

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

4 Antonio Paez^{a,*}

^aSchool of Earth Environment and Society 1280 Main St West Hamilton Ontario L8S 4K1 Canada

5 **Abstract**

6 The emergence of the novel SARS-CoV-2 coronavirus and the global COVID-19 pandemic in 2019 led
7 to explosive growth in scientific research. Alas, much of the research in the literature lacks conditions
8 to be reproducible, and recent publications on the association between population density and the basic
9 reproductive number of SARS-CoV-2 are no exception. Relatively few papers share code and data sufficiently,
10 which hinders not only verification but additional experimentation. In this paper, an example of reproducible
11 research shows the potential of spatial analysis for epidemiology research during COVID-19. Transparency and
12 openness means that independent researchers can, with only modest efforts, verify findings and use different
13 approaches as appropriate. Given the high stakes of the situation, it is essential that scientific findings, on
14 which good policy depends, are as robust as possible; as the empirical example shows, reproducibility is one
15 of the keys to ensure this.

This paper is now published in Geographical Analysis (<https://doi.org/10.1111/gean.12307>)

16 **Keywords:** keyword1, keyword2

*Corresponding author
Email address: paezha@mcmaster.ca (Antonio Paez)

17 **Introduction**

18 The emergence of the novel SARS-CoV-2 coronavirus in 2019, and the global pandemic that followed in
19 its wake, led to an explosive growth of research around the globe. According to Fraser et al. (2021), over
20 125,000 COVID-19-related papers were released in the first ten months from the first confirmed case of the
21 disease. Of these, more than 30,000 were shared in pre-print servers, the use of which also exploded in the
22 past year (Kwon, 2020; Vlasschaert et al., 2020; Añazco et al., 2021).

23 Given the ruinous human and economic cost of the pandemic, there has been a natural tension in the
24 scientific community between the need to publish research results quickly and the imperative to maintain
25 consistently high quality standards in scientific reporting; indeed, a call for maintaining the standards
26 in published research termed the deluge of COVID-19 publications a “carnage of substandard research”
27 (Bramstedt, 2020). Part of the challenge of maintaining quality standards in published research is that,
28 despite an abundance of recommendations and guidelines (e.g., Broggini et al., 2017; Ince et al., 2012;
29 Ioannidis et al., 2014; Brunsdon and Comber, 2020), in practice reproducibility has remained a lofty and
30 somewhat aspirational goal (Konkol and Kray, 2019; Konkol et al., 2019). As reported in the literature, only
31 a woefully small proportion of published research was actually reproducible before the pandemic (Iqbal et al.,
32 2016; Stodden et al., 2018), and the situation does not appear to have changed substantially since (Sumner
33 et al., 2020; Gustot, 2020).

34 The push for open software and data (e.g., Bivand, 2020; Arribas-Bel et al., 2021), along with more
35 strenuous efforts towards open, reproducible research, is simply a continuation of long-standing scientific
36 practices of independent verification. Despite the (at times disproportionate) attention that high profile
37 scandals in science tend to elicit in the media, science as a collective endeavor is remarkable for being a
38 self-correcting enterprise, one with built-in mechanisms and incentives to weed out erroneous ideas. Over the
39 long term, facts tend to prevail in science. At stake is the shorter-term impacts that research may have in
40 other spheres of economic and social life. The case of economists Reinhart and Rogoff comes to mind: by the
41 time the inaccuracies and errors in their research were uncovered (see Herndon et al., 2014), their claims
42 about debt and economic growth had already been seized by policy-makers on both sides of the Atlantic
43 to justify austerity policies in the aftermath of the Great Recession of 2007-2009¹. As later research has
44 demonstrated, those policies cast a long shadow, and their sequels continued to be felt for years (Basu et al.,
45 2017).

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

59 The objective of this paper is to illustrate the importance of reproducibility in research in the context of
60 the flood of COVID-19 papers. For this, I focus on a recent study by Sy et al. (2021) that examined the
61 correlation between the basic reproductive number of COVID-19, R_0 , and population density. The basic
62 reproductive number is a summary measure of contact rates, probability of transmission of a pathogen, and
63 duration of infectiousness. In rough terms, it measures how many new infections each infections begets.

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>

64 The paper of Sy et al. (2021) was selected for being, in the literature examined, almost alone in supporting
65 reproducible research. Accordingly, I wish to be clear that my objective in singling their work for discussion
66 is not to malign their efforts, but rather to demonstrate how open and reproducible research efforts can
67 greatly help to accelerate discovery. More concretely, open data and open code mean that an independent
68 researcher can, with only modest efforts, not only verify the findings reported, but also examine the same
69 data from a perspective which may not have been available to the original researchers due to differences
70 in disciplinary perspectives, methodological traditions, and/or training, among other possible factors. The
71 example, which shows consequential changes in the conclusions reached by different analyses, should serve as
72 a call to researchers to redouble their efforts to increase transparency and reproducibility in their research. In
73 this spirit, the present paper also aims to show how data can be packaged in well-documented, shareable units,
74 and code can be embedded into self-contained documents suitable for review and independent verification.
75 The source for this paper is an R Markdown document which, along with the data package, are available in
76 a public repository².

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

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

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

99 A plausible behavioral adaptation during a pandemic, especially one broadcast as widely and intensely as
100 COVID-19, is risk compensation. Risk compensation is a process whereby people adjust their behavior in
101 response to their *perception* of risk (Noland, 1995; Richens et al., 2000; Phillips et al., 2011). In the case
102 of COVID-19, Chauhan et al. (Chauhan et al., 2021) have found that perception of risks in the US varies
103 between rural, suburban, and urban residents, with rural residents in general expressing less concern about
104 the virus. It is possible that people who listened to the message of leaders saying that they were safe from the
105 virus because of low density may not have taken adequate precautions. Conversely, people in dense places
106 who could more directly observe the impact of the pandemic may have become overly cautious. Both Paez et
107 al. (2020) and Hamidi et al. (2020b) posit this mechanism (i.e., greater compliance with social distancing in
108 denser regions) to explain the results of their analyses. The evidence available does indeed show that there

²<https://github.com/paezha/Reproductive-Rate-and-Density-US-Reanalyzed>

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>

were important changes in behavior with respect to mobility during the pandemic (Jamal and Paez, 2020; Harris and Branon-Calles, 2021; 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 (Praharaj et al., 2020). In addition, there is evidence that changes in mobility correlated with the trajectory of the pandemic (Paez, 2020; Noland, 2021). 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 COVID-19 and population density at the district level in India, however their analysis was bivariate and did not control for other variables, such as income. Similarly, Kadi and Khelfaoui (2020) found a positive and significant correlation between number of cases and population density in cities in Algeria in a series of simple regression models (i.e., without other controls). A question in these relatively simple analyses is whether density is not a proxy for other factors. Other studies have included controls, such as Pequeno et al. (2020), a team that reported a positive association between density and cumulative counts of confirmed COVID-19 cases in state capitals in Brazil after controlling for covariates, including income, transport connectivity, and economic status. In a similar vein, Fielding-Miller et al. (2020) reported a positive relationship between the absolute number of COVID-19 deaths and population density (rate) in rural counties in the US. Roy and Ghosh (2020) used a battery of machine learning techniques to find discriminatory factors, and a positive and significant association between COVID-19 infection and death rates in US states. Wong and Li (2020) also found a positive and significant association between population density and number of confirmed COVID-19 cases in US counties, using both univariate and multivariate regressions with spatial effects. More recently, Sy et al. (2021) reported that the basic reproductive number of COVID-19 in US counties tended to increase with population density, but at a decreasing rate at higher densities.

On the flip side, a number of studies report non-significant or negative associations between population density and COVID-19 outcomes. This includes the research of Sun et al. (2020) who did not find evidence of significant correlation between population density and confirmed number of cases per day *in conditions of lockdown* in China. This finding echoes the results of Paez et al. (2020), who in their study of provinces in Spain reported non-significant associations between population density and infection rates in the early days of the first wave of COVID-19, and negative significant associations in the later part of the first lockdown. Similarly, Skórka et al. (2020) found zero or negative associations between population density and infection numbers/deaths by country. Fielding-Miller et al. (2020) contrast their finding about rural counties with a negative relationship between COVID-19 deaths and population density in urban counties in the US. For their part, in their investigation of doubling time, White and Hébert-Dufresne (2020) identified a negative and significant correlation between population density and doubling time in US states. Likewise, Khavarian-Garmsir et al. (2021) found a small negative (and significant) association between population density and COVID-19 morbidity in districts in Tehran. Finally, two of the most complete studies in the US, by Hamidi et al. (2020a) and Hamidi et al. (2020b), used an extensive set of controls to find negative and significant correlations between density and COVID-19 cases and fatalities at the level of counties in the US.

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

159 **Reproducibility of research**

160 SWN investigated the basic reproductive number of COVID-19 in US counties, and its association with
161 population density, median household income, and prevalence of private mobility. For their multivariate
162 analysis, SWN used mixed linear models. This is an appropriate modelling choice: R_0 is an interval-ratio
163 variable that is suitably modeled using linear regression; further, as SWN note there is a likelihood that the
164 process is not independent “among counties within each state, potentially due to variable resource allocation
165 and differing health systems across states” (p. 3). A mixed linear model accounts for this by introducing
166 random components; in the case of SWN, these are random intercepts at the state level. SWN estimated
167 various models with different combinations of variables, including median household income and prevalence
168 of travel by private transportation. These controls help to account for potential variations in behavior:
169 people in more affluent counties may have greater opportunities to work from home, and use of private
170 transportation reduces contact with strangers. Moreover, they also conducted various sensitivity analyses.
171 After these efforts, SWN concluded that there is a positive association between the basic reproductive number
172 and population density at the level of counties in the US.

173 One salient aspect of the analysis in SWN is that the basic reproductive number can only be calculated
174 reliably with a minimum number of cases, and a large number of counties did not meet such threshold. As
175 researchers do, SWN made modelling decisions, in this case basing their analysis only on counties with valid
176 observations. A modeler with expertise in spatial analysis would likely ask some of the following questions on
177 reading SWN’s paper: how were missing counties treated? What are the implications of the spatial sampling
178 framework used in the analysis? Is it possible to spatially interpolate the missing observations? Was there
179 spatial residual autocorrelation in the models, or was the use of mixed models sufficient to capture spatial
180 dependencies? These questions are relevant and their implications important. Fortunately, SWN are an
181 example of a reasonably open, reproducible research product: their paper is accompanied by (most of) the
182 data and (most of) the code used in the analysis. This means that an independent researcher can, with
183 only a moderate investment of time and effort, reproduce the results in the paper, as well as ask additional
184 questions.

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

186 There are various reasons why a project can fail to be reproducible. In some cases, there might be
187 legitimate reasons to withhold the data, perhaps due to confidentiality and privacy reasons (e.g., Lee et al.,
188 2020). But in many other cases the data are publicly available, which in fact has commonly been the case
189 with population-level COVID-19 information. Typically the provenance of the data is documented, but in
190 numerous studies the data themselves are not shared (Amadu et al., 2021; Bhadra et al., 2021; Cruz et al.,
191 2020; Feng et al., 2020; Fielding-Miller et al., 2020; Hamidi et al., 2020a,b; Inbaraj et al., 2021; Souris and
192 Gonzalez, 2020). As any researcher can attest, collecting, organizing, and preparing data for a project can
193 take a substantial amount of time. Pointing to the sources of data, even when these sources are public, is a
194 small step towards reproducibility—but only a very small one. Faced with the prospect of having to recreate
195 a data set from raw sources is probably sufficient to dissuade all but the most dedicated (or stubborn)
196 researcher from independent verification. This is true even if part of the data are shared (e.g., Wong and Li,
197 2020). In other cases, data are shared, but the processes followed in the preparation of the data are not fully
198 documented (Ahmad et al., 2020; Skórka et al., 2020). These processes matter, as shown by the errors in the
199 spreadsheets of Reinhart and Rogoff (see Herndon et al., 2014, for the discovery of these errors), as well
200 as by the data of biologist Jonathan Pruitt that led to an “avalanche” of paper retractions (see Viglione,
201 2020). Another situation is when papers share well-documented data, but fail to provide the code used in
202 the analysis (Noury et al., 2021; Pequeno et al., 2020; Wang et al., 2021). Making code available only “on
203 demand” (e.g., Brandtner et al., 2021) is an unnecessary barrier when most journals offer the facility to share
204 supplemental materials online. Then there are those papers that more closely comply with reproducibility
205 standards, and share well-documented processes and data, as well as the code used in any analyses reported
206 (Paez et al., 2020; Feyman et al., 2020; Stephens et al., 2021; White and Hébert-Dufresne, 2020; Sy et al.,
207 2021). Even in this case, the pressure to publish “new findings” instead of replication studies can act as a

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]

208 deterrent, perhaps particularly for younger researchers⁴.

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

211 Reproducing SWN

212 SWN examined the association between the basic reproductive number of COVID-19 and population
213 density. The basic reproductive number R_0 is a summary measure of contact rates, probability of transmission
214 of a pathogen, and duration of infectiousness. In rough terms, R_0 measures how many new infections each
215 infections begets. Infectious disease outbreaks generally tend to die out when $R_0 < 1$, and to grow when
216 $R_0 > 1$. Reliable calculation of R_0 requires a minimum number of cases to be able to assume that there
217 is community transmission of the pathogen. Accordingly, SWN based their analysis only on counties that
218 had at least 25 cases or more at the end of the exponential growth phase (see Fig. 1). Their final sample
219 included 1,151 counties in the US, including in Alaska, Hawaii, Puerto Rico, and island territories. SWN
220 used COVID-19 data collected by the New York Times and made available (with versioning) in a GitHub
221 repository⁵. For each county, SWN assumed that the exponential growth period began one week prior to the
222 second daily increase in cases, and assumed that the period of exponential growth lasted approximately 18
223 days.

224 Table 1 reproduces the first three models of SWN (the fourth model did not have any significant variables;
225 see Table 1 in SWN). It is possible to verify that the results match, with only the minor (and irrelevant)
226 exception of the magnitude of the coefficient for travel by private transportation, which is due to a difference
227 in the input (here the variable is changed to one percent units, instead of the ten percent units used by
228 SWN). The mixed linear model gives random intercepts (i.e., the intercept is a random variable), and the
229 standard deviation is reported in the fifth row of Table 1. It is useful to map the random intercepts: as seen
230 in Figure 2, other things being equal, counties in Texas tend to have somewhat lower values of R_0 (i.e., a
231 negative random intercept), whereas counties in South Dakota tend to have higher values of R_0 . The key of
232 the analysis, after extensive sensitivity analysis, is a robust finding that population density has a positive
233 association with the basic reproductive number. But does it?

234 Expanding on SWN

235 The preceding section shows that thanks to the availability of code and data, it is possible to verify
236 the results reported by SWN. As noted earlier, though, an independent researcher might have wondered

⁴The 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.

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

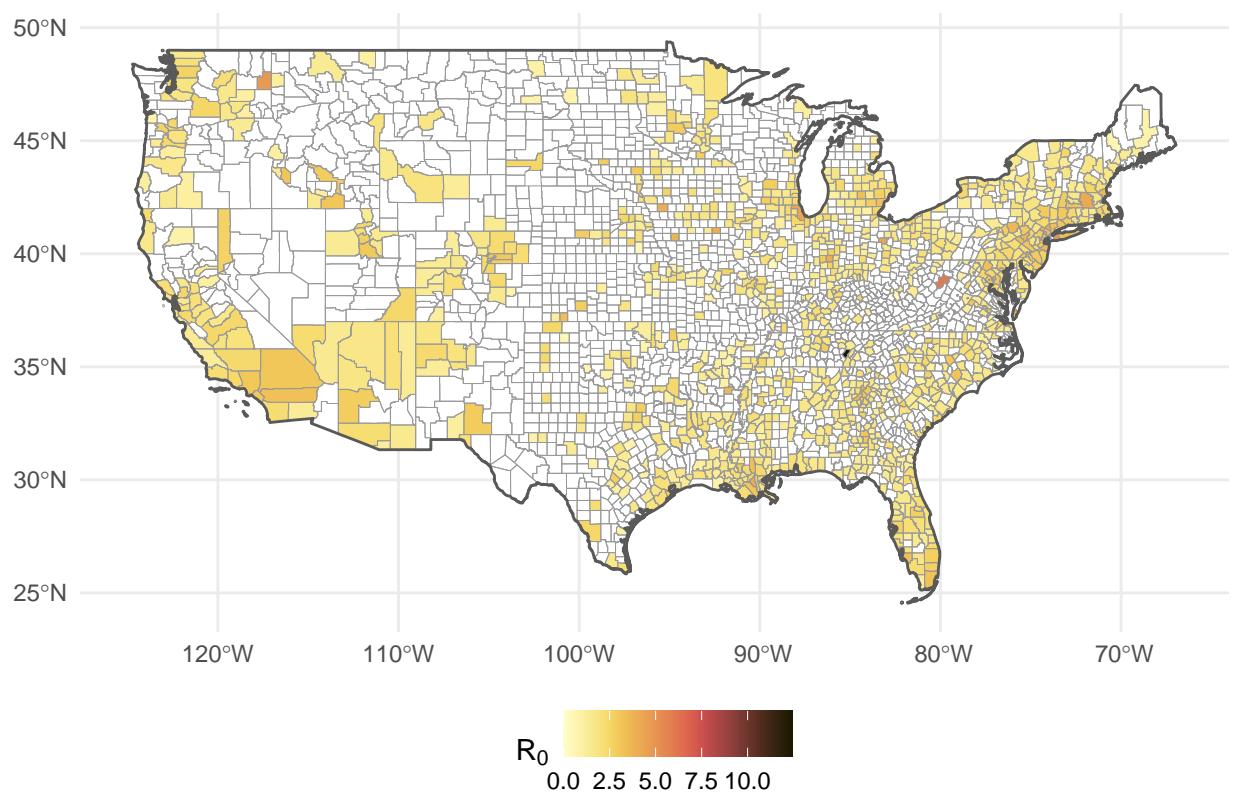


Figure 1: Basic reproductive rate in US counties (Alaska, Hawaii, Puerto Rico, and territories not shown).

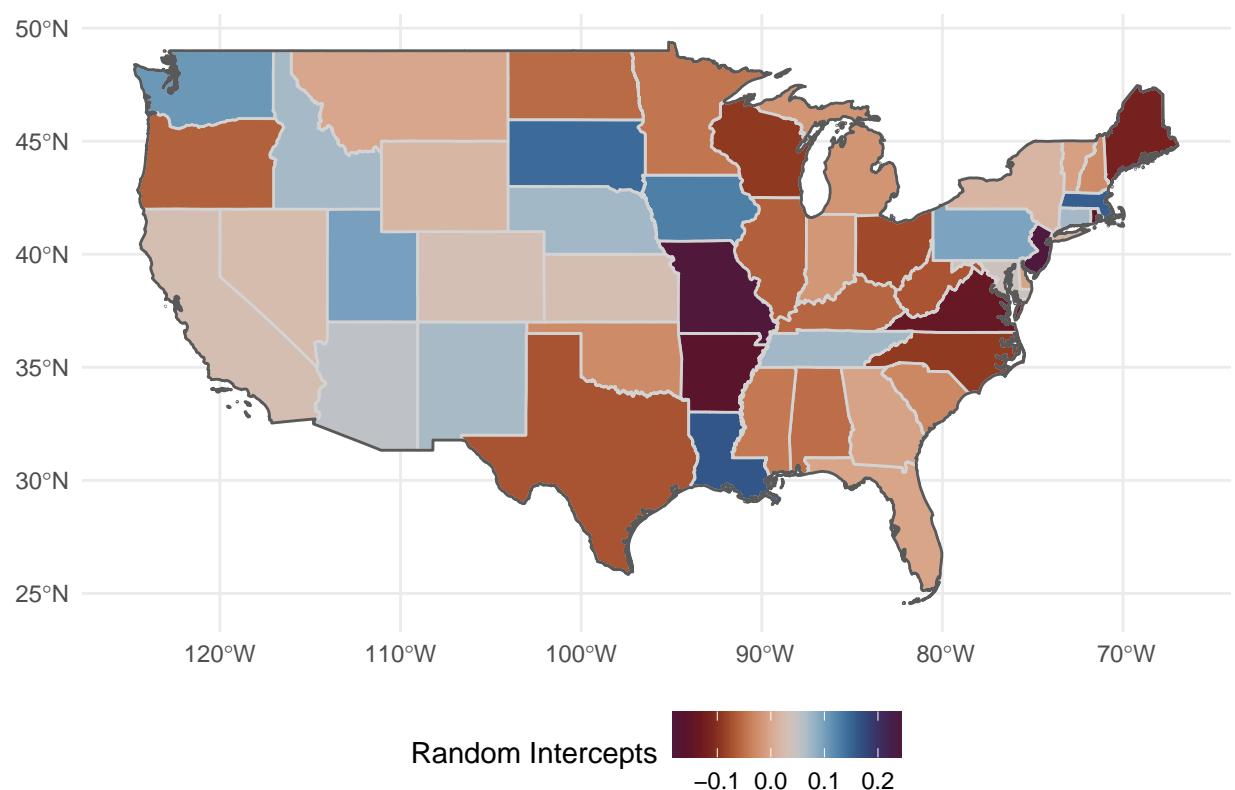
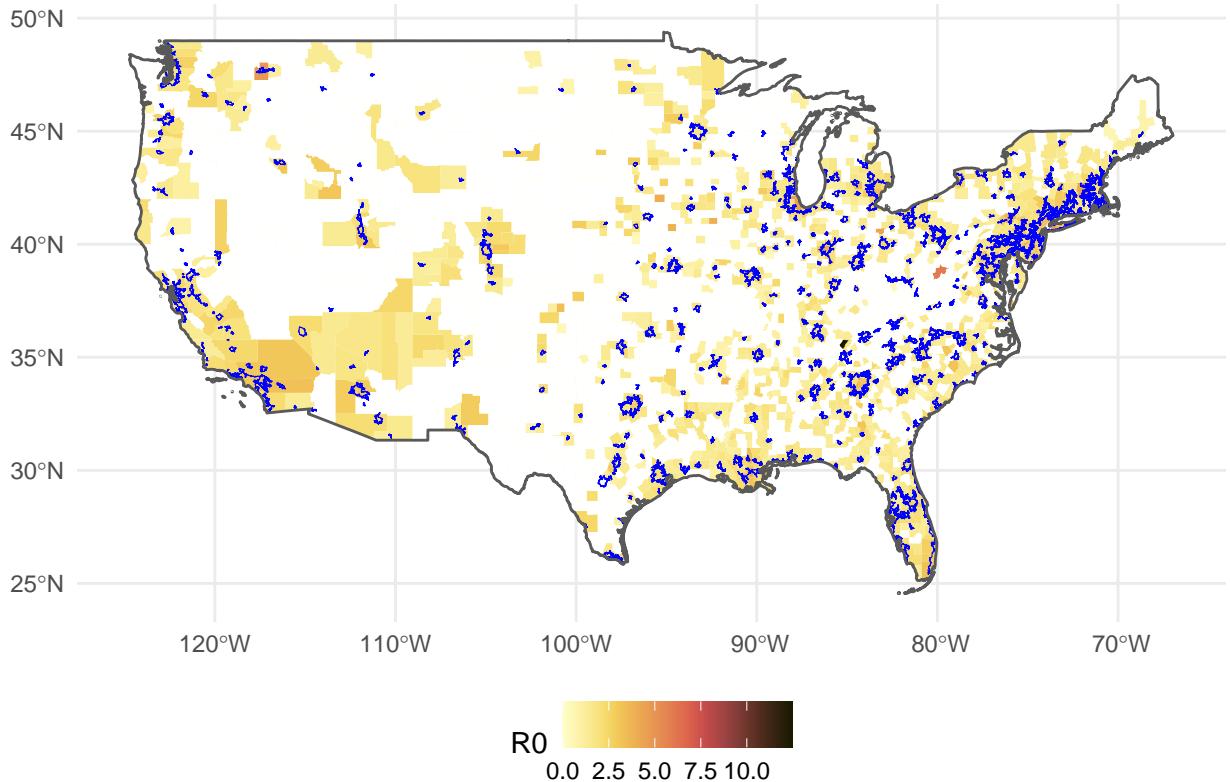


Figure 2: Random intercepts of Model 3 (Alaska, Hawaii, Puerto Rico, and territories not shown).



Note: boundaries of urbanized areas with population > 50,000 are shown in blue

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

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.

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

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

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

Proceed with caution: spatial effects ahead

The results of the selection models, in particular Selection Model 2, make us reassess the original conclusion that density has a positive association with the basic reproductive number of COVID-19. A spatial analyst might still wonder about spatial residual autocorrelation. A challenge here is that spatial models tend to be technically more demanding, and although spatial models for qualitative variables exist, a spatial implementation of the sample selection model does not appear to exist. It might be argued that

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]

292 a reproducible research project can also allow a researcher to be more adventurous with their modeling
 293 decisions: since data and code are shared, other researchers can promptly and with relative ease poke the
 294 methods and see if they appear to be sound.

295 In the present case, it appears that an application of spatial filtering (see Getis and Griffith, 2002; Paez,
 296 2019; Griffith, 2004) can help. Spatial filtering provides an elegant solution to regression problems that may
 297 have difficulties handling the spatial structures of spatial statistical and econometric models (Griffith, 2000).
 298 A key issue in the present example is the fact that there are numerous missing observations, which prevents
 299 the calculation of autocorrelation statistics, let alone the estimation of models with spatial components.

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

- 301 1. Estimate a sample selection model and retrieve the residuals of the outcome. This will be a vector
 302 with missing values for locations that were not sampled.
 303 2. Fit a spatial filter to the residuals. This is done by regressing the estimated residuals of the *observed*
 304 data on the corresponding values of the Moran eigenvectors.
 305 3. The resulting filter will correlate highly with the known residuals, and will provide information in
 306 non-sampled locations that is consistent with the spatial pattern of the known residuals.
 307 4. Test the filter for spatial autocorrelation:
 308 4.1 If significant spatial autocorrelation is detected, this would be indicative of residual spatial pattern.
 309 Introduce the filter as a covariate in the outcome model of the sample selection model and return to
 310 step 1.
 311 4.2 If no significant spatial autocorrelation is detected, this would be indicative of random residual
 312 pattern. Stop.

313 This procedure is implemented using a stopping criterion whereby the search for the filter only stops when
 314 the p-value of Moran's Coefficient of the filter fitted to the residuals is greater than 0.25, which was chosen as
 315 a sufficiently conservative value for testing for autocorrelation. The correlation of the known residuals with

316 the corresponding elements of the filter is consistently high (the correlation coefficient typically is greater
 317 than 0.9). The results of implementing this procedure appear in Table 3 as Selection Model 3. The results
 318 are consistent with Selection Model 2, with two intriguing differences: 1) the variance of Sample Model 3
 319 is smaller; and 2) the sample and outcome processes are no longer correlated (the confidence interval of ρ
 320 includes zero). It appears that by capturing the spatial pattern of the residuals, which is likely strongly
 321 determined by the non-random sampling framework, the outcome model is not only substantially more
 322 precise, but also appears to be independent from the selection process.

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

Variable	Selection Model 3	
	β	95% CI
Sample Selection Model		
Intercept	-7.249	[-8.285, -6.214]
Density (1,000 per sq.km)	2.424	[2.074, 2.774]
Density squared	-0.373	[-0.459, -0.288]
Percent of private transportation	0.056	[0.045, 0.066]
Median household income (10,000)	0.319	[0.282, 0.356]
Outcome Model		
Intercept	2.290	[2.026, 2.553]
Density (1,000 per sq.km)	0.843	[0.786, 0.9]
Density squared	-0.142	[-0.153, -0.131]
Percent of private transportation	-0.010	[-0.012, -0.008]
Median household income (\$10,000)	0.011	[0.003, 0.02]
Spatial filter	1.001	[0.992, 1.011]
σ	0.120	[0.107, 0.133]
ρ	0.495	[0.218, 0.772]

323 Clearly, the various models display some intriguing differences; but how relevant are said differences from
 324 a more substantive standpoint? Figure 4 shows the relationship between density and R_0 implied by SWN
 325 Model 3, Selection Model 2, and Selection Model 3. The left panel of the figure shows the non-linear but
 326 monotonic relationship implied by SWN Model 1. The conclusion is that at higher densities, R_0 is *always*
 327 higher. The two panels on the right, in contrast, shows that Selection Model 2 and Selection Model 3 coincide
 328 that R_0 tends to increase as density grows. This continues until a density of approximately 2.9 (1,000 people
 329 per sq.km). At higher densities than that the relationship between density and R_0 begins to weaken, and the
 330 relationship becomes negative at densities higher than approximately 5.7 (1,000 people per sq.km).

331 To put this into context, other things being equal, the effect of density in a county like Charlottesville in
 332 Virginia (density ~1,639 people per sq.km) is roughly the same as that in a county like Philadelphia (density
 333 ~4,127 people per sq.km). In contrast, the effect of density on R_0 in a county like Arlington in Virginia
 334 (density ~3,093 people per sq.km) is *stronger* than either of the previous two examples. Lastly, the density of
 335 counties like San Francisco in California, or Queens and Bronx in NY, which are among the densest in the
 336 US, contributes even less to R_0 than even the most rural counties in the country.

337 Discussion

338 It is worth at this point to recall Cressie's dictum about modelling: "[w]hat is one person's mean structure
 339 could be another person's correlation structure" (Cressie, 1989, p. 201). There are almost always multiple
 340 ways to approach a modelling situation, as lively illustrated by a recent paper that reports the results of a
 341 crowdsourced modelling experiment (Schweinsberg et al., 2021). In the present case, we would argue that
 342 spatial sampling is an important aspect of the modeling process. Importantly, by adopting high reproducibility
 343 standards, SWN made a valuable contribution to the collective enterprise of seeking knowledge. Their effort,
 344 and subsequent efforts to validate and expand on their work, can potentially contribute to provide clarity to
 345 ongoing conversations about the relevance of density and the spread of COVID-19.

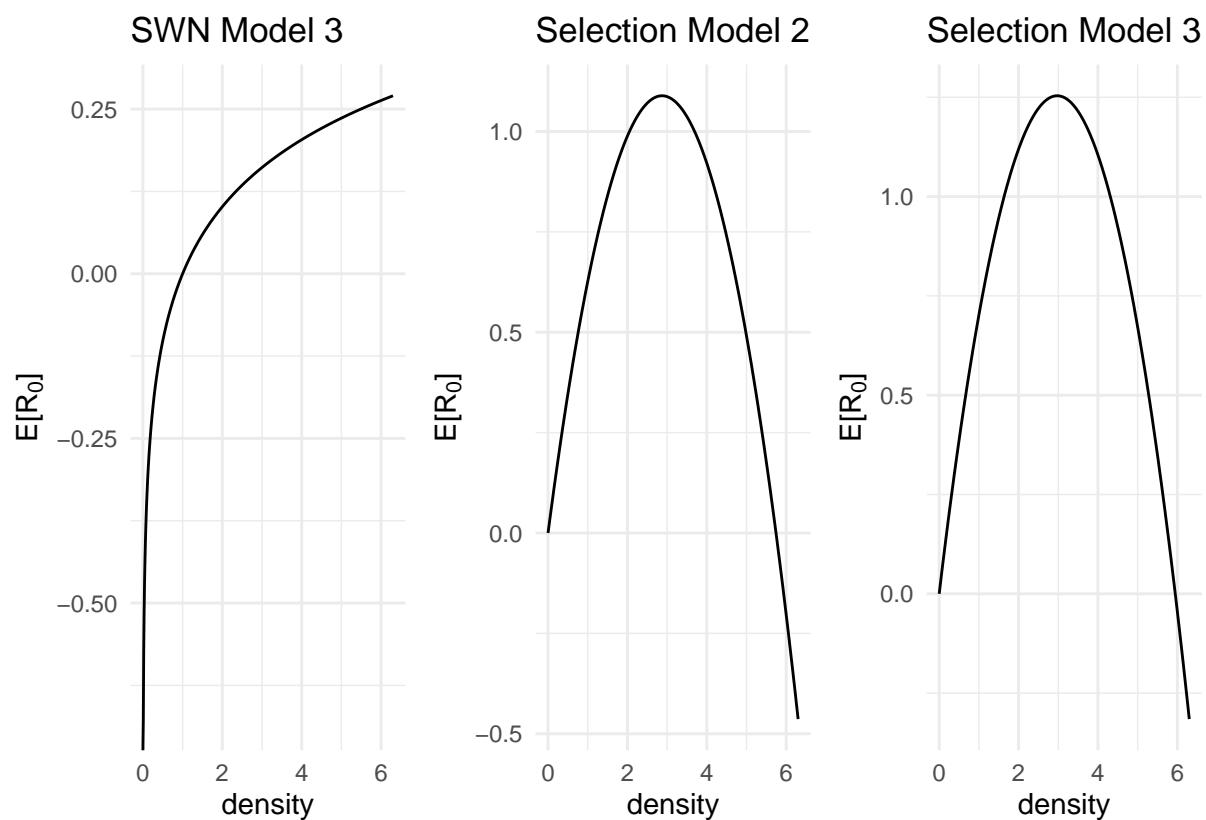


Figure 4: Effect of density according to SWN Model 3 and Sample Selection Model 2.

346 In particular, it is noteworthy that a sample selection model with a different variable transformation
347 does not lend support to the thesis that higher density is *always* associated with a greater risk of spread
348 of the virus [in Wong and Li's words, "‘Density is destiny’ is probably an overstatement"; -Wong and Li
349 (2020)]. At the same time, the results presented here also stand in contrast to the findings of Hamidi et
350 al., who found that higher density was either not significantly associated with the rate of the virus in a
351 cross-sectional study (Hamidi et al., 2020b), or was negatively associated with it in a longitudinal setting
352 [Hamidi et al. (2020a)]. In this sense, the conclusion that density does not aggravate the pandemic may have
353 been somewhat premature; instead, reanalysis of the data of SWN suggests that Fielding-Miller et al. (2020)
354 might be onto something with respect to the difference between rural and urban counties. More generally,
355 there is no doubt that in population-level studies density is indicative of proximity, but it also potentially is
356 a proxy for adaptive behavior. And it is possible that the determining factor during COVID-19, at least in
357 the US, has been variations in perceptions of the risks associated with contagion (Chauhan et al., 2021), and
358 subsequent compensations in behavior in more and less dense regions.

359 Conclusion

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

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

379 More generally, my hope is that the research of Sy et al. (2021), the present paper, and similar reproducible
380 publications, will continue to encourage others to adopt higher reproducibility standards in their research.

381 Acknowledgments

382 The analysis reported in this paper was conducted in the R computing statistical language (?). The source
383 document is an Rmarkdown document (??) processed using knitr (??). The following packages were used
384 in the analysis, and I wish to acknowledge their creators for their generous efforts: adespatial (?), censReg
385 (?), dplyr (?),forcats (?), ggplot2 (?), gmm (?), kableExtra (?), Matrix (?), maxLik (?), miscTools
386 (?), mvtnorm (?), nlme (?), patchwork (?), purrr (?), readr (?), sampleSelection (?), sandwich (???),
387 scico (?), sf (?), sp (??), spatialprobit (?), spData (?), spdep (??), stringr [R-stringr], tibble (?),
388 tidyCensus (?), tidyR (?), tidyverse (?), tmvtnorm (?), units (?). This research was not supported by
389 Canada’s Research Councils.

390 **References**

- 391 Khansa Ahmad, Sehat Erqou, Nishant Shah, Umair Nazir, Alan R. Morrison, Gaurav Choudhary, and Wen-Chih Wu.
392 Association of poor housing conditions with covid-19 incidence and mortality across us counties. *PLOS ONE*, 15(11):e0241327,
393 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0241327. URL <https://dx.doi.org/10.1371/journal.pone.0241327>.
- 394 Iddrisu Amadu, Bright Opoku Ahinkorah, Abdul-Rahaman Afitiri, Abdul-Aziz Seidu, Edward Kwabena Ameyaw, John Elvis
395 Hagan, Eric Duku, and Simon Appah Aram. Assessing sub-regional-specific strengths of healthcare systems associated
396 with covid-19 prevalence, deaths and recoveries in africa. *PLOS ONE*, 16(3):e0247274, 2021. ISSN 1932-6203. doi:
397 10.1371/journal.pone.0247274. URL <https://dx.doi.org/10.1371/journal.pone.0247274>.
- 398 Daniel Arribas-Bel, Mark Green, Francisco Rowe, and Alex Singleton. Open data products: a framework for creating valuable
399 analysis-ready data. *Journal of Geographical Systems*, 2021. ISSN 1435-5930.
- 400 Diego Añazco, Bryan Nicolalde, Isabel Espinosa, Jose Camacho, Mariam Mushtaq, Jimena Gimenez, and Enrique Teran.
401 Publication rate and citation counts for preprints released during the covid-19 pandemic: the good, the bad and the ugly.
402 *PeerJ*, 9:e10927, 2021. ISSN 2167-8359. doi: 10.7717/peerj.10927. URL <https://dx.doi.org/10.7717/peerj.10927>.
- 403 Sanjay Basu, Megan A. Carney, and Nora J. Kenworthy. Ten years after the financial crisis: The long reach of austerity and its
404 global impacts on health. *Social Science & Medicine*, 187:203–207, 2017. ISSN 0277-9536. doi: 10.1016/j.socscimed.2017.06.026.
405 URL <https://dx.doi.org/10.1016/j.socscimed.2017.06.026>.
- 406 Arunava Bhadra, Arindam Mukherjee, and Kabita Sarkar. Impact of population density on covid-19 infected and mortality rate
407 in india. *Modeling Earth Systems and Environment*, 7(1):623–629, 2021. ISSN 2363-6203. doi: 10.1007/s40808-020-00984-7.
408 URL <https://dx.doi.org/10.1007/s40808-020-00984-7>.
- 409 Roger S. Bivand. Progress in the r ecosystem for representing and handling spatial data. *Journal of Geographical Systems*, 2020.
410 ISSN 1435-5930. doi: 10.1007/s10109-020-00336-0. URL <https://dx.doi.org/10.1007/s10109-020-00336-0>.
- 411 Katrina A. Bramstedt. The carnage of substandard research during the covid-19 pandemic: a call for quality. *Journal of Medical
412 Ethics*, 46(12):803–807, 2020. ISSN 0306-6800. doi: 10.1136/medethics-2020-106494. URL <https://dx.doi.org/10.1136/medethics-2020-106494>.
- 413 Christof Brandtner, Luís M. A. Bettencourt, Marc G. Berman, and Andrew J. Stier. Creatures of the state? metropolitan
414 counties compensated for state inaction in initial u.s. response to covid-19 pandemic. *PLOS ONE*, 16(2):e0246249, 2021.
415 ISSN 1932-6203. doi: 10.1371/journal.pone.0246249. URL <https://dx.doi.org/10.1371/journal.pone.0246249>.
- 416 F. Broggini, J. Dellinger, S. Fomel, and Y. Liu. Reproducible research: Geophysics papers of the future - introduction.
417 *Geophysics*, 82(6), 2017. ISSN 0016-8033. doi: 10.1190/geo2017-0918-spseintro.1. URL <[Goto ISI](https://doi.org/10.1190/geo2017-0918-spseintro.1)>://WOS:000417973300069.
- 418 Chris Brunsdon and Alexis Comber. Opening practice: supporting reproducibility and critical spatial data science. *Journal of
419 Geographical Systems*, 2020. ISSN 1435-5949. doi: 10.1007/s10109-020-00334-2. URL <https://doi.org/10.1007/s10109-020-00334-2>.
- 420 Rishabh Singh Chauhan, Denise Capasso da Silva, Deborah Salon, Ali Shamshiripour, Ehsan Rahimi, Uttara Sutradhar, Sara
421 Khoaini, Abolfazl (Kourous) Mohammadian, Sybil Derrible, and Ram Pendyala. Covid-19 related attitudes and risk perceptions
422 across urban, rural, and suburban areas in the united states. *Findings*, 6 2021. doi: 10.32866/001c.23714.
- 423 Noel Cressie. Geostatistics. *The American Statistician*, 43(4):197, 1989. ISSN 0003-1305. doi: 10.2307/2685361. URL
424 <https://dx.doi.org/10.2307/2685361>.
- 425 Christian Joy Pattawi Cruz, Rachel Ganly, Zilin Li, and Stuart Gietel-Basten. Exploring the young demographic profile of
426 covid-19 cases in hong kong: Evidence from migration and travel history data. *PLOS ONE*, 15(6):e0235306, 2020. ISSN
427 1932-6203. doi: 10.1371/journal.pone.0235306. URL <https://dx.doi.org/10.1371/journal.pone.0235306>.
- 428 Steven Farber and Antonio Páez. Running to stay in place: the time-use implications of automobile oriented land-use and
429 travel. *Journal of Transport Geography*, 19(4):782–793, 2011. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2010.09.008. URL
430 <http://www.sciencedirect.com/science/article/pii/S0966692310001559>.
- 431 Yongjiu Feng, Qingmei Li, Xiaohua Tong, Rong Wang, Shuting Zhai, Chen Gao, Zhenkun Lei, Shurui Chen, Yilun Zhou, Jiafeng
432 Wang, Xiongfeng Yan, Huan Xie, Peng Chen, Shijie Liu, Xiong Xv, Sicong Liu, Yanmin Jin, Chao Wang, Zhonghua Hong,
433 Kuifeng Luan, Chao Wei, Jinfu Xu, Hua Jiang, Changjiang Xiao, and Yiyou Guo. Spatiotemporal spread pattern of the
434 covid-19 cases in china. *PLOS ONE*, 15(12):e0244351, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0244351. URL
435 <https://dx.doi.org/10.1371/journal.pone.0244351>.
- 436 Yevgeniy Feyman, Jacob Bor, Julia Raifman, and Kevin N. Griffith. Effectiveness of covid-19 shelter-in-place orders varied by
437 state. *PLOS ONE*, 15(12):e0245008, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0245008. URL <https://dx.doi.org/10.1371/journal.pone.0245008>.
- 438 Rebecca K. Fielding-Miller, Maria E. Sundaram, and Kimberly Brouwer. Social determinants of covid-19 mortality at the
439 county level. *PLOS ONE*, 15(10):e0240151, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0240151. URL <https://dx.doi.org/10.1371/journal.pone.0240151>.
- 440 R Florida, E Glaeser, MM Sharif, K Bedi, TJ Campanella, CH Chee, D Doctoroff, B Katz, R Katz, and J Kotkin. How life in
441 our cities will look after the coronavirus pandemic. *Foreign Policy*, 1, 2020. URL <https://foreignpolicy.com/2020/05/01/future-of-cities-urban-life-after-coronavirus-pandemic/>.
- 442 Nicholas Fraser, Liam Brierley, Gautam Dey, Jessica K. Polka, Máté Pálfy, Federico Nanni, and Jonathon Alexis Coates.
443 The evolving role of preprints in the dissemination of covid-19 research and their impact on the science communication
444 landscape. *PLOS Biology*, 19(4):e3000959, 2021. ISSN 1545-7885. doi: 10.1371/journal.pbio.3000959. URL <https://dx.doi.org/10.1371/journal.pbio.3000959>.
- 445 A. Getis and D. A. Griffith. Comparative spatial filtering in regression analysis. *Geographical Analysis*, 34(2):130–140, 2002.
446 ISSN 0016-7363.
- 447 Jonatan Gomez, Jeisson Prieto, Elizabeth Leon, and Arles Rodríguez. Infekta—an agent-based model for transmission of

- 454 infectious diseases: The covid-19 case in bogotá, colombia. *PLOS ONE*, 16(2):e0245787, 2021. ISSN 1932-6203. doi:
 455 10.1371/journal.pone.0245787. URL <https://dx.doi.org/10.1371/journal.pone.0245787>.
- 456 D. A. Griffith. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, 2(2):
 457 141–156, 2000.
- 458 D. A. Griffith. A spatial filtering specification for the autologistic model. *Environment and Planning A*, 36(10):1791–1811, 2004.
 459 URL ISI:000224552100007C:/Papers/EnvironmentandPlanningA/EnvironmentandPlanningA(2004)36(10)791-181.pdf.
- 460 T. Gustot. Quality and reproducibility during the covid-19 pandemic. *JHEP Rep*, 2(4):100141, 2020. ISSN 2589-5559. doi:
 461 10.1016/j.jhepr.2020.100141.
- 462 Shima Hamidi and Ahoura Zandiataashbar. Compact development and adherence to stay-at-home order during the covid-19
 463 pandemic: A longitudinal investigation in the united states. *Landscape and Urban Planning*, 205:103952, 2021. ISSN 0169-
 464 2046. doi: <https://doi.org/10.1016/j.landurbplan.2020.103952>. URL <https://www.sciencedirect.com/science/article/pii/S016920462030712X>.
- 465 Shima Hamidi, Reid Ewing, and Sadegh Sabouri. Longitudinal analyses of the relationship between development density and
 466 the covid-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the united states. *Health &
 467 Place*, 64:102378, 2020a. ISSN 1353-8292. doi: 10.1016/j.healthplace.2020.102378. URL <https://dx.doi.org/10.1016/j.healthplace.2020.102378>.
- 468 Shima Hamidi, Sadegh Sabouri, and Reid Ewing. Does density aggravate the covid-19 pandemic? *Journal of the American
 469 Planning Association*, 86(4):495–509, 2020b. ISSN 0194-4363. doi: 10.1080/01944363.2020.1777891. URL <https://dx.doi.org/10.1080/01944363.2020.1777891>.
- 470 M. Anne Harris and Michael Braniion-Calles. Changes in commute mode attributed to covid-19 risk in canadian national survey
 471 data. *Findings*, 2 2021. doi: 10.32866/001c.19088.
- 472 T. Herndon, M. Ash, and R. Pollin. Does high public debt consistently stifle economic growth? a critique of reinhart
 473 and rogoff. *Cambridge Journal of Economics*, 38(2):257–279, 2014. ISSN 0309-166X. doi: 10.1093/cje/bet075. URL
 474 <https://dx.doi.org/10.1093/cje/bet075>.
- 475 Leeberk Raja Inbaraj, Caroline Elizabeth George, and Sindhuolina Chandrasingh. Seroprevalence of covid-19 infection in a rural
 476 district of south india: A population-based seroepidemiological study. *PLOS ONE*, 16(3):e0249247, 2021. ISSN 1932-6203.
 477 doi: 10.1371/journal.pone.0249247. URL <https://dx.doi.org/10.1371/journal.pone.0249247>.
- 478 D. C. Ince, L. Hatton, and J. Graham-Cumming. The case for open computer programs. *Nature*, 482(7386):485–488, 2012.
 479 ISSN 0028-0836. doi: 10.1038/nature10836. URL <Go to ISI>://WOS:000300770500042.
- 480 J. P. A. Ioannidis, S. Greenland, M. A. Hlatky, M. J. Khoury, M. R. Macleod, D. Moher, K. F. Schulz, and R. Tibshirani.
 481 Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*, 383(9912):166–175, 2014. ISSN
 482 0140-6736. doi: 10.1016/s0140-6736(13)62227-8. URL <Go to ISI>://WOS:000329579600032.
- 483 S. A. Iqbal, J. D. Wallach, M. J. Khoury, S. D. Schully, and J. P. A. Ioannidis. Reproducible research practices and transparency
 484 across the biomedical literature. *Plos Biology*, 14(1), 2016. ISSN 1545-7885. doi: 10.1371/journal.pbio.1002333. URL
 485 <Go to ISI>://WOS:000371882900003.
- 486 Shaila Jamal and Antonio Paez. Changes in trip-making frequency by mode during covid-19. *Findings*, 11 2020. doi:
 487 10.32866/001c.17977.
- 488 Nadjat Kadi and Mounia Khelfaoui. Population density, a factor in the spread of covid-19 in algeria: statistic study.
 489 *Bulletin of the National Research Centre*, 44(1), 2020. ISSN 2522-8307. doi: 10.1186/s42269-020-00393-x. URL <https://dx.doi.org/10.1186/s42269-020-00393-x>.
- 490 Amir Reza Khavarian-Garmsir, Ayyoob Sharifi, and Nabi Moradpour. Are high-density districts more vulnerable to the
 491 covid-19 pandemic? *Sustainable Cities and Society*, 70:102911, 2021. ISSN 2210-6707. doi: 10.1016/j.scs.2021.102911. URL
 492 <https://dx.doi.org/10.1016/j.scs.2021.102911>.
- 493 M. Konkol and C. Kray. In-depth examination of spatiotemporal figures in open reproducible research. *Cartography and
 494 Geographic Information Science*, 46(5):412–427, 2019. ISSN 1523-0406. doi: 10.1080/15230406.2018.1512421. URL
 495 <Go to ISI>://WOS:000475449900003.
- 496 M. Konkol, C. Kray, and M. Pfeiffer. Computational reproducibility in geoscientific papers: Insights from a series of studies
 497 with geoscientists and a reproduction study. *International Journal of Geographical Information Science*, 33(2):408–429, 2019.
 498 ISSN 1365-8816. doi: 10.1080/13658816.2018.1508687. URL <Go to ISI>://WOS:000450550600009.
- 499 Diana Kwon. How swamped preprint servers are blocking bad coronavirus research. *Nature*, 581(7807):130–132, 2020. ISSN
 500 0028-0836.
- 501 Minha Lee, Jun Zhao, Qianqian Sun, Yixuan Pan, Weiyi Zhou, Chenfeng Xiong, and Lei Zhang. Human mobility trends during
 502 the early stage of the covid-19 pandemic in the united states. *PLOS ONE*, 15(11):e0241468, 2020. ISSN 1932-6203. doi:
 503 10.1371/journal.pone.0241468. URL <https://dx.doi.org/10.1371/journal.pone.0241468>.
- 504 Ruiqi Li, Peter Richmond, and Bertrand M. Roehner. Effect of population density on epidemics. *Physica A: Statistical
 505 Mechanics and its Applications*, 510:713–724, 2018. ISSN 0378-4371. doi: 10.1016/j.physa.2018.07.025. URL <https://dx.doi.org/10.1016/j.physa.2018.07.025>.
- 506 G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, 1983.
- 507 Sarah Micallef, Tonio V. Piscopo, Ramon Casha, Denise Borg, Chantal Vella, Maria-Alessandra Zammit, Janice Borg, Daniela
 508 Mallia, James Farrugia, Sarah Marie Vella, Thelma Xerri, Anette Portelli, Manuel Fenech, Claudia Fsadni, and Charles
 509 Mallia Azzopardi. The first wave of covid-19 in malta; a national cross-sectional study. *PLOS ONE*, 15(10):e0239389, 2020.
 510 ISSN 1932-6203. doi: 10.1371/journal.pone.0239389. URL <https://dx.doi.org/10.1371/journal.pone.0239389>.
- 511 Joseph Molloy, Christopher Tchervenkov, Beat Hintermann, and Kay W. Axhausen. Tracing the sars-cov-2 impact: The first
 512 month in switzerland. *Findings*, 5 2020. doi: 10.32866/001c.12903.
- 513 E. G. Moore. Some spatial properties of urban contact fields. *Geographical Analysis*, 2(4):376–386, 1970.

- 519 E. G. Moore and L. A. Brown. Urban acquaintance fields: an evaluation of a spatial model. *Environment and Planning*, 2(4):
 520 443–454, 1970. URL <http://www.envplan.com/abstract.cgi?id=a020443>.
- 521 R. B. Noland. Perceived risk and modal choice - risk compensation in transportation system. *Accident Analysis and Prevention*,
 522 27(4):503–521, 1995. ISSN 0001-4575. doi: 10.1016/0001-4575(94)00087-3. URL <GotoISI>://WOS:A1995RK39800008.
- 523 Robert B. Noland. Mobility and the effective reproduction rate of covid-19. *Journal of Transport & Health*, 20:101016, 2021.
 524 ISSN 2214-1405. doi: <https://doi.org/10.1016/j.jth.2021.101016>. URL <https://www.sciencedirect.com/science/article/pii/S2214140521000104>.
- 525 Abdul Noury, Abel François, Olivier Gergaud, and Alexandre Garel. How does covid-19 affect electoral participation? evidence
 526 from the french municipal elections. *PLOS ONE*, 16(2):e0247026, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0247026.
 527 URL <https://dx.doi.org/10.1371/journal.pone.0247026>.
- 528 A Paez. Using google community mobility reports to investigate the incidence of covid-19 in the united states. *Findings*, 2020.
 529 doi: <https://doi.org/10.32866/001c.12976>.
- 530 Antonio Paez. Using spatial filters and exploratory data analysis to enhance regression models of spatial data. *Geographical
 531 Analysis*, 51(3):314–338, 2019. ISSN 0016-7363. doi: 10.1111/gean.12180. URL <GotoISI>://WOS:000475464200003.
- 532 Antonio Paez, Fernando A. Lopez, Tatiane Menezes, Renata Cavalcanti, and Maira Galdino da Rocha Pitta. A spatio-temporal
 533 analysis of the environmental correlates of covid-19 incidence in spain. *Geographical Analysis*, n/a(n/a), 2020. ISSN 0016-7363.
 534 doi: 10.1111/gean.12241. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/gean.12241>.
- 535 Pedro Pequeno, Bruna Mendel, Clarissa Rosa, Mariane Bosholn, Jorge Luiz Souza, Fabricio Baccaro, Reinaldo Barbosa, and
 536 William Magnusson. Air transportation, population density and temperature predict the spread of covid-19 in brazil. *PeerJ*,
 537 8:e9322, 2020. ISSN 2167-8359. doi: 10.7717/peerj.9322. URL <https://dx.doi.org/10.7717/peerj.9322>.
- 538 R. O. Phillips, A. Fyhri, and F. Sagberg. Risk compensation and bicycle helmets. *Risk Analysis*, 31(8):1187–1195, 2011. ISSN
 539 0272-4332. doi: 10.1111/j.1539-6924.2011.01589.x. URL <GotoISI>://WOS:000293906500003.
- 540 Sarbeswar Praharaj, David King, Christopher Pettit, and Elizabeth Wentz. Using aggregated mobility data to measure the effect
 541 of covid-19 policies on mobility changes in sydney, london, phoenix, and pune. *Findings*, 10 2020. doi: 10.32866/001c.17590.
- 542 J. Richens, J. Imrie, and A. Copas. Condoms and seat belts: the parallels and the lessons. *Lancet*, 355(9201):400–403, 2000.
 543 ISSN 0140-6736. doi: 10.1016/s0140-6736(99)09109-6. URL <GotoISI>://WOS:000085122100039.
- 544 Joacim Rocklöv and Henrik Sjödin. High population densities catalyse the spread of covid-19. *Journal of Travel Medicine*, 27
 545 (3), 2020. ISSN 1195-1982. doi: 10.1093/jtm/taaa038. URL <https://dx.doi.org/10.1093/jtm/taaa038>.
- 546 Satyaki Roy and Preetam Ghosh. Factors affecting covid-19 infected and death rates inform lockdown-related policymaking.
 547 *PLOS ONE*, 15(10):e0241165, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0241165. URL <https://dx.doi.org/10.1371/journal.pone.0241165>.
- 548 Martin Schweinsberg, Michael Feldman, Nicola Staub, Olmo R. van den Akker, Robbie C.M. van Aert, Marcel A.L.M. van
 549 Assen, Yang Liu, Tim Althoff, Jeffrey Heer, Alex Kale, Zainab Mohamed, Hashem Amireh, Vaishali Venkatesh Prasad,
 550 Abraham Bernstein, Emily Robinson, Kaisa Snellman, S. Amy Sommer, Sarah M.G. Otner, David Robinson, Nikhil Madan,
 551 Raphael Silberzahn, Pavel Goldstein, Warren Tierney, Toshio Murase, Benjamin Mandl, Domenico Viganola, Carolin
 552 Strobl, Catherine B.C. Schaumans, Stijn Kelchtermans, Chan Naseeb, S. Mason Garrison, Tal Yarkoni, C.S. Richard Chan,
 553 Prestone Adie, Paulius Alaburda, Casper Albers, Sara Alspaugh, Jeff Alstott, Andrew A. Nelson, Eduardo Ariño de la
 554 Rubia, Adbi Arzi, Štěpán Bahník, Jason Baik, Laura Winther Balling, Sachin Bunker, David AA Baranger, Dale J. Barr,
 555 Brenda Barros-Rivera, Matt Bauer, Enuh Blaise, Lisa Boelen, Katerina Bohle Carbonell, Robert A. Briers, Oliver Burkhard,
 556 Miguel-Angel Canela, Laura Castrillo, Timothy Catlett, Olivia Chen, Michael Clark, Brent Cohn, Alex Coppock, Natàlia
 557 Cugueró-Escofet, Paul G. Curran, Wilson Cyrus-Lai, David Dai, Giulio Valentino Dalla Riva, Henrik Danielsson, Rosaria
 558 de F.S.M. Russo, Niko de Silva, Curdin Derungs, Frank Dondelinger, Carolina Duarte de Souza, B. Tyson Dube, Marina
 559 Dubova, Ben Mark Dunn, Peter Adriaan Edelsbrunner, Sara Finley, Nick Fox, Timo Gnamb, Yuanyuan Gong, Erin Grand,
 560 Brandon Greenawalt, Dan Han, Paul H.P. Hanel, Antony B. Hong, David Hood, Justin Hsueh, Lilian Huang, Kent N.
 561 Hui, Keith A. Hultman, Azka Javaid, Lily Ji Jiang, Jonathan Jong, Jash Kamdar, David Kane, Gregor Kappler, Erikson
 562 Kaszubowski, Christopher M. Kavanagh, Madian Khabsa, Bennett Kleinberg, Jens Kouros, Heather Krause, Angelos-Miltiadis
 563 Kryptotos, Dejan Lavbič, Rui Ling Lee, Timothy Leffel, Wei Yang Lim, Silvia Liverani, Bianca Loh, Dorte Lønsmann,
 564 Jia Wei Low, Alton Lu, Kyle MacDonald, Christopher R. Madan, Lasse Hjorth Madsen, Christina Maimone, Alexandra
 565 Mangold, Adrienne Marshall, Helena Ester Matskewich, Kimia Mavon, Katherine L. McLain, Amelia A. McNamara, Mhairi
 566 McNeill, Ulf Mertens, David Miller, Ben Moore, Andrew Moore, Eric Nantz, Ziauddin Nasrullah, Valentina Nejkovic,
 567 Colleen S Nell, Andrew Arthur Nelson, Gustav Nilsonne, Rory Nolan, Christopher E. O'Brien, Patrick O'Neill, Kieran
 568 O'Shea, Toto Olita, Jahna Otterbacher, Diana Palsetia, Bianca Pereira, Ivan Pozdniakov, John Protzko, Jean-Nicolas Reyt,
 569 Travis Riddle, Amal (Akmal) Ridhwan Omar Ali, Ivan Ropovik, Joshua M. Rosenberg, Stephane Rothen, Michael Schulte-
 570 Mecklenbeck, Nirek Sharma, Gordon Shotwell, Martin Skarzynski, William Stedden, Victoria Stodden, Martin A. Stoffel,
 571 Scott Stoltzman, Subashini Subbaiah, Rachael Tatman, Paul H. Thibodeau, Sabina Tomkins, Ana Valdivia, Gerrieke B.
 572 Druijff-van de Woestijne, Laura Viana, Florence Villesèche, W. Duncan Wadsworth, Florian Wanders, Krista Watts,
 573 Jason D Wells, Christopher E. Whelpley, Andy Won, Lawrence Wu, Arthur Yip, Casey Youngflesh, Ju-Chi Yu, Arash
 574 Zandian, Leilei Zhang, Chava Zibman, and Eric Luis Uhlmann. Same data, different conclusions: Radical dispersion in
 575 empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and
 576 Human Decision Processes*, 165:228–249, 2021. ISSN 0749-5978. doi: <https://doi.org/10.1016/j.obhd.2021.02.003>. URL
 577 <https://www.sciencedirect.com/science/article/pii/S0749597821000200>.
- 578 Ayyoob Sharifi and Amir Reza Khavarian-Garmsir. The covid-19 pandemic: Impacts on cities and major lessons for urban
 579 planning, design, and management. *Science of The Total Environment*, 749:142391, 2020. ISSN 0048-9697. doi: <https://doi.org/10.1016/j.scitotenv.2020.142391>. URL <https://www.sciencedirect.com/science/article/pii/S0048969720359209>.
- 580 Piotr Skórka, Beata Grzywacz, Dawid Moroń, and Magdalena Lenda. The macroecology of the covid-19 pandemic in the

- 584 anthropocene. *PLOS ONE*, 15(7):e0236856, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0236856. URL <https://dx.doi.org/10.1371/journal.pone.0236856>.
- 585
- 586 Marc Souris and Jean-Paul Gonzalez. Covid-19: Spatial analysis of hospital case-fatality rate in france. *PLOS ONE*, 15(12):
587 e0243606, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0243606. URL <https://dx.doi.org/10.1371/journal.pone.0243606>.
- 588
- 589 Kenton E. Stephens, Pavel Chernyavskiy, and Danielle R. Bruns. Impact of altitude on covid-19 infection and death in the
590 united states: A modeling and observational study. *PLOS ONE*, 16(1):e0245055, 2021. ISSN 1932-6203. doi: 10.1371/journal.
591 pone.0245055. URL <https://dx.doi.org/10.1371/journal.pone.0245055>.
- 592 V. Stodden, J. Seiler, and Z. K. Ma. An empirical analysis of journal policy effectiveness for computational reproducibility.
593 *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):2584–2589, 2018. ISSN 0027-8424.
594 doi: 10.1073/pnas.1708290115. URL <GotoISI>://WOS:000427245400036.
- 595 Josh Sumner, Leah Haynes, Sarah Nathan, Cynthia Hudson-Vitale, and Leslie D. McIntosh. Reproducibility and reporting
596 practices in covid-19 preprint manuscripts. *medRxiv*, page 2020.03.24.20042796, 2020. doi: 10.1101/2020.03.24.20042796.
597 URL <https://www.medrxiv.org/content/medrxiv/early/2020/03/27/2020.03.24.20042796.full.pdf>.
- 598 Zhibin Sun, Hui Zhang, Yifei Yang, Hua Wan, and Yixiang Wang. Impacts of geographic factors and population density on
599 the covid-19 spreading under the lockdown policies of china. *Science of The Total Environment*, 746:141347, 2020. ISSN
600 0048-9697. doi: 10.1016/j.scitotenv.2020.141347. URL <https://dx.doi.org/10.1016/j.scitotenv.2020.141347>.
- 601 Karla Therese L. Sy, Laura F. White, and Brooke E. Nichols. Population density and basic reproductive number of covid-19
602 across united states counties. *PLOS ONE*, 16(4):e0249271, 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0249271. URL
603 <https://dx.doi.org/10.1371/journal.pone.0249271>.
- 604 Giuliana Viglione. 'avalanche'of spider-paper retractions shakes behavioural-ecology community. *Nature*, 578(7794):199–201,
605 2020.
- 606 Caitlyn Vlasschaert, Joel M. Topf, and Swapnil Hiremath. Proliferation of papers and preprints during the coronavirus disease
607 2019 pandemic: Progress or problems with peer review? *Advances in Chronic Kidney Disease*, 27(5):418–426, 2020. ISSN
608 1548-5595. doi: 10.1053/j.ackd.2020.08.003. URL <https://dx.doi.org/10.1053/j.ackd.2020.08.003>.
- 609 Fangxiong Wang, Ziqian Tan, Zaihui Yu, Siqi Yao, and Changfeng Guo. Transmission and control pressure analysis of the
610 covid-19 epidemic situation using multisource spatio-temporal big data. *PLOS ONE*, 16(3):e0249145, 2021. ISSN 1932-6203.
611 doi: 10.1371/journal.pone.0249145. URL <https://dx.doi.org/10.1371/journal.pone.0249145>.
- 612 Easton R. White and Laurent Hébert-Dufresne. State-level variation of initial covid-19 dynamics in the united states. *PLOS
613 ONE*, 15(10):e0240648, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0240648. URL <https://dx.doi.org/10.1371/journal.pone.0240648>.
- 615 David W. S. Wong and Yun Li. Spreading of covid-19: Density matters. *PLOS ONE*, 15(12):e0242398, 2020. ISSN 1932-6203.
616 doi: 10.1371/journal.pone.0242398. URL <https://dx.doi.org/10.1371/journal.pone.0242398>.