

# DIAML: Assignment 2

BY ANTHONY RIZKALLAH

# Table of contents

- 1 Python Packages . . . . . 3**
- 2 Question 1 . . . . . 3**
  - 2.1 Results . . . . . 3
  - 2.2 Insights . . . . . 3
  - 2.3 Results . . . . . 4
  - 2.4 Insights . . . . . 4
  - 2.5 Results . . . . . 5
  - 2.6 Insights . . . . . 5
- 3 Question 2 . . . . . 5**
  - 3.1 Results . . . . . 5
  - 3.2 Insights . . . . . 6
- 4 Question 3 . . . . . 6**
  - 4.1 Results . . . . . 6
  - 4.2 Insights . . . . . 7
  - 4.3 Results . . . . . 7
  - 4.4 Insights . . . . . 7
- 5 Question 4 . . . . . 8**
  - 5.1 Results . . . . . 8
  - 5.2 Insights . . . . . 8
  - 5.3 Results . . . . . 9
  - 5.4 Insights . . . . . 9
- 6 Question 5 . . . . . 10**
  - 6.1 Results . . . . . 10
  - 6.2 Insights . . . . . 10

# 1 Python Packages

OS

Pandas

Matplotlib

Seaborn

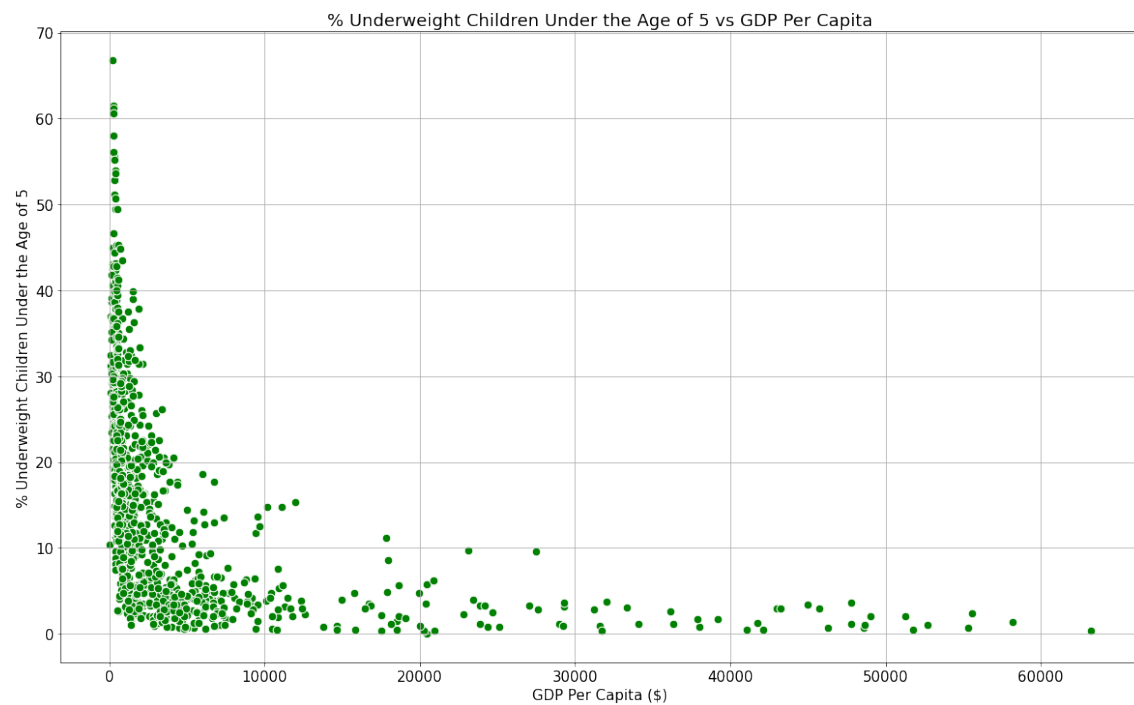
Numpy

Unidecode

Tabulate

## 2 Question 1

### 2.1 Results



**Figure 1.** This graph represents the countries that reported both the percentage of underweight children under 5 and GDP per capita. If a country does not have both values reported, then it wasn't included. Moreover, there is an inconsistent number of reporting among countries- overseas territories and subsidiaries of independent countries did not all report.

### 2.2 Insights

There is a direct association between GDP per capita and malnutrition of children under the age of 5 with some unexpectancies. The section of 0:10,000 dollar GDP per capita has a cluster of data

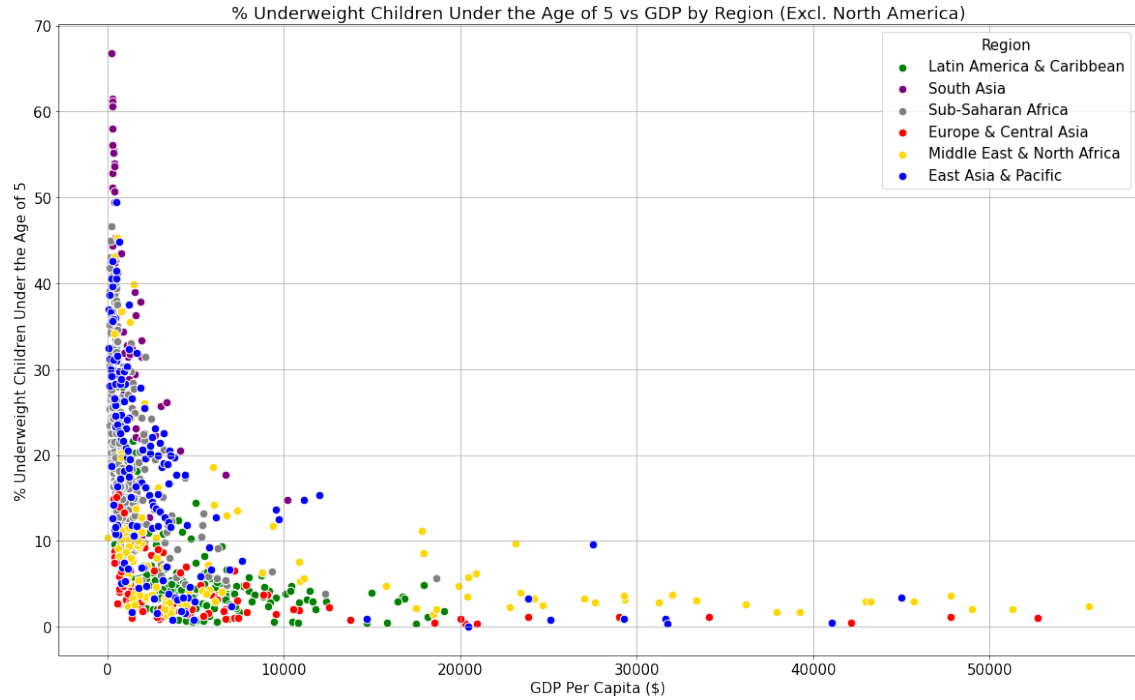
with a range of malnutrition percentages.

Malnutrtion Range	GDP Per Capita Range	Percentage of Countries
0:10%	0:10,000 Dollars	35%
10:70%	0:10,000 Dollars	43%
Rest	Rest	24%

**Table 1.** Data showing the cluster of data between 0% and 10% for the 0 to 10,000 dollar range and the data between 10% and 30% for 10,000 and 30,000 dollars is almost triple the cluster size for all the rest combined.

This table shows that most of the malnutrition is among the 0-10,000 dollar GDP per capita range with almost a 78% of the malnutrition falling in that range. That 78% is split between 35% of countries having a malnutrtition rate of 0:10% and 43% having a malnutrition rate of 10:100%.

## 2.3 Results



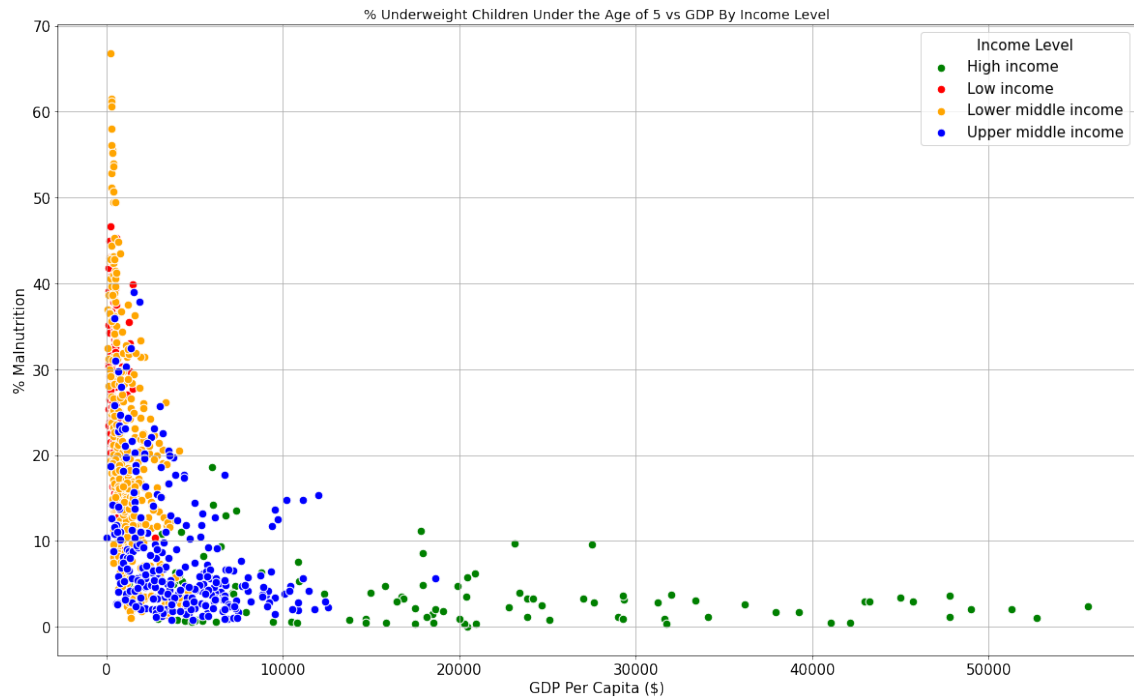
**Figure 2.** Countries and territories are selected by region according to the World Bank's categorization. This graph represents the countries that reported both the percentage of underweight children under 5 and GDP per capita. If a country does not have both values reported, then it wasn't included. Moreover, there is an inconsistent number of reporting among countries- overseas territories and subsidiaries of independent countries did not all report.

## 2.4 Insights

This graph is breaking down the % of underweight children under 5 by regions without including North America. The data shows a cluster of countries falling in that rectangle of 0:10,000 dollars and 0:70% of malnutrition. The countries with the most malnutrition are a combination of different regions with the most prominent being South Asia, East Asia and Pacific, and Sub-Saharan Africa, dominating the 10:70% range among the 0:10,000 dollar range with a few anomalies in the data. The MENA region spans a longer range of GDP per Capitas with a few countries with a malnutrition rate above 10%. Latin America despite its lower GDP per capita rate has a dominantly lower

malnutrition rate ( $<10\%$ ). Europe and Central Asia are mostly within the 0:10,000 dollar GDP per capita range, but remain among the lowest malnutrition rates as well.

## 2.5 Results



**Figure 3.** The income levels are the GDP per capita of each country, where the category represents the purchasing power of that income in the respective country. This graph represents the countries that reported both the percentage of underweight children under 5 and GDP per capita. If a country does not have both values reported, then it wasn't included. Moreover, there is an inconsistent number of reporting among countries- overseas territories and subsidiaries of independent countries did not all report.

## 2.6 Insights

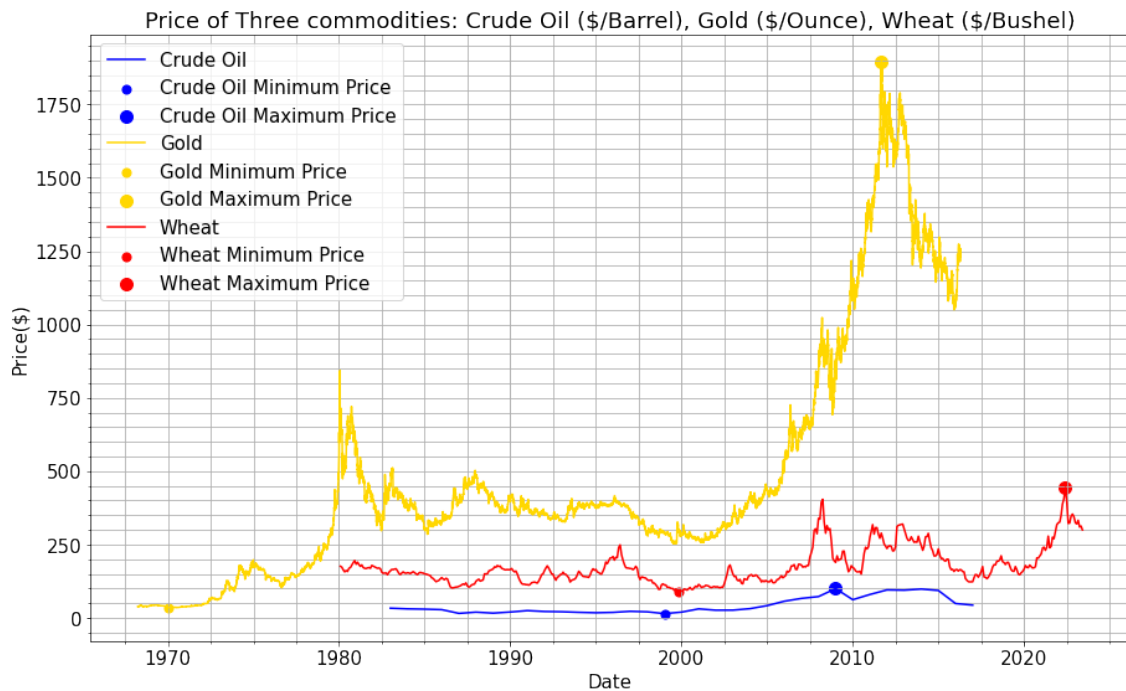
The High Income group has a mostly low malnutrition rate ranging between 0 and 10%. The Low Income group is hiding right behind the Lower-Middle Income, and they are mostly in the 10:70% range of malnutrition. The Upper-Middle Income group is dense in the lowest rates of malnutrition ( $<10\%$ ) with some anomalies  $>10\%$ .

## 3 Question 2

### 3.1 Results

The question asks to synchronize the data, but synchronization is a slightly broad term, and it was not specified whether the requirement is to trim the time series so that they all start and end at the same timestamp, or to make sure that each graph starts and end where the data is available and according to the timestamp. So, I took the liberty to represent them all on the same graph

with various start and end times even though that raises a question if this is an apple-to-apple comparison.



**Figure 4.** Price of Wheat, Gold, and Crude Oil using available data from Nasdaq Data Link. The data is synchronized, showing the minimum and maximum prices over the available data.

### 3.2 Insights

Gold has seen a drastic increase, especially between the years 1978 and 1980, where the price of one ounce of gold almost tripled, only to fall down to almost double the price of what it was in 1978. Then, Gold prices mainly remained slightly volatile, and then sloped down in the late 90's. Between 2000s and 2012, Gold had a huge rally, increasing to its highest price ever, almost 600% increase in 12 years.

A bushel of wheat did not really increase that much, remaining at a reasonable price if put into inflation perspective (not in this graph). It rallied a little bit in 2008 and 2022 possibly due to the financial crisis and post covid-19 trade effects, but came down in 2008 and is trending down in 2023.

A barrel of Crude Oil remained almost constant over the years, slightly varying in price between 1982 and 2018, with price peaks post 2008 financial crisis.

The common rallies for peak prices seem to be financial crises and global events that either destabilize economies directly or through other means.

## 4 Question 3

### 4.1 Results

Some intricacies for the countries that reported the CO<sub>2</sub> Emissions Per Capita was quite interesting. The list of countries that did not report are either subsidiaries of independent nations, states that declared their own independence from their mother nations, or overseas territories and islands. I

compiled them in a list for CO<sub>2</sub>. I could have imputed the missing values by the mean value of the available countries, but due to the difference in sizes, it shifts the standard deviation and median values. So, I omitted the NaN values and I listed the missing countries<sup>1</sup>. I also removed the NaN values from the % Enrolled Students in School because that's a lot of values to impute.

	CO <sub>2</sub> Emissions Per Capita (Metric Tonnes) in 2010
Mean	4.38
Median	2.51
Standard Deviation	5.39
5 <sup>th</sup> Percentile	0.09
25 <sup>th</sup> Percentile	0.63
75 <sup>th</sup> Percentile	6.1
95 <sup>th</sup> Percentile	15.5

**Table 2.** CO<sub>2</sub> Emissions Per Capita (Metric Tonnes) in 2010 for Reported Countries. See footnote 1 for missing countries.

## 4.2 Insights

The average CO<sub>2</sub> emissions are at 4.38 Metric Tonnes for reporting countries with a median of 2.51 Metric Tonnes. This is a big deviation between the two, with the data being positively skewed which means that there is a few extreme values pushing the mean farther away from the median. The standard deviation is at 5.39 Mt which is a relatively high number compared to a mean of 4.38 and median of 2.51. This is shown in the 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentiles, where the 5<sup>th</sup> and 25<sup>th</sup> percentiles are 0.09 and 0.63 compared to 75<sup>th</sup> and 95<sup>th</sup> percentiles at 6.1 and 15.5. This explains the skewness in the data. Particularly, the 75<sup>th</sup> percentile which means that 75% of the values are below this number and only 25% are above it, and then the top 25% are ranging from 6.1 to >15.5. Countries have two extreme directions in their CO<sub>2</sub> emissions either due to size, industrial nature of the country, climate change policy and activism, and economic development.

## 4.3 Results

	% of Enrolled Students in School in 2010
Mean	89.87
Median	92.75
Standard Deviation	10.33
5 <sup>th</sup> Percentile	64.95
25 <sup>th</sup> Percentile	87.80
75 <sup>th</sup> Percentile	96.55
95 <sup>th</sup> Percentile	99.04

**Table 3.** % of Enrolled Students in School in 2010 for Reported Countries.

## 4.4 Insights

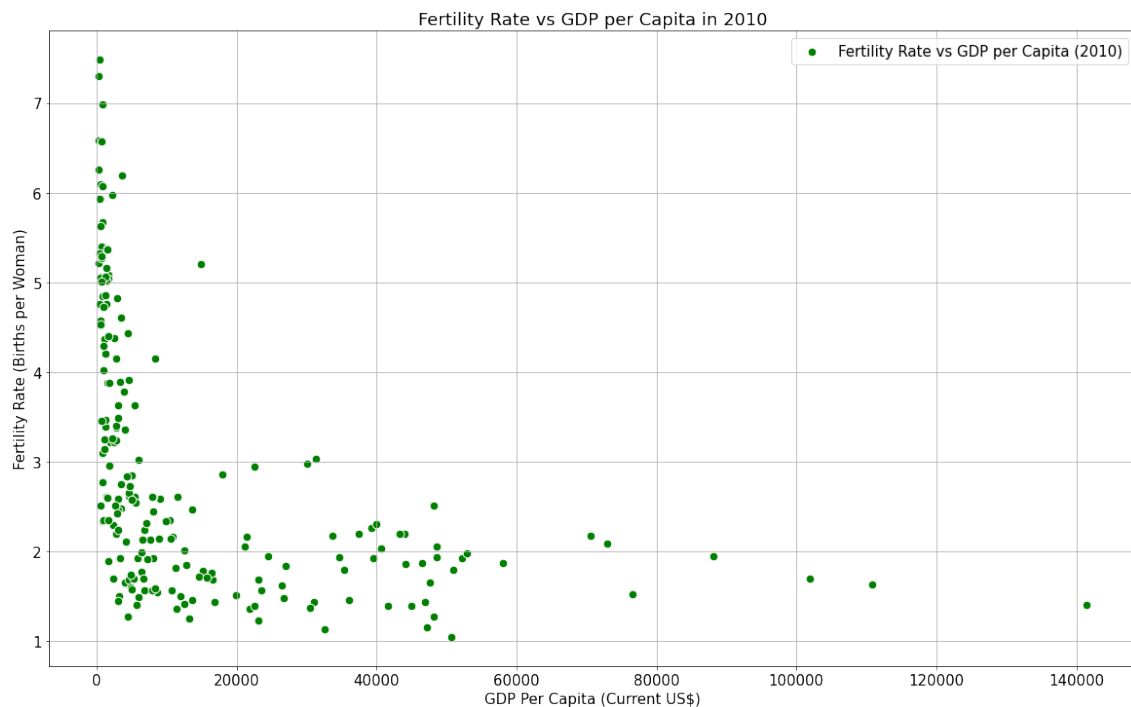
This data is quite interesting with the mean and median being very close to each other with a

1. Aruba, American Samoa, Bermuda, Channel Islands, Curacao, Cayman Islands, Faroe Islands, Gibraltar, Greenland, Guam, Hong Kong SAR, China, Isle of Man, Macao SAR, China, St. Martin (French part), Monaco, Northern Mariana Islands, New Caledonia, Puerto Rico, West Bank and Gaza, French Polynesia, San Marino, Sint Maarten (Dutch part), Turks and Caicos Islands, British Virgin Islands, Virgin Islands (U.S.), Kosovo.

slight skewness to the left of the data, meaning there is very minimal bias. The standard deviation is 10.33 which is not bad compared to the mean and medians. The variance between low and high reporting countries is not that different, and can be verified through the 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentiles. Just comparing the 5<sup>th</sup> and 95<sup>th</sup> percentiles, it is about a 50% increase. However, this is like comparing the top 5 with the bottom 5 in a race, which will obviously yield an obvious difference. Therefore, it is more useful to look at the 25<sup>th</sup> and 75<sup>th</sup> percentiles which show the small margins between the top countries in education. This means that most of the world is roughly educated with some anomalies ofcourse.

## 5 Question 4

### 5.1 Results



**Figure 5.** This graph represents the countries that reported both the fertility rate and GDP per capita. If a country does not have both values reported, then it wasn't included. Moreover, there is an inconsistent number of reporting among countries- overseas territories and subsidiaries of independent countries did not all report.

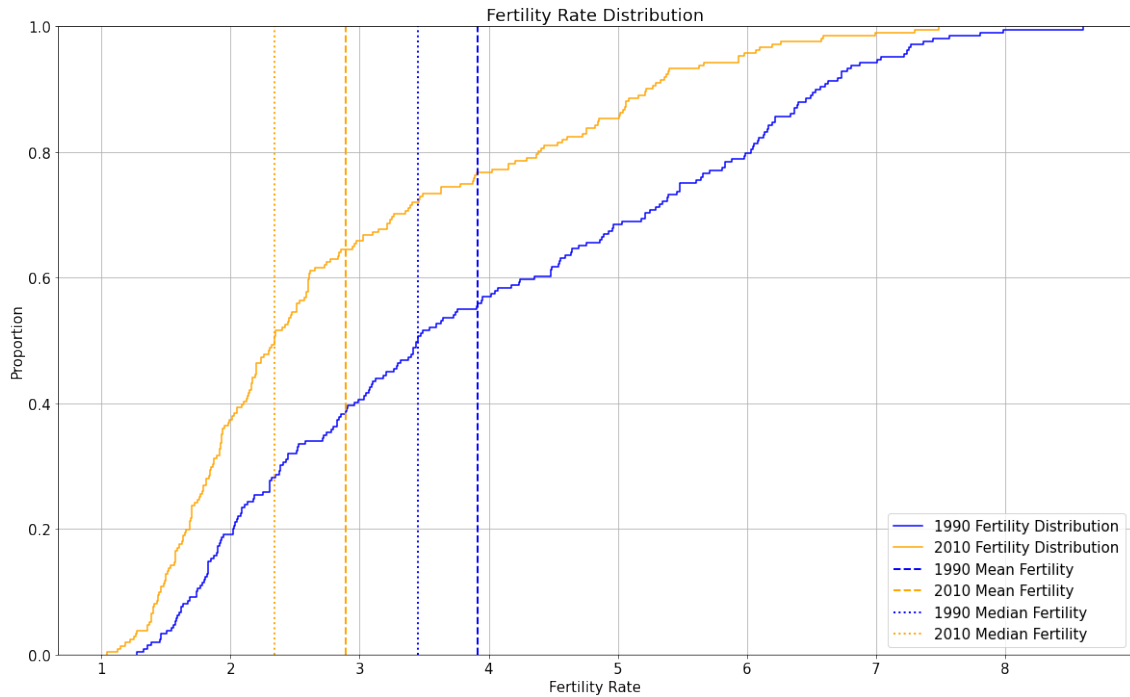
### 5.2 Insights

It seems to be that the higher the GDP per capita, the lower the fertility rate (women giving birth). This is an odd relationship if we just look at the obvious: If you make money, what's stopping you from getting kids? Maybe people who are high-earners don't want to get a lot of kids or if they



do, they get them at a higher age. In poorer countries maybe people get married at an early age and get a lot of kids. There could be a cultural aspect to it as well.

### 5.3 Results



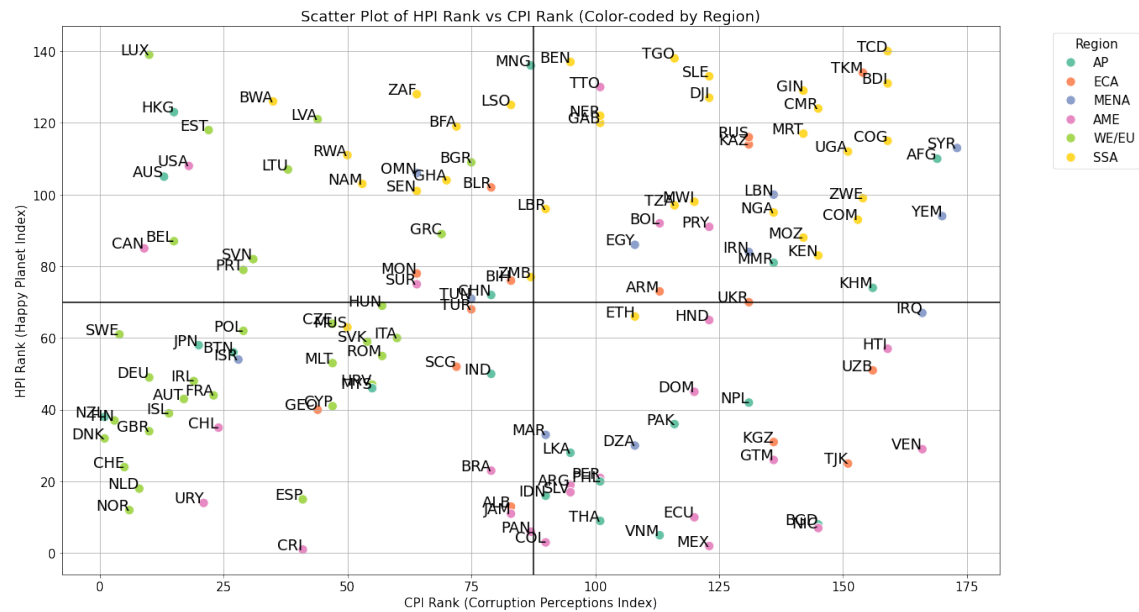
**Figure 6.** Fertility Rate distribution of reporting countries in 1990 and 2010

### 5.4 Insights

First, the fertility rate in 1990 is higher than that in 2010. This can be derived either from looking at the position of the curve, where 1990 is shifted to the right and the 2010 is shifted to the left, or through the mean and median values which follow the same behavior. Again, this could be used to test the hypothesis that when people start making more money, they decide to get less kids. The mean and median shifted from a  $\sim 3.9$  to  $\sim 2.9$  and  $\sim 3.5$  to  $\sim 2.3$ , respectively. This is a clear decrease in fertility between the two years.

## 6 Question 5

## 6.1 Results



**Figure 7.** Scatter plot of HPI Rank vs CPI Rank color-coded by regions: AP: Asia-Pacific, ECA: Europe and Central Asia, MENA: Middle East and North Africa, AME: Americas, WE/EU: Western and Eastern Europe, SSA: Sub-Saharan Africa

## 6.2 Insights

The quadrants on the graph I am most interested in are the top-left and bottom-right or (II and IV in mathematical terms) because they represent low-corruption/low-happiness and high-corruption/high-happiness. Those two quadrants are what would be “weird”. Why would people living in countries where the perception of corruption is very high score higher on happiness and vice versa?

A lot of countries that one wouldn't expect are in quadrant I: United States, Canada, Luxembourg, Australia. Also, countries that would notoriously be considered unpleasant to live in such as Iraq, Mexico, Colombia, and Nepal are scoring in the fourth quadrant, which are corrupt, but have a high-level of happiness.