Data Science Coursework

2025-12-10

Github

Link to Github: https://github.com/Dipesh0303/IDS_Project/tree/main

Github usernames for each member of group: Dipesh(Dipesh0303), Aik Hsiong(AikHsiong), Mia(miahossen), Mahika(mahikachawda), Yinuo(yinuoli871-dot)

Loading libraries and packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- 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
```

```
library(ggplot2)
library(dplyr)
library(sf)
```

```
## Linking to GEOS 3.13.1, GDAL 3.11.4, PROJ 9.7.0; sf_use_s2() is TRUE
```

```
library(rnaturalearth)
library(rnaturalearthdata)
```

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

```
library(readr)
library(stringr)
library(dslabs)
library(english)
library(tinytex)
```

Code for first target (SDG 8.1)

Loading datasets

```
continents <- read.csv("continents-according-to-our-world-in-data.csv")
GDP_per_capita <- read.csv("gdp-per-capita-worldbank.csv")
world <- ne_countries(scale = "medium", returnclass = "sf")
```

Merging datasets

```
full_data <- inner_join(continents, GDP_per_capita, join_by("Code" == "Code", "Entity" == "Entity"))
```

Cleaning the data and calculating the mean GDP per capita in each country for each year

```
full_data <- full_data %>% select(-Year.x) %>%
  rename("Year" = Year.y) %>%
  rename("GDPperCapita" = GDP.per.capita..PPP..constant.2017.international...) %>%
  group_by(Year, Continent) %>%
  mutate(Mean_GDPperCapita = mean(GDPperCapita, na.rm = TRUE)) %>%
  ungroup()
```

Calculate GDP growth for each country

```
full_data <- full_data %>%
  arrange(Entity, Year) %>%
  group_by(Entity) %>%
  mutate(GDP_growth = (GDPperCapita - lag(GDPperCapita)) / lag(GDPperCapita) * 100) %>%
  ungroup()
```

Calculate mean GDP growth for each continent

```
ContinentalGDPgrowth <- full_data %>%
  arrange(Year) %>%
  group_by(Continent, Year) %>%
  summarise(continentalGDPgrowth = mean(GDP_growth, na.rm = TRUE)) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

Filtering data for each continent

```
North_America <- filter(full_data, Continent == "North America")
Asia <- filter(full_data, Continent == "Asia")
Africa <- filter(full_data, Continent == "Africa")
Oceania <- filter(full_data, Continent == "Oceania")
South_America <- filter(full_data, Continent == "South America")
Europe <- filter(full_data, Continent == "Europe")
```

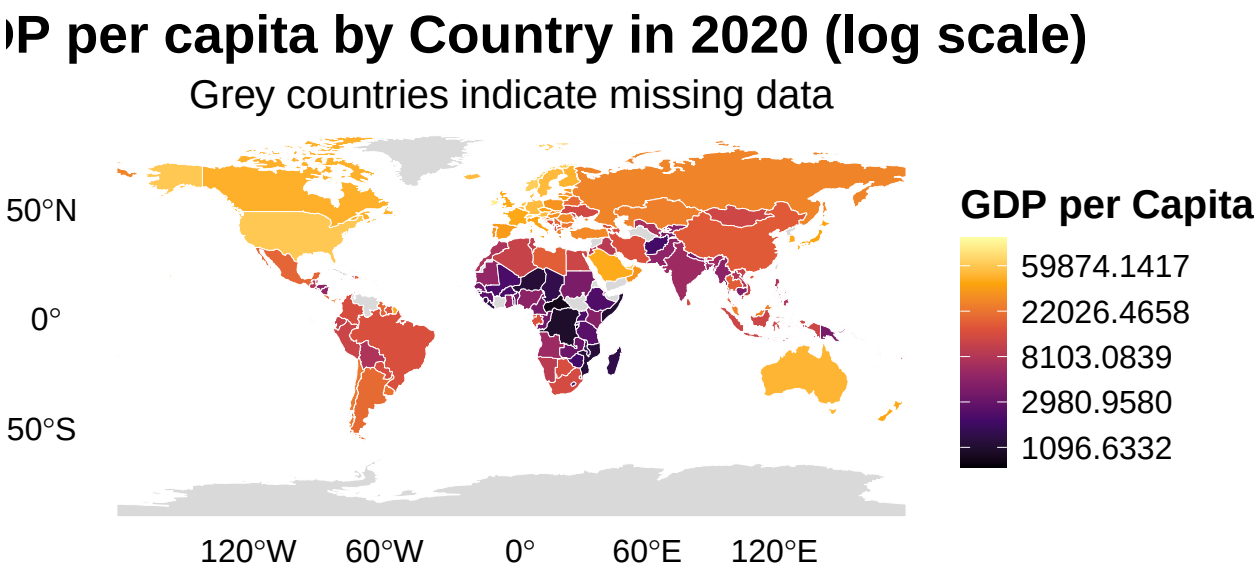
Creating a dataframe to make chloropleth for 2020.

The names are changed for some countries so that when merging the data to create the dataframe these countries are not lost.

```
GDP2020 <- filter(GDP_per_capita, Year == 2020) %>%
  rename("GDPperCapita" = GDP.per.capita..PPP..constant.2017.international...) %>%
  mutate(Entity = recode(Entity,
                         "United States" = "United States of America",
                         "Bosnia and Herzegovina" = "Bosnia and Herz.",
                         "Democratic Republic of Congo" = "Dem. Rep. Congo",
                         "Cape Verde" = "Cabo Verde",
                         "East Timor" = "Timor-Leste",
                         "Micronesia (country)" = "Micronesia",
                         "Eswatini" = "eSwatini",
                         "Antigua and Barbuda" = "Antigua and Barb.",
                         "Central African Republic" = "Central African Rep.",
                         "Congo, Democratic Republic of the" = "Dem. Rep. Congo",
                         "Dominican Republic" = "Dominican Rep.",
                         "Equatorial Guinea" = "Eq. Guinea",
                         "Marshall Islands" = "Marshall Is.",
                         "Solomon Islands" = "Solomon Is.",
                         "Saint Kitts and Nevis" = "St. Kitts and Nevis",
                         "Turks and Caicos Islands" = "Turks and Caicos Is.",
                         "Côte d'Ivoire" = "Ivory Coast",
                         "Curacao" = "Curaçao",
                         "St. Vincent and the Grenadines" = "St. Vincent and the Grenadines",
                         "Sao Tome and Principe" = "São Tomé and Principe",
                         "Cayman Islands" = "Cayman Is."
  ))

map_data2020 <- world %>%
  left_join(GDP2020, by = c("name" = "Entity"))
```

Chloropleth for GDP per capita by Country (2020) using a log scale

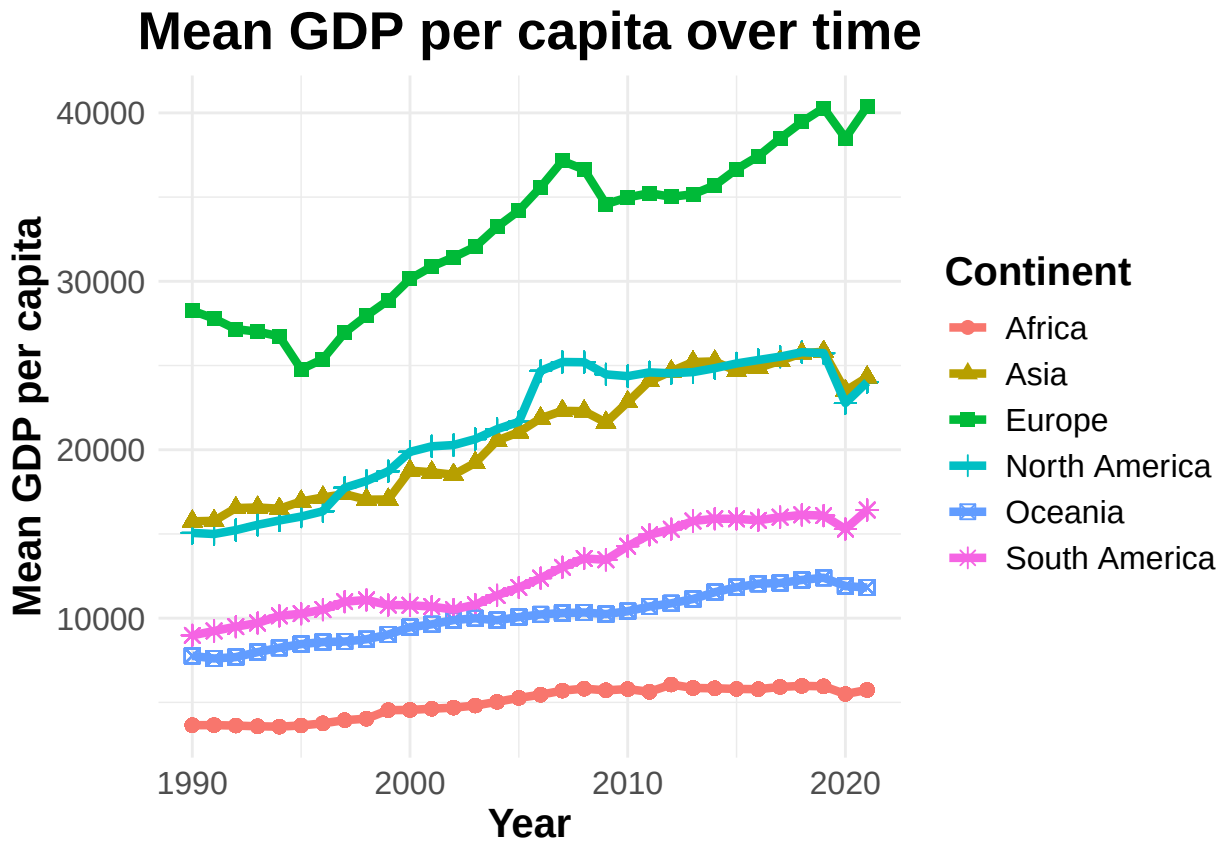
```r
ggplot(map_data2020) +
  geom_sf(aes(fill = GDPperCapita), color = "white", size = 0.15) +
  scale_fill_viridis_c(option = "B",trans = "log", na.value = "grey85") +
  theme_void() +
  labs(
    title = "GDP per capita by Country in 2020 (log scale)",
    subtitle = "Grey countries indicate missing data",
    fill = "GDP per Capita"
  ) +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12))
```



Line chart showing the mean GDP per capita over time for each continent

```r
full_data %>%
  ggplot(aes(x = Year, y = Mean_GDPperCapita, colour = Continent, shape = Continent)) +
  geom_point(size = 2.3) +
  geom_line(size = 1.5) +
```

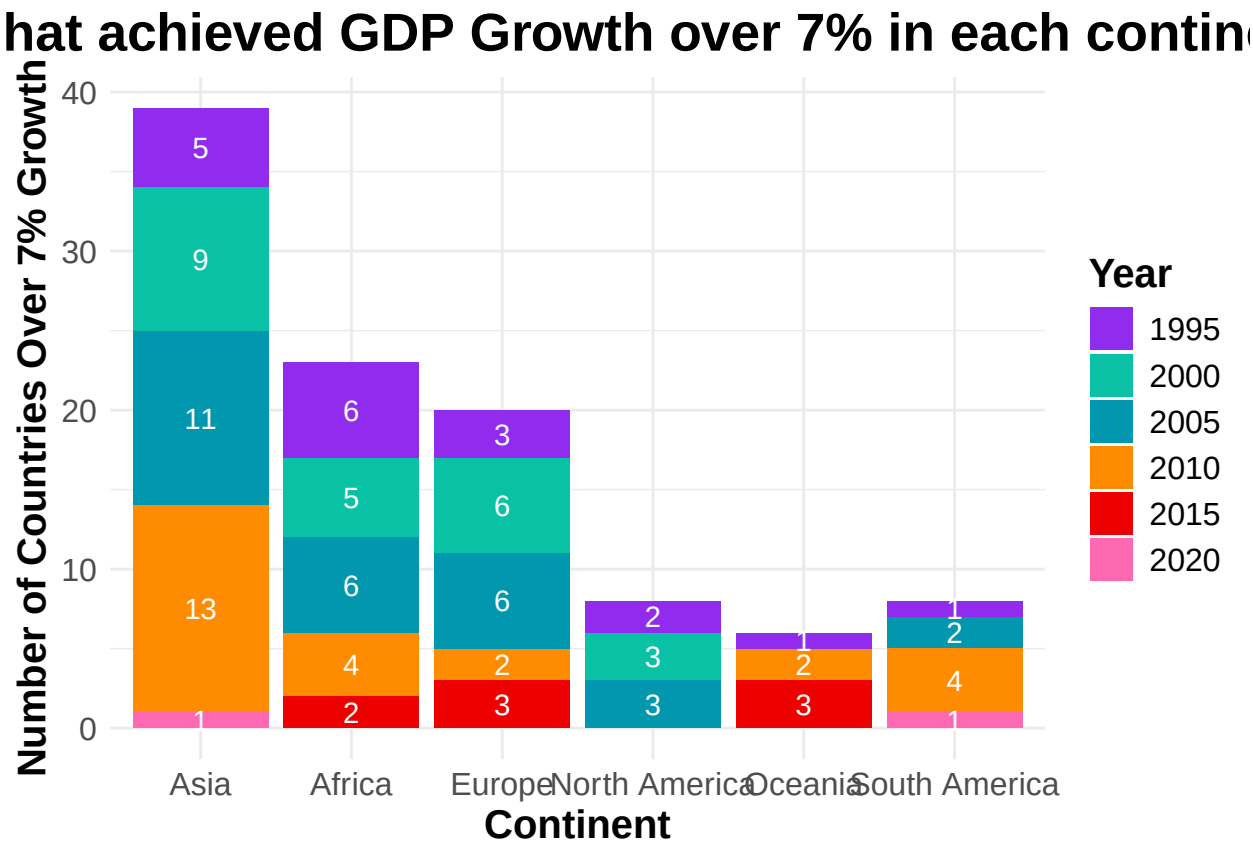
```
  labs(title = "Mean GDP per capita over time",
       x = "Year",
       y = "Mean GDP per capita") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12))
```



Stacked bar chart showing the number of countries where GDP growth is over 7% over time in each continent

```
full_data %>% filter(GDP_growth > 7, Year %in% c(1990, 1995, 2000,2005,2010,2015,2020)) %>%
  group_by(Continent, Year) %>%
  summarise(count = n(), .groups = "drop") %>%
  ggplot(aes(x = reorder(Continent, -count), y = count, fill = factor(Year))) +
  geom_bar(stat = "identity", position = "stack") +
  geom_text(aes(label = count),
            position = position_stack(vjust = 0.5),
            colour = "white",
            size = 4) +
```

```
  labs(
    title = "Number of countries that achieved GDP Growth over 7% in each continent in select years",
    x = "Continent",
    y = "Number of Countries Over 7% Growth",
    fill = "Year") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12)) +
  scale_fill_manual(values = c("1990" = "yellow", "1995" = "purple2", "2000" = "#09c2a5", "2005" = "#009
```



Heatmap globally showing mean GDP growth per continent

```
full_data %>%
  group_by(Continent, Year) %>%
  summarise(Mean_Growth = mean(GDP_growth, na.rm = TRUE)) %>%
  ggplot(aes(x = Year, y = Continent, fill = Mean_Growth)) +
  geom_tile() +
  scale_fill_viridis_c(option = "inferno") +
  labs(title = "Mean GDP Growth by Continent and Year",
```

```
      fill = "Mean GDP growth") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12))
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

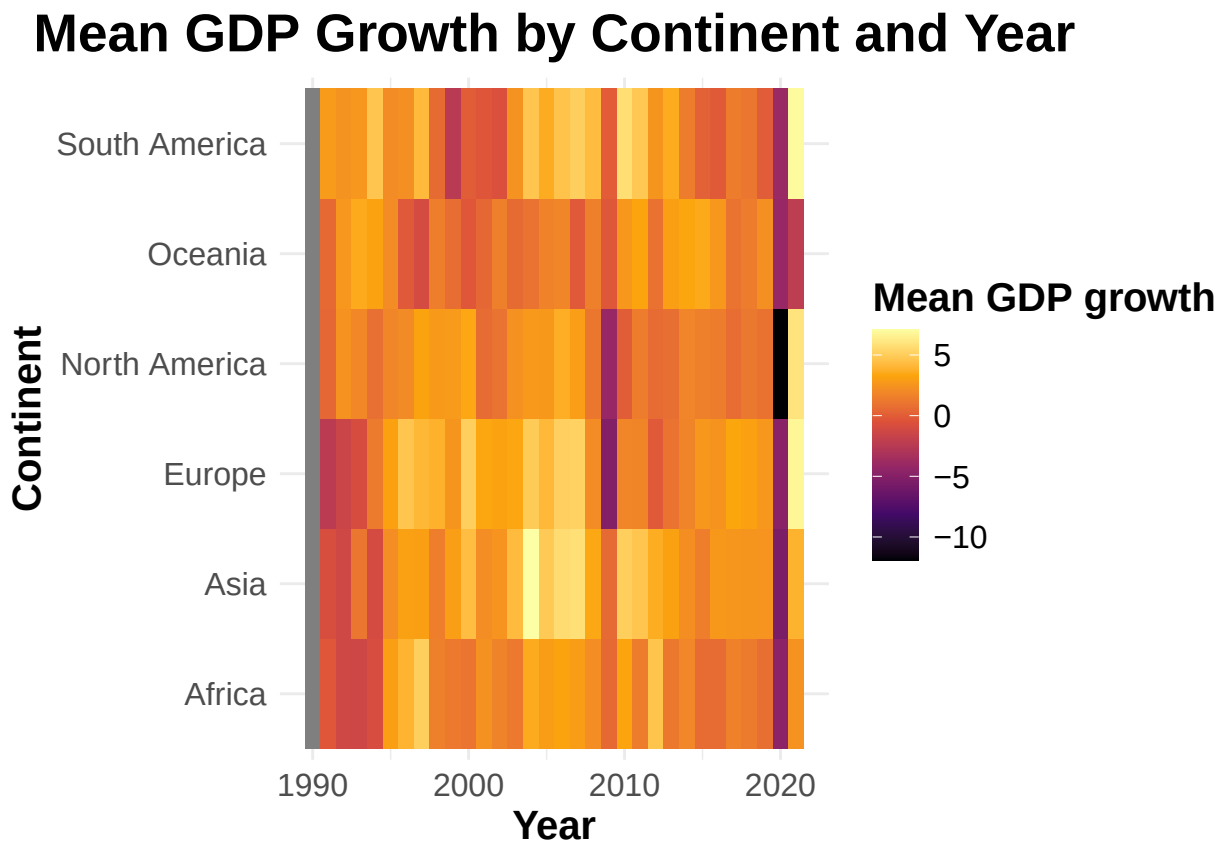


Table counting number of countries in each continent with GDP per capita data

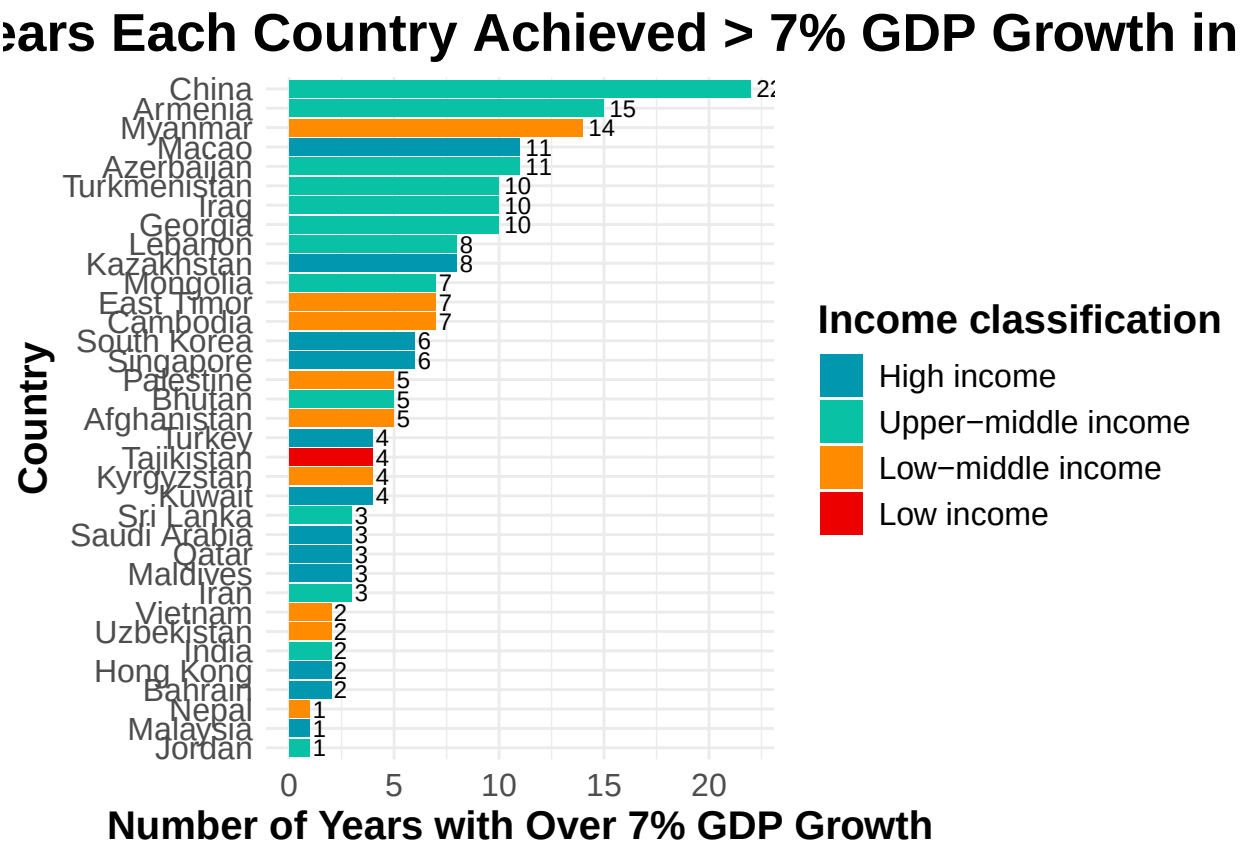
```
full_data %>% filter(Year == 2020) %>%
  group_by(Continent) %>%
  summarise("Number of countries" = n())
```

```
## # A tibble: 6 x 2
##   Continent     `Number of countries`
##   <chr>                         <int>
## 1 Africa                           52
```

```
## 2 Asia                          45
## 3 Europe                        42
## 4 North America                 28
## 5 Oceania                       14
## 6 South America                 11
```

Bar graph showing the number of times GDP growth is over 7% in countries in Asia.

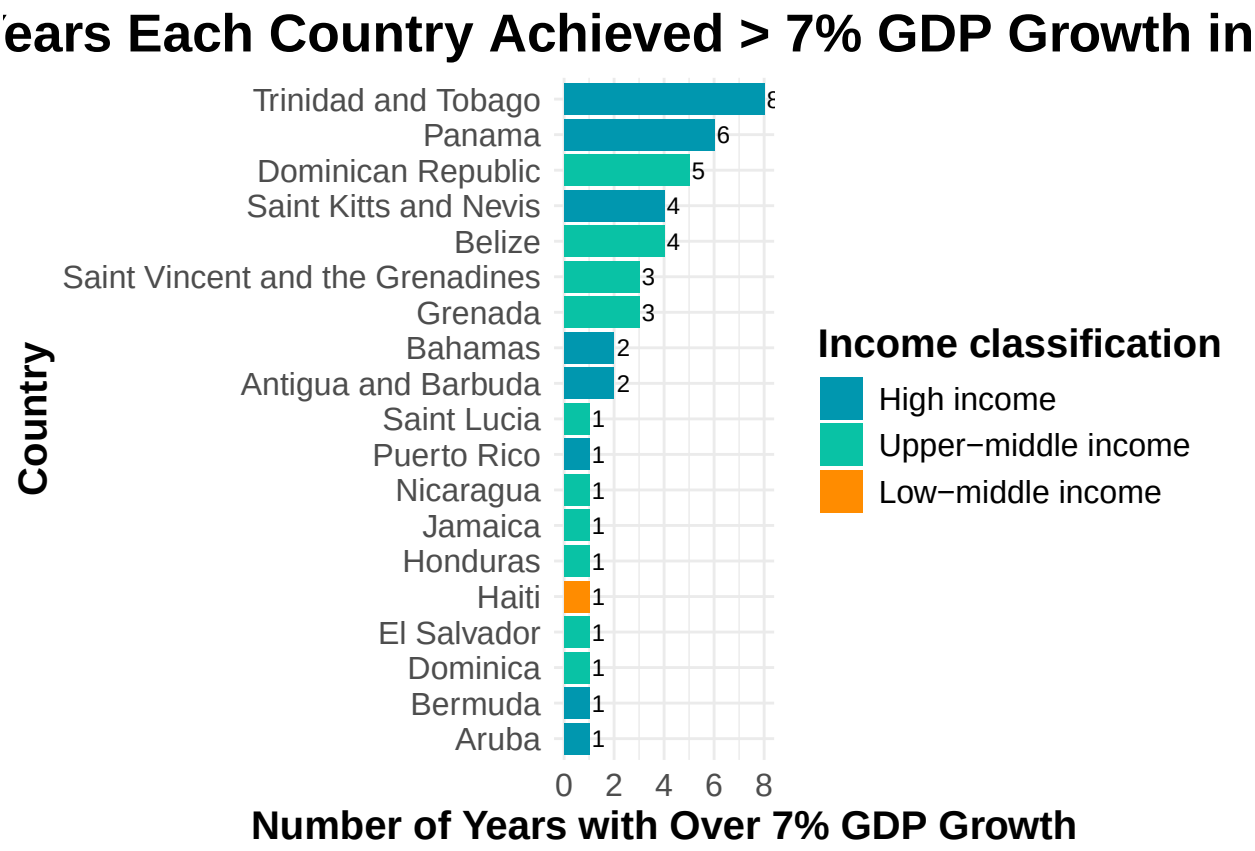
```
Asia %>% filter(GDP_growth > 7) %>%
  group_by(Entity) %>%
  summarise(count = n(),
            avg_gdppc = mean(GDPperCapita, na.rm = TRUE)) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = count, y = reorder(Entity, count), fill = case_when(
    avg_gdppc < 1600 ~ "Low income",
    between(avg_gdppc, 1600, 5000) ~ "Low-middle income",
    between(avg_gdppc, 5000, 14000) ~ "Upper-middle income",
    avg_gdppc >14000 ~ "High income"))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = count), hjust = -0.2, size = 3) +
  labs(
    title = "Number of Years Each Country Achieved > 7% GDP Growth in Asia",
    x = "Number of Years with Over 7% GDP Growth",
    y = "Country",
    fill = "Income classification") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12)) +
  scale_fill_manual(
    values = c("Low income" = "#ed0000",
               "Low-middle income" = "darkorange",
               "Upper-middle income" = "#09c2a5",
               "High income" = "#0097af"),
    breaks = c("High income", "Upper-middle income", "Low-middle income", "Low income")
  )
```



Bar graph showing the number of times GDP growth is over 7% in countries in North America.

```
North_America %>% filter(GDP_growth > 7) %>%
  group_by(Entity) %>%
  summarise(count = n(),
            avg_gdppc = mean(GDPperCapita, na.rm = TRUE)) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = count, y = reorder(Entity, count), fill = case_when(
    avg_gdppc < 1600 ~ "Low income",
    between(avg_gdppc, 1600, 5000) ~ "Low-middle income",
    between(avg_gdppc, 5000, 14000) ~ "Upper-middle income",
    avg_gdppc >14000 ~ "High income"))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = count), hjust = -0.2, size = 3) +
  labs(
    title = "Number of Years Each Country Achieved > 7% GDP Growth in North America",
    x = "Number of Years with Over 7% GDP Growth",
    y = "Country",
    fill = "Income classification") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
```

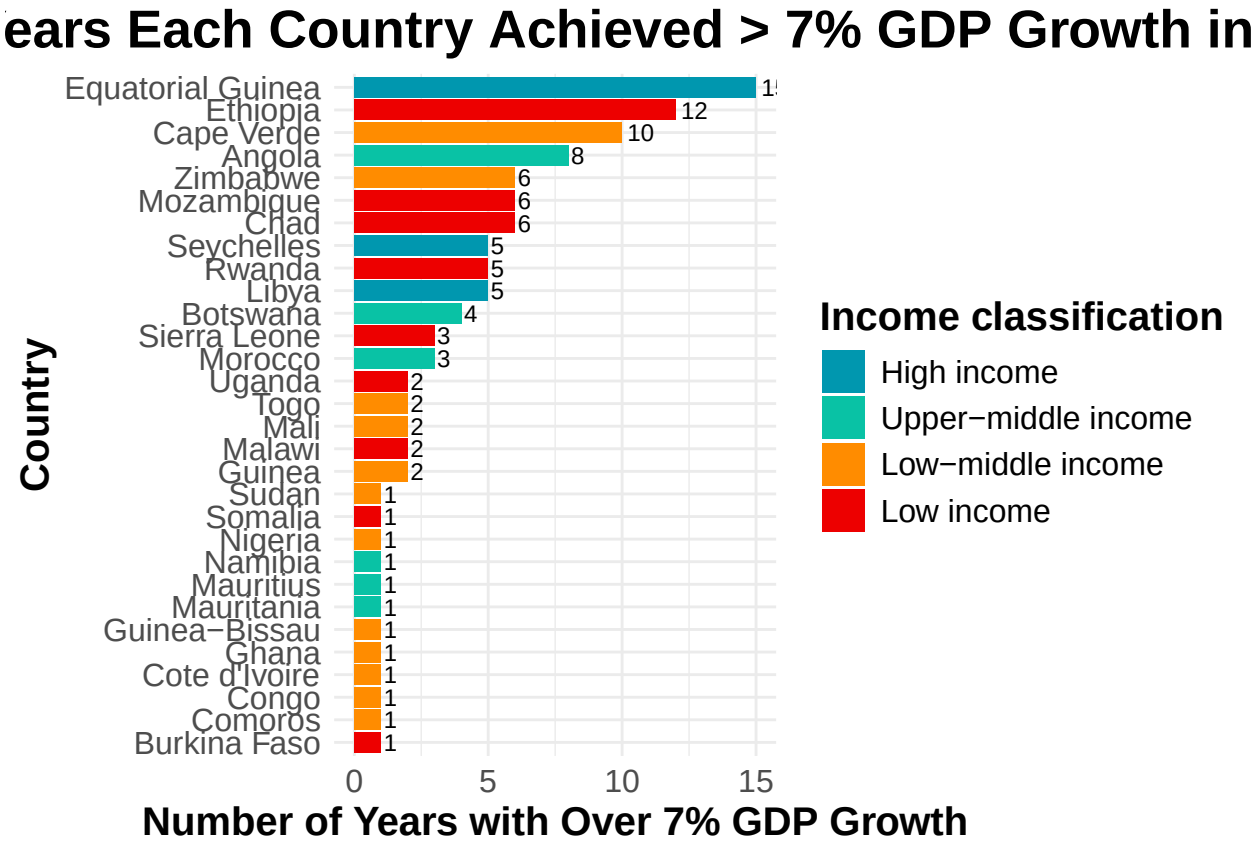
```
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12)) +
  scale_fill_manual(
    values = c("Low income" = "#ed0000",
               "Low-middle income" = "darkorange",
               "Upper-middle income" = "#09c2a5",
               "High income" = "#0097af"),
    breaks = c("High income", "Upper-middle income", "Low-middle income", "Low income")
  )
```



Bar graph showing the number of times GDP growth is over 7% in countries in Africa.

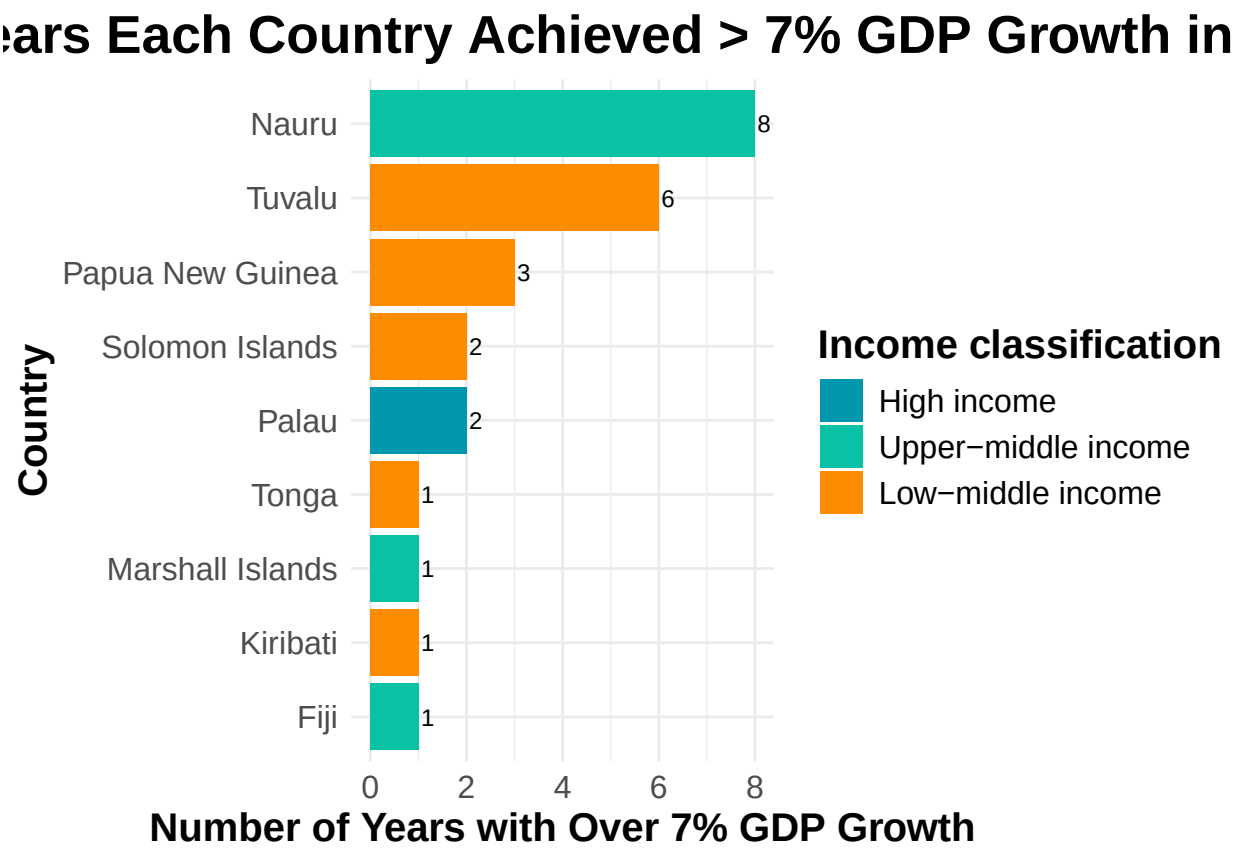
```
Africa %>% filter(GDP_growth > 7) %>%
  group_by(Entity) %>%
  summarise(count = n(),
            avg_gdppc = mean(GDPperCapita, na.rm = TRUE)) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = count, y = reorder(Entity, count), fill = case_when(
    avg_gdppc < 1600 ~ "Low income",
    between(avg_gdppc, 1600, 5000) ~ "Low-middle income",
    between(avg_gdppc, 5000, 14000) ~ "Upper-middle income",
```

```
  avg_gdppc >14000 ~ "High income"))) +
geom_bar(stat = "identity") +
geom_text(aes(label = count), hjust = -0.2, size = 3) +
labs(
  title = "Number of Years Each Country Achieved > 7% GDP Growth in Africa",
  x = "Number of Years with Over 7% GDP Growth",
  y = "Country",
  fill = "Income classification") +
theme_minimal() +
theme(
  axis.title = element_text(size = 15, face = "bold"),
  axis.text = element_text(size = 12),
  plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
  legend.title = element_text(size = 15, face = "bold"),
  legend.text = element_text(size = 12)) +
scale_fill_manual(
  values = c("Low income" = "#ed0000",
             "Low-middle income" = "darkorange",
             "Upper-middle income" = "#09c2a5",
             "High income" = "#0097af"),
  breaks = c("High income", "Upper-middle income", "Low-middle income", "Low income")
)
```



Bar graph showing the number of times GDP growth is over 7% in countries in Oceania.

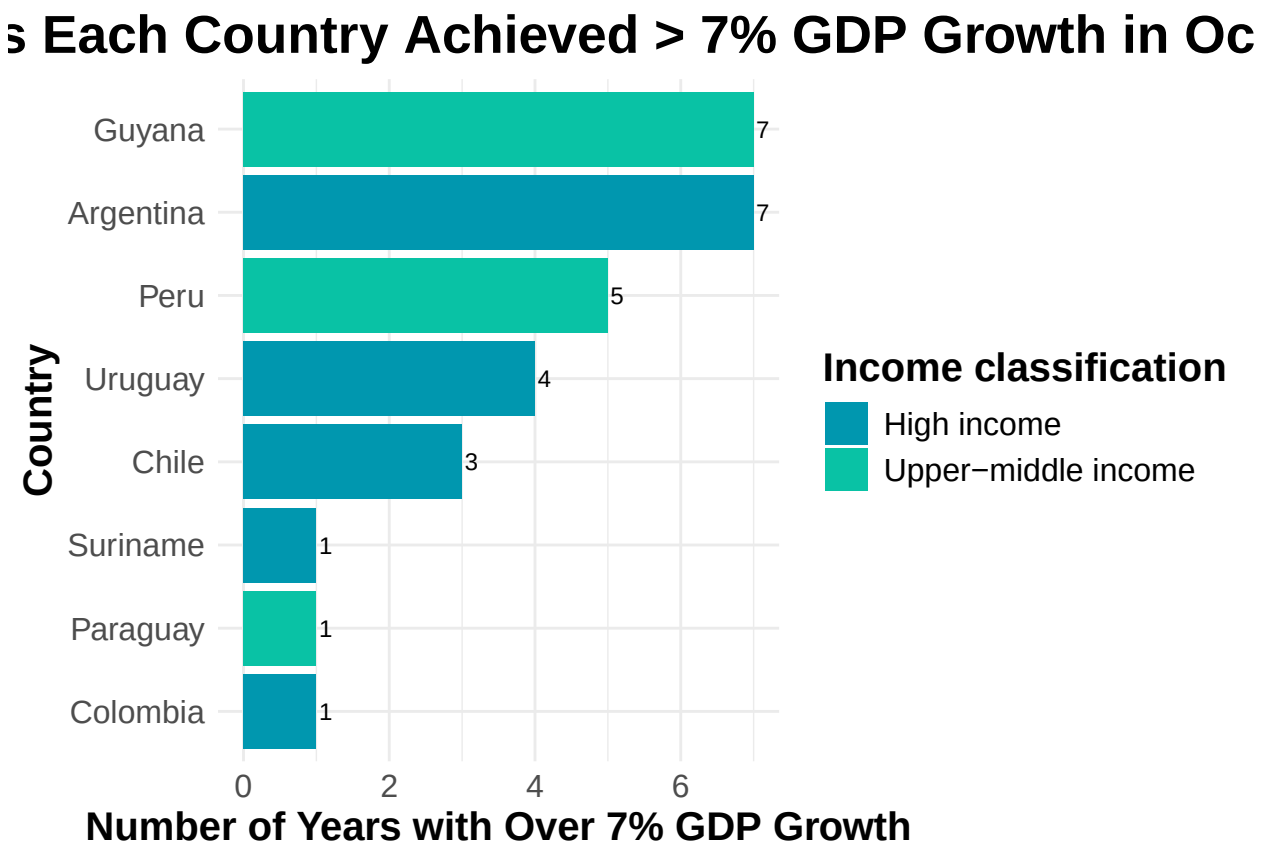
```
Oceania %>% filter(GDP_growth > 7) %>%
  group_by(Entity) %>%
  summarise(count = n(),
            avg_gdppc = mean(GDPperCapita, na.rm = TRUE)) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = count, y = reorder(Entity, count), fill = case_when(
    avg_gdppc < 1600 ~ "Low income",
    between(avg_gdppc, 1600, 5000) ~ "Low-middle income",
    between(avg_gdppc, 5000, 14000) ~ "Upper-middle income",
    avg_gdppc >14000 ~ "High income"))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = count), hjust = -0.2, size = 3) +
  labs(
    title = "Number of Years Each Country Achieved > 7% GDP Growth in Oceania",
    x = "Number of Years with Over 7% GDP Growth",
    y = "Country",
    fill = "Income classification") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12)) +
  scale_fill_manual(
    values = c("Low income" = "#ed0000",
               "Low-middle income" = "darkorange",
               "Upper-middle income" = "#09c2a5",
               "High income" = "#0097af"),
    breaks = c("High income", "Upper-middle income", "Low-middle income", "Low income")
  )
```



Bar graph showing the number of times GDP growth is over 7% in countries in South America

```
South_America %>% filter(GDP_growth > 7) %>%
  group_by(Entity) %>%
  summarise(count = n(),
            avg_gdppc = mean(GDPperCapita, na.rm = TRUE)) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = count, y = reorder(Entity, count), fill = case_when(
    avg_gdppc < 1600 ~ "Low income",
    between(avg_gdppc, 1600, 5000) ~ "Low-middle income",
    between(avg_gdppc, 5000, 14000) ~ "Upper-middle income",
    avg_gdppc >14000 ~ "High income"))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = count), hjust = -0.2, size = 3) +
  labs(
    title = "Number of Years Each Country Achieved > 7% GDP Growth in Oceania",
    x = "Number of Years with Over 7% GDP Growth",
    y = "Country",
    fill = "Income classification") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
```

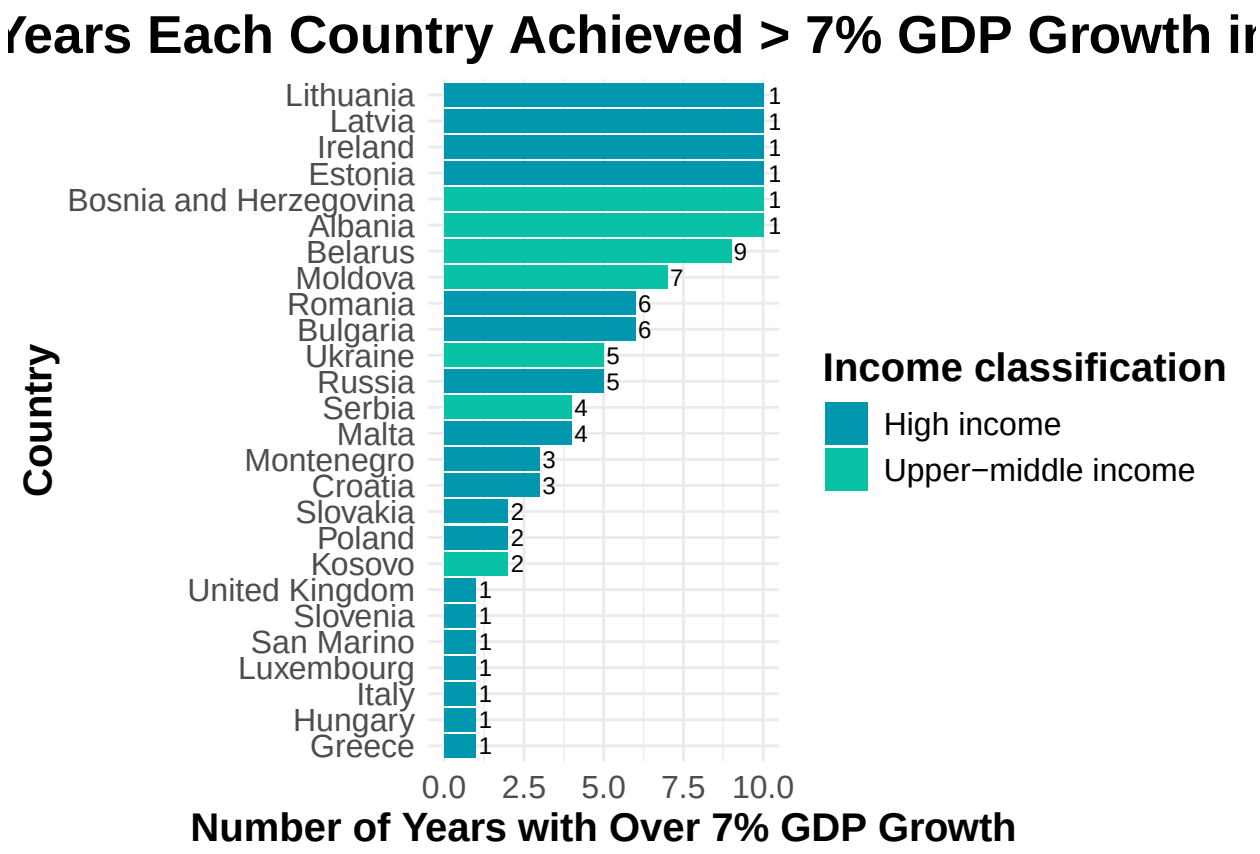
```
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12)) +
  scale_fill_manual(
    values = c("Low income" = "#ed0000",
               "Low-middle income" = "darkorange",
               "Upper-middle income" = "#09c2a5",
               "High income" = "#0097af"),
    breaks = c("High income", "Upper-middle income", "Low-middle income", "Low income")
  )
```



##Bar graph showing the number of times GDP growth is over 7% in countries in Europe

```
Europe %>% filter(GDP_growth > 7) %>%
  group_by(Entity) %>%
  summarise(count = n(),
            avg_gdppc = mean(GDPperCapita, na.rm = TRUE)) %>%
  arrange(desc(count)) %>%
  ggplot(aes(x = count, y = reorder(Entity, count), fill = case_when(
    avg_gdppc < 1600 ~ "Low income",
    between(avg_gdppc, 1600, 5000) ~ "Low-middle income",
    between(avg_gdppc, 5000, 14000) ~ "Upper-middle income",
    avg_gdppc >14000 ~ "High income"))) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = count), hjust = -0.2, size = 3) +
  labs(
```

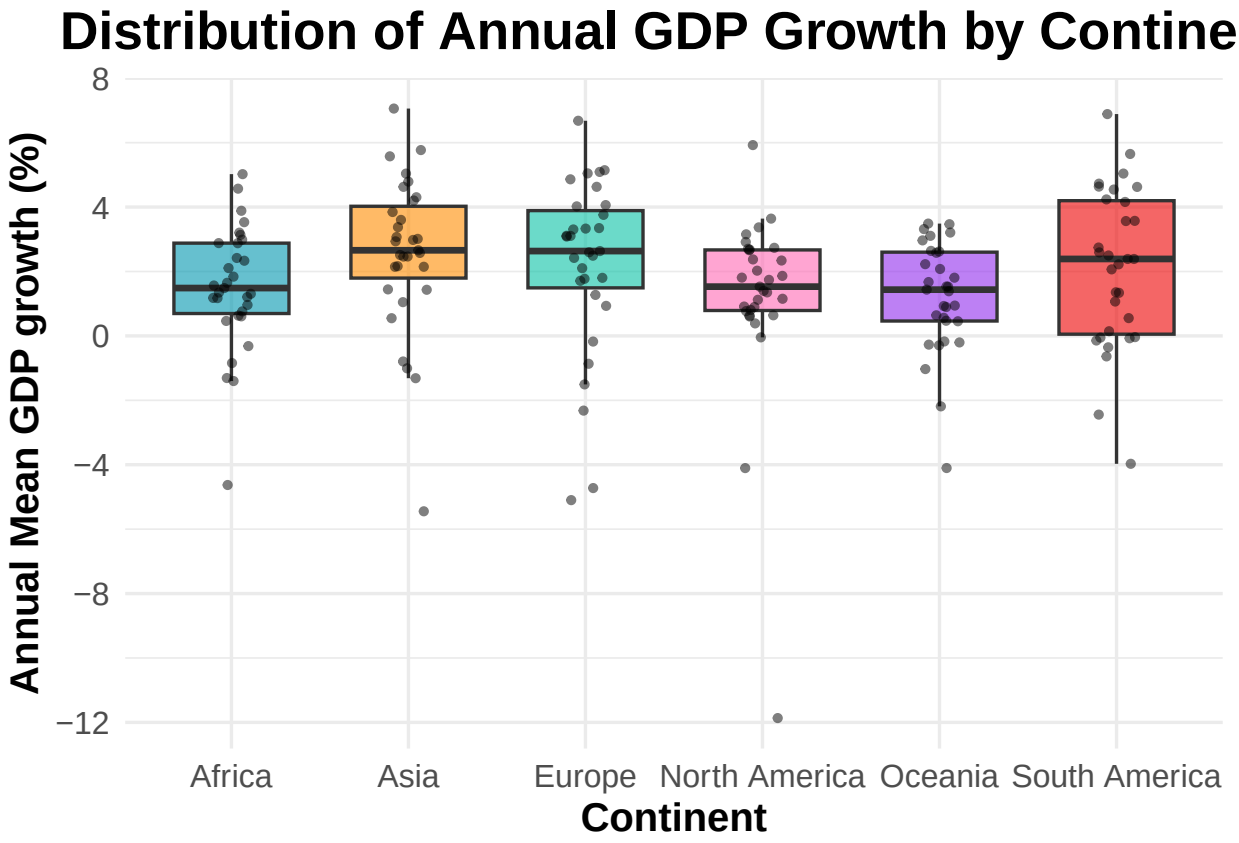
```
    title = "Number of Years Each Country Achieved > 7% GDP Growth in Oceania",
    x = "Number of Years with Over 7% GDP Growth",
    y = "Country",
    fill = "Income classification") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12)) +
  scale_fill_manual(
    values = c("Low income" = "#ed0000",
               "Low-middle income" = "darkorange",
               "Upper-middle income" = "#09c2a5",
               "High income" = "#0097af"),
    breaks = c("High income", "Upper-middle income", "Low-middle income", "Low income")
  )
```



Boxplot showing the distribution of annual GDP growth rates by continent

```
ContinentalGDPgrowth %>%
  ggplot(aes(x = Continent, y = continentalGDPgrowth, fill = Continent)) +
```

```
  geom_boxplot(outlier.shape = NA, alpha = 0.6, width = 0.65) +
  geom_jitter(width = 0.12, alpha = 0.5, size = 1) +
  theme_minimal(base_size = 13) +
  labs(
    title = "Distribution of Annual GDP Growth by Continent",
    x = "Continent",
    y = "Annual Mean GDP growth (%)"
  ) +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    legend.position = "none") +
  scale_fill_manual(values = c("Oceania" = "purple2", "Europe" = "#09c2a5", "Africa" = "#0097af", "Asia
```



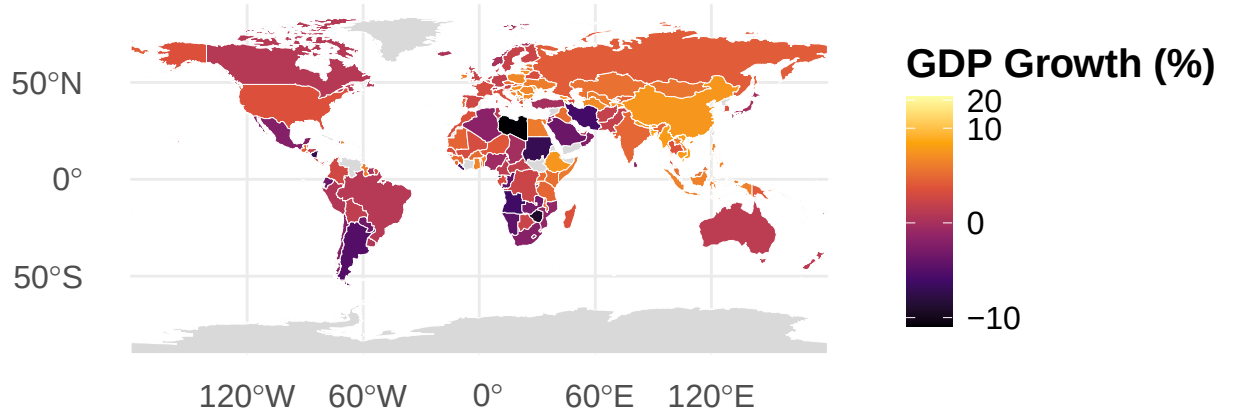
Creating the dataframe for a chloropleth for 2019

```
GDP2019 <- filter(full_data, Year == 2019) %>%
  mutate(Entity = recode(Entity,
                         "United States" = "United States of America",
                         "Bosnia and Herzegovina" = "Bosnia and Herz.",
                         "Democratic Republic of Congo" = "Dem. Rep. Congo",
                         "Cape Verde" = "Cabo Verde",
```

```
                          "East Timor" = "Timor-Leste",
                          "Micronesia (country)" = "Micronesia",
                          "Eswatini" = "eSwatini",
                          "Antigua and Barbuda" = "Antigua and Barb.",
                          "Central African Republic" = "Central African Rep.",
                          "Congo, Democratic Republic of the" = "Dem. Rep. Congo",
                          "Dominican Republic" = "Dominican Rep.",
                          "Equatorial Guinea" = "Eq. Guinea",
                          "Marshall Islands" = "Marshall Is.",
                          "Solomon Islands" = "Solomon Is.",
                          "Saint Kitts and Nevis" = "St. Kitts and Nevis",
                          "Turks and Caicos Islands" = "Turks and Caicos Is.",
                          "Côte d'Ivoire" = "Ivory Coast",
                          "Curacao" = "Curaçao",
                          "St. Vincent and the Grenadines" = "St. Vincent and the Grenadines",
                          "Sao Tome and Principe" = "São Tomé and Principe",
                          "Cayman Islands" = "Cayman Is."
  ))

map_data2019 <- world %>%
  left_join(GDP2019, by = c("name" = "Entity"))
```

Chloropleth for GDP growth by Country (2019)

```
ggplot(map_data2019) +
  geom_sf(aes(fill = GDP_growth), color = "white", size = 0.15) +
  scale_fill_viridis_c(option = "B", trans = scales::pseudo_log_trans(base = 10), na.value = "grey85") +
  theme_minimal() +
  labs(
    title = "GDP Growth by Country (2019, log scaled)",
    subtitle = "Uses pseudo-logarithmic colour scaling to handle extreme values. Grey countries indicat
    fill = "GDP Growth (%)"
  ) +
  theme(
    axis.title = element_text(size = 15, face = "bold"),
    axis.text = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 20, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5, size = 15),
    legend.title = element_text(size = 15, face = "bold"),
    legend.text = element_text(size = 12))
```



Code for second target (SDG 8.6)

Loading of datasets

```
continents <- read.csv("continents-according-to-our-world-in-data.csv")
gdp_per_capita <- read.csv("gdp-per-capita-worldbank.csv")
youth <- read.csv("youth-not-in-education-employment-training.csv")
population <- read.csv("World Population.csv")
```

Cleaning datasets

```
continents <- continents %>%
  select(c(-Year, -Code))

population <- population %>%
  rename(Country = Region..subregion..country.or.area..) %>%
  select(Country, Year, X0.14, X0.24) %>%
  mutate(
    X0.14 = as.numeric(gsub("[^0-9.]", "", X0.14)),
    X0.24 = as.numeric(gsub("[^0-9.]", "", X0.24))
  ) %>%
```

```
  mutate(X0.14 = X0.14*1000, X0.24 = X0.24*1000) %>%
  mutate(Population = X0.24 - X0.14)

# Changing names which are not the same
country_map <- c(
  # population name                          = name in Data set 1
  "Democratic Republic of the Congo" = "Democratic Republic of Congo",
  "Côte d'Ivoire"                    = "Cote d'Ivoire",
  "Lao People's Democratic Republic" = "Laos",
  "Timor-Leste"                      = "East Timor",
  "China, Hong Kong SAR"             = "Hong Kong",
  "China, Macao SAR"                 = "Macao",
  "Russian Federation"               = "Russia",
  "Viet Nam"                         = "Vietnam",
  "Brunei Darussalam"                = "Brunei",
  "State of Palestine"               = "Palestine",
  "United Republic of Tanzania"      = "Tanzania",
  "Bolivia (Plurinational State of)" = "Bolivia",
  "Venezuela (Bolivarian Republic of)" = "Venezuela",
  "Micronesia (Fed. States of)"      = "Micronesia (country)",
  "Türkiye"                          = "Turkey"
)

population <- population %>%
  mutate(
    Country = recode(Country, !!!country_map)
  )
```

Merging datasets

```
df <- youth %>%
  left_join(continents, by ="Entity") %>%
  rename(Country = Entity,
         NEET_percentage = Share.of.youth.not.in.education..employment.or.training..total....of.youth.p
  arrange(Continent, Country, Year) %>%
  drop_na(Continent)

df <- df %>%
  left_join(population, by = c('Country', 'Year'))
# Checking that there are only 6 continents present no NA
unique(df$Continent)
```

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

```
df <- df %>%
  group_by(Continent, Year) %>%
  mutate(continent_population = sum(as.numeric(Population), na.rm = TRUE)) %>%
  mutate(weight = (Population/continent_population)) %>%
  mutate(weighted_NEET_percentage = (weight * NEET_percentage))
```

Graphs colours

```
continent_colors <- c(
  "Asia" = "#1B9E77",
  "Europe" = "#D95F02",
  "Africa" = "#7570B3",
  "North America" = "#E7298A",
  "South America" = "#66A61E",
  "Oceania" = "#E6AB02"
)
```

Bar Graphs

Creating bar dataframe

```
bar <- df %>%
  filter(Year == 2015 | Year == 2020) %>%
  group_by(Continent, Year) %>%
  summarise(continent_avg_NEET = mean(NEET_percentage, na.rm = TRUE))
```

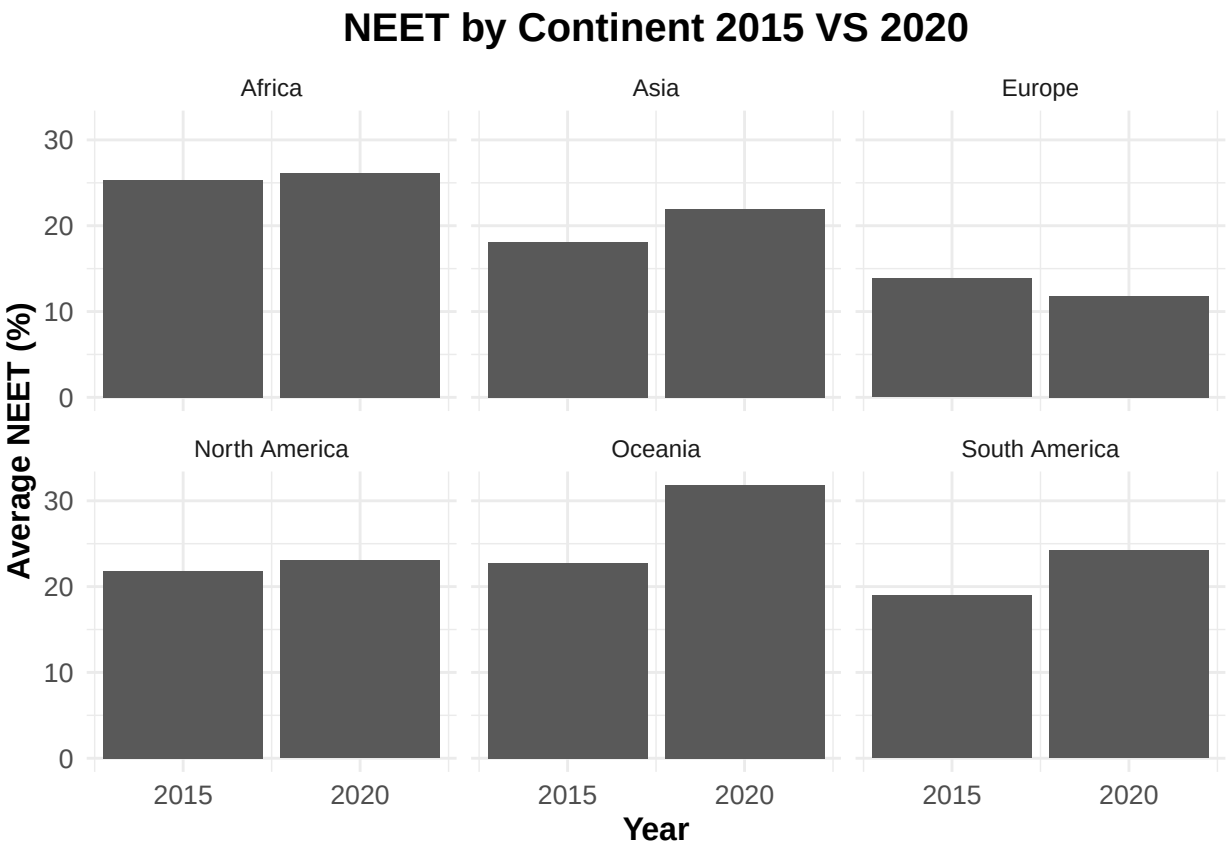
```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

```
weighted_bar <- df %>%
  filter(Year == 2015 | Year == 2020) %>%
  group_by(Continent, Year) %>%
  summarise(continent_weighted_avg_NEET = sum(weighted_NEET_percentage, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

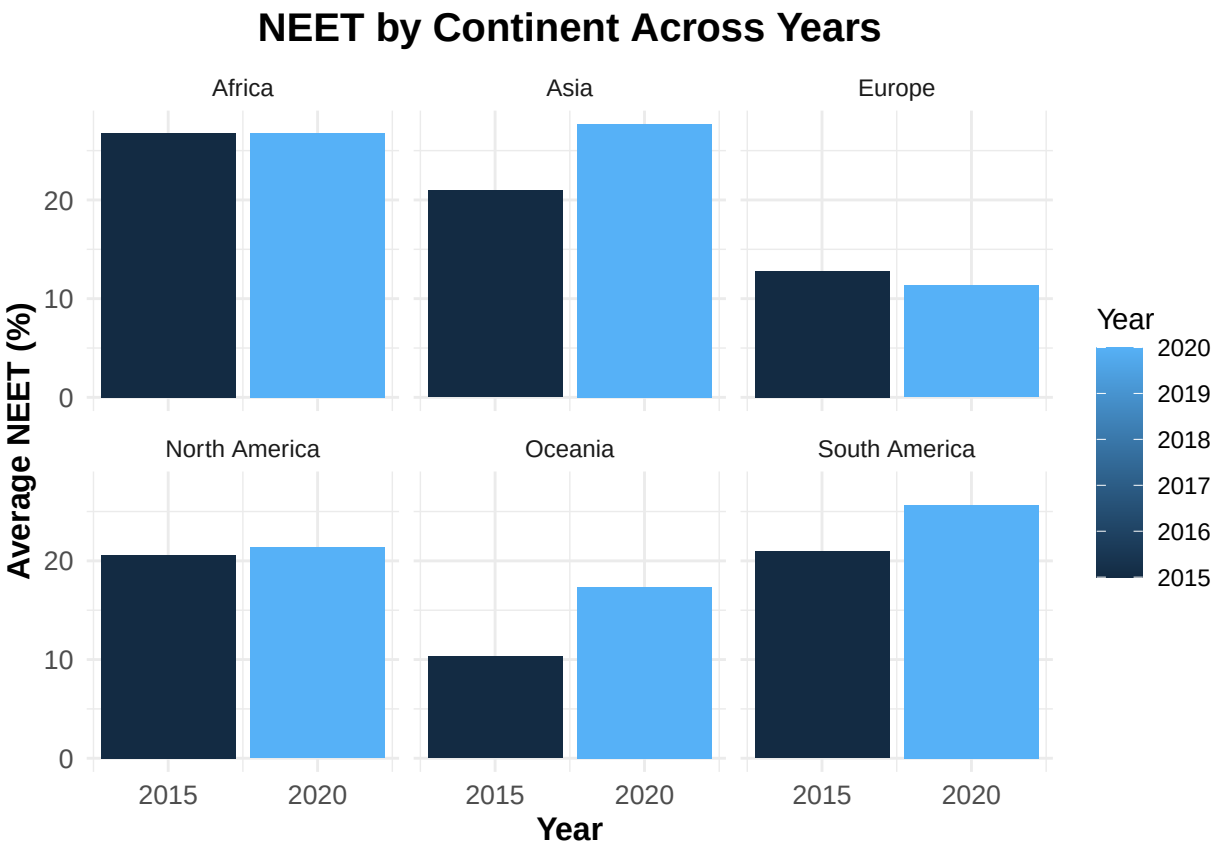
All Continents (2015 vs 2020) – unweighted

```
bar %>%
  ggplot(aes(x = Year, y = continent_avg_NEET)) +
  geom_col() +
  facet_wrap(~ Continent) +
  scale_x_continuous(breaks = c(2015, 2020)) +
  labs(x = "Year",
       y = "Average NEET (%)",
       title = "NEET by Continent 2015 VS 2020") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  )
```



All Continents (2015 vs 2020) – weighted

```
weighted_bar %>%
  ggplot(aes(x = Year, y = continent_weighted_avg_NEET, fill = Year)) +
  geom_col() +
  facet_wrap(~ Continent) +
  scale_x_continuous(breaks = c(2015, 2020)) +
  labs(x = "Year",
       y = "Average NEET (%)",
       title = "NEET by Continent Across Years") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),

  )
```



Line graphs

Creating line dataframe

```
line <- df %>%
  filter(Year >= 2005 & Year<=2020) %>%
  group_by(Continent, Year) %>%
  summarise(continent_avg_NEET = mean(NEET_percentage, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

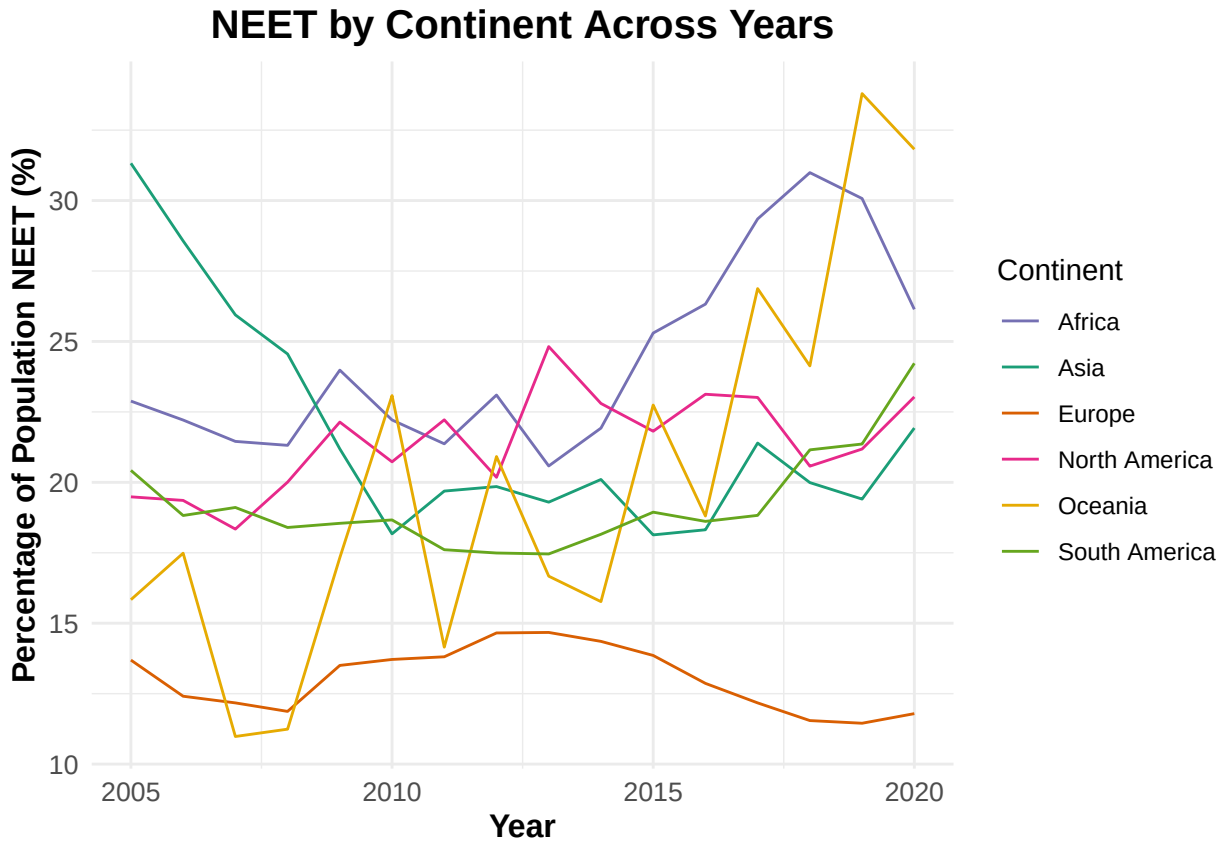
```
weighted_line <- df %>%
  filter(Year >= 2005 & Year<=2020) %>%
  group_by(Continent, Year) %>%
  summarise(continent_weighted_avg_NEET = mean(weighted_NEET_percentage, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

```
trial <- df %>%
  filter(Continent == "Asia")
```

All Continents (2005-2020) – unweigted

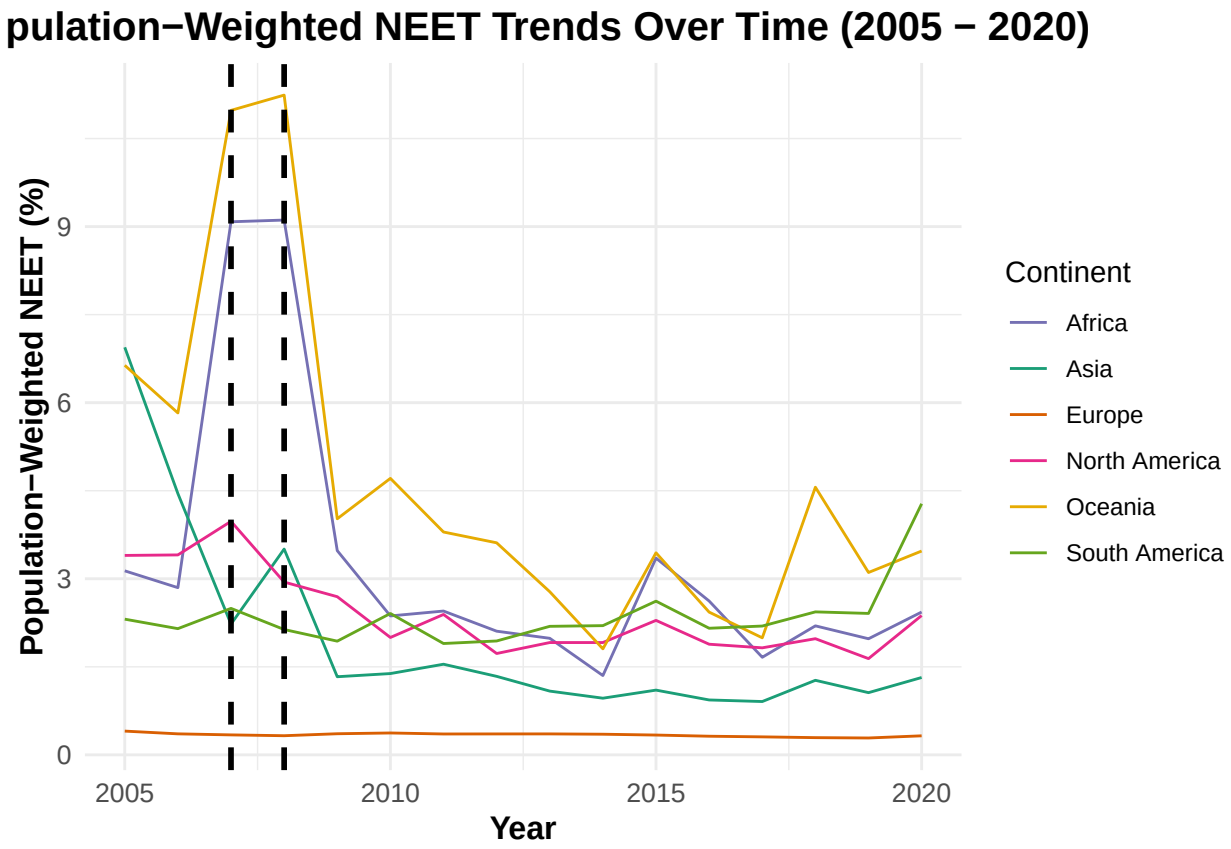
```
line %>%
  ggplot(aes(x = Year, y = continent_avg_NEET)) +
  geom_line(aes(colour = Continent)) +
  labs(x = "Year",
       y = "Percentage of Population NEET (%)",
       title = "NEET by Continent Across Years") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  ) +
  scale_color_manual(values = continent_colors)
```



All Continents (2005 - 2020) – weighted

```
weighted_line %>%
  ggplot(aes(x = Year, y = continent_weighted_avg_NEET)) +
  geom_line(aes(colour = Continent)) +
  geom_vline(xintercept = 2007, linetype = "dashed", colour = "black", linewidth = 1) +
  geom_vline(xintercept = 2008, linetype = "dashed", colour = "black", linewidth = 1) +
```

```
  labs(x = "Year",
       y = "Population-Weighted NEET (%)",
       title = "Population-Weighted NEET Trends Over Time (2005 - 2020)") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  ) +
  scale_color_manual(values = continent_colors)
```



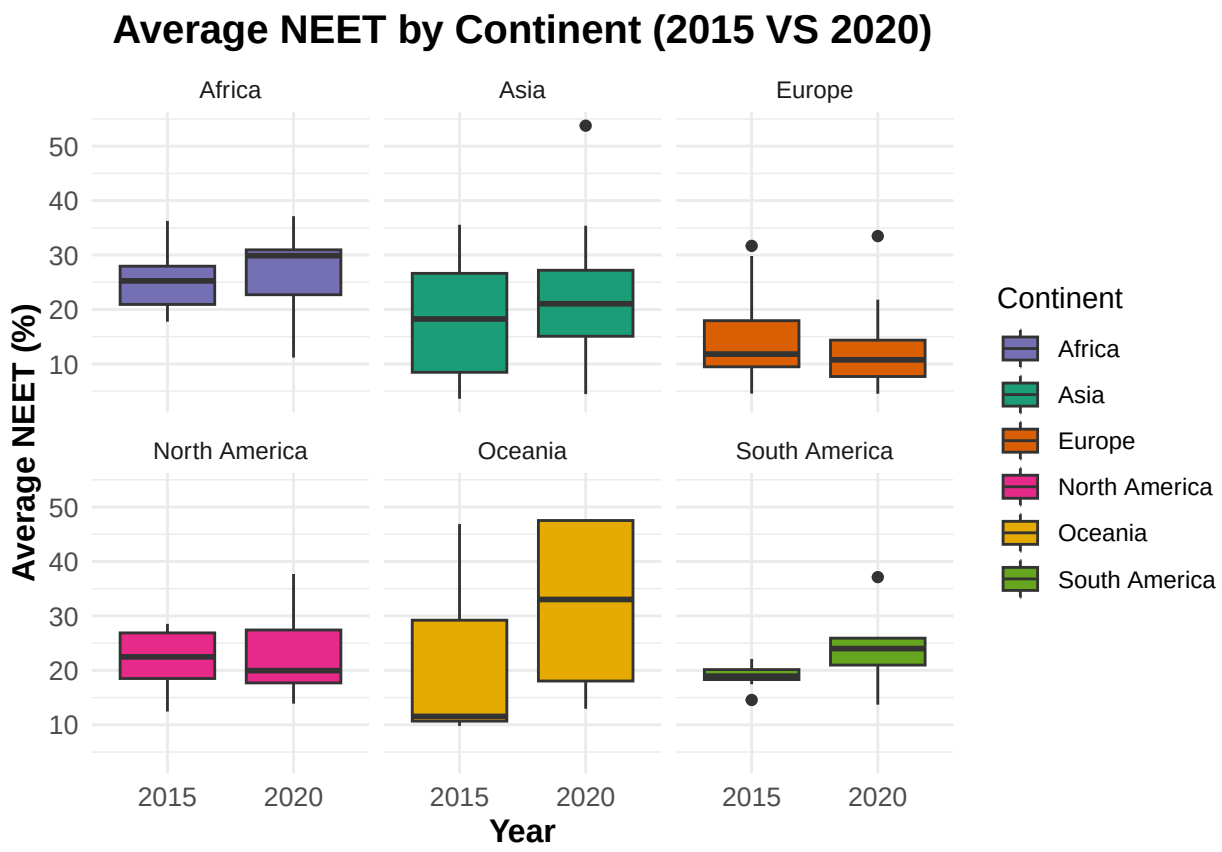
Box Plots

Creating box dataframe

```
box <- df %>%
  filter(Year == 2015 | Year == 2020)
```

All Continents (2015 vs 2020)

```
box %>%
  ggplot(aes(x = factor(Year), y = NEET_percentage, fill = Continent)) +
  geom_boxplot() +
  facet_wrap(~ Continent) +
  labs(x = "Year",
       y = "Average NEET (%)",
       title = "Average NEET by Continent (2015 VS 2020)") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  ) +
  scale_fill_manual(values = continent_colors)
```



Regression graphs

Creating regression dataframe

```
regression <- df %>%
  filter(Year >= 2005 & Year<=2020) %>%
  group_by(Continent, Year) %>%
  summarise(continent_avg_NEET = mean(NEET_percentage, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

```
weighted_regression <- df %>%
  filter(Year >= 2005 & Year<=2020)%>%
  group_by(Continent, Year) %>%
  summarise(continent_weighted_avg_NEET = mean(weighted_NEET_percentage, na.rm = TRUE))
```

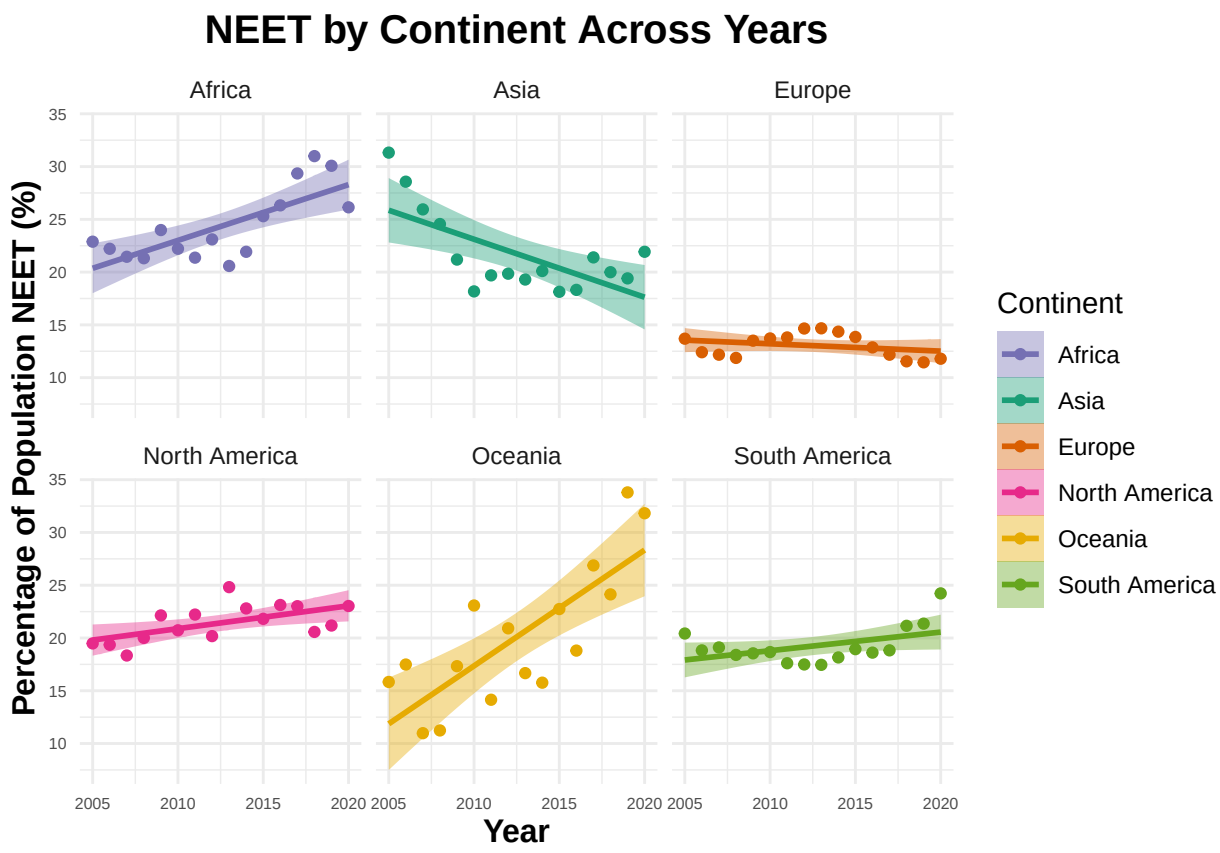
```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

```
weighted_regression_graph <- weighted_regression %>%
  filter(!(Continent == "Africa" & Year == 2007)) %>%
  filter(!(Continent == "Africa" & Year == 2008)) %>%
  filter(!(Continent == "Asia" & Year == 2005)) %>%
  filter(!(Continent == "Oceania" & Year == 2007)) %>%
  filter(!(Continent == "Oceania" & Year == 2008)) %>%
  filter(!(Continent == "South America" & Year == 2020))
```

All continents (2005 to 2020) – unweighted

```
regression %>%
  ggplot(aes(x = Year, y = continent_avg_NEET)) +
  geom_point(aes(colour = Continent), shape = 19) +
  geom_smooth(aes(colour = Continent, fill = Continent), method = "lm") +
  facet_wrap(~ Continent) +
  labs(x = "Year",
       y = "Percentage of Population NEET (%)",
       title = "NEET by Continent Across Years") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 6),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  ) +
  scale_color_manual(values = continent_colors)+
  scale_fill_manual(values = continent_colors)
```

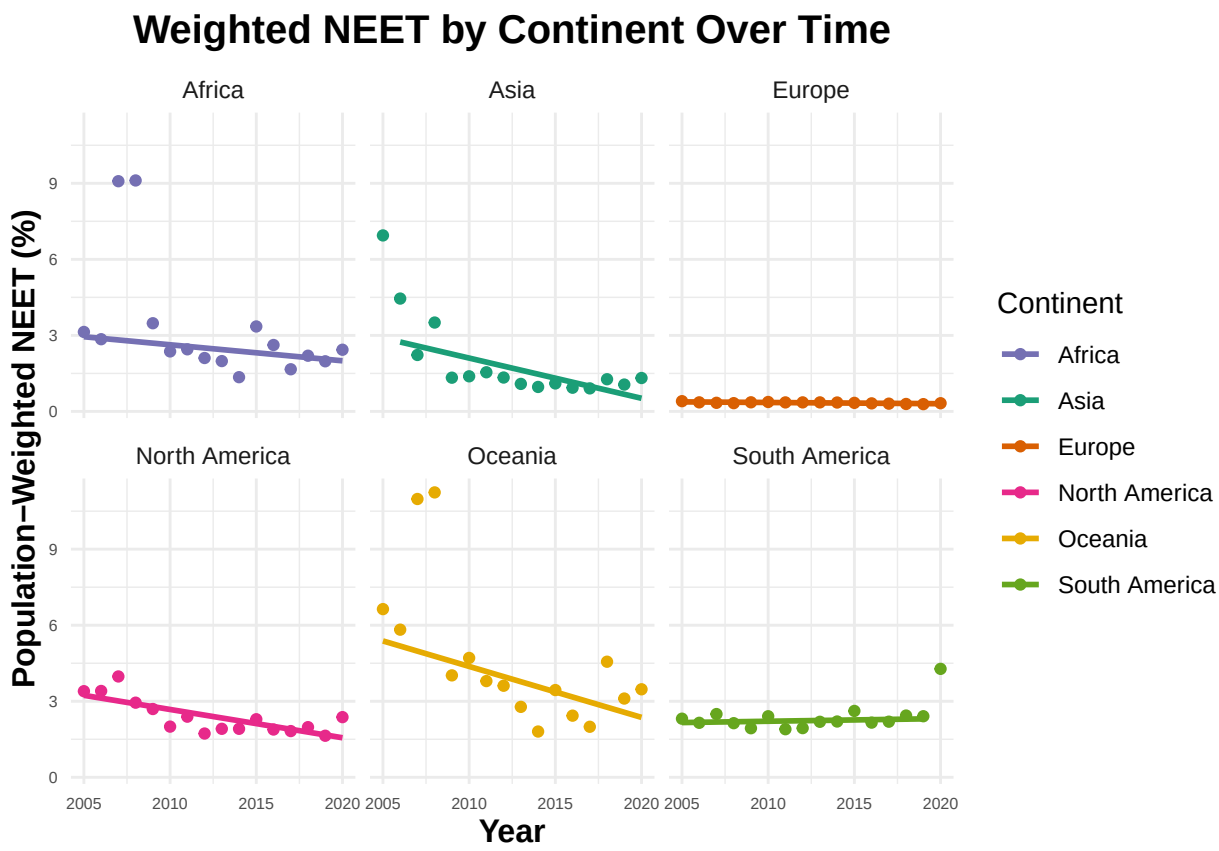
```
## `geom_smooth()` using formula = 'y ~ x'
```



All continents (2005 to 2020) – weighted

```
weighted_regression %>%
  ggplot(aes(x = Year, y = continent_weighted_avg_NEET)) +
  geom_point(aes(colour = Continent), shape = 19) +
  geom_smooth(data = weighted_regression_graph, se = FALSE, aes(colour = Continent),
              method = "lm") +
  facet_wrap(~ Continent) +
  labs(x = "Year",
       y = "Population-Weighted NEET (%)",
       title = "Weighted NEET by Continent Over Time") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 6),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  ) +
  scale_color_manual(values = continent_colors)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
  # scale_fill_manual(values = continent_colors)
```

Heatmaps

Creating Heatmap dataframe

```
heatmap <- df %>%
  filter(Year >= 2005 & Year<=2020) %>%
  group_by(Continent, Year) %>%
  summarise(continent_avg_NEET = mean(NEET_percentage, na.rm = TRUE))
```

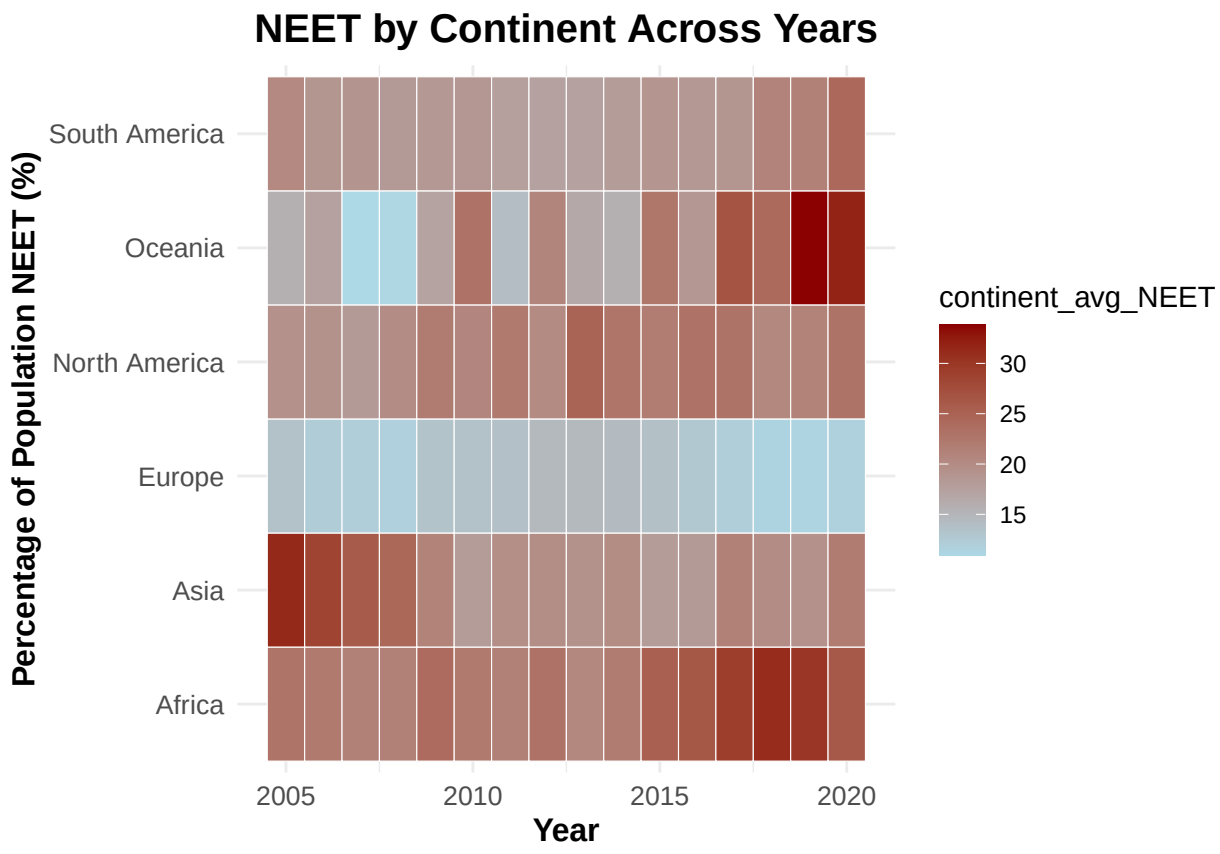
```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

```
weighted_heatmap <- df %>%
  filter(Year >= 2005 & Year<=2020)%>%
  group_by(Continent, Year) %>%
  summarise(continent_weighted_avg_NEET = mean(weighted_NEET_percentage, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

All Continents (2005 to 2020) – unweighted

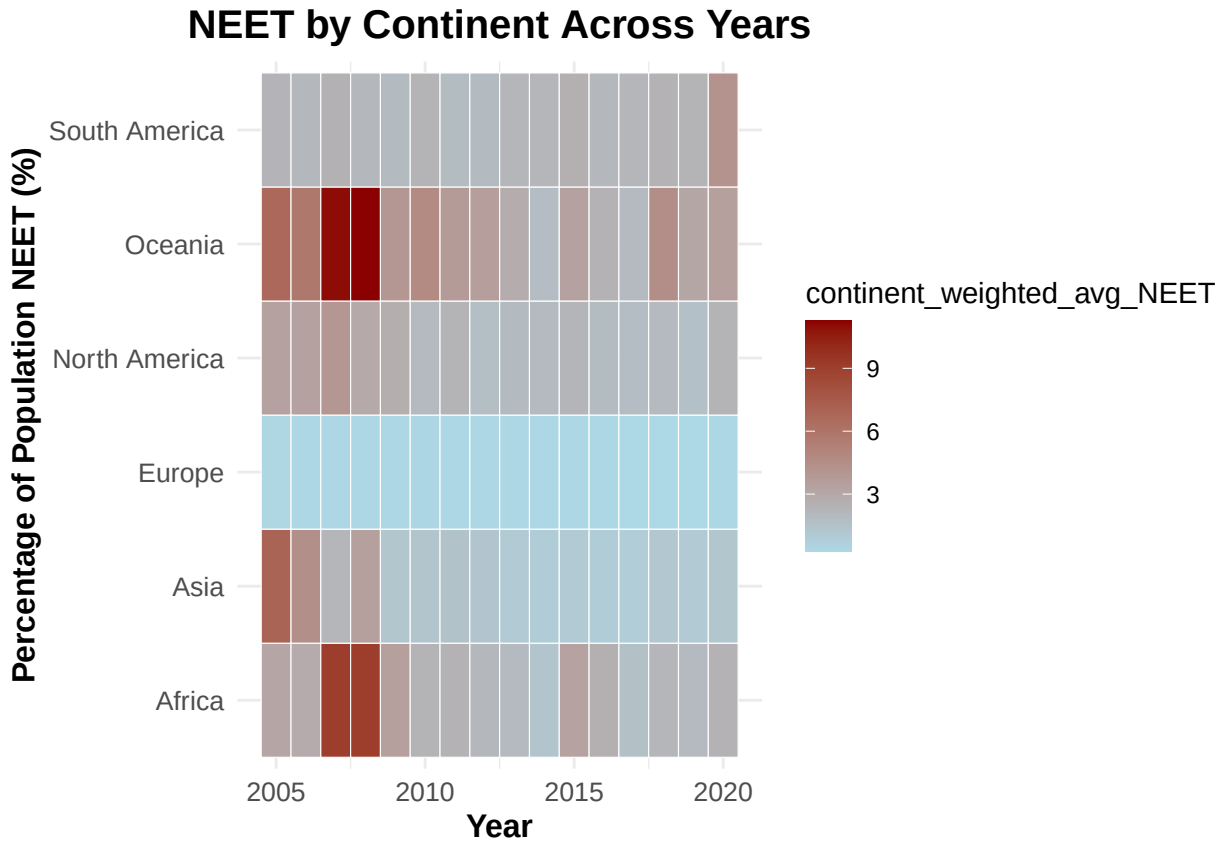
```
heatmap %>%
  ggplot(aes(x = Year,
             y = Continent,
             fill = continent_avg_NEET)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "darkred") +
  labs(x = "Year",
       y = "Percentage of Population NEET (%)",
       title = "NEET by Continent Across Years") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  )
```



All continents (2005 to 2020) – weighted

```
weighted_heatmap %>%
  ggplot(aes(x = Year,
             y = Continent,
```

```
            fill = continent_weighted_avg_NEET)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "lightblue", high = "darkred") +
  labs(x = "Year",
       y = "Percentage of Population NEET (%)",
       title = "NEET by Continent Across Years") +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = 'bold'),
    axis.text = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 15, face = 'bold'),
  )
```



Normalised NEET

Creating normalised NEET data frame

```
NEET_summary <- df %>%
  filter(Year >= 2005 & Year <= 2020) %>%
  group_by(Continent, Country) %>%
  summarise(
    first_year = min(Year),
    last_year = max(Year),
```

```
    neet_first = NEET_percentage[Year == first_year],
    neet_last = NEET_percentage[Year == last_year]
  )
```

```
## `summarise()` has grouped output by 'Continent'. You can override using the
## `.groups` argument.
```

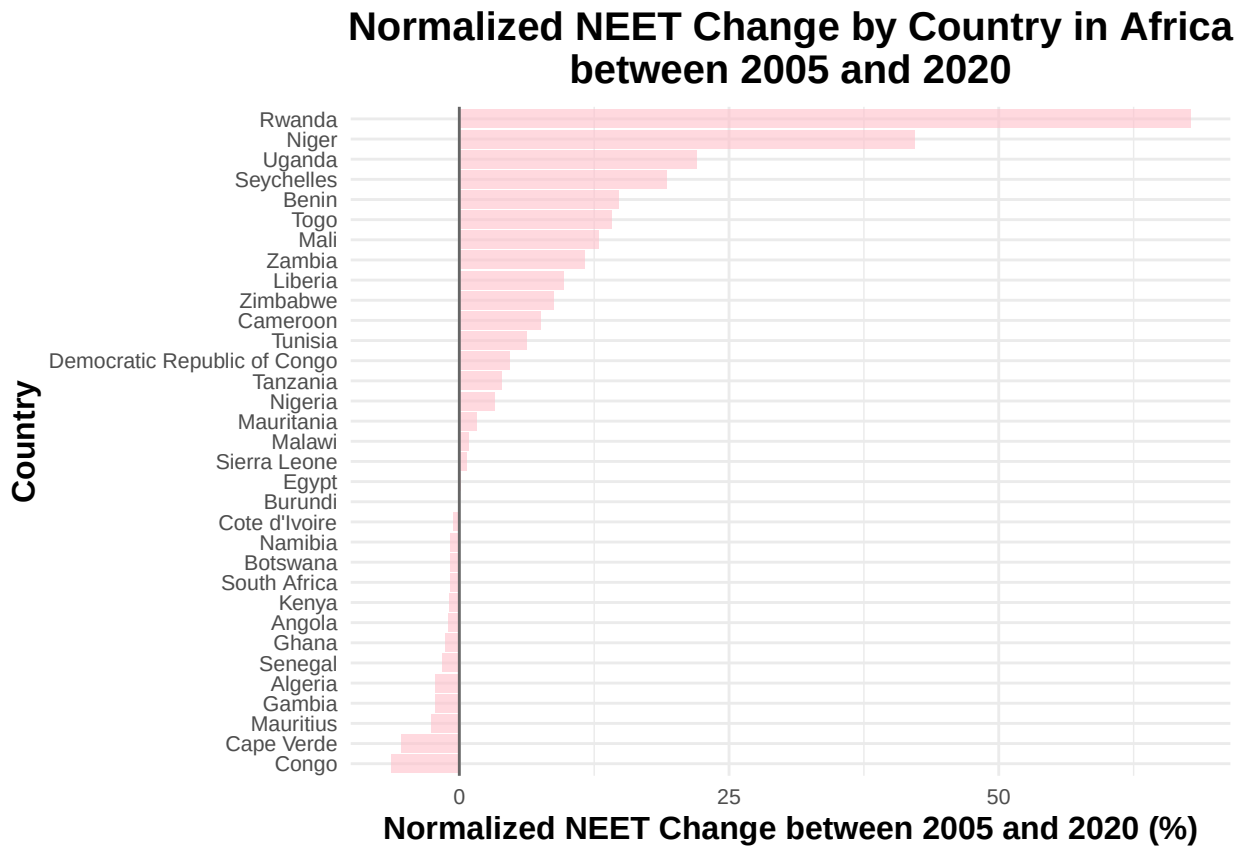
```
normalised_NEET <- NEET_summary %>%
  mutate(
    years_span = last_year - first_year,
    arithmetic_change =  neet_last - neet_first,
    normalised_NEET_decrease = (arithmetic_change / years_span) / neet_first *100
  )



africa_nomalised_NEET <- normalised_NEET %>%
  filter(Continent == 'Africa') %>%
  filter(!is.na(normalised_NEET_decrease))

asia_nomalised_NEET <- normalised_NEET %>%
  filter(Continent == 'Asia') %>%
  filter(!is.na(normalised_NEET_decrease))

europe_nomalised_NEET <- normalised_NEET %>%
  filter(Continent == 'Europe') %>%
  filter(!is.na(normalised_NEET_decrease))

na_nomalised_NEET <- normalised_NEET %>%
  filter(Continent == 'North America') %>%
  filter(!is.na(normalised_NEET_decrease))

oceania_nomalised_NEET <- normalised_NEET %>%
  filter(Continent == 'Oceania') %>%
  filter(!is.na(normalised_NEET_decrease))

sa_nomalised_NEET <- normalised_NEET %>%
  filter(Continent == 'South America') %>%
  filter(!is.na(normalised_NEET_decrease))
```

Plotting graphs

Africa

```
africa_nomalised_NEET %>%
  arrange(normalised_NEET_decrease) %>%
  mutate(Country = factor(Country, levels = Country)) %>%
  ggplot(aes(x = normalised_NEET_decrease, y = Country)) +
  geom_col(fill = "pink", alpha = 0.6) +
  geom_vline(xintercept = 0, colour = "grey40") +
```

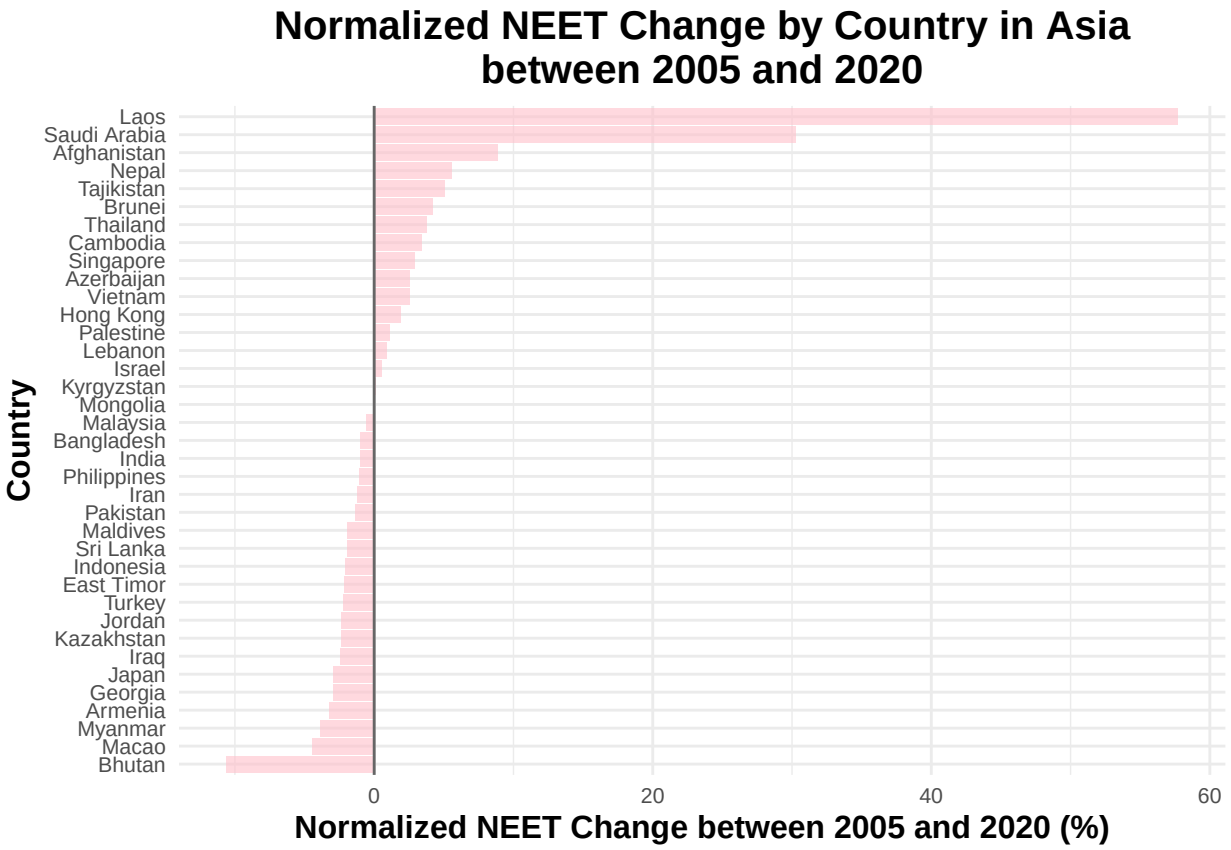
```
  labs(
    x = "Normalized NEET Change between 2005 and 2020 (%)",
    y = "Country",
    title = "Normalized NEET Change by Country in Africa\nbetween 2005 and 2020"
  ) +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = "bold"),
    axis.text  = element_text(size = 8),
    plot.title = element_text(hjust = 0.5, size = 15, face = "bold")
  )
```



Asia

```
asia_nomalised_NEET %>%
  arrange(normalised_NEET_decrease) %>%
  mutate(Country = factor(Country, levels = Country)) %>%
  ggplot(aes(x = normalised_NEET_decrease, y = Country)) +
  geom_col(fill = "pink", alpha = 0.6) +
  geom_vline(xintercept = 0, colour = "grey40") +
  labs(
    x = "Normalized NEET Change between 2005 and 2020 (%)",
    y = "Country",
```

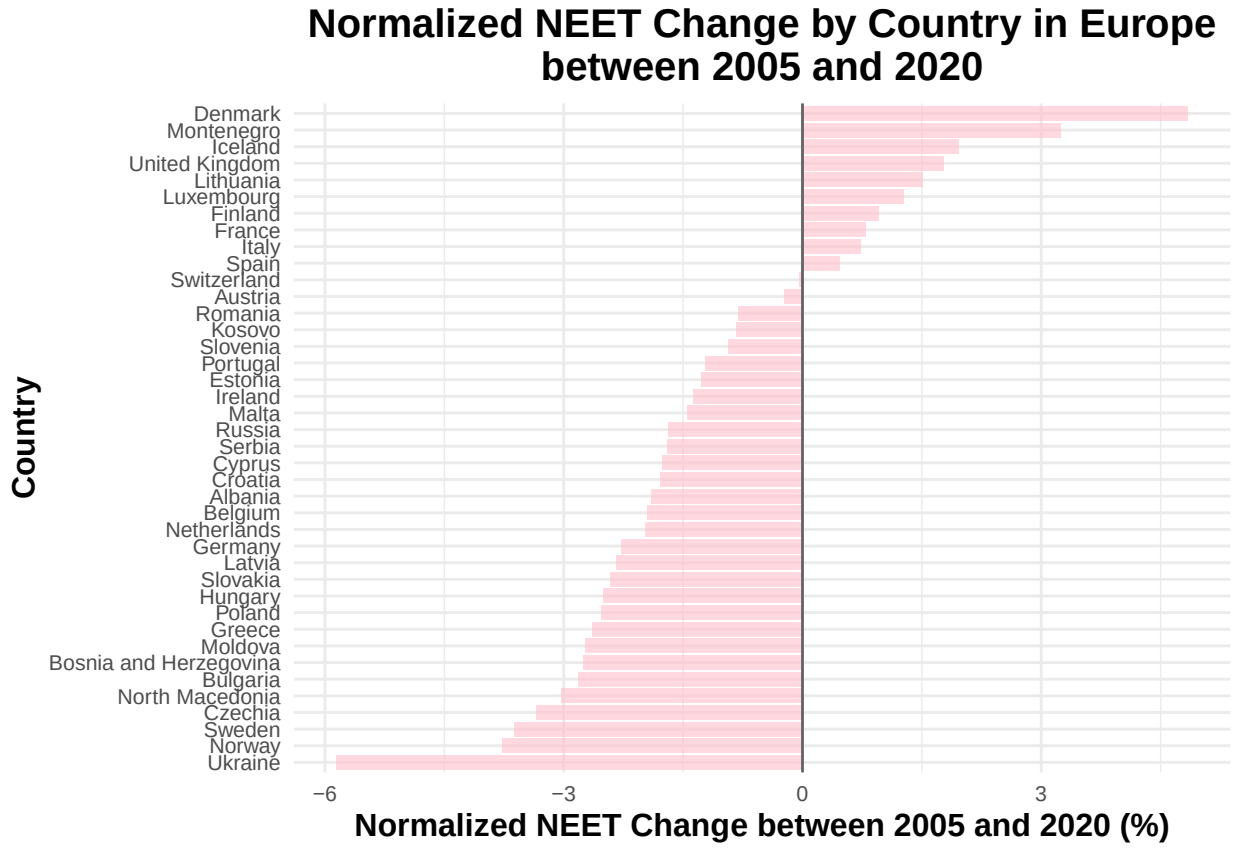
```
    title = "Normalized NEET Change by Country in Asia\nbetween 2005 and 2020"
  ) +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = "bold"),
    axis.text  = element_text(size = 8),
    plot.title = element_text(hjust = 0.5, size = 15, face = "bold")
  )
```



Europe

```
europe_nomalised_NEET %>%
  filter(!(Country == "Belarus")) %>%
  arrange(normalised_NEET_decrease) %>%
  mutate(Country = factor(Country, levels = Country)) %>%
  ggplot(aes(x = normalised_NEET_decrease, y = Country)) +
  geom_col(fill = "pink", alpha = 0.6) +
  geom_vline(xintercept = 0, colour = "grey40") +
  labs(
    x = "Normalized NEET Change between 2005 and 2020 (%)",
    y = "Country",
    title = "Normalized NEET Change by Country in Europe\nbetween 2005 and 2020"
  ) +
```

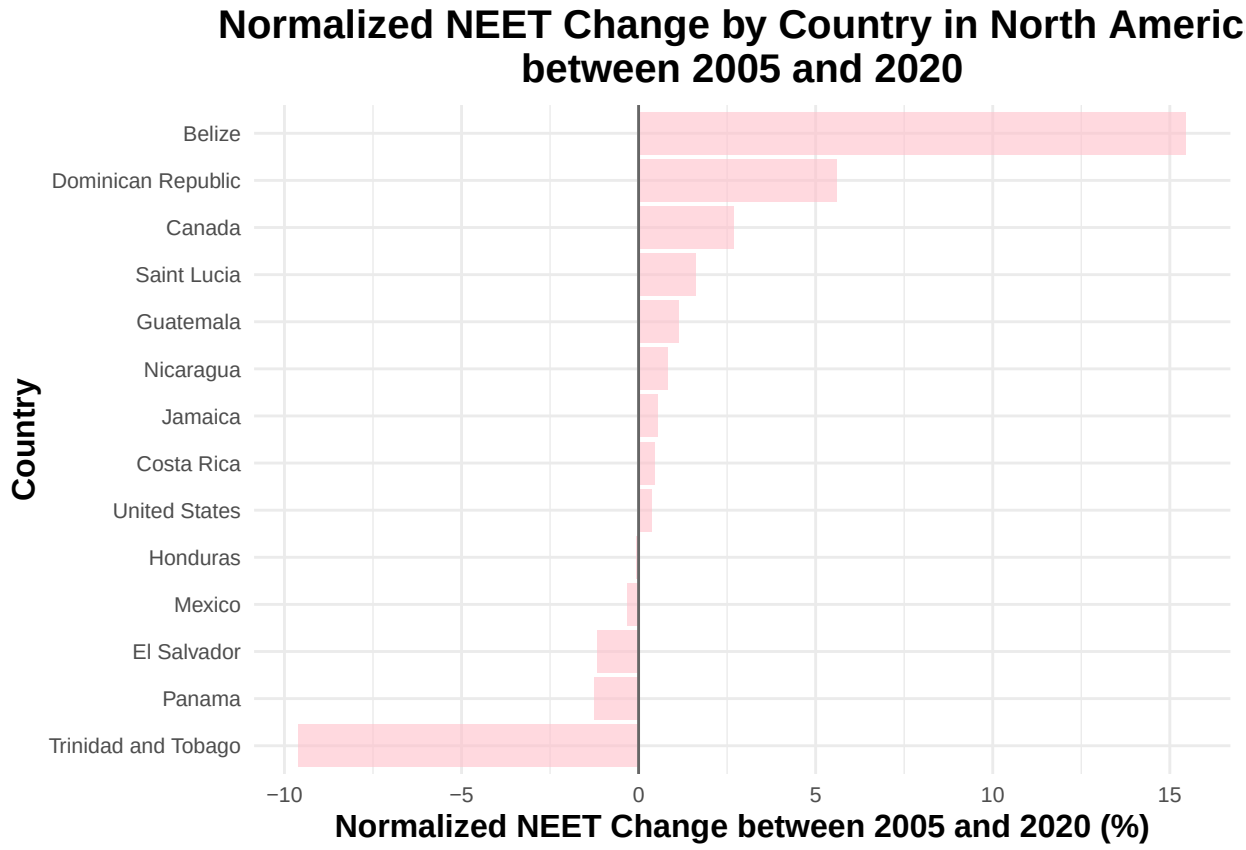
```
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = "bold"),
    axis.text  = element_text(size = 8),
    plot.title = element_text(hjust = 0.5, size = 15, face = "bold")
  )
```



North America

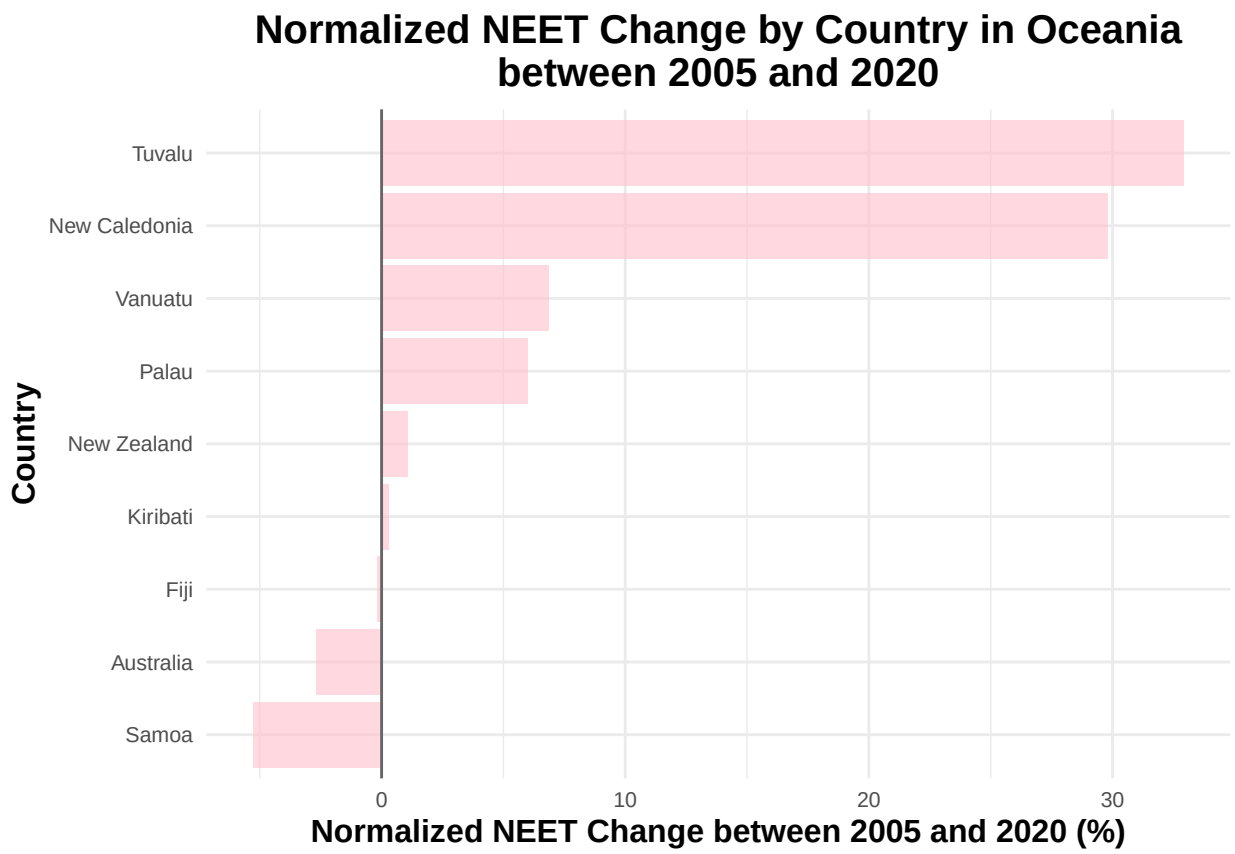
```
na_nomalised_NEET %>%
  filter(!(Country == "Curacao")) %>%
  arrange(normalised_NEET_decrease) %>%
  mutate(Country = factor(Country, levels = Country)) %>%
  ggplot(aes(x = normalised_NEET_decrease, y = Country)) +
  geom_col(fill = "pink", alpha = 0.6) +
  geom_vline(xintercept = 0, colour = "grey40") +
  labs(
    x = "Normalized NEET Change between 2005 and 2020 (%)",
    y = "Country",
    title = "Normalized NEET Change by Country in North America\nbetween 2005 and 2020"
  ) +
  theme_minimal() +
  theme(
```

```
    axis.title = element_text(size = 12, face = "bold"),
    axis.text  = element_text(size = 8),
    plot.title = element_text(hjust = 0.5, size = 15, face = "bold")
  )
```



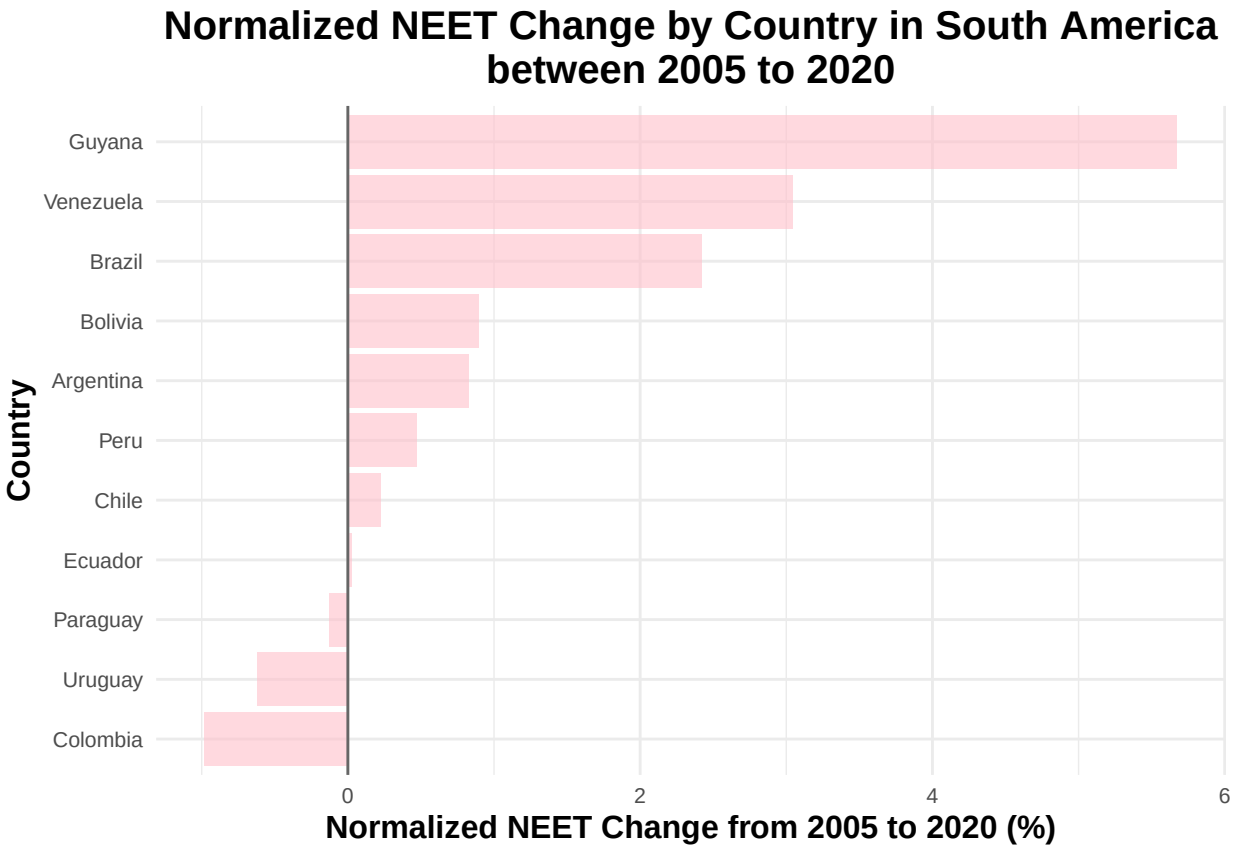
Oceania

```
oceania_nomalised_NEET %>%
  arrange(normalised_NEET_decrease) %>%
  mutate(Country = factor(Country, levels = Country)) %>%
  ggplot(aes(x = normalised_NEET_decrease, y = Country)) +
  geom_col(fill = "pink", alpha = 0.6) +
  geom_vline(xintercept = 0, colour = "grey40") +
  labs(
    x = "Normalized NEET Change between 2005 and 2020 (%)",
    y = "Country",
    title = "Normalized NEET Change by Country in Oceania\nbetween 2005 and 2020"
  ) +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = "bold"),
    axis.text  = element_text(size = 8),
    plot.title = element_text(hjust = 0.5, size = 15, face = "bold")
  )
```



South America

```
sa_nomalised_NEET %>%
  arrange(normalised_NEET_decrease) %>%
  mutate(Country = factor(Country, levels = Country)) %>%
  ggplot(aes(x = normalised_NEET_decrease, y = Country)) +
  geom_col(fill = "pink", alpha = 0.6) +
  geom_vline(xintercept = 0, colour = "grey40") +
  labs(
    x = "Normalized NEET Change from 2005 to 2020 (%)",
    y = "Country",
    title = "Normalized NEET Change by Country in South America\nbetween 2005 to 2020"
  ) +
  theme_minimal() +
  theme(
    axis.title = element_text(size = 12, face = "bold"),
    axis.text  = element_text(size = 8),
    plot.title = element_text(hjust = 0.5, size = 15, face = "bold")
  )
```


Normalized NEET Change by Country in South America between 2005 to 2020
Guyana
Venezuela
Brazil
Bolivia
Argentina
Peru
Chile
Ecuador
Paraguay
Uruguay
Colombia
Country
0
2
4
6
Normalized NEET Change from 2005 to 2020 (%)