A2.2: Ethical Visualizations

Ahmed Raza

2025-02-20

1. Installing and Cleaning

```r
# Loading required libraries
library(tidyverse) # For ggplot, dplyr
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(countrycode) # For country code manipulation
library(scales)     # For nice axis labels
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```r
library(ggridges)
library(gghalves)
library(ggrepel)    # For non-overlapping labels
library(ggtext)     # For fancier text handling
library(forcats)
library(ggthemes) # For news pattern
library(rnaturalearth)
library(rnaturalearthdata)
```

```
##
```

```
## Attaching package: 'rnaturalearthdata'
##
## The following object is masked from 'package:rnaturalearth':
##
##     countries110
```

```r
library(sf)
```

```
## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```r
# Male Life Expectancy dataset
males <- read.csv("Males.csv", skip=4, stringsAsFactors=FALSE, check.names=TRUE)

# Female Life Expectancy dataset
females <- read.csv("Females.csv", skip=4, stringsAsFactors=FALSE, check.names=TRUE)

#Understanding data info
colnames(males)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "X1960"          "X1961"          "X1962"          "X1963"
##  [9] "X1964"          "X1965"          "X1966"          "X1967"
## [13] "X1968"          "X1969"          "X1970"          "X1971"
## [17] "X1972"          "X1973"          "X1974"          "X1975"
## [21] "X1976"          "X1977"          "X1978"          "X1979"
## [25] "X1980"          "X1981"          "X1982"          "X1983"
## [29] "X1984"          "X1985"          "X1986"          "X1987"
## [33] "X1988"          "X1989"          "X1990"          "X1991"
## [37] "X1992"          "X1993"          "X1994"          "X1995"
## [41] "X1996"          "X1997"          "X1998"          "X1999"
## [45] "X2000"          "X2001"          "X2002"          "X2003"
## [49] "X2004"          "X2005"          "X2006"          "X2007"
## [53] "X2008"          "X2009"          "X2010"          "X2011"
## [57] "X2012"          "X2013"          "X2014"          "X2015"
## [61] "X2016"          "X2017"          "X2018"          "X2019"
## [65] "X2020"          "X2021"          "X2022"          "X2023"
## [69] "X"
```

```r
colnames(females)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "X1960"          "X1961"          "X1962"          "X1963"
##  [9] "X1964"          "X1965"          "X1966"          "X1967"
## [13] "X1968"          "X1969"          "X1970"          "X1971"
## [17] "X1972"          "X1973"          "X1974"          "X1975"
## [21] "X1976"          "X1977"          "X1978"          "X1979"
## [25] "X1980"          "X1981"          "X1982"          "X1983"
## [29] "X1984"          "X1985"          "X1986"          "X1987"
## [33] "X1988"          "X1989"          "X1990"          "X1991"
## [37] "X1992"          "X1993"          "X1994"          "X1995"
## [41] "X1996"          "X1997"          "X1998"          "X1999"
## [45] "X2000"          "X2001"          "X2002"          "X2003"
```

```
## [49] "X2004"          "X2005"          "X2006"          "X2007"
## [53] "X2008"          "X2009"          "X2010"          "X2011"
## [57] "X2012"          "X2013"          "X2014"          "X2015"
## [61] "X2016"          "X2017"          "X2018"          "X2019"
## [65] "X2020"          "X2021"          "X2022"          "X2023"
## [69] "X"
```

Data Cleaning and Manipulation

```r
# Removing unnecessary columns
males <- males %>%
  select(-c(Indicator.Name, Indicator.Code, X, X2023))

females <- females %>%
  select(-c(Indicator.Name, Indicator.Code, X, X2023))

# Renaming columns name
colnames(males) <- gsub("^X", "", colnames(males))
colnames(females) <- gsub("^X", "", colnames(females))

# Reshape male data to long format
male <- males %>%
  pivot_longer(cols = starts_with("19") | starts_with("20"),  # Select all year columns
               names_to = "Year",
               values_to = "Life_Expectancy") %>%
  mutate(Year = as.numeric(Year),  # Convert Year to numeric
         Gender = "Male")  # Add gender identifier

# Reshape female data to long format
female <- females %>%
  pivot_longer(cols = starts_with("19") | starts_with("20"),
               names_to = "Year",
               values_to = "Life_Expectancy") %>%
  mutate(Year = as.numeric(Year),
         Gender = "Female")

# Combine male and female data for easier comparison
gender <- bind_rows(male, female)

# Check the reshaped data
colnames(gender)
```

```
## [1] "Country.Name"    "Country.Code"    "Year"            "Life_Expectancy"
## [5] "Gender"
```

Missing Values

```r
# Count missing values (NAs) per country
na_counts <- gender %>%
  group_by(Country.Name) %>%
```

```r
  summarise(NA_Count = sum(is.na(Life_Expectancy)),  # Count NAs in Life_Expectancy
            .groups = "drop") %>%
  arrange(desc(NA_Count))  # Sort by descending order of NA counts

# Countries with NA counts > 12 (Each country has 126 observation, so we will filter the countries with
countries_to_keep <- na_counts %>%
  filter(NA_Count <= 12) %>%
  pull(Country.Name)

#Filter the combined_long data to keep only countries with NA_Count <= 12
filtered_data <- gender %>%
  filter(Country.Name %in% countries_to_keep)

# Calculate the average life expectancy for each country (excluding NAs)
country_avg_life_exp <- filtered_data %>%
  group_by(Country.Name) %>%
  summarise(Avg_Life_Expectancy = mean(Life_Expectancy, na.rm = TRUE),
             .groups = "drop")

# Replace NAs in filtered_data with the country's average life expectancy
gender_data <- filtered_data %>%
  left_join(country_avg_life_exp, by = "Country.Name") %>%
  mutate(Life_Expectancy = ifelse(is.na(Life_Expectancy), Avg_Life_Expectancy, Life_Expectancy)) %>%
  select(-Avg_Life_Expectancy)  # Remove the temporary average column

print("Remaining NAs in Life_Expectancy:")
```

```
## [1] "Remaining NAs in Life_Expectancy:"
```

```r
sum(is.na(gender_data$Life_Expectancy))
```

```
## [1] 0
```

2. Standard Theme

```r
my_theme <- theme_minimal(base_family = "Helvetica", base_size = 11) +
  theme(panel.grid.minor = element_blank(),
        # Bold, bigger title
        plot.title = element_text(face = "bold", size = rel(1.7)),
        # Plain, slightly bigger subtitle that is grey
        plot.subtitle = element_text(face = "plain", size = rel(1.3), color = "grey70"),
        # Italic, smaller, grey caption that is left-aligned
        plot.caption = element_text(face = "italic", size = rel(0.7),
                                    color = "grey70", hjust = 0),
        # Bold legend titles
        legend.title = element_text(face = "bold"),
        # Bold, slightly larger facet titles that are left-aligned for the sake of repetition
        strip.text = element_text(face = "bold", size = rel(1.1), hjust = 0),
        # Bold axis titles
        axis.title = element_text(face = "bold"),
        # Add some space above the x-axis title and make it left-aligned
        axis.title.x = element_text(margin = margin(t = 10), hjust = 0),
```

```r
    # Add some space to the right of the y-axis title and make it top-aligned
    axis.title.y = element_text(margin = margin(r = 10), hjust = 1),
    # Add a light grey background to the facet titles, with no borders
    strip.background = element_rect(fill = "grey90", color = NA),
    # Add a thin grey border around all the plots to tie in the facet titles
    panel.border = element_rect(color = "grey90", fill = NA))
```

2. Team Blue - Ethical Visualization

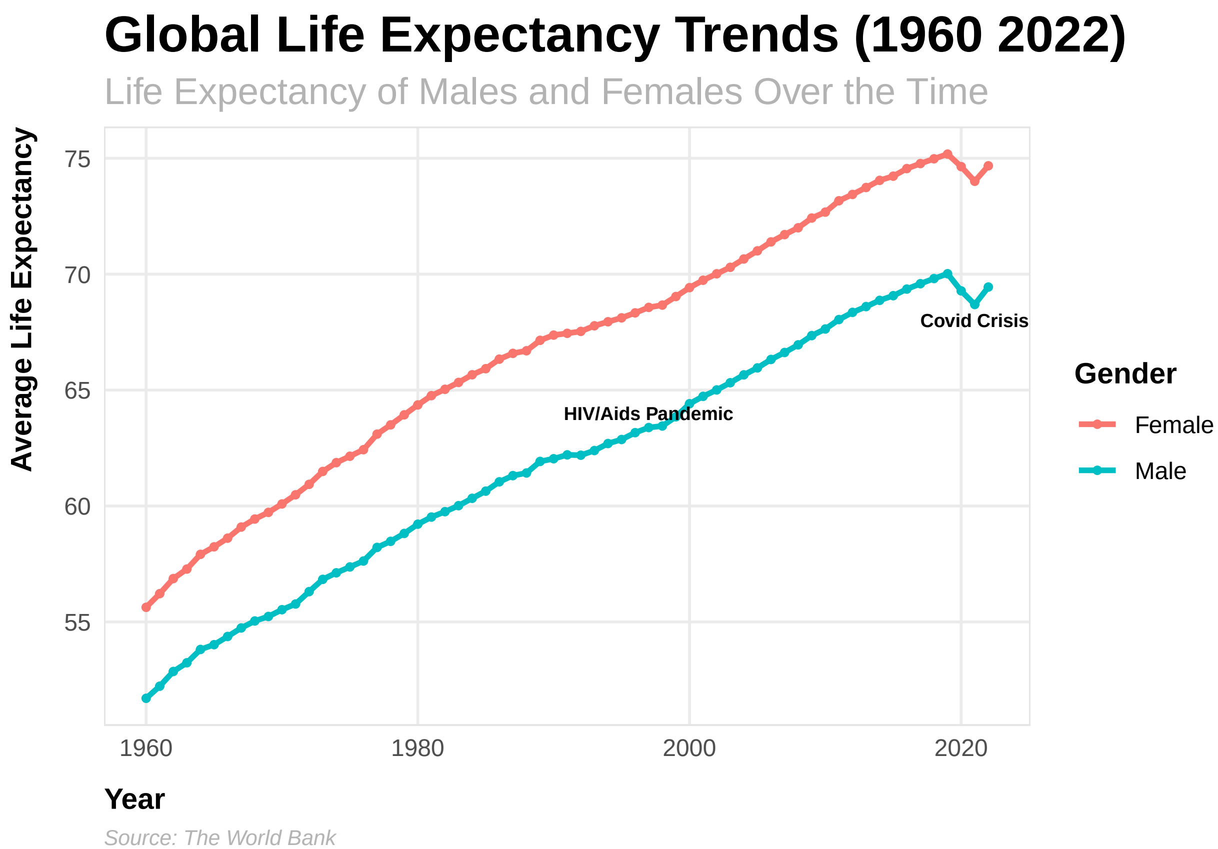
3. Average Life Expectancy Trends across the Globe

We used the basic rules of visualization to view this graph with the intention to sohwcase the average life expectancy across the globe over the past few decades. The data shwos the updward trends in the life expectancy with a few outliers that link with the global pandemics across those years. As we can see a sharp decline in the average life expectancy between 2020 and 2021 which is a time when we faced the deadly pandemics. As we deep dive the data share a quite interesting insights like females consistently outlive males across all the regions throughout these decades.

The data's reliability however is questionablewe also find anamolies in the data as it has a quite a high number of missing values and errors in the output. We used different approaches to clean the data and use ethical data manipulation techniques to make changes in the data structure.

```r
# Calculate global averages by year and gender
global_trends <- gender_data %>%
  group_by(Year, Gender) %>%
  summarise(Avg_Life_Expectancy = mean(Life_Expectancy, na.rm = TRUE),
            .groups = "drop")

# Create the line plot
ggplot(global_trends, aes(x = Year, y = Avg_Life_Expectancy, color = Gender, group = Gender)) +
  geom_line(size = 1) +
  geom_point(size = 1) +
  labs(title = "Global Life Expectancy Trends (1960 2022)",
       subtitle = "Life Expectancy of Males and Females Over the Time",
       x = "Year",
       y = "Average Life Expectancy",
       color = "Gender",
       caption = "Source: The World Bank") +
  my_theme +
  annotate("text", x = 2021, y = 68, label = "Covid Crisis", srt = 0,
           color = "black", size = 2.5, fontface = "bold") +
  annotate("text", x = 1997, y = 64, label = "HIV/Aids Pandemic", srt = 0,
           color = "black", size = 2.5, fontface = "bold")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

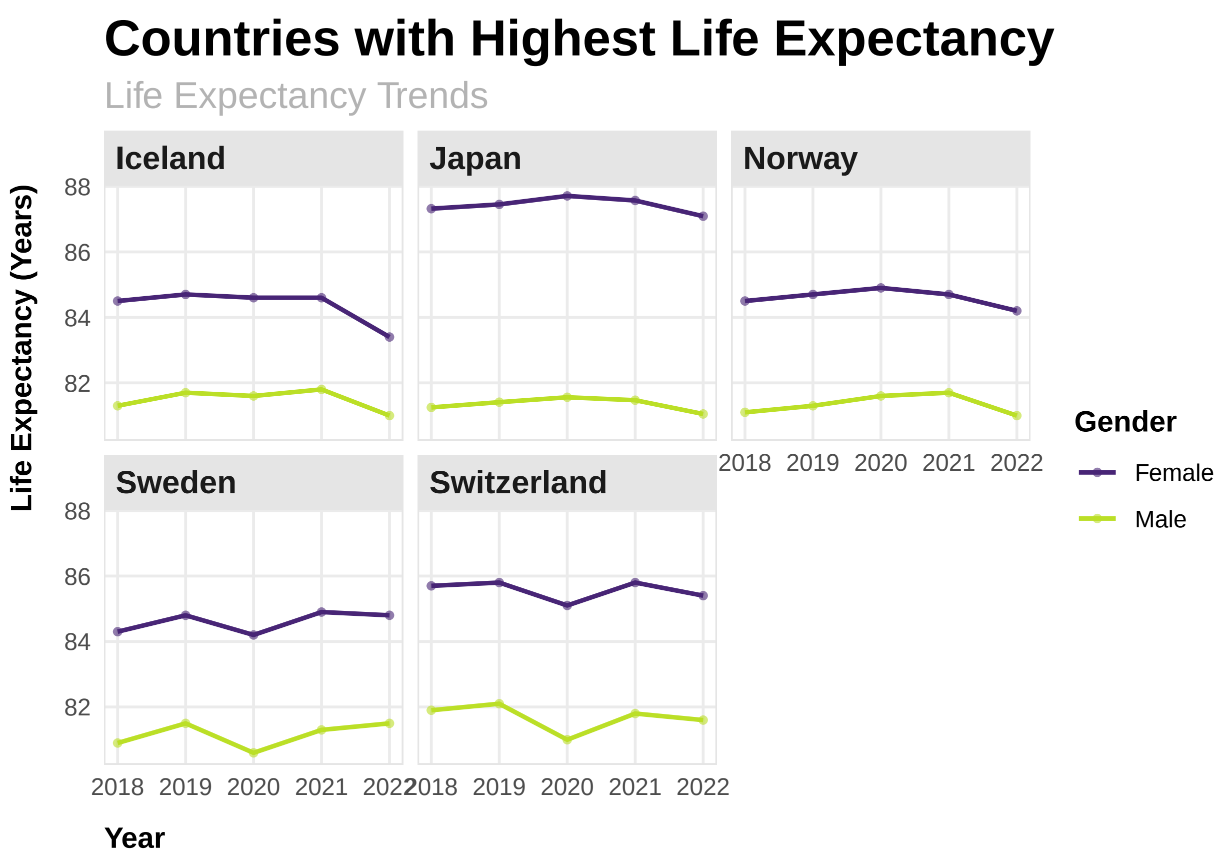


2. Countries with highest Life Expectancy over the Decade

This graph focuses on the countries with the continuos highest life expectancy. We reviewed the trends from 2018 to 2022 for Iceland, Japan, Sweden, and Switzerland, segmented by gender. The grpah also reveals that these conutries have mere or negligible decrease in the averagw life expectancy of the people sharp decline in the global average life expectancy because of the pandemic, reflecting strong healthcare system in the country and administration's effective response to manage the crisis. Iceland and Japan show stable female life expectancy around 86–88 years, while males hover at 82–84 years, with a dip in 2021. Sweden and Switzerland mirror this pattern, but male life expectancy drops more sharply, raising concerns about gender health disparities. We used a few data manipulation techniques to draw this visualization changing the data type for a few columns.

```
# Converting data type
cleaned_data <- gender_data %>%
  mutate(Year = as.numeric(as.character(Year)))

# Identify the top 5 countries by average life expectancy (across all years)
top_countries <- cleaned_data %>%
  group_by(Country.Name) %>%
  summarise(Avg_Life_Expectancy = mean(Life_Expectancy, na.rm = TRUE)) %>%
  arrange(desc(Avg_Life_Expectancy)) %>%
  slice_head(n = 5) %>%
  pull(Country.Name)

# Filter the data for the top 5 countries over the last 5 years
```

```r
top_countries_data <- cleaned_data %>%
  filter(Country.Name %in% top_countries,  # Filter for top 5 countries
         Year >= max(Year, na.rm = TRUE) - 4)  # Filter for last 5 years

# Create the line plot with CRAP Design Principles
ggplot(top_countries_data, aes(x = Year, y = Life_Expectancy, color = Gender)) +
  geom_line(size = 0.8) +  # Thicker lines for better contrast
  geom_point(size = 1, alpha = 0.6) +  # Add points for better readability
  facet_wrap(vars(Country.Name)) +
  scale_color_viridis_d(option = "D", begin = 0.1, end = 0.9) +  # Colorblind-friendly viridis palette
  labs(
    title = "Countries with Highest Life Expectancy",
    subtitle = "Life Expectancy Trends",
    x = "Year",
    y = "Life Expectancy (Years)",
    color = "Gender"  # Improve legend title
  ) +
  my_theme
```

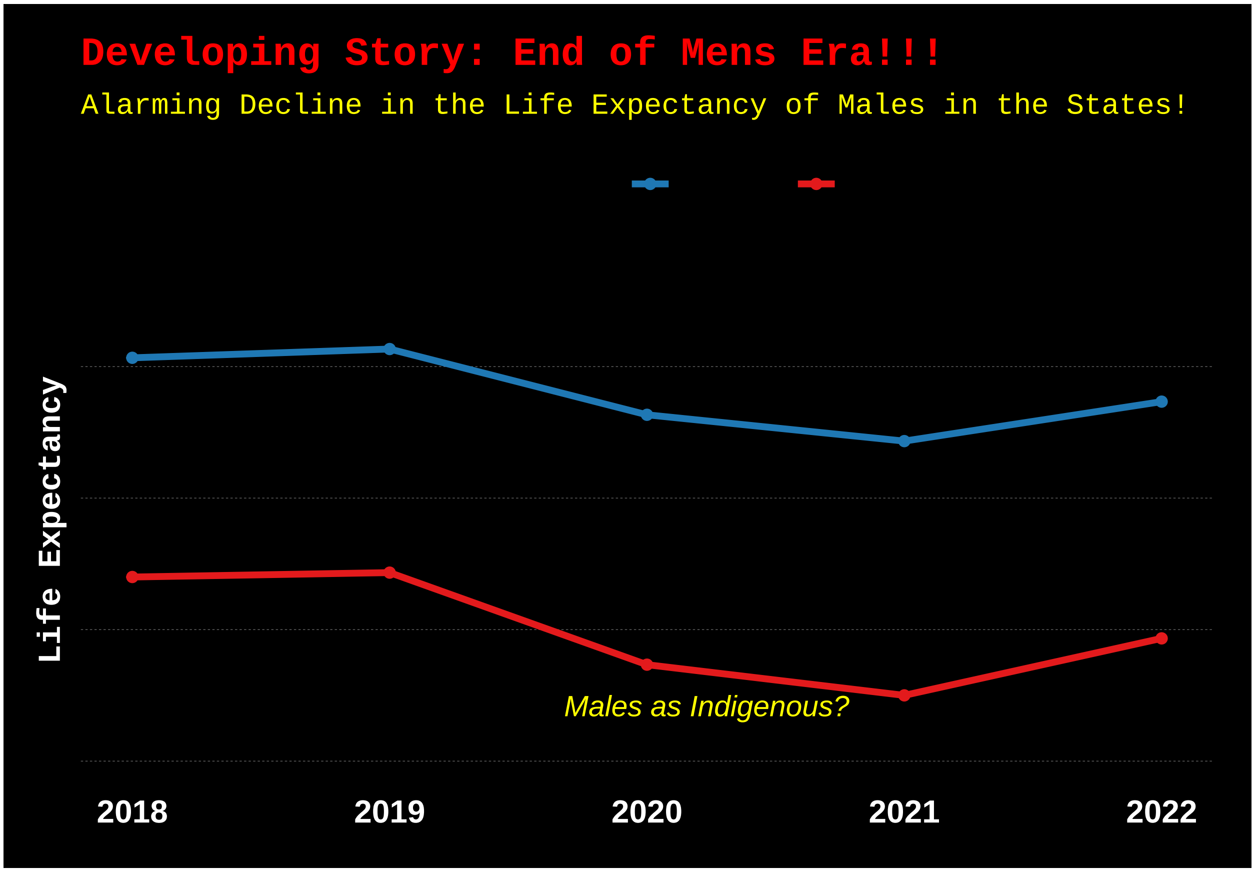


3. Team Red - Decpetive Use of Visualization Tools

4. End of Men's Era

We used deceptive tools to make this graph as it has a sensational title implying a severe crisis, and the reader has no awareness about the scale that we used to draw these conclusions. We created a filter data for the last 05 years knowing that there's decline in these years to support our false narratives. We influenced the reader's attention with the use of dark colors. We also highlighted the data point showing the lowest male's life expectancy and add an annotation to create a false belief. We have used a scale on Y-axis that just touches the beginning and end point of the data so even a small difference in the data scale could be perceived as steep decline/growth. The bright colors and the unexplained annotation "Males as Indigenous?" add confusion rather than clarity. There's also no data source or context about broader trends, so viewers can't assess if this drop is normal or anomalous.

```r
# Filter the data for United States
US_data <- gender_data %>%
  filter(Country.Name == "United States") %>%
  filter(!is.na(Life_Expectancy) & !is.na(Year)) %>%
  filter(Year >= max(Year) - 4) %>% # most recent 5 years data
  select(Year, Life_Expectancy, Gender)

# Convert Year to integer
US_data <- US_data %>%
  mutate(Year = as.integer(Year),
         Life_Expectancy = as.double(Life_Expectancy))  # Convert Life_Expectancy to double

# Create the line plot
ggplot(US_data, aes(x = Year, y = Life_Expectancy, color = Gender)) +
  geom_line(linewidth = 1.2) +  # Thicker lines for clarity
  geom_point(size = 1.5) +  # Add points for readability
  labs(title = "<span style='color:red; font-weight:bold; font-size:20px'>Developing Story: End of Mens
       subtitle = "<span style='color:yellow; font-weight:bold; font-style:italic'>Alarming Decline in 
       x = "Year",
       y = "Life Expectancy",
       color = "Gender") +
  scale_color_manual(values = c("Male" = "#E31A1C", "Female" = "#1F78B4")) +
  scale_y_continuous(limits = c(72, 83),  # Set y-axis range
                     labels = NULL) +  # No y-axis text
  theme_wsj(base_size = 12, color = "white") +  # Use Wall Street Journal theme as a base
  theme(
    plot.background = element_rect(fill = "#000000", color = NA),  # Black background
    panel.background = element_rect(fill = "#000000", color = NA),
    plot.title = element_markdown(lineheight = 1.2, hjust = 0, margin = margin(b = 10),
    size = 13),
plot.subtitle = element_markdown(lineheight = 1.2, hjust = 0, margin = margin(b = 10),
    size = 11),
    axis.text.x = element_text(color = "white", size = 12),
    axis.text.y = element_blank(),
    axis.title.y = element_text(color = "white", size = 12, face = "bold"),
    axis.title.x = element_blank(),
    panel.grid.major = element_line(color = "gray50", linetype = "dashed", size = 0.1),
    panel.grid.minor = element_blank()) +
    geom_text_repel(data = data.frame(Year = 2020.5, Avg_Life_Expectancy = 74,
    Gender = "Male", label = "Males as Indigenous?"),
    aes(x = Year, y = Avg_Life_Expectancy, label = label),
    color = "yellow", size = 4, fontface = "italic", box.padding = 0.5,
    point.padding = 1, max.overlaps = Inf, segment.color = "red",
    segment.size = 1, segment.alpha = 0.8, force = 1, direction = "both")
```

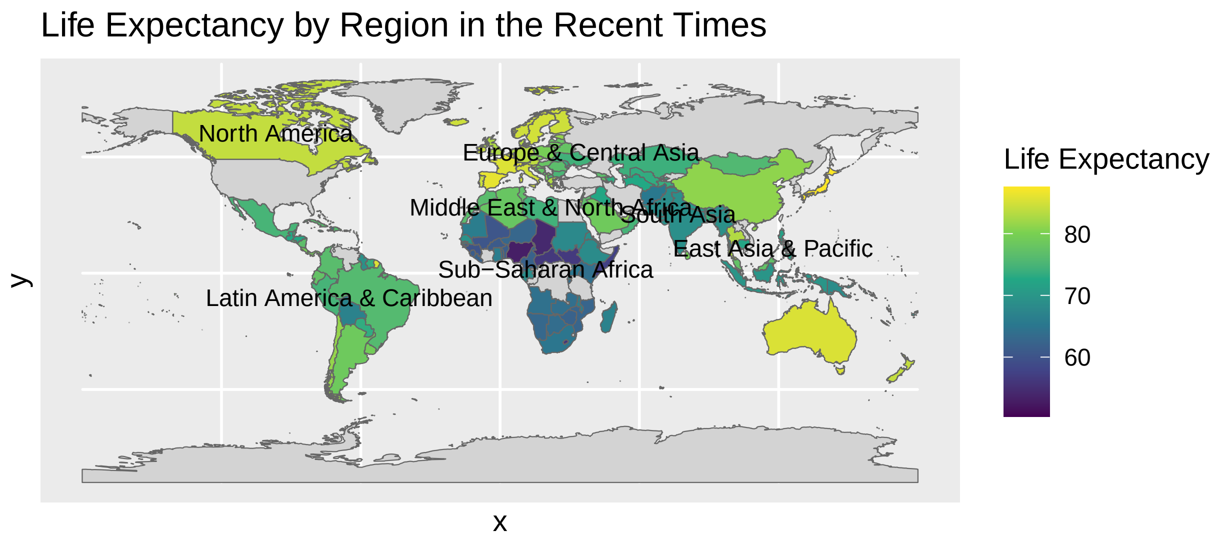


4. Is the concept of a longer life expectancy a fallacy!

It's another example of using unethical tools to decieve the audience. The title of the grpah is deceiving as it says the average life expectancy across the recent times. However, there's no information what does that recent mean as we only used the data for 2021. We intentionally chose this data to show the lowest average life expectancy supporting our false narratives. Further, there's no visible legend scale increments, making it tough to interpret specific differences. The label 'x' on the horizontal axis is unclear, implying the projection system is missing. Regions like 'North America' and 'Latin America & Caribbean' overlap with actual countries, but those country outlines are not drawn distinctly, causing confusion. We also assigned a wrong pattern showing the countries and region with the highest average life expectancy highlighted in the light colors and vice versa.

```
df <- cleaned_data %>%
  mutate(
    Region = countrycode(
      sourcevar   = Country.Name, origin      = "country.name",
      destination = "region")) %>%
  mutate(
    Region = ifelse(is.na(Region), "Others", Region))

# Filter for a single year
df_2021 <- df %>%
  filter(Year == 2021)

# Get a world polygon layer from rnaturalearth
```

```r
world_map <- rnaturalearth::ne_countries(scale = "medium", returnclass = "sf") %>%
  select(iso_a3, admin, geometry) %>%

# We'll rename to allow joining by 'admin' which typically matches "Country.Name"
rename(Country.Name = admin)

# Join with our life expectancy data
# Note: Some countries may not match exactly and will appear as NA or be lost.
world_merged <- world_map %>%
  left_join(df_2021, by = "Country.Name")

# Aggregate by Region
region_polygons <- world_merged %>%
  group_by(Region) %>%
  summarise(geometry = st_union(geometry), .groups = "drop")

# Compute centroids for these region-level polygons
region_centroids <- region_polygons %>%
  mutate(centroid = st_centroid(geometry))

# Create the map with warnings suppressed
suppressWarnings(
  ggplot(data = world_merged)) +

# Color countries by Life Expectancy
geom_sf(aes(fill = Life_Expectancy), color = "gray40", size = 0.2) +

# Add region labels once per region at the centroid
geom_sf_text(
  data = region_centroids,
  aes(geometry = centroid, label = Region), color = "black", size = 3) +

# Continuous color scale for Life Expectancy
  scale_fill_viridis_c(
  option = "viridis",
  name = "Life Expectancy",
  na.value = "lightgray") +
  labs(
  title = "Life Expectancy by Region in the Recent Times",
  subtitle = "Is the concept of a longer life a fallacy!" +
theme_minimal() +
      theme(plot.title = element_text(face = "bold", size = 14),
            plot.subtitle = element_text(size = 10, color = "gray50"),
            axis.title.y = element_blank(),
            axis.title.x = element_blank())
)
```



Life Expectancy by Region in the Recent Times
North America
Europe & Central Asia
Middle East & North Africa
South Asia
East Asia & Pacific
Sub−Saharan Africa
Latin America & Caribbean
y
x
Life Expectancy
80
70
60