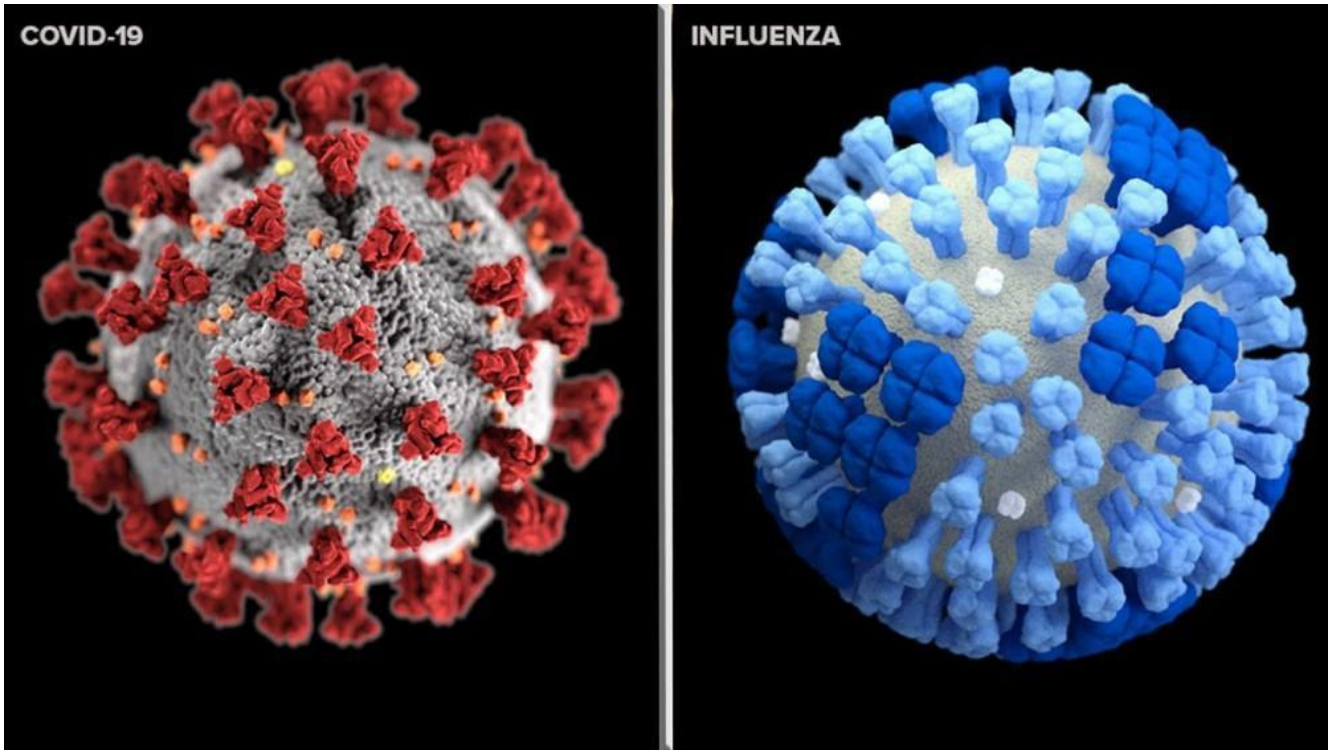

Image 1: Sourced from IM&P: COVID vs. The Flu

PROJECT REPORT

Analysis of the impacts of COVID-19 on Influenza in Australia

Abstract

An investigation into Influenza trends prior to COVID-19 and then proceeding to analyze the overall impact that COVID-19 has made on influenza within Australia. Investigation includes which sex and age group is at most risk of influenza infection and death. Final discovery should there still be concern for influenza while COVID-19 is still present today.

DATA7001_SEM1_2023 Group 19

Alyssa Du-Shane – s4751528
Pawan Parackal – s4792248
Shaodong Cui – s4745684
Xuanyu Qin – s4765132

We give consent for this report to be used as a teaching resource.

Executive Summary

This report presents the findings of an investigation conducted by Group 19, focusing on the effects of COVID-19 and Influenza in Australia. Our primary objective of this study was to generate a positive impact on the wider community by examining the potential effects of these viruses on individuals' lives. The report takes into consideration the interests of key stakeholders and aims to provide valuable insights for their benefit.

The investigation primarily focused on understanding Influenza trends preceding the COVID-19 outbreak and comparing the infections rates of both viruses from 2020 to the present. Through the analyse of reliable data from various government sources, we sought to comprehend the impact of COVID-19 on Influenza and identify the groups at higher risk of contracting Influenza.

Our research revealed that certain gender and age groups exhibit increased vulnerability to infection. Building upon these findings, we have formulated recommendations for the key stakeholders to consider when determining their future course of action. The data collection process involved meticulous sourcing from credible government sources. By utilizing the comprehensive data, we conducted an in-depth analysis, resulting in valuable insights that can contribute to informed decision making.

In conclusion, this report provides an essential examination of the effects of COVID-19 and Influenza in Australia. It offers valuable insights into the relationship between these viruses, their impacts on Influenza trends, and identifies at-risk populations. The recommendations provided aim to assist key stakeholders in developing appropriate strategies moving forward. In closing the reports comprehensive findings and recommendations serve as a foundation for evidence-based decision-making in the realm of public health and wellbeing.

Contents

1. Design Thinking

1.1 Introduction

Influenza is a respiratory illness caused by the influenza virus. Although it is widespread around the world and has symptoms like bad colds, such as headache, runny nose, cough and muscle aches, some people may not think of it as a virus that can lead to serious consequences. According to statistics, seasonal influenza kills as many as 650,000 people each year (World Health Organization, n.d.). However, with the COVID-19 pandemic, healthcare and social systems throughout the world have faced unprecedented challenges. As the COVID-19 and influenza are both respiratory viruses, and their clinical and epidemiological features have many similarities, it is crucial to ensure optimal management of both viruses as we anticipate their continued co-circulation (Chotpitayasunondh et al., 2021). Therefore, we would like to explore the relation between influenza and Covid-19 and analyze whether and how Covid-19 can impact influenza.

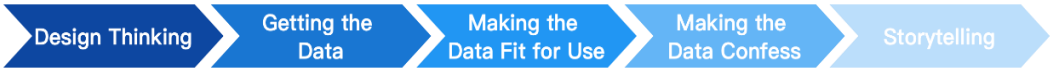

Figure 1: Data Science Process

Our research was carried out in accordance with the five steps of the data science process. The five steps of the data science process are shown in Figure 1. In this report, we will elaborate on our research on this topic, including the formulation of our question, finding datasets based on our question, cleaning the datasets, and using them for visualization. Finally, through our analysis of the results, we present recommendations for the different stakeholders.

1.2 Research Questions

To reach our objectives, we put forward two questions to our further exploration as follows:

1. How did Covid-19 affect influenza?
2. Which gender group and age are at higher risk, when investigating influenza?

1.3 Stakeholders

Based on our research, the key stakeholders are shown in the following table.

Table 1: Key Stakeholders

Stakeholder	Interests
The public	They can understand which gender and age group is sensitive to the influenza virus based on the findings and can get the suggestions to protect themselves.
Government and health departments	They can use the findings to understand the relationship between Covid-19 and influenza, and then they can develop public health policies and resources to control the epidemic and protect public health.

2. Getting the Data

2.1 Data Source

In terms of data source, they are all gathered by government, so we retrieved data from some official websites of Australian Bureau of Statistics, WTO and Australian Department of Health and Aged Care. See Appendix A: Code and datasets for a table which lists the names of the files we found for the following datasets.

1. Data for the total number of infections in influenza including gender from 2008 to 2021

   This source is laboratory-confirmed influenza infection data from the Australian Department of Health and Aged Care website. This file includes statistics on time, state, age group, sex, Indigenous status and type of influenza infection. There are approximately 1,071,548 records.

2. Data for the infection numbers in Covid-19 from 2020 to 2023

   This source is the national notifiable diseases surveillance system (NNDSS) fortnightly datasets from the Australian Department of Health and Aged Care

website., in which we can find the Covid-19 infection numbers. The number of files about Covid-19 infections in 2020, 2021 and 2022 are 26, and the number of files in 2023 are 5. There are approximately 6,230 records in total.

3. Data for the number of deaths due to influenza including gender from 2008 to 2018

   This source is the number of deaths data due to influenza from the Australian Bureau of Statistics website. This file contains the number of deaths including gender by year of occurrence from 2008 to 2018 under different causes of death. There are approximately 1,954 records.

4. Data for the number of deaths due to influenza including gender from 2011 to 2021

   This source is the number of deaths data due to influenza from the Australian Bureau of Statistics website. This file contains the number of deaths including gender by year of occurrence from 2011 to 2021 under different causes of death. There are approximately 1,960 records.

2.2 Data Privacy

The datasets used in our project were obtained from official websites of Australia which are all publicly available information. As a result, privacy concerns were not applicable to our project.

3. Making the Data Fit for Use

3.1 Data Quality Issues

3.1.1 Non-Unique Data and Accuracy

After investigating and checking our cleaned data we confirmed that there were no duplicate entries made within our dataset that could potentially skew our results. There were questions about the accuracy of our data from our Influenza laboratory confirmed cases and we ended up declaring the inaccuracy as incorrect data. The other dataset we ensured that we were investigating the most recent information available to ensure that our discovery would be as accurate as possible. Processes were followed to appropriately clean the data so the focus would be on our variables of interest. See *Appendix B* for Data Cleaning techniques.

3.1.2 Incomplete Data

As this report progresses the reader will discover a significant drop in Influenza cases from the NNDSS fortnightly reporting comparing COVID-19 and influenza cases. From 2020 onwards COVID-19 cases were the primary focus for stakeholders in Laboratory testing and State Governments. Therefore, we suspect that there were likely incomplete and inaccurate recordings of influenza cases. We discover that as COVID-19 becomes the new norm that influenza cases begin to rise again as testing continues for both viruses, not just searching and testing for COVID-19.

3.1.3 Incorrect Data

As mentioned above, the NNDSS data set (Influenza laboratory confirmed cases) it was discovered that there was an inaccuracy of data which flagged as incorrect and outdated. The first dataset included the years 2008-2018 and the second dataset included 2011-2021. The second dataset included higher numbers for the over lapping years of 2011-2018. As the second dataset had been published most recently and has been updated for those years, we decided to cut these years from the first dataset and merged across only 2008-2010 from the first dataset. Therefore, removing the incorrect and outdated data from our investigation.

Additionally, we discovered influenza and COVID-19 infections available on who however we did not use this data as the numbers did not correlate to what was published by the Australia government. We took inconsideration publish date to what numbers were provided and they were significant less. This is understood to be the delay of all laboratories releasing the numbers to the government in time for when the government needs to be sending the data to WHO for release. There WHO data cannot be used in this instance as it is simply not reliable as it hasn't been updated when the government updates with new figures.

3.1.4 Inconsistent Data

During the check the data phase we discovered that there were inconsistencies with our data that was recording COVID-19 and influenza cases from 2020 – 2023(NNDSS fortnightly reports). There were many fortnights were COVID-19 was spelt differently or also included *stars after the word covid. This resulted in us going back to the cleaning phases and amending the sections so there was constancy in spelling which then resulted in the data being able to be merged successfully for each year.

3.2 Data use Considerations

3.2.1 What considerations have been implemented.

As part of our discovery process in "*is our data fit for use*" we took 8 key aspects into consideration to ensure that we are using quality data to provide quality results. The key considerations that we have implemented are:

1. Accuracy: ensuring that our data is correct and from reliable sources which is free from errors and inconsistencies.
2. Completeness: there are no missing values or significant gaps that can negatively impact of research.
3. Consistency: there is consistent values across our datasets with no redundant information or discrepancies. Inconsistent data has been rectified by using the most recent update from the government website.

4. Timeliness: our data has been extracted from the most recently published datasets available to ensure that we are using fresh up to date data in our research.
5. Validity: our data conforms to standards and formats required to reduce errors and inaccuracies.
6. Uniqueness: data sets have been checked after cleaning to ensure that there are no duplicated fields which could potentially distort our discovery in modelling and visualisation.
7. Relevance: datasets have been assessed as to whether the attributes and variables are appropriate for our intended analysis and align with our requirements.
8. Accessibility and Security: all data is freely available to the public and has already been cleaned by the data issuer for any identifiable or sensitive personal information that could put any person at risk. Therefore, our data being used is already protecting all individual's data that has been collected.

4. Making the Data Confess

4.1 Visual Exploratory Data Analysis

When working on a data science project, using statistical methods, and visualizing data through plotting tools, we can make the data confess and find out the trends and relationships behind different data.

4.1.1 Trends of Covid-19 and Influenza infections

In figure 2 we have demonstrated the quarterly trends of both Covid-19 and influenza infection cases from the start of the pandemic in 2020 and up until most recently April 2023. Visually we can see that from the 2nd quartile in 2021 is when Covid-19 cases began to climb and peaked after a year of climbing before cases began to gradually decrease from the 2nd quartile in 2022.

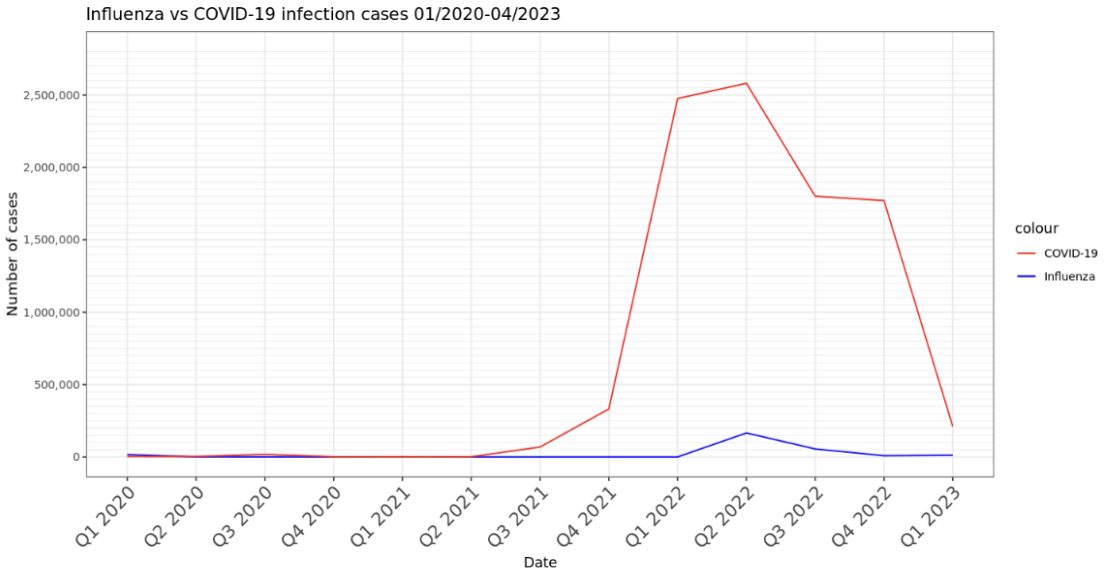

Figure 2: Number of infection influenza and covid-19 cases

To better observe the changing trend of the period from Q1 2020 to Q2 2021 in Figure 2, we zoom in on the image of this period, as shown in Figure 3. As we can see, following the outbreak of COVID-19, the number of influenza infections has significantly decreased and remained at a low level. This seems to indicate an impact of the COVID-19 outbreak on influenza. To confirm our speculation, section 4.1.2 provides a more detailed trend analysis of influenza.

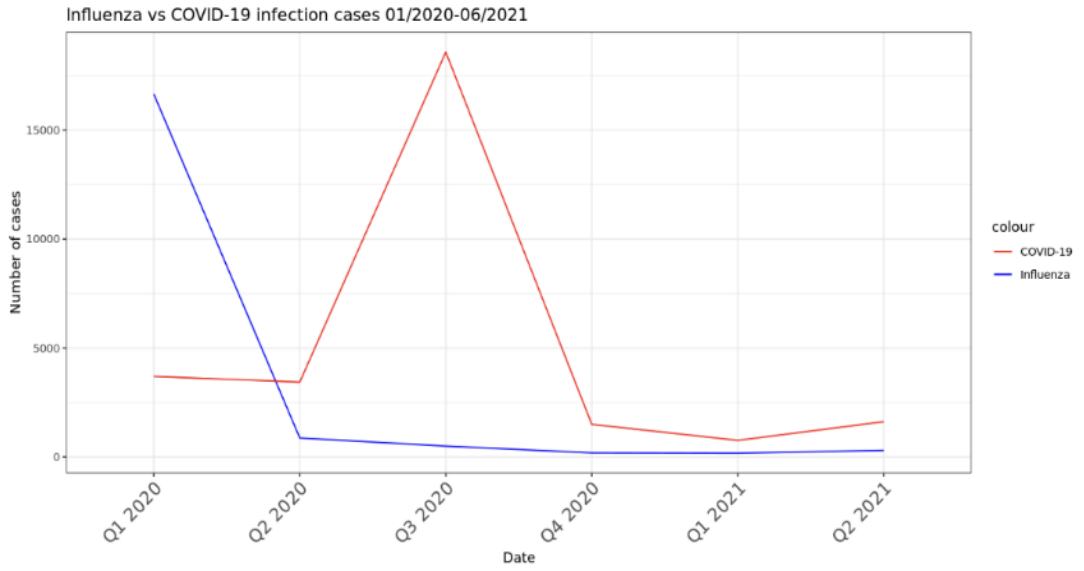

Figure 3: Number of infection influenza and covid-19 cases （Q1 2020 to Q2 2021）

4.1.2 Trends of Influenza

An important part of our discovery to answer our research Question on the affect Covid-19 had on Influenza was to also investigate the trends of Influenza leading up to the Covid-19 pandemic.

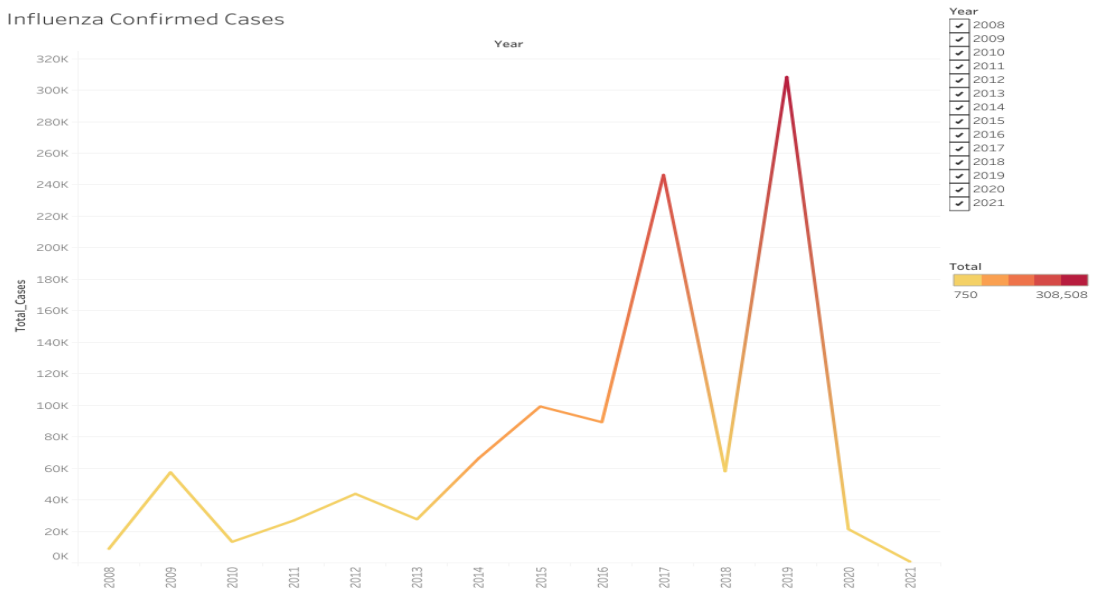

Figure 4: Influenza confirmed cases (2008 to 2021)

Figure 4 shows the annual trend of the number of influenza infections from 2008 to 2021. Based on this graph, we can see that the number of influenza infections dropped dramatically in 2020 when the COVID-19 broke out. Therefore, we will delve

into the data to examine the trend of the number of influenza and COVID-19 infections in 2020 and 2021.

Figure 5, Figure 6, and Figure 7 (zoomed-in view of Figure 6) clearly depict the trends of influenza and COVID-19 in 2020 and 2021. We can observe that within a few months after the outbreak of COVID-19, it demonstrated remarkable spread and dominance. Especially in 2021, when the number of COVID-19 cases sharply increased, the confirmed cases of influenza were so few that they could be almost negligible. Therefore, we can conclude that the emergence of COVID-19 has had a great impact on Influenza.

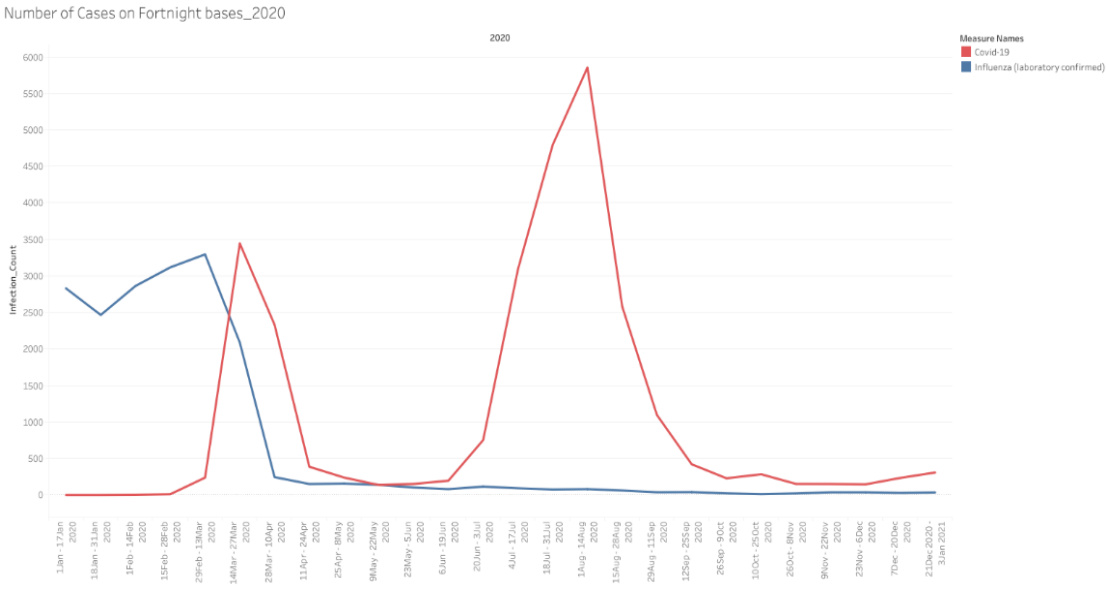

Figure 5: Confirmed cases of influenza and covid-19 (every fortnight in 2020)

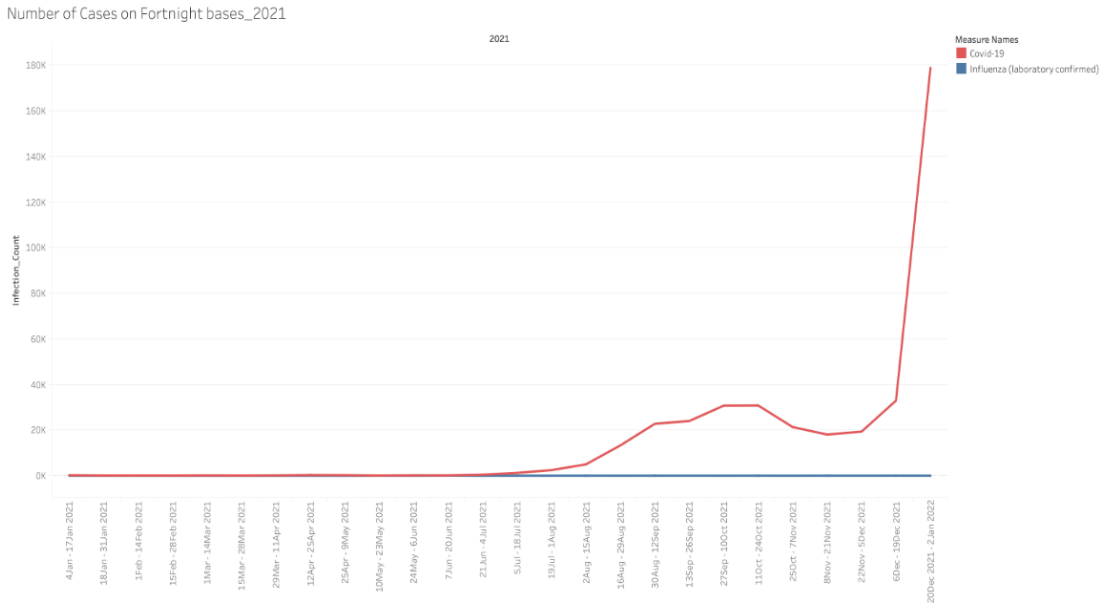

Figure 6: Confirmed cases of influenza and covid-19 (every fortnight in 2021)

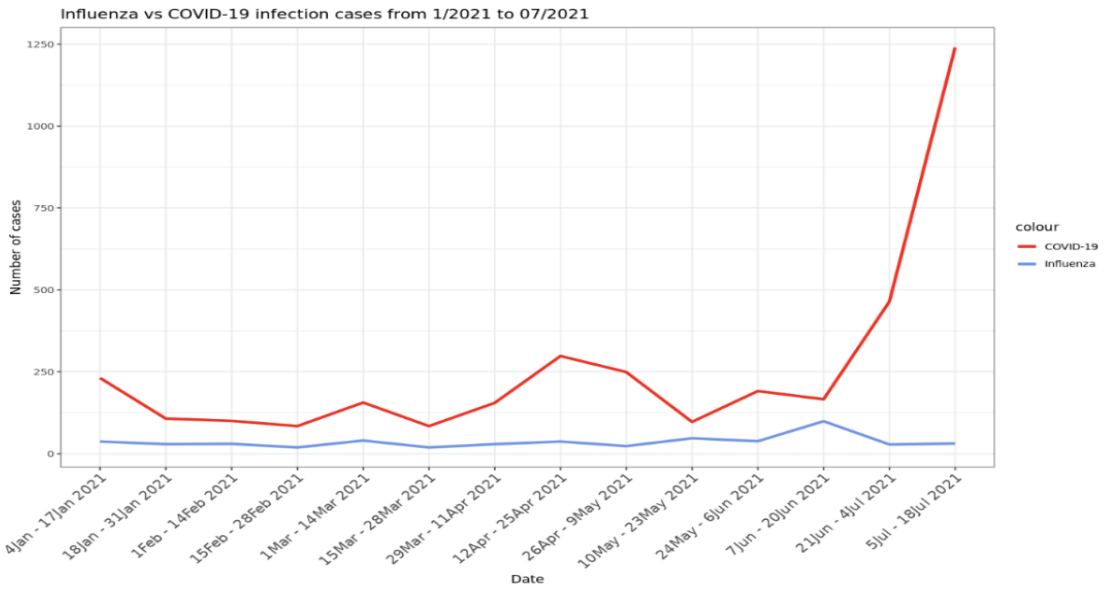

Figure 7: Zoom in view based on Figure 4.1.5 (Jan to July in 2021)

4.1.3 The Impact of COVID-19 on the Gender Ratio of Influenza Infections

Next, we will look at the gender ratio of influenza infections through data. We can see from the Figure 8, from 2008 to 2020, the proportion of women who contracted influenza has always been higher than that of men. Until 2021, the proportion of the two has reversed, and the proportion of men infected exceeds that of women for the first time.

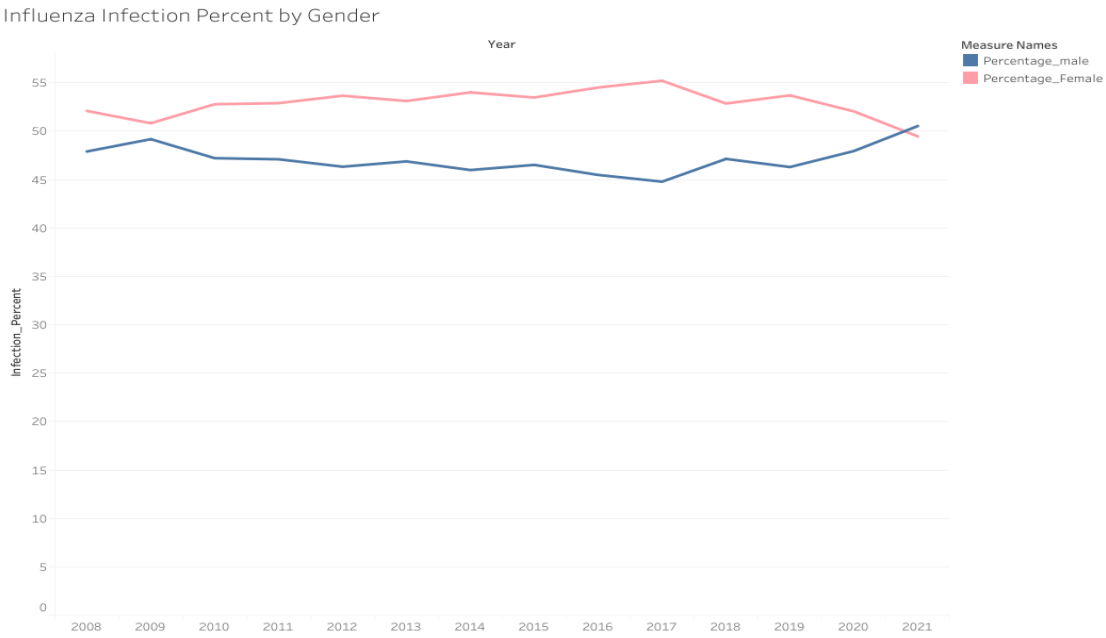

Figure 8: Influenza infection percent by gender

4.1.4 Identifying the age groups among females prone to higher susceptibility to infection.

Based on the above analysis, we found that women are more susceptible to influenza infection, so it is worthwhile to analyse the infection data of women in each age group and find out the most vulnerable groups. We can see from the Figure 9 that each colour represents a different age group, with females aged 0-9 (blue) having the highest number of infections in most years.

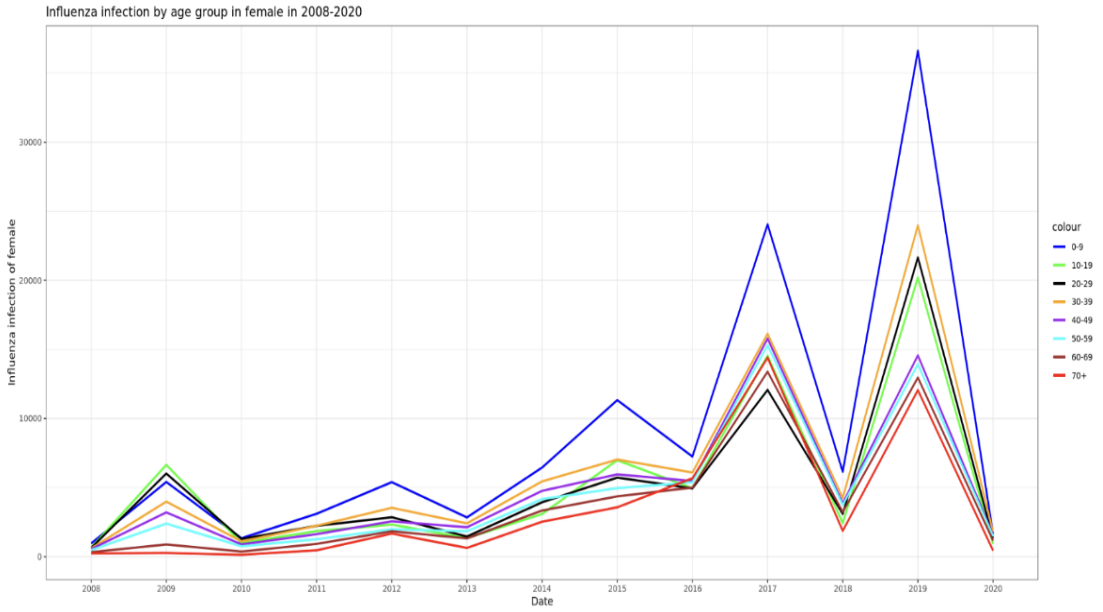

Figure 9: Infection cases in different female ages

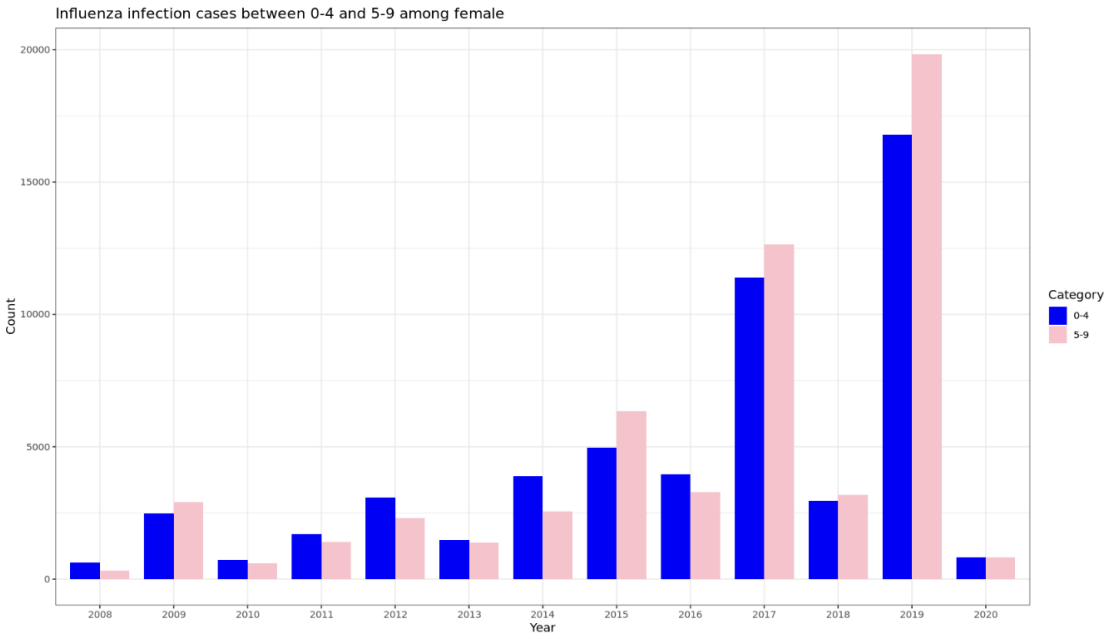

Figure 10: Influenza infection cases between 0-4 and 5-9 among women

For a better visual effect and understanding we visualize the data by dividing the age of each 10 years old, and then we focus on the age group of 0-9 years old to analyse more detailed data. The Figure 10 shows that females aged 5-9 appear to be more susceptible to influenza compared to those aged 0-4. This pattern is particularly pronounced during years of significant influenza outbreaks, such as 2017 and 2019.

4.2 Model Generation

4.2.1 Linear Regression model

We tried to analyse whether the number of men and women infected with influenza is related to age in different years. The following are the model analysis results. We simulated data for all years from 2008-2020, and only show the results for 2009 here. It can be seen from the results that the number of infections gradually decreases with increasing age.

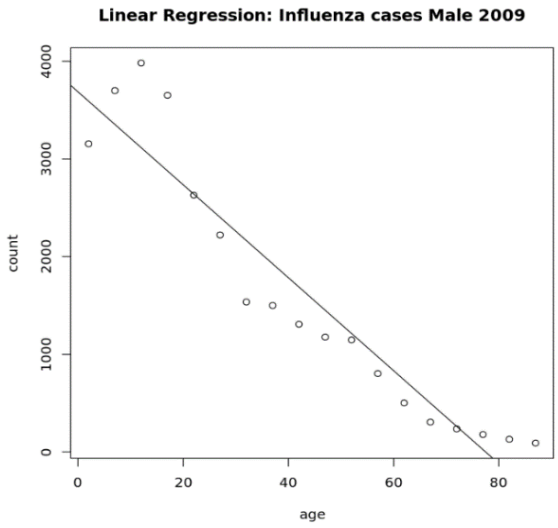

Figure 11: Linear Regression Female '09

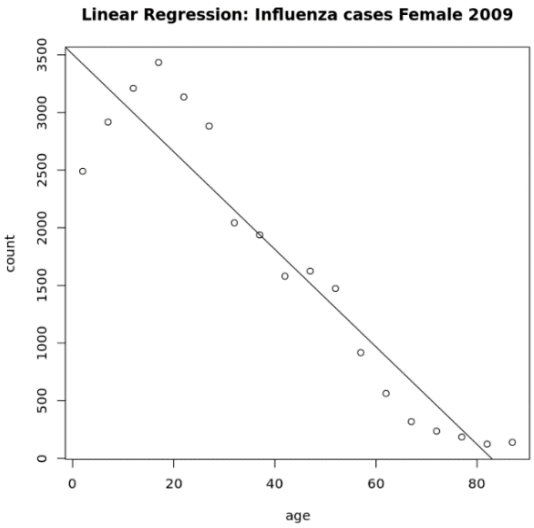

Figure 12: Linear Regression Male '09

```
Call:
lm(formula = count ~ age, data = female2009)

Residuals:
    Min      1Q  Median      3Q     Max
-930.16 -211.94  -32.49  201.47  648.25

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3504.746    184.608   18.98 2.13e-12 ***
age          -42.294      3.584  -11.80 2.62e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 394.4 on 16 degrees of freedom
Multiple R-squared:  0.8969,    Adjusted R-squared:  0.8905
F-statistic: 139.3 on 1 and 16 DF,  p-value: 2.621e-09
```

Figure 13: Linear Regression Coefficients Female '09

```
Call:
lm(formula = count ~ age, data = male2009)

Residuals:
   Min     1Q Median     3Q    Max
-628.4 -265.0 -119.1  296.9  867.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3686.655    204.552   18.02 4.73e-12 ***
age          -47.570      3.971  -11.98 2.11e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 437.1 on 16 degrees of freedom
Multiple R-squared:  0.8997,    Adjusted R-squared:  0.8934
F-statistic: 143.5 on 1 and 16 DF,  p-value: 2.111e-09
```

Figure 14: Linear Regression Coefficients Male '09

4.2.2 Scatter 3D

A scatter point is selected in the figure, and the information is x: age, y: year, z: count. For ease of visualization, we use the middle number of each age range to represent this age group.

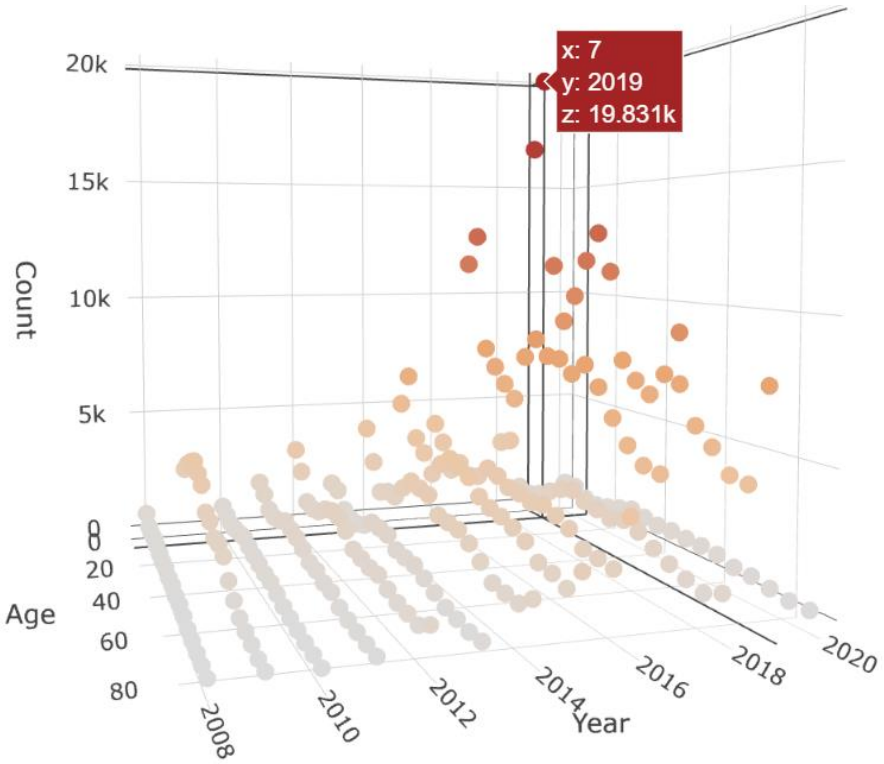

Figure 15: Scatter 3D distribution map of the age group of female influenza infection per year

5. Storytelling

In the storey telling phase of the report, we take a deep dive into discovery and the findings from our COVID-19 and Influenza data to help answer our research question. Our storey telling has been broken down into several sections to highlight our understanding of the research findings and how this can help the interested stakeholders. The goal of our storytelling is to inform and make recommendation based on the significance of our research that has been conducted by Group 19.

5.1 Analysis of the Influenza

In this section, we analyse the infection rates of influenza from the years 2008 to 2021, aiming to uncover significant patterns and shed light on the infection rates during this period leading up to the global pandemic, COVID-19. To gain a comprehensive understanding of the general trend, we plotted a line graph as in figure 4 which depicts the yearly infection counts from 2008 to 2021. Our analysis revealed several noteworthy observations within this period.

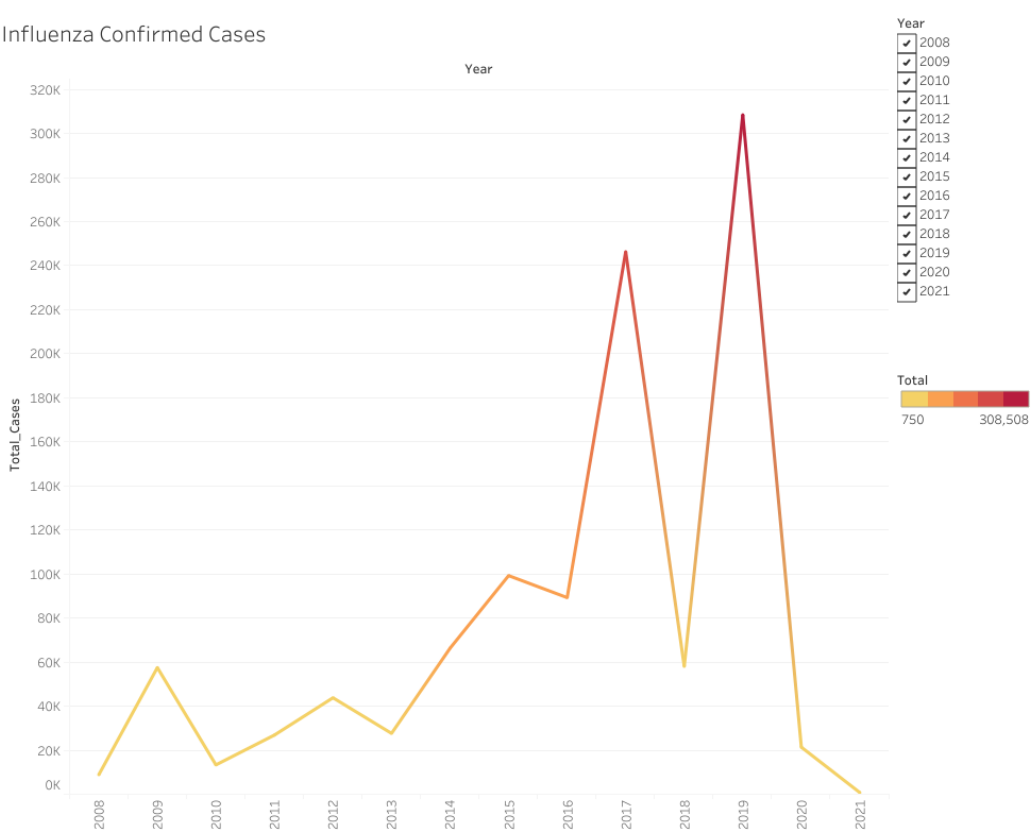

Figure 16: Influenza confirmed cases (2008 to 2021)

Firstly, from 2008 to 2017, the infection rates of influenza demonstrated a consistent upward trend, with a surge in cases observed in specific years. Notably, the year 2017 witnessed a significant peak, recording an infection count of 248,120, a six-fold increase compared to previous years. During this peak period, the high number of influenza cases can be attributed to the prevalence of the influenza strain A(H3N2), which disproportionately affected infection and death rates in the elderly (Australian Government Department of Health, 2017). The 2017 seasonal influenza vaccine had low effectiveness against this strain, partly due to weaker immune responses in the elderly and changes in the virus itself (Australian Government Department of Health, 2017). These factors combined with the presence of virulent strains already circulating in Australia, contributed to the surge in cases, emphasizing the need for improved vaccine strategies moving forward (Dalzell, 2018).

In the following year, 2018, the vaccine was comprehensive enough to cover all strains that had been circulating and was extremely effective (Dalzell, 2018). However, in 2019, there was another extreme peak with a count of 313,033, which was much higher than the previous peak in 2017 and this flu season did commence much earlier than other years. Again, this increase in infection rates can be attributed to presence of virulent strains not adequately covered by the vaccine (CDC, 2021) in addition to the longer flu season as observed in figure 16. As Influenza has an ever-changing nature, which can modify itself, it compromises our immunity developed from previous infections. Therefore, it is required for the flu vaccine to be periodically updated to match circulating strains of the virus (CDC, 2021).

To further our understanding of the infection numbers over the course of the year, we continued to assess the seasonal variation of the influenza virus in a linear graph (Figure 16) illustrating the monthly trends from January to December for the years 2016 to 2021. This visual representation revealed significant patterns and insights into the temporal distribution of influenza infections. Across the five-year period, we observed that influenza cases consistently peaked between the months of August and September, followed by a gradual decline. These months emerged as the most infectious periods for influenza, warranting heightened vigilance and targeted

prevention strategies during this timeframe. This recurring pattern suggests a strong correlation between seasonal factors and the transmission dynamics of the virus.

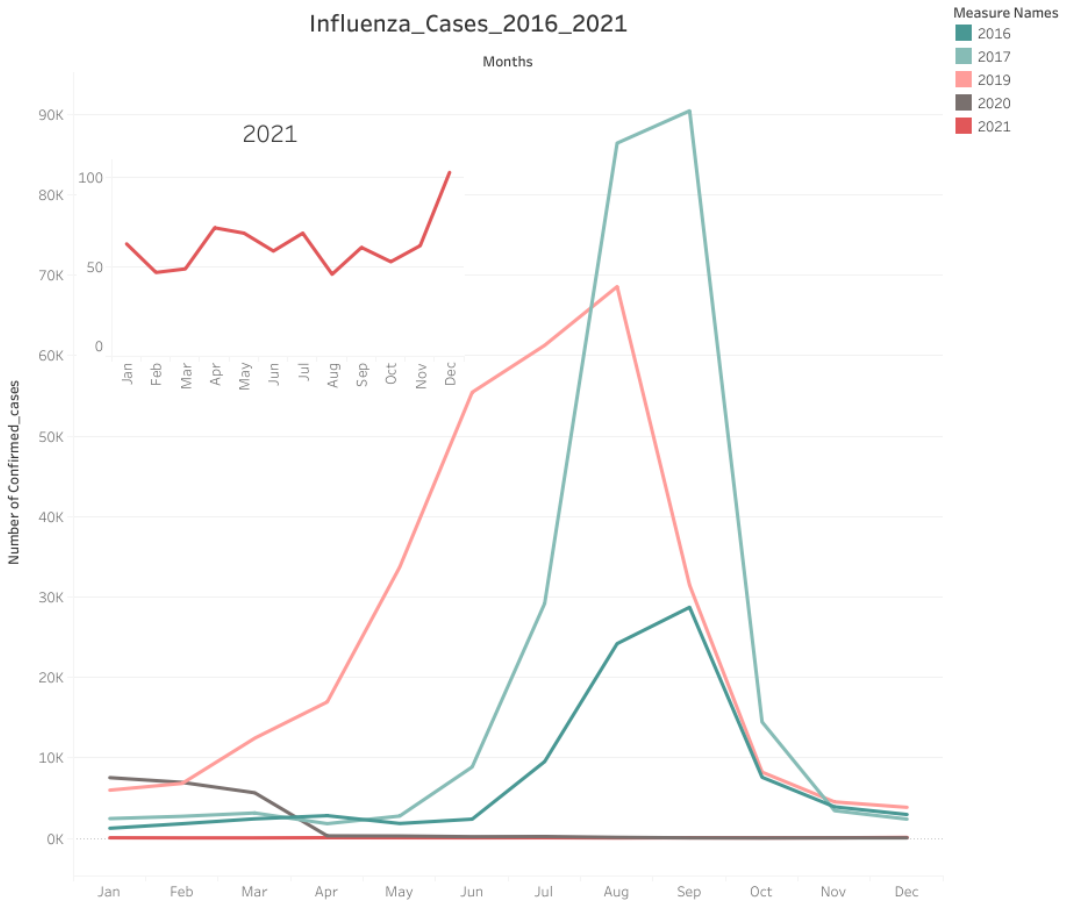

Figure 1176: Seasonal Influenza Confirmed Cases 2016-2021

To provide a comprehensive analysis, we included a separate graph specifically depicting the influenza cases in the year 2021. Notably, the graph for 2021 displayed a substantial decrease in the number of cases compared to the previous years. The availability of this seasonal information is of utmost importance for government and health officials in predicting and preparing for future influenza seasons. By anticipating the peak infection periods, stakeholders can allocate resources, optimize vaccination campaigns, and implement targeted interventions during the most vulnerable months. This enables a more proactive response to minimize the impact of influenza outbreaks on public health.

5.1.1 State-Wise Analysis of the Influenza on peak year

To gain insights into the regional variations, we conducted an analysis of influenza cases across different states in Australia during the peak year of 2019. Our focus was on monthly data, with particular emphasis on the period from May when the flu season typically intensifies. During the peak months, notable variations in infection counts were observed among different states. New South Wales stood out with the highest infection count, reaching a peak of 29,480 cases in August. This finding highlights the heightened risk faced by residents of New South Wales during the peak influenza season. Queensland closely followed, reporting a peak infection count of 21,901 also occurring in August, indicating a significant risk within the state. In Victoria, there was a peak infection count of 13,572 in August, which also emphasizes the need for targeted interventions to protect residents.

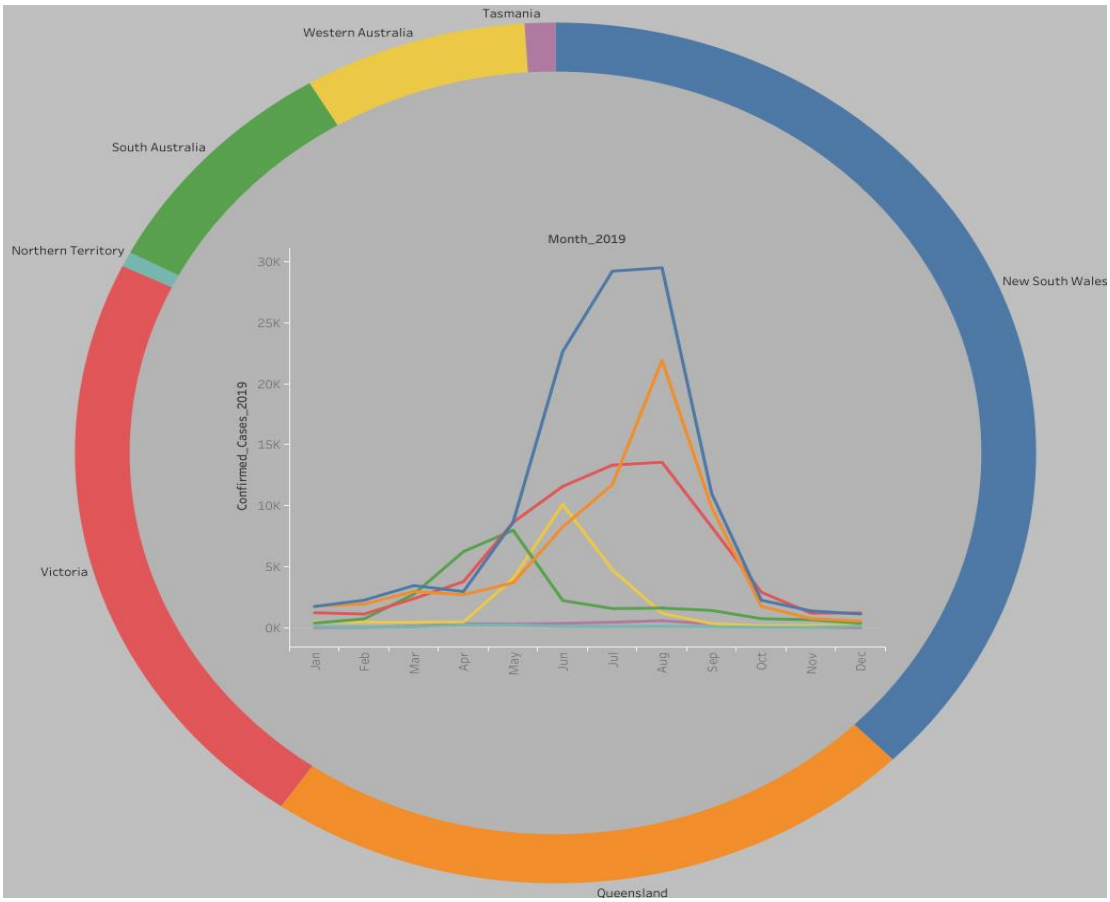

Figure 18: State Wise Influenza Confirmed Cases (2019)

These findings have significant implications for state government stakeholders responsible for public health and well-being. By identifying regions and months with

the highest infection counts, stakeholders can prioritize vaccination efforts and allocate resources strategically. Encouraging residents of New South Wales, Queensland, and Victoria to get vaccinated becomes paramount in reducing the impact of influenza outbreaks within these regions. The availability of state-specific data offers a valuable opportunity for targeted messaging and education campaigns. Government stakeholders can leverage this information to raise awareness about the importance of influenza vaccination, emphasizing the heightened risk during specific months. Collaborating with healthcare providers and community organizations allows stakeholders to amplify vaccination initiatives and promote proactive measures to protect the population.

5.2 Analysis of the effects of COVID-19 on Influenza

In this section, we delve into the impact of the COVID-19 pandemic on the trend of influenza cases, shedding light on the significant shifts observed. Previously on an upward trajectory, the trend of influenza cases underwent notable changes in 2020 and 2021. To thoroughly examine this shift, we closely analyse the graph of influenza cases on a fortnightly basis during these years, drawing comparisons to the rising trend of COVID-19 cases during the same period. In our exploration, we delve in the factors behind the decline in influenza cases and the concurrent surge in COVID-19 cases, substantiating our analysis with relevant references.

Upon observing the graph, in figure 5, of influenza cases on a fortnightly basis in 2020, a distinct pattern emerges. Following the typical influenza season, where cases usually peak from June to September (Australian Government Department of Health, 2021), the trend of influenza instead shows a gradual decline after a summer peak and dropping off at the end of March. This deviation from the anticipated upward trend can be attributed to the profound impact of the COVID-19 pandemic. As depicted in Figure 5, a notable trend emerges where the rise in COVID-19 cases coincides with a significant decline in influenza cases The implementation of vital public health measures, such as social distancing, mask-wearing, and travel restrictions, aimed at mitigating the spread of COVID-19, likely contributed to reducing influenza transmission as well. These comprehensive measures, coupled with heightened public awareness of respiratory hygiene, may have effectively curtailed the circulation of influenza viruses.

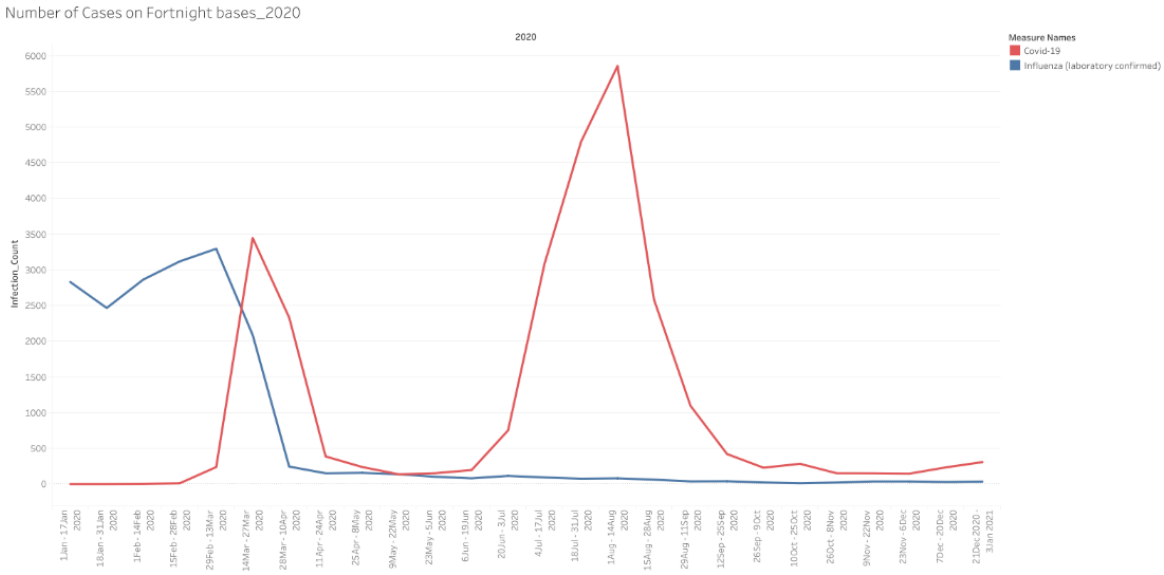

Figure 19: Confirmed cases of influenza and covid-19 (every fortnight in 2020)

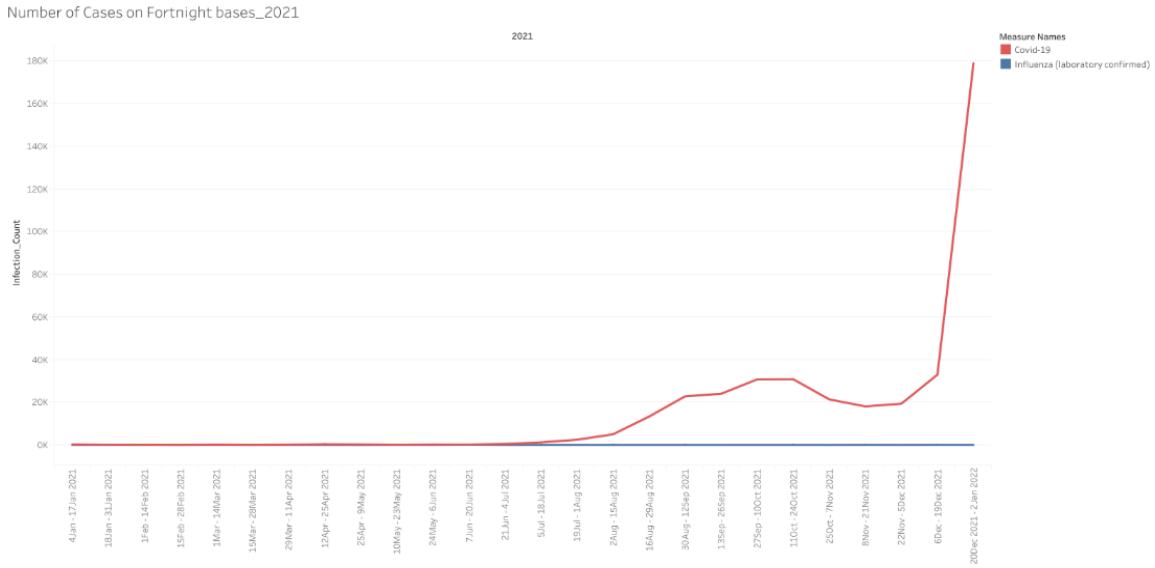

Figure 20: Confirmed cases of influenza and covid-19 (every fortnight in 2021)

Moving to 2021, a continuation of the downward trend in influenza cases becomes evident when examining the graph (Figure 6), despite the usual expectation of an increase during the typical influenza season. In stark contrast, the graph reveals a gradual increase in COVID-19 cases from June onwards, ultimately skyrocketing in November.

The contrasting trend observed can be attributed to the disparities in the transmissibility and impact of COVID-19 and influenza. COVID-19, caused by the novel coronavirus SARS-CoV-2, exhibits higher transmissibility and a more severe disease course compared to seasonal influenza. The implementation of stringent measures to control

COVID-19, such as lockdowns, widespread testing, contact tracing, and vaccination campaigns, likely played a significant role in suppressing the spread of influenza as well (WHO, 2021).

5.3 Analysis of the Influenza based on gender.

In this section, we examine the gender disparities in influenza susceptibility, drawing insights from the analysis of data spanning from 2008 to 2021. The examination of this data reveals that women exhibit a higher susceptibility to influenza compared to men as depicted in figure 18. Multiple factors contribute to this observed disparity. Firstly, hormonal differences between men and women may play a crucial role. Research suggests that estrogen, a hormone predominantly present in women, can affect the immune response and render women more susceptible to viral infections, including influenza (Smith et al., 2020). Also, genetic, and physiological differences contribute to the gender-based differences in influenza susceptibility. Variances in immune responses, cytokine production, and lung structure between men and women may contribute to varying level of vulnerability to influenza (Klein & Flanagan, 2016).

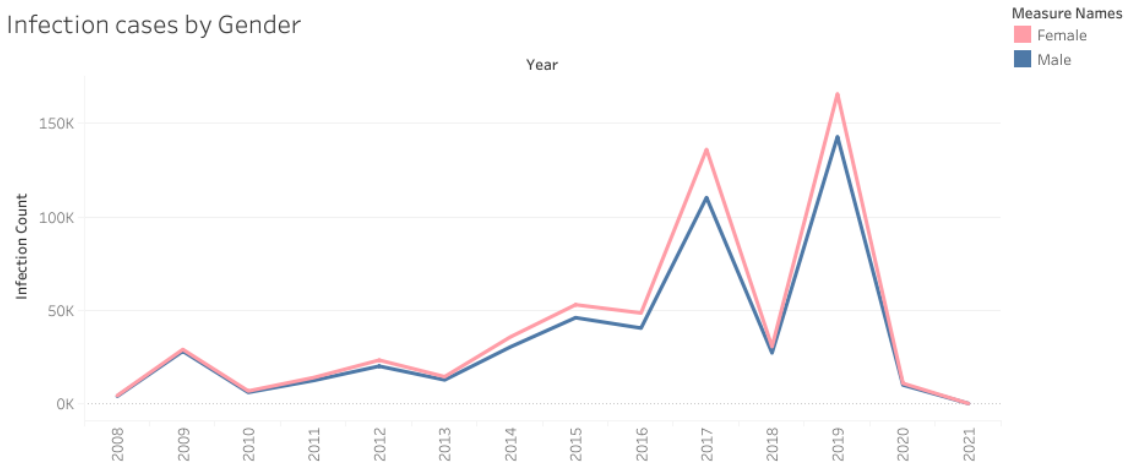

Figure 218: Influenza Confirmed Cases by Gender '08-'21

Furthermore, behavioural, and social factors could further influence the gender disparities in influenza infection rates. Studies have consistently shown that women tend to seek healthcare services more frequently and have more contact with healthcare settings, increasing their exposure to influenza viruses (Simmelink- Mooij et al. 2018). Additionally, societal roles and responsibilities may contribute to differential exposure,

with women often fulfilling caregiving roles and having higher levels of contact with children, who are known to be significant transmitters of respiratory viruses like influenza (CSC, 2021).

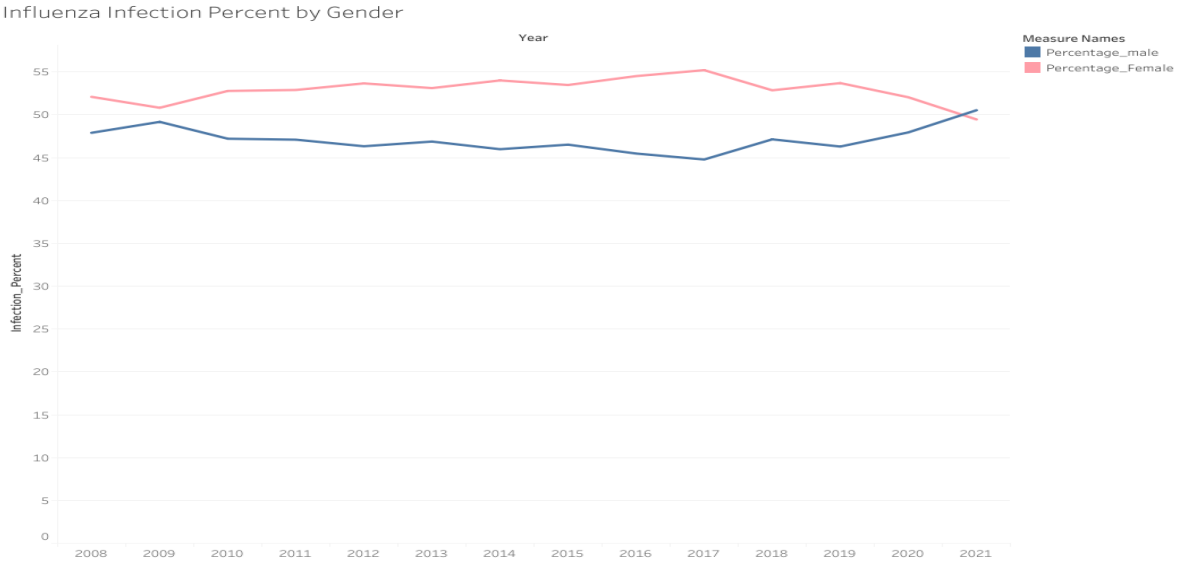

Figure 22: Influenza infection percent by gender

5.3.1 Mortality Rate

Analysis of the mortality rate for influenza cases spanning from 2008 to 2021 (Figure 19) uncovers a notable gender disparity, with females demonstrating a higher susceptibility to influenza-related complications and mortality compared to males. This finding aligns with previous research that has consistently highlighted the increased vulnerability of women to sever outcomes and higher mortality rates associated with influenza (Klein & Flanagan, 2016). Intriguingly, in the years 2016 to 2017, the mortality rate for females surpassed 50%, indicating a particularly elevated risk during that specific period. Understanding the underlying factors contribution to these gender difference in influenza mortality is crucial for healthcare stakeholders and policymakers. Hormonal influences, including the impact of sex hormones on immune response and inflammatory processes, have been suggested as a potential contributing factor to the higher mortality rate among women (Klein & Flanagan, 2016).

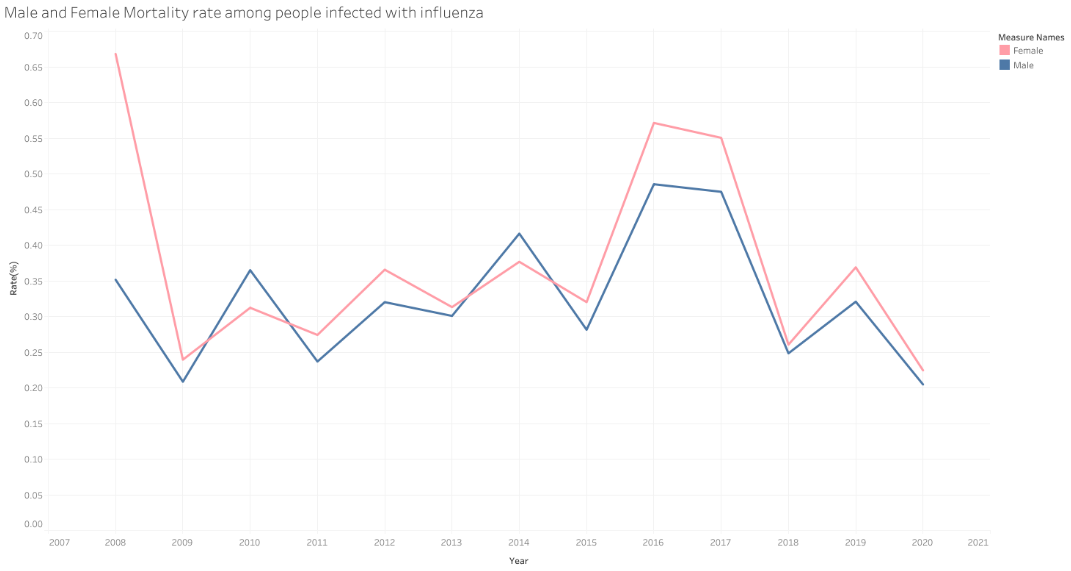

Figure 23: Influenza mortality Rate by Gender '08-'21

Understanding the gender disparities in influenza mortality is crucial for developing targeted interventions and improving public health strategies. Stakeholders, such as the Australian Government Department of Health and Aged Care, play a pivotal role in addressing these disparities and reducing influenza-related mortality among women. Implementing initiatives specifically tailored to women can be effective in mitigating the gender gap. For instance, promoting vaccination campaigns that prioritize and encourage women to receive the flu vaccines can significantly reduce their risk of sever outcomes (Centre for Disease Control and Prevention, 2021). Additionally, raising awareness about the importance of early diagnosis and timely treatment among healthcare providers and the general population can improve outcomes for female influenza patients. Furthermore, conducting further research to elucidate the underlying mechanisms contributing to the gender disparity in influenza mortality rates is essential for informing evidence-based interventions and refining public health strategies. By integrating these comprehensive approaches, stakeholders can work towards minimizing the gender gap in influenza mortality and enhancing overall public health outcomes.

5.4 Analysis of the Influenza based Age group.

Analysis based on age groups provides valuable insights into the vulnerability to influenza across different demographic segments. Examining the data from 2015 to 2019

(Figure 9), we identified several noteworthy trends regarding influenza infection rates among different age groups emerged. Among the eight age groups analysed, the 0 to 9-year-old age group consistently exhibited the highest infection rates, with a significant peak observed in 2019, surpassing 35,000 cases. This finding aligns with previous research highlighting the increased susceptibility of children to influenza infections (Smith et al., 2020). Additionally, the age group of 30 to 39 years demonstrated a considerable number of influenza cases, with approximately 24,000 cases reported during the same period. This observation suggests that individuals in their thirties may also be more susceptible to influenza infections compared to other age groups. Further investigation is warranted to explore the underlying factors contributing to this observation, such as lifestyle, occupational exposure, or specific immunological characteristics within this age range.

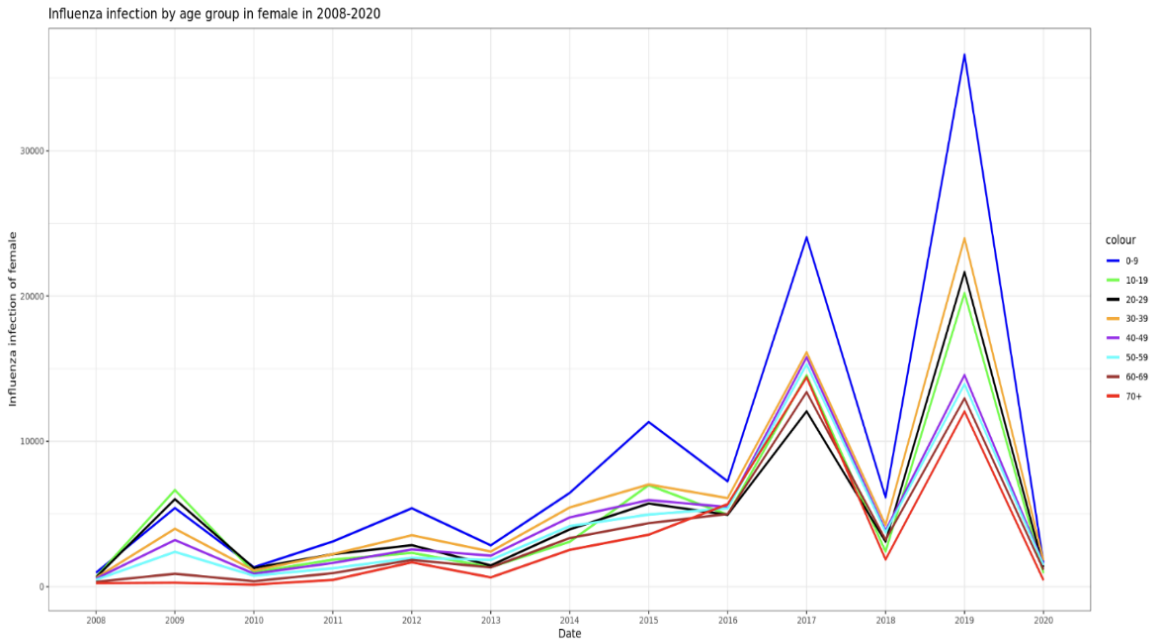

Figure 24: Infection cases in different female ages

Further analysis of the age groups within the 0 to 9-year-old range, as seen in figure 9 above, revealed an interesting pattern, with the subgroup of children aged 5 to 9 appearing to be more vulnerable to influenza infections. This finding provides a more nuanced understanding of the susceptibility of young children to influenza and highlights the importance of considering narrower age ranges when assessing infection rates.

The identification of the 5 to 9-year-old age group as particularly vulnerable to influenza has important implications for public health interventions. Stakeholders, including healthcare providers and public health agencies, can utilize this information to develop targeted prevention and control strategies. Vaccination campaigns specifically targeting children aged 5 to 9 can help reduce influenza transmission within this subgroup and protect their health. Furthermore, implementing educational programs in schools and childcare settings to promote hygiene practices and raise awareness about influenza prevention can further contribute to mitigating the burden of influenza in this age group.

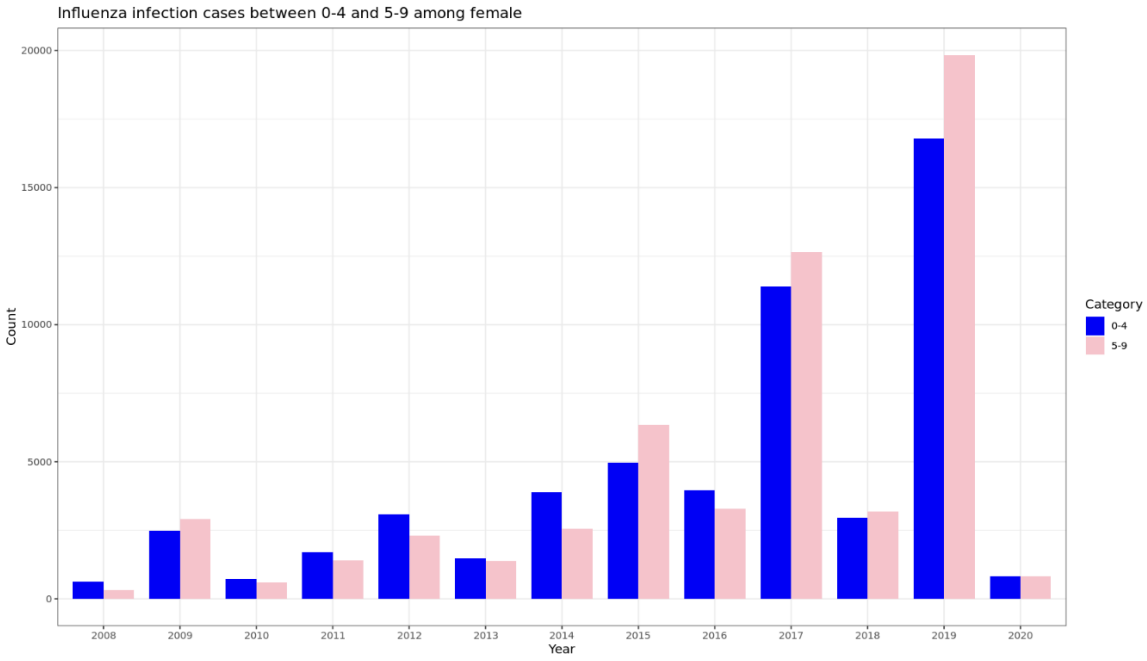

Figure 25: Influenza infection cases between 0-4 and 5-9 among women

Understanding the specific vulnerabilities of different age subgroups allows for more effective and tailored interventions. By focusing on the age group most at risk, resources can be allocated more efficiently to prevent and control influenza outbreaks. It is crucial for stakeholders to collaborate and implement evidence-based strategies to protect those aged 5 to 9, during influenza seasons.

6. Management of Change

Feedback provided by professors and fellow cohort was implemented into the final presentation slides and assisted us in perfecting our report for the reader. This feedback was very useful and was taken onboard with consideration to realistic timeframe of presentation duration and timeliness to make further investigations into our research. As a team we would have appreciated the time to expand on all the great suggestions however time was our biggest constraint. This list consists of the feedback we selected to implement in our presentation and into our report:

1. Adding in values at the peak of our visualisations when discussing making the data confess.
2. Images having a clearer reference from source.
3. Redefining our research questions/ problem we will solve with data.
4. Explaining in more detail the process of making our data fit for use.
5. Problems encountered in inconsistent data.
6. Added a gender and age group at most risk of influenza infection.
7. Removed recommendations and added these to the appropriate section in our story telling phase.
8. Added a 3D visualisation as part of our discovery.

7. Conclusion and Recommendations

In conclusion, the findings of this investigation shed light on the effects of COVID-19 and Influenza in Australia, providing valuable insights for key stakeholders and policymakers. Based on these findings, we propose the following recommendations to improve influenza preventions and control strategies:

1. **Periodic Updates of Influenza Vaccine**: Given the ever-changing nature of influenza and its ability to modify itself, it is crucial to regularly update the flu vaccine to match circulating strains of the virus (CDC, 2021). Government health agencies should collaborate with healthcare providers to ensure the timely development and distribution of updated vaccines, thereby enhancing population immunity against the evolving influenza strains.
2. **Heighted Awareness of Influenza Vaccination**: Government stakeholders should prioritize raising awareness about the importance of influenza vaccination, particularly leading up to the months when the risk of infection is elevated. Targeted messaging campaigns, education initiatives, and collaborations with healthcare providers and community organizations can effectively disseminate information and encourage individuals to "*Go Get Vaccinated",* reducing the impact of influenza outbreaks within respective regions.
3. **Reduction of Influenza-Related Mortality in Women**: Stakeholders, including the Australian Government Department of Health and Aged Care, should consider implementing initiatives aimed at reducing influenza-related mortality in females. This may involve developing and promoting vaccination campaigns specifically tailored to women, emphasizing the importance of early diagnosis and treatment, and conducting further research to understand the underlying factors contributing to the gender disparity in mortality rates.
4. **Targeted Vaccination Campaigns for Children Aged 5 to 9**: Given the vulnerability of children in the 5 to 9-year-old age group to influenza infections, stakeholders should prioritize targeted vaccination campaigns for this subgroup. By focusing resources on ensuring high vaccination coverage among children aged 5 to 9, influenza transmission within this population can be significantly reduced, contributing to their overall health

and well-being. An idea for campaign could include bring the family together to go get vaccinated for the upcoming flu season.

In addition to these recommendations, ongoing surveillance and research efforts should be conducted to monitor influenza trends, identify emerging strains, and assess the effectiveness of prevention and control measures. Collaboration between government agencies, healthcare providers, and research institutions is essential for implementing evidence-based strategies and adapting to the evolving landscape of influenza. By implementing these recommendations, Australia can enhance its influenza prevention and control measures, reduce the burden of influenza-related illnesses, and protect the health and well-being of its population.

References

Australian Government Department of Health. (2021). Influenza (flu): Information for health professionals. Retrieved from https://www.health.gov.au/resources/publications/influenza-flu-information-for-health-professionals

Australian Government Department of Health. (2017). Australian Influenza Surveillance Report - 2017 Season Summary. Australian Government Department of Health. Retrieved from https://webarchive.nla.gov.au/awa/20190510054958/http://www.health.gov.au/internet/main/publishing.nsf/Content/ozflu-surveil-2017-final.htm

Centers for Disease Control and Prevention. (2021). Influenza (flu): Flu activity & surveillance. Retrieved from https://www.cdc.gov/flu/weekly/index.htm

Centers for Disease Control and Prevention. (n.d.). Seasonal influenza (flu). Retrieved from https://www.cdc.gov/flu/index.htm

Chotpitayasunondh, T., Fischer, T. K., Heraud, J. M., et al. (2021). Influenza and COVID-19: What does co-existence mean? Influenza and Other Respiratory Viruses, 15, 407-412. https://doi.org/10.1111/irv.12824

Klein, S. L., & Flanagan, K. L. (2016). Sex differences in immune responses. Nature Reviews Immunology, 16(10), 626–638. https://doi.org/10.1038/nri.2016.90

Nelms, M., (n.d). *COVID vs. The Flu,* Internal Medicine & Pediatrics (IM&P), viewed 20th May 2023, < https://www.coolspringsinternalmedicine.com/blog/covid-vs-the-flu>

Simmelink-Mooij, M., Bunck, A. C. G., van Asten, L., & van der Lubben, I. M. (2018). Gender differences in infectious diseases in the Netherlands: an exploratory study. PLOS ONE, 13(9), e0203461. https://doi.org/10.1371/journal.pone.0203461

Smith, C. K., Vivekanandan-Giri, A., Tang, C., Knight, J. S., Mathew, A. R., & Padmanabhan, A. (2020). Influenza A infection attenuates relaxation responses of mouse bronchial smooth muscle to both cholinergic and β2-adrenergic agonists. American Journal of Physiology-Lung Cellular and Molecular Physiology, 318(1), L63–L74. https://doi.org/10.1152/ajplung.00345.2019

World Health Organization. (2021). Coronavirus disease (COVID-19) pandemic. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

World Health Organization. (n.d). Influenza (Seasonal). Fact Sheet. Retrieved from https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)

Appendix

Appendix A: Code and Datasets

File Name	Description	Source
National-notifiable-diseases-surveillance-system-nndss.csv	Excel had categories influenza lab confirmed infections into three separate sheets which needed to be merged: 2008-2015, 2016-2018, 2019-2021.	NNDSS public dataset- influenza (laboratory confirmed)
Infections_2020.csv	Consolidation of 26 individual csv files from the NNDSS fortnightly reports (2020) after the data had been cleaned and then merged	NNDSS fortnightly reports
Inf_covid_2021.csv	Consolidation of 26 individual csv files from the NNDSS fortnightly reports (2021) after the data had been cleaned and then merged	NNDSS fortnightly reports
Influenza_infection_2022.csv	Consolidation of 26 individual csv files from the NNDSS fortnightly reports (2022) after the data had been cleaned and then merged	NNDSS fortnightly reports
Inf_covid_09Jan_to_19Mar_2023.csv	Consolidation of 5 individual csv files from the NNDSS fortnightly reports (2023) after the data had been cleaned and then merged	NNDSS fortnightly reports
2008_influenza.csv	Death data from 2008-2018 on Influenza case in Australia	Australian Bureau of Statistics, Causes of Death, Australia
influenza_aus.csv	Death data from 2011-2021 on Influenza case in Australia	Australian Bureau of Statistics, Latest Release, Causes of Death, Australia

Appendix B: Data Cleaning

Cleaned Dataset from NNDSS fortnightly reports. (2022 example)

This is our re-cleaned code to include all variations of COVID spelling.

Example: COVID-19, COVID19, COVID-19*

In [1]:
```
df1 = read.csv("Jan3_to_Jan16_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df2 = read.csv("Jan17_to_Jan30_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df3 = read.csv("Jan31_to_Feb13_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df4 = read.csv("Feb14_to_Feb27_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df5 = read.csv("Feb28_to_Mar13_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df6 = read.csv("Mar14_to_Mar27_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df7 = read.csv("Mar28_to_Apr10_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df8 = read.csv("Apr11_to_Apr17_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df9 = read.csv("Apr18_to_May1_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df10 = read.csv("May2_to_May15_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df11 = read.csv("May16_to_May29_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df12 = read.csv("May30_to_Jun12_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df13 = read.csv("Jun13_to_Jun26_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df14 = read.csv("Jun27_to_Jul10_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df15 = read.csv("Jul11_to_Jul24_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df16 = read.csv("Jul25_to_Aug7_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df17 = read.csv("Aug8_to_Aug21_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df18 = read.csv("Aug22_to_Sep4_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df19 = read.csv("Sep5_to_Sep18_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df20 = read.csv("Sep19_to_Oct2_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df21 = read.csv("Oct3_to_Oct16_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df22 = read.csv("Oct17_to_Oct30_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df23 = read.csv("Oct31_to_Nov13_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df24 = read.csv("Nov14_to_Nov27_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df25 = read.csv("Nov28_to_Dec11_2022.csv", header=TRUE, stringsAsFactors=FALSE)
df26 = read.csv("Dec12_to_Jan1_2022_2023.csv", header=TRUE, stringsAsFactors=FALSE)
```

In [2]:
```
df1 <- df1[, -c(1,3:11, 13:23)]
df2 <- df2[, -c(1,3:11, 13:23)]
df3 <- df3[, -c(1,3:11, 13:23)]
df4 <- df4[, -c(1,3:11, 13:23)]
df5 <- df5[, -c(1,3:11, 13:23)]
df6 <- df6[, -c(1,3:11, 13:23)]
df7 <- df7[, -c(1,3:11, 13:23)]
df8 <- df8[, -c(1,3:11, 13:23)]
df9 <- df9[, -c(1,3:11, 13:23)]
df10 <- df10[, -c(1,3:11, 13:23)]
df11 <- df11[, -c(1,3:11, 13:23)]
df12 <- df12[, -c(1,3:11, 13:23)]
df13 <- df13[, -c(1,3:11, 13:23)]
df14 <- df14[, -c(1,3:11, 13:23)]
df15 <- df15[, -c(1,3:11, 13:23)]
df16 <- df16[, -c(1,3:11, 13:23)]
df17 <- df17[, -c(1,3:11, 13:23)]
df18 <- df18[, -c(1,3:11, 13:23)]
df19 <- df19[, -c(1,3:11, 13:23)]
df20 <- df20[, -c(1,3:11, 13:23)]
df21 <- df21[, -c(1,3:11, 13:23)]
df22 <- df22[, -c(1,3:11, 13:23)]
df23 <- df23[, -c(1,3:11, 13:23)]
df24 <- df24[, -c(1,3:11, 13:23)]
df25 <- df25[, -c(1,3:11, 13:23)]
df26 <- df26[, -c(1,3:11, 13:23)]
```

In [3]:
```
inf1 <- df1[grep("Influenza", df1$X),]
inf2 <- df2[grep("Influenza", df2$X),]
inf3 <- df3[grep("Influenza", df3$X),]
inf4 <- df4[grep("Influenza", df4$X),]
inf5 <- df5[grep("Influenza", df5$X),]
inf6 <- df6[grep("Influenza", df6$X),]
inf7 <- df7[grep("Influenza", df7$X),]
inf8 <- df8[grep("Influenza", df8$X),]
inf9 <- df9[grep("Influenza", df9$X),]
inf10 <- df10[grep("Influenza", df10$X),]
inf11 <- df11[grep("Influenza", df11$X),]
inf12 <- df12[grep("Influenza", df12$X),]
inf13 <- df13[grep("Influenza", df13$X),]
inf14 <- df14[grep("Influenza", df14$X),]
inf15 <- df15[grep("Influenza", df15$X),]
inf16 <- df16[grep("Influenza", df16$X),]
inf17 <- df17[grep("Influenza", df17$X),]
inf18 <- df18[grep("Influenza", df18$X),]
inf19 <- df19[grep("Influenza", df19$X),]
inf20 <- df20[grep("Influenza", df20$X),]
inf21 <- df21[grep("Influenza", df21$X),]
inf22 <- df22[grep("Influenza", df22$X),]
inf23 <- df23[grep("Influenza", df23$X),]
inf24 <- df24[grep("Influenza", df24$X),]
inf25 <- df25[grep("Influenza", df25$X),]
inf26 <- df26[grep("Influenza", df26$X),]
```

In [4]:

```
covid1 <- df1[grep("COVID", df1$X),]
covid2 <- df2[grep("COVID", df2$X),]
covid3 <- df3[grep("COVID", df3$X),]
covid4 <- df4[grep("COVID", df4$X),]
covid5 <- df5[grep("COVID", df5$X),]
covid6 <- df6[grep("COVID", df6$X),]
covid7 <- df7[grep("COVID", df7$X),]
covid8 <- df8[grep("COVID", df8$X),]
covid9 <- df9[grep("COVID", df9$X),]
covid10 <- df10[grep("COVID", df10$X),]
covid11 <- df11[grep("COVID", df11$X),]
covid12 <- df12[grep("COVID", df12$X),]
covid13 <- df13[grep("COVID", df13$X),]
covid14 <- df14[grep("COVID", df14$X),]
covid15 <- df15[grep("COVID", df15$X),]
covid16 <- df16[grep("COVID", df16$X),]
covid17 <- df17[grep("COVID", df17$X),]
covid18 <- df18[grep("COVID", df18$X),]
covid19 <- df19[grep("COVID", df19$X),]
covid20 <- df20[grep("COVID", df20$X),]
covid21 <- df21[grep("COVID", df21$X),]
covid22 <- df22[grep("COVID", df22$X),]
covid23 <- df23[grep("COVID", df23$X),]
covid24 <- df24[grep("COVID", df24$X),]
covid25 <- df25[grep("COVID", df25$X),]
covid26 <- df26[grep("COVID", df26$X),]
```

In [5]:

```
df1x <- rbind(inf1, covid1)
df2x <- rbind(inf2, covid2)
df3x <- rbind(inf3, covid3)
df4x <- rbind(inf4, covid4)
df5x <- rbind(inf5, covid5)
df6x <- rbind(inf6, covid6)
df7x <- rbind(inf7, covid7)
df8x <- rbind(inf8, covid8)
df9x <- rbind(inf9, covid9)
df10x <- rbind(inf10, covid10)
df11x <- rbind(inf11, covid11)
df12x <- rbind(inf12, covid12)
df13x <- rbind(inf13, covid13)
df14x <- rbind(inf14, covid14)
df15x <- rbind(inf15, covid15)
df16x <- rbind(inf16, covid16)
df17x <- rbind(inf17, covid17)
df18x <- rbind(inf18, covid18)
df19x <- rbind(inf19, covid19)
df20x <- rbind(inf20, covid20)
df21x <- rbind(inf21, covid21)
df22x <- rbind(inf22, covid22)
df23x <- rbind(inf23, covid23)
df24x <- rbind(inf24, covid24)
df25x <- rbind(inf25, covid25)
df26x <- rbind(inf26, covid26)
```

In [6]:

```
df1x[,2] <- as.numeric(df1x[,2])
df2x[,2] <- as.numeric(df2x[,2])
df3x[,2] <- as.numeric(df3x[,2])
df4x[,2] <- as.numeric(df4x[,2])
df5x[,2] <- as.numeric(df5x[,2])
df6x[,2] <- as.numeric(df6x[,2])
df7x[,2] <- as.numeric(df7x[,2])
df8x[,2] <- as.numeric(df8x[,2])
df9x[,2] <- as.numeric(df9x[,2])
df10x[,2] <- as.numeric(df10x[,2])
df11x[,2] <- as.numeric(df11x[,2])
df12x[,2] <- as.numeric(df12x[,2])
df13x[,2] <- as.numeric(df13x[,2])
df14x[,2] <- as.numeric(df14x[,2])
df15x[,2] <- as.numeric(df15x[,2])
df16x[,2] <- as.numeric(df16x[,2])
df17x[,2] <- as.numeric(df17x[,2])
df18x[,2] <- as.numeric(df18x[,2])
df19x[,2] <- as.numeric(df19x[,2])
df20x[,2] <- as.numeric(df20x[,2])
df21x[,2] <- as.numeric(df21x[,2])
df22x[,2] <- as.numeric(df22x[,2])
df23x[,2] <- as.numeric(df23x[,2])
df24x[,2] <- as.numeric(df24x[,2])
df25x[,2] <- as.numeric(df25x[,2])
df26x[,2] <- as.numeric(df26x[,2])
```

In [7]:
```
combined_table_2022 <- cbind(df1x, df2x[,2], df3x[,2], df4x[,2], df5x[,2], df6x[,2], df7x[,2], df8x[,2], df9x[,2], df10x[,2
                             , df11x[,2], df12x[,2], df13x[,2], df14x[,2], df15x[,2], df16x[,2], df17x[,2], df18x[,2], df19
                             , df20x[,2], df21x[,2], df22x[,2], df23x[,2], df24x[,2], df25x[,2], df26x[,2])
```

In [8]:
```
colnames(combined_table_2022) <- c("Disease", "Jan3_to_Jan16_2022",
"Jan17_to_Jan30_2022",
"Jan31_to_Feb13_2022",
"Feb14_to_Feb27_2022",
"Feb28_to_Mar13_2022",
"Mar14_to_Mar27_2022",
"Mar28_to_Apr10_2022",
"Apr11_to_Apr17_2022",
"Apr18_to_May1_2022",
"May2_to_May15_2022",
"May16_to_May29_2022",
"May30_to_Jun12_2022",
"Jun13_to_Jun26_2022",
"Jun27_to_Jul10_2022",
"Jul11_to_Jul24_2022",
"Jul25_to_Aug7_2022",
"Aug8_to_Aug21_2022",
"Aug22_to_Sep4_2022",
"Sep5_to_Sep18_2022",
"Sep19_to_Oct2_2022",
"Oct3_to_Oct16_2022",
"Oct17_to_Oct30_2022",
"Oct31_to_Nov13_2022",
"Nov14_to_Nov27_2022",
"Nov28_to_Dec11_2022",
"Dec12_to_Jan1_2022_2023")
combined_table_2022
```

A data.frame: 2 × 27

	Disease	Jan3_to_Jan16_2022	Jan17_to_Jan30_2022	Jan31_to_Feb13_2022	Feb14_to_Feb27_2022	Feb28_to_Mar13_2022	Mar14_
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
40	Influenza (laboratory confirmed)	28	10	9	19	49	
24	COVID-19	814323	485404	270766	205213	253471	

Cleaned Dataset from Australian Bureau of Statistics, Causes of Death, Australia

[1]:
```
df = read.csv("influenza_aus.csv", header=TRUE, stringsAsFactors=TRUE)
df_extra = read.csv("2008_influenza.csv", header=TRUE, stringsAsFactors=TRUE)
```

[ ]:
```

```

```
library(dplyr)
## we have multiple columns where there are NA values and in the raw data this was used to separate each year of data
```

[2]:
```
df <- df[,!apply(is.na(df), 2, any)]
head(df)
```

A data.frame: 6 × 34

	X	X2011	X.1	X.2	X2012	X.4	X.5	X2013	X.7	X.8	⋯	X.23	X2019	X.25	X.26	X2020
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>	<fct>
1	Cause of death and ICD-10 code	Males	Females	Persons	Males	Females	Persons	Males	Females	Persons	⋯	Persons	Males	Females	Persons	Males
2	Total deaths	75,377	71,413	146,790	75,677	73,368	149,045	76,555	72,362	148,917	⋯	159,463	85,709	79,081	164,790	85,200
3											⋯					
4	Causes of death										⋯					
5	CHAPTER I Certain infectious and parasitic diseases	1,236	1162	2,398	1240	1197	2,437	1,344	1,344	2,688	⋯	2,422	1,180	1,357	2,537	1,188

[3]:
```
colnames(df) <- c("Cause_Death", "2011_M", "2011_F", "2011_P","2012_M", "2012_F", "2012_P",
                  "2013_M", "2013_F", "2013_P", "2014_M", "2014_F", "2014_P",
                  "2015_M", "2015_F", "2015_P", "2016_M", "2016_F", "2016_P",
                  "2017_M", "2017_F", "2017_P", "2018_M", "2018_F", "2018_P",
                  "2019_M", "2019_F", "2019_P", "2020_M", "2020_F", "2020_P",
                  "2021_M", "2021_F", "2021_P")
```

[4]:
```
head(df)
```

A data.frame: 6 × 34

	Cause_Death	2011_M	2011_F	2011_P	2012_M	2012_F	2012_P	2013_M	2013_F	2013_P	⋯	2018_P	2019_M	2019_F
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>
1	Cause of death and ICD-10 code	Males	Females	Persons	Males	Females	Persons	Males	Females	Persons	⋯	Persons	Males	Females
2	Total deaths	75,377	71,413	146,790	75,677	73,368	149,045	76,555	72,362	148,917	⋯	159,463	85,709	79,081
3											⋯			
4	Causes of death										⋯			
5	CHAPTER I Certain infectious and parasitic diseases (A00-B99)	1,236	1162	2,398	1240	1197	2,437	1,344	1,344	2,688	⋯	2,422	1,180	1,357

[5]:
```
# Now we need to find the columns that match the phrase 'Influenza'
library(data.table)
inf <- df[grep("Influenza", df$Cause_Death),]

head(inf)
```

A data.frame: 4 × 34

	Cause_Death	2011_M	2011_F	2011_P	2012_M	2012_F	2012_P	2013_M	2013_F	2013_P	⋯	2018_P	2019_M	2019_F
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>
789	Influenza and pneumonia (J09-J18)	1121	1,353	2,474	1164	1,606	2,770	1,139	1,366	2,505	⋯	3,124	1,759	2,287
790	Influenza due to certain identified influenza virus (J09)	12	9	21	0	0	0	3	6	9	⋯	8	2	1
791	Influenza due to other identified influenza virus (J10)	6	7	13	34	48	82	21	19	40	⋯	111	352	463
792	Influenza, virus not identified (J11)	12	23	35	31	38	69	15	21	36	⋯	29	104	147

[6]:
```
# now we want to remove row 1 as this row was a sum of all sub catagories of intluenza and pneumonia
df1 <- inf[-1,]
df1
```

A data.frame: 3 × 34

	Cause_Death	2011_M	2011_F	2011_P	2012_M	2012_F	2012_P	2013_M	2013_F	2013_P	⋯	2018_P	2019_M	2019_F	2019_P
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>
790	Influenza due to certain identified influenza virus (J09)	12	9	21	0	0	0	3	6	9	⋯	8	2	1	3
791	Influenza due to other identified influenza virus (J10)	6	7	13	34	48	82	21	19	40	⋯	111	352	463	815
792	Influenza, virus not identified (J11)	12	23	35	31	38	69	15	21	36	⋯	29	104	147	251

[7]:
```
#Now we will clean the other data set which includes from 2008 and make sure our data 2011-2018 is the same before merging
head(df_extra)
```

A data.frame: 6 × 44

	X	X2008	X.1	X.2	X.3	X2009	X.4	X.5	X.6	X2010	⋯	X.25	X.26	X.27	X2017	X.28
	<fct>	<fct>	<fct>	<fct>	<lgl>	<fct>	<fct>	<fct>	<lgl>	<fct>	⋯	<fct>	<fct>	<lgl>	<fct>	<fct>
1	Cause of death and ICD-10 code	Males	Females	Persons	NA	Males	Females	Persons	NA	Males	⋯	Females	Persons	NA	Males	Females
2	Total deaths	72,810	69,699	142,509	NA	72,665	68,639	141,304	NA	73,040	⋯	76,653	158,478	NA	83,997	79,019

[8]:
```
#remove the columns with NA
df_extra<- df_extra[,!apply(is.na(df_extra), 2, any)]
head(df_extra)
```

A data.frame: 6 × 34

	X	X2008	X.1	X.2	X2009	X.4	X.5	X2010	X.7	X.8	⋯	X.23	X2016	X.25	X.26	X2017
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>	<fct>
1	Cause of death and ICD-10 code	Males	Females	Persons	Males	Females	Persons	Males	Females	Persons	⋯	Persons	Males	Females	Persons	Males
2	Total deaths	72,810	69,699	142,509	72,665	68,639	141,304	73,040	69,737	142,777	⋯	157,271	81,825	76,653	158,478	83,997
3											⋯					
4											⋯					
5	Causes of death										⋯					

[9]:
```
#fix the column names so we know the male and females for each year
colnames(df_extra) <- c("Cause_Death", "2008_M", "2008_F", "2008_P","2009_M", "2009_F", "2009_P",
                        "2010_M", "2010_F", "2010_P", "2011_M", "2011_F", "2011_P",
                        "2012_M", "2012_F", "2012_P", "2013_M", "2013_F", "2013_P",
                        "2014_M", "2014_F", "2014P", "2015_M", "2015_F", "2015_P",
                        "2016_M", "2016_F", "2016_P", "2017_M", "2017_F", "2017_P",
                        "2018_M", "2018_F", "2018_P")
```

[10]:
```
head(df_extra)
```

A data.frame: 6 × 34

	Cause_Death	2008_M	2008_F	2008_P	2009_M	2009_F	2009_P	2010_M	2010_F	2010_P	⋯	2015_P	2016_M	2016_F	2016_P
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>
1	Cause of death and ICD-10 code	Males	Females	Persons	Males	Females	Persons	Males	Females	Persons	⋯	Persons	Males	Females	Persons
2	Total deaths	72,810	69,699	142,509	72,665	68,639	141,304	73,040	69,737	142,777	⋯	157,271	81,825	76,653	158,478

[11]:
```
#find the rows that contain the words influenza

inf_extra <- df_extra[grep("Influenza", df_extra$Cause_Death),]
head(inf_extra)
```

A data.frame: 4 × 34

	Cause_Death	2008_M	2008_F	2008_P	2009_M	2009_F	2009_P	2010_M	2010_F	2010_P	⋯	2015_P	2016_M	2016_F	2016_P
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>
790	Influenza and pneumonia (J09-J18)	728	1,056	1,784	766	1,035	1,801	1,048	1,317	2,365	⋯	3,075	1,441	1,890	3,331
791	Influenza due to certain identified influenza virus (J09)	0	0	0	41	37	78	14	7	21	⋯	7	7	4	11
792	Influenza due to other identified influenza virus (J10)	8	10	18	7	8	15	2	1	3	⋯	198	138	186	324
793	Influenza, virus not identified (J11)	7	21	28	11	25	36	7	14	21	⋯	94	51	85	136

[12]:
```
# now we want to remove row 1 as this row was a sum of all sub catagories of intluenza and pneumonia
df2 <- inf_extra[-1,]
df2
```

A data.frame: 3 × 34

	Cause_Death	2008_M	2008_F	2008_P	2009_M	2009_F	2009_P	2010_M	2010_F	2010_P	⋯	2015_P	2016_M	2016_F	2016_P
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>
791	Influenza due to certain identified influenza virus (J09)	0	0	0	41	37	78	14	7	21	⋯	7	7	4	11
792	Influenza due to other identified influenza virus (J10)	8	10	18	7	8	15	2	1	3	⋯	198	138	186	324
793	Influenza, virus not identified (J11)	7	21	28	11	25	36	7	14	21	⋯	94	51	85	136

[14]:
```
#from df2 we want to extract 2008-2010
df2 <- df2[, 1:10]
df2
```

A data.frame: 3 × 10

	Cause_Death	2008_M	2008_F	2008_P	2009_M	2009_F	2009_P	2010_M	2010_F	2010_P
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>
791	Influenza due to certain identified influenza virus (J09)	0	0	0	41	37	78	14	7	21
792	Influenza due to other identified influenza virus (J10)	8	10	18	7	8	15	2	1	3
793	Influenza, virus not identified (J11)	7	21	28	11	25	36	7	14	21

[15]:
```
merged_df <- merge(df2,df1)
merged_df
```

A data.frame: 3 × 43

Cause_Death	2008_M	2008_F	2008_P	2009_M	2009_F	2009_P	2010_M	2010_F	2010_P	⋯	2018_P	2019_M	2019_F	2019_P	2020_M	2020_F	2020_P	2021_M
<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	⋯	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>
Influenza due to certain identified influenza virus (J09)	0	0	0	41	37	78	14	7	21	⋯	8	2	1	3	0	1	1	0
Influenza due to other identified influenza	8	10	18	7	8	15	2	1	3	⋯	111	352	463	815	18	22	40	0

[16]:
```
merged_df <- data.frame(lapply(merged_df, function(x) if(is.factor(x)) as.character(x) else x))
merged_df
merged_df1  <- data.frame(lapply(merged_df[,2:43], function(x) if(is.character(x)) as.numeric(x) else x))
merged_df1
```

A data.frame: 3 × 43

Cause_Death	X2008_M	X2008_F	X2008_P	X2009_M	X2009_F	X2009_P	X2010_M	X2010_F	X2010_P	⋯	X2018_P	X2019_M	X2019_F	X2019_
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	⋯	<chr>	<chr>	<chr>	<chr
Influenza due to certain identified influenza virus (J09)	0	0	0	41	37	78	14	7	21	⋯	8	2	1	
Influenza due to other identified influenza virus (J10)	8	10	18	7	8	15	2	1	3	⋯	111	352	463	81
Influenza, virus not identified (J11)	7	21	28	11	25	36	7	14	21	⋯	29	104	147	25

A data.frame: 3 × 42

X2008_M	X2008_F	X2008_P	X2009_M	X2009_F	X2009_P	X2010_M	X2010_F	X2010_P	X2011_M	⋯	X2018_P	X2019_M	X2019_F	X2019_P
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	⋯	<dbl>	<dbl>	<dbl>	<dbl>
0	0	0	41	37	78	14	7	21	12	⋯	8	2	1	3
8	10	18	7	8	15	2	1	3	6	⋯	111	352	463	815

[17]:
```
# now lets find the sum of each column as we have found all influeza death cases
# first columns need to be changed to numeric not factor
sums_inf <- data.frame(mapply(sum, merged_df1))
head(sums_inf)
sums_inf <- t(sums_inf)
head(sums_inf)

#fix the column names so we know the male and females for each year
colnames(sums_inf) <- c("2008_M", "2008_F", "2008_P","2009_M", "2009_F", "2009_P",
                        "2010_M", "2010_F", "2010_P", "2011_M", "2011_F", "2011_P",
                        "2012_M", "2012_F", "2012_P", "2013_M", "2013_F", "2013_P",
                        "2014_M", "2014_F", "2014P", "2015_M", "2015_F", "2015_P",
                        "2016_M", "2016_F", "2016_P", "2017_M", "2017_F", "2017_P",
                        "2018_M", "2018_F", "2018_P", "2019_M", "2019_F", "2019_P",
                          "2020_M", "2020_F", "2020_P", "2021_M", "2021_F", "2021_P")
sums_inf
```

A data.frame: 6 × 1

	mapply.sum..merged_df1.
	<dbl>
X2008_M	15
X2008_F	31
X2008_P	46
X2009_M	59
X2009_F	70
X2009_P	129

A matrix: 1 × 42 of type dbl

	X2008_M	X2008_F	X2008_P	X2009_M	X2009_F	X2009_P	X2010_M	X2010_F	X2010_P	X2011_M	⋯	X2018_P
mapply.sum..merged_df1.	15	31	46	59	70	129	23	22	45	30	⋯	148

Appendix C: Modelling

[1]:
```
female_total = read.csv("inf-infection-female-age-change-alldata.csv", header=TRUE, stringsAsFactors=FALSE)
male_total = read.csv("inf-infection-male-age-change-alldata.csv", header=TRUE, stringsAsFactors=FALSE)
```

[2]:
```
female2008 <- female_total[1:18, 1:2]
female2009 <- female_total[19:36, 1:2]
female2010 <- female_total[37:54, 1:2]
female2011 <- female_total[55:72, 1:2]
female2012 <- female_total[73:90, 1:2]
female2013 <- female_total[91:108, 1:2]
female2014 <- female_total[109:126, 1:2]
female2015 <- female_total[127:144, 1:2]
female2016 <- female_total[145:162, 1:2]
female2017 <- female_total[163:180, 1:2]
female2018 <- female_total[181:198, 1:2]
female2019 <- female_total[199:216, 1:2]
female2020 <- female_total[217:234, 1:2]
```

[3]:
```
male2008 <- male_total[1:18, 1:2]
male2009 <- male_total[19:36, 1:2]
male2010 <- male_total[37:54, 1:2]
male2011 <- male_total[55:72, 1:2]
male2012 <- male_total[73:90, 1:2]
male2013 <- male_total[91:108, 1:2]
male2014 <- male_total[109:126, 1:2]
male2015 <- male_total[127:144, 1:2]
male2016 <- male_total[145:162, 1:2]
male2017 <- male_total[163:180, 1:2]
male2018 <- male_total[181:198, 1:2]
male2019 <- male_total[199:216, 1:2]
male2020 <- male_total[217:234, 1:2]
```

•[4]:
```
#-------------------------------------------------------------figure 11, 12, 13, 14-------------------------------------------------------------#
model <- lm(count ~ age, data = female2009)
summary(model)
plot(count ~ age, data = female2009)
abline(model)
title(main = "Linear Regression: Influenza cases Female 2009")
```

•••

•[5]:
```
#-------------------------------------------------------------figure 11, 12, 13, 14-------------------------------------------------------------#
model <- lm(count ~ age, data = male2009)
summary(model)
plot(count ~ age, data = male2009)
abline(model)
title(main = "Linear Regression: Influenza cases Male 2009")
```

•[28]:
```
#-------------------------------------------------------------figure 15-------------------------------------------------------------#
male_alldata <- read.csv("inf-infection-male-age-change-alldata.csv", header=TRUE, stringsAsFactors=FALSE)
female_alldata <- read.csv("inf-infection-female-age-change-alldata.csv", header=TRUE, stringsAsFactors=FALSE)
```

[34]:
```
library(scatterplot3d)
library(plotly)
```

[35]:
```
plot_ly(data = female_alldata, x = ~age, y = ~year, z = ~count, type = "scatter3d", mode = "markers",
        marker = list(size = 5, color = ~count)) %>%
  layout(scene = list(xaxis = list(title = "Age"),
                      yaxis = list(title = "Year"),
                      zaxis = list(title = "Count")),
         width = 800,
         height = 600)
```

Appendix D: Visualisations

[1]:
```
inf_covid_merge = read.csv("merge-infection-every quarter.csv", header=TRUE, stringsAsFactors=FALSE)
```

[2]:
```
rownames(inf_covid_merge) <- c("Q1 2020","Q2 2020","Q3 2020","Q4 2020",
                               "Q1 2021","Q2 2021","Q3 2021","Q4 2021",
                               "Q1 2022","Q2 2022","Q3 2022","Q4 2022","Q1 2023")
inf_covid_merge <- inf_covid_merge[,2:3]
colnames(inf_covid_merge) <- c("V1","V2")
```

*[8]:
```
library(ggplot2)
inf_covid_merge$date <- rownames(inf_covid_merge)
inf_covid_merge$date <- factor(inf_covid_merge$date, levels = c("Q1 2020","Q2 2020","Q3 2020","Q4 2020",
                               "Q1 2021","Q2 2021","Q3 2021","Q4 2021",
                               "Q1 2022","Q2 2022","Q3 2022","Q4 2022","Q1 2023"))

options(repr.plot.width = 12, repr.plot.height = 6)
#-----------------------------------------------------------figure 2-----------------------------------------------------------#
ggplot(data = as.data.frame(inf_covid_merge), aes(x = date, y = as.numeric(V1), group = 1)) +
  geom_line(aes(colour = "Influenza")) +
  geom_line(aes(y = as.numeric(V2), colour = "COVID-19")) +
  scale_color_manual(values = c("Influenza" = "blue", "COVID-19" = "red")) +
  ylab("Number of cases") +
  xlab("Date") +
  ggtitle("Influenza vs COVID-19 infection cases 01/2020-04/2023") +
  theme_bw() +
  theme(axis.text.x = element_text(size = 14, angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::comma_format(), breaks = seq(0, 2800000, by = 500000), minor_breaks = seq(0, 2800000, by = 50000), limits = c(0, 2800000))
#-------------------------------------------------------------figure 3-------------------------------------------------------------#
inf_covid_merge_part <- inf_covid_merge[1:6,]
ggplot(data = as.data.frame(inf_covid_merge_part), aes(x = date, y = as.numeric(V1), group = 1)) +
  geom_line(aes(colour = "Influenza")) +
  geom_line(aes(y = as.numeric(V2), colour = "COVID-19")) +
  scale_color_manual(values = c("Influenza" = "blue", "COVID-19" = "red")) +
  ylab("Number of cases") +
  xlab("Date") +
  ggtitle("Influenza vs COVID-19 infection cases 01/2020-06/2021") +
  theme_bw() +
  theme(axis.text.x = element_text(size = 14, angle = 45, hjust = 1))
```

[19]:
```
inf_covid_trend = read.csv("Influenza_infection_2008_2022.csv", header=TRUE, stringsAsFactors=FALSE)
rownames(inf_covid_trend) <- c("2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018","2019","2020","2021","2022")
inf_covid_trend <- inf_covid_trend[,2:3]
colnames(inf_covid_trend) <- c("V1","V2")
inf_covid_trend2008_2021 <- inf_covid_trend[1:14,]
```

*[18]:
```
#-------------------------------------------------------------figure 4-------------------------------------------------------------#
library(ggplot2)
options(repr.plot.width = 12, repr.plot.height = 8)

ggplot(data = as.data.frame(inf_covid_trend2008_2021), aes(x = rownames(inf_covid_trend2008_2021), y = as.numeric(V1), group = 1)) +
  geom_line(aes(colour = "Influenza")) +
  scale_color_manual(values = c("Influenza" = "red")) +
  scale_y_continuous(labels = scales::comma_format(), limits = c(0, 350000)) +
  ylab("Number of infection cases") +
  xlab("Date") +
  ggtitle("Influenza infection cases 2008-2021") +
  theme_bw() +
  theme(axis.text.x = element_text(size = 12, angle = 45, hjust = 1))
```

[3]:
```
inf_covid_2020 = read.csv("2020infection.csv", header=TRUE, stringsAsFactors=FALSE)
```

[4]:
```
# transform the dataframe to make the date to be x
inf_covid_2020_t <- t(inf_covid_2020)
# getrid the desease to avoid the error when convert the string to numeric
inf_covid_2020_t <- inf_covid_2020_t[-1,]
# rename
rownames(inf_covid_2020_t) <- c("1Jan - 17Jan 2020","18Jan - 31Jan 2020","1Feb - 14Feb 2020","15Feb - 28Feb 2020","29Feb - 13Mar 2020","14Mar - 27Mar 2020",
"28Mar - 10Apr 2020","11Apr - 24Apr 2020","25Apr - 8May 2020","9May - 22May 2020","23May - 5Jun 2020","6Jun - 19Jun 2020","20Jun - 3Jul 2020",
"4Jul - 17Jul 2020","18Jul - 31Jun 2020","1Aug - 14Aug 2020","15Aug - 28Aug 2020","29Aug - 11Sep 2020","12Sep - 25Sep 2020","26Sep - 9Oct 2020",
"10Oct - 25Oct 2020","26Oct - 8Nov 2020","9Nov - 22Nov 2020","23Nov - 6Dec 2020","7Dec - 20Dec 2020","21Dec 2020 - 3Jan 2021")
```

*[5]:
```
library(ggplot2)
inf_covid_2020_t <- as.data.frame(inf_covid_2020_t)
inf_covid_2020_t$date <- rownames(inf_covid_2020_t)
inf_covid_2020_t$date <- factor(inf_covid_2020_t$date, levels = c("1Jan - 17Jan 2020","18Jan - 31Jan 2020","1Feb - 14Feb 2020","15Feb - 28Feb 2020",
"29Feb - 13Mar 2020","14Mar - 27Mar 2020","28Mar - 10Apr 2020","11Apr - 24Apr 2020","25Apr - 8May 2020","9May - 22May 2020","23May - 5Jun 2020",
"6Jun - 19Jun 2020","20Jun - 3Jul 2020","4Jul - 17Jul 2020","18Jul - 31Jun 2020","1Aug - 14Aug 2020","15Aug - 28Aug 2020","29Aug - 11Sep 2020",
"12Sep - 25Sep 2020","26Sep - 9Oct 2020","10Oct - 25Oct 2020","26Oct - 8Nov 2020","9Nov - 22Nov 2020","23Nov - 6Dec 2020","7Dec - 20Dec 2020",
"21Dec 2020 - 3Jan 2021"))
options(repr.plot.width = 18, repr.plot.height = 8)
```

```
#-------------------------------------------------------figure 5-------------------------------------------------------#
ggplot(data = as.data.frame(inf_covid_2020_t), aes(x = date, y = as.numeric(V1), group = 1)) +
  geom_line(aes(colour = "Influenza")) +
  geom_line(aes(y = as.numeric(V2), colour = "COVID-19")) +
  scale_color_manual(values = c("Influenza" = "blue", "COVID-19" = "red")) +
  ylab("Number of cases") +
  xlab("Date") +
  ggtitle("Influenza vs COVID-19 infection cases in 2020") +
  theme(axis.text.x = element_text(size = 12, angle = 45, hjust = 1)) +
  theme_bw() +
  scale_y_continuous(breaks = seq(0, 6000, by = 500), minor_breaks = seq(0, 6000, by = 100), limits = c(0, 6000))
```

[9]:
```
inf_covid_2021 = read.csv("inf_covid_2021.csv", header=TRUE, stringsAsFactors=FALSE)
```

[ ]:
```
# transform the dataframe to make the date to be x
inf_covid_2021_t <- t(inf_covid_2021)
# getrid the desease to avoid the error when convert the string to numeric
inf_covid_2021_t <- inf_covid_2021_t[-1,]
# rename
rownames(inf_covid_2021_t) <- c("4Jan - 17Jan 2021","18Jan - 31Jan 2021","1Feb - 14Feb 2021","15Feb - 28Feb 2021","1Mar - 14Mar 2021",
"15Mar - 28Mar 2021","29Mar - 11Apr 2021","12Apr - 25Apr 2021","26Apr - 9May 2021", "10May - 23May 2021","24May - 6Jun 2021","7Jun - 20Jun 2021",
"21Jun - 4Jul 2021","5Jul - 18Jul 2021","19Jul - 1Aug 2021","2Aug - 15Aug 2021","16Aug - 29Aug 2021","30Aug - 12Sep 2021", "13Sep - 26Sep 2021",
"27Sep - 10Oct 2021","11Oct - 24Oct 2021","25Oct - 7Nov 2021","8Nov - 21Nov 2021","22Nov - 5Dec 2021","6Dec - 19Dec 2021","20Dec 2021 - 2Jan 2022")
inf_covid_2021_t <- as.data.frame(inf_covid_2021_t)
```

*[8]:
```
library(ggplot2)
inf_covid_2021_t$date <- rownames(inf_covid_2021_t)
inf_covid_2021_t$date <- factor(inf_covid_2021_t$date, levels = c("4Jan - 17Jan 2021","18Jan - 31Jan 2021","1Feb - 14Feb 2021","15Feb - 28Feb 2021",
"1Mar - 14Mar 2021", "15Mar - 28Mar 2021","29Mar - 11Apr 2021","12Apr - 25Apr 2021","26Apr - 9May 2021","10May - 23May 2021","24May - 6Jun 2021",
"7Jun - 20Jun 2021","21Jun - 4Jul 2021","5Jul - 18Jul 2021", "19Jul - 1Aug 2021","2Aug - 15Aug 2021","16Aug - 29Aug 2021","30Aug - 12Sep 2021",
"13Sep - 26Sep 2021","27Sep - 10Oct 2021","11Oct - 24Oct 2021","25Oct - 7Nov 2021","8Nov - 21Nov 2021","22Nov - 5Dec 2021","6Dec - 19Dec 2021",
"20Dec 2021 - 2Jan 2022"))
#-------------------------------------------------------figure 6-------------------------------------------------------#
options(repr.plot.width = 12, repr.plot.height = 8)
ggplot(data = as.data.frame(inf_covid_2021_t), aes(x = date, group = 1)) +
  geom_line(aes(y = as.numeric(V1), color = "Influenza"), size = 1.2) +
  geom_line(aes(y = as.numeric(V2), color = "COVID-19"), size = 1.2) +
  scale_color_manual(values = c("Influenza" = "blue", "COVID-19" = "red")) +
  ylab("Number of cases") +
  xlab("Date") +
  ggtitle("Influenza vs COVID-19 infection cases in 2021") +
  theme_bw() +
  theme(axis.text.x = element_text(size = 12, angle = 45, hjust = 1))

#-------------------------------------------------------figure 7-------------------------------------------------------#
inf_covid_2021_t_part <- inf_covid_2021_t[1:14,]
ggplot(data = as.data.frame(inf_covid_2021_t_part), aes(x = date, group = 1)) +
  geom_line(aes(y = as.numeric(V1), color = "Influenza"), size = 1.2) +
  geom_line(aes(y = as.numeric(V2), color = "COVID-19"), size = 1.2) +
  scale_color_manual(values = c("Influenza" = "#6495ED", "COVID-19" = "red")) +
  ylab("Number of cases") +
  xlab("Date") +
  ggtitle("Influenza vs COVID-19 infection cases from 1/2021 to 07/2021") +
  theme_bw() +
  theme(axis.text.x = element_text(size = 12, angle = 45, hjust = 1))
```

[1]:
```
inf_infec_male = read.csv("inf-infection-male-age.csv", header=TRUE, stringsAsFactors=FALSE)
inf_infec_female = read.csv("inf-infection-female-age.csv", header=TRUE, stringsAsFactors=FALSE)
male2019 = read.csv("inf-infection-male-age-change-2019.csv", header=TRUE, stringsAsFactors=FALSE)
inf_infec_female_merge = read.csv("inf-infection-female-age-merge.csv", header=TRUE, stringsAsFactors=FALSE)
```

[2]:
```
inf_infec_male <- inf_infec_male[,2:ncol(inf_infec_male)]
inf_infec_female <- inf_infec_female[,2:ncol(inf_infec_female)]
colnames(inf_infec_male) <- c("A","B","C","D","E","F","G","H","I",
                              "J","K","L","M","N","O","P","Q","R")
colnames(inf_infec_female) <- c("A","B","C","D","E","F","G","H","I",
                              "J","K","L","M","N","O","P","Q","R")
# colnames(inf_infec_female) <- c("0-4","5-9","10-14","15-19","20-24","25-29","30-34","35-39","40-44",
#                                 "45-49","50-54","55-59","60-64","65-69","70-74","75-79","80-84","85+")
rownames(inf_infec_male) <- c("2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018","2019","2020")
rownames(inf_infec_female) <- c("2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018","2019","2020")
```

•••

[5]:
```
##### merge age to make the graph more clear
inf_infec_female_merge <- inf_infec_female_merge[, 2:9]
colnames(inf_infec_female_merge) <- c("A","B","C","D","E","F","G","H")
rownames(inf_infec_female_merge) <- c("2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","2018","2019","2020")
```

[6]:
```r
#-------------------------------------------------------------figure 9-------------------------------------------------------------#
library(ggplot2)
inf_infec_female_merge$date <- rownames(inf_infec_female_merge)

options(repr.plot.width = 18, repr.plot.height = 8)

ggplot(data = as.data.frame(inf_infec_female_merge), aes(x = date, group = 1)) +
  geom_line(aes(y = A, color = "0-9"), linewidth = 1) +
  geom_line(aes(y = B, color = "10-19"), linewidth = 1) +
  geom_line(aes(y = C, color = "20-29"), linewidth = 1) +
  geom_line(aes(y = D, color = "30-39"), linewidth = 1) +
  geom_line(aes(y = E, color = "40-49"), linewidth = 1) +
  geom_line(aes(y = F, color = "50-59"), linewidth = 1) +
  geom_line(aes(y = G, color = "60-69"), linewidth = 1) +
  geom_line(aes(y = H, color = "70+"), linewidth = 1) +
  ylab("Influenza infection of female") +
  xlab("Date") +
  ggtitle("Influenza infection by age group in female in 2008-2020") +
  theme(axis.text.x = element_text(size = 20, angle = 45, hjust = 1)) +
  theme_bw() +
  scale_color_manual(values = c("0-9" = "blue", "10-19" = "green", "20-29" = "black", "30-39" = "orange",
                                "40-49" = "purple", "50-59" = "cyan", "60-69" = "brown", "70+" = "red"))
```

[7]:
```r
#### zoom in age 0-9
inf_infec_female_age_zero_nine <- inf_infec_female[,1:2]
inf_infec_female_age_zero_nine$Year <- rownames(inf_infec_female_age_zero_nine)
```

[8]:
```r
#-------------------------------------------------------------figure 10-------------------------------------------------------------#
options(repr.plot.width = 14, repr.plot.height = 8)
data_long <- tidyr::pivot_longer(inf_infec_female_age_zero_nine, cols = c(A, B), names_to = "Category", values_to = "Count")

labels <- c("0-4", "5-9")
colors <- c("A" = "blue", "B" = "pink")

ggplot(data_long, aes(x = Year, y = Count, fill = Category, width=0.8)) +
  geom_col(position = "dodge") +
  scale_fill_manual(values = colors, labels = labels) +
  labs(x = "Year", y = "Count", title = "Influenza infection cases between 0-4 and 5-9 among female") +
  theme_bw()
```