Analytic Plan & Findings

JP

2/25/2021

## Correlations between variables in models

imputed <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/cen\_imputed.rds')  
  
imputed %>%   
 rename(activity = ltpa\_2020c,  
 access = rec\_resource\_2018c,  
 rural = rural\_2020c,  
 white = white\_per\_2016c,  
 black = black\_per\_2016c,  
 latino = latino\_per\_2016c,  
 crime = county\_crime\_2020c,  
 income = med\_inc\_2016\_thousandc) %>%   
 dplyr::select(access:income) %>%   
 cor()

## access activity rural white black  
## access 1.00000000 0.39374063 -0.57070932 0.05334722 -0.14160665  
## activity 0.39374063 1.00000000 -0.31674143 0.16104550 -0.28178974  
## rural -0.57070932 -0.31674143 1.00000000 0.21008278 -0.09931641  
## white 0.05334722 0.16104550 0.21008278 1.00000000 -0.85916023  
## black -0.14160665 -0.28178974 -0.09931641 -0.85916023 1.00000000  
## latino 0.07481133 0.15660431 -0.25059427 -0.09193183 -0.09040296  
## crime 0.10049914 -0.09906138 -0.34933464 -0.38950109 0.36708721  
## income 0.34573423 0.50281965 -0.33330540 0.14784569 -0.26924761  
## latino crime income  
## access 0.07481133 0.10049914 0.3457342  
## activity 0.15660431 -0.09906138 0.5028196  
## rural -0.25059427 -0.34933464 -0.3333054  
## white -0.09193183 -0.38950109 0.1478457  
## black -0.09040296 0.36708721 -0.2692476  
## latino 1.00000000 0.09163311 0.0694618  
## crime 0.09163311 1.00000000 -0.1442696  
## income 0.06946180 -0.14426963 1.0000000

imputed %>%   
 rename(activity = ltpa\_2020c,  
 access = rec\_resource\_2018c,  
 rural = rural\_2020c,  
 white = white\_per\_2016c,  
 black = black\_per\_2016c,  
 latino = latino\_per\_2016c,  
 crime = county\_crime\_2020c,  
 income = med\_inc\_2016\_thousandc) %>%   
 dplyr::select(access:income) %>%  
 inspectdf::inspect\_cor()

## # A tibble: 28 x 7  
## col\_1 col\_2 corr p\_value lower upper pcnt\_nna  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 black white -0.859 0. -0.868 -0.850 100  
## 2 rural access -0.571 2.43e-224 -0.594 -0.547 100  
## 3 income activity 0.503 1.31e-174 0.476 0.528 100  
## 4 activity access 0.394 7.65e-108 0.364 0.423 100  
## 5 crime white -0.390 1.42e-105 -0.419 -0.359 100  
## 6 crime black 0.367 5.45e- 94 0.336 0.397 100  
## 7 crime rural -0.349 2.67e- 85 -0.380 -0.318 100  
## 8 income access 0.346 1.37e- 83 0.315 0.376 100  
## 9 income rural -0.333 8.05e- 78 -0.364 -0.302 100  
## 10 rural activity -0.317 1.85e- 70 -0.348 -0.285 100  
## # ... with 18 more rows

Notes.

* All the variables were centered to reduce potential multicollinearity with the inclusion of the interactions for the 2nd and 3rd models.
* The percentage of White residents within each county was removed from the models due to its high correlation with Black/African American residents within each county (*r* = -.859, *p* < .001).

## Regression Assumptions

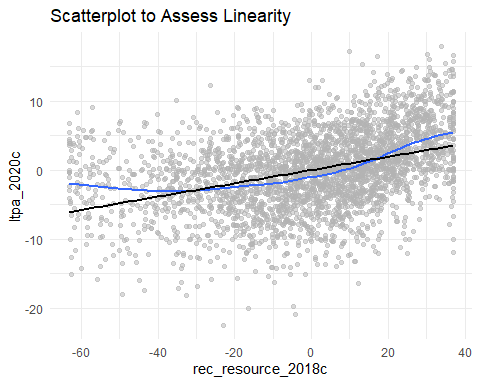
imputed %>%   
 rename(activity = ltpa\_2020c,  
 access = rec\_resource\_2018c,  
 rural = rural\_2020c,  
 white = white\_per\_2016c,  
 black = black\_per\_2016c,  
 latino = latino\_per\_2016c,  
 crime = county\_crime\_2020c,  
 income = med\_inc\_2016\_thousandc) %>%   
 dplyr::select(access:income,  
 -white) %>%   
 psych::describe(., na.rm = TRUE)

## vars n mean sd median trimmed mad min max range  
## access 1 3142 0 23.14 3.27 1.81 22.50 -63.02 36.98 100.00  
## activity 2 3142 0 5.65 0.11 0.06 5.49 -22.49 17.91 40.40  
## rural 3 3142 0 31.20 0.89 1.44 38.36 -58.59 41.41 100.00  
## black 4 3142 0 13.33 -5.50 -3.38 2.55 -7.72 78.47 86.18  
## latino 5 3142 0 12.73 -4.40 -3.07 3.00 -8.31 90.65 98.96  
## crime 6 3142 0 186.49 -44.39 -26.02 136.60 -248.37 1571.14 1819.51  
## income 7 3142 0 11.24 -1.29 -1.02 7.88 -28.42 78.28 106.70  
## skew kurtosis se  
## access -0.66 -0.04 0.41  
## activity -0.12 0.19 0.10  
## rural -0.16 -1.11 0.56  
## black 2.71 7.76 0.24  
## latino 3.47 14.24 0.23  
## crime 1.97 6.75 3.33  
## income 1.47 4.81 0.20

linearity(imputed, rec\_resource\_2018c, ltpa\_2020c, se = FALSE, size = 1)

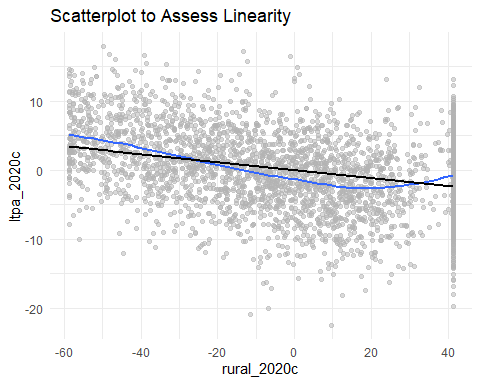
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## `geom\_smooth()` using formula 'y ~ x'



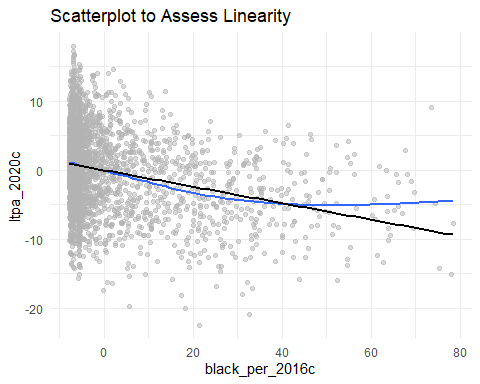
linearity(imputed, rural\_2020c, ltpa\_2020c, se = FALSE, size = 1)

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## `geom\_smooth()` using formula 'y ~ x'



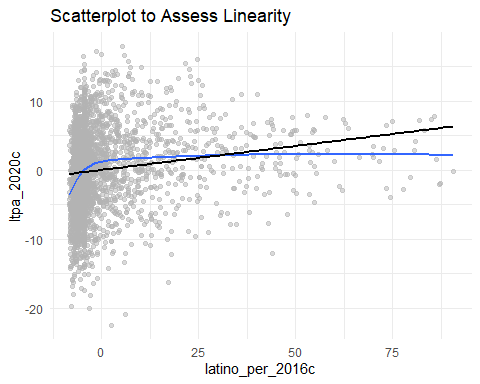
linearity(imputed, black\_per\_2016c, ltpa\_2020c, se = FALSE, size = 1)

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## `geom\_smooth()` using formula 'y ~ x'



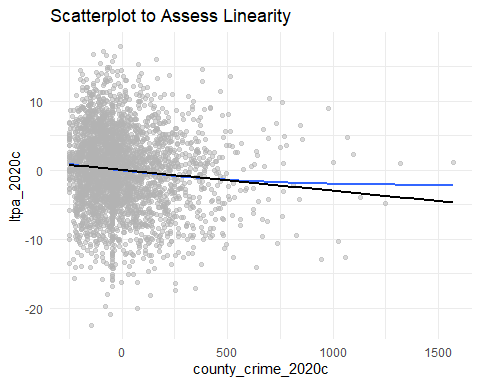
linearity(imputed, latino\_per\_2016c, ltpa\_2020c, se = FALSE, size = 1)

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## `geom\_smooth()` using formula 'y ~ x'



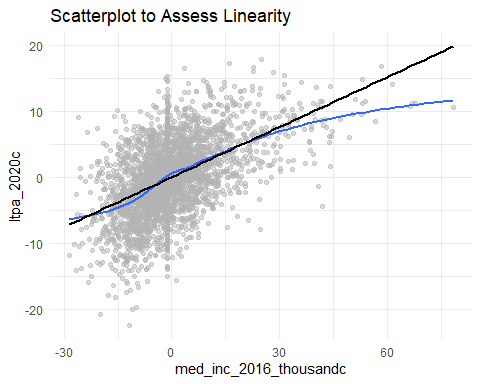
linearity(imputed, county\_crime\_2020c, ltpa\_2020c, se = FALSE, size = 1)

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## `geom\_smooth()` using formula 'y ~ x'



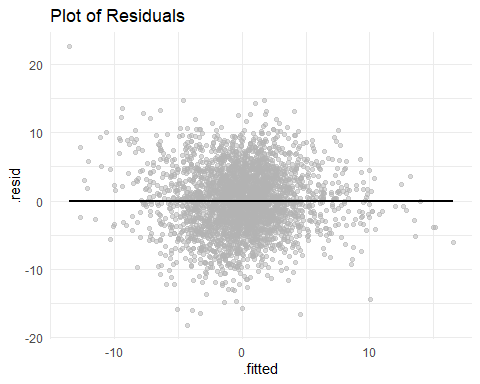
linearity(imputed, med\_inc\_2016\_thousandc, ltpa\_2020c, se = FALSE, size = 1)

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'  
## `geom\_smooth()` using formula 'y ~ x'

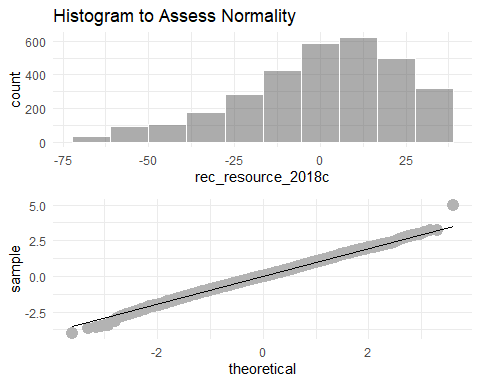


predictors <- c('rec\_resource\_2018c', 'rural\_2020c', 'black\_per\_2016c', 'latino\_per\_2016c', 'county\_crime\_2020c', 'med\_inc\_2016\_thousandc')  
residual\_view(data = imputed, x = predictors, y = 'ltpa\_2020c', se = FALSE, size = 1)

## `geom\_smooth()` using formula 'y ~ x'

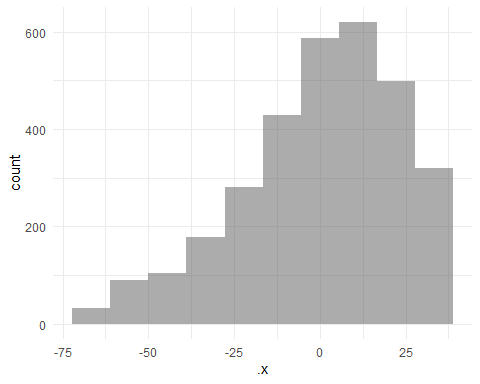


normality\_view(data = imputed, x = rec\_resource\_2018c, model\_x = predictors, y = 'ltpa\_2020c',   
 alpha = .5, bins = 10, se = FALSE)

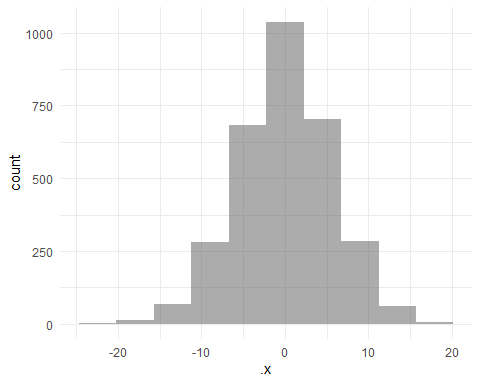


num\_only <- imputed %>%   
 dplyr::select(rec\_resource\_2018c:med\_inc\_2016\_thousandc,  
 -white\_per\_2016c)  
  
map(num\_only, ~ggplot(num\_only, aes(.x)) +  
 geom\_histogram(alpha = .5, bins = 10) +   
 theme\_minimal())

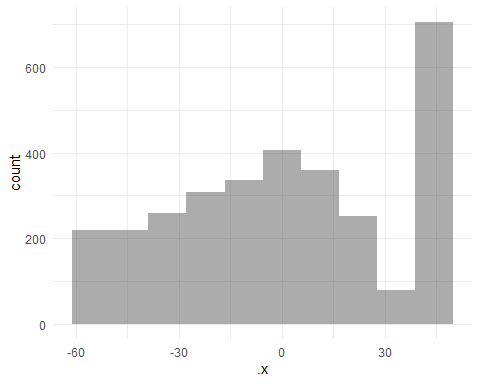
## $rec\_resource\_2018c



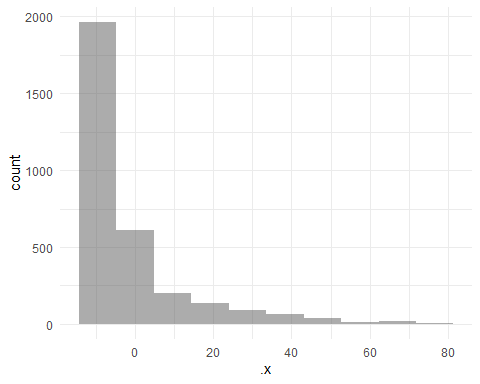
##   
## $ltpa\_2020c



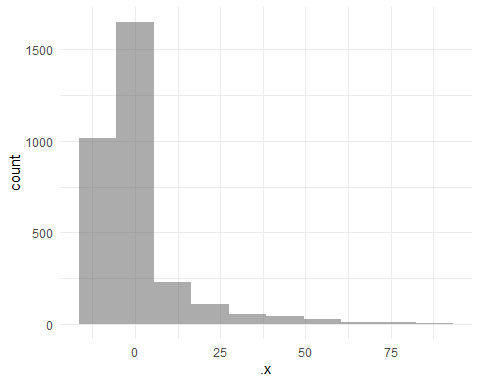
##   
## $rural\_2020c



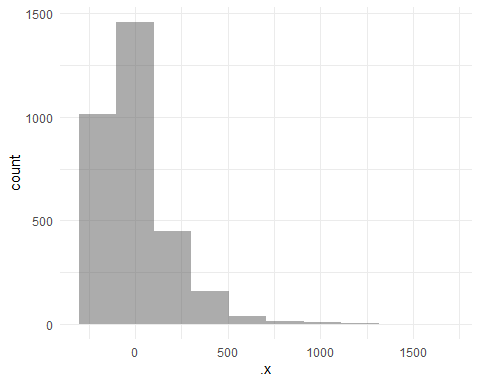
##   
## $black\_per\_2016c



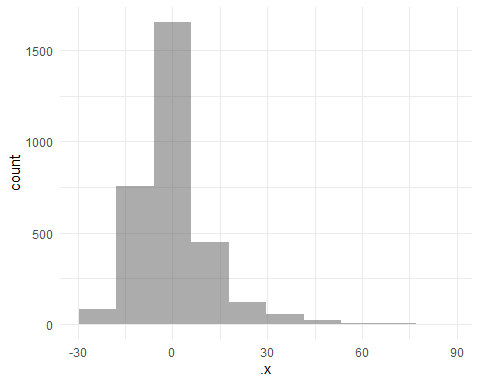
##   
## $latino\_per\_2016c



##   
## $county\_crime\_2020c



##   
## $med\_inc\_2016\_thousandc



Notes.

* The residual plot may not be random. This could indicate spatial autocorrelation.
* There doesn’t seem to be any indication of quadratic relationships.
* Q-Q plot appears to be good.
* Histograms and descriptive statistics suggests non-normal distributions of rurality, Black/African resident population, Latino resident population, violent crime, and median household income.

**For Raoul** Should there be transformations to the variables with non-normal distributions (rurality, Black/African resident population, Latino resident population, violent crime, and median household income)? Is this a problem if this is how these variables are actually distributed in the United States?

## Models Tested

## Creating Neighbors

poly\_nb\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/poly\_nb\_queenc.rds')  
poly\_nb\_queen

## Neighbour list object:  
## Number of regions: 2625   
## Number of nonzero links: 14866   
## Percentage nonzero weights: 0.2157424   
## Average number of links: 5.663238   
## 12 regions with no links:  
## 68 69 70 71 72 73 521 522 524 1170 1174 2519

poly\_listw\_queenc <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/poly\_listw\_queenc.rds')  
  
# poly\_nb\_queenc <- poly2nb(acs\_polyc, row.names = acs\_polyc$rowid, queen = TRUE)  
  
# poly\_listw\_queenc <- nb2listw(poly\_nb\_queenc, style = "W", zero.policy = TRUE)

Notes.

* Queen contiguity weights were used. I used these weights because I read from several sources that these are more accurate for examining irregular polygons, such as counties.
* Under the section function, it calls for a weights list for neighbors. I chose W since it made the most sense to me about standardizing the rows to equal 1. Below are the options.

From Starting from a binary neighbours list, in which regions are either listed as neighbours or are absent (thus not in the set of neighbours for some definition), the function adds a weights list with values given by the coding scheme style chosen. B is the basic binary coding, W is row standardised (sums over all links to n), C is globally standardised (sums over all links to n), U is equal to C divided by the number of neighbours (sums over all links to unity), while S is the variance-stabilizing coding scheme proposed by Tiefelsdorf et al. 1999, p. 167-168 (sums over all links to n).

**For Raoul** For the weights, is there any weight list that is better than one another? Are queen contiguity weights appropriate or should I use rook contiguity weights?

## Moran Values for Each Variable

ltpa\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/moran\_ltpa\_queenc.rds')  
ltpa\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$ltpa\_2020c   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 42.129, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.5044269422 -0.0003828484 0.0001435828

rec\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/moran\_rec\_queenc.rds')  
rec\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$rec\_resource\_2018c   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 26.563, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.3179206102 -0.0003828484 0.0001435918

crime\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/crime\_queenc.rds')  
crime\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$county\_crime\_2020c   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 22.982, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.2746004510 -0.0003828484 0.0001431700

income\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/income\_queenc.rds')  
income\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$med\_inc\_2016\_thousandc   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 50.693, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.6066609012 -0.0003828484 0.0001433999

rural\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/rural\_queenc.rds')  
rural\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$rural\_2020c   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 25.418, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.3042673886 -0.0003828484 0.0001436514

black\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/black\_queenc.rds')  
black\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$black\_per\_2016c   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 64.737, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.7744779521 -0.0003828484 0.0001432669

latino\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/latino\_queenc.rds')  
latino\_queen

##   
## Moran I test under randomisation  
##   
## data: acs\_polyc$latino\_per\_2016c   
## weights: poly\_listw\_queenc n reduced by no-neighbour observations  
##   
##   
## Moran I statistic standard deviate = 66.996, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Moran I statistic Expectation Variance   
## 0.8006784976 -0.0003828484 0.0001429672

Notes.

* These are the Moran’s I values for each variable included in the models.

## 

ols\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/ols\_modelc.rds')  
summary(ols\_queen)

##   
## Call:  
## lm(formula = model\_variablesc, data = acs\_polyc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.0577 -3.0116 -0.0101 3.0452 22.0188   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.1683509 0.0902417 -1.866 0.0622 .   
## rec\_resource\_2018c 0.0474110 0.0050487 9.391 < 0.0000000000000002 \*\*\*  
## county\_crime\_2020c -0.0011048 0.0005735 -1.926 0.0542 .   
## med\_inc\_2016\_thousandc 0.1881766 0.0086670 21.712 < 0.0000000000000002 \*\*\*  
## rural\_2020c -0.0104783 0.0041171 -2.545 0.0110 \*   
## black\_per\_2016c -0.0527295 0.0074179 -7.108 0.00000000000151 \*\*\*  
## latino\_per\_2016c 0.0461964 0.0069531 6.644 0.00000000003699 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.586 on 2618 degrees of freedom  
## Multiple R-squared: 0.3717, Adjusted R-squared: 0.3702   
## F-statistic: 258.1 on 6 and 2618 DF, p-value: < 0.00000000000000022

AIC(ols\_queen)

## [1] 15454.55

Notes.

* This model is to be able to test the residuals for the moran’s test and the LM test.

ols\_model\_moran\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/ols\_model\_moran\_queenc.rds')  
ols\_model\_moran\_queen

##   
## Global Moran I for regression residuals  
##   
## data:   
## model: lm(formula = model\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## Moran I statistic standard deviate = 23.531, p-value <  
## 0.00000000000000022  
## alternative hypothesis: two.sided  
## sample estimates:  
## Observed Moran I Expectation Variance   
## 0.2797964378 -0.0014635423 0.0001428736

**For Raoul** I’m curious on how to report this Moran’s I value compared to the Moran’s I for each variable. This tells me that the residuals of the model for the OLS model has evidence of spatial dependence while the other values represent evidence that each variable is spatially dependent. Is that how these values would be interpreted?

lm\_spa\_tests\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/lm\_spa\_tests\_queenc.rds')  
lm\_spa\_tests\_queen

##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = model\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## LMerr = 549.06, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = model\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## LMlag = 627.98, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = model\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## RLMerr = 14.043, df = 1, p-value = 0.0001786  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = model\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## RLMlag = 92.959, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = model\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## SARMA = 642.02, df = 2, p-value < 0.00000000000000022

**For Raoul** Since the LM tests indicate that a spatial lag model has a higher value, is this enough evidence of running this model over the spatial error model? Also the SARMA model has a higher value than the others but the resources I used stated that this is rarely the most appropriate model to run. I have also not found anything on how to run this particular model in R.

The model tested was a spatial lag model

sar\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/model\_sar\_queenc.rds')  
summary(sar\_queen, zstats = TRUE, Nagelkerke = TRUE)

##   
## Call:  
## lagsarlm(formula = model\_variablesc, data = acs\_polyc, listw = poly\_listw\_queenc,   
## zero.policy = TRUE)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.488998 -2.541086 0.016888 2.627849 18.318443   
##   
## Type: lag   
## Regions with no neighbours included:  
## 68 69 70 71 72 73 521 522 524 1170 1174 2519   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.07888684 0.08026250 -0.9829 0.3256761  
## rec\_resource\_2018c 0.02839286 0.00451497 6.2886 0.0000000003203  
## county\_crime\_2020c -0.00098036 0.00050990 -1.9227 0.0545234  
## med\_inc\_2016\_thousandc 0.12266398 0.00837442 14.6475 < 0.00000000000000022  
## rural\_2020c -0.01272117 0.00366181 -3.4740 0.0005127  
## black\_per\_2016c -0.02288565 0.00668743 -3.4222 0.0006212  
## latino\_per\_2016c 0.01878377 0.00623362 3.0133 0.0025842  
##   
## Rho: 0.4757, LR test value: 486.84, p-value: < 0.000000000000000222  
## Asymptotic standard error: 0.020935  
## z-value: 22.723, p-value: < 0.000000000000000222  
## Wald statistic: 516.33, p-value: < 0.000000000000000222  
##   
## Log likelihood: -7475.855 for lag model  
## ML residual variance (sigma squared): 16.621, (sigma: 4.0769)  
## Nagelkerke pseudo-R-squared: 0.47804   
## Number of observations: 2625   
## Number of parameters estimated: 9   
## AIC: 14970, (AIC for lm: 15455)  
## LM test for residual autocorrelation  
## test value: 22.922, p-value: 0.0000016872

AIC(sar\_queen)

## [1] 14969.71

sar\_model\_find <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/sar\_model\_find.rds')  
sar\_model\_find

## Impact measures (lag, exact):  
## Direct Indirect Total  
## rec\_resource\_2018c 0.029892347 0.0241437230 0.054036070  
## county\_crime\_2020c -0.001032132 -0.0008336418 -0.001865774  
## med\_inc\_2016\_thousandc 0.129142129 0.1043066922 0.233448821  
## rural\_2020c -0.013393002 -0.0108173817 -0.024210383  
## black\_per\_2016c -0.024094288 -0.0194606942 -0.043554982  
## latino\_per\_2016c 0.019775782 0.0159726839 0.035748466  
## ========================================================  
## Simulation results (asymptotic variance matrix):  
## Direct:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## rec\_resource\_2018c 0.033019 0.0049188 0.0028399 0.0028399  
## county\_crime\_2020c -0.001167 0.0006486 0.0003745 0.0003745  
## med\_inc\_2016\_thousandc 0.131485 0.0063163 0.0036467 0.0036467  
## rural\_2020c -0.014526 0.0037472 0.0021634 0.0021634  
## black\_per\_2016c -0.021893 0.0103155 0.0059557 0.0059557  
## latino\_per\_2016c 0.014837 0.0049078 0.0028335 0.0028335  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## rec\_resource\_2018c 0.029855 0.030188 0.0305566 0.0346199 0.0382768  
## county\_crime\_2020c -0.001861 -0.001376 -0.0008377 -0.0007937 -0.0007542  
## med\_inc\_2016\_thousandc 0.125872 0.128142 0.1306647 0.1344183 0.1377965  
## rural\_2020c -0.017195 -0.016662 -0.0160708 -0.0131621 -0.0105442  
## black\_per\_2016c -0.030485 -0.027542 -0.0242708 -0.0174336 -0.0112802  
## latino\_per\_2016c 0.010838 0.012127 0.0135591 0.0169085 0.0199229  
##   
## ========================================================  
## Indirect:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## rec\_resource\_2018c 0.0250860 0.0022602 0.0013049 0.0013049  
## county\_crime\_2020c -0.0008905 0.0004868 0.0002811 0.0002811  
## med\_inc\_2016\_thousandc 0.1004106 0.0068831 0.0039740 0.0039740  
## rural\_2020c -0.0112047 0.0033878 0.0019559 0.0019559  
## black\_per\_2016c -0.0169196 0.0087515 0.0050527 0.0050527  
## latino\_per\_2016c 0.0112349 0.0031712 0.0018309 0.0018309  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## rec\_resource\_2018c 0.023144 0.023866 0.024668 0.0260971 0.0273830  
## county\_crime\_2020c -0.001407 -0.001069 -0.000693 -0.0006133 -0.0005415  
## med\_inc\_2016\_thousandc 0.094991 0.096567 0.098319 0.1032080 0.1076080  
## rural\_2020c -0.013282 -0.013159 -0.013023 -0.0101594 -0.0075823  
## black\_per\_2016c -0.025079 -0.021380 -0.017270 -0.0126342 -0.0084617  
## latino\_per\_2016c 0.008230 0.009645 0.011217 0.0128161 0.0142550  
##   
## ========================================================  
## Total:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## rec\_resource\_2018c 0.058105 0.007031 0.0040595 0.0040595  
## county\_crime\_2020c -0.002058 0.001134 0.0006548 0.0006548  
## med\_inc\_2016\_thousandc 0.231896 0.009990 0.0057677 0.0057677  
## rural\_2020c -0.025731 0.007100 0.0040991 0.0040991  
## black\_per\_2016c -0.038813 0.019002 0.0109706 0.0109706  
## latino\_per\_2016c 0.026072 0.008031 0.0046369 0.0046369  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## rec\_resource\_2018c 0.053664 0.054054 0.054487 0.060348 0.065623  
## county\_crime\_2020c -0.003268 -0.002445 -0.001531 -0.001407 -0.001296  
## med\_inc\_2016\_thousandc 0.221238 0.228463 0.236491 0.237626 0.238648  
## rural\_2020c -0.030231 -0.029821 -0.029366 -0.023458 -0.018140  
## black\_per\_2016c -0.055565 -0.048922 -0.041541 -0.030068 -0.019742  
## latino\_per\_2016c 0.019068 0.021772 0.024776 0.029725 0.034178  
##   
## ========================================================  
## Simulated standard errors  
## Direct Indirect Total  
## rec\_resource\_2018c 0.0049188102 0.0022601821 0.007031291  
## county\_crime\_2020c 0.0006486039 0.0004868309 0.001134134  
## med\_inc\_2016\_thousandc 0.0063163396 0.0068831050 0.009990001  
## rural\_2020c 0.0037471861 0.0033877619 0.007099856  
## black\_per\_2016c 0.0103155048 0.0087514973 0.019001615  
## latino\_per\_2016c 0.0049078297 0.0031712215 0.008031339  
##   
## Simulated z-values:  
## Direct Indirect Total  
## rec\_resource\_2018c 6.712890 11.099099 8.263834  
## county\_crime\_2020c -1.799783 -1.829262 -1.814501  
## med\_inc\_2016\_thousandc 20.816698 14.587984 23.212807  
## rural\_2020c -3.876494 -3.307391 -3.624101  
## black\_per\_2016c -2.122354 -1.933343 -2.042605  
## latino\_per\_2016c 3.023196 3.542759 3.246309  
##   
## Simulated p-values:  
## Direct Indirect   
## rec\_resource\_2018c 0.000000000019081 < 0.000000000000000222  
## county\_crime\_2020c 0.07189495 0.06736036   
## med\_inc\_2016\_thousandc < 0.000000000000000222 < 0.000000000000000222  
## rural\_2020c 0.00010597 0.00094169   
## black\_per\_2016c 0.03380800 0.05319399   
## latino\_per\_2016c 0.00250120 0.00039596   
## Total   
## rec\_resource\_2018c 0.00000000000000022204  
## county\_crime\_2020c 0.06960067   
## med\_inc\_2016\_thousandc < 0.000000000000000222  
## rural\_2020c 0.00028997   
## black\_per\_2016c 0.04109152   
## latino\_per\_2016c 0.00116912

Notes.

**For Raoul** I really would like help in how to interpret the findings. So just so I’m clear, the first summary cannot be interpreted due to spillover between the estimates on physical activity. So it relies on the impact measures of direct, indirect, and total effects (the second file loaded [e.g., sar\_model\_find]).

My understanding of the interpretations are: The estimate (*b* = 0.030) of the direct effect between access to recreational resources and physical activity would be that for a one unit increase in access in any given county, there would be a 0.028 increase in physical activity in that county.

The indirect effect would be for a one unit increase in access in neighboring counties, there would be a 0.024 increase in physical activity in that county.

The total effect would be the addition of these two effects. So for a one unit increase in access in a county and it’s neighboring counties, there would be a 0.054 increase in physical activity in the neighboring counties.

Any feedback on this would be greatly appreciated.

## Crime Interaction Model

**For Raoul** For the following two models, I ran everything exactly the same as the previous model. I’m not sure how to interpret the interactions. Specifically, I’m not sure how to interpret the significant indirect, direct, and total effects of the access to recreational resources and median household income interaction. Lastly, the spatial lag model apparently does not provide p values for these effects and they are simulated by running the model how ever many times requested in the code. I was able to run it three times on a remote server with 64gb of memory and it took a while to run with the number of observations in the data. I’m curious to know if three simulations is enough since it crashed trying to run anymore than that.

I also included my analytic plan for my study below.

int\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_modelc.rds')  
summary(int\_queen)

##   
## Call:  
## lm(formula = int\_variablesc, data = acs\_polyc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.1321 -3.0153 0.0107 3.0629 22.7570   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) -0.18345860 0.09075708 -2.021  
## rec\_resource\_2018c 0.04791451 0.00505817 9.473  
## county\_crime\_2020c -0.00114093 0.00057383 -1.988  
## med\_inc\_2016\_thousandc 0.18873124 0.00867238 21.762  
## rural\_2020c -0.00988857 0.00413404 -2.392  
## black\_per\_2016c -0.05265151 0.00741613 -7.100  
## latino\_per\_2016c 0.04634914 0.00695199 6.667  
## rec\_resource\_2018c:county\_crime\_2020c 0.00002884 0.00001885 1.530  
## Pr(>|t|)   
## (Intercept) 0.0433 \*   
## rec\_resource\_2018c < 0.0000000000000002 \*\*\*  
## county\_crime\_2020c 0.0469 \*   
## med\_inc\_2016\_thousandc < 0.0000000000000002 \*\*\*  
## rural\_2020c 0.0168 \*   
## black\_per\_2016c 0.00000000000161 \*\*\*  
## latino\_per\_2016c 0.00000000003172 \*\*\*  
## rec\_resource\_2018c:county\_crime\_2020c 0.1260   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.585 on 2617 degrees of freedom  
## Multiple R-squared: 0.3722, Adjusted R-squared: 0.3706   
## F-statistic: 221.7 on 7 and 2617 DF, p-value: < 0.00000000000000022

AIC(int\_queen)

## [1] 15454.21

int\_moran\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_moran\_queenc.rds')  
int\_moran\_queen

##   
## Global Moran I for regression residuals  
##   
## data:   
## model: lm(formula = int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## Moran I statistic standard deviate = 23.688, p-value <  
## 0.00000000000000022  
## alternative hypothesis: two.sided  
## sample estimates:  
## Observed Moran I Expectation Variance   
## 0.2816464686 -0.0014889525 0.0001428619

int\_spa\_tests\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_spa\_tests\_queenc.rds')  
int\_spa\_tests\_queen

##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## LMerr = 556.35, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## LMlag = 630.67, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## RLMerr = 14.611, df = 1, p-value = 0.0001321  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## RLMlag = 88.935, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## SARMA = 645.28, df = 2, p-value < 0.00000000000000022

int\_sar\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_sar\_queenc.rds')  
summary(int\_sar\_queen, zstats = TRUE, Nagelkerke = TRUE)

##   
## Call:  
## lagsarlm(formula = int\_variablesc, data = acs\_polyc, listw = poly\_listw\_queenc,   
## zero.policy = TRUE)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.5209216 -2.5426639 0.0013284 2.6345570 18.6325979   
##   
## Type: lag   
## Regions with no neighbours included:  
## 68 69 70 71 72 73 521 522 524 1170 1174 2519   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value  
## (Intercept) -0.085332466 0.080765259 -1.0565  
## rec\_resource\_2018c 0.028626452 0.004527008 6.3235  
## county\_crime\_2020c -0.000995685 0.000510273 -1.9513  
## med\_inc\_2016\_thousandc 0.122973417 0.008398214 14.6428  
## rural\_2020c -0.012471023 0.003676899 -3.3917  
## black\_per\_2016c -0.022887812 0.006688441 -3.4220  
## latino\_per\_2016c 0.018879947 0.006235052 3.0280  
## rec\_resource\_2018c:county\_crime\_2020c 0.000012105 0.000016812 0.7200  
## Pr(>|z|)  
## (Intercept) 0.2907174  
## rec\_resource\_2018c 0.0000000002557  
## county\_crime\_2020c 0.0510240  
## med\_inc\_2016\_thousandc < 0.00000000000000022  
## rural\_2020c 0.0006945  
## black\_per\_2016c 0.0006216  
## latino\_per\_2016c 0.0024615  
## rec\_resource\_2018c:county\_crime\_2020c 0.4715012  
##   
## Rho: 0.47514, LR test value: 485.02, p-value: < 0.000000000000000222  
## Asymptotic standard error: 0.021009  
## z-value: 22.616, p-value: < 0.000000000000000222  
## Wald statistic: 511.49, p-value: < 0.000000000000000222  
##   
## Log likelihood: -7475.594 for lag model  
## ML residual variance (sigma squared): 16.62, (sigma: 4.0768)  
## Nagelkerke pseudo-R-squared: 0.47814   
## Number of observations: 2625   
## Number of parameters estimated: 10   
## AIC: 14971, (AIC for lm: 15454)  
## LM test for residual autocorrelation  
## test value: 21.927, p-value: 0.0000028324

int\_model\_find <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_model\_find.rds')  
int\_model\_find

## Impact measures (lag, exact):  
## Direct Indirect  
## rec\_resource\_2018c 0.03013398847 0.0242890173  
## county\_crime\_2020c -0.00104812031 -0.0008448205  
## med\_inc\_2016\_thousandc 0.12944948696 0.1043406792  
## rural\_2020c -0.01312777667 -0.0105814335  
## black\_per\_2016c -0.02409313843 -0.0194198871  
## latino\_per\_2016c 0.01987420971 0.0160192874  
## rec\_resource\_2018c:county\_crime\_2020c 0.00001274251 0.0000102709  
## Total  
## rec\_resource\_2018c 0.05442300573  
## county\_crime\_2020c -0.00189294085  
## med\_inc\_2016\_thousandc 0.23379016616  
## rural\_2020c -0.02370921014  
## black\_per\_2016c -0.04351302557  
## latino\_per\_2016c 0.03589349714  
## rec\_resource\_2018c:county\_crime\_2020c 0.00002301341  
## ========================================================  
## Simulation results (asymptotic variance matrix):  
## Direct:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE  
## rec\_resource\_2018c 0.02719449 0.00269191 0.00155418  
## county\_crime\_2020c -0.00147122 0.00030916 0.00017849  
## med\_inc\_2016\_thousandc 0.13192862 0.01004108 0.00579722  
## rural\_2020c -0.01423213 0.00154884 0.00089422  
## black\_per\_2016c -0.02045811 0.00287457 0.00165963  
## latino\_per\_2016c 0.02164551 0.00184919 0.00106763  
## rec\_resource\_2018c:county\_crime\_2020c 0.00000828 0.00001951 0.00001127  
## Time-series SE  
## rec\_resource\_2018c 0.00155418  
## county\_crime\_2020c 0.00017849  
## med\_inc\_2016\_thousandc 0.00579722  
## rural\_2020c 0.00089422  
## black\_per\_2016c 0.00165963  
## latino\_per\_2016c 0.00106763  
## rec\_resource\_2018c:county\_crime\_2020c 0.00001127  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50%  
## rec\_resource\_2018c 0.025037209 0.025699509 0.026435397  
## county\_crime\_2020c -0.001798095 -0.001588928 -0.001356521  
## med\_inc\_2016\_thousandc 0.124264110 0.126277041 0.128513630  
## rural\_2020c -0.015673321 -0.015023113 -0.014300660  
## black\_per\_2016c -0.023257992 -0.021850742 -0.020287131  
## latino\_per\_2016c 0.020426430 0.020582583 0.020756087  
## rec\_resource\_2018c:county\_crime\_2020c -0.000004296 -0.000002954 -0.000001462  
## 75% 97.5%  
## rec\_resource\_2018c 0.02830992 0.02999699  
## county\_crime\_2020c -0.00129616 -0.00124184  
## med\_inc\_2016\_thousandc 0.13587270 0.14249586  
## rural\_2020c -0.01347541 -0.01273269  
## black\_per\_2016c -0.01897999 -0.01780356  
## latino\_per\_2016c 0.02226372 0.02362059  
## rec\_resource\_2018c:county\_crime\_2020c 0.00001464 0.00002914  
##   
## ========================================================  
## Indirect:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE  
## rec\_resource\_2018c 0.019981692 0.00080045 0.000462140  
## county\_crime\_2020c -0.001073314 0.00012204 0.000070460  
## med\_inc\_2016\_thousandc 0.097534814 0.01234578 0.007127840  
## rural\_2020c -0.010452110 0.00052111 0.000300865  
## black\_per\_2016c -0.015136970 0.00281612 0.001625890  
## latino\_per\_2016c 0.016046261 0.00256695 0.001482027  
## rec\_resource\_2018c:county\_crime\_2020c 0.000006588 0.00001525 0.000008807  
## Time-series SE  
## rec\_resource\_2018c 0.000462140  
## county\_crime\_2020c 0.000070460  
## med\_inc\_2016\_thousandc 0.007127840  
## rural\_2020c 0.000300865  
## black\_per\_2016c 0.001625890  
## latino\_per\_2016c 0.001482027  
## rec\_resource\_2018c:county\_crime\_2020c 0.000008807  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50%  
## rec\_resource\_2018c 0.01920447 0.019592190 0.0200229891  
## county\_crime\_2020c -0.00119984 -0.001124710 -0.0010412284  
## med\_inc\_2016\_thousandc 0.08585828 0.091331831 0.0974135499  
## rural\_2020c -0.01095020 -0.010710844 -0.0104448940  
## black\_per\_2016c -0.01814701 -0.015976674 -0.0135651916  
## latino\_per\_2016c 0.01365812 0.014735131 0.0159318039  
## rec\_resource\_2018c:county\_crime\_2020c -0.00000329 -0.000002191 -0.0000009702  
## 75% 97.5%  
## rec\_resource\_2018c 0.02039184 0.02072381  
## county\_crime\_2020c -0.00100588 -0.00097406  
## med\_inc\_2016\_thousandc 0.10367717 0.10931442  
## rural\_2020c -0.01018977 -0.00996015  
## black\_per\_2016c -0.01351138 -0.01346294  
## latino\_per\_2016c 0.01730016 0.01853168  
## rec\_resource\_2018c:county\_crime\_2020c 0.00001159 0.00002289  
##   
## ========================================================  
## Total:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE  
## rec\_resource\_2018c 0.04717618 0.00304126 0.00175587  
## county\_crime\_2020c -0.00254453 0.00043078 0.00024871  
## med\_inc\_2016\_thousandc 0.22946343 0.02088963 0.01206063  
## rural\_2020c -0.02468424 0.00187460 0.00108230  
## black\_per\_2016c -0.03559508 0.00551996 0.00318695  
## latino\_per\_2016c 0.03769177 0.00433425 0.00250238  
## rec\_resource\_2018c:county\_crime\_2020c 0.00001487 0.00003477 0.00002007  
## Time-series SE  
## rec\_resource\_2018c 0.00175587  
## county\_crime\_2020c 0.00024871  
## med\_inc\_2016\_thousandc 0.01206063  
## rural\_2020c 0.00108230  
## black\_per\_2016c 0.00318695  
## latino\_per\_2016c 0.00250238  
## rec\_resource\_2018c:county\_crime\_2020c 0.00002007  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50%  
## rec\_resource\_2018c 0.044278566 0.045660552 0.047196092  
## county\_crime\_2020c -0.002997938 -0.002713638 -0.002397749  
## med\_inc\_2016\_thousandc 0.214148255 0.217608872 0.221454001  
## rural\_2020c -0.026144810 -0.025733957 -0.025277454  
## black\_per\_2016c -0.041399619 -0.037773602 -0.033744693  
## latino\_per\_2016c 0.034084555 0.035317714 0.036687891  
## rec\_resource\_2018c:county\_crime\_2020c -0.000007585 -0.000005144 -0.000002433  
## 75% 97.5%  
## rec\_resource\_2018c 0.04870176 0.05005686  
## county\_crime\_2020c -0.00230204 -0.00221590  
## med\_inc\_2016\_thousandc 0.23731327 0.25158662  
## rural\_2020c -0.02393113 -0.02271944  
## black\_per\_2016c -0.03249136 -0.03136337  
## latino\_per\_2016c 0.03956388 0.04215227  
## rec\_resource\_2018c:county\_crime\_2020c 0.00002623 0.00005203  
##   
## ========================================================  
## Simulated standard errors  
## Direct Indirect Total  
## rec\_resource\_2018c 0.00269191481 0.00080045071 0.0030412581  
## county\_crime\_2020c 0.00030915806 0.00012203984 0.0004307830  
## med\_inc\_2016\_thousandc 0.01004107661 0.01234578133 0.0208896304  
## rural\_2020c 0.00154883785 0.00052111374 0.0018745959  
## black\_per\_2016c 0.00287457051 0.00281612440 0.0055199615  
## latino\_per\_2016c 0.00184919347 0.00256694587 0.0043342541  
## rec\_resource\_2018c:county\_crime\_2020c 0.00001951337 0.00001525439 0.0000347677  
##   
## Simulated z-values:  
## Direct Indirect Total  
## rec\_resource\_2018c 10.1022835 24.9630506 15.5120600  
## county\_crime\_2020c -4.7587970 -8.7947858 -5.9067668  
## med\_inc\_2016\_thousandc 13.1388915 7.9002545 10.9845615  
## rural\_2020c -9.1889091 -20.0572531 -13.1677663  
## black\_per\_2016c -7.1169274 -5.3751071 -6.4484290  
## latino\_per\_2016c 11.7053765 6.2511099 8.6962522  
## rec\_resource\_2018c:county\_crime\_2020c 0.4243011 0.4318767 0.4276256  
##   
## Simulated p-values:  
## Direct   
## rec\_resource\_2018c < 0.000000000000000222  
## county\_crime\_2020c 0.0000019475020343   
## med\_inc\_2016\_thousandc < 0.000000000000000222  
## rural\_2020c < 0.000000000000000222  
## black\_per\_2016c 0.0000000000011036   
## latino\_per\_2016c < 0.000000000000000222  
## rec\_resource\_2018c:county\_crime\_2020c 0.67135   
## Indirect   
## rec\_resource\_2018c < 0.000000000000000222  
## county\_crime\_2020c < 0.000000000000000222  
## med\_inc\_2016\_thousandc 0.0000000000000028866   
## rural\_2020c < 0.000000000000000222  
## black\_per\_2016c 0.0000000765371801492   
## latino\_per\_2016c 0.0000000004075459969   
## rec\_resource\_2018c:county\_crime\_2020c 0.66583   
## Total   
## rec\_resource\_2018c < 0.000000000000000222  
## county\_crime\_2020c 0.00000000348887   
## med\_inc\_2016\_thousandc < 0.000000000000000222  
## rural\_2020c < 0.000000000000000222  
## black\_per\_2016c 0.00000000011302   
## latino\_per\_2016c < 0.000000000000000222  
## rec\_resource\_2018c:county\_crime\_2020c 0.66892

AIC(int\_sar\_queen)

## [1] 14971.19

## Income Interaction Model

int\_income\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_income\_modelc.rds')  
summary(int\_income\_queen)

##   
## Call:  
## lm(formula = income\_int\_variablesc, data = acs\_polyc)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.1460 -2.9713 -0.0091 3.0763 20.7205   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) -0.2769006 0.0952525 -2.907  
## rec\_resource\_2018c 0.0508513 0.0051332 9.906  
## county\_crime\_2020c -0.0009035 0.0005751 -1.571  
## med\_inc\_2016\_thousandc 0.1748200 0.0094549 18.490  
## rural\_2020c -0.0093724 0.0041205 -2.275  
## black\_per\_2016c -0.0606674 0.0077425 -7.836  
## latino\_per\_2016c 0.0449463 0.0069474 6.470  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0011414 0.0003265 3.496  
## Pr(>|t|)   
## (Intercept) 0.003680 \*\*   
## rec\_resource\_2018c < 0.0000000000000002 \*\*\*  
## county\_crime\_2020c 0.116310   
## med\_inc\_2016\_thousandc < 0.0000000000000002 \*\*\*  
## rural\_2020c 0.023011 \*   
## black\_per\_2016c 0.00000000000000673 \*\*\*  
## latino\_per\_2016c 0.00000000011694814 \*\*\*  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.000481 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.577 on 2617 degrees of freedom  
## Multiple R-squared: 0.3746, Adjusted R-squared: 0.3729   
## F-statistic: 223.9 on 7 and 2617 DF, p-value: < 0.00000000000000022

AIC(int\_income\_queen)

## [1] 15444.32

int\_income\_moran\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_income\_moran\_queenc.rds')  
int\_income\_moran\_queen

##   
## Global Moran I for regression residuals  
##   
## data:   
## model: lm(formula = income\_int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## Moran I statistic standard deviate = 23.513, p-value <  
## 0.00000000000000022  
## alternative hypothesis: greater  
## sample estimates:  
## Observed Moran I Expectation Variance   
## 0.2793984480 -0.0015711677 0.0001427879

int\_income\_spa\_tests\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_income\_spa\_tests\_queenc.rds')  
int\_income\_spa\_tests\_queen

##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = income\_int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## LMerr = 547.5, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = income\_int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## LMlag = 620.94, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = income\_int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## RLMerr = 15.862, df = 1, p-value = 0.00006814  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = income\_int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## RLMlag = 89.297, df = 1, p-value < 0.00000000000000022  
##   
##   
## Lagrange multiplier diagnostics for spatial dependence  
##   
## data:   
## model: lm(formula = income\_int\_variablesc, data = acs\_polyc)  
## weights: poly\_listw\_queenc  
##   
## SARMA = 636.8, df = 2, p-value < 0.00000000000000022

int\_income\_sar\_queen <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_income\_sar\_queenc.rds')  
summary(int\_income\_sar\_queen, zstats = TRUE, Nagelkerke = TRUE)

##   
## Call:  
## lagsarlm(formula = income\_int\_variablesc, data = acs\_polyc, listw = poly\_listw\_queenc,   
## zero.policy = TRUE)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.554445 -2.578803 0.058043 2.637410 17.413217   
##   
## Type: lag   
## Regions with no neighbours included:  
## 68 69 70 71 72 73 521 522 524 1170 1174 2519   
## Coefficients: (asymptotic standard errors)   
## Estimate Std. Error z value  
## (Intercept) -0.15643241 0.08485716 -1.8435  
## rec\_resource\_2018c 0.03092620 0.00460163 6.7207  
## county\_crime\_2020c -0.00083797 0.00051195 -1.6368  
## med\_inc\_2016\_thousandc 0.11347994 0.00896123 12.6634  
## rural\_2020c -0.01192491 0.00366756 -3.2515  
## black\_per\_2016c -0.02866521 0.00699726 -4.0966  
## latino\_per\_2016c 0.01802364 0.00623078 2.8927  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.00081097 0.00029102 2.7866  
## Pr(>|z|)  
## (Intercept) 0.065259  
## rec\_resource\_2018c 0.00000000001809  
## county\_crime\_2020c 0.101666  
## med\_inc\_2016\_thousandc < 0.00000000000000022  
## rural\_2020c 0.001148  
## black\_per\_2016c 0.00004192040151  
## latino\_per\_2016c 0.003820  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.005326  
##   
## Rho: 0.47348, LR test value: 482.38, p-value: < 0.000000000000000222  
## Asymptotic standard error: 0.020945  
## z-value: 22.606, p-value: < 0.000000000000000222  
## Wald statistic: 511.02, p-value: < 0.000000000000000222  
##   
## Log likelihood: -7471.97 for lag model  
## ML residual variance (sigma squared): 16.58, (sigma: 4.0719)  
## Nagelkerke pseudo-R-squared: 0.47958   
## Number of observations: 2625   
## Number of parameters estimated: 10   
## AIC: 14964, (AIC for lm: 15444)  
## LM test for residual autocorrelation  
## test value: 21.427, p-value: 0.0000036746

int\_income\_model\_find <- read\_rds('C:/Users/cpppe/Desktop/github\_projects/dissertation/rds\_files/centered\_folder/int\_income\_model\_find.rds')  
int\_income\_model\_find

## Impact measures (lag, exact):  
## Direct Indirect  
## rec\_resource\_2018c 0.0325410288 0.0260684472  
## county\_crime\_2020c -0.0008817269 -0.0007063468  
## med\_inc\_2016\_thousandc 0.1194053493 0.0956549980  
## rural\_2020c -0.0125475729 -0.0100517947  
## black\_per\_2016c -0.0301619807 -0.0241626042  
## latino\_per\_2016c 0.0189647571 0.0151925673  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0008533178 0.0006835884  
## Total  
## rec\_resource\_2018c 0.058609476  
## county\_crime\_2020c -0.001588074  
## med\_inc\_2016\_thousandc 0.215060347  
## rural\_2020c -0.022599368  
## black\_per\_2016c -0.054324585  
## latino\_per\_2016c 0.034157324  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.001536906  
## ========================================================  
## Simulation results (asymptotic variance matrix):  
## Direct:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE  
## rec\_resource\_2018c 0.0320981 0.00786524 0.00454100  
## county\_crime\_2020c -0.0009290 0.00048916 0.00028242  
## med\_inc\_2016\_thousandc 0.1220438 0.01135205 0.00655411  
## rural\_2020c -0.0083152 0.00293671 0.00169551  
## black\_per\_2016c -0.0297945 0.00455125 0.00262766  
## latino\_per\_2016c 0.0243119 0.00860907 0.00497045  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0009069 0.00006797 0.00003924  
## Time-series SE  
## rec\_resource\_2018c 0.00454100  
## county\_crime\_2020c 0.00028242  
## med\_inc\_2016\_thousandc 0.00655411  
## rural\_2020c 0.00169551  
## black\_per\_2016c 0.00262766  
## latino\_per\_2016c 0.00497045  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.00003924  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50%  
## rec\_resource\_2018c 0.0237155 0.0293458 0.0356018  
## county\_crime\_2020c -0.0013054 -0.0012037 -0.0010908  
## med\_inc\_2016\_thousandc 0.1145451 0.1155196 0.1166023  
## rural\_2020c -0.0112903 -0.0096391 -0.0078043  
## black\_per\_2016c -0.0339898 -0.0321384 -0.0300813  
## latino\_per\_2016c 0.0169412 0.0196568 0.0226741  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0008401 0.0008745 0.0009126  
## 75% 97.5%  
## rec\_resource\_2018c 0.0366022 0.0375026  
## county\_crime\_2020c -0.0007351 -0.0004150  
## med\_inc\_2016\_thousandc 0.1258473 0.1341678  
## rural\_2020c -0.0067359 -0.0057742  
## black\_per\_2016c -0.0275939 -0.0253553  
## latino\_per\_2016c 0.0281482 0.0330749  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0009423 0.0009689  
##   
## ========================================================  
## Indirect:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE  
## rec\_resource\_2018c 0.0246987 0.00670374 0.00387041  
## county\_crime\_2020c -0.0006991 0.00034663 0.00020013  
## med\_inc\_2016\_thousandc 0.0938597 0.01371420 0.00791790  
## rural\_2020c -0.0063256 0.00207894 0.00120028  
## black\_per\_2016c -0.0227487 0.00272716 0.00157453  
## latino\_per\_2016c 0.0188336 0.00762185 0.00440048  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0006936 0.00001885 0.00001088  
## Time-series SE  
## rec\_resource\_2018c 0.00387041  
## county\_crime\_2020c 0.00020013  
## med\_inc\_2016\_thousandc 0.00791790  
## rural\_2020c 0.00120028  
## black\_per\_2016c 0.00157453  
## latino\_per\_2016c 0.00440048  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.00001088  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50%  
## rec\_resource\_2018c 0.0178164 0.0217815 0.0261871  
## county\_crime\_2020c -0.0009611 -0.0008947 -0.0008208  
## med\_inc\_2016\_thousandc 0.0857851 0.0859425 0.0861174  
## rural\_2020c -0.0084897 -0.0071875 -0.0057405  
## black\_per\_2016c -0.0255529 -0.0239298 -0.0221265  
## latino\_per\_2016c 0.0127295 0.0145999 0.0166780  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0006795 0.0006829 0.0006868  
## 75% 97.5%  
## rec\_resource\_2018c 0.0283601 0.0303159  
## county\_crime\_2020c -0.0005645 -0.0003337  
## med\_inc\_2016\_thousandc 0.0979058 0.1085154  
## rural\_2020c -0.0051712 -0.0046588  
## black\_per\_2016c -0.0212565 -0.0204734  
## latino\_per\_2016c 0.0219896 0.0267700  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.0007008 0.0007135  
##   
## ========================================================  
## Total:  
##   
## Iterations = 1:3  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 3   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE  
## rec\_resource\_2018c 0.056797 0.0144948 0.00836857  
## county\_crime\_2020c -0.001628 0.0008358 0.00048253  
## med\_inc\_2016\_thousandc 0.215904 0.0250298 0.01445099  
## rural\_2020c -0.014641 0.0050102 0.00289261  
## black\_per\_2016c -0.052543 0.0072237 0.00417060  
## latino\_per\_2016c 0.043146 0.0162174 0.00936310  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.001601 0.0000857 0.00004948  
## Time-series SE  
## rec\_resource\_2018c 0.00836857  
## county\_crime\_2020c 0.00048253  
## med\_inc\_2016\_thousandc 0.01445099  
## rural\_2020c 0.00289261  
## black\_per\_2016c 0.00417060  
## latino\_per\_2016c 0.00936310  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.00004948  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50%  
## rec\_resource\_2018c 0.041532 0.051127 0.061789  
## county\_crime\_2020c -0.002267 -0.002098 -0.001912  
## med\_inc\_2016\_thousandc 0.200645 0.201462 0.202370  
## rural\_2020c -0.019780 -0.016827 -0.013545  
## black\_per\_2016c -0.059543 -0.056068 -0.052208  
## latino\_per\_2016c 0.029671 0.034257 0.039352  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.001520 0.001557 0.001599  
## 75% 97.5%  
## rec\_resource\_2018c 0.064962 0.0678185  
## county\_crime\_2020c -0.001300 -0.0007487  
## med\_inc\_2016\_thousandc 0.223578 0.2426657  
## rural\_2020c -0.011907 -0.0104330  
## black\_per\_2016c -0.048850 -0.0458287  
## latino\_per\_2016c 0.050138 0.0598449  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.001643 0.0016824  
##   
## ========================================================  
## Simulated standard errors  
## Direct Indirect  
## rec\_resource\_2018c 0.00786524455 0.00670374247  
## county\_crime\_2020c 0.00048915864 0.00034663406  
## med\_inc\_2016\_thousandc 0.01135204816 0.01371419874  
## rural\_2020c 0.00293670836 0.00207893741  
## black\_per\_2016c 0.00455124690 0.00272715905  
## latino\_per\_2016c 0.00860907191 0.00762185237  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.00006796991 0.00001884554  
## Total  
## rec\_resource\_2018c 0.01449479689  
## county\_crime\_2020c 0.00083575838  
## med\_inc\_2016\_thousandc 0.02502984900  
## rural\_2020c 0.00501015587  
## black\_per\_2016c 0.00722369853  
## latino\_per\_2016c 0.01621735875  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 0.00008569841  
##   
## Simulated z-values:  
## Direct Indirect Total  
## rec\_resource\_2018c 4.081004 3.684317 3.918427  
## county\_crime\_2020c -1.899095 -2.016941 -1.948050  
## med\_inc\_2016\_thousandc 10.750820 6.843983 8.625844  
## rural\_2020c -2.831455 -3.042705 -2.922214  
## black\_per\_2016c -6.546437 -8.341537 -7.273718  
## latino\_per\_2016c 2.823991 2.471006 2.660457  
## rec\_resource\_2018c:med\_inc\_2016\_thousandc 13.343334 36.803520 18.676282  
##   
## Simulated p-values:  
## Direct   
## rec\_resource\_2018c 0.000044841536117   
## county\_crime\_2020c 0.0575520   
## med\_inc\_2016\_thousandc < 0.000000000000000222  
## rural\_2020c 0.0046337   
## black\_per\_2016c 0.000000000058926   
## latino\_per\_2016c 0.0047430   
## rec\_resource\_2018c:med\_inc\_2016\_thousandc < 0.000000000000000222  
## Indirect   
## rec\_resource\_2018c 0.00022932   
## county\_crime\_2020c 0.04370166   
## med\_inc\_2016\_thousandc 0.0000000000077021   
## rural\_2020c 0.00234462   
## black\_per\_2016c < 0.000000000000000222  
## latino\_per\_2016c 0.01347337   
## rec\_resource\_2018c:med\_inc\_2016\_thousandc < 0.000000000000000222  
## Total   
## rec\_resource\_2018c 0.00008912856009236   
## county\_crime\_2020c 0.0514090   
## med\_inc\_2016\_thousandc < 0.000000000000000222  
## rural\_2020c 0.0034755   
## black\_per\_2016c 0.00000000000034972   
## latino\_per\_2016c 0.0078035   
## rec\_resource\_2018c:med\_inc\_2016\_thousandc < 0.000000000000000222

AIC(int\_income\_sar\_queen)

## [1] 14963.94

## Comparing model fit between the main effects model and the interaction models

LR.sarlm(int\_sar\_queen, sar\_queen)

##   
## Likelihood ratio for spatial linear models  
##   
## data:   
## Likelihood ratio = 0.52105, df = 1, p-value = 0.4704  
## sample estimates:  
## Log likelihood of int\_sar\_queen Log likelihood of sar\_queen   
## -7475.594 -7475.855

LR.sarlm(int\_income\_sar\_queen, sar\_queen)

##   
## Likelihood ratio for spatial linear models  
##   
## data:   
## Likelihood ratio = 7.7687, df = 1, p-value = 0.005316  
## sample estimates:  
## Log likelihood of int\_income\_sar\_queen Log likelihood of sar\_queen   
## -7471.970 -7475.855

## Analytic Plan

Analytic Plan

All analyses for the present study were conducted in R 4.0.3 (R Core Team, 2020), using the following packages: tidyverse (Wickham et al., 2019), psych (Revelle, 2020), inspectdf (Rushworth, 2020), sf (Bivand et al., 2018), spdep (Rogers et al., 2013), and tidycensus (Walker & Herman, 2020). With limited missing values in the dataset, missing data were addressed by imputing the median. Model assumptions for ordinary least squares (OLS) regressions were tested, including univariate normality, linearity, homoscedasticity, residuals of the models, and multicollinearity. Correlations of .7 or above were used as a threshold indicating multicollinearity. Variables with values over this threshold were removed from the analyses. Due to the spatial nature of the data, Moran’s I tests were conducted to examine spatial dependence in all the variables of interest. These tests examine if variable values are clustered together with nearby counties and show spatial dependence. Moran’s I values range from -1 to 1, with higher values indicating that spatial clusters are similar and a zero-value indicating a random pattern in spatial clusters. Since counties are irregular shapes in the United States, queen contiguity weights were used. These weights are more accurate when examining irregular polygons. All continuous variables were centered for moderation analyses.

Study Aim 1 Analytic Plan

An OLS model was tested to examine the association between recreational resources and LTPA while adjusting for violent crime, median household income, percentage of a given county that is rural, and estimates of Black/African American, and Latina/o populations in each county. The residuals were used to examine if the Moran’s test for the model suggested spatial dependence. Lagrange Multiplier (LM) diagnostic tests were conducted on this model to suggest whether to conduct a spatial error or spatial lag model. Based on the findings of the LM test, a spatial regression model was conducted that mirrored that of the OLS model while accounting for the spatial dependence in the OLS model.

Study Aim 2 Analytic Plan

A second OLS model was tested to investigate violent crime rates as a potential moderator of the association between access to recreational resources and LTPA. The residuals of the moderation analysis were used to assess spatial dependence. LM tests were also conducted for the interaction model. The spatial regression model included the interaction between access to recreational resources and violent crime rates while adjusting for median household income, percentage of a given county that is rural, and percentages of Black/African American and Latina/o populations within a given county.

Exploratory Analyses

A post-hoc OLS regression was conducted investigating median household income as a moderator of the association between access to recreational resources and LTPA. Spatial dependence and selection of the correct spatial regression were assessed by Moran’s I and LM tests. From the LM tests, a spatial lag model was conducted that included the interaction between access to recreational resources and median household income while adjusting for violent crime rates, percentage of a given county that is rural, and percentages of Black/African American and Latina/o populations within a given county.