

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 (**GlossaryNomenclatureTerritorial2021?**). 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 (**StatisticsEurostat2022a?**) (**StatisticsEurostat2022?**). 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:

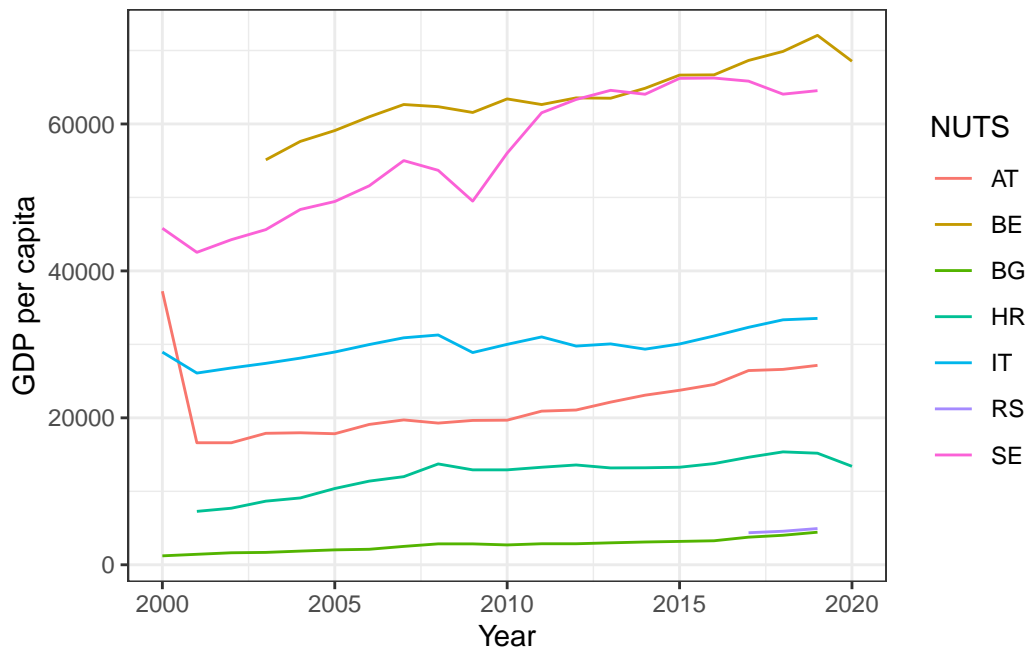
$$y_i = GDP_i / population_i$$

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 (**StatisticsEurostata?**)

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

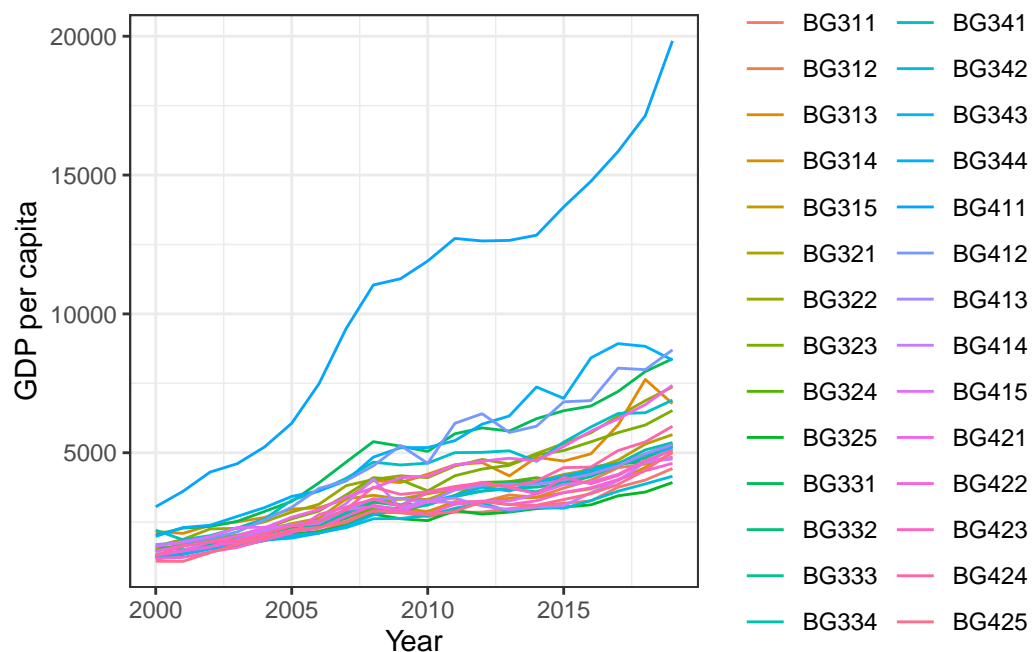
From the GG plot we can see that none of the countries have data from the whole time frame, some are missing data from the first years and others are missing from the last years. Serbia only have data from around 2017 to 2019. This might come from the fact that Serbia became an independent republic in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

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 (**EconomicRecoveryEU?**)

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



The table shows the GDP per capita on a NUTS3 level for Bulgaria alone. One Region sticks out from the rest (BG411) with a higher GDP per capita and a higher growth from 2000 to 2020, almost reaching a GDP per capita on 20 000. BG411 is a Region in Sofia, capital city of Bulgaria. The reason for this might be because of a better work situation in Sofia with less unemployed people and more business going around (**StatisticalRegionsEuropean?**).

From year 2000 to around 2005 the other regions had around the same GDP per capita with a similar growth. The following years after 2005 some of the regions have split a bit from others and had a faster growth which laded them to surpass a GDP per capita over 5000 in 2019, some reaching a GDP per capita over 7500 by year 2019.

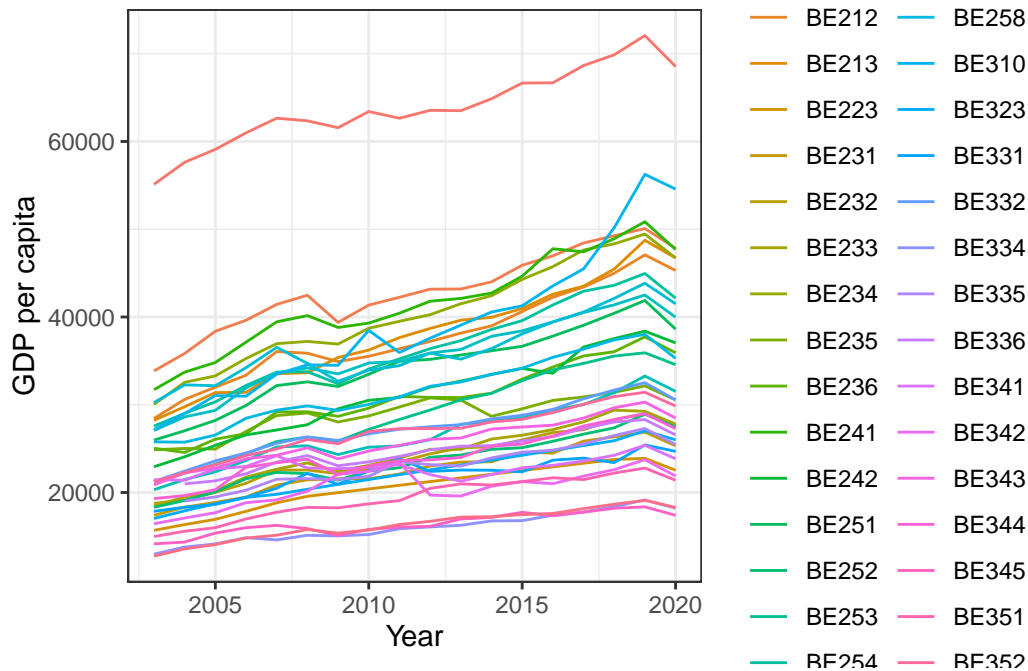
Region	GDP__capita
BG411	11904.952
BG344	5186.760
BG331	5048.607

Using 2010 as a base point which is after the financial crises, we found that the three regions with highest GDP per capita is BG411 (Sofia), BG344 (Stara Zagora) and BG331 (Varna), all major cities in Bulgaria. These Regions also managed to maintain a good growth in GDP per capita through the years till 2019

Region	GDP__capita
BG325	2555.374
BG311	2701.086
BG342	2734.754

The three regions with lowest GDP Per capita is BG325 (Silistra), BG311 (Vidin) and BG342 (Silven) where BG325 and BG311 both are cities located north in Bulgaria. BG342 is a larger city located more in the middle of Bulgaria a bit southeast.

Belgium GDP



In Belgium the table shows that it is more variance in GDP per capita from region to region. Most regions have had somewhat of a similar growth trough year 2000 until 2018 before a dip from 2018 to 2020. The dip could be explained by the government formation that started to take place in 2018. We can see that BE100 has a noticeable higher GDP per capita then the closest region. Region BE100 includes the capital city of Belgium, Brussel (**brzozowskiBelgiumBreaks16month2020?**).

Region	GDP_capita
BE100	63408.55
BE211	41353.08
BE241	39306.81

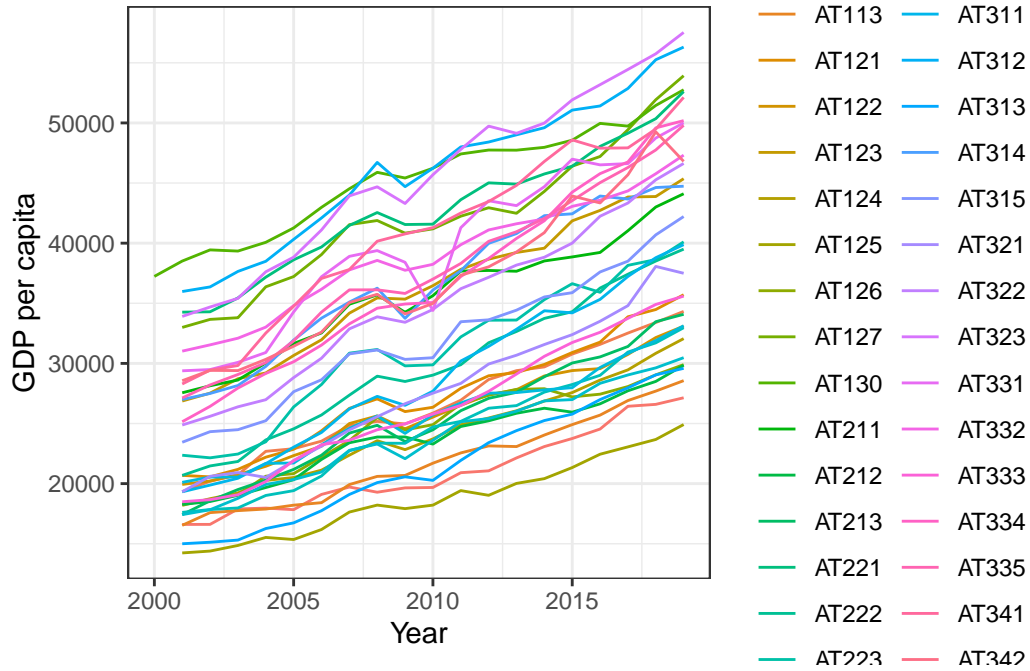
The three regions with the largest GDP in Belgium in 2010, after the financial crises, is BE100 (Brussel), BE211 (Antwerpen) and BE241 (Halle-Vilvoorde). Brussel being the capital and Halle-Vilvoorde being a region located around Brussel indicates that it is generally lower unemployment and more business in this area. Antwerpen is located more north in Belgium is a major city when it comes to seaports. Its also well known for its diamond industry (**AntwerpHistoryDiamonds?**)

Region	GDP_capita
BE334	15211.81
BE353	15727.11
BE345	15796.64

The three regions with the smallest GDP per capita in Belgium by 2010 was BE334 (Waremmme), BE353 (Philippeville) and BE345 (Virton). Philippeville and Virton is both

located north and border to France. Waremmе is located more mid-east

Austria GDP



Austria has a more aggregated graphs, even though it is a gap between highest and lowest, the GDP per capita development is more equal. We can see that some regions has experience a bit more ups and downs trough the years and most regions has a dip around the financial crises in 2008.

Region	GDP_capita
AT312	46232.41
AT130	46230.94
AT323	45690.26

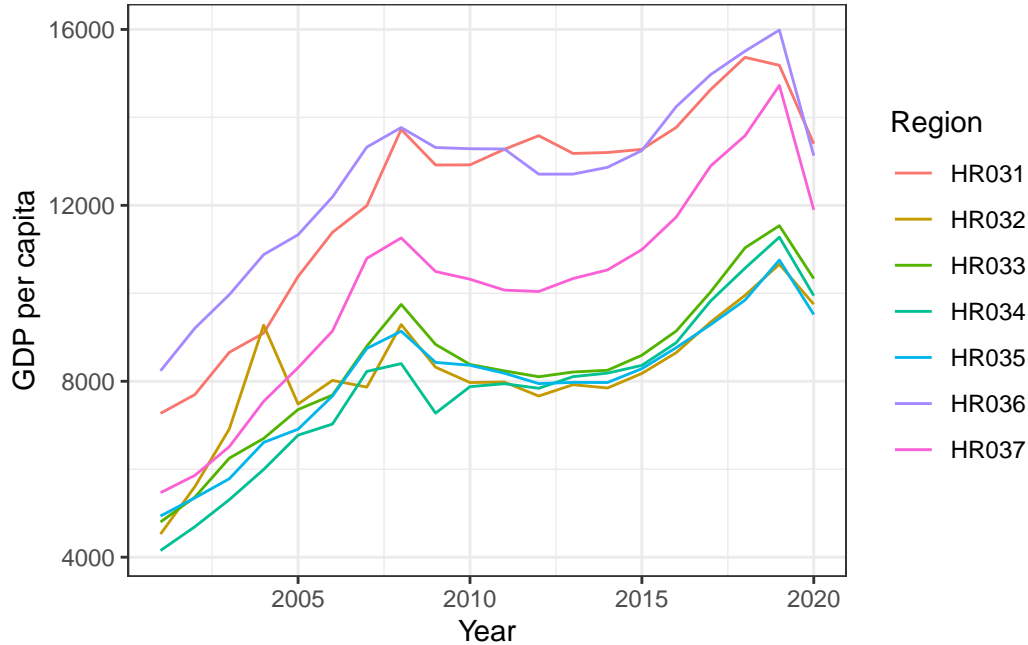
AT312 (Linz-Wels), AT130 (Wien) and AT323 (Salzburg) is the Regions with the highest scoring GDP per capita in 2010. They have held a similar pattern and GDP per capita from 2000 to 2020.

Region	GDP_capita
AT125	18212.92
AT111	19680.49
AT313	20272.85

The three lowest scoring regions are AT125 (Weinviertel), AT111 (Mittelburgenland) and AT313 (Mühlviertel) witch all is more greener areas. Both Weinviertel and Mühlviertel are

located north and borders to the Czech Republic while Mittelburgenland southeast and borders to Hungary.

Croatia GDP



The graphs for Croatia shows that the financial crises also had an impact on the GDP per capita in 2008 until 2012/2013 before the GDP per capita startet do climb again before it dips again around 2018. We can see that region HR036 and HR031 has GDP per capita that lays a bit higher than the other regions for the most part.

However, region HR037 and HR032 also had some good years. Croatia has also fewer regions with available DGP data on a nuts3 level then the other countries.

Region	GDP_capita
HR036	13287.36
HR031	12920.32
HR037	10320.23

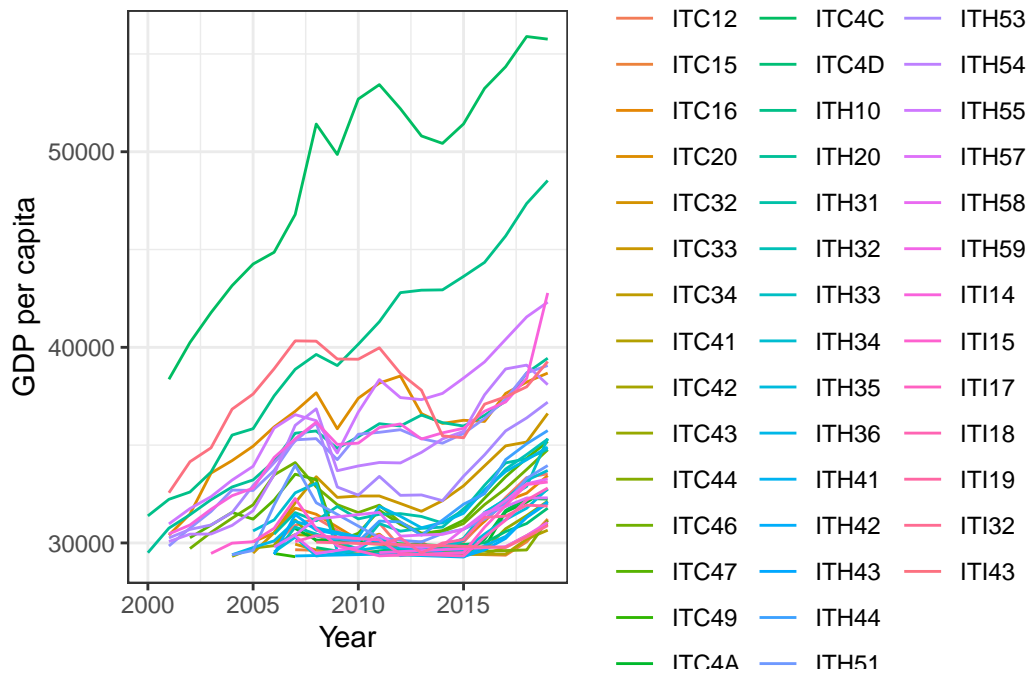
The topvthree regions in 2010 measured by GDP per capita is HR036 (istarska županija), HR031 (Primorsko-goranska županija) and HR037 (dubrovnik-neretva). istarska županija is big on processing industry, tourism and trade (**Economy?**).

Dubrovnik-Neretva is the capital of Croatias and are also big on tourism, but also makes a living out of hospitality industry, agriculture and maritime indisutry (**Economya?**).

Region	GDP_capita
HR034	7877.665
HR032	7971.199
HR035	8363.857

On the other side, the three regions with lowest GDP per capita by 2010 is HR034 (Šibensko-kninska županija), HR032 (Ličko-senjska županija) and HR035 (Splitsko-dalmatinska županija). All are located south in Croatia by the coast and borders to Bosnia-Herzegovina.

Italy GDP



Italy consists of many NUTS3 Regions. We picked out only the 500 biggest and used them in our plot. Some of the regions in the plot don't have GDP per capita data for the whole-time frame. We can see that there is a big difference in the Regions at the bottom and the one on top. The top Region ITC4C (Milano) also stands out from the rest with a higher GDP per capita at all times between year 2000 and 2020. We can also see that it is somewhat of a dip around 2011 too. This could come from Sovereign Debt (**romano2011CrisisItaly2021?**) and later on tension between Russia and Ukraine which impacted exports (**ewingItalyFallsBack2014?**).

Region	GDP_capita
ITC4C	52696.50
ITH10	40159.02
ITI43	39391.67

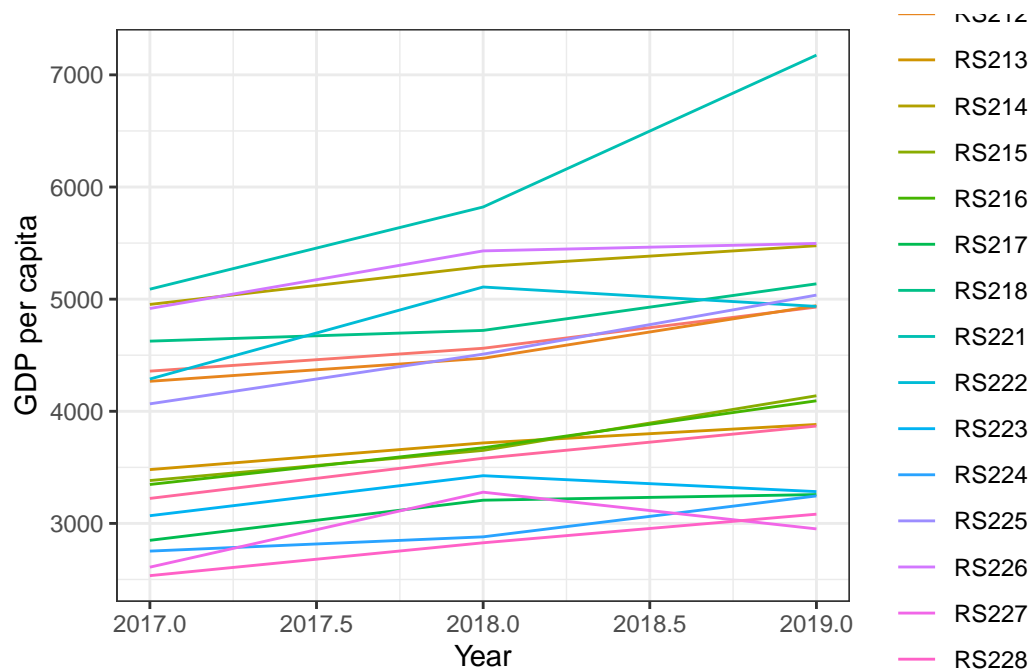
ITC4C (Milano), ITH10 (Bolzano-Bozen) and ITI43 (Rome) are the three regions in Italy with the highest GDP per capita by 2010. Milano are one of the biggest fashion cities in

the world, and arranges “Fashion-week” every year which draws visitors from all over the world (**MilanHistoryPopulation?**). Bolzano-Bozen is a rich city and focus on quality agriculture, producing wine, fruit and dairy products. Tourism is also a source of income for the city (**BolzanoBozenTrentinoAlto?**). Rome is a powerful economic force in Italy and are the center for politics. The main industry in Rome are service (**EconomyItaly?**).

Region	GDP_capita
ITG14	14502.91
ITF48	14831.01
ITF64	15105.24

Out of the 500 Regions we picked out, ITG4 (Agrigento), ITF48 (Barletta-Andria-Trani) and ITF64 (Catanzaro) had the lowest GDP per capita in 2010. All three regions are located South in Italy. Agrigento’s economic is focused around Sulfur and potash mining, tourism and agriculture (**AgrigentoItalyBritannica?**). Barletta-Andria-Trani have cities such as Barletta where the economy comes from producing cement and other similar products (**WhereBarlettaBarlettaAndriaTrani?**). The lowest scoring region, Catanzaro, have farming as one of their main industries (**CalabriaRegionItaly?**).

Serbia GDP



As mentioned earlier, Serbia only provides us GDP data from 2017-2019. Looking at the graphs, 2018 seems to be a breaking point for the GDP per capita where some regions experiences growth, while other experiences a decline in GDP Per capita.

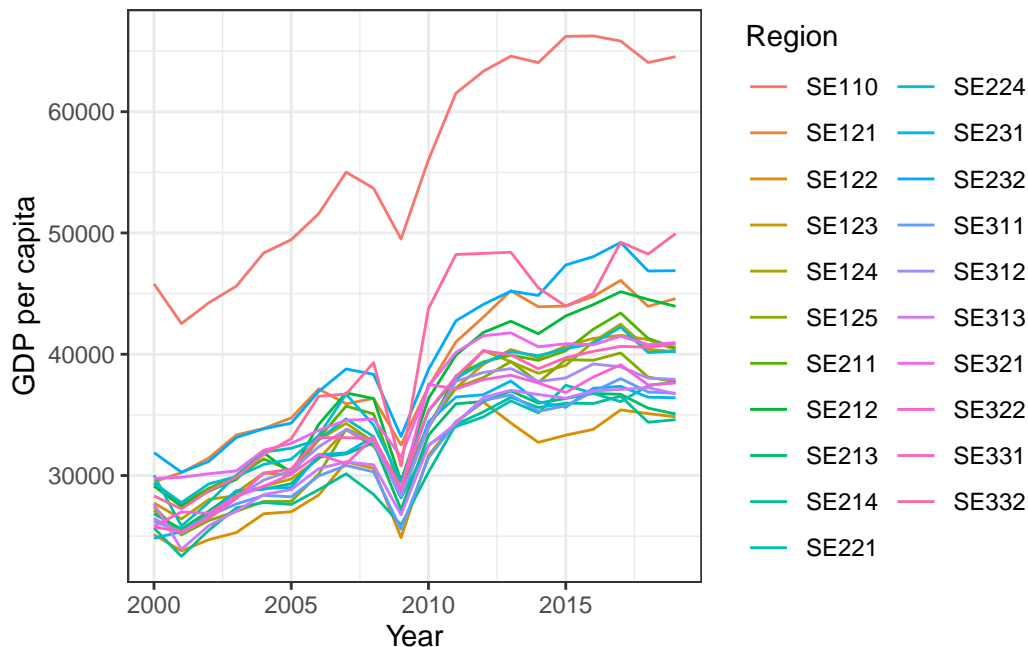
Region	GDP_capita
RS221	5088.925
RS214	4952.876
RS226	4916.722

The three regions with the highest GDP per capita in Serbia, here measured for year 2017 was RS221 (Borska oblast), RS214 (Moravička oblast) and RS226 (Pirotška oblast)

Region	GDP_capita
RS228	2534.092
RS227	2609.823
RS224	2753.131

The three lowest scoring regions in Serbia in 2017 was RS228 (Pčinjska oblast), RS227 (Pirotška oblast) and RS224 (Jablanička oblast)

Sweden GDP



The Swedish statistics show Stockholm as superior to the other regions both in GDP, GDP per capita and population. As expected from the capital of Sweden. This is where the main industrial clusters are located, which draws population and GDP higher. What is notable here is that the neighbour region SE122 (Södermanland) which borders Stockholm, is second to last at the bottom with half the GDP per Capita compared to Stockholm. We believe this has a connection to the employment rates and there would be a bigger attraction for workers to employ in Stockholm. Another interesting bit about the Swedish GDP per capita is that their smallest region by population (2019) is Gotland (SE214) with 59 249, comparing Gotland to Stockholm there is almost 40x more people in Stockholm (2 344 124).

The GDP is 2050,4 against 151 284,4 which is aprox a 74x difference in favor Stockholm. This makes the GDP per capita 34 603 against Stockholm's 64 537, less then a 2x difference.

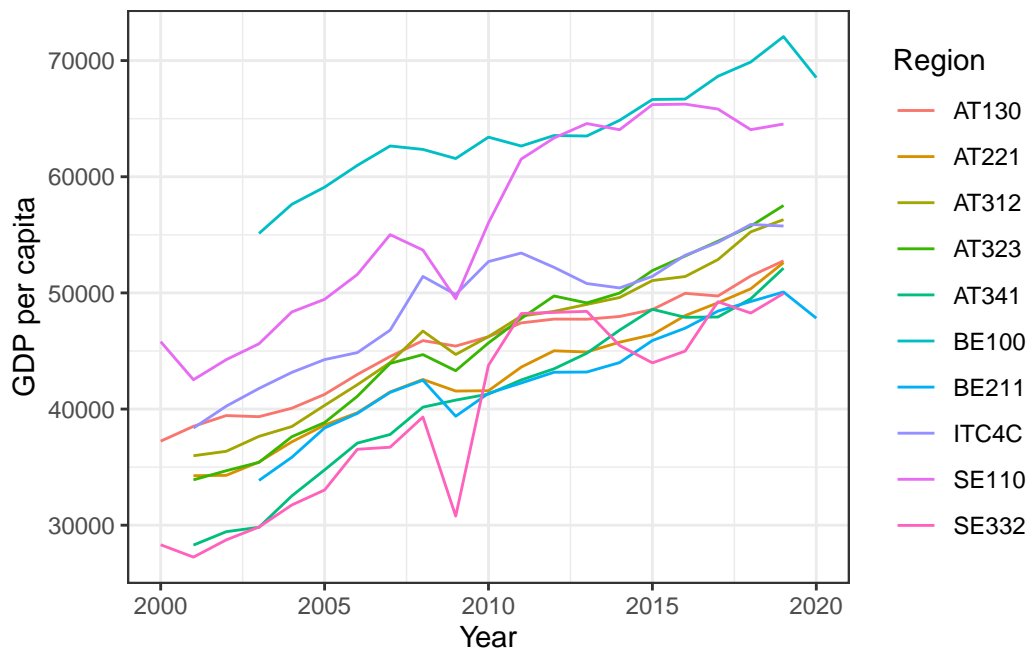
Region	GDP_capita
SE110	56036.50
SE332	43768.31
SE232	38740.05

As previously we chosed 2010 as point, where the recovery from the financial crisis was in route. Stockholm had a GDP per capita of 56036,50 at that time while Norrbottens came in second with 43768,31 and Vastra Gotaland at third with 38740,05.

Region	GDP_capita
SE214	30273.33
SE311	31466.64
SE122	31673.87

The regions with the lowest GDP per capita in 2010 were fairly close, and little that differed them from eachoter aswell as some other regions.

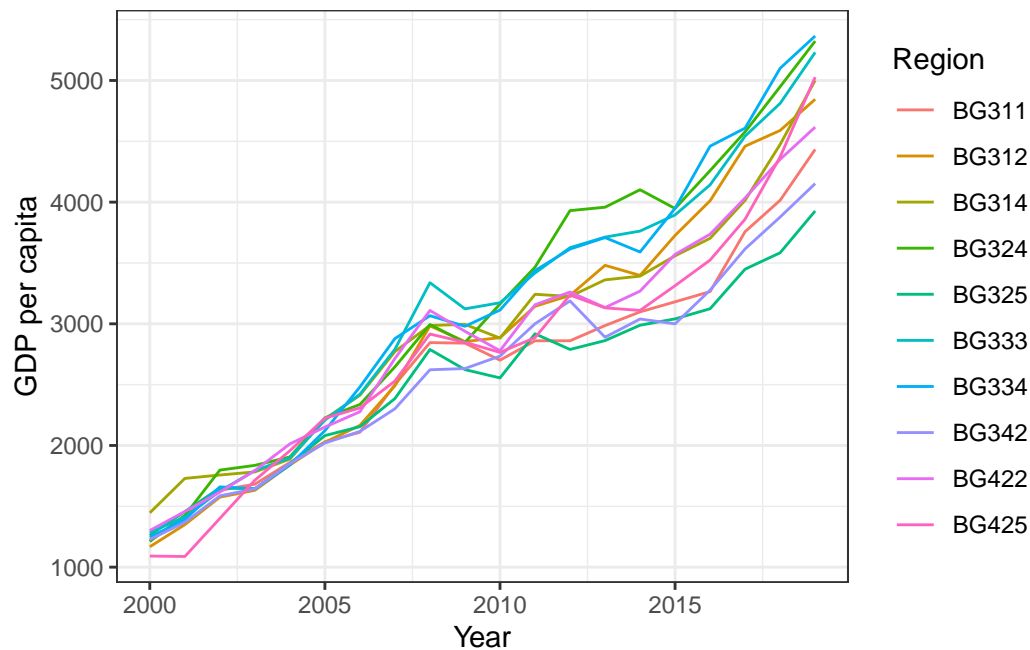
10 regions with highest GDP



Comparing the regions with the highest GDP per capita between the countries, we see BE100 (Brussels) with a GDP that was rising well above 70 000 before the corona pandemic, and was slightly lower than 70 000 in 2020. From this dataset among the highest gdp per capita regions, it was only Brussels and Antwerp that had data available until 2020. AT323 which

is Salzburg, more concretely Salzburg und Umgebung (@NUTSStatisticalRegions2022), had a population of 366'866 in 2019, which is a great difference compared to SE110 - Stockholm with a population of 2'344'124 in 2019. A considerable difference, showing how the GDP per capita is fairly close to each other even though the aforementioned population difference.

10 regions with lowest GDP



We want to clarify that there was insufficient data when it came to Bosnia & Hercegovina, which could have been a contender for the lowest GDP per Capita in some regions. Same goes for Serbia, where we chose to withhold them from this top 10 list because a lack of information pre- 2017.

The 10 regions with the lowest GDP per capita shows us they are *all* located in Bulgaria. Eight of the lowest GDP per capita regions are located in Nuts 1 - BG3 (Northern & Eastern Bulgaria). BG 3 is divided into 4 Nuts 2 regions : (BG31)Northwestern, (BG32)Northern Central, (BG33) Northeastern, (BG34) Southeastern. BG31 has three provinces that are on our top 10list; Vidin Province(311), Montana Province(312) and Pleven Province(314). It is clear to see that the pattern for where the GDP per capita is low, is around the Eastern areas, Northeast, central east and southeast. These are also some of the areas that are furthest away from the Capital. The interesting part of this diagram is that around 2004, the GDP per capita of the regions were fairly aligned with each other at 2000, before they started ascending and increasing the gap.

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:

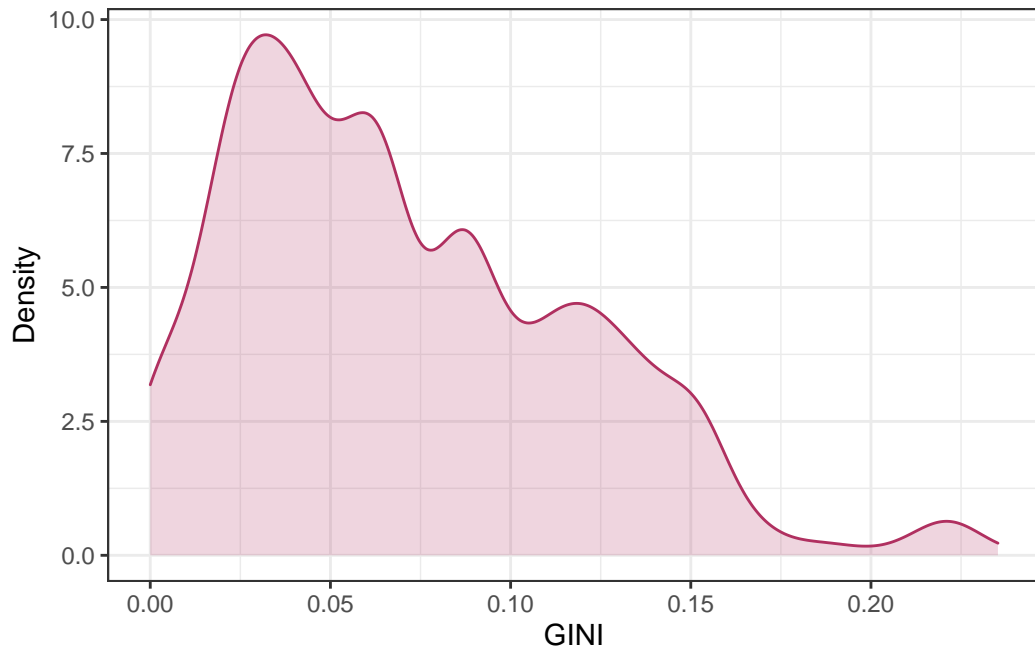
$$GINIW_j = \frac{1}{2\bar{y}_j} \sum_i^{n_j} \sum_l^{n_j} \frac{p_i}{P_j} \frac{p_l}{P_j} |y_i - y_l|$$

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

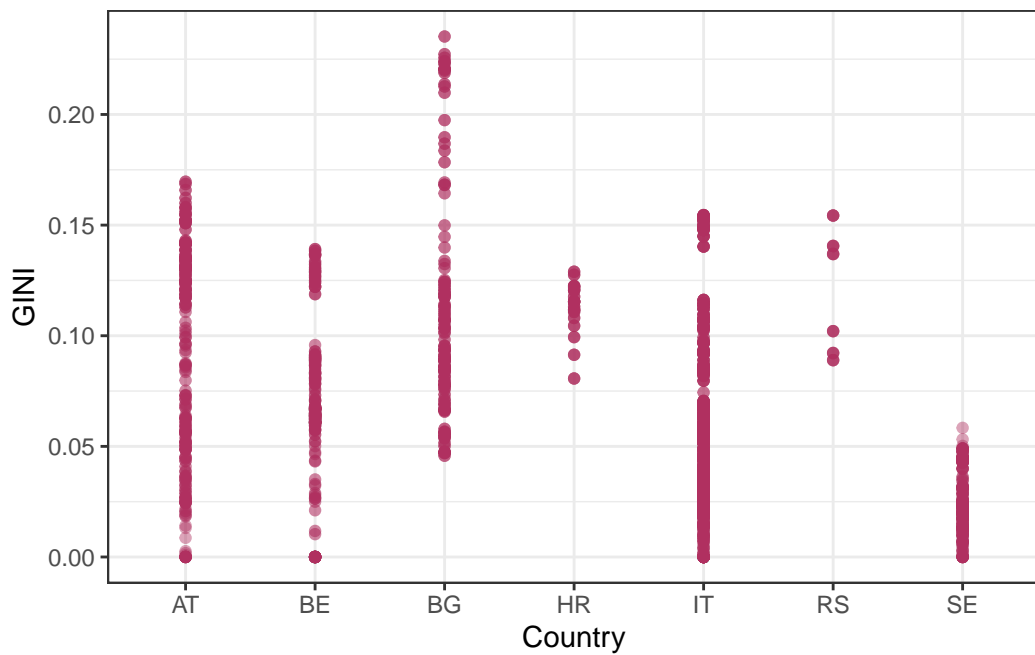
The summary statistics above give us the descriptive figures for all our selected countries combined, throughout the whole period of 2000 - 2020. Based on the statistics, we can note that the lowest coefficient is equal to 0, meaning that there are some regions where inequality is absent. We can also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

Based on this data we see that all of the countries are collectively within adequate equality or better. where the min, 1st q, median and mean are all below the 0,1 point. The max is at 0,23524 which you can see on the graphs under, and is within the Bulgarian regions. Seeing how all of the top 10 regions that scored the lowest were from Bulgaria, it does not come of as a shock, that they also have the highest GINI score. The capital of Bulgaria, Sofia, has more than twice of the average GDP per capita of the remaining regions, and for some of the lower regions, more than 5times the amount.

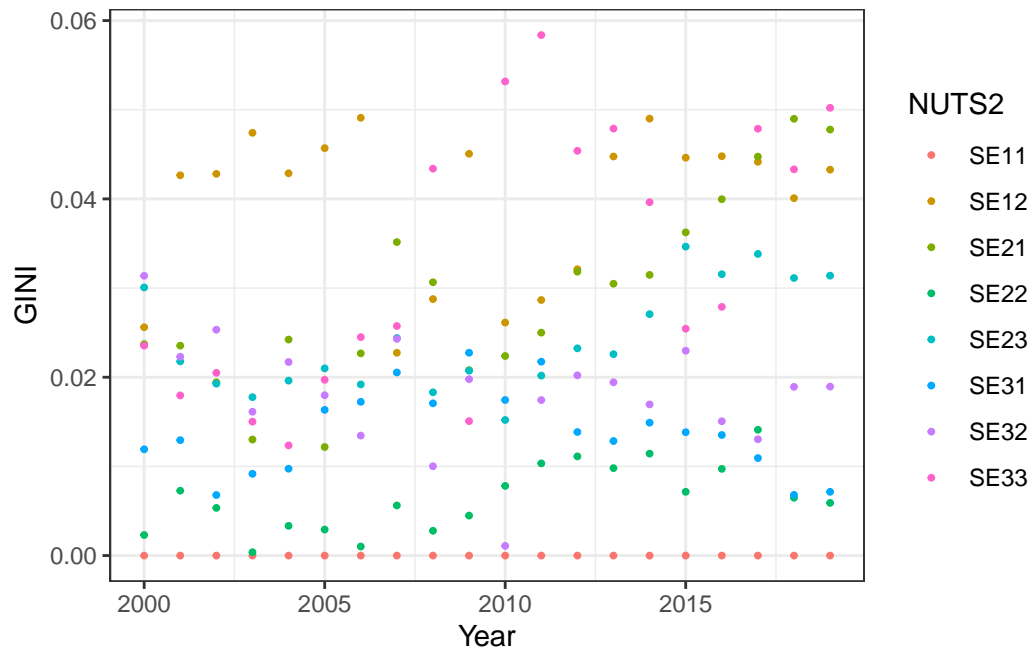


Density distribution of all coefficients in all selected countries. We can observe that the density plot above matches our summary statistics of the coefficients. The density is at peak, at approximately 0,03 GINI, reaching almost 10 in density. this signalizes that there is adequate equality between regions.

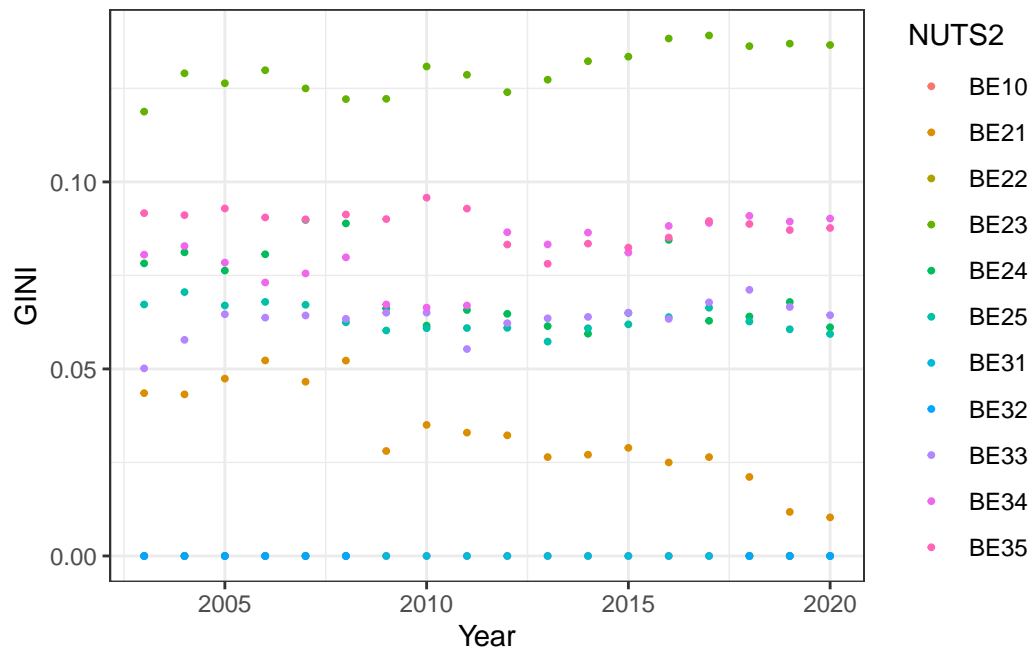


In this plot we see 4 countries touching the 0,00 GINI point, symbolizing perfect equality for some regions, with one stretching beyon 0,2 going towards adequate equality. Further

on we see Sweden being the country that scored the best with having the biggest density closest to 0,00 and we can with that confirm Sweden being the country that is closest to perfect equality between its own regions.

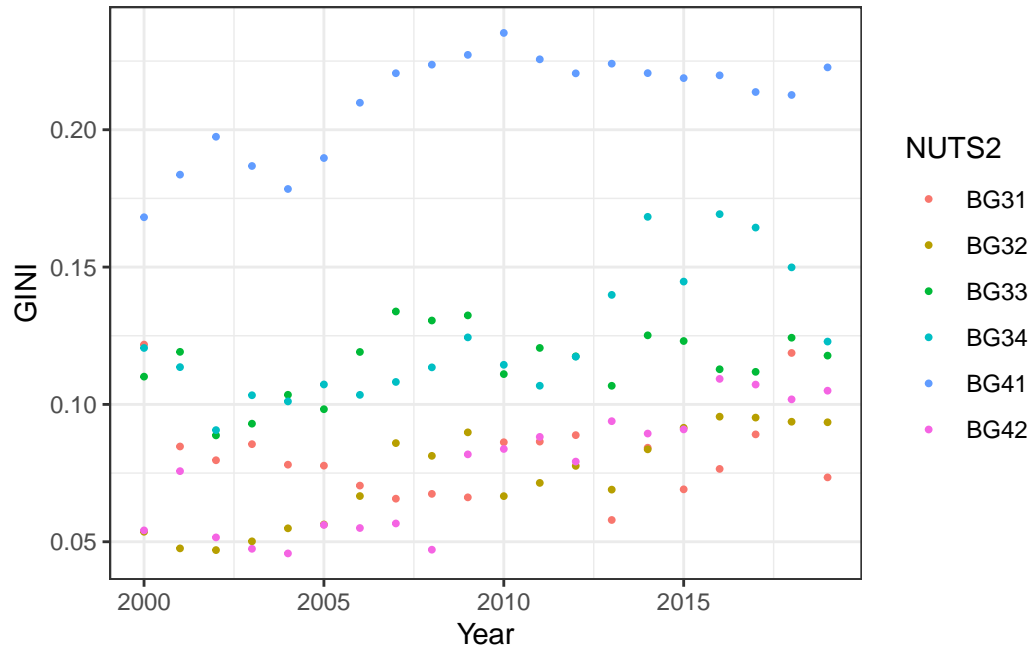


A closer look at Sweden and we see their regions and how their GINI index has been throughout the years.



Comparing this plot to the one above you see there is a bigger diversity and inequality

compared to Sweden. This is still well within limits of adequate equality, and only one region that crosses the 0,1 line.



The Bulgarian GINI plot which is the one with the biggest spread between our three plots, where the range is from 0,04 to 0,235. This visualization shows us how the plot excelled during the years and created a bigger gap with the years. The overall mean also increased with the years implying they are moving further away from perfect equality.