MATH 210 Assignment 7

Data Science with Python and pandas

INSTRUCTIONS

◇ Create a new Jupyter notebook, set the kernel to Python 3, present your solutions in the notebook and clearly label the solutions

◇ There are 4 questions and each is worth 3 points for 12 total points

◇ Submit the `.ipynb` file to Connect by **11pm Monday April 4**

◇ You may work on these problems with others but you must write your solutions on your own

QUESTIONS

1. (a) Import the Canadian Chronic Disease Surveillance System dataset from the url:

    http://www.math.ubc.ca/~pwalls/data/a7_diseases.csv

    using `pandas.read_csv`. (The dataset is available from the Government of Canada's Open Data Portal. See this `link` to find the dataset and more information.)
   (b) Apply the `.head()`, `.tail()` and `.info()` methods to get an overview of the data.
   (c) Select the column `Fiscal Year` and apply the `.unique()` method to list the years included in the dataset.
   (d) Select the column `Age Group` and apply the `.unique()` method to list the age groups included in the dataset.
   (e) Select the column `Diseases` and apply the `.unique()` method to list the diseases included in the dataset.

2. Plot the prevalence in 2010 of each disease in the dataset. See:

    http://www.math.ubc.ca/~pwalls/data/a7_q2.png

3. Plot the average number of incidences per year by age group separately for asthma, diabetes and heart failure. See:

    http://www.math.ubc.ca/~pwalls/data/a7_q3.png

    It may be necessary to use the `.reindex()` method to order the age groups properly in the figures.

4. Plot the total incidences of osteoporosis in 2010 by gender. See:

    http://www.math.ubc.ca/~pwalls/data/a7_q4.png