HUMBER INSTITUTE OF TECHNOLOGY
AND ADVANCED LEARNING
(HUMBER COLLEGE)

Project: Quality of Life Canada Analysis

Submitted to:

Professor Ammar Al-Qaraghuli

Submission Date: 08/12/2023

Submitted by: Vaishali

1. Prosperity Domain- Job Satisfaction by Province

```r
#loading necessary libraries

library(tidyverse)

library(ggplot2)

library(dplyr)

#reading data

job_data = read.csv("C:/Users/Simran/Desktop/Big Data2/jobdata.csv")

print(job_data)

#filtering required data

job_data_set<- subset(job_data, Gender == "Total, all persons" & Indicators == "Very satisfied or satisfied
with job")

print(job_data_set)

#tidying data

job_new <- job_data_set[, c("GEO", "Indicators", "VALUE")]

colnames(job_new) <- c("Province", "Satisfaction", "Percentage")

jobdata1 <- subset(job_new, Province != "Canada (excluding territories)")

print(jobdata1)

#Plotting the graph

ggplot(jobdata1, aes(x = reorder(Province, Percentage), y = Percentage, fill = Province)) +

  geom_bar(stat = "identity") +

  geom_text(aes(label = paste0(Percentage, "%")), vjust = -0.5, color = "black", size = 3) +

  labs(title = "Job Satisfaction by Province",

      x = "Provinces",

      y = "Percentage (%)",

      fill = "Provinces") +

  scale_fill_brewer(palette = "Paired") +

  theme_minimal() +

  theme(

    plot.title = element_text(size = 16, face = "bold"),

    legend.position = "none",
```
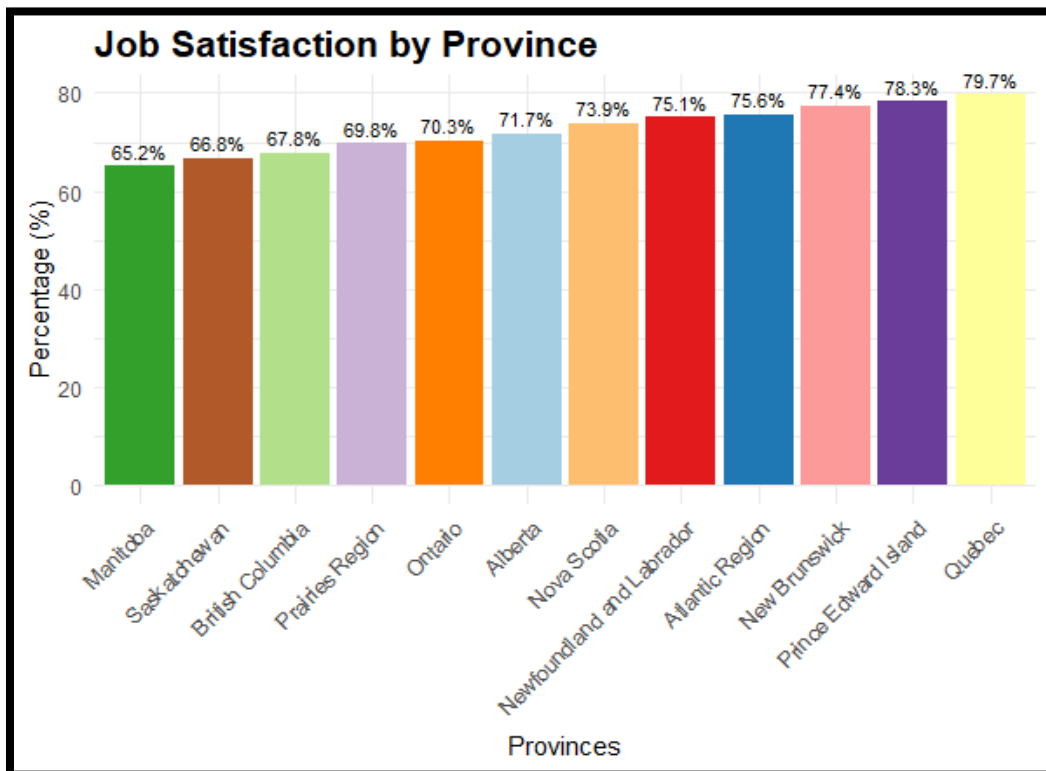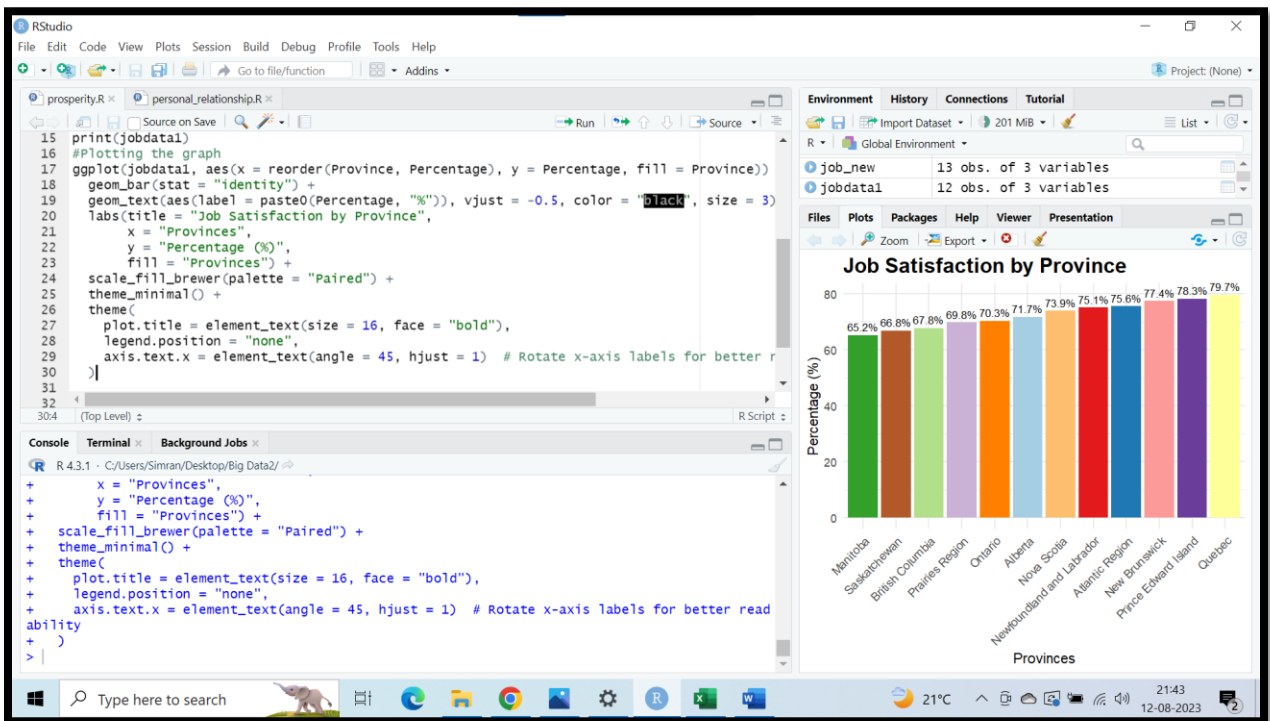
axis.text.x = element_text(angle = 45, hjust = 1) # Rotate x-axis labels for better readability

)

2. Society Analysis- Personal Relationship Satisfaction by Province

```r
#loading necessary libraries

library(tidyverse)

library(ggplot2)

library(dplyr)

#reading data

personal_data = read.csv("C:/Users/Simran/Desktop/Big Data2/personaldata.csv")

print(personal_data)

#filtering required data

personal_data_set<- subset(personal_data, Gender == "Total, all persons" & Indicators == "Satisfaction
with personal relationships rating of 8, 9 or 10")

print(personal_data_set)

#tidying data

personal_new <- personal_data_set[, c("GEO", "Indicators", "VALUE")]

 colnames(personal_new) <- c("Province", "Satisfaction", "Percentage")

life_data <- subset(personal_new, Province != "Canada (excluding territories)")

print(life_data)

library(ggplot2)

library(RColorBrewer)

#library(RColorBrewer)


# Define the base colors for the ombre effect

base_colors <- brewer.pal(12, "Set1")  # Using Brewer palette as base colors

num_shades <- 5  # Number of shades for each color

ombre_palette <- character(length(base_colors) * num_shades)

for (i in seq_along(base_colors)) {

  shades <- colorRampPalette(c("white", base_colors[i]))(num_shades + 1)

  ombre_palette[((i - 1) * num_shades + 1):(i * num_shades)] <- shades[-1]

}
```

```
ggplot(life_data, aes(x = reorder(Province, Percentage), y = Percentage, fill = Province)) +

  geom_bar(stat = "identity") +

  geom_text(aes(label = paste0(Percentage, "%")), vjust = -0.5, color = "black", size = 3) +

  labs(title = "Personal Satisfaction by Province",

       x = "Provinces",

       y = "Percentage (%)",

       fill = "Provinces") +

  scale_fill_manual(values = ombre_palette) +  # Use the ombre palette here

  theme_minimal() +

  theme(

    plot.title = element_text(size = 16, face = "bold"),

    legend.position = "none",

    axis.text.x = element_text(angle = 45, hjust = 1)

  )
```

Personal Satisfaction by Province
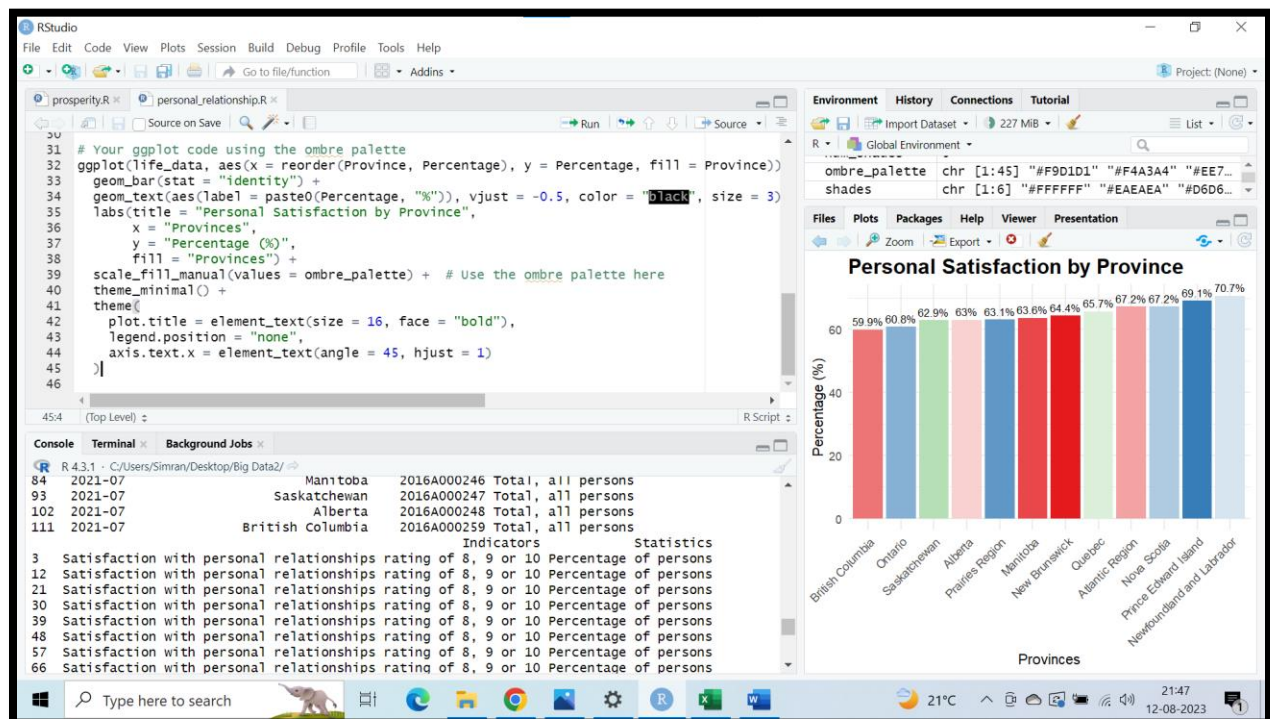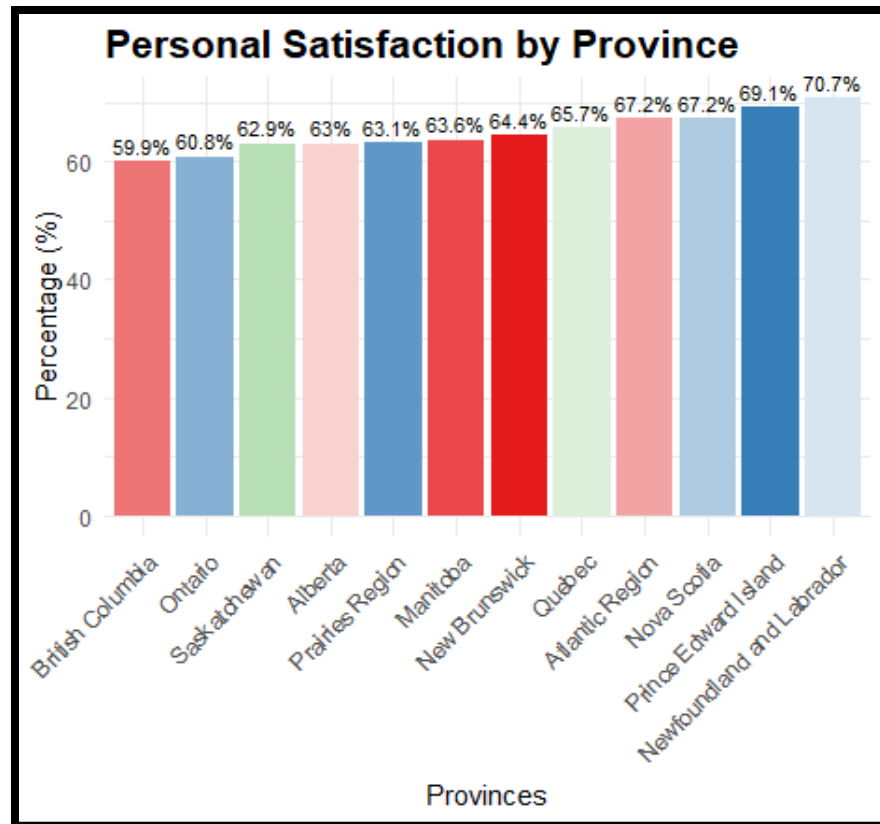
3. Health Domain Analysis

```
# Load the necessary packages

library('tidyverse')

library(lubridate)

library(ggplot2)

library(RColorBrewer)

library(ggthemes)

#get Working directory

getwd()

############################################### HEALTH-DATA-ANALYSIS
#################################################################
```

```r
# Read the first CSV file for Health data at Birth time

Birth_Data <- read.csv("C:/Users/Simran/Desktop/Big Data2/Health_Birth_Data.csv")

#Replace all missing values with NA

Birth_Data <- Birth_Data %>%

  mutate_all(~ ifelse(is.na(.), NA, .))

head(Birth_Data)

Birth_Data

# Read the second CSV file for Health data at 65 Yrs above age.

Above65_Data <- read.csv("C:/Users/Simran/Desktop/Big Data2/Health_65above_data.csv")


#Replace all missing values with NA

Above65_Data <- Above65_Data %>%

  mutate_all(~ ifelse(is.na(.), NA, .))

head(Above65_Data)

Above65_Data

# Check data types of common columns in Birth_Data

cat("Data types of common columns in Birth_Data:\n")

str(Birth_Data[c("REF_DATE", "GEO", "Age.group", "Sex", "Income.group", "VALUE")])

# Check data types of common columns in Above65_Data

cat("\nData types of common columns in Above65_Data:\n")

str(Above65_Data[c("REF_DATE", "GEO", "Age.group", "Sex", "Income.group", "VALUE")])

# Merge the files based on multiple common columns

Health_merged_data <- merge(Birth_Data, Above65_Data, by = c("REF_DATE", "GEO", "Age.group",
"Sex", "Income.group", "VALUE"), all.x=TRUE, all.y = TRUE)

print(Health_merged_data)

#Selecting only useful columns

Health_merged_data <- Health_merged_data %>%

  select(REF_DATE,GEO,Age.group,Sex,Income.group,VALUE)
```

```r
print(Health_merged_data)


# Create a custom color palette for Age.group categories
custom_colors <- c("#E41A1C", "#377EB8")


# Create a new PDF device with a custom width and height
pdf("Health Data Analysis QOL.pdf", width = 20, height = 10)


# Create the bar plot using ggplot with the custom color palette
ggplot(Health_merged_data, aes(x = GEO, y = VALUE, fill = factor(Age.group))) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(rows = vars(Sex), cols = vars(Income.group)) +
  labs(title = "Health Data Analysis",
      x = "GEO",
      y = "Health Adjusted Life Expectancy by Sex",
      fill = "Age group") +
  scale_fill_manual(values = custom_colors) +  # Specify the custom color palette
  theme_minimal()


# Close the PDF device
dev.off()
```
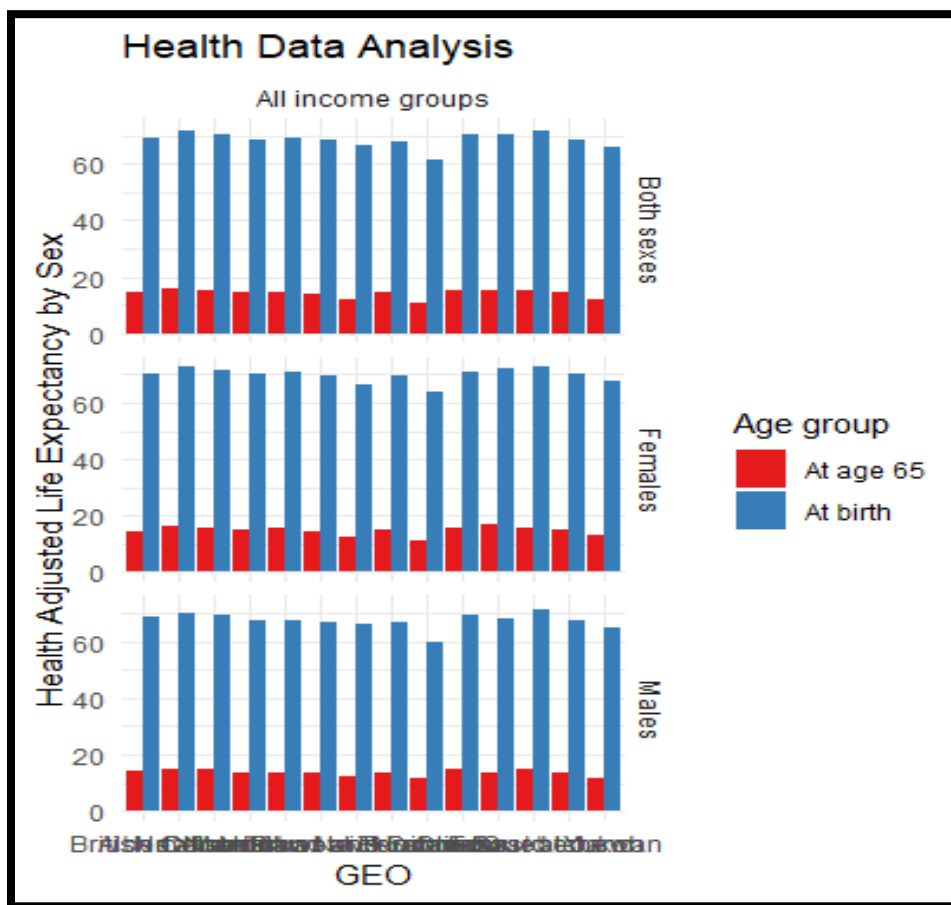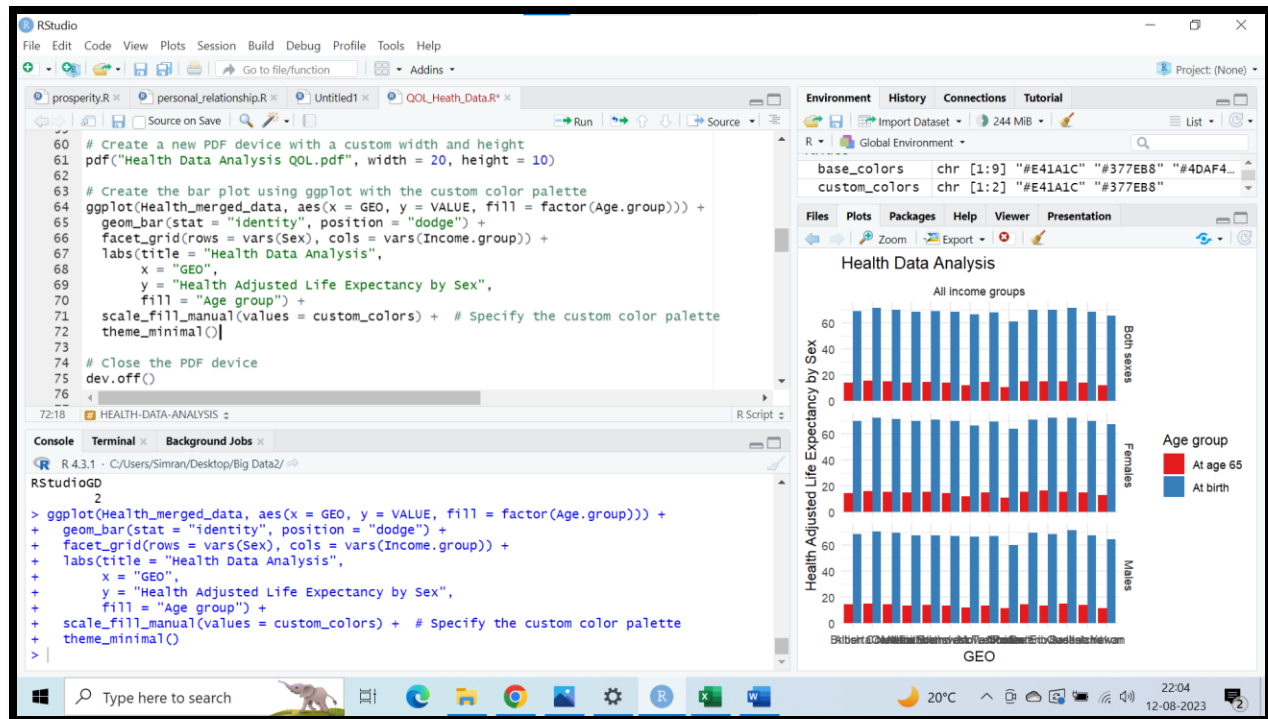
4. Good-Governance Data Analysis

```
# Load the necessary packages
library('tidyverse')
library(lubridate)
library(ggplot2)
library(RColorBrewer)
library(ggthemes)

#get Working directory
getwd()
```

Read the first CSV file for Governance data for age above 18 yrs

Above18_Data <- read.csv("C:/Users/Simran/Desktop/Big Data2/Governance_18above.csv")

Get the coloum names

head(Above18_Data)

Read the second CSV files for Governance data for Men

Men_Data <- read.csv("C:/Users/Simran/Desktop/Big Data2/Governance_Men.csv")

Get the coloum names

head(Men_Data)

Read the third CSV file for Governance data for Women

Women_Data <- read.csv("Governance_Women.csv")

Get the column names of the Women_Data data frame

head(Women_Data)

Check data types of common columns in Above18_Data

cat("Data types of common columns in Above18_Data:\n")

str(Above18_Data[c("REF_DATE", "GEO", "Visible.minority", "Selected.sociodemographic.characteristics", "Indicators", "VALUE")])

Check data types of common columns in Governance_Men

```r
cat("Data types of common columns in Above18_Data:\n")

str(Men_Data[c("REF_DATE", "GEO", "Visible.minority", "Selected.sociodemographic.characteristics",
"Indicators", "VALUE")])

# Check data types of common columns in Governance_Women

cat("Data types of common columns in Above18_Data:\n")

str(Women_Data[c("REF_DATE", "GEO", "Visible.minority", "Selected.sociodemographic.characteristics",
"Indicators", "VALUE")])

# Merge Above18_Data and Men_Data

merged_data1 <- merge(Above18_Data, Men_Data,

              by = c("REF_DATE", "GEO", "Visible.minority", "Selected.sociodemographic.characteristics",
"Indicators", "VALUE"),

              all.x = TRUE, all.y = TRUE)

# Merge merged_data1 and Women_Data

Governance_merged_data <- merge(merged_data1, Women_Data,

                by = c("REF_DATE", "GEO", "Visible.minority",
"Selected.sociodemographic.characteristics", "Indicators", "VALUE"),

                all.x = TRUE, all.y = TRUE)

#Print the merged database

Governance_merged_data


#Selecting only useful columns

Governance_merged_data <- Governance_merged_data %>%

  select(REF_DATE,GEO,Visible.minority,Selected.sociodemographic.characteristics,Indicators,VALUE)

#Print the merged database

Governance_merged_data

# Create a custom color palette for Visible.minority categories

#custom_colors <- c("#E41A1C", "#377EB8", "#E69F00",
"#F0E442","#0072B2","#D55E00","#CC79A7","#000000", "#999999", "#8DD3C7")

# Create a new PDF device with a custom width and height

pdf("Good Governance Data Analysis QOL.pdf", width = 20, height = 10)
```

```r
# Create a custom color palette

custom_colors <- c("#E41A1C", "#377EB8", "#4DAF4A", "#984EA3", "#FF7F00", "#FFFF33")

# Create the bar plot using ggplot

ggplot(Governance_merged_data, aes(x = Visible.minority, y = VALUE, fill = Indicators)) +

  geom_bar(stat = "identity", position = "dodge") +

  facet_grid(rows = vars(Selected.sociodemographic.characteristics)) +

  labs(title = "Good Governance Data Analysis",

      x = "Visible Minority",

      y = "Value",

      fill = "Indicators") +

  scale_fill_manual(values = custom_colors) +

  theme_minimal()

# Close the PDF device

dev.off()
```
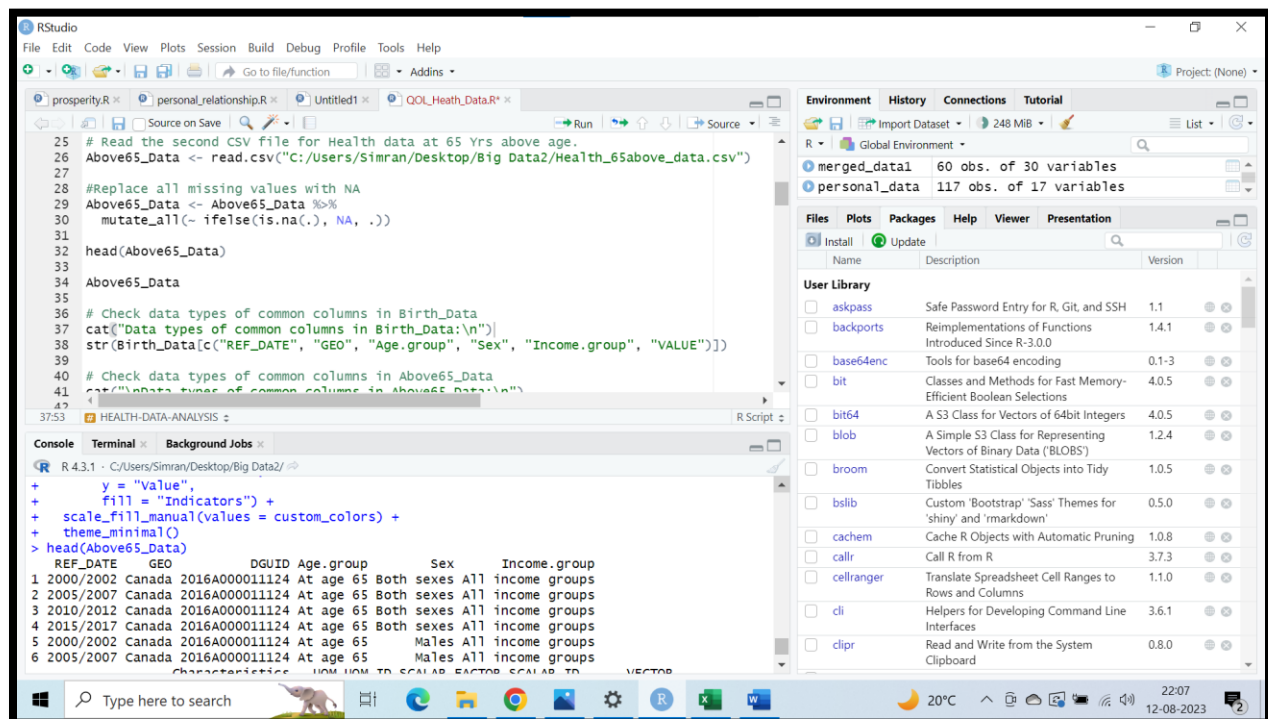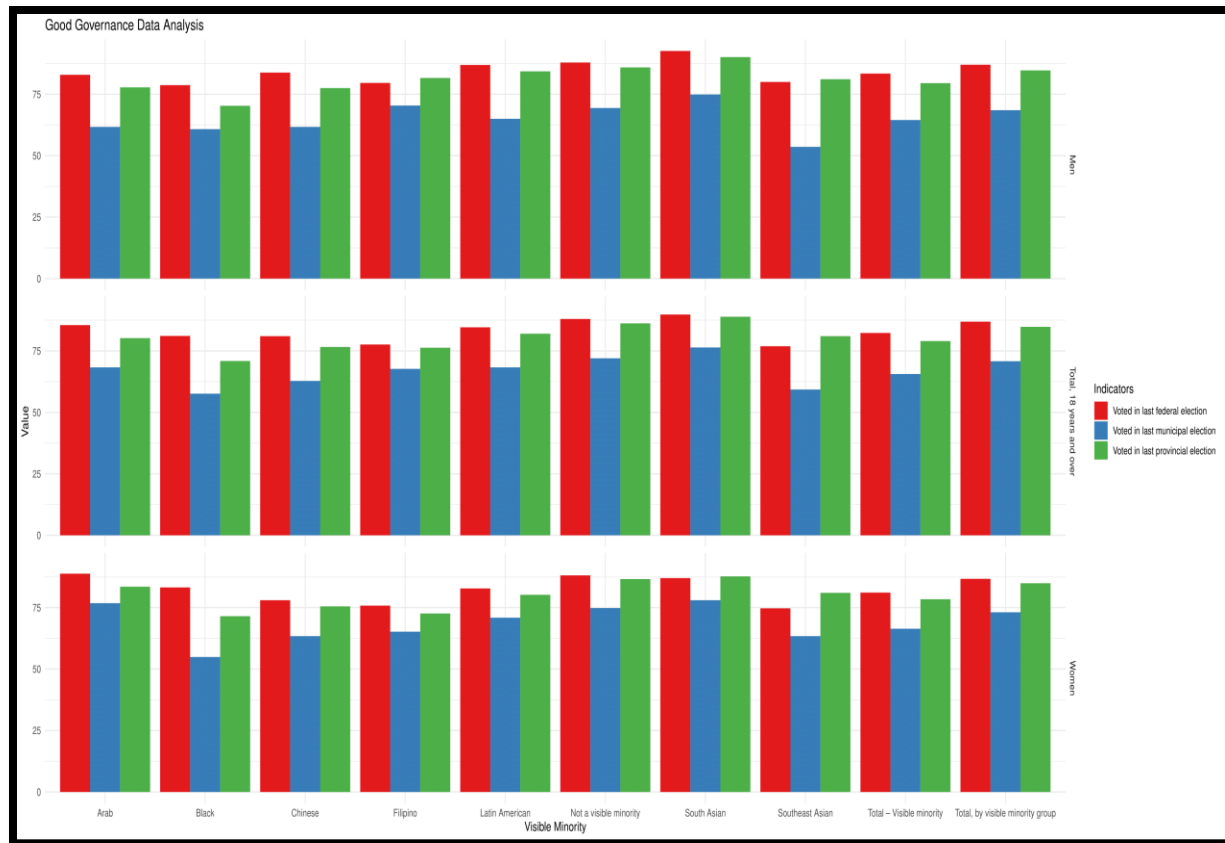
Good Governance Data Analysis

5. Environment Analysis- Satisfaction with local environment by province Q4 2022

#load dataset

provincedata <- read.csv("environment_by_province.csv")

#load ggplot2 for bar graph

install.packages("ggplot2")

library(ggplot2)

#filter the data to keep only what is needed

provincedata_filtered <- subset(provincedata, Gender == "Total, all persons" & Indicators == "Satisfaction with local environment rating of 8, 9 or 10")

#rename the columns for clarity

df_new <- provincedata_filtered[, c("GEO", "Indicators", "VALUE")]

colnames(df_new) <- c("Province", "Satisfaction_Score", "Percentage")

#print the new dataframe to check results

print(df_new)

#remove the row with "Canada (excluding territories)" as we don't need it

df_new_filtered <- subset(df_new, Province != "Canada (excluding territories)")

#print the new filtered dataframe

print(df_new_filtered)

#create the bar chart

ggplot(df_new_filtered, aes(x = Province, y = Percentage, fill = Province)) +

  geom_bar(stat = "identity") +

  labs(title = "Satisfaction Scores 8,9, or 10 by Province",

      x = "Province",

      y = "Percentage",

      fill = "Province") +

  theme_minimal() +

  theme(axis.text.x = element_blank(),
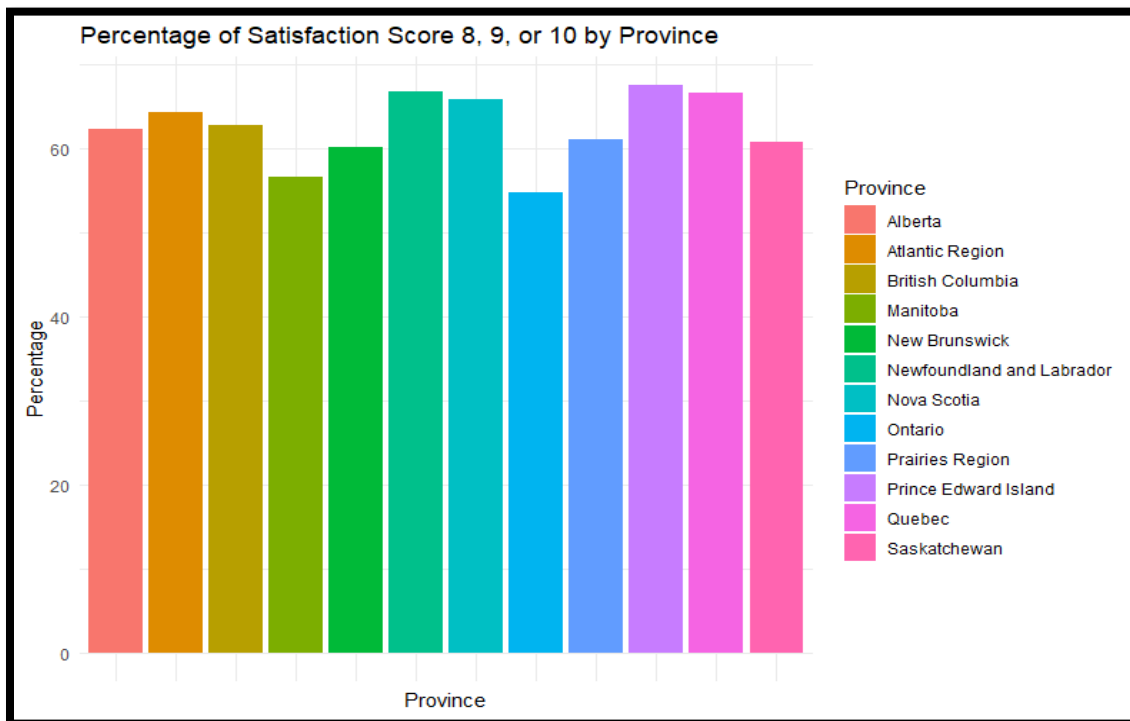
INFOGRAPHIC

QUALITY OF LIFE
CANADA

WHAT IS IT?

Quality of life refers to the wealth and comfort of individuals, communities and society based on both material and non-material factors.

Health
Refers to physical well-being, mental health, social interactions, and various aspects of daily functioning

Prosperity
Includes financial well-being, personal fulfillment, and overall quality of life.

Society
Equality, justice, and inclusivity within a society can significantly impact the quality of life experienced by its members.

Good Governance
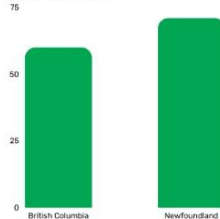Transparency, efficient services, economic opportunities, and citizen participation within a society.
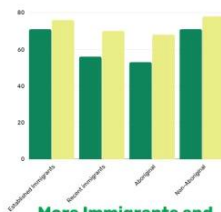
Environment
Includes factors such as access to clean air and water, natural beauty, ecological sustainability

73%

Canadians are **highly satisfied** with their jobs.

British Columbia | Newfoundland

Personal Satisfaction **10% higher** in Newfoundland (71%) as compared to the lowest in BC (60%)

Established Immigrants | Recent Immigrants | Aboriginal | Non-Aboriginal

More Immigrants and Aboriginals voted in the last federal elections as compared to the one before
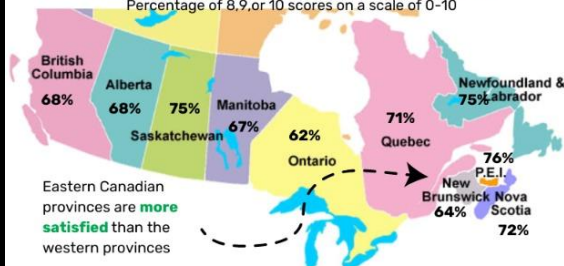
70 Years
Life expectancy at birth in Canada

60 Years
Lowest in Nunavut

MORE THAN 60% CANADIANS ARE HAPPY WITH THEIR LOCAL ENVIRONMENT

Percentage of 8,9,or 10 scores on a scale of 0-10

British Columbia 68%
Alberta 68%
75%
Manitoba 67%
Saskatchewan
62% Ontario
71% Quebec
Newfoundland & Labrador 75%
76% P.E.I.
New Brunswick 64%
Nova Scotia 72%

Eastern Canadian provinces are **more satisfied** than the western provinces