

Postgraduate Certificate in Software Design with Artificial Intelligence

Data Visualisation - (AL_KSAIG_9_1)

Final Assignment - Covid-19 Visualisation

Student ID: A00267948

Student Name: Daniel Foth

Git Repository:

https://github.com/DanielsHappyWorks/data-vis-fianl-covid-19-sk

Visualisation Dashboard:

https://danielshappyworks.github.io/data-vis-fianl-covid-19-sk/

Contents

Declaration	4
Introduction	5
Data	6
Datasets	6
The Covid-19 Data Set	6
The Country Data Set	6
The Country and Continent Codes List	6
Pre-Processing	7
Software	8
Technology Stack	8
Source Control	8
Deployment	8
Customisation	8
Temporal Analysis	10
Statistics	10
Global	10
South Korea and The Countries Surrounding It	10
South Korea	11
Cumulative Case Visualisation	11
Global Cases	11
South Korea and The Countries Surrounding It Cases	12
South Korea Cases	12
Analysis of Cumulative Cases	12
Cumulative Death Visualisations	13
Global Deaths	13
South Korea and The Countries Surrounding It Deaths	13
South Korea Deaths	14
Analysis of Cumulative Deaths	14
Daily Case Visualisations	14
Global Cases	15
South Korea and The Countries Surrounding It Cases	15
South Korea Cases	16
Analysis of Daily Cases	16

Daily Death Visualisations	16
Global Deaths	16
South Korea and The Countries Surrounding It Deaths	17
South Korea Deaths	17
Analysis of Daily Deaths Cases	18
Analysis of all visualisations	18
Geospatial Analysis	19
Maps	19
Global Cumulative Cases	19
Global Cumulative Deaths	19
Global Recent Cases	19
Global Recent Deaths	20
Global Map Analysis	20
South Korea and The Countries Surrounding It Cumulative Cases.	20
South Korea and The Countries Surrounding It Cumulative Deaths	21
South Korea and The Countries Surrounding It Recent Deaths	21
South Korea and The Countries Surrounding It Recent Deaths	21
South Korea and The Countries Surrounding It Map Analysis	21
South Korea	22
South Korea Map Analysis	22
Analysis of Map Visualisations	22
Continent Visualisations	24
Analysis of Continent Visualisations	24
Country Visualisations	25
Analysis of Country Visualisations	27
Analysis of Continent and Countries with Biggest Impact Visualisation	ns27
Conclusions	28
References	29
Appendix	30
Link to get global time-based diagrams:	30

Declaration

I declare that this report was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or processional qualification except as specified.

Introduction

Coronavirus, better known as Covid-19, is spread in sneeze or cough droplets. (1) Covid-19 can survive for up to 72 hours on plastic and stainless steel, less than 4 hours on copper and less than 24 hours on cardboard but household disinfectants will kill the virus on surfaces (1) where it's present.

It can take up to 14 days for symptoms of Covid-19 to appear. They can be like the symptoms of cold and flu. (2) The most common symptoms include a fever, cough and shortness of breath. The consensus around the world is that if you have any of these symptoms, you should behave as if you have the virus and self-isolate for 14 days (2) and contact a doctor as needed.

As part of this report a custom dashboard has been created to visualise the Covid-19 pandemic. The dashboard uses the web stack to create visualisations based on time a geographic location data sourced from ECDC, Google and Datahub.

This report will look at the global effects of the pandemic, zoning in specifically on South Korea and its general location. Throughout the report we will see how much South Korea was affected and what stage of the pandemic the country is at. We will also discuss the future to come.

Data

Three different data sets are used for the analysis of Covid-19. These datasets have been refined and used to create visualisations. All the datasets were sourced on the 13th of April 2020.

Datasets

The Covid-19 Data Set

The following dataset has been sourced from ECDC. (3)

It contains global information about the Covid-19 Pandemic ranging from 31st December 2019 to the 13th of April 2020.

Data of Interest:

- 1. Date (Day, Month, Year) Can be used to plot cases and deaths against time.
- 2. Cases Can be used to get total, daily and cumulative cases.
- 3. Deaths Can be used to get total, daily and cumulative deaths.
- 4. CountriesAndTerritories Can be used as the locations display name.
- 5. geold Can be used to map locations geographically.
- 6. countryterritoryCode Can be used to map locations geographically.

The Country Data Set

The following dataset has been sourced from Google. (4)

This dataset contains latitude and longitude data that allows us to accurately plot Covid-19 information on a map.

Data of Interest:

- 1. Country Can be used to join this dataset with the covid-19 dataset by geold.
- 2. Latitude Can be used to plot data on a map.
- 3. Longitude Can be used to plot data on a map.

The Country and Continent Codes List

The following dataset has been sourced from JohnSnowLabs on Datahub. (5)

This dataset contains information that allows us to map continent to country data which allows for plotting of Covid-19 information based on continents.

Data of Interest:

- Continent_Name Can be used as the display name for data grouped by continents.
- Continent_Code Can be used to visualise data by continent.
- Two_Letter_Country_Code Can be used to join this dataset with the covid-19 dataset by geold.
- 4. Three_Letter_Country_code Can be used to join this dataset with the covid-19 dataset by countryterritoryCode.

Pre-Processing

Missing fields in the data such as country codes and latitude and longitude data meant that some pre-processing needed to be done. Any missing field was manually added to the dataset by use of Google Maps.

For some fields like the population data for 2018 in the Covid-19 dataset (3) too many fields were missing. This meant that it wasn't worth trying to fill in the gaps to make visualisations with those fields.

Since the bespoke software developed for the purposes of visualising the data is written in JavaScript, the data had to be processed into Json objects. This was done for ease of importing and processing of the data in the specific environment. Three Json files can be found in the ./src/_data directory in the project structure.

Software

All the visualisations used in this report are created using custom software which employs graphing frameworks and allows for creation of an interactive dashboard. The code is generic enough to use any country's data to display visualisations if it's present in the dataset.

The Dashboard can be accessed via the following link: https://danielshappyworks.github.io/data-vis-fianl-covid-19-sk/.

It's highly recommended to look at the visualisations on the dashboard as they have useful tooltips that help with readability of the exact figures present.

Technology Stack

The web development stack, leveraging a set of useful libraries, was used to develop the dashboard. The stack consists of HTML, JavaScript and CSS.

The used libraries include:

- ReactJs (6) React leverages components to allow encapsulation of code. This
 makes it easy to create single page user interfaces over multiple files with proper
 separation and templating.
- 2. Materialize (7) A Bootstrap like library, that comes with pre-built HTML and CSS components that allow for quick and responsive web design.
- 3. Chart.js (8) This library provides a way to create a multitude of interactive graphs with tooltips and legends.
- 4. Leaflet (9) Leaflet provides interactive maps that can be customised with markers to display data geographically.

Source Control

GitHub has been used as the source control for this project.

The code is available here: https://github.com/DanielsHappyWorks/data-vis-fianl-covid-19-sk

Deployment

The dashboard has been deployed on GitHub pages. GitHub pages were used as it's easy to use for single page applications that run on the web stack. It only takes one command "npm run deploy" to deploy any changes after the package.json file is configured correctly.

Customisation

It's possible to pass in parameters via to URL to customise the content. The parameters are:

- 1. live=true by default it displays cached data from the 13th of April 2020
- 2. country=Ireland,United_Kingdom by default the country data is South Korea

Opening the dashboard with the live=true parameter will pull the latest Covid-19 data from ECDC. This takes a bit of time as the data is pulled live from the ECDC servers.

Opening the dashboard with the country= Ireland, United_Kingdom will use the specified countries in the graphs Country Data section of the Dashboard. Countries need to be "," separated and use "_" instead of spaces to work properly.

Sample customisation:

https://danielshappyworks.github.io/data-vis-fianl-covid-19-sk/?live=true&country=Ireland,United_Kingdom

Temporal Analysis

In this section, we will look at the status of the pandemic on 13th of April 2020 and how it evolved over time for South Korea, the countries surrounding it and the world with the data leading up to the 13th of April 2020.

Note all the graphs are inside the visualisations.zip in the Temporal folder at their normal size. The images in the document are scaled down for readability purposes.

Statistics

Global

Total Cases		Recent Cases 72,395
Total Deaths	Recent Deaths	Mortality Rate

Globally we can see that the pandemic is still very active with 72,395 new cases and 5,321 new deaths recorded for all countries on the 13th of April. Out of the total 1,1806,612 cases and 113,506 deaths, we have a mortality rate of about 6.2% for every person that has been affected.

South Korea and The Countries Surrounding It Total Cases 101,389 Recent Cases 647 Total Deaths Recent Deaths 3,670 Recent Deaths 9

The countries that are included in this data are South Korea, Japan, China and Taiwan. They are the only countries surrounding South Korea that have provided us with data. It's hard to tell how accurate this is due to the belief that some countries aren't disclosing all information available to them.

Overall, the area looks like it's handling the pandemic well as of the 13th of April. The overall recent cases and deaths are 647 and 9 which is only 0.89% and 0.16% of the global cases and deaths. Population data would be very useful here as we would be able to more accurately compare world data versus the areas data.

South Korea



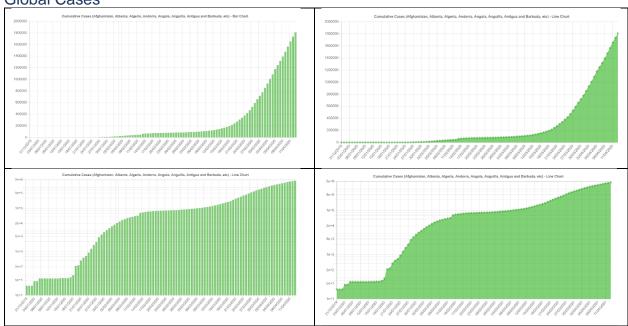
South Korea makes up a 10th of the cases from itself and the surrounding areas. The estimated population of South Korea is 51,269,185 in 2020 (10). This means that only 0.02% of the total population has been infected and that only 0.00042% have died. When compared to the population most of the figures for South Korea become a lot less significant.

We can also see that the deaths and cases on April 13th are low but from data like this it's hard to tell if this is because the country is still in the early stages of the pandemic or if they have levelled off already.

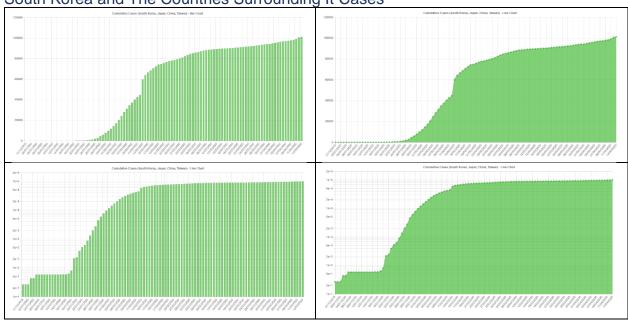
Cumulative Case Visualisation

The graphs below are analysed together in the "Analysis of Cumulative Cases" section below.

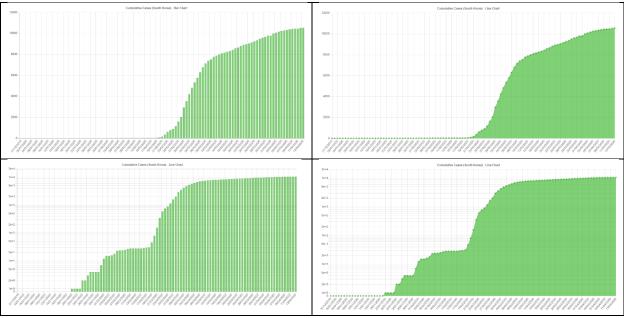




South Korea and The Countries Surrounding It Cases



South Korea Cases



Analysis of Cumulative Cases

Globally we can see that cases started being recognised and dated as far back as the 31st of December 2019. It wasn't until the 21st of January that the amount of cases started to rapidly show up. This rapid rise is only noticeable on the logarithmic graphs. An educated guess would mean that from the 21st, more people were being tested due to improvements in the procedures and wider spread of the pandemic.

From the Global data we can also tell that the pandemic hasn't reached a peak around the world yet. The linear graph is increasing at a rapid rate.

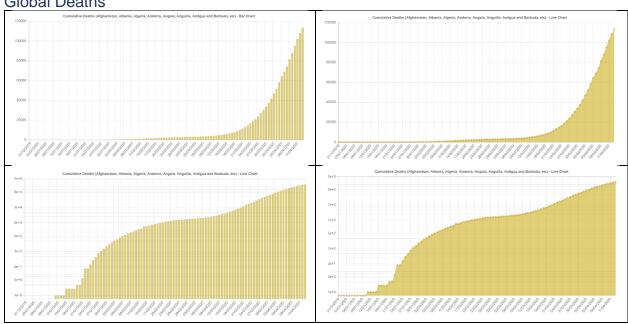
In the area of South Korea, we see very similar behaviour to the global graphs at the beginning but then the new cases start to flatten out. This is probably due to the general area being close to the source of the pandemic.

Lastly, South Korea itself exhibits the same behaviours as the countries it's close to. The only difference is that the event seems to be staggered by a few days. It seems that the pandemic didn't properly reached the country until the 20th of January.

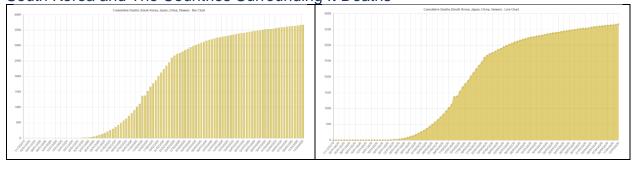
Cumulative Death Visualisations

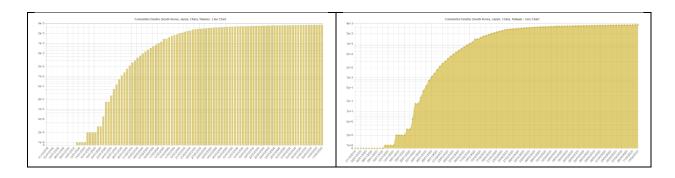
The graphs below are analysed together in the "Analysis of Cumulative Deaths" section below.

Global Deaths

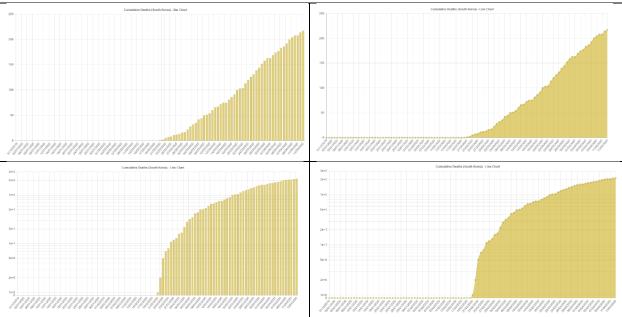








South Korea Deaths



Analysis of Cumulative Deaths

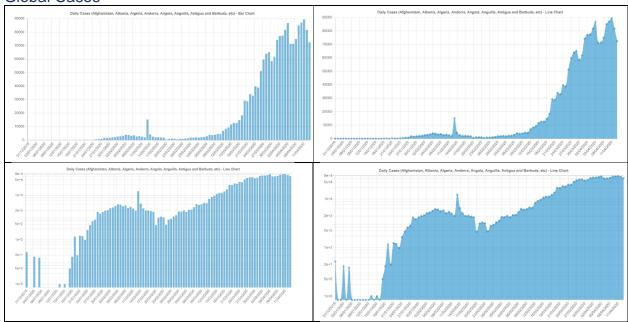
Comparing the deaths to cases, the curves are quite similar. They are just staggered and are much smaller in terms of figures. South Korea's death rate hasn't flattened quite yet. This is most likely because it can take up to two weeks for Covid-19 to start taking affect on the body. The prediction is that the value will flatten soon.

Comparing the death to case rate we can see that the mortality rate isn't very high but seems to be somewhat relative to the amount of cases.

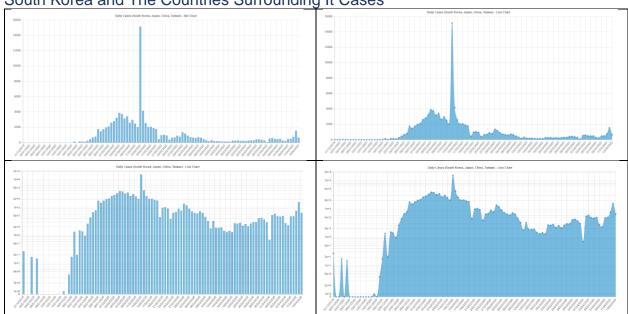
Daily Case Visualisations

The graphs below are analysed together in the "Analysis of Daily Cases" section below.

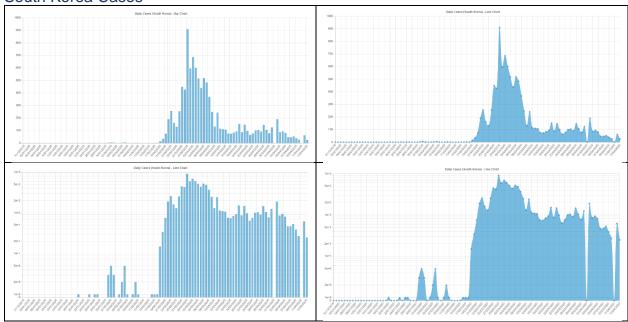
Global Cases







South Korea Cases



Analysis of Daily Cases

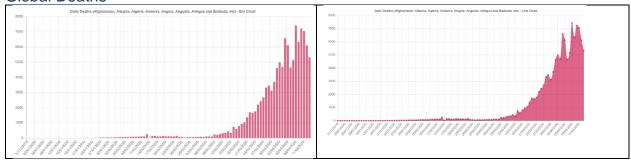
These graphs show us the exact peaks over time by displaying the amount of discovered cases that day. From the global data we can see the general peak that happens in South Korea and the countries around it in the middle of the graph. At that point, the peak is small, but it can still be distinguished on the linear graph. The next peak, from the 12th of March to 13th of April, causes the graph to skyrocket which implies that the disease started spreading around the whole world very quickly. We can identify the peaks between all graphs just by looking at some of the very high figures without needing the dates, the most prominent one occurring on the 13th of February.

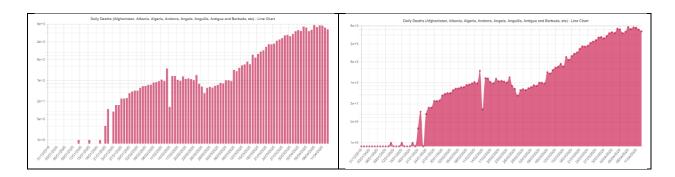
From this graph we can definitely see that the spread of Covid-19 in South Korea has slowed down as the figures drop off very quickly after the biggest peak.

Daily Death Visualisations

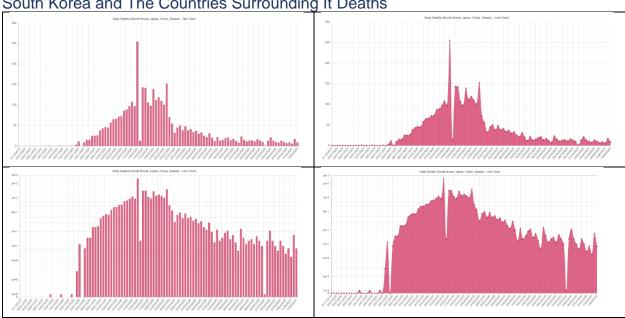
The graphs below are analysed together in the "Analysis of Daily Deaths" section below.



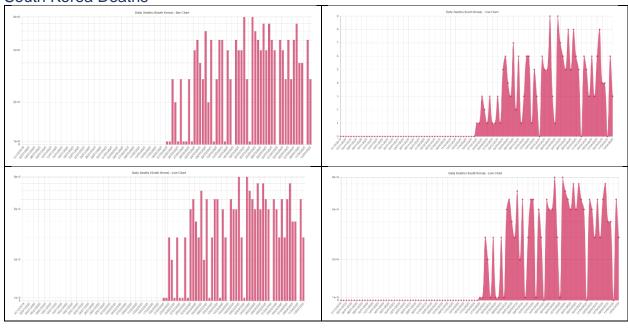




South Korea and The Countries Surrounding It Deaths



South Korea Deaths



Analysis of Daily Deaths Cases

Once again, as with the cumulative data, the daily deaths are staggered by a few days from the cases, allowing us to make similar assumptions to the ones above.

Analysis of all visualisations

All the visualisations come in the form of two bar and line graphs, a linear representation of the data and a logarithmic one. The line graphs allow us to see the growth from one day to another by looking at the incline but it's easier to get exact values from bar charts. Both are very good for visualising data over time.

The linear representation gives us an accurate growth over time for cumulative values and an accurate depiction of the actual values every day in the daily visualisations. These are easy to read and understand. By looking at these graphs we can pull figures out quickly and see different behaviours in data.

The "logarithmic scale shows the rate of change of the number" (11). Logarithmic graphs are harder to read but they can show us a lot about data that might not be easily read on a linear graph since respond to skewness towards large values. In our case it sheds a lot of light on the cumulative graphs as they tend to grow into very large numbers over time.

Colour coding the graphs also allows us to quickly identify what kind of data we are looking at. This lets us sort visualisations by the nature of the data they they display without having to read text.

Geospatial Analysis

In this section we will look at how the pandemic is impacting the world geographically on the 13th of April 2020 for South Korea, the countries surrounding it and the world.

Note all the graphs are inside the visualisations.zip in the Geospatial folder at their normal size. The images in the document are scaled down for readability purposes.

Maps





Global Cumulative Deaths



Global Recent Cases



Global Recent Deaths



Global Map Analysis

By looking at the plotted maps, we can see that in all cases the EU and US have been hit the most by the pandemic. The two seem to be switching positions between graphs. When it comes to the most recent cases, US seems to be larger while Cumulative death data implies that Europe has had a lot more fatalities so far. The larger circles in these areas especially in the recent cases implies that the pandemic is still growing a lot in the EU and US.

There are some other large circles in areas like China or Brazil but in general they are not as big as the EN or US. This could mean that those countries are handling the pandemic better or are yet to experience growth as it's still in the early stages. Going by the general lack of reported deaths in Africa and Oceania, it's probably still in the early stages of the pandemic at this point.

South Korea and The Countries Surrounding It Cumulative Cases



South Korea and The Countries Surrounding It Cumulative Deaths



South Korea and The Countries Surrounding It Recent Cases



South Korea and The Countries Surrounding It Recent Deaths

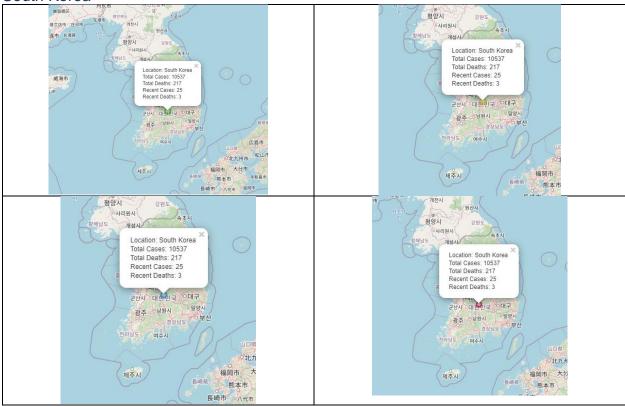


South Korea and The Countries Surrounding It Map Analysis

Not many countries surround South Korea. These include China, Japan, Taiwan and North Korea. South Korea is cut off most of them by water as the only mainland its connected to is North Korea. What makes tracking how Covid-19 made it there is even harder when we are provided with no data by North Korea.

With its general proximity to China which got many cases of Covid-19 in the early days of the pandemic, it's very likely that the first cases in South Korea originated from China due to their geographic proximity.

South Korea



South Korea Map Analysis

South Korea is surrounded mainly by water. This puts it at an advantage when it comes to containing a pandemic. Unlike the EU or the US where people can easily travel, in South Korea, your only choices are to go via Ship or Airplane. This is further endorsed by the fact that the cases in the country are limited and in the Temporal section above we can see that the curve is already starting to level of for the country.

Analysis of Map Visualisations

All the circle markers on the map are relative in size to each other depending on how big the number total number is. This is useful when trying to identity countries with a very large impact. This could be further improved if all circles scaled only based on the ones you see when you zoom in as the countries with the biggest impact can easily make small differences unnoticeable. Africa is a perfect example of this as it has many areas with different amounts of cases but when put up against the US all the countries in Africa look like they have very similar figures.

The tooltip is very helpful as it lets us easily see the exact figures for each country just by clicking on the marker on the dashboard.

Colour coding the maps with the same colours as the temporal visualisations makes the data more easily identifiable and signals to the user that they are looking at different data. Adding opacity lets us better visualise how much compact areas are affected as it results in a darker colour signifying higher numbers in the area. This is especially visible in Europe for the "Global Cumulative Deaths" Map

Continent Visualisations



Analysis of Continent Visualisations

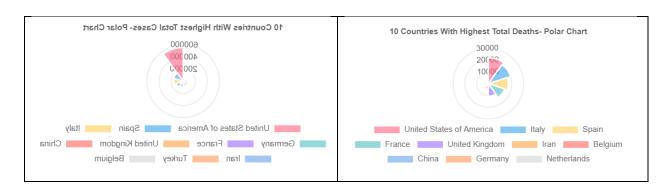
Building on top of what we gathered from the world map, these visualisations further prove that Europe and South America are currently struggling the most with Covid-19. We can see the breakdown in terms of total and recent cases and deaths. South America has listed more new cases on April 13 than Europe, but overall Europe still has the most cumulative cases deaths and daily deaths. This could change in the next few days as it seems like the pandemic is starting to spread more in South America.

South America, Oceania and Africa's cases and deaths are negligible compared to all other continents and on certain area-based representations (Pie, Polar and Doughnut Charts) they are unnoticeable.

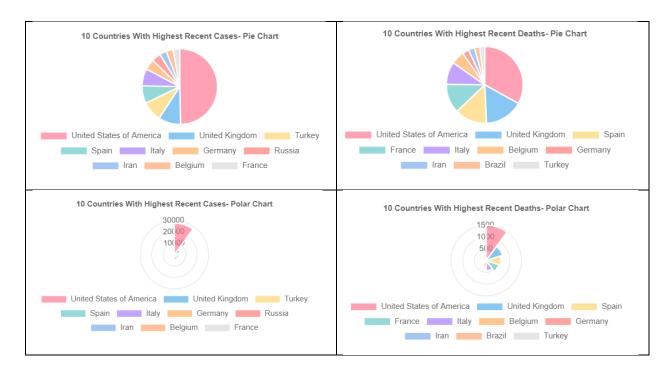
Lastly, we have Asia, which initially had the worst case of the pandemic as it has originated from there, taking the middle ground in all areas of the pandemic. At this point the area has generally flattened in terms of the cases due to restrictions and awareness which led to other countries with faster growth to overtake the continent.

Country Visualisations









Analysis of Country Visualisations

These charts have been limited to the 10 countries with the highest cases and deaths as displaying any more made the graphs hard to display on the dashboard.

All these graphs are dominated by the United States of America which matches its depiction on the map. It also makes sense due to the country's size compared to most others around earth. A lot of the other countries that show up generally belong to Europe, with some other appearances like Turkey, Brazil and Iran.

Analysis of Continent and Countries with Biggest Impact Visualisations

These visualisations are very distinct with multiple colours to identify all the data. The colours may be an issue for the colour blind as some of them are very similar. As an improvement it could be a good consideration to add a panel to the dashboard to alter colour schemes on demand to help with this. Also using the same colour for the country as in the continent could have made the visualisation even more readable as it would have nicely linked the graphs together.

The tooltips and ability to isolate data out by clicking on the legend makes the visualisations more interactive and allows for some level of configuration.

In general, these graphs do a great job at showing us how geographic locations compare to each other in terms of the pandemic but 5 graphs per dataset is overkill. One bar and pie chart would probably have been enough but having multiple versions of similar graphs can be useful and it can help change perspective on the data ever so slightly.

Conclusions

From the cumulative temporal visualisations, we can see that the Covid-19 pandemic is still going strong and could do so for the foreseeable future as it spreads to all the continents that haven't been as affected. The Asian continent and especially South Korea are already reaching a point where the curves are flattening, meaning that restrictions in some areas around the globe might start easing soon.

Even though countries like South Korea have seen a decline in new cases there is still a chance of a second wave for the pandemic if restrictions are lifted too soon. Therefore, it is necessary to keep social distancing and isolating for as long as possible.

In terms of the visualisations, there are some glaring issues that could be fixed mentioned in their respective Analysis sections above. Even though these issues have been identified, the charts, maps and graphs are very useful and can keep us informed on the current state of the pandemic very well. When evaluating and analysing this data its best to use multiple different graphs with both temporal and geographical data as it gives us a lot more information to work with. Therefore, the dashboard created for this project makes data analysis more effective as all of the charts, maps and graphs are in one place.

References

- 1. HSE. How coronavirus is spread. [Online].; 2020 [cited 2020 May 06. Available from: https://www2.hse.ie/conditions/coronavirus/how-coronavirus-is-spread.html.
- 2. HSE. Symptoms of coronavirus. [Online].; 2020 [cited 2020 May 06. Available from: https://www2.hse.ie/conditions/coronavirus/symptoms.html.
- 3. ECDC. COVID-19 situation update worldwide. [Online].; 2020 [cited 2020 May 06. Available from: ecdc.europa.eu.
- 4. Google. countries.csv. [Online]. [cited 2020 May 06. Available from: https://developers.google.com/public-data/docs/canonical/countries_csv.
- 5. JohnSnowLabs. Country and Continent Codes List. [Online]. [cited 2020 May 06. Available from: https://datahub.io/JohnSnowLabs/country-and-continent-codes-list#data-cli.
- 6. Facebook. React A JavaScript library for building user interfaces. [Online].; 2020 [cited 2020 May 06. Available from: https://reactjs.org/.
- 7. Materialize. Materialize A modern responsive front-end framework based on Material Design. [Online].; 2020 [cited 2020 May 06. Available from: https://materializecss.com/.
- 8. Chart.js. Chart.js Simple yet flexible JavaScript charting for designers & developers. [Online]. [cited 2020 May 06. Available from: https://www.chartjs.org/.
- 9. Leaflet. Leaflet an open-source JavaScript library. [Online].; 2019 [cited 2020 May 06. Available from: https://leafletjs.com/.
- 10. worldometers. South Korea Population. [Online].; 2020 [cited 2020 May 07. Available from: https://www.worldometers.info/world-population/south-korea-population/.
- Robbins N. When Should I Use Logarithmic Scales in My Charts and Graphs? [Online].; 2012 [cited 2020 May 07. Available from: https://www.forbes.com/sites/naomirobbins/2012/01/19/when-should-i-use-logarithmic-scales-in-my-charts-and-graphs/.

Appendix

Link to get global time-based diagrams:

https://danielshappyworks.github.io/data-vis-fianl-covid-19-

sk/?country=Afghanistan,Albania,Algeria,Andorra,Angola,Anguilla,Antigua and Barbud a, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Bonaire Saint Eustati us_and_Saba,Bosnia_and_Herzegovina,Botswana,Brazil,British_Virgin_Islands,Brunei_ Darussalam, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Cape Verde ,Cayman_Islands,Central_African_Republic,Chad,Chile,China,Colombia,Congo,Costa_ Rica, Cote divoire, Croatia, Cuba, Cura%C3%A7ao, Cyprus, Czechia, Democratic Republi c_of_the_Congo,Denmark,Djibouti,Dominica,Dominican_Republic,Ecuador,Egypt,El_Sa Ivador, Equatorial_Guinea, Eritrea, Estonia, Eswatini, Ethiopia, Falkland_Islands_(Malvinas) Faroe Islands, Fiji, Finland, France, French Polynesia, Gabon, Gambia, Georgia, Germany, Ghana, Gibraltar, Greece, Greenland, Grenada, Guam, Guatemala, Guernsey, Guinea, Guin ea_Bissau,Guyana,Haiti,Holy_See,Honduras,Hungary,Iceland,India,Indonesia,Iran,Iraq, Ireland, Isle_of_Man, Israel, Italy, Jamaica, Japan, Jersey, Jordan, Kazakhstan, Kenya, Kosov o, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Liberia, Libya, Liechtenstein, Lithuania, Luxemb ourg, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Mauritania, Mauritius, Mexico, Mo Idova, Monaco, Mongolia, Montenegro, Montserrat, Morocco, Mozambique, Myanmar, Namib ia, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, North Ma cedonia, Northern_Mariana_Islands, Norway, Oman, Pakistan, Palestine, Panama, Papua_ New_Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto_Rico, Qatar, Romania, Russia, Rwanda, Saint_Kitts_and_Nevis, Saint_Lucia, Saint_Vincent_and_the_Grenadine s,San_Marino,Sao_Tome_and_Principe,Saudi_Arabia,Senegal,Serbia,Seychelles,Sierr a Leone, Singapore, Sint Maarten, Slovakia, Slovenia, Somalia, South Africa, South Kore a, South_Sudan, Spain, Sri_Lanka, Sudan, Suriname, Sweden, Switzerland, Syria, Taiwan, T hailand, Timor Leste, Togo, Trinidad and Tobago, Tunisia, Turkey, Turks and Caicos isl ands, Uganda, Ukraine, United_Arab_Emirates, United_Kingdom, United_Republic_of_Tan zania, United_States_of_America, United_States_Virgin_Islands, Uruguay, Uzbekistan, Ve nezuela, Vietnam, Yemen, Zambia, Zimbabwe