Table of contents

Introduction

In America, unequal access to education is a widely recognized issue that appears to particularly plague impoverished districts and States. Multiple factors contribute to this problem, including the school's funding, whether it is a public or private institution, the student-to-teacher ratio, distance from the school, geographic location, familial income, and racial/ethnic background. This disparity in educational access is a deeply structural issue that spatially affects low-income states and districts.

Educational funding in America is primarily Local with the Federal government only contributing approximately 9% of their total budget (cite Atlantic). Because of this, high-poverty districts are far less able to contribute quality budgets for schooling, resulting in them spending 15.6% less per student than low-poverty districts (cite Atlantic). Lower-income families residing in impoverished districts are particularly affected by this inequality, as it has the potential to impact their children's future access to higher education and potentially, their socioeconomic status. (socio-economic status). The acknowledgement of unequal access to education in America is and has been widely understood in discourse for quite some time. However, a nation which appears to receive little attention in regard to this topic of study is Canada.

In this paper, I aim to investigate the spatial distribution of educational facilities in Canada and specifically Ontario to address the two following research claims: (1) What is the distribution of educational facilities across Canada & (2) Do they differ spatially and numerical in accordance with Provincial & municipal wealth. This paper was conducted using the statistical programming language R (R Core Team 2022). The analysis of my data was enabled using the following packages: readr (Wickham, Hester, and Bryan 2023), ggplot (Wickham 2016), tidyverse (Wickham et al. 2019), dplyr (**dplr?**), janitor (Firke 2021), knitr (Xie 2023), ggmap (Kahle and Wickham 2013), marginaleffects (Arel-Bundock 2023), gridtext (Wilke and Wiernik 2022), gridextra(**gridextra?**).

This paper commences with a discussion of the data used for analysis, it's methodologies and features. I then present a logistic regression model and an analysis of the data will follow which includes data visualization in the form of tables, graphs and maps. I conclude with a discussion of my results and a call for future research within the Canadian context.

Data

2.1 Source

This paper will follow an analysis of the distribution of education facilities across a series of Provinces in Canada. The dataset used for this paper was sourced from Statistics Canada, which provides users access to a series of Open Data bases for their public use. The source of data used for this analysis is provided by The Open Database of Education Facilities (ODEF) which is a collection of open databases concerning the types and locations of education facilities across Canada (cite statcan). The ODEF gathers data from open data portals and webpages that are managed by municipal and provincial governments to facilitate access to topics of public interest (cite statcan).

2.2 Methodology

The ODEF collects microdata on education facilities from open data portals, provincial/territorial websites, and federal departments. The individual datasets are sourced from their respective original sources and standardized into the ODEF (cite statcan). The data frame was collected between August 2019 and March 2021 (cite statcan). The target audience of this database is any educational institution that is considered a physical location where the primary purpose is to provide instruction to a group of students or participants. This thus includes all levels of education, with no exclusion for funding arrangement, subject area, denomination or student type (cite statcan). As a result, this database comprises facilities such as those for early childhood education, kindergarten, elementary, secondary, post-secondary institutions, and training centres. (cite statcan). The database mainly consists of geocoordinates obtained from its sources. The education facilities were located by searching their names, city, and province, and were included if the resulting school name closely matched the original name. (cite statcan).

2.3 Features

The obtained database comprises 18,982 observations with 28 variables, collected between 2019 and 2021. It encompasses Canadian education facilities spanning from Pre-K to Post-Secondary institutions. My examination involved removing variables that were of non-interest in my analysis and including only variables that I believed to be necessary to the research objective. Additionally, I regrouped values that referred to the same facility type but had different nomenclature. Below I have grouped the 13 variables of interest used in my analysis with the aim to describe them further.

Facility Name

- This includes the name of each facility included in the database.

Facility Type

- This variable refers to the type of institution. For this analysis, I have chosen to only include the following: Public Schools, Private Schools, Francophone Schools, Catholic Schools and First Nations Schools.

ISCED Levels

- This refers to the International Standard Classification of Education which denotes grade range to provide a standard definition of education level.
- I have chosen to only include the ISCED levels that are in accordance with Grades K - 12 (ISCED020, ISCED1, ISCED 2, ISCED 2).

Full Address

City

Province

- As a result of the provinces I wish to focus on for this paper and the need to remove missing variables, this paper will only analyze the following Provinces: Ontario, Quebec, British Columbia, Alberta, Nova Scotia and New Brunswick.

Postal Code

Census Subdivision (CSD)

- The CSD names are derived from geographic coordinates (latitude and longitude)

Longitude

Latitude

Model

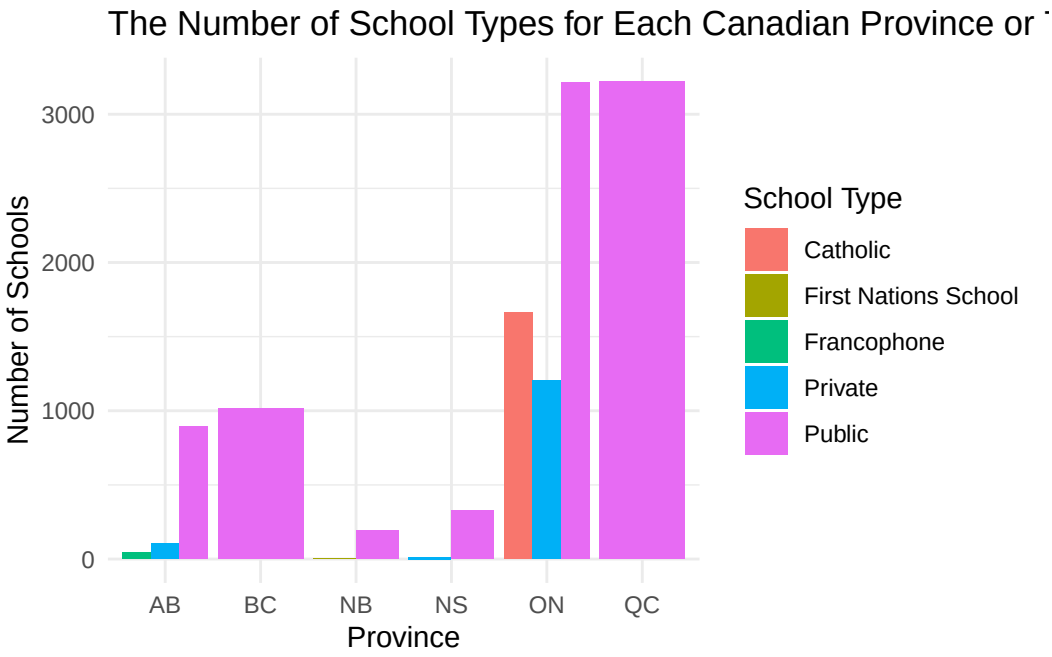
Results

To enable the visualization of my data I utilized the statistical coding language R (R Core Team 2022). The exploratory analysis began by observing the number of school types for each of the Canadian provinces of interest: Alberta, British Columbia, New Brunswick, Nova Scotia, Ontario and Quebec. The purpose of this is to examine the numerical distribution of school types in Canada and identify the province with the highest distribution of public to private schools.

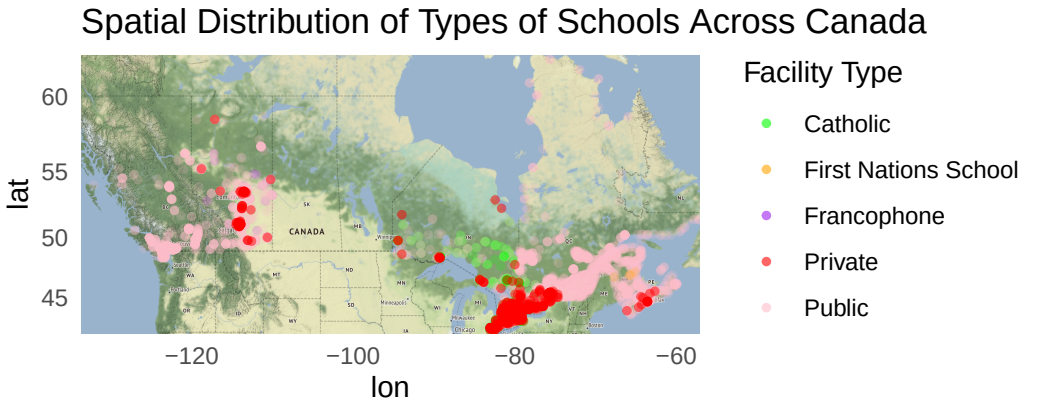
Table 1: The Number of School Types for Each Canadian Province

Province	School Type	Number of School Type
AB	Francophone	42
AB	Private	105
AB	Public	895
BC	Public	1018
NB	First Nations School	4
NB	Public	193
NS	Private	14
NS	Public	327
ON	Catholic	1662
ON	Private	1206
ON	Public	3213
QC	Public	3220

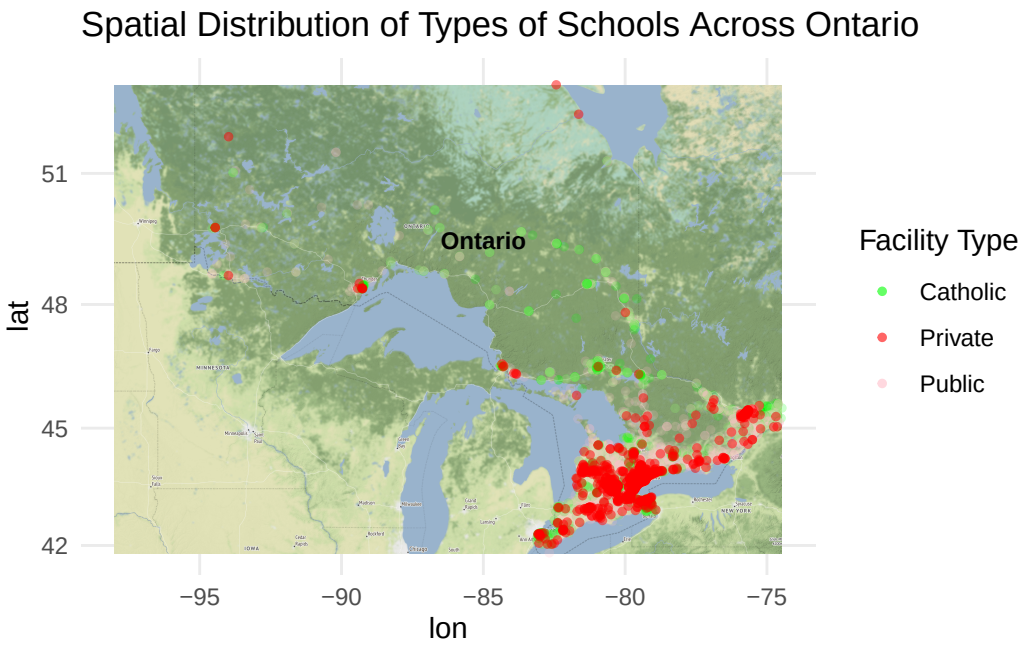
(**table_1?**) illustrates the numerical difference between each of the provinces of interest and their associated number of school types. It is apparent that Ontario maintains the largest number of schools total across all the Provinces of interest. This distribution appears accurate as Ontario has the largest population among all the provinces in Canada. Ontario also has the highest numerical values of Private schools, with 1206. Given that Ontario has both the largest population among all Canadian provinces and is the richest province with extreme wealth disparity, the observed distribution appears to be accurate. Alberta maintains the second highest variation of public to private schools with Nova Scotia being the third.



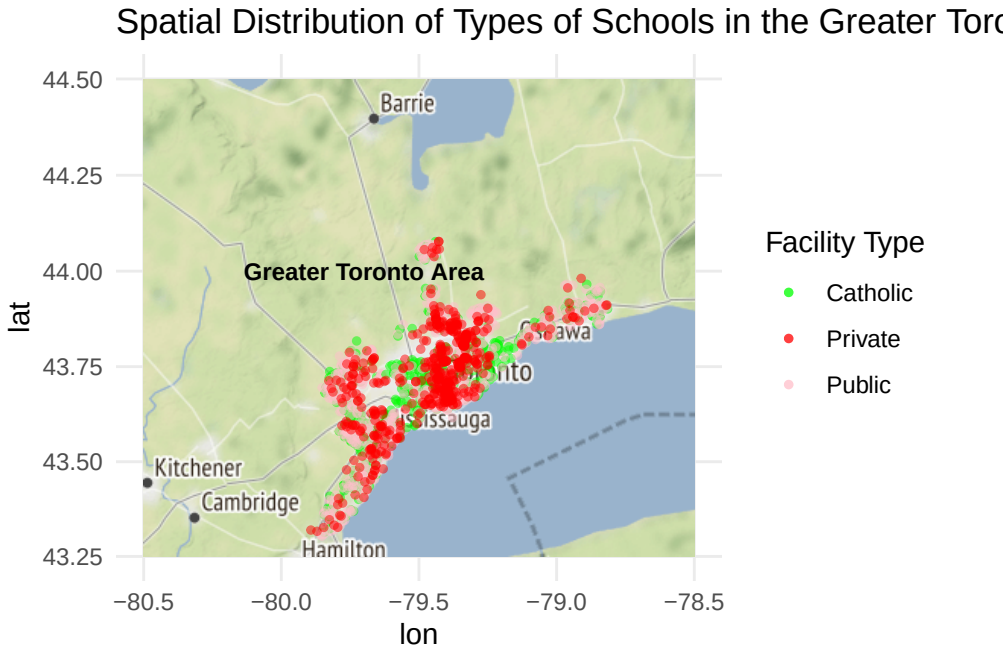
(**fig_1?**) is a visual distribution of Table 1 to better aid in the comprehension of the data. Here the ratio of private to public schools is far more digestible. It also becomes apparent that Ontario appears to be the only Province with Catholic school. This is due to the fact that Ontario is the only province that has a separate publicly funded Catholic school system. In other provinces and territories, these types of educational institutions would be classified as private schools.



(**map_1?**) showcases the spatial distribution of school types across the map of Canada. In turning your attention toward the red dots, which denote public schools, It is visually evident from the map that public schools are concentrated in the highest volume in Ontario. Upon observing the distribution of school types across Ontario, and given its status as the richest province with the highest income inequality, I have chosen to further examine this distribution within Ontario.



(**map_2?**) again is a spatial distribution of school types across the map of Ontario. Again using the same colours to denote the school types, it is apparent here that the highest volume of private schools appears to cluster in Northern Ontario among the Greater Toronto Area (GTA), Ottawa, London and Kitchener. Further, we see the scattering of private schools in the South East of Ontario, and then very few in the West of Ontario. While Southern Ontario is far more popular this relationship is notable as there is a growing Urban and Rural divide in Ontario. Furthermore, the distribution of private and public schools in Toronto is noteworthy because Toronto is not only the wealthiest city in Canada but also the most unequal with an increasing wealth gap between its rich and poor residents. To further address this distribution, I have graphed the distribution of school types across the Greater Toronto Area (gta).



(**map_3?**) highlights the spacial distribution of school types across the following cities in the GTA: Toronto,

Mississauga, Brampton, Markham, Vaughn, Oakville, Burlington, Richmond Hill, Newmarket, Pickering, Ajax, Whitby and Oshawa. From this map, it is apparent that private schools are most heavily populated in the richest areas of Toronto, and as it moves area from Toronto, the number of public schools appears to dwindle. We see that public schools cluster most heavily around Toronto, Oakville and Richmond hill which are some of the wealthiest cities in Toronto. The cause of this distribution is likely a result of the differences in wealth within these cities.

Discussion

Findings

Canadian Educational Disparities: A case study

https://www.theglobeandmail.com/news/national/time-to-lead/a-tale-of-two-schools-the-correlation-between-income-and-education/article15463950/ (cite 1)

https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2015367-eng.htm (cite 2)

https://www-tandfonline-com.myaccess.library.utoronto.ca/doi/full/10.1080/02680939.2017.1346203 (cite 3)

To better support the findings of this paper, I now turn my efforts toward a review of research on the educational disparities within Canada in relation to their geographical wealth. I have chosen to first focus my efforts on Ontario, as this was the main focus of my analysis. Toronto, Canada's largest city exhibits a growing divide between low and high-income individuals and thus exhibits a high wealth disparity. During an investigation into the high school education offered in two different neighbourhoods in Toronto The Globe and Mail argues that the difference between these schools, and the student's access to education is largely due to the socioeconomic backgrounds of the students who attend these schools (cite 1). One high school, Collegiate Institute is located in one of Toronto's wealthiest neighbourhoods, Forest Hill and the other, Central Technical School, is located in Merish Village, a far less affluent area. Students who attend education facilities in Forest Hill have higher academic achievement and see a greater amount of students pursuing post-secondary education and endeavours, Central Tech however does not showcase the same sentiments (cite 1). Students who reside in higher-income areas, such as forest hill have greater access to better educational resources in opposition to students in low-income families and areas who face far more barriers in regard to their access to quality education (cite 1). The Globe and Mail further discuss how these inequalities are perpetuated by the Canadian education system where schools in lower-income neighbourhoods receive less funding and resources than those of wealthier areas (cite 1).

To enhance my findings on geographical wealth and access to private and public schools I will now discuss a paper published by Statistic Canada titled Difference in Academic Performance Between Public and Private School Students by Geographic Region. Using the National Housing Survey to compare the academic performances of elementary to high school-aged students who attend public and private schools across Canada, their findings found that students attending private schools had higher academic performances than those in public schools (cite 2). Their findings were especially prominent in Atlantic provinces and Quebec (cite 2). The authors suggest that students who are able to attend private schools tend to come from high-income families and because of such, these types of private institutions are far more prevalent in high-income areas and neighbourhoods (cite 2).

A 2017 paper, titled The Geography of School Choice in a City with Growing Inequality: The Case of Vancouver sought to examine the socio-spatial impact of school choice on students' secondary school enrollment patterns (cite 3). Their investigation found that students who opt into over-subscribed schools (define this later) come from affluent areas and neighbourhoods with above-average levels of capital in Vancover (cite 3). Their findings found increasing levels of segregation among student mobility and the possibility that parents choose schools based on their socioeconomic characteristics within their affluent locations (cite 3). These finding indicate that schools in affluent neighbourhoods with dominant racial and ethnic groups are perceived as better than schools in less affluent neighborhoods (cite 3). → reword this

##Limitations and Bias These findings are in impart limited by the ODEF database from Statistics Canada, structural education policies, my manipulation and cleaning of variables, and biases imposed from my analysis.

To begin, the ODEF database is potentially limited in its accuracy and completeness. Such that before my manipulation, the database failed to include any Private school facilities for the Provinces of British Columbia and Quebec. This is something I know to be untrue and has thus impacted the capabilities of my analysis of this Province and the potential misrepresentation of their educational systems and facilities. Further, as a result of numerous missing values, ample data had to be removed, resulting in a failure to include smaller Canadian Provinces and Territories. The database simultaneously failed to include virtual and homeschooled education facilities, which erases a part of the population. Further, with the data collection period occurring between 2019-2021, the database in not representative of education facilities over time. Further, it is possible that the database is not entirely representative of the shift in educational facilities, policies and virtual schools which occurred following the COVID-19 pandemic.

Additionally, the removal of variables and regrouping of values during the data cleaning process has inevitably led to the loss and possible misinterpretation of data and findings. Additionally, my understanding of the original variable values for educational facility types could have been misinformed. This is a result of absent definitions included in the database and may have thus led to potential biases in my analysis.

Other limitations include the methodology and tools used for analysis and potential implicit biases which ultimately impact the way in which the data is explored, presented and discussed.

Future Research

This analysis has worked to provide insight into the distribution of educational facilities across Canadian Provinces. While the current analysis delivers valuable insights, I believe it is important to acknowledge the avenues of future research which could help strengthen this area of research within the Canadian context.

Firstly, I believe that future research should investigate the level of school funding in relation to geographical municipalities and their respective level of wealth. In doing so, I believe such research would provide more insight into the relationship between geographical wealth and school funding and act to enrich my current analysis.

Further research could also act to investigate student outcomes and academic achievements like graduation rates, and grades, again in relation to geographical location. This could provide insight into the effectiveness of different educational facilities and how they differ across municipalities of varying wealth.

Finally, future research can investigate and explore the relationship between access to certain educational facilities and demographic factors such as income, ethnicity, and race.

Appendix

Additional details

References

Arel-Bundock, Vincent. 2023. *Marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. https://CRAN.R-project.org/package=marginaleffects.

Auguie, Baptiste. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2023. *Rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. https://CRAN.R-project.org/package=rgdal.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with Ggplot2." *The R Journal* 5 (1): 144–61. https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wilke, Claus O., and Brenton M. Wiernik. 2022. *Gridtext: Improved Text Rendering Support for 'Grid' Graphics*. https://CRAN.R-project.org/package=gridtext.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.