Canada_UnMet_HealthCare_2018_2022

Ahmed Abdou

2023-08-30

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     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 error
```

```r
library(dplyr)
library(ggplot2)

data<- read.csv("C:/Users/Admin/Desktop/UnMet Healtcare Analysis- canada- 16 and over/13100836.csv")
head(data)
```

```
##   REF_DATE    GEO           DGUID        Sex
## 1     2018 Canada 2016A000011124 Both sexes
## 2     2018 Canada 2016A000011124 Both sexes
## 3     2018 Canada 2016A000011124      Males
## 4     2018 Canada 2016A000011124      Males
## 5     2018 Canada 2016A000011124    Females
## 6     2018 Canada 2016A000011124    Females
##                                           Statistics     UOM UOM_ID
## 1     Number of persons with unmet health care needs  Number    223
## 2 Percentage of persons with unmet health care needs Percent    239
## 3     Number of persons with unmet health care needs  Number    223
## 4 Percentage of persons with unmet health care needs Percent    239
## 5     Number of persons with unmet health care needs  Number    223
## 6 Percentage of persons with unmet health care needs Percent    239
##   SCALAR_FACTOR SCALAR_ID      VECTOR COORDINATE  VALUE STATUS SYMBOL
## 1     thousands         3 v1288345915      1.1.1 1536.0      B     NA
## 2         units         0 v1288345916      1.1.2    5.1      B     NA
## 3     thousands         3 v1288345917      1.2.1  683.0      C     NA
## 4         units         0 v1288345918      1.2.2    4.6      C     NA
## 5     thousands         3 v1288345919      1.3.1  853.0      C     NA
## 6         units         0 v1288345920      1.3.2    5.6      C     NA
##   TERMINATED DECIMALS
## 1         NA        0
## 2         NA        1
## 3         NA        0
## 4         NA        1
## 5         NA        0
## 6         NA        1
```

```r
#Select required rows to do the analysis

data_selected<-data[c("REF_DATE","GEO","Sex","Statistics", "VALUE","UOM")]
head(data_selected)
```

```
##   REF_DATE    GEO        Sex                                         Statistics
## 1     2018 Canada Both sexes     Number of persons with unmet health care needs
## 2     2018 Canada Both sexes Percentage of persons with unmet health care needs
## 3     2018 Canada      Males     Number of persons with unmet health care needs
## 4     2018 Canada      Males Percentage of persons with unmet health care needs
## 5     2018 Canada    Females     Number of persons with unmet health care needs
## 6     2018 Canada    Females Percentage of persons with unmet health care needs
##    VALUE     UOM
## 1 1536.0  Number
## 2    5.1 Percent
## 3  683.0  Number
## 4    4.6 Percent
## 5  853.0  Number
## 6    5.6 Percent
```

```r
# Summarize missing values using dplyr

missing_summary <- data_selected %>%
  summarise_all(~ sum(is.na(.)))
missing_summary
```

```
##   REF_DATE GEO Sex Statistics VALUE UOM
## 1        0   0   0          0     1   0
```

```r
# Removing the row that includes no values

cleaned_data <- na.omit(data_selected)
head(cleaned_data)
```

```
##   REF_DATE    GEO        Sex                                         Statistics
## 1     2018 Canada Both sexes     Number of persons with unmet health care needs
## 2     2018 Canada Both sexes Percentage of persons with unmet health care needs
## 3     2018 Canada      Males     Number of persons with unmet health care needs
## 4     2018 Canada      Males Percentage of persons with unmet health care needs
## 5     2018 Canada    Females     Number of persons with unmet health care needs
## 6     2018 Canada    Females Percentage of persons with unmet health care needs
##    VALUE     UOM
## 1 1536.0  Number
## 2    5.1 Percent
## 3  683.0  Number
## 4    4.6 Percent
## 5  853.0  Number
## 6    5.6 Percent
```

```r
#Filtering Sex column and remove "both Sex" and keep only Males & Females

library(dplyr)
```

```
filtered_data <- cleaned_data %>%
  filter(trimws(Sex) %in% c("Males", "Females"))
unique(filtered_data$Sex)
```

```
## [1] "Males "  "Females"
```

```
#Filtering UOM column and remove Number(I WILL WORK WITH %)

filtered_data<- filtered_data%>%
  filter(trimws(UOM) %in% c("Percent"))
head(filtered_data)
```

```
##   REF_DATE                       GEO     Sex
## 1     2018                    Canada  Males
## 2     2018                    Canada Females
## 3     2018        Atlantic provinces  Males
## 4     2018        Atlantic provinces Females
## 5     2018 Newfoundland and Labrador  Males
## 6     2018 Newfoundland and Labrador Females
##                                            Statistics VALUE     UOM
## 1 Percentage of persons with unmet health care needs   4.6 Percent
## 2 Percentage of persons with unmet health care needs   5.6 Percent
## 3 Percentage of persons with unmet health care needs   5.7 Percent
## 4 Percentage of persons with unmet health care needs   7.6 Percent
## 5 Percentage of persons with unmet health care needs   4.9 Percent
## 6 Percentage of persons with unmet health care needs   6.8 Percent
```

```
df<- filtered_data%>%
  select(-Statistics,-UOM)
head(df)
```

```
##   REF_DATE                       GEO     Sex VALUE
## 1     2018                    Canada  Males    4.6
## 2     2018                    Canada Females   5.6
## 3     2018        Atlantic provinces  Males    5.7
## 4     2018        Atlantic provinces Females   7.6
## 5     2018 Newfoundland and Labrador  Males    4.9
## 6     2018 Newfoundland and Labrador Females   6.8
```

```
#Renaming
df <- df %>%
  rename(
    Year = REF_DATE,
    Province = GEO,
    Sex = Sex,
    "%" = VALUE)
head(df)
```

```
##   Year                  Province     Sex   %
## 1 2018                    Canada  Males  4.6
## 2 2018                    Canada Females 5.6
## 3 2018        Atlantic provinces  Males  5.7
## 4 2018        Atlantic provinces Females 7.6
## 5 2018 Newfoundland and Labrador  Males  4.9
## 6 2018 Newfoundland and Labrador Females 6.8
```

```r
####################################################
#Delete non provinces records(eg.Canada)


unique(df$Province)
```

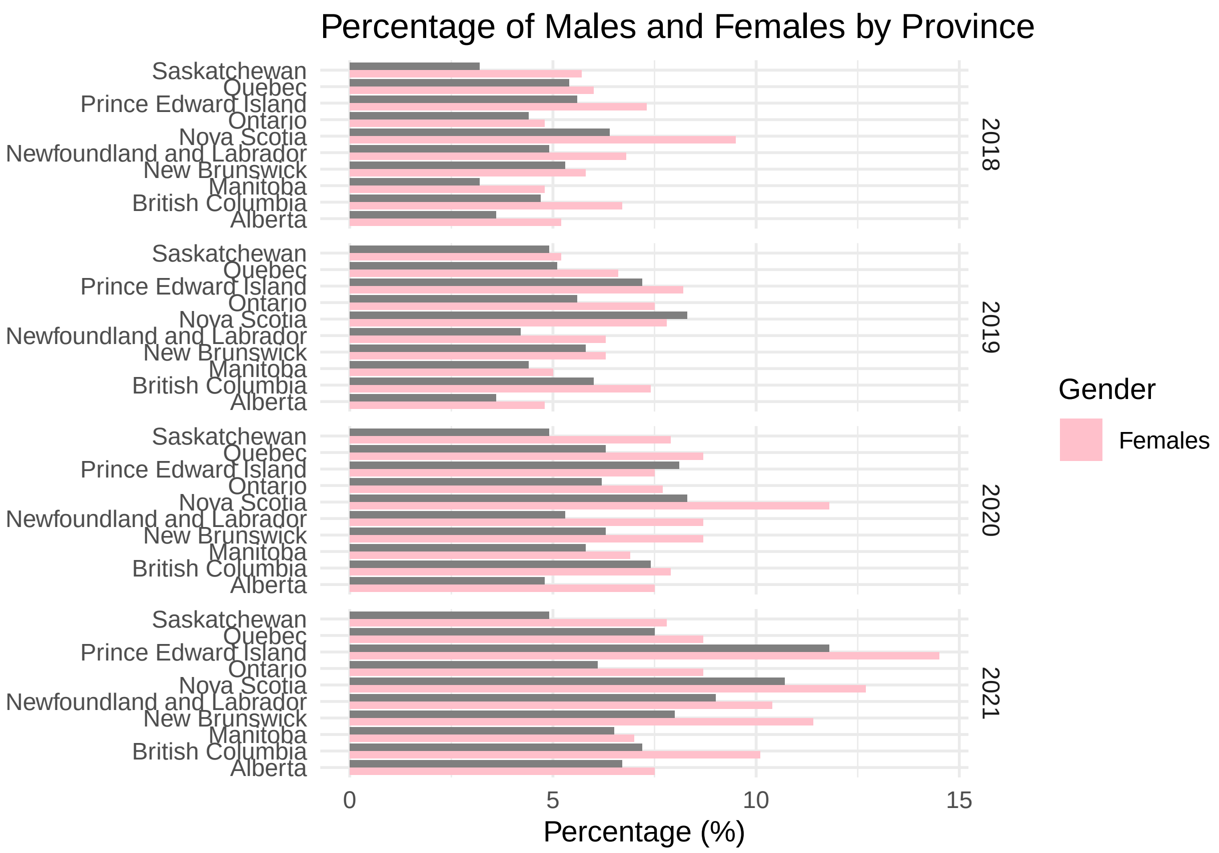
```
##  [1] "Canada"                    "Atlantic provinces"
##  [3] "Newfoundland and Labrador"  "Prince Edward Island"
##  [5] "Nova Scotia"               "New Brunswick"
##  [7] "Quebec"                    "Ontario"
##  [9] "Prairie provinces"         "Manitoba"
## [11] "Saskatchewan"              "Alberta"
## [13] "British Columbia"          "Canada, 16 to 24 years"
## [15] "Canada, 25 to 54 years"    "Canada, 25 to 34 years"
## [17] "Canada, 35 to 44 years"    "Canada, 45 to 54 years"
## [19] "Canada, 55 to 64 years"    "Canada,  65 years and over"
```

```r
selected_provinces <- c("Prince Edward Island", "Quebec", "Manitoba","British Columbia","Nova Scotia",
                        "Ontario","Saskatchewan","Newfoundland and Labrador","New Brunswick","Alberta")

unmet_df<- df %>%
  filter(Province %in% selected_provinces)

#########################################################################
library(dplyr)
library(ggplot2)

ggplot(unmet_df, aes(x = `%`, y = Province, fill = Sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(Year ~ ., scales = "free_x", space = "free_x") +
  labs(title = "Percentage of Males and Females by Province",
       x = "Percentage (%)",
       y = "Province",
       fill = "Gender") +
  theme_minimal() +
  theme(axis.text.y = element_text(angle = 0, hjust = 1),
        axis.title.y = element_blank()) +
  scale_fill_manual(values = c("Males" = "black", "Females" = "pink"))
```

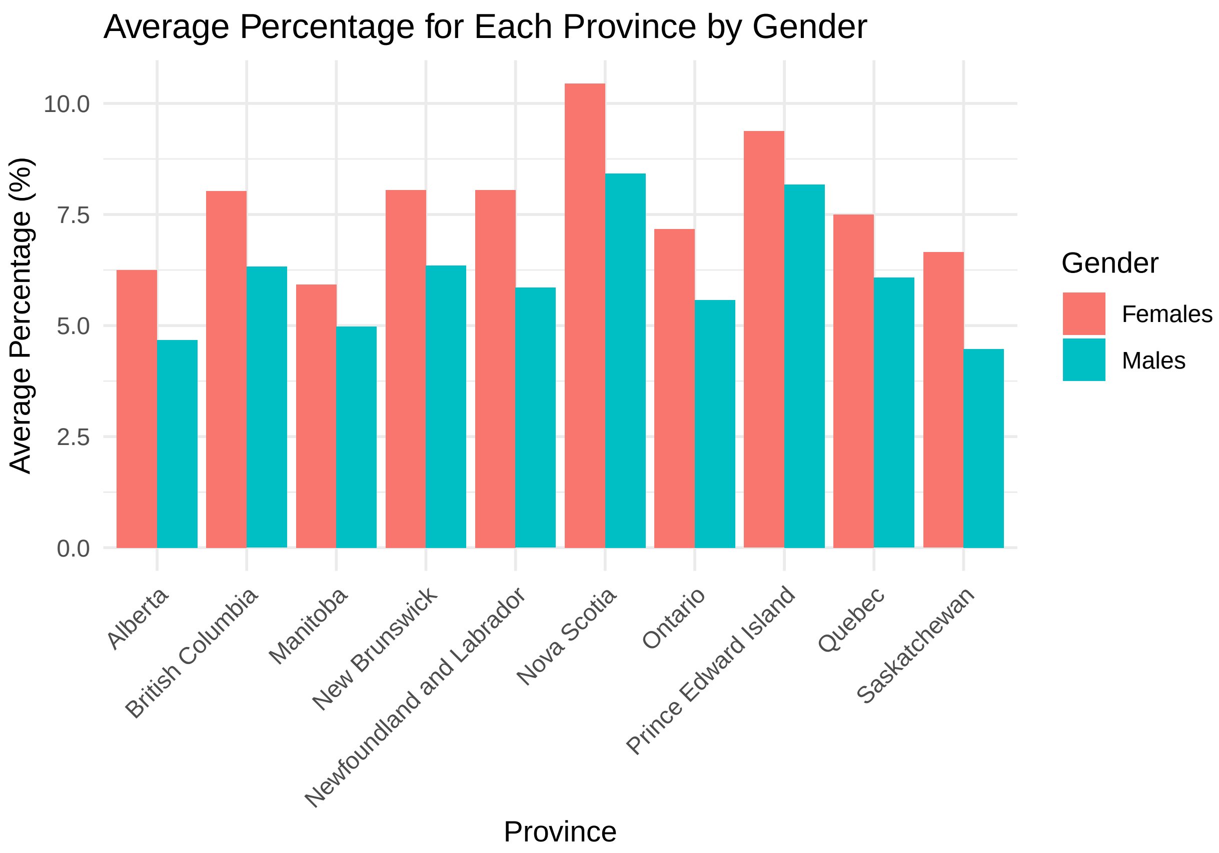


```
# Calculate average percentage for each province and gender
library(dplyr)
library(ggplot2)

avg_data <- unmet_df %>%
  group_by(Province, Sex) %>%
  summarise(`Avg %` = mean(`%`))
```

```
## `summarise()` has grouped output by 'Province'. You can override using the
## `.groups` argument.
```

```
# Create the bar chart
ggplot(avg_data, aes(x = Province, y = `Avg %`, fill = Sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Average Percentage for Each Province by Gender",
       x = "Province",
       y = "Average Percentage (%)",
       fill = "Gender") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

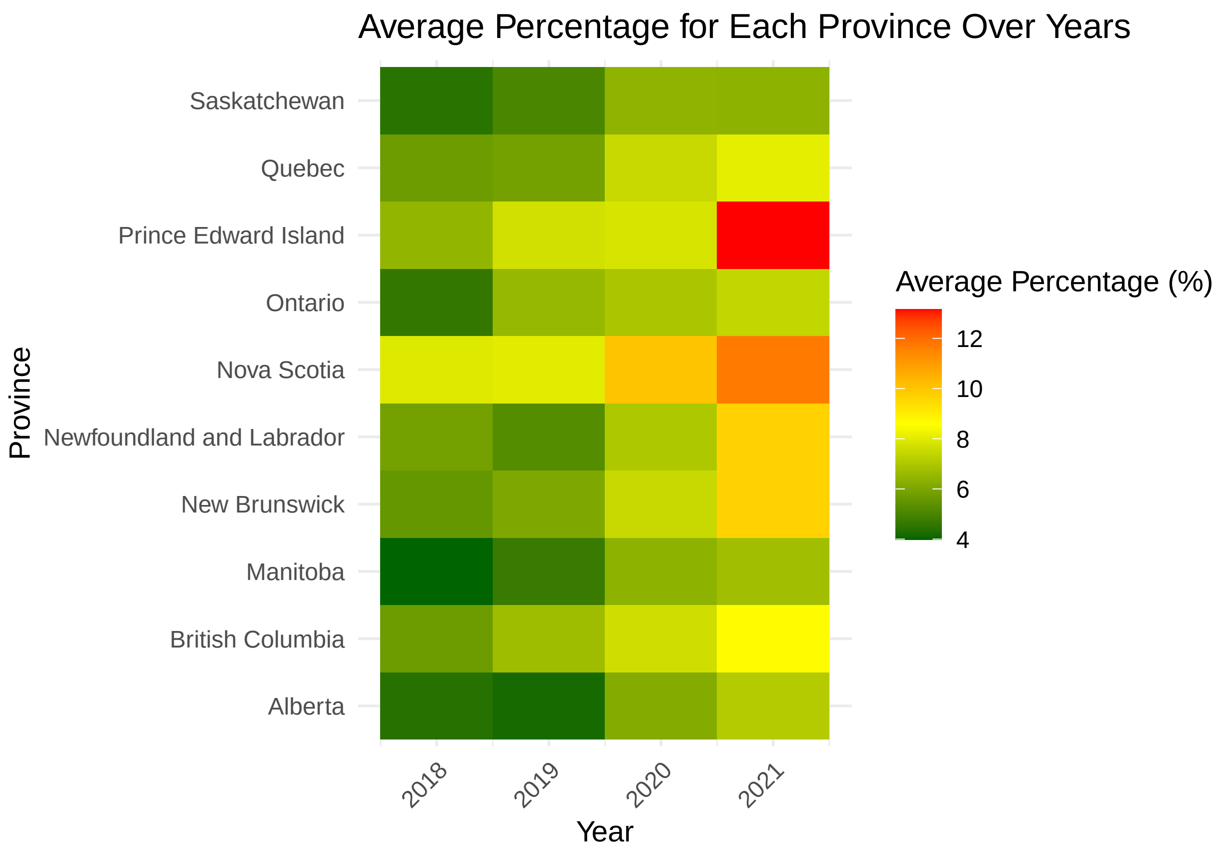


```
##########################################################################
library(dplyr)
library(ggplot2)

# Calculate average percentage for each province and year
avg_data <- unmet_df %>%
  group_by(Province, Year) %>%
  summarise(`Avg %` = mean(`%`))
```

```
## `summarise()` has grouped output by 'Province'. You can override using the
## `.groups` argument.
```

```
# Create the heatmap with red for higher values and green for lower values
ggplot(avg_data, aes(x = Year, y = Province, fill = `Avg %`)) +
  geom_tile() +
  labs(title = "Average Percentage for Each Province Over Years",
       x = "Year",
       y = "Province",
       fill = "Average Percentage (%)") +
  scale_fill_gradientn(colors = c("darkgreen", "yellow", "red")) +  # Adjust the color gradient
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
##############################################
# Load necessary libraries
library(dplyr)
library(ggplot2)

# Calculate average percentage for each province and gender
avg_data <- unmet_df %>%
  group_by(Province, Sex) %>%
  summarise(`Avg %` = mean(`%`))
```

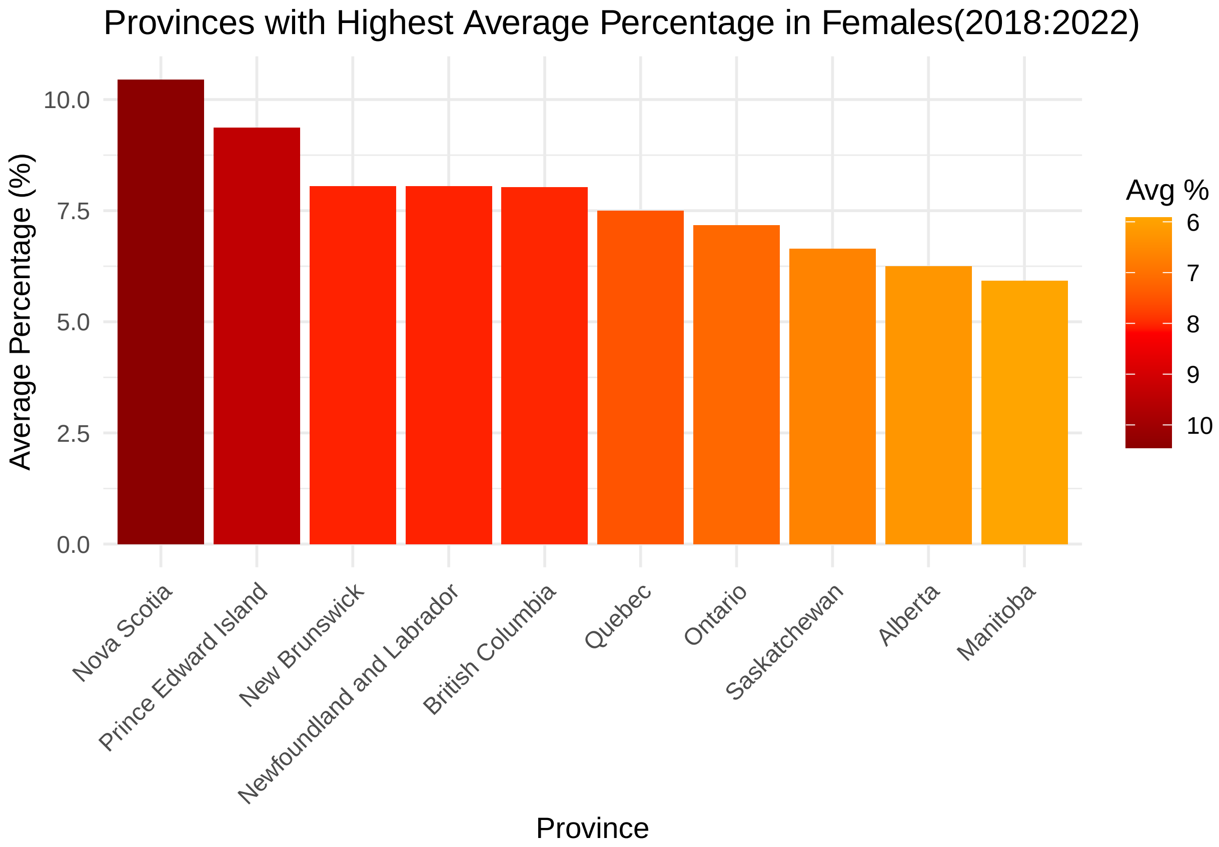
```
## `summarise()` has grouped output by 'Province'. You can override using the
## `.groups` argument.
```

```
# Sort avg_data by `Avg %` in descending order for females
sorted_avg_females <- avg_data %>%
  filter(Sex == "Females") %>%
  arrange(desc(`Avg %`))

# Create a custom color scale ranging from dark red to light red
color_scale <- scale_fill_gradientn(
  colors = c("orange","red", "darkred"),  # Adjust color values
  guide = guide_colorbar(reverse = TRUE)
)

# Create a bar plot for provinces with highest average % in females
p_females <- ggplot(sorted_avg_females, aes(x = reorder(Province, -`Avg %`), y = `Avg %`)) +
  geom_bar(stat = "identity", aes(fill = `Avg %`)) +  # Use fill aesthetic
  labs(title = "Provinces with Highest Average Percentage in Females(2018:2022)",
       x = "Province",
       y = "Average Percentage (%)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  color_scale

# Display the females plot
print(p_females)
```



```
################################################################
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.