD1 showed a high-high clustering in Huari, a high-low outlier in site 42, and sites 44-45 Laguna Purhuay. Santa Cruz only showed a series of outliers: a high-low outlier at site 6 and a low-high outlier at site 10. Local Getis Gi **complimented** the I statistics as the Huari clustering is a coldspot at the 95% confidence level. Local statistics for the Huarás layer produced only an outlier at the Laguna Purhuay sites 44-45. Getis Gi showed these sites and site 33 in Huari as coldspots.

The Classic Recuay dataset produced a Global Moran’s I value of 0.236909 and a Z-score of 2.084085. The P-value of 0.037152 indicates the autocorrelation is statistically different from random. The global Getis Gi value is 0.033478 which gives a z-score of 0.602661. The P-value of 0.546734 indicates the hot and cold clustering is not significantly distinct from random.

Local statistics showed significant low-low clustering in Santa Cruz with high-low outliers to the north at sites 1,5,6, and 8. A single low-high outlier was present at site 25 in Huaraz. Getis Gi **complimented** these results showing coldspots around Santa Cruz at the 95% confidence. Isolated hotspots were also identified at site 21 in Huaraz as well as at 30 and 35 in Huari.

The Late Recuay dataset produced a Global Moran’s I value of 0.304899 and a Z-score of 3.204363. The P-value of 0.001354 implies the autocorrelation is significantly distinct form random. The global Getis Gi value is 0.024980 and a z-score of 1.895379. The p-value of 0.058042 indicates the hot and cold clustering is on the cusp of significance at the .05 level.

Local Stats showed significant low-low clustering in Santa Cruz and a high-high clustering in the south of Huari. North of Santa Cruz were high-low outliers at sites 5,7, and 8. Getis Gi **complimented,** albeit dubiously significant. There was a visible coldspot in Santa Cruz and a hotspot in Huari. There were also a few isolated hotspots north of Huaraz and on the Río Ancash.

Post Recuay produced only a single low-high outlier at Santa Cruz number 10. There was no significant hot or cold spots from the local Getis Gi.

All three layers tested for global significance showed significant autocorrelation but insignificant hotspot and coldspots. Locally the results tended to compliment one another.

**Trial 4**

The fourth trial used the coded values from the Ceramic\_Type\_Num. Trial 3 assessed clustering with respect to presence or absence of ceramic, Trial 4 assesses clustering with respect to the classes of ceramic (Local plain, Local fineware, foreign made).

The Recuay Dataset produced a global Moran’s I value of 0.225355 with a Z-score of 2.450499. The P-value of 0.014266 indicates the autocorrelation is statistically distinct from random. The global Getis Gi value is 0.023621 with a Z-score of 1.436216. The P-value of 0.150941 indicates the clustering of hot and cold spots is not distinct from random.

Local clusters included a low-low clustering in Santa Cruz, a high-high clustering in Huaraz with some significant low-high outliers at sites 23 and 24 near Huaraz. 44-45 Llamacorral site and 42 in Huari are also outliers. Local Getis Gi **complimented** these results with coldspots in Santa Cruz and Huaraz at the 95% significance level.

Local tests on the existent sites in AD1 showed only outliers, high-low at 44-45 and site 42 in Huari, as well as site 6 in Santa Cruz. There was also a low-high outlier at site 10 in Santa Cruz. Local Getis Gi showed no significant hotspot or coldspots. The Huaras layer similarly showed only one high-low outlier at Llamacorral 44-45. **Complimentary** there was a coldspot at the laguna Purhuay.

The Classic Recuay dataset produced a global Moran’s I value of 0.416340 and a Z-score of 3.543422. The P-value of 0.000395 indicates significant autocorrelation in the data. The global Getis Gi value is 0.041017 and z-score of 2.119290. With a p-value of 0.034066 the hot and cold spot clustering is statistically distinct from random. **BOTH SIGNIFICANT .05**

Local statistics on the Classic Recuay showed high-high clustering in Huari and Huaraz, with a high-low outlier at site 39 in Huari. There was a low-low cluster near Santa Cruz with high-low outliers along the periphery at sites 15, 6, and 3. There was a high-low outlier at site 18 along the Río Ancash. Getis Gi **complimented** the results showing a coldspot in Santa Cruz and a hotspot near Huaraz at the 95% confidence level.

The Late Recuay dataset produced a global Moran’s I value of 0.249135 and a z-score of 2.694425. The p-value of 0.007051 indicates the autocorrelation is significantly different than random. The global Getis Gi statistic was 0.026335 and z-score of 1.787636. The p-value of 0.073835 indicates the hot and cold spot clustering is on the cusp of being distinct from random.

Local tests showed low-low clustering in Santa Cruz, with high-low outliers n the north. Huaráz exhibited a high cluster with sites 23 and 28 as low-high outliers. At Huari there was a low-low cluster in the 44-45 Llamacorral site. Getis Gi values corroborate with hotspots in Huaráz and coldspots at Santa Crux at the 95% significance level.

Finally the post Recuay era showed only a single low-high outlier at site 10 in Santa Cruz. Local Getis Gi showed nothing significant.

The global tests show significant autocorrelation on all three layers as well as one global Getis Gi. In all these values are complimentary.

**Trial 5**

Trial 5 tested the Metal\_Num variable. This variable is binary, indicating presence or absence of metal work at a site.

The Recuay Dataset produced a global Moran’s I value of 0.462331 and a z-score of 4.861910. The p-value of 0.000001 implies the data is extremely autocorrelated across the dataset. The global Getis Gi value was 0.065857 and z-score of 4.326386. The P-value was 0.000015 implying the hot and cold spot clusters are significantly different from random. **Both significant**

Local tests showed a low-low clustering in Huari and at Santa Cruz. Huaraz showed high-high clustering as well. Local Getis Gi values **complimented** these results with a cold spot in Huari and a hotspot in Huaraz at both the 95 and 99% significance. There was a weaker cold spot at Caraz near Santa Cruz.

Local tests on the existent sites in AD1 showed only outliers, both in Santa Cruz. There was a high-low outlier at number six and a low-high outlier at number 10. Local Getis Gi values showed a cold spot in Huari and a hotspot including site number 10. The local statistics could not be run on the Huarás Recuay layer because all the values in the Huarás layer contained no metal. As such all the sites were coded as zeroes and no meaningful clustering or hotspots are statistically possible.

The Classic Recuay dataset produced a global Moran’s I value of 0.348017 and a z-score of 3.084549. The p-value of 0.002039 indicates the data is significantly autocorrelated. Global Getis Gi value was 0.084501 with a z-score of 2.563004. The p-value of 0.010377 indicates the hot and cold clusters are distinct from random. **Both Significant.**

The local Moran’s I showed a low-low cluster in Santa Cruz and a high-high clustering in Huaraz. Sites 21 and 22 near Huaraz were low-high outliers and site 6 near Santa Cruz was a high-low outlier. Getis Gi **complimented** these results showing hotspots at Huaraz into the 99% confidence level, and a coldspot at Santa Cruz at the 95% level.

In the Late Recuay dataset, global Moran’s I statistic is 0.538277 and the z-score 5.651898. The P-value is 0.00 indicating the autocorrelation is significantly different from random. The global Getis Gi value was 0.083341 with a Z-Score of 5.065333. The P-value is 0.00 implying hot and cold spot clustering is different from zero. **Both Significant.**

Local statistics were calculated showing low-low clustering in Huari and high-high clustering in Huaraz. There was an low-high outlier at site 23 near Huaraz. Getis Gi **complimented** these results well showing a coldspot in Huari and a hotspot in Huaraz at the 95 and 99% confidence levels. There was a single site at number 17 near Santa Cruz which was a coldspot at the 90%.

**Post Recuay?**

Global statistics for trial 5 showed significant autocorrelation and extreme value clusters. Local tests were generally complimentary.

**Trial 6**

Trial 6 preformed the same tests for autocorrelation and hotspots on the MNI variable. This is the only variable in the first series of tests not generated internally. The value is the number of individuals as published by the author, no coding or binary assessment.

The Recuay Dataset produced a global Moran’s I value of 0.006486 and a Z-score of 0.381770. The P-value was 0.702632 indicating the autocorrelation was not distinct from random. The global Getis Gi produced a value of 0.023217 and a Z-score of 0.072228. The P-value of 0.942421 implies the hot and cold spot clustering is not significantly different from random. **Both insignificant.**

Local tests on the Recuay dataset showed low-low clustering in Santa Cruz with high-low outlier at number 6 and low-high outlier at number 7. There was also high-high clustering at Huaraz. There was significant low-low clustering in Huari at the Laguna Purhuay sites 44-45. Getis Gi produced only a single hotspot in Santa Cruz only at the 90% level.

The Existent sites in AD1 showed only outliers at Santa Cruz. There was a high-low outlier at number 6 and a low-high outlier at number 10. Getis gi showed a hotspot in Santa Cruz as well.

Huarás Recuay local Moran’s I produced only outliers. There was a low-high outlier at site 33 and a high-low outlier at sites the Laguna Purhuay sites 44-45. Local Getis Gi test showed nothing significant.

In the Classic Recuay dataset, the global Moran’s I value is -0.017189 which gave a Z-score of 0.390948. The P-value of 0.695836 indicates the autocorrelation is not distinct from random. The global Getis Gi value is 0.015810 with a Z-score of -0.450429. The P-value of 0.652401 implies the hot and cold spot clusters are not distinct from random. **Both insignificant**

There were significant local low-low clustering in Santa Cruz with a high-low outlier at number 6. There was a single high-low outlier at site 41 in Huari. Getis Gi showed a hotspot in Santa Cruz, a result **not complimentary** with the local Moran’s I. **Both are not distinct global stats.**

In the Late Recuay dataset, the global Moran’s I statistic is 0.001522 with a Z-score of 0.330361. This generated a P-value of 0.741127, implying the autocorrelation is not significantly different from random. Similarly the global Getis Gi value of 0.018013 and Z-score of -0.115273 produced a P-value of 0.908229 implying insignificant hot and cold spot clustering. **Both insignificant**

Local Moran’s I showed only outliers in Santa Cruz, a high-low outlier at number 6 and a low-high outlier at number 7. In Huaraz there was another high-low outlier at site 28. Local Getis Gi showed nothing significant.

**Discussion**

Problem with outliers at site 44 and 45. The Llamacorral site is an anomaly built in the lakebed of glacial Laguna Purhuay. At the summit of a cliff is the residential settlement Ishla Ranra built at the top of a cliff. These two sites literally on top of one another separated by mere meters of latitude and longitude but distances in elevation. The two sites were often outliers on one another. I don’t think the results are invalid, they are two distinct sites with distinct attributes, it is logical they are commonly autocorrelated or outliers.

I’m sorry, I bit off a bit more than I could actually get done. Above is a raw discussion of the data. All the layers in bold, showing complimentary or statistically significant results will be accompanied by a basic map and chart to show what I am talking about.

Once I get done with exams I’ll be back on it.