

Homework 4

Due 11:55PM March 11, 2016. **Only PDF** will be accepted for the write up. **No scans** of handwritten work will be accepted. For this homework you will be working in groups of two, a group of three may be allowed with approval. All programs will be evaluated on the CSIF. This specification is subject to change.

The National Household Travel Survey (NHTS) is a comprehensive data set on travel and transportation patterns in the United States. The most current update is from 2009. The datasets are available in CSV format online at

<http://nhts.ornl.gov/2009/download/Ascii.zip> and also on the smartsite in Homework/NHTS_2011_Ascii.zip. You will find the users guide online at <http://nhts.ornl.gov/2009/pub/UsersGuideV2.pdf> and also on the smartsite at Homework/NHTS_2011_UsersGuideV2.pdf. You will also find the codebook available online at <http://nhts.ornl.gov/2009/pub/Codebook.pdf> and also on the smartsite in Homework/NHTS_2011_Codebook.pdf very useful for decoding the CSV files.

The U.S. Energy Information Administration (EIA) produces reports of energy consumption and emissions produced across all sectors. Three monthly reports: millions of kWh of electricity produced, millions of metric tons of CO₂ produced by electrical sector, and millions of metric tons of CO₂ produced by transportation sector are available in the Homework directory on the smartsite. They are EIA_MkWh_2015.csv, EIA_CO2_Electric_2015.csv, and EIA_CO2_Transportation_2015.csv.

- 1) Create a database schema for the NHTS and EIA data. Describe your schema in your write up; specifically describe how you expect to be able to update your database with future updates of NHTS and EIA data.
- 2) Write a program to load the NHTS and EIA data into a PostgreSQL that follows your schema. Your program can be written in C++ or python. If you choose to make a C++ program, you must include a makefile. Include a readme file with directions of how to use your program. If you choose to make a python program you must specify which version of python you used.
- 3) Write another program to query your database to calculate the following values, put the results in your write up, some may be best described with a chart instead of raw values:
 - a) Calculate the percent of individuals that travel less than 5 - 100 miles a day for every 5 mile increments (e.g. 5, 10, 15, ..., 95, 100).
 - b) Calculate the average fuel economy of all miles traveled for trips less than specific distances from previous problem. Only consider trips that utilize a household vehicle (VEHID is 1 or larger), use the EPA combined fuel economy (EPATMPG) for the particular vehicle.

- c) Calculate the percent of transportation CO₂ emissions should be attributed to household vehicles for each month of the survey (3/2008 – 04/2009). For this problem assume that the data is an accurate representation of the estimated 117,538,000 households in the US. Assume that 8.887×10^{-3} metric tons CO₂ is produced from each gallon of gasoline combusted.
- d) Plug-in hybrid vehicles have recently become commercially available; these vehicles operate in a purely electric mode, or in a hybrid mode. Assume that every vehicle is a plug-in hybrid that has X miles of all electric range. The first X miles in a given day will be driven all electric and the remainder will be at the fuel economy listed for the particular vehicle. The Energy Efficiency Ratio (EER) is used to estimate the amount of electricity an equivalent electric vehicle will consume. Assuming an EER of 3.0 and 33.1 kWh per gallon of gasoline, you can calculate the equivalent energy efficiency in miles/kWh by multiplying the EPA combined fuel economy by 0.090634441. Calculate the change of CO₂ over the months of the survey if every household vehicle were plug-in hybrids with 20 mile electric range. Calculate for 40 and 60 mile electric ranges as well.
- 4) Extra credit: The Hierarchical Data Format 5 (HDF5) is a file format that makes it possible to manage extremely large and complex data collections. MATLAB's version 7.3 file format is an HDF5 file with a proprietary header. The CSV files provided for this assignment have been consolidated into a single MATLAB v7.3 file (HW4Data.mat) that is available in the resources section of smartsite. Implement your program so that it can load the database from the MATLAB file. You may find the HDF5 command-line tools h5ls and h5dump helpful when trying to figure out the layout of the data.
- 5) Extra credit: Additional queries.
- a) The Nissan Leaf has a range of 84 to 107 miles and the Tesla Model S has a range of 208 to 270 miles. Repeat problem 3d except that all trips less than 84 miles are handled electric and all trips greater than 84 miles are handled conventionally (not just the first 84 miles). Assume that if the trip is handled in a conventional vehicle if the trip is greater than 84 miles. Calculate also for 107, 208, and 270 miles.
- b) Repeat 5a on a monthly basis for year 2014; assume transportation in a particular month is equivalent to those months in the survey. Also assume that all additional electricity will be met with a combination of Nuclear, Wind, and Natural Gas. These will be proportional to the amount of their relative energy production in a particular month. Calculate the change of CO₂ in each of the months.

Some useful tips:

- Each household has only one test day, you will need to weigh the test data by the number of days in that month for several of the queries.
- When loading the tuples into the database, insert them in batches. Inserting one tuple at a time may cause the program to take on the order of hours instead of minutes.
- Test using the subset of the data set first.