**Exercise 2.** Estimating Zika case counts in Colombia during the 2015-2016 outbreak

**Data sources:**

1. Healthmap feed of Zika related news alerts associated to the word Zika (with time stamps)

File name: *HM\_Zika\_04282016.csv*

1. Search fraction of the word “Zika” in Google search engine

File name: *Zika\_Colombia\_GT\_Apr29\_2016.csv*

**Suggestions on how to proceed:**

1. Download data cut file for **Healthmap** news alerts (link here)
2. Build a time line of cumulative cases as reported by the news alerts (you may need to go directly to the websites and obtain the reported number of cumulative cases)
3. Once you have a time line of cumulative cases per week, download the time series of search fractions for the word “Zika” from the tool “Google trends”.
4. Plot this time series.
5. Assuming that the Zika activity (number of weekly cases) is proportional to the search activity on the word “Zika”, construct a cumulative “curve” of search activity over the time period of interest. (May 2015 – April 2016).
6. Find the appropriate proportionality constant so that the cumulative search activity matches the cumulative number of cases as reported in the Healthmap news.
7. Multiply the time series of cumulative search activity by this constant.

You have now constructed an estimate of cumulative Zika cases with two non-traditional data sources.

**Validation**

1. Finally, compare your results with the number of cases as reported by PAHO.

File name: *PAHO\_Colombia\_Zika\_2016.csv*

Know that these values were only available at least a couple of months after the fact.

1. Plot all the cumulative curves in one figure.