Quality of Life in Canada

Analytics by R

Nouman Ahmed Shah Khan
Thi Mong Thuy Vo
Dmytro Tretiakov
Yu-Shen Ma

TABLE OF CONTENTS

INTRODUCTION

- Quality of Life includes economic, psychological, social, and existential well-being.
- Understanding Quality of Life helps identify inequalities and guide better public policy.

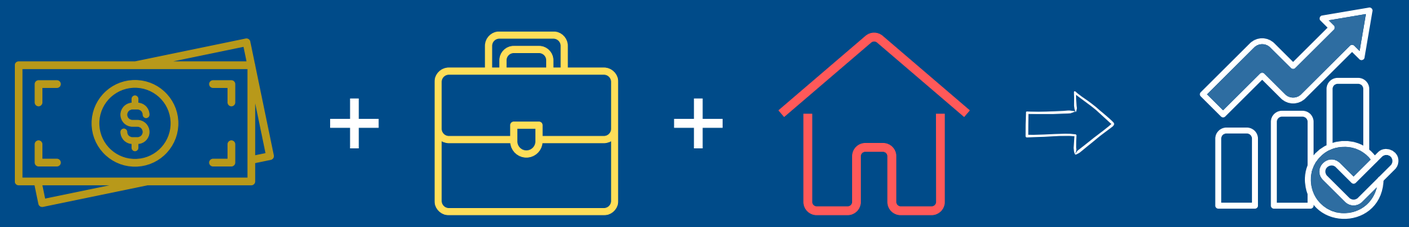
- Poverty: captures material living conditions and access to basic needs.
- Mental Health: reflects emotional well-being and levels of stress across society.
- Future Outlook: measures optimism, confidence, and expectations for the near future.
- Sense of Meaning & Purpose: explores deeper fulfillment and life satisfaction.

Prosperity

By Nouman

Definition :

- Prosperity refers to the economic well-being and financial success of individuals and society. It's about having the resources, opportunities, and security needed to live a comfortable life.

Prosperity = 

POVERTY

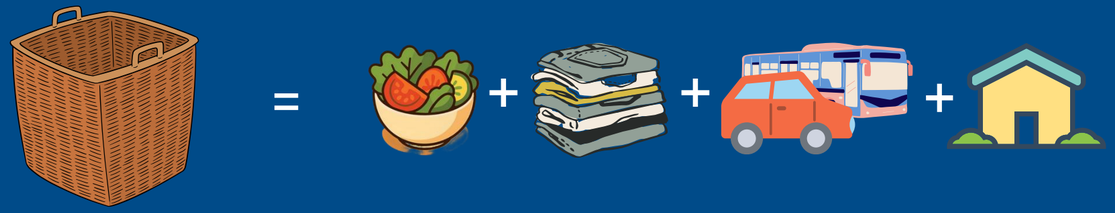
A condition where individuals or communities lack the financial resources and essentials for a basic standard of living

Poverty as an Indicator

- Definition: The percentage of the population living below Canada's official poverty line based on the Market Basket Measure

MARKET BASKET MEASURE (MBM)

A family lives in poverty if they cannot afford a specific basket of goods and services representing a modest standard of living.

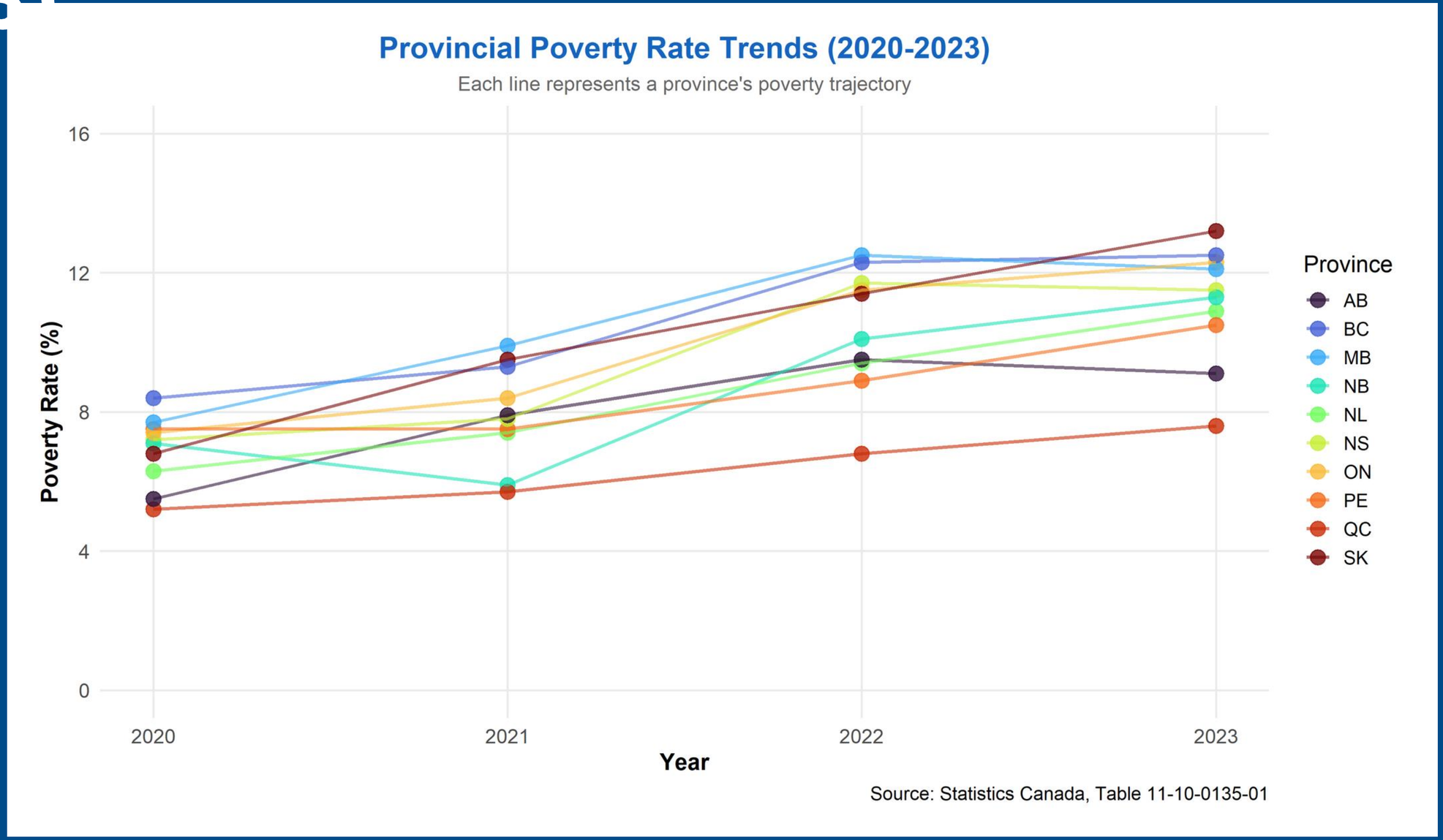


10.9%

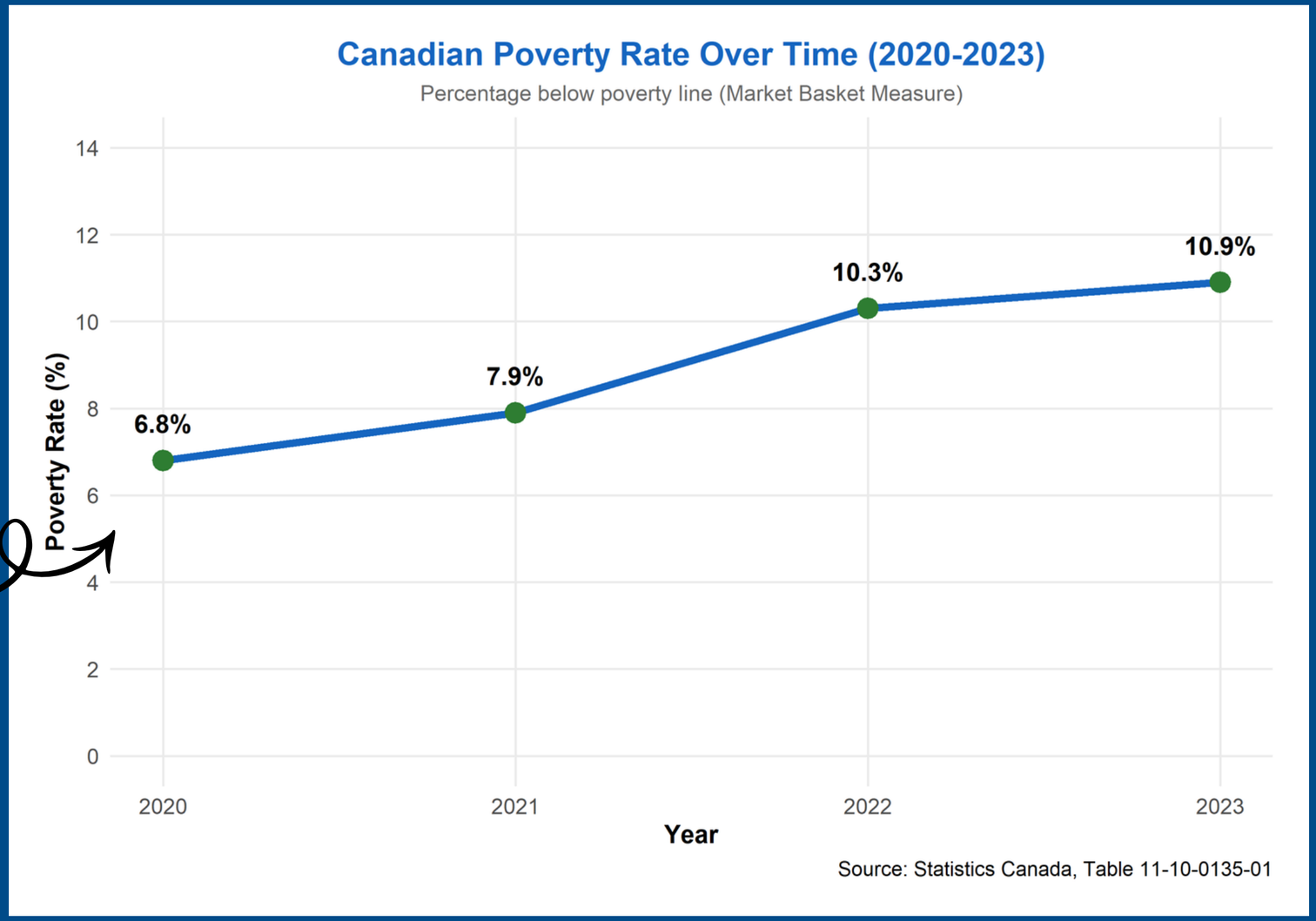
of Canadians live below the poverty line

Data Source: Canadian Income Survey

Poverty Rate Over Time (2020-2023)



Poverty Rate Over Time (2020-2023)

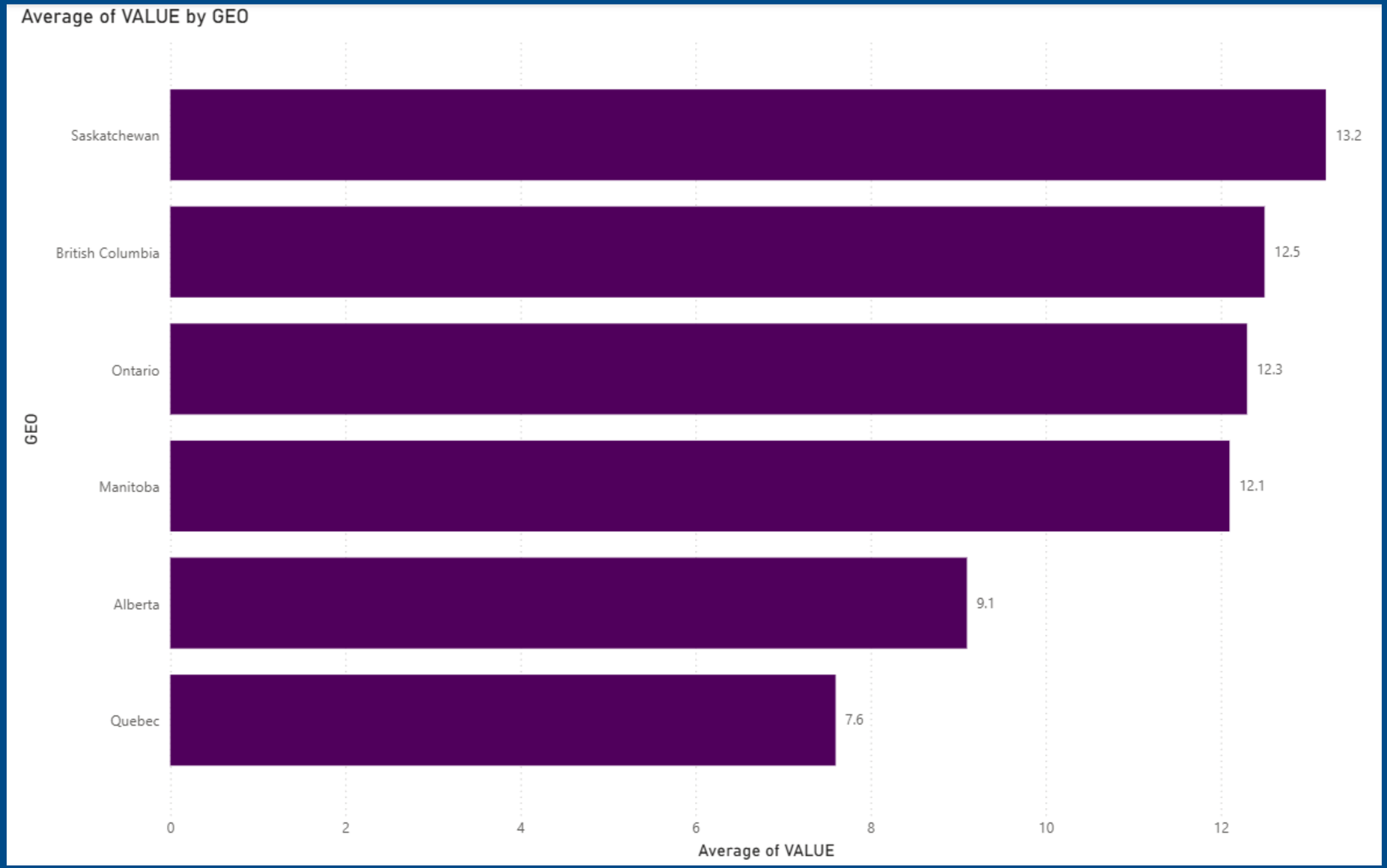


2020's low rate 6.8% was due to COVID-19 emergency benefits (CERB), not actual poverty reduction.

Steepest increase occurred between 2021-2022 2.4%

Overall Change of +4.1 % points

Poverty Rate by Province

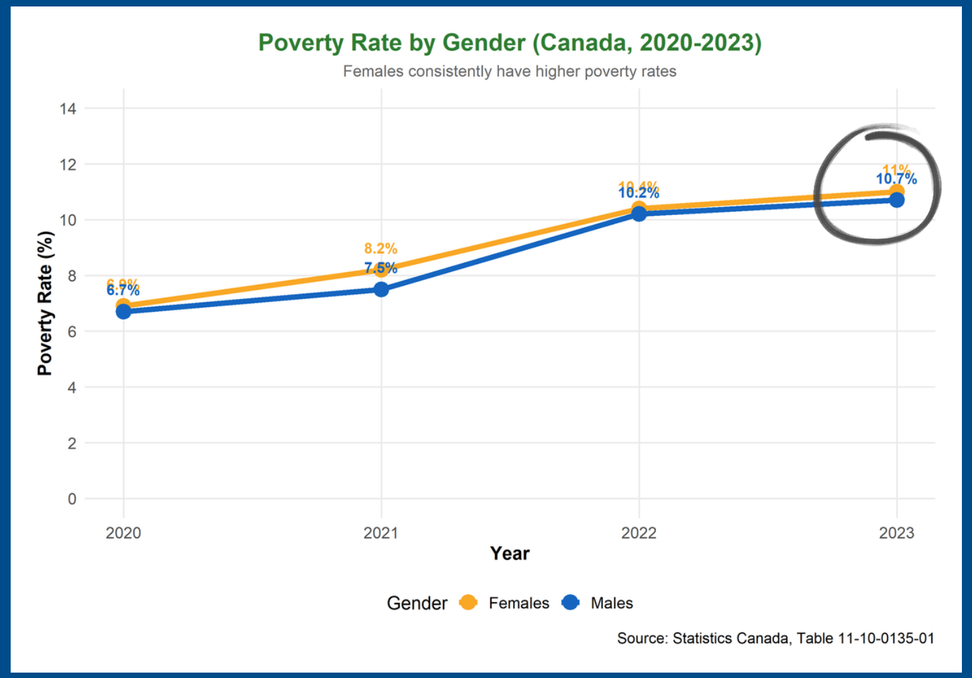


MAX- 13.2

5.6 percentage points separate highest and lowest provinces

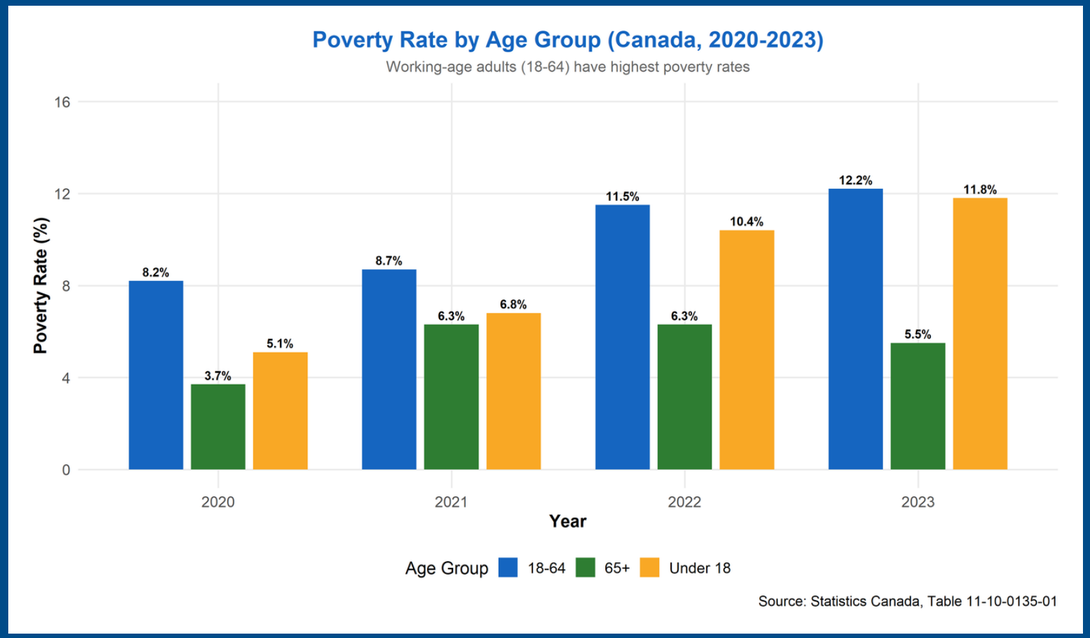
MIN- 7.6

Poverty by Demographics



Females : 11.0%
Males: 10.7%

Gender gap: 0.3 percentage points



Under 18 : 11.8%

18-64 : 12.2%

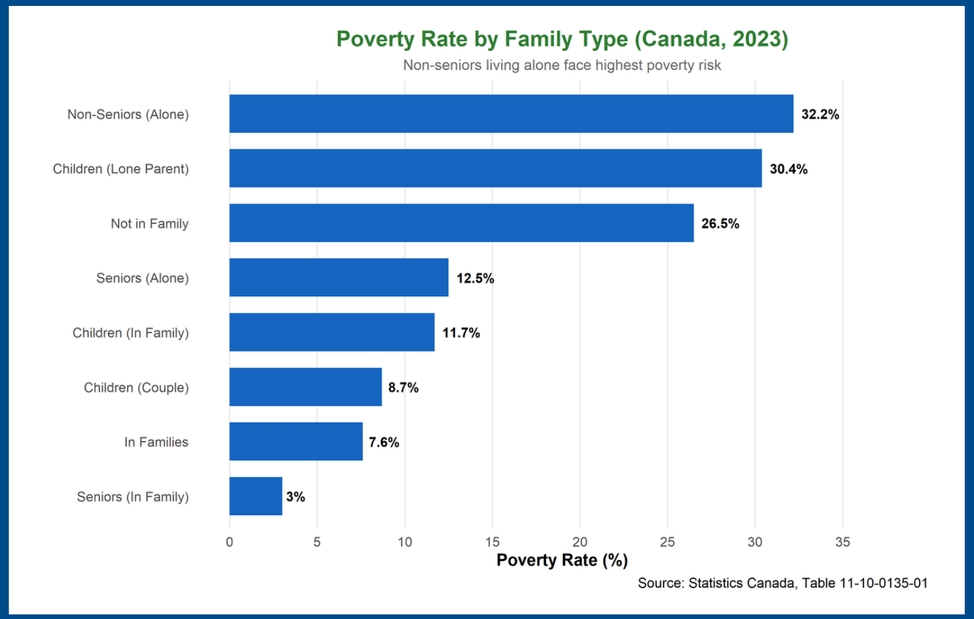
65+ : 5.5%

Working-age adults face **highest poverty risk gap**

⚠ HIGHEST RISK

Non-seniors alone: 32.2%

Lone-parent children: 30.4%



Key Findings

RISING TREND

+4.1%

Poverty rose from 6.8% to 10.9% as COVID benefits ended

CHILDREN AT RISK

30.4%

Children in lone-parent families most vulnerable

LIVING ALONE

32.2%

Non-seniors alone face the highest poverty rate

REGIONAL GAP

4.3%

Gap between highest (BC) and lowest (QC) province

Mental Health

By Thuy

Mental Health in CANADA

Overview:

- Good mental health declined from 2020–2022, largely influenced by the pandemic.
- Stress levels increased, with the strongest impact on young adults (18–34) and women.
- Regional differences are clear, some provinces recovered faster while others saw deeper declines.

Study Period: 2016 - 2022

Data Sources: Statistics Canada

```
> summary_table
# A tibble: 14 × 3
   Location                Total_Person Avg_Percentage
   <chr>                   <chr>              <dbl>
 1 Alberta                 13,720,700          32.1
 2 British_Columbia        15,597,100          31.4
 3 Canada                  118,962,100         32
 4 Manitoba                3,891,200           30.9
 5 New_Brunswick           2,338,900           30.4
 6 Newfoundland_Labrador   1,598,300           29.9
 7 Northwest_Territories   118,300             29.5
 8 Nova_Scotia             3,038,700           31.2
 9 Nunavut                 78,600              27.5
10 Ontario                 46,261,200          32
11 Prince_Edward           481,500             30.7
12 Quebec                  28,409,500          33.2
13 Saskatchewan            3,305,700           30.4
14 Yukon                   120,900             30.7
```

Table: Gender Categories

Gender
Both_Genders
Males
Females

Table: Indicator Types

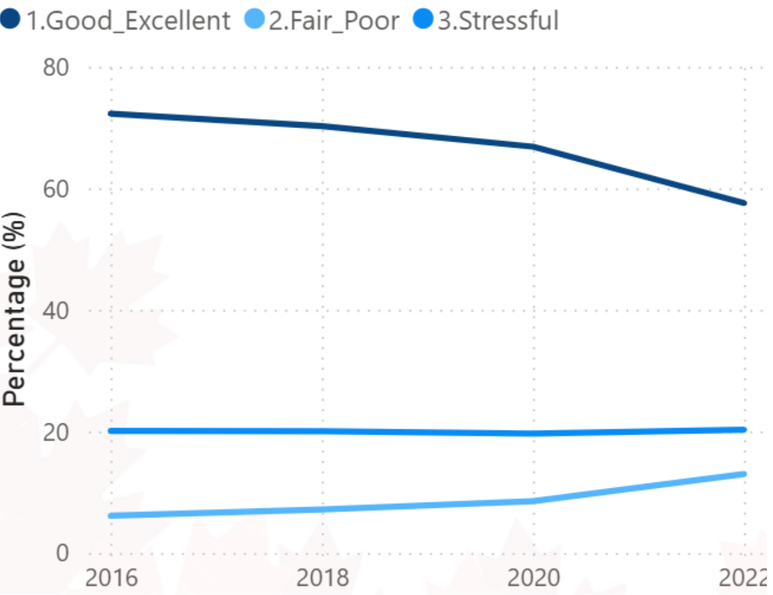
Indicators
Good_Excellent
Fair_Poor
Stressful

Table: Age Groups in Dataset

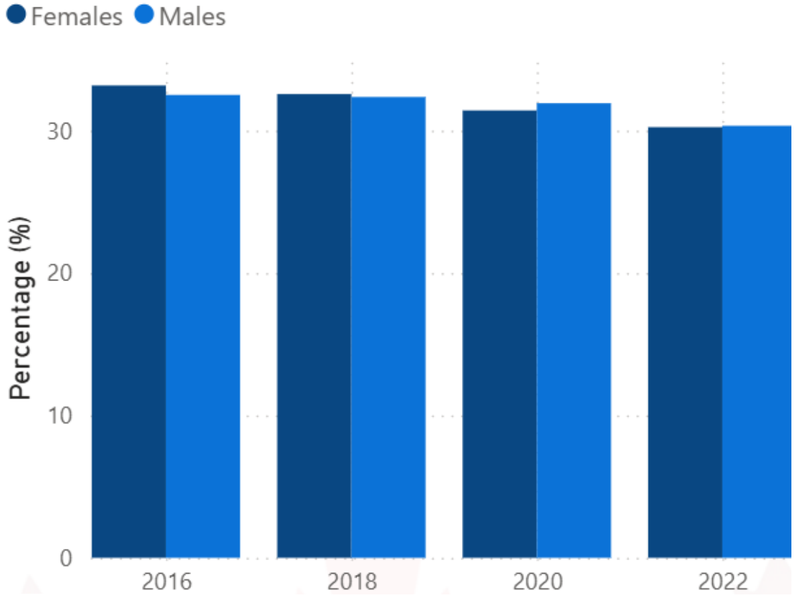
Age_Group	Age_Range
Total	12 years and over
Teen	12 to 17 years
Young	18 to 34 years
Midlife	35 to 49 years
Older	50 to 64 years
Senior	65 years and over

```
> summary(health)
   Location          Age_group           Gender           Indicators
 Length:2657        Length:2657        Length:2657        Length:2657
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character

      Year          Person          Percentage
 Min.   :2016   Min.   :     200   Min.   : 2.50
 1st Qu.:2016   1st Qu.:    9200   1st Qu.:10.50
 Median :2018   Median :   45350   Median :21.15
 Mean   :2019   Mean   :  363857   Mean   :32.83
 3rd Qu.:2020   3rd Qu.:  223450   3rd Qu.:61.80
 Max.   :2022   Max.   :21243500   Max.   :89.40
                NA's   :43         NA's   :43
>
```

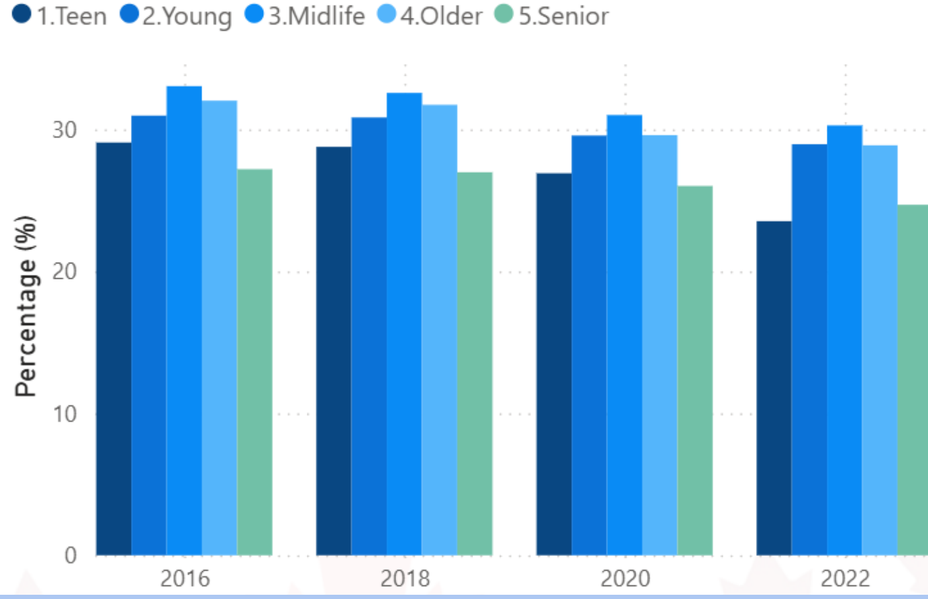
Mental Health Status 2016-2022



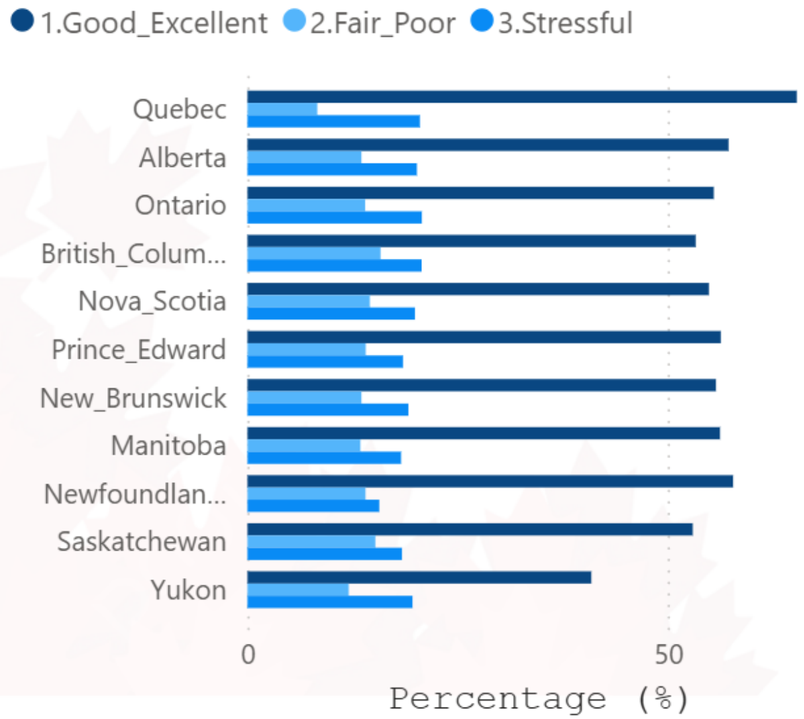
Mental Health By Gender 2016-2022



Mental Health by Age_group



Mental Health by Location in 2022

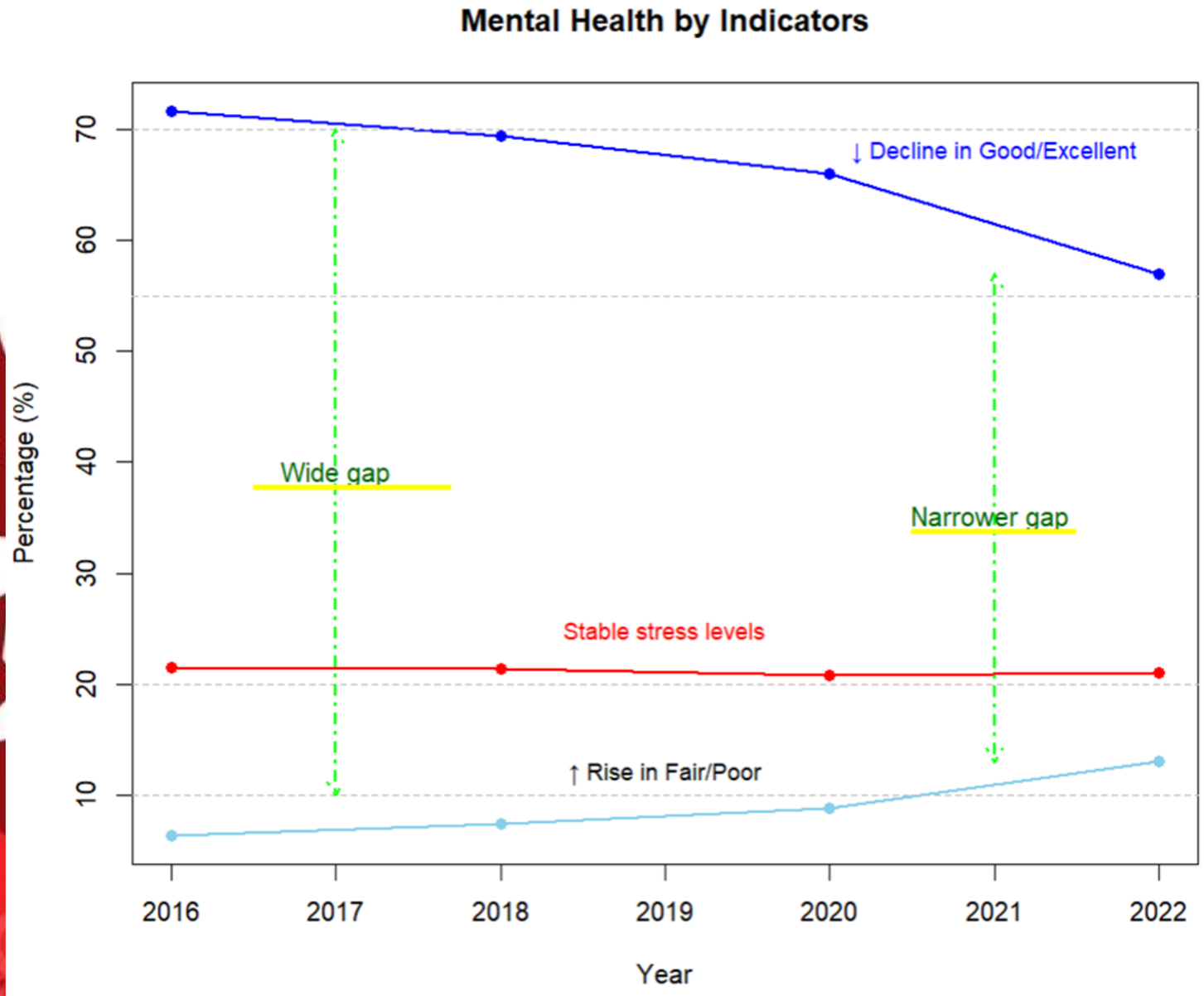


% by Location

- Quebec
- Alberta
- Canada
- Ontario
- British_Columbia
- Nova_Scotia
- Manitoba
- New_Brunswick

% by Gender

- Both_Genders
- Males
- Females

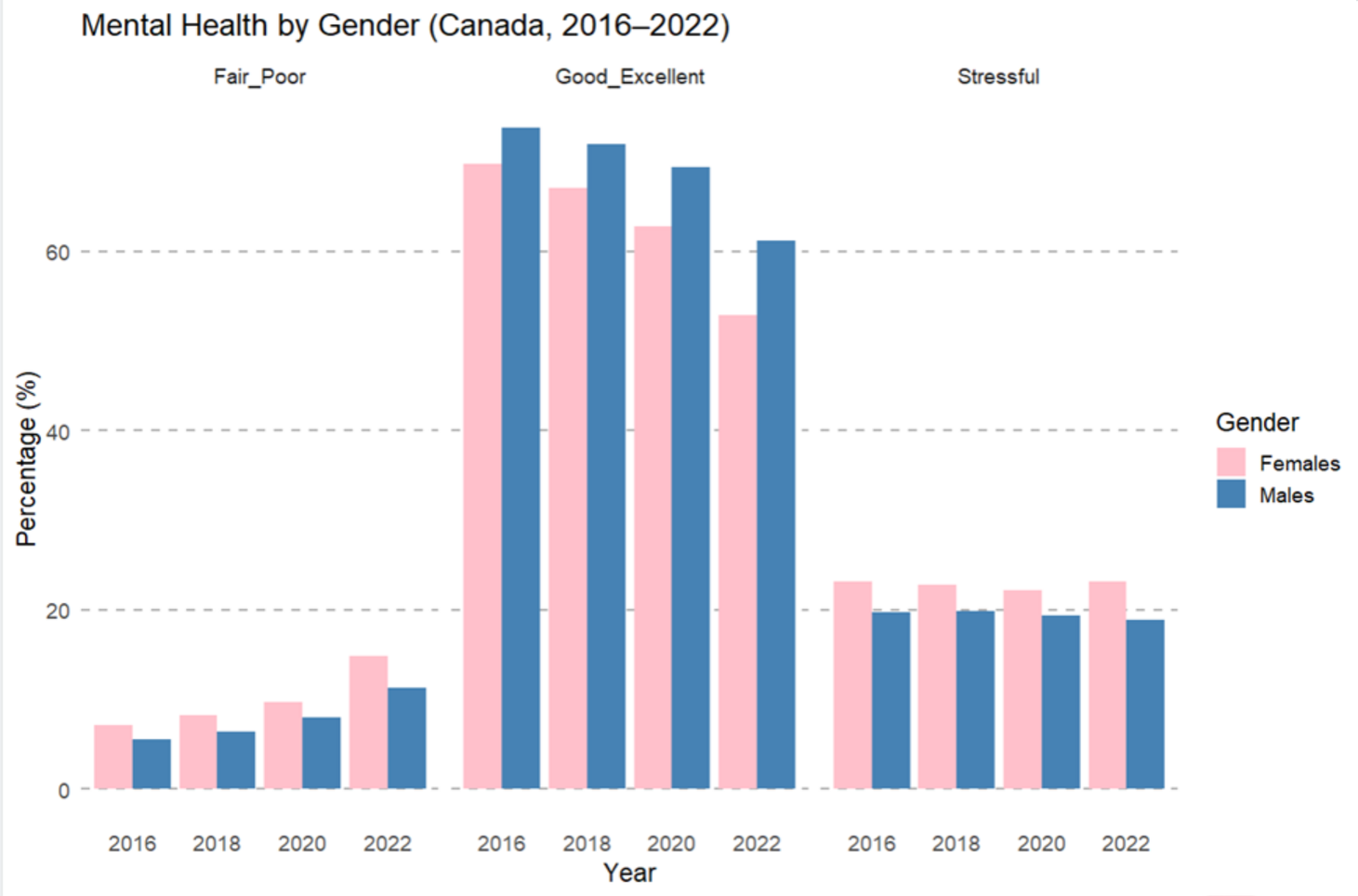


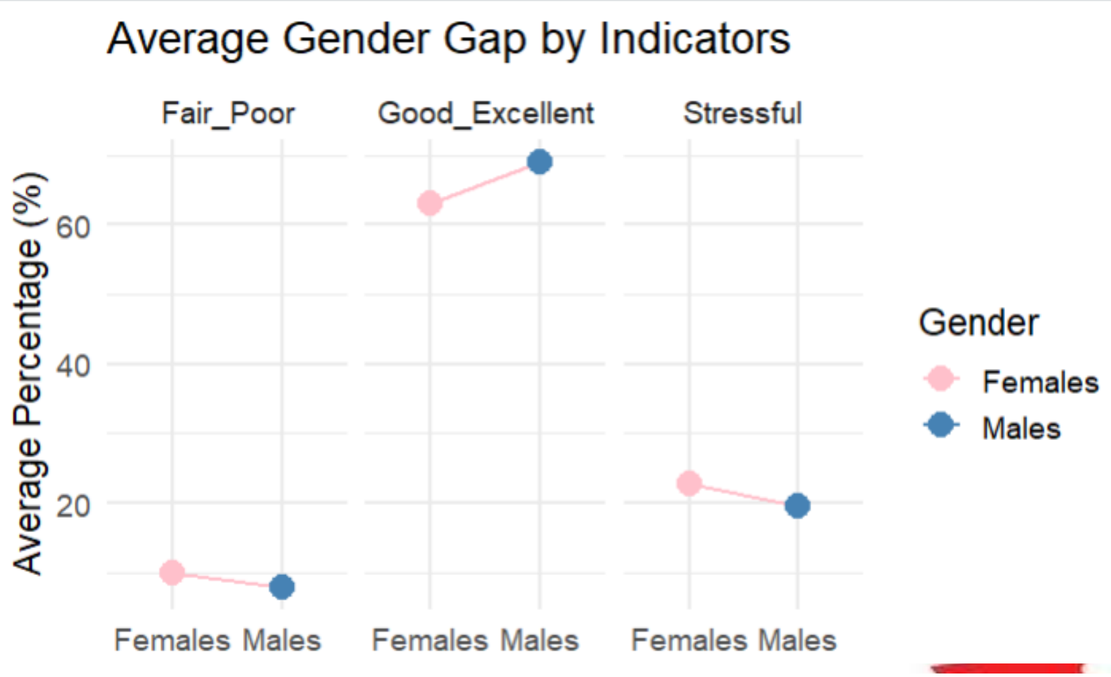
Mental Health Trends (2016–2022)

1. Good/Excellent dropped from ~70% to ~60%.
2. Fair/Poor ratings increased slightly.
3. Stress levels stayed stable around 20%.
4. Gap narrowed between positive & negative .

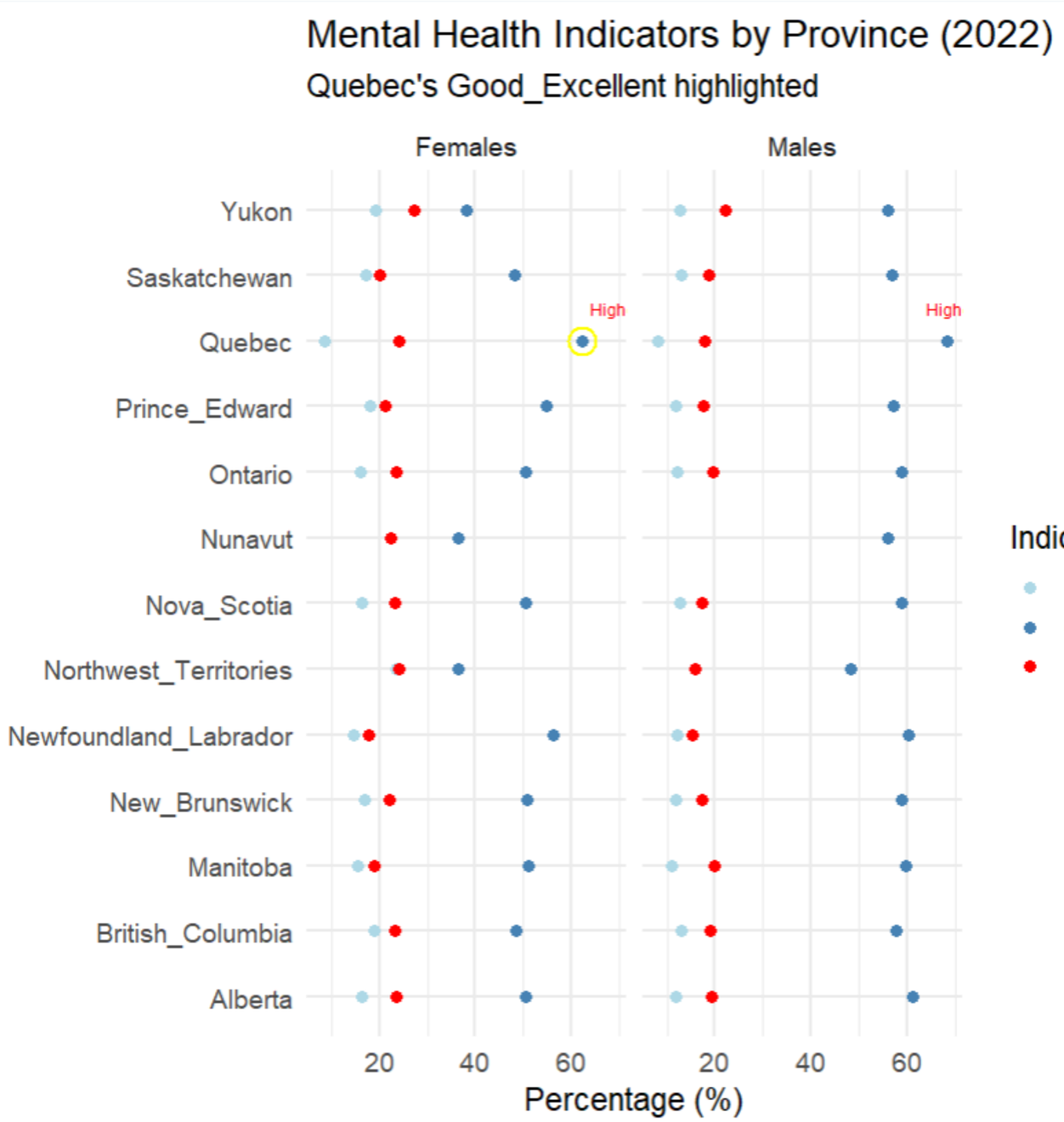
Visualization:

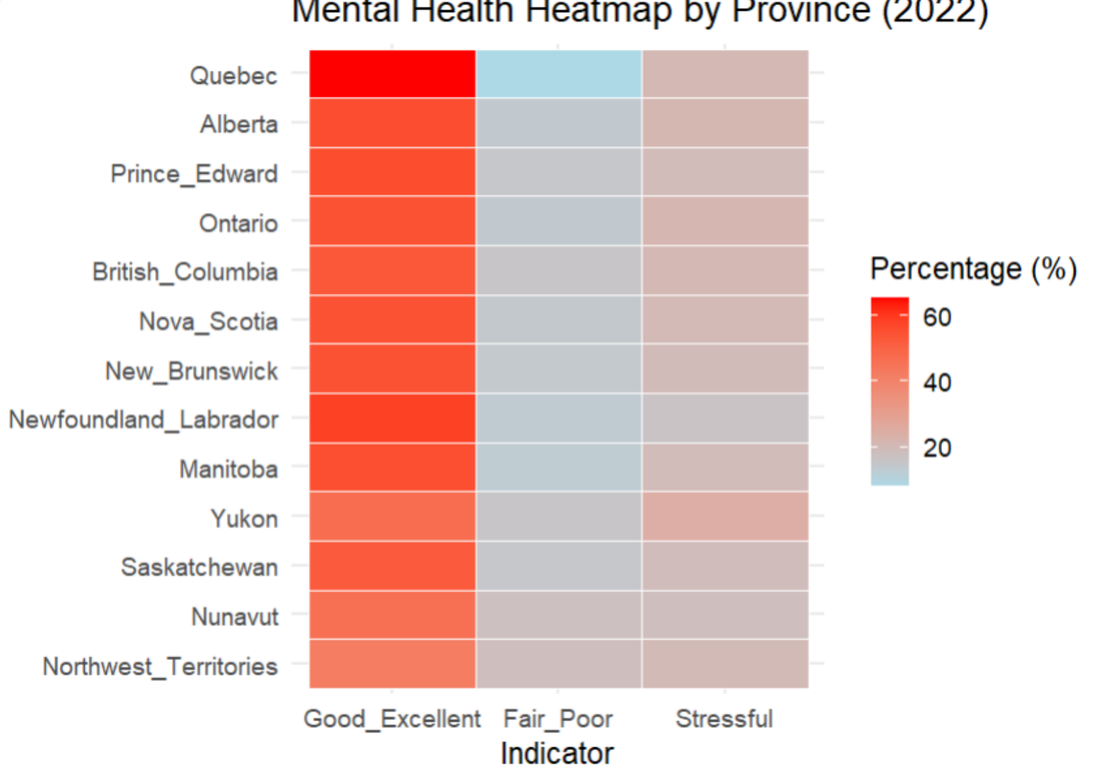
1. Plot functions: plot(), lines(), points()
2. Highlights: abline(), text(), arrows()





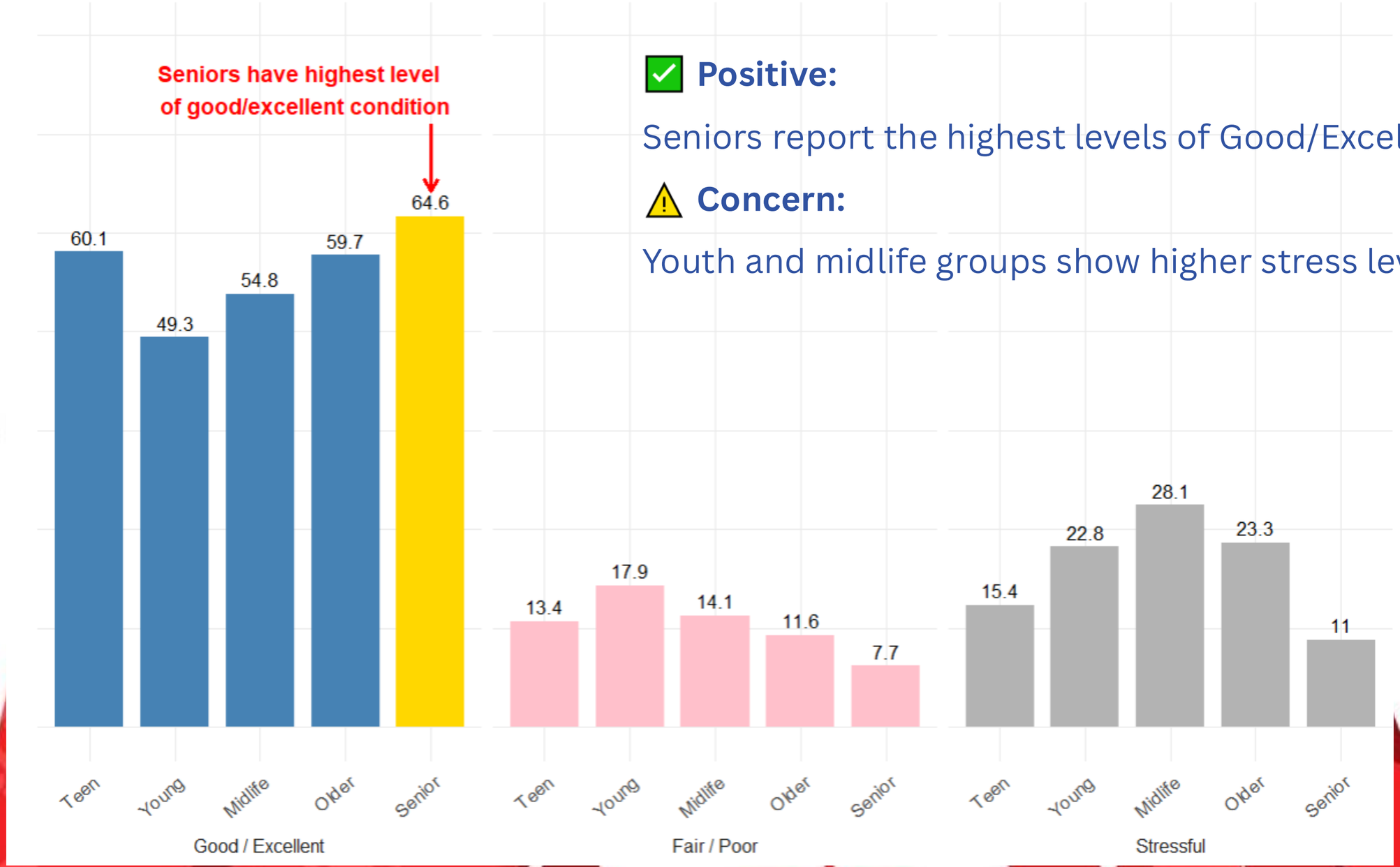
Visualization: ggplot2 (bar chart, line chart and boxplot)





Visualization: ggplot2 (dotmap, heatmap)

Mental Health by Age Groups in 2022



✅ Positive:

Seniors report the highest levels of Good/Excellent .

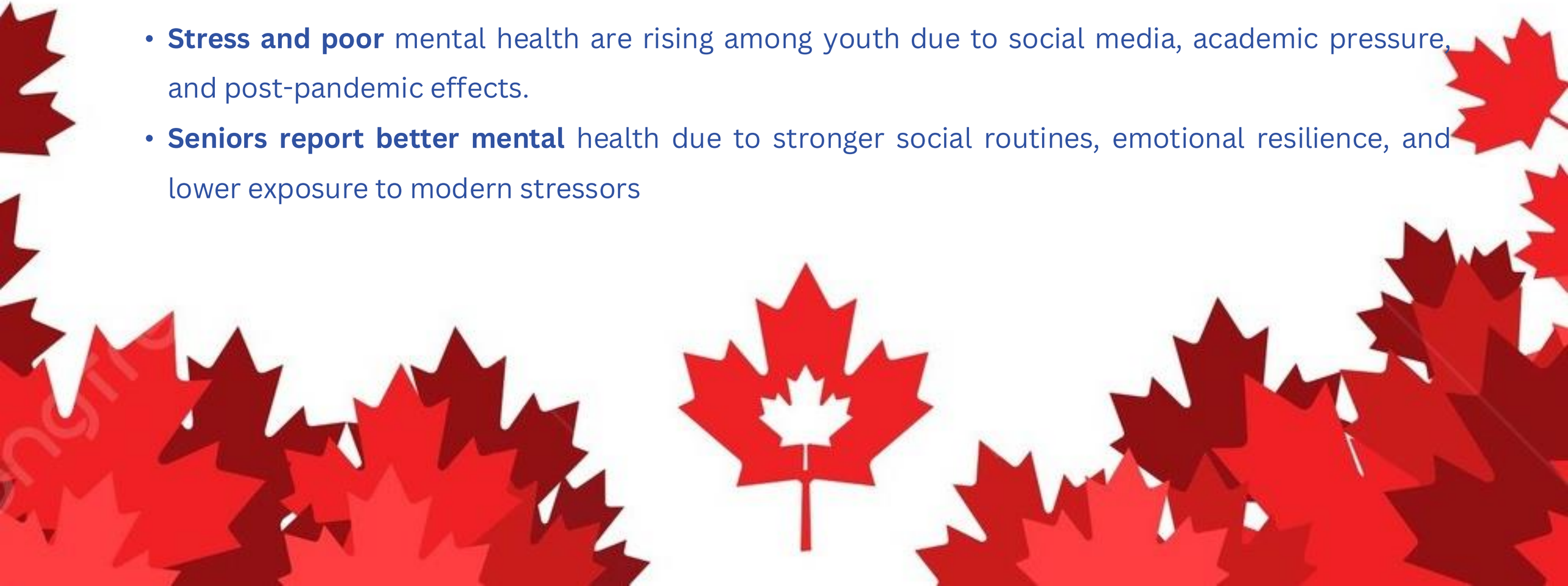
⚠ Concern:

Youth and midlife groups show higher stress levels.

CONCLUSION:

Mental health remains a critical public health priority, requiring support across all age groups.

- **Young people** (especially young women) are the most likely to meet criteria for mood or anxiety disorders.
- **Stress and poor** mental health are rising among youth due to social media, academic pressure, and post-pandemic effects.
- **Seniors report better mental** health due to stronger social routines, emotional resilience, and lower exposure to modern stressors

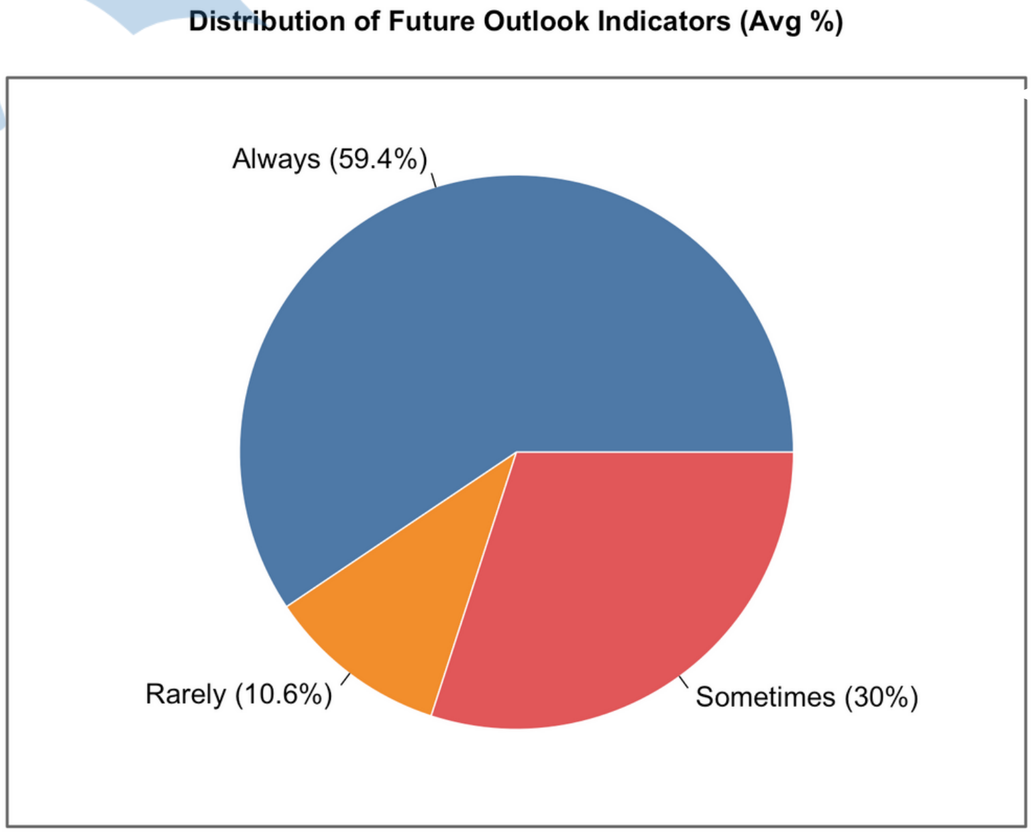


Future Outlook

By
Dmytro

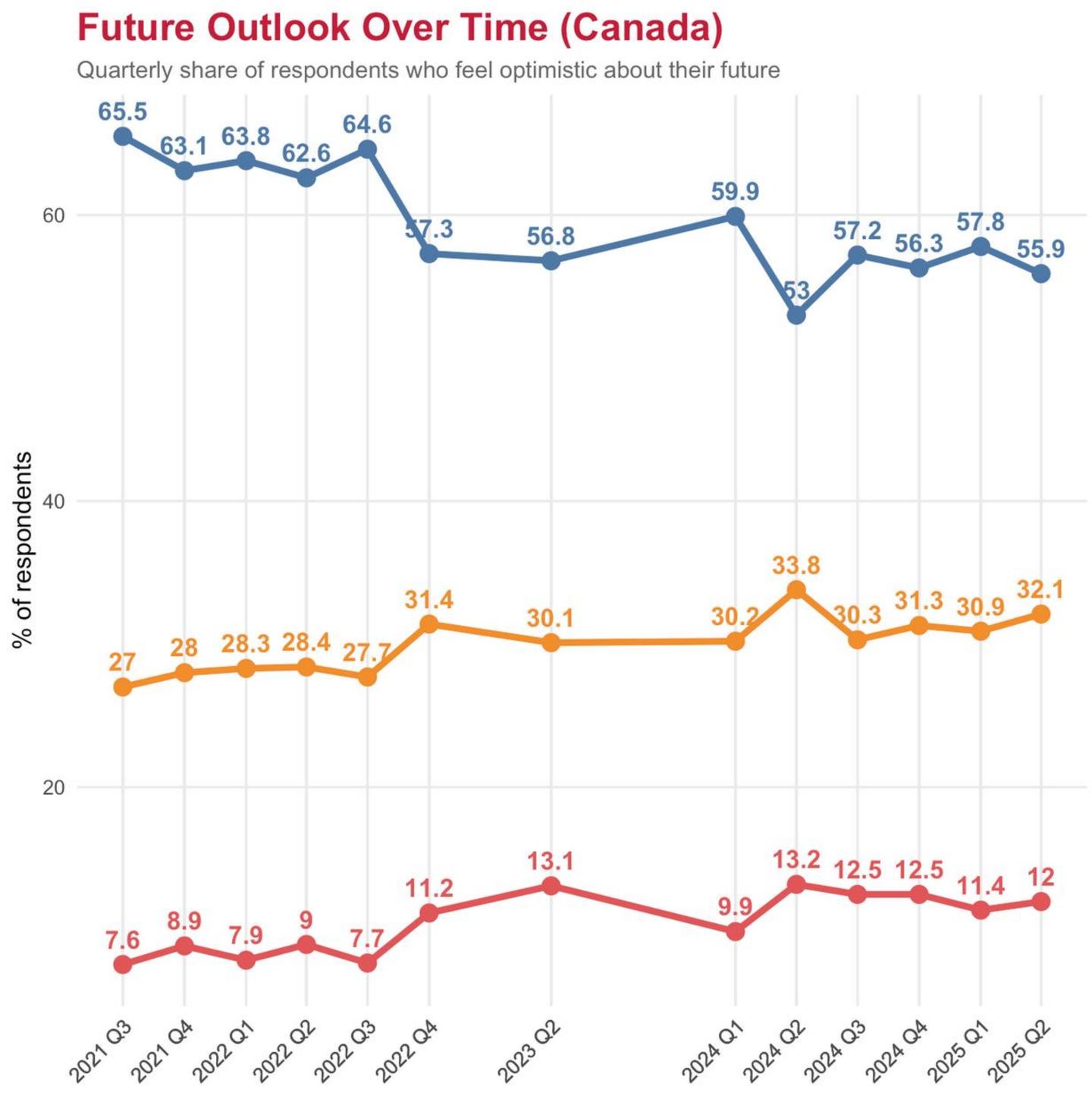
Future Outlook

Future outlook shows how optimistic Canadians feel about their life in the near future, grouped into three indicators: Always/Often, Sometimes, and Rarely/Never. Measured quarterly, this dataset explores future outlook levels across Canadian provinces and genders.




```
> print(summary_table)
# A tibble: 9 × 4
  Indicators Gender Avg_Number Avg_Percent
  <fct>      <fct>       <dbl>       <dbl>
1 Always     Total      19430.        59.5
2 Always     Men         9424.        58.2
3 Always     Women      10005.        60.8
4 Sometimes  Total       9811.        30
5 Sometimes  Men         4857.        29.9
6 Sometimes  Women       4954.        30
7 Rarely     Total       3458.        10.5
8 Rarely     Men         1946.        11.9
9 Rarely     Women       1512.         9.1
```

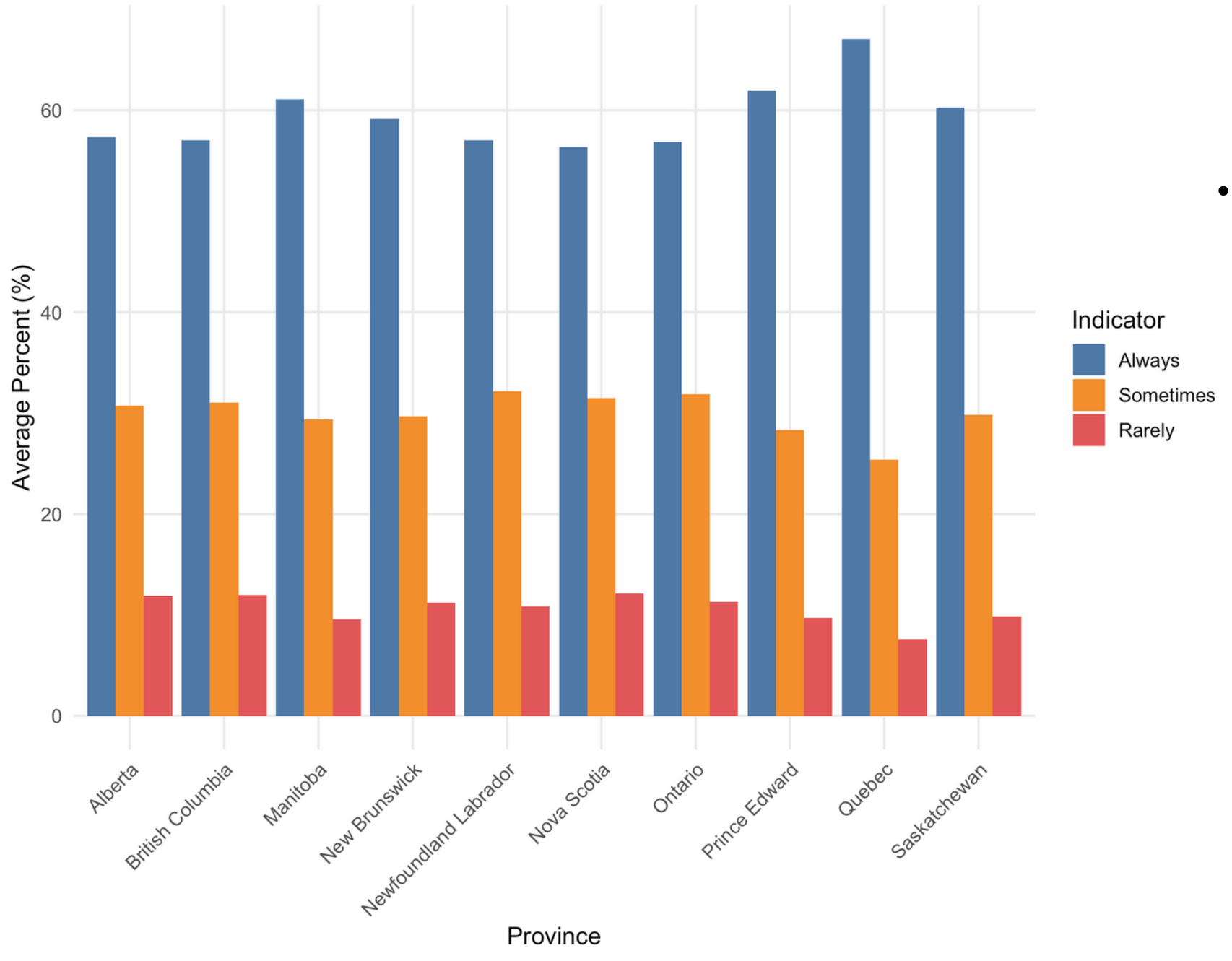
```
> print(summary_stats)
# A tibble: 3 × 6
  Indicators Min_Percent Max_Percent Mean_Percent Median_Percent SD_Percent
  <fct>           <dbl>       <dbl>        <dbl>          <dbl>      <dbl>
1 Always           51.1        66.3         59.5           58.9       4.02
2 Sometimes        26.7        34.2         30.0           30.2       1.98
3 Rarely            6.3        14.7         10.5           10.6       2.38
```



- **Strong** optimism has fallen by about 10 points.
- Both **moderate** and **low** optimism increased by 4-5 points over the past 4 years.

Average Future Outlook by Province

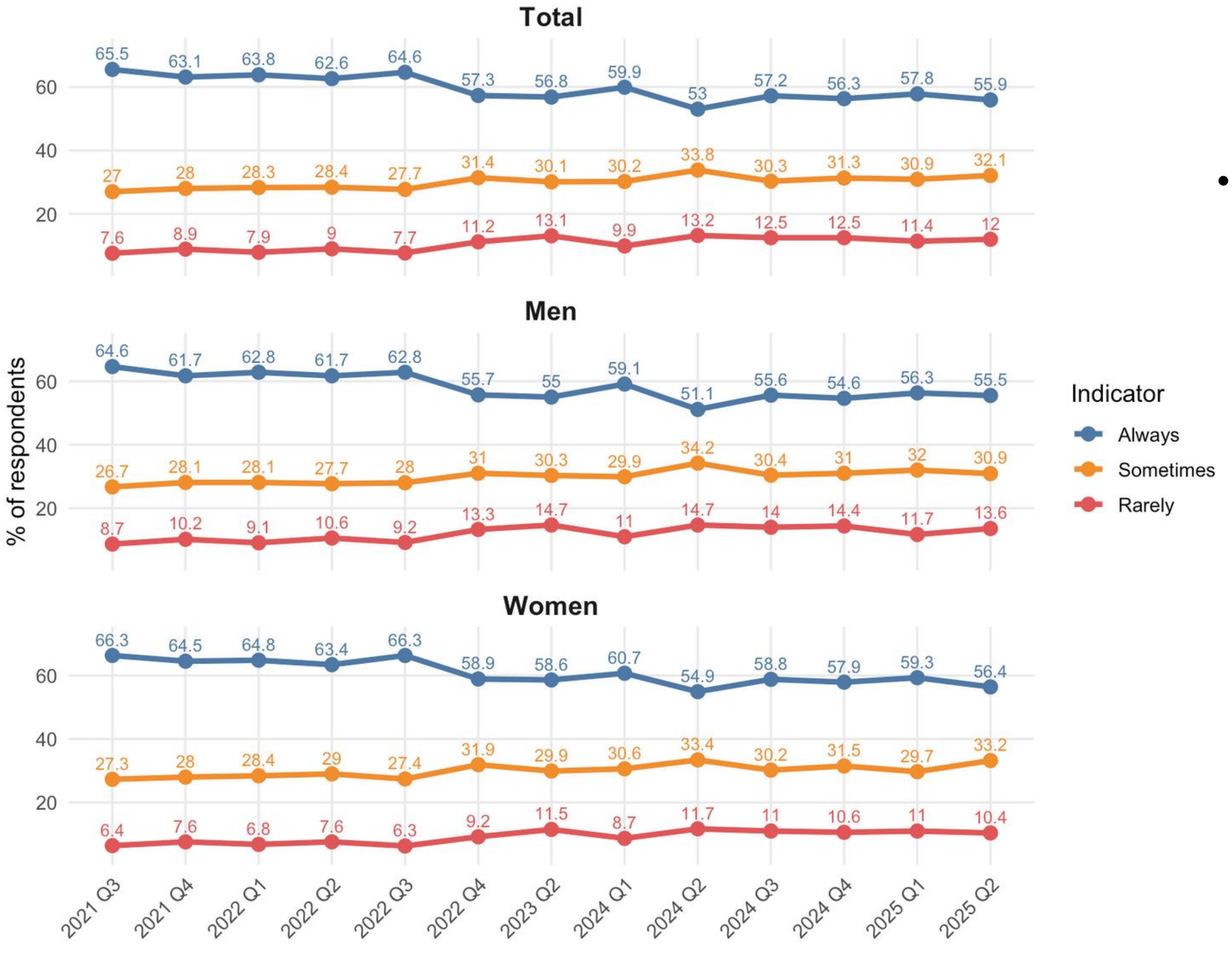
Average percentage of respondents who feel optimistic about their future



- **Quebec** shows the highest average strong optimism (~67%)
- **Ontario** has the lowest level of strong optimism (~57%)

Future Outlook Over Time by Gender (Canada)

Quarterly share of Canadians who feel optimistic, compared across gender groups



- **Women** consistently report slightly higher strong optimism than **men**.
- **Both** genders show a similar long-term decline of around 8–10 points.

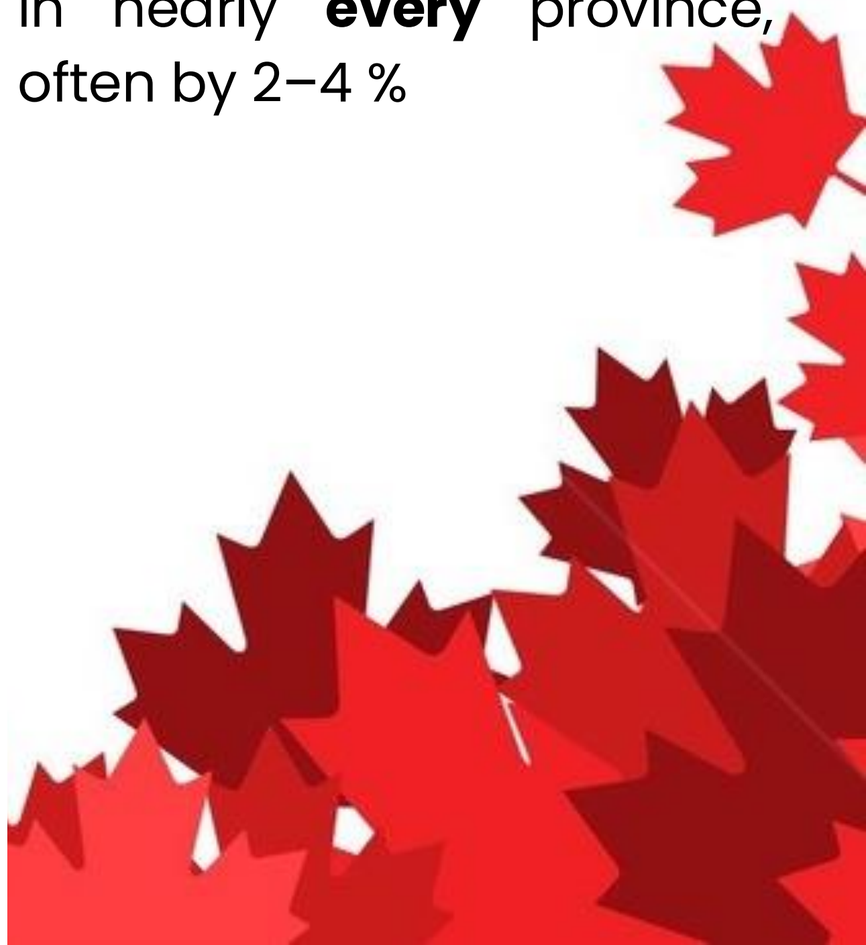
Average Future Outlook by Province and Gender

Heatmap comparing optimism patterns across demographic groups

	Always			Sometimes			Rarely		
	Total	Men	Women	Total	Men	Women	Total	Men	Women
Saskatchewan	60.3	57.5	63.1	29.8	31.4	28.2	9.9	11.2	8.6
Quebec	67.1	65.5	68.6	25.4	25.7	25	7.6	8.8	6.4
Prince Edward	61.9	59.9	63.9	28.3	28.9	27.9	9.7	11.2	8.3
Ontario	56.9	56.2	57.6	31.9	31.4	32.3	11.3	12.4	10.1
NovaScotia	56.4	54.4	58.2	31.5	31.7	31.4	12.1	13.9	10.4
Newfoundland Labrador	57	55.8	58.2	32.2	31.5	32.8	10.8	12.7	9
New Brunswick	59.1	57.3	60.9	29.6	29.9	29.5	11.2	12.9	9.6
Manitoba	61.1	58.4	63.8	29.4	30.3	28.5	9.6	11.4	7.7
British Columbia	57	55.4	58.6	31	31.2	30.8	12	13.5	10.5
Alberta	57.4	55.6	59.1	30.7	30.1	31.3	11.9	14.2	9.6

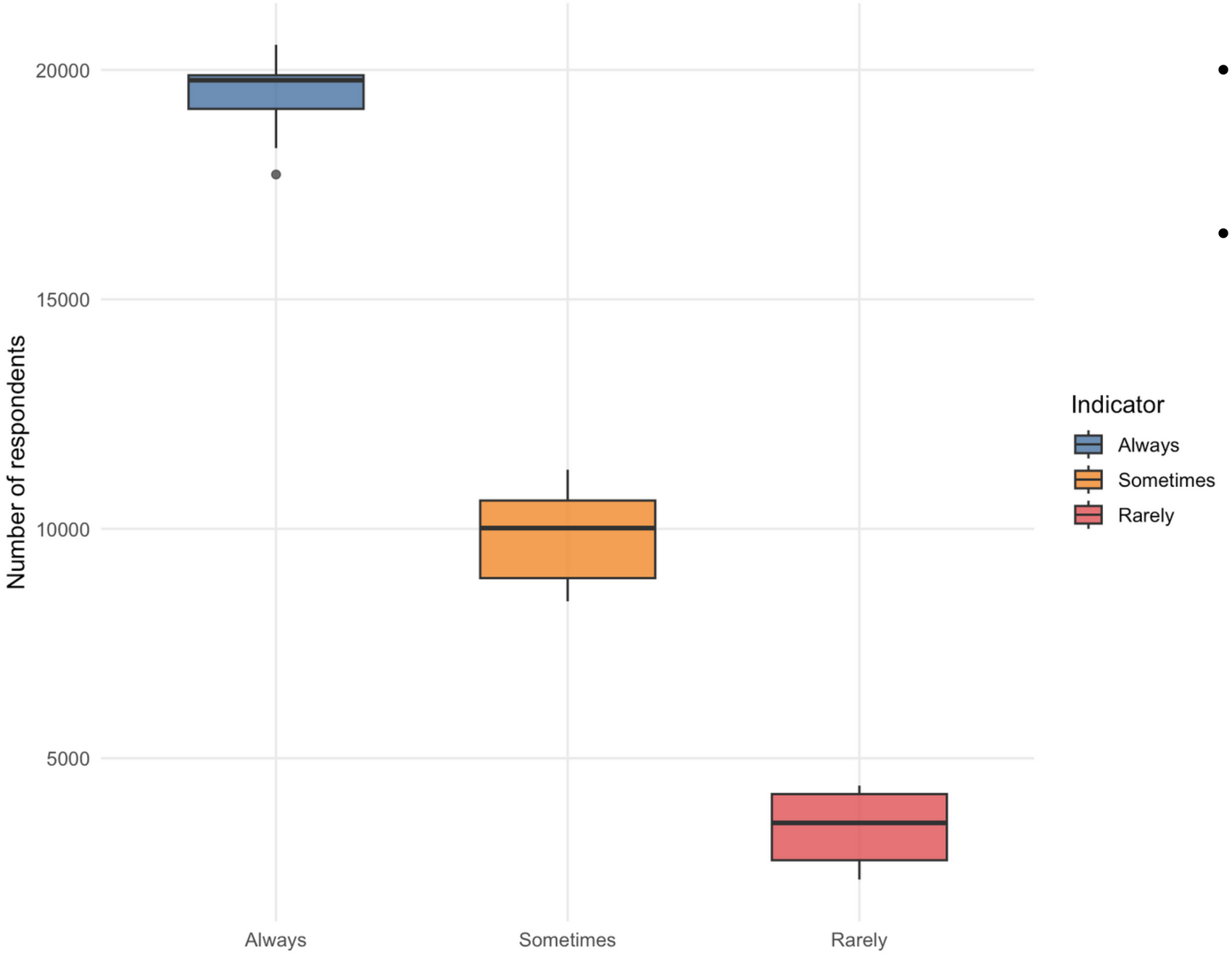
Indicator
- Always
- Sometimes
- Rarely

- **Quebec** is the most optimistic province overall across men and women
- Women report higher strong optimism than men in nearly **every** province, often by 2–4 %



Respondent Count Variation by Indicator

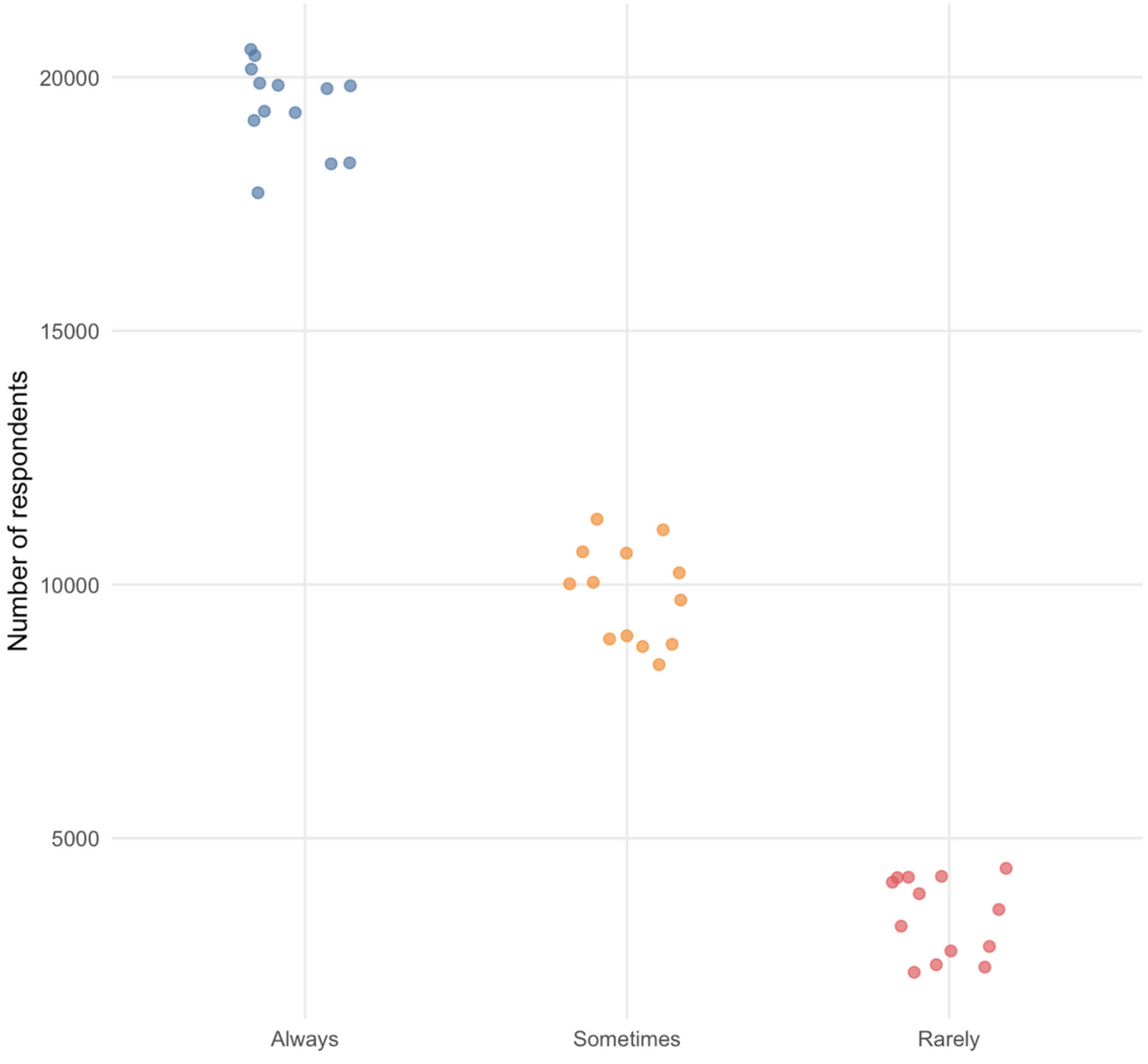
Boxplot showing median, quartiles, and spread of optimism response counts in Canada



- **Always** median ~19.5k
- **Sometimes** median ~10k with more variability.
- **Rarely** median ~3.5k with the widest relative spread

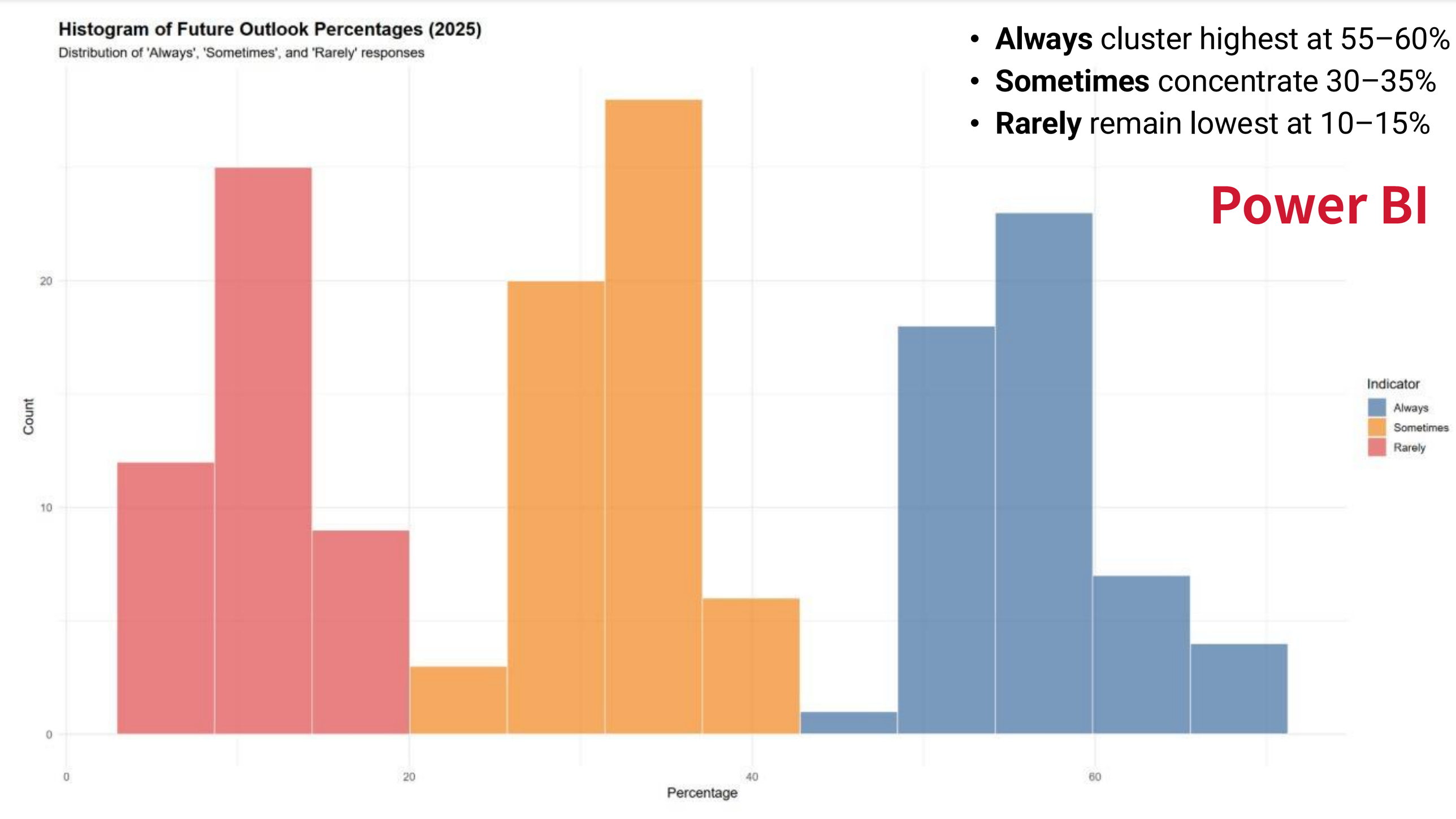
Respondent Count Distribution by Indicator

Swarm plot showing quarterly total respondent counts for optimism levels in Canada



- **Respondent counts** cluster clearly by indicator, showing a consistent pattern across quarters

- **Always** responses highest (≈18–20k)
- **Sometimes** in the mid range (≈9–11k)
- **Rarely** lowest (≈3–4k)



- **Always** cluster highest at 55–60%
- **Sometimes** concentrate 30–35%
- **Rarely** remain lowest at 10–15%

Power BI

Future Outlook – Summary

Key national insights on optimism, gender differences, geography, and trends

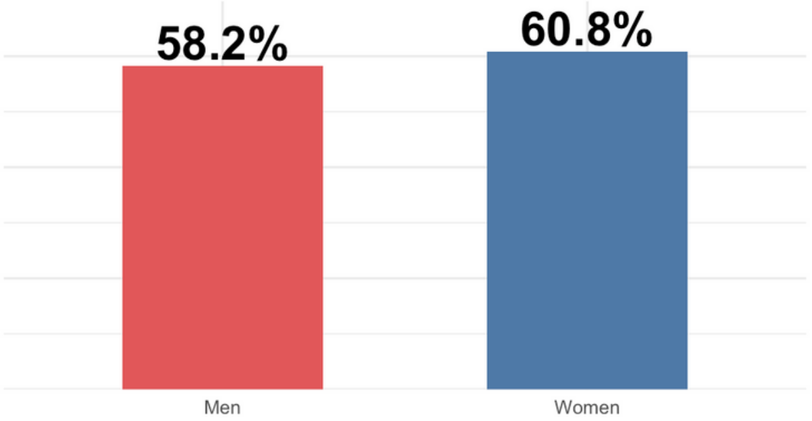
Canada's Average Optimism

Most Canadians feel optimistic about their future.

59.5%

Women More Optimistic

Women report slightly higher optimism than men.



Low Level of Pessimism

Only a small share of Canadians feel pessimistic.

10.5%

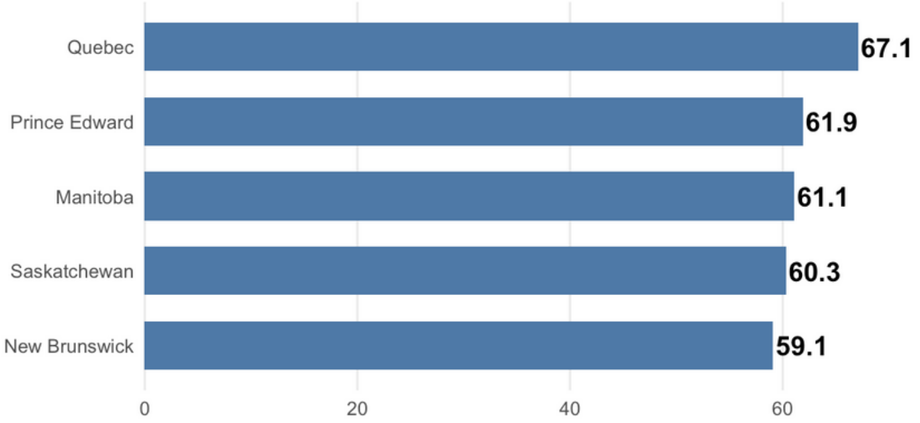
Moderate Optimism Stable

Around 30% consistently feel moderately optimistic.

30%

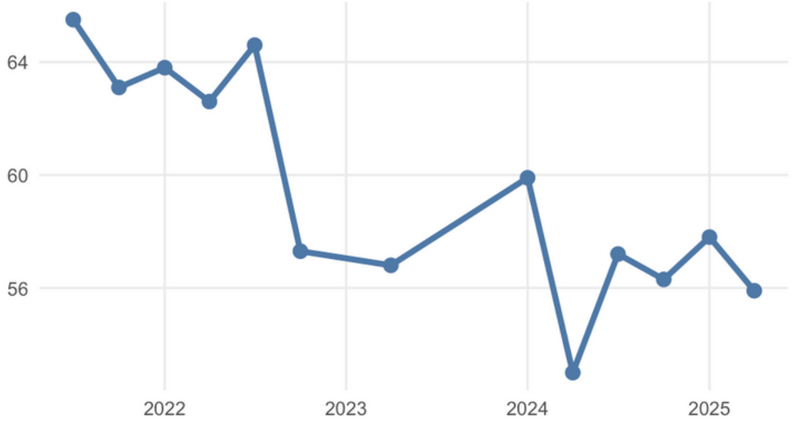
Top Provinces by Optimism

Quebec leads Canada in optimism.



Optimism Over Time

Optimism fluctuates but remains high overall.



Sense of Meaning and Purpose

By Yu-Shen

Sense of Meaning and Purpose

Definition:

Proportion of the population by level of feeling as if the things they do in life are worthwhile.

Measurement:

Using a scale of 0 to 10, where 0 means "Not at all" and 10 means "Completely", to what extent do you feel the things you do in your life are worthwhile?

Dataset Sample:

Geography	Gender	Indicators	Reference period	
				Percent
Canada (excluding territories) (map)	Total, all persons[7, 8]	Sense of meaning and purpose rating between 0 and 5[4]	Q2 2025	17.2
		Sense of meaning and purpose rating of 6 or 7[4]	Q2 2025	27.0
		Sense of meaning and purpose rating of 8, 9 or 10[4]	Q2 2025	55.7
	Men	Sense of meaning and purpose rating between 0 and 5[4]	Q2 2025	18.5
		Sense of meaning and purpose rating of 6 or 7[4]	Q2 2025	26.8
		Sense of meaning		

Summary – High sense of MP by Geography in CA

We classified the Sense of Meaning and Purpose ratings into three levels:

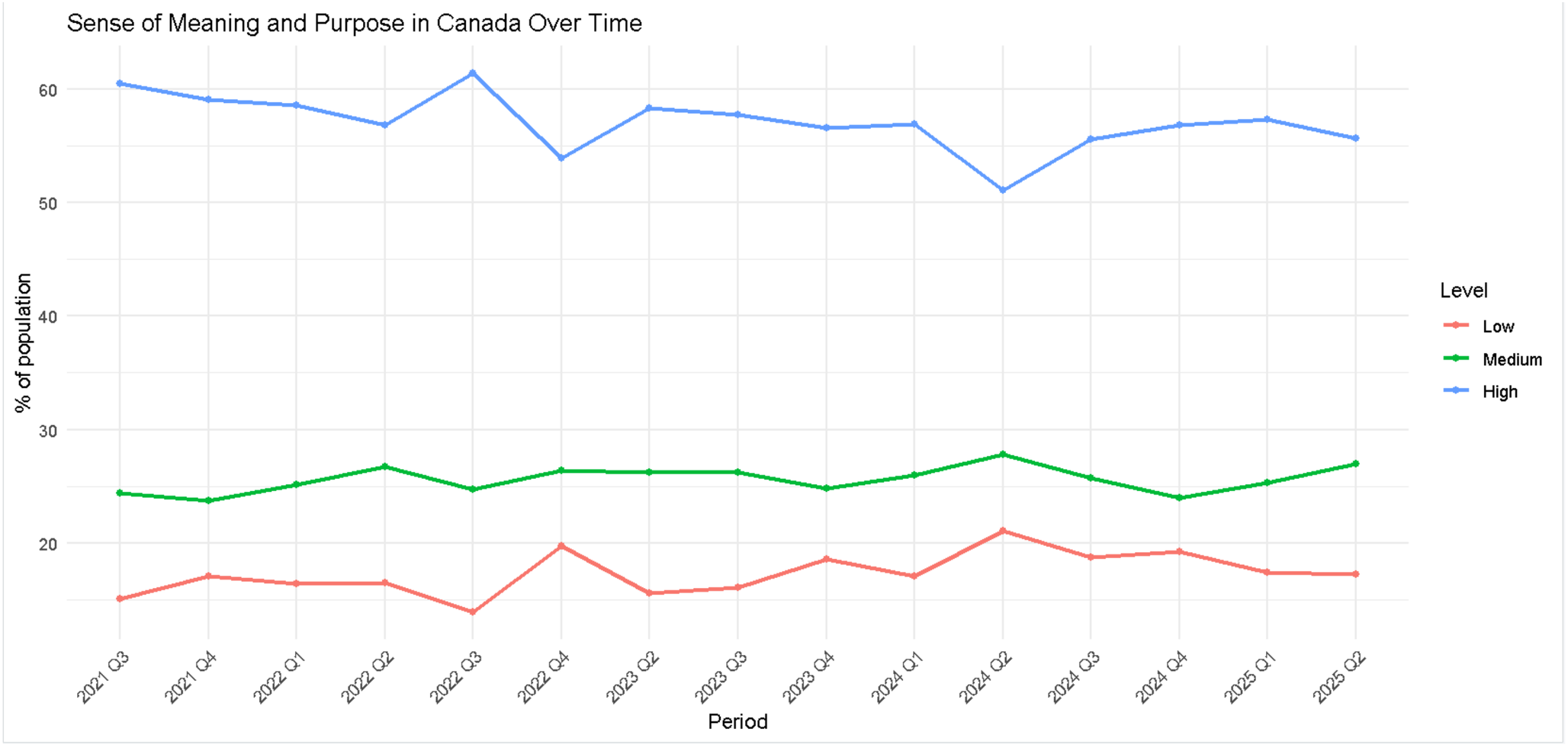
1. Rating between 0 and 5: **Low**
2. Rating of 6 or 7: **Medium**
3. Rating between 8 to 10: **High**

The form is percentage of people having high rating of MP in various regions, and its sample size.

Release date: 2025-09-10
Geography: Canada, Geographical region of Canada, Province or territory

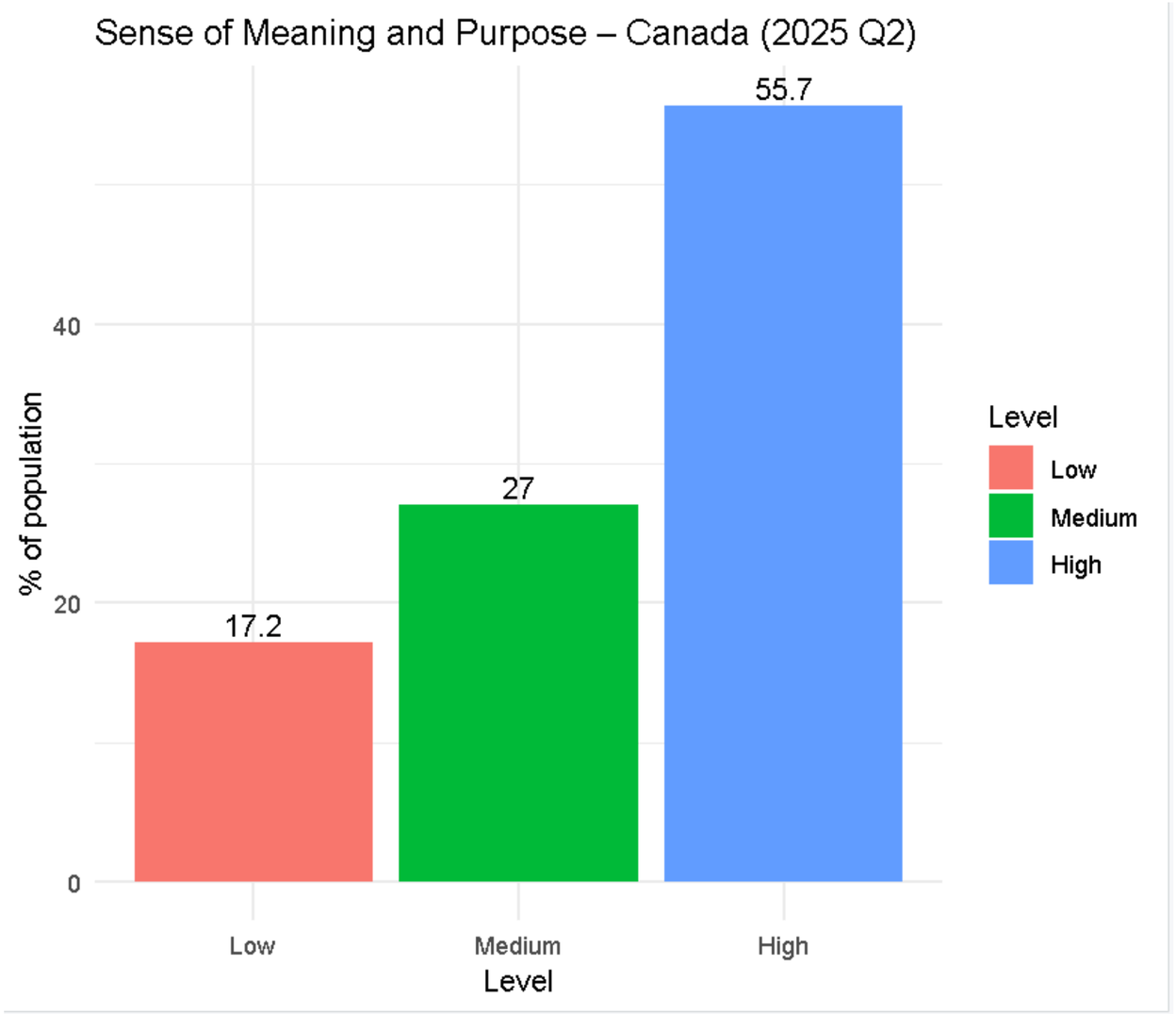
As of - 2025 Q2	High Sence of MP	Sample
Prince_Edward	64.1	98
Quebec	63.7	4832
New_Brunswick	63	457
Nova_Scotia	57.6	529
Manitoba	57.1	671
Saskatchewan	57	549
Newfoundland_Labrador	56.8	265
Canada	55.7	19275
British_Columbia	55.6	2688
Ontario	52.3	7173
Alberta	49.7	2013

Sense of Meaning and Purpose in Canada – trend over time



Sense of Meaning and Purpose in Canada of 2025 Q2

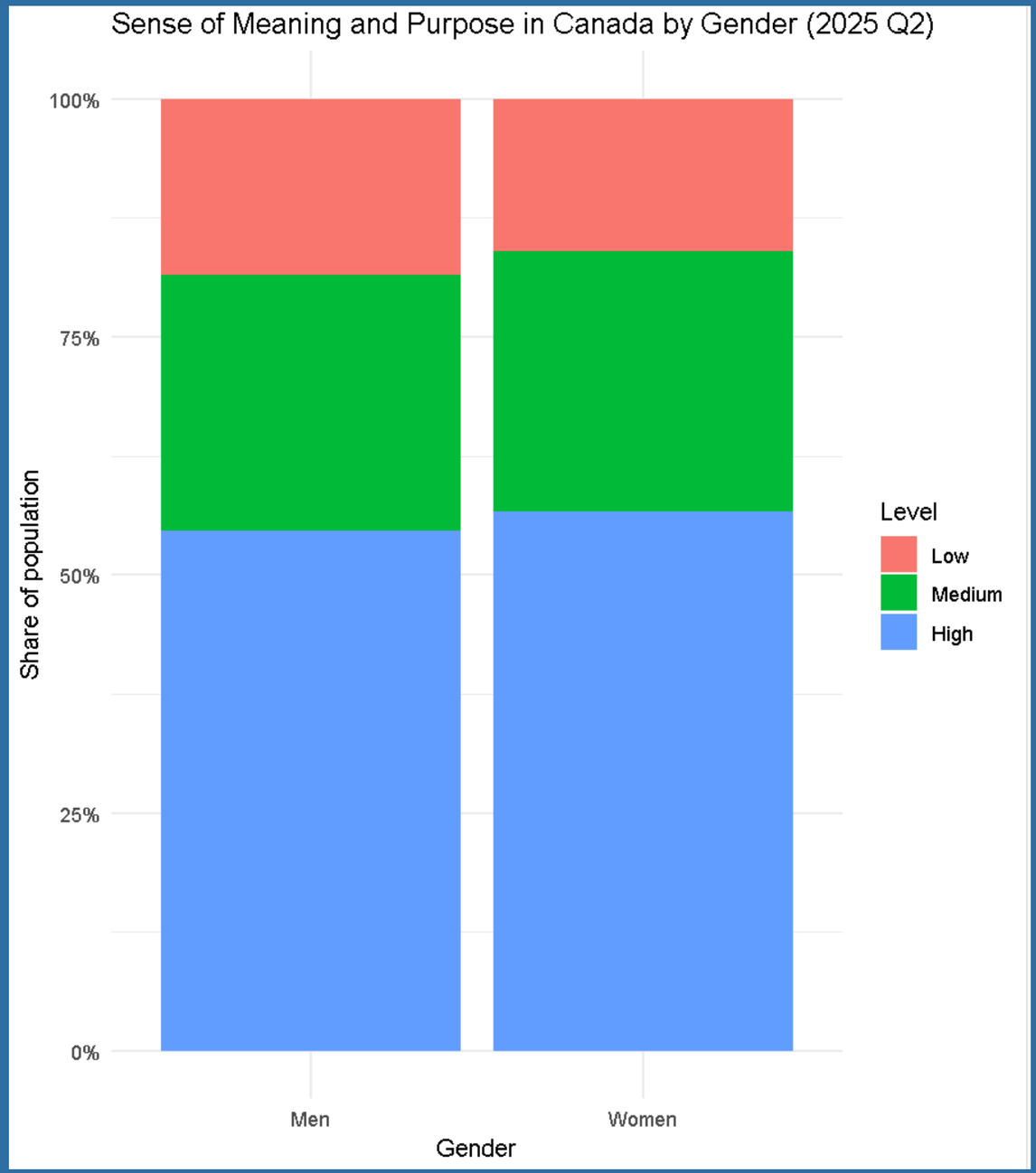
In the most recent period, most Canadians fall into the **High** meaning and purpose category, with a smaller group in Low. This suggests generally strong perceived purpose nationally, but a non-negligible group at risk of low purpose.



Sense of Meaning and Purpose in Canada by Gender

Overall, Canadian women tend to have higher sense of meaning and purpose compare to men, but medium proporsion seems to be about the same.

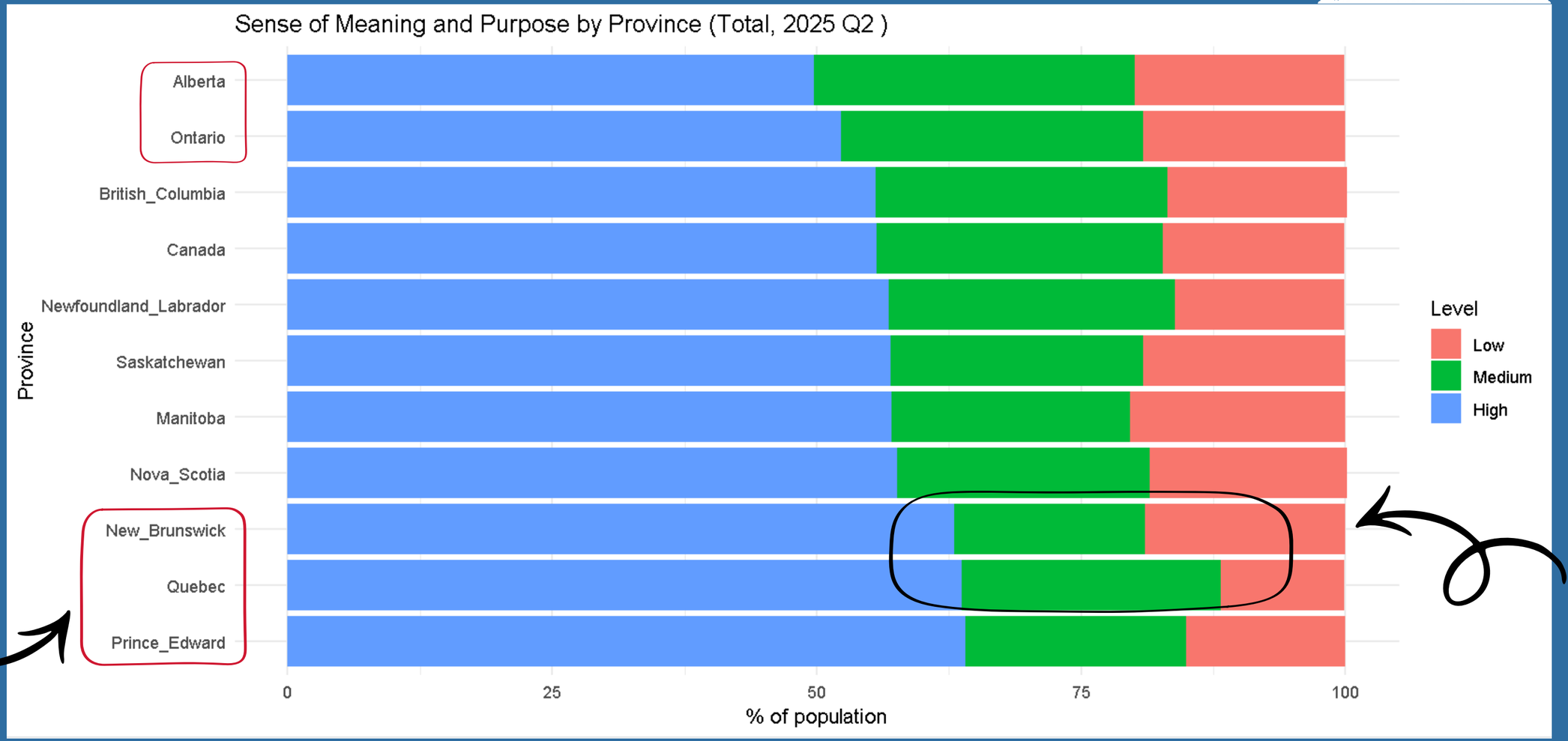
– (In the latest period)



Sense of Meaning and Purpose by Province

65% people reported high level MP at **Prince Edward Island**

20% people reported high level MP at **Monitaba**



Sense of Meaning and Purpose by Provinces &Gender

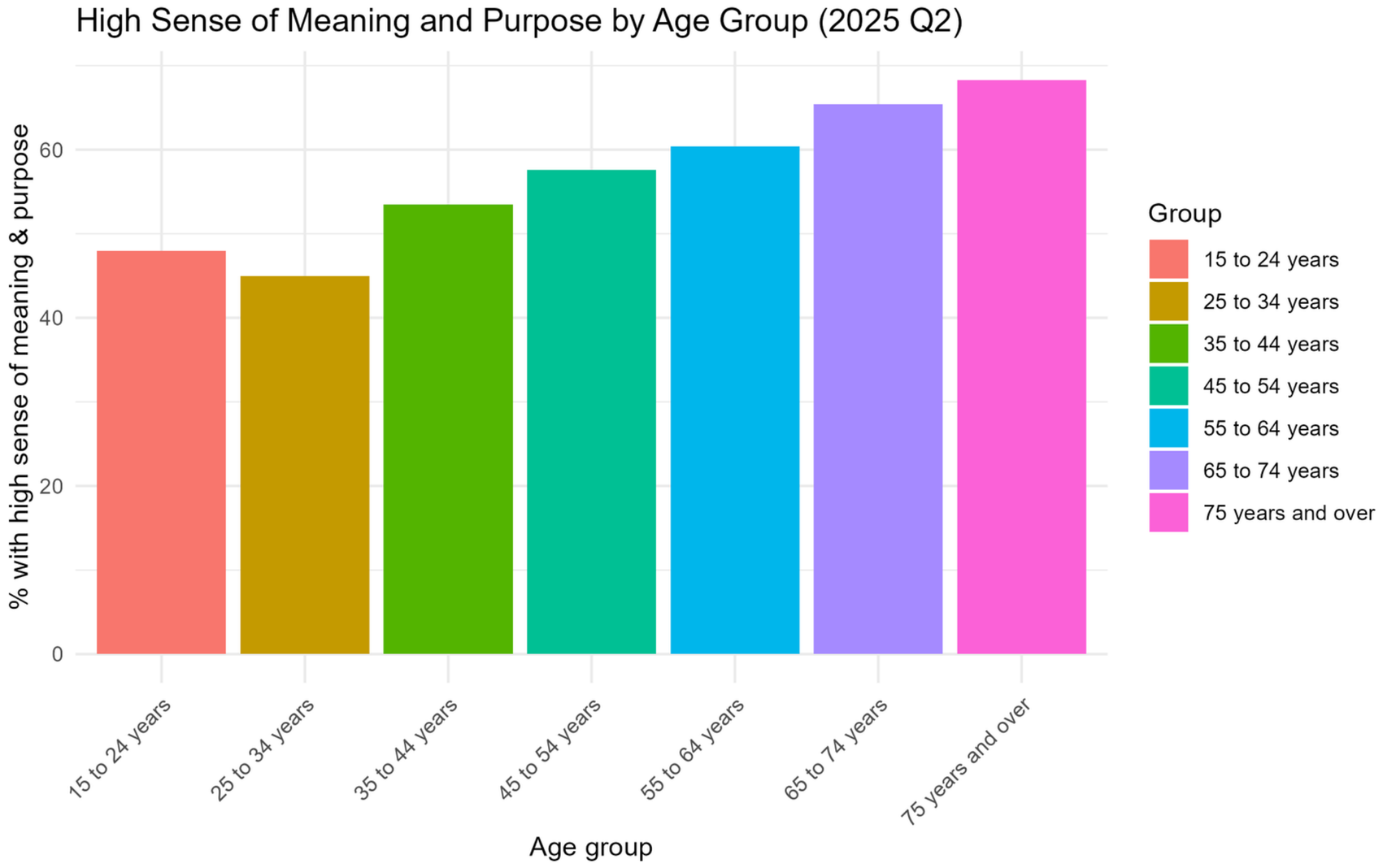
65% Men are in high level MP at **Quebec**

66% Women are in high level MP at **Quebec**

25% Men are in low level MP at **Manitoba**

20% Women are in low level MP at **Alberta**

Sense of Meaning and Purpose in Provinces by Age



Older age groups (especially 65–74 and 75+) tend to report the highest levels of meaning and purpose.

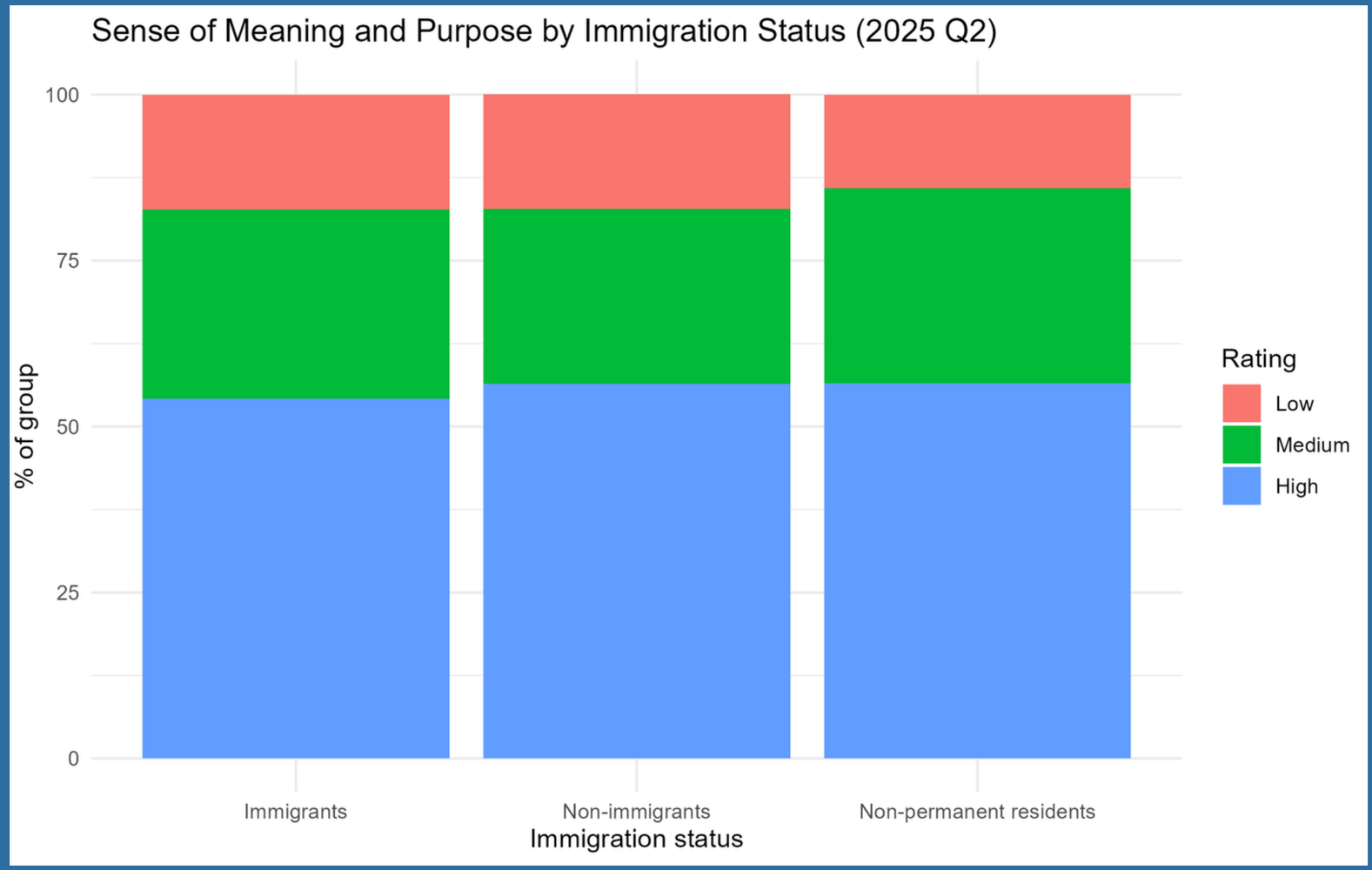
While younger adults (e.g., 15–24, 25–34) show lower proportions in the high category.

This suggests a possible increase in perceived purpose with age and life experience.

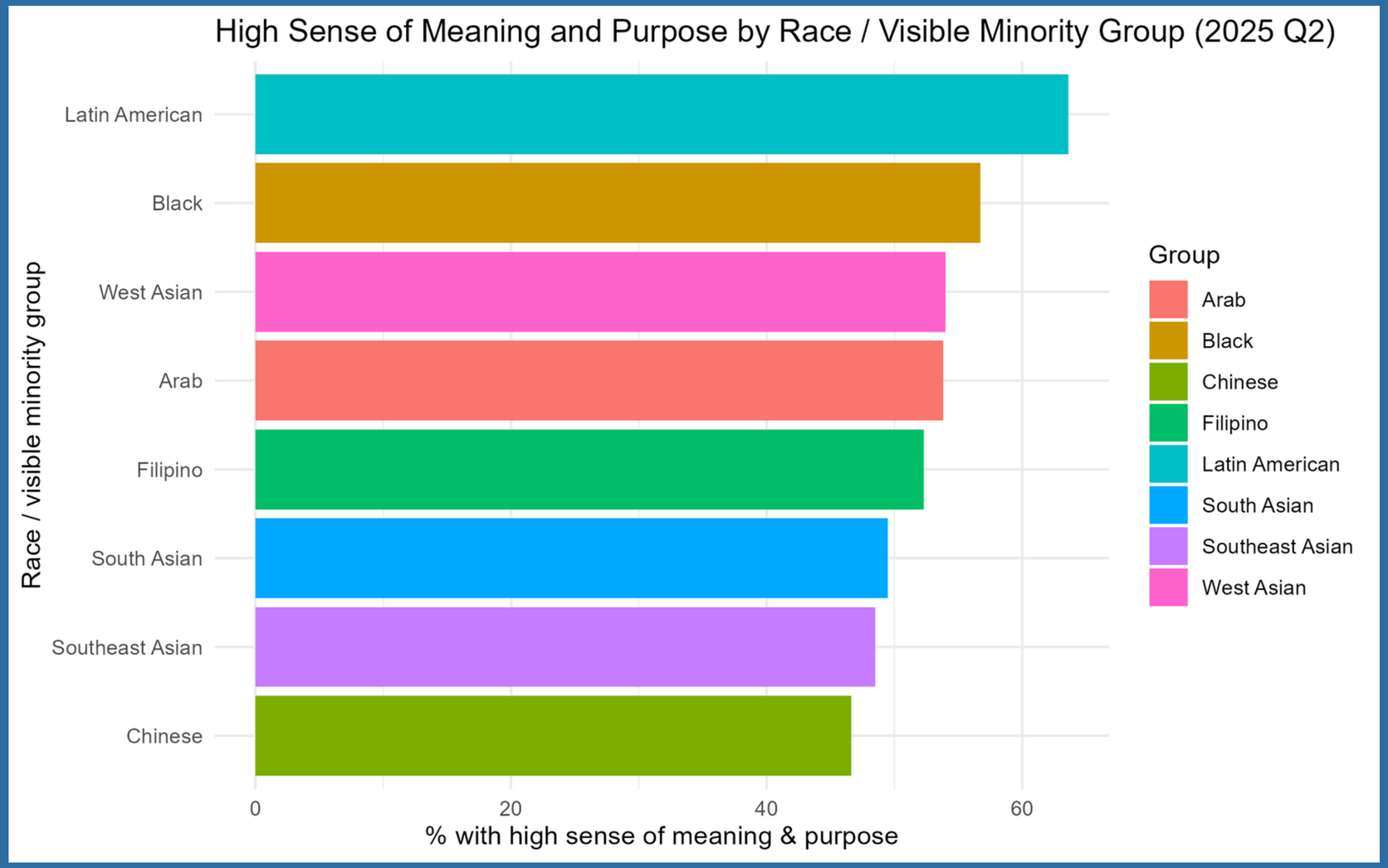
Sense of Meaning and Purpose in Canada by Immagration Status

Non-immigrants typically have the largest share in the High category, while immigrants residents show slightly higher proportions in the Low and Medium levels.

Non-permanent residents show similar or even better traits than Non-immagrants.



Sense of Meaning and Purpose in Canada by Visible Minority



There are meaningful differences across visible minority groups.

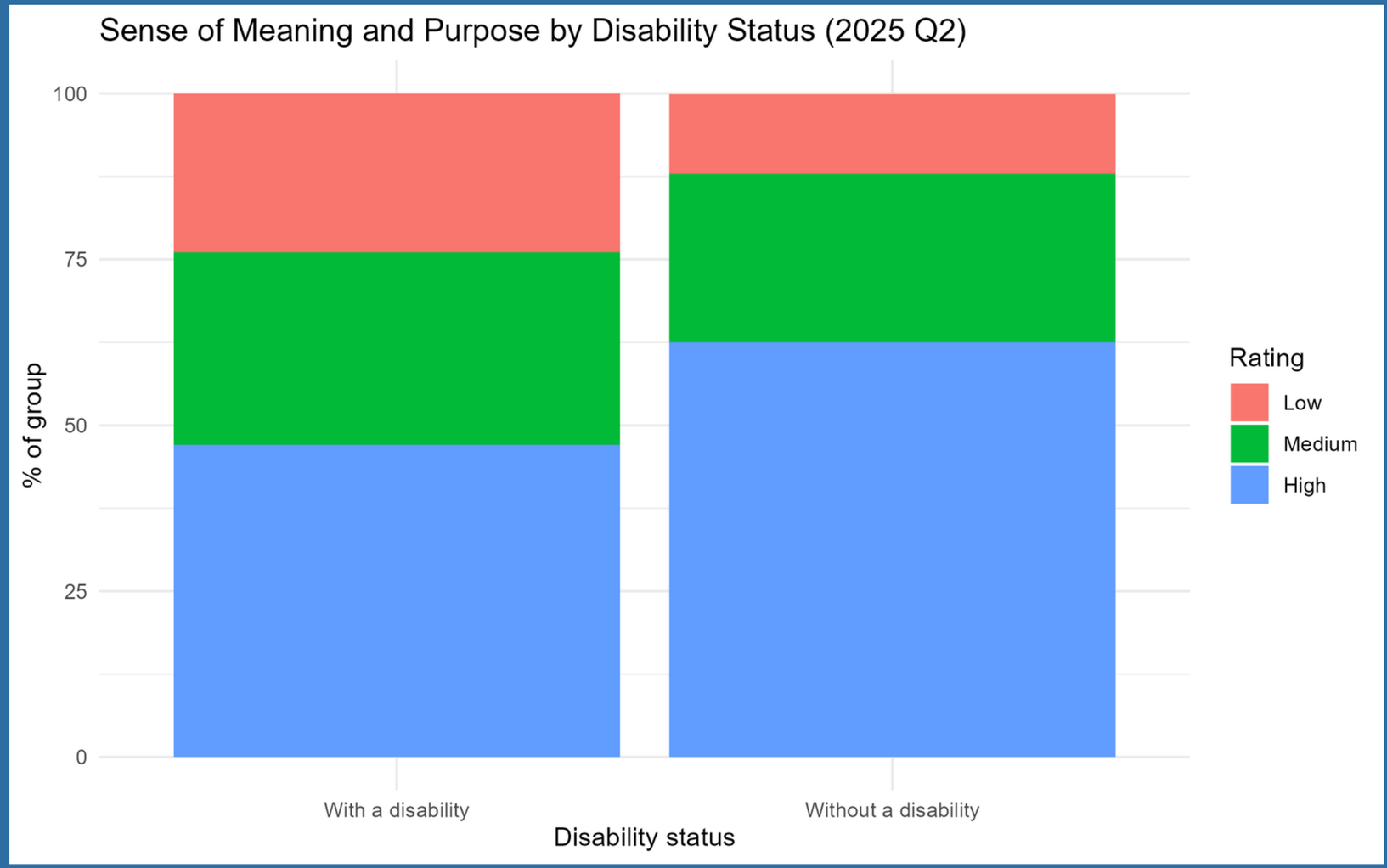
Latin Americans report very high levels of meaning and purposethan other racial communities.

Among all, Chinese live in Canada shows lower Sense of MP.

Sense of Meaning and Purpose in Canada by Disability Status

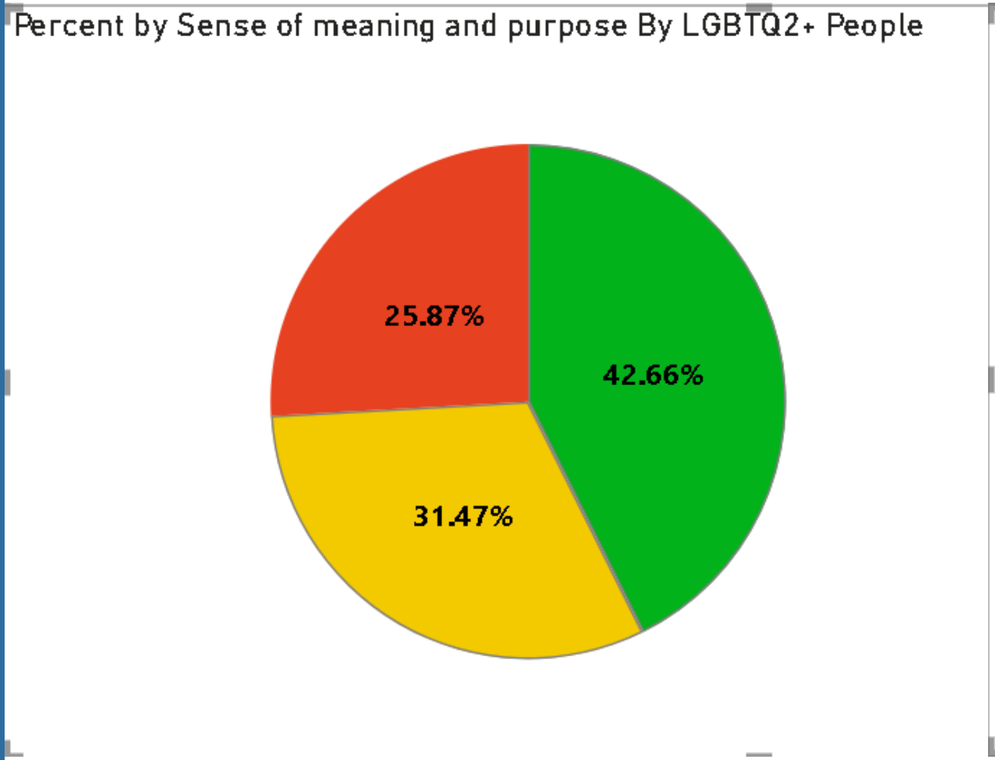
People without a disability are more likely to fall into the High category, whereas those with a disability show higher shares in Low and Medium.

This indicates an important wellbeing gap associated with disability.

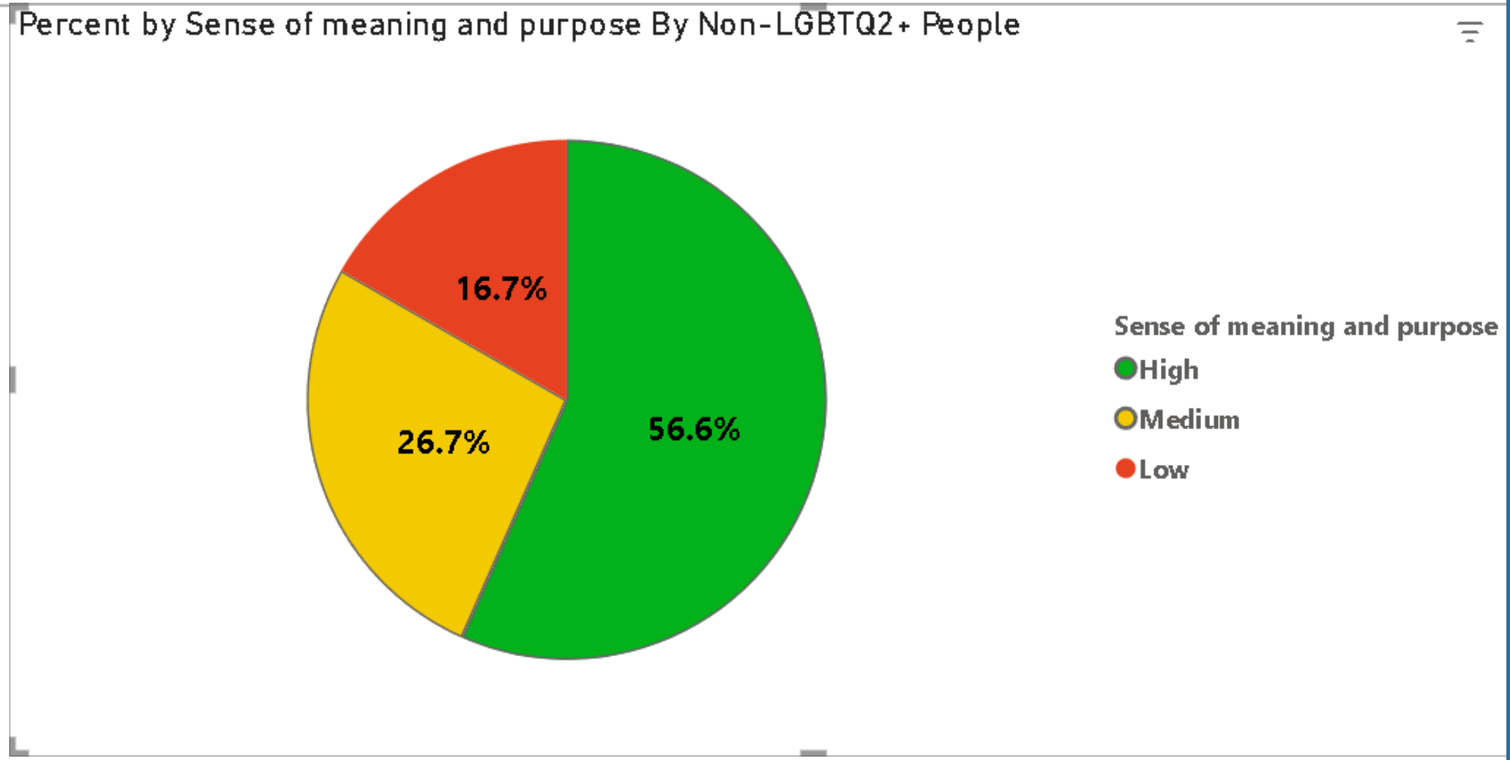


Sense of Meaning and Purpose by LGBTQ2+

LGBTQ2+ people

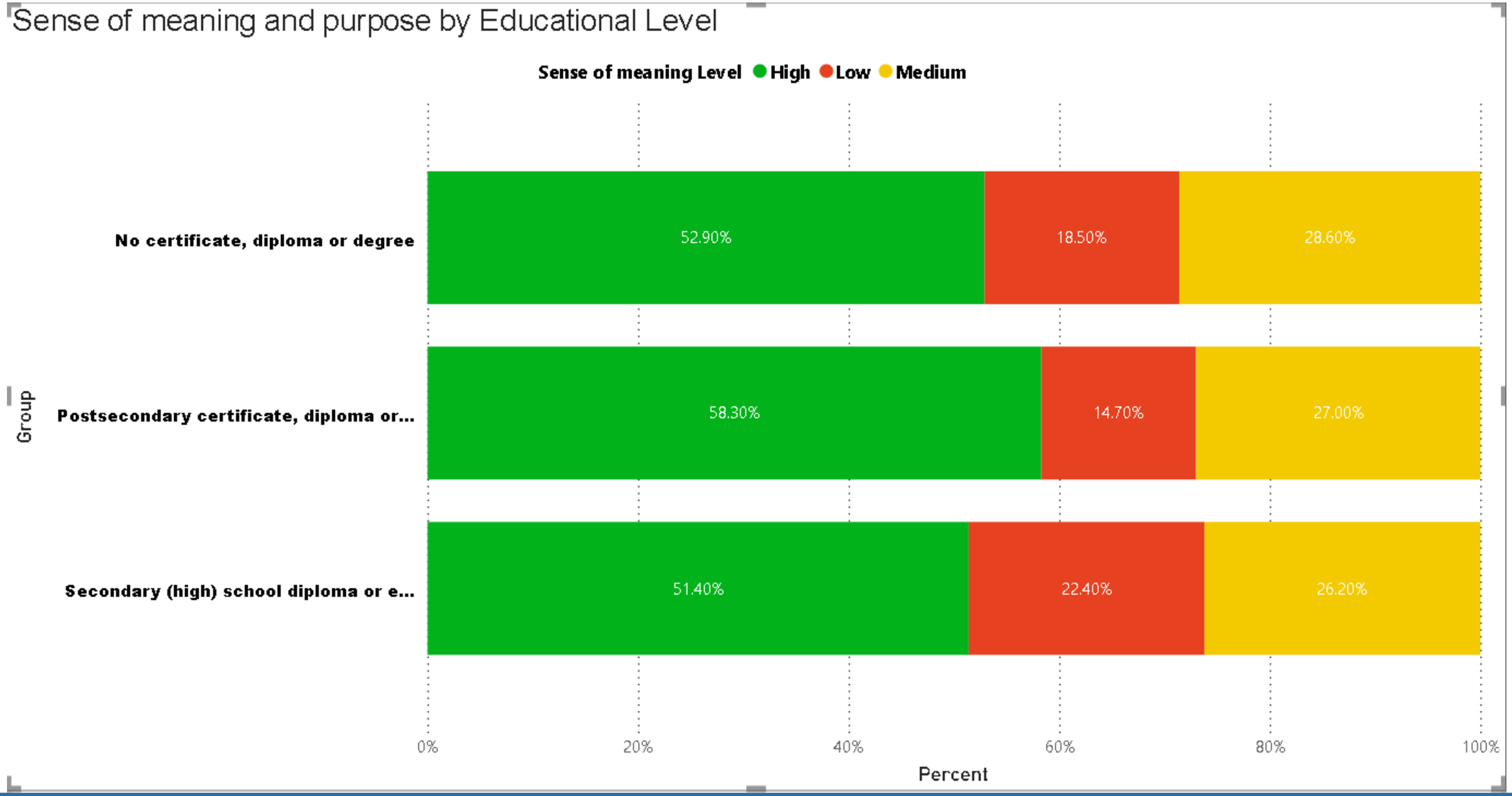


Non-LGBTQ2+ people

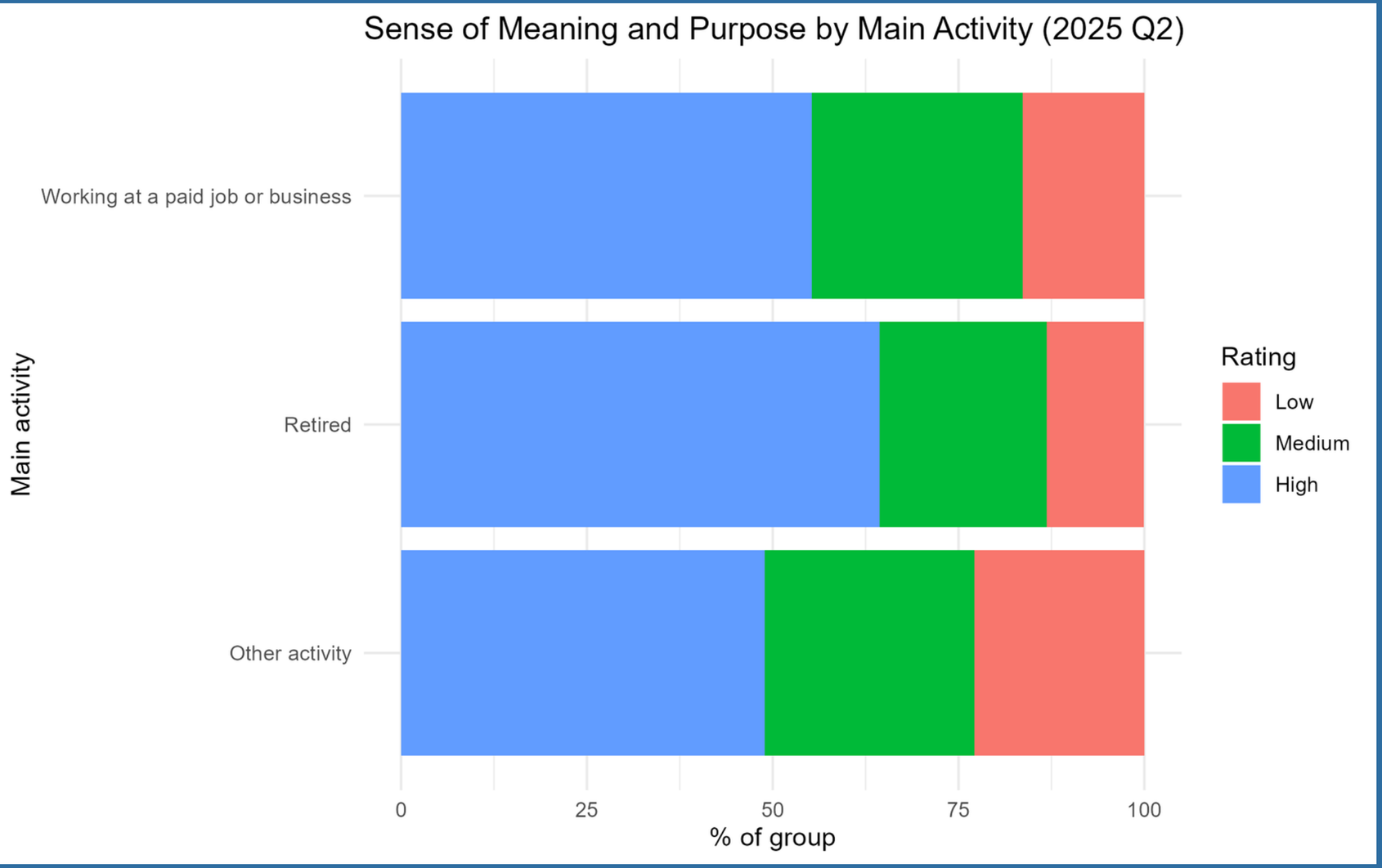


Non-LGBTQ2+ people tend to have a higher proportion with High meaning and purpose, while LGBTQ2+ people have a larger share in the Low category. This suggest that discrimination or social exclusion may negatively impact their sense of purpose.

Sense of Meaning and Purpose in Provinces by Educational Level



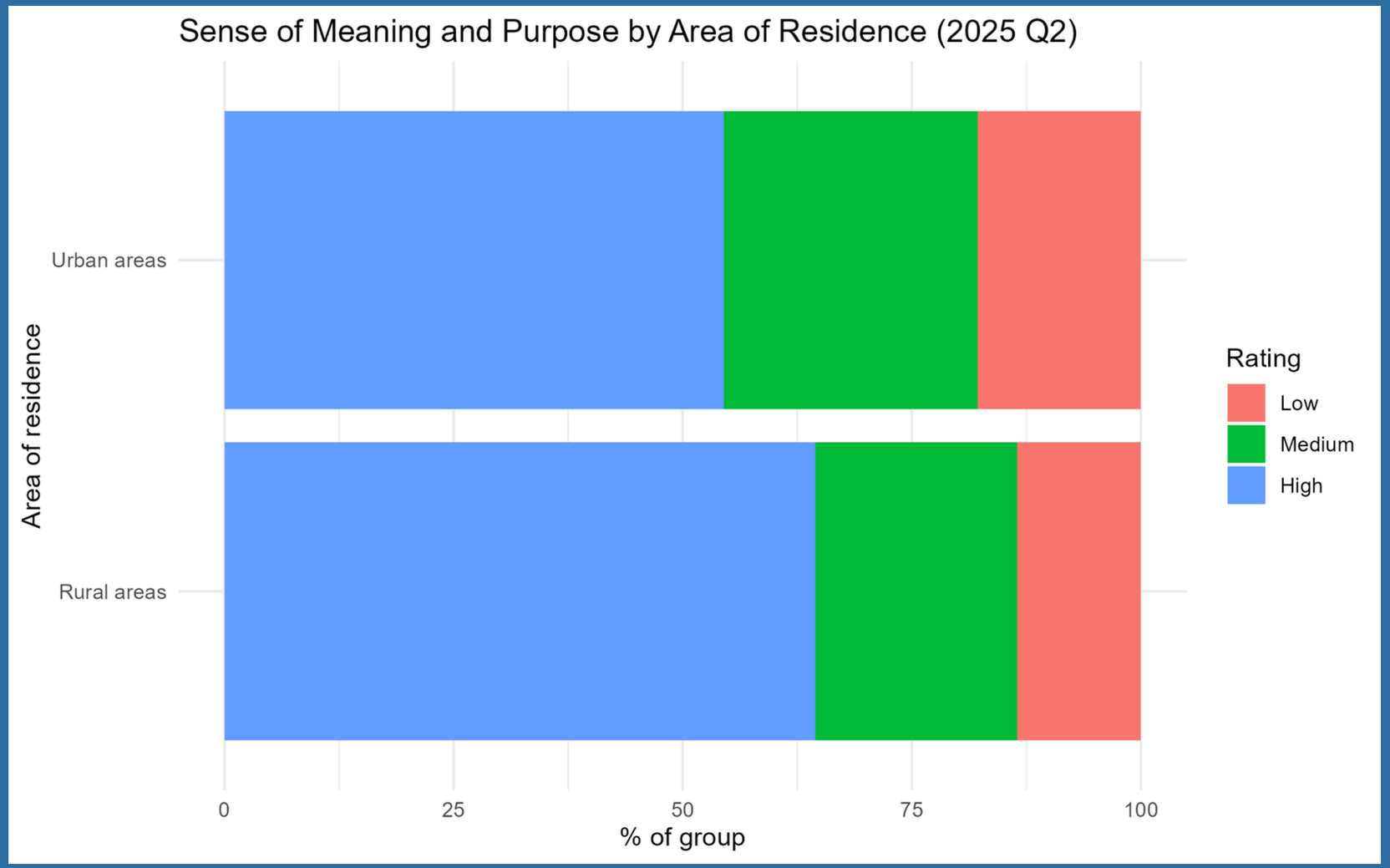
Sense of Meaning and Purpose in Canada by Career Status



Retired people often show very high levels of meaning and purpose.

While those in the “other activity” category may have a larger Low segment, perhaps reflecting instability (e.g., unemployment, caregiving, studying) in that group.

Sense of Meaning and Purpose in Canada by Area of Residence



Rural areas tend to have slightly higher shares in the High category than urban areas, which may reflect tighter community ties or different lifestyles.

Urban areas show a somewhat larger low segment.

Summary - Sence of Meaning and Purpose

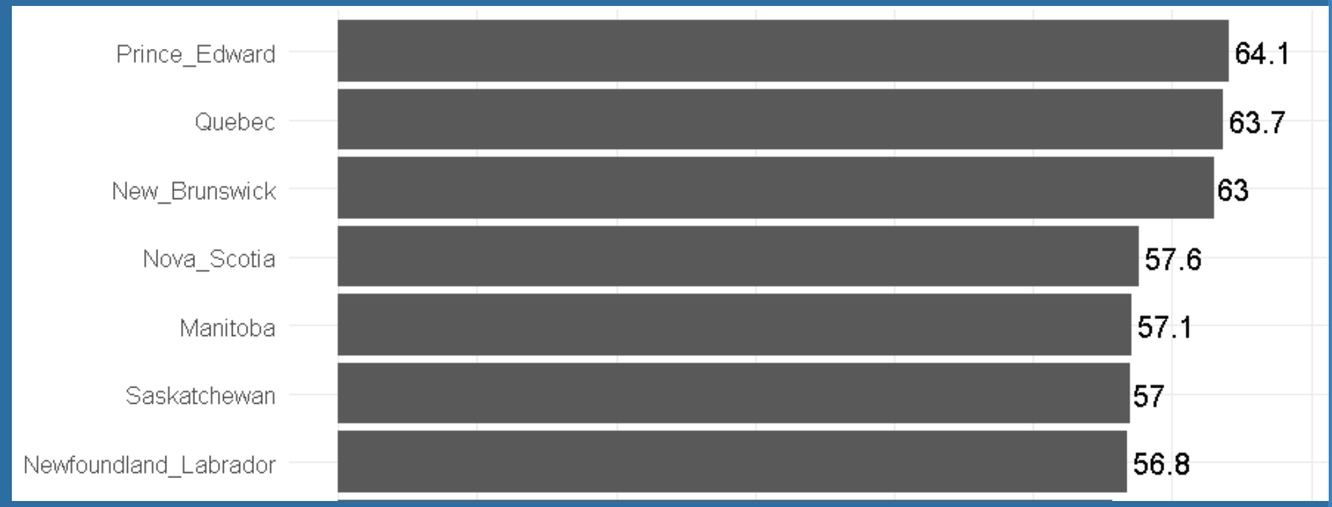
- More women find meaning and purpose in their Canadian life than men, but not much difference.

- Prince Edward Island and Quebec have more people with high sence of MP.

- The older the people gets, the more likely they feel more meaningful of their life.

W 56.7%

M 54.7%



Summary - Sence of Meaning and Purpose

- Disability matters a lot in terms of sense of meaning and purpose.
- Non-LGBTQ2+ people tend to feel more meaning and purpose in Canadian life.
- Education Level may not directly lead to higher sense of MP.
- Retired Canadian feel much more meaning in their life compared to people without a stable career.
- Rural area has more people reported a meaningful life with purpose than urban area in Canada.

CONCLUSION

- Quality of life in Canada is shaped by economic, emotional, and social factors working together.
- Poverty rates have risen again after pandemic supports ended, leaving many groups financially vulnerable.
- Mental health has declined, especially among young adults and women, while seniors show stronger resilience.
- Optimism about the future is dropping across provinces and genders, with fewer Canadians feeling confident about what lies ahead.
- Sense of meaning and purpose varies widely—older adults, rural residents, and non-disabled groups report higher purpose, while others face greater challenges.
- Improving quality of life requires addressing financial security, mental well-being, optimism, and deeper personal fulfillment.

REFERENCES

Prosperity (Poverty)

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110013501

Mental Health

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310080502

Future outlook by gender and province

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310084701

Sense of meaning and purpose by gender and province

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310084501

Sense of meaning and purpose by gender

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310084601

APPENDICES

1.Appendix 1_ Prosperity

2.Appendix 2_ Mental Health

3.Appendix 3_ Future Outlook

4.Appendix 4_ Sense of Meaning and Purpose

QUESTIONS

THANK YOU

Appendix 1_ Prosperity

```r
# Load required libraries
library(readxl)
library(dplyr)
library(ggplot2)
library(tidyr)

# COLOR THEME: Green, Blue, Yellow
color_green <- "#2E7D32"
color_blue <- "#1565C0"
color_yellow <- "#F9A825"
color_light_green <- "#81C784"
color_light_blue <- "#64B5F6"

# ============================================================
# PART 1: LOAD DATA
# ============================================================

# Read the Excel file
poverty_raw <- read_excel("poverty_data_csv.xlsx", sheet = "poverty_data")

print("RAW DATA LOADED")
str(poverty_raw)
print(paste("Total rows:", nrow(poverty_raw)))

# ============================================================
# PART 2: DATA CLEANING
# ============================================================

print("DATA CLEANING")

# Select and rename columns
poverty_clean <- poverty_raw %>%
  select(
    Year = REF_DATE,
    Geography = GEO,
    Category = `Persons in low income`,
    Poverty_Rate = VALUE,
    Status = STATUS
  ) %>%
  mutate(
    Year = as.numeric(Year),
    Poverty_Rate = as.numeric(Poverty_Rate)
  ) %>%
  filter(!is.na(Poverty_Rate))

print("Cleaned Data:")
str(poverty_clean)
print(paste("Rows after cleaning:", nrow(poverty_clean)))

print("Unique Years:")
print(sort(unique(poverty_clean$Year)))

print("Unique Categories:")
print(unique(poverty_clean$Category))

# ============================================================
# PART 3: DIVIDE INTO SEPARATE DATAFRAMES (SHEETS)
# ============================================================

print("DIVIDING DATA INTO SHEETS")

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

# SHEET: TOTAL (All persons)
total <- poverty_clean %>% filter(Category == "All persons")
print(paste("Total sheet rows:", nrow(total)))

# SHEET: GENDER
gender <- poverty_clean %>% filter(Category %in% c("Males", "Females"))
print(paste("Gender sheet rows:", nrow(gender)))

# SHEET: AGE
age <- poverty_clean %>% filter(Category %in% c("Persons under 18 years",
                                                "Persons 18 to 64 years",
print(paste("Age sheet rows:", nrow(age)))

# SHEET: FAMILY
family <- poverty_clean %>% filter(Category %in% c(
  "Persons in economic families",
  "Persons not in an economic family",
  "Seniors in economic families",
  "Seniors not in an economic family",
  "Non-seniors not in an economic family",
  "Persons under 18 years in economic families",
  "Persons under 18 years in couple families with children",
  "Persons under 18 years in female lone-parent families"))
print(paste("Family sheet rows:", nrow(family)))

# SHEET: PROVINCES
province <- total %>% filter(Geography %in% provinces_list)
print(paste("Province sheet rows:", nrow(province)))

# SHEET: CANADA
canada <- total %>% filter(Geography == "Canada")
print(paste("Canada sheet rows:", nrow(canada)))

# ============================================================
# PART 4: SUMMARY STATISTICS
# ============================================================

print("SUMMARY STATISTICS")

canada_summary <- canada %>%
  summarise(
    N = n(),
    Min = min(Poverty_Rate),
    Max = max(Poverty_Rate),
    Mean = round(mean(Poverty_Rate), 2),
    Median = median(Poverty_Rate),
    SD = round(sd(Poverty_Rate), 2)
  )
print("Canada Summary:")
print(canada_summary)

# ============================================================
# PART 5: PLOT 1 - LINE CHART: Poverty Over Time
# ============================================================

print("PLOT 1: Poverty Over Time (Line Chart)")

plot1 <- ggplot(canada, aes(x = Year, y = Poverty_Rate)) +
  geom_line(color = color_blue, linewidth = 2) +
  geom_point(color = color_green, size = 5) +
  geom_text(aes(label = paste0(Poverty_Rate, "%")),
            vjust = -1.5, hjust = 0.5, size = 5, fontface = "bold", color = "black") +
  scale_y_continuous(limits = c(0, 14), breaks = seq(0, 14, 2)) +
  scale_x_continuous(breaks = c(2020, 2021, 2022, 2023)) +
  labs(title = "Canadian Poverty Rate Over Time (2020-2023)",
       subtitle = "Percentage below poverty line (Market Basket Measure)",
       x = "Year",
       y = "Poverty Rate (%)",
       caption = "Source: Statistics Canada, Table 11-10-0135-01") +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = color_blue, hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "gray40", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12),
    panel.grid.minor = element_blank(),
    plot.margin = margin(20, 20, 20, 20)
  )

print(plot1)
ggsave("01_poverty_over_time.png", plot1, width = 10, height = 7, dpi = 300)

# ============================================================
# PART 6: PLOT 2 - BAR CHART: Province (FOR POWER BI - Skip in R)
# ============================================================
```

Appendix 1_ Prosperity

```r
# province_2023 <- province %>% filter(Year == 2023) %>% arrange(desc(Poverty_Rate))
# plot2 <- ggplot(province_2023, aes(x = reorder(Geography, Poverty_Rate), y = Poverty_Rate)) +
#   geom_bar(stat = "identity", fill = color_green, width = 0.7) +
#   coord_flip() + ...

print("PLOT 2: Province chart - CREATE IN POWER BI")

# ============================================================
# PART 7: PLOT 3 - LINE CHART: Gender Comparison
# ============================================================

print("PLOT 3: Poverty by Gender (Line Chart)")

gender_canada <- gender %>%
  filter(Geography == "Canada") %>%
  arrange(Year, Category)

plot3 <- ggplot(gender_canada, aes(x = Year, y = Poverty_Rate, color = Category, group = Category)) +
  geom_line(linewidth = 2) +
  geom_point(size = 5) +
  geom_text(aes(label = paste0(Poverty_Rate, "%")),
            vjust = -1.5, size = 4, fontface = "bold", show.legend = FALSE) +
  scale_color_manual(values = c("Females" = color_yellow, "Males" = color_blue),
                     name = "Gender") +
  scale_y_continuous(limits = c(0, 14), breaks = seq(0, 14, 2)) +
  scale_x_continuous(breaks = c(2020, 2021, 2022, 2023)) +
  labs(title = "Poverty Rate by Gender (Canada, 2020-2023)",
       subtitle = "Females consistently have higher poverty rates",
       x = "Year",
       y = "Poverty Rate (%)",
       caption = "Source: Statistics Canada, Table 11-10-0135-01") +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = color_green, hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "gray40", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12),
    axis.text = element_text(size = 12),
    legend.position = "bottom",
    legend.text = element_text(size = 12),
    panel.grid.minor = element_blank(),
    plot.margin = margin(20, 20, 20, 20)
  )

print(plot3)
ggsave("02_poverty_by_gender.png", plot3, width = 10, height = 7, dpi = 300)

# ============================================================
# PART 8: PLOT 4 - GROUPED BAR CHART: Age Groups
# ============================================================

print("PLOT 4: Poverty by Age Group (Grouped Bar Chart)")

age_canada <- age %>%
  filter(Geography == "Canada") %>%
  mutate(Age_Group = case_when(
    Category == "Persons under 18 years" ~ "Under 18",
    Category == "Persons 18 to 64 years" ~ "18-64",
    Category == "Persons 65 years and over" ~ "65+",
    TRUE ~ Category
  ))

plot4 <- ggplot(age_canada, aes(x = factor(Year), y = Poverty_Rate, fill = Age_Group)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), width = 0.7) +
  geom_text(aes(label = paste0(Poverty_Rate, "%")),
            position = position_dodge(width = 0.8),
            vjust = -0.5, size = 3.5, fontface = "bold") +
  scale_fill_manual(values = c("Under 18" = color_yellow,
                               "18-64" = color_blue,
                               "65+" = color_green),
                    name = "Age Group") +
  scale_y_continuous(limits = c(0, 16), breaks = seq(0, 16, 4)) +
  labs(title = "Poverty Rate by Age Group (Canada, 2020-2023)",
       subtitle = "Working-age adults (18-64) have highest poverty rates",
       x = "Year",
       y = "Poverty Rate (%)",
       caption = "Source: Statistics Canada, Table 11-10-0135-01") +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = color_blue, hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "gray40", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12),
    legend.position = "bottom",
    legend.text = element_text(size = 12),
    panel.grid.minor = element_blank(),
    plot.margin = margin(20, 20, 20, 20)
  )

print(plot4)
ggsave("03_poverty_by_age.png", plot4, width = 12, height = 7, dpi = 300)

# ============================================================
# PART 9: PLOT 5 - HORIZONTAL BAR CHART: Family Type
# ============================================================

print("PLOT 5: Poverty by Family Type (Bar Chart)")

family_canada_2023 <- family %>%
  filter(Geography == "Canada", Year == 2023) %>%
  mutate(Family_Short = case_when(
    Category == "Persons in economic families" ~ "In Families",
    Category == "Persons not in an economic family" ~ "Not in Family",
    Category == "Persons under 18 years in female lone-parent families" ~ "Children (Lone Parent)",
    Category == "Persons under 18 years in couple families with children" ~ "Children (Couple)",
    Category == "Persons under 18 years in economic families" ~ "Children (In Family)",
    Category == "Seniors in economic families" ~ "Seniors (In Family)",
    Category == "Seniors not in an economic family" ~ "Seniors (Alone)",
    Category == "Non-seniors not in an economic family" ~ "Non-Seniors (Alone)",
    TRUE ~ Category
  )) %>%
  arrange(desc(Poverty_Rate))

plot5 <- ggplot(family_canada_2023, aes(x = reorder(Family_Short, Poverty_Rate), y = Poverty_Rate)) +
  geom_bar(stat = "identity", fill = color_blue, width = 0.7) +
  geom_text(aes(label = paste0(Poverty_Rate, "%")),
            hjust = -0.2, size = 4, fontface = "bold") +
  coord_flip() +
  scale_y_continuous(limits = c(0, 38), breaks = seq(0, 35, 5)) +
  labs(title = "Poverty Rate by Family Type (Canada, 2023)",
       subtitle = "Non-seniors living alone face highest poverty risk",
       x = "",
       y = "Poverty Rate (%)",
       caption = "Source: Statistics Canada, Table 11-10-0135-01") +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = color_green, hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "gray40", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 11),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_blank(),
    plot.margin = margin(20, 30, 20, 20)
  )

print(plot5)
ggsave("04_poverty_by_family.png", plot5, width = 11, height = 7, dpi = 300)

# ============================================================
# PART 10: PLOT 6 - SCATTER PLOT: Province vs Year
# ============================================================

print("PLOT 6: Scatter Plot - Province Trends")

province_scatter <- province %>%
  mutate(Province_Short = case_when(
    Geography == "Newfoundland and Labrador" ~ "NL",
    Geography == "Prince Edward Island" ~ "PE",
    Geography == "Nova Scotia" ~ "NS",
    Geography == "New Brunswick" ~ "NB",
    Geography == "Quebec" ~ "QC",
    Geography == "Ontario" ~ "ON",
    Geography == "Manitoba" ~ "MB",
    Geography == "Saskatchewan" ~ "SK",
    Geography == "Alberta" ~ "AB",
    Geography == "British Columbia" ~ "BC",
    TRUE ~ Geography
  ))

plot6 <- ggplot(province_scatter, aes(x = Year, y = Poverty_Rate, color = Province_Short)) +
  geom_point(size = 4, alpha = 0.8) +
  geom_line(aes(group = Province_Short), linewidth = 1, alpha = 0.6) +
  scale_color_viridis_d(option = "turbo", name = "Province") +
  scale_x_continuous(breaks = c(2020, 2021, 2022, 2023)) +
  scale_y_continuous(limits = c(0, 16), breaks = seq(0, 16, 4)) +
  labs(title = "Provincial Poverty Rate Trends (2020-2023)",
       subtitle = "Each line represents a province's poverty trajectory",
       x = "Year",
       y = "Poverty Rate (%)",
       caption = "Source: Statistics Canada, Table 11-10-0135-01") +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = color_blue, hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "gray40", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12),
    legend.position = "right",
    panel.grid.minor = element_blank(),
    plot.margin = margin(20, 20, 20, 20)
  )

print(plot6)
ggsave("05_poverty_scatter_province.png", plot6, width = 12, height = 7, dpi = 300)

# ============================================================
# PART 11: PLOT 7 - HISTOGRAM: Distribution of Poverty Rates
# ============================================================

print("PLOT 7: Histogram - Distribution of Poverty Rates")

all_2023 <- poverty_clean %>% filter(Year == 2023)

plot7 <- ggplot(all_2023, aes(x = Poverty_Rate)) +
  geom_histogram(binwidth = 3, fill = color_green, color = "white", alpha = 0.8) +
  geom_vline(aes(xintercept = mean(Poverty_Rate)),
             color = color_yellow, linewidth = 1.5, linetype = "dashed") +
  annotate("text", x = mean(all_2023$Poverty_Rate) + 3, y = 40,
           label = paste("Mean:", round(mean(all_2023$Poverty_Rate), 1), "%"),
           fontface = "bold", size = 5, color = color_yellow) +
  scale_x_continuous(breaks = seq(0, 35, 5)) +
  labs(title = "Distribution of Poverty Rates (2023)",
       subtitle = "Histogram showing frequency of poverty rate values",
       x = "Poverty Rate (%)",
       y = "Frequency",
       caption = "Source: Statistics Canada, Table 11-10-0135-01") +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = color_blue, hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "gray40", hjust = 0.5),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text = element_text(size = 12),
    plot.margin = margin(20, 20, 20, 20)
  )

print(plot7)
ggsave("06_poverty_histogram.png", plot7, width = 10, height = 7, dpi = 300)

# ============================================================
# PART 12: KEY FINDINGS
# ============================================================

print("----------------------------------------")
print("KEY FINDINGS")
print("----------------------------------------")

canada_2023_rate <- canada %>% filter(Year == 2023) %>% pull(Poverty_Rate)
canada_2020_rate <- canada %>% filter(Year == 2020) %>% pull(Poverty_Rate)
print(paste("Canada Poverty Rate (2023):", canada_2023_rate, "%"))
print(paste("Change 2020-2023:", round(canada_2023_rate - canada_2020_rate, 1), "percentage points"))

province_2023 <- province %>% filter(Year == 2023) %>% arrange(desc(Poverty_Rate))
print(paste("Highest Province:", province_2023$Geography[1], "-", province_2023$Poverty_Rate[1], "%"))
print(paste("Lowest Province:", tail(province_2023$Geography, 1), "-", tail(province_2023$Poverty_Rate, 1), "%"))

gender_2023 <- gender_canada %>% filter(Year == 2023)
print(paste("Females (2023):", gender_2023$Poverty_Rate[gender_2023$Category == "Females"], "%"))
print(paste("Males (2023):", gender_2023$Poverty_Rate[gender_2023$Category == "Males"], "%"))

print("----------------------------------------")
print("ANALYSIS COMPLETE! 6 plots saved as PNG.")
```

Appendix 2_ Mental Health

```r
#==========================================================================
#Libirary
library(tidyverse)
library(tidyr)
library(stringr)
library(readr)
library(dplyr)
library(ggplot2)

#==========================================================================
#Load CSV file
health <- read.csv("Final_Dataset_Thuy.csv")
head(health,10)
summary(health)

#==========================================================================
# Create Gender Table:
genders <- c("Both_Genders", "Males", "Females")
gender_table <- data.frame(Gender = genders)
print(gender_table)
library(knitr)
kable(gender_table, caption = "Gender Categories")
#--------------------------------------------------------------------------
# Create Indicators Table
indicators <- c("Good_Excellent", "Fair_Poor", "Stressful")
indicators_table <- data.frame(Indicators = indicators)
print(indicators_table)
kable(indicators_table, caption = "Indicator Types")
#--------------------------------------------------------------------------
# Create Age Group Table:
age_group <- data.frame(
  Age_Group = c("Total", "Teen", "Young", "Midlife", "Older", "Senior"),
  Age_Range = c("12 years and over",
                "12 to 17 years",
                "18 to 34 years",
                "35 to 49 years",
                "50 to 64 years",
                "65 years and over"))
print(age_group)
kable(age_group, caption = "Age Groups in Dataset")
```

```r
#==========================================================================
#Data Cleaning:
#--------------------------------------------------------------------------
##1.Standardize Text fields:
health <- health %>%
  mutate(
    Location   = str_squish(str_trim(Location)),
    Age_group  = str_squish(str_trim(Age_group)),
    Gender     = str_squish(str_trim(Gender)),
    Indicators = str_squish(str_trim(Indicators))
  )

## Remove undefined value (F value, NA value):
health <- health %>% filter(if_all(everything(), ~ . != "F"))
health <- health %>% drop_na(Location, Age_group, Gender, Indicators, Year, Percentage)

#==========================================================================
# Filter relevant data
health_summary <- health %>%
  filter(Location != "Canada",
         Age_group == "Total",
         Gender %in% c("Both_Genders", "Males", "Females"),
         Indicators %in% c("Good_Excellent", "Fair_Poor", "Stressful"),
         Year == 2022) %>%
  mutate(Percentage = as.numeric(Percentage)) %>%
  filter(!is.na(Percentage))

# Create summary table
summary_table <- health_summary %>%
  select(Location, Indicators, Gender, Percentage) %>%
  pivot_wider(names_from = Gender, values_from = Percentage) %>%
  arrange(Location, Indicators)

# View the table
print(summary_table)

#
```

```r
#SECTION 1: "Mental Health by Indicator":
# 1) Filter Canada, Total age, Both genders
health_canada <- health %>%
  filter(
    Location == "Canada",
    Age_group == "Total",
    Gender == "Both_Genders")

health_canada$Year <- as.numeric(as.character(health_canada$Year))
health_canada$Percentage <- as.numeric(health_canada$Percentage)

# 2) Split by Indicators
good   <- subset(health_canada, Indicators == "Good_Excellent")
fair   <- subset(health_canada, Indicators == "Fair_Poor")
stress <- subset(health_canada, Indicators == "Stressful")

# 3) Ensure Year is numeric
good$Year   <- as.numeric(good$Year)
fair$Year   <- as.numeric(fair$Year)
stress$Year <- as.numeric(stress$Year)

# 4) Start empty plot :
plot(good$Year, good$Percentage,
     type = "n",
     xlab = "Year",
     ylab = "Percentage (%)",
     ylim = range(health_canada$Percentage, na.rm = TRUE),
     main = "Mental Health by Indicators")

# 5) Add 3 indicator lines
lines(good$Year,   good$Percentage,   col = "blue",     lwd = 2)
lines(fair$Year,   fair$Percentage,   col = "skyblue", lwd = 2)
lines(stress$Year, stress$Percentage, col = "red",      lwd = 2)

points(good$Year,   good$Percentage,   col = "blue",     pch = 16)
points(fair$Year,   fair$Percentage,   col = "skyblue", pch = 16)
points(stress$Year, stress$Percentage, col = "red",      pch = 16)
```

```r
# 6) Add ablines:
abline(h=c(10,20,55,70), col="grey", lty = 2, lwd =1.5)

# 7) Add comments
text(x = 2021, y = 68, labels = "↓ Decline in Good/Excellent", col = "blue", cex = 0.9)
text(x = 2019, y = 12, labels = "↑ Rise in Fair/Poor", col = "black", cex = 0.9)
text(x = 2019, y = 25, labels = "Stable stress levels", col = "red", cex = 0.9)

# 8) Add Arrows
arrows(x0 = 2017, y0 = 70, x1 = 2017, y1 = 10, col = "green", lwd = 2, length = 0.1, code=3,lty=4)
text(x = 2017, y = 39, labels = "Wider gap", col = "darkgreen", cex = 1)
rect(xleft = 2016.5, xright = 2017.7, ybottom = 37.5, ytop = 38,
     col = "yellow", border = NA)

arrows(x0 = 2021, y0 = 57, x1 = 2021, y1 = 13, col = "green", lwd = 2, length = 0.1, code =3, lty=4)
text(x = 2021, y = 35, labels = "Narrower gap ", col = "darkgreen", cex = 1)
rect(xleft = 2020.5, xright = 2021.5, ybottom = 33.5, ytop = 34,
     col = "yellow", border = NA)
```

Appendix 2_ Mental Health

```r
# SECTION 2: Mental Health by GENDER:
# Prepare dataset
health_gender <- health %>%
  filter(Location == "Canada",
         Age_group == "Total",
         Gender %in% c("Males", "Females"),
         Indicators %in% c("Good_Excellent", "Fair_Poor", "Stressful")) %>%
  mutate(Percentage = as.numeric(Percentage)) %>%
  filter(!is.na(Percentage))

# Plot1:
p1 <- ggplot(health_gender, aes(x = factor(Year), y = Percentage, fill = Gender)) +
  geom_col(position = "dodge") +
  facet_wrap(~Indicators, nrow = 1) +
  scale_fill_manual(values = c("Males" = "steelblue", "Females" = "pink")) +
  labs(title = "Mental Health by Gender",
       x = "Year", y = "Percentage (%)") +
  theme_minimal(base_size = 14) +
  theme(
    panel.grid.major.y = element_line(color = "gray70", size = 1, linetype = "dashed"),
    panel.grid.minor.y = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank())
print(p1)
```

```r
# Plot 2. Distribution of indicators:
p2 <- ggplot(health_gender, aes(x = Gender, y = Percentage, fill = Gender)) +
  geom_boxplot() +
  facet_wrap(~Indicators) +
  scale_fill_manual(values = c("Males" = "steelblue", "Females" = "pink")) +
  labs(title = "Distribution by Gender",
       x = "Gender", y = "Percentage (%)") +
  theme_minimal(base_size = 14)
print(p2)

# Plot 3. Gender gap avr years:
gender_avg <- health_gender %>%
  group_by(Gender, Indicators) %>%
  summarise(Average = mean(Percentage, na.rm = TRUE), .groups = "drop")
p3 <- ggplot(gender_avg, aes(x = Gender, y = Average, group = Indicators, color = Gender)) +
  geom_line(aes(group = Indicators), position = position_dodge(width = 0.2)) +
  geom_point(size = 4) +
  facet_wrap(~Indicators) +
  scale_color_manual(values = c("Males" = "steelblue", "Females" = "pink")) +
  labs(title = "Average Gender Gap by Indicators",
       x = "Gender", y = "Average Percentage (%)") +
  theme_minimal(base_size = 14)
print(p3)
```

```r
#SECTION 3: Mental Health by LOCATION:
# HEATMAP BY LOCATION
# Prepare dataset:
health_heatmap <- health %>%
  filter(Location != "Canada",
         Age_group == "Total",
         Gender == "Both_Genders",
         Indicators %in% c("Good_Excellent", "Fair_Poor", "Stressful"),
         Year == 2022) %>%
  mutate(Percentage = as.numeric(Percentage)) %>%
  filter(!is.na(Percentage))

# Heatmap: Province vs Indicator
health_heatmap$Indicators <- factor(health_heatmap$Indicators,
                                    levels = c("Good_Excellent", "Fair_Poor", "Stressful"))
ggplot(health_heatmap, aes(x = Indicators, y = reorder(Location, Percentage), fill = Percentage)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "red") +
  labs(title = "Mental Health Heatmap by Province (2022)",
       x = "Indicator", y = "Province", fill = "Percentage (%)") +
  theme_minimal(base_size = 14)
```

```r
#DOT MAP
# Prepare dataset
health_province <- health %>%
  filter(Location != "Canada",
         Age_group == "Total",
         Gender %in% c("Males", "Females"),
         Indicators %in% c("Good_Excellent", "Fair_Poor", "Stressful"),
         Year == 2022) %>%
  mutate(Percentage = as.numeric(Percentage),
         Location_num = as.numeric(factor(Location))) %>%
  filter(!is.na(Percentage),
         !is.na(Gender), Gender != "",
         !is.na(Location), Location != "")

highlight_quebec <- health_province %>%
  filter(Location == "Quebec", Indicators == "Good_Excellent", Gender == "Females")

highlight_female <- highlight_quebec %>% filter(Gender == "Females") %>% slice(1)
highlight_male <- highlight_quebec %>% filter(Gender == "Males") %>% slice(1)
```

```r
# Plot
ggplot(health_province, aes(x = Percentage, y = Location, color = Indicators)) +
  geom_point(size = 2) +
  geom_point(data = highlight_quebec, aes(x = Percentage, y = Location),
             shape = 21, size = 5, stroke = 1.2, color = "yellow", fill = NA) +
  annotate("text",
           x = highlight_female$Percentage + 2,
           y = highlight_female$Location_num + 0.5,
           label = "Highest         ", color = "red", hjust = 0, size = 2.5) +
  annotate("text",
           x = highlight_male$Percentage + 2,
           y = highlight_male$Location_num + 0.1,
           label = "Highest", color = "red", hjust = 0, size = 3) +
  facet_wrap(~Gender) +
  scale_color_manual(values = c("Good_Excellent" = "steelblue",
                                "Fair_Poor" = "lightblue",
                                "Stressful" = "red")) +
  labs(title = "Mental Health Indicators by Province (2022)",
       subtitle = "Quebec's Good_Excellent highlighted",
       x = "Percentage (%)", y = "Province", color = "Indicator") +
  theme_minimal(base_size = 14)
```

Appendix 2_ Mental Health

```r
#SECTION 4: Mental Health by Age Group:
ggplot(health_2022_total,
       aes(x = Age_group,
           y = Percentage,
           fill = Indicators)) +
geom_col(width = 0.8) +

#Highlight only Senior in Good/Excellent
geom_col(
    data = subset(health_2022_total,Age_group == "Senior" & Indicators == "Good / Excellent"),
    fill = "gold", width = 0.8) +

# Value labels
geom_text(aes(label = Percentage),vjust = -0.5,size = 3.5) +

#TEXT LABEL
geom_text(
    data = health_2022_total %>%
      filter(Indicators == "Good / Excellent",
             Age_group  == "Senior") %>%
      mutate(y_text = Percentage * 1.25),
    aes(x = 3.5,
        y = y_text,
        label = "Seniors have highest level \n of good/excellent condition"),
    inherit.aes = FALSE,
    color = "red",
    size = 4,       # smaller text
    fontface = "bold",
    hjust = 0.5) +

#ARROW
geom_segment( data = health_2022_total %>%
      filter(Indicators == "Good / Excellent",
             Age_group  == "Senior") %>%
      mutate(y_start = Percentage * 1.18,
             y_end   = Percentage * 1.05),
    aes(x = Age_group, xend = Age_group,
        y = y_start,yend = y_end),
    inherit.aes = FALSE,
    colour = "red",        # red arrow
    size = 0.9,
    arrow = arrow(length = unit(0.25, "cm"))) +

# y-axis
scale_y_continuous( limits = c(0, max(health_2022_total$Percentage, na.rm = TRUE) * 1.35)) +

# custom indicator colors
scale_fill_manual(values = c( "Good / Excellent" = "steelblue","Fair / Poor" = "pink",
"Stressful" = "gray70")) +

facet_wrap(~ Indicators, nrow = 1,strip.position = "bottom",scales = "fixed") +

labs(title = "Mental Health by Age Groups in 2022") +

theme_minimal(base_size = 12) + theme(
    axis.text.x      = element_text(angle = 40, hjust = 1),
    strip.background = element_blank(),
    strip.placement  = "outside",
    axis.text.y      = element_blank(),
    axis.ticks.y     = element_blank(),
```

Appendix 3_Future Outlook

```r
# Big Data 2 - Course Project
# Central Indicators Domain: Future Outlook
# Data Source: Statistics Canada Quality of Life Hub
# https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310084701
# Table 13-10-0847-01: Future outlook by gender and province

rm(list = ls()); cat("\014")

# Load required libraries
library(stringr)
library(readr)
library(dplyr)

# Step 1: Load dataset
futout_raw <- read_csv("Future_Outlook.csv")

# Step 2: View dataset structure
print("Data Structure")
str(futout_raw)
print(paste("Total rows in raw data:", nrow(futout_raw)))
print(paste("Total columns:", ncol(futout_raw)))

print("Column data types")
str(futout_raw)

# Step 3: Data Cleaning
futout_clean <- futout_raw %>%

  # Convert Number to numeric (remove commas if present)
  mutate(
    Number = as.numeric(gsub(",", "", Number)),
  ) %>%


  # Convert 'Reference period' to date type
  mutate(
    Year = as.numeric(str_extract(`Reference period`, "\\d{4}")),

    Date = case_when(
      str_detect(`Reference period`, "Q1") ~ as.Date(paste0(Year, "-01-01")),
      str_detect(`Reference period`, "Q2") ~ as.Date(paste0(Year, "-04-01")),
      str_detect(`Reference period`, "Q3") ~ as.Date(paste0(Year, "-07-01")),
      str_detect(`Reference period`, "Q4") ~ as.Date(paste0(Year, "-10-01"))
    )
  ) %>%

  # Remove helper column Year (we don't need it anymore)
  select(-Year, -`Reference period`)

print("First 5 rows AFTER cleaning:")
head(futout_clean, 5)

str(futout_clean)

# Step 4: Save cleaned dataset
write_csv(futout_clean, "Future_Outlook_Clean.csv")

print("Cleaned dataset saved as 'Future_Outlook_Clean.csv'")
```

```r
# Slide 1: Summary Table + Pie Chart
rm(list = ls()); cat("\014")

# Load libraries
library(readr)
library(dplyr)

# Load cleaned dataset
futout_clean <- read_csv("Future_Outlook_Clean.csv")

# Summary Table

summary_table <- futout_clean %>%
  filter(Geography == "Canada") %>%
  group_by(Indicators, Gender) %>%
  summarise(
    Avg_Number = round(mean(Number, na.rm = TRUE), 1),
    Avg_Percent = round(mean(Percent, na.rm = TRUE), 1),
    .groups = "drop"
  ) %>%
  mutate(
    Indicators = factor(Indicators, levels = c("Always", "Sometimes", "Rarely")),
    Gender = factor(Gender, levels = c("Total", "Men", "Women"))
  ) %>%
  arrange(Indicators, Gender)

print(summary_table)
```

```r
# Summary Statistics

summary_stats <- futout_clean %>%
  filter(Geography == "Canada") %>%
  group_by(Indicators) %>%
  summarise(
    Min_Percent = min(Percent, na.rm = TRUE),
    Max_Percent = max(Percent, na.rm = TRUE),
    Mean_Percent = round(mean(Percent, na.rm = TRUE), 2),
    Median_Percent = median(Percent, na.rm = TRUE),
    SD_Percent = round(sd(Percent, na.rm = TRUE), 2),
    .groups = "drop"
  ) %>%

  mutate(
    Indicators = factor(Indicators,
                        levels = c("Always", "Sometimes", "Rarely"))
  ) %>%
  arrange(Indicators)

print(summary_stats)
```

```r
# Pie Chart

pie_data <- futout_clean %>%
  group_by(Indicators) %>%
  summarise(AvgPercent = mean(Percent), .groups = "drop")

labels <- paste0(
  pie_data$Indicators,
  " (", round(pie_data$AvgPercent, 1), "%)"
)

colors <- c("#4E79A7", "#F28E2B", "#E15759")

pie(
  pie_data$AvgPercent,
  labels = labels,
  col = colors,
  border = "white",
  cex = 1.2,
  main = "Distribution of Future Outlook Indicators (Avg %)"
)
box(lwd = 2, col = "gray40")
```

Appendix 3_Future Outlook

```r
# Slide 2: Future Outlook Over Time (Canada Only)
rm(list = ls()); cat("\014")

library(dplyr)
library(ggplot2)
library(lubridate)

# Load cleaned dataset
futout_clean <- read_csv("Future_Outlook_Clean.csv")

# Filter only Canada & Total, create quarter labels
canada_time <- futout_clean %>%
  filter(Geography == "Canada", Gender == "Total") %>%
  mutate(
    QuarterLabel = paste0(year(Date), " Q", quarter(Date)),
    Indicators = factor(Indicators, levels = c("Always", "Sometimes", "Rarely"))
  )

# Plot
plot_time <- ggplot(canada_time,
                    aes(x = Date, y = Percent,
                        color = Indicators, group = Indicators)) +
  geom_line(linewidth = 1.5) +
  geom_point(size = 3.5) +

  # Bigger percentage labels
  geom_text(aes(label = round(Percent, 1)),
            vjust = -0.9, size = 4.2, fontface = "bold",
            show.legend = FALSE) +

  # Use custom quarter labels on x-axis
  scale_x_date(
    breaks = canada_time$Date,
    labels = canada_time$QuarterLabel
  ) +

  scale_color_manual(values = c(
    "Always" = "#4E79A7",
    "Sometimes" = "#F28E2B",
    "Rarely" = "#E15759"
  )) +

  labs(
    title = "Future Outlook Over Time (Canada)",
    subtitle = "Quarterly share of respondents who feel optimistic about their future ,
    x = "Period",
    y = "% of respondents",
    color = "Indicator"
  ) +

  theme_minimal(base_size = 12) +
  theme(
    axis.title.x = element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 18, face = "bold", color = "#C41E3A"),
    plot.subtitle = element_text(size = 11, color = "gray40"),
    panel.grid.minor = element_blank(),
    legend.position = "right"
  )

print(plot_time)
```

```r
# Slide 3: Average Future Outlook by Province
rm(list = ls()); cat("\014")

library(readr)
library(dplyr)
library(ggplot2)

futout_clean <- read_csv("Future_Outlook_Clean.csv")

prov <- futout_clean %>%
  # select provinces only, total population
  filter(Geography != "Canada", Gender == "Total") %>%
  mutate(
    # fix names
    Geography = gsub("_", " ", Geography),
    Geography = ifelse(Geography == "NovaScotia", "Nova Scotia", Geography),
    Indicators = factor(Indicators,
                        levels = c("Always", "Sometimes", "Rarely"))
  ) %>%
  group_by(Geography, Indicators) %>%
  summarise(AvgPercent = mean(Percent, na.rm = TRUE), .groups = "drop")

# Plot
ggplot(prov, aes(x = Geography, y = AvgPercent, fill = Indicators)) +
  geom_col(position = "dodge") +
  scale_fill_manual(values = c(
    "Always" = "#4E79A7",
    "Sometimes" = "#F28E2B",
    "Rarely" = "#E15759"
  )) +
  labs(
    title = "Average Future Outlook by Province",
    subtitle = "Average percentage of respondents who feel optimistic about their future",
    x = "Province",
    y = "Average Percent (%)",
    fill = "Indicator"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#C41E3A"),
    plot.subtitle = element_text(size = 11, color = "gray40"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "right",
    panel.grid.minor = element_blank()
  )
```

```r
# Slide 4: Future Outlook Over Time by Gender (Canada)
rm(list = ls()); cat("\014")

library(readr)
library(dplyr)
library(ggplot2)
library(lubridate)

# Load cleaned dataset
futout_clean <- read_csv("Future_Outlook_Clean.csv")

# Prepare data for Canada, with quarter labels
gender_line <- futout_clean %>%
  filter(Geography == "Canada",
         Gender %in% c("Total", "Men", "Women")) %>%
  mutate(
    QuarterLabel = paste0(year(Date), " Q", quarter(Date)),
    Indicators = factor(Indicators,
                        levels = c("Always", "Sometimes", "Rarely")),
    Gender = factor(Gender, levels = c("Total", "Men", "Women"))
  )

# Line chart: 3 panels (Total, Men, Women)
plot_gender_line <- ggplot(
  gender_line,
  aes(x = QuarterLabel, y = Percent,
      color = Indicators, group = Indicators)
) +
  geom_line(linewidth = 1.3) +
  geom_point(size = 2.8) +
  geom_text(aes(label = round(Percent, 1)),
            vjust = -0.8, size = 3, show.legend = FALSE) +

  scale_y_continuous(expand = expansion(mult = c(0.10, 0.15))) +

  scale_color_manual(values = c(
    "Always" = "#4E79A7",
    "Sometimes" = "#F28E2B",
    "Rarely" = "#E15759"
  )) +

  facet_wrap(~ Gender, ncol = 1) +

  labs(
    title = "Future Outlook Over Time by Gender (Canada)",
    subtitle = "Quarterly share of Canadians who feel optimistic, compared across gender groups",
    x = "Quarter",
    y = "% of respondents",
    color = "Indicator"
  ) +

  theme_minimal(base_size = 12) +
  theme(
    strip.text = element_text(size = 13, face = "bold"),
    axis.title.x = element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 18, face = "bold", color = "#C41E3A"),
    plot.subtitle = element_text(size = 11, color = "gray40"),
    panel.grid.minor = element_blank(),
    legend.position = "right"
  )

print(plot_gender_line)
```

Appendix 3_Future Outlook

```r
# Slide 5: Average Future Outlook by Province and Gender
rm(list = ls()); cat("\014")

library(dplyr)
library(ggplot2)
library(readr)

# Load cleaned dataset
futout_clean <- read_csv("Future_Outlook_Clean.csv")

# Prepare province-level average data by gender
prov_heatmap_gender <- futout_clean %>%
  filter(Geography != "Canada",
         Gender %in% c("Total", "Men", "Women")) %>%
  mutate(
    Geography = gsub("_", " ", Geography),
    Indicators = factor(Indicators, levels = c("Always", "Sometimes", "Rarely")),
    Gender = factor(Gender, levels = c("Total", "Men", "Women"))
  ) %>%
  group_by(Geography, Indicators, Gender) %>%
  summarise(
    AvgPercent = round(mean(Percent, na.rm = TRUE), 1),
    .groups = "drop"
  ) %>%

  # Create internal combined column for positioning
  mutate(
    Column = paste(Indicators, Gender, sep = "_"),

    # Simplify label shown on the x-axis
    Label = Gender
  )

# Order columns properly
column_order <- c(
  "Always_Total", "Always_Men", "Always_Women",
  "Sometimes_Total", "Sometimes_Men", "Sometimes_Women",
  "Rarely_Total", "Rarely_Men", "Rarely_Women"
)
```

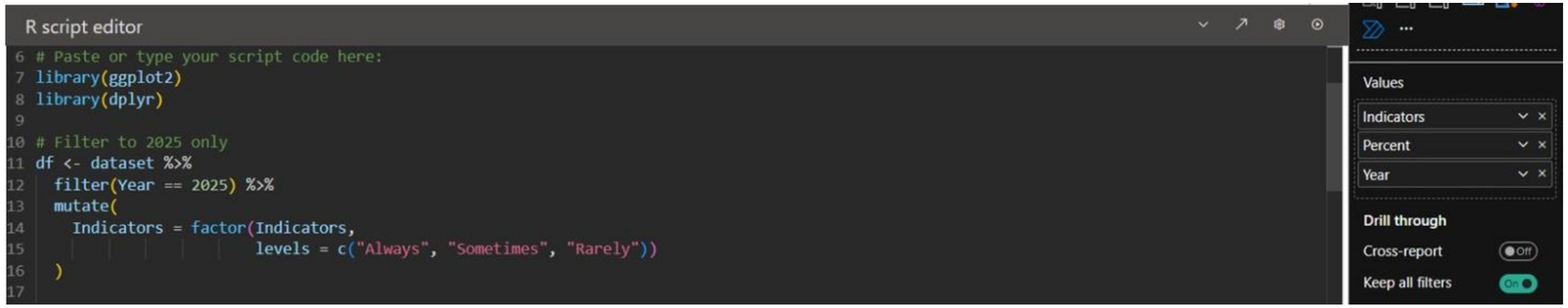
```r
prov_heatmap_gender$Column <- factor(prov_heatmap_gender$Column,
                                     levels = column_order)

# Plot heatmap
plot_heat_gender <- ggplot(prov_heatmap_gender,
                           aes(x = Column, y = Geography, fill = Indicators)) +

  geom_tile(color = "white", linewidth = 0.4) +

  geom_text(aes(label = AvgPercent),
            color = "black", size = 3.5, fontface = "bold") +

  scale_fill_manual(values = c(
    "Always"    = "#4E79A7",
    "Sometimes" = "#F28E2B",
    "Rarely"    = "#E15759"
  )) +

  # Replace x-axis text with simplified Gender labels
  scale_x_discrete(labels = prov_heatmap_gender$Label) +

  labs(
    title = "Average Future Outlook by Province and Gender",
    subtitle = "Heatmap comparing optimism patterns across demographic groups",
    fill = "Indicator"
  ) +

  theme_minimal(base_size = 11) +
  theme(
    axis.title = element_blank(),
    axis.text.x = element_text(angle = 45, hjust = 1, size = 10),
    plot.title = element_text(size = 18, face = "bold", color = "#C41E3A"),
    plot.subtitle = element_text(size = 11, color = "gray40"),
    panel.grid = element_blank(),
    legend.position = "right"
  )

print(plot_heat_gender)
```

```r
# Slide 6: Respondent Count Variation by Indicator
rm(list = ls()); cat("\014")

library(dplyr)
library(ggplot2)
library(readr)

futout_clean <- read_csv("Future_Outlook_Clean.csv")

canada_box <- futout_clean %>%
  filter(Geography == "Canada",
         Gender == "Total",
         !is.na(Number)) %>%
  mutate(
    Indicators = factor(Indicators,
                        levels = c("Always", "Sometimes", "Rarely"))
  )

box_plot <- ggplot(canada_box, aes(x = Indicators, y = Number, fill = Indicators)) +
  geom_boxplot(alpha = 0.85, width = 0.6, outlier.color = "gray30") +

  scale_fill_manual(values = c(
    "Always" = "#4E79A7",
    "Sometimes" = "#F28E2B",
    "Rarely" = "#E15759"
  )) +

  labs(
    title = "Respondent Count Variation by Indicator",
    subtitle = "Boxplot showing median, quartiles, and spread of optimism response counts in Canada",
    x = "Indicator",
    y = "Number of respondents",
    fill = "Indicator"
  ) +
```

```r
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#C41E3A"),
    plot.subtitle = element_text(size = 11, color = "gray40"),
    axis.title.x = element_blank(),
    panel.grid.minor = element_blank()
  )

print(box_plot)
```

Appendix 3_Future Outlook

```r
# Slide 7: Respondent Count Distribution by Indicator
rm(list = ls()); cat("\014")

library(dplyr)
library(ggplot2)
library(readr)

# Load cleaned dataset
futout_clean <- read_csv("Future_Outlook_Clean.csv")

# Prepare data
swarm_data <- futout_clean %>%
  filter(
    Geography == "Canada",
    Gender == "Total",       # <<-- KEEP ONLY TOTAL
    !is.na(Number)
  ) %>%
  mutate(
    Indicators = factor(Indicators,
                        levels = c("Always", "Sometimes", "Rarely"))
  )

# Swarm plot using jitter
swarm_plot <- ggplot(
  swarm_data,
  aes(x = Indicators, y = Number, color = Indicators)
) +
  geom_jitter(width = 0.18, alpha = 0.7, size = 2) +

  scale_color_manual(values = c(
    "Always" = "#4E79A7",
    "Sometimes" = "#F28E2B",
    "Rarely" = "#E15759"
  )) +

  labs(
    title = "Respondent Count Distribution by Indicator",
    subtitle = "Swarm plot showing quarterly total respondent counts for optimism levels in Canada",
    x = "Indicator",
    y = "Number of respondents",
    color = "Indicator"
  ) +

  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(size = 18, face = "bold", color = "#C41E3A"),
    plot.subtitle = element_text(size = 11, color = "gray40"),
    axis.title.x = element_blank(),
    panel.grid.minor = element_blank()
  )

print(swarm_plot)

```



Appendix 3_Future Outlook

```r
# Slide 9: Future Outlook - Summary
rm(list = ls()); cat("\014")

library(dplyr)
library(ggplot2)
library(readr)
library(patchwork)
library(lubridate)

futout_clean <- read_csv("Future_Outlook_Clean.csv")

col_always    <- "#4E79A7"
col_sometimes <- "#F28E2B"
col_rarely    <- "#E15759"

# Unified Theme - all tiles use same title style
tile_theme <- theme_void() +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 10, hjust = 0.5),
    plot.margin = margin(10, 10, 10, 10)
  )

# TILE 1 – Canada (Always)
tile1_data <- futout_clean %>%
  filter(Geography == "Canada", Gender == "Total", Indicators == "Always") %>%
  summarise(Avg = round(mean(Percent), 1))

tile1 <- ggplot(tile1_data, aes(1, 1)) +
  geom_text(aes(label = paste0(Avg, "%")),
            size = 18, color = col_always, fontface = "bold") +
  labs(
    title = "Canada's Average Optimism",
    subtitle = "Most Canadians feel optimistic about their future."
  ) +
  tile_theme
```

```r
# TILE 2 – Gender comparison
tile2_data <- futout_clean %>%
  filter(Geography == "Canada", Indicators == "Always",
         Gender %in% c("Women", "Men")) %>%
  group_by(Gender) %>%
  summarise(Avg = round(mean(Percent), 1))

tile2 <- ggplot(tile2_data, aes(Gender, Avg, fill = Gender)) +
  geom_col(width = 0.55) +
  geom_text(aes(label = paste0(Avg, "%")),
            vjust = -0.2, size = 7, fontface = "bold") +

  scale_y_continuous(expand = expansion(mult = c(0, 0.15))) +
  scale_fill_manual(values = c("Women" = col_always, "Men" = col_rarely)) +

  labs(
    title = "Women More Optimistic",
    subtitle = "Women report slightly higher optimism than men."
  ) +

  theme_minimal(base_size = 10) +
  theme(
    legend.position = "none",
    axis.title = element_blank(),
    axis.text.y = element_blank(),

    plot.title = element_text(size = 14, face = "bold", hjust = 0.5,
                              margin = margin(t = 6, b = 2)),
    plot.subtitle = element_text(size = 10, hjust = 0.5,
                                 margin = margin(t = 0, b = 5)),

    plot.margin = margin(t = 5, r = 10, b = 5, l = 10)
  )
```

```r
# TILE 3 – Rarely (Low pessimism)
tile3_data <- futout_clean %>%
  filter(Geography == "Canada", Gender == "Total", Indicators == "Rarely") %>%
  summarise(Avg = round(mean(Percent), 1))

tile3 <- ggplot(tile3_data, aes(1, 1)) +
  geom_text(aes(label = paste0(Avg, "%")),
            size = 18, color = col_rarely, fontface = "bold") +
  labs(
    title = "Low Level of Pessimism",
    subtitle = "Only a small share of Canadians feel pessimistic."
  ) +
  tile_theme

# TILE 4 – Sometimes
tile4_data <- futout_clean %>%
  filter(Geography == "Canada", Gender == "Total", Indicators == "Sometimes") %>%
  summarise(Avg = round(mean(Percent), 1))

tile4 <- ggplot(tile4_data, aes(1, 1)) +
  geom_text(aes(label = paste0(Avg, "%")),
            size = 18, color = col_sometimes, fontface = "bold") +
  labs(
    title = "Moderate Optimism Stable",
    subtitle = "Around 30% consistently feel moderately optimistic."
  ) +
  tile_theme
```

Appendix 3_Future Outlook

```r
# TILE 5 – Top Provinces (Always)
tile5_data <- futout_clean %>%
  filter(Indicators == "Always", Gender == "Total", Geography != "Canada") %>%
  mutate(Geography = gsub("_", " ", Geography)) %>%
  group_by(Geography) %>%
  summarise(Avg = round(mean(Percent), 1)) %>%
  arrange(desc(Avg)) %>% slice(1:5)

tile5 <- ggplot(tile5_data, aes(reorder(Geography, Avg), Avg)) +
  geom_col(fill = col_always, width = 0.65) +
  geom_text(aes(label = Avg), hjust = -0.05,   # closer, less overhang
            size = 4, fontface = "bold") +
  coord_flip(clip = "off") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05))) +  # LESS empty right space
  labs(
    title = "Top Provinces by Optimism",
    subtitle = "Quebec leads Canada in optimism."
  ) +
  theme_minimal(base_size = 10) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.4),
    plot.subtitle = element_text(size = 10, hjust = 0.4),
    axis.title = element_blank(),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_blank(),
    plot.margin = margin(10, 10, 10, 10)
  )

```

```r
# TILE 6 – Sparkline (Trend)
tile6_data <- futout_clean %>%
  filter(Geography == "Canada", Gender == "Total", Indicators == "Always")

tile6 <- ggplot(tile6_data, aes(Date, Percent)) +
  geom_line(linewidth = 1.2, color = col_always) +
  geom_point(size = 2.5, color = col_always) +
  labs(
    title = "Optimism Over Time",
    subtitle = "Optimism fluctuates but remains high overall."
  ) +
  theme_minimal(base_size = 10) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 10, hjust = 0.5),
    axis.title = element_blank(),
    panel.grid.minor = element_blank()
  )

# FINAL LAYOUT
final_plot <- (tile1 | tile2 | tile3) /
  (tile4 | tile5 | tile6) +
  plot_annotation(
    title = "Future Outlook – Summary",
    subtitle = "Key national insights on optimism, gender differences, geography, and trends",
    theme = theme(
      plot.title = element_text(size = 24, face = "bold", hjust = 0),
      plot.subtitle = element_text(size = 14, hjust = 0)
    )
  )

print(final_plot)
```

Appendix 4_ Sense of Meaning and Purpose

```r
library(tidyverse)
library(readxl)
sense_raw <- read_excel("Final_Dataset_Morys (version 2).xlsx",
sheet = 1)
sense <- sense_raw %>%
mutate(
# keep period order as it appears in the file
Period = factor(Period, levels = unique(Period)),
# turn 0–5, 6–7, 8–10 into Low / Medium / High
Level = case_when(
str_detect(Indicators, "0-5")  ~ "Low",
str_detect(Indicators, "6-7")  ~ "Medium",
str_detect(Indicators, "8-10") ~ "High",
TRUE ~ Indicators
),
Level = factor(Level, levels = c("Low", "Medium", "High"))
)
latest_period <- sense %>%
summarise(latest = max(Period)) %>%
pull(latest)
latest_period
canada_latest <- sense %>%
filter(Geography == "Canada",
Gender == "Total",
Period == latest_period) %>%
group_by(Level) %>%
summarise(Percent = sum(Percent), .groups = "drop")
ggplot(canada_latest,
aes(x = Level, y = Percent, fill = Level)) +
geom_col() +
geom_text(aes(label = round(Percent, 1)),
vjust = -0.3) +
labs(
title = paste("Sense of Meaning and Purpose – Canada (", latest_period,
x = "Level",
y = "% of population"
) +
theme_minimal()
```

```r
canada_time <- sense %>%
filter(Geography == "Canada",
Gender == "Total") %>%
group_by(Period, Level) %>%
summarise(Percent = sum(Percent), .groups = "drop")
head(canada_time, 10)
ggplot(canada_time,
aes(x = Period, y = Percent, color = Level, group = Level)) +
geom_line(size = 1) +
geom_point() +
labs(
title = "Sense of Meaning and Purpose in Canada Over Time",
x = "Period",
y = "% of population",
color = "Level"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
ggplot(canada_time,
aes(x = Period, y = Percent, color = Level, group = Level)) +
geom_line(linewidth =  = 1) +
ggplot(canada_time,
aes(x = Period, y = Percent, color = Level, group = Level)) +
geom_line(linewidth = 1) +
geom_point() +
labs(
title = "Sense of Meaning and Purpose in Canada Over Time",
x = "Period",
y = "% of population",
color = "Level"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
prov_latest <- sense %>%
filter(Gender == "Total",
Period == latest_period) %>%
group_by(Geography, Level) %>%
summarise(Percent = sum(Percent), .groups = "drop")
# order provinces by High share
prov_order <- prov_latest %>%
filter(Level == "High") %>%
arrange(desc(Percent)) %>%
pull(Geography)
prov_latest <- prov_latest %>%
                                    ıy, levels = prov_order))
```

```r
ggplot(prov_latest,
aes(x = Geography, y = Percent, fill = Level)) +
geom_col() +
coord_flip() +
labs(
title = paste("Sense of Meaning and Purpose by Province (Total,", latest_period, ")"),
x = "Province",
y = "% of population",
fill = "Level"
) +
theme_minimal()
prov_gender <- sense %>%
filter(Period == latest_period,
Gender %in% c("Men", "Women")) %>%
group_by(Geography, Gender, Level) %>%
summarise(Percent = sum(Percent), .groups = "drop") %>%
mutate(Geography = factor(Geography, levels = prov_order))
prov_gender
ggplot(prov_gender,
aes(x = Geography, y = Percent, fill = Level)) +
geom_col(position = "dodge") +
coord_flip() +
facet_wrap(~ Gender) +
labs(
title = paste("Sense of Meaning and Purpose by Province & Gender (", latest_period, ")", sep = ""),
x = "Province",
y = "% of population",
fill = "Level"
) +
theme_minimal()
canada_gender <- sense %>%
filter(Geography == "Canada",
Gender %in% c("Men", "Women"),
Period == latest_period) %>%
group_by(Gender, Level) %>%
summarise(Percent = sum(Percent), .groups = "drop")
canada_gender <- sense %>%
filter(Geography == "Canada",
Gender %in% c("Men", "Women"),
Period == latest_period) %>%
group_by(Gender, Level) %>%
summarise(Percent = sum(Percent), .groups = "drop")
canada_gender
```

```r
ggplot(canada_gender,
aes(x = Gender, y = Percent, fill = Level)) +
geom_col(position = "fill") +
scale_y_continuous(labels = scales::percent) +
labs(
title = paste("Sense of Meaning and Purpose in Canada by Gender (", latest_period, ")", sep = ""),
x = "Gender",
y = "Share of population",
fill = "Level"
) +
theme_minimal()
summary_high <- sense %>%
filter(Gender == "Total",
Level == "High",
Period == latest_period) %>%
select(Geography, Percent, People) %>%
arrange(desc(Percent))
summary_high
write_csv(summary_high, "summary_high_meaning_by_province.csv")
```

Appendix 4_ Sense of Meaning and Purpose

```r
# Install once if needed
install.packages(c("tidyverse", "readxl"))
library(tidyverse)
library(readxl)
library(stringr)
file <- "Sense_socialDemo_data_BD.xlsx"
# Helper to clean any sheet
clean_sense_sheet <- function(df) {
df %>%
mutate(
Group = str_trim(Group),
`Sense of meaning and purpose rating` =
factor(`Sense of meaning and purpose rating`,
levels = c("Low", "Medium", "High")),
Period = factor(Period, levels = unique(Period))
)
}
all_group <- read_excel(file, sheet = "All Group") %>% clean_sense_sheet()
age       <- read_excel(file, sheet = "Age")       %>% clean_sense_sheet()
immigrant <- read_excel(file, sheet = "Immagrant") %>% clean_sense_sheet()
minority  <- read_excel(file, sheet = "Minority")  %>% clean_sense_sheet()
indig     <- read_excel(file, sheet = "Indigenous")%>% clean_sense_sheet()
disab     <- read_excel(file, sheet = "Disability")%>% clean_sense_sheet()
lgbtq     <- read_excel(file, sheet = "LGBTQ2+")   %>% clean_sense_sheet()
educ      <- read_excel(file, sheet = "Eduaction") %>% clean_sense_sheet()
activity  <- read_excel(file, sheet = "Activity")  %>% clean_sense_sheet()
area      <- read_excel(file, sheet = "Area")      %>% clean_sense_sheet()
overall_trend <- all_group %>%
filter(Group == "Total by all Ages") %>%         # whole sample baseline
group_by(Period, `Sense of meaning and purpose rating`) %>%
summarise(Percent = sum(Percent), .groups = "drop")
ggplot(overall_trend,
aes(x = Period,
y = Percent,
color = `Sense of meaning and purpose rating`,
group = `Sense of meaning and purpose rating`)) +
geom_line(size = 1) +
geom_point() +
labs(
title = "Sense of Meaning and Purpose – Overall Trend (18+)",
x = "Period",
y = "% of population",
color = "Rating"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
age_high <- age %>%
filter(Period == latest_period,
Group != "Total",
`Sense of meaning and purpose rating` == "High") %>%
group_by(Group) %>%
summarise(Percent = sum(Percent), .groups = "drop") %>%
arrange(Group)
ggplot(age_high,
aes(x = Group, y = Percent, fill = Group)) +
geom_col() +
labs(
title = paste("High Sense of Meaning and Purpose by Age Group (", latest_period, ")", se
x = "Age group",
y = "% with high sense of meaning & purpose"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
imm_latest <- immigrant %>%
filter(Period == latest_period,
Group != "Total") %>%
group_by(Group, `Sense of meaning and purpose rating`) %>%
summarise(Percent = sum(Percent), .groups = "drop")
ggplot(imm_latest,
aes(x = Group,
y = Percent,
fill = `Sense of meaning and purpose rating`)) +
geom_col() +
labs(
title = paste("Sense of Meaning and Purpose by Immigration Status (", latest_period, ")"
x = "Immigration status",
y = "% of group",
fill = "Rating"
) +
theme_minimal()
race_high <- minority %>%
filter(Period == latest_period,
Group != "Total",
`Sense of meaning and purpose rating` == "High") %>%
group_by(Group) %>%
summarise(Percent = sum(Percent), .groups = "drop") %>%
arrange(Percent)
ggplot(race_high,
aes(x = reorder(Group, Percent), y = Percent, fill = Group)) +
geom_col() +
coord_flip() +
labs(
title = paste("High Sense of Meaning and Purpose by Race / Visible Minority Group (", latest_period, ")", sep = ""),
x = "Race / visible minority group",
y = "% with high sense of meaning & purpose"
) +
theme_minimal()
indig_latest <- indig %>%
filter(Period == latest_period,
Group != "Total") %>%
mutate(Group = str_trim(Group)) %>%
group_by(Group, `Sense of meaning and purpose rating`) %>%
summarise(Percent = sum(Percent), .groups = "drop")
ggplot(indig_latest,
aes(x = Group,
y = Percent,
fill = `Sense of meaning and purpose rating`)) +
geom_col() +
labs(
title = paste("Sense of Meaning and Purpose by Indigenous Identity (", latest_period, ")",
x = "Group",
y = "% of group",
fill = "Rating"
) +
theme_minimal()
disab_latest <- disab %>%
filter(Period == latest_period,
Group != "Total") %>%
group_by(Group, `Sense of meaning and purpose rating`) %>%
summarise(Percent = sum(Percent), .groups = "drop")
ggplot(disab_latest,
aes(x = Group,
y = Percent,
fill = `Sense of meaning and purpose rating`)) +
geom_col() +
labs(
title = paste("Sense of Meaning and Purpose by Disability Status (", latest_period, ")", sep
x = "Disability status",
y = "% of group",
fill = "Rating"
) +
theme_minimal()
# Aggregate for latest period, exclude the category total
lgbtq_pie <- lgbtq %>%
filter(Period == latest_period,
Group != "Total") %>%
group_by(Group, `Sense of meaning and purpose rating`) %>%
summarise(Percent = sum(Percent), .groups = "drop")
ggplot(lgbtq_pie,
aes(x = "", y = Percent,
fill = `Sense of meaning and purpose rating`)) +
geom_col(width = 1) +
coord_polar(theta = "y") +
facet_wrap(~ Group) +
# labels = the rating text (Low / Medium / High)
geom_text(
aes(label = paste0(`Sense of meaning and purpose rating`, "\n", round(Percent, 1), "%")),
position = position_stack(vjust = 0.5),
size = 3
) +
labs(
title = paste("Sense of Meaning and Purpose by LGBTQ2+ Status (",
latest_period, ")", sep = ""),
fill  = "Rating"
) +
theme_void()
```