Task & Data Analysis

Name: **The World of(f) Data**
Date: **2006-03-08**
Author: **ChD**
File: **World.xls**
Format: **MS Excel, covertible to tab-delimited *.csv**

Scenario

The Story:
In co-operation with national statistical agencies, the UN is collecting masses of data that attempt to describe the life of people throughout the world. These data cover aspects ranging from physiological and medical indicators to indicators concerned with the economic and cultural profile of the world's nations. For this sonification, a data set was set up that comprises 190 countries (from Afghanistan to Zimbabwe).

The Variables:
For each country, the following variables are reported:

Name of variable	Description	Min	Max
Country	Name of the country	-	-
Region	Global region where the country is located	-	-
	AF...Africa,	-	-
	AS...Asia including Russia,	-	-
	AU...Australia and Oceania,	-	-
	CA...Central America & the Caribbean region	-	-
	EU...Europe,	-	-
	ME...Middle East,	-	-
	NA...North America,	-	-
	SA...South America.	-	-
Area	in km²	2	17075200
Population	Population of the country	7460	1306313812
GDP	Gross national product in $	12200000	11750000000000
GDP/capita	GDP per inhabitant, in $	400,00	58900,00
Capital	Name of the capital	-	-
Latitude	Latitude coordinates of the capital	-	-
Longitude	Longitude coordinates of the capital	-	-
Sanitation	Estimated % of population with access to improved sanitation facilities, 2000:	1,0	100,0
	total, urban, rural.		
Drinking	Estimated % of population with access to improved drinking water sources, 2000:	11,0	100,0
	total, urban, rural.		

Life Expect.	Life expectancy at birth: male, female.	32,5	85,3
School	School Life Expectancy: expected number of years of formal schooling: total, male, female	2,0	23,0
Housing	Average number of persons per room: total, urban, rural.	,5	3,3
Unempl.	Unemployment rate. The percentage of the total labor force that is unemployed but actively seeking employment and willing to work. both, male, female.	1,0	39,0
Marriage	Singulate mean age at Marriage: male, female	18,0	35,0
Contraceptives	Contraceptive use, married women.	4,0	86,0
Fertility	Total fertility rate	,9	7,9
Youth fert.	Births per 1000 women aged 15-19	3,0	271,0
HIV/AIDS:	Estimated number living with HIV/AIDS, end 2003; adults, i. e. persons aged 15-49.	,0	5300,0
HIV women	Estimated women's share (in %) of HIV/AIDS.	,0	65,0
Illiteracy	illiteracy rate (in %), population aged 15- 24; male and fermale.	,1	86,0
Employees	Share (in %) of economic active population who are employees male, female.	5,0	100,0
Employers	Share (in %) of economic active population who are employers: male, female.	,0	20,0
Own-acc.	Share (in %) of economic active population who are own-account workers: male, female.	1,0	93,0
Family w.	Share (in %) of economic active population who are contributing family workers: male, female.	0	73,0
Density	Density of Population: Population per Area	1,8	16204,5

Challenges:

Sonifying these social data offers some interesting challenges. The first challenge lies within the fact that, as with most sets of aggregate social data, this set has a certain amount of missing values. The question is how to cope with these missing data within the sonification.

The second challenge is to choose the variables that shall be included in the sonification. There should be a rationale why specific variables are chosen. This is aggravated by the fact that many of these variables are not independent from others. Some correlate: the GDP per capita, for instance, is positively related with the school life expectancy - the more money is available, the longer individuals go to school. Some may be comparable as they both seem to reflect a more universal (cultural) factor

not directly depicted by the data, like e. g. gender differences. Others, like population and GDP per capita, are related by mathematical reasons that do not necessarily deliver substantial insights into the structure of societies. These points should be considered when variables are chosen. Further, it might be sensible to make preliminary theoretical assumptions about the nature and direction of the relation between the chosen indicators in order to have a rationale for including it. The less trivial these assumptions, the more interesting the sonification.

Third, the data set can be approached from several points. It offers several variables that can be used to construct an order that can be mapped on the time dimension (presumably the core dimension of the sonification). One approach is to put the countries in a geographical order, using the coordinates of the capitals given for each country. Another is to leave the geographical distribution aside and order the countries along one of the other variables, like e.g. the GDP per capita. As this would lead to statements in the form of: "The richer/poorer a country, the more/less...", it is clear that the choice of this starting point is of crucial importance for the extent to which the sonification is able to deliver the wanted insights. (As the data set is non-historic in the sense that the data themselves only report the actual status, these two possibilities seem to be the main ways to approach these data. As data values we always took the most recent figures available.)

The Keys:

Question:	EITHER: Formulate a question you are interested in, and decide how to display it. OR: Answer the questions given below.
Answers:	- -
Subject:	Social data
Sounds:	- -

Questions:

Remark: Each of the questions comprises two parts. The first part a.) is concerned with the data, the second b.) with the interpretation. As the interpretation is in principle up to experts, it can be skipped. But, however, the interesting aspects of the investigative path lined out here is essentially based on the aspects of interpretation.

1. Fertility rate:
    a. The fertility rate is under public discussion for about 15 years now: but are there really mentionable differences? Where is it high, and where low?
    b. Can you find a pattern and make a hypothesis about the fertility rat in this world?
2. Fertility rate & Life expectancy
    a. Do fertility and the life expectancy of females co-vary?
    b. How can the relation between these two variables be interpreted?
3. Fertility rate & Life expectancy female & Regions
    a. Can you find regional differences in this relationship?
    b. How can these relations be interpreted? What are possible reasons for the existence / inexistence of such differences?
4. An assumption might be that the more children a women has, the more her body is exhausted. The independent variable here is fertility rate, and the life expectancy becomes the dependent. How could this assumption be refuted?