

## Table of contents

<b>Introduction</b>	<b>3</b>
<b>Data</b>	<b>3</b>
2.1 Source . . . . .	3
2.2 Methodology . . . . .	4
2.3 Features . . . . .	4
<b>Model</b>	<b>5</b>
<b>Results</b>	<b>5</b>
<b>Discussion</b>	<b>9</b>
Findings . . . . .	9
Case Study of sorts . . . . .	9
Future Research . . . . .	9
<b>Appendix</b>	<b>10</b>
<b>Additional details</b>	<b>10</b>
<b>References</b>	<b>11</b>

```
library(readr)
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.2.3

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

```
intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

Warning: package 'tidyverse' was built under R version 4.2.3

Warning: package 'ggplot2' was built under R version 4.2.3

Warning: package 'tibble' was built under R version 4.2.3

Warning: package 'tidyverse' was built under R version 4.2.3

Warning: package 'forcats' was built under R version 4.2.3

Warning: package 'lubridate' was built under R version 4.2.3

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v forcats    1.0.0      v stringr    1.5.0
v ggplot2     3.4.2      v tibble     3.2.1
v lubridate   1.9.2      v tidyverse  1.3.0
v purrr      1.0.1
```

```
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)
library(janitor)

Warning: package 'janitor' was built under R version 4.2.3
Attaching package: 'janitor'

The following objects are masked from 'package:stats':
  chisq.test, fisher.test

library(knitr)

Warning: package 'knitr' was built under R version 4.2.3
library(ggmap)

Warning: package 'ggmap' was built under R version 4.2.3
i Google's Terms of Service: <https://mapsplatform.google.com>
i Please cite ggmap if you use it! Use `citation("ggmap")` for details.

library(gridtext)

Warning: package 'gridtext' was built under R version 4.2.3
library(grid)
library(gridExtra)

Warning: package 'gridExtra' was built under R version 4.2.3
Attaching package: 'gridExtra'

The following object is masked from 'package:dplyr':
  combine

library(sp)
library(rgdal)

Warning: package 'rgdal' was built under R version 4.2.3
Please note that rgdal will be retired during 2023,
plan transition to sf/stars/terra functions using GDAL and PROJ
at your earliest convenience.
See https://r-spatial.org/r/2022/04/12/evolution.html and https://github.com/r-spatial/evolution
rgdal: version: 1.6-5, (SVN revision 1199)
Geospatial Data Abstraction Library extensions to R successfully loaded
Loaded GDAL runtime: GDAL 3.5.2, released 2022/09/02
Path to GDAL shared files: C:/Users/gsell/AppData/Local/R/win-library/4.2/rgdal/gdal
```

```
GDAL binary built with GEOS: TRUE
Loaded PROJ runtime: Rel. 8.2.1, January 1st, 2022, [PJ_VERSION: 821]
Path to PROJ shared files: C:/Users/gsell/AppData/Local/R/win-library/4.2/rgdal/proj
PROJ CDN enabled: FALSE
Linking to sp version:1.6-0
To mute warnings of possible GDAL/OSR exportToProj4() degradation,
use options("rgdal_show_exportToProj4_warnings"="none") before loading sp or rgdal.
```

```
library(sf)
```

```
Warning: package 'sf' was built under R version 4.2.3
Linking to GEOS 3.9.3, GDAL 3.5.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

## Introduction

P1: Get a source on education facilities in Canada

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.

P2: More context and the gap my work fills

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.

P3: Introduce the dataset and the purpose of the paper and the research question

In this paper, I aim to investigate the spatial distribution of educational facilities in Canada to address the two following research claims: (1) What is the distribution of educational facilities across Canada & (2) Do they differ spatially and numerically in accordance with the wealth of the Province & municipality wealth. Using an Open Source Data made available by Statistics Canada, I have conducted an exploration of this topic and found that \_\_\_\_\_ INSERT FINDINGS HERE \_\_\_\_\_.

P4: Provide a layout for the paper (what was done, what was found and why its important)

This paper will first begin by \_\_\_\_\_ TALK ABOUT DATA SECTION\_\_\_\_\_. I will then go onto present \_\_\_\_\_ TALK ABOUT FIGURES AND TABLE\_\_\_\_\_. Next I discuss the impacts of my findings on \_\_\_\_\_ TALK ABOUT DISCUSSION/ LITERATURE \_\_\_\_\_. Where I commence by discussion implication of my study on future research and push for policy changes.

## 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

```
#reading my cleaned data for analysis

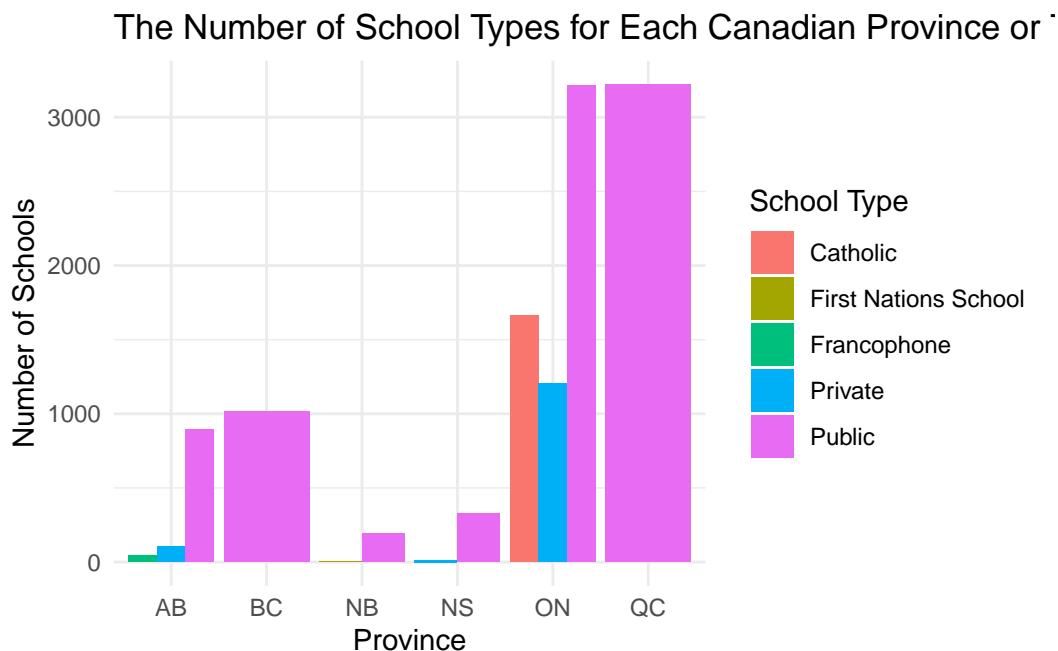
cad_data <- read_csv(here::here("inputs/data/cad_school_remove.csv"))

Rows: 11899 Columns: 13
-- Column specification -----
Delimiter: ","
chr (7): Facility_Name, Facility_Type, Full_Addr, City, Prov_Terr, Postal_Co...
dbl (6): ISCED020, ISCED1, ISCED2, ISCED3, Longitude, Latitude

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

#using ggplot to create a barchart

cad_data |>
  count(Prov_Terr, Facility_Type) |>
  ggplot(aes(x = Prov_Terr, y = n, fill = Facility_Type))+
  geom_col(position = "dodge")+
  theme_minimal()+
  labs(
    title = "The Number of School Types for Each Canadian Province or Territory",
    x = "Province",
    y = "Number of Schools",
    fill = "School Type")
```



```

#Regrouping variables of interest to create a table

number_facility_type <- cad_data |>
  count(Prov_Terr, Facility_Type)

#Using Knitr to create a table

knitr::kable(number_facility_type,
  "simple",
  caption = "The Number of School Types for Each Canadian Province",
  col.names = c("Province", "School Type", "Number of School Type"))

```

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

```

#creating dimensions of the map
bbox <- c(left = -133.68, bottom = 41.80, right = -57.129, top = 62.41)
canada_map <- get_stamenmap(bbox,
  zoom = 5,
  maptype = "terrain")

```

i Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

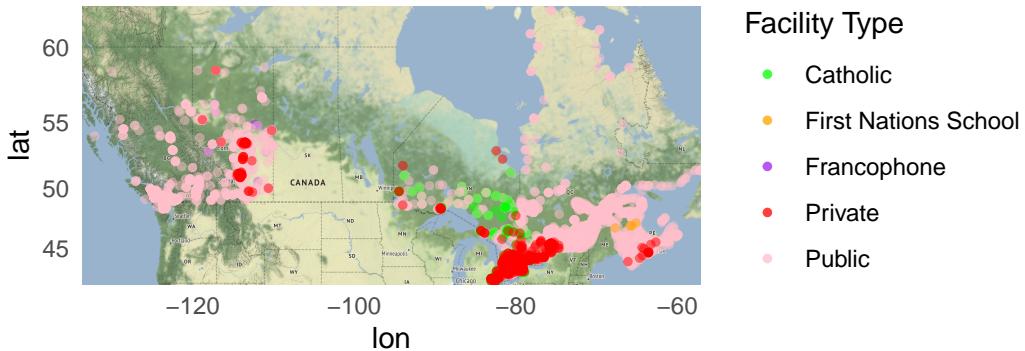
```

#plotting the map
ggmap(canada_map,
  extent = "device",
  maprange = FALSE) +
  geom_point(data = cad_data %>% filter(Facility_Type != "Private"), aes(x = Longitude, y = Latitude),
  geom_point(data = cad_data %>% filter(Facility_Type == "Private"), aes(x = Longitude, y = Latitude),
  scale_color_manual(values = c("green", "orange", "purple", "red", "pink")) +
  labs(
    title = "Spatial Distribution of Types of Schools Across Canada",
    color = "Facility Type") +
  theme_minimal()

```

Warning: Removed 3 rows containing missing values (`geom\_point()`).

## Spatial Distribution of Types of Schools Across Canada



```
pdf("map.pdf", width = 20, height = 20)
print(ggplot2::last_plot())

Warning: Removed 3 rows containing missing values (`geom_point()`).

dev.off()

pdf
2

#Filtering for only Ontario
ontario_schools <- cad_data %>%
  filter(Prov_Terr == "ON")

#Get the latitude and longitude data for Ontario schools
ontario_locations <- ontario_schools |>
  select(Longitude, Latitude, City, Facility_Type)

#making dimensions of the map

bbox <- c(left = -98, bottom = 41.80, right = -74.471, top = 52.92)
ontario_map <- get_stamenmap(bbox,
  zoom = 7,
  maptype = "terrain")

i Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.
i 63 tiles needed, this may take a while (try a smaller zoom?)
```

```

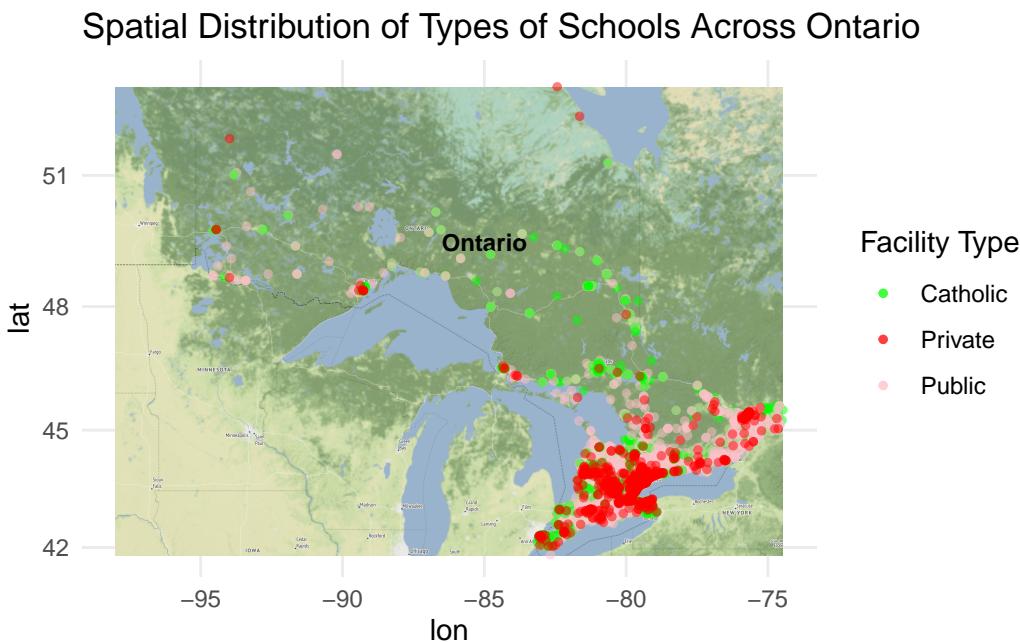
#plotting the map
ggmap(ontario_map,
      extent = "nomral",
      maprange = FALSE) +
  annotate("text", x = -85, y = 49.5, label = "Ontario", size = 3, fontface = "bold", color = "black")

  geom_point(data = ontario_schools %>% filter(Facility_Type != "Private"), aes(x = Longitude, y = Latitude),
             color = Facility_Type, size = 2)

  geom_point(data = ontario_schools %>% filter(Facility_Type == "Private"), aes(x = Longitude, y = Latitude),
             color = Facility_Type, size = 2)

  scale_color_manual(values = c("green", "red", "pink")) +
  labs(
    title = "Spatial Distribution of Types of Schools Across Ontario",
    color = "Facility Type") +
  theme_minimal()

```



```

# Define GTA cities
gta_cities <- c("Toronto", "Mississauga", "Brampton", "Markham", "Vaughan", "Oakville", "Burlington", "GTA")

# Filter for only Ontario schools
ontario_schools <- cad_data |>
  filter(Prov_Terr == "ON")

# Get the latitude and longitude data for Ontario schools in the GTA
ontario_gta_schools <- ontario_schools |>
  filter(City %in% gta_cities) |>
  select(Longitude, Latitude, City, Facility_Type)

# Define map dimensions for the GTA
gta_bbox <- c(left = -80.5, bottom = 43.25, right = -78.5, top = 44.5)

```

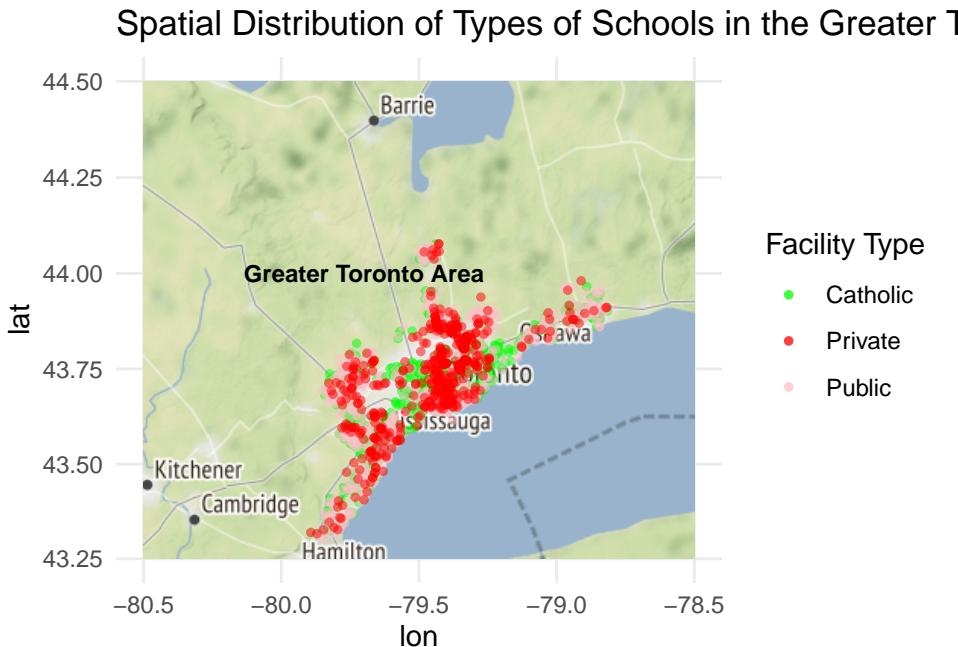
```

# Get map data for the GTA
gta_map <- get_stamenmap(bbox = gta_bbox,
                           zoom = 8,
                           maptype = "terrain")

i Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

# Plot the map with GTA schools
ggmap(gta_map,
      extent = "normal",
      maprange = FALSE) +
  annotate("text", x = -79.7, y = 44., label = "Greater Toronto Area", size = 3, fontface = "bold",
  geom_point(data = ontario_gta_schools %>% filter(Facility_Type != "Private"), aes(x = Longitude, y = Latitude), color = "green") +
  geom_point(data = ontario_gta_schools %>% filter(Facility_Type == "Private"), aes(x = Longitude, y = Latitude), color = "red") +
  scale_color_manual(values = c("green", "red", "pink")) +
  labs(title = "Spatial Distribution of Types of Schools in the Greater Toronto Area", color = "Facility Type",
       theme_minimal())

```



## Discussion

### Findings

#### Case Study of sorts

##Limitations and Bias

### Future Research

## **Appendix**

### **Additional details**

## References