Mini-Project #1: World Energy Use Data Reduction and Analysis

Computer tools like MATLAB can greatly simplify the work of processing massive amounts of data. Many experiments or design projects generate large data sets, and writing a computer program to handle the data can *greatly* reduce the amount of repetitious and tedious things you have to do by hand.

In this project, you will be given a set of data from the World Bank Organization regarding the use of energy derived from crude oil, for 108 countries. Fossil fuels are a hot-topic in the media, but it is difficult to produce a reasonable argument without understanding the data. As scientist and engineers we need to base our design decisions on quantifiable measures.

This document has three sections to describe the activities of this project. The first section describes how the data were acquired, the second section describes what you must do with the data, and the third section describes the deliverables for this project.

1. HOW THE DATA WERE ACQUIRED

The World Bank is a vital source of financial and technical assistance to developing countries around the world. They have 2 goals for the world to achieve by 2030:

- end extreme poverty by decreasing the percentage of people living on less than $1.25 per day to no more than 3%;
- promote shared prosperity by fostering the income growth of the bottom 40% for every country.

The website for The World Bank (www.worldbank.org) contains data for a plethora of measurement factors used to assess the state of most countries: population, income, energy use, waste production, etc.

The data you will be using for this project have been taken directly from the World Bank database. You will be analyzing the data associated with country population and energy use in the form of oil. There are 108 countries listed; I removed any country which did not have a complete list of data between 1971 and 2011.

2. WHAT YOU NEED TO DO WITH DATA

This section describes the tasks you will do for this project. You will plot raw data, compute total oil use for each country, determine the combined yearly oil use of these countries, and compute the average increase in oil usage from year to year.

The descriptions are organized as milestones for each project day in class. Of course you may work faster than what is suggested here, but to keep from falling too far behind, you should try to keep up with this schedule.

Day 1 – Download the data and plot raw oil usage

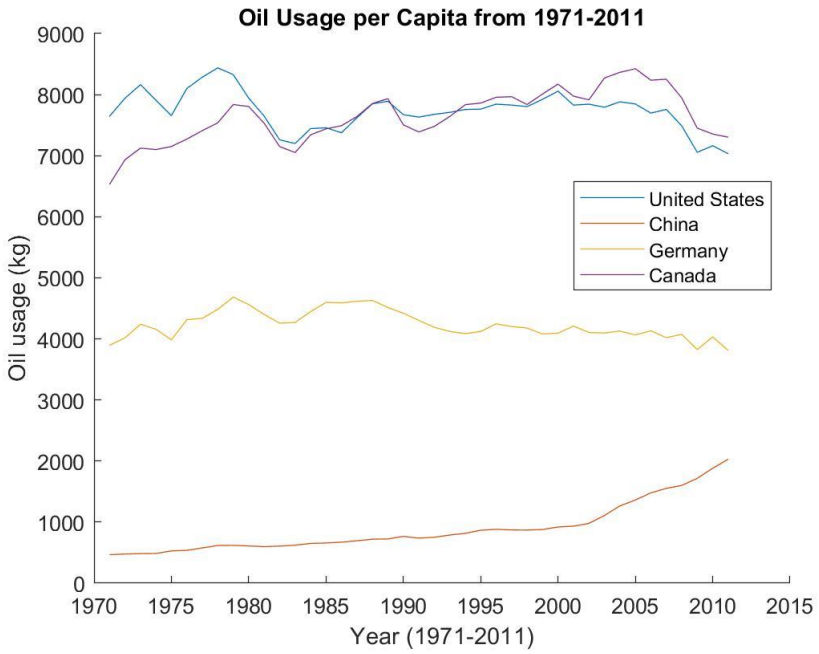
Download the three data files from the course web site: Country_Names.xlsx; Per_Capita_Data.xlsx; Population_Data.xlsx. You should right-click on each file and save it to your Project 1 folder.

As a first step we will investigate the historical oil usage of 4 countries: USA, China, Germany, and Canada. Open the file Country_Names.xlsx in Excel and determine the row index for these for countries. You can add them to the table below for quick reference:

Country	Index number
United States	
China	
Germany	
Canada	

Now load the file Per_Capita_Data.xlsx into Matlab. This matrix has 41 columns – each column corresponds to the amount of oil used per capita for the years 1971 to 2011; the units are in kilograms. The matrix has 108 rows corresponding to the countries listed in Country_Names.xlsx.

Plot the per capita oil usage vs. year for the 4 countries listed above (hint: the data is on the rows you wrote down on the previous page). You will need to extract this data from the main matrix, and you will need to generate a time vector (*i.e.* a list of years from 1971-2011) for your plot command. Plot all 4 curves on the same figure being sure to use axis labels (with units), a legend, and a title. Your plot should look similar to the one shown below (note that you can click on the legend to drag it to a more convenient location).



Day 2 – Compute total oil usage

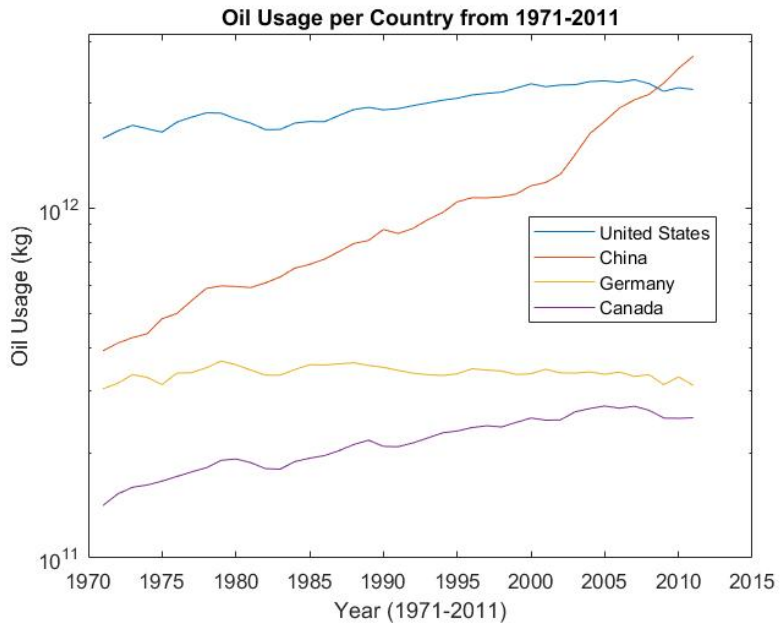
The plots from Day 1 show the per capita oil usage, but these plots may be deceiving. China has a much larger population than Canada; what is the total amount of oil consumed in each country?

To compute the total volume of oil used yearly in each country we must multiply the per capita oil usage with the country's population for each separate year of data. Import the population for each country from the file Population_Data.xlsx. The population data matrix has 41 columns – each column corresponds to the yearly population. The matrix has 108 rows corresponding to the countries listed in Country_Names.xlsx. The mathematical representation of this computation is shown below:

$$m^{\text{oil}}(t_j) = \text{m}^{\text{per capita}}(t_j) \cdot P(t_j)$$

where $m^{\text{oil}}(t_j)$ is the mass of oil used by a given country during some year t_j, $m^{\text{per capita}}(t_j)$ is the per capita oil used by that country in year t_j, and $P(t_j)$ is the population of that country in year t_j.

Create a new figure to plot the total oil usage vs. year for the 4 countries listed above; this can be done using the `figure` command (type `help figure` to learn how it works). Since the values are so large, it would be useful to plot the oil usage using a log scale. This can be done using the `semilogy` command (type `doc semilogy` to learn how it works). Plot all 4 curves on the same figure being sure to use axis labels (with units), a legend, and a title. Your plot should look similar to the one shown below.



Look at the magnitudes of oil usage on this new plot; that is a lot of oil! Let us compute the total amount of oil used by all 108 countries combined on a yearly basis: create a new vector which stores this information. The mathematical equivalent of this is shown below:

$$m^{\text{world oil}}(t_j) = \sum_{i=1}^{108} m_i^{\text{per capita}}(t_j) \cdot P_i(t_j)$$

where $m^{\text{world oil}}(t_j)$ is the total mass of oil used by all 108 countries during some year t_j, $m_i^{\text{per capita}}(t_j)$ is the per capita oil used by a single country i in year t_j, and $P_i(t_j)$ is the population of country i in year t_j.

Create a third figure and plot this total mass of oil vs. year being sure to use axis labels (with units) and a title (note: you'll add to this plot in your Day 3 work, so there will be an example plot for you to check against in that section).

Day 3 – Compute a best-fit line for your data

The plot of world oil use vs. year is surprisingly linear. To get a sense for how much more oil we are using from year to year, let's compute a best-fit linear relationship for Figure 3 to determine the slope. The best-fit line will have the form:

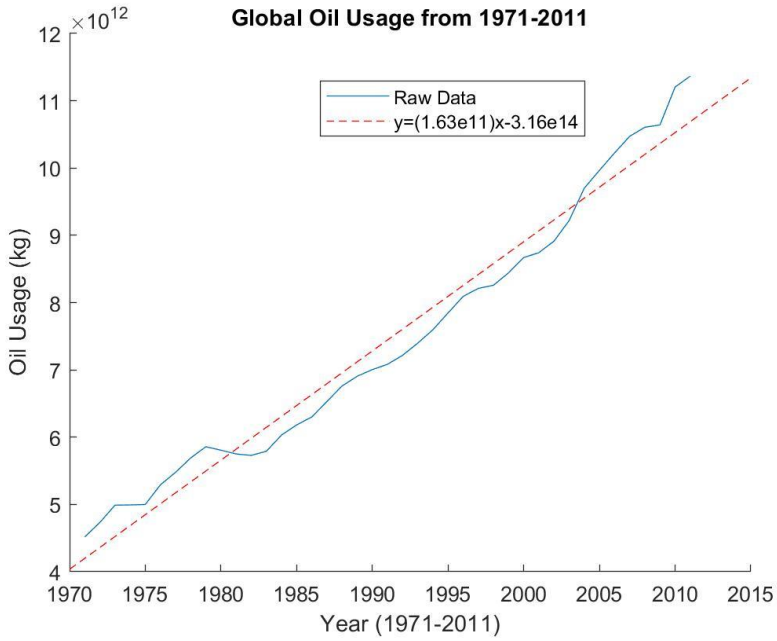
$$y = mx + b$$

where m is the slope and b is the intercept value. We can compute the numerical value of m and b using basic statistics and some calculus. The resulting equations are given below:

$$m = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad b = \frac{\sum_{i=1}^{n} y_i}{n} - m\frac{\sum_{i=1}^{n} x_i}{n}$$

where x_i corresponds to the x-coordinate of the i^{th} data point, y_i corresponds to the y-coordinate of the i^{th} data point, and n is the total number of points (*i.e.* 41 for the span of 1971-2011). Plot the best-fit line on Figure 3 (*i.e.* with the raw data) using a red dashed line and add an appropriate legend including the equation for the best-fit line. You may find the use of the `hold`

command to be useful (type `doc hold` to learn how it works). Your plot should look similar to the one shown below.



Day 4 – Qualify the oil usage and write memo

According to Figure 3, we can see that approximately 11 trillion kilograms of oil were used by the 108 countries in 2011. Let's see if we can qualify how much oil that is to add some perspective.

If we were to pour all of the oil used in 2011 onto the state of Indiana, assuming that the oil doesn't flow across the state lines or into the ground, how deep do you think the oil would be?

The state of Indiana has an area of 94,321 km^2 and oil has a density of approximately 920 kg/m^3. Compute the depth of the oil and include a statement in your memo explaining your result.

3. DELIVERABLES FOR THIS PROJECT

The following deliverables must be included in your electronic submission (please submit items 1-4 as part of a single PDF):

1. The oil use per capita plot for the 4 countries from Day 1
2. The total oil use plot for the 4 countries from Day 2
3. The total combined oil use plot, including the best-fit line, from Days 2 and 3
4. A statement explaining how deep the oil would be if poured onto Indiana from Day 4.
5. Your MATLAB code with good commenting throughout—that is, a good header section, definitions and units for all variables, good section labelling

You will submit your PDF with Deliverables 1-4 and a working copy of your Matlab code as a .m file (Deliverable 5), to the Moodle drop-box by Friday April 10rd, 2020 at 5 PM EDT.