# Using Benford's Law to Analyze Covid-19 Data

With cases of Covid-19 rising everywhere, and so many lives lost to the pandemic, one might ask if the data for confirmed cases or deaths we are provided are accurate. Do we have cases of countries that are consistently manipulating the data in order to show that they are on top of the pandemic? Can we trust these data in order to forecast cases in near future or provide necessary assistance?

In this project, we will use a dataset from John Hopkins University to assess the veracity or accuracy  of covid-19 data as provided per country. To do that we will use Benford's law that stipulates that  frequencies of first digits in naturally occurring numerical distributions, with multiple changes of magnitude, follow a particular logarithmic distribution. Since the pandemic was a natural occurrence, the number that we receive would reflect that. Benford’s law shows that in such a case, the leading digit 1 occurs about 30.1% of the time, the leading digit 2 occurs about 17.6% of the time, etc with the leading digit 9 occurring lastly about 4.6% of the time.

To check the accuracy of the data we will use a chi-square test with 95% confidence level. The test will  show how close observed values in our data  are with respect to the expected values provided by the Benford distribution. We will assume that the values are pretty close and hence the data accurate, if the chi-square statistic calculated is less than or equal to the critical chi-square value that is obtained by taking into account the degree of freedom of our data .