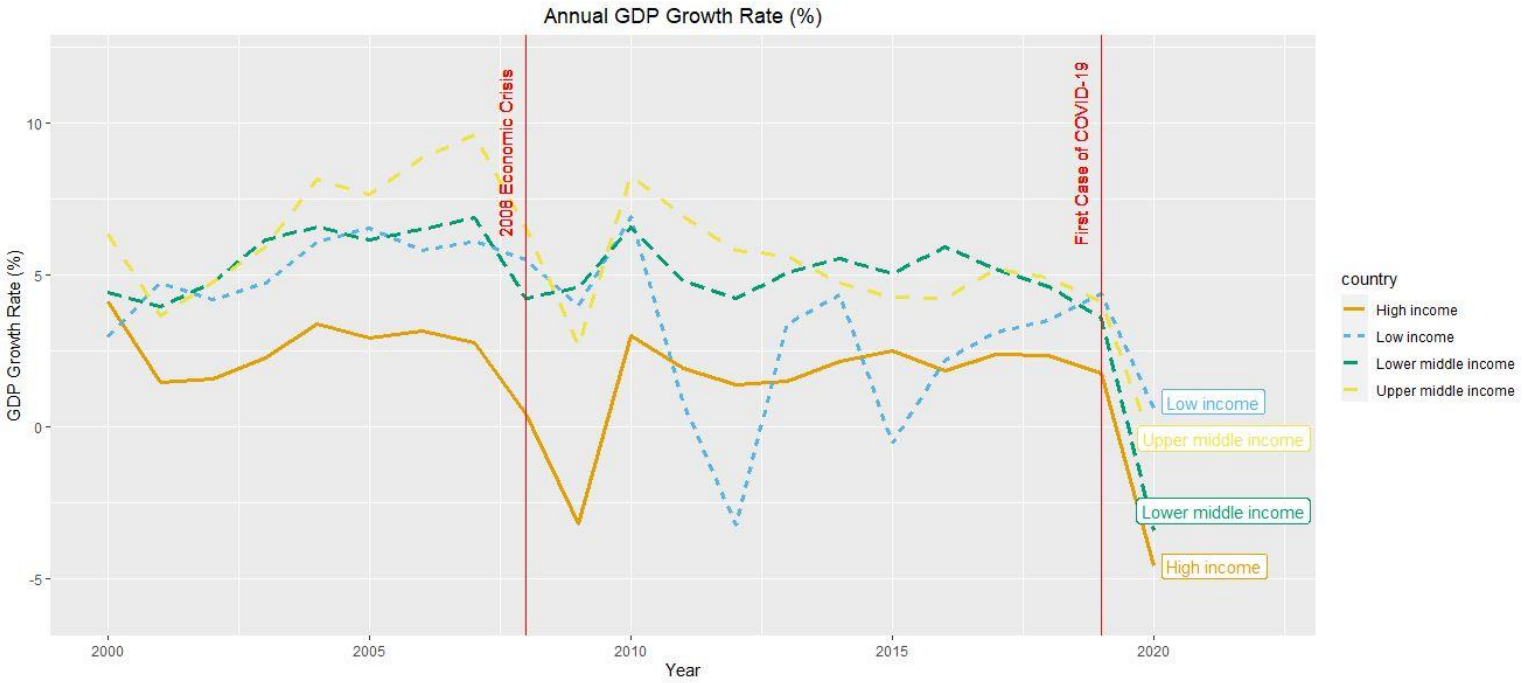
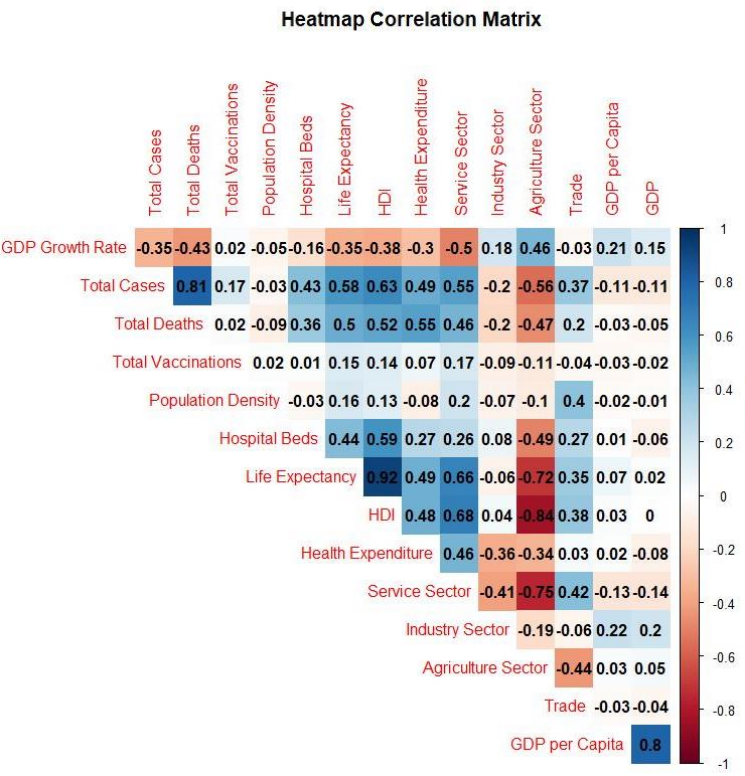
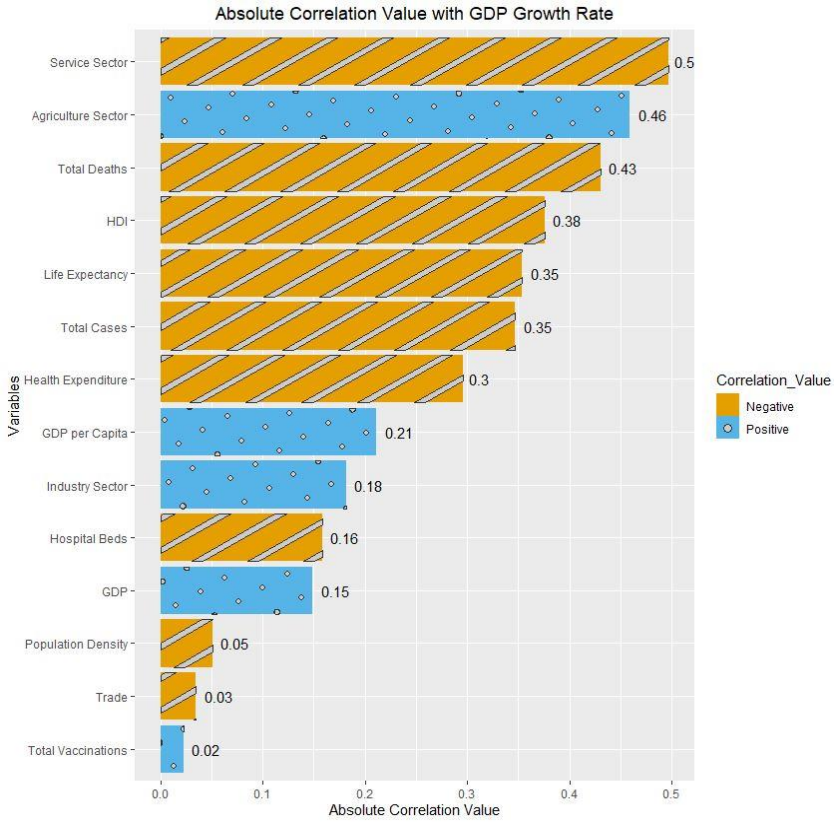
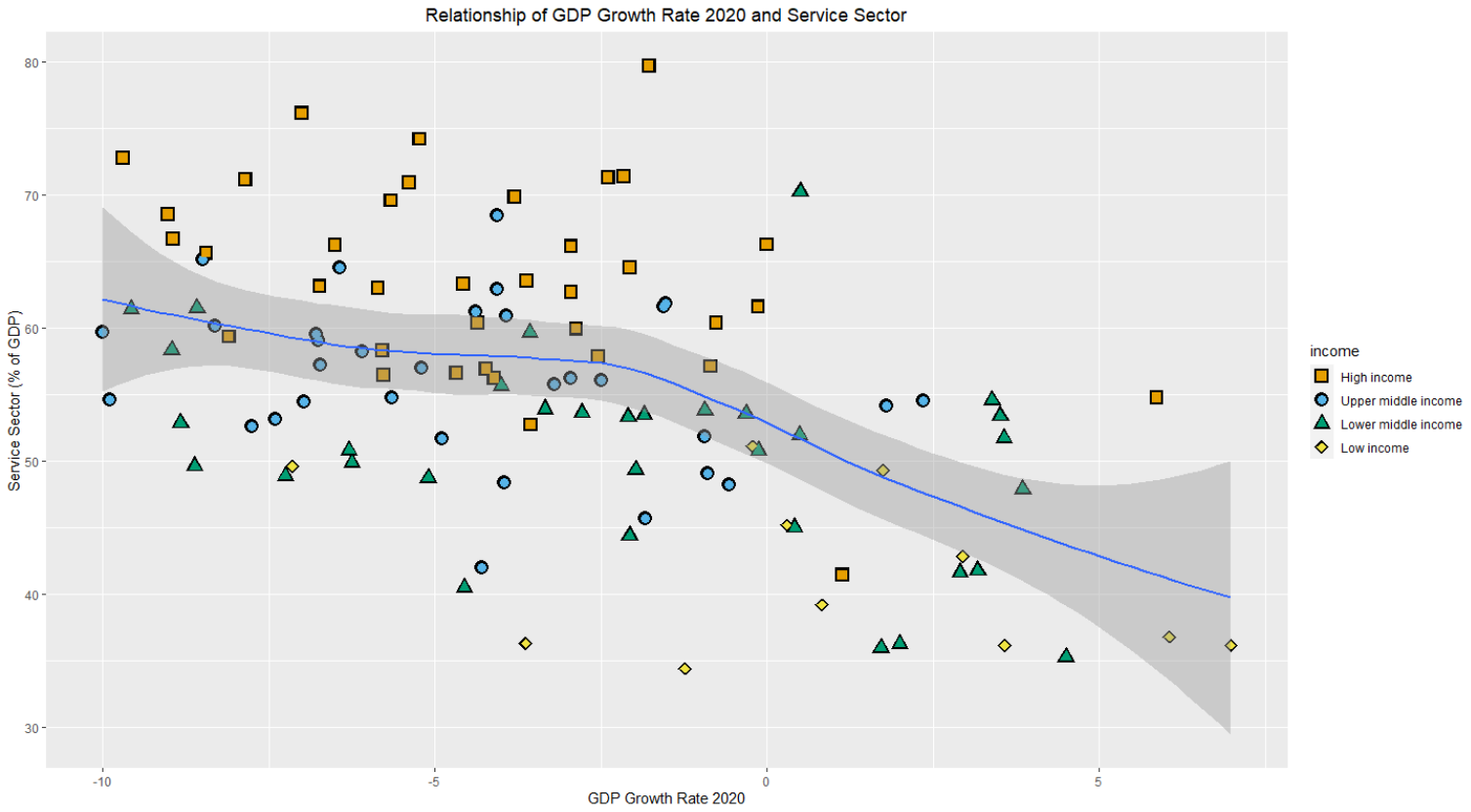
INF4000 Data Visualisation – Report
No Reg: 210132797
Words: 3051

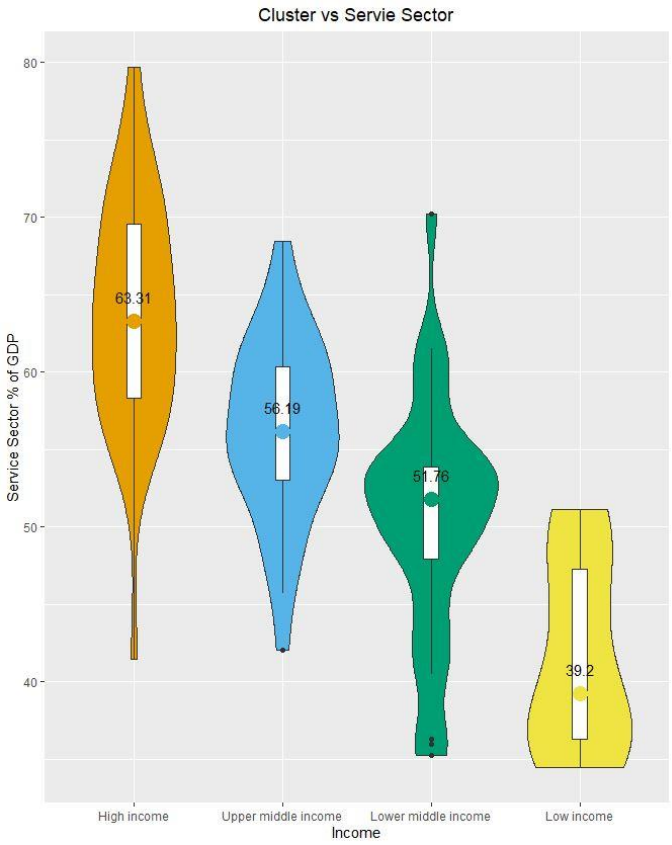


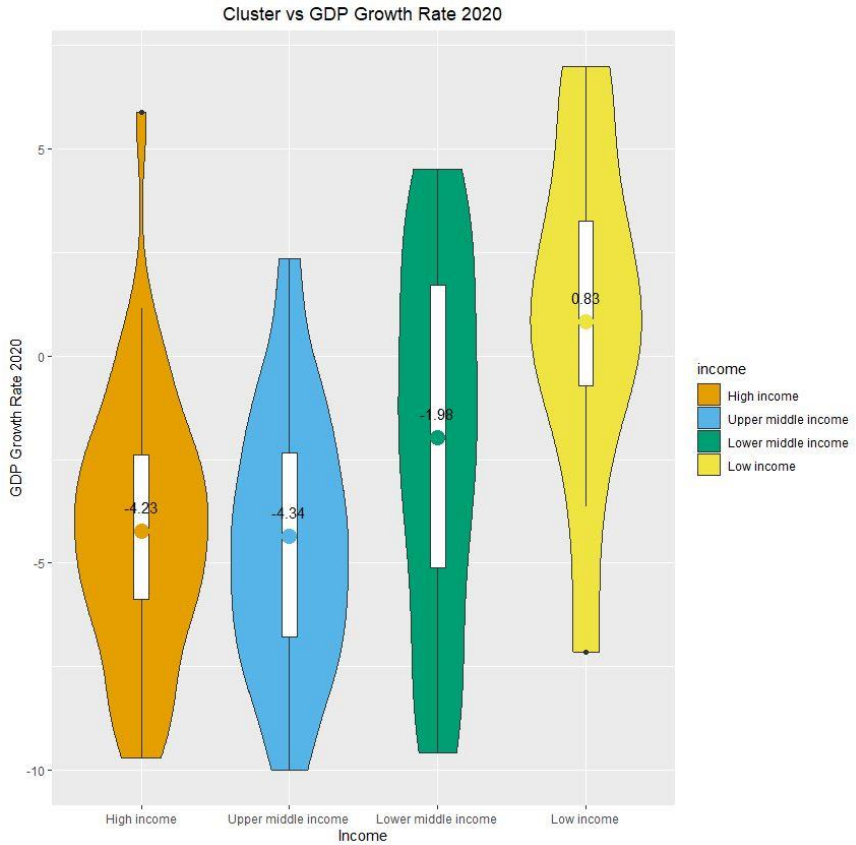




Top Centre: Figure 1, Annual GDP Growth Rate
Bottom Left: Figure 2, Heatmap Correlation Matrix
Bottom Right: Figure 3, Bar Chart Absolute Correlation Value







Top Centre: Figure 4, Scatter Plot Relationship of GDP Growth Rate and Service Sector
Bottom Left: Figure 5, Box Plot Service Sector by Income Level
Bottom Right: Figure 6, Box Plot GDP Growth Rate by Income Level

1. Knowledge Building

The spread of SARS-COV-2 virus that brought the COVID-19 pandemic has affected not only the healthcare system of countries, but the impact can be seen across critical sectors, including food supply (Aday & Aday, 2020), education (Daniel, 2020), global diplomacy (Davies & Wenham, 2020), and especially economy(Verma et al., 2021). The implications of the economic impact cannot be underestimated as it can affect health and wellbeing (Winters et al., 2012), families (Boushey & Cherry, 2003), leisure and tourism (Smeral, 2009), and businesses (Peric & Vitezic, 2016). With the sheer impact of COVID-19 to the economic sector that affects almost every aspect of our lives, a comprehensive assessment needs to be conducted to measure how big the impact of COVID-19 pandemic brought to our economy and the variables that contribute to it. This study will use the visualisation method to analyse the economic downturn during the covid-19 pandemic and its associated factors.

As the first goal is to assess whether COVID-19 caused an economic crisis or not, Figure 1 shows the same pattern of the 2008 economic crisis (Kotz, 2009) and covid-19 that appears in 2020. Both show a downturn pattern in the GDP growth rate for all income levels. As asserted by Prager et al. (2017) and Smith, A.T (2009), a pandemic might negatively affect the GDP by 3-4%; the graph also shows the relatively similar pattern, high-income countries were hit hardest with a 4.6% decrement, followed by lower-middle-income countries with -3.4%, and upper-middle-income and low income each one having -0.7% and 0.6% for their GDP growth rate respectively.

Following the first figure, an assessment of factors that impact the economic downturn during covid is crucial. By conducting this assessment, governments can more focus on the related factors. Figure 2 and 3 shows factors that most correlated with the economic downturn during the covid-19 pandemic. It can be seen from the picture that the Service sector has the most significant correlation score with the GDP Growth rate. This finding is in line with the study by Pak et al. (2020), suggesting that the service sector is one of the factors that highly associated with the economic downturn. However, correlation does not always mean causal association between the variables; hence further research outside this study needs to be conducted to assess the causal relationship between the variables.

Figure 4 shows the further relationship between the service sector and the GDP growth rate 2020. The inverse relationship between the sector service and GDP growth rate can be seen from this figure. Countries with a more significant contribution of sector service to their GDP tend to have a lower GDP growth rate. The findings support the claims of Fernandes (2020), who argues that the service sector will suffer the most from this pandemic. As several countries perform lockdown, the global travel industry, such as airlines and cruise companies, face a massive reduction in their activities. Hospitality sectors like tourism, restaurants, hotels, casinos, and hotels are deserted as people avoid public spaces. Moreover, stock markets collapsed in early 2020, which suggests that the pandemic has negatively affected financial service sectors like banks and stock markets (Fernandes, 2020)

Based on Figures 1 and 4, we can see a separation between income levels regarding their service sector contribution and GDP growth rate. The separation between each income is presented in Figures 5 and 6, showing that countries with higher income tend to have more significant service sector contributions towards their GDP. In contrast, lower-income countries tend to rely more on the agricultural sector. The results align with studies from Estrada, G. et

al. (2013) and Buera & Kaboski (2012), suggesting that countries with higher income tend to have more significant service sector contributions towards their GDP.

Concluding the prior visualisations, it is now clear that indeed the COVID-19 cause an economic downturn that affects countries at every income level with a significant association with the service sector. Countries that rely more on the service sectors, like most high-income countries, are more vulnerable to the economic impact of the pandemic. With this idea in mind, governments can start to mitigate by implementing several actions such as strengthening international cooperation, increasing infrastructure construction, and creating more innovative reform (Mou, 2020) to overcome the adversity.

2. Theoretical Frameworks

2.1 Ask a Question

The visualisation realised using ASSERT model. There is different question answered by each visualisation. Each question is formed by asking more detailed and critical question regarding the previous visualisation. However, this part only discuss the theoretical frameworks of the scatter plot and the heat map.

The purpose of the heatmap is to tackle the question of "What factors that highly associated with the worlds' economy during the covid 19 pandemic?". Answering this question is vital since there are a lot of factors and variables from the data. Without a proper knowledge on the corelation between the GDP growth rate and the factors, the solution to this issue might not concluded. Based on the visualisation, there are other question emerges. "How's the service sector affect GDP growth rate in during the covid-19 pandemic?". Although the question might be partially answered by the first visualisation, a common audience might find difficulty in visualising the relationship between service sector and GDP growth rate. In order to show the direct relationship between those variables, a scatter plot that show GDP growth rate on X-Axis and service sector on the Y-axis is added.

2.2 Search the data

The data is gathered from 2 sources, the first data is from World Bank Indicators (WDI, 2019) that comprises of GDP, GDP Growth rate, Service sector, Industrial Sector, Agriculture Sector, Trade, and healthcare expenditure. The second data is gathered from the OWIDCOVID19 (Owidbot, 2021) that contain the rest of the variables. Both data is an open data that can be accessed (not edited) by everyone, thus satisfy the accessibility and data credibility requirement.

2.3 Structure the data

The data consists of several data types. The first is continuous data that describes most variables (e.g., GDP growth rate, GDP, total cases, and deaths). The second type is categorical, which describe the income level and region of the countries. Before conducting the visualisation and visual analysis. The data needs to be cleansed from missing values and outliers due to missing values, and outliers can be somewhat troublesome, primarily when we conduct mathematical or statistical analysis to our data. The missing values may lead to computational error, unexpected values, and incorrect conclusions. (Zainal Abidin et al., 2018).

2.4 Envision the data

Due to the sheer number of factors from the data set, correlation value is used to determine which factors that has the most significant association with our variable of concern (GDP Growth rate). To measure the strength of the relationship between variables, the Pearson correlation coefficient is used. Correlation values near zero indicate a weaker correlation between the variables. As the correlation value approaches 1, it indicates a strong positive correlation between two variables, which suggests that if one value is increased, the other will also increase. In contrast, if the number gets closer to -1, there is a strong negative correlation, so if one value increases the other decreases. Finally, the correlation value leads to the deeper understanding on how one variable behaves when there are change in other variables.

2.5 Representing the visualisation

Heatmap is considered the best at representing the correlation between variables. There are two axes in heatmap with each contains the same variables which form a matrix like structure. Each cell in the matrix contains a correlation value of the relationship between variable in that specific row and column. Due to the design and accessibility consideration, the correlation value is presented with both numbers and colour.

After knowing the most significant correlation, the next step is to visualise the relationship between the variables. Since all variables used are non-logarithmic (relatively same range) and non-polar (no angles or complex numbers), a cartesian coordinate is used as the base coordinate system. In addition, points are considered the best at representing multiple data points compared to lines. Considering the aesthetic, I choose different colours and shape to represent different income levels. In addition to that, a smoothed regression line is shown to emphasize the relationship between variables.

2.6 Tell a story

The overall visualisation is designed sequentially as if the subsequent visualisation is an answer to the question that emerges after critically assessing the prior visualisation. The first example is shown to answer whether an economic downturn happened during the pandemic. The second visualisation is shown to answer what factors affect the economic downturn during the pandemic. After knowing the most associated factors, the third question visualises the relationship between the factor and the GDP growth rate. Finally, the last visualisation is shown to conclude the prior visualisations of how the economic impact of the pandemic is different between various income levels.

3. Accessibility

3.1 Colour-blindness and Colour-related issues

All visualisation provided in this report is built utilising Okabe-Ito colour pallet, a colourblind-friendly colour pallet (Okabe & Ito, 2002). This colour pallet has been adjusted to fulfil the needs of three types of colour-blindness, protanope, deuteranope, and tritanope. However, due to the limited range of colours that audiences can distinguish, especially those with colour-blindness, additional properties are introduced to the visualisation.

Shapes are added to the scatter plot to distinguish one income level from another. This additional property is introduced to accommodate audiences with total colour-blindness. In addition to that, the shape is relatively easier to be distinguished from one shape to another since everyone can better distinguish a square from a triangle shape compared to colours, which have broader terms, names, and different people might call one colour different from one another, especially colourblind people.

In addition to colours and shapes, numbers are also used for the heatmap. Although a colourblind-friendly pallet has already been introduced for the heatmap, a value from a continuous variable's colour representation is relatively harder to distinguish. This shortcoming affects people with colour-blindness and non-colourblind people who sometimes find difficulty in differentiating between light and dark colours. To overcome this issue, a numerical value is added to the heatmap to represent the correlation value; hence, the difference between the values is distinguishable.

Lastly, texture and line type properties also added into the visualisation for the bar chart and the line chart.

3.2 Data Accessibility and Credibility

The data was retrieved from two sources. The first one was retrieved from WDI data that included in the WDI package in R. This data contains variables like GDP growth rate, GDP per capita, service, industrial, and agriculture percentage of GDP, as well as income classification and countries' iso code. The second dataset was retrieved from the OWID Covid-19 dataset. These datasets contain COVID-19 disease prevalence factors such as the number of cases, deaths, and vaccination tolls. These data comply with Few's (2009) attributes of meaningful data that are high in volume, historically accurate, consistent in structure and criteria, multivariate, atomic, clear, and accurate, structured dimensions as well as credible and accessible data source.

3.3 Knowledge on Statistic

Knowledge on statistics. People with adequate knowledge on statistics might find it easy to understand the boxplot and correlation value shown by heatmap and bar chart in vis 1 and 2. However, people without the basic knowledge on statistic might find it difficult to read and understand both visualisations.

Although the range, quartiles, and median already shown explicitly in the boxplot, audiences without knowledge on basic statistics might have difficulties interpreting the chart. In order to overcome this shortcoming, I added the median value of each income level into the boxplot. In addition to that, the violin plot was also added to emphasize the density property that the boxplot has not addressed. However, introducing more features and properties to the graph may result in an overly complex and overcrowded graph.

4. Visualisation Choices

4.1 Line Chart

A line chart is used to visualise multiple trend lines due to its popularity in visualising time-series data (Wang et al., 2018). This form of graph is chosen to visualise GDP growth rate, which has negative values and line-cross pattern. Due to the variable's negative value and line-cross pattern, a line chart is preferred over the area or stacked area chart due to its ability to visualise a crossing line and a negative value that is impossible to visualise by area chart. Another approach of using multiple single-line chart is unnecessary for this visualisation due to the small number of trendline shown in the graph. However, If the number of trendlines that want to be visualised is numerous (>5), multiple single-line charts are considered better to make the visualisation less crowded, so the audience can better grasp the graph. This visualisation is also useful for visually comparing trends in a span of time. For example, we can directly see the same pattern in GDP growth rate for both the economic crisis of 2008 and during the covid-19 pandemic that shows both similarities in terms of the economic downturn during the crisis.

4.2 Heatmap and Bar chart

To better understand the significant correlation value of multiple variables, the Correlation heatmap was chosen due to its ability to describe data visually and statistically with correlation value. However, instead of using a traditional heatmap that only describes the correlation value by colour, I also visualise the value of the correlation value, which is coloured according to the heatmap colour spectrum. By using this approach, audiences, especially those who are colourblind, can distinguish and understand the correlation matrix better. Other opinions might argue that a dendrogram also can visualise this form of data, however dendrograms mainly used to show relationship in a hierarchical clustering or classification. In addition to that, a bar chart is also presented to complement the heatmap. With a bar chart, we can exploit audiences' pre-attentive visual, so they may quickly grasp the quantitative ratio and comparison between correlation values since the difference between correlation values is presented by how high the bars are, not only based on numbers and colours which might take longer to be processed.

4.3 Scatter Plot

A scatterplot can be used to understand the relationship between two variables better (Wang et al., 2018). The GDP growth rate is chosen as the independent variable in the X-axis and the service sector as the dependent variable in the Y-axis. In addition to the scattered points, a smoothed regression line also emphasises the relationship between both variables. This combination of scatterplot and a regression line is chosen over several other visualisation options such as connected scatter and line chart. A connected scatter plot and a line chart will show overly complicated and crowded visualisation as too many points, and lines are shown in a single graph that may cause confusion and multi-interpretation on the audiences.

4.4 Box plot and Violin plot

A visualisation that combines both box plot and violin plot is utilised to combine the strength of both graphs. Boxplot is great at visualising numerical variables. A boxplot shows four main factors of a variable: the centre or median, spread, asymmetry, and outliers. However, the boxplot lacks the ability to show the density of the data points. In addition to that, although the median of the categories is shown in the boxplot, they are not shown explicitly and may cause audiences to estimate the value. Hence, combining both forms of visualisation create a

comprehensive visualisation in which the centre, spread, asymmetry, outliers, and the variable's density is described. In addition, a point and a label of the median are shown to emphasise and give numerical detail of the centre point of the data.

5. Implications and Improvements

5.1 Implications

Even though the goal of visualisation provided in this study is to deliver knowledge and information as objective as possible, there will always be an unintentional misleading information embedded in the visualisation, or a built-in bias (Huff, 1954). Although there are probably several misinterpretations from the visualisation, let us focus more on the most crucial and common misinterpretation and bias in the visualisation.

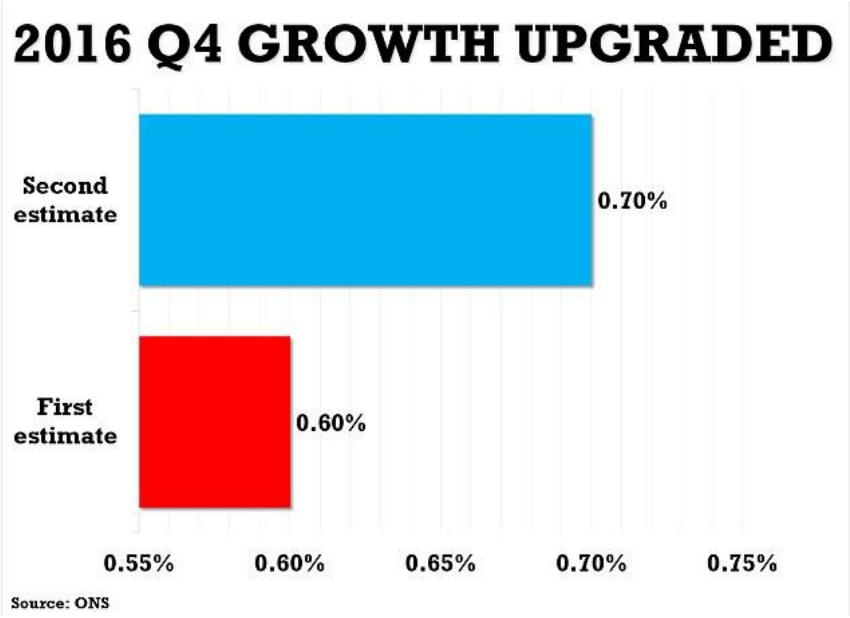


Figure 7: Truncated Bar Graph

The first potential bias is from the bar chart. The most common bias presented in the bar chart is the truncation bias. This form of practice often misleads viewers into exaggerating the difference or ratio in the graph (Yang et al., 2021). An example of this misleading graph can be seen in Figure 7 retrieved from the dailymail.co.uk by Sculthorpe (2017). The picture visualises a 0.1% difference between the first and the second estimate of the 2016 Q4 economic upgrade; the difference between the categories is exaggerated due to the usage of truncation, narrowing the full range of the scale. Considering the potential bias, I provide a bar chart without truncation that shows the full range of the correlation value. However, instead of showing negative values to the opposite direction of the positive direction, the negative values are shown by colour and texture. Although this idea initially aims to emphasise the absolute correlation value between variables, it must be explained further as it may lead to misunderstanding.

Second, as Stigler S (2005) argued, correlation value does not imply causation. It simply tells us about the significance of mathematical association between the variables. Further study needs to be conducted to identify causation between variables, which is not covered by this report. To overcome the potential bias and misinterpretation of the correlation value, a scatter plot is provided to emphasise the relationship between the variables.

5.2 Improvements

The line chart can be improved by considering the line blocking aspect. It is better to place dashed lines on top of the full line so there will be no visual blocking on the display. In addition to this, an additional indicator or visual cues can also be introduced to emphasise the similarity between the line pattern in 2008 economic crisis and in 2019 COVID-19.

Adding a more contrast background gives more highlight to the data points; hence the audiences may find it easier to spot the data points, especially on the scatterplot.

The third improvement can be done on the box and violin plot. Although the information visualised are already complete, a simpler approach without reducing information can be implemented on this chart. For example, instead of stacking boxplot and violin plot together, a range, median, and quantile marker can be inserted to the violin plot. In addition to that, introducing a more suitable font size and font style also can highlight the median text label.

Lastly, the overall quality of the image and its components (i.e., axis label, axis title, data points label) needs to be improved. An improvement on the image's resolution, quality, and aspect ratio are crucial in giving audiences a pleasant visual journey when exploring the visualisation.

6. Bibliography

Aday, S., & Aday, M. S. (2020). Impact of COVID-19 on the food supply chain. In *Food Quality and Safety* (Vol. 4, Issue 4, pp. 167–180). Oxford University Press. https://doi.org/10.1093/fqsafe/fyaa024

Boushey, H., & Cherry, R. (2003). The Severe Implications of the Economic Downturn on Working Families. *WorkingUSA*, *6*(3), 35–54.

Daniel, S. J. (2020). Education and the COVID-19 pandemic. *Prospects*, *49*(1–2), 91–96. https://doi.org/10.1007/s11125-020-09464-3

Davies, S. E., & Wenham, C. (2020). Why the COVID-19 response needs international relations. *International Affairs*, *96*(5), 1227–1251. https://doi.org/10.1093/ia/iiaa135

Estrada, G., Acharya, Y., Batten, A., Brimble, P., Doung, P., Gulamov, M. I., Hossain, M. Z., Mellor, D., Park, D., & Rahman, S. (2013). *The Service Sector in Lower-Income Asian Economies ADB Economics Working Paper Series*. http://ssrn.com/abstract=2262577www.adb.org/economicsPrintedonrecycledpaperPrintedinthePhilippinesElectroniccopyavailableat:https://ssrn.com/abstract=2262577

Fernandes, N. (2020). *Economic effects of coronavirus outbreak (COVID-19) on the world economy*. https://ssrn.com/abstract=3557504

Few, S. (2021). *Now You See It*. Amsterdam University Press.

Kotz, D. M. (2009). The financial and economic crisis of 2008: A systemic crisis of neoliberal capitalism. *Review of Radical Political Economics*, *41*(3), 305–317. https://doi.org/10.1177/0486613409335093

Mou, J. (2020). Research on the Impact of COVID19 on Global Economy. *IOP Conference Series: Earth and Environmental Science*, *546*(3). https://doi.org/10.1088/1755-1315/546/3/032043

Okabe, M., & Ito, K. (2002, November 20). *Color Universal Design (CUD) / Colorblind Barrier Free*. Color Universal Design (CUD) - How to Make Figures and Presentations That Are Friendly to Colorblind People -. Retrieved 30 January 2022, from https://jfly.uni-koeln.de/color/

Owidbot. (2021). *Data on COVID-19 (coronavirus) by Our World in Data* [Data set]. https://github.com/owid/covid-19-data/tree/master/public/data

Peric, M., & Vitezic, V. (2016). Impact of global economic crisis on firm growth. *Small Business Economics*, *46*(1), 1–12. https://doi.org/10.1007/s11187-015-9671-z

Prager, F., Wei, D., & Rose, A. (2017). Total Economic Consequences of an Influenza Outbreak in the United States. *Risk Analysis*, *37*(1), 4–19. https://doi.org/10.1111/risa.12625

Sculthorpe, T. (2017, February 22). *British economy grew by 0.7% in final 3 months of 2016*. Mail Online. Retrieved 30 January 2022, from https://www.dailymail.co.uk/news/article-4248690/Economy-grew-0-7-final-three-months-2016.html

Smeral, E. (2009). The impact of the financial and economic crisis on European tourism. *Journal of Travel Research*, *48*(1), 3–13. https://doi.org/10.1177/0047287509336332

Stigler, S. M. (2005). Correlation and Causation: A Comment. *Perspectives in Biology and Medicine*, *48*(1x), 88-S94. https://doi.org/10.1353/pbm.2005.0045

The World Bank, World Development Indicators (2012). *GDP growth (annual %)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

The World Bank, World Development Indicators (2012). *Manufacturing, value added (% of GDP)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NV.IND.MANF.ZS

The World Bank, World Development Indicators (2012). *Services, value added (% of GDP)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NV.SRV.TOTL.ZS

The World Bank, World Development Indicators (2012). *Industry (including construction), value added (% of GDP)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NV.IND.TOTL.ZS

The World Bank, World Development Indicators (2012). *Agriculture, forestry, and fishing, value added (% of GDP)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NV.AGR.TOTL.ZS

The World Bank, World Development Indicators (2012). *Trade (% of GDP)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NE.TRD.GNFS.ZS

The World Bank, World Development Indicators (2012). *GDP (constant LCU)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NY.GDP.MKTP.KN

The World Bank, World Development Indicators (2012). *GDP per capita (current LCU)* [Data file]. Retrieved from http://data.worldbank.org/indicator/NY.GDP.PCAP.CN

Tilly, C. (2011). The impact of the economic crisis on international migration: A review. In *Work, Employment and Society* (Vol. 25, Issue 4, pp. 675–692). https://doi.org/10.1177/0950017011421799

Verma, P., Dumka, A., Bhardwaj, A., Ashok, A., Kestwal, M. C., & Kumar, P. (2021). A Statistical Analysis of Impact of COVID19 on the Global Economy and Stock Index Returns. *SN Computer Science*, *2*(1). https://doi.org/10.1007/s42979-020-00410-w

Wang, Y., Han, F., Zhu, L., Deussen, O., & Chen, B. (2018). Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2), 1141–1154. https://doi.org/10.1109/TVCG.2017.2653106

Winters, L., Mcateer, S., & Scott-Samuel, A. (2012). *Assessing the Impact of the Economic Downturn on Health and Wellbeing*.

Yang, B. W., Vargas Restrepo, C., Stanley, M. L., & Marsh, E. J. (2021). Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition*, *10*(2), 298–311. https://doi.org/10.1016/j.jarmac.2020.10.002

Zainal Abidin, N., Ritahani Ismail, A., & Emran, N. A. (2018). Performance Analysis of Machine Learning Algorithms for Missing Value Imputation. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 9, Issue 6). www.ijacsa.thesai.org