R instead of GIS

Using RStudio for location data analytics

Ryan Garnett

1.0 Overview

Analyzing, evaluating and visualizing geographic data is a common requirement within municipal governments. Typically these tasks have been undertaken within commercial or open source GIS software application (i.e. ArcMap, GRASS, MapInfo, QGIS, etc.). GIS applications are typically graphic user interfaces (GUI) that are comprised of a number of tools that can be used to perform analysis and visualization tasks.

R is an open source language that is typically considered to be used for statistical analysis. As a language R has been around for approximately 25 years, experiencing significant advancements. R Studios is a GUI based interface to R that allows for analysis, visualization, modelling, and more.

Within the domain of GIS R provides an excellent platform for data analysis, exploration, modelling, and visualization. The R community is very active in the development and advancement of geospatial packages (tools/libraries) that are available. This workshop will focus on providing an introduction to the understanding of capabilities of using R Studios as a GIS application. Within the workshop examples will focus on vector data, specifically around visualization, data exploration, and geoprocessing.

2.0 Prerequisites

R Studio, similar to other software application, has the capability to extend functionality via add-ins. Within the world of R add-ins are called "packages". A package is similar to a library in other programming languages (i.e. Python), or extensions in GIS applications (i.e. ArcMap or QGIS).

As of August 17, 2018 there were 12,851 packages available on the official repository. In order to access a package it must first be installed, which can be done with the following line of code: install.packages("packageName") i.e. install.packages("tidyverse"). Once a package is installed it must be accessed using a command called *library*. Within the tutorial we will be using four packages: tidyverse, sf, and mapview.

tidyverse

The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures. The core tidyverse includes the packages that you're likely to use in everyday data analyses. As of tidyverse 1.2.0, the following packages are included in the core tidyverse:

- ggplot2
- dplyr
- tidyr
- readr
- purrr
- tibble
- stringr
- forcats

sf

Support for simple features, a standardized way to encode spatial vector data. Binds to 'GDAL' for reading and writing data, to 'GEOS' for geometrical operations, and to 'PROJ' for projection conversions and datum transformations.

mapview

An R package created to help researchers during their spatial data analysis workflow. It provides functions to very quickly and conveniently create interactive visualizations of spatial data. It was created to fill the gap of quick (not presentation grade) interactive plotting to examine and visually investigate both aspects of spatial data, the geometries and their attributes.

2.1 Packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.2.1 --
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(sf)
```

```
## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3
```

```
library(mapview)
```

2.2 Loading Vector Data

```
# Data downloaded from City of Toronto's Open Data Catalog
# https://www.toronto.ca/city-government/data-research-maps/open-data/open-
data-catalogue/
```

```
# Load vector (shapefile) information
setwd("C:/WorkDocs/rGIS/vector")

schools <- st_read("SCHOOL.shp")

## Reading layer `SCHOOL' from data source
`C:\WorkDocs\rGIS\vector\SCHOOL.shp' using driver `ESRI Shapefile'
## Simple feature collection with 1130 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 295167.7 ymin: 4827588 xmax: 334260.2 ymax: 4854324
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs

subway <- st_read("TTC_SUBWAY_LINES.shp")

## Reading layer `TTC_SUBWAY_LINES' from data source
`C:\WorkDocs\rGIS\vector\TTC_SUBWAY_LINES.shp' using driver `ESRI Shapefile'
## Simple feature collection with 4 features and 3 fields
## geometry type:  LINESTRING
## dimension:      XY
## bbox:           xmin: 301927.6 ymin: 4832875 xmax: 324823.1 ymax: 4850535
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs

wards <- st_read("WARD.shp")

## Reading layer `WARD' from data source `C:\WorkDocs\rGIS\vector\WARD.shp'
using driver `ESRI Shapefile'
## Simple feature collection with 25 features and 10 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: 293568.7 ymin: 4826564 xmax: 335747.2 ymax: 4857107
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs

#schools download url:
http://opendata.toronto.ca/gcc/school_frm_police_mtm3.zip
#subway download url:
http://opendata.toronto.ca/gcc/TTC_subway%20lines_mtm3.zip
#wards download url:
http://opendata.toronto.ca/gcc/WARD25_OpenData_08072018_mtm3.zip
```

3.0 Exploring Data

Within most analytics projects, prior to performing analysis, there is a need to explore the data. Typically this involves looking at the attribute information, assessing for quality, trends, unique values, and determining if new attributes should be created. With location

data there is another element that can be explored, the geometry. Within this section data exploration techniques, both attribute and geometry, will covered.

3.1 Exploring Attributes

One of the first steps inthe data exploration phase is to understand the structure of the attribute data, specifically knowing the column names and their data type (i.e. string, number, date, etc.). The ***str*** function will return the structure of the data table.

```
#Explore schools data
#View the structure of the data
str(schools)
```

```
## Classes 'sf' and 'data.frame':   1130 obs. of  26 variables:
##  $ NAME      : Factor w/ 1130 levels "A G B U ZAROUKIAN",..: 122 123 124
125 127 129 130 131 133 134 ...
##  $ SCL_LVL   : Factor w/ 0 levels: NA NA NA NA NA NA NA NA NA NA ...
##  $ SCL_TP    : Factor w/ 7 levels "C","EP","ES",..: 3 2 2 2 2 2 2 2 2 2
...
##  $ BRD_NAME  : Factor w/ 4 levels "Conseil scolaire de district catholique
Centre-Sud",..: 3 4 4 4 4 4 4 4 4 4 ...
##  $ SCL_TP_DSC: Factor w/ 7 levels "College","English Public",..: 3 2 2 2 2
2 2 2 2 2 ...
##  $ ADD_PT_ID : num  565242 11467213 8344927 312084 313306 ...
##  $ ADD_NUM   : Factor w/ 473 levels "1","1 1/2","10",..: 144 463 389 389
80 330 10 458 455 414 ...
##  $ LN_NAM_FUL: Factor w/ 736 levels "Albion Rd","Alexander St",..: 645 85
705 87 91 163 92 430 165 97 ...
##  $ ADD_FULL  : Factor w/ 1071 levels "1 1/2 Garfield Ave",..: 381 1050 888
875 224 735 72 1044 1035 934 ...
##  $ POSTAL_CD : Factor w/ 1013 levels "M1B 1B3","M1B 1H3",..: 312 280 934
199 207 930 352 773 764 810 ...
##  $ MUN       : Factor w/ 6 levels "East York","Etobicoke",..: 4 4 2 5 5 2
4 3 3 4 ...
##  $ CITY      : Factor w/ 1 level "Toronto": 1 1 1 1 1 1 1 1 1 1 ...
##  $ GEN_USE_CD: int  102004 102001 102001 102001 102001 102001 102001
102001 102002 102001 ...
##  $ CNTL_ID   : num  14228166 11467035 8344924 14198001 103922 ...
##  $ LO_NUM    : int  211 95 60 60 151 45 106 93 90 70 ...
##  $ LO_NUM_SUF: Factor w/ 2 levels "1/2","A": NA NA NA NA NA NA NA NA NA NA
...
##  $ HI_NUM    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HI_NUM_SUF: Factor w/ 0 levels: NA NA NA NA NA NA NA NA NA NA ...
##  $ LN_NAM_ID : num  6792 5028 2752 7427 7438 ...
##  $ X         : num  312586 318323 298326 319668 322959 ...
##  $ Y         : num  4851066 4848876 4836061 4848776 4852218 ...
##  $ LATITUDE  : num  43.8 43.8 43.7 43.8 43.8 ...
##  $ LONGITUDE : num  -79.4 -79.3 -79.6 -79.3 -79.3 ...
##  $ OBJECTID  : num  3075317 2985494 3837769 2527855 2225897 ...
##  $ RID       : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
##  $ geometry  :sfc_POINT of length 1130; first list element:  'XY' num
312586 4851066
##  - attr(*, "sf_column")= chr "geometry"
##  - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",..: NA NA NA
NA NA NA NA NA NA NA ...
##   ..- attr(*, "names")= chr  "NAME" "SCL_LVL" "SCL_TP" "BRD_NAME" ...
```

```
#Unique attributes of a column
unique(schools$SCL_TP_DSC)
```

```
## [1] English Separate English Public   Private          College
## [5] French Public    French Separate  University
## 7 Levels: College English Public English Separate ... University
```

```
#Count unique attributes of a column
schools %>%
  group_by(BRD_NAME) %>%
  count(BRD_NAME, sort = TRUE)
```

```
## Simple feature collection with 5 features and 2 fields
## geometry type:  MULTIPOINT
## dimension:      XY
## bbox:           xmin: 295167.7 ymin: 4827588 xmax: 334260.2 ymax: 4854324
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## # A tibble: 5 x 3
## # Groups:   BRD_NAME [5]
##   BRD_NAME                     n                                geometry
##   <fct>                    <int>                  <MULTIPOINT [m]>
## 1 Toronto District Schoo~   629 (295167.7 4843000, 296873.3 4845103, 2970~
## 2 <NA>                      257 (296177 4843029, 296239.9 4842922, 296269~
## 3 Toronto Catholic Distr~   223 (295197.7 4843000, 296327.3 4845518, 2969~
## 4 Conseil scolaire Viamo~    12 (299830.8 4838640, 306755.6 4841764, 3083~
## 5 Conseil scolaire de di~     9 (299889.9 4843871, 304292.8 4832669, 3095~
```

3.1.1 Selecting Subsets of Data

The abilitiy to select a subset of information is essential when analyzing data, as sometimes you may be interested in a portion of the dataset. Within R the ***select*** and ***filter*** functions in the ***tidyverse*** package allows for easy data selection and subsetting.

```
#Selecting columns
#Select specific columns
schools %>%
  select(NAME, SCL_TP, BRD_NAME,SCL_TP_DSC, ADD_FULL, MUN, CITY)
```

```
## Simple feature collection with 1130 features and 7 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 295167.7 ymin: 4827588 xmax: 334260.2 ymax: 4854324
```

```
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## First 10 features:
##                                       NAME SCL_TP
## 1                    BREBEUF COLLEGE SCHOOL     ES
## 2                       BRIAN PUBLIC SCHOOL     EP
## 3                  BRIARCREST JUNIOR SCHOOL     EP
## 4           BRIDLEWOOD JUNIOR PUBLIC SCHOOL     EP
## 5  BRIMWOOD BOULEVARD JUNIOR PUBLIC SCHOOL     EP
## 6                  BROADACRES JUNIOR SCHOOL     EP
## 7                  BROADLANDS PUBLIC SCHOOL     EP
## 8                BROCK JUNIOR PUBLIC SCHOOL     EP
## 9                  BROCKTON LEARNING CENTRE     EP
## 10                 BROOKHAVEN PUBLIC SCHOOL     EP
##                                    BRD_NAME       SCL_TP_DSC
## 1  Toronto Catholic District School Board English Separate
## 2           Toronto District School Board   English Public
## 3           Toronto District School Board   English Public
## 4           Toronto District School Board   English Public
## 5           Toronto District School Board   English Public
## 6           Toronto District School Board   English Public
## 7           Toronto District School Board   English Public
## 8           Toronto District School Board   English Public
## 9           Toronto District School Board   English Public
## 10          Toronto District School Board   English Public
##                ADD_FULL            MUN    CITY                  geometry
## 1    211 Steeles Ave E     North York Toronto POINT (312586.4 4851066)
## 2          95 Brian Dr     North York Toronto POINT (318322.5 4848876)
## 3    60 Wellesworth Dr      Etobicoke Toronto POINT (298326.5 4836061)
## 4   60 Bridlewood Blvd    Scarborough Toronto POINT (319667.8 4848776)
## 5    151 Brimwood Blvd    Scarborough Toronto POINT (322959.2 4852218)
## 6        45 Crendon Dr      Etobicoke Toronto POINT (298989.2 4834017)
## 7  106 Broadlands Blvd     North York Toronto   POINT (319553 4844899)
## 8    93 Margueretta St former Toronto Toronto POINT (309797.9 4834569)
## 9        90 Croatia St former Toronto Toronto POINT (309789.1 4835137)
## 10    70 Brookhaven Dr     North York Toronto POINT (305131.2 4840062)
```

```
#Selecting attributes
#Filter operators (i.e. greater than/less than, equal to, multiple equal, in
range of)
#Value greater than
schools %>%
  filter(LO_NUM > 50)
```

```
## Simple feature collection with 818 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 295167.7 ymin: 4827588 xmax: 334260.2 ymax: 4854324
## epsg (SRID):    NA
```

```
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## First 10 features:
##                                       NAME SCL_LVL SCL_TP
## 1                  BREBEUF COLLEGE SCHOOL    <NA>     ES
## 2                     BRIAN PUBLIC SCHOOL    <NA>     EP
## 3                BRIARCREST JUNIOR SCHOOL    <NA>     EP
## 4          BRIDLEWOOD JUNIOR PUBLIC SCHOOL    <NA>     EP
## 5  BRIMWOOD BOULEVARD JUNIOR PUBLIC SCHOOL    <NA>     EP
## 6                BROADLANDS PUBLIC SCHOOL    <NA>     EP
## 7               BROCK JUNIOR PUBLIC SCHOOL    <NA>     EP
## 8                BROCKTON LEARNING CENTRE    <NA>     EP
## 9                BROOKHAVEN PUBLIC SCHOOL    <NA>     EP
## 10                BROOKSIDE PUBLIC SCHOOL    <NA>     EP
##                                  BRD_NAME       SCL_TP_DSC ADD_PT_ID
## 1  Toronto Catholic District School Board English Separate    565242
## 2           Toronto District School Board   English Public  11467213
## 3           Toronto District School Board   English Public   8344927
## 4           Toronto District School Board   English Public    312084
## 5           Toronto District School Board   English Public    313306
## 6           Toronto District School Board   English Public    492477
## 7           Toronto District School Board   English Public   9780839
## 8           Toronto District School Board   English Public    788987
## 9           Toronto District School Board   English Public    493354
## 10          Toronto District School Board   English Public  14248664
##    ADD_NUM      LN_NAM_FUL            ADD_FULL POSTAL_CD            MUN
## 1      211   Steeles Ave E   211 Steeles Ave E   M2M 3Y6     North York
## 2       95        Brian Dr         95 Brian Dr   M2J 3Y6     North York
## 3       60  Wellesworth Dr   60 Wellesworth Dr   M9C 4R3      Etobicoke
## 4       60 Bridlewood Blvd  60 Bridlewood Blvd   M1T 1P7    Scarborough
## 5      151   Brimwood Blvd   151 Brimwood Blvd   M1V 1E5    Scarborough
## 6      106 Broadlands Blvd 106 Broadlands Blvd   M3A 1J7     North York
## 7       93  Margueretta St   93 Margueretta St   M6H 3S4 former Toronto
## 8       90      Croatia St       90 Croatia St   M6H 1K9 former Toronto
## 9       70   Brookhaven Dr    70 Brookhaven Dr   M6M 4N8     North York
## 10      75      Oasis Blvd       75 Oasis Blvd   M1X 0A3    Scarborough
##       CITY GEN_USE_CD  CNTL_ID LO_NUM LO_NUM_SUF HI_NUM HI_NUM_SUF
## 1  Toronto     102004 14228166    211       <NA>      0       <NA>
## 2  Toronto     102001 11467035     95       <NA>      0       <NA>
## 3  Toronto     102001  8344924     60       <NA>      0       <NA>
## 4  Toronto     102001 14198001     60       <NA>      0       <NA>
## 5  Toronto     102001   103922    151       <NA>      0       <NA>
## 6  Toronto     102001   441603    106       <NA>      0       <NA>
## 7  Toronto     102001 14017889     93       <NA>      0       <NA>
## 8  Toronto     102002  1144293     90       <NA>      0       <NA>
## 9  Toronto     102001   446535     70       <NA>      0       <NA>
## 10 Toronto     102001 14248995     75       <NA>      0       <NA>
##    LN_NAM_ID        X       Y LATITUDE LONGITUDE OBJECTID RID
## 1       6792 312586.4 4851066 43.80152 -79.40304  3075317   1
## 2       5028 318322.5 4848876 43.78173 -79.33181  2985494   2
```

```
## 3       2752 298326.5 4836061 43.66647 -79.58008  3837769   3
## 4       7427 319667.8 4848776 43.78080 -79.31510  2527855   4
## 5       7438 322959.2 4852218 43.81171 -79.27409  2225897   5
## 6       5039 319553.0 4844899 43.74591 -79.31663  2108916   7
## 7       3919 309797.9 4834569 43.65306 -79.43785  1813936   8
## 8       3241 309774.1 4835137 43.65817 -79.43814  3108903   9
## 9       5048 305131.2 4840062 43.70251 -79.49570  2101191  10
## 10     16908 326209.3 4854324 43.83057 -79.23361  3214631  11
##                    geometry
## 1  POINT (312586.4 4851066)
## 2  POINT (318322.5 4848876)
## 3  POINT (298326.5 4836061)
## 4  POINT (319667.8 4848776)
## 5  POINT (322959.2 4852218)
## 6    POINT (319553 4844899)
## 7  POINT (309797.9 4834569)
## 8  POINT (309789.1 4835137)
## 9  POINT (305131.2 4840062)
## 10 POINT (326209.3 4854324)
```

```
#Value less than
schools %>%
  filter(LO_NUM < 50)
```

```
## Simple feature collection with 283 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 297186.3 ymin: 4828134 xmax: 333976.7 ymax: 4854247
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## First 10 features:
##                                    NAME SCL_LVL SCL_TP
## 1             BROADACRES JUNIOR SCHOOL    <NA>     EP
## 2                   BETH TORAH HEBREWS    <NA>     PR
## 3               BLAYDON PUBLIC SCHOOL    <NA>     EP
## 4              BUCHANAN PUBLIC SCHOOL    <NA>     EP
## 5           C R MARCHANT MIDDLE SCHOOL    <NA>     EP
## 6                CALICO PUBLIC SCHOOL    <NA>     EP
## 7  CHARLES GORDON SENIOR PUBLIC SCHOOL    <NA>     EP
## 8           CITY ADULT LEARNING CENTRE    <NA>     EP
## 9  CITY VIEW ALTERNATIVE SENIOR SCHOOL    <NA>     EP
## 10             CLAIRLEA PUBLIC SCHOOL    <NA>     EP
##                          BRD_NAME     SCL_TP_DSC ADD_PT_ID ADD_NUM
## 1  Toronto District School Board English Public   7515119      45
## 2                            <NA>        Private    520627      47
## 3  Toronto District School Board English Public    490361      25
## 4  Toronto District School Board English Public    314022       4
## 5  Toronto District School Board English Public   8837557       1
## 6  Toronto District School Board English Public    495446      35
```

```
## 7  Toronto District School Board English Public   6436438      25
## 8  Toronto District School Board English Public   3841913       1
## 9  Toronto District School Board English Public  10759447      38
## 10 Toronto District School Board English Public   6511806      25
##       LN_NAM_FUL         ADD_FULL POSTAL_CD            MUN    CITY
## 1     Crendon Dr    45 Crendon Dr   M9C 3G6      Etobicoke Toronto
## 2  Glenbrook Ave 47 Glenbrook Ave   M6B 2L7     North York Toronto
## 3    Blaydon Ave   25 Blaydon Ave   M3M 2C9     North York Toronto
## 4    Bucannan Rd    4 Bucannan Rd   M1R 3V3    Scarborough Toronto
## 5       Ralph St       1 Ralph St   M9N 3A8           York Toronto
## 6      Calico Dr     35 Calico Dr   M3L 1V5     North York Toronto
## 7    Marcos Blvd   25 Marcos Blvd   M1K 5A7    Scarborough Toronto
## 8   Danforth Ave   1 Danforth Ave   M4K 1M8 former Toronto Toronto
## 9     Shirley St    38 Shirley St   M6K 1S9 former Toronto Toronto
## 10 Rosalind Cres 25 Rosalind Cres   M1L 2X1    Scarborough Toronto
##    GEN_USE_CD  CNTL_ID LO_NUM LO_NUM_SUF HI_NUM HI_NUM_SUF LN_NAM_ID
## 1      102001  7515109     45       <NA>      0       <NA>      1409
## 2      109001   446062     47       <NA>      0       <NA>      5696
## 3      102001   443191     25       <NA>      0       <NA>      4974
## 4      102001   109942      4       <NA>      0       <NA>      7464
## 5      102001  8837555      1       <NA>      0       <NA>       829
## 6      102001   442948     35       <NA>      0       <NA>      5092
## 7      102001   110208     25       <NA>      0       <NA>      8421
## 8      102002  3841911      1       <NA>      0       <NA>        86
## 9      102001 14017192     38       <NA>      0       <NA>      4400
## 10     102001  6511804     25       <NA>      0       <NA>      8824
##           X       Y LATITUDE LONGITUDE OBJECTID RID
## 1  298974.2 4834017 43.64808 -79.57202  1585500   6
## 2  308650.8 4840820 43.70933 -79.45202  2960051  25
## 3  305891.4 4843712 43.73537 -79.48626  1789838  38
## 4  320899.4 4845527 43.75153 -79.29990  2761395  40
## 5  303741.8 4839975 43.70173 -79.51294  2844008  43
## 6  304495.0 4843995 43.73791 -79.50360  1602783  45
## 7  324036.3 4844972 43.74646 -79.26096  1969589  47
## 8  316068.5 4837006 43.67492 -79.36006  2211953  59
## 9  309888.8 4834067 43.64853 -79.43673  2616872  61
## 10 321207.3 4841647 43.71660 -79.29619  2185042  64
##                    geometry
## 1  POINT (298989.2 4834017)
## 2  POINT (308650.8 4840820)
## 3  POINT (305891.4 4843712)
## 4  POINT (320899.4 4845527)
## 5  POINT (303741.8 4839975)
## 6    POINT (304495 4843995)
## 7  POINT (324036.3 4844972)
## 8  POINT (316068.5 4837006)
## 9  POINT (309903.8 4834067)
## 10 POINT (321207.3 4841647)
```

```r
#Equal to a value
schools %>%
  filter(LO_NUM == 10)
```

```
## Simple feature collection with 13 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 297530.8 ymin: 4833025 xmax: 327970.4 ymax: 4851621
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## First 10 features:
##                                              NAME SCL_LVL SCL_TP
## 1                   CENTRAL ETOBICOKE HIGH SCHOOL    <NA>     EP
## 2                GREENHOLME JUNIOR MIDDLE SCHOOL    <NA>     EP
## 3          ST COLUMBA CATHOLIC ELEMENTARY SCHOOL    <NA>     ES
## 4                   MADRESATUL BANAAT ALMUSLIMAAT    <NA>     PR
## 5               BLISS CARMAN SENIOR PUBLIC SCHOOL    <NA>     EP
## 6                         BLOORDALE MIDDLE SCHOOL    <NA>     EP
## 7                J B TYRRELL SENIOR PUBLIC SCHOOL    <NA>     EP
## 8                                       THE LINDEN    <NA>     PR
## 9  OUR LADY OF WISDOM CATHOLIC ELEMENTARY SCHOOL    <NA>     ES
## 10                  ELMBANK JUNIOR MIDDLE ACADEMY    <NA>     EP
##                                    BRD_NAME       SCL_TP_DSC ADD_PT_ID
## 1           Toronto District School Board   English Public    996285
## 2           Toronto District School Board   English Public   9178016
## 3  Toronto Catholic District School Board English Separate    349085
## 4                                      <NA>          Private   1041845
## 5           Toronto District School Board   English Public    306063
## 6           Toronto District School Board   English Public   8432037
## 7           Toronto District School Board   English Public    322672
## 8                                      <NA>          Private    861476
## 9  Toronto Catholic District School Board English Separate    348652
## 10          Toronto District School Board   English Public   8128665
##    ADD_NUM      LN_NAM_FUL           ADD_FULL POSTAL_CD           MUN
## 1       10     Denfield St     10 Denfield St   M9R 3H1      Etobicoke
## 2       10  Jamestown Cres  10 Jamestown Cres   M9V 3M5      Etobicoke
## 3       10  John Tabor Trl  10 John Tabor Trl   M1B 1M9    Scarborough
## 4       10       Vulcan St       10 Vulcan St   M9W 1L2      Etobicoke
## 5       10    Bellamy Rd S    10 Bellamy Rd S   M1M 3N8    Scarborough
## 6       10       Toledo Rd       10 Toledo Rd   M9C 2H3      Etobicoke
## 7       10 Corinthian Blvd 10 Corinthian Blvd   M1W 1B3    Scarborough
## 8       10    Rosehill Ave    10 Rosehill Ave   M4T 1G5 former Toronto
## 9       10     Japonica Rd     10 Japonica Rd   M1R 4R7    Scarborough
## 10      10    Pittsboro Dr    10 Pittsboro Dr   M9V 3R4      Etobicoke
##       CITY GEN_USE_CD  CNTL_ID LO_NUM LO_NUM_SUF HI_NUM HI_NUM_SUF
## 1  Toronto     102002   909523     10       <NA>      0       <NA>
## 2  Toronto     102001  9178005     10       <NA>      0       <NA>
## 3  Toronto     102003   104405     10       <NA>      0       <NA>
## 4  Toronto     113001   908227     10       <NA>      0       <NA>
```

```
## 5  Toronto     102001 14245815     10       <NA>      0       <NA>
## 6  Toronto     106007 30009321     10       <NA>      0       <NA>
## 7  Toronto     102001   106292     10       <NA>      0       <NA>
## 8  Toronto     112001 14021216     10       <NA>      0       <NA>
## 9  Toronto     102003   109091     10       <NA>      0       <NA>
## 10 Toronto     102001 20053315     10       <NA>      0       <NA>
##    LN_NAM_ID        X       Y LATITUDE LONGITUDE OBJECTID RID
## 1       1457 300310.2 4837454 43.67902 -79.55549  3220356 110
## 2       1870 298105.7 4843683 43.73508 -79.58291  2534446 212
## 3       8200 327970.3 4851621 43.80620 -79.21183  2437358 309
## 4       2724 298446.6 4840372 43.70527 -79.57864  3792626 353
## 5       7314 326889.0 4843447 43.73266 -79.22561  1516289 379
## 6       2640 298979.3 4833025 43.63915 -79.57195  2700776 381
## 7       7651 318935.1 4849673 43.78889 -79.32418  2438936 402
## 8       4320 313433.7 4838292 43.68653 -79.39272  3091547 592
## 9       8189 320228.5 4846701 43.76212 -79.30819  2653925 792
## 10      2278 297530.8 4843843 43.73650 -79.59005  3837891 901
##                    geometry
## 1  POINT (300310.2 4837454)
## 2  POINT (298105.7 4843683)
## 3  POINT (327970.3 4851621)
## 4  POINT (298446.6 4840372)
## 5    POINT (326889 4843447)
## 6  POINT (298979.3 4833025)
## 7  POINT (318935.1 4849673)
## 8  POINT (313433.7 4838292)
## 9  POINT (320228.5 4846701)
## 10 POINT (297530.8 4843843)
```

```
#Multiple records equal to a value
schools %>%
  filter(LO_NUM %in% c(10, 20, 30))
```

```
## Simple feature collection with 59 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 297530.8 ymin: 4829605 xmax: 333976.7 ymax: 4853342
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## First 10 features:
##                                                NAME SCL_LVL SCL_TP
## 1                    CORVETTE JUNIOR PUBLIC SCHOOL    <NA>     EP
## 2                  DANFORTH GARDENS PUBLIC SCHOOL     <NA>     EP
## 3         D'ARCY MCGEE CATHOLIC ELEMENTARY SCHOOL     <NA>     ES
## 4                    CENTRAL ETOBICOKE HIGH SCHOOL    <NA>     EP
## 5      DONVIEW MIDDLE HEALTH AND WELLNESS ACADEMY     <NA>     EP
## 6                     GENERAL CRERAR PUBLIC SCHOOL    <NA>     EP
## 7             GENERAL MERCER JUNIOR PUBLIC SCHOOL     <NA>     EP
## 8  SENHOR SANTO CRISTO CATHOLIC ELEMENTARY SCHOOL     <NA>     ES
```

```
## 9                       SHAUGHNESSY PUBLIC SCHOOL    <NA>     EP
## 10         ST BARNABAS CATHOLIC ELEMENTARY SCHOOL    <NA>     ES
##                                 BRD_NAME       SCL_TP_DSC ADD_PT_ID
## 1           Toronto District School Board   English Public    323066
## 2           Toronto District School Board   English Public   2370028
## 3  Toronto Catholic District School Board English Separate     51593
## 4           Toronto District School Board   English Public    996285
## 5           Toronto District School Board   English Public    513470
## 6           Toronto District School Board   English Public   4605063
## 7           Toronto District School Board   English Public    877576
## 8  Toronto Catholic District School Board English Separate   8713910
## 9           Toronto District School Board   English Public  10538488
## 10 Toronto Catholic District School Board English Separate    396328
##    ADD_NUM       LN_NAM_FUL            ADD_FULL POSTAL_CD            MUN
## 1       30     Corvette Ave     30 Corvette Ave   M1K 3G2    Scarborough
## 2       20 Santamonica Blvd 20 Santamonica Blvd   M1L 4H4    Scarborough
## 3       20      Bansley Ave      20 Bansley Ave   M6E 2A2           York
## 4       10      Denfield St      10 Denfield St   M9R 3H1      Etobicoke
## 5       20      Evermede Dr      20 Evermede Dr   M3A 2S3     North York
## 6       30      McGregor Rd      30 McGregor Rd   M1P 1C8    Scarborough
## 7       30    Turnberry Ave    30 Turnberry Ave   M6N 1P8 former Toronto
## 8       30       Humbert St       30 Humbert St   M6J 1M5 former Toronto
## 9       30 Shaughnessy Blvd 30 Shaughnessy Blvd   M2J 1H5     North York
## 10      30     Washburn Way     30 Washburn Way   M1B 1H3    Scarborough
##       CITY GEN_USE_CD  CNTL_ID LO_NUM LO_NUM_SUF HI_NUM HI_NUM_SUF
## 1  Toronto     102001   111593     30       <NA>      0       <NA>
## 2  Toronto     102001  2370032     20       <NA>      0       <NA>
## 3  Toronto     102003 14068332     20       <NA>      0       <NA>
## 4  Toronto     102002   909523     10       <NA>      0       <NA>
## 5  Toronto     102001   439470     20       <NA>      0       <NA>
## 6  Toronto     102001 14668247     30       <NA>      0       <NA>
## 7  Toronto     102001  1140783     30       <NA>      0       <NA>
## 8  Toronto     102003 14018848     30       <NA>      0       <NA>
## 9  Toronto     102001 10538470     30       <NA>      0       <NA>
## 10 Toronto     102003   105391     30       <NA>      0       <NA>
##    LN_NAM_ID        X       Y LATITUDE LONGITUDE OBJECTID RID
## 1       7657 323946.0 4842753 43.72649 -79.26216  1970561  74
## 2       8879 323241.1 4840748 43.70846 -79.27098  3095927  87
## 3        429 309386.2 4838925 43.69226 -79.44291  1495158  90
## 4       1457 300310.2 4837454 43.67902 -79.55549  3220356 110
## 5       5540 318356.1 4846928 43.76420 -79.33144  2434505 132
## 6       8460 322537.0 4844883 43.74570 -79.27958  3143507 144
## 7       4576 307685.8 4837458 43.67906 -79.46401  3837902 145
## 8       3712 311165.6 4833793 43.64606 -79.42090  2662746 167
## 9       6691 316724.9 4847613 43.77039 -79.35169  2515873 169
## 10      9193 326631.2 4850622 43.79725 -79.22852  1860950 204
##                    geometry
## 1    POINT (323946 4842753)
## 2  POINT (323241.1 4840748)
## 3  POINT (309386.2 4838925)
```

```
## 4  POINT (300310.2 4837454)
## 5  POINT (318356.1 4846928)
## 6    POINT (322537 4844883)
## 7  POINT (307685.8 4837458)
## 8  POINT (311165.6 4833793)
## 9  POINT (316724.9 4847613)
## 10 POINT (326631.2 4850622)
```

```
#Multiple text records equal to a value
schools %>%
  filter(SCL_TP_DSC %in% c("College", "English Public"))
```

```
## Simple feature collection with 647 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 295167.7 ymin: 4828047 xmax: 334260.2 ymax: 4854324
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## First 10 features:
##                                       NAME SCL_LVL SCL_TP
## 1                      BRIAN PUBLIC SCHOOL    <NA>     EP
## 2                  BRIARCREST JUNIOR SCHOOL    <NA>     EP
## 3           BRIDLEWOOD JUNIOR PUBLIC SCHOOL    <NA>     EP
## 4  BRIMWOOD BOULEVARD JUNIOR PUBLIC SCHOOL    <NA>     EP
## 5                  BROADACRES JUNIOR SCHOOL    <NA>     EP
## 6                  BROADLANDS PUBLIC SCHOOL    <NA>     EP
## 7                BROCK JUNIOR PUBLIC SCHOOL    <NA>     EP
## 8                  BROCKTON LEARNING CENTRE    <NA>     EP
## 9                  BROOKHAVEN PUBLIC SCHOOL    <NA>     EP
## 10                  BROOKSIDE PUBLIC SCHOOL    <NA>     EP
##                          BRD_NAME     SCL_TP_DSC ADD_PT_ID ADD_NUM
## 1  Toronto District School Board English Public  11467213      95
## 2  Toronto District School Board English Public   8344927      60
## 3  Toronto District School Board English Public    312084      60
## 4  Toronto District School Board English Public    313306     151
## 5  Toronto District School Board English Public   7515119      45
## 6  Toronto District School Board English Public    492477     106
## 7  Toronto District School Board English Public   9780839      93
## 8  Toronto District School Board English Public    788987      90
## 9  Toronto District School Board English Public    493354      70
## 10 Toronto District School Board English Public  14248664      75
##         LN_NAM_FUL            ADD_FULL POSTAL_CD            MUN    CITY
## 1         Brian Dr         95 Brian Dr   M2J 3Y6     North York Toronto
## 2   Wellesworth Dr   60 Wellesworth Dr   M9C 4R3      Etobicoke Toronto
## 3  Bridlewood Blvd  60 Bridlewood Blvd   M1T 1P7    Scarborough Toronto
## 4    Brimwood Blvd   151 Brimwood Blvd   M1V 1E5    Scarborough Toronto
## 5       Crendon Dr       45 Crendon Dr   M9C 3G6      Etobicoke Toronto
## 6  Broadlands Blvd 106 Broadlands Blvd   M3A 1J7     North York Toronto
## 7   Margueretta St   93 Margueretta St   M6H 3S4 former Toronto Toronto
```

```
## 8       Croatia St       90 Croatia St   M6H 1K9 former Toronto Toronto
## 9    Brookhaven Dr    70 Brookhaven Dr   M6M 4N8     North York Toronto
## 10      Oasis Blvd       75 Oasis Blvd   M1X 0A3    Scarborough Toronto
##    GEN_USE_CD  CNTL_ID LO_NUM LO_NUM_SUF HI_NUM HI_NUM_SUF LN_NAM_ID
## 1      102001 11467035     95       <NA>      0       <NA>      5028
## 2      102001  8344924     60       <NA>      0       <NA>      2752
## 3      102001 14198001     60       <NA>      0       <NA>      7427
## 4      102001   103922    151       <NA>      0       <NA>      7438
## 5      102001  7515109     45       <NA>      0       <NA>      1409
## 6      102001   441603    106       <NA>      0       <NA>      5039
## 7      102001 14017889     93       <NA>      0       <NA>      3919
## 8      102002  1144293     90       <NA>      0       <NA>      3241
## 9      102001   446535     70       <NA>      0       <NA>      5048
## 10     102001 14248995     75       <NA>      0       <NA>     16908
##           X       Y LATITUDE LONGITUDE OBJECTID RID
## 1  318322.5 4848876 43.78173 -79.33181  2985494   2
## 2  298326.5 4836061 43.66647 -79.58008  3837769   3
## 3  319667.8 4848776 43.78080 -79.31510  2527855   4
## 4  322959.2 4852218 43.81171 -79.27409  2225897   5
## 5  298974.2 4834017 43.64808 -79.57202  1585500   6
## 6  319553.0 4844899 43.74591 -79.31663  2108916   7
## 7  309797.9 4834569 43.65306 -79.43785  1813936   8
## 8  309774.1 4835137 43.65817 -79.43814  3108903   9
## 9  305131.2 4840062 43.70251 -79.49570  2101191  10
## 10 326209.3 4854324 43.83057 -79.23361  3214631  11
##                    geometry
## 1  POINT (318322.5 4848876)
## 2  POINT (298326.5 4836061)
## 3  POINT (319667.8 4848776)
## 4  POINT (322959.2 4852218)
## 5  POINT (298989.2 4834017)
## 6    POINT (319553 4844899)
## 7  POINT (309797.9 4834569)
## 8  POINT (309789.1 4835137)
## 9  POINT (305131.2 4840062)
## 10 POINT (326209.3 4854324)

#Selecting attributes
#Filter attributes in multiple columns
schools %>%
  filter(LO_NUM < 50, SCL_TP_DSC == "College")

## Simple feature collection with 2 features and 25 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 303219 ymin: 4841814 xmax: 303988.5 ymax: 4846283
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
##                              NAME SCL_LVL SCL_TP BRD_NAME SCL_TP_DSC
```

```
## 1     SENECA COLLEGE (JANE CAMPUS)    <NA>      C     <NA>    College
## 2 SENECA COLLEGE (YORKGATE CAMPUS)    <NA>      C     <NA>    College
##   ADD_PT_ID ADD_NUM       LN_NAM_FUL            ADD_FULL POSTAL_CD
## 1  20092290      21 Beverly Hills Dr 21 Beverly Hills Dr   M3L 1A2
## 2  11297202       1   York Gate Blvd    1 York Gate Blvd   M3N 3A1
##          MUN    CITY GEN_USE_CD  CNTL_ID LO_NUM LO_NUM_SUF HI_NUM
## 1 North York Toronto     102006 10145179     21       <NA>      0
## 2 North York Toronto     115001 30020932      1       <NA>      0
##   HI_NUM_SUF LN_NAM_ID        X       Y LATITUDE LONGITUDE OBJECTID RID
## 1       <NA>      4949 303988.5 4841814 43.71828 -79.50988  3185452 161
## 2       <NA>      7143 303219.0 4846283 43.75851 -79.51944  3020433 164
##                   geometry
## 1 POINT (303988.5 4841814)
## 2   POINT (303219 4846283)

#Combining column and attribute selection
schools %>%
  select(NAME, SCL_TP, BRD_NAME, SCL_TP_DSC, LO_NUM, ADD_FULL, MUN, CITY) %>%
  filter(LO_NUM < 50, SCL_TP_DSC == "College")

## Simple feature collection with 2 features and 8 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 303219 ymin: 4841814 xmax: 303988.5 ymax: 4846283
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
##                              NAME SCL_TP BRD_NAME SCL_TP_DSC LO_NUM
## 1     SENECA COLLEGE (JANE CAMPUS)      C     <NA>    College     21
## 2 SENECA COLLEGE (YORKGATE CAMPUS)      C     <NA>    College      1
##              ADD_FULL        MUN    CITY                  geometry
## 1 21 Beverly Hills Dr North York Toronto POINT (303988.5 4841814)
## 2    1 York Gate Blvd North York Toronto   POINT (303219 4846283)
```

3.2 Exploring Geometry

Location data has another dimension of information, geometry. Exploring properties and components of the geometry can provide valuable insights into the data. Within this section we will explore common geometry exploration techniques.

```
#Count the number of nodes
npts(schools)

## [1] 1130

npts(subway)

## [1] 1497

npts(wards)

## [1] 25434
```

```r
#Calculate area of polygon
st_area(wards)               #area of each polygon
```

```
## Units: [m^2]
##  [1] 22964601 40022969 30296208 30437109 13597720 18702653 12095504
##  [8] 24979346 24421053 22662279 37349157 48365656 15328717 13122348
## [15] 21399284 28205058 26117022 54089586 28190607 16804389 19762924
## [22] 35360126 30701851 21801465  5863100
```

```r
sum(st_area(wards))          #sum of all polygons
```

```
## 642640730 [m^2]
```

```r
#Add attribute to polygon with calculated area
wards$area_m <- st_area(wards)
head(wards)
```

```
## Simple feature collection with 6 features and 11 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: 299266.8 ymin: 4826564 xmax: 327661.7 ymax: 4855605
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
##   AREA_ID AREA_TYPE AREA_S_CD AREA_L_CD           AREA_NAME        X
## 1 2551040      WD18        16        16     Don Valley East 318237.3
## 2 2551044      WD18        03        03 Etobicoke-Lakeshore 303099.5
## 3 2551048      WD18        15        15     Don Valley West 314825.9
## 4 2551052      WD18        23        23   Scarborough North 324522.1
## 5 2551056      WD18        11        11 University-Rosedale 313306.5
## 6 2551035      WD18        10        10   Spadina-Fort York 313874.8
##         Y LONGITUDE LATITUDE OBJECTID                       geometry
## 1 4844000 -79.33298 43.73972 18150849 POLYGON ((319824.6 4841688,...
## 2 4831000 -79.52087 43.62165 18150785 POLYGON ((304964.2 4834449,...
## 3 4843000 -79.37536 43.72840 18150721 POLYGON ((316684.2 4841542,...
## 4 4852000 -79.25467 43.80967 18150657 POLYGON ((326825.3 4855329,...
## 5 4837000 -79.39432 43.67114 18150593 POLYGON ((313648.6 4838734,...
## 6 4833000 -79.38733 43.63580 18150929 POLYGON ((314159.7 4834870,...
##           area_m
## 1 22964601 [m^2]
## 2 40022969 [m^2]
## 3 30296208 [m^2]
## 4 30437109 [m^2]
## 5 13597720 [m^2]
## 6 18702653 [m^2]
```

```r
#Calculate length of line
st_length(subway)
```

```
## Units: [m]
## [1] 38894.062 26192.382  6622.532  5366.335
```

```r
#Add attribute to line with length
subway$length_m <- st_length(subway)
head(subway)
```

```
## Simple feature collection with 4 features and 4 fields
## geometry type:  LINESTRING
## dimension:      XY
## bbox:           xmin: 301927.6 ymin: 4832875 xmax: 324823.1 ymax: 4850535
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
##   OBJECTID                ROUTE_NAME RID                       geometry
## 1    53420 LINE 1 (YONGE-UNIVERSITY)   1 LINESTRING (302520.8 485053...
## 2    53421 LINE 2 (BLOOR - DANFORTH)   2 LINESTRING (301927.6 483287...
## 3    53422      LINE 3 (SCARBOROUGH)   3 LINESTRING (323850.7 484343...
## 4    53423         LINE 4 (SHEPPARD)   4 LINESTRING (311939.6 484661...
##        length_m
## 1 38894.062 [m]
## 2 26192.382 [m]
## 3  6622.532 [m]
## 4  5366.335 [m]
```

```r
#Check for valid geometry (OGC)
ogcGeom <- st_is_valid(wards)     #test layer for valid geometry
length(ogcGeom[ogcGeom == FALSE])  #count the number of invalid geometry
```

```
## [1] 0
```

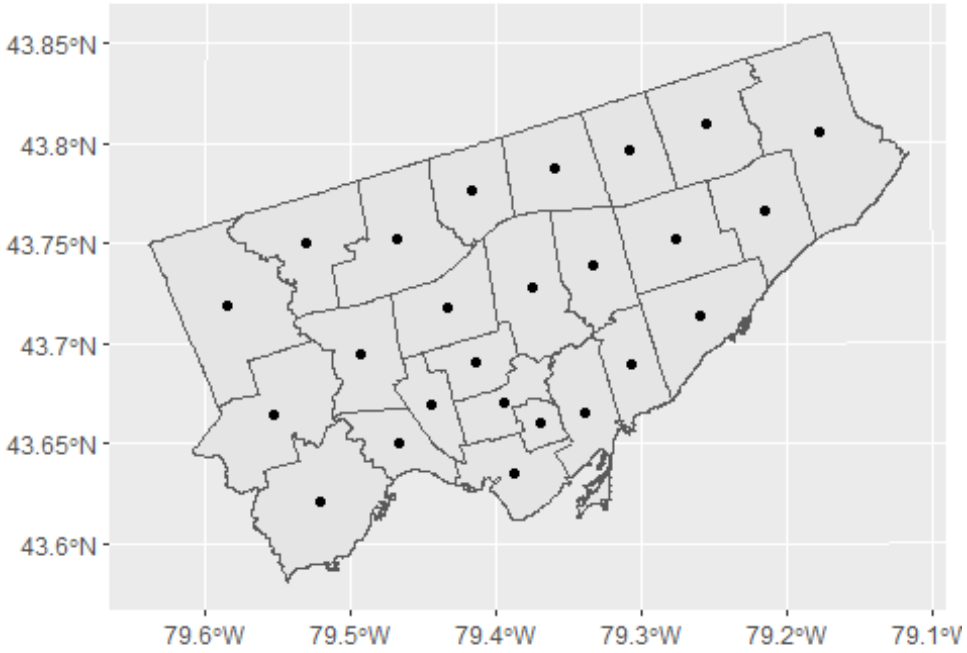
```r
length(ogcGeom[ogcGeom == TRUE])   #count the number of valid geometry
```

```
## [1] 25
```

```r
#Create centroid of polygon
wards_cent <- st_centroid(wards)
```

```
## Warning in st_centroid.sf(wards): st_centroid assumes attributes are
## constant over geometries of x
```

```r
#Visualize centroids using ggplot

ggplot() +
  geom_sf(data = wards) +
  geom_sf(data = wards_cent)
```



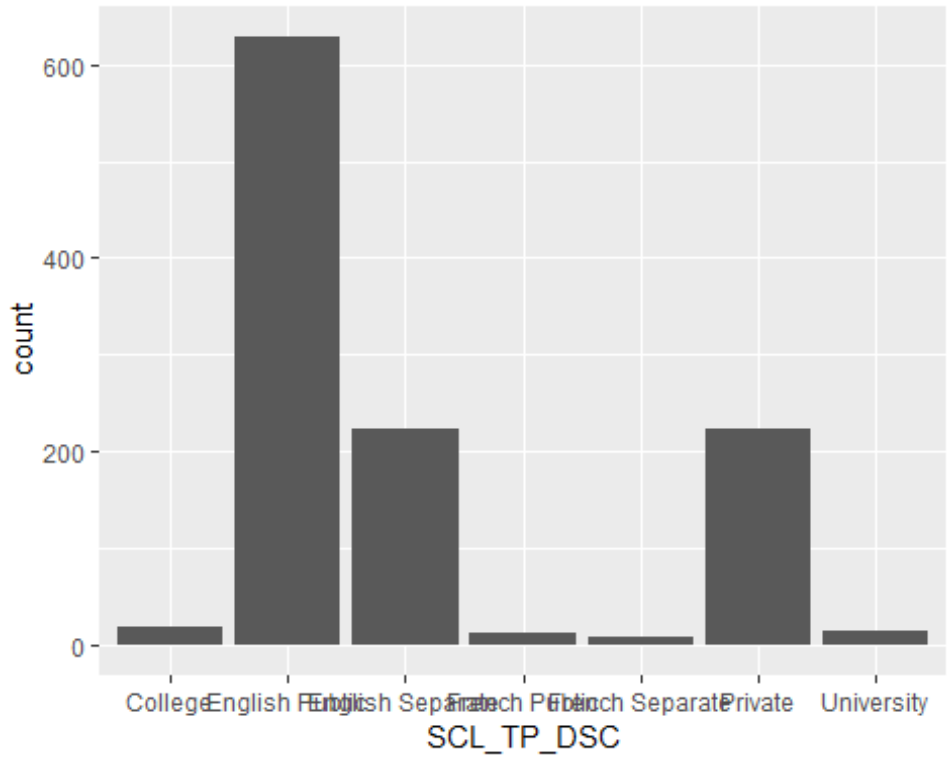
4.0 Plotting Data

Displaying data visually within R can be accomplished with the use of many different packages. Within this workshop we will be focusing on the powerful ggplot2 package. ggplot2 allows for the visualization of both attribute and geographic information, r esulting in a very useful tool for GIS users.

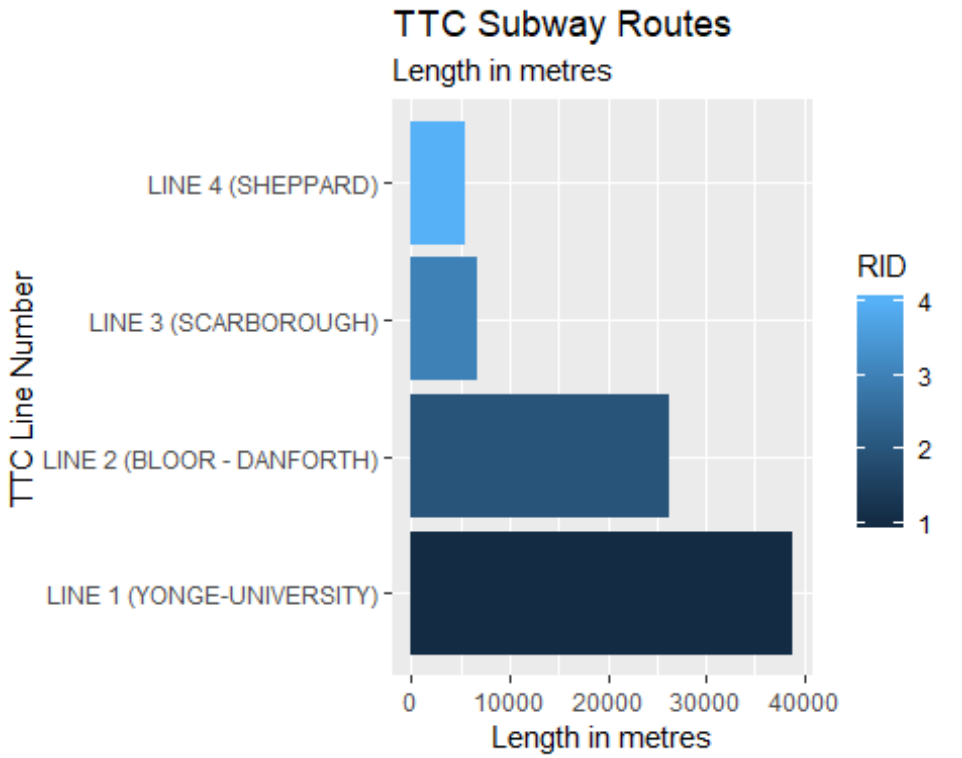
4.1 Charts and Graphs

Location data typically has attribute information, allowing for the creation of insightful visualization. Selecting an effective visualization is dependant on the available attributes. For further information on selecting effective visualization, refer to this resource on designing graphs and charts.

```
#Using ggplot2 to view attribute information (bar chart)
#Visualizing the frequency of school group
ggplot(schools, aes(SCL_TP_DSC)) +
  geom_bar()
```



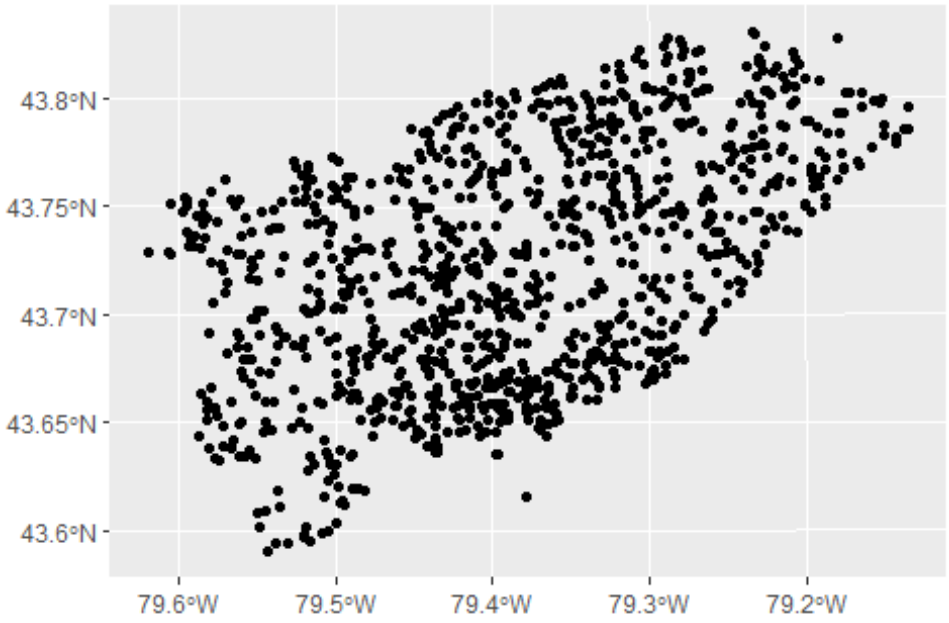

```
#Visualizing the length of subway route length by TTC line
ggplot(subway, aes(x = ROUTE_NAME, y = as.numeric(length_m), label = 'Length
in metres')) +
  geom_bar(stat = 'identity', aes(fill = RID)) +
  labs(subtitle = "Length in metres",
  title = "TTC Subway Routes") +
  labs(x = "TTC Line Number") +
  labs(y = "Length in metres") +
  coord_flip()
```



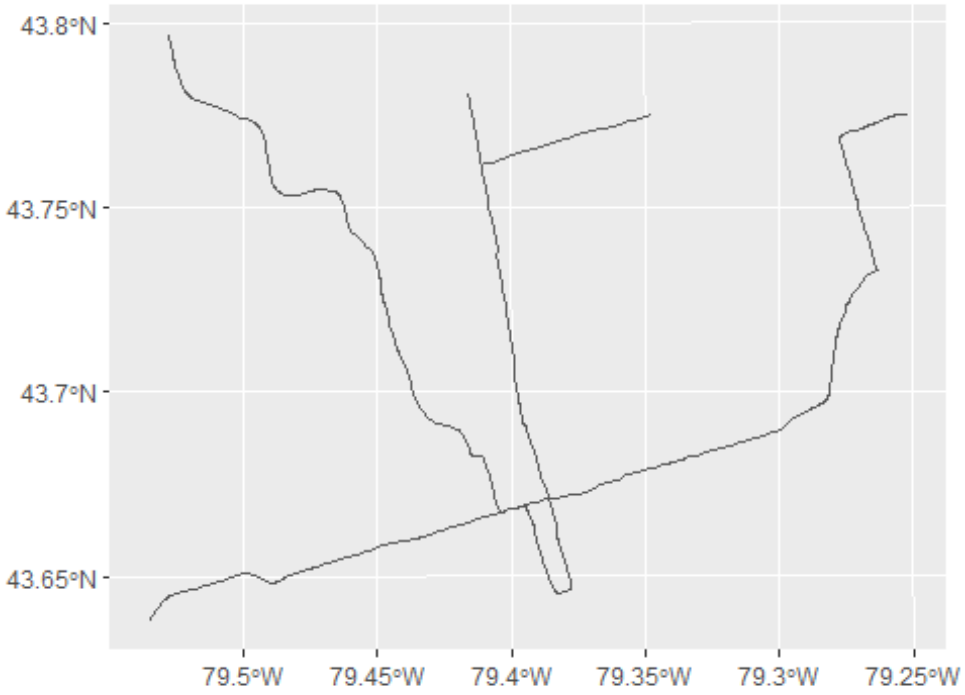
4.2 Static Maps - Simple Symbology

Maps are an effective visualization for conveying geographic information. A map can server numerous purposes and can range in complexity. In many cases a basic map will suffice when initially exploring a dataset. The following examples illustrate how to display points, lines and polygons within ggplot2.

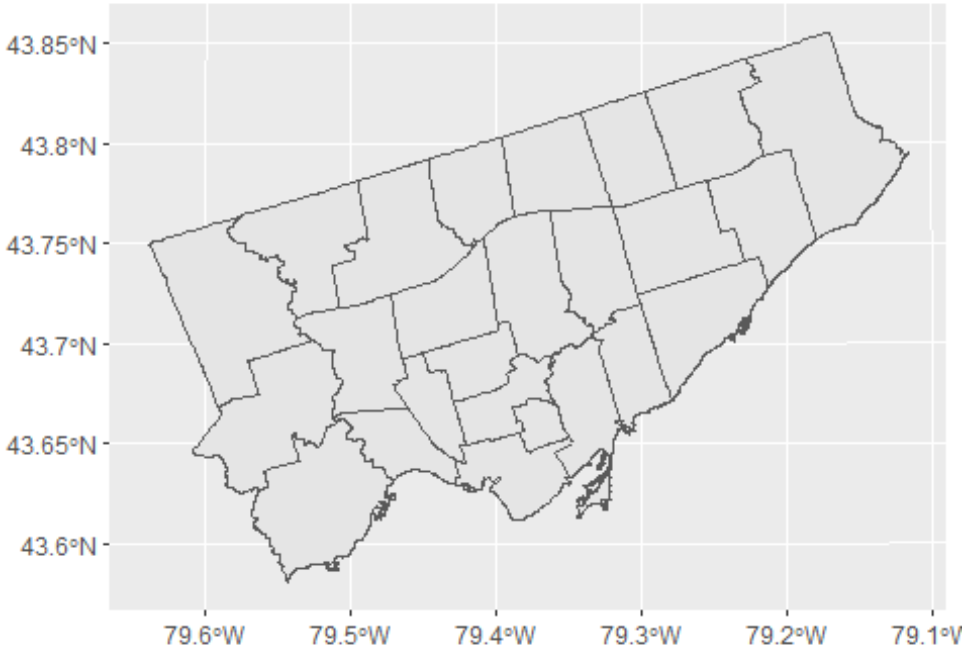
```
#Plot point vector using ggplot2
#Plot without symbology
ggplot(data = schools) +
  geom_sf()
```




```
#Plot line vector using ggplot2
#Plot without symbology
ggplot(data = subway) +
  geom_sf()
```




```
#Plot polygon vector using ggplot2
#Plot without symbology
ggplot(data = wards) +
  geom_sf()
```



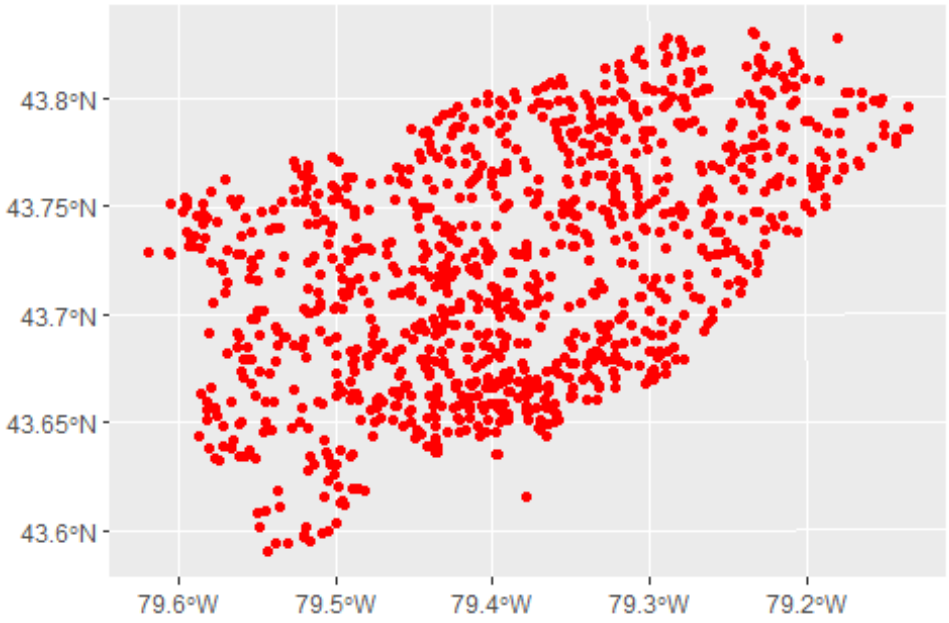
4.3 Static Maps - Custom Symbology

While basic maps can be useful for initial data exploration, understanding tends and relationships from a basic map can be difficult. A method to gain visual insights is to use different symbology techniques. Within the workshop we will be focusing on symbology techniques, specifically: colours, transparency, size, shapes, labels, and legends. Examples with points, lines and polygons will be illustrated.

4.3.1 Static Maps - Custom Symbology: Colours

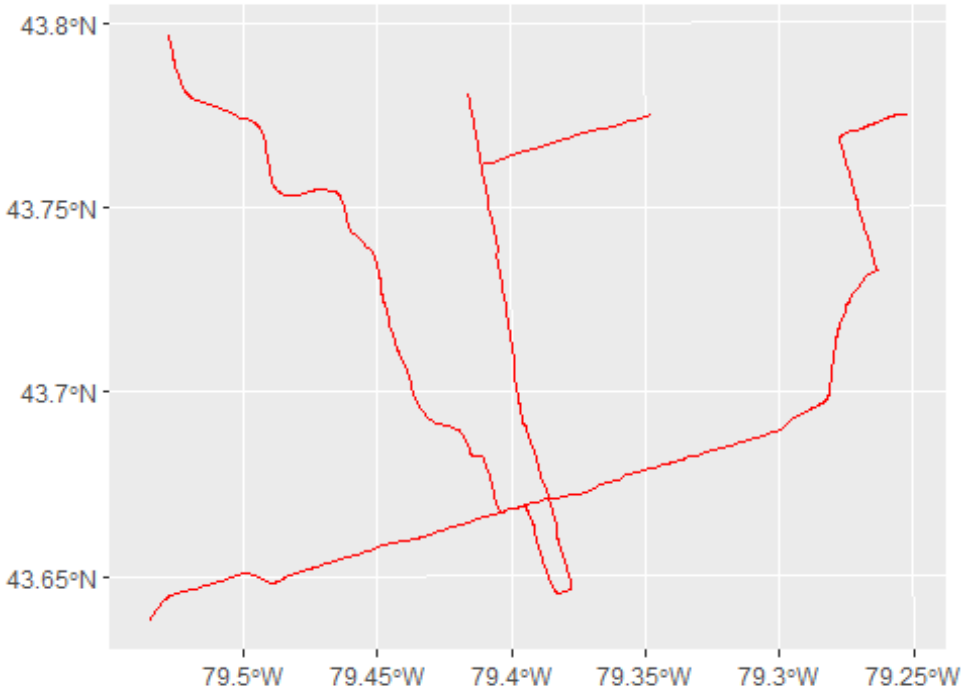
Symbology - colour by text

```
#Map setting the color using color text
#Map schools
ggplot(data = schools) +
  geom_sf(color = "red")
```



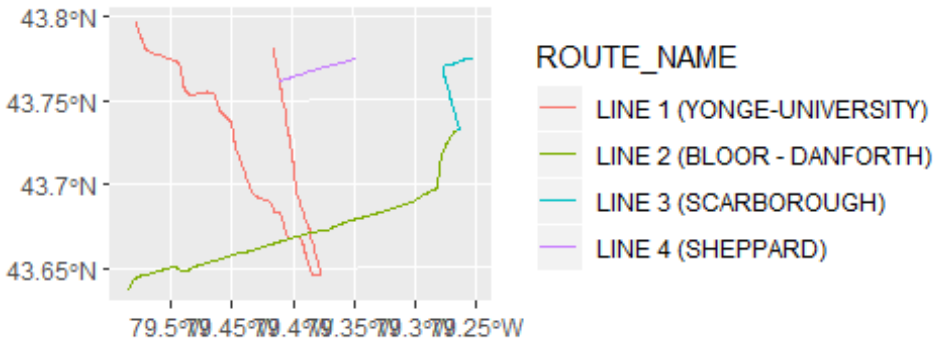
Symbology - colour by value

```
#Map setting the color using hex code
#Map subway
ggplot(data = subway) +
  geom_sf(color = "#FF0000")
```



Symbology - colour by attribute

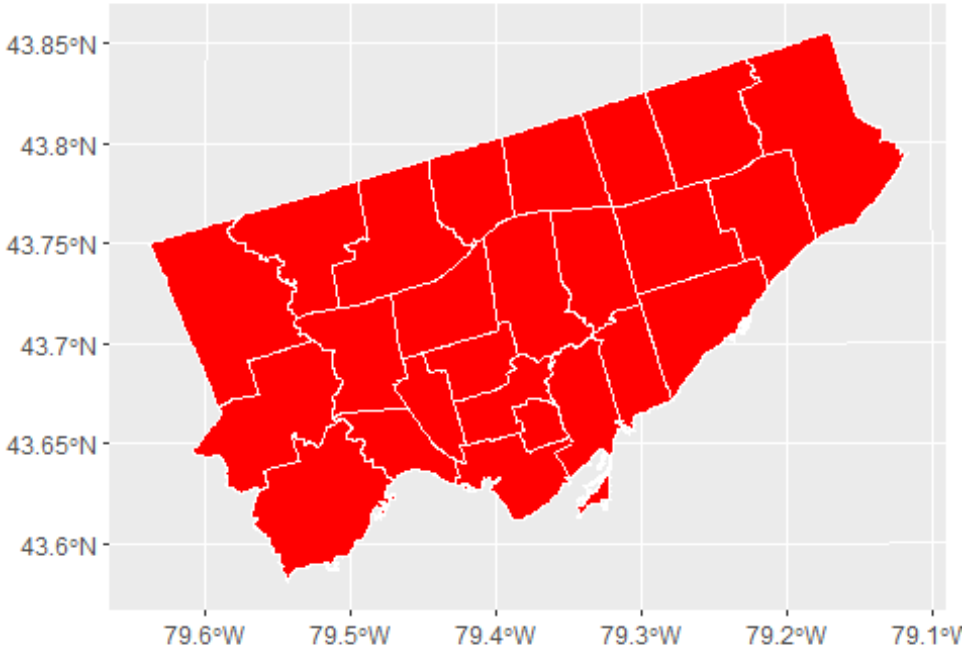
```
#Map setting the color based on attribute values
#Map subway lines
ggplot(data = subway) +
  geom_sf(aes(color = ROUTE_NAME), show.legend = "line")
```




```
#The show.legend = "line" piece changes the legend symbology from polygons to
line. Replacing "line" to "point" will use points in the legend
```

Symbology - polygon fill

```
#Map setting the color using color text
#Map wards
ggplot(data = wards) +
  geom_sf(fill = "red", color = "white")
```



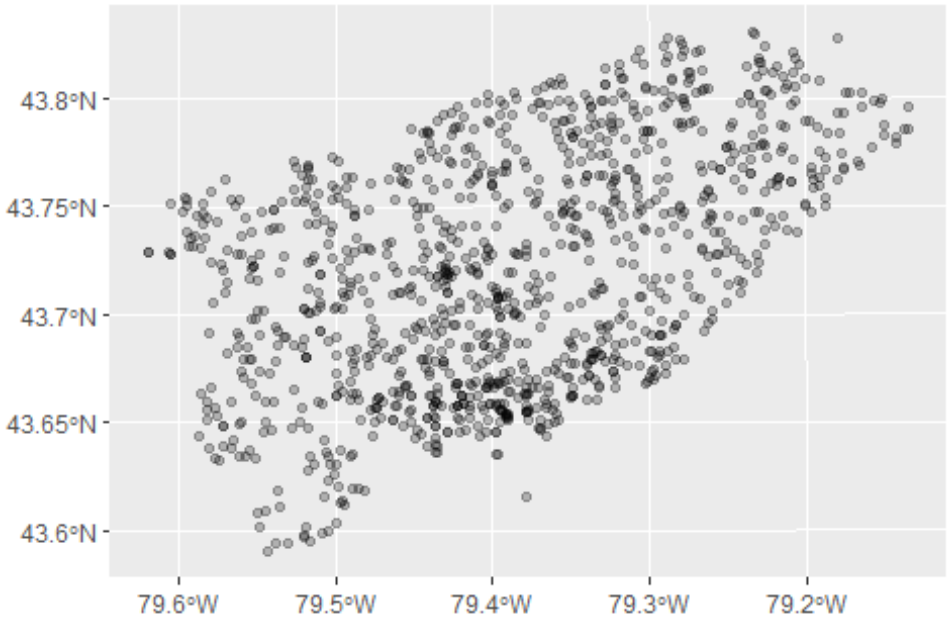

```
# fill controls the internal polygon colour; color controls the outline
polygon colour
```

4.3.2 Static Maps - Custom Symbology: Transparency

Transparceny is an effective technique when diplaying datasets with a significant amount of data. Transparency can be applied with there is a need to layer where data points overlap each other, or when layers are stacked, or to visually deprioritize one layer over another.

Symbology - point transparency

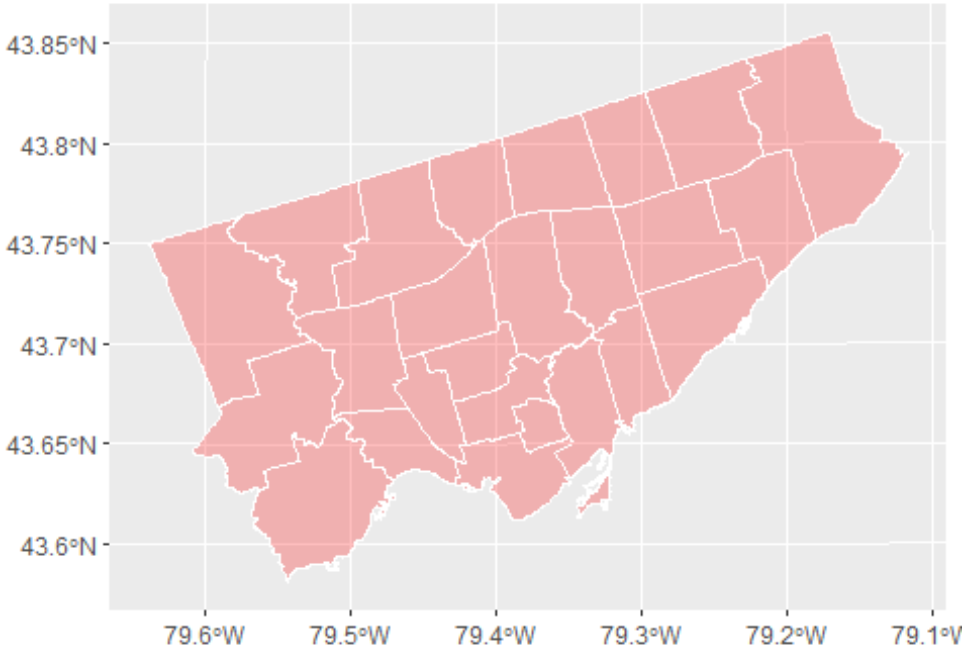
```
#Map setting the transparceny of a point layer
#Map schools
ggplot(data = schools) +
  geom_sf(alpha = 0.25)
```




```
#Alpha is the means for controlling transparency. Values from 1 to 0 are used
to set the transparency level, with 0 being fully transparent and 1 having no
transparency.
```

Symbology - polygon transparency

```
#Map setting the transparceny of a polygon
#Map wards
ggplot(data = wards) +
  geom_sf(alpha = 0.25, fill = "red", color = "white")
```



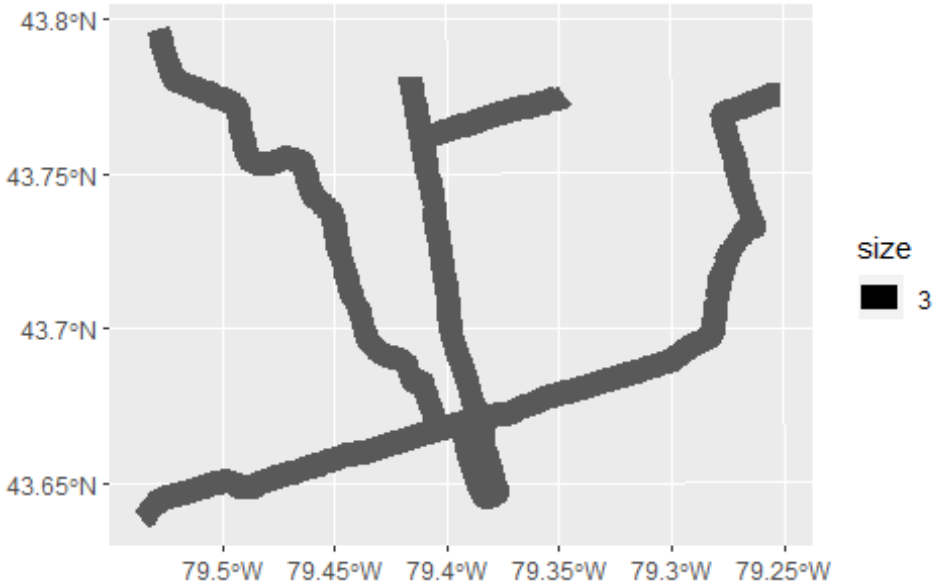

```
#Alpha is the means for controlling transparency. Values from 1 to 0 are used
to set the transparency level, with 0 being fully transparent and 1 having no
transparency.
```

4.3.3 Static Maps - Custom Symbology: Sizes

Modifying the size of an object can be useful in many scenarios, such as illustrating importance, or ensuring features are visible.

Symbology - static sizes

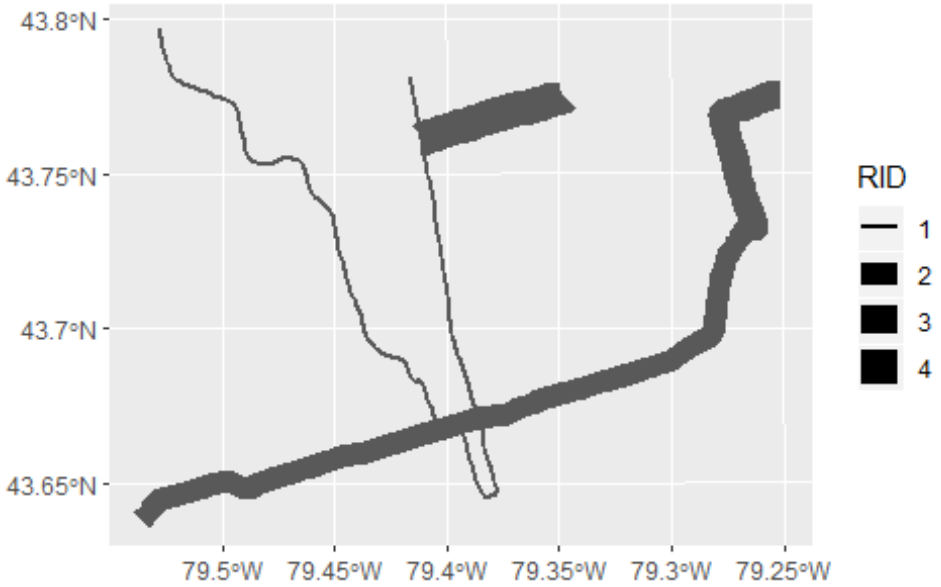
```
#Map setting the visual size (width) of the lines using a static value
#Map subways
ggplot(data = subway) +
  geom_sf(aes(size = 3), show.legend = "line")
```




```
#The show.legend = "line" piece changes the legend symbology from polygons to
line. Replacing "line" to "point" will use points in the legend
```

Symbology - sizes by attribute

```
#Map setting the visual size (width) of the lines using an attribute value
#Map subways
ggplot(data = subway) +
  geom_sf(aes(size = RID), show.legend = "line")
```

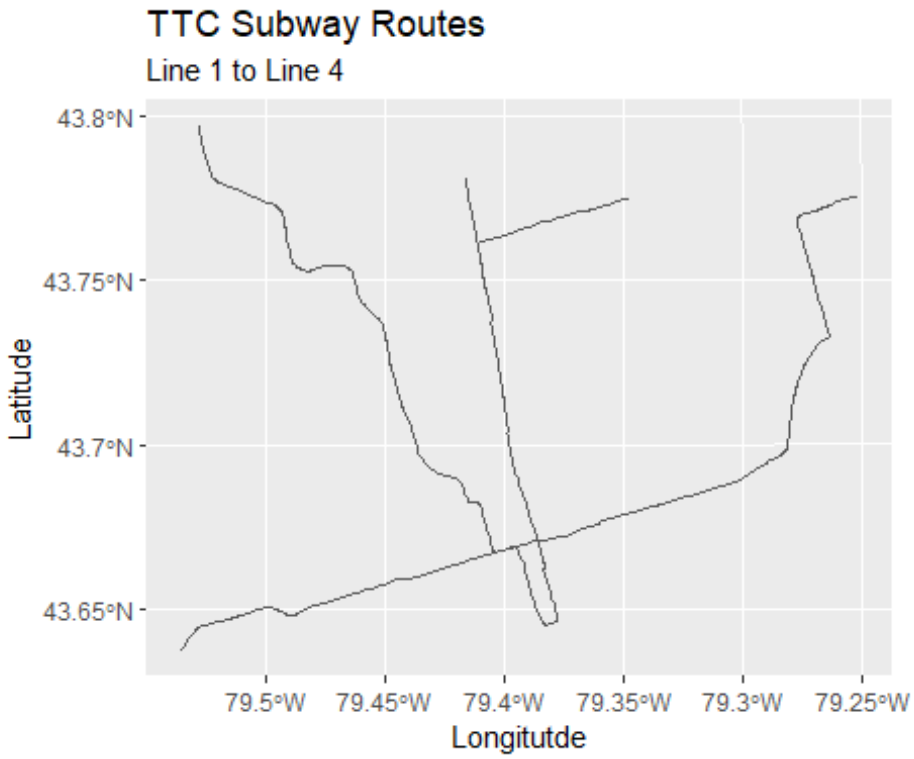



```
#The show.legend = "line" piece changes the legend symbology from polygons to
line. Replacing "line" to "point" will use points in the legend
```

4.3.4 Static Maps - Custom Symbology: Labels

Visualizations are a source of media to convey information that would be too complex for written text. The addition of labels to a visualization can improve the communication, providing context to the visualization.

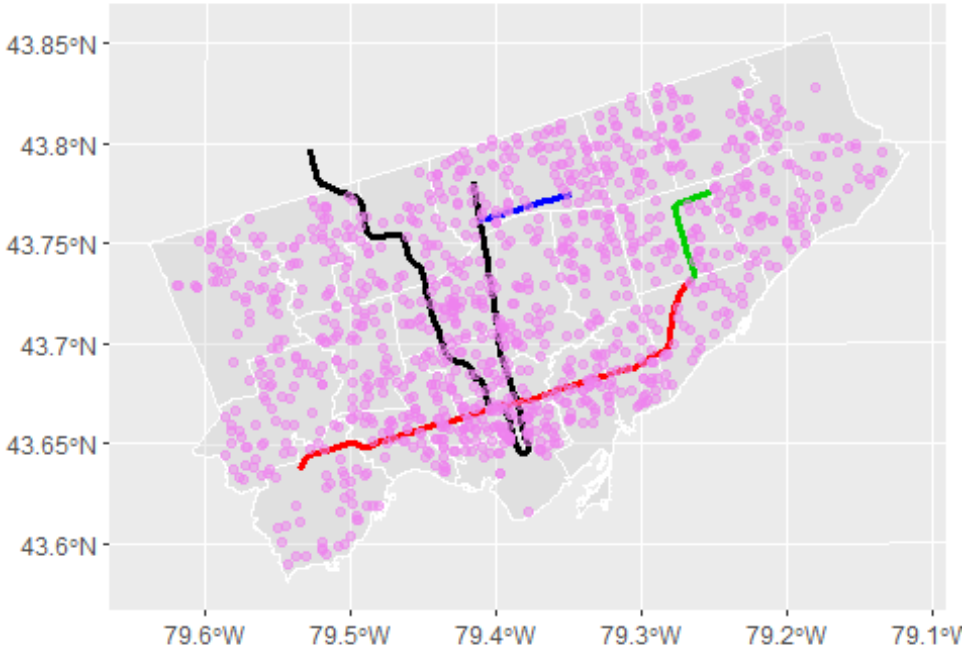
```
ggplot() +
  geom_sf(data = subway) +
  labs(title = "TTC Subway Routes",      #using labs function to add the plot
title
       subtitle = "Line 1 to Line 4") +  #using labs function to add a plot
subtitle
  labs(x = "Longitutde") +               #using labs function to add a label
on the x-axis
  labs(y = "Latitude")                   #using labs function to add a label
on the y-axis
```



4.3.5 Static Maps - Custom Symbology: Multiple vectors in one map

Rarely within location data analytics are visualizations created representing a single dataset. Typically multiple datasets are used in conjunction to illustrate the complexity of the analysis.

```
#Map with point, line and polygon vector together
ggplot() +
  geom_sf(data = wards, alpha = 0.25, fill = "grey", color = "white") +
  geom_sf(data = subway, color = subway$RID, size = 1.25) +
  geom_sf(data = schools, color = "violet", alpha = 0.5)
```



5.0 Geoprocessing

A significant component to location data analytics is intermediate stage, typically consisting of generating new datasets from existing data that will support analysis. There are many different geoprocessing operations, with varying parametres. This section will focus on one geoprocessing task that will be used in the next section, spatial analysis. For more geoprocessing examples see the sf cheatsheet.

Buffering

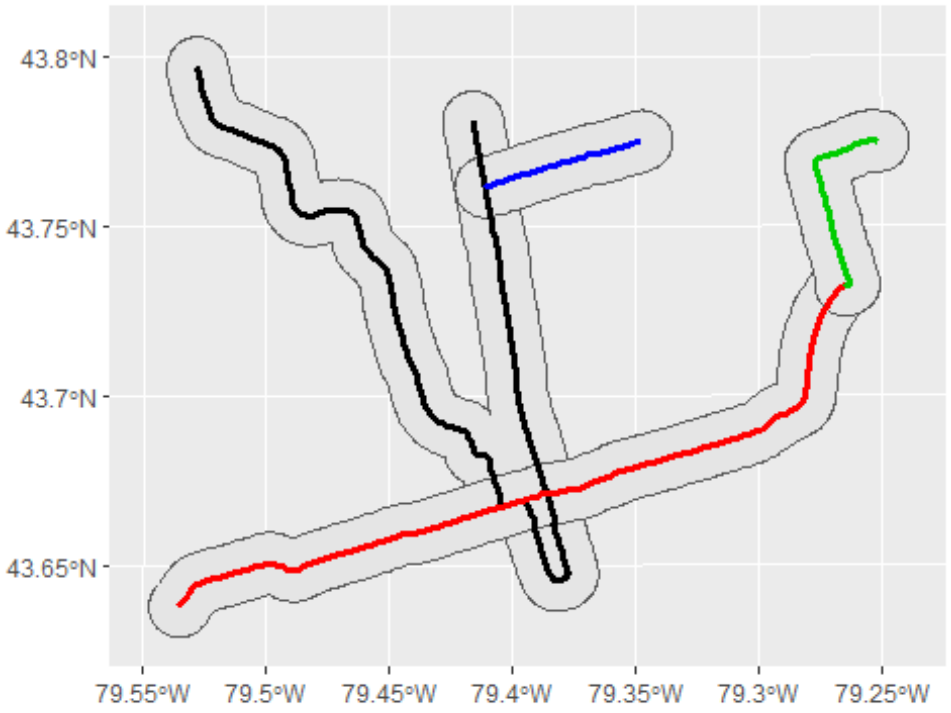
```
#Buffer
#Buffer subway lines - not storing geometry
st_buffer(subway, 1000)
```

```
## Simple feature collection with 4 features and 4 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: 300927.8 ymin: 4831875 xmax: 325823 ymax: 4851535
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
##   OBJECTID                ROUTE_NAME RID                       geometry
## 1    53420 LINE 1 (YONGE-UNIVERSITY)   1 POLYGON ((303600.7 4850208,...
## 2    53421 LINE 2 (BLOOR - DANFORTH)   2 POLYGON ((301261.9 4833636,...
## 3    53422      LINE 3 (SCARBOROUGH)   3 POLYGON ((321750.5 4847255,...
## 4    53423         LINE 4 (SHEPPARD)   4 POLYGON ((311751.3 4847603,...
##        length_m
```

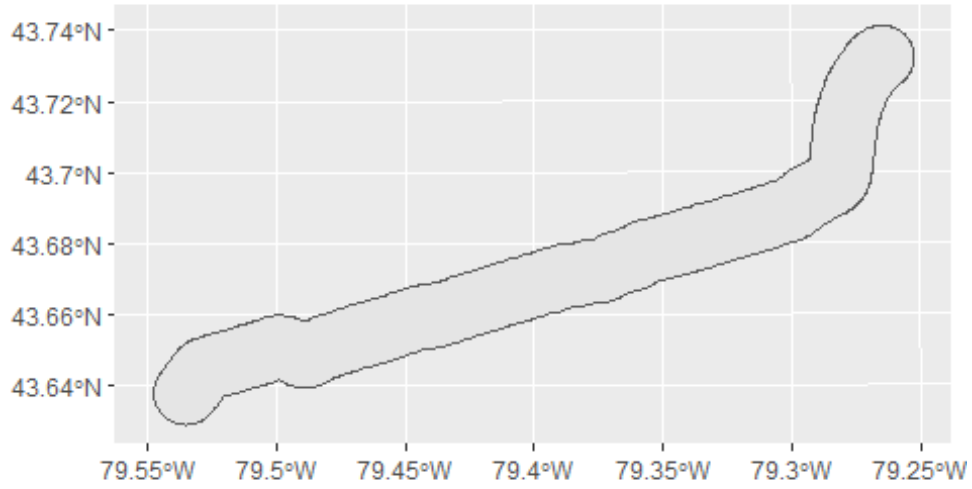
```
## 1 38894.062 [m]
## 2 26192.382 [m]
## 3  6622.532 [m]
## 4  5366.335 [m]
```

```
#Buffer subway lines - storing geometry
subway_buffer <- st_buffer(subway, 1000)

#Map of buffered subway lines
ggplot() +
  geom_sf(data = subway_buffer)
```

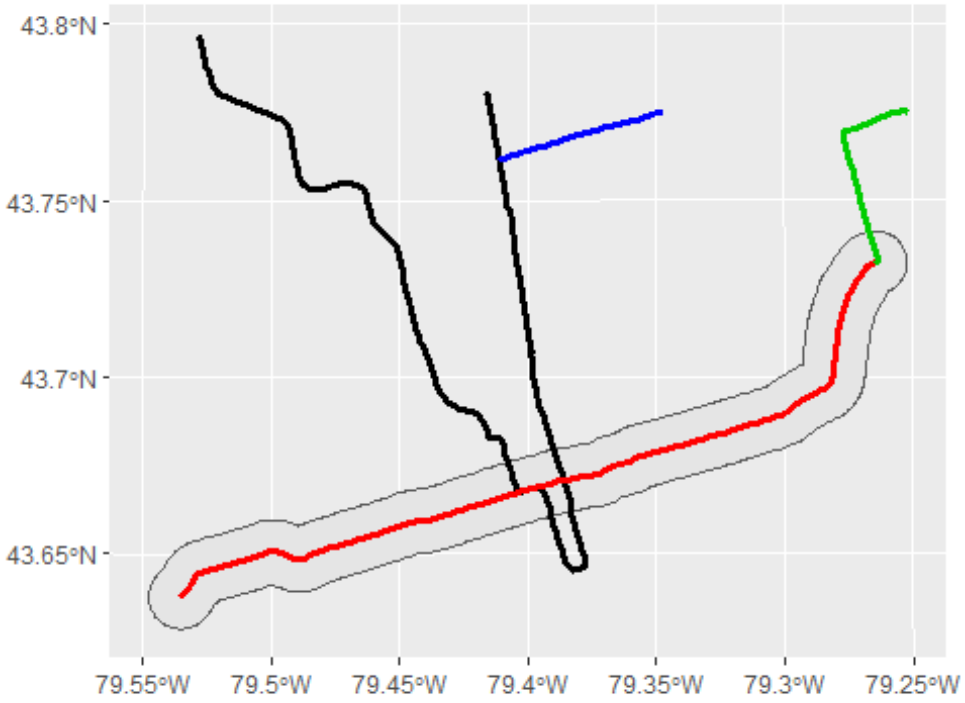

```
#Map of buffered subway lines with subway lines
ggplot() +
  geom_sf(data = subway_buffer) +
  geom_sf(data = subway, color = subway$RID, size = 1.25)
```




```r
#Buffer selected record
#Buffer subway line #2 (Bloor - Danforth)
subway_buffer_line2 <- subway %>%
  filter(RID == 2) %>%
  st_buffer(1000)

#Map of buffered subway line 2 (Bloor -Danforth)
ggplot() +
  geom_sf(data = subway_buffer_line2)
```




```
#Map of buffered subway line 2 (Bloor -Danforth) with subway lines
ggplot() +
  geom_sf(data = subway_buffer_line2) +
  geom_sf(data = subway, color=subway$RID, size = 1.25)
```



5.1 Spatial Analysis

Location data analytics is rooted in spatial analysis, leveraging tools and techniques for understanding the physical environment. Spatial analysis typically requires multiple datasets, and utilizes a combination of geoprocessing and analysis techniques. This section will focus on one spatial analysis problem – identifying the schools that are within 1km of a subway line. The example will provide two alternatives: one where new data layers are created, and a second where all analysis is performed in the computers memory. For more spatial analysis examples see the sf cheatsheet.

```
#Count points within polygon(s) - st_intersection
#Requires geometry to be created
count(st_intersection(schools, subway_buffer))
```

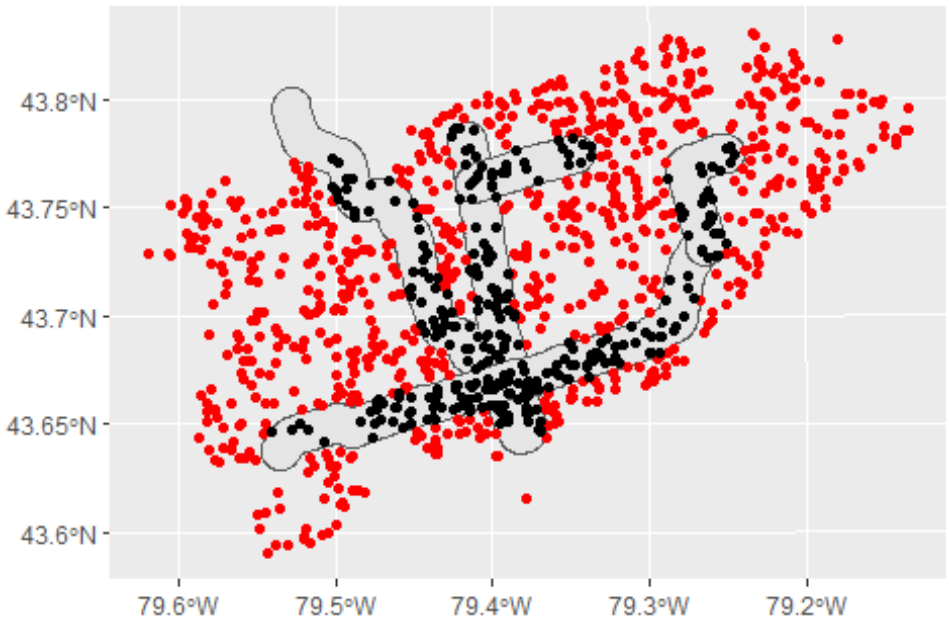
```
## Warning: attribute variables are assumed to be spatially constant
## throughout all geometries

## Simple feature collection with 1 feature and 1 field
## geometry type:  MULTIPOINT
## dimension:      XY
## bbox:           xmin: 301470.1 ymin: 4833389 xmax: 325324.6 ymax: 4849450
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## # A tibble: 1 x 2
##       n                                                          geometry
```

```
##   <int>                                                    <MULTIPOINT [m]>
## 1   404 (301470.1 4833842, 302529.5 4833950, 302985.4 4834239, 303305.7 4~
```

```
#Points in polygons - st_intersection
schoolsInBuffer <- st_intersection(schools, subway_buffer)
```

```
## Warning: attribute variables are assumed to be spatially constant
## throughout all geometries
```

```
#Map points that intersect with subway buffer
ggplot() +
  geom_sf(data = subway_buffer) +
  geom_sf(data = schools, color = "red") +
  geom_sf(data = schoolsInBuffer, color = "black")
```



Analysis without creating objects

```
#Count points within polygon(s) - st_intersection
#Does not require geometry to be created
count(st_intersection(schools, st_buffer(subway, 1000)))
```

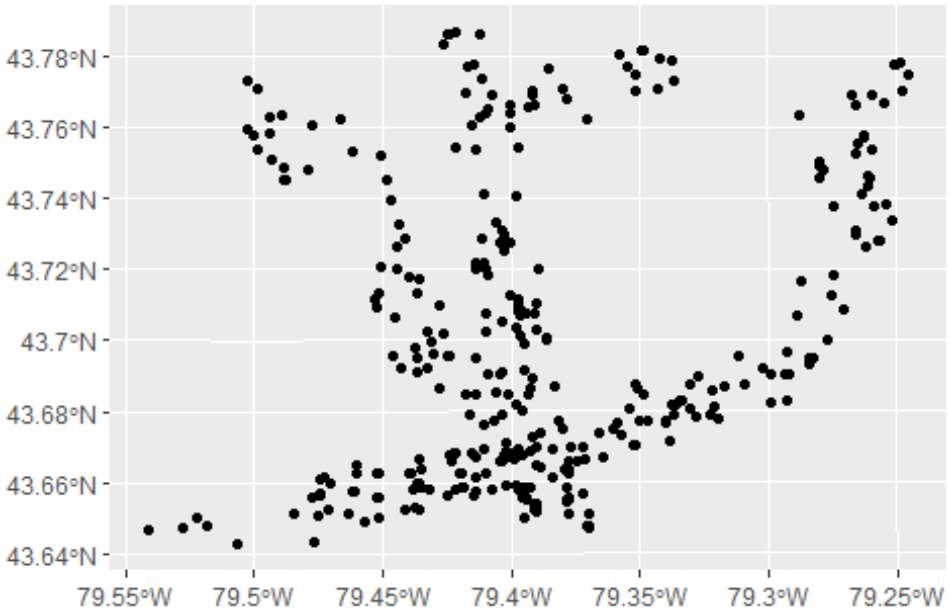
```
## Warning: attribute variables are assumed to be spatially constant
## throughout all geometries
```

```
## Simple feature collection with 1 feature and 1 field
## geometry type:  MULTIPOINT
## dimension:      XY
## bbox:           xmin: 301470.1 ymin: 4833389 xmax: 325324.6 ymax: 4849450
```

```
## epsg (SRID):    NA
## proj4string:    +proj=tmerc +lat_0=0 +lon_0=-79.5 +k=0.9999 +x_0=304800
+y_0=0 +datum=NAD27 +units=m +no_defs
## # A tibble: 1 x 2
##       n                                                          geometry
##   <int>                                                  <MULTIPOINT [m]>
## 1   404 (301470.1 4833842, 302529.5 4833950, 302985.4 4834239, 303305.7 4~
```

```
#Map points that intersect with subway buffer
#Does not require geometry to be created
ggplot() +
  geom_sf(data = st_intersection(schools, st_buffer(subway, 1000)))
```

```
## Warning: attribute variables are assumed to be spatially constant
## throughout all geometries
```



6.0 Resources

GIS in R

Spatial Data in R: New Directions

Geocomputation with R

R for Data Science

ggplot2 book

ggplot2 Colours Cookbook

Data visualization 101

sf cheatsheet