# Methods

## Epidemiological data

We obtained Brazil’s weekly cases from Brazil’s Health Ministry (Secretaria de Vigilância em Saúde, Ministério da Saude), which are available at: <http://portalms.saude.gov.br/boletins-epidemiologicos>

Mexico’s weekly cases were retrieved from Mexico’s Official Epidemiological Office (Dirección General de Epidemiología), which are available at: <https://www.gob.mx/salud/acciones-y-programas/direccion-general-de-epidemiologia-boletin-epidemiologico>

Colombia’s weekly cases were retrieved from Colombia’s National Health Institute (Instituto Nacional de Salud), which are available at: <http://www.ins.gov.co/buscador-eventos/BoletinEpidemiologico/Forms/AllItems.aspx>

Brazil population data: <https://cidades.ibge.gov.br/brasil/panorama>

Colombia

Mexico population data: <http://www.beta.inegi.org.mx/temas/estructura/>

## Google Search Queries

Since wanted to develop a model that would be able to be trained on one country and tested using another country, we needed to select a search term that would be general between countries, as different keywords may introduce a different bias depending on the country.

We decided to select “zika” as our only search term, as it is the most general term used between countries, we also assumed that any other zika related terms, such as “zika symptoms” and “zika virus”, are reflected on the more general “zika” search index.

To keep the model as general as possible, we decided to use

|  |  |  |  |
| --- | --- | --- | --- |
|  | Brazil | Colombia | Mexico |
| Cumulative Cases | 233659 |  |  |
| Population | 207660929 (2017) | 48202617 (2015) | 119938473 |