Big Data Visualisation Report - 16164571

Visualization report:

Visualizing data plays a big role in understanding data that has been collected. It helps to find relationships between data attributes as well as understand the structure of the data a lot better. These visualizations can also be used to capture an audience's attention and help to tell a story at a presentation or in an important meeting.

It is important to note that when you want to visualize data, it can help to stick to some fundamental principles to ensure that your Visualisation is clear, effective, and informative. These principles include 1) Simplifying 2) Understand Magnitudes 3)Use of Color and 4) Use of Structure

1) It is important to keep the data that has been visualized in a bite-sized format that is easy to take in for the viewer. Even though the data shown is complex this still doesn't mean that the visualization can be **simplified**. You can remove items from the graph which don't help to explain the visualization such as grids. Removing any unnecessary bars around the edges of the graph that don't provide useful information.

   Based on human perception, when we limit the number of objects displayed at once the easier and quicker the visualization is to understand. It can be better to make 2 or 3 graphs instead of merging them all into one.

   Avoid using 3D diagrams, shadows, and animations unless it helps to add value to your explanation.

   Graph1): In order to simplify this graph I removed any minor tick labeling on the x and y-axis using methods like fig.y-axis.minor_tick_line_color = None . I also provided a legend to help the user better understand what each data piece meant.
   I removed any frame around the graph with fig.outline_line_color = None

   Graph 2): To help simply this graph users can select or deselect what they would like to see in order to compare certain total cases over time. I also removed any frame and grid lines as they did not provide much value. I also didn't want to overload the graph so I provided just 5 data pieces

   Graph 3): For this graph I took out a lot of unnecessary visual data. I kept the metrics to a minimum only showing two figures of the total deaths,25000 and 75000 deaths. I also removed the x-axis labels as the graphs tell enough without them. It is easy to compare each graph as they are all lined up next to each other. This looks neat and tidy compared to stacking multiple plots on one graph. I also removed visual noise such as grids and graphs borders

Graph 4): I removed all grid lines and other visual noise such as the grid frame. I also adjusted the y-axis tickers to only show major tickers 0, 500000, 1000000,1500000,2000000.
I also adjusted the names of the countries at the bottom from horizontal, where they were overlapping and hard to read, to slanted and well-formatted for easy viewing.

2) In order to understand **magnitudes**, we can use Weber's Law to help us. The understanding of magnitudes helps us to make sure that for example, the two colors we pick for something have a noticeable, clear, visual difference. The change in magnitude can have a large effect on how we perceive the information. Humans find it hard to visualize an area, so it is important that we understand what type of representation is easier for humans to understand and also harder to understand. Color hue-saturation-density is the least accurate way of representing data while the length and different shapes are the most accurate means of displaying differences in magnitudes.

   Graph 1): In this graph I allowed the data to fill the whole graph so that the data and its magnitude can show the full impact of the increase and decreases in cases over the lockdown period. It is clear the higher the data point on the map, the higher the the number of confirmed cases.

   Graph 2): As there is such a variance in the max number of confirmed cases from the top country to the bottom country, it helps us understand how many more cases there are in other countries compared to Ireland. There is not a lot of change in the magnitude of cases in Ireland, making it almost seem like there are no cases where there are actually still 75,000 cases in Ireland. This graph's variety in magnitude change shows how big the virus is in some countries compared to others.

   Graph 3): The change in the magnitude of these individual graphs who times when the the virus was under control and times when the virus is rapidly growing and killing people.
   It also shows that a lot of countries have managed to keep cases under control until the last quarter of the virus period where deaths are now rising in 7 out of 8 of the graphed countries.

   Graph 4): It is clear that the magnitude helps the users to perceive the information clearly in this graph. 3 countries have a very high bar while the rest seem almost irrelevant compared to them. The visual data of this graph is represented by the shape and not the color.

3) **Color** is an important feature of visualization. It helps to make a clear distinction between different data types, differences in the data, and is visually appealing for the viewer to look at and examine. It can be important to use a color pallet for your data. Color can be used on certain data types such as Sequential, Diverging, and Categorical data.
   Sequential coloring can be used when data is from high to low e.g 0 - 100, diverging Colors can be used when you have data that meets at a midpoint e.g temperature and categorical colors allows for the use of completely different colors.

   Graph 1): In this graph, I used color to help the lines and circle plots stand out from Eachother. This helps the user to see the average cases but shows how the average is affected by an influx in the daily cases.
   For this graph, I also colored in the lockdown period to help define this time.

   Graph 2): To help define each country I selected a different color for each and kept the background as plain white to keep the graph simple

   Graph 3): For these graphs, I kept the color scheme simple as I felt it best suited these graphs. I added a fill to under the original red line in a slightly lighter red to help the data to stand out and have more of a visual impact.

   Graph 4): In the graph, I used a color palette with 20 colors to allow for each of my bars in the graph that had the most visual effect to stand out. Each color is very different from the color either side and it is clear which countries have the highest statistics

4) Lastly, the **structure** impacts the viewers' visual perception of the data. It is important to properly structure the graphs as the grouping and connection between these data pieces can give a perception about the data and it is important for these to be done correctly, to ensure the viewer properly understands what they are looking at easily and quickly. e.g Items which may be grouped together might give a natural impression that they belong together even though they may not be telling the same thing

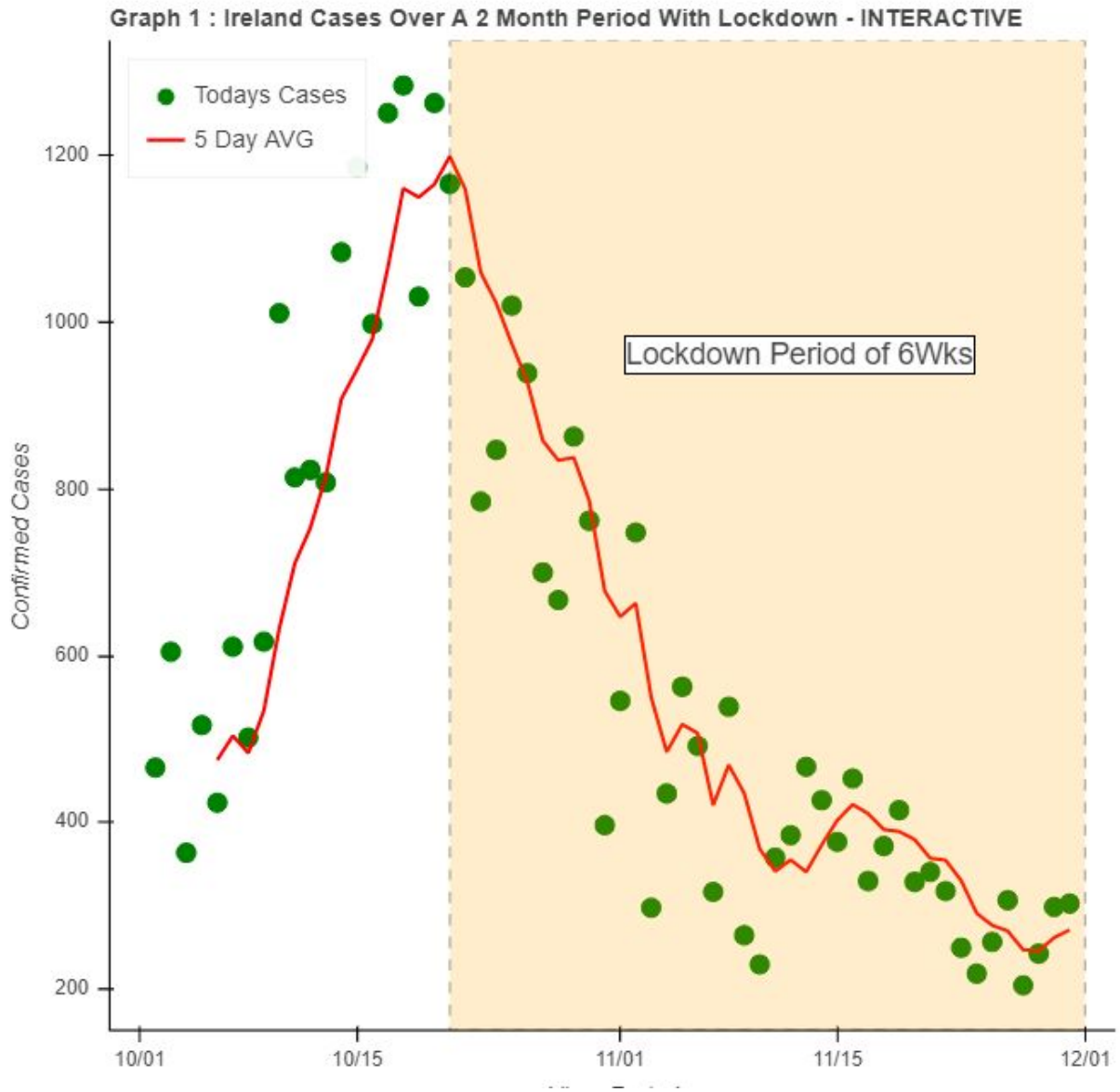
   Graph 1): This graph shows structure by having a continuous line of the same data such as the 5 days moving average. It gives the perception that all this data is telling the same thing. Today's cases points also are all of the same circle structure and color implying they are all belonging to the same data category.

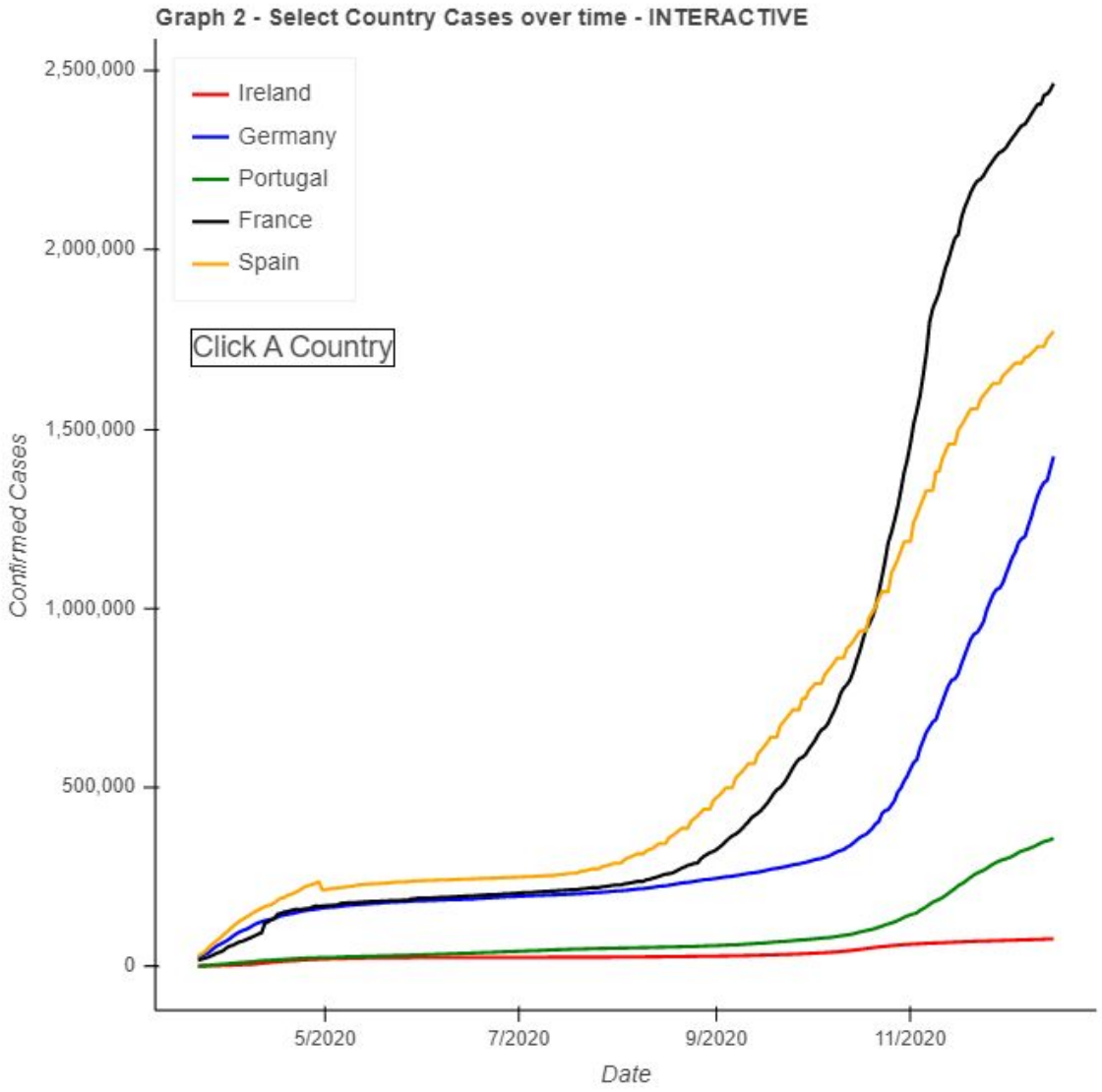
   Graph 2): Each line in this data is continuous and even though they cross over at certain points it is still clear what country the data is representing.

Graph 3): For these graphs, I structured the graphs scaled from 0-75,000 at the top and the other graphs from 0-25,000 underneath. We can also see symmetry in this graph.

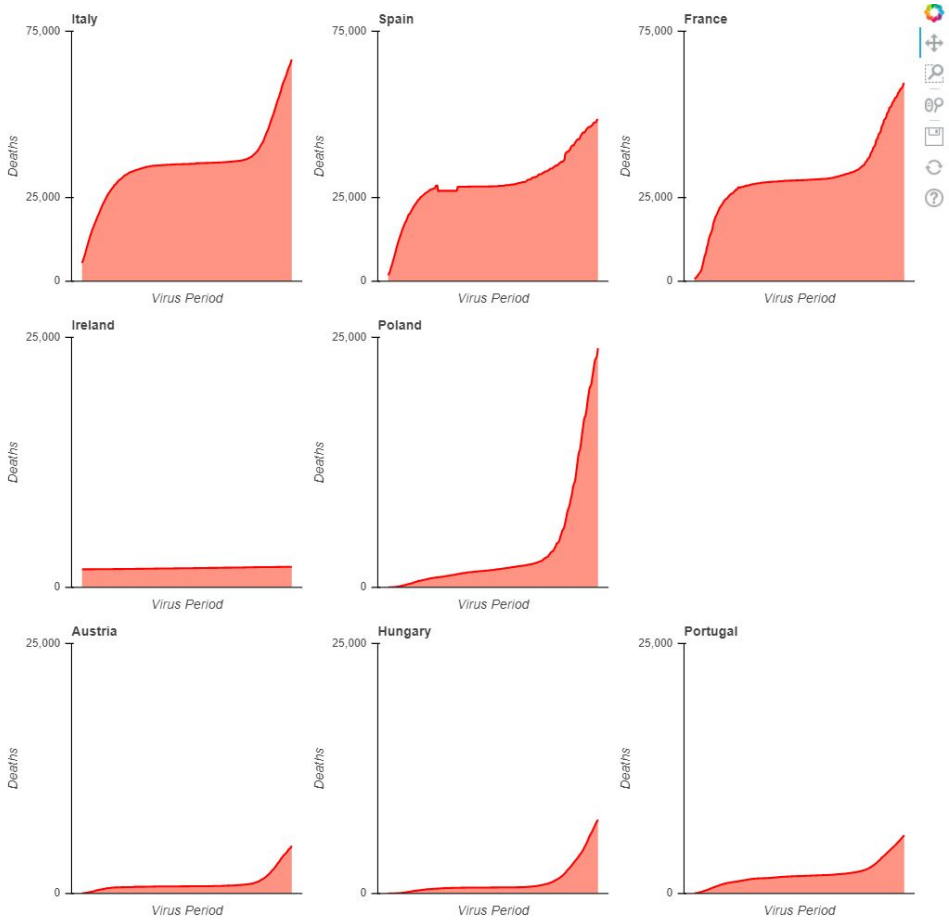
Graph 4): Lastly the bars in this graph are all in close proximity to each other which helps show that they are the same data type. If they were spaced far apart from each other at random intervals it might give the impression that they are of different data.

My Graphs

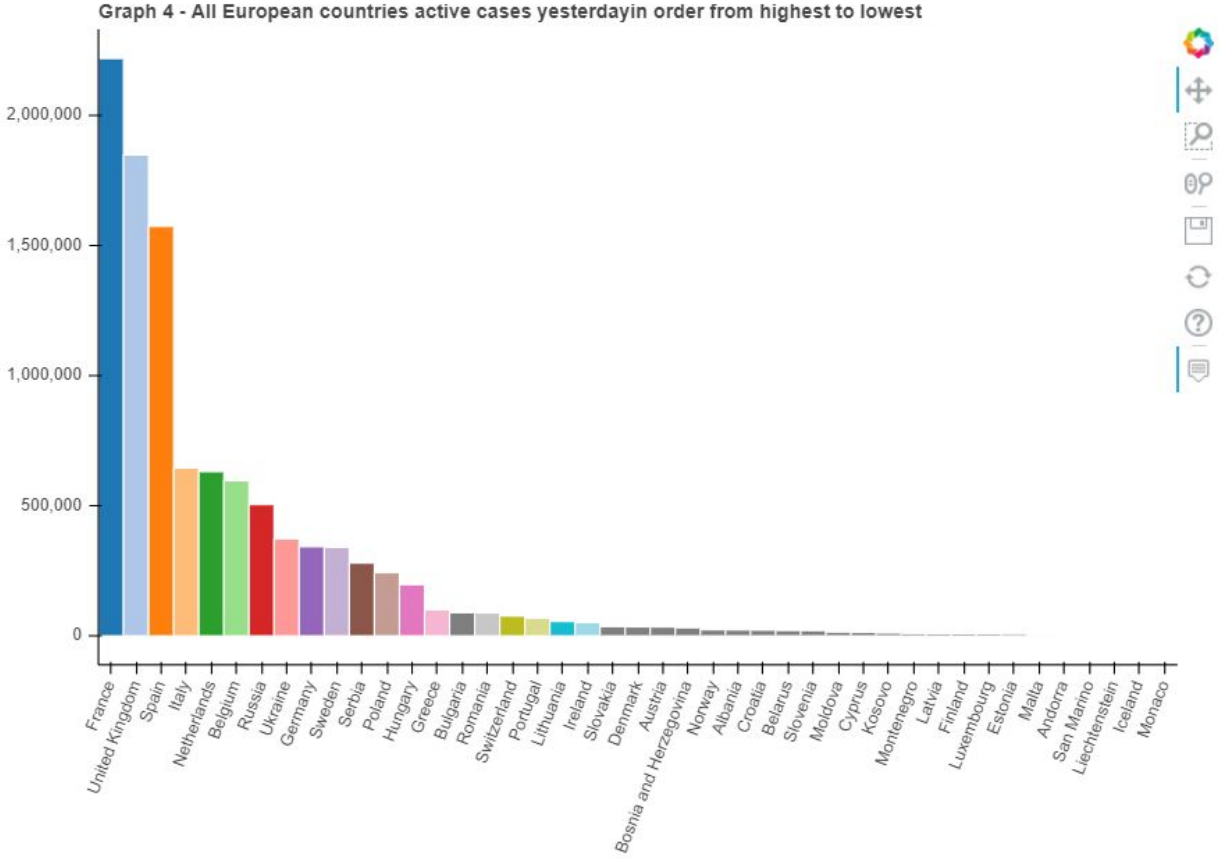

Graph 1 : Ireland Cases Over A 2 Month Period With Lockdown - INTERACTIVE
Todays Cases
5 Day AVG
Lockdown Period of 6Wks
Confirmed Cases
1200
1000
800
600
400
200
10/01
10/15
11/01
11/15
12/01


Graph 2 - Select Country Cases over time - INTERACTIVE
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
Confirmed Cases
Ireland
Germany
Portugal
France
Spain
Click A Country
5/2020
7/2020
9/2020
11/2020
Date

Graph 3 - Death Cases and Spike detection in several European countries



Graph 4 - All European countries active cases yesterdayin order from highest to lowest



Countries grouped by Active cases for 12-16-2020