Eurostat EDA

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

Eurostat EDA

In this paper we are going to present an explanatory data analysis of different statistics of selected European countries. We are more specifically going to analyze the European countries Austria (AT), Belgium (BE), Bulgaria (BG), Croatia (HR), Italy (IT), Serbia (RS), and Sweden (SE).

The paper will consist of four parts (assignments), where in the first part we will explore sub-national GDP and regional inequity in our selected countries. For this part we will use data collected from Eurostat. The data for the first part of the paper consists of GDP and population statistics for the years 2000 – 2020, on a sub-regional level i.e., at NUTS 3 level. NUTS (Nomenclature of territorial units for statistics) is the geographical nomenclature subdividing the economic territory of countries in the European Union. These levels consist of NUTS 1, 2 and 3, with 3 representing the smallest territorial units in a country (*Glossary*, 2021). The remaining parts of this paper i.e., 2, 3 and 4, will be explained continuously and gradually when we eventually get to them later in the paper.

Sub-National GDP

To start our analysis of sub-national GDP and regional inequity for our selected countries we must, as mentioned, collect data from Eurostat. We download population by broad age group and sex, as well as gross domestic product at current markets prices, at NUTS 3 level (*Population on 1 January by Broad Age Group, Sex and NUTS 3 Region*, 2022) (*Gross Domestic Product (GDP) at Current Market Prices by NUTS 3 Regions*, 2022). After we have added our two datasets to our RStudio project, we can calculate GDP per capita at the NUTS 3 level for the separate countries. This is achieved with dividing the GDP on the number of population figures, and can be presented with the following formula:

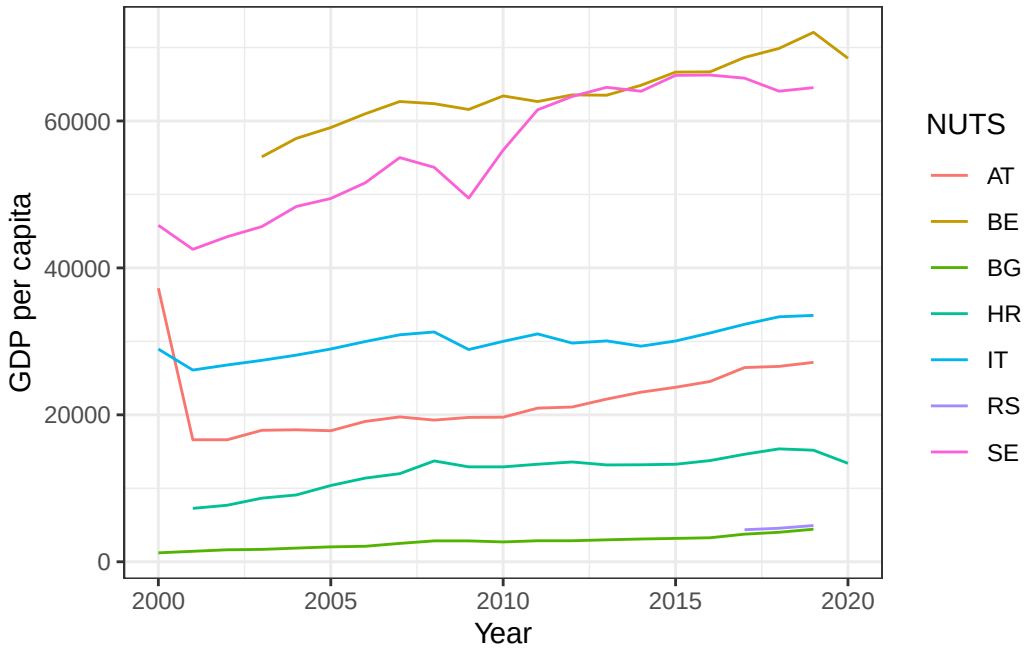
$$y_i = GDP_i / population_i$$

	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	NA's :771	NA's :771

Looking at the GDP per capita for all countries 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.



In the gg plot above we have taken out the countries that belongs to our selection - Austria (AT), Belgium (BE), Bulgaria (BG), Croatia (HR), Italy (IT), Serbia (RS) and Sweden (SE) - and added them in to a time series from year 2000 - 2020 (X-axis) to compare the evolution of GDP per capita (Y-axis) over the given time period. The data is found and downloaded from Eurostat (*Statistics | Eurostat*, n.d.)

Seven out of eight countries we where given was available for download trough Eurostat, the country witch we are missing is Bosnia and Herzegovina.

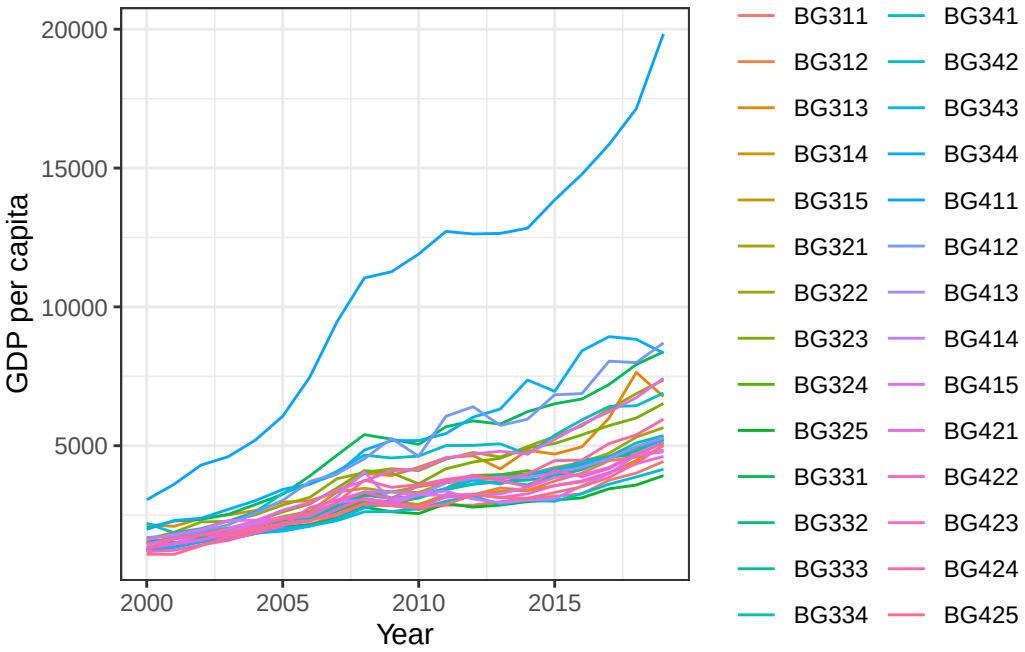
From the GG plot we can see that none of the countries have data from the whole time frame, some are missing data from the first years and others are missing from the last years. Serbia only have data from around 2017 to 2019. This might come from the fact that Serbia became an independent republic in 2006 and for that reason has not been collecting og establishing much info around GDP (*Serbia*, n.d.).

The country´s with the highest GDP per capita is Belgium and Sweden, where both countries has experienced growth in GDP per capita over time. Sweden also has had the biggest growth in GDP Per capita compared to the other countries.

From the graphs, the three countries with highest GDP per capita had a noticeable decline around 2008, which is related to the financial crises at this time. Sweden also experiences the largest dip around this time (*Economic Recovery of EU Regions After 2008*, n.d.)

Four of the five countris on the bottom also has experienced growth, just not as significant as Sweden and Belgium. Austria had a big downfall in GDP per capita from year 2000 to 2001, but has had a stable growth since then.

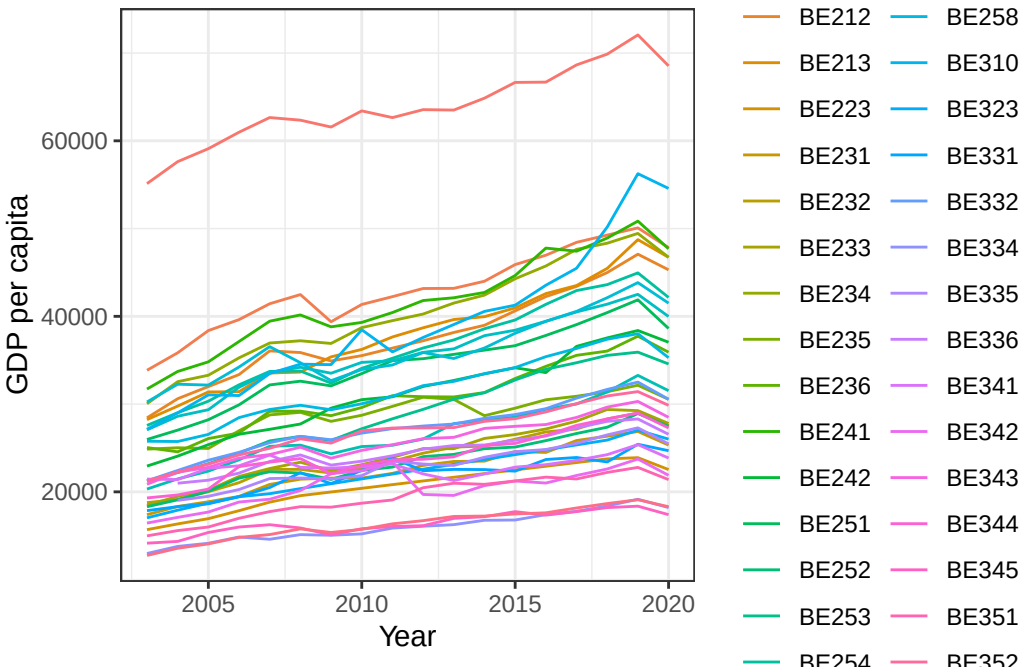
Bulgaria GDP



Region	GDP_capita
BG411	11904.952
BG344	5186.760
BG331	5048.607

Region	GDP_capita
BG325	2555.374
BG311	2701.086
BG342	2734.754

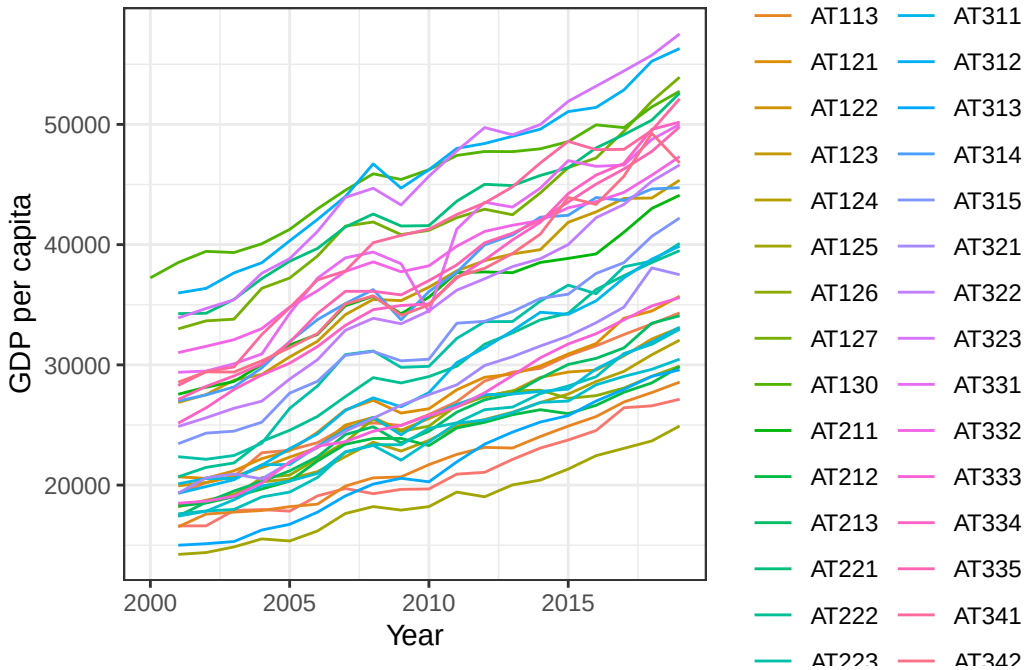
Belgium GDP



Region	GDP_capita
BE100	63408.55
BE211	41353.08
BE241	39306.81

Region	GDP_capita
BE334	15211.81
BE353	15727.11
BE345	15796.64

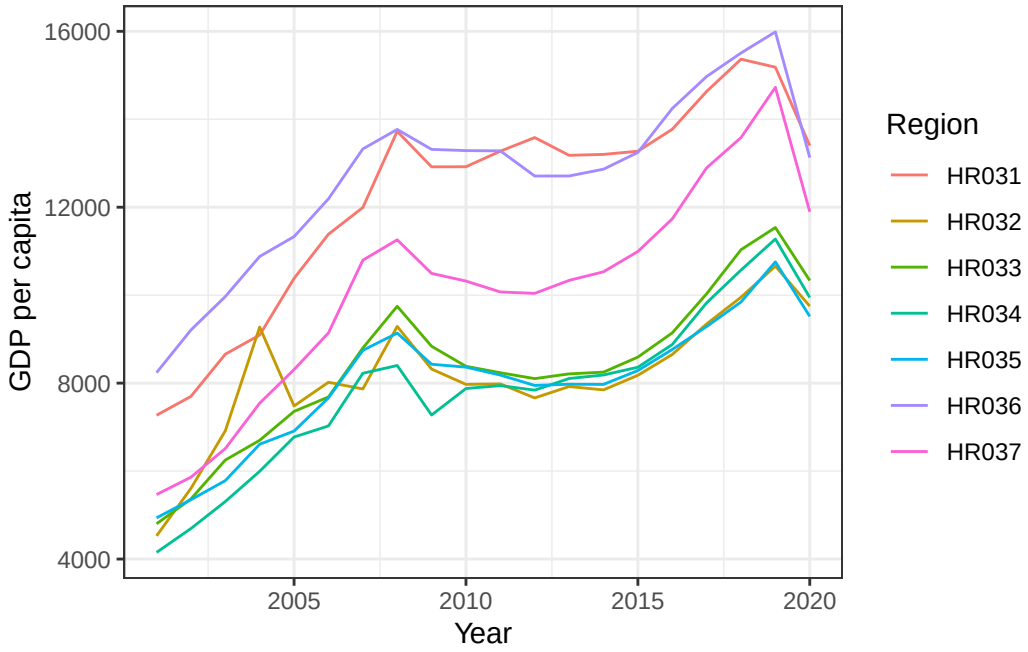
Austria GDP



Region	GDP_capita
AT312	46232.41
AT130	46230.94
AT323	45690.26

Region	GDP_capita
AT125	18212.92
AT111	19680.49
AT313	20272.85

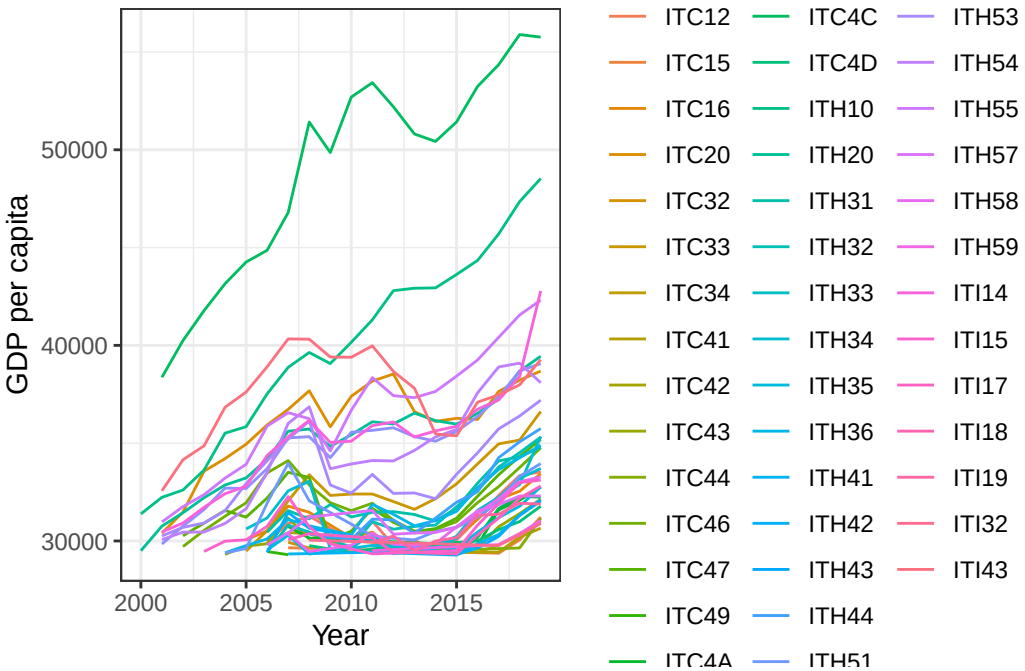
Croatia GDP



Region	GDP_capita
HR036	13287.36
HR031	12920.32
HR037	10320.23

Region	GDP_capita
HR034	7877.665
HR032	7971.199
HR035	8363.857

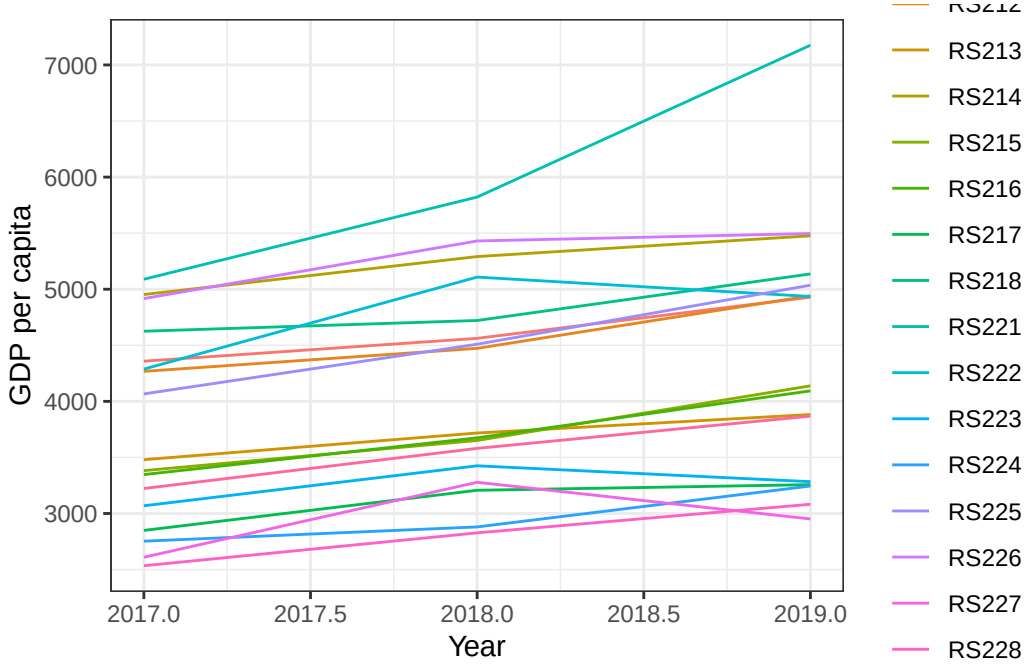
Italy GDP



Region	GDP_capita
ITC4C	52696.50
ITH10	40159.02
ITI43	39391.67

Region	GDP_capita
ITG14	14502.91
ITF48	14831.01
ITF64	15105.24

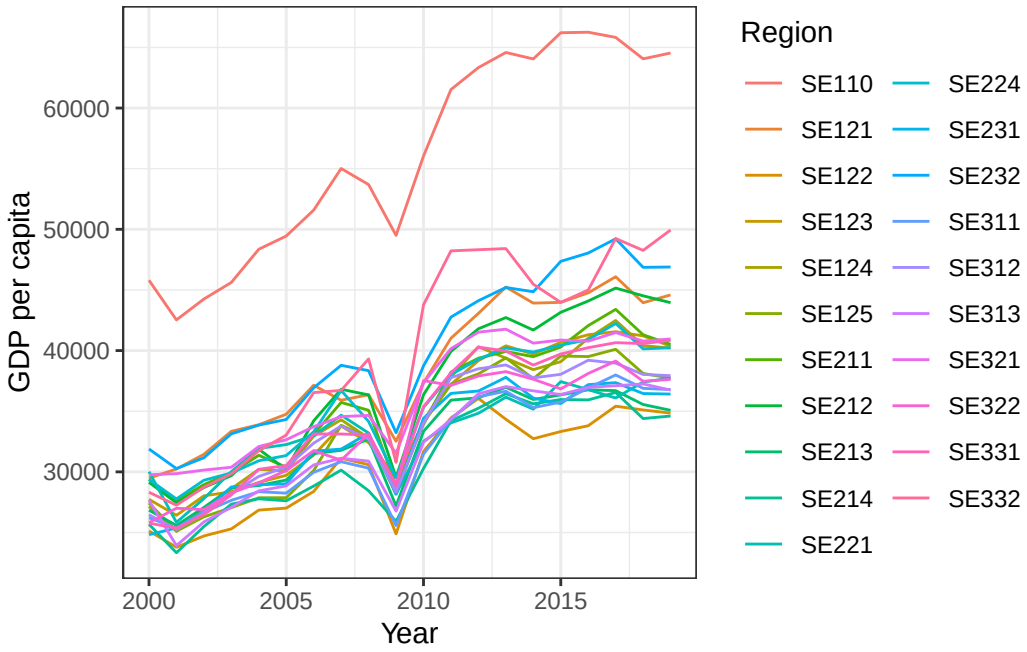
Serbia GDP



Region	GDP_capita
RS221	5088.925
RS214	4952.876
RS226	4916.722

Region	GDP_capita
RS228	2534.092
RS227	2609.823
RS224	2753.131

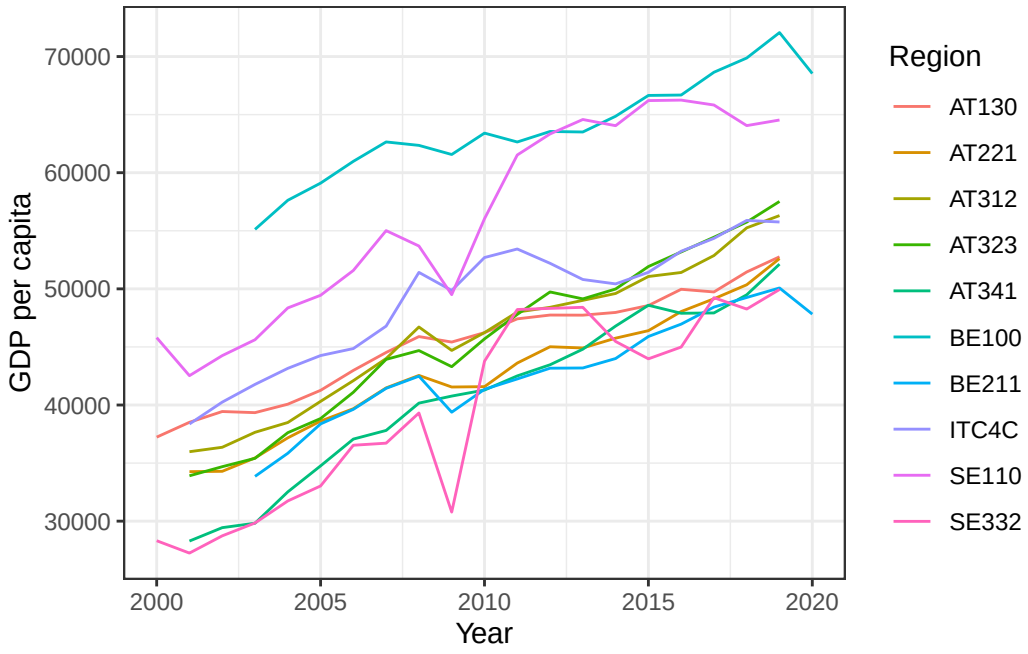
Sweden GDP



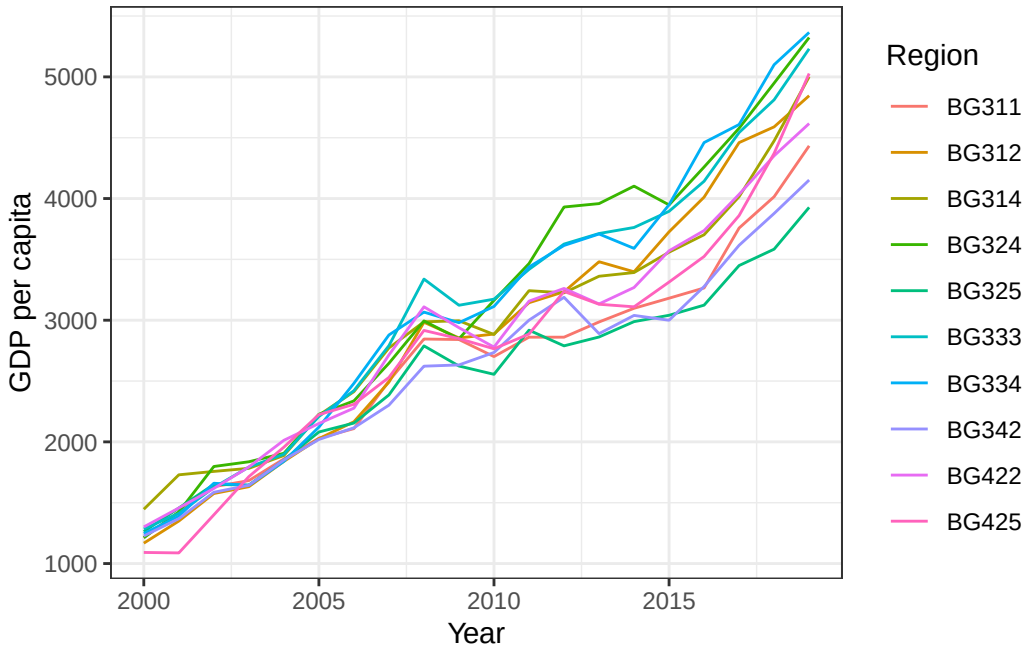
Region	GDP_capita
SE110	56036.50
SE332	43768.31
SE232	38740.05

Region	GDP_capita
SE214	30273.33
SE311	31466.64
SE122	31673.87

10 regions with highest GDP



10 regions with lowest GDP



Regional inequity

Regional inequity refers to skewed distribution and the differences in quality of life, wealth and general living standards among individuals in a region. To determine regional inequity for our selected countries, we calculate the population-weighted GDP Gini coefficients ($GINIW_j$). The coefficients can have values ranging from 0 to 1. The value 0 corresponds to no inequity, and the higher the coefficient, the higher inequity appears in the respective country/ region. The value 1 corresponds to absolute inequity. The coefficients are calculated by the following formula:

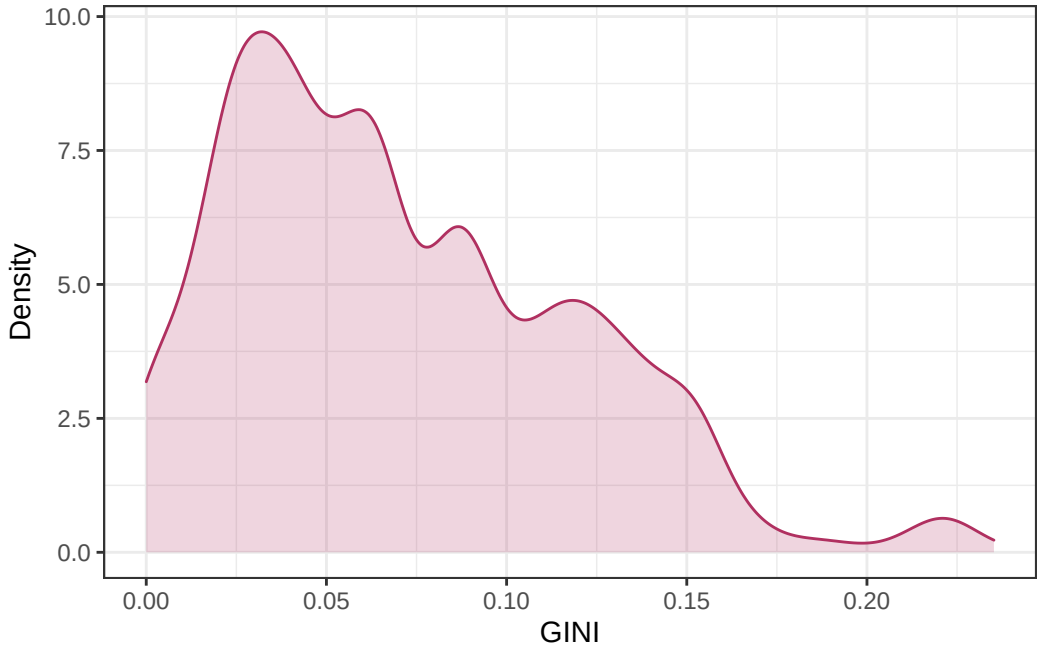
$$GINIW_j = \frac{1}{2\bar{y_j}} \sum_{i}^{n_j} \sum_{l}^{n_j} \frac{p_i}{P_j} \frac{p_l}{P_j} |y_i - y_l|$$

	GINI_NUTS2
	Min. :0.00000
	1st Qu.:0.03295
	Median :0.06233
	Mean :0.07098
	3rd Qu.:0.10441
	Max. :0.23524

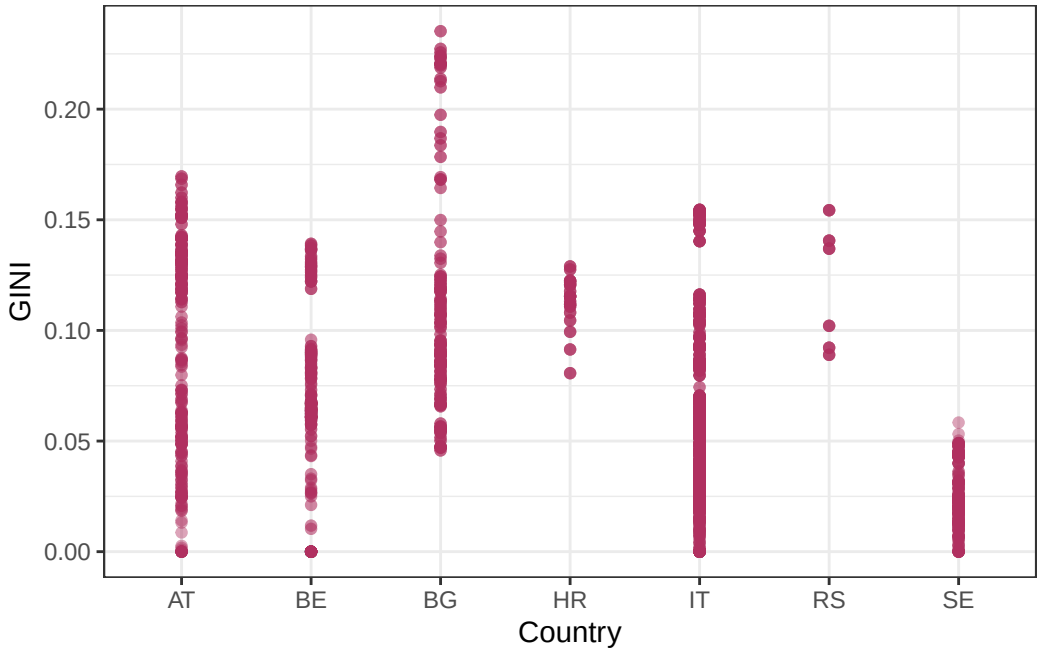
The summary statistics above give us the descriptive figures for all our selected countries combined, throughout the whole period of 2000 - 2020. Based on the statistics, we can note that the lowest coefficient is equal to 0, meaning that there are regions where inequity is absent. We can also observe that the highest inequity throughout the entire period is 0.235. The mean of the Gini coefficients are 0.071.

	GINI_NUTS2
	Min. :0.00000
	1st Qu.:0.03579
	Median :0.06505
	Mean :0.07238
	3rd Qu.:0.10428
	Max. :0.23524

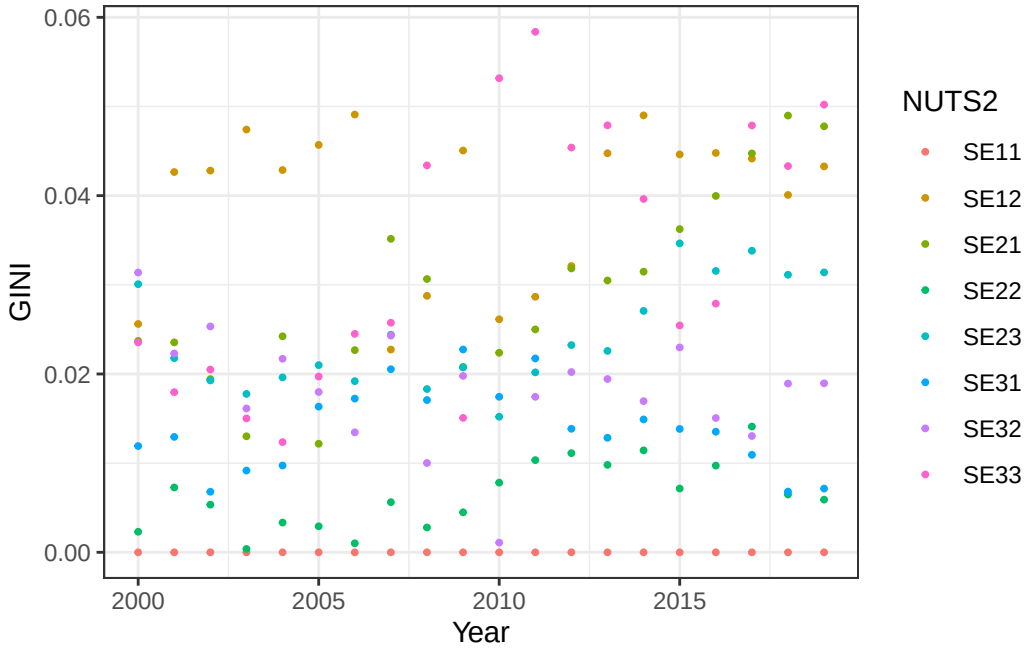
To compare to our previous on the Gini coefficient to sub-national GDP, we have taken a closer look at the year 2010 as a measurement for the two

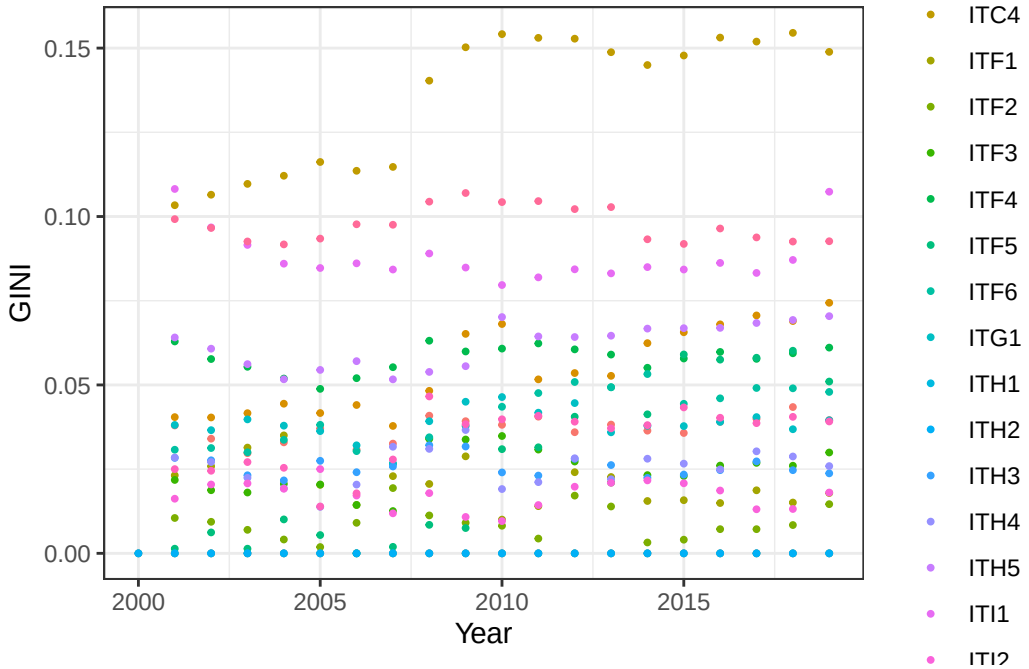


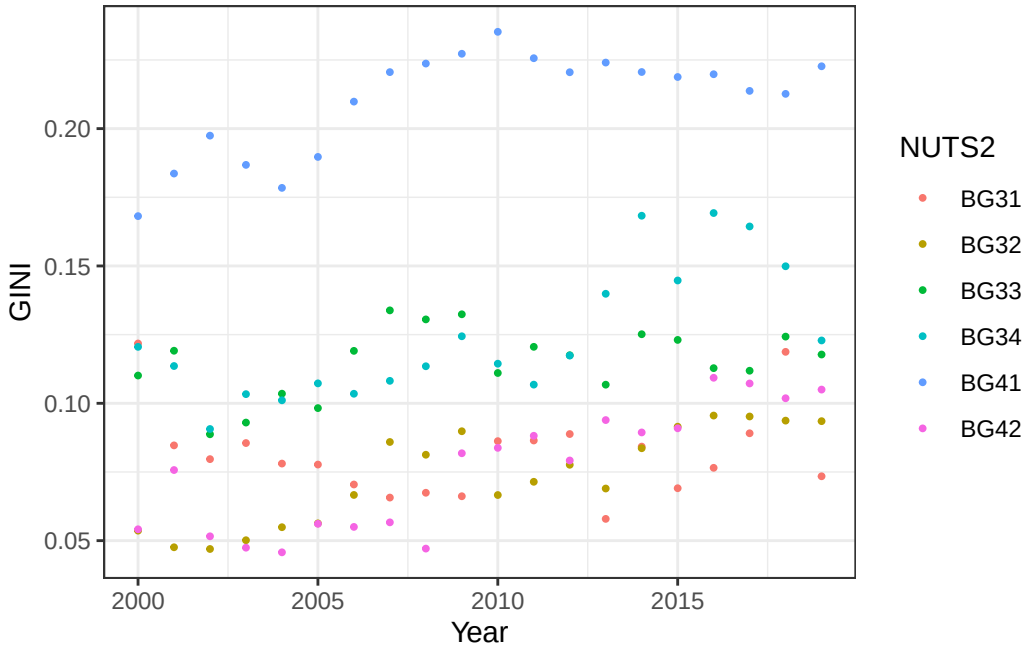
Density distribution of all coefficients in all selected countries. We can observe that the density plot above matches our summary statistics of the coefficients.



Based on the summary statistics, the density plot and the scatter plot, we can see noteable outliers. This applies in particular to Bulgaria (BG), where there are some observations if the Gini coefficient above 0.2 up to 0.235. This is noteworthy as only Bulgaria's coefficient is at this level.


0.06
0.04
0.02
0.00
GINI
2000
2005
2010
2015
Year
NUTS2
SE11
SE12
SE21
SE22
SE23
SE31
SE32
SE33


0.15
0.10
0.05
0.00
GINI
2000
2005
2010
2015
Year
ITC4
ITF1
ITF2
ITF3
ITF4
ITF5
ITF6
ITG1
ITH1
ITH2
ITH3
ITH4
ITH5
ITI1
ITI2



Economic recovery of EU regions after 2008. (n.d.). https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20191023-1.

Glossary:Nomenclature of territorial units for statistics (NUTS). (2021). https://ec.europa.eu/eurostat/statis explained/index.php?title=Glossary:Nomenclature_of_territorial_units_for_statistics_(NUTS).

Serbia. (n.d.). https://www.fn.no/Land/serbia.

Statistics | Eurostat. (n.d.). https://ec.europa.eu/eurostat/databrowser/view/nama_10r_3gdp/default/tabl

Statistics | Eurostat. (2022). https://ec.europa.eu/eurostat/databrowser/view/nama_10r_3gdp/default/tab

Statistics | Eurostat. (2022). https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjanaggr3/default/t