KWANIL KIM (994961022)
QIANHUI FAN (912391768)

Write Up

1.
We created a table for each cvs file and add all the attributes of each cvs file to its corresponding table. We've mapped each attribute to it's type in make_datatypes function for each table, if we need to add new columns in the future, we just need to add new columns and it's type to the mapping in corresponding make_datatypes function. If values are updated in EIA or NHTS file, we can't add the same table again when it's already been created, so we need to manually drop the table, and reinsert, then program will load the updated data in the files and insert into corresponding table.

2.
We created a make_datatypes function for each NHTS file and one for all EIA file, since all EIA files have the same attributes. We create a dictionary for CSV files, and use attributes as fieldnames, and use copy_expert function to insert data into table.

3 a)
For this part, we get the total number of people from counting the individuals who has a valid TRPMILES data, and getting the total miles travelled by all individuals in all households, then we can count the number of individuals who's total miles is less than 5, 10, 15 up to 100 miles, divided by the total number of number, giving the percentage. We use TRPMILES instead of VMT_MILES because we not only want the driver miles, but also miles traveled by other transportation.

Result:
Total number of people: 260446
Percentage of people who travel less than 5 miles: 12.62%
Percentage of people who travel less than 10 miles: 23.20%
Percentage of people who travel less than 15 miles: 35.35%
Percentage of people who travel less than 20 miles: 43.88%
Percentage of people who travel less than 25 miles: 52.98%
Percentage of people who travel less than 30 miles: 58.97%
Percentage of people who travel less than 35 miles: 65.30%
Percentage of people who travel less than 40 miles: 69.56%
Percentage of people who travel less than 45 miles: 73.99%
Percentage of people who travel less than 50 miles: 76.91%
Percentage of people who travel less than 55 miles: 80.07%
Percentage of people who travel less than 60 miles: 82.09%
Percentage of people who travel less than 65 miles: 84.39%
Percentage of people who travel less than 70 miles: 85.88%
Percentage of people who travel less than 75 miles: 87.45%
Percentage of people who travel less than 80 miles: 88.58%
Percentage of people who travel less than 85 miles: 89.74%
Percentage of people who travel less than 90 miles: 90.57%
Percentage of people who travel less than 95 miles: 91.45%
Percentage of people who travel less than 100 miles: 92.05%
****

Percentage of People who travel less than X miles
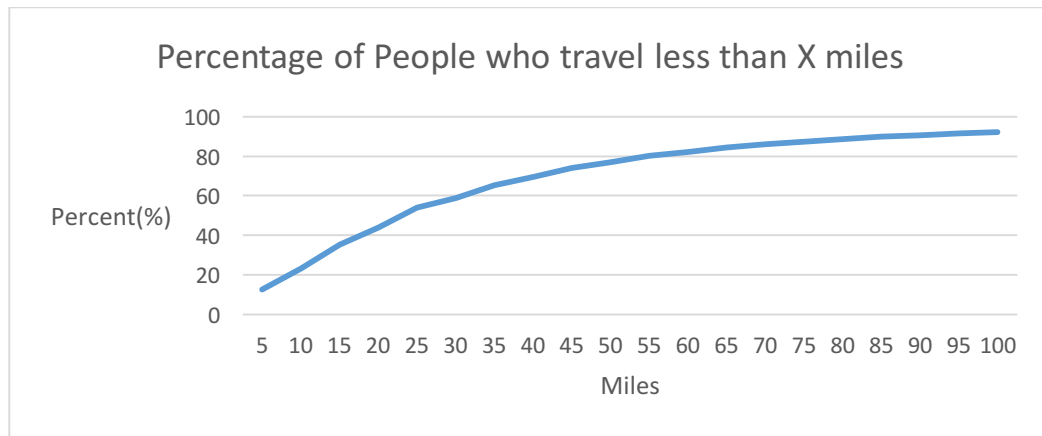
3 b)
We want to use EPATMPG from VEHV2PUB, and TRPMILES from DAYV2PUB, so we inner join the two table on HOUSEID, we didn't use natural join because inner join won't create new table, the resulting table will select all the trips with TRPMILES smaller than a desired value(5, 10, 15…100) and it's corresponding fuel economy, then we just need to use the average function to get the average of EPATMPG value.
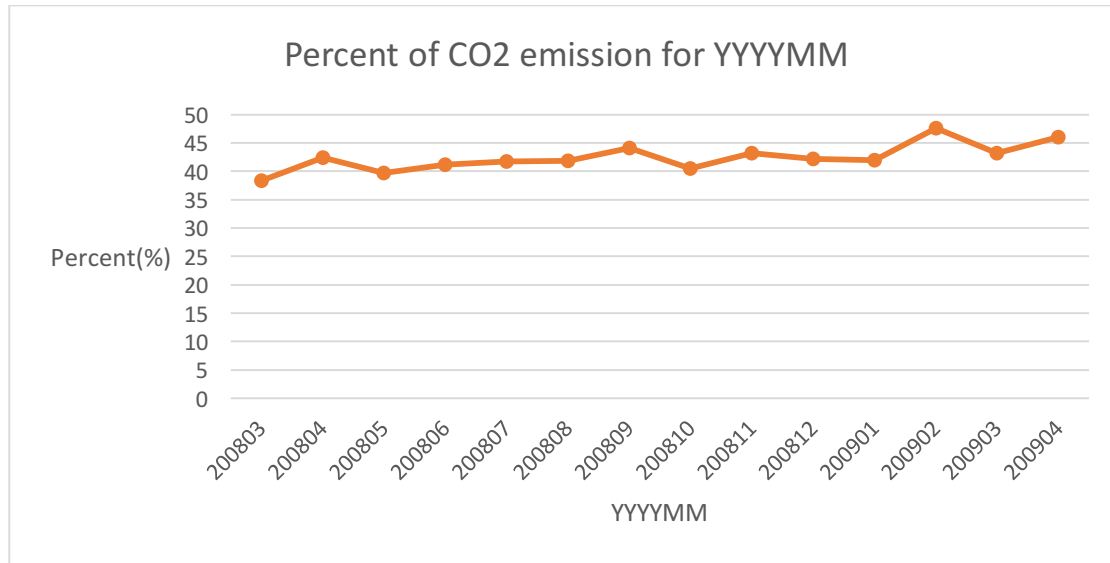
Result:
Average fuel less than 5 miles: 26.39
Average fuel less than 10 miles: 26.35
Average fuel less than 15 miles: 26.33
Average fuel less than 20 miles: 26.32
Average fuel less than 25 miles: 26.32
Average fuel less than 30 miles: 26.33
Average fuel less than 35 miles: 26.33
Average fuel less than 40 miles: 26.33
Average fuel less than 45 miles: 26.33
Average fuel less than 50 miles: 26.33
Average fuel less than 55 miles: 26.33
Average fuel less than 60 miles: 26.33
Average fuel less than 65 miles: 26.33
Average fuel less than 70 miles: 26.33
Average fuel less than 75 miles: 26.33
Average fuel less than 80 miles: 26.33
Average fuel less than 85 miles: 26.33
Average fuel less than 90 miles: 26.33
Average fuel less than 95 miles: 26.32
Average fuel less than 100 miles: 26.32
****

Average fuel for trips less than X miles
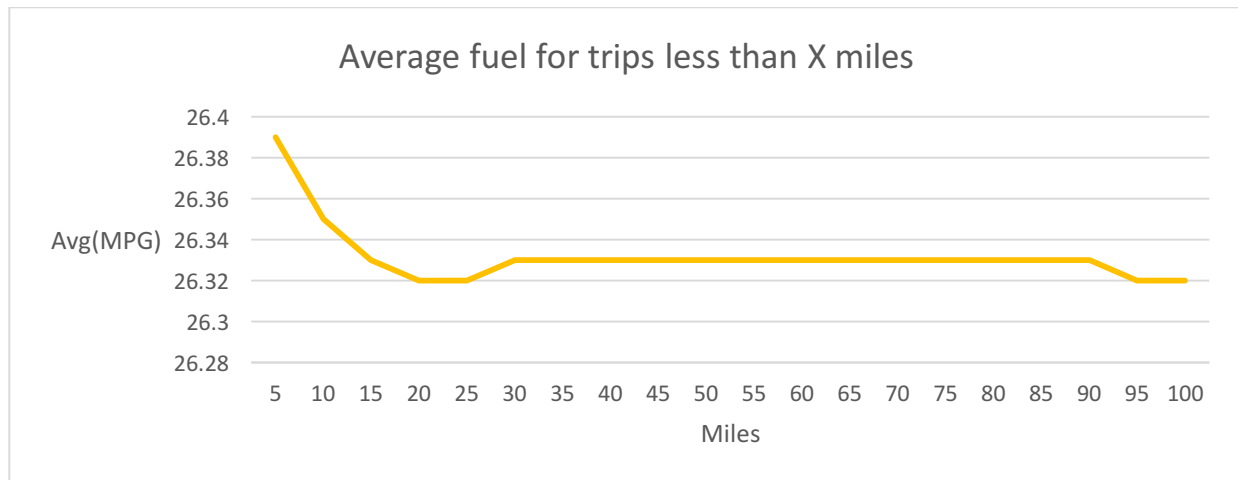
3 c)
We first natural join DAYV2PUB and VEHV2PUB to get the legit value of TRPMILE and EPATMPG we could use, the natural join will join on the HOUSED and VEHID. Then we can iterate through each month, and find sum of TRPMILE / EPATMPG, the gallon of fuel consumed, for each trip, scale up to whole month by multiply 30 and also multiply 8.887*0.001 which give the total amount of CO_2 emission attributed to a household's vehicles in a month, we then have to scale this up to all the household in U.S. by multiplying total household divided by number of distinct household ID, then we compare the value we get to the "Total Energy Transportation Sector CO_2 Emissions" which corresponding YYYYMM from EIA_CO2_Transportation_2015, and get the portion of household vehicle emission.

Percent of CO_2 emission for 200803: 38.34%
Percent of CO_2 emission for 200804: 42.39%
Percent of CO_2 emission for 200805: 39.65%
Percent of CO_2 emission for 200806: 41.15%
Percent of CO_2 emission for 200807: 41.75%
Percent of CO_2 emission for 200808: 41.80%
Percent of CO_2 emission for 200809: 44.11%
Percent of CO_2 emission for 200810: 40.47%
Percent of CO_2 emission for 200811: 43.25%
Percent of CO_2 emission for 200812: 42.19%
Percent of CO_2 emission for 200901: 41.98%
Percent of CO_2 emission for 200902: 47.55%
Percent of CO_2 emission for 200903: 43.21%
Percent of CO_2 emission for 200904: 46.06%
****

Percent of CO2 emission for YYYYMM

Percent(%)

50
45
40
35
30
25
20
15
10
5
0

200803 200804 200805 200806 200807 200808 200809 200810 200811 200812 200901 200902 200903 200904

YYYYMM

3 d)

We first natural join DAYV2PUB and VEHV2PUB to get the legit value of TRPMILES and EPATMPG we could use, the natural join will join on the HOUSED and VEHID. Then for each range we iterate through each month of the survey, and compare a TRPMILE value to the range, if it's smaller than that range, we'll use TRPMILE value divided by energy efficiency of fuel economy in miles/KWh, and get the corresponding KWhs, then we can use this value to get the amount of CO2 emission generated, scale this value to a month. On the other hand, if the TRPMILES value is larger than the range, then we use the range to calculate the CO2 emission for using electricity, and use the remainder miles divided by EPATMPG to get gallon, then covert to CO2 emission and scale up to whole month.

Miles: '20'
200803, Hybrid, Conventional, Percent Change
[(65050759.6103263, 82961567.9407219, 21.5892837791992)]
200804, Hybrid, Conventional, Percent Change
[(71631144.1384628, 90924081.3860625, 21.218732104295)]
200805, Hybrid, Conventional, Percent Change
[(72656012.2460329, 92601913.702655, 21.5394052445486)]
200806, Hybrid, Conventional, Percent Change
[(75093498.1943122, 94631328.615859, 20.6462602896104)]
200807, Hybrid, Conventional, Percent Change
[(82136042.4528045, 100775580.561657, 18.496086060698)]
200808, Hybrid, Conventional, Percent Change
[(81931066.567098, 100520548.257922, 18.4932155792924)]
200809, Hybrid, Conventional, Percent Change
[(70783961.0040368, 87793298.8665457, 19.3743008659063)]
200810, Hybrid, Conventional, Percent Change
[(72189679.7883013, 90280710.287521, 20.0386444032223)]
200811, Hybrid, Conventional, Percent Change
[(74402910.76852, 91921245.1101761, 19.0579819938892)]
200812, Hybrid, Conventional, Percent Change
[(76043666.4073178, 94604710.2242603, 19.6195768402478)]
200901, Hybrid, Conventional, Percent Change

[(69993096.4369462, 89031983.5903342, 21.3843232348863)]
200902, Hybrid, Conventional, Percent Change
[(69273395.2331057, 89550199.856152, 22.6429473698749)]
200903, Hybrid, Conventional, Percent Change
[(72548589.7044436, 94498001.1097834, 23.2273816880423)]
200904, Hybrid, Conventional, Percent Change
[(79457898.8593686, 102825803.505486, 22.7257204412418)]
****

Miles: '40'
200803, Hybrid, Conventional, Percent Change
[(62014929.8080802, 82961567.9407219, 25.2486044473129)]
200804, Hybrid, Conventional, Percent Change
[(68458383.9207088, 90924081.3860625, 24.7081929483177)]
200805, Hybrid, Conventional, Percent Change
[(69293058.8949157, 92601913.702655, 25.1710292754684)]
200806, Hybrid, Conventional, Percent Change
[(71669274.8980077, 94631328.615859, 24.2647483171902)]
200807, Hybrid, Conventional, Percent Change
[(78648384.9231439, 100775580.561657, 21.9569021733145)]
200808, Hybrid, Conventional, Percent Change
[(78462954.2746604, 100520548.257922, 21.9433681625621)]
200809, Hybrid, Conventional, Percent Change
[(68058267.9718105, 87793298.8665457, 22.4789718002674)]
200810, Hybrid, Conventional, Percent Change
[(69195309.4354806, 90280710.287521, 23.3553776713639)]
200811, Hybrid, Conventional, Percent Change
[(71429840.9296859, 91921245.1101761, 22.2923483640037)]
200812, Hybrid, Conventional, Percent Change
[(72851903.4491018, 94604710.2242603, 22.9933654715432)]
200901, Hybrid, Conventional, Percent Change
[(66974164.6095189, 89031983.5903342, 24.775162914836)]
200902, Hybrid, Conventional, Percent Change
[(65951109.2738495, 89550199.856152, 26.3529178273311)]
200903, Hybrid, Conventional, Percent Change
[(68738705.7792635, 94498001.1097834, 27.2590901691074)]
200904, Hybrid, Conventional, Percent Change
[(75100644.8296273, 102825803.505486, 26.9632307559643)]
****

Miles: '60'
200803, Hybrid, Conventional, Percent Change
[(60919984.203215, 82961567.9407219, 26.5684271459963)]
200804, Hybrid, Conventional, Percent Change
[(67237964.0365768, 90924081.3860625, 26.0504334917773)]
200805, Hybrid, Conventional, Percent Change
[(67940562.2475643, 92601913.702655, 26.6315786240427)]
200806, Hybrid, Conventional, Percent Change
[(70165587.202188, 94631328.615859, 25.8537439677994)]
200807, Hybrid, Conventional, Percent Change
[(77098345.6092154, 100775580.561657, 23.4950122048217)]
200808, Hybrid, Conventional, Percent Change

[(76932346.7385379, 100520548.257922, 23.4660494079878)]
200809, Hybrid, Conventional, Percent Change
[(66965357.1387755, 87793298.8665457, 23.723839970327)]
200810, Hybrid, Conventional, Percent Change
[(67964918.066559, 90280710.287521, 24.7182284564354)]
200811, Hybrid, Conventional, Percent Change
[(70119525.162726, 91921245.1101761, 23.7178248851163)]
200812, Hybrid, Conventional, Percent Change
[(71571297.3771239, 94604710.2242603, 24.3470042797401)]
200901, Hybrid, Conventional, Percent Change
[(65764891.2354557, 89031983.5903342, 26.1334089353082)]
200902, Hybrid, Conventional, Percent Change
[(64667835.0921392, 89550199.856152, 27.78593995768)]
200903, Hybrid, Conventional, Percent Change
[(67124185.0354382, 94498001.1097834, 28.9676138678781)]
200904, Hybrid, Conventional, Percent Change
[(73210769.3457179, 102825803.505486, 28.8011696968533)]
****

Percent CO_2 Change Over Month
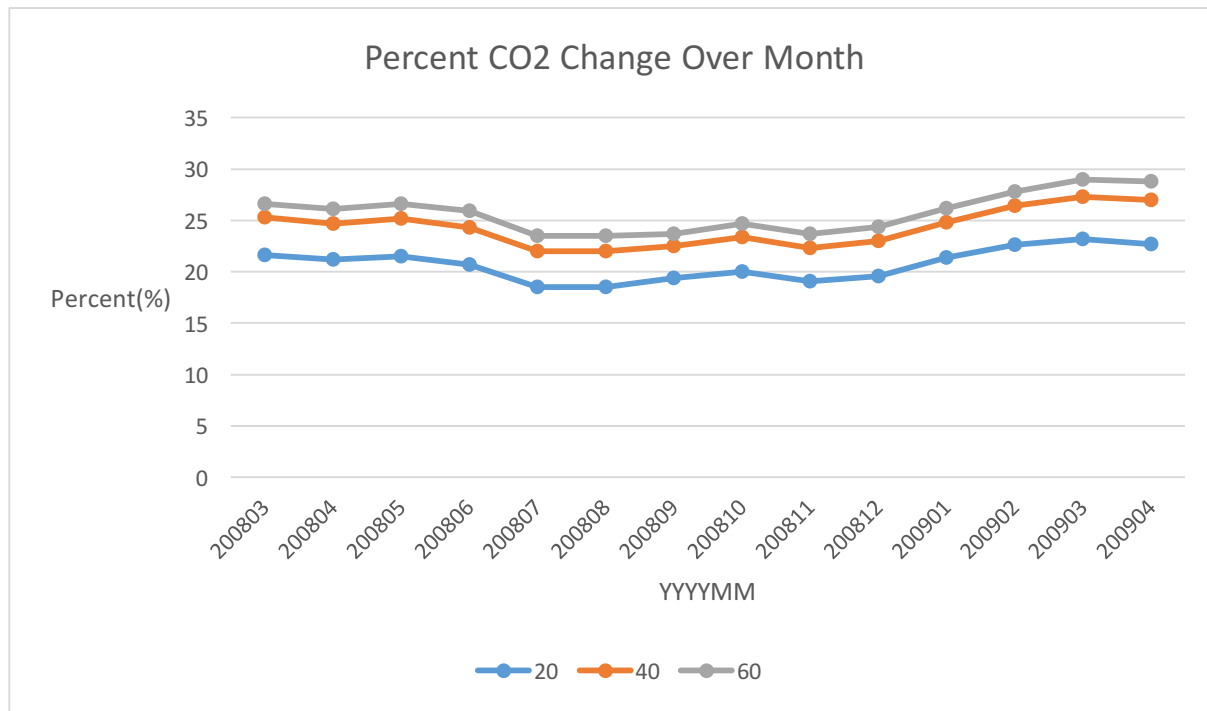
EXTRA CREDIT
4
Works for the EIA files but then NHTS files it reads the row but somehow it goes beyond the existing rows and keeps inserting
5 a)
same as 3d except for miles larger than the range, we'll use the whole TRPMILE to divided by EPATMPG.
b)
We were able to get the natural gas, nuclear, and wind electricity to sum up at one time but it gives negative numbers.