DATA CORE - TEST

Instructions

1. You have 24 hours to complete the test from the start time requested.
2. In your response email please include:
   a. Your answers to the test questions as a word document attachment.
   b. Full analytic software code as an attachment.
   c. State the time it took to take the test.
3. Answers to the questions can be brief and bullets are permissible for answers. However, do support your answer with pertinent data, statistics and visualizations as noted.
4. If you make any assumptions while answering a specific question, please make those clear at the beginning of your answer. You are also welcome to put in logic for what is being done in each step as comments in the code.
5. Copy-paste the output to your answers in the word document as directed.
6. Visualizations can be created with any software/tool as per your choice. Indicate the tool you employed in your answer.

About the Datasets

The dataset 'Test_DataCore.csv' contains line-level data of patient encounters at H + H facilities for a number of years. Each row in the dataset corresponds to a visit at an H + H facility with visit dates and times and the patient's demographic information.

The dataset 'Test_DataCore_VitalStats.csv' contains patient level information from vital statistics. This data corresponds to deaths in NYC in the time period of interest.

Please note that none of these datasets contain any PHI. All the patient and visit information in this dataset is fictional, created solely for the purposes of this test and does not correspond to real patients.

Test questions

Q1: Import the Dataset 'Test_DataCore.csv' into your preferred statistical analysis program and include the code you used to import the dataset in your submitted code. (No *.doc* answer necessary for this question, only the code)

Q2: How many patients are in the dataset?

Q3: How many of the patients died?

Q4: Vital statistics recently released the data on deaths in 2015. For H+H patients, these are available in the file 'Test_DataCore_VitalStats.csv'. Merge the missing dates of death received from vital statistics 'vitalstats' for the patients into the previous dataset. How many patients died in total?

Q5: When was the last visit date for the patients who have died? Create a table which lists number of days from last visit date to date of death. Paste the output table in your answer.

Q6: Create a table with the race distribution of patients in the dataset. Paste the output table in your answer.

Q7: Create a visualization with a software/tool of your choice for age and age by gender distribution of patients in the dataset. Paste the visualizations in your answer. Also, briefly describe the distributions.

Q8: Calculate total days spent at the hospital for each patient, visualize (with a software/tool of your choice), and describe the distribution. Paste the visualizations in your answer.

Q9: A) Investigate inpatient visits (patients who stayed for more than a day at the hospital) for each patient. Output a list of the top 10 patient_ids who spent maximum days in the hospital in the past year (2015). Paste the output table in your answer.
B) How many inpatient and outpatient visits did the previous list of patients have in 2015? Paste the output table in your answer.