Abstract

The aim of this project is to analyse the effect of the Covid-19 vaccination on the number of positive Covid-19 cases, as well as its effect on the number of hospitalizations in Ontario.. All data used was from `https://data.ontario.ca/dataset/`. The available vaccination data begins from January 2021. Routine VOC PCR testing of positive Covid-19 samples ceased on November 12th 2021 [1], and so to compare the effect of vaccinations to positive Covid-19 cases, we use analyse data from 1st January 2021 to 31st October 2021. The effect to be analysed is whether administration of the Covid-19 vaccination led to a reduction of positive Covid-19 cases, as well as to analyse if administration of the vaccine reduced the number of intensive care unit hospitalizations caused by Covid-19 (in Ontario). As the effect of the vaccination is being analysed, rather than using data of total number of first doses administered per day, we instead want to look at total number of people who have received their first dose of vaccination for least 14 days ago. For this reason, the vaccination data will be shifted by 14 days. As a cross-correlation accounts for shifts in time-series being compared, for completeness sake a cross-correlation will be taken between the positive Covid-19 cases data and unshifted vaccination data, to see where on the time-lag access the strongest correlation is.

Data Acquisition and Filtering

The data from the data.ontario website was downloaded and plotted as discrete time series. Before using techniques to compare different data sets (i.e. to see correlation between vaccine doses administered and positive Covid-19 cases etc), the data was filtered using techniques used in class, to remove any unwanted noise. In order to filter the data, techniques from lab 3 were used, where our original time-series was de-trended by removing a numpy.polyfit line (of sufficiently high degree, at least degree 7).The de-trended data was Fourier Transformed (using Numpy's built in FFT), with the Fourier spectrum in the frequency domain set to zero for high frequencies. This filtered detrended time-series was then inverse Fourier transformed back to the time domain (using Numpy fft.ifft), with the original trend finally added back in. Initially, filtering this way led to a bit of an unexpected effect; an unnatural uptick was introduced to the end of the data. It appears that this uptick was due to the periodic nature of the inverse fourier transform, leading to the end values of the data being mirrored (offset due to frequency changes when filtering). In order to rectify this, the data used was taken for upto November 14th 2021, with a boxcar window applied before filtering (setting data after October 31st 2021 to 0). The data is then shown upto October 31st 2021. The data for number of total first dose vaccine administered had minimal high frequency noise (as this data was cumulative, not day-to-day), and so this data was not filtered. The data was however shifted, in order for it to represent total people who had been vaccinated for 14 days, rather than just the total number of people vaccinated who had been vaccinated by that day. The plots of this data (raw versus filtered) is visible on the next page.

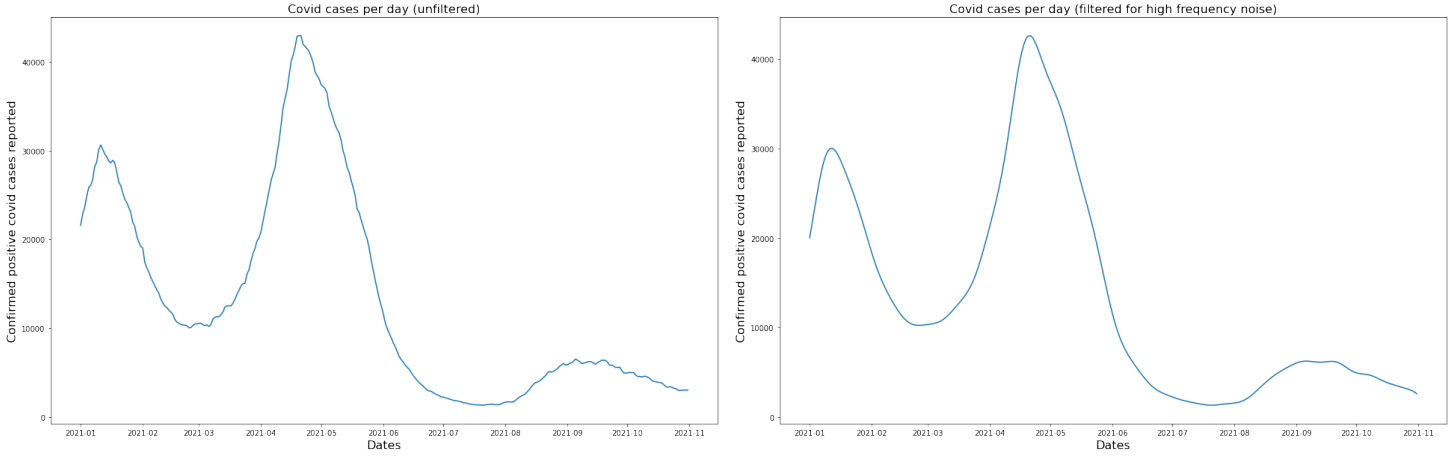


Figure 1: Covid-19 cases f-domain filtering comparison; unfiltered (left) versus filtered (right)

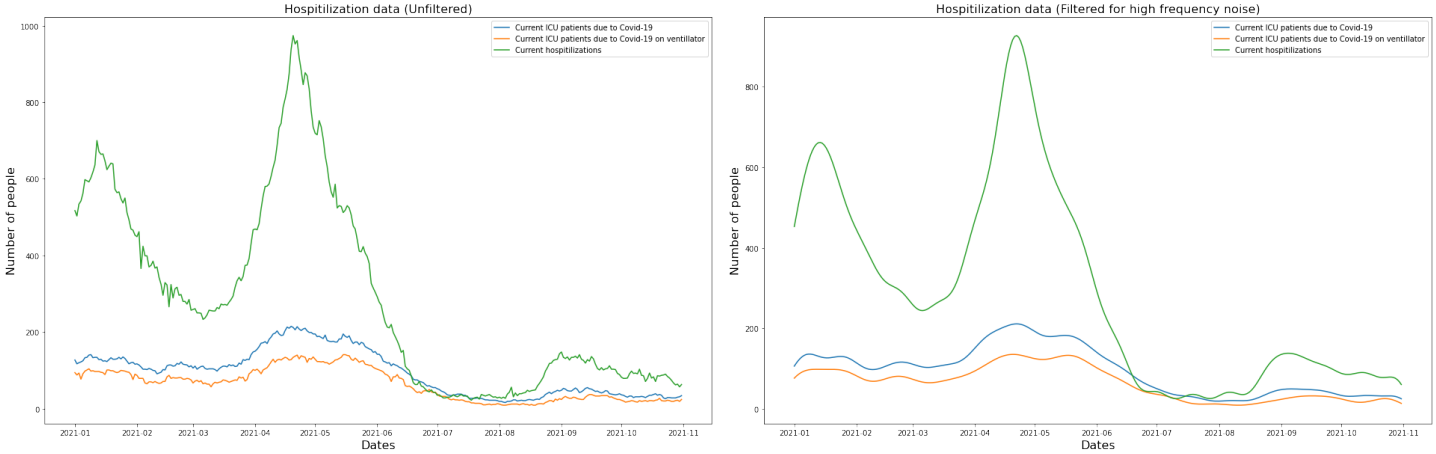


Figure 2: Hospitalization data f-domain filtering comparison; unfiltered (left) versus filtered (right)

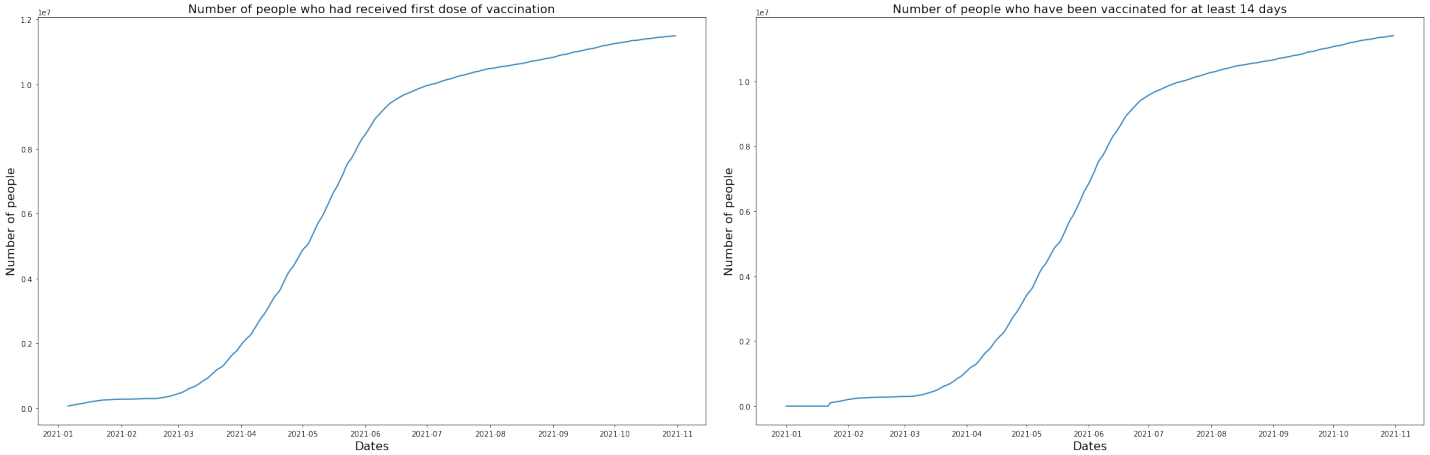
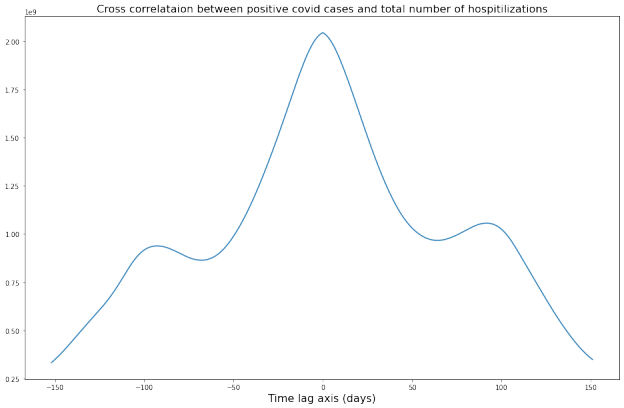


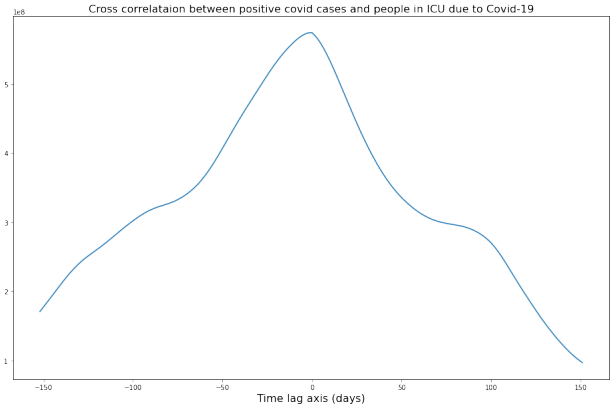
Figure 3: First dose vaccination data; unshifted (left), shifted (right)

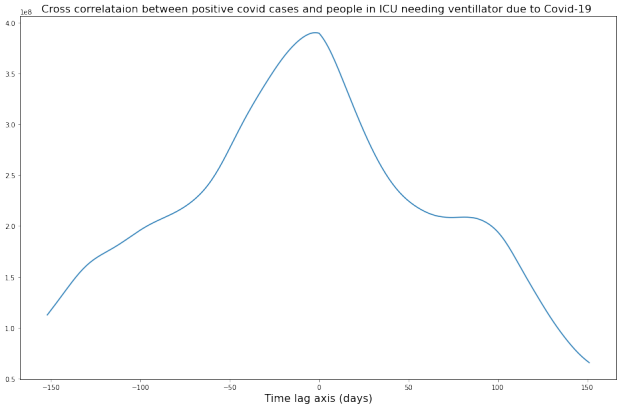
Analysis

From the graphs of the data, a few trends are immediately obvious. The shape of the graphs for total hospitalizations and positive Covid-cases per day are almost identical. The number of people in the ICU due to Covid, as well as the number of people requiring the ventilator in the ICU also seem to spike with increases in positive cases. To analyse these relations more precisely, we use plot the cross correlation of this data. We do this using Numpy's auto-correlate function (i.e. numpy.correlate with mode = 'same'). We use this function directly rather than computation directly from definition as we showed in lab 4 that these are identical.

For the cross-correlation graphs between positive covid cases and the different hospitalization data, we except to see a maximum at 0, with some smaller peaks fading to 0 away from the origin. The reason for this is that cross-correlation is a measure of overlap between two time-series, where this overlap is as one of the series is shifted across the other (hence the time-*lag*) axis. The maximum occurs when the series maximally overlap (from class, we know that the exact value of this maximum is arbitrary - as long as it is sufficiently large and of order magnitude comparable to the data sets, the shape is what determines the correlation). We except this peak to occur at t = 0, as these series maximally overlap without any shifting (which is what we except; positive covid cases lead to immediate hospitalizations).







References

[1] data.ontario.ca. Status of covid-19 cases in ontario. https://data.ontario.ca/dataset/status-of-covid-19-cases-in-ontario/resource/ed270bb8-340b-41f9-a7c6-e8ef587e6d11, 2022. [Online; accessed 19-April-2022].