# An Analysis of Socio-Economic Factors in Economic Growth Theory

## Assignment 2 - ETO5513

Cameron Picton & Gustav McRae-Johns

2025-08-11

# *Introduction*

Neo-classical economic growth models (such as the Cobbs-Douglas production function) suggest that GDP can be forecasted using a linear combination of national technology and capital levels, and labour growth (Dykas, 2023). Within the scope of "capital", endogenous growth theory suggests that traditional physical capital has given way to include both human capital (Martin, 1998) and social infrastructure as essential ingredients to a nations' economic potential (Saygılı, 2021). Specifically education directly effects human capital - Hanushek & Woessmann (2010) argue that investment in education leads to long-term productivity growth -, and factors such as environmental protections and healthcare access contribute to social infrastructure (Trinh, 2023) (Saygılı, 2021). Emerging from the traditional view of labour growth, gender diversity and female engagement in the workforce directly impacts labour growth rates and positively impacts on output growth rates (Dahlum, 2022). Similarly, the role of industry has been emphasised in growth theory, particularly through the work of Kaldor (1966) and Kuznets (1966), who linked industrialisation to structural transformation and economic development. This report will thus discuss and analyse the four questions:

1. Is there a relationship between education expenditure and GDP per capita?
2. How does the proportion of employment in industry relate to GDP per capita?
3. Does greater access to healthcare positively impact on economic growth?
4. Does a greater female workforce increase economic growth?

Using cross-country data, we aim to explore these relationships.

---

# *Data Set Introduction*

The data being used to investigate this problem records 96 variables across 66 nations for the year 2017. The data identifies 93 quantitative socio-economic variables, 2 categorical variables (nation & region), and a country code variable. The data is contained within one comma delimited file and was downloaded from:

https://www.kaggle.com/datasets/nishanthsalian/socioeconomic-country-profiles/data (https://www.kaggle.com/datasets/nishanthsalian/socioeconomic-country-profiles/data)

```
# read in data file
Socio_Econ <- as_tibble(read.csv("soci_econ_country_profiles.csv"))
Socio_Econ %>%
  head(n=10) %>%
  kable(caption = "Socio-economic variables for 66 Nations, 2017") %>%
  kable_styling(bootstrap_options = "striped", full_width = FALSE)
```

Socio-economic variables for 66 Nations, 2017

| X | country | Region | Surface.area..km2. | Population.in.thousands..2017. | Population.density..per.km2..2017. | Sex.ratio..m.per.100. |
|---|---|---|---|---|---|---|
| 0 | Argentina | SouthAmerica | 2780400 | 44271 | 16.2 | |
| 1 | Australia | Oceania | 7692060 | 24451 | 3.2 | |
| 2 | Austria | WesternEurope | 83871 | 8736 | 106.0 | |
| 3 | Belarus | EasternEurope | 207600 | 9468 | 46.7 | |
| 4 | Belgium | WesternEurope | 30528 | 11429 | 377.5 | |
| 5 | Bosnia and Herzegovina | SouthernEurope | 51209 | 3507 | 68.8 | |
| 6 | Brazil | SouthAmerica | 8515767 | 209288 | 25.0 | |
| 7 | Bulgaria | EasternEurope | 111002 | 7085 | 65.3 | |
| 8 | Canada | NorthernAmerica | 9984670 | 36624 | 4.0 | |
| 9 | Chile | SouthAmerica | 756102 | 18055 | 24.3 | |

---

# *Analysis - Education & Industry Employment*

The first part of this analysis will focus on how education - as human capital - and industry employment - as a traditional determinant of GDP - impact economic growth.

# *Data Exploration*

```
#load data
Econ_data <- read_csv("soci_econ_country_profiles.csv")

# Clean and rename variables
Econ_clean <- Econ_data %>%
  select(
    Country = country,
    GDP_per_Capita = `GDP per capita (current US$)`,
    Education_Expenditure = `Education: Government expenditure (% of GDP)`,
    Industry_Employment = `Employment: Industry (% of employed)`,
  ) %>%
  drop_na() %>%
  filter(
    GDP_per_Capita > 0,
    Education_Expenditure > 0,
    Industry_Employment > 0,
  )
```

```
# GDP per Capita
Econ_clean %>%
  summarise(
    Mean = mean(GDP_per_Capita, na.rm = TRUE),
    SD   = sd(GDP_per_Capita, na.rm = TRUE),
    Min  = min(GDP_per_Capita, na.rm = TRUE),
    Max  = max(GDP_per_Capita, na.rm = TRUE)
  ) %>%
  mutate(across(everything(), as.numeric)) %>%
  pivot_longer(
    cols = everything(),
    names_to = "Statistic",
    values_to = "Value"
  ) %>%
  kable(
    digits = 2,
    caption = "GDP per Capita — Descriptive Statistics"
  )
```

GDP per Capita — Descriptive Statistics

| Statistic | Value |
|---|---:|
| Mean | 25617.42 |
| SD | 21182.08 |
| Min | 1410.40 |
| Max | 80831.10 |

```
# Education Expenditure
Econ_clean %>%
  summarise(
    Mean = mean(Education_Expenditure, na.rm = TRUE),
    SD   = sd(Education_Expenditure, na.rm = TRUE),
    Min  = min(Education_Expenditure, na.rm = TRUE),
    Max  = max(Education_Expenditure, na.rm = TRUE)
  ) %>%
  mutate(across(everything(), as.numeric)) %>%
  pivot_longer(
    cols = everything(),
    names_to = "Statistic",
    values_to = "Value"
  ) %>%
  kable(
    digits = 2,
    caption = "Education Expenditure (% of GDP) — Descriptive Statistics"
  )
```

Education Expenditure (% of GDP) — Descriptive Statistics

| Statistic | Value |
|---|---:|
| Mean | NA |
| SD | 1.29 |
| Min | 2.20 |
| Max | 8.60 |

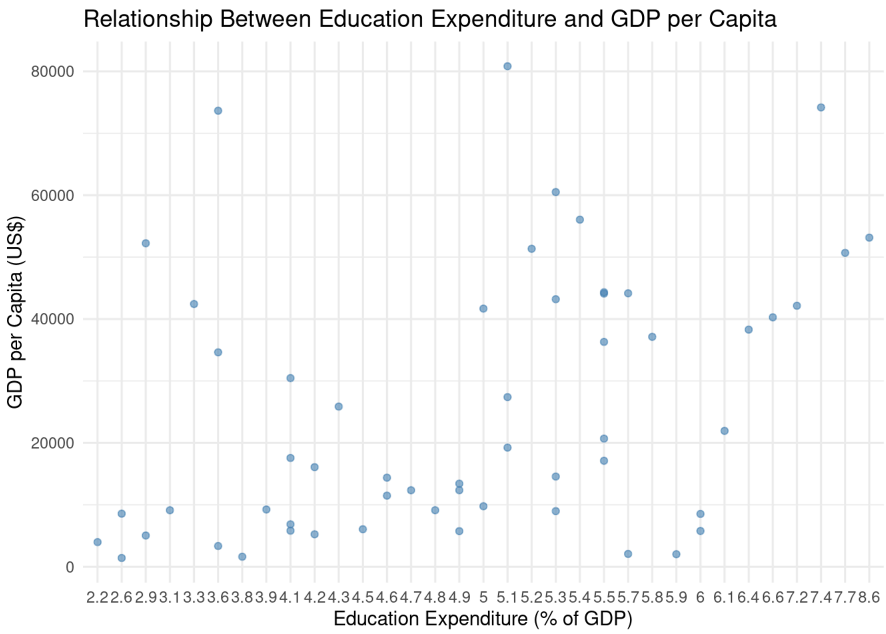
```
# Industry Employment
Econ_clean %>%
  summarise(
    Mean = mean(Industry_Employment, na.rm = TRUE),
    SD   = sd(Industry_Employment, na.rm = TRUE),
    Min  = min(Industry_Employment, na.rm = TRUE),
    Max  = max(Industry_Employment, na.rm = TRUE)
  ) %>%
  mutate(across(everything(), as.numeric)) %>%
  pivot_longer(
    cols = everything(),
    names_to = "Statistic",
    values_to = "Value"
  ) %>%
  kable(
    digits = 2,
    caption = "Industry Employment (% of Employed) — Descriptive Statistics"
  )
```

Industry Employment (% of Employed) — Descriptive Statistics

| Statistic | Value |
|---|---:|
| Mean | 24.27 |
| SD | 6.36 |
| Min | 14.90 |
| Max | 54.10 |

# *Analysis: Education Expenditure and GDP per Capita*

```
ggplot(Econ_clean, aes(x = Education_Expenditure, y = GDP_per_Capita)) +
  geom_point(alpha = 0.6, colour = "steelblue") +
  geom_smooth(method = "lm", se = TRUE, colour = "darkred") +
  labs(
    title = "Relationship Between Education Expenditure and GDP per Capita",
    x = "Education Expenditure (% of GDP)",
    y = "GDP per Capita (US$)"
  ) +
  theme_minimal()
```




# *Discussion of Education on GDP*

The scatter plot shows the relationship between education expenditure as a percentage of GDP and GDP per capita across countries. While a few high-income countries do invest more in education (6–8%), the overall trend appears weak and diffuse. Countries with similar education spending (e.g., 4–6% of GDP) show vastly different economic outcomes, with GDP per capita ranging from under $5,000 to over $80,000.

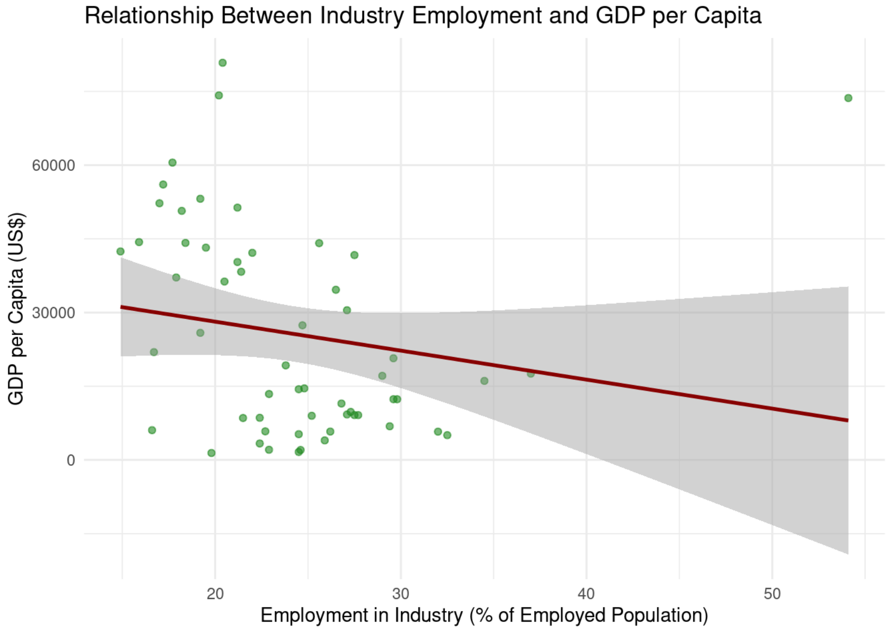
This pattern suggests that education expenditure alone is not a strong predictor of national income. Although economic theory (2010) supports the idea that investment in education fosters long-term growth, this may depend on additional factors such as education quality, governance, efficiency of spending, or the stage of economic development. Some low-income countries may invest in education but face barriers in translating that investment into productivity gains.

Overall, the findings highlight that while education is an important component of human capital, expenditure as a share of GDP may not capture the full picture. Future analysis could benefit from including more nuanced variables—such as test scores, enrolment rates, or lagged effects—to better reflect education's contribution to growth.

---

# *Analysis: Industry Employment and GDP per Capita*

```
ggplot(Econ_clean, aes(x = Industry_Employment, y = GDP_per_Capita)) +
  geom_point(alpha = 0.6, colour = "forestgreen") +
  geom_smooth(method = "lm", se = TRUE, colour = "darkred") +
  labs(
    title = "Relationship Between Industry Employment and GDP per Capita",
    x = "Employment in Industry (% of Employed Population)",
    y = "GDP per Capita (US$)"
  ) +
  theme_minimal()
```




---

# Discussion of Industry Employment on GDP

The scatter plot examining the relationship between industry employment share and GDP per capita reveals a slightly negative linear trend. Countries with a higher proportion of their workforce employed in industry tend to have lower GDP per capita, while countries with higher incomes tend to show lower industrial employment, typically between 15–25%.

This finding may seem counterintuitive at first but is well supported by structural transformation theory (1966). In the early stages of development, countries tend to shift workers from agriculture into industry, leading to productivity gains. However, as countries continue to develop, their economies tend to move toward the service sector, where value-added and innovation often drive further income growth. As a result, post-industrial economies may show both lower industrial employment and higher GDP per capita.

The slight negative trend also reflects the current global economic landscape, where advanced economies have largely deindustrialised and rely more on services and knowledge-based industries. In contrast, many middle- and low-income countries still depend on industrial jobs but may lack the infrastructure or productivity to translate that into sustained income growth.

These results suggest that industrial employment is a development stage indicator, not a direct driver of GDP per capita in high-income contexts.

---

# *Analysis - Healthcare Access & Gender Diversity*

The second part of this analysis will focus on how healthcare access - as social capital - and female workforce engagement - as labour - impact economic growth.

## *Data Exploration*

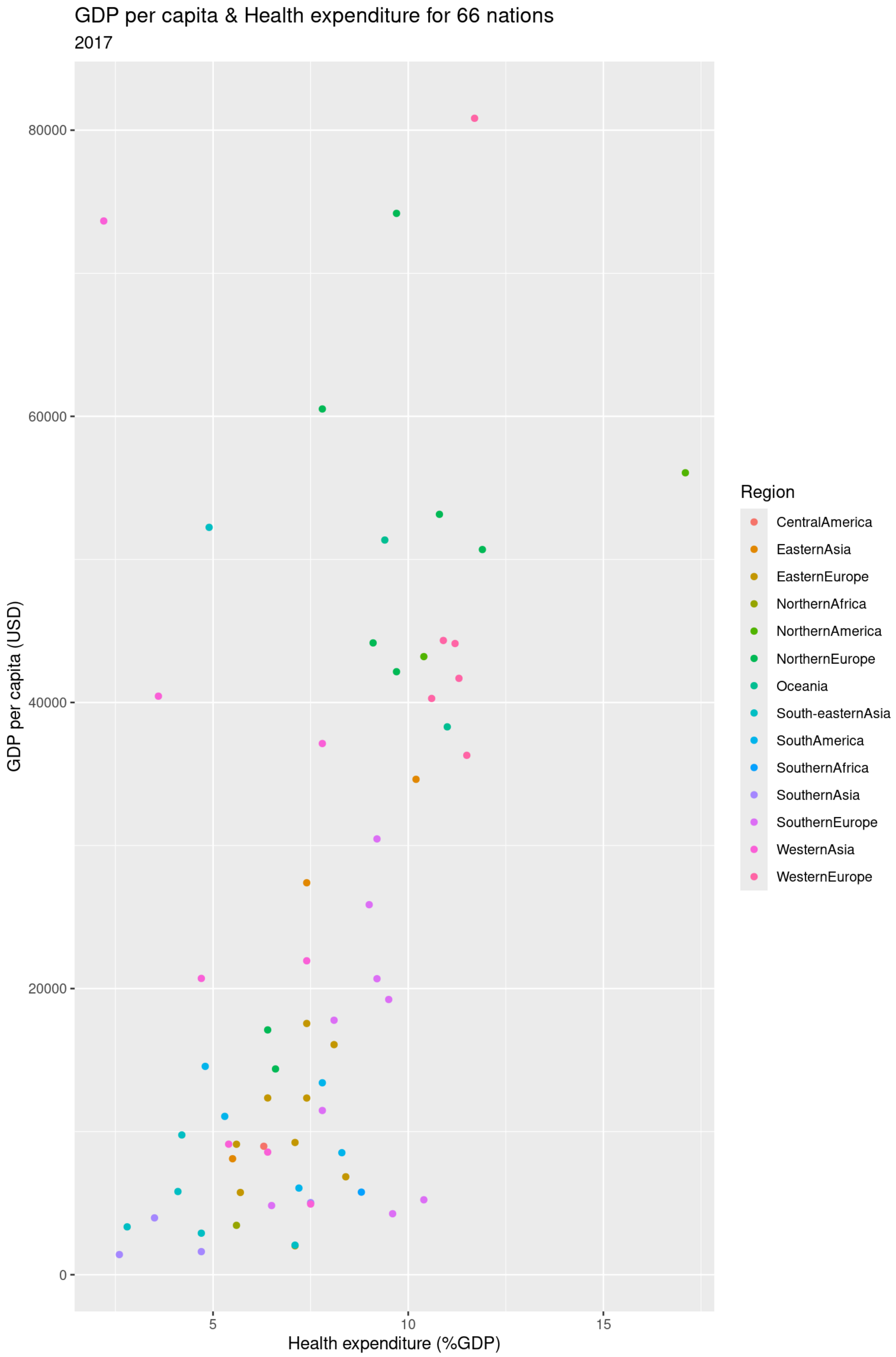
Each question will focus on two dependent variables and will use GDP per capita (USD) to represent economic growth. Healthcare access will use healthcare expenditure as a percentage of GDP and life expectancy at birth to represent a nations investment in healthcare access. Gender diversity will use labour force participation and the number of seats women hold in parliament as measures of female engagement in the economy.

```
Analysis_df <- Socio_Econ %>%
  select("country", "Region", "GDP.per.capita..current.US..", "Health..Total.expenditure....of.GDP.", "Life.expectancy.at.bi
rth..total..years.", "Labour.force.participation..female.male.pop....", "Seats.held.by.women.in.national.parliaments..") %>%
  rename(Nation = "country", "GDP per capita (USD)" = GDP.per.capita..current.US.., "Health expenditure (%GDP)" = Health..To
tal.expenditure....of.GDP., "Life expectancy (Years)" = Life.expectancy.at.birth..total..years., "Labour force participation
(%)" = Labour.force.participation..female.male.pop...., "Parliament seats held by women" = Seats.held.by.women.in.national.p
arliaments..) %>%
  separate("Labour force participation (%)", into = c("Female labour participation (%)", "Male labour force"), convert = TRU
E) %>%
  select(-("Male labour force"))
Analysis_df
```

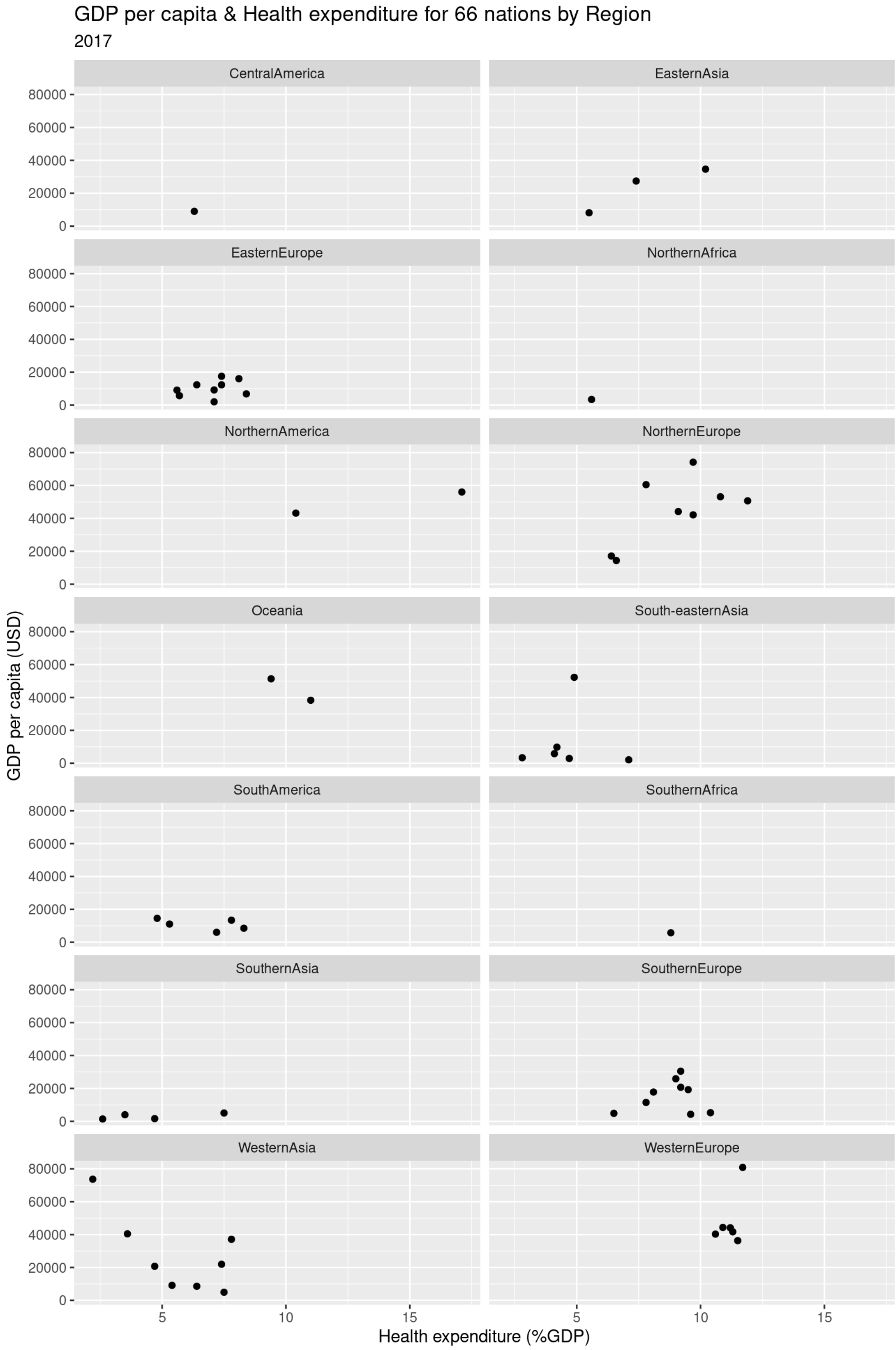
```
## # A tibble: 66 × 7
##    Nation                 Region   `GDP per capita (USD)` Health expenditure (…¹
##    <chr>                  <chr>                     <dbl>                  <dbl>
##  1 Argentina              SouthAm…                 14564.                    4.8
##  2 Australia              Oceania                  51352.                    9.4
##  3 Austria                Western…                 44118.                   11.2
##  4 Belarus                Eastern…                  5751.                    5.7
##  5 Belgium                Western…                 40278.                   10.6
##  6 Bosnia and Herzegovina Souther…                  4265                     9.6
##  7 Brazil                 SouthAm…                  8528.                    8.3
##  8 Bulgaria               Eastern…                  6847.                    8.4
##  9 Canada                 Norther…                 43206.                   10.4
## 10 Chile                  SouthAm…                 13416.                    7.8
## # ℹ 56 more rows
## # ℹ abbreviated name: ¹`Health expenditure (%GDP)`
## # ℹ 3 more variables: `Life expectancy (Years)` <dbl>,
## #   `Female labour participation (%)` <int>,
## #   `Parliament seats held by women` <dbl>
```

## *Analysis: Healthcare Access*

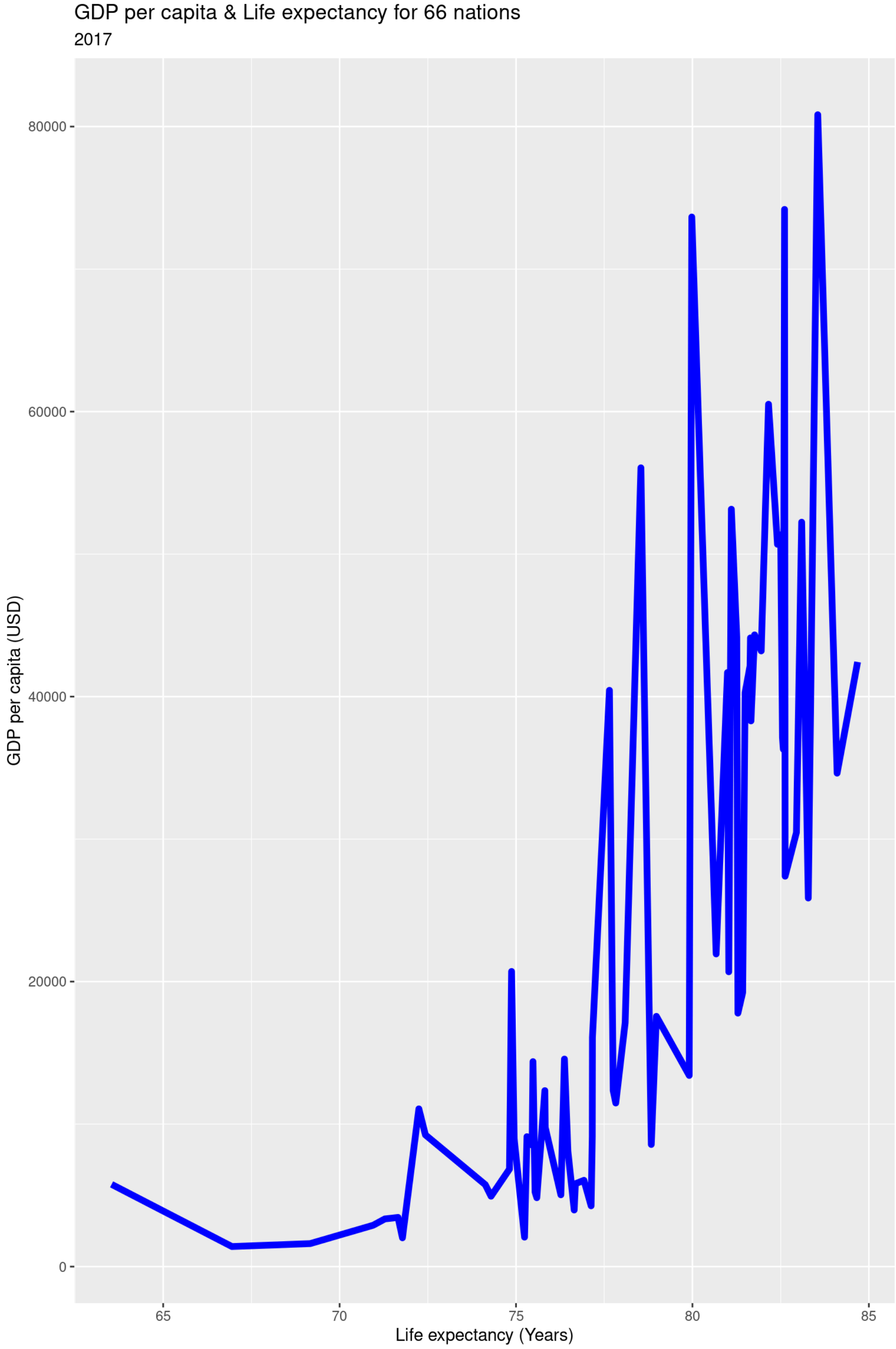
```
# The analysis for health expenditure may be distorted by values which are below 0%. As this is non-sensical, rows containin
g values less then 0 are removed.
Health_df <- Analysis_df %>%
  filter(`Health expenditure (%GDP)` >= 0)

ggplot(data = Health_df) +
  geom_point(mapping = aes(x = `Health expenditure (%GDP)`, y = `GDP per capita (USD)`, colour = Region)) +
  labs(
    title = "GDP per capita & Health expenditure for 66 nations",
    subtitle = "2017")
```




```
ggplot(data = Health_df) +
  geom_point(mapping = aes(x = `Health expenditure (%GDP)`, y = `GDP per capita (USD)`)) +
  facet_wrap(~ Region, nrow = 7) +
  labs(
    title = "GDP per capita & Health expenditure for 66 nations by Region",
    subtitle = "2017")
```

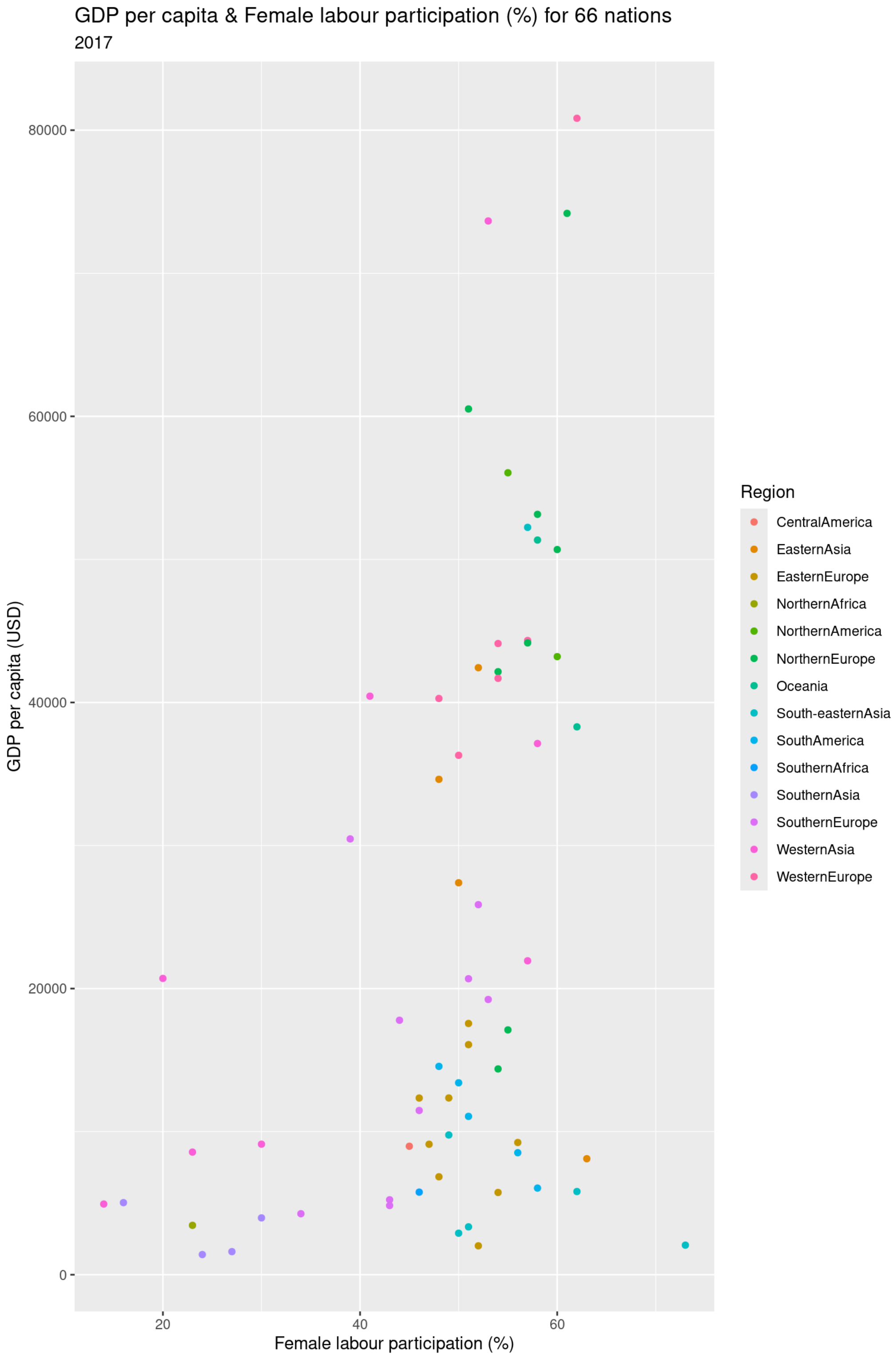



```
ggplot(data = Analysis_df) +
  geom_line(mapping = aes(x = `Life expectancy (Years)`, y = `GDP per capita (USD)`), color = "blue", linewidth=2) +
  labs(
    title = "GDP per capita & Life expectancy for 66 nations",
    subtitle = "2017")
```

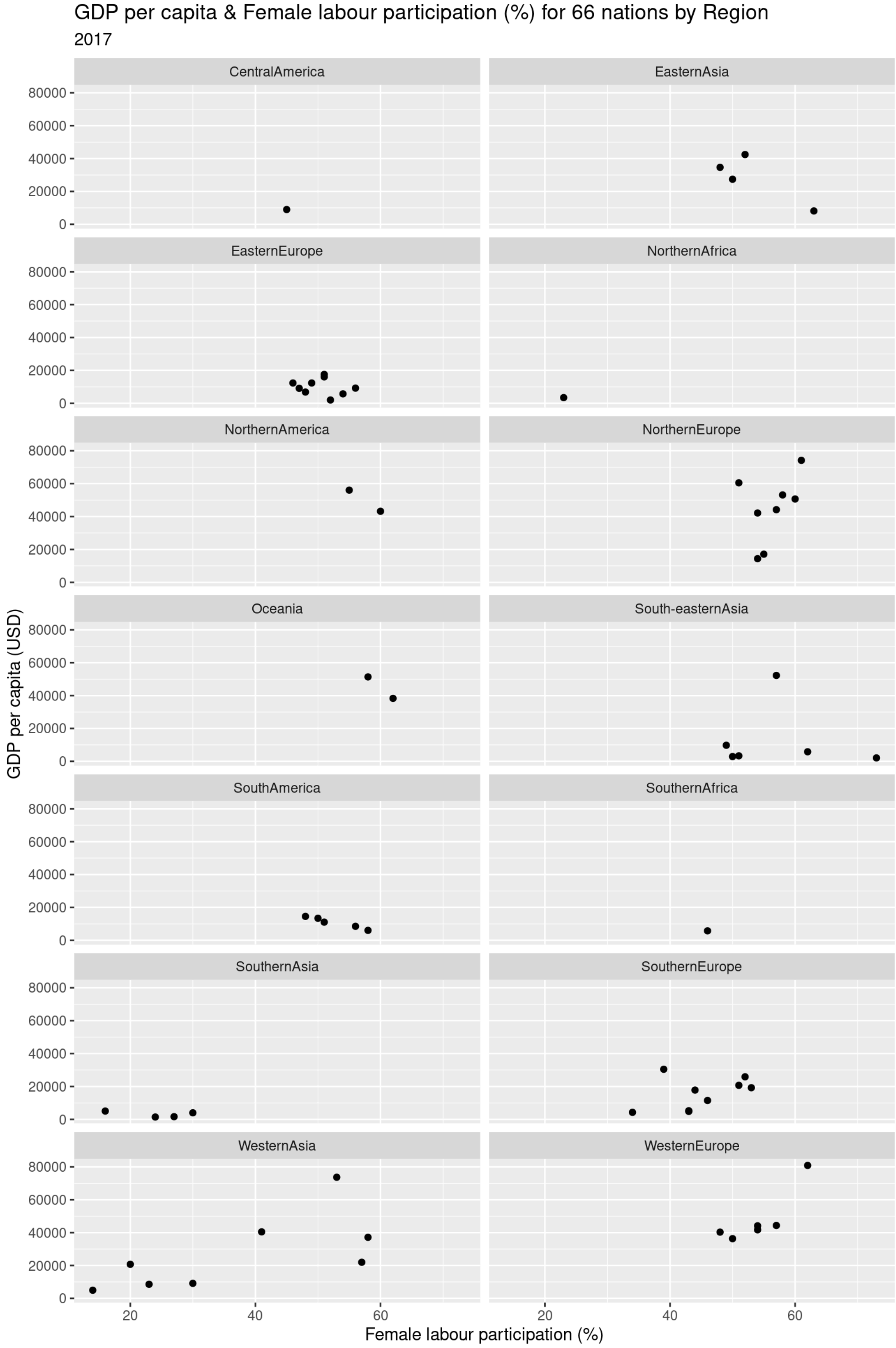



## *Analysis: Gender Diversity*

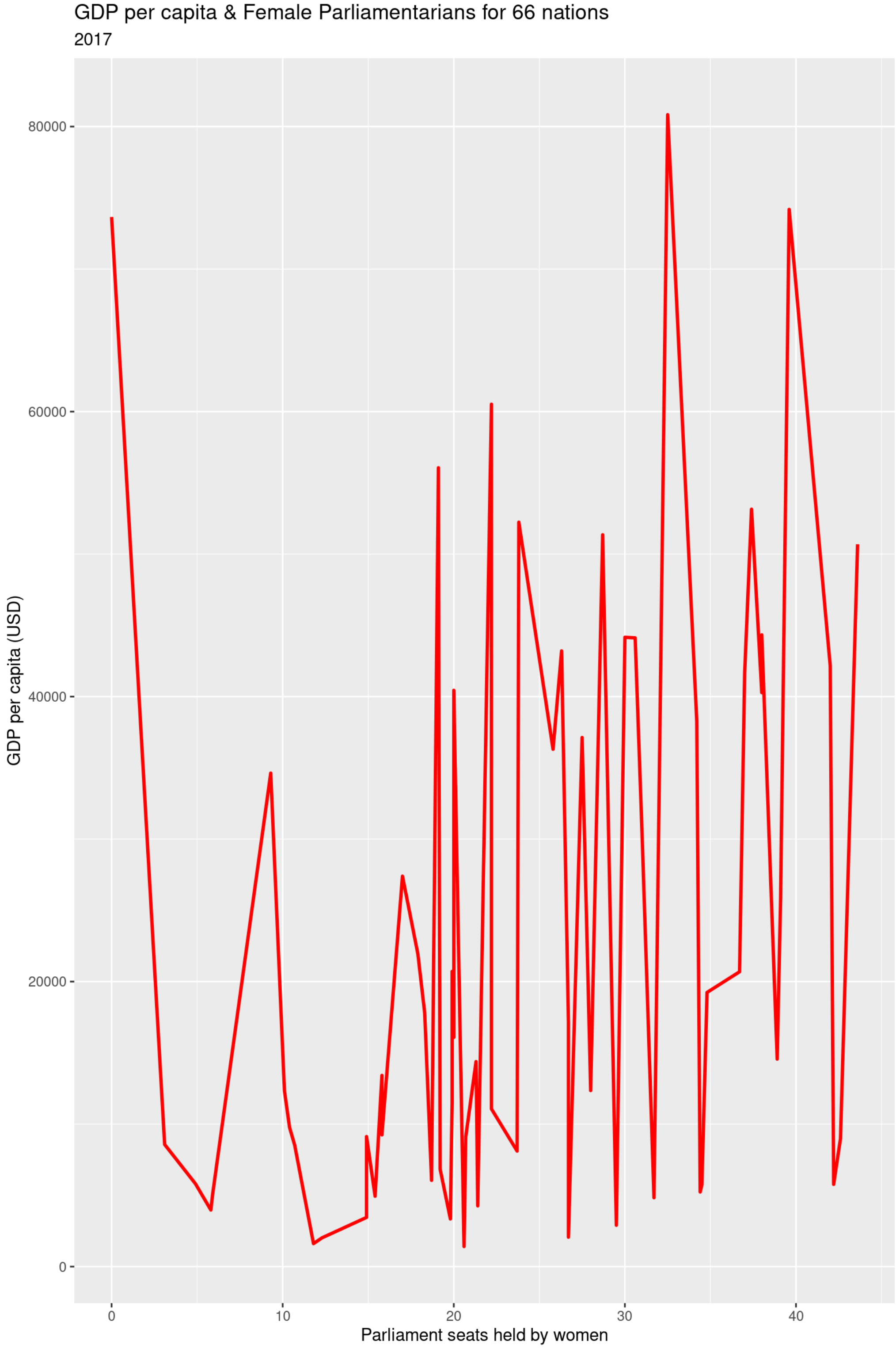
```
ggplot(data = Analysis_df) +
  geom_point(mapping = aes(x = `Female labour participation (%)`, y = `GDP per capita (USD)`, colour = Region)) +
  labs(
    title = "GDP per capita & Female labour participation (%) for 66 nations",
    subtitle = "2017")
```




```
ggplot(data = Analysis_df) +
  geom_point(mapping = aes(x = `Female labour participation (%)`, y = `GDP per capita (USD)`)) +
  facet_wrap(~ Region, nrow = 7) +
  labs(
    title = "GDP per capita & Female labour participation (%) for 66 nations by Region",
    subtitle = "2017")
```




```
Parliament_df <- Analysis_df %>%
  filter(`Parliament seats held by women` >= 0)

ggplot(data = Parliament_df) +
  geom_line(mapping = aes(x = `Parliament seats held by women`, y = `GDP per capita (USD)`), color = "red", linewidth = 1) +
  labs(
    title = "GDP per capita & Female Parliamentarians for 66 nations",
    subtitle = "2017")
```




## *Regression Modelling*

```
Econ_model <- lm(`GDP per capita (USD)` ~ `Health expenditure (%GDP)` + `Life expectancy (Years)` + `Female labour participa
tion (%)` + `Parliament seats held by women`, data = Analysis_df)
tidy(Econ_model)
```

```
## # A tibble: 5 × 5
##   term                               estimate std.error statistic  p.value
##   <chr>                                 <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                       -232124.    32085.     -7.23  9.15e-10
## 2 `Health expenditure (%GDP)`          -167.      260.     -0.644 5.22e- 1
## 3 `Life expectancy (Years)`            3097.      439.      7.06  1.83e- 9
## 4 `Female labour participation (%)`     230.      168.      1.37  1.75e- 1
## 5 `Parliament seats held by women`      192.      191.      1.00  3.20e- 1
```

```
summary(Econ_model)
```

```
##
## Call:
## lm(formula = `GDP per capita (USD)` ~ `Health expenditure (%GDP)` +
##     `Life expectancy (Years)` + `Female labour participation (%)` +
##     `Parliament seats held by women`, data = Analysis_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -19547  -9657  -2620   6187  46258
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -232124.2    32085.3  -7.235 9.15e-10 ***
## `Health expenditure (%GDP)`         -167.3      259.8  -0.644    0.522
## `Life expectancy (Years)`           3096.8      438.7   7.060 1.83e-09 ***
## `Female labour participation (%)`    230.2      167.6   1.374    0.175
## `Parliament seats held by women`     191.5      191.0   1.003    0.320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14240 on 61 degrees of freedom
## Multiple R-squared:  0.551,  Adjusted R-squared:  0.5215
## F-statistic: 18.71 on 4 and 61 DF,  p-value: 4.416e-10
```

# *Discussion of Healthcare Access & Gender Diversity on GDP*

Regarding healthcare access, the plots analysing the relationship between healthcare expenditure and GDP per capita show a random scatter in the data but with a general - yet weak - positive trend overall. Faceted for specific regions the data shows only positive correlations in Eastern Asia and Northern Europe, whereas other regions do not illustrate a linear relationship. With respect to life expectancy, the line-plot shows a positive trending relationship between increased life expectancy and increased GDP per capita, however strong variation exists around a strictly linear relationship, indicating a weak correlation. Thus whilst a weak relationship does seem to exist in the data between healthcare access and GDP, this is contrary to the findings in the literature, which suggest a stronger correlation (Saygılı, 2021).

In analysing gender diversity data, the scatterplots of female labour participation and GDP per capita showed an overall random scatter, suggesting a lack of linear correlation. However when faceted by region, Western and Northern Europe and Western Asia, showed a mildly positive correlation between female labour participation and increased GDP. Conversely the line graph illustrating the dependency of GDP per capita on the amount of female parliamentarians showed no linear correlation whatsoever. Given Dahlum (2022) strongly conclude that female economic participation is determinative of economic growth - at least in the long-run -, these results are contradictory to the literature.

Furthermore, multiple linear regression was constructed to model GDP per capita as a function of health expenditure, life expectancy, female labour participation and the amount of female parliamentarians. Whilst these 4 variables accounted for 55.1% of the variation in the GDP per capita data (with a R-squared value of 0.551), only life expectancy was statistically significant (below a 0.01 threshold).

---

# *Conclusions*

This report purported to affirm the validity of endogenous growth theories which suggest that traditional variables of economic growth - capital and labour - can be expanded to include human and social infrastructure and gender diversity.

When examining how education expenditure and industrial employment relate to GDP per capita across countries, the findings suggest that neither variable alone offers a strong or consistent explanation of income levels. The relationship between education spending and GDP appears highly variable, likely due to differences in education quality, governance, or lag effects. Similarly, higher employment in industry tends to occur in lower-income, industrialising economies, whereas high-income countries show lower industrial employment, reflecting structural transformation. Overall, the visual trends lack clarity, indicating that simple bivariate analysis may be insufficient. A more robust approach—such as multiple or non-linear regression models, or inclusion of additional variables (e.g., productivity, service sector size, or education quality)—may yield deeper insights into the complex drivers of GDP.

In analysing the relationships between economic growth measured by GDP with healthcare access and female economic participation, the results indicated that life expectancy had a strong correlation to GDP per capita, whereas healthcare expenditure only showed linear relationships in particular geographical regions: Eastern Asia and Northern Europe. Likewise female labour participation also only showed a linear relationship in specific regions: Western and Northern Europe and Western Asia; whereas the rate of female parliamentarians was not associated to GDP per capita variation. Thus whilst these variables mostly showed mild positive associations, the results were mixed and could be better analysed by expanding the scope of study to include more than one year, which limited this analysis.

---

# *References*

Dahlum, K., S. (2022). Women's political empowerment and economic growth. *World Development*, *156*, 105822. https://doi.org/10.1016/j.worlddev.2022.105822 (https://doi.org/10.1016/j.worlddev.2022.105822)

Dykas, T., P. (2023). *The solow model of economic growth: Application to contemporary macroeconomic issues*. Routledge, Taylor & Francis Group. 10.4324/9781003323792 (https://10.4324/9781003323792)

Hanushek, E. A., & Woessmann, L. (2010). *The economics of international differences in educational achievement* (NBER Working Paper 15949). National Bureau of Economic Research. https://doi.org/10.3386/W15949 (https://doi.org/10.3386/W15949)

Kaldor, N. (1966). *Causes of the slow rate of economic growth of the united kingdom*. Cambridge University Press.

Kuznets, S. (1966). *Modern economic growth: Rate, structure and spread*. Yale University Press.

Martin, &. S., R. (1998). Slow convergence? The new endogenous growth theory and regional development. *Economic Geography*, *74*(3), 201--227. https://doi.org/10.1111/j.1944-8287.1998.tb00113.x (https://doi.org/10.1111/j.1944-8287.1998.tb00113.x)

Saygılı, &. Ö., H. (2021). Regional economic growth in turkey: The effects of physical, social and financial infrastructure investment. *Empirical Economics*, *60*(4), 2039--2061. https://doi.org/10.1007/s00181-020-01828-0 (https://doi.org/10.1007/s00181-020-01828-0)

Trinh, M., H. H. (2023). Do green growth and technological innovation matter to infrastructure investments in the era of climate change? Global evidence. *Applied Economics*, *55*(35), 4108--4129. https://doi.org/10.1080/00036846.2022.2125493 (https://doi.org/10.1080/00036846.2022.2125493)