Cindy Chao, Dehowe Feng
Professor Nitta
ECS 165A Winter 2016

## Schema:

**Explain the schema and how to update the db with future updates of NHTS and EIA data**

The Schema comprised of 7 relations named after each of the 7 files given. For example, the DAYV2PUB file was turned into a relation DAY( HOUSEID , PERSONID , … ), with all of the attributes being the columns from the CSV files.

The NTHS values were casted as FLOATs, and the non numerical values were transformed into a placeholder value '66666'. The reasoning behind this is that none of the querying done required these non-numerical values. The EIA values were casted as VARCHARs due to the large number of attributes that were strings in the CSV files.

To update the tables, assuming that the new data is stored in exactly the same fashion as the previous CSV files, the loader.py script can be modified to take the new files and insert the tuples into each respective table. For value insertions, we used for loops that iterated through the values of each table and strung together the tuples in bunches of 1000 into a 'sql' variable that executed after 1000 and then restarted again until all the insertions were done.

## Results:

**a) The percentage of individuals that travel less than 5-100 miles a day for 5 mile increments:**

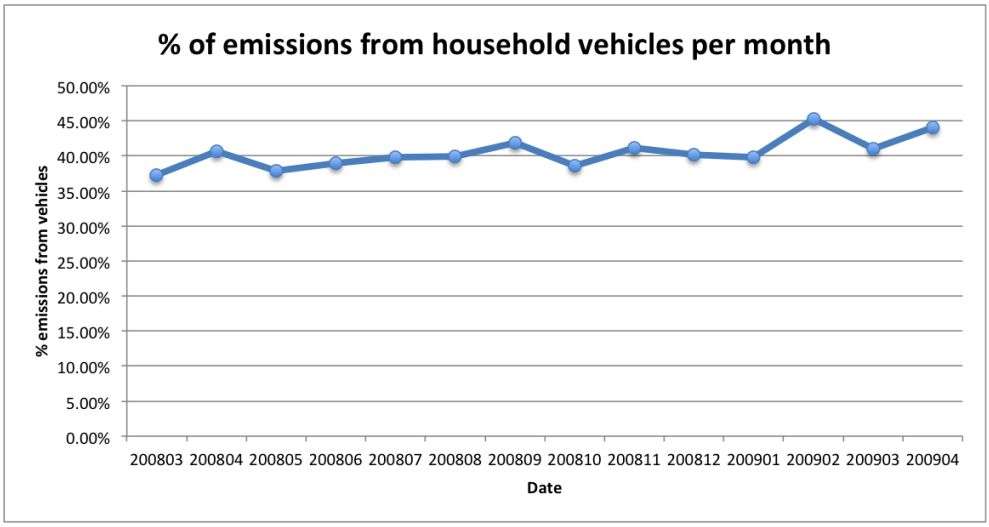
In order to calculate the percentages, the total number of miles a person traveled in their selected travel date was added up by using SUM(TRPMILES) and GROUP BY(HOUSEID, PERSONID, TDAYDATE) to extract each individual. This query was executed in a loop that would increment by 5 each time. The result from the query was then divided by the total number of individuals who travel less than 100 miles a day and multiplied by the number of days in the month.

**b) The average fuel economy of all miles traveled for trips less than 5-100 miles for every 5 mile increments:**

For this query, only the NTHS file was used. The Day trip table and Vehicle table were inner joined on HouseID to get a table of vehicles that traveled on trips. Vehicles with VEHID greater than or equal to 1 (Household Vehicles) were needed for this calculation. Then, all the ETAMPG's for each vehicle per day were averaged out using the AVG command and retrieved using python's `fetchone()` function.

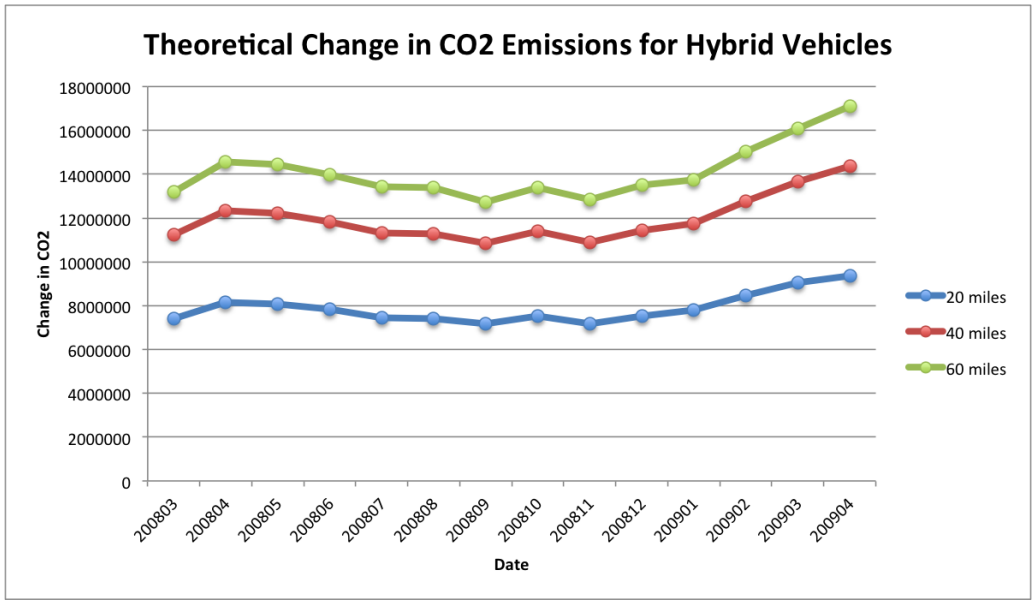
**c) The percent of transportation $CO_2$ emissions from household vehicles for each month of the survey (3/2008 - 4/2009):**

For this query, both the NHTS files and EIA files had to be used. The tables DAY and VEHICLE were joined and for each household vehicle that traveled, the total number of trip miles (TRPMILES) were divided by the fuel economy (EPATMPG) and then converted to metric tons of $CO_2$. Each quotient was then summed up to get the total amount of $CO_2$ emissions from household vehicles for each month and then divided by the total amount of $CO_2$ emissions from the entire transportation sector. This was found by selecting the VALUE where MSN = 'TEACEUS' for the correct travel date. The final percentage was also scaled up according to the number of houses surveyed for that month.




**d) The change of $CO_2$ emissions for each month of the survey if every household vehicle were plug in hybrids with ranges of 20, 40, and 60 miles:**

For each range, the total kWh consumed by the vehicle was calculated by dividing up the vehicles into two groups: ones that traveled under the range and ones that traveled over. For the vehicles that traveled under the range, it could be assumed those miles were driven all electric and so the total miles were divided by the energy efficiency: the fuel economy multiplied by 0.09063441 (the ratio of EER to 33.1 kWh per gallon). This number was then multiplied by conversion ratio found by dividing the TXEIEUS by the ELETPUS values for the specific month. For the vehicles that traveled over the range, the remaining miles over the range were calculated similarly to how $CO_2$ emissions were calculated in problem 3c and then added to the $CO_2$ emissions calculated for the electric portion of the trip. Both groups were then added together and then scaled up depending on the number of houses, found by COUNT(DISTINCT(HOUSEID)) in a separate query. The change was then calculated by subtracting the $CO_2$ emissions of hybrid vehicles from the $CO_2$ emissions of normal vehicles.


Theoretical Change in CO2 Emissions for Hybrid Vehicles
18000000
16000000
14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
Change in CO2
200803
200804
200805
200806
200807
200808
200809
200810
200811
200812
200901
200902
200903
200904
Date
20 miles
40 miles
60 miles

Data Tables:

Query 3A

| Maximum Miles Travelled | % of people |
|---|---|
| 5 | 15.21% |
| 10 | 26.47% |
| 15 | 39.43% |
| 20 | 48.54% |
| 25 | 58.24% |
| 30 | 64.64% |
| 35 | 71.39% |
| 40 | 75.95% |
| 45 | 80.69% |
| 50 | 83.80% |
| 55 | 87.17% |
| 60 | 89.34% |
| 65 | 91.80% |
| 70 | 93.39% |
| 75 | 95.07% |
| 80 | 96.28% |
| 85 | 97.53% |
| 90 | 98.42% |
| 95 | 99.36% |
| 100 | 100.00% |

Query 3B

| Maximum Miles Travelled | EPA Combined Fuel Economy |
|---|---|
| 5 | 26.38624139 |
| 10 | 26.34663983 |
| 15 | 26.33112015 |
| 20 | 26.32426447 |
| 25 | 26.32410026 |
| 30 | 26.3272874 |
| 35 | 26.325997 |
| 40 | 26.32598759 |
| 45 | 26.32591479 |
| 50 | 26.32541467 |
| 55 | 26.32579555 |
| 60 | 26.32672768 |
| 65 | 26.32736446 |
| 70 | 26.32675702 |
| 75 | 26.32625388 |
| 80 | 26.32556612 |
| 85 | 26.32511556 |
| 90 | 26.32473613 |
| 95 | 26.32423579 |
| 100 | 26.32423456 |

Query 3C

| Date | % emissions from vehicles |
|---|---|
| 200803 | 37.21% |
| 200804 | 40.64% |
| 200805 | 37.82% |
| 200806 | 38.92% |
| 200807 | 39.75% |
| 200808 | 39.94% |
| 200809 | 41.87% |
| 200810 | 38.54% |
| 200811 | 41.19% |
| 200812 | 40.14% |
| 200901 | 39.77% |
| 200902 | 45.24% |
| 200903 | 41.01% |
| 200904 | 44.08% |

QUERY 4D

| Date | Miles | CO2 Change | Date | Miles | CO2 Change | Date | Miles | CO2 Change |
|---|---|---|---|---|---|---|---|---|
| 200803 | 20 | 7407595.394 | 200803 | 40 | 11239013.76 | 200803 | 60 | 13209309.56 |
| 200804 | 20 | 8138243.515 | 200804 | 40 | 12336690.24 | 200804 | 60 | 14541421.95 |
| 200805 | 20 | 8058145.948 | 200805 | 40 | 12211565.79 | 200805 | 60 | 14433992.5 |
| 200806 | 20 | 7824797.519 | 200806 | 40 | 11826502.11 | 200806 | 60 | 13965478.28 |
| 200807 | 20 | 7440305.847 | 200807 | 40 | 11312995.21 | 200807 | 60 | 13425309.46 |
| 200808 | 20 | 7398411.248 | 200808 | 40 | 11287508.66 | 200808 | 60 | 13390106.69 |
| 200809 | 20 | 7170186.062 | 200809 | 40 | 10839091.95 | 200809 | 60 | 12727052.31 |
| 200810 | 20 | 7530206.473 | 200810 | 40 | 11401015.39 | 200810 | 60 | 13387067.07 |
| 200811 | 20 | 7178232.226 | 200811 | 40 | 10894419.5 | 200811 | 60 | 12852850.87 |
| 200812 | 20 | 7514920.788 | 200812 | 40 | 11435243.57 | 200812 | 60 | 13512564.31 |
| 200901 | 20 | 7809708.308 | 200901 | 40 | 11740639.93 | 200901 | 60 | 13758310.72 |
| 200902 | 20 | 8452123.5 | 200902 | 40 | 12755023.69 | 200902 | 60 | 15025083.33 |
| 200903 | 20 | 9032382.363 | 200903 | 40 | 13675041.57 | 200903 | 60 | 16086938.89 |
| 200904 | 20 | 9370189.924 | 200904 | 40 | 14363295.16 | 200904 | 60 | 17090377.1 |