

# *Obtaining census spatial files*

*Bruno Santos & Antonio Paez*

*2025-07-03*

## *Introduction*

This Rmarkdown file is part of the **CommuteCA** package. This package was created in conjunction with the office of the *Research Data Center* at *McMaster University*, the *Sherman Centre for Digital Scholarship* and the *Mobilizing Justice*<sup>1</sup>.

The main objective of the *CommuteCA* package is to develop standardized methods for transport analysis in research, especially for studies using Statistics Canada surveys. Among the available surveys, we focused our efforts on the *2021 Census of Population*, which contain valuable variables for transportation research.

After learning the basics of **R**, the concepts of data and performing some exploratory data analysis, we moved on to the next steps to perform accessibility analysis. This R markdown aims to organize the census spatial data files required to perform our analysis. In this Rmarkdown, we'll demonstrate how to obtain the census spatial data files from the **cancensus** library.

## *Data*

To perform the next steps of our accessibility analysis, we need to download/obtain the census spatial files. More specifically, we need to have the following data sets:

- **Census Divisions:** Census division (CD) is the general term for provincially legislated areas (such as county, municipalité régionale de comté (MRC) and regional district) or their equivalents in Canada. Census divisions are intermediate geographic areas between the province/territory level and the municipality (census subdivision).
- **Dissemination areas:** A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons based on data from the previous Census of Population Program. It is the smallest standard geographic area for which all census data are disseminated. *Only for analysis at the Dissemination area's level.*
- **Census tracts:** Census tracts (CTs) are small, relatively stable geographic areas that usually have a population of fewer than 7,500 persons, based on data from the previous Census of Population

<sup>1</sup> The Mobilizing Justice project is a multidisciplinary and multi-sector collaboration with the objective of understand and address transportation poverty in Canada and to improve the well-being of Canadians at risk of transport poverty. The Social Sciences and Humanities Research Council (SSRHC) has provided funding for the project, which was created by an unprecedented alliance of academics from various Canadian provinces and institutions, transportation firms, and nonprofit organizations

Program. They are located in census metropolitan areas (CMAs) and in census agglomerations (CAs) that had a core population of 50,000 or more in the previous census.

There are two ways to obtain these data sets: by downloading the shapefile directly from the 2021 Census of Population’s website, or by accessing through the `cancensus` library.

We will prepare our data set using the `cancensus` library, but you can also access the data directly from the Census website, by clicking in this link.

For `Census Divisions`, you will need to:

- Set the *language* of file (English or French).
- Define the *type* as *Cartographic Boundary Files (CBF)*.
- In *Administrative boundaries*, set *Census divisions*.
- Define the *format* as *Shapefile (.shp)* and click and *Continue*.
- Click in the link to download file.

For `Census tracts and Dissemination areas`, you will need to:

- Set the *language* of file (English or French).
- Define the *type* as *Cartographic Boundary Files (CBF)*.
- In *Statistical boundaries*, set *Census tracts and Dissemination areas*.
- Define the *format* as *Shapefile (.shp)* and click and *Continue*.
- Click in the link to download file.

### *Obtaining data files using the `cancensus` library*

First, you need install the `cancensus` library [von Bergmann et al., 2022]:

```
library('cancensus')
library('geojsonsf')
library('sf')
library('dplyr')
library('here')
```

`cancensus` requires a valid `CensusMapper` API key to use. You can obtain a free API key by signing up for a `CensusMapper` account. To check your API key, just go to “Edit Profile” (in the top-right of the `CensusMapper` menu bar). Once you have your key, you can store it in your system environment so it is automatically used in API calls. To do so just enter `set_cancensus_api_key('<your_api_key>', install = TRUE)`.

```
#set_cancensus_api_key('<your API ket>', install = TRUE, overwrite=TRUE)
```

You can visualize all available data set in the library by doing:

```
list_census_datasets()

## # A tibble: 29 x 6
##   dataset description      geo_dataset attribution reference reference_url
##   <chr>    <chr>          <chr>      <chr>      <chr>      <chr>
## 1 CA1996  1996 Canada Census  CA1996      StatCan 19~ 92-351-U https://www1~
## 2 CA01    2001 Canada Census  CA01        StatCan 20~ 92-378-X https://www1~
## 3 CA06    2006 Canada Census  CA06        StatCan 20~ 92-566-X https://www1~
## 4 CA11    2011 Canada Census  CA11        StatCan 20~ 98-301-X~ https://www1~
## 5 CA16    2016 Canada Census  CA16        StatCan 20~ 98-301-X https://www1~
## 6 CA21    2021 Canada Census  CA21        StatCan 20~ 98-301-X https://www1~
## 7 CA01xSD 2001 Canada Census  CA01        StatCan 20~ 92-378-X https://www1~
## 8 CA06xSD 2006 Canada Census  CA06        StatCan 20~ 92-566-X https://www1~
## 9 CA11xSD 2011 Canada Census  CA11        StatCan 20~ 98-301-X https://www1~
## 10 CA16xSD 2016 Canada Census  CA16        StatCan 20~ 98-301-X https://www1~
## # i 19 more rows
```

For this demonstration, we will access the data set related to the City of Toronto (Census Divisions (CD) = 3520).

```
code <- "3520" # As a character! you can change for your city code
```

In our case, we are interested on the Census Divisions:

```
census_divisions <- get_census(dataset='CA21', regions=list(CSD=code), level='CSD',
                             geo_format = "sf")
```

```
## Downloading: 11 kB      Downloading: 11 kB      Downloading: 11 kB      Downloading: 11 kB
```

Filtering and adjusting variable names:

```
census_divisions <- census_divisions %>%
  dplyr::select(GeoUID, name, PR_UID, `Shape Area`, geometry) %>%
  dplyr::rename(CDUID = GeoUID,
               CDNAME = name,
               PRUID = PR_UID)

plot(census_divisions["geometry"])
```



Census tracts (it is important to remember that if your city of analysis is not within a Census Agglomeration or Census Metropolitan Area, you will not have census tract available for your city!):

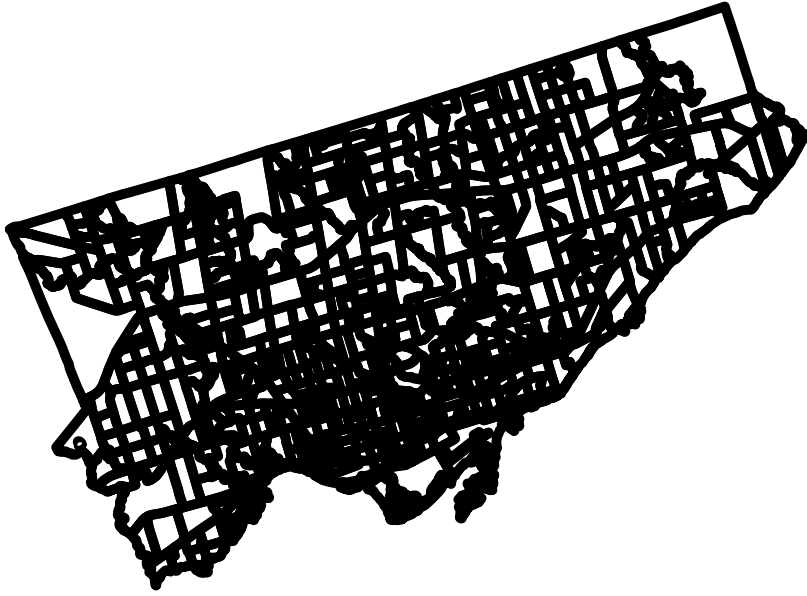
```
census_tracts <- get_census(dataset='CA21', regions=list(CSD=code), level='CT', geo_format = "sf")
```

```
## Downloading: 3.6 kB      Downloading: 3.6 kB      Downloading: 12 kB      Downloading: 12 kB      Downloading: 12 kB
```

Filtering and adjusting variable names:

```
census_tracts <- census_tracts %>%
  dplyr::select(GeoUID, name, `Shape Area`, CD_UID, CMA_UID, geometry) %>%
  dplyr::rename(CTUID = GeoUID,
               CTNAME = name,
               PCD = CD_UID,
               LANDAREA = `Shape Area`,
               CMAUID = CMA_UID) %>%
  mutate(PRUID = substr(CMAUID, start = 1, stop = 2),
         CMAUID = substr(CMAUID, start = 3, stop = 5)) %>%
  dplyr::select(CTUID, CTNAME, LANDAREA, PCD, CMAUID, PRUID, geometry)

plot(census_tracts["geometry"])
```



And Dissemination areas:

```
dissemination_areas <- get_census(dataset='CA21', regions=list(CSD=code), level='DA', geo_format = "sf")
```

```
## Downloading: 32 kB      Downloading: 32 kB      Downloading: 32 kB      Downloading: 32 kB      Downloading: 32 kB
```

Filtering and adjusting variable names:

```
dissemination_areas <- dissemination_areas %>%
  dplyr::select(GeoUID, `Shape Area`, CD_UID, CT_UID, CMA_UID, geometry) %>%
  dplyr::rename(DAUID = GeoUID,
                PCD = CD_UID,
                LANDAREA = `Shape Area`,
                CMAUID = CMA_UID,
                CTUID = CT_UID) %>%
  mutate(PRUID = substr(CMAUID, start = 1, stop = 2),
         CMAUID = substr(CMAUID, start = 3, stop = 5)) %>%
  dplyr::select(DAUID, LANDAREA, PCD, CTUID, CMAUID, PRUID, geometry)

plot(dissemination_areas["geometry"])
```



### *Saving the files*

We will now download these datasets as shapefiles for use in the next steps. First, we will create a folder within the `raw-data` directory using the city code of our study area. This organization makes it easier to manage files if we decide to analyze multiple study areas.

```
directory <- paste0(here(), "/data-raw/output/PCD", code, "/spatial-files/")
```

```
if(!dir.exists(directory)){
  dir.create(directory, recursive = TRUE)}
```

Saving the files:

```
sf::st_write(census_divisions, paste0(directory, "census_divisions.shp"), append=FALSE)
```

```
## Deleting layer `census_divisions' using driver `ESRI Shapefile'
```

```
## Writing layer `census_divisions' to data source
```

```
## `C:/Bruno/CommuteCA/data-raw/output/PCD3520/spatial-files/census_divisions.shp' using driver `ESRI Shapefile'
```

```
## Writing 1 features with 4 fields and geometry type Multi Polygon.
```

```
sf::st_write(census_tracts, paste0(directory, "census_tracts.shp"), append=FALSE)
```

```
## Deleting layer `census_tracts' using driver `ESRI Shapefile'
```

```
## Writing layer `census_tracts' to data source
```

```
## `C:/Bruno/CommuteCA/data-raw/output/PCD3520/spatial-files/census_tracts.shp' using driver `ESRI Shapefile'
```

```
## Writing 585 features with 6 fields and geometry type Multi Polygon.
```

```
sf::st_write(dissemination_areas, paste0(directory, "dissemination_areas.shp"), append=FALSE)
```

```
## Deleting layer `dissemination_areas' using driver `ESRI Shapefile'  
## Writing layer `dissemination_areas' to data source  
##   `C:/Bruno/CommuteCA/data-raw/output/PCD3520/spatial-files/dissemination_areas.shp' using driver `E  
## Writing 3743 features with 6 fields and geometry type Multi Polygon.
```

## *References*

Jens von Bergmann, Dmitry Shkolnik, and Aaron Jacobs. *cancensus*:  
*R package to access, retrieve, and work with Canadian Census data*  
*and geography*, 2022. URL [https://mountainmath.github.io/](https://mountainmath.github.io/cancensus/)  
[cancensus/](https://mountainmath.github.io/cancensus/). R package version 0.5.7.