IDS Group Work

Jing Qian Low

2025-12-01

```
# Load libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readr)
library(tidyr)

# Set working directories
setwd("~/Library/Mobile Documents/com~apple~CloudDocs/IDS Group Assignment/data sets")
continents <- read.csv("./continents-according-to-our-world-in-data.csv")
gdp <- read.csv("./gdp-per-capita-worldbank.csv")
youth <- read.csv("./youth-not-in-education-employment-training.csv")
```

Target 1

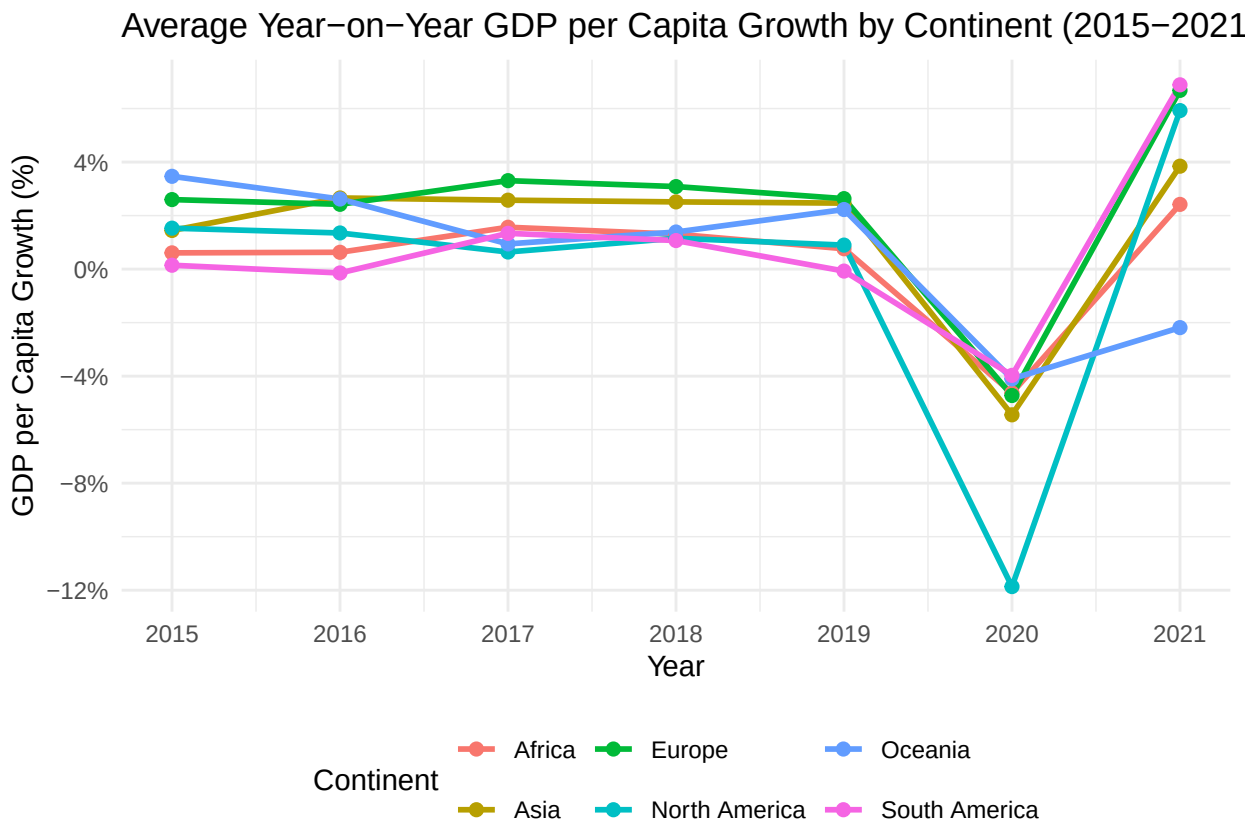
Average Year-on-Year GDP per Capita Growth by Continent

```
continents_clean <- continents %>%
  select(Entity, Code, Continent)

# Keep needed columns and rename GDP per capita
gdp_cont <- gdp %>%
  select(
    Entity, Code, Year,
    gdp_pc = `GDP.per.capita..PPP..constant.2017.international...`
  ) %>%
  left_join(continents_clean, by = c("Entity", "Code")) %>%
```

```r
  # Drop Antarctica and any rows with missing continent info
  filter(!is.na(Continent), Continent != "Antarctica") %>%
  arrange(Entity, Year)

# Compute year-on-year GDP per capita growth for each country
gdp_growth <- gdp_cont %>%
  group_by(Entity) %>%
  mutate(growth_pct = (gdp_pc - lag(gdp_pc)) / lag(gdp_pc) * 100) %>%
  ungroup() %>%
  filter(Year >= 2015, Year <= 2021, !is.na(growth_pct))

# mean growth across countries in each continent
continent_growth <- gdp_growth %>%
  group_by(Continent, Year) %>%
  summarise(
    mean_growth = mean(growth_pct, na.rm = TRUE),
    .groups = "drop"
  )

# Plot graph
p_continents <- ggplot(
  continent_growth,
  aes(x = Year, y = mean_growth, color = Continent, group = Continent)
) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  scale_x_continuous(breaks = 2015:2021) +
  scale_y_continuous(labels = function(x) paste0(round(x, 1), "%")) +
  labs(
    title = "Average Year-on-Year GDP per Capita Growth by Continent (2015-2021)",
    x = "Year",
    y = "GDP per Capita Growth (%)",
    color = "Continent"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    legend.position  = "bottom",
    legend.direction = "horizontal",
    axis.text.x      = element_text(angle = 0)
  )
```

```
## 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.
```

```r
print(p_continents)
```

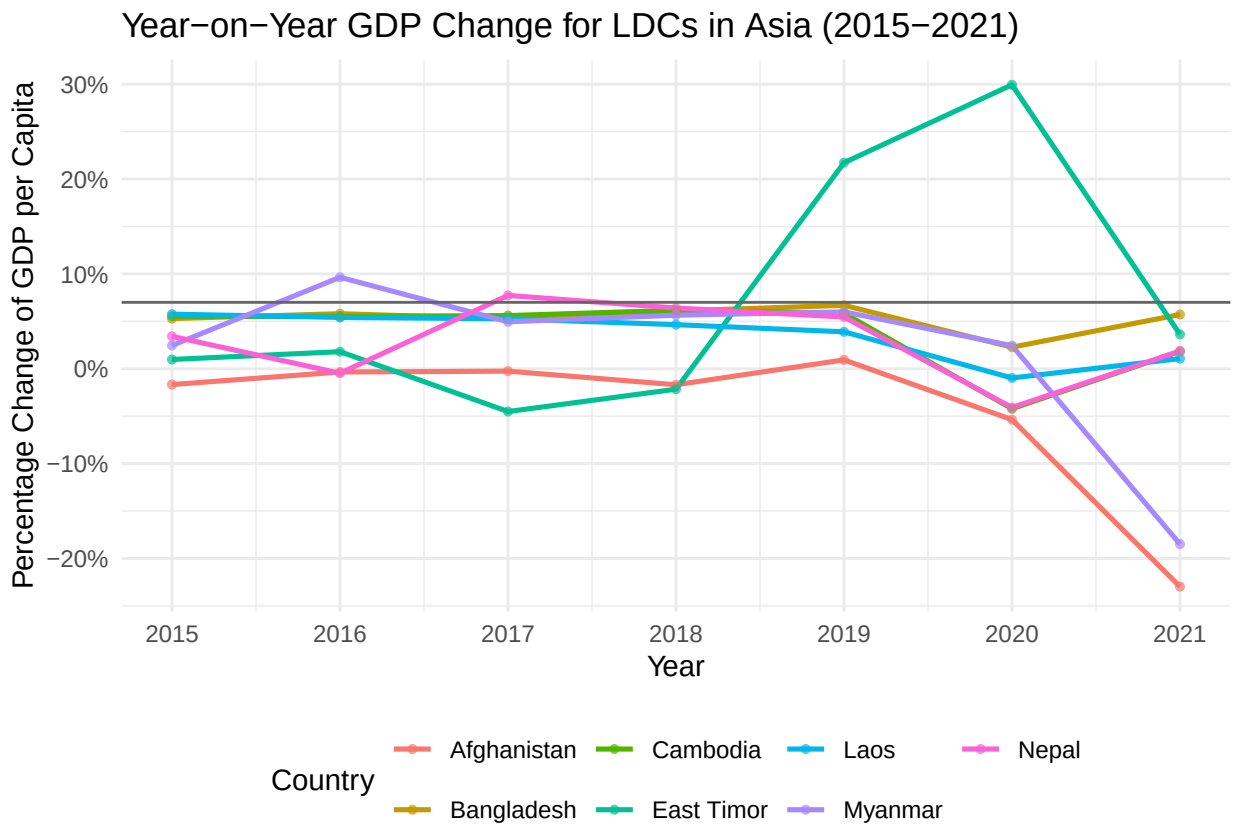



```
ggsave(
  "continents_gdp_pc_growth_2015_2021.png",
  p_continents,
  width = 11,
  height = 6,
  dpi = 300
)
```

LDC GDP Change According to Continent

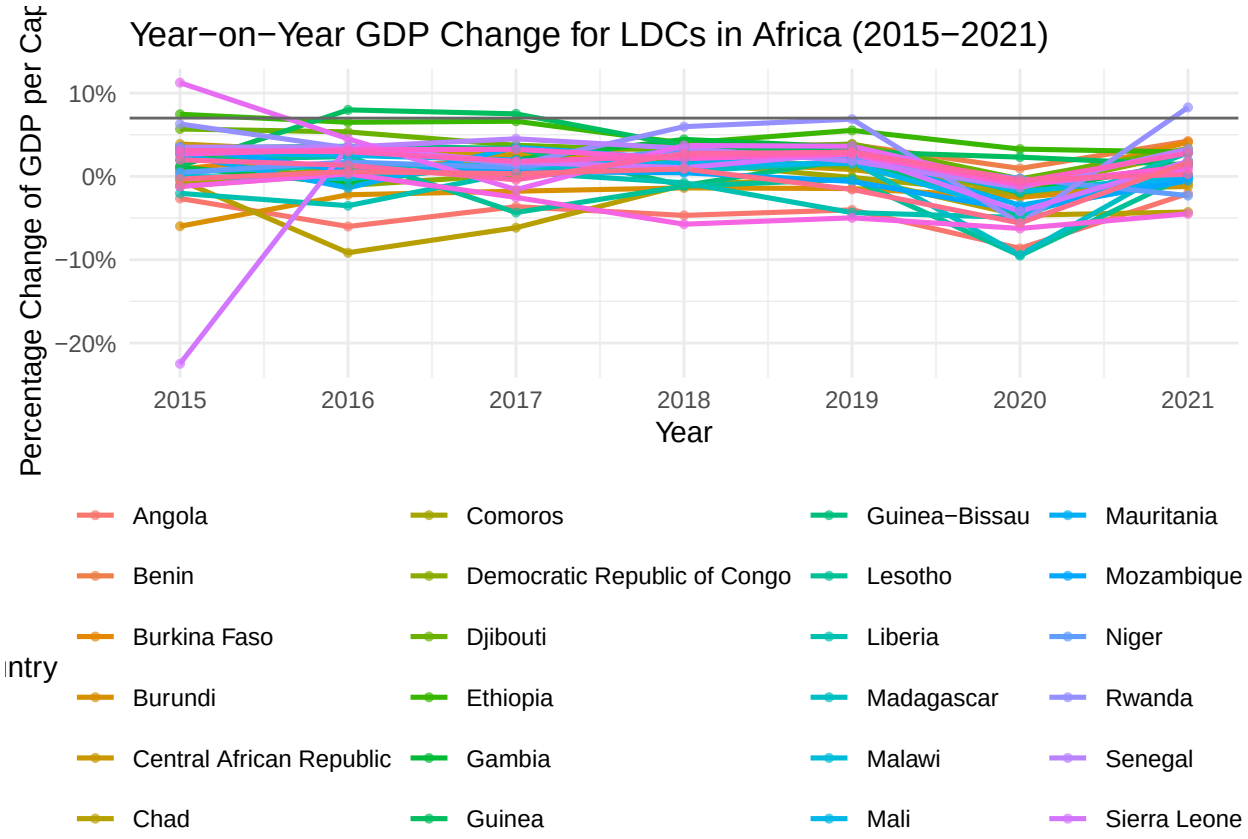
```
##Asia
# Keep needed columns and rename the GDP column
gdp_asia <- gdp %>%
  select(
    Entity, Code, Year,
    gdp_pc = `GDP.per.capita..PPP..constant.2017.international...`
  ) %>%
  arrange(Entity, Year)

# Asian LDC ISO-3 codes
asia_ldc_codes <- c(
  "AFG", # Afghanistan
  "BGD", # Bangladesh
  "KHM", # Cambodia
  "LAO", # Lao People's Democratic Republic
  "MMR", # Myanmar
```

```
  "NPL", # Nepal
  "TLS", # Timor-Leste
  "YEM"  # Yemen
)

asia_ldc <- gdp_asia %>%
  filter(Code %in% asia_ldc_codes)

# Year-on-year percent change
asia_ldc_growth <- asia_ldc %>%
  group_by(Entity) %>%
  mutate(growth_pct = (gdp_pc - lag(gdp_pc)) / lag(gdp_pc) * 100) %>%
  ungroup() %>%
  filter(Year >= 2015, Year <= 2021, !is.na(growth_pct))

# Line graph (one line per LDC country)
p_asia <- ggplot(
  asia_ldc_growth,
  aes(x = Year, y = growth_pct, color = Entity, group = Entity)
) +
  geom_line(size = 0.9) +
  geom_point(size = 1.1, alpha = 0.7) +
  scale_x_continuous(breaks = 2015:2021) +
  scale_y_continuous(labels = function(x) paste0(round(x, 1), "%")) +
  labs(
    title = "Year-on-Year GDP Change for LDCs in Asia (2015-2021)",
    x = "Year",
    y = "Percentage Change of GDP per Capita",
    color = "Country"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    legend.position  = "bottom",
    legend.direction = "horizontal",
    axis.text.x      = element_text(angle = 0)
  ) +
  geom_hline(yintercept = 7, linetype = "solid", color = "grey40")

print(p_asia)
```

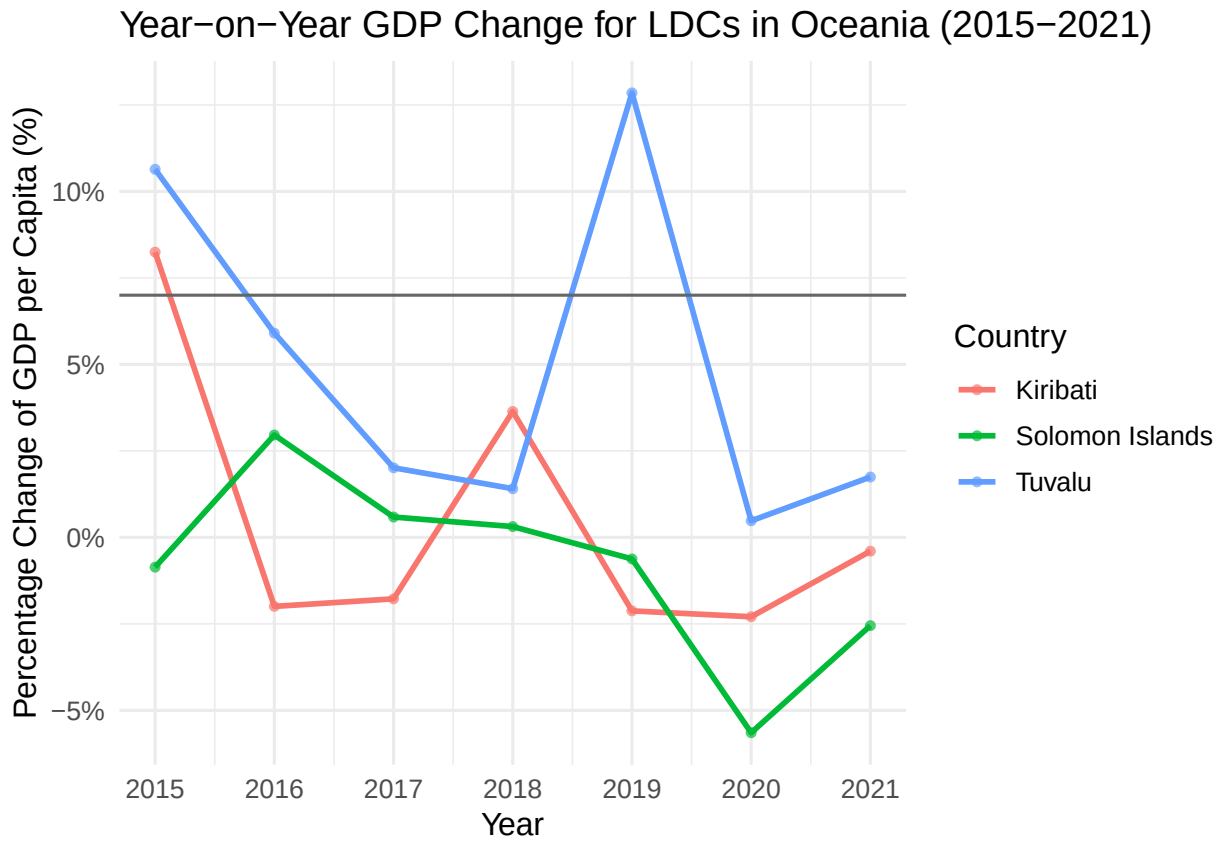



```
ggsave("asia_ldc_gdp_change_2015_2021.png",
       p_asia, width = 11, height = 6, dpi = 300)

## Africa
# Keep needed columns and rename the GDP column
gdp_africa <- gdp %>%
  select(
    Entity, Code, Year,
    gdp_pc = `GDP.per.capita..PPP..constant.2017.international...`
  ) %>%
  arrange(Entity, Year)

# African LDC ISO-3 codes (UN list)
africa_ldc_codes <- c(
  "AGO","BEN","BFA","BDI","CAF","TCD","COM","COD","DJI","ERI","ETH","GMB",
  "GIN","GNB","LSO","LBR","MDG","MWI","MLI","MRT","MOZ","NER","RWA","SEN",
  "SLE","SOM","SSD","SDN","TZA","TGO","UGA","ZMB"
)

africa_ldc <- gdp_africa %>%
  filter(Code %in% africa_ldc_codes)

# Year-on-year percent change
africa_ldc_growth <- africa_ldc %>%
  group_by(Entity) %>%
  mutate(growth_pct = (gdp_pc - lag(gdp_pc)) / lag(gdp_pc) * 100) %>%
```

```r
  ungroup() %>%
  filter(Year >= 2015, Year <= 2021, !is.na(growth_pct))

# Line graph (one line per LDC country)
p_africa <- ggplot(africa_ldc_growth,
                   aes(x = Year, y = growth_pct, color = Entity, group = Entity)) +
  geom_line(size = 0.9) +
  geom_point(size = 1.1, alpha = 0.7) +
  scale_x_continuous(breaks = 2015:2021) +
  scale_y_continuous(labels = function(x) paste0(round(x, 1), "%")) +
  labs(
    title = "Year-on-Year GDP Change for LDCs in Africa (2015-2021)",
    x = "Year",
    y = "Percentage Change of GDP per Capita",
    color = "Country"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    legend.position  = "bottom",
    legend.direction = "horizontal",
    axis.text.x      = element_text(angle = 0)
  ) +
  geom_hline(yintercept = 7, linetype = "solid", color = "grey40")

print(p_africa)
```

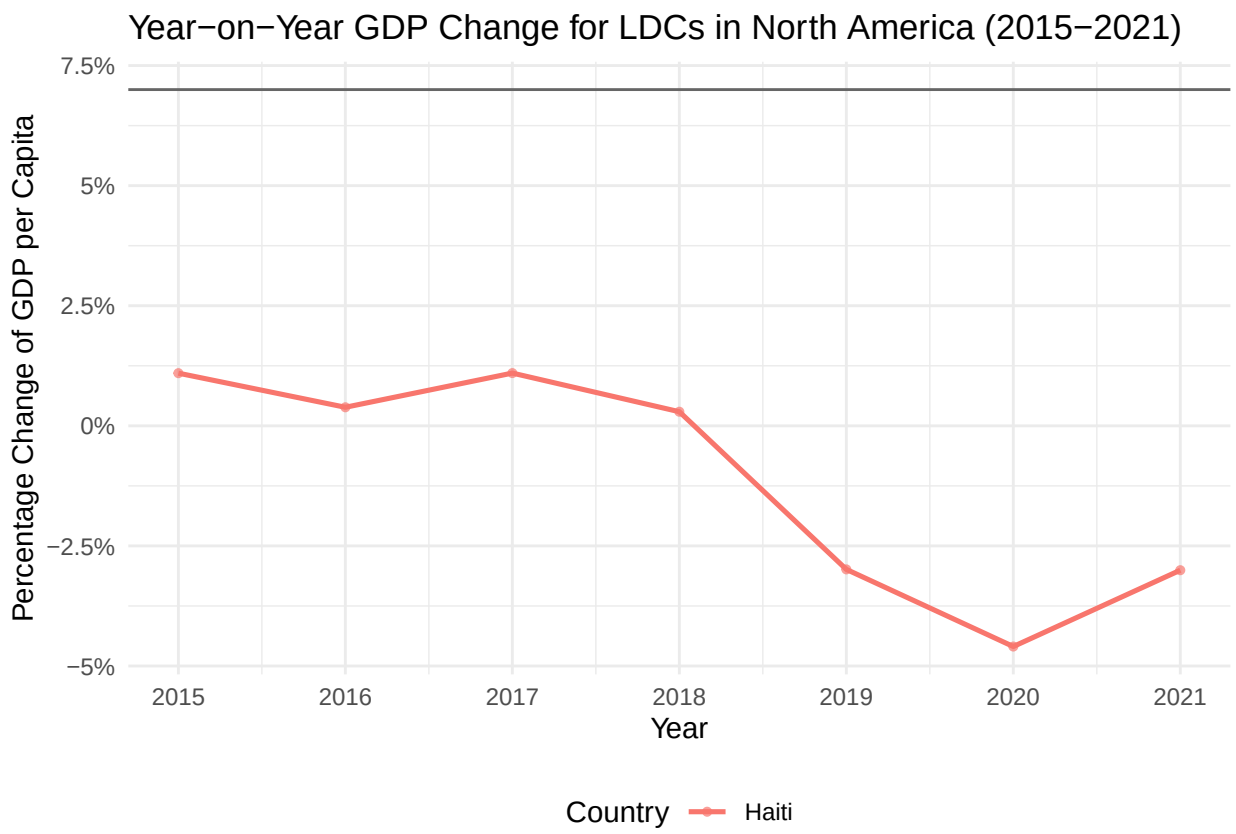


```r
ggsave("africa_ldc_gdp_change_2015_2021.png", p_africa, width = 11, height = 6, dpi = 300)

## Oceania
# Keep needed columns and rename the GDP column
gdp_oceania <- gdp %>%
  select(Entity, Code, Year, gdp_pc = GDP.per.capita..PPP..constant.2017.international...) %>%
  arrange(Entity, Year)

# African LDC ISO-3 codes
oceania_ldc_codes <- c("KIR", "SLB", "TUV")

gdp_oceania_ldc <- gdp_oceania %>%
  filter(Code %in% oceania_ldc_codes)

# Year-on-year percent change
plot_df <- gdp_oceania_ldc %>%
  group_by(Entity) %>%
  mutate(
    growth_yoy = (gdp_pc - lag(gdp_pc)) / lag(gdp_pc),
    growth_pct = growth_yoy * 100
  ) %>%
  ungroup() %>%
  filter(Year >= 2015, Year <= 2021, !is.na(growth_pct))

# Line graph (one line per LDC country)
p <- ggplot(plot_df, aes(x = Year, y = growth_pct, color = Entity, group = Entity)) +
  geom_line(size = 1) +
  geom_point(size = 1.2, alpha = 0.7) +
  scale_x_continuous(breaks = 2015:2021) +
  scale_y_continuous(labels = function(x) paste0(round(x, 1), "%")) +
  labs(
    title = "Year-on-Year GDP Change for LDCs in Oceania (2015-2021)",
    x = "Year",
    y = "Percentage Change of GDP per Capita (%)",
    color = "Country"
  ) +
  theme_minimal(base_size = 12) +
  geom_hline(yintercept = 7, linetype = "solid", color = "grey40")

print(p)
```




```
ggsave(
  filename = "oceania_ldc_gdp_change_2015_2021.png",
  plot = p,
  width = 9, height = 5
)


##North America
# Keep needed columns and rename the GDP column
gdp_na <- gdp %>%
  select(
    Entity, Code, Year,
    gdp_pc = `GDP.per.capita..PPP..constant.2017.international...`
  ) %>%
  arrange(Entity, Year)

# North American LDC ISO-3 codes (Haiti only)
na_ldc_codes <- c("HTI")  # Haiti

na_ldc <- gdp_na %>%
  filter(Code %in% na_ldc_codes)

# Year-on-year percent change
na_ldc_growth <- na_ldc %>%
  group_by(Entity) %>%
  mutate(growth_pct = (gdp_pc - lag(gdp_pc)) / lag(gdp_pc) * 100) %>%
```

```r
  ungroup() %>%
  filter(Year >= 2015, Year <= 2021, !is.na(growth_pct))

# Line graph (Haiti only)
p_na <- ggplot(
  na_ldc_growth,
  aes(x = Year, y = growth_pct, color = Entity, group = Entity)
) +
  geom_line(size = 0.9) +
  geom_point(size = 1.1, alpha = 0.7) +
  scale_x_continuous(breaks = 2015:2021) +
  scale_y_continuous(labels = function(x) paste0(round(x, 1), "%")) +
  labs(
    title = "Year-on-Year GDP Change for LDCs in North America (2015-2021)",
    x = "Year",
    y = "Percentage Change of GDP per Capita",
    color = "Country"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    legend.position  = "bottom",
    legend.direction = "horizontal",
    axis.text.x      = element_text(angle = 0)
  ) +
  geom_hline(yintercept = 7, linetype = "solid", color = "grey40")

print(p_na)
```

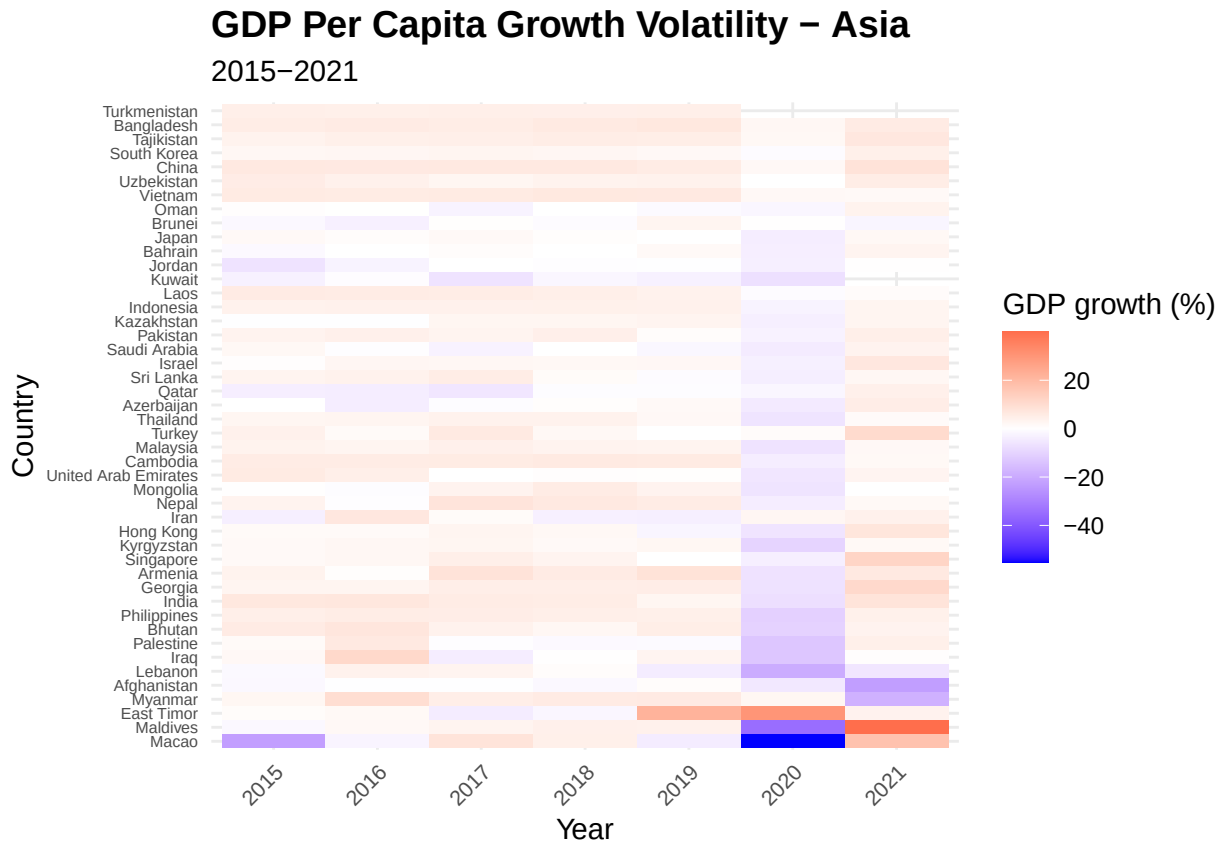



```
ggsave("north_america_ldc_gdp_change_2015_2021.png",
       p_na, width = 11, height = 6, dpi = 300)
```

Volatility

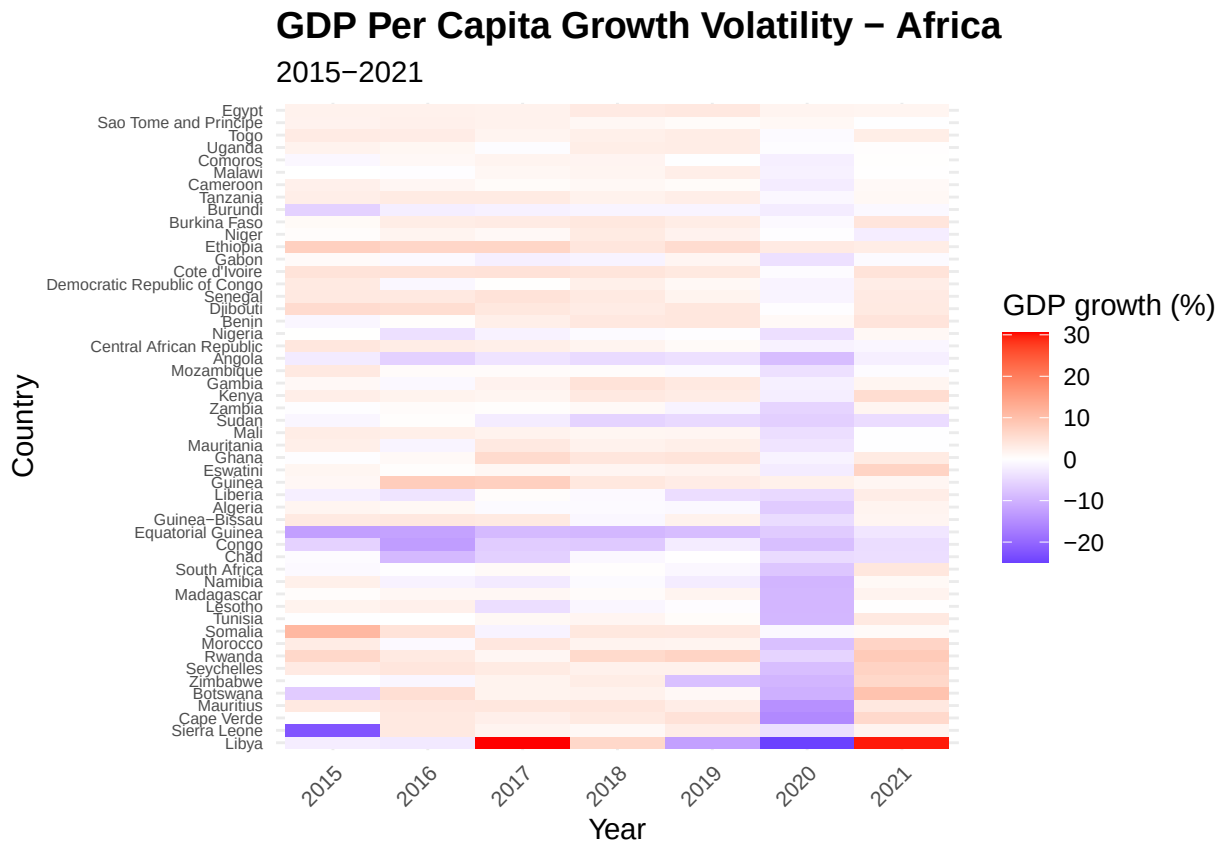
```
# Rename important columns
gdp_clean <- gdp %>%
  rename(
    country = Entity,
    code    = Code,
    year    = Year,
    gdp_per_capita = GDP.per.capita..PPP..constant.2017.international...  # <-- CHANGE THIS
  ) %>%
  filter(!is.na(gdp_per_capita))

# Compute year-on-year GDP per capita growth (%) for each country
gdp_growth <- gdp_clean %>%
  arrange(country, year) %>%
  group_by(country) %>%
  mutate(
    gdp_growth = (gdp_per_capita / lag(gdp_per_capita) - 1) * 100
  ) %>%
  ungroup()

continents_clean <- continents %>%
```

```r
  rename(
    country  = Entity,
    continent = Continent
  )

gdp_growth_cont <- gdp_growth %>%
  left_join(continents_clean, by = "country") %>%
  filter(!is.na(continent)) %>%
  filter(year >= 2015)

# Heatmap- Asia
asia <- gdp_growth_cont %>%
  filter(continent == "Asia", year >= 2015)

# Order countries by volatility (std dev of growth)
asia <- asia %>%
  group_by(country) %>%
  mutate(volatility = sd(gdp_growth, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(volatility)) %>%
  mutate(country = factor(country, levels = unique(country)))

# Plot heatmap
p_asia <- ggplot(asia, aes(x = factor(year), y = country, fill = gdp_growth)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "blue",
    mid = "white",
    high = "red",
    midpoint = 0,
    na.value = "grey90",
    name = "GDP growth (%)"
  ) +
  labs(
    title = "GDP Per Capita Growth Volatility - Asia",
    subtitle = "2015-2021",
    x = "Year",
    y = "Country"
  ) +
  theme_minimal() +
  theme(
    axis.text.y = element_text(size = 6),
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    plot.title  = element_text(face = "bold", size = 14)
  )

print(p_asia)
```



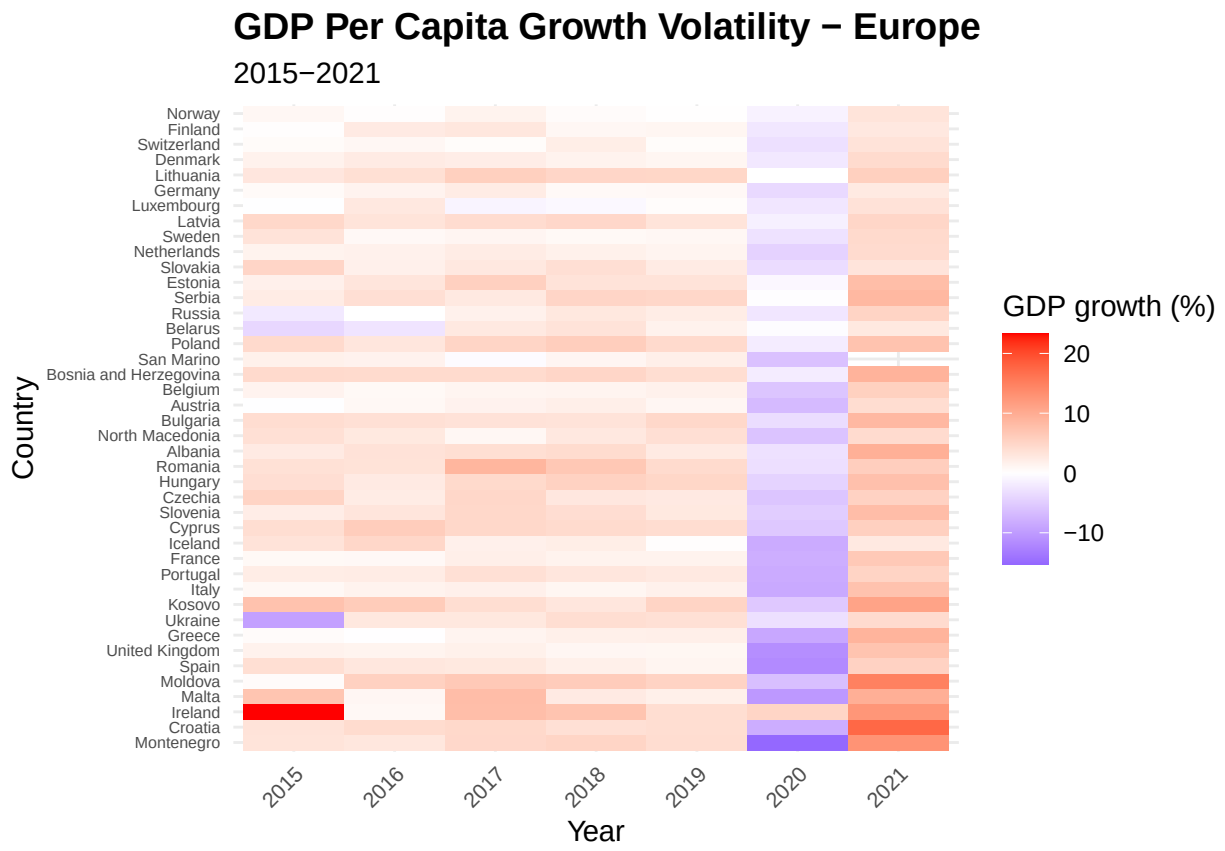

```r
ggsave(
  filename = "heatmap_volatility_Asia.png",
  plot = p_asia,
  width = 10,
  height = max(6, n_distinct(asia$country) * 0.25),   # auto-adjust height
  dpi = 300,
  units = "in"
)

#Heatmap - Africa
africa <- gdp_growth_cont %>%
  filter(continent == "Africa", year >= 2015)

# Order countries by volatility (std dev of growth)
africa <- africa %>%
  group_by(country) %>%
  mutate(volatility = sd(gdp_growth, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(volatility)) %>%
  mutate(country = factor(country, levels = unique(country)))

# Plot heatmap
p_africa <- ggplot(africa, aes(x = factor(year), y = country, fill = gdp_growth)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "blue",
```

```
    mid = "white",
    high = "red",
    midpoint = 0,
    na.value = "grey90",
    name = "GDP growth (%)"
  ) +
  labs(
    title = "GDP Per Capita Growth Volatility - Africa",
    subtitle = "2015-2021",
    x = "Year",
    y = "Country"
  ) +
  theme_minimal() +
  theme(
    axis.text.y = element_text(size = 6),
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    plot.title  = element_text(face = "bold", size = 14)
  )

print(p_africa)
```



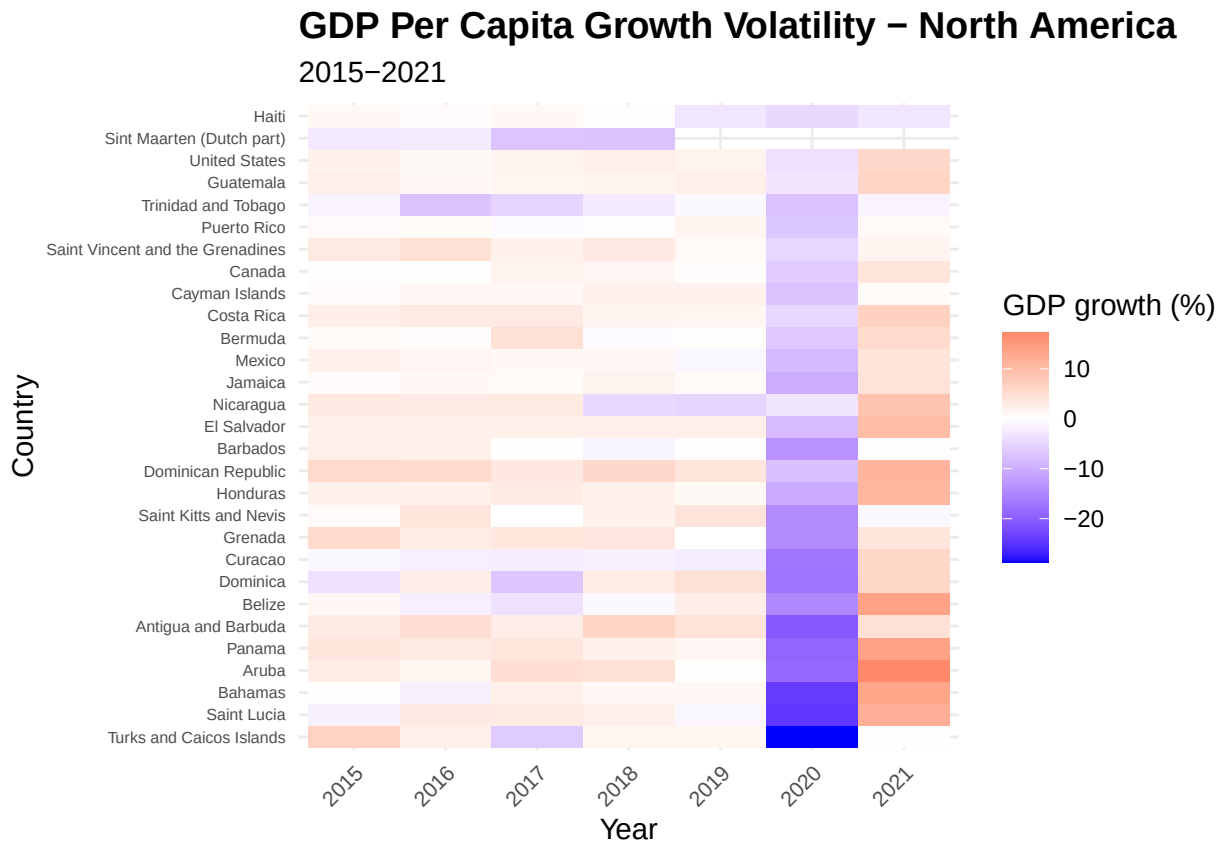

```
ggsave(
  filename = "heatmap_volatility_Africa.png",
  plot = p_africa,
  width = 10,
```

```
  height = max(6, n_distinct(asia$country) * 0.25),   # auto-adjust height
  dpi = 300,
  units = "in"
)

#Heatmap - Europe
europe <- gdp_growth_cont %>%
  filter(continent == "Europe", year >= 2015)

# Order countries by volatility (std dev of growth)
europe <- europe %>%
  group_by(country) %>%
  mutate(volatility = sd(gdp_growth, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(volatility)) %>%
  mutate(country = factor(country, levels = unique(country)))

# Plot heatmap
p_europe <- ggplot(europe, aes(x = factor(year), y = country, fill = gdp_growth)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "blue",
    mid = "white",
    high = "red",
    midpoint = 0,
    na.value = "grey90",
    name = "GDP growth (%)"
  ) +
  labs(
    title = "GDP Per Capita Growth Volatility - Europe",
    subtitle = "2015-2021",
    x = "Year",
    y = "Country"
  ) +
  theme_minimal() +
  theme(
    axis.text.y = element_text(size = 6),
    axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
    plot.title  = element_text(face = "bold", size = 14)
  )

print(p_europe)
```



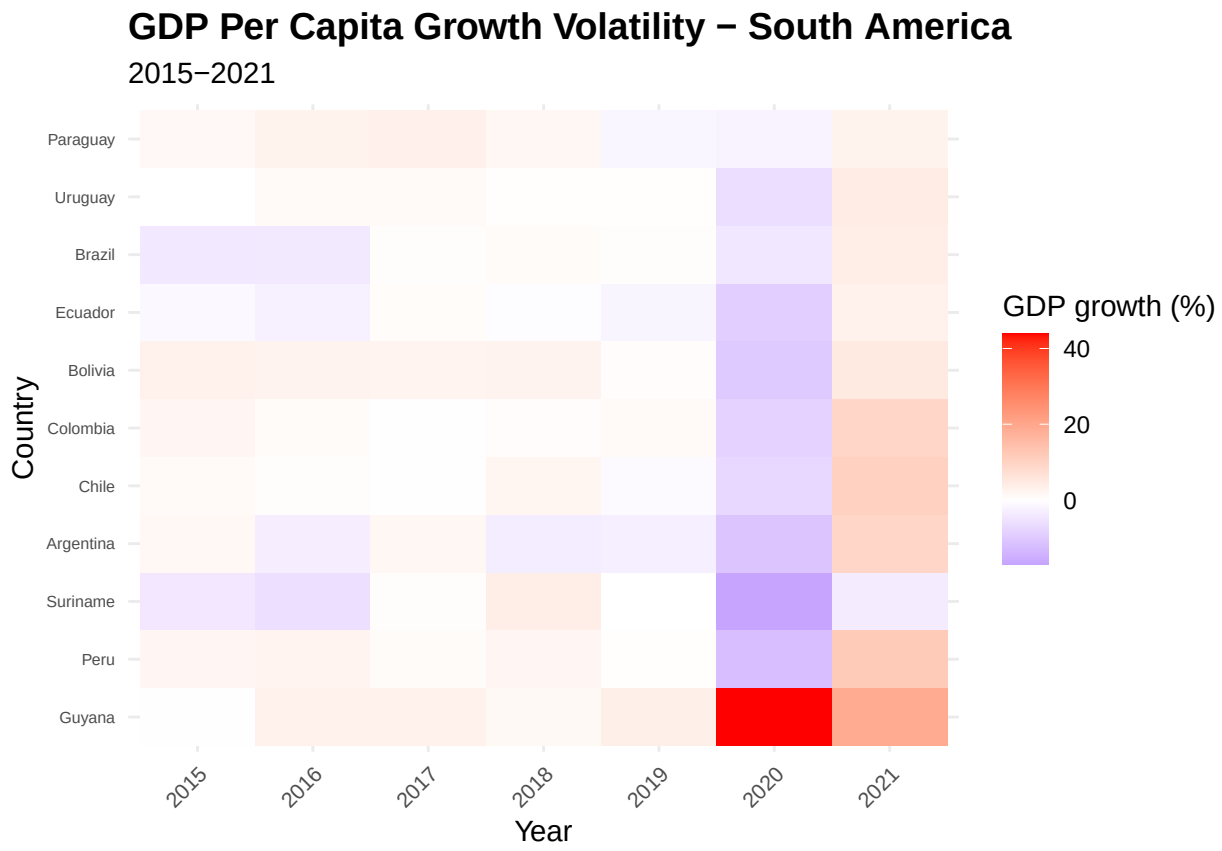

```r
ggsave(
  filename = "heatmap_volatility_Europe.png",
  plot = p_asia,
  width = 10,
  height = max(6, n_distinct(asia$country) * 0.25),   # auto-adjust height
  dpi = 300,
  units = "in"
)

#Heatmap - north america
north_america <- gdp_growth_cont %>%
  filter(continent == "North America", year >= 2015)

# Order countries by volatility (std dev of growth)
north_america <- north_america %>%
  group_by(country) %>%
  mutate(volatility = sd(gdp_growth, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(volatility)) %>%
  mutate(country = factor(country, levels = unique(country)))

# Plot heatmap
p_north_america <- ggplot(north_america, aes(x = factor(year), y = country, fill = gdp_growth)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "blue",
```

```
    mid = "white",
    high = "red",
    midpoint = 0,
    na.value = "grey90",
    name = "GDP growth (%)"
  ) +
  labs(
    title = "GDP Per Capita Growth Volatility - North America",
    subtitle = "2015-2021",
    x = "Year",
    y = "Country"
  ) +
  theme_minimal() +
  theme(
    axis.text.y  = element_text(size = 6),
    axis.text.x  = element_text(angle = 45, hjust = 1, size = 8),
    plot.title   = element_text(face = "bold", size = 14)
  )

print(p_north_america)
```



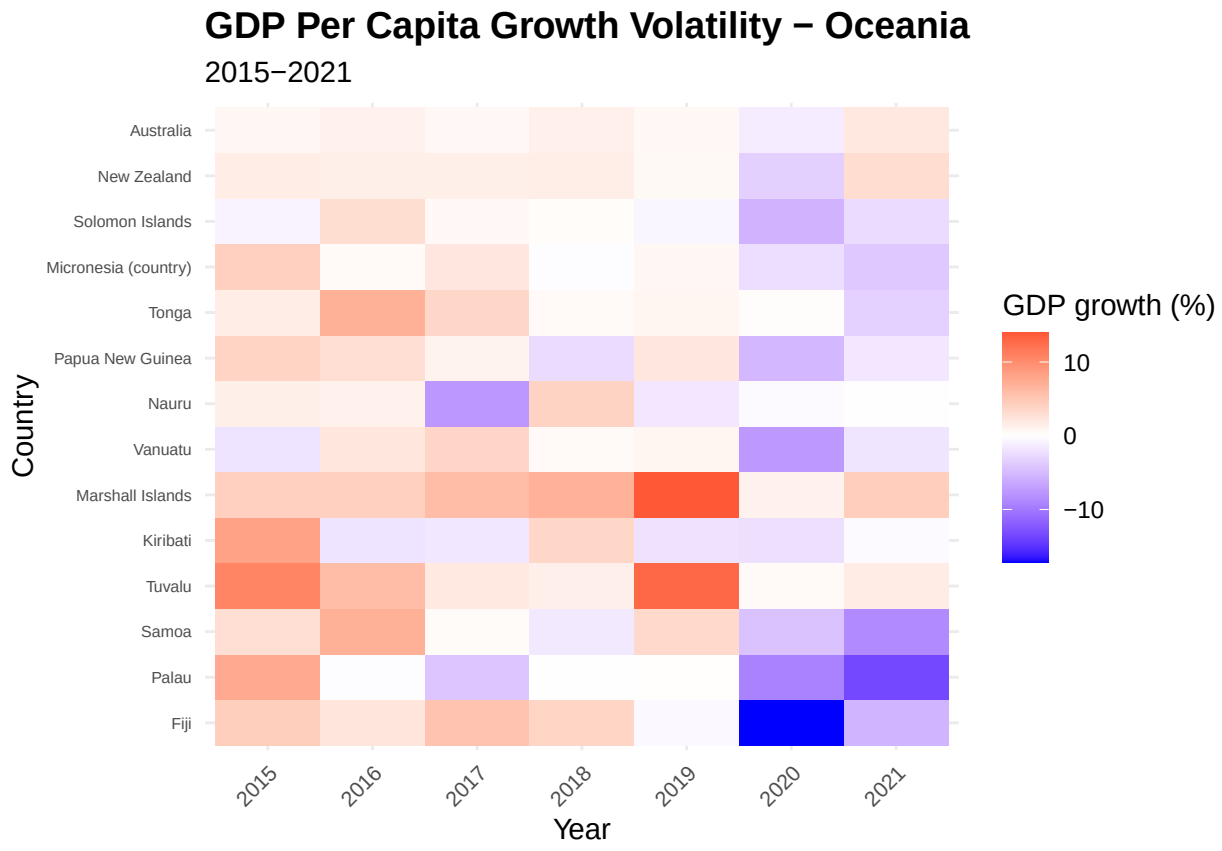

```
ggsave(
  filename = "heatmap_volatility_North_America.png",
  plot = p_north_america,
  width = 10,
```

```
  height = max(6, n_distinct(north_america$country) * 0.25),   # auto-adjust height
  dpi = 300,
  units = "in"
)

#heatmap - south america
south_america <- gdp_growth_cont %>%
  filter(continent == "South America", year >= 2015)

# Order countries by volatility
south_america <- south_america %>%
  group_by(country) %>%
  mutate(volatility = sd(gdp_growth, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(volatility)) %>%
  mutate(country = factor(country, levels = unique(country)))

# Plot heatmap
p_south_america <- ggplot(south_america, aes(x = factor(year), y = country, fill = gdp_growth)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "blue",
    mid = "white",
    high = "red",
    midpoint = 0,
    na.value = "grey90",
    name = "GDP growth (%)"
  ) +
  labs(
    title = "GDP Per Capita Growth Volatility - South America",
    subtitle = "2015-2021",
    x = "Year",
    y = "Country"
  ) +
  theme_minimal() +
  theme(
    axis.text.y  = element_text(size = 6),
    axis.text.x  = element_text(angle = 45, hjust = 1, size = 8),
    plot.title   = element_text(face = "bold", size = 14)
  )

print(p_south_america)
```




```r
ggsave(
  filename = "heatmap_volatility_South_America.png",
  plot = p_south_america,
  width = 10,
  height = max(6, n_distinct(south_america$country) * 0.25),   # auto-adjust height
  dpi = 300,
  units = "in"
)

#heatmap - oceania
oceania <- gdp_growth_cont %>%
  filter(continent == "Oceania", year >= 2015)

# Order countries by volatility
oceania <- oceania %>%
  group_by(country) %>%
  mutate(volatility = sd(gdp_growth, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(volatility)) %>%
  mutate(country = factor(country, levels = unique(country)))

# Plot heatmap
p_oceania <- ggplot(oceania, aes(x = factor(year), y = country, fill = gdp_growth)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "blue",
```

```
    mid = "white",
    high = "red",
    midpoint = 0,
    na.value = "grey90",
    name = "GDP growth (%)"
  ) +
  labs(
    title = "GDP Per Capita Growth Volatility - Oceania",
    subtitle = "2015-2021",
    x = "Year",
    y = "Country"
  ) +
  theme_minimal() +
  theme(
    axis.text.y  = element_text(size = 6),
    axis.text.x  = element_text(angle = 45, hjust = 1, size = 8),
    plot.title   = element_text(face = "bold", size = 14)
  )

print(p_oceania)
```




```
ggsave(
  filename = "heatmap_volatility_Oceania.png",
  plot = p_oceania,
  width = 10,
```

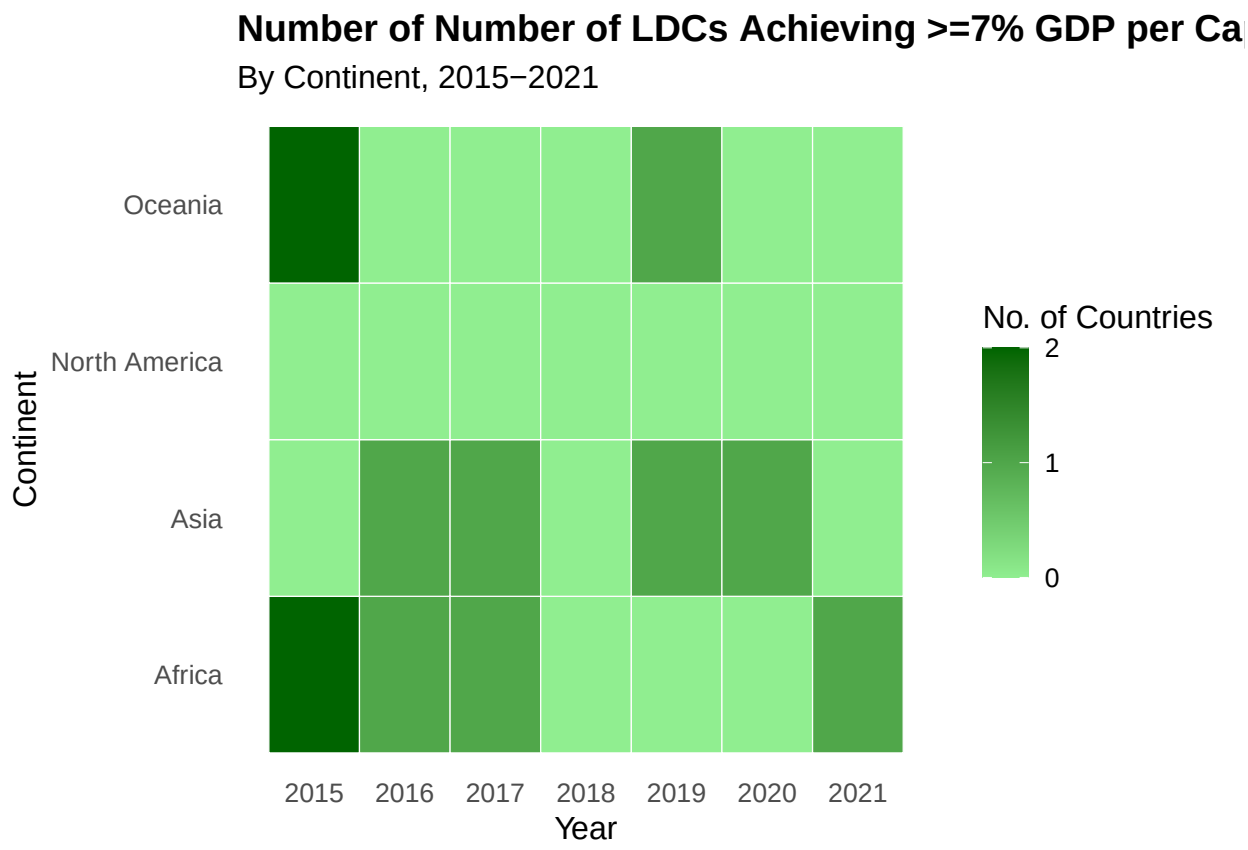
```
  height = max(6, n_distinct(oceania$country) * 0.25),   # auto-adjust height
  dpi = 300,
  units = "in"
)
```

Time Series Heatmap

```
# LDC ISO codes by continent

africa_ldc_codes <- c(
  "AGO","BEN","BFA","BDI","CAF","TCD","COM","COD","DJI","ERI","ETH","GMB",
  "GIN","GNB","LSO","LBR","MDG","MWI","MLI","MRT","MOZ","NER","RWA","SEN",
  "SLE","SOM","SSD","SDN","TZA","TGO","UGA","ZMB"
)

asia_ldc_codes <- c(
  "AFG","BGD","KHM","LAO","MMR","NPL","TLS","YEM"
)

north_america_ldc_codes <- c("HTI")  # Haiti

oceania_ldc_codes <- c(
  "KIR",  # Kiribati
  "SLB",  # Solomon Islands
  "TUV"   # Tuvalu
)

continents_clean <- continents %>%
  select(Entity, Code, Continent)

gdp_cont <- gdp %>%
  select(
    Entity, Code, Year,
    gdp_pc = `GDP.per.capita..PPP..constant.2017.international...`
  ) %>%
  left_join(continents_clean, by = c("Entity", "Code")) %>%
  # Drop Antarctica and any rows with missing continent info
  filter(!is.na(Continent), Continent != "Antarctica") %>%
  arrange(Entity, Year)

# Compute year-on-year GDP per capita growth for each country
gdp_growth <- gdp_cont %>%
  group_by(Entity) %>%
  mutate(growth_pct = (gdp_pc - lag(gdp_pc)) / lag(gdp_pc) * 100) %>%
  ungroup() %>%
  filter(Year >= 2015, Year <= 2021, !is.na(growth_pct))

# Tag LDCs with continent and filter for SDG period (2015-2021)

ldc_growth <- gdp_growth %>%   # must contain: Code, Year, growth_pct
  mutate(Continent_LDC = case_when(
    Code %in% africa_ldc_codes        ~ "Africa",
    Code %in% asia_ldc_codes          ~ "Asia",
```

```
    Code %in% north_america_ldc_codes ~ "North America",
    Code %in% oceania_ldc_codes       ~ "Oceania",
    TRUE                              ~ NA_character_
  )) %>%
  filter(!is.na(Continent_LDC),
         Year >= 2015, Year <= 2021)

# Count number of LDCs hitting >= 7% GDP per capita growth

ldc_count_7pct <- ldc_growth %>%
  group_by(Continent_LDC, Year) %>%
  summarise(
    num_ldc_over7 = sum(growth_pct >= 7, na.rm = TRUE),
    .groups = "drop"
  )

# Order the continents nicely for the heatmap
ldc_count_7pct$Continent_LDC <- factor(
  ldc_count_7pct$Continent_LDC,
  levels = c("Africa", "Asia", "North America", "Oceania")
)

# Plot Time-Series Heatmap

p_ldc_heat <- ggplot(ldc_count_7pct,
                     aes(x = Year,
                         y = Continent_LDC,
                         fill = num_ldc_over7)) +
  geom_tile(color = "white") +
  scale_x_continuous(breaks = 2015:2021) +

  labs(
    title    = "Number of Number of LDCs Achieving >=7% GDP per Capita Growth",
    subtitle = "By Continent, 2015-2021",
    x        = "Year",
    y        = "Continent",
    fill     = "No. of Countries"
  ) +

  theme_minimal(base_size = 12) +
  theme(
    panel.grid      = element_blank(),
    axis.text.x     = element_text(angle = 0),
    plot.title      = element_text(face = "bold"),
    legend.position = "right"
  )+
  scale_fill_gradient(
    low = "lightgreen",
    high = "darkgreen",
    breaks = 0:max(ldc_count_7pct$num_ldc_over7),
    labels = 0:max(ldc_count_7pct$num_ldc_over7)
  )
```

```
print(p_ldc_heat)
```




```
ggsave("ldc_7pct_heatmap_2015_2021_all_continents.png",
       p_ldc_heat, width = 10, height = 5.5, dpi = 300)
```

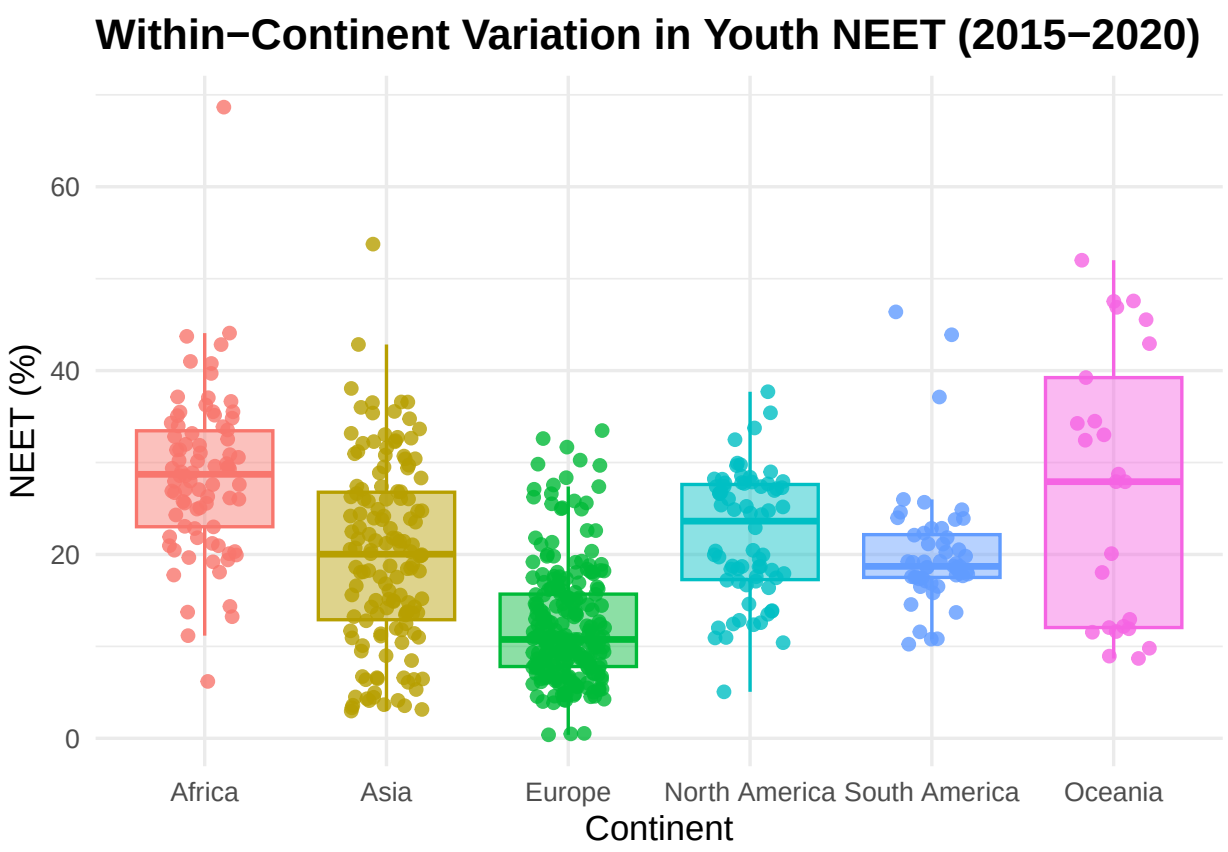
Target 2

box plot NEET

```
# rename NEET column
youth_neet <- youth %>%
  select(
    Entity,
    Code,
    Year,
    neet_pct = "Share.of.youth.not.in.education..employment.or.training..total....of.youth.population."
  )

continents_clean <- continents %>%
  select(Entity, Code, Continent)

# Join NEET with continent, remove Antarctica
neet_with_cont <- youth_neet %>%
```

```r
  left_join(continents_clean, by = c("Entity", "Code")) %>%
  filter(!is.na(Continent), Continent != "Antarctica")

# Filter to SDG period (2015-2020)
neet_2015_2020 <- neet_with_cont %>%
  filter(Year >= 2015, Year <= 2020, !is.na(neet_pct))

# Optional: set a nice continent order
neet_2015_2020$Continent <- factor(
  neet_2015_2020$Continent,
  levels = c("Africa", "Asia", "Europe", "North America", "South America", "Oceania")
)

# Boxplot: within-continent NEET variation (2015-2020 pooled)
p_neet_box <- ggplot(neet_2015_2020,
                     aes(x = Continent, y = neet_pct, fill = Continent)) +
  geom_boxplot(alpha = 0.7, outlier.alpha = 0.5) +
  labs(
    title = "Within-Continent Variation in Youth NEET (2015-2020)",
    subtitle = "Distribution of NEET (% of youth 15-24) by continent",
    x = "Continent",
    y = "NEET (%)"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    legend.position = "none",
    plot.title = element_text(face = "bold"),
    axis.text.x = element_text(size = 11)
  )

p_neet_box_points <- ggplot(neet_2015_2020,
       aes(x = Continent,
           y = neet_pct,
           fill = Continent,
           color = Continent)) +

  # Boxplot (median, IQR, whiskers)
  geom_boxplot(alpha = 0.45, outlier.shape = NA) +

  # Points (jitter), using same continent colors
  geom_jitter(width = 0.2, size = 1.8, alpha = 0.8) +

  labs(
    title = "Within-Continent Variation in Youth NEET (2015-2020)",
    x = "Continent",
    y = "NEET (%)"
  ) +

  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold"),
    legend.position = "none"
  )
```

```
print(p_neet_box_points)
```




```
ggsave("neet_within_continent_boxplot_2015_2020.png",
       p_neet_box_points, width = 11, height = 6, dpi = 300)
```