Target 2

First, We load all required packages and define a color palette for continents.

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

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

Load datasets:

We import the NEET dataset, the continent mapping, youth unemployment, 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.
```

```
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`
```

```
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.
```

---

3. Clean NEET data (limit to 2020)

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

Clean continent dataset

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

Attach continent to NEET

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

Clean youth unemployment (WDI)

```r
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
```

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

Clean population data

```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)
  )
```
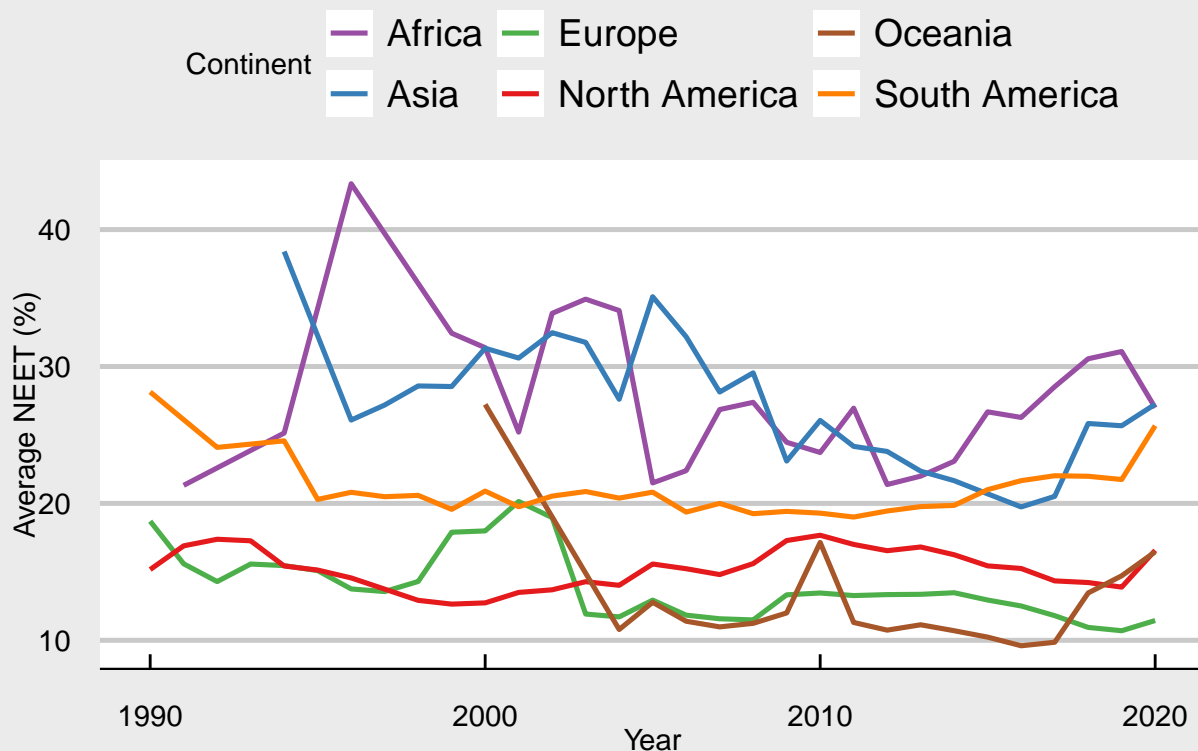
Combine NEET + population, compute weighted averages

```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
  )
```

Graph 1 - NEET trend (population-weighted)**

```r
ggplot(neet_continent_year,
       aes(x = Year, y = neet_avg, colour = Continent)) +
  geom_line(linewidth = 0.9) +
  theme_economist_white() +
  scale_color_manual(values = continent_colors) +
  ggtitle("Are continents reducing NEET by 2020?") +
  xlab("Year") +
  ylab("Average NEET (%)")
```

Are continents reducing NEET by 2020?

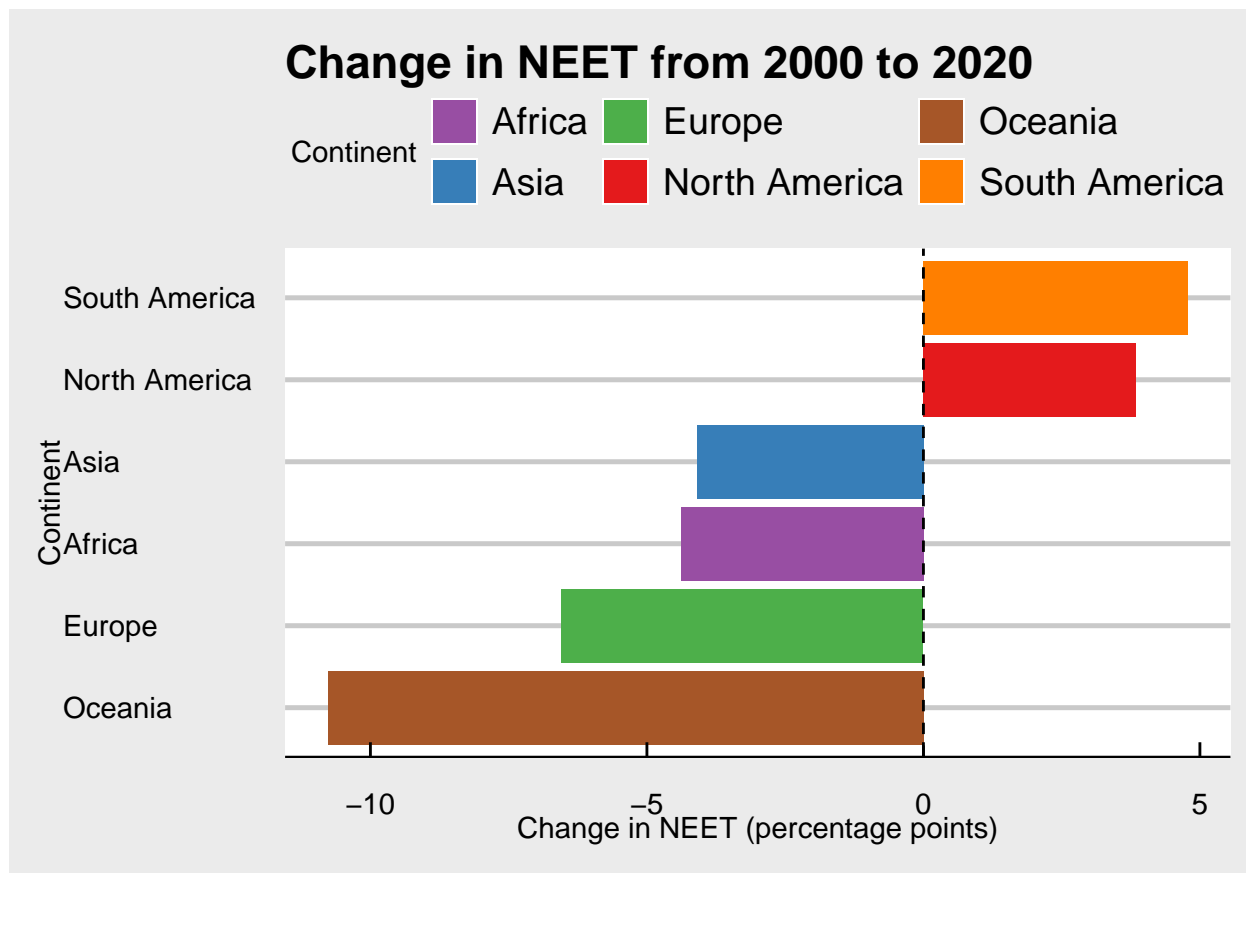
NEET values for 2000, 2015, 2020

```
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 - NEET change 2000 → 2020**

```
neet_change_2000 <- neet_2000 %>%
  inner_join(neet_2020, by = "Continent") %>%
  mutate(change_pp = neet_2020 - neet_2000)

ggplot(neet_change_2000,
       aes(x = reorder(Continent, change_pp), y = change_pp, fill = Continent)) +
  geom_col() + coord_flip() +
  theme_economist_white() +
  scale_fill_manual(values = continent_colors) +
  ggtitle("Change in NEET from 2000 to 2020") +
  xlab("Continent") + ylab("Change in NEET (percentage points)") +
  geom_hline(yintercept = 0, linetype = "dashed")
```

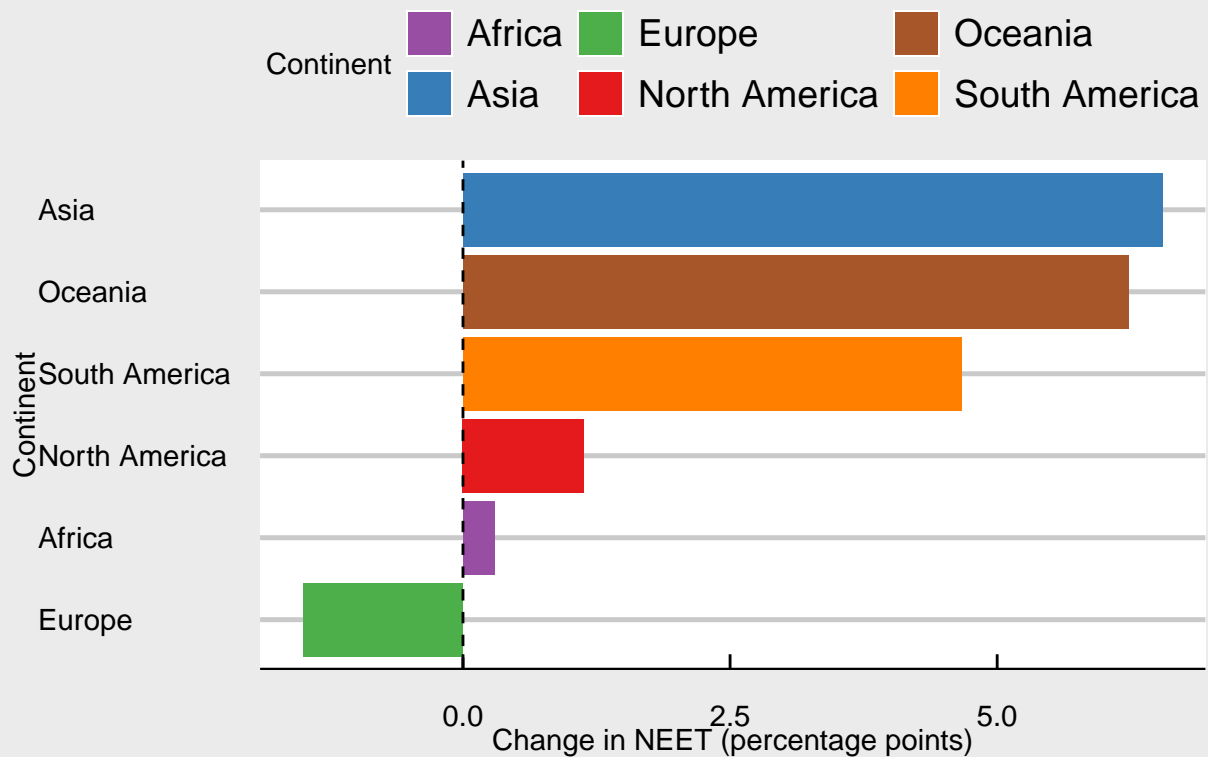**Change in NEET from 2000 to 2020**

Graph 3 - NEET change 2015 → 2020**

```
neet_change_2015 <- neet_2015 %>%
  inner_join(neet_2020, by = "Continent") %>%
  mutate(change_pp = neet_2020 - neet_2015)

ggplot(neet_change_2015,
       aes(x = reorder(Continent, change_pp), y = change_pp, fill = Continent)) +
  geom_col() + coord_flip() +
  theme_economist_white() +
  scale_fill_manual(values = continent_colors) +
  ggtitle("Change in NEET from 2015 to 2020") +
  xlab("Continent") + ylab("Change in NEET (percentage points)") +
  geom_hline(yintercept = 0, linetype = "dashed")
```

Change in NEET from 2015 to 2020

Combine unemployment + population, compute weighted averages

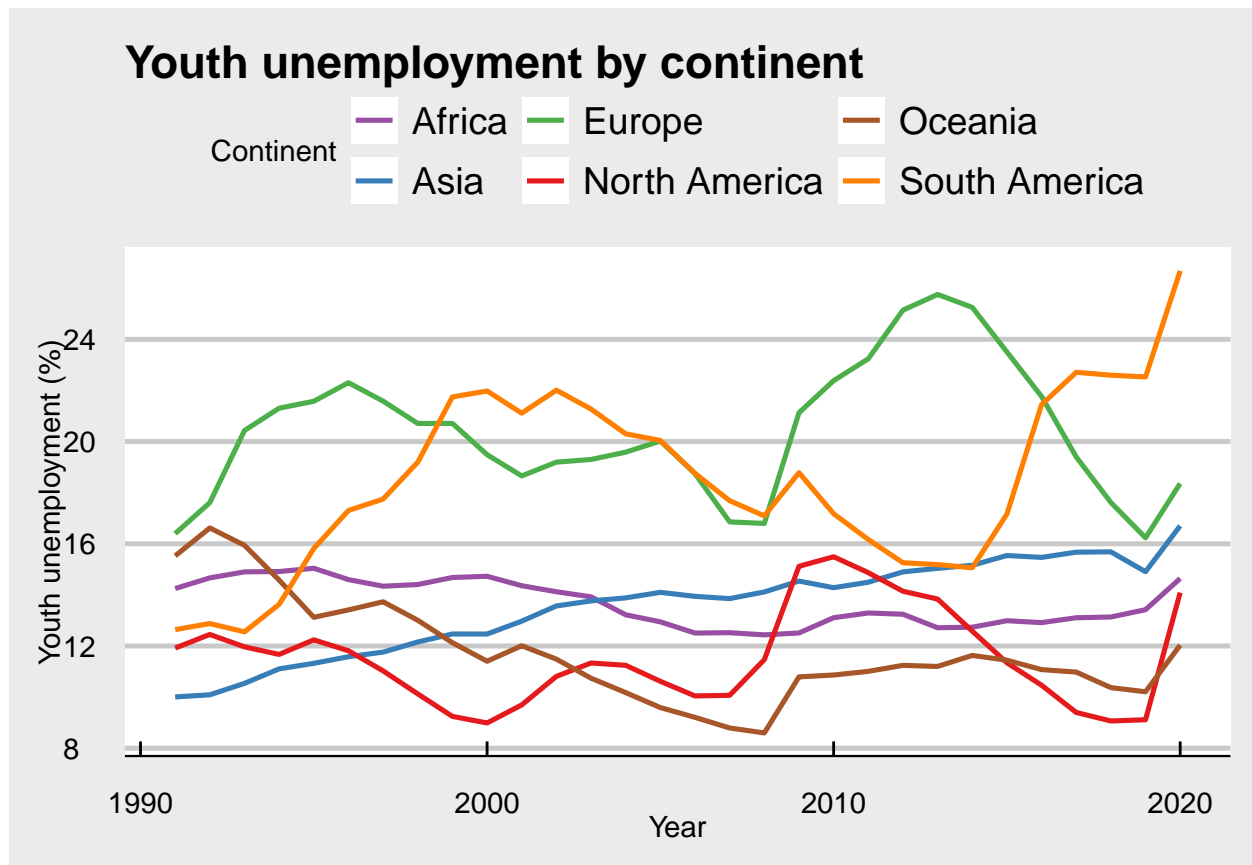```r
unemp_pop <- youth_unemp_continent %>%
  inner_join(population, by = c("Country_Code", "Year"))

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
)
```

---

Graph 4 - Youth unemployment trend (weighted)

```r
ggplot(unemp_continent_year,
       aes(x = Year, y = unemp_avg, colour = Continent)) +
```

```
geom_line(linewidth = 0.9) +
theme_economist_white() +
scale_color_manual(values = continent_colors) +
ggtitle("Youth unemployment by continent ") +
xlab("Year") +
ylab("Youth unemployment (%)")
```

Youth unemployment by continent

Youth unemployment values for 2000, 2015, 2020

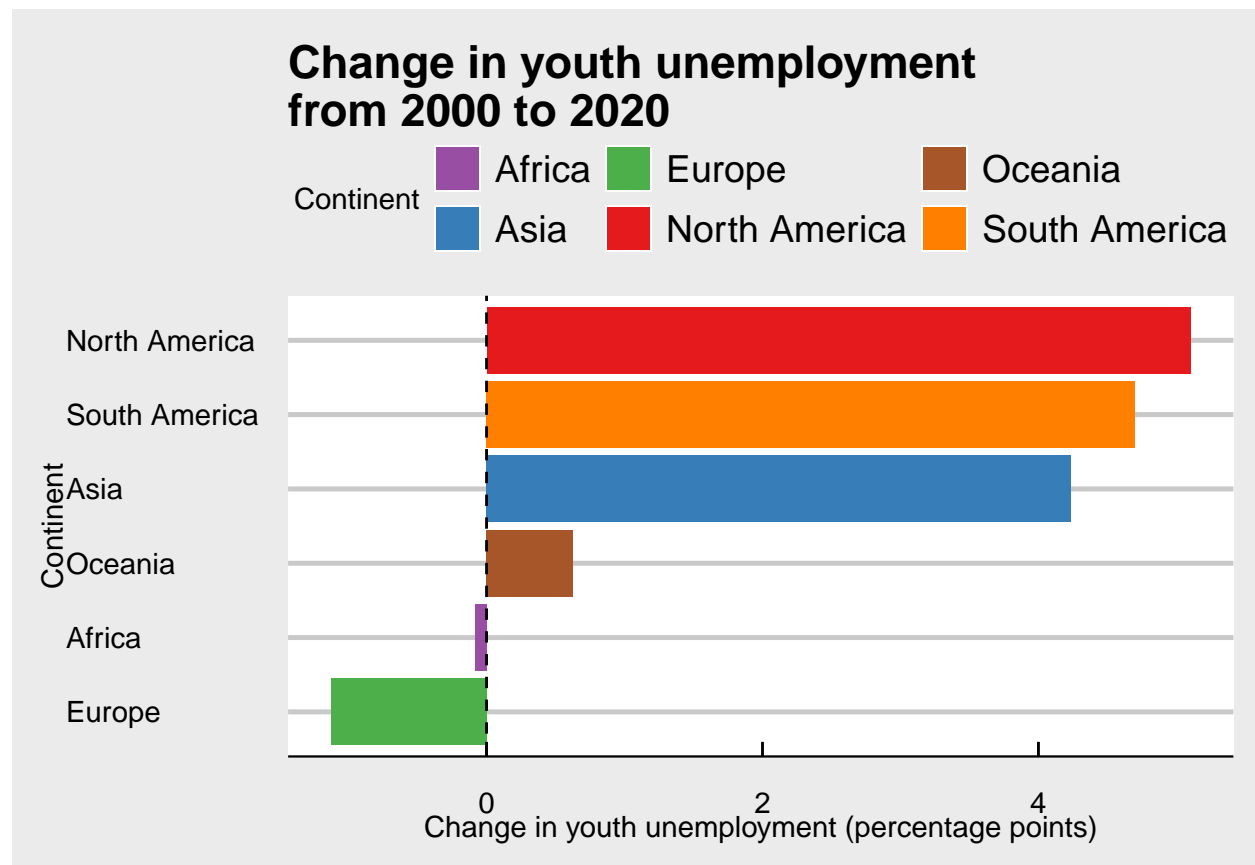```
unemp_2000 <- unemp_continent_year %>% filter(Year == 2000) %>% select(Continent, unemp_2000 = unemp_avg
unemp_2015 <- unemp_continent_year %>% filter(Year == 2015) %>% select(Continent, unemp_2015 = unemp_avg
unemp_2020 <- unemp_continent_year %>% filter(Year == 2020) %>% select(Continent, unemp_2020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

Graph 5: Change in youth unemployment 2000 → 2020

```
unemp_change_2000 <- unemp_2000 %>%
  inner_join(unemp_2020, by = "Continent") %>%
  mutate(change_pp = unemp_2020 - unemp_2000)
```

```
ggplot(unemp_change_2000,
       aes(x = reorder(Continent, change_pp), y = change_pp, fill = Continent)) +
geom_col() + coord_flip() +
theme_economist_white() +
scale_fill_manual(values = continent_colors) +
ggtitle("Change in youth unemployment\nfrom 2000 to 2020") +
xlab("Continent") + ylab("Change in youth unemployment (percentage points)") +
geom_hline(yintercept = 0, linetype = "dashed")
```


Graph 6: Change in youth unemployment 2015 → 2020

```
unemp_change_2015 <- unemp_2015 %>%
  inner_join(unemp_2020, by = "Continent") %>%
  mutate(change_pp = unemp_2020 - unemp_2015)

ggplot(unemp_change_2015,
       aes(x = reorder(Continent, change_pp), y = change_pp, fill = Continent)) +
geom_col() + coord_flip() +
theme_economist_white() +
scale_fill_manual(values = continent_colors) +
ggtitle("Change in youth unemployment\nfrom 2015 to 2020") +
xlab("Continent") + ylab("Change in youth unemployment (percentage points)") +
geom_hline(yintercept = 0, linetype = "dashed")
```

Change in youth unemployment from 2015 to 2020