DESCRIPTION OF THE DATABASE

This database corresponds to the article: Royuela, V. and García, G. (2010)
"Economics and Social Convergence in Colombia" [1]

Data available at: http://riscd2.eco.ub.es/~vroyuela

- The database contains information about the economic and social variables at the departmental level in Colombia for the period 1975 through 2005.

- The departments in the sample are:
  - Arauca (Ara), Antioquia (Ant), Atlántico (Atl), Bolívar (Bol), Boyacá (Boy), Caldas (Cal), Caquetá (Caq), Casanare (Cas), Cauca (Cau), Cesar (Ces), Córdoba (Cór), Cundinamarca (Cun), Chocó (Cho), Huila (Hui), La Guajira (La Gua), Magdalena (Mag), Meta (Met), Nariño (Nar), Norte de Santander (Nors), Putumayo (Put), Quindío (Qui), Risaralda (Ris), Santander (San), Sucre (Suc), Tolima (Tol), Valle (Val),
  - the capital district of Bogotá (Bog),
  - and the group we label as Amazonía Group (GA), formed by Amazonas, Guainía, Guaviare, Vichada and Vaupés.

  Thus, the dataset contains information of 28 spatial units during 31 years.

- First we list the original variables we have used to build our data set, together with the source, description and a link from where we have download the data. These variables are included in Sheet1 and Sheet2 of the excel file.

- Description of the procedure we have used to build the final dataset (Sheet3 and Sheet4):
  - The data sources for build the departmental GDP were DANE and CEGA. The first one only provides homogeneous data of GDP between 1990 to 2005 at disaggregate level for all the 33 departments (including Bogotá), while the CEGA even though provides data of GDP and income since 1975, only includes 23 departments, the capital district of Bogotá, and the nine "New Departments" grouped into a single observation (a total of 25 departments). Finally, CEGA database finishes in 2000.

    Taking into account that departmental results coincide between CEGA and DANE from 1990 onwards (both use the System of National Accounts of 1993, CEGA, 2006), we try to build a consistent series of GDP that considers the heterogeneity of departments. Two procedures have been followed. The first one consisted in using as baseline the data of CEGA, and then using the GDP computed by DANE (from 2000 to 2005) for calculating department growth rates. Subsequently we applied these growth rates to the CEGA database for

updating the series until to 2005. The second procedure consisted in assign values of GDP to each of the nine "New Departments". We used the data computed by DANE of GDP for the period 1990 to 2005 to find the relative position of the new departments, and subsequently we filled the DANE data between 1975 and 1989 maintaining the relative positions between these new departments in 1990 CEGA data. This way we consider a data set ranging from 1975 to 2005 (31 years) for 26 departments, Bogotá and the Amazonía Group (GA) (thus, a total of 28 spatial units).

- For income variable we only used the data of CEGA because it is not supplied by DANE, and consequently it is not possible to enlarge the database for "New Departments". Hence, we prefer excluded to the nine "New Departments" to avoid bias by omission of regional heterogeneity. Consequently, for the income variable we have data of 23 departments and Bogotá for the period 1975 through 2000: 24 units during 26 years.

- Variables illiteracy rate, infant mortality rate and murder rate were defined positively. Then literacy rate is defined as the complement of illiteracy rate, so that measure the percentage of literate population greater than age 15. The infant survival rate is calculated as 1000 minus the infant mortality rate and it measure the number of infants surviving their first year of life over 1000 birth. Finally, the non-murder rate is computed as the complementary measure of the murder rate and it is computed as the amount of people who is not murdered per 10,000 inhabitants.

- Finally, variables were expressed in terms relatives for graph kernels. By relative we mean that the variables are expressed as ratios to the national average of the corresponding year. We additionally take the logarithm of the relative variable, as it facilitates the comparison to the national level. Expressed in logs, a relative value that is equal to 0 means that the department has the same value as the country, while a value that is, for example, equal to -0.1 means that the value of that department is 10% lower than the national value. We used the following formula for each variable: Ln(value_year)-Ln(mean_department).

List of original variables

Sheet1

Name: Population
Source: National Planning Department (DNP)
Description: Population
Link: www.dnp.gov.co

Name: GDP_r (CEGA)
Source: Centro de Estudios Ganaderos y Agrícolas (CEGA)
Description: Real gross domestic product in pesos, constant prices of 1994
Link: http://economia.uniandes.edu.co/investigaciones_y_publicaciones/CEGA

Name: GDP_r (DANE)
Source: National Administrative Department of Statistics (DANE)
Description: Real gross domestic product in pesos, constant prices of 1994
Link: www.dane.gov.co

Name: GDP_pc
Source: DANE and CEGA
Description: GDP per capita = GDP_r (CEGA)/Population

Name: INCOME_pc
Source: Centro de Estudios Ganaderos y Agrícolas (CEGA)
Description: Real per capita household available income in pesos, constant prices of 1994. This variable has data only for the period 1975-2000
Link: http://economia.uniandes.edu.co/investigaciones_y_publicaciones/CEGA

Name: ILLITERACY
Source: DANE
Description: Illiteracy rate that measure the percentage of illiterate population greater than age 15. This variable was taken of Census facts by DANE in the years 1973, 1985, 1993 and 2005.
Link: www.dane.gov.co

Name: MURDER
Source: DANE
Description: Number of homicides. This variable is computed yearly for the period 1990-2005
Link: www.dane.gov.co

Name: MURDER_r
Source: DANE

Description: Murder rate by ten thousand (using DANE and DNP data): (MURDER/Population)x10,000

List of final variables

Sheet2

These variables were considered for the periods 1985-1990, 1990-1995, 1995-2000 and 2000-2005

Name: LEB
Source: DANE
Description: Life expectancy at birth in years
Link: www.dane.gov.co

Name: IMR
Source: DANE
Description: Infant mortality rate by thousand. It is number of deaths infants under one year old per thousand live births
Link: www.dane.gov.co

Sheet3

Name: Population
Source: National Planning Department (DNP)
Description: Population
Link: www.dnp.gov.co

Name: GDP_r (CEGA-DANE)
Source: CEGA-DANE
Description: Real gross domestic product in pesos, constant prices of 1994

Name: GDP_pc
Source: DANE
Description: GDP per capita = GDP_r (CEGA-DANE)/Population

Name: INCOME_pc
Source: Centro de Estudios Ganaderos (CEGA)
Description: Real per capita household available income in pesos, constant prices of 1994. This variable has data only for the period 1975-2000
Link: http://economia.uniandes.edu.co/investigaciones_y_publicaciones/CEGA

Name: LITERACY
Source: DANE
Description: 100-ILLITERACY
Link: www.dane.gov.co

Name: NON-MURDER_r
Source: DANE
Description: 10,000-MURDER_r

List of final variables

Sheet4

These variables were considered for the periods 1985-1990, 1990-1995, 1995-2000 and 2000-2005

Name: LEB
Source: DANE
Description: Life expectancy at birth in years
Link: www.dane.gov.co

Name: ISR
Source: DANE
Description: 1000-IMR: infant survival rate. It is number of survived infants under one year old per thousand deaths births
 Link: www.dane.gov.co