Effects of Tertiary Education Enrolment in Europe

Maximilian Grotz

11 3 2021

Setup

In the following code I import relevant packages and the datasets used for this analysis. Additionally, I will rename some countries and geographical regions to shorter names.

```
TertEduEU_table <- read_csv("../sourcedata/TertiaryEduEU_EuroStat_Correct.csv",
    col_types = cols(SEX = col_skip(), AGE = col_skip(),
        UNIT = col_skip()))

colnames(TertEduEU_table) <- c("Year",
                               "Country",
                               "Type of Education",
                               "Percentage")
TertEduEU_table$Year <- as.integer(TertEduEU_table$Year)

TertEduEU_table$Year <- as.integer(TertEduEU_table$Year)

TertEduEU_table$Percentage <- as.numeric(TertEduEU_table$Percentage)
```

```
## Warning: NAs durch Umwandlung erzeugt
```

```
TertEduEU_table[TertEduEU_table$Country ==
                  "Germany (until 1990 former territory of the FRG)",
                "Country"] <- "Germany"
TertEduEU_table[TertEduEU_table$Country ==
                  "European Union - 28 countries (2013-2020)",
                "Country"] <- "EU 28"
TertEduEU_table[TertEduEU_table$Country ==
                  "Euro area - 19 countries  (from 2015)",
                "Country"] <- "Euro area"

TertEduEU_table <- TertEduEU_table[!TertEduEU_table$Country
                                   %in%
                                     c("European Union - 27 countries (from 2020)",
                                       "European Union - 15 countries (1995-2004)"), ]

TertEduEU_table <- TertEduEU_table %>%
  mutate(EU = as.factor(ifelse(Country == "EU 28", 1, 0)))
```

Test for NAs

I received the message that some values were transformed into NAs. This worried me, so I printed all rows with NAs. Those are only 4 entries from Montenegro in 2010, so I decided to keep the data frame clean and delete those rows.

```
TertEduEU_table %>%
  filter_all(any_vars(is.na(.)))
```

```
## # A tibble: 4 x 5
##    Year Country   `Type of Education`                                Percentage EU
##   <int> <chr>     <chr>                                                   <dbl> <fct>
## 1  2010 Monteneg~ Less than primary, primary and lower seconda~            NA 0
## 2  2010 Monteneg~ Upper secondary, post-secondary non-tertiary~            NA 0
## 3  2010 Monteneg~ Upper secondary and post-secondary non-terti~            NA 0
## 4  2010 Monteneg~ Tertiary education (levels 5-8)                          NA 0
```

```
TertEduEU_table <- TertEduEU_table  %>%
  drop_na()
```

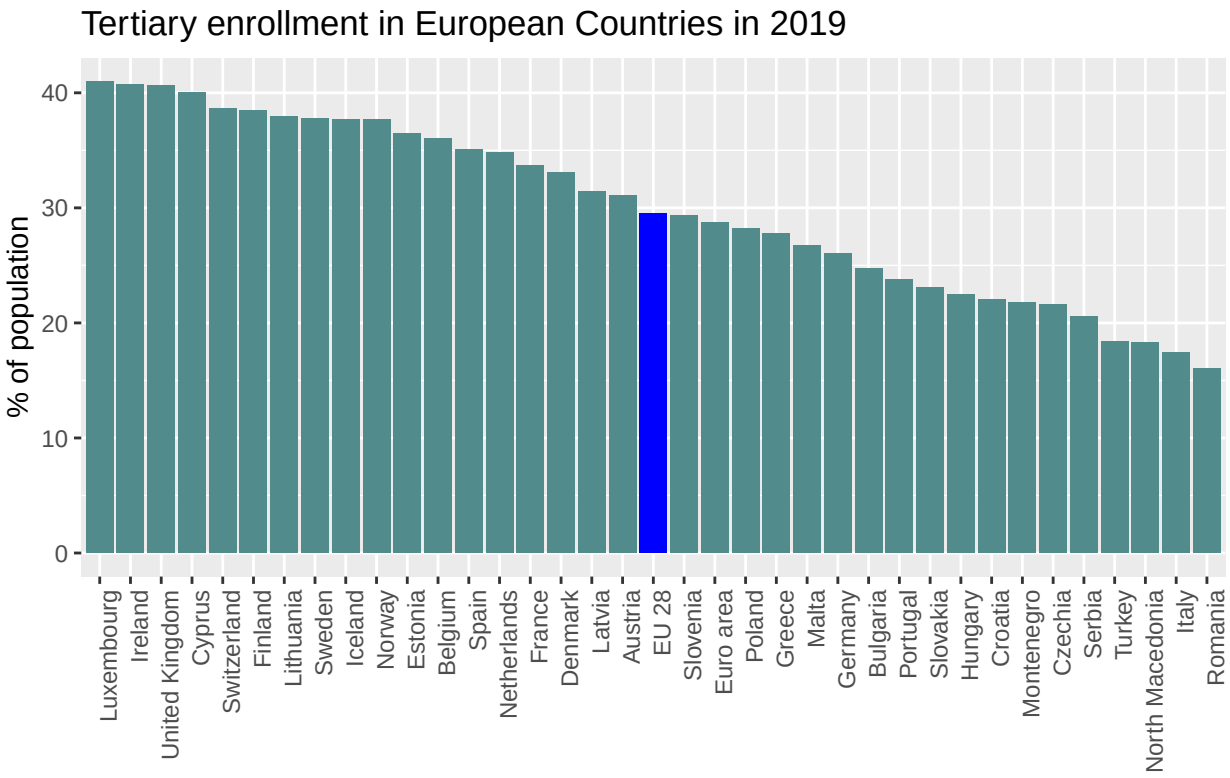
Exploring the dataset

Current level of Tertiary Enrollment

The following graph gives an overview over the most current data on tertiary education enrollment in European countries. The EU wide figure can be seen as well

```
TertEduEU_2019 <- TertEduEU_table[((TertEduEU_table$Year == 2019) &
                                     (TertEduEU_table$`Type of Education` ==
                                        "Tertiary education (levels 5-8)")), ]

ggplot() +
  geom_bar(data = TertEduEU_2019, aes(x = reorder(Country, -Percentage),
                                      weight = Percentage,
                                      fill = EU)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Tertiary enrollment in European Countries in 2019",
       caption = "Source: EuroStat",
       x = "",
       y = "% of population") +
  scale_fill_manual(values = c("darkslategray4", "blue")) +
  theme(legend.position = "none")
```



Source: EuroStat

EU and Eurozone

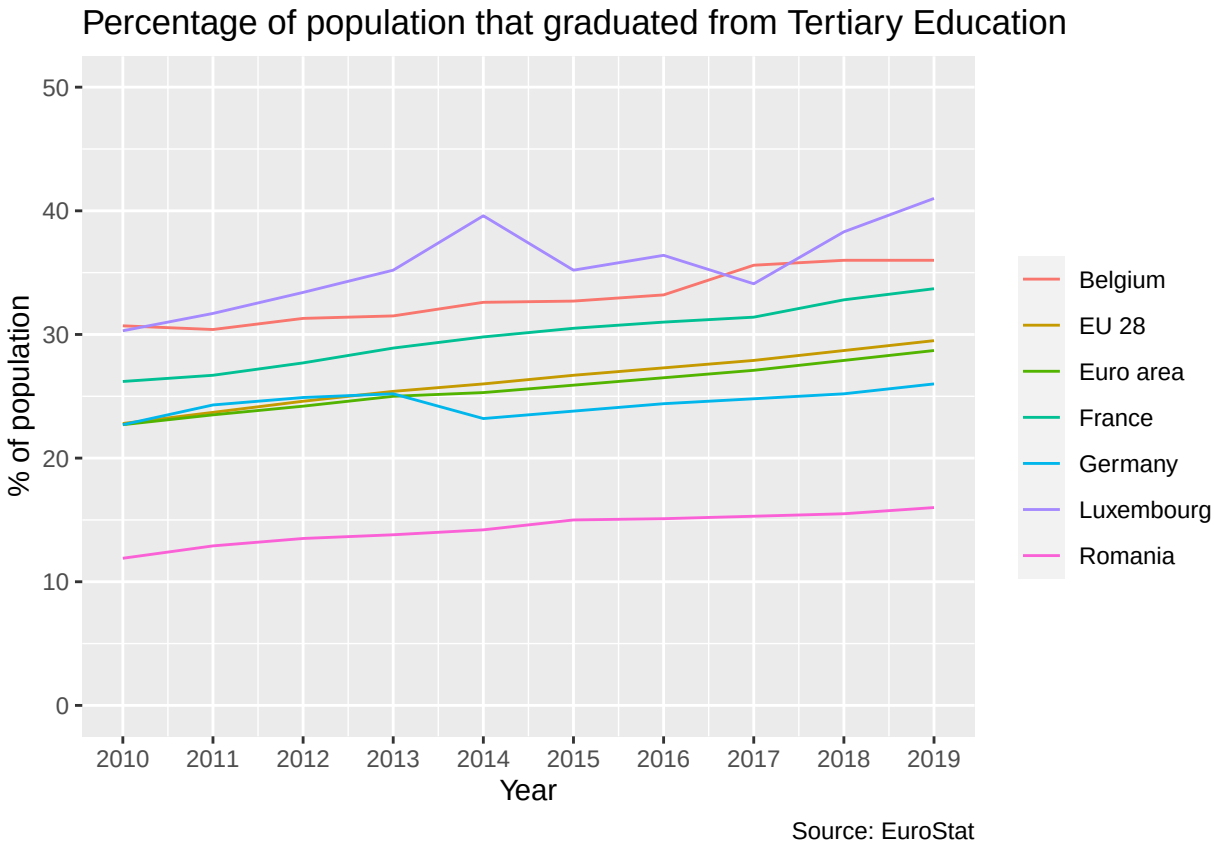
The focus of this analysis should be the tertiary education. We have the data for tertiary enrollment for all European countries, but also for EU and Euro area countries as a whole. Those might be interesting to compare future data to. Additionally, I will add the change over time of other selected countries.

```
TertEduEU_EU_EA <- TertEduEU_table[TertEduEU_table$Country %in% c("EU 28",
                                                                  "Euro area",
                                                                  "Luxembourg",
                                                                  "Germany",
                                                                  "Belgium",
                                                                  "France",
                                                                  "Romania"), ]

TertEduEU_EU_EA <- TertEduEU_EU_EA[TertEduEU_EU_EA$`Type of Education` ==
                                     "Tertiary education (levels 5-8)", ]

ggplot() +
  geom_line(data = TertEduEU_EU_EA, aes(x = Year,
                                        y = Percentage,
                                        color = Country)) +
  labs(title = "Percentage of population that graduated from Tertiary Education",
       y = "% of population",
       x = "Year",
       color = "",
```

```
      caption = "Source: EuroStat") +
  scale_x_continuous(breaks = c(2010:2019)) +
  ylim(0, 50)
```



Adding GDP to the analysis

Setup

Here I import the data, rename some countries/areas and prepare it for the analysis.

```
GDP_table <- read_excel("../sourcedata/GDP_EU_EuroStat.xlsx",
    sheet = "Sheet 1", skip = 8, n_max = 40)
```

```
## New names:
## * `` -> ...3
## * `` -> ...5
## * `` -> ...7
## * `` -> ...9
## * `` -> ...11
## * ...
```

```r
GDP_table <- GDP_table[! GDP_table$TIME == "GEO (Labels)" &
                         ! GDP_table$TIME ==
                         "European Union - 27 countries (from 2020)", ]

GDP_table[GDP_table$TIME == "Germany (until 1990 former territory of the FRG)",
          "Country"] <- "Germany"
GDP_table[GDP_table$TIME == "European Union - 28 countries (2013-2020)",
          "Country"] <- "EU 28"
GDP_table[GDP_table$TIME == "Euro area - 19 countries  (from 2015)",
          "Country"] <- "Euro area"

colnames(GDP_table)[1] <- "Country"

GDP_table <- GDP_table  %>%
  select("Country", "2019")

GDP_table$GDPperCap <- as.integer(GDP_table$`2019`)
```

```
## Warning: NAs durch Umwandlung erzeugt
```

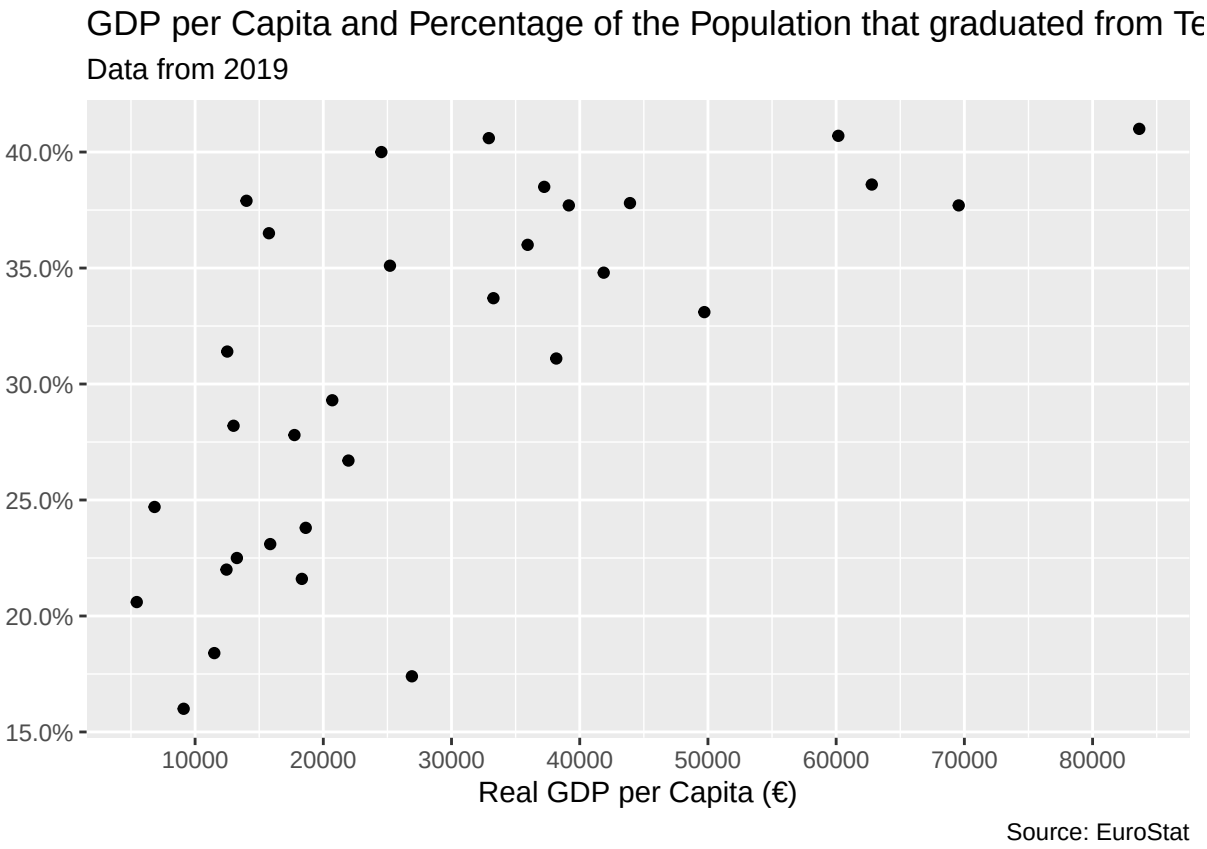
Correlation between real GDP per capita and enrollment in tertiary education

There is a strong correlation. It is doubtfult hat there is high university enrollment causes higher income. Rather, welthier countries simply spend more on educaiton and more students are interested in a university education.

```r
TertEdu_GDP_EU <- merge(TertEduEU_2019, GDP_table, by = "Country")
TertEdu_GDP_EU$Percentage <- TertEdu_GDP_EU$Percentage /100

ggplot() +
  geom_point(data = TertEdu_GDP_EU, aes(x = GDPperCap, y = Percentage)) +
  labs(title = "GDP per Capita and Percentage of the Population that graduated from Tertiary Education"
       subtitle = "Data from 2019", x = "Real GDP per Capita (€)",
       y = "",
       caption = "Source: EuroStat") +
  scale_y_continuous(labels = percent) +
  scale_x_continuous(breaks = seq(0, 85000, 10000))
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```


GDP per Capita and Percentage of the Population that graduated from Te
Data from 2019
40.0%
35.0%
30.0%
25.0%
20.0%
15.0%
10000
20000
30000
40000
50000
60000
70000
80000
Real GDP per Capita (€)
Source: EuroStat