

# Writing About the Data

(Regression Modeling in Practice Week 1 Assignment – Step 2)

## Expected Activities

Describe the procedures that were used to collect the data.

- a) Report the study design that generated that data (for example: data reporting, surveys, observation, experiment).

*The data were collected from various sources, so no single study design is used. Below table captures the source for each of the variables in the dataset (as defined in the codebook)*

Variable Name	Description of Indicator	Source
<b>incomeperperson</b>	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
<b>alconsumption</b>	2008 alcohol consumption per adult (age 15+), litres. Recorded and estimated average alcohol consumption, adult (15+) per capita consumption in litres pure alcohol	WHO
<b>armedforcesrate</b>	Armed forces personnel (% of total labor force)	Work Development Indicators
<b>breastcancerper100th</b>	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)
<b>co2emissions</b>	2006 cumulative CO2 emission (metric tons), Total amount of CO2 emission in metric tons since 1751.	CDIAC (Carbon Dioxide Information Analysis Center)
<b>femaleemployrate</b>	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
<b>hivrate</b>	2009 estimated HIV Prevalence % - (Ages 15-49). Estimated number of people living with HIV per 100 population of age group 1549.	UNAIDS online database
<b>internetuserate</b>	2010 Internet users (per 100 people) Internet users are people with access to the worldwide network.	World Bank
<b>lifeexpectancy</b>	2011 life expectancy at birth (years). The average number of years a newborn child	Human Mortality Database, World Population

	would live if current mortality patterns were to stay the same.	Prospects, Publications and files by history prof. James C Riley Human Lifetable Database
<b>oilperperson</b>	2010 oil Consumption per capita (tonnes per year and person)	BP
<b>polityscore</b>	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
<b>relectricperperson</b>	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
<b>suicideper100th</b>	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.
<b>urbanrate</b>	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

b) Describe the original purpose of the data collection.

*Gapminder (<https://www.gapminder.org/>) 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.*

c) Describe how the data were collected.

*The data was collected from various authentic sources like WHO, World Bank, ILO, etc. Explicit source details for each variable is explained in the table provided in response to (a).*

d) Report when the data were collected.

*The data is collected during different time periods. Explicit details for each variable is explained in the table provided in response to (a).*

e) Report where the data were collected.

*The dataset contains information pertaining to all the countries in the world. They are collected from various international agencies who possess the authentic data in that area.*