Assessing Minnesota's Public Transit

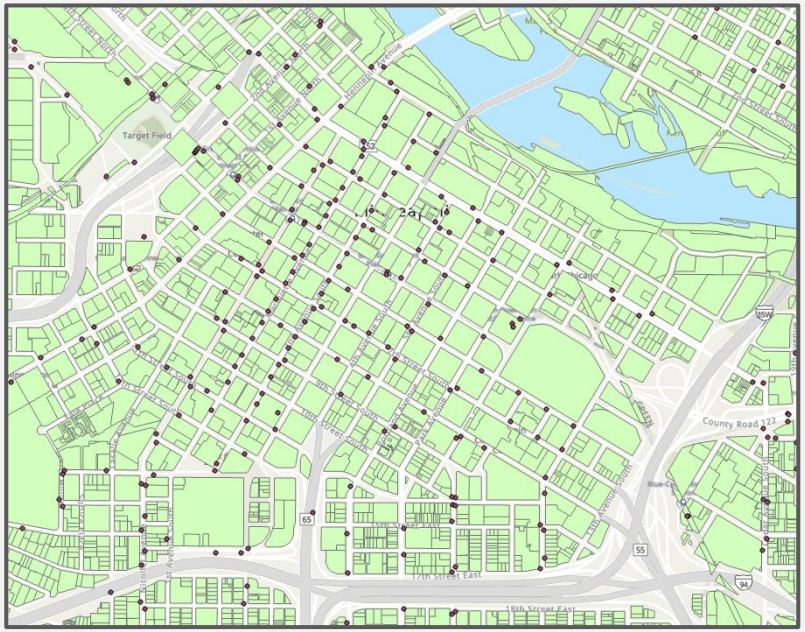
Lucas Major

Background

- Transit Score
  - Measure of a location’s accessibility to public transit methods
  - Important for measuring future transit developments
- Current state
  - Sites like Zillow, Redfin have scores for many properties
  - Scores are limited to Minneapolis/Saint Paul
  - Many other cities in Minnesota have access to transit
    - Green line Expansion
    - Bus lines running to Burnsville, Woodbury
- Goal
  - Create a model that scores every parcel within the MET 7 region based on accessibility to transit

Data

- Via API
  - Parcel data for 7-county region
    - Contains location information
  - Transit data for Minnesota
    - Point for every transit stop in Minnesota
    - Dataset of high frequency routes



- Sensitivity Analysis/Results Verification
  - Current Zillow transit scores from parcels in Minneapolis

Methodology

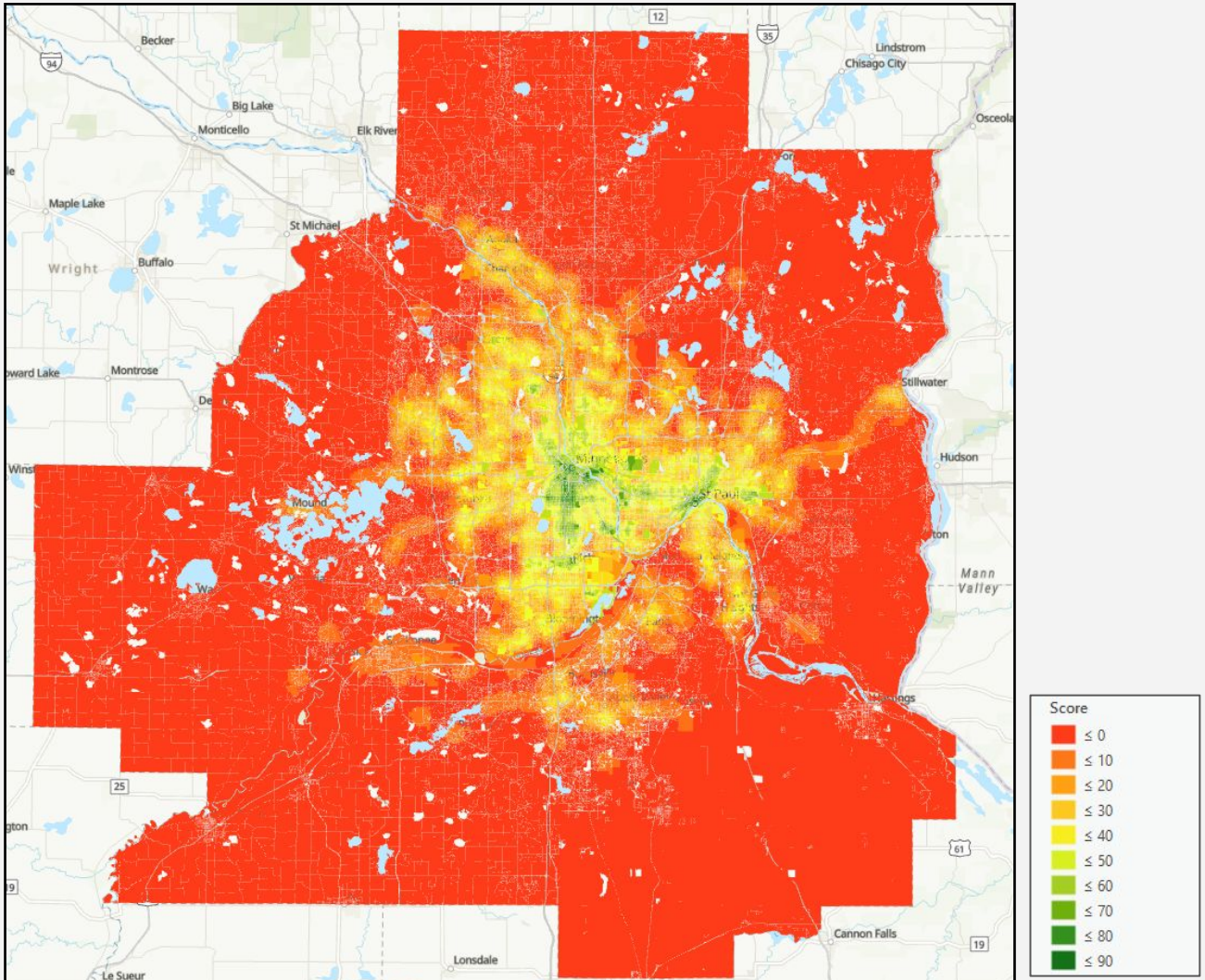
- Weighted Linear combination
  - Distance to transit stop
  - Type of transit stop (Bus/Light Rail)
  - Frequency of route
  - Sums the calculated scores for each transit stop

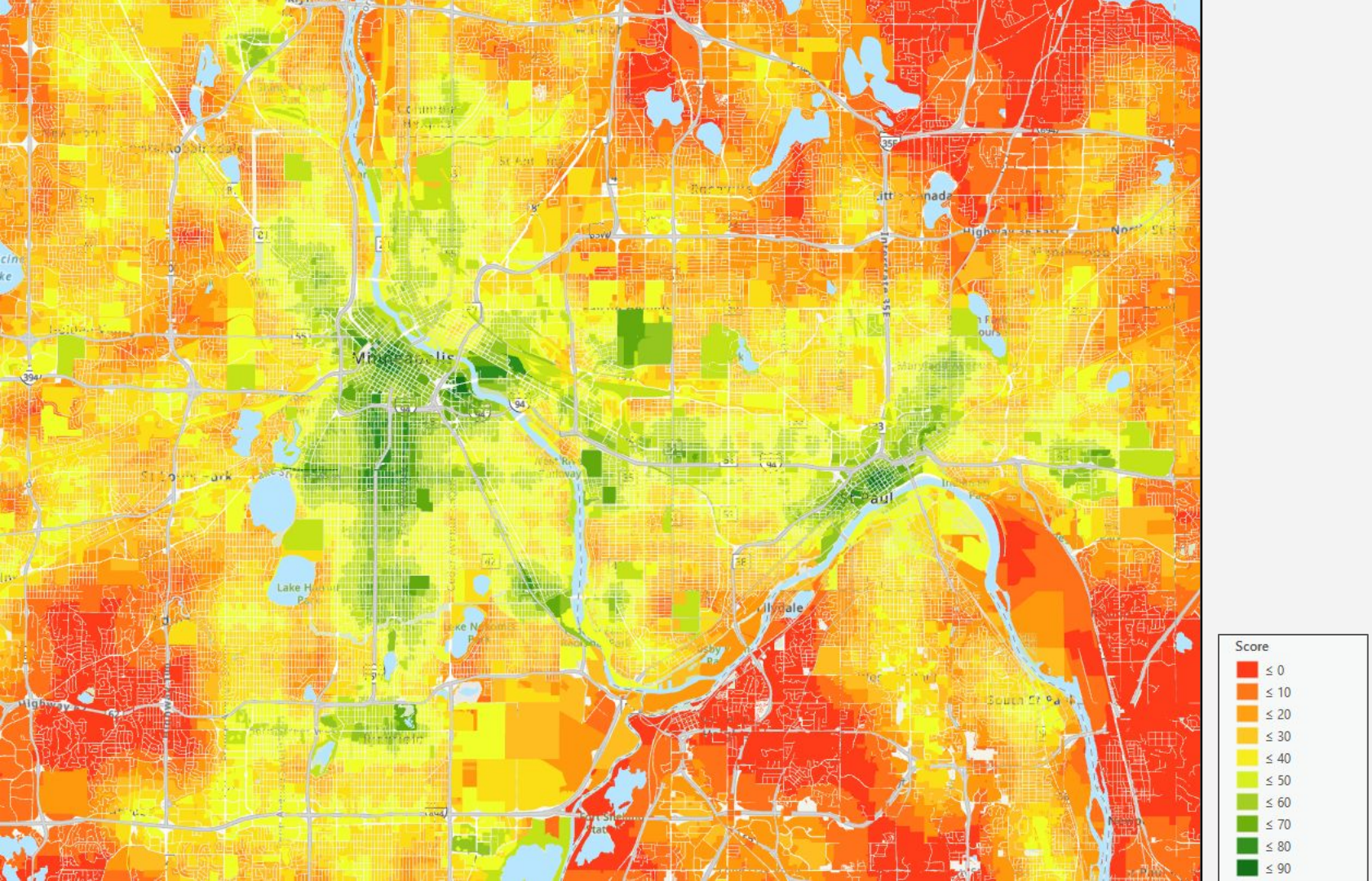
$$Score = \sum_{i=1}^{n} (\omega_L(x_{i1}) + \omega_B(x_{i2}) + \omega_F(x_{i3})) * d_i$$

$$\omega_L = 0.5 \quad \omega_B = 0.2 \quad \omega_F = 0.3 \quad n = 45$$

- Accuracy Assessment/Sensitivity Analysis
  - Sample 50 random parcels
  - Compare Zillow transit score with calculated score
  - Compute R^2 and MSE
  - Tune model parameters for highest accuracy

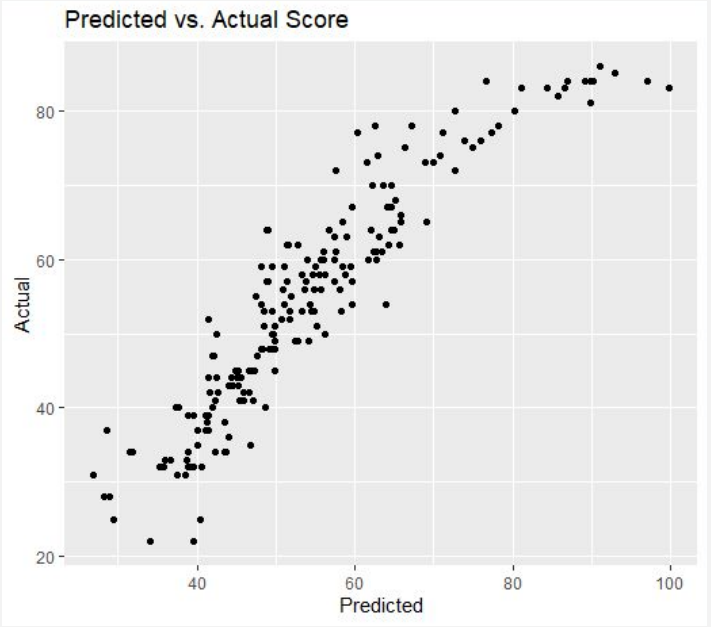
R^2	RMSE
.75	.21


Becker
Big Lake
Monticello
Elk River
Maple Lake
St Michael
Wright
Buffalo
Montrose
Lindstrom
Chisago City
Osceola
Stillwater
Hudson
Mound
St Paul
Mann Valley
Cannon Falls
Lonsdale
Le Sueur
Score
≤ 0
≤ 10
≤ 20
≤ 30
≤ 40
≤ 50
≤ 60
≤ 70
≤ 80
≤ 90


Score
≤ 0
≤ 10
≤ 20
≤ 30
≤ 40
≤ 50
≤ 60
≤ 70
≤ 80
≤ 90

Methodology

- Linear Model
  - Trained from existing Zillow data
  - Parameters
    - Distance to nearest
    - Number of transit stops within 1 KM
    - Frequency
    - Type(Bus/Light rail)
    - Home Sale value
  - Number, Distance, and Frequency were significant

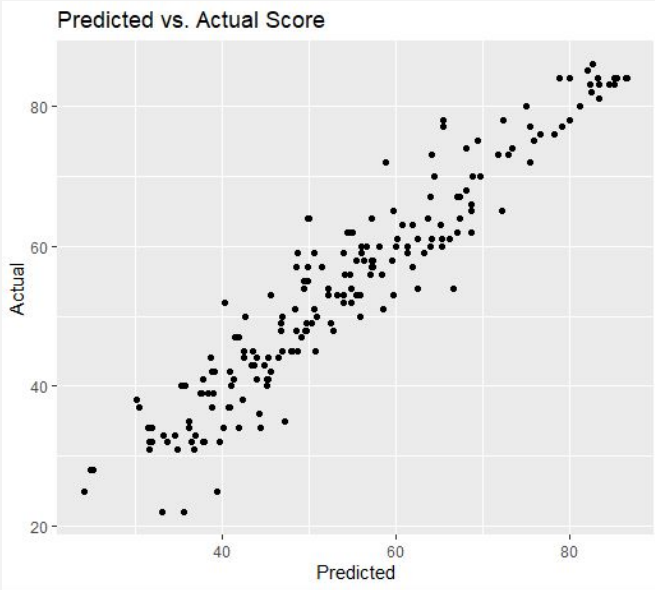


$$Score = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}$$

R^2	RMSE
0.86	0.09

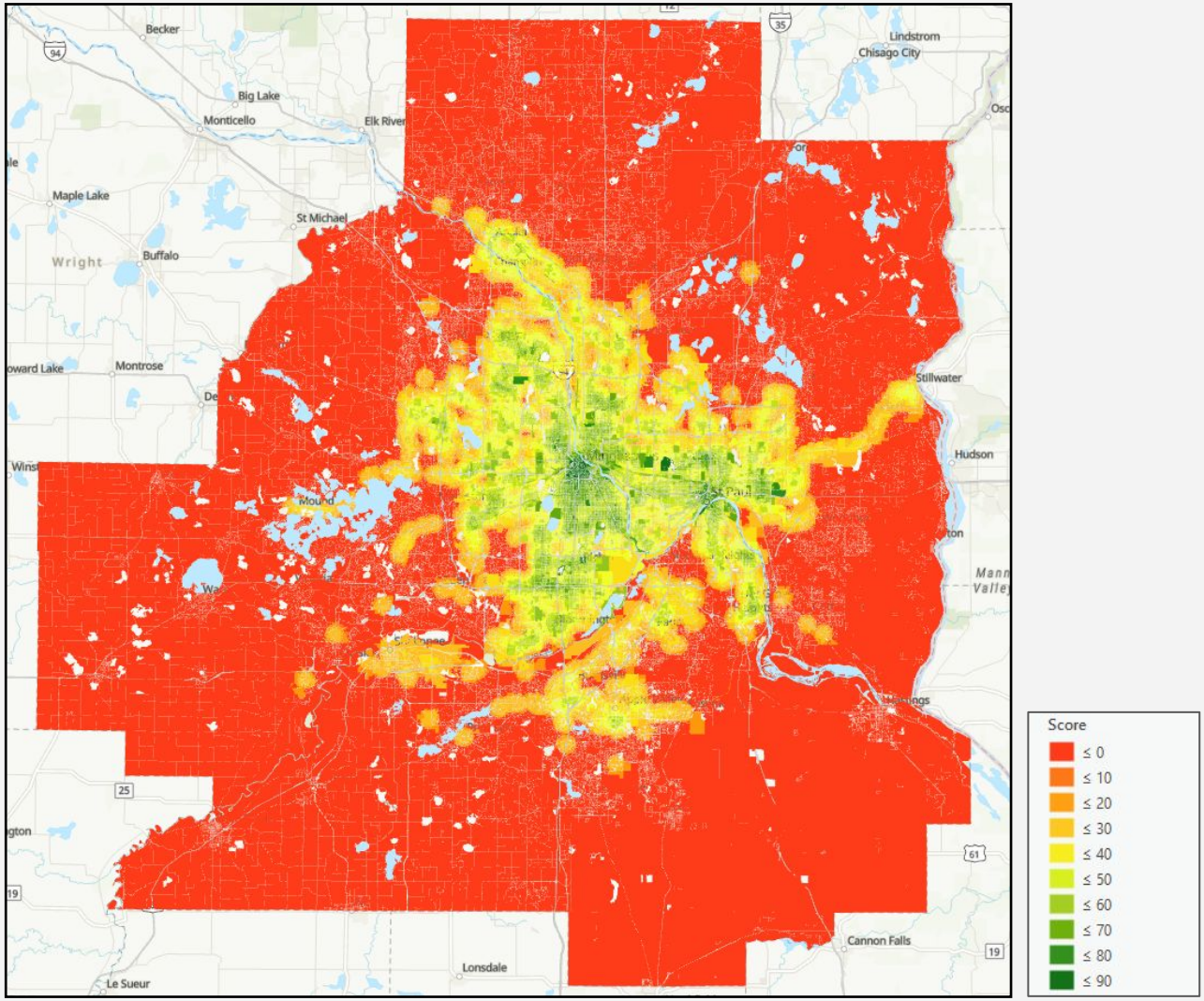
Methodology

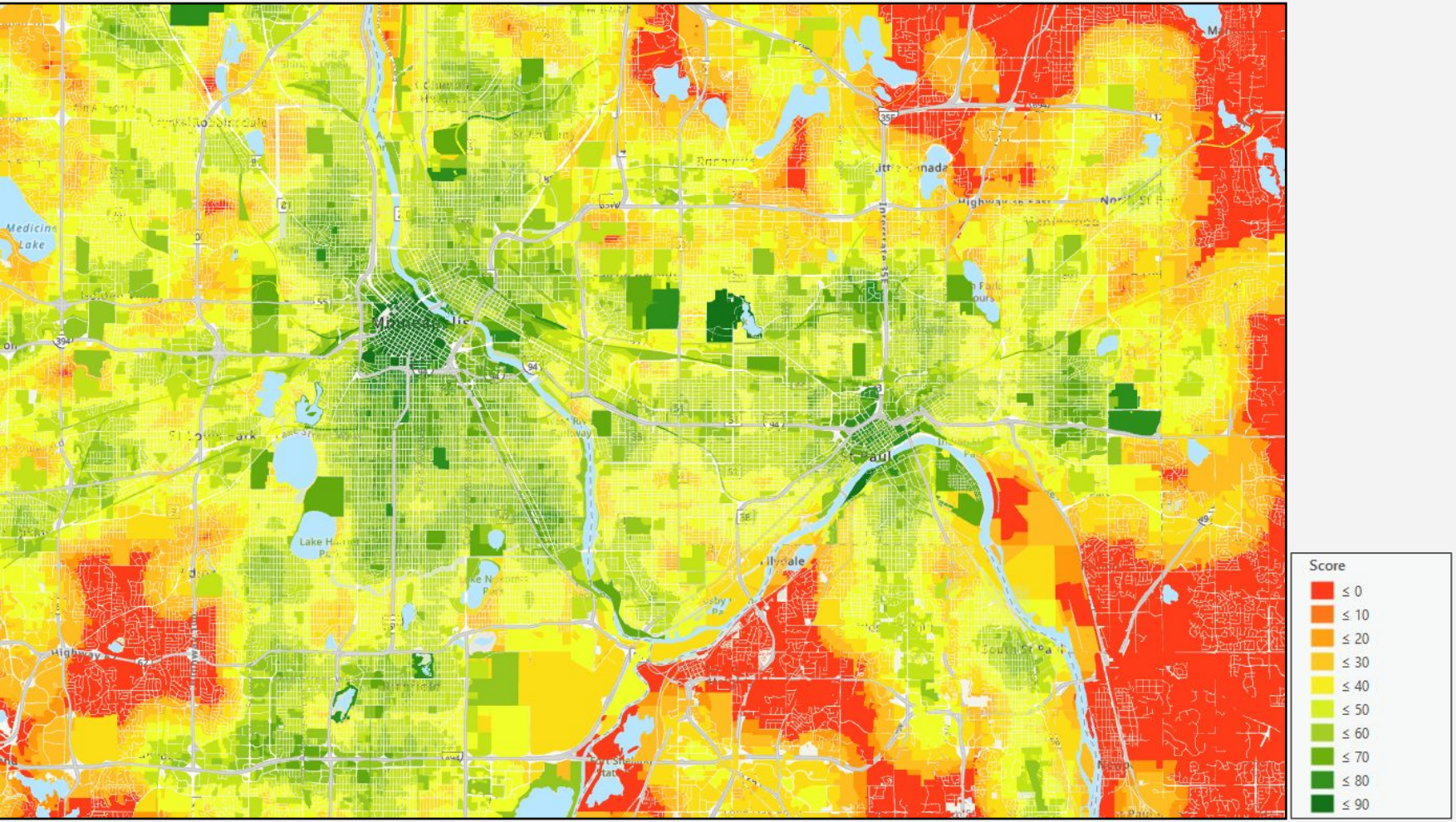
- Polynomial Model
  - Same parameters
  - Num and Distance set to polynomial
  - Other parameters kept linear
  - Order of 3 gives best results
  - ANOVA table to see if R^2 is statistically higher

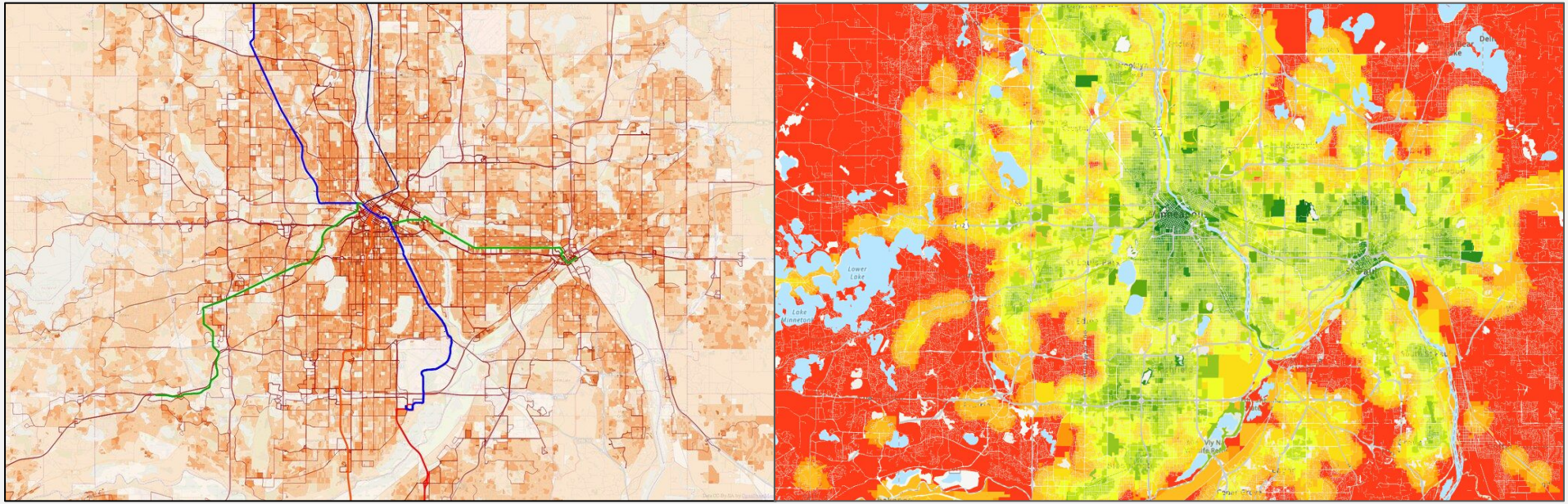


$$Score = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \beta_3 x_{i1}^3 + \beta_4 x_{i2} + \beta_4 x_{i2}^2 + \beta_5 x_{i2}^3 + \beta_6 (x_{i1} * x_{i2}) + \beta_7 x_{i3} + \beta_8 x_{i4} + \beta_9 x_{i5}$$

R^2	RMSE
0.91	0.08


Becker
Big Lake
Monticello
Elk River
Maple Lake
St Michael
Wright
Buffalo
Montrose
Lindstrom
Chisago City
Stillwater
Hudson
Mound
St Paul
Cannon Falls
Lonsdale
Le Sueur
Score
≤ 0
≤ 10
≤ 20
≤ 30
≤ 40
≤ 50
≤ 60
≤ 70
≤ 80
≤ 90


Score
≤ 0
≤ 10
≤ 20
≤ 30
≤ 40
≤ 50
≤ 60
≤ 70
≤ 80
≤ 90



Conclusion

- Polynomial model
  - Better accuracy
  - Less computationally intensive
- Weighted linear combination
  - Much more detailed parameters
  - Higher potential to give more accurate results
- Future analysis
  - More training data
  - Measure accessibility to each transit stop
  - Sample more parcels in accuracy assessment
  - Reduce computational cost