HUMBER INSTITUTE OF TECHNOLOGY

AND ADVANCED LEARNING

Logo, company name

Description automatically generated

**BUSE 5303 – Big Data 2**

Group Project

Data Visualization for Decision Making

Submitted by: Ansh Gangdev

Deepanshi Garg

Janna Avila

Kapil Sharma

Kivan Ilangakoon

Saransh Vaid

**Video Presentation Link**

<https://drive.google.com/file/d/1mcqZbK2JzDX2t5ApZTjwpDSve5sA1D1C/view>

**RStudio Code**

*#installing and loading packages*

install.packages("dplyr")

install.packages("lubridate")

install.packages("stringr")

install.packages("tidyr")

install.packages("plotly")

install.packages("rnaturalearth")

install.packages("rnaturalearthdata")

install.packages("ggmap")

install.packages("devtools")

install.packages("sf")

devtools::install\_github("ropensci/rnaturalearthhires")

remove.packages("tidyverse")

library(plotly)

library(readr)

library(dplyr)

library(lubridate)

library(stringr)

library(ggplot2)

library(tidyr)

library(reshape2)

library(rnaturalearth)

library(sf)

library(rnaturalearthdata)

library(ggmap)

*#------------------------------------------------------------------------------*

*#------------------------------------------------------------------------------*

*#------------------------------------------------------------------------------*

*#importing data files into DFs*

Prosperity\_Poverty <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**ProsPov.csv")

Prossperty\_Unemployment <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**ProsUnem.csv")

Health\_Life <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**HealLif.csv")

Health\_Mental <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**HealMen.csv")

Social\_Belonging <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**SocBel.csv")

Environment\_Gas <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**EnvGas.csv")

Governance\_Victim <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**GoodVic.csv")

*#Selecting relevant cols for DFs and renaming*

Prosperity\_Poverty <- Prosperity\_Poverty[, c("REF\_DATE", "GEO", "Individual MBM poverty status (7)", "VALUE")]

names(Prosperity\_Poverty)[names(Prosperity\_Poverty) == 'Individual MBM poverty status (7)'] <- 'CATEGORY'

Prossperty\_Unemployment <- Prossperty\_Unemployment[, c("REF\_DATE", "GEO", "Labour force characteristics", "VALUE")]

names(Prossperty\_Unemployment)[names(Prossperty\_Unemployment) == 'Labour force characteristics'] <- 'CATEGORY'

Health\_Life <- Health\_Life[, c("REF\_DATE", "GEO", "Age group", "VALUE")]

names(Health\_Life)[names(Health\_Life) == 'Age group'] <- 'CATEGORY'

Health\_Mental <- Health\_Mental[, c("REF\_DATE", "GEO", "Indicators", "VALUE")]

Social\_Belonging <- Social\_Belonging[, c("REF\_DATE", "GEO", "Indicators", "VALUE")]

names(Environment\_Gas)[names(Environment\_Gas) == '1990 greenhouse gas emissions (megatonnes of carbon dioxide equivalent)'] <- 'Year\_1990'

names(Environment\_Gas)[names(Environment\_Gas) == '2005 greenhouse gas emissions (megatonnes of carbon dioxide equivalent)'] <- 'Year\_2005'

names(Environment\_Gas)[names(Environment\_Gas) == '2020 greenhouse gas emissions (megatonnes of carbon dioxide equivalent)'] <- 'Year\_2020'

names(Environment\_Gas)[names(Environment\_Gas) == 'Province or territory'] <- 'Province'

names(Governance\_Victim)[names(Governance\_Victim) == 'Percentage of persons reporting victimization (%)'] <- 'Victimization'

names(Governance\_Victim)[names(Governance\_Victim) == 'Percentage of persons reporting violent victimization (%)'] <- 'Violent\_Victimization'

*#------------------------------------------------------------------------------*

*#------------------------------------------------------------------------------*

*#------------------------------------------------------------------------------*

*#importing data files into DFs*

Prosperity\_Poverty <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**ProsPov.csv")

Prossperty\_Unemployment <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**ProsUnem.csv")

Health\_Life <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**HealLif.csv")

Health\_Mental <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**HealMen.csv")

Social\_Belonging <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**SocBel.csv")

Environment\_Gas <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**EnvGas.csv")

Governance\_Victim <- read\_csv("C:**\\**Users**\\**deepa**\\**OneDrive**\\**Desktop**\\**Big Data 2**\\**Project**\\**GoodVic.csv")

*#Selecting relevant cols for DFs and renaming*

Prosperity\_Poverty <- Prosperity\_Poverty[, c("REF\_DATE", "GEO", "Individual MBM poverty status (7)", "VALUE")]

names(Prosperity\_Poverty)[names(Prosperity\_Poverty) == 'Individual MBM poverty status (7)'] <- 'CATEGORY'

Prossperty\_Unemployment <- Prossperty\_Unemployment[, c("REF\_DATE", "GEO", "Labour force characteristics", "VALUE")]

names(Prossperty\_Unemployment)[names(Prossperty\_Unemployment) == 'Labour force characteristics'] <- 'CATEGORY'

Health\_Life <- Health\_Life[, c("REF\_DATE", "GEO", "Age group", "VALUE")]

names(Health\_Life)[names(Health\_Life) == 'Age group'] <- 'CATEGORY'

Health\_Mental <- Health\_Mental[, c("REF\_DATE", "GEO", "Indicators", "VALUE")]

Social\_Belonging <- Social\_Belonging[, c("REF\_DATE", "GEO", "Indicators", "VALUE")]

names(Environment\_Gas)[names(Environment\_Gas) == '1990 greenhouse gas emissions (megatonnes of carbon dioxide equivalent)'] <- 'Year\_1990'

names(Environment\_Gas)[names(Environment\_Gas) == '2005 greenhouse gas emissions (megatonnes of carbon dioxide equivalent)'] <- 'Year\_2005'

names(Environment\_Gas)[names(Environment\_Gas) == '2020 greenhouse gas emissions (megatonnes of carbon dioxide equivalent)'] <- 'Year\_2020'

names(Environment\_Gas)[names(Environment\_Gas) == 'Province or territory'] <- 'Province'

names(Governance\_Victim)[names(Governance\_Victim) == 'Percentage of persons reporting victimization (%)'] <- 'Victimization'

names(Governance\_Victim)[names(Governance\_Victim) == 'Percentage of persons reporting violent victimization (%)'] <- 'Violent\_Victimization'

*#splitting DF into 3 DFs*

Prosperity\_Poverty1 <- Prosperity\_Poverty[Prosperity\_Poverty$CATEGORY %**in**% c("Total - Individual MBM poverty status"), ]

Prosperity\_Poverty2 <- Prosperity\_Poverty[Prosperity\_Poverty$CATEGORY %**in**% c("In poverty"), ]

Prosperity\_Poverty3 <- Prosperity\_Poverty[Prosperity\_Poverty$CATEGORY %**in**% c("Not in poverty"), ]

*#assigning common ID so the DFs can now be merged*

Prosperity\_Poverty1 <- cbind(ID = 1:nrow(Prosperity\_Poverty1), Prosperity\_Poverty1)

Prosperity\_Poverty2 <- cbind(ID = 1:nrow(Prosperity\_Poverty2), Prosperity\_Poverty2)

Prosperity\_Poverty3 <- cbind(ID = 1:nrow(Prosperity\_Poverty3), Prosperity\_Poverty3)

*#merging the 3 DFs by ID's*

Prosperity\_Poverty4 <- merge(Prosperity\_Poverty1, Prosperity\_Poverty2, by = c("ID"))

Prosperity\_Poverty5 <- merge(Prosperity\_Poverty3, Prosperity\_Poverty4, by = c("ID"))

*#Rename*

names(Prosperity\_Poverty5)[names(Prosperity\_Poverty5) == 'VALUE'] <- 'NOT\_IN\_POVERTY'

names(Prosperity\_Poverty5)[names(Prosperity\_Poverty5) == 'VALUE.x'] <- 'TOTAL'

names(Prosperity\_Poverty5)[names(Prosperity\_Poverty5) == 'VALUE.y'] <- 'IN\_POVERTY'

*#Selecting relevant cols for DFs and renaming*

Prosperity\_Poverty6 <- Prosperity\_Poverty5[, c("REF\_DATE", "GEO", "TOTAL", "NOT\_IN\_POVERTY", "IN\_POVERTY")]

*#splitting DF into 2 DFs*

Prossperty\_Unemployment1 <- Prossperty\_Unemployment[Prossperty\_Unemployment$CATEGORY %**in**% c("Population"), ]

Prossperty\_Unemployment2 <- Prossperty\_Unemployment[Prossperty\_Unemployment$CATEGORY %**in**% c("Unemployment rate"), ]

*#assigning common ID so the DFs can now be merged*

Prossperty\_Unemployment1 <- cbind(ID = 1:nrow(Prossperty\_Unemployment1), Prossperty\_Unemployment1)

Prossperty\_Unemployment2 <- cbind(ID = 1:nrow(Prossperty\_Unemployment2), Prossperty\_Unemployment2)

*#merging the 3 DFs by ID's*

Prossperty\_Unemployment3 <- merge(Prossperty\_Unemployment1, Prossperty\_Unemployment2, by = c("ID"))

*#Rename*

names(Prossperty\_Unemployment3)[names(Prossperty\_Unemployment3) == 'REF\_DATE.x'] <- 'REF\_DATE'

names(Prossperty\_Unemployment3)[names(Prossperty\_Unemployment3) == 'GEO.x'] <- 'GEO'

names(Prossperty\_Unemployment3)[names(Prossperty\_Unemployment3) == 'VALUE.x'] <- 'POPULATION'

names(Prossperty\_Unemployment3)[names(Prossperty\_Unemployment3) == 'VALUE.y'] <- 'UNEMPLOYMENT\_RATE'

*#Selecting relevant cols for DFs and renaming*

Prossperty\_Unemployment4 <- Prossperty\_Unemployment3[, c("REF\_DATE", "GEO", "POPULATION", "UNEMPLOYMENT\_RATE")]

*#splitting DF into 3 DFs*

Social\_Belonging1 <- Social\_Belonging[Social\_Belonging$Indicators %**in**% c("Very strong or somewhat strong sense of belonging to local community"), ]

Social\_Belonging2 <- Social\_Belonging[Social\_Belonging$Indicators %**in**% c("Somewhat weak or very weak sense of belonging to local community"), ]

Social\_Belonging3 <- Social\_Belonging[Social\_Belonging$Indicators %**in**% c("No opinion on sense of belonging to local community"), ]

*#assigning common ID so the DFs can now be merged*

Social\_Belonging1 <- cbind(ID = 1:nrow(Social\_Belonging1), Social\_Belonging1)

Social\_Belonging2 <- cbind(ID = 1:nrow(Social\_Belonging2), Social\_Belonging2)

Social\_Belonging3 <- cbind(ID = 1:nrow(Social\_Belonging3), Social\_Belonging3)

*#merging the 3 DFs by ID's*

Social\_Belonging4 <- merge(Social\_Belonging1, Social\_Belonging2, by = c("ID"))

Social\_Belonging5 <- merge(Social\_Belonging3, Social\_Belonging4, by = c("ID"))

*#Rename*

names(Social\_Belonging5)[names(Social\_Belonging5) == 'VALUE'] <- 'NO\_OPINION'

names(Social\_Belonging5)[names(Social\_Belonging5) == 'VALUE.x'] <- 'VERYSTRONG\_SOMEWHATSTRONG'

names(Social\_Belonging5)[names(Social\_Belonging5) == 'VALUE.y'] <- 'SOMEWHATWEAK\_VERYWEAK'

*#Selecting relevant cols for DFs and renaming*

Social\_Belonging6 <- Social\_Belonging5[, c("REF\_DATE", "GEO", "VERYSTRONG\_SOMEWHATSTRONG", "SOMEWHATWEAK\_VERYWEAK", "NO\_OPINION")]

*#splitting DF into 3 DFs*

Health\_Mental1 <- Health\_Mental[Health\_Mental$Indicators %**in**% c("Excellent or very good perceived mental health"), ]

Health\_Mental2 <- Health\_Mental[Health\_Mental$Indicators %**in**% c("Good perceived mental health"), ]

Health\_Mental3 <- Health\_Mental[Health\_Mental$Indicators %**in**% c("Fair or poor perceived mental health"), ]

*#assigning common ID so the DFs can now be merged*

Health\_Mental1 <- cbind(ID = 1:nrow(Health\_Mental1), Health\_Mental1)

Health\_Mental2 <- cbind(ID = 1:nrow(Health\_Mental2), Health\_Mental2)

Health\_Mental3 <- cbind(ID = 1:nrow(Health\_Mental3), Health\_Mental3)

*#merging the 3 DFs by ID's*

Health\_Mental4 <- merge(Health\_Mental1, Health\_Mental2, by = c("ID"))

Health\_Mental5 <- merge(Health\_Mental3, Health\_Mental4, by = c("ID"))

*#Rename*

names(Health\_Mental5)[names(Health\_Mental5) == 'VALUE'] <- 'FAIR\_OR\_POOR'

names(Health\_Mental5)[names(Health\_Mental5) == 'VALUE.x'] <- 'EXCELLENT\_OR\_VERYGOOD'

names(Health\_Mental5)[names(Health\_Mental5) == 'VALUE.y'] <- 'GOOD'

*#Selecting relevant cols for DFs and renaming*

Health\_Mental5 <- Health\_Mental5[, c("REF\_DATE", "GEO", "EXCELLENT\_OR\_VERYGOOD", "GOOD", "FAIR\_OR\_POOR")]

*#splitting DF into 2 DFs*

Health\_Life1 <- Health\_Life[Health\_Life$CATEGORY %**in**% c("At birth"), ]

Health\_Life2 <- Health\_Life[Health\_Life$CATEGORY %**in**% c("At age 65"), ]

*#assigning common ID so the DFs can now be merged*

Health\_Life1 <- cbind(ID = 1:nrow(Health\_Life1), Health\_Life1)

Health\_Life2 <- cbind(ID = 1:nrow(Health\_Life2), Health\_Life2)

*#merging the 3 DFs by ID's*

Health\_Life3 <- merge(Health\_Life1, Health\_Life2, by = c("ID"))

*#Rename*

names(Health\_Life3)[names(Health\_Life3) == 'REF\_DATE.x'] <- 'REF\_DATE'

names(Health\_Life3)[names(Health\_Life3) == 'GEO.x'] <- 'GEO'

names(Health\_Life3)[names(Health\_Life3) == 'VALUE.x'] <- 'AT\_BIRTH'

names(Health\_Life3)[names(Health\_Life3) == 'VALUE.y'] <- 'AT\_AGE65'

*#Selecting relevant cols for DFs and renaming*

Health\_Life4 <- Health\_Life3[, c("REF\_DATE", "GEO", "AT\_BIRTH", "AT\_AGE65")]

*#---------------------------------------------------------------*

*#---------------------------------------------------------------*

*#---------------------------------------------------------------*

*#Prosperity Poverty --------------------------------------------*

df1 <- Prosperity\_Poverty6[, c("GEO", "TOTAL", "IN\_POVERTY", "NOT\_IN\_POVERTY")]

df1$POVERTY\_RATE <- (df1$IN\_POVERTY / df1$TOTAL) \* 100

*#Creating another dummy for sorting GEO by Alphabetical order*

df1\_S <- df1

*#Sorting GEO by Alphabetical order*

df1\_Sorted <- df1\_S[order(df1\_S$GEO), ]

*#rounding off average rate columns to 2 DP*

df1\_Sorted <- df1\_Sorted %>%

mutate\_if(**is**.numeric,

round, digits = 2)

*# Create named vector of province abbreviations*

PROVINCE\_drop <- c(

"Newfoundland and Labrador" = "NL",

"Prince Edward Island" = "PE",

"Nova Scotia" = "NS",

"New Brunswick" = "NB",

"Quebec" = "QC",

"Ontario" = "ON",

"Manitoba" = "MB",

"Saskatchewan" = "SK",

"Alberta" = "AB",

"British Columbia" = "BC"

)

df1\_Sorted <- df1\_Sorted %>%

mutate(PROVINCE\_drop = PROVINCE\_drop[GEO])

df1B <- df1\_Sorted[, c("PROVINCE\_drop", "TOTAL", "IN\_POVERTY", "NOT\_IN\_POVERTY", "POVERTY\_RATE")]

df1C <- df1B

*#--------------------------------------------------------*

*#--------------------------------------------------------*

*#--------------------------------------------------------*

*#Renaming the PROVINCE\_drop to postal for merging the dataframe*

names(df1C)[names(df1C) == 'PROVINCE\_drop'] <- 'postal'

*# load the data*

canada\_provinces <- ne\_states(country = "canada", returnclass = "sf")

*#1mturkData <- read.csv("alldata2.csv", header = TRUE)*

*#Creating dummy*

canada\_provincesA <- canada\_provinces

*#Merging the data to have same number of rows*

*#merged\_data <- merge(canada\_provincesA, df1BSorted, by = "postal", all.x = TRUE)*

merged\_data <- merge(canada\_provincesA, df1C, by = c("postal"))

*#binding the data since it has same number of rows now*

df1D <- cbind(df1C, merged\_data)

*#dropping one of the postal code column and creating a new dataframe*

df1E <- df1D[, -6]

*#--------------------------------------------------------*

*#--------------------------------------------------------*

*#--------------------------------------------------------*

*# convert df1E to sf object*

df1E\_sf <- st\_as\_sf(df1E, coords = c("longitude", "latitude"), crs = 4326)

*# plot choropleth map*

ggplot() +

geom\_sf(data = df1E\_sf, aes(fill = POVERTY\_RATE)) +

scale\_fill\_gradient(low = "white", high = "red", na.value = "grey") +

theme\_void()

*#--------------------------------------------------------*

*#--------------------------------------------------------*

*#--------------------------------------------------------*

library(plotly)

library(sf)

*#Prosperity Unemployment Rate --------------------------------------------*

df2 <- Prossperty\_Unemployment4[, c("GEO", "POPULATION", "UNEMPLOYMENT\_RATE")]

*# Add new column with province abbreviations to existing data frame*

df2 <- df2 %>%

mutate(PROVINCE = PROVINCE\_drop[GEO])

Prossperty\_Unem <- read\_csv("ProsUnem1.csv")

df2A <- Prossperty\_Unem[, c("GEO", "VALUE")]

names(df2A)[names(df2A) == 'VALUE'] <- 'UNEMPLOYMENT'

df2B <- cbind(df2, df2A$UNEMPLOYMENT)

names(df2B)[names(df2B) == 'df2A$UNEMPLOYMENT'] <- 'UNEMPLOYMENT'

*#df2B$POVERTY\_RATE <- (df1$IN\_POVERTY / df1$TOTAL) \* 100*

df2C <- df2B[, c("PROVINCE", "POPULATION", "UNEMPLOYMENT\_RATE")]

*# Create the plot*

ggplot(df2C, aes(x = PROVINCE, y = POPULATION)) +

geom\_bar(stat = "identity", fill = "#FA9203") +

geom\_text(aes(label = paste0(UNEMPLOYMENT\_RATE, "%")), vjust = -0.5, size = 4) +

labs(x = "Province", y = "Population (Thousands)", title = "Population vs Unemployment Rate by Province") +

theme(plot.title = element\_text(hjust = 0.5, margin = margin(b = 20), size = 18))

ggsave("Population vs Unemployment Rate by Province.jpeg", width = 8, height = 6, dpi = 300)

*#Health-adjusted life expectancy --------------------------------------------*

df3 <- Health\_Life4

*# Add new column with province abbreviations to existing data frame*

df3 <- df3 %>%

mutate(PROVINCE = PROVINCE\_drop[GEO])

df3A <- df3[, c("PROVINCE", "AT\_BIRTH", "AT\_AGE65")]

df3B <- df3A %>%

select(PROVINCE, AT\_BIRTH, AT\_AGE65) %>%

pivot\_longer(cols = c(AT\_BIRTH, AT\_AGE65),

names\_to = "Variable", values\_to = "Value") %>%

mutate(Variable = factor(Variable, levels = c("AT\_BIRTH", "AT\_AGE65")))

*# create grouped bar chart*

ggplot(df3B, aes(x = PROVINCE, y = Value, fill = Variable)) +

geom\_bar(position = "dodge", stat = "identity") +

xlab("Province") +

ylab("Years") + *# Change y-axis label*

ggtitle("Health-adjusted life expectancy") +

theme\_classic() +

scale\_fill\_discrete(name = "Year") +

labs(fill = "Year") +

scale\_fill\_manual(values = c("#E14F08", "#6D0E4E"),

guide = guide\_legend(title = "Health-adjusted life expectancy")) +

theme(plot.title = element\_text(hjust = 0.5, margin = margin(b = 20), size = 18))

ggsave("Health-adjusted life expectancy.jpeg", width = 8, height = 6, dpi = 1080)

*#Health-Perceived mental health --------------------------------------------*

df4 <- Health\_Mental5

*# Add new column with province abbreviations to existing data frame*

df4 <- df4 %>%

mutate(PROVINCE = PROVINCE\_drop[GEO])

df4A <- df4[, c("PROVINCE", "EXCELLENT\_OR\_VERYGOOD", "GOOD", "FAIR\_OR\_POOR")]

df4B <- df4A %>%

select(PROVINCE, EXCELLENT\_OR\_VERYGOOD, GOOD, FAIR\_OR\_POOR) %>%

pivot\_longer(cols = c(EXCELLENT\_OR\_VERYGOOD, GOOD, FAIR\_OR\_POOR),

names\_to = "Variable", values\_to = "Value") %>%

mutate(Variable = factor(Variable, levels = c("EXCELLENT\_OR\_VERYGOOD", "GOOD", "FAIR\_OR\_POOR")))

ggplot(df4B, aes(x = Value, y = reorder(PROVINCE, Value), fill = Variable)) +

geom\_col(position = "stack") +

geom\_text(aes(label = paste0(round(Value), "%")), position = position\_stack(vjust = 0.5)) +

labs(x = "Percentage", y = "Province",

title = "Perceived Mental Health") +

scale\_fill\_manual(values = c("#0072B2", "#462969", "#B31A15"),

name = "Health Perception") +

theme(plot.title = element\_text(hjust = 0.5, margin = margin(b = 20), size = 18))

ggsave("Perceived Mental Health.jpeg", width = 8, height = 6, dpi = 1080)

*#Society - Sense of belonging to local community ---------------------------*

df5 <- Social\_Belonging6

df5 <- df5 %>%

mutate(PROVINCE = PROVINCE\_drop[GEO])

df5A <- df5[, c("PROVINCE", "VERYSTRONG\_SOMEWHATSTRONG", "SOMEWHATWEAK\_VERYWEAK", "NO\_OPINION")]

df5B <- df5A %>%

select(PROVINCE, VERYSTRONG\_SOMEWHATSTRONG, SOMEWHATWEAK\_VERYWEAK, NO\_OPINION) %>%

pivot\_longer(cols = c(VERYSTRONG\_SOMEWHATSTRONG, SOMEWHATWEAK\_VERYWEAK, NO\_OPINION),

names\_to = "Variable", values\_to = "Value") %>%

mutate(Variable = factor(Variable, levels = c("VERYSTRONG\_SOMEWHATSTRONG", "SOMEWHATWEAK\_VERYWEAK", "NO\_OPINION")))

ggplot(df5B, aes(x = Value, y = reorder(PROVINCE, Value), fill = Variable)) +

geom\_col(position = "stack") +

geom\_text(aes(label = paste0(round(Value), "%")), position = position\_stack(vjust = 0.5)) +

labs(x = "Percentage", y = "Province",

title = "Sense of Belonging to Society") +

scale\_fill\_manual(values = c("#FA9203", "#0072B2", "#B31A15"),

name = "Belonging Perception") +

theme(plot.title = element\_text(hjust = 0.5, margin = margin(b = 20), size = 18))

ggsave("Sense of Belonging to Society.jpeg", width = 8, height = 6, dpi = 1080)

*#Environment - Greenhouse Gas Emissions ------------------------------------*

df6 <- Environment\_Gas

*# drop last 3 rows*

df6 <- head(df6, -3)

*#add province abbr*

df6$Abbr <- c("NL", "PE", "NS", "NB", "QC", "ON", "MB", "SK", "AB", "BC")

df6A <- df6[, c("Abbr", "Year\_1990", "Year\_2005", "Year\_2020")]

names(df6A)[names(df6A) == 'Abbr'] <- 'PROVINCE'

df6A$Year\_1990 <- **as**.numeric(df6A$Year\_1990)

df6B <- df6A %>%

select(PROVINCE, Year\_1990, Year\_2005, Year\_2020) %>%

pivot\_longer(cols = c(Year\_1990, Year\_2005, Year\_2020),

names\_to = "Variable", values\_to = "Value") %>%

mutate(Variable = factor(Variable, levels = c("Year\_1990", "Year\_2005", "Year\_2020")))

*# create grouped bar chart*

ggplot(df6B, aes(x = PROVINCE, y = Value, fill = Variable)) +

geom\_bar(position = "dodge", stat = "identity") +

xlab("Province") +

ylab("Emissions (CO2 Mt)") + *# Change y-axis label*

ggtitle("Greenhouse Gas Emission") +

theme\_classic() +

scale\_fill\_discrete(name = "Year") +

labs(fill = "Year") + *# Change legend title (alternative method)*

scale\_fill\_manual(values = c("#E14F08", "#462969", "#FA9203"),

guide = guide\_legend(title = "Greenhouse Gas Emission")) +

theme(plot.title = element\_text(hjust = 0.5, margin = margin(b = 20), size = 18))

ggsave("Greenhouse Gas Emission.jpeg", width = 8, height = 6, dpi = 1080)

*#Good Governance - Personal Safety ------------------------------------*

df7 <- Governance\_Victim

*# drop last 3 rows and top row*

df7 <- head(df7, -3)

df7 <- slice(df7, -1)

*#add province abbr*

df7$Abbr <- c("NL", "PE", "NS", "NB", "QC", "ON", "MB", "SK", "AB", "BC")

df7A <- df7[, c("Abbr", "Victimization", "Violent\_Victimization")]

names(df7A)[names(df7A) == 'Abbr'] <- 'PROVINCE'

df7A

my\_colors <- c("#E69F00", "#ff6e40", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7","#26495c", "#c4a35a", "#c66b3d")

ggplot(df7A, aes(x = Victimization, y = Violent\_Victimization, color = PROVINCE)) +

geom\_point(size = 5, aes(color = PROVINCE)) +

geom\_text(aes(label = PROVINCE), size = 3, vjust = 2, color = "black") +

labs(title = "Reports Made of Violent and Non-Violent Victimization", x = "Non-Violent", y = "Violent") +

theme(plot.title = element\_text(hjust = 0.5, margin = margin(b = 20), size = 15)) +

scale\_color\_manual(values = my\_colors) +

guides(color = "none")

ggsave("Reports Made of Violent and Non-Violent Victimization.jpeg", width = 8, height = 6, dpi = 1080)

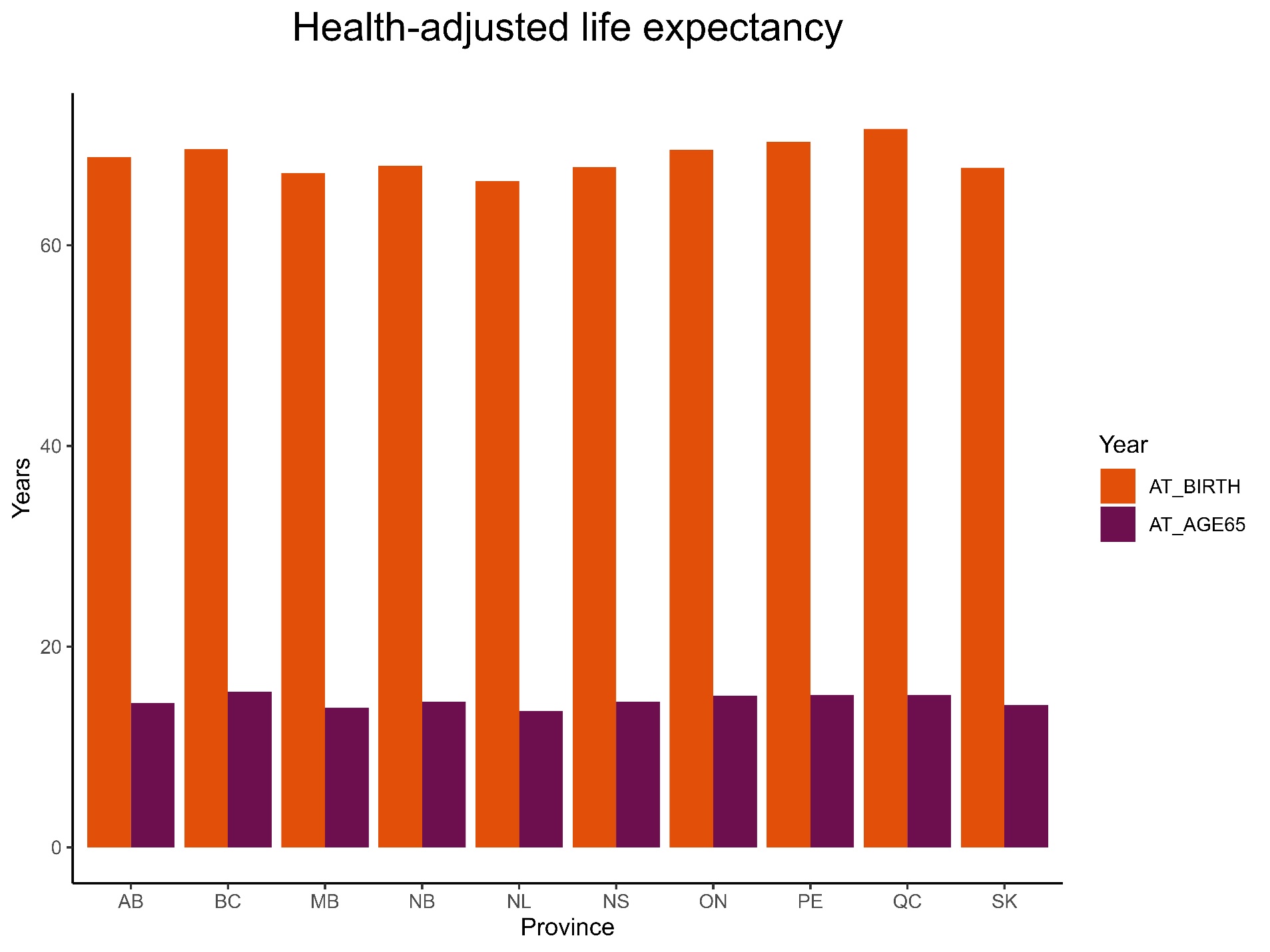
*############################################################################*

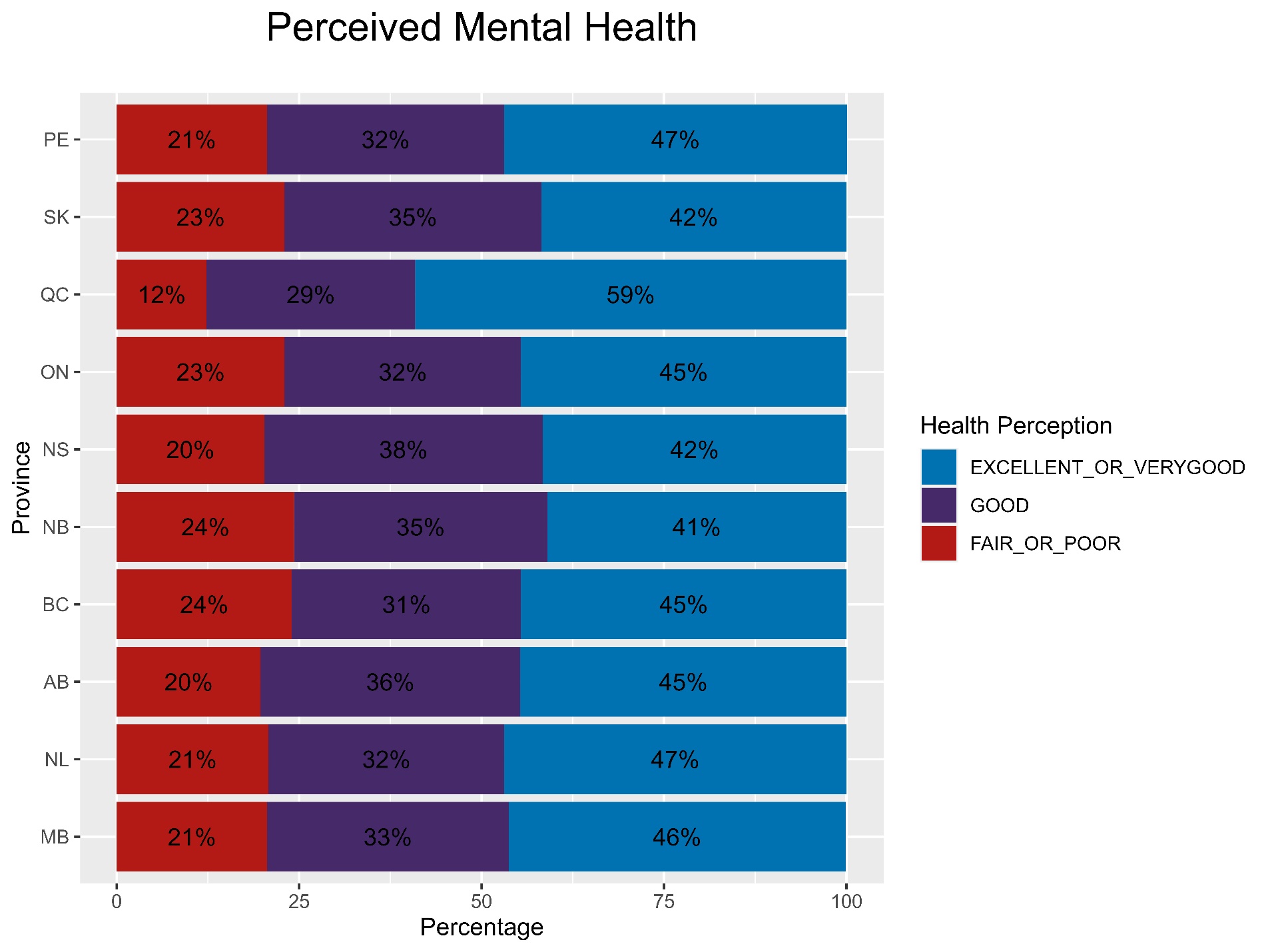
*############################################################################*

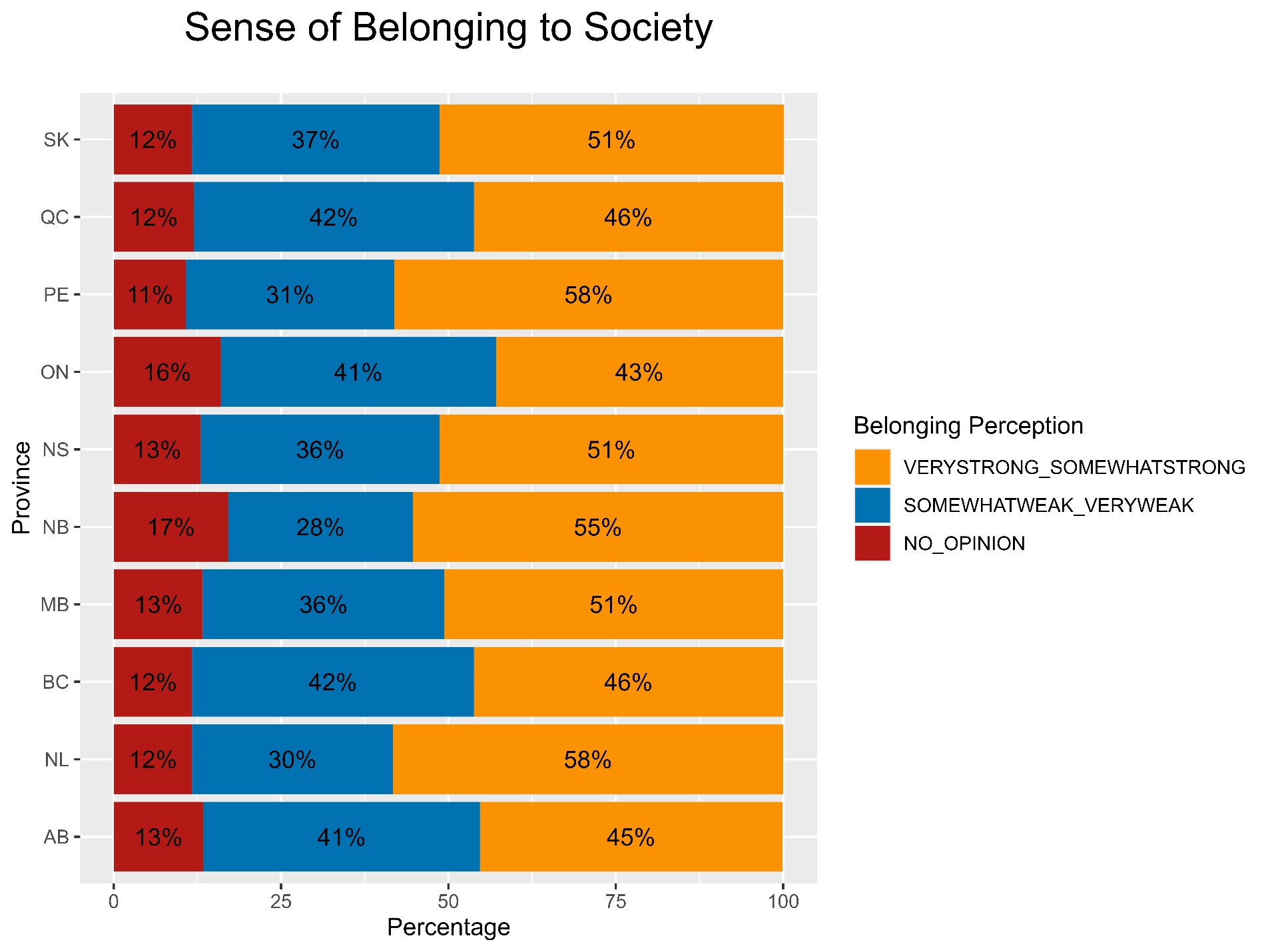
*############################################################################*

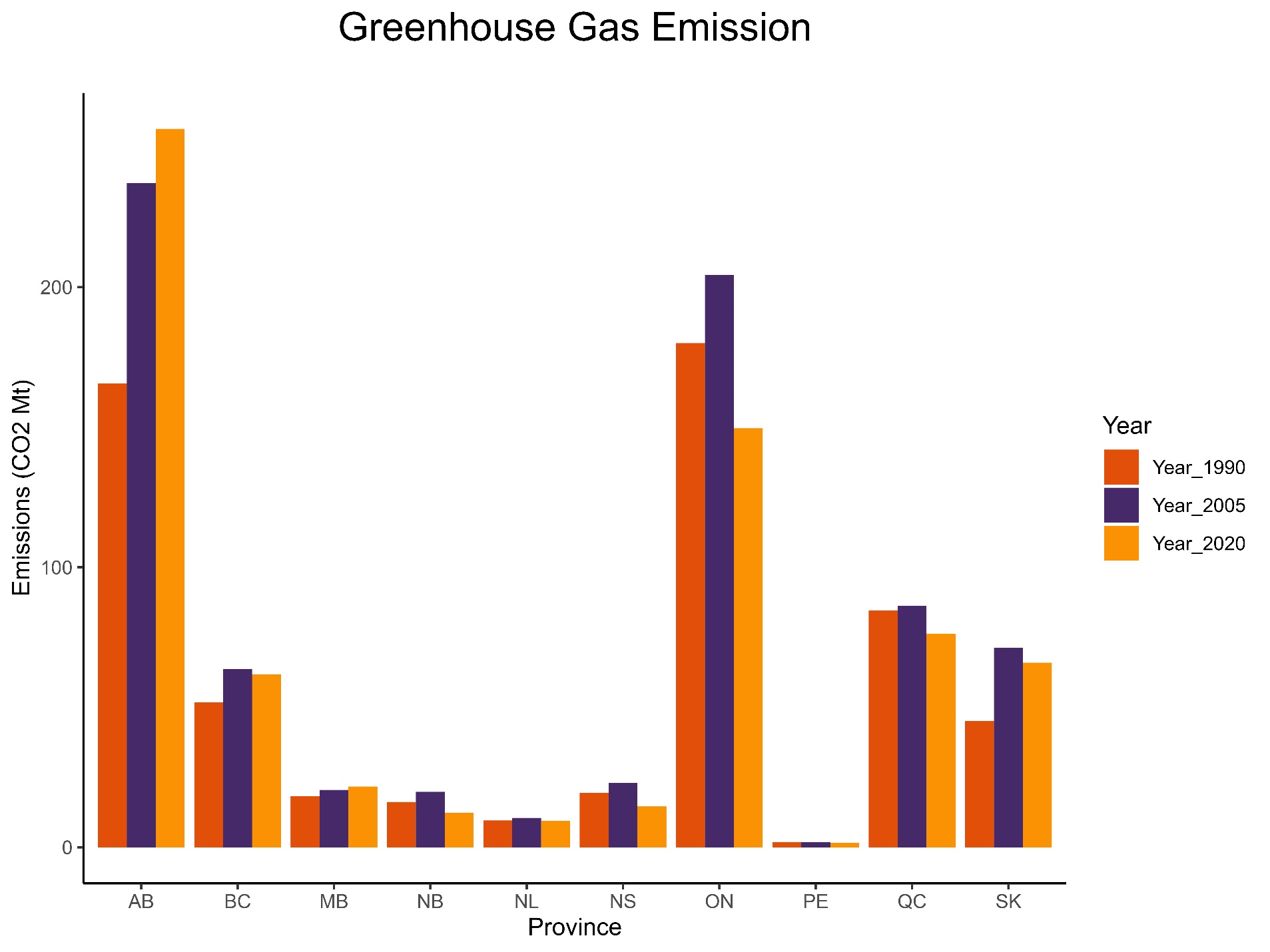
Chart, bar chart

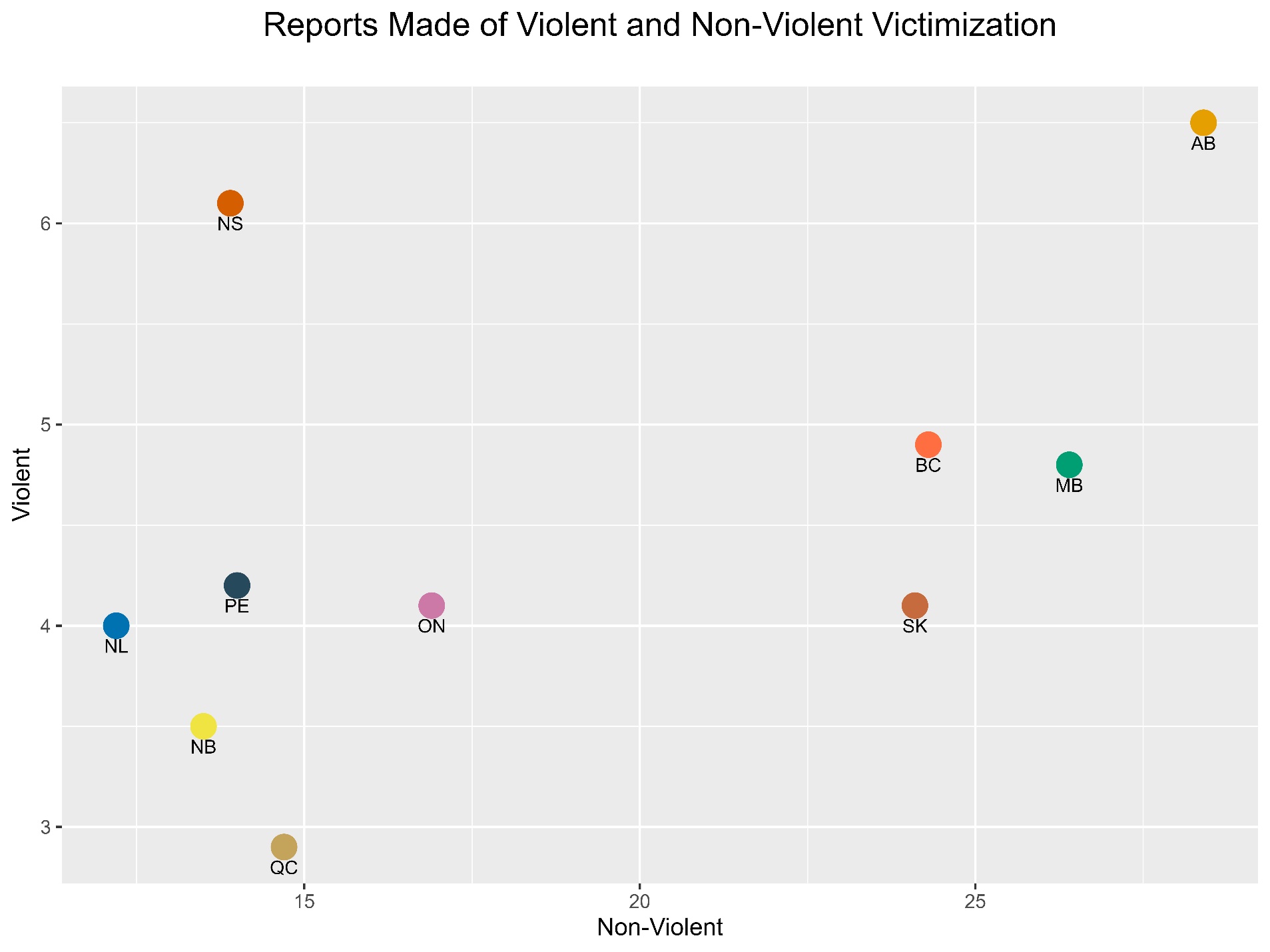
Description automatically generated











**Canada of Life - Infograph**

