

CSC8002 – Big Data Management – Final Project

**Analysis of Mobility Data and Covid 19 Infection Rates in Australia and Sweden
in 2020**

Name: Diana Kither

Table of Contents

i) Introduction	3
ii) Data Exploration	3
iii) Literature Review	4
iv) Research Question	6
v) Method	7
vi) Connection between Data Sets	7
vii) Data Analysis	7
viii) Findings	12
ix) Ethics and Privacy	12
x) References	13

i) Introduction

The Google COVID-19 mobility data set was created to investigate changes in mobility of populations in various parts of the world during the COVID-19 pandemic (Google 2020). The data set contains the travel data of people to common locations collected from the location history of personal mobile devices (Google, 2020). Data was collected between February 15th 2020 and December 1st 2020 (Google, 2020). Numerous types of locations were grouped into categories including residential, workplace and retail. Daily movement data to these location categories was collected, combined and then compared to a baseline mobility value to determine the percentage change in movement of people to those locations as the pandemic progressed (Google 2020). Data was collected across multiple geographic regions in 135 countries around the world. This report will be focusing on data collected from Australia and Sweden. Google has made the data set publicly available on the web and available for download at <https://www.google.com/covid19/mobility/>.

The COVID-19 cases data set is a database created by the team at Our World in Data at Oxford University in the United Kingdom. Given the multitude of data sources around the world and lack of world authority with respect to data reporting during the pandemic, the Our World in Data research group created a database with the aim of centralising this international data to make it easily accessible for further investigation (Hassel et al, 2020). The database is comprised of data pertaining to case numbers, rates of PCR testing, patient deaths, population numbers, vaccination rates and other health statistics from 94 countries in the world relating to the COVID-19 pandemic (Hassel et al, 2020). Data has been collected since January 2020 from government and published data in addition to validated data from other sources (Hassel et al, 2020). Again this report will be focusing on data collected from Australian and Sweden. Our World in Data has made this dataset publicly available for download at <https://ourworldindata.org/coronavirus-data>.

The COVID-19 pandemic has had devastating effects around the world. Social distancing and limiting movement are crucial elements in managing the spread of this highly contagious virus. Monitoring movement data of the population to different types of locations in relation to rates of infection in different countries with different management strategies is important with respect to Australia's response to this pandemic. Our knowledge of population movement combined with medical data enables us to assess the efficacy of movement restrictions in managing the spread of infection and to identify types of locations where the risk of infection may be higher. In addition, this information can also be used develop recommendations to manage population movement, identify infection hotspots and in turn limit further infections in this pandemic and for disease control in the future.

ii) Data Exploration

The Google COVID-19 mobility data set is comprised of 234.1MB of data in a comma separated (.csv) file. For initial assessment of the data set, the data file was loaded into Jupyter Notebooks and analysed using PySpark. The data set contains 14 variables and 3,342,175 observations. For this report, only data from Australia and Sweden was examined.

A total of 76,688 observations were collected in Australia and 65,171 observations from Sweden. Data was collected from February 15th 2020 till December 1st 2020. After removal

of variables which contain no values 10 variables remain. The four initial variables in the data set identify the location from which movement data was collected and are all categorical variables. The first two variables in the data set identify the country where the data was collected, by code and name respectively. The next two variables further identify the region within the country from which the data was collected. Sub_region_1 indicates the state or territory and sub_region_2 the council or shire within the state or territory. The variable Iso_3166_2_code represents the International Organisation for Standardisation (ISO) country and subdivision identification code, an internationally recognised code to identify countries and their regions (International Organisation for Standardisation, 2020). The variable date indicates the date that location data was collected. Data was collected daily.

The final six numerical variables measure the change in movement across six different categories of location, retail and recreation, grocery and pharmacy, parks, transit stations, workplaces and residential. Data was collected from a variety of location types and similar places were grouped together in each category however some locations may fall into more than one category (Google 2020). Each of these variables measures the percentage change in movement to that location category which is calculated daily (Google 2020). Movement to areas was measured from location data saved in the location history from individual mobile devices (Google 2020). This data was combined and then compared to a baseline movement value for that particular day of the week and a percentage change between the two values was calculated giving the percentage change in movement for that date for each category (Google 2020). A baseline day value was calculated for each day of the week and was considered a typical movement value for that week day for each location category (Google 2020). This typical value was calculated from the median values for that week day from January 3rd 2020 till February 6th 2020 for the individual categories (Google 2020). Google (2020) also acknowledges that the baseline values do not consider periodic movement changes that may result due to seasons or other events.

The Our World in Data COVID-19 cases data set is comprised of 15.98MB of data in a comma separated (.csv) file. The original data set contains 62,407 entries and 59 variables. The Australian data set contains 367 entries with data collected from January 26th 2020 till January 26th 2021. The Swedish data set contains 361 entries with data collected from February 1st, 2020 till January 26th, 2021. After the removal of variables with no values the Australian and Swedish data sets contained 41 and 51 variables respectively. The first three variables iso_code, continent and location identify the country location where data was collected which is then followed by the date on which data was collected. The next variables identify the number of cases diagnosed on each day and the cumulative number of cases over time. Similar values are calculated for the number of deaths. Given the differences in how countries report daily case numbers and the number of deaths, a smoothed value was calculated for each of these values in an effort to minimise the effect of these differences (Hassel et al, 2020). The smoothed value is calculated as a 7 day average of each variable (Hassel et al, 2020). The next group of variables pertain to numbers of PCR tests conducted and cumulative numbers of tests, again including a smoothed value.

The next group of variables in the Swedish data set describes the rates of vaccination, including new and total vaccination numbers, vaccination numbers per head of population and smoothed values. These figures are not included in the Australian data set as vaccination programs have not commenced. The final group of variables in the COVID-19 cases data set is comprised of variables associated with population number, GDP, living standards, hospital and ICU admissions and other generalised health data. ICU data however is not included in the Australian data set.

iii) Literature Review

Since the release of the Google COVID-19 mobility data set, several studies have utilised this information to analyse the effects of population mobility and lockdown policies on managing the COVID-19 pandemic.

Yilmazkuday (2020) performed a study soon to be published in the Journal of Human Behaviour and Social Environment comparing international mobility data from the Google COVID-19 mobility data set with data of COVID-19 case numbers and deaths. The aim of this study was to investigate the relationship between COVID-19 infection rates and mobility changes as a result of lockdown restrictions (Yilmazkuday, 2020). Yilmazkuday (2020) used a difference in difference study design to investigate relationships between the two data sets across 130 countries from February to May 2020, making considerations for the effects of particular days, countries and timing of the hundredth COVID-19 case in each area. This analysis showed very positive results (Yilmazkuday, 2020). Yilmazkuday (2020) found that increased time at places of residence and decreased time in places outside the home including work, transit, retail and recreation locations all reduced rates COVID-19 cases and deaths. Furthermore increased time spent at places of residence had the greatest impact on COVID-19 statistics (Yilmazkuday, 2020).

Wang et al (2020) conducted a study using the Google COVID-19 mobility data set to investigate the relationship between changes in population mobility and COVID-19 infection rates across all states and territories in Australia. Mobility data from all regions in Australia was examined for a six month period from February 15th 2020 till August 15th 2020 (Wang, 2020). Furthermore Wang et al (2020) also went on to investigate the time delay between instituting lockdown measures and the point at which these restrictions start to affect COVID-19 infection rates. The study utilised Google COVID-19 mobility data and COVID-19 health data obtained from the Australian Department of Health (Wang et al 2020). Similarly to Yilmazkuday (2020), Wang et al (2020) also found that mobility restrictions had a positive effect on reducing the rate of COVID-19 infections in Australia. Wang et al (2020) also found that restrictions had a delay of between 7 and 14 days before their impact on COVID-19 infections was seen. Given the documented incubation period of COVID-19 of up to 14 days this result is to be expected (Wang et al 2020). In addition Wang et al (2020) also looked at mobility prior to the first and second waves of infection, focusing on Victoria and found the relationship between mobility and COVID-19 infection rates varied with area, time and mobility type, highlighting that other measures for preventing infection are also influencing infection rates in addition to mobility restrictions. Furthermore changes in some mobility types differed between Victoria's first and second wave of infections (Wang et al 2020).

Various studies have also been conducted utilising the Our World in Data COVID-19 cases data set. Cao et al (2020) used the Our World in Data COVID-19 database to investigate the influence of demographics and socioeconomics on the COVID-19 global case fatality rate (CFR), the ratio between COVID-19 deaths and the number of positive infections up to July 2nd, 2020. Cao et al (2020) found that the average CFR amongst countries in the world was higher than previously estimated and that CFR increased in countries with larger populations particularly in countries with greater incomes. It was suggested that this result may be influenced by factors including the strain on health care services with larger numbers of people, higher risks of spread of infection and poorer health (Cao et al 2020). When examining various socioeconomic factors, it was also found that increased rates of smoking in females was associated with an increase in CFR (Cao et al 2020). The authors also highlight that the study is assessing these results on a global scale and that further investigation at a country level may yield varied findings (Cao et al 2020)

Phipps et al (2020) used the Our World in Data COVID-19 cases data set to investigate the use of backcasting statistical methods to estimate the true infection rate of COVID-19 in the population. Differences in testing, tracing and reporting between countries make estimating the actual infection rate in a population difficult (Phipps et al, 2020). With their method, Phipps et al (2020) estimate that the true COVID-19 infection rate in the population is around 6.2 times greater than the actual infection rate and that these results were similar to other epidemiological studies. The study also found that there was an inverse relationship between true detection rate and the proportion of positive test results highlighting the importance of testing on a large scale and also testing those in the population who are asymptomatic for disease (Phipps et al 2020)

iv) Research Question

The goal of the research question of this report is to investigate the differences in population mobility and COVID-19 new case numbers between Australia and Sweden to examine for any significant similarities or differences between the two countries. Australia and Sweden were selected for this investigation due to their similar living standards and position in the world economically but also for the differences in their COVID-19 response. Australia has adopted a firm response to the COVID-19 pandemic, instituting strict lockdown measures during the first and second waves in the country as well as other social distancing and hygiene measures to curb the spread of infection. Sweden however has been less stringent favouring a herd immunity approach, bringing in some lawful restrictions including reducing numbers of people at gatherings and other measures however it has not had a strict lockdown, relying on general and voluntary restrictions to manage the spread of COVID-19 (Claeson & Hanson 2021; Duckett & Mackey 2020; Franks 2020; LePage 2020). On the Lowy Institute poll rating countries response to the COVID-19 pandemic, Australia ranked at 8th in the world whilst Sweden was 37th (Lowy Institute, 2020).

For this report COVID-19 mobility and case data was explored from February 15th 2020 till December 1st 2020. Through investigating the trends of the numbers of new cases in both countries and changes in population mobility we can assess the response of each approach to managing the pandemic. The insights gained from this data and from comparing Australia's management of COVID-19 with that of other nations can in turn assist in improving our country's response to the COVID-19 pandemic as it continues.

v) Method

To investigate the research question, Spark was used to explore and analyse the data set. Spark is a distributed processing system which is gaining in popularity for analysing big data sets (Gupta 2016; Kalron 2020; Stoica 2014). Spark can read data from a number of field systems and is able to process data queries quickly and in parallel within memory which facilitates its speed (Gupta 2016; Kalron 2020; Stoica 2014).

The challenge with processing big data sets is their sheer size, requiring large amounts of processing power to achieve this. The Hadoop ecosystem is a software platform for distributed processing, enabling data to be processed across a cluster of computers (Bhandari 2020; Kalron 2020). Processing queries in parallel across multiple clusters simultaneously is advantageous for several reasons (Guptak 2016; Kalron 2020). Parallel computing enables problems to be processed quickly (Guptak 2016; Kalron 2020). Processing is spread across a number of machines enabling a more efficient use of hardware and reducing hardware requirements which in turn reduces cost (Guptak 2016; Kalron 2020). Given some of the limitations of the Hadoop system, Spark was subsequently developed to improve processing times and efficiency (Gupta, 2020). Spark can operate on top of the Hadoop system strengthening its performance or it can operate independently (Gupta 2016; Kalron 2020; Stoica 2014).

Hadoop MapReduce is a programming system used to process and analyse data in the Hadoop ecosystem (Bhandari 2020). For this project Spark was selected given its increased speed and efficiency with processing and analysing large data sets when compared to Map Reduce (Gupta 2016; Kalron 2020).

vi) Connection between Data Sets

To investigate the research question, the Google COVID-19 mobility data set and the Our World in Data COVID-19 cases data set were utilised. New case numbers were compared between the two countries as well as changes in mobility in all of the six location categories to see if any significant similarities or differences were present.

The data sets were each analysed with PySpark and visualisations created using Pandas for further analysis within a Google Colab Notebook. Data was selected for both Sweden and Australia from both data sets for the time period from February 15th, 2020 till December 1st, 2020. Using Pyspark the data subset for Australia as whole was selected from the COVID-19 mobility data set for each mobility category and the same method was applied to select data from Sweden. From the COVID-19 cases data set population, total and new case data was selected for both Australia and Sweden individually for comparative analysis. The code for data analysis can be found in the file *Final_Project_Colab.ipynb*

vii) Data Analysis

The goal of the research question was to compare COVID-19 new case numbers and population mobility statistics between Australia and Sweden from February 15th 2020 to December 1st 2020. As western countries with similar living standards, Australia and Sweden had different responses to the COVID-19 pandemic. Whilst Australia opted for a hard lockdown approach to curb the spread of COVID-19, Sweden has been more relaxed with their restrictions but has still experienced a voluntary lockdown amongst its citizens as

a result (Claeson & Hanson 2021; Duckett & Mackey 2020; Franks 2020; LePage 2020). During the pandemic, Sweden has experienced a much greater number of cases of COVID-19 compared to Australia. As of December 1st, 2020 Sweden had a total case number per million of population of 25,819 cases with population of 10.09 million people. Australia had a total case number of 1095 cases per million of population by that time with a population of 25.49 million.

The first step in data analysis was to compare new case numbers for each nation over the prescribed time period. To compare new case numbers between the two countries, data was plotted on a logarithmic (log 10) scale in an effort to better visualise trends in the data as shown in Figure 1 below. Plotting data on a logarithmic scale enables a more efficient comparison of case numbers of very different magnitudes from different sized populations (Fleming 2020). Furthermore it also allows for better assessment of the rates of change in new case numbers when compared to plotting on a linear scale which is useful in assessing the effectiveness of infection control measures as they are instituted (Fleming 2020).

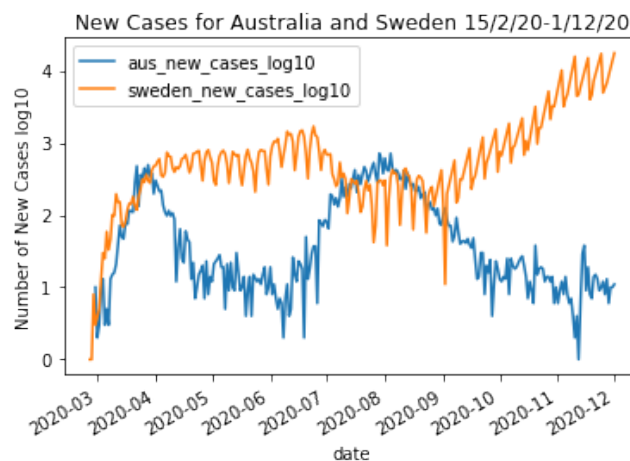


Figure 1 – New COVID-19 case numbers (log 10) in Australia and Sweden from 15.2.20 – 1.12.20

When examining Figure 1 it is evident that both countries had a rapid rise in case numbers at the start of the pandemic. In April the rate of new cases in Australia began to fall whilst in Sweden new case numbers continued at not dissimilar levels until July. As Australia's second wave took hold in August the rate of new cases in Australia began to rise sharply as Sweden's level began to decline to less than that of Australia. By September however this trend had reversed with new case numbers in Sweden increasing rapidly with the second wave of infection in Europe to the highest levels the country had seen. Rates of new cases in Australia have continued decline since August with a large disparity between new case numbers of the two countries clearly evident just before Christmas.

The data from the six categories of mobility measured in the Google COVID-19 mobility data set was then examined. Figure 2 below illustrates the change in mobility between the populations of Australia and Sweden from places of residence.

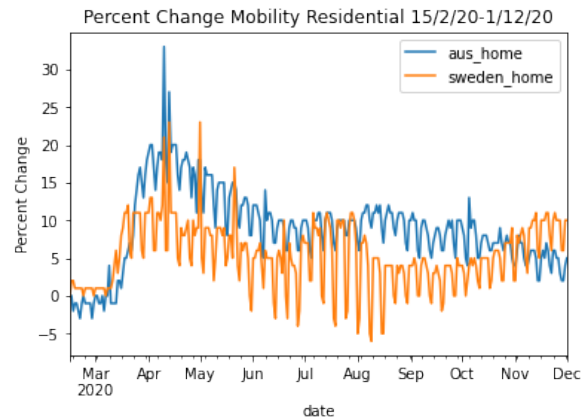


Figure 2 – Percentage change in mobility to places of residence in Australia and Sweden from 15.2.20 – 1.12.20

As is evident in Figure 2 above, Australia had a general increase in time spent at home compared to Sweden across the year. Australia had an average increase in the time spent at home of 8.9% with a standard deviation of 5.36% slightly higher than that of Sweden at 5.35% with a standard deviation of 4.41%. Both countries experienced a peak increase in time spent at home in April likely due to Easter. Sweden also experienced peaks in May and June likely associated with the public holidays May Day and Ascension Day (Tomlin 2019). However in November, time spent at home in Sweden increased to greater than that in Australia likely secondary to start of the second wave of infections in Europe in conjunction with the relaxation of lockdown restrictions in Australia.

The next mobility category assessed was movement to the workplace. As is evident in Figure 3 below both Australians and Swedes spent less time in the workplace with movement to the places of work generally below 0%. Sweden had a much greater variation in movement to the workplace compared to Australia however the mean change in mobility to the work place was actually greater than that of Australia at -22% compared to -17%. Sharp decreases in workplace movement in April in both countries are likely the result of Easter. The declines in Sweden for May and June all coincide with public holidays for May Day, Ascension Day and the Midsummer celebrations (Tomlin 2019). Australia experienced another sharp decline in October likely secondary to the Queen's Birthday holiday and Thank You day holiday in some states.

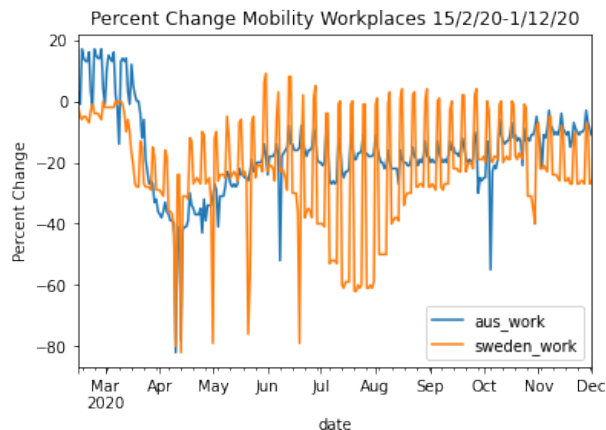


Figure 3 – Percentage change in mobility to workplaces in Australia and Sweden from 15.2.20 – 1.12.20

The third mobility category examined was movement to recreational and retail areas. This category also includes movement to restaurants and cafes. As is evident in Figure 4, movement to these areas was greater in Sweden compared with Australia until November at the commencement of the second wave in Europe. Australia had a mean decrease in mobility of -18.6% to retail and recreation areas compared to that of Sweden at -8.5%. While both countries generally had mobility of less than 0% this was not the case for Sweden during the summer months where mobility percentages were greater than baseline. Sharp declines were seen in both countries around Easter and in Sweden around the time of midsummer and summer solstice celebrations (Tomlin 2019).

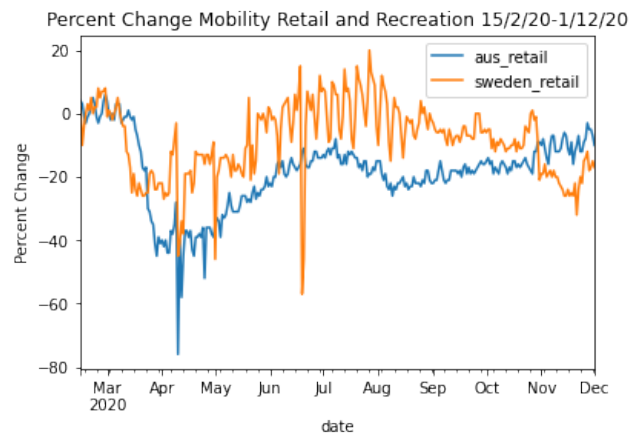


Figure 4 – Percentage change in mobility to retail and recreational areas in Australia and Sweden from 15.2.20 – 1.12.20

Movement for groceries and pharmacies was similar in both countries which is not unexpected given the need for these essential items. The mean change in mobility to grocery and pharmacy for Australia was -1.65% and for Sweden -0.93%. Australia had a spike in mobility for grocery and pharmacy in March in conjunction with panic buying experienced in the country at the start of the pandemic. Sweden had a sharp decrease in mobility for grocery and pharmacy at the end of June which may be in conjunction with the large celebrations experienced for midsummer and the summer solstice, a major event and holiday in the country (Tomlin, 2019).

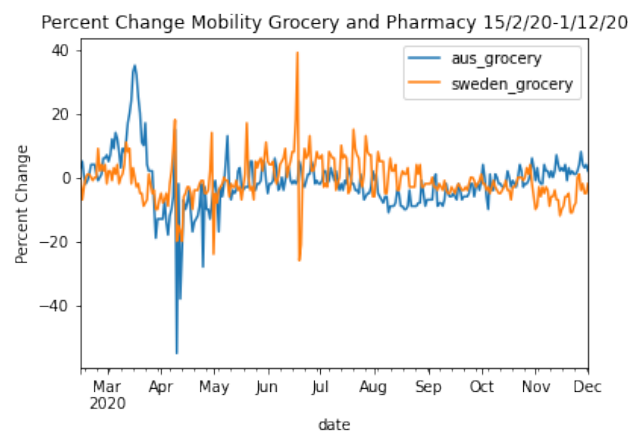


Figure 5 – Percentage change in mobility for grocery and pharmacy in Australia and Sweden from 15.2.20 – 1.12.20

Movement to public transit stations was reduced for both countries compared to baseline often significantly however this reduction was greater for Australia when compared to Sweden. Australia had a mean reduction of -39.06% and Sweden of -25.51%. Sweden had a greater variation in movement, particularly over the summer months when people are likely to venture out and when restrictions in Europe were more relaxed. However movement in Sweden then reduced to less than that of Australia in November coinciding with winter and the second wave on infection in Europe where restrictions were increased.

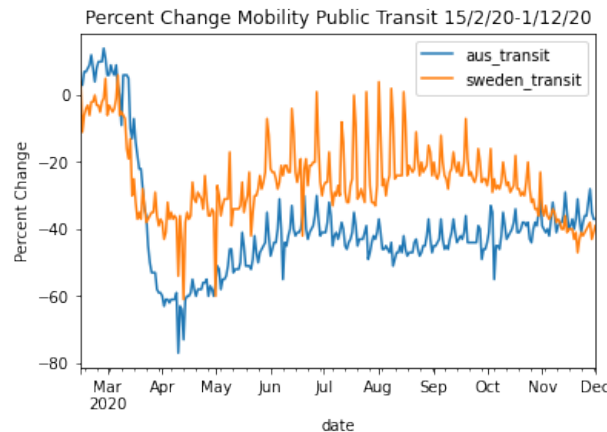


Figure 6 – Percentage change in mobility to public transit stations in Australia and Sweden from 15.2.20 – 1.12.20

The final mobility category examined was movement to parks which includes city and national parks as well as beach areas. Some significant differences were seen in mobility to these areas between the two countries. Movement to parks in Australia decreased by an average of -14.86% across the year with a maximum increase of 20%. Sweden however had a marked increase in movement to parks from baseline. The country experienced an average of 102% increase from baseline with a maximum increase of 380%. It should be noted that these increases were predominantly noted in the summer months. During this time better weather conditions will encourage people to venture out compared to the cold winters experienced in Europe and furthermore restrictions were generally more relaxed in Europe during the summer. However it should also be noted that the baseline mobility level for the Google mobility data was calculated during winter. Given the significant differences in weather between Swedish summers and winters this may in turn affect results.

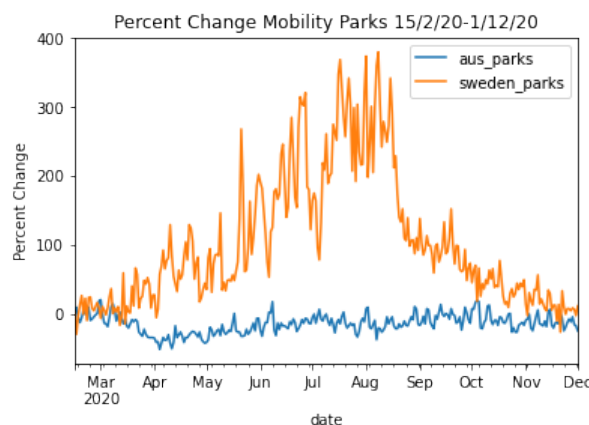


Figure 7 – Percentage change in mobility to the parks in Australia and Sweden from 15.2.20 – 1.12.20

When examining the COVID-19 case data between Australia and Sweden up until December 1st, 2020, it is clear that Australia has been more successful at managing the rate of COVID-19 infections. The total number of cases in Australia at that time had reached 27,923 cases with a total cases per million of population of 1095.03. Sweden however had a total case number of 260,758 with total cases per million of 25,819.49. Australia in turn has a population of around 2.5 times that of Sweden at approximately 26 million compared to Sweden at 10 million however it is important to remember that Australia's population is spread widely over a large land mass when considering its average population density of 3.202 people per square kilometre compared with Sweden of 24.178 which will affect management of the spread of COVID-19. The data also highlights that despite Sweden having some restrictions which in turn did lead to voluntary lockdowns as a result of public compliance to recommendations, more of the population in Australia were staying at home when compared to Sweden and mobility outside the home in Sweden was greater than that of Australia.

Experts acknowledge that limiting mobility will have a positive effect on reducing the spread of COVID-19 and Australia's ability to achieve this is evident in the reduction of new case numbers. When examining data from Sweden, more relaxed restrictions resulted in greater population movement and higher rates of new case numbers when compared to Australia. Despite the presence of some restrictions in Sweden and other public health recommendations this less stringent approach coincides with higher rates of new COVID-19 cases with levels being at their highest by the start of December. These results favour a stricter approach with respect to movement restrictions in managing the spread of COVID-19. It is also important to note that other recommendations with regards to managing the spread of infection including social distancing, hygiene recommendations and mask wearing also have a positive effect on the prevention of COVID-19 infections in addition to mobility restrictions (Wang et al 2020).

viii) Findings

The Google COVID-19 mobility data set and the Our World in Data COVID-19 cases data set were used to investigate patterns and differences between new cases of COVID-19 infection and changes in mobility for Australia and Sweden. Australia and Sweden adopted different policies in their management of the COVID-19 pandemic with Sweden adopting a less stringent approach with regard to restrictions compared with Australia. Australia has been more successful compared to Sweden with managing the spread of COVID-19 and differences in mobility were seen in several areas. In both countries, mobility was generally reduced with the exception of parks in Sweden. Population mobility however was reduced to a greater extent in Australia compared to Sweden with Australians generally spending more time at home. This has likely influenced Australia's success in managing COVID-19 highlighting that reducing population movement is important in managing the spread of COVID-19. Other measures however including hygiene, masking wearing, social distancing and other factors are likely to also have also influenced this result (Wang et al 2020).

ix) Ethics and Privacy

With the explosion of information available to us from data, we have an abundance of material from which we can gain beneficial knowledge that can aid decision making. However maintaining user privacy and utilising this information ethically is of paramount

importance (Lee et al 2016). Data usage is also governed by standards and legislation and it is imperative that these guidelines are adhered to (Lee et al 2016). When using the Google COVID-19 mobility data set considerations in relation to both privacy and ethics whilst using this information are imperative. Data collection is legislated by Australian Privacy Law. Australian privacy law dictates that an organisation can only collect data that is directly in relation to its activities, the identification of a person must be protected, a person must consent to having data collected and the information must be collected lawfully and directly from that person (Australian Government, 2018). In permitted health situations the law does account for exemptions for collecting data that may be necessary for public health and safety without an individual's consent (Australian Government, 2018).

The Google COVID-19 mobility data set was created from data stored in the location history of personal mobile devices (Google 2020). Location data is collected by Google for various services including Google Maps (Google 2020). This data was obtained anonymously and combined so as no personal individual information is disclosed and to limit any identifiable data. Google (2020) utilises differential privacy techniques to prevent identification of individuals from collected data. Furthermore Google (2020) also acknowledges if low levels information were obtained from particular location areas, this information was excluded from the data set to reduce the risk of individuals being identified. Data was also collected with consent of the user given through turning on location services on a device (Google 2020). Google also acknowledges that the default setting for location services is off, therefore the user has to actively give consent for data usage by turning on location services. The user also has the option to clear or delete their location history at any time (Google, 2020)

The Our World in Data COVID-19 cases data set is comprised of publicly reported data supplied by governments and other validated sources (Hassel et al, 2020). Data is being collected for its primary purpose of managing the pandemic. Data is reported anonymously so as to not identify individuals and collected from countries as a whole so further limiting individual identification through location (Australian Government, 2018). Australian Law dictates that data can be collected without an individual's consent if required for a public health emergency (Australian Government, 2018). Given the public health crisis of COVID-19 this legislation would apply in this case (Australian Government, 2018).

Using data ethically is also of huge importance, information must be used in an honourable way and the effect of using data must be considered for all parties involved (Lawler, 2019). Given that individuals are giving consent for data to be collected through Google location services and that data is being used for an important cause and not maliciously is essential in ensuring this information is used in a proper way (Lawler, 2019).

x) References

1. Australian Government, 2020, '*Guide to Data Analytics and the Australian Privacy Principles*' web article, 21st March, Office of the Australian Information Commissioner - Australian Government, Sydney, Australia, viewed 14th January 2020, <https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-data-analytics-and-the-australian-privacy-principles/#s2-2-collection-of-personal-information-app-3>

2. Bhandari A 2020 'Introduction to Hadoop Ecosystem for Big Data and Data Engineering', *Analytics Vidhya*, weblog, October 23rd, viewed February 7th 2021, <https://www.analyticsvidhya.com/blog/2020/10/introduction-hadoop-ecosystem/>
3. Cao Y, Hiyoshi A, Montgomery S, 2020 'COVID-19 case fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data', *BMJ Open*, Vol 10, No e043560, <https://bmjopen.bmj.com/content/10/11/e043560>
4. Claeson M, Hanson S 2021, 'COVID-19 and the Swedish Enigma', *The Lancet*, Vol 397, No 10271, pp 259-261
5. Covidlive 2020, 'Covidlive', Australia, viewed 19th December 2020, <https://covidlive.com.au>
6. Duckett S, Mackey W 2020, 'No, Australia should not follow Sweden's approach to Coronavirus', *The Conversation*, weblog post, July 29th, viewed February 14th, 2021, <https://theconversation.com/no-australia-should-not-follow-swedens-approach-to-coronavirus-143540>
7. Franks P 2020, 'Coronavirus: Why the Nordics are the best for comparing out strategies' *The Conversation*, weblog post, April 4th, viewed February 14th 2021, <https://theconversation.com/coronavirus-why-the-nordics-are-our-best-bet-for-comparing-strategies-135344>
8. Fleming S 2020, 'What's a logarithmic graph and how does it help explain the spread on COVID-19?', *World Economic Forum*, weblog post, April 3rd, viewed February 11th 2021, <https://www.weforum.org/agenda/2020/04/covid-19-spread-logarithmic-graph/>
9. Google 2020, *Community Mobility Reports Help*, Google, California USA, viewed 13th December 2020, <<https://support.google.com/covid19-mobility#topic=9822927>>
10. Gupta A 2016, 'Comprehensive Introduce to Apache Spark, RDDs and Data Frames (using PySpark)', *Analytics Vidhya*, weblog post, September 19th, viewed February 7th 2021, https://www.analyticsvidhya.com/blog/2016/09/comprehensive-introduction-to-apache-spark-rdds-dataframes-using-pyspark/?utm_source=blog&utm_medium=hadoop-ecosystem
11. Hassel J, Mathieu E, Beltekian D, Macdonald B, Gianttino C, Ortiz-Ospina E, Roser M, Ritchie H, 2020 'A cross-country database of COVID-19 testing' *Scientific Data*, Vol 7, No 345, pp 1-7, <https://www.nature.com/articles/s41597-020-00688-8#citeas>
12. International Organisation for Standardisation, 2020, 'Popular Standards – ISO 3166 Country Codes', International Organisation for Standardisation, Geneva, Switzerland, viewed 13th December 2020, <https://www.iso.org/iso-3166-country-codes.html>
13. Kalron A, 2020, 'How do Hadoop and Spark Stack Up?', *logz.io*, weblog post, January 16th, viewed February 7th 2021, <https://logz.io/blog/hadoop-vs-spark/#:~:text=It's%20also%20a%20top%2Dlevel,an%20RDD%2C%20Resilient%20Distributed%20Dataset.>
14. Lawler B, 2019 'Five Global Trends in Data Ethics and Privacy in 2019' *Looker*, weblog post, 28th January, viewed 19th December 2020, <https://looker.com/blog/big-data-ethics-privacy#:~:text=Data%20privacy%20is%20responsibly%20collecting,based%20on%20your%20brand%20values>
15. Lee WW, Zanki W, Chang H, 2016 'An Ethical Approach to Data Privacy Protection', *ISACA Journal*, Vol 6, pp 1-9

16. LePage M, 2020, 'Is Sweden's Coronavirus strategy a cautionary tale or a success story?' *New Scientist*, No 3296, <https://www.newscientist.com/article/2251615-is-swedens-coronavirus-strategy-a-cautionary-tale-or-a-success-story/>
17. Lowy Institute 2020, 'Lowy Institute COVID Performance Index', Lowy Institute, Sydney Australia, viewed 11th February, <https://interactives.lowyinstitute.org/features/covid-performance/#rankings>
18. Phipps SJ, Grafton RQ, Kompas T, 2020, 'Robust estimates of the true (population) infection rate of COVID-19: a backcasting approach', *The Royal Society of Open Science*, Vol 7, No 200909, pp 1-12, <https://royalsocietypublishing.org/doi/10.1098/rsos.200909#d1e1736>
19. Stoica I, 2014, 'Apache Spark and Hadoop: Working Together', *Databricks*, weblog post, January 21st, viewed February 7th 2021, <https://databricks.com/blog/2014/01/21/spark-and-hadoop.html>
20. Tomlin A, 2019 'A quick guide to Swedish holidays and traditions', *Routes North*, weblog post, October 30th 2019, viewed February 14th 2021
21. Wang S, Liu Y, Hu T, 2020, 'Effect of the Change of Human Mobility Adherent to Social Restriction Policies and its Effect on COVID-19 Cases in Australia', *International Journal of Environmental Research and Public Health*, Vol 17, No 21, pp 7930 - 7946
22. Yilmazkuday H 2020 'Stay at Home Works to Fight Against COVID-19: International Evidence from Google Mobility Data', *Journal of Human Behaviour in the Social Environment*, Forthcoming, <<http://dx.doi.org/10.2139/ssrn.3571708>>