Geospatial Study of COVID-19 vis-à-vis Medical Care in India

By Amrita Iyer

1. Introduction

On 30 January, WHO declared COVID-19 a Public Health Emergency of International Concern (PHEIC). India is the second most populous country of the world. According to the 2019 revision of the World Population Prospects, India's population has crossed 1.3 billion.

This population is spread out in 29 states and 7 union territories. Healthcare is a state subject, which means that states are free to cater to their people's needs as they see fit, on a case-by-case basis. Since Independence, India has focused on the larger picture. The priority in a developing country is primary care at the peripheral level, preventive measures, immunisation, maternity and paediatric care as well as dealing with common infections such as tuberculosis. Public health centres and community health centres are the usual providers of these services. However, not enough hospital beds and specialised facilities have been provided by the public sector during this time.

Hospital-population ratio and the hospital bed-population ratio are good metrics indicating healthcare infrastructure. The World Health Organization provides the rationale behind the hospital beds to population ratio: "Hospital beds are used to indicate the availability of inpatient services. There is no global norm for the density of hospital beds in relation to total population". As there is no global norm to follow, every country, state and district must build the necessary infrastructure on a case-by-case basis, based on metrics like population, last mile reach. With COVID 19 into the mix, the number of testing centres has also emerged as a new metric to understand how equipped states are dealing with the COVID-19 crisis. Also apparent is the disparateness between rural hospitals/beds and urban hospitals/beds vis-à-vis the population. All this results in a mixed level of public health service in India.

1.1 Aim

Public health infrastructure of a country must be equipped to treat its people. This project aims to visualize healthcare infrastructure data of the state containing the highest number of COVID cases as of now, to illustrate why it is necessary to overhaul the public healthcare in India.

1.2 Stakeholders

In general, each and every one of us is a stakeholder in a discussion on healthcare. However, to be more specific, such analysis and visualizations can help administrators and law-makers explore options on how to make healthcare available and affordable to everyone at the national, state and district levels.

2. Data

2.1 Data sources

The data to be used is summarised below:

1. Statewise COVID-19 data is available at the Ministry of Health and Family Welfare website rendered into a csv file. It is a public dataset available to all.
2. Nation-wide hospital data and ICMR (Indian Council of Medical Research) labs are available from Kaggle here and here. These are also public datasets as indicated by the creator. These datasets will be primarily used to visualize the status of healthcare infrastructure in India.

3. Pincode data with corresponding latitude and longitude is available from Kaggle here. It is required to create a database of ICMR labs with respect to their locations. This is also a public dataset.

4. Google Places API can be leveraged to get a closer look at hospitals within the state containing highest number of COVID19 cases.

5. Folium based maps will be used to visualize the data geospatially.

2.2 Data description and cleaning

The dataset containing COVID-19 patients per state contains the name of the state, the number of active cases, the number of discharged patients, the number of deaths and the total number of cases (which is the aggregate of 'active', 'discharged' and 'death' columns). Firstly, the 'NaN' have been identified and removed. Two rows in this dataset contain only NaN values and so have been dropped. The 'S.no.' column is redundant and can be dropped. The shape of the dataframe shows 36 values, which corresponds to the total of states and union territories. Finally, the columns are renamed as follows:

	State/UT	Active	Discharged	Deaths	Total
0	Andaman and Nicobar Islands	10.0	35.0	0.0	45.0
1	Andhra Pradesh	3948.0	3917.0	96.0	7961.0
2	Arunachal Pradesh	92.0	11.0	0.0	103.0
3	Assam	1856.0	3039.0	9.0	4904.0
4	Bihar	2033.0	5098.0	50.0	7181.0

The healthcare data contains the name of the state, and the numbers of primary healthcare centres (PHC), community healthcare centres (CHC), sub district hospitals (SDH), total public health facilities (TPHF), district hospitals (DH), rural hospitals (RH), urban hospitals (UH), the total number of public beds (PublicBeds), the number of rural beds (RuralBeds) and the number of urban beds (UrbanBeds). The original column names have been changed to those in brackets. For the purpose of this study, the columns of PHC, CHC, TPHF have been dropped as they do not have specialised facilities meant to deal with a crisis such as COVID 19. This is not to suggest that such places are not contributing in dealing with the crisis. Their primary capability includes basic healthcare, prevention and educating the public. Second, the column 'SDH' have to be dropped as there were missing values. Third, the column 'DH' also has been dropped as the main aim is to compare the numbers of rural and urban hospitals. DH can be categorised under either category and so complicate the analysis.

	State/UT	PublicBeds	RH	RuralBeds	UH	UrbanBeds
0	Andaman and Nicobar Islands	1246	27	575	3	500
1	Andhra Pradesh	60799	193	6480	65	16658
2	Arunachal Pradesh	2320	208	2136	10	268
3	Assam	19115	1176	10944	50	6198
4	Bihar	17796	930	6083	103	5936

The ICMR data contains the names of the labs, their full addresses, pincodes, cities, states and the type of institution. The columns containing addresses have been dropped as pincode, city and state are enough. The column containing 'type' has been dropped as the entire list is of ICMR labs. Out of the 170 rows, 2 rows are missing this data. They have been dropped as well.

	lab	pincode	State/UT
0	ICMR-Regional Medical Research Centre, Port Blair	744103	Andaman and Nicobar Islands
1	Tomo Riba Institute of Health & Medical Scienc...	791110	Arunachal Pradesh
2	Sri Venkateswara Institute of Medical Sciences...	517507	Andhra Pradesh
3	Rangaraya Medical College, Kakinada	533001	Andhra Pradesh
4	Sidhartha Medical College, Vijaywada	520008	Andhra Pradesh

The dataset containing pincodes has the pincodes in relation to the name of the place, the city, latitude, longitude and accuracy. Excepting the pincodes, latitude and longitude, all other columns have been dropped.

	pincode	latitude	longitude
0	110001	28.6333	77.2167
1	110002	28.6333	77.2500
2	110003	28.6500	77.2167
3	110004	28.6500	77.2167
4	110005	28.6500	77.2000

The final dataframe is a merged dataset of the cleaned datasets.

	State/UT	Active	Discharged	Deaths	Total	PublicBeds	RH	RuralBeds	UH	UrbanBeds	lab
0	Andaman and Nicobar Islands	10.0	35.0	0.0	45.0	1246	27	575	3	500	1
1	Andhra Pradesh	3948.0	3917.0	96.0	7961.0	60799	193	6480	65	16658	5
2	Assam	1856.0	3039.0	9.0	4904.0	19115	1176	10944	50	6198	4
3	Bihar	2033.0	5098.0	50.0	7181.0	17796	930	6083	103	5936	2
4	Chandigarh	60.0	315.0	6.0	381.0	3756	0	0	4	778	2

Apart from this, Google Places API has been used to get the location of hospitals in the concerned 'State/UT' for analysis.

3. Methodology

3.1 Seaborn Relplot

I plotted the dataframe into a relplot from seaborn's library taking the features 'PublicBeds', 'TotalCases', 'ActiveCases' and 'CovidDeaths'. This is Figure 1. A relplot is essentially a scatterplot. However, its features 'hue' and 'size' allow for a visual depiction of four features in total in a single figure. Thus, from the resulting figure, we see that most of the states have public beds upto 20000. These states have active cases ranging from 0 to 20000 and deaths ranging from 0 to 2000. A few states have public beds ranging from 40000 to 60000 and active cases and deaths the same as above.

There are three notable outliers. The first (in dark blue) has around 20000 public beds and nearly 60000 COVID cases (in the range 40000-60000), with about 20000 active cases and more than 2000 deaths. This is New Delhi. The fact that the public beds are nearly equal to the active cases does not bode well because of factors like new admissions, relapses, etc. Also, it is unknown how many people have been discharged because of lack of public beds.

On the other end of the spectrum is the second outlier (in dark blue), which has over 70000 public beds, less than 60000 cases out of which nearly 20000 are active with less than 2000 deaths. This is Tamil Nadu.

The third outlier (in yellow) has nearly 70000 beds with more than 120000 cases of which more than 60000 are active with more than 6000 deaths. This is Maharashtra, arguably the worst affected state in India.

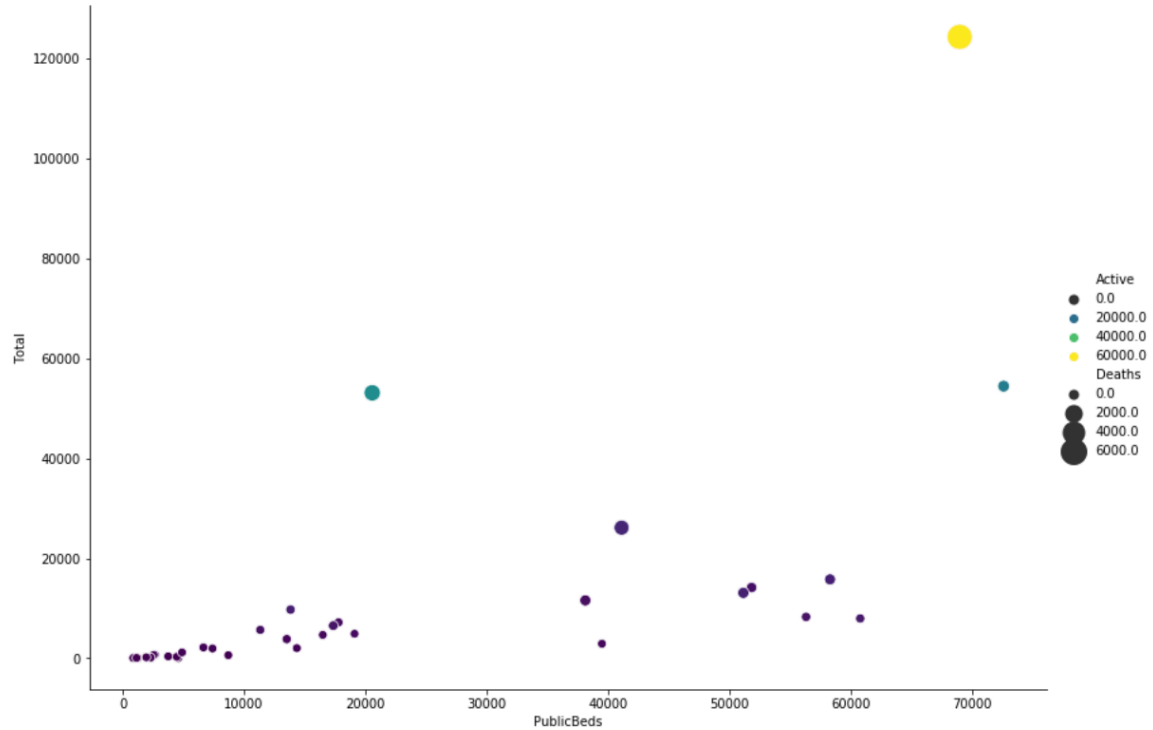

Figure 1: Relplot showing Total COVID cases ('Total') in relation to total no. of public beds ('PublicBeds') , along with active COVID cases ('Active') and no. of deaths ('Deaths')

3.2 Combined bar graph

After zeroing in on the most affected state, the next step is to check the healthcare metrics of the state vis-à-vis the other states. That has been done using bar graphs.

A. The statewise comparison of rural hospitals to urban hospitals shows that in most states rural hospitals are more than urban hospitals. This is seen in Figure 2. The exceptions are Delhi, which is completely categorised under 'urban', and most notably, Maharashtra, where urban hospitals are more. Even the number of hospitals in comparison to other states is dismal, with atleast 14 states being ahead of Maharashtra. This is indicative of improvement that can be made in healthcare in this state.

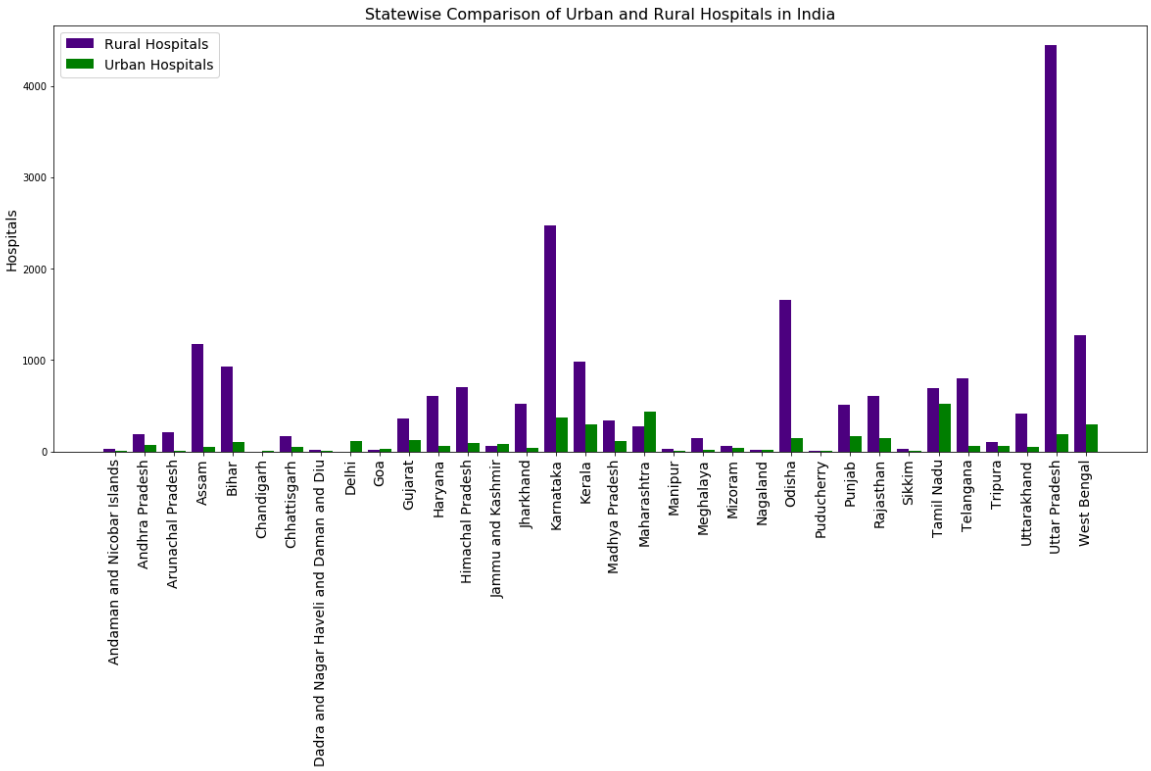

Figure 2: Combined bar graph showing statewise rural and urban hospitals

B. The next step is to compare the number of rural beds with those of urban beds. Out of 34 states/UTs[1] in the graph (Figure 3), 21 states/UTs have more have more urban beds than rural beds**, including Maharashtra.** The exceptions include: Arunachal Pradesh, Assam, Bihar, Chattisgarh, Haryana, Jammu & Kashmir, Jharkhand, Manipur, Rajasthan, Uttar Pradesh.

[1] The MoHFW data contains 35 States and UTs (that is, without Lakshadweep). Out of these, 'Daman and Diu' and 'Dadra and Nagar Haveli' have been clubbed together. Thus, we get 34 States/UTs

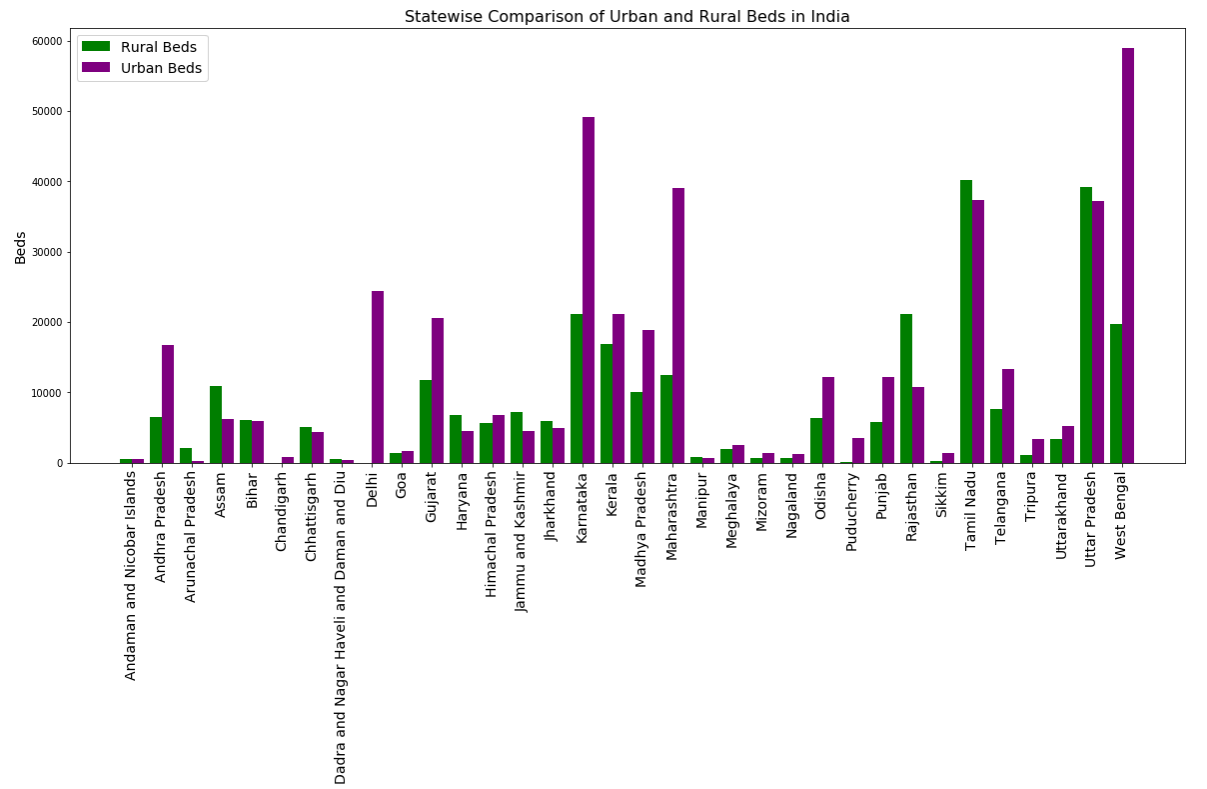

Figure 3: Combined bar graph showing statewise rural and urban beds

C. The testing centre per state graph (Figure 4) shows that Maharashtra has the highest number of testing centres in India, which may account for the higher number of cases in that the samples are being processed faster. However, it is also true that the highest number of COVID related deaths have been in this state, which raises questions about its healthcare.

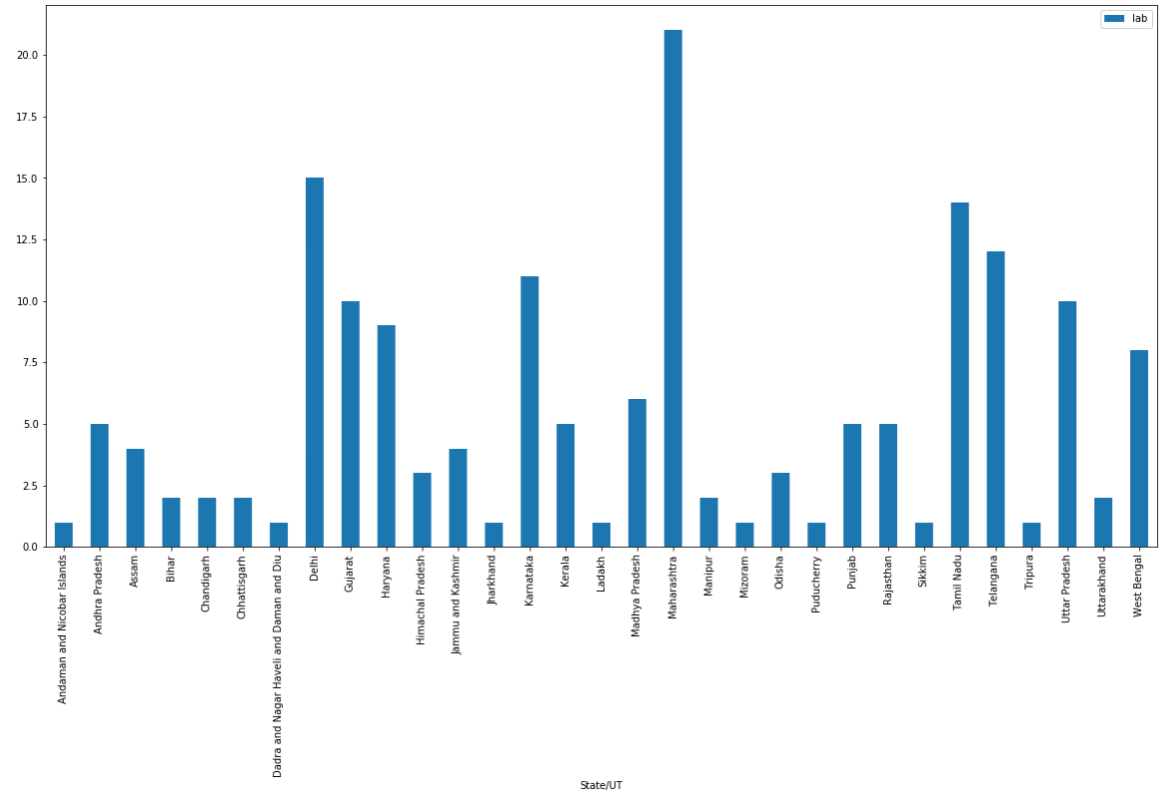

Figure 4: Bar graph showing number of testing labs per state

3.3 Google Places API and Maps

A. Figure 5 is a map showing the locations of ICMR testing labs in Maharashtra. Since clusters of them have the same latitude and longitude figures, they have been plotted in the same spot. The map gives an idea of labs being concentrated in urban areas.

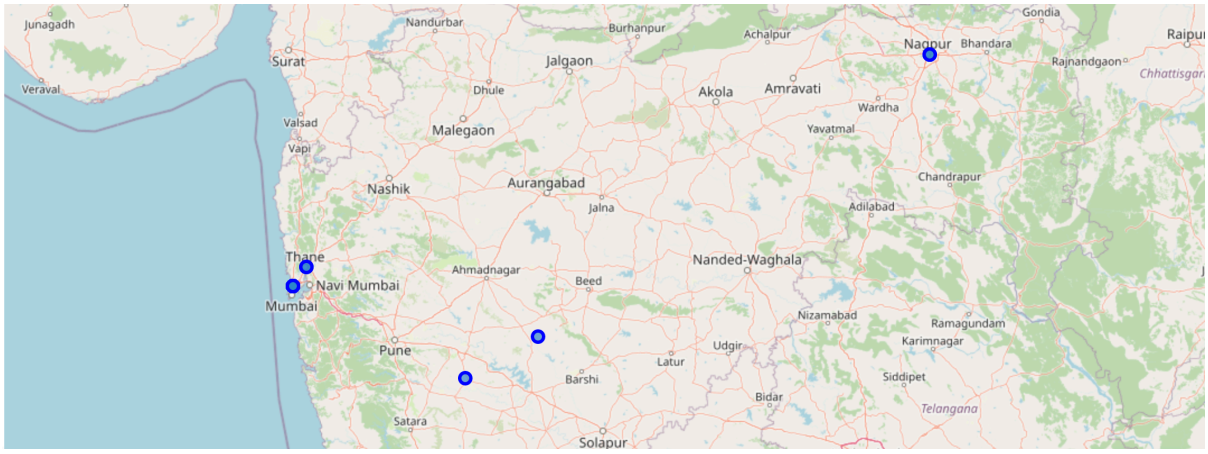

Figure 5: Map showing clustered ICMR testing labs in Maharashtra

B. Figure 6 is a map showing locations of hospitals obtained using the Google Places API in the city of Jalna, Maharashtra. Google Places API has been used to find hospitals around these locations. Folium has been used to render these locations onto a map. The image shows a cluster of hospitals in the city of Jalna.

Notably, the areas surrounding Jalna have very few hospitals, of which two clusters are visible to the south.

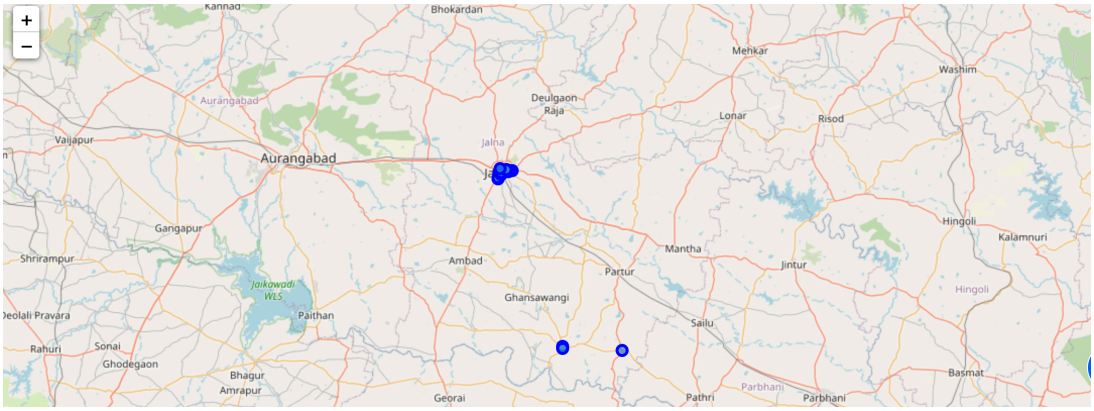

Figure 6: Map showing locations of hospitals in the city of Jalna, Maharashtra.

C. Figure 7 is a zoomed in image of Jalna to show the clustering of hospitals.

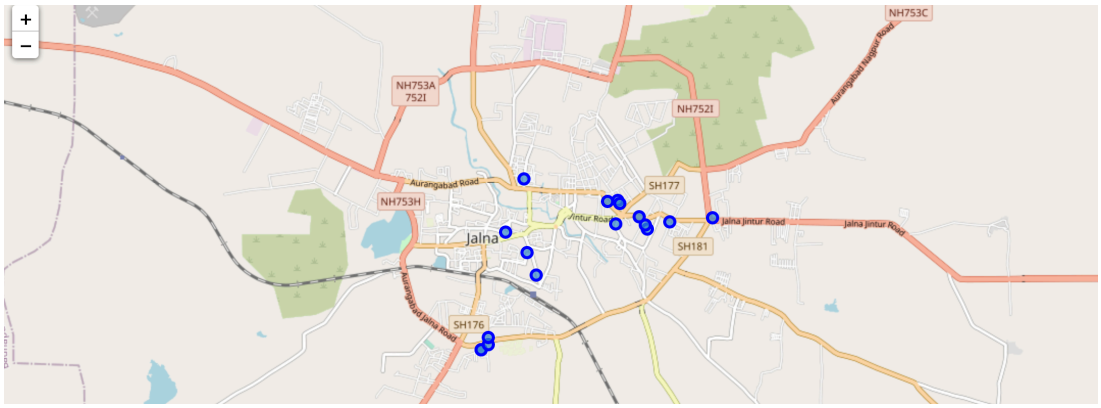

Figure 7: Map showing zoomed in image of hospitals in the city of Jalna, Maharashtra.

D. For comparison's sake, Google Places API has been used to get locations of hospitals in Mumbai. Figure 8 shows a map of Mumbai's hospitals.

.

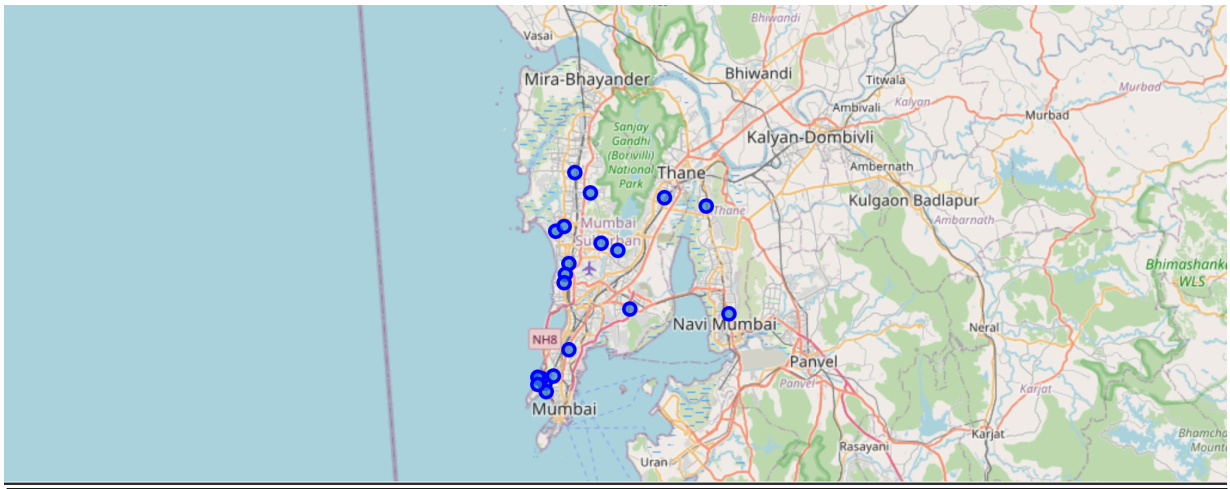

Figure 8: Map showing locations of hospitals in the city of Mumbai, Maharashtra.

4. Results

1. The relplot (Figure 1) shows that Maharashtra is spades ahead in terms of COVID-19 cases: in both active cases and deaths. This shows that the state's healthcare infrastructure needs to be investigated.
2. The bar graph comparing number of rural to urban hospitals (Figure 2) shows that Maharashtra is an exception to the norm and has more urban than rural hosptials. Also, the number of hospitals is less as compared to other states.
3. The bar graph comparing number of rural to urban beds (Figure 3) shows that Maharashtra's urban beds are more than three times its rural beds.
4. The testing centre per state shows that Maharashtra has the highest number of testing centres in India (Figure 4) and the map (Figure 5) shows that these testing centres are concentrated in urban areas like Mumbai, Thane, Pune and Nagpur.
5. When the two maps (Figures 6 and 8) are compared, it becomes apparent that hospital clusters in the rural vicinity of Jalna, hospital clusters are few. Within Jalna, the clusters increase. In Mumbai, these clusters are very high. The aim was to show a comparison between urban and rural areas. Iteratively, this process can be performed for the entire state
6. Maps plotting hospitals in Jalna (Figure 6) and Mumbai (Figure 7) show that the density of hospitals decreases as we move away from the urban centre.

5. Discussion

1. The relplot (Figure 1) shows that the number of public beds are hardly enough to host COVID-19 patients. This neither accounts for existing vacant beds nor new non-COVID patients who are afflicted by other contagious and non-contagious diseases. These diseases could be anything from cardiovascular diseases and cancers to bacterial infections, to other viral infections to accidents.
2. The fact that in most states rural hospitals are more than urban hospitals should be a positive indication (Figure 2). However, comparing Figure 2 with Figure 3, we get different results.
3. Rural beds are much less than urban beds in most states. This seems to be because urban areas have a higher population. However, this does not bode well for people in rural areas. They would have to travel to urban areas to get specialist treatment. With COVID19 in the mix, and lockdowns and locked borders in place, rural people will have more difficulty in getting treated. This also leads to people being quarantined within their homes, which increases the chances of spreading the infection if proper care is not taken. Maharashtra has almost double urban beds to rural beds. The state has nearly 39000 urban beds to 12000 rural beds.
4. The states of Arunachal Pradesh, Goa, Nagaland and Meghalaya have no ICMR labs as evident by Figure 4. This means that samples must be sent outside the state to be assessed. This not only adds to the time but also increases the susceptibility of the sample getting corrupted.
5. The maps plotted using Google Places API also indicates geospatially that the density of hospitals decreases as we move away from the urban centres.

6. Conclusion

COVID-19 has been a rude shock to humanity itself. Healthcare must be focused on and invested in to ensure that adequate care reaches everybody. Hospitals to population ratio and hospital beds to population ratio are good metrics to gauge the status of medical healthcare in a country, state, city and district. However, they do not allow us to spatially view them the way a map does. This is what this project has attempted to do with respect to the Indian state of Maharashtra, currently home to the highest number of COVID cases in India. Data science visualizations (both analytical and geospatial) can help figure out areas where more investment (in terms of finance, manpower, infrastructure) is required, such that resources can be optimally allocated.