Observation of UK districts' population density and its effect on COVID-19 spread in Python

Author: Constantinos Constantinou, June 2020

This simple exercise is carried out to observe the effect of the population density within the UK on the spread of the COVID-19. To assist this exercise, the development of coding in Python was created to gather the COVID-19 accumulated infections data from the government's health services in the four countries England, Scotland, Wales and Northern Ireland. With the aid of data map projection and graphs the accumulated COVID-19 infections can be observed by district.

When thinking of the COVID-19 transmission one may argue that it is anticipated that in high population density areas the spread of the virus will be more evident than in areas where the population density is lower. Allegedly this is a very logical argument as in these areas more people are prone to interact with others while doing their grocery shopping, while walking or exercising outdoors or while commuting via public transport. However, it is interesting to observe the degree of this argument being true.

The demographics data are made available on the individual country's national records of statistics [1] [2] [3] [4]. The records used include the most recent data from the previous mid-year 2018 population density updates for all countries except for Scotland which has issued a mid-year 2019 population density estimation. However, the data gathered give a very good representation of the population density to our date.

The COVID-19 accumulated infections data by district are made available by the individual country's health services [5] [6] [7] [8] and infections are added up since the virus outbreak detection in UK.

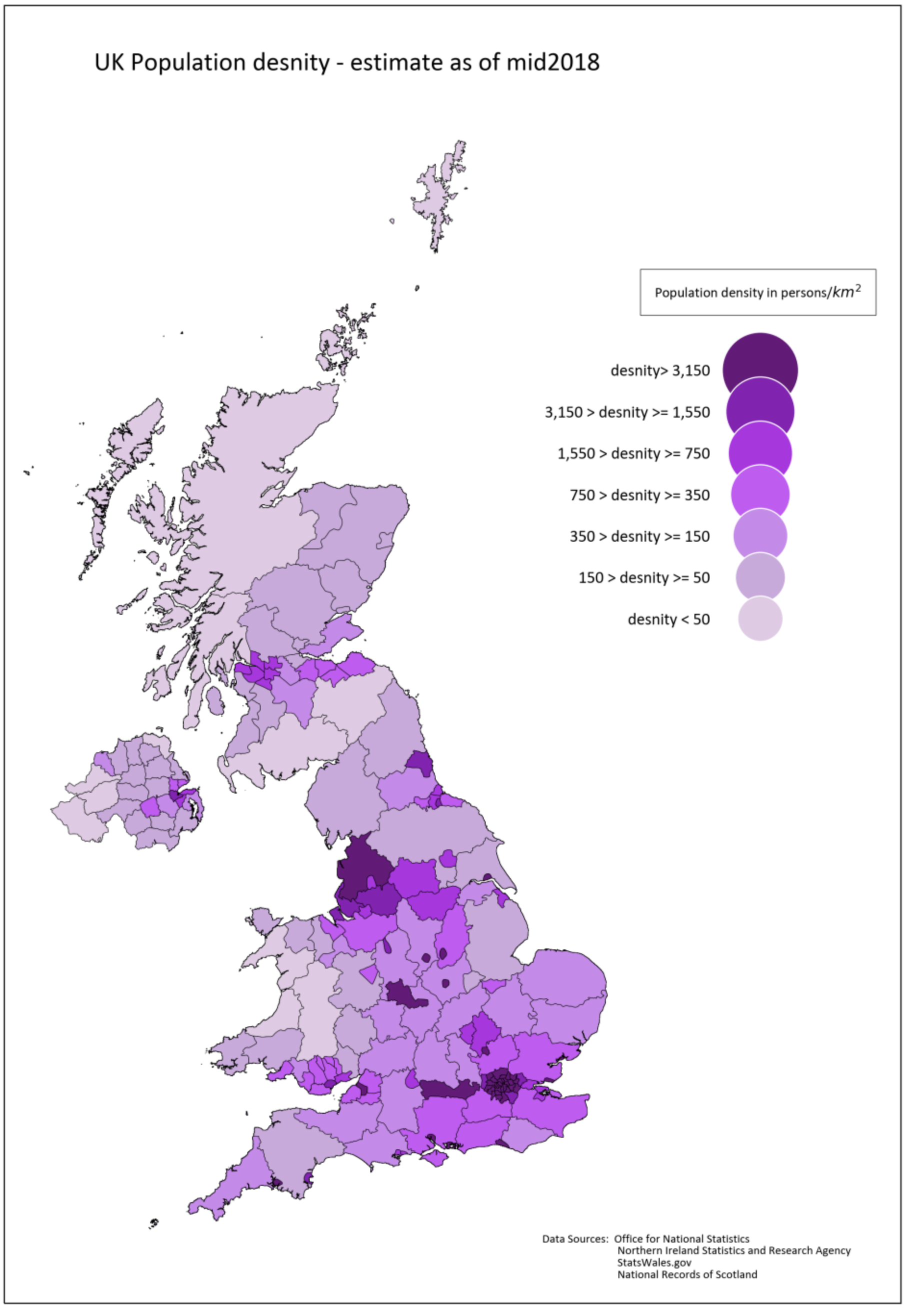
For the purpose of this exercise only the COVID-19 accumulated infections cases have been used for each district and not the deaths. The data collection was carried out on the following dates: 17-03-20, 08-04-20, 09-05-20 and 07-06-20. The idea behind these dates selection is to observe how did the accumulated infections grew at crucial points, the early stages of the virus spread, at its peak and towards the lower levels of the infection spread.

Implementation in Python and analysis

To start with, the packages used in the coding exercise are the *Cartopy* Python package for geospatial data processing and creation of maps, *matplotlib* package for the creation of graphs and plot diagrams and the *Shapely* package to be able to read and import the spatial geometric files onto maps. The actual source code for the exercise can be found in GitHub repository https://github.com/Constantino-C

The map in figure 1 helps to visualise the population density distribution throughout the UK. It is mostly evident that each London district has among the highest population densities, as well as Berkshire, West Midlands and Lancashire.

Figure 1. Population Desnity Distribution Map



The following figures 2,3,4 and 5 show the recorded incidents of the COVID-19 infections throughout UK districts projected on the UK geographical map at the selected dates 17-03-20, 08-04-20, 09-05-20 and 07-06-20.

The district with the maximum recorded accumulative infections is identified and highlighted on the map with the recorded cases shown on the figures.

The legend shows the number and the percentage of the districts which have recorded cases within the limits set in the algorithm which are expressed as a percentage of the highest infected district recorded cases. Therefore, the districts are categorised based on the COVID-19 accumulative cases and coloured as per the legend.

Note that Northern Ireland is shown as one unbroken district which is since data for the individual districts have not been made available in accessible format by the NI health services. In this respect, when comparing the COVID-19 accumulative cases to the population density, NI has been excluded in order to derive more meaningful conclusions.

Figure 2. UK COVID-19 accumulative incidents as of March 17th 2020

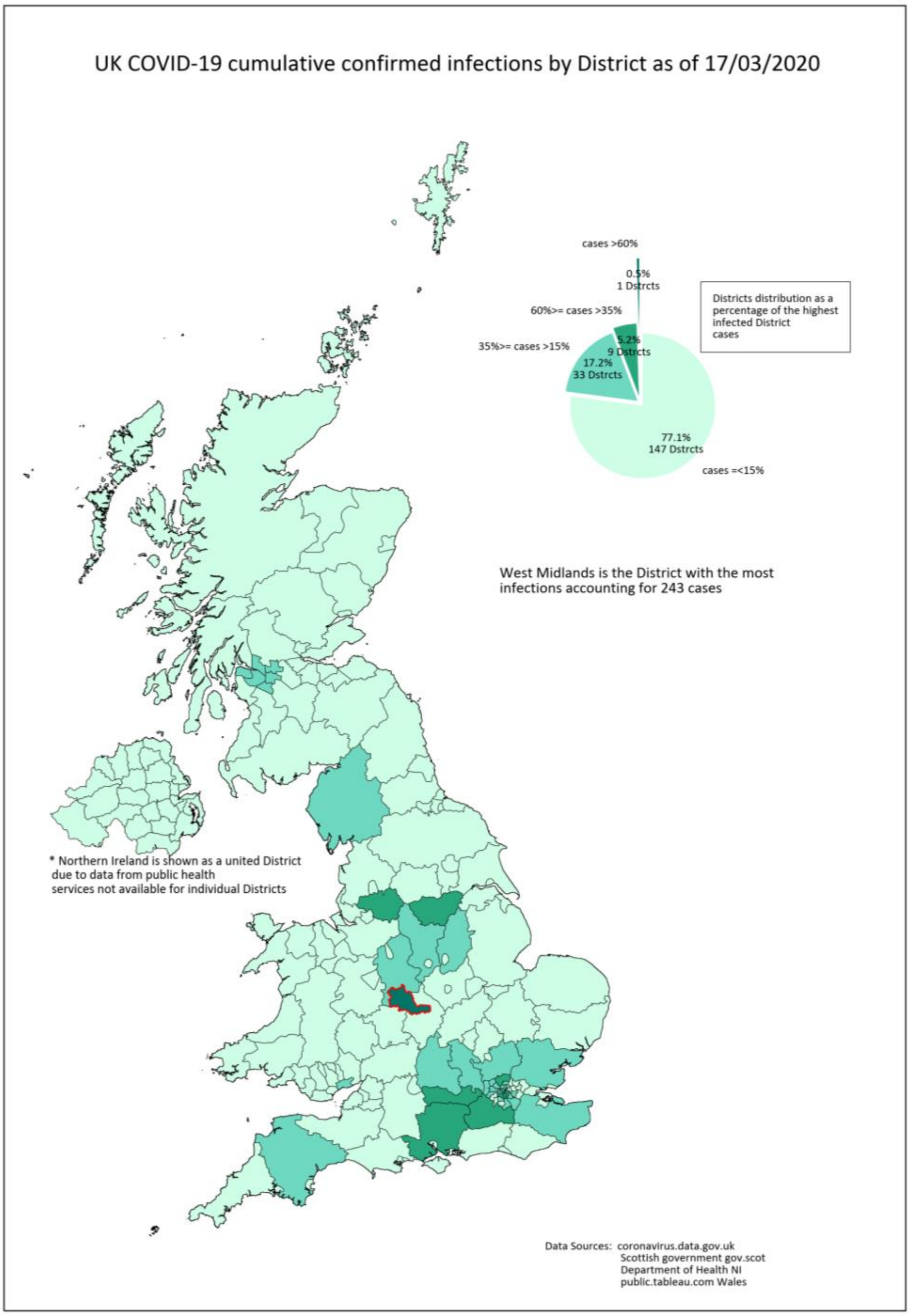


Figure 3. UK COVID-19 accumulative incidents as of April 8th 2020

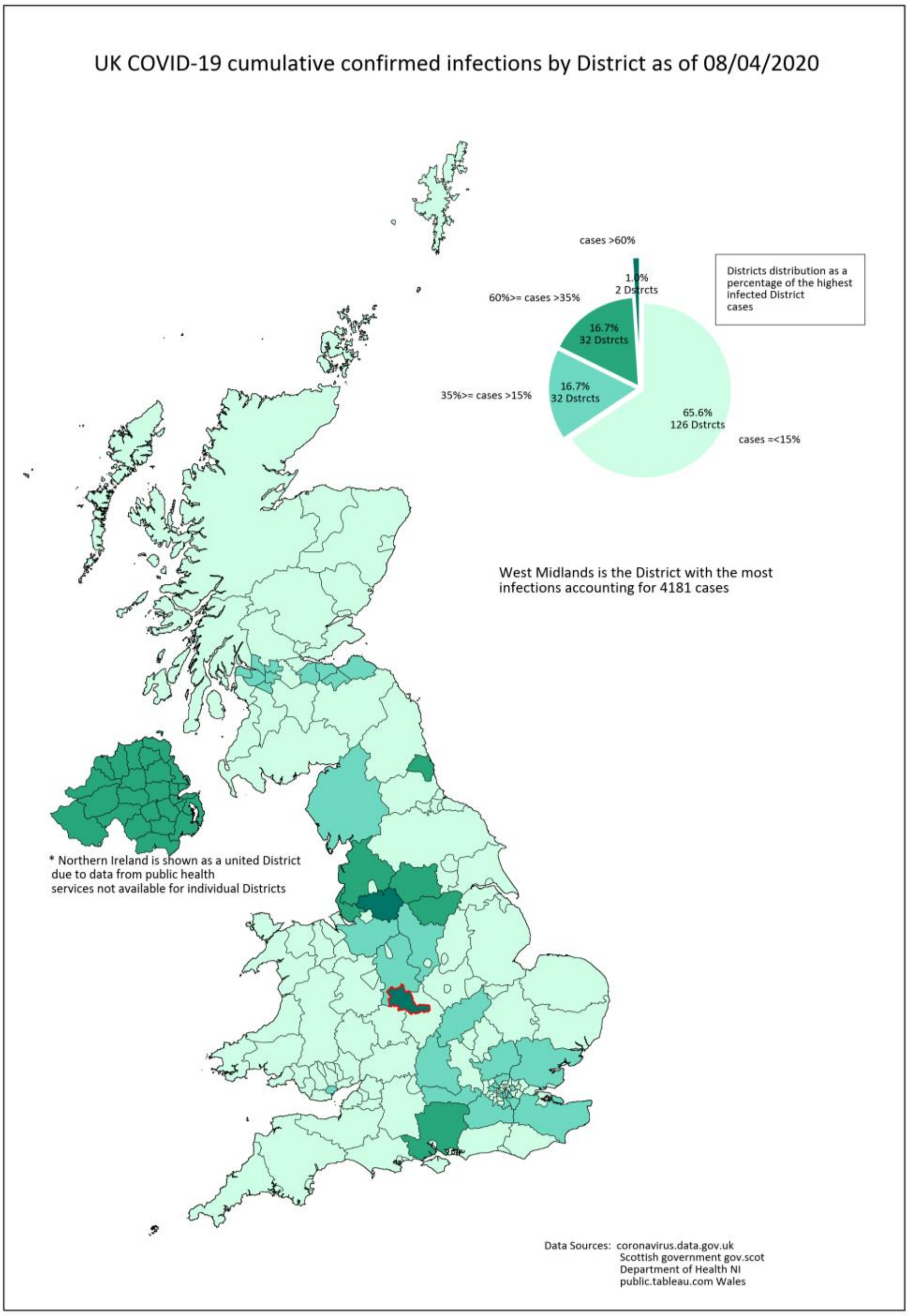


Figure 4. UK COVID-19 accumulative incidents as of May 9th 2020

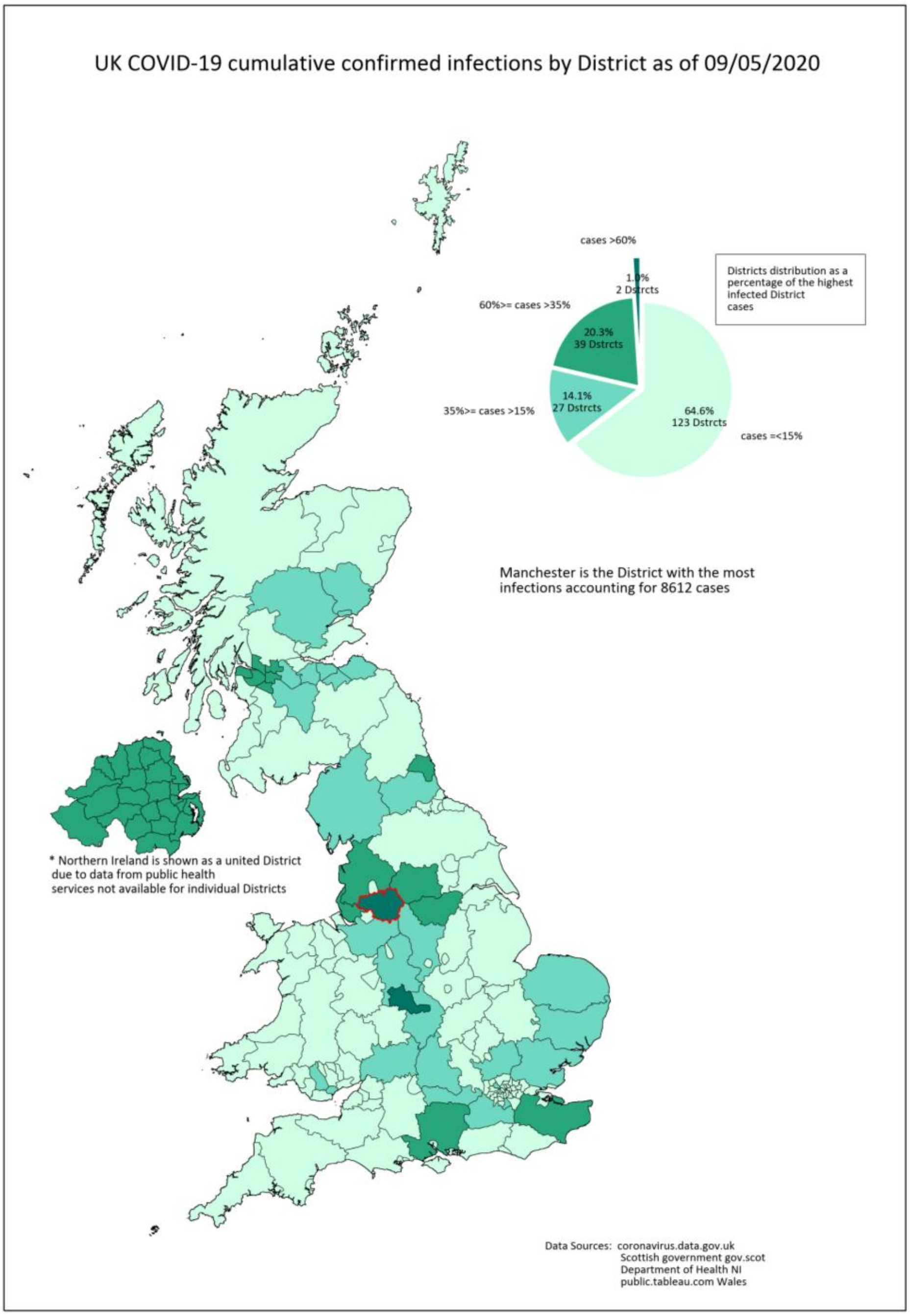
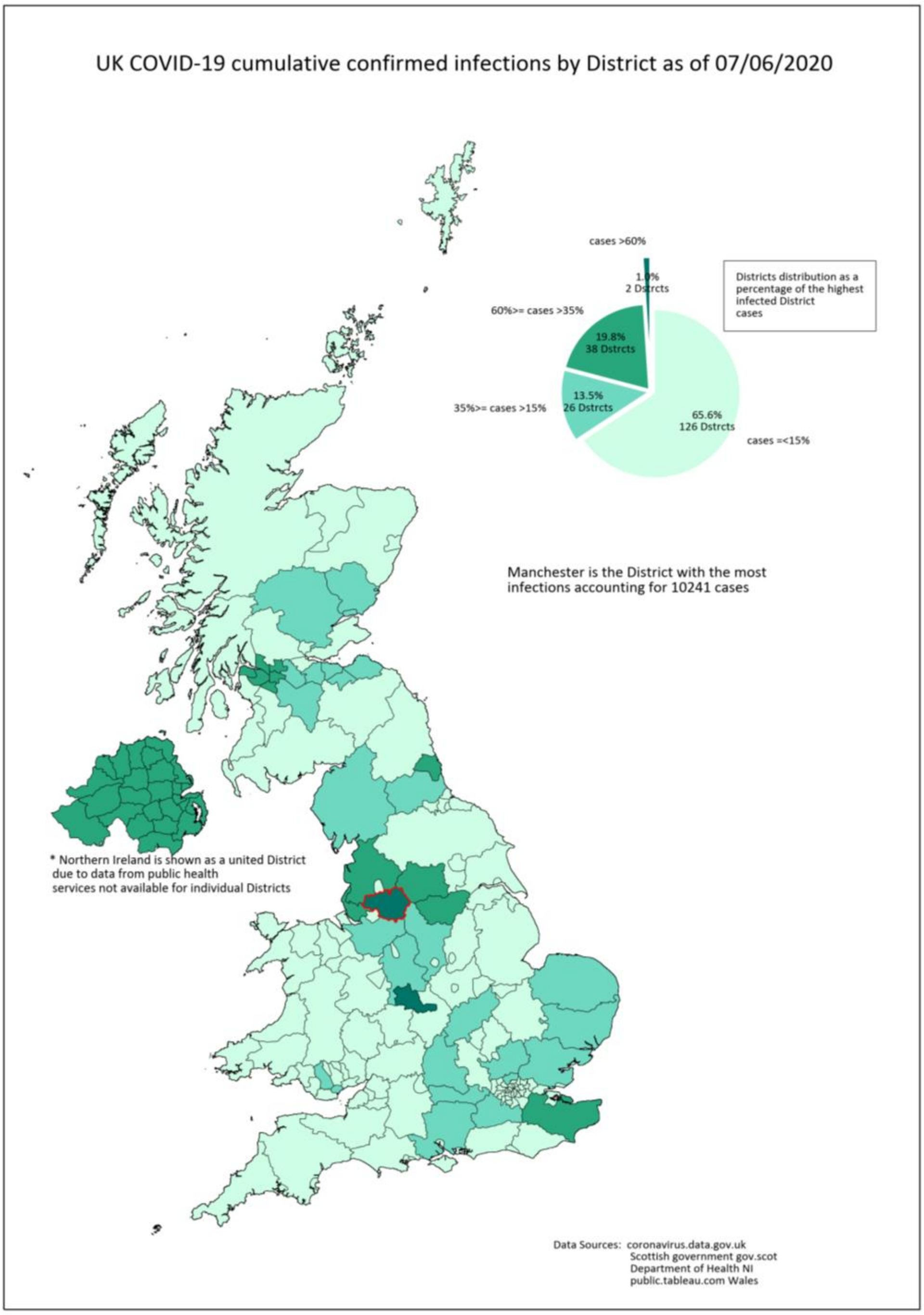


Figure 5. UK COVID-19 accumulative incidents as of June 7th 2020



Now we have the four map figures showing the COVID-19 accumulative infections at different stages of the virus outbreak. Observing the figures above the following is understood:

- Considering London and precisely the individual districts forming up the area of London which are clearly among the highest population density areas, it can be seen that the individual boroughs are not among the hardest hit regions by the virus outbreak.

- If we were to aggregate the London districts to account for one London district, then we would have similar effect to that of Northern Ireland and this would not be particularly helpful in comparing the individual districts.

- From the onset of the virus outbreak the most infected England regions have been almost steady in leading the infection cases with these being West Midlands, North West and South East, with Scotland's Glasgow and the Clyde, Edinburgh and the Lothians and North East Scotland cases becoming more evident after the virus spread peak after May 9th.

- Comparing the population density map in figure 1 with the COVID-19 accumulative infections by district we can see that approximately the most populated districts have accumulated the most recoded cases. This is expected as more people are prone to interact with each other.

- It appears that the hardest hit districts by the virus infections are Manchester and West Midlands.

Another question that arises regarding the virus spread is how the virus accumulative infections varies district by district in comparison with the population density. Let's see if the following figures help us understand this further.

In the following figures, the COVID-19 accumulative infections and the population density for all UK districts and in individual countries have been projected on the same graphs for the four selected dates.

It is difficult to identify similarities of the COVID-19 cases and population density trends looking at the UK and England graph since there are many data adding to the obscurity of the peak points.

Looking at the less dense graphs those of Wales and Scotland, it is most evident that the both COVID-10 and population density follow almost always the same trend. This helps to get a meaningful idea of the relationship between the two parameters examined.

However, if we take a closer look at the UK graph, we can see that the two parameters do always follow the same trend; in fact, there are occasions where they follow opposite trends. Let's investigate the trend relationship of them closer by quantifying the trends and categorise them. See this analysis section after the following graphs.

UK COVID-19 accumulative infections and population density graphs

Figure 6. UK COVID-19 infections and population density by district – 17.03.20

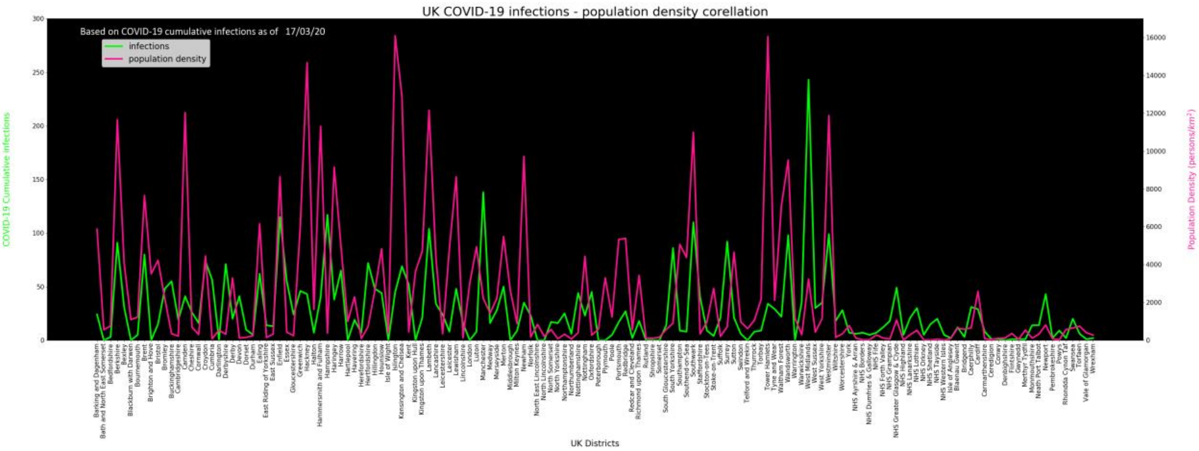


Figure 7. UK COVID-19 infections and population density by district – 08.04.20

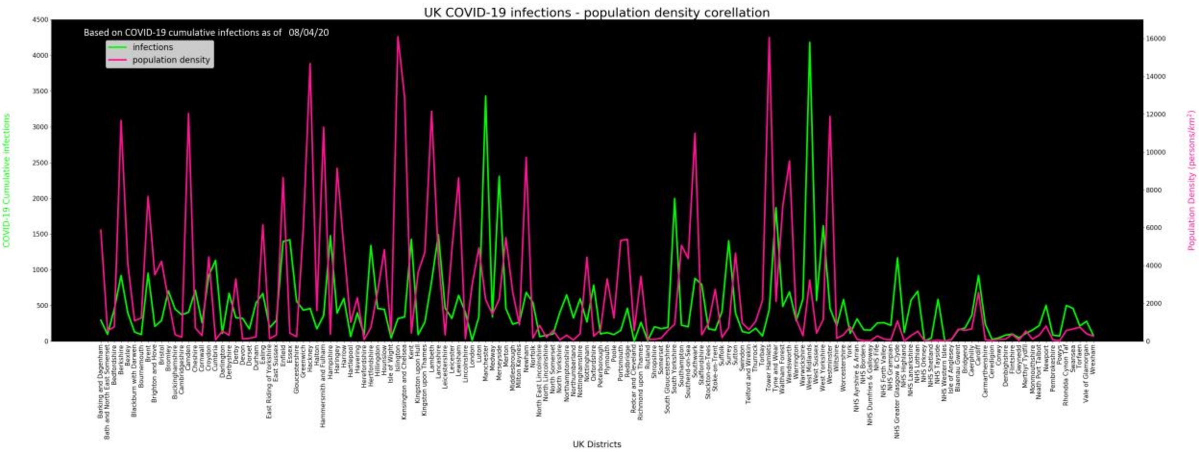


Figure 8. UK COVID-19 infections and population density by district – 09.05.20

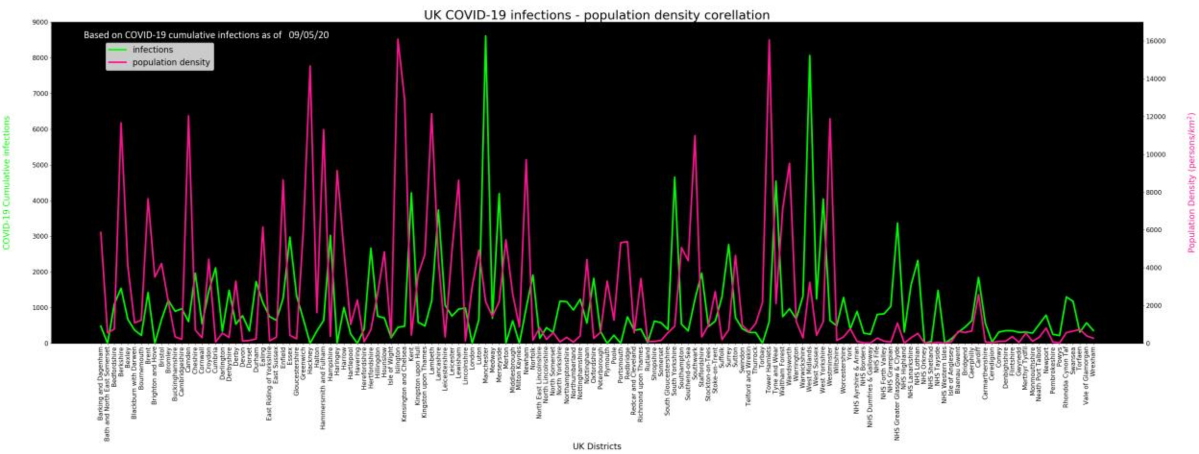
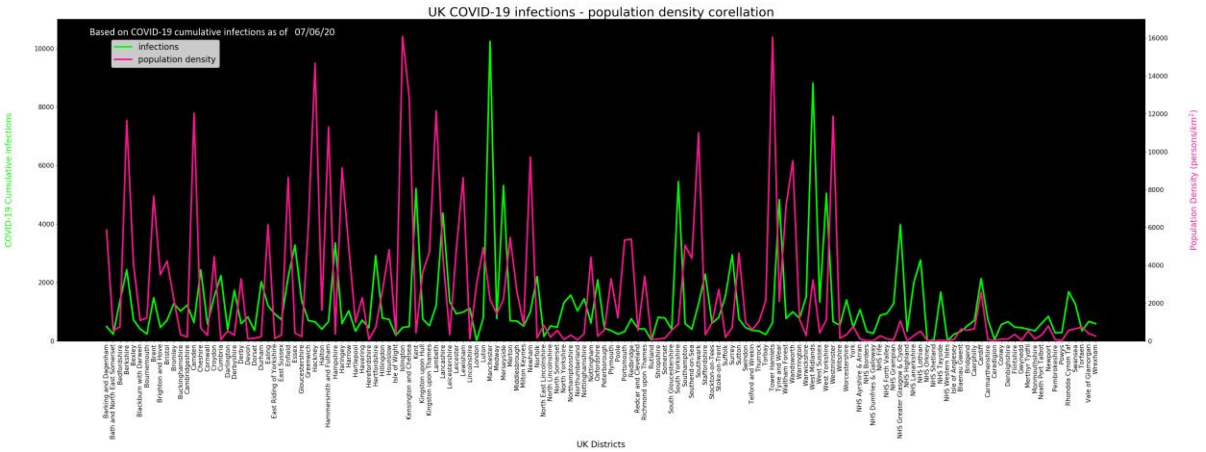


Figure 9. UK COVID-19 infections and population density by district – 07.06.20



England COVID-19 accumulative infections and population density graphs

Figure 10. England COVID-19 infections and population density by district – 17.03.20

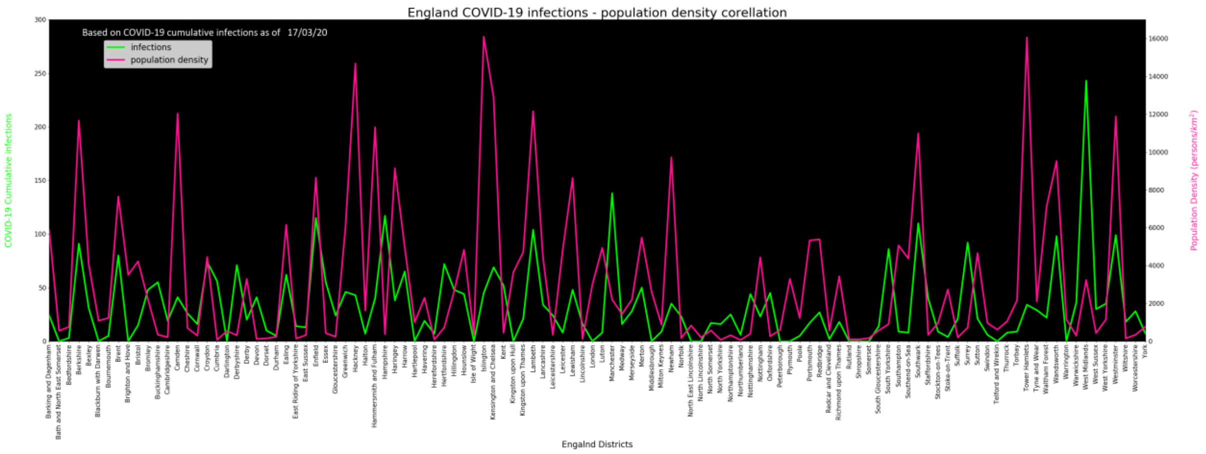


Figure 11. England COVID-19 infections and population density by district – 08.04.20

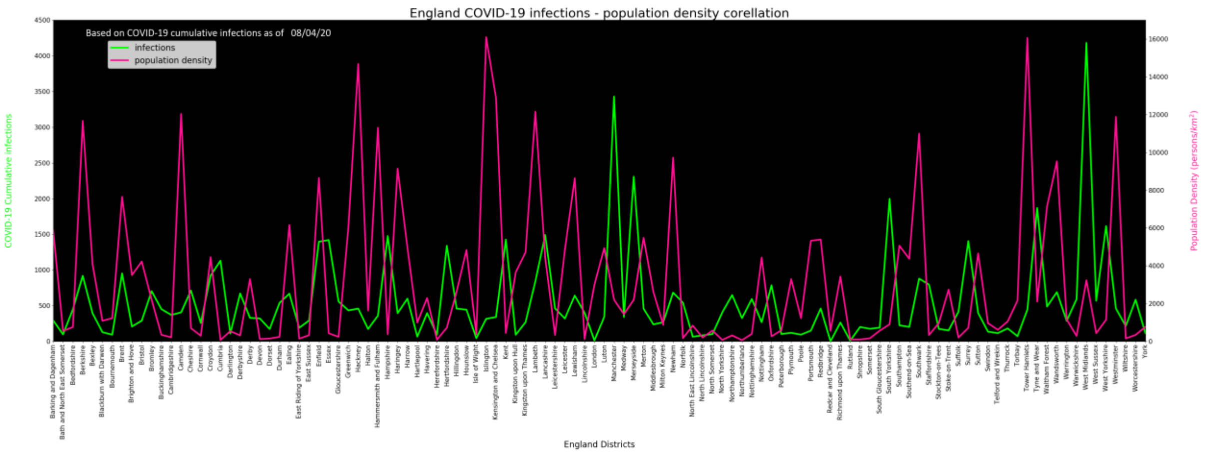


Figure 12. England COVID-19 infections and population density by district – 09.05.20

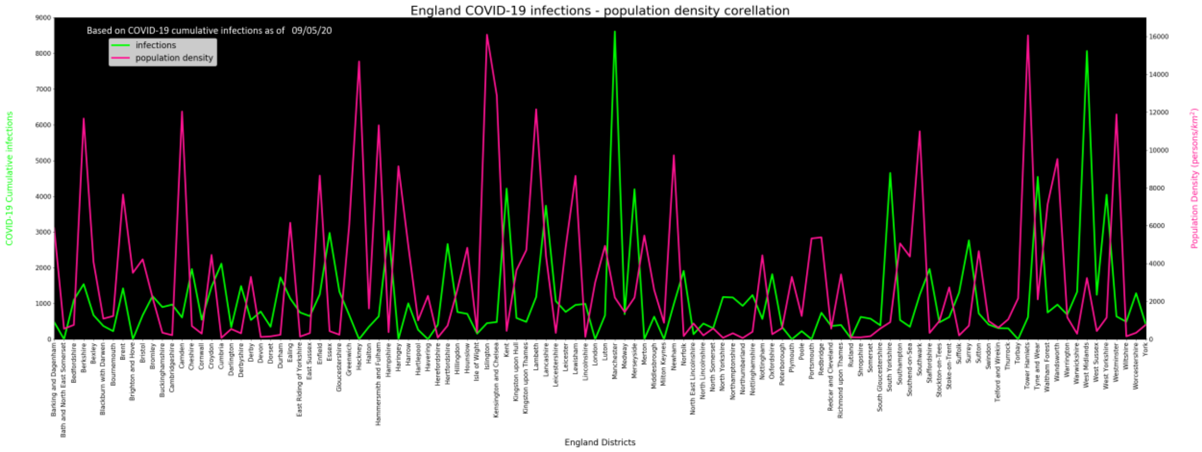
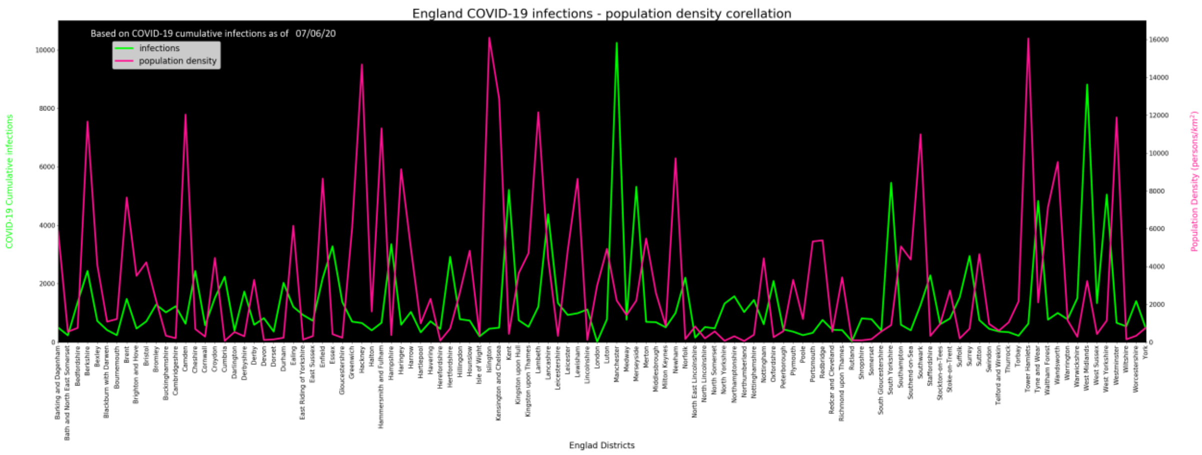


Figure 13. England COVID-19 infections and population density by district – 07.06.20



Scotland COVID-19 accumulative infections and population density graphs

Figure 14. Scotland COVID-19 infections and population density by district – 17.03.20

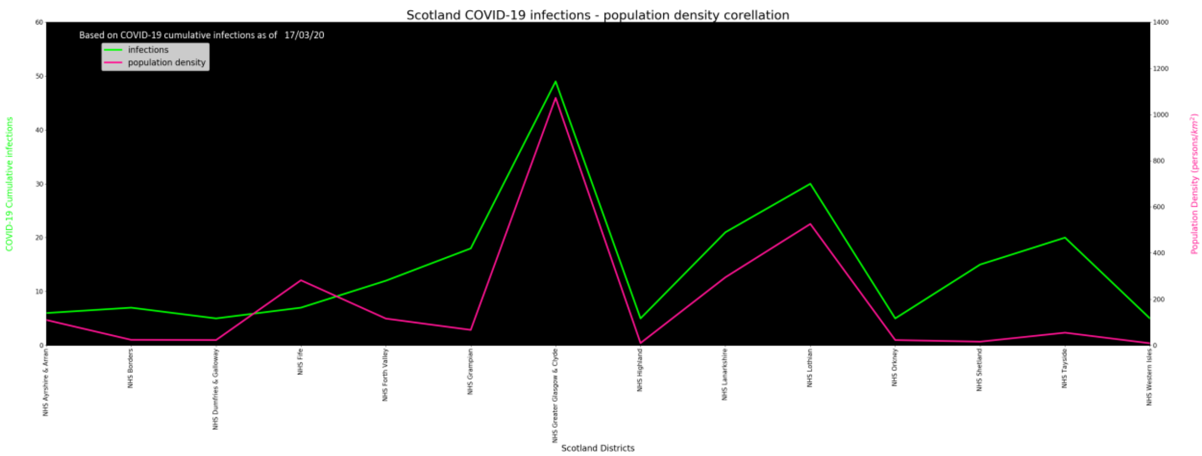


Figure 15. Scotland COVID-19 infections and population density by district – 08.04.20

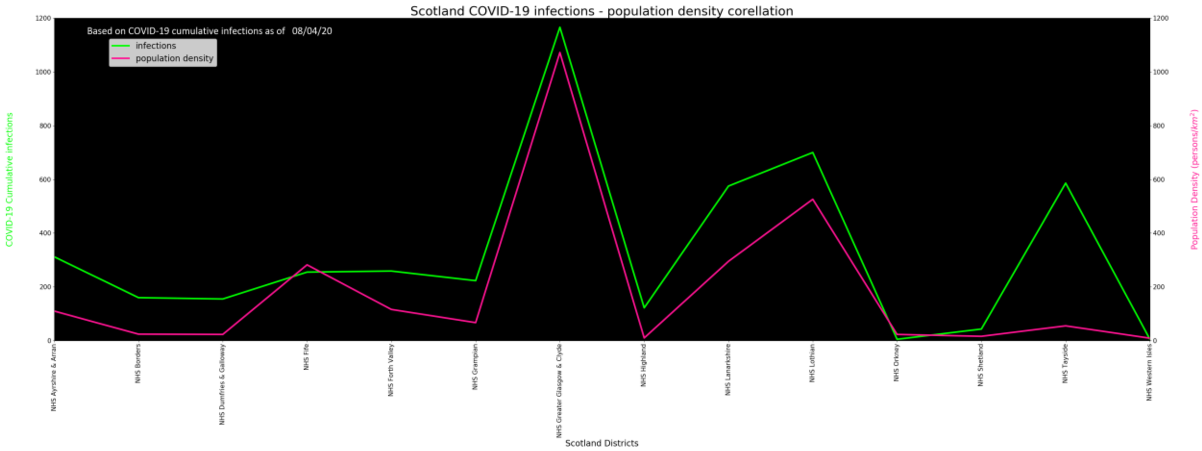


Figure 16. Scotland COVID-19 infections and population density by district – 09.05.20

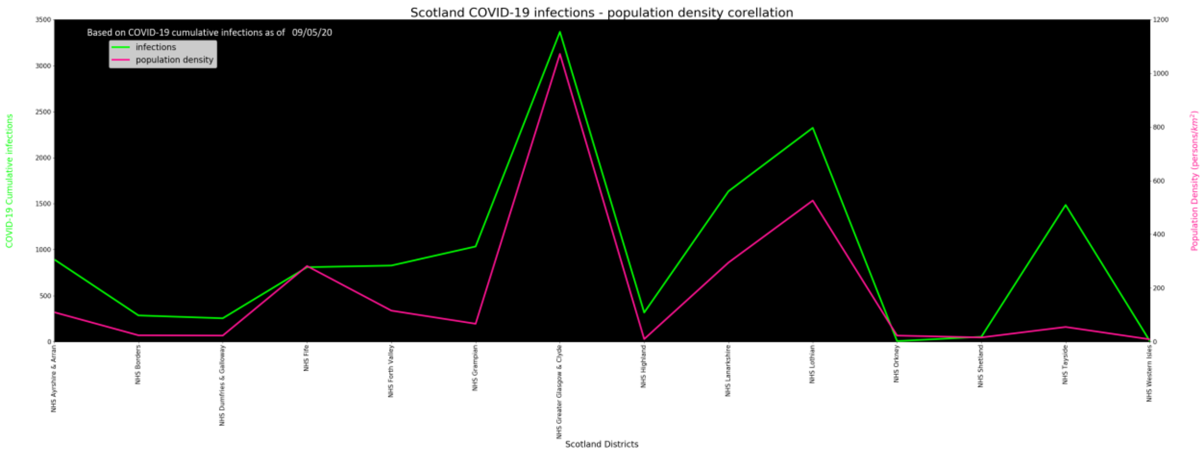
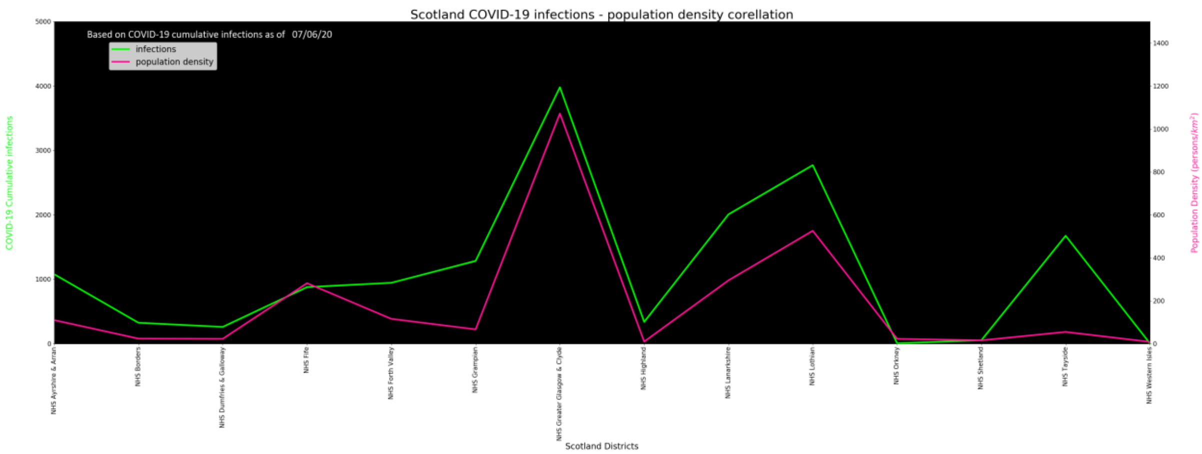


Figure 17. Scotland COVID-19 infections and population density by district – 07.06.20



Wales COVID-19 accumulative infections and population density graphs

Figure 18. Wales COVID-19 infections and population density by district – 17.03.20

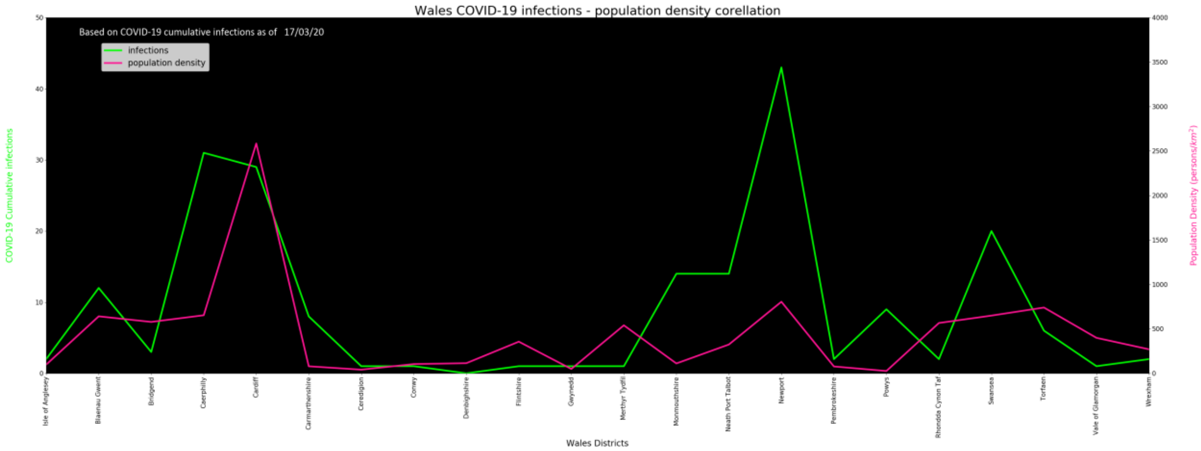


Figure 19. Wales COVID-19 infections and population density by district – 08.04.20

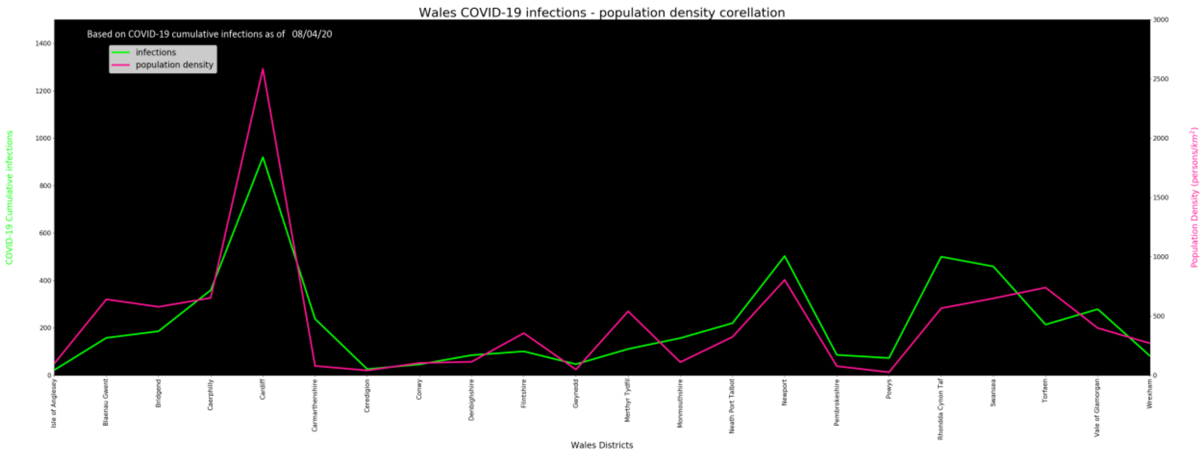


Figure 20. Wales COVID-19 infections and population density by district – 09.05.20

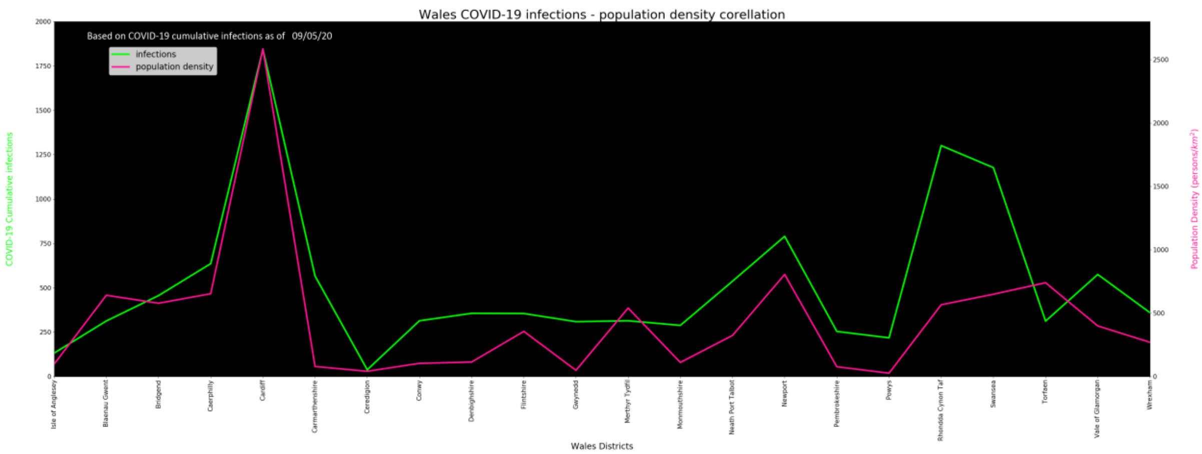


Figure 21. Wales COVID-19 infections and population density by district – 07.06.20

Normalised Correlation of trending data

We have seen previously that in some cases the COVID-19 cases and population density data follow the same trend and in some other cases follow opposite trends. In this section we will aim to quantify the cases of alike and opposite trends. This will help understand better the degree of the relationship of the two parameters.

If we look at the following figures – for the same periods of time as applied before – we will see the trends have been normalised into three categories as follows:

- ***Category 1:*** The population density and COVID-19 infection cases follow the same trends. This means the two parameters follow either the same upwards direction or downwards direction moving from district to district.

- ***Category 0:*** The population density and COVID-19 infection cases follow different trends moving from district to district: *a)* The population density follows an upwards trend and the COVID-19 cases follow a downwards trend *or* b) The population density remains stable and the COVID-19 cases increase or decrease.

- ***Category -1:*** The population density and COVID-19 infection cases follow different trends moving from district to district: *a)* The COVID-19 infection cases follow an upwards trend and population density follows a downwards trend *or* b) The COVID-19 cases remain stable and population density increase or decrease.

Figure 22. UK COVID-19 & population density normalised correlation 17.03.2020

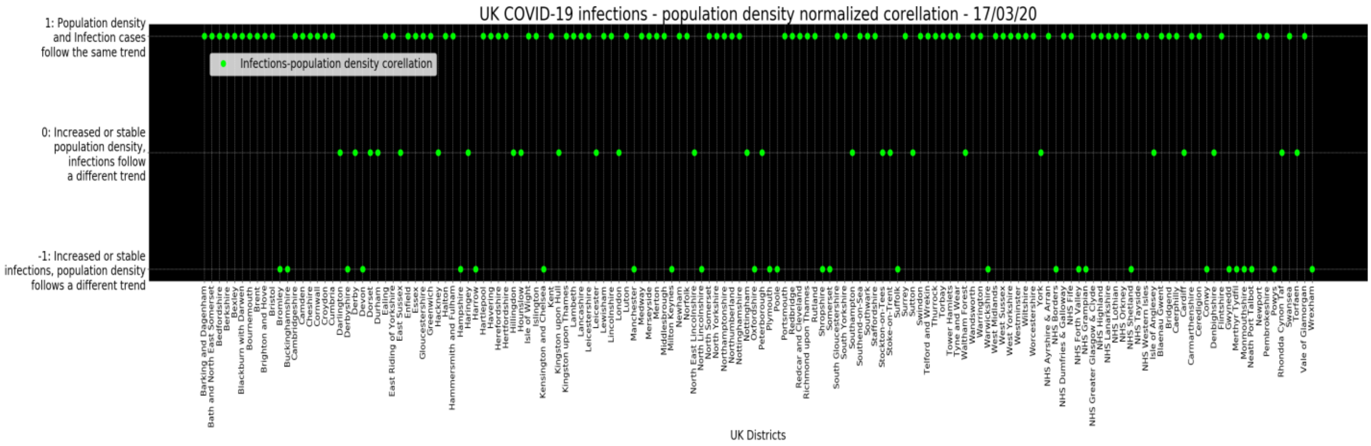


Figure 23. UK COVID-19 & population density normalised correlation 08.04.2020

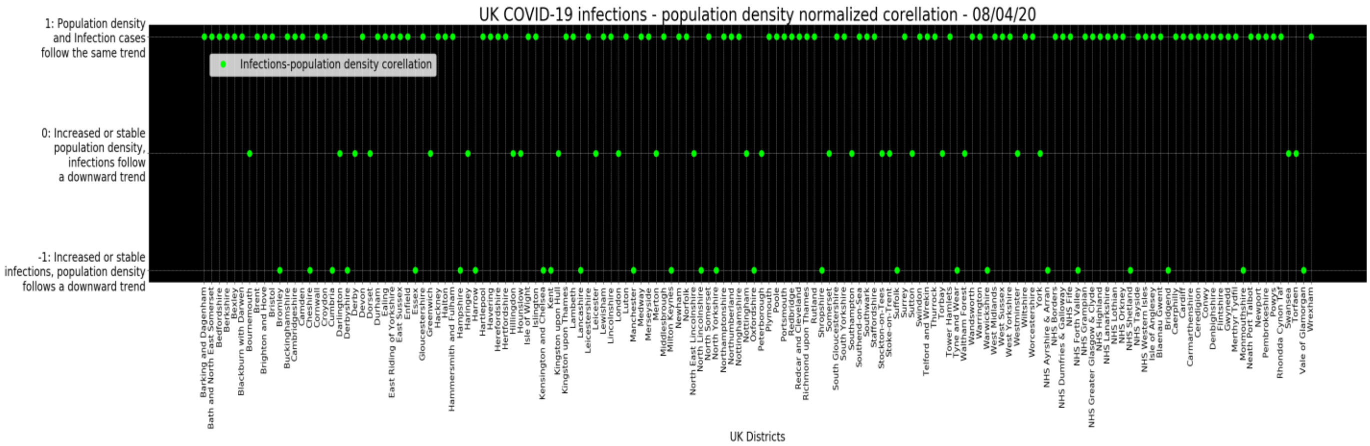


Figure 24. UK COVID-19 & population density normalised correlation 09.05.2020

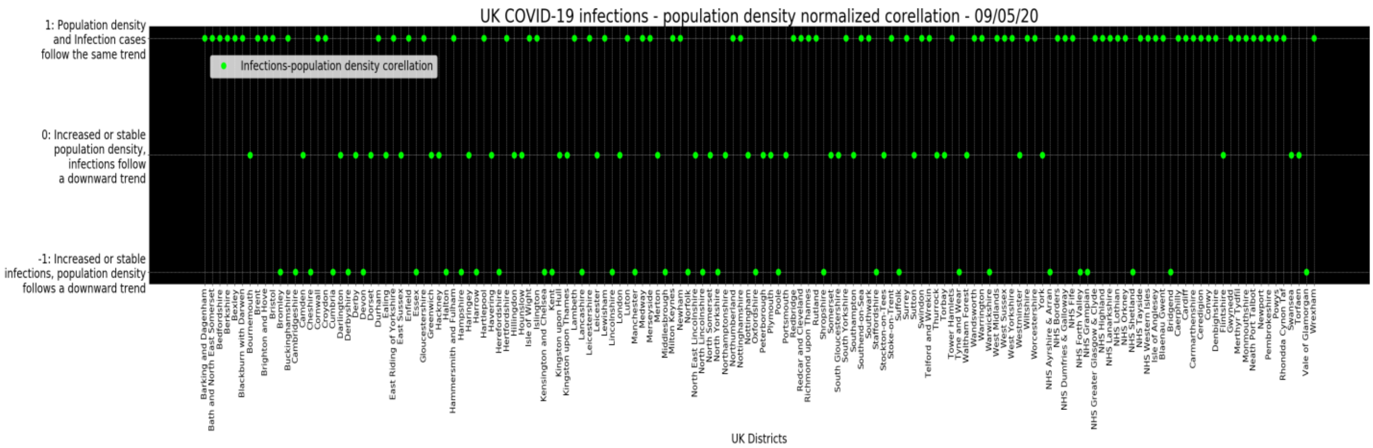
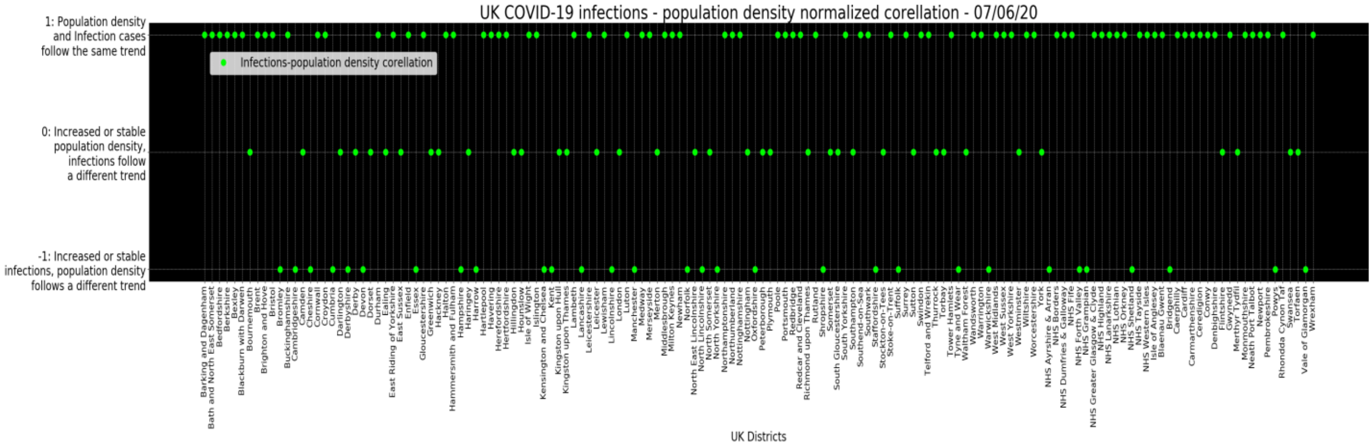


Figure 25. UK COVID-19 & population density normalised correlation 07.06.2020



Let’s put the above quantities in table for a better perception.

Figure 26. Normalised COVID-19 & population density trend quantities

Categories	17.03.20	08.04.20	09.05.20	07.06.20
C[1]	94	97	77	81
C[0]	26	26	38	37
C[-1]	28	25	33	30
C[1]/(C[0]+C[-1])	1.741	1.902	1.085	1.208
C[1] as % of total	0.635	0.655	0.520	0.547
C[0]+C[-1] as % of total	0.365	0.345	0.48	0.453

At first look, it appears from table 1, that the COVID-19 infection cases and population density category 1 trends are the majority out of all categories. In the first months of the virus outbreak this is very obvious with category C[1] trends being almost double than the other two categories C[0] and C[-1]. At the peak of the virus outbreak (around 08.04.20) category C[1] is 1.9 times more than the categories C[0] and C[-1] combined.

However, from the data above, it is realised that the relationship between COVID-19 infection cases and population density is not interchangeable. It may be that C[1] in the first months has reached up to 65.5% of the total trending data, yet in the following months when the accumulative COVID-19 infection cases have increased in all districts, C[1] has fallen down to 54.7%.

The above observation tells us that in fact population density has a role to play in the COVID-19 infection cases spread; however, there are other factors that have a significant impact on the virus spread.

Some ideas as to what these factors can be - and a thought for a future work – are the people obedience to the official health guidance, numbers of people travelling throughout the country for work or other reasons, proven efficiency of protective measures throughout the public, locations of mandatory application of protective measures.

Conclusions

This simple exercise attempted to verify the significance of the population density impact on the COVID-19 spread.

As we have seen above, the population density has a major role to play in the virus spread but there are also other parameters which have a crucial part in the spread and these are far from negligible.

Scenarios of virus outbreak throughout the globe as the COVID-19, are very hard to manage with measures that we can be sure they will be efficient enough and this is mainly due to the rarity of such virus outbreaks and lack of experience. This is a good chance for humanity to study in detail COVID-19 and its spread, the protection measures used and their efficiency in order to establish a proven primary response method for any future similar scenario.

Eventually, data analysis in the case of COVID-19 is a field which has a lot to offer and provide its benefits by analysis even more complex data to help understand and fight such virus outbreaks.

References

[1] https://www.nrscotland.gov.uk

[2] https://statswales.gov.wales

[3] https://www.nisra.gov.uk

[4] https://www.ons.gov.uk

[5] https://coronavirus.data.gov.uk

[6] https://www.gov.scot

[7] https://www.health-ni.gov.uk

[8] https://public.tableau.com

[9] http://www.diva-gis.org/gdata