## 

ICT702 Task 2

Exercises

2018 Semester 1 SouthBank, Data Wrangling

# Theme: Child Mortality

Your task is to investigate mortality levels of children, infants and babies around the world. Which countries have a high rate of child deaths and which countries have low rates? Are there connections between child mortality rates and the income levels of the country, or the region where they are situated? Are the child mortality rates improving in recent years or getting worse? Reduction of child mortality was Goal 4 of the **United Nations' Millennium Development Goals** (2000-2015):

* <http://www.un.org/millenniumgoals/>
* <http://www.un.org/millenniumgoals/childhealth.shtml>

The data you will need is available from **WHO:** **The World Health Organisation**.

The two files you need to download are available from these web pages (or can be downloaded from the ICT702 Blackboard page):

Figure 1. <http://www.un.org/millenniumgoals/bkgd.shtml>

* child mortality data from: <http://apps.who.int/gho/data/view.main.182?lang=en>. (Download COMPLETE Data as CSV Dataset).
* country metadata from: <http://apps.who.int/gho/data/node.metadata.COUNTRY?lang=en>. (To get all the data you need, you must download the JSON version of this file, or the CSV XMART version).

## Learning Objectives

In this task you will learn how to:

* Apply relevant Python programming concepts to a data analysis challenge
* Read data from real sources and wrangle it into the form you need.
* Develop creative approaches to solving the wrangling/analysis problems.
* Adhere to the recommended Python programming styles
* Write programs that produce correct and useful output
* Organise and present a data analysis report
* Give an insightful analysis of the given problem.

Task 2 is broken into two parts (each worth 20% of the course). *Note: The following deadlines are one week later than the course outline, because this semester we missed the Week 5 lecture and workshop.*

1. Due Week 10. Use Python to read and analyse the child mortality data and generate various useful graphs that give insight into the trends.
2. Due Week 13. Use Python to combine the child mortality data and the country metadata, to give higher-level analyses of child mortality in relation to income grouping and regions of the world.

# Child Mortality

In this first part of Task 2, you should write a Python script that reads and analyses the child mortality data file (WHOSIS\_MDG\_000003.csv) and produce at least FIVE useful graphs that give insight into the data trends.

You should produce the following graphs:

1. graph the infant mortality rates of all countries in 1990, sorted from lowest rate to highest rate. (If the graph is too crowded, you can show just each 10th country, or show the first 10 and last 10 countries).
2. graph the infant mortality rates as above, but for 2015.
3. choose three representative countries in different areas of the world, then create a line graph showing the *trends* in infant mortality in those countries over the years from 1990 to 2015.
4. for those same three countries, graph the *reduction* in infant mortality rates over the 1990/2015 period - that is, the 1990 rate divided by the 2015 rate.
5. a graph of your choice... (explain why you chose that graph, and what conclusions you draw from the graph).

Hints:

1. Some of the columns contain multiple values (a mortality rate, plus a confidence interval), so you will need to split these up into separate columns.
2. You can either use standard Python data structures to store and manipulate the data, or use the Pandas library if you prefer.
3. Use markup and headings to break your Jupyter notebook into sections and give commentary about what you doing, and discussion of your results. This Jupyter notebook will be what you submit.

# Child Mortality and Country Types

In this second part of Task 2, you should write another Python script that reads and analyses the country metadata (COUNTRY.json) and merges it with the child mortality data from Part 1, to allow you to do some higher-level analysis of **infant mortality** trends.

Your report should include at least two graphs that display or compare **infant mortality** in different *regions* of the world (using the 'WHO\_REGION' string to group the countries).

You report should include at least two graphs that compare infant mortality across different *income groupings* (using the 'WORLD\_BANK\_INCOME\_GROUP' string to classify the countries).

Hints:

1. You can use the 'json' library to read the .json file. The resulting object is quite deeply nested, so you will need to explore which substructures contain the data that you want, and then extract that substructure into a dictionary or list that is easier to use. Or write a function that extracts the data that you need.
2. You can either use standard Python data structures to store and manipulate the data, or use the Pandas library if you prefer.
3. Use markup and headings to break your Jupyter notebook into sections and give commentary about what you doing, and discussion of your results. This Jupyter notebook will be what you submit.

# Marking Criteria

Your submission for each part of this task will be assessed according to the following criteria:

* Presentation and organisation of your report [30%]
* Data analysis and program output [30%]
  + Correctness of the program output and graphs
  + Insightful analysis and discussion of the data
* Programming style [40%]
  + Creative approaches to solving the problem
  + Good use of relevant programming concepts, especially good use of functions
  + Good use of appropriate Python data structures, such as lists and dictionaries
  + Adherence to the recommended Python programming styles (indentation, documentation, naming conventions, etc.)