Taxdata Tables for Index Generation

2019-05-17

Workflow for economic index generation:

The workflow for generating an economic index consists of devising metrics based on available StatCan annual taxdata at individual and family levels. The following document contains fields from individual tables and their usage for index generation.

The data field considerations for generating a simple index include:

1. levels of geography: The levels of geography are divided into 14 categories as listed here:

- 3 FSA: (URBAN FSA)

- 6 RURPC: (RURAL POSTAL CODE AREAS)
- 7 OUA: (OTHER URBAN AREAS)
- 8 CITY: (CITY TOTAL)
- 9 RURC: (RURAL COMMUNITIES)
- 10 OPROV: (OTHER PROVINCIAL TOTAL)
- 11 PROV: (PROVINCE/TERRITORY TOTAL)

- 12 CANADA: (CANADA)
- 21 CD: (CENSUS DIVISION)
- 31 FED: (FEDERAL ELECTORAL DISTRICT)
- 51 ER: (ECONOMIC REGION)
- 41 CMA: (CENSUS METROPOLITAN AREA)
- 42 CA: (CENSUS AGGLOMERATION)
- 61 CT: (CENSUS TRACT)

Based on initial scans of data and maps, the levels of geography that are usable for this project include those that are small enough to hold the smallest number of taxfilers in a region and is inclusive of all of BC. The challenge here is that not one geography can be chosen and so economic indices needs to be generated for rural and urban communities separately. Another challenge is that a same city, e.g. Abbotsford, can repeat multiple times with the same level of geography but different postal regions. Since a cityID is also provided, we need to roll up the taxfilers information based on both level of geography and common cityIDs.

Levels of geography that can be used include: #9 for RURAL COMMUNITIES and #61 for CENSUS TRACT. RURAL POSTAL CODE AREAS (#6) is also usable as they are small enough and exclusive of rural communities. CENSUS DIVISION (#21) also provides small enough regions <50,000 which is exclusive of census tracts.

2. cityID: Since levels of geography and postal names can have repetitions, we would need to consider names of places, geography and postal codes, as well as cityID to roll-up the taxfilers' information into one row per region of interest.

3. quintile ratio: The simplest strategy for quintile ratio generation is dividing the income into 5 brackets. Currently, the income is divided in <5k+, 5k+, 10k+, 15k+, 20k+, 25k+, 35k+, 50k+, 75k+, 75k+, 100k+, 150k+, 200k+, 250k+. Therefore, a ratio can vary depending on the regions of geography and the normal distribution of taxfilers in each region. The index will vary depending on the distribution of taxfilers based on year and region.

Milestones:

We have understood the complexities of the taxdata and the need for clarifying the business needs of partner ministries prior to providing the relevant indices. The codes will be shared within a centralized github repository and shared with ministry partners as needed prior to July 2019. We will spend the month of June-July 2019 on fine-tuning the indices for both individual and family tables and share on a centralized github page.

Deliverables:

The codes for generating simple economic index will be shared on the github page. Once a strategy for individuals is confirmed, a similar document will be generated for family tables.

Stakeholder interactions:

Any ministry partner who is interested in economic indices for their projects will require an understanding of the geographic regions of the province and year(s) of interest. These ministries need to consult with DPDD about their own data ranges in terms of years of data availability, geographical concepts in their own datasets, and the type of index they are looking to create.

The simplest linkage will include use of table #1 from taxdata. This table holds information about taxfilers and dependents, summary table, income and demographics of individuals. Table #5 is the next simplest table for more detailed index generation that sheds light on gender and age distribution as well as income. Tables #6-12 include additional complexities and details that can be used depending on ministry needs.

Popdata will likely not be affected by the changing demands of ministry partners after the first linkage is supported. However, Popdata will also need to know the needs of ministry as far as it is concerned with the usage of tables, in order to use the right table for geographical concept conversions.

The index code will be stored in a central repository such that any data scientist could run a similar iteration of it depending on need.

Appendix I:

Individual Taxdata Tables structures:

Table I-01

This table holds information about taxfilers and dependents, summary table, income and demographics of individuals.

```
ind_1 <- fread(here("input-data", "1_IND.csv"))

Fields relevant to index:

postal|area|
year
cityid|
level|of|geo|
place|me|geo|
```

```
taxfilers|#/
total|income|median|males
total|income|median|females
total|income|median|total
```

Table I-02

This table includes information on taxfilers and dependents by sex, marital status and age groups. There are **no** fields for index generation, directly.

```
ind_2 <- fread(here("input-data", "2_IND.csv"))
```

Table I-3A-B-C

These tables includes information on male (3A) and female (3B) and combined (3C) taxfilers and dependents by single years of age. There are **no** fields for index generation, directly.

```
ind_3A <- fread(here("input-data", "3A_IND.csv"))
ind_3B <- fread(here("input-data", "3B_IND.csv"))
ind_3C <- fread(here("input-data", "3C_IND.csv"))
```

Table I-04

This table includes information on taxfilers and dependents with income by source of income. There are **no** fields for index generation, directly.

```
ind_4 <- fread(here("input-data", "4_IND.csv"))
```

Table I-5A-B-C

These tables include male, female, and all taxfilers and dependents with income by total Income and age groups. The breakdowns for age and income can help with index generation.

```
ind_5A <- fread(here("input-data", "5A_IND.csv"))
ind_5B <- fread(here("input-data", "5B_IND.csv"))
ind_5C <- fread(here("input-data", "5C_IND.csv"))

Fields of interest include:

postal|area|
cityid|
level|of|geo|
place|me|
  ## taxfilers in the brackets below:
  all|males|with|income|by|total|income|0-24|years|<$5k
    all|males|with|income|by|total|income|0-24|years|$5k+
    all|males|with|income|by|total|income|0-24|years|$10k+
    all|males|with|income|by|total|income|0-24|years|$15k+
    all|males|with|income|by|total|income|0-24|years|$20k+
    all|males|with|income|by|total|income|0-24|years|$25k+
    all|males|with|income|by|total|income|0-24|years|$35k+
    all|males|with|income|by|total|income|0-24|years|$50k+
```

```
    all|males|with|income|by|total|income|0-24|years|$75k+
    all|males|with|income|by|total|income|0-24|years|$100k+
    all|males|with|income|by|total|income|0-24|years|$150k+
    all|males|with|income|by|total|income|0-24|years|$200k+
    all|males|with|income|by|total|income|0-24|years|$250k+
    all|males|with|income|by|total|income|0-24|years|total

- Other age brackets include:
  25-34, 35-44, 45-54, 55-64, 65-74, 75+

- This information also exist for females and both genders combined.
```

Table I-06

Taxfilers and dependents with income by sex, income taxes, selected deductions and benefits.

```
ind_6 <- fread(here("input-data", "6_IND.csv"))

postal|area|
year
cityid|
level|of|geo|
place|me|

total|income|$'000|males
total|income|$'000|females
total|income|$'000|total
```

Table I-7A-B-C

This table includes information on male, female, and all taxfilers and dependents with income by after-tax income and age groups

```
ind_7 <- fread(here("input-data", "7_IND.csv"))

postal|area|
year
cityid|
level|of|geo|
place|me|
    ## taxfilers in the brackets below:
  all|males|with|after-tax|income|by|total|income|0-24|years|<$5k
  all|males|with|after-tax|income|by|total|income|0-24|years|$5k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$10k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$15k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$20k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$25k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$35k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$50k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$75k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$100k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$150k+
  all|males|with|after-tax|income|by|total|income|0-24|years|$200k+
```

```
  all|males|with|after-tax|income|by|total|income|0-24|years|$250k+
  all|males|with|after-tax|income|by|total|income|0-24|years|total

Other age brackets include:
  25-34, 35-44, 45-54, 55-64, 65-74, 75+

- This information also exist for females and both genders combined.
```

Table I-08

This table includes information on taxfilers and dependents with income by income taxes and after-tax income, sex and age groups.

```
ind_8 <- fread(here("input-data", "8_IND.csv"))

postal|area|
year
cityid|
postal|walk|
level|of|geo|
place|me|
  ## taxfilers in the brackets below:
  all|males|total|income|#|0-24|years
  all|males|total|income|#|$'000|0-24|years

- Other age brackets include:
  25-34, 35-44, 45-54, 55-64, 65+

- This information also exist for females and both genders combined.
```

Table I-09

This table includes economic dependency profile of individuals.

```
ind_9 <- fread(here("input-data", "9_IND.csv"))

postal|area|
year
cityid|
level|of|geo|
place|me|
empl.|income|#|males
empl.|income|#|females
empl.|income|#|total
empl.|income|$'000|males
empl.|income|$'000|females
empl.|income|$'000|total

- Other sources of income include:
- government and non-government transfers.
```

Table I-10

Labour income profile of individuals in this table can be used for index generation:

```
ind_10 <- fread(here("input-data", "10_IND.csv"))
postal|area|
year
cityid|
level|of|geo|
place|me|
all|persons|males|
all|persons|females|
all|persons|total|
taxfilers|males|
taxfilers|females|
taxfilers|total|
total|income|number|males
total|income|number|females
total|income|number|total
total|income|$'000|males
total|income|$'000|females
total|income|$'000|total
total|income|median|males
total|income|median|females
total|income|median|total

- Other sources of income include:
labour, employment, wages, salaries, self employment, income,
```

Table I-11

Table including taxfilers and dependents at 15 years of age and over with labour income by sex and age groups. Fields of interest include:

```
ind_11 <- fread(here("input-data", "11_IND.csv"))
postal|area|
year
cityid|
level|of|geo|
place|me|

- Other sources of income include:
- female and male age ranges within income brackets
```

Table I-12

Table includes taxfilers and dependents at 15 years of age and over receiving employment insurance by age groups and sex. Fields of interest include:

```
ind_12 <- fread(here("input-data", "12_IND.csv"))

postal|area|
year
```

```
cityid|
level|of|geo|
place|me|

- Other sources of income include:
- labour, employment insurance income by age and sex categories
```

Table I-13

Table includes after-tax low income status of tax filers and dependents, census family low income measure, for couple and lone parent families by family composition. There are **no** fields for index generation, directly.

```r
ind_13 <- fread(here("input-data", "13_IND.csv"))
```

Appendix II:

Income inequality in Canada is defined in more detail, here. Based on this, the average income in Canada stands at 44000 Canadian dollars. The total income explorer released by Statistics Canada provides additional measures, here, and here. Additional socioeconomic reports can be found, here.