

Visualisation Report

Jordan Marshall - 18256716

What is Data Visualisation?

Data visualisation aims to convert data into easy to interpret graphs/diagrams. It makes it much easier to understand the trends in the data. Instead of reading line by line of a dataset, the user can view a graph that summarises a section of that dataset or even the whole dataset. The graphs have structure, are easy to read and easy to understand. The user can see the magnitude of data compared to each other too. Data visualisation uses many techniques to achieve the end goal of a clear and concise view of the data. These include data cleaning, exploring data structures, identifying outliers, identifying trends/patterns and presenting the end results to the user.

Simplifying

Simplifying relates to summarising a large volume of data and removing unnecessary features. A dataset can contain hundreds or even thousands of data points and it can be very difficult to understand trends and patterns. For example, a large business might have 3 million orders in a year, the business wants to see trends related to where products are being shipped to, which products have higher profit margins or which product is least popular. In a business, decisions are very important. Instead of going through all 3 million entries, a few simple graphs can show the trends and margins between products. Overall the principle of simplifying lowers the complexity of the data.

In my COVID-19 visualisation, I used daily reports to view the distribution of cases over most countries worldwide. The user is able to specify a date that they want to view. I wanted to show how a country was handling COVID-19 based on case numbers and compare them to other countries. These case numbers include confirmed cases and death cases. I also compared the incident rate for each country. Incident rate is the rate at which a new case occurs over time in relation to a certain population size. As well as this, I also provided additional statistics to reinforce the graphs provided.

I plotted a choropleth map of the world where the number of confirmed cases defines the colour of the country. Choropleth maps are useful for displaying geographic data. This is also known as a heatmap. Heatmaps are used to display a range among values. This made it easier to compare countries with each other based on confirmed cases. In some of my graphs, I selected the top 20 entries. This made it easier to view a selection of the data without showing all the countries.

The benefit of using Dash and Plotly allowed me to make my graphs interactive. The user can zoom into the graphs and even select certain countries if wanted to view. The user can also save the graph state into an image file, which can be useful for documents or presentations.

Understand Magnitudes

Magnitude means the sheer quantity or distance of an object or thing. In the case of COVID-19, it is important to understand the population size of countries and the cases in relation to them. The United States of America has a population of approximately 330 million people. That means they will have a large number of confirmed cases and death cases in comparison to other countries of smaller population. Incident rate is the rate at which a new case occurs over time in relation to a certain population size. I use a bar chart to compare the top 20 highest incident rate countries and also the lowest incident rate countries. As seen in the “Top 20 Highest Incident Rate” graph, the US had a much higher incident rate compared to countries like Russia or Brazil.

Another way of displaying the magnitude of the data is using pie charts. Pie charts are able to show the proportion of each attribute given. Each element in the pie chart is relevant to each other and its corresponding size. It is simple and easy to understand. I used the pie charts to show the difference between the top 20 counties with the highest confirmed cases and death cases (graphed separately).

I used a scatter map to compare each country in respect to confirmed and death cases. Each country is also coloured based on a heatmap. The heatmap represents confirmed cases. Heatmaps are used to display a range among values. The more yellowish the dot, the more confirmed cases it had compared to other countries. The more blueish the dot, the less confirmed cases it had compared to other countries. As again shown, the US had more confirmed and death cases in respect to other countries.

Use of Color

The use of colour makes visualisations more visually appealing. It is a crucial method to convey quantitative information from a dataset. Colours allow for important data to be highlighted. I used bright colours in my graphs to make key values more important. This is very visible in my use of scatter graphs, choropleth maps and pie charts. The use of a heatmap allowed me to represent the data and map to a colour. The brighter the colour the more important it is. Brighter colours represent more important attributes while darker colours represent less important data. I decided to use a dark theme for my web application as it would make the graphs more visible. Plotly had a built in dark theme

which I used. Using bright colours in contrast with the dark theme made the graphs stand out more. Overall, the use of contrasting colours is a technique used to contrast and compare different data entries. If there was a use of a single colour, it is a technique for showing continuous data.

I tried to avoid using too many colours as it would be too confusing. By using a limited number of colours it allows the user to comprehend the data more. The user could become overwhelmed with too many colours.

Use Structure

Using structure in any data visualisation dashboard is very important. Having relevant information near to each other can help with understanding trends and patterns in the dataset. My dashboard is designed with simplicity in mind. Graphs are laid out in two columns, where each row's graphs are related to each other. Take for example both bar charts, one representing confirmed cases and the other representing the death cases. Both graphs are similar as they select the top 20 countries in each category, but also are both bar graphs.

At the top of my dashboard I display statistics from the dataset. This includes total confirmed cases, total death cases, mean confirmed cases, mean death cases and ratio between total confirmed and total death cases. This is all relevant to understanding the graphs below. It is a quick overall summary of the dataset.

Each type of graph has a specific use when conveying data. Pie charts are used to show proportions. They show the user the breakdown of each value when combined as a whole. One disadvantage of using a pie chart is that you can only use a few values. For example, I use the top 20 countries in each case as it can fit on the pie chart without being confusing. Bar charts are useful for comparing multiple values. Take for example comparing all countries by confirmed cases. Scatter charts are used to show distribution and relationships between two attributes. In my case, I show the distribution of confirmed and death cases among all the countries. Choropleth maps show distribution among countries of a value. They are good for showing geographical data such as COVID-19 data.

Overall having good structure leads to better decisions in relation to the data. It makes the data more understandable and having graphs interactive is an added benefit.