Programming Languages for Data Engineering (AC50002)

R programming Assignment

Yashash Thittamaranahalli Venkatesh (2546871)

Contents

Part2- Data exploration

Selection of Dataset

The selection of the COVID-19 dataset for European countries between 2020-2023 was deliberate and driven by my keen interest and personal experience during the pandemic. This dataset was chosen due to its relevance and significance in understanding the dynamics of COVID-19 across European nations. The decision was influenced by a desire to explore and analyse a dataset that aligns with real-world experiences and sheds light on the patterns and trends in the region. The dataset's source from the (European Centre for Disease Prevention and Control (2022)).

Graphs for my data

Bar Plot

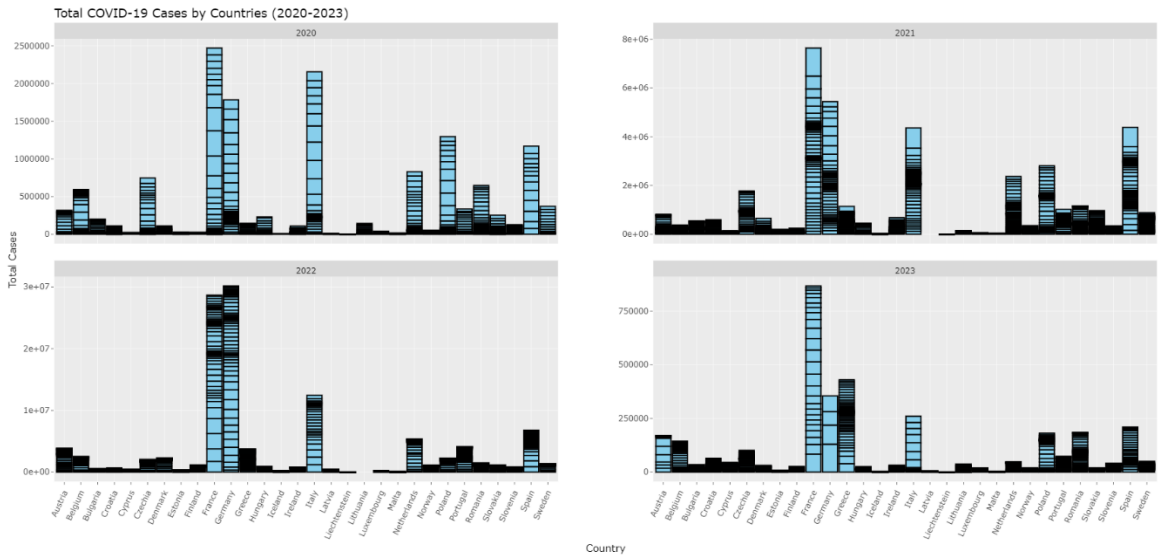

Figure 1Bar Graph showing total cases vs Country (2020-2023)

I have collected data for four distinct years, which shows the annual total of COVID-19 instances. These educational bar graphs, created with R Studio's ggplot tool, show how Covid-19 has affected different nations and how trends have changed over time.

It is clear from looking at the graphs which countries were more affected by COVID-19 year. With the most instances overall in 2020, France claimed the lead, followed by Italy and Germany. In 2021, France continued to have high numbers, while Germany outpaced Italy overall in terms of instances. There was a noticeable increase in instances related to the Netherlands and Spain.

Germany surpassed all other nations in 2022 to become the country with the highest number of COVID-19 cases, followed by France. In other countries, there was a notable decline in the number of cases. By 2023, France had regained its lead, and there had been a noticeable decline in Germany's instances and a rise in Greece's.

These graphs show the overall number of instances for each nation while also illuminating the dynamic variations from year to year. France has a high incidence rate all the time, although there are significant variations in Germany and Italy. The graphs interactive features improve our analytical skills and offer a useful resource for comprehending how COVID-19 instances are spreading over the European countries.

Line graph

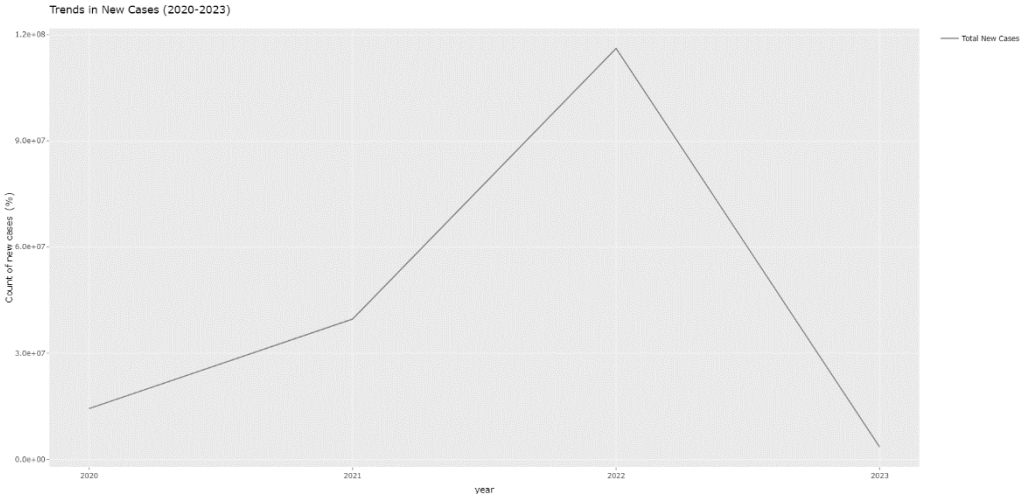

Figure 2Line graph count of new cases vs year.

The line graph illustrating the count of COVID-19 cases over the past three years reveals a consistent upward trajectory, marked by a significant surge in 2022 concurrent with the emergence of new virus variants. Although there is a decline in cases noted in 2023, the levels remain elevated. This observed trend, as depicted in the graph, offers valuable insights into the dynamic evolution of the European spread of the COVID-19 pandemic. The graphical representation serves as a critical tool for understanding the temporal progression of the pandemic, enabling ongoing monitoring of the effectiveness of implemented public health measures. This graph provides essential information for making informed decisions concerning strategic responses to the ongoing pandemic. Through a thorough analysis of the presented data, a comprehensive understanding of the pandemic's trajectory is attained, allowing for a nuanced approach to public health management. The graph facilitates adaptive measures to effectively address the challenges posed by COVID-19 on a global scale.

Bubble plot

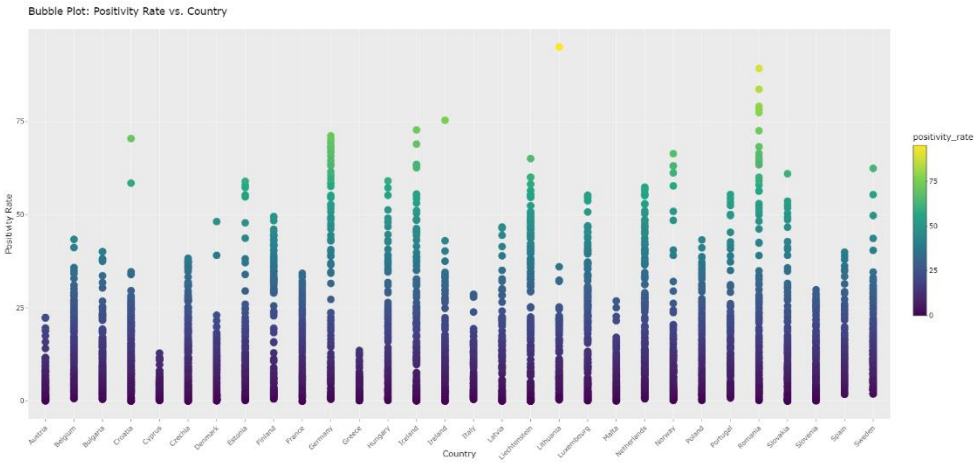

Figure 3Bubble plot for positivity vs Countries

The bubble plot illustrates the variation in the positivity rate among different countries, with the size of each bubble indicating the magnitude of this difference. The colour of each bubble corresponds to the specific positivity rate, providing a visual representation of the disparities across countries. This graphical representation serves as a comprehensive tool for discerning and comparing positivity rates, enabling a nuanced understanding of the global landscape. The interactive nature of the bubble plot enhances the viewer's ability to explore and interpret the data, making it a valuable resource for gaining insights into the diverse impacts of COVID-19 across various European countries.

Pie Chart

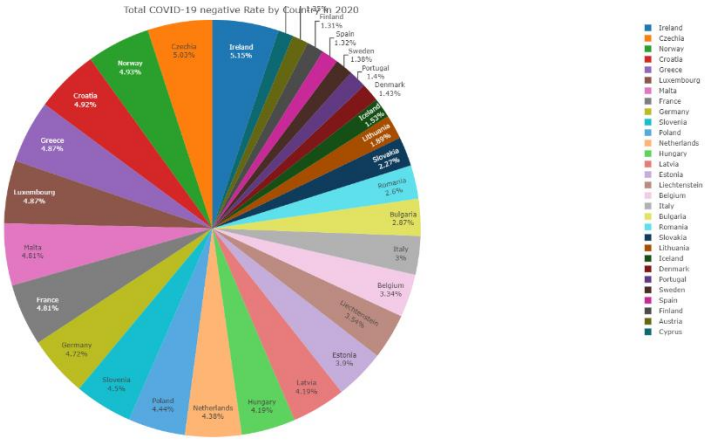

Figure 4Pie Chatr for Countries with negative rate

The pie chart is like an informative wheel representing different countries and how many people tested negative for COVID-19. Each slice of the pie is a country, and the bigger the slice, the higher the percentage of people who tested negative there.

When you look at the pie, you can see that some countries, like Ireland, have a big slice, meaning they have a high percentage of negative cases. On the other hand, countries like Finland have a small slice, indicating a lower percentage of negative cases.

It's a quick way to figure out which places might need more help to control the virus by identifying countries with a lower percentage of negative cases. It's also useful for checking if the rules in different countries are working well to keep the virus in check.

Line Graph

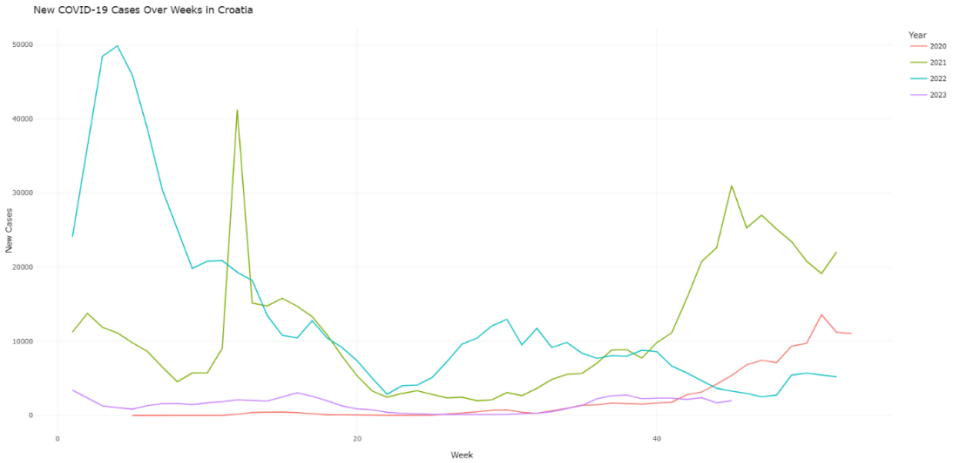

Figure 5 Number of Covid-19 cases with Croatia ,2020-2023

The graph illustrates a declining trend in new COVID-19 cases in Croatia over time, with notable spikes in early 2021 and late 2022, possibly attributed to emerging virus variants and relaxed public health measures. Despite these fluctuations, Croatia's case count remains comparatively lower than many other countries, influenced by factors such as its smaller population, robust public health system, and high vaccination rate.

Continuing preventive measures, such as vaccination, mask-wearing, social distancing, and frequent handwashing, is essential to curb the virus's spread. Immediate testing is recommended for individuals displaying COVID-19 symptoms. Overall, the graph reflects positive trends in Croatia's COVID-19 situation, emphasizing the importance of sustained public health efforts.

Line Graph

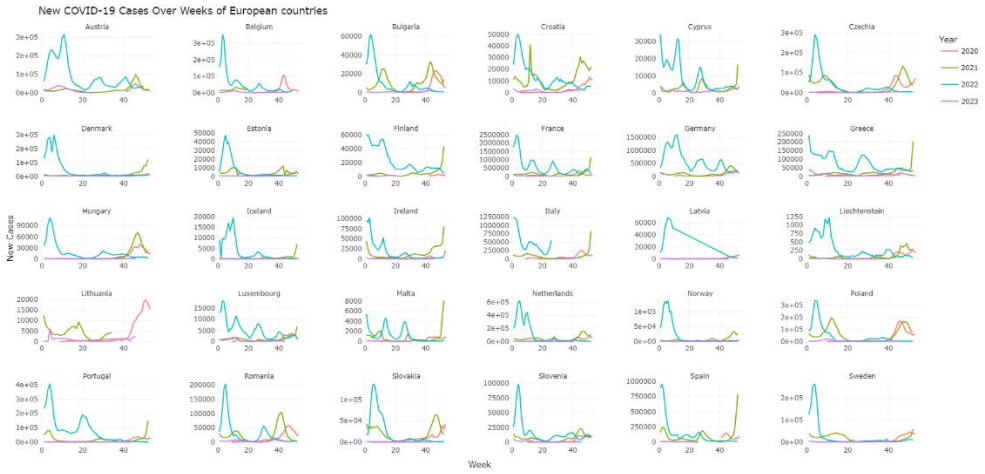

Figure 6Graph for all 30 countries with new cases vs week

Comparable to Figure 5, which I completed for all 30 countries, this figure will display the COVID 19 increase and decrease with new cases each week, with the year indicated by the lines. Required country we want to see the effects of covid-19 we can just see the country name and we will be having the details of that country

Scatter plot

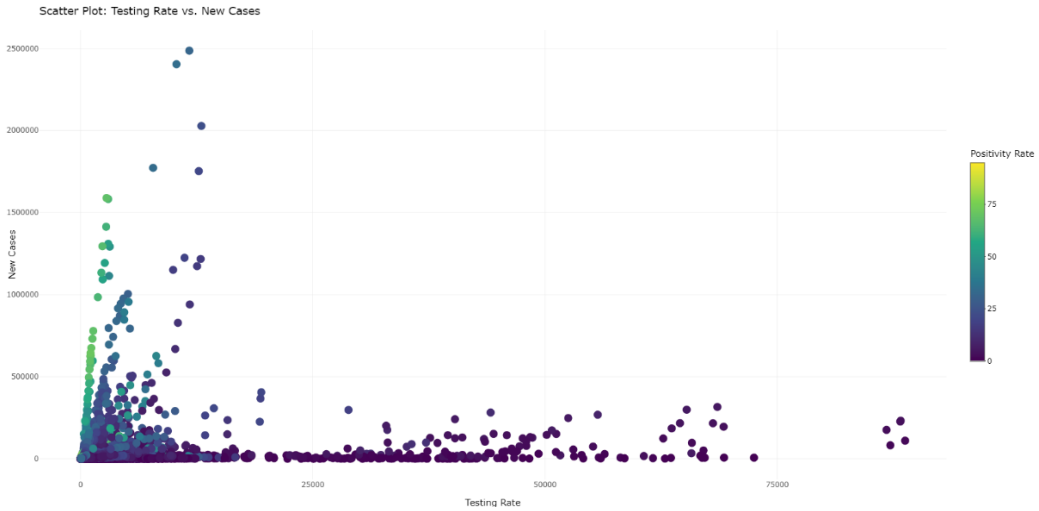

Figure 7Scatter Plot: Tesiting Rate vs Cases

This graph illustrates the interplay between testing rates, new COVID-19 cases, and positivity rates across European countries. The x-axis represents the testing rate, the y-axis denotes the number of new cases, and each point's colour signifies the positivity rate. A positive correlation emerges between testing rates and new cases, indicating that higher testing rates are associated with increased case identification. Interestingly, countries with elevated testing rates demonstrate lower positivity rates, suggesting adept case identification and isolation strategies.

Box plot

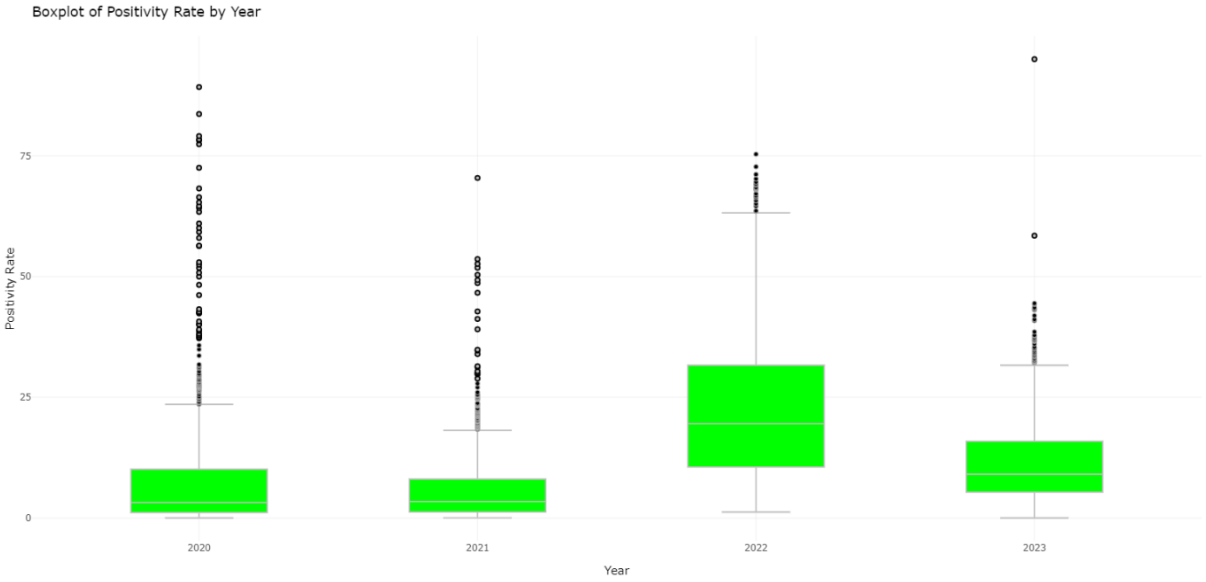

Figure 8Boxplot of positivity rate by year

The graph shows a steady decline in the COVID-19 positivity rate since 2020, which translates into a lower annual percentage of positive tests. This decline can be attributed to multiple factors. A higher rate of vaccination is essential for preventing COVID-19 infections, which in turn lowers the number of positive test results. Furthermore, improvements in test availability and accuracy have made it easier to identify infected people. The declining positivity rate could also be caused by modifications to the virus's transmissibility or modifications to public behaviour, such as increased mask use and social distancing. Although this trend shows that the pandemic is being contained, sustained control still depends on vaccination and adherence to preventive measures.

Suggested Hypothesis

The examination of the COVID-19 dataset for European nations (2020–2023) generates theories that need more research. The bar graph (Figure 1) indicates that vaccination rates, healthcare infrastructure, and population density may have an impact on variations in the impact. The line graph (Figure 2) suggests that new virus variants may be responsible for the 2022 spike, while improved public health interventions may be the cause of the subsequent 2023 declines. The hypothesis that variations in positivity rates could be a sign of variations in testing methods and healthcare accessibility is brought up by the bubble plot (Figure 3). Higher percentages of negative tests may be associated with more successful testing techniques, according to Figure 4's pie chart. Based on the examination of Croatia's pattern (Figure 5), it is hypothesised that a smaller population, strong public health policies, and a higher vaccination rate may be the causes of the country's lower case count. The hypothesis that patterns in case fluctuations may reveal common trends or specific country-level factors is prompted by the graph of all 30 countries (Figure 6). Higher testing rates may result in lower positivity rates, which are a sign of early case identification, according to the scatter plot (Figure 7). Lastly, the boxplot (Figure 8) implies that more accurate testing and successful vaccination campaigns might be associated with the drop in positivity rates. These succinct theories serve as a basis for more thorough examinations of the variables influencing COVID-19 dynamics in European nations.

Table of Figures

References

European Centre for Disease Prevention and Control (2022). *Data on testing for COVID-19 by week and country*. [online] European Centre for Disease Prevention and Control. Available at: https://www.ecdc.europa.eu/en/publications-data/covid-19-testing.