Milestone 2 – Group 16

**Motivation of the Idea**

Local mosquito-borne Zika virus transmission and Zika virus outbreaks have been reported in the United States and in United States territories. We want to provide a platform where users can understand their Zika risk, from where they live, their travel history, the travel history of their family by having access to a visual map that contains the total number of cases, the rate of cases per week, CDC travel notices, and a rough number of Zika in the future in a given region.

**Features that will definitely be implemented in the application**

* Interactive D3.js map where users can click on a US state or territory to see the total number of cases, rate of cases/month, and the predicted number of cases in the future
* Have a sidebar that links to: (stored in NoSql)
  + CDC website for more information about Zika virus
  + CDC website for more information about Zika virus symptoms
  + CDC website for more information about how it’s contracted
  + Infographic about Zika transmission

**Features that might implemented in the application, given enough time**

* Twitter feed on the bottom of the website that collects tweets that have the hashtag “zikavirus”
* Login feature

**Technology and tools to be used**

NoSQL – MongoDB – for pictures, videos, links and other nontabular data, query data live through Wikipedia

MySQL

Meanstack to communicate between NoSQL and MySQl

Node.js

JavaScript – make actual website

Python to put data in usable format

Amazon Web Services - hosting database and deploying application

**Description of the complimentary sources you intend to use for data, and how you intend to ingest the data into your database**

Sources for the data

CDC Github - Zika total case count by state and US territory (csv)

Manual data scraping from climate.gov - State average 12-month temperature data (xlsx)

Manual data scraping from <http://www.latlong.net/> - Latitude and longitude of states (xlsx)

Predicted temperature data – still have to run the simulation in MATLAB to get the data (csv)

* Still working on getting predicated temperature
* Temperatures expected to increase by 1 degree in all states
* So the latitude and longitude of where the mosquitos can go will change (can travel farther north)
* Base on count data we have from the states we will probably run a regression using lat/long data, total case count data, Zika case rate data to get a prediction of the number of cases in the state that currently does not have the Zika virus

CDC infographic – pdf

* <https://www.cdc.gov/zika/pdfs/zika-transmission-infographic.pdf>
* <https://www.cdc.gov/zika/pdfs/flu_or_zika_infographic.pdf>

We will ingest the data into the database by using python or java cursors to insert the data.

**Division of responsibilities**

Collect data – Frankie

* Development: Frankie
* Check if complete: Xilei

Exploratory data analysis - All

Develop with MeanStack (Angular, Express, MongoDB, Node.js)

* Development: All
* Check if complete:

Design user interface

* Development: All
* Check if complete:  All

Get data into usable format

* Development:
* Check if complete:

Write SQL and NoSQL queries - Lanqing, Frankie

* Development: Frankie
* Check if complete:  Lanqing

Design schema - Xilei, Frankie - meet Wednesday talk to TA on Friday

* Development: Frankie, Xilei
* Check if complete:  Lanqing

Put data into SQL and NoSQL

* Development: Frankie
* Check if complete:  Lanqing

Set up website – Xilei, Lanqing

* Development: Lanqing
* Check if complete: Xilei, Will

Data Visualization for website (D3)

* Development: Xilei
* Check if complete: Frankie, Lanqing

Make sure server/stack works

* Development: Will
* Check if complete:  Lanqing

**SQL DDL**

CREATE TABLE State (

)

CREATE TABLE CaseRate (

)

CREATE TABLE CaseCount (

)

CREATE TABLE CasePredict(

)

**NoSql – MongoDB**

Collections : notices, links

**ER Diagram**