Occupational Table Explaratory Data Analysis

Summary

Stina sent two datasets. One is occupational data for all Canadian immigrants, from 1980 to 2016 (10 and 5 year increments), while the other is a series of archived web-only tables (prepared datasets) for the Haitian Canadian community specifically. I will explore the former here. I don't want to resort to web-scraping for the latter. So I will try to gather the Haitian Canadian datasets that underlie the tables with an API in another notebook.

I tried to download the raw data, rather than the table data, for this dataset. Unfortunately, downloading 1gb+ failed multiple times. I will look into gathered the equivalent data with an API in the other notebook.

I was able to wrangle the table data into some sample visualizations without styling (below). It will be relatively easy now to put any of the table data into this form. Moreover, this kind of visualization is probably among the best for this table data.

Data import

The csv file for the table that Stina linked had additional information that needed to be remvoed to be parsed correctly:

"Occupation - National Occupational Classification (NOC) 2016 (693A), Highest Certificate, Diploma or Degree (15), Admission Category and Applicant Type (31), Period of Immigration (7), Age (5A) and Sex (3) for the Immigrant Population Who Landed Between 1980 and 2016 Aged 15 Years and Over, in Private Households of Canada, Provinces and Territories, 2016 Census - 25% Sample Data"

Different views with filtering

```
# Top-level categories
imm_occ %>%
  filter(str_detect(code, "\\b[1-9]{1}\\b")) %>%
  select(c("code", "occupation"))
```

```
## # A tibble: 17 x 2
##    code  occupation
##    <chr> <chr>
##  1 1     "Business, finance and administration occupations"
##  2 2     "Natural and applied sciences and related occupations"
##  3 3     "Health occupations"
##  4 4     "Occupations in education, law and social, community and government se~
##  5 5     "Occupations in art, culture, recreation and sport"
##  6 6     "Sales and service occupations"
##  7 7     "Trades, transport and equipment operators and related occupations"
##  8 8     "Natural resources, agriculture and related production occupations"
##  9 9     "Occupations in manufacturing and utilities"
## 10 [1]   "Excludes census data for one or more incompletely enumerated Indian r~
```

```
## 11 [2]   "'Admission category' refers to the name of the immigration program or~
## 12 [2]   "Users are advised to consult data quality comments for 'Highest certi~
## 13 [3]   "Period of immigration refers to the period in which the immigrant fir~
## 14 [4]   "Includes immigrants who landed in Canada on or prior to May 10, 2016."
## 15 [5]   "Refers to the kind of work performed by persons aged 15 years and ove~
## 16 [6]   "Includes persons aged 15 years and over who never worked for pay or s~
## 17 [7]   "Includes persons aged 15 years and over who have worked at some point~
```

```
# EX: "Occupations in manufacturing and utilities"
imm_occ %>%
  filter(str_detect(code, "^9[1-9]$")) %>%
  select(c("code", "occupation"))
```

```
## # A tibble: 4 x 2
##   code  occupation
##   <chr> <chr>
## 1 92    Processing, manufacturing and utilities supervisors and central control~
## 2 94    Processing and manufacturing machine operators and related production w~
## 3 95    Assemblers in manufacturing
## 4 96    Labourers in processing, manufacturing and utilities
```

```
# EX: "Processing, manufacturing and utilities supervisors and central control operators"
imm_occ %>%
  filter(str_detect(code, "^92[1-9]$")) %>%
  select(c("code", "occupation"))
```

```
## # A tibble: 4 x 2
##   code  occupation
##   <chr> <chr>
## 1 921   Supervisors, processing and manufacturing occupations
## 2 922   Supervisors, assembly and fabrication
## 3 923   Central control and process operators in processing and manufacturing
## 4 924   Utilities equipment operators and controllers
```

```
imm_trans <- imm_occ %>%
  filter(str_detect(code, "751")) %>%
  select(matches("\\d"))
```

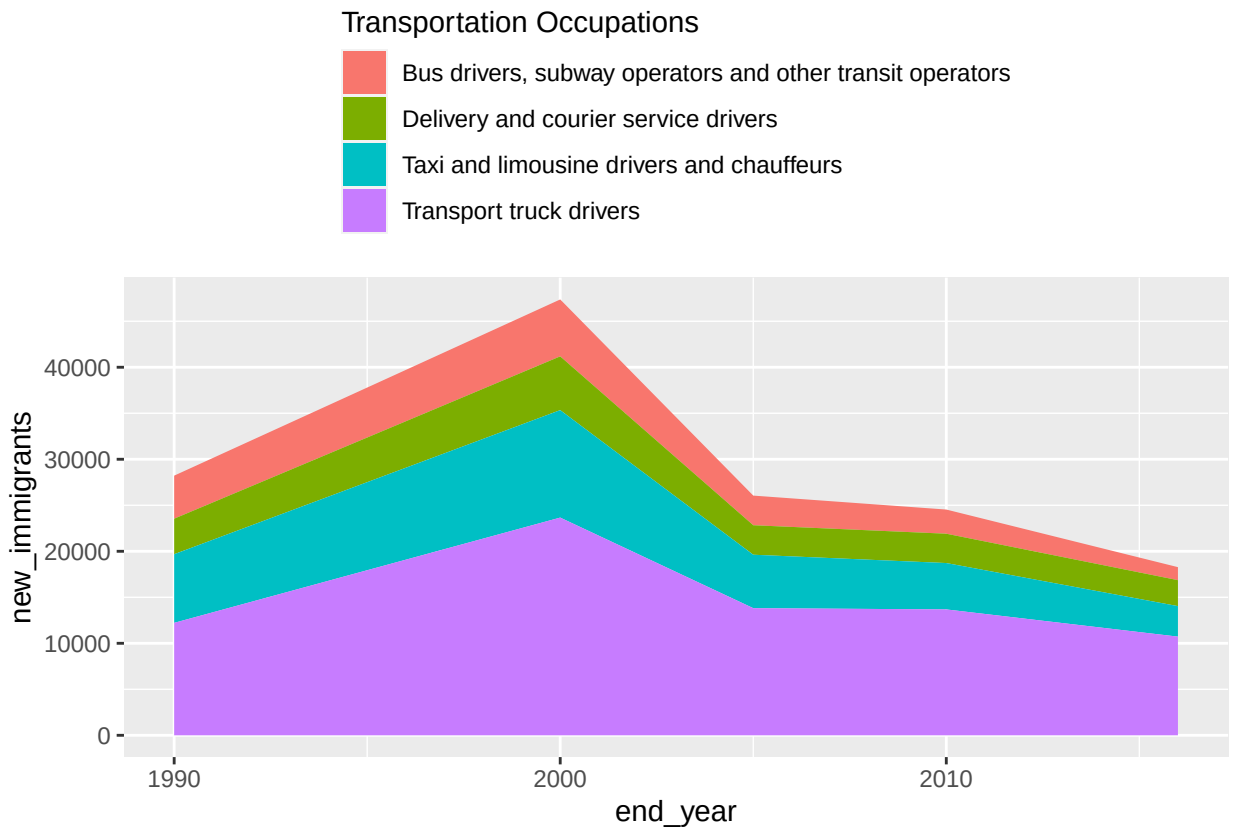
Demo Visualization

Work in transportation over time among new immigrants to Canada.

```
# get transport occupations per year
imm_trans <- imm_occ %>%
  filter(str_detect(code, "751"))

imm_trans[-1,] %>% # remove total transport occupations row
  select(-"total")  %>% # remove total years immigration column
  pivot_longer(!c("code", "occupation"),
               names_to = "years",
               values_to = "new_immigrants") %>%
  separate(years,
           into = c("start_year", "end_year"),
           sep = "_",
           convert = TRUE) %>%
  ggplot(aes(x = end_year, y = new_immigrants, fill = occupation)) +
```

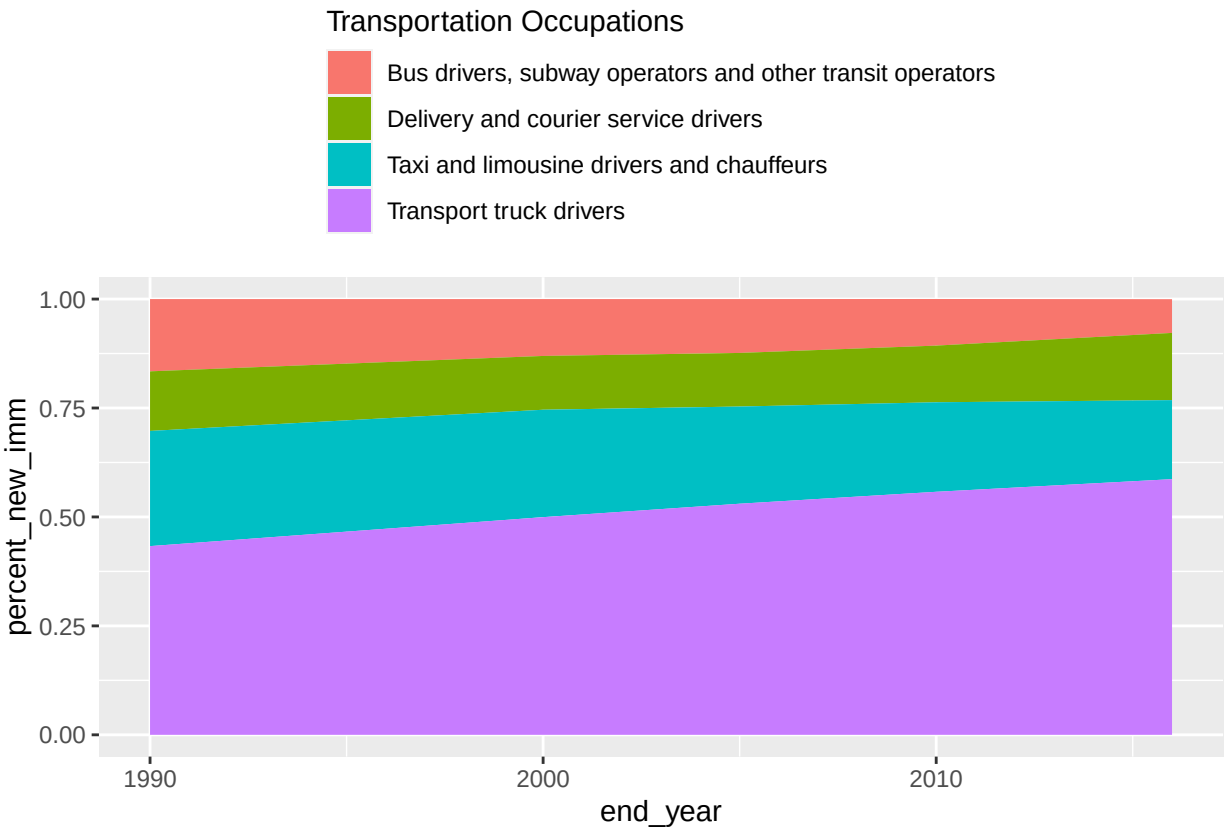
```
  geom_area() +
  theme(legend.position="top") +
  guides(fill = guide_legend(title = "Transportation Occupations",
                             direction = "vertical"))
```



Work in transportation over time among new immigrants to Canada, percentages.

```
imm_rel <- bind_cols(imm_trans[c("code", "occupation", "total")], imm_trans %>%
             transmute(across(matches("\\d"), ~ ./ .[1])))

imm_rel[-1,] %>% # remove total transport occupations row
  select(-"total")  %>% # remove total years immigration column
  pivot_longer(!c("code", "occupation"),
               names_to = "years",
               values_to = "percent_new_imm") %>%
  separate(years,
           into = c("start_year", "end_year"),
           sep = "_",
           convert = TRUE) %>%
  ggplot(aes(x = end_year, y = percent_new_imm, fill = occupation)) +
  geom_area() +
  theme(legend.position="top") +
  guides(fill = guide_legend(title = "Transportation Occupations",
                             direction = "vertical"))
```


Transportation Occupations
Bus drivers, subway operators and other transit operators
Delivery and courier service drivers
Taxi and limousine drivers and chauffeurs
Transport truck drivers
1.00
0.75
0.50
0.25
0.00
percent_new_imm
1990
2000
2010
end_year