

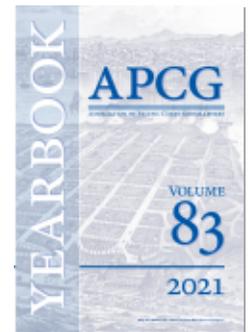


PROJECT MUSE®

COVID-19 in Los Angeles: A Multivariate Analysis of Disease
Infection Rates

Steven M. Graves, Petra Nichols

Yearbook of the Association of Pacific Coast Geographers, Volume 83,
2021, pp. 64-80 (Article)



Published by University of Hawai'i Press

DOI: <https://doi.org/10.1353/pcg.2021.0008>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/841796>

COVID-19 in Los Angeles: A Multivariate Analysis of Disease Infection Rates

STEVEN M. GRAVES
California State University, Northridge

PETRA NICHOLS
California State University, Northridge

ABSTRACT

By the time the World Health Organization declared COVID-19 a pandemic in March 2020, Los Angeles was already showing troubling signs of what was to come. Cities with international airports, significant tourist economies, and ample jet-setting populations quickly ascended to the top of case and morbidity counts. Los Angeles County, with its exposure to international tourism and massive population, easily outpaced all counterparts in the U.S. for both cases and deaths. Los Angeles County's vast spatial extent, enormous population, complex ethnic diversity, and deep economic disparities make it an ideal laboratory for the study of human behaviors. The ample and trustworthy COVID-19 data make it an excellent location for statistical modeling of infection rates. Our neighborhood-level analysis offers a powerful perspective into the causal associations that county- or state-level analyses cannot. Using COVID-19 case rate data from the LA Public Health Department and 2018 Census block group data, we constructed a series of statistical models measuring the association between COVID-19 infection rates, ethnicity, income, housing, household density, and a host of socioeconomic variables. Our exceptionally robust data model (Adjusted $R^2 = 0.93$) demonstrates that neighborhood housing characteristics were the most statistically significant factor associated with elevated neighborhood case rates, followed by income and ethnicity (percent Hispanic and Asian) characteristics. Implications for both public policy and methodology are discussed.

Keywords: COVID-19, pandemic, Los Angeles, GIS, regression

WHEN THE VENERABLE *Forbes* magazine ran an article titled “Why Geography Is a Key Part of Fighting the COVID-19 Coronavirus Outbreak” (Shepherd 2020) early in the pandemic, they reminded the public of something geographers and epidemiologists have known since the days of John

Snow. Not only have geographers regularly contributed to the battle against COVID-19, but the spatial nature of the disease and the unusual availability of rich datasets have provided geographers with a bounty of research opportunities capable of shedding light on broader questions concerning a host of human behaviors and conditions. Los Angeles County's unique geography and its diversity of income, housing, and ethnic neighborhoods provided an especially fruitful case study into COVID-19 infection rates and their causes. By creating several dozen large neighborhood units of analysis and subjecting them to ordinary least squares regression testing, we found very strong evidence that family size and housing characteristics had very robust associations with infection rates, while income and ethnic characteristics also had statistically significant associations with infection rates. The public policy implications are multiple but challenging. Geography is indeed a key part of the arsenal of defense against COVID-19.

Background

There is no shortage of research on the spatial aspects of infectious diseases. Medical geographers offer textbooks with full chapters on the subject (see e.g., Meade and Emch 2010). Real and Biek note that the "...spatial analysis of infectious disease processes recognizes that host-pathogen interactions occur in specific locations at specific times, and that often the nature, direction, intensity, and outcome of specific interactions depend upon the specific location and identity of both host and pathogen" (2007, 935). COVID-19 inspired several journals to publish special volumes dedicated to Coronavirus research, and several review articles have appeared as well. The edition published by the Royal Dutch Geographical Society (Aalbers, Beerepoot, and Martijn 2020) included more than twenty-five articles on a wide range of responses to and consequences of the outbreak. The special edition included an equally wide range of theoretical and topical approaches to the study of the pandemic across the globe. Franch-Pardo, et al. (2020) reviewed sixty-three articles that examined and/or analyzed the various aspects of the pandemic, grouping the research into five broad categories: Spatiotemporal analysis, Health and Social Geography, Environment Variables, Data Mining, and Web-based Mapping. UNAM also produced a special COVID-19 volume in their online cartography journal *terra digitalis*, which featured a series of interactive maps of infection rates and vulnerability at various scales in the U.S. and Mexico (see e.g., Graves and He 2020).

We drew inspiration and methodological guidance from articles that focused on data modeling of the socioeconomic factors associated with the spread of the disease. An early unpublished paper by data modelers from MIT (Wheaton and Thompson 2020) found that after two months of the first known case of COVID-19 in the United States, they could model the spread of the disease at the county level with some confidence. They found, not surprisingly, that case numbers tracked proportionately with population, but also that residential density and per capita income were related at the county level. They found weaker relationships with income, housing, and age variables at the MSA level. Wheaton and Thompson (2020) note a surprising positive relationship was found between income and case rates. We noticed a similar association early in the pandemic in Los Angeles County but found that it reversed with time. Our speculation is that wealthier Los Angelenos were exposed earlier to the virus through international travel and visits to ski resorts in the early months of 2020, when the pandemic was in its early stages. We found that as 2020 wore on, Los Angelenos living in high-income neighborhoods were, by virtue of their occupation or reserved wealth, apparently able to shelter at home away from others, allowing infection rates to plunge in wealthier neighborhoods while rates soared in poorer parts of Los Angeles.

Another study analyzed the effect of social distancing behaviors on the spread of COVID-19 at the state level using data from the IPUMS American Time Use Survey and the COVID Tracking Project during the early months of the pandemic (Price and von Holm 2020). Based on the results of their data model, Price and von Holm conclude that social distancing had a significant effect on the spread of the disease, and they offered policy suggestions to save lives. They also found “infections increase with the percentage of a state’s population that is African American, foreign born, self-employed, and living in metropolitan areas/cities...” (2020, 8). Their findings also provided us with guidance toward likely covariates to use in our model. Mollalo, Vahedi, and Rivera (2020) used a battery of regression techniques to model a spectrum of thirty-five variables on infection rates at the county level for the continental United States in early April of 2020. They found that income, income inequality, ethnicity, and health care provision were highly associated with infection rates. Also of note, they found the OLS model, which we employ in this study, a misspecified model at the county level. Like many others, they chose to use sophisticated Multiscale Geographically Weighted Regression, rather than reconsidering their scale

of analysis or a modification of the units of analysis. Similar conclusions were reached by Sun, Matthews, and Yang (2020) who also used county-level infection rates from June 2020 as their dependent variable against a battery of demographic independent variables in their OLS and various geographically weighted models. They concluded that all the models offered insight.

We argue that most of the early research on COVID-19 suffered from an incomplete record of cases. In June 2020, many places had not yet experienced the first wave of the pandemic, while other places, more connected to international travel and ski resorts, etc. had already experienced a first wave of infections, skewing the data. This was true at both micro and macro levels of analysis. Zalla et al. (2021) argue, as do we, that structural racism is a key element in the risk factors associated with both COVID-19 infections and deaths among U.S. citizens. However, we disagree with some of their policy recommendations regarding the reporting practices of health departments. They claim that the weighting formulas used by various health agencies create a tendency to underestimate racial disparities in the rate of infection. We believe they miss the mark by discouraging the practice of “controlling for geography.” We argue instead that the problem is the *scale* of analysis, not the use of geographically based units of analysis. County- and state-level analyses tend to hide the effect of ethnicity, as well as the effects of income, housing, and employment on infection rates. Micro-level analyses of data, as we conduct in this study, not only demonstrate far more clearly the disproportionate burden COVID-19 has placed on communities of color, but also expose quite clearly the statistically significant roles various aspects of racism (employment, housing, income, etc.) have on the disproportionate rates of infection. We also argue that a variety of statistical approaches to significant social and medical problems is the ideal approach because there is value in multiple measures of infection and racial disparity.

Data quality and the politicization of data have concerned many researchers. African researchers examined the effect of government measures on the differing rates of infection and transmission, as well as the economic impact of restrictions on business and travel (Asongu, Diop, and Nnanna 2020). Published in July 2020, their study found that lockdown measures were beneficial in Europe but not so much elsewhere, except in part the Americas, where mobility restrictions had slowed the spread. With the benefit of hindsight, it is easy to question the results of this early study because the major spike in infections in the Americas and elsewhere had yet to materialize. More importantly, the occasionally poor quality of infection

and death rate data was perhaps not as evident in 2020. The April 2021 spike in infections that ravaged India highlighted the problem of data collection and reliability in the developing world, where isolated rural populations and vast urban populations both proved difficult for overstretched public health officials to monitor. The politicization of many aspects of the COVID-19 pandemic created numerous challenges for the medical community, medical geographers, and other epidemiologists (see e.g., Schragger 2020). The peculiar political climate of 2020, an election year fraught with misinformation campaigns, added hard-to-estimate political variables at the state and county level in the United States. Governor DeSantis of Florida was accused of ordering officials to misdirect the public on the extent of disease in Florida. One GIS analyst for the state was fired, ostensibly, for claiming the official numbers published by the state were falsified (Smith 2021). Officials from other states and counties across the United States have been accused on social media and in the popular press of manipulating public COVID-19 data for political and/or economic reasons. Though it is difficult to prove or reject such accusations, numerous data dashboards, including those created by *The New York Times* and Johns Hopkins University regularly highlighted dramatic difference in infection rates in neighboring counties situated across state lines. These differences were sometimes starkly evident even in locations where no physical, transportation, or political barriers between neighboring states could possibly explain the abrupt and consistent shift in infection rates across a highly permeable political boundary. This suggests data quality problems. In addition to the suspect reliability of public data, a hodgepodge application of hygiene mandates and policy guidance on fighting the infection were unevenly proscribed for schools and businesses across the country, making statistical analyses of their effects on COVID-19 infection rates a major challenge.

Many of the papers we reviewed seemed to suffer from an uncritical examination of the unit of analysis. Compounding the concerns about data reliability and the obscuring effects of the uneven policy measures, we concluded that states and most counties are generally too internally heterogeneous to be used as the basis for the type of analysis that seeks to understand the effects of income, ethnicity, and housing on infection rates. This is especially true of urban counties, such as Los Angeles, where the population is exceptionally large and heterogeneous.

In many ways, Los Angeles County is an ideal laboratory for analyses of a variety of phenomena. Home to roughly ten million people, it is nearly

as populous as the state of Ohio. By some measures, it is the most ethnically diverse county in the United States, thereby providing an excellent opportunity to study the effects of culture, ethnicity, and racism on a variety of phenomena. Los Angeles County has a steep wealth gradient and vast income inequality as well that expresses itself unevenly in a diversity of residential arrangements. Rural districts in the far northern corners of the county have extremely low population densities. Industrial districts and wealthy enclaves in the Santa Monica Mountains also have low population densities, but in parts of downtown, some neighborhoods have densities over 40,000 persons per square mile; it is like a small country. The only analytically useful element Los Angeles truly lacks is a large region of poor, non-Hispanic Whites.

Data and Methods

The Los Angeles County, Pasadena, and Long Beach health departments published case data by neighborhood (N = 350) daily throughout the pandemic (County of Los Angeles Public Health 2021). The *Los Angeles Times* collected and collated data from the Pasadena, Long Beach, and Los Angeles County health departments and served it via GitHub to the public. We assume the data is of good quality, and to our knowledge there were no credible accusations of data tampering for political gain. Our nearly daily inspection of the data and maps during the study period suggested a high degree of data consistency across neighborhoods and health departments. The health departments made great efforts to provide timely, transparent, and accessible data. They also provided occasional corrections to mis-entered data, even when the news was not good. We did not see substantial “cross border” inconsistencies between the data provided by the three health agencies as were evident elsewhere. Therefore, we are confident in both the accuracy and reliability of the data collection.

The *Los Angeles Times* (2021) also provided via GitHub a digital map of 346 “places” in the county, a close copy of the one used by the Los Angeles Public Health Department and other Los Angeles County governmental agencies. These places ranged in size from neighborhoods that were around a few thousand square meters (e.g., Little Tokyo) to the massive Angeles National Forest. Some places in the database were municipalities (e.g., Burbank, Inglewood); others were small “vanity neighborhood” designations scattered across the City of Los Angeles (e.g., Wellington Square).

We mapped the COVID-19 infection rates from the 350 neighborhoods reported by the *Los Angeles Times* using the provided coordinates onto the

map of 346 neighborhood polygons. Several “neighborhoods” had no data throughout the course of the study. Because many of the 350 places were very small, we assumed cases occurring in these micro-neighborhoods were included in the reported cases of adjacent neighborhoods or cities, and thus included their case counts in the larger, encompassing neighborhood polygons. This was the first step in an iterative process of integrating smaller neighborhoods into larger ones. Our next step was to merge places with fewer than 1,000 persons into larger, adjacent or encompassing places or neighborhoods. Most of these small neighborhoods were tiny enclaves, vanity designations, or simply unincorporated portions of an adjacent city that could be logically combined into their incorporated neighbors. This process reduced the number of neighborhoods in our study to ninety.

The preliminary map with these ninety “super neighborhoods” permitted an exploratory round of data modeling using ordinary least squares regression in ArcPro GIS. In each of our super neighborhoods in Los Angeles County, we calculated COVID-19 rates of infection, as well as multiple candidate variables, each of which were variants of measures of ethnicity, income, family structure, housing, and educational attainment characteristics, as suggested by our review of existing literature on COVID-19 and similar infectious diseases. For example, we tried both median household income and median disposable income as candidate variables in our efforts to refine the model. A rate of infection per 100,000 persons was calculated for each regional neighborhood using the 2018 population estimates from ESRI/US Census. The population figures supplied by the *Los Angeles Times*/Los Angeles County Health Department were estimated to be slightly dated or otherwise less accurate. These preliminary models allowed us to evaluate promising candidate variables, and via a stepwise process, to narrow candidates to a list of roughly six variables that were both logically associated with COVID-19 rates and statistically viable for use in subsequent models.

Although the preliminary models regularly produced promising results, the models all suffered from levels of spatial autocorrelation among the model residuals exceeding the 1.96 threshold, according to Moran’s I tests. The z-score associated with the Moran’s I test for our base model (ninety neighborhoods, Percent Hispanic, Median Household Income, Average Household Size) using the traditional “Queen’s Contiguity” rules was 4.96. Excessive spatial autocorrelation violates the assumption of independence that is built into regression models. We anticipated this problem because it commonly affects linear regression models in urban settings when the

model uses small areal units of analysis (e.g., census block groups). The outcome is predictable because people travel regularly and freely across the invisible and often randomly created boundaries of an urban area for work, shopping, etc., taking with them behaviors and infectious diseases, while their demographic characteristics remain tied to their place of residence. Spatial autocorrelation tends to decrease as the size of the units of analysis increases by decreasing the likelihood people venture across boundaries, carrying with them their behaviors and, in this case, their infectious disease.

Spatial autocorrelation is associated with the Modifiable Areal Unit Problem or MAUP (see, e.g., Dark and Bram 2007 or Openshaw 1977), and it characterizes the problem arbitrary boundaries create for those trying to analyze a host of geographic questions. A common and statistically effective option to reduce excessive spatial autocorrelation among residuals is to use one of the geographically weighted regression models, such as those reviewed in the background section above. However, for important policy matters, where it is essential to communicate research findings to the media and policy makers with a minimum of statistical jargon and caveats, we prefer the simplicity of well-known, common, linear regression models. Therefore, we chose to attempt to treat the MAUP as a modifiable areal *solution* rather than to see it as only a problem.

To reduce the level of spatial autocorrelation to acceptable levels, it was necessary to merge small neighborhoods with adjacent ones to form units of analysis that were larger. At the same time, we sought to create units of analysis that were internally homogenous to avoid diminishing our ability to measure the effect of test variables, particularly ethnicity and income. To create fewer units of analysis that were somewhat internally homogenous, we merged data from an additional twenty-one neighborhoods with adjacent neighborhoods. The selection process was straightforward, but admittedly included a measure of subjectivity. Neighborhoods selected for merging were selected based on the following criteria: (1) areal size; (2) population; (3) income; (4) ethnicity; and (5) physical geography and/or infrastructure. Neighborhoods with small populations and/or small areal extent were given priority to be merged with an adjacent neighborhood because people living in small neighborhoods are most likely to travel across neighborhood borders to work, shop, or attend school in an adjacent region. If multiple adjacent neighborhoods had similar ethnic and income profiles, multiple neighborhoods were merged at once. If a significant difference in ethnicity (e.g., percent Asian) or income was evident, those regions were not merged

even if they were small. So, for example, Chinatown was not merged with any adjacent neighborhood because it was ethnically distinct from its neighbors despite its size, but El Monte and South El Monte, which had similar ethnicity and income profiles, were merged into a single unit of analysis though they were much larger. We avoided combining adjacent regions that were separated by a physical barrier, such as the Verdugo Mountains or Palos Verdes Hills, because such barriers restrict travel and contact between populations.

By merging several adjacent neighborhoods into fewer units of analysis, avoiding where possible any merging of neighborhoods with sharp differences in ethnicity or income, we created a map of sixty-nine super neighborhoods. Many were municipal units, such as Burbank, Carson, or Santa Monica. By reducing the number of units of analysis, we created regions with increased internal homogeneity. This permitted the modeling of data using common linear regression (OLS) models that were robust yet lacked excessive spatial autocorrelation among the residuals. We believe this reduction in the number of neighborhoods included in the study offers a more logical division of Los Angeles County than the jigsaw puzzle of unincorporated regions, micro-municipalities, political enclaves, and vanity neighborhoods. The resultant units of analysis represent one possible reconfiguration of Los Angeles County into regions that feature some degree of internally homogeneity and, we maintain, reflect a more reasonable approximation of “neighborhoods” as they are perceived and experienced by the residents of Los Angeles County. Drawing and redrawing polygons that reflect the lived experience of residents, or “communities of interest,” is a challenging task that vexes geographers seeking to identify and delineate natural communities suitable for election redistricting.

Preliminary models with the reduced number of neighborhoods revealed three substantial outliers that were withdrawn from the model: Catalina Island, Downtown Los Angeles, and an industrial region called Bandini Islands. Catalina Island was eliminated because of the limited interaction it had with the mainland. Downtown Los Angeles and Bandini Islands appeared to suffer from data inaccuracies in the US Census data that appear to have made their case rates far exceed what we believed were possible. We suspect that the population of the Downtown is far larger than reported by the census, probably a consequence of the exceptionally large, unhoused population in the region. Also, because housing conditions are a key element of our model, including many thousands of unhoused persons living on or near Los Angeles’ Skid Row made little sense. Future models

may include a measure of homelessness. We suspect too that the Bandini Islands neighborhood had far more households and/or persons than what our database of US Census data indicated. We suspect a glitch in our census figures or in the map of the region provided by the County of Los Angeles/ *Los Angeles Times*.

The final model used COVID-19 case data for May 15, 2021, a date just after Los Angeles County entered a period of the least restrictive tier of “lockdown.” This date was also marked by a relatively even distribution of infection rates across the county, a suggestion, at the time, that the pandemic had largely run its course throughout the county. Another wave of infections erupted in July 2021 due in part to the arrival of the so-called Delta Variant, which invites additional study and perhaps the inclusion of new variables that indicate rates of vaccine hesitancy or political affiliation. Our final Ordinary Least Squares model used a combination of variables that yielded a close fit of estimates to the calculated infection rates that we found surprising (Adjusted $R^2 = 0.93$). The remaining model diagnostics, including the Jarque-Bera value and Moran’s I value for residual spatial autocorrelation, came back within acceptable parameters, suggesting a robust and stable specification for the model. Multicollinearity among the independent variables was likewise below the diagnostic threshold for concern ($VIF < 5.0$). Notable among the discarded variables was percent persons with a bachelor’s degree because it was repeatedly flagged for excessive multicollinearity with income and housing variables. In the end, our model included six variables with significant association with COVID-19 rates of infection in Los Angeles County.

Discussion

Rates of COVID-19 infection in Los Angeles County are strongly associated with five or six easily calculated variables. The most robust association was with Mean Household Size (Table 1). This finding is logical because exposure to an infectious disease would be most extreme for people living under the same roof. Exposure to COVID-19 would, in many cases, increase exponentially if those household members regularly left a shared dwelling for work, play, or other activities where the dangers of community spread are manifest. Two related income variables, Median Household Income, and the percent of rental housing in each neighborhood, also had robust statistical association with COVID-19 infection rates. Income is a proxy measure of a family’s ability to shelter-in-place, away from the public dur-

Table 1. OLS Coefficient Table

Variable	Coefficient	Standard Error	t-Statistic	Probability	Robust Std. Err.	Robust - t	Robust Prob.	VIF
Intercept	-175.451916	2101.496315	-0.08348	0.933743	2663.592593	-0.065870	0.947702	-----
Mean Median Household Income	-0.037101	0.011570	-3.20682	0.002171*	0.013090	-2.834394	0.006277*	4.356166
Percent Hispanic	4752.991782	1406.386225	3.379578	0.001295*	1621.336164	2.931528	0.004794*	4.266980
Percent Asian	-4087.53108	1142.203419	-3.57863	0.000701*	958.939723	-4.26255	0.000075*	1.301972
Mean Household Size	3212.230512	351.805921	9.130689	0.000000*	477.960936	6.720697	0.000000*	4.767042
Percent Living in Group Quarters	-6647.86683	3600.917117	-1.84616	0.069890	5267.212876	-1.26212	0.211872	4.845415
Percent Living in Rental Housing	4779.058174	1387.167010	3.445193	0.001060*	1615.929463	2.957467	0.004457*	2.588527

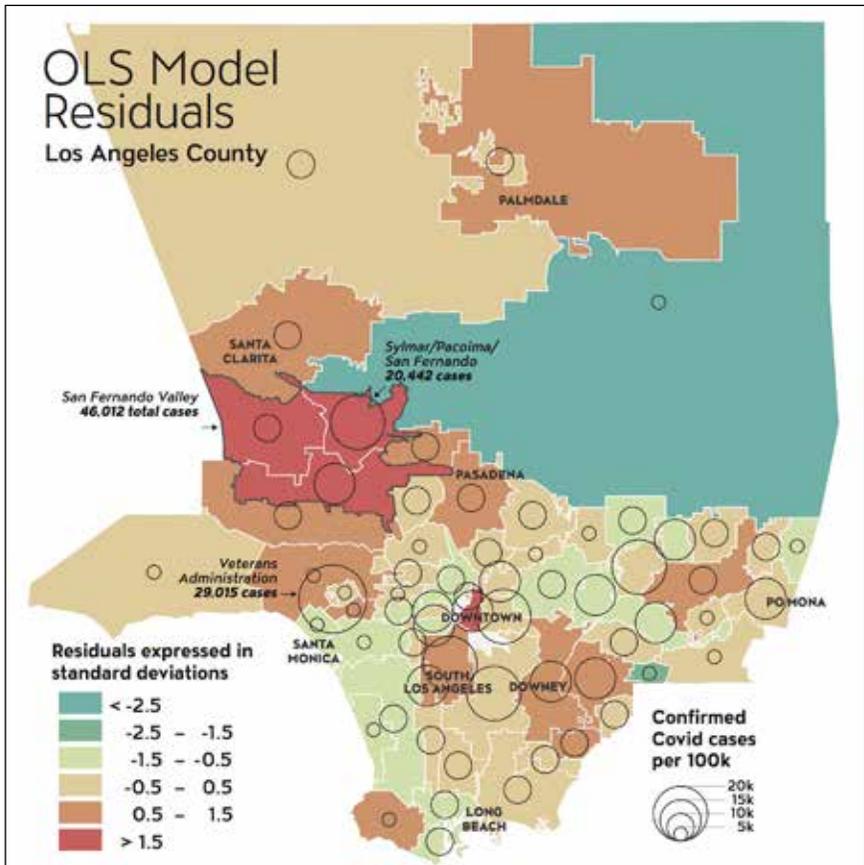
Adjusted R² = 0.93

ing the pandemic. Presumably, wealthier Los Angelenos were able to stay at home and work from home while avoiding public places (for shopping, etc.) during the pandemic, reducing their exposure to the virus. Because so many housing units, especially in poorer neighborhoods, are multi-family dwellings, the percent renter variable also probably captures some of the dangers of exposure to an airborne pathogen in common hallways, air circulation systems, etc. that are characteristic of such dwellings. A third housing variable, “Percent of Persons Living in Group Quarters,” was also significant at the .10 level, indicating the small but important effect group homes, prisons, nursing facilities, etc. had on the neighborhood infection rates. Had such facilities been more common, and more evenly distributed throughout the county, we suspect that the group home variable would have shown an even stronger association.

Finally, we identified two ethnicity variables with robust associations with COVID-19 infection rates. The variable Percent Asian, despite its troublesome breadth of inclusion, was inversely associated with COVID-19 infection rates. We suspect that the inverse relationship may be a byproduct of the rapid adoption of hygiene protocols, especially mask wearing, by the largest ethnic groups that contribute to the broad category “Asian” in Los Angeles County. Even before the current pandemic, it was not uncommon to see Los Angelenos of Chinese, Korean, and Japanese descent wearing masks during flu season. This finding does not echo other studies, many conducted earlier in the pandemic, that note elevated case and death rates among Asians (see e.g., Yee 2021). Our model very closely estimated the case rate in Chinatown, an ethnic enclave that was *not* merged into any adjacent regional neighborhood. The model fit was also reasonably close in the Asian ethnoburbs of the San Gabriel Valley (Figure 1).

The percentage of Hispanics of all racial identifications had a significant, positive association with COVID-19 infection rates, even after controlling for income and housing characteristics. We hesitate to speculate as to why this is so, but a blunt opinion piece in *The Los Angeles Times* penned by Gustavo Arrelano, a long-time resident of Los Angeles and a son of Mexican immigrants, speculated that part of the reason for the elevated case rate among Latinos was a reluctance among Latinos to forego important communal gatherings, family celebrations, and practices that put Latinx people at risk during the pandemic. Also, it is very logical that so much of Los Angeles’ essential workforce is comprised of people of Latin American heritage. They worked, often without interruption, during the entirety of the harsh

lockdown period (Arrelano 2020). Arleta, Pacoima, and San Fernando were the hardest-hit neighborhoods in Los Angeles. Those three were among six (of sixty-nine) neighborhoods with an infection rate surpassing twenty percent. Still, the map of model residuals indicates that the Northeastern San Fernando Valley (Arleta, Pacoima, San Fernando, etc.) was the most severely underestimated region in our model. In the rest of the San Fernando Valley, the map of residuals indicates small but consistent under-prediction of the case rates in adjacent regions, including those regions of the Valley with high incomes and low infection rates. This clearly points to the effect of proximity to vulnerable neighborhoods to residents in the more exclusive neighborhoods. Similar conditions and dynamics were evident in East LA /Boyle



Heights as well as Vernon/Florence-Firestone neighborhoods, although the model better predicted the rates of infection in the Los Angeles Basin.

The map of model residuals is also an effective aid for inquiry into further research. The regions where the model underpredicted infection rates, particularly the beach communities stretching from Santa Monica to the Redondo Beach as well as the vast region of the northwestern county, suggest several avenues for future research and model building. One possible missing variable that might prove worthy of including in future models would be the effect of climate and wind. Cooler and breezier conditions common near the beach, in the mountains, and in the Mojave may have suppressed infection rates by a small amount. The percent of non-Hispanic White people in these same regions also may have helped suppress infection rates, though in our models, percent White consistently suffered from excessive multi-collinearity with percent Hispanic and various income variables. More-sophisticated models, or clever indices of ethnic identification, may produce a statistically viable way to include percent White in a model alongside percent Hispanic.

The percent of Black people was also excluded from the model because it was generally statistically insignificant after controlling for income. Indeed, some of the lowest infection rates in Los Angeles County were in the wealthy African American enclaves in and around View Park and Windsor Hills. Other portions of the Crenshaw District and Inglewood also had moderate infection rates. We acknowledge that in ethnically integrated neighborhoods, especially those with heterogenous income levels, our model hides some of the burden of disease. This concern was raised by Zalla, et al. (2021). Our methodology was designed to minimize that tendency, but we admit that small “pocket neighborhoods” with specific conditions or ethnic characteristics may be washed out by values of neighboring communities within a neighborhood. On the other hand, it is possible that microethnic communities may adopt some of the cultural and economic practices of their more numerous neighbors, and as a result blend rather seamlessly into the larger neighborhood context.

Better data regarding specific categories of employment may also produce interesting results. Our attempts to include employment variables in our model were unsuccessful. Low significance levels were the most common issue. We suspect it is because the categories available to us were too broad. Different approaches to measuring the impact of employment are warranted. Other likely variables we would like to see modeled include things like pro-

pensity to follow public health guidelines, political affiliations and belief, vaccine hesitancy, and measures of access to health care and testing facilities.

The implications for this research are numerous and serious. While we can hope for another century-long respite from a global pandemic that we had enjoyed prior to 2019, we cannot expect to be so lucky. The infiltration of wilderness regions that were formerly remote and the increase in international travel increase the likelihood for novel viruses to infect humans and the likelihood of rapid, global spread. Clearly, public health officials should be preparing for the next pandemic now as the lessons of the continuing COVID-19 crisis are manifesting themselves in the consciousness of political leaders and policy makers both inside and outside government. Knowing where future diseases are likely to be most severe and having proactive plans in place to help the most vulnerable will be key to preventing widespread illness and death from occurring again. Critical to any effective plan will be proactive measures to help vulnerable populations who live in crowded housing to find temporary housing, or at the very least, improved protocols for minimizing intra-household transmission. This demands a robust knowledge of how the human and physical geography of a region influences the articulation of disease across space and through time. In the case of Los Angeles, and other large metropolitan regions, preventive education, medical supplies, and vaccination plans should be developed, particularly for neighborhoods where numerous essential workers live, especially if those workers live in crowded households, in large apartment blocks, and have large percentages of people with linguistic or cultural barriers that reduce the effectiveness of hygienic protocols that slow the spread of infections.

Acknowledgments

The authors would like to thank the team at the Health Equity Research and Education (HERE) Center for their generous support of health equity research (NIH UL1GM118976). The authors would also like to acknowledge the support provided by the National Science Foundation for fostering student research and data science instruction. (NSF Data Science Program with Career Support and Connections to Industry, Award #1842386).

Literature Cited

- Aalbers, Manuel B., Niels Beerepoot, and Gerritsen Martijn. 2020. The Geography of the COVID-19 Pandemic. *Tijdschrift voor economische en sociale geografie* 111 (3): 201–204. doi:10.1111/tesg.12456
- Asongu, Simplice A., Samba Diop, and Joseph Nnanna. 2020. The geography of the effectiveness and consequences of Covid-19 measures: Global Evidence. *Journal of Public Affairs* 1–9. <https://doi.org/10.1002/pa.2483>
- County of Los Angeles Public Health. 2021. COVID-19 Homepage. *County of Los Angeles Public Health*. May 1. Accessed 14 May 2021. <http://publichealth.lacounty.gov/media/coronavirus/locations.htm>
- Franch-Pardo, Ivan, Brian M. Napoletano, Fernando Rosete-Verges, and Lawal Billa. 2020. Spatial Analysis and GIS in the Study of COVID-19. A review. *Science of the Total Environment* 739: 1–10. <https://doi.org/10.1016/j.scitotenv.2020.140033>
- Graves, Steven M., and Lauren He. 2020. Covid-19 Mapping with Microsoft Power BI. *Terra Digitalis* 4 (2): 1–5. <https://doi.org/10.22201/igg.25940694e.2020.2.74>
- Los Angeles Times*. 2021. [datadesk/california-coronavirus_data](https://github.com/datadesk/california-coronavirus-data). May 1. Accessed 1 May 2021. <https://github.com/datadesk/california-coronavirus-data>
- Meade, Melinda S., and Michael Emch. 2010. *Medical Geography*. Guilford Press.
- Mollalo, Abolfazl, Behzad Vahedi, and Kiara M. Rivera. 2020. GIS-based Spatial Modeling of COVID-19 Incidence Rate in the Continental United States. *Science of the Total Environment* 728: 1–8. <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Price, Gregory N., and Eric J von Holm. 2020. *The Effect of Social Distancing On The Spread of Novel Coronavirus: Estimates From Linked State-Level Infection And American Time Use Survey Data*. doi:10.13140/RG.2.2.30632.32007
- Real, Leslie A., and Roman Biek. 2007. Spatial dynamics and genetics of infectious diseases on heterogenous landscapes. *Journal of the Royal Society Interface* 4: 935–948. doi:10.1098/rsif.2007.1041
- Schrager, Benjamin. 2020. The geography of the US's mishandling of COVID-19: A commentary on the politics of science in democracies. *The Geographical Journal* 187: 51–56. doi: 10.1111/geoj.12363
- Shepherd, Marshall. 2020. *Why Geography is a Key Part of Fighting the COVID-19 Coronavirus Outbreak*. March 5. Accessed 5 May 2021. <https://www.forbes.com/sites/marshallshepherd/2020/03/05/why-the-discipline-of-geography-is-a-key-part-of-the-coronavirus-fight/>
- Smith, Shaqualia. 2021. Is Florida Under-Counting COVID-19 Deaths? *The Famuan*. April 12. Accessed 11 May 2021. <http://www.thefamuanonline.com/2021/04/12/is-florida-under-counting-covid-19-deaths/>

- Sun, Feinuo, Stephen A. Matthews, and Tse-Chuan Yang. 2020. A Spatial Analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: Where geography matters? *Annals of Epidemiology* 52: 54–59. <https://doi.org/10.1016/j.annepidem.2020.07.014>
- Wheaton, Will C., and Anne Kinsella Thompson. 2020. *The Geography of Covid-19 Growth in the US: Counties and Metropolitan Areas*. April 6. Accessed 5 May 2021. <https://mitcre.mit.edu/news/the-geography-of-covid-19-growth-in-the-us-counties-and-metropolitan-areas>
- Zalla, Lauren C., Chantel L. Martin, Jessie K. Edwards, Danielle R. Gartner, and Grace A. Noppert. 2021. A geography of risk: Structural racism and COVID-19 mortality in the United States. *American Journal of Epidemiology*. <https://doi.org/10.1093/aje/kwab>