

¹ Reproducibility of research during COVID-19:
² examining the case of population density and the basic
³ reproductive rate from the perspective of spatial analysis

⁴ Author^{*,a}

⁵ *^aDepartment, Street, City, State ZIP*

⁶ **Abstract**

The emergence of the novel SARS-CoV-2 coronavirus and the global COVID-19 pandemic has led to explosive growth in scientific research. Alas, much of the research in the literature lacks conditions to be reproducible, and recent publications on the association between population density and the basic reproductive number of SARS-CoV-2 are no exception. Relatively few papers share code and data sufficiently, which hinders not only verification but additional experimentation. In this paper, an example of reproducible research shows the potential of spatial analysis for epidemiology research during COVID-19. Transparency and openness means that independent researchers can, with relatively modest efforts, verify findings and use different approaches as appropriate. Given the high stakes of the situation, it is essential that scientific findings, on which good policy depends, are as robust as possible; as the empirical example shows, reproducibility is one of the keys to ensure this.

*Corresponding Author
Email address: author@institution.edu (Author)

⁷ **Introduction**

⁸ The emergence of the novel SARS-CoV-2 coronavirus in 2019, and the global
⁹ pandemic that followed in its wake, led to an explosive growth of research around
¹⁰ the globe. According to Fraser et al. (2021), over 125,000 COVID-19-related
¹¹ papers were released in the first ten months from the first confirmed case of
¹² the disease. Of these, more than 30,000 were shared in pre-print servers, the
¹³ use of which also exploded in the past year (Añazco et al., 2021; Kwon, 2020;
¹⁴ Vlasschaert et al., 2020).

¹⁵ Given the ruinous human and economic cost of the pandemic, there has
¹⁶ been a natural tension in the scientific community between the need to publish
¹⁷ research results quickly and the imperative to maintain consistently high quality
¹⁸ standards in scientific reporting; indeed, a call for maintaining the standards in
¹⁹ published research has termed the deluge of COVID-19 publications a “carnage
²⁰ of substandard research” (Bramstedt, 2020). Part of the challenge of maintaining
²¹ quality standards in published research is that, despite an abundance of recom-
²² mendations and guidelines (e.g., Broggini et al., 2017; Brunsdon and Comber,
²³ 2020; Ince et al., 2012; Ioannidis et al., 2014), in practice reproducibility has
²⁴ remained a lofty and somewhat aspirational goal (Konkol et al., 2019; Konkol
²⁵ and Kray, 2019). As reported in the literature, only a woefully small proportion
²⁶ of published research was actually reproducible before the pandemic (Iqbal et al.,
²⁷ 2016; Stodden et al., 2018), and the situation does not appear to have changed
²⁸ substantially since (Gustot, 2020; Sumner et al., 2020).

²⁹ The push for open data and software, along with more strenuous efforts
³⁰ towards open, reproducible research, is simply a continuation of long-standing
³¹ scientific practices of independent verification. Despite the (at times dispro-
³²portionate) attention that high profile scandals in science tend to elicit in the
³³ media, science as a collective endeavor is remarkable for being a self-correcting
³⁴ enterprise, one with built-in mechanisms and incentives to weed out erroneous
³⁵ ideas. Over the long term, facts tend to prevail in science. At stake is the
³⁶ shorter-term impacts that research may have in other spheres of economic and
³⁷ social life. The case of economists Reinhart and Rogoff comes to mind: by the
³⁸ time the inaccuracies and errors in their research were uncovered (see Herndon
³⁹ et al., 2014), their claims about debt and economic growth had already been
⁴⁰ seized by policy-makers on both sides of the Atlantic to justify austerity policies

41 in the aftermath of the Great Recession of 2007-2009¹. As later research has
42 demonstrated, those policies cast a long shadow, and their sequels continued to
43 be felt for years (Basu et al., 2017).

44 In the context of COVID-19, a topic that has grabbed the imagination of
45 numerous thinkers has been the prospect of life in cities after the pandemic
46 (Florida et al., 2020); the implications of the pandemic for urban planning,
47 design, and management are the topic of ongoing research (Sharifi and Khavarian-
48 Garmsir, 2020). The fact that the worst of the pandemic was initially felt in dense
49 population centers such as Wuhan, Milan, Madrid, and New York, unleashed a
50 torrent of research into the associations between density and the spread of the
51 pandemic. he answers to some important questions hang on the results of these
52 research efforts. For example, are lower density regions safer from the pandemic?
53 Are de-densification policies warranted, even if just in the short term? And in
54 the longer term, will the risks of life in high density regions presage a flight from
55 cities? What are the implications of the pandemic for future urban planning and
56 practice? Over the past year, numerous papers have sought to throw light on
57 the underlying issue of density and the pandemic; nonetheless the results, as will
58 be detailed next, remain mixed. Further, to complicate matters, precious few of
59 these studies appear to be sufficiently open to support independent verification.

60 The objective of this paper is to illustrate the importance of reproducibility
61 in research in the context of the flood of COVID-19 papers. To this end,
62 a recent study by Sy et al. (2021) is chosen as an example of reproducible
63 research. The objective is not to malign the analysis of these researchers, but
64 rather to demonstrate the value of openness to allow for independent verification
65 and further analysis. Open data and open code mean that an independent
66 researcher can, with only modest efforts, not only verify the findings reported,
67 but also examine the same data from a perspective which may not have been
68 available to the original researchers due to differences in disciplinary perspectives,
69 methodological traditions, and/or training, among other possible factors. The
70 example, which shows consequential changes in the conclusions reached by
71 different analyses, should serve as a call to researchers to redouble their efforts
72 to increase transparency and reproducibility in research. The present paper, in

¹Nobel Prize in Economics Paul Krugman noted that “Reinhart–Rogoff may have had more immediate influence on public debate than any previous paper in the history of economics” <https://www.nybooks.com/articles/2013/06/06/how-case-austerity-has-crumbled/?pagination=false>

73 addition, aims to show how data can be packaged in well-documented, shareable
74 units, and code can be embedded into self-contained documents suitable for
75 review and independent verification. The source for this paper is an R Markdown
76 document which, along with the data package, will be available in a public
77 repository².

78 **Background: the intuitive relationship between density and spread of
79 contagious diseases**

80 The concern with population density and the spread of the virus during the
81 COVID-19 pandemic was fueled, at least in part, by dramatic scenes seen in
82 real-time around the world from large urban centers such as Wuhan, Milan,
83 Madrid, and New York. In theory, there are good reasons to believe that higher
84 density may have a positive association with the transmission of a contagious
85 virus. It has long been known that the potential for inter-personal contact is
86 greater in regions with higher density (see for example the research on urban
87 fields and time-geography, including Farber and Páez, 2011; Moore, 1970; Moore
88 and Brown, 1970). Mathematically, models of exposure and contagion indicate
89 that higher densities can catalyze the transmission of contagious diseases (Li et
90 al., 2018; Rocklöv and Sjödin, 2020). The idea is intuitive and likely at the root
91 of messages, by some figures in positions of authority, that regions with sparse
92 population densities faced lower risks from the pandemic³.

93 As Rocklöv and Sjödin (Rocklöv and Sjödin, 2020) note, however, mathematical
94 models of contagion are valid at small-to-medium spatial scales (and
95 presumably, small temporal scales too, such as time spent in restaurants, concert
96 halls, cruises), and the results do not necessarily transfer to larger spatial units
97 and different time scales. There are solid reasons for this: while in a restaurant,
98 one can hardly avoid being in proximity to other customers. On the other
99 hand, a person can choose to (or be forced to as a matter of policy) not go to a
100 restaurant in the first place. Nonetheless, the idea that high density correlates
101 with high transmission is so seemingly sensible that it is often taken for granted

²For peer-review purposes, the data package and code are currently in an anonymous Drive folder: <https://drive.google.com/drive/folders/1cT6tcUc1pJ4aT5ajQ0emO0lyS46P8Ige?usp=sharing>

³Governor Kristi Noem of South Dakota, for example, claimed that sparse population density allowed her state to face the pandemic down without the need for strict policy interventions <https://www.inforum.com/lifestyle/health/5025620-South-Dakota-is-not-New-York-City-Noem-defends-lack-of-statewide-COVID-19-restrictions>

even at larger scales (e.g., Cruz et al., 2020; Micallef et al., 2020). At larger scales, however, there exists the possibility of behavioral adaptations, which are difficult to capture in the mechanistic framework of differential equations (or can be missing in agent-based models, e.g., Gomez et al., 2021); these adaptations, in fact, can be a key aspect of disease transmission.

A plausible behavioral adaptation during a pandemic, especially one broadcast as widely and intensely as COVID-19, is risk compensation. Risk compensation is a process whereby people adjust their behavior in response to their *perception* of risk (Noland, 1995; Phillips et al., 2011; Richens et al., 2000). In the case of COVID-19, Chauhan et al. (Chauhan et al., 2021) have found that perception of risks in the US varies between rural, suburban, and urban residents, with rural residents in general expressing less concern about the virus. It is possible that people who listened to the message of leaders saying that they were safe from the virus because of low density may not have taken adequate precautions. Conversely, people in dense places who could more directly observe the impact of the pandemic may have become overly cautious. Both Paez et al. (2020) and Hamidi et al. (2020b) posit this mechanism (i.e., greater compliance with social distancing in denser regions) to explain the results of their analyses. The evidence available does indeed show that there were important changes in behavior with respect to mobility during the pandemic (Harris and Braniion-Calles, 2021; Jamal and Paez, 2020; Molloy et al., 2020); furthermore, shelter in place orders may have had greater buy-in from the public in higher density regions (Feyman et al., 2020; Hamidi and Zandiatashbar, 2021), and the associated behavior may have persisted beyond the duration of official social-distancing policies (Prahraj et al., 2020). In addition, there is evidence that changes in mobility correlated with the trajectory of the pandemic (Noland, 2021; Paez, 2020). Given the potential for behavioral adaptation, the question of density becomes more nuanced: it is not just a matter of proximity, but also of human behavior, which is better studied using population-level data and models.

Background: but what does the literature say?

When it comes to population density and the spread of COVID-19, the international literature to date remains inconclusive.

On the one hand, there are studies that report positive associations between population density and various COVID-19-related outcomes. Bhadra (2021), for example, reported a moderate positive correlation between the spread of

137 COVID-19 and population density at the district level in India, however their
138 analysis was bivariate and did not control for other variables, such as income.
139 Similarly, Kadi and Khelfaoui (2020) found a positive and significant correlation
140 between number of cases and population density in cities in Algeria in a series
141 of simple regression models (i.e., without other controls). A question in these
142 relatively simple analyses is whether density is not a proxy for other factors.
143 Other studies have included controls, such as Pequeno et al. (2020), a team
144 that reported a positive association between density and cumulative counts
145 of confirmed COVID-19 cases in state capitals in Brazil after controlling for
146 covariates, including income, transport connectivity, and economic status. In
147 a similar vein, Fielding-Miller et al. (2020) reported a positive relationship
148 between the absolute number of COVID-19 deaths and population density (rate)
149 in rural counties in the US. Roy and Ghosh (2020) used a battery of machine
150 learning techniques to find discriminatory factors, and a positive and significant
151 association between COVID-19 infection and death rates in US states. Wong and
152 Li (2020) also found a positive and significant association between population
153 density and number of confirmed COVID-19 cases in US counties, using both
154 univariate and multivariate regressions with spatial effects. More recently, Sy
155 et al. (2021) reported that the basic reproductive number of COVID-19 in US
156 counties tended to increase with population density, but at a decreasing rate at
157 higher densities.

158 On the flip side, a number of studies report non-significant or negative
159 associations between population density and COVID-19 outcomes. This includes
160 the research of Sun et al. (2020) who did not find evidence of significant
161 correlation between population density and confirmed number of cases per day
162 *in conditions of lockdown* in China. This finding echoes the results of Paez et
163 al. (2020), who in their study of provinces in Spain reported non-significant
164 associations between population density and infection rates in the early days of
165 the first wave of COVID-19, and negative significant associations in the later
166 part of the first lockdown. Similarly, (2020) found zero or negative associations
167 between population density and infection numbers/deaths by country. Fielding-
168 Miller et al. (2020) contrast their finding about rural counties with a negative
169 relationship between COVID-19 deaths and population density in urban counties
170 in the US. For their part, in their investigation of doubling time, White and
171 Hébert-Dufresne (2020) identified a negative and significant correlation between
172 population density and doubling time in US states. Likewise, (2021) found
173 a small negative (and significant) association between population density and

174 COVID-19 morbidity in districts in Tehran. Finally, two of the most complete
175 studies in the US (by Hamidi et al. 2020a, and 2020b) used an extensive set
176 of controls to find negative and significant correlations between density and
177 COVID-19 cases and fatalities at the level of counties in the US.

178 As can be seen, these studies are implemented at different scales in different
179 regions of the world. They also use a range of techniques, from correlation
180 analysis, to multivariate regression, spatial regressions, and machine learning
181 techniques. This is natural and to be expected: individual researchers have only
182 limited time and expertise. This is why reproducibility is important. To pick an
183 example (which will be further elaborated in later sections of this paper), the
184 study of Sy et al. [(2021); hereafter SWN] would immediately grab the attention
185 of a researcher with expertise in spatial analysis.

186 **Reproducibility of research**

187 SWN investigated the basic reproductive number of COVID-19 in US counties,
188 and its association with population density, median household income, and
189 prevalence of private mobility. For their multivariate analysis, SWN used mixed
190 linear models. This is a reasonable modelling choice: R_0 is an interval-ratio
191 variable that is suitably modeled using linear regression; further, as SWN note
192 there is a likelihood that the process is not independent “among counties
193 within each state, potentially due to variable resource allocation and differing
194 health systems across states” (p. 3). A mixed linear model accounts for this by
195 introducing random components; in the case of SWN, these are random intercepts
196 at the state level. SWN estimated various models with different combinations
197 of variables, including median household income and prevalence of travel by
198 private transportation. These are sensible controls, given potential variations
199 in behavior: people in more affluent counties may have greater opportunities
200 to work from home, and use of private transportation reduces contact with
201 strangers. Moreover, they also conducted various sensitivity analyses. After
202 these efforts, SWN concluded that there is a positive association between the
203 basic reproductive number and population density at the level of counties in the
204 US.

205 One salient aspect of the analysis in SWN is that the basic reproductive
206 number can only be calculated reliably with a minimum number of cases, and a
207 large number of counties did not meet such threshold. As researchers do, SWN
208 made modelling decisions, in this case basing their analysis only on counties

with valid observations. A modeler with expertise in spatial analysis would likely ask some of the following questions on reading SWN's paper: how were missing counties treated? What are the implications of the spatial sampling framework used in the analysis? Is it possible to spatially interpolate the missing observations? Was there spatial residual autocorrelation in the models, or was the use of mixed models sufficient to capture spatial dependencies? These questions are relevant and their implications important. Fortunately, SWN are an example of a reasonably open, reproducible research product: their paper is accompanied by (most of) the data and (most of) the code used in the analysis. This means that an independent expert can, with only a moderate investment of time and effort, reproduce the results in the paper, as well as ask additional questions.

Alas, reproducibility is not necessarily the norm in the relevant literature. There are various reasons why a project can fail to be reproducible. In some cases, there might be legitimate reasons to withhold the data, perhaps due to confidentiality and privacy reasons (e.g., Lee et al., 2020). But in many other cases the data are publicly available, which in fact has commonly been the case with population-level COVID-19 information. Typically the provenance of the data is documented, but in numerous studies the data themselves are not shared (Amadu et al., 2021; Bhadra et al., 2021; Cruz et al., 2020; Feng et al., 2020; Fielding-Miller et al., 2020; Hamidi et al., 2020a, 2020b; Inbaraj et al., 2021; Souris and Gonzalez, 2020). As any researcher can attest, whether a graduate student or a seasoned scientist, collecting, organizing, and preparing data for a project can take a substantial amount of time. Pointing to the sources of data, even when these sources are public, is a small step towards reproducibility—but only a very small one. Faced with the prospect of having to recreate a data set from raw sources is probably sufficient to dissuade all but the most dedicated (or stubborn) researcher from independent verification. This is true even if part of the data are shared (e.g., Wong and Li, 2020). In other cases, data are shared, but the processes followed in the preparation of the data are not fully documented (Ahmad et al., 2020; Skórka et al., 2020). These processes matter, as shown by the errors in the spreadsheets of Reinhart and Rogoff (Herndon et al., 2014) and the data of biologist Jonathan Pruitt that led to an “avalanche” of paper retractions⁴. Another situation is when papers share well-documented data, but

⁴<https://doi.org/10.1038/d41586-020-00287-y>

fail to provide the code used in the analysis (Noury et al., 2021; Pequeno et al., 2020; Wang et al., 2021). Making code available only “on demand” (e.g., Brandtner et al., 2021) is an unnecessary barrier when most journals offer the facility to share supplemental materials online. Then there are those papers that more closely comply with reproducibility standards, and share well-documented processes and data, as well as the code used in any analyses reported (Feyman et al., 2020; Paez et al., 2020; Stephens et al., 2021; Sy et al., 2021; White and Hébert-Dufresne, 2020). Even in this case, the pressure to publish “new findings” instead of replication studies can act as a deterrent, perhaps particularly for younger researchers⁵.

In the following sections, the analysis of RWN is reproduced, some relevant questions from the perspective of an independent researcher with expertise in spatial analysis are asked, and the data are reanalyzed.

Reproducing SWN

SWN examined the association between the basic reproductive number of COVID-19 and population density. The basic reproductive number R_0 is a summary measure of contact rates, probability of transmission of a pathogen, and duration of infectiousness. In rough terms, R_0 measures how many new infections each infection begets. Infectious disease outbreaks generally tend to die out when $R_0 < 1$, and to grow when $R_0 > 1$. Reliable calculation of R_0 requires a minimum number of cases to be able to assume that there is community transmission of the pathogen. Accordingly, SWN based their analysis only on counties that had at least 25 cases or more at the end of the exponential growth phase (see Fig. 1). Their final sample included 1,151 counties in the US, including in Alaska, Hawaii, Puerto Rico, and island territories.

Table 1 reproduces the first three models of SWN (the fourth model did not have any significant variables; see Table 1 in SWN). It is possible to verify that the results match, with only the minor (and irrelevant) exception of the magnitude of the coefficient for travel by private transportation, which is due to a difference in the input (here the variable is changed to one percent units,

⁵The present paper was desk rejected by three journals that had previously published research on population density and the spread of COVID-19; in one case, the paper was too opinionated for the journal, in the other two cases, the paper was not a “good fit” despite dealing with a nearly identical issue as papers previously published in said journals. This does not inspire much confidence in the commitment of journals to reproducibility in research.

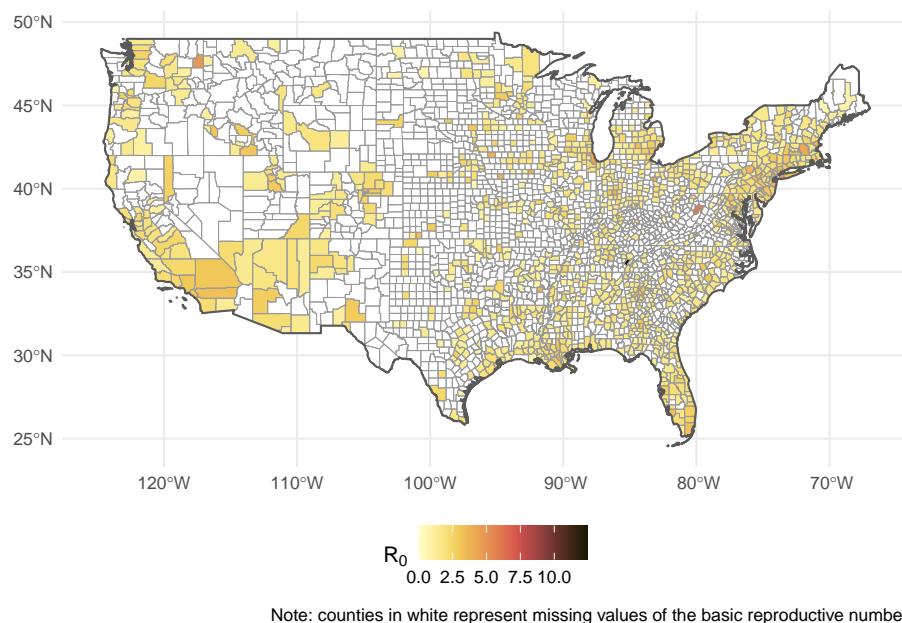


Figure 1: Basic reproductive rate in US counties (Alaska, Hawaii, Puerto Rico, and territories not shown).

Table 1: Reproducing SWN: Models 1-3

Variable	Model 1		Model 2		Model 3	
	beta	95% CI	beta	95% CI	beta	95% CI
Intercept	2.274	[2.167, 2.381]	3.347	[2.676, 4.018]	3.386	[2.614, 4.157]
Log of population density	0.162	[0.133, 0.191]	0.145	[0.115, 0.176]	0.147	[0.113, 0.18]
Percent of private transportation			-0.013	[-0.02, -0.005]	-0.013	[-0.021, -0.005]
Median household income (\$10,000)					-0.003	[-0.033, 0.026]
Standard deviation (Intercept)	0.166	[0.108, 0.254]	0.136	[0.081, 0.229]	0.137	[0.081, 0.232]
Within-group standard error	0.665	[0.638, 0.693]	0.665	[0.638, 0.693]	0.665	[0.638, 0.694]

instead of the ten percent units used by SWN). The mixed linear model gives random intercepts (i.e., the intercept is a random variable), and the standard deviation is reported in the fourth row of Table 1. It is useful to map the random intercepts: as seen in Figure 2, other things being equal, counties in Texas tend to have somewhat lower values of R_0 (i.e., a negative random intercept), whereas counties in South Dakota tend to have higher values of R_0 . The key of the analysis, after extensive sensitivity analysis, is a robust finding that population density has a positive association with the basic reproductive number. But does it?

Expanding on SWN

The preceding section shows that thanks to the availability of code and data, it is possible to verify the results reported by SWN. As noted earlier, though, an independent researcher might have wondered about the implications of the spatial sampling procedure used by SWN. The decision to use a sample of counties with reliable basic reproductive numbers, although apparently sensible, results in a non-random spatial sampling scheme. Turning our attention back to Figure 1, we form the impression that many counties without reliable values of R_0 are in more rural, less dense parts of the United States. This impression is reinforced when we overlay the boundaries of urban areas with population greater than 50,000 on the counties with valid values of R_0 (see Figure 3). The fact that R_0 could not be accurately computed in many counties without large urban areas does not mean that there was no transmission of the virus: it simply means that we do not know with sufficient precision to what extent that was the case. The low number of cases may be related to low population and/or low population density. This is intriguing, to say the least: by excluding cases based on the ability to calculate R_0 we are potentially *selecting* the sample in a non-random way.

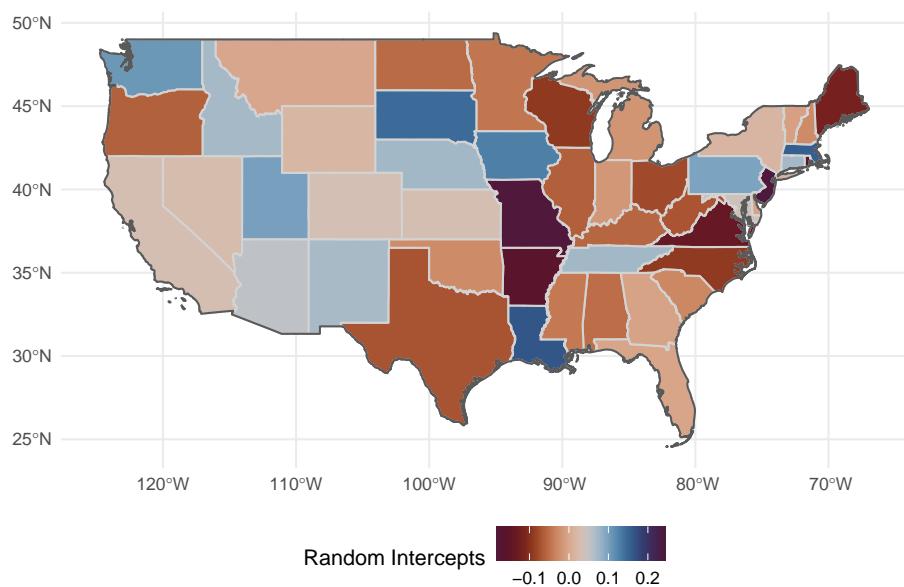


Figure 2: Random intercepts of Model 3 (Alaska, Hawaii, Puerto Rico, and territories not shown).

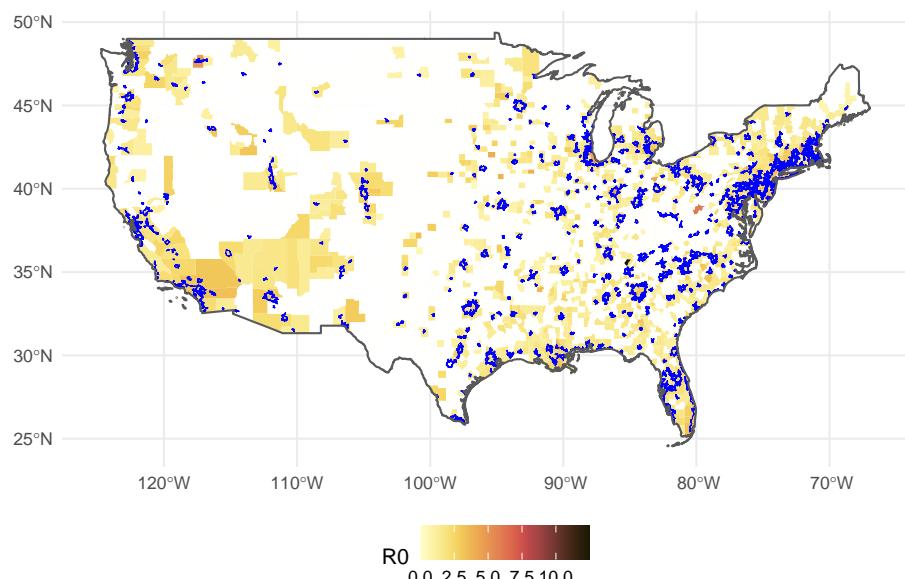


Figure 3: Urban areas with population > 50,000 (Alaska, Hawaii, Puerto Rico, and territories not shown).

A problematic issue with non-random sample selection is that parameter estimates can become unreliable, and numerous techniques have been developed to address this. A model useful for sample selection problems is Heckman's selection model (see Maddala, 1983). The selection model is in fact a system of two equations, as follows:

$$\begin{aligned} y_i^{S*} &= \beta^{S'} x_i^S + \epsilon_i^S \\ y_i^{O*} &= \beta^{O'} x_i^O + \epsilon_i^O \end{aligned}$$

where y_i^{S*} is a latent variable for the sample selection process, and y_i^{O*} is the latent outcome. Vectors x_i^S and x_i^O are explanatory variables (with the possibility that $x_i^S = x_i^O$). Both equations include random terms (i.e., ϵ_i^S and ϵ_i^O). The first equation is designed to model the *probability* of sampling, and the second equation the outcome of interest (say R_0). The random terms are jointly distributed and correlated with parameter ρ .

What the analyst observes is the following:

$$y_i^S = \begin{cases} 0 & \text{if } y_i^{S*} < 0 \\ 1 & \text{otherwise} \end{cases}$$

and:

$$y_i^O = \begin{cases} 0 & \text{if } y_i^S = 0 \\ y_i^{O*} & \text{otherwise} \end{cases}$$

In other words, the outcome of interest is observed *only* for certain cases ($y_i^S = 1$, i.e., for sampled observations). The probability of sampling depends on x_i^S . For the cases observed, the outcome y_i^O depends on x_i^O .

A sample selection model is estimated using the same selection of variables as SWN Model 3. This is Sample Selection Model 1 in Table 2. The first thing to notice about this model is that the sample selection process and the outcome are correlated ($\rho \neq 0$ with 5% of confidence). The selection equation indicates that the probability of a county to be in the sample increases with population density (but at a decreasing rate due to the log-transformation), when travel by private modes is more prevalent, and as median household income in the county is higher. This is in line with the impression made by Figure 3 that counties with reliable values of R_0 tended to be those with larger urban centers. Once that the selection probabilities are accounted for in the model, several things happen with the outcomes model. First, the coefficient for population density is still positive,

320 but the magnitude changes: in effect, it appears that the effect of density is more
321 pronounced than what SWN Model 3 indicated. The coefficient for percent of
322 private transportation changes signs. And the coefficient for median household
323 income is now significant.

324 The second model in Table 2 (Selection Model 2) changes the way the
325 variables are entered into the model. The log-transformation of density in SWN
326 and Selection Model 1 assumes that the association between density and R_0 is
327 monotonically increasing (if the sign of the coefficient is positive) or decreasing
328 (if the sign of the coefficient is negative). There are some indications that the
329 relationship may actually not be monotonical. For example, Paez et al. (2020)
330 found a positive (if non-significant) relationship between density and incidence
331 of COVID-19 in the provinces of Spain at the beginning of the pandemic. This
332 changed to a negative (and significant) relationship during the lockdown. In
333 the case of the US, Fielding-Miller et al. (2020) found that the association
334 between COVID-19 deaths and population density was positive in rural counties,
335 but negative in urban counties. A variable transformation that allows for non-
336 monotonic changes in the relationship is the square of the density.

337 As seen in the table, Selection Model 2 replaces the log-transformation of
338 population density with a quadratic expansion. The results of this analysis
339 indicate that with this variable transformation, the selection and outcome
340 processes are still correlated ($\rho \neq 0$ with 5% of confidence). But a few other
341 interesting things emerge. When we examine the outcomes model, we see that
342 the quadratic expansion has a positive coefficient for the first order term, but a
343 negative coefficient for the second order term. This indicates that R_0 initially
344 tends to increase as density grows, but only up to a point, after which the
345 negative second term (which grows more rapidly due to the square), becomes
346 increasingly dominant. Secondly, the sign of the coefficient for travel by private
347 transportation becomes negative again. This, of course, makes more sense
348 than the positive sign of Selection Model 1: if people tend to travel in private
349 transportation, the potential for contact should be lower instead of higher. And
350 finally median household income is no longer significant, similar to SWN Model
351 3.

352 **Proceed with caution: spatial effects ahead**

353 The results of the selection models, in particular Selection Model 2, make
354 us reassess the original conclusion that density has a positive association with

Table 2: Estimation results of sample selection models

Variable	Selection Model 1		Selection Model 2	
	β	95% CI	β	95% CI
Sample Selection Model				
Intercept	-2.237	[-3.109, -1.365]	-7.339	[-8.381, -6.297]
Log of population density	0.385	[0.352, 0.418]		
Density (1,000 per sq.km)			2.484	[2.13, 2.838]
Density squared			-0.387	[-0.473, -0.3]
Percent of private transportation	0.025	[0.016, 0.034]	0.057	[0.046, 0.067]
Median household income (10,000)	0.202	[0.168, 0.235]	0.32	[0.283, 0.357]
Outcome Model				
Intercept	0.605	[-0.257, 1.466]	2.784	[1.652, 3.915]
Log of population density	0.39	[0.354, 0.426]		
Density (1,000 per sq.km)			0.758	[0.509, 1.008]
Density squared			-0.132	[-0.187, -0.077]
Percent of private transportation	0.01	[0.001, 0.018]	-0.011	[-0.021, -0.001]
Median household income (\$10,000)	0.126	[0.094, 0.159]	0.002	[-0.033, 0.037]
σ	0.954	[0.904, 1.003]	0.684	[0.652, 0.716]
ρ	0.971	[0.961, 0.98]	-0.199	[-0.377, -0.022]

355 the basic reproductive number of COVID-19. A spatial analyst might still
 356 wonder about spatial residual autocorrelation. A challenge here is that spatial
 357 models tend to be technically more demanding, and although spatial models
 358 for qualitative variables exist, a spatial implementation of the sample selection
 359 model does not appear to exist. It might be argued that a reproducible research
 360 project can also allow a researcher to be more adventurous with their modeling
 361 decisions: since data and code are shared, other researchers can promptly and
 362 with relative ease poke the methods and see if they appear to be sound.

363 In the present case, it appears that an application of spatial filtering (see
 364 Getis and Griffith, 2002; Griffith, 2004; Paez, 2019) can help. Spatial filtering
 365 provides an elegant solution to regression problems that may have difficulties
 366 handling the spatial structures of spatial statistical and econometric models
 367 (Griffith, 2000). A key issue in the present example is the fact that there are
 368 numerous missing observations, which prevents the calculation of autocorrelation
 369 statistics, let alone the estimation of models with spatial components.

370 The following is an unorthodox, but potentially effective use of filters in a
 371 sample selection model:

- 372 1. Estimate a sample selection model and retrieve the residuals of the outcome.
 373 This will be a vector with missing values for locations that were not sampled.
 374 2. Fit a spatial filter to the residuals. This is done by regressing the estimated

375 residuals of the *observed* data on the corresponding values of the Moran
376 eigenvectors.

377 3. The resulting filter will correlate highly with the known residuals, and will
378 provide information in non-sampled locations that is consistent with the
379 spatial pattern of the known residuals.

380 4. Test the filter for spatial autocorrelation:
381 4.1 If significant spatial autocorrelation is detected, this would be indicative
382 of residual spatial pattern. Introduce the filter as a covariate in the outcome
383 model of the sample selection model and return to step 1.
384 4.2 If no significant spatial autocorrelation is detected, this would be
385 indicative of random residual pattern. Stop.

386 This procedure is implemented using a stopping criterion whereby the search
387 for the filter only stops when the p-value of Moran's Coefficient of the filter
388 fitted to the residuals is greater than 0.25, which was chosen as a sufficiently
389 conservative value for testing for autocorrelation. The correlation of the known
390 residuals with the corresponding elements of the filter is consistently high (the
391 correlation coefficient typically is greater than 0.9). The results of implementing
392 this procedure appear in Table 3 as Selection Model 3. The results are consistent
393 with Selection Model 2, with two intriguing differences: 1) the variance of
394 Sample Model 3 is smaller; and 2) the sample and outcome processes are
395 no longer correlated (the confidence interval of ρ includes zero). It appears
396 that by capturing the spatial pattern of the residuals, which is likely strongly
397 determined by the non-random sampling framework, the outcome model is not
398 only substantially more precise, but also appears to be independent from the
399 selection process.

400 **Discussion**

401 How relevant are the differences between the various model specifications
402 presented above? Figure 4 shows the relationship between density and R_0 implied
403 by SWN Model 3, Selection Model 2, and Selection Model 3. The left panel
404 of the figure shows the non-linear but monotonic relationship implied by SWN
405 Model 1. The conclusion is that at higher densities, R_0 is *always* higher. The
406 two panels on the right, in contrast, shows that Selection Model 2 and Selection
407 Model 3 coincide that R_0 tends to increase as density grows. This continues until
408 a density of approximately 2.9 (1,000 people per sq.km). At higher densities

Table 3: Estimation results of sample selection model with spatial filter

Variable	Selection Model 3	
	β	95% CI
Sample Selection Model		
Intercept	-7.304	[-8.346, -6.262]
Density (1,000 per sq.km)	2.445	[2.089, 2.802]
Density squared	-0.380	[-0.468, -0.292]
Percent of private transportation	0.056	[0.046, 0.067]
Median household income (10,000)	0.318	[0.28, 0.356]
Outcome Model		
Intercept	2.563	[2.497, 2.629]
Density (1,000 per sq.km)	0.760	[0.746, 0.774]
Density squared	-0.133	[-0.135, -0.13]
Percent of private transportation	-0.011	[-0.012, -0.011]
Median household income (\$10,000)	0.002	[-0.001, 0.004]
Spatial filter	1.000	[0.998, 1.001]
σ	0.017	[0.015, 0.019]
ρ	-0.304	[-0.957, 0.349]

409 than that the relationship between density and R_0 begins to weaken, and the
 410 relationship becomes negative at densities higher than approximately 5.7 (1,000
 411 people per sq.km).

412 To put this into context, other things being equal, the effect of density in
 413 a county like Charlottesville in Virginia (density ~1,639 people per sq.km) is
 414 roughly the same as that in a county like Philadelphia (density ~4,127 people
 415 per sq.km). In contrast, the effect of density on R_0 in a county like Arlington in
 416 Virginia (density ~3,093 people per sq.km) is *stronger* than either of the previous
 417 two examples. Lastly, the density of counties like San Francisco in California, or
 418 Queens and Bronx in NY, which are among the densest in the US, contributes
 419 even less to R_0 than even the most rural counties in the country.

420 It is worth at this point to recall Cressie's dictum about modelling: "[w]hat
 421 is one person's mean structure could be another person's correlation structure"
 422 (Cressie, 1989, p. 201). There are almost always multiple ways to approach a
 423 modelling situation. In the present case, we would argue that spatial sampling
 424 is an important aspect of the modeling process, but one that perhaps required
 425 different technical skills than those available to SWN. There is nothing wrong
 426 with that. What matters is that, by adopting relatively high reproducibility
 427 standards, these researchers made a valuable and honest contribution to the
 428 collective enterprise of seeking knowledge. Their effort, and subsequent efforts
 429 to validate and expand on their work, can potentially contribute to provide

430 clarity to ongoing conversations about the relevance of density and the spread of
431 COVID-19.

432 In particular, it is noteworthy that a sample selection model with a different
433 variable transformation does not lend support to the thesis that higher density
434 is *always* associated with a greater risk of spread of the virus [as put by Wong
435 and Li, “‘Density is destiny’ is probably an overstatement”; (2020)]. At the
436 same time, this also stands in contrast to the findings of Hamidi et al., who
437 found that higher density was either not significantly associated with the rate
438 of the virus in a cross-sectional study (Hamidi et al., 2020b), or was negatively
439 associated with it in a longitudinal setting [Hamidi et al. (2020a)]. In this
440 sense, the conclusion that density does not aggravate the pandemic may have
441 been somewhat premature; instead, reanalysis of the data of SWN suggests
442 that Fielding-Miller et al. (2020) might be onto something with respect to
443 the difference between rural and urban counties. More generally, there is no
444 doubt that in population-level studies density is indicative of proximity, but it
445 also potentially is a proxy for adaptive behavior. And it is possible that the
446 determining factor during COVID-19, at least in the US, has been variations in
447 perceptions of the risks associated with contagion (Chauhan et al., 2021), and
448 subsequent compensations in behavior in more and less dense regions.

449 Conclusion

450 The tension between the need to publish research potentially useful in dealing
451 with a global pandemic, and a potential “carnage of substandard research”
452 (Bramstedt, 2020), highlights the importance of efforts to maintain the quality of
453 scientific outputs during COVID-19. An important part of quality control is the
454 ability of independent researchers to verify and examine the results of materials
455 published in the literature. As previous research illustrates, reproducibility in
456 scientific research remains an important but elusive goal (Gustot, 2020; e.g.,
457 Iqbal et al., 2016; Stodden et al., 2018; Sumner et al., 2020). This idea is
458 reinforced by the review conducted for this paper in the context of research
459 about population density and the spread of COVID-19.

460 Taking one recent example from the literature [Sy et al., Sy et al. (2021);
461 SWN], the present paper illustrates the importance of good reproducibility
462 practices. Sharing data and code can catalyze research, by allowing independent
463 verification of findings, as well as additional research. After verifying the results of
464 SWN, experiments with sample selection models and variations in the definition

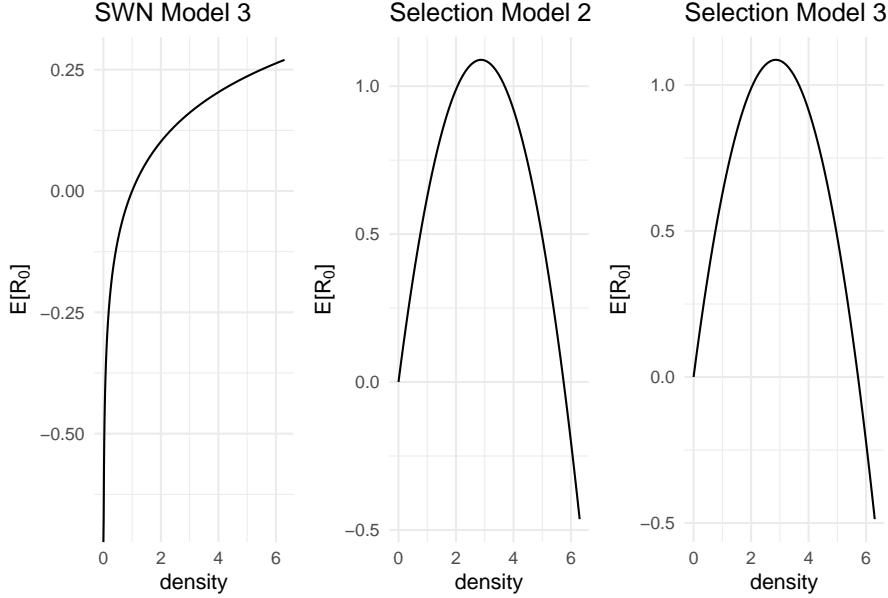


Figure 4: Effect of density according to SWN Model 3 and Sample Selection Model 2.

465 of model inputs, lead to an important reappraisal of the conclusion that high
 466 density is associated with greater spread of the virus. Instead, the possibility
 467 of a non-monotonical relationship between population density and contagion is
 468 raised.

469 In the spirit of openness, this paper is prepared as an R Markdown document,
 470 an a companion data package is provided. The data package contains the relevant
 471 documentation of the data, and all data pre-processing is fully documented.
 472 Hopefully this, and similar reproducible papers, will continue to encourage others
 473 to adopt reproducible standards in their research.

474 **References**

- 475 Ahmad, K., Erqou, S., Shah, N., Nazir, U., Morrison, A.R., Choudhary, G.,
 476 Wu, W.-C., 2020. Association of poor housing conditions with COVID-
 477 19 incidence and mortality across US counties. PLOS ONE 15, e0241327.
 478 doi:10.1371/journal.pone.0241327
- 479 Amadu, I., Ahinkorah, B.O., Afitiri, A.-R., Seidu, A.-A., Ameyaw, E.K., Hagan,
 480 J.E., Duku, E., Aram, S.A., 2021. Assessing sub-regional-specific strengths of

- 481 healthcare systems associated with COVID-19 prevalence, deaths and recoveries in africa. PLOS ONE 16, e0247274. doi:10.1371/journal.pone.0247274
- 482
- 483 Añazco, D., Nicolalde, B., Espinosa, I., Camacho, J., Mushtaq, M., Gimenez, J.,
- 484 Teran, E., 2021. Publication rate and citation counts for preprints released
- 485 during the COVID-19 pandemic: The good, the bad and the ugly. PeerJ 9,
- 486 e10927. doi:10.7717/peerj.10927
- 487 Basu, S., Carney, M.A., Kenworthy, N.J., 2017. Ten years after the financial
- 488 crisis: The long reach of austerity and its global impacts on health. Social
- 489 Science & Medicine 187, 203–207. doi:10.1016/j.socscimed.2017.06.026
- 490 Bhadra, A., Mukherjee, A., Sarkar, K., 2021. Impact of population density on
- 491 covid-19 infected and mortality rate in india. Modeling Earth Systems and
- 492 Environment 7, 623–629. doi:10.1007/s40808-020-00984-7
- 493 Bramstedt, K.A., 2020. The carnage of substandard research during the COVID-
- 494 19 pandemic: A call for quality. Journal of Medical Ethics 46, 803–807.
- 495 doi:10.1136/medethics-2020-106494
- 496 Brandtner, C., Bettencourt, L.M.A., Berman, M.G., Stier, A.J., 2021. Creatures
- 497 of the state? Metropolitan counties compensated for state inaction
- 498 in initial u.s. Response to COVID-19 pandemic. PLOS ONE 16, e0246249.
- 499 doi:10.1371/journal.pone.0246249
- 500 Broggini, F., Dellinger, J., Fomel, S., Liu, Y., 2017. Reproducible research: Geo-
- 501 physics papers of the future - introduction. Geophysics 82. doi:10.1190/geo2017-
- 502 0918-spseintro.1
- 503 Brunsdon, C., Comber, A., 2020. Opening practice: Supporting reproducibility
- 504 and critical spatial data science. Journal of Geographical Systems.
- 505 doi:10.1007/s10109-020-00334-2
- 506 Chauhan, R.S., Capasso da Silva, D., Salon, D., Shamshiripour, A., Rahimi,
- 507 E., Sutradhar, U., Khoeini, S., Mohammadian, A.(Kouros)., Derrible, S.,
- 508 Pendyala, R., 2021. COVID-19 related attitudes and risk perceptions
- 509 across urban, rural, and suburban areas in the united states. Findings.
- 510 doi:10.32866/001c.23714
- 511 Cressie, N., 1989. Geostatistics. The American Statistician 43, 197. doi:10.2307/2685361
- 512 Cruz, C.J.P., Ganly, R., Li, Z., Gietel-Basten, S., 2020. Exploring the young de-
- 513 mographic profile of COVID-19 cases in hong kong: Evidence from migration
- 514 and travel history data. PLOS ONE 15, e0235306. doi:10.1371/journal.pone.0235306
- 515 Farber, S., Páez, A., 2011. Running to stay in place: The time-use implications
- 516 of automobile oriented land-use and travel. Journal of Transport Geography
- 517 19, 782–793. doi:10.1016/j.jtrangeo.2010.09.008

- 518 Feng, Y., Li, Q., Tong, X., Wang, R., Zhai, S., Gao, C., Lei, Z., Chen, S., Zhou,
519 Y., Wang, J., Yan, X., Xie, H., Chen, P., Liu, S., Xv, X., Liu, S., Jin, Y.,
520 Wang, C., Hong, Z., Luan, K., Wei, C., Xu, J., Jiang, H., Xiao, C., Guo, Y.,
521 2020. Spatiotemporal spread pattern of the COVID-19 cases in china. PLOS
522 ONE 15, e0244351. doi:10.1371/journal.pone.0244351
- 523 Feyman, Y., Bor, J., Raifman, J., Griffith, K.N., 2020. Effectiveness of COVID-19
524 shelter-in-place orders varied by state. PLOS ONE 15, e0245008. doi:10.1371/journal.pone.0245008
- 525 Fielding-Miller, R.K., Sundaram, M.E., Brouwer, K., 2020. Social determinants
526 of COVID-19 mortality at the county level. PLOS ONE 15, e0240151.
527 doi:10.1371/journal.pone.0240151
- 528 Florida, R., Glaeser, E., Sharif, M., Bedi, K., Campanella, T., Chee, C., Doctoroff,
529 D., Katz, B., Katz, R., Kotkin, J., 2020. How life in our cities will look after
530 the coronavirus pandemic. Foreign Policy 1.
- 531 Fraser, N., Brierley, L., Dey, G., Polka, J.K., Pálfy, M., Nanni, F., Coates,
532 J.A., 2021. The evolving role of preprints in the dissemination of COVID-19
533 research and their impact on the science communication landscape. PLOS
534 Biology 19, e3000959. doi:10.1371/journal.pbio.3000959
- 535 Getis, A., Griffith, D.A., 2002. Comparative spatial filtering in regression analysis.
536 Geographical Analysis 34, 130–140.
- 537 Gomez, J., Prieto, J., Leon, E., Rodríguez, A., 2021. INFECTA—an agent-based
538 model for transmission of infectious diseases: The COVID-19 case in bogotá,
539 colombia. PLOS ONE 16, e0245787. doi:10.1371/journal.pone.0245787
- 540 Griffith, D.A., 2000. A linear regression solution to the spatial autocorrelation
541 problem. Journal of Geographical Systems 2, 141–156.
- 542 Griffith, D.A., 2004. A spatial filtering specification for the autologistic model.
543 Environment and Planning A 36, 1791–1811.
- 544 Gustot, T., 2020. Quality and reproducibility during the COVID-19 pandemic.
545 JHEP Rep 2, 100141. doi:10.1016/j.jhepr.2020.100141
- 546 Hamidi, S., Ewing, R., Sabouri, S., 2020a. Longitudinal analyses of the rela-
547 tionship between development density and the COVID-19 morbidity and
548 mortality rates: Early evidence from 1,165 metropolitan counties in the united
549 states. Health & Place 64, 102378. doi:10.1016/j.healthplace.2020.102378
- 550 Hamidi, S., Sabouri, S., Ewing, R., 2020b. Does density aggravate the COVID-
551 19 pandemic? Journal of the American Planning Association 86, 495–509.
552 doi:10.1080/01944363.2020.1777891
- 553 Hamidi, S., Zandiatashbar, A., 2021. Compact development and adherence
554 to stay-at-home order during the COVID-19 pandemic: A longitudinal

- 555 investigation in the united states. *Landscape and Urban Planning* 205,
556 103952. doi:<https://doi.org/10.1016/j.landurbplan.2020.103952>
- 557 Harris, M.A., Braniion-Calles, M., 2021. Changes in commute mode attributed to
558 COVID-19 risk in canadian national survey data. *Findings*. doi:[10.32866/001c.19088](https://doi.org/10.32866/001c.19088)
- 559 Herndon, T., Ash, M., Pollin, R., 2014. Does high public debt consistently stifle
560 economic growth? A critique of reinhart and rogoff. *Cambridge Journal of
561 Economics* 38, 257–279. doi:[10.1093/cje/bet075](https://doi.org/10.1093/cje/bet075)
- 562 Inbaraj, L.R., George, C.E., Chandrasingh, S., 2021. Seroprevalence of COVID-19
563 infection in a rural district of south india: A population-based seroepidemi-
564 logical study. *PLOS ONE* 16, e0249247. doi:[10.1371/journal.pone.0249247](https://doi.org/10.1371/journal.pone.0249247)
- 565 Ince, D.C., Hatton, L., Graham-Cumming, J., 2012. The case for open computer
566 programs. *Nature* 482, 485–488. doi:[10.1038/nature10836](https://doi.org/10.1038/nature10836)
- 567 Ioannidis, J.P.A., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R.,
568 Moher, D., Schulz, K.F., Tibshirani, R., 2014. Increasing value and reduc-
569 ing waste in research design, conduct, and analysis. *Lancet* 383, 166–175.
570 doi:[10.1016/s0140-6736\(13\)62227-8](https://doi.org/10.1016/s0140-6736(13)62227-8)
- 571 Iqbal, S.A., Wallach, J.D., Khoury, M.J., Schully, S.D., Ioannidis, J.P.A., 2016.
572 Reproducible research practices and transparency across the biomedical
573 literature. *Plos Biology* 14. doi:[10.1371/journal.pbio.1002333](https://doi.org/10.1371/journal.pbio.1002333)
- 574 Jamal, S., Paez, A., 2020. Changes in trip-making frequency by mode during
575 COVID-19. *Findings*. doi:[10.32866/001c.17977](https://doi.org/10.32866/001c.17977)
- 576 Kadi, N., Khelfaoui, M., 2020. Population density, a factor in the spread of
577 COVID-19 in algeria: Statistic study. *Bulletin of the National Research
578 Centre* 44. doi:[10.1186/s42269-020-00393-x](https://doi.org/10.1186/s42269-020-00393-x)
- 579 Khavarian-Garmsir, A.R., Sharifi, A., Moradpour, N., 2021. Are high-density
580 districts more vulnerable to the COVID-19 pandemic? *Sustainable Cities
581 and Society* 70, 102911. doi:[10.1016/j.scs.2021.102911](https://doi.org/10.1016/j.scs.2021.102911)
- 582 Konkol, M., Kray, C., 2019. In-depth examination of spatiotemporal figures
583 in open reproducible research. *Cartography and Geographic Information
584 Science* 46, 412–427. doi:[10.1080/15230406.2018.1512421](https://doi.org/10.1080/15230406.2018.1512421)
- 585 Konkol, M., Kray, C., Pfeiffer, M., 2019. Computational reproducibility in
586 geoscientific papers: Insights from a series of studies with geoscientists and
587 a reproduction study. *International Journal of Geographical Information
588 Science* 33, 408–429. doi:[10.1080/13658816.2018.1508687](https://doi.org/10.1080/13658816.2018.1508687)
- 589 Kwon, D., 2020. How swamped preprint servers are blocking bad coronavirus
590 research. *Nature* 581, 130–132.
- 591 Lee, M., Zhao, J., Sun, Q., Pan, Y., Zhou, W., Xiong, C., Zhang, L., 2020. Human

- 592 mobility trends during the early stage of the COVID-19 pandemic in the
593 united states. PLOS ONE 15, e0241468. doi:10.1371/journal.pone.0241468
- 594 Li, R., Richmond, P., Roehner, B.M., 2018. Effect of population density on
595 epidemics. Physica A: Statistical Mechanics and its Applications 510, 713–724.
596 doi:10.1016/j.physa.2018.07.025
- 597 Maddala, G.S., 1983. Limited-dependent and qualitative variables in econometrics.
598 Cambridge University Press, Cambridge.
- 599 Micallef, S., Piscopo, T.V., Casha, R., Borg, D., Vella, C., Zammit, M.-A.,
600 Borg, J., Mallia, D., Farrugia, J., Vella, S.M., Xerri, T., Portelli, A., Fenech,
601 M., Fsadni, C., Mallia Azzopardi, C., 2020. The first wave of COVID-
602 19 in malta; a national cross-sectional study. PLOS ONE 15, e0239389.
603 doi:10.1371/journal.pone.0239389
- 604 Molloy, J., Tchervenkov, C., Hintermann, B., Axhausen, K.W., 2020. Trac-
605 ing the sars-CoV-2 impact: The first month in switzerland. Findings.
606 doi:10.32866/001c.12903
- 607 Moore, E.G., 1970. Some spatial properties of urban contact fields. Geographical
608 Analysis 2, 376–386.
- 609 Moore, E.G., Brown, L.A., 1970. Urban acquaintance fields: An evaluation of a
610 spatial model. Environment and Planning 2, 443–454.
- 611 Noland, R.B., 1995. PERCEIVED RISK AND MODAL CHOICE - RISK
612 COMPENSATION IN TRANSPORTATION SYSTEM. Accident Analysis
613 and Prevention 27, 503–521. doi:10.1016/0001-4575(94)00087-3
- 614 Noland, R.B., 2021. Mobility and the effective reproduction rate of COVID-19.
615 Journal of Transport & Health 20, 101016. doi:<https://doi.org/10.1016/j.jth.2021.101016>
- 616 Noury, A., François, A., Gergaud, O., Garel, A., 2021. How does COVID-19
617 affect electoral participation? Evidence from the french municipal elections.
618 PLOS ONE 16, e0247026. doi:10.1371/journal.pone.0247026
- 619 Paez, A., 2019. Using spatial filters and exploratory data analysis to en-
620 hance regression models of spatial data. Geographical Analysis 51, 314–338.
621 doi:10.1111/gean.12180
- 622 Paez, A., 2020. Using google community mobility reports to investigate the
623 incidence of COVID-19 in the united states. Findings. doi:<https://doi.org/10.32866/001c.12976>
- 624 Paez, A., Lopez, F.A., Menezes, T., Cavalcanti, R., Pitta, M.G. da R., 2020.
625 A spatio-temporal analysis of the environmental correlates of COVID-19
626 incidence in spain. Geographical Analysis n/a. doi:10.1111/gean.12241

- 629 Pequeno, P., Mendel, B., Rosa, C., Bosholn, M., Souza, J.L., Baccaro, F.,
630 Barbosa, R., Magnusson, W., 2020. Air transportation, population density
631 and temperature predict the spread of COVID-19 in brazil. PeerJ 8, e9322.
632 doi:10.7717/peerj.9322
- 633 Phillips, R.O., Fyhr, A., Sagberg, F., 2011. Risk compensation and bicycle
634 helmets. Risk Analysis 31, 1187–1195. doi:10.1111/j.1539-6924.2011.01589.x
- 635 Praharaj, S., King, D., Pettit, C., Wentz, E., 2020. Using aggregated mobility
636 data to measure the effect of COVID-19 policies on mobility changes in
637 sydney, london, phoenix, and pune. Findings. doi:10.32866/001c.17590
- 638 Richens, J., Imrie, J., Copas, A., 2000. Condoms and seat belts: The parallels
639 and the lessons. Lancet 355, 400–403. doi:10.1016/s0140-6736(99)09109-6
- 640 Rocklöv, J., Sjödin, H., 2020. High population densities catalyse the spread of
641 COVID-19. Journal of Travel Medicine 27. doi:10.1093/jtm/taaa038
- 642 Roy, S., Ghosh, P., 2020. Factors affecting COVID-19 infected and death
643 rates inform lockdown-related policymaking. PLOS ONE 15, e0241165.
644 doi:10.1371/journal.pone.0241165
- 645 Sharifi, A., Khavarian-Garmsir, A.R., 2020. The COVID-19 pandemic: Impacts
646 on cities and major lessons for urban planning, design, and management.
647 Science of The Total Environment 749, 142391. doi:<https://doi.org/10.1016/j.scitotenv.2020.142391>
- 648 Skórka, P., Grzywacz, B., Moroń, D., Lenda, M., 2020. The macroecology of
649 the COVID-19 pandemic in the anthropocene. PLOS ONE 15, e0236856.
650 doi:10.1371/journal.pone.0236856
- 651 Souris, M., Gonzalez, J.-P., 2020. COVID-19: Spatial analysis of hospital case-
652 fatality rate in france. PLOS ONE 15, e0243606. doi:10.1371/journal.pone.0243606
- 653 Stephens, K.E., Chernyavskiy, P., Bruns, D.R., 2021. Impact of altitude on
654 COVID-19 infection and death in the united states: A modeling and obser-
655 vational study. PLOS ONE 16, e0245055. doi:10.1371/journal.pone.0245055
- 656 Stodden, V., Seiler, J., Ma, Z.K., 2018. An empirical analysis of journal pol-
657 icy effectiveness for computational reproducibility. Proceedings of the Na-
658 tional Academy of Sciences of the United States of America 115, 2584–2589.
659 doi:10.1073/pnas.1708290115
- 660 Sumner, J., Haynes, L., Nathan, S., Hudson-Vitale, C., McIntosh, L.D., 2020.
661 Reproducibility and reporting practices in COVID-19 preprint manuscripts.
662 medRxiv 2020.03.24.20042796. doi:10.1101/2020.03.24.20042796
- 663 Sun, Z., Zhang, H., Yang, Y., Wan, H., Wang, Y., 2020. Impacts of geographic
664 factors and population density on the COVID-19 spreading under the lock-
665

- 666 down policies of china. *Science of The Total Environment* 746, 141347.
667 doi:10.1016/j.scitotenv.2020.141347
- 668 Sy, K.T.L., White, L.F., Nichols, B.E., 2021. Population density and basic
669 reproductive number of COVID-19 across united states counties. *PLOS ONE*
670 16, e0249271. doi:10.1371/journal.pone.0249271
- 671 Vlasschaert, C., Topf, J.M., Hiremath, S., 2020. Proliferation of papers and
672 preprints during the coronavirus disease 2019 pandemic: Progress or prob-
673 lems with peer review? *Advances in Chronic Kidney Disease* 27, 418–426.
674 doi:10.1053/j.ackd.2020.08.003
- 675 Wang, F., Tan, Z., Yu, Z., Yao, S., Guo, C., 2021. Transmission and control pres-
676 sure analysis of the COVID-19 epidemic situation using multisource spatio-
677 temporal big data. *PLOS ONE* 16, e0249145. doi:10.1371/journal.pone.0249145
- 678 White, E.R., Hébert-Dufresne, L., 2020. State-level variation of initial COVID-19
679 dynamics in the united states. *PLOS ONE* 15, e0240648. doi:10.1371/journal.pone.0240648
- 680 Wong, D.W.S., Li, Y., 2020. Spreading of COVID-19: Density matters. *PLOS*
681 *ONE* 15, e0242398. doi:10.1371/journal.pone.0242398