

Eurostat EDA

Sindre H. Øveraas, Alen Colakovic & Sebastian M. Fløysand

Eurostat EDA

Meta-data description

Nama 10r

Regional accounts are based on the same definitions and concepts as national accounts, this is because they are a specification of the national accounts.

Gross domestic product (GDP) is the standard measure of the value added created through the production of goods and services in a country during a certain period. As such, it also measures the income earned from that production, or the total amount spent on final goods and services (less imports). While GDP is the single most important indicator to capture economic activity, it falls short of providing a suitable measure of people's material well-being for which alternative indicators may be more appropriate.

@GDPSpendingGross

Our goal with this task is to compare the GDP's of the different regions within nations. We will dig deeper into the evolution of the GDP to each of our given nations as well as some of their most intriguing regions.

Sub-national GDP

Country codes: BE - Belgium, BG - Bulgaria, HR - Croatia, IT - Italy, AT - Austria, SE - Sweden, RS - Serbia

These are the RStudio packages we install and load to do our calculations.

```
library(tidyverse)
library(vtable)
library(dineq)
library(dplyr)
```

We downloaded from Eurostat our subset of countries sub-regional GDP (nama_10r_3gdp__custom_356493 and population (demo_r_pjanaggr3__custom_3579517_linear.csv) for the years

2000 - 2020. We then calculated the GDP per capita, and named the new data set “GDP_Per_Capita”.

```
library(readr)
GDP <- read_csv('nama_10r_3gdp__custom_3564935_linear.csv')
Population <- read_csv("demo_r_pjanaggr3__custom_3579517_linear.csv")

gdpdata <- GDP %>%
  rename(Year = TIME_PERIOD, GDP = OBS_VALUE, Region = geo)

populationdata <- Population %>%
  rename(Year = TIME_PERIOD, Population = OBS_VALUE, Region = geo)

GDP_Per_Capita <- gdpdata %>%
  left_join(populationdata, by=c("Region", "Year")) %>%
  select(Region, Year, GDP, Population) %>%
  mutate(
    GDP_capita = (GDP * 1000000)/Population)

```

To report our descriptive statistics on GDP per capita, we used the summary command. This gave us the following descriptive statistics on the variables GDP, population and GDP capita.

GDP		Population		GDP_capita	
Min.	: 74.55	Min.	: 20320	Min.	: 1087
1st Qu.	: 1738.28	1st Qu.	: 164518	1st Qu.	:17180
Median	: 5614.05	Median	: 273920	Median	:25185
Mean	: 10238.24	Mean	: 406217	Mean	:24191
3rd Qu.	: 10640.23	3rd Qu.	: 429030	3rd Qu.	:31351
Max.	:181212.88	Max.	:4355725	Max.	:72062
		NA's	:771	NA's	:771

Looking at the GDP per capita for all country’s in the dataset, we find that the difference in population from min to max is big.

The difference between Median and Mean is also relatively big witch can indicate that some of the biggest regions has much larger amount of population then the rest and therefore affects the mean and pulls it higher.

The fact that the 3rd quartile is just a few thousands away from the mean amplifies our suspicion that we have some outliers with a very high population compared to the rest.

Its reason to believe that high population equals high GDP based on the fact that it is more people that contributes to the GDP.

However, this can not be applied in every circumstances.

For example Monaco with a population in 2020 of just above 39.000 [[@MonacoPopulation2022](#)] had a GDP on 6.25 billion USD the same year versus Burundi with a GDP on 3.22 billion [[@GDPConstant2015](#)] and a population just above 11.89 million [[@BurundiPopulation2022](#)].

The GDP per capita gives us a more accurate measure.

in the model above we can see that its a big difference in GDP per capita in each region, this can be caused by a population or a cluster with rich/poor individuals.

We also had some NA values witch we chose to take out of the dataset.

NA values may come from the fact that Population or the GDP was not measured this year or was not available when the dataset was made.

To calculate the population watertight GDP Ginie coefficient, we used the following command.

```
[1] 0.2603924
```

To calculate the population watertight GDP Ginie coefficients for the European NUTS2 level, we used the following commands.

```
GDP_Per_Capita <- GDP_Per_Capita %>%
  mutate(NUTS2 = substr(GDP_Per_Capita$Region,1,4))

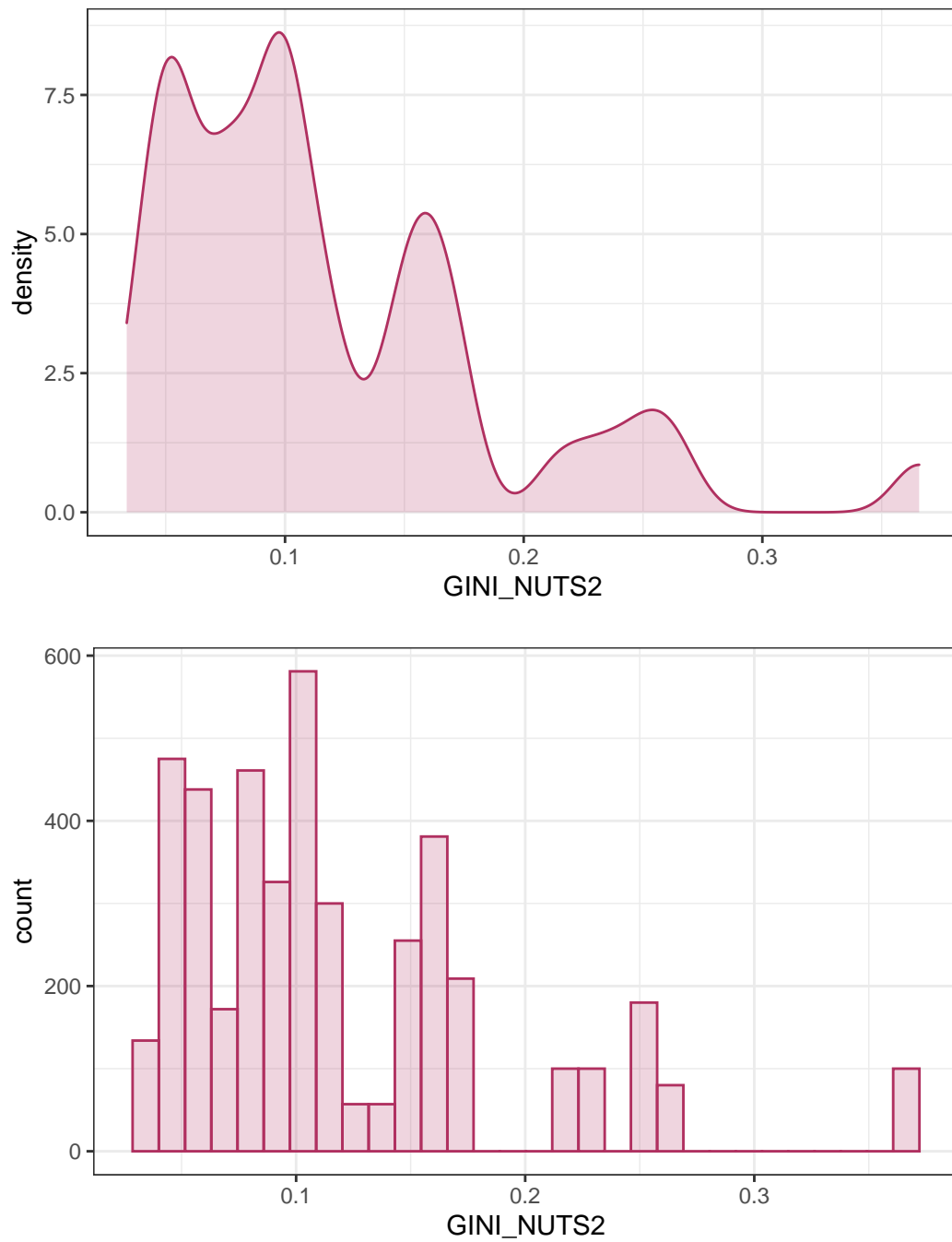
GDP_Per_Capita <- GDP_Per_Capita %>%
  mutate(NUTS = substr(GDP_Per_Capita$Region,1,2))

GDP_Per_Capita %<>%
  group_by(NUTS2) %>%
  na.exclude(GDP_Per_Capita) %>%
  mutate(GINI_NUTS2 = gini.wtd(GDP_capita, weights = Population)) %>%
  ungroup()
```

To summarize the data of the Ginie coefficients, we once more used the summarize command. We see here the gap between minimum and maximum is more then 10 times difference from 0.03367 to 0.36569 but the mean and median is at 0.09839 and 0.11800 which implies that the majority is located in closer to the minium and middle.

```
GINI_NUTS2
Min.      :0.03367
1st Qu.   :0.07065
Median    :0.09839
Mean      :0.11800
3rd Qu.   :0.15525
Max.      :0.36569
```

For visualization of our data (the Ginie coefficients for the European NUTS2 level of our selected countries), we produced a density plot and histogram, by the use of the ggplot2 package.



Looking at the plot above there is one outlier up against 0.4 with around 100 observations. The same result also seem to occur in the density plot. The majority is seen around 0.1 with a descending decline.

We made a variety of diagrams to show the difference between some countries and regions.

For this diagram we picked and divided the countries in to their respected sub-regions and catagorized them, before comparing.

Central and Eastern Europe: Croatia, Serbia and Bulgaria

South Europe: Italy

Northern Europe: Sweden

Western Europe: Austria and Belgium

(Catagorized collected from wikipedia : (**FileEuropeanSubregions?**))

With this diagram, we can see the GDP of the different areas and their rise and fall from the 2000 to 2020.

****explain what we see****

Further on we wish to highlight some of the regions within countries to see differences in GDP. There are some countries with wider gaps between them, and makes this a insightful visual representation. We dig deeper into Sweden where we used the regions: Stockholm, Uppsala and Gotland. With these we se some big gaps when it comes to GDP and how Stockholm alone raises the country´s collective GDP. In the diagram below we also show the GDP Per Capita in each region.*** The simple reason behind the difference is the location of Stockholm and that it is the capital of sweden, as well as being one of the major industrial cities in Sweden also called industrial cluster area.

Do more calulations on differences on regional levels and give further explanations

**** show 2 diagrams of gdp aswell as gdp per capita ****