Target 2

First, we load the packages we need and define a custom colour palette for continents that will be reused in all plots.

```
library(tidyverse)
library(scales)
library(ggthemes)

continent_colors <- c(
  "Asia" = "#377EB8",
  "Europe"= "#4DAF4A",
  "Africa" = "#984EA3",
  "North America"= "#E41A1C",
  "South America"= "#FF7F00",
  "Oceania"= "#A65628"
)
```

Then, we import all four raw datasets: NEET, continent classification, youth unemployment (WDI) and population.

```
neet_raw <- read_csv("youth-not-in-education-employment-training.csv")
```

```
## Rows: 1908 Columns: 4
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (2): Entity, Code
## dbl (2): Year, Share of youth not in education, employment or training, tota...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
continents_raw <- read_csv("continents-according-to-our-world-in-data.csv")
```

```
## Rows: 285 Columns: 4
## -- Column specification --------------------------------------------------
## Delimiter: ","
## chr (3): Entity, Code, Continent
## dbl (1): Year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
youth_unemp_raw <- read_csv("youth_unemployment.csv", skip = 3)
```

```
## New names:
## Rows: 266 Columns: 70
## -- Column specification
## ---------------------------------------------------------- Delimiter: "," chr
## (4): Country Name, Country Code, Indicator Name, Indicator Code dbl (34): 1991,
## 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, ... lgl (32): 1960,
## 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, ...
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * '' -> '...70'
```

```r
population_raw <- read_csv("population.csv")
```

```
## Rows: 16930 Columns: 4
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr (2): Country Name, Country Code
## dbl (2): Year, Value
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Here we clean the NEET data: we keep only the relevant columns, convert year and NEET values to numeric, and restrict the sample to 1990–2020 with non-missing values.

```r
neet <- neet_raw %>%
  select(Entity, Code, Year, neet_share = `Share of youth not in education, employment or training, tota
  ) %>%
  mutate(
    Year = as.integer(Year),
    neet_share = as.numeric(neet_share)
  ) %>%
  filter(!is.na(Code), !is.na(neet_share),
    Year >= 1990,
    Year <= 2020
  )
```

We cleans the continent lookup table: we keep country name, ISO3 code and continent, drop duplicate codes, and remove Antarctica which we do not analyse.

```r
continents <- continents_raw %>% select(Entity, Code, Continent) %>%
  distinct(Code, .keep_all = TRUE) %>% filter(Continent != "Antarctica")
```

We now attach the continent information to each NEET observation by joining on both country code and name.

```
neet_continent <- neet %>% inner_join(continents, by = c("Code", "Entity"))
```

We also need to reshape and clean the WDI youth unemployment data: we go from wide (one column per year) to long format, rename columns, convert year and unemployment rate to numeric, restrict to 1990–2020, and then attach each country to its continent.

```
year_cols <- names(youth_unemp_raw)[5:ncol(youth_unemp_raw)]

youth_unemp_long <- youth_unemp_raw %>%
  rename(
    Country_Name   = `Country Name`,
    Country_Code   = `Country Code`,
    Indicator_Name = `Indicator Name`,
    Indicator_Code = `Indicator Code`
  ) %>%
  pivot_longer(
    cols = all_of(year_cols),
    names_to = "Year",
    values_to = "youth_unemployment"
  ) %>%
  mutate(
    Year = as.integer(Year),
    youth_unemployment = as.numeric(youth_unemployment)
  ) %>%
  filter(
    Year >= 1990,
    Year <= 2020,
    !is.na(youth_unemployment)
  )
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `Year = as.integer(Year)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
youth_unemp_continent <- youth_unemp_long %>%
  inner_join(continents, by = c("Country_Code" = "Code", "Country_Name" = "Entity")
  )
```

Here we clean the population data: we rename variables for clarity, convert year and population to numeric, and keep observations between 1990 and 2020 with valid population values.

```r
population <- population_raw %>%
  rename(
    Country_Name = `Country Name`,
    Country_Code = `Country Code`,
    population = Value
  ) %>%
  mutate(
    Year = as.integer(Year),
    population = as.numeric(population)
  ) %>%
  filter(
    Year >= 1990,
    Year <= 2020,
    !is.na(population)
  )
```

This chunk merges NEET with population and computes population-weighted NEET averages by continent and year. We also store the maximum NEET value for later use in graph scaling.

```r
neet_pop <- neet_continent %>%
  inner_join(
    population,
    by = c("Code" = "Country_Code", "Year" = "Year")
  )

neet_continent_year <- neet_pop %>%
  group_by(Continent, Year) %>%
  summarise(
    total_pop = sum(population),
    total_neet_pop = sum(neet_share/100 * population),
    .groups = "drop"
  ) %>%
  mutate(
    neet_avg = (total_neet_pop / total_pop) * 100
  )

max_neet <- max(neet_continent_year$neet_avg, na.rm = TRUE)
```

Graph 1: trend in population-weighted NEET by continent We shade the SDG period (2015–2020), draw a dashed line at 2015, highlight the SDG years with a thicker line, and label the pre-SDG and SDG implementation periods.

```r
neet_key_years <- neet_continent_year %>%
  filter(Year %in% c(2000, 2015, 2019, 2020))

neet_sdg <- neet_continent_year %>%
  filter(Year >= 2015, Year <= 2020)
```
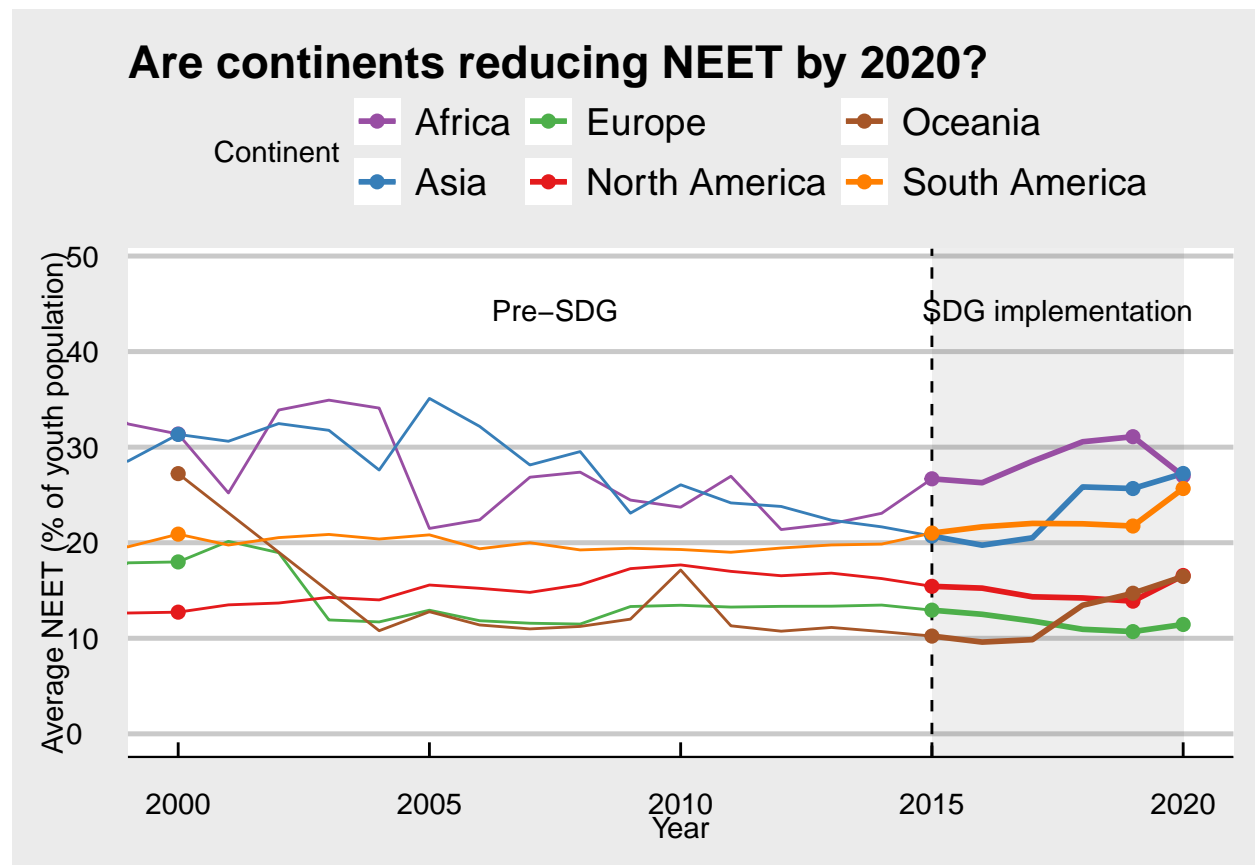
```
ggplot(neet_continent_year,
       aes(x = Year, y = neet_avg, colour = Continent)) +
  annotate("rect",
           xmin = 2015, xmax = 2020,
           ymin = -Inf, ymax = Inf,
           alpha = 0.1) +
  geom_vline(xintercept = 2015, linetype = "dashed") +
  geom_line(linewidth = 0.5) +
  geom_line(data = neet_sdg, linewidth = 1) +
  geom_point(data = neet_key_years, size = 2) +
  annotate("text",
           x = 2007.5, y = max_neet + 1,
           label = "Pre-SDG") +
  annotate("text",
           x = 2017.5, y = max_neet + 1,
           label = "SDG implementation") +
  scale_color_manual(values = continent_colors) +
  ggtitle("Are continents reducing NEET by 2020?") +
  xlab("Year") +
  ylab("Average NEET (% of youth population)") +
  coord_cartesian(xlim = c(2000, 2020),
                  ylim = c(0, max_neet + 5)) +
  theme_economist_white()
```
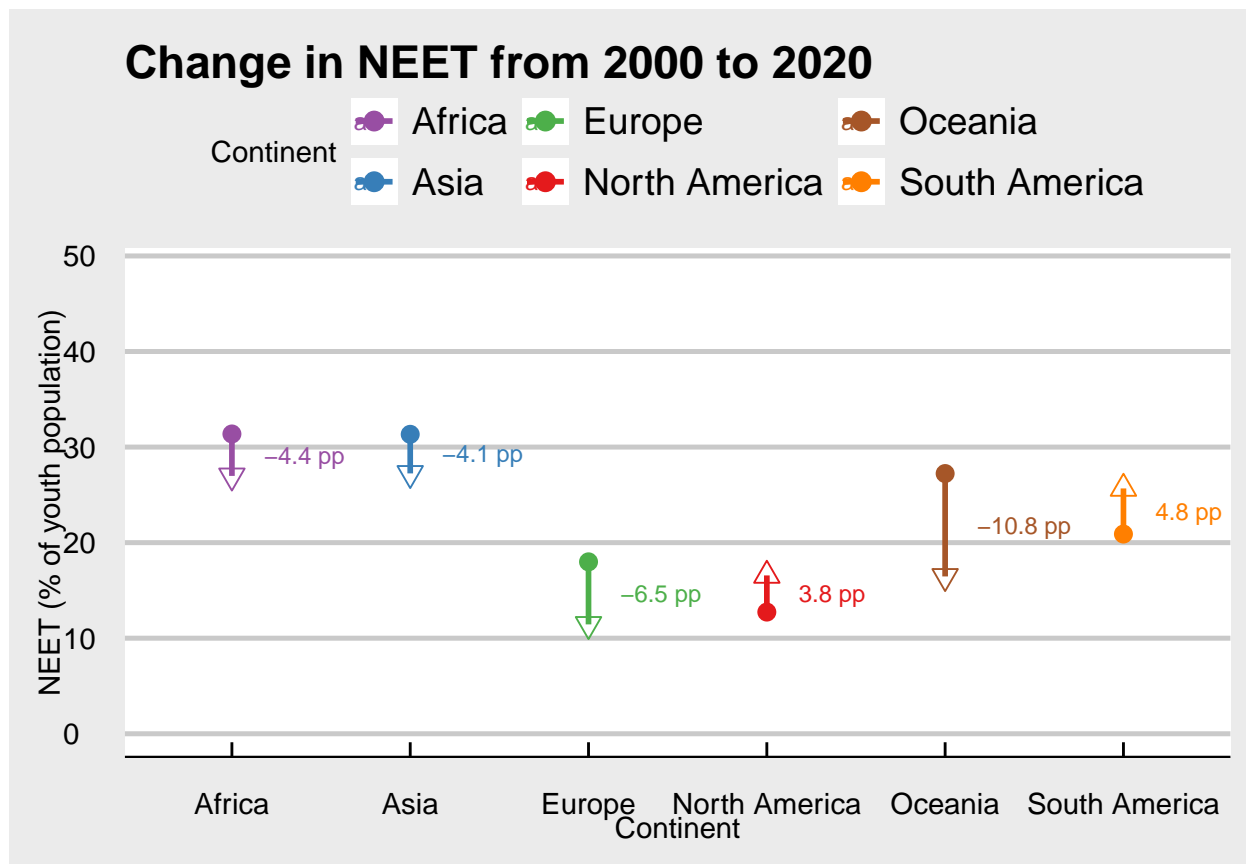
Here we extract the population-weighted NEET levels for three key years (2000, 2015, 2020) by continent.
These values are used to compute changes for the dumbbell plots.

```
neet_2000 <- neet_continent_year %>% filter(Year == 2000) %>% select(Continent, neet_2000 = neet_avg)
neet_2015 <- neet_continent_year %>% filter(Year == 2015) %>% select(Continent, neet_2015 = neet_avg)
neet_2020 <- neet_continent_year %>% filter(Year == 2020) %>% select(Continent, neet_2020 = neet_avg)
```

Graph 2: We visualise the change in NEET between 2000 and 2020. For each continent we draw a vertical
line from 2000 to 2020, put a circle at the starting value, a triangle (up/down) at the end value, and place
a label on the right showing the change in percentage points.

```
neet_2000_2020 <- neet_2000 %>%
  inner_join(neet_2020, by = "Continent") %>%
  mutate(
    change_2000_2020 = neet_2020 - neet_2000,
    mid_neet = (neet_2000 + neet_2020) / 2,
    end_shape = ifelse(change_2000_2020 >= 0, 24, 25)
  )

ggplot(neet_2000_2020, aes(x = Continent, colour = Continent)) +
  geom_segment(aes(xend = Continent,
                   y = neet_2000, yend = neet_2020),
               linewidth = 1) +
  geom_point(aes(y = neet_2000), size = 3, shape = 16) +
  geom_point(aes(y = neet_2020, shape = end_shape), size = 3) +
  scale_shape_identity() +
  geom_text(aes(y = mid_neet,
                label = paste0(round(change_2000_2020, 1), " pp")),
            position = position_nudge(x = 0.18),
            hjust = 0,
            size = 3) +
  scale_color_manual(values = continent_colors) +
  ggtitle("Change in NEET from 2000 to 2020 ") +
  xlab("Continent") +
  ylab("NEET (% of youth population)") +
  coord_cartesian(ylim = c(0, max_neet + 5)) +
  theme_economist_white()
```
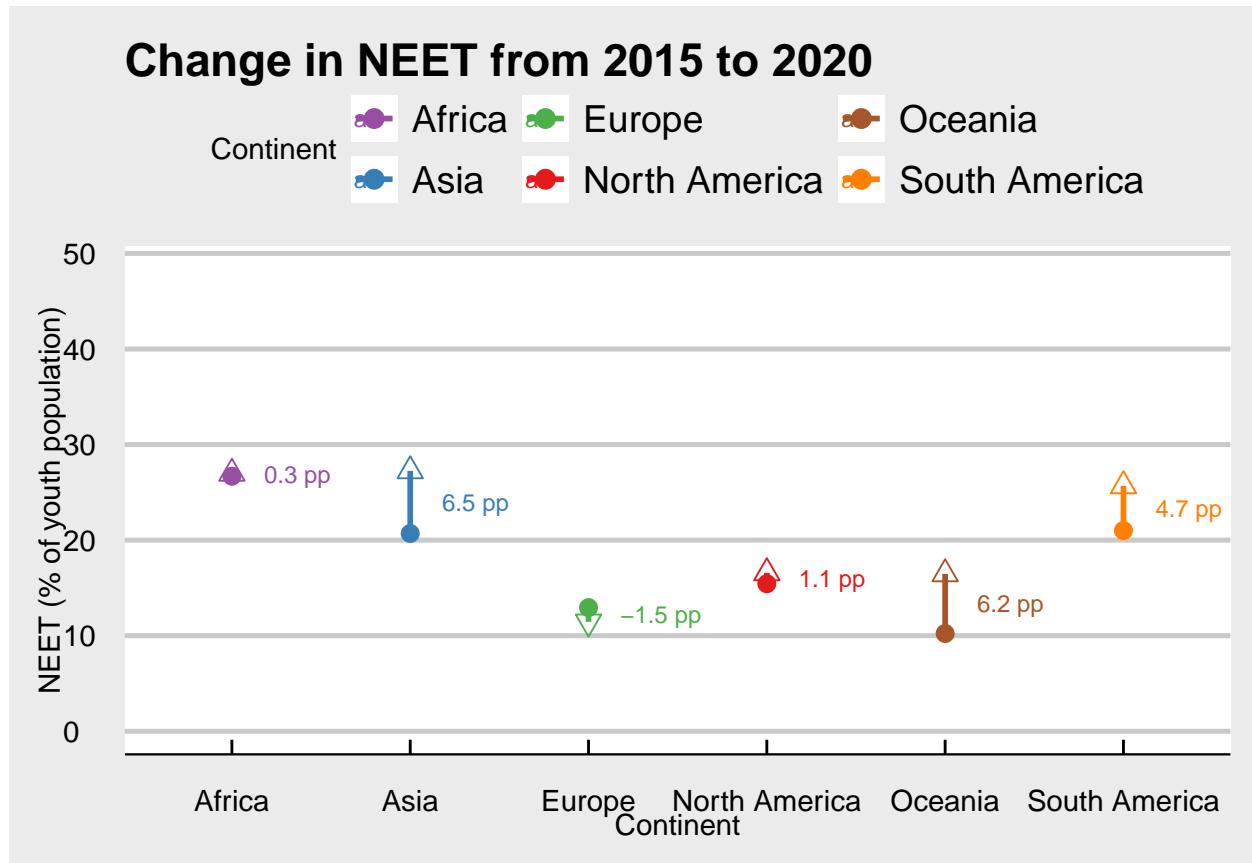
Graph 3: We repeat the same dumbbell design but now compares NEET levels between 2015 and 2020. It highlights whether NEET improved or worsened during the SDG period.

```r
neet_2015_2020 <- neet_2015 %>%
  inner_join(neet_2020, by = "Continent") %>%
  mutate(
    change_2015_2020 = neet_2020 - neet_2015,
    mid_neet         = (neet_2015 + neet_2020) / 2,
    end_shape        = ifelse(change_2015_2020 >= 0, 24, 25)
  )

ggplot(neet_2015_2020, aes(x = Continent, colour = Continent)) +
  geom_segment(aes(xend = Continent,
                   y = neet_2015, yend = neet_2020),
               linewidth = 1) +
  geom_point(aes(y = neet_2015), size = 3, shape = 16) +
  geom_point(aes(y = neet_2020, shape = end_shape), size = 3) +
  scale_shape_identity() +
  geom_text(aes(y = mid_neet,
                label = paste0(round(change_2015_2020, 1), " pp")),
            position = position_nudge(x = 0.18),
            hjust = 0,
            size = 3) +
```

```
scale_color_manual(values = continent_colors) +
ggtitle("Change in NEET from 2015 to 2020") +
xlab("Continent") +
ylab("NEET (% of youth population)") +
coord_cartesian(ylim = c(0, max_neet + 5)) +
theme_economist_white()
```


This chunk merges youth unemployment with population and computes population-weighted unemployment rates by continent and year. We again store the maximum value to control the y-axis in the plots.

```
unemp_pop <- youth_unemp_continent %>%
  inner_join(population, by = c("Country_Code", "Year"))

unemp_continent_year <- unemp_pop %>%
  group_by(Continent, Year) %>%
  summarise(
    total_pop       = sum(population),
    total_unemp_pop = sum(youth_unemployment/100 * population),
    .groups         = "drop"
  ) %>%
  mutate(
    unemp_avg = (total_unemp_pop / total_pop) * 100
```
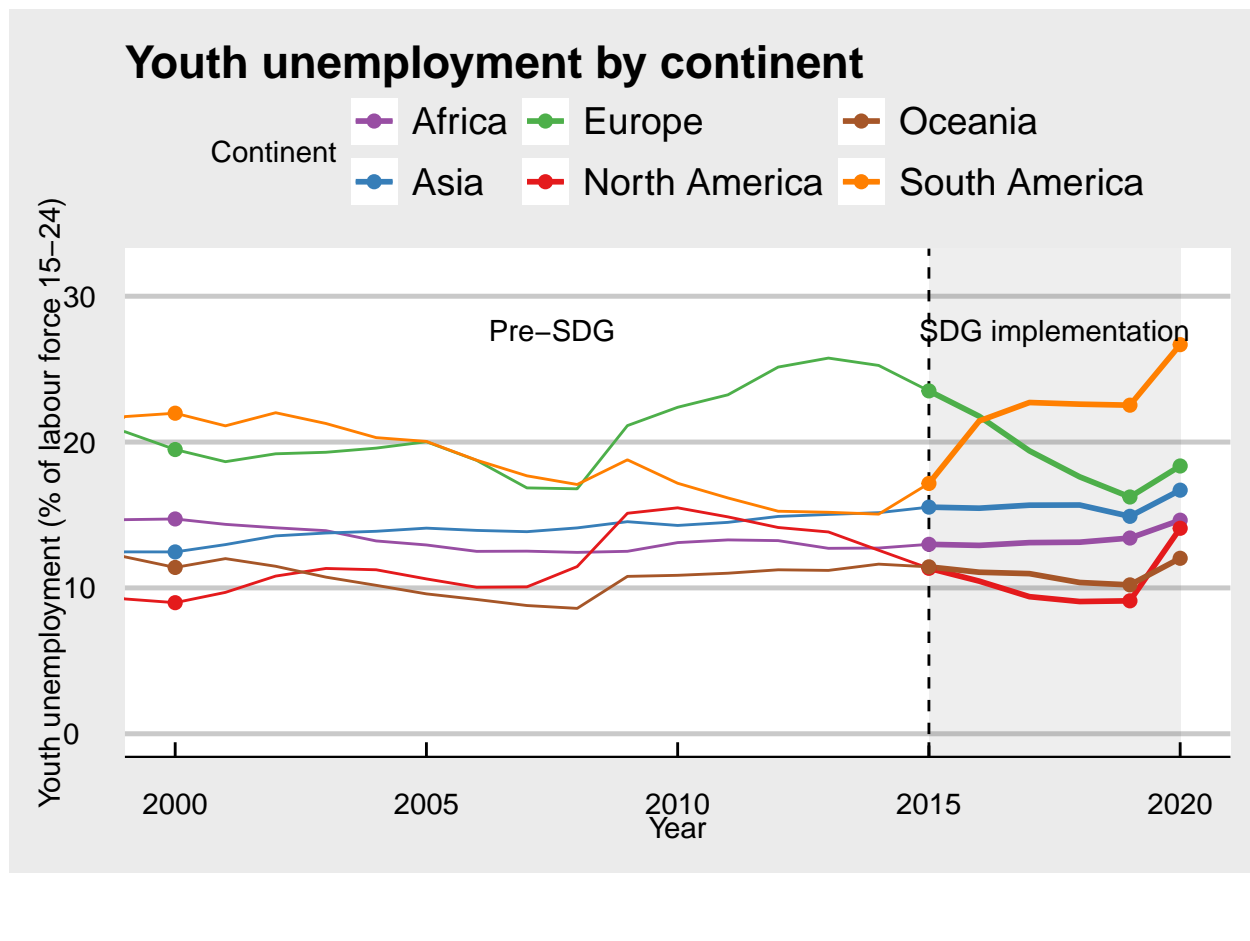
```
  )

max_unemp <- max(unemp_continent_year$unemp_avg, na.rm = TRUE)
```

Graph 4: The evolution of population-weighted youth unemployment by continent, using the same SDG visual structure as Graph 1 (pre-SDG vs SDG implementation period).

```
unemp_key_years <- unemp_continent_year %>%
  filter(Year %in% c(2000, 2015, 2019, 2020))

unemp_sdg <- unemp_continent_year %>%
  filter(Year >= 2015, Year <= 2020)

ggplot(unemp_continent_year,
       aes(x = Year, y = unemp_avg, colour = Continent)) +
  annotate("rect",
           xmin = 2015, xmax = 2020,
           ymin = -Inf, ymax = Inf,
           alpha = 0.1) +
  geom_vline(xintercept = 2015, linetype = "dashed") +
  geom_line(linewidth = 0.5) +
  geom_line(data = unemp_sdg, linewidth = 1) +
  geom_point(data = unemp_key_years, size = 2) +
  annotate("text",
           x = 2007.5, y = max_unemp + 1,
           label = "Pre-SDG") +
  annotate("text",
           x = 2017.5, y = max_unemp + 1,
           label = "SDG implementation") +
  scale_color_manual(values = continent_colors) +
  ggtitle("Youth unemployment by continent") +
  xlab("Year") +
  ylab("Youth unemployment (% of labour force 15-24)") +
  coord_cartesian(xlim = c(2000, 2020),
                  ylim = c(0, max_unemp + 5)) +
  theme_economist_white()
```

Youth unemployment by continent

Continent
- Africa
- Asia
- Europe
- North America
- Oceania
- South America

Here we extract population-weighted unemployment levels for 2000, 2015 and 2020 by continent, to compute changes for the remaining dumbbell plots.

```
unemp_2000 <- unemp_continent_year %>% filter(Year == 2000) %>% select(Continent, unemp_2000 = unemp_avg
unemp_2015 <- unemp_continent_year %>% filter(Year == 2015) %>% select(Continent, unemp_2015 = unemp_avg
unemp_2020 <- unemp_continent_year %>% filter(Year == 2020) %>% select(Continent, unemp_2020 = unemp_avg
```

Graph 5: We want to see the change in youth unemployment between 2000 and 2020. As in Graph 2, each continent has a vertical line from start to end, a circle at the 2000 value, a triangle at the 2020 value (up or down), and a label to the right with the change in percentage points.
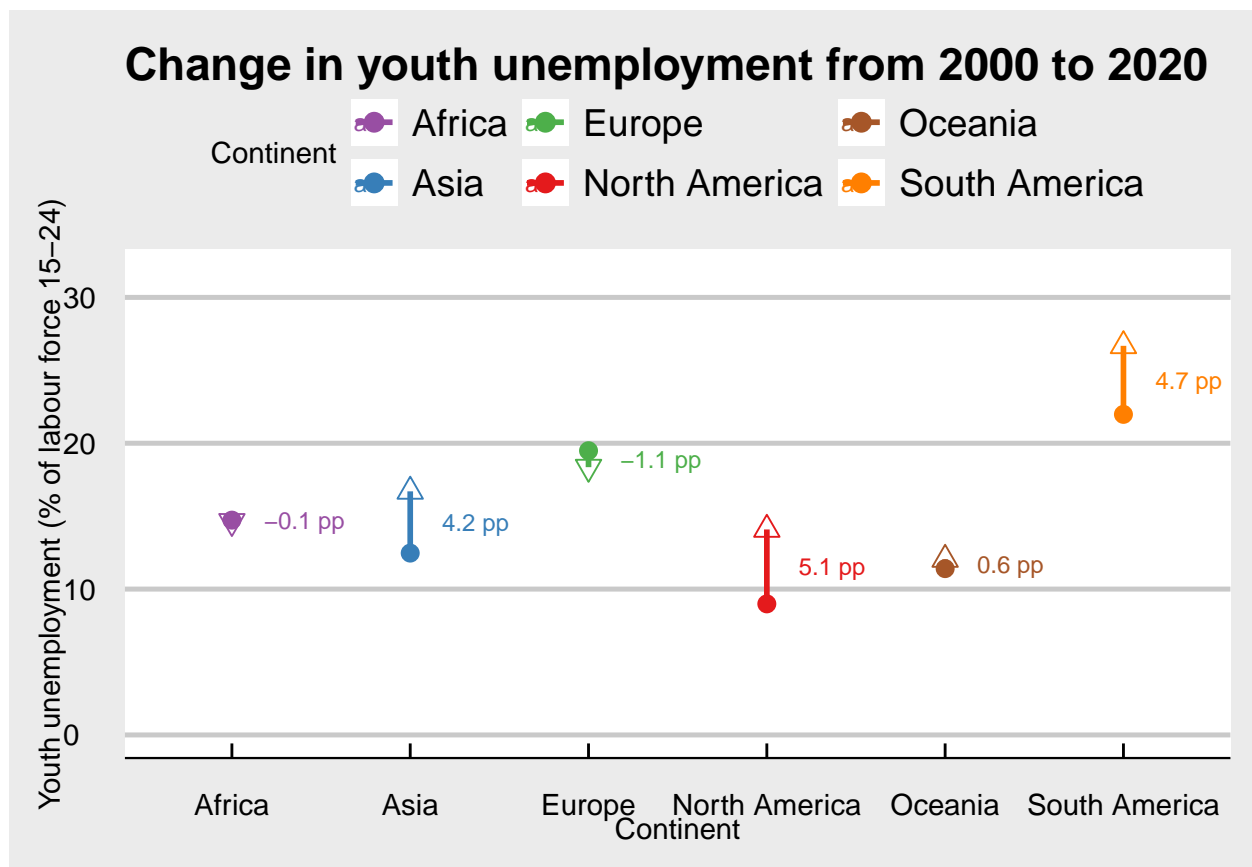
```
unemp_2000_2020 <- unemp_2000 %>%
  inner_join(unemp_2020, by = "Continent") %>%
  mutate(
    change_2000_2020 = unemp_2020 - unemp_2000,
    mid_unemp        = (unemp_2000 + unemp_2020) / 2,
    end_shape        = ifelse(change_2000_2020 >= 0, 24, 25)
  )

ggplot(unemp_2000_2020, aes(x = Continent, colour = Continent)) +
```

```r
geom_segment(aes(xend = Continent,
                 y = unemp_2000, yend = unemp_2020),
             linewidth = 1) +
geom_point(aes(y = unemp_2000), size = 3, shape = 16) +
geom_point(aes(y = unemp_2020, shape = end_shape), size = 3) +
scale_shape_identity() +
geom_text(aes(y = mid_unemp,
              label = paste0(round(change_2000_2020, 1), " pp")),
          position = position_nudge(x = 0.18),
          hjust = 0,
          size = 3) +
scale_color_manual(values = continent_colors) +
ggtitle("Change in youth unemployment from 2000 to 2020") +
xlab("Continent") +
ylab("Youth unemployment (% of labour force 15-24)") +
coord_cartesian(ylim = c(0, max_unemp + 5)) +
theme_economist_white()
```


Graph 6: We apply the same visual logic to the SDG period only, comparing youth unemployment between 2015 and 2020 for each continent.

```r
unemp_2015_2020 <- unemp_2015 %>%
  inner_join(unemp_2020, by = "Continent") %>%
  mutate(
    change_2015_2020 = unemp_2020 - unemp_2015,
    mid_unemp = (unemp_2015 + unemp_2020) / 2,
    end_shape = ifelse(change_2015_2020 >= 0, 24, 25)
  )

ggplot(unemp_2015_2020, aes(x = Continent, colour = Continent)) +
  geom_segment(aes(xend = Continent,
                   y = unemp_2015, yend = unemp_2020),
               linewidth = 1) +
  geom_point(aes(y = unemp_2015), size = 3, shape = 16) +
  geom_point(aes(y = unemp_2020, shape = end_shape), size = 3) +
  scale_shape_identity() +
  geom_text(aes(y = mid_unemp,
                label = paste0(round(change_2015_2020, 1), " pp")),
            position = position_nudge(x = 0.18),
            hjust = 0,
            size = 3) +
  scale_color_manual(values = continent_colors) +
  ggtitle("Change in youth unemployment from 2015 to 2020") +
  xlab("Continent") +
  ylab("Youth unemployment (% of labour force 15-24)") +
  coord_cartesian(ylim = c(0, max_unemp + 5)) +
  theme_economist_white()
```

Change in youth unemployment from 2015 to 2020