General Instructions

The purpose of this group project is to analyze a data set, using Python to process the data and to generate graphs, figures, and visualizations as necessary, with the ultimate goal of producing a single written report which presents your results and explains their significance.

You may work in groups of two or three for this project. Each week, you should submit the code that you wrote (there may be multiple .py files generated), a written report of your results including the graphs that your code generated, and a small statement of each group member’s contribution to that week’s results. The deadline for turning in all files is 11:59PM on the due date.

The report should include the following:

1. Overview: A description of what this section of the analysis is about.
2. Methodology: How you performed the analysis.
3. Results: The results that your analysis produced.
4. Discussion: Interpreting the results.

The above is a general outline of what you should present, but you are not required to “number” or outline your report using these terms. Just be sure that it is clear that you are providing the required parts.

It should also be noted that there may be multiple ways to interpret the given prompts. The purpose of this assignment is to explore the possibilities of what you can do. Do not be concerned with trying to figure out a single “right” answer. The prompt is a suggested starting point for investigation. You are not being graded on whether or not you come up with a specific answer, but rather how correct your methodology is and on the quality of execution.

For all written work (i.e., the reports, not the .py files), you may use any word processor that you like, but please make sure to export the report as a PDF, so that there will not be any software compatibility issues in grading.

This project will make use of the World Bank Nutrition and Population data set, which is available via a link on the course website.
Due: April 1, 2019

Data for a single year

Choose a single metric that can be compared across multiple countries, and visualize that data in several ways. You may choose to compare multiple countries or groupings of countries. The intent is that you may demonstrate a multi-faceted perspective of a particular metric.

Longitudinal study

Choose several metrics and visualize them across multiple years. You may choose to use statistics for related individual countries, or by groupings provided by the dataset.

Add some historical insight to the analysis (e.g., on the graph, show that in year XXXX a governmental change happened or some world event that may be correlated with a change in the metric values).

Due: April 23, 2019

SQL Queries

Using the hnp.stats.sqlite3 SQLite database, write SQL queries to answer the following questions. You can submit all of the queries in a single document. Each question must be answered with a single query. When asking questions about countries, do not include the country groupings (unless specifically instructed). Give the answer as well as the query used to generate the answer. Remember that you can use aggregate functions and subqueries.

1. How many countries are in the data set?
2. How many currencies are in the data set?
3. How many countries use the U.S. Dollar as their currency?
4. How many countries use the Euro as their currency?
5. List all the currencies that are used by more than one country and a count of how many countries use that currency. Order them by the number of countries that use that currency, with the largest number on top.
6. List the top 10 currencies by the total population (of all countries that use that currency) in 2016 in descending order by total population. (e.g., add up the total population for all countries that use the Euro, etc.)
7. What are the different “country groupings”? Do not include the list here. Only give the query used to generate a list.

8. How many country groupings are there?

9. Which 5 countries in the “High income” group had the largest population growth percentage in 2016? Include the country and their population growth percentage.

10. Which 5 countries in the “High income” group had the smallest population growth in 2016? Include the country and their population growth percentage.

11. What is the number of hospital beds (per 1,000) for each country for the years 1960, 1970, 1980, 1990, 2000, and 2010? Do not include the list here, as it is quite large. Only give the query used to generate a list.

12. Of the countries for which this data is available, in 2016, calculate the total number of unemployed males, the total number of unemployed females in each Region, as well as their respective percentage of the whole.

13. In 2015, which 5 countries had the highest incidence of HIV?

14. In 2015, which 10 countries had the highest mortality rate (under-5)? Include their rural population (%) and Unemployment (%).

15. Repeat the above query, but this time for the 10 countries with the lowest mortality rate.

16. In each Income group, what is the highest life expectancy at birth (total) in 2016, and what country had that value?