

Meaningful visualization of data

Niko Sauer

Centre for the Advancement of Scholarship
University of Pretoria

April 30, 2020

1 Background

In the media data concerning the development of the SARS-CoV-2 epidemic is visually presented in ways that confound meaningful interpretation. For example, absolute numbers are presented which may be sensational, but is not compared to any meaningful standard. Hypothetically, 100 infections per million of a population, appears to be mild compared to 5000 infections in a population of 50 million. On the other hand, the milder figure could indicate the sensitivity of, say, the economy to unusual phenomena.

Data such as number of infections and number of tests are presented in a cumulative way, perhaps in the hope of deducing in some way the rate at which the infection is spreading. This rate is the concern of the present note.

2 Rate of increase

To obtain an idea of the rate of increase, the reported daily number of new infections is helpful, but also misleading. The reason for that is that the number of tests vary from day to day. It is reasonable to argue that an increase in the number of tests is likely to increase the number of positive identifications. We should be more concerned with the ratio: *new cases to number of new tests* which may be given a probabilistic meaning. The data for such a calculation can be derived from published data. The ratio is quite small. The suggested index is *number of new cases per 1000 tests* for the same day.

If presented as a time series derived from observations, the suggested index turns out to be quite noisy. On the many possible reasons for the observed noise we shall not speculate here. The important point is that it is difficult to form an idea of the rate of increase from “jumpy” figures. A way to approach this difficulty would be to use moving averages of noisy data. In Fig.1 below moving averages with equal weight were used over

consecutive time intervals of 5 days and the average plotted at the middle of the interval. The plot was then slightly smoothed by fitting a Bézier curve to the calculated points. Of course, the first two and the last two observations would be absent. Gaps in the bar chart indicate missing data.

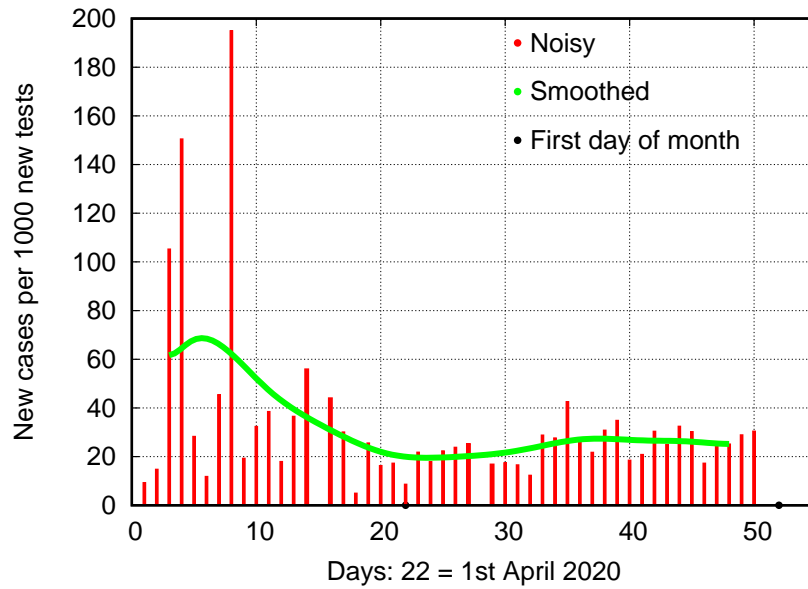


Figure 1: Growth

When reliable data on recoveries become available it could potentially be used to give a different picture.

3 Sources

Data can be sourced from the National Institute for Communicable Diseases: [NICD](#) — updates.

Two websites where “new cases” are presented in a way similar to the description above are [WITS](#) and [UCT](#).