

The Mental Health Impact of Youth Unemployment in the Netherlands

Collin Veenstra 2856836, Margarida Birne 2861977, Marit Springer 2860040, Naofal Adlouni 2860056, Valentina Caruso 2860056, Xenia Drude 2860056

2025-05-26

Set-up your environment

```
require(tidyverse)
```

The Mental Health Impact of Youth Unemployment in the Netherlands

Collin Veenstra, Margarida Birne, Marit Springer, Naofal Adlouni, Valentina Caruso, Xenia Drude

Tutorial 2

Chantal Schouwenaar

Introduction

1.1 Describe the Social Problem

In the Netherlands, depression has proven to be the most prevalent mental health disorder, with almost one in five individuals expressing depressive symptoms (de Graaf et al., 2010, Trimbos Institute, 2020). Unemployment has been identified as a main factor that aggravates this issue. Beyond its financial implications, being out of work has substantial psychological consequences. It often has an affect on self-esteem, increased physical and cognitive issues, and, in severe cases, suicidal tendencies (Ringdal & Rootjes, 2022). These aspects are not only a result of economic stress but also due to a loss of other benefits such as a sense of purpose, the absence of routine and decreased social connections (ElMorsy, 2024).

There is an extensive amount of research surrounding the correlation among mental health and unemployment. This analysis is focused on how this relationship affects young people in the Netherlands. For emerging adults, working is important as it is associated with obtaining greater independence, forming an identity and starting a career. In the Netherlands, individuals are legally allowed to work from the age of 13 with certain conditions and freely from the age of 16 (Netherlands Enterprise Agency, n.d). Employment from a young age is widely normalized and many times expected during adolescence. Therefore, not having a job can be unsettling and a disruption of deeply rooted social expectations, leading to higher risk of mental health implications.

Part 2 - Data Sourcing

2.1 Load in the data

```
#import of mental data gemeentes 2020
Aatillgoereemental2020 = read.csv("data/Aatillgoereemental2020.csv")
goestillnissewaardmental2020 = read.csv("data/goestillnissewaardmental2020.csv")
noardtillvlielandmental2020 = read.csv("data/noardtillvlielandmental2020.csv")
vlissingentillzwollemental2020 = read.csv("data/vlissingentillzwollemental2020.csv")

#import of mental data gemeentes 2022
Aatillgoereemental2022 = read.csv("data/Aatillgoereemental.csv")
goestillnissewaardmental2022 = read.csv("data/goestillnissewaardmental2022.csv")
noardtillvlielandmental2022 = read.csv("data/noardtillvlielandmental2022.csv")
vlissingentillzwollemental2022 = read.csv("data/vlissingentillzwollemental2022.csv")

#import of unemployment data 2020-2023
gem_unemployment = read.csv("standard data/gem_unemployment.csv")
```

2.2 Provide a short summary of the dataset(s)

```
#head(dataset)
```

For this project, we used all the gemeentes in the Netherlands, that could be roughly considered municipalities. Our data came from the national government's website (CBS StatLine). There is missing data because there was no information available in for certain years - the item 2.3 explains in more detail each variable of our analysis.

Regarding the unemployment data: the observed population were registered in the Basisregistratie Personen (BRP) as of October 1st of the reference year, which is the date when data is collected yearly. The amount of people going to school, for example, is determined based on their status on October 1st. In contrast, the total number of people receiving benefits is calculated throughout the entire month. The classification of the municipalities is based on their situation as of January 1st, 2024. The data on individuals looking for a job and registered with the UWV comes from two sources: from 2005 to 2018, it is taken from UWV Werkbedrijf, and from 2019 onwards, it is based on the Geregistreerd Werkzoekende UWV system. All the numbers are absolute.

Regarding the mental health data: the information is retrieved from a model of calculation used by the National Institute for Public Health and the Environment (RIVM) based on data from the Health Monitor for Adults and the Elderly for the year of 2020 and the the Corona Health Monitor 2022 from GGDs, CBS and RIVM. Every four years, new data from the Health Monitor for Adults and the Elderly become available and that is obtained through surveys with an anonymous nature. The format of the questions follow the MHI-5 and the Kessler-10 (K10) questionnaires.

2.3 Describe the type of variables included

MHI-5: Percentage of people reporting psychological complaints

The analysis surrounding psychological complaints is measured by the Mental Health Inventory 5 (MHI-5), which is the international standard procedure to measure mental health. It includes five questions

related to the well being of an individual over the past month, related to displays of anxiety, depression, behavioral/emotional control and general positive effect. Based on the results of the questionnaire, people received a score ranging from 0 (extremely unhealthy) to 100 (perfectly healthy). A person is considered psychologically unhealthy by the Statistics Netherlands (CBS) if they have received a score lower than 60. The only data we had available was for the year of 2022 because the data of 2020, 2021 and 2023 were not available and that explains the N/A markings in the graph.

High Risk Anxiety and Depression: Percentage of people aged 18 or older at high risk of anxiety disorder or depression

The data was based on the Kessler-10 (K10) questionnaire: a very commonly used screening questionnaire for anxiety and depression. It includes 10 questions concerning emotional states like worthlessness, tiredness and nervousness (during the past 4 weeks). Each question is answered on a 5 point scale, ranging from “Always” (highest score of 5) to “Never” (lowest score of 1). If 3 or more answers are marked as “Impossible to answer”, the indicator receives the value Impossible. If 1 or 2 items are Impossible, their values are estimated based on the average score for those items. The total score ranges from 10 to 50, where 10-50 suggests that there’s no or low risk, 16-29 that there is moderate risk, and 30-50 that there is high risk of anxiety disorder or depression. We could find data for 2020 and 2022.

Unemployment Variables

The data inside the table shows the labor market situation of young people who are between 15 and 27 years old in the Netherlands (total youth), who were registered in the Basisregistratie Personen (BRP). It indicates whether a person is working, studying, searching for work and registered as job seekers at the UWV, or receiving benefits. The total unemployed youth represents the people that are not actively looking for a job or could not find work. The people working can be divided into self employed or employees. This data can be split in different regions.

Part 3 - Quantifying

3.1 Data cleaning

```
#make 1 large mental data set of all the small partial mental datasets for 2020 and 2022
gemeentesmentaldata2020 = bind_rows(Aatillgoereemental2020,goestillnissewaardmental2020,noardtillvlielardmental2020)
gemeentesmentaldata2022 = bind_rows(Aatillgoereemental2022,goestillnissewaardmental2022,noardtillvlielardmental2022)

write.csv(gemeentesmentaldata2020,"data/gemeentesmentaldata2020.csv")
write.csv(gemeentesmentaldata2022,"data/gemeentesmentaldata2022.csv")

#####
####data cleaning, get rid of the useless variables in mental data set####
#####

#filtering useless categories like smoking, alcohol consumption and other health statistics, so that we can focus on mental health
filtered_gemeentesmentaldata2020 = gemeentesmentaldata2020 %>%
  select(Leeftijd,Marges,Perioden,Wijken.en.buurtten,Regioaanduiding.Soort.regio..omschrijving.,Psychisch.toestand)

write.csv(filtered_gemeentesmentaldata2020,"important data/filtered_gemeentesmentaldata2020.csv")

#the same for the data of 2022
```

```

filtered_gemeentesmentaldata2022 = gemeentesmentaldata2022 %>%
  select(Leeftijd,Marges,Perioden,Wijken.en.buurtten,Regioaanduiding.Soort.regio..omschrijving.,Psychisch)

write.csv(filtered_gemeentesmentaldata2022,"important data/filtered_gemeentesmentaldata2022.csv")

#####
####data cleaning, get rid of the useless variables in unemployment data set####
#####

#to calculate unemployment% we need total population and the unemployed population. Therefore we can fi
fltrd_unemployment = gem_unemployment %>%
  filter(Arbeidskenmerken %in% c("Totaal ", "Arb.pos.: niet-werkzame bevolking"))

#this category was not useful for our analysis so to make the data set less complex we filtered this va
fltrd_unemployment = fltrd_unemployment %>%
  filter(Geregistreerd.werkzoekende.bij.UWV %in% "Totaal")

write.csv(fltrd_unemployment,"important data/fltrd_unemployment")

#####
####prepare mental and unemployment set so that we can merge them together in 1 data set####
#####

# unemployment data set arranged at Period
fltrd_unemployment = fltrd_unemployment %>%
  arrange(Perioden)

# variable name changed from regio.s to gemeente
fltrd_unemployment = fltrd_unemployment %>%
  rename(Gemeente = Regio.s)

#Period made as character so that we could combine the data sets
fltrd_unemployment$Perioden = as.character(fltrd_unemployment$Perioden)

#instead of no value, we made all missing values of 2020 NA
filtered_gemeentesmentaldata2020$Psychische.klachten.... = NA_real_

#combined the 2020 and 2022 mental data set
Mental2020_2022 = bind_rows(filtered_gemeentesmentaldata2020,filtered_gemeentesmentaldata2022)

#variable name changed from wijken en buurten to gemeente
Mental2020_2022 = Mental2020_2022 %>%
  rename(Gemeente = Wijken.en.buurten)

#Period made as character so thaaf we could combine the data sets
Mental2020_2022$Perioden = as.character(Mental2020_2022$Perioden)

#####
####
#####

totaal_unemp <- fltrd_unemployment %>% filter(Arbeidskenmerken == "Totaal ")
totaal_unemp$aantal_jongeren <- totaal_unemp$Jongeren..15.tot.27.jaar...aantal.

```

```

totaal_unemp$Jongeren..15.tot.27.jaar...aantal. <- NULL
totaal_unemp$X <- NULL
totaal_unemp$Arbeidskenmerken <- NULL
nietwerkzaam <- fltrd_unemployment %>% filter(Arbeidskenmerken != "Totaal ")
nietwerkzaam$nietwerkzame_jongeren <- nietwerkzaam$Jongeren..15.tot.27.jaar...aantal.
nietwerkzaam$Jongeren..15.tot.27.jaar...aantal. <- NULL
nietwerkzaam$X <- NULL
nietwerkzaam$Arbeidskenmerken <- NULL

unemployment <- full_join(totaal_unemp, nietwerkzaam, by = c("Onderwijsvolgend", "Uitkering", "Geregistreerd.werkzoekende.bij.UWV"))

write.csv(unemployment,"important data/unemployment.csv")

#####
###now merge the 2 data sets together###
#####

#merge the mental and unemployment data set
Combined_data = full_join(unemployment,Mental2020_2022, by = c("Gemeente","Perioden"))

write.csv(Combined_data,"important data/Combined_data.csv")

#####
###clean the combined data set so that all useless variables are filtered out###
#####

#clean up0 all useless variables so that the data set becomes more clear
Data_Cleancombined = Combined_data %>%
  select(-Onderwijsvolgend, -Uitkering, -Geregistreerd.werkzoekende.bij.UWV, -Marges, -Leeftijd)

#change the variable names to easier names
names(Data_Cleancombined) = c(
  "Periode",
  "Gemeente",
  "Aantal_jongeren",
  "Niet_werkzame_jongeren",
  "Regio_type",
  "Psychische_klachten",
  "Hoog_risico_angst_depressie"
)

write.csv(Data_Cleancombined,"important data/Data_Cleancombined.csv")

```

Please use a separate 'R block' of code for each type of cleaning. So, e.g. one for missing values, a new one for removing unnecessary variables etc.

3.2 Generate necessary variables

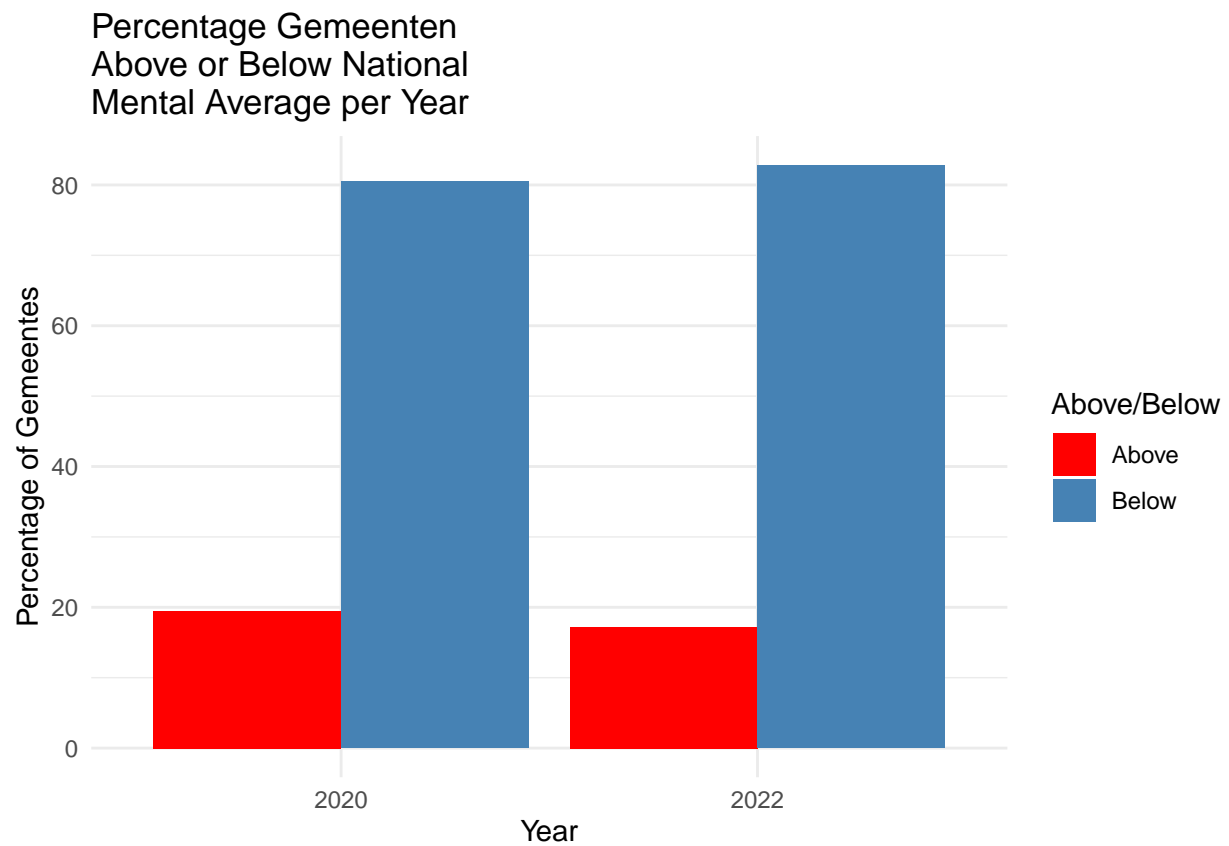
Variable 1: Unemployment percentage

```

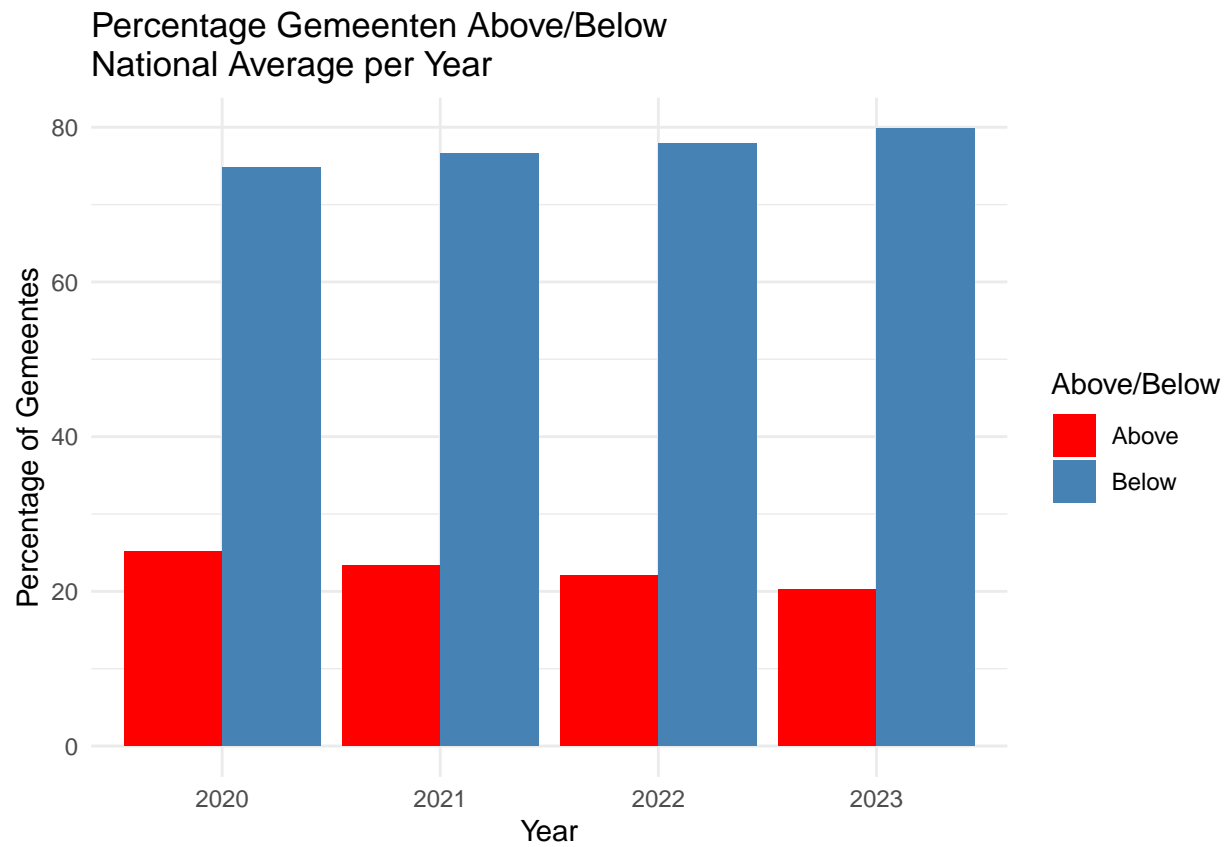
Data_Cleancombined$unemployment_percentage = (Data_Cleancombined$Niet_werkzame_jongeren/Data_Cleancombined$Aantal_jongeren)*100
Data_Cleancombined$unemployment_percentage = round(Data_Cleancombined$unemployment_percentage,2)

```

Variable 2: Unemployment change per year

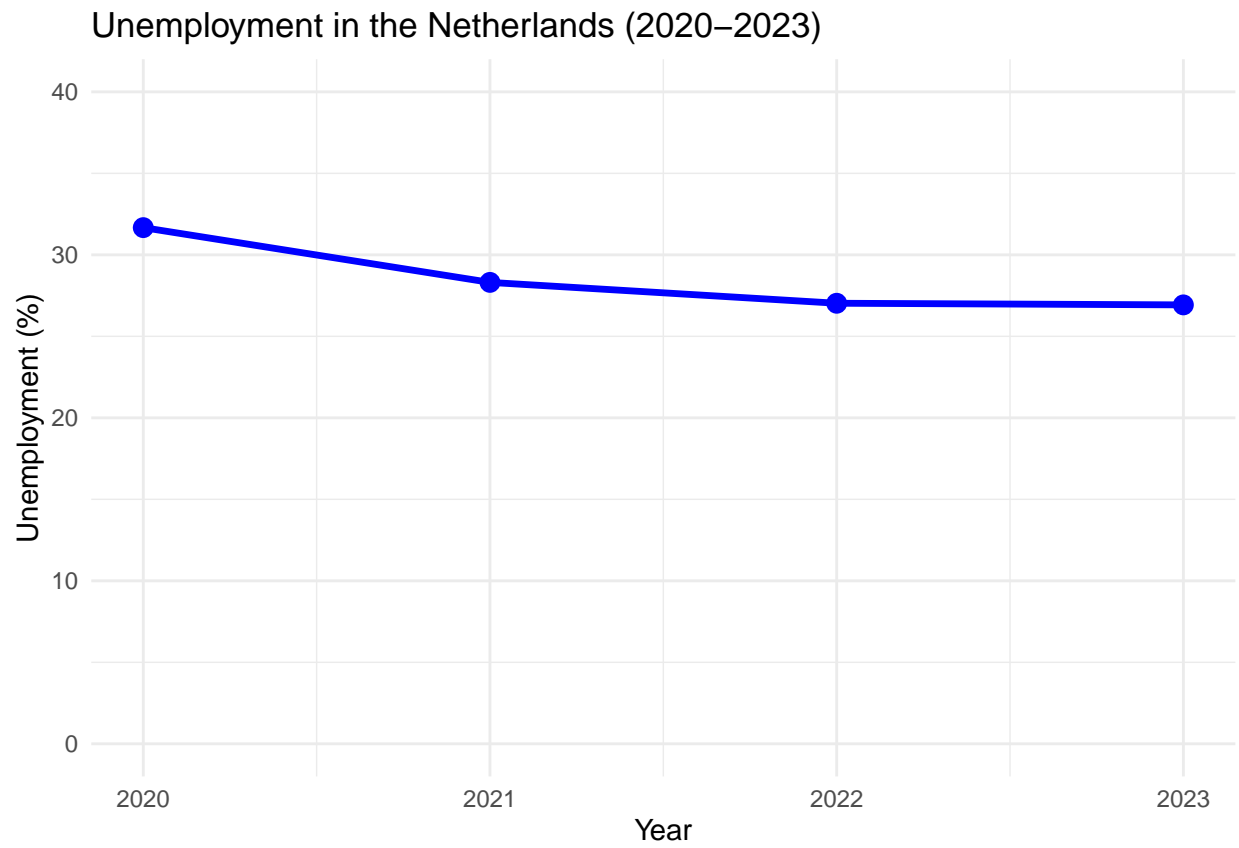


Variable 3: National Unemployment Averages of 2020-2023

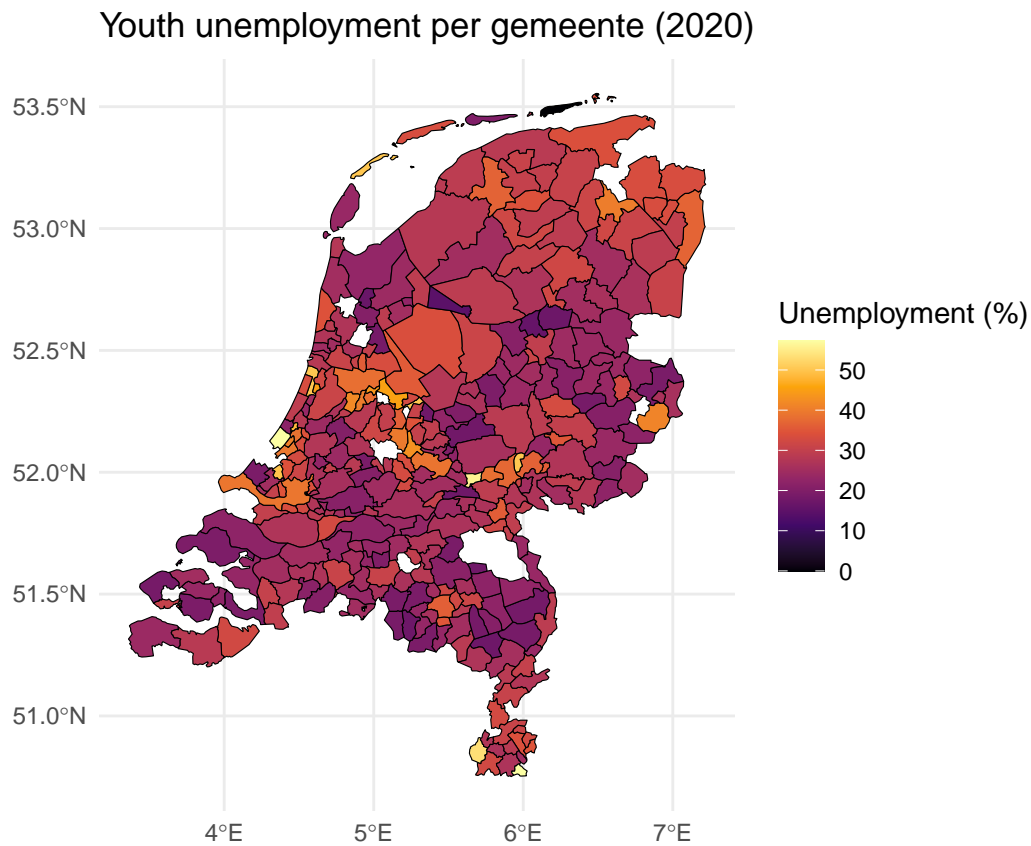


Variable 4: Unemployment Percentage Change

3.3 Visualize temporal variation

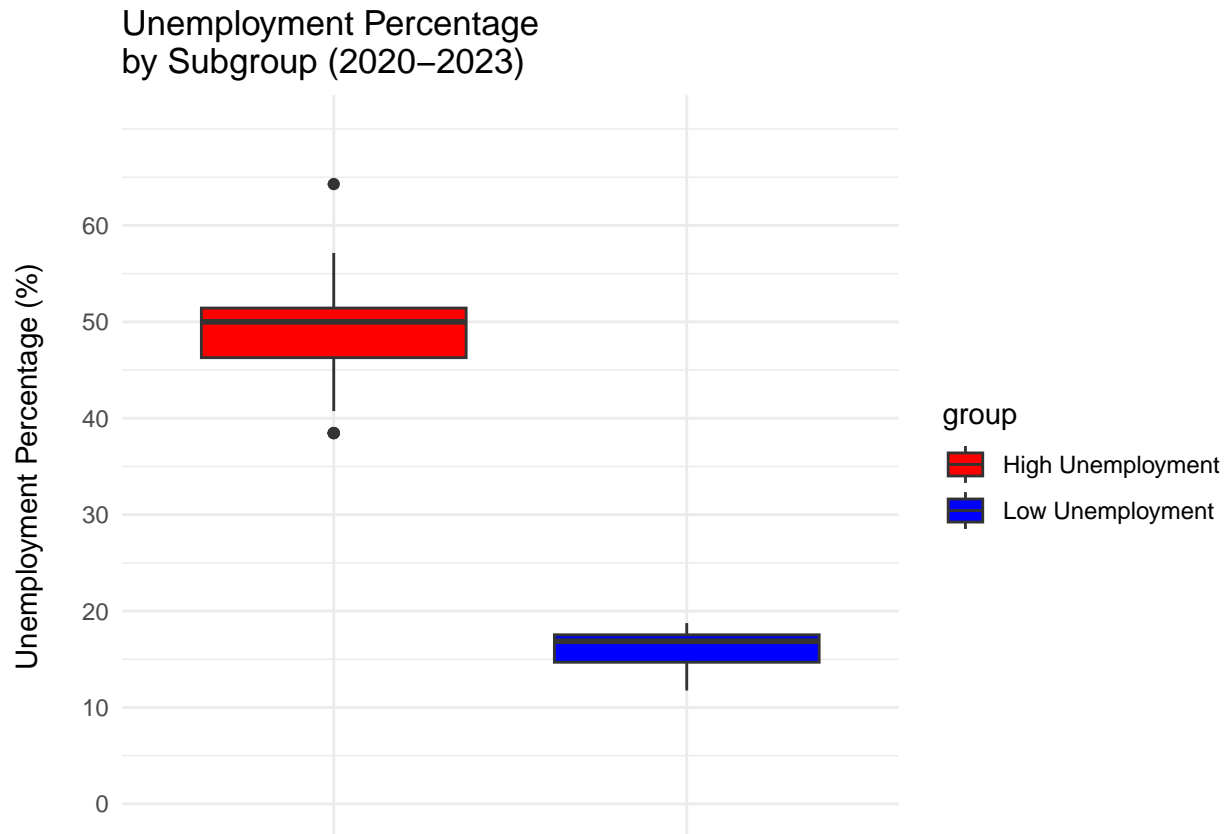


3.4 Visualize spatial variation



The heat map above shows the corresponding youth unemployment percentage per dutch municipality in the year 2020 from 0% to 50%. It visualizes the spatial differences the country is facing regarding this issue, helping the overall understanding of the multitude of factors, that may have an impact on the social problem overall. The contrasting data can be explained by the very low population in some areas, that then has a stronger impact on the overall percentage.

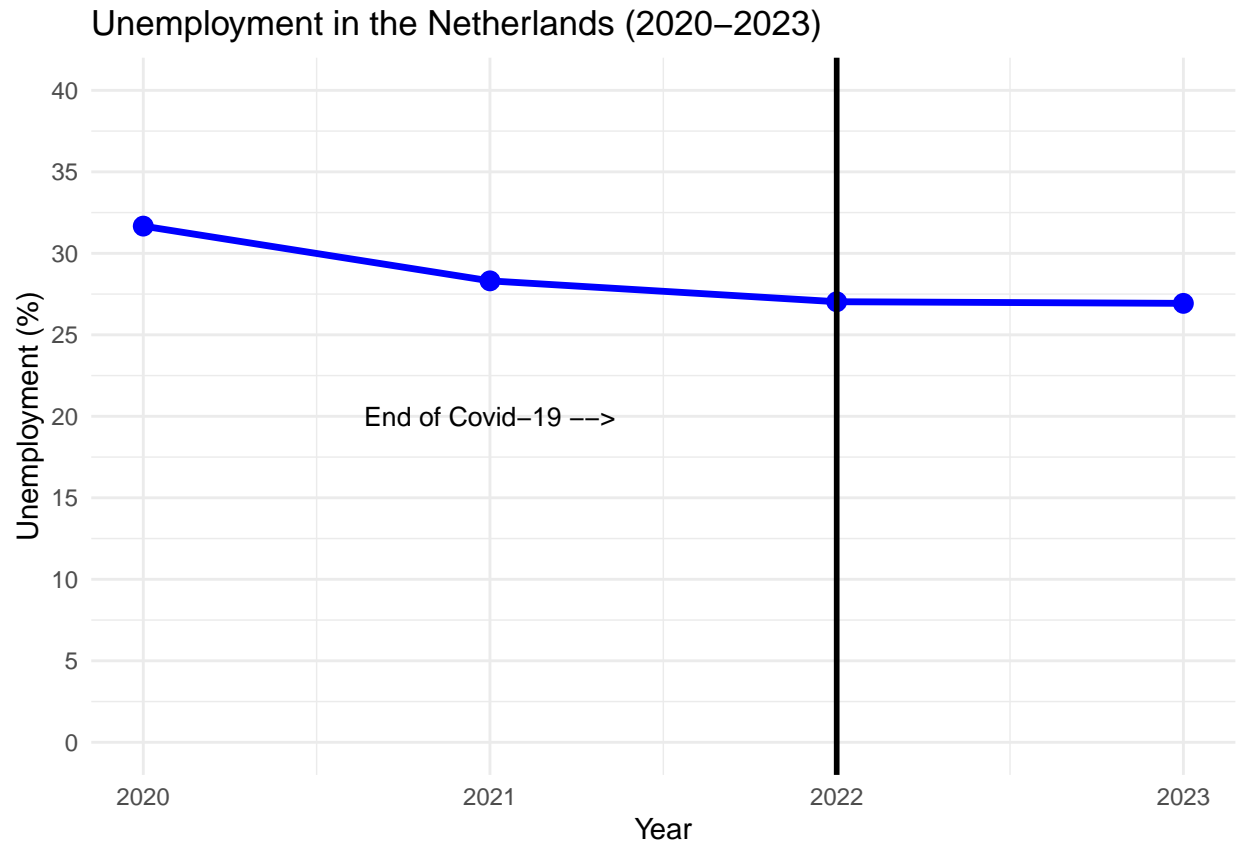
3.5 Visualize sub-population variation



Firstly, the gemeente unemployment data from 2020-2023 was compared, to define the top 10 gemeentes with the lowest and highest percentages of unemployment. The gemeentes Vlieland and Schiermonnikoog were then excluded, due to being islands with low levels of population, which were therefore giving the wrong idea of the overall percentages, since they had zero unemployment in some years. Finally, the results were compiled into the box plot, to visualise the findings. This is relevant to mental health and unemployment in the Netherlands, because it gives us an accurate visual of one of the variables.

3.6 Event analysis

Analyze the relationship between two variables.



```
names(Data_Cleancombined) = c(
  "Period",
  "Gemeente",
  "Total_Youth",
  "Total_Unemployed_Youth",
  "Region_Type",
  "MHI_5",
  "High_Risk_Anxiety_Depression",
  "Unemployment(%)",
  "Above_or_Below_National_Risk",
  "Above_or_Below_National_Unemployment",
  "Unemployment_Change_per_Year"
)
```

We decided to chose the end of corona as an event. As it is possible to see in the graph, unemployment decreased after 2022...

Part 4 - Discussion

4.1 Discuss your findings

Part 5 - Reproducibility

5.1 Github repository link

Provide the link to your PUBLIC repository here: ...

5.2 Reference list

Centraal Bureau voor de Statistiek. (2024). CBS Statline. Rijksinstituut voor Volksgezondheid en Milieu. Retrieved from: <https://statline.rivm.nl/#/RIVM/nl/dataset/50120NED/table>

Centraal Bureau voor de Statistiek. (2025). CBS Statline. Retrieved from: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/86087NED/table>

ElMorsy, M. (2024, January 23). The silent struggle: The impact of unemployment on mental health. Career Edge. Retrieved from: <https://www.careeredge.ca/the-silent-struggle-the-impact-of-unemployment-on-mental-health/>

Netherlands Enterprise Agency. (n.d.). Employment of minors and young adults. Business.gov.nl. Retrieved from: <https://business.gov.nl/regulation/employment-young-people/>

Ringdal, C., & Rootjes, F. (2022). Depression and labor supply: Evidence from the Netherlands. *Economics & Human Biology*, 45, 101103. Retrieved from: <https://doi.org/10.1016/j.ehb.2021.101103>