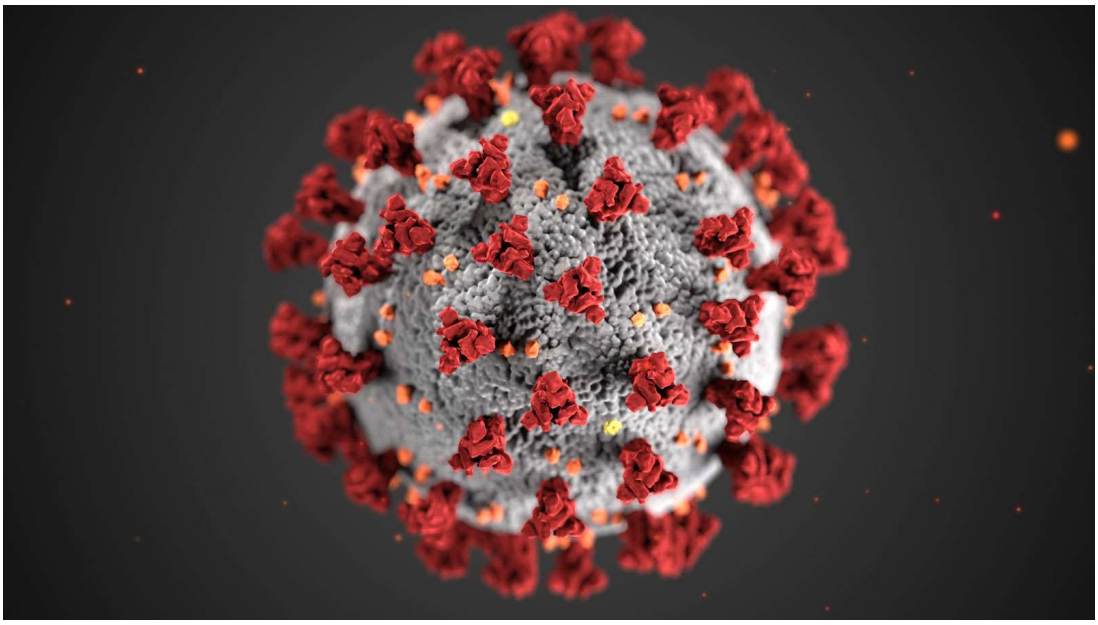

Visualising The Covid-19 Pandemic.

VISUALISATION AND ANALYSIS OF COVID-19 DATA

Joel Foster | COSC3000 | 26/04/2021

Contents

Introduction

The COVID-19 pandemic has had a vastly significant impact on the entire world, with ramifications being seen in several industries, such as health care and technology production. With the recent release of COVID-19 vaccinations, several concerns have been raised over the safety and effectiveness of such measures. In order to alleviate some concerns, data visualisations can be utilised to demonstrate the purpose of such measures and also provide a clearer understanding of the underlying factors involved in this pandemic. Additionally, such visualisations can reveal patterns or external influencing factors which have not been considered before, allowing a deeper understanding of the problem. As such, it seems extremely pertinent to generate such visualisations, not only for informing any future response but also conveying the influences of complex factors concisely to the wider public.

Aims

The primary aim of this report is to provide context for and summarise the current and past COVID-19 situation in Australia. In pursuit of this goal, this report will visualise connections between vaccination rates and the amount of positive COVID-19 cases in Australia. This will be achieved by:

- Examining case rates before and after the introduction of vaccinations.
- Examining other external factors such as geographical location, in relation to case rates.
- Visualising trends in the frequency of new cases/cases.

Through these visualisations, the report will clarify the main factors involved in the pandemic and indicate flaws/strengths in the current response, in turn helping to inform any future response.

Methods

Data Source

There are two primary data sources:

1. OWID Coronavirus Source Data (updated as of 08-03-2021) (University of Oxford, 2021) [1]
2. COVID-19 In Australia: Real-Time Report (updated as of 08-03-2021) (Real Time Report, 2021) [2]

The following necessary statistics are included in [1]:

- Date of record
- Total Cases
- New Cases
- Total Vaccinations
- New Vaccinations
- Total Tests

[1] was used in performing comparisons and generating most visualisations. [2] was used as a supplementary dataset, to get data related to specific regions of Australia which was needed for the generated heatmaps.

Both data sets were stored in .csv file for ease of use and are released under the Creative Commons BY license.

Data Processing

Data set [2], once imported, needed no formatting.

Data set [1] processing steps (can be found in [a]):

1. Importing into MATLAB.
2. Selecting required rows & columns
    a. Excluding non-Australia tuples
    b. Only including necessary variables
3. 'Cleaning' data
    a. Filling missing NaN values (for plotting and summaries)
        i. NaN were replaced with previous non-NaN value or 0 if no such value existed
4. Data set was split into the required subsets (tables) for each visualisation
5. Summary (average) variables were created and calculated in the associated data table
    a. Examples:
        i. Days Since Start of Vaccination Period (conversion between date and int types).
        ii. Factor shifted variables: Total vaccinations per thousand people.

6. Generating matrices and vectors for plots.

Tools:

Created in MATLAB:

- Figure 1:
  - Utilising:
    - bar()
      - For stacked bars
    - errorbar()
      - For specificity and sensitivity of tests
    - handles
      - For changing colouring, linestyle, etc.
- Figure 2:
  - Utilising:
    - bar()
    - plot()
    - handles
- Figure 3:
  - Utilising:
    - scatter()
    - plot()
    - handles
- Figure 4:
  - Utilising:
    - plot()
    - handles
- Figure 5:
  - Utilising:
    - plot()
      - For month lines
    - polar()
      - For spiral
    - plot3
      - For rising points
    - handles

Created in Excel:

- Animation

Results

Infectivity of COVID-19 virus

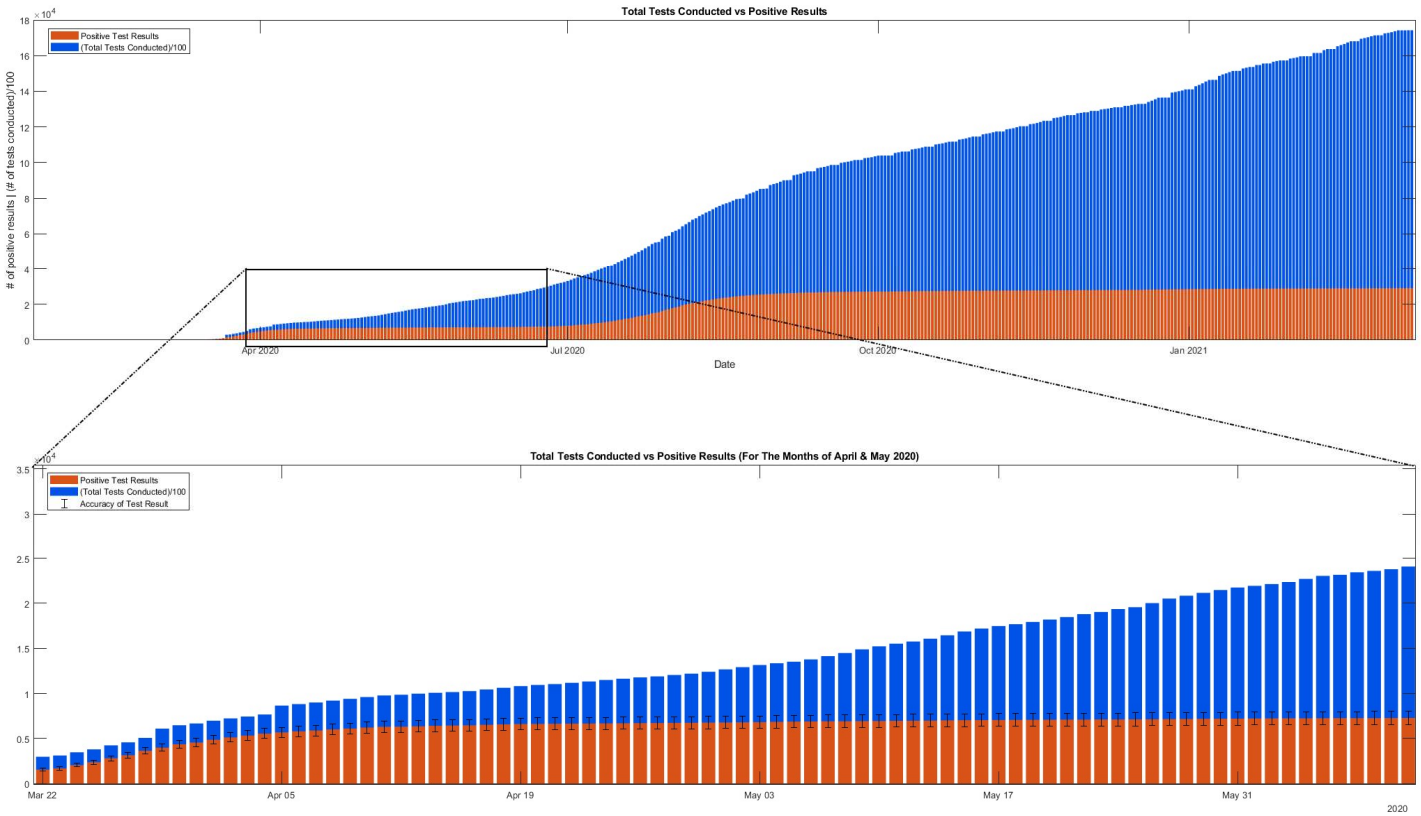

Figure 1 : A histogram showing total number of tests (blue) and total number of COVID cases (red) per day from March 2020 to March 2021

The figure above compares the total number of tests vs the total number of positive results (cases of COVID19). It does this using a stacked bar graph, displaying COVID19 cases as a subset of all COVID19 tests. This figure shows the progression of the COVID19 over the first year of quarantine. It highlights the relationship between the number of tests conducted and the number of COVID19 cases. Additionally, it draws attention to JUL-SEP 2020 period as a point of interest (linear fit of tests vs cases). The tail end of the plot indicates a decline in new cases despite a significant increase in tests being conducted. The top plot highlights two points of extreme ascent of new cases (MAR and AUG), with near constant behaviour at all other points. Ultimately, the purpose of this visualisation is to demonstrate how the number of cases are not dependent upon the number of tests performed (other then by the inequality # of tests > # of cases).

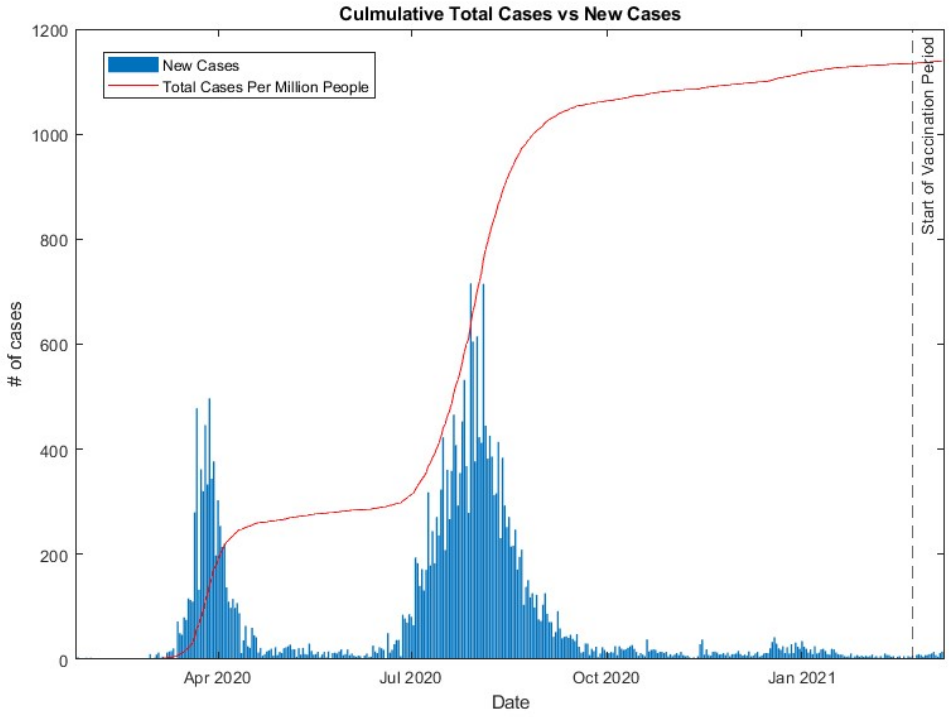

Figure 2: A line graph (total cases per million people) against a histogram (new cases) per day from March 2020 to March 2021

The figure above compares the total number of cases vs the number of new cases by calendar day. It does this using a bar graph of new cases overlayed by a line graph showing the cumulative total by date. The figure draws attention to MAR 2020 & JUL 2020 as regions of extreme increases in the number of infected individuals. Conversely, APR 2020 & SEP 2020 are highlighted as regions of extreme decreases. The tail end of the plot highlights the difference between the number of new cases before and after the vaccination period.

Relation Between Vaccination Rates and Positive Cases

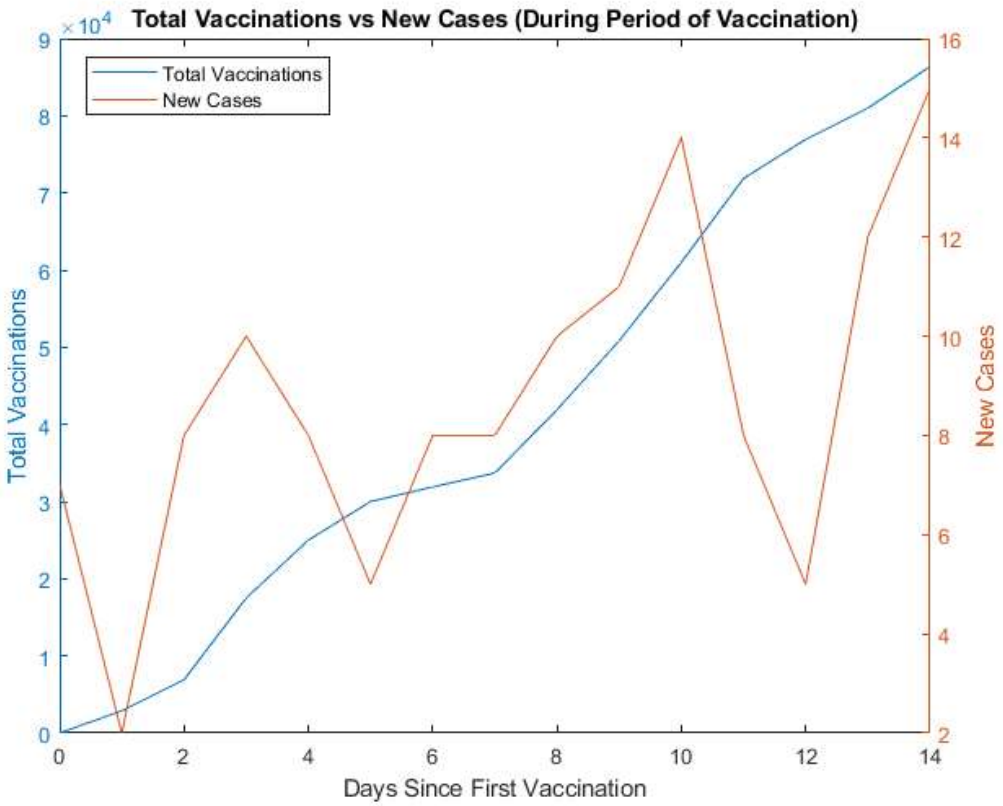

Figure 3: Overlayed line graphs (total vaccinations, total cases) per day after the start of vaccination period.

The figure above compares the number of new cases vs the number of total vaccinations over the vaccination period. It does this by plotting two line graphs, one showing the total number of vaccinations and the other showing the number of new cases. The figure highlights the increasing oscillation of new cases over the vaccination period, despite closely linear increase of total vaccinations. This indicates a positive relationship or non-relationship between the number of total vaccinations and the number of new cases; however, the data set is too small to be certain of this.

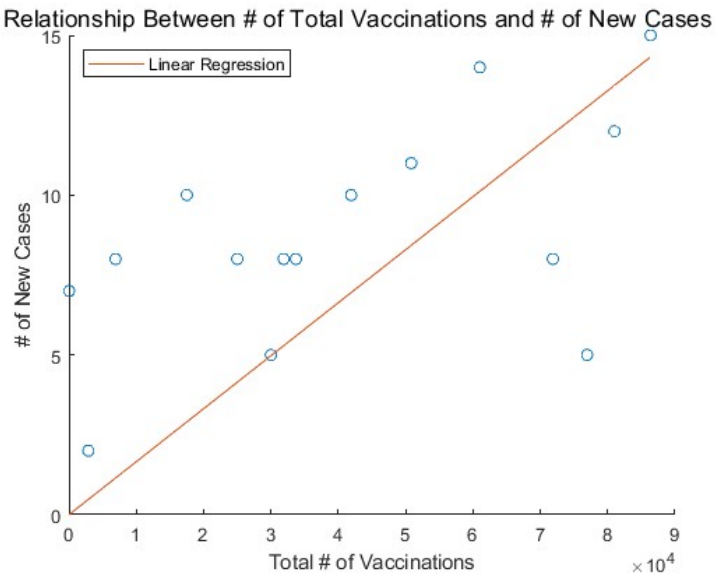

Figure 4: Scatter plot of new cases vs total vaccinations with applied linear regression fit

The figure above highlights the relationship between the total number of vaccines and the number of new cases. It does this by using a scatter graph to plot each element tuple then performing a linear regression, then plotting the linear fit over the scatter graph. The figure suggests a weak positive linear relation between the two variables. The figure also highlights outliers between [0 1] and [7 8] on the x axis.

External Factors

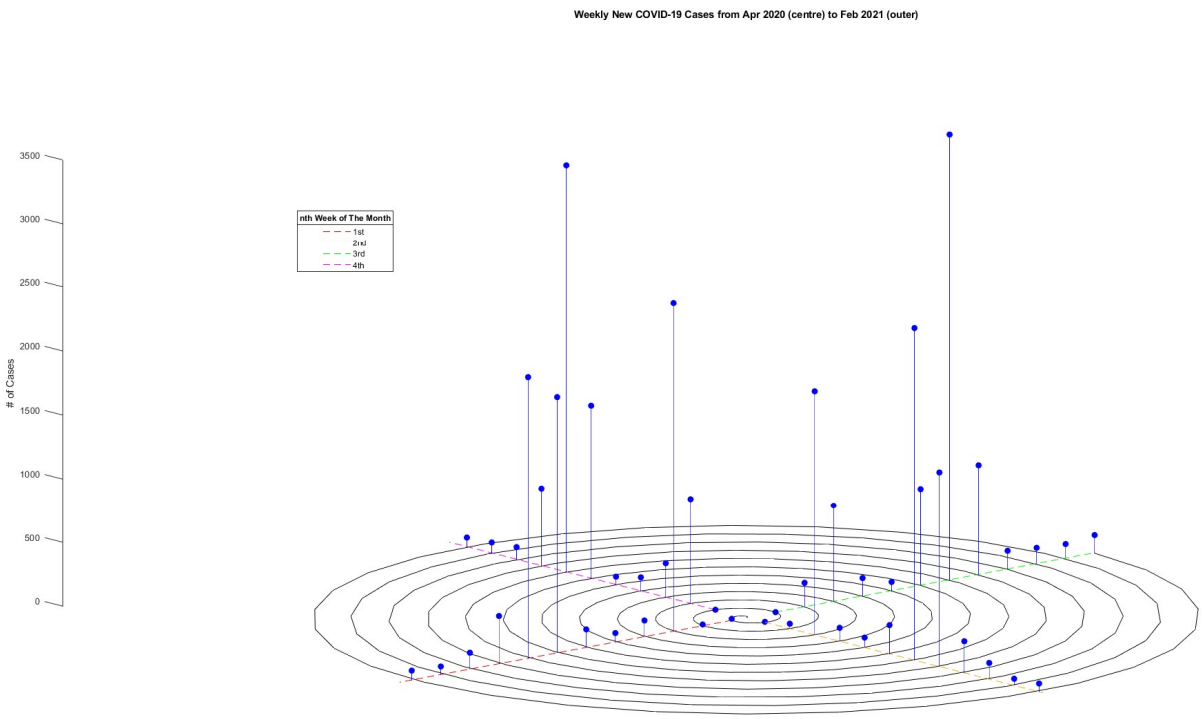

Figure 5: Spiral plot of new cases for each week of a month (month is represented by a full revolution/ring)

The figure above compares the number of new cases per week. It does this using a 3d spiral graph, where 90-degree rotations indicate each week of a month and the number of new cases is indicated by the height of the point above the spiral graph. The figure shows months of consistent high new cases (OCT) and months of consistent low new cases (JAN). Additionally, it is shown that the number of new cases per week can vary drastically every week. In relation to this aspect, the tail end of the plot is significantly juxtaposed with the rest of the plot. Additionally, a decreasing and slightly oscillating trend can be observed in the tail end, indicating a significant change which has influenced the frequency new cases (vaccinations or external factor).

Animation:

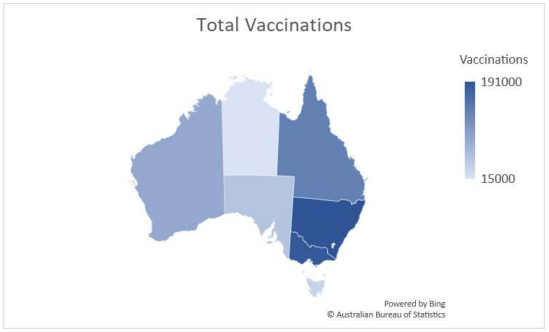

Figure 6 – 1st frame of animation

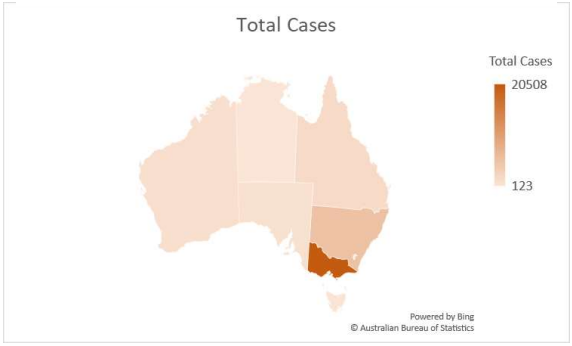

Figure 7 – 2nd frame of animation

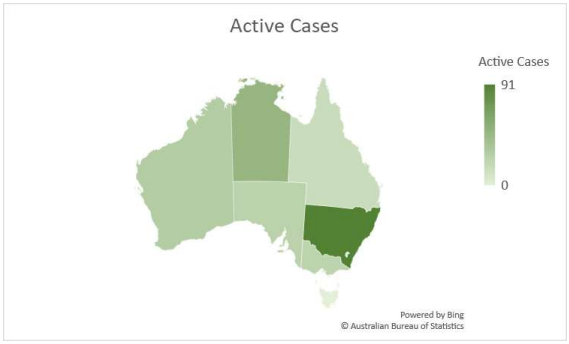

Figure 8 – 3rd frame of animation

The figures above depict each frame of the attached animated GIF [b]. Each frame shows a heatmap of Australia based on different variables. Darker shades were used to indicate higher values. Figures 6, 7 and 8 show that despite Victoria's high vaccination it has the highest number of total cases, however, has one of the lowest numbers of active cases. Similarly, figures 6 and 8 show that despite New South Wales' high vaccination total, it has the highest active case totals. These two observations insinuate that there are other geographical related reasons (legislation/lockdown procedure) that have influenced COVID19 rates.

Discussion

Findings

Trends & Observations from figures 1 & 2:

- The number of new cases has large deviances before the introduction of public vaccination.
- The period after vaccinations maintained low new case rates.

Trends & Observations from figures 3 & 4:

- The number of total vaccinations and new cases share a similar linear trend (over the vaccination period)

Trends & Observations from figures 5 & 6:

- The number of new cases exhibits local oscillating behaviour
- The number of new cases remained low after start of vaccination period
- Vaccinations seem to be a lesser factor to the number of new cases than geographical region (legislation, region's individual COVID19 plan).

These findings suggest that vaccinations have little effect on the rate of new COVID19 cases. This result is likely due to the limited data set and the effects of an unexplored factor.

Design Decisions

Shared Design Techniques:

- Use of contrasting colours (excluding red/green contrasts) in order to facilitate clear contrasts (whether the person is colorblind or not)
- Optimising aspect ratio, model size, viewing angle, legend placement, amount of whitespace etc. in order to accurately portray the data whilst also making it visually stimulating. (in particular Figure 4)

Figure 1:

- Stacked Bar Chat to emphasise the relation between total tests and total cases.
- Zoom in sub plot to highlight key sub region.
- Error Bars used to indicate the possible effect of

Figure 4:

- Plotted linear regression over data points to examine validity of line of best fit.

Figure 5:

- Exploring the possible cyclic nature of the data as well as confirming suppositions from previous models

- Excellent choice of model (since the data is dependent upon time)

Figure 6:

- Using relatable and understandable imagery (map) to convey important information.
- Using one color of varying lightness to ensure it is interpretable by all (and using strong shades for the extremes)

Limitations/Areas for Improvement

The main limitation encountered, as shown in Figure 2, was the lack of data after the introduction of vaccination. This significantly limited my ability to produce any meaningful model on the effects of vaccination on the spread of COVID19. Due to this limitation, figures 3 and 4 suffer greatly and don't clearly reflect the situation.

Additionally, as alluded to in the report, there are influencing factors (such as quarantining) which aren't included in the data set and as such cannot be accounted for. Examining such influences and how they interact with vaccination rates to influence case rates would be beneficial (multivariate data). Additionally, a cluster examination of the total vaccinations vs new cases plot may provide interesting insight.

The animation could have been improved on by sourcing new data which used more specific location data. Doing this could lead to understanding more about the influence of location on COVID19 spread. Additionally, given enough time visualising the differences between countries would prove insightful. Additionally, by utilising a more robust data set the visualisation could be extended over time (allowing it to demonstrate similar capabilities to the spiral plot).

Personal Reflection

Despite these areas for improvement, I am proud of the visualisations I have produced. In particular, I am proud of my ability to learn and use MATLAB to generate a range of different yet appropriate graphical representations. As such, the visualisations produced are clear, concise and demonstrate trends within the data.

References

Real Time Report. (2021, 3 8). Retrieved from COVID-19 in Australia: https://covid-19-au.com/

University of Oxford. (2021, 3 8). *Coronavirus Pandemic.* Retrieved from Our World In Data: https://ourworldindata.org/

Appendix

List of attachments:

- visualise.m (MATLAB script for generating plots) [a]
- aus_heat.gif (Animated GIF) [b]