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 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(str_replace_all(befolkning, " ", ""))
  )

# Omgjør BNP til langt format

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(str_replace_all(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**), samnt 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.

Det finnes ulike hoved metoder å beregne brytto nasjonalt produkt på. Eurostat har valgt å benytte den såkalte «inntekstmetoden». Eurostat velger denne metoden ovenfor utgifts metoden grunnet mangel på data over gode overføringer mellom regioner.

Etter inntektsmetoden regnes BNP på følgende måte: Lønn som utbetales til ansatte + bedrifter sin fortjeneste + skatte og avgifter minus subsidier gitt fra staten + avskrivinger knyttet til industri.

BNP per innbygger

```
# Kombinerer tabellene og filtrerer bort NA-verdiene


KombinertRen <- Kombinert %>%
  filter(!is.na(BNPPI))
```

```
[1] 4680
```

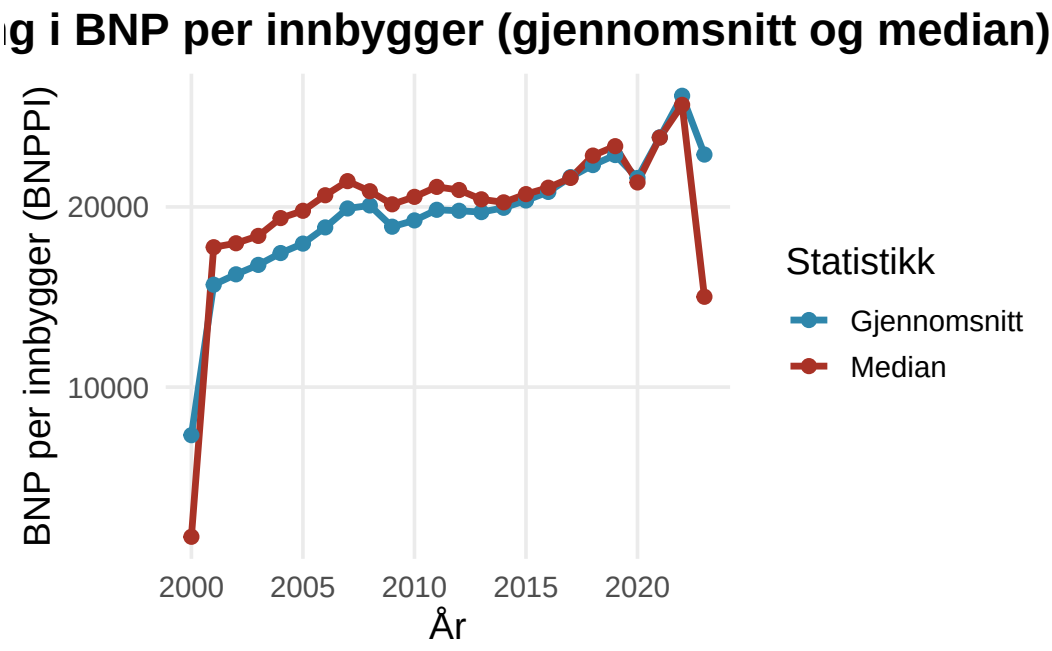
```
# A tibble: 1 x 6
    geo region   aar befolkning   BNP BNPPI
  <int>  <int> <int>      <int> <int> <int>
1     0      0     0        358   515   841
```

I tabellene over har vi kombinert de to datasettene hentet fra Eurostat. Vi har da en tabell som viser BNP per innbygger i hvert av NUTS3 områdene for delt på år. I hver observasjon hvor det forekommer en NA verdi i populasjon og eller total BNP vil også få BNPPI som NA verdi. Datasettet består av totalt 4680 observasjoner, og 841 av dem ender opp ned NA verdi i BNPPI. For å videre kunne jobbe mer effektivt i utarbeidelsen av tabellene, lager vi et nytt datasett hvor alle observasjoner med NA verdi er fjernet.

Table 1: Utvikling i BNP per innbygger (gjennomsnitt, median, minimum og maksimum) per år

År	Gjennomsnitt BNPPI	Median BNPPI	Minste BNPPI	Største BNPPI
2000	7 332	1 693	825	36 461
2001	15 681	17 764	1 025	38 652
2002	16 259	17 984	1 066	40 238
2003	16 783	18 392	1 235	41 704
2004	17 440	19 381	1 338	43 010
2005	17 965	19 786	1 534	44 057
2006	18 864	20 647	1 998	44 605
2007	19 915	21 431	2 460	47 618
2008	20 082	20 872	3 142	50 967
2009	18 903	20 146	2 583	49 307
2010	19 251	20 559	2 606	52 021
2011	19 842	21 112	2 760	52 664
2012	19 786	20 939	3 074	51 194
2013	19 713	20 422	3 381	50 043
2014	19 952	20 254	3 507	50 629
2015	20 357	20 710	3 663	51 592
2016	20 832	21 068	3 833	53 373
2017	21 655	21 609	4 349	54 406
2018	22 323	22 850	4 982	55 923
2019	22 879	23 372	5 274	57 074
2020	21 605	21 358	5 250	55 637
2021	23 863	23 849	5 685	61 275
2022	26 170	25 664	6 070	66 633
2023	22 906	15 014	6 989	61 545

I tabellen over ser vi utviklingen av gjennomsnitt, median, minste og høyeste verdi for BNP per inbygger for fra år 2000 til år 2023. Verdiene er basert på alle observasjonene fra datasettet hvor observasjoner med NA verdier er fjernet. Videre skal vi lage en deskriptiv analyse utifra statistikken som fremkommer av datasettet. For å enklere visualisere dette lager vi linediagrammer for sentrale mål.




```
[1] "geo"        "region"     "aar"        "befolkning" "BNP"
[6] "BNPPI"
```

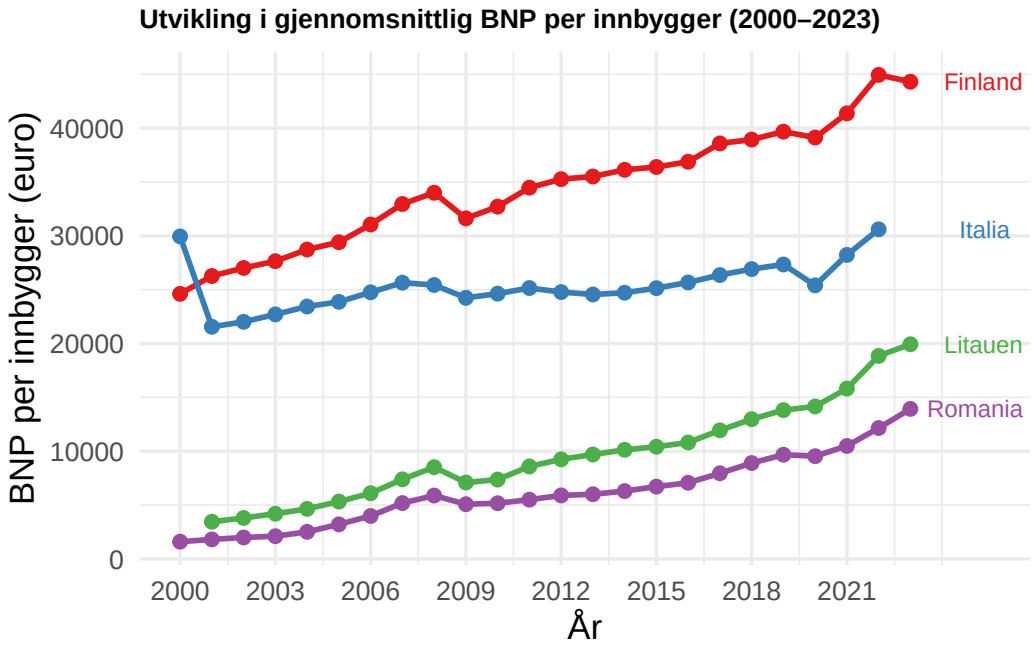
```
# 1. Beregn gjennomsnitt per land og år
KombinertRen <- KombinertRen %>%
  mutate(
    land = substr( geo, 1, 2)
  )
SnittTidLand <- KombinertRen %>%
  group_by(land, aar) %>%
  summarise(
    gjennomsnitt_bnp = mean(BNPPI, na.rm = TRUE),
    .groups = "drop"
  )

# 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",
  )
```



I tabellene over ser vi utviklingen av gjennomsnitt og median verdien fra år 2000 til 2023. Begge de statistiske målene følger i stor grad samme utvikling. Det er en stor økning fra 2000 til 2001. Etter dette er det i hovedsak en jevn generelt vekst fra år til år, med et par avvik. I både 2008 og 2020 ser vi reduksjon i BNPPI sammenlignet med det forekommende året. I 2023 ser vi igjen en drastisk endring, spesielt i Median hvor BNPPI reduseres fra ca 25 000 dollar til ca 15 000 dollar. Reduksjonen i år 2008 og 2020 er ikke overraskende og kan forklares med henholdsvis finanskrisen og Covid-19.

En annen faktor som kan påvirke utviklingen av BNPPI er utartingen av NA verdier i data settet vårt. Som tidligere vist er det totalt 841 observasjoner som ikke kommer med i denne tabellen. I hvilken år disse fremkommer og hvilken NUTS3 regioner som forsvinner fra disse årene kan forventes å ha en effekt på gjennomsnittet for det året. Dette kommer av at noen land generelt sett har høyere BNP enn andre.

aar	antall_NA	total_obs
2000	140	195
2001	31	195
2002	26	195
2003	26	195
2004	26	195
2005	26	195
2006	26	195
2007	26	195

2008	26	195
2009	26	195
2010	26	195
2011	26	195
2012	26	195
2013	26	195
2014	26	195
2015	26	195
2016	26	195
2017	26	195
2018	26	195
2019	26	195
2020	26	195
2021	26	195
2022	26	195
2023	124	195

I denne tabellen ser vi at i BNPPI datasettet er det 140 observasjoner fra 2000 og 124 fra 2023 som har NA verdier. Dette skiller seg ut fra årene i mellom hvor det jevnt gjennom er 26 observasjoner med NA verdi. Totalt er det 195 observasjoner hvert år. En teori som kan forklare at median og gjennomsnittsverdiene er veldig lave i forhold er at NUTS3 regioner med NA verdier i disse årstallene i hovedsak stammer fra land med relativt høyest BNPPI.

Part B: Regional ulikhet

Ginis koeffisient

Oppsumering av artikkel

I 2017 publiserte Christian Lessmann og Andre Seidel en artikkel hvor de ser på hvor godt egnet «lysdata» er til å estimere regional inntekt. Sentrale spørsmål som tas opp i artikkelen er hvordan regionale økonomiske ulikheter utvikler seg over tid, og hvilke faktorer som fører til endringer i regional ulikhet.

I en 2012 artikkel som Lessmann og Seidel er inspirert av presenteres ideen om at «lysdata» er en indikasjon på hvor økonomisk utviklet et land er. Tankegangen er at aktiviteter som koster penger å gjennomføre på kvelden som regel trenger lys, og jo rikere et land/ region er jo flere har råd til å delta på slike aktiviteter. Ved å analysere satellitt bilder på nattestid kan mengden lys være en indikator på velstanden i området.

Nytteverdien fra denne vinklingen kommer av at det er stor variasjon på hvor dokumentert regionale ulikheter er fra land til land. Artikkelen påpeker at tidligere studier på området fokuserer på vel- utviklede nasjoner, da det ikke finnes tilgjengelige gode nok data for mindre utviklete land på regionalt nivå. Dersom det kan bevises at dette konseptet holder vann får forskere enn ny måte å analysere mindre velstående land sin regionale ulikhet.

For å undersøke hvor valid teorien er, analyserer artikkelen først velstående land hvor det allerede finnes godt dekkende data over regionale ulikheter. Ved å analysere bilder fra NASA satellitter og rangere regioner fra 1992 og utover, skaper de estimater over hvor mye økonomisk utvikling regioner har hatt i perioden.

I studien brukes Gini-koeffisient. Dette er et mål som måler hvor skjevt ressurser i en befolkning fordeles, for eksempel mellom regioner. Gini-koeffisient har en skala fra 0 til 1. En 0-verdi betyr at alt er likt fordelt, og 1 betyr at alle ressursene ligger hos en eller et veldig få antall personer. I studien beregnes og måles koeffisienten i regioner over tid, som viser om det er divergens eller konvergens. Dersom koeffisient øker over tid vil det si at forskjellene mellom de rike og fattige øker over tid, og det er divergens. Hvis koeffisienten reduseres over tid reduseres forskjellen mellom rik og fattige, dvs konvergens.

Ved å sammenligne estimatene med faktiske BNP data kommer det frem at teorien i stor grad stemmer overens i middels rike og rike land, mens i fattigere land man har data på er konseptet dårlig egnet.

Beregne Gini-koeffisient på vårt datasett

Videre i oppgaven skal vi beregne Gini-koeffisienten på NUTS2 nivå for landene Italia, Finland, Romania og Litauen. Første steg er å transformere BNNPPIRen datasettet våres for NUTS3 regioner til en oversikt over NUTS2 regioner. Hver NUTS3 region har en tilhørende "geo" kode. Disse består av 5 tegn, for eksempel ITC31 for Imperia. De fire første sympolene er det samme som NUTS 2 geo koden. Ved å samle alle obervasjoner hvor de firste første sympolene er like kan vi lage en datasett fordelt på NUTS 2 regioner.

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.

Part B: Regional Inequity

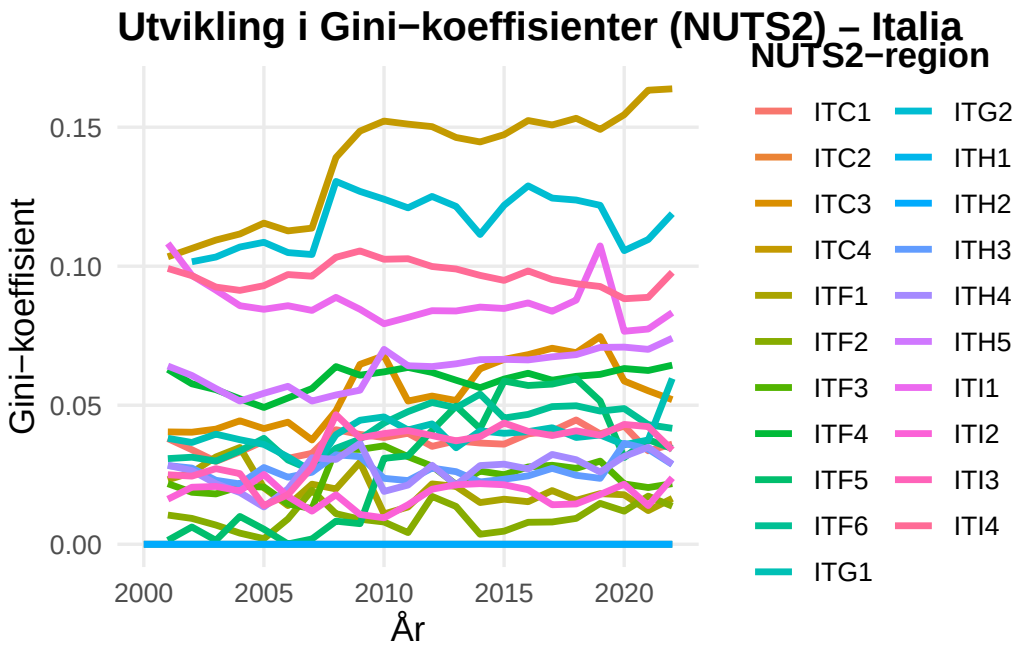
litarture Review

• 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

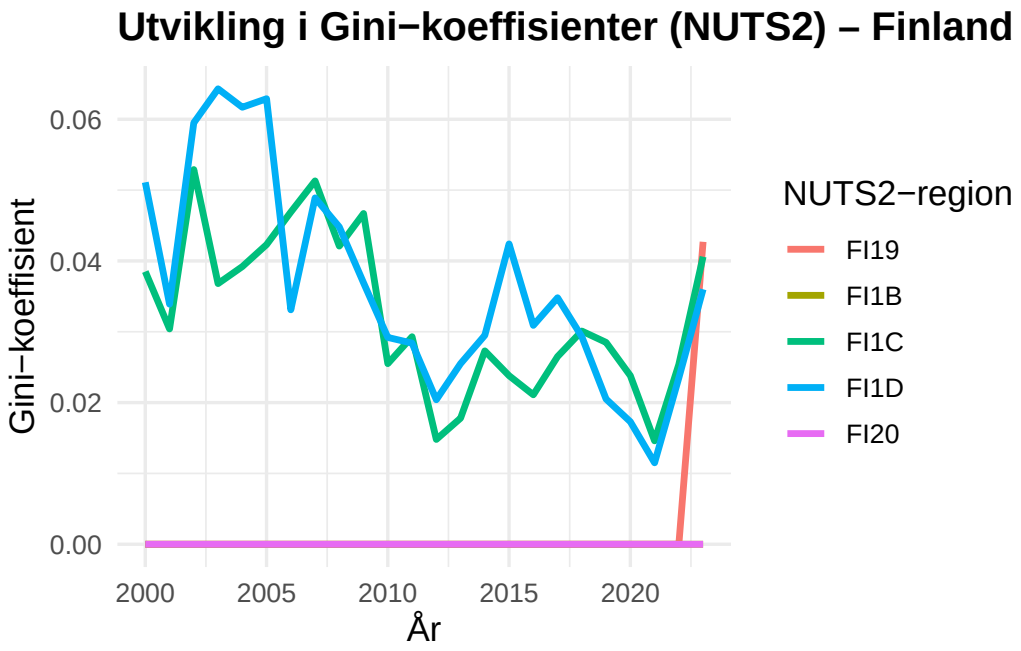
```
# A tibble: 4 x 2
  land  antall_nuts2
  <chr>        <int>
1 FI               5
2 IT              21
3 LT               2
4 RO               8
```

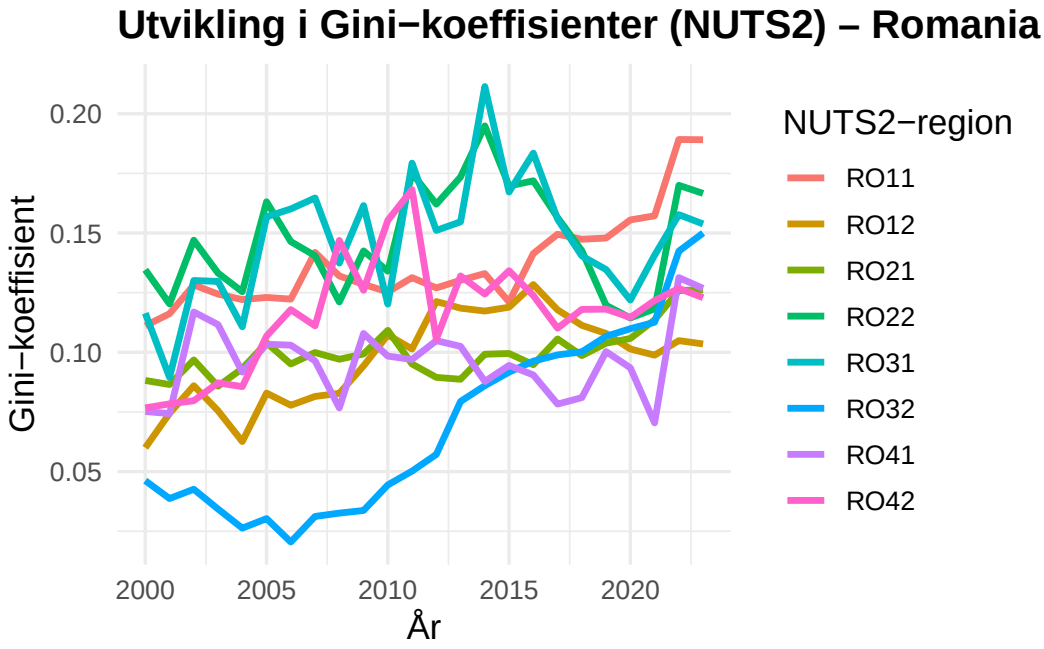
I datasettet er det tilsammen 36 NUTS2 regioner, men hvilken land disse faller innenfor er ganske skjevfordelt. Det er 21 NUTS2 regioner fra Italia, mens Litauen har 2. Finland og Romania har henholdsvis 5 og 8 NUTS2 regioner.

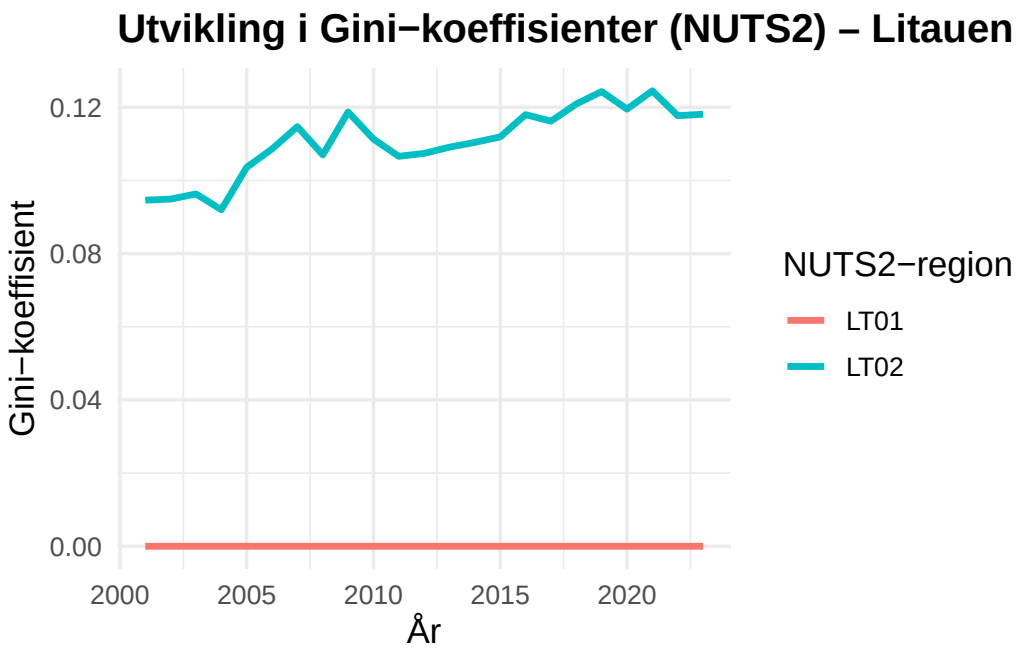



```
<theme> List of 1
 $ plot.title: <ggplot2::element_text>
  ..@ family       : NULL
  ..@ face         : chr "bold"
  ..@ italic       : chr NA
  ..@ fontweight   : num NA
  ..@ fontwidth    : num NA
  ..@ colour       : NULL
  ..@ size         : num 15
  ..@ hjust        : NULL
  ..@ vjust        : NULL
  ..@ angle        : NULL
  ..@ lineheight   : NULL
  ..@ margin       : NULL
  ..@ debug        : NULL
  ..@ inherit.blank: logi FALSE
 @ complete: logi FALSE
 @ validate: logi TRUE
```

Regionene ITC 4 og ITG2 skiller seg ut her da de gjennom hele rekken har en koeffisient mellom 0,10 og 0,17. Dette tyder på at i disse regionene er det en større skjevfordeling av BNPPI enn i resten av landet. I en gjennomsnittsberegning av den gjennomsnittlige Koeffisienten per år ligger den på rundt 0,05.


Utvikling i Gini−koeffisienter (NUTS2) – Finland
Gini−koeffisient
0.06
0.04
0.02
0.00
2000
2005
2010
2015
2020
År
NUTS2−region
FI19
FI1B
FI1C
FI1D
FI20


Utvikling i Gini−koeffisienter (NUTS2) – Romania
Gini−koeffisient
0.20
0.15
0.10
0.05
2000
2005
2010
2015
2020
År
NUTS2−region
RO11
RO12
RO21
RO22
RO31
RO32
RO41
RO42


Utvikling i Gini−koeffisienter (NUTS2) – Litauen
Gini−koeffisient
0.12
0.08
0.04
0.00
2000
2005
2010
2015
2020
År
NUTS2−region
LT01
LT02

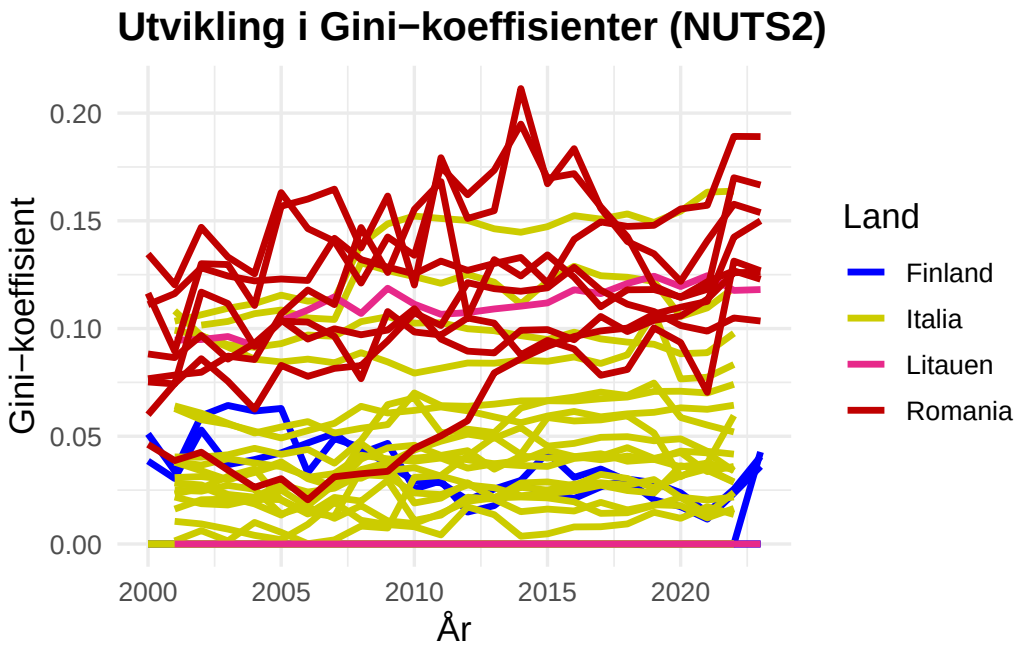
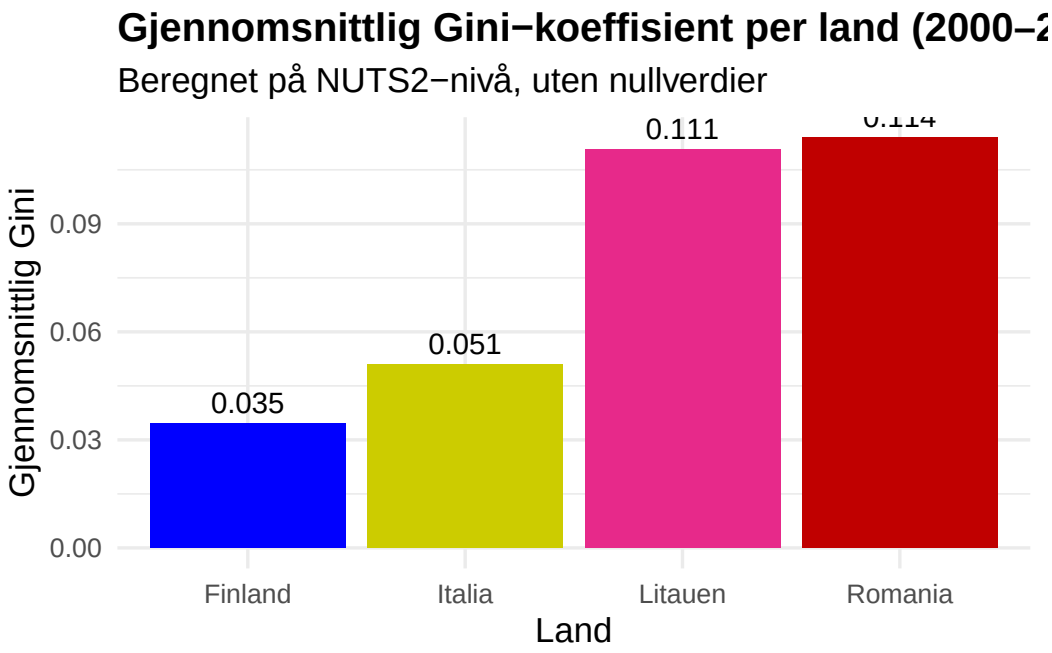
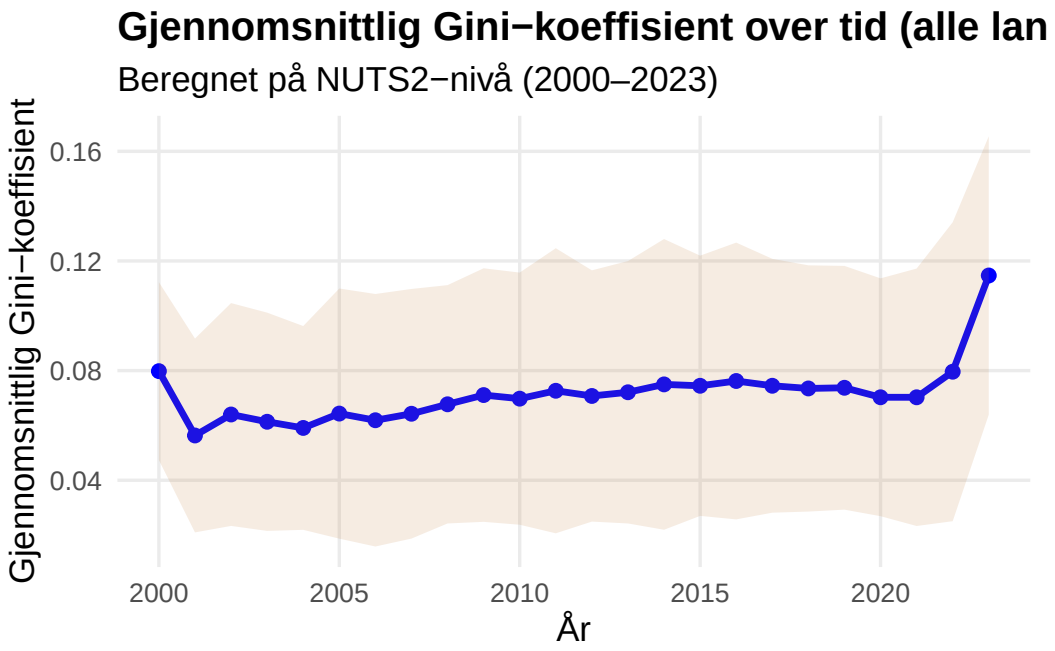

Utvikling i Gini−koeffisienter (NUTS2)
Gini−koeffisient
0.20
0.15
0.10
0.05
0.00
2000
2005
2010
2015
2020
År
Land
Finland
Italia
Litauen
Romania

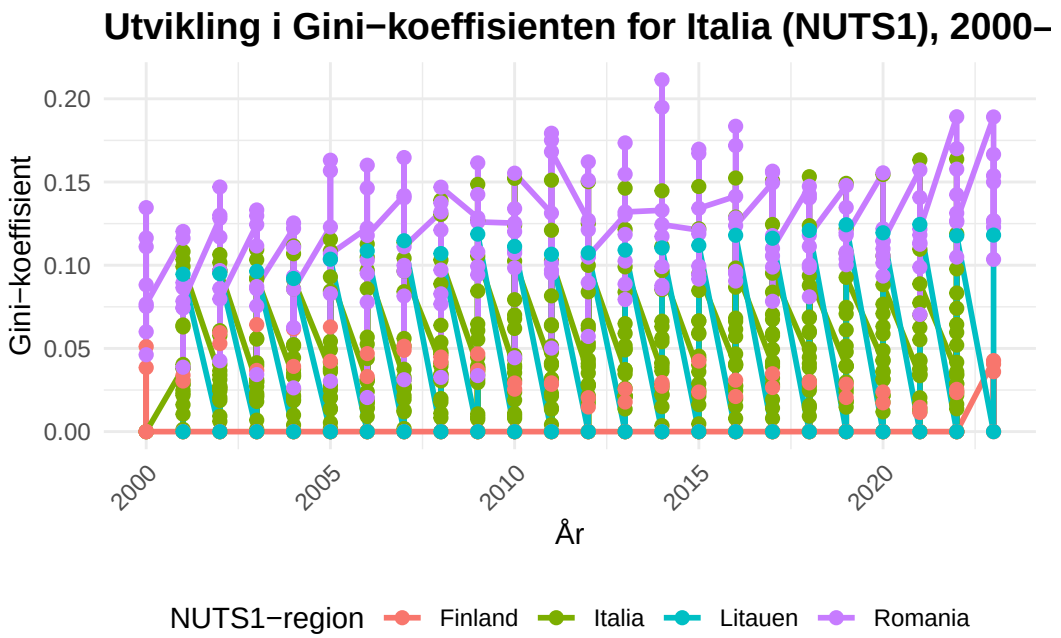
Table 3: Gjennomsnittlig Gini-koeffisient per land (uten 0-verdier)

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





I tabellen over ser vi GINI koeffisienten for alle NUTS2 regionene som tilhører vårt datasett. GINI verdien beskriver hvor jevn fordelt BNPPI er fordelt mellom NUTS3 regionene som havner i samme NUTS2 region. Koeffisienten ligger i mellom 0,02 og 0,11 for majoriteten av regionene gjennom hele årrekken. Dette indikerer at i hver enkelt NUTS2 region er det en jevn fordeling av BNPPI i de tilhørende NUTS3 regionene.



Her er det laget et linediagram som viser utviklingen av GINI koefisient på NUTS1 region basert på fordelingen av BNPPI i NUTS2 regionene. Dette tallet sier noe om fordelingen av BNPPI mellom NUTS2 regioner sett opp mot hele landet. Koefisienten ligger mellom 0,17 og 0,19. Dette indikerer at det i Italia er en større skjevfordeling mellom NUTS2 regionene, enn det er i NUTS 3 regionene satt opp mot regioner i samme NUT2.

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 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