

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

<sup>4</sup> Author<sup>\*,a</sup>

<sup>5</sup> *<sup>a</sup>Department, Street, City, State ZIP*

<sup>6</sup> **Abstract**

The emergence of the novel SARS-CoV-2 coronavirus and the global COVID-19 pandemic ~~has in 2019~~ 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 only~~ 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](mailto:author@institution.edu) (Author)

<sup>7</sup> **Introduction**

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

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

<sup>29</sup> The push for open ~~data and software~~, software and data (Arribas-Bel et  
<sup>30</sup> al., 2021; e.g., Biyand, 2020), along with more strenuous efforts towards open,  
<sup>31</sup> reproducible research, is simply a continuation of long-standing scientific practices  
<sup>32</sup> of independent verification. Despite the (at times disproportionate) attention  
<sup>33</sup> that high profile scandals in science tend to elicit in the media, science as a  
<sup>34</sup> collective endeavor is remarkable for being a self-correcting enterprise, one with  
<sup>35</sup> built-in mechanisms and incentives to weed out erroneous ideas. Over the long  
<sup>36</sup> term, facts tend to prevail in science. At stake is the shorter-term impacts that  
<sup>37</sup> research may have in other spheres of economic and social life. The case of  
<sup>38</sup> economists Reinhart and Rogoff comes to mind: by the time the inaccuracies and  
<sup>39</sup> errors in their research were uncovered (see Herndon et al., 2014), their claims  
<sup>40</sup> about debt and economic growth had already been seized by policy-makers  
<sup>41</sup> on both sides of the Atlantic to justify austerity policies in the aftermath of

42 the Great Recession of 2007-2009<sup>1</sup>. As later research has demonstrated, those  
43 policies cast a long shadow, and their sequels continued to be felt for years (Basu  
44 et al., 2017).

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

62 The objective of this paper is to illustrate the importance of reproducibility  
63 in research in the context of the flood of COVID-19 papers. ~~To this end, For this,~~  
64 I focus on a recent study by Sy et al. (2021) ~~is chosen as an example of that~~  
65 ~~examined the correlation between the basic reproductive number of COVID-19,~~  
66 ~~R<sub>0</sub>, and population density. The basic reproductive number is a summary~~  
67 ~~measure of contact rates, probability of transmission of a pathogen, and duration~~  
68 ~~of infectiousness. In rough terms, it measures how many new infections each~~  
69 ~~infections begets. The paper of Sy et al. (2021) was selected for being, in~~  
70 ~~the literature examined, almost alone in supporting~~ reproducible research. ~~The~~  
71 ~~objective~~ Accordingly, I wish to be clear that my objective in singling their work  
72 ~~for discussion~~ is not to malign ~~the analysis of these researchers~~their efforts, but  
73 rather to demonstrate ~~the value of openness to allow for independent verification~~

---

<sup>1</sup>Nobel 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>

74 ~~and further analysis.~~ Open how open and reproducible research efforts can  
75 greatly help to accelerate discovery. More concretely, open data and open code  
76 mean that an independent researcher can, with only modest efforts, not only  
77 verify the findings reported, but also examine the same data from a perspective  
78 which may not have been available to the original researchers due to differences  
79 in disciplinary perspectives, methodological traditions, and/or training, among  
80 other possible factors. The example, which shows consequential changes in the  
81 conclusions reached by different analyses, should serve as a call to researchers to  
82 redouble their efforts to increase transparency and reproducibility in ~~research~~  
83 ~~The present paper , in addition, their research. In this spirit, the present paper~~  
84 ~~also~~ aims to show how data can be packaged in well-documented, shareable units,  
85 and code can be embedded into self-contained documents suitable for review and  
86 independent verification. The source for this paper is an R Markdown document  
87 which, along with the data package, will be available in a public repository<sup>2</sup>.

88 **Background: the intuitive relationship between density and spread of  
89 contagious diseases**

90 The concern with population density and the spread of the virus during the  
91 COVID-19 pandemic was fueled, at least in part, by dramatic scenes seen in real-  
92 time around the world from large urban centers such as Wuhan, Milan, Madrid,  
93 and New York. In theory, there are good reasons to believe that higher density  
94 ~~may could~~ have a positive association with the transmission of a contagious  
95 virus. It has long been known that the potential for inter-personal contact is  
96 greater in regions with higher density (see for example the research on urban  
97 fields and time-geography, including Farber and Páez, 2011; Moore, 1970; Moore  
98 and Brown, 1970). Mathematically, models of exposure and contagion indicate  
99 that higher densities can catalyze the transmission of contagious diseases (Li et  
100 al., 2018; Rocklöv and Sjödin, 2020). The idea is intuitive and likely at the root  
101 of messages, by some figures in positions of authority, that regions with sparse  
102 population densities faced lower risks from the pandemic<sup>3</sup>.

---

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

<sup>3</sup>Governor 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>

103 As Rocklöv and Sjödin (Rocklöv and Sjödin, 2020) note, however, mathematical  
104 models of contagion are valid at small-to-medium ~~spatial scales~~spaces (and  
105 presumably, ~~small temporal scales~~smaller time intervals too, such as time spent  
106 in restaurants, concert halls, cruises), and the results do not necessarily transfer  
107 to larger spatial units and ~~different time scales~~longer time periods. There are  
108 solid reasons for this: while in a restaurant, one can hardly avoid being in proximity  
109 to other customers. On the other hand, a person can choose to (or be forced  
110 to as a matter of policy) not go to a restaurant in the first place. Nonetheless, the  
111 idea that high density correlates with high transmission is so seemingly sensible  
112 that it is often taken for granted even at ~~larger scales~~the scale of large spaces  
113 (e.g., Cruz et al., 2020; Micallef et al., 2020). ~~At larger scales~~In such conditions,  
114 however, there exists the possibility of behavioral adaptations, which are difficult  
115 to capture in the mechanistic framework of differential equations (or can be  
116 missing in agent-based models, e.g., Gomez et al., 2021); these adaptations, in  
117 fact, can be a key aspect of disease transmission.

118 A plausible behavioral adaptation during a pandemic, especially one broadcast  
119 as widely and intensely as COVID-19, is risk compensation. Risk compensation  
120 is a process whereby people adjust their behavior in response to their *perception*  
121 of risk (Noland, 1995; Phillips et al., 2011; Richens et al., 2000). In the case of  
122 COVID-19, Chauhan et al. (Chauhan et al., 2021) have found that perception  
123 of risks in the US varies between rural, suburban, and urban residents, with  
124 rural residents in general expressing less concern about the virus. It is possible  
125 that people who listened to the message of leaders saying that they were safe  
126 from the virus because of low density may not have taken adequate precautions.  
127 Conversely, people in dense places who could more directly observe the impact  
128 of the pandemic may have become overly cautious. Both Paez et al. (2020) and  
129 Hamidi et al. (2020b) posit this mechanism (i.e., greater compliance with social  
130 distancing in denser regions) to explain the results of their analyses. The evidence  
131 available does indeed show that there were important changes in behavior with  
132 respect to mobility during the pandemic (Harris and Braniion-Calles, 2021; Jamal  
133 and Paez, 2020; Molloy et al., 2020); furthermore, shelter in place orders may  
134 have had greater buy-in from the public in higher density regions (Feyman et al.,  
135 2020; Hamidi and Zandiataashbar, 2021), and the associated behavior may have  
136 persisted beyond the duration of official social-distancing policies (Prahraj et  
137 al., 2020). In addition, there is evidence that changes in mobility correlated with  
138 the trajectory of the pandemic (Noland, 2021; Paez, 2020). Given the potential  
139 for behavioral adaptation, the question of density becomes more nuanced: it

<sup>140</sup> is not just a matter of proximity, but also of human behavior, which is better  
<sup>141</sup> studied using population-level data and models.

<sup>142</sup> **Background: but what does the literature say?**

<sup>143</sup> When it comes to population density and the spread of COVID-19, the  
<sup>144</sup> international literature to date remains inconclusive.

<sup>145</sup> On the one hand, there are studies that report positive associations between  
<sup>146</sup> population density and various COVID-19-related outcomes. Bhadra (2021),  
<sup>147</sup> for example, reported a moderate positive correlation between the spread of  
<sup>148</sup> COVID-19 and population density at the district level in India, however their  
<sup>149</sup> analysis was bivariate and did not control for other variables, such as income.  
<sup>150</sup> Similarly, Kadi and Khelfaoui (2020) found a positive and significant correlation  
<sup>151</sup> between number of cases and population density in cities in Algeria in a series  
<sup>152</sup> of simple regression models (i.e., without other controls). A question in these  
<sup>153</sup> relatively simple analyses is whether density is not a proxy for other factors.  
<sup>154</sup> Other studies have included controls, such as Pequeno et al. (2020), a team  
<sup>155</sup> that reported a positive association between density and cumulative counts  
<sup>156</sup> of confirmed COVID-19 cases in state capitals in Brazil after controlling for  
<sup>157</sup> covariates, including income, transport connectivity, and economic status. In  
<sup>158</sup> a similar vein, Fielding-Miller et al. (2020) reported a positive relationship  
<sup>159</sup> between the absolute number of COVID-19 deaths and population density (rate)  
<sup>160</sup> in rural counties in the US. Roy and Ghosh (2020) used a battery of machine  
<sup>161</sup> learning techniques to find discriminatory factors, and a positive and significant  
<sup>162</sup> association between COVID-19 infection and death rates in US states. Wong and  
<sup>163</sup> Li (2020) also found a positive and significant association between population  
<sup>164</sup> density and number of confirmed COVID-19 cases in US counties, using both  
<sup>165</sup> univariate and multivariate regressions with spatial effects. More recently, Sy  
<sup>166</sup> et al. (2021) reported that the basic reproductive number of COVID-19 in US  
<sup>167</sup> counties tended to increase with population density, but at a decreasing rate at  
<sup>168</sup> higher densities.

<sup>169</sup> On the flip side, a number of studies report non-significant or negative  
<sup>170</sup> associations between population density and COVID-19 outcomes. This includes  
<sup>171</sup> the research of Sun et al. (2020) who did not find evidence of significant  
<sup>172</sup> correlation between population density and confirmed number of cases per day  
<sup>173</sup> *in conditions of lockdown* in China. This finding echoes the results of Paez et  
<sup>174</sup> al. (2020), who in their study of provinces in Spain reported non-significant

175 associations between population density and infection rates in the early days  
176 of the first wave of COVID-19, and negative significant associations in the  
177 later part of the first lockdown. Similarly, Skórka et al. (2020) found zero or  
178 negative associations between population density and infection numbers/deaths  
179 by country. Fielding-Miller et al. (2020) contrast their finding about rural  
180 counties with a negative relationship between COVID-19 deaths and population  
181 density in urban counties in the US. For their part, in their investigation of  
182 doubling time, White and Hébert-Dufresne (2020) identified a negative and  
183 significant correlation between population density and doubling time in US  
184 states. Likewise, Khavarian-Garmsir et al. (2021) found a small negative (and  
185 significant) association between population density and COVID-19 morbidity  
186 in districts in Tehran. Finally, two of the most complete studies in the US~~t~~,  
187 by Hamidi et al. (2020a,~~and~~) and Hamidi et al. (2020b), used an extensive  
188 set of controls to find negative and significant correlations between density and  
189 COVID-19 cases and fatalities at the level of counties in the US.

190 As can be seen, these studies are implemented at different scales in different  
191 regions of the world. They also use a range of techniques, from correlation  
192 analysis, to multivariate regression, spatial regressions, and machine learning  
193 techniques. This is natural and to be expected: individual researchers have only  
194 limited time and expertise. This is why reproducibility is important. To pick an  
195 example (which will be further elaborated in later sections of this paper), the  
196 study of Sy et al. -(2021); ~~hereafter SWN~~, hereafter referred to as SWN, would  
197 immediately grab the attention of a researcher with expertise in spatial analysis.

## 198 Reproducibility of research

199 SWN investigated the basic reproductive number of COVID-19 in US counties,  
200 and its association with population density, median household income, and  
201 prevalence of private mobility. For their multivariate analysis, SWN used mixed  
202 linear models. This is ~~a reasonable an appropriate~~ modelling choice:  $R_0$  is an  
203 interval-ratio variable that is suitably modeled using linear regression; further,  
204 as SWN note there is a likelihood that the process is not independent “among  
205 counties within each state, potentially due to variable resource allocation and  
206 differing health systems across states” (p. 3). A mixed linear model accounts for  
207 this by introducing random components; in the case of SWN, these are random  
208 intercepts at the state level. SWN estimated various models with different  
209 combinations of variables, including median household income and prevalence

210 of travel by private transportation. These ~~are sensible controls, given controls~~  
211 help to account for potential variations in behavior: people in more affluent  
212 counties may have greater opportunities to work from home, and use of private  
213 transportation reduces contact with strangers. Moreover, they also conducted  
214 various sensitivity analyses. After these efforts, SWN concluded that there is  
215 a positive association between the basic reproductive number and population  
216 density at the level of counties in the US.

217 One salient aspect of the analysis in SWN is that the basic reproductive  
218 number can only be calculated reliably with a minimum number of cases, and a  
219 large number of counties did not meet such threshold. As researchers do, SWN  
220 made modelling decisions, in this case basing their analysis only on counties  
221 with valid observations. A modeler with expertise in spatial analysis would  
222 likely ask some of the following questions on reading SWN's paper: how were  
223 missing counties treated? What are the implications of the spatial sampling  
224 framework used in the analysis? Is it possible to spatially interpolate the missing  
225 observations? Was there spatial residual autocorrelation in the models, or was  
226 the use of mixed models sufficient to capture spatial dependencies? These  
227 questions are relevant and their implications important. Fortunately, SWN are  
228 an example of a reasonably open, reproducible research product: their paper is  
229 accompanied by (most of) the data and (most of) the code used in the analysis.  
230 This means that an independent ~~expert researcher~~ can, with only a moderate  
231 investment of time and effort, reproduce the results in the paper, as well as ask  
232 additional questions.

233 Alas, reproducibility is not necessarily the norm in the relevant literature.

234 There are various reasons why a project can fail to be reproducible. In some  
235 cases, there might be legitimate reasons to withhold the data, perhaps due to  
236 confidentiality and privacy reasons (e.g., Lee et al., 2020). But in many other  
237 cases the data are publicly available, which in fact has commonly been the case  
238 with population-level COVID-19 information. Typically the provenance of the  
239 data is documented, but in numerous studies the data themselves are not shared  
240 (Amadu et al., 2021; Bhadra et al., 2021; Cruz et al., 2020; Feng et al., 2020;  
241 Fielding-Miller et al., 2020; Hamidi et al., 2020a, 2020b; Inbaraj et al., 2021;  
242 Souris and Gonzalez, 2020). As any researcher can attest, ~~whether a graduate~~  
243 ~~student or a seasoned scientist,~~ collecting, organizing, and preparing data for a  
244 project can take a substantial amount of time. Pointing to the sources of data,  
245 even when these sources are public, is a small step towards reproducibility-but  
246 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 ([see](#) Herndon et al., 2014 ~~and for the discovery of these errors~~, as well as by the data of biologist Jonathan Pruitt that led to an “avalanche” of paper retractions <sup>4</sup>([see](#) Viglione, 2020). Another situation is when papers share well-documented data, but 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 researchers<sup>4</sup>.

In the following sections, the analysis of ~~RWN-SWN~~ 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

---

<sup>4</sup>

The 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.](#)

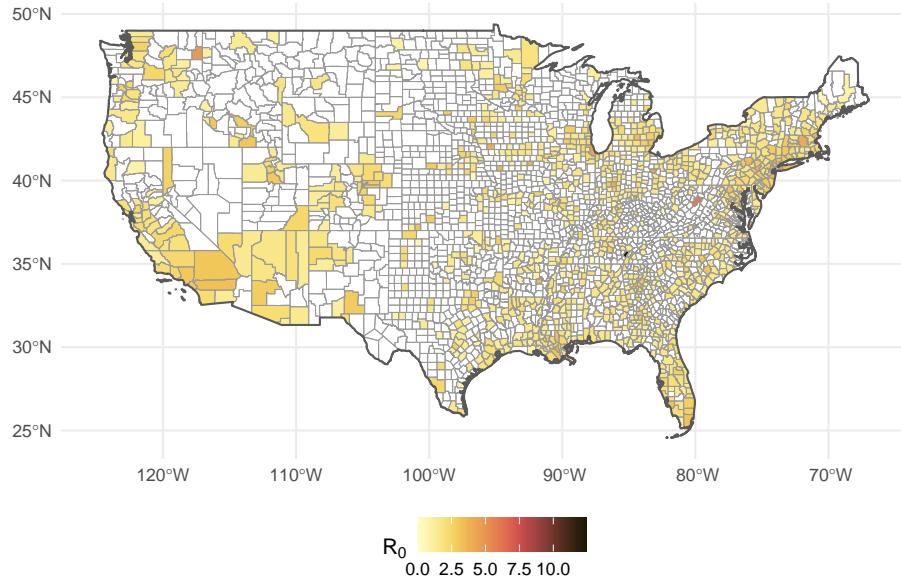


Figure 1: Basic reproductive rate in US counties (Alaska, Hawaii, Puerto Rico, and territories not shown).

277 community transmission of the pathogen. Accordingly, SWN based their analysis  
 278 only on counties that had at least 25 cases or more at the end of the exponential  
 279 growth phase (see Fig. 1). Their final sample included 1,151 counties in the  
 280 US, including in Alaska, Hawaii, Puerto Rico, and island territories. SWN used  
 281 COVID-19 data collected by the New York Times and made available (with  
 282 versioning) in a GitHub repository<sup>5</sup>. For each county, SWN assumed that the  
 283 exponential growth period began one week prior to the second daily increase in  
 284 cases, and assumed that the period of exponential growth lasted approximately  
 285 18 days.

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

<sup>5</sup><https://github.com/nytimes/covid-19-data>

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~~ fifth 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.

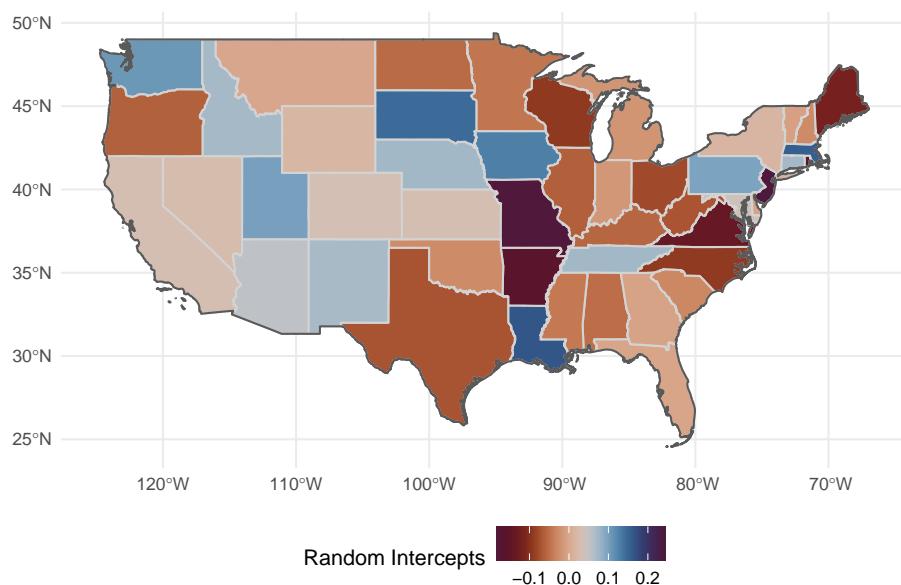


Figure 2: Random intercepts of Model 3 (Alaska, Hawaii, Puerto Rico, and territories not shown).

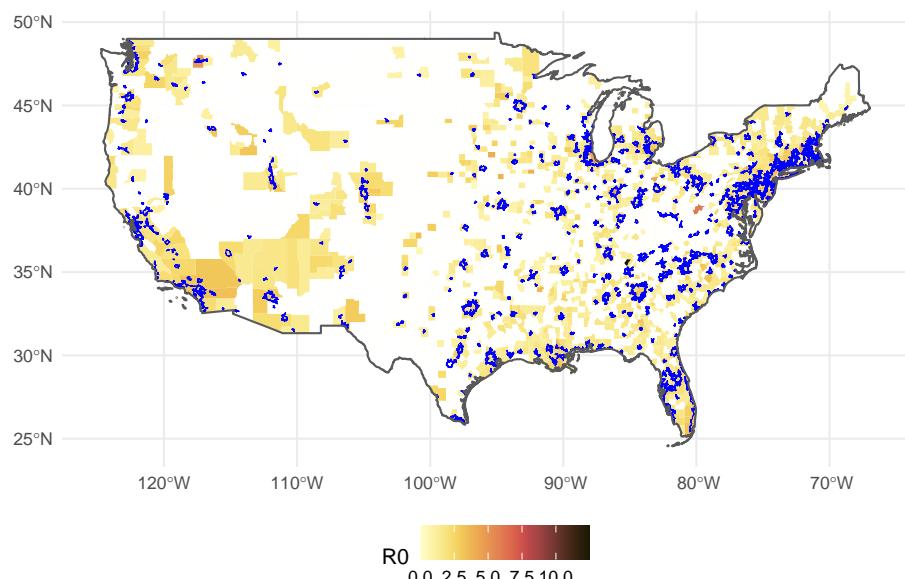


Figure 3: Urban areas with population > 50,000 (Alaska, Hawaii, Puerto Rico, and territories not shown).

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.,  $\epsilon_i^S$  and  $\epsilon_i^O$ ). 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  $\rho$ .

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,

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

342 The second model in Table 2 (Selection Model 2) changes the way the  
343 variables are entered into the model. The log-transformation of density in SWN  
344 and Selection Model 1 assumes that the association between density and  $R_0$  is  
345 monotonically increasing (if the sign of the coefficient is positive) or decreasing  
346 (if the sign of the coefficient is negative). There are some indications that the  
347 relationship may actually not be monotonical. For example, Paez et al. (2020)  
348 found a positive (if non-significant) relationship between density and incidence  
349 of COVID-19 in the provinces of Spain at the beginning of the pandemic. This  
350 changed to a negative (and significant) relationship during the lockdown. In  
351 the case of the US, Fielding-Miller et al. (2020) found that the association  
352 between COVID-19 deaths and population density was positive in rural counties,  
353 but negative in urban counties. A variable transformation that allows for non-  
354 monotonic changes in the relationship is the square of the density.

355 As seen in the table, Selection Model 2 replaces the log-transformation of  
356 population density with a quadratic expansion. The results of this analysis  
357 indicate that with this variable transformation, the selection and outcome  
358 processes are still correlated ( $\rho \neq 0$  with 5% of confidence). But a few other  
359 interesting things emerge. When we examine the outcomes model, we see that  
360 the quadratic expansion has a positive coefficient for the first order term, but a  
361 negative coefficient for the second order term. This indicates that  $R_0$  initially  
362 tends to increase as density grows, but only up to a point, after which the  
363 negative second term (which grows more rapidly due to the square), becomes  
364 increasingly dominant. Secondly, the sign of the coefficient for travel by private  
365 transportation becomes negative again. This, of course, makes more sense  
366 than the positive sign of Selection Model 1: if people tend to travel in private  
367 transportation, the potential for contact should be lower instead of higher. And  
368 finally median household income is no longer significant, similar to SWN Model  
369 3.

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

371 The results of the selection models, in particular Selection Model 2, make  
372 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	
	$\beta$	95% CI	$\beta$	95% CI
<b>Sample Selection Model</b>				
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]
<b>Outcome Model</b>				
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]
$\sigma$	0.954	[0.904, 1.003]	0.684	[0.652, 0.716]
$\rho$	0.971	[0.961, 0.98]	-0.199	[-0.377, -0.022]

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

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

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

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

393 residuals of the *observed* data on the corresponding values of the Moran  
394 eigenvectors.

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

398 4. Test the filter for spatial autocorrelation:  
399 4.1 If significant spatial autocorrelation is detected, this would be indicative  
400 of residual spatial pattern. Introduce the filter as a covariate in the outcome  
401 model of the sample selection model and return to step 1.  
402 4.2 If no significant spatial autocorrelation is detected, this would be  
403 indicative of random residual pattern. Stop.

404 This procedure is implemented using a stopping criterion whereby the search  
405 for the filter only stops when the p-value of Moran's Coefficient of the filter  
406 fitted to the residuals is greater than 0.25, which was chosen as a sufficiently  
407 conservative value for testing for autocorrelation. The correlation of the known  
408 residuals with the corresponding elements of the filter is consistently high (the  
409 correlation coefficient typically is greater than 0.9). The results of implementing  
410 this procedure appear in Table 3 as Selection Model 3. The results are consistent  
411 with Selection Model 2, with two intriguing differences: 1) the variance of  
412 Sample Model 3 is smaller; and 2) the sample and outcome processes are  
413 no longer correlated (the confidence interval of  $\rho$  includes zero). It appears  
414 that by capturing the spatial pattern of the residuals, which is likely strongly  
415 determined by the non-random sampling framework, the outcome model is not  
416 only substantially more precise, but also appears to be independent from the  
417 selection process.

418 ~~How relevant are the differences between the various model specifications  
419 presented above~~ Clearly, the various models display some intriguing differences;  
420 but how relevant are said differences from a more substantive standpoint? Figure  
421 4 shows the relationship between density and  $R_0$  implied by SWN Model 3,  
422 Selection Model 2, and Selection Model 3. The left panel of the figure shows  
423 the non-linear but monotonic relationship implied by SWN Model 1. The  
424 conclusion is that at higher densities,  $R_0$  is *always* higher. The two panels on the  
425 right, in contrast, show that Selection Model 2 and Selection Model 3 coincide  
426 that  $R_0$  tends to increase as density grows. This continues until a density of  
427 approximately 2.9 (1,000 people per sq.km). At higher densities than that the  
428 relationship between density and  $R_0$  begins to weaken, and the relationship

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

Variable	Selection Model 3	
	$\beta$	95% CI
<b>Sample Selection Model</b>		
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]
<b>Outcome Model</b>		
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]
$\sigma$	0.017	[0.015, 0.019]
$\rho$	-0.304	[-0.957, 0.349]

429 becomes negative at densities higher than approximately 5.7 (1,000 people per  
 430 sq.km).

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

## 439 Discussion

440 It is worth at this point to recall Cressie's dictum about modelling: "[w]hat  
 441 is one person's mean structure could be another person's correlation structure"  
 442 (Cressie, 1989, p. 201). There are almost always multiple ways to approach  
 443 a modelling situation—as lively illustrated by a recent paper that reports the  
444 results of a crowdsourced modelling experiment (Schweinsberg et al., 2021). In  
 445 the present case, we would argue that spatial sampling is an important aspect  
 446 of the modeling process, but one that perhaps required different technical skills  
447 than those available to SWN. There is nothing wrong with that. What matters is

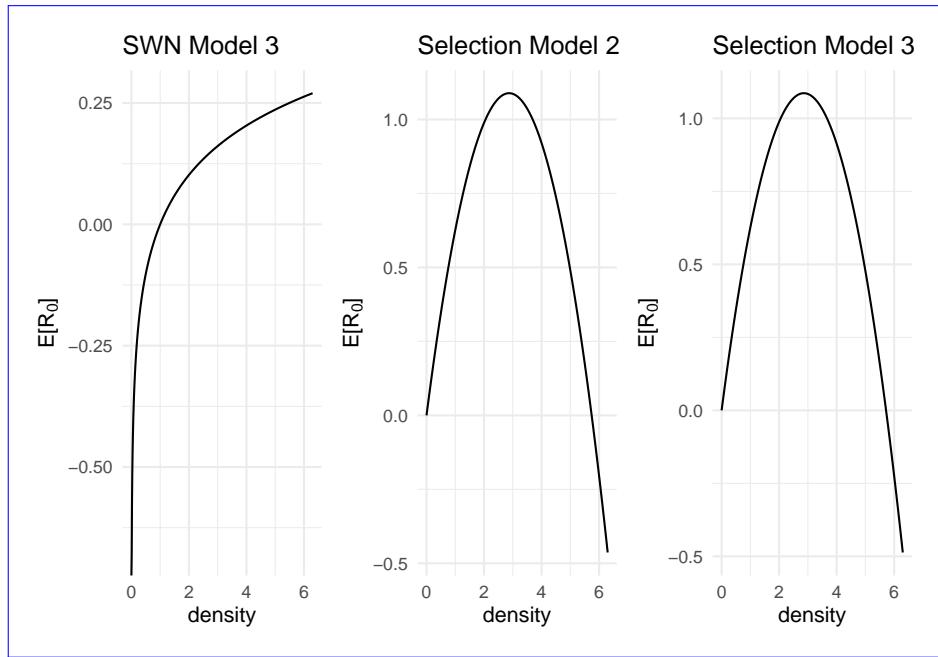


Figure 4: Effect of density according to SWN Model 3 and Sample Selection Model 2.

448 ~~that. Importantly~~, by adopting ~~relatively~~ high reproducibility standards, ~~these~~  
449 ~~researchers~~ SWN made a valuable ~~and honest~~ contribution to the collective  
450 enterprise of seeking knowledge. Their effort, and subsequent efforts to validate  
451 and expand on their work, can potentially contribute to provide clarity to ongoing  
452 conversations about the relevance of density and the spread of COVID-19.

453 In particular, it is noteworthy that a sample selection model with a different  
454 variable transformation does not lend support to the thesis that higher density  
455 is *always* associated with a greater risk of spread of the virus [~~as put by in~~ Wong  
456 and Li's words, “‘Density is destiny’ is probably an overstatement”; (2020)]. At  
457 the same time, ~~this also stands~~ the results presented here also stand in contrast  
458 to the findings of Hamidi et al., who found that higher density was either not  
459 significantly associated with the rate of the virus in a cross-sectional study  
460 (Hamidi et al., 2020b), or was negatively associated with it in a longitudinal  
461 setting [Hamidi et al. (2020a)]. In this sense, the conclusion that density does not  
462 aggravate the pandemic may have been somewhat premature; instead, reanalysis  
463 of the data of SWN suggests that Fielding-Miller et al. (2020) might be onto  
464 something with respect to the difference between rural and urban counties. More  
465 generally, there is no doubt that in population-level studies density is indicative  
466 of proximity, but it also potentially is a proxy for adaptive behavior. And it is  
467 possible that the determining factor during COVID-19, at least in the US, has  
468 been variations in perceptions of the risks associated with contagion (Chauhan  
469 et al., 2021), and subsequent compensations in behavior in more and less dense  
470 regions.

471 ~~Effect of density according to SWN Model 3 and Sample Selection Model 2.~~

## 472 Conclusion

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

483 Taking one recent example from the literature [Sy et al., Sy et al. (2021);  
484 SWN], the present paper illustrates the importance of good reproducibility  
485 practices. Sharing data and code can catalyze research, by allowing independent  
486 verification of findings, as well as additional research. After verifying the results of  
487 SWN, experiments with sample selection models and variations in the definition  
488 of model inputs, lead to an important reappraisal of the conclusion that high  
489 density is associated with greater spread of the virus. Instead, the possibility  
490 of a non-monotonical relationship between population density and contagion  
491 is raised. I do not claim that the analysis presented here is the last word  
492 on the topic of density and the spread of COVID-19, and there is always the  
493 possibility that someone else will be better equipped to analyze these data with  
494 greater competence. By opening up the analysis, documenting the way data  
495 were pre-processed, and by sharing analysis ready data, my hope would be that  
496 others will be able to discover the limitations of my own analysis and improve  
497 on it, as appropriate.

498 In the spirit of openness, this paper is prepared as an R Markdown document,  
499 an a companion data package is provided. The data package contains the  
500 relevant documentation of the data. More generally, my hope is that the research  
501 of Sy et al. (2021), the present paper, and all datapre-processing is fully  
502 documented. Hopefully this, and similar reproducible paperspublications, will  
503 continue to encourage others to adopt reproducible higher reproducibility stan-  
504 dards in their research.

## 505 References

- 506 Ahmad, K., Erqou, S., Shah, N., Nazir, U., Morrison, A.R., Choudhary, G.,  
507 Wu, W.-C., 2020. Association of poor housing conditions with COVID-  
508 19 incidence and mortality across US counties. PLOS ONE 15, e0241327.  
509 doi:10.1371/journal.pone.0241327
- 510 Amadu, I., Ahinkorah, B.O., Afitiri, A.-R., Seidu, A.-A., Ameyaw, E.K., Hagan,  
511 J.E., Duku, E., Aram, S.A., 2021. Assessing sub-regional-specific strengths of  
512 healthcare systems associated with COVID-19 prevalence, deaths and recov-  
513 eries in africa. PLOS ONE 16, e0247274. doi:10.1371/journal.pone.0247274
- 514 Añazco, D., Nicolalde, B., Espinosa, I., Camacho, J., Mushtaq, M., Gimenez, J.,  
515 Teran, E., 2021. Publication rate and citation counts for preprints released  
516 during the COVID-19 pandemic: The good, the bad and the ugly. PeerJ 9,  
517 e10927. doi:10.7717/peerj.10927

- 518 Arribas-Bel, D., Green, M., Rowe, F., Singleton, A., 2021. Open data products:  
519     A framework for creating valuable analysis-ready data. *Journal of Geographical*  
520     *Systems*.
- 521 Basu, S., Carney, M.A., Kenworthy, N.J., 2017. Ten years after the financial  
522     crisis: The long reach of austerity and its global impacts on health. *Social*  
523     *Science & Medicine* 187, 203–207. doi:10.1016/j.socscimed.2017.06.026
- 524 Bhadra, A., Mukherjee, A., Sarkar, K., 2021. Impact of population density on  
525     covid-19 infected and mortality rate in india. *Modeling Earth Systems and*  
526     *Environment* 7, 623–629. doi:10.1007/s40808-020-00984-7
- 527 Bivand, R.S., 2020. Progress in the r ecosystem for representing and handling  
528     spatial data. *Journal of Geographical Systems*. doi:10.1007/s10109-020-00336-0
- 529 Bramstedt, K.A., 2020. The carnage of substandard research during the COVID-  
530     19 pandemic: A call for quality. *Journal of Medical Ethics* 46, 803–807.  
531     doi:10.1136/medethics-2020-106494
- 532 Brandtner, C., Bettencourt, L.M.A., Berman, M.G., Stier, A.J., 2021. Creatures  
533     of the state? Metropolitan counties compensated for state inaction  
534     in initial u.s. Response to COVID-19 pandemic. *PLOS ONE* 16, e0246249.  
535     doi:10.1371/journal.pone.0246249
- 536 Broggini, F., Dellinger, J., Fomel, S., Liu, Y., 2017. Reproducible research: Geo-  
537     physics papers of the future - introduction. *Geophysics* 82. doi:10.1190/geo2017-  
538     0918-spseintro.1
- 539 Brunsdon, C., Comber, A., 2020. Opening practice: Supporting reproducibility  
540     and critical spatial data science. *Journal of Geographical Systems*.  
541     doi:10.1007/s10109-020-00334-2
- 542 Chauhan, R.S., Capasso da Silva, D., Salon, D., Shamshiripour, A., Rahimi,  
543     E., Sutradhar, U., Khoeini, S., Mohammadian, A.(Kouros)., Derrible, S.,  
544     Pendyala, R., 2021. COVID-19 related attitudes and risk perceptions  
545     across urban, rural, and suburban areas in the united states. *Findings*.  
546     doi:10.32866/001c.23714
- 547 Cressie, N., 1989. Geostatistics. *The American Statistician* 43, 197. doi:10.2307/2685361
- 548 Cruz, C.J.P., Ganly, R., Li, Z., Gietel-Basten, S., 2020. Exploring the young de-  
549     mographic profile of COVID-19 cases in hong kong: Evidence from migration  
550     and travel history data. *PLOS ONE* 15, e0235306. doi:10.1371/journal.pone.0235306
- 551 Farber, S., Páez, A., 2011. Running to stay in place: The time-use implications  
552     of automobile oriented land-use and travel. *Journal of Transport Geography*  
553     19, 782–793. doi:10.1016/j.jtrangeo.2010.09.008
- 554 Feng, Y., Li, Q., Tong, X., Wang, R., Zhai, S., Gao, C., Lei, Z., Chen, S., Zhou,

- 555 Y., Wang, J., Yan, X., Xie, H., Chen, P., Liu, S., Xv, X., Liu, S., Jin, Y.,  
556 Wang, C., Hong, Z., Luan, K., Wei, C., Xu, J., Jiang, H., Xiao, C., Guo, Y.,  
557 2020. Spatiotemporal spread pattern of the COVID-19 cases in china. PLOS  
558 ONE 15, e0244351. doi:10.1371/journal.pone.0244351
- 559 Feyman, Y., Bor, J., Raifman, J., Griffith, K.N., 2020. Effectiveness of COVID-19  
560 shelter-in-place orders varied by state. PLOS ONE 15, e0245008. doi:10.1371/journal.pone.0245008
- 561 Fielding-Miller, R.K., Sundaram, M.E., Brouwer, K., 2020. Social determinants  
562 of COVID-19 mortality at the county level. PLOS ONE 15, e0240151.  
563 doi:10.1371/journal.pone.0240151
- 564 Florida, R., Glaeser, E., Sharif, M., Bedi, K., Campanella, T., Chee, C., Doctoroff,  
565 D., Katz, B., Katz, R., Kotkin, J., 2020. How life in our cities will look after  
566 the coronavirus pandemic. Foreign Policy 1.
- 567 Fraser, N., Brierley, L., Dey, G., Polka, J.K., Pálfy, M., Nanni, F., Coates,  
568 J.A., 2021. The evolving role of preprints in the dissemination of COVID-19  
569 research and their impact on the science communication landscape. PLOS  
570 Biology 19, e3000959. doi:10.1371/journal.pbio.3000959
- 571 Getis, A., Griffith, D.A., 2002. Comparative spatial filtering in regression analysis.  
572 Geographical Analysis 34, 130–140.
- 573 Gomez, J., Prieto, J., Leon, E., Rodríguez, A., 2021. INFEKTA—an agent-based  
574 model for transmission of infectious diseases: The COVID-19 case in bogotá,  
575 colombia. PLOS ONE 16, e0245787. doi:10.1371/journal.pone.0245787
- 576 Griffith, D.A., 2000. A linear regression solution to the spatial autocorrelation  
577 problem. Journal of Geographical Systems 2, 141–156.
- 578 Griffith, D.A., 2004. A spatial filtering specification for the autologistic model.  
579 Environment and Planning A 36, 1791–1811.
- 580 Gustot, T., 2020. Quality and reproducibility during the COVID-19 pandemic.  
581 JHEP Rep 2, 100141. doi:10.1016/j.jhepr.2020.100141
- 582 Hamidi, S., Ewing, R., Sabouri, S., 2020a. Longitudinal analyses of the rela-  
583 tionship between development density and the COVID-19 morbidity and  
584 mortality rates: Early evidence from 1,165 metropolitan counties in the united  
585 states. Health & Place 64, 102378. doi:10.1016/j.healthplace.2020.102378
- 586 Hamidi, S., Sabouri, S., Ewing, R., 2020b. Does density aggravate the COVID-  
587 19 pandemic? Journal of the American Planning Association 86, 495–509.  
588 doi:10.1080/01944363.2020.1777891
- 589 Hamidi, S., Zandiatashbar, A., 2021. Compact development and adherence  
590 to stay-at-home order during the COVID-19 pandemic: A longitudinal  
591 investigation in the united states. Landscape and Urban Planning 205,

- 592 103952. doi:<https://doi.org/10.1016/j.landurbplan.2020.103952>
- 593 Harris, M.A., Branić-Calles, M., 2021. Changes in commute mode attributed to  
594 COVID-19 risk in Canadian national survey data. Findings. doi:10.32866/001c.19088
- 595 Herndon, T., Ash, M., Pollin, R., 2014. Does high public debt consistently stifle  
596 economic growth? A critique of Reinhart and Rogoff. Cambridge Journal of  
597 Economics 38, 257–279. doi:10.1093/cje/bet075
- 598 Inbaraj, L.R., George, C.E., Chandrasingh, S., 2021. Seroprevalence of COVID-19  
599 infection in a rural district of south India: A population-based seroepidemiological  
600 study. PLOS ONE 16, e0249247. doi:10.1371/journal.pone.0249247
- 601 Ince, D.C., Hatton, L., Graham-Cumming, J., 2012. The case for open computer  
602 programs. Nature 482, 485–488. doi:10.1038/nature10836
- 603 Ioannidis, J.P.A., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R.,  
604 Moher, D., Schulz, K.F., Tibshirani, R., 2014. Increasing value and reducing  
605 waste in research design, conduct, and analysis. Lancet 383, 166–175.  
606 doi:10.1016/s0140-6736(13)62227-8
- 607 Iqbal, S.A., Wallach, J.D., Khoury, M.J., Schully, S.D., Ioannidis, J.P.A., 2016.  
608 Reproducible research practices and transparency across the biomedical  
609 literature. Plos Biology 14. doi:10.1371/journal.pbio.1002333
- 610 Jamal, S., Paez, A., 2020. Changes in trip-making frequency by mode during  
611 COVID-19. Findings. doi:10.32866/001c.17977
- 612 Kadi, N., Khelfaoui, M., 2020. Population density, a factor in the spread of  
613 COVID-19 in Algeria: Statistic study. Bulletin of the National Research  
614 Centre 44. doi:10.1186/s42269-020-00393-x
- 615 Khavarian-Garmsir, A.R., Sharifi, A., Moradpour, N., 2021. Are high-density  
616 districts more vulnerable to the COVID-19 pandemic? Sustainable Cities  
617 and Society 70, 102911. doi:10.1016/j.scs.2021.102911
- 618 Konkol, M., Kray, C., 2019. In-depth examination of spatiotemporal figures  
619 in open reproducible research. Cartography and Geographic Information  
620 Science 46, 412–427. doi:10.1080/15230406.2018.1512421
- 621 Konkol, M., Kray, C., Pfeiffer, M., 2019. Computational reproducibility in  
622 geoscientific papers: Insights from a series of studies with geoscientists and  
623 a reproduction study. International Journal of Geographical Information  
624 Science 33, 408–429. doi:10.1080/13658816.2018.1508687
- 625 Kwon, D., 2020. How swamped preprint servers are blocking bad coronavirus  
626 research. Nature 581, 130–132.
- 627 Lee, M., Zhao, J., Sun, Q., Pan, Y., Zhou, W., Xiong, C., Zhang, L., 2020. Human  
628 mobility trends during the early stage of the COVID-19 pandemic in the

- 629        united states. PLOS ONE 15, e0241468. doi:10.1371/journal.pone.0241468
- 630        Li, R., Richmond, P., Roehner, B.M., 2018. Effect of population density on
- 631        epidemics. Physica A: Statistical Mechanics and its Applications 510, 713–724.
- 632        doi:10.1016/j.physa.2018.07.025
- 633        Maddala, G.S., 1983. Limited-dependent and qualitative variables in economet-
- 634        rics. Cambridge University Press, Cambridge.
- 635        Micallef, S., Piscopo, T.V., Casha, R., Borg, D., Vella, C., Zammit, M.-A.,
- 636        Borg, J., Mallia, D., Farrugia, J., Vella, S.M., Xerri, T., Portelli, A., Fenech,
- 637        M., Fsadni, C., Mallia Azzopardi, C., 2020. The first wave of COVID-
- 638        19 in malta; a national cross-sectional study. PLOS ONE 15, e0239389.
- 639        doi:10.1371/journal.pone.0239389
- 640        Molloy, J., Tchervenkov, C., Hintermann, B., Axhausen, K.W., 2020. Trac-
- 641        ing the sars-CoV-2 impact: The first month in switzerland. Findings.
- 642        doi:10.32866/001c.12903
- 643        Moore, E.G., 1970. Some spatial properties of urban contact fields. Geographical
- 644        Analysis 2, 376–386.
- 645        Moore, E.G., Brown, L.A., 1970. Urban acquaintance fields: An evaluation of a
- 646        spatial model. Environment and Planning 2, 443–454.
- 647        Noland, R.B., 1995. PERCEIVED RISK AND MODAL CHOICE - RISK
- 648        COMPENSATION IN TRANSPORTATION SYSTEM. Accident Analysis
- 649        and Prevention 27, 503–521. doi:10.1016/0001-4575(94)00087-3
- 650        Noland, R.B., 2021. Mobility and the effective reproduction rate of COVID-19.
- 651        Journal of Transport & Health 20, 101016. doi:<https://doi.org/10.1016/j.jth.2021.101016>
- 652        Noury, A., Fran ois, A., Gergaud, O., Garel, A., 2021. How does COVID-19
- 653        affect electoral participation? Evidence from the french municipal elections.
- 654        PLOS ONE 16, e0247026. doi:10.1371/journal.pone.0247026
- 655        Paez, A., 2019. Using spatial filters and exploratory data analysis to en-
- 656        hance regression models of spatial data. Geographical Analysis 51, 314–338.
- 657        doi:10.1111/gean.12180
- 658        Paez, A., 2020. Using google community mobility reports to investigate the
- 659        incidence of COVID-19 in the united states. Findings. doi:<https://doi.org/10.32866/001c.12976>
- 660        Paez, A., Lopez, F.A., Menezes, T., Cavalcanti, R., Pitta, M.G. da R., 2020.
- 661        A spatio-temporal analysis of the environmental correlates of COVID-19
- 662        incidence in spain. Geographical Analysis n/a. doi:10.1111/gean.12241
- 663        Pequeno, P., Mendel, B., Rosa, C., Bosholn, M., Souza, J.L., Baccaro, F.,

- 666 Barbosa, R., Magnusson, W., 2020. Air transportation, population density  
667 and temperature predict the spread of COVID-19 in brazil. PeerJ 8, e9322.  
668 doi:10.7717/peerj.9322
- 669 Phillips, R.O., Fyhri, A., Sagberg, F., 2011. Risk compensation and bicycle  
670 helmets. Risk Analysis 31, 1187–1195. doi:10.1111/j.1539-6924.2011.01589.x
- 671 Praharaj, S., King, D., Pettit, C., Wentz, E., 2020. Using aggregated mobility  
672 data to measure the effect of COVID-19 policies on mobility changes in  
673 sydney, london, phoenix, and pune. Findings. doi:10.32866/001c.17590
- 674 Richens, J., Imrie, J., Copas, A., 2000. Condoms and seat belts: The parallels  
675 and the lessons. Lancet 355, 400–403. doi:10.1016/s0140-6736(99)09109-6
- 676 Rocklöv, J., Sjödin, H., 2020. High population densities catalyse the spread of  
677 COVID-19. Journal of Travel Medicine 27. doi:10.1093/jtm/taaa038
- 678 Roy, S., Ghosh, P., 2020. Factors affecting COVID-19 infected and death  
679 rates inform lockdown-related policymaking. PLOS ONE 15, e0241165.  
680 doi:10.1371/journal.pone.0241165
- 681 Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O.R., van Aert,  
682 R.C.M., van Assen, M.A.L.M., Liu, Y., Althoff, T., Heer, J., Kale, A.,  
683 Mohamed, Z., Amireh, H., Venkatesh Prasad, V., Bernstein, A., Robinson,  
684 E., Snellman, K., Amy Sommer, S., Otner, S.M.G., Robinson, D., Madan, N.,  
685 Silberzahn, R., Goldstein, P., Tierney, W., Murase, T., Mandl, B., Viganola,  
686 D., Strobl, C., Schaumans, C.B.C., Kelchtermans, S., Naseeb, C., Mason  
687 Garrison, S., Yarkoni, T., Richard Chan, C.S., Adie, P., Alaburda, P., Albers,  
688 C., Alspaugh, S., Alstott, J., Nelson, A.A., Ariño de la Rubia, E., Arzi, A.,  
689 Bahnik, Š., Baik, J., Winther Balling, L., Bunker, S., AA Baranger, D., Barr,  
690 D.J., Barros-Rivera, B., Bauer, M., Blaise, E., Boelen, L., Bohle Carbonell,  
691 K., Briers, R.A., Burkhard, O., Canela, M.-A., Castrillo, L., Catlett, T.,  
692 Chen, O., Clark, M., Cohn, B., Coppock, A., Cugueró-Escofet, N., Curran,  
693 P.G., Cyrus-Lai, W., Dai, D., Valentino Dalla Riva, G., Danielsson, H., F. S.  
694 M. Russo, R. de, de Silva, N., Derungs, C., Dondelinger, F., Duarte de Souza,  
695 C., Tyson Dube, B., Dubova, M., Mark Dunn, B., Adriaan Edelsbrunner,  
696 P., Finley, S., Fox, N., Gnambs, T., Gong, Y., Grand, E., Greenawalt,  
697 B., Han, D., Hanel, P.H.P., Hong, A.B., Hood, D., Hsueh, J., Huang, L.,  
698 Hui, K.N., Hultman, K.A., Javaid, A., Ji Jiang, L., Jong, J., Kamdar, J.,  
699 Kane, D., Kappler, G., Kaszubowski, E., Kavanagh, C.M., Khabsa, M.,  
700 Kleinberg, B., Kouros, J., Krause, H., Kryptotos, A.-M., Lavbić, D., Ling  
701 Lee, R., Leffel, T., Yang Lim, W., Liverani, S., Loh, B., Lönsmann, D.,  
702 Wei Low, J., Lu, A., MacDonald, K., Madan, C.R., Hjorth Madsen, L.,

- 703 Maimone, C., Mangold, A., Marshall, A., Ester Matskewich, H., Mavon,  
704 K., McLain, K.L., McNamara, A.A., McNeill, M., Mertens, U., Miller, D.,  
705 Moore, B., Moore, A., Nantz, E., Nasrullah, Z., Nejkovic, V., Nell, C.S.,  
706 Arthur Nelson, A., Nilsonne, G., Nolan, R., O'Brien, C.E., O'Neill, P.,  
707 O'Shea, K., Olita, T., Otterbacher, J., Palsetia, D., Pereira, B., Pozdniakov,  
708 I., Protzko, J., Reyt, J.-N., Riddle, T., (Akmal) Ridhwan Omar Ali, A.,  
709 Ropovik, I., Rosenberg, J.M., Rothen, S., Schulte-Mecklenbeck, M., Sharma,  
710 N., Shotwell, G., Skarzynski, M., Stedden, W., Stodden, V., Stoffel, M.A.,  
711 Stoltzman, S., Subbaiah, S., Tatman, R., Thibodeau, P.H., Tomkins, S.,  
712 Valdivia, A., Druijff-van de Woestijne, G.B., Viana, L., Villesèche, F., Duncan  
713 Wadsworth, W., Wanders, F., Watts, K., Wells, J.D., Whelpley, C.E., Won,  
714 A., Wu, L., Yip, A., Youngflesh, C., Yu, J.-C., Zandian, A., Zhang, L.,  
715 Zibman, C., Luis Uhlmann, E., 2021. Same data, different conclusions:  
716 Radical dispersion in empirical results when independent analysts operationalize  
717 and test the same hypothesis. *Organizational Behavior and Human Decision  
718 Processes* 165, 228–249. doi:<https://doi.org/10.1016/j.obhdp.2021.02.003>
- 719 Sharifi, A., Khavarian-Garmsir, A.R., 2020. The COVID-19 pandemic: Impacts  
720 on cities and major lessons for urban planning, design, and management.  
721 *Science of The Total Environment* 749, 142391. doi:<https://doi.org/10.1016/j.scitotenv.2020.142391>
- 722 Skórka, P., Grzywacz, B., Moroń, D., Lenda, M., 2020. The macroecology of  
723 the COVID-19 pandemic in the anthropocene. *PLOS ONE* 15, e0236856.  
724 doi:[10.1371/journal.pone.0236856](https://doi.org/10.1371/journal.pone.0236856)
- 725 Souris, M., Gonzalez, J.-P., 2020. COVID-19: Spatial analysis of hospital case-  
726 fatality rate in france. *PLOS ONE* 15, e0243606. doi:[10.1371/journal.pone.0243606](https://doi.org/10.1371/journal.pone.0243606)
- 727 Stephens, K.E., Chernyavskiy, P., Bruns, D.R., 2021. Impact of altitude on  
728 COVID-19 infection and death in the united states: A modeling and obser-  
729 vational study. *PLOS ONE* 16, e0245055. doi:[10.1371/journal.pone.0245055](https://doi.org/10.1371/journal.pone.0245055)
- 730 Stodden, V., Seiler, J., Ma, Z.K., 2018. An empirical analysis of journal pol-  
731 icy effectiveness for computational reproducibility. *Proceedings of the Na-  
732 tional Academy of Sciences of the United States of America* 115, 2584–2589.  
733 doi:[10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115)
- 734 Sumner, J., Haynes, L., Nathan, S., Hudson-Vitale, C., McIntosh, L.D., 2020.  
735 Reproducibility and reporting practices in COVID-19 preprint manuscripts.  
736 medRxiv 2020.03.24.20042796. doi:[10.1101/2020.03.24.20042796](https://doi.org/10.1101/2020.03.24.20042796)
- 737 Sun, Z., Zhang, H., Yang, Y., Wan, H., Wang, Y., 2020. Impacts of geographic  
738 factors and population density on the COVID-19 spreading under the lock-  
739

- 740 down policies of china. Science of The Total Environment 746, 141347.  
741 doi:10.1016/j.scitotenv.2020.141347
- 742 Sy, K.T.L., White, L.F., Nichols, B.E., 2021. Population density and basic  
743 reproductive number of COVID-19 across united states counties. PLOS ONE  
744 16, e0249271. doi:10.1371/journal.pone.0249271
- 745 [Viglione, G., 2020. 'Avalanche' of spider-paper retractions shakes behavioural-ecology  
746 community. Nature 578, 199–201.](#)
- 747 Vlasschaert, C., Topf, J.M., Hiremath, S., 2020. Proliferation of papers and  
748 preprints during the coronavirus disease 2019 pandemic: Progress or prob-  
749 lems with peer review? Advances in Chronic Kidney Disease 27, 418–426.  
750 doi:10.1053/j.ackd.2020.08.003
- 751 Wang, F., Tan, Z., Yu, Z., Yao, S., Guo, C., 2021. Transmission and control pres-  
752 sure analysis of the COVID-19 epidemic situation using multisource spatio-  
753 temporal big data. PLOS ONE 16, e0249145. doi:10.1371/journal.pone.0249145
- 754 White, E.R., Hébert-Dufresne, L., 2020. State-level variation of initial COVID-19  
755 dynamics in the united states. PLOS ONE 15, e0240648. doi:10.1371/journal.pone.0240648
- 756 Wong, D.W.S., Li, Y., 2020. Spreading of COVID-19: Density matters. PLOS  
757 ONE 15, e0242398. doi:10.1371/journal.pone.0242398