# Appendix A: Calculating, Visualizing and Checking Distances

## Appendix A1: Calculating Distances to Polling Stations from Representative Point

In this project, we use several distance metrics. To ensure the continuation of the previous literature, we calculated the Euclidian distance between the points and with referencing the elements of geospatial coordinates (northing and easting), i.e.

And the Manhattan distance oriented with north, i.e.

It is apparent that the Manhattan distance can be strongly tied to the orientation of the coordinate reference grid. We also use the Maximum Manhattan distance, which can be viewed as the maximum possible distance between two points following city block routing. Geometrically, it is the length of the sides of an equilateral right triangle. Mathematically it is:

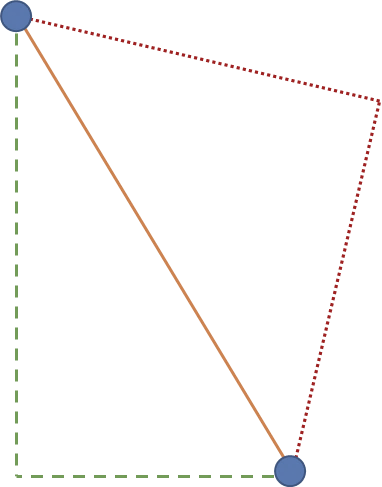
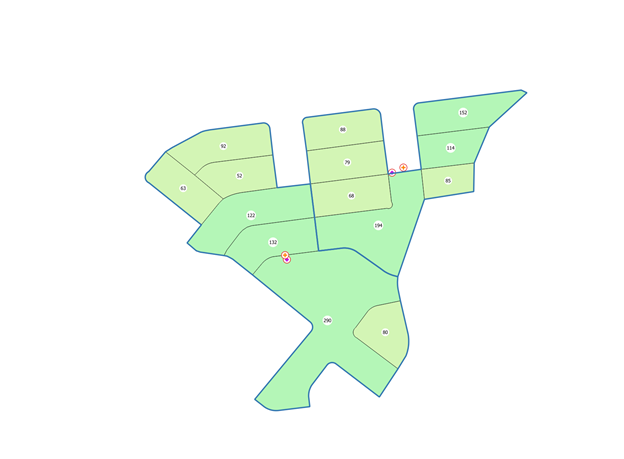
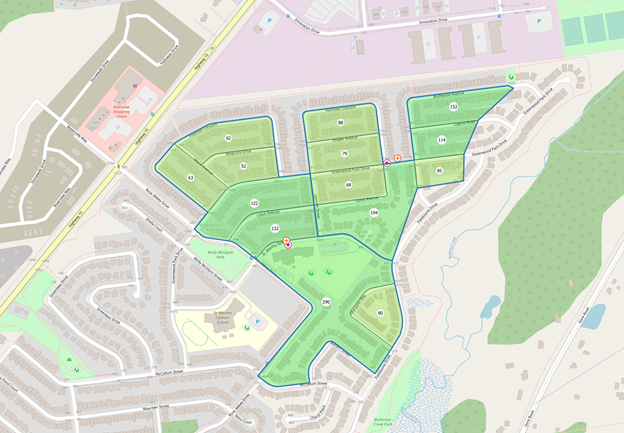


Figure 2 : Euclidian distance (Orange, solid), Manhattan distance (Green, dashed), Maximum Manhattan (Red, dotted)

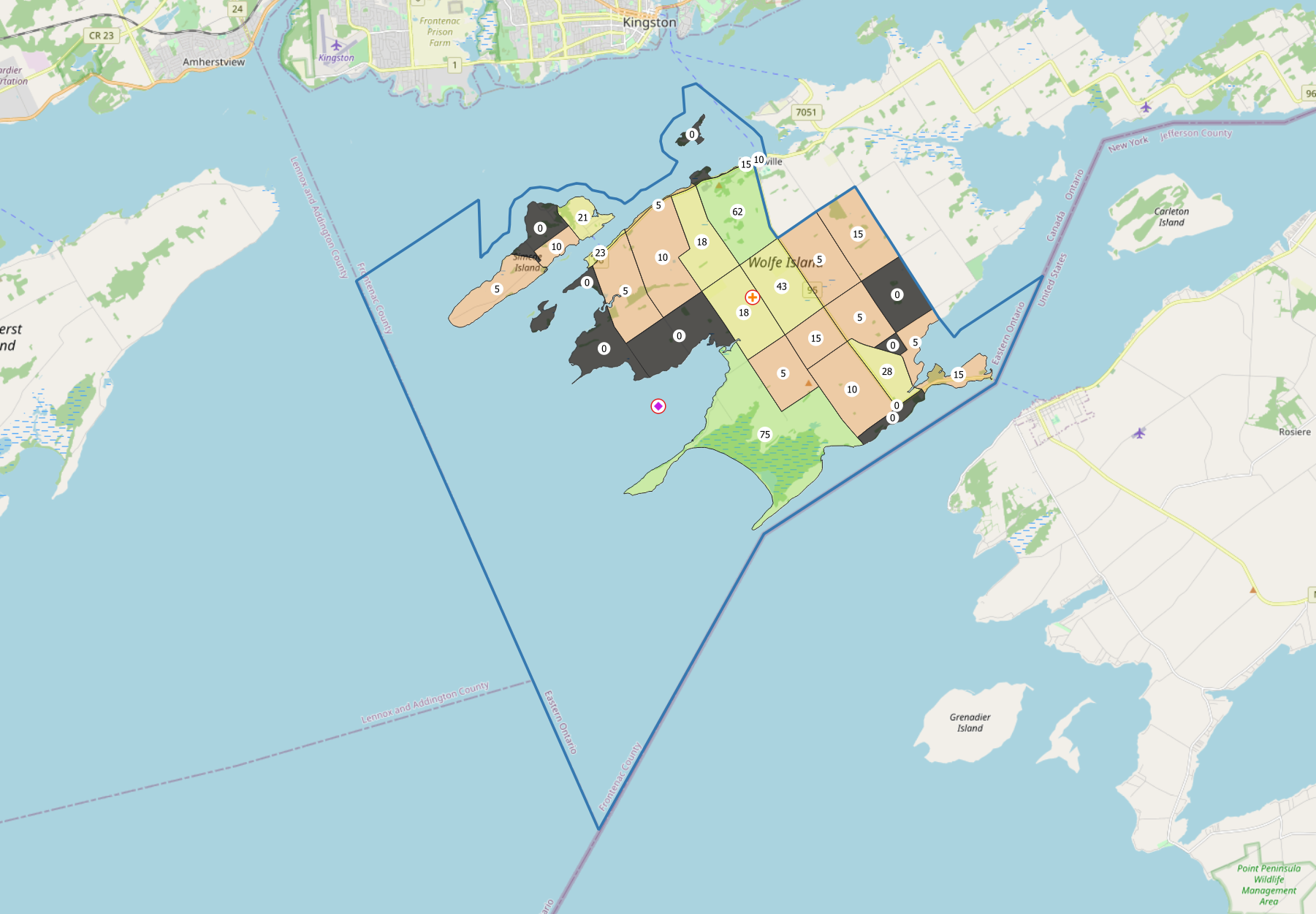
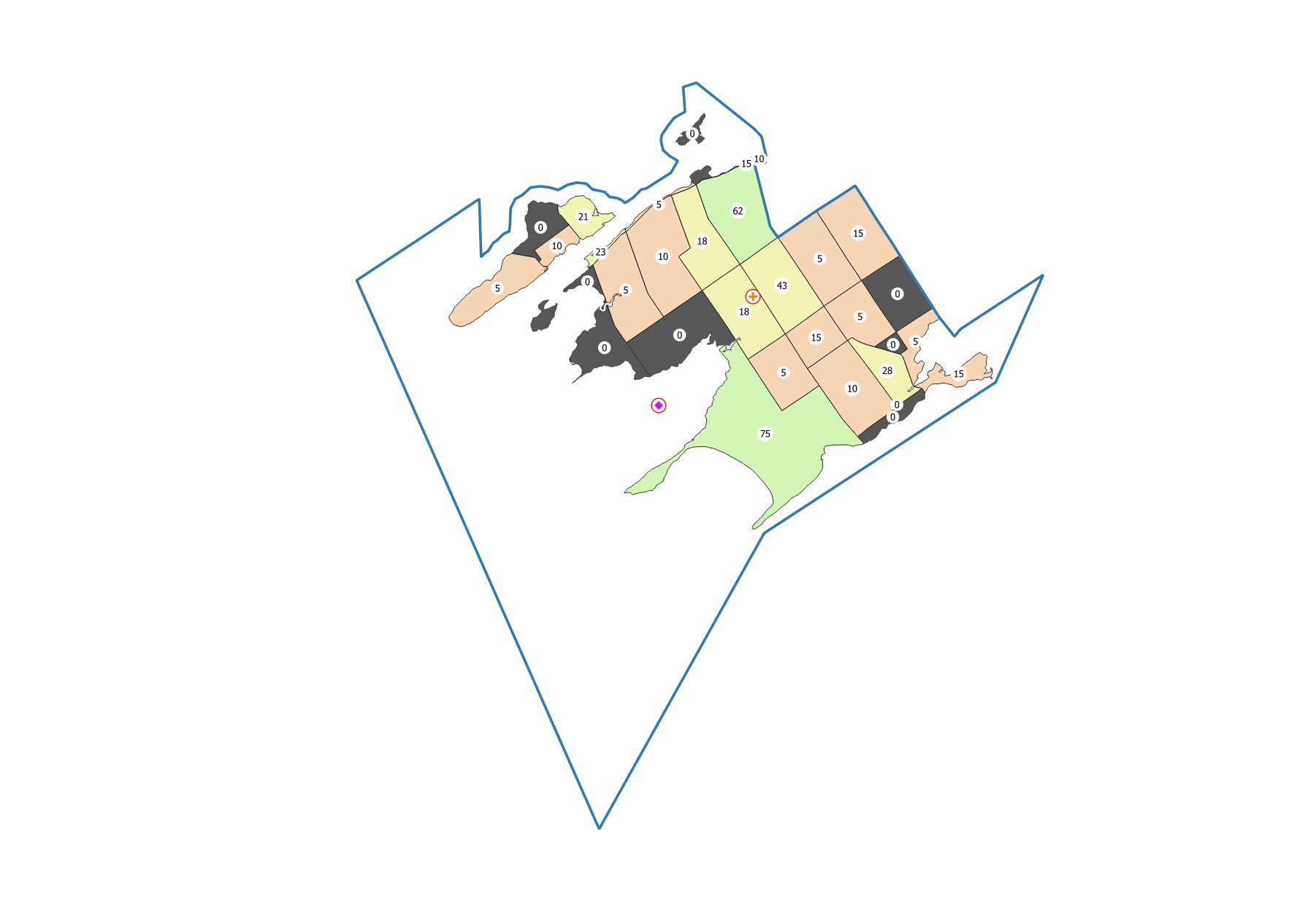
## Appendix A2: Maps

Many centroids of polling divisions are close to the geographic centroid of the polling division, as is often in the case of uniformly dense areas. The following figures depict two polling divisions in the Kingston, Ontario area. These are Polling Divisions 9 and 10 in federal riding 35044. The geographic centroid is represented by a purple diamond, while the mean center of the population is represented by an orange cross. Within each polling division we have the dissemination blocks (DB). They are labeled by the population count of the DB and the label is placed in their geographic centroid.





However, in cases of polling divisions on the coast of the great lakes or in areas with large unpopulated areas, the geographic centroid may be several hundred or thousand meters apart. An example of another polling division in the Kingston area is represented below, which is polling division 212 in federal riding 35044. Again, the geographic centroid is represented by a purple diamond, while the mean center of the population is represented by an orange cross. In this case, these points are further apart with the geographic centroid being in Lake Ontario.



## Appendix A3: Notes on manual correction of distances

After inspection of the coordinates of several polling locations, some pairs of polling locations and polling divisions were erroneously large. When passing an address to the tool to convert addresses to latitude and longitude coordinates, the data reports back a ‘confidence’ score that ArcGIS REST believes it got the address correct, a numeric value of 1 to 10. A select few scored 8 or 9 but are incorrect, in a sense ArcGIS REST was confidently incorrect.

We have spent extra time looking through the data to correct or remove these errors. The procedure for correcting or removing these errors were as follows:

1. We inspected the polling locations where the distance between the polling location and the representative point of the polling division was greater than 10km (Euclidean).
2. We searched the latitude and longitude coordinates in google maps and turned on the satellite imagery
3. Using a mix of map labels, satellite imagery, judgement, regular internet searches, and Google Street View we either accepted the coordinates as correct or rejected the coordinates.
4. If the coordinates were incorrect, we attempted to locate the geospatial coordinates of the polling location (also using searching the address on google maps, satellite imagery, regular internet searches).
5. If we could confidently find the place where voting took place in 2016, we updated the data set to reflect the correct location

After conducting this procedure, we realized many first nations reserves were incorrect. We then applied a similar procedure above to any polling location with the following terms in the city name, site name, or site address: first nation, nation, band office, reserve, friendship center (centre), and indian. We checked all these places regardless of distances.

We checked 842 pairs of polling divisions and polling locations. Of those, 627 were correct, 215 were incorrect. We were able to correct 189 of those.

# Appendix B: Breakdown of Calculations for Census Data

A complete breakdown of calculations is as follows:

For each row entry in the Dissemination Area data, the following information was computed:

|  |  |  |
| --- | --- | --- |
| Column Name | Procedure | Notes |
| CENSUS\_YEAR | Keep | Should be identical for all entries |
| GEO\_CODE (POR) | Replace Geocode | This item was replaced with the polling division identification information, i.e. a geocode with the format “{FedNum}-{PolDivNum}-{PolDivSfx}” |
| GEO\_LEVEL | “POLLDIV” | We overwrote this tag to our new custom geo level tag |
| GNR | Max | Took the maximum global non-response rate for the items in the PolDiv to represent a worst case scenario |
| GNR\_LF | Max | Ibid |
| DATA\_QUALITY\_FLAG | Max | Ibid, represented as an integer value with the minimum 0 being the best and the maximum 9 being the worst |

And the following methods were used to calculate the calculated members

|  |  |  |
| --- | --- | --- |
| Type | Notes | Members |
| Proportional population | Simply, the sum of the proportion of the DA data in the PolDiv | 1, 2, 4 |
| Proportional Dwelling |  |  |
| Weighted Average | A weighted average of the population to the item being calculated |  |

# Appendix C: Histograms

## Figure C1: Distances (Manhattan Meters, in km) from representative point to polling stations, full set of data



## Figure C2: Distances (Manhattan Meters, in km) from representative point to polling stations, bottom 75% of data.



# Appendix D: Robustness Checks with 100% of Data

## Table D1: Predictors of Distance and Travel Times

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|  | Manhattan Meter (kms) | Walking Time (Min) | Driving Time (Min) | Transit Time (Min) | Manhattan Meter (kms) | Walking Time (Min) | Driving Time (Min) | Transit Time (Min) |
| Land Area (100 000s sq kms) | >0.00\*\* | >0.00\*\* | >0.00\*\* | >0.00\*\* | >0.00\*\* | >0.00\*\* | >0.00\*\* | >0.00\*\* |
|  | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Number of Electors | <0.00\*\* | -0.02\*\* | <0.00\*\* | >0.00\*\* | <0.00\*\* | -0.01\*\* | <0.00\*\* | 0.01\*\* |
|  | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Average Age | 0.01\*\* | 0.14\*\* | >0.00 | <0.00 | 0.04\*\* | 0.46\*\* | 0.02\*\* | 0.10\*\* |
|  | 0.00 | 0.04 | 0.00 | 0.01 | 0.00 | 0.04 | 0.00 | 0.01 |
| Percentage Immigrant | -3.43\*\* | -44.03\*\* | -2.61\*\* | -14.93\*\* |  |  |  |  |
|  | 0.08 | 1.38 | 0.10 | 0.35 |  |  |  |  |
| Percentage Indigenous |  |  |  |  | 3.99\*\* | 99.96\*\* | 6.68\*\* | 15.83\*\* |
|  |  |  |  |  | 0.26 | 4.58 | 0.32 | 1.40 |
| Average value of Dwellings (1000s of CAD) | <0.00 | <0.00\* | <0.00\*\* | <0.00\*\* | <0.00\*\* | <0.00\*\* | <0.00\*\* | <0.00\*\* |
|  | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Constant | 2.37\*\* | 31.76\*\* | 3.87\*\* | 14.30\*\* | 0.32\* | 3.01 | 2.13\*\* | 5.56\*\* |
|  | 0.14 | 2.46 | 0.17 | 0.63 | 0.14 | 2.38 | 0.17 | 0.62 |
| N (Polling Divisions) | 24143 | 24140 | 24141 | 22175 | 24143 | 24140 | 24141 | 22175 |
| R2 | 0.10 | 0.06 | 0.04 | 0.09 | 0.04 | 0.04 | 0.03 | 0.03 |

Percent Indigenous and Percent Immigrant not included in the same model due to collinearity.

OLS Regression Models, Standard errors in second row, \* p<0.05, \*\* p<0.01

Data at the polling division level. Case of Ontario is used only. Distances are from a representative point to the polling station.

## Table D2: Distance to the Polling Station and Turnout

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | Turnout | Turnout | Turnout | Turnout |
| Land Area (100 000s sq kms) | >0.00\*\* | >0.00\*\* | >0.00\*\* | >0.00 |
|  | 0.00 | 0.00 | 0.00 | 0.00 |
| Number of Electors | -0.02\*\* | -0.02\*\* | -0.02\*\* | -0.02\*\* |
|  | 0.00 | 0.00 | 0.00 | 0.00 |
| Average Age | 0.16\*\* | 0.17\*\* | 0.17\*\* | 0.17\*\* |
|  | 0.01 | 0.01 | 0.01 | 0.01 |
| Percentage with Post-Secondary Education | 23.20\*\* | 22.72\*\* | 22.93\*\* | 23.76\*\* |
|  | 0.74 | 0.74 | 0.74 | 0.77 |
| Percentage Indigenous | 2.81\* | 3.75\*\* | 4.75\*\* | -6.69\*\* |
|  | 1.30 | 1.29 | 1.29 | 1.60 |
| Percentage Immigrant | -9.30\*\* | -9.13\*\* | -9.06\*\* | -9.58\*\* |
|  | 0.70 | 0.70 | 0.70 | 0.71 |
| Average value of Dwellings (1000s of CAD) | <0.00\*\* | <0.00\*\* | <0.00\*\* | <0.00\*\* |
|  | 0.00 | 0.00 | 0.00 | 0.00 |
| Manhattan Meter (kms) | -0.50\*\* |  |  |  |
|  | 0.04 |  |  |  |
| Squared Term - Manhattan Meter (kms) | 0.01\*\* |  |  |  |
|  | 0.00 |  |  |  |
| Walking time (minutes) |  | <0.00 |  |  |
|  |  | 0.00 |  |  |
| Squared Term - Walking time (minutes) |  | 0.00\* |  |  |
|  |  | 0.00 |  |  |
| Driving Time (minutes) |  |  | -0.13\*\* |  |
|  |  |  | 0.02 |  |
| Squared Term - Driving Time (minutes) |  |  | >0.00\*\* |  |
|  |  |  | 0.00 |  |
| Transit – Depart by 1200 (minutes) |  |  |  | -0.37\*\* |
|  |  |  |  | 0.02 |
| Squared Term - Transit – Depart by 1200 (minutes) |  |  |  | 0.01\*\* |
|  |  |  |  | 0.00 |
| \_cons | 48.19\*\* | 47.75\*\* | 48.07\*\* | 49.92\*\* |
|  | 0.97 | 0.97 | 0.97 | 1.00 |
| N | 24031 | 24037 | 24038 | 22060 |
| R-sq | 0.21 | 0.21 | 0.21 | 0.23 |

OLS Regression Models, Standard errors in second row, \*p<0.05, \*\* p<0.01

Data at the polling division level. Case of Ontario is used only. Distances are from a representative point to the polling station.

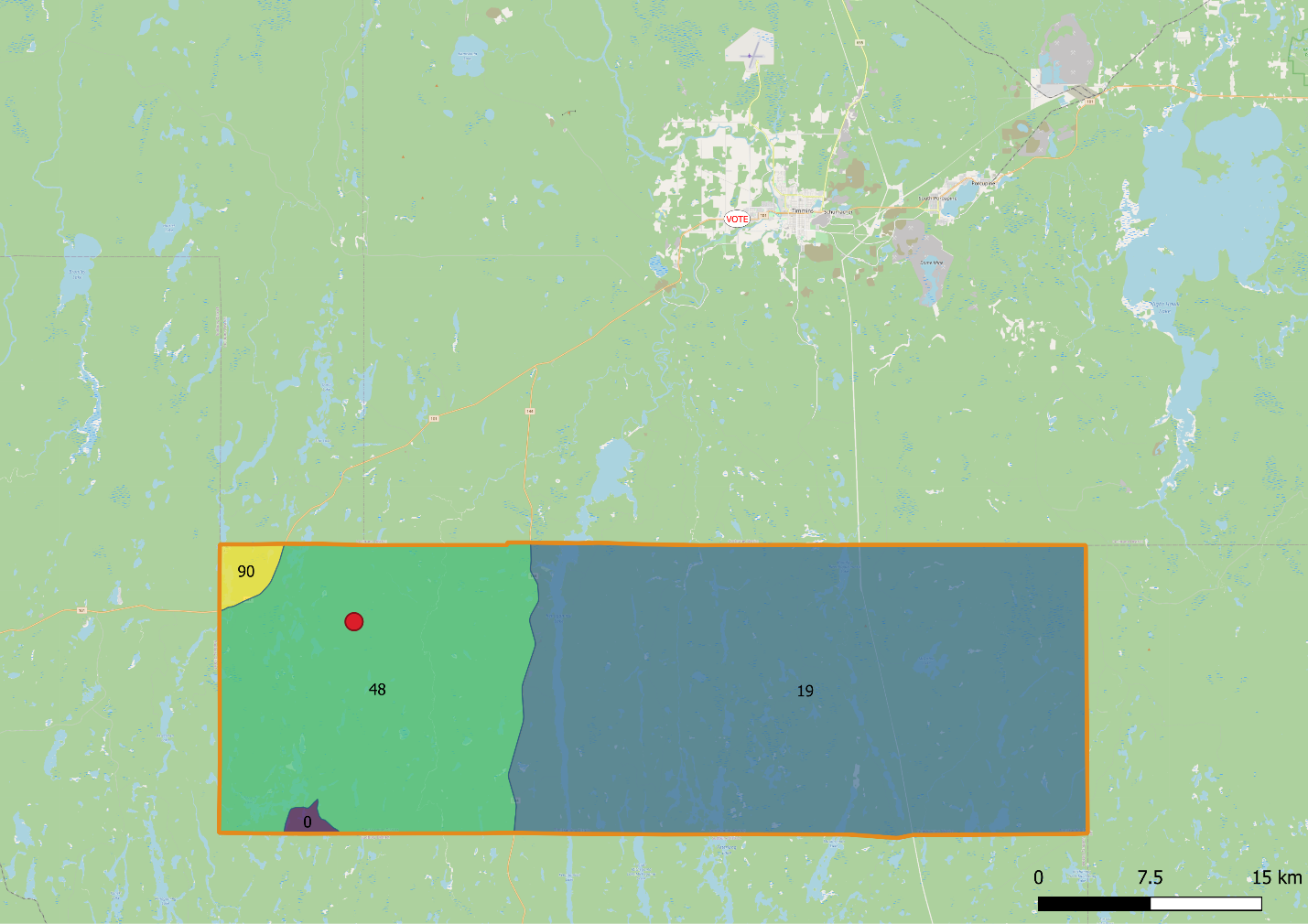
Riding Fixed Effects Included but not reported here.

# Appendix E: Summary of Distance and Travel Time Data

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Number of Observations | Mean Distance/Time (Meters or Seconds) | Mean Distance/Time (Minutes or Kilometres) | Standard Deviation | Minimum Value | 0.25 Percentile | Median | 0.75 Percentile | Maximum Value |
| Euclidean Meter | 25250 | 1837 | 1.84 | 21721 | 0.77 | 305 | 557 | 1059 | 1400000 |
| Manhattan Meter | 25250 | 2323 | 2.32 | 26989 | 0.87 | 383 | 706 | 1348 | 1700000 |
| Maximum Manhattan Meter | 25250 | 2598 | 2.60 | 30718 | 1.10 | 432 | 787 | 1498 | 2000000 |
| Driving Distance | 25207 | 2572 | 2.57 | 29474 | 0.00 | 521 | 922 | 1704 | 2000000 |
| Walking Distance | 25207 | 2414 | 2.41 | 28066 | 0.00 | 459 | 795 | 1481 | 1800000 |
| Driving Time | 25207 | 212 | 3.53 | 1092 | 0.00 | 91 | 142 | 217 | 70326 |
| Walking Time | 25207 | 1765 | 29.41 | 20382 | 0.00 | 338 | 586 | 1097 | 1300000 |
| Transit Arrive by 930 | 23048 | 681 | 11.34 | 617 | 0.00 | 310 | 505 | 806 | 7031 |
| Transit Depart by 1200 | 23045 | 685 | 11.41 | 623 | 0.00 | 309 | 507 | 812 | 6602 |
| Transit Arrive by 2130 | 23045 | 695 | 11.59 | 631 | 0.00 | 312 | 514 | 827 | 9844 |

*Times in Seconds, Distances in Meters, unless otherwise indicated. Full set of data. Case of Ontario only.*

# Appendix F: Extreme Distance



The above image is an example of a Polling Division-location pair that is approximately 37km (Euclidean distance) apart. The orange polygon represents a polling division for the federal riding 35107 (Timmins–James Bay) with a polling division number 119-0 that is just south of the municipality of Timmins, ON. The polling division is comprised of four dissemination blocks that are labeled with their population as reported in the 2016 Canadian census (i.e., 90, 48, 19, and 0 people). The representative point is computed to be at the red dot in the North-west quadrant of the polling division. The people in this polling division vote where it is indicated by a white oval icon labeled with VOTE that is in the city limits of Timmins.

# Appendix G: Full Set of Predictive Margins from Table 2

### Table G1: Walking Time and Turnout (Predictive Margins)



*Predictive Margins with 95% Confidence Intervals, from results in Table 2.*

*Only bottom 75% of data included.*

### Figure G2: Driving Time and Turnout (Predictive Margins)



*Predictive Margins with 95% Confidence Intervals, from results in Table 2.*

*Only bottom 75% of data included.*

### Figure G3: Transit Time and Turnout (Predictive Margins)



*Predictive Margins with 95% Confidence Intervals, from results in Table 2.*

*Only bottom 75% of data included.*