SOK-2008-2022-oppgave1

2022-09-06

Utfordring 1.3

Oppgave 5

```
#Download the Excel file "GCIPrawdatatest.xlsx".
#I have taken away data from Norway 1980-1990 as it was faulty
#Save it in an easily accessible location, such as a folder on your Desktop or in your personal folder.

library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.1.3
```

```
library(ineq)
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.1.3
```

```
# Set your working directory to the correct folder.
# Insert your file path for 'YOURFILEPATH'.
#setwd("YOURFILEPATH")
getwd()
```

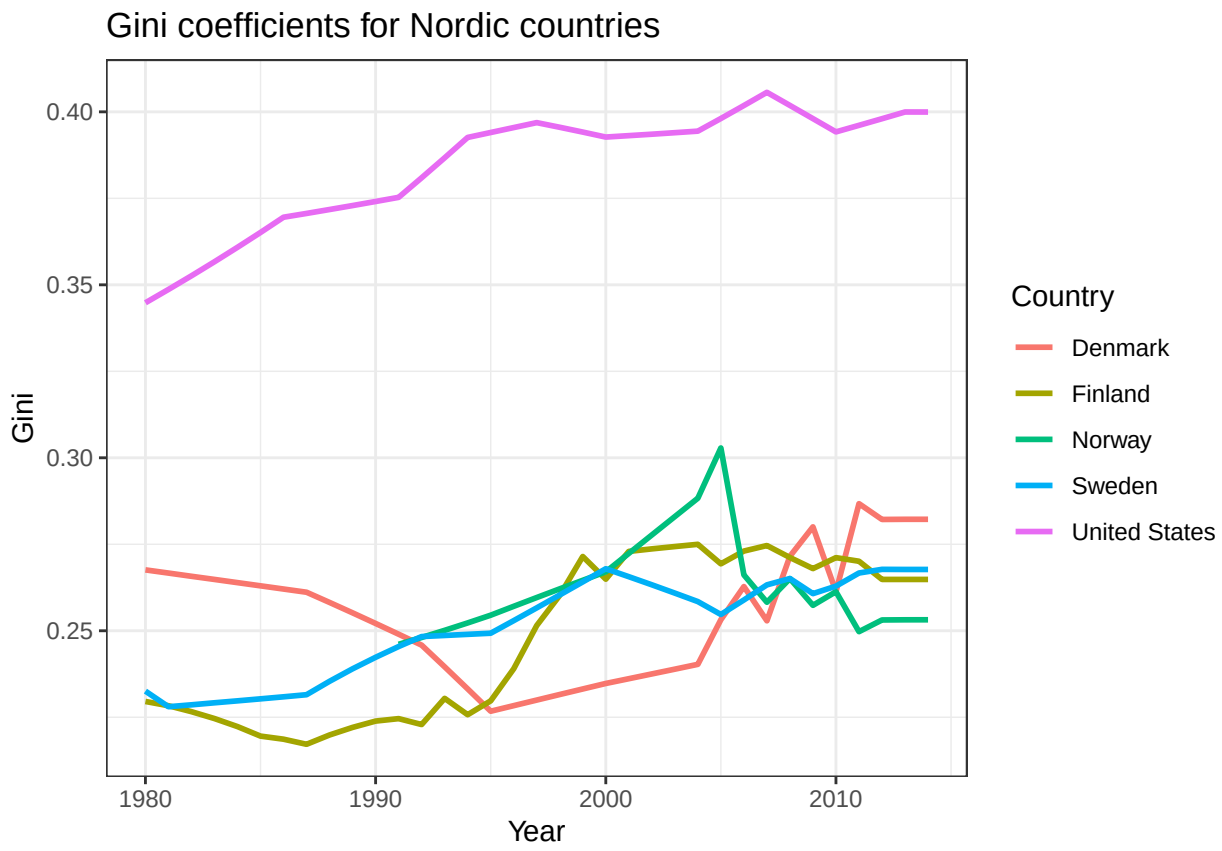
```
## [1] "C:/Users/OddVi/OneDrive/Dokumenter/Studier/SOK-2008/Innleveringer"
```

```
decile_data <- read_excel("GCIPrawdatatest.xlsx", skip = 2)
#The data is now in a 'tibble' (like a spreadsheet for R). Let's use the head function to look at the f
head(decile_data)
```

```
## # A tibble: 6 x 14
##   Country      Year `Decile 1 Income` `Decile 2 Income` `Decile 3 Income`
##   <chr>       <dbl>             <dbl>             <dbl>             <dbl>
## 1 Afghanistan  1980               206               350               455
## 2 Afghanistan  1981               212               361               469
## 3 Afghanistan  1982               221               377               490
## 4 Afghanistan  1983               238               405               527
## 5 Afghanistan  1984               249               424               551
## 6 Afghanistan  1985               256               435               566
## # ... with 9 more variables: Decile 4 Income <dbl>, Decile 5 Income <dbl>,
## #   Decile 6 Income <dbl>, Decile 7 Income <dbl>, Decile 8 Income <dbl>,
## #   Decile 9 Income <dbl>, Decile 10 Income <dbl>, Mean Income <dbl>,
## #   Population <dbl>
```

```
#Now we use loops to complete our task. We begin by creating a new variable in our dataset, gini, which
decile_data$gini <- 0
#Now we use a loop to run through all the rows in our dataset (country-year combinations). For each row
#The function that calculates Gini coefficients from a vector of numbers is called Gini,and we apply it
# Give us the number of rows in decile_data
noc <- nrow(decile_data)

for (i in seq(1, noc)){
  # Go to Row I to get the decile data
  decs_i <- unlist(decile_data[i, 3:12])
  decile_data$gini[i] <- Gini(decs_i)
}
#With this code, we calculated 4,799 Gini coefficients without having to manually run the same command
#First we use the subset function to select Nordic countries and save their data as temp_data. As an ex
temp_data <- subset(
  decile_data, Country %in% c("United States","Sweden","Finland","Norway",
                              "Denmark"))
#Now we plot the data using ggplot.

ggplot(temp_data,
       aes(x = Year, y = gini, color = Country)) +
  geom_line(size = 1) +
  theme_bw() +
  ylab("Gini") +
  ggtitle("Gini coefficients for Nordic countries")
```

```
## Warning: Removed 11 row(s) containing missing values (geom_path).
```




```
#This example is based on great webpages of CORE: https://www.core-econ.org/doing-economics/book/text/0
```

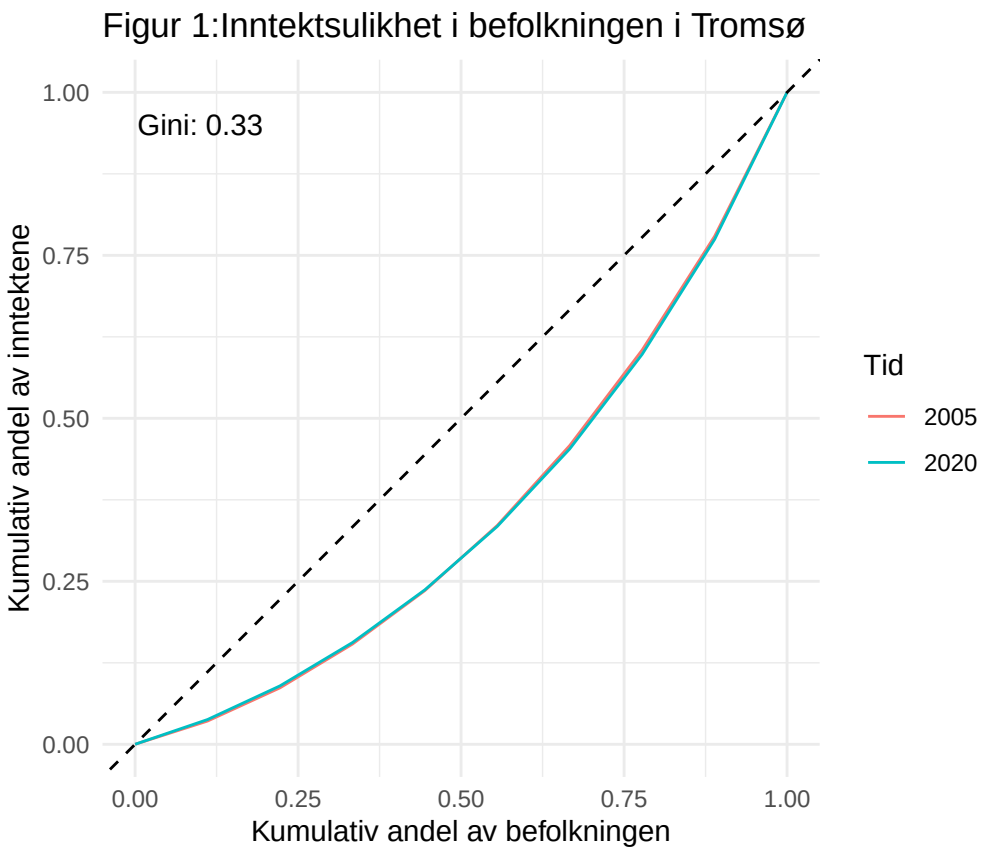
Oppgave 6:

```
# Laster ned nødvendige pakker
library(gglorenz)
library(PxWebApiData)

#Hvilke variabler som finnes i tabellen
variables <- ApiData("https://data.ssb.no/api/v0/en/table/12558/",
                     returnMetaFrames = TRUE)
names(variables)
```

```
## [1] "Region"       "InntektSkatt" "Desiler"      "ContentsCode" "Tid"
```

```
#hvilke verdier har ulike variablene
values <- ApiData("https://data.ssb.no/api/v0/en/table/12558/",
                  returnMetaData = TRUE)
#Kommunekoder
# values[[1]]$values

#Inntekt før/etter skatt
```

```r
#values[[2]]$values # 00 = Samlet inntekt, 00S=Inntekt etter skatt

#Desiler
#values[[3]]$values

#Statistikkvariabel
#values[[4]]$values

#År
#values[[5]]$values

data <- ApiData("https://data.ssb.no/api/v0/en/table/12558/",
                Tid =c("2005", "2020"), # Velg årene 2005 og 2020
                Desiler=c("01", "02", "03" ,"04", "05", "06" ,"07", "08" ,"09", "10"), #Vi velger alle
                InntektSkatt="00", #Vi velger samlet inntekt
                ContentsCode="VerdiDesil", #Velger den høyeste verdien i desilen
                Region=c("5401","1902")) #Tromsø endret kommunenummer i 2020


# Henter fram tabellen
tabell<- data[[2]] %>%
  drop_na(value)

# Lager Lorenz-kurve
tabell %>%
   ggplot(aes(value, color = Tid)) +
  stat_lorenz(desc = FALSE) +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "Kumulativ andel av befolkningen",
       y = "Kumulativ andel av inntektene",
       title = "Figur 1:Inntektsulikhet i befolkningen i Tromsø") +
  annotate_ineq(tabell$value)
```


Figur 1:Inntektsulikhet i befolkningen i Tromsø
Gini: 0.33
Kumulativ andel av inntektene
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Kumulativ andel av befolkningen
Tid
2005
2020