Financial vulnerability of young people due to student debt

By Mick Bakker (2861052), Alex Emmerig (2855406), Loek Graven (2863472), Yaiden de Jong (2806624), Obbe Rozier (2870631), Jenno Schmersal (2853582)

Tutorial 2, Group 4

Tutor: Chantal Schouwenaar

# Part 1 - Identify a Social Problem

## 1.1 Describe the Social Problem

Student debt is becoming a serious social problem, especially for young people who are just starting their adult lives. According to Jongeren Informatie Punt (n.d.), a high student debt can significantly reduce the chances of getting a mortgage. This makes it harder for young people to buy a house, which contributes to the broader housing shortage. As housing becomes less accessible, students are often forced to rent longer, live with their parents, or move to less desirable areas. This has long-term effects on their independence, stability, and ability to build up wealth. Moreover, research from Vox Magazine (Noij, 2021) shows that the current loan system increases inequality in society. Students from lower-income families are more likely to accumulate high debts, while students with wealthier parents can avoid borrowing altogether. This deepens the divide between social groups, as those with debt may postpone major life decisions like buying a home, starting a business, or even having children. The Nibud Studentenonderzoek 2024 (Groen et al., 2025) also highlights the psychological burden of debt. Many students report experiencing stress, anxiety, and financial insecurity, which can negatively affect their academic performance and mental health. Together, these findings show that student debt is not just a financial issue but also a far-reaching social and it has emotional consequences that affect young people's wellbeing and future prospects.

We will also visualize the ratio of student debts to GDP per capita of the U.S., the UK and the Netherlands. We personally couldn't find this comparison between the three countries in a source, but this will give a better idea of the severity of the debts compared to each other.

# Part 2 - Data Sourcing

## 2.1 Load in the data

Data Netherlands: https://www.cbs.nl/nl-nl/nieuws/2024/41/minder-mensen-met-studieschuld (CBS, 2024)

Data UK: https://www.gov.uk/government/statistics/student-loans-in-england-2023-to-2024/student-loans-in-england-financial-year-2023-24 (GOV.UK, 2024)

Data USA: https://educationdata.org/average-student-loan-debt (Education Data Initiative, 2024)

Data GDP per Capita https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2023&locations=NL-GB-US&start=2007 (World Bank Open Data, n.d.)

```
# Making dataframes of the excel-files
df1 = read_excel("Studieschuld_2011-2024(goeie).xlsx", skip = 1)
df2 = read_excel("England_figure_10.xlsx")
df3 = read_excel("Average Total Student Debt USA.xlsx")
df4 = read_excel("gdp_per_capita2.xlsx")
```

## 2.2 Provide a short summary of the dataset(s)

```
head(df1)
```

```
## # A tibble: 6 x 6
##   ...1    ...2  Studieschuld ...4     ...5       ...6
##   <chr>   <lgl> <chr>        <chr>    <chr>      <chr>
## 1 Periode NA    Personen     Som      Gemiddeld  Mediaan
## 2 <NA>    NA    x 1 000      mld euro 1 000 euro <NA>
## 3 <NA>    NA    <NA>         <NA>     <NA>       <NA>
## 4 2011    NA    <NA>         12.6     12600      <NA>
## 5 2012    NA    <NA>         11.8     11800      <NA>
## 6 2013    NA    <NA>         11.9     11900      <NA>
```

In this case we see 6 variables, but we miss some information on what units they are in. We also don't know anything about the year/moment in which this data has been captured yet.

### Dataset 1:

Source: Centraal Bureau voor de Statistiek (CBS). Credible because it is an independent government organization with the duty to provide statistics. Metadata: Title: Studieschulden 2011–2024 Publisher: CBS (Statistics Netherlands) Language: Dutch Format: Excel (.xlsx) Temporal coverage: Annually, from 2011 to 2024 Variables: Year (reference date: January 1) Age groups (e.g., 18–24, 25–29, etc.) Number of individuals with student debt Average student debt per person Total student debt in the age group Update frequency: Annually License: CC BY 4.0 (free to use with attribution) Data type: Aggregated, administrative and survey-based

### Dataset 2

Source: Student Loans Company (SLC), published on GOV.UK. Credible because The SLC is a government-owned organization responsible for administering the student loans in the UK, so these are official statistics Metadata: Title: Average higher education loan balance on entry into repayment by financial year 2006-07 to 2023-24: England & EU (£) Publisher: Student Loans Company (SLC) / GOV.UK Language: English Format: Excel (.xlsx) Temporal coverage: Annually, from 2006-07 - 2023-24 Variables: Financial year Average student debt per person at entry into repayment Update frequency: Annually License: Open Government Licence v3.0 (free to use with attribution) Data type: Aggregated, based on administrative data from Student Finance England

### Dataset 3:

Source: EducationData.org, maintained by Education Data Initiative. Credible because EducationData.org is a non-profit organization that provides education-related data from credible sources, like the U.S. Department of Education, NCES, and the Federal Reserve. Metadata: Title: Average Student Loan Debt in the U.S. Publisher: Education Data Initiative Language: English Format: Excel (.xlsx) Temporal coverage: Annually, from 2004 to 2021 Variables: Year Average total student loan debt Update frequency: Annually License: Not clearly stated. Data type: Aggregated, compiled from administrative government data

### Dataset 4:

Source: The World Bank Open Data. Credible because The World Bank is one of the most trusted organizations for global development data, including GDP per capita. Metadata: Title: GDP per capita (current US) $-$ $UnitedKingdom, UnitedStates, NetherlandsPublisher$ : $TheWorldBankLanguage$ : $EnglishFormat$ : $Excel(.xlsx)Temporalcoverage$ : $Annually, from2007to2023Variables$ : $YearCountry(UnitedKingdom,$ Update frequency: Annually License: Open data (CC BY 4.0), free to use with attribution. Data type: Aggregated, sourced from national accounts and World Bank calculations.

These are things that are usually included in the metadata of the dataset. For your project, you need to provide us with the information from your metadata that we need to understand your dataset of choice.

### 2.3 Describe the type of variables included

The first 3 datasets are primarily administrative data that is kept by the respective (semi-)governments agencies. The GDP is calculated by national accounts data and population estimates.

## Part 3 - Quantifying

### 3.1 Data cleaning

```
# ---- DATA CLEANING EN MERGING ----


# 1. Importing Excel-files
df1 = read_excel("Studieschuld_2011-2024(goeie).xlsx", skip = 1)
df2 = read_excel("England_figure_10.xlsx")
df3 = read_excel("Average Total Student Debt USA.xlsx")
df4 = read_excel("gdp_per_capita2.xlsx")


# 2. Removing unnecessary rows and columns
df1 = df1[-c(1, 2, 3), ]
df1 = df1[, -c(2, 3, 6)]
df1 = df1[, -c(2) ]
df2 = df2[, -c(3)]


# 3. Converting academic year to calendar year of UK data
df2$'Financial year' <- 2007:2024


# 4. Renaming columns for merging
colnames(df1)[1] <- "Year"
colnames(df1)[2] <- "Studieschuld_NL"
colnames(df2)[1] <- "Year"
colnames(df2)[2] <- "Studieschuld_UK"
colnames(df3)[1] <- "Year"
colnames(df3)[2] <- "Studieschuld_US"

# 5. Converting 'Year' to a character
```

```r
df2$Year <- as.character(df2$Year)
df3$Year <- as.character(df3$Year)

# 6. Cleaning 'Year' values, remove '*' in Dutch data
df1$Year[df1$Year == "2023*"] <- 2023
df1$Year[df1$Year == "2024*"] <- 2024

# 7. Merging df1, df2 and df3
df_merge <- full_join(df1, df2, by = "Year")
df_merge <- full_join(df_merge, df3, by = "Year")

# 8. Sorting data by year (ascending)
df_sorted <- df_merge[order(df_merge$Year, decreasing = FALSE), ]


# 9. Converting debt columns to numeric
df_sorted$Studieschuld_NL <- as.numeric(df_sorted$Studieschuld_NL)
df_sorted$Studieschuld_UK <- as.numeric(df_sorted$Studieschuld_UK)
df_sorted$Studieschuld_US <- as.numeric(df_sorted$Studieschuld_US)

# (making variable, also noted in 3.2 but code wouldnt work if i didnt place it here)
df_sorted$Growth_NL <- c(NA, diff(df_sorted$Studieschuld_NL) / head(df_sorted$Studieschuld_NL, -1) * 10
df_sorted$Growth_UK <- c(NA, diff(df_sorted$Studieschuld_UK) / head(df_sorted$Studieschuld_UK, -1) * 10
df_sorted$Growth_US <- c(NA, diff(df_sorted$Studieschuld_US) / head(df_sorted$Studieschuld_US, -1) * 10


df_sorted <- df_sorted %>%
  mutate(across(c(Studieschuld_NL, Studieschuld_UK, Studieschuld_US, Growth_NL, Growth_UK, Growth_US),
         Year = as.numeric(Year))


# 10. converting valuta to euro
df_sorted <- df_sorted %>%
  mutate(
    Studieschuld_US = Studieschuld_US * 0.86,
    Studieschuld_UK = Studieschuld_UK * 1.17
  )

# 11. Read and clean GDP per capita data
gdp_per_capita_data <- read_xlsx("gdp_per_capita2.xlsx")


jaar_kolommen <- as.character(unlist(gdp_per_capita_data[3, ]))

jaar_kolommen[is.na(jaar_kolommen)] <- paste0("V", which(is.na(jaar_kolommen)))
df_gdp <- gdp_per_capita_data[4:6, ]
colnames(df_gdp) <- jaar_kolommen
df_gdp <- df_gdp %>%
  select(`Country Name`, `2007`:`2023`)

df_gdp_long <- df_gdp %>%
  pivot_longer(
    cols = -`Country Name`,
```

```
    names_to = "Year",
    values_to = "GDP_PC"
  )

df_gdp_long$GDP_PC <- as.numeric(df_gdp_long$GDP_PC)
df_gdp_long$Year <- as.integer(df_gdp_long$Year)
```

Explain why you merged the datasets in the way you did: We merged the data from the student debts in the UK, US and the Netherlands together because all the data came from single datasets. In all the datasets, "years" was the variable being in all of them, so we used this variable to combine all 3 of the datasets together. In the GDP per capita dataset we already have all the data from the US, UK and the Netherlands in one dataframe, so we merged this one to the student debt one by year as well. Why was this merge option the best choice: This merge option gives us a way to produce a detailed temporal view of the student debt and the GDP and seemed the only logical one with these datasets. Suggest a way in which this could have been done better in an ideal world: In an ideal world we would have found a dataset with all student debts per year for a lot more countries. In reality we found that surprisingly this is not very well documented for a lot of countries. Also it was hard to find reliable datasets from the same source for the 3 different countries that included the same years. This is why in our dataset the student debt of the UK starts in 2007, NL in 2011, and US in 2014. For clearancy we used 2014 as the starting point for our analyses. Discuss searches for errors and document any found + how you fixed them: About halfway through the analyses, when we wanted to make the axes have a clear variable, we noticed that the student debts were all in different currencies (e.g. UK was pounds, US was dollars and NL was euros). So, we had to correct all these numbers to make it one currency (we chose euros). The GDP dataset was in dollars for all countries, so we converted all of them to euros. We did however choose the current conversion rate, and not the historical conversion rate per year, as this would make this assignment a bit too extensive. This could be something we could work on in future research.

## 3.2 Generate necessary variables

### Variable 1

```
df_sorted$Growth_NL <- c(NA, diff(df_sorted$Studieschuld_NL) / head(df_sorted$Studieschuld_NL, -1) * 10
df_sorted$Growth_UK <- c(NA, diff(df_sorted$Studieschuld_UK) / head(df_sorted$Studieschuld_UK, -1) * 10
df_sorted$Growth_US <- c(NA, diff(df_sorted$Studieschuld_US) / head(df_sorted$Studieschuld_US, -1) * 10
```

Explain which variables you used/combined to create your new variable: -We used the student debts and the years to create the year-to-year growth of the studet debt per country.

Why is the creation of this variable informative and useful for your analyses & your topic: - This new variable gives us the information about how the student debts have been changing over time compared to the economy around it. What do you intent to analyse/visualize using this variable: - We intend to visualize how economic disruptions like COVID-19 influence the student debts. With this new variable we can analyze how these economic disruptions influence the student debt.

### Variable 2

```
#Preparing GDP data

jaar_kolommen <- as.character(unlist(df4[3, ]))
```

```r
jaar_kolommen[is.na(jaar_kolommen)] <- paste0("V", which(is.na(jaar_kolommen)))
df_gdp <- df4[4:6, ]
colnames(df_gdp) <- jaar_kolommen
df_gdp <- df_gdp %>%
  select(`Country Name`, `2007`:`2023`)

df_gdp_long <- df_gdp %>%
  pivot_longer(
    cols = -`Country Name`,
    names_to = "Year",
    values_to = "GDP_PC"
  )
df_gdp_long$GDP_PC <- as.numeric(df_gdp_long$GDP_PC)
df_gdp_long$Year <- as.integer(df_gdp_long$Year)

# Renaming and changing to wide format
df_gdp_wide <- df_gdp_long %>%
  pivot_wider(
    names_from = `Country Name`,
    values_from = GDP_PC
  ) %>%
  rename(
    GDP_PC_UK = `United Kingdom`,
    GDP_PC_NL = Netherlands,
    GDP_PC_US = `United States`
  ) %>%
  arrange(Year)

# Valuta conversion to euro
df_gdp_wide <- df_gdp_wide %>%
  mutate(
    GDP_PC_UK = GDP_PC_UK * 0.86,
    GDP_PC_NL = GDP_PC_NL * 0.86,
    GDP_PC_US = GDP_PC_US * 0.85
  )

# Merging and making variable
df_merged <- full_join(df_sorted, df_gdp_wide, by = "Year")
df_merged <- df_merged %>%
  mutate(
    Schuld_GDP_NL = 100 * Studieschuld_NL / GDP_PC_NL,
    Schuld_GDP_UK = 100 * Studieschuld_UK / GDP_PC_UK,
    Schuld_GDP_US = 100 * Studieschuld_US / GDP_PC_US
  )
```

Explain which variables you used/combined to create your new variable: -We used student debt and the GDP per capita ratio to create a visualisation of student debt as a percentage of GDP per capita. We also used the year-on-year procentual growth of student debt to create a visualisation for the growth in student debt per country. We used both these plots for our event analysis.

Why is the creation of this variable informative and useful for your analyses & your topic: - Our topic is about how student debt affects students, and these variables help us understand how big the debt is, how much of a burden it puts on students, and how it's been changing over time compared to the economy around them.

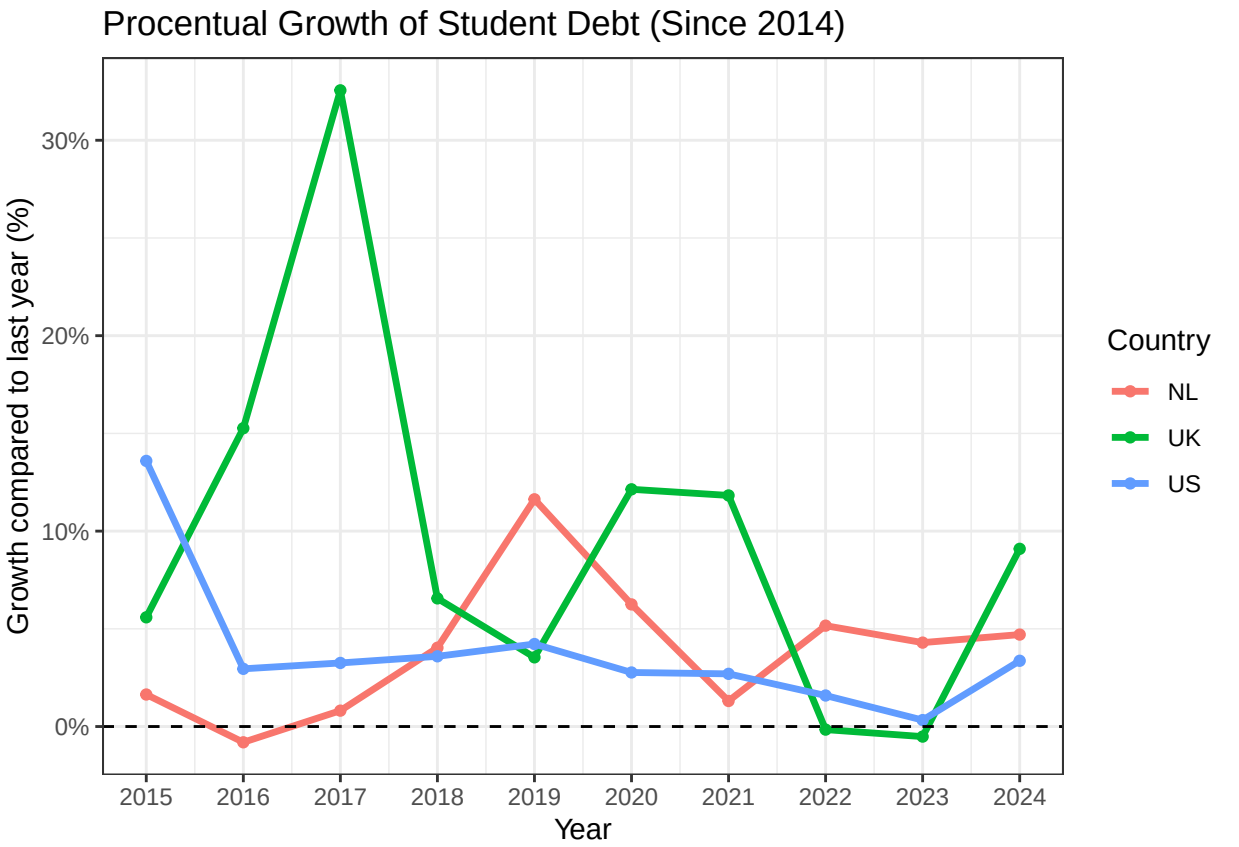
What do you intent to analyse/visualize using this variable: - We intend to visualize how economic disruptions like COVID-19 influence our plots. With our new variables we can analyze how these economic disruptions influence the student debt and the GDP per capita. COVID-19 caused for example major job losses all over the world and Brexit caused economic uncertainty in the UK. Both have an effect on our variables.

### 3.3 Visualize temporal variation

```
#Procentual growth per year
df_growth_long <- df_sorted %>%
  select(Year, Growth_NL, Growth_UK, Growth_US) %>%
  pivot_longer(
    cols = starts_with("Growth_"),
    names_to = "land",
    names_prefix = "Growth_",
    values_to = "groei_percentage"
  )

df_growth_long %>%
  filter(Year >= 2015) %>%
  ggplot(aes(x = Year, y = groei_percentage, color = land)) +
  geom_line(size = 1.2) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed", color = "black") +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  scale_x_continuous(breaks = 2015:2024) +
  labs(
    title = "Procentual Growth of Student Debt (Since 2014)",
    x = "Year",
    y = "Growth compared to last year (%)",
    color = "Country"
  ) +
  theme_bw()
```

```
## 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.
```




```
ggsave("Temporal_Visualization2.png", width = 8, height = 5)
```

Short description of how this visualization aligns with your topic: - This graph aligns with our problem because it visualizes the percentage growth of student debt in the Netherlands, the UK, and the US. It shows that the UK has had the largest growth in student debt since 2016, followed by the Netherlands, while the US has experienced the lowest growth. The increase in student debt in the UK is likely due to Brexit, while in the Netherlands it may be attributed to a change in the student finance system, which did not work as intended. In recent years, under the new system, the growth has slowed down. Meanwhile, the growth of student debt in the US has remained relatively consistent.

### 3.4 Visualize sub-population variation

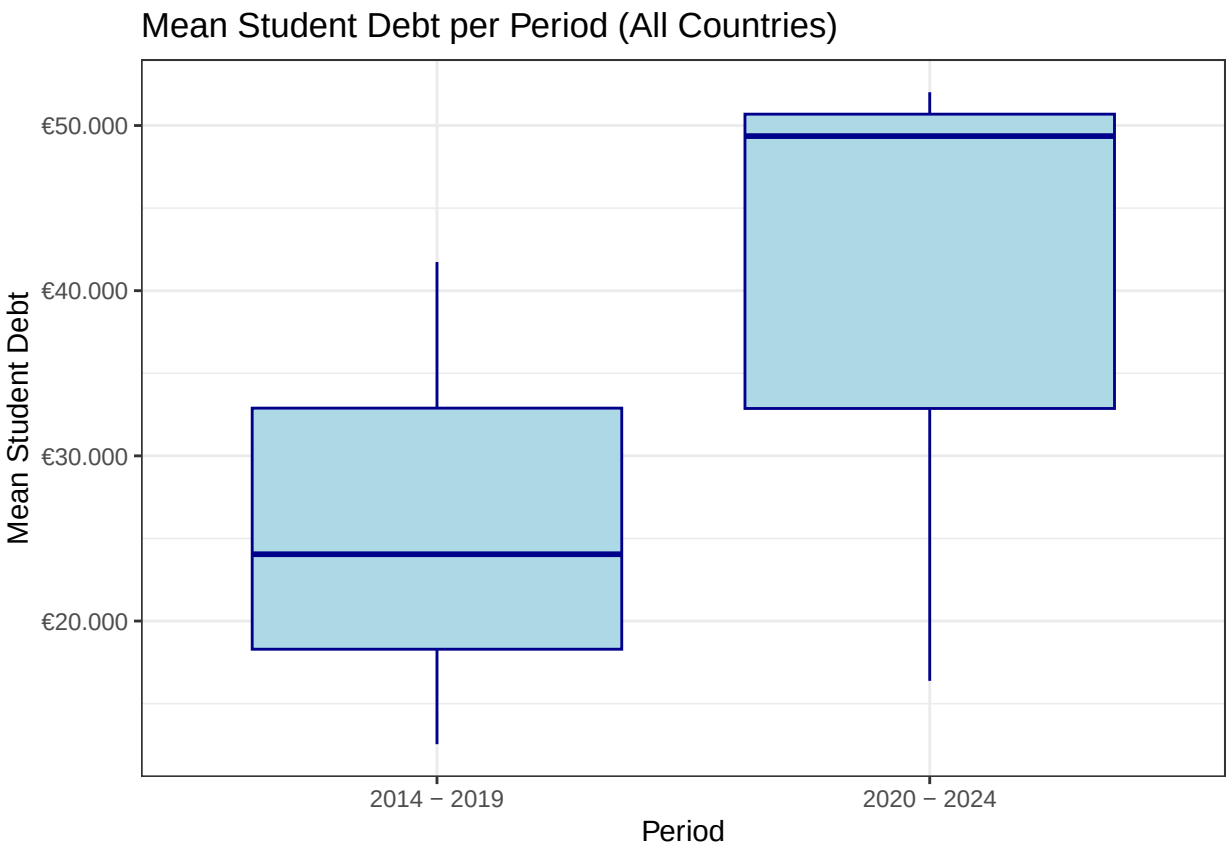
```
data_selected <- df_sorted %>%
  select(Year, starts_with("Studieschuld"))

# Converting data to long format
data_long <- data_selected %>%
  pivot_longer(
    cols = -Year,
    names_to = c("variabele", "land"),
    names_pattern = "(Studieschuld)_(.*)"
  ) %>%
  rename(studieschuld = value)
```

```r
# Label time period
data_long <- data_long %>%
  mutate(
    period = case_when(
      Year <= 2019 ~ "2014 - 2019",
      Year > 2019 ~ "2020 - 2024"
    )
  )


# Calculate mean per country per period
data_avg <- data_long %>%
  group_by(land, period) %>%
  summarise(studieschuld_mean = mean(studieschuld, na.rm = TRUE), .groups = "drop")

# Create boxplot
ggplot(data_avg, aes(x = period, y = studieschuld_mean)) +
  geom_boxplot(fill = "lightblue", color = "darkblue") +
  labs(title = "Mean Student Debt per Period (All Countries)",
       x = "Period",
       y = "Mean Student Debt") +
  scale_y_continuous(labels = label_dollar(prefix = "€", big.mark = ".", decimal.mark = ",")) +
  theme_bw()
```



```r
ggsave("Sub_Population.png", width = 8, height = 5)
```

Short description of how this visualization aligns with your topic: - This boxplot shows that the average student debt in all countries clearly went up between 2020 and 2024, compared to 2014 to 2019. The median is higher, and the total range of debt is also bigger. This means that student debt has become a bigger problem for young people, especially during and after the COVID-19 period. Because of this, it has become harder for many to get a mortgage, which makes the housing shortage worse. The higher debt also causes more stress and worry, which affects young people's mental health. On top of that, it increases the gap between rich and poor, because students from richer families are less affected.

### 3.5 Visualize spatial variation

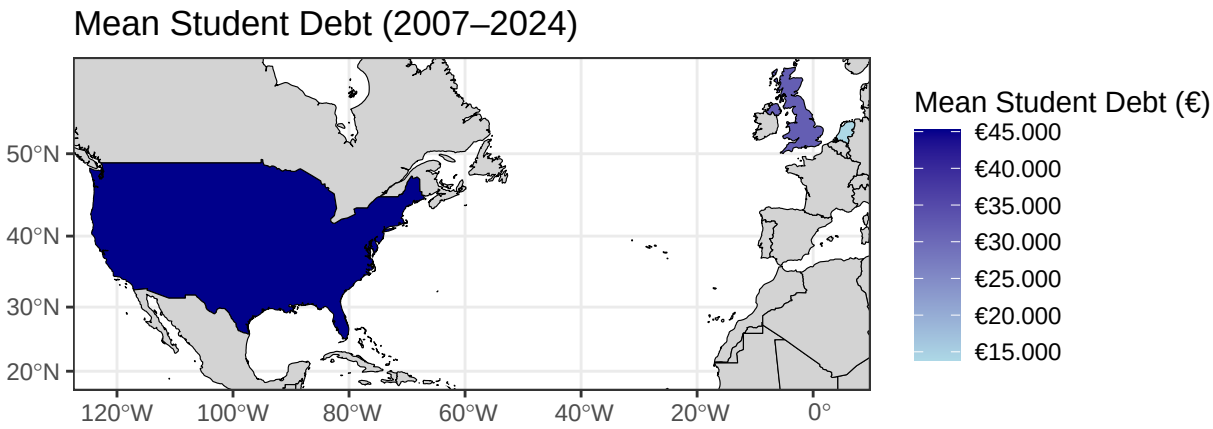
```r
#means
gemiddelde <- data_selected %>%
  summarise(
    Studieschuld_NL = mean(Studieschuld_NL, na.rm=TRUE),
    Studieschuld_UK = mean(Studieschuld_UK, na.rm=TRUE),
    Studieschuld_US = mean(Studieschuld_US, na.rm=TRUE)
  ) %>%
  pivot_longer(everything(), names_to = "land", values_to = "gemiddelde_schuld") %>%
  mutate(
    land = recode(land,
        Studieschuld_NL = "Netherlands",
        Studieschuld_UK = "United Kingdom",
        Studieschuld_US = "United States of America"
    )
  )


# Worldmap
wereldkaart <- ne_countries(scale = "medium", returnclass = "sf")

wereldkaart_met_data <- wereldkaart %>%
  left_join(gemiddelde, by = c("admin" = "land"))


#Plot
ggplot(wereldkaart_met_data) +
  geom_sf(aes(fill = gemiddelde_schuld), color = "black") +
  scale_fill_gradient(
    low = "lightblue",
    high = "darkblue",
    na.value = "lightgrey",       # <- Grey for countries without data
    name = "Mean Student Debt (€)",
    labels = label_dollar(prefix = "€", big.mark = ".", decimal.mark = ",")
  ) +
  coord_sf(
    crs = 3857,
    xlim = c(-1.35e7, 0.4e6),    # Left (US) to right (NL)
    ylim = c(2.2e6, 8e6)        # Down (South US) to up (North UK)
  ) +
```

```
  labs(title = "Mean Student Debt (2007-2024)") +
  theme_bw()
```




```
ggsave("Spatial_Visualization.png", width = 8, height = 5)
```

Short description of how this visualization aligns with your topic: - This map aligns with our topic because this visualises that the United States have the highest average debt, and the Netherlands the lowest average debt. This visualises that not all student loan systems have an equal financial risk, and the place where you study has a big impact on the height of the debts. This shows where students are most vulnerable for debts.

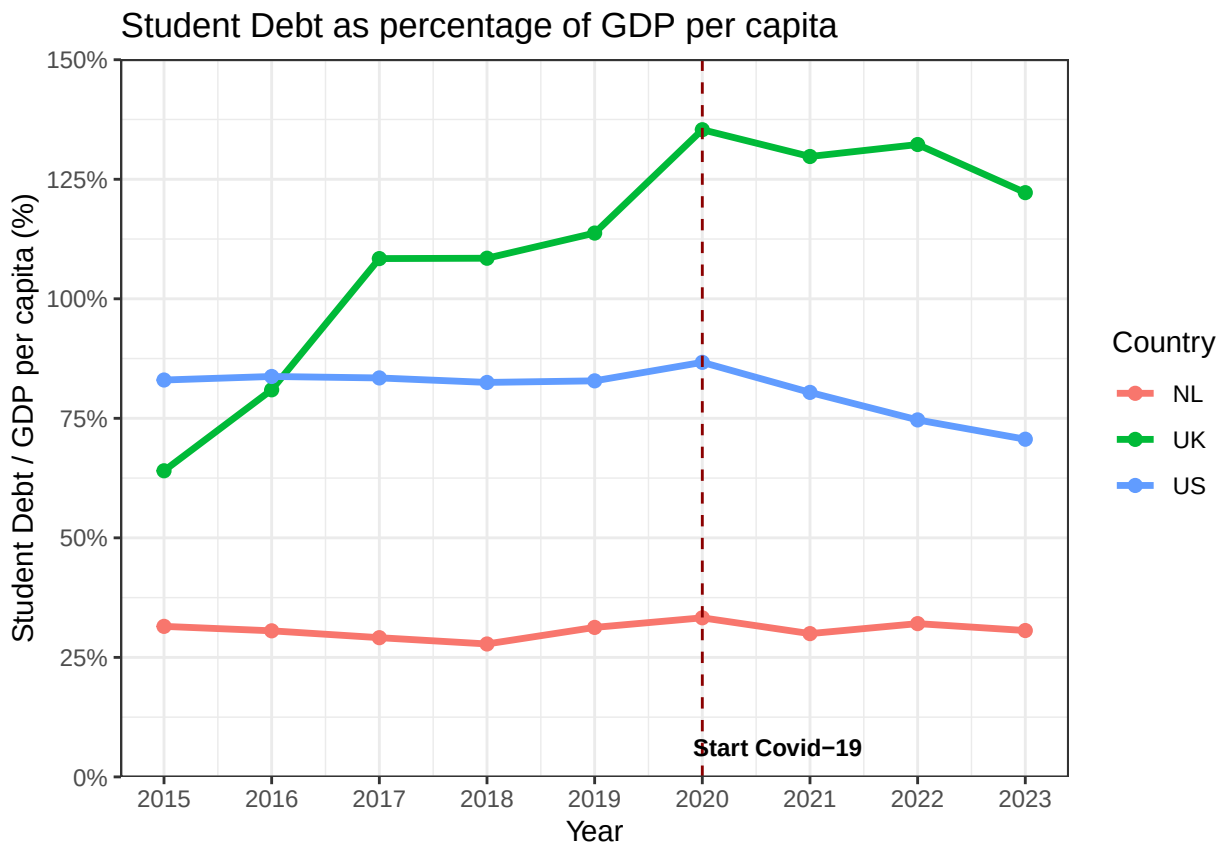
### 3.6 Event analysis

Analyze the relationship between two variables.

```
df_long <- df_merged %>%
  select(Year, Schuld_GDP_NL, Schuld_GDP_UK, Schuld_GDP_US) %>%
  pivot_longer(
    cols = starts_with("Schuld_GDP"),
    names_to = "Land",
    values_to = "Schuld_GDP"
  ) %>%
  mutate(
    Land = case_when(
```

```
      Land == "Schuld_GDP_NL" ~ "NL",
      Land == "Schuld_GDP_UK" ~ "UK",
      Land == "Schuld_GDP_US" ~ "US"
    )
  )

df_long %>%
  filter(Year >= 2015, Year <= 2023) %>%
  ggplot(aes(x = Year, y = Schuld_GDP, color = Land)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  geom_vline(xintercept = 2020, linetype = "dashed", color = "darkred") +
  labs(
    title = "Student Debt as percentage of GDP per capita",
    x = "Year",
    y = "Student Debt / GDP per capita (%)",
    color = "Country"
  ) +
  annotate(
    "text",
    x = 2020.7,
    y = 6.250,
    label = paste0("Start Covid-19"),
    color = "black",
    fontface = "bold",
    size = 3
  ) +
  scale_y_continuous(
    labels = scales::percent_format(scale = 1),
      limits = c(0, 150),
      expand = c(0, 0),
      breaks = seq(0,150,25)
    ) +
  scale_x_continuous(breaks = 2014:2023) +
  theme_bw()
```




```
ggsave("Event_Analysis.png", width = 8, height = 5)
```

What does this visualization mean for your topic: - This visualization shows how student debt changed in the Netherlands, the UK, and the US when COVID-19 started in 2020. In the UK, student debt rose quickly after the pandemic, but it became a smaller percentage of GDP per capita, however students still owed more compared to the average income in the country. In the US, student debt continued to grow steadily, but the change during COVID-19 was smaller. In the Netherlands the increase was slower, but debt still went up slightly after 2020. This shows that COVID-19 had a different impact on student debt in each country, depending on how their systems responded.

# Part 4 - Discussion

## 4.1 Discuss your findings

Our findings from this research are pretty clear if we look at the plots. Looking at our spatial visualisation, the students in the US have the most student debt, around 45.000 dollars, followed by the UK and the Dutch students. In the temporal visualisation we see big fluctuations for the UK procentual growth of student debt (since 2014). This can be explained through different factors that influence the debt of students such as tuition fee hikes, inflation-linked interest, COVID-19 disruptions and mostly Brexit's impact. The US student debt is steadily growing since 2014. The Netherlands has a bit of a fluctuation in 2019, that can be explained by the full implementation of the loan system ('leenstelsel'), replacing grants with loans, leading students to borrow significantly more.

Our sub-population is divided in two periods, from 2014 till 2019 and from 2020 till 2024. If we look at the plot clearly we see that the students from 2020 till 2024 have more debt than the other period. This could

be caused by COVID-19 and the rising inflation. Lastly, we looked at our event analysis using two plots that show how student debt changed over time and compared to income. The first plot shows that student debt kept growing in all countries, but especially in the UK and the US. In the Netherlands, the increase was slower, but debt still went up each year. After COVID-19 started in 2020, debt in the UK rose even faster, likely because students had fewer job options and needed to borrow more. The second plot shows student debt as a percentage of GDP per capita, which means how big the debt is compared to the average income. Since 2017, the UK has had the highest percentage, and it actually became less after COVID-19. The US stayed in second place, and the Netherlands remained the lowest. This shows that student debt became a bigger problem in countries where costs were already high, and COVID-19 made it even harder for students to manage their finances. While our visualizations and variables provide strong insights, there are some limitations to our data. For example, we worked with aggregated data, which means we couldn't work with more personal data, such as gender, income level or education type. Also, the data from different countries were sometimes based on slightly different definitions. For example, in the UK, the reported student debt figures only include students who have already started repaying their loans, excluding the students who are still studying or not yet required to repay. This makes direct comparisons more difficult. Finally, the GDP figures we used are expressed in current dollars, they have not been adjusted for inflation, which reduces the accuracy of comparisons across different years and countries. In future research, we should adjust all of our numbers of GDP per capita and student debt for inflation and historical currency conversions. This could give a very different but above all a more accurate description of our problem Based on what we found, we think that governments should look more closely at how student debt relates to what students can earn later in life. This is important when they make or change loan systems, because it shows how much stress the debt can cause for young people. In the future, it would be helpful to use more detailed information about people, like their family situation and income. This can help us understand better how student debt leads to unfair differences and who needs support the most.

# Part 5 - Reproducibility

## 5.1 Github repository link

Provide the link to your PUBLIC repository here: https://github.com/yaidendejong/groep4

## 5.2 Reference list

Een torenhoge studieschuld bij DUO: wat zijn nou écht de gevolgen? | Jongeren Informatie Punt. (n.d.). https://www.jiphaarlemmermeer.nl/een-torenhoge-studieschuld-bij-duo-wat-zijn-nou-%C3%A9cht-de-gevolgen Groen, A., Van den Enden, G., & Veerman, N. (2025). Nibud Studentenonderzoek 2024: Onderzoek naar de geldzaken van mbo-, hbo en wo-studenten. In Nibud. https://www.tweedekamer.nl/downloads/document?id=2024D32954 Noij, M. (2021, July 14). Studenten na evaluatie leenstelsel: 'Het vergroot de ongelijkheid in de samenleving' Vox Magazine. https://www.voxweb.nl/nieuws/studenten-na-evaluatie-leenstelsel-het-vergroot-de-ongelijkheid-in-de-samenleving Centraal Bureau voor de Statistiek. (2024, October 9). Minder mensen met studieschuld. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/nieuws/2024/41/minder-mensen-met-studieschuld Student loans in England: financial year 2023-24. (2024, June 20). GOV.UK. https://www.gov.uk/government/statistics/student-loans-in-england-2023-to-2024/student-loans-in-england-financial-year-2023-24 Hanson, M. (2024, August 16). Average Student Loan Debt [2024]: by Year, Age & More. Education Data Initiative. https://educationdata.org/average-student-loan-debt World Bank Open Data. (n.d.). World Bank Open Data. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2023&locations=NL-GB-US&start=2007