Introduction to Geographic Information Systems

Final Project— Arjun Grover

Introduction

Research Problem: Is there any relationship between population density and the spread of the coronavirus in India?

The aim of this study is to geospatially examine high population density areas and compare them to states that have been highly impacted by the COVID-19 virus visually in India.

Using a quantile classification, we sought to analyze whether there exists a positive or negative correlation between the two sets of data (Population Density & Total Confirmed Coronavirus Cases per state) during a singular timeframe.

In this study we used population density calculated by dividing the total population per state by the area (per square kilometer) and used QGIS software to classify the data to provide us with a visual representation of high density and low-density states in India. Further, we obtained data regarding the total reported confirmed coronavirus cases per state and used GIS software to visually represent the data using the same classification as population density.

This study should be helpful in providing directional inputs for communication, legislative, social and economic policy and processes going ahead as India deals with the second wave of coronavirus pandemic and possibly more such contagious pandemics in the future. It will help governmental officials and city planners to better understand the different requirements

for different states to deal with such future ordeals. Does a state with higher population density have relatively higher number of confirmed cases?

Hypothesis: After reviewing the two GIS created maps and comparing the high-density zones and prior reading of literature in this area, it is my hypothesis that in India there is a positive relationship between population density and COVID infections. **However, such a relationship may not be very strong** owing to a variety of factors; such as the nature of how the virus spreads from person to person, the way the crisis was handled by the government or due to negligence by the people of a state who may not have followed proper safety guidelines.

During the course of this study, we prepared data tables, reference maps created using GIS software as well as an empirical analysis to confirm our hypothesis.

Methodology

Study Area

- **Country:** India

- India is a country in south-east Asia. The second most populous country in the world after China. The seventh largest country by land area and the most populous democracy in the world. India accounts for the bulk of the Indian subcontinent, lying atop the Indian tectonic plate, a part of the Indo-Australien Plate. India's coastline measures 7,517 kilometers (4,700 mi) in length; of this distance, 5,423 kilometers (3,400 mi) belongs to peninsular India and 2,094 kilometers (1,300 mi) to the Andaman, Nicobar, and Lakshadweep Island chains (both of which have been highlighted in our visual analysis). India is predominantly an agricultural economy.

- Total Number of States Surveyed: 36 (including Telangana and Jammu and Kashmir)

- As per the 2011 Census report, the total number of residents in India equals 1,210,193,422. The **total** geo-graphical **area of** the country is 3,287,240 sq.km.

- However, since we wanted to provide a more updated point of view, we have obtained current statewide population updates as per the Unique Identification Authority of India in the period of 2019-20 as part of their nationwide implementation of the AADHAR identification. The source is cited in references for your viewing.

- The updated total number of residents in 2019-20 stands at 137,05,08,600. The state-wise distribution for both population and area per square kilometer has been attached in Table 1 below (including calculated population density):

Table 1

ST_NM	Pop_Max	ABRV	AREA_SQ.KM	POP_DEN
Arunachal Pradesh	1570458	AR	83743	19
Assam	35607039	AS	78438	454
Chandigarh	1158473	CH	114	10162
Karnataka	67562686	KA	191791	352
Manipur	3091545	MN	22327	138
Meghalaya	3366710	ML	22429	150
Mizoram	1239244	MZ	21081	59
Nagaland	2249695	NL	16579	136
Punjab	30141373	PB	50362	598
Rajasthan	81032689	RJ	342239	237
Sikkim	690251	SK	7096	97
Tripura	4169794	TR	10486	398
Uttarakhand	11250858	UT	53483	210
Telangana	38510982	TG	112077	344
Bihar	124799926	BR	94163	1325
Kerala	35699443	KL	38863	919

Madhya Pradesh	85358965	MP	308252	277
Andaman & Nicobar	417036	AN	8249	51
Gujarat	63872399	GJ	196024	326
Lakshadweep	73183	LD	32	2287
Odisha	46356334	OR	155707	298
Dadra and Nagar Haveli and Daman and Diu	615724	DN	603	1021
Ladakh	289023	LA	59673	5
Jammu & Kashmir	13606320	JK	42241	322
Chhattisgarh	29436231	CG	135191	218
Delhi	18710922	DL	1483	12617
Goa	1586250	GA	3702	428
Haryana	28204692	HR	44212	638
Himachal Pradesh	7451955	HP	55673	134
Jharkhand	38593948	JH	79714	484
Tamil Nadu	77841267	TN	130058	599
Uttar Pradesh	237882725	UP	240928	987
West Bengal	99609303	WB	88752	1122
Andhra Pradesh	53903393	AP	160205	336
Puducherry	1413542	PY	479	2951
Maharashtra	123144223	MH	307713	400

COVID Impact on India

- The data collected on total confirmed cases across different states in India has been obtained from Ministry of Health and Family Welfare, India and has been accurately updated until **24 April 2021, 16:51 PM.**

- COVID-19, a worldwide pandemic originated from China. Purportedly, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); India reported its first case on 30 January 2019. As of 24th April, 2021 - India is the second most affected country after the USA with a total number of confirmed cases at 17.3 million (and counting) with over 195,000 deaths.

- A statewide distribution of total reported confirmed cases in India is presented in Table 2 below:

ST_NM	COV-04/21
Arunachal Pradesh	17430
Assam	233453
Chandigarh	37232
Karnataka	1274659
Manipur	30151
Meghalaya	15631
Mizoram	5283
Nagaland	12889
Punjab	326694
Rajasthan	483273
Sikkim	7037
Tripura	34429
Uttarakhand	142349
Telangana	387106
Bihar	378442
Kerala	1350501
Madhya Pradesh	472785
Andaman & Nicobar	5596
Gujarat	467640
Lakshadweep	1805

Odisha	394694
Dadra and Nagar Haveli and Daman and Diu	6142
Ladakh	13089
Jammu & Kashmir	156344
Chhattisgarh	622965
Delhi	980679
Goa	73644
Haryana	402843
Himachal Pradesh	84065
Jharkhand	190692
Tamil Nadu	1051487
Uttar Pradesh	1013370
West Bengal	713780
Andhra Pradesh	1009228
Puducherry	51372
Maharashtra	4161676

Table 2

GIS Methodology

Software: QGIS - a free and open-source cross-platform desktop geographic information system application that supports viewing, editing, and analysis of geospatial data.

Steps taken to perform GIS Analysis:

1. Obtained a shape file of India with state division that was rightfully updated and accepted as of 2021. This shape file includes the states administrative boundaries of Jammu & Kashmir as well as the newly formed state of Telangana.

2. Verified the source of the shape file and examined the attribute table to find that the current shapefile did not have accurate abbreviations for the individual states. Therefore, edited the attribute table by clicking the edit button —> Create New Field —> Added Abbreviations for each state —> Based on available data, specified the length and precision of each dataset so that QGIS accepts the newly inputted data —> Saved.

3. Used the Labels tab in the properties option to represent each state using the newly created abbreviations from our new field in the attribute table. Changed the font size, color and frame to suit the needs of the map.

4. Repeated Step 2 for adding Population Data (POP_MAX) and Total area per square kilometer (AREA_SQ.KM) for each state in the attribute table and finally saved the layer.

5. In excel, obtained the population density by dividing “POP_MAX” by “AREA_SQ.KM” for each state, and subsequently followed Step 2 to add a total of 4 new fields to our attribute layer.

6. Then, we obtained the base data for our first map, so we saved this layer as “Population Density India - 2019” and proceeded with the classification process.

7. Then, using QGIS tool, we used the Symbology tab and changed the data representation from “Single Symbol” to “Graduated” through data driven classification—> After initial view of our data, we tried to use both Natural Breaks as well as Quantile classification.

8. **Justification for Choosing Quantile Classification**: At first glance, when we used Natural breaks, the visual representation was extremely one class oriented, with a majority of states coming under the second bracket. Therefore, quantile classification, is useful for "spreading out" the variation that is clustered toward the low end of the scale and "collapsing" the variation at the high end. This gave us a good estimation of how population density has been spread across India by spreading out the variation clustered at the lower end as a majority of the states are at the lower end of the population density spectrum.

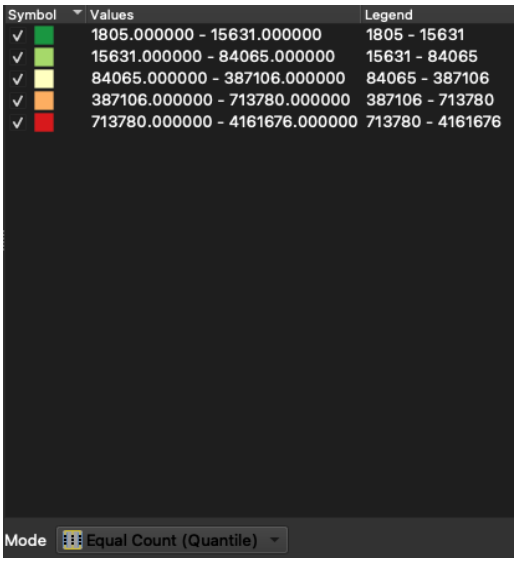
9. Once our map classification was done, we chose a yellow to deep red color spread and applied it on our map.

10. MAP MAKING PROCESS: New Print Layout —> Add Map —> Added a grid and a frame for our map —> Moved Items to center the map —> Added two other maps to provide a zoomed image of smaller Lakshadweep and Andaman & Nicobar —> Add labels for the title, source as well as the zoomed in states —> Add North Arrow and changed from default —> Add SCALEBAR - found a limitation in QGIS, as even after changing the CRS of the map layer and print layout, the scale for India was inaccurate. Therefore, decided to include a scalebar for Andaman & Nicobar only. —> Add Legend and removed the auto-update feature to specify our legend for POP_DENS —> Finally, we exported the map to an image for reference viewing in the analysis part of this study.

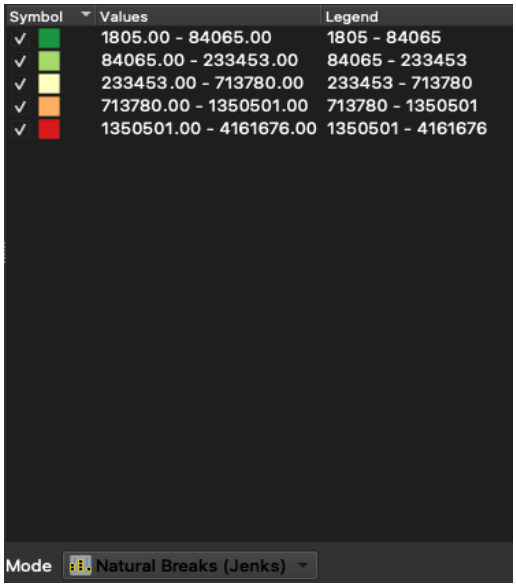
11. We followed the same process as Steps 1-10 for our second map, except this time we added COVID-19 data obtained from the Ministry of health and welfare in India. Again, we decided to use Quantile classification for the same reasons:

12. We also used a different color band to see a difference between population density and covid impact. We saved this second layer as “COVID-19 Cases India”.

Quantile



Natural Breaks



Thus, we completed our GIS analysis. The next part of this paper, presents our inferences and conclusions for both maps!

Literature Review

- Covid-19 is a highly contagious disease which has become a serious global health concern. Residents living in areas with high population density, such as big or metropolitan cities, have a higher probability to come into close contact with others and consequently any contagious disease is expected to spread rapidly in dense areas. However, recently, after analyzing Covid-19 cases in the USA researchers at the Johns Hopkins Bloomberg School of Public Health, London school of economics, and IZA—Institute of Labor Economics conclude that the spread of Covid-19 is not linked with population density (Carozzi et al, 2020). Another paper that compares population density and COVID mortality/infection rates DISTRICT wise found a moderately positive relationship (Bhadra et al, 2020).

- Therefore, our research is based on these two research articles:

1) https://www.iza.org/publications/dp/13440/urban-density-and-covid-19

2) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7553801/

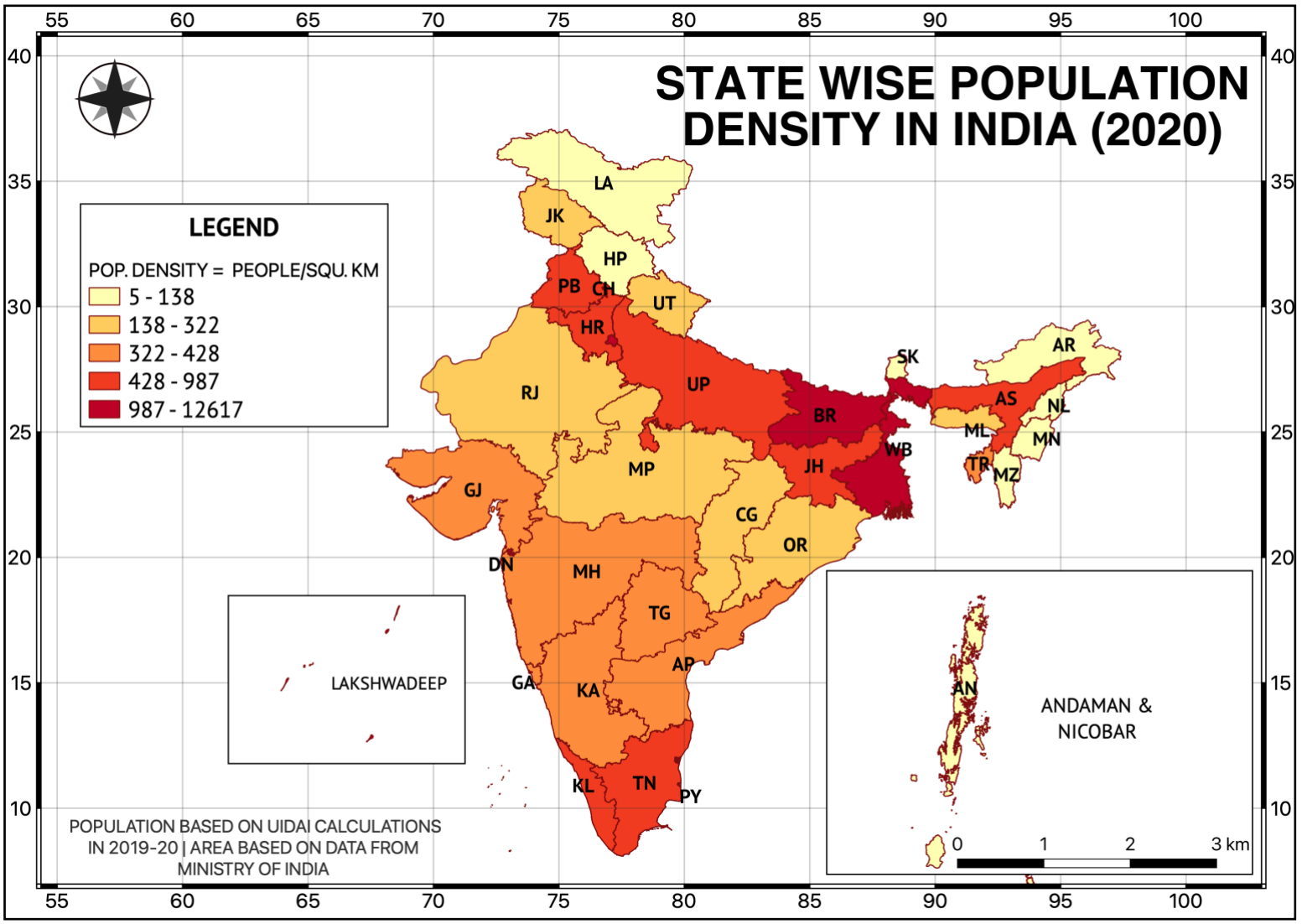
As evident from the literature review above research across geographies have indicated low to moderate correlation between population density and spread of COVID 19. Of course, these researches have been undertaken in different time bands since the onset of the pandemic. Some during and after the first wave and others during the onset of the second wave.

Our research considers an expanded time span from the date of first reporting in January 2020 until the end of April, 2021. The investigation, based on a linear and visual geo-spatial correlation between population density of a state and reported incidence of COVID seeks to confirm or reject the hypothesis that there does exist a correlation between the two factors. The eventual outcome should then enable us to delve deeper in to factors contributing to the spread of the virus.

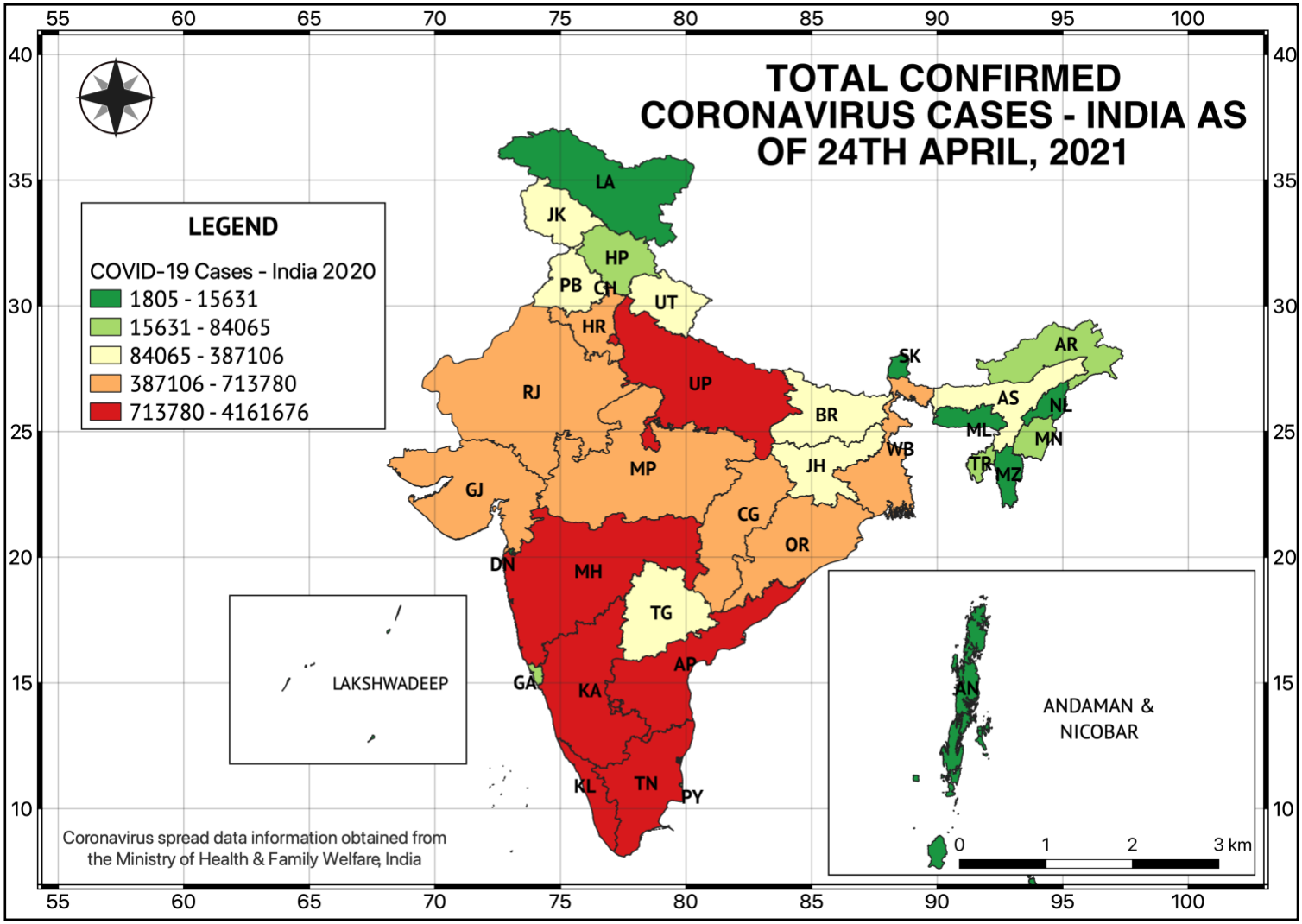
Analysis

Presented below are the maps of both population density and COVID impact across states in India. Although, both maps can be accessed here. We advise the reader to view the attached maps in PDF form for better viewing.

Map 1: Population Density



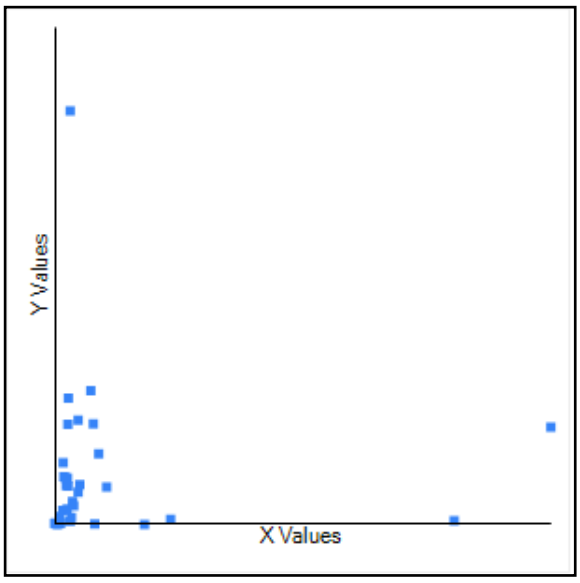
Map 2 - COVID Infection Rates



NOTE: We first used statistical methods to find a relationship between Population Density and Total Covid Confirmed Cases and reach a conclusion. We then compared our conclusion with our visual graphics and look at geo-spatial data to prove our findings and conclude with recommendations for future healthcare measures that should be taken in India.

Correlation Analysis:

Number of Observations: 36



	POP_DEN (X)	COV-04/21 (Y)
POP_DEN (X)	1	
COV-04/21 (Y)	0.0315224	1
P Value		(0.8552)

Based on these results, we estimate the correlation coefficients considering infection rates as the dependent variables and population density as the independent variable which is found as **0.0315**, indicating an **extremely weak positive correlation**. This conclusion supports the findings by the researchers at the Johns Hopkins Bloomberg School of Public Health, London school of economics, and IZA—Institute of Labor Economics**,** based on US data.

However, it is important to examine whether the association is genuine or not which can be done by considering the null hypothesis test. We applied the *F* test, which compares variances of the two variables and the t-test for the purpose and estimated the ***p* value** which essentially gives the probability that the results from the sample data occurred by chance.

Our Null Hypothesis:	**There is a minimal to no statistically significant relationship between Population Density and COVID infection rates.**
Alternate Hypothesis:	**There is a statistically significant relationship between Population Density and COVID infection rates.**
Significance Value (Alpha):	**0.05 or 95%**

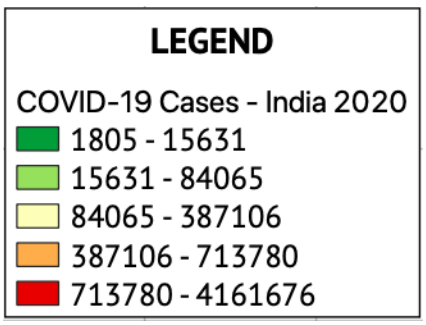
As we obtained a p-value of **0.8552** which is greater than the alpha, we can conclude that there is no statistically significant relationship between the variables. Therefore, we retain our null hypothesis and confirm that population density and COVID infection rates have minimal impact on each other.

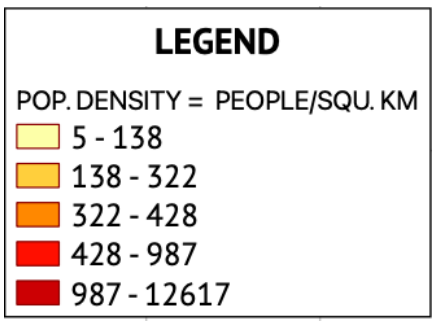
Geospatial Analysis

To confirm our results geospatially, we did a visual state by state analysis, using the maps created using GIS software as above.

As shown in the maps, we have classified our data into 5 bands each based on quantile classification. Therefore, for both population density and COVID infection rates - there is a scale of 1-5. For the sake of simplicity, we divided each comparative result into three categories:

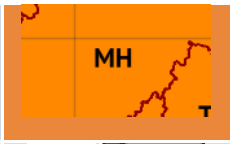
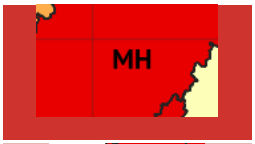
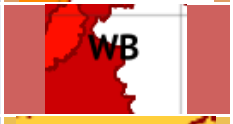
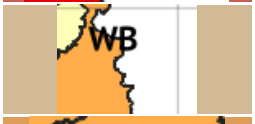
CATEGORIES	MEANING	MARGIN OF ERROR
Correctly Predicted	The band of PD is correctly predicted to be the same band of COVID Infection Rates.	0
Closely Predicted	The band of PD is closely predicted, within 2 band classifications, but not the exact same.	+2, -2





Incorrectly Predicted	The band of PD is incorrectly predicted to be the same band of COVID Infection Rates, outside of 2 MOE.	>2, <2

Top 10 highest affected in India as of 24th April

No.	STATE	OUTCOME	MOE	PD Visual	COVID Visual
1	Maharashtra	Closely Predicted	2		
2	Kerala	Closely Predicted	1		
3	Karnataka	Closely Predicted	2		
4	Tamil Nadu	Closely Predicted	1		
5	Delhi	Correctly Predicted	-		
6	Uttar Pradesh	Closely Predicted	1		
7	Andhra Pradesh	Closely Predicted	2		
8	West Bengal	Closely Predicted	-1		
9	Chattisgarh	Closely Predicted	2		
10	Rajastan	Closely Predicted	2		

Key Observations:

- 90% were closely predicted - from which 55.5% were within a margin of error of 2 for our Quantile Bands.

- 3 or 33.3% of decisions were within an MOE of 1, while one outlier was found in the Case of West Bengal, which exhibited a reverse causation of -1 (MOE), where in WB actually had the highest population density band of 5, while its COVID cases were 1 band behind.

- There were 0 cases of incorrectly predicted.

Conclusion

- Thus, these geospatial results support our statistical analysis that although a majority of observations were closely predicted, there is not enough evidence to conclude that the relationship between COVID infection rates and population density is statistically significant in India. That is, the positive correlation is somewhat weak. This is supported by the first research paper by Carrozi et al.

- This outcome is however in contrast to earlier research by Bhadra et al, where a moderately positive correlation was found.

- Some of the reasons to account for the spread of COVID 19 in India could be as follows:

  1) There is a vast difference in the living conditions of people in the USA and in India which may be responsible for the different behavior of the infected/mortality cases due to Covid-19 with population density in the two countries. The (large) density in India is reflected through the pressing of people against each other in the street, public vehicles, trains, queue for ration, etc. The average area occupied by a family for living in cities of India is also much smaller. So containing a highly infectious disease like Covid-19 is a serious challenge for a country like India. However, our results don't reflect this behavior. So, if touch could not be the main reason for the spread, we believe air and the weather to be greater factors in increasing the spread of the virus. This can be verified through the emergence of the 2nd wave, which has hit its peak during the summer heat.

2) The study by Bhadhra & Mukherjee showed a moderately positive relationship between infection rates and population density when comparing different DISTRICTS across India. Difference in units of comparison (district versus State ; Urban versus rural) may also lead to different outcomes

3) We also need to consider a widely held belief that cases in India are highly under reported due to the lack of testing , cultural resistance to seek medical attention as well as on the ground lapses in local governance

4) Population density of a city or country does not capture the finer points of how people actually gather within smaller spaces, such as those on college campuses or during Weddings or religious events. Most often when people talk about density and COVID-19, they're really talking about crowding. Crowding can also result from socioeconomic conditions that force many people to live in a small space or from cultural preferences for living in multigenerational households. Buses and other forms of mass transit can also get crowded, even in smaller urban areas. The rise in infections post the 'Kumbh Mela' and other religious gatherings, massive political rallies and onset of the wedding season further substantiate the impact of crowding rather than density in spreading the infection.

Recommendations

- The research clearly suggests that measures to contain the spread of the virus need to be dealt with at multiple levels and not as fait accompli just because India has a huge population density.

- This means that we need to strictly adhere to safety precautions, mandatory follow-ship of social distancing norms including total abstinence from overcrowding.

- Concurrently the Pandemic needs to be dealt with at central level rather than be treated as a state subject. At the apex level, the government needs to create a strong and dedicated team of scientists, doctors, statisticians, GIS specialist to monitor and map the spread of the virus: isolating specific factors that may be predominantly contributing to the infection and repeatedly projecting the likely numbers as well as possible red zone areas as feeder to the communications team.

- Furthermore, India needs a clear and comprehensive top down healthcare policy implementation in terms of prevention through social behavior and policing, rapid vaccination production and enrolment. Quick and seamless setting up of infrastructure for treatment.

- More importantly, the country needs a very detailed and thought-out communication program using influencers at various levels within the hierarchy to motivate the common man to stay safe, healthy and vaccinated.

References

- Carozzi F, Provenzano S, Roth S (2020) Urban density and COVID-19, discussion paper series, IZA Institure of Labor economics, IZA DP No. 13440

- Bhadra, A., Mukherjee, A., & Sarkar, K. (2020). Impact of population density on Covid-19 infected and mortality rate in India. *Modeling earth systems and environment*, 1–7. Advance online publication. https://doi.org/10.1007/s40808-020-00984-7

- UIDAI, G. (2020, December 31). Statewide Aadhar Saturation Overall. Retrieved from https://uidai.gov.in/images/state-wise-aadhaar-saturation.pdf

- *Corona Virus Micro Site State wise Detail page*. Zee News. (2020, April 30). https://zeenews.india.com/coronavirus-news/statewise.

- Jude N. (2021, January 4). *Choropleth Maps - A Guide to Data Classification*. GIS Geography. https://gisgeography.com/choropleth-maps-data-classification/.

- *India*. Worldometer. (n.d.). https://www.worldometers.info/coronavirus/country/india/.

Appendix

Note - Kindly refer to the attached Excel Sheet for complete data calculations.

Correlation Calculation

POP_DEN (X VALUE)	COV-04/21 (Y VALUE)	$X - M_x$	$Y - M_y$	$(X - M_x)^2$	$(X - M_x)(Y - M_y)$
19	17430	-1122.639	-443971.53	1260318.08	498419702.6
454	233453	-687.639	-227948.53	472847.242	156746272.4
10162	37232	9020.361	-424169.53	81366914.6	-3826162313
352	1274659	-789.639	813257.472	623529.575	-642179726.7
138	30151	-1003.639	-431250.53	1007291.02	432819800.5
150	15631	-991.639	-445770.53	983347.686	442043390.9
59	5283	-1082.639	-456118.53	1172106.96	493811656.1
136	12889	-1005.639	-448512.53	1011309.58	451041640.1
598	326694	-543.639	-134707.53	295543.242	73232250.73
237	483273	-904.639	21871.472	818371.519	-19785784.33
97	7037	-1044.639	-454364.53	1091270.41	474646855.4
398	34429	-743.639	-426972.53	552998.797	317513376.1
210	142349	-931.639	-319052.53	867951.019	297241742.5
344	387106	-797.639	-74295.528	636227.797	59261002.23
1325	378442	183.361	-82959.528	33621.297	-15211551.19
919	1350501	-222.639	889099.472	49568.075	-197948118.6
277	472785	-864.639	11383.472	747600.408	-9842592.774
51	5596	-1090.639	-455805.53	1189493.19	497119234.4
326	467640	-815.639	6238.472	665266.797	-5088340.552
2287	1805	1145.361	-459596.53	1311852.08	-526403989.7

298	394694	-843.639	-66707.528	711726.575	56277064.62
1021	6142	-120.639	-455259.53	14553.742	54922003.59
5	13089	-1136.639	-448312.53	1291947.96	509569453.4
322	156344	-819.639	-305057.53	671807.908	250037013.1
218	622965	-923.639	161563.472	853108.797	-149226306
12617	980679	11475.361	519277.472	131683913	5958896511
428	73644	-713.639	-387757.53	509280.464	276718851.3
638	402843	-503.639	-58558.528	253652.13	29492351.87
134	84065	-1007.639	-377336.53	1015336.13	380218959.6
484	190692	-657.639	-270709.53	432488.908	178029113.1
599	1051487	-542.639	590085.472	294456.964	-320203325
987	1013370	-154.639	551968.472	23913.186	-85355791.25
1122	713780	-19.639	252378.472	385.686	-4956432.774
336	1009228	-805.639	547826.472	649054.019	-441350310.4
2951	51372	1809.361	-410029.53	3273787.63	-741891482
400	4161676	-741.639	3700274.47	550028.242	-2744267448
		Mx: 1141.639	My: 461401.528	Sum: 238386870.306	Sum: 2158184732.861