Module 2 Assignment Questions

Note that the answers to each of these questions should be the direct result of running appropriate commands and not involve any further processing, including manual work. Answers without the commands used to achieve them will not get any grade.

1. The Exercise 1 Dataset (located in your assignment prompt in Blackboard) contains a portion of the data from NYC about causes of death for the year 2010.
   Dataset format: CSV
   Field names (in order): Year, Ethnicity, Sex, Cause of Death, Death Count.
   Answer the following questions from this data using UNIX commands.
   - 1.1: How many male record groups and how many female record groups does the data have? **(8 points)**
   - 1.2: How many white female groups are there? Copy entire records of females to a new text file where the records are organized by death count in descending order. **(8 points)**
   - 1.3: List all causes of death by their frequencies in descending order. What are the three most frequent causes of death for black males, and five least frequent causes of death for hispanic females? **(10 points)**

2. Obtained from UNICEF, the Exercise 2 Dataset (located in your assignment prompt in Blackboard) contains data related to the population of 70+ countries for the year 2017.
   Dataset format: CSV
   Field names (in order): Country, Population, Urban Population, Percentage of Urban Population.
   Answer the following questions from this data using UNIX commands:
   - 2.1: Which country has the lowest percentage of urban population? **(8 points)**
   - 2.2: List the countries where the urban population is more than 10 million and yet they comprise less than half of the population. **(10 points)**

3. For the following exercise, use the Exercise 3 Dataset (located in your assignment prompt in Blackboard), which contains availability of essential medicines in 38 countries for the years 2007 - 2013, obtained from the World Health Organization (WHO).
   Dataset format: CSV
   Field names (in order): Country, Median availability of selected generic medicines (%) - Private, Median availability of selected generic medicines (%) - Public
   - 3.1: Which country had the lowest percentage median availability of selected generic medicines in private. **(8 points)**
   - 3.2: List the top five countries with the highest percentage of public and private median availability of selected medicines in 2007–2013. **(8 points)**

- 3.3: List the top three countries where it is best to rely on the private availability of selected generic medicines than public. Explain your answer with valid reasons. **(10 points)**

4. Write a Python script that assigns the list [25,18,9,13,34,15,22,17,12,37,15] to a variable "age" and uses that information to:
   - 4.1: Determine if a person is in high school or not. Assume that for a person to be in high school, their age should be between 14 and 18, inclusive. **(5 points)**
   - 4.2: From the list, calculate the percentage of people not going to high school. **(5 points)**