

IDS Group Work Code Justin Li

Justin Li

2025-12-07

Contents

1	1. Absolute Change in Youth NEET Rate (2015-2020) by Continent	1
---	--	---

1 1. Absolute Change in Youth NEET Rate (2015-2020) by Continent

This section calculates the absolute change in youth NEET rates across six continents between 2015 and 2020, and visualizes the results using a bar chart.

```
# Load required packages (only once for the entire document)
if (!require(countrycode)) install.packages("countrycode")
library(readr)
library(dplyr)
library(ggplot2)
library(countrycode)
library(tidyr)
library(gridExtra)
library(purrr)

# Load NEET dataset
df_neet <- read_csv("~/Desktop/youth-not-in-education-employment-training.csv", show_col_types = FALSE)
colnames(df_neet)[4] <- "NEET_rate"

# Custom continent mapping for specific countries
custom_continent <- c(
  "United States" = "North America", "Canada" = "North America", "Mexico" = "North America",
  "Australia" = "Oceania", "New Zealand" = "Oceania",
  "Brazil" = "South America", "Argentina" = "South America",
  "China" = "Asia", "India" = "Asia",
  "Germany" = "Europe", "France" = "Europe",
  "Nigeria" = "Africa", "South Africa" = "Africa",
  "Kosovo" = "Europe", "Micronesia (country)" = "Oceania"
)

# Data cleaning and continent assignment
df_neet_clean <- df_neet %>%
  filter(!grepl("\\(WB\\)|EU|Central Asia|Latin America", Entity)) %>%
  mutate(
```

```

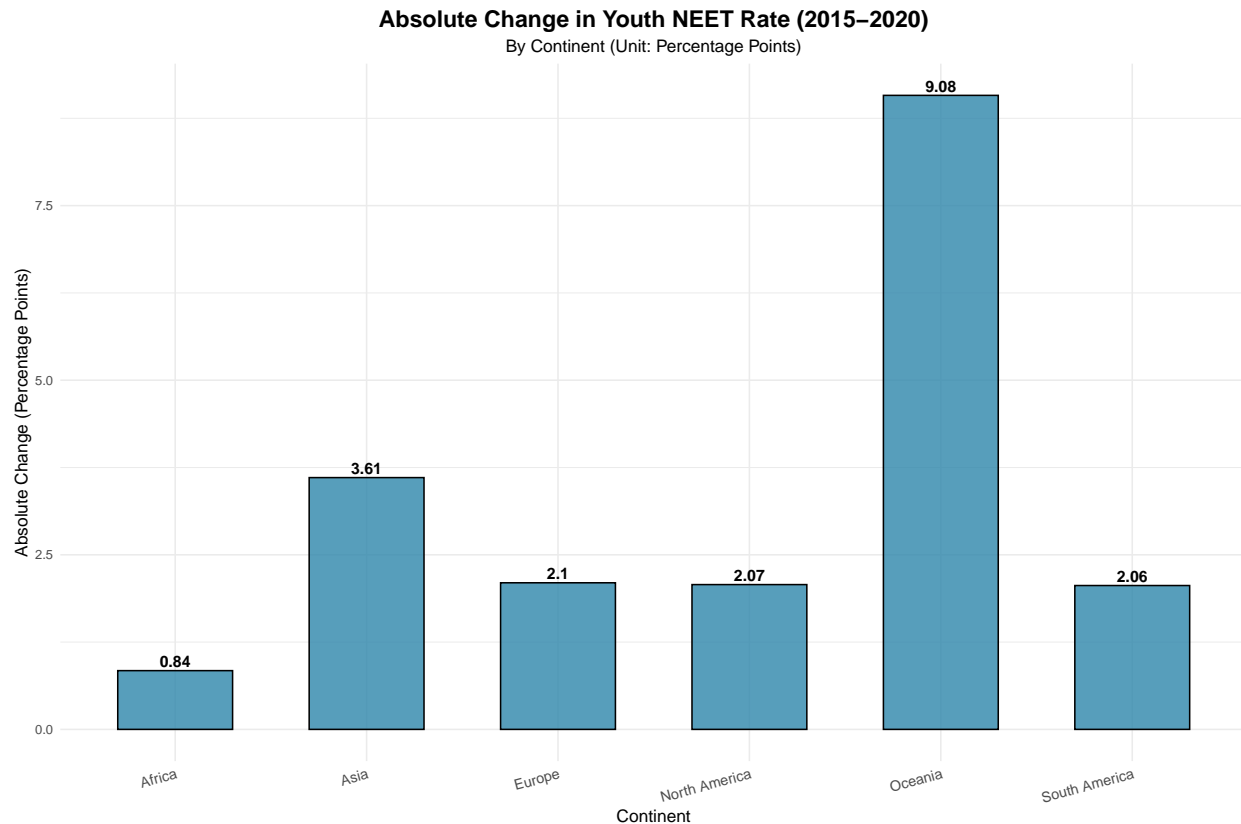
Continent = suppressWarnings(countrycode(Entity, origin = "country.name", destination = "continent"))
Continent = case_when(
  Entity %in% names(custom_continent) ~ custom_continent[Entity],
  TRUE ~ Continent
)
)

# Define target continents
target_continents <- c("Asia", "Europe", "Africa", "North America", "South America", "Oceania")

# Calculate yearly average NEET rates and absolute change
continent_change <- df_neet_clean %>%
  filter(
    Continent %in% target_continents,
    Year %in% c(2015, 2020),
    !is.na(NEET_rate)
  ) %>%
  group_by(Continent, Year) %>%
  summarise(NEET_rate = mean(NEET_rate, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = Year, values_from = NEET_rate, names_prefix = "NEET_") %>%
  mutate(Absolute_Change = abs(NEET_2020 - NEET_2015)) %>%
  complete(Continent = target_continents, fill = list(Absolute_Change = 0))

# Create bar chart
p <- ggplot(continent_change, aes(x = Continent, y = Absolute_Change)) +
  geom_col(
    fill = "#2E86AB",
    width = 0.6,
    alpha = 0.8,
    color = "black"
  ) +
  geom_text(
    aes(label = round(Absolute_Change, 2)),
    vjust = -0.3,
    size = 4,
    fontface = "bold"
  ) +
  labs(
    title = "Absolute Change in Youth NEET Rate (2015-2020)",
    subtitle = "By Continent (Unit: Percentage Points)",
    x = "Continent",
    y = "Absolute Change (Percentage Points)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 12, hjust = 0.5),
    axis.text.x = element_text(angle = 15, hjust = 1, size = 10),
    axis.title = element_text(size = 12)
  )
print(p)

```



```
# Define file paths (Mac Desktop)
gdp_file <- "gdp-per-capita-worldbank.csv"
continent_file <- "continents-according-to-our-world-in-data.csv"
gdp_df <- read_csv(file.path("~/Desktop", gdp_file), show_col_types = FALSE)
continent_df <- read_csv(file.path("~/Desktop", continent_file), show_col_types = FALSE)

# Merge data and rename column
merged_df <- gdp_df %>%
  left_join(continent_df %>% select(Entity, Continent), by = "Entity") %>%
  filter(
    !is.na(Continent),
    !is.na(`GDP per capita, PPP (constant 2017 international $)`),
    Year >= 1990
  ) %>%
  rename(gdp_per_capita = `GDP per capita, PPP (constant 2017 international $)`)

continents_list <- c("Asia", "Africa", "Europe", "North America", "South America", "Oceania")

# LDC lists by continent
ldc_asia <- c("Afghanistan", "Bangladesh", "Nepal", "Myanmar")
ldc_africa <- c("Ethiopia", "Rwanda", "Mozambique", "Malawi")
ldc_europe <- c("Moldova", "Bosnia and Herzegovina", "Serbia", "Albania")
ldc_north_america <- c("Haiti", "Guatemala", "Honduras", "Nicaragua")
ldc_south_america <- c("Bolivia", "Paraguay", "Ecuador", "Guyana")
ldc_oceania <- c("Solomon Islands", "Vanuatu", "Fiji", "Papua New Guinea")

ldc_by_continent <- list(
```

```

Asia = ldc_asia,
Africa = ldc_africa,
Europe = ldc_europe,
`North America` = ldc_north_america,
`South America` = ldc_south_america,
Oceania = ldc_oceania
)

growth_df <- merged_df %>%
  group_by(Entity, Continent) %>%
  mutate(
    base_year = ifelse(any(Year == 1990), 1990, min(Year, na.rm = TRUE)),
    base_gdp = first(gdp_per_capita[Year == base_year])
  ) %>%
  filter(!is.na(base_gdp)) %>%
  mutate(
    cumulative_growth = (gdp_per_capita / base_gdp) * 100,
    base_year_label = paste("Base Year:", base_year)
  ) %>%
  ungroup() %>%
  filter(Continent %in% continents_list)

# Iterate over each continent to generate plots
purrr::walk(continents_list, function(continent) {
  continent_data <- growth_df %>% filter(Continent == continent)
  ldc_countries <- ldc_by_continent[[continent]]
  ldc_data <- continent_data %>% filter(Entity %in% ldc_countries)

  continent_avg_growth <- continent_data %>%
    group_by(Year) %>%
    summarise(avg_cumulative_growth = mean(cumulative_growth, na.rm = TRUE), .groups = "drop")

  p <- ggplot() +
    geom_area(data = continent_avg_growth, aes(x = Year, y = avg_cumulative_growth),
      fill = "steelblue", alpha = 0.6) +
    geom_line(data = continent_avg_growth, aes(x = Year, y = avg_cumulative_growth),
      color = "darkblue", linewidth = 1.2) +
    geom_line(data = ldc_data, aes(x = Year, y = cumulative_growth, color = Entity),
      linewidth = 1, alpha = 0.8) +
    geom_hline(yintercept = 487, linetype = "dashed", color = "red", linewidth = 1, alpha = 0.8) +
    annotate("text", x = max(continent_avg_growth$Year), y = 500,
      label = "UN Target: 7% Annual Growth (Cumulative 387%)",
      color = "red", hjust = 1, size = 3) +

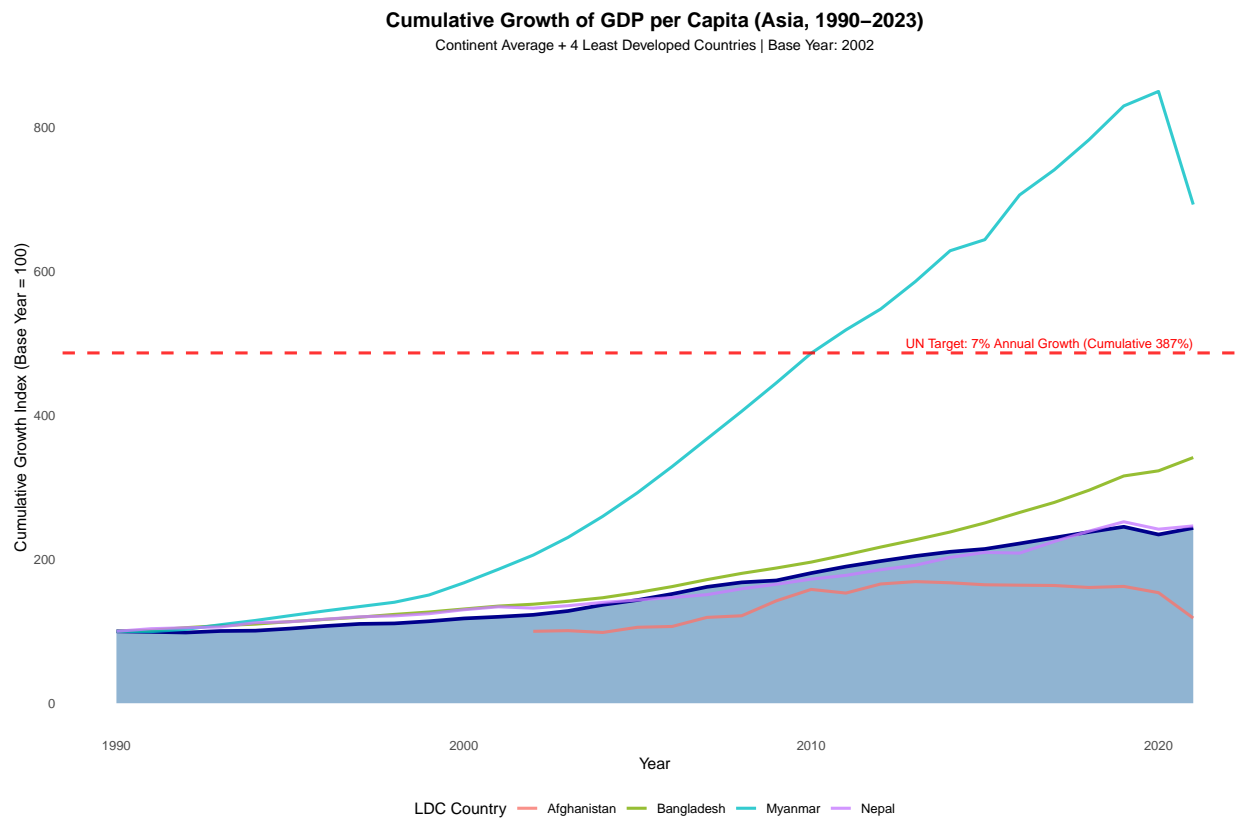
    labs(
      title = paste("Cumulative Growth of GDP per Capita (", continent, ", 1990-2023)", sep = ""),
      subtitle = paste("Continent Average + 4 Least Developed Countries |", unique(continent_data$base_year)),
      x = "Year", y = "Cumulative Growth Index (Base Year = 100)",
      color = "LDC Country"
    ) +
    theme_minimal() +
    theme(
      plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
      plot.subtitle = element_text(hjust = 0.5, size = 10),

```

```

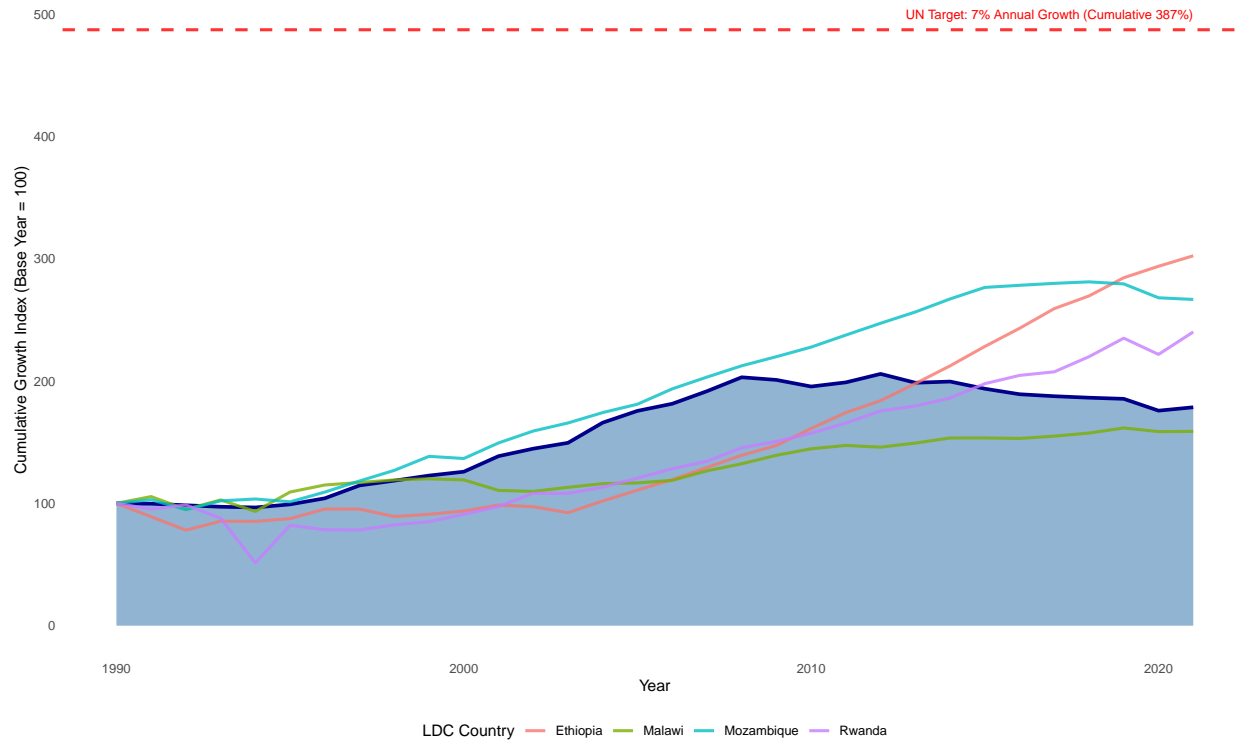
legend.position = "bottom",
panel.grid = element_blank(),
plot.background = element_rect(fill = "white", color = NA)
)
print(p)
})

```



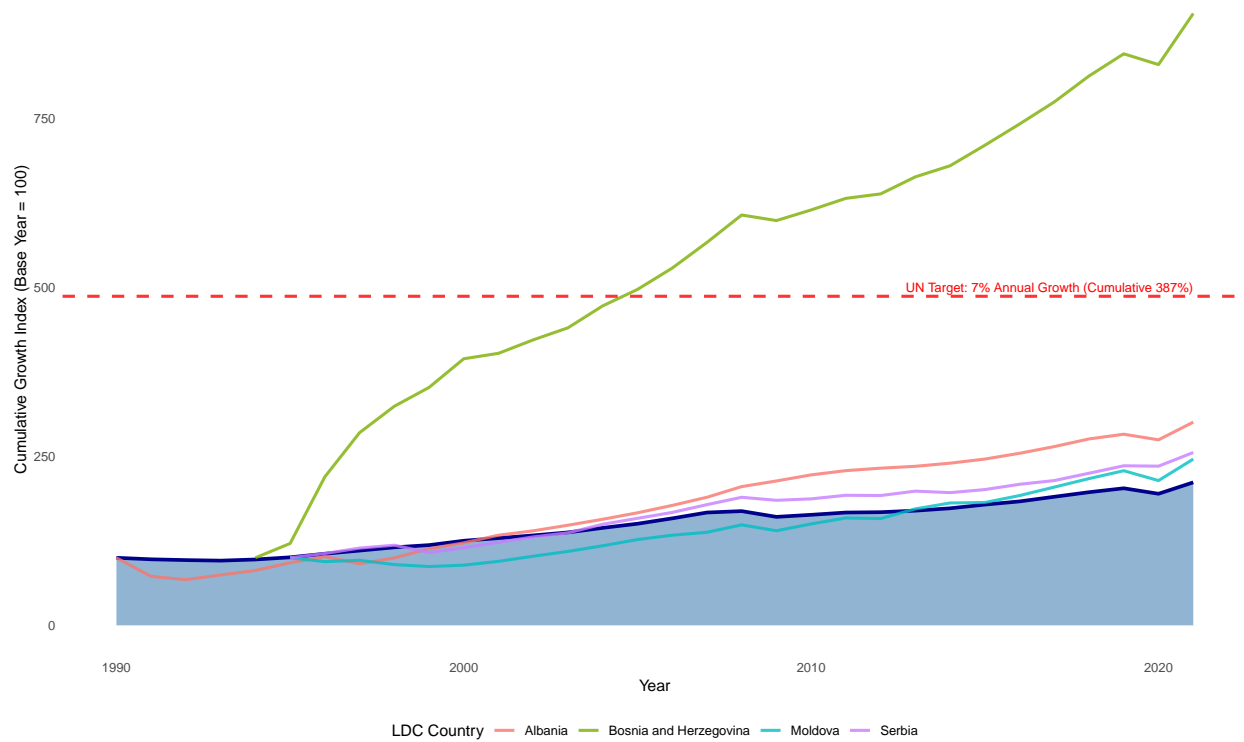
Cumulative Growth of GDP per Capita (Africa, 1990–2023)

Continent Average + 4 Least Developed Countries | Base Year: 1990



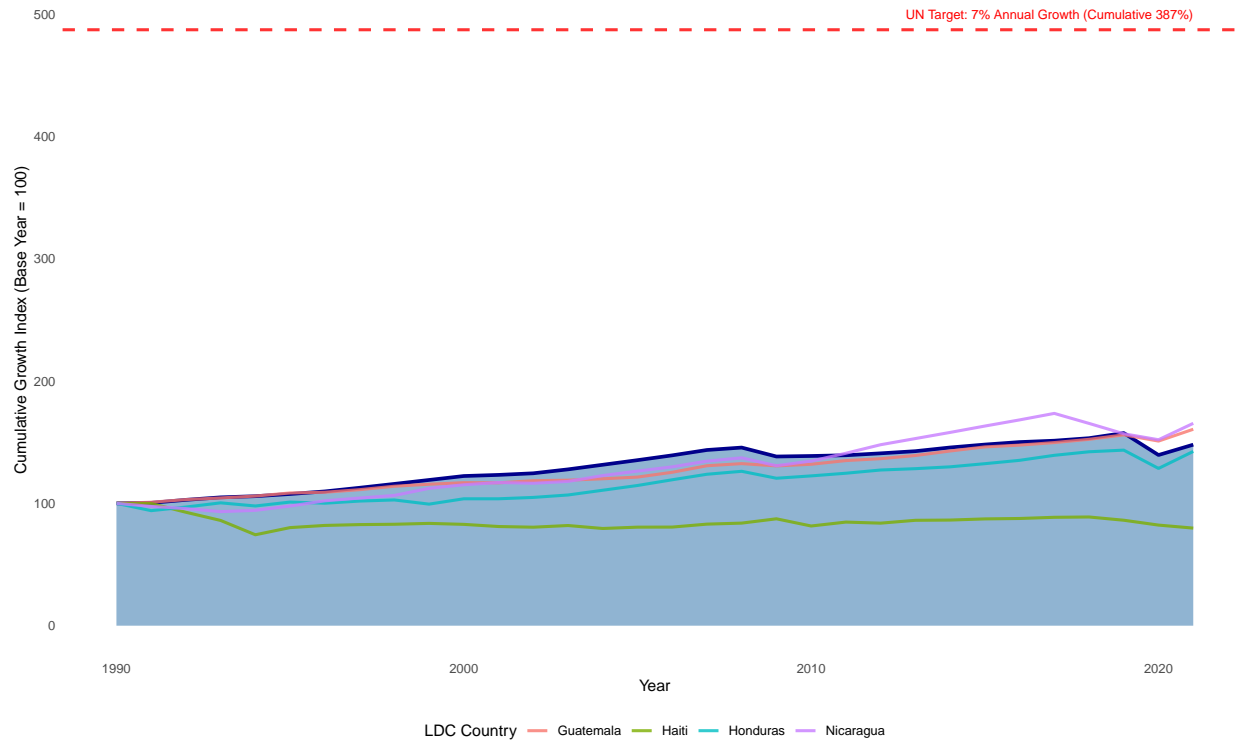
Cumulative Growth of GDP per Capita (Europe, 1990–2023)

Continent Average + 4 Least Developed Countries | Base Year: 1990



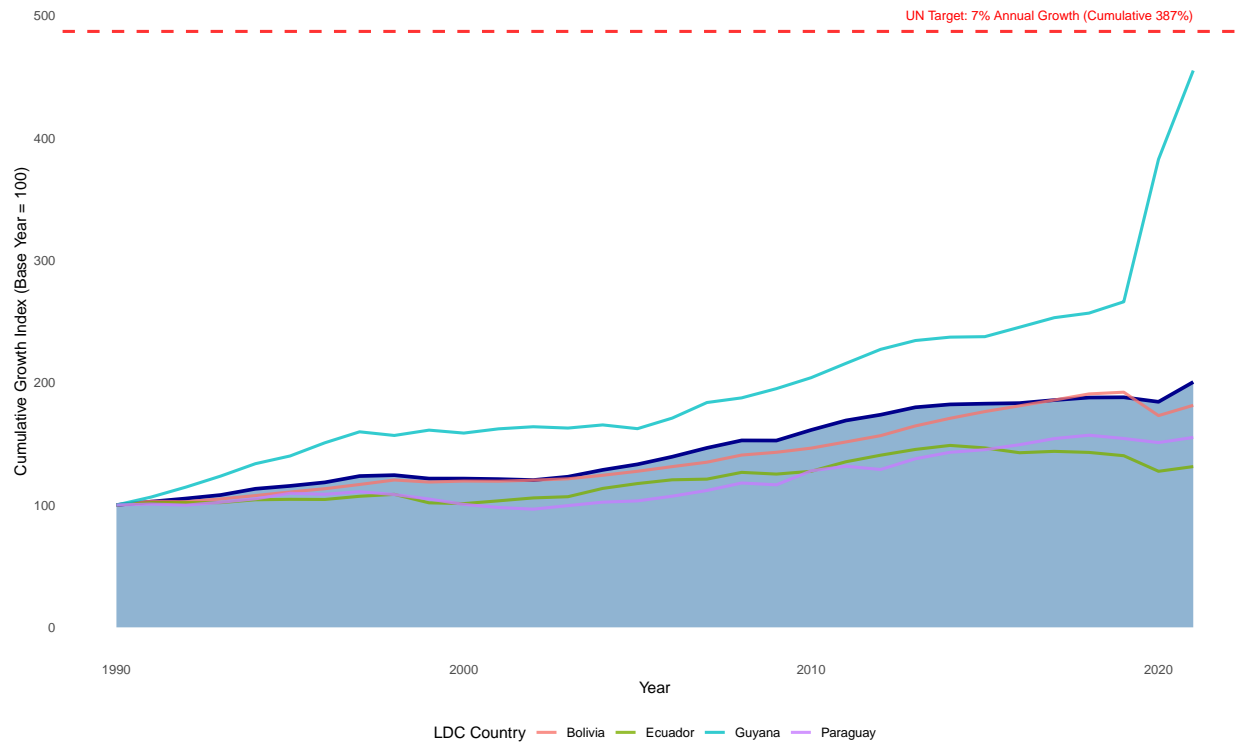
Cumulative Growth of GDP per Capita (North America, 1990–2023)

Continent Average + 4 Least Developed Countries | Base Year: 1990



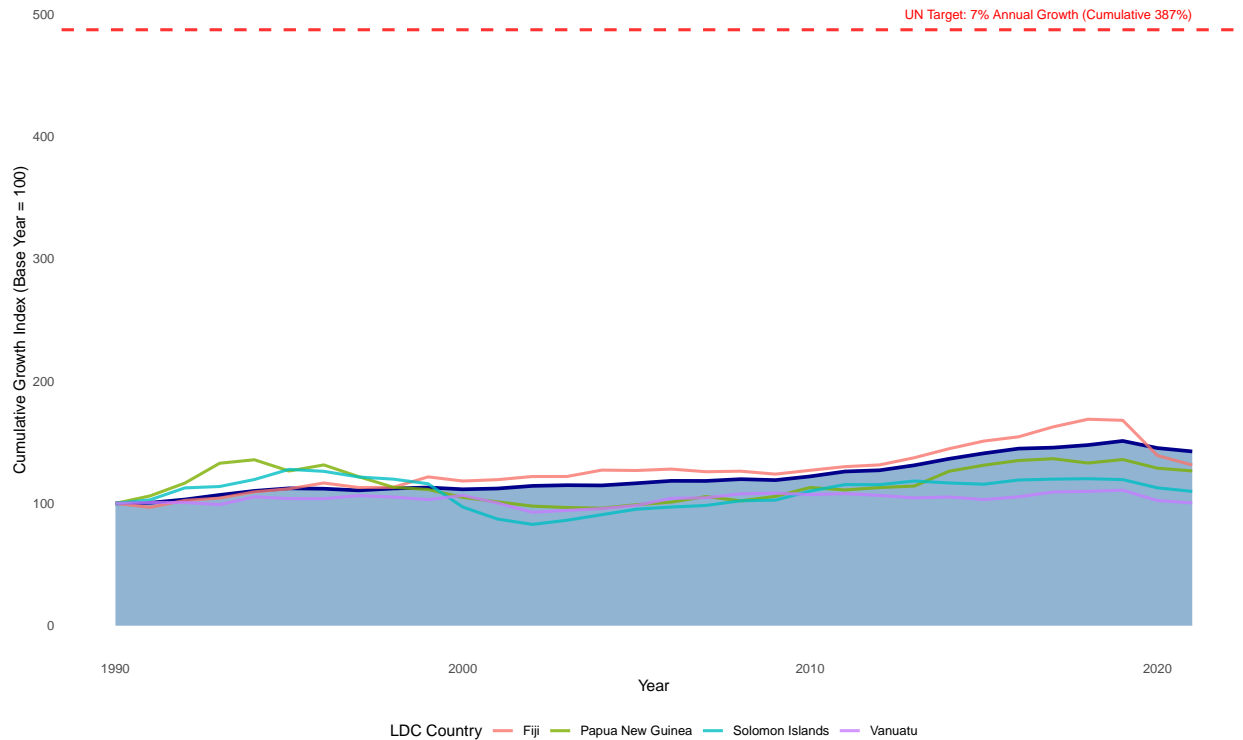
Cumulative Growth of GDP per Capita (South America, 1990–2023)

Continent Average + 4 Least Developed Countries | Base Year: 1990



Cumulative Growth of GDP per Capita (Oceania, 1990–2023)

Continent Average + 4 Least Developed Countries | Base Year: 1990



```
# Final summary table
growth_df %>%
  filter(Year == max(Year)) %>%
  group_by(Continent) %>%
  summarise(avg_cumulative_growth = mean(cumulative_growth, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(avg_cumulative_growth)) %>%
  print()
```

```
## # A tibble: 6 x 2
##   Continent      avg_cumulative_growth
##   <chr>          <dbl>
## 1 Asia          244.
## 2 Europe        212.
## 3 South America 200.
## 4 Africa        178.
## 5 North America 148.
## 6 Oceania       142.
```

```
# Reuse merged_df and growth_df from Section 2 (no need to reload data)
purrr::walk(continent_list, function(continent) {
  continent_data <- growth_df %>% filter(Continent == continent)
  ldc_countries <- ldc_by_continent[[continent]]
  ldc_data <- continent_data %>% filter(Entity %in% ldc_countries)

  # Calculate median cumulative growth
  continent_median_growth <- continent_data %>%
```



```

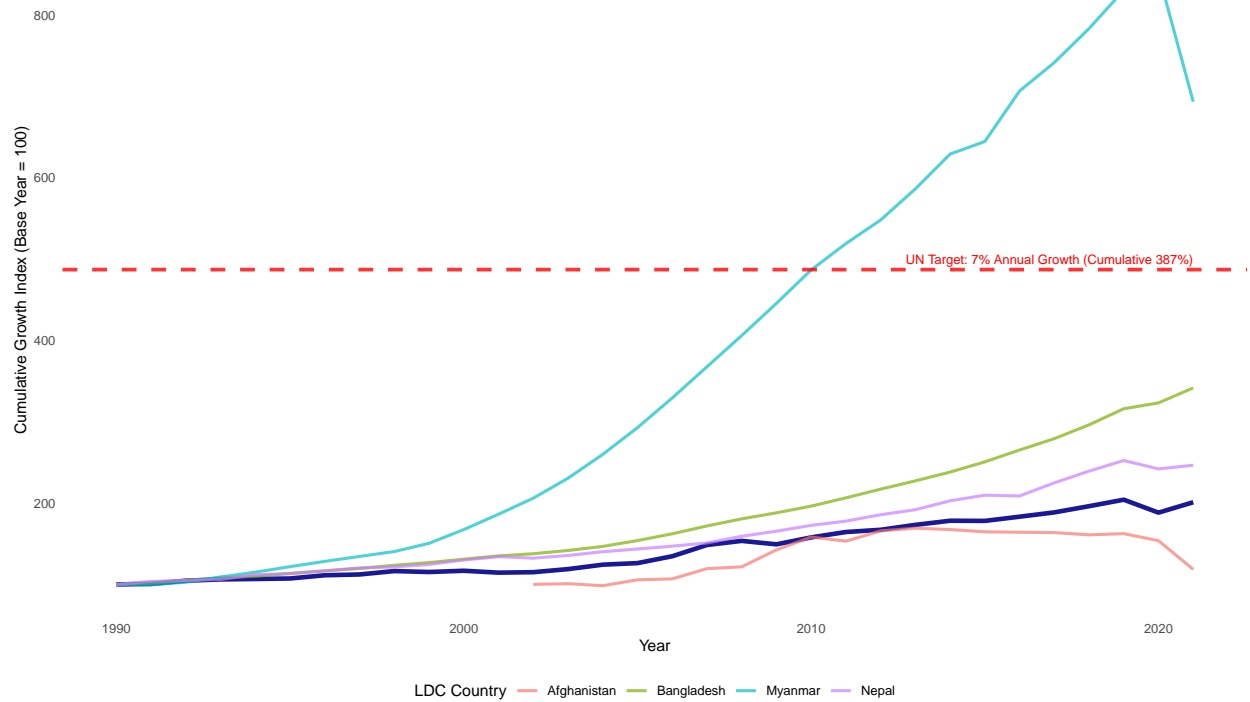
group_by(Year) %>%
  summarise(median_cumulative_growth = median(cumulative_growth, na.rm = TRUE), .groups = "drop")

# Create line plot
p <- ggplot() +
  geom_line(data = continent_median_growth,
            aes(x = Year, y = median_cumulative_growth),
            color = "darkblue", linewidth = 1.5, alpha = 0.9) +
  geom_line(data = ldc_data,
            aes(x = Year, y = cumulative_growth, color = Entity),
            linewidth = 1, alpha = 0.7) +
  geom_hline(yintercept = 487, linetype = "dashed", color = "red", linewidth = 1.2, alpha = 0.8) +
  annotate("text", x = max(continent_median_growth$Year), y = 500,
          label = "UN Target: 7% Annual Growth (Cumulative 387%)",
          color = "red", hjust = 1, size = 3) +
  labs(
    title = paste("Median Cumulative Growth of GDP per Capita (", continent, ", 1990-2023)", sep = " "),
    subtitle = paste("Continent Median + 4 Least Developed Countries |", unique(continent_data$base_y), sep = " "),
    x = "Year", y = "Cumulative Growth Index (Base Year = 100)",
    color = "LDC Country"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, size = 10),
    legend.position = "bottom",
    panel.grid = element_blank(),
    plot.background = element_rect(fill = "white", color = NA)
  )
print(p)
})

```

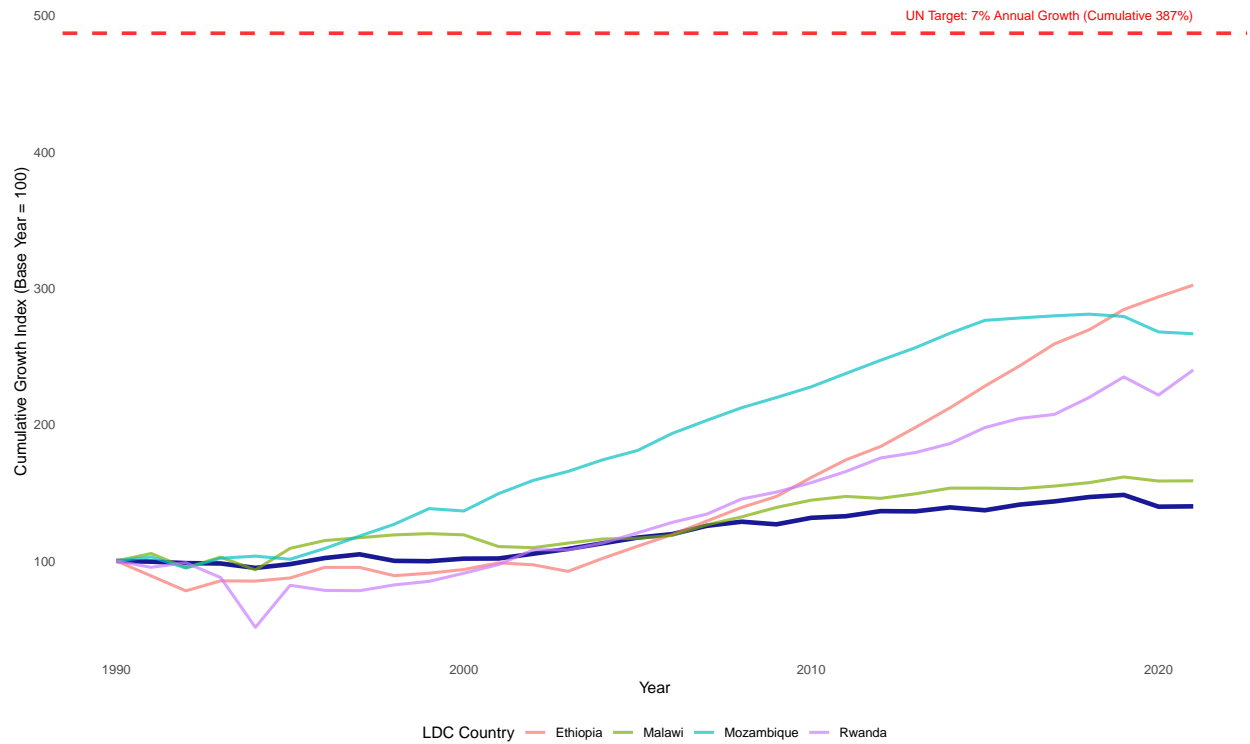
Median Cumulative Growth of GDP per Capita (Asia, 1990–2023)

Continent Median + 4 Least Developed Countries | Base Year: 2002



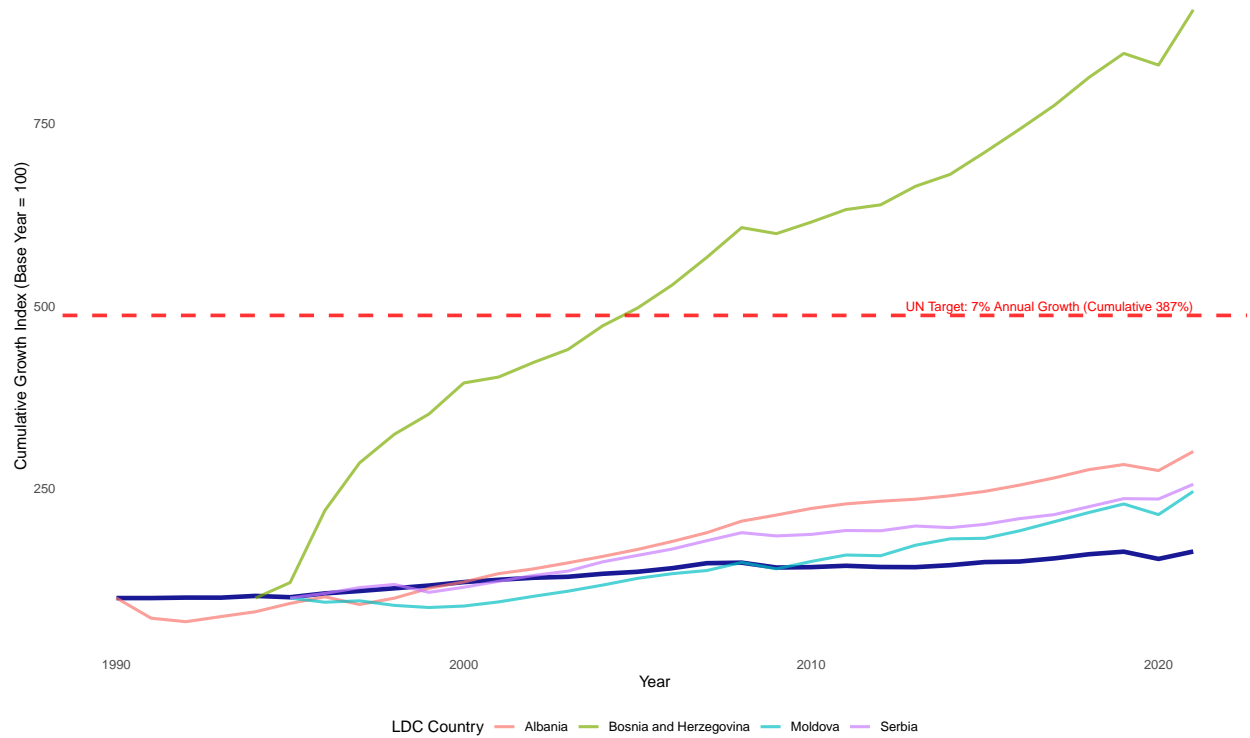
Median Cumulative Growth of GDP per Capita (Africa, 1990–2023)

Continent Median + 4 Least Developed Countries | Base Year: 1990



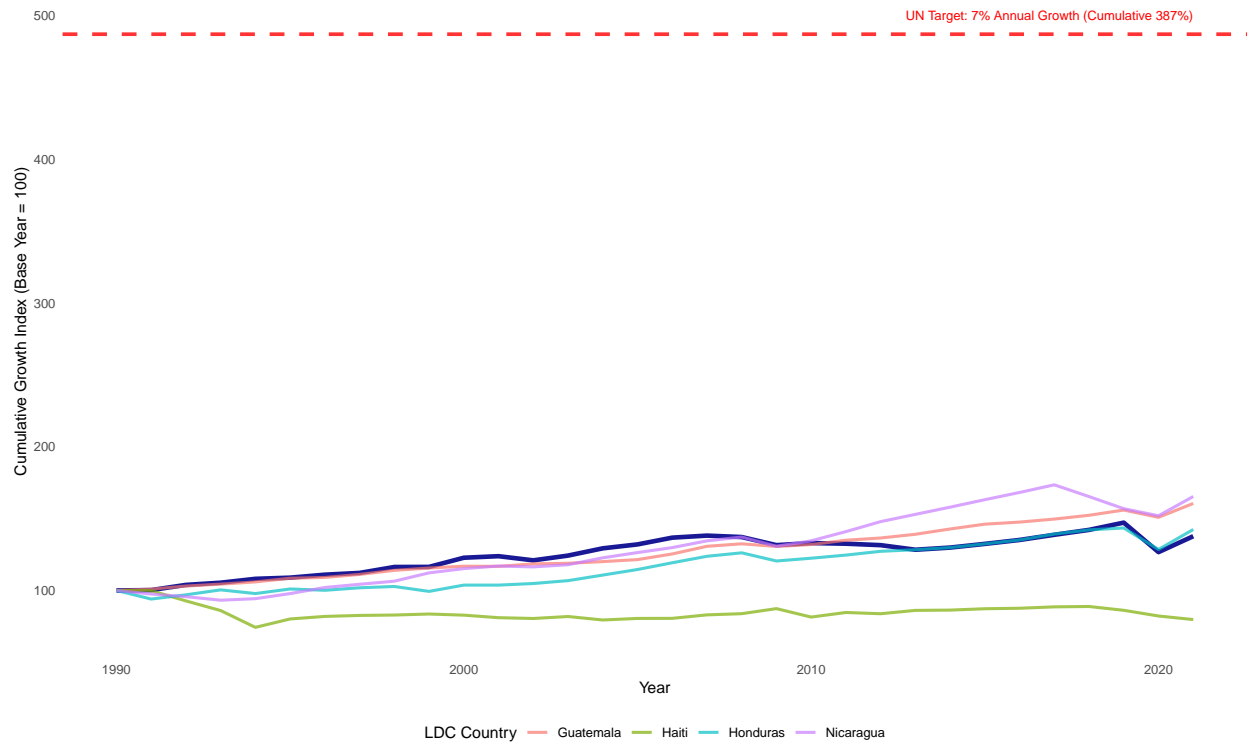
Median Cumulative Growth of GDP per Capita (Europe, 1990–2023)

Continent Median + 4 Least Developed Countries | Base Year: 1990



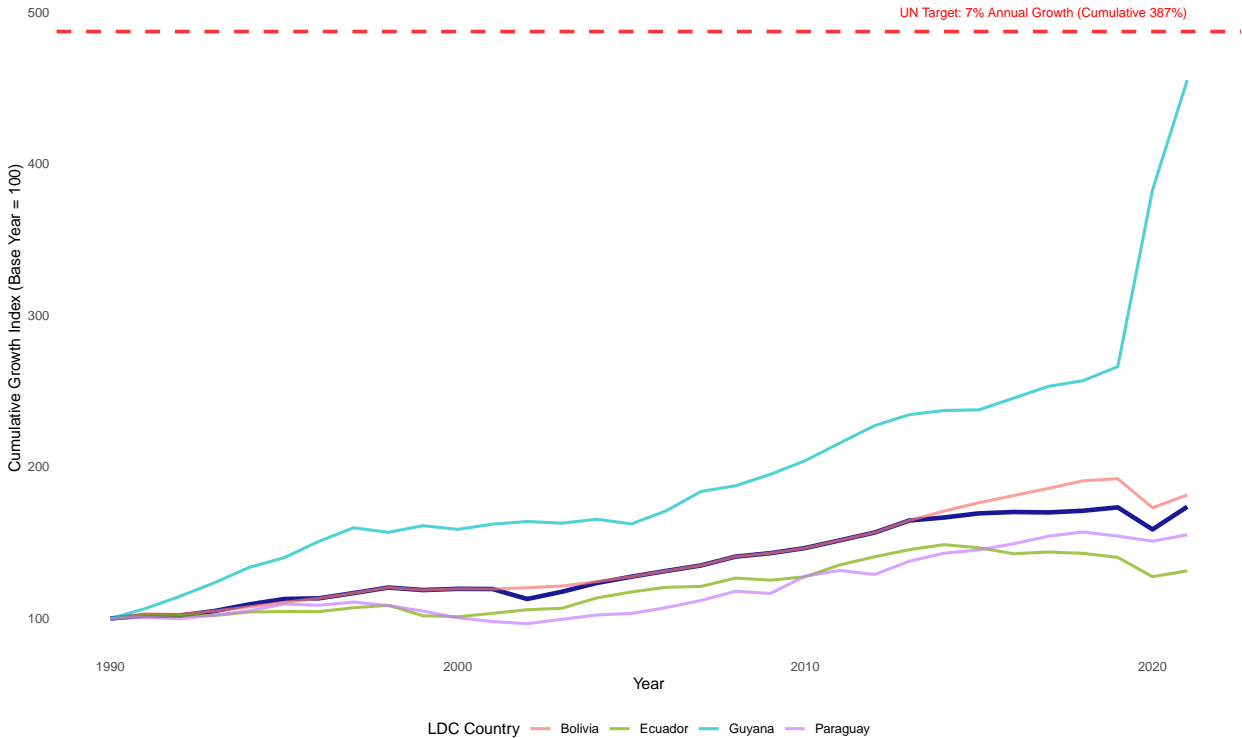
Median Cumulative Growth of GDP per Capita (North America, 1990–2023)

Continent Median + 4 Least Developed Countries | Base Year: 1990



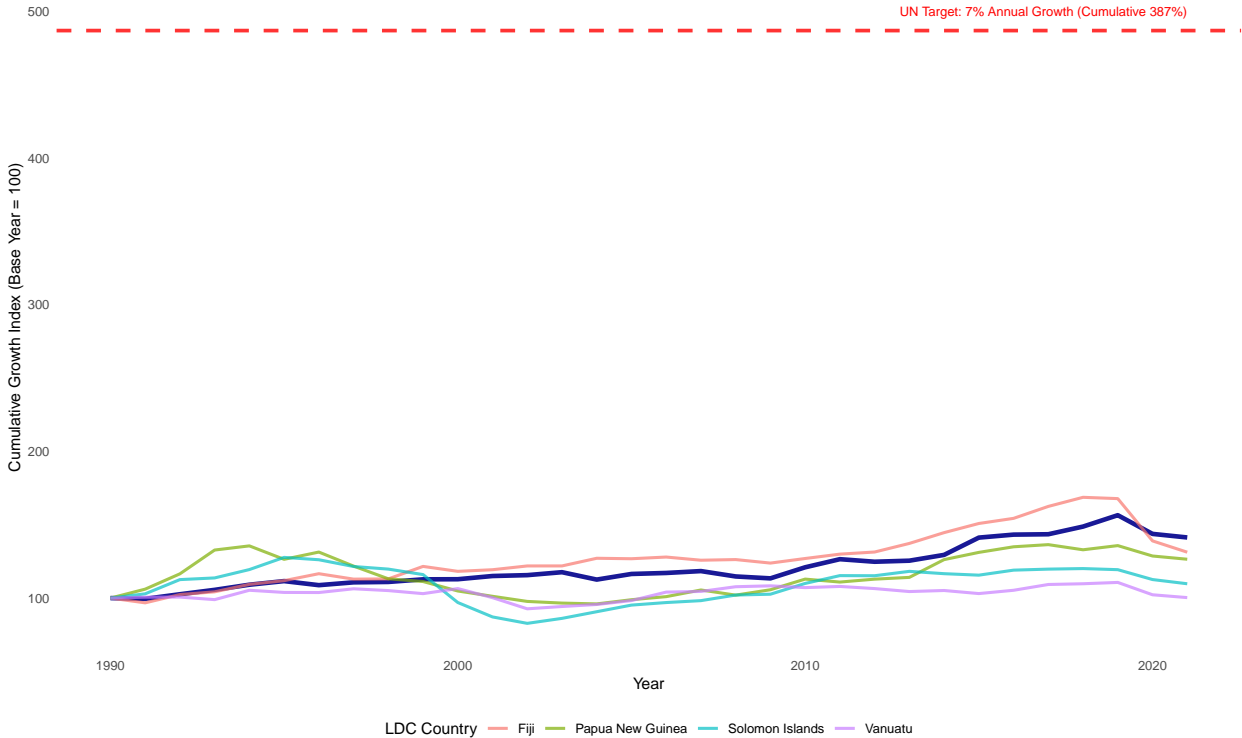
Median Cumulative Growth of GDP per Capita (South America, 1990–2023)

Continent Median + 4 Least Developed Countries | Base Year: 1990



Median Cumulative Growth of GDP per Capita (Oceania, 1990–2023)

Continent Median + 4 Least Developed Countries | Base Year: 1990



```

# Median summary table
growth_df %>%
  filter(Year == max(Year)) %>%
  group_by(Continent) %>%
  summarise(median_cumulative_growth = median(cumulative_growth, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(median_cumulative_growth)) %>%
  print()

## # A tibble: 6 x 2
##   Continent      median_cumulative_growth
##   <chr>          <dbl>
## 1 Asia          201.
## 2 South America 174.
## 3 Europe        164.
## 4 Oceania       141.
## 5 Africa        140.
## 6 North America 138.

# Reuse NEET dataset from Section 1 (df_neet)
colnames(df_neet)[4] <- "NEET_rate"

# Custom continent mapping for NEET-unemployment analysis
custom_continent_unemp <- c(
  "United States" = "North America", "Canada" = "North America", "Mexico" = "North America",
  "Jamaica" = "North America", "El Salvador" = "North America",
  "Brazil" = "South America", "Argentina" = "South America", "Chile" = "South America",
  "Australia" = "Oceania", "New Zealand" = "Oceania", "Fiji" = "Oceania",
  "Kazakhstan" = "Asia", "Japan" = "Asia", "South Korea" = "Asia", "Thailand" = "Asia",
  "Vietnam" = "Asia", "China" = "Asia", "Nigeria" = "Africa", "South Africa" = "Africa",
  "Ethiopia" = "Africa", "Germany" = "Europe", "France" = "Europe", "United Kingdom" = "Europe",
  "Italy" = "Europe", "Spain" = "Europe", "Switzerland" = "Europe", "Kosovo" = "Europe"
)

# Clean NEET data: 2015-2020 average
df_neet_clean_unemp <- df_neet %>%
  filter(Year >= 2015 & Year <= 2020, !is.na(NEET_rate)) %>%
  group_by(Entity) %>%
  summarise(NEET_rate = mean(NEET_rate, na.rm = TRUE), .groups = "drop") %>%
  mutate(
    Continent = suppressWarnings(countrycode(Entity, origin = "country.name", destination = "continent")),
    Continent = case_when(
      Entity %in% names(custom_continent_unemp) ~ custom_continent_unemp[Entity],
      TRUE ~ Continent
    )
  ) %>%
  filter(!is.na(Continent))

# Unemployment data
unemp_data <- data.frame(
  Entity = c(
    "Kazakhstan", "Switzerland", "Japan", "South Korea", "Thailand", "Nigeria",
    "Germany", "Moldova", "Iceland", "Hong Kong", "Netherlands", "Vietnam",
    "Australia", "Malta", "Jamaica", "United States", "Czech Republic",
  )

```

```

"El Salvador", "Taiwan", "Austria", "Latvia", "Cyprus", "Poland",
"United Kingdom", "Lithuania", "Denmark", "Bulgaria", "Ireland",
"New Zealand", "Slovenia", "Canada", "Hungary", "Norway", "Saudi Arabia",
"Euro Area", "Albania", "European Union", "Rwanda", "Turkey", "Slovakia",
"China", "Georgia", "Luxembourg", "Croatia", "Belgium", "Portugal",
"Greece", "France", "Estonia", "Kosovo", "Italy", "Iran", "Cape Verde",
"Sri Lanka", "Finland", "Serbia", "Sweden", "Spain", "New Caledonia",
"Romania", "Ethiopia", "Morocco", "Palestine", "Angola", "South Africa",
"Fiji", "Brazil", "Argentina", "Chile", "Papua New Guinea"
),
Unemp_Rate = c(
  3, 3.1, 3.5, 5.3, 5.9, 6.5, 6.6, 6.8, 6.9, 7.9, 9, 9.03, 9.6, 10.1, 10.2, 10.4,
  10.6, 11.8, 11.81, 12.1, 12.1, 12.6, 12.6, 12.7, 12.9, 13.2, 13.3, 13.4, 13.8,
  13.8, 14.1, 14.1, 14.4, 14.7, 14.8, 15.1, 15.2, 15.5, 15.6, 16.2, 17.3, 17.5,
  17.6, 17.8, 18, 18.3, 18.8, 18.9, 19.4, 19.5, 19.8, 20.2, 20.3, 20.8, 22.5, 23.4,
  24.6, 25.3, 26.2, 26.9, 27.2, 35.8, 37.6, 48.7, 58.5,
  11.7, 14.2, 13.5, 9.2, 12.5
)
)

# Merge NEET and unemployment data
combined_data <- df_neet_clean_unemp %>%
  inner_join(unemp_data, by = "Entity") %>%
  filter(!is.na(NEET_rate) & !is.na(Unemp_Rate))

# Target continents
target_continents <- c("Asia", "Europe", "Africa", "North America", "South America", "Oceania")
combined_data <- combined_data %>%
  bind_rows(tibble(Continent = target_continents)) %>%
  distinct(Continent, Entity, .keep_all = TRUE)

# Continent color palette
continent_colors <- c(
  "Asia" = "#2E86AB", "Europe" = "#A23B72", "Africa" = "#F18F01",
  "North America" = "#C73E1D", "South America" = "#3F88C5", "Oceania" = "#43AA8B"
)

# Initialize plot list
plot_list <- list()

# Create plot for each continent
for (cont in target_continents) {
  plot_data <- filter(combined_data, Continent == cont) %>%
    filter(!is.na(NEET_rate) & !is.na(Unemp_Rate)) %>%
    mutate(Entity = factor(Entity))

  if(nrow(plot_data) >= 2) {
    p <- ggplot(plot_data, aes(x = Unemp_Rate, y = NEET_rate, label = Entity)) +
      geom_point(
        color = continent_colors[cont], fill = adjustcolor(continent_colors[cont], alpha.f = 0.3),
        size = 4, shape = 21, stroke = 1
      ) +
      geom_smooth(

```

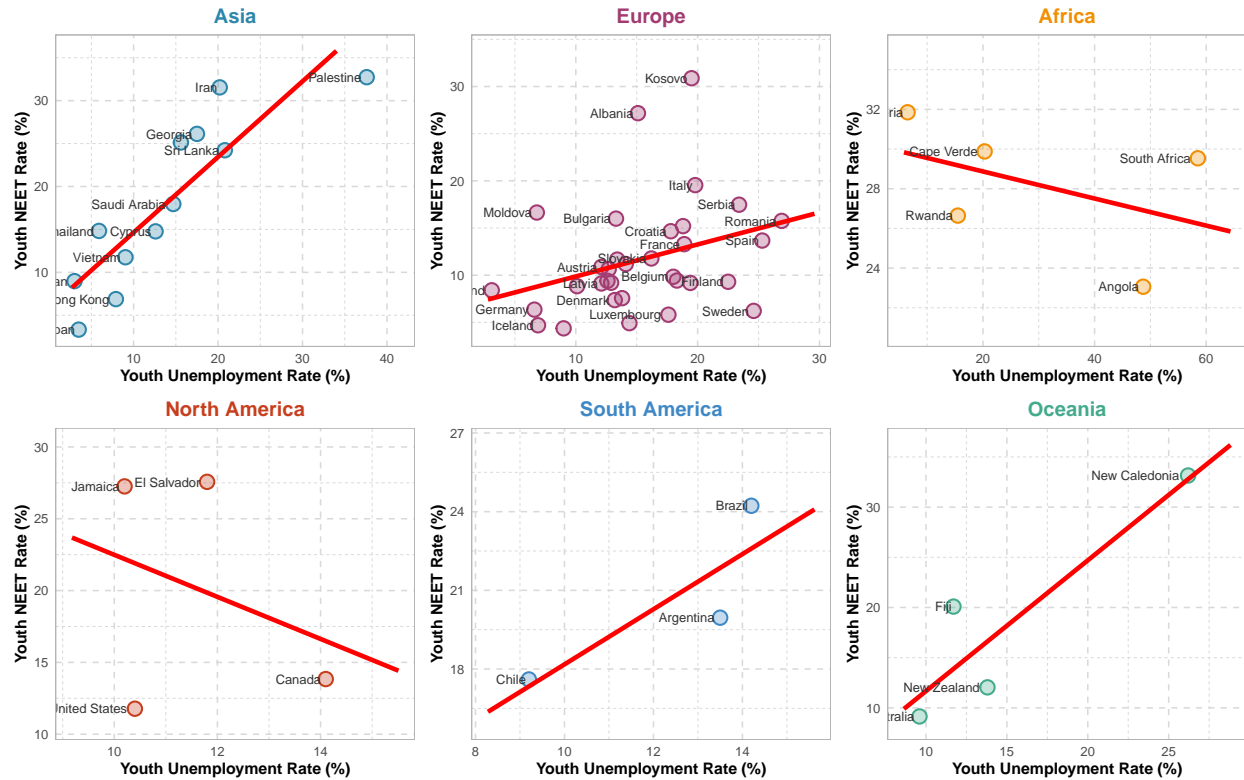
```

    method = "lm", se = FALSE, color = "red",
    linewidth = 1.5, fullrange = TRUE
  ) +
  geom_text(hjust = 1.1, vjust = 0.5, size = 3, color = "#333333", check_overlap = TRUE) +
  labs(
    title = cont,
    x = "Youth Unemployment Rate (%)",
    y = "Youth NEET Rate (%)"
  ) +
  scale_x_continuous(limits = c(min(plot_data$Unemp_Rate)*0.9, max(plot_data$Unemp_Rate)*1.1)) +
  scale_y_continuous(limits = c(min(plot_data$NEET_rate)*0.9, max(plot_data$NEET_rate)*1.1)) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5, color = continent_colors[cont]),
    axis.title = element_text(size = 11, face = "bold"),
    axis.text = element_text(size = 9),
    panel.grid = element_line(linetype = "dashed", color = "gray85"),
    panel.border = element_rect(color = "gray70", linewidth = 0.5, fill = NA),
    plot.margin = unit(c(0.2, 0.2, 0.2, 0.2), "cm")
  )
} else {
  p <- ggplot(data.frame()) +
    labs(title = cont, x = "Youth Unemployment Rate (%)", y = "Youth NEET Rate (%)") +
    theme_minimal() +
    theme(plot.title = element_text(size = 14, face = "bold", hjust = 0.5))
}
plot_list[[cont]] <- p
}

# Combine plots into panel
final_plot <- grid.arrange(
  plot_list[["Asia"]], plot_list[["Europe"]], plot_list[["Africa"]],
  plot_list[["North America"]], plot_list[["South America"]], plot_list[["Oceania"]],
  nrow = 2, ncol = 3,
  top = "Relationship between Youth NEET Rate and Unemployment Rate (2015-2020)",
  bottom = "Data Source: Original CSV + Supplementary 2015-2020 Average Data"
)

```

Relationship between Youth NEET Rate and Unemployment Rate (2015–2020)



Data Source: Original CSV + Supplementary 2015–2020 Average Data

```
print(final_plot)
```

```
## TableGrob (4 x 3) "arrange": 8 grobs
##   z      cells   name      grob
## 1 1 (2-2,1-1) arrange  gtable[layout]
## 2 2 (2-2,2-2) arrange  gtable[layout]
## 3 3 (2-2,3-3) arrange  gtable[layout]
## 4 4 (3-3,1-1) arrange  gtable[layout]
## 5 5 (3-3,2-2) arrange  gtable[layout]
## 6 6 (3-3,3-3) arrange  gtable[layout]
## 7 7 (1-1,1-3) arrange text[GRID.text.1050]
## 8 8 (4-4,1-3) arrange text[GRID.text.1051]
```

```
# Print data count
cat("=== Continent Data Count ===\n")
```

```
## === Continent Data Count ===
```

```
print(combined_data %>% filter(!is.na(NEET_rate)) %>% count(Continent, name = "Country_Count"))
```

```
## # A tibble: 6 x 2
##   Continent    Country_Count
##   <chr>         <int>
## 1 Africa             5
## 2 Asia              12
```



```
## 3 Europe          33
## 4 North America   4
## 5 Oceania         4
## 6 South America   3
```

```
# Reuse NEET dataset from Section 1 (df_neet)
colnames(df_neet)[4] <- "NEET_rate"

# Custom continent mapping for trend analysis
custom_continent_trend <- c(
  "United States" = "North America", "Canada" = "North America", "Mexico" = "North America",
  "Costa Rica" = "North America", "Guatemala" = "North America", "Panama" = "North America",
  "Australia" = "Oceania", "New Zealand" = "Oceania", "Fiji" = "Oceania",
  "Brazil" = "South America", "Argentina" = "South America", "Chile" = "South America",
  "China" = "Asia", "India" = "Asia", "Germany" = "Europe", "France" = "Europe",
  "Nigeria" = "Africa", "South Africa" = "Africa",
  "Kosovo" = "Europe", "Micronesia (country)" = "Oceania"
)

# Data cleaning and continent assignment
df_neet_clean_trend <- df_neet %>%
  filter(!grepl("\\(WB\\)|EU|Central Asia|Latin America", Entity)) %>%
  mutate(
    Continent = suppressWarnings(
      countrycode(Entity, origin = "country.name", destination = "continent")
    ),
    Continent = case_when(
      Entity %in% names(custom_continent_trend) ~ custom_continent_trend[Entity],
      TRUE ~ Continent
    )
  )

# Filter target continents and time period, calculate yearly average
target_continents <- c("Asia", "Europe", "Africa", "North America", "South America", "Oceania")
continent_yearly <- df_neet_clean_trend %>%
  filter(
    Continent %in% target_continents,
    Year >= 2015 & Year <= 2020,
    !is.na(NEET_rate)
  ) %>%
  group_by(Continent, Year) %>%
  summarise(NEET_rate = mean(NEET_rate, na.rm = TRUE), .groups = "drop")

# Print unique continents
print(unique(continent_yearly$Continent))
```

```
## [1] "Africa"          "Asia"            "Europe"          "North America"
## [5] "Oceania"         "South America"
```

```
# Create NEET trend plot
p <- ggplot(continent_yearly, aes(x = Year, y = NEET_rate, color = Continent, group = Continent)) +
  geom_line(linewidth = 1.5) +
  geom_point(size = 3, alpha = 0.8) +
```

```

scale_x_continuous(breaks = 2015:2020, limits = c(2015, 2020)) +
scale_color_manual(values = c(
  "Asia" = "#E63946", "Europe" = "#1D3557",
  "Africa" = "#F77F00", "North America" = "#0077B6",
  "South America" = "#2A9D8F", "Oceania" = "#9C27B0"
)) +
labs(
  title = "NEET Trends (2015-2020) - Six Continents",
  subtitle = "Youth NEET Rate by Continent",
  x = "Year",
  y = "NEET Rate (%)",
  color = "Continent"
) +
theme_minimal() +
theme(
  plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
  plot.subtitle = element_text(size = 12, hjust = 0.5),
  legend.position = "bottom",
  legend.key.size = unit(1.2, "cm")
)
print(p)

```

