HUANYU MONTHLY UPDATE

Outline

- Research Idea Overview
- Scores Data
  - Methods & Results
  - Exploratory Data Analysis
- Census Data
  - Methods & Results
- SLR Exploration
  - Introduction
  - Model Performance
- GWR Exploration
  - Introduction
  - Data Visualization
  - Model Performance

Research Idea Overview

- This study explored **UHF-level population data** to understand the **similarities and differences** across various socio-economic and demographic factors and how these factors influenced **decision-making patterns** and response to health crises during and after the COVID-19 pandemic.

Scores Data: Methods & Results

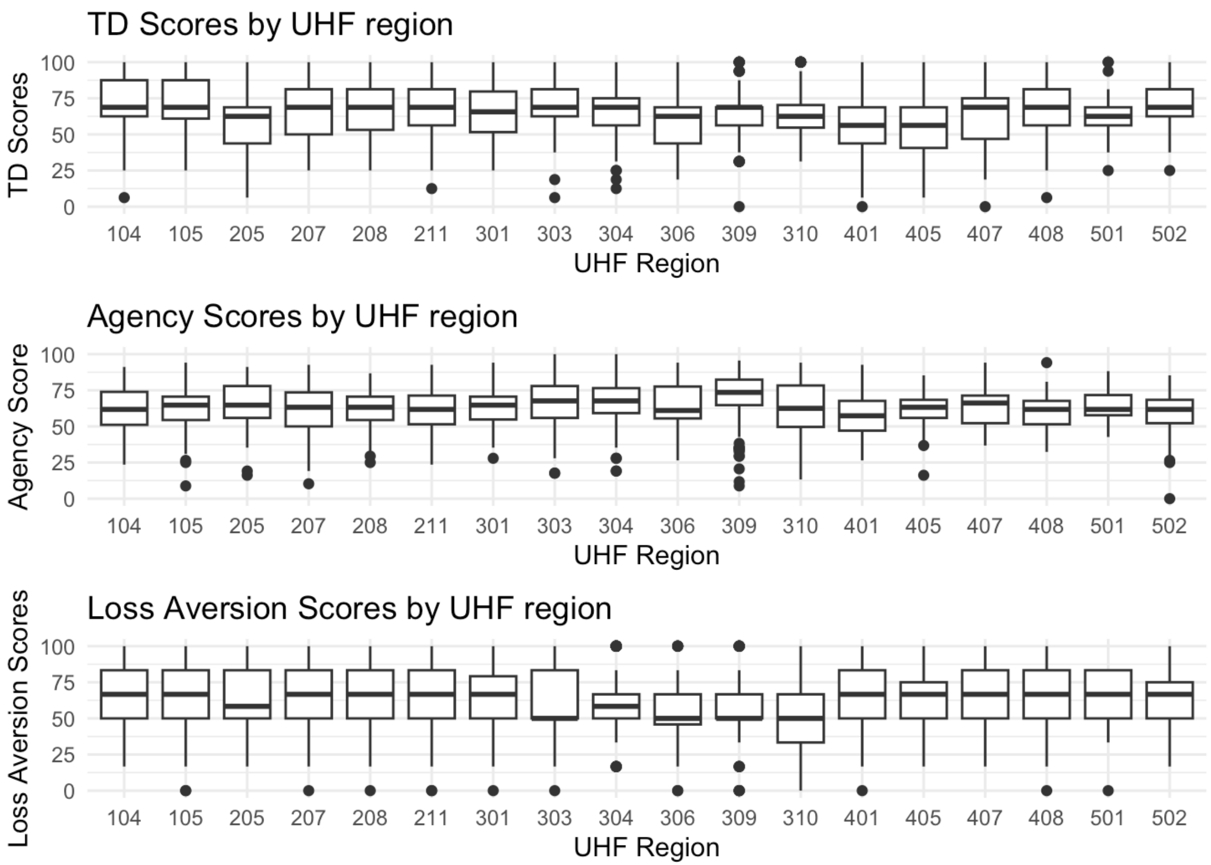
- Merge 3 Scores:
  - Agency Score
  - Temporal Discounting Score
  - Loss Aversion Score

- Standardization:
  - Scaled all scores to a range of 1-100

- Summarizing Data:
  - Combined scores and participants at the UHF level
  - Calculated the median score for each UHF

- Filtering:
  - Filtered UHF levels where participants > 15 and got 18 regions in total
  - Exported the results into 1_output/dat_merge_2_UHF>=15.csv

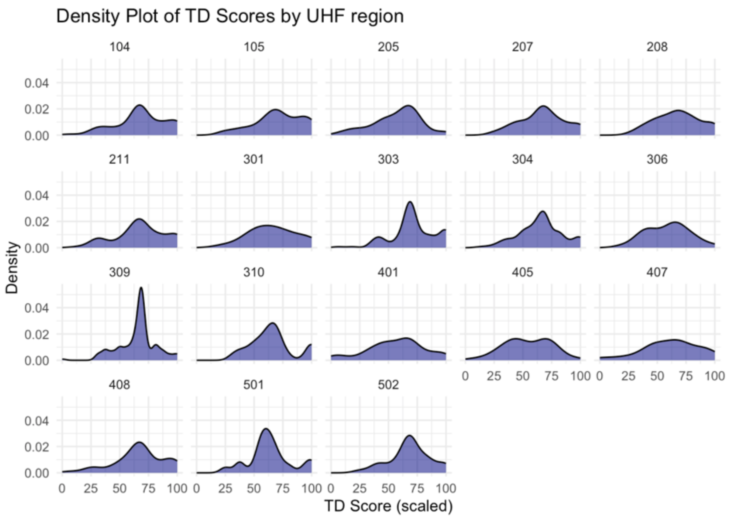
Scores Data: Exploratory Data Analysis

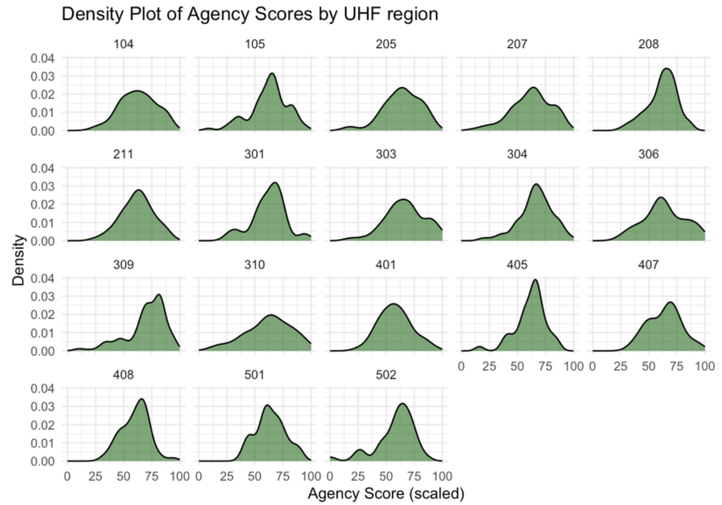


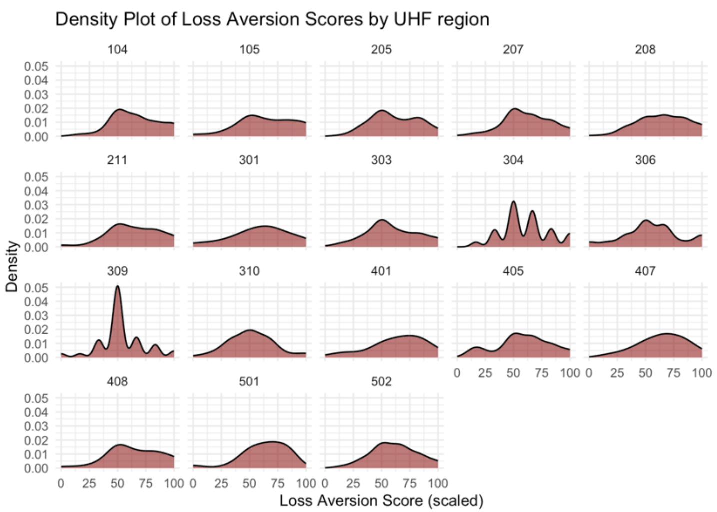
Comment:

UHF 401 (Long Island City - Astoria) has the lowest median scores for both temporal discounting and agency, while UHF 309 (Union Square - Lower East Side) has the highest median score for agency.

Scores Data: Exploratory Data Analysis







Comment:

All three scores follow normal distributions on UHF level in general, including those with smaller sample sizes (region 301 - Washington Heights, 407 - Southwest Queens, 501 - Port Richmond).

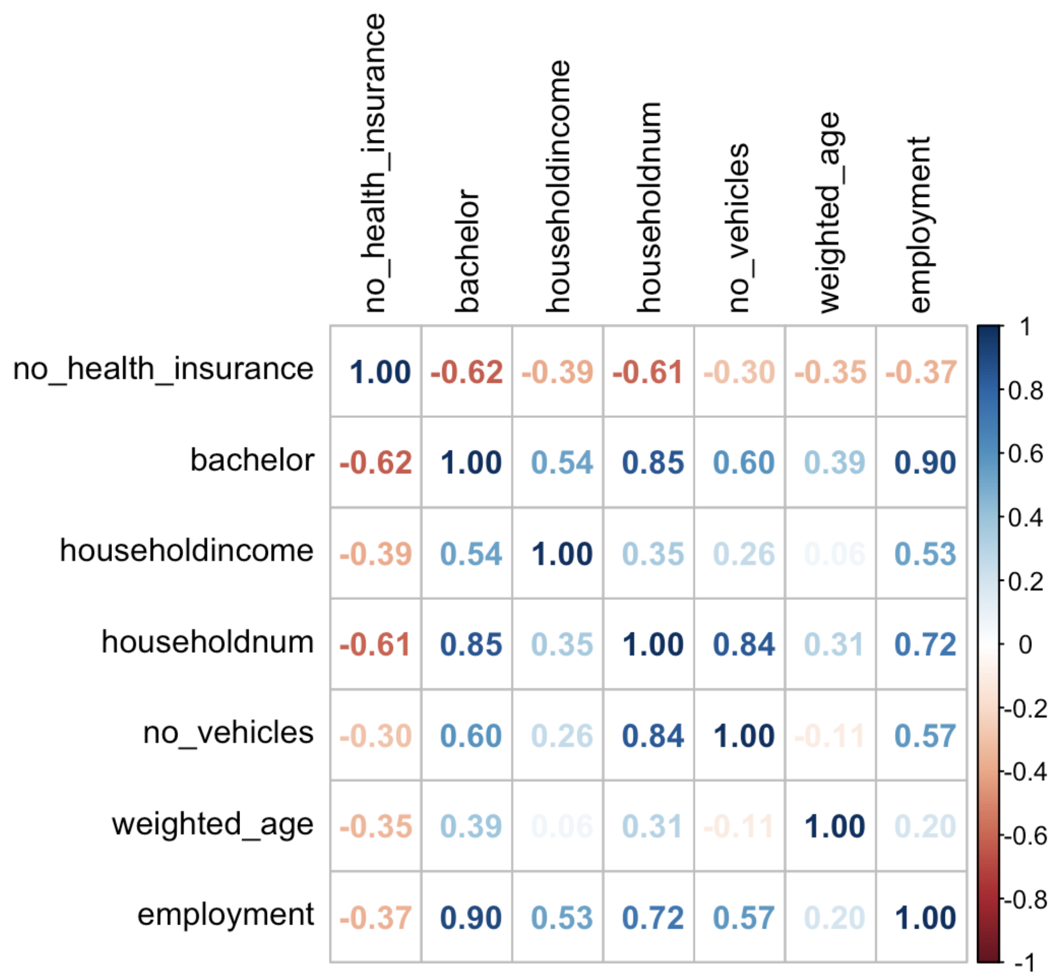
Census Data: Methods & Results

- Merge ZCTA - ZIP Code - UHF mapping to one dataset
    - Exported the results into 2_output/mapping_guide.csv

- Download census data
    - Include population, individuals without health insurance, education levels, household income, household numbers, households without vehicles, age distribution, and employment status; then calculate the percentage of these variables based on population in UHF-level
    - Converted the census data from ZCTA level to UHF level and got 42 regions in total; 2_output/uhf_summary.csv

Census Data: Methods & Results

- Correlation Plot

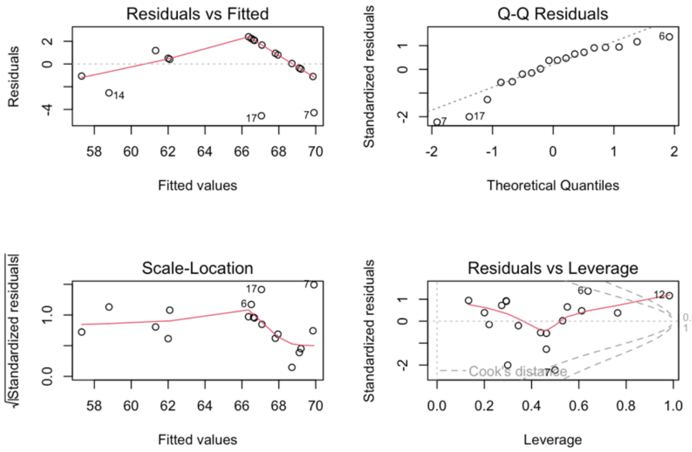


SLR Exploration: Introduction

- Used Ordinary Least Squares (OLS) methods
- Aimed to fit a SLR model about scores based on census data from 18 UHF areas.
- Then used it to predict scores for the remaining 24 UHF regions.

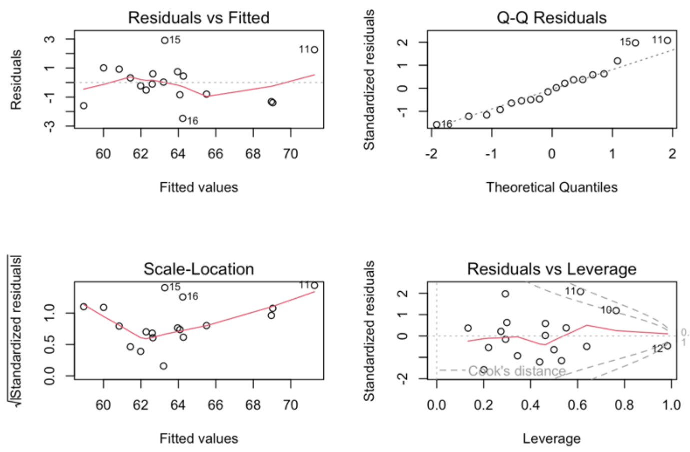
```
##
## Call:
## lm(formula = median_td_score ~ no_health_insurance + bachelor +
##     householdincome + householdnum + no_vehicles + weighted_age +
##     employment, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5563 -0.9092  0.4560  1.5451  2.3923
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         118.247278  26.074434   4.535  0.00108 **
## no_health_insurance  -3.343602   0.880388  -3.798  0.00350 **
## bachelor             -0.641565   0.555444  -1.155  0.27492
## householdincome      -0.005186   0.002767  -1.874  0.09041 .
## householdnum         -1.494042   0.469305  -3.184  0.00976 **
## no_vehicles           0.677998   0.179913   3.768  0.00367 **
## weighted_age          0.145357   0.411473   0.353  0.73123
## employment            0.407859   0.456678   0.893  0.39279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.718 on 10 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.7703, Adjusted R-squared:  0.6095
## F-statistic: 4.791 on 7 and 10 DF,  p-value: 0.01327
```

SLR Exploration: Model Performance



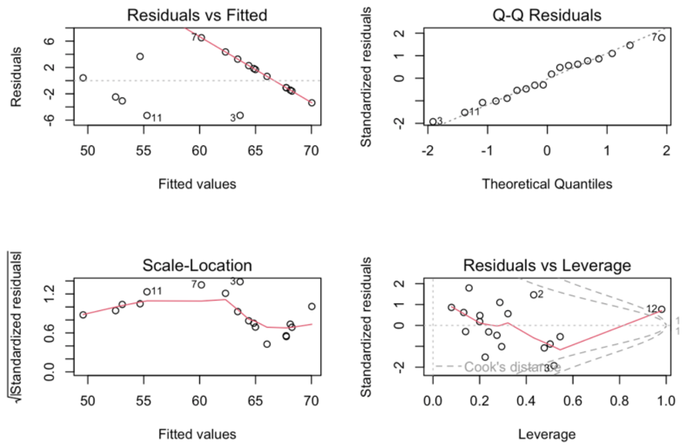

```
mean(model_td$residuals^2)
```

```
## [1] 4.105642
```




```
mean(model_ag$residuals^2)
```

```
## [1] 1.702577
```




```
mean(model_los_3$residuals^2)
```

```
## [1] 10.39861
```

Comment:

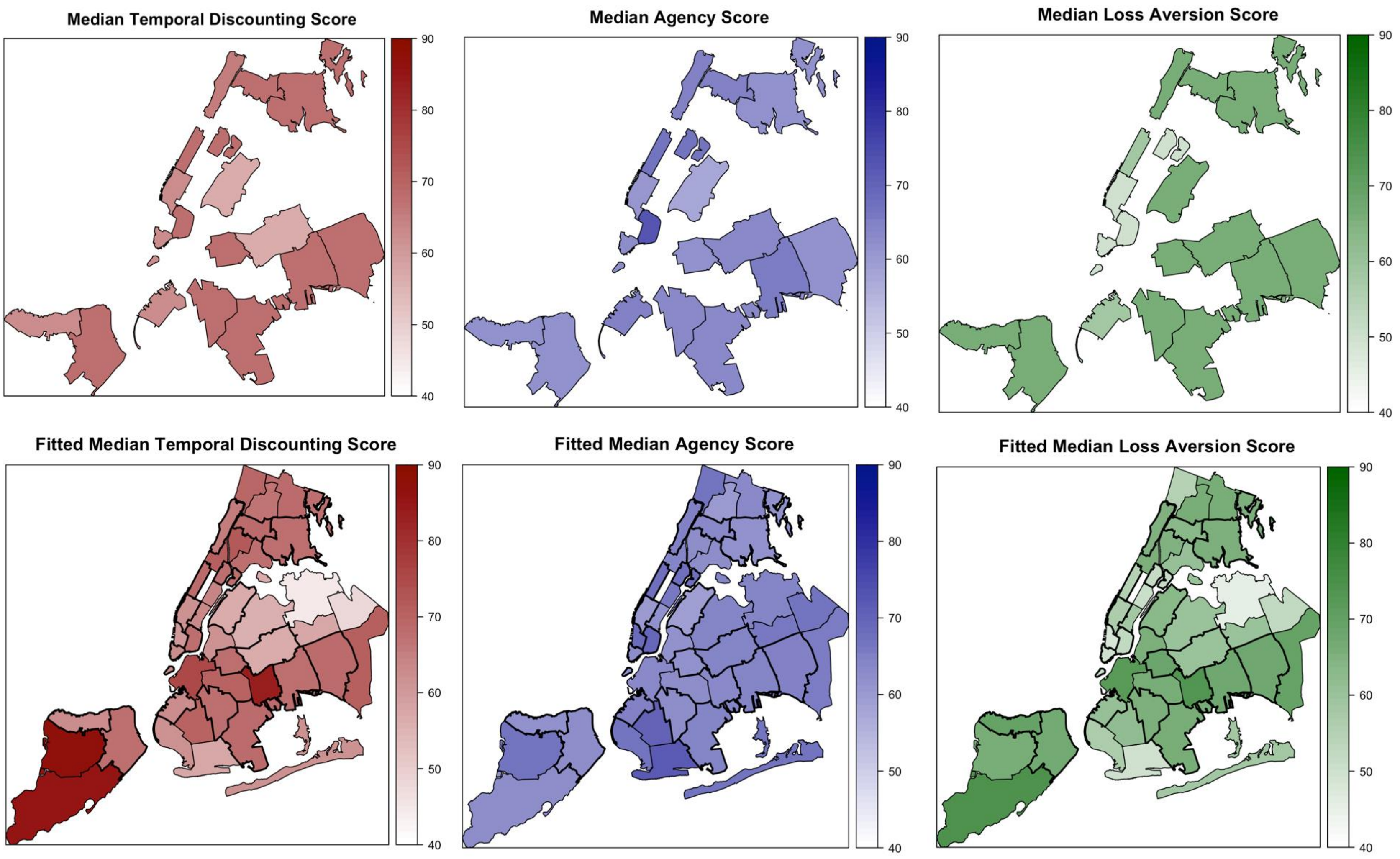
- Detailed code and results:
  - model.html
- Whole dataset after prediction:
  - 3_output/whole_dataset.csv

GWR Exploration: Introduction

- Aimed to fit a GWR model about scores based on census data from 18 UHF areas.
- Then used it to predict scores for the remaining 24 UHF regions.

```
## Call:
## gwr(formula = median_ag_score ~ no_health_insurance + bachelor +
##     householdincome + householdnum + no_vehicles + weighted_age +
##     employment, data = sp_data_train, bandwidth = bwG_ag, gweight = gwr.Gauss,
##     hatmatrix = TRUE)
## Kernel function: gwr.Gauss
## Fixed bandwidth: 144355.1
## Summary of GWR coefficient estimates at data points:
##                           Min.     1st Qu.      Median     3rd Qu.        Max.
## X.Intercept.        98.4617527  99.7803921 100.3350745 100.5905142 100.9987732
## no_health_insurance  0.1599872   0.1660066   0.1697520   0.1783941   0.1908162
## bachelor             1.2324980   1.2460650   1.2551338   1.2572380   1.2803131
## householdincome     -0.0014954  -0.0014125  -0.0013940  -0.0013400  -0.0012480
## householdnum        -1.0558823  -1.0394310  -1.0343447  -1.0317301  -1.0221307
## no_vehicles          0.3413946   0.3423886   0.3444250   0.3458659   0.3484888
## weighted_age         0.5415082   0.5487896   0.5554236   0.5650097   0.5807190
## employment          -1.0723346  -1.0644593  -1.0627551  -1.0570106  -1.0481244
##                      Global
## X.Intercept.        99.8300
## no_health_insurance  0.1829
## bachelor             1.2492
## householdincome     -0.0013
## householdnum        -1.0275
## no_vehicles          0.3426
## weighted_age         0.5566
## employment          -1.0594
## Number of data points: 37
## Effective number of parameters (residual: 2traceS - traceS'S): 8.42462
## Effective degrees of freedom (residual: 2traceS - traceS'S): 28.57538
## Sigma (residual: 2traceS - traceS'S): 1.455421
## Effective number of parameters (model: traceS): 8.219785
## Effective degrees of freedom (model: traceS): 28.78021
## Sigma (model: traceS): 1.450233
## Sigma (ML): 1.279039
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 148.69
## AIC (GWR p. 96, eq. 4.22): 131.4333
## Residual sum of squares: 60.52982
## Quasi-global R2: 0.7915936
```

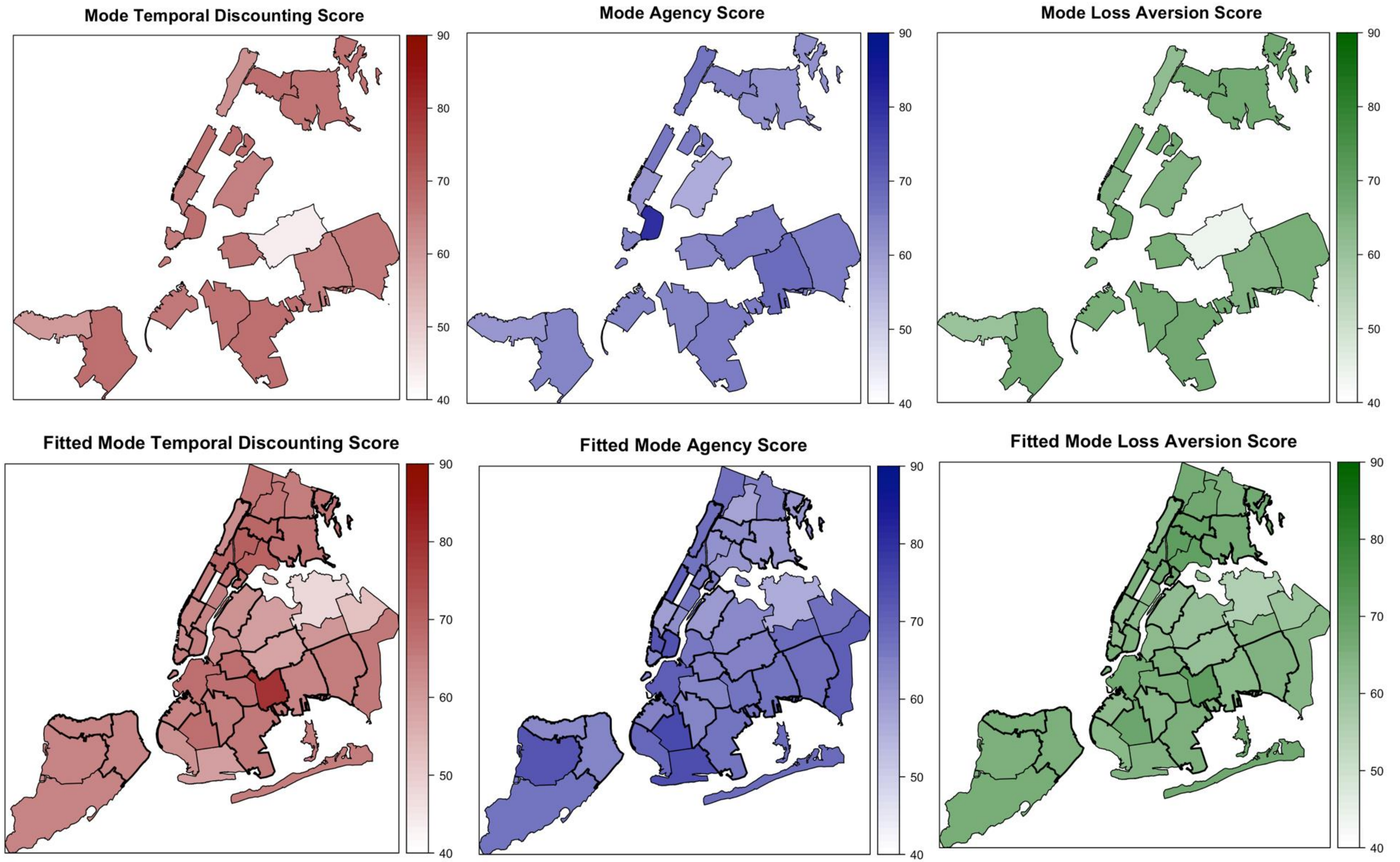
GWR Exploration: Data Visualization (Median)



Comment:

The first row visualized the median scores across 18 UHF areas, with deeper colors indicating higher scores. On the second row, GWR predicted these scores in remaining 24 UHF areas and then presented on the map.

GWR Exploration: Data Visualization (Mode)



Comment:

The first row visualized the mode scores across 18 UHF areas, with deeper colors indicating higher scores.
On the second row, GWR predicted these scores in remaining 24 UHF areas and then presented on the map.

GWR Exploration: Model Performance (Median)

```
##
##  Brunsdon, Fotheringham & Charlton (1999) ANOVA
##
## data:  gwrG_td
## F = 142.62, df1 = 22.107, df2 = 19.810, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## SS GWR improvement   SS GWR residuals
##         124.284685           1.361613
```

```
LMZ.F3GWR.test(gwrG_td)
```

```
##
## Leung et al. (2000) F(3) test
##
##                    F statistic Numerator d.f. Denominator d.f.     Pr(>)
## (Intercept)            27.7957         8.0673            19.81 3.650e-09 ***
## no_health_insurance    46.6695        10.8598            19.81 1.048e-11 ***
## bachelor               60.6928        10.6747            19.81 9.340e-13 ***
## householdincome       171.3753        10.1307            19.81 < 2.2e-16 ***
## householdnum          168.3985        14.4766            19.81 < 2.2e-16 ***
## no_vehicles           177.1026         7.8675            19.81 < 2.2e-16 ***
## weighted_age           70.2326         5.6036            19.81 3.700e-12 ***
## employment             91.4077         9.9474            19.81 2.438e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##  Brunsdon, Fotheringham & Charlton (1999) ANOVA
##
## data:  gwrG_ag
## F = 1.4412, df1 = 15.243, df2 = 28.963, p-value = 0.193
## alternative hypothesis: greater
## sample estimates:
## SS GWR improvement   SS GWR residuals
##           1.296247          60.529822
```

```
LMZ.F3GWR.test(gwrG_ag)
```

```
##
## Leung et al. (2000) F(3) test
##
##                    F statistic Numerator d.f. Denominator d.f.    Pr(>)
## (Intercept)            2.42064        4.96514           28.963 0.059999 .
## no_health_insurance    0.63634       18.81581           28.963 0.845530
## bachelor               1.60231       14.04152           28.963 0.137868
## householdincome        4.44322        7.08409           28.963 0.001798 **
## householdnum           1.89044       13.38983           28.963 0.074013 .
## no_vehicles            0.80142        6.50345           28.963 0.585188
## weighted_age           2.27995        3.51705           28.963 0.091817 .
## employment             0.79568       16.36952           28.963 0.680850
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##  Brunsdon, Fotheringham & Charlton (1999) ANOVA
##
## data:  gwrG_los
## F = 3.7203, df1 = 15.243, df2 = 28.963, p-value = 0.001152
## alternative hypothesis: greater
## sample estimates:
## SS GWR improvement   SS GWR residuals
##           11.79092          213.28765
```

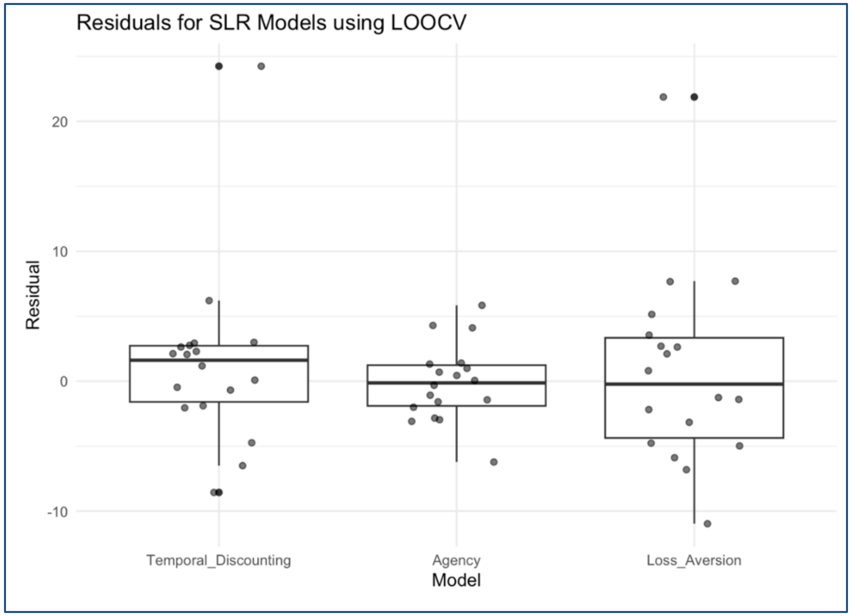
```
LMZ.F3GWR.test(gwrG_los)
```

```
##
## Leung et al. (2000) F(3) test
##
##                    F statistic Numerator d.f. Denominator d.f.     Pr(>)
## (Intercept)             1.4488         4.9651           28.963 0.2371267
## no_health_insurance     3.6732        18.8158           28.963 0.0008438 ***
## bachelor                1.0159        14.0415           28.963 0.4651130
## householdincome         3.4878         7.0841           28.963 0.0076523 **
## householdnum            5.1174        13.3898           28.963 0.0001169 ***
## no_vehicles             2.9684         6.5035           28.963 0.0197870 *
## weighted_age            1.9994         3.5170           28.963 0.1279498
## employment              2.0665        16.3695           28.963 0.0426815 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

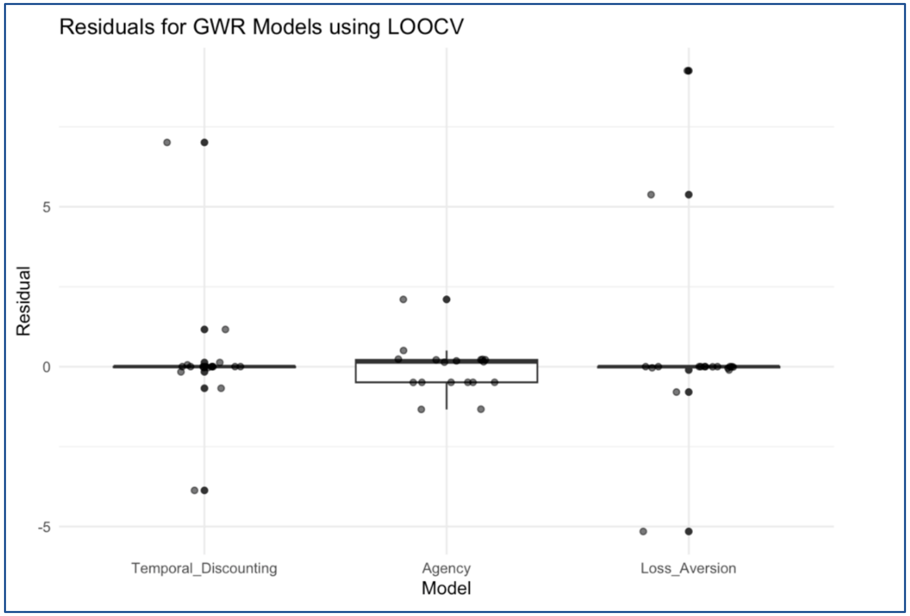
Comment:

- Model fit tests for all three models show an improvement in the explanatory power of GWR over SLR.
- The spatial heterogeneity variables are shown.
- ***AIC** can also be used to test model fit, but it is not possible to infer whether the differences between models are statistically different.

LOOCV (Median)




```
## Temporal Discounting          Agency     Loss Aversion
##             45.55767         8.31241          51.22241
```




```
## Temporal Discounting          Agency     Loss Aversion
##            3.6685299       0.5561021         7.8802843
```