# The Association of Dengue Disease with Temperature, Precipitation and Vegetation index in Tropical and Sub-tropical Parts of the World

Final report prepared for the Data Analysis and Interpretation Specialization

March 10, 2017

## Introduction to the Research Question

The purpose of this study was to predict the number of dengue cases each week for each location (city, year and week of year in San Juan and Iquitos) based on environmental variables and find best predictor factors describing how changes in maximum, minimum and average air temperatures, total precipitation, relative and specific humidity and satellite measured vegetation index affect number of disease cases.

As data analyst, my current task is to predict the number of dengue cases each week (in each location) based on environmental variables describing changes in temperature, precipitation, vegetation and other factors.

The research to identify factors associated to epidemic diseases is very important to the public health and may help to better understand if in these case factors are related to climate change. These days many of the nearly half billion dengue cases per year occurring in Latin America. In many cases dengue disease causes severe health problems and even death. Accurate dengue predictions would help public health workers and people around the world take steps to reduce the impact of these epidemics. Although the relationship with climate is complex, a growing number of scientists argue that climate change is likely to produce distributional shifts that will have significant public health implications worldwide.

## Methods

### Sample

The sample included N=1111 weekly environmental measurements for San Juan (Puerto Rico) tropical city and Iquitos (Peru) sub-tropical cities. Data, indicators and measurement provided by NOAA's GHCN (daily climate data weather station measurements), NOAA's NCEP (Climate Forecast System Reanalysis measurements), NOAA's CDR (Normalized Difference Vegetation Index), PERSIANN satellite precipitation measurements.

San Juan is the capital of Puerto Rico with population around 395,236 (based on 2010 census) with tropical monsoon climate with well distributed rainfall but the months of January, February, and March are the driest. Annual rainfall has historically ranged from 35.53 in (902 mm) in 1991 to 89.50 in (2,273 mm) in 2010.

Iquitos, capital of the Peruvian Amazon with population of 471,993 inhabitants, with equatorial or sub-tropical climate and constant rainfall throughout the year, without a distinct dry season, but a wetter summer. The rainy summer arrives in November and ends in May. March and April have the heaviest rains and humidity, with precipitations of about 300 and 280 millimeters (12 and 11 in), respectively. In May, the Amazon River, one of the rivers surrounding the city, reaches its highest levels. It falls about 9 or 12 meters (30 or 39 ft.) at its lowest point in October, and then steadily rises again cyclically according to rainfall.

### Measures

The response variable TOTAL_CASES represents weekly counts of dengue cases for up to 52 weeks per year for each location, San Juan (SJ) and Iquitos (IQ) from 1990 to 2010.
Predictors for San Juan (SJ) and Iquitos (IQ) included:

1. Week of the year (weekofyear), quantitative variable, week id's ranging from 1 to 52
2. Mean of specific humidity (specific_humidity_g_per_kg), quantitative variable
3. Total millimeters precipitation amount (station_precip_mm)
4. Satellite average vegetation index (Satellite average vegetation index (vegitation_index_avg)
5. Maximum temperature Celsius(station_max_temp_c )
6. Minimum temperature Celsius(station_min_temp_c )
7. Average temperature Celsius (station_avg_temp_c )

8. Mean dew point temperature in Kelvins (dew_point_temp_k)

**Analysis.**

The distribution of dengue cases and all predictors were evaluated by examining mean, standard deviation and minimum and maximum values for all quantitative variables, including univariate analysis for outliners. Based on analysis results, outliners were kept in dataset.

The Pearson correlation was used to test correlation coefficient between variables. The ANOVA used to test analysis of variance to conduct bivariate analyses as well. The General linear model (GLM) was used to test basic linear regression model for the association between explanatory variables and response variable to test strength of relationship between variables. The multiple regression including STEPWISE variable selection was also used to test possible relationship between primary variable and additional confounders (variables) in the model, various graphs and plots were also used for analysis of data distribution.

To predict total dengue cases Penalized regression method (Lasso - Least Absolute Selection and Shrinkage Operator) regression was used to test model and to provide greater prediction accuracy for both locations (N=1111) and for each separately, San Juan N=724 and Iquitos N=387. Prior to conducting LASSO regression all predictor variables were standardized with mean=0 and standard deviation=1. The estimation of LASSO regression model was performed with 70% of training set and 30% of test set for both and each location separately.

In addition, K-MEANS cluster analysis were applied on training set to create K=1-10 clusters using Euclidean distance to partition observations into smaller set of clusters based on similarity of responses on multiple variables. All clustering variables were standardized using STANDARD procedure to have also a mean of 0 and standard deviation of 1. The training and test sets created with 70% in training and 30% in test. Observations with missing values removed prior creation of both sets. Iquitos (IQ) training set N=271), test N=116, San Juan (SJ) training set N=507, test N=217.

## Results

### Descriptive statistics

The geographical locations of sampled cities location are different. Table 1 shows descriptive statistics for San Juan with average maximum temperature in Celsius of 31.6 (mean=31.6, std=1.5), maximum of 329 Dengue cases (mean=30 and std=36), precipitation maximum of 163.1mm (mean=26.07, std=26.80) which is significantly different from Iquitos descriptive statistics shown in Table 2 with average maximum temperature in Celsius of 34 (mean=34.6, std=1.3), maximum Dengue cases of 83 (mean=9, std=10) and precipitation maximum of 543.3mm (mean=68.18, std=68.12). The maximum temperature in Celsius

**Table 1. San Juan (SJ) – Weekly Descriptive statistics for reported Dengue Cases**

| Label | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| Dengue Cases | 724 | 30.3397790 | 36.1541271 | 1.0000000 | 329.0000000 |
| Weekly precipitation mm | 724 | 26.0745856 | 26.8046525 | 0 | 163.1000000 |
| Mean specific humidity | 724 | 16.5879282 | 1.5522984 | 11.7157143 | 19.4400000 |
| Satelite Vegitation index average | 724 | 0.1170797 | 0.0585700 | -0.0925646 | 0.3423375 |
| Maximum temperature Celsius | 724 | 31.6244475 | 1.7000436 | 26.7000000 | 35.6000000 |
| Minimum temperature Celsius | 724 | 22.6241713 | 1.5045744 | 17.8000000 | 25.6000000 |
| Average temperature Celsius | 724 | 27.0368587 | 1.4182717 | 22.8428571 | 30.0714286 |
| Mean dew point temperature Kelvins | 724 | 295.1447415 | 1.5618508 | 289.6428571 | 297.7957143 |

**Table 2. Iquitos (IQ) - Weekly Descriptive statistics for reported Dengue Cases**

| Label | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| Dengue Cases | 387 | 8.9147287 | 9.8117232 | 1.0000000 | 83.0000000 |
| Weekly precipitation mm | 387 | 66.1813953 | 68.1285231 | 0 | 543.3000000 |
| Mean specific humidity | 387 | 17.3530786 | 1.3526227 | 12.1114286 | 20.4614286 |
| Satelite Vegitation index average | 387 | 0.2535655 | 0.0736128 | 0.0841554 | 0.5039822 |
| Maximum temperature Celsius | 387 | 34.0142119 | 1.3410921 | 30.1000000 | 42.2000000 |
| Minimum temperature Celsius | 387 | 21.3824289 | 1.1662798 | 16.4000000 | 24.2000000 |
| Average temperature Celsius | 387 | 27.5813664 | 0.9490614 | 21.4000000 | 30.8000000 |
| Mean dew point temperature Kelvins | 387 | 295.7400628 | 1.3188935 | 290.0885714 | 298.4500000 |

The analysis of reported weekly total Dengue cases revealed difference in distribution between San Juan and Iquitos. Figure 1 – San Juan, shows bimodal distribution with noticeable increase in Dengue cases (means) starting from week 28 (means=35), maximum mode at weeks 34, 35, 40 (means=58, 59 and 59), weeks 52, 2, 3 (means=42, 35, 36) and lowest observed values between weeks 8 and 22 (means=15), lowest in weeks 17, 19 and 20 (means=10). Distribution of Dengue cases means corresponded to start and end rainfall season.

**Figure 1. Variables Association: weekofyear and total_cases (means) - SanJuan (SJ)**

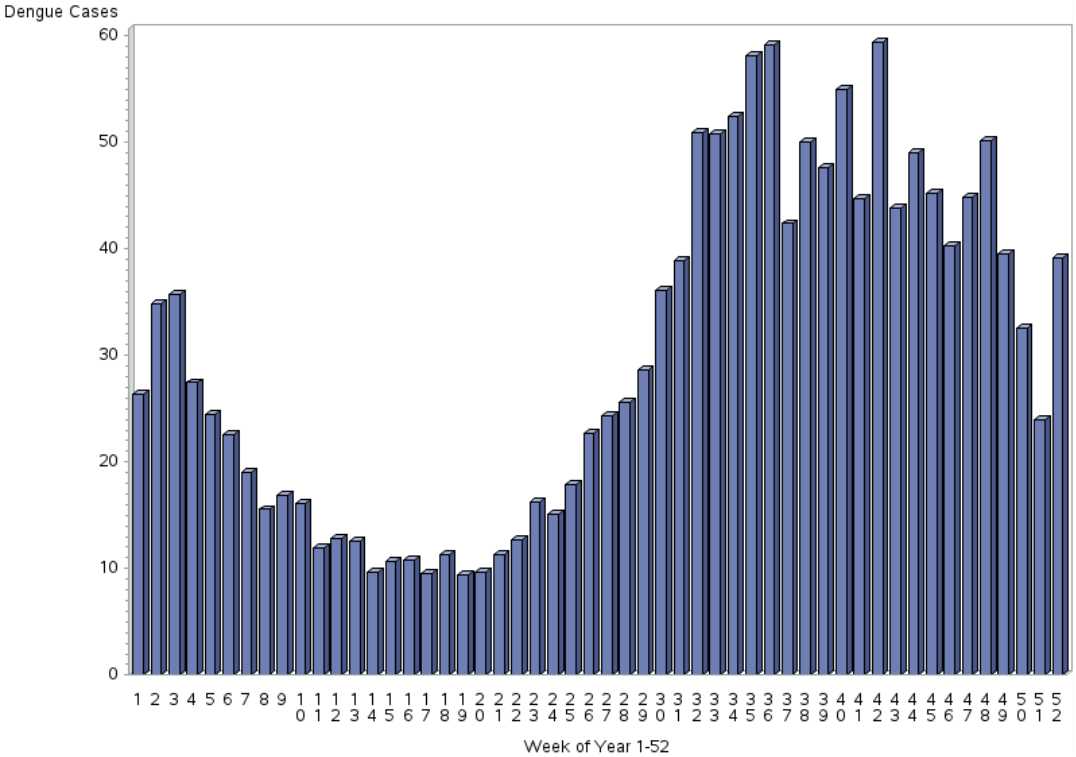
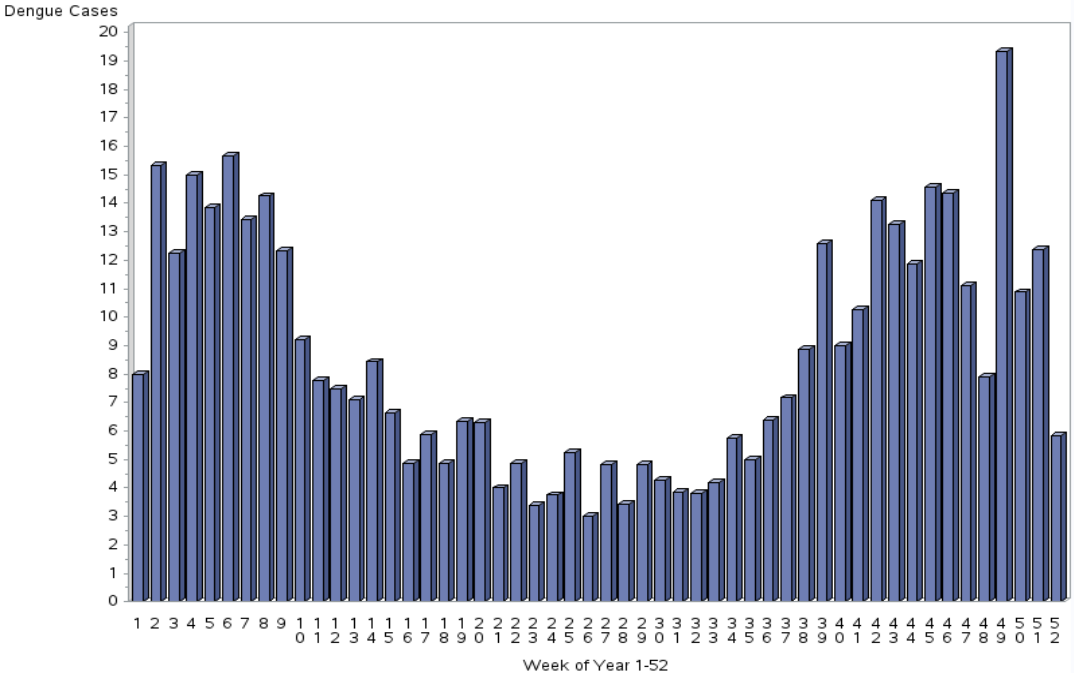



Figure 2 – Iquitos, also revealed bimodal distribution and associatiation of detected Dengue cases, with rainfall season and week of year but different highest and lowest means. First highest spike (means=16, 15, 14) weeks 2-8, and at second spike (means=14, 15, 49) weeks 38 – 49, having lowest (means=5, 3, 5) weeks 15 – 32.

**Figure 2. Variables Association: weekofyear and total_cases(means) - Iquitos (IQ)**




Based on observed weekly distribution we may conclude that number of Dengue cases is associated with rainfall season at specific geographical location. Because of slight difference in rainfall periods and it's duration, association of predictors with number of Dengue cases have to be analyzed separately for each location. Bases on climatological general climate observations Iquitos (IQ) almost does not have dry season with weekly precipitations (mm) much higher (mean=66, max=543) than San Juan (mean=26, max=163) and association of predictors with reported Dengue cases is affected.

**Bivariate analyses**

Figure 3 – San Juan N=724. Reveals positive Pearson correlation between total number of Dengue Cases and selected predictors showing that number of Dengue cases had significant correlation when following quantitative predictors increased: specific humidity (Pearson r=0.28, p< .0001), dew point temperature (Pearson r=0.27, p< .0001), station average temperature (Pearson r=0.22, p< .0001) and corresponding minimum temperature (Pearson r=0.21, p< .0001), maximum temperature (Pearson r=0.17, p< .0001).
Figure 4 – Iquitos N=387, also reveals positive Pearson correlation between total number of Dengue Cases and selected predictors, that number of Dengue cases had significant correlation when following quantitative predictors increased: specific humidity (Pearson r=0.18, p< .0003), dew point temperature (Pearson r=0.18, p< .0004), and corresponding minimum temperature (Pearson r=0.16, p< .0008).
Based on observed statistics, San Juan data correlates better with selected predictors than Iquitos data, but both locations have common predictors: specific humidity with strongest correlation coefficient ( SJ: r=30, IQ: r=0.18), station minimum temperature and mean dew point temperature and corresponding predictors p-values confirm significant correlation with weekly Dengue cases.

Figure 3. – San Juan

| | total_cases |
|---|---|
| total_cases<br>Dengue Cases | 1.00000 |
| weekofyear<br>Week of Year 1-52 | 0.28779<br><.0001 |
| station_precip_mm<br>Weekly precipitation mm | 0.07090<br>0.0566 |
| specific_humidity_g_per_kg<br>Mean specific humidity | 0.27945<br><.0001 |
| vegitation_index_avg<br>Satelite Vegitation index average | 0.04610<br>0.2153 |
| station_max_temp_c<br>Maximum temperature Celsius | 0.17083<br><.0001 |
| station_min_temp_c<br>Minimum temperature Celsius | 0.21470<br><.0001 |
| station_avg_temp_c<br>Average temperature Celsius | 0.22418<br><.0001 |
| dew_point_temp_k<br>Mean dew point temperature Kelvins | 0.27116<br><.0001 |

Figure 4. - Iquitos

| | _cases |
|---|---|
| total_cases<br>Dengue Cases | 1.00000 |
| weekofyear<br>Week of Year 1-52 | 0.01232<br>0.8091 |
| station_precip_mm<br>Weekly precipitation mm | 0.04460<br>0.3816 |
| specific_humidity_g_per_kg<br>Mean specific humidity | 0.18257<br>0.0003 |
| vegitation_index_avg<br>Satelite Vegitation index average | -0.02462<br>0.6292 |
| station_max_temp_c<br>Maximum temperature Celsius | 0.08794<br>0.0840 |
| station_min_temp_c<br>Minimum temperature Celsius | 0.16962<br>0.0008 |
| station_avg_temp_c<br>Average temperature Celsius | 0.07661<br>0.1325 |
| dew_point_temp_k<br>Mean dew point temperature Kelvins | 0.17792<br>0.0004 |

The analysis of variance (ANOVA) indicated that total weekly Dengue cases between San Juan and Iquitos are differ significantly as a function of mean specific humidity in Figure 5 (F=66.88, p<. 0001) and mean dew point temperature in Figure 6 (F=4071, p< .0001), indicating that the model as a whole accounts for a significant portion of the variability in predictor variable.

**Figure 5: Association of Dengue Cases with mean specific humidity**

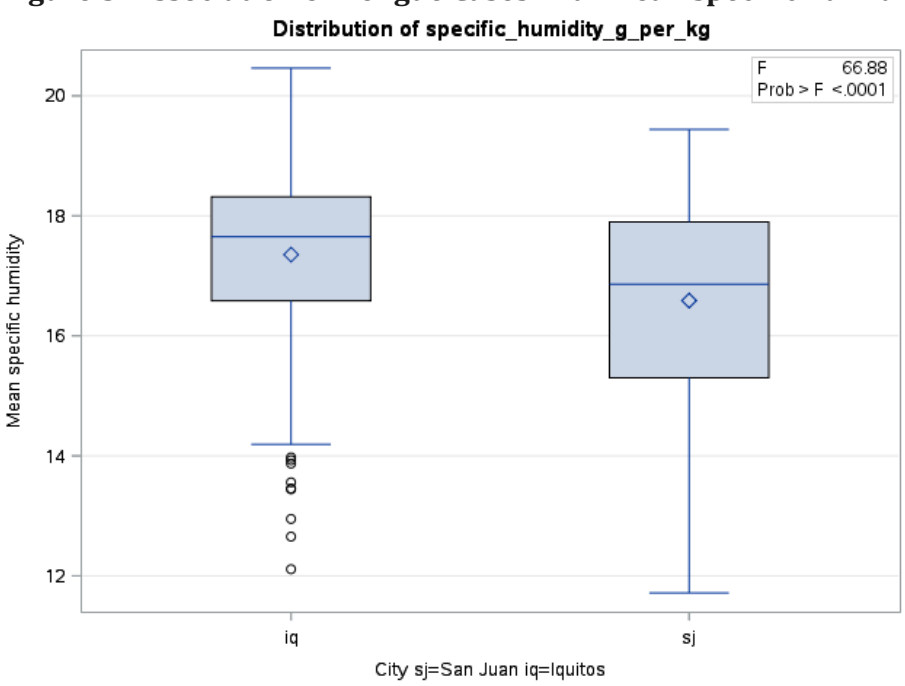
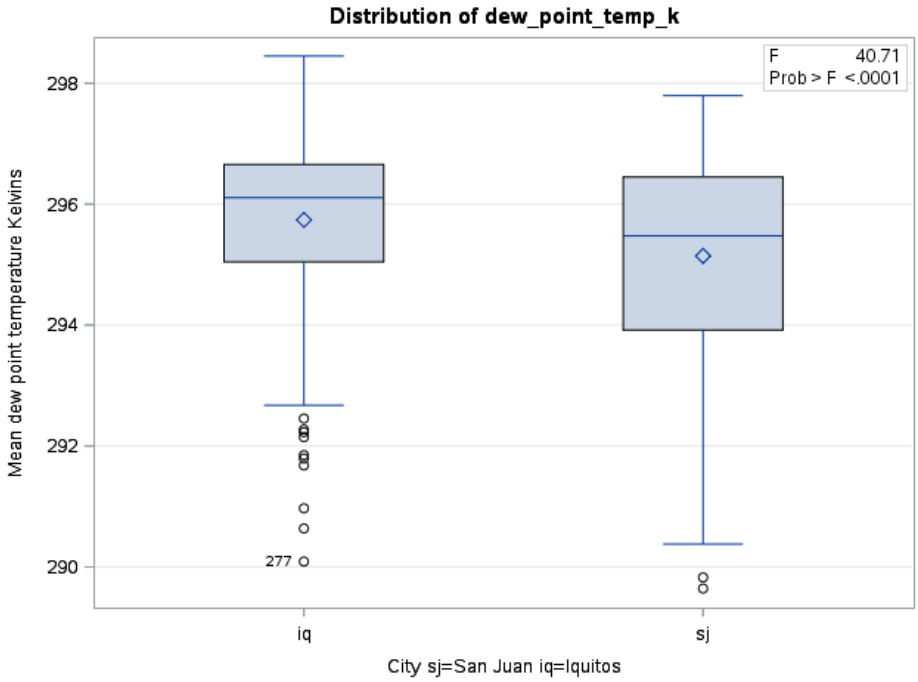



**Figure 6: Association of Dengue Cases with mean specific humidity**




## Multivariate Analyses

### Multiple regression analyses

Figure 7 - San Juan N=724, shows results of multiple regression analysis with stepwise selection. Based on displayed statistics, only mean specific humidity (F=61.16, p< .0001) and mean dew point temperature (F=16.24, p< .0001) are strongly associated with Dengue cases. Plots of residuals for the association of Dengue cases with mean specific humidity and mean dew point temperature also revealed that Dengue cases increased (290 highest) with increase of means specific humidity to its highest (18-19 Kelvins) and/or mean dew point temperature increased to its highest at about 296.5 Kelvins.

**Figure 7 – San Juan. Regression Stepwise selection of predictors**

Summary of Stepwise Selection

| Label | Number Vars In | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
|---|---|---|---|---|---|---|
| Mean specific humidity | 1 | 0.0781 | 0.0781 | 17.3331 | 61.16 | <.0001 |
| Mean dew point temperature Kelvins | 2 | 0.0203 | 0.0984 | 3.0958 | 16.24 | <.0001 |

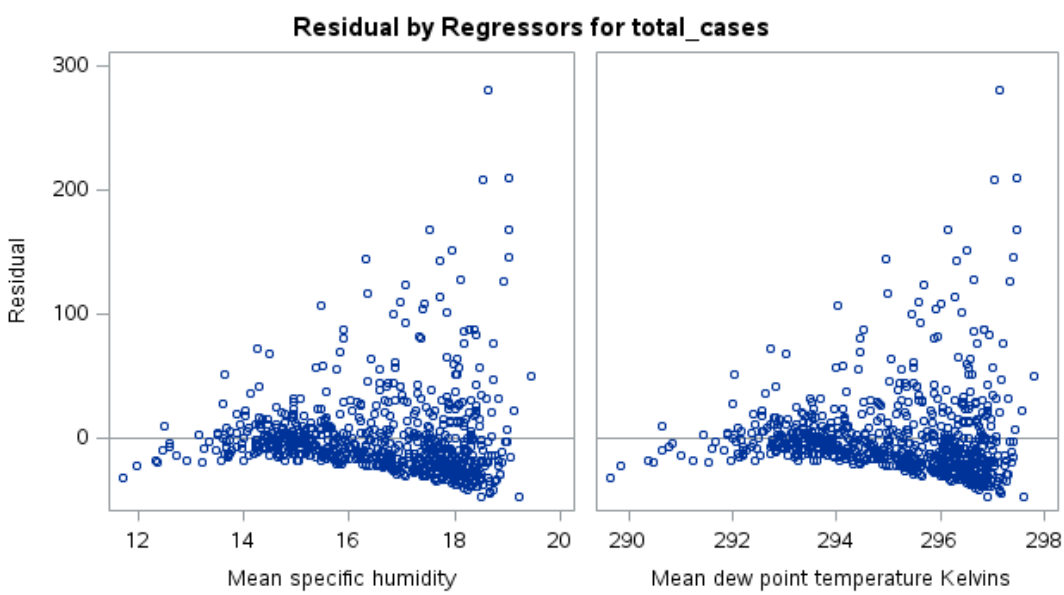


Figure 8 – Iquitos N=387, shows results of multiple regression analysis with stepwise selection. Based on displayed statistics, only mean specific humidity (F=13.28, p= .0003) is strongly positive associated with Dengue cases, selected minimum temperature (F=2.48, p=0.1161) does not have strong association with Dengue cases and will be removed from model. Plots of residuals for the association of Dengue cases with mean specific humidity also revealed that Dengue cases increased (65 highest) with increase of means specific humidity to its highest to about 18-19 Kelvins and dramatically reduced at 20 Kelvins. Separate regression test for association between Iquitos Dengue cases and mean specific humidity (Figure 9 – Iquitos model fit plot) proved strong positive association F=13.28, p= .0003 and most of residuals are in 95% confidence limits.

**Figure 8 – Iquitos. Regression Stepwise selection of predictors**

Summary of Stepwise Selection

| Label | Number Vars In | Partial R-Square | Model R-Square | C(p) | F Value | Pr > F |
|---|---|---|---|---|---|---|
| Mean specific humidity | 1 | 0.0333 | 0.0333 | 4.4119 | 13.28 | 0.0003 |
| Minimum temperature Celsius | 2 | 0.0062 | 0.0395 | 3.9251 | 2.48 | 0.1161 |

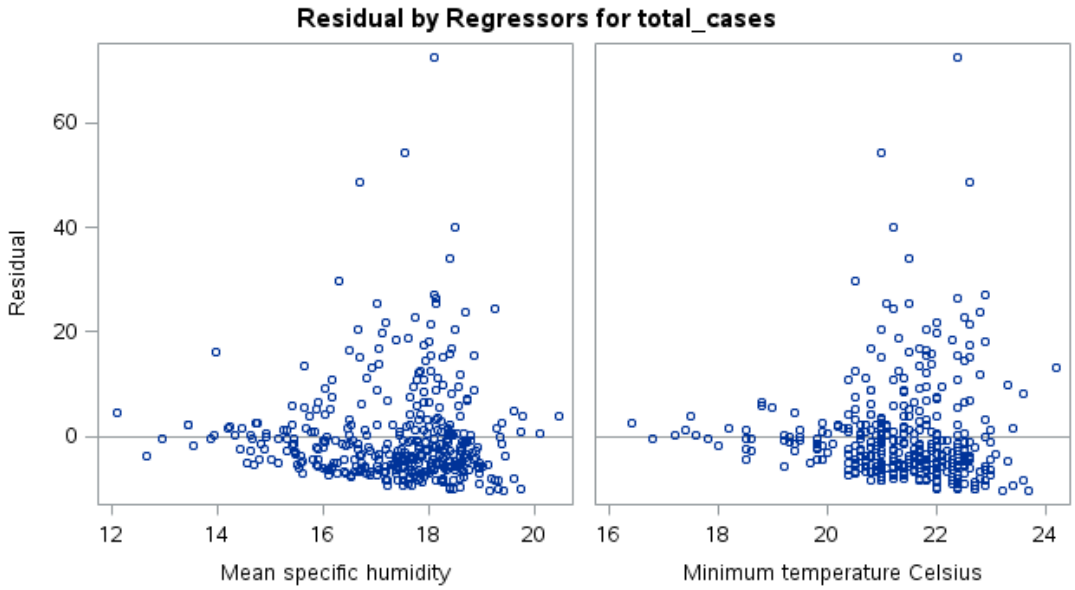



**Figure 9 – Iquitos Model fit plot**

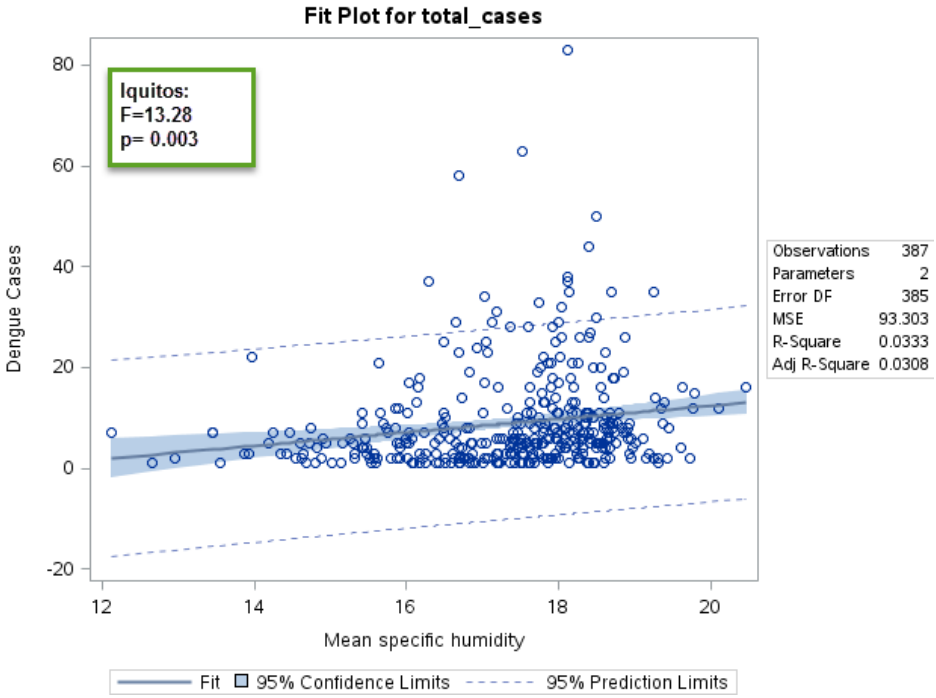


Figure 9a – Parameter estimates shoes that number of Dengue cases increasing by one point when mean specific humidity increased by 66.54 (San Juan) and 1.23 (Iquitos).

Figure 9a Parameter estimates.

**Parameter Estimates**

| Label (San Juan) | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Variance Inflation |
|---|---|---|---|---|---|---|
| Intercept | 1 | 16565 | 4130.45498 | 4.01 | <.0001 | 0 |
| Mean specific humidity | 1 | 66.54724 | 14.92332 | 4.46 | <.0001 | 328.31091 |
| Mean dew point temperature Kelvins | 1 | -59.76245 | 14.83205 | -4.03 | <.0001 | 328.31091 |

**Parameter Estimates**

| Label (Iquitos) | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Variance Inflation |
|---|---|---|---|---|---|---|
| Intercept | 1 | -14.06712 | 6.32652 | -2.22 | 0.0268 | 0 |
| Mean specific humidity | 1 | 1.32437 | 0.36348 | 3.64 | 0.0003 | 1.00000 |

**Lasso Regression analysis**

The result of Lasso regression analyses results for San Juan (Figure 10) and Iquitos (Figure 11) revealed that mean specific humidity is one best predictor of Dengue cases, which corresponds with performed analyses.
Figure 10 – San Juan, shows that only average temperature and dew point temperature dropped from model.
Figure 11 – Iquitos, shows that mean specific humidity, station minimum temperature, minimum temperature kelvins, satellite south-west, north-west, south-east vegetation index and air temperature in kelvins selected.

**Figure 10 – San Juan Lasso regression analysis**

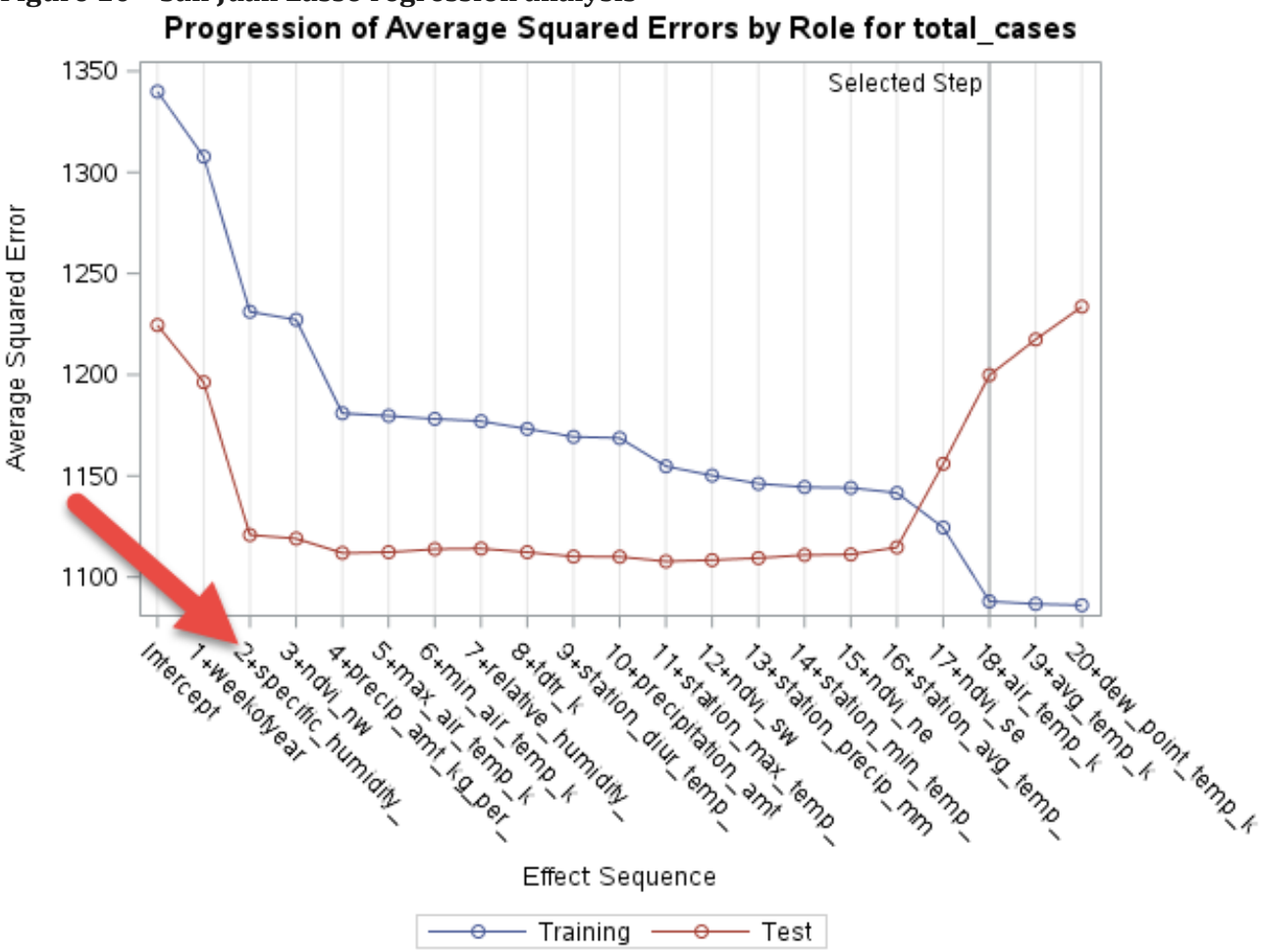
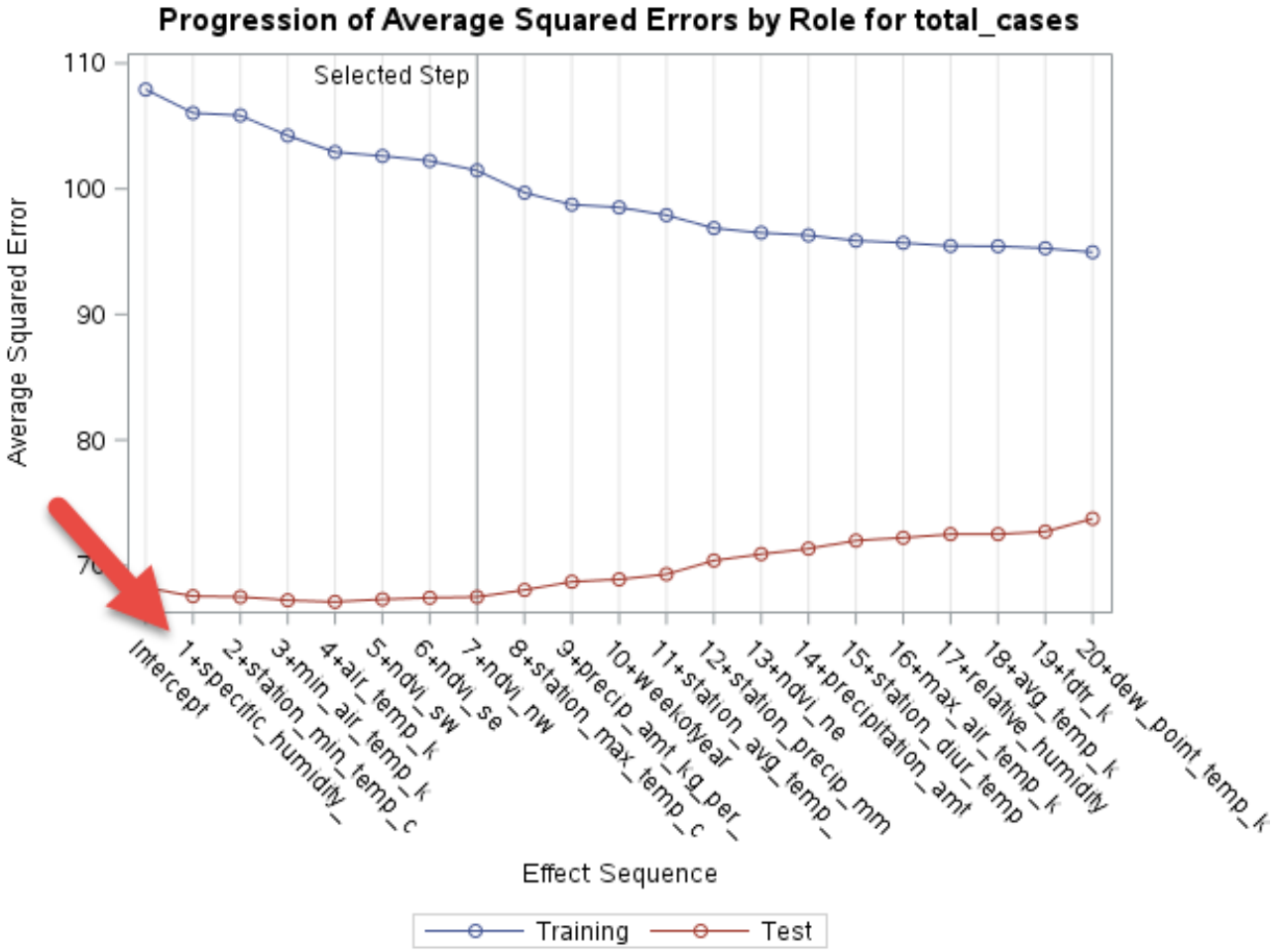


**Figure 11 – Iquitos Lasso regression analysis**




## Conclusions/Limitations

The purpose of this study was to identify best environmental predictors of Dengue disease in tropical and sub-tropical regions to identify common effects of environmental variables based on reported cases of Dengue disease in San Juan, Puerto Rico (N=724) and Iquitos, Peru (N=387)
This study shows that mean specific humidity parameter is one of the best predictors and number of Dengue disease cases increasing with increase of mean specific humidity (Figure 9a – Parameter estimates). The mean dew point temperature has also strong association with San Juan reported Dengue cases, but does not have such association with Iquitos data.

From effects suggested by Lasso regression only mean specific humidity and mean dew point temperature (San Juan only) have significant association with reported Dengue cases in both cases, testing different combination of suggested predictors degraded performance of main predictor mean specific humidity and mean dew point temperature.

This study successfully created predictive algorithm limited to predict missing or possible Dengue cases, but limited to mean specific humidity and mean dew point temperature and limited to provided factors. To create more robust and reliable algorithm, other. To increase prediction accuracy and to achieve best possible results, more distinct factors specific locations have to be added.