CARNEGIE MELLON UNIVERSITY
DATA, INFERENCE & APPLIED MACHINE LEARNING (COURSE 18-785)
ASSIGNMENT 2

You should submit, using Canvas, a report in the form of a PDF document (Student_ID-Name-DIAMLassignment2.pdf). Include a cover-sheet on the assignment with your name and the required details. Number the pages, graphs, tables and answers carefully to correspond with the questions. Each answer should be supported by Matlab code, graphs and calculations. The submission deadline is 23h59 Rwandan Time (CAT) on Monday 16th September 2019.

No.	Question	Format	Value
1	Using the World Bank Indicators, download data for "GDP per capita (current US$)" and "Malnutrition prevalence, weight for age (% of children under 5)." What kind of relationship do you expect? Make a scatter plot of malnutrition against GDP per capita (using all available years and countries). What kind of relationship do you see? Make a graph for developing regions (six types and use one color for each). Make a graph for income levels (four types and one color for each). Carefully label all graphs and provide legends.	Three Graphs	20%
2	Using Quandl, download data, synchronize the time stamps and plot time series for the prices of Wheat, Crude Oil and Gold in $ on the same graph. Indicate the maximum and minimum prices in all three time series using coloured dots. Use a legend to explain each one.	Graph	20%
3	Download "CO2 emissions (metric tons per capita)" data from the World Bank Indicators. Select the emissions for all countries in 2010 and calculate summary statistics. Provide a table giving the mean, median, standard deviation, 5, 25, 75, and 95 percentiles. Repeat the same process for "School enrolment, primary (% net)."	Two Tables	20%
4	The World Bank Indicators provide variables called "Fertility rate, total (births per woman)" and "GDP per capita (current US$)". Make a scatter plot of Fertility rate versus GDP per capita for all countries in 2010. Produce cumulative distribution functions for the fertility rate variable using data from 1990 and 2010 respectively. Use vertical lines to indicate the mean and median. Use a legend to explain which is the mean and which is the median. Have fertility rates changed over this twenty-year period?	Two graphs	20%
5	Download the latest data for the "Happy Planet Index" from http://www.happyplanetindex.org for year 2016, and "Corruption Perceptions Index" from https://www.transparency.org/news/feature/corruption_perceptions_index_2016 . Both datasets are available as excel spread-sheets. Find matching countries for both indices and make a carefully labelled scatter plot to demonstrate the relationship using ranks in both cases. Are there any countries that stand out as being unusual?	Graph	20%