

Population-Level Data in NorthEastern Ontario

Sharing our Challenges &
Successes

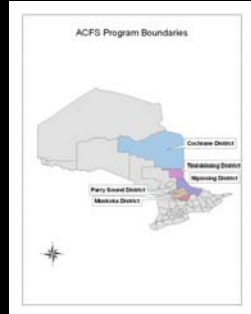
Types of Population-Level Data

Statistics Canada Census /
Tax Filer Data

Early Development Instrument (EDI)

Ontario Early Years

Challenge



Geographic Level

Which Level Best Suits Needs (e.g. CMA, DA, Postal FSA, Custom defined)? Areas are sometimes either too big or too small.

Level of Geography – Area too large – not informative enough. Area too small – not enough information.

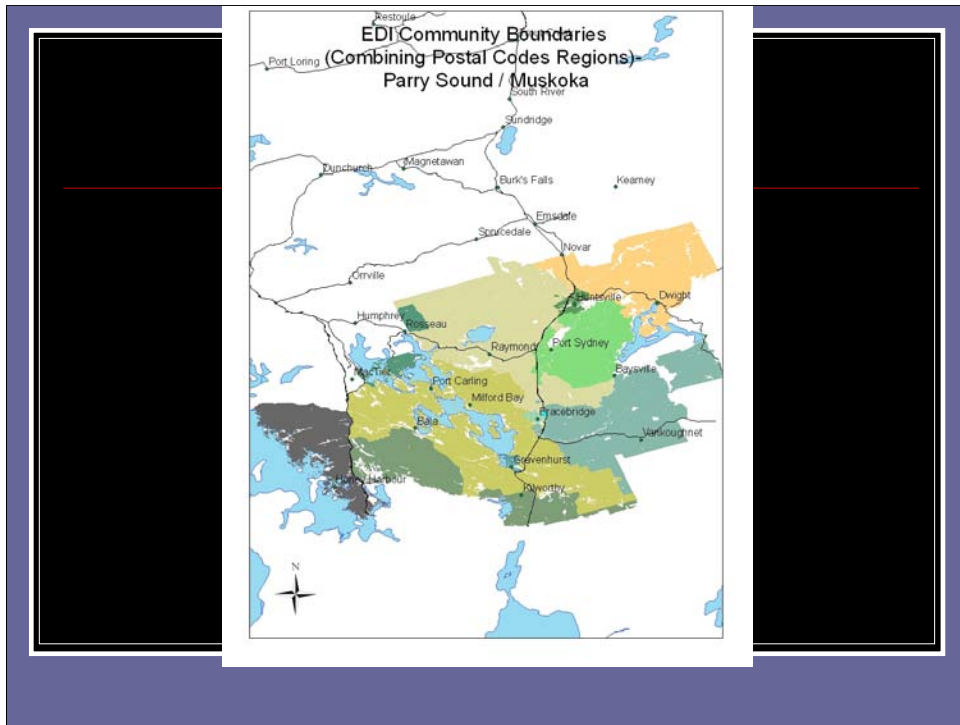
Level of Geography for each project – eg. Census data and Early Years Demographic project (compare Census district or CM level to District level demo. Data or relate Census DA / Early Years program level)? What is the purpose of the comparisons?

Purpose might vary – might need to have multiple levels of data

For EDI – we defined our own communities; however, when doing this, we ensured the boundaries aligned with a geographic level that already existed (e.g. 6 digit level postal code). Aggregated Postal Codes



Boundaries at FSA postal level Muskoka District – Too big (7 areas). Trends in Honey Harbour may be different than Severn Bridge.



Boundaries using 6 digit level postal code and aggregated together to create communities Muskoka district (11 areas).

Challenges Continued

Geographic Level

With multiple data sources, boundaries don't always align

Multiple Projects – combined analyses?

Not only thinking about each project in isolation (e.g. Census Analyses only) but will you want to analyze data from different sources together? EDI and Census – layer together on maps.

Strategy

Went with micro-level of data minimum when ordering data

➔ Census – DA level minimum

➔ Shape Files for GIS – 6 digit Postal Code level minimum (polygon/point file)

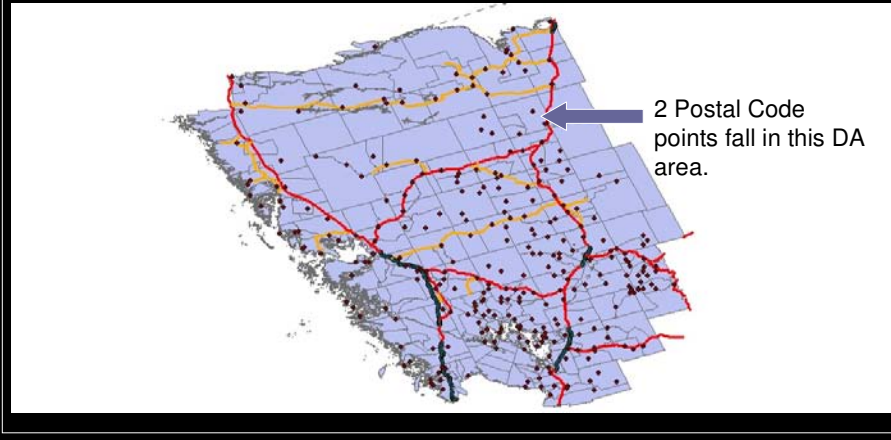
Boundaries – Used GIS Mapping Technology

Micro level – when ordering, ensured we would also have macro levels (CMA's; FSA's) so could meet needs of projects with larger geographic levels (e.g. district).

Statistics Canada – time to worked with us to determine needs. Had one main contact could call directly.

Different Boundaries – used GIS mapping to overlay one source with another (e.g. locations of EY programs, EDI custom communities and Census DA). When overlaying, used point GIS postal code file to see which DA the point falls or overlay polygon postal code file with DA polygon – assign Postal Code to DA area most postal code area falls). Haven't been through this process fully yet but thought of.

Strategy – Aligning Boundaries

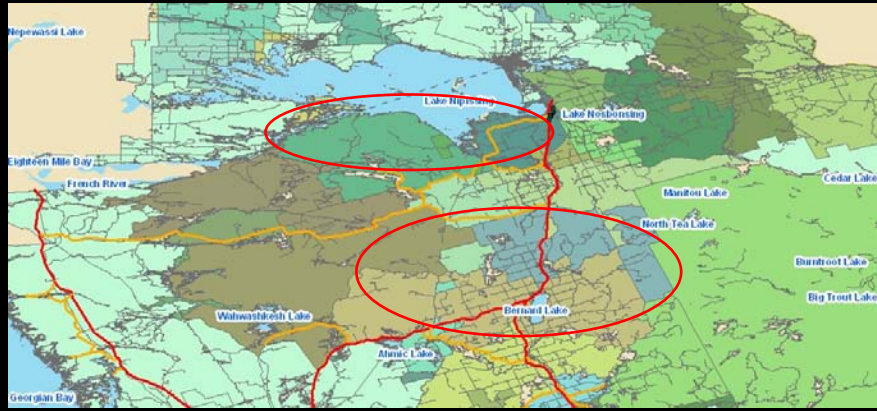


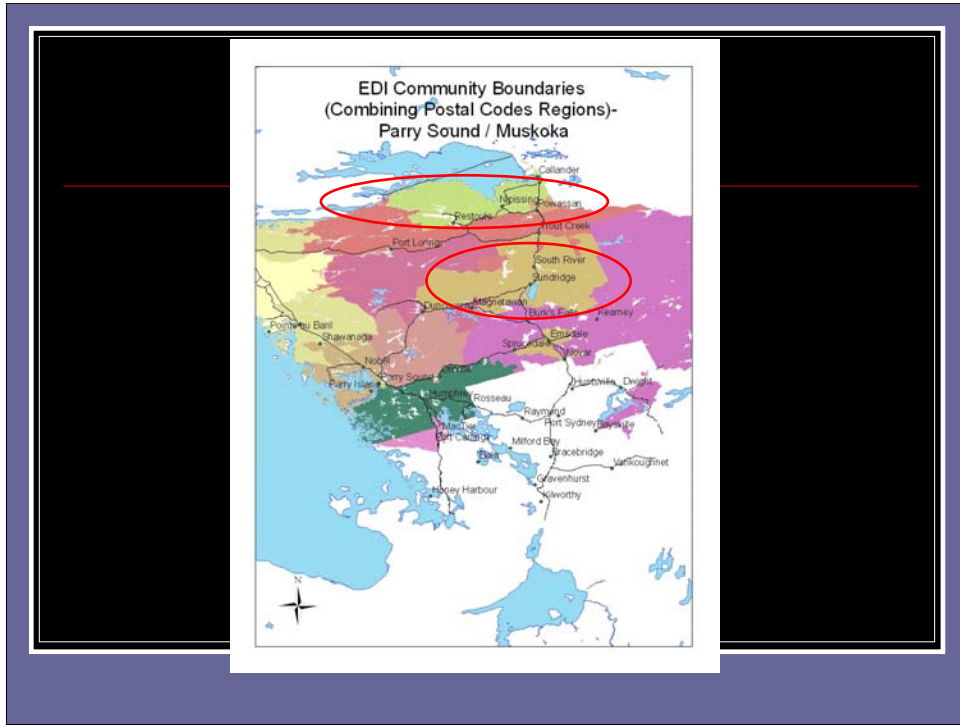
Found the postal code conversion file from Stats Canada too difficult / time consuming to use as there are many duplicate values (many to many relationship).

Strategy Con't

Creating New Boundaries – ensured new boundaries aligned with existing boundary (e.g. merging areas)

Boundaries Before Merge

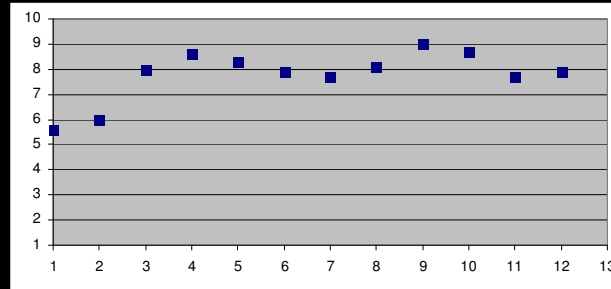




3 or more postal code areas combined to make one.

Challenge

Small Samples



Need to ensure confidentiality as most important (small sample size can breach this)

Strategy

Working with Population or sample?

What Type of Analysis Do You Need?

Create larger sample sizes (merge areas)

Merge multiple years of data together for one region

Accept low sample size and report as such

(e.g. missing data vs. low sample size / represented when mapping)

Populations of 30 or more is more acceptable for more advanced analysis as minimum but more is better. For EDI, we couldn't even meet this level in a few regions we created. We used population of 10 or more just to create a community snap shot (not using for comparisons) – remember, the EDI isn't a sample but the total SK population – the results do indicate the situation in that community. In this situation, however, important to document skews/limitations etc... that could affect the data. In essence, in these reports, more appropriate to focus attention on the sample description more so than in other reports with larger samples.

For EDI, developed communities using postal code boundaries – merged several areas together to increase sample sizes.

ALSO, we created our communities based on one year of EDI data (so that we would have at least 10 as sample size when creating). In following years, one area, with significant population drops according to Census, didn't have a sample size of 10 (8). So, we have 2 options – merge this community with another (thus creating an area that's getting too large and forcing us to re-analyze older data if want to do comparisons from one data set to the next) OR not reporting data for that area because of low sample size for that year (hoping the sample size will increase again then next round). In essence, do we sit and wait OR take action?? This is the question we are currently struggling with. Something to consider when creating boundaries and working with low sample sizes.

Strategy for EDI particularly...

Can still conduct analysis but be cautious...

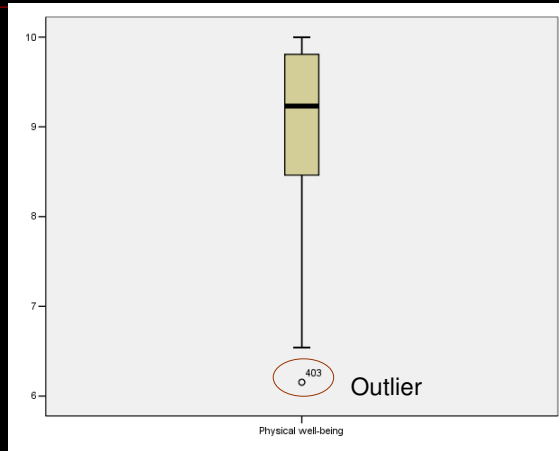
Apply Transformations

Remove “outliers” from the data set – may result in “normal distributed” without them.

If not “normally distributed”, is skew in the same direction (e.g. comparing one round to the next) ?

Can find out information about outliers and skew using the “Explore” procedure in SPSS.

Example



In case, I would remove case number 403 from analysis.

Challenge



Expense/Cost

Micro-level data

Custom Orders

Data Types GIS Mapping

Time

Strategy

- Submit Proposals
- Share the Cost
- Hire Research Assistants
- Partnerships
- Find free data (e.g. Census, street files)

Sharing Expense – Best Start Proposal – Cost shared between Muskoka/Parry Sound Best Start and my own budget (1/3 each). Could have also obtained by accessing budget through a Data Working Group I participate.

Excellent Summer Students – Recruited and Interviewed upper year University students (Sociology/Psych.) or recent grads. Good experience for graduate studies – very capable/organized/reliable/experienced.

Partnerships – 1) Relationships with Nipissing University Sociology Profs. – they can recruit for us as well. 2) School Boards – share base shape file maps / data

Share the workload with other professionals according to expertise (DWG).

Free Data – Free membership to the “Land Registry” as long as you share your files / “meta data” with other members. Can access base shape file maps with some boundary files as well.

Summary

Working with Population-level data in Northern Ontario is definitely challenging

IT REQUIRES...

planning by considering current & future uses/users
(e.g. build in flexibility at least)

time (e.g. planning; developing partnerships;
proposals)

Attention to detail (e.g. merging areas, make sure one area doesn't go with
two different regions / organization of Census raw data / ensuring correct order)

Time to do analysis (because have to do leg work to merge communities, try and line up boundaries as may not have as many political ones as larger urban areas – more investigation into skews of data etc...). In Urban areas with larger sample sizes, don't have to worry about details of skew or merging boundaries to create larger populations.

Ensuring correct order when merging data sources together into one large data set (e.g. GIS mapping). Easy to make mistakes. Need full focus and fresh mind.

Successful when...

The community indicates the data was useful and was used for planning (e.g. they loved your report and the maps)

The templates / methodology you used in past works in future with very little adjustment

The data / information makes sense to the community or can be explained

Partners feel the benefits from investment of time / money