Built environment | Demographic | Socioeconomic | Transit

Spatiotemporal patterns of drug activity in Chicago

Luna Zhiyue Xia

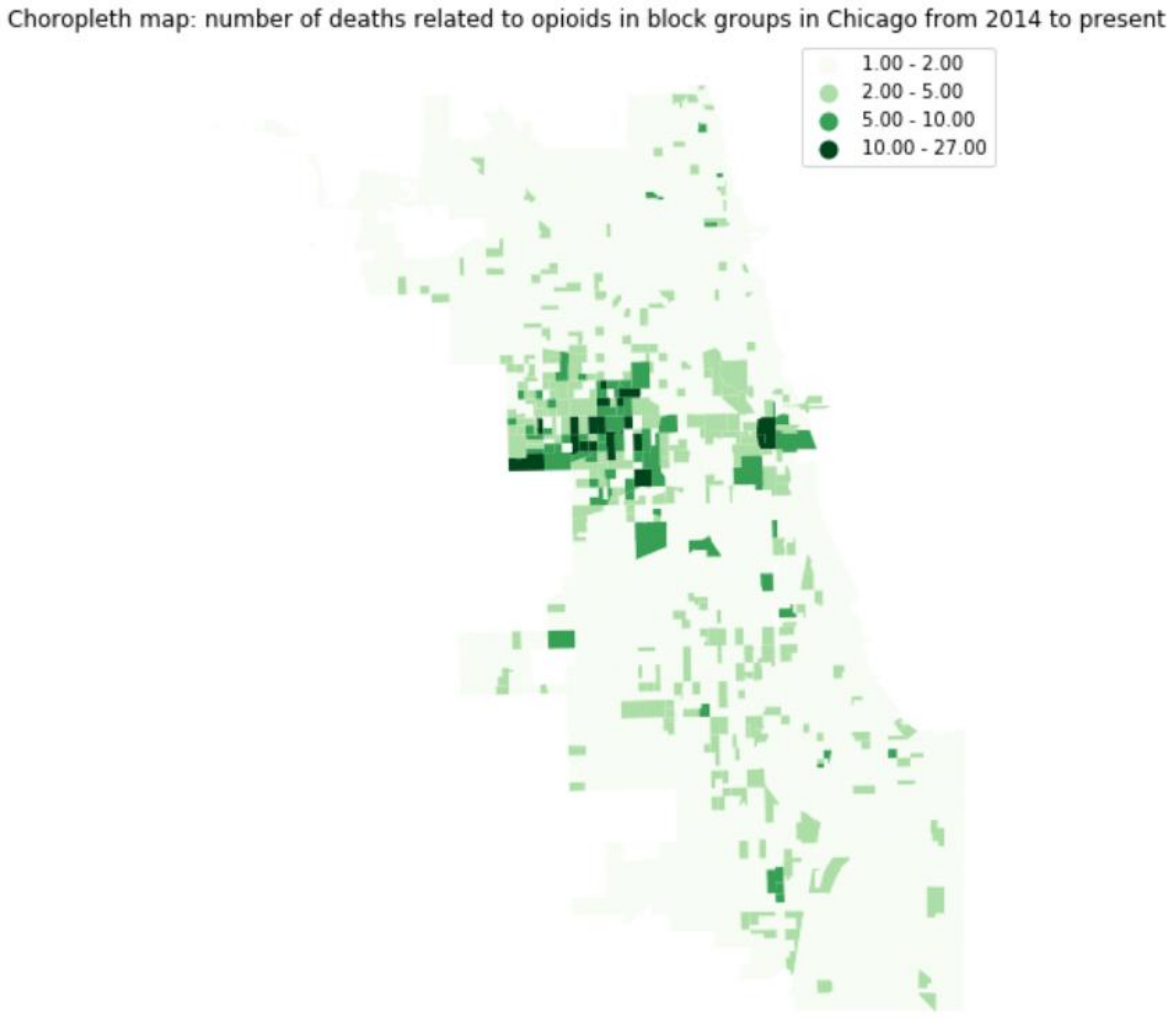
Three hypothesizes:

1) The spatial patterns of drug activities in city areas are significantly clustered (tested by ESDA and reported in last update)

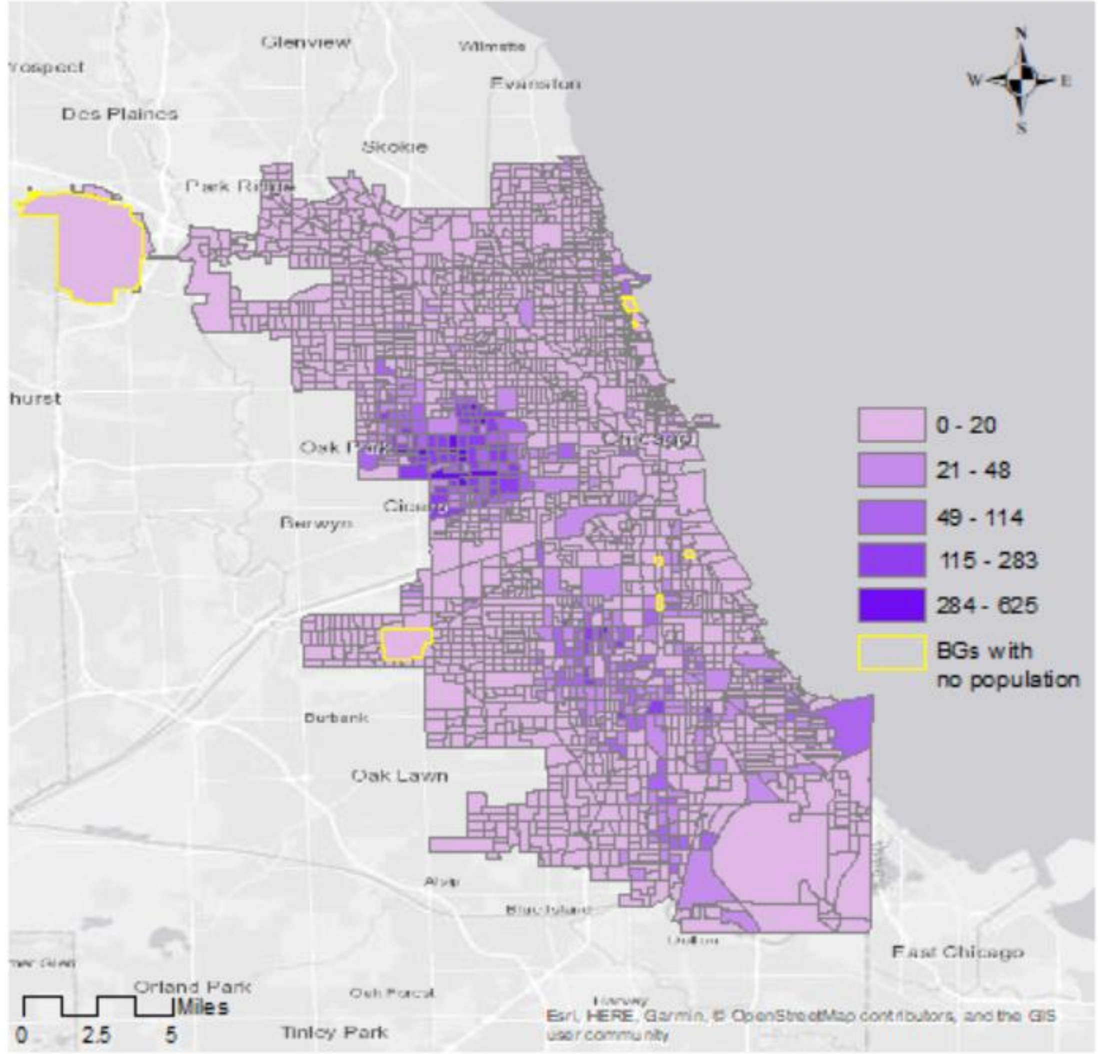
2) Drug arrests (which record times and locations) can be used as a proxy of drug activities

3) The spatial patterns of drug activities are correlated with built environment, demographic and socioeconomic.

Bivariate Moran Statistic

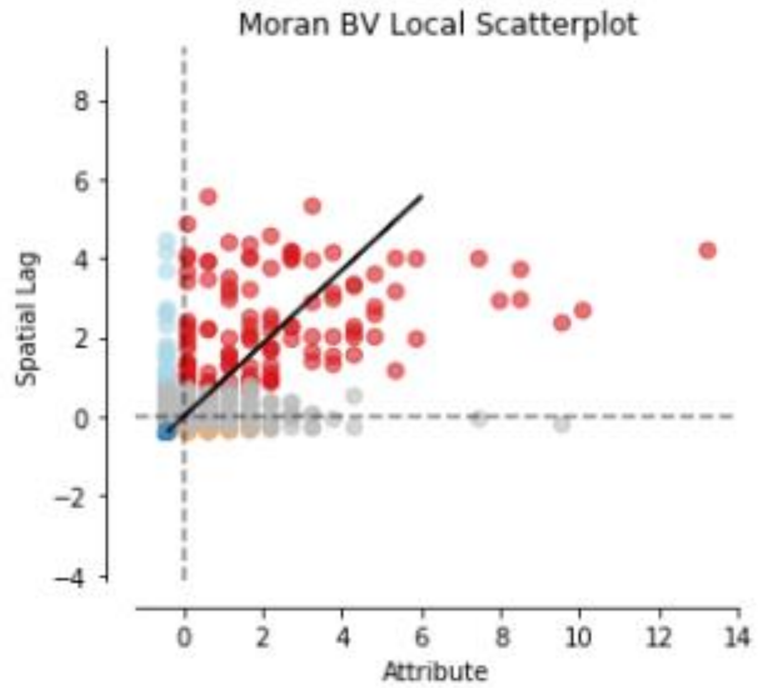


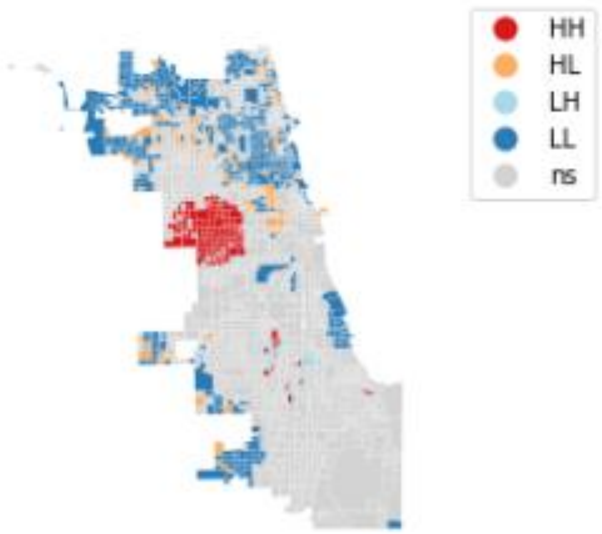
Bivariate ESDA between Opioids related death and Drug arrest



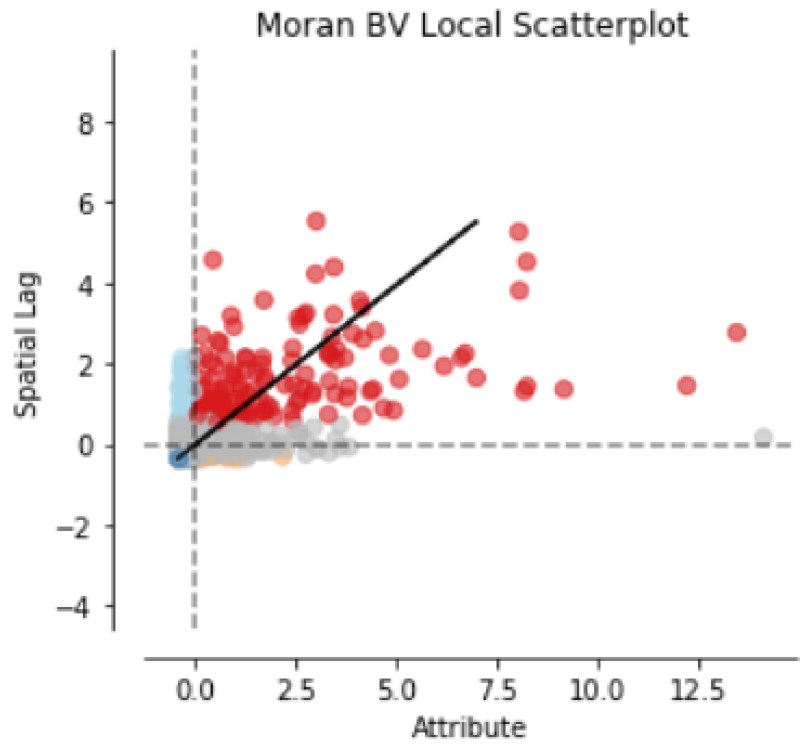
Bivariate Moran Statistic

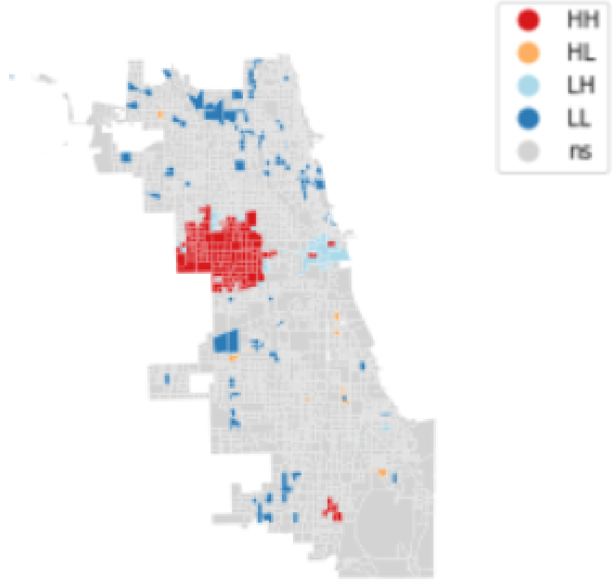
Global bivariate Moran’s I is 0.42 with p-value smaller than 0.001
→ Positive correlation between Death and Arrest in nearby areas





Bivariate analysis:
Death surrounded by arrest
HH cluster indicate high death rate surrounded by high drug arrest





Arrest surrounded by death
HH cluster indicate high frequency of arrest surrounded by high death rate

GWR and MGWR

Variables:
Total number: 182
Sources: EPA, ACS, ESRI Business, Chicago open data portal

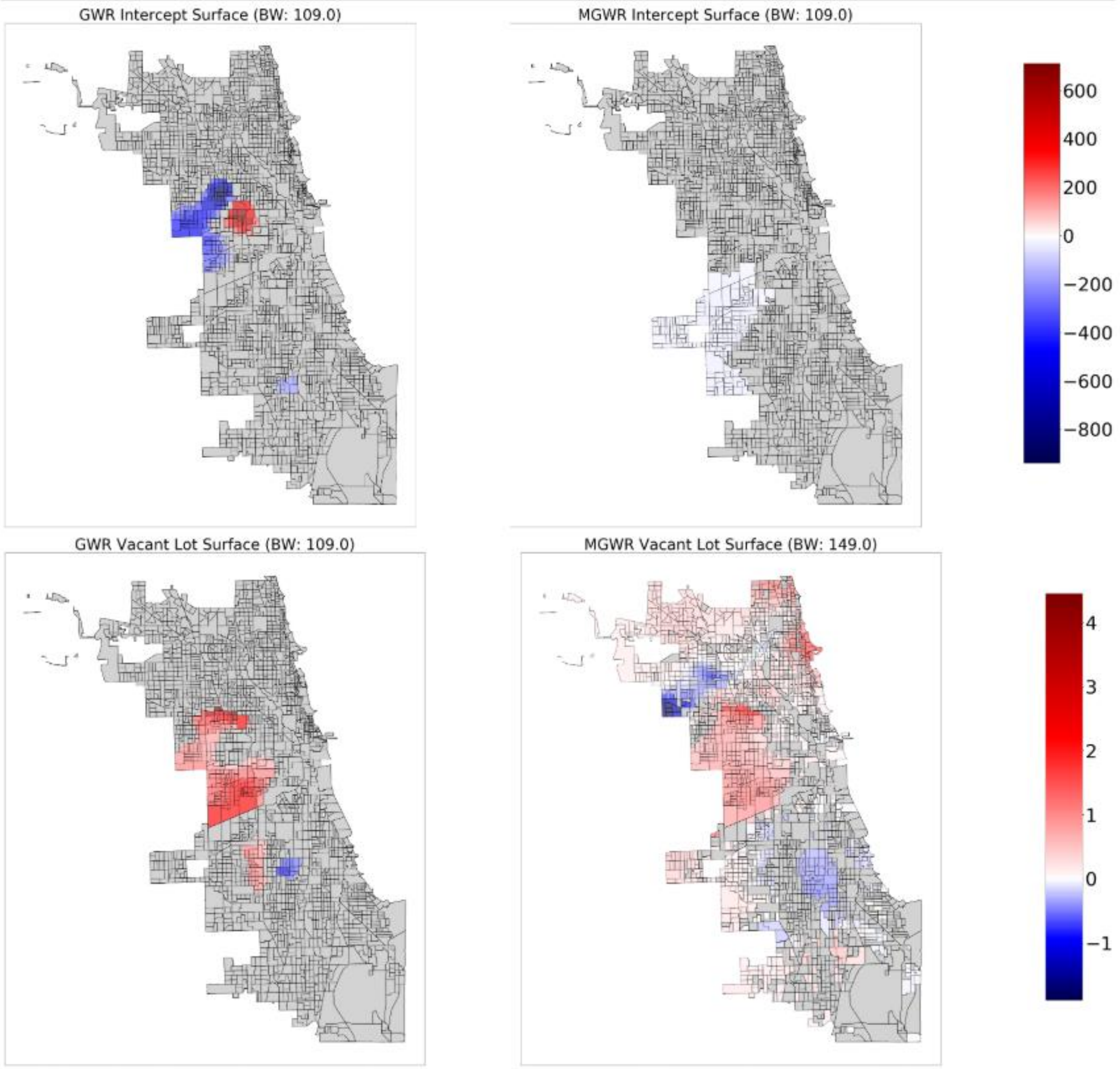
```
list_var = chicago_variable.columns
print(*list_var, sep=", ")
```

```
GEOID10_x, GEOID_Data, geometry, DEATH, FID_x, GEOID, arrest_per1000, FID_y, GEOID10_y, P_WRKAGE, AUTOOWN0, PCT_AO
0, AUTOOWN1, PCT_AO1, AUTOOWN2P, PCT_AO2P, WORKERS, R_LOWWAGEW, R_MEDWAGEW, R_HIWAGEWK, R_PCTLOWWA, EMPTOT, E5_RET
10, E5_OFF10, E5_IND10, E5_SVC10, E5_ENT10, E8_RET10, E8_OFF10, E8_IND10, E8_SVC10, E8_ENT10, E8_ED10, E8_HLTH10,
E8_PUB10, E_FEDT10, E_FEDRET10, E_FEDOFF10, E_FEDIND10, E_FEDSVC10, E_FEDENT10, E_LOWWAGEW, E_MEDWAGEW, E_HIWAGEW
K, E_PCTLOWWA, AC_TOT, AC_WATER, AC_LAND, AC_UNPR, D1A, D1B, D1C, D1C5_Ret10, D1C5_Off10, D1C5_Ind10, D1C5_Svc10,
D1C5_Ent10, D1C8_Ret10, D1C8_Off10, D1C8_Ind10, D1C8_Svc10, D1C8_Ent10, D1C8_Ed10, D1C8_Hlth1, D1C8_Pub10, D1D, D1
_flag, D2A_JPHH, D2B_E5MIX, D2B_E5MIXA, D2B_E8MIX, D2B_E8MIXA, D2A_EPHHM, D2C_TRPMX1, D2C_TRPMX2, D2C_TRIPEQ, D2R_
JOBPOP, D2R_WRKEMP, D2A_WRKEMP, D2C_WREMIX, D3a, D3aao, D3amm, D3apo, D3b, D3bao, D3bmm3, D3bmm4, D3bpo3, D3bpo4,
D4a, D4b025, D4b050, D4c, D4d, D5ar, D5ae, D5br, D5br_Flag, D5be, D5be_Flag, D5cr, D5cri, D5ce, D5cei, D5dr, D5dr
i, D5de, D5dei, NatWalkInd, TrAccess_I, Pop_byTr, Pop_byTr_m, Pop_byTr_1, Pop_byTr_a, Pct_Pop_by, Pct_Pop__1, Pct_
Pop__2, Pct_Pop__3, HU_byTr, HH_byTr, Wrks_byTr, Wrks_byTr_, Wrks_byTr1, Wrks_byT_1, Pct_Wrks_b, Pct_Wrks_1, Pct_W
rks_2, Pct_Wrks_3, LoWgWrks_b, LoWgWrks_1, LoWgWrks_2, MeWgWrks_b, HiWgWrks_b, LoMeWgWrks, LoMeWgWr_1, LoMeWgWr_2,
Pct_LoWgWr, Pct_LoWg_1, Pct_LoWg_2, Pct_LoWg_3, Pct_MeWgWr, Pct_LoMeWg, Pct_LoMe_1, Pct_LoMe_2, Pct_LoMe_3, Jobs_b
yTr, Jobs_byTr_, Jobs_byTr1, Job_byTr_A, Pct_Jobs_b, Pct_Jobs_1, Pct_Jobs_2, Pct_Jobs_3, FID_x, GEOID10_x, TotPop,
TotalHouse, PMale, PFemale, PWhite, PBlack, PAsian, Phispanic, PBachelorHigher, MedianHouseIncome, PPovertyHouse,
PPovertyIndv, PEmploy, Punemploy, PFulltimeEmploy, PParttimeEmploy, POccupyHouse, POwner, PRenter, FID_y, GEOID10_
y, GasStation, VacLot, VacBldg, ParkingLot, AlleyCount, AlleyLen, ParkArea, HighSch
```

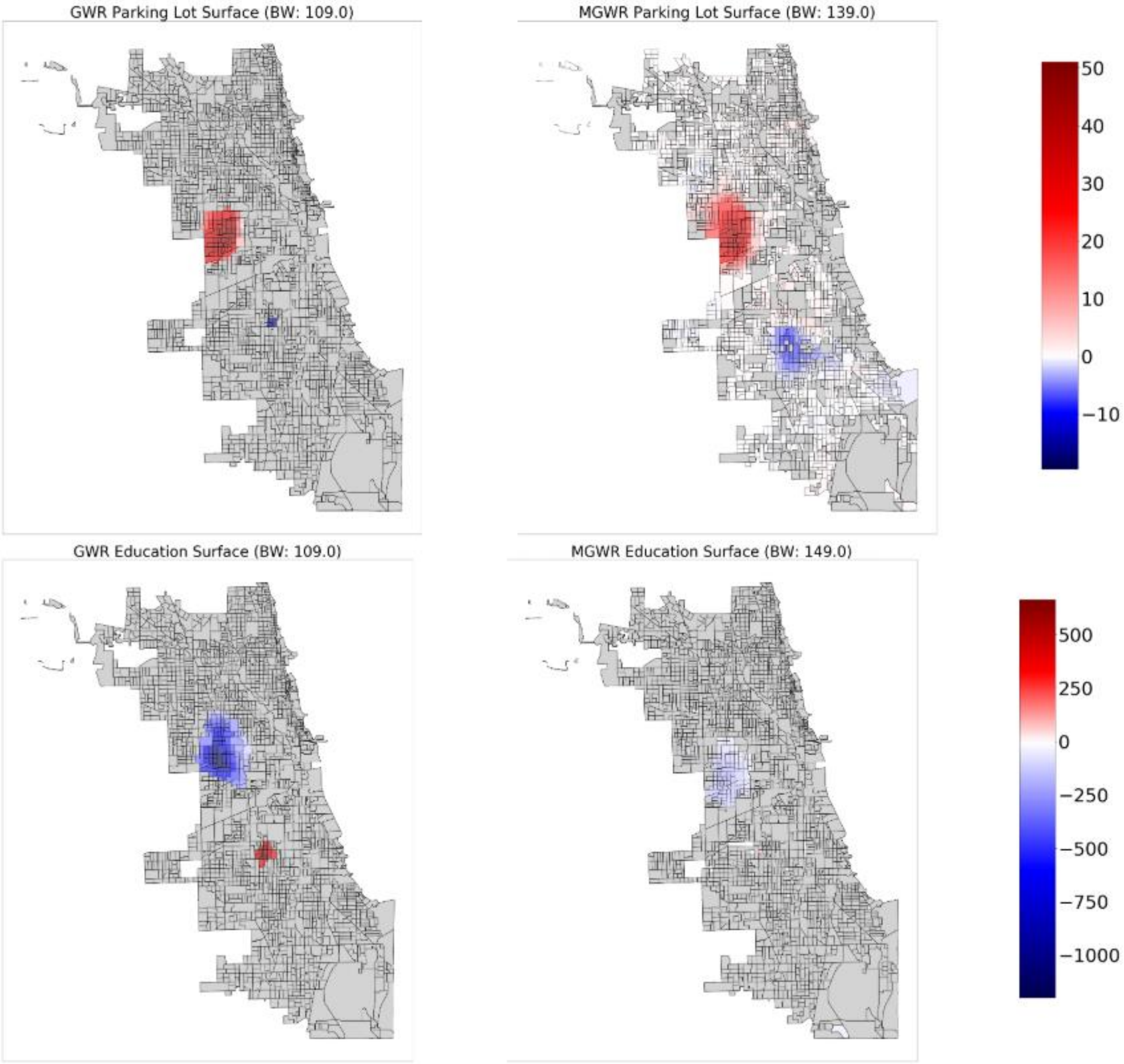
Using Random Forest to select 5 important variables to predict frequency of drug arrests
Model : frequency of drug arrest per 1000 population ~ VacantLot + ParkingLot + PBachelorHigher + Pct_LoMeWg + D5ar

Pct_LoMeWg: percent of low-medium wage workers have accessibility to public transit
D5ar: jobs accessible by 45 min transit

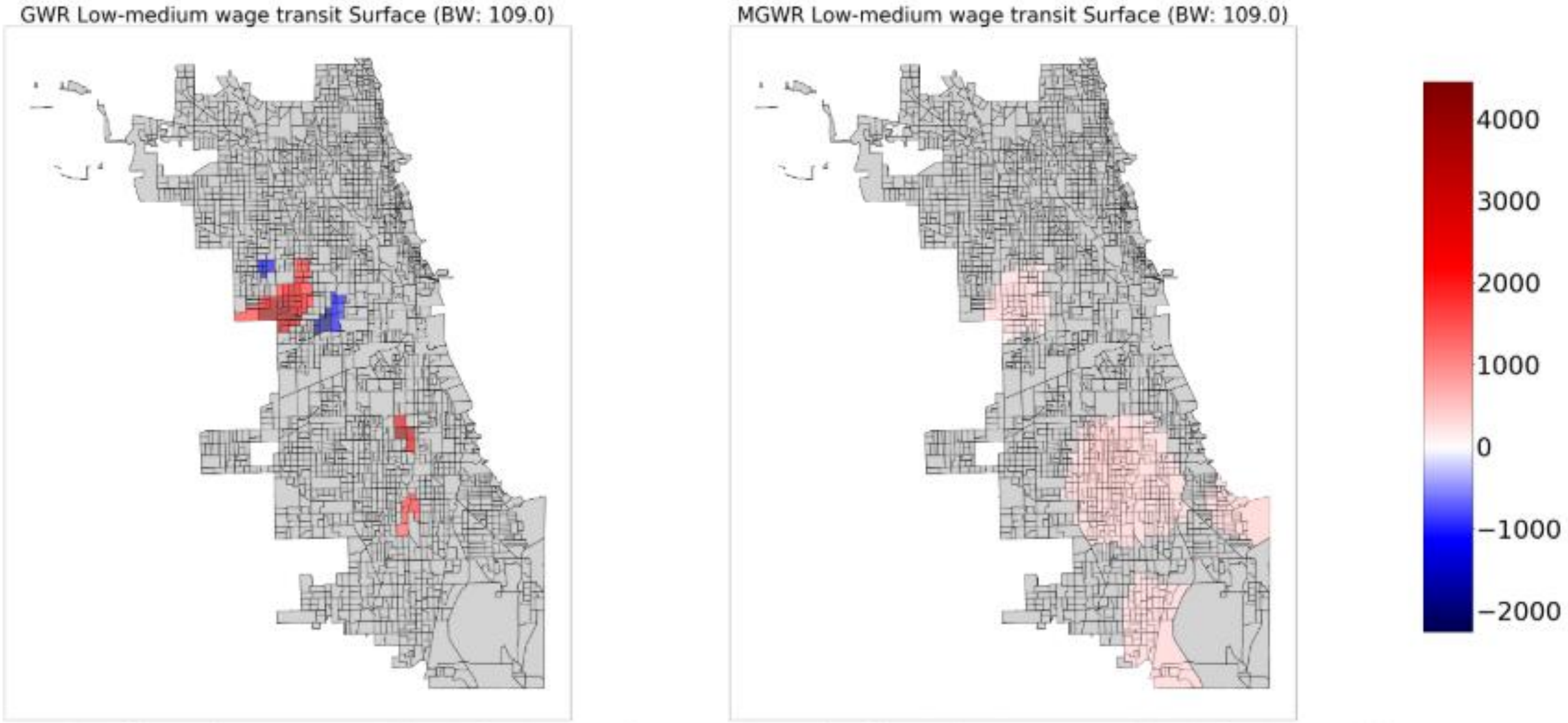
GWR and MGWR

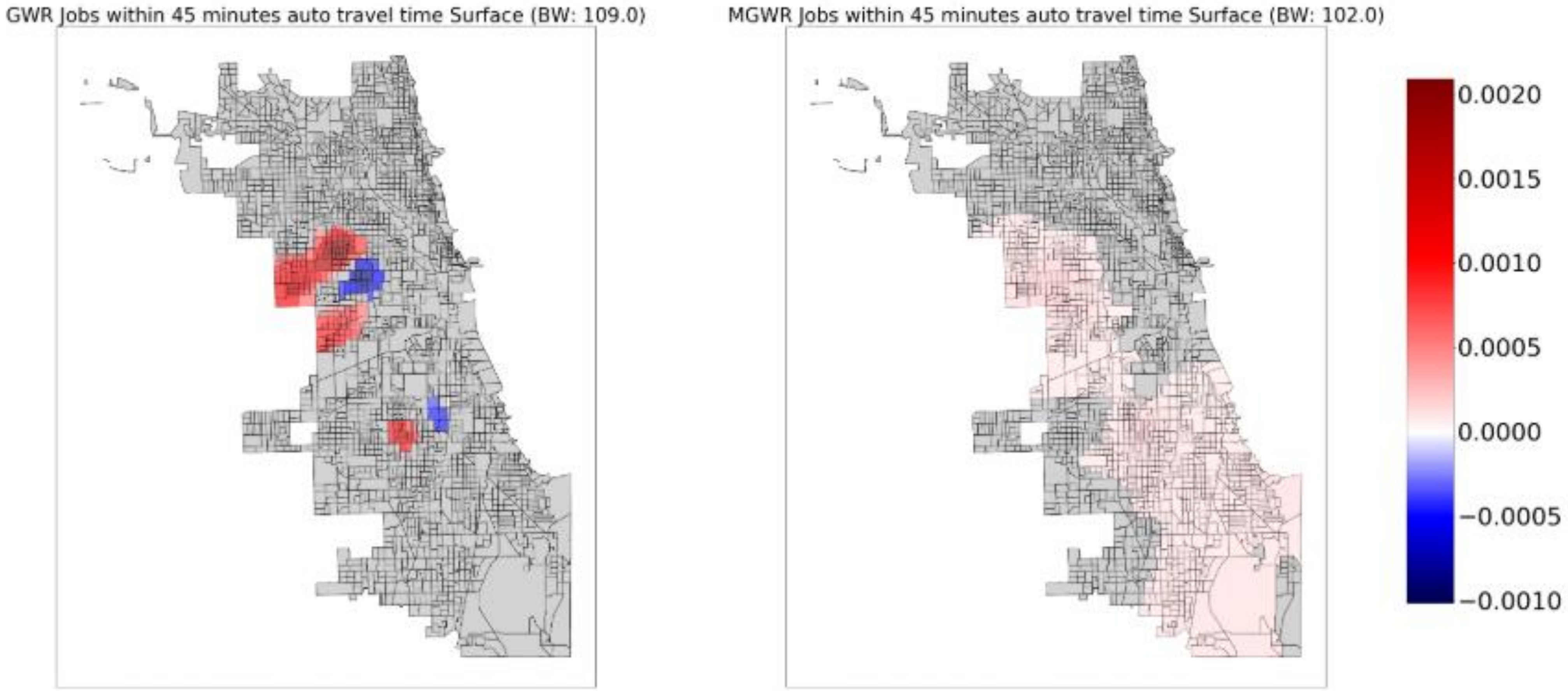

GWR Intercept Surface (BW: 109.0)
MGWR Intercept Surface (BW: 109.0)
600
400
200
0
−200
−400
−600
−800
GWR Vacant Lot Surface (BW: 109.0)
MGWR Vacant Lot Surface (BW: 149.0)
4
3
2
1
0
−1

GWR and MGWR


GWR Parking Lot Surface (BW: 109.0)
MGWR Parking Lot Surface (BW: 139.0)
50
40
30
20
10
0
−10
GWR Education Surface (BW: 109.0)
MGWR Education Surface (BW: 149.0)
500
250
0
−250
−500
−750
−1000

GWR and MGWR






```
LinAlgError: Matrix is singular.
```

Backfitting: 100% 200/200 [22:03<00:00, 6.62s/it]

```
[109. 135. 211. 123. 133. 107.]
```

Inference: 0% 0/1 [00:00<?, ?it/s]

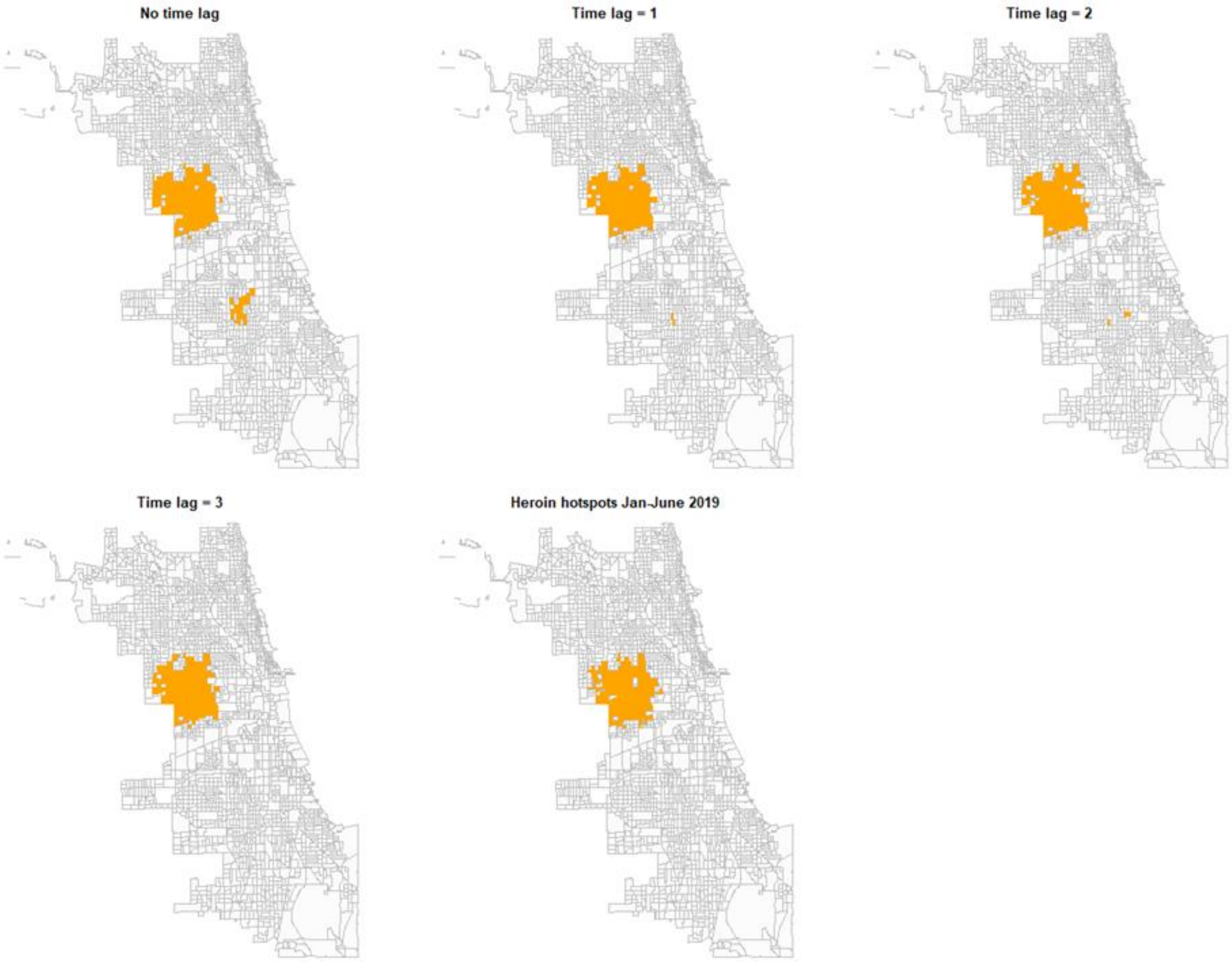
Use random forest to predict hotspots

- 3 type of additional variables create
  - Spatiotemporal lag variables – nb (time lag = 1, 2, 3)
    - No. of block groups in neighborhood (Queen adjacency) that belonged to a hotspot in the past time periods (6 mos)
  - Temporal autocorrelation variables – hotspot (time lag = 1, 2, 3)
    - Was this block group in a hotspot in the past time periods? Y/N
  - Trend variables – nb_trend, hotspot_trend
    - Difference between ‘nb’ for different time lag (eg. ‘nb_t-1’ – ‘nb_t-2’)
    - Difference between ‘hotspot’ for different time lag (eg. ‘hotspot_t-2’ – ‘hotspot_t-3’)

Time series analysis

Train Random Forest model using data between 2016 and 2018

Predict in 2019



Next steps

- Try Geographical Random Forest (package ‘spatialML’ only in R CRAN, any python package?)
- (potential) estimate the accessibility to opioid treatment center