The relationship between unemployment and crime in the Netherlands

Romeo Schoonderbeek, 2868407 Paul Claessens, 2859603 Furkan Kadir Öztürk, 2852640

2025-06-24

Set-up your environment

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- 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 error
```

```
require(cbsodataR)
```

```
## Loading required package: cbsodataR
```

```
require(sf)
```

```
## Loading required package: sf
## Linking to GEOS 3.13.0, GDAL 3.5.3, PROJ 9.5.1; sf_use_s2() is TRUE
```

Part 1 - Identify a Social Problem

1.1 Describe the Social Problem

Our Social problem is to look at the relation between crime and unemployment, since these are often believed to be related with each other. Unemployment can cause financial stress and social exclusion, these factors can lead to criminal activity. Criminal records can also lower the chances of being accepted for a job, which makes criminal activity an easier option. Why this is a relevant issue, even though crime rates in

the Netherlands have seen a steady decline. Financial and online crime are relatively growing (Nieuws Sociaal en Groen, 2023). Whilst youth unemployment remains a prominent issue in vulnerable communities. Understanding the relationship between unemployment and crime is highly important for the design of social and economic policies. A study by Verbruggen(2014), reviewed by Fischer(2015) shows a critical perspective on the relationship between crime and unemployment. It shows that employment is generally associated with lower crime rates, however it also shows that unemployment doesn't necessarily lead to more crime. Whilst Verbruggen (2014) gives a good insight into the vulnerable Dutch population, broader analyses of this problem are still very limited. How structural shifts affect this relationship for example. Additionally, the understanding of spatial and subgroup variations in the Netherlands are also limited. We aim to contribute towards the limited understanding of the special and subgroup variations.

Part 2 - Data Sourcing

2.1 Load in the data

```
# provincelevel
DS_Panel_Unemployment <- read.csv("data/raw_data/unemployment_data_transform.csv")
DS_Panel_Crime <- read.csv("data/raw_data/panel_data_crime.csv")
DS_Population <- read.csv("data/population_data.csv")
area_data_working <- read.csv("data/raw_data/area_data.csv")

# aggragate

agg_unemployment <- read.csv("data/raw_data/aggragate_total_unemployment.csv")
agg_crime <- read.csv("data/raw_data/aggragate_total_crime.csv")
agg_population <- read.csv("data/raw_data/population_data_raw.csv")

# subpopulation

criminaliteit_per_leeftijd <- read.csv("data/raw_data/criminaliteit_per_leeftijd.csv")

unemployment_age <- read.csv("data/raw_data/unemployment_per_age.csv")
```

2.2 Provide a short summary of the dataset's

The data sets are categorized within 3 categories: province level, aggregate and sub population.

2.3 limitations

For crime statistics, Only crimes that are registered are accounted for, however there are also crimes that have not been reported and thus not counted in this dataset. The data does not distinguish if perpratrators were unemployed. So conclusions drawn from this data can never be entirely valid. Additionally, people reported as unemployed might do undeclared work.

Part 3 - Quantifying

3.1 Data cleaning

CBS datasets already come in panel data form. Data cleaning will include removing unnecessary rows and columns, adjusting data frames and merging data sets. 3 main data sets are used. Province panel data, aggregate panel data and data sets used for sub population analysis.

```
# province panel data cleaning

merged_data <- cbind(DS_Panel_Unemployment, DS_Panel_Crime [, 5])

merged_data <-cbind(merged_data, DS_Population [, 4])

merged_data <- merged_data [, -c(1)]
colnames(merged_data) <- c("Year", "province", "Unemployment_Rate", "total_Crimes", "Population")

merged_data <- mutate(merged_data, crimes_per_capita = (total_Crimes/Population)*100)

write.csv(merged_data, "data/working_data/merged_panel_data_final.csv")

area_data_working <- read.csv("data/raw_data/area_data.csv")

colnames(area_data_working) = c("X", "province", "Area(KM2)")

area_data_working <- area_data_working[6:17 ,]

area_data_working[[3]]  = as.numeric(area_data_working[[3]])

area_data_working[, 3] = area_data_working[, 3]/ 100

flevoland <- data.frame(name = "Flevoland", value = 1417 )
colnames(flevoland) = c("province", "Area(KM2)")

area_data_working <- bind_rows(
  area_data_working[1:5, ], flevoland,
  area_data_working[6:nrow(area_data_working), ]
)
```

```
## New names:
## New names:
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
```

```
area_data_working <- area_data_working %>%
  slice(rep(1:n(), each = 12))

merged_panel_data_final <- cbind(merged_data, area_data_working[, 3])

colnames(merged_panel_data_final) <- c("year", "province", "unemployment_rate", "total_crimes", "popula
```

```r
# aggregate data cleaning

agg_unemployment <- agg_unemployment[-c(51,52) ,]
agg_unemployment <- agg_unemployment[, -1]

agg_crime <- agg_crime[-82 , ]
agg_crime <- agg_crime[-c(1:31) ,]
agg_crime <- agg_crime[, -1]

colnames(agg_unemployment) = c("year", "unemployment")
colnames(agg_crime) = c("year", "total_crimes")

agg_merged <- cbind(agg_unemployment, agg_crime[, 2])

colnames(agg_merged) = c("year", "unemployment", "total_crime")

agg_merged <- agg_merged %>%
  mutate(total_crime = as.numeric(total_crime)) %>%
  mutate(year = as.numeric(year))

str(agg_merged)
```

```
## 'data.frame':    50 obs. of  3 variables:
##  $ year        : num  1975 1976 1977 1978 1979 ...
##  $ unemployment: num  3.5 4.2 4.3 4.2 4.1 4.3 5.7 8 9.6 9.6 ...
##  $ total_crime : num  486670 564025 590880 612370 668235 ...
```

```r
agg_population <-agg_population[, -4]
agg_population <- agg_population[, -1]
agg_population <- agg_population[-c(1:75) ,]

agg_merged <- cbind(agg_merged, agg_population[, 2])

colnames(agg_merged) = c("year", "unemployment", "total_crime", "population")

str(agg_merged)
```

```
## 'data.frame':    50 obs. of  4 variables:
##  $ year        : num  1975 1976 1977 1978 1979 ...
##  $ unemployment: num  3.5 4.2 4.3 4.2 4.1 4.3 5.7 8 9.6 9.6 ...
##  $ total_crime : num  486670 564025 590880 612370 668235 ...
##  $ population  : int  13600000 13730000 13810000 13900000 13990000 14090000 14210000 14290000 143400
```

```r
# subpopulation

crime2021 <- criminaliteit_per_leeftijd %>% filter(Perioden == "2021")

crime2021 <- crime2021 %>% filter(Geboorteland == "Totaal")

crime2021 <- crime2021[, -c(1:2, 4:8)]

colnames(crime2021) <- c("Age", "Crime")
```

```r
crime2021$Crime = (crime2021$Crime/10000)*100

unemployment_age <- read.csv("data/raw_data/unemployment_per_age.csv")

unemployment_age <- unemployment_age[-c(1:6, 19) ,]

unemployment_age <- unemployment_age[, -1]

colnames(unemployment_age) = c("Age", "Unemployment")

unemployment_age$Unemployment <- gsub(",", ".", unemployment_age$Unemployment)

unemployment_age$Unemployment <- as.numeric(unemployment_age$Unemployment)
```

3.2 Generate necessary variables

Variable 1 and 2 (Crimes per capita and population density)

```r
agg_merged <- agg_merged %>%
  mutate(crime_per_capita = (total_crime/population)*100)

merged_panel_data_final <- mutate(merged_panel_data_final, poplation_density = population/area_km2)
```

Variable 3 Age categories

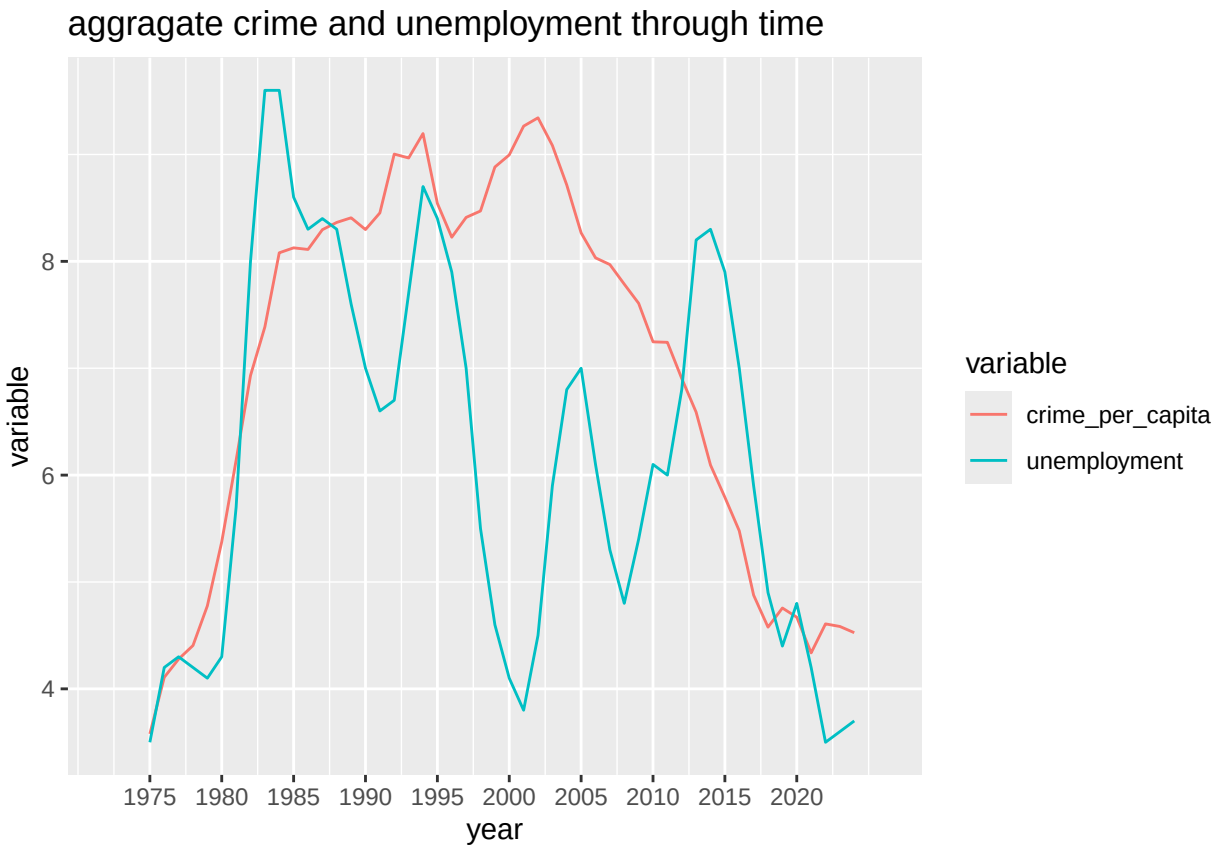
```r
crime2021$Age <- ifelse(crime2021$Age %in% c("12 tot 18 jaar", "18 tot 23 jaar"),
                                    "Youth (12-23)", "Adult+ (23+)")

unemployment_age$Age <- ifelse(unemployment_age$Age %in% c("15 tot 20 jaar", "20 tot 25 jaar"),
                               "Youth (15-25)", "Adult (25+)")
```

3.3 Visualize temporal variation

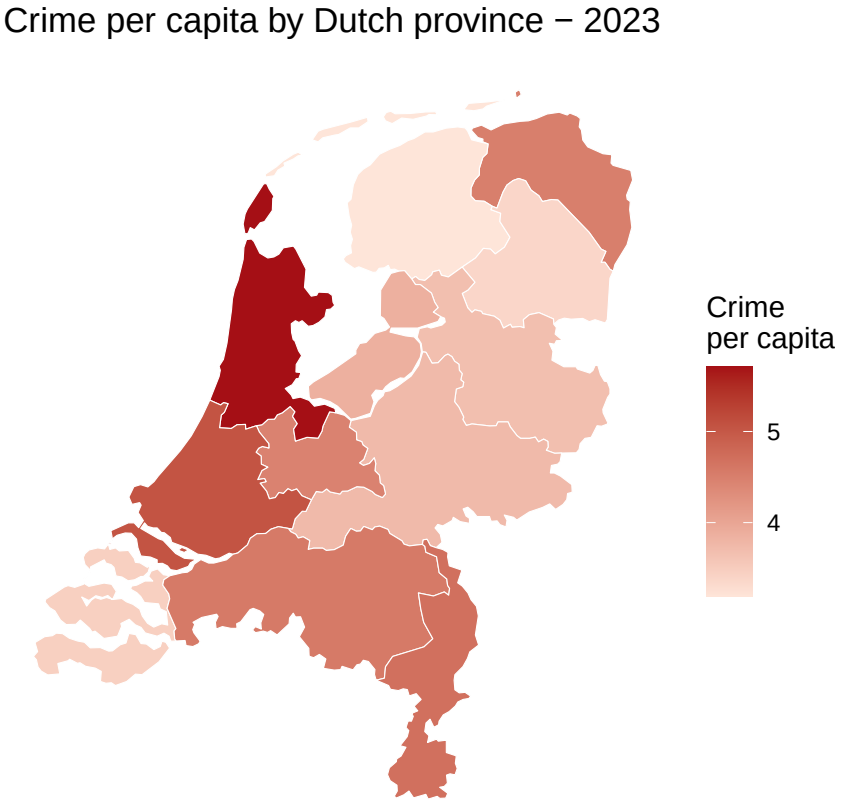
```r
pivot_agg_data <- agg_merged %>%
  pivot_longer(cols = c(unemployment, crime_per_capita), names_to = "variable", values_to = "value")

ggplot(pivot_agg_data, aes(x = year, y = value, colour = variable )) +
  geom_line() +
  labs(title = "aggragate crime and unemployment through time", x = "year", y = "variable") +
  scale_x_continuous(breaks = seq(1975, 2024, by = 5), limits = c(1972, 2026))
```

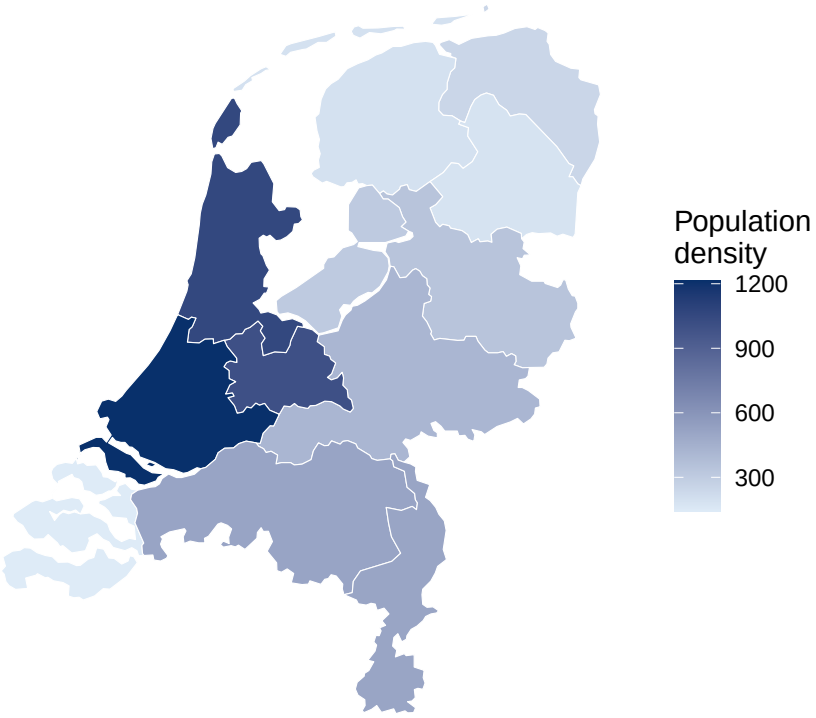


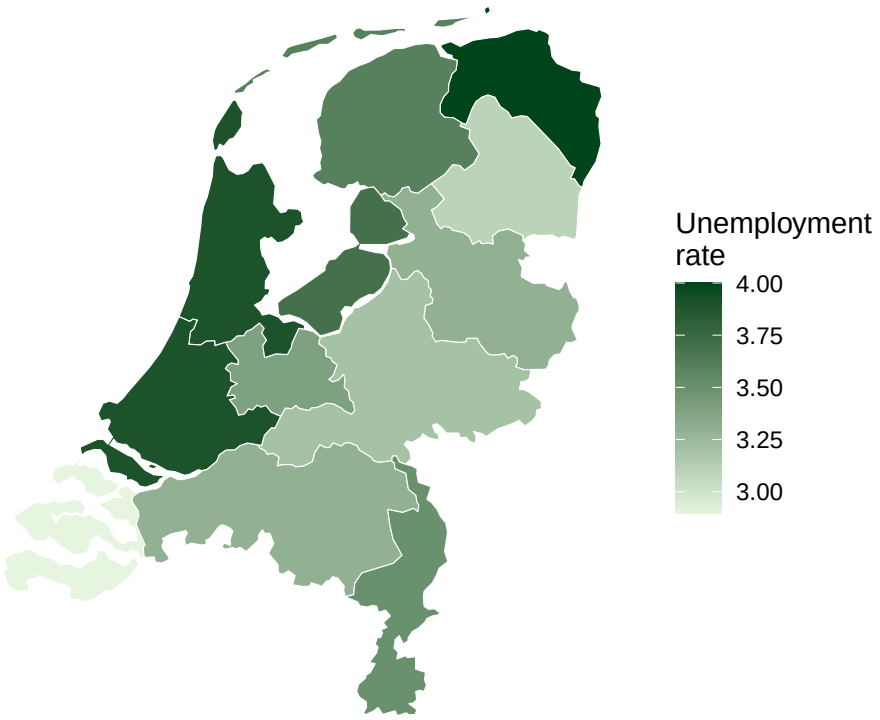
This time series can be seperated into two larger periods. Pre 2000 where crime seems to follow unemployment, and post 2000 where crime seems less corralated to unemployment, apart from certain periods. Surprisingly, unemployment seems to be lagging behind crime in certain periods.

3.4 Visualize spatial variation



Population density by Dutch province − 2023

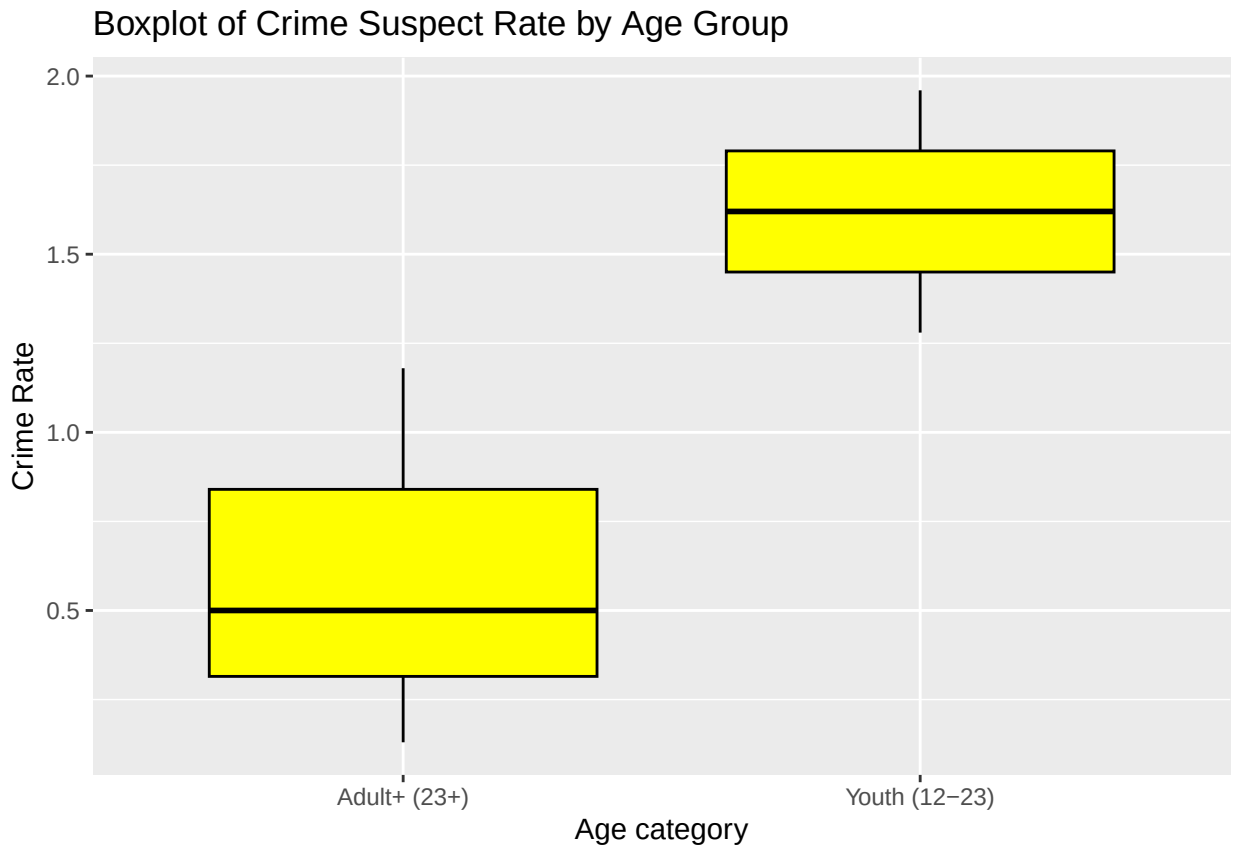




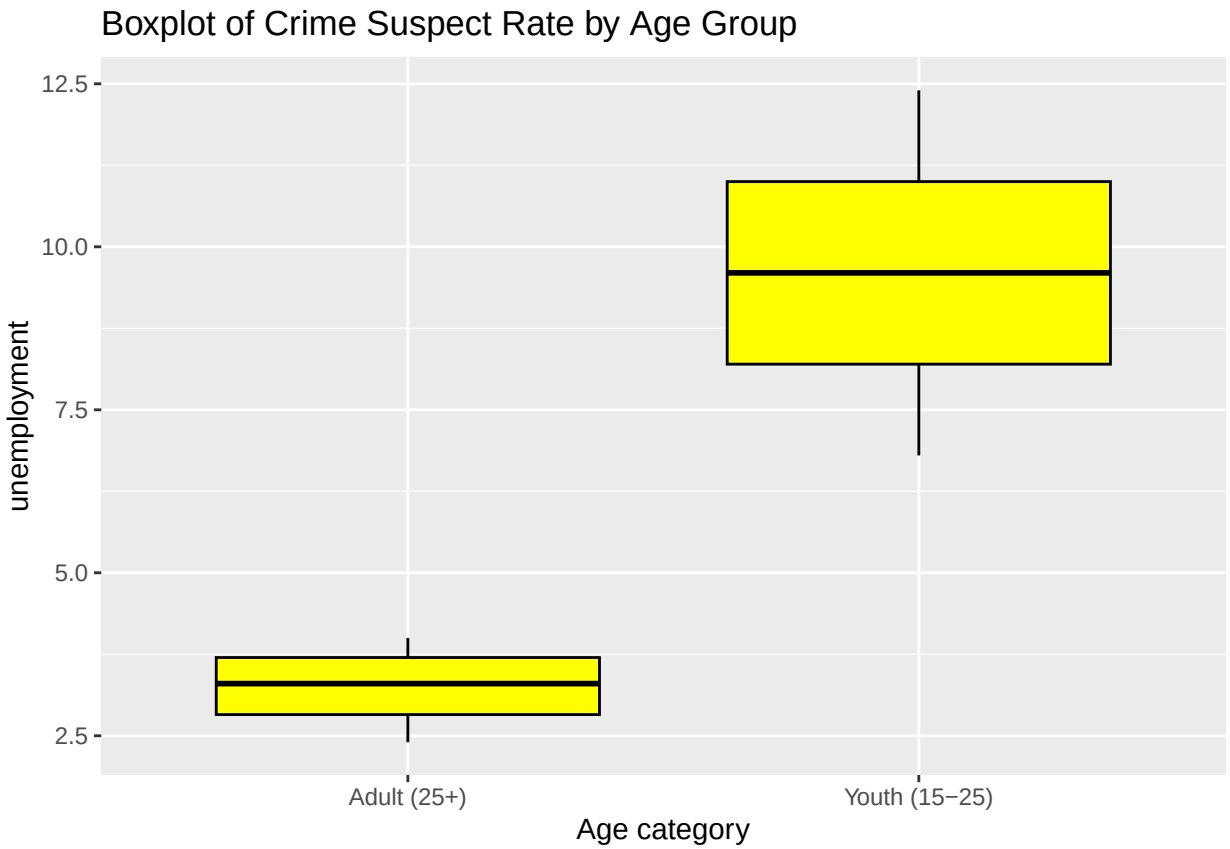
We can generally see the pattern that crime per capita follows unemployment. An anomaly is Groningen. Its crime unemployment is fairly high, yet crime per capita is low. Similarly, Utrecht's unemployment is low yet its crime per capita is somewhat higher. If population density is considered, than this pattern migh be better explained.Furthermore, Other factors can also affect cause these anomalies, which shows that the connection is not as straighforward. Individual cities and munciplaties were not plotted, which impacts our abilitiy to draw conclusions.

3.5 Visualize sub-population variation

```
ggplot(crime2021, aes(x = Age, y = Crime)) +
  geom_boxplot(fill = "yellow", color = "black") +
  labs(
    title = "Boxplot of Crime Suspect Rate by Age Group",
    x = "Age category",
    y = "Crime Rate"
  )
```




```
ggplot(unemployment_age, aes(x = Age, y = Unemployment)) +
  geom_boxplot(fill = "yellow", color = "black") +
  labs(
    title = "Boxplot of Crime Suspect Rate by Age Group",
    x = "Age category",
    y = "unemployment"
  )
```



In the boxplots above, it is clearly visible that unemployment and crime are linked. The Youth category is more vulnarable to unemployment than the adult category (Yeung & Yang, 2020), which in turn correlates with higher crime rates. Kessler et al. (2021) finds that reducing youth unemployment reduces youth crimes. Data limitation limit causality conclusions. Since the data does not distinguish if youth perpratrators were unemployed. Furthermore, youth might also be more prone to criminality regardless of employment status.

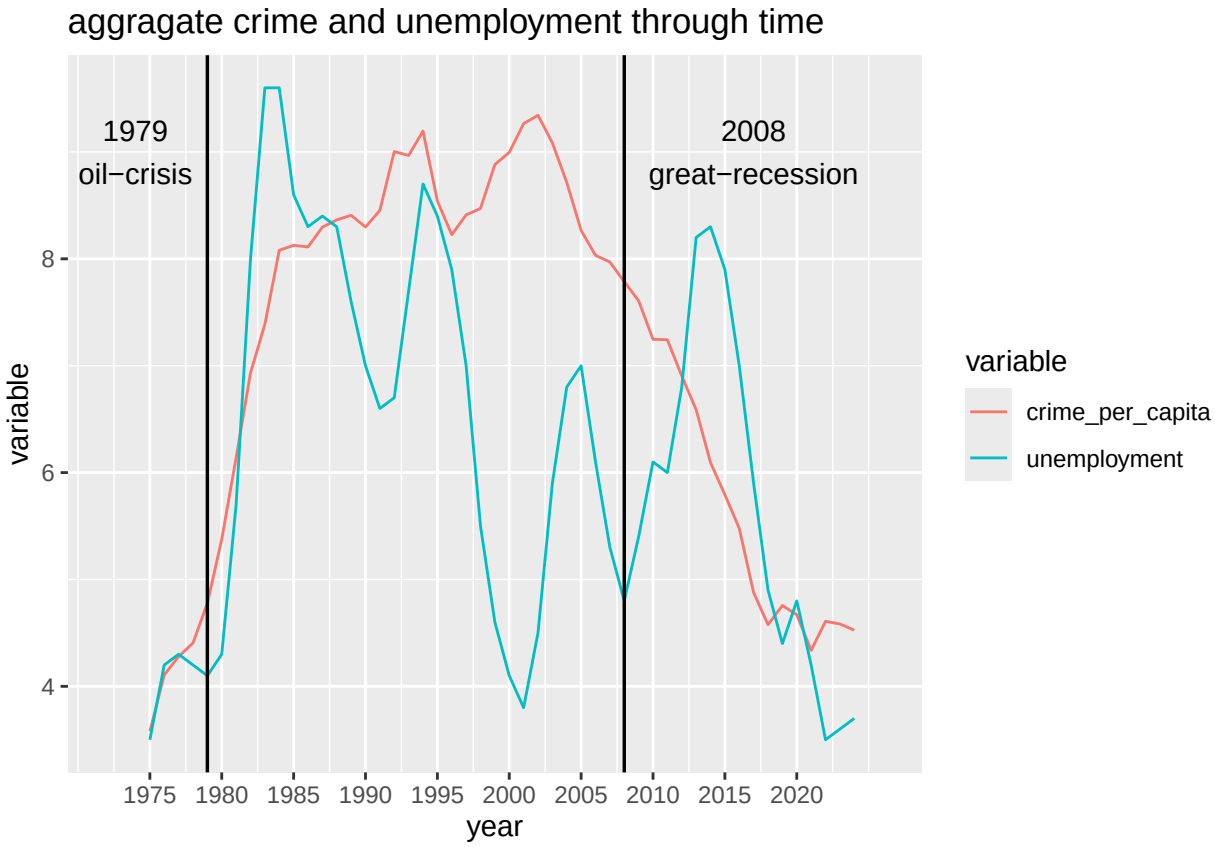
3.6 Event analysis

```
pivot_agg_data <- agg_merged %>%
  pivot_longer(cols = c(unemployment, crime_per_capita), names_to = "variable", values_to = "value")

ggplot(pivot_agg_data, aes(x = year, y = value, colour = variable )) +
  geom_line() +
  labs(title = "aggragate crime and unemployment through time", x = "year", y = "variable") +
  scale_x_continuous(breaks = seq(1975, 2024, by = 5), limits = c(1972, 2026)) +
  geom_vline(xintercept = c(1979, 2008),
             linetype = "solid",
             color = "black",
             size = 0.6) +
  annotate("text", x = 1974, y = 9, label = "1979\noil-crisis") +
  annotate("text", x = 2017, y = 9, label = "2008\ngreat-recession")
```

```
## 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.
```



Both 1979 oil-crises and the 2008 recession see large spikes in unemployment, yet only 1979 sees an increase in crimp per capita as well. A possible explanation for the movement of crime per capita in accordance with unemployment in 1980 to an inverse movement in 2008 could be the fact that the population is much younger in 1980 compared to 2008 and thus more people are in their peak-crime committing age 18-35 (Bezuinigen En Hervormingen in De Jaren '80, 2022). Another explanation is that the welfare safety net was much more robust in 2008 than in the 1980's which means that people who were unemployed where less inclined to do crime (Bezuinigen En Hervormingen in De Jaren '80, 2022). This shows that while unemployment may be linked in some cases, other factors play a significant role as well. Furthermore, trends in crime per capita predates trends in unemployment during both events, which negatively impacts the validity of our conclusions.

Part 4 - Discussion

4.1 Discuss your findings

The clearly shows that unemployment and crime are correlated, but also implicates that there is not a direct causality between unemployment and crime. Unemployment is likely a catalyst for crime. That is, unemployment is related to various social factors which might directly impact potency to commit crimes. These results can be of interest for policy makers, as not only battling unemployment helps prevent crime, but setting up other social programs might reduce blowbacks of unemployment. Allocating funds towards vulnarable groups, such as youth, and areas with a high population density will achieve the greatest results.

Part 5 - Reproducibility

5.1 Github repository link

Provide the link to your PUBLIC repository here:https://github.com/Autelius/Crime-and-Unemployment-Group-Project-FR-GIT.git

5.2 Reference list

Fischer, T. (2015). De relatie tussen werk en criminaliteit uitgeplozen. Tijdschrift voor Criminologie, 57(3), 334–338. https://www.boomportaal.nl/tijdschrift/TvC/TvC_0165-182X_2015_057_003_008

Nieuws Sociaal en Groen. (2023). Criminaliteit: de stand van zaken in Nederland. Retrieved from https://nieuwsociaalengroen.nl/criminaliteit/

Verbruggen, J. (2014). Previously institutionalized youths on the road to adulthood: A longitudinal study on employment and crime. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

Statline. (2025). Arbeidsdeelname; provincie, 2013–2024 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/85268NED/table?dl=BE1FB

Cbs. (2025). Geregistreerde criminaliteit; soort misdrijf, regio, 2010-2024 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/85268NED/table?dl=BE1FB

Statline. (2025). Bevolkingsontwikkeling; regio per maand, 2002-2024 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/37230ned/table

Statline. (2006). Volkstelling; oppervlakten, 1930 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/71118NED/table

Statline. (2025). Geregistreerde criminaliteit, 1948-2024 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/83723NED/table

Statline. (2025). Verdachten; geslacht, leeftijd, herkomst, opleiding, huishoudensinkomen, 2010-2024 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/85658NED/table

Statline. (2025). Arbeidsdeelname; kerncijfers, 2013-2024 [Data set]. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/85264NED/table?dl=BE1F8

Bezuinigen en hervormingen in de jaren '80. (2022, June 21). IsGeschiedenis. https://isgeschiedenis.nl/nieuws/bezuinigen-en-hervormingen-in-de-jaren-80