**The report of information of vislization of Covid-19 in United Kingdom**

**Introduction**

The COVID-19 pandemic is a global pandemic which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The epidemic was first discovered in Wuhan in December 2019, and then quickly spread to many countries around the world in early 2020, gradually turning into a global pandemic. As of April 9, 2021, there have been more than 134 million reported cases of COVID-19 worldwide; more than 2.9 million people have died and more than 76.188 million have recovered. This pandemic is still influencing people daily life.

For control the pandemic of Covid19, understanding the virus by data analysis is very necessary, therefore data visualizations for Covid-19 provides great opportunities to understand how the virus spread out and exam is the government’s action useful or not. This will support the government or other health organization to improved their strategies.

In this report, the visualization is focusing on three parts. Firstly, the Covid-19 varied from time will be studied. Secondly, the Covid varied with different areas will be discussed and finally the government actions will be measured by visualization.

**Dataset description**

The dataset is sourced from GOV.UK1 (url: <https://coronavirus.data.gov.uk/>), which is developed and maintained by [Public Health England](https://www.gov.uk/government/organisations/public-health-england).

This data set include information as below:

The Cases by specimen date: The number of people who was test with least one positive result.

The number of deaths: Number of deaths of people who had had a positive test result.

The number of patients: Daily count of confirmed COVID-19 patients in hospital at 8am

The number of virus test: Number of confirmed positive, negative or void COVID-19 virus test results.

From this website2 it also provides cases number reported in different areas in UK.

There is no dataset for the government actions, but the GOV.UK also provide the information of that.

**Question 1- Covid-19 vs Time**

**How is the Covid-19 varied in the whole period?**

First of all, it is important to understand the general variation of the epidemic whole period. The variation of the number of cases is an important factor to show the how epidemic is going well or worse. The number of new cases by specimen cases is the most important factor to show how severe the situation is. In addition, United Kingdom government did not publish the cured cases, the daily new cases are the most directly factor to tell the situation of epidemic. Therefore, analysis of the new cases and other related data with timeseries is able to tell how the Covid-19 vary with time in the whole period. It can be also clear to deal some sub-question below:

*When is the epidemic most severe?*

**Visualization strategies**

For express the variation of the daily reported new cases and other factors, the line graph is suitable to show changes and trends of epidemic in UK. The dataset also provides quantitative values (the number of cases etc.) and date (continuous period of time), which is easily to used ggplot2 to achieve the vislization.

**Data processing and Visual encoding**

For achieving the data vislization, the type of data(variable) is required to be confirmed.

Data: Number of Cases (int), Date (Date)

But in the dataset, the number of cases character, and also the date also requires to be converted to date in R firstly and then they are ready to visualization. In the x-lab of vislization, it is unnecessary to show every single date due to there is too much data, therefore, the 2-month date break interval was set in R to make the x-lab to visualized more clearly. In R, scientific notation is set as default to express numbers that are too large, which was not except in this report, to deal with that, options (scipen = 200) were pre-set to pre-processing all number in this report to express as normal way.

Firstly, as mentioned above, the daily new cases reported is the most important factor to measure the general condition of Covid-19 epidemic, the variation with date is shown as below.

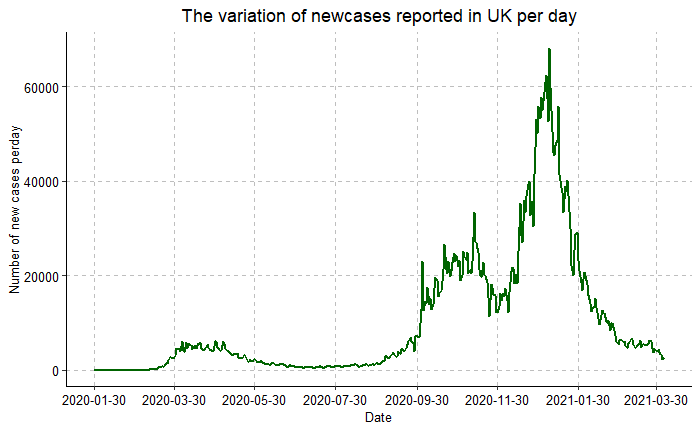


Fig1.The variation of new cases reported number

The epidemic can be split into two parts by there is 2 separated picks shown on the line graph. The first one is around April and the second one is a larger peak start from September 2020 and end with late March.

In the first period, the most new cases reported date is 6th May, with 6111 new cases reported. For the second period, the number of new cases reported is much larger than in the first period. It could be seen that the epidemic in UK increase very fast at September 2020, and December 2020 and January 2021. The date with most new cases reported is 08th Jan 2021 with 68053 cases reported. By measurement of the severe condition by the number of new cases daily, the most severe period for the Covid-19 epidemic is September 2020 - February 2021.

From the view of deaths number reported, the conclusion may be different. Similar with the line graph of new cases reported, the line of deaths reported also has two peaks at similar time with most cases reported point. There is point should be noticed that the value of deaths is similar that only with difference of hundreds, although the daily new cases differences is large with the second peak is nearly 10 times than the first one.

The similar of death number and the large differences of new cases number can tell us in the first period epidemic, the death rate is much higher than second period. This is possible to caused by lack of experiences for facing a s sudden epidemic of new virus never meet.

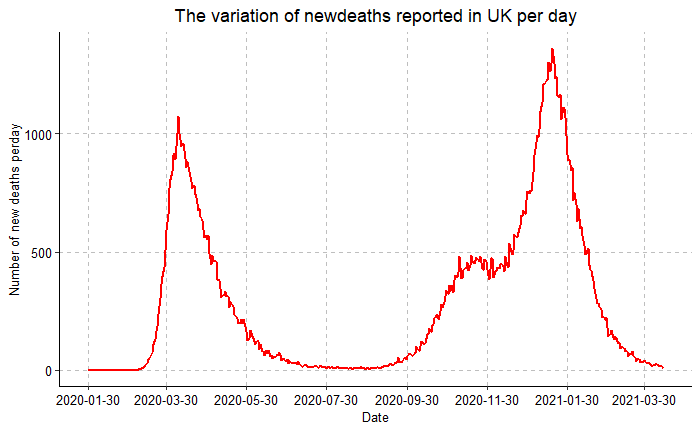
In the second period, although the new cases number is high, but the medical condition is 

Fig2. The variation of new deaths cases

better and the virus is not danger as the first period. So, for answer the question when is the epidemic most severe, the answer is the that April 2020 is the most dangerous part and the winter of 2020 and 2021 is the time the epidemic increased most fast.

The interactive 3D graph is also a possible solution to explain the trend of Covid -19 variation. Using R to build a new data frame combine both new cases and death cases with data, and the 3D scatter is able to build based on the new data frame. The vislization is shown below.

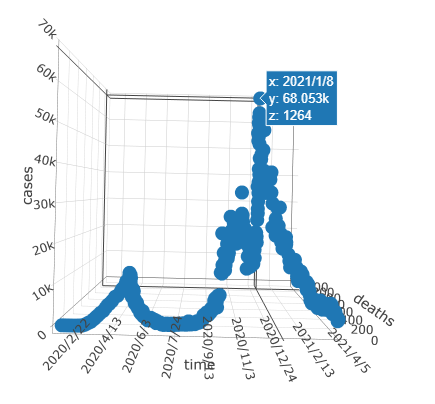
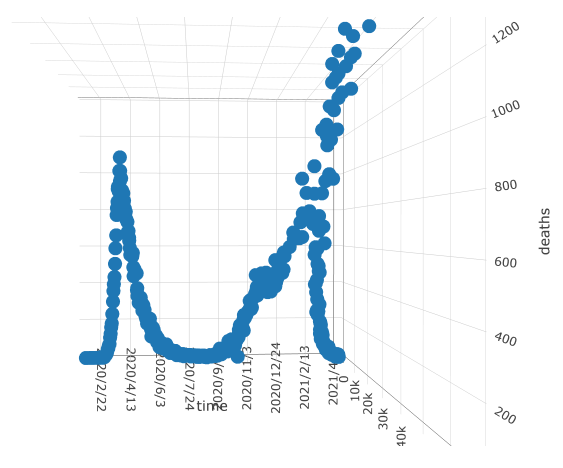


Fig3. The 3D model

This 3D plot is interactive and can be rotate and click the point to check detailed information, this plot also can be evidence that there is two severe period and the first one is more danger and the second one is more spreadable.

**Question 2 – Covid -19 vs Areas**

**How Covid-19 performances in different areas in UK?**

The Covid-19 epidemic in different area has different number of cases reported. There are some places control the epidemic well and the number of cases is at lower level of average. They can be good example to others areas which provide successful experiences to others. Therefore, analysis of the epidemic in different areas is essential to find the different condition of areas which is helpful to measure how the rules or actions performed by the governments in different places. Not only the number of cases, the geological and populational factors should also be considered in analysis. The answers of those sub-questions are also useful to understand the Covid-19 performances in different areas:

*Which areas are suffering most severe condition of epidemic?*

*Which areas are good at control of epidemic?*

**Visualization strategies**

For each different area has individual values, bar chart is effective to emphasizes individual values and also comparing individual values. Therefore, bar chart to show different values in different areas brings great expressiveness and effectiveness to understand the condition in different areas.

To enhanced the expressiveness, a bubble plot on map makes easier to show the epidemic on certain areas, because only the name of the areas in the bar chart is indirectly to tell people to know where the area is. By this method, the different numbers of cases will be encoded to the size of bubble which corresponds to the real location of the areas in map. The map with bubble is able to express the information of Covid-19 in the appealing and interpretive way.

**Data processing and** **Visual encoding**

Data (Types) in dataset: Areas (Nominal), Number of cases (Quantitative)

For vislization on map, it is required the geographic coordinate information, the information is from LatLong.net. And information was added into correspond areas in the csv file.

Firstly, the cases in different areas were shown as below.

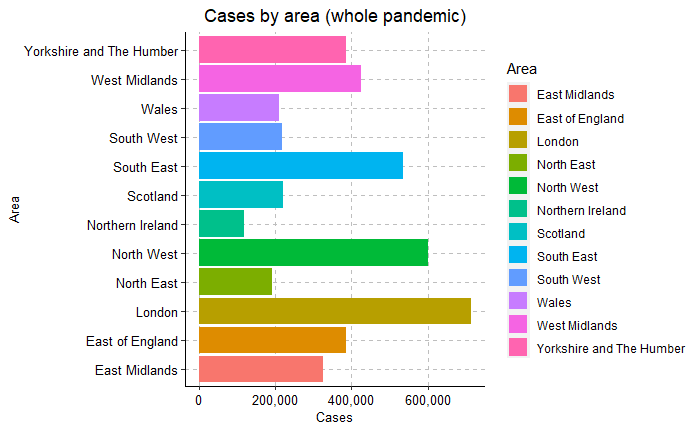


Fig4. The total cases in different areas in UK

By this plot, it is clear to see that London has the highest number of cases, and South East and North West also have higher level of cases. However, it is still not reasonable to conclude that those areas are preformed worse than other areas. Due to that the different area has different population, the cases per 100,000 population is a more important factor to measure the epidemic situation for different area. The vislization is shown as below.

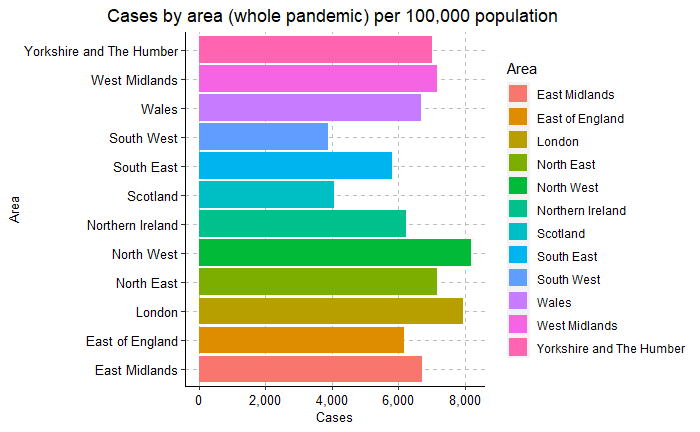
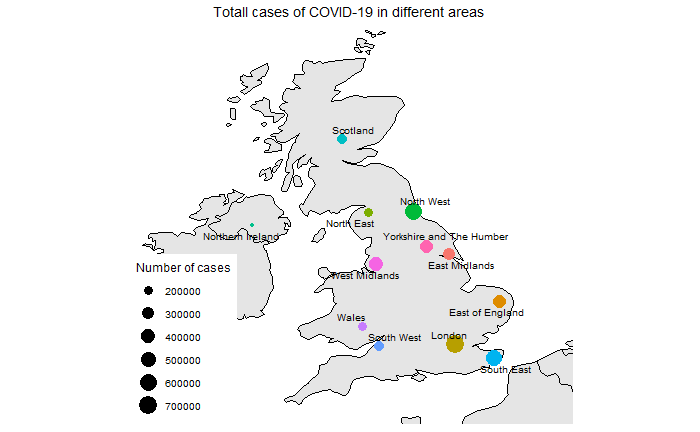


Fig5. The cases per 100,000 population in different areas in UK

By this graph, it is clear to see that the average cases per 100,000 population is closer than comparation of the total cases. It can be also to see that North West has the highest cases per 100,000 population, London and North West also have higher level of cases. This bar plot indicates that those areas should improve their action to control the epidemic. It also indicates that South West and Scotland clearly have lower-level cases per 100,000 population than others.

To enhance the expressiveness, the vislization of bubble plot on map is list below with both total cases and cases per 100,000 population. This graph is able to make people to directly understand of Covid-19 situation on map. As same for the bar chart, London, North West and South East are suffering more severe condition. It can be also conducted that Scotland and Northern Ireland have better control of Covid-19 than most of areas in England.



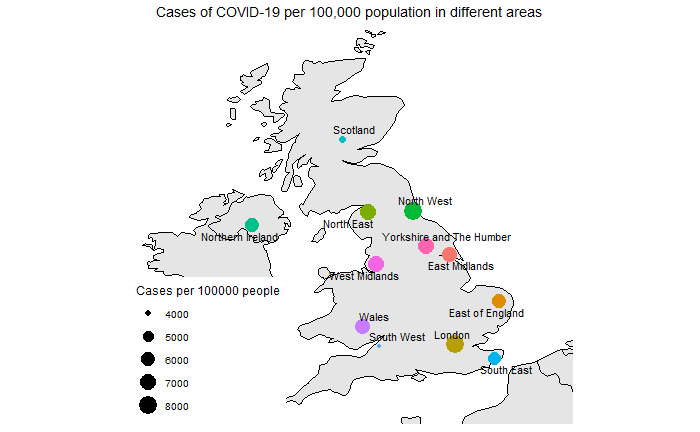


Fig6. The total cases/ cases per 100,000 population in different areas in UK

**Question 3 – Covid-19 vs Government Actions**

**How dose UK government affect the Covid-19?**

In order to control the spread out of Covid-19, the UK government have taken much efforts to control the epidemic. The UK have experienced three times national lockdown and publish some rules to keep the social distance in both indoor and outdoor activities. It is valuable to summarized to measure are those actions have positive impact on epidemic or not.

To analyzed the action, the report will focus on those sub-question:

*Are those national lock down all effective and find the reason to caused that?*

**Visualization strategies**

The action of governments and the cases number of Covid-19 are all time serious data but they all independent variables to each other. To express the relationship between them is more complex, therefore, Gantt Chart might be a possible way to express the internal relationship of the epidemic situation and government actions and other time-dependent events.

**Data processing and Visual encoding**

There is no avaliable dataset contained all information required from open sources. Therefore, the whole process of data is required.

About the the epidemic situation, only the relationship(decrsing, incresing) is required but not the exact value. From previous dataset, the excel tools is easily to extract the increasing and decresing period of date.

The government actions information is from *Institute for governmen*, these action are classifed as national lockdown and other events, R can be used to creat a data frame contains the information for all of them.

The data frame with event and start columns has been prepared, and it is easy to create a timeline with the epidemic relation and other government actions. The Gantt Chart was shown as below.

From the Gantt Chart, it clear to see that the first and the third national lockdown have changed the increasing trend of daily new cases reported, however, the second national lockdown seemed it has not significant effect on control the increasing trend of Covid-19. However, it can be also to noticed that the time period of second lockdown is shorter than the first and third lockdown. That might limit the effect of control the increase of Covid-19.

It can be also noticed that social distance and rule of six did not prevent the increasing period of Covid-19. This might be a evidence to prove that the government announcement or Promotions and slogans for those rules are not strictly followed. The start of vaccine can be believed a great news for control the virus; the visualization also shown that the decreasing period was coming later than the first vaccine in UK.

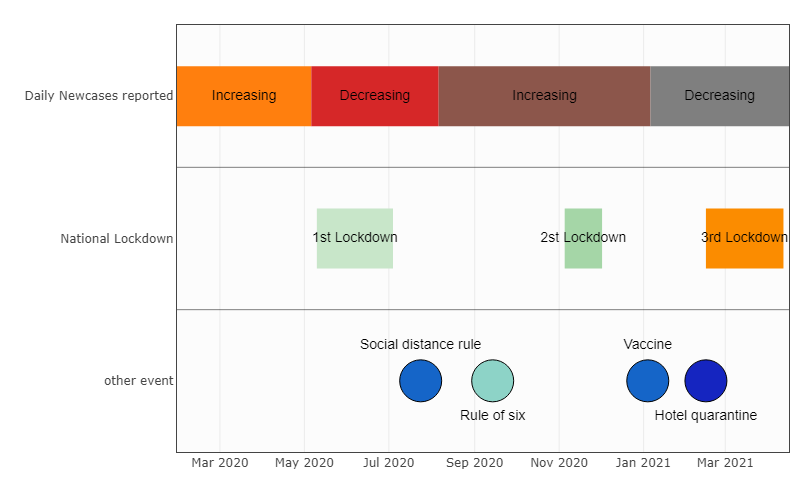
****

Fig7. The trend with time series with government actions

**Future development**

This report provides vislization solution to the general variation of covid-19 on time, areas and government action individually. To improve the effectiveness, the complex interaction is able to more represent detailed information. For example, the variation of number of cases can be represent by animation with time, this provides a more effective way to show the variation more directly. An integration of these information visualization should also be considered, a dashboard on web could be a good platform to represent all information in a more elegant layout which can also improve the effectiveness for users.

**Conclusion**

After data analysis and information vislization, it is possible to conclude that give the answer to the question.

**Question 1 - Covid-19 vs Time**

*1.1 When is the epidemic most severe?*

Generally, the Covid-19 epidemic experienced two severe period, the first one is around April in 2020 and second one is in the winter between 2020 and 2021. From the number of new cases, the second period has significantly higher number which means that the virus was extremely spreadable at that time. However, the deaths cases were similar which means in first period, the virus was much more dangerous. Both of period could be treated as the most severe period.

**Question 2 – Covid-19 vs Areas**

*2.1 Which areas are suffering most severe condition of epidemic?*

North West and London are suffering more severe condition of epidemic than others.

*2.2 Which areas are good at control of epidemic?*

South West and Scotland have lower cases than other areas.

**Question 3 - Covid 19 vs Government Actions**

*3.1 Are those national lock down all effective and find the reason to caused that?*

The first and third lockdown action are useful to control the trendy of increase of Covid -19 cases. The second one is not that useful because the time period is shorter than others.

In general, UK have experienced very severe period of Covid-19, it have been reached to a very high point. Fortunately, the government actions is proved to be useful to control the spread out of virus, the lockdown caused very positive impact on decreasing the new cases number. Although the general trendy for the virus is positive, but people should still be careful about it and do not make the third peak come again. Anyway, the vaccination have been put in used and this can bring confidence to people to finish the epidemic.

**Reference**

1. GOV.UK (2021) Coronavirus (COVID-19) in the UK, 15 April. Available at: https://coronavirus.data.gov.uk/ (Accessed: 15 April 2020).
2. Office for National Statics (2021) Latest data and analysis on coronavirus (COVID-19) in the UK and its effect on the economy and society., 15 April. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases (Accessed: 15 April 2020).