

Land-use in the Amazon basin and its effect on vector-borne disease rates

Abstract

Vector-borne diseases are among the leading causes of death around the world. As health experts and officials continually battle these instances of infection, obstacles such as land-use may be among the leading causes of the ever-present perpetuation of these diseases. In this study, we look to identify the potential interactions between land-use, land types, and environmental affects and disease infection rates. We found that although there is no widespread generalization of the interactions between land-use and vector-borne diseases as a whole, we can identify potential indicators of increasing rates of disease for individual diseases.

Introduction

Background

Vector-borne diseases like malaria, dengue, and Zika are among some of the most infectious (and lethal) in the world; they account for around 17% of all infectious disease and cause over 700,000 global deaths annually. They can be spread to humans and animals via blood-feeding anthropods, such as mosquitos, ticks, fleas, and other parasites (see right). Curbing the spread of infection is a top priority for many health experts and officials around the globe, but obstacles often stand in the way of doing so. For example, in the Amazon basin, land-use change such as deforestation disturbs the natural habitats of these disease-bearing vectors, agitating them and potentially increasing the infection rate of those present at the site. In addition, the spread of disease can be facilitated and amplified through medical malpractice, such as the sharing of needles or untested blood for transfusions.

The following analysis serves as an initial exploration for a summer NSF EEID funded REU that a member of the group is participating in. As background, the REU investigates the effect of gold mining on the transmission of leishmaniasis (a



vector-borne disease) in the Peruvian Amazon. This study is especially important as these areas are less developed and have less access to medicine.

The details of the project can be seen [here](#). Research projects under the NSF EEID funding feature an intersection of geospatial datasets and statistical modeling in a computational biology and (disease) ecology setting.

Aims

We sought to explore the effects of land-use change on the dynamics of vector-borne diseases, including risk and transmission, based on data from a case study on land-use change in the Amazon basin (primarily Brazil, Colombia, and Peru). We sought to answer the following questions: 1) How do disease rates relate to precipitation for a given year? (*i.e.*, what, if any, is the strength and direction of the correlation between disease vector activity and amount of rain?) and 2) Which biomes are at a higher risk of disease incidence? After completing our initial exploratory analysis of the data, we decided to focus more intensely on certain variables and create more specified graphs based off of those variables.

To answer the first question we proposed, we created scatterplots comparing disease incidence estimates with precipitation rate. Our initial interpretations and reactions to these were that there did not appear to be any significant differences in the environment in which diseases thrived; upon further inspection, however, we determined that specific diseases tend to thrive in medium-wet environments. To answer our second question, we created line plots comparing median disease rates with different land types, as well as countries. Unlike with our first plots, we were able to initially identify evidence that certain diseases thrive in different land types.

Materials and methods

Datasets

The dataset was provided to our group by [Dr. Andrew MacDonald](#), one of the supervisors of a NSF EEID funded project (DEB-2011147). There are various sources for the data, depending on the variable. The case data are derived from reported cases from the population of municipalities and their hospitals and clinics. The rest are outlined in the Peru Data Column Key, but are the following:

- MODIS satellite imagery
- Landsat satellite imagery
- DMSP-OLS nighttime lights satellite imagery
- JRC Global Surface Water Mapping Data: Jean-Francois Pekel, Andrew Cottam, Noel Gorelick, Alan S. Belward, High-resolution mapping of global surface water and its long-term changes. Nature 540, 418-422(2016). doi:10.1038/nature20584
- MapBiomas land use and land cover mapping project

The data consists of disease estimates sampled from different year-municipality combinations, where each combination acts as its own observation and has different land-use properties. The municipalities are from Colombia, Peru, and Brazil and the years span from 2000-2019. The data itself was collected from reported cases of disease incidences from the population of these municipalities and their hospitals and clinics. Below is a complete table of the variable descriptions:

Name	Description	Type	Units of measurement
Code*	a unique identifier for each municipality	Numeric	N/A
Year*	year in which the measurements were taken	Numeric	N/A
Country*	country in which the measurements were taken	Character	N/A
Name*	name of the municipality	Character	N/A
Population*	population in the municipality in a given year	Numeric	people
Chikungunya:Zika*	disease incidence estimates for a given disease	Numeric	per 1000 people
LST	Land Surface Temperature	Numeric	degrees Celsius
OptTemp	optimal transmission temperature for each disease	Numeric	degrees Celsius
NDVI / EVI	vegetation index	Numeric	N/A
Precip*	precipitation rate	Numeric	inches per year
StableLights / AvgRad	light emission	Numeric	N/A

Name	Description	Type	Units of measurement
SWOccurrence	occurrence of standing water (water bodies)	Numeric	N/A
pland*	proportion of land	Numeric	N/A
te	configuration of the landscape, total edge	Numeric	N/A
area_mn	area of an average patch of a given land type	Numeric	N/A
enn_mn	average distance between patches of the same land type	Numeric	N/A

Given the large amount of variables and observations in the original dataset, we took a subset of the dataset based on the variables we deemed the most important (indicated with an asterisk (*) in the table above), as well as years with the most non-missing values. In our subsetting dataset, each row is an observation (a specific municipality/year combination) and are observing at all municipalities (~8,500) in Colombia, Peru, and Brazil in 2010-2017. The relevant population in our case is all locations globally in which both land-use changes and vector-borne diseases exist. The sampling frame and the sample are the same: data from all municipalities in Colombia, Peru, and Brazil over 8 years (2010-2017). As such, the sampling mechanism is most likely administrative data. While no reliable methods of extrapolation exist, perhaps we are able to generalize the findings of our analysis based on particular properties of the land, such as land type or precipitation rate.

Below are some example rows of our completed tidy dataset:

Code	Year	Country	Name	Population	Chikungunya	Cutaneous.Leishmaniasis	Dengue	Mal
5001	2010	Colombia	MEDELLIN	1937458	NaN	0.003097	8.036303	0.018
5001	2011	Colombia	MEDELLIN	1960132	NaN	0.003061	0.409156	0.005
5001	2012	Colombia	MEDELLIN	2019943	NaN	0.001485	0.352980	0.005
5001	2013	Colombia	MEDELLIN	2075273	NaN	0.001927	1.128526	0.004
5001	2014	Colombia	MEDELLIN	2117947	0.032579	0.002833	1.545364	0.002

Methods

For each question, we took the original tidied data set and melted it several times for the purpose of plotting the variables of interest using Altair. In both questions, we followed the popular method of taking the median of the variable of interest of each group we were interested in. For example, we took the median estimated disease rate of each disease and land type for the second question. To quantify the relationship between disease rate and land type, we only focused on observations with a significant proportion of each land type. That is, we subsetting each melted data frame so that only observations whose proportion of land type is greater than the median proportion of each land type then plotted a line plot of each line plot by median rate of disease. Similarly for the first question, we took a scatter plot of precipitation levels by rate of each disease

Results

Initial exploratory analysis

In our initial investigation with our subsetting dataframe, we created several plots that we often referenced in our analysis.

Below is a correlation heatmap of the variables in our subsetting dataframe. This figure allows us to see which the strength and direction of the correlation between any two of the variables.

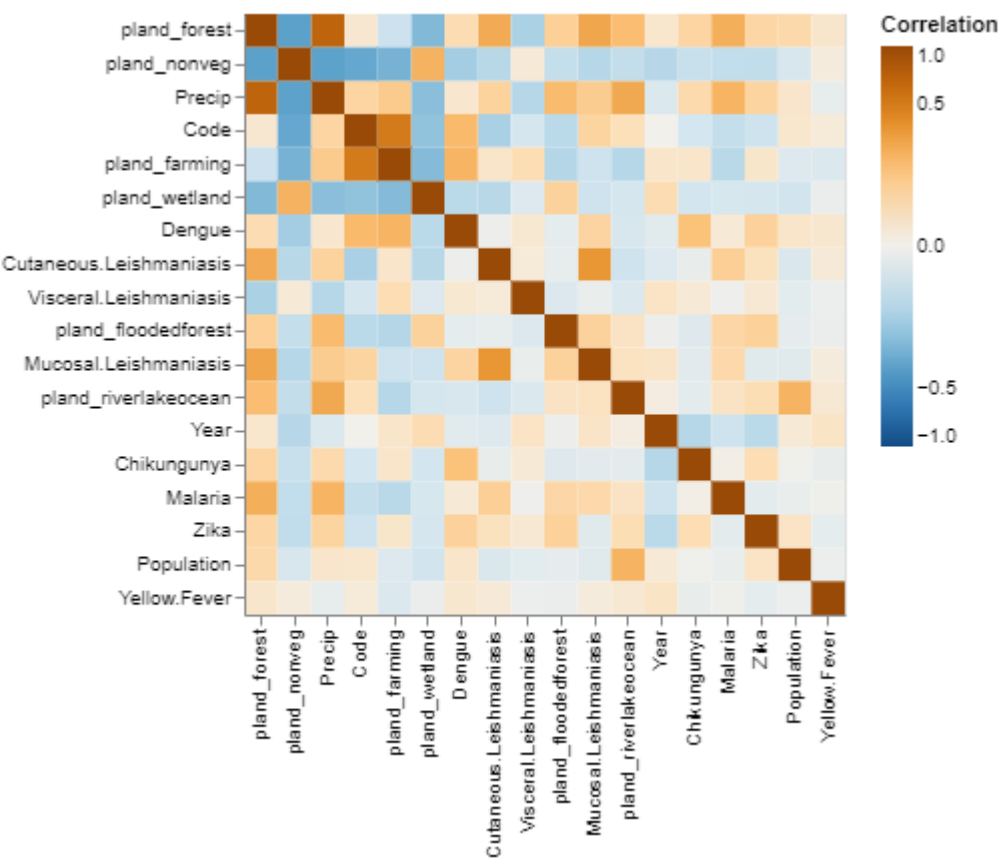


Fig. 1: correlation heatmap of all the relevant variables

Below are faceted plots of Malaria rates vs. precipitation. We often referenced this figure when identifying potential patterns in disease rate interactions with environmental factors.

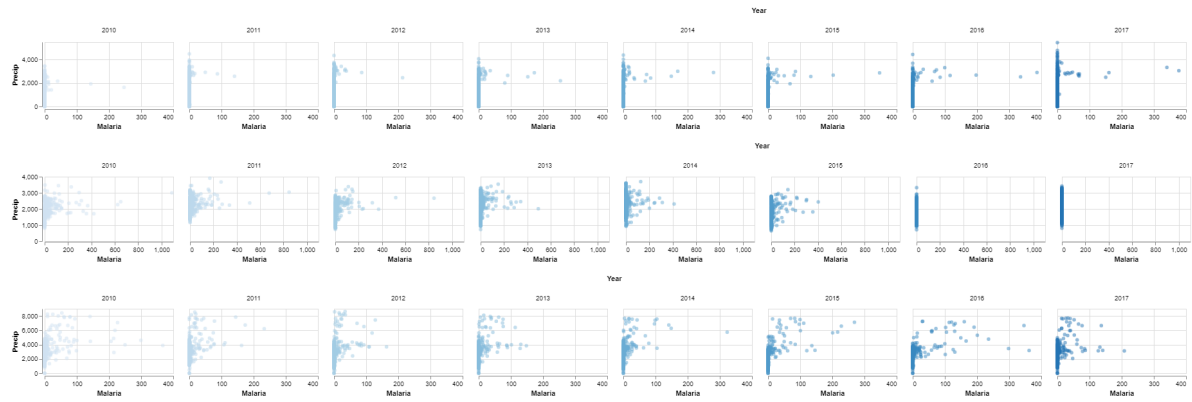


Fig. 2: faceted plots (by country and year) of Malaria rates vs. precipitation

Disease and precipitation rates

We created the following graph to help explore the potential relationship between disease rates and precipitation. Displayed below, this figure shows the individual disease rates for each observation vs. the corresponding precipitation rate for that municipality/year combination.

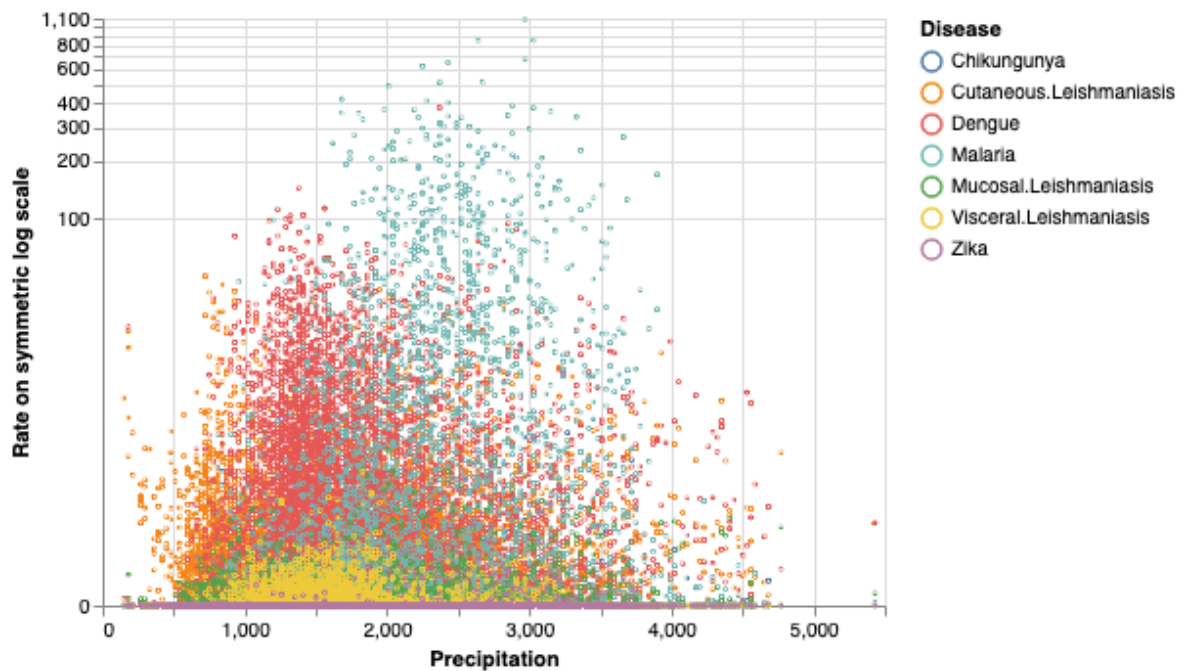
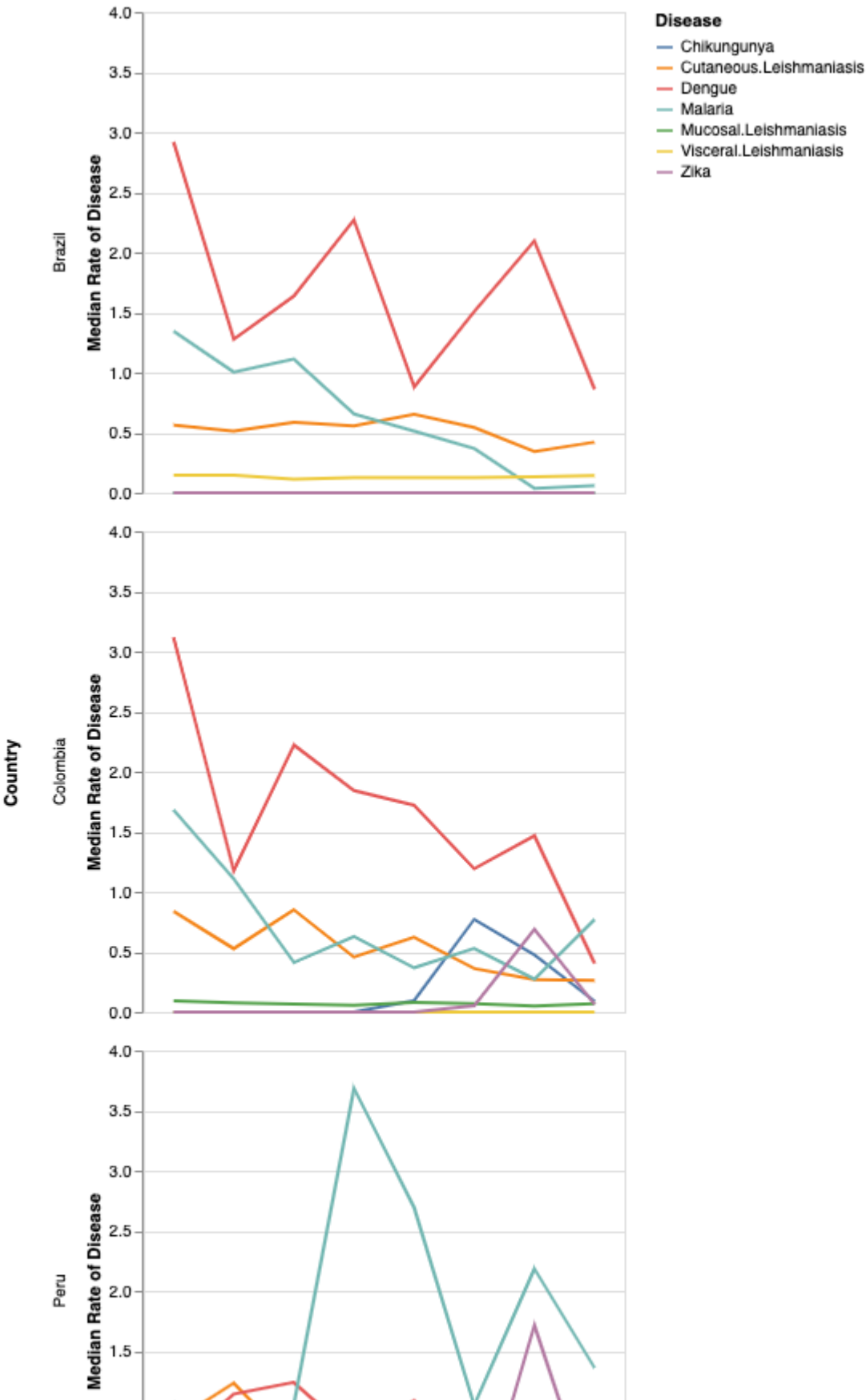


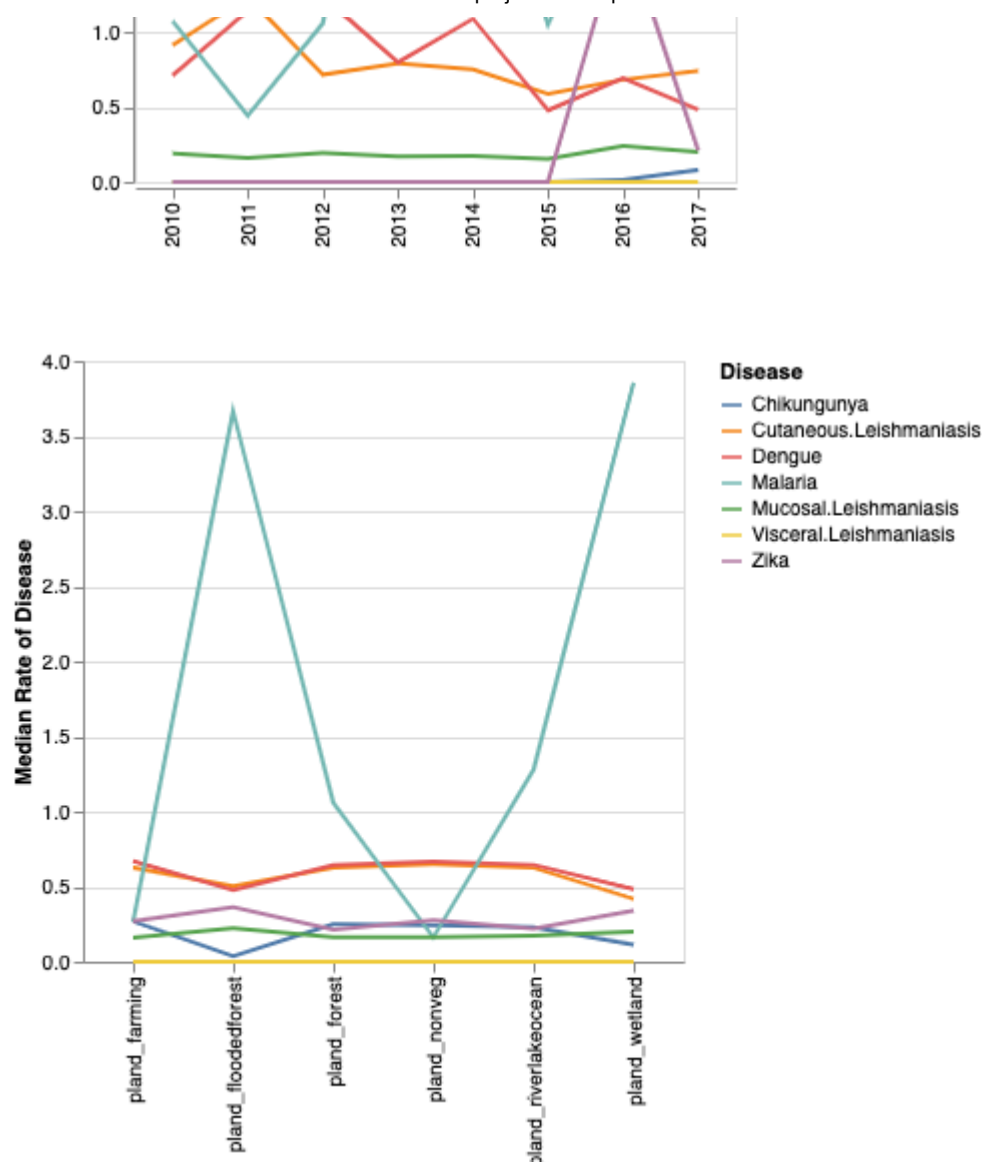
Fig. 3: scatterplot comparing disease rates (on a log scale) and precipitation rates

This graph provides evidence of precipitation potentially influencing individual disease rates. For example, Malaria was highly present in relatively high levels of precipitation. Additionally, Dengue was slightly prevalent in mid levels of precipitation. We do, however, see the opposite with other diseases: Leishmaniasis did not appear to be too dependent on precipitation levels. These findings are also supported by Fig. 1; the 'Precip' and 'Malaria' variables have a slightly strong positive correlation, while the three types of Leishmaniasis as a collective are balanced between positive and negative correlations with precipitation.

Biomes at risk

We created the following two graphs to help explore the potential relationship between land types and disease rates. Displayed below, the first three plots show the disease rates as time continues (from 2010 to 2017) in each country in our study. Contrastingly, the plot on the right shows disease rates vs. land type.





Figs. 4 and 5: line plots depicting the interaction between disease rates and land types/countries

From these graphs, we found that Malaria stuck out as a primary vector-borne disease in the Amazon basin. It seemed to be most prevalent in land with significant flooded forest area and wetland area. Visceral Leishmaniasis is almost nonexistent in the above areas, while Dengue and Cutaneous Leishmaniasis were pretty steady and prevalent in each area. On the other hand, we found that there were strong decreasing rates of disease from 2010 to 2017 in Brazil and Colombia, and slight decreasing rates of disease in Peru. This is possibly indicative of increasing quality of healthcare in these countries, or even a growing population resilience to these diseases.

Discussion

The results for the first question make less sense than the second, as we will see. Dengue is said to thrive in heavy precipitation, while malaria thrives in moderate precipitation. However, the scatter plot suggests the opposite. Perhaps the population of each observation is key in explaining this phenomenon and is a potential point of further investigation. Another point of contention may be the primary carrier of the disease. Which animals primarily hold Malaria, and which primarily hold Dengue? Does this have any impact on whether precipitation truly acts as an indicator for disease rates of certain diseases?

On the other hand, the results for the second question are interesting and make slightly more sense since Malaria is spread by mosquitos and standing bodies of water are breeding grounds for mosquitos. However, dengue thrives in heavy precipitation, so it is weird how rates of disease for dengue aren't as maliciously high as malaria in each land type. Perhaps it is the means of spread for each disease that is responsible for the disparity. This may be another topic to explore.