**Week 1: Getting Your Research Project Started**

**Chosen dataset**

The GapMinder was the data set choosing. That data set is an independent educational non-proﬁt ﬁghting global misconceptions, it was created by Ola Rosling, Anna Rosling Rönnlund and Hans Rosling in 2005.

**Research Question**

How the economic indicators is related with the social and environmental indicators.

**Hypothesis**

The country with higher economic indicators can present goods social indicator, however the lowest sustainability indicators.

**Search terms/keywords used**

Economy indicators

incomeperperson - Gross Domestic Product per capita in constant 2000 US$.

Social indicators

lifeexpectancy - life expectancy at birth (years)

femaleemployrate - female employees age 15+ (% of population)

employrate - total employees age 15+ (% of population)

urbanrate - urban population (% of total)

internetuserate - Internet users (per 100 people)

Environmental indicators

co2emissions - cumulative CO2 emission (metric tons).

oilperperson - oil Consumption per capita (tonnes per year and person)

relectricperperson - residential electricity consumption, per person (kWh)

**Literature Review**

In the literature are many works with correlation between the economy, social and environmental indicators. Kummu and Varis (2021) use the Gapminer data set to analyse the indicators along the word latitude. Gallego (2005) use that indicators to analyze the developed in Spain. Fiorito (2013) study the ratio between the Electricity Consumption with Gross Domestic Product is related with the sustainability.

**Reference**

Kummu, M., Varis, O., 2011. The world by latitudes: a global analysis of human population, development level and environment across the north–south axis over the past half century. Appl. Geogr. 31 (2), 495–507

Gallego, I. (2006), The use of economic, social and environmental indicators as a measure of sustainable development in Spain. Corp. Soc. Responsib. Environ. Mgmt, 13: 78-97.

Fiorito, G., 2013. Can we use the energy intensity indicator to study “decoupling” in modern economies? J. Clean. Prod. 47, 465–473.

https://lcs-andrade.tumblr.com/post/665058629824708608/week-1-getting-your-research-project-started

Week 2 - Running Your First Program

The GapMinner data set have information about the social, economic and environmental development separated by countries. The Variables separated in the data set are continuous quantitative variables to looking for the answers we have proposed in the work was done a scaling in the values, to better understanding the distribution of these.

The scaling adopted to “*incomeperperson*” variable are:

|  |  |
| --- | --- |
| X = Gross Domestic Product per capita in constant 2000 US$ | |
| X < 1000 | 1 |
| 1000< X < 10000 | 2 |
| 10000< X < 100000 | 3 |
| X > 100000 | 4 |

The scaling adopted to “*lifeexpectancy*” variable are:

|  |  |
| --- | --- |
| X = life expectancy at birth (years) | |
| X < 60 | 1 |
| 60<= X < 70 | 2 |
| 70<= X < 80 | 3 |
| X >= 80 | 4 |

The scaling adopted to “*femaleemployrate*” variable are:

|  |  |
| --- | --- |
| X = female employees age 15+ (% of population) | |
| X < 50 | 1 |
| X >= 50 | 2 |

The scaling adopted to “*employrate*” variable are:

|  |  |
| --- | --- |
| X = total employees age 15+ (% of population) | |
| X < 50 | 1 |
| X >= 50 | 2 |

The scaling adopted to “*urbanrate*” variable are:

|  |  |
| --- | --- |
| X = urban population (% of total) | |
| X < 50 | 1 |
| X >= 50 | 2 |

The scaling adopted to “*internetuserate*” variable are:

|  |  |
| --- | --- |
| X = Internet users (per 100 people) | |
| X < 50 | 1 |
| X >= 50 | 2 |

The scaling adopted to “*co2emissions*” variable are:

|  |  |
| --- | --- |
| X = cumulative CO2 emission (metric tons) | |
| X < 10000 | 1 |
| 10000 <= X < 100000 | 2 |
| 100000<= X < 1000000 | 3 |
| 1000000<= X < 10000000 | 4 |
| 10000000<= X < 100000000 | 5 |
| X>=100000000 | 6 |

The scaling adopted to “*oilperperson*” variable are:

|  |  |
| --- | --- |
| X = oil Consumption per capita (tonnes per year and person) | |
| X < 0.1 | 1 |
| 0.1 <= X < 1 | 2 |
| 1 <= X < 10 | 3 |
| X>=10 | 4 |

The scaling adopted to “*relectricperperson*” variable are:

|  |  |
| --- | --- |
| X = residential electricity consumption, per person (kWh) | |
| X < 10 | 1 |
| 10 <= X < 100 | 2 |
| 100<= X < 1000 | 3 |
| 1000<= X < 10000 | 4 |
| X>=10000 | 5 |

The Code:

The Figures:

Summary:

The scaling applying in the values allow seeing how the values are distributed.

- The “*incomeperperson”* distribution show the of Gross Domestic Product per capita of 24% countries are leastwise 10 times bigger than 75% of countries, and 1% of countries are more rich (100 times) than 28% of countries. That variable have a 23 missing data.

- The “*lifeexpectancy*” distribution look similar to the “*incomeperperson”*. The most of countries the life expectancy are between 70 and 80 years. In 12% of countries the life expectancy (>80) are 20 year bigger than 20% of countries (<60). That variable have a 22 missing data.

- The “*femaleemployrate*” distribution shows the most countries, 57%, less than half of woman population are unemployed. That variable have a 35 missing data.

- The “*empoyrate*” distribution shows the most countries, 79%, over half of population are employed. That variable have a 35 missing data.

- The “*urbanrate*” distribution shows the most countries, 60%, over half of population living in the urban area. That variable have a 10 missing data.

- The “*internetuserate*” distribution shows the most countries, 60%, less than half of the population does not have an internet connection. That variable have a 70 missing data.

- The “*co2emission*” distribution shows how are discrepant in the countries. 8% of countries emit more CO2 than 1000 times than 40% of countries. That variable have a 13 missing data.

- The “*oilperperson*” distribution look similar to a co2emission. That variable have a 150 missing data.

- The “*relectricperperson*” distribution shows the 34% of countries the consumption of electricity are leastwise 10 times than .65% of countries That variable have a 82 missing data.

The propose of that work is see if the variable “*incomeperperson”* explain the discrepancy perceived in the social and environmental indicators.

https://lcs-andrade.tumblr.com/post/665579807750356992/week-2-running-your-first-program