The Zika Virus

Richa Deshpande (rd382), Il Kwon (ik226), Harini Kannan (hk473)

Because of the recent media coverage on the Zika virus we decided we wanted to investigate it. Our visualization represents six different data sets, each with particular information based on the countries, number of cases, and economic factors.

South America:

Datasets:

We collected several data sets in order to create displays of various comparisons across South American countries.

confirmed.json, suspected.json
The first data set we collected was the confirmed and suspected number of zika virus outbreaks for a list of South American countries. This data was found from http://ais.paho.org/phip/viz/ed_zika_epicurve.asp. This data was pulled in by hand from the information in the bar graphs. Because the data was over months, we decided to pull the confirmed and suspected cases for each country for the most recent months. Note: If the suspected number is 0 and the confirmed number is > 0 that means that all the cases were confirmed zika virus cases.

economicStats.csv
Ideally, we wanted to assess the statistics of the patients, such as income, where they were from, and other socio-economic factors. However, because the outbreak is so new we do not have patient statistics. Instead, we found the socioeconomic factors data on the countries themselves and compared them.

For these countries, we wanted to assess different economic factors with which to compare the countries. The data for these economic factors was found on http://databank.worldbank.org/data/home.aspx, http://www.tradingeconomics.com/ , and http://knoema.com ,which for each country lists the numbers for different economic factors. Because we just wanted to get the most recent numbers for the socio-economic factors, we had to go to different sources of data because often one site would have the most recent data while the others didn't. One of the challenges we faced was that data on the socio-economic factors on some of the latin american countries was not there. This translated into filtering our data based on the availability of these socioeconomic

factors. Therefore, a compromise we made was choosing the more prominent countries with economic information that was available.

The variables for this part of the data set are the following:
States: Weirdly labelled, these are the countries themselves
GDPPC: This is the GDP Per Capita for each country in USD
Poverty headcount ratio: This is the headcount at National Poverty Lines (% of Population)
Births Attended by Skilled Staff (%): This is the percent of women who gave birth who did so by a skilled staff
Prenatal Care: This is the percent of pregnant women who were given some sort of prenatal care
Mortality: These are the mortality rates for infants (per 1000 births)

Mosquito.json
Mosquito is the main vector of zika transmission. There are two species of mosquitos that are subject to the transmission, Aedes aegypti and Ae.albopictus. The data which has been retrieved from Nature.com (http://www.nature.com/articles/sdata201535#f3) contains occurrences of these two species worldwide. We treated both species equally since distinguishing is not necessary. However, we filtered only those occur within the Americas.

Focusing on Brazil:

monthlyCases.csv:
We also collected data on Brazil specifically to analyze the spread of Zika through this country. We chose Brazil because it one of the countries that has been given more attention by the media because of the rapid spread of the virus in the past few months. The data came from a Brazilian Medical websource. We also painstakingly searched for and rendered a JSON file for the display of Brazil and its states. With the vast amount of data provided by the medical websource, we decided to break up the suspected cases per state by months to show the spreading of Zika over time. The variables included the States, corresponding months, and number of suspected cases. We also assigned each state and ID number because the JSON file did not contain information on which state was being graphed. Theses variables were organized in a .csv to be used for the second half of the visualization. We had to reform some of this data because during some months, some states reported very few, if any new suspected cases. We used this as an advantage to compare the states that did not have many new reported cases to those that did.

brasilregions.csv:
Here we introduced the bar graph comparing the GDP of each state to the others. This way the user can analyze which states frequently had many suspected cases as compared to their GPD per capita. The GDP data was found on Wikipedia. In general, we collected the data for each month by looking at the reported suspected cases at the last week of that month. This was organized into a CSV and the GDP data was a separate csv file used to create the bar graph.

The variables included:
States: A list of the current infected states in each iteration.
Month: The current month on that iteration.
Cases: Number of cases for each State in States, zero if not in States.
ID: Special ID to identify which state was being mapped.

Visual Elements:
The visual elements in this part of the visualization include the map of Brazil states, and GDP graph. We integrated the two visualizations together by highlighting the bars in the GDP graph for the states which had a significant number of suspected cases that month. This way the user is able to analyze how the GDP of states frequently having suspected cases relates. The slider allows the user to move through the months to see the spread of Zika from the rightmost state Parabai to eventually all of Brazil in March 2016. As the user slides through the months, the bars of the GDP highlight according to which states have suspected cases for that month.

Brazil Takeaway:
From the higher level, this part of the visualization shows how quickly Zika has spread in the past few months, especially in Brazil. We also see how geography played a role in spread as the nearby affected states caught the virus usually in the next following month. Also, states that were smaller, and therefore more condensed usually had more suspected cases. By integrating the map with the GDP bar graph, we see that states with lower GDP are among the first to catch the virus as well as have a higher number of suspected cases each month.

Overall Takeaway:
From the data it is clear that there is a heavy correlation between the outbreaks and the regions with the most mosquitos. In terms of the economic factors, the data shows that there is not a heavy correlation between GDP per Capita and poverty of a country and the number of confirmed cases. At first, we anticipated a correlation given a heavy

association between the two in the media, however it doesn't seem to be strong in the data. Additionally, we learned factors such as births by a skilled health staff and pregnant women receiving prenatal care were not as severe as the media also portrays. For most of the countries, the values for both those economic metrics were in the high 90s.

When we looked at the progression of suspected cases in Brazil, it is definitely apparent that regardless of whether economic factors have an impact on the country that the Zika Virus is a growing epidemic. From only a couple affected states in Brazil last year, every state has now been afflicted.