

TP2: Étude de l'effet du taux de chômage sur l'appui à l'immigration au Canada : Comparaison entre 2019 et 2021 (Option 1)

Svetlana Zhuk

2024-02-04

Introduction

Les opinions publiques sur l'immigration évoluent au fil du temps. Si le Canada est généralement connu pour ses attitudes positives à l'égard de l'immigration, même ici, ces attitudes varient. L'explication de ces changements est un domaine de recherche actif, mais les facteurs sociétaux ou économiques à l'origine de cette évolution ne sont toujours pas clairs. Par exemple, Mayda (2006) affirme que les attitudes anti-immigration peuvent être rationalisées et sont souvent enracinées dans la compétition pour des ressources rares avec les immigrants [Mayda, 2006](#). Dans ce travail, j'explore une question connexe, celle de l'influence économique sur la perception de l'immigration par le public.

J'émetts l'hypothèse qu'une augmentation du chômage peut diminuer les perceptions positives à l'égard de l'immigration au Canada. Étant donné que le Canada a connu un changement dans son taux de chômage, passant de 5,66 % en 2019 à 7,46 % en 2021, il sera intéressant d'étudier si ce changement dans le taux de chômage a influencé les changements d'attitudes à l'égard de l'immigration.

J'aimerais comparer les attitudes entre ces deux années, en me concentrant sur la question de recherche exposée.

- Comment l'augmentation du taux de chômage entre 2019 et 2021 a-t-elle influencé les attitudes des Canadiens à l'égard de l'immigration ?

Pour répondre à ma question de recherche, j'utilise des données provenant de deux sources : les enquêtes Democracy Checkup et les données de la Banque mondiale. The Democracy Checkup surveys menées en [2019](#) (*5067 observations*) and [2021](#) (*7949 observations*) au Canada sont des enquêtes en ligne qui recueillent l'opinion du grand public sur les valeurs démocratiques, les politiques publiques et les questions d'actualité. Un échantillon en ligne de la population

générale canadienne a été obtenu par l'intermédiaire de Leger, en utilisant des quotas indépendants pour les catégories d'âge, le sexe et la province, conformément au recensement canadien de 2016. Cette méthodologie de collecte de données me permet d'incorporer la variable **immigration_level** et de décrire la généralisation à l'opinion publique.

En outre, j'incorpore des données de la [World Bank](#), en me concentrant spécifiquement sur les données d'observation du taux de chômage au Canada en 2019 et 2021. La Banque mondiale recueille des ensembles de données internationales basées sur les données générées par les systèmes statistiques nationaux, ce qui en fait une source essentielle de données macroéconomiques.

Données et méthodes

J'ai utilisé trois variables clés : **"immigration_level"**, **"unemployment"**, **"proportions"**. Les variables **"immigration_level"** and **"unemployment"** proviennent toutes deux de trois ensembles de données. La variable **"proportions"** a été créée pour tenir compte de la différence de nombre d'observations entre deux ensembles de données de Democracy Checkup. La variable **"year"** a permis de fusionner ces ensembles de données.

1. Pour commencer, la variable **"immigration_level"** provient de deux bases de données d'enquêtes en ligne : Democracy Checkup 2019 (dat_2019) et Democracy Checkup 2021 (dat_2021), qui ont été fusionnées en un seul ensemble de données (Data_clean). L'enquête posait la question suivante : 'Do you think Canada should admit... [immigrants]?' Initialement, cette variable était numérique. Je l'ai convertie en variable catégorielle. Cela me permet de visualiser la différence entre les catégories.
 - 1 - 'More'
 - 2 - 'Fewer'
 - 3 - 'About the same'

J'ai filtré les réponses catégorisées comme "Je ne sais pas" et "-99". En outre, j'ai créé une nouvelle colonne **"year"** pour faciliter la fusion de cet ensemble de données avec celui contenant la variable "chômage".

2. La variable **"unemployment"** pour les années 2019 et 2021 a été obtenue à partir des ensembles de données d'observation de la Banque mondiale (metadata).

Pour obtenir les données sur le chômage pour 2019 et 2021, j'ai sélectionné ces années spécifiques et filtré sur le Canada. (metadata_can_unemployment). J'ai ensuite utilisé la fonction *pivot_longer* pour transformer les données, en créant de nouvelles colonnes pour **"unemployment"** and **"year"**. La colonne "year" a servi de colonne clé entre les ensembles de données "Data_clean" et "metadata_can_unemployment" pour produire un nouvel ensemble de données "Data_combined".

La variable “unemployment” étant numérique, je l’ai traitée comme une variable catégorielle pour l’un des graphiques en utilisant la fonction *as.factor()*.

3. La variable “**proportions**” a été créée pour tenir compte des différences dans le nombre total d’observations entre les deux ensembles de données de Democracy Checkup.

Pour ce faire, le nombre total d’observations pour chaque catégorie de “immigration_level” (count) a été divisé par le nombre total d’observations dans les deux ensembles de données, puis multiplié par 100. La nouvelle colonne “proportions” a été générée. Cette colonne sera utilisée pour la visualisation sur l’axe Y.

Pour étudier la question de recherche, une analyse du croisement entre deux variables a été réalisée : unemployment rates and the proportions within the immigration_level categories.

Une description plus détaillée du processus de nettoyage des données se trouve dans l’Annexe : code de nettoyage

Les résultats

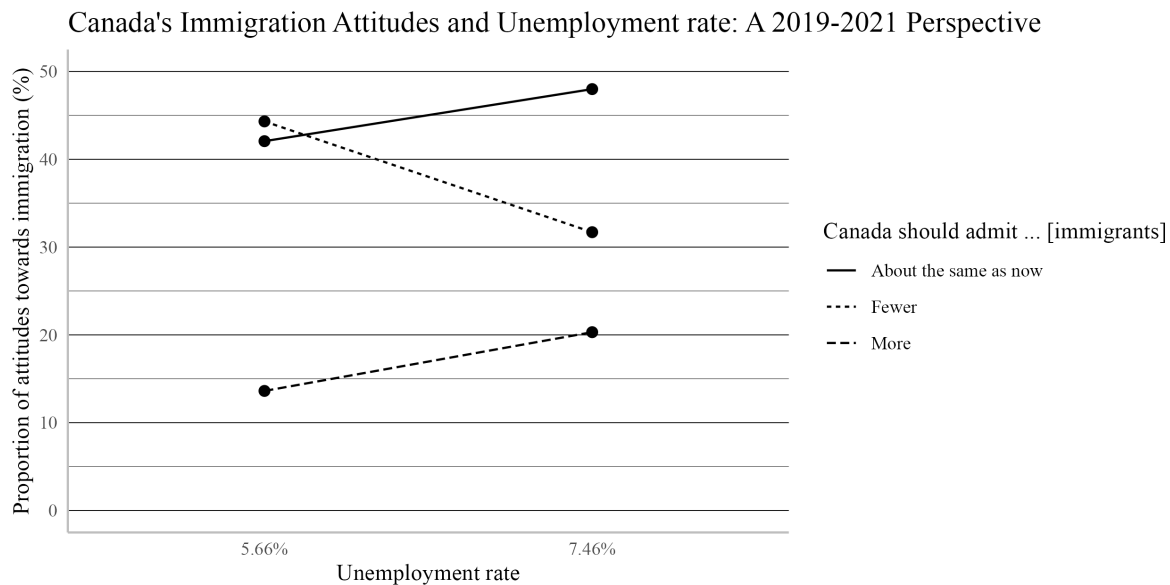


Figure 1: Option 1

Pour produire ce graphique, j’ai utilisé la variable unemployment sur l’axe X, qui provient des métadonnées d’observation des données de la Banque mondiale. Sur l’axe des ordonnées se trouve la variable proportions, qui a été calculée sur la base des données de la variable immigration_level, dérivée des données de l’enquête en ligne Democracy CheckUp. Les changements

d'attitude sont représentés par des lignes, chacune correspondant à une catégorie au sein de la variable `immigration_level`.

Les résultats sont intrigants : l'augmentation du taux de chômage de 5,66 % à 7,46 % n'a pas entraîné une diminution du soutien à l'immigration. Au contraire, dans la catégorie "Canada should admit more immigrants," nous avons observé une augmentation de 13,6 % à 20,3 %. Pour la catégorie "Canada should admit fewer immigrants," il y a eu une diminution de 44,3 % à 31,6 %. En outre, la catégorie "Canada should admit about the same number of immigrants as now" a augmenté, passant de 42 % à 48 %. Ces résultats remettent en question mon hypothèse initiale selon laquelle un taux de chômage plus élevé diminuerait le soutien à l'immigration. Une analyse plus approfondie est nécessaire pour comprendre si les changements d'attitude à l'égard de l'immigration pendant les périodes de chômage accru sont différenciés par classe, en particulier parmi les travailleurs peu qualifiés, comme le suggère l'étude de [Finseraas et al. 2014](#).

Une analyse similaire est présentée dans le graphique à barres, sans la variable du chômage.

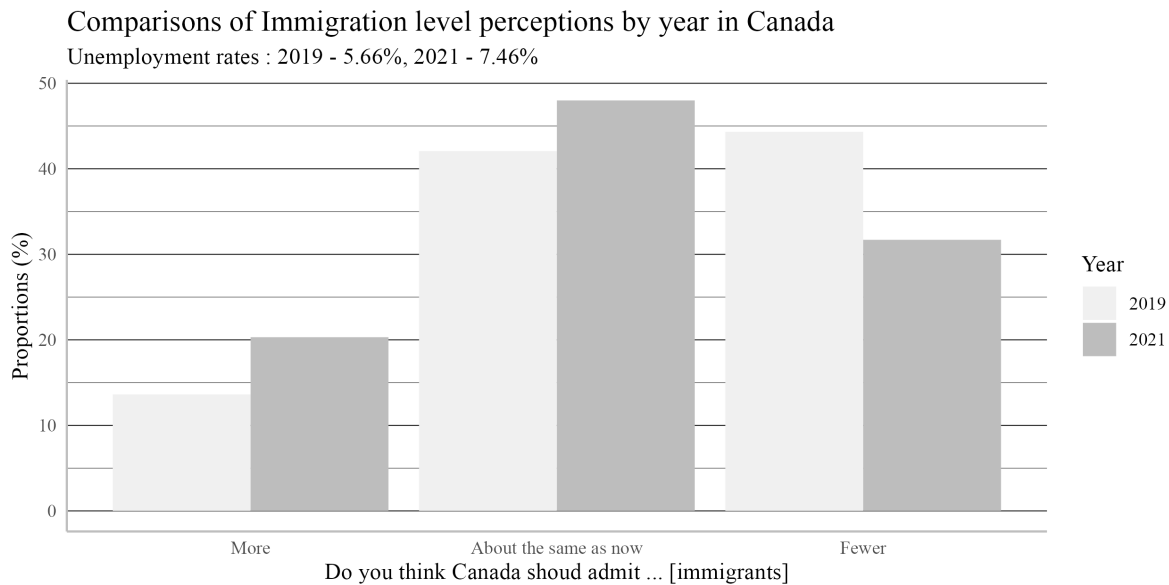


Figure 2: Option 2

Annexe : code de nettoyage

```
#1. Libraries  
  
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.0
v purrr      1.0.2

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(dplyr)
```

```
#2. Data
```

```
dat_2019 <- read_csv("~/Dropbox/fas_1001_Zhuk/_tp/_tp2/Data/cora-cdem-2019_F1.csv") #Democ
```

Warning: One or more parsing issues, call `problems()` on your data frame for details, e.g.:

```
dat <- vroom(...)
problems(dat)
```

```
Rows: 5067 Columns: 257
```

```
-- Column specification -----
Delimiter: ","
chr   (35): UserAgent, Q_Language, ResponseId, age_cat, vote_intention_7_TEX...
dbl   (218): Duration_in_seconds0, survey_minutes, wave, RecordedDate, conse...
lgl    (2): language_70_TEXT, ethnicity_33
date   (2): StartDate, EndDate
```

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```
dat_2021 <- read_csv("~/Dropbox/fas_1001_Zhuk/_tp/_tp2/Data/cora-cdem-2021_F1.csv") #Democ
```

Warning: One or more parsing issues, call `problems()` on your data frame for details, e.g.:

```
dat <- vroom(...)
problems(dat)
```

```
Rows: 7949 Columns: 304
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr (32): ResponseId, UserLanguage, dc21_vote_choice_6_TEXT, dc21_provvote...
```

```
dbl (264): Duration_in_seconds_, dc21_consent_t_First_Click, dc21_consent_...
```

```
lgl (5): dc21_yob_2003_age, dc21_provvote_D0_13, dc21_0, dc21_1, religon_fr
```

```
date (3): StartDate, EndDate, RecordedDate
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
metadata_can <- read_csv("~/Dropbox/fas_1001_Zhuk/_tp/_tp2/Data/API_SL.UEM.TOTL.ZS_DS2_en_
```

```
New names:
```

```
Rows: 266 Columns: 68
```

```
-- Column specification
```

```
----- Delimiter: "," chr
```

```
(4): Country Name, Country Code, Indicator Name, Indicator Code dbl (32): 1991,
```

```
1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, ... lgl (32): 1960,
```

```
1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, ...
```

```
i Use `spec()` to retrieve the full column specification for this data. i
```

```
Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
* `` -> `...68`
```

```
# 3 Nettoyage de variables
```

```
## 3.1 Democracy Checkup, 2019 [Canada]
```

```
# Vérification des observations
```

```
dat_2019 |> select(immigration_level) |>  
table()
```

```
immigration_level
```

```
 1    2    3    4  
653 2125 2017 272
```

```
# Je change les nombres en noms de catégories dans la variable "immigration_level" avec un
```

```
dat_2019_imm <- dat_2019 |> mutate(immigration_level = case_when(
```

```
immigration_level == 1 ~ "More",
immigration_level == 2 ~ "Fewer",
immigration_level == 3 ~ "About the same as now")) |> select(immigration_level) |>
filter(immigration_level != 4)
```

```
# Vérification des observations
```

```
dat_2019_imm |> table()
```

immigration_level		
About the same as now	Fewer	More
2017	2125	653

```
# Création d'une nouvelle colonne "year"
```

```
dat_2019_imm <- dat_2019_imm |> mutate(year = "2019")
```

```
# 3 Nettoyage de variables
```

```
## 3.2 Democracy Checkup, 2021 [Canada]
```

```
# Vérification de la variable
```

```
table(dat_2021$dc21_imm_level)
```

```
-99    1    2    3
  2 1614 2519 3814
```

```
# Je change le nom de la variable par le même nom que dans la base de données 2019 "immigr
```

```
dat_2021_imm <- dat_2021 |> mutate(immigration_level = case_when(
  dc21_imm_level == 1 ~ "More",
  dc21_imm_level == 2 ~ "Fewer",
  dc21_imm_level == 3 ~ "About the same as now")) |>
select(immigration_level) |>
filter(immigration_level != "-99")
```

```
# Vérification de la variable
```

```
dat_2021_imm |> table(useNA ="always")
```

```
immigration_level
About the same as now      Fewer      More
      3814                2519      1614
      <NA>
      0
```

```
# Création d'une nouvelle colonne "year"
```

```
dat_2021_imm <- dat_2021_imm |> mutate(year = "2021")
```

```
# 3 Nettoyage de variables
```

```
## 3.3 The World Bank Data Unemployment
```

```
# Je choisis les données basées sur le Canada pour les années 2019 et 2021
```

```
metadata_can_unemployment <- metadata_can |>
  select(`Country Name`, `2019`, `2021`) |>
  filter(`Country Name` == "Canada")
```

```
# Changer les données de plus larges à plus longues
```

```
metadata_can_unemployment_longer <- metadata_can_unemployment |>
  pivot_longer(cols = starts_with(c("2019", "2021")),
    names_to = "year",
    values_to = "unemployment")
```

```
# Verification
```

```
metadata_can_unemployment_longer |> head()
```

```
# A tibble: 2 x 3
```

```
  `Country Name` year  unemployment
  <chr>          <chr>          <dbl>
1 Canada      2019          5.66
2 Canada      2021          7.46
```



```
#4. Fusion des données: Democracy Checkup, 2019, 2021 [Canada]
```

```
Data_clean <- bind_rows(dat_2019_imm, dat_2021_imm)
```

```
# Vérifier que les variables se sont bien jointes
```

```
Data_clean |> select(year, immigration_level) |>  
table()
```

	immigration_level			
year	About the same as now	Fewer	More	
2019		2017	2125	653
2021		3814	2519	1614

```
# 4.1 Fusion des données: Democracy Checkup, 2019, 2021 [Canada] + The World Bank Data Une
```

```
Data_combined <- left_join(Data_clean, metadata_can_unemployment_longer, by = c("year"))
```

```
# Verification
```

```
Data_combined |> select(immigration_level, unemployment) |>  
table()
```

		unemployment	
immigration_level		5.66	7.46
About the same as now	2017	3814	
Fewer	2125	2519	
More	653	1614	

```
# 5: créer un nouvel ensemble de données avec une nouvelle colonne pour calculer les propor
```

```
# This code was based on ChatGPT suggestion
```

```
data_clean_proportions <- Data_combined |>  
  filter(!is.na(immigration_level)) |>  
  group_by(year, immigration_level) |>  
  summarize(count = n(), unemployment = first(unemployment), .groups = "keep") |>  
  group_by(year) |>  
  mutate(total = sum(count), proportion = count / total * 100)
```