# Goal2

2025-12-01

## R Markdown

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     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/Group project")
youth_neet <- read_csv("data sets/Youth_not_in_education_employment_training_with_continent.csv")
```

```
## Rows: 1908 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): Entity, Code, Continent
## 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.
```

```
head(youth_neet)
```

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

```r
avg_NEET_by_continent_year <- youth_neet %>%
  group_by(Continent, Year) %>%
  summarise(
    avg_NEET = mean(Share.of.youth.not.in.education..employment.or.training..total....of.youth.populati
                    na.rm = TRUE)
  )
```

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

```r
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
```

```r
youth_neet2 <- read_csv("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.
```

```r
head(youth_neet2)
```

```
## # 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.
```

```r
colnames(youth_neet2)
```

```
## [1] "Entity"
## [2] "Code"
## [3] "Year"
## [4] "Share.of.youth.not.in.education..employment.or.training..total....of.youth.population."
## [5] "pop_15_24"
## [6] "Continent"
```

```
youth_neet2 <- youth_neet2 %>%
  rename(
    NEET = "Share.of.youth.not.in.education..employment.or.training..total....of.youth.population.",
    population = "pop_15_24"
  )
```

```
neet_weighted <- youth_neet2 %>%
  group_by(Continent, Year) %>%
  summarise(
    weighted_NEET = weighted.mean(NEET, population, na.rm = TRUE),
    total_population = sum(population, na.rm = TRUE),
    n_countries = sum(!is.na(NEET) & !is.na(population)),
    .groups = "drop"
  )

neet_weighted
```

```
## # A tibble: 217 x 5
##    Continent  Year weighted_NEET total_population n_countries
##    <chr>     <dbl>         <dbl>            <dbl>       <int>
##  1 Africa     1991          21.3          3173483           1
##  2 Africa     1994          25.1           326596           1
##  3 Africa     1996          43.4           330806           1
##  4 Africa     1999          32.4          6357672           1
##  5 Africa     2000          31.4         13976220           2
##  6 Africa     2001          25.4         17649356           3
##  7 Africa     2002          34.1         12313110           3
##  8 Africa     2003          35.1         11960519           3
##  9 Africa     2004          34.3         14836139           4
## 10 Africa     2005          21.9         41364523           7
## # i 207 more rows
```

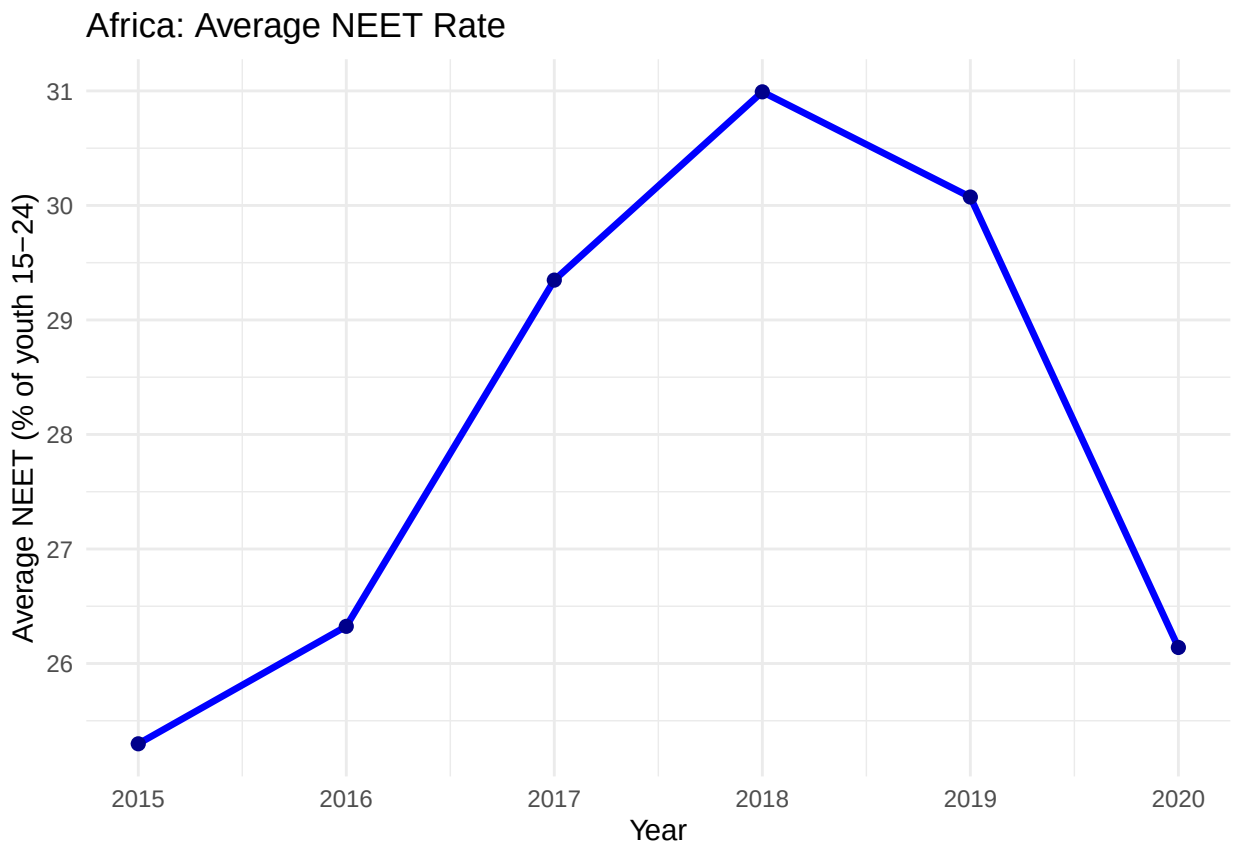
```
colnames(youth_neet2)
```

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

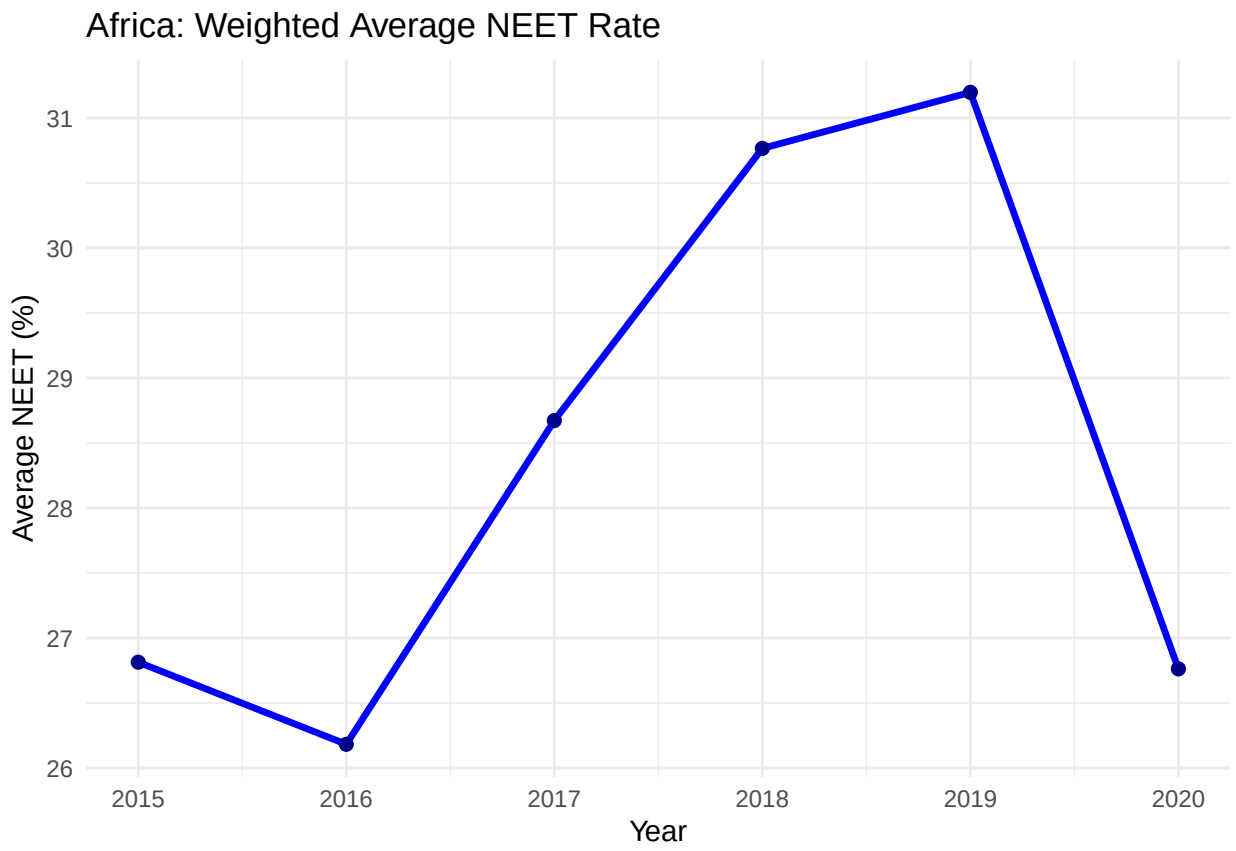
```
# Filter Africa's data
africa_neet <- avg_NEET_by_continent_year %>%
  filter(Continent == "Africa", Year >= 2015, Year <= 2020) %>%
  group_by(Year)

# Line plot
ggplot(africa_neet, aes(x = Year, y = avg_NEET)) +
  geom_line(color = "blue", size = 1.2) +
```

```
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Africa: Average NEET Rate",
    x = "Year",
    y = "Average NEET (% of youth 15-24)"
  ) +
  theme_minimal()
```

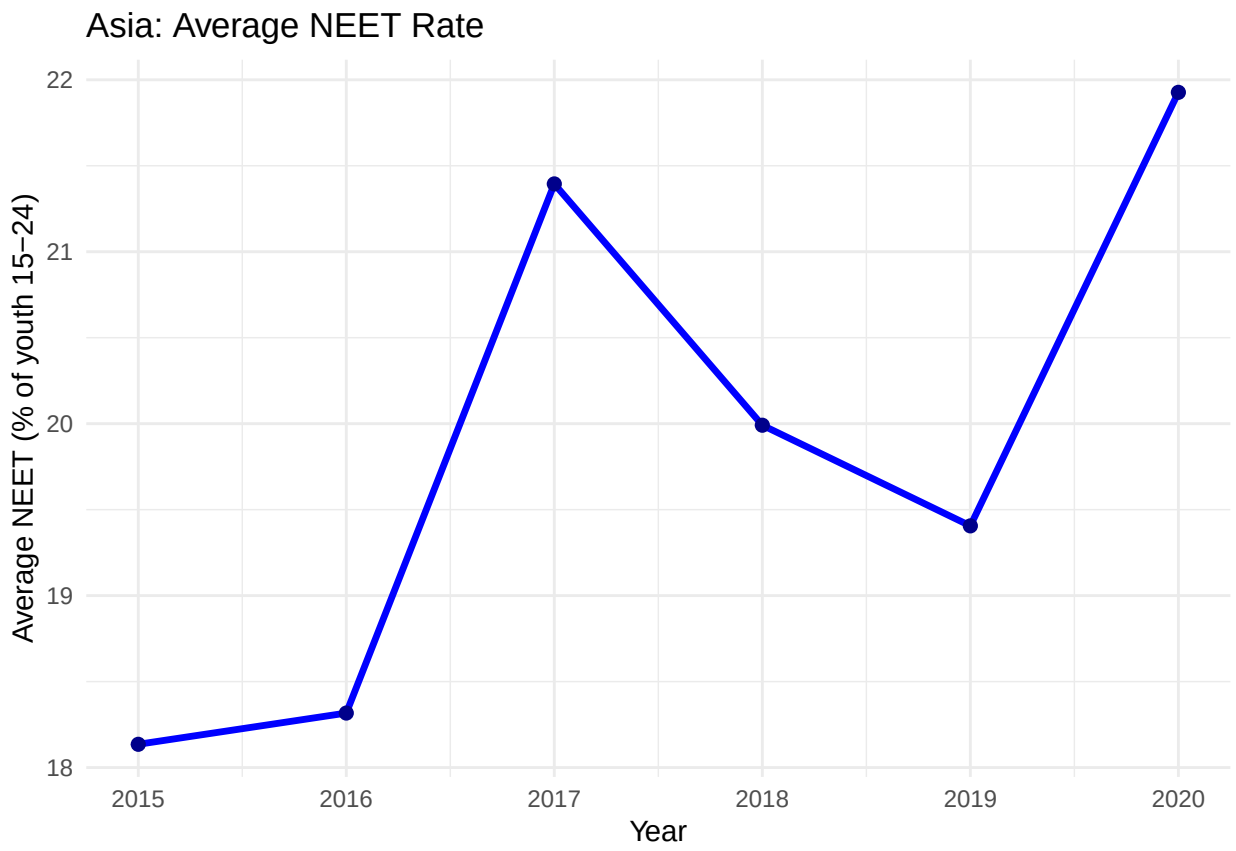
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



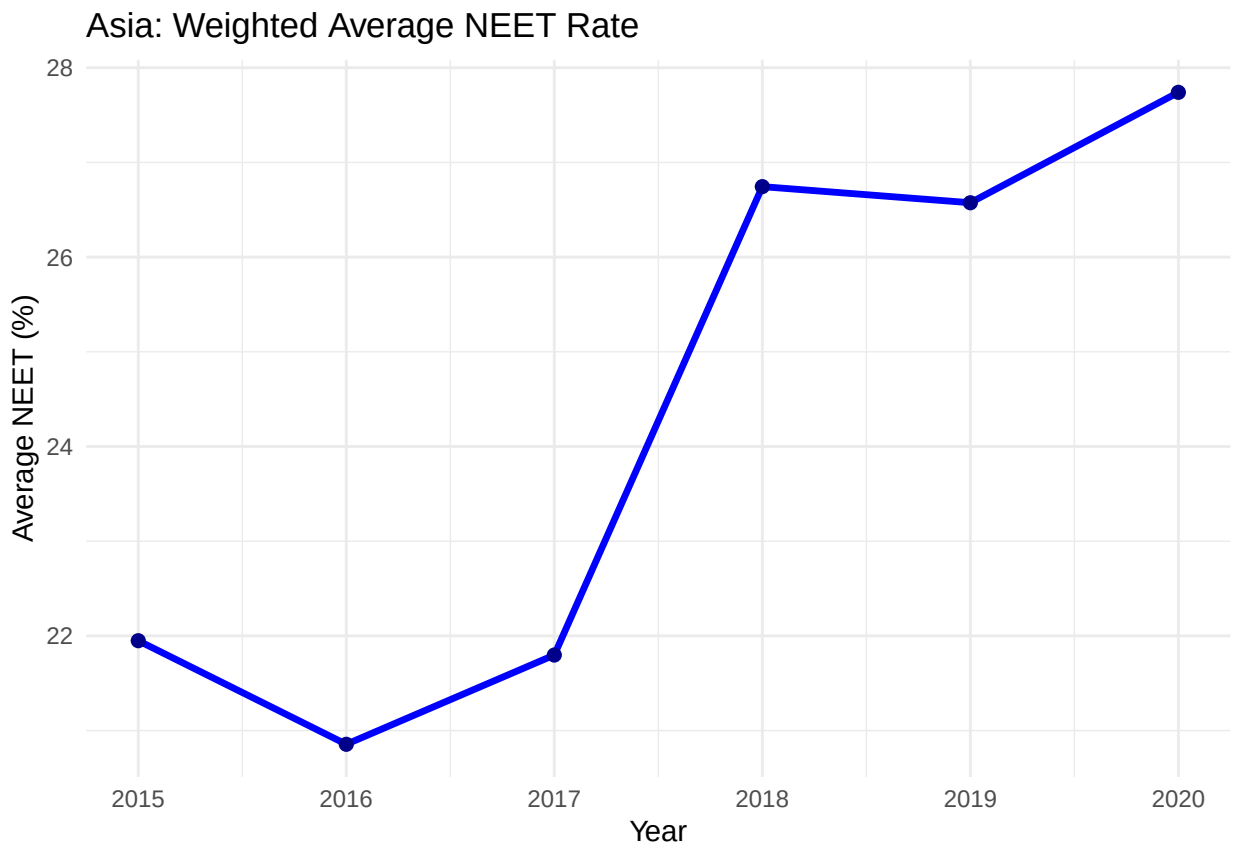

```
ggplot(neet_weighted %>% filter(Continent == "Africa",Year >= 2015, Year <= 2020),
       aes(x = Year, y = weighted_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Africa: Weighted Average NEET Rate",
    x = "Year",
    y = "Average NEET (%)"
  ) +
  theme_minimal()
```



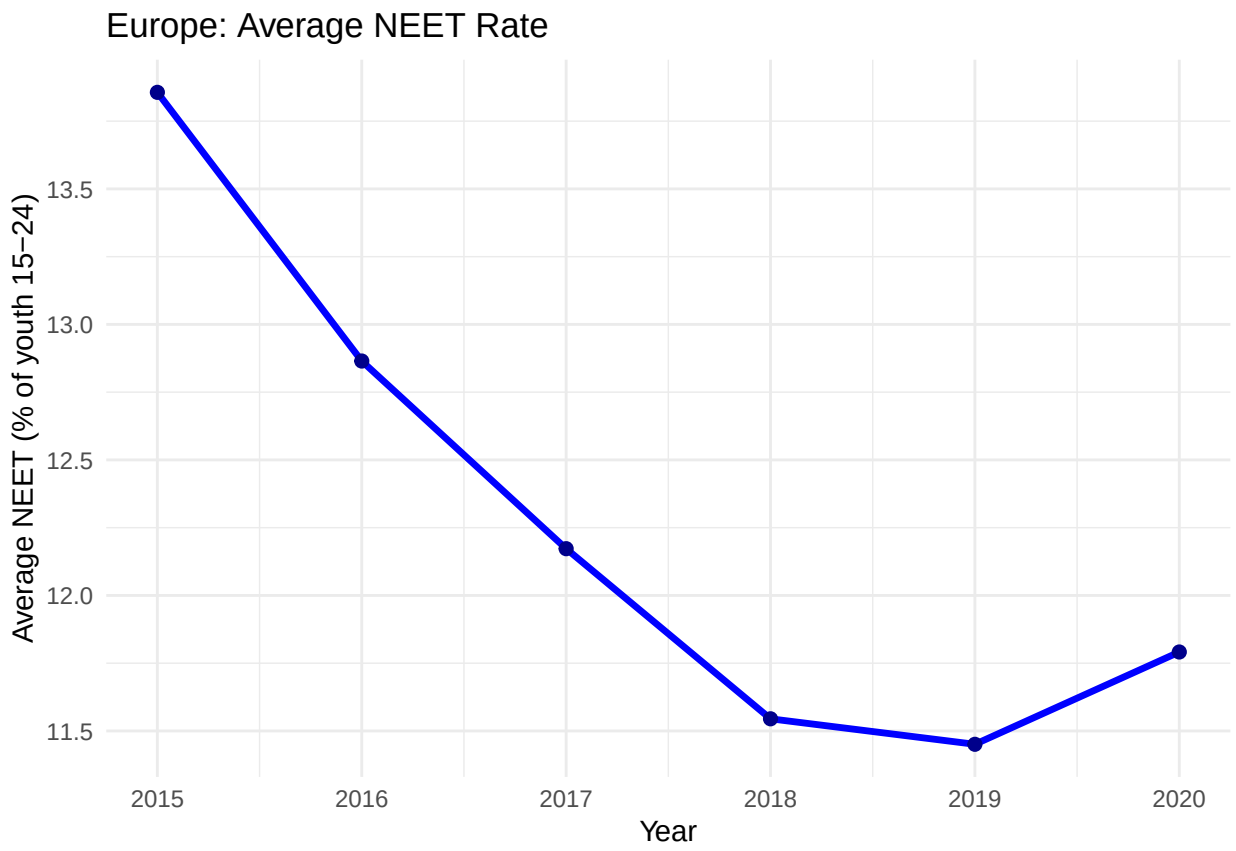

```r
asia_neet <- avg_NEET_by_continent_year %>%
  filter(Continent == "Asia", Year >= 2015, Year <= 2020) %>%
  group_by(Year)

# Line plot
ggplot(asia_neet, aes(x = Year, y = avg_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Asia: Average NEET Rate",
    x = "Year",
    y = "Average NEET (% of youth 15-24)"
  ) +
  theme_minimal()
```



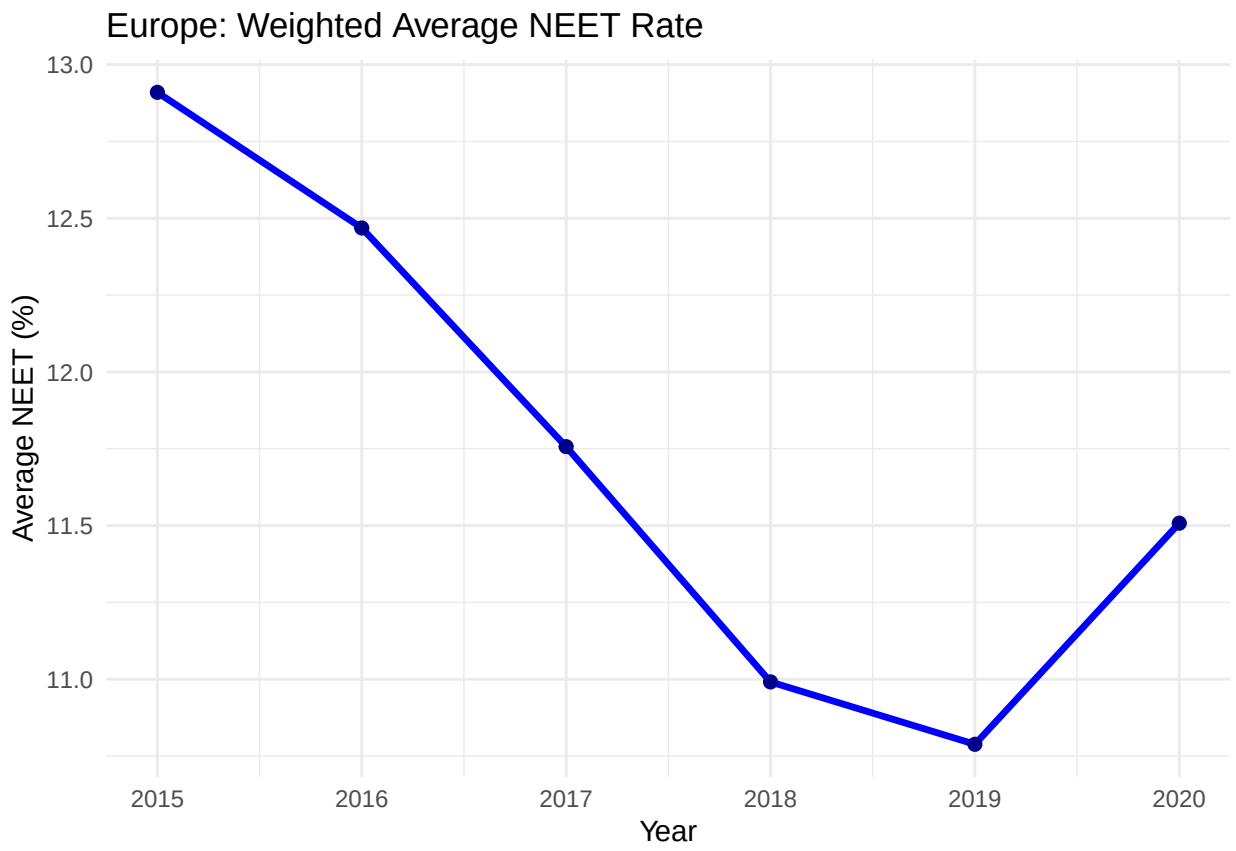

```
ggplot(neet_weighted %>% filter(Continent == "Asia",Year >= 2015, Year <= 2020),
       aes(x = Year, y = weighted_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Asia: Weighted Average NEET Rate",
    x = "Year",
    y = "Average NEET (%)"
  ) +
  theme_minimal()
```



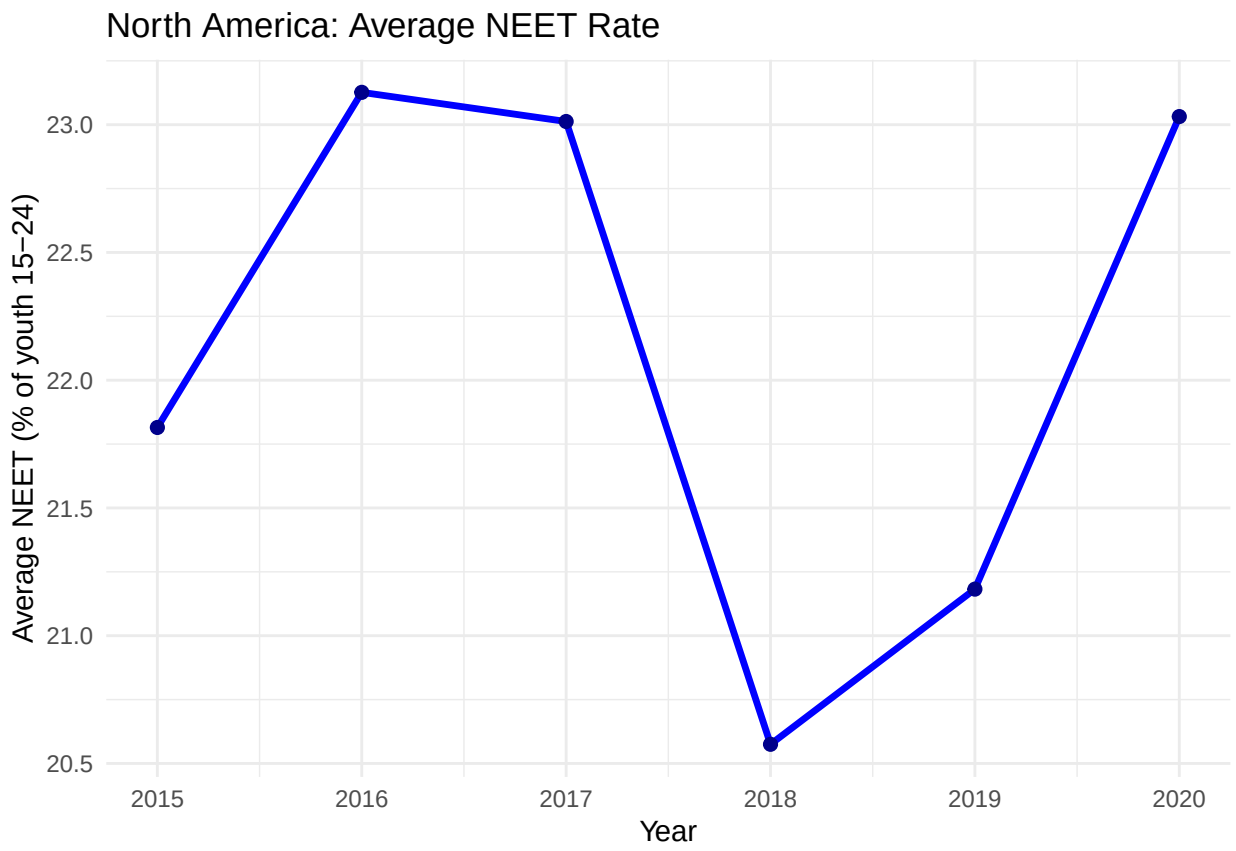

```r
# Filter Europe's data
europe_neet <- avg_NEET_by_continent_year %>%
  filter(Continent == "Europe", Year >= 2015, Year <= 2020) %>%
  group_by(Year)

# Line plot
ggplot(europe_neet, aes(x = Year, y = avg_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Europe: Average NEET Rate",
    x = "Year",
    y = "Average NEET (% of youth 15-24)"
  ) +
  theme_minimal()
```



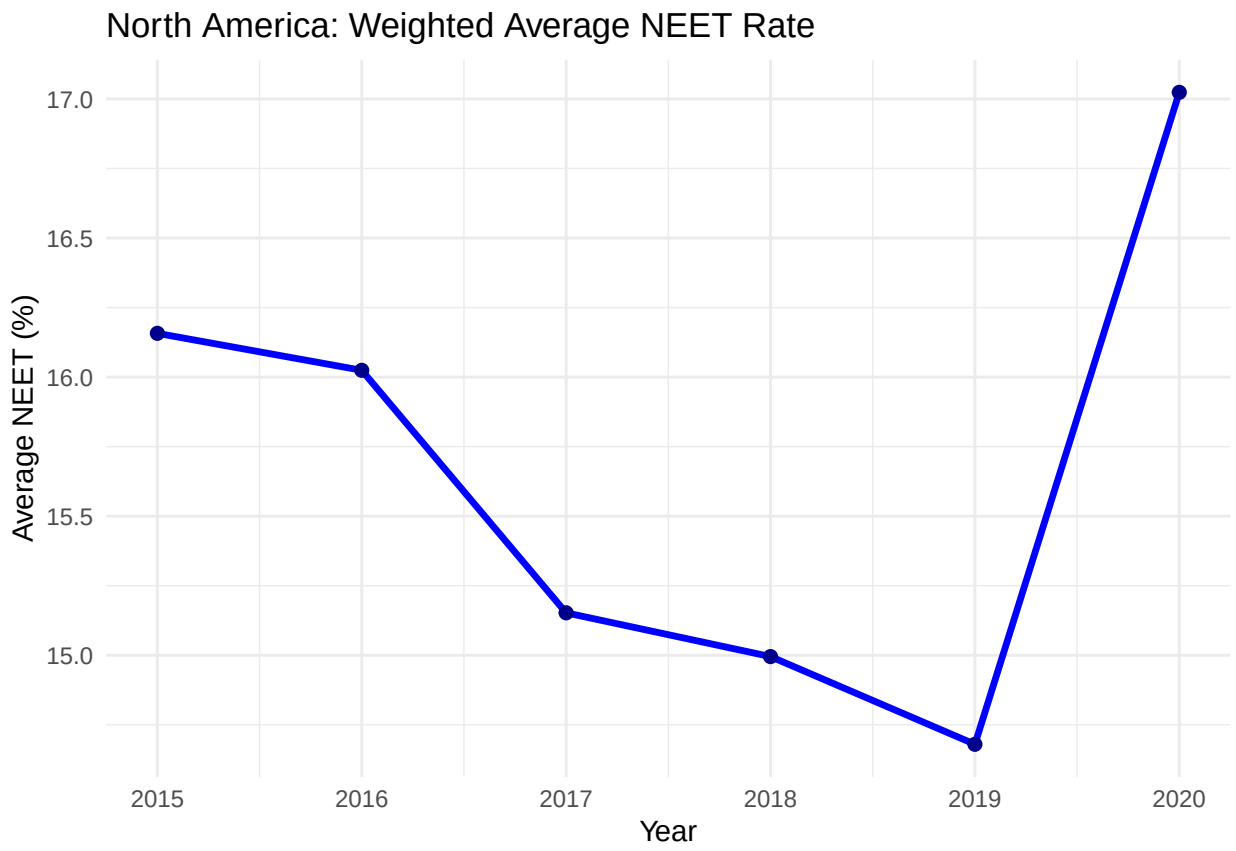

```
ggplot(neet_weighted %>% filter(Continent == "Europe",Year >= 2015, Year <= 2020),
       aes(x = Year, y = weighted_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Europe: Weighted Average NEET Rate",
    x = "Year",
    y = "Average NEET (%)"
  ) +
  theme_minimal()
```



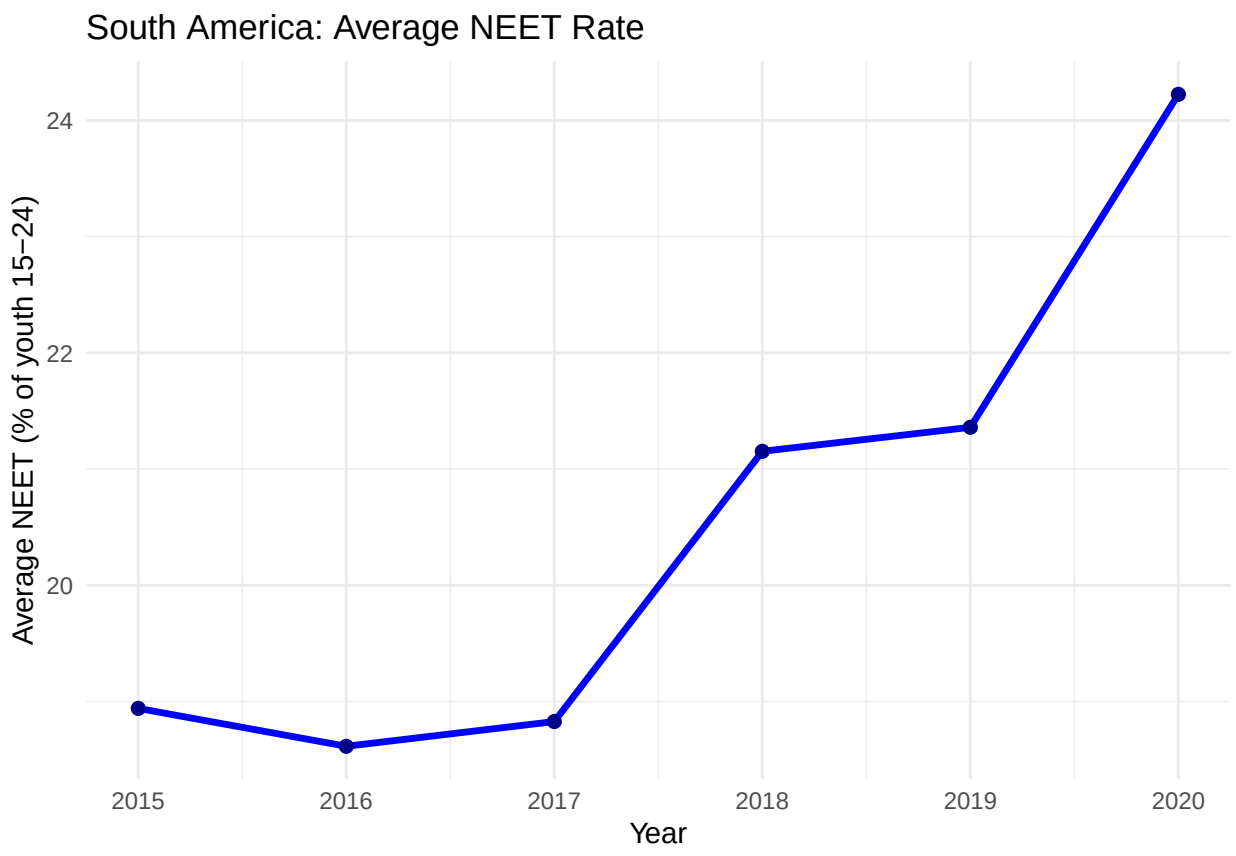

```r
# Filter North America's data
north_america_neet <- avg_NEET_by_continent_year %>%
  filter(Continent == "North America", Year >= 2015, Year <= 2020) %>%
  group_by(Year)

# Line plot
ggplot(north_america_neet, aes(x = Year, y = avg_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "North America: Average NEET Rate",
    x = "Year",
    y = "Average NEET (% of youth 15-24)"
  ) +
  theme_minimal()
```



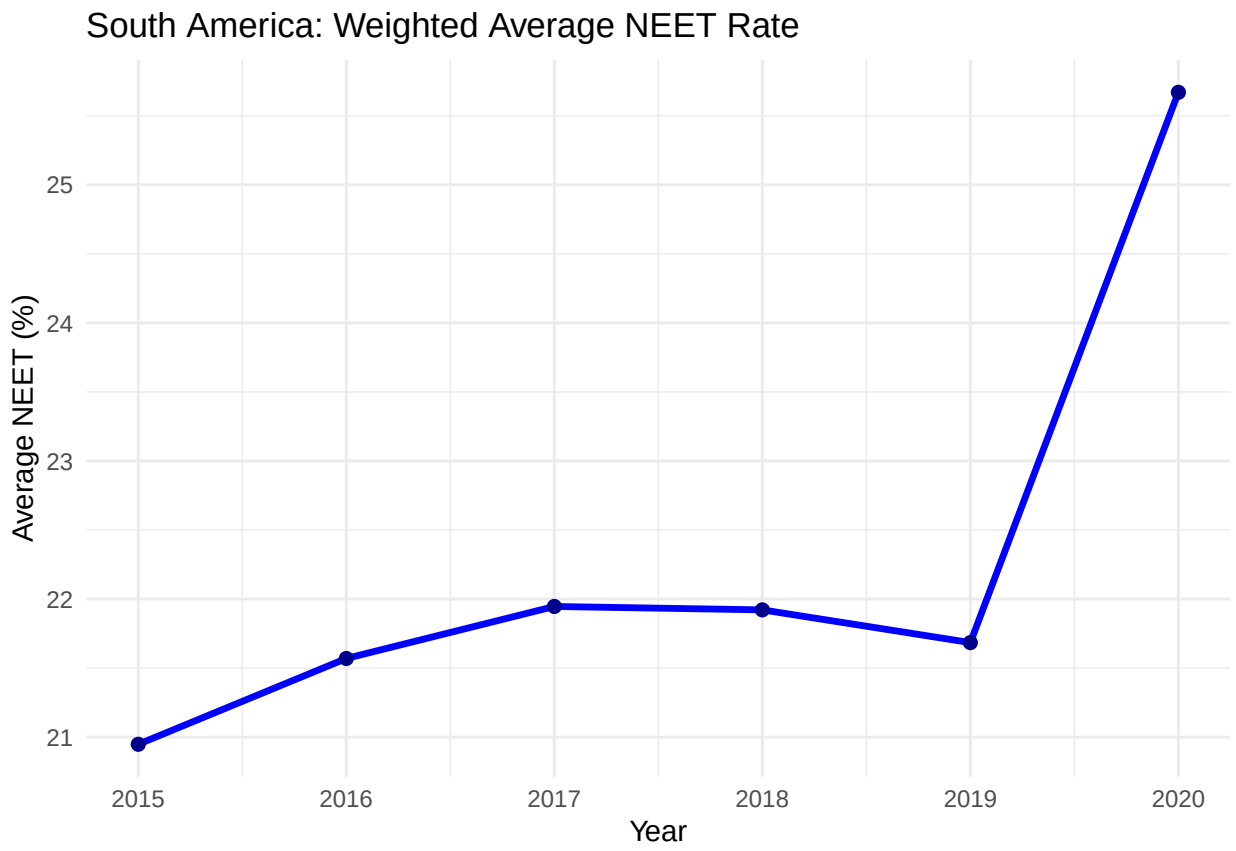

```
ggplot(neet_weighted %>% filter(Continent == "North America",Year >= 2015, Year <= 2020),
       aes(x = Year, y = weighted_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "North America: Weighted Average NEET Rate",
    x = "Year",
    y = "Average NEET (%)"
  ) +
  theme_minimal()
```



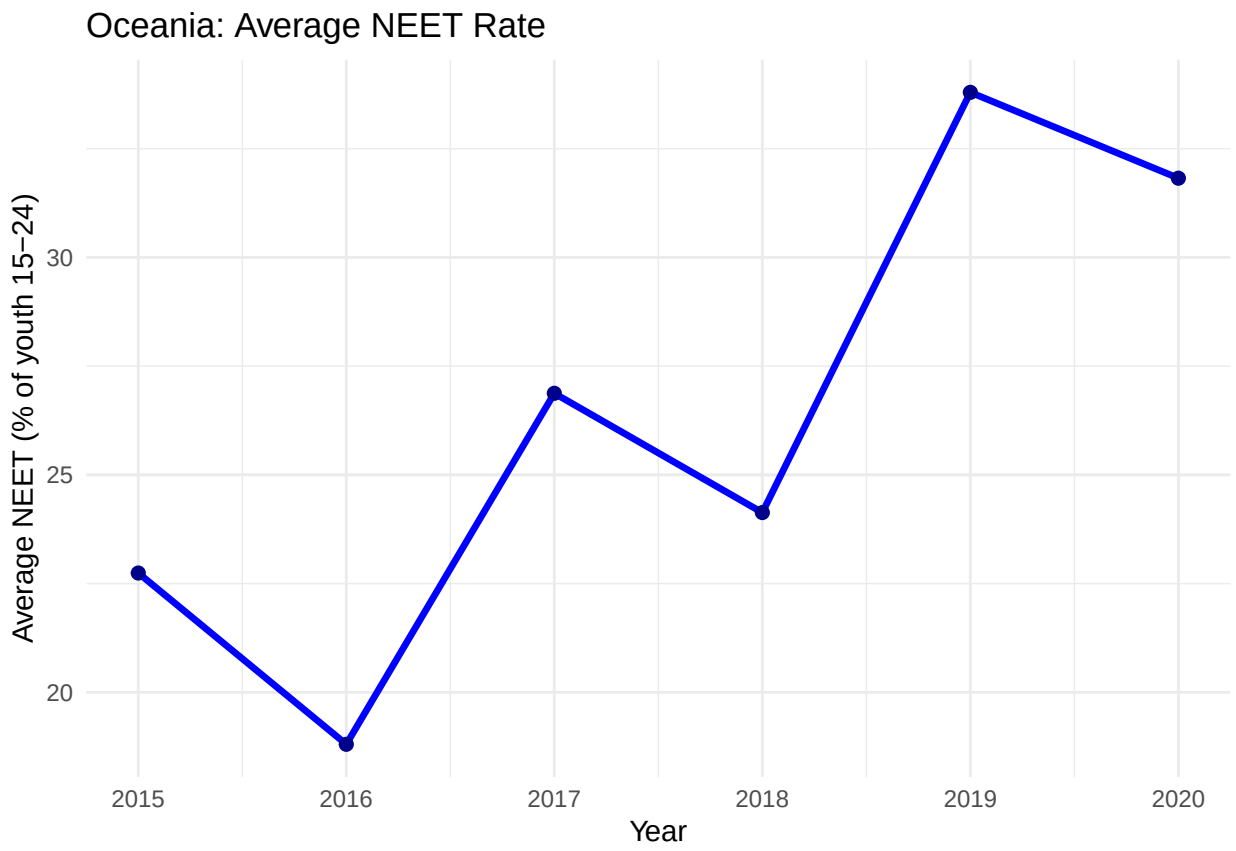

```r
# Filter South America's data
south_america_neet <- avg_NEET_by_continent_year %>%
  filter(Continent == "South America", Year >= 2015, Year <= 2020) %>%
  group_by(Year)

# Line plot
ggplot(south_america_neet, aes(x = Year, y = avg_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "South America: Average NEET Rate",
    x = "Year",
    y = "Average NEET (% of youth 15-24)"
  ) +
  theme_minimal()
```



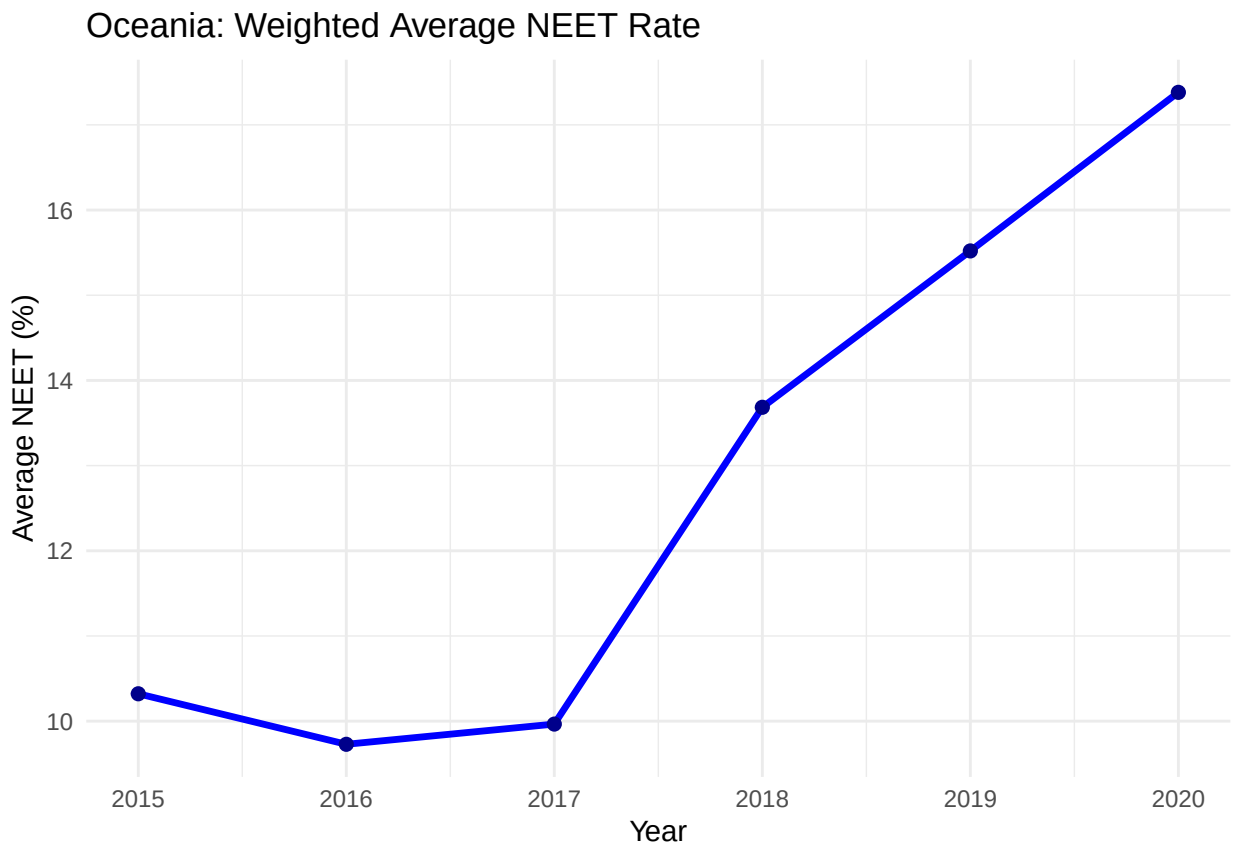

```
ggplot(neet_weighted %>% filter(Continent == "South America",Year >= 2015, Year <= 2020),
       aes(x = Year, y = weighted_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "South America: Weighted Average NEET Rate",
    x = "Year",
    y = "Average NEET (%)"
  ) +
  theme_minimal()
```




```r
# Filter Oceania's data
oceania_neet <- avg_NEET_by_continent_year %>%
  filter(Continent == "Oceania", Year >= 2015, Year <= 2020) %>%
  group_by(Year)

# Line plot
ggplot(oceania_neet, aes(x = Year, y = avg_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Oceania: Average NEET Rate",
    x = "Year",
    y = "Average NEET (% of youth 15-24)"
  ) +
  theme_minimal()
```

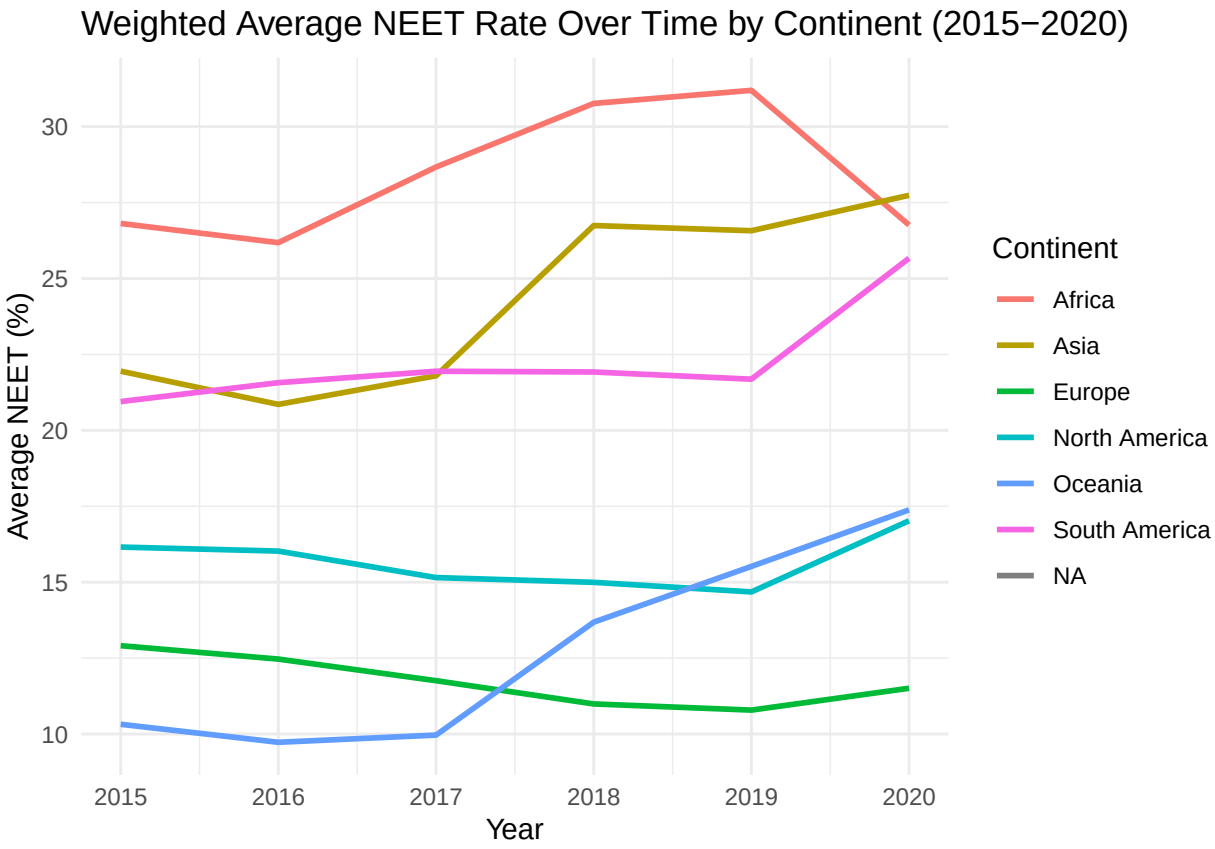



```
ggplot(neet_weighted %>% filter(Continent == "Oceania",Year >= 2015, Year <= 2020),
       aes(x = Year, y = weighted_NEET)) +
  geom_line(color = "blue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "Oceania: Weighted Average NEET Rate",
    x = "Year",
    y = "Average NEET (%)"
  ) +
  theme_minimal()
```




```
ggplot(neet_weighted %>% filter(Year >= 2015, Year <= 2020),
       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()
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
## (`geom_line()`).
```


Weighted Average NEET Rate Over Time by Continent (2015−2020)
Average NEET (%)
30
25
20
15
10
2015
2016
2017
2018
2019
2020
Year
Continent
Africa
Asia
Europe
North America
Oceania
South America
NA