Reproducibility Exercise

This code organizes the data from CSV files into a data frame and creates new columns

Load the required package for reading data

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

1) Read the data into data frame

```
data1 <- read_csv('inputs/gdp_52-67.csv')
```

```
## Rows: 568 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): continent, country
## dbl (3): year, pop, gdp
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data2 <- read_csv('inputs/le_52-67.csv')
```

```
## Rows: 568 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): continent, country
## dbl (2): year, lifeExp
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

2) Merge data frames based on continent, country, and year

```
merged_df <- merge(data1, data2, by = c("continent", "country", "year"))
stargazer(merged_df, type = "text")
```

```
##
## ================================================================================================
## Statistic  N         Mean             St. Dev.            Min                 Max
## ------------------------------------------------------------------------------------------------
## year       568     1,959.500             5.595            1,952               1,967
## pop        568   19,698,280.000     68,188,340.000        60,011           754,550,000
## gdp        568 72,511,306,869.000 274,437,023,827.000 52,784,691.000 3,880,918,003,146.000
## lifeExp    568       52.463             12.287            28.801              74.160
## ------------------------------------------------------------------------------------------------
```

3) Create a new column with GDP per capita

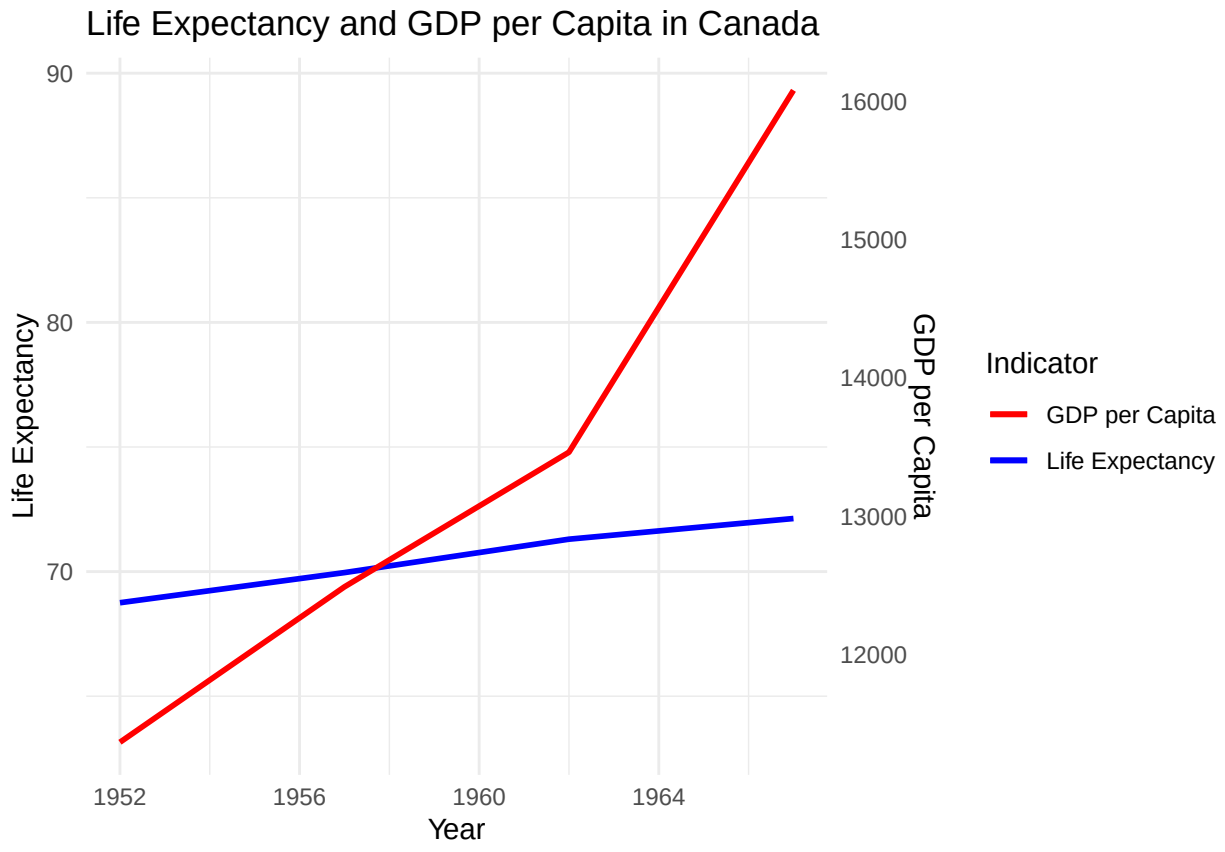
```
merged_df$gdpPercap <- merged_df$gdp / merged_df$pop
```

4) Visualize life expectancy and gdp per capita over time for Canada in the 1950s and 1960s using a line plot

```
# Filter the data for Canada
canada_data <- merged_df[merged_df$country == "Canada", ]

# Create a line plot with two separate vertical axes
ggplot(data = canada_data, aes(x = year)) +
  geom_line(aes(y = lifeExp, color = "Life Expectancy"), size = 1) +
  geom_line(aes(y = gdpPercap / 180, color = "GDP per Capita"), size = 1) +

  # Primary y-axis for lifeExp
  scale_y_continuous(
    name = "Life Expectancy",
    sec.axis = sec_axis(trans=~.*180, name = "GDP per Capita")
  ) +

  # Customize colors
  scale_color_manual(values = c("Life Expectancy" = "blue", "GDP per Capita" = "red")) +

  # Set labels and title
  labs(
    x = "Year",
    y = "Life Expectancy",
    color = "Indicator",
    title = "Life Expectancy and GDP per Capita in Canada"
  ) +

  # Customize theme
  theme_minimal()
```

```
## 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.
```



5) Regress life expectancy on gdp per capita and display the regression table

```
# Fit a linear regression model
model <- lm(lifeExp ~ gdpPercap, data = merged_df)

# Display the regression table
stargazer(model, type = "text", style = "aer")
```

```
##
## ===============================================================
##                                        lifeExp
## ---------------------------------------------------------------
## gdpPercap                             0.000***
##                                       (0.000)
##
## Constant                             50.225***
##                                       (0.540)
##
## Observations                            568
## R2                                     0.130
```

```
## Adjusted R2                          0.128
## Residual Std. Error            11.472 (df = 566)
## F Statistic                 84.367*** (df = 1; 566)
## ------------------------------------------------------------
## Notes:               ***Significant at the 1 percent level.
##                      **Significant at the 5 percent level.
##                      *Significant at the 10 percent level.
```

Moving the PDF file to the output directory

```
#markdown::render("reproducibility_exercise.Rmd", output_file = "/output/reproducibility_exercise.pdf")
```