Analysis report exercise

Michael Tang

05/07/2020

The Atlantic region might not have a significant population size (around 2.3 million measured in 2011), yet it remains an integral part of Canada that we should not overlook. Using the 2011 Census data, we looked at the characteristics and composition of families and households in that region. Below are some major findings that could be important to the client,

- 46.4% of couple (married + common-law) families had children living at home
- 55.2% of census families had a household size of 2 persons
- 16.4% of the total census families were lone-parent families, of which 80.1% were mother-parent families
- 17.2% of the total census couple families were common-law couples, of which 42.2% had children living at home
- Children under 18 accounted for 68.0% of the total children in census families
- Average number of persons per household was 2.36

Evidently, it would be more likely to find families living in 2-person households with less 50% chance of finding couples who had children living with them at home, Also, a significant portion of lone-parent families were maintained by mothers. With only data from 2011, it was not possible to deduce how the percentage of common-law couples within the total census couple population would evolve with time. Nonetheless, close to half of the common-law couples had children living with them at home in 2011. In addition, 60% of the children population were under 18, suggesting that there could be some demand for children's products, school supplies, or day care services. Depending on what industry the client is working on, the data above could offer some insights on the way that business strategies could be drafted.

The statistics were obtained by first trimming the input data down to the relevant columns interested, then assigning an unique id to each row of the data frame, which was then transformed from long to wide format.

```
# select data under topic of "family characteristics",
# then use subset to get data in provinces in the Atlantic region,
# get columns interested (i.e. prov., charac., and total)
# assign trimmed data frame to lfamtable
lfamtable <- inputdata %>% .[.$"Topic" == "Family characteristics", ] %>%
                subset(., .$"Prov_Name" %in% c("New Brunswick", "Nova Scotia",
                        "Prince Edward Island", "Newfoundland and Labrador")) %>%
                          .[, names(.) %in% c("Prov_Name", "Characteristic", "Total")]

# get the number of rows for each province, call it len_ent
len_ent = length(lfamtable[lfamtable$"Prov_Name" == "Newfoundland and Labrador",
                           "Prov_Name"])

# create a list that goes from 1 to the end of list len_ent
list_ent = 1:len_ent

# as there are 4 provinces interested, replicate the list 4 times
```

```r
# to generate a set of Ids for rows in each province
id = rep(list_ent, times = 4)

lfamtable$"Id" <- id

# use reshape() to transform lfamtable to wide format
wfamtable <- reshape(lfamtable, idvar = c("Id", "Characteristic"), timevar = "Prov_Name",
                     v.names = "Total", direction = "wide")
```

Then, useful information embedded in the data was conveniently extracted by using the id assigned to the specific rows. For example, to obtain the percentage of couple families who have children at home, we could perform the following calculations to arrive at the results

```r
# get sum of number of married couples with children at home (id = 10)
# get sum of number of common-law couples with children at home (id = 16)
# get number of couple families having children at home
numCouFamChildHome = sum(wfamtable[10, 3:6]) + sum(wfamtable[16, 3:6])

# get total number of couple families (id = 6)
totalNumFCouFam = sum(wfamtable[7, 3:6])

# get percentage of couple families having children living at home
perCouFamChildHome = (numCouFamChildHome/totalNumFCouFam)*100
```

Similarly, other calculations were conducted using ids on the relevant rows.