

Assignment 1 — MSB104 — Group 3

Irjan & Magnus

Part A Sub-national GDP and GDP per capita

Data aquisition

```
# Henter inn populasjons datasett fra excell
Populasjon <- read_excel("DEMO_Ass1.xlsx", sheet = 2, col_types = "text") %>%
  clean_names()

# Henter inn regional BNP datasett fra excell
BNP <- read_excel("GDP_Ass1.xlsx", sheet = 2, col_types = "text") %>%
  clean_names()

# Omgjør BNP og populasjonen til langt format

PopulasjonLang <- Populasjon %>%
  pivot_longer(
    cols = starts_with("x"),
    names_to = "aar",
    values_to = "befolknning"
  ) %>%
  mutate(
    aar = as.integer(str_remove(aar, "^\w")),
    befolkning = as.numeric(befolkning)
  )

BNPLang <- BNP %>%
  pivot_longer(
    cols = starts_with("x"),
    names_to = "aar",
    values_to = "BNP"
```

```

) %>%
mutate(
  aar = as.integer(str_remove(aar, "^\u00c5")),
  BNP = as.numeric(BNP)
)

```

Kort gjennomgang av datasett og variabler

Datasette *demo_r_pjanggr3* som er hentet fra Eurostat inneholder årlige befolkningestimerer på NUTS3-nivå for EU-, EFTA- og kandidatland. Variablene **values** viser totalt antall bosatte personer per 1. januar, målt i antall personer. Hver observasjon identifiseres ved regionkode(**geo**) og år (**time**), sammt kjønn (**sex**) og alder (**age**). I denne analysen benyttes kun total befolkning (**sex = T, age = TOTAL**), slik at dataene ikke er splittet etter kjønn eller alder.

GDP per Capita Calculation

```

# Kombinerer tabellene og filtrerer bort NA-verdiene

Kombinert <- PopulasjonLang %>%
  left_join(BNPLang, by = join_by(region, aar)) %>%
  mutate(bnp_per_innbygger = (BNP * 1e6) / befolkning)

Kombinert <- Kombinert %>%
  filter(!is.na(bnp_per_innbygger))

```

Descriptive Analysis

Table 1: Sammendrag av BNP per innbygger (2000–2023)

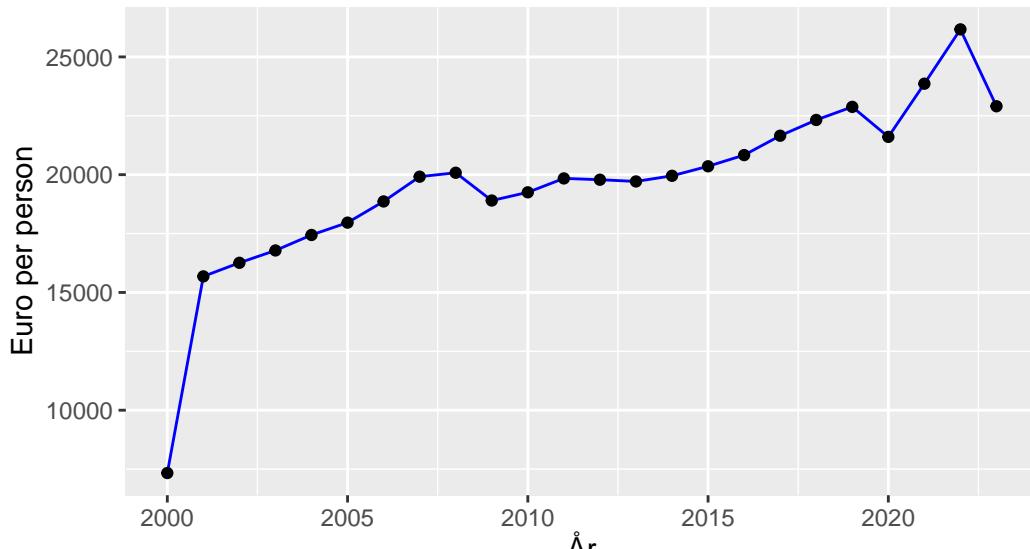
gjennomsnitt	median	standardavvik	minimum	maksimum
19 883,13	20 461,87	11 386,48	825,34	66 632,91

Basert på Eurostat-dataene, er BNP per innbygger beregnet som total BNP delt på antall innbyggere i hver NUTS3-region. Datasettet dekker perioden 2000-2023 for Italia, Romania, Finland og Litauen. BNP-verdiene er målt i millioner euro og er konvertert til euro per person for å gjøre regionene sammenlignbare både mellom og innenfor landene.

Gjennomsnittlig BNP per innbygger i utvalget er **19 883 euro**, med et standardavvik på **11 386 euro**, noe som viser en tydelig variasjon i økonomisk velstand mellom regionene. Minimumsverdien er **825 euro**, mens maksimum når opp mot **66 633 euro**, som reflekterer store forskjeller mellom utviklede regioner i Italia og Finland sammenlignet med mindre utviklede områder i Romania og Litauen. Medianverdien på **20 462 euro** ligger nær gjennomsnittet, noe som indikerer en tilnærmet symmetrisk fordeling, men noen enkelte regioner med svært høyt BNP per innbygger trekker gjennomsnittet litt opp.

Over tid viser dataene en jevn vekst i BNP per innbygger i alle fire land. Finland og Italia hadde de høyeste nivåene gjennom hele perioden, mens Romania og Litauen viser sterkt vekst og gradvis konvergens mot de rike landene etter 2010. Figurene nedenfor illustrerer både tidsutviklingen i gjennomsnittlig BNP per innbygger og fordelingen mellom regionene i det siste observasjonsåret, som danner grunnlaget for videre analyse av regional ulikhet i del B.

Utvikling i gjennomsnittlig BNP per innbygger (2000–2023)



Kilde: Eurostat – nama_10r_3gdp og demo_r_pjanaggr3

```
# Fordeling per land

# Samler alle regionene basert på de 2 første bokstavene i geo variabelen for å samle alle regionene
Kombinert <- Kombinert %>%
  mutate(land = substr(geo, 1, 2))

# Beregner gjennomsnittlig BNP per innbygger per land
SnittLand <- Kombinert %>%
  group_by(land) %>%
```

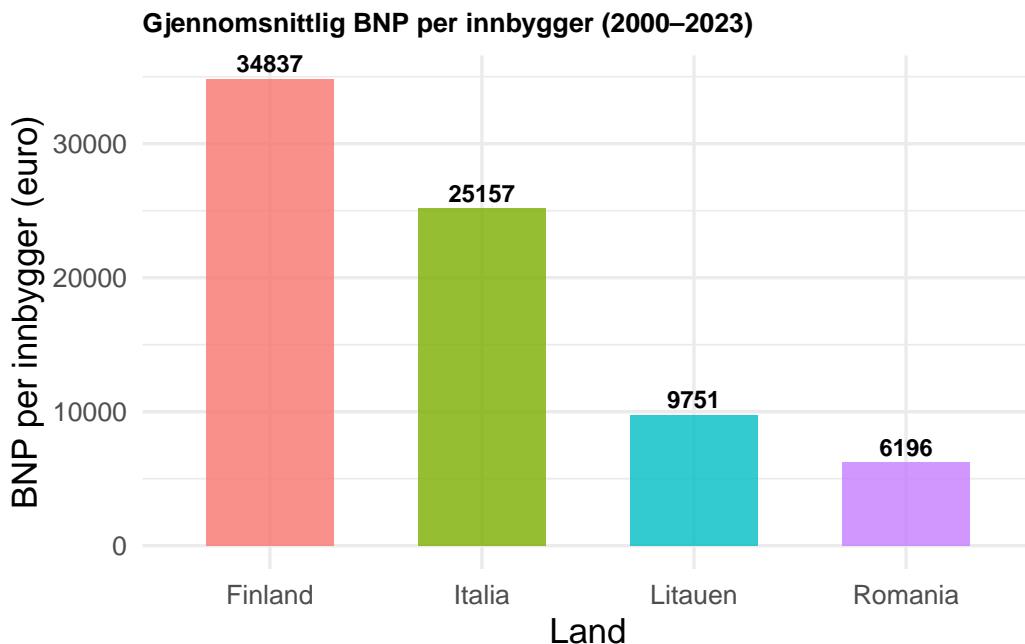
```

summarise(gjennomsnitt_bnp = mean(bnp_per_innbygger, na.rm = TRUE))

# Lager en til variabel slik at det blir enklere å lese av grafen
SnittLand <- SnittLand %>%
  mutate(
    land_navn = recode(land,
      "IT" = "Italia",
      "FI" = "Finland",
      "LT" = "Litauen",
      "RO" = "Romania"
    )
  )

ggplot(SnittLand, aes(x = land_navn, y = gjennomsnitt_bnp, fill = land_navn)) +
  geom_col(alpha = 0.8, width = 0.6) +
  geom_text(aes(label = round(gjennomsnitt_bnp, 0)),
            vjust = -0.4, size = 3, fontface = "bold") +
  labs(
    title = "Gjennomsnittlig BNP per innbygger (2000–2023)",
    x = "Land",
    y = "BNP per innbygger (euro)"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    legend.position = "none",
    plot.title = element_text(face = "bold", size = 10)
  )

```



```

library(tidyverse)
library(ggrepel)

# 1. Beregn gjennomsnitt per land og år
SnittTidLand <- Kombinert %>%
  group_by(land, aar) %>%
  summarise(
    gjennomsnitt_bnp = mean(bnp_per_innbygger, na.rm = TRUE),
    .groups = "drop"
  )

# 2. Bytt ut landkodene med fulle navn
SnittTidLand <- SnittTidLand %>%
  mutate(
    land_navn = recode(land,
      "IT" = "Italia",
      "FI" = "Finland",
      "LT" = "Litauen",
      "RO" = "Romania"
    )
  )

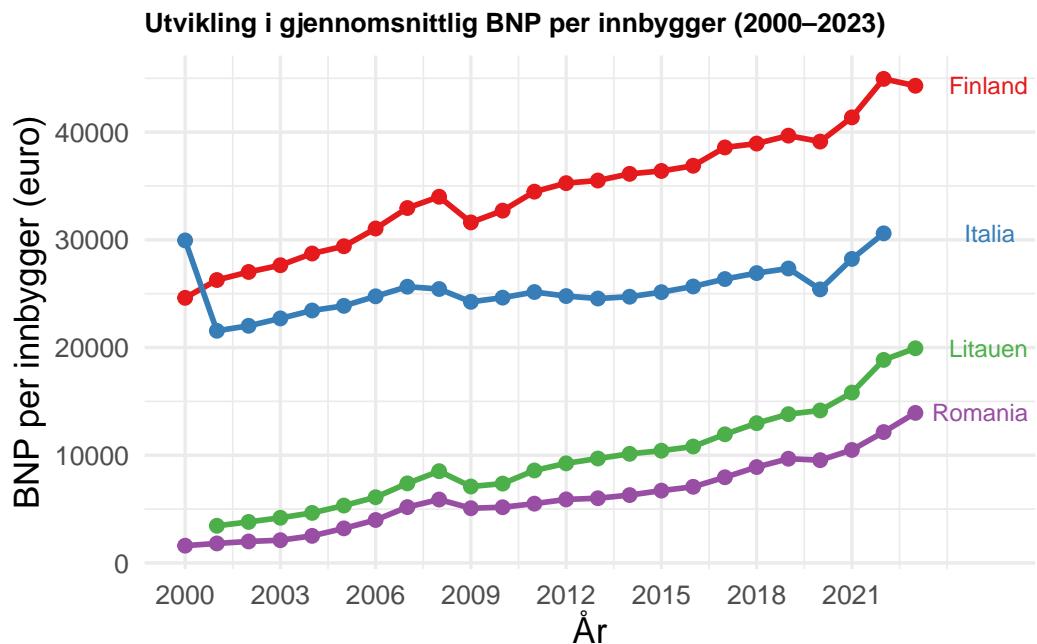
# 3. Finn siste observasjon for etikettpllassering
  
```

```

etiketter <- SnittTidLand %>%
  group_by(land_navn) %>%
  filter(aar == max(aar))

# 4. Lag linjediagram med etiketter
ggplot(SnittTidLand, aes(x = aar, y = gjennomsnitt_bnp, color = land_navn)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  geom_text_repel(
    data = etiketter,
    aes(label = land_navn),
    nudge_x = 2.5,
    size = 3,
    direction = "y",
    hjust = 0,
    segment.color = NA
  ) +
  scale_x_continuous(
    breaks = seq(2000, 2023, 3)
  ) +
  scale_color_brewer(palette = "Set1") +
  labs(
    title = "Utvikling i gjennomsnittlig BNP per innbygger (2000-2023)",
    x = "År",
    y = "BNP per innbygger (euro)",
    color = NULL
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold", size = 10),
    legend.position = "none",
  )

```



Part B: Regional Inequity

Literature Review

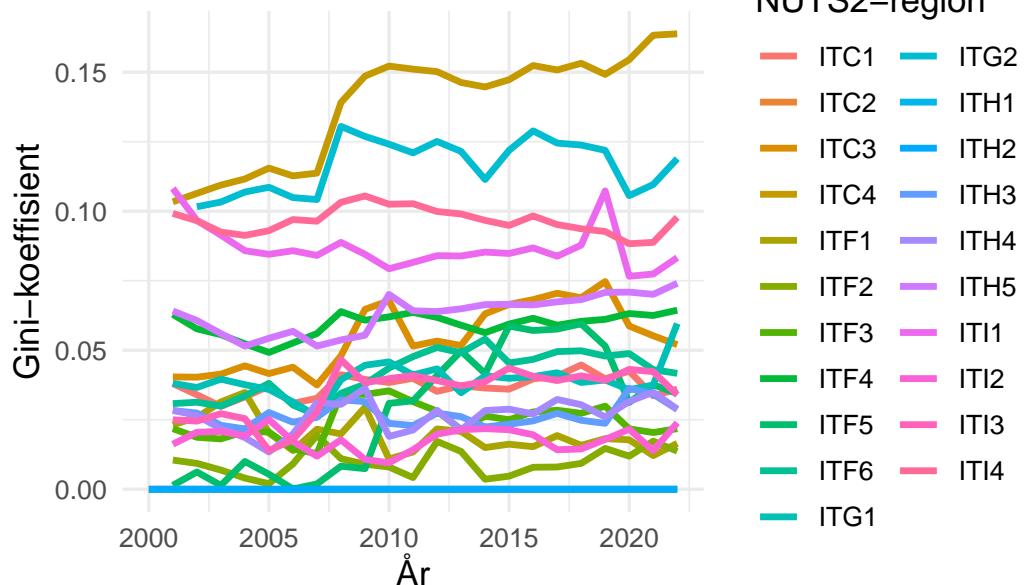
- Read the paper by Lessmann & Seidel (2017). You will find a copy under Filer/pdf at the Canvas site of the course. Give a short summary of the paper (200 – 400 words).

Gini Coefficient Calculation

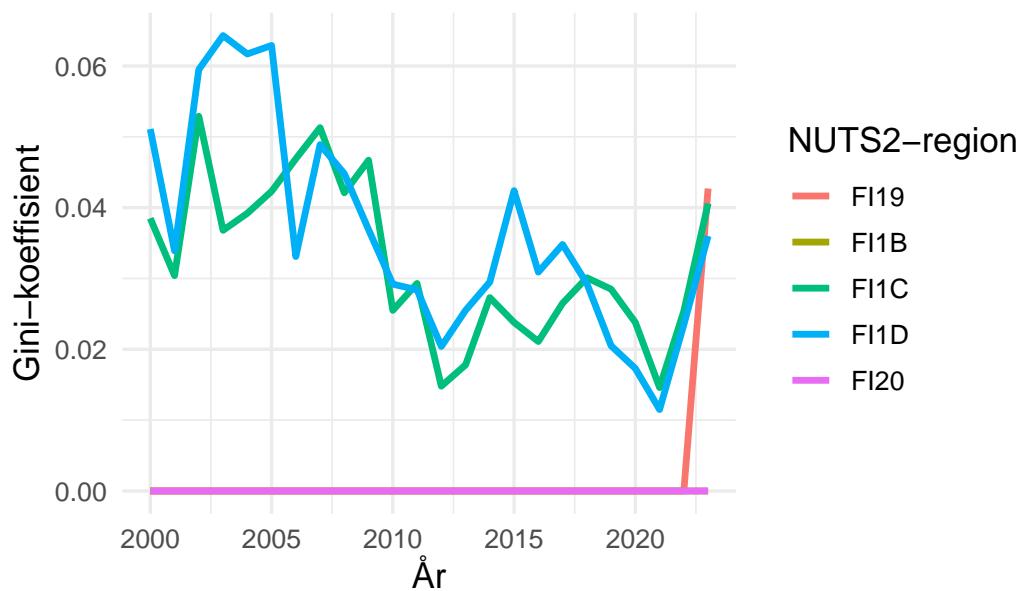
```
# A tibble: 4 x 2
  land   antall_nuts2
  <chr>      <int>
1 FI          5
2 IT         21
3 LT          2
4 RO          8
```

Utvikling i Gini-koeffisienter (NUTS2) – Italia

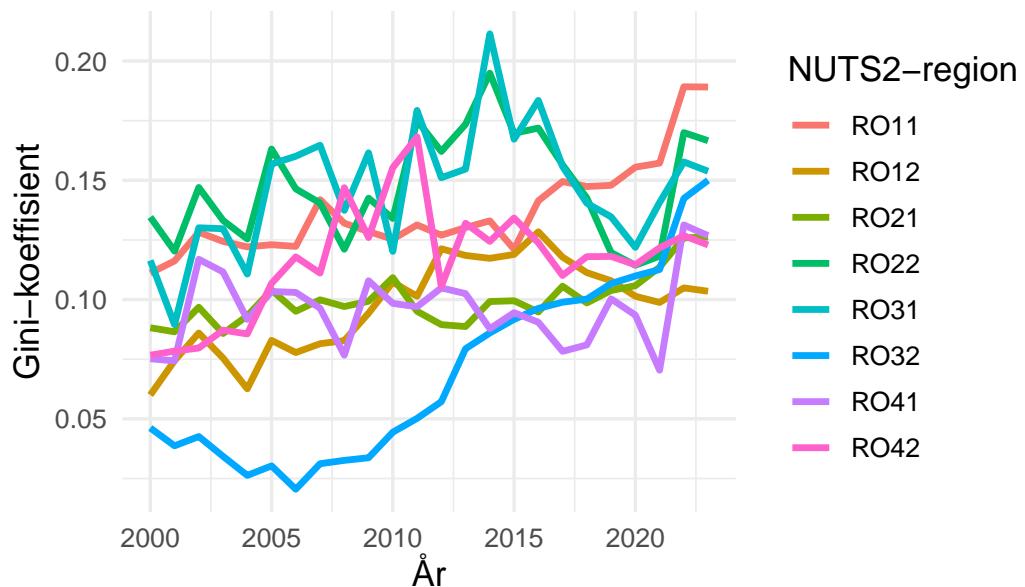
NUTS2-region



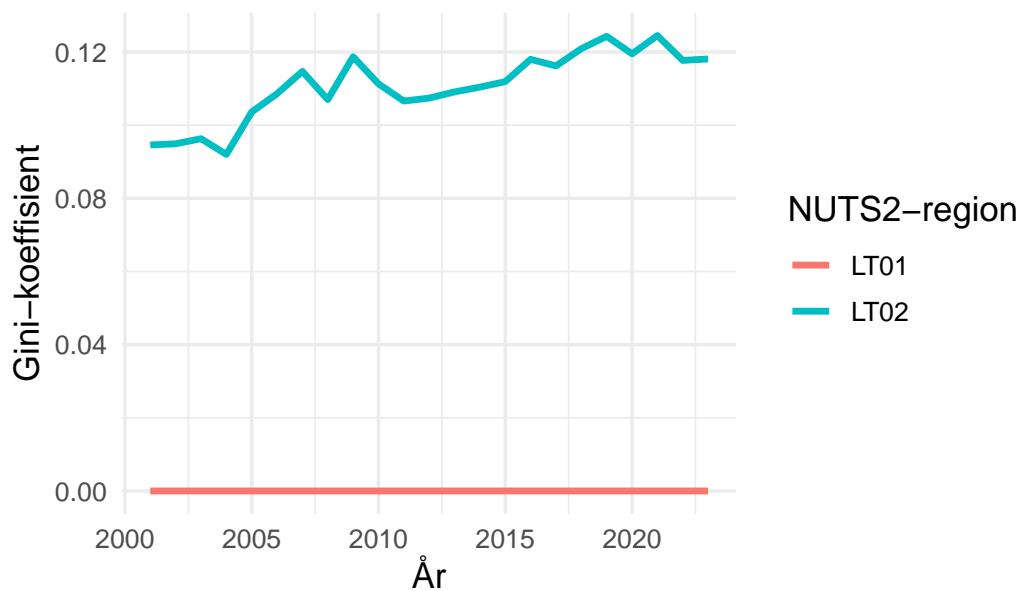
Utvikling i Gini-koeffisienter (NUTS2) – Finland



Utvikling i Gini–koeffisienter (NUTS2) – Romania



Utvikling i Gini–koeffisienter (NUTS2) – Litauen



Utvikling i Gini-koeffisienter (NUTS2)

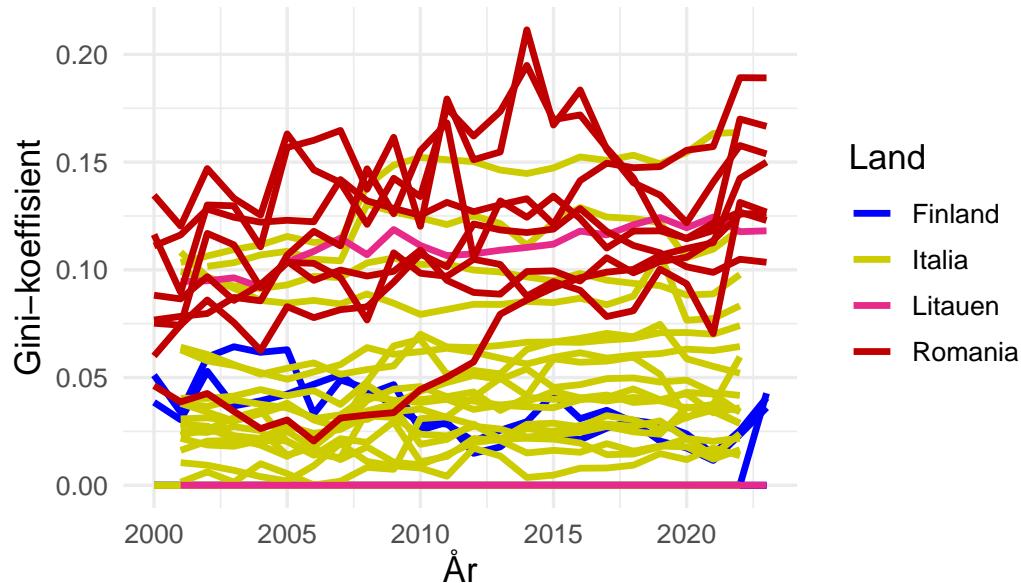
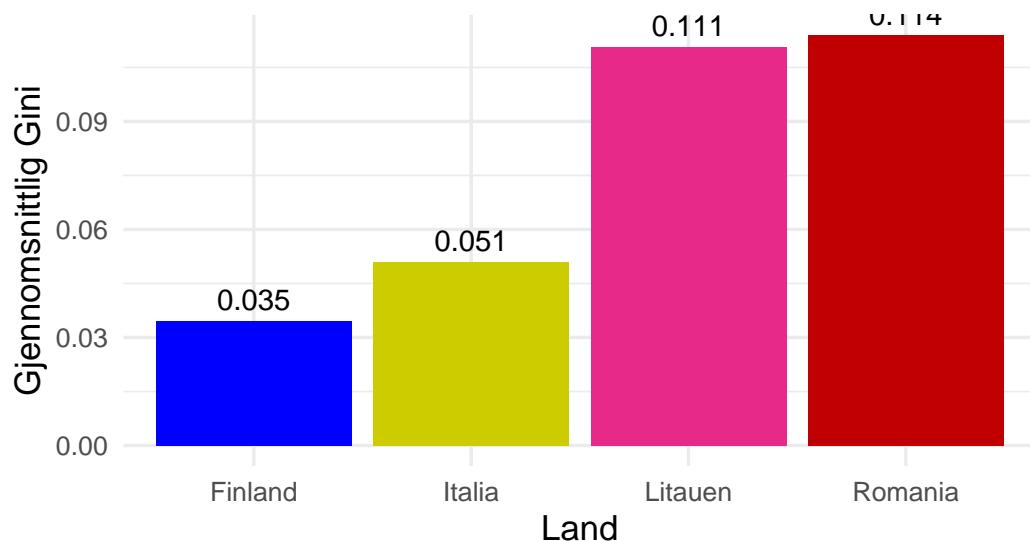


Table 2: Gjennomsnittlig Gini-koeffisient per land (uten 0-verdier)

land	gjennomsnitt_gini	sd_gini	min_gini	max_gini
Romania	0.1139	0.0350	0.0205	0.2114
Litauen	0.1107	0.0095	0.0920	0.1245
Italia	0.0510	0.0364	0.0001	0.1638
Finland	0.0346	0.0132	0.0115	0.0643

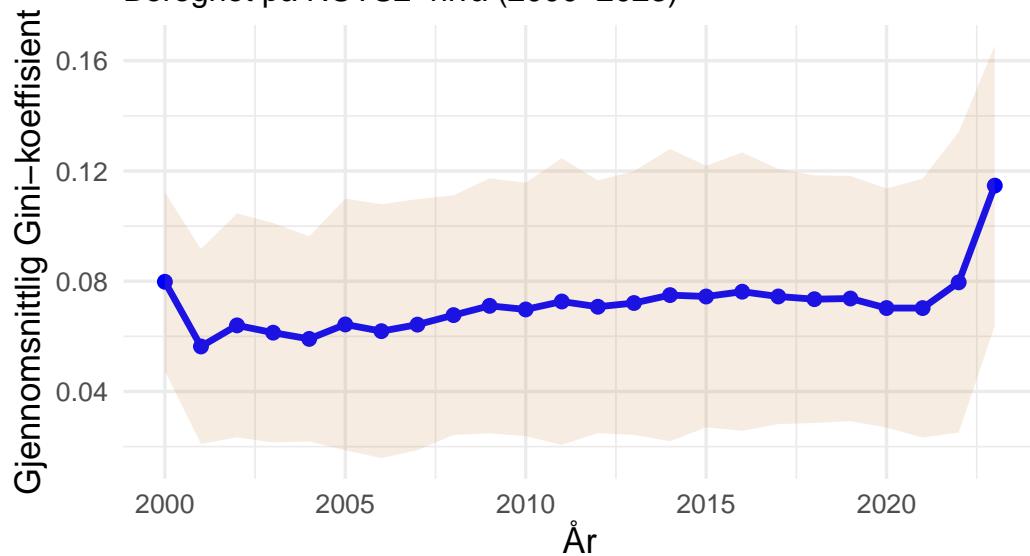
Gjennomsnittlig Gini–koeffisient per land (2000–2)

Beregnet på NUTS2-nivå, uten nullverdier



Gjennomsnittlig Gini–koeffisient over tid (alle lan

Beregnet på NUTS2-nivå (2000–2023)



Diskusjon

De beregnede Gini-koeffisientene for BNP per innbygger på NUTS2-nivå viser tydelige forskjeller i graden av regional ulikhet mellom landene i utvalget. Romania og Litauen har de høyeste gjennomsnittlige Gini-verdiene på henholdsvis 0,1139 og 0,1107, mens Italia (0,0510) og Finland (0,0346) viser langt lavere nivåer av regional ulikhet. Dette innebærer at de østeuropeiske landene, som i større grad har vært gjennom en overgangsøkonomi og rask omstilling etter EU-innlemmelsen, fortsatt opplever mer betydelige forskjeller i økonomisk utvikling mellom regionene sammenlignet med de mer etablerte økonomiene i Vest- og Nord-Europa.

Romania skiller seg ut som landet med størst regional ulikhet. Både gjennomsnittet og variasjonen (standardavvik = 0,0350) er høye, noe som indikerer store forskjeller mellom regionene og betydelig endring over tid.

Resultatene fra analysen viser tydelige mønstre i den regionale økonomiske ulikheten mellom de fire landene. Italia og Finland, som begge har modne og diversifiserte økonomier, viser lave og stabile Gini-koeffisienter. Dette indikerer at inntektsnivået mellom regionen er relativt balansert, og at den økonomiske veksten har vært jevnt fordelt. Italia har likevel små, men medvarende forskjeller mellom nord og sør, noe som gjenspeiles i at enkelte NUTS2-regioner ligger over landsgjennomsnittet.

Romania og Litauen viser derimot høyere Gini-koeffisienter, som peker på større regionale forskjeller. I Romania skyldes dette i stor grad forskjellen mellom hovedstadregionen Bucuresti-Ilfov og de litt mer landlige regionene. Litauen viser et lignende mønster, med rask vekst i Vilnius-regionen sammenlignet med de sørlige og østlige områdene.

Over tid ser vi altså en økende konvergens mellom landene, men fortsatt betydelig intern ulikhet i de nye EU-medlemmene. Denne utviklingen kan