**Supporting Information 2: Reproducible R code**

Please note: You should locate the Supporting\_Information\_Data.RData file (provided by the authors upon request) in the same directory of this code.

* Datasets are in courtesy of GeoDa Center, NHGIS, and Denver Open Data Catalog.
* Datasets provided by GeoDa Center are kindly permitted to be re-distributed and can also be downloaded from GeoDa Center webpage.
* The redistribution of “Denver Crime dataset” shapefile (saved as an object denv.crime in the attached RData) is permitted under open data license of Denver Open Data Catalog.
* A 1990 Census data table from the NHGIS used in this code is legally protected not to be re-distributed for any other purpose excluding a submission to an academic journal. Please contact NHGIS to request further usage of the data.

# If you do not have pacman package in your machine, please install it first.

# p\_load in pacman package will automatically find whether required packages are available in your machine and if not, it will install all required packages from the internet.

# Please make sure your machine is connected to the internet.

# The generalized function (e.g., sf\_analysis) to analyze each dataset is contained in the attached RData file.

install.packages('pacman')

library(pacman)

p\_load(tidyverse, spdep, sf, spatialreg, sp, stringr)

# Note: please make sure to set your working directory as the directory with this script and the RData file before running the next line

load('Supporting\_Information\_Data\_1.RData')

load('Supporting\_Information\_Data\_2.RData')

## 1. AirBnB ####

airbnb <- airbnb %>% st\_transform(crs = 3857) %>%

mutate(crimes = num\_crimes / population \* 1000,

theft = num\_theft / population \* 1000) %>%

filter(!is.na(accept\_r) & !is.na(rev\_rating)) %>%

mutate\_if(is.numeric, funs(scale))

airbnb.yvec <- c('response\_r', 'accept\_r', 'rev\_rating', 'price\_pp')

airbnb.xvec <- c('poverty', 'crowded', 'dependency', 'without\_hs', 'unemployed', 'income\_pc', 'harship\_in', 'crimes', 'theft')

airbnb.result2 <- sf\_analysis(airbnb, yvec = airbnb.yvec, xvec = airbnb.xvec, mode = 'area')

airbnb.xm <- X.sac(airbnb, airbnb.xvec, 'area')

## 2. Baltimore Housing ####

balt <- balt %>% st\_transform(crs = 3857) %>%

mutate(PRICE\_LIV = PRICE / SQFT,

PRICE\_LOT = PRICE / LOTSZ) %>%

mutate\_if(is.numeric, list(~scale(.)))

balt.yvec <- c('PRICE\_LIV', 'PRICE\_LOT')

balt.xvec <- colnames(balt)[3:13]

balt.result2 <- sf\_analysis(balt, yvec = balt.yvec, xvec = balt.xvec, mode = 'point')

balt.xm <- X.sac(balt, balt.xvec, 'point')

## 3. Boston Housing ####

bost <- bost %>% st\_transform(crs = 3857)# %>%

bost.sp <- bost %>% mutate\_at(.vars = vars(9:23), .funs = list(~scale(.)))

bost.yvec <- c('CMEDV')

bost.xvec <- bost.sp %>% colnames %>% .[11:23]

bost.result2 <- sf\_analysis(bost.sp, bost.yvec, bost.xvec, 'point')

bost.xm <- X.sac(bost, bost.xvec, 'point')

## 4. Industry mixes ####

## Charleston Industry Mix

## Hickory Industry Mix

## Orlando Industry Mix

## Sacramento Industry Mix

## Seattle Industry Mix

## Lansing Industry Mix

zipstat.s <- zipstat %>%

mutate(ZIP = sprintf('%05s', ZIPA),

rateuniv = (E33006 + E33007) / (E4H001 + E4H002) \* 100,

hin = E4U001,

youth = (E1L013 + E1L014 + E1L015 + E1L016 + E1L017) / (E4H001 + E4H002) \* 100) %>%

dplyr::select(ZIP, rateuniv, hin, youth)

imix.ll <- imix.l %>%

lapply(function(x) x %>%

st\_transform(crs = 3857) %>%

left\_join(zipstat.s, by = 'ZIP') %>%

mutate\_at(.vars = vars(rateuniv:youth),

.funs = list(~ifelse(is.na(.), median(., na.rm = TRUE), .))))

for (i in c(1:4, 6)){ imix.ll[[i]] <- imix.ll[[i]] %>% dplyr::select(-CBSA\_CODE)}

imix.ll <- imix.ll %>%

lapply(function(x) x %>% dplyr::select(-52:-53))

colnames(imix.ll[[5]])[c(1:50)] <- colnames(imix.ll[[4]])[1:50]

imix.lp <- imix.ll %>% lapply(function(x) x %>% mutate\_if(is.numeric, list(~scale(.))))

imix.lp[[1]] %>% colnames %>% .[c(14:49)[-grep('.US.\*', .[14:49])]] -> imix.yvec

imix.lp[[1]] %>% colnames %>% .[c(53:55, 57:59)] -> imix.xvec

imix.result.l <- imix.lp %>%

lapply(function(x) sf\_analysis(x, imix.yvec, imix.xvec, 'area'))

imix.result2 <- imix.result.l %>% do.call(rbind, .)

## 5. Chicago Health and Socio-Economic ####

chhs.std <- chhs %>% mutate\_at(.vars = vars(62:65, 70:86), .funs = list(~scale(.)))

chhs.yvec <- colnames(chhs)[70:86]

chhs.xvec <- colnames(chhs)[62:65]

chhs.results2 <- sf\_analysis(chhs.std, yvec = chhs.yvec, xvec = chhs.xvec, mode = 'area')

chhs.xm <- X.sac(chhs, colnames(chhs)[62:65], mode = 'area')

## 6. Chicago Health Indicators ####

hein <- hein %>% st\_transform(crs = 3857)

hein.sp <- hein %>% mutate\_at(.vars = vars(5:31), .funs = funs(scale))

hein.yvec <- colnames(hein)[5:25]

hein.xvec <- colnames(hein)[26:31]

hein.result2 <- sf\_analysis(hein.sp, hein.yvec, hein.xvec, 'area')

hein.xm <- X.sac(hein.sp, hein.xvec, 'area')

## 7. Cincinnati Crime ####

cinc <- cinc %>% st\_transform(crs = 3857) %>%

mutate(BURG = BURGLARY / POPULATION \* 1000,

ASSA = ASSAULT / POPULATION \* 1000,

THEF = THEFT / POPULATION \* 1000,

HH\_NON = HH\_NONFAMI / HOUSEHOLDS \* 100,

HU\_VACANT = HU\_VACANT / HSNG\_UNITS \* 100,

GROUP\_QUAR = GROUP\_QUAR / POPULATION \* 100,

GQ\_NONINST = GQ\_NONINST / POPULATION \* 100,

NONWHITE = (BLACK + AMINDIAN + ASIAN + HAWAIIAN + OTHER\_RACE)/ POPULATION \* 100,

JUVENILE = (AGE\_15\_19 + AGE\_20\_24) / POPULATION \* 100) %>%

filter(!is.na(HH\_NON)) %>%

mutate\_if(is.numeric, funs(scale))

cinc.yvec <- c('BURG', 'ASSA', 'THEF')

cinc.xvec <- c('HH\_NON', 'HU\_VACANT', 'GROUP\_QUAR', 'GQ\_NONINST', 'NONWHITE', 'JUVENILE')

cinc.result2 <- sf\_analysis(cinc, cinc.yvec, cinc.xvec, 'area')

cinc.xm <- X.sac(cinc, cinc.xvec, 'area')

## 8. Columbus Crime ####

colu <- colu %>% st\_transform(3857)

colu.sp <- colu %>%

mutate\_at(.vars = vars(HOVAL:DISCBD), .funs = funs(scale))

colu.yvec <- c('CRIME')

colu.xvec <- c('INC', 'HOVAL', 'OPEN', 'PLUMB', 'DISCBD', 'NSA', 'NSB', 'EW', 'CP')

colu.result2 <- sf\_analysis(colu.sp, colu.yvec, colu.xvec, 'area')

colu.xm <- X.sac(colu, colu.xvec, 'area')

## 9. Denver Crime ####

denv <- denv %>%

mutate(NEIGHBOR = stringr::str\_replace(tolower(NBRHD\_NAME), ' ', '-'))

denv.crime.s <- st\_set\_geometry(denv.crime, NULL) %>%

mutate(year = substr(REPORTED\_D, 1, 4)) %>%

filter(year != '2019') %>%

mutate(OFFENSE\_YEAR = paste(OFFENSE\_TY, year, sep = '')) %>%

group\_by(NEIGHBORHO, OFFENSE\_YEAR) %>%

summarize(N = n()) %>%

ungroup %>%

spread(key = OFFENSE\_YEAR, value = N)

denv.a <- denv %>%

left\_join(denv.crime.s, by = c('NEIGHBOR' = 'NEIGHBORHO')) %>%

mutate\_at(.vars = vars(130:(ncol(.)-1)), .funs = funs(ifelse(is.na(.), 0, .))) %>%

.[,sapply(., FUN = function(x) length(unique(x)) >= 0.2 \* length(x))] %>%

mutate\_at(.vars = vars(123:321), .funs = funs(./POPULATION \* 100000)) %>%

mutate(pnonwhite = 100 - PCT\_WHITE,

pvacant = 100 \* VACANTUNIT / HOUSINGUNI,

prent = 100 \* HU\_RENTED / HOUSINGUNI,

psinglef = 100 \* (MALE\_HHL\_1 + FEMALE\_H\_1) / FAMILY\_HHL,

pjuvenile = 100 \* (AGE\_18\_AND + AGE\_20 + AGE\_21 + AGE\_22\_TO\_) / POPULATION

) %>%

mutate\_at(.vars = vars(pnonwhite, pvacant, prent, psinglef, pjuvenile, 123:321),

.funs = list(~scale(.)))

denv.yvec <- stringr::str\_replace\_all(colnames(denv.a)[123:321], '-', '.')

denv.xvec <- c('pnonwhite', 'pvacant', 'prent', 'psinglef', 'pjuvenile')

denv.result2 <- sf\_analysis(denv.a, denv.yvec, denv.xvec, 'area')

denv.xm <- X.sac(denv.a, denv.xvec, 'area')

## 10. Natregimes (National Crime) ####

nat <- nat %>% st\_transform(3857)

nats <- nat %>% split(., .$STATE\_NAME)

# Excluding states with less than 30 counties

for (i in 1:length(nats)){

if (nrow(nats[[i]]) < 30) { nats[[i]] <- NA}

}

nats[which((nats %>% lapply(length) %>% do.call(c, .)) <= 1)] <- NULL

nats.soc <- nats %>% lapply(function(x){

x %>%

mutate\_if(is.numeric, list(~scale(.)))

return(x)})

nats.yvec <- c('HR60', 'HR70', 'HR80', 'HR90')

nats.xvec <- expand.grid(c('RD', 'PS', 'UE', 'DV', 'MA'), seq(60, 90, 10)) %>%

mutate(fnm = str\_c(Var1, Var2, sep = '')) %>%

.$fnm

nats.res.l60 <- nats.soc %>%

lapply(function(x) sf\_analysis(x, nats.yvec[1], nats.xvec[1:5], 'area')) %>%

do.call(rbind, .)

nats.res.l70 <- nats.soc %>%

lapply(function(x) sf\_analysis(x, nats.yvec[2], nats.xvec[6:10], 'area')) %>%

do.call(rbind, .)

nats.res.l80 <- nats.soc %>%

lapply(function(x) sf\_analysis(x, nats.yvec[3], nats.xvec[11:15], 'area')) %>%

do.call(rbind, .)

nats.res.l90 <- nats.soc %>%

lapply(function(x) sf\_analysis(x, nats.yvec[4], nats.xvec[16:20], 'area')) %>%

do.call(rbind, .)

nats.f60 <- HR60 ~ RD60 + PS60 + UE60 + DV60 + MA60

nats.f70 <- HR70 ~ RD70 + PS70 + UE70 + DV70 + MA70

nats.f80 <- HR80 ~ RD80 + PS80 + UE80 + DV80 + MA80

nats.f90 <- HR90 ~ RD90 + PS90 + UE90 + DV90 + MA90

nats.xm60 <- X.sac(nat, xvars = as.character(nats.f60)[3] %>% strsplit(., split = ' + ', fixed = T) %>% .[[1]], mode = 'area')

nats.xm70 <- X.sac(nat, xvars = as.character(nats.f70)[3] %>% strsplit(., split = ' + ', fixed = T) %>% .[[1]], mode = 'area')

nats.xm80 <- X.sac(nat, xvars = as.character(nats.f80)[3] %>% strsplit(., split = ' + ', fixed = T) %>% .[[1]], mode = 'area')

nats.xm90 <- X.sac(nat, xvars = as.character(nats.f90)[3] %>% strsplit(., split = ' + ', fixed = T) %>% .[[1]], mode = 'area')

nats.xm2 <- data.frame(matrix(c(nats.xm60, nats.xm70, nats.xm80, nats.xm90), nrow = 4, byrow = T)) %>%

.[rep(seq\_len(nrow(.)), 35),]

nats.xm22 <- nats.xm2 %>%

mutate\_each(funs = funs(do.call(c, .)))

colnames(nats.xm22) <- c('RD', 'PS', 'UE', 'DV', 'MA')

nats.result2 <- bind\_rows(

nats.res.l60,

nats.res.l70,

nats.res.l80,

nats.res.l90

)

## 11. US Elections ####

elec <- elec %>% st\_transform(crs = 3857) %>%

mutate\_if(is.numeric, funs(replace\_na(., 0)))

elec.sp <- elec %>%

mutate(ELDER14 = AGE775214,

WHITE14 = RHI825214,

FOREIGN14 = POP645213,

FEMALE14 = SEX255214,

LATINO14 = RHI725214,

VETERAN14 = VET605213 / PST045214 \* 100,

RETAIL07 = RTN131207,

ESTAB13 = NES010213,

INCOME13 = INC910213,

POVERTY13 = PVY020213,

FIRMNW07 = (SBO001207 - (SBO315207 + SBO115207 + SBO215207 + SBO515207 + SBO415207)) / SBO001207 \* 100) %>%

filter(PST045214 > 0)

elec.sp1 <- elec.sp %>% mutate\_at(.vars = vars(85:95), .funs = funs(scale))

elec.yvec1 <- c('diff\_2016')

elec.yvec2 <- c('diff\_2012')

elec.xvec <- c('ELDER14', 'WHITE14', 'FEMALE14', 'LATINO14',

'RETAIL07', 'ESTAB13', 'INCOME13', 'POVERTY13')

# Warning: it may take some time (it takes longer than an hour in a decent laptop; the duration also depends on the performance of your machine)

elec16.result2 <- sf\_analysis(elec.sp, elec.yvec1, elec.xvec, 'area')

elec12.result2 <- sf\_analysis(elec.sp, elec.yvec2, elec.xvec, 'area')

elec.xm <- X.sac(elec.sp, elec.xvec, 'area')

## 12. Phoenix ACS ####

phx <- phx %>% st\_transform(crs = 3857)

phx.yvec <- c('inc')

phx.xvec <- c('renter\_rt', 'vac\_hsu\_rt', 'white\_rt', 'black\_rt', 'hisp\_rt', 'fem\_nh\_rt')

phx.result2 <- sf\_analysis(phx, phx.yvec, phx.xvec, 'area')

phx.xm <- X.sac(phx, phx.xvec, 'area')

## 13. MSA Employment ####

## MSA Employment (Charleston, South Carolina)

## MSA Employment (Hickory, North Carolina)

## MSA Employment (Lansing, Michigan)

## MSA Employment (Milwaukee, Wisconsin)

## MSA Employment (Orlando, Florida)

## MSA Employment (Sacramento, California)

## MSA Employment (Savannah, Georgia)

## MSA Employment (Seattle, Washington)

## MSA Employment (Tampa, Florida)

msas.s <- msas

msas.st <- msas.s %>%

lapply(function(x) {x <- x %>%

mutate(emp\_away\_p = EMP\_AWAY / EMPL16 \* 100,

emp\_home\_p = EMP\_HOME / EMPL16 \* 100,

emp\_29\_p = EMP\_29 / EMPL16 \* 100,

emp\_30\_p = EMP\_30 / EMPL16 \* 100,

emp\_civ\_p = (EMP16\_2) / EMPL16 \* 100,

occ\_man\_p = OCC\_MAN / EMPL16 \* 100,

occ\_off1\_p = OCC\_OFF1 / EMPL16 \* 100,

occ\_info\_p = OCC\_INFO / EMPL16 \* 100,

pov\_tot1k = POV\_TOT / 1000,

hh\_inc1k = HH\_INC / 1000,

hsg\_val1k = HSG\_VAL / 1000) %>%

mutate\_if(is.numeric, .funs = list(~replace\_na2(.))) %>%

mutate\_if(is.numeric, .funs = list(~replace\_inf(.)))

return(x)})

msas.sts <- msas.st %>%

lapply(function(x){

x2s <- grep('emp\_away\_p', colnames(x))

x2e <- grep('hsg\_val1k', colnames(x))

x2 <- x %>%

mutate\_at(.vars = vars(x2s:x2e), .funs = list(~scale(.)))

return(x2)

})

msas.yvec <- c('emp\_away\_p', 'emp\_home\_p', 'emp\_29\_p', 'emp\_30\_p', 'emp\_civ\_p', 'occ\_man\_p', 'occ\_off1\_p', 'occ\_info\_p')

msas.xvec <- c('pov\_tot1k', 'hh\_inc1k', 'hsg\_val1k')

msas.result.l <- msas.sts %>%

lapply(function(x) sf\_analysis(x, msas.yvec, msas.xvec, 'area'))

msas.result2 <- msas.result.l %>% do.call(rbind, .)

msas.xm <- msas.sts %>%

lapply(function(x) X.sac(x, msas.xvec, 'area'))

msas.xm2 <- msas.xm %>% do.call(bind\_rows, .) %>% .[rep(seq\_len(nrow(.)), each = 8),]

## 14. New York Education ####

nye <- st\_read(str\_c(ddir, 'GeoDa/nyc\_2000Census/NYC\_2000Census.shp'))

nye <- nye %>% st\_transform(3857)

nye <- nye %>%

mutate(

hs\_d = hs / over25 \* 1000,

somecol\_d = somecol / over25 \* 1000,

col\_d = college / over25 \* 1000,

master\_d = master / over25 \* 1000,

prof\_d = prof / over25 \* 1000,

phd\_d = phd / over25 \* 1000,

school\_d = SCHOOL\_CT / population \* 1000

) %>%

mutate\_if(is.numeric, funs(replace\_na2)) %>%

mutate\_if(is.numeric, funs(replace\_inf))

nyet <- nye %>% split(., .$BoroName) %>%

lapply(function(x) {x %>% mutate\_if(is.numeric, .funs = list(~scale(.)))

return(x)})

nyet.xvec <- c('school\_d', 'GENDER\_PAR', 'PER\_PRV\_SC', 'YOUTH\_DROP', 'PER\_MNRTY', 'mean\_inc', 'HS\_DROP', 'PER\_ASIAN')

nyet.yvec <- c('hs\_d', 'somecol\_d', 'col\_d', 'master\_d', 'prof\_d', 'phd\_d', 'COL\_DEGREE')

# Warning: it may take some time

nyet.result.l <- nyet %>%

lapply(function(x) sf\_analysis(x, nyet.yvec, nyet.xvec, 'area'))

nyet.xm.l <- nyet %>%

lapply(function(x) X.sac(x, nyet.xvec, 'area'))

nyet.result <- nyet.result.l %>% do.call(rbind, .)

nyet.xm <- nyet.xm.l %>% do.call(rbind, .)

## 15. New York Unemployment ####

st\_crs(nyt) <- 4326

nyt <- nyt %>% st\_transform(3857)

nyt <- nyt %>%

mutate(

other.une = otherunemp / otherethni \* 100,

mixed.une = mixedunemp / mixed \* 100,

male.une = maleunempl/ male \* 100,

his.une = hispanicun / hispanic \* 100,

fem.une = femaleunem / female \* 100,

euro.une = europeanun/ european \* 100,

ameind.une = americanun / american \* 100,

asian.une = asianunemp / asian \* 100,

afr.une = africanune / african \* 100,

withssi\_p = withssi / households \* 100,

withpubass\_p = withpubass / households \* 100,

p\_professionb = profession / poptot \* 100,

medianage = medianage %>% as.character %>% as.numeric,

gini = gini %>% as.character %>% as.numeric,

medianinco = medianinco %>% as.character %>% as.numeric) %>%

mutate\_if(is.numeric, list(~replace\_na2(.))) %>%

mutate\_if(is.numeric, list(~replace\_nan(.))) %>%

mutate\_if(is.numeric, list(~replace\_inf(.)))

nyts <- nyt %>% split(., .$boroname) %>%

lapply(function(x) {

x %>%

mutate\_if(is.numeric, .funs = list(~replace\_nan(.))) %>%

mutate\_if(is.numeric, .funs = list(~replace\_na2(.))) %>%

mutate\_if(is.numeric, .funs = list(~scale(.)))

return(x)})

nyts.yvec <- c('other.une', 'male.une', 'his.une', 'fem.une', 'euro.une', 'ameind.une', 'afr.une')

nyts.xvec <- c('medianage', 'medianinco', 'onlylessth', 'poor', 'withssi\_p', 'withpubass\_p', 'p\_professionb', 'struggling')

nyts.result.l <- nyts %>%

lapply(function(x) sf\_analysis(x, nyts.yvec, nyts.xvec, 'area'))

nyts.xm.l <- nyts %>%

lapply(function(x) X.sac(x, nyts.xvec, 'area'))

nyts.result <- nyts.result.l %>% do.call(rbind, .)

nyts.xm <- nyts.xm.l %>% do.call(rbind, .)

## To gather each result

all.results <-

bind\_rows(

airbnb.result2,

balt.result2,

bost.result2,

imix.result2,

chhs.result2,

hein.result2,

cinc.result2,

colu.result2,

denv.result2,

nats.result2,

elec16.result2,

elec12.result2,

phx.result2,

msas.result2,

nyet.result,

nyts.result

)

all.results.a <-

bind\_rows(

airbnb.xm,

balt.xm,

bost.xm,

imix.xm,

chhs.xm,

hein.xm,

cinc.xm,

colu.xm,

denv.xm,

nats.xm,

elec.xm,

elec.xm,

phx.xm,

msas.xm,

nyet.xm,

nyts.xm

)

all.results <- bind\_cols(all.results, all.results.a)