

Exploring Youth NEET

2025-12-01

R Markdown

```
#Setting the working directory, loading all necessary libraries and data

library(sf)

## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE

library(rnaturalearth)
library(rnaturalearthdata)

## 
## Attaching package: 'rnaturalearthdata'

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

library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.6
## vforcats   1.0.1      v stringr   1.6.0
## v ggplot2   4.0.1      v tibble    3.3.0
## v lubridate 1.9.4      v tidyrr    1.3.1
## v purrr    1.2.0

## -- 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

setwd("~/Desktop/Uni/Intro to Data Science")

youth_neet <- read_csv("Group project/data sets/Youth_neet_final.csv")

## Rows: 1908 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (3): Entity, Code, Continent
```

```

## dbl (3): 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.

#checking the data has loaded
head(youth_neet)

## # A tibble: 6 x 6
##   Entity     Code  Year Share.of.youth.not.in.education.~1 pop_15_24 Continent
##   <chr>      <chr> <dbl>                               <dbl>      <dbl> <chr>
## 1 Afghanistan AFG    2014                         35.1    6914571 Asia
## 2 Afghanistan AFG    2017                         42.8    7704034 Asia
## 3 Afghanistan AFG    2020                         53.8    8444268 Asia
## 4 Afghanistan AFG    2021                         62.8    8621571 Asia
## 5 Albania       ALB    2002                         41.8    548992  Europe
## 6 Albania       ALB    2005                         35.2    546133  Europe
## # i abbreviated name:
## # 1: Share.of.youth.not.in.education..employment.or.training..total....of.youth.population.

#changing the column names for simplification
youth_neet <- youth_neet %>%
  rename(
    NEET = "Share.of.youth.not.in.education..employment.or.training..total....of.youth.population.",
    population = "pop_15_24"
  )
colnames(youth_neet)

## [1] "Entity"      "Code"        "Year"         "NEET"        "population"
## [6] "Continent"

#checking for missing values
sum(is.na(youth_neet))

## [1] 324

#checking where missing values occur
colSums(is.na(youth_neet))

##      Entity     Code      Year      NEET population Continent
##      0          108        0        0        108        108

#calculating average NEET rates by year and continent
avg_NEET_by_continent_year <- youth_neet %>%
  group_by(Continent, Year) %>%
  summarise(
    avg_NEET = mean(NEET,
                    na.rm = TRUE)
  )

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

```

#confirming the output
head(avg_NEET_by_continent_year)

## # A tibble: 6 x 3
## # Groups:   Continent [1]
##   Continent Year avg_NEET
##   <chr>     <dbl>    <dbl>
## 1 Africa      1991    21.3
## 2 Africa      1994    25.1
## 3 Africa      1996    43.4
## 4 Africa      1999    32.4
## 5 Africa      2000    31.1
## 6 Africa      2001    26.9

#exploring the NEET rates by calculating summary data
continent_summary <- youth_neet %>%
  filter(!is.na(Continent), !is.na(NEET)) %>%
  group_by(Continent) %>%
  summarise(
    mean_neet      = mean(NEET, na.rm = TRUE),
    weighted_mean = sum(NEET * population, na.rm = TRUE) /
      sum(population, na.rm = TRUE),
    median_neet   = median(NEET, na.rm = TRUE),
    min_neet       = min(NEET, na.rm = TRUE),
    max_neet       = max(NEET, na.rm = TRUE),
    sd_neet        = sd(NEET, na.rm = TRUE),
    iqr_neet       = IQR(NEET, na.rm = TRUE)
  )
continent_summary

## # A tibble: 6 x 8
##   Continent   mean_neet weighted_mean median_neet min_neet max_neet sd_neet
##   <chr>        <dbl>        <dbl>        <dbl>      <dbl>      <dbl>    <dbl>
## 1 Africa        25.8        26.1        26.9      3.79      68.7     9.04
## 2 Asia          21.4        27.3        21.4      2.97      62.8    11.0
## 3 Europe         13.1        12.6        11.5      0.38      41.8    6.73
## 4 North America  20.6        15.7        19.9      5.06      52.0    6.50
## 5 Oceania        21.5        12.1        14.0      7.04      52.0    12.7
## 6 South America  19.6        20.6        19.2      9.18      46.4    4.83
## # i 1 more variable: iqr_neet <dbl>

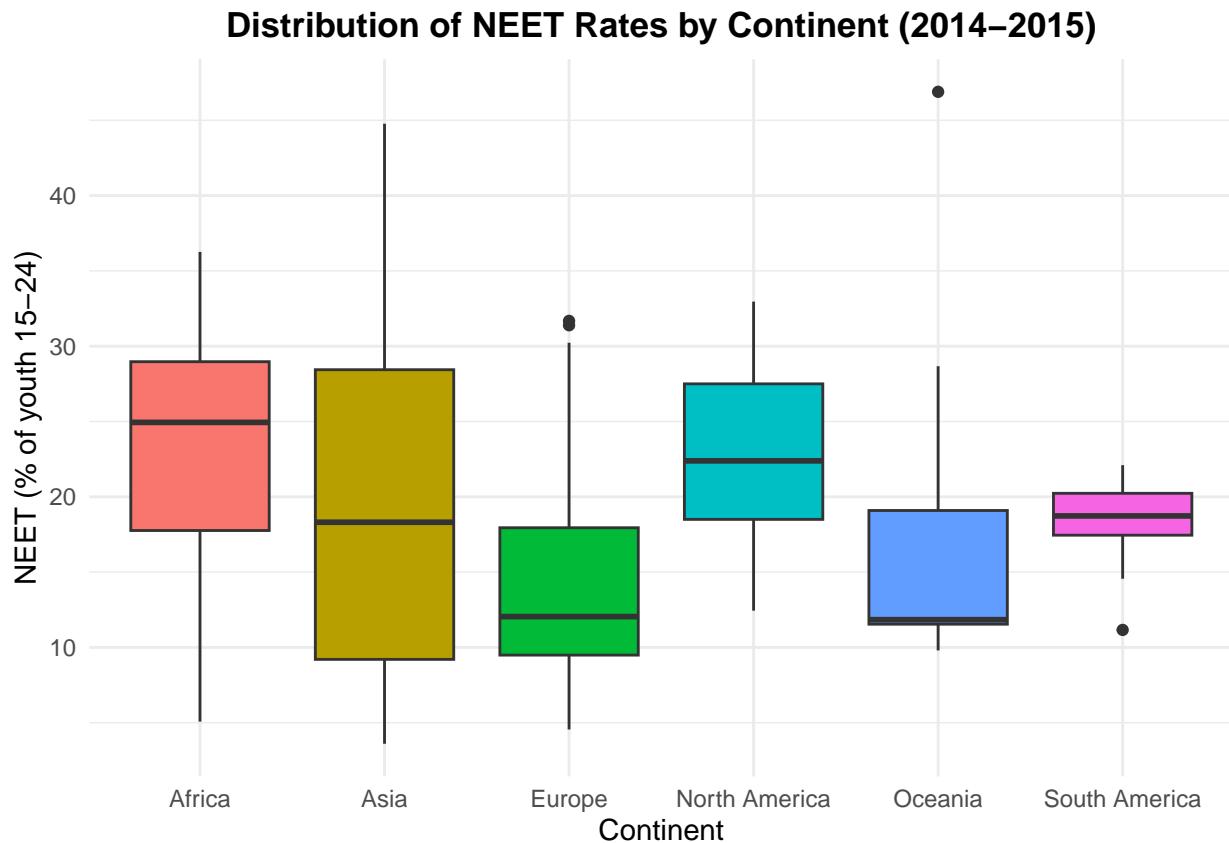
#Boxplot: NEET by continent for 2014 & 2015
ggplot(
  youth_neet %>%
    filter(!is.na(Continent), Year %in% c(2014, 2015)),
  aes(x = Continent, y = NEET, fill = Continent)
) +
  geom_boxplot() +
  labs(
    title = "Distribution of NEET Rates by Continent (2014–2015)",
    x = "Continent",
    y = "NEET (% of youth 15–24)"
)

```

```

) +
theme_minimal() +
theme(
  legend.position = "none",
  plot.title = element_text(hjust = 0.5, face = "bold")
)

```

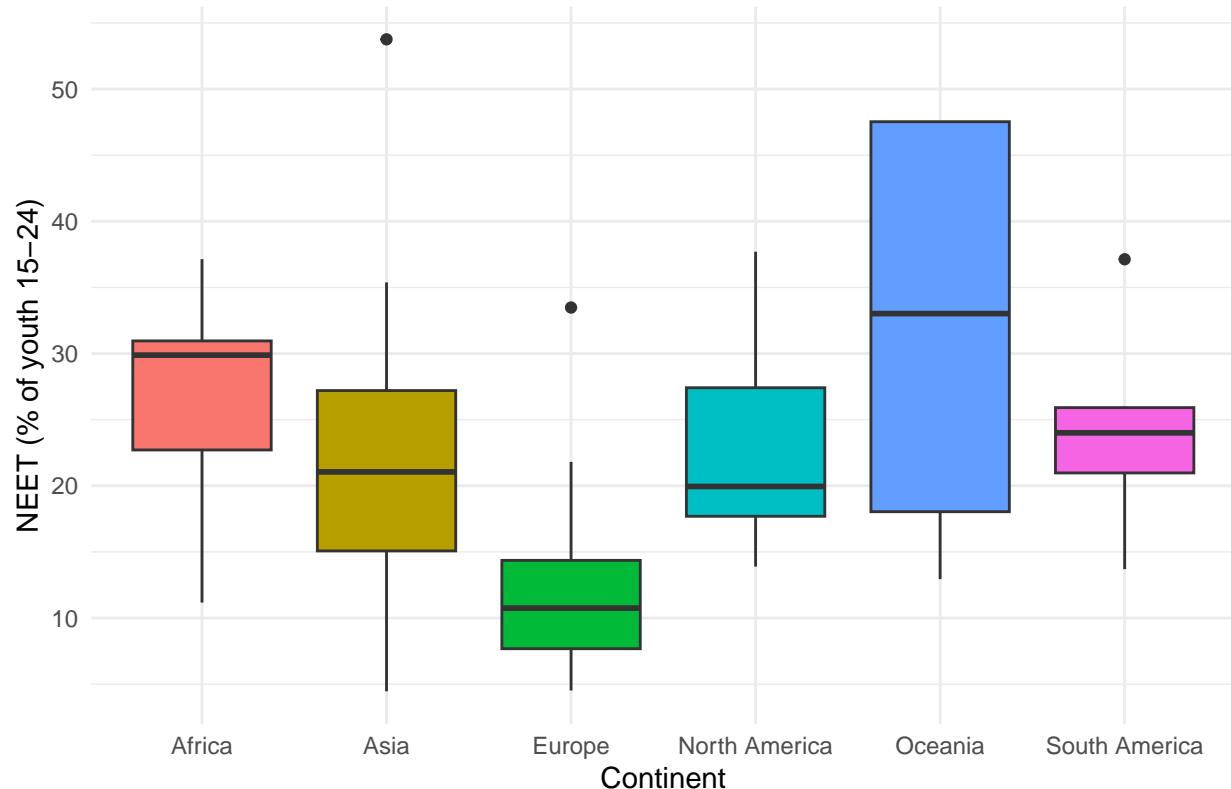


```

#Boxplot: NEET by continent for 2020
ggplot(
  youth_neet %>%
    filter(!is.na(Continent), Year %in% c(2020)),
  aes(x = Continent, y = NEET, fill = Continent)
) +
  geom_boxplot() +
  labs(
    title = "Distribution of NEET Rates by Continent (2020)",
    x = "Continent",
    y = "NEET (% of youth 15-24)"
  ) +
  theme_minimal() +
  theme(
    legend.position = "none",
    plot.title = element_text(hjust = 0.5, face = "bold")
)

```

Distribution of NEET Rates by Continent (2020)



```
#calculating population weighted NEET for each continent by year
neet_weighted <- youth_neet %>%
  group_by(Continent, Year) %>%
  summarise(
    weighted_NEET = weighted.mean(NEET, population, na.rm = TRUE),
    total_population = sum(population, na.rm = TRUE),
  )
```

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

```
#checking the output
head(neet_weighted)
```

```
## # A tibble: 6 x 4
## # Groups:   Continent [1]
##   Continent Year weighted_NEET total_population
##   <chr>     <dbl>        <dbl>            <dbl>
## 1 Africa     1991        21.3            3173483
## 2 Africa     1994        25.1            326596
## 3 Africa     1996        43.4            330806
## 4 Africa     1999        32.4            6357672
## 5 Africa     2000        31.4            13976220
## 6 Africa     2001        25.4            17649356
```

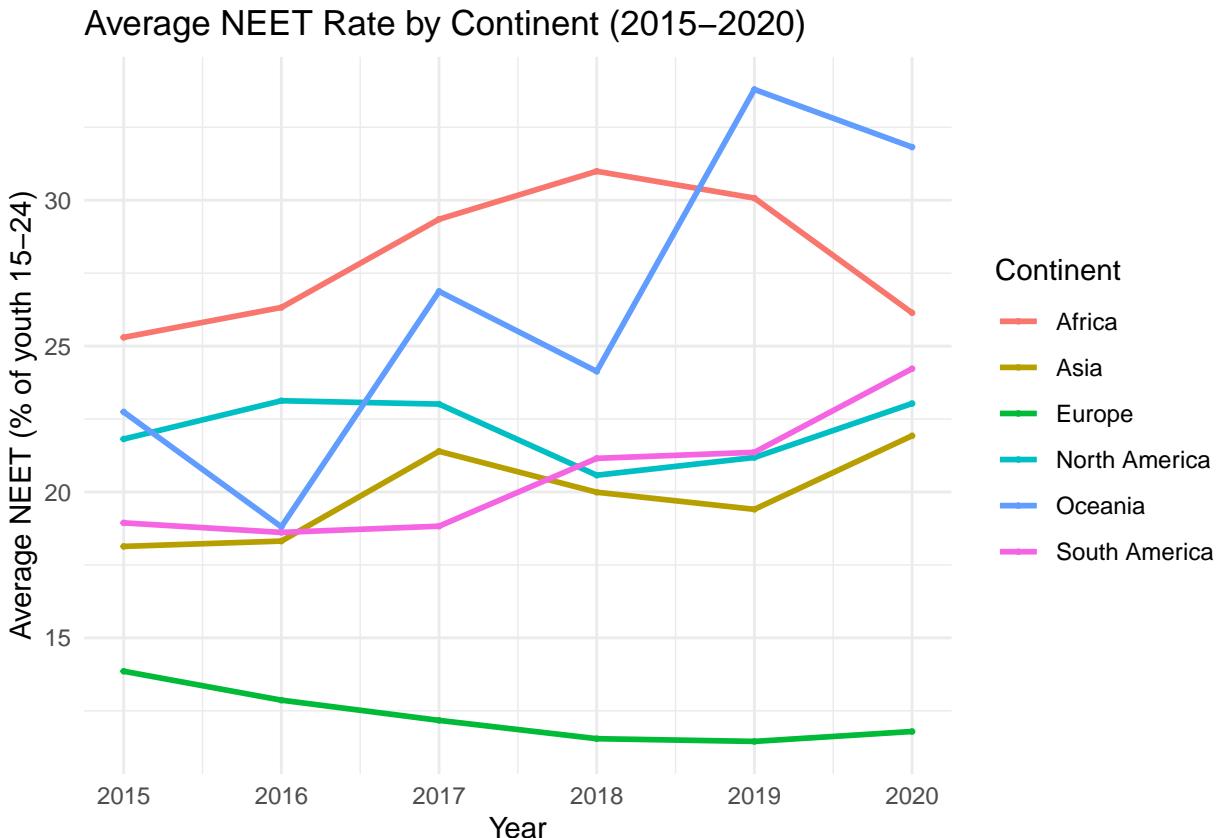
```

#line graph of each continent between 2015 and 2020 showing the average NEET
neet_plot_data <- avg_NEET_by_continent_year %>%
  filter(Year >= 2015, Year <= 2020) %>%
  filter(!is.na(Continent))    # <-- Remove NA continent

ggplot(neet_plot_data, aes(x = Year, y = avg_NEET, color = Continent)) +
  geom_line(size = 1) +
  geom_point(size = 0.5) +
  labs(
    title = "Average NEET Rate by Continent (2015–2020)",
    x = "Year",
    y = "Average NEET (% of youth 15–24)",
    color = "Continent"
  ) +
  theme_minimal()

## 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.

```



```

#line graph of each continent between 2015 and 2020 showing the average NEET weighted by the population o...
ggplot(neet_weighted %>% filter(Year >= 2015, Year <= 2020) %>%
  filter(!is.na(Continent)) ,

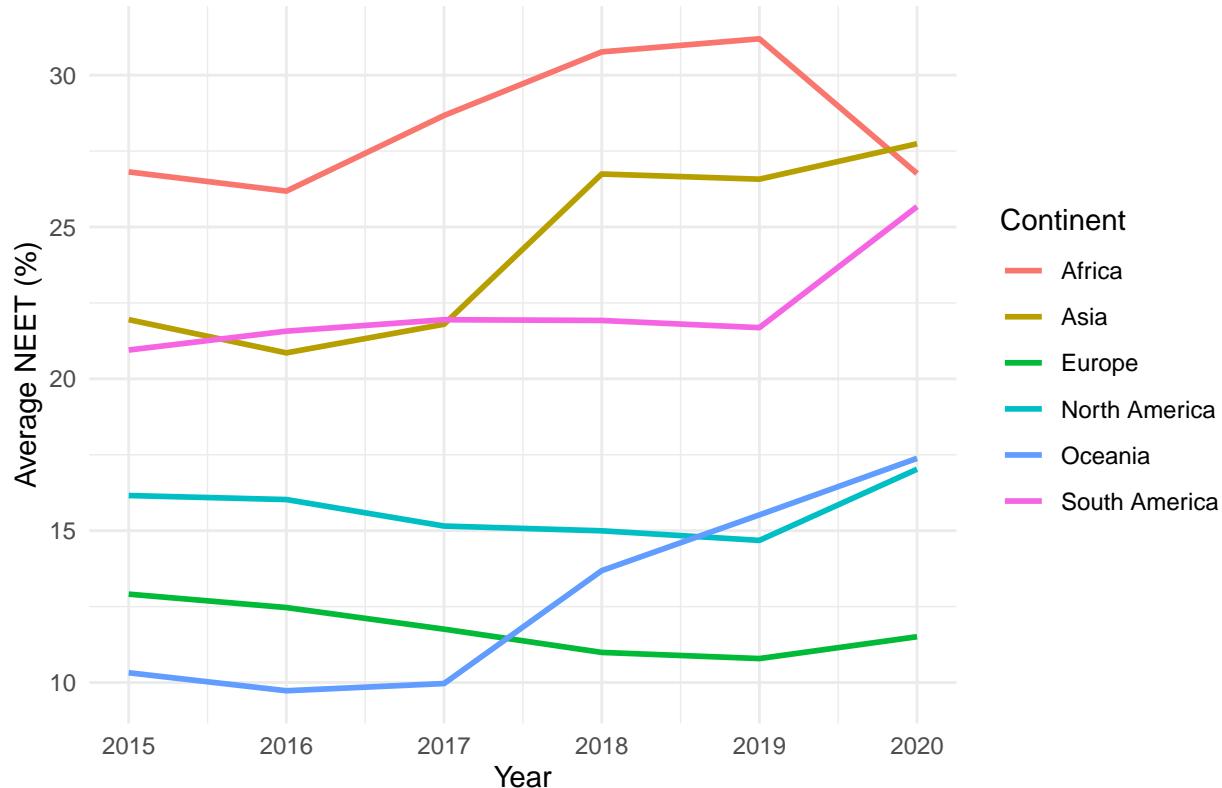
```

```

aes(x = Year, y = weighted_NEET, color = Continent)) +
geom_line(size = 1) +
labs(
  title = "Weighted Average NEET Rate Over Time by Continent (2015–2020)",
  x = "Year",
  y = "Average NEET (%)"
) +
theme_minimal()

```

Weighted Average NEET Rate Over Time by Continent (2015–2020)



```

#calculating the percentage point change between 2010 to 2015 average and the 2020 average
neet_change_continent <- avg_NEET_by_continent_year %>%
  filter(!is.na(Continent),
         Year >= 2010, Year <= 2020) %>% # keep only 2010–2020
  group_by(Continent) %>%
  summarise(
    avg_2010_2015 = mean(avg_NEET[Year >= 2010 & Year <= 2015], na.rm = TRUE),
    neet_2020 = avg_NEET[Year == 2020][1], # NEET value in 2020
    .groups = "drop"
  ) %>%
  mutate(
    percentage_point_change = neet_2020 - avg_2010_2015
  )

#checking the output
head(neet_change_continent)

```

```

## # A tibble: 6 x 4
##   Continent      avg_2010_2015 neet_2020 percentage_point_change
##   <chr>          <dbl>       <dbl>                  <dbl>
## 1 Africa           22.4        26.1                  3.73
## 2 Asia              19.2        21.9                  2.72
## 3 Europe            14.2        11.8                 -2.39
## 4 North America     22.1        23.0                  0.939
## 5 Oceania           18.9        31.8                  12.9
## 6 South America     18.1        24.2                  6.17

#load world map as an simple features object
world <- ne_countries(scale = "medium", returnclass = "sf") %>%
  select(name, continent, geometry)

#keep only the 6 continents
world <- world %>%
  filter(continent %in% c("Africa", "Asia", "Europe",
                         "North America", "South America", "Oceania"))

#join NEET % difference data to every country by continent
world_neet <- world %>%
  left_join(neet_change_continent,
            by = c("continent" = "Continent"))

#plotting the percentage point difference by continent on the world map
ggplot(world_neet) +
  geom_sf(aes(fill = percentage_point_change), color = "grey30", size = 0.1) +
  scale_fill_gradient2(
    low = "lightgreen",
    mid = "lightyellow",
    high = "red",
    midpoint = 0,
    name = "Percentage \nPoint Difference"
  ) +
  labs(
    title = "Percentage Point Difference in Average NEET Rates (2015-2020)",
    subtitle = "Colour shows change in continental averages"
  ) +
  theme_minimal() +
  theme(
    axis.text = element_blank(),
    panel.grid = element_blank()
  )

```

Percentage Point Difference in Average NEET Rates (2015–2020)

Colour shows change in continental averages

