

# Assignment 1: Data and Descriptive, MSB104-gruppe 6 2025

Karin Liang

## Introduksjon

Denne oppgaven studerer regionale økonomiske forskjeller i Belgia (BE), Nederland (NL), Bulgaria (BG) og Norge (NO). Ved å bruke Eurostat-data om BNP og befolkning beregnes BNP per innbygger for å sammenligne regioner på tvers av land og over tiden fra 2000 til 2023.

## Del A: Subnasjonalt BNP og BNP per innbygger

Denne analysen bruker to regionale datasett fra Eurostat: GDP (nama\_10r\_3gdp) og population (demo\_r\_pjanagr3).

## Datainnsamling

```
# Lurt å gi denne label setup. Vil da automatisk
# bli kjørt før en chunk kjøres. Dette er dårlig dokumentert
# for quarto, men ser ut til å være arvet fra R Markdown
# documents
# pakker
library(dineq)
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':
```

```
filter, lag
```

```
The following objects are masked from 'package:base':
```

```
intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(stringr)
library(tidyverse)
```

```
— Attaching core tidyverse packages ————— tidyverse
2.0.0 —
```

✓ forcats	1.0.0	✓ readr	2.1.5
✓ lubridate	1.9.4	✓ tibble	3.3.0
✓ purrr	1.1.0	✓ tidyr	1.3.1

— Conflicts —  
tidyverse\_conflicts() —  
\* dplyr::filter() masks stats::filter()  
\* dplyr::lag() masks stats::lag()  
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

```
# install datapakken "eurostat"
# install.packages("eurostat")
library(eurostat)
```

## Laster ned datasettet GDP

```
# laster ned datasettet GDP (nama_10r_3gdp)
gdp <- get_eurostat("nama_10r_3gdp", time_format = "num")
```

```
indexed 0B in 0s, 0B/s
indexed 1.00TB in 0s, 131.76TB/s
```

```
Table nama_10r_3gdp cached at /var/folders/6z/
cg500fzs0w38623jddbr4c1c0000gn/T/RtmpoGRc7g/
eurostat/1b518072620676dc260f0a735a7f8094.rds
```

```
# velger tidperiod 2000-2023, unit i MIO_EUR (Million euro), og 4 land
("BE", "NL", "BG", "NO")
gdp <- gdp %>%
  filter(unit == "MIO_EUR") %>%
  mutate(TIME_PERIOD = as.numeric(TIME_PERIOD), geo = as.character(geo))
%>%
  filter(TIME_PERIOD >= 2000, TIME_PERIOD <= 2023) %>% #velger år 2000-2023
  filter(substr(geo, 1, 2) %in% c("BE", "NL", "BG", "NO")) %>% # velger 4
land
  filter(nchar(geo) == 5) # velger NUTS3

# viser BNP-datastruktur
glimpse(gdp)
```

```
Rows: 2,591
Columns: 5
$ freq      <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A", "A", "A",
```

```
"A"...
$ unit      <chr> "MIO_EUR", "MIO_EUR", "MIO_EUR", "MIO_EUR", "MIO_EUR",
"MI..."
$ geo       <chr> "BE100", "BE100", "BE100", "BE100", "BE100", "BE100",
"BE1..."
$ TIME_PERIOD <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
2012...
$ values    <dbl> 54681.25, 57619.35, 59500.55, 62125.09, 64604.73,
65375.89...
```

Datasettet gdp (nama\_10r\_3gdp) er hentet fra Eurostat og inneholder årlige verdier for BNP (bruttonasjonalprodukt) på NUTS3-nivå. Etter filtrering ifølge oppgaven består datasettet gdp av 2 591 observasjoner og 5 variabler fra 2000 til 2023 for landene Belgia (BE), Nederland (NL), BG Bulgaria (BG), Norge (NO).

### Laster ned datasettet population

```
# laster ned datasettet population (demo_r_pjanaggr3)
population <- get_eurostat("demo_r_pjanaggr3", time_format = "num")
```

```
indexed 0B in 0s, 0B/s
indexed 1.00TB in 0s, 36.24TB/s
```

```
Table demo_r_pjanaggr3 cached at /var/folders/6z/
cg500fzs0w38623jddbr4c1c0000gn/T//RtmpoGRc7g/
eurostat/960037bae2045dd1b6910fbf408cac7b.rds
```

```
# velger tidperiod 2000-2023, age Total population (Million euro), og 4
land ("BE", "NL", "BG", "NO")
population <- population %>%
  filter(sex == "T", age == "TOTAL") %>%
  mutate(TIME_PERIOD = as.numeric(TIME_PERIOD), geo = as.character(geo))
%>%
  filter(between(TIME_PERIOD, 2000, 2023)) %>% #velger år 2000-2023
  filter(substr(geo, 1, 2) %in% c("BE", "NL", "BG", "NO")) %>% # velger 4
land
  filter(nchar(geo) == 5)      # velger NUTS3

# viser population-datastruktur
glimpse(population)
```

```
Rows: 3,023
Columns: 7
$ freq      <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A", "A", "A",
"A"..."
$ unit      <chr> "NR", "NR", "NR", "NR", "NR", "NR", "NR", "NR", "NR",
```

```

"NR"...
$ sex      <chr> "T", "T", "T", "T", "T", "T", "T", "T", "T", "T", "T", "T",
"T"...
$ age      <chr> "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL",
"TOT"...
$ geo      <chr> "BE100", "BE100", "BE100", "BE100", "BE100", "BE100",
"BE1"...
$ TIME_PERIOD <dbl> 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,
2009...
$ values    <dbl> 959318, 964405, 978384, 992041, 999899, 1006749,
1018804, ...

```

Datasettet population (demo\_r\_pjanaggr3) er hentet fra Eurostat og viser befolkningen per 1. januar for hver NUTS3-region. Etter filtrering ifølge oppgaven består datasettet av 3 023 observasjoner og 7 variabler fra 2000 til 2023 for landene Belgia (BE), Nederland (NL), BG Bulgaria (BG), Norge (NO).

### Informasjon om dataene:

Begge datasettene er hentet fra Eurostat, og oppdateres årlig. Informasjonen om variablene bygger på metadata [1], [2]. Datadekningen for EFTA-land (inkludert Norge) starter i 2010. Derfor starter dataene for Norge i denne oppgaven fra 2010. BNP-datasettet (nama\_10r\_3gdp) ifølge metadatabeskrivelsen [1] er en del av Regional Economic Accounts og viser verdien av bruttonasjonalprodukt for de relevante europeiske regionene. Tallene er målt i millioner euro (MIO\_EUR), rapporteres årlig (freq = «A») på NUTS3-nivå, og dekker perioden 2000–2023. Datasettet inneholder fem variabler: freq, unit, geo, TIME\_PERIOD og values. Variablen geo representerer regionkoder, TIME\_PERIOD viser år, og values gir BNP-verdi i millioner euro. Dataene rapporteres av nasjonale statistiske institutter og samles inn og standardiseres i Eurostat. Datasettet gdp dekker nå regionale BNP-data for BE (Belgia), NL (Nederland), BG (Bulgaria), NO (Norge) som knyttet til oppgaven. Population (demo\_r\_pjanaggr3) ifølge metadatabeskrivelsen [2] kommer fra Eurostats Demography Database og viser befolkningen per 1. januar hvert år. Tallene måles i antall personer (NR) og er tilgjengelige etter kjønn og alder, men her brukes bare total befolkning (kjønn = «T», alder = «TOTAL»). Datasettet inneholder variablene freq, unit, sex, age, geo, TIME\_PERIOD og values, der values viser antall personer i hver region og år. Datasettet brukes til å beregne BNP per innbygger ved å koble befolkningsdataene med de regionale BNP-tallene. Dette muliggjør sammenligninger av økonomisk velstand mellom regioner og mellom landene Belgia (BE), Bulgaria (BG), Nederland (NL) og Norge (NO) i perioden 2000–2023.

## Beregning av BNP per innbygger (GDPC)

### Merger datasetter GDP og population, og fjerner manglende verdier

```

# Merger datasetter GDP og population
gdp_pop <- left_join(
  gdp,
  population,
  by = join_by(geo, TIME_PERIOD),
  suffix = c("_gdp", "_pop"))

```

```
# Teller antall manglende verdier (NA) i hver variabel
colSums(is.na(gdp_pop))
```

```
freq_gdp    unit_gdp      geo TIME_PERIOD  values_gdp  freq_pop
      0         0         0         0         0      462
unit_pop    sex         age  values_pop
      462      462      462      462
```

```
# Viser totalt antall manglende verdier i hele datasettet
sum(is.na(gdp_pop))
```

```
[1] 2310
```

```
# Dette steget brukes for å sjekke om datasettet har manglende verdier. De
to metodene viser det samme resultatet – 2 310 manglende verdier i
datasettet etter merged.
```

```
# Fjerner manglende verdier i BNP- og befolkningskolonnene før videre
beregninger
```

```
gdp_pop <- gdp_pop %>%
  drop_na(values_gdp, values_pop)
```

Samslåing av datasett ble utført her. Datasettet BNP og population ble slått sammen ved hjelp av regionkode (geo) og år (TIME\_PERIOD). Så sjekker og fjerner 2 310 manglende verdier i BNP- og befolkningskolonnene før videre beregninger.

### Beregner BNP per innbygger (GDPC) etter å merge dataene

```
# Beregner BNP per innbygger etter å merge dataene
gdp_pop <- gdp_pop %>%
  mutate(GDPC = (values_gdp * 1e6) / values_pop) %>% # BNP per innbygger
  (GDPC) er beregnet i euro per person
  filter(values_pop > 0, is.finite(GDPC)) # Fjerner ugyldige eller ekstreme
observasjoner
```

```
# Sjekker datastrukturen og verdiene etter sammenslåing og beregning
# Viser strukturen til datasettet etter beregningen
glimpse(gdp_pop)
```

```
Rows: 2,119
Columns: 11
$ freq_gdp    <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A", "A", "A",
"A"...
$ unit_gdp    <chr> "MIO_EUR", "MIO_EUR", "MIO_EUR", "MIO_EUR", "MIO_EUR",
"MI...
$ geo         <chr> "BE100", "BE100", "BE100", "BE100", "BE100", "BE100",
"BE1...
```

```

$ TIME_PERIOD <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
2012...
$ values_gdp <dbl> 54681.25, 57619.35, 59500.55, 62125.09, 64604.73,
65375.89...
$ freq_pop <chr> "A", "A", "A", "A", "A", "A", "A", "A", "A", "A", "A",
"A"...
$ unit_pop <chr> "NR", "NR", "NR", "NR", "NR", "NR", "NR", "NR", "NR", "NR",
"NR"...
$ sex <chr> "T", "T", "T", "T", "T", "T", "T", "T", "T", "T", "T",
"T"...
$ age <chr> "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL", "TOTAL",
"TOT...
$ values_pop <dbl> 992041, 999899, 1006749, 1018804, 1031215, 1048491,
106853...
$ GDPC <dbl> 55119.95, 57625.17, 59101.67, 60978.45, 62649.14,
62352.36...

```

```

# Oppsummerer sentrale statistiske mål for å kontrollere verdiene
summary(gdp_pop)

```

freq_gdp	unit_gdp	geo	TIME_PERIOD
Length:2119	Length:2119	Length:2119	Min. :2000
Class :character	Class :character	Class :character	1st Qu.:2008
Mode :character	Mode :character	Mode :character	Median :2013
			Mean :2013
			3rd Qu.:2018
			Max. :2023
values_gdp	freq_pop	unit_pop	
sex			
Min. : 168.8	Length:2119	Length:2119	
Length:2119			
1st Qu.: 1068.3	Class :character	Class :character	
Class :character			
Median : 3828.0	Mode :character	Mode :character	
Mode :character			
Mean :			
8393.3			
3rd Qu.:			
10733.5			
Max. :103285.0			
age	values_pop	GDPC	
Length:2119	Min. : 41906	Min. : 1087	
Class :character	1st Qu.: 125404	1st Qu.: 6061	
Mode :character	Median : 217969	Median : 25632	
	Mean : 283701	Mean : 24551	
	3rd Qu.: 365834	3rd Qu.: 35051	
	Max. :1296615	Max. :105721	

BNP per innbygger (GDPC) ble beregnet som total BNP delt på befolkningstall. Videre ble observasjoner med null eller ugyldige verdier ekskludert for å sikre et rent og pålitelig datasett for videre analyse. `glimpse(gdp_pop)` viser at datasettet inneholder 2 119 observasjoner og 11 variabler etter sammenslåing og rensing. Variablene inkluderer blant annet regionkode (geo), år (TIME\_PERIOD), BNP i millioner euro (values\_gdp), befolkning (values\_pop) og beregnet BNP per innbygger (GDPC). `summary(gdp_pop)` gir et overblikk over fordelingen i dataene. Tidsperioden dekker årene 2000–2023. BNP-verdiene varierer fra om lag 168 til 103 285 millioner euro, mens befolkningen i regionene varierer fra under 42 000 til over 1,29 millioner personer. Variabelen GDPC viser de beregnede verdiene for BNP per innbygger, med minst 1 087 og maks 105 721, som danner grunnlaget for den videre analysen.

## Beskrivende analyse

### Grunnleggende deskriptiv statistikk

```
# Grunnleggende deskriptiv statistikk for merged datasettet
gdp_pop_summary <- gdp_pop %>%
  filter(values_pop > 0, is.finite(GDPC)) %>%
  summarise(
    n_obs = n(),
    Min = min(GDPC, na.rm = TRUE),
    Q1 = quantile(GDPC, 0.25, na.rm = TRUE),
    Median = median(GDPC, na.rm = TRUE),
    Mean = mean(GDPC, na.rm = TRUE),
    Q3 = quantile(GDPC, 0.75, na.rm = TRUE),
    Max = max(GDPC, na.rm = TRUE),
    SD = sd(GDPC, na.rm = TRUE)
  )
gdp_pop_summary
```

```
# A tibble: 1 × 8
  n_obs   Min    Q1 Median   Mean    Q3    Max    SD
  <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  2119 1087. 6061. 25632. 24551. 35051. 105721. 17452.
```

Resultatene for grunnleggende deskriptiv statistikk for merged datasettet `gdp_pop` viser følgende nøkkeltall: antall observasjoner: 2 119 regioner, minimumsverdi 1 087 euro per person, første kvartil (Q1) 6 061 euro, median 25 632 euro, gjennomsnitt (Mean) 24 551 euro, tredje kvartil (Q3) 35 051 euro, maksimumsverdi 105 721 euro, og standardavvik: 17 452 euro.

```
bnp_per_innbygger <- gdp_pop %>%
  ggplot(aes(x = GDPC)) +
  geom_histogram(bins = 40, fill = "steelblue", color = "white") +
  labs(
    title = "Fordeling av BNP per innbygger (alle regioner)",
    x = "BNP per innbygger (euro per person)",
    y = "Antall regioner"
  )
bnp_per_innbygger
```

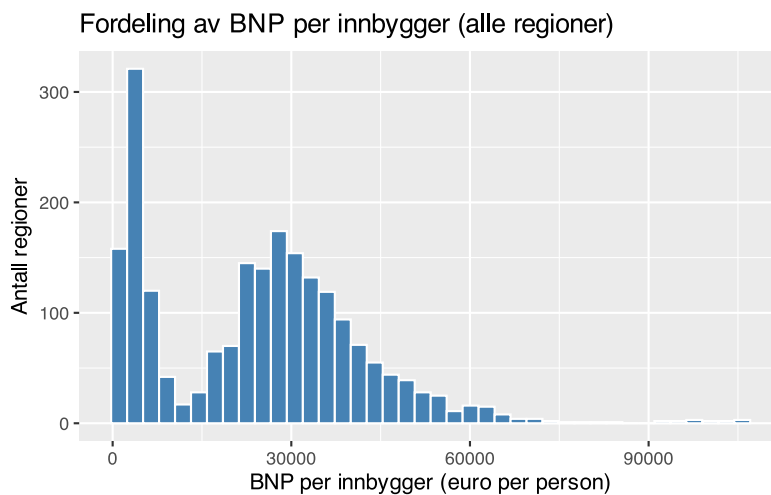


Figure 1: Fordeling av BNP per innbygger (alle regioner)

Figure 1 viser at halvparten av regionene har et BNP per innbygger mellom omtrent 5 000 og 40 000 euro, og at spredningen er ganske stor. Gjennomsnittet ligger litt lavere enn medianen, noe som tyder på at fordelingen er høyreskjev, slik figuren viser. Det betyr at de fleste regionene har relativt lavt BNP per innbygger, mens noen få regioner har svært høy verdiskaping.

### Deskriptiv statistikk for hvert land

```
# Deskriptiv statistikk for hvert land
gdp_pop_land_summary <- gdp_pop %>%
  filter(values_pop > 0, is.finite(GDPC)) %>%
  mutate(country = substr(geo, 1, 2)) %>%
  group_by(country) %>%
  summarise(
    n_obs = n(),
    Min = min(GDPC, na.rm = TRUE),
    Q1 = quantile(GDPC, 0.25, na.rm = TRUE),
    Median = median(GDPC, na.rm = TRUE),
    Mean = mean(GDPC, na.rm = TRUE),
    Q3 = quantile(GDPC, 0.75, na.rm = TRUE),
    Max = max(GDPC, na.rm = TRUE),
    SD = sd(GDPC, na.rm = TRUE)
  )
gdp_pop_land_summary
```

```
# A tibble: 4 × 9
  country n_obs   Min    Q1 Median   Mean    Q3   Max   SD
  <chr>   <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 BE       850 12756. 22665. 27979. 30263. 35751. 82418. 11092.
2 BG       672 1087.  2531.  3763.  4572.  5496. 31697.  3286.
3 NL       506 21694. 29189. 33612. 35413. 39395. 104524.  9812.
4 NO        91 37623. 48020. 53533. 58337. 60404. 105721. 16748.
```

Tabellen viser beskrivende statistikk for BNP per innbygger (i euro) i de fire landene Belgia (BE), Bulgaria (BG), Nederland (NL) og Norge (NO). Resultatene viser store forskjeller mellom



landene. Belgia har en moderat inntektsfordeling, der medianen ligger på rundt 28 000 euro per person. Maksimumsverdien er omtrent tre ganger enn median. Bulgaria har den laveste verdien blant de fire landene. Medianen er kun 3 763 euro, og både gjennomsnitt og maksimum er langt lavere enn i de andre landene. Maksimumsverdien er omtrent åtte ganger enn median. Nederland ligger betydelig høyere, med en median på omtrent 33 600 euro og et maksimum over 100 000 euro. Maksimumsverdien er omtrent tre ganger enn median. Norge har de høyeste verdiene, med en median på 53 533 euro og et gjennomsnitt på 58 337 euro. Maksimumsverdien er omtrent to ganger enn median.

```
bnp_per_innbygger_land <- gdp_pop %>%
  mutate(
    land_full = recode(substr(geo, 1, 2),
                        "BE" = "Belgia",
                        "BG" = "Bulgaria",
                        "NL" = "Nederland",
                        "NO" = "Norge")) %>%
  ggplot(aes(x = land_full, y = GDPC, fill = land_full)) +
  geom_boxplot() +
  labs(
    title = "BNP per innbygger etter land",
    x = "Land",
    y = "BNP per innbygger (euro)"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    legend.position = "none")
bnp_per_innbygger_land
```

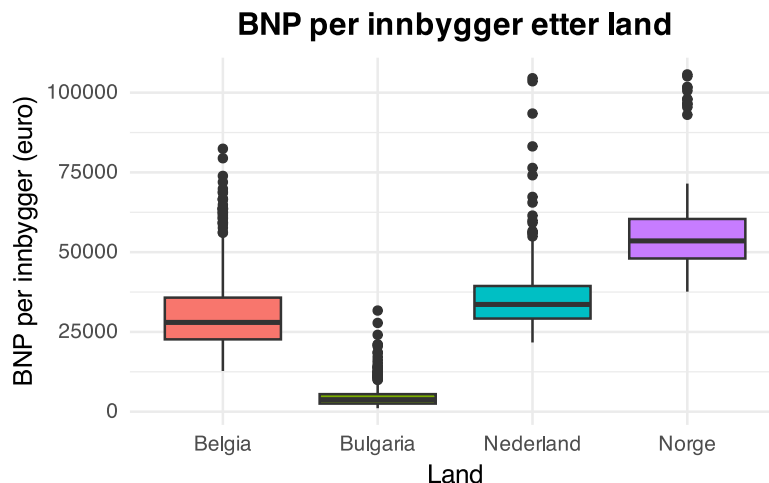


Figure 2: Sammenligning av BNP per innbygger etter land

Figure 2 gir et tydelig visuelt bilde av forskjellene mellom landene. Bulgaria skiller seg ut med svært lave verdier og liten variasjon mellom regionene, noe som indikerer et generelt lavt, men jevnt inntektsnivå. Belgia har et bredere spenn, med enkelte regioner som ligger betydelig over medianen, men ingen ekstreme avvik. Nederland viser både høyere nivåer og større spredning, spesielt enkelte rike regioner med høy BNP per innbygger. Norge har høyest

median og et bredt spenn, med flere regioner som ligger over 100 000 euro per person. Både tabellen og Figure 2 viser at Norge har den høyeste nivåen av BNP per innbygger. Nederland har nesten maks BNP per innbygger som Norge, men har lavere median og gjennomsnitt BNP per innbygger enn Norge. Bulgaria ligger lavest. Belgia befinner seg mellom disse, med moderate verdier og begrenset variasjon mellom regionene.

### Median BNP per innbygger etter land over tid

```
bnp_over_tid_median <- gdp_pop %>%
  mutate(land = substr(geo, 1, 2)) %>%
  group_by(land, TIME_PERIOD) %>%
  summarise(
    median_BNPI = median(GDPC, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(land_full = recode(land,
                             "BE" = "Belgia",
                             "BG" = "Bulgaria",
                             "NL" = "Nederland",
                             "NO" = "Norge")) %>%
  ggplot(aes(x = TIME_PERIOD, y = median_BNPI, color = land_full)) +
  geom_line(linewidth = 1.2) +
  scale_y_continuous(limits = c(0, 70000),
                     labels = scales::comma) +
  labs(
    title = "Utvikling i median BNP per innbygger (2000–2023)",
    x = "År",
    y = "BNP per innbygger (euro)",
    color = "Land"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    legend.title = element_text(face = "bold"),
    legend.position = "right"
  )
bnp_over_tid_median
```

### Utvikling i median BNP per innbygger (2000–2023)

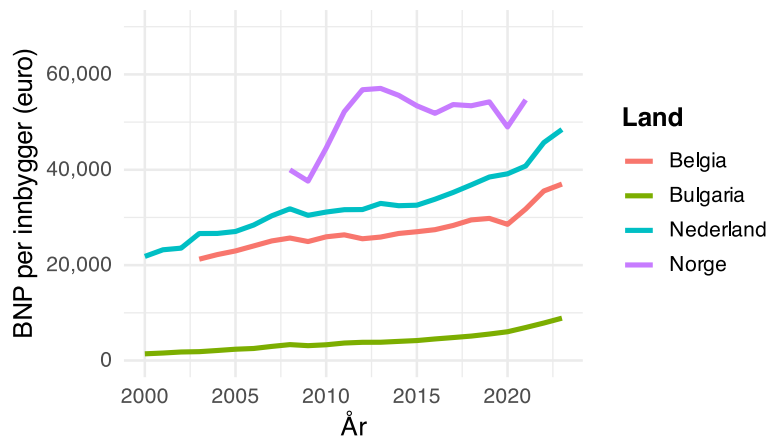


Figure 3: Utvikling i median BNP per innbygger (2000–2023)

Figure 3 er basert på median BNP per innbygger fra 2000 til 2023, og viser tydelig forskjeller i økonomisk utvikling mellom de fire landene: Norge ligger høyest gjennom hele perioden, med median BNP per innbygger mellom omtrent 40 000 og 60 000 euro. Nederland følger som nummer to, med tydelig jevn vekst fra rundt 20 000 til over 45 000 euro. Belgia ligger noe lavere, men viser også en stabil, gradvis økning som Nederland. Bulgaria starter svært lavt, men øker opp mot omtrent 10 000 euro i 2023. Median BNP i Norge har høyest nivå og størst variasjon, etterfulgt av Nederland, mens Belgia ligger midt i gruppen og Bulgaria betydelig lavere. Utviklingen over tid viser likevel positiv vekst i alle land. Resultatene til Figure 3 er i samsvar med resultatene til Figure 2 og resultatene av beskrivende statistikk for BNP per innbygger.

## Del B: Regional ulikhet

### Litteraturgjennomgang

Ifølge artikkelen 'Regional inequality, convergence, and its determinants– A view from outer space' C. Lessmann and A. Seidel [3] forsøker regional ulikhet, konvergens og dens determinanter, og finnes det en tydelig sammenheng mellom økonomisk utvikling og forskjeller mellom regioner. I artikkelen C. Lessmann and A. Seidel [3] bekrefter først forholdet mellom lysstyrkedata og regional inntekt, deretter bruker satellittbaserte nattlysdata for å måle økonomisk aktivitet i 180 land og 3166 regioner i perioden 1992–2012. Metoden bidrar å analysere regionale ulikheter også i land der det mangler pålitelige inntektsdata. Resultatene til C. Lessmann and A. Seidel [3] viser at omtrent 67–70 prosent av landene har opplevd sigma-konvergens, det vil si at inntektsforskjellene mellom regioner har blitt mindre over tid. Samtidig finner C. Lessmann and A. Seidel [3] at forholdet mellom økonomisk utvikling og ulikhet følger et N-formet mønster, der ulikheten øker i lavinntektsland, reduseres i mellominntektsland og stiger svakt igjen i høyninntektsland. Videre viser C. Lessmann and A. Seidel [3] at utvikling, mobilitet, åpenhet, ressurser, institusjoner, overføringer og utdanning, og etnisitet er syv grupper som kan påvirke ulike determinanter for regional ulikhet. Derimot kan geografiske faktorer over tid også ha en innvirkning for regional ulikhet. C. Lessmann and A. Seidel [3] peker på at regionale ulikheter ikke bare har økonomiske, men også politiske implikasjoner, fordi økonomisk aktivitet ofte er sterkt konsentrert i enkelte byområder og

ressursrike regioner. Dette kan påvirke både sosial stabilitet og utformingen av nasjonal politikk. Dermed gir studien et globalt rammeverk for å forstå hvorfor enkelte land opplever regional konvergens mens andre får økende ulikhet, og hvordan geografiske, økonomiske og institusjonelle forhold påvirker utviklingsmønsteret. Kort sagt fra M. Bandeira Morais, J. Swart, and J. A. Jordaan [4], Gini-koeffisienten er et av de mest brukte målene på økonomisk ulikhet, og varierer fra 0 (full likhet) til 1 (total ulikhet). Jo høyere Gini-koeffisienten indikerer jo større forskjeller i økonomisk aktivitet mellom ulike regioner.

## Beregning av Gini-koeffisienten på NUTS2-nivå

```
# Beregner befolkningsvektet Gini på NUTS2-nivå (basert på GDPC for NUTS3)
gini <- gdp_pop %>%
  mutate(
    NUTS2 = substr(geo, 1, 4), # Henter NUTS2-kode fra NUTS3 og legger til
    landkode
    country = substr(geo, 1, 2)
  ) %>%
  filter(values_pop > 0, is.finite(GDPC)) %>% # Fjerner ugyldige
  observasjoner (befolkning ≤ 0 eller manglende)
  group_by(country, NUTS2, TIME_PERIOD) %>%
  summarise(
    n_nuts3 = dplyr::n(), # Teller hvor mange NUTS3-regioner som inngår i
    hver NUTS2-gruppe
    Gini_weighted = if (dplyr::n() >= 2) {
      # Beregner befolkningsvektet Gini ved hjelp av funksjonen gini.wtd()
      as.numeric(dineq::gini.wtd(GDPC, weights = values_pop))
    } else {
      NA_real_
    }, # Hvis NUTS2 bare består av én NUTS3, Gini settes til NA
    .groups = "drop"
  )

# Sjekker struktur og summary
# Sjekker strukturen til Gini-resultatene
glimpse(gini)
```

```
Rows: 646
Columns: 5
$ country      <chr> "BE", "BE", "BE", "BE", "BE", "BE", "BE", "BE", "BE",
"B..."
$ NUTS2        <chr> "BE10", "BE10", "BE10", "BE10", "BE10", "BE10",
"BE10", ...
$ TIME_PERIOD  <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011,
20...
$ n_nuts3      <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,...
$ Gini_weighted <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
NA, ...
```

```
# Oppsummerer Gini-fordelingen
summary(gini$Gini_weighted)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.000948	0.038088	0.067481	0.076888	0.096178	0.282296	173

I denne delen beregnet jeg den befolkningsvektede Gini-koeffisienten på NUTS2-nivå basert på BNP per innbygger (GDPC) for NUTS3-regionene. Jeg hentet NUTS2- og landskoder fra NUTS3-koden og fjernet ugyldige observasjoner der befolkningen var  $\leq 0$  eller verdiene ikke var endelige. Deretter grupperte jeg dataene etter land, NUTS2 og år (TIME\_PERIOD), og brukte funksjonen `gini.wtd(GDPC, weights = values_pop)` fra pakken `dineq` til å beregne Gini-koeffisienten med befolkningsvekter. Beregningen av oppgaven bygger på flere antakelser og forenklinger. Det antas at NUTS3-regionene gir et representativt bilde av den økonomiske aktiviteten innenfor hver NUTS2-region, og at ulikheten mellom disse kan tolkes som intern regional ulikhet. Videre antas det at befolkningen er en egnet vekt, slik at regioner med flere innbyggere får større innflytelse på den samlede ulikheten. Observasjoner med ugyldige eller manglende verdier er fjernet, og for NUTS2-regioner som kun består av en NUTS3-region, ble Gini satt til NA som ikke brukes i beregningen, siden intern ulikhet ikke kan defineres i slike tilfeller. Resultatet er en tidsserie med Gini-verdier for hver NUTS2-region i perioden 2000–2023, som danner grunnlaget for videre analyse av utviklingen i regional økonomisk ulikhet. `glimpse(gini)` viser at datasettet med Gini-resultater inneholder fem variabler: land (country), regionkode (NUTS2), år (TIME\_PERIOD), antall NUTS3-regioner (n\_nuts3) og den beregnede Gini-koeffisienten (Gini\_weighted). `summary(gini$Gini_weighted)` viser at Gini-verdiene varierer mellom omtrent 0.001 og 0.28, med et gjennomsnitt på rundt 0.08, og 173 NA-verdier som skyldes NUTS2-regioner med bare én NUTS3. Dette bekrefter at beregningen fungerer som forventet, og at resultatene ligger innenfor et realistisk område for regional ulikhet.

## Datapresentasjon

### Statistikk Gini totalt og etter land

```
# statistikk totalt gini
gini_totalt <- gini %>%
  summarise(
    n_obs = sum(!is.na(Gini_weighted)),
    Min   = min(Gini_weighted, na.rm = TRUE),
    Q1    = quantile(Gini_weighted, 0.25, na.rm = TRUE),
    Median= median(Gini_weighted, na.rm = TRUE),
    Mean  = mean(Gini_weighted, na.rm = TRUE),
    Q3    = quantile(Gini_weighted, 0.75, na.rm = TRUE),
    Max   = max(Gini_weighted, na.rm = TRUE),
    SD    = sd(Gini_weighted, na.rm = TRUE)
  )
gini_totalt
```

```
# A tibble: 1 × 8
  n_obs      Min      Q1 Median   Mean     Q3    Max     SD
  <int>    <dbl>  <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl>
1   473 0.000948 0.0381 0.0675 0.0769 0.0962 0.282 0.0496
```

```
# statistikk gini etter land
gini_etter_land <- gini %>%
  group_by(country) %>%
  summarise(
    n_obs = sum(!is.na(Gini_weighted)),
    Min   = min(Gini_weighted, na.rm = TRUE),
    Q1    = quantile(Gini_weighted, 0.25, na.rm = TRUE),
    Median= median(Gini_weighted, na.rm = TRUE),
    Mean  = mean(Gini_weighted, na.rm = TRUE),
    Q3    = quantile(Gini_weighted, 0.75, na.rm = TRUE),
    Max   = max(Gini_weighted, na.rm = TRUE),
    SD    = sd(Gini_weighted, na.rm = TRUE),
    .groups = "drop"
  )
gini_etter_land
```

```
# A tibble: 4 × 9
  country n_obs      Min      Q1 Median   Mean     Q3    Max     SD
  <chr>   <int>    <dbl>  <dbl> <dbl>  <dbl>  <dbl> <dbl> <dbl>
1 BE      171 0.00992 0.0625 0.0781 0.0783 0.0899 0.153 0.0297
2 BG      144 0.0458 0.0818 0.103 0.118 0.126 0.282 0.0565
3 NL      147 0.000948 0.0249 0.0334 0.0385 0.0452 0.117 0.0232
4 NO       11 0.0213 0.0292 0.0342 0.0330 0.0366 0.0447 0.00679
```

Totalt består datasettet av 473 observasjoner av Gini-koeffisient på NUTS2-nivå for Belgia, Nederland, Bulgaria og Norge mellom 2000 til 2023. Gini-verdiene varierer fra 0.0009 til 0.2823, med en median på 0.0675 og et gjennomsnitt på 0.0769, noe som tyder på moderat regional ulikhet. Videre ser på landene hver for seg, finner vi følgende nøkkeldata: Belgia har ca. minimum 0.010, median 0.078 og maksimum 0.153. Bulgaria har ca. minimum 0.046, median 0.103 og maksimum 0.282. Nederland har ca. minimum 0.001, median 0.033 og maksimum 0.117. Norge har ca. minimum 0.021, median 0.034 og maksimum 0.045. Disse standardverdiene viser at forskjellene i Gini-koeffisienten mellom regionene generelt er små i alle landene. Norge fremstår som det mest jevne landet, mens Bulgaria har tydeligere regionale forskjeller. Resultatene viser at ulikheten er størst i Bulgaria og minst i Norge og Nederland, mens Belgia ligger midt imellom. Dette betyr at økonomisk aktivitet i Nederland og Norge er jevnere fordelt mellom regionene, mens Bulgaria har større forskjeller i økonomisk aktivitet mellom ulike regionene.

### Gjennomsnittlig Gini per land per år

```
# Gjennomsnittlig Gini
gini_gjennomsnittlig <- gini %>%
  filter(!is.na(Gini_weighted)) %>%
```

```
group_by(country, TIME_PERIOD) %>%
summarise(mean_gini = mean(Gini_weighted), .groups = "drop") %>%
mutate(land_full = recode(country,
                           "BE" = "Belgia", "BG" = "Bulgaria",
                           "NL" = "Nederland", "NO" = "Norge"))

gini_gjennomsnittlig
```

```
# A tibble: 77 × 4
  country TIME_PERIOD mean_gini land_full
  <chr>      <dbl>      <dbl> <chr>
1 BE          2003      0.0757 Belgia
2 BE          2004      0.0794 Belgia
3 BE          2005      0.0790 Belgia
4 BE          2006      0.0797 Belgia
5 BE          2007      0.0798 Belgia
6 BE          2008      0.0800 Belgia
7 BE          2009      0.0712 Belgia
8 BE          2010      0.0736 Belgia
9 BE          2011      0.0720 Belgia
10 BE         2012      0.0784 Belgia
# i 67 more rows
```

```
# Høyeste og laveste gjennomsnittlig Gini
gini_max_min <- gini_gjennomsnittlig %>%
  group_by(country) %>%
  summarise(
    max_gini = max(mean_gini, na.rm = TRUE),
    year_max = TIME_PERIOD[which.max(mean_gini)],
    min_gini = min(mean_gini, na.rm = TRUE),
    year_min = TIME_PERIOD[which.min(mean_gini)])
gini_max_min
```

```
# A tibble: 4 × 5
  country max_gini year_max min_gini year_min
  <chr>      <dbl>      <dbl>      <dbl>      <dbl>
1 BE          0.0820      2016      0.0712      2009
2 BG          0.151       2022      0.0925      2002
3 NL          0.0454      2006      0.0305      2010
4 NO          0.0447      2020      0.0213      2017
```

Tabellene viser gjennomsnittlig Gini og de høyeste og laveste verdiene for gjennomsnittlig Gini per land og tilhørende år i perioden 2000–2023. Bulgaria har hatt den høyeste ulikheten blant landene, med maksimum 0.151 i 2022 og minimum 0.093 i 2002. Belgia viser små endringer over tid, med maksimum 0.082 i 2016 og minimum 0.071 i 2009. Nederland har hatt lave verdier, mellom 0.031 i 2010 og 0.045 i 2006, mens Norge har de laveste forskjellene, fra 0.021 i 2017 til 0.045 i 2020. Samlet viser resultatene at ulikheten har vært lav og stabil i de tre av fire landene, men med en tydelig økende tendens i Bulgaria de siste årene.

## Median Gini per land per år

```
# Median Gini per land per år
gini_median <- gini %>%
  filter(!is.na(Gini_weighted)) %>%
  group_by(country, TIME_PERIOD) %>%
  summarise(median_gini = median(Gini_weighted), .groups = "drop") %>%
  mutate(land_full = recode(country,
                             "BE" = "Belgia", "BG" = "Bulgaria",
                             "NL" = "Nederland", "NO" = "Norge"))

gini_median
```

```
# A tibble: 77 × 4
  country TIME_PERIOD median_gini land_full
  <chr>      <dbl>      <dbl> <chr>
1 BE          2003      0.0782 Belgia
2 BE          2004      0.0812 Belgia
3 BE          2005      0.0763 Belgia
4 BE          2006      0.0731 Belgia
5 BE          2007      0.0755 Belgia
6 BE          2008      0.0798 Belgia
7 BE          2009      0.0661 Belgia
8 BE          2010      0.0650 Belgia
9 BE          2011      0.0657 Belgia
10 BE         2012      0.0806 Belgia
# i 67 more rows
```

```
# Høyeste og laveste median Gini per land (med år)
gini_median_max_min <- gini_median %>%
  group_by(country) %>%
  summarise(
    max_median_gini = max(median_gini, na.rm = TRUE),
    year_max = TIME_PERIOD[which.max(median_gini)],
    min_median_gini = min(median_gini, na.rm = TRUE),
    year_min = TIME_PERIOD[which.min(median_gini)])
gini_median_max_min
```

```
# A tibble: 4 × 5
  country max_median_gini year_max min_median_gini year_min
  <chr>      <dbl>      <dbl>      <dbl>      <dbl>
1 BE          0.0844      2016          0.0650      2010
2 BG          0.130       2022          0.0842      2002
3 NL          0.0408      2019          0.0193      2010
4 NO          0.0447      2020          0.0213      2017
```

Tabellene viser median Gini og høyeste og laveste verdier for median Gini per land og tilhørende år i perioden 2000–2023. Tredene og resultatene er det nesten samme som gjennomsnittlig Gini per land per år. Bulgaria har hatt de høyeste verdiene, med maksimum 0.130 i 2022 og minimum 0.084 i 2002. Belgia har hatt moderate og stabile verdier mellom 0.065



(2010) og 0.084 (2016). Nederland og Norge har hatt de laveste forskjellene, med verdier fra henholdsvis 0.019–0.041 og 0.021–0.045.

## Visualisering

### Sammenligning gjennomsnitt og median Gini

```
figuren_gini_mean_median <- ggplot() +  
  # lager linjer for gjennomsnitt-verdier  
  geom_line(  
    data = gini_gjennomsnittlig,  
    aes(x = TIME_PERIOD, y = mean_gini, color = "Gjennomsnitt"),  
    linewidth = 1.1  
  ) +  
  # lager linjer for median-verdier  
  geom_line(  
    data = gini_median,  
    aes(x = TIME_PERIOD, y = median_gini, color = "Median"),  
    linewidth = 1.1,  
    linetype = "dashed"  
  ) +  
  facet_wrap(~ land_full, ncol = 2) +  
  scale_color_manual(  
    values = c("Gjennomsnitt" = "red", "Median" = "cyan")  
  ) +  
  labs(  
    title = "Utvikling i Gini: Gjennomsnitt og median",  
    x = "År",  
    y = "Gini",  
    color = ""  
  ) +  
  theme_minimal(base_size = 14) +  
  theme(  
    plot.title = element_text(hjust = 0.5, face = "bold"),  
    legend.position = "bottom",  
    legend.text = element_text(size = 12)  
  )  
figuren_gini_mean_median
```

## Utvikling i Gini: Gjennomsnitt og median

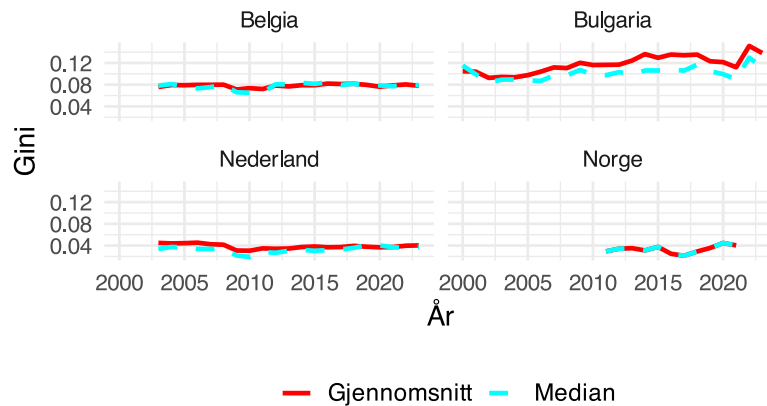


Figure 4: Gjennomsnittlig og median Gini (NUTS2) per land, 2000–2023

Figure 4 viser at gjennomsnittet og medianen Gini følger nesten samme utvikling over tid for. noe som indikerer at fordelingen av Gini-koeffisienten mellom regionene er forholdsvis symmetrisk, uten tydelige ekstremverdier. Bulgaria ligger klart høyest gini og har hatt en svak økning i ulikheten etter 2010, spesielt 2022. Belgia har et stabilt nivå uten store endringer. Nederland og Norge ligger lavt gjennom hele perioden, med små svingninger og generelt jevn regional fordeling. Norge viser den lavste Gini i 2017, og høyeste i 2020, både gjennomsnitt og median.

Resultatene til Figure 4 viser at ulikheten mellom regionene har vært mest stabil i Belgia og lavest i Nederland og Norge, spesielt Norge har den laveste ulikheten mellom de fire landene. Bulgaria har hatt den tydeligste økningen i gjennomsnitt Gini, betyr det at Bulgaria har økt ulikheten mellom regionale.

### Figuren for fordeling av Gini (NUTS2) etter land

```
figuren_gini_etter_land <- gini %>%
  filter(!is.na(Gini_weighted)) %>%
  mutate(land_full = recode(country,
                             "BE" = "Belgia", "BG" = "Bulgaria",
                             "NL" = "Nederland", "NO" = "Norge")) %>%
  ggplot(aes(land_full, Gini_weighted, fill = land_full)) +
  geom_boxplot(outlier.alpha = 0.5) +
  labs(title = "Fordeling av Gini (NUTS2) etter land",
       x = "Land", y = "Gini-koeffisient") +
  theme_minimal(base_size = 13) +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"),
        legend.position = "none")
figuren_gini_etter_land
```

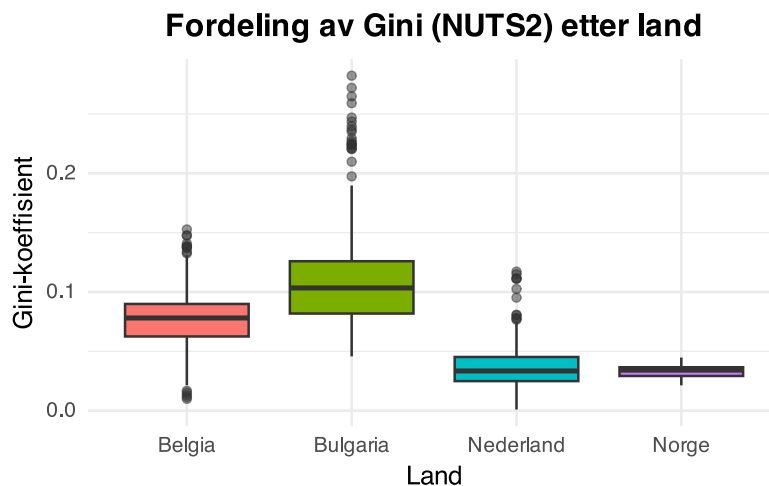


Figure 5: Fordeling av Gini-koeffisienter (NUTS2) etter land, 2000–2023

Figure 5 viser sprednings situasjonen til de Gini-koeffisientene for Belgia, Bulgaria, Nederland og Norge i perioden 2000–2023. Bulgaria har høyest nivå og størst variasjon, noe som viser store forskjeller i regional økonomisk aktivitet. Belgia ligger på et middels nivå med moderat spredning. Nederland har lavere verdier og mer jevn fordeling, mens Norge har de laveste verdiene og minst variasjon mellom regionene. Sprednings situasjonen til Figure 5 viser stemmer overens med resultatene til både `?@fig-gini_gjennomsnittlig` og `?@fig-gini_median`, der Bulgaria har størst ulikhet og Norge og Nederland viser den mest balanserte regionale fordelingen.

#### Figuren for Gini per NUTS2 over tid (fasettert etter land)”

```
figuren_gini_over_tid <- gini %>%
  filter(!is.na(Gini_weighted)) %>% # Fjerner rader med manglende
  Gini (NA)
  ggplot(aes(TIME_PERIOD, Gini_weighted, group = NUTS2)) + # Setter opp
  tidsserie per NUTS2-region
  geom_line(alpha = 0.25) + # Tegner tynne linjer for hver
  NUTS2-region (gjennomsiktighet 25%)
  facet_wrap(~ country, ncol = 2, # Lager separate paneler
  (fasetter) for hvert land
    labeller = as_labeller(c(
      BE = "Belgia",
      BG = "Bulgaria",
      NL = "Nederland",
      NO = "Norge")))) +
  labs(title = "Gini per NUTS2 over tid (fasettert etter land)", # Tittel
  og aksetitler
    x = "År",
    y = "Gini-koeffisient") +
  theme_minimal(base_size = 12) +
  theme(plot.title = element_text(hjust = 0.5, face = "bold")) #
  Midtstiller og gjør tittelen fet
figuren_gini_over_tid
```

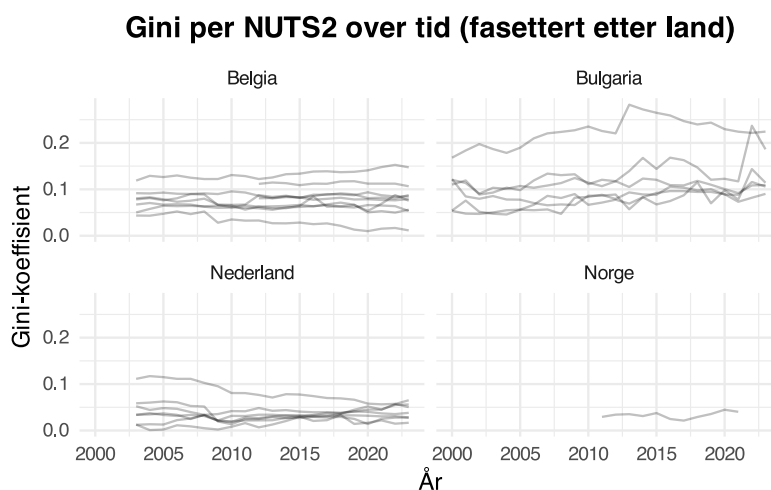


Figure 6: Utvikling i Gini per NUTS2 over tid (2000–2023), fasettert etter land

Figure 6 viser utviklingen i Gini-koeffisienten for hver NUTS2-region i Belgia, Bulgaria, Nederland og Norge i perioden 2000–2023. Hver linje representerer en region, slik at variasjonen innad i landene kan observeres. For Norge vises kun en linje, fordi flere norske NUTS2-regioner består av bare en NUTS3-region, dermed ikke får beregnet intern Gini-verdi. I tillegg starter de norske dataene først i 2010, nevnte i informasjon om dataene. Bulgaria har høyest over 0.2 Gini, viser størst intern ulikhet. Belgia har Gini med relevant stabilte mellom 0.05-0.1 mesteparten av tiden. Mens Norge og Nederland har nesten lavere enn 0.1 Gini, jevn og stabil regional fordeling. Resultatene til Figure 6 samsvarer med Figure 5, `?@fig-gini_gjennomsnittlig` og `?@fig-gini_median`.

## Diskusjon

### Analyse og sammenligning av Gini-resultater

Basert på analysen av Gini-koeffisientene kan man se at det finnes betydelige forskjeller i regional økonomisk ulikhet mellom Belgia, Bulgaria, Nederland og Norge i perioden 2000–2023. Bulgaria har de høyeste Gini-verdiene, Nederland og Norge de laveste, mens Belgia ligger mellom disse. Dette mønsteret er konsistent på tvers av resultatene for Gini totalt, Gini etter land, gjennomsnittlig Gini og median Gini, noe som viser at analysen har høy stabilitet og konsistens.

Når man ser på utviklingen over tid, har Bulgaria opplevd en tydelig økning i regional ulikhet etter 2010. Maksimumsverdien for Gini er rundt 0.28, noe som indikerer at de økonomiske forskjellene mellom regionene har økt og at den regionale utviklingen er lite balansert. Belgia har Gini-verdier som varierer mellom 0.07 og 0.15, og viser et generelt stabilt nivå, noe som tyder på moderate regionale forskjeller. Nederland har verdier mellom 0.03 og 0.11, med små variasjoner over tid, som indikerer en jevn og balansert regional fordeling. Norge har de laveste og mest stabile verdiene, mellom 0.02 og 0.05, noe som viser svært små økonomiske forskjeller mellom regionene og en balansert regional struktur. Som tidligere nevnt har de norske dataene en spesiell karakter, ettersom flere NUTS2-regioner består av kun én NUTS3-region. Dette fører til at figuren for Norge bare viser én Gini-linje, men dette påvirker ikke gyldigheten av konklusjonene.

Statistisk sett er Bulgarias maksimale Gini-verdi omtrent 0.28, betydelig høyere enn de andre landenes øvre grenser, mens Norges maksimale verdi er bare rundt 0.045, noe som skaper en tydelig kontrast. Gjennomsnitts- og medianverdiene er svært like i alle land, noe som viser at dataene er symmetrisk fordelt uten alvorlig skjevhet eller ekstreme verdier. Denne stabiliteten bekrefter ytterligere påliteligheten i beregningen av Gini-koeffisientene. Videre viser SD-verdiene at variasjonen mellom regionene generelt er lav, spesielt i Norge og Nederland, noe som indikerer en jevn regional struktur uten store avvik. Boksplottet støtter dette funnet, ettersom de fleste land har få eller ingen tydelige uteliggere, med unntak av Bulgaria, som viser enkelte regioner med markant høyere ulikhet enn resten.

Basert på betydningen av Gini-koeffisienten [4], der en verdi nær 1 indikerer høy ulikhet og en verdi nær 0 indikerer full likhet, kan vi se følgende situasjoner: Bulgaria har den høyeste Gini-verdien blant de fire landene, noe som viser at de regionale økonomiske forskjellene er størst og at ulikheten er mest markert. Belgia ligger på et middels nivå og viser en moderat grad av ulikhet. Nederland og Norge har de laveste Gini-verdiene, spesielt Norge, som viser den mest balanserte økonomiske fordelingen mellom regionene og det laveste nivået av ulikhet.

Analysen bekrefter at graden av regional ulikhet varierer betydelig mellom landene og følger et mønster som samsvarer med [3]. Resultatene viser at økonomisk utvikling og ulikhet henger tett sammen, der lavinntektsland som Bulgaria har høyere og økende Gini-verdier, mens høyninntektsland som Nederland og Norge har lavere og mer stabile nivåer. Dette støtter den N-formede sammenhengen mellom utvikling og ulikhet som [3] beskriver.

[5] peker på at Norge er et ressursrikt og velstående land, der staten har brukt inntektene fra naturressursene til å utvikle og stabilisere økonomien gjennom målrettet politikk og solide institusjoner. I lys av de bestemmende faktorene for regional ulikhet som fremheves av [3] som utvikling, mobilitet, åpenhet, ressurser, institusjoner, overføringer og utdanning, og etnisitet, kan man se at Norges økonomiske styrke og institusjonelle kvalitet bidrar til lav regional ulikhet. Dette gjenspeiles tydelig i landets lave Gini-koeffisient, som indikerer en balansert og inkluderende økonomisk utvikling.

## Refleksjon

Denne analysen viser at det i perioden 2000–2023 finnes tydelige forskjeller i regional økonomisk ulikhet mellom Belgia, Bulgaria, Nederland og Norge. I henhold til forklaringen av Gini-koeffisienten i [4] har Bulgaria de høyeste Gini-verdiene, som har økt etter 2010, noe som indikerer større ulikhet mellom regionene. Belgia har en middels Gini-verdi med små endringer over tid, noe som tyder på en stabil grad av regional ulikhet. Nederland og Norge har de laveste og mest stabile Gini-verdiene, noe som viser at ulikheten i disse landene er relativt liten. Som [3] påpeker, har lavinntektsland ofte høyere grad av ulikhet, mens høyninntektsland har lavere nivåer. De høye Gini-verdiene i Bulgaria og de lave verdiene i Norge bekrefter dette mønsteret. I tillegg påpeker [5] at Norge er et ressursrikt og institusjonelt velfungerende land som har oppnådd økonomisk stabilitet gjennom målrettet og effektiv politikk. Dette har bidratt til å opprettholde lave regionale forskjeller. Dette viser også at land med rikelige naturressurser og sterke institusjoner har bedre forutsetninger for å redusere regional ulikhet. Funnene bekrefter dermed at regional ulikhet henger tett sammen med de syv bestemmende faktorene som foreslås av [3], spesielt ressurser, utvikling og institusjoner. Samlet sett viser analysen at graden av regional ulikhet ikke bare reflekterer økonomisk utviklingsnivå, men også hvordan land forvalter sine ressurser og institusjoner. Effektiv ressursforvaltning og et

velfungerende politisk-institusjonelt system fremstår som avgjørende for å redusere regionale forskjeller, fremme bærekraftig utvikling og motvirke økende ulikhet.

## **Del C: Dokumenter bruken min av AI**

I denne A1-oppgaven brukte jeg OpenAI ChatGPT (GPT-5-modellen) som et støtteverktøy i arbeidsprosessen. Verktøyet ble brukt til å rette kodefeil, oversette språk, forstå fagbegreper, lese og oppsummere litteratur, tolke datainnhold samt forbedre struktur og språklig klarhet i teksten. Bidrar å styrke min forståelse av analysen, utvikle resonnementet og øke kvaliteten på den skriftlige framstillingen. I Del A brukte jeg ChatGPT til å forstå Eurostat-datasettet og metadata-beskrivelsene, inkludert forklaring av variabler, filtreringskriterier og tilpasning av kode. Bidrar å identifisere og rette feil i R-koden, slik at resultatene stemte bedre overens med analysens formål. I Del B brukte jeg ChatGPT til å forklare betydningen av Gini-koeffisienten og andre indikatorer, sammenligne verdier og trender, og formulere resultatene på et mer presist og naturlig norsk. Videre brukte jeg ChatGPT til å lese, forstå og oppsummere metadata informasjon, og relevant litteratur [3], samt til å finne tilleggsreferanser [4], [5]. Under skriveprosessen skrev jeg noen ganger først tekstavsnitt på mitt morsmål, og brukte deretter ChatGPT til å oversette og forbedre den norske versjonen slik at grammatikk og logikk ble mer sammenhengende. Etter hver bruk kontrollerte og redigerte jeg innholdet selv for å forhindre overredigering, og sikre at alle analyser og konklusjoner bygget på min egen forståelse og mine data.

## **Kilder**

### **Bibliography**

- [1] Eurostat, “Regional economic accounts by NUTS 3 regions (reg\_eco10).” 2025.
- [2] Eurostat, “Population on 1 January by broad age group, sex and NUTS 3 region (demo\_r\_pjanaggr3).” 2025.
- [3] C. Lessmann and A. Seidel, “Regional inequality, convergence, and its determinants – A view from outer space,” *European Economic Review*, vol. 92, pp. 110–132, 2017, doi: <https://doi.org/10.1016/j.eurocorev.2016.11.009>.
- [4] M. Bandeira Morais, J. Swart, and J. A. Jordaan, “Economic Complexity and Inequality: Does Regional Productive Structure Affect Income Inequality in Brazilian States?,” *Sustainability*, vol. 13, no. 2, 2021, doi: 10.3390/su13021006.
- [5] Ø. Cappelen and L. Mjøset, “Can Norway Be a Role Model for Natural Resource Abundant Countries?,” *Discussion Papers in Economics*, vol. 9, pp. 1–26, 2009, [Online]. Available at: [https://www.researchgate.net/publication/228983025\\_Can\\_Norway\\_Be\\_a\\_Role\\_Model\\_for\\_Natural\\_Resource\\_Abundant\\_Countries](https://www.researchgate.net/publication/228983025_Can_Norway_Be_a_Role_Model_for_Natural_Resource_Abundant_Countries)