

Writing About the Data

(Regression Modeling in Practice Week 1 Assignment)

Expected Activities

- Describe the sample, the data collection procedure, and how you managed your data

Gapminder is a non-profit venture that seeks to promote sustainable global development and achievement of the United Nations Millennium Development Goals. It seeks to increase the use and understanding of statistics about social, economic, and environmental development at local, national, and global levels.

The social and economic aspects of the data provided by Gapminder through their dataset, made me interested in it and I have decided to choose the same for further analysis and research.

Gapminder contains data for all 192 UN members, aggregating data for Serbia and Montenegro. Additionally, it includes data for 24 other areas, generating a total of 215 areas. GapMinder collects data from a handful of sources, including the Institute for Health Metrics and Evaluation, US Census Bureau's International Database, United Nations Statistics Division, and the World Bank.

The dataset contains the following parameters:

Variable Name	Description of Indicator	Main Source
incomeperperson	2010 Gross Domestic Product per capita in constant 2000 US\$. The inflation but not the differences in the cost of living between countries has been taken into account.	World Bank Work Development Indicators
alconsumption	2008 alcohol consumption per adult (age 15+), litres. Recorded and estimated average alcohol consumption, adult (15+) per capita consumption in litres pure alcohol	WHO
armedforcesrate	Armed forces personnel (% of total labor force)	Work Development Indicators
breastcancerper100th	2002 breast cancer new cases per 100,000 female. Number of new cases of breast cancer in 100,000 female residents during the certain year.	ARC (International Agency for Research on Cancer)

co2emissions	2006 cumulative CO2 emission (metric tons), Total amount of CO2 emission in metric tons since 1751.	CDIAC (Carbon Dioxide Information Analysis Center)
femaleemployrate	2007 female employees age 15+ (% of population). Percentage of female population, age above 15 that has been employed during the given year.	International Labour Organization
hivrate	2009 estimated HIV Prevalence % - (Ages 15-49). Estimated number of people living with HIV per 100 population of age group 15-49.	UNAIDS online database
internetuserate	2010 Internet users (per 100 people) Internet users are people with access to the worldwide network.	World Bank
lifeexpectancy	2011 life expectancy at birth (years). The average number of years a newborn child would live if current mortality patterns were to stay the same.	<ol style="list-style-type: none"> 1. Human Mortality Database, 2. World Population Prospects: 3. Publications and files by history prof. James C Riley 4. Human Lifetable Database
oilperperson	2010 oil Consumption per capita (tonnes per year and person)	BP
polityscore	2009 Democracy score (Polity). Overall polity score from the Polity IV dataset, calculated by subtracting an autocracy score from a democracy score. The summary measure of a country's democratic and free nature. -10 is the lowest value, 10 the highest.	Polity IV Project

relectricperperson	2008 residential electricity consumption, per person (kWh). The amount of residential electricity consumption per person during the given year, counted in kilowatt-hours (kWh).	International Energy Agency
suicideper100th	2005 Suicide, age adjusted, per 100,000. Mortality due to self-inflicted injury, per 100 000 standard population, age adjusted	Combination of time series from WHO Violence and Injury Prevention (VIP) and data from WHO Global Burden of Disease 2002 and 2004.
urbanrate	2008 urban population (% of total). Urban population refers to people living in urban areas as defined by national statistical offices (calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects)	World Bank

After examining the codebook for the Gapminder dataset, I have decided that I am particularly interested in the following parameters and intending to perform regression analysis of those.

- Life Expectancy ("lifeexpectancy")
- Urbanization Rate ("urbanrate")
- Income per Person ("incomeperperson")