

Assignment 1 — MSB104 — Group 3

Irjan & Magnus

Part A_ Sub-national GDP and GDP per capita

Data acquisition

```
Sys.setlocale("LC_ALL", "Norwegian")
```

```
[1] "LC_COLLATE=Norwegian_Norway.1252;LC_CTYPE=Norwegian_Norway.1252;LC_MONETARY=Norwegian_Norway.1252;LC_NUMERIC=Norwegian_Norway.1252;LC_TIME=Norwegian_Norway.1252"
```

```
library(readxl)
library(dplyr)
library(tidyr)
library(janitor)

# 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 = "befolkning"
```

```

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

```

Kort gjennomgang av datasett og variabler

Datasette *demo_r_pjanggr3* som er hentet fra Eurostat inneholder årlige befolkningestimater 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**), samt 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))

```

Decriptive Analysis

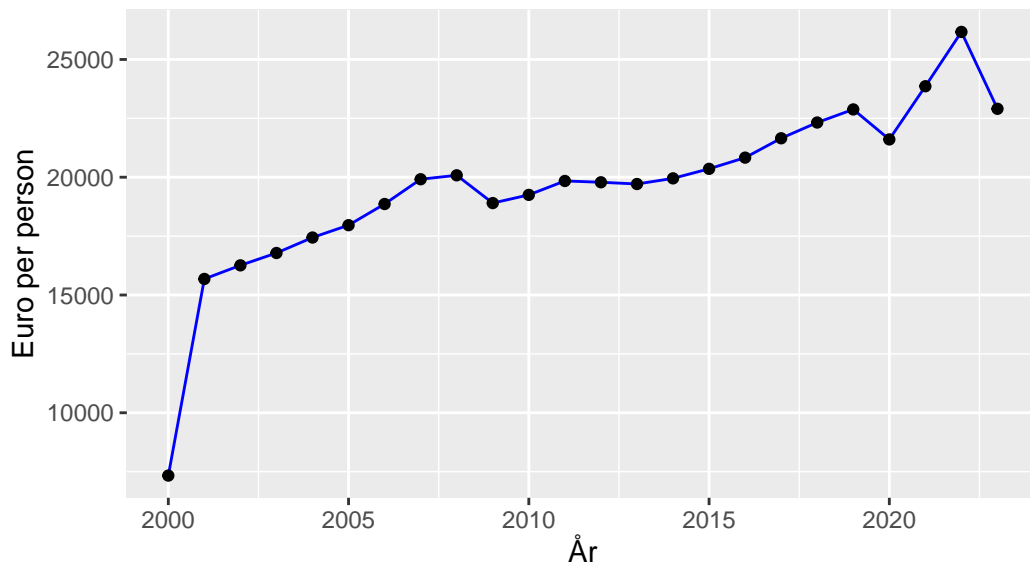
```
# A tibble: 1 x 5
  gjennomsnitt median standardavvik minimum maksimum
    <dbl>   <dbl>         <dbl>   <dbl>     <dbl>
1    19883. 20462.         11386.    825.    66633.
```

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 sterk 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 r
```

```
Kombinert <- Kombinert %>%
```

```
  mutate(land = substr(gео, 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",
```

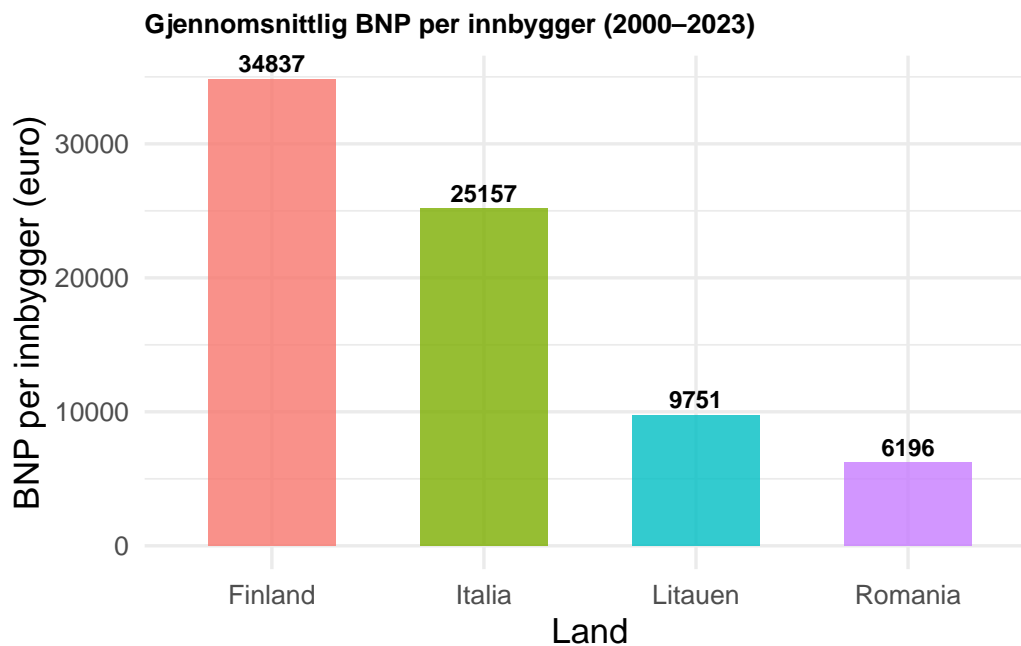
```
      "LT" = "Litauen",
```

```
      "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"
  )

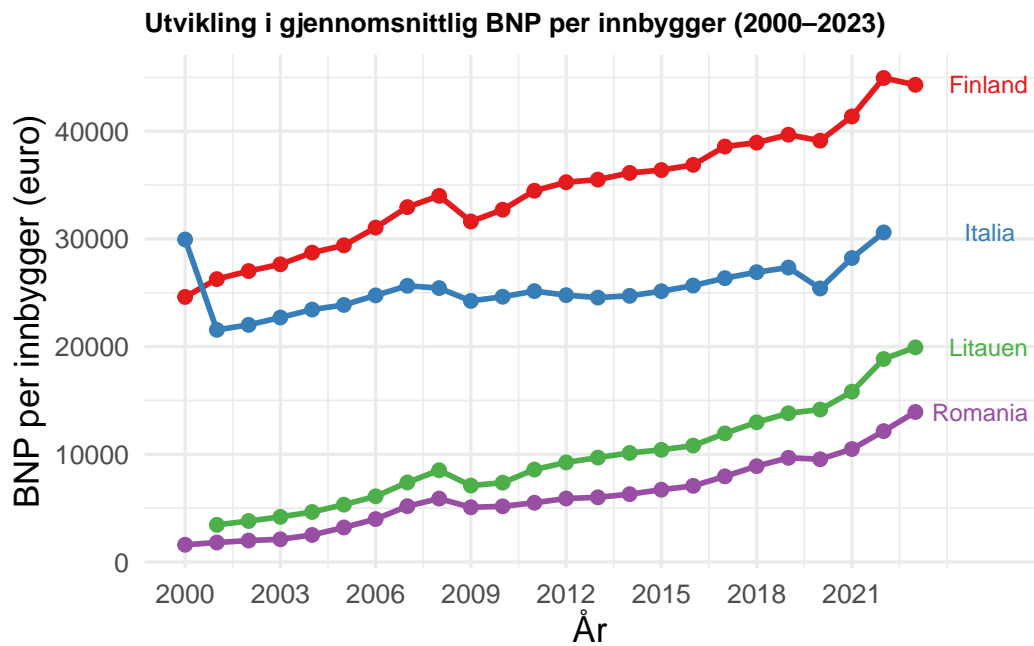
# 2. Bytt ut landkodene med fulle navn
SnittTidLand <- SnittTidLand %>%
  mutate(
    land_navn = recode(land,
      "IT" = "Italia",
      "FI" = "Finland",
      "LT" = "Litauen",
      "RO" = "Romania"
    )
  )

# 3. Finn siste observasjon for etikettplassering
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

litarture 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
```

```

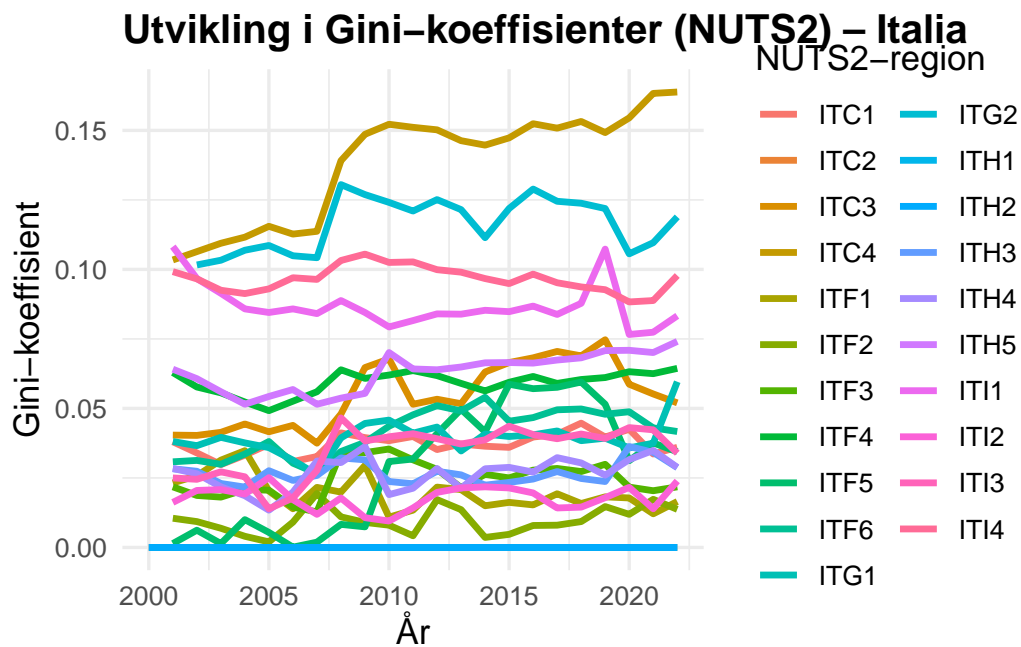
2 IT      21
3 LT      2
4 RO      8

```

```

# A tibble: 464 x 3
  nuts2  aar  gini
  <chr> <int> <dbl>
1 ITC1   2001 0.038
2 ITC1   2002 0.034
3 ITC1   2003 0.0298
4 ITC1   2004 0.0332
5 ITC1   2005 0.0371
6 ITC1   2006 0.0309
7 ITC1   2007 0.0327
8 ITC1   2008 0.0412
9 ITC1   2009 0.0395
10 ITC1  2010 0.0384
# i 454 more rows

```



```

# A tibble: 120 x 3
  nuts2  aar  gini
  <chr> <int> <dbl>

```

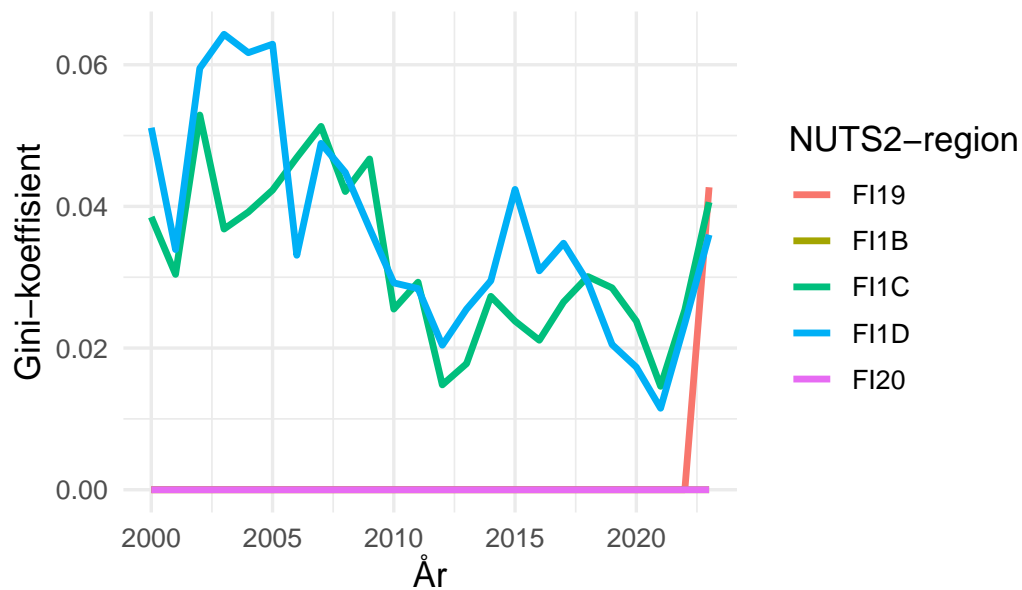


```

1 FI19 2000 0
2 FI19 2001 0
3 FI19 2002 0
4 FI19 2003 0
5 FI19 2004 0
6 FI19 2005 0
7 FI19 2006 0
8 FI19 2007 0
9 FI19 2008 0
10 FI19 2009 0
# i 110 more rows

```

Utvikling i Gini-koeffisienter (NUTS2) – Finland



```

# A tibble: 192 x 3
  nuts2  aar  gini
  <chr> <int> <dbl>
1 R011  2000 0.111
2 R011  2001 0.116
3 R011  2002 0.128
4 R011  2003 0.124
5 R011  2004 0.122
6 R011  2005 0.123
7 R011  2006 0.122

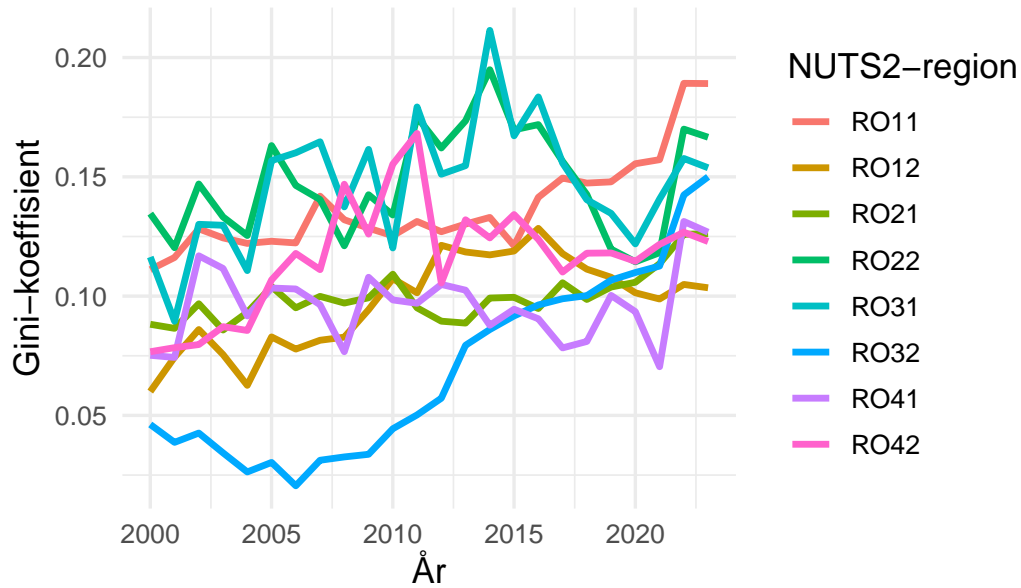
```

```

8 RO11 2007 0.142
9 RO11 2008 0.132
10 RO11 2009 0.128
# i 182 more rows

```

Utvikling i Gini-koeffisienter (NUTS2) – Romania

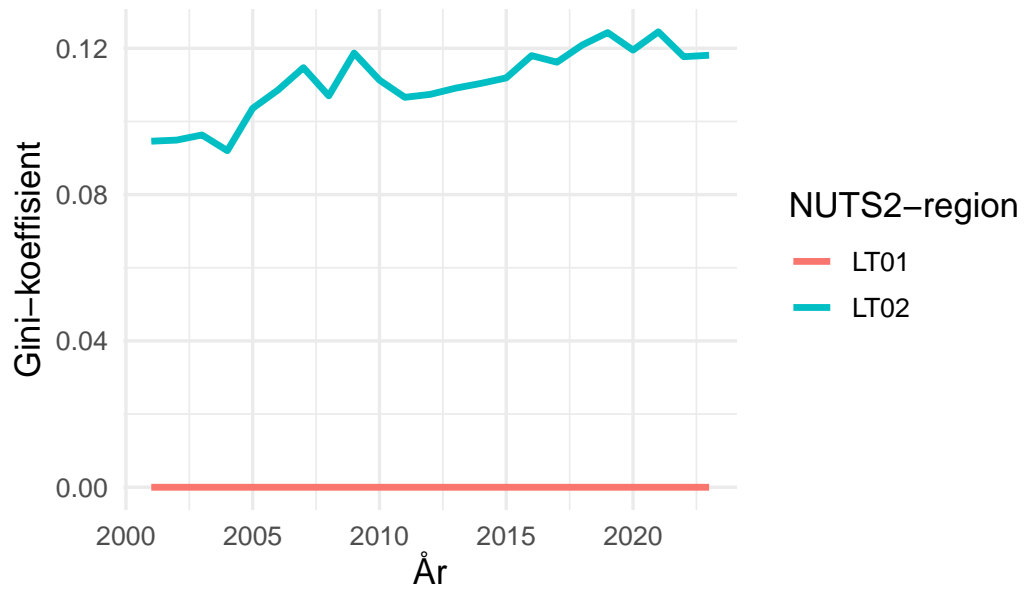


```

# A tibble: 46 x 3
  nuts2  aar  gini
  <chr> <int> <dbl>
1 LT01  2001    0
2 LT01  2002    0
3 LT01  2003    0
4 LT01  2004    0
5 LT01  2005    0
6 LT01  2006    0
7 LT01  2007    0
8 LT01  2008    0
9 LT01  2009    0
10 LT01  2010    0
# i 36 more rows

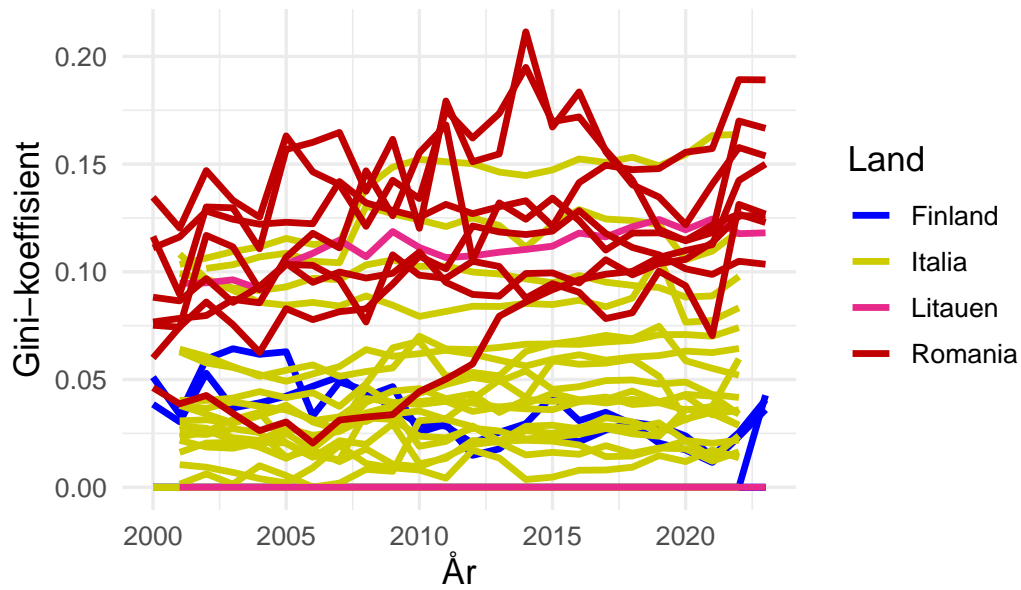
```

Utvikling i Gini-koeffisienter (NUTS2) – Litauen



```
# A tibble: 6 x 4
  nuts2  aar  gini land
<chr> <int> <dbl> <chr>
1 ITC1   2001 0.038 Italia
2 ITC1   2002 0.034 Italia
3 ITC1   2003 0.0298 Italia
4 ITC1   2004 0.0332 Italia
5 ITC1   2005 0.0371 Italia
6 ITC1   2006 0.0309 Italia
```

Utvikling i Gini-koeffisienter (NUTS2)



A tibble: 4 x 5

	land	gjennomsnitt_gini	sd_gini	min_gini	max_gini
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Romania	0.114	0.035	0.0205	0.211
2	Litauen	0.111	0.0095	0.092	0.124
3	Italia	0.051	0.0364	0.0001	0.164
4	Finland	0.0346	0.0132	0.0115	0.0643

Gjennomsnittlig Gini-koeffisient per land (2000–2

Beregnet på NUTS2-nivå, uten nullverdier

