A Spatial Analysis of Property Values and Recreational Access in Calgary Communities: Final Project Report

GEOS 406 - Geospatial Project Foundations

Prof. Adrian Faraguna

GEOS 409 - Data Capture I

Prof: Oscar Andrade Rios

GEOS 410 - Data Manipulation and Transformation

Prof: Jamie Johnston-Stewart

GEOS 418 - GIS Data Modelling

GEOS 419 - GIS Data Analysis and Output

Prof: Carlos Serpas

Group 2

Vincent Chirchir 000954522

Ryden Kungl 000608801

Marcin Zachara 000977141

April 11th, 2025

XX Pages

TABLE OF CONTENTS

Introduction

Background

The purchase of a new home can often carry a promise of a better life. A better life can mean many different things to many people, but there's a growing body of studies that point to the importance of access to green spaces as something that is perceived as beneficial to a healthier, happier life (Kwon et al., 2021). Given this positive association of access to green spaces and health, does this mean that better access to green spaces translates into higher property values? Studies conducted over the years have determined there is an overall consensus that parks increase the property value of nearby housing (Crompton, 2001, Nicholls, 2004). The question already suggests a spatial relationship exists, so applying this spatial analysis to our study area is a perfect opportunity to test the validity of claims like this, as well as offers different options of looking at this relationship.

Geographic Question

We set out with a simple geographic question: "Is the average cost of residential property higher in communities with greater access to parks and leisure centres?" in order to see if there is a socio-economic relationship associated with purchasing housing that is close to city parks and leisure centres in Calgary on a community level.

Data and Methods

Data Collection

Since our area of study consists of the communities of the City of Calgary, the city became our main source of data. From the city's Open Data Portal we were able to obtain the boundaries of Calgary's communities, roads network and pathways.

Park polygons were taken from a Parks Sites dataset and include those green spaces that are maintained by Parks and Open Spaces department of the City of Calgary (City of Calgary, 2025). The municipal parks were then merged with a Provincial Parks dataset to include areas that are under the jurisdiction of the Province but fall within the city limits, specifically to account for Fish Creek Provincial Park (Alberta Parks, 2009).

Our main dataset consisted of the City of Calgary's assessed values of residential, non-residential and farm land properties in the city (Roll_CPID_2025). The dataset includes all properties that are on The City of Calgary annual property tax assessment rolls, so all parcels registered with the Alberta Land Titles Office with unique numbers assigned a roll number for properties and business identifier for businesses, which correspond to physical locations (City of Calgary, 2025). The important thing to remember is that these assessed values are not market values and that they were produced for tax purposes, however, they are based on neighbourhoods sales comparisons through a mass appraisal method, so the market value is taken into consideration, but we are dealing with more stable numbers, not influenced by current market trends and unique characteristics of particular houses (Findyourcalgary.ca). This makes the dataset more suitable for our analysis than the dataset based on market values from a particular month would be for example. The data available is up to date with the present year. Every year the City assesses the property values as of July 1 of the previous year through a "mass evaluation process that determines a general value of how much your home could sell for under typical conditions as compared to others" (City of Calgary, 2025). The current year data is updated weekly; The 2024 data was last updated on Dec.1 , 2024; The 2023 data was last updated on Dec. 1, 2023, etc.

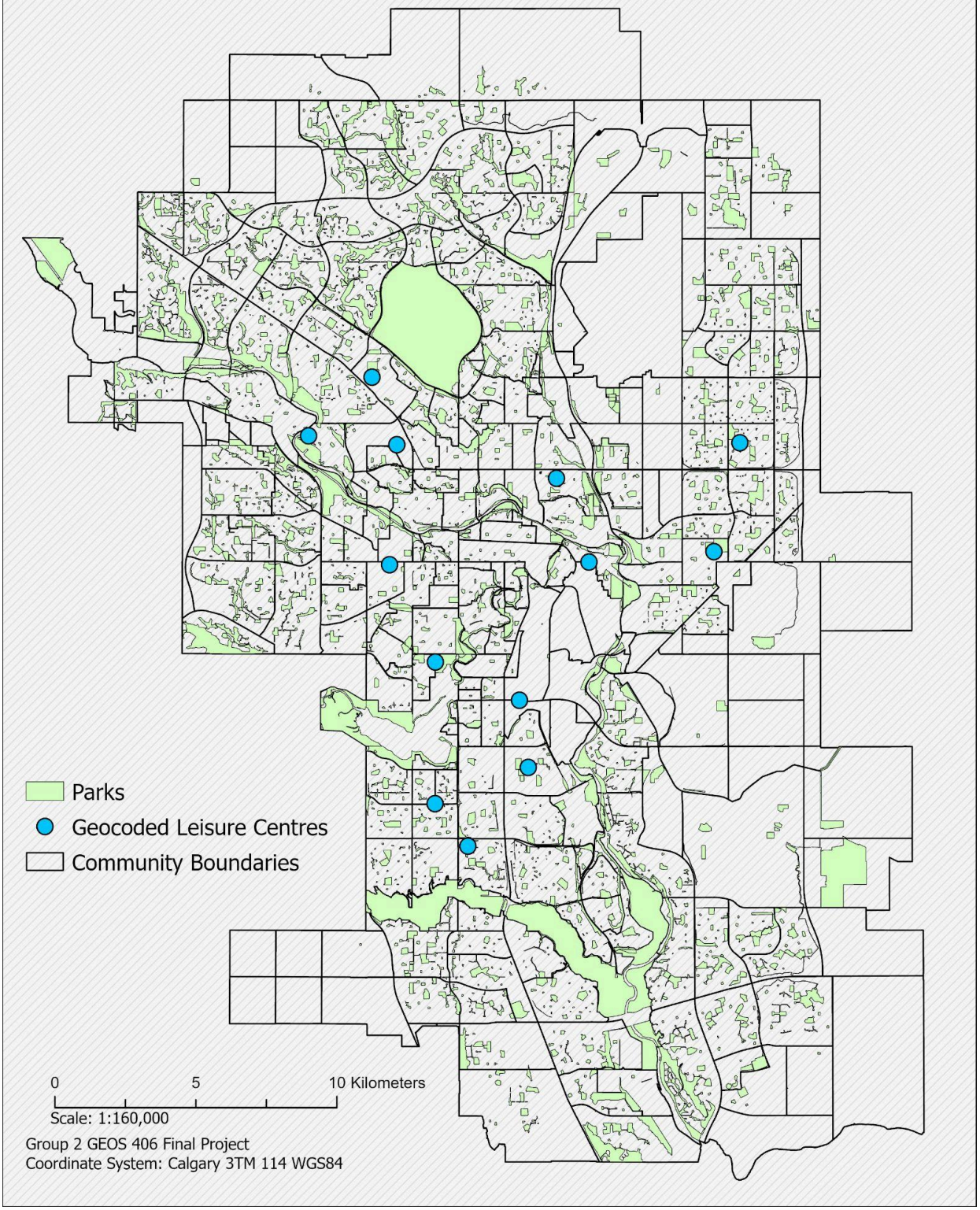
Coordinate Systems and Projections

All our datasets came in a shapefile format which we then exported to our Group2_Final_Project geodatabase using custom Calgary 3TM -114 WGS1984 Projected Coordinate System to be consistent with the projection used by the City of Calgary.

Geocoding (GEOS 409)

We geocoded leisure centres as points to see if, together with park proximity, they work to influence the properties' values. The addresses of leisure centres were taken from the list on the City of Calgary website and then transferred into an Excel table, which was brought into the project table using the Excel to Table tool. Locator was created based on the City of Calgary shapefile CALGIS_TRANSNET_CENTERLINE that was brought into our geodatabase as Calgary_Roadnet feature class and based on the fields that allowed for the matching of building numbers and street names with Local Extra High precision. The Geocode Table tool matched all of the locations as point feature class, Leisure_Centres_addr_Geocoded, which were then merged with park centroids into network analysis's Service Areas calculations of 5-minute walking distance coverage per neighbourhood.

Parks and Leisure Centres in Calgary Communities, 2025

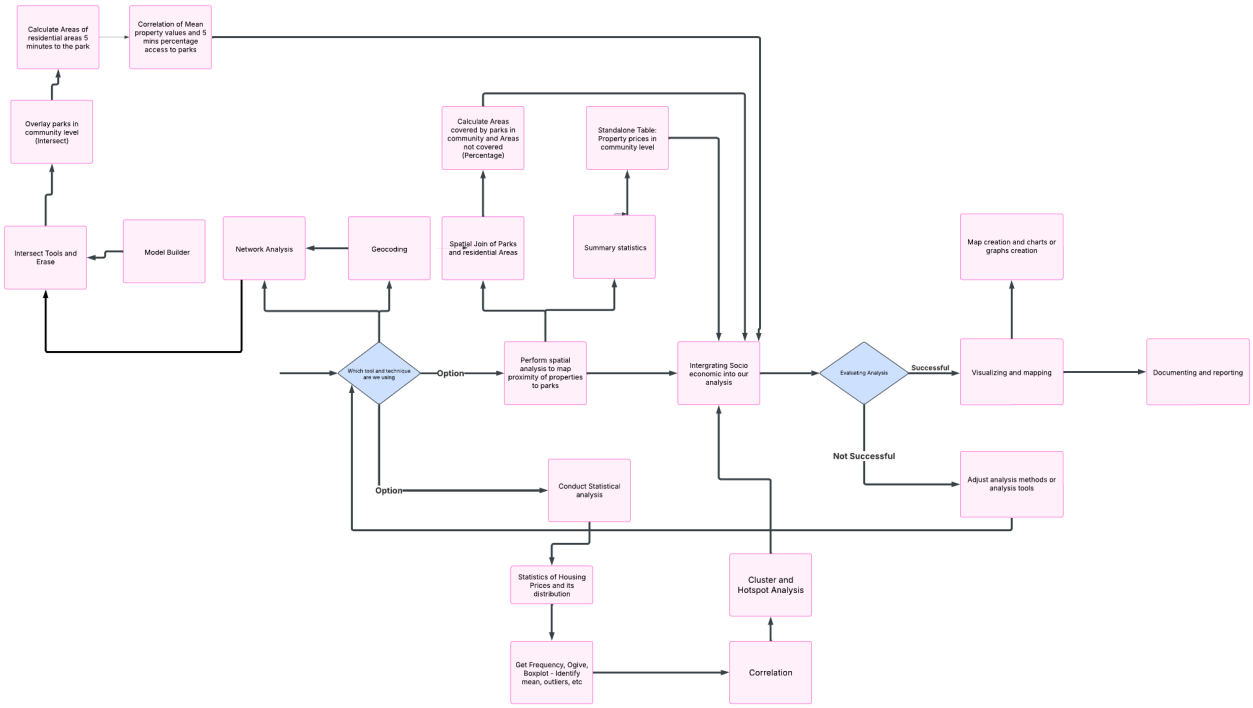


Methodology

This phase focused on integrating spatial analysis and statistical analysis to understand the relationship between our two variables. i.e, average normalized residential property prices and 5 minutes walking access to public parks and leisure centres.

This methodology was designed to ensure a step-by-step process from data input, data analysis, data management, and data output and finally to documenting reports and producing map layouts

All analyses, from data preparation, statistical analysis, and spatial analysis, integrated our socio-economic problem for evaluation and reporting



Spatial and Statistical Analysis (GEOS 418 & 419)

We set out by querying the property values table ROLL_CPID_2025 to select only Residential properties using field Assessment_Class_Description. We further queried the table using the field Land_Use_Designation to get rid of the properties such as strip malls, parking lots, industrial and commercial buildings which are not classified as residential properties, as these were outside of the scope of our analysis. We exported this selection to our geodatabase as Community_Properties_Values to perform statistical analysis on the properties of interest to us.

Statistical Pattern in Residential House Prices in Calgary

There are 191 residential communities in Calgary with a combined total house price value of $ 5,989,262,500.00 and an average value of $1.61 billion. The deviation of house prices from the average is $1.75 billion, indicating that, on average, the house price in each community differs from the overall mean by about $1.75 billion. With such a standard deviation of around $1.75 billion, this indicates significant variability in house prices in Calgary. The high value suggests that the house prices in various communities are widely dispersed from the average. This means some communities may have significantly higher or lower prices.

	STATISTICAL VALUES
MEAN	$1,599,955,202.71
MEDIAN	$1,278,073,040.00
MODE	#N/A
MAX	$5,989,262,500.00
MIN	$87,429,000.00
RANGE	$5,901,833,500.00
VARIANCE	$1,380,437,954,509,550,000.00
STD	$1,174,920,403.48
1ST QUARTILE	$744,046,497.00
2ND QUARTILE	$1,278,073,040.00
3RD QUARTILE	$2,149,959,784.50
IQR	$1,405,913,287.50
LOWER LIMIT	-$1,364,823,434.25
UPPER LIMIT	$4,258,829,715.75

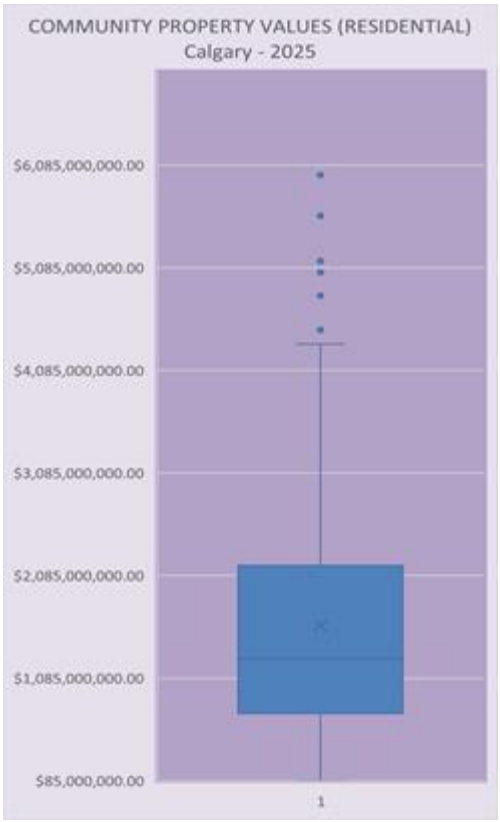


Figure X: Summary Statistics

Frequency Distribution

A graph of the frequency distribution for the community property values appears below. As you can see from the graph, 36 communities fall under $930,548,071 - $1,353 107,607. Because most communities fall under that category, the average property value is quite high ($1.6 billion). This is a positively skewed distribution.

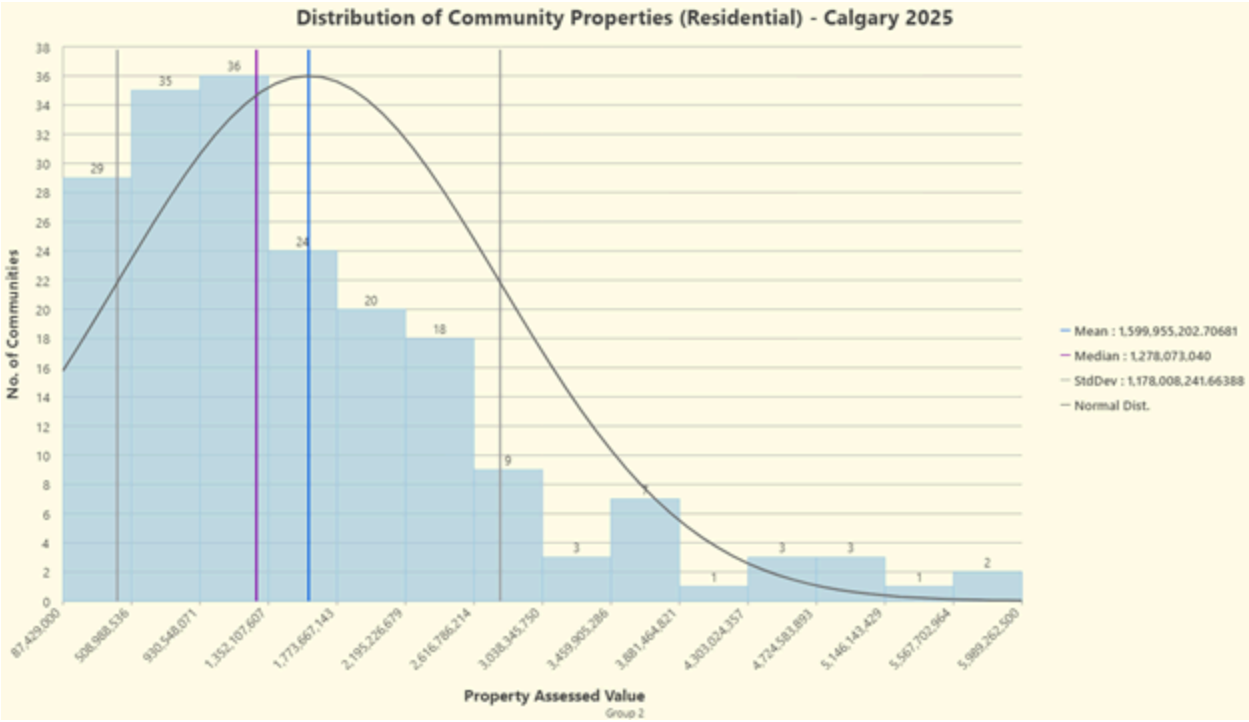


Figure 3: Frequency Distribution of community property values

Now that we know our distribution is clustered towards the higher end, with fewer communities with higher property values, we may want to assume that those 36 communities are close to a park. Additional questions may be raised:

What are the chances that our house prices per community are random?

Where are the clusters of house prices with unusually high prices?

Which community has unusually high house prices?

Using the Summary Statistics tool, we then obtained a standalone table Community_Property_Clean_Statistics with several calculations on the field Re-assessed_value like sum total and central tendency (average, median, minimum and maximum values, range) by community using the Comm_Code field as a case field.

After exporting the selection from the Community_Boundaries feature, including only Residential communities, we named it Residential_Boundaries, and then we performed

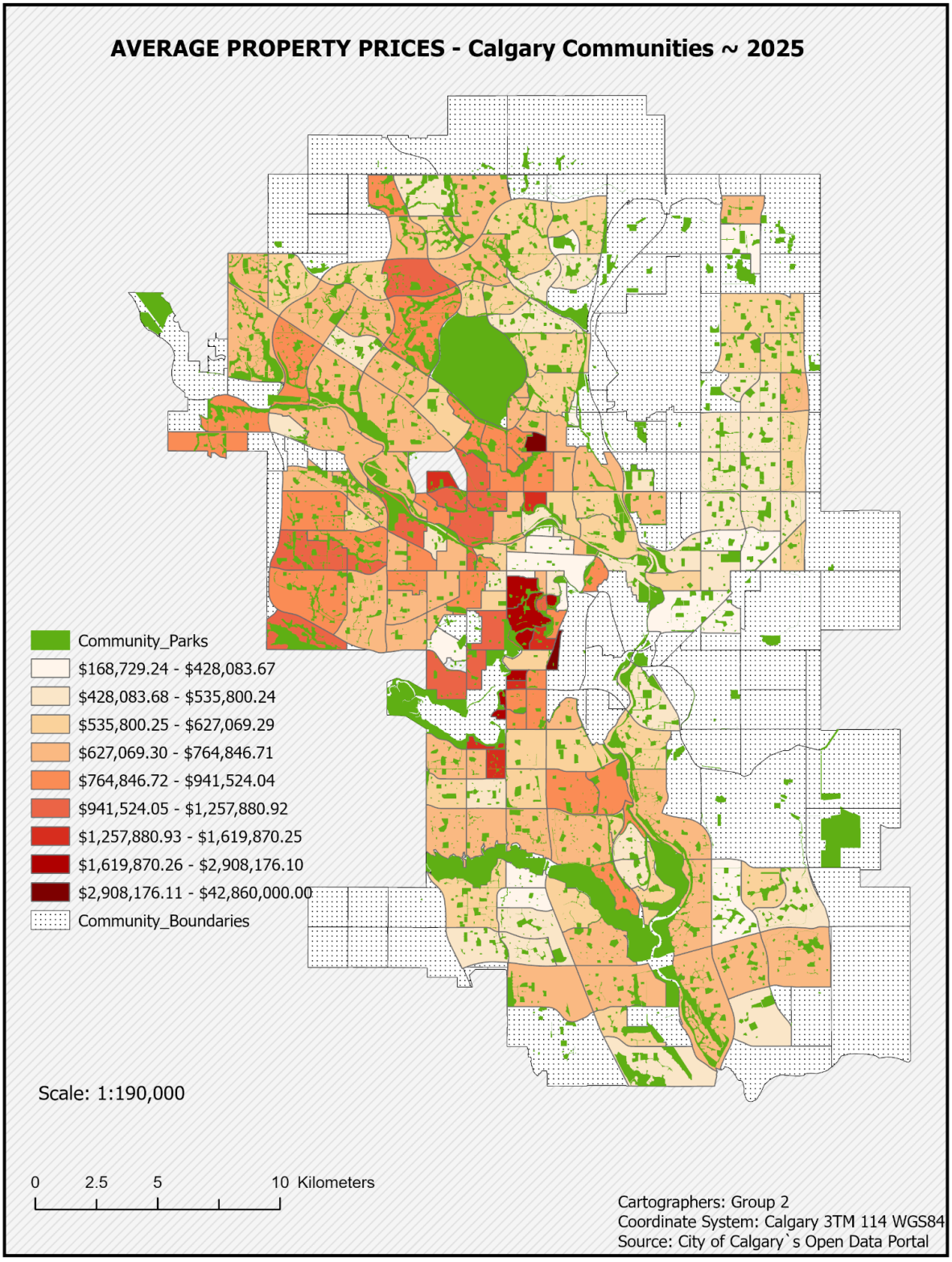
a one-to-one cardinality join with the summarized statistics table to be able to map the calculated statistics.

We repeated this procedure with the new added numeric field: NORM_BY_AREA to our property selection table, using Double data type. We created this field in order to normalize our property values by area, to correct for the fact that some of our properties are large polygons like apartment buildings and can distort the reality of how expensive housing can be in a given community. The additional field then was calculated to give us the price of all properties per square meter of its parcel area (the field Re_assessed_value was divided by Land_Size_Sm).

Another standalone table was created, Community_Property_Avg_Normalized, with the series of statistics using this new normalized value by community: sum total and central tendency (average, median, minimum and maximum values, range). This table, in turn, was joined with the Residential_Communities attribute table.

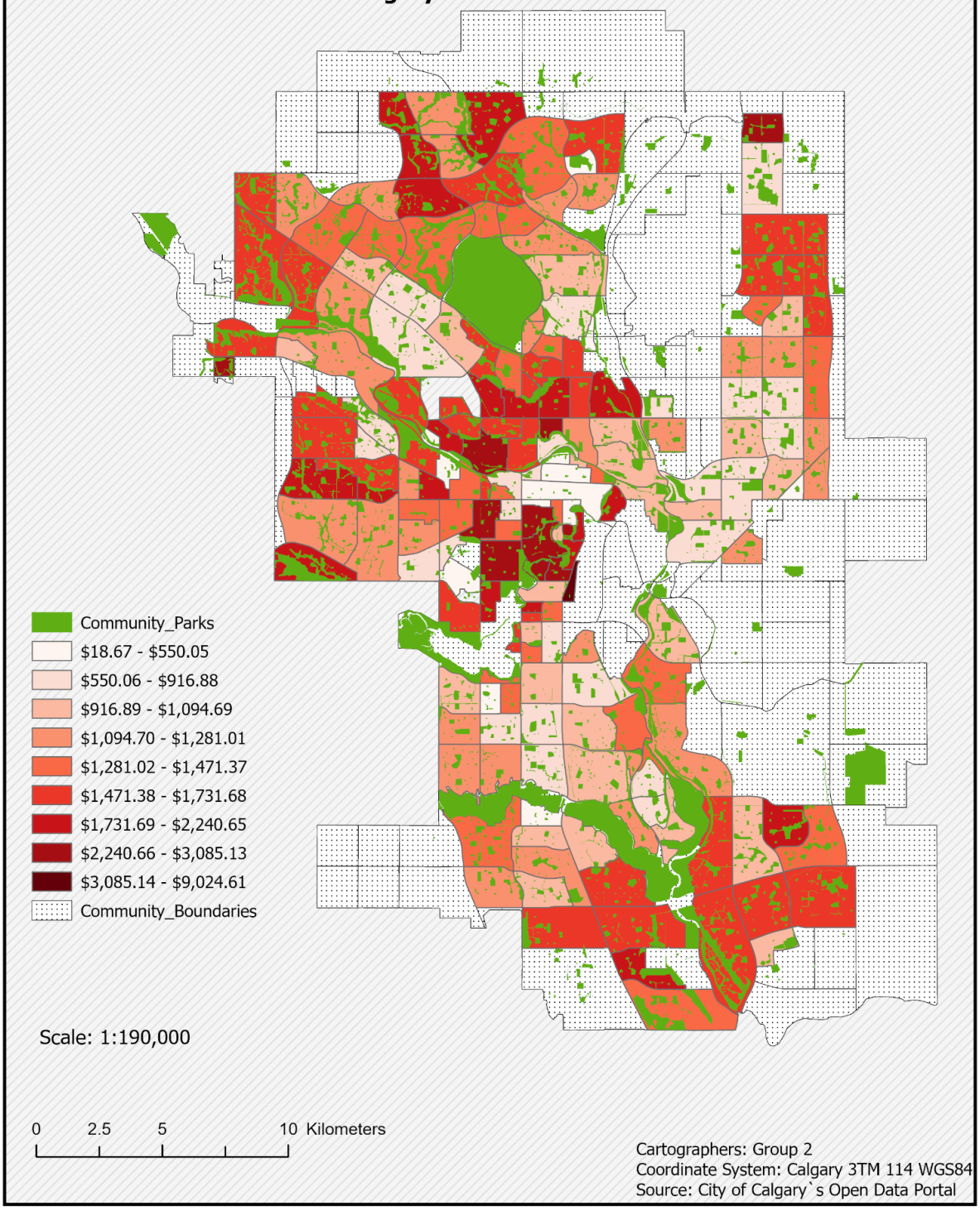
The results are two choropleth maps below showing the average prices of properties by communities, using graduated colours for respective Mean values for each community. The Sturges formula was used to determine the 9 classes with the Natural Breaks (Jenks) method.

Average property prices by community



Average normalized property prices choropleth map

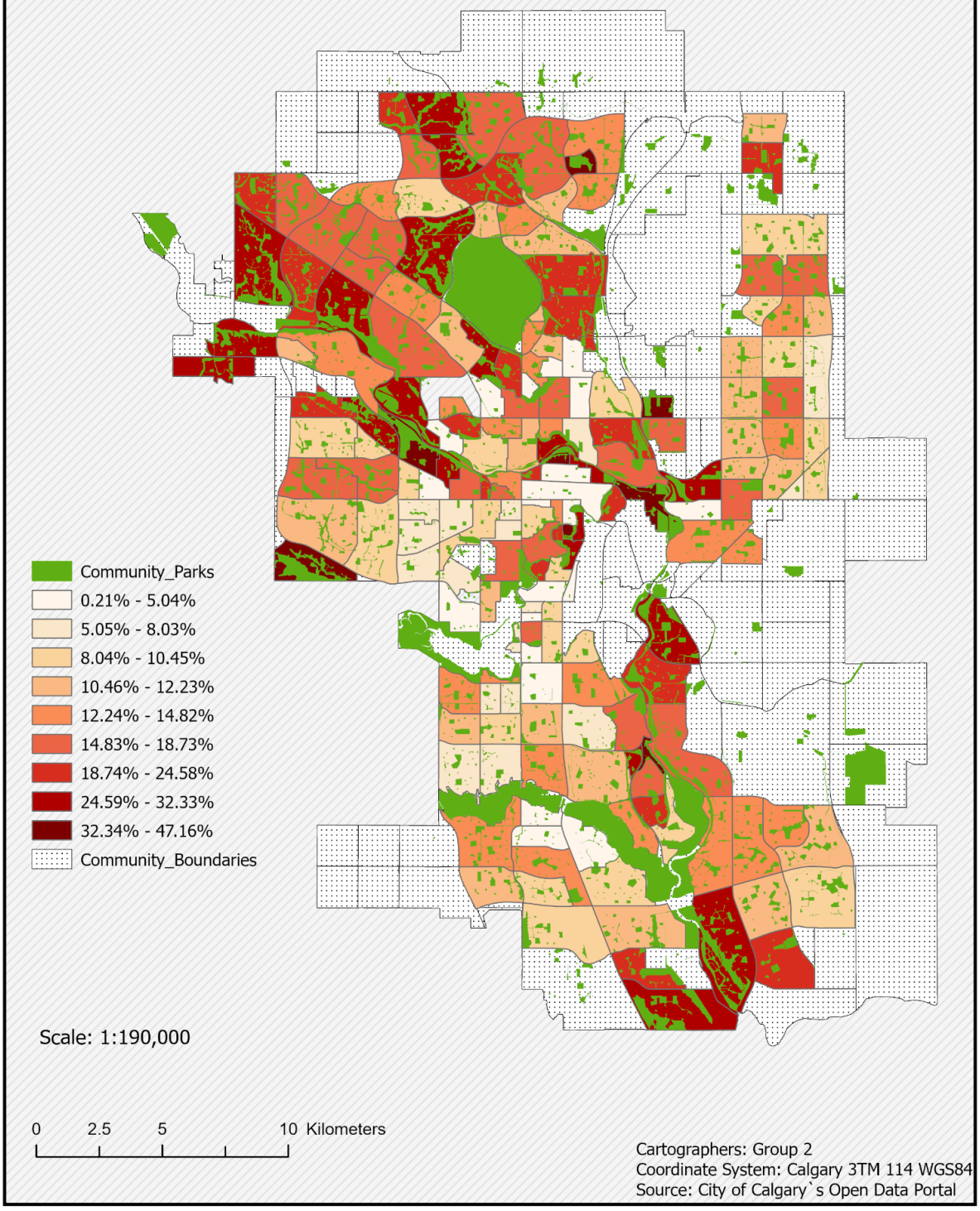
AVERAGE NORMALIZED HOUSE PRICES
Calgary Communities ~ 2025



The next step was to add a numeric Double type field to Public Parks and calculate the parks' polygon areas in square kilometers. The same was done with communities to have their areas in square kilometers. Public parks in Calgary stretch across many communities, so the Intersect tool was used to have parks polygons fit into the communities. The resulting Community_Parks_Area dataset had the parks divided by communities, but the areas had to be re-calculated. Another field was added to calculate what percentage each park polygon constitutes within the community. Another standalone table was obtained through the Summary Statistics tool, Community_Park_Statistics1, where we were able to obtain total percentages of all parkland area by each community, with the comm_code field used again as a case field.

So that we can map parkland area percentages per community, this standalone table was joined to the Residential_Boundaries feature class with the resulting choropleth map as a visual representation below:

Community Park Distribution (Percentage) - Calgary Communities ~ 2025

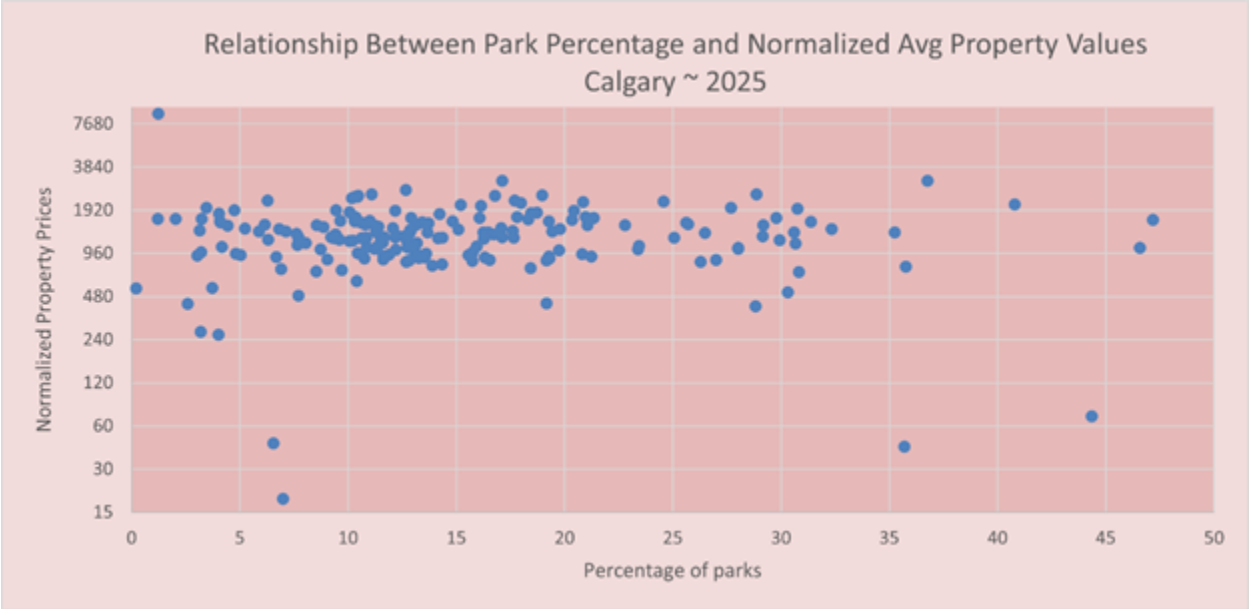


Correlation

Our goal is to study whether park percentages influence property values at the community level and whether the percentage of parks in a community affects the price of residential properties. To determine that, we are going to investigate the correlation between the two variables. We suspected that the higher the percentage of parks in a community, the higher the prices of properties.

A new feature class was created that includes fields with normalized average property values with park area percentages by community through another join operation so that correlation could be shown in a graph. The graph below indicates that there's no correlation between the percentage of park area within the community and the average normalized property value. The calculated correlation coefficient is -0.03778. The communities covered with more parkland are not more expensive compared to other communities as a whole. This is because, as we could see earlier from park statistics, Calgary communities generally have a pretty good park area coverage across the city.

The scatter plot graph below suggests that there is small to no shared variance between the two variables.



The correlation is -0.03778, suggesting that all the variance in each variable is independent of the other variable. The magnitude of the relationship between the two variables is indicated by a value r of, which is a weak negative correlation.

Let us look at the **coefficient of determination** to understand how much of our variance in normalized average house prices can be explained by park percentage.

Determined by squaring the correlation r.

$$R = -0.03778$$

$$R^2 = 0.001426$$

This indicates that the park percentage only explains 0.14% of the variance in normalized property prices. The remaining 99.86% is likely to be explained by other factors.

The coefficient of determination (R^2) value of **0.001426** suggests that park percentage has an extremely minimal effect on normalized average property prices. It explains almost none of the variability in property prices, reinforcing the conclusion that park percentage is not a significant predictor of property prices in this dataset.

Let us look at whether our correlation is **statistically significant**. We will begin with a null hypothesis that:

1. There is no relationship between property prices and park percentage.
2. There is a relationship between the two variables.

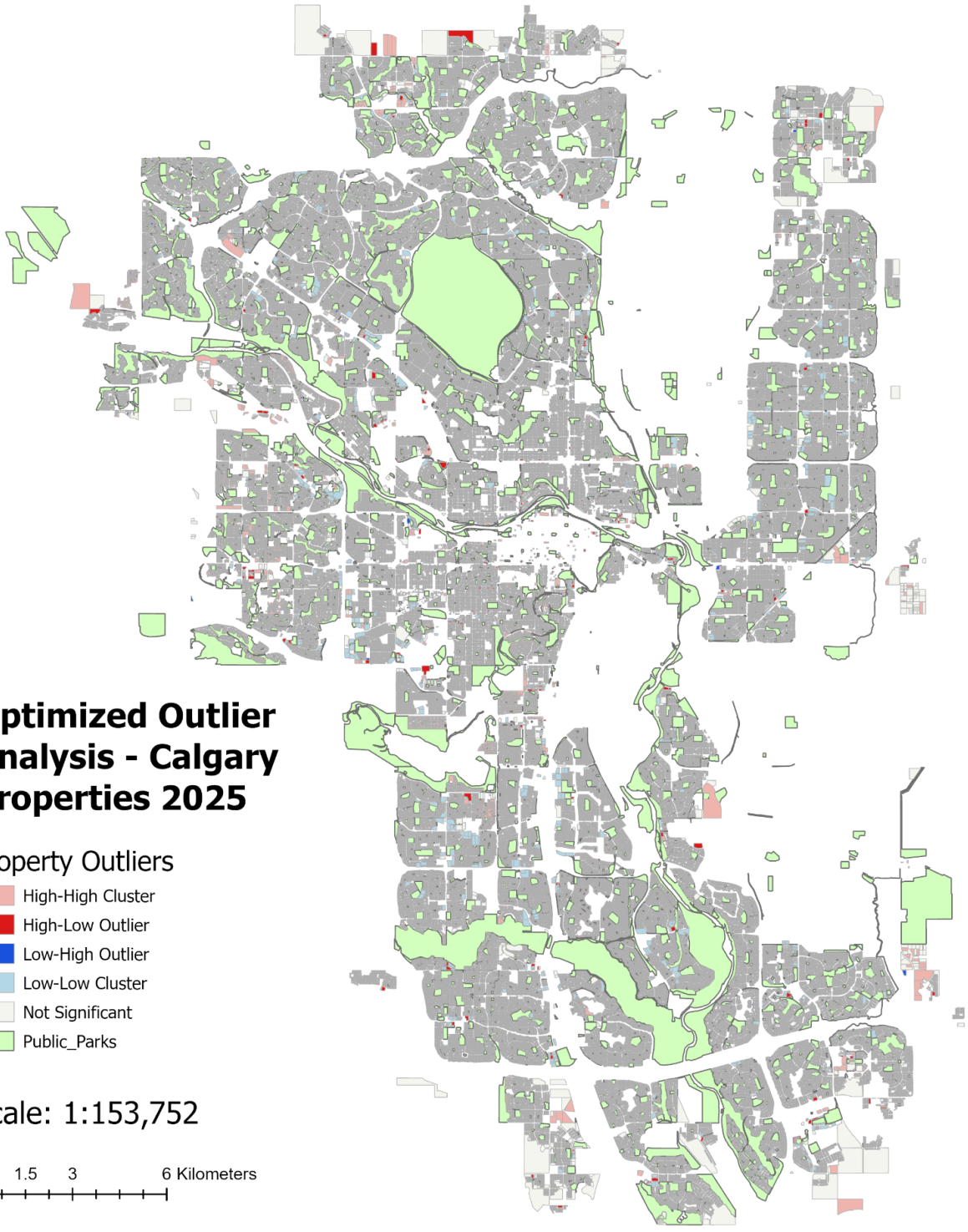
This formula was used to calculate the t values to test if there is a significant difference between the two variables. A larger T value will indicate a strong relationship, and a smaller value indicates a weak relationship.

$$t = (r)\sqrt{\frac{N-2}{1-r^2}}$$

Our t value of -0.52111 is low, which shows that the correlation is weak and not significant.

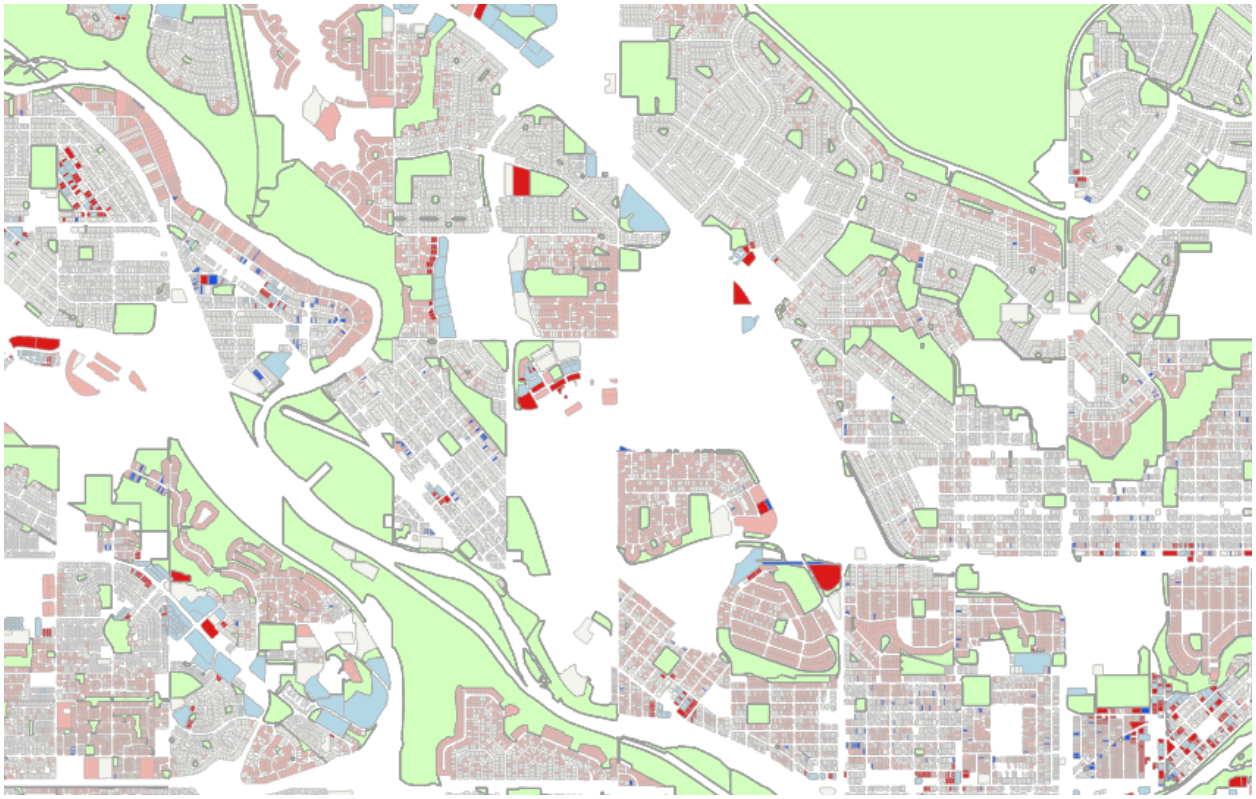
Optimized Hotspot and Optimized Outlier Analysis

Optimized Hot Spot Analysis and Optimized Outlier Analysis were performed on selected property values, before and after the price normalization by area, with Distance Band of 94 meters and then 64 meters for normalized values, then performed again with Hot Spot Analysis (Getis-Ord Gi*) and Cluster and Outlier Analysis (Anselin Local Moran's I) with Distance Band of 150 meters. Although the Outlier analysis suggests in certain areas high price values for properties bordering green spaces or water like in the example below, it is difficult to talk about patterns, and the results are generally hard to interpret. Still, it led us to test our question through network analysis by looking at the properties that are closer to parks rather than to stay on the community level.

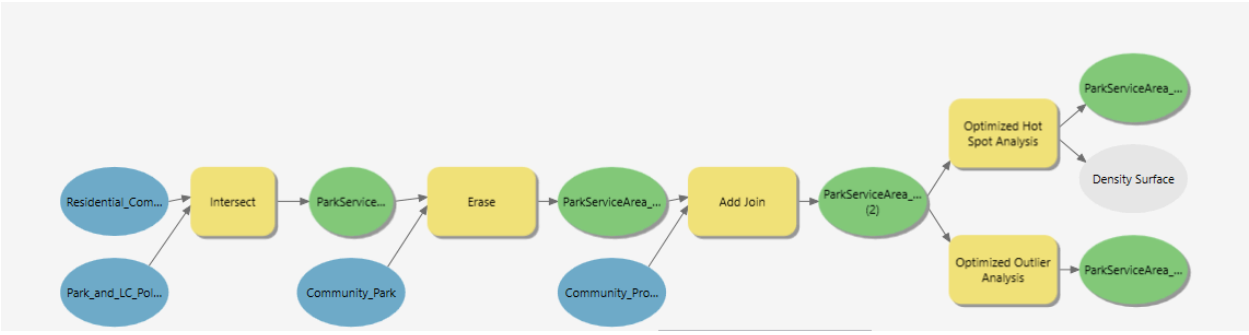


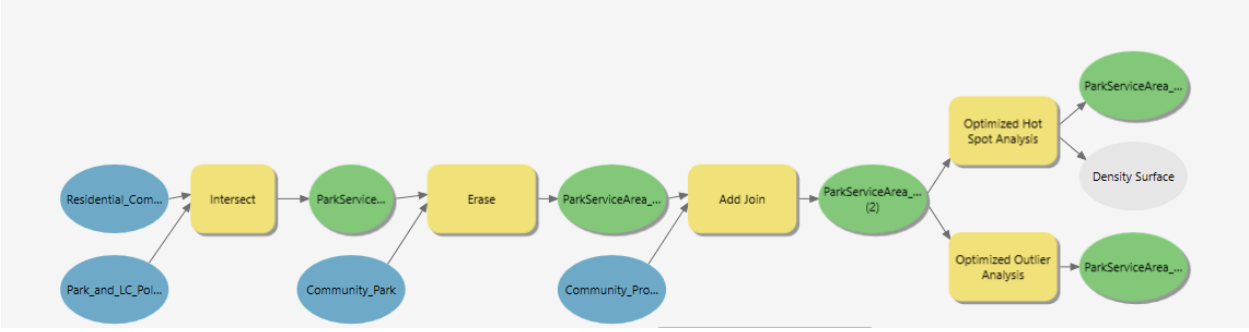
Cartographers: Group 2
Coordinate System: Calgary 3TM 114 WGS84

Below is a zoomed image of the fragment from the map above:



Model for creating Optimized Hot spot analysis and optimized outlier analysis (GEOS410)



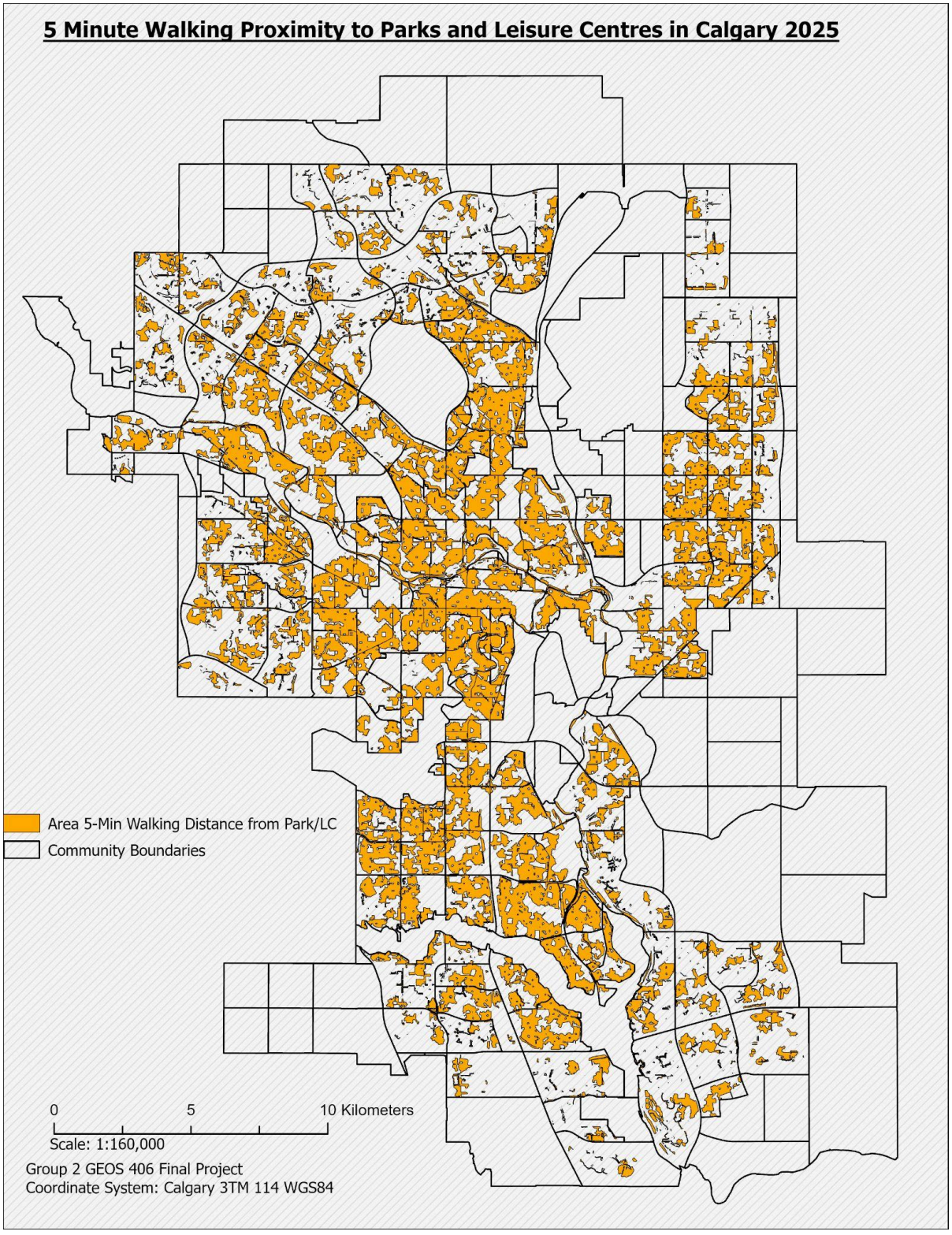


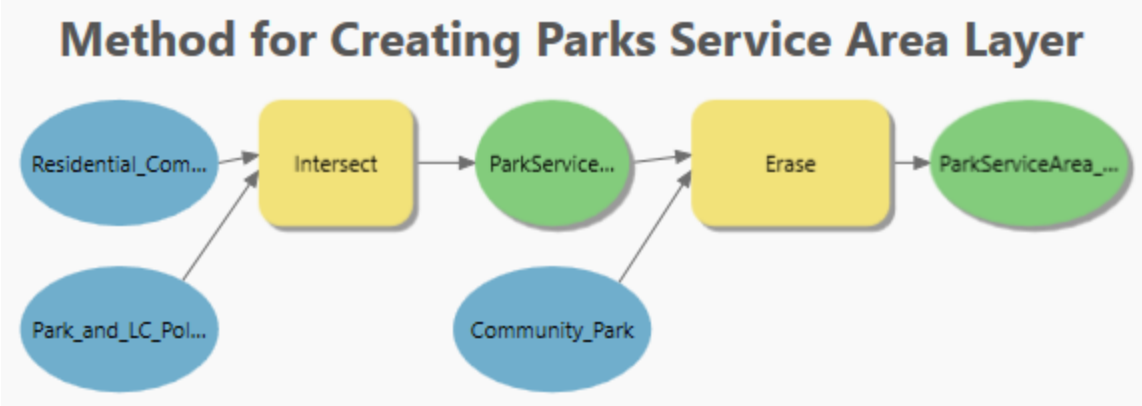
Network Analysis (GEOS 409)

Analysis of 5-minute proximity to Parks

The level of spatial proximity between residential areas and the park is measured using the 5-minute distance to the park centroid. The service area under network analysis was used to create 5-minute polygon areas between every community. The map below shows the service area map distribution of communities that have access to the park in 5 minutes. This also includes 5 minutes to access a neighbouring community park. From the figure, it can be shown that most of the Calgary residential areas have more access to parks within 5 minutes with a concentrated distribution at the centre quadrant.

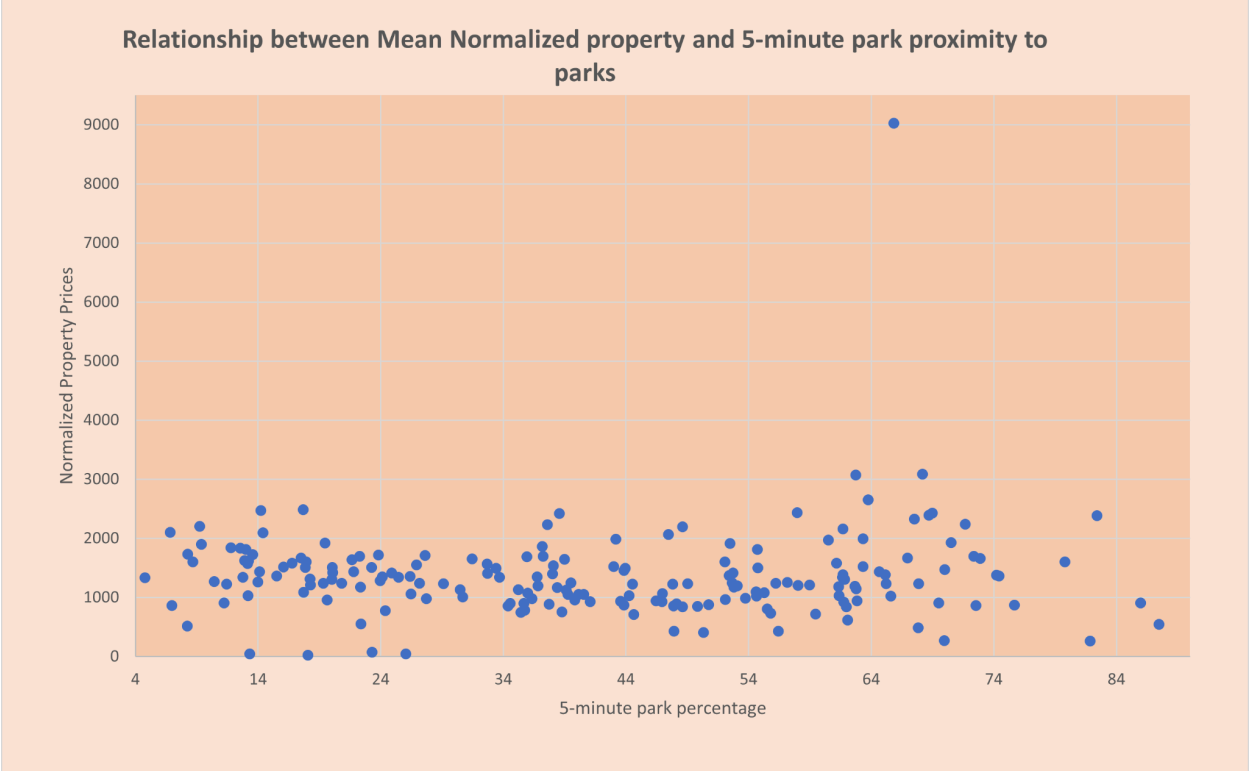
5 Minute Walking Proximity to Parks and Leisure Centres in Calgary 2025





Correlation of overall proximity to parks in 5 minutes

Again, correlation was done after the analysis to quantify relationship between park access and property values using the Pearson method between percentage of each community within a 5 minutes walk tp public parks and the mean normalized property valuest. The results were a weak positive correlation of **0.055573**, meaning there is no linear relationship between the two variables.

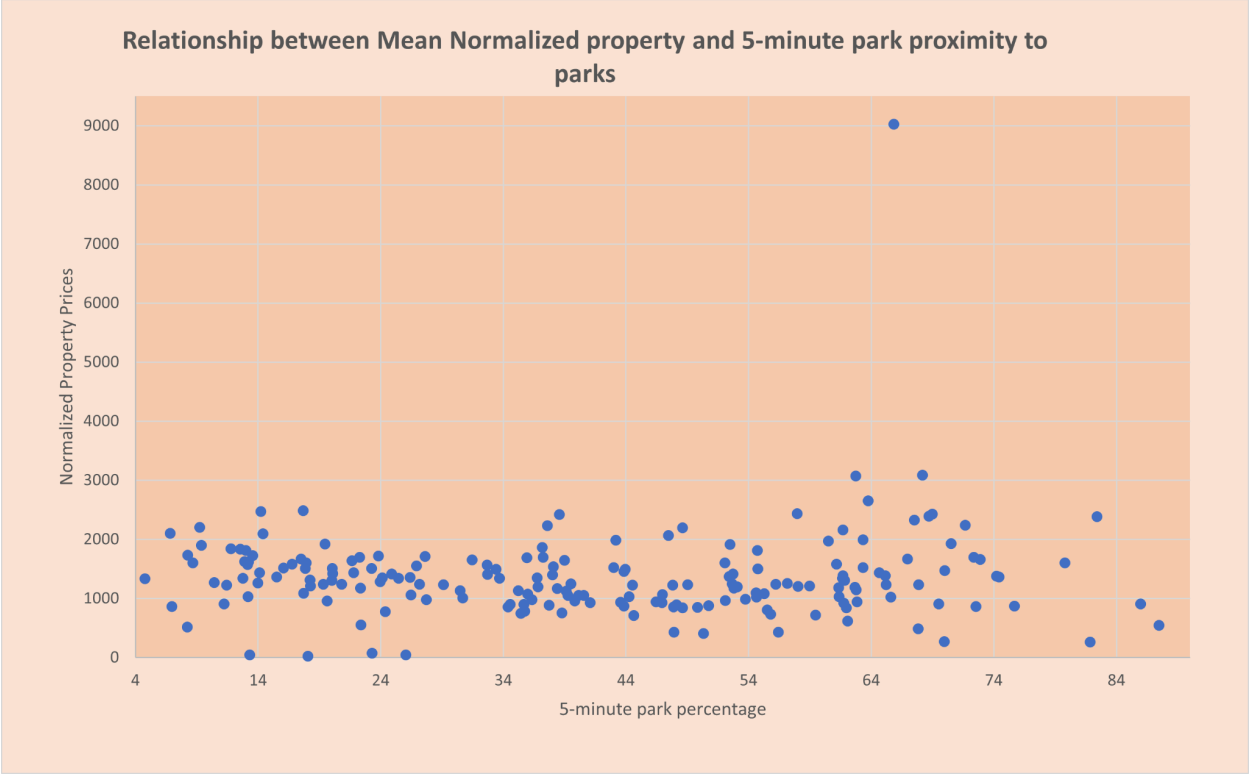


The coefficient determination of **0.003088** means that only **0.31%** of the variation in the percentage of area within 5 minutes to parks can be explained by the mean normalized property values. This is a low value indicating that the mean normalized property values are responsible by other factors.

These findings support support the conclusion that proximity to parks is not a variable that determines the property values. Other variables play more influence in property prices

Correlation of overall proximity to parks in 5 minutes

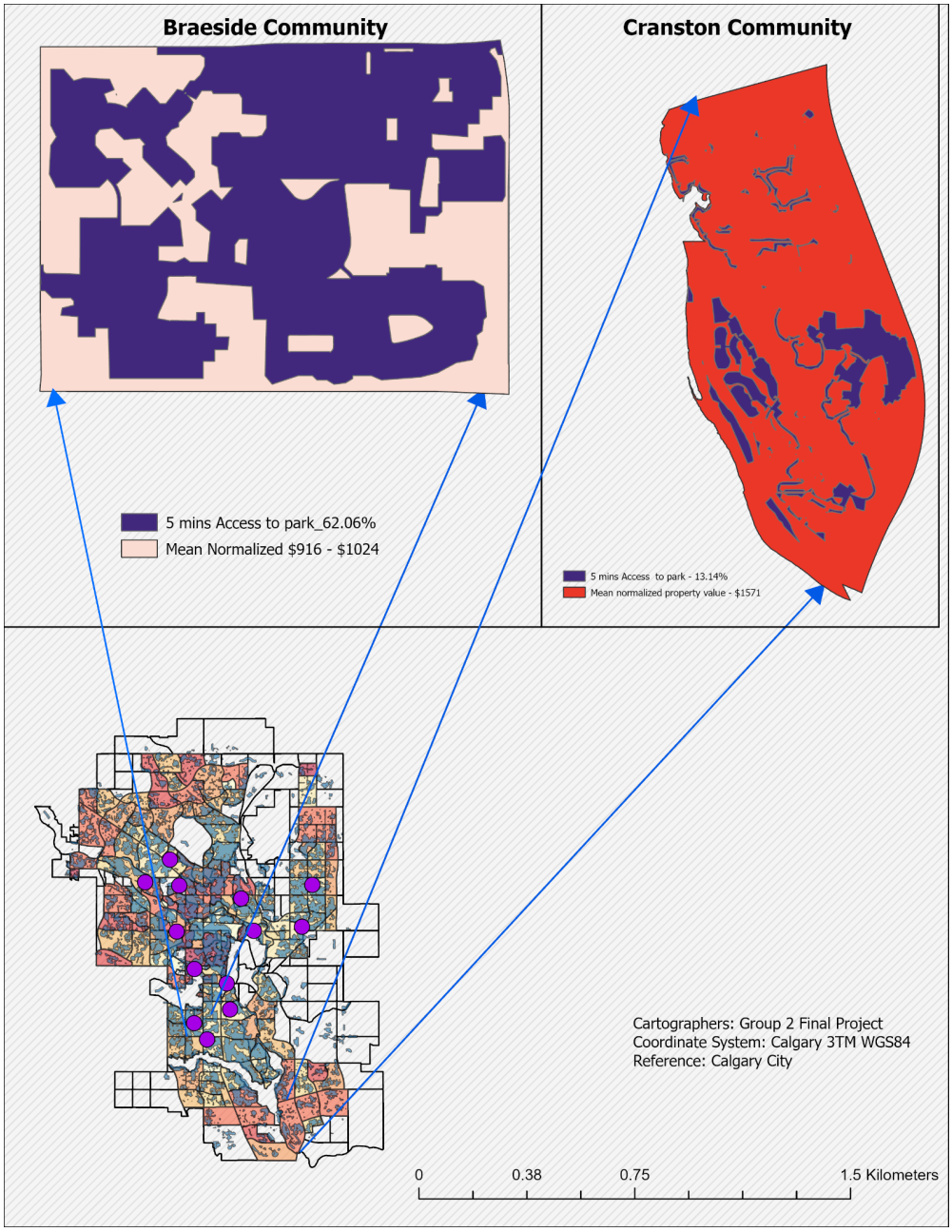
Again, correlation was done after the analysis to quantify relationship between park access and property values using the Pearson method between percentage of each community within a 5 minutes walk tp public parks and the mean normalized property valuest.The results were a weak positive correlation of **0.055573**, meaning there is no linear relationship between the two variables.



The coefficient determination of **0.003088** means that only **0.31%** of the variation in the percentage of area within 5 minutes to parks can be explained by the mean normalized property values. This is a low value indicating that the mean normalized property values are responsible by other factors.
This findings support support the conclusion that proximity to parks while it is good, is not a variable that determines the property values. Other variables play more influence in property prices.

Conclusion

From the figure below, residential areas prices with high proximity are not majorly dependent on parks. For example, the community of Cranston is 25.6% covered by parks, and 13.14% of the community has access to parks in 5 minutes. The same community has a mean normalized property price of $1571.80 per square ft. The Braeside community, by contrast, has 62.06% access to parks in 5 minutes and has a mean normalized property price of $613.40 sqft.
This shows that higher park accessibility does not correspond to higher property values. This is supported by statistical analysis showing a very weak correlation between the two variables.



There was no correlation found for any of the statistical methods we applied in answering our geographic question. Despite the findings of other research in our literature review, there must be conditions different in Calgary than the study areas of those papers. Parks must have a wide distribution in Calgary, and thus no matter where one chooses to purchase property, they will always be in a relative close proximity to a park. Since no correlation arises between parks proximity and property values, property values must be influence largely by other contributing factors in Calgary.

References:

City of Calgary (2025), www.calgary.ca

City of Calgary (n.d.) *Indoor Pools, Fitness & Leisure Centres* https://www.calgary.ca/rec-locations/pools.html

City of Calgary (2025) *Park_Sites* (Shapefile) https://data.calgary.ca/Recreation-and-Culture/Parks-Sites/i9fu-gjqj

City of Calgary (2025) *Parcel_with_Roll_2025* [Shapefile] retrieved from ArcGIS Online

City of Calgary (2025) *Community Boundaries* [Shapefile] https://data.calgary.ca/Base-Maps/Community-Boundaries/ab7m-fwn6

City of Calgary. (2024). *Calgary Equity Index – Methodology Documentation*. Retrieved from https://www.calgary.ca/research/calgary-equity-index.html

Crompton, J. L. (2001). The impact of parks on property values: A review of the empirical evidence. *Journal of Leisure Research*; Urbana Vol. 33, Iss. 1, pp. 1-31. Retrieved from https://www.proquest.com/docview/

Kwon et al. (2021) Urban green space and happiness in developed countries. EPJ Data Science. 10, 28. https://doi.org/10.1140/epjds/s13688-021-00278-7

Meng, Y. & Malczewski, J. (2015). A GIS-based Multicriteria Decision Making Approach for Evaluating Accessibility to Public Parks in Calgary, Alberta, *Human Geographies* Vol. 9 (No. 1), pp. 29-41. Retrieved from https://www.researchgate.net

National Recreation and Park Association (NRPA). (2020). The impact of proximate proportion of park-like space and views on housing prices. *Parks & Recreation Magazine*. Retrieved from https://www.nrpa.org/parks-recreation-magazine/

Nicholls, S. (2004). Measuring the impact of parks on property values: New research shows that green spaces increase the value of nearby housing. (Research Update). *Parks & Recreation* (Arlington, Va.), vol. 39, no. 3, National Recreation and Park Association, pp. 24-32.

O'Sullivan, A. (2012) *Urban Economics* (8th ed.). McGraw-Hill Irwin

Alberta Parks (2009) *Downloadable Data Sets*. [Shapefile] https://www.albertaparks.ca/albertaparksca/library/downloadable-data-sets/

Findyourcalgary.ca

Appendices

Appendix A: Geodatabase Schema

GROUP 2 GEODATABASE SCHEMA

Residential Boundaries Schema	
Community Code	Text (3)
Community Name	Text (50)
Sector Name	Text (14)
Community Area (SQKM)	Double
Area_5_Minutes_Park_SQKM	Double
SUM_PARK_SQKM	Double
Park_Perc	Double
Perc_5_Minute_Park	Double
Sum_Norm_By_Area	Double
Mean_Norm_Area	Double
Sum_Re_Assessed_Value	Double
Mean_Re_Assessed_Value	Double
Sum_Land_Size_SQKM	Double
Mean_Land_Size_SQKM	Double

Parks Database	
Class	Text (20)
Comm_Code	Text(3)
Name	Text(50)
Sector_Name	Text(14)
Comm_Are_SQKM	Double
FID_Public_Parks	Long
Park_Name	Text(100)
Park_SQKM	Double
Park_SQKM_Intersect	Double

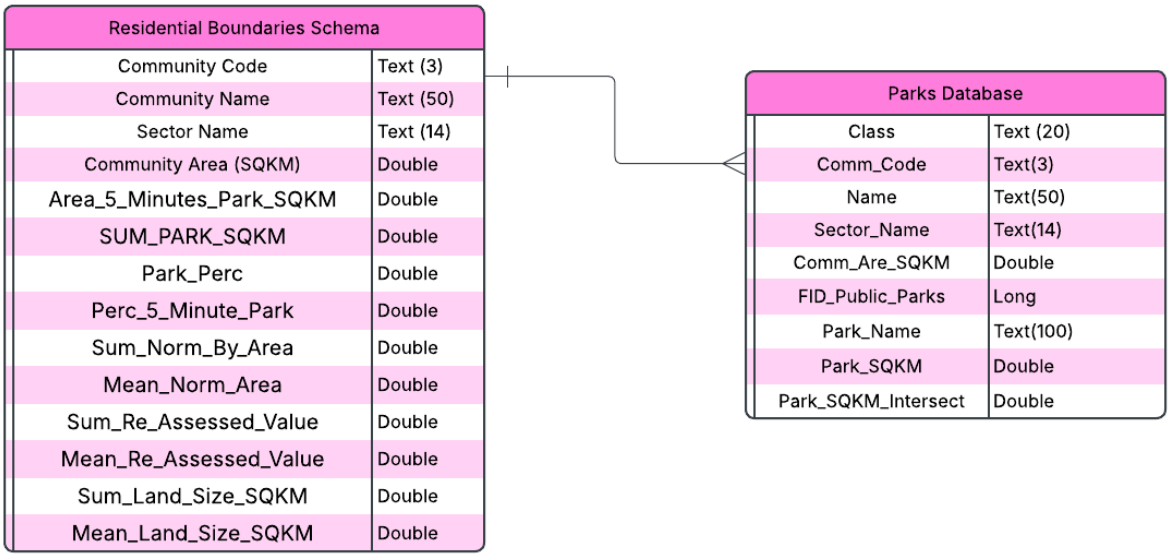
Walkable Roads Pathways	
Road Name	Text (52)
Street Type	Text (4)
Sector Name	Text (14)
Road Category	Text (46)
Speed Limit KMH	Double
Kilometres	Double

Park5MinsWalk	
Comm_Code	Text(3)
Comm_Name	Text(50)
FID_Park and LC_Polygons	Text(46)
Name	Text(
FromBreak	Double
ToBreak	Double
Area_5_Minutes_Park_SQM	Double

Leisure Centres	
Name	Text(60)
Adress	Text(50)
Match Address	Text(50)
Address Type	Text (50)
Y	Text(1)

Residential Property Values	
Community Code	Text (3)
Community Name	Text (50)
Unique Key	Text(90)
Roll Year	Long
Roll Number	Text(10)
CPID	Double
Address	Text (60)
RE_ASSESSED VALUE	Double
Land Size	Double

Group 2 Relationship Table



Appendix B: Risk Management

ID	RISK Description	Likelihood of Occurrence	Impact of Occurrence	Severity	Mitigation Measures	Updated Risk Management Plan
1	Have not done this type of project before Unfamiliar with GIS technology and processes can lead to technical issues. Not managing project before can lead to delays in deliverables	High	High	Med	Learning material in classes and practicing processes. Consulting with instructors whenever issues arise	Expected outcomes of unfamiliar processes differed from the actual capabilities of the tools once they were learned and it was necessary to fit them to the scope of the project and find a way to utilize them in a way that would contribute to our project, specifically Network Analysis was necessary to be used in a different way that was planned.
2	Availability of required data	Low	High	High	Our geographic question would need to be adjusted to	We found that some of our initial data like the Calgary

	and resources. The data can be out of date. We can have not enough or too much data making statistical analysis difficult.				fit the available data, further refinement of the project scope. Monitoring for house prices data availability on the internet outside of City of Calgary data.	network was not suitable for geocoding so we had to use for a different one. The suitable one available on Calgary's website was not free. We ended up using the one utilized by SAIT students.
3	Time management. School and exam schedule demands on our ability to find enough time to work on the final project.	High	High	High	Building time management skills, working outside school hours.	At the end of semester the school workload piled up with tests and assignments and despite our schedule adjustments and meeting more often to work on the project, instead of weekly meetings to 3 or 4 times a week, we still run out of time to be able to refine our analysis.
4	Computer Hardware failure, damage to data due to computer viruses	Low	Med	Low	Emergency plan to reassign work to team members who don't experience technical issues to avoid interruption to the schedule. Developing a schedule of backing up data on regular basis.	Through challenges of sharing the results of our work between three members of the team we developed the system of entering a date in the file's name after every update. Some data loss was not entirely avoided and on few occasions we had to recreate some of the processes for the second time.
5	Errors in executing data capture or data analysis	Med	Med	Med	Quality control/quality check list document developed	We used our quality control checklist to ensure our data standards met the needs of the project.
6	Organization of work, duplicate files updated separately, changes not finding its way	Med	Med	Low	Work will be done in shared documents on google drive where changes by team members can be tracked and only one person, project manager, responsible for	Occasionally, not all files were transferred during transfer processes, but we had good documentation on original sources we could consult.

	to the final report or maps				preparing final version for delivery.	
7	Parks polygon may not be suitable for network analysis	Low	Low	Med	Learning network analysis to be able to see how utilize Parks polygon best	Through attribute and spatial queries, as well as overlay techniques park polygons feature class as well as property polygons problems with datasets were mitigated.
8	Unclear communication and collaboration. Conflicts within the team, members leaving the team.	Low	Med	Med	Establish clear communication and collaboration channels, clearly define roles and responsibilities and expectations of each team member. Having a platform to share information, feedback and updates. Prepare plan to reassign tasks in case someone leaves the group due to unforeseen circumstances.	During meetings we assign roles with tasks to each other and set the due dates for tasks to be completed.