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

**Introduction to the Research Question**

The purpose of this study was to 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](https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/) weather station measurements), NOAA's NCEP ([Climate Forecast System Reanalysis](http://rda.ucar.edu/datasets/ds093.0/#metadata/detailed.html?_do=y) measurements), NOAA's [CDR (Normalized Difference Vegetation Index](https://www.ncdc.noaa.gov/cdr)), PERSIANN [satellite precipitation measurements](http://www.ncdc.noaa.gov/cdr/operationalcdrs.html).

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](https://en.wikipedia.org/wiki/Rainfall) throughout the year, without a distinct [dry season](https://en.wikipedia.org/wiki/Dry_season), but a wetter summer. The [rainy](https://en.wikipedia.org/wiki/Rainy) [summer](https://en.wikipedia.org/wiki/Wet_season) 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 )

A**nalysis**.

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 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 maximum Dengue cases of 83 (mean=9, std=10) and precipitation maximum of 543.3mm (mean=68.18, std=68.12).

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

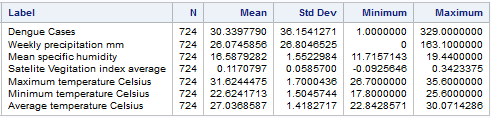
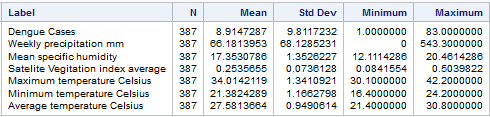


Table 2. Iquitos (IQ) - Weekly Descriptive statistics for reported Dengue Cases



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.

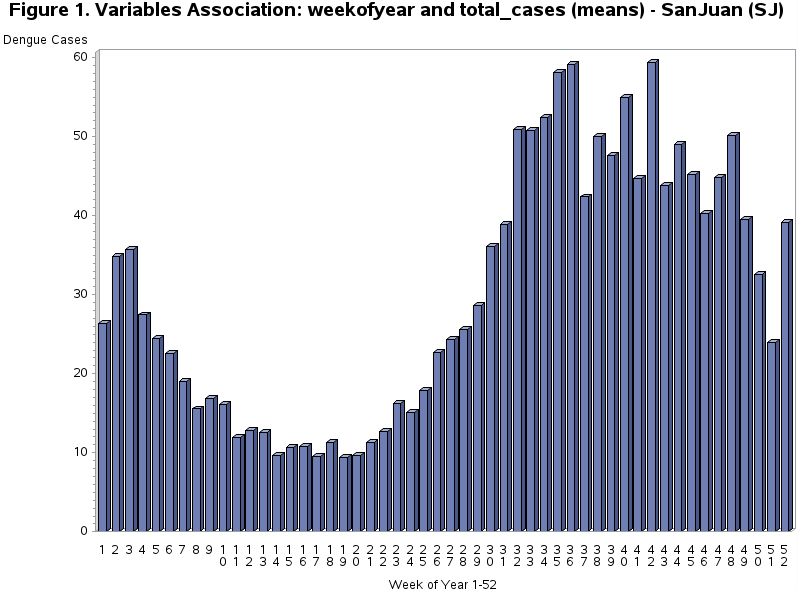
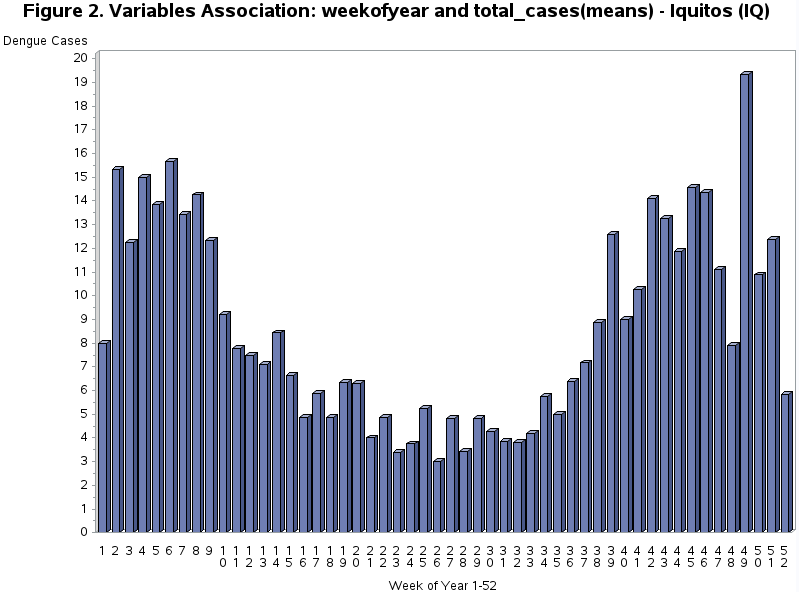


Figure 2 – Iquitos, also revealed bimodal distribution and associatiation of detected Dengue cases, with rainfall season and week of year but different highest and lowest means. First highest spike (means=16, 15, 14) weeks 2-8, and at second spike (means=14, 15, 49) weeks 38 – 49, having lowest (means=5,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



Based on observed distribution we may conclude that week of year can be a major predictor because it is associated with rainfall season at specific geographical location and data/factors of both locations have to be analyzed separately.

**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, station minimum temperature and mean dew point temperature and corresponding predictors p-values confirm significant correlation with weekly Dengue cases.

