

1 **Geographic Disparities and Predictors of COVID-19 Hospitalization Risk in the St.**
2 **Louis Area, Missouri (USA).**

3
4 Morganne Igoe¹, Praachi Das², Suzanne Lenhart¹, Alun L. Lloyd², Lan Luong³, Dajun
5 Tian³, Cristina Lanzas⁴, Agricola Odoi^{5,§}
6
7
8
9

10
11
12
13
14
15 ¹Department of Mathematics, University of Tennessee, Knoxville, 37996
16

17 ²Biomathematics Graduate Program and Department of Mathematics, North Carolina
18 State University, Raleigh NC 27695
19

20 ³BJC Healthcare, St. Louis, Missouri, 63110
21

22 ⁴Department of Population Health and Pathobiology and Comparative Medicine
23 Institute, North Carolina State University, Raleigh, 27607
24

25 ⁵Department of Biomedical and Diagnostics Sciences, University of Tennessee,
26 Knoxville, 37996
27
28
29
30
31
32
33
34
35
36
37
38

39 **§Corresponding Author:**

40 Agricola Odoi, aodoi@utk.edu
41

42 **Abstract**

43 **Background:** COVID-19 has overwhelmed the US healthcare system, with over 44
44 million cases and over 700,000 deaths as of October 6, 2021. There is evidence that
45 some communities are disproportionately affected. This may result in geographic
46 disparities in COVID-19 hospitalization risk that, if identified, could guide control efforts.
47 Therefore, the objective of this study is to investigate Zip Code Tabulation Area (ZCTA)-
48 level geographic disparities and identify predictors of COVID-19 hospitalization risk in
49 the St. Louis area.

50 **Methods:** Hospitalization data for COVID-19 and several chronic diseases were
51 obtained from the Missouri Hospital Association. ZCTA-level data on socioeconomic
52 and demographic factors were obtained from the US Census Bureau American
53 Community Survey. Age-adjusted COVID-19 and several chronic disease
54 hospitalization risks were calculated. Geographic disparities in distribution of COVID-19
55 age-adjusted hospitalization risk, socioeconomic and demographic factors as well as
56 chronic disease risks were investigated using choropleth maps. Predictors of ZCTA-
57 level COVID-19 hospitalization risks were investigated using global negative binomial
58 and local geographically weighted negative binomial models.

59 **Results:** There were geographic disparities of COVID-19 hospitalization risks. COVID-
60 19 hospitalization risks were significantly higher in ZCTAs with high diabetes
61 hospitalization risks ($p < 0.0001$), high risks of COVID-19 cases ($p < 0.0001$), as well as
62 high percentages of black population ($p = 0.0416$) and populations with some college
63 education ($p = 0.0005$). The coefficients of the first three predictors varied across ZCTAs,
64 implying that the associations between COVID-19 hospitalization risks and these

65 predictors varied by geographic location. This implies that a “one-size-fits-all” approach
66 may not be appropriate for management and control.

67 **Conclusions:** There is evidence of geographic disparities in COVID-19 hospitalization
68 risks that are driven by differences in socioeconomic, demographic and health-related
69 factors. The impacts of these factors vary by geographical location with some factors
70 being more important predictors in some locales than others. Use of both global and
71 local models leads to a better understanding of the determinants of geographic
72 disparities in health outcomes and utilization of health services. These findings are
73 useful for informing health planning to identify geographic areas likely to have high
74 numbers of individuals needing hospitalization as well as guiding vaccination efforts.

75

76

77

78

79

80

81 **Key Words:** Severe Acute Respiratory Syndrome Coronavirus 2; SARS-CoV-2; Coronavirus
82 Disease 2019; COVID-19; Disparities; Hospitalization Risks; Predictors; Negative Binomial
83 Models; Geographically Weighted Regression Models; Epidemiology; Missouri.

84 **Background**

85 There have been over 184 million confirmed Coronavirus Disease 2019 (COVID-19)
86 cases and over 4 million deaths worldwide as of July 8, 2021, with over 33 million
87 confirmed cases and over 600,000 deaths in the United States [1]. As of the same date,
88 the state of Missouri has reported over 633,000 cases and 9,700+ deaths. The disease
89 has overwhelmed many United States hospital systems, with large numbers of patients
90 requiring critical care and interventions such as mechanical ventilation [2–4]. This surge
91 of COVID-19 patients has put strain on hospital resources, potentially impacting the
92 care available to both COVID-19 and non-COVID-19 patients [5].

93

94 There is evidence of geographic disparities in the severity of the disease with certain
95 population groups experiencing more severe disease than others. These disparities
96 might be driven by population characteristics such as socioeconomic, demographic, and
97 chronic disease factors [6–10]. For instance, there is evidence that Non-Hispanic
98 American Indian, non-Hispanic Black, and Hispanic people have higher hospitalization
99 risks than their non-Hispanic Asian and non-Hispanic White counterparts [11]. There are
100 also reports that conditions such as diabetes mellitus, obesity, chronic lung conditions,
101 renal disease, cancer, and cardiovascular disease may increase the severity of the
102 condition and risk of hospitalizations among COVID-19 patients with these co-
103 morbidities [6, 12–14]. Identifying geographic disparities in COVID-19 hospitalization
104 risks and determinants of these disparities is important in providing information to guide
105 hospital preparedness to handle the patient surge and for targeting resources for public
106 health efforts to control the condition at the community level.

107
108 Improved understanding of the geographic disparities and predictors of COVID-19
109 hospitalization risk at the local level, such as the Zip Code Tabulation Areas (ZCTA),
110 would help identify local geographic areas with higher needs for hospital beds and other
111 healthcare resources. This is useful information for healthcare planning and service
112 provision at the local level. It may also inform vaccination efforts by helping to identify
113 areas where higher vaccination coverage may have the largest effect in reducing
114 hospital burden. Therefore, the objective of this study is to identify geographic
115 disparities and predictors of COVID-19 hospitalization risk in the St. Louis Area,
116 Missouri, United States.

117 **Methods**

118 **Study Area**

119 This study was performed in an area of Missouri that includes 108 ZCTAs located in St.
120 Charles county, St. Louis county, and St. Louis City as well as parts of Jefferson,
121 Franklin and Warren counties (**Figure 1**). The area had a population of approximately 2
122 million people comprising 74% white, 20% black, 3% Hispanic/Latino, and 3% Asian.
123 Forty-eight percent of the population was male, while 52% was female. Thirty-one
124 percent of the population was older than 25 years old, 53% was 25-64, and 16% was
125 older than 65 years of age. The ZCTA-level population density varied from 10 people
126 per square mile in St. Charles County to 9,368 people per square mile in St. Louis City
127 County.

128

129 **Data Sources**

130 *COVID-19 and Chronic Disease Hospitalization Data*

131 Data on COVID-19 and chronic condition hospitalizations were obtained from the
132 Hospital Industry Data Institute, a not-for-profit organization founded by the Missouri
133 Hospital Association that summarizes state level discharge data. The COVID-19 data
134 provided information on the numbers of COVID-19 discharges by patient ZCTA and by
135 age group from April 1, 2020 to September 30, 2020. In addition to the COVID-19
136 hospitalization data, data on the number of confirmed COVID-19 cases were obtained
137 from the county Departments of Health. COVID-19 case data were included as one of
138 the potential predictors of COVID-19 hospitalization risks. Data on chronic conditions
139 were extracted for the time period 2019-2020 based on ICD-10 codes and included the

140 following conditions/behaviors: obesity, tobacco use, cancer (breast, colorectal,
141 prostate, lung, endometrial, leukemia and lymphoma), chronic obstructive pulmonary
142 disease, chronic kidney disease, heart failure, and diabetes. These conditions were
143 selected due to their potential association with COVID-19 severity. All data were
144 aggregated to the ZCTA-level to facilitate subsequent ZCTA-level analyses. The ZCTA-
145 level risks of these conditions were then computed and presented as disease-specific
146 hospitalizations per 100 population.

147

148 *Socioeconomic, demographic and base map data*

149

150 ZCTA-level data on socioeconomic and demographic factors including age, sex, race,
151 population density, education level, and median income, were obtained from the U.S.
152 Census Bureau 2018 American Community Survey (ACS) 5-year estimates [15]. The
153 cartographic boundary files, used for generating maps, were obtained from the US
154 Census Bureau Website [16].

155

156 **Descriptive Analyses**

157

158 All descriptive analyses were performed in R version 4.1.0 [17] using RStudio version
159 1.4.1103 [18]. Normality of distribution of continuous variables was assessed using
160 Shapiro-Wilk test. Medians as well as 1st and 3rd quartiles were used as measures of
161 central tendency and spread for all continuous variables since they were all non-
162 normally distributed. The COVID-19 ZCTA-level hospitalization risks were directly age-
163 standardized, in STATA version 16 [19], using the 2010 US population as the standard
164 population.

165

166 **Investigation of Predictors of COVID-19 Hospitalizations**

167 *Global Models*

168 Univariable global Poisson models were used to assess simple associations between
169 each of the potential predictors and COVID-19 hospitalization risk. As opposed to local
170 models that estimate as many regression coefficients per predictor as the number of
171 geographic units, global models assume constant relationships between each potential
172 predictor and the outcome and therefore estimate one regression coefficient per
173 potential predictor. These models were fit under the Generalized Linear Model (GLM)
174 framework in R [17] specifying log link. The dependent variable was the expected count
175 of COVID-19 hospitalizations in each ZCTA based on the age-adjusted/standardized
176 number of COVID-19 hospitalizations and the offset was the natural log of the
177 population in each ZCTA. A relaxed α of 0.15 was used to assess potentially significant
178 predictors in univariable models. The linearity of the relationships between the log risk
179 and potential predictor variables was assessed graphically.

180

181 Spearman rank correlation coefficients were computed to assess bivariate correlations
182 among the potential predictor variables identified during the univariable analyses using
183 R [17]. Variables with correlation coefficients greater than 0.7 were considered highly
184 correlated. To avoid multicollinearity, only one of a pair of highly correlated variables
185 was retained for further assessments using multivariable Poisson model. Biological and
186 statistical considerations were used to determine which of a pair of highly correlated
187 variables to retain. Non-correlated variables that had univariable $p \leq 0.15$ were assessed

188 in a multivariable global Poisson model, built using manual backwards elimination and
189 an α of 0.05. Variables were considered as confounders if their removal from the model
190 resulted in changes of at least 20% of the coefficients of any of the variables in the
191 model. Confounders were retained in the final main effects model. Biologically plausible
192 two-way interactions between variables in the final main effects multivariable model
193 were assessed and the significant ones retained in the final model. The final Poisson
194 regression model was then assessed for overdispersion.

195
196 Due to the presence of overdispersion in the final Poisson model (based on comparison
197 of model deviance and degrees of freedom), a global negative binomial model was built.
198 The process of building the negative binomial (NB) model was similar to that of the
199 Poisson model described above. However, the `glm.nb` function of the MASS package
200 [20] of R [17] was used instead of the `glm` function. The rest of the model specifications
201 were similar to the Poisson model. Goodness-of-fit of the final global NB model was
202 assessed using Pearson and Deviance chi-square goodness-of-fit tests.

203 204 *Local Models*

205 To assess if the association between each of the predictors and hospitalization risks
206 varied by geographic location, a geographically weighted negative binomial (GWNB)
207 model, proposed by Silva and Rodrigues [21], was fit to the data specifying the same
208 outcome, link function, offset and significant predictors as in the final global
209 multivariable NB model. However, unlike the global multivariable NB model that
210 estimates one regression coefficient for each predictor and thus assumes a constant

211 strength of association across all ZCTAs, the GWNB theoretically estimates as many
212 regression coefficients as the number of ZCTAs. Essentially, GWNB evaluates a local
213 model of the outcome by fitting a regression equation to every ZCTA in the dataset.
214 These separate model equations are constructed by incorporating the outcome and
215 predictor variables of the ZCTAs that fall within the neighborhood of each target ZCTA.
216 Therefore, the GWNB allows identification of local variations in the strength of
217 associations and therefore giving the importance of specific predictors in different local
218 areas (ZCTAs). This implies that some factors may be more important predictors of
219 hospitalization risk in some ZCTAs than others.

220

221 The local GWNB model was fit in SAS version 9.4 [22] using a SAS/IML macro [23].
222 The estimation of local regression coefficients was based on biquadratic kernel
223 weighting function [23], while the bandwidth was estimated using the adaptive method
224 which allows the size of the bandwidth to vary based on the density of observations.
225 Bias-corrected Akaike Information Criteria (AICc) was used to determine the optimum
226 kernel bandwidth and for comparing the goodness-of-fit of the global NB and GWNB
227 models. The better fitting model was identified as the one with the lower AICc value.

228

229 Stationarity of the GWNB coefficients was assessed using: (a) randomization non-
230 stationarity test based on 999 replications [24]; (b) comparison of the interquartile range
231 of the local GWNB model coefficients with the standard error estimates of the global NB
232 model. Local coefficients whose interquartile ranges were larger than twice the standard

233 error of the regression coefficient from the global NB model were considered non-
234 stationary [25, 26].

235

236 **Cartographic Displays**

237 Choropleth maps showing the geographic distributions of ZCTA-level age-adjusted
238 COVID-19 hospitalization risks, the socioeconomic, demographic and chronic disease
239 factors as well as local regression coefficients from the GWNB models were generated
240 using QGIS 3.16.6 [27]. Jenk's optimization classification scheme [28–30] was used to
241 determine the critical intervals of the choropleth maps.

242 **Results**

243 **Descriptive Statistics**

244 The ZCTA-level median percentage of males was 48.6% while that of black and
245 Hispanic populations were 3.7% and 2.2%, respectively (**Table 1**). For education
246 variables, 38.2% of the population had high school education, 23% had some college
247 education while 8.6% had associate's degree and 18% had bachelor's degree. The
248 median household income was just over \$59,400 with 9.5% of the ZCTA-level
249 population living below the poverty line. Among chronic conditions investigated, Chronic
250 Kidney Disease had the highest hospitalization risk (7.6%) followed by diabetes (7.5%)
251 and heart failure had the lowest (3.2%) (**Table 1**). The ZCTA-level median number of
252 confirmed cases of COVID-19 was 360 cases which was equivalent to 2.2% of the
253 population at ZCTA-level (**Table 1**).

254

255 **Predictors of COVID-19 Hospitalization Risks**

256 *Global Model*

257 A number of variables had univariable associations with COVID-19 hospitalization risks
258 at a relaxed $p < 0.2$ (**Table 2**). Of the assessed demographic variables, only percentage
259 of black population had a univariable association ($p < 0.001$) with hospitalization risk. By
260 contrast, all the assessed educational, economic, health behavior and chronic disease
261 variables had significant univariable associations with COVID-19 hospitalization risks
262 (**Table 2**). The ZCTA-level risk of confirmed COVID-19 cases had a significant
263 ($p < 0.001$) univariable association with COVID-19 hospitalization risk but the raw count
264 of COVID-19 cases per ZCTA did not ($p = 0.9$) (**Table 2**).

265

266 Based on the final global multivariable NB model, the ZCTA-level hospitalization risks

267 were higher in ZCTAs that were high in the following predictors: percentage of black

268 population ($p=0.0416$), percentage of population with some college education

269 ($p=0.0005$), percentage of individuals hospitalized with diabetes ($p<0.0001$), and

270 number of ZCTA-level COVID-19 cases per 100 population ($p=0.0001$) (**Table 3**). A

271 map of the distribution of age-adjusted COVID-19 hospitalization risks and each of the

272 significant predictors is shown in **Figure 2**. High age-adjusted COVID-19 hospitalization

273 risks tended to occur in the Northeast of the study area and included ZCTAs in parts of

274 St. Charles, St. Louis and Louis City counties. These areas also had high percentages

275 of black population, individuals with some college education, those with high diabetes

276 hospitalization risks as well as high risks of COVID-19 cases (**Figure 2**). High

277 hospitalization risks were also evident in the Southwest areas of the study area that

278 included parts of Warren and Franklin counties. It is worth noting that these areas also

279 tended to have high diabetes hospitalization risks and percentages of the population

280 with some college education (**Figure 2**). Since the global model did not show evidence

281 of good fit based on both Deviance ($p=0.03$) and Pearson ($p=0.01$) goodness-of-fit

282 tests, stationarity of the regression coefficients was assessed using a local GWNB

283 model.

284 *Local Model*

285 The p-values of the stationarity tests from the GWNB model indicate that the
286 coefficients for the association between COVID-19 hospitalization risks and percentage
287 of black population ($p=0.001$) and number of hospitalized patients with diabetes per 100
288 population ($p=0.032$) were non-stationary (**Table 4**). Additionally, comparison of the
289 $2\times$ SEs of the global coefficients and IQR of the local coefficients showed evidence of
290 non-stationarity of the coefficients of the above two predictors as well as the population
291 adjusted cases of COVID-19 (**Table 4**).

292
293 The spatial distribution of the local coefficients of the three predictors whose relationship
294 with COVID-19 hospitalizations were identified as non-stationary provides visual
295 evidence for variability of the local relationships (**Figure 3**). Thus, associations of
296 COVID-19 hospitalization risks with percentage of black population, diabetes
297 hospitalization risks and COVID-19 adjusted cases varied considerably across the study
298 area, with a strong East-West gradient. The association between percentage of black
299 population and COVID-19 hospitalization risks, for instance, was positive in the
300 Northeast and negative in the West and Southwest. Moreover, the strength of the
301 association was higher in ZCTAs in the West compared to those near the center of the
302 study area. The strength of association between COVID-19 hospitalization risks with
303 diabetes hospitalization risks was also higher in the Northeast and lower in the West. All
304 ZCTAs, except one (63025), showed evidence of positive association between COVID-
305 19 hospitalization risks and diabetes hospitalization risks. Finally, the association
306 between COVID-19 hospitalization and population adjusted cases of COVID-19

307 increased from West to East, with the association staying positive in all ZCTAs of the
308 study area. It is worth noting that the local GWNB model had a much better goodness-
309 of-fit (AICc=986.4) than the global model (AICc=1002.7).

310 **Discussion**

311 The goal of this study was to investigate geographic disparities and identify predictors of
312 ZCTA-level COVID-19 hospitalization risks in the St. Louis area. The findings of this
313 study can be used to identify areas where the population is at higher risk of
314 hospitalization due to COVID-19 in order to guide planning and control efforts and to
315 reduce potential overburdening of hospitals during COVID-19 surges.

316

317 There was evidence of geographic disparities in COVID-19 age-adjusted hospitalization
318 risks in the study area. Urban ZCTAs in St. Louis City and St. Louis County exhibited
319 high hospitalization risks. These ZCTAs also have high percentages of the population
320 that are black, some as high as 98.4%. Some of these ZCTA also had high diabetes
321 hospitalization risks. It is worth noting that some rural ZCTAs in Franklin and Warren
322 counties had high COVID-19 hospitalization risk but very low percentages of the
323 population that were black. However, these ZCTAs had high diabetes hospitalization
324 risks implying that the COVID hospitalization risks in these rural ZCTAs was more
325 driven by diabetes burden than demographic factors. Thus, although COVID-19
326 hospitalization risks in the more urban areas seems to be driven by the demographic
327 composition of the population, the risks in the more rural areas seem to be driven more
328 by diabetes burden. Geographic areas with intermediate to high COVID-19
329 hospitalization risks tended to have high percentages of the population with some
330 college education and included ZCTAs in Franklin, Jefferson, St. Louis City and St.
331 Louis County. It is worth noting that St. Louis City and St. Louis County tend to have
332 similar restrictions, but often time, the restrictions from the other counties are more

333 laxed. Previous ecological and individual level studies have not considered education
334 level as a predictor of hospitalization risk. The level of education may be a proxy of
335 occupation and other sociodemographic factors that may impact both the risk of
336 infection and the resulting severity of the disease and hence hospitalization risks.

337
338 The above findings are consistent with those from previous ecological studies that
339 investigated risk factors and predictors of COVID-19 hospitalization. For instance, an
340 ecological study by Nguyen et al. reported that diabetes was a significant predictor of
341 increased COVID-19 hospitalization risk at the county level in Georgia, USA even after
342 controlling for sociodemographic and economic factors [31]. On the contrary, an Iranian
343 study conducted at the provincial level did not find a significant association when only
344 controlling for chronic disease factors [32]. This may indicate differences of relationships
345 in different geographic areas and populations as well as the importance of controlling for
346 sociodemographic factors when evaluating the impact of chronic disease variables on
347 COVID-19 hospitalization risk. Considering the fact that ZCTAs with high diabetes
348 hospitalization risks also tended to have high COVID-19 hospitalization risks, COVID-19
349 mitigation efforts may need to be targeted to these ZCTAs to reduce the potential
350 burden of the disease.

351
352 Interestingly, the study by Nguyen et al., which also considered sociodemographic and
353 economic factors as well as comorbidities, did not find a significant association between
354 percentage of black population with COVID-19 hospitalization risk [31]. However, it did
355 identify other socioeconomic factors that this study did not consider, including:

356 percentage of children in poverty and percentage of the population with severe housing
357 problems. Although not directly comparable with the current ecological study, previous
358 individual level studies have also identified associations between black race and
359 COVID-19 hospitalization risk [8, 13, 33].

360
361 The local GWNB model allowed modeling of geographically varying associations
362 between COVID-19 hospitalization risks and its predictors instead of assuming constant
363 associations across the study area. Although the coefficients of percentage of black
364 population, diabetes hospitalization risk and risk of COVID-19 infections varied spatially,
365 the coefficients of percentage of the population with some college education did not and
366 hence was modeled as stationary. This suggests that the coefficients for the percentage
367 of the population with some college education are generalizable to all ZCTAs in the
368 study area. In contrast, geographic variations in the associations between COVID-19
369 hospitalization risks and the percentage of black population, diabetes hospitalization
370 risks and risks of COVID-19 infections suggest that global coefficients do not
371 adequately describe the associations between COVID-19 hospitalization risks and these
372 predictors across the study area. These findings have health planning and service
373 provision implications. For instance, a “one size fits all” approach would not be suitable
374 for addressing geographic disparities in COVID-19 hospitalization risks across the study
375 area since some predictors are more important in some locales than others. Thus,
376 different locales may require slightly different strategies depending on the most
377 important predictors driving COVID-19 hospitalization risk in the location. Therefore,
378 planning for hospital capacity and other disease management and control efforts will

379 need to use evidence-based approaches informed by empirical evidence from both
380 global and local models.

381

382 *Strengths and limitations*

383 This study used both global and local models to investigate geographic disparities and
384 identify predictors of COVID-19 hospitalization risks in St. Louis region of Missouri. The
385 use of local models to investigate stationarity of regression coefficients of significant
386 predictors and model non-stationary coefficients is a key strength of the study. This
387 approach is particularly important in guiding local health planning since the importance
388 of different predictors are not constant across the study area implying that different
389 management and control strategies may need to be used in different areas. Therefore,
390 modeling approaches that use both global and local models help to better understand
391 the relationships between the outcome and predictors and may be more useful in
392 guiding control efforts at the local level. However, the study is not without limitations.
393 The hospital data has limitations associated with diagnostic classifications of COVID-19
394 in situations when the patient has co-morbidities that may have contributed to
395 hospitalization. Additionally, there may be geographic differences in COVID-19 case
396 ascertainment and reporting.

397

398 *Conclusions*

399 There is evidence of geographic disparities in COVID-19 hospitalization risks in the St.
400 Louis area of Missouri. These disparities are driven by socioeconomic, demographic
401 and health-related factors. The impacts of these factors vary by geographical location

402 with some factors being more important predictors of COVID-19 hospitalization risk in
403 some locales than others. This demonstrates the importance of using not only global but
404 also local models to investigate determinants of geographic disparities in health
405 outcomes and utilization of health services. This study's findings are useful for informing
406 healthcare system planning to identify geographic areas likely to have high numbers of
407 individuals needing hospitalization as well as in guiding vaccination efforts.

408

409 **Declarations**

410 **Ethical approval and consent to participate**

411 The study was approved by the North Carolina State University Institutional Review
412 Board (IRB number: 22342) and all study methods were carried out in accordance with
413 relevant guidelines and regulations. The study used anonymized secondary data
414 provided to the investigators, by the Hospital Industry Data Institute, in such a manner
415 that the identify of human subjects cannot be ascertained directly or through identifiers
416 linked to the subjects. The investigators did not contact the subjects and did not re-
417 identify subjects.

418

419 **Consent for publication**

420 Not applicable.

421

422 **Availability of data and materials**

423 The investigators cannot share the datasets, which are not publicly available, because
424 they do not have legal ownership nor authority to share the data. However, the data are
425 available upon request from the Hospital Industry Data Institute at 4712 Country Club
426 Drive P.O. Box 60 Jefferson City, MO 65102-0060, telephone (573)893-3700.

427

428 **Competing interests**

429 The authors declare that they have no competing interests.

430

431

432 **Funding**

433 This work was supported by CDC U01CK000587-01M001. The funders had no role in
434 study design, data collection and analysis, decision to publish, or preparation of the
435 manuscript.

436

437 **Authors' contributions**

438 SL, ALL, CL, and AO conceptualized research idea

439 LL, DT collected and curated the data

440 MI, PD, SL, ALL, CL and AO analyzed data

441 MI, PD and AO wrote the manuscript

442 All authors edited the manuscript

443 All authors read and approved the final manuscript.

444

445 **Acknowledgements**

446 We thank the St. Louis Comparative Modeling Network for facilitating access to data

447 and useful discussions.

448

449 **References**

- 450
- 451 1. World Health Organization (WHO). WHO Coronavirus (COVID-19) Dashboard.
452 Accessed on July 8, 2021. <https://covid19.who.int/>. WHO. 2021.
- 453 2. Keene AB, Shiloh AL, Eisen L, Berger J, Karwa M, Fein D, et al. Critical Care Surge
454 During the COVID-19 Pandemic: Implementation and Feedback From Frontline
455 Providers.
- 456 3. Hassan EM, Mahmoud HN. Impact of multiple waves of COVID-19 on healthcare
457 networks in the United States. *PLoS One*. 2021;16 3 March.
- 458 4. Khera R, Jain S, Lin Z, Ross JS, Krumholz HM. Evaluation of the anticipated burden
459 of COVID-19 on hospital-based healthcare services across the United States. *medRxiv*.
460 2020.
- 461 5. Ghanchi H, Takayanagi A, Savla P, Hariri OR, Tayag EC, Schiraldi M, et al. Effects of
462 the COVID-19 Pandemic on Stroke Patients. *Cureus*. 2020.
- 463 6. Garg S, Kim L, Whitaker M, Cummings C, Holstein R, Prill M, et al. MMWR -
464 Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-
465 Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020.
466 2019.
- 467 7. CDC COVID-19 Response Team. Severe Outcomes Among Patients with
468 Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020.
469 *Morb Mortal Wkly Rep*. 2020;69:343–6. <https://www.cdc.gov>.
- 470 8. Gao Y dong, Ding M, Dong X, Zhang J jin, Kursat Azkur A, Azkur D, et al. Risk
471 factors for severe and critically ill COVID-19 patients: A review. *Allergy: European*
472 *Journal of Allergy and Clinical Immunology*. 2020;76:428–55.
- 473 9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients
474 infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497–506.
- 475 10. Fadl N, Ali E, Salem TZ. COVID-19: Risk Factors Associated with Infectivity and
476 Severity. *Scand J Immunol*. 2021;;e13039.
- 477 11. Prevention C for DC and. COVID-19 Racial and Ethnic Health Disparities | CDC.
478 Centers for Disease Control and Prevention. 2021.
479 [https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-](https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/disparities-hospitalization.html#print)
480 [disparities/disparities-hospitalization.html#print](https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/racial-ethnic-disparities/disparities-hospitalization.html#print). Accessed 30 Jun 2021.
- 481 12. Zhou Y, Chi J, Lv W, Wang Y. Obesity and diabetes as high-risk factors for severe
482 coronavirus disease 2019 (Covid-19). *Diabetes/Metabolism Research and Reviews*.
483 2020;37.
- 484 13. Harlem G. Descriptive analysis of social determinant factors in urban communities
485 affected by COVID-19. *J Public Heal (United Kingdom)*. 2020;42:466–9.
- 486 14. Robilotti E V., Babady NE, Mead PA, Rolling T, Perez-Johnston R, Bernardes M, et
487 al. Determinants of COVID-19 disease severity in patients with cancer. *Nat Med*.

- 488 2020;26:1218–23.
- 489 15. US Census Bureau. ACS Demographic and Housing Estimates, 2018: 5-Year
490 Estimates. 2018. <https://www.census.gov/programs-surveys/acs>. Accessed 30 Jan
491 2020.
- 492 16. United States Census Bureau. TIGER/Line Shapefiles. 2019.
493 [https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-](https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.html)
494 [file.html](https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.html). Accessed 1 Dec 2019.
- 495 17. R Core Team. R: A language and environment for statistical computing. R
496 Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
497 2021. <https://www.r-project.org/>.
- 498 18. RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA
499 URL <http://www.rstudio.com/>. 2021.
- 500 19. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp
501 LLC. 2019.
- 502 20. W. N. Venables; B. D. Ripley. Modern Applied Statistics with S. Fourth. New York:
503 Springer; 2002. <https://www.stats.ox.ac.uk/pub/MASS4/>.
- 504 21. da Silva AR, Rodrigues TCV. Geographically Weighted Negative Binomial
505 Regression-incorporating overdispersion. *Stat Comput.* 2014;24:769–83.
- 506 22. SAS Institute. Statistical Analysis System (SAS) Version 9.4. Cary, NC, USA. SAS
507 Institute Inc. 2016.
- 508 23. Ricardo Da Silva A. A SAS ® Macro for Geographically Weighted Negative Binomial
509 Regression.
- 510 24. Hope ACA. A Simplified Monte Carlo Significance Test Procedure on JSTOR. *J R*
511 *Stat Soc Ser B Stat Methodol.* 1968;30:582–598.
512 <https://www.jstor.org/stable/2984263?seq=1>. Accessed 27 May 2021.
- 513 25. Brunson C, Fotheringham AS, Charlton ME. Geographically weighted regression: a
514 method for exploring spatial nonstationarity. *Geogr Anal.* 1996;28:281–98.
515 doi:10.1111/j.1538-4632.1996.tb00936.x.
- 516 26. Fotheringham AS, Brunson C, Charlton M. Geographically Weighted Regression:
517 The Analysis of Spatially Varying Relationships. West Sussex, England: John Wiley &
518 Sons Ltd; 2002.
- 519 27. QGIS.org. QGIS Geographic Information System. 2021. <http://www.qgis.org>.
- 520 28. AD Cliff; P Haggett. Atlas of Disease Distributions: Analytic Approaches to
521 Epidemiological Data. Oxford, United Kingdom: Basil Blackwell Ltd; 1988.
- 522 29. Jenks GF. The Data Model Concept in Statistical Mapping. *Int Yearb Cartogr.*
523 1967;7:186–90.
- 524 30. McMaster R. In Memoriam: George F. Jenks (1916–1996). *Cartogr Geogr Inf Sci.*

525 1997;24:56–9.

526 31. Nguyen TH, Shah GH, Schwind JS, Richmond HL. Community Characteristics and
527 COVID-19 Outcomes: A Study of 159 Counties in Georgia, United States. J Public Heal
528 Manag Pract. 2021;27:251–7.

529 32. Khorrami Z, Nili S, Sharifi H, Eybpoosh S, Shokoohi M. Association of cigarette
530 smoking, obesity, and underlying medical conditions with COVID-19 hospitalization and
531 mortality in Iran: A nationwide retrospective ecological study. Med J Islam Repub Iran.
532 2020;34:133.

533 33. Ko JY, Danielson ML, Town M, Derado G, Greenlund KJ, Daily Kirley P, et al. Risk
534 Factors for COVID-19-associated hospitalization: COVID-19-Associated Hospitalization
535 Surveillance Network and Behavioral Risk Factor Surveillance System.

536

537

538 **Tables**

539

540 **Table 1: Descriptive Statistics of ZCTA-level Potential Predictors of COVID-19**

541 **Hospitalization Risks in the St. Louis Area, Missouri**

Type of Variable	Variable	Median	First Quartile	Third Quartile
Demographic Factors				
	% male population	48.6	47.4	50.3
	% black population	3.7	0.9	34.2
	% Hispanic/Latino population	2.2	1.0	3.2
Educational Variables				
	% with ≤ high school education	38.2	23.9	47.6
	% with some college	23.0	19.8	25.6
	% with associate's degree	8.6	6.6	10.6
	% with bachelor's degree	18.0	9.4	25.9
Economic Variables				
	% below poverty level	9.5	5.8	16.8
	median household income	59,768.5	46005.3	77504.0
Health Behavior (ZCTA-level number of hospitalized patients that use tobacco per 100)				
	% tobacco ¹	10.4	6.8	14.5
Co-morbidities (ZCTA-level number of hospitalized patients with specific condition per 100)				
	% obesity	7.0	5.6	8.0
	% diabetes	7.5	6.5	9.2
	% cancer	3.8	3.3	4.3
	% COPD ²	4.0	3.1	4.9
	% CKD ³	7.6	6.4	8.9
	% heart failure	3.2	2.6	3.8
COVID-19 Cases				
	total cases	360	118	607
	cases per 100 population	2.2	1.9	2.5

542 ¹ZCTA-level % of COVID-19 hospitalized patients that were tobacco users

543 ²Chronic Obstructive Pulmonary Disease

544 ³Chronic Kidney Disease

545

546 **Table 2: Univariable Associations of Sociodemographic, Economic, and Chronic**
 547 **Disease Potential Predictors of COVID-19 Hospitalization Risk in the St. Louis**
 548 **Area, Missouri**

Type of Variable	Variable	Coefficient	95% Confidence Interval	P-values
Demographic Factors				
	% male population	0.004	-0.015, 0.024	0.727
	% black population	0.007	0.005, 0.009	<0.0001
	% Hispanic/Latino population	-0.021	-0.056, 0.016	0.292
Educational Variables				
	% with ≤ high school education	0.015	0.010, 0.020	<0.0001
	% with some college education	0.035	0.022, 0.049	<0.0001
	% with associate's degree	0.028	0.0002, 0.055	0.036
	% with bachelor's degree	-0.021	-0.028, -0.014	<0.0001
Economic Variables				
	% below poverty level	0.017	0.009, 0.024	<0.0001
	median income	-5.00E-06	-0.000008, -0.000003	<0.0001
Health Behavior (ZCTA-level number of hospitalized patients that use tobacco per 100 Population)				
	% tobacco ¹	0.046	0.034, 0.058	<0.0001
Co-morbidities (ZCTA-level number of hospitalized patients with specific condition per 100 Population)				
	% obesity	0.138	0.104, 0.171	<0.0001
	% cancer	-0.029	-0.096, 0.032	0.377
	% COPD ²	0.11	0.055, 0.164	<0.0001
	% CKD ³	0.1	0.068, 0.131	<0.0001
	% heart failure	0.194	0.123, 0.265	<0.0001
	% diabetes	0.11	0.083, 0.136	<0.0001
Confirmed COVID-19 Cases				
	total cases	1.40E-05	-0.0002, 0.00023	0.9
	cases per 100 population	0.277	0.192, 0.3621	<0.0001

549 ¹ZCTA-level % of COVID-19 hospitalized patients that were tobacco users

550 ²Chronic Obstructive Pulmonary Disease

551 ³Chronic Kidney Disease

552

553 **Table 3: Final Global Negative Binomial Model Showing Significant Determinants**
554 **of COVID-19 Hospitalization Risk in the St. Louis area**

Name	Coefficient	95 % Confidence Interval	p-value
% black population	0.0014	0.0001, 0.0027	0.0416
% with some college education	0.0180	0.0078, 0.0281	0.0005
% diabetes ¹	0.0628	0.0397, 0.0860	<0.0001
Confirmed COVID-19 cases per 100 population	0.2623	0.2027, 0.3218	<0.0001

555 ¹Number of hospitalized patients with diabetes per 100 Population

556

557 **Table 4: Results of assessment of stationarity of the coefficients of the predictors of the COVID-19 hospitalization**
 558 **risks in the St. Louis Area, Missouri**

	Global SE¹	Global SE¹x2	IQR² of Local Coefficient	IQR-2(SE)	Stationarity test p-value	Is Coefficient Stationary?
% Black Population	0.0007	0.0014	0.0148676	0.0134676	0.001	No ⁴
% With Some College Education	0.0052	0.0104	0.0092401	-0.0011599	0.406	Yes
% diabetes ³	0.0118	0.0236	0.0416903	0.0180903	0.032	No ⁴
Cases of COVID-19 per 100 population	0.0304	0.0608	0.1164299	0.0556299	0.109	No ⁵

559 ¹Standard Error

560 ²Interquartile Range

561 ³Number of hospitalized patients with diabetes per 100 Population

562 ⁴Coefficients are non-stationary based on both the p-value of the stationarity test and IQR-2(SE) assessment

563 ⁵Coefficient is non-stationary based on IQR-2(SE) assessment

564 **Figure Titles**

565

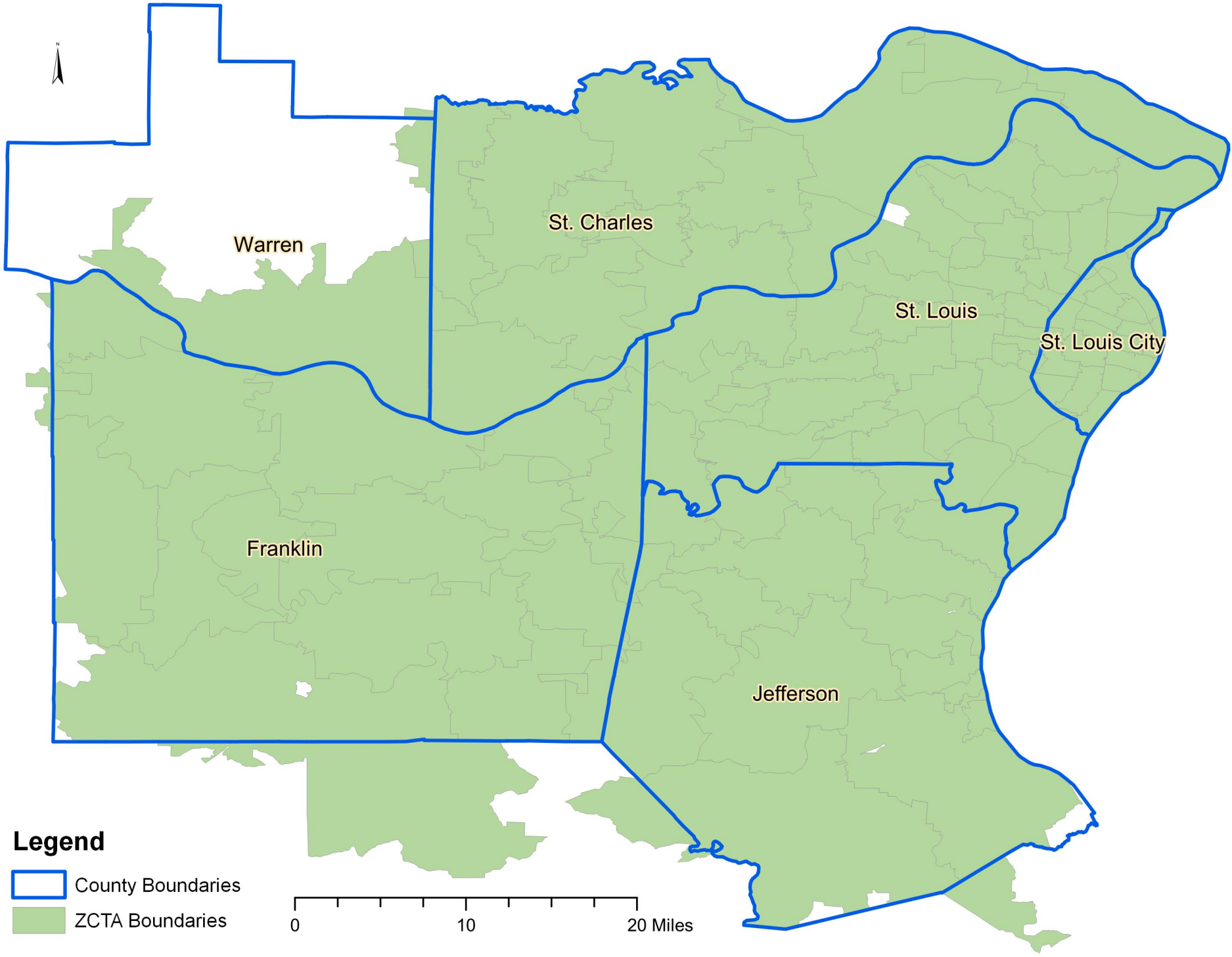
566 **Figure 1: Map of study area showing geographic distribution of Zip Code Tabulation**
567 **Areas and Counties**

568

569 **Figure 2: Geographic distribution of ZCTA-level COVID-19 age-adjusted**
570 **hospitalization risks and its significant predictors in the St. Louis area, Missouri.**

571

572 **Figure 3: Geographically Varying Coefficients of the Local Geographically**
573 **Weighted Negative Binomial Model Predicting COVID-19 Age-adjusted**
574 **Hospitalization Risks in the St. Louis area**



Warren

St. Charles



St. Louis

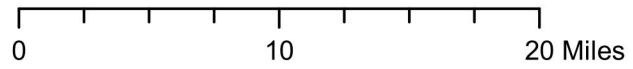
St. Louis City

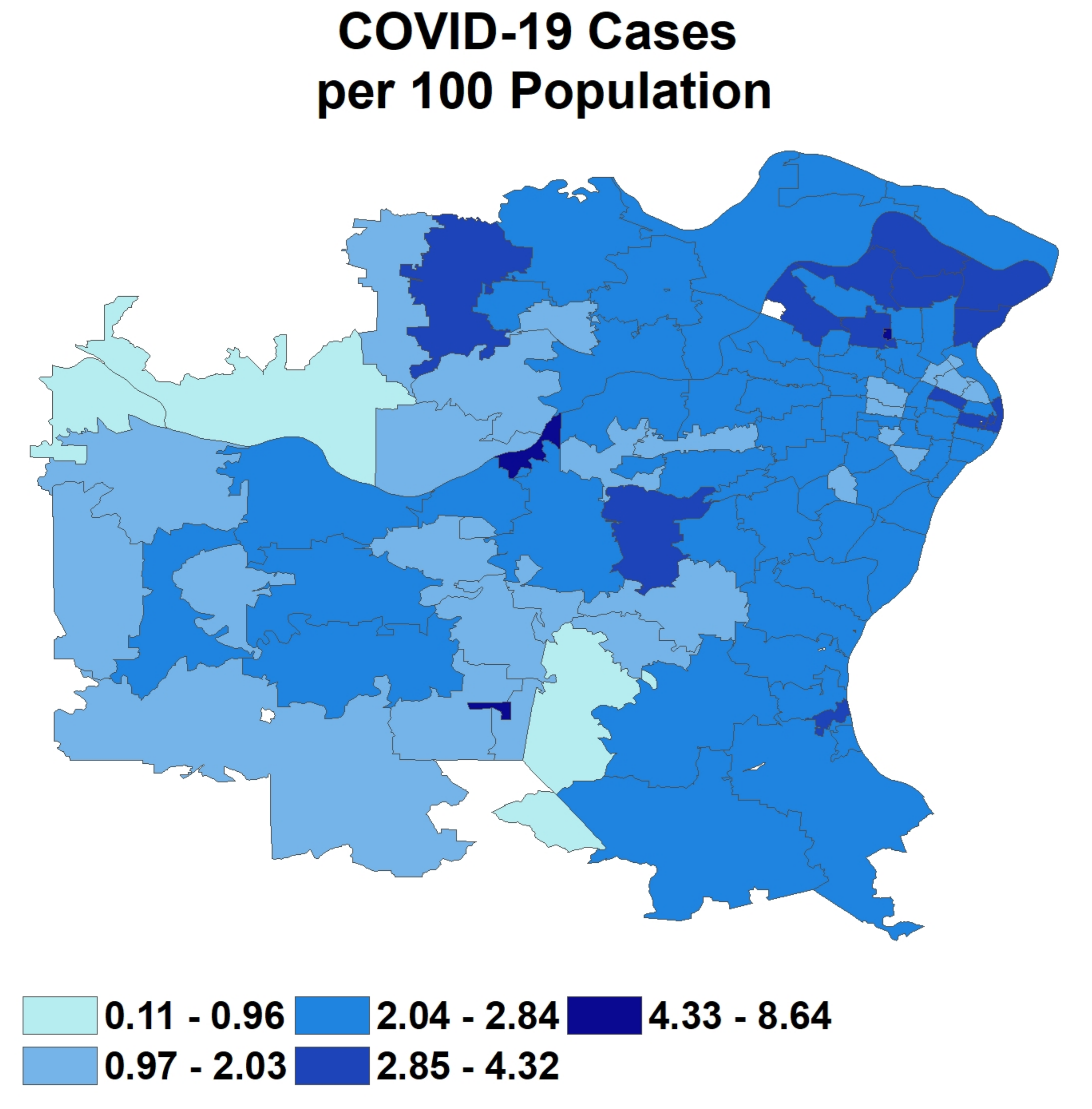
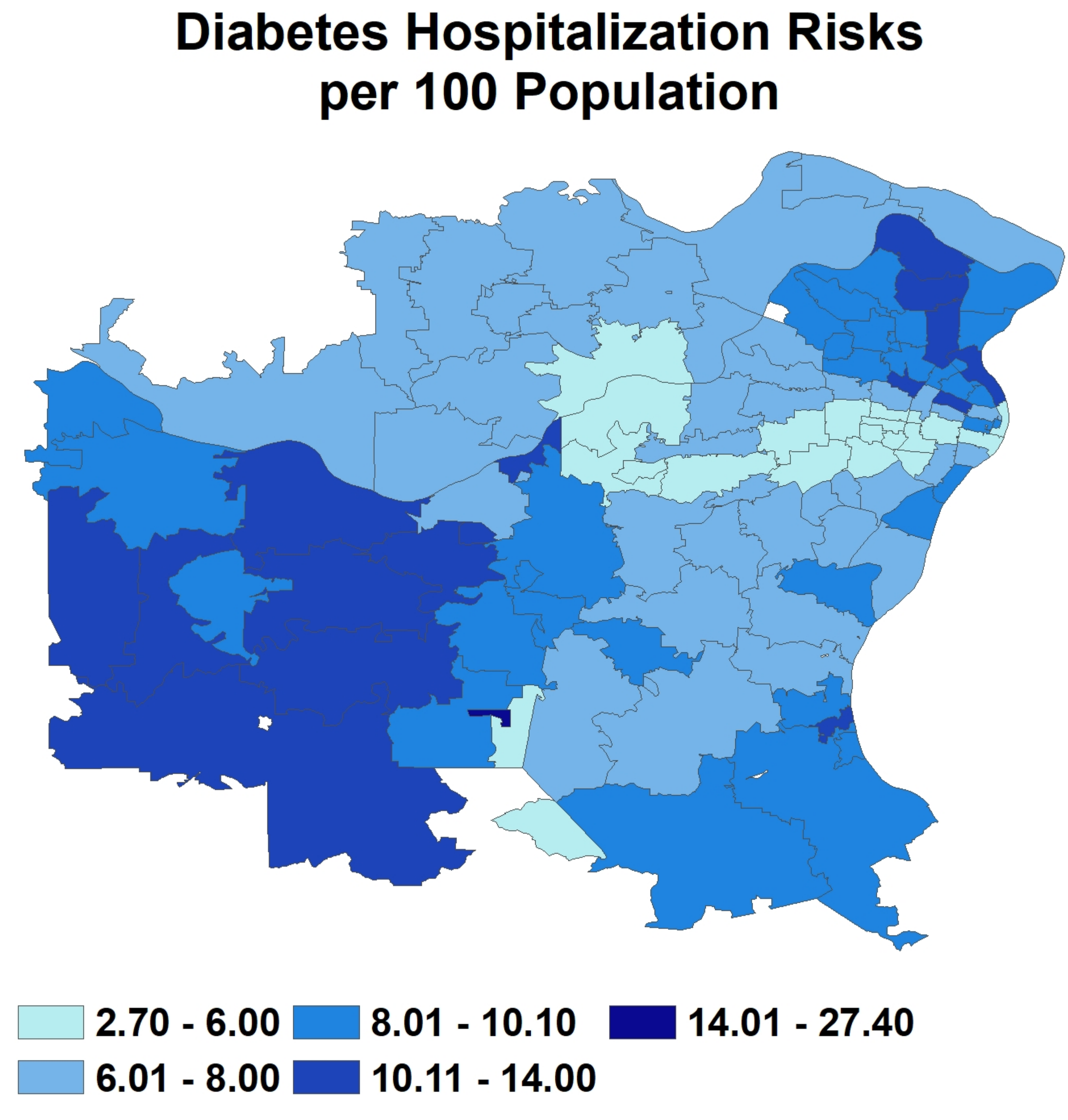
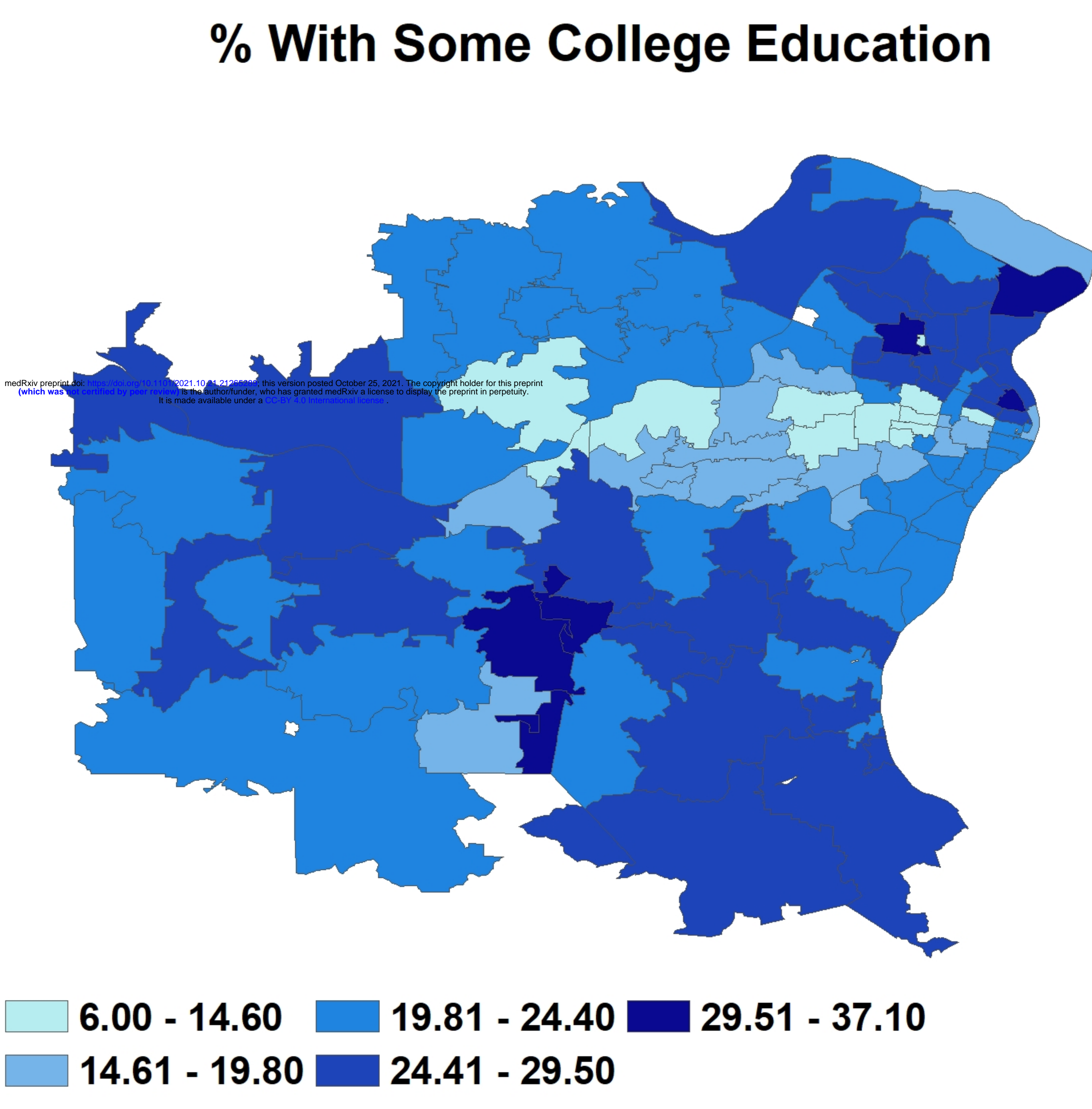
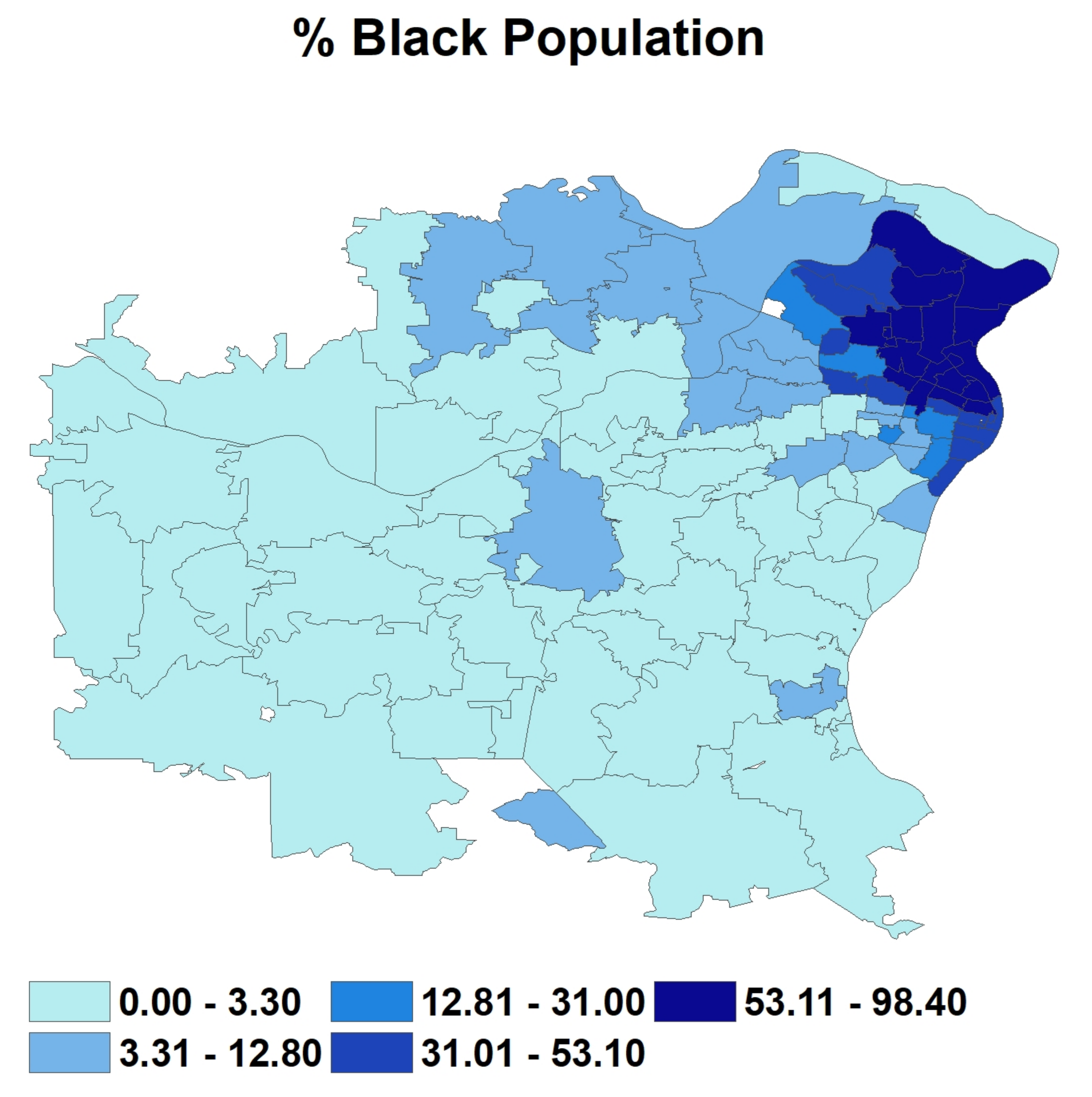
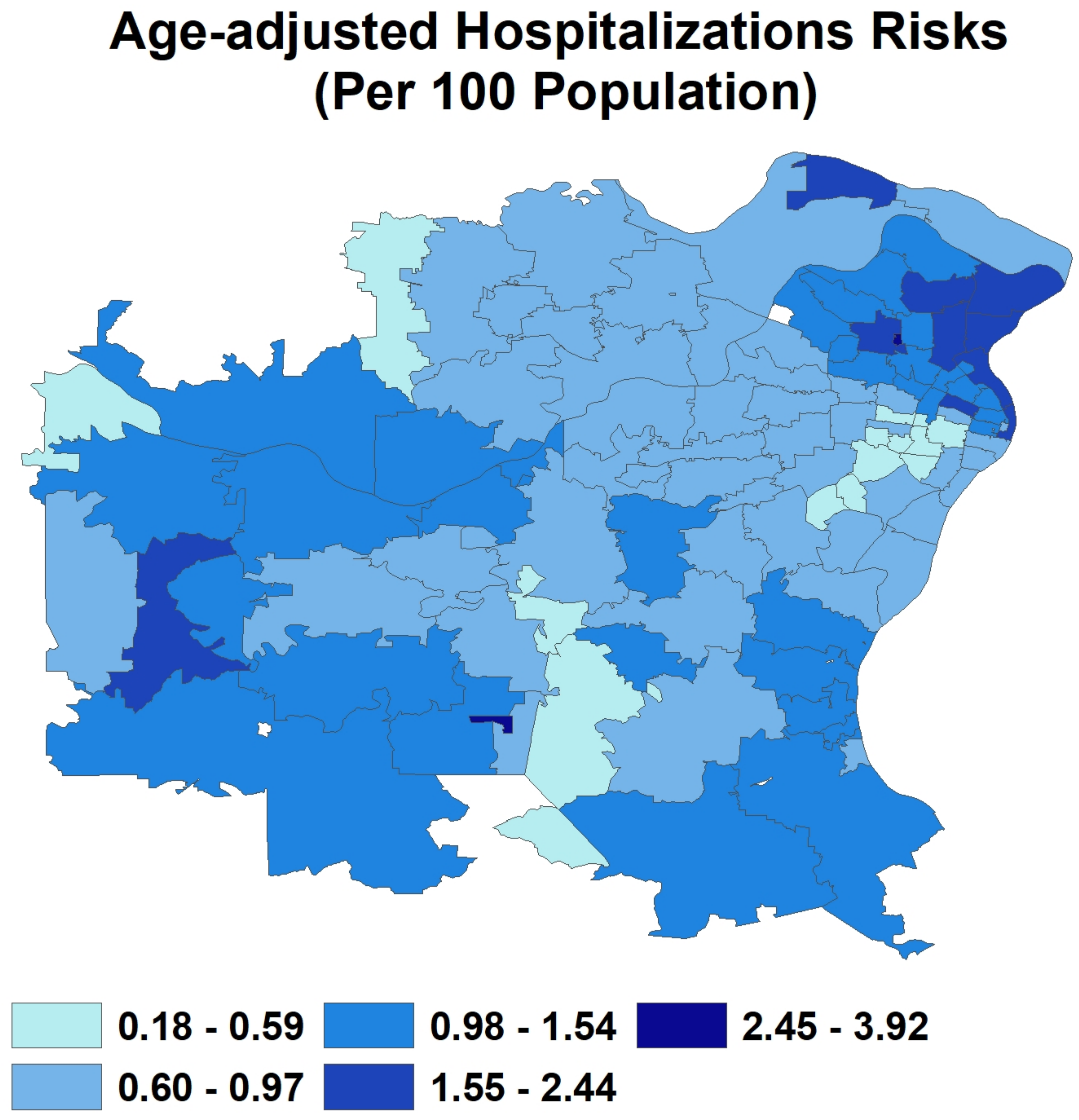
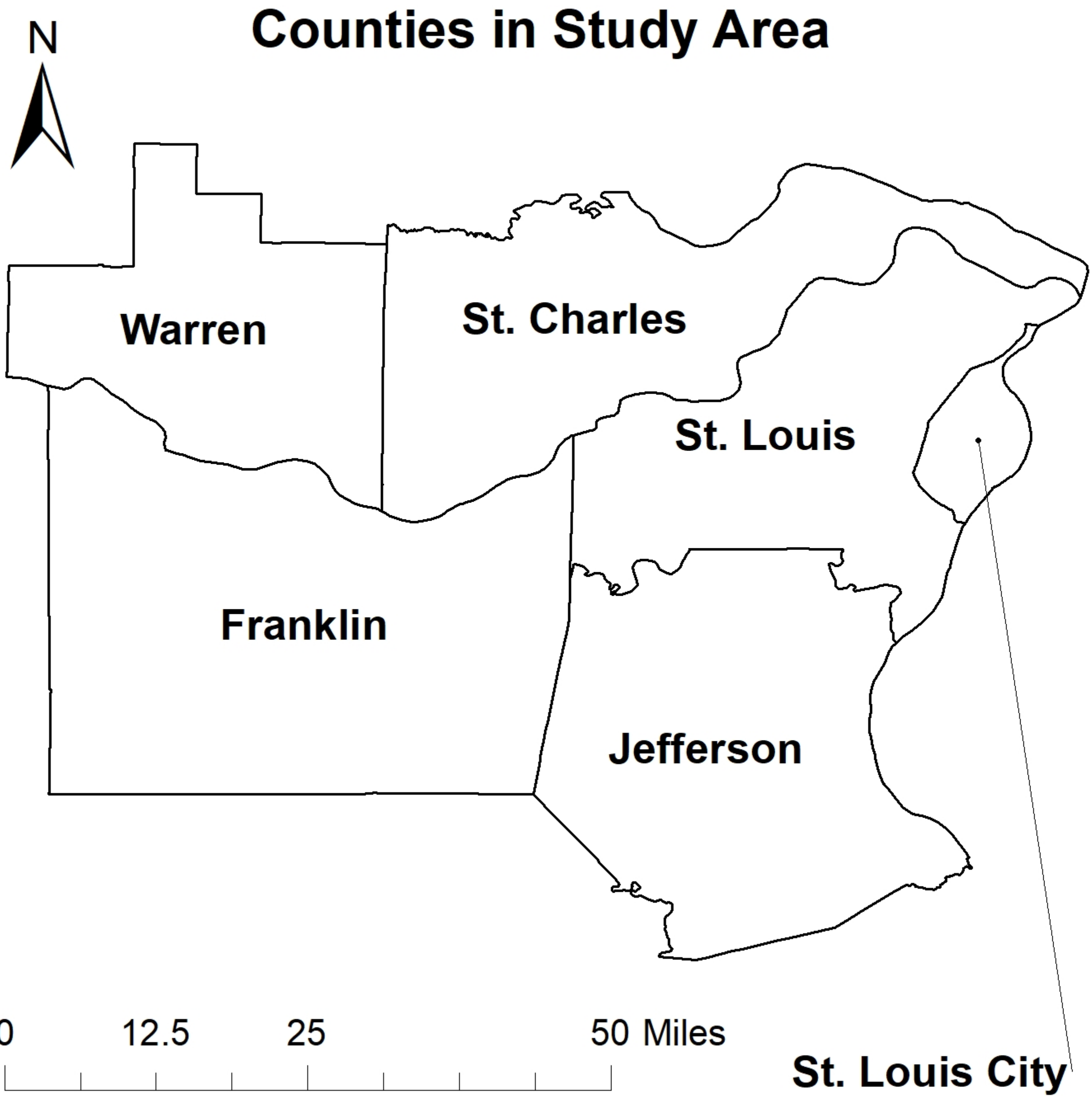
Franklin

Jefferson

Legend

-  County Boundaries
-  ZCTA Boundaries



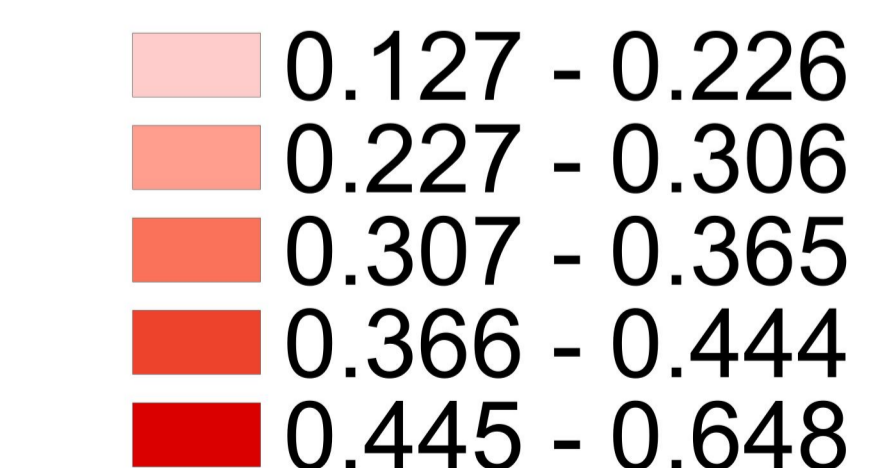
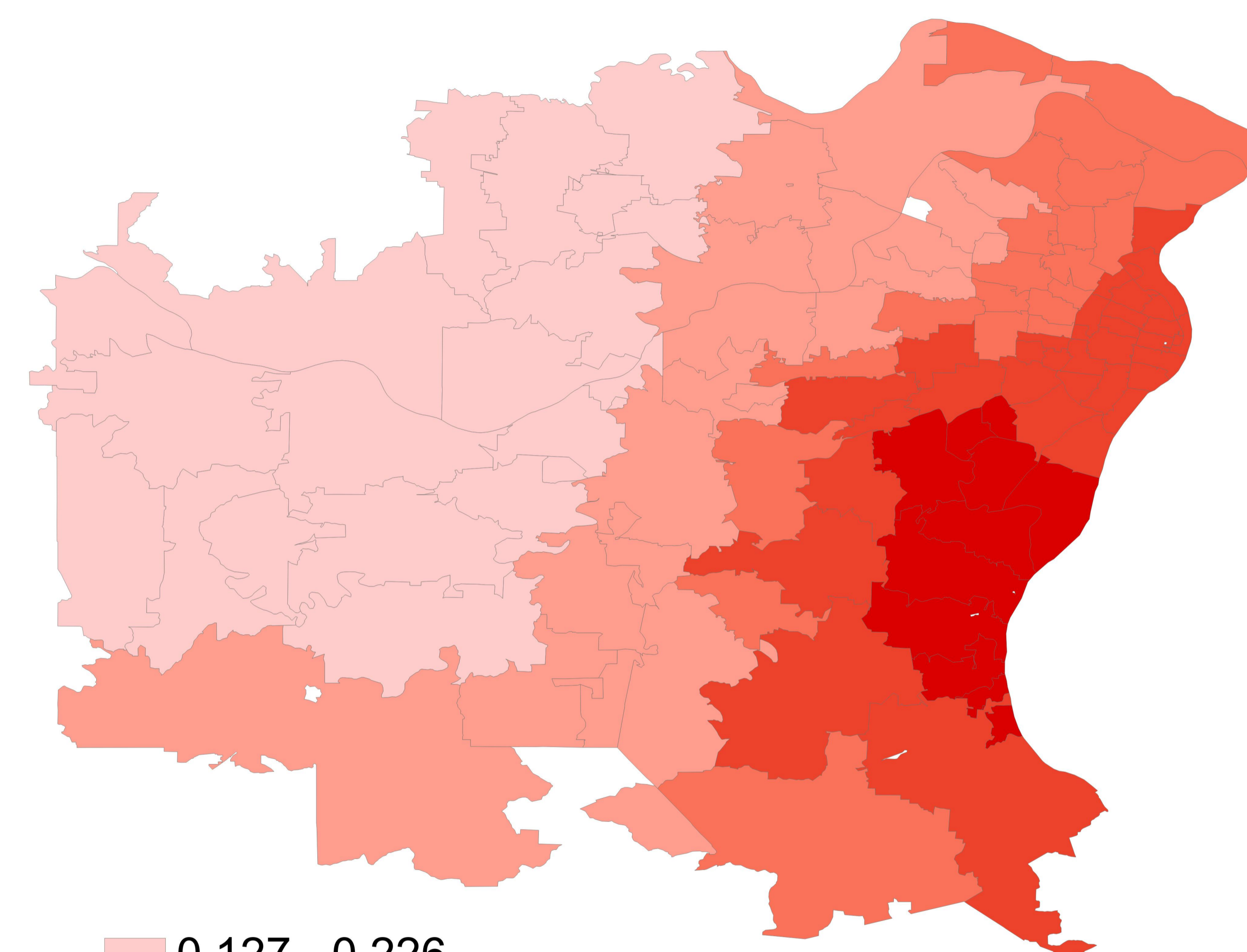
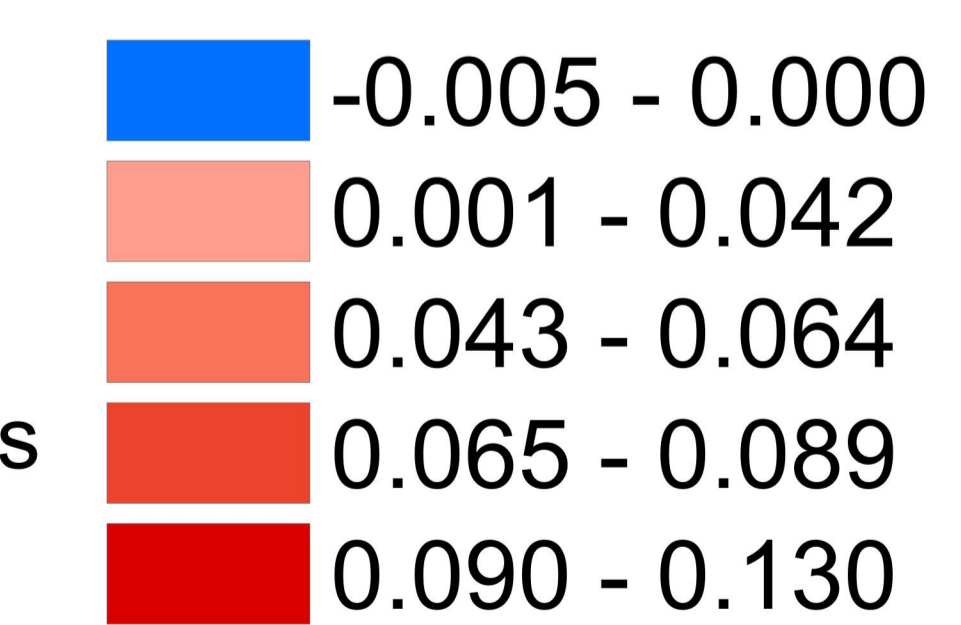
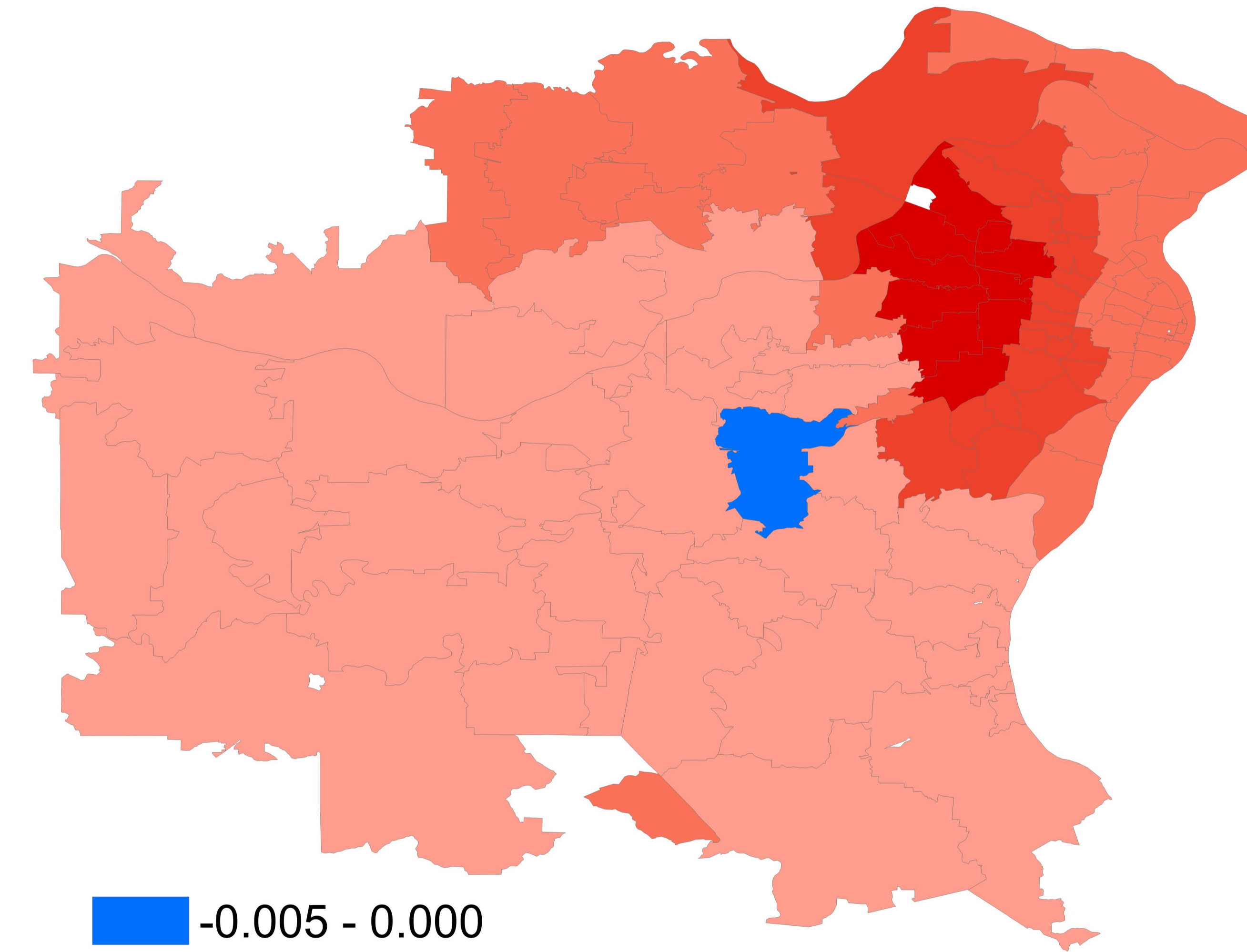
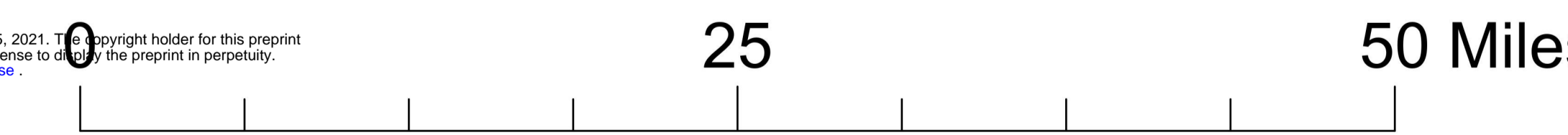
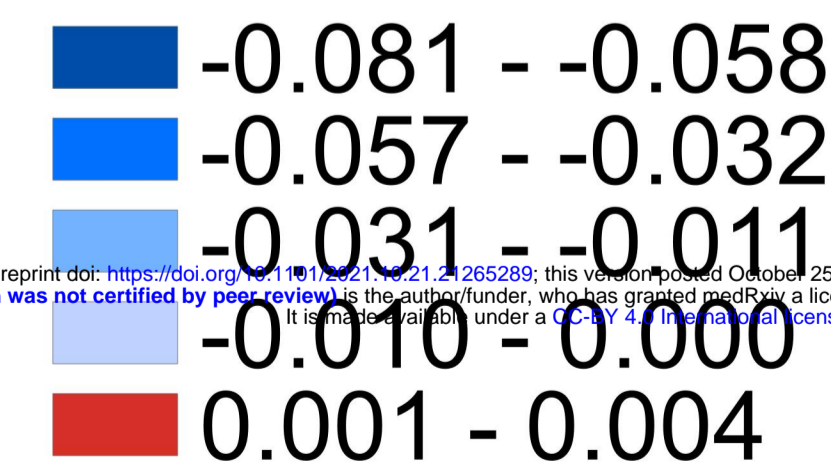
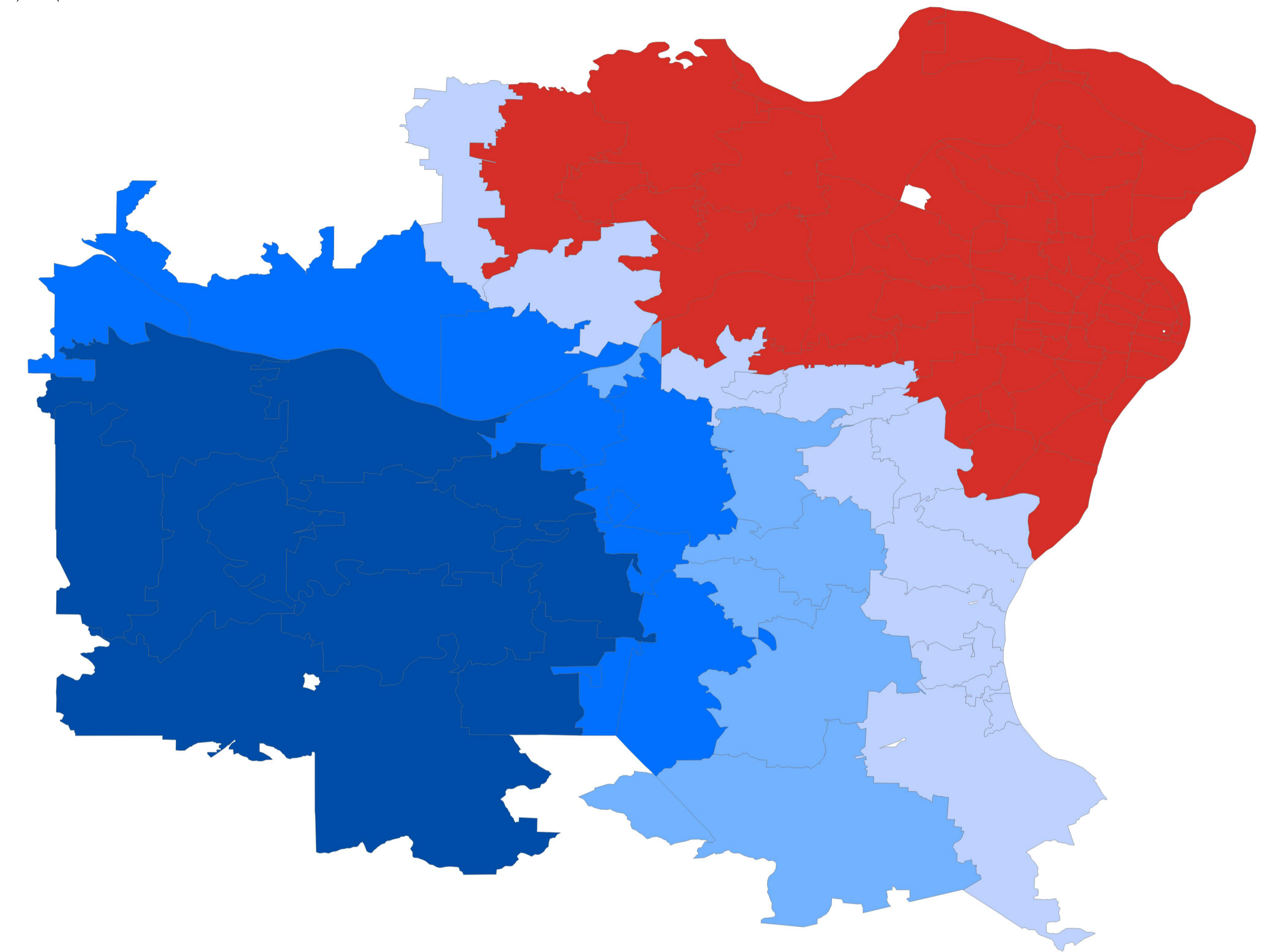


Source: U. S. Census Bureau, 2018 ACS; COVID Case Data from MO Depts, of Health; COVID Hosp. Data from Missouri Hospital Association

% Black Population Coefficient

% Diabetes Hospitalization Coefficient

COVID-19 Cases per 100 Population Coefficient



Source: U. S. Census Bureau, 2018 ACS; COVID Case Data from MO Depts, of Health; COVID Hosp. Data from Missouri Hospital Association